The Unfolding Structure of Arguments in Online Debates:
The case of a No-Deal Brexit

Carlo R. M. A. Santagiustina∗ and Massimo Warglien†
Department of Management, Ca’ Foscari University of Venice, Cannaregio 873, 30123 Venice, Italy
(Dated: March 31, 2021)

In the last decade, political debates have progressively shifted to social media. Rhetorical devices employed by online actors and factions that operate in these debating arenas can be captured and analysed to conduct a statistical reading of societal controversies and their argumentation dynamics. In this paper, we propose a five-step methodology, to extract, categorize and explore the latent argumentation structures of online debates. Using Twitter data about a “no-deal” Brexit, we focus on the expected effects in case of materialisation of this event. First, we extract cause-effect claims contained in tweets using regular expressions that exploit verbs related to Creation, Destruction and Causation. Second, we categorise extracted “no-deal” effects using a Structural Topic Model estimated on unigrams and bigrams. Third, we select controversial effect topics and explore within-topic argumentation differences between self-declared partisan user factions, i.e., Brexiteers and Remainers. We hence type topics using estimated covariate effects on topic propensities, then, using the topics correlation network, we study the aggregate topological structure of the debate to identify coherent topical constellations. Finally, we analyse the debate time dynamics and infer lead/follow relations among factions. Results show that the proposed methodology can be employed to perform a statistical rhetorics analysis of debates, and map the architecture of controversies across time. In particular, the “no-deal” Brexit debate on Twitter is shown to have a multifaceted assortative argumentation structure heavily characterized by factional constellations of arguments, as well as by polarized narrative frames invoked through verbs related to Creation and Destruction. Our findings highlight the benefits of implementing a systemic approach to the analysis of debates, which allows the unveiling of topical and factional dependencies between causal arguments and rhetoric devices employed in online debates.

I. INTRODUCTION

Online debates have become a major component of contemporary democratic life, involving millions of people in the expression of opinions on a vast range of topics[1, 2]. Debates are strictly associated with argumentation to speech acts that express participants’ opinions and try to affect other participants’ views by offering reasons, triggering frames, eliciting emotions[3]. Debates can be finalized to deliberation, as in public assemblies, or more loosely directed to communicate and shape opinions on controversial subjects or issues. Online debates are typically of this second type. Online Social Media (OSM) platforms, like Twitter, Facebook and Reddit, are arenas where these lively debates nowadays take place. These virtual spaces are incessantly used by wide communities of users to gather information, communicate their thoughts and views concerning (realized or possible) events occurring at the national or global scale. Partisan and non-partisan participants to these online debates often state publicly their opinions about these events, their likelihood and expected effects. Therefore, online debates offer a privileged window on the expressed arguments of a large sample of the politically and socially active population.

In this paper, we introduce a methodology for analysing the argumentation patterns and structural properties of online debates and demonstrate its use by analyzing the no-deal Brexit controversy. Our fundamental assumption is that arguments do not come insulated in a debate, nor they can be analyzed assuming that they are independent. They are fundamental components of a debate architecture and have to be understood in their interaction with such structure. Debates are about issues that are discussed in specific communication arenas to persuade an audience, orient opinions or decisions. Issues in debates always incorporate a controversial component, associated with partisan factions that support specific views or resolutions. As debates are meant to be persuasive, they imply arguments and counterarguments. Arguments do not come alone but display different degrees of cohesiveness - they are correlated and assembled in coherent blocks. Such correlations provide structure and composite arguments to the debate. Arguments also have internal structure and are expressed through a multitude of semantic components that offer different nuances and interpretations – these are often related to the faction expressing them. Finally, debates happen in time, and are characterized by interaction dynamics among participant factions, with leading and following relations, agenda-setting tentatives, attacks and defenses, shifts of dominant topics and opinions. These can be related both to the endogenous dynamics of the
debate and to external events that can steer them. Online debates are no exception: they often involve a very large number of participants and offer remarkable opportunities to be observed and analyzed over long periods. To apprehend them, it is important to develop a coherent framework of observations and analysis to reveal how the different architectural components of a debate are composed and linked.

**An integrated roadmap to analysing online debates**

The last decade has been characterised by a sharp rise of interest concerning the use of computational methods for the automated analysis of online debates[4]. This rising interest in the subject has been accompanied by the rapid development of text mining and machine learning methods[5]. Here we develop a hybrid framework to analyze arguments in online debates. The proposed framework builds on front edge literature by combining and extending in innovative ways existing statistical, text mining and network analysis methods. This, to offer a more consistent and systemic characterization and interpretation of debates which existing tools cannot capture, and to tie together the macro-level (argumentative) and micro-level (phrasal) features of debates. In a way, our approach moves the first steps into the statistical rhetoric of (online) debates. In brief, our approach can be summarized in five steps:

1. **Argument extraction.** In this paper we focus on causal arguments. Most conventional approaches[6] rely on Part-Of-Speech recognition, Relation Extraction and exploit (potentially ambiguous) causal connectives to capture causal statements. Instead, we focus only on verbs as causal markers which express unambiguously the semantics of causation, and for which cause-effect relations can be identified and extracted using simple regular expressions (RegEx). Besides being more robust for the type of OSM data employed, this method offers a rich set of possibilities in differentiating types of causal (as well as other modal) arguments.

2. **Aggregating causal arguments in classes (of effects).** The second step consists in aggregating arguments via Structural Topic Modeling (STM) to obtain a limited number of them. In this paper, we focus on the effects which pertain to a single cause/event (i.e. the no-deal). By exploiting the metadata of the tweet and that of the extracted causal relations, we show differences within and between single arguments in terms of factional characterization and types of causal relations employed.

3. **Comparing faction rhetoric and phrases.** We analyze through an innovative method the internal structure of each argument. We transform each topic into an oriented, weighted graph of words and their associations using uni- and bigrams distributions. This provides considerable additional information on how words are used and phrases are constructed inside the topic. For example, this enables us to see how analogous arguments (e.g. the economic effects of the no-deal) are differently articulated by different factions.

4. **Mapping the structure of the debate.** Arguments are correlated. We filter the network of arguments correlations to uncover the ‘building blocks’ of argumentation and characterize their relationships in terms of types of causal verbs, factions, and inhibition/activation relationships among arguments.
Identifying lead and follow faction dynamics.

We look at the time series of arguments proportions to explore their evolution and identify debate leader/follower dynamics among partisan factions taking part in the debate.

The output of each step is represented in Figure 1. We focus on causal relations as a prominent example of argumentation, but our approach can be extended to other types of relation -such as permission, possibility and influence- used in argumentation[7].

The case of a No-Deal Brexit

In this paper, we implement the proposed methodology to analyze the anatomy of the online debate on the “no-deal” hard-Brexit by extracting arguments, uncovering their correlation structure, and analyzing the semantics of different factions’ arguments. The Brexit online debate is an appealing case to investigate with the proposed methodology, because, since the 2016 referendum, the “no-deal” has polarized public opinions and received great media attention, both offline and online. Besides, the attitude towards the “no-deal” can be considered among the most polarizing dimensions of the Brexit debate, and is certainly one of the drivers of the 2019 general UK elections results.

For analysing the “no-deal” Brexit online debate we use Twitter data in English published from February 2019 to May 2020, which are directly referring to the “no-deal” scenario1. The time frame of this work includes the first and second extension granted by the EU, covering the period in which the debate has been more active and intense. By applying our methodology to Twitter posts about the “no-deal”, we can identify and map online arguments about the expected effects of a hard Brexit and to understand in which terms opposing partisan factions, i.e. Brexiteers and Remainers, confront each other and try to influence non-partisan online public through distinct argumentation and persuasion strategies.

II. RESULTS

The no-deal debate on Twitter

As preliminary step, using RegEx conditions we identify self-declared Brexit faction partisans among users who tweeted about a no-deal Brexit2. Whereas, users whose Twitter profile Bio don’t match neither partisan faction (Brexiteer and Remainer) RegEx conditions are considered part of a residual group, called Others. Table V in the Appendix reports the number of tweets and retweets referring to the no-deal by faction. We observe

---

1 See methods Section IV A for details.
2 See methods Section IV B for details.
that both partisan factions represent around 2% of the total volume of activity. Interestingly, Remainers exhibit a higher retweet share (79.7%) with respect to Brexiteers (76.2%) and Others (75.2%).

The dynamics of no-deal tweets counts by day (see Figure 2) allow to single out five stages of activity associated with the phases of the Brexit process. The first stage ends in mid-April 2019, with the approval of the EU first “flexible” extension of the UK’s membership. This period is characterized by high volatility and extreme activity peaks, which generally last less than a week. From the end of April to the end of May there is very low activity – with a rapid increase in the week before the EU Parliamentary elections. The third phase, with mid-low volumes of activity and some peaks, follows the success of the Brexit Party at the EU elections and ends with the nomination of Boris Johnson as Tory leader. The fourth and highly active phase corresponds to the first Johnson government and ends with the 2nd extension accorded by the EU. Finally, the subsequent period is characterized by extremely low volumes of tweeting activity about the no-deal.

These phases are marked not only by different volumes of tweets but also by different levels of activity by partisan factions with respect to Others (see Figure 3).

The first phase is characterized by a higher initiative by Brexiteers, while phase 4 sees an initial burst of activity by Remainers followed by a more balanced debate activity of the two partisan factions.

Our approach to argument extraction is based on verbs. Remarkably, Remainers use in their tweets (almost 28%) more causal verbs with respect to Brexiteers. This difference is attenuated, but persists, for retweets (see Table VI in the Appendix). It is possible to capture semantic differences among causal arguments. In particular, we focus on three groups of causal verbs. Two groups clearly express the polarity of the causation relation, we label them as Destruction and Construction verbs. Whereas, the third group, which has a neutral polarity, is simply called Causation. By applying our relation extraction method we obtain for each group of verbs ordered pairs of causes and effects, with the associated metadata (see examples in table I).

| Example I | cause-side | effect-side | rel.type | negated |
|-----------|------------|-------------|----------|---------|
| No deal will not cause anything like the problems Remainers are predicting | No Deal | anything [...] predicting | Causation | TRUE |

| Example II | cause-side | effect-side | rel.type | negated |
|------------|------------|-------------|----------|---------|
| Which of course rules out any free trade deal with the US since no deal will create a hard border between the two Irelands. | Which [...] no deal | a hard [...] Irelands. | Creation | FALSE |

3 See methods Section IV B for details.
We have 204 648 relations, of which 36.116 contain “no-deal” inside the cause side of the relation. From now on we analyse only this set of relations. Further argumentation style differences among factions exist. In particular, Brexiteers put stronger emphasis than Remainers on Creation relationships, while the converse is true for Destruction relations (see Table IV in Appendix). This is true also when we consider only the relations which are not negated (see Table VI in Appendix).

Aggregating arguments in classes (of effects)

Since extracted relations include a large and varied population of effects, there is a need to aggregate the phrases describing the effects of the no-deal in a manageable number of effect topics. We do so through a STM applied on the effect side of extracted relations. One major advantage of using a STM, is that it allows to include metadata as covariates affecting topic proportions and topic contents. This makes it possible to capture, for example, differences in how (topic contents) and how much (topic proportions) factions speak about the different expected effects of the no-deal. These differences cannot be captured with classical topic modeling techniques, like LDA or CTM. The covariates that we allow to affect topic proportions are time (t), the user faction (fct), the relation verb negation (neg) and the relation verb group (rel.type). By so doing, the model can be used to identify which topics characterise each partisan faction, i.e. topics that are more likely to be observed conditionally on the user belonging to one partisan faction with respect to the other. For explaining the topic content we consider only the user faction (fct). This allows to capture differences between groups, in terms of how they debate about each topic, by giving different probability weight to words (unigrams) and collocations (bigrams) that are used to express causal beliefs about the expected effects of a no-deal scenario. Table II shows the 10 most probable tokens (unigrams and bigrams) for the top 10 topics, ranked in decreasing order by overall topic proportion. For example, topic 16, concerning trade agreements, (which is more likely used by Brexiteers) clearly shows differences in the evaluation of Brexiteers (stressing clean terms and the opportunities to use WTO trading agreements) and Remainers (which see the same issue as a nightmare). In topic 32 (which is characterizing Brexiteers), which is about the economic consequences of the no deal, Remainers stress damages to jobs and the stress on the health system (“nhs” token), which disappear from the top list of words for the Brexiteers. Interestingly, for topic 28, which is characterizing neither partisan faction, that focuses on the sense (and non sense) of a no-deal scenario, the Brexiteers and Others appear to claim that a no-deal will likely produce no difference (see rank of bigram token: “no → difference”), whereas for Remainers “uncertainty” is a more high-ranked effect. Finally, Remainers appear to be more concerned than Brexiteers and Others by the scenario of Scotland leaving UK as a result of a no-deal (see topic 9).

Exploring faction rhetoric and narratives of no-deal arguments

Figure 4, shows the argumentative network of topic 2, which has been reconstructed using estimated token probabilities for that topic.4

To see in which terms the two partisan factions intervene in the non-partisan debate concerning topic 2, we create the topic network for the faction Others and filter separately nodes and edges keeping only the 80th percentile, to prune the network from terms (unigrams) and collocations (bigrams) less frequently employed by non-self-declared partisan users within this topic. We subsequently overlay to this network5, the differences in token probabilities between the Brexiteers and Remainers, for each token in the former network. The color of edges and nodes hence represents the partisan faction by which a specific term or collocation is more employed in relation to the selected topic.6

The terms ”shortages”, ”food” and ”medicine” are among the most relevant terms for non partisan users (see box n.1 figure 4). The probabilities of these terms and collocations are very different for Brexiteers and Remainers. For example, while Brexiteers focus more on the term ”food” and the collocation ”food → shortages”, Remainers focus relatively more on the terms ”medicine”, ”medication”, and, the collocations ”medicine → shortages” and ”medication → shortages”.

The Remainers also tend to focus more on the possible effects on unemployment of a no-deal scenario (see box n.2), both the unigrams ”job” and ”losses”, and the bigram ”job → losses” have higher probability for Remainers w.r.t. Brexiteers. Finally, box n. 4 shows that members of partisan factions are more inclined to talk about the party representing the other faction with respect to their own: Remainers use relatively more frequently ”tory” and ”tory → party”, whereas Brexiteers use more frequently ”labour” and ”labour → party”. This signals that while debating about the effects of no-deal, partisan faction members also argue about the no-deal narratives employed by the opposing faction.

---

4 unigrams’ probability is used to weight nodes and bigrams’ probability is used to weight edges for a selected topic
5 using a red-gray-blue color scale, where gray is centered at 0, i.e. no difference between faction probabilities for that token (unigram/bigram)
6 See methods Section IV D for details.
TABLE II: Summary table of top 10 topics (by overall topic proportion), with top 10 tokens (by token probability) by topic and by faction. Row color scale represents significant (at the 0.01 significance level) differences of the estimated topic proportion coefficients of the two partisan factions. The closer is the row color to the corresponding faction color, the more characterizing is a specific topic for one of the two partisan factions. Characterizing topics for Remainers are in blue, whereas characterizing topics for Brexiteers are in red. For the full list of topics we refer to Table 2 in the Supplement.

| Topic | Topic Prop. | Brexiteers (top 10 tokens) | Others (top 10 tokens) | Remainers (top 10 tokens) |
|-------|-------------|-----------------------------|------------------------|---------------------------|
| 24    | 4.9%        | now, just, know, voting, see, happy, ever, well, tea, table | now, know, remain, sense, see, well, far, fact, last, remainers | far, now, know, right, anything, must, staff, year, saying, thinking |
| 23    | 4.5%        | no-deal, borisjohnson, illegal, says, likely, no-deal->brexit, pm, scotland, says->borisjohnson, may | no-deal, likely, borisjohnson, says, pm, says->borisjohnson, likely->says | no-deal, borisjohnson, likely, pm, says, outcome, likely->says, really, perhaps, says->borisjohnson |
| 9     | 4.4%        | unitekingsdom, businesses, leaving, imports, tariffs, competitive, 1, recession, outside, loss | unitekingsdom, recession, leaving, less, massive, businesses, crisis, tariffs, exports, unitekingsdom->economy | unitekingsdom, recession, leaving, massive, tariffs, living, & unemployment, scotland->leaving, unitekingsdom->fishing |
| 16    | 4.3%        | europeunion, law, who, deal-europeunion, end, clean, terms, europeunion-law, 2017, trading | europeunion, back, law, end, some, whole, terms, working, within, deal-europeunion | europeunion, law, unitekingsdom-europeunion, nightmare, europeunion-grae, backstop, neighbours, sounds end apply |
| 20    | 4%          | deal, get, leave, without, good, mps, thereclosely, @, good, difficult | deal, get, good, without, or, getting, negotiate, done, trying, good->deal | deal, remain, get, thereclosely, best, revoke, good, done, put, travel |
| 3     | 4%          | brexit, likely, happen, less, finally, wrong, talk, +, @, patients | brexit, want, happen, really, possible, voters, less, +, idea, less->likely | brexit, peoplesvote, voted, brexit-europeunion, wrong, constituents, vote->no-deal, failure, happen, finally |
| 32    | 4%          | economy, break, lives, world, harm, union, break-union, pound, rejoin, third | economy, damage, jobs, union, harm, lives, nhs, put, 10, risk | economy, jobs, damage, union, harm, break, nhs, huge, break-union, decides |
| 28    | 3.9%        | no, sense, difference, long, no->difference, actually, no-sense, vote, default, change | no, one, sense, difference, long, power, no-sense, change, no->difference, little | no, sense, uncertainty, change, difference, harder, long, longer, no->sense, position |
| 33    | 3.8%        | better, off, much, poorer, worse, better->off, threat, road, poorest, democratic | country, off, worse, much, poorer, things, difficult, worse-off, north, mean | less, poorer, off, leave, much, even, terms, lives, trade, democratic |
| 11    | 3.4%        | border, northernireland, months, chaos, hard, public, ireland, land, operation, northernireland->border | border, chaos, hard, public, hard->border, ireland, months, problem, minister, northernireland | chaos, public, months, border, disorder, hard, chaos->public, months->chaos, public->disorder, hard->border |

The structure of the no-deal debate

In a debate, classes of effects (i.e., topics) related to a common cause are not independent. Some topics are likely occurring together in sentences, whereas others mutually inhibit each other. For example, some effects are semantically related through the verbs expressing causal relations. Topics tend to be clustered according to verbs’ polarity. Effects related to Destruction tend to be positively correlated among them and negatively correlated with those related to Creation. The same is true for Creation verbs. Figure 5, displaying the correlations among topics, shows the near-decomposability of causal arguments in blocks around the two diagonals of the matrix. One can capture a finer-grained structure of the debate by filtering the correlation network. Filtration creates a continuum of networks resulting from the deletion of edges whose weight is below a given threshold varying over the range of observed weights. In our setting weights represent the value of correlations, and these are filtered according to their absolute values. At higher levels of the threshold the network displays strong relations between arguments. The positively connected parts of the graph capture the conceptual building blocks of the debate’s arguments, i.e. topic constellations. As the threshold is relaxed, a weaker (but still significant) set of relations among topics emerges and "assembles" the building blocks (see Figure 6).

Additional structural information is gained by displaying significant topic covariates in a topic correlation network, by typing the nodes by dominant partisan faction (shape), dominant polar verb group (color), and negation of the relation (shadow), whenever the difference among topics is significant.
FIG. 4: Topic 2 argumentation network. Unigrams (nodes) and bigrams (edges) have been filtered (80th percentile threshold) on their probabilities for the non partisan faction *Others*. Node size and edge width represent probabilities of unigrams and bigrams for the non partisan faction *Others*. Node/edge color scales are used to represent the differences in unigram/bigram probabilities between partisan factions (prob. Brexiteers minus prob. Remainers). Numbered boxes display zoomed areas of interest.
8 See Section 1 in the Supplement.
9 We remark that the only pair of Creation-Destruction topics that is positively correlated at the filtration level 0.10 is {topic 24, topic 38}. This pair of topics contains topic 24 that is one of the few topics for which the negation has a significant positive effect on the topic’s expected proportion, i.e. the topic is more likely to appear conditional on the presence of a negation of the relation’s verb phrase (e.g. “[No deal] does not create [topic 24]”).

FIG. 5: Heat-map of correlations between topics, in red positively correlated topics, in blue negatively correlated topics. Non significant correlations have been identified (and set to 0) using the method proposed by Meinshausen and Buhlmann[8]. The topics have been reordered on the basis of the estimated difference between the coefficients of polarized verb groups’ (Creation Vs Destruction) effects on topic proportions ($\hat{\beta}_{rel-Destruction} - \hat{\beta}_{rel-Creation}$). If this difference is significant at the 0.01 level, the label (Creation or Destruction) of the verb group characterizing the topic is contained in brackets after the topic number, followed by the value of the estimated difference (sign adjusted). If non significant, the label None follows the topic number (in brackets).

types is significant. Filtration of such typed network allows highlights that the topic correlation network exhibits type-based assortativity, i.e. assortative-mixing of the topics for multiple (simple and composite) typing dimensions.

Figure 6 shows clear evidence of polar group verb assortativity (up to the 0.14 threshold), and partisan faction assortativity (up to the 0.10 threshold).

Moreover, up to the 0.10 threshold, positively correlated triangles (3-cliques with non dashed edges) are formed by topics of non opposing types. These positively correlated triangles are also coherent in terms of polar faction types, being made either by topics that are not characterizing any of the partisan factions or by topics that are not of opposing factions.

Up to the 0.10 threshold, the only triangle containing both positively and negatively correlated topics is {topic 23, topic 32, topic 9}, which is balanced and coherent being formed by two positively correlated Creation topics (32 and 9), that are both negatively correlated with the third one that is a Destruction topic (23).

More generally, as we lower the threshold to 0.05, the connected components of the graph grow respecting a basic triangle balance principle, i.e. connected triplets of topics are always made of an odd number of positive edges. At thresholds below 0.05 most of the triangles are still balanced, but, as the threshold further lowers some imbalanced triangles appear, maybe displaying the appearance of “relational noise”.

The lead and follow dynamics of the no-deal debate

Debate’s lead and follow relations among partisan factions are important because they reveal the capacity of a faction of setting the thematic agenda for a specific debate in a specific moment in time. Leader-follower relations among factions can be seen as processes of highly coordinated debate activity: when a faction starts discussing more or less about specific effect topics the other faction then follows similar topic proportion variations. In our setting, a leading faction can be seen as a synchronized group of partisan users initiating a debate shift, whereas a following faction can be seen as a rival group of partisan users pushed to respond to the other faction by discussing in the following days the topics proposed by the latter.
FIG. 6: Filtered topics correlation network (absolute value) and faction/type covariates. Node color represents predominant partisan (Brexiteer/Remainer) faction (p.value<0.01), see legend for color details. Node shape represents predominant Destruction/Creation relation type (p.value<0.01), see legend for shape details. Node shadow if negated relation is predominant (p.value<0.01). Solid lines represent positive correlations whose absolute value is higher or equal to the filtration threshold. Dashed lines represent negative correlations.

To identify possibly dynamic lead-follow relations among Brexiteers and Remainers, we apply a Dynamic Time Warping[10] method called FLICA[9]. This method allows us to infer time-varying lead-follow relations between pairs of multidimensional time series, in our case between the faction-specific 38-dimensional topic proportion series, each dimension representing the average daily proportion of a specific topic for a specific partisan faction. We use a window of 30 days, a max lag window of 6 days, and a window time shift of 1 day. Our results are robust to changes of the three parameters, in particular to the max lead/lag window which is the most relevant parameter representing the max range of the warping. With max warping window values from 3 to 10 days, lead/follow patterns remain consistent. While, as expected, reducing the warping range mostly affects the amplitude of the observed oscillations of the index.
FIG. 7: Central plot: lead/follow relation index among partisan factions. The index has been constructed using the method proposed by [9], it is based on Dynamic Time Warping of the 38-dimensional daily topics proportions series by faction, using a time window of 30 days, a max lag window of 6 days, and a window shift of 1 day. The closer the index value is to 1 the more intense is the lead of Brexiteers on Remainers on that day, the converse is true for values close to -1.

Upper and lower plots: Top topics by absolute weekly topic-proportion variation\(^a\). Inside gold, silver and bronze boxes are contained the IDs, of the top 3 topics, in terms of largest weekly positive variations, representing respectively the 1st, 2nd and 3rd ranked topics. Inside the dark violet, violet and light violet boxes are contained the IDs of the top 3 topics in terms of largest weekly negative variations, representing respectively the 1st, 2nd and 3rd ranked topics. Some boxes may be missing when for less than 3 topics non-null (positive/negative) weekly topic proportions changes are observed.

\[^a\]|w_{x,k,i} - w_{x-7,k,i}|\ see methods section IV F for details

As we can see from the central plot in Figure 7, which shows the lead/follow index obtained by applying FLICA, lead-follow relations among the two partisan factions exhibit multiple cycles with different intensities and degrees of persistence. In particular, by relating the peaks and troughs of the index to political events in the UK, we notice that peaks correspond to moments in which politicians and political parties close to Brexiteers were in a position of strength, whereas the converse is true for troughs. Interestingly, rapid shifts in leadership occur after the following events:

- **MPs Amber Rudd ultimatum (22 Feb. 2019):** Brexiteers lead the debate in the first two decades of February, then Remainers take the lead until the beginning of March 2019;

- **1st extension (10 Apr. 2019):** Brexiteers lead the debate from the beginning of March to mid April, then Remainers take the lead until the end of the second decade of May 2019;

- **EU elections (23 May 2019) and the resignation of Theresa May (7 Jun. 2019):** Brexiteers lead the debate from May 21st to the day in which T. May’s resignation becomes effective;

- **Yellowhammer plan leak (18 Aug. 2019):** Remainers lead the debate from mid August until mid October 2019 (excluding 25 Aug. 2019);

- **2019 United Kingdom General Election (12 Dec. 2019):** Remainers lead the debate from the beginning of November until two days before the GE, then Brexiteers take the lead until the end of January 2020 (excluding 21-23 Dec. 2019);

Moreover, by jointly analysing the lead/follow index and the weekly topic variations (see upper and lower plots in Fig. 7), we observe some extremely interesting patterns. First (i), periods where Remainers lead or where leadership starts shifting towards them are often characterised by the presence among the top 3 topics (by weekly positive-sign variation) of topic 32 and topic 9, which are both related to the economic effects of a no-deal scenario. These two topics are also domi-
nated by *Destruction* causal verbs and form one of the strongest macro-argument components identified in Figure 6. Moreover, topic 32 was also found to be characterizing for Remainers. Second (ii), periods where *Brexiteers* lead or where leadership starts shifting towards them are similarly characterised by the presence among the top 3 topics (by weekly positive-sign variation) of the topic 20 and topic 24, which are both dominated by *Creation* causal verbs, and trivialize the alternative to a no deal scenario by ridiculing the difficulties and delays in the negotiations of a trade deal with the EU. Topic 20 is characterizing for *Brexiteers* and contains many emojis that jokingly refer to the extremely remote possibility of reaching an agreement and hence avoiding a no-deal scenario.

Both points suggest that the structural properties identified in the effect topic-correlation network analysis also play a role in the dynamics of lead/follow relations among factions – further supporting the importance to consider the interdependence of arguments and of levels analysis.

**III. DISCUSSION**

We have shown that the structure and dynamics of online debates connect arguments in a coherent way different levels and roles. Our reconstruction of causal arguments in the No Deal debate allows to unveil how the debate dynamics relate to external events, how the different causal arguments mutually connect, and how they are shaped by factional interests. For example, *Remainers* resort more extensively to causal arguments than *Brexiteers* do, they emphasize the potential destructive causal effects of the no deal Brexit, and stress its (negative) economic implications. In particular, we show how the network of arguments displays polarized assortativity around structurally balanced building blocks that aggregate topics by type and factional orientation. Correlated constellations of arguments also play an important role in the debate dynamics, as they often mark, in conjunction with external events, shifts in the factional leadership of the debate.

Thus, better insights can be obtained by addressing the complex architecture of debates through approaches that combine different tools in a coordinated way. In this paper we explore a multi-step approach that traces a methodological roadmap through the different architectural components of the debate – pointing in the direction of the development of the statistical rhetorics of debates. In order to do this, we had to adapt pre-existing tools in innovative ways, e.g. by combining structural topic modeling with network modelling. This allows to ask new questions about the structural properties of debates that generalize beyond the specificity of our case study. For example, the filtration of the correlation network of effect topics suggest some interesting structural regularities – e.g. argument triangles are all structural balanced until very low level of correlation (maybe corresponding to relational noise) are considered.

Besides this general contribution, our paper further contributes to more specific streams of literature.

In relation to Structural Topic Models, this work highlights how classification algorithms that jointly exploit text and metadata can be fruitfully used not only with short online social media posts, like Tweets[11], but also with subsets of short posts which may represent phrases or causal relations extracted with RegEx or NLP algorithms from the former. Moreover, besides using classical document-level covariates, like author and publishing date, this work shows that one can transform otherwise non-exploited textual data from a post or from its metadata in valuable categorical covariates, like the faction and causal relation verb type.

From the point of view of network science, this paper clearly demonstrates that from the micro to the macro scale, graphs rather than sets, appear to be the most appropriate way to analyse and model debates: At the micro level, the inclusion of bigrams in modelling setting allowed us to reconstruct faction-specific argumentation networks, which highlight in which terms single factions intervene in a topical debate; At the macro level, the combination of the topics correlation matrix with predominant covariate-level labels allow us to model the architecture of the debate, and identify its salient dimensions and topological properties. From a broader perspective this work contributes to the emerging field of narrative economics[12] and constructivist approaches to socio-economic issues[13] by offering an extensive framework for studying online causal debates and their dynamics.

Our analysis of the No Deal debate considers only causal arguments that have a single cause (the “No Deal” causal factor) and focuses on different effects associated to it. This constraint was introduced to keep our analysis simple enough, but is not a an intrinsic limitation of our approach. Indeed, one may consider multiple causes of a single effect, or multiple causes of multiple effects as well. A second limitation is that we reconstructed the phrasal networks only within each effect topic. However, as we have shown, arguments are correlated and form detectable constellations of topics. It is possible to reconstruct phrasal networks for such constellations as well. This would potentially bring to light more connections among words and the phrasal constituents of more complex arguments. Thirdly, more systematic information could be extracted by looking at structural indicators of networks, such as centrality, betweenness, modularity and community structure. We explore such issues in a separate paper[14]. Finally, in our analysis we didn’t consider retweets which provide natural indicators of the social resonance of arguments. As this deviates from the main focus of this paper, we defer this analysis to a different one.
IV. METHODS

The data acquisition, preparation and modelling procedures used in this study are as follows.

A. Data collection and pre-processing

Tweets about the no-deal have been collected through Twitter’s Stream API, from the beginning of February 2019 to beginning of May 2020. In total, more than 9 million tweets were downloaded and archived in a SQL database. The downloaded data were then pre-processed to remove tweets that are unrelated to the no-deal Brexit. In particular, tweets related to the US-China trade war (and its no-deal scenario) where identified and hence removed using a RegEx containing references to these two countries and their leaders. The cleaned data-set contains 9004927 tweets about the no-deal Brexit and their metadata.

B. Arguments and argument-specific covariates extraction

To find and extract from the no-deal Tweets segments of text that identify cause-effect relations we employed a RegEx algorithm that exploits verbs. In particular, a list of verbs and verb phrases related to Causation, Creation and Destruction (see Table 7 in the Supplement) are used to build a set of RegEx functions, which are hence used to identify and isolate the two sides of each cause-effect relation contained in the corpus of tweets. The steps of the the cause-effect extraction process can be summarised as follows:

1. Tweets are segmented in sentences using punctuation characters.

2. For each sentence:
   2.1 Verb phrases related to cause-effect relations are identified through RegEx functions.
   2.2 If there are one or more RegEx matching, for each matching we identify and record the verb phrase’s:
      2.2.1 position (start char. and end char.);
      2.2.2 type (causation, creation, or destruction);
      2.2.3 negation (e.g., "Y will not cause X" implies neg=TRUE) if any;
      2.2.4 verbal form, which can be:
         • active (e.g., X will cause Y);
         • passive (e.g., Y will be caused by X);
         • end-of-sentence (e.g., Y that X will cause);

2.3 For each sentence with at least one RegEx matching we use the previously extracted information together with a set of verbal form specific functions to split (and reorder) the different components of each sentence in one ore more cause-effect relation triplets, each of which includes the following elements:
   • a subject representing the cause-side of the relation (e.g., A no-deal Brexit ...);
   • a predicate representing the relation type to which the verb phrase corresponds (e.g., ... would certainly destroy ... → rel.type=Destruction);
   • an object representing the effect-side of the relation (e.g., ... UK’s economy and labour market.).
   2.4 We associate to each cause-effect relation triplet a set of covariates characterising the relation:

   We hence obtain a set of cause-effect relation triplets and their covariates. Each triplet i is characterised by:
   • ti - date-time of relation i (which is the publishing date-time UTC of the tweet from where the relation i was extracted)
   • rel.typei - verb phrase type of relation i (3-levels categorical: Causation, Creation, Destruction)
   • negi - verb negation dummy of relation i (binary):
     • negi = TRUE if the verb phrase of the relation i contains a negation;
     • otherwise negi = FALSE;
   • fcti - faction (3-levels categorical: Brexiteer, Remainder, Other):
     • fcti = Brexiteer if the biography of the user that has posted the tweet from which the relation i was extracted matches at least one case insensitive- RegEx condition contained in the Brexiteer dictionary (dictBrexiteer = {brexiteer, vote brexit, voted...})

11 Twitter’s V1 filtered Stream API endpoint has been queried through DMI-TCAT[15], using as stream filters ‘no deal’ and ‘no-deal’
12 Regex with case sensitive matching: “China| china| u.|[.]s|[.]a| u.|[.]s| U.|[S] US| USA| UNITEDSTATES| UnitedStates| UNITED STATES| United States| united states| trump| Trump| TRUMP| Xi| Xi”
13 For RegEx algorithm details and commented code we refer to Table 8 in the Supplement.
14 the Quanteda library[16] was used for this purpose.

15 The following examples are based on the sentence: A no-deal Brexit would certainly destroy UK’s economy and labour market.
To aggregate the extracted no-deal effects in classes of effects (i.e., topics) and to see in which terms the propensity to speak about these classes may depend on covariates, we estimate a Structural Topic Model (STM) using only the previously extracted 36116 effects related to the no-deal. We selected STM for its unique combination of features required to fulfill our objectives. (i) First, being an extension of the Dirichlet-Multinomial Regression topic model[17],STM allows for the inclusion of covariate information in the estimation process. This affects the estimation through informative priors and, more importantly, allows us to evaluate the effects of extracted covariates on topic proportions. (ii) Second, being constructed upon the Correlated Topic Model[18] it allows us to infer the interdependence structure among topics that co-occur in (the effect-side of) relations that have no-deal as a subject. (iii) Finally, being a generalization of the Sparse Additive Generative[19] topic model, it allows covariates to affect the contents of a topic, through sparse deviations with respect to a baseline distribution. This feature is here applied to the fct covariate, to model and analyse in which terms, for a given topic, faction-specific argumentative styles can be distinguished from one another.

Each effect $d \in \{1, ..., D\}$ (where $D = 36116$) is represented as a set of tokens from a vocabulary of unigrams and bigrams, indexed by $v \in \{1, ..., N\}$. Effects are hence transformed in matrix called $ExT$ of size $D - by - K$ containing the counts of the number of tokens by effect. As a modelling strategy, we allow covariates contained in the $(D) - by - (4)$ matrix $X = \{t, rel, neg, fct\}$ to affect topic proportions, whereas only the covariate vector $Y = \{fct\}$ is allowed to affect the contents of topics. The choice of having the faction covariate $fct$ affect contents is related to our objectives of understanding if and in which terms partisan factions taking part in the Brexit debate use different words (unigrams) and associations (bigrams) to speak about an inferred class of no-deal effects (i.e., topic). As tokens, we include unigrams and bigrams which appear at least 10 times in the final collection of no-deal effects. This leaves us with a vocabulary $V$ made of $N = 3505$ tokens, of which $N_u = 2462$ are unigrams and $N_b = 1043$ are bigrams. The matrices ExT and X are hence used, together with the vector Y, as inputs to estimate our model using the Stm package for R[20].

A Structural Topic Model with $K$ topics is defined as:

**Topic proportion**

\[
\mu_{d,k} = X_d \gamma_k \\
\gamma_k \sim N(0, \sigma_k^2) \\
\sigma_k^2 \sim \text{Gamma}(s^\gamma, r^\gamma)
\]  

\[\text{(1)}\]

---

among others, the cause side must match the RegEx "no-[\?deal", see Table 6 in the Supplement for details.
Language model

\[ \theta_d \sim \text{LogisticNormal} (\mu_d, \Sigma) \]
\[ z_{d,n} \sim \text{Mult} (\theta_d) \]
\[ v_{d,n} \sim \text{Mult} (\beta_{d}^{v=zd,n}) \] (2)

Topic content

\[ \beta_{d,v}^{k} \propto \exp (m_v + \kappa_v^{k} + \kappa_v^{y} + \kappa_v^{y,k}) \]
\[ \kappa_v^{y,k} \sim \text{Laplace} (0, \tau^{y,k}) \]
\[ \tau^{y,k} \sim \text{Gamma} (s^k, r^k) \] (3)

Where topics are index by \( k \), \( X_d \) is a \( 1 \text{-by-} 4 \) vector, \( \gamma_k \) is a \( 4 \text{-by-} K \) matrix of coefficients, and \( \Sigma \) is a \( K \text{-by-} K \) topic proportion covariance matrix. The distribution over tokens \( n \) is the combination of three effects: a topic effect \((\kappa_v^k)\); a costate \( \text{fct} \) covariate effect \((\kappa_v^y)\), and a topic-covariate interaction effect \((\kappa_v^{y,k})\). These three effects are modelled as sparse deviations from a baseline token \((m_v)\). To choose the number of topics \( K \), we estimate the model for different values of \( K \) ranging from 3 to 70. For each value of \( K \), we repeat 50 times the following procedure: (i) split the corpus in a random training set and a test set\(^{17}\) using a different random seed at each repetition; (ii) estimate the STM model\(^{18}\); (iii) compute the lower bound and the mean likelihood of the STM to evaluate its performance. Then for each \( K \in \{3, ..., 70\} \), the average values of the lower bound and the mean likelihood are computed. The aforementioned model performance indicators suggest that \( K = 38 \) is a good candidate number of topics for estimating an STM with our corpus of no-deal effects\(^{19}\). Finally, the STM is re-estimated for \( K = 38 \) with the whole set of effects\(^{20}\).

D. Constructing faction-specific narrative networks

To construct faction specific narrative networks for a topic \( k \) we use the posteriors of \( \tilde{m}_v, \kappa_v^k, \kappa_v^y, \kappa_v^{y,k} \), which are respectively called \( \tilde{m}_v, \kappa_v^k, \kappa_v^y, \kappa_v^{y,k} \). In particular, we first separate tokens in the vocabulary \( V \) in two disjoint sets \( V_u \) and \( V_b \), where \( V_u \) contains only the unigrams from the vocabulary \( V \) and \( V_b \) only the bigrams. For all token \( v \in V_u \), we filter out the less relevant unigrams for the \( k \)th topic and for faction Others. We do so by keeping only tokens above the 80th percentile rank, in terms of the following posteriors sum: \( \tilde{m}_v + \kappa_v^k + \kappa_v^y = \text{Others}, + \kappa_v^{y,k} = \text{Others,k} \). We then apply the same procedure for bigrams (all \( v \in V_b \)). Finally we filter out bigrams that are not connecting unigrams pairs in the 20% top percentile. From the resulting unigram and bigram sets and their weights, which are given by the exponential of the above mentioned posteriors sum, we can construct the narrative network of the topic \( k \) for non partisan users \((\text{fct} = \text{Others})\). We construct it by using unigrams as nodes and bigrams as edges, and by representing unigrams’ weights through the node size and bigrams’ weights through the edge width. The resulting network can be seen as a graphical representation of the phrasal microstructure of the debate about a topic, for non partisan users. Since we are interested in analysing how partisan factions intervene in this debate we overlay the topic content differences between the two partisan factions using a continuous color scale ranging from blue (for negative values) to light-gray (for zero) to red (for positive values). For any token \( v \), the posterior probability difference between faction \( i \) and faction \( j \) is given by:

\[ \hat{\delta}_{v,i,j} = \exp (\tilde{m}_v + \kappa_v^y = i + \kappa_v^{y,j,k} + \kappa_v^{y,j,k}) - \exp (\tilde{m}_v - \kappa_v^y = j + \kappa_v^{y,j,k}) \] (4)

With \( i = \text{Brexit} \) and \( j = \text{Rem.} \), we obtain the difference between the Brexiters and Remainers partisan factions for the topic \( k \). We can hence overlay the colors representing partisan factions’ differences to the narrative network of non-partisan users \((i.e., \text{Others})\), as shown in Figure 4 for topic 2.

E. Filtering the network structure of a debate and identifying constellations of effect-classes

To explore the relationship between covariate values and topic proportions we use the \textit{estimateEffect} function of the STM library. This function allows to estimate the effects of one or more covariates included in the STM estimation phase on expected topic proportions. For each effect \( d \), the proportions of a topic \( k \) are modelled as a function of the faction \((\text{fct}_d)\), relation type \((\text{rel.type}_d)\), and verb negation \((\text{neg}_d)\) covariate values:

\[ \text{propensity}_{k,d} = f(\text{fct}_d, \text{neg}_d, \text{rel.type}_d) \] (5)

This method also allows to assess which covariate coefficients are statistically significant\(^{21}\). To analyse if there are significant differences in topic prevalence among the two partisan factions (Brexiters Vs Remainers) and

\[ \text{see Table 3 in the Supplement for regression results} \]

\[ \text{\footnote{The training set contains a random sample containing 25\% of the total number of no-deal effects}} \]

\[ \text{\footnote{See Section 2 in the Supplement for details about STM parameter values.}} \]

\[ \text{\footnote{See Section 2 in the Supplement.}} \]

\[ \text{\footnote{Using spectral initialization, which allows the estimated STM to be deterministic conditionally on parameters and covariates values.}} \]
among the two polarized verb types (Creation Vs Destruction), we compute the coefficients’ differences and their variance to test if the former are statistically different.

To represent the aggregate structure of the debate about no-deal Brexit effects, we transform the $\Sigma$ matrix in a topics propensity correlation matrix. We hence obtain a pruned correlation matrix that can be visualized as an undirected graph, where nodes represent topics and edges represent correlations between them. We then label the correlation graph on the basis of covariate-levels that have predominant effects (i.e., significantly larger coefficients) with respect to their opposing type (Brex Vs Rem., Creation Vs Destruction). More specifically, for each topic we label its node on the basis of the covariate level that implies a significantly higher propensity for that topic (if any), otherwise we label that predominance relation property with the “none” label. The topic’s predominant relation type ($\tilde{\beta}_{rel.type=\text{Creation}}$ Vs $\tilde{\beta}_{rel.type=\text{Destruction}}$) is represented through the node color. Whereas, the topic’s predominant partisan faction ($\tilde{\beta}_{rel.type=\text{Brex.}}$ Vs $\tilde{\beta}_{rel.type=\text{Rem.}}$) is represented through the node shape. Finally, we represent significant positive effects on a topic’s propensity related to the presence of a negated relation (i.e., $\tilde{\beta}_{neg=T\text{RUE}} > 0$) by applying a shadow around the topic node. To highlight the main structural relations among topics used in the no-deal effects debate, and identify topic constellations that attract or repulse each other, we filter the correlation graph using different threshold levels applied to the absolute value of the correlations, which are represented through the edges’ width. These threshold values are progressively lowered, and at each step isolated nodes are removed to show only the backbone of the debate for that specific filtration level. This allows to analyse the building blocks of a debate and how these blocks grow as we lower the threshold. Moreover, by analysing the topology of this network (e.g., balanced and unbalanced triangles or cliques) one can see if these building blocks are coherent either in terms of the sign of the correlations that characterise them, or in terms of the property labels associated to predominant covariate levels.

F. Identifying time-varying faction lead/follow relations

Many methods to analyse lead-follow relations among time series exist, like cross-correlations among faction-specific daily topic proportions series. Despite their usefulness, these methods have several limits, in particular the resulting lead-follow relations are by construction static. As a result, we apply a method based on Dynamic Time Warping [10] (DTM), which allows inferred lead-follow relations among factions to change across time.

As a first step to identify the time varying faction lead/follow relations, the $D-$by–$K$ matrix containing the distribution of topics by effect, called $\text{ExT}$, is extracted from the estimated STM. We have that $\text{ExT}_{d,k}$ represents the estimated propensity of topic $k$ in the no-deal effect $d$, and $\sum_{k=1}^{K} \text{ExT}_{d,k} = 1$. Using $\text{ExT}$ together with the $t$ and $fct$ covariates contained in $X$, for each faction $i \in \{\text{Brex.},...,\text{Rem.}\}$, for each topic $k \in \{1,...,K\}$, and for each day $x \in \{01-02-2019,02-02-2019,...,01-05-2020\}$, we compute the average daily propensity of topic $k$ for faction $i$ on day $x$ and call it $w_{x,k,i}$. Where for a specific day $x$ and faction $i$, we have that $w_{x,k,i} > 0 \Leftrightarrow \sum_{j} w_{x,j,i} = 1$ if there is at least one effect $d$ that has $x$ as date ($t_{d}$ covariate equal to $x$) and $i$ as faction ($fct_{d}$ covariate equal to $i$) [22], and $w_{x,k,i} = 0 \forall k$ otherwise [23].

Each matrix $w_{...,i}$ is of size $T-$by–$K$ and contains as column vectors $K$ time-series with the average estimated topic propensities of extracted effects posted by users belonging to faction $i$. Dynamic lead-follow relations among factions are identified with a DTM method called FLICA [9]. In this framework the notion of leading entity (i.e., leading faction) corresponds to the initiation of topical proportion patterns that other factions hence follow. Given a set of time series representing average topic proportions for each faction, one can use this method to identify periods of coordinated activity between factions, and infer the dynamics across time of lead and follow relations between them. The algorithm takes as input the $w_{...,i}$ matrices for two or more factions, for example Brexiteers ($i = \text{Brex.}$) and Remainers ($i = \text{Rem.}$), each of which can be seen as a 38-dimensional time series (at the daily frequency), and through a DTM algorithm gives as output a dynamic directed network, for which nodes represent the factions and edges represent following relations between them. Each frame of this dynamic network represents a day. For each frame, inferred lead-follow relations between pairs of nodes are mutually exclusive, so, or Brexiteers follow Remainers (Brex. $\rightarrow$ Rem. and Brex. $\leftarrow$ Rem.) or Remainers follow Brexiteers (Brex. $\rightarrow$ Rem. and Brex. $\leftarrow$ Rem.) or no lead-follow relation is observed (Brex. $\rightarrow$ Rem. and Brex. $\leftarrow$ Rem.). $f_{i,j,x} \in \{0,1\}$ is the weight of the edge $ij$ at the date $x$, and represents the strength of the follow relation (if any) between node $i$ and $j$ at a specific day. The values of $f_{\text{Brex.},\text{Rem.},x}$ and $f_{\text{Rem.},\text{Brex.},x}$ are used to build our partisan faction lead/follow relation index (central plot in Figure 7), which is defined as follows:

$$f_{\text{Rem.},\text{Brex.},x} = f_{\text{Rem.},\text{Brex.},x} - f_{\text{Brex.},\text{Rem.},x} \in [-1,1]$$

(6)

To implement the FLICA algorithm, we employ mFLICA function from the mFLICA library for R [21], using a window of one month (30 days), a max lag window of 6 days, and a window time shift of 1 day. Results appear to be robust, and small and medium changes in the aforementioned parameters give similar results.
[1] J. A. Gottfried, B. W. Hardy, R. L. Holbert, K. M. Winneg, and K. H. Jamieson, The changing nature of political debate consumption: Social media, multitasking, and knowledge acquisition, Political Communication 34, 172 (2017).

[2] D. Lazer, A. S. Pentland, L. Adamic, S. Aral, A. L. Barabasi, D. Brewer, N. Christakis, N. Contractor, J. Fowler, M. Gutmann, et al., Life in the network: the coming age of computational social science, Science (New York, NY) 323, 721 (2009).

[3] K. P. a. Claudia Wiesner, Taru Haapala, Debates, Rhetoric and Political Action: Practices of Textual Interpretation and Analysis, 1st ed., Rhetoric, Politics and Society (Palgrave Macmillan UK, 2017).

[4] A. T. Barron, J. Huang, R. L. Spang, and S. DeDeo, Individuals, institutions, and innovation in the debates of the french revolution, Proceedings of the National Academy of Sciences 115, 4607 (2018).

[5] D. M. Blei, Probabilistic topic models, Communications of the ACM 55, 77 (2012).

[6] T. Willaert, P. Van Eecke, K. Beuls, and L. Steels, Building social media observatories for monitoring online opinion dynamics, Social Media+ Society 6, 2056305119898778 (2020).

[7] L. Talmy, Force dynamics in language and cognition, Cognitive science 12, 49 (1988).

[8] N. Meinshausen, P. Bühlmann, et al., High-dimensional graphs and variable selection with the lasso, Annals of statistics 34, 1436 (2006).

[9] C. Amornbunchornvej and T. Berger-Wolf, Framework for inferring leadership dynamics of complex movement from time series, in Proceedings of the 2018 SIAM International Conference on Data Mining (SIAM, 2018) pp. 549–557.

[10] D. J. Berndt and J. Clifford, Using dynamic time warping to find patterns in time series., in KDD workshop, Vol. 10 (Seattle, WA, USA:, 1994) pp. 359–370.

[11] T. A. Curry and M. P. Fix, May it please the twitterverse: The use of twitter by state high court judges, Journal of Information Technology & Politics 16, 379 (2019).

[12] R. J. Shiller, Narrative economics, American Economic Review 107, 967 (2017).

[13] J. R. Searle, S. Willis, et al., The construction of social reality (Simon and Schuster, 1995).

[14] G. Minello, C. Santagiustina, and M. Warglien, Improving lda with bigrams: a network-based approach, in preparation (2021).

[15] E. Borra and B. Rieder, Programmed method: Developing a toolset for capturing and analyzing tweets, Aslib Journal of Information Management (2014).

[16] K. Benoit, K. Watanabe, H. Wang, P. Nulty, A. Obeng, S. Müller, and A. Matsuo, quanteda: An r package for the quantitative analysis of textual data, Journal of Open Source Software 3, 774 (2018).

[17] D. M. Mimno and A. McCallum, Topic models conditioned on arbitrary features with dirichlet-multinomial regression., in UAI, Vol. 24 (Citeseer, 2008) pp. 411–418.

[18] D. Blei and J. Lafferty, Correlated topic models, Advances in neural information processing systems 18, 147 (2006).

[19] J. Eisenstein, A. Ahmed, and E. P. Xing, Sparse additive generative models of text, in Proceedings of the 28th international conference on machine learning (ICML-11) (Citeseer, 2011) pp. 1041–1048.

[20] M. E. Roberts, B. M. Stewart, and D. Tingley, Stm: An r package for structural topic models, Journal of Statistical Software 91, 1 (2019).

[21] C. Amornbunchornvej, mifica: An r package for inferring leadership of coordination from time series, arXiv preprint arXiv:2004.06092 (2020).

ACKNOWLEDGEMENTS

The authors acknowledge financial support from the European Union ODYCCEUS Horizon 2020 project, grant agreement number 732942.

AUTHOR CONTRIBUTIONS STATEMENT

M.W. and C.S. conceived the paper, the methodology, analysed the results and wrote the paper. C.S. downloaded the data and undertook the data analysis. M.W. and C.S. reviewed the manuscript.

22 i.e., card(t == x ⊗ fct == i) > 0
23 w_{x,k,i} = 0 \forall k \Rightarrow \sum_j w_{x,j,i} = 0
### ADDITIONAL INFORMATION

**TABLE III:** Counts (and shares) of extracted relations that have "no deal" in cause-side, by relation type and by faction, followed by Pearson’s Chi-squared test.

| Faction    | Creation | Causation | Destruction | TOT. |
|------------|----------|-----------|-------------|------|
| Brexiteers | 436 (49.7%) | 325 (37.1%) | 116 (13.2%) | 877 (2.4%) |
| Others     | 15817 (46.0%) | 13181 (38.3%) | 5398 (15.7%) | 34396 (95.2%) |
| Remainers  | 361 (42.8%) | 334 (39.6%) | 148 (17.6%) | 843 (2.3%) |
| **TOT.**   | 16614 (46.0%) | 13840 (38.3%) | 5662 (15.7%) | 36116 |

Pearson’s Chi-squared test (vars: *faction* and *rel. type*, data: *causal relations that "no deal" in cause-side*):

X-squared = 10.482, df = 4, p-value = 0.03305

**TABLE IV:** Count (and share) of tweets and retweets about the "no-deal" containing one or more *Causal Markers* (CM), by faction, followed by Pearson’s Chi-squared tests.

| Faction    | TWEETS with CM | TWEETS without CM |
|------------|----------------|-------------------|
| Brexiteers | 7407 (11.9%)   | 54900 (88.1%)     |
| Others     | 265879 (12.5%) | 1853757 (87.5%)   |
| Remainers  | 6304 (15.2%)   | 35161 (84.8%)     |

Pearson’s Chi-squared test (vars: *faction* and *contains.causal.marker*, data: *tweets only*):

X-squared = 289.18, df = 2, p-value < 10\(^{-15}\)

| Faction    | RETWEETS with CM | RETWEETS without CM |
|------------|------------------|---------------------|
| Brexiteers | 28928 (14.5%)    | 170617 (85.5%)      |
| Others     | 1024830 (16.0%)  | 5394586 (84.0%)     |
| Remainers  | 27630 (17.0%)    | 134928 (83.0%)      |

Pearson’s Chi-squared test (vars: *faction* and *contains.causal.marker*, data: *retweets only*):

X-squared = 448.21, df = 2, p-value < 10\(^{-15}\)

**TABLE V:** Counts (and shares) of collected Twitter posts about "no deal" by type and by faction.

| Faction    | TWEETS | RETWEETS | FACTION TOT. |
|------------|--------|----------|--------------|
| Brexiteers | 62307 (23.8%) | 199545 (76.2%) | 261852 (2.9%) |
| Others     | 2119636 (24.8%) | 6419416 (75.2%) | 8539052 (94.8%) |
| Remainers  | 41465 (20.3%) | 162558 (79.7%) | 204023 (2.3%) |
| **TYPE TOT.** | 2223408 (24.7%) | 6781519 (75.3%) | 9004927 |

Pearson’s Chi-squared test (vars: *faction* and *is.retweet*):

X-squared = 2285.7, df = 2, p-value < 10\(^{-15}\)
TABLE VI: Counts of extracted relations that have "no deal" as cause, by relation type, negation dummy and faction, followed by Mantel-Haenszel chi-squared tests

| fct       | N. (%)  | negated | N. (%)  | rel.type | N. (%) |
|-----------|---------|---------|---------|----------|--------|
|           |         | FALSE   | 856 (97.6%) | Creation | 423 (49.4%) |
|           |         | TRUE    | 21 (2.4%)  | Causation| 319 (37.3%) |
| Brexiteers| 877 (2.4%) |         |          | Destruction| 114 (13.3%) |
|           |         |         |          | Creation | 13 (61.9%) |
|           |         |         |          | Causation| 6 (28.6%) |
|           |         |         |          | Destruction| 2 (9.5%) |
| Others    | 34396 (95.2%) |         | 33343 (96.9%) | Creation | 15162 (45.5%) |
|           |         | TRUE    | 1.053 (3.1%) | Causation| 12892 (38.7%) |
|           |         |         |          | Destruction| 5289 (15.8%) |
|           |         |         |          | Creation | 655 (62.2%) |
|           |         |         |          | Causation| 289 (27.4%) |
|           |         |         |          | Destruction| 109 (10.4%) |
| Remainers | 843 (2.3%) |         | 813 (96.4%) | Creation | 344 (42.3%) |
|           |         |         |          | Causation| 326 (40.1%) |
|           |         |         |          | Destruction| 143 (17.6%) |
|           |         | TRUE    | 30 (3.6%)  | Causation| 8 (26.7%) |
|           |         |         |          | Destruction| 5 (16.6%) |

Mantel-Haenszel chi-squared tests:
- strata: fct, vars: rel.type and negated, $M^2 = 118.81$, df = 2, p-value < $10^{-14}$
- strata: rel.type, vars: negated and fct, $M^2 = 2.5141$, df = 2, p-value = 0.2845
- strata: negated, vars: rel.type and fct, $M^2 = 10.994$, df = 4, p-value = 0.02663
Supplement Section 1 contains additional materials related to the results. It includes: the values of the parameters used for the final STM estimation (Table 1); a summary table of top 10 tokens by topic and by faction (Table 2); a table containing estimated topic propensity covariate effects (Table 3); three figures representing respectively significant differences in topic proportions as a function of polar verb type (Figure 1), partisan faction (Figure 2), and negation (Figure 3) covariates.

Robustness checks and description of parameters used in the STM are described in Section 2 of the Supplement. The section includes: the values of the parameters used in experiments for choosing the number of topics $K$ (Table 4); two Figures representing the STM’s Lower bound (Figure 6) and Heldout likelihood (Figure 5) as a function of $K$.

Finally, details about the methods and code are given in Section 3, which contains: a summary of the workflow (Figure 7); the list of used stopwords (Table 5); a RegEx used to identify relations that mention the no-deal in the cause-side (Table 6); the list of verbs used to build verb phrases related to Destruction, Causation and Creation (Table 7); and the R code used to identify extract cause-effect relations from texts (Table 8).

### 1. Results supplement

The results in the main paper and in the Supplement have been obtained using the parameter values in Table 1 for estimating the STM. For not listed parameters, Stm (V.1.3.5) library defaults have been used. The model converged after 32 iterations. The value of the Lower Bound at the final iteration is $-1203368.488309473032132$.

| parameter            | value           |
|----------------------|-----------------|
| $K$                  | 38              |
| Topic prevalence     | $rel.type + neg + fct + s(t)$ |
| Topic content        | $fct$           |
| max.em.its           | 500             |
| entol                | $1e-05$         |
| init.type            | Spectral        |
| gamma.maxits         | 5000            |
Table 2. Summary table of topics (by overall topic proportion), with top 10 tokens (by token probability) by topic and by faction. Row color scale represents significant (at the 0.01 significance level) differences of the estimated topic proportion coefficients of the two partisan factions. The closer is the row color to the corresponding faction color, the more characterizing is a specific topic for one of the two partisan faction. Characterizing topics for Remainers are in blue, whereas characterizing topics for Brexiteers are in red.

| Topic | Topic Prop. | Brexiteers (top 10 tokens) | Others (top 10 tokens) | Remainers (top 10 tokens) |
|-------|-------------|---------------------------|------------------------|--------------------------|
| 24    | 4.5%        | now, just, know, voting, see, happy, ever, well, unlike, table | now, know, remain, sense, see, well, far, fast, last, remain | far, now, know, right, anything, must, staff, year, saying, thinking |
| 23    | 4.5%        | no-deal, borisjohnson, illegal, says, likely, no-deal->brexit, pm, scotland, says->borisjohnson, may | no-deal, likely, borisjohnson, says, pm, says->borisjohnson, likely->says, illegal, add | no-deal, borisjohnson, likely, pm, says, outcome, likely->says, really, perhaps, says->borisjohnson |
| 9     | 4.4%        | unitedkingdom, businesses, leaving, imports, tariffs, competitive, 1, recession, outside, loss | unitedkingdom, recession, leaving, less, massive, businesses, crisis, tariffs, exports, unitedkingdom->economy | unitedkingdom, recession, leaving, massive, tariffs, living, E, unemployment, scotland->leaving, unitedkingdom->fishing |
| 16    | 4.3%        | europeunion, law, veto, deal->europeunion, end, clean, terms, europeunion->law, 2017, trading | europeunion, back, law, end, come, whole, terms, working, within, deal->europeunion | europeunion, law, unitedkingdom->europeunion, nightmare, europeunion->give, deadlock, neighbours, sound, end, apply |
| 20    | 4%          | deal, get, leave, without, good, mps, thereismay, deal, trying, good->deal | deal, remain, get, mememays, best, revoke, good, done, put, travel | deal, remain, get, mememays, best, revoke, good, done, put, travel |
| 3     | 4%          | brexit, likely, happen, less, finally, wrong, talk, ->, patients | brexit, want, happen, really, possible, voters, less, +, idea, less->likely | brexit, people->vote, voted, brexit->europeunion, wrong, constituents, vote->no-deal, failure, happen, finally |
| 32    | 4%          | economy, break, lives, world, harm, union, breakunion, pound, remain, third | economy, damage, jobs, union, harm, lives, nhs, put, 10, nhs | economy, jobs, damage, union, harm, break, nhs, huge, breakunion, decades |
| 29    | 3.8%        | no, sense, difference, long, no->difference, actually, no->sense, voice, default, change | no, one, sense, difference, long, power, no->sense, change, no->difference, little | no, sense, uncertainty, change, difference, harder, long, longer, no->sense, position |
| 33    | 3.8%        | better, off, much, poorer, worse, better->off, threat, road, poorest, democratic | country, off, worse, much, poorer, things, difficult, worse->off, north, mean | less, poorer, off, leave, much, even, terms, lives, trade, democratic |
| 11    | 3.4%        | border, northernireland, months, chaos, hard, public, ireland, land, operation, northernireland->border | border, chaos, hard, public, hard->border, ireland, months, problem, minister, northernireland | chaos, public, months, border, disorder, hard, chaos->public, months->chaos, public->disorder, hard->border |
| 26    | 3.4%        | ireland, unitedireland, united, damage, government, labour, minimal, minimal->damage, labour->government, or | government, ireland, jeremycorbyn, toxi, northernireland, or, labour, damage, united, years | jeremycorbyn, northernireland, ireland, scotland, farmm, mess, government, jeremycorbyn->government, less->wite, wales |
| 35    | 2.9%        | £, going, win, give, money, millions, power, exit, little, control | money, said, going, millions, tax, harder, give, win, pay, others | money, going, benefit, £, spending, financial, tax, fortune, bubbles, hide |
| 25    | 2.8%        | people, want, still, believe, claiming, make, isn't, understand, thinks, die | people, still, believe, understand, national, million, die, emergency, customs, checks | people, want, still, believe, understand, lack, national, stopthecoup, make, die |
| 13    | 2.8%        | years, take, sun, or, trade, minimal, government, jeremycorbyn, labour, damage | years, trade, deals, new, world, another, civil, free, trade->deals, unrest | trade, years, civil, free, another, risk, unrest, friends, civil->unrest, nation |

Table 1 – Continued on next page
### Table 1 – Continued from previous page

| Topic Prop. | Brexiteers (top 10 tokens) | Others (top 10 tokens) | Remainers (top 10 tokens) |
|-------------|---------------------------|------------------------|---------------------------|
| 21 2.8%     | like, look, theressmey, look->like, left, warning, fishing, brexitees->fishing, fishing->warning, sacrifice->brexitees, theressmey->sacrifice | like, look, theressmey, 1, look->like, thing, shit, seem, big, best | like, look, look->like, seem->like, seem, poor, fault, thing, extremely, fool |
| 34 2.7%     | offer, votes, position, million, things, way, result, bed, 17-4, negotiation | better, way, bad, work, life, lose, option, anyone, everything, less | better, life, worse, less, anyone, situation, bed, plan, way, things |
| 37 2.7%     | already, many, real, term, sides, said, deaths, mess, proper, short | many, deaths, already, term, short, shortage, herdship, mess, short-term, pain | many, deaths, misery, med, proper, avoidable->deaths, avoidable, real, short, officials->say |
| 31 2.6%     | economic, hit, non, year, german, economic->problems, poor, problems, no-deal->default | economic, time, say, united kingdom, disaster, companies, actually, hit, recession | economic, crisis, impossible, companies, uter, disruption, social, economic->crisis, norway, german |
| 15 2.4%     | let, shortage, huge, thus, disruption, dogging, alone, water, let-alone, problems | problems, disruption, huge, 3, medicines, supply, delays, including, price, serious | disruption, medical, supplies, let, supply, cancer, drugs, medicines, increase, petrol |
| 5 2.4%      | 50, article, article->50, party, labour, heard, no-deal, tory, default, mp | party, tory, election, labour, majority, general, must, tory->party, brexitees-party, conservative | vote, party, labour, tory, tory->party, ok, changes, leader, place, general |
| 17 2.4%     | great britain, great, great britain->great, country, suffer, 2019, democracy, great britain->people, save, stupid | great britain, great, break, left, democracy, agree, farming, though, suffer, england | country, great britain, take, democracy, single, uk labour, much->damage, 2019, account, entire |
| 7 2.2%      | industry, business, stock, car->industry, products, global, car, entire, getting, experts | business, right, industry, great britain, thousands, state, manufacturing, car, services, orchids | manufacturing, industry, global, thousands, reputation, international, manufacturing->industry, investment, car, cars |
| 10 2.2%     | impossible, stop, revolution, peaceful, parliament, extension, ge, fucking, gov, main | parliament, stop, negotiations, position, impossible, ge, negotiating, least, expensive, taking | negotiations, stop, parliament, decade, expensive, lying, uncertainty, always, block, 100 |
| 36 2.2%     | even, billions, tories, blame, half, bad->deal, even->through, though, super, even->if | even, tories, everyone, billions, rich, blame, care, richer, else, generation | even, tories, vulnerable, rich, already, though, blame, generation, billions, everyone |
| 19 2.1%     | fail, laws, prices, 30, rights, sky, market, high, lower, shops | £, year, prices, pound, billion, rights, fall, lower, 100, higher | problems, businesses, citizens, terrible, terrible->problems, problem->citizens, citizens->businesses, prices, lower, kids |
| 2 2.1%      | food, as, shortages, riots, food->shortages, co2, cost, government, held, apparently | food, shortages, job, medicine, losses, job->losses, collapse, food->shortages, medicine->shortages, riots | shortages, job, food, medicine, food->shortages, medication, co2, medicine->shortages, job->losses, losses |

Table 1 – Continued on next page
Table 1 – Continued from previous page

| Topic | Topic Prop. | Brexiteers (top 10 tokens) | Others (top 10 tokens) | Remainders (top 10 tokens) |
|-------|-------------|-----------------------------|------------------------|---------------------------|
| 1     | 1.9%        | election, go, general, next, majority, go, away, be, next-election, next->ge | vote, go, referendum, away, decision, election, next, along, ge, campaign | go, ge, next, decision, away, suffering, campaign, next->ge, fully, go->away |
| 30    | 1.9%        | conservatives, point, 3, need, 10, unmentionable, exactly, conservatives->need, exactly->conservatives, ago | voted, need, mps, 2, point, who, first, place, something, rules | need, time, mean, point, lose->jobs, home, lose, fantasy, first, illusion |
| 27    | 1.8%        | new, two, germany, happened, europe, collapse, changing, reality, becoming, france | scotland, europe, uncertainty, 1, political, two, politics, sick, full, zero | or, getting, wants, sick, night, new, political, europe, existing, need |
| 38    | 1.8%        | general->election, or, not, just, europeanunion, deal, vote, no-deal, people, unitedkingdom | not, or, just, general->election, confirm, early->election, continued, future, keep, might | not, election, general->election, or, europeanunion, unitedkingdom, deal, people, just, no-deal |
| 6     | 1.6%        | if, one, happens, every, parties, art50, wonder, day, every->day, sport | if, every, day, mo, wales, wonder, avoid, happens, quickly, course | if, every, one, happens, apart, day, wonder, every->day, avoid, ministers |
| 8     | 1.6%        | leave, 31, leave->europeanunion, worst, benn, october, ready, case, mild, disruption | leave, case, worst, issue, show, foreign, office, pass, says, voted->leave | act, case, momentum, benn, benn->act, designed, boss, roll, worst, facts |
| 12    | 1.8%        | act, backstop, agreed, withdrawal, choice, agreement, situation, agree, withdraw->act, ask | extension, backstop, agreement, choice, act, withdrawal, since, revoke, ask, drop | everything, 50, fall, career, choice, article, pound, agreement, limit, force |
| 22    | 1.6%        | outbreak, scurry, outbreak->scurvy, wait, death, enough, available, streets, across, mailonline | 50, article, article->50, death, across, outbreak, scurry, outbreak->scurvy, wait, streets | prime, prime->minister, minister, available, easily, across->country, agree, streets, across, cameron |
| 13    | 1.5%        | sure, remainers, date, leavers, maybe, almost, come, certain, massive, hopefully | sure, never, ever, enough, almost, leavers, gets, march, certain | default, government, revokes50, option, referendum, legal, unitedkingdom->government, come, sure, revokes50->default |
| 4     | 1.1%        | live, blocking->roads, farmers->blocking, blocking, roads->protest, protest, farmers, roads, news, move | news, next, live, move, farmers, live->news, 4, stage, blocking, next->stage | business, stage, lost, news, live, next, roads, move->next, next->stage, stage->live |
| 14    | 1.1%        | clear, referendum, accept, nothing, name, behind, ridiculous, second, mode, perfectly->clear | clear, nothing, second, accept, major, ref, second->referendum, hands, behind,indy | clear, nothing, 2nd, fact, accept, table, ref, behind, second, ridiculous |
| 29    | 0.9%        | not, time, losing, claims, pointless, votes, or->not, society, something, commons | default, disaster, legal, votes, greater, damage->economy, position, whether, society, legislation | just, not->just, society, revocation, greater, not, less->competitive, competitive, economically, power |
Table 3: Estimated topic propensity parameters

| topic | term          | estimate | std.error | statistic | p.value |
|-------|---------------|----------|-----------|-----------|---------|
| 1     | Constant      | 0.0160   | 0.0023    | 7.0273    | 0.0000  |
| 1     | rel.type=Creation | 0.0083   | 0.0008    | 10.8517   | 0.0000  |
| 1     | rel.type=Destruction | 0.0078   | 0.0011    | 7.0078    | 0.0000  |
| 1     | neg=TRUE      | 0.0165   | 0.0025    | 6.4734    | 0.0000  |
| 1     | fct=Other     | 0.0006   | 0.0023    | 0.2592    | 0.7955  |
| 1     | fct=Remainer  | 0.0012   | 0.0034    | 0.3577    | 0.7206  |
| 2     | Constant      | 0.0405   | 0.0032    | 12.5478   | 0.0000  |
| 2     | rel.type=Creation | -0.0394  | 0.0013    | -31.0073  | 0.0000  |
| 2     | rel.type=Destruction | -0.0387  | 0.0016    | -24.5399  | 0.0000  |
| 2     | neg=TRUE      | -0.0038  | 0.0028    | -1.3641   | 0.1725  |
| 2     | fct=Other     | 0.0088   | 0.0029    | 2.9790    | 0.0029  |
| 2     | fct=Remainer  | 0.0062   | 0.0045    | 1.3652    | 0.1722  |
| 3     | Constant      | 0.0221   | 0.0023    | 9.7882    | 0.0000  |
| 3     | rel.type=Creation | 0.0100   | 0.0008    | 12.4265   | 0.0000  |
| 3     | rel.type=Destruction | 0.0244   | 0.0011    | 22.7367   | 0.0000  |
| 3     | neg=TRUE      | -0.0038  | 0.0028    | -1.3641   | 0.1725  |
| 3     | fct=Other     | 0.0020   | 0.0023    | 0.8617    | 0.3889  |
| 3     | fct=Remainer  | -0.0054  | 0.0032    | -1.7141   | 0.0865  |
| 4     | Constant      | 0.0157   | 0.0028    | 5.6770    | 0.0000  |
| 4     | rel.type=Creation | 0.0016   | 0.0009    | 1.7799    | 0.0751  |
| 4     | rel.type=Destruction | -0.0019  | 0.0013    | -1.4109   | 0.1583  |
| 4     | neg=TRUE      | -0.0073  | 0.0021    | -3.4211   | 0.0006  |
| 4     | fct=Other     | 0.0020   | 0.0029    | 0.6942    | 0.4875  |
| 4     | fct=Remainer  | -0.0028  | 0.0038    | -0.7466   | 0.4553  |
| 5     | Constant      | 0.0640   | 0.0040    | 15.8701   | 0.0000  |
| 5     | rel.type=Creation | -0.0036  | 0.0011    | -3.3920   | 0.0010  |
| 5     | rel.type=Destruction | 0.0297   | 0.0016    | 18.6011   | 0.0000  |
| 5     | neg=TRUE      | -0.0026  | 0.0023    | -1.1385   | 0.2549  |
| 5     | fct=Other     | -0.0405  | 0.0040    | -10.0968  | 0.0000  |
| 5     | fct=Remainer  | -0.0416  | 0.0049    | -8.5345   | 0.0000  |
| 6     | Constant      | 0.0193   | 0.0020    | 9.5565    | 0.0000  |
| 6     | rel.type=Creation | 0.0033   | 0.0006    | 5.1408    | 0.0000  |
| 6     | rel.type=Destruction | 0.0042   | 0.0009    | 4.6849    | 0.0000  |
| 6     | neg=TRUE      | -0.0039  | 0.0016    | -2.4214   | 0.0155  |
| 6     | fct=Other     | -0.0028  | 0.0020    | -1.4046   | 0.1601  |
| 6     | fct=Remainer  | -0.0034  | 0.0031    | -1.1154   | 0.2647  |
| 7     | Constant      | 0.0106   | 0.0025    | 4.2538    | 0.0000  |
| 7     | rel.type=Creation | -0.0005  | 0.0009    | -0.6007   | 0.5480  |
| 7     | rel.type=Destruction | 0.0412   | 0.0014    | 29.5135   | 0.0000  |
| 7     | neg=TRUE      | -0.0081  | 0.0022    | -3.6805   | 0.0002  |
| 7     | fct=Other     | 0.0085   | 0.0025    | 3.4044    | 0.0007  |
| 7     | fct=Remainer  | 0.0057   | 0.0037    | 1.5426    | 0.1229  |
| 8     | Constant      | 0.0246   | 0.0029    | 8.4942    | 0.0000  |
| 8     | rel.type=Creation | 0.0046   | 0.0009    | 5.2425    | 0.0000  |
| 8     | rel.type=Destruction | -0.0051  | 0.0010    | -4.9672   | 0.0000  |
| 8     | neg=TRUE      | 0.0008   | 0.0021    | 0.3784    | 0.7051  |
| 8     | fct=Other     | -0.0059  | 0.0029    | -2.0451   | 0.0408  |
| 8     | fct=Remainer  | -0.0146  | 0.0036    | -4.0617   | 0.0000  |
| 9     | Constant      | 0.0388   | 0.0027    | 14.2883   | 0.0000  |
| 9     | rel.type=Creation | -0.0175  | 0.0009    | -18.9675  | 0.0000  |
| 9     | rel.type=Destruction | -0.0034  | 0.0012    | -2.8020   | 0.0051  |
| 9     | neg=TRUE      | -0.0043  | 0.0023    | -1.8385   | 0.0660  |

Continued on next page
Table 3 – Continued from previous page

| topic | term | estimate | std.error | statistic | p.value |
|-------|------|----------|-----------|-----------|---------|
| 9     | fct=Other | 0.0063  | 0.0028    | 2.2785    | 0.0227  |
| 9     | fct=Remainer | 0.0091  | 0.0040    | 2.2853    | 0.0223  |
| 10    | Constant | 0.0252  | 0.0026    | 9.6413    | 0.0000  |
| 10    | rel.type=Creation | 0.0080  | 0.0007    | 10.8194   | 0.0000  |
| 10    | rel.type=Destruction | 0.0006  | 0.0010    | 0.5435    | 0.5868  |
| 10    | neg=TRUE | 0.0018  | 0.0023    | 0.7796    | 0.4356  |
| 10    | fct=Other | -0.0057 | 0.0026    | -2.2104   | 0.0271  |
| 10    | fct=Remainer | -0.0150 | 0.0034    | -4.4076   | 0.0000  |
| 11    | Constant | 0.0252  | 0.0026    | 9.6413    | 0.0000  |
| 11    | rel.type=Creation | 0.0080  | 0.0007    | 10.8194   | 0.0000  |
| 11    | rel.type=Destruction | 0.0006  | 0.0010    | 0.5435    | 0.5868  |
| 11    | neg=TRUE | 0.0018  | 0.0023    | 0.7796    | 0.4356  |
| 11    | fct=Other | -0.0057 | 0.0026    | -2.2104   | 0.0271  |
| 11    | fct=Remainer | -0.0150 | 0.0034    | -4.4076   | 0.0000  |
| 12    | Constant | 0.0252  | 0.0026    | 9.6413    | 0.0000  |
| 12    | rel.type=Creation | 0.0080  | 0.0007    | 10.8194   | 0.0000  |
| 12    | rel.type=Destruction | 0.0006  | 0.0010    | 0.5435    | 0.5868  |
| 12    | neg=TRUE | 0.0018  | 0.0023    | 0.7796    | 0.4356  |
| 12    | fct=Other | -0.0057 | 0.0026    | -2.2104   | 0.0271  |
| 12    | fct=Remainer | -0.0150 | 0.0034    | -4.4076   | 0.0000  |
| 13    | Constant | 0.0252  | 0.0026    | 9.6413    | 0.0000  |
| 13    | rel.type=Creation | 0.0080  | 0.0007    | 10.8194   | 0.0000  |
| 13    | rel.type=Destruction | 0.0006  | 0.0010    | 0.5435    | 0.5868  |
| 13    | neg=TRUE | 0.0018  | 0.0023    | 0.7796    | 0.4356  |
| 13    | fct=Other | -0.0057 | 0.0026    | -2.2104   | 0.0271  |
| 13    | fct=Remainer | -0.0150 | 0.0034    | -4.4076   | 0.0000  |
| 14    | Constant | 0.0252  | 0.0026    | 9.6413    | 0.0000  |
| 14    | rel.type=Creation | 0.0080  | 0.0007    | 10.8194   | 0.0000  |
| 14    | rel.type=Destruction | 0.0006  | 0.0010    | 0.5435    | 0.5868  |
| 14    | neg=TRUE | 0.0018  | 0.0023    | 0.7796    | 0.4356  |
| 14    | fct=Other | -0.0057 | 0.0026    | -2.2104   | 0.0271  |
| 14    | fct=Remainer | -0.0150 | 0.0034    | -4.4076   | 0.0000  |
| 15    | Constant | 0.0252  | 0.0026    | 9.6413    | 0.0000  |
| 15    | rel.type=Creation | 0.0080  | 0.0007    | 10.8194   | 0.0000  |
| 15    | rel.type=Destruction | 0.0006  | 0.0010    | 0.5435    | 0.5868  |
| 15    | neg=TRUE | 0.0018  | 0.0023    | 0.7796    | 0.4356  |
| 15    | fct=Other | -0.0057 | 0.0026    | -2.2104   | 0.0271  |
| 15    | fct=Remainer | -0.0150 | 0.0034    | -4.4076   | 0.0000  |
| 16    | Constant | 0.0252  | 0.0026    | 9.6413    | 0.0000  |
| 16    | rel.type=Creation | 0.0080  | 0.0007    | 10.8194   | 0.0000  |
| 16    | rel.type=Destruction | 0.0006  | 0.0010    | 0.5435    | 0.5868  |
| 16    | neg=TRUE | 0.0018  | 0.0023    | 0.7796    | 0.4356  |
| 16    | fct=Other | -0.0057 | 0.0026    | -2.2104   | 0.0271  |
| 16    | fct=Remainer | -0.0150 | 0.0034    | -4.4076   | 0.0000  |
| 17    | Constant | 0.0252  | 0.0026    | 9.6413    | 0.0000  |
| 17    | rel.type=Creation | 0.0080  | 0.0007    | 10.8194   | 0.0000  |
| 17    | rel.type=Destruction | 0.0006  | 0.0010    | 0.5435    | 0.5868  |
| 17    | neg=TRUE | 0.0018  | 0.0023    | 0.7796    | 0.4356  |
| 17    | fct=Other | -0.0057 | 0.0026    | -2.2104   | 0.0271  |
| 17    | fct=Remainer | -0.0150 | 0.0034    | -4.4076   | 0.0000  |
| 18    | Constant | 0.0252  | 0.0026    | 9.6413    | 0.0000  |
| 18    | rel.type=Creation | 0.0080  | 0.0007    | 10.8194   | 0.0000  |
| 18    | rel.type=Destruction | 0.0006  | 0.0010    | 0.5435    | 0.5868  |
| 18    | neg=TRUE | 0.0018  | 0.0023    | 0.7796    | 0.4356  |
| 18    | fct=Other | -0.0057 | 0.0026    | -2.2104   | 0.0271  |
| 18    | fct=Remainer | -0.0150 | 0.0034    | -4.4076   | 0.0000  |

Continued on next page
| topic | term | estimate | std.error | statistic | p.value |
|-------|------|----------|-----------|-----------|---------|
| 18    | neg=TRUE | 0.0101  | 0.0027   | 3.7413  | 0.0002  |
| 18    | fct=Other | -0.0083 | 0.0035  | -2.3435 | 0.0191  |
| 18    | fct=Remainer | -0.0111 | 0.0043  | -2.5963 | 0.0094  |
| 19    | Constant | 0.0285  | 0.0034  | 8.3680  | 0.0000  |
| 19    | rel.type=Creation | -0.0088 | 0.0011  | -7.9786 | 0.0000  |
| 19    | rel.type=Destruction | -0.0102 | 0.0014  | -7.2765 | 0.0000  |
| 19    | neg=TRUE | -0.0085 | 0.0026  | -3.3223 | 0.0009  |
| 19    | fct=Other | 0.0035  | 0.0033  | 1.0435  | 0.2967  |
| 19    | fct=Remainer | 0.0108  | 0.0043  | 2.5015  | 0.0124  |
| 20    | Constant | 0.0412  | 0.0032  | 12.9566 | 0.0000  |
| 20    | rel.type=Creation | -0.0088 | 0.0011  | -7.9786 | 0.0000  |
| 20    | rel.type=Destruction | -0.0102 | 0.0014  | -7.2765 | 0.0000  |
| 20    | neg=TRUE | -0.0085 | 0.0026  | -3.3223 | 0.0009  |
| 20    | fct=Other | 0.0035  | 0.0033  | 1.0435  | 0.2967  |
| 20    | fct=Remainer | 0.0108  | 0.0043  | 2.5015  | 0.0124  |
| 21    | Constant | 0.0285  | 0.0034  | 8.3680  | 0.0000  |
| 21    | rel.type=Creation | -0.0088 | 0.0011  | -7.9786 | 0.0000  |
| 21    | rel.type=Destruction | -0.0102 | 0.0014  | -7.2765 | 0.0000  |
| 21    | neg=TRUE | -0.0085 | 0.0026  | -3.3223 | 0.0009  |
| 21    | fct=Other | 0.0035  | 0.0033  | 1.0435  | 0.2967  |
| 21    | fct=Remainer | 0.0108  | 0.0043  | 2.5015  | 0.0124  |
| 22    | Constant | 0.0288  | 0.0029  | 9.9498  | 0.0000  |
| 22    | rel.type=Creation | -0.0047 | 0.0011  | -23.0179 | 0.0000  |
| 22    | rel.type=Destruction | -0.0200 | 0.0014  | -14.4377 | 0.0000  |
| 22    | neg=TRUE | -0.0050 | 0.0024  | -2.0774 | 0.0378  |
| 22    | fct=Other | 0.0071  | 0.0029  | 2.4084  | 0.0160  |
| 22    | fct=Remainer | -0.0061 | 0.0038  | -1.6008 | 0.1094  |
| 23    | Constant | 0.0238  | 0.0034  | 7.0785  | 0.0000  |
| 23    | rel.type=Creation | -0.0119 | 0.0028  | -4.3161 | 0.0000  |
| 23    | rel.type=Destruction | -0.0061 | 0.0038  | -1.6008 | 0.1094  |
| 23    | neg=TRUE | 0.0022  | 0.0034  | 0.6454  | 0.5186  |
| 23    | fct=Other | 0.0017  | 0.0046  | 0.3641  | 0.7158  |
| 23    | fct=Remainer | 0.0017  | 0.0046  | 0.3641  | 0.7158  |
| 24    | Constant | 0.0354  | 0.0029  | 12.3901 | 0.0000  |
| 24    | rel.type=Creation | 0.0157  | 0.0008  | 20.1294 | 0.0000  |
| 24    | rel.type=Destruction | 0.0047  | 0.0011  | 4.3537  | 0.0000  |
| 24    | neg=TRUE | 0.0195  | 0.0026  | 7.4187  | 0.0000  |
| 24    | fct=Other | -0.0065 | 0.0028  | -2.3026 | 0.0213  |
| 24    | fct=Remainer | -0.0074 | 0.0035  | -2.1327 | 0.0330  |
| 25    | Constant | 0.0221  | 0.0023  | 9.5598  | 0.0000  |
| 25    | rel.type=Creation | -0.0010 | 0.0007  | -1.4067 | 0.1595  |
| 25    | rel.type=Destruction | 0.0142  | 0.0012  | 12.1294 | 0.0000  |
| 25    | neg=TRUE | -0.0070 | 0.0019  | -3.6501 | 0.0003  |
| 25    | fct=Other | 0.0030  | 0.0023  | 1.3099  | 0.1902  |
| 25    | fct=Remainer | 0.0059  | 0.0031  | 1.9232  | 0.0545  |
| 26    | Constant | 0.0395  | 0.0032  | 12.1924 | 0.0000  |
| 26    | rel.type=Creation | -0.0303 | 0.0012  | -24.6338 | 0.0000  |
| 26    | rel.type=Destruction | -0.0192 | 0.0017  | -11.1510 | 0.0000  |
| 26    | neg=TRUE | -0.0037 | 0.0030  | -1.2225 | 0.2215  |
| 26    | fct=Other | 0.0131  | 0.0032  | 4.1555  | 0.0000  |
| 26    | fct=Remainer | 0.0066  | 0.0043  | 1.5310  | 0.1258  |
| 27    | Constant | 0.0171  | 0.0024  | 7.0375  | 0.0000  |
| 27    | rel.type=Creation | -0.0028 | 0.0009  | -3.2322 | 0.0013  |

Continued on next page
Table 3 – Continued from previous page

| topic term | estimate | std.error | statistic | p.value |
|------------|----------|-----------|-----------|---------|
| 27 rel.type = Destruction | -0.0076 | 0.0011 | -6.6079 | 0.0000 |
| 27 neg = TRUE | -0.0040 | 0.0021 | -1.9116 | 0.0559 |
| 27 fct = Other | 0.0073 | 0.0024 | 3.0478 | 0.0023 |
| 27 fct = Remainer | 0.0066 | 0.0034 | 1.9263 | 0.0541 |
| 28 Constant | 0.0284 | 0.0033 | 8.5978 | 0.0000 |
| 28 rel.type = Creation | 0.0228 | 0.0011 | 21.6190 | 0.0000 |
| 28 rel.type = Destruction | -0.0040 | 0.0013 | -3.1698 | 0.0015 |
| 28 neg = TRUE | 0.0187 | 0.0036 | 5.1774 | 0.0000 |
| 28 fct = Other | -0.0030 | 0.0034 | -0.8910 | 0.3729 |
| 28 fct = Remainer | 0.0014 | 0.0043 | 0.3351 | 0.7376 |
| 29 Constant | 0.0211 | 0.0023 | 9.3169 | 0.0000 |
| 29 rel.type = Creation | 0.0030 | 0.0006 | 5.1266 | 0.0000 |
| 29 rel.type = Destruction | -0.0030 | 0.0008 | -3.7888 | 0.0002 |
| 29 neg = TRUE | -0.0013 | 0.0017 | -0.7728 | 0.4397 |
| 29 fct = Other | -0.0030 | 0.0034 | -0.8910 | 0.3729 |
| 29 fct = Remainer | 0.0014 | 0.0043 | 0.3351 | 0.7376 |
| 30 Constant | 0.0132 | 0.0019 | 6.9358 | 0.0000 |
| 30 rel.type = Creation | 0.0031 | 0.0007 | 4.1969 | 0.0000 |
| 30 rel.type = Destruction | -0.0005 | 0.0010 | -0.5347 | 0.5929 |
| 30 neg = TRUE | -0.0034 | 0.0018 | -1.8495 | 0.0644 |
| 30 fct = Other | 0.0064 | 0.0018 | 3.4516 | 0.0006 |
| 30 fct = Remainer | -0.0041 | 0.0025 | -1.5969 | 0.1103 |
| 31 Constant | 0.0367 | 0.0029 | 12.5020 | 0.0000 |
| 31 rel.type = Creation | -0.0219 | 0.0011 | -20.8501 | 0.0000 |
| 31 rel.type = Destruction | -0.0253 | 0.0013 | -19.0093 | 0.0000 |
| 31 neg = TRUE | -0.0013 | 0.0017 | -0.7728 | 0.4397 |
| 31 fct = Other | 0.0063 | 0.0029 | 2.1692 | 0.0301 |
| 31 fct = Remainer | -0.0018 | 0.0041 | -0.4389 | 0.6607 |
| 32 Constant | 0.0315 | 0.0021 | 12.2313 | 0.0000 |
| 32 rel.type = Creation | -0.0247 | 0.0009 | -26.9047 | 0.0000 |
| 32 rel.type = Destruction | 0.0189 | 0.0014 | 13.8506 | 0.0000 |
| 32 neg = TRUE | -0.0053 | 0.0025 | -2.1541 | 0.0312 |
| 32 fct = Other | 0.0113 | 0.0026 | 4.4257 | 0.0000 |
| 32 fct = Remainer | 0.0278 | 0.0046 | 6.0781 | 0.0000 |
| 33 Constant | 0.0075 | 0.0021 | 2.9559 | 0.0031 |
| 33 rel.type = Creation | 0.0230 | 0.0009 | 26.6673 | 0.0000 |
| 33 rel.type = Destruction | 0.0296 | 0.0013 | 23.6445 | 0.0000 |
| 33 neg = TRUE | -0.0117 | 0.0021 | -5.4961 | 0.0000 |
| 33 fct = Other | 0.0108 | 0.0025 | 4.3929 | 0.0000 |
| 33 fct = Remainer | 0.0213 | 0.0038 | 5.6308 | 0.0000 |
| 34 Constant | 0.0232 | 0.0022 | 10.5312 | 0.0000 |
| 34 rel.type = Creation | 0.0100 | 0.0007 | 15.1427 | 0.0000 |
| 34 rel.type = Destruction | 0.0009 | 0.0010 | 0.9221 | 0.3565 |
| 34 neg = TRUE | 0.0021 | 0.0021 | 1.0075 | 0.3137 |
| 34 fct = Other | -0.0034 | 0.0022 | -1.5807 | 0.1140 |
| 34 fct = Remainer | -0.0021 | 0.0029 | -0.7178 | 0.4729 |
| 35 Constant | 0.0109 | 0.0021 | 5.2107 | 0.0000 |
| 35 rel.type = Creation | 0.0189 | 0.0008 | 24.1147 | 0.0000 |
| 35 rel.type = Destruction | 0.0021 | 0.0009 | 2.1888 | 0.0286 |
| 35 neg = TRUE | -0.0056 | 0.0020 | -2.7087 | 0.0056 |
| 35 fct = Other | 0.0074 | 0.0021 | 3.4534 | 0.0006 |
| 35 fct = Remainer | 0.0009 | 0.0031 | 0.2990 | 0.7649 |
| 36 Constant | 0.0069 | 0.0021 | 3.2615 | 0.0011 |

Continued on next page
### Table 3 – Continued from previous page

| topic | term | estimate | std.error | statistic | p.value |
|-------|------|----------|-----------|-----------|---------|
| 36    | rel.type=Creation | 0.0127 | 0.0008 | 15.9608 | 0.0000 |
| 36    | rel.type=Destruction | 0.0062 | 0.0011 | 5.8633 | 0.0000 |
| 36    | neg=TRUE | -0.0073 | 0.0019 | -3.8505 | 0.0001 |
| 36    | fct=Other | 0.0098 | 0.0021 | 4.7018 | 0.0000 |
| 36    | fct=Remainer | 0.0179 | 0.0033 | 5.4723 | 0.0000 |
| 37    | Constant | 0.0302 | 0.0023 | 13.1362 | 0.0000 |
| 37    | rel.type=Creation | -0.0262 | 0.0009 | -30.1670 | 0.0000 |
| 37    | rel.type=Destruction | -0.0225 | 0.0012 | -18.9386 | 0.0000 |
| 37    | neg=TRUE | -0.0001 | 0.0023 | -0.0411 | 0.9672 |
| 37    | fct=Other | 0.0118 | 0.0023 | 5.0894 | 0.0000 |
| 37    | fct=Remainer | 0.0131 | 0.0035 | 3.7157 | 0.0002 |
| 38    | Constant | 0.0074 | 0.0004 | 16.5572 | 0.0000 |
| 38    | rel.type=Creation | 0.0002 | 0.0002 | 1.0797 | 0.2803 |
| 38    | rel.type=Destruction | -0.0003 | 0.0002 | -1.2309 | 0.2184 |
| 38    | neg=TRUE | -0.0001 | 0.0004 | -0.3539 | 0.7234 |
| 38    | fct=Other | 0.0039 | 0.0005 | 8.6132 | 0.0000 |
| 38    | fct=Remainer | 0.0016 | 0.0006 | 2.5763 | 0.0100 |

Figure 1. Significant estimated differences in argument $k$ proportions (with 99% C.I.) of as a function of causal relation type $rel$: $\hat{\beta}_{k,rel=Destruction} - \hat{\beta}_{k,rel=Creation}$. 

"CREATION" ARGUMENT(S)  
"DESTRUCTION" ARGUMENT(S)
Figure 2. Significant estimated differences in topic $k$ proportions (with 99% C.I.) of as a function of partisan faction $fct$: $\hat{\beta}_{k,fct}^{\text{Brexiteer}} - \hat{\beta}_{k,fct}^{\text{Remaininer}}$.

Figure 3. Significant estimated differences in topic $k$ proportions (with 99% C.I.) of as a function of relation verb negation $neg$: $\hat{\beta}_{k,neg=\text{TRUE}} - \hat{\beta}_{k,neg=\text{FALSE}}$. 
2. Robustness Checks

Table 4. Parameters used in experiments with different values for $K$ from 3 to 70

| Parameter                                      | Value                  |
|------------------------------------------------|------------------------|
| $K$                                            | c(3:70)                |
| Number of experiment replications              | 50                     |
| Topic prevalence                              | rel.type + neg + fct + s(t) |
| Topic content                                  | fct                    |
| Maximum Expectation Maximization Iterations   | 200                    |
| Expectations Maximization Tolerance           | 1e-05                  |
| Initialization Type                           | Spectral               |
| Proportion of docs to be held out              | 25%                    |
| $M$                                            | 10                     |

Figure 4. Mean values of Lower bound and Heldout likelihood (based on 50 replications) of estimated STM different values of $K$. Labels represent the number of topics $K$. At each replication a different random sample containing a 25% of the population of no-deal effects has been heldout. The heldout sample was hence used to compute the Heldout likelihood.
Figure 5. Mean values of the STM *Lower bound* as a function of $K$. At each replication a different random sample containing a 25% of the population of no-deal effects has been heldout. The heldout sample was hence used to compute the *Heldout likelihood*.

Figure 6. Mean values of the STM *Heldout likelihood* as a function of $K$. At each replication a different random sample containing a 25% of the population of no-deal effects has been heldout. The heldout sample was hence used to compute the *Heldout likelihood*. 
3. Methodology

Figure 7. Workflow summary

Table 5. List of stopwords (based on Quanteda stopword list) used in STM estimation step

| stopwords |
|-----------|
| "i", "me", "my", "myself", "we", "our", "ours", "ourselves", "you", "your", "yours", "yourself", "yourselves", "he", "him", "his", "himself", "she", "her", "hers", "herself", "it", "its", "itself", "they", "them", "their", "theirs", "themselves", "what", "which", "who", "whom", "this", "that", "these", "those", "am", "is", "are", "was", "were", "be", "been", "being", "have", "has", "had", "having", "do", "does", "did", "doing", "would", "should", "could", "ought", "i'm", "you're", "he's", "she's", "it's", "we're", "they're", "i've", "you've", "we've", "they've", "i'd", "you'd", "he'd", "she'd", "we'd", "they'd", "i'll", "you'll", "he'll", "she'll", "we'll", "they'll", "isn't", "aren't", "wasn't", "weren't", "hasn't", "haven't", "hadn't", "doesn't", "don't", "didn't", "won't", "wouldn't", "shan't", "shouldn't", "can't", "cannot", "couldn't", "mustn't", "let's", "let's", "that's", "who's", "what's", "here's", "there's", "when's", "where's", "why's", "how's", "a", "an", "the", "and", "but", "because", "as", "until", "while", "of", "at", "by", "for", "with", "about", "against", "between", "into", "through", "during", "before", "after", "above", "below", "to", "from", "up", "down", "in", "out", "on", "over", "under", "again", "further", "then", "once", "here", "there", "when", "where", "why", "how", "all", "any", "both", "each", "few", "more", "most", "other", "some", "such", "only", "own", "same", "so", "than", "too", "very", "will", "a", "an", "the", "&", "\&", "e", "\&", "via", "also", "think", "simply", "\&", "re", "youwatch", "whomost", "nah", "itbe", "amongst", "lrast", "youto", "theof", "mayhem", "amlo", "oui", "oh", "kap", "ect", "bla", "try", "yet", "thats", "got", "hoe", "lot", "it"
Table 6. RegEx used to identify causes with "no-deal" in the last four tokens of the cause-side

| RegEx                                                                 |
|----------------------------------------------------------------------|
| destroy/destroyed/destroying/demolishing/destroying/oblitrate/obliterate/cancel/cancelled/canceling/cancelling/eradicating/truncate/eradicate/eradicating/extricate/extirpate/eliminate/eliminates/eliminating/eradicate/extirpates/wipe/wipe/uni2423out/wipes/wipe/uni2423out/wiped/wipe/uni2423out/wiping/wipe/uni2423out/kill/kills/killed/losing |
SUPPLEMENT - STATISTICAL RHETORIC OF ONLINE ARGUMENTS & DEBATES 17
perl = T,
ignore.case = T)
contains_end_form[i] = TRUE)

# identify and extract end form
end_form[i]["effect"] = gsub(".*\([^,:\]\{1,\}[^,:\]\{1,\}.*\[^,:\]\{1,\}"), "\1\$", \1\$"), perl = T,
ignore.case = T)

# end form pruning
end_form[i] = gsub("(.*\([^,:\]\{1,\}[^,:\]\{1,\}.*\[^,:\]\{1,\}"), "\1\$", \1\$"), sep = ",", end_form[i]), perl = T,
ignore.case = T)

# identify different parts of the end form
are_end_form[i] = grepl("that\_passive\_end\_regex\_",
end_form[i]["effect"]
perl = T,
ignore.case = T)

# extract matches of the end form
define_end_form[i] = regcapturedmatches\(Define\_end\_form[i]["are\_end\_form[i]["effect"]\]
if(\# Define the effect_end to the effect
end_form[i]["effect"] = paste(\# Define the effect_end to the effect
define_end_form[i]["effect"] = paste(\# Create relation metadata
negated_end_form[i] = grepl(\# Create relation metadata
negated_end_form[i]["effect"] = grepl(\# Clean the effect side
define_end_form[i]["effect"]
perl = T)

# clean the cause side
define_end_form[i]["cause"] = gsub("\[^,:\]\{1,\}"), "\1\$", \1\$"), perl = T,
ignore.case = T)

# clean the cause side
define_end_form[i]["cause"] = gsub("\[^,:\]\{1,\}"), "\1\$", \1\$"), perl = T,
ignore.case = T)

# identify the sentence fragment in which the end_form appears
end_form[i]["effect"] = base::strsplit\(y[i]["effect"]", "\(<[^,:\]\{1,\}>\)"), perl = TRUE)\)
contained_end_form[i] = isqdf\(\# Identify the sentence fragment in which the end_form appears
end_form[i]["effect"]
perl = T,
ignore.case = T)

if(length\(end_form[i]["effect"]\) > 1 &
sum\(contained_end_form[i]["effect"]\) >= 1){

# if multiple fragments contained in sentence prune end_form fragment before next processing step
y[i]["effect"] = paste\(end_form[i]["effect"]", sep = ",", collapse = "\"")
else{
  contained_end_form[i]["effect"] = FALSE
}
else{
  contained_end_form[i]["effect"] = FALSE
}

### passive and normal forms ###
if(grepl(pattern = "[^:;,!]{4,}",y[i],perl = T, ignore.case = T)){

#extract relation's operators
operators[i][] = regmatches(y[i],
  gregexpr(  
    paste(passive_pattern, pattern, sep = ""),
    y[i],
    perl = T,
    ignore.case = T
  ))[1]

#name the operators NORMAL as default
names(operators[i][]) = rep("NORMAL", length(operators[i][]))

#identify which ones are passive
are_passive[i] = grepl(passive_pattern, unlist(operators[i][]),
  perl = T,
  ignore.case = T)

#name these ones PASSIVE
names(operators[i][are_passive[i]]) = "PASSIVE"

#if there is at least one passive process string as follows
if(length(operators[i][are_passive[i]]) > 0) {
  splits[i][] = regmatches(  
    y[i],
    gregexpr(  
      RegEx(operators[i][are_passive[i]], operators[i][are_passive[i]]),
      y[i],
      perl = T,
      ignore.case = T
    ),
    invert = T
  )[1]
}
else{  
  #if no passive process string as follows
  splits[i][] = regmatches(  
    y[i],
    gregexpr(  
      RegEx(operators[i]),
      y[i],
      perl = T,
      ignore.case = T
    ),
    invert = T
  )[1]
}

#####BUILD PAIRS#####
if(length(operators[i][]) > 0) {
  splits[i][] = gsub("^[[:punct:][:space:]]{1}:([[:punct:][:space:]]{1}$","
  splits[i][]
  negation[i][] = grepl(  
    negators_modals_verbs_regex_,
    splits[i][],
    ignore.case = T,  
    perl = T,
    invert = T
  )
  splits[i][] = gsub(  
    negators_modals_verbs_regex_,
    "",  
    splits[i][],
    ignore.case = T,
    perl = T
  )
  splits[i][] = gsub(modals_verbs_regex_,
  "",  
  splits[i][],
  ignore.case = T,
  perl = T
  )
  splits[i][] = gsub(and_regex,
  "",  
  splits[i][],
  perl = T,
  ignore.case = T
  )
  and[i][] = grepl(and_regex,  
    splits[i][],
    perl = T,
    ignore.case = T
  )
  splits[i][] = gsub(  
    and_regex,
    "",  
    splits[i][],
    perl = T,
    ignore.case = T
  )
  which[i][] = grepl("which\[[:space:]]\{0,\}$",  
    splits[i][],
    ignore.case = T,
    perl = T
  )
  splits[i][] = gsub(  
    "which\[[:space:]]\{0,\}\{0,\}\{0,\}$",  
    "",  
    splits[i][],
    ignore.case = T,
    perl = T
  )
splits[[i]], perl = T, ignore.case = T)
splits[[i]] = gsub("^[[:space:]]+:punct:[[:space:]]+:punct:[0-9]*", "+", splits[[i]], perl = T, ignore.case = T)
before[[i]] = splits[[i]][1:length(operators[[i]])]
after[[i]] = splits[[i]][2:length(operators[[i]])] + 1]
for (j in 1:n_relations[[i]]) {
  k = j
  if (!are_passive[[i]][j]) {
    if (j == 2 & (and[[i]][j] | which[[i]][j] - 1)) {
      k = 1
      before[[i]][j] = splits[[i]][k]
      after[[i]][j] = splits[[i]][j + 1]
    } else {
      if (j > 2 & (and[[i]][j] | which[[i]][j] - 1)) {
        k = max(which(!and[[i]][1:j - 1] & !which[[i]][1:j - 1]))
        before[[i]][j] = splits[[i]][k]
        after[[i]][j] = splits[[i]][j + 1]
      } else {
        after[[i]][j] = splits[[i]][k]
        before[[i]][j] = splits[[i]][j + 1]
      }
    } else {
      if (j == 2 & (and[[i]][j] | which[[i]][j] - 1)) {
        k = 1
        after[[i]][j] = splits[[i]][k]
        before[[i]][j] = splits[[i]][j + 1]
      } else {
        if (j > 2 & (which[[i]][j] - 1)) {
          k = max(which(!which[[i]][1:j - 1]))
          after[[i]][j] = splits[[i]][k]
          before[[i]][j] = splits[[i]][j + 1]
        } else {
          after[[i]][j] = splits[[i]][k]
          before[[i]][j] = splits[[i]][k + 1]
        }
      }
    }
  }
}
causal_data[[i]] = data.frame(
id = names[i],
  cause = before[[i]][1:n_relations[[i]]],
  rel_negation = negation[[i]][1:n_relations[[i]]],
  rel_operator = operators[[i]][1:n_relations[[i]]],
  rel_passive_form = are_passive[[i]][1:n_relations[[i]]],
  rel_creation = grep(creation_verb_regex_match_all, operators[[i]][1:n_relations[[i]]], ignore.case=T, perl=T),
  rel_destruction = grep(destruction_verb_regex_match_all, operators[[i]][1:n_relations[[i]]], ignore.case=T, perl=T),
  rel_causation = grep(cause_verb_regex_match_all, operators[[i]][1:n_relations[[i]]], ignore.case=T, perl=T),
  rel_coref_res = and[[i]][1:n_relations[[i]]] | which[[i]][1:n_relations[[i]]],
  effect = after[[i]][1:n_relations[[i]]],
  stringsAsFactors = F, check.names = F, fix.empty.names = F, row.names = NULL
if(contains_end_form[[i]]){
  causal_data[[i]] = tryCatch(rbind(are_end_form[[i]], causal_data[[i]]))
} else{
  if(contains_end_form[[i]]){
    causal_data[[i]] = tryCatch(rbind(are_end_form[[i]]))
  } else{
    if(contains_end_form[[i]]){
      causal_data[[i]] = tryCatch(rbind(are_end_form[[i]]))
    }
  }
}))
tryCatch(dplyr::bind_rows(causal_data))

Ca' Foscari University of Venice, Cannaregio 873, 30123 Venice, Italy
Email address: carlo.santagiustina@unive.it

Ca' Foscari University of Venice, Cannaregio 873, 30123 Venice, Italy
Email address: warglien@unive.it