Research Collection Yong Pung How School Of Law

Yong Pung How School of Law

4-2024

Legal hypergraphs

Corinna COUPETTE

Dirk HARTUNG
Singapore Management University, dirkhartung@smu.edu.sg

Daniel Martin KATZ

Citation

Legal hypergraphs

Corinna Coupette[1,2], Dirk Hartung[2,3] and
Daniel Martin Katz[2,3,4]

[1]Max Planck Institute for Informatics, Saarbrucken, Germany
[2]Center for Legal Technology and Data Science, Bucerius Law School, Hamburg, Germany
[3]CodeX, The Stanford Center for Legal Informatics, Stanford Law School, Stanford, CA, USA
[4]Illinois Tech, Chicago Kent College of Law Chicago, IL, USA

CC, 0000-0001-9151-2092; DH, 0000-0002-1916-4879

Complexity science provides a powerful framework for understanding physical, biological and social systems, and network analysis is one of its principal tools. Since many complex systems exhibit multilateral interactions that change over time, in recent years, network scientists have become increasingly interested in modelling and measuring *dynamic* networks featuring *higher-order relations*. At the same time, while network analysis has been more widely adopted to investigate the structure and evolution of law as a complex system, the utility of dynamic higher-order networks in the legal domain has remained largely unexplored. Setting out to change this, we introduce *temporal hypergraphs* as a powerful tool for studying legal network data. Temporal hypergraphs generalize static graphs by (i) allowing any number of nodes to participate in an edge and (ii) permitting nodes or edges to be added, modified or deleted. We describe models and methods to explore *legal hypergraphs* that evolve over time and elucidate their benefits through case studies on legal citation and collaboration networks that change over a period of more than 70 years. Our work demonstrates the potential of dynamic higher-order networks for studying complex legal systems, and it facilitates further advances in legal network analysis.

This article is part of the theme issue 'A complexity science approach to law and governance'.

THE ROYAL SOCIETY PUBLISHING

1. Introduction

In law, myriad entities—e.g. individuals, organizations and texts—interact in intricate ways, producing emergent phenomena, feedback loops and nonlinear change [1]. As such, legal systems are *complex systems* [2–4], and complex systems have long been successfully modelled as *networks* using the language of *graphs* [5].

While considerable progress has been made using simple graphs as models of complex networks, in recent years, the network-science community has embraced more expressive *higher-order network models* [6], allowing us, for example, to distinguish multiple types of nodes and edges (*multilayer networks* [7]) or capture change over time (*temporal networks* [8]). A comparatively recent trend is the systematic study of networks beyond pairwise interactions, allowing interactions between more than two nodes at a time [9]. These networks are naturally modelled as *hypergraphs*, which generalize graphs by allowing edges to contain any nonempty subset of nodes [10]. Acknowledging that n-ary interactions may also evolve over time, we arrive at *temporal hypergraphs* [11–13].

Our work starts from the observation that in the legal domain, interactions constantly change *and* often involve more than two entities [14,15]. For example, in the *legislature* [16,17], lawmakers constantly negotiate new laws, which once passed may repeal or amend old laws, and the composition of the lawmaking body changes with every election. Likewise, in the *judiciary*, individual judges are affiliated with courts and judicial bodies, but they may retire or move between institutions [18], and panels of judges decide cases based on prior decisions, which they often cite in their opinions [19,20]. Hence, legal systems are naturally modelled as *temporal legal hypergraphs*—but as far as we know, this option has not been explored in the literature.[1]

(a) Contributions

In this work, we make three contributions. First, we introduce *hypergraphs* and *temporal hypergraphs* as mathematical representations of legal network data. Second, we describe classic methods to study temporal hypergraphs in the legal domain. Third, we demonstrate the impact of data-modelling decisions on the output of these methods in case studies on legal information networks and legal collaboration networks, sharing two novel datasets in the process. To the best of our knowledge, we are the first to model legal network data as temporal hypergraphs and highlight the potential of hypergraph methods for understanding legal phenomena.

(b) Related work

Our work is closely related to two bodies of work: *legal network analysis*, which focuses on our *domain* of interest, and *higher-order network science*, especially the literature studying *hypergraphs*, which engages with our *models* and *methods*.

In *legal network analysis*, inspired by the seminal work of [22], scholars have studied both *information networks*, such as judicial citation networks [19,20,23–26] or legislation networks [14,16,27,28], and *social networks*, such as legislative collaboration networks [29–31] or judicial collaboration networks [32]. While most investigations have focused on individual countries [33–36], others have compared several countries [37,38], or studied networks at the European level [39–42] or the international level [43–46]. Although researchers are increasingly using *temporal graphs* in their modelling [47,48], they have yet to systematically consider higher-order interactions, as has recently been done in other domains [49].

In *higher-order network science*, much work has been done to port concepts and methods from graphs to *hypergraphs*, leading, e.g. to generalizations of walks [50,51], centralities [52,53], motifs [12,54–56], clustering [57–60] and cores [61]. Exploiting edge cardinality as a new source

---

[1]The only (partial) exception we are aware of is the work by Boulet *et al.* [21], who model sequences of multilateral-treaty ratifications as directed hypergraphs, encoding time in hyperedge directions (based on a non-standard definition of directed hypergraphs). However, they only use the incidence matrix of this hypergraph to embed their nodes (representing countries) in $\mathbb{R}^k$ for some chosen (undisclosed) k and then use methods tailored to point-cloud data to perform the rest of their analyses.

Table 1. Metadata describing the datasets used as our running examples. We state the network type, the semantics of our raw data along with their count, the temporal coverage of the data along with the number of distinct timestamps T in the dataset, and the type of network evolution.

dataset	network type	raw data	cases	coverage	T	evolution type
GFCC	citation	case texts	3618	1951–2022 (72 years)	2109	point-aggregation
ICSID	collaboration	case metadata	742	1974–2023 (50 years)	1077	interval-event

of variation in hypergraphs, researchers are increasingly adapting concepts and methods from *topology* to analyse higher-order network data [13,62,63], and they have also made some progress by leveraging (structurally more constrained) *simplicial complexes* [64–67]. Recent surveys have consolidated our knowledge of when and how to use higher-order network models or collected the mathematical and computational tools currently available for their study [9,68–71], and the software landscape for working with higher-order network data has improved dramatically over the past few years [72–74]. Nevertheless, the study of higher-order network data is still in its relative infancy, especially when compared with traditional network analysis.

(c) Structure

After describing the two datasets that serve as our running examples in §2, we discuss how these datasets can be modelled as static or temporal (hyper)graphs in §3. In §4, we then introduce different methods for investigating temporal legal hypergraphs and show how these can be used on our datasets, before concluding with a discussion in §5.

2. Data

As running examples, we consider two types of network data: judicial decisions and their citations as a classic example of information networks (*legal citation networks*), and arbitrators and their tribunals as a classic example of social networks (*legal collaboration networks*). In particular, we study citation networks derived from the official collection of Germany's Federal Constitutional Court (GFCC), and collaboration networks derived from the cases registered at the International Center for Settlement of Investment Disputes (ICSID). Notably, these networks also feature different types of temporal evolution: GFCC decisions are associated with one specific date and aggregate over time (*point-aggregation*), and ICSID cases form intervals of collaboration events, associated with a start date and an end date (*interval-event*). Providing a high-level overview of our datasets in table 1, in the following, we give more information on each of our datasets.

(a) Legal citation networks (GFCC)

We derive our legal citation network data from the texts of judicial decisions published by the Federal Constitutional Court of Germany (*the Court*) in volumes 1–160 of its official collection, BVerfGE. The Court settles questions surrounding the interpretation of the German constitution (*Grundgesetz*), and it is generally known to rely heavily on citations to its own case law to develop its arguments [75]. We source the texts of volumes 1–140 from the online appendix made available by Coupette [19] and collect the texts of the remaining volumes from the Court's official website (https://www.bundesverfassungsgericht.de/DE/Entscheidungen/Entscheidungen/Amtliche%20Sammlung%20BVerfGE.html).[2] The result is a *GFCC corpus* of 3618

---

[2]Although the Court provides an overview of the decisions included in its official collection on its website, its documentation is still incomplete (e.g. some decisions included in the collection are missing from the overview) and the text of most older decisions is not currently available from the website.
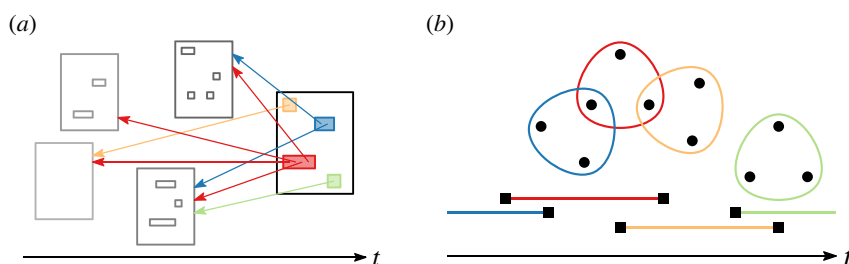
Figure 1. Schematic depiction of the phenomena captured by our datasets. With GFCC, we seek to understand the development of constitutional jurisprudence, with a particular view to how new decisions (large rectangles) create meaning by combining ideas from previous decisions (arrows, drawn here only for the most recent decision, and small rectangles). With ICSID, we would like to explore how the social structure of an elite group, arbitrators (circles) appointed to panels in international-arbitration cases (lines and rounded polygons), evolves over time. (a) GFCC, (b) ICSID.

decisions covering the Court's most prominent jurisprudence from its inception in 1951 to its latest collection-included decisions rendered in early 2022.

For each decision in the GFCC corpus, we extract the other GFCC decisions it cites using a procedure previously introduced in the literature [19]. Notably, our extraction process records the cited decisions at the level of individual *citation blocks*, i.e. uninterrupted strings of cited decisions contained in a citing decision. This allows us to distinguish between decisions cited to illustrate the same legal argument (co-citation in a citation block) and decisions cited (only) in the same decision, providing a more nuanced perspective on the relationships that citations establish between co-cited decisions. We provide a schematic depiction of the phenomenon captured by our GFCC data in figure 1a.

(b) Legal collaboration networks (ICSID)

We derive our legal collaboration network data from the metadata of cases registered at the World Bank's International Center for Settlement of Investment Disputes (*the Center*) since its establishment in 1966 until mid-June 2023. Among the signatory states of the ICSID Convention, the Center is designed to provide an impartial forum, independent of national courts, for the settlement of investment-related disputes between foreign investors and investee states. All ICSID disputes involve investors as claimants and states as respondents, and they are typically administered by panels of three arbitrators: one appointed by the claimants, one appointed by the respondents, and one jointly appointed by the parties' selected arbitrators.[3] We obtain the case metadata from the Center's website (https://icsid.worldbank.org/cases) and restrict ourselves to cases for which arbitrator names and date information are complete.[4] This leaves us with 742 cases (corresponding to 2226 tribunal-member slots) registered between 6 March 1974 and 14 June 2023, with participation by 441 unique arbitrators.

In addition to arbitrator names and case-registration dates, we collect and clean the available case metadata on arbitrators' nationalities, appointing entities, claimants and their nationalities, respondent states, subject and economic sector of the dispute, legal instruments and applicable rules involved, as well as conclusion status, conclusion date, and the date at which the tribunal

---

[3]The procedural details and arbitration realities are more complicated. For a review of the rules governing ICSID's operations, including the selection of arbitrators, see [76]. For a summary of concerns surrounding ICSID's operations, see [77].

[4]By completeness of *arbitrator names*, we mean that we can identify one arbitrator in each role. We do not currently account for changes in panel composition while a case is ongoing (identified arbitrators correspond to the *latest* entries in the ICSID database), which will be of interest to scholars specializing in international arbitration. By completeness of *date information*, we mean that (i) the case is either still pending or we know when it was concluded and (ii) the tribunal constitution date is available. For all cases fulfilling these conditions, a case registration date is also available (which usually, but not always, lies before the tribunal constitution date), but there are three cases that fulfil condition (i) for which we know only the registration date but not the tribunal constitution date. For consistency, we exclude those cases from our analysis.

was constituted. This allows us to analyse several classes of collaboration and affinity relations, including party roles, industry specialization and nationality. A schematic depiction of the phenomenon captured by our ICSID data is given in figure 1*b*.

3. Models

Having described our example datasets in §2, we now show how these datasets can be modelled as graphs of varying expressiveness. To this end, we first introduce the relevant abstract concepts, along with their associated notation and basic terminology. We then proceed to model our datasets using these concepts and provide basic statistics of the resulting concrete mathematical objects. As there are two ways to go from static graphs to temporal hypergraphs, depicted in figure 2, in our exposition, we take the route via temporal graphs for the GFCC data and the route via static hypergraphs for the ICSID data.

(a) Concepts

Static graphs. A *simple graph* $G = (V, E)$ is a tuple containing n nodes (vertices) $V = \{v_1, \ldots, v_n\}$ and m edges $E = \{e_1, \ldots, e_m\}$. We call n the *order* and m the *size* of G. If the graph is *undirected*, $e_i \in \binom{V}{2}$ for all $i \in [m]$, where for a set S and a positive integer $k \leq |S|$, $\binom{S}{k}$ denotes the set of all k-element subsets of S, and for $x \in \mathbb{N}$ with $0 \notin \mathbb{N}$, $[x] = \{i \in \mathbb{N} \mid i \leq x\}$. If the graph is *directed*, $E \subseteq (V \times V) \setminus E_{ii}$, where $E_{ii} = \{(i, i) \mid i \in V\}$. In *multi-graphs*, edges can occur multiple times, and hence, $E = (e_1, \ldots, e_m)$ is an indexed family of sets, with $e_i \in \binom{V}{2}$ (undirected graphs) resp. $e_i \in (V \times V) \setminus E_{ii}$ (directed graphs) for all $i \in [m]$. We write $u \sim v$ if $\{u, v\} \in E$ in undirected graphs, where u is said to be *adjacent* to v if $u \sim v$ and $u \to v$ if $(u, v) \in E$ in directed graphs, where u is to be *adjacent* to v if $u \to v$ or $v \to u$.

In undirected graphs, we denote the *neighbourhood* of node u as $\mathcal{N}(u) = \{v \mid u \sim v\}$ and define the *degree* of u as $\delta(u) = |\{e \in E \mid u \in e\}|$. In directed graphs, we denote the *out-neighbourhood* of node u as $\mathcal{N}^+(u) = \{v \mid u \to v\}$, the *in-neighbourhood* of u as $\mathcal{N}^-(u) = \{v \mid v \to u\}$, and the *neighbourhood* of u as $\mathcal{N}(u) = \mathcal{N}^+(u) \cup \mathcal{N}^-(u)$. Further, we define the *out-degree* of u as $\delta^+(u) = |\{(i, v) \in E \mid i = u\}|$, the *in-degree* of u as $\delta^-(u) = |\{(v, j) \in E \mid j = u\}|$, and the *degree* of u as $\delta(u) = \delta^+(u) + \delta^-(u)$. Finally, the *line graph* of G is $L(G) = (E, \{\{e, f\} \mid e \cap f \neq \emptyset\})$, i.e. each edge in G becomes a node in $L(G)$, and two nodes e and f are connected in $L(G)$ if the edges e and f intersect in G.

Static hypergraphs. Generalizing simple undirected graphs, a simple undirected *hypergraph* $H = (\mathcal{V}, \mathcal{E})$, also called a *set system*, is a tuple containing n nodes $\mathcal{V}$ and m hyperedges $\mathcal{E} \subseteq \mathcal{P}(\mathcal{V}) \setminus \emptyset$, where $\mathcal{P}(\mathcal{V}) = \{S \mid S \subseteq \mathcal{V}\}$ is the power set of $\mathcal{V}$, i.e. in contrast to ordinary edges, a hyperedge e can have any cardinality $|e| \in [n]$. In an undirected *multi-hypergraph*, $\mathcal{E} = (e_1, \ldots, e_m)$ is an indexed family of sets, with $e_i \subseteq \mathcal{V}$ for all $i \in [m]$. We define adjacency, node degrees and node neighbourhoods as detailed for simple graphs, and further denote the *edge-neighbourhood* of an edge as $\mathcal{N}(e) = \{f \in \mathcal{E} \mid e \neq f \wedge e \cap f \neq \emptyset\}$, as well as its *node-neighbourhood* as $\mathcal{N}_\mathcal{V}(e) = \{v \mid v \in \bigcup \mathcal{N}(e)\}$. Observe that while $\delta(u) = |\mathcal{N}(u)|$ in simple graphs and $\delta(u) \geq |\mathcal{N}(u)|$ in multi-graphs for all $u \in V$, these relations do not generally hold in hypergraphs. Further, as the cardinality of hyperedges is generally variable, we call a hypergraph *r-uniform* if all hyperedges have the same cardinality, i.e. $|e| = r$ for all $e \in \mathcal{E}$, such that graphs are 2-uniform hypergraphs. As the size of *hyperedge intersections* can vary as well, we can further define an *s-line graph*, which generalizes the line graph by placing edges between two nodes e and f in $L(H)$ only if $|e \cap f| \geq s$.

Lastly, there are two natural *projections* that transform a hypergraph $H = (\mathcal{V}, \mathcal{E})$ into a graph: the *bipartite projection*, also known as the *star expansion*, defines $G^\star = (V^\star, E^\star)$ with $V^\star = \mathcal{V} \dot\cup \mathcal{E}$ and $E^\star = \{\{v, e\} \mid v \in \mathcal{V}, e \in \mathcal{E}, v \in e\}$, and it is lossless if we remember the partition of $V^\star$ into the original node and edge sets. The *clique projection*, also known as the *clique expansion*, defines $G^\circ = (V^\circ, E^\circ)$ with $V^\circ = \mathcal{V}$ and $E^\circ = \{\{u, v\} \mid \exists e \in \mathcal{E} : \{u, v\} \subseteq e\}$, i.e. two nodes are adjacent in G° if and only if they are adjacent in H. It can optionally be equipped with a *weighting function*, $w : E^\circ \to \mathbb{R}$ reflecting the intensity of node-to-node associations. The simplest weighting function is $w : E^\circ \to \mathbb{N}$ with $w(e) = |\{e \in E \mid \{u, v\} \subseteq E\}|$ for each $e \in E^\circ$, i.e. an edge $\{u, v\}$ in E° is weighted by how often u and

static graph $\longrightarrow$ temporal graph

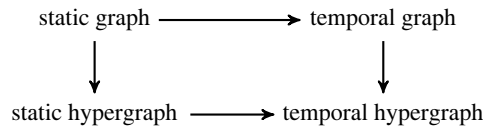static hypergraph $\longrightarrow$ temporal hypergraph

Figure 2. Routes from static graphs to temporal hypergraphs. We will go *via temporal graphs* when modelling the GFCC data, and *via static hypergraphs* when modelling the ICSID data.

v co-occur in edges from {H}. Even when equipped with a weighting function, clique expansions are generally lossy, i.e. we cannot uniquely reconstruct H from G°.

Temporal (hyper)graphs. There are multiple ways to capture information on temporal evolution in (hyper)graph data. We use the definition of a *snapshot-graph sequence* from Gauvin *et al.* [78], slightly adapted to more naturally accommodate changes to the node set. That is, we define a *temporal graph* $G_\mathcal{T} = (\mathcal{T}, \boldsymbol{\Gamma})$, where $\mathcal{T} = t_1, \ldots, t_T$ is a sequence of monotonically increasing time stamps, $\boldsymbol{\Gamma} = (\Gamma_1, \ldots, \Gamma_T)$ is a sequence of snapshot graphs $\Gamma_i = (V_i, E_i)$ for $i \in [T]$, V_i is the set of nodes present in graph snapshot i and E_i is the set (or indexed family) of edges present in graph snapshot i.[5] To define a *temporal hypergraph* $H_\mathcal{T} = (\mathcal{T}, \boldsymbol{\Xi})$, we simply allow hyperedges in $\mathcal{E}_i$.

(b) Instantiations

Legal citation networks (GFCC). A legal citation network is classically modelled as a *static directed graph* $G = (V, E)$. In a *judicial* citation network, each node $v \in V$ represents a decision, and each edge $e = (u, v) \in E$ represents a citation $u \rightarrow v$. When an edge indicates the *binary observation* that decision u cites decision v at least once, the result is a *simple* directed graph, whereas when an edge indicates an *observed citation instance*, the result is a directed *multi-graph* (or equivalently, a *count-weighted* directed graph). Observe that judicial decisions (unlike, e.g. statutory texts) cannot typically be altered or removed after they have been published, and that each decision is associated with a decision date. Hence, the graph $G = (V, E)$ used in the static model is simply the last snapshot graph Γ_T of a *temporal graph* $G_\mathcal{T} = (\mathcal{T}, \boldsymbol{\Gamma})$ whose time stamps are the ordered sequence of unique decision dates found in the data. Thus, denoting as $t(u)$ the time stamp of decision u, for a snapshot graph $\Gamma_i = (V_i, E_i)$ with $i \in [T]$, the node set is $V_i = \{u \in V \mid t(u) \leq t_i\}$, and the edge sequence is $E_i = ((u, v) \in E \mid t(u) \leq t_i \wedge t(v) \leq t_i)$. Note that the resulting snapshot-graph sequence $G_\mathcal{T}$ is a *growing* graph, i.e. we have $V_i \subseteq V_j$ and $E_i \subseteq E_j$ for i, j with $t_i \leq t_j$.

While temporal graphs retain more information than static graphs, they do not explicitly capture higher-order interactions in legal citation network data. In judicial citation networks in particular, when modelling decisions as nodes and citations as edges, we can only recover which decisions are cited together *in the same decision* by considering the out-neighbourhoods of the nodes (equivalently, we could define a hypergraph in which each citing decision is a hyperedge and each cited decision is a node). However, we lose valuable information regarding which decisions get cited together *in the same legal context*. To capture this information, we can model each decision as a sequence of hyperedges, i.e. $s(u) = (s_1, \ldots, s_c)$, where $s_i \subseteq V$ contains the decisions that are cited in context i, and c is the number of distinct contexts in the decision. This requires us to define a *context model* to describe when we deem decisions to be cited in the same context. The practical zeroth-order approach we adopt in this paper is to consider each sequence of consecutive citations (i.e. citations not interrupted by prose) as a separate hyperedge.

Independent of the context model, the result is a *temporal hypergraph* $H_\mathcal{T} = (\mathcal{T}, \boldsymbol{\Xi})$ with the same node sets as the temporal graph defined above, and hyperedge sequences corresponding to the contexts that occurred on or before t_i, i.e. $\mathcal{E}_i = (*s(u) \mid t(u) \leq t_i)$, where we employ $*$ to denote sequence unpacking in a slight abuse of notation. In this framework, individual decisions are represented not only as nodes but also as *(sub-)hypergraphs*, as illustrated in figure 3. Assuming

---

[5]The difference to the definition from Gauvin *et al.* [78] is that we allow the node sets to differ between snapshots.
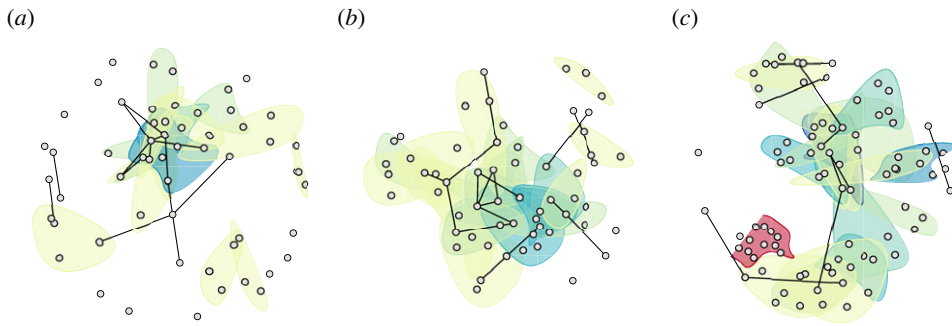
Figure 3. Three prominent decisions from the GFCC corpus depicted as hypergraphs (labeled volume, page). The decisions concern data privacy (*a*), European integration (*b*) and religious freedom (*c*), respectively. In each panel, nodes represent decisions cited at least once by the visualized decision, and (hyper)edges indicate unique citation blocks. Hyperedge colours progress, by increasing cardinality, from light yellow to dark blue and then red, with binary edges (indicating that exactly two decisions share a citation block) drawn in black. To limit visual clutter, we encode neither the sequential order of hyperedges nor their multiplicities. (*a*) 125, 260—Retention of Data, (*b*) 135, 317—ESM Treaty, (*c*) 153, 1—Headscarf III.

Table 2. Basic statistics of our data modelled as static graphs ($\tau_E = *g$) or static hypergraphs ($\tau_E = *h$), allowing binary edges only ($\tau_E = b*$) or permitting multi-edges ($\tau_E = m*$). For each (hyper)graph, we state its edge semantics, its number of edges m, as well as its mean degree $\overline{\delta}$ and its median degree δ_M. For the directed GFCC graphs, we report the average out-degree (equal to the average in-degree) under $\overline{\delta}$, and separate values for the median out- and in-degrees under δ_M.

(a) GFCC					(b) ICSID				
edges	τ_E	m	$\overline{\delta}$	δ_M	edges	τ_E	m	$\overline{\delta}$	δ_M
citations	bg	39 428	10.9	$6^+|7^-$	case sharing	bg	1 869	8.5	4
	mg	77 284	21.4	$9^+|10^-$		mg	2 226	10.1	4
citation blocks	bh	13 541	10.2	6	shared cases	bh	722	4.9	2
	mh	46 257	21.4	10		mh	742	5.0	2

that $t(v) \leq t(u)$ for all $v \in \bigcup s(u)$ (i.e. citations do not point into the future),[6] just like $G_{\mathcal{T}}$ is a growing graph, $H_{\mathcal{T}}$ is a growing hypergraph.

In table 2a, we provide the basic statistics of the GFCC data when modelled as static (multi-)graphs or (multi-)hypergraphs, supplementing the static description with a temporal perspective in figure 4a. We see that the statistics differ widely depending on whether we allow or disallow multi-edges. This indicates considerable redundancy in our data, which can be leveraged to distinguish different intensities of binary or n-ary relationships between the judicial decisions in our corpus. Furthermore, note that under the same edge model, the degree statistics in the graph and the hypergraph are very similar even though the edge counts differ widely—which is a consequence of our definitions. We also see that our median degrees are consistently much smaller than our mean degrees under the same (hyper)graph model. This suggests that our degree distributions are skewed, which can be confirmed by an inspection of the empirical complementary cumulative distribution functions (CCDFs) of node degrees in $\Gamma_T \equiv G$. In figure 4b, we additionally show the empirical CCDFs of edge sizes at different points in time in our temporal multi-hypergraph model $H_{\mathcal{T}}$. Here, we witness a noticeable shift to the right,

---

[6]Our GFCC dataset contains 38 instances of citations forward in time and 55 instances of citations where source decision and target decision were rendered on the same date. While forward citations mostly result from editorial choices concerning the collection, lateral citations can also follow from scheduling decisions within the Court.
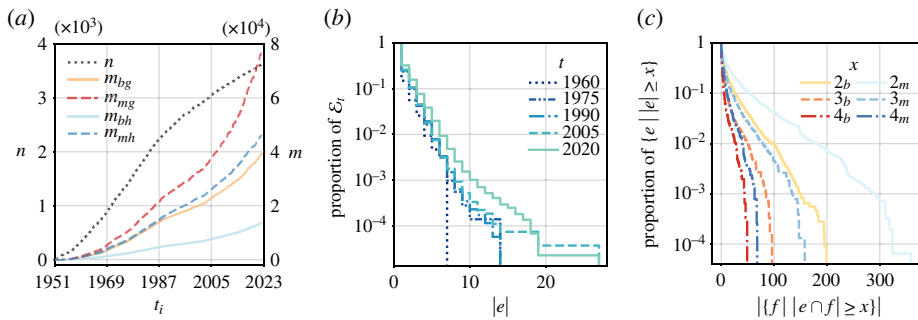
Figure 4. Descriptive statistics based on different network models of the GFCC data. We show (i) the evolution of n and m in the temporal graph and hypergraph models, (ii) empirical CCDFs of edge sizes at different time points in the temporal multi-hypergraph model (*b*), and (iii) empirical CCDFs of thresholded edge-neighbourhood sizes for the last hypergraph in the temporal hypergraph model with binary edges (*b*) or multi-edges (*m*) (*c*). (*a*) Evolution of $G_{\mathcal{T}}$ and $H_{\mathcal{T}}$, (*b*) edge sizes in Ξ_t, (*c*) neighbourhood sizes in Ξ_T.

indicating that citation blocks grow larger over time. Finally, in figure 4*c*, we show the empirical CCDFs of thresholded edge-neighbourhood sizes for hyperedges in $\Xi_T \equiv H$, modelled as a simple hypergraph (*b*) or a multi-hypergraph (*m*). That is, we ask: for a given edge e with $|e| = x$, how many other edges share at least x nodes with e? We observe that nontrivial nonempty hyperedge intersections (i.e. $|e \cap f| > 1$) exist in our data, but also that a large fraction of these intersections is contributed by multi-hyperedges (i.e. citation blocks that occur more than once).

Legal collaboration networks (ICSID). A legal collaboration network is classically modelled as a *static undirected graph* $G = (V, E)$. In an *arbitrator* collaboration network, each node $v \in V$ represents an arbitrator sitting on at least one tribunal, and each edge $e = \{u, v\} \in E$ indicates that arbitrators u and v share a tribunal. The result is an undirected *simple* graph when an edge models the *binary observation* that u and v sat on the same tribunal at least once, and an undirected *multi-graph* (or equivalently, an undirected *count-weighted* graph) when an edge models an *observed collaboration instance*. In this modelling framework, every ICSID case induces three (potential) edges, one between each pair of arbitrators that is part of the case tribunal. That is, the static graph G is a *clique expansion* of the static 3-uniform hypergraph $H = (V, E)$ with the same node set as G, and the edge sequence containing one hyperedge for each constituted tribunal of a case.

While moving from static graphs to static hypergraphs allows us to investigate higher-order interactions in ICSID arbitration tribunals, we have yet to account for temporal evolution. To this end, observe that unlike the citations in the GFCC data, which enter the citation network and then stay there, collaborations on ICSID tribunals only exist while the case is active. We proxy the activity timespan of a case by the closed interval between its tribunal constitution date and its conclusion date. Therefore, when we model the ICSID data as a temporal hypergraph $H_{\mathcal{T}} = (\mathcal{T}, \Xi)$, our time stamps are the ordered sequence of unique dates on which a tribunal is constituted *or* a case is closed. Denoting as $t_{\min}(e)$ the tribunal constitution date of a case e, and as $t_{\max}(e)$ its conclusion date, for a snapshot hypergraph $\Xi_i = (\mathcal{V}_i, \mathcal{E}_i)$ with $i \in [T]$, our edge sequence is $\mathcal{E}_i = (e \mid t_{\min}(e) \leq t_i \wedge t_{\max}(e) \geq t_i)$, and our node set is $\mathcal{V}_i = \bigcup_{e \in \mathcal{E}_i} e$. This construction implies that $\delta(v) > 0$ for all $v \in \mathcal{V}_i$. In contrast to our GFCC data, due to the interval nature of ICSID collaborations, the temporal hypergraph derived from our ICSID data is *not* growing.

In table 2b, we provide the basic statistics of the ICSID data when modelled as static (multi-) graphs or (multi-)hypergraphs. We see that in the representations allowing multi-edges, the graph representation has exactly three times as many edges as the hypergraph representation, and its average and median degrees are twice as high. This is a consequence of the fact that H is 3-uniform and G is a clique expansion of H. Furthermore, the differences between the edge counts of our binary and multi-edge models show that while there is a considerable fraction of
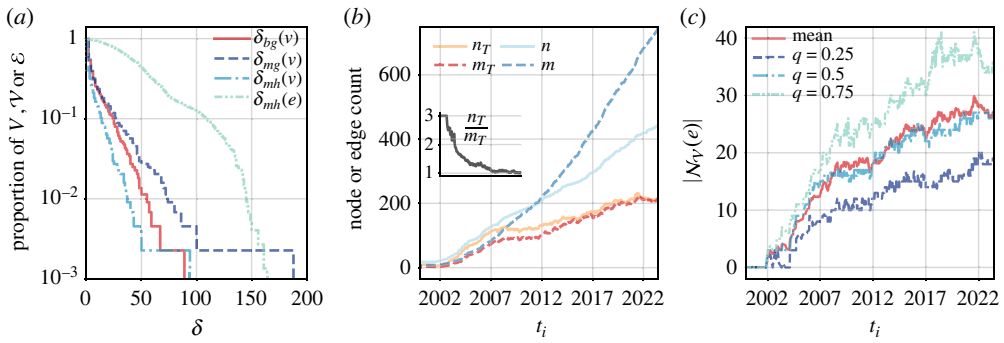
Figure 5. Descriptive statistics based on different network models of the ICSID data. We show (i) the empirical CCDFs of degrees for the static graph G and the static hypergraph H (a), (ii) the evolution of n and m in $H_{\mathcal{T}}$ in contrast to a static H cut off at the same time stamp (b), and (iii) statistics characterizing how the distribution of the number of nodes (arbitrators) in the neighbourhood of an edge (case) in $H_{\mathcal{T}}$ changes over time (c). (a) Degrees in G and H, (b) order and size in $H_{\mathcal{T}}$ and H, (c) neighbourhood sizes in $H_{\mathcal{T}}$.

repeated *collaboration pairs* ($\frac{357}{2\,226} = 0.16$), the fraction of repeated *collaboration trios* is much smaller ($\frac{20}{742} = 0.03$)—but both fractions are significantly higher than expected under a random model.[7] Similar to our findings for the GFCC data, we again observe median degrees that are much smaller than mean degrees, with $\bar{\delta} > 2\delta_M$ in all network models, and as illustrated in figure 5a, the degree distributions are skewed, too.[8] To demonstrate the impact of temporal modelling on our ICSID data, in figure 5b, we show the order and size of hypergraphs in $H_{\mathcal{T}}$ as compared to those of a static hypergraph H that, at time t_i, includes all cases whose tribunal was constituted on a date no later than t_i (effectively following a point-aggregation instead of an interval-event evolution model). We see that the total number of cases (m) grows about twice as fast as the number of active cases ($m_{\mathcal{T}}$), and the total number of arbitrators (n) grows more than three times as fast as the number of active arbitrators ($n_{\mathcal{T}}$). Notably, the number of active arbitrators per active case ($n_{\mathcal{T}}/m_{\mathcal{T}}$) drops from 3 in the early 2000s to around 1 in mid-2023 (i.e. on average, each active case shares two arbitrators on its tribunal with other active cases), indicating increasing concentration of cases among arbitrators. This concentration is also reflected in figure 5c, where we depict the temporal evolution of point statistics showing how many arbitrators are at most one hop away from an individual case (i.e. $|\mathcal{N}_{\mathcal{V}}(e)|$ for $e \in \mathcal{E}_i$ at time stamp i). Figure 5c further reveals that the mean and median node-neighbourhood sizes of edges remain consistently close to each other, which contrasts with the discrepancy between mean and median node degrees recorded in table 2b. Interestingly, while the mean node-neighbourhood size of edges has roughly doubled between 2007 and 2023, with an active case now having more than 25 arbitrators in its immediate vicinity, we also find that although over 90% of nodes have been part of the largest connected component, the diameter of this component has remained relatively large (i.e. between 6 and 9).

4. Methods

In §3, we found that the representation choice, and especially the modelling decisions (i) *graph* or *hypergraph*, (ii) *static* or *temporal* and (iii) *binary* edges or *multi*-edges, can heavily influence our results when we analyse legal citation or legal collaboration network data—at least if we look at basic statistics. Now, we investigate whether this also holds when applying established network-analysis methods. Starting from the micro level and gradually zooming out to the meso level,

---

[7]For 1000 draws from the hypergraph configuration model (see §4), the z-score for observing 357 redundant collaboration *pairs* is 4.81, and the z-score for observing 20 redundant collaboration *trios* is 10.40.

[8]Much of the skew in node degree distributions is due to one arbitrator, *Brigitte Stern*, who sat on the tribunals of 94 cases in our corpus, serving as the president in 2 cases, and appointed by (or on behalf of) respondents in the remaining 92 cases.

we demonstrate how generalizations of three classic network-analysis concepts can be applied to (temporal) legal hypergraphs: *centralities*, *motifs* and *communities*.

(a) Centralities

Centralities are fundamental tools for assessing the *importance* of individual nodes or edges in a graph, whether importance is considered to mean sheer connectedness (*degree*), the ability to quickly disseminate information (*closeness*) or gatekeeping power (*betweenness*). A number of centrality measures have been generalized from graphs to hypergraphs, including the aforementioned centralities [50] as well as eigenvector centralities [52]. Since hyperedges can intersect in more than one node, walk-based centrality measures can additionally be parametrized by a required intersection size s, which determines how many nodes must be shared between two hyperedges for them to be considered connected in the line graph $L(H)$. Furthermore, *hyperedge* centralities play a more prominent role in hypergraph analysis than *edge* centralities in classic network analysis.

ICSID example. To illustrate the impact of temporal (hyper)graph modelling on micro-level centrality assessment, in figure 6, we show the highest-centrality edges—informally termed the *centrality backbone*—of the ICSID data, modelled as a temporal hypergraph, a temporal graph, or a static hypergraph, as judged by either shortest-path betweenness centrality or shortest-path closeness centrality. Here, edge betweenness is the classic shortest-path betweenness centrality for edges, hyperedge betweenness is *1-betweenness centrality*, i.e. shortest-path betweenness centrality of nodes in the 1-line graph associated with the hypergraph, and hyperedge closeness is *1-closeness centrality*, i.e. shortest-path closeness centrality of nodes in the 1-line graph.[9] We see that the backbone obtained from the temporal graph representation does not contain any triangles, i.e. it does not respect the *higher-order* structure of the data. Similarly, the backbone obtained from the static hypergraph representation contains cases that overlap in their tribunal members but not in time, i.e. it ignores the *temporal* structure of the data.

(b) Motifs

Motifs are small, connected, non-isomorphic subgraphs that are statistically overrepresented in an observed graph when compared with some randomized null model [79]. This idea straightforwardly translates to the hypergraph setting, with the modifications that (i) *k*-motifs can contain hyperedges of cardinality up to *k*, and (ii) the null model needs to be a randomized *hypergraph* model. Here, the simplest choice is the *hypergraph configuration model*, which keeps the node-degree and hyperedge-cardinality distributions and randomizes node-to-hyperedge affiliations [56,80].

ICSID example. To show how temporal (hyper)graph modelling impacts meso-level motif analysis, we again focus on the ICSID data. Recall that our ICSID hypergraphs are 3-uniform, such that there are only three 4-motifs, displayed in figure 7a, that could theoretically occur in our data. Calling these motifs Y, T and O to reflect their shape, we count how often each of these potential motifs occurs, both for individual snapshot-hypergraphs of the temporal hypergraph $H_{\mathcal{T}}$, and for the static hypergraph H that aggregates all case observations.[10] We then compare the empirical counts with the counts we observe in 1000 hypergraphs drawn from a hypergraph configuration model. The results, depicted in figure 7b and 7c for the Y-motif in Ξ_T and in H, indicate that we

---

[9]Note that temporality is reflected in figure 6a,b via the *structure* of the underlying hypergraph: Only cases active in the last snapshot of $H_{\mathcal{T}}$ are considered present for the purposes of our centrality assessment. This contrasts with the *aggregated* input data underlying figure 6d, which was never observed as a snapshot (and hence, should *not* be analysed using a static centrality measure). Developing *temporal* centrality measures that operate on the *entire* $H_{\mathcal{T}}$ presents a great opportunity for future work.

[10]In the terminology of Gauvin *et al.* [78], we randomize both hypergraphs using degree-constrained link shufflings (based on their implicit link-timeline representations). Note that for the purposes of this exposition, we focus on how data-modelling decisions concerning our input data impact our results. Hence, we do not require motifs to be temporally connected (e.g. a Y-motif can consist of two cases that share two arbitrators, even if these cases are disjoint in time).
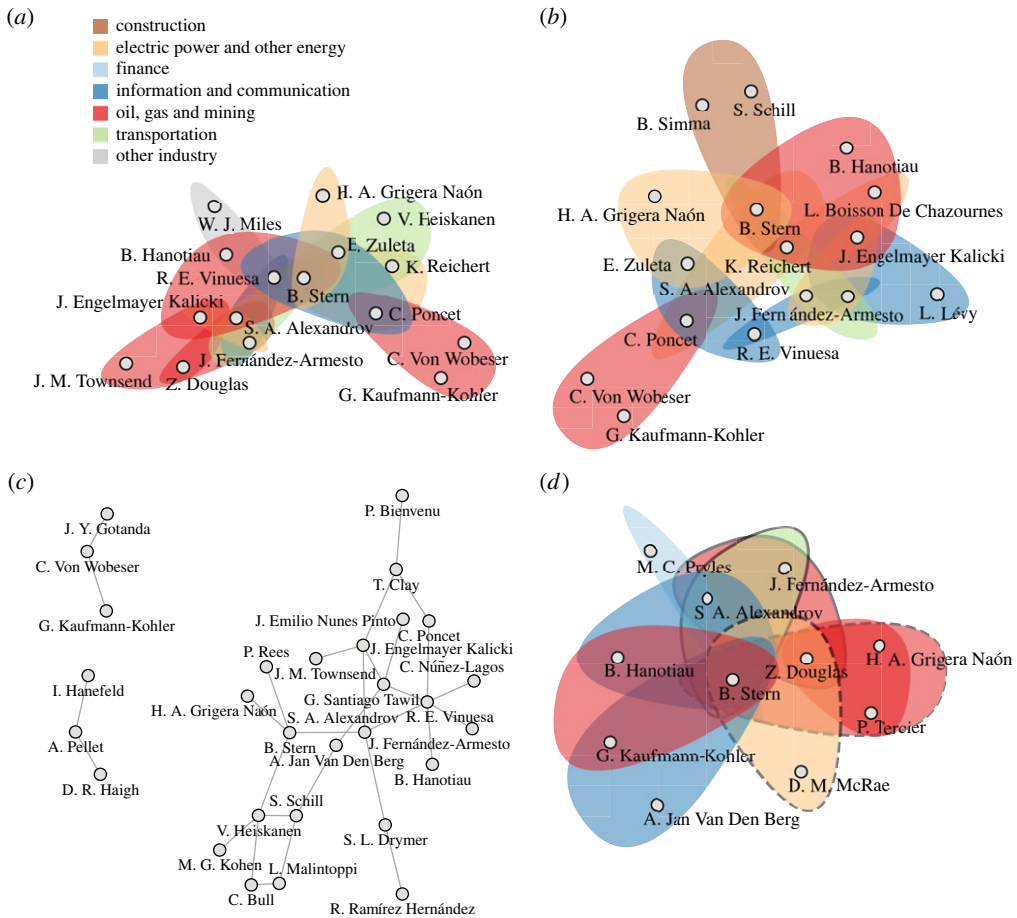
Figure 6. Centrality backbone of the ICSID data as assessed by different centrality measures. Hyperedges represent ICSID cases and nodes represent arbitrators participating in their tribunals. We show the top-ranked hyperedges based on the last snapshot of the *temporal* hypergraph representation, Ξ_T, using either hyperedge 1-betweenness centrality ((a), hyperedges ranked ≤ 10) or hyperedge 1-closeness centrality ((b), hyperedges ranked ≤ 10), the top-ranked edges (indicating that two arbitrators shared a tribunal) based on the last snapshot of the temporal *graph* representation, Γ_T, as judged by edge betweenness centrality ((c), edges ranked ≤ 30), and the top-ranked hyperedges based on the *static* hypergraph representation, H, as judged by hyperedge 1-closeness centrality ((d), hyperedges ranked ≤ 7). In (a,b,d), cases are coloured by their economic sector. In figure 6d, we additionally surround cases concluded *before* 2017 with dashed black lines and cases whose tribunal was constituted *after* 2017 with solid black lines, thus highlighting that not all top-ranked cases in H had overlapping activity times. (a) Hyperedge 1-betweenness centrality on Ξ_T. (b) Hyperedge 1-closeness centrality on Ξ_T. (c) Edge betweenness centrality on Γ_T. (d) Hyperedge 1-closeness centrality on H.

observe far more case pairs that share two arbitrators than we would expect if active arbitrators were assigned to cases randomly (figure 7b), and aggregating cases over time only makes the discrepancy between the observed hypergraph and the null model more extreme (figure 7c). While we do not observe the *O*-motif—either in the ICSID data or in the hypergraphs sampled from the configuration model—we *do* observe individual occurrences of the *T*-motif in some snapshot hypergraphs, and seven occurrences of the *T*-motif in the aggregated, static H ($z = 6.95$). This is remarkable because the *T*-motif is not compatible with complete *role specialization* among arbitrators: with three cases each sharing two arbitrators, it is impossible to assign the roles president, claimant and respondent to arbitrators such that each arbitrator has the same role in all cases.
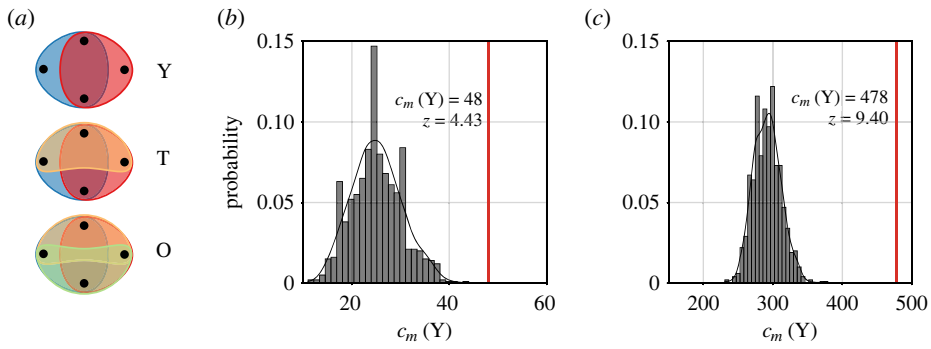
Figure 7. Theoretically possible 4-motifs in 3-uniform hypergraphs (*a*), and frequency of the *Y*-motif in a population of 1000 hypergraphs drawn from the configuration model, with the same node-degree distribution and edge-cardinality distribution as our ICSID hypergraph (red line), for the last hypergraph snapshot in our temporal hypergraph Ξ_T (*b*) and the statically modelled hypergraph H (*c*), with annotated counts $c_m(Y)$ and *z*-scores *z*. In the ICSID data, the *Y*-motif occurs much more frequently than expected under the null model. (*a*) Feasible 4-motifs. (*b*) *Y*-motif in Ξ_T. (*c*) *Y*-motif in H.

(c) Communities

Moving beyond the local structures captured by motifs, identifying *communities* (or *clusters*) of closely related nodes is one of the most important meso-level tasks in network analysis [81]. While some hypergraph-specific clustering methods exist, many of them implicitly work with a projection onto some kind of graph [82], and the toolkit for community detection in graphs is currently much more flexible. Thus, we would like to encode the information contained in a legal hypergraph H in a *weighted graph* G_w that can be used as an input to existing community-detection algorithms. How exactly this translation should proceed depends on the semantics of the underlying data. Hence, our goal here is merely to illustrate one plausible approach, tailored to judicial citation data, and to demonstrate that this approach works well in practice.

To see how we can transform a hypergraph representing legal citation data into a weighted graph reflecting its higher-order relationships, we focus on data that are structured like our GFCC data. Recall that in these data, nodes correspond to judicial decisions and hyperedges correspond to citation blocks, which are themselves associated with the decision in which they occur (i.e. we can think of citing decisions as *hyperedge colours*). Therefore, a citation block provides evidence for pairwise associations between the decisions contained in the block as well as between the citing decision and each of the cited decisions, and we would like to define the edge weights of our weighted graph to reflect the strength of this evidence. We make the following assumptions: (i) For associations between co-cited decisions, the smaller the hyperedge, the stronger the evidence.[11] (ii) For associations between the citing decision and the co-cited decisions, if we want to include them at all, the more a decision gets cited and the fewer other decisions get cited, the stronger the evidence. Hence, we begin by defining $G_w = G^\circ = (V^\circ, E^\circ) = (V, \{\{u, v\} \mid \exists\, e \in \mathcal{E} : \{u, v\} \subseteq e\})$, i.e. the skeleton of G_w is the clique expansion of $H = (\mathcal{V}, \mathcal{E})$. We then define $w : E^\circ \to \mathbb{R}$ with

$$w(\{u, v\}) = \sum_{\{u,v\} \subseteq e} \frac{1}{|e| - 1},$$

[11]This assumption is *self-contained*, i.e. it does not require any knowledge regarding the relationship between co-cited decisions that is not expressed in the data. While testing it empirically could yield an interesting contribution to the study of judicial citation practice, our theoretical justification of this assumption is that (1) one function of citation blocks is to back up contentious points, (2) the more contentious a point, the more supporting evidence should be provided for an argument to be convincing and (3) the more decisions are cited together, the fewer features will be naturally shared by *all* of them. Alternative assumptions, e.g. that each co-occurrence of two decisions in a citation block provides the same level of evidence (which would also be self-contained), can easily be incorporated by adjusting the weight-allocation procedure of our graph G_w.
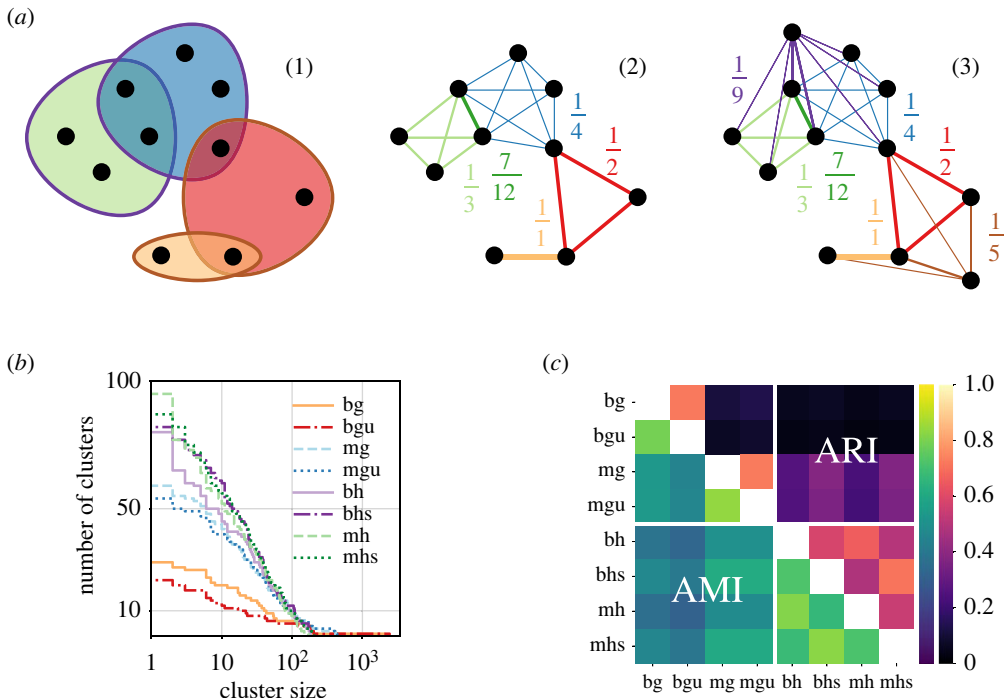
Figure 8. In the association-graph construction, we transform (1) a hypergraph H with coloured hyperedges (indicating citation blocks that belong to the same decision) into (2) a weighted clique expansion of H, optionally (3) augmenting this expansion with star-expansion-like self-affiliations (a). *Infomap* clusterings are more balanced when based on an association-graph representation derived from the GFCC hypergraph than when based on one of the traditional directed graph representations, as indicated by the cluster-size distribution of the AMI-determined medoids of 50 differently seeded clusterings using the same model (b). Clusterings based on association graphs also capture node-to-node relationships differently, as indicated by the AMI and ARI scores of pairwise comparisons between the medoid clusterings of each model (c). In figures (b,c), labels encode the underlying (hyper)graph representation as (*b*inary|*m*ulti)(*g*raph|*h*ypergraph)[self-association (versus none)|*u*ndirected flow (versus directed)]. (a) Association-graph construction, (b) medoid cluster-size distribution. (c) medoid clustering similarity.

i.e. for each hyperedge e of cardinality at least 2 in H, we add $1/|e| - 1$ to the weight of each edge $\{u, v\}$ with $\{u, v\} \subseteq e$. This implies that the sum of the weights added to the edges incident with node u in G_w as a consequence of its membership in a hyperedge e is exactly 1. It also ensures that a random walk on the resulting graph that picks its next destination proportionally to the edge weights corresponds to a non-lazy random walk on the hypergraph H that, when at node u, first picks a hyperedge e uniformly at random and then picks a node $v \in e \setminus \{u\}$ uniformly at random (this is called an *equal-edges random walk* in [62]).

Note that G_w as defined above will contain relatively little information about recent decisions, as these decisions did not get many opportunities to be cited by other decisions. To enable clustering algorithms to also integrate these decisions, we can opt to further include the evidence provided by the *self-association* of all decisions in the form of their own citations (at the expense of a *clean* correspondence between a random walk on H and a random walk on G_w). That is, we can add all citing decisions to V°, and for a decision u containing citation blocks $\mathcal{E}_u$, to each edge $\{u, v\}$ with $v \in \mathcal{E}_u$, we can add the weight $\frac{|\{e | v \in e \wedge e \in \mathcal{E}_u\}|}{\sum_{e \in \mathcal{E}_u} |e|}$, i.e. the fraction of citations by u that go to v, which implies that the sum of the weights added to edges by the self-association of u is exactly 1. Illustrating the construction in figure 8a, we call the resulting graph, which is a hybrid between a weighted clique expansion and a weighted star expansion, the *association graph* of H.

GFCC Example. To see how using an association-graph representation instead of a classic graph representation impacts clustering results, we compute 50 clusterings for each of eight different GFCC data representations, four hypergraph-derived association-graph representations and four traditional graph representations. We use *Infomap* [83] as our clustering algorithm because (i) its theoretical foundation based on random walks mimics the legal search process, (ii) it does not require us to specify the number of communities but rather derives it from the data based on information-theoretic principles and (iii) it has been shown to work well on classic representations of legal citation data [14,19,32]. Moreover, the group developing *Infomap* has tested the algorithm for hypergraph clustering based on clique expansions similar to those defined above [82]. Evaluating our clustering results is not straightforward, especially because (i) our clusterings are based on *different* (hyper)graph representations of the same data, (ii) there is no universally accepted *ground truth* and (iii) unsupervised quality measures do not necessarily capture clustering quality from a *legal* perspective. To understand how our clustering results are related and gauge clustering quality, we thus combine objective quantitative measures and subjective qualitative assessments.

We begin by investigating the quantitative structure of the partitions associated with each clustering. To this end, in figure 8*b*, we show the cluster-size distribution of the *medoid* of 50 differently seeded clusterings for each of our eight (hyper)graph representations. Here, the *medoid* of a set of clusterings using the same underlying representation is the clustering with the largest sum of within-model adjusted mutual information (AMI) scores, indicating that on average, it is closest to all other clusterings using the same model (at least when judged by AMI). We observe a steep step from cluster size 1 to larger cluster sizes in both association-graph representations not using self-affiliation (*bh* and *mh*), which is absent in the results using self-affiliation (*bhs* and *mhs*), demonstrating the importance of including self-affiliation when working with the hypergraph-based model. Furthermore, we see that clusterings based on association-graph representations tend to be rather balanced, with the three largest clusters each holding between 4% and 6% of all nodes, whereas clusterings based on classic graph representations (*bg*, *bgu*, *mg*, *mgu*) tend to have one to three large dominant clusters containing between 10% and 70% of all nodes. Hence, from a *balancedness* perspective, clusterings leveraging affiliation-graph representations with self-affiliation (*bhs*, *mhs*) appear to be preferable over all other clusterings. Notably, clusterings based on affiliation-graph representations also achieve *consistently* higher *performance*[12] than those based on classic graph representations—i.e. the former exhibit higher performance than the latter *even* when both are evaluated on *classic* graph representations.

To further explore the relationships between our clusterings, in figure 8*c*, we show the pairwise similarities between the medoid clusterings, as assessed by their AMI scores and their adjusted rand index (ARI) scores. We see that the similarities between clusterings based on association-graph representations on the one hand and clusterings based on classic graph representations on the other hand are relatively low (upper right and lower left quadrant), which suggests that association graphs capture node-to-node relationships differently from classic graph representations. However, this can also be said, more generally, for each pair of representations using dissimilar underlying models, as outside the submatrices *bg*, *bgu*, *mg*, *mgu*, *bh*, *mh*, *bhs*, *mhs*, both AMI and ARI are relatively low. Notably, the scores also reflect the large variation in cluster sizes between the different medoid clusterings (cf. figure 8*b*).

Lastly, we *manually* inspect our medoid clusterings, assigning labels to their individual clusters based on the titles (or rather: taglines) of the clustered decisions provided by the Court (available in the online materials). Here, we focus on the most promising representations based on the affiliation graph and the classic graph, namely, *mhs* and *mbu*. We find that while both clusterings capture the high-level structure of German Federal Constitutional Court jurisprudence, the hypergraph-based clustering tends to provide a more fine-grained map of German constitutional

---

[12]*Performance* is defined as the number of intra-community edges plus the number of inter-community non-edges, divided by the total number of potential edges.

law. As a result, we can observe the evolution of distinct lines of case law at a higher level of detail, unearthing, for example, a sequence of decisions on the legal status of Berlin as a special case in a more general line of decisions on state-to-state and state-to-federation relations.

5. Conclusion

In this work, we proposed *hypergraphs* and *temporal hypergraphs* as mathematical representations of legal network data. We showcased how generalizations of established network-analysis methods can be used to study temporal hypergraphs in the legal domain, and we demonstrated how data modelling decisions impact the results of these methods in case studies on legal citation networks (GFCC data: citations in German Federal Constitutional Court decisions) and legal collaboration networks (ICSID data: arbitration tribunals of cases at the International Center for Settlement of Investment Disputes). While we provided some evidence for the potential of temporal legal hypergraphs as representations of legal network data, we have arguably only scratched the surface—especially in terms of our data, methods and applications. Therefore, we see avenues for future work in several directions. First, regarding *data*, it would be interesting to see existing legal network datasets modelled as temporal hypergraphs, and the results of prior legal network studies could be re-evaluated based on the resulting representations. Second, regarding *methods*, we see great opportunities in translating further concepts from graphs to hypergraphs, and in exploring novel approaches to hypergraph data analysis based on representations of hypergraphs as filtered simplicial complexes or partially ordered sets—which will allow us to better understand, inter alia, the evolution of legal argumentation. And third, regarding *applications*, we are curious to see what further insights in-depth analyses of temporal legal hypergraphs can reveal about the structure and dynamics of complex legal systems.

References

1. Ruhl J, Katz DM, Bommarito MJ. 2017 Harnessing legal complexity. *Science* **355**, 1377–1378. (doi:10.1126/science.aag3013)
2. Murray J, Webb T, Wheatley S. 2018 *Complexity theory and law: mapping an emergent jurisprudence*. Oxfordshire: Routledge.
3. Ruhl J, Katz DM. 2015 Measuring, monitoring, and managing legal complexity. *Iowa Law Rev.* **101**, 191–244.
4. Ruhl JB. 2007 Law's complexity: a primer. *Ga. St. UL Rev.* **24**, 885–911. (doi:10.58948/0738-6206.1052)
5. Thurner S, Hanel R, Klimek P. 2018 *Introduction to the theory of complex systems*. Oxford, UK: Oxford University Press.
6. Lambiotte R, Rosvall M, Scholtes I. 2019 From networks to optimal higher-order models of complex systems. *Nat. Phys.* **15**, 313–320. (doi:10.1038/s41567-019-0459-y)
7. Kivelä M, Arenas A, Barthelemy M, Gleeson JP, Moreno Y, Porter MA. 2014 Multilayer networks. *J. Complex Netw.* **2**, 203–271. (doi:10.1093/comnet/cnu016)

16

royalsocietypublishing.org/journal/rsta *Phil. Trans. R. Soc. A* **382**: 20230141

8. Holme P, Saramäki J. 2012 Temporal networks. *Phys. Rep.* **519**, 97–125. (doi:10.1016/j. physrep.2012.03.001)

9. Battiston F, Cencetti G, Iacopini I, Latora V, Lucas M, Patania A, Young J-G, Petri G. 2020 Networks beyond pairwise interactions: structure and dynamics. *Phys. Rep.* **874**, 1–92. (doi:10.1016/j.physrep.2020.05.004)

10. Berge C. 1970 *Graphes et hypergraphes*. Paris, France: Dunot.

11. Ceria A, Wang H. 2023 Temporal-topological properties of higher-order evolving networks. *Sci. Rep.* **13**, 5885. (doi:10.1038/s41598-023-32253-9)

12. Lee G, Shin K. 2023 Temporal hypergraph motifs. *Knowl. Inf. Syst.* **65**, 1549–1586. (doi:10.1007/s10115-023-01837-2)

13. Myers A, Joslyn C, Kay B, Purvine E, Roek G, Shapiro M. 2023 Topological analysis of temporal hypergraphs. In *Int. Workshop on Algorithms and Models for the Web-graph*, pp. 127–146. Cham, Switzerland: Springer.

14. Coupette C, Beckedorf J, Hartung D, Bommarito M, Katz DM. 2021 Measuring Law Over Time: A Network Analytical Framework With an Application to Statutes and Regulations in the United States and Germany. *Front. Phys.* **9**, 269:1–269:31. (doi:10.3389/fphy.2021.658463)

15. Karpen U, Xanthaki H (eds). 2020 *Legislation in Europe: a country by country guide*. London, UK: Bloomsbury Publishing.

16. Katz DM, Coupette C, Beckedorf J, Hartung D. 2020 Complex Societies and the Growth of the Law. *Sci. Rep.* **10**, 18737. (doi:10.1038/s41598-020-73623-x)

17. Page EC, Dimitrakopoulos D. 1997 The dynamics of EU growth: a cross-time analysis. *J. Theor. Polit.* **9**, 365–387. (doi:10.1177/0951692897009003006)

18. Katz DM, Stafford DK. 2010 Hustle and flow: a social network analysis of the American federal judiciary. *Ohio St. LJ* **71**, 457–509.

19. Coupette C. 2019 *Juristische Netzwerkforschung: Modellierung, Quantifizierung und Visualisierung relationaler Daten im Recht [Legal Network Science: Modeling, Measuring, and Mapping Relational Data in Law]*. Tübingen, Germany: Mohr Siebeck.

20. Derlén M, Lindholm J. 2014 Goodbye van Gend en Loos, hello Bosman? Using network analysis to measure the importance of individual CJEU judgements. *Eur. Law J.* **20**, 667–687. (doi:10.1111/eulj.12077)

21. Boulet R, Barros-Platiau AF, Mazzega P. 2019 Environmental and trade regimes: comparison of hypergraphs modeling the ratifications of un multilateral treaties. In *Law, public policies and complex systems: networks in action*, pp. 221–242. New York, NY: Springer.

22. Fowler JH, Johnson TR, Spriggs JF, Jeon S, Wahlbeck PJ. 2007 Network analysis and the law: measuring the legal importance of precedents at the US Supreme Court. *Polit. Anal.* **15**, 324–346. (doi:10.1093/pan/mpm011)

23. Alschner W, Charlotin D. 2018 The growing complexity of the International Court of Justice's self-citation network. *Eur. J. Int. Law* **29**, 83–112. (doi:10.1093/ejil/chy002)

24. Bommarito II MJ, Katz DM, Zelner JL, Fowler JH. 2010 Distance measures for dynamic citation networks. *Physica A* **389**, 4201–4208. (doi:10.1016/j.physa.2010.06.003)

25. Leicht EA, Clarkson G, Shedden K, Newman ME. 2007 Large-scale structure of time evolving citation networks. *Eur. Phys. J. B* **59**, 75–83. (doi:10.1140/epjb/e2007-00271-7)

26. Schmid CS, Chen THY, Desmarais BA. 2022 Generative dynamics of Supreme Court citations: analysis with a new statistical network model. *Polit. Anal.* **30**, 515–534. (doi:10.1017/pan.2021.20)

27. Coupette C, Fleckner AM. 2019 Das Wertpapierhandelsgesetz (1994–2019)—Eine quantitative juristische Studie [The Securities Trading Act (1994–2019)—A Quantitative Legal Study]. In *Festschrift 25 Jahre WpHG*, pp. 53–85. Berlin, Germany: De Gruyter.

28. Li W, Azar P, Larochelle D, Hill P, Lo AW. 2015 Law is code: a software engineering approach to analyzing the United States code. *J. Bus. Tech. L.* **10**, 297. (doi:10.2139/ssrn.2511947)

29. Kirkland JH, Gross JH. 2014 Measurement and theory in legislative networks: the evolving topology of congressional collaboration. *Social Netw.* **36**, 97–109. (doi:10.1016/j. socnet.2012.11.001)

30. Ringe N, Victor JN, Cho WT. 2017 Legislative networks. In *The Oxford handbook of political networks*, pp. 471–490. Oxford, UK: Oxford University Press.

31. Tam Cho WK, Fowler JH. 2010 Legislative success in a small world: network analysis and the dynamics of congressional legislation. *J. Polit.* **72**, 124–135. (doi:10.1017/S002238160999051X)

32. Katz DM, Stafford DK. 2010b Hustle and flow: a social network analysis of the American federal judiciary. *Ohio St. LJ* **71**, 457–509.

33. Bommarito II MJ, Katz DM. 2010 A mathematical approach to the study of the United States Code. *Physica A* **389**, 4195–4200. (doi:10.1016/j.physa.2010.05.057)

34. Boulet R, Mazzega P, Bourcier D. 2018 Network approach to the French system of legal codes part II: the role of the weights in a network. *Artif. Intell. Law* **26**, 23–47. (doi:10.1007/s10506-017-9204-y)

35. Mazzega P, Bourcier D, Boulet R. 2009 The network of French legal codes. In *Proc. of the Int. Conf. on Artificial Intelligence and Law (ICAIL)*, pp. 236–237. New York, NY: Association for Computing Machinery.

36. Robaldo L, Boella G, Caro LD, Violato A. 2014 Exploiting networks in law. In *Proc. of the Ninth Int. Conf. on Language Resources and Evaluation* (*LREC*). Luxemburg City, Luxemburg: European Language Resources Association (ELRA).

37. Alschner W. 2023 Network analysis for the comparative study of judicial behaviour. In *Oxford handbook of comparative judicial behaviour*. Forthcoming.

38. Coupette C, Hartung D. 2022 Rechtsstrukturvergleichung [Structural Comparative Law]. *RabelsZ—Rabel J. Comp. Int. Priv. Law* **86**, 935–975. (doi:10.1628/rabelsz-2022-0082)

39. Derlén M, Lindholm J. 2017 Peek-a-boo, it's a case law system! Comparing the European Court of Justice and the United States Supreme Court from a network perspective. *Ger. Law J.* **18**, 647–686. (doi:10.1017/S2071832200022100)

40. Koniaris M, Anagnostopoulos I, Vassiliou Y. 2018 Network analysis in the legal domain: a complex model for European Union legal sources. *J. Complex Netw.* **6**, 243–268. (doi:10.1093/comnet/cnx029)

41. Lupu Y, Voeten E. 2012 Precedent in international courts: a network analysis of case citations by the European Court of Human Rights. *Br. J. Polit. Sci.* **42**, 413–439. (doi:10.1017/S0007123411000433)

42. Siems M. 2023 A network analysis of judicial cross-citations in Europe. *Law Soc. Inquiry* **48**, 881–905. (doi:10.1017/lsi.2022.22)

43. Alschner W, Skougarevskiy D. 2016 Mapping the universe of international investment agreements. *J. Int. Econ. Law* **19**, 561–588. (doi:10.1093/jiel/jgw056)

44. Boulet R, Barros-Platiau AF, Mazzega P. 2016 35 years of multilateral environmental agreements ratifications: a network analysis. *Artif. Intell. Law* **24**, 133–148. (doi:10.1007/s10506-016-9180-7)

45. Kim RE. 2013 The emergent network structure of the multilateral environmental agreement system. *Glob. Environ. Change* **23**, 980–991. (doi:10.1016/j.gloenvcha.2013.07.006)

46. Lee B, Lee K-M, Yang J-S. 2019 Network structure reveals patterns of legal complexity in human society: the case of the constitutional legal network. *PLoS ONE* **14**, e0209844:1–e0209844:15. (doi:10.1371/journal.pone.0209844)

47. Clark TS, Lauderdale BE, Katz JN. 2012 The genealogy of law. *Pol. Anal.* **20**, 329–350. (doi:10.1093/pan/mps019)

48. Tarissan F, Nollez-Goldbach R. 2015 Temporal properties of legal decision networks: a case study from the International Criminal Court. In *Int. Conf. on Legal Knowledge and Information Systems (JURIX)*, pp. 111–120. Amsterdam, The Netherlands: IOS Press.

49. Citraro S, Warner-Willich J, Battiston F, Siew CS, Rossetti G, Stella M. 2023 Hypergraph models of the mental lexicon capture greater information than pairwise networks for predicting language learning. *New Ideas Psychol.* **71**, 101034. (doi:10.1016/j.newideapsych.2023.101034)

50. Aksoy SG, Joslyn C, Marrero CO, Praggastis B, Purvine E. 2020 Hypernetwork science via high-order hypergraph walks. *EPJ Data Sci.* **9**, 16. (doi:10.1140/epjds/s13688-020-00231-0)

51. Carletti T, Battiston F, Cencetti G, Fanelli D. 2020 Random walks on hypergraphs. *Phys. Rev. E* **101**, 022308. (doi:10.1103/PhysRevE.101.022308)

52. Benson AR. 2019 Three hypergraph eigenvector centralities. *SIAM J. Math. Data Sci.* **1**, 293–312. (doi:10.1137/18M1203031)

53. Estrada E, Rodríguez-Velázquez JA. 2006 Subgraph centrality and clustering in complex hyper-networks. *Physica A* **364**, 581–594. (doi:10.1016/j.physa.2005.12.002)

54. Benson AR, Gleich DF, Leskovec J. 2016 Higher-order organization of complex networks. *Science* **353**, 163–166. (doi:10.1126/science.aad9029)

55. Lee G, Ko J, Shin K. 2020 Hypergraph motifs: concepts, algorithms, and discoveries. *Proc. VLDB Endowment* **13**, 2256–2269. (doi:10.14778/3407790.3407823)

56. Lotito QF, Musciotto F, Montresor A, Battiston F. 2022 Higher-order motif analysis in hypergraphs. *Commun. Phys.* **5**, 79. (doi:10.1038/s42005-022-00858-7)

57. Amburg I, Veldt N, Benson A. 2020 Clustering in graphs and hypergraphs with categorical edge labels. In *Proc. of the ACM Web Conf.*, pp. 706–717. New York, NY: Association for Computing Machinery.

58. Contisciani M, Battiston F, De Bacco C. 2022 Inference of hyperedges and overlapping communities in hypergraphs. *Nat. Commun.* **13**, 7229. (doi:10.1038/s41467-022-34714-7)

59. Veldt N, Wirth A, Gleich DF. 2020 Parameterized correlation clustering in hypergraphs and bipartite graphs. In *Proc. of the ACM Int. Conf. on Knowledge Discovery and Data Mining (KDD)*, pp. 1868–1876. New York, NY: Association of Computing Machinery.

60. Yin H, Benson AR, Leskovec J. 2018 Higher-order clustering in networks. *Phys. Rev. E* **97**, 052306. (doi:10.1103/PhysRevE.97.052306)

61. Mancastroppa M, Iacopini I, Petri G, Barrat A. 2023 Hyper-cores promote localization and efficient seeding in higher-order processes. *Nat. Commun.* **14**, 6223. (doi:10.1038/s41467-023-41887-2)

62. Coupette C, Dalleiger S, Rieck B. 2023 Ollivier-Ricci Curvature for Hypergraphs: A Unified Framework. In *Proc. of the Int. Conf. on Learning Representations (ICLR)*. Appleton, WI: International Conference on Learning Representations.

63. Zhou Y, Rathore A, Purvine E, Wang B. 2022 Topological simplifications of hypergraphs. *IEEE Trans. Vis. Comput. Graph.* **29**, 3209–3225. (doi:10.1109/TVCG.2022.3153895)

64. Baccini F, Geraci F, Bianconi G. 2022 Weighted simplicial complexes and their representation power of higher-order network data and topology. *Phys. Rev. E* **106**, 034319. (doi:10.1103/PhysRevE.106.034319)

65. Benson AR, Abebe R, Schaub MT, Jadbabaie A, Kleinberg J. 2018 Simplicial closure and higher-order link prediction. *Proc. Natl Acad. Sci. USA* **115**, E11221–E11230. (doi:10.1073/pnas.1800683115)

66. Bianconi G. 2021 *Higher-order networks*. Cambridge, UK: Cambridge University Press.

67. Sharma A, Moore TJ, Swami A, Srivastava J. 2017 Weighted simplicial complex: a novel approach for predicting small group evolution. In *Pacific-Asia Conf. on Knowledge Discovery and Data Mining (PAKDD)*, pp. 511–523. New York, NY: Springer.

68. Battiston F *et al.* 2021 The physics of higher-order interactions in complex systems. *Nat. Phys.* **17**, 1093–1098. (doi:10.1038/s41567-021-01371-4)

69. Battiston F, Petri G. 2022 *Higher-order systems*. New York, NY: Springer.

70. Majhi S, Perc M, Ghosh D. 2022 Dynamics on higher-order networks: a review. *J. R. Soc. Interface* **19**, 20220043. (doi:10.1098/rsif.2022.0043)

71. Torres L, Blevins AS, Bassett D, Eliassi-Rad T. 2021 The why, how, and when of representations for complex systems. *SIAM Rev.* **63**, 435–485. (doi:10.1137/20M1355896)

72. Badie-Modiri A, Kivelä M. 2023 Reticula: a temporal network and hypergraph analysis software package. *SoftwareX* **21**, 101301. (doi:10.1016/j.softx.2022.101301)

73. Landry NW, Lucas M, Iacopini I, Petri G, Schwarze A, Patania A, Torres L. 2023 XGI: a Python package for higher-order interaction networks. *J. Open Source Softw.* **8**, 5162. (doi:10.21105/joss.05162)

74. Lotito QF, Contisciani M, De Bacco C, Di Gaetano L, Gallo L, Montresor A, Musciotto F, Ruggeri N, Battiston F. 2023 Hypergraphx: a library for higher-order network analysis. *J. Complex Netw.* **11**, cnad019. (doi:10.1093/comnet/cnad019)

75. Weber R, Wittmann L. 2022 The role of precedents and case law in the jurisprudence of the German federal constitutional court. In *Constitutional law and precedent*, pp. 83–105. Abingdon, UK: Routledge.

76. Reed L, Paulsson J, Blackaby N. 2011 *Guide to ICSID arbitration*. The Hague, Netherlands: Kluwer Law International.

77. Rao W. 2021 Are arbitrators biased in ICSID arbitration? A dynamic perspective. *Int. Rev. Law Econ.* **66**, 105980. (doi:10.1016/j.irle.2021.105980)

78. Gauvin L, Génois M, Karsai M, Kivelä M, Takaguchi T, Valdano E, Vestergaard CL. 2022 Randomized reference models for temporal networks. *SIAM Rev.* **64**, 763–830. (doi:10.1137/19M1242252)

79. Milo R, Shen-Orr S, Itzkovitz S, Kashtan N, Chklovskii D, Alon U. 2002 Network motifs: simple building blocks of complex networks. *Science* **298**, 824–827. (doi:10.1126/science.298.5594.824)

80. Chodrow PS. 2020 Configuration models of random hypergraphs. *J. Complex Netw.* **8**, cnaa018. (doi:10.1093/comnet/cnaa018)

81. Fortunato S. 2010 Community detection in graphs. *Phys. Rep.* **486**, 75–174. (doi:10.1016/j.physrep.2009.11.002)

82. Eriksson A, Edler D, Rojas A, de Domenico M, Rosvall M. 2021 How choosing random-walk model and network representation matters for flow-based community detection in hypergraphs. *Commun. Phys.* **4**, 133. (doi:10.1038/s42005-021-00634-z)

83. Rosvall M, Bergstrom CT. 2008 Maps of random walks on complex networks reveal community structure. *Proc. Natl Acad. Sci. USA* **105**, 1118–1123. (doi:10.1073/pnas.0706851105)

84. Coupette C, Hartung D, Katz DM. 2024 Legal hypergraphs (Software). *Zenodo*. (doi:10.5281/zenodo.8081507)

85. Coupette C, Hartung D, Katz DM. 2024 Legal hypergraphs (GFCC Dataset). *Zenodo*. (doi:10.5281/zenodo.8081511)

86. Coupette C, Hartung D, Katz DM. 2024 Legal hypergraphs (ICSID Dataset). *Zenodo*. (doi:10.5281/zenodo.8081513)