Artificial Intelligence and Law
https://doi.org/10.1007/s10506-026-09523-8

ORIGINAL RESEARCH

Tracking the evolution of case law with main path analysis

Xiaorui Jiang[1] · **Maria Breskaya**[2] · **Gijs van Dijck**[2]

Received: 5 March 2025 / Accepted: 6 May 2026
© The Author(s) 2026

Abstract

Understanding case law evolution contributes to the comprehension of legal principles, the identification of trends in judicial reasoning, and testing consistency between judgments. Automation of the analysis has the potential to make it easier to keep up with relevant case law. While legal network analysis typically focuses on identifying precedents, it has largely overlooked tracking historical developments in judicial reasoning. This study applies Main Path Analysis (MPA) to legal citation networks to uncover how key decisions have shaped the law over time. This study illustrates MPA with consumer protection law cases decided by the Court of Justice of the European Union (CJEU). MPA was applied to a citation network of 446 consumer rights cases from 1980 to Oct 2023. The method involves two steps: (1) weighting citations to measure knowledge flow and (2) extracting significant paths that trace the evolution of judicial reasoning. The Top-3 key-route main paths from five time slices were merged into an overlay network and analysed for alignment with judicial interpretations and evolving consumer protection themes. The topics on the paths are compared to those prevalent in the literature on consumer protection. The analysis uncovered key paths in consumer protection jurisprudence shaped by pivotal cases. Organised around five themes—unfair terms, equivalent effect, comparative advertising, free movement of goods, and product liability—paths align closely with the CJEU's interpretative practices. While MPA highlights significant pathways, it has caveats and challenges, particularly in achieving comprehensive and representative results and in large networks. Future improvements could include integrating semantic adjustments and key case identification to enhance both aspects, as well as developing complementary methods to identify derivative or branching paths in domain development.

Keywords Main path analysis · Network analysis · Case law · Evolution · Consumer protection

Extended author information available on the last page of the article

Published online: 01 July 2026

1 Introduction

In law, network analysis has been applied to identify landmark decisions or precedents, which is the practice of using a rule, set in a prior legal case, to decide subsequent cases (Derlén and Lindholm 2015, pp. 1075–1076; Lupu and Voeten 2012, pp. 416–417). A famous precedent in the EU context is the Bosman case, where invoking the freedom of movement for workers resulted in the ban of restrictions on foreign EU football players within national leagues and in allowing players in the EU to transfer to another club at the end of their contract without clubs having to pay a transfer fee.

Network analysis relies on two key measures: nodes (also called vertices) and links (also called arcs or edges). The nodes can be individuals, groups, or information, whereas the links connect the nodes. Treating court decisions as nodes and court citations as links allows for creating a precedent network and for exploring how central precedents are in a given network (Fowler et al. 2007, p. 325). Analysts subsequently assume that, for instance, decisions cited more frequently are more likely to be more important than decisions cited less frequently.

Studies that applied network analysis to the legal field commonly focus on identifying precedents (Whalen 2016, p. 547ff). For example, in his seminal work, Fowler et al. conducted a network analysis on 26,681 majority opinions written by the US Supreme Court and the cases that cite them from 1791 to 2005 (Fowler et al. 2007). The authors described a method for identifying the most important precedents, with their rankings outperforming several commonly used case importance measures. A subsequent publication found that (1) reversed cases tend to be more important than other decisions, (2) cases that overrule the reversed cases 'quickly become and remain even more important', and (3) the Supreme Court carefully grounds overruling decisions in past precedent (Fowler and Jeon 2008).

Other courts that have been analysed by means of network analysis include the European Court of Human Rights (ECtHR) (Christensen et al. 2016; Lupu and Voeten 2012; Olsen and Küçüksu 2017), the Court of Justice of the European Union (CJEU) (Derlén and Lindholm 2014, 2015; Knops and Schaper 2014; Schaper 2014), the International Criminal Court (Tarissan and Nollez-Goldbach 2015, 2016), and courts in European member states (Agnoloni and Pagallo 2015; van Dijck and van Kuppevelt 2019; Winkels et al. 2011) and Canada (Neale 2013). Legal network analysis studies have also focused on other areas than case law, such as legal services (Beaverstock 2004; Heinz and Laumann 1982; Lazega 2001; Lazega and Krackhardt 2000; Lazega and van Duijn 1997), statutes and regulatory codes (Bommarito and Katz 2010; Boulet et al. 2011; Katz and Bommarito 2014; Koniaris et al. 2015), patent citations (Whalen 2016, pp. 550–551 (providing an overview)), criminal behaviour (Whalen 2016, pp. 553–554 (providing an overview)), and terrorists (Qin et al. 2005). Methodological studies have focused on improving the use of legal network analysis by focusing on, among other things, community detection methods and centrality measures that capture the relevance of precedents (Lettieri et al. 2016; Page et al. 1999; Smith 2007; Tarissan et al. 2016; van Kuppevelt et al. 2020; van Opijnen 2012; Zhang and Koppaka 2007).

A topic that has remained underexplored is the historical development of case law. Tracking the evolution of case law helps understand legal principles, identify trends in judicial reasoning, and assess consistency across rulings. Automating this analysis can simplify staying up to date on relevant case law, yet automation methods have not been explored. A strategy explored in this paper, main path analysis (MPA), is to determine the paths through which information in case law flows and to identify the paths that most likely contain decisions that have played an important role in the evolution of a network while excluding decisions that have played an insignificant role.

MPA was introduced in Hummon and Doreian (1989). Batagelj (2003) developed efficient algorithms for it. MPA uses traversal paths and traversal weights to assess which paths information flows through in a citation network. MPA serves a distinct purpose from conventional network measures such as in-degree or centrality. Centrality metrics identify nodes that receive many citations or occupy influential positions within the network, whereas MPA traces citation sequences that connect earlier and later cases. Consequently, MPA reconstructs the flow of doctrinal development rather than measuring individual influence. In other words, a case may have high centrality yet not appear on the main path if it does not function as a key link in the transmission of legal reasoning over time, while a moderately cited case may prove central in connecting major doctrinal shifts. Thus, MPA complements centrality analysis by highlighting the pathways through which doctrinal influence flows.

MPA has since been used to map technological trajectories (Fontana et al. 2009; Verspagen 2007; Liu et al. 2019), detect technological changes (Lucio-Arias and Leydesdorff 2008; Martinelli 2012), and conduct literature reviews (Calero-Medina and Noyons 2008; Harris et al. 2011; Bhupatiraju et al. 2012; Colicchia and Strozzi 2012; Lu et al. 2012; Liu et al. 2013a, 2013b). As far as the authors are aware of, the only application of MPA in the legal domain used this method to compare its results to cases mentioned in an authoritative trademark book (Liu et al. 2014). The current paper presents a case study on the network of cases about consumer right protection in the European Court of Justice, to uncover how key decisions or laws about consumer protection have evolved over time and reveal the structure of judicial reasoning and the lineage of influential rulings.

This paper continues by providing a brief introduction to MPA, which we believe is necessary for a proper understanding of the subsequent analysis (Section 2; an illustration can be found in Appendix 1), and providing an analysis (Section 3) of how it can be applied in the legal domain, case law in particular. In the discussion (Section 4) and conclusion (Section 5) sections, we evaluate the potential of MPA and outline future work to advance its use for the partly automated tracking of case law evolution.

2 Main path analysis

Main path analysis, MPA for short, is an approach to examining (the amount of) knowledge flows along the links in a citation network "with a view to finding the main stream of research through that network" (Hummon and Doreian 1989). MPA not only traces the evolution of a topic, but also significantly compresses the underly-

ing intellectual structure by highlighting the critical events in the development of the topic and uncovering the major knowledge flows among them. MPA is performed in two steps. The first step is to measure the amount of knowledge flow along each citation link (*citation weighting*, Section 2.2). The second step is to extract the most significant path of knowledge flows (*path extraction*, Section 2.3).

2.1 Preliminaries

In theory, a case citation network should have no cycle, called *acyclic*, because citations should go from new to old cases. The approaches discussed in this section apply to acyclic networks. Practically, however, citation networks often contain small cycles, for instance when judgments cite each other because they were made on the same day or drafted in the same period by the same judges. This is also true for the case study presented in this paper. Cycles are usually manually broken based on simple rules such as deleting citations from older to newer nodes or randomly deleting some links until no cycle exists. In practice, MPA is often only applied to the largest connected component of a citation network, i.e., a subnetwork where every node (case) can be reached from some other nodes (cases).

In MPA, there are three types of nodes. *A source* is an old node without outgoing citations (the oldest judgment does not cite other judgments), while a *sink* is a new node having no incoming citations (the newest judgment cannot be cited). Both are boundary nodes of the citation network under analysis. Other nodes are called *intermediate*s. Because the purpose of MPA is to uncover how old nodes influence the new ones, the citation network under analysis is typically reversed into an *influence network*. This makes it more natural to explain the meaning of a main path. Figure 1 illustrates such an example, where smaller-numbered nodes represent older cases, and links going from old to new cases represent the direction of influence flows. In Fig. 1, sources and sinks are light- and dark-shaded, and other nodes are all intermediates. Finally, without loss of generality, given a link between two nodes u and v in

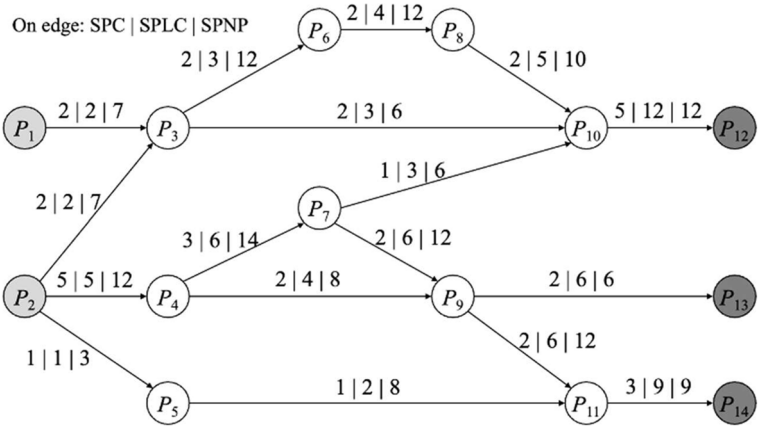


Fig. 1 A toy citation network and the traversal counts of all its edges

the influence network, denoted by $u \rightarrow v$, the starting point (u) and the ending point (v) are called the *tail* and *head* of the link respectively.

2.2 Citation weighting

According to the foundational work by Hummon and Doreian (1989), each link[1] is weighted by the number of knowledge diffusion paths (a.k.a. search paths) in the network that pass the link, called *traversal count*. In the literature, three main types of traversal counts have been employed. They are the Search Path Link Count (SPLC) and the Search Path Node Pair (SPNP) by Hummon and Doreian (1989), and the Search Path Count (SPC) defined in Batagelj (2003). Their definitions are similar, with the difference lying mainly in the origins and destinations of the search paths. In essence, SPC only considers the (search) paths between the boundary nodes. Given a link $u \rightarrow v$, SPLC expands it to relevant ancestors (of node *u*) and SPNP goes even further by involving relevant descendants (of node *v*). In the context of MPA, ancestors of a node *v* are the cases that directly or indirectly cite *v* while descendants are the cases that are directly and indirectly cited by *v*. Table 1 summarises the three types of traversal counts for a link $u \rightarrow v$.

Table 1 Summary of three traversal count-based citation weighting schemes

Method	Definition (given link $u \rightarrow v$)	Origins	Destinations	Standardisation
SPC	The number of search paths between all sources and sinks	Sources	Sinks	Connect pseudo-source s^* to all sources and all sinks to pseudo-sink t^*
SPLC	The number of search paths from all ancestors of u to all sinks	All nodes	Sinks	Connect pseudo-source s^* to all nodes and all sinks to pseudo-sink t^*
SPNP	The number of search paths from all ancestors of u to all descendants of v	All nodes	All nodes	Connect pseudo-source s^* to all nodes and all nodes to pseudo-sinkt^*

[1] Note that we followed Hummon and Doreian's convention (1989) to reverse the original citation network into influence network before performing main path analysis. Therefore, "citation link" here and in the remaining text essentially means a directed link from the tail to head.

Figure 1 illustrates the calculation of the SPX values, where X is either "C", "LC", or "NP". We use $P_3 \to P_{10}$ as an example. To calculate SPC, the origins and destinations are $\{P_1, P_2\}$ and $\{P_{12}, P_{13}, P_{14}\}$ respectively. The search paths that pass $P_3 \to P_{10}$ are $P_1 \to P_3 \to P_{10} \to P_{12}$, and $P_2 \to P_3 \to P_{10} \to P_{12}$. Therefore, $SPC\,(P_3 \to P_{10}) = 2$. Regarding SPLC, there will be one more origin node, P_3 itself, thus one more search path $P_3 \to P_{10} \to P_{12}$ will be counted. So, $SPLC\,(P_3 \to P_{10}) = 3$. Considering SPNP, not only the sink nodes but all nodes reachable from P_{10} are treated as traversal destinations, so the destination set is $\{P_{10}, P_{12}, P_{13}, P_{14}\}$. Compared to SPLC, three more search paths will be counted: $P_1 \to P_3 \to P_{10}$, $P_2 \to P_3 \to P_{10}$, and $P_3 \to P_{10}$, so $SPNP\,(P_3 \to P_{10}) = 6$.

2.3 Path exploration

Several ways exist to explore paths. Hummon and Doreian (1989) suggested a "priority first search" approach, which starts from all origin nodes and explores the next node iteratively by following the most weighted out-going link. Imagine a city planner who needs to identify the important roads to decide where to improve the city's infrastructure. If we want to find the most used route (path) from the mall to the train station, we may start at the mall, find all possible roads, and select, at every intersection, the road with the highest number of cars. In this road network analogy, we define road weight as the number of cars using each road per day (i.e., how many people use the road), with roads with a higher number of cars per day receiving more weights than roads with a lower number of cars. The planner aims at finding out the most traffic-heavy route, which is the main path. In Hummon and Doreian's approach, we always make the "best" choice locally at each intersection. The resultant route is likely to be important because each road is important at the corresponding intersection, although it does not guarantee that global optimality (we refer to Section 3.2 for an example of this case).

Hummon and Doreian's quasi-exhaustive path exploration method will enumerate paths based on SPNP. Using Fig. 1 for illustration, note that, when starting from P_3, the next link to explore is $P_3 \to P_6$ because $SPNP\,(P_3 \to P_6) > SPNP(P_3 \to P_{10})$. Similarly, the next link to explore starting from P_7 will be $P_7 \to P_9$, and so on. The collection of paths listed below were called "network of main paths". According to Hummon and Doreian, analysts may make use of their domain knowledge to determine which path is the most representative one, i.e., the *main path*.

$P_{10} \to P_{12}$
$P_8 \to P_{10} \to P_{12}$
$P_6 \to P_8 \to P_{10} \to P_{12}$
$P_3 \to P_6 \to P_8 \to P_{10} \to P_{12}$
$P_1 \to P_3 \to P_6 \to P_8 \to P_{10} \to P_{12}$
$P_9 \to P_{11} \to P_{14}$
$P_7 \to P_9 \to P_{11} \to P_{14}$
$P_4 \to P_7 \to P_9 \to P_{11} \to P_{14}$
$P_2 \to P_4 \to P_7 \to P_9 \to P_{11} \to P_{14}$
$P_{11} \to P_{14}$
$P_5 \to P_{11} \to P_{14}$

Considering the acyclic nature of a citation network, Batagelj (2003) suggested using the most weighted path among all possible paths as the main path, also known as *critical path* or *longest path*. In the road example, the most weighted path would con-

sist of the roads that most people take, i.e., the route with highest total road weight, which would be $P_2 \rightarrow P_4 \rightarrow P_7 \rightarrow P_9 \rightarrow P_{11} \rightarrow P_{14}$, according to SPNP. An efficient polynomial algorithm exists for finding the critical path in an acyclic network (Cormen et al. 2009, §24.2). Chen et al. (2022) integrated the critical path approach with the "network of main paths" concept. A candidate set of main paths were formed by extracting the single source critical path from each node, and further analysis was done by domain experts to choose a representative subset.

Different methods yield different results. In a forward search, one would start from the mall and keep selecting the busiest roads until the train station is reached. In backward search, one starts at the train station and traces backward to the mall, using the most-used roads. Finally, the key-route method would find the busiest road section and then expand in both directions (forward and backward). Liu and Lu (2012) proposed these three widely used local approaches for main path extraction. Formally, a *forward local main path* is extracted by starting path exploration from source nodes to sinks in a priority-first fashion (recall that a citation network is typically reversed), where "forward" means from old to new cases in our context, while a *backward local main path* is done in the reverse direction, starting from sink nodes to sources. The third and possibly the most widely used method is the *key-route local main path* approach. Liu and Lu argued that the most highly weighted link(s), called *key-route(s)*, should be maintained in the main path, so path exploration starts from the *key-route* and expands in both directions, backward until reaching sources and forward until reaching sinks.

The case study (Section 4) presented in this paper was done using the key-route main path approach, so it is demonstrated in more detail here. In Fig. 1, according to SPNP, the most significant link is $P_4 \rightarrow P_7$ ($SPNP(P_4 \rightarrow P_7) = 14$). Starting from P_4 to explore backward, the partial key-route main path is $P_4 \leftarrow P_2$ ("$\leftarrow$" represents backward). Exploring in the forward direction from P_7, we get the partial main path $P_7 \rightarrow P_9 \rightarrow P_{11} \rightarrow P_{14}$. So, the key-route main path is again $P_2 \rightarrow P_4 \rightarrow P_7 \rightarrow P_9 \rightarrow P_{11} \rightarrow P_{14}$. It is easy to verify that, according to SPNP, the forward main path of this small network is also $P_2 \rightarrow P_4 \rightarrow P_7 \rightarrow P_9 \rightarrow P_{11} \rightarrow P_{14}$, where the starting point is the source P_2 because, among all sources, P_2 has the most heavily weighted out-going link, i.e., $P_2 \rightarrow P_4$ ($SPNP(P_2 \rightarrow P_4) = 12$). Differently, the backward main path of this network is $P_1 \rightarrow P_3 \rightarrow P_6 \rightarrow P_8 \rightarrow P_{10} \rightarrow P_{12}$, because P_{12} is the starting point for backward path exploration as it has the most significant in-coming link, i.e., $P_{10} \rightarrow P_{12}$ ($SPNP(P_{10} \rightarrow P_{12}) = 12$) among all sinks.

2.4 Caveats

There has been discussion about which path exploration method to choose. Longest path was recommended by Batagelj (2003) because it shares the same concept as task scheduling optimisation in operational science, called *critical path*. Other names include top path (Fortana et al. 2009, Martinelli 2012) and global main path (Liu and Lu 2012). However, a single path can hardly represent all the important development pathways of a large domain, which is the rationale for integrating the global main path with other three local main paths (Liu and Lu 2012). An integrated view can

also be obtained by extracting and merging several main paths of the same type, such as forward main path in Verspagen (2007) and critical main path in Fontana et al. (2009). As for extracting the top-K critical paths, fortunately, efficient algorithms of polynomial time and space consumption exist (Chen et al. 2022). Concerning the other three best-effort path exploration methods, Liu and Lu (2012) argued that key-route main path seemed to produce the most relevant results, although all four types of main paths collectively provide more comprehensive and complementary views. As far as we can see, critical path (e.g., Chen et al. 2022) and key-route main path (Jiang et al. 2020; Jiang and Liu 2023) have been the more popular methods.

There is also a debate around citation weighting. While some studies reported similar results, many studies also reported on the discriminative impacts of different SPX methods on the analysis result. From a knowledge flow point of view, Batagelj (2003) suggested SPC as a "first choice" due to its "nice" property that the total inflow to a node is equal to its total outflow. Liu et al. (2019) argued that SPLC "is the closest to the knowledge diffusion scenario in science and technology development" because intermediate nodes also produce knowledge and transmit the knowledge to the current nodes (i.e., sinks). Using a similar knowledge production and diffusion analogy, Kuan (2020) argued treating each intermediate as a knowledge consumer (as being influenced) and conjectured SPNP as a more comprehensive measure is reasonable. The community has not reached a consensus on the most proper SPX method.

Improvements to traversal count-based citation weighting exist. Long knowledge diffusion paths may cause a bias towards old nodes. To counteract this, Liu and Kuan (2016) proposed to use (length-based) *knowledge decay* to adjust the significance of search paths of different lengths. Jiang and Zhuge (2019) adopted similar ideas using an exponential knowledge decay that is empirically supported by the phenomenon of "attention decay in science" (Parolo et al. 2015): longer paths contribute significantly less to traversal counting compared to shorter paths. The two studies differ in that, given a link, Liu and Kuan only decay partial search paths from origins to the link, while Jiang and Zhuge decay the complete search paths from origins to destinations that pass the link. These methods offer a way to parameterise the length and speed of knowledge diffusion. Analysts can cap long diffusion paths, assuming they can obtain more topically coherent main paths. Different thresholds of path length cap can be tested, as well as the types and parameters of the knowledge decay function.

Another caveat concerns the measurement and balancing of comprehensiveness and representativeness. We define *comprehensiveness* of a main path as how many important events of a field are identified, in other words, recall of important nodes. It is related to the *representativeness* of a main path, which is defined as how many identified events are indeed important events of a field, i.e., the precision of main path nodes. Barberá-Tomás et al. (2011) showed that the proportion of real significant events about scientific or technological development on a main path, most of the time, is higher than that of the whole network, which is empirically corroborated by the current paper (refer to Section 4.4 for the detailed experimental results). However, it is not guaranteed that this method can identify all or most of the significant events that form the domain under analysis. Nor does the main path guarantee to include only significant events. The situation drastically worsens for large networks. Indeed, Filippin (2021) applied the critical path method to an extremely large cita-

tion network of patents in the semiconductor industry and concluded that "the main path cannot be said to represent the backbone of the technological trajectory of the semiconductor industry". This was not because most of the patents on the main path were unimportant, but the low representativeness was mainly a byproduct of the low comprehensiveness: A lot of key players were either absent (e.g., ASML, TSMC, Samsung) or severely underrepresented (e.g., Intel and Toshiba), while some players were overrepresented (like IBM).

Filippin (2021) was one of the few studies that performed an objective evaluation of the quality of a main path relying on a Likert-scale expert evaluation of main path patents' relevance to the real technological development of the semiconductor industry. MPA resulted in a relatively low precision (about 31–37% by two experts), whereas recall was not reported. Jiang and Liu (2023) performed a more comprehensive quantitative evaluation on three popular topics of computational linguistics research. The recall of their method was significantly improved but still not satisfactory, achieving around 39.7% on automated text summarisation, 49.7% on natural language parsing, and only 24.4% on machine translation – a much larger network of publications than the other two.

In most MPA studies, representativeness is often overseen by comprehensiveness by extracting a network of main paths. A simple way is to extract several main paths, as we did in the analysis below (c.f., Verspagen 2007; Ho et al. 2014, Jiang et al. 2020, Jiang and Liu 2023). A more subtle issue is that more recent developments may be missed, for example, by the key-route main path approach because recent nodes may have received fewer citations, and key-routes often appear in the middle of a citation network. Ho et al. commented that while multiple main paths may converge into the main stream, additional *branch paths* should be added to compensate for often underrepresented recent developments. They proposed to extract branch paths from a few top-cited recent nodes about certain still vibrant topics in a bidirectional way. Similar claims have been made in Kim and Shin (2018), arguing that a domain may go through "several generations of major technological changes, emergence of derivative technologies, and hundreds of minor improvements", so the technical historiograph "must have a main path, derivative paths and technology junctures." We expect to see more profound research on this aspect – developing and validating novel methods for finding meaningful branch paths or derivative paths to improve the comprehensiveness of MPA results.

3 Consumer right protection: an MPA-based qualitative investigation

The case study presented in this paper concerns consumer protection in EU law. We selected CJEU judgments that carried the label 'consumer protection' in the 'subject matter' category of the metadata. EUR-Lex and its underlying database CELLAR provide metadata that includes subject matter. It is not known who exactly tagged whether judgments concerned 'consumer protection'. An inspection by the authors found the labelling to be accurate. The judgments and their metadata were retrieved through (1) a SPARQL endpoint that allows querying RDF data from the CELLLAR

repository, which holds metadata and full-text documents related to EU legal publications in a machine-readable format,[2] and (2) web scraping. We extracted basic information about the cases, including CELEX ID, ECLI ID, Title, CELEX Keywords, and Operative Part, as well as the citations to other cases within the collection. Each judgment in our collection contained CELEX identifiers for the judgment itself as well as for other judgments cited within it. Using this metadata, we were able to capture the network of citations without performing full-text parsing.

Next, the citation network was constructed, which contained 446 cases (nodes) and 1231 citations (edges). The network is not dense. In addition to one large component comprising 338 cases and 1226 citations, it also included 93 small isolated acyclic components, 1 component of four nodes, 1 component of three nodes, 10 components of two nodes, and 81 isolated nodes. Our analysis was performed on the largest component. This contained one strongly connected component of three cases that mutually cite each other, i.e., 3 cases (62000CJ0052, 62000CJ0154, 62000CJ0183) and 6 citations. The three judgements were made on the same day (25 April 2002) by the same chamber (the Fifth Chamber). It was not straightforward for these cases to decide which citations to remove to break cycles. Luckily, JMPA package (Java package for MPA) (Jiang et al. 2020) can handle citation networks with small cycles like this.

To illustrate the result of MPA analysis on the case network about consumer right protection, SPC was used to calculate citation weights and *top-3 key-route main paths* were extracted. Figure 2 illustrates the network of extracted main paths, where the main path cases extracted from different time slices are represented in different colours and shapes. The operative parts, which are a short form of the interpretations of the rules, are made available in the supplementary materials (https://doi.org/10.34894/OETCQ1). To better uncover the dynamics of the domain evolution, the original citation network was split into five time slices based on the following time spans with a changing end year: [1980, 2000] (extracted main paths are visualised in circular shape), [1980, 2010] (in diamond shape), [1980, 2015] (in rectangular shape), [1980, 2020] (in triangular shape), [1980, 2023] (in hexagonal shape). On each time slice, the top-3 key-route main paths were extracted by exploring bidirectionally from the three links with the highest traversal counts. Then, the main paths extracted from each time slice were merged into one network, called *main path network* in Jiang et al. (2020).

3.1 Key topics in the EU consumer protection scholarship

To identify the key concepts and thematical developments in the EU consumer protection law, we reviewed leading doctrinal scholarship such as textbooks on general EU law and topic-specific articles dedicated to the analysis of a distinct legal development in this field of the Internal Market law. Rather than aiming to map individual cases, our review allowed us to trace the core themes and regulatory developments that structure the formation and evolution of the EU consumer protection law. These

[2] https://publications.europa.eu/webapi/rdf/sparql.

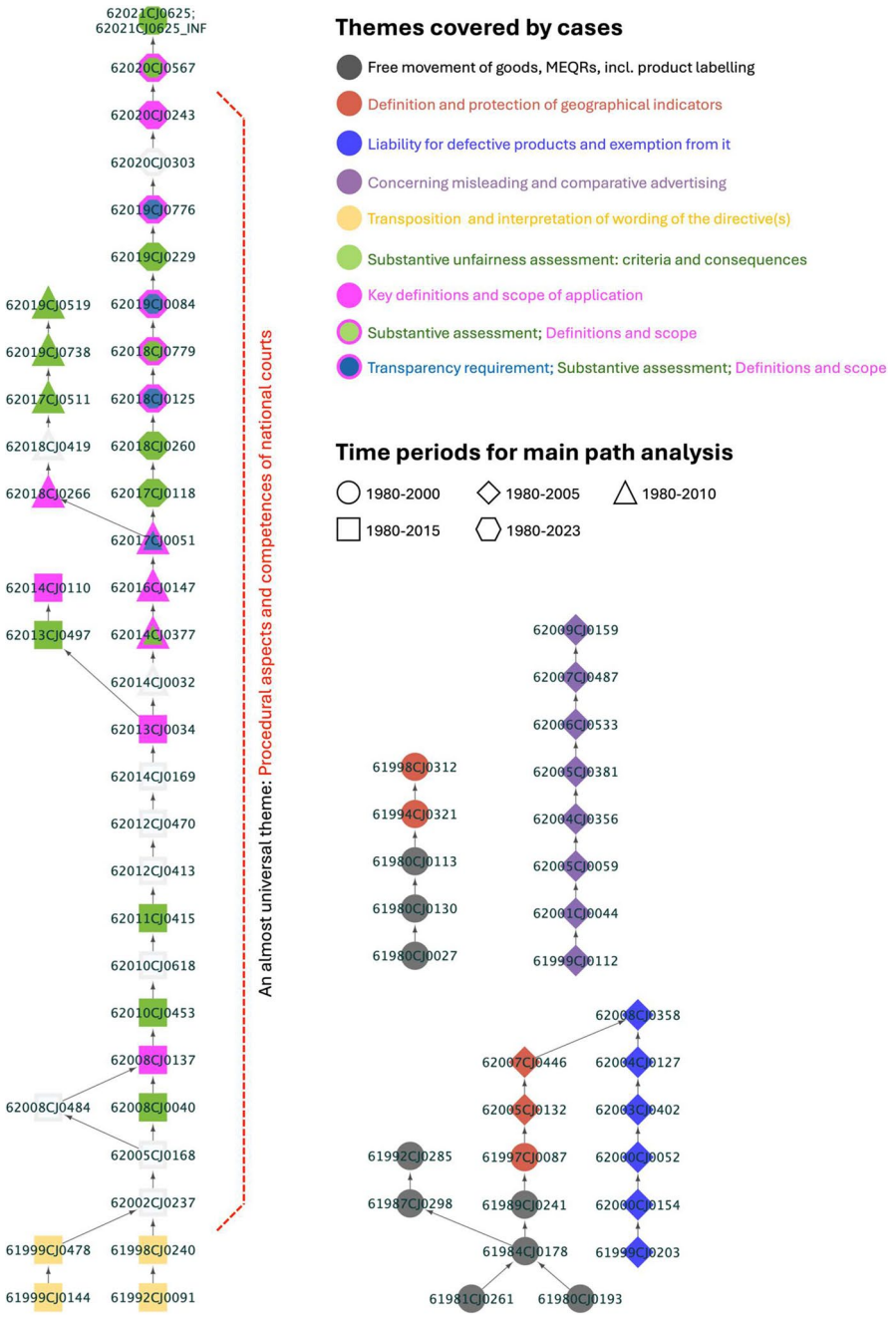


Fig. 2 Main path network of the consumer rights protection cases in the court of justice of the EU. The paths of circular nodes (on the right), were extracted from the 1980–2000 time slice of the complete case citation network. One path extracted from the 1985–2025 time slice converged with a circular path covering cases 61984CJ0l93, 61981CJ0261, 61984CJ0l 78, 61989CJ0241, and 61997CJ0087 (bottom right), so only the new cases (i.e., 62005CJ0l32 and 62007CJ0446) were represented in a diamond shape

core themes were then used as an analytical baseline for the assessment of the main paths determined through the MPA methodology.

3.1.1 Consumer protection through market integration, negative harmonisation, and the 'average consumer'

Barnard & Peers (2020) identify the adoption of the Unfair Terms in Consumer Directive 1993/13, which introduced minimum controls in consumer contracts and Consumer Sales Directive 1999/44, which set the minimum quality standards for goods as the **key moments that started to affect core areas of private law**, sales law in particular, within national commercial law regimes. Firstly, the authors addressed the "**negative harmonization**" impact of EU law on national consumer protection rules, which mostly concerned the **removal of national consumer protection rules** that impeded the development of the internal market without having any real consumer protection purpose (Barnard and Peers 2020). The CJEU developed a line of case law extending the prohibition of **quantitative restrictions** under Article 34 TFEU to **measures having equivalent effect** (MEQRs), which significantly shaped the limits of national measures regulating market access (Barnard 2022; Lianos 2014). This includes, for example, **national product and marketing** rules (e.g., German Beer Purity Law[3]; Italian durum wheat pasta rules[4]; Cassis de Dijon[5]) as well as **advertising** rules (e.g. Clinique case[6]).

In the context of market integration, the literature also addresses the development of **the 'average consumer'** concept as the central benchmark in the EU law on market access and consumer protection law (Leczykiewicz and Weatherill, 2016). The concept has been shaped and widely used by the CJEU to assess when national measures are overly restrictive (Barnard and Peers 2020; Barnard 2022). At the same time, scholarship highlights the tension between the abstract and moving benchmark and the realities of consumer behaviour, as well as the difficulties the Court faced in balancing national consumer protection with the objective of the freedom of movement of goods in the EU internal market (Mak 2012; Twigg-Flesner 2016; Weatherill 2021).

3.1.2 Information obligations and consumer rights

Secondly, Barnard and Peers (2020) define **information obligations** as the earliest EU consumer law measures, for example, those resulting in the Unfair Commercial

[3] Court of Justice of the European Union. (1987). Commission of the European Communities v. Federal Republic of Germany, Case 178/84, ECLI:EU:C:1987:126, European Court Reports, 1987–01227.

[4] Court of Justice of the European Union. (1988). *3 Glocken GmbH and Gertraud Kritzinger v. USL Centro-Sud and Provincia autonoma di Bolzano*, Case 407/85, ECLI:EU:C:1988:401, European Court Reports, 1988–04239.

[5] Court of Justice of the European Union. (1979). *Rewe-Zentral AG v. Bundesmonopolverwaltung für Branntwein*, Case 120/78, ECLI:EU:C:1979:42, European Court Reports, 1979–00649.

[6] Court of Justice of the European Union. (1994). *Verband Sozialer Wettbewerb eV v. Clinique Laboratoires SNC and Estée Lauder Cosmetics GmbH*, Case C-315/92, ECLI:EU:C:1994:34, European Court Reports, 1994 I-00317.

Practices Directive (UCPD, Directive 2005/29/EC). Thirdly, the authors identify the **right of withdrawal** as one of the core consumer protection instruments. The right first appeared in 1985 Doorstep Selling Directive, and was subsequently standardised in Consumer Rights Directive for distance and off-premises contracts.

The rules on information obligations and withdrawal are defined as the EU measures adopted to ensure freedom of contact, enabled through a full-knowledge consumer equipped with substantive, non-excludable rights (Howells et al. 2017). From this perspective, the literature conceptualises the EU consumer protection law as being framed by a particular model of consumer behaviour, in which informed and reasonable decision-making plays a central role, but also underlines the limitations of such approach in practice as national expectations from such a standard continues to differ (Twigg-Flesner 2016; Weatherill 2021).

3.1.3 Product safety, technical harmonisation and liability

Barnard and Peers (2020) then proceed to more specific topics. The new approach to **technical harmonization** is discussed including general **product safety** (General Product Safety Directive, currently a Regulation); and specialised legislation (e.g. protective equipment, toys, etc.). **Product liability** is another limb of the EU approach towards a safe internal market by ensuring compensation schemes for those injured by defective products. The product liability regime was established to remove obstacles created by the limitations of contract (privity) and tort (fault requirement) within national legal systems (Verheyen 2018).

3.1.4 Unfair contract terms[7] and judicial control

The authors (Barnard and Peers 2020; Barnard 2022; Josipović, 2024) subsequently addresses the topic of **unfair terms** including such principles as good faith; significant imbalance of bargaining power; procedural and substantive fairness. The scholarship underlines the effects that Unfair Terms in Consumer Contracts Directive has caused on national procedural law: (1) national rules should not impede the effectiveness of the enforcement of Directive's aims; (2) the introduction of the *ex officio* doctrine requiring courts to consider unfairness of terms on their initiative (Barnard and Peers 2020). The Directive also shows the EU preference for preventive controls and injunctions against unfair terms brought by public authorities or consumer groups. In this context, the literature also highlights the increasing importance of harmonised procedural mechanisms and judicial control as integral components of consumer protection (Twigg-Flesner 2016).

[7] A more detailed review of the relevant doctrinal developments is provided in the Special Issue on 30 Years of Unfair Contract Terms Directive (2023) Issue 4, Volume 19, *European Review of Contract Law.* The issue identified a wide range of topics related to the consumer protection under this directive, including transparency requirements; standard contract terms in the digital age; blacklisted terms in digital markets; the interplay between national courts and the CJEU; and the concept of a weaker or vulnerable party in consumer contracts.

3.1.5 Harmonisation techniques

Barnard and Peers (2020) continue by addressing **the Consumer Sales Directive 1999 as a minimal harmonisation** tool, following the failure to agree on a Common European Sales Law. The more recent **Consumer Sales Directive 2019 (CSD)** represented a shift towards **maximum harmonisation**.Under this legal regime, goods must comply with the contract in terms of quality of goods; delivery; characteristics of goods; and reasonable expectations. The authors illustrate how minimum harmonisation sets a minimum set of expectations, whereas maximum harmonisation constrains the offerings on the market and costs associated with them for all market participants. The **Digital content and digital services directive** (DCSD) 2019 supplements the legal framework by addressing similar issues in the digital environment and specifies the burden of proof and remedies (repair; replacement; termination or price reduction for the CSD; cure as the primary remedy for DCSD). The authors specifically focused on air passenger rights as an example of where the EU provides minimum rights that cannot be deviated from.

3.1.6 Access to justice, enforcement and effective redress

Finally, the authors (Barnard and Peers 2020; Twigg-Flesner 2016) address **enforcement and redress**. Member States can enforce regulation involving the criminal or administrative sanctions. The EU requires them to provide 'adequate and effective' means of enforcement.[8] The Consumer Protection Cooperation Regulation strengthens the legal framework and requires each Member State to nominate a competent consumer authority to be part of the EU-wide enforcement network. The Consumer Injunctions Directive further requires Member States to grant qualified entities from other states access to their injunction procedures (too complex as in Duchesne case). The authors also cover **access to justice and collective redress** for consumers (Barnard and Peers 2020).

3.2 Main themes

Figure 2 shows the extracted paths, where node shape represented the time slice from which the corresponding cases were extracted to form the main paths and node colour coded the themes that were covered by the main path cases (corresponding to the colour scheme for the UCTD themes that we have developed at the end of this subsection). The main trunk in the left part of Fig. 2 covering the rectangular, triangular and hexagonal paths can be labelled as unfair terms in consumer contracts and are about several directives of European council, including mainly Directive 93/13/EEC but also 1999/44/EC, 2005/29/EC and 2008/48/EC. Overall, the **rectangular** path and sub-paths (extracted from the 1980–2010 time slice) include judgments on:

[8] A detailed review of case law on the application of the principles of effectiveness, proportionality and Article 47 of Charter of Fundamental Rights (right to effective remedy and fair trial) has been provided in the framework of the research project Roadmap to European Effective Justice (Re-Jus): Judicial Training Ensuring Effective Redress to Fundamental Rights Violations, and published in Re-Jus Casebook – Effective Justice in Consumer Protection (2018).

- (Light-yellow: Theme 1) Incomplete transposition, indirect effect, and national interpretation of the wording of the directive (e.g., incorrect or incomplete transposition of the directive; duty of national courts to interpret national law in conformity with the directive even when it has not been (correctly) transposed (indirect effect); non-contractual state liability for incomplete transposition and implementation; guidance on the interpretation of national law provisions and the link between national law and UCTD obligations).
- (A universal theme in red: Theme 2) The scope of competences of national courts when reviewing contractual terms (who is allowed to bring actions when and/or at which court; what national legislators and courts are allowed and not allowed to assess; the evaluation of national legislation in light of the unfair terms directive; ex officio review of unfair terms by national courts; burden of proof).
- (Pink colour: Theme 3) Definitions (e.g., consumer status; seller or supplier; standard terms of contracts; individually negotiated terms; other concepts related to the scope of the directive's application).
- (Green colour: Theme 4) Criteria to be considered when assessing contractual terms (e.g., what qualifies as a consumer; significant imbalance of contract terms; overall assessment of contracting circumstances).
- (Light grey: Out of the nine main themes) Consumer protection more generally (e.g., procedural rules and access to justice; whether contract amendments reflecting national law requiring sellers to propose changes falls under the directive).

The **triangular** path (extracted from the 1980–2015 time slice) provides further clarification of:

- (A universal theme in red: Theme 2) Procedural aspects (e.g., national procedural rules affecting consumer rights protection; role of national courts and the limits of their intervention; assessment standards for fairness).
- (Pink colour: Theme 3) Key definitions.
- Consumer rights (e.g., terms may be maintained even if unfair provided that a consumer agrees).

The **hexagonal** path (extracted from the 1980–2023 time slice) subsequently addresses the:

- (A universal theme in red: Theme 2) Role and competences of the national courts in applying the UCPD.
- (Green colour: Theme 4) Consequences of unfair terms (removal of unfairness; consequences for the validity of the remaining terms; how unfair terms affect obligations)).
- (Dark blue: Theme 5) Transparency criteria in consumer contracts (transparency test for consumer contracts; clarity and intelligibility of contract terms; effects of the formulation of terms on the consumer decision-making).
- Effects of mandatory national legal provisions.

The paths of **circular** and **diamond-shaped** nodes address different areas of consumer protection law, covering themes outside the scope of UCTD. The judgments on the **circular** paths concern the area of *free movement of goods* and *measures equivalent to quantitative restrictions* (MEQRs, prohibited by Article 30 of the Treaty), more specifically what can constitute MEQRs and balancing national consumer interests with the notion of a European Single Market (e.g., shifts in legal principles regarding trade restrictions, product labelling, and market access within the EU) (Both in dark black: Theme 6).

The **diamond-shaped** path then continues with judgments about other consumer protection legal frameworks, such as the *definition and protection of geographical indicators* (continuation of free movement and MEQRs) (Brown colour: Theme 7), *product liability* or liability for defective products (meaning of maximum harmonisation; interpretation of national transposition laws; what national law is allowed and not allowed to regulate regarding the substitution of defendants) (Light blue: Theme 8), and *comparative advertising* (aspects of what can be compared; what can be considered as fair use of the symbolics of the competitors and the limits of such use as regards to consumer protection) (Purple colour: Theme 9).

We developed a set of thematic codes for the themes that have been discussed above and manually annotated the operative parts of all main path cases. The code list and the corresponding colour scheme are summarised below. Themes of the main trunk were coded using different background colours while themes of the other smaller paths isolated from the main trunk were coded using text colours (see Supplementary File 1 at https://doi.org/10.34894/OETCQ1).

UCTD Themes:

Theme 1 Transposition of the directive and the interpretation of the wording of the directive
Theme 2 Procedural aspects and competences of national courts
Theme 3 Key definitions and scope of application
Theme 4 Substantive unfairness: criteria and consequences (Note that this theme merged the criteria to be considered when assessing contractual terms, the consequences of unfair terms and the effects of mandatory national legal provisions)
Theme 5 Transparency and information requirements

Other themes outside the scope of UCTD (typically in earlier years):

Theme 6: Free movement of goods and measures equivalent to quantitative restrictions (including product labelling)
Theme 7 Definitions and protections of geographical indicators
Theme 8 State's liability for defective goods and exemption from product liability
Theme 9 Misleading and comparative advertising

3.3 Topics in literature versus topics in MPA paths

When comparing the topics in the literature to the ones observed through MPA, it can be concluded that most of the topics mentioned in the literature can be found on

the paths. Table 2 provides a mapping of the topics mentioned in the literature to the ones produced by MPA. It is important to underline that the alignment occurred at the conceptual level: the paths traced by MPA captured the same themes, though not necessarily through the identical cases addressed or most cited in the scholarship. Conversely, MPA did not yield additional topics, at least not general topics, compared to the literature.

Although the MPA approach provided an acceptable coverage of the topics in the literature, we believe the first path (combining rectangular, triangular, and hexagonal paths) can be shortened if one aims to trace the evolution of how the Court interprets and applies consumer protection law. Several cases address very similar issues. This concerns, for instance, judgments about whether sellers are allowed to claim compensation on the basis of alternative national legal provisions after compensation under contract is no longer possible due to that contractual term being considered unfair and non-binding by the court. Another example concerns the brown nodes in the earlier paths on the Geographic Indicators, which may be merged with the path with judgments about MEQRs.

Only occasionally it turned out to be difficult to find the exact relation of the identified case to the path (e.g., 62013CJ0497 (2015)), which only marginally addresses the UCTD and focuses mostly on other questions and legislation. In most instances, it is easy to follow the connections between the cases.

Note that in Fig. 2, some branches that were extracted from the network slices corresponding to earlier time periods are disconnected from the main trunk. This phenomenon usually implies a shift of focus in the legal debate in the underlying domain. Nevertheless, disconnectedness is ultimately an undesirable trait as it may cause misconceptions that various branches evolved independently and separately without interaction, which is often not true. Surprisingly, this issue has been repeatedly overlooked by the MPA community.

Additionally, we measured the time spent on the interpretation of the MPA. If interpreting the paths takes too long, MPA may not be practical for real-world applications, it may produce too many paths or paths that are difficult to understand, it may indicate that the extracted paths are not intuitive or representative, and the process may be difficult to scale as it would require too much human intervention.

Regarding the time, it turned out challenging to precisely calculate the time spent on each case, as the human analysis required not only interpreting the operative parts in isolation, but also in relation to preceding and subsequent judgments on the paths. Nevertheless, we were able to draw some general observations and, to a certain extent, time indications. One general observation was that the operative parts in some of the cases were rather easy to summarise. This particularly concerned judgements where there was only one question, where the answer to the preliminary question was formulated in a plain manner, and where it mostly restated the phrasing of the directive itself (e.g., 62020CJ0243). In such cases, which constituted the minority of cases included on the paths, two–three minutes were sufficient to grasp the judgement and its impact. Cases that took more time were instances where the court answered multiple questions (the court sometimes considered up to five or six different questions), interpreted the application of the UCTD in relation to other pieces of EU law, making it necessary to identify how the interpretation of different pieces of legisla-

Table 2 Mapping literature to MPA topics, and vice versa

Key themes identified in the literature	Related legal concepts mentioned in the literature	How the themes appear in the paths
Market integration and negative harmonisation	Quantitative restrictions (QRs), measures having equivalent effect (MEQRs), mutual recognition, distinction between product requirements and selling arrangements	Case law identified in the paths thoroughly address the theme of negative integration. The path with judgements on MEQRs included several landmark and most-referenced cases in the scholarship on this topic (such as '*Dassonville Formula'* 8–74; '*The Vinegar'* case 193/80; '*The Margarine'* 261/81, or 'The Beer Purity Law' 178/84 cases) while the landmark case on mutual recognition established in *Cassis de Dijon* 120/78 has not been identified by the paths. Some other important cases that cover the exemptions for MEQRs, and the difficulty the Court had in distinguishing between 'product requirements' and 'selling arrangements' have not been captured
Concept of 'Average Consumer'	Market access; fairness assessment in the application of the UCTD and UCPD	The case law identified in the paths cover the concept of an average consumer in relation to the application of the UCTD and UCPD
Positive Integration and consumer protection	Unfair Contract Terms Directive (UCTD); Unfair Commercial Practices Directive (UCPD); Consumer Rights Directive; withdrawal rights; transparency and fairness	The case law identified in the the paths cover the application of the UCTD (scope of application; concepts of fairness and transparency; procedural aspects) in detail; certain elements of the UCPD are addressed too (assessment of fairness)
Harmonisation of consumer law	Minimal vs maximal harmonisation; standardisation of consumer rights	The case law that addressed the type of harmonisation is present in the paths that cover application the UCTD and the PLD
Procedural rules and judicial control	*Ex officio* control; effectiveness; procedural autonomy; injunctions; role of national courts; indirect effect of directives; liability of Member States for failing to (correctly) transpose directives	The case law on procedural rules of consumer protection is present in the path, particularly, the principle of 'indirect effect' is addressed by the cases identified in the paths in the context of the UCTD. The path also captured the landmark decision in *Faccini Dori* 91/92 (CELEX No. 61992CJ0091). Case law on *ex officio* judicial control is present on the paths too
Non-Discrimination and Fundamental Rights	Effective remedies, procedural rights	Case law on the topic of fundamental rights protection is identified in the paths in the context of the cases on application of the UCTD (consumer procedural rights)
Emergence of e-commerce and legislative responses (distance contracts, platform regulation, digital content)	Distance contracts; platform regulation; online contracting	The case law identified in the paths does not directly address the concept of e-commerce, however, much of the case law identified in the paths on the UCPD is often referenced by the court when addressing the issues related to online contracting. Platform regulation and digital content are not explicitly discussed in the identified cases; however, the paths on the UCTD cover the major topics of transparency, which play a fundamental role in cases on digital content platforms

Table 2 (continued)

Key themes identified in the literature	Related legal concepts mentioned in the literature	How the themes appear in the paths
(Product Safety and Liability;	Defective products; product safety and risk allocation	The process of harmonisation of product safety and liability is captured in the paths through the clarification of certain central definitions of the relevant legislation (e.g. 'putting into circulation'); interpreting scope of liability; and identifying liable parties
Access to Justice	Alternative Dispute Resolution and Online Dispute Resolution	Case law addressing access to justice is present in the pathes, in particular in relation to the application of the UCTD
Digitalisation and new technologies	Digital content; digital services; online platforms regulation; algorithmic decision-making	These topics are partially reflected in the paths through case law on online commercial practices, No cases explicitly related to platform liability or digital content are identified. These themes are relateively recent in the EU law, and, thus, the case law remains limited

tion affected the scope of the application of the UCTD, or addressed not only procedural questions but also how case law impacted substantive law provisions (e.g., cases related to consumer credit). In such cases, the analysis took more time: around five-seven minutes per case in shorter cases or 10 min in more complicated cases. Additional time was sometimes required because it was necessary to revisit either the questions asked to the court or the judgement to understand the context.

For some paths, particularly the smaller ones, we could obtain approximations of the time spent reviewing the operative parts and tracing the interconnections. The circular paths could be analysed in approximately 10 min (path in the middle) and 30 min (path on the bottom), respectively. The circular path in the middle (see Fig. 2) was completed relatively quickly as the case law was straightforward and commonly addressed one main question only. The interpretation of the diamond-shaped path connected to the circular path on the definition and protection of geographical indicators and product liability took approximately 30 min, whereas a similar amount of time was spent on the diamond-shaped path on comparative advertising.

3.4 Key-node exploration

In addition to analysing whether the paths include the topics mentioned in the literature (key-path exploration), we were interested in whether the most important judgments are covered on the paths (key-node exploration). Contrary to what Liu and Lu (2012) claimed, Filippin (2021) showed that path exploration from key-routes might be less useful than expected in improving comprehensiveness. The fact that many key patents issued by key players were missing inspired an alternative way to explore main paths from pre-identified key nodes. Kuan (2020, 2024) recently formalised this idea as *key-node search*. Kuan suggested using prior knowledge of a domain to identify some key events (or milestones) and argued that *key-node main paths* may better

reflect the real technical trajectories initiated from the milestones.[9] Similar ideas have been applied earlier, e.g., in Park and Magee (2017), Lu et al. (2021).

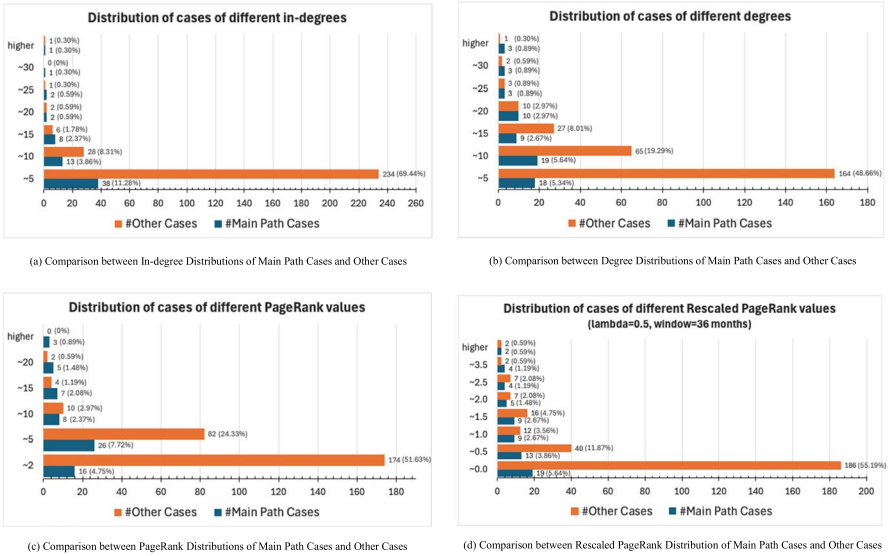
A challenge in a key-node exploration in a case citation network is determining the nodes' (i.e., cases') relevance or importance. For a legal expert, what is relevant depends on the legal question. Hence, cases can be or cannot be relevant, depending on what the expert is seeking or interested in. Moreover, measuring whether each case is relevant can be time-consuming and resource-intensive. This study uses nodes' centrality scores as a proxy for relevance. As explained in the introduction, cases with lower centrality scores can generally be considered less relevant, while those with higher scores are deemed more important.

On the citation network under MPA analysis (i.e., the largest connected component of the original case citation network), we calculated various centrality measures, including in-degree (number of incoming citations), out-degree (number of outgoing citations), degree (incoming + outgoing citations), betweenness (Freeman 1977), and PageRank (Page et al. 1999). Particularly in-degree and degree have proven to correlate well with importance and relevance (van Dijck et al. 2023). A variant of PageRank that has been proven successful in identifying scientific milestones and resolving the "time bias" (Jiang et al. 2016) was also applied: Rescaled PageRank, which normalises the PageRank values of each case using all the cases that are decided on in a time window centring around it (Mariani et al. 2016). MPA does not provide a ranked list as centrality measures do, so there is no straightforward way to quantify the correlations. Instead, we compared the proportions of nodes with high and low centrality scores on the paths to those in the complete network.

Ideally, we would like to observe that the proportion of judgments with a high centrality is substantially higher than the proportion of judgments with a low centrality score. The data indeed reveal such a pattern. The histograms reveal that when using in-degree, PageRank, and degree, the high-centrality cases identified as main path cases outnumber those missed. These three methods seem to correlate well with main path analysis in identifying the important cases. In contrast, using betweenness or Rescaled PageRank, the number of missed high-centrality cases outnumber the main path cases by a nonnegligible margin, which suggests that the incorporation of betweenness analysis or Rescaled PageRank may have a significant impact on the quality of main path analysis (Fig. 3).

The observation that the main path mostly includes judgments with a high centrality and few with a low centrality is confirmed by Table 7 in Appendix 1. This table lists the centrality analysis results of all 65 main path cases displayed in Fig. 2. Different font styles are used to indicate how well main path analysis 'correlate' with different centrality analyses: The **bold underlined** and **bold** fonts indicate that the main path case is ranked as the top-5 and top-65, respectively, by the corresponding

[9] Historically, Batagelj (2003) first mentioned summing link weights into node weights, which was later applied in his study about the formation of peer review as a research field (Batagelj et al. 2017), while Liu and Lu (2012) suggested averaging instead of summing. Further Jiang and Zhuge (2019) validated that the N^+ values of a reversed citation network, called *forward search path count*, outperformed PageRank and its variants as a strong citation impact indicator. Jiang and Zhuge's account proved that MPA is an integrative solution to citation network analysis because it is feasible to use MPA itself to identify the key-nodes and perform key-node main path analysis.



(a) Comparison between In-degree Distributions of Main Path Cases and Other Cases

(b) Comparison between Degree Distributions of Main Path Cases and Other Cases

(c) Comparison between PageRank Distributions of Main Path Cases and Other Cases

(d) Comparison between Rescaled PageRank Distribution of Main Path Cases and Other Cases

Fig. 3 Histograms with distribution of cases across centrality measures

centrality measure (because the main path network contains 65 cases); the *italic* font indicates that the main path case is in the top-110 ranked list according to the corresponding centrality, where 110 is used to approximate the first quantile, similar to JCR Q1 (Journal Citation Report, Quantile 1).

Several insights can be derived from Table 7. Firstly, the main path cases are often also the top cases ranked by different centrality measures (comprehensiveness). Vice versa, a good portion of top-ranked cases by different centralities are indeed on the main path network (representativeness), which may indicate the feasibility of combining centrality analysis and main path analysis in the future. Secondly, degree-based centrality measures are better at finding new main path cases, but PageRank and its variants perform relatively weakly at new cases. In contrast, PageRank performs better at identifying relatively old main paths before 2000, which may cause time bias due to the nature of the PageRank calculation (Jiang et al. 2016). However, most main path cases identified by PageRank are corroborated by either or both of the two Rescaled PageRank instances (window = 12 months or window = 36 months), which might indicate the local importance of these main path cases during a certain period after the decision dates of the judgments on the paths, whereas they might have ceased to be prominent cases today. Thirdly, there are eight main path cases identified by only one centrality measure (we treat the two variants of Rescaled PageRank as one method as their results do not differ a lot) and 12 main path cases identified by only two centrality measures. The voting results indicate the potential of fusing multiple centrality measures and using different thresholds for the voter to identify a potential list of important cases to guide or adjust the main path analysis.

We further analysed the coverage of high-centrality cases and the latter's relationships to the main path. With five centrality measures at hand, we defined high-centrality cases threshold h as those top-ranked by at least h centrality measures.

Table 3 shows the numbers of high-centrality cases, and those on and off the main path network at threshold 1, 2 and 3. The extracted main path network contains 66 cases. So, thresholds 2 and 3 are both reasonable choices for evaluation. A significant percentage of high-centrality cases (at thresholds 2 and 3) were covered by the main path network. Especially, at threshold 3, the coverage was close to 60%. A further analysis of the off-path high-centrality cases was presented in Fig. 4. To better reduce visual clutter threshold 3 is used, and the main path network is represented in solid grey nodes and black edges, with other cases and case citations slightly shaded away. High-centrality cases on and off the main paths are represented by bold black and red node boundaries, respectively. Details of the off-path high-centrality cases can be found in the supplementary materials (https://doi.org/10.34894/OETCQ1).

Most off-path high-centrality cases are directly linked to the main path network. For demonstrative purposes, some of them are labelled with their case IDs, plus the IDs of the connected main path cases (in a smaller font size). In this network, two off-path high-centrality cases are not directly linked to the main path network, highlighted in yellow colour and a bold red boundary. This concerns cases where national law provides for a general prohibition on sales with bonuses that pursues other objectives in addition to consumer protection (62008CJ0540) or a general prohibition of combined offers made by a vendor to a consumer (62007CJ0261) (both prohibitions were considered unlawful), which are further operationalisations of cases on the main path.

4 Discussion

The results indicate good precision (representativeness) and good recall (comprehensiveness) on the topic level. The topics found on the paths produced by MPA have a large overlap with those in the literature, and vice versa. This suggests that, at least for EU consumer protection, MPA can produce substantively meaningful topics with high coverage and, in effect, find topic clusters, also when the nodes are single-topic nodes.

Interestingly, the chronological evolution of the interpretation of the Unfair Contract Terms Directive (UCTD) is commonly not addressed in legal scholarship, which is mostly structured around main topics relevant to consumer protection or around the seminal cases that have formed the core of the EU consumer law framework. MPA may, therefore, assist in informing legal practitioners and academics about developments and directions in their respective fields. One way to do so is to create casebooks for national judges to better inform them about the development of EU law. One of the attempts to supply judges with such instruments has been the Re-Jus

Table 3 Mapping literature to MPA topics, and vice versa

High-centrality threshold	#High-centrality cases	#High-centrality cases on main path	#High-centrality cases off main path
1	146	47 (32.2%)	99 (67.8%)
2	87	39 (44.8%)	48 (55.2%)
3	32	19 (59.4%)	13 (30.6%)

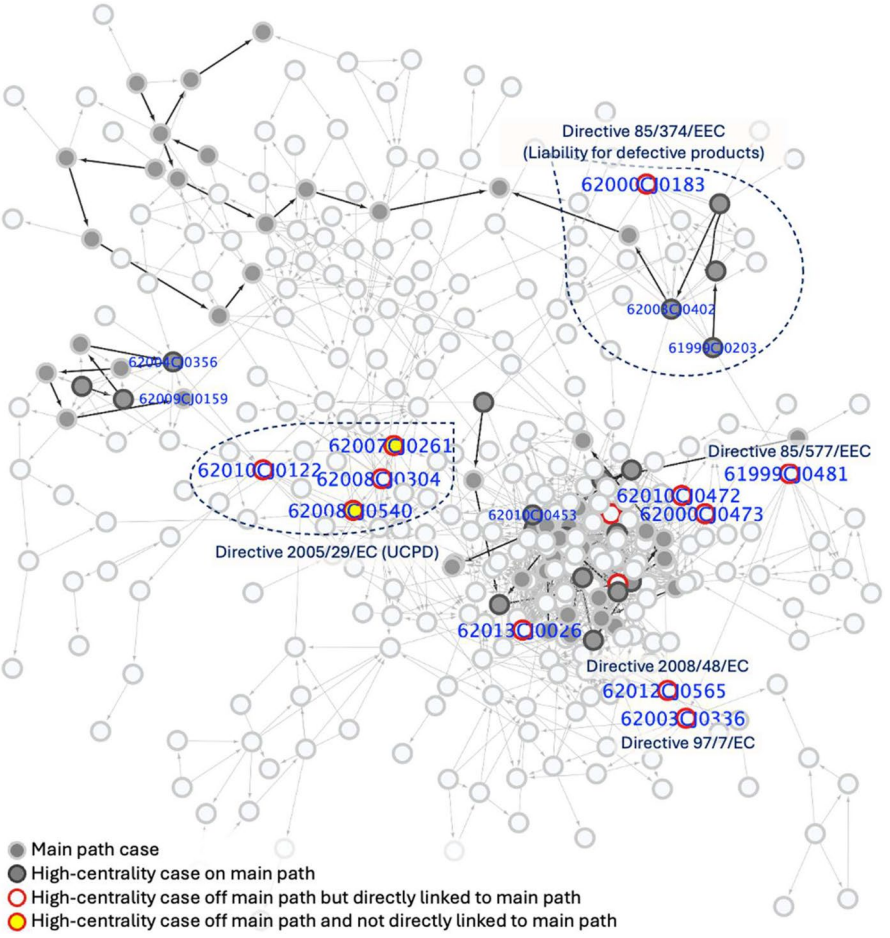


Fig. 4 Main path network cases in the largest connected component of the case citation network, where high-centrality cases are those cases top-ranked by at least three out of the five centrality metrics

Casebook (Effective Justice in Consumer Protection),[10] which aimed to summarise the main CJEU approaches to apply the principles of effectiveness and proportionality in line with Art 47 of the Charter of Fundamental Rights on matters concerning consumer protection. The research group included a large number of participants, and the time required to complete the project could be substantially reduced through the creation of such paths, and it could significantly enlarge the scope of the covered material as manual work does not physically allow for such a nuanced analysis that the MPA-created paths suggest.

[10] Re-Jus Casebook – Effective Justice in Consumer Protection (2018), published in the framework of the project: Roadmap to European Effective Justice (Re-Jus): Judicial Training Ensuring Effective Redress to Fundamental Rights Violations.

The time spent on interpreting the paths was substantive and limited, depending on one's perspective. MPA significantly reduces the time spent on analysing the evolution of a legal topic compared to having to study numerous books, journal articles, and case law on the subject. Nevertheless, the interpretation took considerable time – hours – even by a subject matter expert. Future research could focus on further reducing the process of interpreting the paths produced by MPA.

Although the results are promising, our research also has limitations. Firstly, our research focused on EU consumer protection, not other topics. The quality of the results may be different for different topics. Since this study is the first in its kind, we were initially mostly interested in whether the approach has any promise. Secondly, we used the operative parts of the judgments to determine their topics and how they relate to previous judgments. Operative parts are dense and significantly shorter than the full judgments. If the operative parts are unavailable, interpreting the paths takes substantially longer. Thirdly, the interpretations were made by a subject matter expert. MPA becomes less appealing if such an expert is not available to do the interpretation. Future research may need to examine how to automate the process, both regarding providing summaries (as an alternative to the operative parts) and the automated interpretation of the paths. Fourthly, the analysis identified instances in which cases on the path could have been merged or that did not constitute a significant legal advancement.

Moreover, this study finds that the paths, to an important extent, include nodes with high centrality scores and a low(er) proportion of nodes with a low(er) score. However, not all high-centrality cases are on the path. Off-path high-centrality cases have two types: (1) those that substantively connect to main path cases, sharing common themes with main path cases, and (2) those that do not substantively connect to main path cases, sit in different regions of the case network, and thus may represent additional themes. This shows that MPA and centrality measure different phenomena (doctrinal shifts and central precedents) and can be complementary. Testing how judgments with high centrality scores can be included on the path is a direction for future research. MPA identifies the most frequently traversed citation pathways but can overlook judgments, including those with influence. Incorporating central cases into or alongside the main path helps integrate influential "off-path" judgments and help reveal how landmark decisions intersect with the primary line of reasoning identified by MPA. Similarly, relying solely on the main path may preserve the internal coherence of citation flow but risk underrepresenting judgments that shape the doctrine's overall direction. The comprehensiveness and representativeness of MPA results may be improved by, for example, first identifying cases as starting points based on a certain centrality measure (e.g., betweenness) or using the measure to adjust citation weights to ensure that MPA does not miss them.

Finally, the current research has tested only the classical topology-only search path-counting method, overlooking the topics of cases and the semantic relations between cases. Regarding case semantics, several prior studies exist. For example, Liu et al. (2014) pioneered the use of human-annotated relevancy between citing and cited cases to weight citation links and found that the average importance of the retrieved legal opinions (i.e., main path cases) was improved (i.e., obtaining better representativeness). Huang et al. (2022) measured semantic similarity by the Jaccard

coefficient between cases' key-phrase sets and argued that "articles of low relatedness on a topic can be removed from the main paths", which leads to more coherence main paths. Corroborating this, Chen et al. (2022) claimed a dramatic improvement in the topical coherence of the documents along the main path by incorporating semantic relevance. A unique study is Jiang and Liu (2023), where citation functions were used to construct multiple networks for MPA, resulting in complementary views and improved representativeness and comprehensiveness. Work on citation functions (Sadeghian et al. 2018) or semantic relations (Sulis et al. 2022) also exists in the case law. Therefore, they are promising directions for our future work.

5 Conclusion

This study explored whether MPA can assist in tracking the evolution of case law within legal citation networks. We applied MPA to consumer protection cases decided by the CJEU and identified thematic areas that showed considerable overlap with the topics emerging in the literature. MPA significantly reduced the effort required to analyse the evolution of case law compared to traditional methods, yet interpretation remained time-intensive and required subject matter experts. Future research may explore ways to reduce the efforts needed to interpret MPA paths and test whether MPA produces acceptable results in legal domains beyond consumer protection. Despite its limitations, this study suggests MPA holds promise for legal scholars and practitioners and warrants further exploration.

Appendix 1

Illustration: the DNA network

We use the DNA network compiled in Garfield et al. (1964) to illustrate how different MPA methods work and differ on a real-world citation network. The network contains 62 citation links among 40 nodes (publications about gene and the discovery of DNA). The first application of main path analysis concerned an analysis of the DNA network to reveal how the early research on genetic technologies evolved into the groundbreaking discovery of the double-helix structure of DNA, which later won the Nobel Prize in 1962 (Hummon and Doreian 1989). Hummon and Doreian present a detailed analysis of the development of the DNA theory by extracting and checking the main path (of publications) against the list of the important DNA works manually compiled by Asimov's historical account (Asimov 1963). Hummon and Doreian have shown that the nodes on the main path do indicate the significant events in the history of DNA works. In the presentation below, the JMPA package (Java package for MPA) (Jiang et al. 2020) was used to extract the main path(s) of all the methods introduced above (Fig. 5).

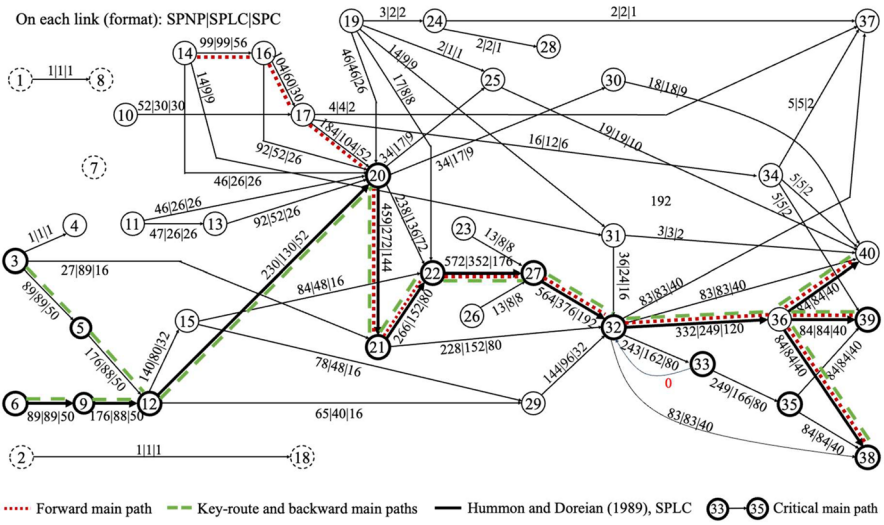


Fig. 5 The DNA theory network reproduced and adapted from Hummon and Doreian (1989)

Traversal count calculation A visual inspection immediately reveals that the SPX (X is either "C", "LC" or "NP") counts of the links in the lower part of the network should be higher because the lower part is more densely connected than the other part(s). Especially, the link $20 \rightarrow 21$ is reachable from many sources (node 3, 6, 11, 14, 19) and to many sinks (37, 38, 39, 40). Consequently, $20 \rightarrow 21$ is expected to have a high weight. Similarly, link $22 \rightarrow 27$ tends to have a high weight, and $27 \rightarrow 32$ may have even higher SPX counts because all paths through $22 \rightarrow 27$ must pass $27 \rightarrow 32$. Therefore, it is expected that the main path is likely in the lower part of the network and it may go through links $20 \rightarrow 21 \rightarrow 22 \rightarrow 27 \rightarrow 32$. The SPX counts are recorded next to the links in Fig. 5

Main path extraction Figure 5 also shows that different path exploration methods may result in slightly different main paths. The key-route, forward, and backward main paths are represented in bold black solid lines, bold red dotted lines, and bold dashed lines respectively. In this network, all three citation weighting schemes result in the same main path for each path exploration method. The critical main path is represented by bold solid circles. It is $\langle 3 \rightarrow 5|6 \rightarrow 9\rangle \rightarrow 12 \rightarrow 20 \rightarrow 21 \rightarrow 22 \rightarrow 27 \rightarrow 32 \rightarrow 33 \rightarrow 35 \rightarrow \langle 38|39\rangle$, where $\langle x \rightarrow \cdots \rightarrow y|x\prime \rightarrow \cdots \rightarrow y\prime\rangle$ means two branches $x \rightarrow \cdots \rightarrow y$ and $x\prime \rightarrow \cdots \rightarrow y\prime$ that have the same SPX weight, while $\langle x|y\rangle$ means two nodes that have the same weight on their links from the common tail node (or to the common head node). Four critical paths exist with the same weight in this network, which start from either link $3 \rightarrow 5$ or $6 \rightarrow 9$ and end at either node 38 or 39

The forward main path starts from link $14 \rightarrow 16$, which receives the highest weight among all out-going links of the six sources. There is some additional complexity in finding the backward main path(s). There are five links ending at sinks

38, 39, and 40 that share the same highest weight, so all the five links $35 \rightarrow 38$, $35 \rightarrow 39$, $36 \rightarrow 38$, $36 \rightarrow 39$, and $36 \rightarrow 40$ will be selected for path exploration in the first iteration. The exploration in the second iteration follows node 36 because its incoming link $32 \rightarrow 36$ has a higher weight than those of node 35. The partial path towards node 35 ends prematurely and is deleted. Thus, the backward main path is $\langle 3 \rightarrow 5 | 6 \rightarrow 9 \rangle \rightarrow 12 \rightarrow 20 \rightarrow 21 \rightarrow 22 \rightarrow 27 \rightarrow 32 \rightarrow 36 \rightarrow \langle 38 | 39 | 40 \rangle$, where the underlined part differs from the global main path. The key-route main path(s) based on all three SPX methods are the same as the backward main path, although the key-route based on SPNP is $22 \rightarrow 27$ but SPLC and SPC result in $27 \rightarrow 32$ as the key-route

According to SPNP and SPC, the main path extracted using Hummon and Doreian's approach also happens to be the same as the backward and key-route main path. Table 4 lists the candidate main paths starting from the six sources: 3, 6, 10, 11, 14, and 19. The final main path(s) is $\langle 3 \rightarrow 5 | 6 \rightarrow 9 \rangle \rightarrow 12 \rightarrow 20 \rightarrow 21 \rightarrow 22 \rightarrow 27 \rightarrow 32 \rightarrow 33 \rightarrow 35 \rightarrow \langle 38 | 39 \rangle$. The only complexity occurs at node 3 with SPLC, where two out-going links, $3 \rightarrow 5$ and $3 \rightarrow 21$, have the same weight. In the second iteration of priority-first search, the link $21 \rightarrow 22$, which has a higher weight than $5 \rightarrow 12$, will be followed, so the candidate path becomes $3 \rightarrow 21 \rightarrow 22 \rightarrow 27 \rightarrow 32 \rightarrow 33 \rightarrow 35 \rightarrow \langle 39 \rangle$. In this case, node 6 is selected as the starting point, and the main path under SPLC is $6 \rightarrow 9 \rightarrow 12 \rightarrow 20 \rightarrow 21 \rightarrow 22 \rightarrow 27 \rightarrow 32 \rightarrow 33 \rightarrow 35 \rightarrow \langle 38 | 39 \rangle$

DNA events on the main path Based on several past reviews, main path analysis has been proved capable of identifying the mainstream of major genetic discoveries.

Table 4 Network of main paths of the DNA network using Hummon and Doreian's approach (1989). SPNP and SPC always search the same paths, as listed inthe upper half of the table, and generate the same main paths according to Hummon and Doreian's approach, but SPLC will search different paths andgenerate different main paths, as listed in the lower half of the table. * $\langle x \mid y \rangle$ means two branches to (or from) node x and y respectively of the same SPX weight

SPX	Origin	Candidate main paths (nodes and links)	In final main path?
	3	$3 \rightarrow 5 \rightarrow 12 \rightarrow 20 \rightarrow 21 \rightarrow 22 \rightarrow 27 \rightarrow 32 \rightarrow 33 \rightarrow 35 \rightarrow \langle 38 \mid 39 \rangle^*$	$\surd$
	6	$6 \rightarrow 9 \rightarrow 12 \rightarrow 20 \rightarrow 21 \rightarrow 22 \rightarrow 27 \rightarrow 32 \rightarrow 33 \rightarrow 35 \rightarrow \langle 38 \mid 39 \rangle$	$\surd$
SPNP	10	$10 \rightarrow 17 \rightarrow 34 \rightarrow \langle 37 \mid 39 \rangle$	$\times$
SPC	11	$11 \rightarrow 20 \rightarrow 21 \rightarrow 22 \rightarrow 27 \rightarrow 32 \rightarrow 33 \rightarrow 35 \rightarrow \langle 38 \mid 39 \rangle$	$\times$
	14	$14 \rightarrow 16 \rightarrow 17 \rightarrow 34 \rightarrow \langle 37 \mid 39 \rangle$	$\times$
	19	$19 \rightarrow 20 \rightarrow 21 \rightarrow 22 \rightarrow 27 \rightarrow 32 \rightarrow 33 \rightarrow 35 \rightarrow \langle 38 \mid 39 \rangle$	$\times$
	3	$3 \rightarrow 21 \rightarrow 22 \rightarrow 27 \rightarrow 32 \rightarrow 33 \rightarrow 35 \rightarrow \langle 38 \mid 39 \rangle$	$\times$
	6	$6 \rightarrow 9 \rightarrow 12 \rightarrow 20 \rightarrow 21 \rightarrow 22 \rightarrow 27 \rightarrow 32 \rightarrow 33 \rightarrow 35 \rightarrow \langle 38 \mid 39 \rangle$	$\surd$
SPLC	10	$10 \rightarrow 17 \rightarrow 34 \rightarrow \langle 37 \mid 39 \rangle$	$\times$
	11	$11 \rightarrow 20 \rightarrow 21 \rightarrow 22 \rightarrow 27 \rightarrow 32 \rightarrow 33 \rightarrow 35 \rightarrow \langle 38 \mid 39 \rangle$	$\times$
	14	$14 \rightarrow 16 \rightarrow 17 \rightarrow 34 \rightarrow \langle 37 \mid 39 \rangle$	$\times$
	19	$19 \rightarrow 20 \rightarrow 21 \rightarrow 22 \rightarrow 27 \rightarrow 32 \rightarrow 33 \rightarrow 35 \rightarrow \langle 38 \mid 39 \rangle$	$\times$

For example, Miescher discovered DNA in 1969 (Event 3) (see Table {7} for details of all main path events). Asimov (1963) identified Avery et al.'s important discovery that "*DNA carries genetic information that could transform bacteria strain into another from which the DNA was extracted*". It is well-known that James Dewey Watson and Francis Harry Compton Crick (Event 27) discovered the double helix structure of nucleic acids and were jointly awarded the Nobel Prize in Physiology or Medicine 1962,[11] together with Wilkins (Event 26), for "its significance for information transfer in living material". Severo Ochoa (Event 32) and Arthur Kornberg (Event 33) were jointly awarded the Nobel Prize in Physiology or Medicine 1959[12] for "their discovery of the mechanisms in the biological synthesis of ribonucleic acid and deoxyribonucleic acid". All these events sit on the critical, key-route and backward main paths using either SPNP or SPLC. In 1964, Institute of Scientific Information (ISI) published the Science Citation Index. In the same year, the ISI founder Eugene Garfield and his colleagues analysed the history of science and identified additional important events using citation weights (Garfield et al. 1964): Event 20 (discovered DNA carried genetic information that can change a strain into another), 22 (discovered different nucleotides in the chain are in random order), 21 (discovered Purines and pyrimidines present in unequal quantities within nucleic acids), 36 (manufactured Messenger RNA in test tube (from DNA, nucleotides, enzymes), and 40 (Ultimately verified triplet code, using method of 32). All these events are on all types of main paths

Efficient traversal count calculation

Batagelj (2003) proposed a unified efficient algorithm for calculating traversal counts. The conceptual idea of the algorithm is illustrated in Fig. 6. Given a link $u \to v$, $SPX(u \to v)$ can be determined by the number of search paths from all origins to node u, denoted by $N^-(u)$, and the number of search paths from node v to all destinations, denoted by $N^+(v)$. More specifically, we have $SPX\,(u \to v) = N^-(u) \times N^+(v)$. Now the question becomes, for each node *u*, how to calculate $N^-(u)$ and $N^+(v)$. According to Fig. 6, the search paths to node *u* must go through one of *u*'s incoming links, so obviously $N^-(u)$ is equal to the total number of search paths passing the incoming links of *u*. Mathematically, $N^-\,(u) = \sum_{\forall p:p\to u} N^-\,(p)$, where $p \to u$ means p is a direct ancestor of u. Similarly, we can derive $N^+\,(v) = \sum_{\forall q:v\to q} N^+\,(q)$, where $v \to q$ means q is a direct descendant of v

Batagelj put forward an interesting method to unify the approach for SPX computation (X is either "NP", "LC" or "C"), or more precisely, the computation of the N^- and N^+ values for all the nodes. It is done by *standardising* a citation network. The idea is illustrated in Fig. 7. Firstly, a pseudo-source s^* and a pseudo-sink t^* are added to the network. Then, additional links are added to connect the pseudo-source and pseudo-sink to other nodes, as summarised in Table 5 and visualised in Fig. 7. For SPC, s^* is linked to all sources and all sinks are linked to t^*. For calculating the

[11] https://www.nobelprize.org/prizes/medicine/1962/summary/.

[12] https://www.nobelprize.org/prizes/medicine/1959/summary/.

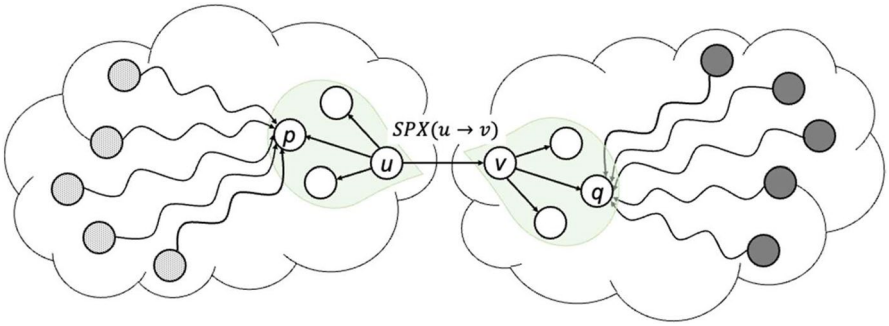


Fig. 6 Schematic network for illustrating the calculation of traversal counts for each link

Table 5 Summary of three traversal count-based citation weighting schemes

Method	Definition (given link $u \rightarrow v$)	Origins	Destinations	Standardisation
SPC	The number of search paths between all sources and sinks	Sources	Sinks	Connect pseudo-source s^* to all sources and all sinks to pseudo-sink t^*
SPLC	The number of search paths from all ancestors of u to all sinks	All nodes	Sinks	Connect pseudo-source s^* to all nodes and all sinks to pseudo-sink t^*
SPNP	The number of search paths from all ancestors of u to all descendants of v	All nodes	All nodes	Connect pseudo-source s^* to all nodes and all nodes to pseudo-sink t^*

SPLC of link $u \rightarrow v$, additional links are added between the pseudo-source s^* and every ancestor of u. In practice, for the purpose of calculating the SPLC values for all the links, s^* will be connected to all the nodes in the original network. Compared to SPLC, for the calculation of SPNP, additional links are added to connect all the nodes to the pseudo-sink t^*. Note that, a citation network is first *reversed* to reflect the direct of scientific influence.

On a standardised network, SPX of link $u \rightarrow v$ is calculated as follows:

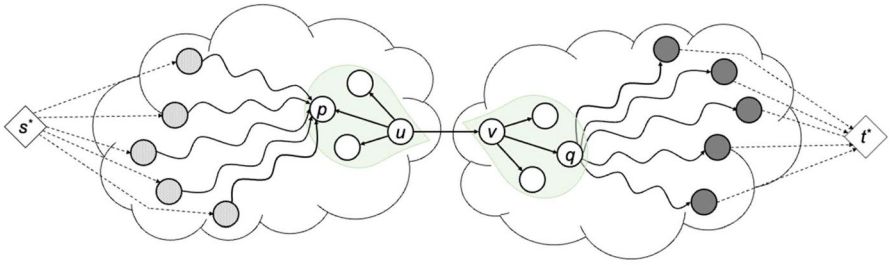
$$SPX\left(u \rightarrow v\right) = N^{-}\left(u\right) \times N^{+}\left(v\right), \tag{A1}$$

where

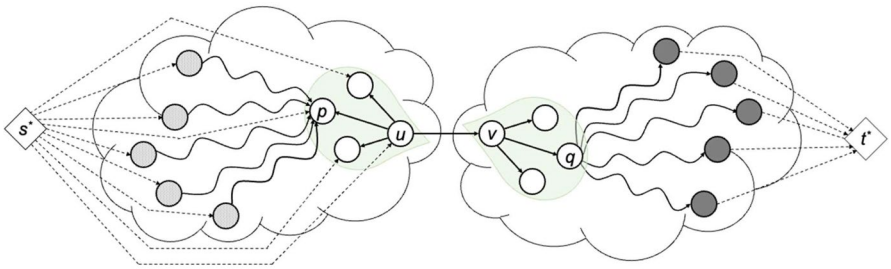
$$N^{-}\left(u\right) = \begin{cases} \sum_{\forall p:p \rightarrow u} N^{-}\left(p\right) & \text{if} u \text{is not} s^* \\ 1 & \text{otherwise} \end{cases}, \tag{A2}$$

$$N^{+}\left(v\right) = \begin{cases} \sum_{\forall q:v \rightarrow q} N^{+}\left(q\right) & \text{if} v \text{is not} t^* \\ 1 & \text{otherwise} \end{cases}. \tag{A3}$$

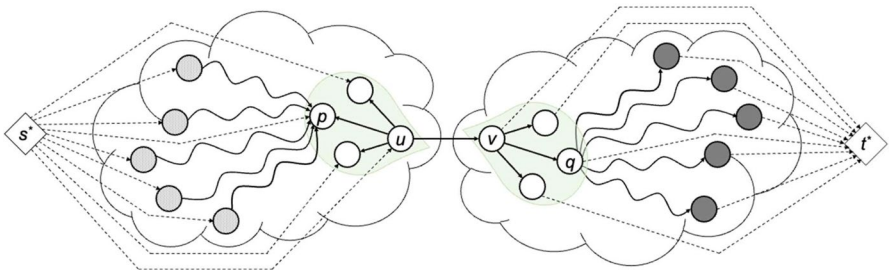
Table 6 illustrates the N^{-} and N^{+} values of each node for the calculation of traversal count-based citation weights for the network in Fig. 5. According to Fig. 5, $SPNP\left(P_7 \rightarrow P_{10}\right) = 6 = N^{-}\left(uP_7\right) \times N^{+}\left(P_{10}\right)$. Thanks to the recursive nature



(a) Standardisation for Search Path Count (SPC)



(b) Standardisation for Search Path Link Count (SPLC)



(c) Standardisation for Search Path Node Pair (SPNP)

Fig. 7 Citation network standardisation for the purpose of unifying the method for traversal count calculation

of the definitions of N^- and N^+, an efficient algorithm exists. The main idea is to sort the nodes in such a way that when calculating $N^-(u)$, the N^- of all the tail nodes p of the links ending at u have already been computed. Similarly, when calculating $N^+(v)$, the N^+ values of all the head nodes q of the out-going links from v are all ready. Algorithmically speaking, this is done by a topological sort of the nodes of the citation network, which is possible because a citation network is often acyclic. According to Equation (A2), N^- and N^+ are computed iteratively for each node following the *topological order* and *reverse topological order* of nodes, respectively. Specifically, N^- computation starts by initialising that of the pseudo-source s^* to 1, followed by the oldest nodes that are only linked from s^*, followed by the nodes that are only linked from these nodes, and so on, until the pseudo-sink t^*. On the contrary, the calculation of N^+ starts from the pseudo-sink t^* by initialising $N^+(t^*) = 1$, followed by the newest nodes that are only linked to the pseudo-sink, followed by the nodes that are only linked to these nodes, and so on, until the pseudo-source s^*

Table 6 The calculation of traversal counts in Fig. 1

	$N^-(u)$			$N^+(v)$		
	SPC	SPLC	SPNP	SPC	SPLC	SPNP
P_1	1	1	1	2	2	7
P_2	1	1	1	8	8	22
P_3	2	3	3	2	2	7
P_4	1	2	2	5	5	12
P_5	1	2	2	1	1	3
P_6	2	4	4	1	1	4
P_7	1	3	3	3	3	7
P_8	2	5	5	1	1	3
P_9	2	6	6	2	2	4
P_{10}	5	12	12	1	1	2
P_{11}	3	9	9	1	1	2
P_{12}	5	6	6	1	1	1
P_{13}	2	12	12	1	1	1
P_{14}	3	9	9	1	1	1

Details of the main path DNA events

Table 7 Events on the main path of the DNA network

Event	Author(s)	Year	Discovery	Nobel Prize
3	Miescher	1869	Isolation of nucleic acid	
5	Kossel	1886	Study of purine and pyrimidine content of nucleic acid	
*6**	*Fisher & Piloty*	*1891*	*Isolation and synthesization of ribose as a freely occurring sugar*	
*9**	*Levene*	*1909*	*Identified the 5 Carbone sugar ribose as a component of nucleic acid*	
12	Levene, Mori & London	1929	Discovery that certain nucleic acids contain deoxyribose (DNA)	
14	Alloway	1935	Proof that genetic material from a dead strain influences characteristics of a live strain	
16	Bawden & Pirie	1936	Discovered the virus (cf. 14) was also a nucleoprotein	
17	Casperson & Schultz	1938	RNA concentration is highest in cells where the rate of protein synthesis is highest	
20	Avery, MacLeod & McCarty	1944	Discovered DNA carried genetic information that can change a strain into another	
21	Chargaff	1947	Purines and pyrimidines present in unequal quantities within nucleic acids	
22	Chargaff	1950	Different nucleotides in the chain are in random order	
27	Watson & Crick	1953	Constructed model of spatial molecular configuration of DNA (via method of Event 26**)	Laureates (w/Wilkins**)
32	Ochoa	1955	Isolated a bacterial enzyme producing polynucleotide strands of RNA	Laureates (w/Homberg***)
36*	Hurwitz	1961	Manufactured Messenger RNA in test tube (from DNA, nucleotides, enzymes)	
*38**	*Norvelli*	*1961*	*Extended 36* via *DNA nucleotides, ribosomes and amino acides*	
39	*Mirsky& Allbrey*	*1962*	*Messenger RNA separated from mammalian cells*	
40	Nirenberg & Maltaei	1961	Ultimate verification of triplet code (using method of 32)	

* These are main path events extracted by automated algorithms without the human intervention based on domain knowledge. Note that, the forward main path is slightly different from what Hummon and Doreian (1989) reported, but we believe their main path was finally decided base on human knowledge about the most important events of the domain, such as the discovery of DNA (node 3). Manual investigation is indeed

** Event 26 by Wilkins in 1953: Developed X-ray diffraction methods for studies of nucleic acid

*** Event 33 by Komberg in 1956: Produced synthetic polyneucleotides of RNA from an enzyme

Centrality rankings of main path cases

The table is made available as dataset at https://doi.org/10.34894/OETCQ1

Table 8 Centrality analysis of main path cases. λ is the damping factor in the PageRank algorithm

Case ID	Degree	In-deg (citations)	Out-deg (references)	Betweenness	PageRank	λ = 0.5 12 months	λ = 0.5 36 months
61980CJ0027	226	118	310	204	**33**	*80*	**59**
61980CJ0113 [–]	226	147	229	161	*100*	106	*84*
61980CJ0130	188	118	229	204	**44**	**48**	**35**
61980CJ0193 [–]	226	118	310	204	*92*	169	118
61981CJ0261	188	*89*	310	204	**51**	**42**	**47**
61984CJ0178*	135	*89*	178	*81*	*90*	**39**	**17**
61987CJ0298 [–]	188	200	127	179	226	*79*	163
61989CJ0241	*100*	*89*	*98*	**30**	**65**	*84*	**43**
61992CJ0091	188	*89*	310	204	**2**	**21**	**8**
61992CJ0285 [–]	260	284	178	204	284	318	368
61994CJ0321 [–]	156	118	178	141	138	147	146
61997CJ0087	*85*	**41**	178	**28**	**61**	**36**	*93*
61998CJ0240	**43**	**15**	229	**62**	**1**	**26**	**7**
61998CJ0312 [–]	135	118	127	141	150	408	227
61999CJ0112	*100*	**37**	310	204	**13**	**51**	**28**
61999CJ0144	156	*66*	310	204	**15**	119	**52**
61999CJ0203	156	*66*	310	204	**7**	*66*	**24**
61999CJ0478*	156	*89*	229	204	**23**	150	128
62000CJ0052	*85*	**41**	178	179	**8**	**23**	**23**
62000CJ0154*	*100*	*66*	127	*89*	**6**	*88*	*66*
62001CJ0044	*100*	**41**	229	204	**26**	**54**	**61**
62002CJ0237*	123	*66*	178	*103*	**12**	121	127
62003CJ0402	*72*	**41**	127	**19**	**32**	**22**	**12**
62004CJ0127 [–]	188	147	178	194	196	188	173
62004CJ0356	*72*	*66*	*79*	**17**	**35**	*81*	*73*
62005CJ0059 [–]	188	147	178	204	*105*	208	196
62005CJ0132	*85*	*66*	*98*	**15**	*81*	**27**	*96*
62005CJ0168	**22**	**11**	127	*81*	**3**	**12**	**21**
62005CJ0381*	123	118	*98*	*95*	**60**	142	153
62006CJ0533 [–]	156	147	127	204	112	176	209
62007CJ0446*	*100*	*66*	127	**11**	*93*	131	124
62007CJ0487	*100*	*89*	*98*	**51**	**29**	*83*	**65**
62008CJ0040	**8**	**4**	178	115	**5**	**9**	**9**

Table 8 (continued)

Case ID	Degree	In-deg (citations)	Out-deg (references)	Betweenness	PageRank	λ = 0.5 12 months	λ = 0.5 36 months
62008CJ0137	**36**	**22**	*98*	*93*	**18**	**59**	*85*
62008CJ0358 [–]	188	200	127	137	213	312	302
62008CJ0484	**22**	**15**	*79*	**32**	**19**	*73*	*77*
62009CJ0159	*100*	*89*	*98*	**29**	**62**	116	140
62010CJ0453	**5**	**5**	**49**	**18**	**11**	**33**	**22**
62010CJ0618	**1**	**2**	**49**	**31**	**9**	**7**	**3**
62011CJ0415	**3**	**3**	**61**	**58**	**16**	**16**	**4**
62012CJ0413	*72*	**41**	127	161	**55**	175	144
62012CJ0470	**17**	**20**	**49**	**44**	**43**	115	*88*
62013CJ0034	**9**	**9**	**28**	**16**	**28**	**24**	**30**
62013CJ0497	**59**	**52**	*79*	**41**	*68*	**63**	*68*
62014CJ0032	**28**	**22**	**61**	*86*	*74*	**47**	**60**
62014CJ0110*	*100*	*89*	*98*	**36**	127	158	158
62014CJ0169	**22**	**25**	**37**	**55**	**48**	118	*104*
62014CJ0377	**3**	**7**	**1**	**2**	**31**	**3**	**6**
62016CJ0147	**17**	**37**	**12**	**8**	*107*	*67*	**57**
62017CJ0118	**33**	**29**	**49**	**55**	119	*68*	**54**
62017CJ0511	**59**	147	**28**	125	250	254	207
62018CJ0125	**11**	**14**	**20**	**65**	*109*	**28**	**25**
62018CJ0260	**50**	**41**	*79*	*95*	148	*104*	*91*
62018CJ0266	**36**	**41**	**37**	*66*	137	*96*	*74*
62018CJ0419	**13**	**22**	**7**	**5**	123	**52**	**45**
62018CJ0779	**33**	*89*	**7**	**49**	224	163	145
62019CJ0084	**36**	*66*	**20**	**33**	219	144	138
62019CJ0229	**33**	*89*	**7**	*77*	217	122	116
62019CJ0519*	*85*	200	**37**	*102*	275	268	233
62019CJ0738	**59**	147	**28**	*71*	258	216	203
62019CJ0776	**11**	**41**	**1**	**20**	197	*74*	*81*
62020CJ0243	**22**	*66*	**4**	**24**	179	**13**	**51**
62020CJ0303	**50**	200	**12**	*95*	273	249	261
62020CJ0567 [–]	156	200	*98*	152	268	129	175
62021CJ0625; 62021CJ0625_INF [–]	156	284	*79*	204	284	240	272
Match top 1	√		√		√		
Total top-5 matches	2	4	3	2	4	1	2
Total top-65 matches (+*top-66*)	28	27 (+*9*)	22	27 (+*1*)	32	24 (+*2*)	27 (+*1*)
Total top-110 matches (Quantile 1)	42	46	35	38	39	37	39

Table 8 (continued)

Case ID	Degree	In-deg (citations)	Out-deg (references)	Betweenness	Pag-eRank	λ = 0.5 12 months	λ = 0.5 36 months
Total top-65 matches by voting (4/7)	15						
Total top-65 matches by voting (3/7)	33						
Total top-65 matches by voting (2/7)	45						
Total top-65 matches by voting (1/7)	53						
Total non-matches by any centrality	12						

Bold figures indicate a match of main path cases against the top 65 by the corresponding centrality measure, while **bold underlined** indicate matches against the top 5. Table cells in the *italic* font indicate a match against the top 110 (to approximate Quantile 1 as in Journal Citation Report, e.g., JCR Q1). A special case is the *italic underlined* font, which indicates the main path cases that narrowly miss the top-65 ranked list of a corresponding centrality measure. The symbol "[–]" at the end of each Case ID indicates that this case cannot be discovered by any centrality measure as the top-65. The asterisk "*" at the shoulder of a figure in a cell indicates the only centrality measure which is able to identify the corresponding case as a top-65 case. The second to fourth last row summarises the how many main path cases can be identified by voting of five centralities, treating two Rescaled PageRank instances as a single centrality. Each centrality measure's top 65 cases are used for voting because the main path network consists of 65 cases. 4/7 means 4 centrality measures out of 7 voted that a case as an important case, i.e., in the top-65 ranked list by the centrality

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s10506-026-09523-8.

Author contribution Conceptualization, XJ+GD; methodology, XJ+GD; software, XJ; investigation, XJ+MB+GD; resources, XJ+MB+GD; data curation, XJ; writing—original draft preparation, XJ+MB+GD; writing—review and editing, XJ+MB+GD; supervision, XJ+GD; funding acquisition, XJ+GD.

Funding This research was supported by The Royal Society, UK, International Exchanges grant (IES\R1\231175).

Data availability Available athttps://doi.org/10.34894/OETCQ1.

Declarations

Ethical approval and consent to participate Not applicable.

Consent for publication All authors have read and agreed to the published version of the manuscript.

Conflict of interest No conflict of interest was declared.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.

References

Agnoloni T, Pagallo U (2015) The case law of the Italian constitutional court, its power laws, and the web of scholarly opinions. Paper presented at the Proceedings of the 15th International Conference on Artificial Intelligence and Law (ICAIL'15), San Diego California, USA. https://doi.org/10.1145/2746090.2746108

Asimov I (1963) The genetic code. Signet.

Barberá-Tomás D, Jiménez-Sáez F, Castelló-Molina I (2011) Mapping the importance of the real world: the validity of connectivity analysis of patent citations networks. Res Policy 40(3):473–486. https://doi.org/10.1016/j.respol.2010.11.002

Barnard C, Peers S (2020) European Union Law. Oxford University Press

Barnard C (2022) The Substantive Law of the EU: The Four Freedoms. Oxford University Press

Batagelj V, Ferligoj A, Squazzoni F (2017) The emergence of a field: a network analysis of research on peer review. Scientometrics 113:503–532. https://doi.org/10.1007/s11192-017-2522-8

Batagelj V (2003) Efficient algorithms for citation network analysis. arXiv: /cs/0309023. https://doi.org/10.48550/arXiv.cs/0309023

Beaverstock JV (2004) 'Managing across Borders': knowledge management and expatriation in professional service legal firms. J Econ Geogr 4(2):157–179. https://doi.org/10.1093/jeg/4.2.157

Bhupatiraju S, Nomaler Ö, Triulzi G, Verspagen B (2012) Knowledge flows—Analyzing the core literature of innovation, entrepreneurship and science and technology studies. Res Policy 41(7):1205–1218. https://doi.org/10.1016/j.respol.2012.03.011

Bommarito MJII II, Katz DM (2010) A mathematical approach to the study of the United States Code. Phys A 389(19):4195–4200. https://doi.org/10.1016/j.physa.2010.05.057

Boulet R, Mazzega P, Bourcier D (2011) A network approach to the French system of legal codes—part I: analysis of a dense network. Artif Intell Law 19(4):333–355. https://doi.org/10.1007/s10506-011-9116-1

Calero-Medina C, Noyons ECM (2008) Combining mapping and citation network analysis for a better understanding of the scientific development: The case of the absorptive capacity field. J Informetr 2(4):272–279. https://doi.org/10.1016/j.joi.2008.09.005

Chen L, Xu S, Zhu L, Zhang J, Xu H, Yang G (2022) A semantic main path analysis method to identify multiple development trajectories. J Informetr 12:101281. https://doi.org/10.1016/j.joi.2022.101281

Christensen ML, Palmer Olsen H, Tarissan F (2016) Identification of case content with quantitative network analysis: an example from the ECtHR. Paper presented at the 29th International Conference on Legal Knowledge and Information Systems (JURIX'16), Nice, France. https://doi.org/10.3233/978-1-61499-726-9-53

Colicchia C, Strozzi F (2012) Supply chain risk management: a new methodology for a systematic literature review. Supply Chain Manag 17(4):403–418. https://doi.org/10.1108/13598541211246558

Cormen TH, Leiserson CE, Rivest RL (2009) Introduction to Algorithms (3ed). MA, Usa, Cambridge

Derlén M, Lindholm J (2014) Goodbye van Gend en Loos, Hello Bosman? Using Network Analysis to Measure the Importance of Individual CJEU Judgments. Eur Law J 20(5):667–687. https://doi.org/10.1111/eulj.12077

Derlén M, Lindholm J (2015) Characteristics of Precedent: The Case Law of the European Court of Justice in Three Dimensions. Ger Law J 16(5):1073–1098. https://doi.org/10.1017/S2071832200021040

van Dijck G, van Kuppevelt D (2019) Deelonderwerpen en precedenten in het kader van werkgeversaansprakelijkheid ex art. 7:658/7:611 BW: een netwerkanalyse. Ned Juristenbl 10:652–661

Filippin F (2021) Do main paths reflect technological trajectories? Applying main path analysis to the semiconductor manufacturing industry. Scientometrics 126:6443–6477. https://doi.org/10.1007/s11192-021-04023-9

Fontana R, Nuvolaric A, Verspagen B (2009) Mapping technological trajectories as patent citation networks. An application to data communication standards. Econ Innov New Tech 18(4):311–336. https://doi.org/10.1080/10438590801969073

Fowler JH, Jeon S (2008) The authority of Supreme Court precedent. Soc Networks 30(1):16–30. https://doi.org/10.1016/j.socnet.2007.05.001

Fowler JH, Johnson TR, Spriggs JF II, Jeon S, Wahlbeck PJ (2007) Network analysis and the law: measuring the legal importance of precedents at the U.S. supreme court. Polit Anal 15(3):324–346. https://doi.org/10.1093/pan/mpm011

Freeman L (1977) A set of measures of centrality based on betweenness. Sociometry 40(1):35–41. https://doi.org/10.2307/3033543

Garfield E, Sher IH, Torpie RJ (1964) The use of citation data in writing the history of science. Philadelphia, PA: Institute for Scientific Information. Availabel at https://garfield.library.upenn.edu/papers/useofcitdatawritinghistofsci.pdf. Accessed 23 Feb 2025.

Harris JK, Beatty KE, Lecy JD, Cyr JM, Shapiro R 2nd (2011) Mapping the multidisciplinary field of public health services and systems research. Am J Prev Med 41(1):105–111. https://doi.org/10.1016/j.amepre.2011.03.015

Heinz JP, Laumann EO (1982) Chicago Lawyers: The Social Structure of the Bar. Russell Sage Foundation

Ho JC, Saw EC, Lu LYY, Liu JS (2014) Technological barriers and research trends in fuel cell technologies: A citation network analysis. Technol Forecast Soc 82:66–79. https://doi.org/10.1016/j.techfore.2013.06.004

Howells G, Twigg-Flesner C, Wilhelmsson T (2017) Rethinking EU Consumer Law. Routledge. https://doi.org/10.4324/9781315164830

Huang CH, Liu JS, Ho MHC, Chou TC (2022) Towards more convergent main paths: A relevance-based approach. J Informetr 16:101317. https://doi.org/10.1016/j.joi.2022.101317

Hummon NP, Doreian P (1989) Connectivity in a citation network: the development of DNA theory. Soc Networks 11:39–63. https://doi.org/10.1016/0378-8733(89)90017-8

Jiang X, Liu J (2023) Extracting the evolutionary backbone of scientific domains: The semantic main path network analysis approach based on citation context analysis. J Assn Inf Sci Tech 74(5):546–569. https://doi.org/10.1002/asi.24748

Jiang X, Zhuge H (2019) Forward search path count as an alternative indirect citation impact indicator. J Informetr 13(4):10977. https://doi.org/10.1016/j.joi.2019.100977

Jiang X, Sun X, Yang Z, Zhuge H, Yao J (2016) Exploiting heterogeneous scientific literature networks to combat ranking bias: Evidence from the computational linguistics area. J Assoc Inf Sci Tech 67(7):1679–1702. https://doi.org/10.1002/asi.23463

Jiang X, Zhu X, Chen J (2020) Main path analysis on cyclic citation networks. J Assn Inf Sci Tech 71:578–595. https://doi.org/10.1002/asi.24258

Josipović T (2024) Scope of application of the unfair contract terms directive: Is it time for a new compromise? Eur Rev Private Law 32(3):335–386. https://doi.org/10.54648/erpl2024025

Katz DM, Bommarito MJ II (2014) Measuring the Complexity of the Law: The United States Code. Artif Intell Law 22(4):337–374. https://doi.org/10.1007/s10506-014-9160-8

Kim J, Shin J (2018) Mapping extended technological trajectories: integration of main path, derivative paths, and technology junctures. Scientometrics 116:1439–1459. https://doi.org/10.1007/s11192-018-2834-3

Knops M, Schaper M (2014) Holding Complexity: Analysing the CJEU's VAT Case Law as a Network. World Journal of VAT/GST Law 3(3):141–165. https://doi.org/10.5235/20488432.3.3.141

Koniaris M, Anagnostopoulos I, Vassiliou Y (2015) Network Analysis in the Legal Domain: A complex model for European Union legal sources. Journal of Complex Networks 6(2):243–268. https://doi.org/10.1093/comnet/cnx029

Kuan CH (2020) Regarding weight assignment algorithms of main path analysis and the conversion of arc weights to node weights. Scientometrics 124:775–782. https://doi.org/10.1007/s11192-020-03468-8

Kuan CH (2024) Integrating prior field knowledge as key documents with main path analysis utilizing key-node path search. J Informetr 18(3):101569. https://doi.org/10.1016/j.joi.2024.101569

Lazega E (2001) The Collegial Phenomenon: The Social Mechanisms of Cooperation Among Peers in a Corporate Law Partnership. Oxford University Press, Oxford, UK

Lazega E, Krackhardt D (2000) Spreading and Shifting Costs of Lateral Control Among Peers: A Structural Analysis at the Individual Level. Qual Quant 34(2):153–175. https://doi.org/10.1023/A:1004759418226
Lazega E, van Duijn M (1997) Position in Formal Structure, Personal Characteristics and Choices of Advisors in a Law Firm: A Logistic Regression Model for Dyadic Network Data. Soc Networks 19(4):375–397. https://doi.org/10.1016/S0378-8733(97)00006-3
Lettieri N, Altamura A, Faggiano A, Malandrino D (2016) A computational approach for the experimental study of EU case law: analysis and implementation. Soc Netw Anal Min 7:58. https://doi.org/10.1007/s13278-017-0477-7
Leczykiewicz D, Weatherill S (2016) The images of the consumer in EU law. In: Leczykiewicz D, Weatherill S (eds) The Images of the Consumer in EU Law: Legislation, Free Movement and Competition Law. Oxford: Hart Publishing, pp 1–20. https://doi.org/10.5040/9781474202510.ch-001
Lianos I (2014) In memoriam Keck: The reformation of the EU law on the free movement of goods. *CLES Research Paper Series* No. 5/2014.
Liu JS, Kuan CH (2016) A new approach for main path analysis: Decay in knowledge diffusion. J Assoc Inf Sci Tech 67(2):465–476. https://doi.org/10.1002/asi.23384
Liu JS, Lu LYY (2012) An integrated approach for main path analysis: Development of the Hirsch index as an example. J Am Soc Inf Sci 63:528–542. https://doi.org/10.1002/asi.21692
Liu JS, Lu LYY, Lu W-M, Lin BJY (2013a) Data envelopment analysis 1978–2010: A citation-based literature survey. Omega 41(1):3–15. https://doi.org/10.1016/j.omega.2010.12.006
Liu JS, Lu LYY, Lu W-M, Lin BJY (2013b) A survey of DEA applications. Omega 41(5):893–902. https://doi.org/10.1016/j.omega.2012.11.004
Liu JS, Chen HH, Ho MHC, Li YC (2014) Citations with different levels of relevancy: Tracing the main paths of legal opinions. J Assoc Inf Sci Tech 65(12):2479–2488. https://doi.org/10.1002/asi.23135
Liu JS, Lu LYY, Ho M (2019) A few notes on main path analysis. Scientometrics 119:379–391. https://doi.org/10.1007/s11192-019-03034-x
Lu LYY, Lin BJY, Liu JS, Yu C-Y (2012) Ethics in nanotechnology: what's being done? What's missing? J Bus Ethics 109:583–598. https://doi.org/10.1007/s10551-012-1432-1
Lu Z, Ma Y, Song L (2021) Patent citation network analysis based on improved main path analysis: mapping key technology trajectory. In: Sun X, Zhang X, Xia Z, Bertino E (eds) Advances in Artificial Intelligence and Security. ICAIS 2021. Communications in Computer and Information Science, vol 1423. Springer, Cham. https://doi.org/10.1007/978-3-030-78618-2_13
Lucio-Arias D, Leydesdorff L (2008) Main-path analysis and path-dependent transition in HistCite-based historiograms. J Am Soc Inf Sci Tech 27(1):25–45. https://doi.org/10.1002/asi.20903
Lupu Y, Voeten E (2012) Precedent in International Courts: A Network Analysis of Case Citations by the European Court of Human Rights. Brit J Polit Sci 42(2):413–439. https://doi.org/10.1017/S0007123411000433
Mak V (2012) The 'average consumer' of EU law in domestic litigation: Examples from consumer credit and investment cases. Tilburg Law School Research Paper No. 004/2012.
Mariani MS, Medo M, Zhang YC (2016) Identification of milestone papers through time-balanced network centrality. J Informetr 10(4):1207–1223. https://doi.org/10.1016/j.joi.2016.10.005
Martinelli (2012) An emerging paradigm or just another trajectory? Understanding the nature of technological changes using engineering heuristics in the telecommunications switching industry. Res Policy 41(2):414–429. https://doi.org/10.1016/j.respol.2011.10.012
Neale T (2013) Citation analysis of Canadian case law. J Open Access Law 1(1). Retrieved from https://ojs.law.cornell.edu/index.php/joal/article/view/20. Accessed 6 June 2026
Olsen HP, Küçüksu A (2017) Finding hidden patterns in ECtHR's case law: on how citation network analysis can improve our knowledge of ECtHR's article 14 practice. Int J Discrim Law 17(1):4–22. https://doi.org/10.1177/1358229117693715
Page L, Brin S, Motwani R, Winograd T (1999) The PageRank citation ranking: bringing order to the web. Retrieved from http://ilpubs.stanford.edu:8090/422. Accessed 6 June 2026
Park H, Magee CL (2017) Tracing technological development trajectories: a genetic knowledge persistence-based main path approach. PLoS ONE 12(1):e0170895. https://doi.org/10.1371/journal.pone.0170895
Parolo PDB, Pan RK, Ghosh R, Huberman BA, Kaski K, Fortunato S (2015) Attention decay in science. J Informetr 9:734–745. https://doi.org/10.1016/j.joi.2015.07.006
Qin J, Xu JJ, Hu D, Sageman M, Chen H (2005). Analyzing terrorist networks: a case study of the global Salafi Jihad network. In: Kantor P et al. (ed) Intelligence and Security Informatics. ISI 2005. Lecture Notes in Computer Science vol 3495. Springer, Berlin, Heidelberg. https://doi.org/10.1007/11427995_24

Sadeghian A, Sundaram L, Wang DZ, Hamilton WF, Branting K, Pfeifer C (2018) Automatic semantic edge labeling over legal citation graphs. Artif Intell Law 26:127–144. https://doi.org/10.1007/s10506-018-9217-1

Schaper MGH (2014) A computational legal analysis of acte clair rules of EU law in the field of direct taxes. World Tax J 6(1):77–108

Smith TA (2007) The web of law. San Diego L Rev 44:309–354. Retrieved from: https://digital.sandiego.edu/sdlr/vol44/iss2/5. Accessed 6 June 2026

Sulis E, Humphreys L, Vernero F, Amantea IA, Audrito D, Caro LD (2022) Exploiting co-occurrence networks for classification of implicit inter-relationships in legal texts. Inf Syst 106:101821. https://doi.org/10.1016/j.is.2021.101821

Tarissan F, Nollez-Goldbach R (2015) Temporal properties of legal decision networks: a case study from the international criminal court. Paper presented at the 28th International Conference on Legal Knowledge and Information Systems (JURIX'15), Braga, Portugal. https://doi.org/10.3233/978-1-61499-609-5-111

Tarissan F, Panagis Y, Sadl U (2016) Selecting the cases that defined Europe: complementary metrics for a network analysis. Paper presented at the 2016 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM'16), Davis California, USA. https://doi.org/10.1109/ASONAM.2016.7752308

Tarissan F, Nollez-Goldbach R (2016) Analysing the first case of the International Criminal Court from a network-science perspective. J Complex Netw 4(4):616–634. https://doi.org/10.1093/comnet/cnw002

Twigg-Flesner C (ed) (2016). Research handbook on EU consumer and contract law. Edward Elgar Publishing

van Dijck G, Rodrigues de Miranda B, Crombach C (2023) Centrality scores and precedent value in legal network analysis. In: Sileno G et al. (eds) JURIX: Legal Knowledge and Information Systems. Legal Knowledge and Information Systems, IOS Press, pp 247–256. https://doi.org/10.3233/FAIA230970

van Kuppevelt D, van Dijck G, Schaper M (2020) Purposes and challenges of legal network analysis on case law. In: Whalen R (ed) Computational Legal Studies: The Promise and Challenge of Data-Driven Legal Research. Cheltenham, UK / Northampton, MA, USA: Edward Elgar, pp 265–292. Retrieved from https://doi.org/10.4337/9781788977456.00017. Accessed 6 June 2026

van Opijnen M (2012) Citation Analysis and beyond: in search of indicators measuring case law importance. Frontiers in Artificial Intelligence and Applications. Volume 250: Legal Knowledge and information systems (JURIX 2012): 95-104. https://doi.org/10.3233/978-1-61499-167-0-95

Verheyen T (2018) Full harmonization, consumer protection and products liability: A fresh reading of the case law of the ECJ. Eur Rev Private Law 26(1):119–140. https://doi.org/10.54648/erpl2018006

Verspagen B (2007) Mapping technological trajectories as patent citation networks: A study on the history of fuel cell research. Adv Complex Syst 10(1):93–115. https://doi.org/10.1142/S0219525907000945

Weatherill S (2021) Consumer policy. In Craig P and de Búrca G (eds) The Evolution of EU Law (3rd edn). Oxford University Press. https://doi.org/10.1093/oso/9780192846556.003.0028

Whalen R (2016) Legal networks: the promises and challenges of legal network analysis. Mich St L Rev 2:539–565. Retrieved from http://digitalcommons.law.msu.edu/lr/vol2016/iss2/6. Accessed 6 June 2026

Winkels R, de Ruyter J, Kroese H (2011) Determining authority of Dutch case law. Front Artif Intell Appl 235:103–112. https://doi.org/10.3233/978-1-60750-981-3-103

Zhang P, Koppaka L (2007) Semantics-based legal citation network. Paper presented at the Proceedings of the 11th international conference on Artificial intelligence and law

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Authors and Affiliations

Xiaorui Jiang[1] · Maria Breskaya[2] · Gijs van Dijck[2]

✉ Gijs van Dijck
gijs.vandijck@maastrichtuniversity.nl

Xiaorui Jiang
xiaorui.jiang@sheffield.ac.uk

Maria Breskaya
m.breskaya@maastrichtuniversity.nl

[1] School of Information, Journalism and Communication, University of Sheffield, Sheffield, UK

[2] Maastricht Law & Tech Lab, Maastricht University, Maastricht, The Netherlands