

War of Words: The Competitive Dynamics of Legislative Processes

Victor Kristof
EPFL

Matthias Grossglauser
EPFL

Patrick Thiran
EPFL

ABSTRACT

A body of law is an example of a dynamic corpus of text documents that are jointly maintained by a group of editors who compete and collaborate in complex constellations. Our goal is to develop predictive models for this process, thereby shedding light on the competitive dynamics of parliamentarians who make laws. For this purpose, we curated a dataset of 450 000 legislative edits introduced by European parliamentarians over the last ten years. An *edit* modifies the status quo of a law, and could be in competition with another edit if it modifies the same part of that law. We propose a model for predicting the success of such edits, in the face of both the *inertia* of the status quo and the *competition* between overlapping edits. The parameters of this model can be interpreted in terms of the influence of parliamentarians and of the controversy of laws.

ACM Reference Format:

Victor Kristof, Matthias Grossglauser, and Patrick Thiran. 2020. War of Words: The Competitive Dynamics of Legislative Processes. In *Proceedings of The Web Conference 2020 (WWW '20)*, April 20–24, 2020, Taipei, Taiwan. ACM, New York, NY, USA, 7 pages. <https://doi.org/10.1145/3366423.3380041>

1 INTRODUCTION

The past decade has seen the emergence of several open-government initiatives, such as the one proposed by the former US president, Barack Obama, on the first day of his presidency [6]. Similar efforts followed in France, Belgium, Sweden, the United Kingdom, Tunisia, and the European Union. Open-government data published on the Web are of great interest to citizens, companies, sub- and supra-government entities, and researchers. These initiatives increase the transparency and trust associated with government, and enable novel analyses to be carried out about their processes.

We are specifically interested in the process by which the laws of a jurisdiction are maintained and in how they evolve over time. Not surprisingly, the dynamics of the legislative process is complex, given the confluence of many stakeholders, topics, special interests, and lobbying groups. Until open-government was introduced, the work of parliaments had not been systematically accessible to the general public, and internal documents – when they existed – were difficult to find. We propose (i) a new dataset of legislative edits obtained via rich, openly accessible data from the European Parliament, and (ii) a methodology to better understand the dynamics of legislative processes. Using edits proposed by parliamentarians on

legislative texts, we study the competitive dynamics of collaborations and conflicts between parliamentarians.

We curate our dataset from the European Parliament’s online document repository. It is composed of edits, proposed by parliamentarians, on laws under consideration by the Parliament. Each data point consists of edit metadata, such as the nationality and the political affiliation of its author(s), the type of edit, its length, and which law it is modifying. The dataset contains 449 493 edits proposed by 1 214 parliamentarians on 1 889 dossiers over ten years (two legislature periods). In Section 2, we set the framework by giving some background on the European legislative process. In Section 3, we describe our dataset in detail. In Section 4, we use our dataset to describe the evolution of a law via a graph-theoretical viewpoint.

Our model focuses on the interplay of collaboration and competition between parliamentarians as they modify laws. They can collaborate on a proposed modification of a law by jointly submitting an *edit* for consideration. An important feature of our model accounts for the way an edit benefits from the support of multiple parliamentarians. We posit a measure of *strength* for each parliamentarian, and an edit inherits the strengths of its supporters. There are two sources of competition in the process. First, a proposed edit competes with the status quo, because the edit can be rejected in favor of not changing the existing state of a law. Our model incorporates this by endowing each law with a measure of *inertia* that represents the level of controversy of a law. Second, proposed edits of a law are frequently mutually exclusive, because they overlap and are incompatible. These edits then compete against each other, as well as against the status quo. This parsimonious set of assumptions underlies our model, formulated in Section 5; and we will show, in Section 6, that it is sufficient to capture the salient features of the law-making dynamics.

2 THE EUROPEAN LEGISLATIVE PROCESS

2.1 Representative Democracies

In representative democracies, citizens elect politicians to represent them in the various branches of the government. The executive branch is in charge of executing and enforcing the laws. Representatives of the executive branch can also propose new laws, but, to avoid a concentration of power, they cannot pass new legislation without the approval of the legislative branch. The legislative branch, typically a parliament, represents both the people and the sub-governmental entities (such as states and municipalities). Parliamentarians can propose new legislation or amend propositions made by the executive branch. Finally, the judicial branch balances the power of the executive branch and the legislative branch through its ability to decide whether the laws are constitutional.

Here, we focus on the European Union (EU). The EU is a political and economic union of 28 countries called member states.

This paper is published under the Creative Commons Attribution 4.0 International (CC-BY 4.0) license. Authors reserve their rights to disseminate the work on their personal and corporate Web sites with the appropriate attribution.

WWW '20, April 20–24, 2020, Taipei, Taiwan

© 2020 IW3C2 (International World Wide Web Conference Committee), published under Creative Commons CC-BY 4.0 License.

ACM ISBN 978-1-4503-7023-3/20/04.

<https://doi.org/10.1145/3366423.3380041>

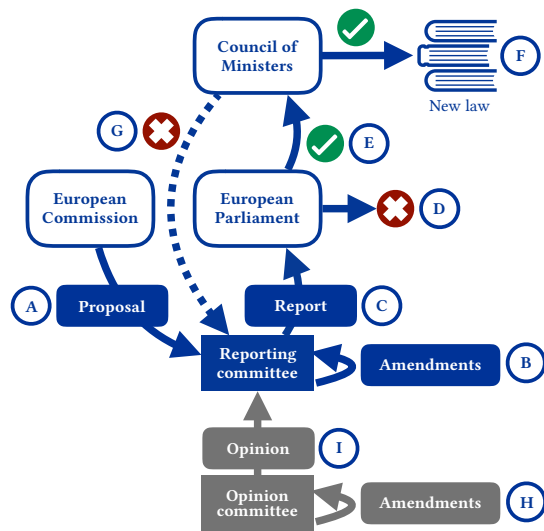


Figure 1: Sketch of the ordinary legislative procedure. (A) The Commission submits a legislative proposal to one of the Parliament committees. (B) The proposal is amended and (C) submitted to vote to the whole Parliament. (D) If it is rejected, the proposal is abandoned. (E) If it is accepted, it is transferred to the Council. (F) If the Council accepts the amended proposal, a new law is adopted. (G) If the Council amends it, it is sent back to the committee. (H) Other committees can optionally make amendments and (I) suggest them to the reporting committee.

This union enables them to share their markets, to ease mobility across borders, to favor economic development, and to harmonize laws. The EU covers an estimated population of 513 million, and up to 84% of member states' national laws emanate from the EU [9]. Hence, EU laws have a significant impact on the life of many people. European institutions make efforts to be transparent. They make a lot of valuable data available online: parliamentary amendments, meetings by the commissioners with civil society, and a transparency register to monitor interest groups.

The EU political system is broadly similar to that of a regular state. The 751 parliament representatives (MEPs, for Member of the European Parliament) are elected every five years by universal suffrage. The executive branch is called the *European Commission*. The legislative branch consists of the *European Parliament* and of the *Council of Ministers*. The Parliament is divided into 20 committees, comprising sub-sets of MEPs and specialized in some particular policy area (such as fisheries, judiciary affairs, transportation, and trade). Each MEP is a member of at least one committee. The myriad of national parties aggregate into a small number of political groups.

2.2 The Ordinary Legislative Procedure

We now describe the EU legislative process in some detail, leading up to our modeling assumptions. Under the Treaty of Lisbon [10], which marks the beginning of the 7th legislature in 2009, the Parliament's powers were increased. The Parliament became central

in the process through which new laws are created. This process can take the form of various procedures, the main one being the *ordinary legislative procedure* (OLP) [11]. Through the OLP, the Commission initiates a legislative proposal, and the Parliament must adopt it in order for the proposal to become a law. Other procedures exist, where the Parliament is not necessarily involved. Since 2009, the Parliament has dealt with 90% of all new laws via the OLP. In this regard, we focus on the dynamics of the legislative process in the Parliament. A sketch of the OLP is illustrated in Figure 1 and described in the next paragraphs.

To create a new law, (A) the Commission drafts a legislative *proposal* and transfers it to the corresponding committee of the Parliament. For instance, if the proposal introduces regulations on greenhouse-gas emissions, it is transferred to the Environment Committee. The committee appoints a *rapporteur* to lead the debate. The role of the committee is to write a *report* in the form of *amendments* to the proposal, i.e., insertions in or deletions of parts of the proposal. The rapporteur first seeks external expertise to draft a report. Then, (B) other MEPs on the committee can in turn propose amendments to the proposal. To constitute the final report to be submitted to the whole Parliament, each amendment by the rapporteur or by other MEPs is therefore voted on within the committee. Once the committee finds a consensus, (C) they transfer the report to the whole Parliament.

In the plenary session, the Parliament holds a vote on the report. (D) If rejected, the proposal is abandoned; (E) if accepted, the report, establishing the Parliament's position on the proposal, is transferred to the Council of Ministers. The report is therefore an important document and the rapporteur has an important role to play. The ministers (of the different EU countries) can accept the report, (F) in which case, the proposal is adopted with the Parliament's amendments and a new law is created; or they can make amendments, (G) in which case it is transferred back to the parliamentary committee. At this stage, we say that a law has gone through the first *reading*.

Other committees can also independently decide to address an *opinion* to the reporting committee. For instance, the Transportation Committee might consider that it is also concerned by greenhouse-gas emissions and that it is entitled to give its opinion to the Environment Committee. An opinion is similar to a report in that it contains amendments to the proposal. It is created similarly to a report, i.e., (H) the opinion committee appoints a rapporteur to draft an opinion, and other MEPs can propose amendments. (I) The opinion committee then transfers its opinion to the reporting committee. An opinion differs from a report in that it is not voted by the whole Parliament. Amendments to the opinions are, however, valuable to the reporting committee that often takes them into consideration. We refer to reports and opinions as *dossiers*.

This iterative process can be repeated up to three times (three readings). The third reading, called conciliation, involves a negotiation between the Parliament and the Council. During the 8th legislature for example, 99% of all laws were adopted after the first reading, i.e., after amendments made by both the Parliament and the Council, and 89% were adopted directly after amendments by the Parliament, i.e., the Council accepted it without making amendments.

Table 1: Descriptive statistics of the WAR OF WORDS dataset.

| Legislature | # amendments | # edits | # conflicts | # MEPs | # dossiers | # data points | % accepted |
|-----------------|--------------|---------|-------------|--------|------------|---------------|------------|
| EP7 (2009–2014) | 108 292 | 200 407 | 40 302 | 761 | 1 089 | 126 417 | 37.7% |
| EP8 (2014–2019) | 128 885 | 249 086 | 56 298 | 791 | 800 | 141 034 | 25.7% |

| | | | |
|---|--|--|--|
| <p>Amendment 108 Jolanta Emilia Hibner, Małgorzata Handzlik, Bogusław Sonik Recital 16</p> <p><i>Text proposed by the Commission</i></p> <p>(16) [...] The use of additives necessary for manufacturing of tobacco products <i>should</i> be allowed, <i>as long as they do not result in a characterising flavour</i>. The Commission should ensure uniform conditions for the implementation of the provision on characterising flavour. [...]</p> | | <p><i>Amendment</i></p> <p>(16) [...] The use of additives necessary for manufacturing of tobacco products <i>shall</i> be allowed. The Commission should ensure uniform conditions for the implementation of the provision on characterising flavour. [...]</p> | |
| <p>Amendment 5 Klaus-Heiner Lehne Recital 16</p> <p><i>Text proposed by the Commission</i></p> <p>(16) [...] The use of additives necessary for manufacturing of tobacco products <i>should be</i> allowed, <i>as long as they do not result in a characterising flavour</i>. The Commission should ensure uniform conditions for the implementation of the provision on characterising flavour. [...]</p> | | <p><i>Amendment</i></p> <p>(16) [...] The use of additives necessary for manufacturing of tobacco products <i>is</i> allowed. The Commission should ensure uniform conditions for the implementation of the provision on characterising flavour. [...]</p> | |

Figure 2: Example of two conflicting amendments in their raw format. (Top) Am. 108 is proposed by three MEPs on a text legislating tobacco-related products. (Bottom) Am. 5 is proposed by another MEP on the same text. The two edits of Am. 108, replacing "should" by "shall" and removing the end of the first sentence, are rejected. The first edit is in conflict with the first edit of Am. 5, proposing to replace "should be" by "is". This edit is also rejected.

3 DATA

We collected a dataset of 237 177 legislative amendments from the European Parliament website.¹ The dataset spans the 7th legislature (referred to as EP7), from 2009 to 2014, and the 8th legislature (EP8), from 2014 to 2019. MEPs come from 28 different countries, and they belong to one of the 8 (EP7) or 9 (EP8) political groups. We show in Figure 2 an example of a raw amendment. An amendment consists of (i) one or several authors, (ii) the original text by the European Commission, and (iii) the amended text by the author(s). MEPs propose amendments on a specific article of the legislation, and they can modify several parts within a single amendment. As a result, we decompose the difference between the original and the amended text into one or several *edits*, as defined below. We summarize our dataset in Table 1 and we refer to it as the WAR OF WORDS dataset. In the next paragraphs, we describe the data that we extract from amendments and that we use for the subsequent analysis. Technical details about data processing are given in Appendix A.

¹Data and code publicly available on <https://github.com/indy-lab/war-of-words>.

Edits. An edit is a sequence of words that are inserted or deleted or both. We extract edits by computing the *diff*, i.e., the difference between the words in two texts, between the original and the amended text of each amendment. We normalize the texts by removing special characters and by putting the words in lower case. We keep punctuation because the structure of sentences is important in legal texts. We merge identical edits proposed by different MEPs, thus considering them as one edit proposed by all authors together. This is in line with the Rules of Procedure of the Parliament [12]. We extract 200 407 edits for EP7 and 249 086 edits for EP8. On average, there are 1.85 and 1.93 edits per amendment for EP7 and EP8, respectively. There are also more dossiers in EP7 than in EP8, which means that there are proportionally more edits per dossier in EP8.

Conflicts. There exists an inherent competition between the MEPs in the amending process, as amendments are vehicles of political ideas and interests. We are therefore interested in the conflicts between edits. We define a *conflict* as a set of edits that overlap. Edits overlap because they modify parts of the text at the same position. We extract 40 302 conflicts for EP7 and 56 298 for EP8. Adding the conflicts to isolated edits, we obtain a dataset of 126 417 data points for EP7 and 141 034 data points for EP8.

Labels. The votes on each edit are not publicly available, and we need to infer their outcomes from the raw data. Reports and opinions contain only the amendments accepted within the committees. Draft reports, draft opinions, and other documents containing all proposed amendments are published separately. Therefore, if the edits extracted from the latter documents appear in the former documents, we label them as *accepted*, i.e., the committee votes to include these edits in their report or opinion. Otherwise, we label them as *rejected*. Out of the proposed edits, 37.7% are accepted for EP7 and 25.7% for EP8.

Timestamps. The timeline of the legislative process described in Section 2 varies from one dossier to another. Depending on the dossier, MEPs can propose edits during a window of one to six months, after which all the edits related to that dossier are published together. As a result, the actual, detailed chronology of the edits is unfortunately hidden. Furthermore, there is a delay between the time an edit is proposed and the time it is voted: recent edits might be voted *before* older ones. The timestamps associated with each edit are, therefore, noisy.

Example of Amendment. We show in Figure 2 an example of two conflicting amendments in their raw format. Amendment 108 is proposed on Recital 16 of a legislation on tobacco-related products. Its authors are three Polish MEPs: Jolanta Emilia Hibner, Małgorzata Handzlik, and Bogusław Sonik. It consists of two edits: The first one deletes "should" and inserts "shall"; the second one deletes the end of the first sentence. Amendment 5 is authored by a German MEP:

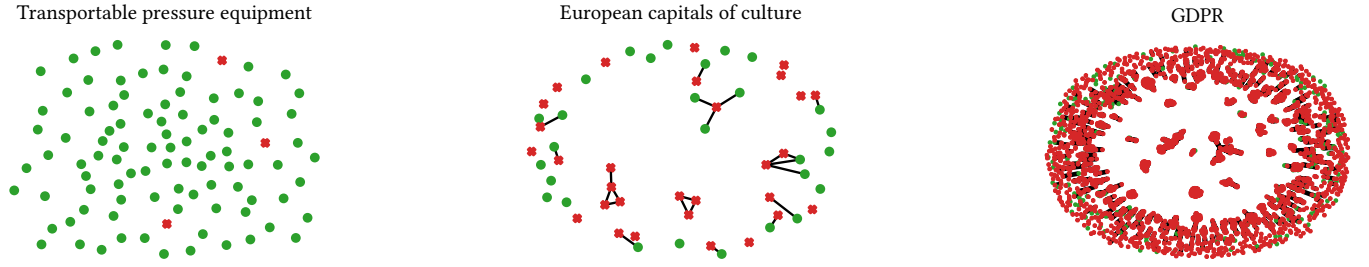


Figure 3: (Left) The "transportable pressure equipment" edit graph contains 96 edits (97% accepted) and no conflicts. (Center) The "European capitals of culture" edit graph contains 58 edits (48% accepted) and 16 conflicts. (Right) The GDPR edit graph contains 3154 edits (9% accepted) and 1298 conflicts.

Klaus-Heiner Lehne. It consists of two edits: The first one replaces "should be" by "is"; the second one is identical to the second edit of Amendment 108. Consequently, the first edits of Amendment 5 and 108 are in conflict, whereas the second edits are identical and are therefore merged, as they are proposed by the four MEPs together. All these edits were rejected in this case. This example illustrates nonetheless the subtlety of legislative texts: The difference between "should", "shall", and "is" is crucial [21].

4 EDIT GRAPH

We describe the dynamics of the legislative process in terms of the conflicts between edits. For each dossier, we construct the edit graph $G = (V_G, E_G)$, such that each node $v \in V_G$ is an edit and such that there is an undirected edge $(u, v) \in E_G$ if edits u and v overlap. A component of size at least 2 in G is therefore a group of overlapping edits. An isolated node corresponds to an edit that does not overlap with any other edit.

In Figure 3, we show the edit graphs of three regulations of EP7. We depict each node with a green dot if the edit is accepted, and with a red cross if the edit is rejected. The "transportable pressure equipment" (left), a very specific legislation, exhibits a graph with 96 nodes, among which 97% are accepted. The graph contains only isolated nodes, meaning that no edits overlap: all its components are size 1. The "European capitals of culture" (center), which can affect some cities of member states, exhibits a graph with 58 nodes, among which 48% are accepted. The graph contains 16 cliques and the average component size is 1.49. The GDPR (right), with high stakes for both businesses and consumers, exhibits a graph with 3154 nodes, among which only 9% are accepted. The graph contains 1298 cliques, meaning that many edits are conflicting, and has an average component size of 3.44.

Conflicts are inherent in the ordinary legislative procedure defined in Section 2, as every proposed edit reflects a disagreement with the initial law proposal. A first class of conflicts occur between the proposal and each edit proposed by MEPs. These conflicts appear as components of any size in G . Hence, every isolated node and every clique in G are such conflicts. We call them "conflicts with the status quo", as they are in disagreement with the proposal. For example, each edit of Amendments 108 and 5 in Figure 2 is such a conflict. In Figure 3 (left), each green node is an edit accepted over the status quo, and each red node is an edit rejected over the

status quo. Similarly, in Figure 3 (center), the cliques with all red nodes are rejected over the status quo.

Another class of conflicts occur between two or more edits proposed by MEPs. If several MEPs propose different edits on the same part of a text, they compete with each other for the acceptance of their suggestions. In this case, the edits conflict with the status quo and with edits proposed by other MEPs. These conflicts appear as a clique of size at least 2 in G , as there is an edge between overlapping edits. For example, in Figure 2, the first edit in Amendment 108 and the first edit in Amendment 5 form such a conflict. It corresponds to a clique of size 2. In Figure 3 (left), there are no such conflicts. As no edge links any two nodes, all conflicts are only with the status quo. In Figure 3 (center), however, the cliques with one green node and one or more red nodes are conflicts between several edits, where one edit is accepted over the others and over the status quo.

In G , two green nodes cannot appear at both ends of the same edge, as only one edit can be accepted among those that are conflicting. Hence, green nodes can only appear as an independent set on the components. Two red nodes, however, can appear at both ends of the same edge, as they can both be rejected: this is the case with the first edit in Amendments 108 and 5.

5 STATISTICAL MODELS

We propose a statistical model of edit outcomes from conflicts. We incorporate assumptions reminiscent of the Bradley-Terry model [2] and of the Rasch model [14], as follows. We model the amending process as a "game" between (a) the MEPs themselves (similar to the Bradley-Terry model) and (b) the MEPs and the status quo (similar to the Rasch model). For simplicity, let us suppose that an edit proposed by MEP u is accepted on dossier i over a conflicting edit proposed by MEP v . As an example, a MEP from one party might propose a modification favoring economic interests, whereas another MEP from another party proposes a modification at the same position in the proposal favoring social interests. We model the probability of the edit proposed by MEP u to be accepted over the edit proposed by MEP v on dossier i , i.e., the probability of MEP u "winning" over MEP v on dossier i as

$$\begin{aligned}
 p(u >_i v) &= \frac{\exp(s_u)}{\exp(s_u) + \exp(s_v) + \exp(d_i + b)} \\
 &= \frac{1}{1 + \exp[-(s_u - s_v)] + \exp[-(s_u - d_i) + b]}, \quad (1)
 \end{aligned}$$

where $s_u, s_v \in \mathbf{R}$ are the *skills* of MEPs u and v , $d_i \in \mathbf{R}$ is the *inertia* of dossier i , and $b \in \mathbf{R}$ is a global bias parameter. The first exponential in the denominator of (1) encodes the MEP-MEP interaction. The second exponential encodes the MEP-dossier interaction. If an edit proposed by MEP u does not conflict with any other edits, the MEP-MEP term vanishes, leaving only the MEP-dossier term.

The parameters in this model enable interpretation. The skill s_u quantifies the ability of MEP u to pass an edit representing their views. We interpret a high skill as a high *influence*. The inertia d_i quantifies the resistance to change of dossier i . This resistance is not due to the dossier resisting *per se* but rather to the effect of other MEPs voting the edits or proposing conflicting edits. In this sense, we interpret a high inertia as a sign of possible high *controversy*. The general bias term b tunes the importance that the model gives to the MEP-MEP term relative to the MEP-dossier term. We conduct an in-depth analysis of the parameters in Section 6.

Multiple Authors and Multiple Conflicts. As explained in Section 3 and Section 4, one or more MEPs can propose an edit, and an edit can be in conflict with one or more other edits. It is easy to generalize (1) to multiple authors and multiple conflicts. To model multiple authors, we simply sum the skills of each author of an edit. To model multiple conflicts, we observe that each conflict generates a new MEP-MEP interaction term. Call $C = \{a, b, \dots\}$ the set of conflicting edits proposed by authors $\mathcal{A}_a, \mathcal{A}_b, \dots$. Note that C forms a clique in the edit graph G of Section 4. The probability of edit a being accepted over edits b, \dots on dossier i is given by

$$p(a >_i C - \{a\}) = \frac{\exp(s_a)}{\sum_{c \in C} \exp(s_c) + \exp(d_i + b)}, \quad (2)$$

where $s_a = \sum_{u \in \mathcal{A}_a} s_u$ is the cumulated skill of all authors of edit a . We refer to this model as the WoW model. The probability that all edits are rejected, i.e., the status quo of dossier i wins, is given by

$$p(i > C) = 1 - \sum_{a \in C} p(a >_i C - \{a\}) = \frac{\exp(d_i + b)}{\sum_{a \in C} \exp(s_a) + \exp(d_i + b)}.$$

Rapporteur Feature. We focus on the role of *rapporteur*, explained in Section 2. A rapporteur is a MEP with a special role in shaping a dossier, which plausibly confers additional influence compared to other MEPs. In order to validate this hypothesis, we add a parameter $r \in \mathbf{R}$ to the skill s_u of a MEP u if they are the rapporteur for the dossier i , i.e., we replace s_a in (2) by

$$s_a = \sum_{u \in \mathcal{A}_a} s_u + r \mathbf{1}_{\{u \text{ is rapporteur for } i\}}.$$

We refer to this model as the WoW(R) model.

Learning the Model. Each observation k is a triplet (C_k, i_k, ℓ_k) of (a) a set of conflicting edits C_k with $|C_k| = c_k > 0$, (b) a dossier i_k on which the edits are proposed, and (c) a label $\ell_k \in C_k \cup \{i_k\}$ indicating which of the c_k edits or the status quo is accepted. Given a dataset of K independent triplets $\mathcal{D} = \{(C_k, i_k, \ell_k) \mid k = 1, \dots, K\}$, we learn the parameters by maximizing their log-likelihood under \mathcal{D} . That is, by collecting all the parameters into a single vector θ ,

Table 2: Average cross entropy of a baseline and our model.

| Legislature | Random | WoW | WoW(R) |
|-------------|--------|-------|--------------|
| EP7 | 0.886 | 0.714 | 0.690 |
| EP8 | 0.942 | 0.748 | 0.726 |

we seek to minimize the negative log-likelihood

$$-\ell(\theta; \mathcal{D}) = \sum_{k=1}^K \sum_{a \in C_k} \left[\mathbf{1}_{\{\ell_k = a\}} \log p(a >_{i_k} C_k - \{a\}) + \mathbf{1}_{\{\ell_k = i_k\}} \log p(i_k > C_k) \right], \quad (3)$$

where $p(a >_{i_k} C_k - \{a\})$ and $p(i_k > C_k)$ depend on θ . In order to avoid overfitting, we add L_2 -regularization to the negative log-likelihood. We pre-process our dataset by keeping only the dossiers for which more than 10 edits were proposed and by keeping only the MEPs who proposed more than 10 edits. Hence, we obtain a dataset of $K = 125733$ data points for EP7 and $K = 140763$ data points for EP8. We split them into 70% for training and validation, and we keep 30% as a test set. The log-likelihood (3) is convex, and we find optimal parameters by using a convex optimizer, such as L-BFGS-B [3].

6 RESULTS

We use the average cross-entropy loss to measure the predictive power of our probabilistic models. Let (C_k, i_k, ℓ_k) be an observation. We compute $-\log p(\ell_k >_{i_k} C_k - \{\ell_k\})$, and we report the average value for all points in our test set. A lower value of the loss means better calibrated probabilities. We compare our models against a random predictor that randomly chooses one of the edits or the status quo as the winner. We show in Table 2 the overall performance of our model over EP7 and EP8.

The WoW model outperforms the random predictor, and including the rapporteur feature r in the WoW(R) model provides a greater decrease in loss. The value of r is positive for both EP7 ($r = 1.18$) and EP8 ($r = 1.31$). This "rapporteur advantage" complements the findings of [4], conducted by interviewing key informants over EP5 (1999–2004) and EP6 (2004–2009). They show that the rapporteur, with their particular role, has some influence on the legislative process, although constrained. The value of r is nonetheless higher in EP8 than in EP7. This suggests that the rapporteur's influence increased in EP8.

Influence and Inertia. Table 3 provides a list of the three dossiers in EP8 with the highest inertia parameter d_i and the three dossiers with the lowest d_i . The values of d_i correlate well with the number of nodes, the number of cliques, the average size of cliques, and the edit acceptance rate. The top-three dossiers include laws with high stakes: The "Screening of foreign direct investments" sets a framework to better equip the EU for investments from non-EU countries. It has crucial implications for companies, workers, governments, and citizens. The infamous "Copyright in the Digital Single Market", considered to be a threat to freedom of expression on the Web by its opponents, sparked public protests in several

Table 3: Top-3 and bottom-3 dossiers in EP8 according to their inertia parameters d_i .

| d_i | Type | Title | # nodes | # cliques | avg. clique size | % accepted |
|--------|---------|---|---------|-----------|------------------|------------|
| 3.304 | report | Screening of foreign direct investments | 1040 | 272 | 3.1 | 2.6 |
| 3.204 | report | Copyright in the Digital Single Market | 2657 | 577 | 4.3 | 2.6 |
| 3.106 | report | Energy efficiency labelling | 1292 | 319 | 3.4 | 6.0 |
| -2.611 | opinion | Financial support for customs control equipment | 60 | 1 | 2.0 | 90.0 |
| -2.644 | opinion | Establishing the supervisory authorities on financial markets | 69 | 0 | 0.0 | 98.6 |
| -2.849 | opinion | Unfair trading practices in the food supply chain | 63 | 6 | 2.0 | 84.1 |

cities. The reporting committee publicized that "MEPs have rarely or never been subject to a similar degree of lobbying before"[13]. Finally, the "Energy efficiency labelling" updated famous labels for electrical appliances, which guide consumers in their purchases. The bottom-three dossiers are all opinions, which are intrinsically less important than reports, as explained in Section 2.

7 RELATED WORK

Amendment analysis was pioneered by [7]. The author compares the influence of the Parliament—as an institution rather than individual MEPs—over the Commission during EP3 (1989–1994) and EP4 (1994–1999). They do so by modeling the acceptance rate of 500 amendments. Similar analyses are developed in [20] and [8] with datasets of, respectively, 1 000 and 5 000 amendments. Our work introduces a large dataset of more than 450 000 amendments spanning EP7 and EP8.

Predicting the success of edits has been widely studied in the context of Wikipedia [1, 5, 22]. Similarly, a whole body of literature covers the conflicts between two Wikipedia edits [17, 23] and the quantification of controversy of Wikipedia articles [15, 16]. The notion of conflict is, however, different in our setting, where multiple edits can be in conflict at the same time. In this case, the task of predicting which edit will be accepted out of all the conflicting edits is more complex, and classic approaches cannot be used.

Our model draws inspiration from the discrete choice models. First, it borrows from the Bradley-Terry model in the pairwise-comparisons literature [2, 18, 24] to model the competitive dynamics between MEPs. These approaches learn a real-valued score for individuals and model the probability that one individual wins over another as a function of the difference of their scores. Second, it borrows from the Rasch model in the item-response theory [14] to model the competitive dynamics between MEPs and the status quo. These approaches learn a real-valued strength for each individual and a real-valued difficulty for each item, and they model the probability that an individual wins over the item as a function of the difference of the strength and the difficulty. Our model unifies both approaches by learning a strength for each MEP and a difficulty for each dossier, considering (i) conflicts between MEPs and (ii) conflicts between MEPs and the status quo.

8 CONCLUSION

In this paper, we have introduced a new dataset of legislative edits and a model of edit outcomes. Our dataset provides rich information on a long-term, dynamical process of interactions between

parliamentarians. Our proposed model learns a skill parameter for MEPs who propose edits and an inertia parameter for the law proposals that resist to change. We have provided an interpretation of the parameters, in terms of the influence of MEPs and of the controversy of the laws. We have also shown that MEPs in the role of rapporteur, hence in charge of a particular dossier, have more influence than other MEPs on the committee.

Future Work. First, a limitation of our approach is that our model is agnostic to the actual text of the edits. A cosmetic edit correcting a typo is obviously not equivalent to a more substantial change of the law. It is however complex to discriminate these two types of edits, as even one word can have critical legal implications (e.g., "shall" versus "should" in the example of Figure 2). We plan to investigate this aspect more deeply. Second, the inclusion of the rapporteur feature, and its subsequent improvement in predictive performance, opens the perspective of including additional features related to the MEPs, the edits, and the dossiers. This would help improve the performance of our model and better understand what contributes to the success of edits. Finally, our model assumes that if MEP u is more influential than MEP v , then $p(u >_i v) > p(v >_i u)$ for all dossiers i . This strong assumption is clearly not always realistic: dossiers span a vast amount of different topics, and MEPs have their own specializations and interests. We plan to improve our model by capturing these dependencies.

A DATA PROCESSING

We extract data from official legislative documents in DOCX format fetched online². We convert the original documents to HTML by using the PyDocX³ library. We scrape the HTML documents in order to extract structured data. We extract edits by using the `difflib`⁴ library. Finally, we map proposed edits to reported edits using the edit distance as defined by `difflib` with a threshold of 0.9.

We identify conflicts between edits by casting this problem as the Max-Cliques problem in the edit graph G defined in Section 4. Although the Max-Clique problem is NP-complete, there exist algorithms for solving it efficiently. We use the Bron-Kerbosch algorithm with pivot [19], which has worst-case time complexity of $O(3^{N/3})$ in the number of nodes of the connected component from which we seek to extract cliques. This is efficient in practice, as the size of the connected components in G is typically small. We successfully identify approximately 100 000 conflicts in less than 20 minutes.

²<http://www.europarl.europa.eu/committees/en/search-in-documents.html>

³<https://pydocx.readthedocs.io/en/latest/>

⁴<https://docs.python.org/3.6/library/difflib.html>

REFERENCES

- [1] B. Thomas Adler and Luca de Alfaro. 2007. A Content-Driven Reputation System for the Wikipedia. In *Proceedings of WWW'07*. Banff, AB, Canada.
- [2] Ralph Allan Bradley and Milton E. Terry. 1952. Rank Analysis of Incomplete Block Designs: I. The Method of Paired Comparisons. *Biometrika* 39, 3/4 (1952), 324–345.
- [3] Richard H Byrd, Peihuang Lu, Jorge Nocedal, and Ciyu Zhu. 1995. A limited memory algorithm for bound constrained optimization. *SIAM Journal on Scientific Computing* 16, 5 (1995), 1190–1208.
- [4] Rory Costello and Robert Thomson. 2010. The policy impact of leadership in committees: Rapporteurs' influence on the European Parliament's opinions. *European Union Politics* 11, 2 (2010), 219–240.
- [5] Gregory Druck, Gerome Miklau, and Andrew McCallum. 2008. Learning to Predict the Quality of Contributions to Wikipedia. In *Proceedings of WikiAI 2008*. Chicago, IL, USA.
- [6] President Barack Obama's White House. 2018. Open Government Initiative. <https://obamawhitehouse.archives.gov/open> Accessed: 2020-02-02.
- [7] Amie Kreppel. 1999. What affects the European Parliament's legislative influence? An analysis of the success of EP amendments. *JCMS: Journal of Common Market Studies* 37, 3 (1999), 521–537.
- [8] Amie Kreppel. 2002. Moving beyond procedure: an empirical analysis of European Parliament legislative influence. *Comparative Political Studies* 35, 7 (2002), 784–813.
- [9] Vaughne Miller. 2010. How much legislation comes from Europe? *Economic Indicators* 7, 10 (2010).
- [10] Conference of the Representatives of the Governments of the Member States. 2007. *Treaty of Lisbon Amending the Treaty on European Union and the Treaty Establishing the European Community*. European Union.
- [11] European Parliament. 2018. The Ordinary Legislative Procedure. <http://www.europarl.europa.eu/ordinary-legislative-procedure/en/ordinary-legislative-procedure.html> Accessed: 2020-02-02.
- [12] European Parliament. 2018. Rules of Procedure of the European Parliament - Rule 174. <http://www.europarl.europa.eu/sides/getDoc.do?pubRef=//EP//TEXT+RULES-EP+20180731+RULE-174+DOC+XML+V0//EN&language=EN&navigationBar=YES> Accessed: 2020-02-02.
- [13] European Parliament. 2019. Questions and Answers on issues about the digital copyright directive. <https://www.europarl.europa.eu/news/en/press-room/20190111IPR23225/> Accessed: 2020-02-02.
- [14] Georg Rasch. 1960. *Probabilistic Models for Some Intelligence and Attainment Tests*. Danmarks Pædagogiske Institut.
- [15] Hoda Sepehri Rad and Denilson Barbosa. 2012. Identifying controversial articles in Wikipedia: A comparative study. In *Proceedings of WikiSym'12*. Linz, Austria.
- [16] Hoda Sepehri Rad, Aibek Makazhanov, Davood Rafiei, and Denilson Barbosa. 2012. Leveraging editor collaboration patterns in Wikipedia. In *Proceedings of the 23rd ACM conference on Hypertext and social media*. 13–22.
- [17] Róbert Sumi, Taha Yasseri, András Rung, András Kornai, and János Kertész. 2011. Edit wars in Wikipedia. In *Proceedings of SocialCom 2011*. Boston, MA, USA.
- [18] Louis L. Thurstone. 1927. The method of paired comparisons for social values. *The Journal of Abnormal and Social Psychology* 21, 4 (1927), 384–400.
- [19] Etsuji Tomita, Akira Tanaka, and Haruhisa Takahashi. 2006. The worst-case time complexity for generating all maximal cliques and computational experiments. *Theoretical Computer Science* 363, 1 (2006), 28–42.
- [20] George Tsebelis, Christian B Jensen, Anastassios Kalandrakis, and Amie Kreppel. 2001. Legislative procedures in the European Union: An empirical analysis. *British Journal of Political Science* 31, 4 (2001), 573–599.
- [21] John Vidal. 2015. How a 'typo' nearly derailed the Paris climate deal. <https://www.theguardian.com/environment/blog/2015/dec/16/how-a-typo-nearly-derailed-the-paris-climate-deal> Accessed: 2020-02-02.
- [22] Ali Batuhan Yardim, Victor Kristof, Lucas Maystre, and Matthias Grossglauser. 2018. Can Who-Edits-What Predict Edit Survival?. In *Proceedings of KDD'18*. London, United Kingdom.
- [23] Taha Yasseri, Robert Sumi, András Rung, András Kornai, and János Kertész. 2012. Dynamics of conflicts in Wikipedia. *PLoS one* 7, 6 (2012).
- [24] Ernst Zermelo. 1928. Die Berechnung der Turnier-Ergebnisse als ein Maximumproblem der Wahrscheinlichkeitsrechnung. *Mathematische Zeitschrift* 29, 1 (1928), 436–460.