Analysing EU Treaty-Making and Litigation With Network Analysis and Natural Language Processing

3 authors, including:

Michal Ovádek
University of Gothenburg
45 PUBLICATIONS **34** CITATIONS

SEE PROFILE

Arthur Dyevre
KU Leuven
67 PUBLICATIONS **202** CITATIONS

SEE PROFILE

Some of the authors of this publication are also working on these related projects:

Project   Trade law View project

Project   EUTHORITY View project

Analysing EU Treaty-Making and Litigation With Network Analysis and Natural Language Processing

Michal Ovádek, Arthur Dyevre* and Kyra Wigard

KU Leuven Centre for Empirical Jurisprudence, Leuven, Belgium

We apply network analysis and topic modeling techniques to explore the evolution of the European Union's treaty making activity and the patterns of litigation they have given rise to. Our analysis reveals that, despite the expansion of the bloc's policy remit, its treaty-making activity retains a strong economic focus. Among the many agreements negotiated by EU institutions, the European Economic Agreement, the Ankara Agreement with Turkey and the World Trade Organization Agreement form the largest clusters of litigated cases. EU international agreements are disproportionately litigated in cases pertaining to residence rights and competition law.

Keywords: network analysis, natural language processing, topic modeling, international agreements, European union, litigation

1. INTRODUCTION

Negotiations with the United Kingdom over the post-Brexit deal have highlighted the role of the European Union (EU) as a global treaty-making powerhouse. Member states have delegated treaty-making powers to EU institutions over an expanding set of policy domains, starting with trade, and later matters of defense and security. While the EU is not a state, it is a major international actor and a full member of the World Trade Organization (WTO).

The Brexit negotiations further highlighted the sensitivity of the issue of judicial review. The Leave campaign, including Boris Johnson himself, explicitly referred to the European Court of Justice (ECJ), as a reason for withdrawing from the bloc and the UK government insisted on keeping the ECJ out of the trade deal [1]. (It was eventually decided that legal disputes would be entrusted to an *ad hoc* arbitration panel.) This reflects the broader phenomenon that litigation, by bringing judges into the picture, can decisively influence the effective operation of international treaties.

While treaty negotiation dynamics and aspects of the ECJ's case law have received attention from political scientists and legal scholars [2–4], it is difficult to get a general sense of the variety of agreements and the extent to which they have given rise to litigation.

The sheer number of agreements and EU court cases rules out manual analysis. So we attempt to provide such an overview using machine learning and network analysis methods. We use probabilistic topic modeling to analyse the contents of EU international agreements and network analysis to identify the main clusters of citations to international agreements.

What our data exploration reveals is that, despite the expansion of the bloc's policy remit, its treaty-making activity retains a strong economic focus. It also shows that, among the many agreements negotiated by EU institutions, the European Economic Agreement, the Ankara Agreement with Turkey and the WTO Agreement form the largest clusters of litigated cases. EU international agreements are disproportionately litigated in cases relating to residence rights and,

more surprisingly, competition law, while the opposite is true in cases relating to internal market themes such as public procurement and VAT.

2. RELATED WORK

Our paper relates to the growing literature applying machine learning and Natural Language Processing (NLP) methods to the study of law and legal documents [5–7]. It also relates to the legal and political science literature applying network analysis methods to the analysis of case citation dynamics [8–11] as well as the evolution and structure of legislation and networks of judges and law professors [12–14].

3. DATA

First, we collected data on EU international agreements[1] (N ≈ 10, 000), including full texts where available in English, from the `EUR-Lex` website, the official EU legal database, using a dedicated data collection R package [15]. `EUR-Lex` is well-curated and the data can be assumed to be close to complete, if not so.

Although international agreements can take on various forms, from formal treaties to agreements made through letters, we distinguish two main categories of legal acts based on the metadata in the database. The first comprises agreements in the form of stand-alone documents, regardless of whether these take the form of a treaty, a formalized exchange of letters or of a new protocol to a preceding agreement. The second category is formed by "joint decisions," which are acts produced by a body which was itself set up under a pre-existing international agreement.

Figure 1 shows that joint decisions account for an increasing number of new international legal texts. Ovádek and Raina [16] explain that this trend is likely grounded in the heightened ambition and scope of EU international agreements, most notably exemplified by the Agreement on the European Economic Area which de facto extends the EU internal market to Norway, Iceland and Liechtenstein. Joint decisions then serve to deal with various technical issues arising from the operation of the legal relationship. This governance model has been applied in many EU international agreements. In subsequent analysis, we focus predominantly on the stand-alone agreements rather than joint decisions, as these constitute international agreements in the stricter sense.

To explore litigation patterns involving international agreements, we gathered the entire universe of rulings rendered by EU courts[2] up to 2020 (N ≈ 26, 000). The collected metadata included information on the legal acts cited, which we used to identify citations to EU international agreements.

---

[1]EU international agreements are those agreements to which the EU is a party in its own right. This excludes at present some well-known international documents such as the European Convention on Human Rights and Fundamental Freedoms.
[2]Historically, there have been three EU courts: the European Court of Justice (since 1953), the General Court (founded in 1988 as the Court of First Instance) and the Civil Service Tribunal (established in 2005 and dissolved in 2016).

Figure 2 shows the proportion of EU court cases that contain a reference to at least one international act. We see a steady rise until about 2010 when the trend reverses. Historically, cases referring to EU international agreements account for around 5% of the case law.

4. METHODOLOGY
4.1. Topic Modeling

Topic modeling is a suite of unsupervised document-clustering techniques developed to generate thematic annotations automatically, whereby topics are modeled as probability over words and documents as probability over topics [17, 18]. We use the structural topic model developed by [19] and implemented in the `stm` package for R to generate topics measuring issue attention in international agreements. The implementation builds on the Correlated Topic Model developed by [20].

This topic modeling approach is preferred over the more conventional Latent Dirichlet Allocation (LDA) for three reasons. First, we want to account for temporal variations in the number and thematic focus of agreements and legal disputes. Whereas, LDA is oblivious to the order in which documents appear in the corpus, we group documents by year, whereby topics in year t are assumed to have evolved from to topics prevalent in year $t - 1$. Dynamic topic models have been shown to better fit temporal dynamics in issue attention than LDA variants [21]. Third, our approach ensures that the resulting topics are not overly skewed toward years with more documents, which can occur when there are significant variations in the number of documents over time—which is the case for both EU international agreements and EU court cases. Third, simultaneously with temporal changes, we want to compare topic prevalence in cases citing international agreements to cases where no such reference is made. Our approach allows to model this difference directly.

To allow an assessment of temporal dynamics, we specify a covariate interacting with topic prevalence:

$$\theta_{1:D}|t_{1:D}\gamma, \Sigma \sim LogisticNormal\mu = t_{1:D}\gamma, \Sigma). \qquad (1)$$

where t_d is the year in which document d was issued; γ is a $p \times (K - 1)$ matrix of coefficients for topic proportion and Σ is a $(K - 1) \times (K - 1)$ covariance matrix.

To investigate the variance in topic proportion between cases citing EU international agreements and cases containing no such reference, we estimate a dynamic topic model of EU court cases in which also specify a dummy variable capturing reference to international agreements.

To calculate topic proportion conditional on covariates, the method originally implemented in the `stm` package relied on OLS regression. To constrain topic proportion within the $(0, 1)$ interval, we model topic proportion conditional on covariates using quasi-binomial regression.

Our text-mining approach is based on the bag-of-words paradigm. Accordingly, punctuation, numbers, html tags, rare words, and words common to many documents (including stop-words) were removed from the raw texts—all these are standard pre-processing steps in bag-of-words studies.

FIGURE 1 | Categories of international acts in the dataset (1953–2020).
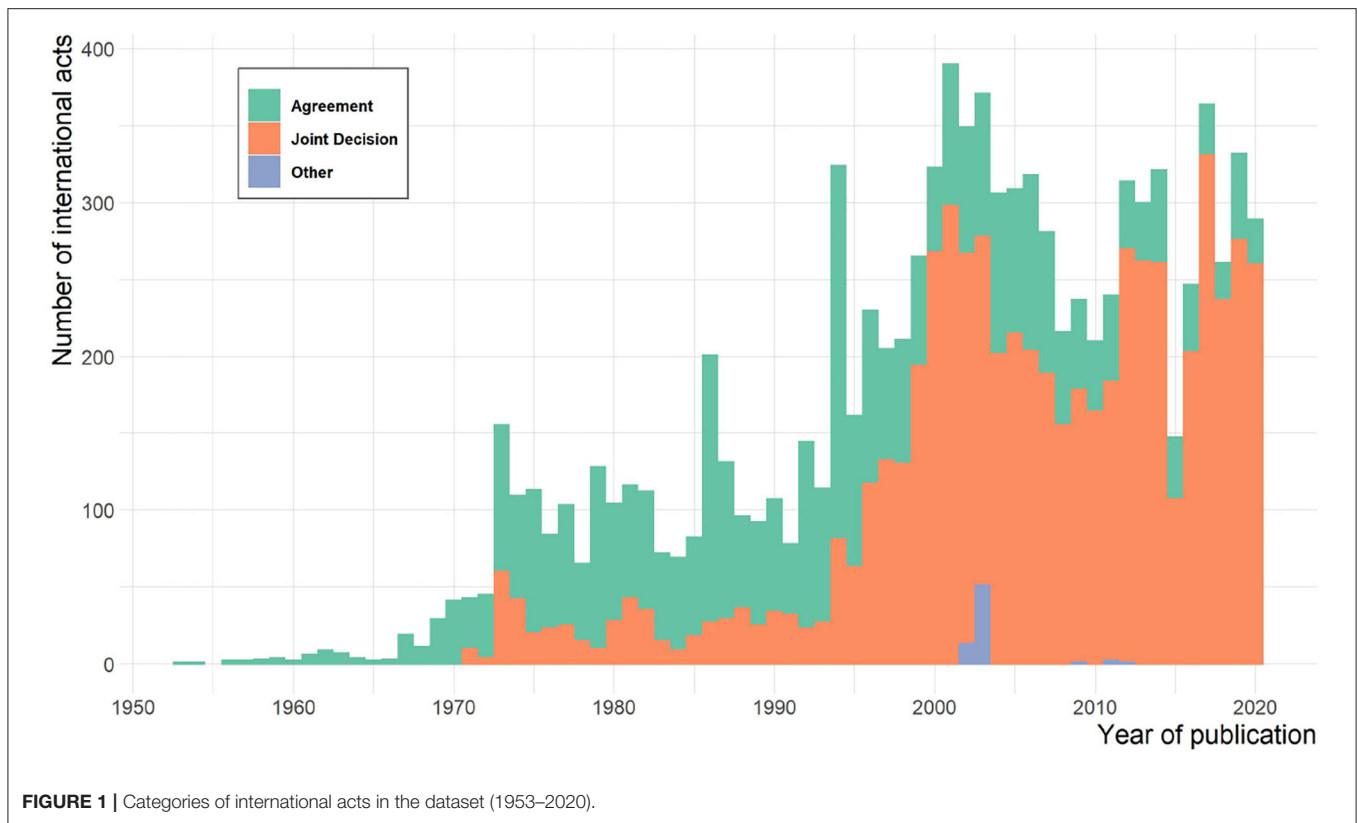
4.2. Network Analysis

We use network analysis to model patterns of citations in EU court cases. Recent years have seen network analysis proliferate in social sciences and legal studies [9, 10, 13, 22]. In network analysis, a network consists of nodes (also known as "vertices") and edges (or links). In our analysis, a node is either an EU court case or an EU international agreement while edges represent citations either to an agreement or to another case. As with applications of network analysis to citation patterns in judicial opinions, we model cases, agreements and citations as directed networks. The directed nature of our legal citation networks results from the fact that agreements do not cite cases while a new case can only cite an older case[3].

To generate our citation networks, we construct an adjacency matrix of court rulings and EU international agreements. A node is adjacent to another if an edge connects them. Formally, if U is the set of all nodes $u_1, ..., u_n$ in a network, the adjacency matrix A_{ij} is a square $n \times n$ matrix connecting nodes u_i and u_j. The elements of the matrix take on value one if two nodes are adjacent and zero otherwise.

Our analysis is primarily concerned with the network centrality of agreements and cases citing agreements. A basic measure of the importance of a treaty or precedent in judicial opinions is in-degree centrality [8, 9, 23]. In-degree centrality simply measures the number of inward citations. In-degree
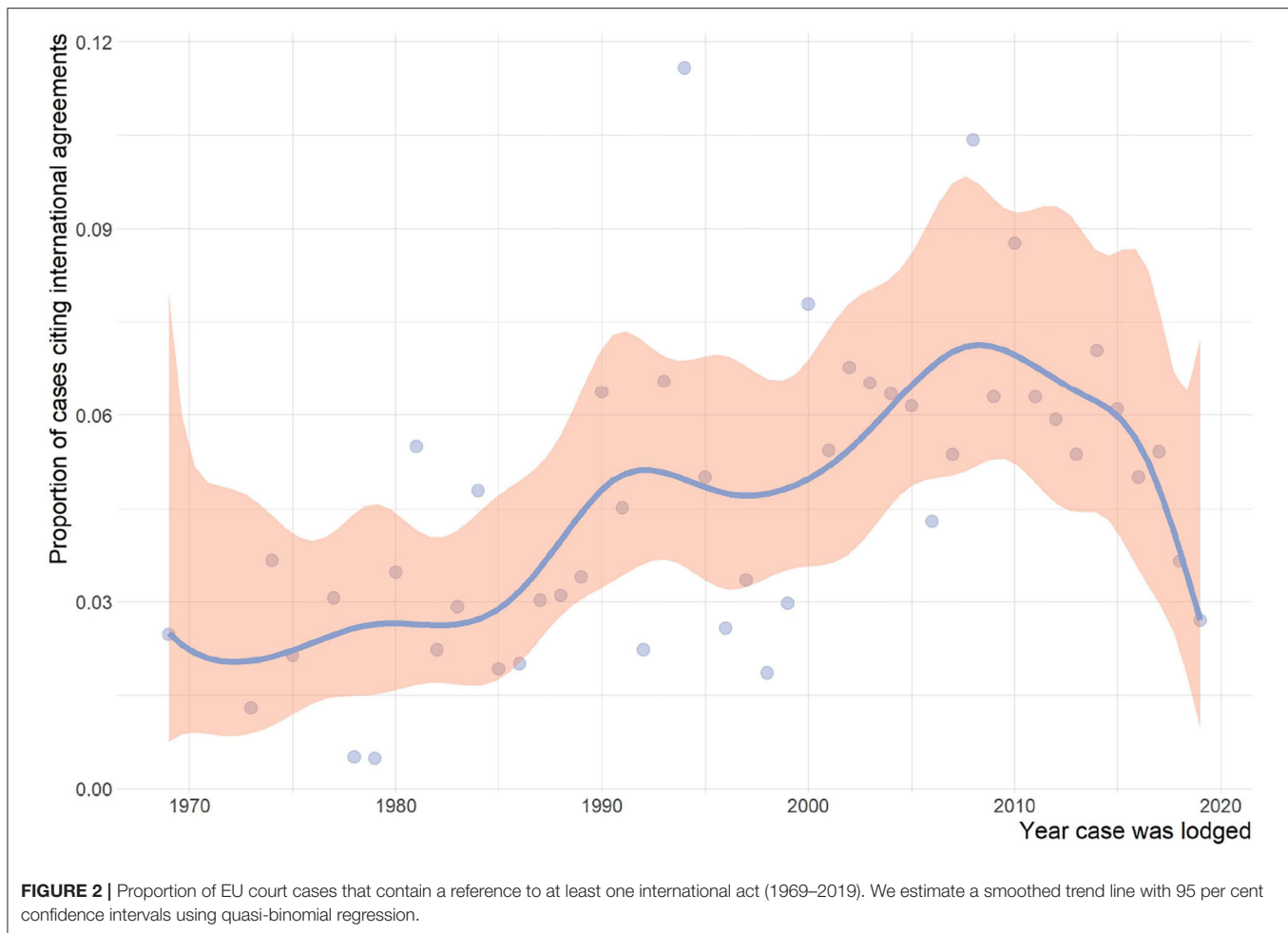
centrality gives equal weight to all inward citations, regardless of the position of the citing node in the network. An alternative measure of importance is eigenvector centrality. Unlike degree centrality, eigenvector centrality takes into account the position occupied by citing cases in the network. The metric assigns greater weight to inward citations cases that are themselves cited more frequently [8].

Finally, we use the fast-greedy community detection algorithm to identify clusters of densely connected cases and agreements [24]. The underlying intuition behind this community-detection algorithm is that cases form a community if they refer more to cases (and agreements) inside the community than to cases (and agreements) outside the community [9]. In mathematical terms, Newman [24] defines the problem of community detection in networks as one of optimizing the value Q in the following function:

$$Q = \sum_i (e_{ii} - \sum_j e_{ij}^2) \qquad (2)$$

where e_{ij} is the proportion of edges between nodes in community i and community j. Values of $Q \neq 0$ have the interpretation of indicating a network division (into communities) where some degree of community structure is present. Calculating Q for all possible network divisions is computationally expensive, however, even with just a few dozen nodes. The fast-greedy algorithm uses hierarchical clustering to solve the problem approximately. Using this technique we obtain a classification of

---

[3]Our analysis ignores references to other documents and agreements which occasionally occur in EU international agreements.

FIGURE 2 | Proportion of EU court cases that contain a reference to at least one international act (1969–2019). We estimate a smoothed trend line with 95 per cent confidence intervals using quasi-binomial regression.

all nodes in our network as belonging to one of S communities where S is determined computationally by finding the maximal value of Q.

5. RESULTS

5.1. Topics in EU International Agreements

To set k, which determines the number of topics, we relied primarily on interpretability and our domain knowledge. We found that $K = 9$ resulted in the most interpretable model. Because agreements sometimes contain technical nomenclatures with numerous acronyms resulting in less interpretable models, we excluded terms with fewer than three characters.
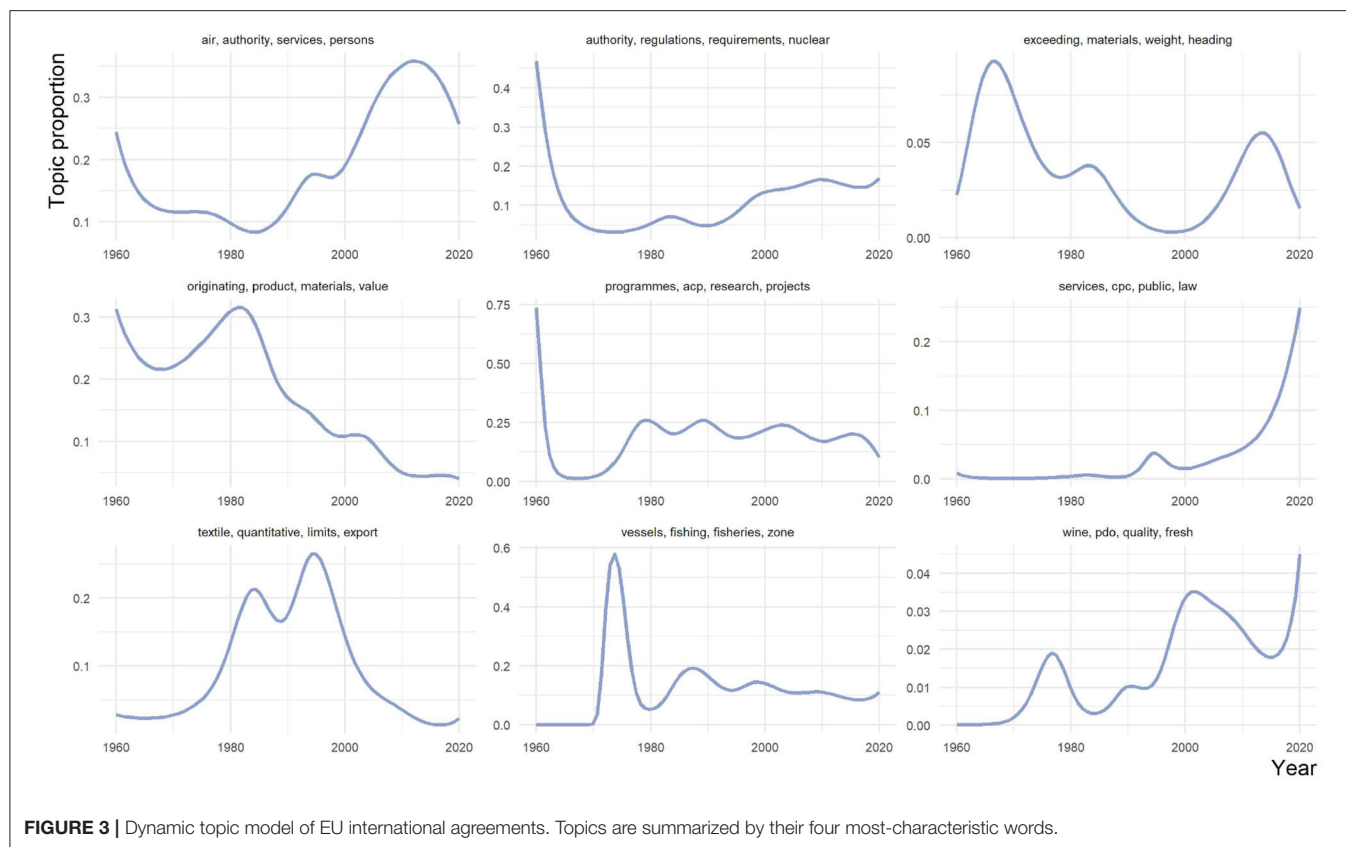
Plotted in **Figure 3** are the nine topics (summarized by their most characteristic words) and their proportion in the corpus over time. EU international agreements largely pertain to trade in goods ("products," "materials," "textile," "wine") and arrangements relating to their shipment and labeling ("originating," "weight," "pdo"). Three product categories can be discerned from the topmost characteristic words: textiles, fish, and other agricultural products (such as wine), including the protection of geographical

indications ("pdo" stands for "Protected designation of origin").

Except for wine and protection of geographical indications, topics relating to products and product shipment (topic "originating, product, materials, value" and topic "exceeding, materials, weight, heading") have seen their relative importance decline over time. Textile ("textile, quantitative, limits, export") experienced a surge in attention in the 1980s and 1990s but later reverted to relative obscurity. Topics relating to services and air transport ("services, cpc, public, law" and "air, authority, services, persons") have steadily grown in importance.

Fishing—an issue that featured prominently in Brexit negotiations—saw a blip in the 1970s. After that, topic proportion remained more or less constant at around ten per cent.

Of great historical importance is the EU's relationship with its member states' former African, Caribbean and Pacific (referred to as "ACP") colonies (topic "programmes, acp, research, projects"). Development cooperation between the EU and ACP countries has given rise to a succession of agreements, starting with the Yaoundé Agreements (1969), followed by the Lomé Conventions (1974) and the Cotonou Agreement (2000).

FIGURE 3 | Dynamic topic model of EU international agreements. Topics are summarized by their four most-characteristic words.

The temporal shifts in topic proportion visible in **Figure 3** reflect, for a part, the evolution of market integration and the growing emphasis on services in later stages of the construction of the internal market. But, while new treaties from the Single European Act (1987) to Lisbon (2009) have granted the EU competences in new policy areas, such environmental protection, immigration and security, our topic model indicate that EU treaty-making continues to concentrate on trade.

5.2. Citation Patterns in Litigation

After examining issue attention in the agreements, we now consider references to these agreements in EU court cases.

References to EU international agreements in EU court cases exhibit marked disparities. Just seven agreements—the Agreement on the European Economic Area (EEA), the WTO Agreements, the Ankara Agreement, the 1970 Additional Protocol to the Ankara Agreement, the International Convention on the Harmonized Commodity Description and Coding System (HS Convention), the Aarhus Convention and the EU-Switzerland Agreement on free movement of persons—are cited in more than 20 rulings (see the **Supplementary Material**). Most cited is the EEA Agreement (mentioned in 288 rulings); followed by WTO Agreements (mentioned in 139 rulings), the Ankara Agreement (mentioned in 66 rulings) and its 1970 Additional Protocol (41 rulings).
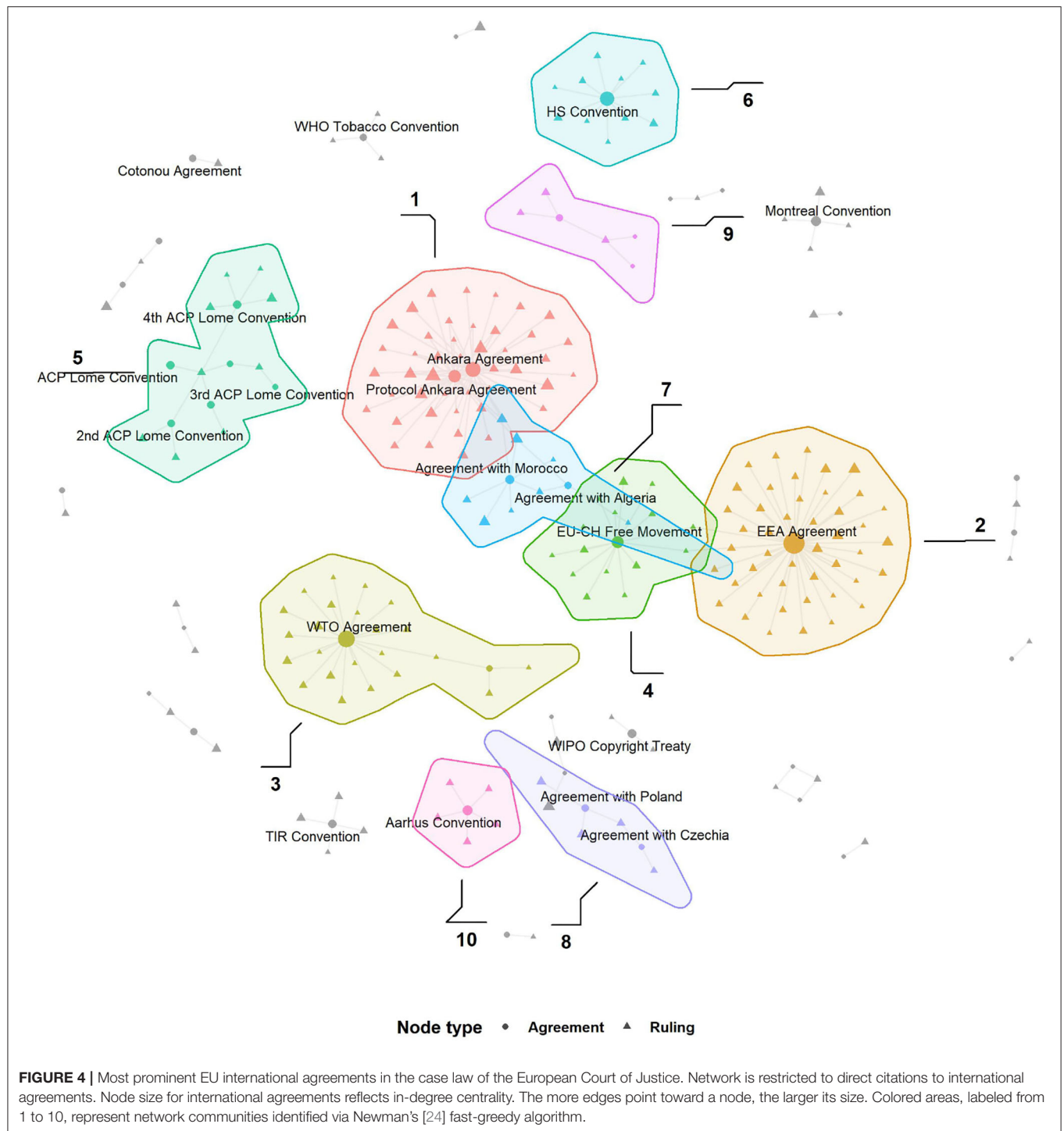
Figure 4 conveys the same point more systematically. The network is restricted to edges representing direct citations to

the agreement. The communities, as identified by the fast-greedy algorithm, are largely isolated from each other, which signifies that two distinct treaties rarely have direct legal bearing on the same case. We employ the in-degree node centrality metric to show the importance of a node in the citation network.

Interestingly, even thematically proximate agreements, such as the WTO agreements and the HS Convention, both of which address international trade in goods, form case clusters that are almost completely separate. The main exception in the network is the EEC-Turkey Ankara Agreement, which is accompanied by the 1970 Additional Protocol. However, given the explicit legal connection between these two international agreements, we should expect that both will often be relevant to the same legal dispute. Similarly, a sparser chain of rulings connects the successive Lomé Conventions between the EEC and the ACP countries, though curiously these are not connected to the Cotonou Agreement which succeeded the Lomé framework in 2000.

Our second network goes beyond direct citations and considers the centrality of both citing and cited rulings. Properties of these networks—including number of nodes, edges, diameter, average degree, modularity, connectance, and transitivity—are reported in **Table 1**. **Figure 5** displays the most prominent agreements and rulings along with the main clusters identified via our Newman's [24] fast-greedy algorithm.

To facilitate data visualization, we use eigen centrality to reduce the size of the network. As explained section 4, eigen

FIGURE 4 | Most prominent EU international agreements in the case law of the European Court of Justice. Network is restricted to direct citations to international agreements. Node size for international agreements reflects in-degree centrality. The more edges point toward a node, the larger its size. Colored areas, labeled from 1 to 10, represent network communities identified via Newman's [24] fast-greedy algorithm.

centrality captures the importance of cases from which citations originate[4]. The network plotted in **Figure 5** reflect this definition of case importance.

---

[4]Let x_i be the eigen centrality of node i in network Q. Then $x_i = \frac{1}{\lambda} \sum_{j \in M(i)} x_j$ where $M(i)$ is a set of all neighbors of i and λ is a constant.

Compared to **Figure 4**, there is overall greater overlap between communities, although communities two (EEA Agreement), three (Ankara Agreement and Protocol to Ankara Agreement), five, seven, and eight clearly stand somewhat apart from the overlapping core formed by, in particular communities one (WTO Agreement) and four. The cases in the community clustered around the Ankara Agreement (community 3), such as

TABLE 1 | Summary metrics of citation network.

Property	Value
Number of nodes	3,863
Number of edges	8,198
Average node degree	4.24
Average node in-degree	2.12
Average node out-degree	2.12
Network diameter	11
Connectance	0.0005495041
Modularity coefficient	0.725
Transitivity	0.071

Metrics correspond to the network displayed in Figure 5. Diameter measures the shortest path between the two most distant nodes. Connectance measures the ratio of realized to possible links (computed as $\frac{L}{N*(N-1)}$). The modulatority coefficient is calculated from the vector of community membership generated via Newmans's [24] fast-greedy algorithm for community detection. Transivity measures the probability that adjacent nodes are connected.

C-561/14 relate predominantly to the rights of Turkish citizens to live and work in the EU. The WTO community (community 1) encompass some landmark cases concerning the interpretation and application of international law in the EU legal order, such as *ATAA*, *Rosneft*, and *Krizan*, the latter dealing specifically with access to information and justice in environmental matters as regulated by the Aarhus Convention. For the first two of these cases, the network seems to reflect well the general nature of the international law questions they raised, as they stand in the center of the entire network with connections spanning across communities.

5.3. Litigation Topic and Incidence of References to EU International Agreements

To explore how cases citing international agreements may differ from cases that do not, we estimated a topic model of EU court cases with covariates for time and reference to international acts as explained in section 4. It is important to bear in mind that EU cases far outnumber EU international agreements. So, with a substantially larger set of documents, we found that, for this task, $K = 30$ produced the most interpretable model.

Figure 6 illustrates the comparative evolution in topic proportion for the two case categories. Some topics indicate a clear divergence between cases with and cases without reference to international acts. For example, cases concerning residence and family reunion rights (topic "residence, country, family, nationals") have an obvious international dimension, which has emerged early in litigation. The number of cases not referencing international agreements in this area has been catching up, however, possibly spurred by the creation of important EU rules such as Directive 2004/38/EC on the right of citizens of the Union and their family members to move and reside freely within the territory of the Member States and the ECJ's interpretation of the notion of EU citizenship [25].

Surprisingly, international agreements seem to be disproportionately invoked in competition law cases ("fine, undertakings, cartel, fines"). This may reflect the globalization of antitrust regulation promoted by the European Commission, which has resulted in the insertion of competition provisions in several agreements [26].

International agreements seem to have become increasingly less relevant in cases pertaining to public procurement ("contracts, tender, award, consumer") and indirect taxation ("vat, tax, sixth, taxable"). These legal areas, along with trademarks ("mark, trademark, euipo, board") and road safety ("insurance, vehicles, vehicle, freedom") have seen increasing regulatory harmonization at EU level; a development that seems to been accompanied by intensifying litigation [7]. To the extent that these topics are highly prevalent in recent years, they may explain the pattern seen in **Figure 2**, which shows a declining proportion of EU court citing EU international agreements.

That staff cases ("staff, officials, competition, post"), most of which employment disputes between EU civil servants and EU institutions, almost never cite international law appears banal, although it provides face validity for our methodological approach.
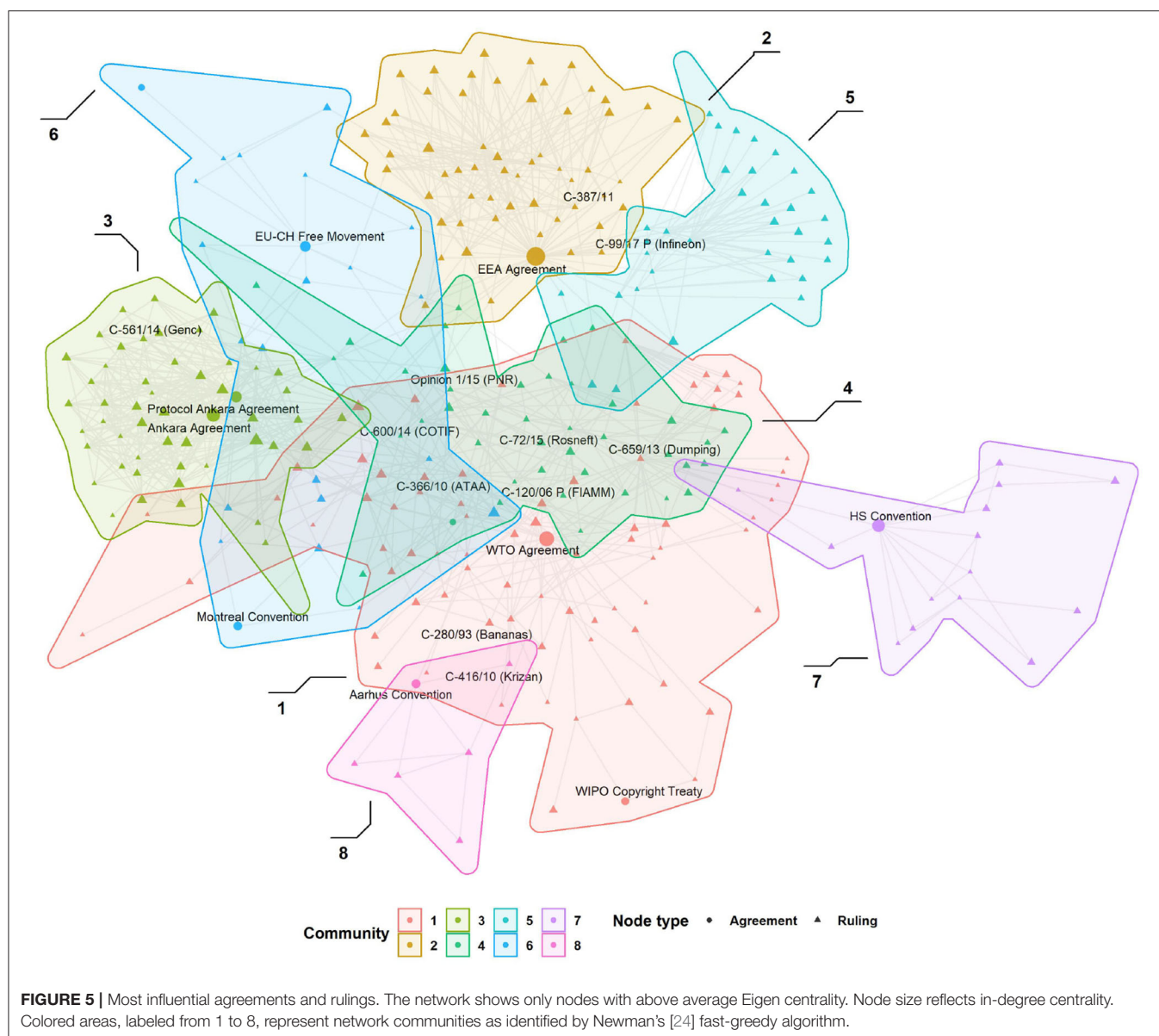
5.4. Case Clusters and Topic Proportion

Finally, we combine network analysis and topic modeling to assess variations in thematic focus in cases belonging to distinct communities. For this purpose, we averaged topic proportion across cases belonging to the same community. Depicted in **Figure 7** is a radial plot comparing topical distribution in the EEA and Ankara Agreement clusters, corresponding to, respectively, community two and three in the network illustrated in **Figure 5**. Average topic proportion for other communities is reported in **Supplementary Material**.

Whereas cases in community two, which are centered around the EEA treaty, concern primarily the free circulation of products and capital, cases in community three, which are clustered around agreements with Turkey, deal mostly with rights of Turkish citizens to live and work in the EU. Combining the two results provides further validation of both approaches. Given how relatively little overlap there is between the two communities, we would expect the topical content of the cases to be quite different, which is precisely what we observe in the topic model.

6. DISCUSSION

What the law is in a given domain or on a given question is typically the expression of information scattered across a large web of texts connected in complex ways [13, 27]. The larger the web, the more difficult it becomes to comprehend its general structure and dynamics using the tools lawyers and legal academics have traditionally applied to study and research the law—manual parsing of documents. While they still require domain knowledge, network analysis and NLP methods provide a scalable alternative to explore the complexity of law.

We applied these two techniques to examine three aspects of the large body of international agreements concluded by EU institutions: (1) their dominant theme, (2) their comparative

FIGURE 5 | Most influential agreements and rulings. The network shows only nodes with above average Eigen centrality. Node size reflects in-degree centrality. Colored areas, labeled from 1 to 8, represent network communities as identified by Newman's [24] fast-greedy algorithm.

centrality in EU court cases, and (3) in the area of litigation in which they are more likely to be involved. We found that economic issues continue to dominate the EU's treaty-making activity; that the EEA, Ankara Agreement, and the WTO Agreement form the largest litigation clusters; and that references to international agreements is proportionally higher in disputes pertaining to antitrust and residence and family reunion rights.

The particular salience of international agreements in residence and family reunion rights speaks directly to the British government's insistence on excluding both mobility rights and the ECJ's jurisdiction from a post-Brexit free trade agreement [1]. Private litigant's standing combined with justiciable mobility rights seem to operate as a powerful litigation catalyst, inviting judges to step in.

Our analysis is primarily conceived as exploratory, but we are confident that it achieves its goal of providing an overview of the EU's treaty-making activity and litigation. Still, we point out two limitations which similar studies may seek to address in the future. The first is that our analysis of litigation is restricted to treaties and agreements to which the EU is formally party. However, international agreements to which the EU is not party—such as the Vienna Convention on the Law of Treaties and the United Nations Charter—have been invoked in ECJ decisions. Future work may seek to map these dynamics. Second, our text-mining procedure follows a bag-of-word approach, which disregards synonymy as well as polysemy and co-reference resolution. Future research may seek to apply distributed semantic and transformer models, which implement

FIGURE 6 | Topic proportion in EU court cases (1980–2019) with and without reference to EU international agreements. Topics are summarized by their four most-characteristic words.
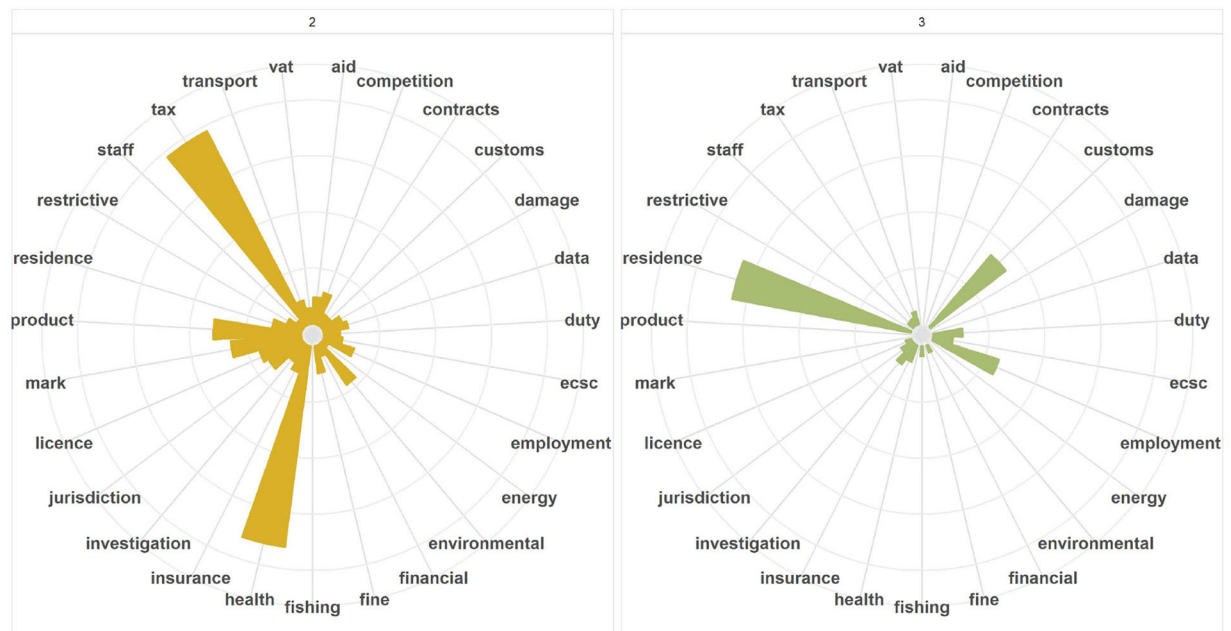

FIGURE 7 | Average topic proportion in network community two (EEA Agreement) and three (Ankara Agreement) identified via fast-greedy algorithm [24]. Values closer to the origin of the circle indicate lower topic proportion. Topics are labeled using the topmost characteristic word.

word embeddings capturing more of the context in which words and even sentences occur [28, 29].

DATA AVAILABILITY STATEMENT

Publicly available datasets were analyzed in this study. This data can be found at: www.eur-lex.eu.

AUTHOR CONTRIBUTIONS

MO conducted the data analysis with input and supervision from AD. AD and MO wrote most of the article. KW contributed suggestions and syntactical improvements on the last draft. All authors contributed to the article and approved the submitted version.

FUNDING

ACKNOWLEDGMENTS

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fphy.2021.657607/full#supplementary-material

REFERENCES

1. Leinarte E, Barnard C. From constitutional adjudication to trade arbitration enforcing mobility rights post-brexit. *Eur Foreign Affairs Rev.* (2020) 25: 589–612. doi: 10.17863/CAM.58501

2. Young AR. European trade policy in interesting times. *J Eur Integr.* (2017) 39:909–23. doi: 10.1080/07036337.2017.1371705

3. Lenaerts K, De Smijter E. The European community's treaty-making competence. *Yearb Eur Law.* (1996) 16:1. doi: 10.1093/yel/16.1.1

4. Teodosiu MJ. The treaty-making powers of the european economic community. *J Int'l L Prac.* (1993) 2:501.

5. Frankenreiter J, Livermore MA. Computational methods in legal analysis. *Annu Rev Law Soc Sci.* (2020) 16:39–57. doi: 10.1146/annurev-lawsocsci-052720-121843

6. Ash E, Chen DL. Case vectors: spatial representations of the law using document embeddings. In: *Law as Data*. Livermore M, Rockmore D, editors. Santa Fe: Santa Fe Institute Press (2019). p. 313–37. doi: 10.2139/ssrn.204926

7. Dyevre A, Lampach N. Issue attention on international courts: evidence from the European court of justice. *Rev Int Organ.* (2020) 15:1–23. doi: 10.1007/s11558-020-09391-0

8. Fowler JH, Johnson TR, Spriggs JF, Jeon S, Wahlbeck PJ. Network analysis and the law: measuring the legal importance of precedents at the US Supreme Court. *Polit Anal.* (2007) 15:324–46. doi: 10.1093/pan/mpm011

9. Lupu Y, Voeten E. Precedent in international courts: a network analysis of case citations by the european court of human rights. *Br J Polit Sci.* (2012) 42:413–39. doi: 10.1017/S0007123411000433

10. Larsson O, Naurin D, Derlén M, Lindholm J. Speaking law to power: the strategic use of precedent of the court of justice of the European Union. *Compar Polit Stud.* (2017) 50:879–907. doi: 10.1177/0010414016639709

11. Sadl U, Olsen HP. Can quantitative methods complement doctrinal legal studies: using citation network and corpus linguistic analysis to understand international courts. *LJIL.* (2017) 30:327. doi: 10.1017/S0922156517000085

12. Ruhl J, Katz DM. Measuring, monitoring, and managing legal complexity. *Iowa L Rev.* (2015) 101:191.

13. Katz DM, Coupette C, Beckedorf J, Hartung D. Complex societies and the growth of the law. *Sci Rep.* (2020) 10:1–14. doi: 10.1038/s41598-020-73623-x

14. Katz DM, Stafford DK. Hustle and flow: a social network analysis of the American federal judiciary. *Ohio St LJ.* (2010) 71:457.

15. Ovádek M. Facilitating access to data on European Union laws. *Polit Res. Exchange.* (2021) 4:e1870150. doi: 10.1080/2474736X.2020.1870150

16. Ovádek M, Raina A. The evolution of EU trade law through the prism of competence: a quantitative, longitudinal perspective. *J World Trade.* (2019) 53:489–508.

17. Blei DM, Ng AY, Jordan MI. Latent dirichlet allocation. *J Mach Learn Res.* (2003) 3:993–1022. doi: 10.1162/jmlr.2003.3.4-5.993

18. Blei DM. Probabilistic topic models. *Commun ACM.* (2012) 55:77–84. doi: 10.1145/2133806.2133826

19. Roberts ME, Stewart BM, Tingley D, Airoldi EM. The structural topic model and applied social science. In: *Advances in Neural Information Processing Systems Workshop on Topic Models: Computation, Application, and Evaluation.* Vol. 4. Lake Tahoe: Harrahs and Harveys (2013).

20. Blei DM, Lafferty JD. A correlated topic model of science. *Ann Appl Stat.* (2007) 1:17–35. doi: 10.1214/07-AOAS136

21. Blei DM, Lafferty JD. Dynamic topic models. In: *Proceedings of the 23rd International Conference on Machine Learning.* Pittsburgh, PA (2006). p. 113–20. doi: 10.1145/1143844.1143859

22. Borgatti SP, Mehra A, Brass DJ, Labianca G. Network analysis in the social sciences. *Science.* (2009) 323:892–5. doi: 10.1126/science.1165821

23. Freeman LC, Roeder D, Mulholland RR. Centrality in social networks: II. Experimental results. *Soc Netw.* (1979) 2:119–41. doi: 10.1016/0378-8733(79)90002-9

24. Newman ME. Fast algorithm for detecting community structure in networks. *Phys Rev E.* (2004) 69:066133. doi: 10.1103/PhysRevE.69.066133

25. Šadl U, Madsen MR. Did the financial crisis change European citizenship law? An analysis of citizenship rights adjudication before and after the financial crisis. *Eur Law J.* (2016) 22:40–60. doi: 10.1111/eulj.12167

26. Aydin U. Promoting competition: European union and the global competition order. *J Eur Integr.* (2012) 34:663–81. doi: 10.1080/07036337.2012.707366

27. Ruhl J, Katz DM, Bommarito MJ. Harnessing legal complexity. *Science.* (2017) 355:1377–8. doi: 10.1126/science.aag3013

28. Mikolov T, Chen K, Corrado G, Dean J. Efficient estimation of word representations in vector space. *arXiv [Preprint]. arXiv:13013781.* (2013).

29. Devlin J, Chang MW, Lee K, Toutanova K. Bert: pre-training of deep bidirectional transformers for language understanding. *arXiv [Preprint]. arXiv:181004805.* (2018).