Simulating subject communities in case law citation networks

Jerrold Tsin Howe SOH
Singapore Management University, jerroldsoh@smu.edu.sg

Simulating Subject Communities in Case Law Citation Networks

Jerrold Soh Tsin Howe *

Yong Pung How School of Law, Singapore Management University, Singapore, Singapore

We propose and evaluate generative models for case law citation networks that account for legal authority, subject relevance, and time decay. Since Common Law systems rely heavily on citations to precedent, case law citation networks present a special type of citation graph which existing models do not adequately reproduce. We describe a general framework for simulating node and edge generation processes in such networks, including a procedure for simulating case subjects, and experiment with four methods of modelling subject relevance: using subject similarity as linear features, as fitness coefficients, constraining the citable graph by subject, and computing subject-sensitive PageRank scores. Model properties are studied by simulation and compared against existing baselines. Promising approaches are then benchmarked against empirical networks from the United States and Singapore Supreme Courts. Our models better approximate the structural properties of both benchmarks, particularly in terms of subject structure. We show that differences in the approach for modelling subject relevance, as well as for normalizing attachment probabilities, produce significantly different network structures. Overall, using subject similarities as fitness coefficients in a sum-normalized attachment model provides the best approximation to both benchmarks. Our results shed light on the mechanics of legal citations as well as the community structure of case law citation networks. Researchers may use our models to simulate case law networks for other inquiries in legal network science.

Keywords: case law citation networks, legal network science, physics of law, network modelling, community detection

1 INTRODUCTION

Citations between cases form the bedrock of Common Law reasoning, organizing the law into directed graphs ripe for network analysis. A growing number of complexity theorists and legal scholars have sought to exploit legal networks to uncover insights about complex systems in general and legal systems in particular. Clough et al. [1] show that transitive reduction produces different effects on a citation network of judgments from the United States Supreme Court ("USSC") as compared to academic paper and patent networks. Fowler et al. [2, 3] pioneered using centrality analysis to quantify the authority of USSC precedent. This inquiry has been since been extended and applied to other courts [4] such as the Court of Justice of the European Union [5], the European Court of Human Rights [6], and the Singapore Court of Appeal ("SGCA") [7]. Examining case law citation networks ("CLCN"s) from the Supreme Courts of the United States, Canada, and India, Whalen et al. [8] find that cases whose citations have low average ages, but high variance within those ages are significantly more likely to later become highly influential. Beyond case law networks, Bommarito, Katz, and colleagues [9–11] have exploited the network structure of US and German legislation to study the growth of legal systems as well as the law's influence on society.

The community structure of CLCNs has received significantly less attention. This, however, is a rich area of research that retraces to seminal works in network science [12, 13]. Understanding communities broadly as connected subgraphs with denser within-set connectivity than without [14] allows us to automatically uncover network communities by iteratively removing links between otherwise dense subgraphs [13] or stochastically modelling link probabilities [15, 16]. A wide range of community detection techniques [17–21] as well as measures for evaluating community quality [22–24] have been studied. To our knowledge, two prior works have examined community structures in case law. Bommarito et al. [25] develop a distance measure for citation networks which they exploit to uncover communities in USSC judgments. Mirshahvalad and colleagues [26] use a network of European Court of Justice judgments to empirically benchmark a proposed method for identifying the significance of detected communities through random link perturbation.

Studying community structures in CLCNs can reveal deeper insights for both legal studies and network science. For legal studies, how far communities in CLCNs mirror legal doctrinal areas (e.g., torts and contracts) is telling of judicial (citation) practices. A judge who cites solely on doctrinal considerations should generate likewise doctrinal communities; one who cites for other (legal or political) reasons would transmit noisier signals. Community detection algorithms could also further the task of legal topic classification. Thus far, this has primarily been studied from a text-classification approach [27, 28].

For network science, CLCNs present a special case of the citation networks that have been studied extensively by the field. Studies mapping scientific papers as complex networks have demonstrated that they exhibit classic scale-free degree distributions [29] (but *cf* [30]). This has been attributed to preferential attachment, in that papers which have been cited more will be cited more. Other factors shaping paper citations include age [31] and text similarity [32]. These variables' interacting influences on citation formation yield rich structural dynamics in citation networks. For instance, over time, some papers come to be entrenched as central graph nodes while others fade into obsolescence, showing that age alone does not determine centrality [33]. Numerous generative models, discussed further in Part 2.2, have thus been proposed for citation networks, including for web hyperlinks [33–35].

As [1]'s findings suggest, however, the structure of CLCNs may differ from those of these traditional citation networks. In law, judges must consider the authority and relevance of precedent, amongst other things, when citing cases in their judgments. The doctrine of precedent further requires them to prefer certain citations to others. It is thus worth studying how CLCNs relate to traditional citation networks.

To this end, we examine how far generative models proposed for traditional citation networks can successfully replicate CLCNs. After a brief review of existing models (**Section 2.2**), we propose and evaluate a CLCN-tailored model that attempts to account for the unique mechanics of legal citations. The model is premised on an attachment function that attempts to capture aspects of legal authority, subject relevance, and time decay

(**Section 2.3**). As measures for legal authority and decay are well-established, we focus on how subject relevance may be modelled. We devise a method for simulating node-level subjects and experiment with alternative attachment functions that incorporate these vectors in four different ways: using subject cosine similarity as a standalone linear feature in the attachment model; using the same as fitness coefficients [36]; constraining nodes to citing within subject-conditional "local worlds" [37]; using subjects to generate subject-sensitive PageRank scores [38] (**Section 2.3.3**). We then study by simulation the topological and community properties of networks produced by these alternative models (**Section 3.1**) and benchmark promising models (and baselines) against two empirical CLCNs: early decisions of the United States Supreme Court and of the Singapore Court of Appeal (**Section 3.2**).[1]

We find that using subject similarity scores as fitness coefficients within a sum-normalized probability function best approximates these actual networks. However, key differences remain between the simulated and actual networks, suggesting that other factors influencing legal citations are remain unaccounted for. Nonetheless, our work represents a first step towards better capturing and studying the mechanics of case law citation networks.

2 MATERIALS AND METHODS

Section 2.1 sets the legal theory and context behind case law citation formation. **Section 2.2** explores how far these are captured by existing models. **Section 2.3** explains our proposed models. **Section 2.4** describes the simulation protocol. **Section 2.5** explains the graph metrics used to evaluate the simulations. **Section 2.6** details the benchmark datasets.

2.1 Legal Context
We define a CLCN as a graph $G(N, E)$ where *all* nodes $n \in N$ are legal case judgments and all edges $e \in E$ citations between them.[2] Let k, k^- and k^+ respectively denote the degree, in-degree, and out-degree distributions of G. Nodes may have attributes such as the authoring judge, decision date, legal subject, and the text of the judgment. Edges may be weighted (e.g., if a judgment cites another more than once), and have attributes such as whether the citation affirms or overrules the cited case.

Like all citation networks, CLCNs are time-directed and acyclic.[3] CLCNs are unique, however, because the probability that a new node n_t cites an earlier n_{t-1} (denoted $P((n_t, n_{t-1}))$ and the entire distribution P) is shaped by legal theory and practice.

---

[1]All models and simulation methods are implemented in Python and will subsequently be made available in a GitHub repository.

[2]Though judgments often cite to other forms of law such as statutes. We leave these such hybrid networks aside for now.

[3]Judgments published close in time *could* cite one another. This occurs, albeit rarely, in the Singapore dataset below.

Posner [39] identifies five overarching reasons for legal *paper* citations, namely to:

1. Acknowledge priority or influence of prior art
2. Provide bibliographic or substantive information
3. Focus disagreements
4. Appeal to authority
5. Reinforce the prestige of one's own or another's work

Reason (4) is particularly pertinent to *case* citations in Common Law systems characterized by the doctrine of binding precedent. The doctrine, in brief, means propositions of law central to a court's essential holding are taken as binding law for future purposes. Lower courts are bound to follow these holdings. While courts at the same level of hierarchy are not technically bound the same way, great deference is generally accorded to past cases nonetheless.

Recent studies have thus sought to measure legal authority with network centrality measures calibrated for the legal domain [2, 3]. Beyond citation counts, a judgment's authority is further shaped by its subject areas and time context [40]. Lawyers do not think of judgments as simply authoritative in the abstract, but within a given doctrinal subject area (i.e., torts or contracts) and at a given time. Precedential value waxes and wanes as a judgment gets entrenched by subsequent citations, ages into obsolescence over time, and as other complementary or substitute judgments emerge [7, 39]. Relevance, authority, and age are thus key, interconnected drivers of CLCN link formation [8].

2.2 Existing Models
How far are these legal mechanics captured by existing network models? In this section, we review existing citation network generative models and consider how they may be used to simulate CLCNs. Note that this paper is not a comprehensive review and will only highlight illustrative examples.

2.2.1 Degree-based Models
Classic Barabasi-Albert ("BA") [41] preferential attachment sets $P_{BA} = \frac{k_i}{\sum_j k_j}$. This model famously recovers scale-free degree distributions observed in empirical networks. However, in this model earlier nodes acquire a significant and permanent advantage over later ones, particularly if the former are cited early on. Thus, a known limitation [42] of using BA model for simulating citation networks is that, since $P_{BA} = 0$ whenever $k = 0$, new nodes (which necessarily have $k^- = 0$) are very unlikely to gain citations. Thus, the final graph may be such that most subsequent nodes cite the root node. This, of course, does not occur in empirical CLCNs (see also **Section 3.2**).

The "copying" model [34] offers a partial workaround. Links are determined by first randomly choosing one node from N as a "prototype", denoted n_p. Destinations are then either selected randomly from N by a coin toss with manually-specified probability α or copied from n_p otherwise. While the model does not explicitly include k in its attachment process, notice that nodes with zero in-degree may only be cited under the former branch, while nodes with high in-degree are more likely to be

cited under the latter branch. The copying model can therefore be broadly understood as a mixture between the Erdos-Renyi and BA models with mixture intensity controlled by α. Setting $\alpha = 1$ recovers Erdos-Renyi completely, though the model with $\alpha = 0$ is not completely equivalent to BA. This allows the copying model to produce scale-free degree distributions while leaving open the possibility for $k = 0$ nodes to be cited. However, these nodes are still less likely (depending on α) to be cited, as they cannot be cited under the copying branch. Moreover, the random process used for selecting prototypes and deciding whether to copy does not accord with legal intuitions. We do not expect new judgments (or papers) to randomly choose older judgments to either cite or copy citations from.

Another alternative proposed by Bommarito et al. ("BEA") [43] is a generalizable attachment function which considers in- and out-degree separately. More precisely,

$$P_{BEA}\left((n_t, N_{t-1})\right) = \frac{e^{\alpha k_{N_{t-1}}^- + \beta k_{N_{t-1}}^+}}{\sum_{i=1}^{|N_{t-1}|} e^{\alpha k_i^- + \beta k_i^+}} \quad (1)$$

where k_i^- and k_i^+ are node i's in- and out-degree respectively, and $\{\alpha, \beta\} \in \mathbb{R}$ are parameters for tuning their influences on P. Denoting $\{k_i^-, k_i^+\}$ as a single feature vector X_i and $\{\alpha, \beta\}$ as weight vector B, (1) may be rewritten as

$$P_{BEA}\left((n_t, N_{t-1})\right) = \frac{e^{BX_{N_{t-1}}}}{\sum_{i=1}^{|N_{t-1}|} e^{BX_i}} = softmax(BX) \quad (2)$$

Because the softmax has a vector smoothing effect, using it over conventional sum normalization ensures that non-zero citation probabilities are assigned to all nodes, even for nodes where $k = 0$. Seen this way, BEA provides a readily-extensible framework for modelling citation networks. BX is capable of encompassing an arbitrary range of weights and features.[4]

The BA and BEA models may be seen as instances of what Pham et al. [44] call the General Temporal ("GT") model. GT generalizes k into an arbitrary function of node degree $A(k)$, known as the "attachment kernel", such that $P_{GT} \propto A(k)$. The GT framework allows a large class of degree-based attachment models to be specified and estimated by maximum likelihood. For instance [44], simulate networks with $A(k) = 3\left((log(max(k, 1)))^2 + 1\right)$. GT thus offers an attractive framework for modelling legal authority in CLCN link formation. But, despite its name, GT attachment does not explicitly model node age. Yet, age has been identified as a factor driving citation networks, including CLCNs [8].

2.2.2 Aging Models
More generally, degree-based models generally ignore the well-documented influence of node age on citation formation [31, 42, 45]. By contrast, "aging" models [31, 45] propose introducing a decay vector $w(n, t)$ such that $P \propto A(k) \times w(n, t)$. Here, $w(n, t)$ can be any standard decay function which takes maps every n to weights bounded by $[0, 1]$ based on their arrival time t_n. Decay

---

[4]Notice that this effectively models link probabilities with a multi-class logistic model.

functions are further monotonically non-increasing with item age $a_n = t - t_n$, and assign weight 1 to nodes with $a_n = 0$ [46]. For instance, a simple sliding window assigns all items younger than a cut-off age to weight 1 and all other items to weight 0.

The specific aging model proposed by Wang et al. [31] uses node in-degree and exponential decay such that $P_{aging} \propto k^- \times \exp(-\tau a)$. In exponential decay functions, the parameter τ controls decay rate and induces a fixed half-life of $\frac{ln(2)}{\tau}$. Thus, the aging model gives younger nodes a higher chance to be cited than older ones with the same in-degree. However, nodes with $k^- = 0$ still have zero probability of being cited, regardless of age.

One extension of the aging model which incorporates intuitions from copying models are Singh et al.'s "relay" models [33]. Like copying models, relay models first choose a prototype n_p and performs a coin toss to determine the next step. But unlike copying models, prototypes are chosen by BA preferential attachment instead of randomly. The first coin's head probability is given by $\exp(-\tau a_n)$ (that is, exponential decay). On heads, n_p itself is cited and the process ends. On tails, a *second* coin toss with manually-specified head probability θ decides if citations are "relayed" (on heads) or if n_p will be cited nonetheless (on tails). In a "relay", a new prototype is selected from within the set of nodes *citing n_p* via a specified distribution D (Singh et al. use either uniform-random or preferential attachment). The process repeats until broken by a coin toss or the maximum specified relay depth is reached (in which case the final prototype is cited).

The exponential decay which parameterizes the first coin toss means aged papers are less likely to be cited themselves than they are to relay citations to younger papers citing them. At the same time, since prototypes are chosen initially by preferential attachment and subsequently re-chosen by D, relay models incorporate aspects of degree-based, scale-free models [33]. show that relay models better fit empirically-observed distributions of paper citation age gaps (i.e. the age difference between citing and cited papers) than the classic aging model. Relay models thus provide a more sophisticated method to account for both degree and age simultaneously. What remains missing, however, is a way to incorporate subject relevance as well. We thus turn to examine "fitness" models.

2.2.3 Fitness Models

Fitness models [47] attempt to account for each node's innate ability to compete for citations. This is generally achieved by introducing a vector of node fitness coefficients η_i. For instance, the Bianconi-Barabasi model [36] introduces a vector of uniform-randomly sampled ηs to classic preferential attachment such that $P_{BB} = \frac{\eta_i k_i}{\sum_j \eta_j k_j}$. Introducing η weakens the monopoly k holds over citation probabilities in P_{BA}. A fit node has a good chance of being cited even if its degree is low (though not if its degree is zero) [48]. further propose introducing a time-decay vector w, such that the final attachment function becomes $P = \frac{\eta_i A(k_i) w}{\sum_j \eta_j A(k_j) w}$. Notice that fitness in this regard may represent any arbitrary attribute other than degree which is believed to influence citation probabilities. For instance [42], use the ratio between (a) the

theoretical number of citations a node should receive under BA and (b) the actual number received to measure the "relevance" of a node.

Likewise, if we conceptualize η, $A(k)$, and w as capturing legal relevance, authority, and time effects respectively, this three-variable model appears ideal for modelling legal citations. Here, degree-based centrality scores (an instance of $A(k)$) have been shown to capture legal authority well (see **Section 2.3.1** below). Modelling time effects with w is also relatively standard. The crux, therefore, is devising *eta*s that capture subject relevance. This turns on the distribution it is sampled from. Drawing fitness uniformly from [0, 1], as in [36], yields in expectation an evenly distributed node ranking inconsistent with the intuition that nodes sharing subjects with the citing node should be fitter (that is, more relevant) than others.

One workaround is to calculate *eta* empirically from the *text content* of actual papers [35]. use the cosine similarity between (stopped and lemmatized) word frequency distributions of two papers' texts to their content similarity. They then propose a "three-feature model" which places content similarity scores alongside in-degree preferential attachment and power-law time decay as competing node attachment distributions. The model randomly chooses one of the three (with probability α, β, γ respectively, $\alpha + \beta + \gamma = 1$) as the final attachment function. This creates a probabilistic mixture between the content similarity, degree, and aging models, although only one model is ultimate used to generate any given edge.

Using text similarity measures to capture content overlap is intuitively logical and allows us to exploit the growing literature on text embeddings [49] (which find increasing representation in legal studies as well [7, 50]). The main drawback is that because text is difficult to simulate, we are limited to simulating *edges* between actual, existing nodes. To illustrate, for CLCNs, we can compute text similarity between empirically-observed case judgments and simulate citations between them. This would, of course, reveal important insights about CLCNs. However, generating the case judgments themselves would be difficult. We would not be able to deviate from empirically-observed node attributes.

To summarise, existing literature provides a wealth of citation network generation models. Each have their own strengths and weaknesses when theoretically applied to CLCNs. At the same time, we are not aware of any study that attempts to do this. Building on this literature, the next Section proposes a model tailored for CLCNs.

2.3 Modelling Case Law Citation Networks

Following [43], we start at time $t = 0$ with G_0 comprising $|N_0| = 1$ node and $|E_0| = 0$ edges. For each t till a specified stop T, $|N_t| \sim Po(\lambda)$ new nodes are added. Each n_t cites $|E_{n_t}| \sim Po(\mu)$ prior nodes (re-drawn independently per node). $\{\lambda, \mu\} \in \mathbb{R}$ thus control network growth rates. The main innovation of our model lies in the attachment function. In the abstract, we use a probability function $P = f(\Lambda, \rho, w)$ where Λ generalizes $A(k)$ above to encompass any function, including functions not based on k alone, capable of measuring legal authority, ρ measures subject relevance, and w is a time-decay function.

The goal is to calibrate these variables in a legally-contextualized manner. Below we expand on each variable in turn.

2.3.1 Authority

In place of k^- and k^+ above, we use Kleinberg's [51] authority and hub scores. These have been shown to accurately recover legally-significant cases from CLCNs [2, 3, 5, 40, 52]. We denote authority and hub scores as $A(k^-)$ and $A(k^+)$ respectively. While we might intuitively expect in-degree to be more representative of authority than out-degree, legal scholars have found that out-degree-based scores can be a better predictor of future case influence [40]. Cases which discuss and synthesize a large number of authorities tend to represent important disputes into which significant legal and financial resources are poured. For similar reasons, they also tend to become important legal checkpoints themselves. We therefore include both score types in the model. In any event, α or β could be set to zero to remove either score.

2.3.2 Time

Following the aging model, we use a standard exponential decay where $w(n, t) = \exp(-\tau a_n)$. This suits the legal context because citations to centuries-old judgments are not uncommon. Thus, discrete decay functions like the sliding window that apply a standard discount to all papers above a certain age would not model this observation well. Given the law's respect for old authority, we assume throughout this paper that $\tau = 0.01$, resulting in a precedent half-life of about 70 periods. Of course, future work could explore how τ may be empirically estimated (see [31]) and how it may vary over time, jurisdiction, subject, or even judge.

2.3.3 Subject Relevance

To derive relevance, we first need to simulate subjects for each node. Drawing inspiration from Latent Dirichlet Allocation ("LDA") [53], we assign each node a vector of m subjects $\phi_i \sim Dir(\psi, m)$. ψ is a m-sized vector that controls subject skew. If we want some subjects to occur more than others, possibly following a power-law, a similarly skewed ψ may be used. As a null model, however, we may set $\psi = \frac{1}{m}$.

One drawback of the Dirichlet is that non-zero probabilities may occur across many subjects. This is inconsistent with how legal cases generally discuss only a few subjects. Thus, we set a minimum cut-off of 0.1 below which subject values are floored to 0. The vector is then normalized to sum back to 1. Should this cutoff result in an entirely zero vector, one randomly-chosen subject is assigned weight 1. Because LDA treats documents as finite mixtures over m latent overlapping 'topics' that are in turn multinomial distributions across words, such a cut-off is intuitively similar to assuming that any subject generating less than 10% of the words in a judgment is not a subject that should be ascribed to it.

These subject vectors are analogous to overlapping community belonging coefficients [24], though it is always possible to partition nodes into discrete subjects by taking, for instance, $max(\phi_i)$. Here, we interpret ϕ as non-fuzzy subject *proportions* rather than probabilities. That is, a case with $\psi_i =$

{0.51, 0.25, 0.24} has subjects 1, 2, and 3, each with probability 1. But it is *primarily about* subject 1, in that 51% of its content is expected to come from the same.

Given ϕ, subject relevance can be modelled in at least four ways:

As Linear Features: First, we can derive subject similarity scores $\rho^{sim}_{n_t, N_{t-1}} = g(\phi_{n_t}, \phi_{N_{t-1}})$, where g is some vector similarity measure. Many options for g exist, but for now we default to cosine similarity given its established use in document clustering, including for legal documents [54]. The simplest way to model relevance is then to include ρ^{sim} as a standalone linear feature with its own weight γ such that $P_{linear} \propto w \times (\alpha A(k^-) + \beta A(k^+) + \gamma \rho^{sim})$.

As Fitness Coefficients: Including ρ^{sim} linearly is attractively simple, but may fail to account for interaction effects between authority, relevance, and time. Thus, a second approach is to model ρ^{sim} as fitness coefficients, so $P_{fitness} \propto \rho_{sim} \times w \times (\alpha A(k^-) + \beta A(k^+))$. This is broadly similar to the model proposed in [48], except that fitness values are computed from simulated subjects. Notice that, unlike with the linear features approach, using subject similarities as coefficients ensures that prior nodes with zero subject overlap with the citing node will be assigned $P = 0$ as well.

As Locality Constraints: Another more direct to enforce this is to limited nodes to citing within subject-conditional "local words" [37]. Within each locality, nodes can be selected by any subject-conditional probability distribution [37].'s original paper used the uniform distribution. To account for legal authority and time effects, we continue to choose nodes using HITS scores and exponential decay. To approximate the idea of nodes being authoritative within subjects, HITS scores are re-computed within the subgraph of nodes sharing at least one subject with the citer. More precisely then, $P_{local} \propto w \times (\alpha A_{local}(k^-) + \beta A_{local}(k^+))$, with the subscript 'local' denoting that the vector is computed on a subject-local subgraph.

For Subject-Sensitive PageRank: Another way to interact subject overlap with degree-based authority is to use ϕ to compute so-called topic-sensitive PageRank ("TSPR") scores [38] that may be used in place of HITS scores. While conventional PageRank [55] produces one global ranking that disregards node topic, TSPR first calculates m (the total number of topics) different rankings by setting non-uniform personalization vectors for each topic given by

$$v(n_i) = \begin{cases} \dfrac{1}{|N_j|} & \text{if } n_i \in N_j \\ 0 & \text{otherwise} \end{cases} \qquad (3)$$

where N_m is the set of nodes with subject j. Given a query node q with topic weights ϕ_q, TSPR then returns $TSPR_q = \sum \phi_q \cdot PR_j$, with PR_j being personalized PageRank scores for topic j. While TSPR has not been studied in legal networks literature, it offers a promising way to simultaneously account for authority and subject relevance using just one centrality measure. Thus, $P_{TSPR} \propto w \times TSPR_q$.

The subject models above are, of course, based on the literature reviewed in **Section 2.2**. The best approach for modelling subject

relevance is not obvious. Neither are the approaches mutually exclusive. For instance, after constraining the citable node set by subject, we may still include ρ as a linear feature while also using TSPR scores to model authority. Other combinations are also theoretically possible. But doing so may lead to contradictions. For example, calculating TSPR within a subject-constrained subgraph will return the simple PageRank score of that subgraph. It may also overplay the importance of subject relevance. For now, we study the properties of the networks produced by each approach independently.

To summarise, the proposed subject models begin with one root node and, at each time step t, adds $|N_t| \sim Po(\lambda)$ nodes with $|E_{n_t}| \sim Po(\mu)$ edges per node. Attachment probabilities are specified generally by $P = f(A(k), w(t), \phi)$, where $A(k)$ is some degree-based centrality measure including simple in/out-degree, HITS scores, and TSPR, $w(t)$ is an exponential decay function, and $\phi \sim Dir(\psi, m)$ a vector of simulated node subjects which may be incorporated into P in four different ways proposed here (though we do not rule out alternatives).

2.4 Model Simulations

To study the properties of our proposed models, we simulate 50 iterations of $T = 500$ steps for each subject model. Building on [43], we experimented with softmax as well as sum normalization for each model. For ease of reference, below we refer to the four proposed subject models as Linear, Fitness, Locality, and TSPR respectively. We use brackets to identify the normalization scheme. To illustrate, TSPR (sum) refers to a sum-normalized attachment model based on exponentially-decayed subject-sensitive PageRank scores.

For baseline comparison, we also simulated the BA, BEA, aging, copying, and relay models. We ran BA with *degree*-based preferential attachment (not in-degree), following the original model. Likewise, BEA was simulated with α and β both equal to 1. The aging baseline follows [31]'s specification, using only in-degree and an exponential decay with $tau = 0.01$. A softmax-normalized alternative was tested as well. The copying model was run with copying probability $\alpha = 0.5$ (not to be confused with the in-degree weight α in our models). Finally, we used preferential attachment relay and set relay depth at 1, $\tau = 0.3$, and $\theta = 0.9$. This follows optimal parameters found by Singh et al. for approximating the scientific paper networks they studied.

Including 5 baselines, subject models, and alternative normalizations, a total of 16 different models are run for 50 iterations each.[5] To promote comparisons across models, we fix a few key parameters in our simulations. First, $|N_t|$ is fixed universally at 1. Thus, exactly one node is added at every step for every simulation. Second, the number of subjects is fixed at 30. Third, within each iteration we first draw all $|E_{n_t}|$s from $Po(5)$ and all ϕ_{n_t}s from $Dir\left(\frac{1}{30}, 30\right)$ and use the same inputs across all models/approaches. This means the out-degree distribution of all models in the same iteration will be similar. Further, because the same subject vectors are used across all parameterizations

within the same iteration, only 50×501 individual node subject vectors are generated.[6] Fourth, all weights $\{\alpha, \beta, \gamma\}$ are set at 1 whenever relevant (though recall that γ is only used by the linear feature subject model). Finally, an exponential decay with $\tau = 0.01$ is used for all models (except the relay model).

A few implementation details are worth noting. First, because $|E_{n_t}|$ is randomly-drawn, it can exceed the total number of nodes in the existing, citable graph. Further, some attachment models result in zero citation probabilities for certain nodes, further limiting the citable node set. Thus, whenever $|N_{t-1}| < |E_{n_t}|$, we draw only $|N_{t-1}|$ destinations (while still using P). As a result, a node's realized out-degree can be lower than its initially-drawn out-degree. This is more likely to occur in the Locality models since nodes may only cite within subject-local subgraphs. This accounts for minor differences in total edge counts across model simulations. Second, nodes and edges are added in batches *after* attachment probabilities and edge destinations for every new node at a given t is determined. All computations are based solely on G_{t-1}, so nodes and edges added at the same t do not influence computations for each other. Third, after P is calculated, destination nodes are selected *without* replacement, so $|E_{n_t}|$ *unique* destination nodes are always drawn. This follows prior literature which (implicitly) samples without replacement [43].

Finally, we use NetworkX's [56] Python package to compute HITS scores. Since convergence is not guaranteed, we allow the algorithm to run for a maximum of 300 power iterations, three times the package default. We modify the package slightly to return prevailing scores if convergence is not achieved by then. To facilitate convergence (and save computational resources), we exploit the intuition that HITS scores for step $t + 1$ should not differ too much from those of step t and provide HITS scores from previous steps as warm starts. Note that this cannot be done for Locality. Because the citable node set in that model varies from node to node, relevant prior HITS scores vary.

2.5 Model Properties

An important preliminary question is whether the subject models yield scale-free degree distributions even as they seek to model time and relevance effects. As our simulation protocol fixes out-degree distributions, so comparing out-degree or total degree distribution is less meaningful. Thus, we begin by examining each model's realised in-degree distributions. To compute the average distribution across 50 iterations of the same model, we stack distributions on each other to produce a 50×501 matrix of in-degree counts. We then take the column-wise *mode* of this matrix[7] as the average distribution and compute the frequency-rank distribution of the same.

To further examine how the baseline and subject models differ in subject structure, we also derive subject signatures for each network. These are broadly inspired by [33]'s temporal bucket signatures. More precisely, denote the subject edge histogram of a

---

[5]Other parameterizations for the copying ($\alpha \in \{0.1, 0.9\}$) and relay models(uniform relay) were run but do not change our results.

[6]This includes the root node plus 500 node-time steps.

[7]Mean and median are not suitable here since they yield decimal values that would render frequency counts problematic. Where more than one mode exists we take the smallest value.

graph G whose nodes fall within m subjects as a size $m \times m$ matrix $H(i, j)$ where each entry is the total number of times nodes with subject j have cited nodes with subject i. Because one node may have many subjects, a single edge can add to many entries.[8] Thus, $|H| \geq |E|$.

The global-sum-normalized matrix $H_{norm} = \frac{H}{|H|}$ then yields the unconditional probability distribution for the cited and citing subject pairings of an arbitrary edge. Meanwhile, normalizing H column-wise so that $H_{col} = \frac{H}{|H(j)|}$ yields in each column the probability of subject i being cited conditional on the citing edge having subject j. In this way, H, H_{norm} and H_{col} offer different insights on the subject signature of a single network. Subject signatures for each *model* may then be computed by averaging these matrices across model iterations.

Finally, we compute a range of network density and community quality metrics for selected models. These include intra/inter-community edge ratios [57] and link modularity, being [18]'s modularity scores extended to the case of overlapping, directed communities [22]. We compute these metrics against (1) the simulated ϕs themselves (as ground truth subject labels) and (2) communities recovered by k-clique percolation. k-Clique percolation is a useful baseline because it is an established method for *overlapping* community detection [14]. Briefly, it recovers communities by percolating k-sized cliques to adjacent k-cliques (i.e., those sharing k-1 nodes). Its time-efficiency also means running the algorithm on all simulated networks is more practical than other equally established but less efficient algorithms, such as Girvan-Newman edge-betweenness [13]. We fix k at 3, the smallest meaningful input, to allow more, smaller communities to be returned. Though a directed k-clique algorithm exists [58], we could not find any open-source implementation. As an accessible baseline was desired, we relied on NetworkX's undirected implementation instead. Code for k-clique percolation and all community quality metrics are from CDLIB [59].

2.6 Empirical Benchmarks

Studying the structural properties and subject signatures of the simulations identifies certain more promising approaches for modelling CLCNs. As a final step, we benchmark selected approaches against two empirical CLCNs. The first is the internal network of USSC judgments that is well-studied in legal network science. To obtain legal subjects, we join Fowler et al.'s [3] edgelist with metadata from the Spaeth database [60], particularly the "issue areas" identified for each case. The second is an internal network of SGCA judgments that has also been studied in prior work [7]. The dataset covers citations between all reported decisions of the SGCA from 1965 (the year Singapore gained independence) to 2017. Judgments are assigned to subjects using catchwords provided by the Singapore Law Reports, the authoritative reporter of SGCA judgments. Following [28], we map these subject labels to 31 unique subject areas (including the "Others" category). Note that in both datasets subjects are

overlapping in that the same case may belong to more than one subject.

These networks represent different CLCN archetypes. Although both the United States and Singapore inherited English law, the legal, social, and time contexts in which each system originated and developed is vastly different. Further, while the USSC primarily (and selectively) *reviews* cases of federal and constitutional significance, the SGCA, as its name suggests, routinely considers *appeals* on matters of substantive (including private) law. The datasets are also practically usefully because both provide human-labelled legal subjects. To be sure, this also implies that comparisons between the two networks must be made with caution. On top of their different legal contexts, the legal subjects provided by each database also differ. The Spaeth database uses broad issue areas such as "Civil Rights" and "Due Process". The Singapore dataset uses specific doctrinal areas such as torts and contracts. Below we refrain from drawing comparisons between the two networks except on broad properties such as degree distributions.

As we are primarily interested in network generation, we focus on the first 2001 nodes of the USSC network and the first 1001 of the SGCA network (making allowance for one root node). More USSC nodes were used because the early USSC graph was sparser. The resultant USSC and SGCA benchmark graphs had 777 and 779 edges each. The USSC benchmark spans from the year 1791 (the first node) to 1852 (the 2001th node), while the SGCA benchmark spans from 1970 to 1999. More detailed properties of both graphs are discussed in **Section 3.2**.
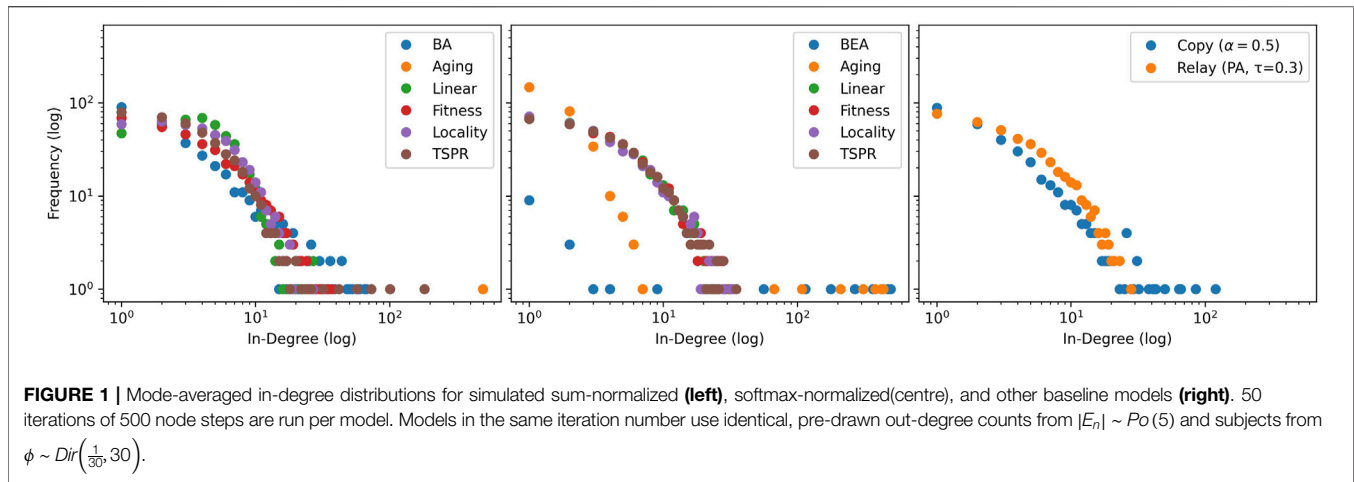
To set up the comparison, we first calibrated the model with the empirical properties of each network. Specifically, USSC simulations used an average edge rate per case of $\frac{777}{2001} = 0.388$. Subject vectors were drawn across 14 issue areas from $Dir(\psi_{us}, 14)$, with ψ_{us} being the normalized issue frequency distribution of the benchmark graph. Likewise, SGCA simulations were run using edge rates of $\frac{779}{1001} = 0.778$ with subjects drawn from $Dir(\psi_{sg}, 31)$. As with the initial simulations, we fix $|N_t| = 1$ for all t and pre-draw all $|E_t|$s and ϕs per iteration.

All benchmark simulations are run and assessed using the same protocols and metrics described in **Section 2.4**. The implementation details noted there apply. Additionally, we exploit the subject signatures to compute vector distances between the simulated and empirical graphs. In particular, we take the distance between H_{norm} as an indicator of the aggregate differences between the subject structure of two graphs. We also measure distances between the main diagonals/off-diagonals alone for insight into differences between intra/inter-subject structure. Column-wise distances between H_{col}s can be taken as a measure of per-subject differences in structure.

Amongst the wide range of graph distance measures available (see [61] for a review) we use the L1 distance because of its simple, intuitive interpretation as the sum of absolute differences.[9] More precisely, $L1(H_1, H_2) = \sum_j \sum_i |H_1(i, j) - H_2(i, j)|$. At the same time,

---

[8]To illustrate, if a node with subjects 1 and 2 cites a node with subjects 3 and 4, entries(3,1), (4,1), (3,2) and (4,2) are all incremented.

[9]We also experimented with Kullbeck-Leibler divergence and obtained similar results, though that measure often returns infinity.

FIGURE 1 | Mode-averaged in-degree distributions for simulated sum-normalized **(left)**, softmax-normalized(centre), and other baseline models **(right)**. 50 iterations of 500 node steps are run per model. Models in the same iteration number use identical, pre-drawn out-degree counts from $|E_n| \sim Po\,(5)$ and subjects from $\phi \sim Dir\left(\frac{1}{30}, 30\right)$.

classic measures such as the Hamming and Jaccard distances are meant primarily for binary (adjacency) matrices. Future work could explore more tailored distance measures. In particular, since H and H_{norm} may also be interpreted as adjacency matrices for a weighted and directed meta-graph between the *subjects*, a distance measure specialized for such graphs (e.g. [62, 63]) may be useful.

Importantly, note that subject indexes must be aligned before computing most vector distances, including the L1, as entry order affects results. In our context, this is akin to ensuring that subject i of H_1 is comparable to subject i of H_2. As our simulated subjects are arbitrary, we cannot order them by substantial content: say, to place torts at index 0 and contracts at index 1. Instead, we place the most frequent subject at index 0 and the least frequent at $m - 1$. Since the most common subjects are often also more likely to be cited (simply as a function of frequency), frequency indexing desirably concentrates probabilities towards the top-left quarter of the matrix, presenting a visually-readable signature (see **Section 3.2**). When computing signature distances, therefore, we are comparing citation patterns across subjects as ranked by frequency. If the most common subjects in graphs G_1 and G_2 both primarily cite other common subjects, signature distances would be relatively small. But if common subjects in G2 primarily cite its least common subjects, signature distances would be larger.

3 RESULTS

3.1 Model Simulations

As shown in **Figure 1**, most subject models successfully generate scale-free in-degree distributions similar to baselines. We also observe that Aging (sum) produces an average in-degree distribution with one node monopolizing most edges. Because aging models consider only in-degree, sum-normalization leads exactly to the problem, discussed in **Section 2.2**, where new nodes are never cited. This is partially addressed in by softmax normalization, but the Aging (softmax) model still manifests a visibly more imbalanced in-degree distribution than others. The
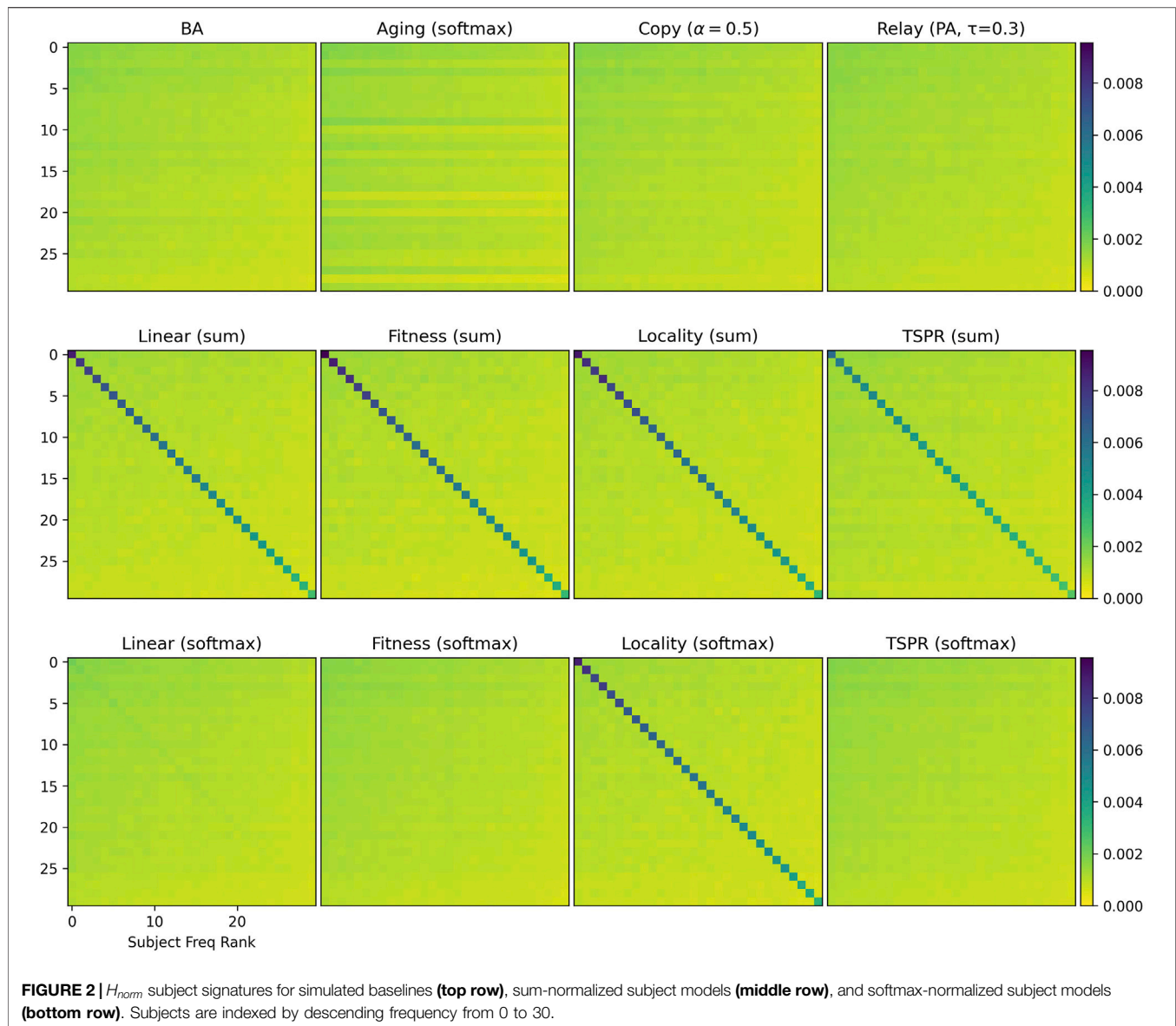
same is also true for the BEA model. This is because even though the softmax function generally assigns non-zero probabilities, small-valued input elements are quickly assigned vanishingly small values. Consider for example that $softmax\,(5, 5, 1) = 0.495, 0.495, 0.009$ at three decimal places. This affects models like BEA and Aging, which use simple degree counts as inputs, more directly because differences in feature values are larger.

The models are further differentiated by subject signatures. **Figure 2** shows that the baseline models do not reproduce intra-subject citations. Instead, citation densities are evenly spread across all subjects. As expected, this also applies to the softmax-normalized subject models (except Locality). The softmax function's smoothing effect is particularly significant here because we use features — such as authority score and cosine similarities — that range within $[0, 1]$. Much of the potentially differentiating information captured by these variables are expressed in terms of small decimal differences which are easily smoothened away. Therefore, except when used in the context of the Locality model (which would impose strict subject constraints to begin with), softmax normalization appears unsuited for our purposes.

The remaining subject models have largely similar subject signatures. To further distinguish them, we examine specific graph properties presented in **Tables 1** and **2**. While our subject models are similar to baselines in some respects like connectivity, they have clear structural differences, particularly in terms of subject communities.

General Properties: Clustering coefficients were generally low (save in the BEA and Aging (softmax) models). All models produced giant components encompassing most (around 99.9%) of the graph. However, in-degree distributions in baselines are generally more imbalanced than in the subject models. Gini coefficients for the baselines ranged between 0.52 (relay) to 0.998 (Aging (sum)) whereas those for the subject models ranged between 0.417 (Linear (sum)) to 0.598 (TSPR (sum)).

Subject Structure: Relative to all others, sum-normalized subject models yielded more intra-subject edges, lower expansion and conductance scores for communities defined by

FIGURE 2 | H_{norm} subject signatures for simulated baselines **(top row)**, sum-normalized subject models **(middle row)**, and softmax-normalized subject models **(bottom row)**. Subjects are indexed by descending frequency from 0 to 30.

the gold labels, as well as higher link modularities. These networks therefore exhibit stronger within-subject clustering (an attribute which, to recall, should in theory characterize CLCNs). Conversely, the softmaxed subject models differed only slightly from the baselines. To be sure, stronger conformity with subject labels is not *necessarily* better, since legal citations are influenced by more than subject relevance alone. Depending on the extent of subject clustering desired, therefore, TSPR approaches may offer a middle-ground.

k-Clique Structure: The k-clique metrics paint a less coherent picture. There is no clear correlation between how well the models retrace gold label subjects and the average number, size, and quality of the communities recovered by k-clique percolation. The average number (size) of communities uncovered amongst sum-normalized subject models varies from around 28 to 82 (7–18) whereas the

softmaxed subject models tend to yield around 70 k-clique communities of 5–8 nodes. Insofar as our models approximate the legal citation process, these results suggest that k-clique percolation may be less useful for clustering (and classifying) case law by legal subjects. Communities recovered by the algorithm do not appear to reflect actual legal areas, suggesting that legal subject clustering does not follow the assumptions of k-clique percolation. Nonetheless, the observed k-clique communities may be the result of other clustering mechanics inherent in legal citation networks. To this extent, they offer an independent basis for assessing the structural similarity of different simulated models. For instance, it is clear from **Table 2** that the BEA model, which always results in exactly 1 large k-clique community that encapsulates the whole network, is structurally distinct from the rest. The sum-normalized Fitness and TSPR models

TABLE 1 | Simulated properties for sum-normalized and relay models.

	BA	Aging	Relay	Linear	Fitness	Locality	TSPR
General Structure							
Avg Clust Coef	0.043 (0.0)	0.0 (0.0)	0.018 (0.0)	0.059 (0.0)	0.086 (0.01)	0.058 (0.0)	0.07 (0.0)
Giant Comp %	0.994 (0.0)	0.994 (0.0)	0.994 (0.0)	1.0 (0.0)	0.994 (0.0)	0.999 (0.0)	0.999 (0.0)
Gini (In-Deg)	0.739 (0.01)	0.998 (0.0)	0.52 (0.01)	0.417 (0.01)	0.597 (0.01)	0.456 (0.01)	0.598 (0.01)
Subject Structure							
#Intra Edges	420.08 (29.29)	83.84 (33.3)	405.86 (25.59)	2228.54 (52.4)	2507.18 (60.4)	2449.08 (59.06)	1740.94 (46.04)
Expansion	4.559 (0.11)	0.917 (0.03)	4.435 (0.11)	2.981 (0.08)	2.594 (0.07)	2.626 (0.08)	3.388 (0.09)
Conductance	0.865 (0.01)	0.923 (0.03)	0.866 (0.01)	0.437 (0.01)	0.38 (0.01)	0.387 (0.01)	0.528 (0.01)
Link Modularity	0.005 (0.0)	0.005 (0.0)	0.005 (0.0)	0.031 (0.0)	0.035 (0.0)	0.034 (0.0)	0.023 (0.0)
k-Clique Structure							
#Comms Recov'd	32.66 (7.64)	0.0 (0.0)	84.74 (10.69)	72.36 (9.53)	28.14 (5.45)	81.92 (10.15)	52.48 (6.4)
Avg Comm Size	13.485 (2.92)	–	5.205 (0.38)	8.929 (0.98)	18.058 (3.09)	7.661 (0.81)	10.02 (1.04)
#Intra Edges	1819.5 (94.97)	–	983.42 (85.87)	1915.8 (124.49)	2107.3 (89.04)	1740.84 (116.24)	1890.7 (103.13)
Expansion	4.563 (0.23)	–	4.911 (0.19)	3.864 (0.14)	4.097 (0.27)	4.01 (0.15)	4.078 (0.2)
Conductance	0.663 (0.01)	–	0.68 (0.01)	0.598 (0.01)	0.62 (0.02)	0.617 (0.01)	0.628 (0.01)
Link Modularity	0.044 (0.0)	–	0.019 (0.0)	0.042 (0.0)	0.056 (0.0)	0.038 (0.0)	0.047 (0.0)

Notes: Values represent the mean (standard deviation) of the relevant metric over 50 iterations. Relay was simulated with $\theta = 0.9$, $\tau = -0.3$, relay depth 1, and preferential relay. Relay is not a sum-normalized model but because the relay mechanism relies on sum-normalized preferential attachment, it is expedient to present its properties here instead of in a separate table.

TABLE 2 | Simulated properties for softmax-normalized and copying models.

	BEA	Aging	Copy	Linear	Fitness	Locality	TSPR
General Structure							
Avg Clust Coef	0.469 (0.01)	0.341 (0.02)	0.044 (0.0)	0.015 (0.0)	0.015 (0.0)	0.051 (0.0)	0.016 (0.0)
Giant Comp %	0.994 (0.0)	0.997 (0.0)	0.999 (0.0)	0.999 (0.0)	0.999 (0.0)	0.999 (0.0)	0.999 (0.0)
Gini (In-Deg)	0.988 (0.0)	0.905 (0.01)	0.719 (0.01)	0.544 (0.01)	0.548 (0.01)	0.564 (0.01)	0.549 (0.01)
Subject Structure							
#Intra Edges	417.82 (64.92)	411.02 (48.24)	423.82 (24.52)	476.0 (25.78)	420.42 (25.01)	2448.34 (58.85)	424.32 (26.85)
Expansion	4.561 (0.14)	4.565 (0.12)	4.557 (0.11)	4.511 (0.1)	4.559 (0.1)	2.642 (0.08)	4.555 (0.1)
Conductance	0.878 (0.02)	0.874 (0.01)	0.863 (0.01)	0.847 (0.01)	0.864 (0.01)	0.385 (0.01)	0.862 (0.01)
Link Modularity	0.005 (0.0)	0.005 (0.0)	0.005 (0.0)	0.006 (0.0)	0.005 (0.0)	0.034 (0.0)	0.005 (0.0)
k-Clique Structure							
#Comms Recov'd	1.0 (0.0)	1.32 (0.87)	27.94 (5.58)	77.34 (8.28)	76.52 (8.07)	62.96 (7.16)	77.46 (8.91)
Avg Comm Size	480.68 (4.38)	419.163 (109.05)	14.826 (2.86)	5.245 (0.27)	5.293 (0.31)	8.565 (0.89)	5.257 (0.35)
#Intra Edges	2452.9 (56.38)	2388.24 (61.4)	1680.34 (95.53)	986.3 (81.47)	1000.64 (84.98)	1742.48 (110.98)	997.5 (83.89)
Expansion	0.0 (0.0)	0.353 (0.74)	4.035 (0.27)	4.502 (0.18)	4.476 (0.16)	4.027 (0.2)	4.469 (0.15)
Conductance	0.0 (0.0)	0.064 (0.13)	0.629 (0.02)	0.663 (0.01)	0.662 (0.01)	0.629 (0.01)	0.661 (0.01)
Link Modularity	0.072 (0.0)	0.07 (0.0)	0.042 (0.0)	0.019 (0.0)	0.02 (0.0)	0.04 (0.0)	0.02 (0.0)

Notes: Values represent the mean (standard deviation) of the relevant metric over 50 iterations. Copy was simulated with $\alpha = 0.5$. Note that Copy is not a softmax-normalized model but is tabulated here for brevity.

also stand out from the other sum-normalized subject models as they tend to produce fewer but larger k-clique communities (see **Table 1**).

In sum, preliminary simulations demonstrate that the specific approach used to incorporate subject relevance induces significantly different network structures. While using subject cosine similarities as fitness values in a sum-normalized model may be *theoretically* similar to setting subject constraints on citable localities, the resulting networks differ in key aspects such as in-degree distribution and k-clique structure. The normalization function used also makes a key difference. This in turn underscores the importance of carefully selecting the subject model. Here we do not declare any one model as correct or best. To the extent that an ideal model exists, its parameterization likely turns on the specific type of CLCN we are trying to simulate. The legal and institutional context underlying the network would be relevant. Citation practices

in one court at a given time may fall closer to the Fitness model, whereas another court may cite more in line with the TSPR model.

That said, it is also clear that certain subject models, especially those using softmax normalization, are unlikely to successfully capture the nuances of legal citations for the reasons given above. For benchmarking purposes, therefore, we focus on approaches that appear more promising, being Linear (sum), Fitness (sum), Locality (softmax), and TSPR (sum). For brevity, we only present results for more promising baselines as well (being BA, Copying, and Relay).[10]

---

[10]Results for the other baselines and varying parameterizations of Copying and Relay(on file) do not affect our conclusions.
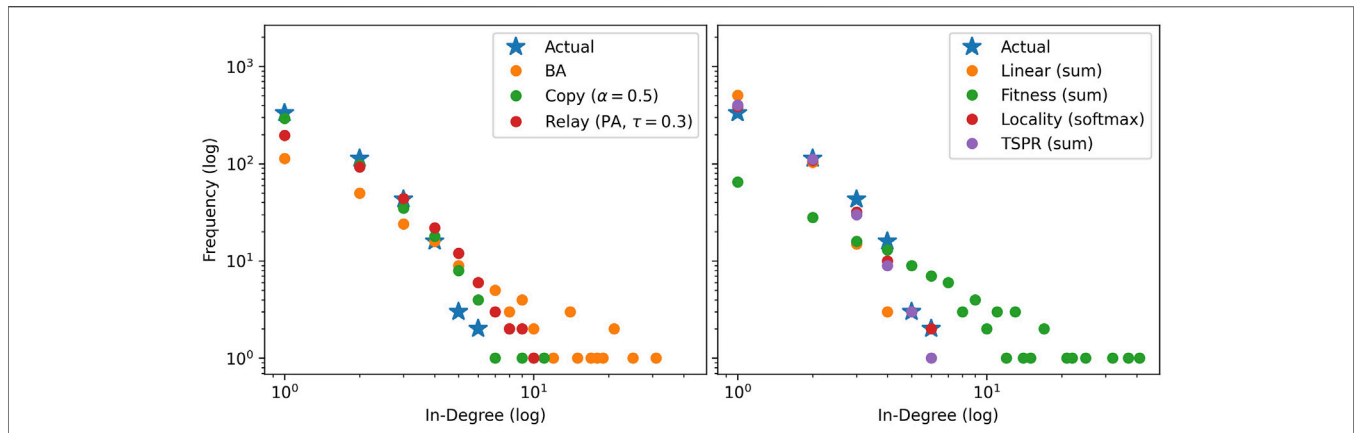
FIGURE 3 | Mode-averaged in-degree distributions for USSC-calibrated baselines **(left)** and subject models **(right)** relative to the actual. Mirroring the actual network, 50 iterations of 2000 node steps are run per model. Models in the same iteration use identical, pre-drawn out-degree counts from $|E_n| \sim Po\,(0.388)$ and subjects from $\phi \sim Dir\,(\psi_{us}, 14)$, mirroring the USSC network's actual properties.
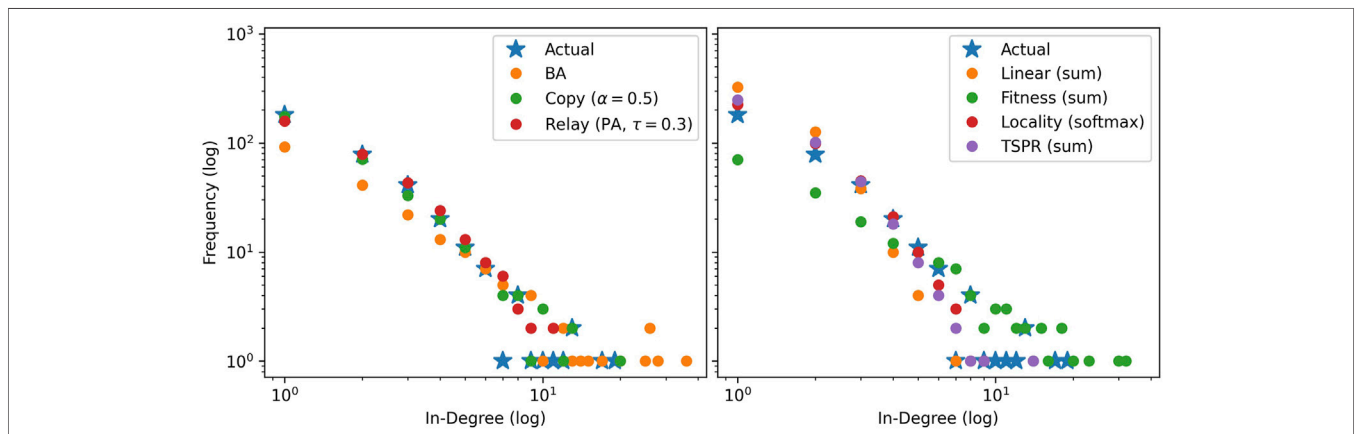

FIGURE 4 | Mode-averaged in-degree distributions for SGCA-calibrated baselines **(left)** and subject models **(right)** relative to the actual. Mirroring the actual network, 50 iterations of 1000 node steps are run per model. Models in the same iteration use identical, pre-drawn out-degree counts from $|E_n| \sim Po\,(0.778)$ and subjects from $\phi \sim Dir\,(\psi_{sg}, 31)$.
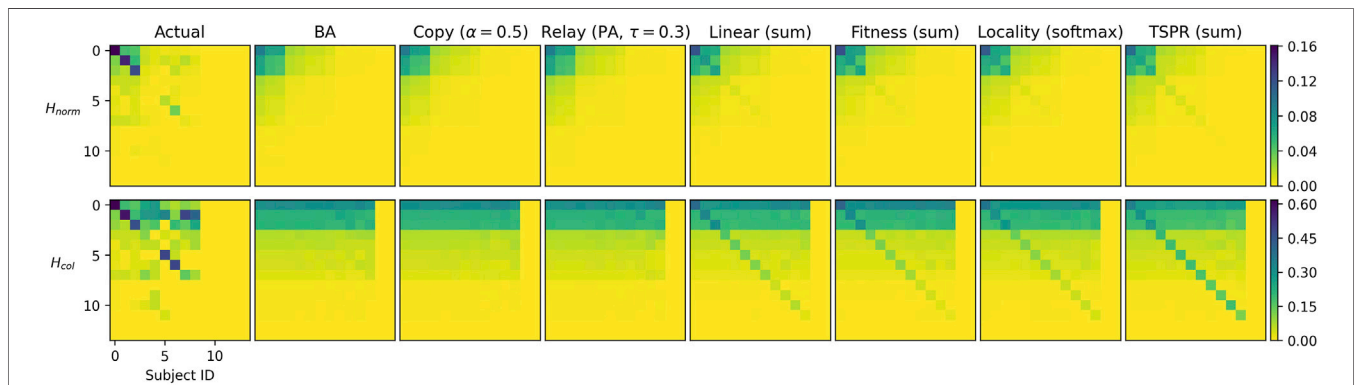

FIGURE 5 | H_{norm} **(top row)** and H_{col} **(bottom row)** subject signatures for USSC-calibrated baselines and selected subject models. Subjects are indexed by descending frequency from 0 to 14. Note that color scales differ across rows.

FIGURE 6 | H_{norm} **(top row)** and H_{col} **(bottom row)** subject signatures for SGCA-calibrated baselines and selected subject models. Subjects are indexed by descending frequency from 0 to 31. Note that color scales differ across rows.
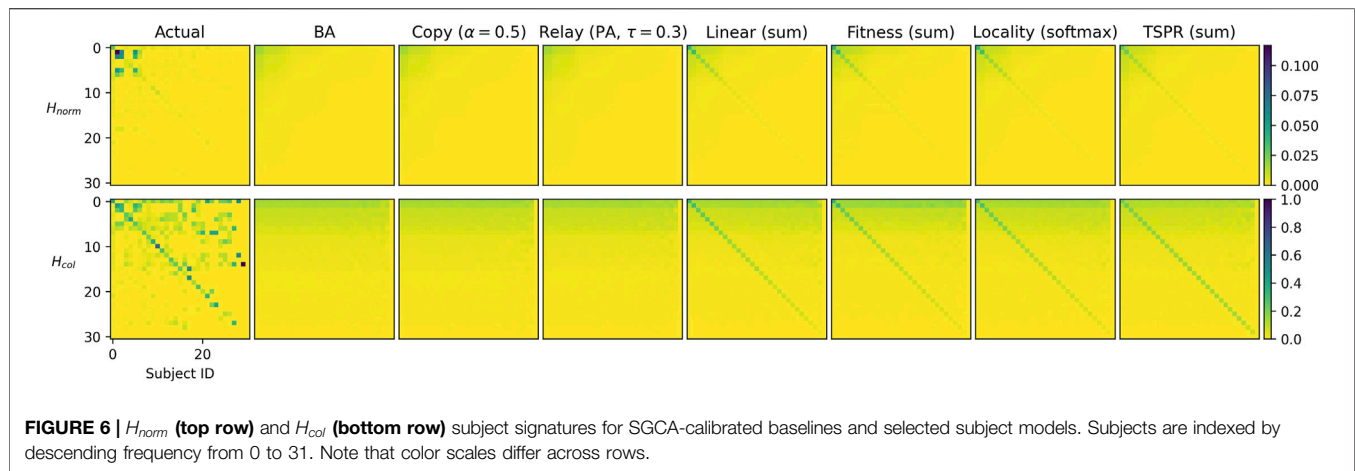
3.2 Empirical Benchmarks

Figures 3 and **4** compare in-degree distributions produced by the calibrated baseline and subject models against actual distributions from the USSC and SGCA benchmarks respectively. Both benchmark distributions are broadly reproduced by all models (including baselines), though BA and Fitness (sum) produce slightly gentler-sloping distributions in both cases. At the same time, **Figures 5** and **6** show that only the subject models successfully reproduce empirical subject signatures. To see this, first notice that since subject distributions for both the USSC and SGCA are imbalanced, citation densities are naturally concentrated within the top left quarter (which indexes the most frequent subjects). Second, observe that main diagonals for both benchmarks are also denser than their off-diagonals, showing that subject overlap materially influences actual legal citations. Although the general top-left concentration is reproduced by all models (recall that simulated subjects are drawn from a Dirichlet parameterized by actual subject frequencies), only subject models exhibit the benchmarks' distinctive signature.

These observations apply to both H_{norm} and H_{col} signatures. As expected, H_{col} densities for both the actual and simulated graphs are concentrated upwards across all rows, reflecting how nodes with the most frequent subjects are relatively likely to be cited, regardless of citing subject. This may be explained by subject frequency imbalance. To illustrate, most nodes, by definition, will have subject 0. Assuming node i has subjects $\{0, 1\}$ and node j has $\{0, 30\}$, j may cite i under the subject models due to the shared subject 0. The edge (j, i) would count not only towards $H(0, 0)$ but $H(30, 0)$ and $H(30, 1)$ as well.

Downward density gradients for the benchmarks' signatures are noticeably less smooth than for the simulations. This is clearest for the SGCA benchmark, where (within the top 15 subjects) some less frequent subjects are more likely to be cited than the most frequent. This is suggestive of specific correlations between legal subjects. For example, subject 15 may be legally very relevant to and therefore often cited by subject 5 even though the former rarely occurs. Our models do not currently account for such subject covariance and future work could explore this further.

Figure 7 tabulates L1 distances between the benchmark and simulated networks and provides further confirmation that the subject models offer a closer approximation of CLCNs than baselines. Total, diagonal, off-diagonal, and column-wise distances are consistently smaller for the subject models than for the baselines. Note that results are similar if we include other (previously discussed) baselines, including alternative parameterizations of Copying and Relay, and also if we use Kullbeck-Leibler divergence instead of L1 distance.[11] Surprisingly, none of the subject models are clearly superior to the others. All clock in comparable numbers for every metric.

Therefore, to determine which models produce better empirical fit, we look into detailed network properties for each benchmark as presented in **Tables 3** and **4**. Results here are consistent with the preliminary simulations. Clustering coefficients and giant component percentages for all models are broadly similar, but in-degree distributions vary across models.

As expected, subject models yield networks with stronger subject communities than all baselines. The subject models fit the SGCA benchmark well, each producing around 800 intra-subject edges compared to the benchmark's 889. They also yield expansion, conductance, and link modularity scores indicating better subject community quality. However, the community quality metrics for the benchmark's network are *even* better, suggesting that our models can place even more weight on subject overlap.

Fit for the USSC benchmark is less clear. The USSC network has relatively few intra-subject edges. Resultantly, the baselines fall closer to the benchmark on this metric. However, community quality for the subject models are significantly closer to the actual. Re-creating the USSC network may therefore require model parameterizations which create smaller but even more distinctive communities.

In terms of k-clique structure, the model closest to both benchmarks is Fitness (sum). The model produces on average 13.62 communities of 3.85 nodes (compared to 24 and 4.042) and

---

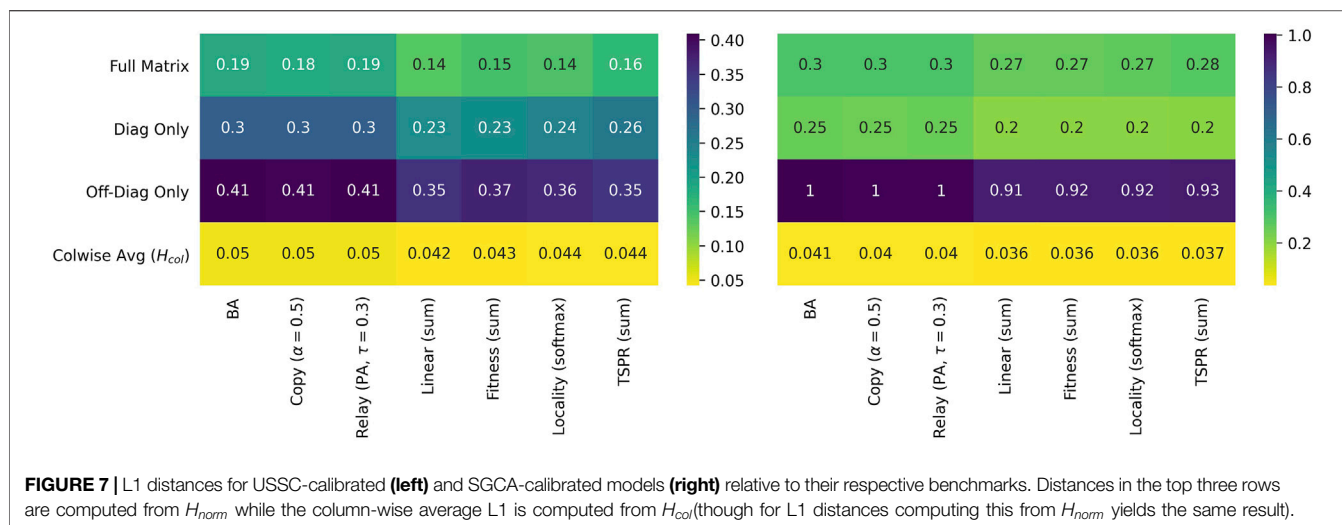[11]Though Kullbeck-Leibler returns infinity on some graphs.

FIGURE 7 | L1 distances for USSC-calibrated **(left)** and SGCA-calibrated models **(right)** relative to their respective benchmarks. Distances in the top three rows are computed from H_{norm} while the column-wise average L1 is computed from H_{col} (though for L1 distances computing this from H_{norm} yields the same result).

TABLE 3 | Actual versus simulated properties for the USSC network.

	Actual	BA	Copy	Relay	Linear (sum)	Fitness (sum)	Locality (softmax)	TSPR (sum)
General Structure								
Avg Clust Coef	0.009	0.001 (0.0)	0.0 (0.0)	0.0(0.0)	0.0(0.0)	**0.002(0.0)**	0.0(0.0)	0.001(0.0)
Giant Comp %	0.259	0.32(0.01)	0.041(0.02)	0.32(0.01)	0.025(0.01)	**0.319(0.01)**	0.101(0.04)	0.037(0.02)
Gini(In-Deg)	0.811	0.94(0.0)	0.841(0.01)	0.876(0.01)	0.736(0.01)	0.962(0.0)	**0.792(0.01)**	0.785(0.01)
Subject Structure								
#Intra Edges	384	623.76(42.38)	**618.4(31.41)**	618.54(35.13)	956.0(42.03)	964.0(46.54)	909.98(38.74)	986.66(43.5)
Expansion	0.206	0.325(0.03)	0.325(0.03)	0.323(0.03)	0.263(0.03)	0.266(0.02)	0.27(0.03)	**0.206(0.03)**
Conductance	0.356	0.764(0.02)	0.759(0.02)	0.758(0.02)	0.562(0.04)	0.585(0.04)	0.585(0.04)	**0.382(0.04)**
Link Modularity	0.036	**0.028(0.0)**	**0.028(0.0)**	**0.028(0.0)**	0.046(0.0)	0.045(0.0)	0.043(0.0)	0.044(0.0)
k-Clique Structure								
#Comms Recov'd	24	6.12(2.15)	0.14(0.5)	2.66(1.62)	1.0(0.9)	**13.62(2.93)**	0.52(0.68)	1.4(1.07)
Avg Comm Size	4.042	3.429(0.47)	3.0(0.0)	3.204(0.5)	3.095(0.26)	**3.852(0.3)**	3.091(0.29)	3.009(0.05)
#Intra Edges	124	23.26(8.63)	4.2(2.68)	9.082(5.34)	4.6(2.51)	**63.8(13.67)**	3.727(1.55)	5.73(2.42)
Expansion	1.925	**0.596(0.15)**	0.267(0.28)	0.591(0.22)	0.252(0.3)	0.468(0.13)	0.321(0.35)	0.296(0.24)
Conductance	0.378	0.207(0.04)	0.107(0.11)	**0.213(0.06)**	0.095(0.1)	0.162(0.04)	0.12(0.12)	0.114(0.08)
Link Modularity	0.011	0.002(0.0)	0.0(0.0)	0.001(0.0)	0.0(0.0)	**0.004(0.0)**	0.0(0.0)	0.001(0.0)

Notes: Except in the "actual" column, values are the mean(standard deviation) of the relevant statistic over 50 simulations per model. For the actual network properties the case issue areas assigned by the Spaeth database are used as subject labels. k-Clique expansion, conductance, and link modularity scores are only computed for and averaged within simulations which return at least 1 k-clique community and should be interpreted in this light. 45, 15, 28, and 13 iterations of the Copy, Linear, Locality, and TSPR models respectively returned 0 k-clique communities. All other models tabulated always returned at least 1. In each row, results numerically closest to the benchmark are bolded.

20.52 communities of size 4.452 (compared to 45 and 4.978) for the actual SGCA network). All other models yield significantly fewer k-clique communities. Indeed, 15 iterations of Linear (sum), 28 iterations of Locality (softmax), and 13 iterations of TSPR produced 0 k-clique communities. Conversely, Fitness (sum) always returns at least one k-clique community (including when calibrated with the SGCA network).

4 DISCUSSION

The mechanics of case law citations involve complex interactions between the legal authority of a case, its relevance to the citer's subjects, as well as the case's age. We may represent the case law

citation function in abstract as a probability distribution $P = f(\Lambda, \rho, w)$, where each input variable captures each of these attributes respectively. Determining the exact functional form to be used, however, is difficult. Numerous reasonable approaches can be imagined. For subject relevance alone, we experimented with using subject similarities as linear features, fitness values, to constrain citable horizons by subject, and to use subject-sensitive centrality scores. These all draw on existing literature on citation networks, but other approaches may be possible.

In this light, this paper studied by simulation the expected properties of networks generated by the four approaches above and compared them against established network models. We then compared more promising models against two actual CLCNs

TABLE 4 | Actual versus simulated properties for the SGCA network.

	Actual	BA	Copy	Relay	Linear(sum)	Fitness(sum)	Locality(softmax)	TSPR(sum)
General Structure								
Avg Clust Coef	0.044	**0.005(0.0)**	0.001(0.0)	0.002(0.0)	0.003(0.0)	0.009(0.0)	0.002(0.0)	**0.005(0.0)**
Giant Comp %	0.347	0.535(0.02)	0.557(0.03)	0.535(0.02)	0.572(0.04)	**0.534(0.02)**	0.584(0.03)	0.589(0.03)
Gini(In-Deg)	0.789	0.903(0.0)	0.804(0.01)	**0.789(0.01)**	0.624(0.01)	0.911(0.01)	0.721(0.01)	0.71(0.01)
Subject Structure								
#Intra Edges	889	259.96(24.28)	260.7(19.54)	257.22(21.85)	815.82(32.9)	**831.18(41.47)**	813.04(31.06)	768.5(27.58)
Expansion	0.332	0.712(0.03)	0.713(0.03)	0.709(0.03)	0.534(0.03)	0.531(0.03)	0.533(0.03)	**0.479(0.03)**
Conductance	0.449	0.87(0.01)	0.869(0.02)	0.872(0.02)	0.564(0.02)	0.577(0.03)	0.563(0.03)	**0.462(0.03)**
Link Modularity	0.047	0.011(0.0)	0.011(0.0)	0.011(0.0)	**0.037(0.0)**	**0.037(0.0)**	0.036(0.0)	0.033(0.0)
k-Clique Structure								
#Comms Recov'd	45	11.98(3.03)	2.06(1.46)	7.08(2.66)	6.14(2.15)	**20.52(4.19)**	4.74(2.08)	9.86(2.85)
Avg Comm Size	4.978	4.482(0.93)	3.659(1.5)	3.233(0.25)	3.448(0.42)	**4.457(0.64)**	3.172(0.22)	3.167(0.21)
#Intra Edges	386	69.44(15.58)	9.545(6.04)	24.56(9.53)	23.64(8.11)	**122.0(24.07)**	15.88(7.24)	33.12(10.63)
Expansion	1.139	0.922(0.15)	0.391(0.24)	**1.079(0.18)**	0.63(0.19)	0.78(0.15)	0.61(0.24)	0.519(0.14)
Conductance	0.273	**0.283(0.03)**	0.14(0.08)	0.323(0.04)	0.21(0.05)	0.242(0.04)	0.207(0.07)	0.183(0.04)
Link Modularity	0.031	0.005(0.0)	0.001(0.0)	0.002(0.0)	0.002(0.0)	**0.009(0.0)**	0.001(0.0)	0.003(0.0)

Notes: Except in the "actual" column, values are the mean(standard deviation) of the relevant statistic over 50 simulations per model. For the actual network properties, case catchwords assigned by the Singapore Law Reports are used as subject labels. k-Clique expansion, conductance, and link modularity scores are only computed for and averaged within simulations which return at least 1 k-clique community and should be interpreted in this light. 6 iterations of the Copy model returned 0 k-clique communities. All other models tabulated always returned at least 1. In each row, results numerically closest to the benchmark are bolded.

from the USSC and SGCA respectively. Our findings underscore the importance of a legally-informed model of link generation process. The properties of all proposed subject models were substantially different from those of the baseline models, (provided that sum-normalization was used), and closer to empirical benchmarks in terms of both graph properties and subject signature.

Amongst the range of approaches studied, we found softmax normalization generally unsuited for the task because it smoothens away distinguishing differences in case attributes. We also see that model properties vary substantially when alternative means of modelling subject relevance were tested. All subject models performed comparably in terms of subject signature distance from both benchmarks. However, while three of the four subject models studied yielded average in-degree distributions very close to both benchmarks, the Fitness (sum) model yielded a noticeably gentler degree-rank slope. Nonetheless, Fitness (sum) best fits the k-clique structure of both benchmarks.

The emergent picture is that all subject models provide plausible (and superior to baseline) approaches for modelling CLCNs. For our two benchmarks, however, Fitness (sum) slightly edges out the other approaches as the most promising method for modelling CLCNs. To recall, this model takes the cosine similarity between the subject belonging vectors of two cases as a fitness coefficient that modifies the weight computed from each case's authority and age. Notice that this model is similar to the General Temporal model in many respects, and may be seen as an extension of that model tailored to the legal citation context. An important caveat here is that although Fitness (sum) is *relatively* better when compared to alternative models, it is not in *absolute* terms a perfect approximation of either benchmark. Key differences remain between the actual and fitness-simulated networks. Further, alternative models may be better suited for CLCNs from other courts.

Our findings hint at possible universality in terms of the way courts think about subject relevance when deciding which cases to cite, though we hasten to add that the two benchmarks ran do not offer sufficient evidence. Universal or otherwise, the alternative models proposed may be useful for examining how courts and possibly individual judges differ when selecting cases to cite. If court/judge A's citation network is better approximated by model X while court/judge B's network is better approximated by model Y. Our models are also helpful for generating better simulated data that may be used to research other questions in legal network science. For instance, researchers could benchmark centrality algorithms against networks simulated with these models to study how far these algorithms recover legally-significant nodes. This, to recall **Section 1**, is a rich area of legal networks research. Since the data is simulated, it becomes possible to specifically dictate which nodes are legally-significant.

5 LIMITATIONS AND FUTURE WORK

This paper is the first to study how the unique mechanics of case law citations may be simulated and studied using network models. As a first step, it necessarily leaves a number of important questions unexplored.

First, although we have identified promising approaches for modelling CLCNs, alternative models for legal authority, subject relevance, and time decay remain to be studied. We have focused on comparing different methods for modelling subject relevance because this is the least explored question. Nonetheless, the effect of varying authority and time-decay models are worth studying further. Varying time-decay models in particular may yield insights on how quickly the value of precedent depreciates (a question often raised by legal scholars [8, 39]), as well as how much deference different courts accord to antiquated precedent.

Second, future work can explore a larger space of model parameterizations. An immediate extension would be using time-variant network growth rates. That is, both λ and μ may vary across time. Further, the feature weights α, β, and γ may be adjusted to generate and also reflect different judicial attitudes towards assessing authority and relevance. It may be possible to learn these weights from empirical data, whether using exponential random graph models or machine learning techniques. This would provide a means of quantitatively measuring which factors most influence legal citation decisions, providing a common metric for comparing how these differ (or remain the same) across judges, time, and space.

Third, this study was limited by data availability. Despite growing literature in case law citations analysis, few publicly available edgelists can be linked to case-level (subject) metadata. For now, we have benchmarked our models against two empirical networks produced by apex Common Law courts. Future work can consider how closely these models approximate the citation mechanics of courts in other jurisdictions, particularly those of Civil Law jurisdictions where the doctrine of precedent theoretically holds less sway.

Fourth, while we have focused primarily on network structure, the microscopic properties of our proposed networks remain largely unexplored. Future work could use node-level metrics such as centrality and accessibility to study what kinds of cases and subjects are most likely to become entrenched in the core of such networks (e.g. [64]). Additionally, the task of replicating and studying CLCNs would also benefit from also exploiting the textual content of case judgments, as was done in [35]. Methods that exploit both network structure and node attributes for community detection (e.g. [65]) could be

explored. This would connect our work to the growing literature on legal language processing (e.g. [66, 67]).

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/**Supplementary Material**, further inquiries can be directed to the corresponding author.

AUTHOR CONTRIBUTIONS

JS was in charge of all coding, data analysis, and writing.

ACKNOWLEDGMENTS

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fphy.2021.665563/full#supplementary-material

REFERENCES

1. Clough JR, Gollings J, Loach TV, and Evans TS. Transitive Reduction of Citation Networks. *J Complex Networks* (2015) 3:189–203. doi:10.1093/comnet/cnu039

2. Fowler JH, Johnson TR, Spriggs JF, Jeon S, and Wahlbeck PJ. Network Analysis and the Law: Measuring the Legal Importance of Precedents at the U.S. Supreme Court. *Polit Anal* (2007) 15:324–46. doi:10.1093/pan/mpm011

3. Fowler JH, and Jeon S. The Authority of Supreme Court Precedent. *Soc. Networks* (2008) 30:16–30. doi:10.1016/j.socnet.2007.05.001

4. Leibon G, Livermore M, Harder R, Riddell A, and Rockmore D. Bending the Law: Geometric Tools for Quantifying Influence in the Multinetwork of Legal Opinions. *Artif Intell L* (2018) 26:145–67. doi:10.1007/s10506-018-9224-2

5. Derlén M, and Lindholm J. Characteristics of Precedent: The Case Law of the European Court of Justice in Three Dimensions. *Ger Law J* (2015) 16:1073–98. doi:10.1017/S2071832200021040

6. Lupu Y, and Voeten E. Precedent in International Courts: A Network Analysis of Case Citations by the European Court of Human Rights. *Br J. Polit. Sci.* (2012) 42:413–39. doi:10.1017/S0007123411000433

7. Soh J. A Network Analysis of the Singapore Court of Appeal's Citations to Precedent. *SSRN J* (2019) 31:246–84. doi:10.2139/ssrn.3346422

8. Whalen R, Uzzi B, and Mukherjee S. Common Law Evolution and Judicial Impact in the Age of Information. *Elon L Rev* (2017) 9:115–70.

9. Bommarito MJ, II, and Katz DM. *Properties of the United States Code Citation Network* (2010). Available from: https://https://arxiv.org/abs/0911.1751 (Accessed 09 July, 2021).

10. Katz DM, Coupette C, Beckedorf J, and Hartung D. Complex Societies and the Growth of the Law. *Sci Rep* (2020) 10:18737, 2020 . ArXiv: 2005.07646. doi:10.1038/s41598-020-73623-x

11. Coupette C, Beckedorf J, Hartung D, Bommarito M, and Katz DM. Measuring Law over Time: A Network Analytical Framework with an Application to Statutes and Regulations in the United States and Germany. *Front Physics* (2021) 9:269. doi:10.3389/fphy.2021.658463

12. Watts D, and Strogatz S. Collective Dynamics of 'Small-World' Networks. *Nature* (1998) 393:3. doi:10.1038/30918

13. Girvan M, and Newman MEJ. Community Structure in Social and Biological Networks. *Proc Natl Acad Sci* (2002) 99:7821–6. Institution: National Academy of Sciences Publisher: National Academy of Sciences Section: Physical Sciences. doi:10.1073/pnas.122653799

14. Porter MA, Onnela JP, and Mucha PJ. Communities in Networks. *Notices Am Math Soc* (2009) 56:1082–166.

15. Holland PW, Laskey KB, and Leinhardt S. Stochastic Blockmodels: First Steps. *Soc Networks* (1983) 5:109–37. doi:10.1016/0378-8733(83)90021-7

16. Karrer B, and Newman MEJ. Stochastic Blockmodels and Community Structure in Networks. *Phys Rev E* (2011) 83:016107. doi:10.1103/PhysRevE.83.016107

17. Newman MEJ. Fast Algorithm for Detecting Community Structure in Networks. *Phys Rev E* (2004) 69:066133. doi:10.1103/PhysRevE.69.066133

18. Newman MEJ, and Girvan M. Finding and Evaluating Community Structure in Networks. *Phys Rev E* (2004) 69:026113. doi:10.1103/PhysRevE.69.026113

19. Clauset A, Newman MEJ, and Moore C. Finding Community Structure in Very Large Networks. *Phys Rev E* (2004) 70:066111. doi:10.1103/PhysRevE.70.066111

20. Derényi I, Palla G, and Vicsek T. Clique Percolation in Random Networks. *Phys Rev Lett* (2005) 94:160202. doi:10.1103/PhysRevLett.94.160202

21. Newman MEJ. Finding Community Structure in Networks Using the Eigenvectors of Matrices. *Phys Rev E* (2006) 74:036104. doi:10.1103/PhysRevE.74.036104

22. Nicosia V, Mangioni G, Carchiolo V, and Malgeri M. Extending the Definition of Modularity to Directed Graphs with Overlapping Communities. *J Stat Mech* (2009) 2009:P03024. doi:10.1088/1742-5468/2009/03/P03024

23. Mucha PJ, Richardson T, Macon K, Porter MA, and Onnela J-P. Community Structure in Time-dependent, Multiscale, and Multiplex Networks. *Science* (2010) 328:876–8. doi:10.1126/science.1184819

24. Chen M, and Szymanski BK. Fuzzy Overlapping Community Quality Metrics. *Soc Netw Anal Min* (2015) 5:40. doi:10.1007/s13278-015-0279-8

25. Bommarito MJ, Katz DM, Zelner JL, and Fowler JH. Distance Measures for Dynamic Citation Networks. *Physica A: Stat Mech its Appl* (2010) 389:4201–8. doi:10.1016/j.physa.2010.06.003

26. Mirshahvalad A, Lindholm J, Derlén M, and Rosvall M. Significant Communities in Large Sparse Networks. *PLoS One* (2012) 7:e33721. Public Library of Science. doi:10.1371/journal.pone.0033721

27. Sulea OM, Zampieri M, Malmasi S, Vela M, Dinu LP, and van Genabith J. Exploring the Use of Text Classification in the Legal Domain (2017) Minneapolis, Minnesota, 2-7 June 2019 CoRR abs/1710.09306.

28. Soh J, Lim HK, and Chai IE. Legal Area Classification: A Comparative Study of Text Classifiers on Singapore Supreme Court Judgments. In: Proceedings of the Natural Legal Language Processing Workshop 2019. Minneapolis, Minnesota: Association for Computational Linguistics (2019). 67–77. doi:10.18653/v1/W19-2208

29. de Solla Price DJ. Networks of Scientific Papers. *Science* (1965) 149:510–5. Institution: American Association for the Advancement of Science Section: Articles. doi:10.1126/science.149.3683.510

30. Broido AD, and Clauset A. Scale-free Networks Are Rare. *Nat Commun* (2019) 10(1):1017. Institution: Nature Publishing Group. doi:10.1038/s41467-019-08746-5

31. Wang M, Yu G, and Yu D. Effect of the Age of Papers on the Preferential Attachment in Citation Networks. *Physica A: Stat Mech its Appl* (2009) 388:4273–6. doi:10.1016/j.physa.2009.05.008

32. Menczer F. Evolution of Document Networks. *Proc Natl Acad Sci* (2004) 101:5261–5. doi:10.1073/pnas.0307554100

33. Singh M, Sarkar R, Goyal P, Mukherjee A, and Chakrabarti S. Relay-Linking Models for Prominence and Obsolescence in Evolving Networks. In: Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Halifax, Nova Scotia, 13-17 August 2017 (2017). 1077–86. ArXiv: 1609.08371. doi:10.1145/3097983.3098146

34. Kumar R, Raghavan P, Rajagopalan S, Sivakumar D, Tomkins A, and Upfal E. Stochastic Models for the Web Graph. In: Proceedings 41st Annual Symposium on Foundations of Computer Science. 12-14 November 2000, Redondo Beach, CA: IEEE Comput. Soc (2000). 57–65. doi:10.1109/SFCS.2000.892065

35. Amancio DR, Oliveira ON, and da Fontoura Costa L. Three-feature Model to Reproduce the Topology of Citation Networks and the Effects from Authors' Visibility on Their H-index. *J Informetrics* (2012) 6:427–34. doi:10.1016/j.joi.2012.02.005

36. Bianconi G, and Barabási A-L. Competition and Multiscaling in Evolving Networks. *Europhys Lett* (2001) 54:436–42. doi:10.1209/epl/i2001-00260-6

37. Li X, and Chen G. A Local-World Evolving Network Model. *Physica A: Stat Mech its Appl* (2003) 328:274–86. doi:10.1016/S0378-4371(03)00604-6

38. Haveliwala TH. Topic-sensitive Pagerank. In: Proceedings of the 11th International Conference on World Wide Web, 7-11 May 2002, New York, NY, USA: Association for Computing Machinery (2002). 517–526. doi:10.1145/511446.511513

39. Posner R. An Economic Analysis of the Use of Citations in the Law. *Am L Econ Rev* (2000) 2:381–406. doi:10.1093/aler/2.2.381

40. Carmichael I, Wudel J, Kim M, and Jushchuk J. Examining the Evolution of Legal Precedent Through Citation Network Analysis. *N. C. Law Rev.* (2017) 96:44. https://scholarship.law.unc.edu/nclr/vol96/iss1/6/

41. Barabasi AL, and Albert R. Emergence of Scaling in Random Networks. *Science* (1999) 286:4. doi:10.1126/science.286.5439.509

42. Medo M, Cimini G, and Gualdi S. Temporal Effects in the Growth of Networks. *Phys Rev Lett* (2011) 107:238701. doi:10.1103/PhysRevLett.107.238701

43. Ii MJB, Katz DM, and Zelner JL. On the Stability of Community Detection Algorithms on Longitudinal Citation Data. *Proced - Soc Behav Sci* (2010) 4:26–37. doi:10.1016/j.sbspro.2010.07.480

44. Pham T, Sheridan P, and Shimodaira H. PAFit: A Statistical Method for Measuring Preferential Attachment in Temporal Complex Networks. *PLoS One* (2015) 10:e0137796. doi:10.1371/journal.pone.0137796

45. Hajra KB, and Sen P. Aging in Citation Networks. *Physica A: Stat Mech its Appl* (2005) 346:44–8. doi:10.1016/j.physa.2004.08.048

46. Cormode G, Shkapenyuk V, Srivastava D, and Xu B. Forward Decay: A Practical Time Decay Model for Streaming Systems. In: 2009 IEEE 25th International Conference on Data Engineering, 29 March-2 April 2009, Shanghai, China: IEEE (2009). 138–49. doi:10.1109/ICDE.2009.65

47. Caldarelli G, Capocci A, De Los Rios P, and Muñoz MA. Scale-Free Networks from Varying Vertex Intrinsic Fitness. *Phys Rev Lett* (2002) 89:258702. doi:10.1103/PhysRevLett.89.258702

48. Sun J, Staab S, and Karimi F. Decay of Relevance in Exponentially Growing Networks. In: Proceedings of the 10th ACM Conference on Web Science, 30 June-3 July 2019, Boston, MA (2018) 343–51. doi:10.1145/3201064.3201084

49. Mikolov T, Sutskever I, Chen K, Corrado GS, and Dean J. Distributed Representations of Words and Phrases and Their Compositionality. In: CJC Burges, L Bottou, M Welling, Z Ghahramani, and KQ Weinberger, editors. *Advances in Neural Information Processing Systems 26*. Curran Associates, Inc.) (2013). 3111–9.

50. Chen DL, Ash E, and Livermore MA. Case Vectors: Spatial Representations of the Law Using Document Embeddings. In: DN Rockmore, editor. *Law as Data: Computation, Text, & the Future of Legal Analysis*. Santa Fe: SFI Press (2019). HOLLIS number: 99153793511303941

51. Kleinberg JM. Authoritative Sources in a Hyperlinked Environment. *J ACM* (1999) 46:604–32. doi:10.1145/324133.324140

52. Cross FB, and Spriggs JFI. The Most Important(And Best) Supreme Court Opinions and Justices. *Emory Law J.* (2010) 60:407–502. https://scholarlycommons.law.emory.edu/elj/vol60/iss2/5/

53. Blei DM, Ng AY, and Jordan MI. Latent Dirichlet Allocation. *J Machine Learn Res* (2003) 3:993–1022. doi:10.5555/944919.944937

54. Bhattacharya P, Ghosh K, Pal A, and Ghosh S. *Methods for Computing Legal Document Similarity: A Comparative Study* . Madrid, Spain: 5th Workshop on Legal Data Analysis (2019).

55. Brin S, and Page L. *The PageRank Citation Ranking: Bringing Order to the Web* (1998). Available from: http://ilpubs.stanford.edu:8090/422/1/1999-66.pdf (Accessed 09 July, 2021).

56. Hagberg AA, Schult DA, and Swart PJ. Exploring Network Structure, Dynamics, and Function Using Networkx. In: G Varoquaux, T Vaught, and J Millman, editors. Proceedings of the 7th Python in Science Conference. Pasadena, CA: (2008). 11–5.

57. Radicchi F, Castellano C, Cecconi F, Loreto V, and Parisi D. Defining and Identifying Communities in Networks. *Proc Natl Acad Sci* 101. Institution: National Academy of Sciences Section: Physical Sciences (2004). 2658–63. doi:10.1073/pnas.0400054101

58. Palla G, Farkas IJ, Pollner P, Derényi I, and Vicsek T. Directed Network Modules. *New J Phys* (2007) 9:186. doi:10.1088/1367-2630/9/6/186

59. Rossetti G, Milli L, and Cazabet R. CDLIB: a python Library to Extract, Compare and Evaluate Communities from Complex Networks. *Appl Netw Sci* (2019) 4(1):1–26. Institution: Springer. doi:10.1007/s41109-019-0165-9

60. Spaeth HJ, Epstein L, Martin AD, Segal JA, Ruger TJ, and Benesh SC. *The Supreme Court Database*. Version 2017 Release 1. St Louis, Missouri: Publication Title: The Supreme Court Database (2017).

61. Donnat C, and Holmes S. Tracking Network Dynamics: A Survey Using Graph Distances. *Ann Appl Stat* (2018) 12:971–1012. doi:10.1214/18-AOAS1176

62. Hammond DK, Gur Y, and Johnson CR. Graph Diffusion Distance: A Difference Measure for Weighted Graphs Based on the Graph Laplacian Exponential Kernel. in: 2013 IEEE Global Conference on Signal and Information Processing, 3-5 December 2013, Austin, TX (2013). 419–22. doi:10.1109/GlobalSIP.2013.6736904

63. Xu Y, Salapaka SM, and Beck CL. A Distance Metric between Directed Weighted Graphs. In: 52nd IEEE Conference on Decision and Control, 10-

13 December 2013, Firenze: IEEE (2013). p. 6359–64. doi:10.1109/CDC.2013.6760895

64. Amancio DR, Nunes MGV, Oliveira ON, and Costa Ld. F. Extractive Summarization Using Complex Networks and Syntactic Dependency. *Physica A: Stat Mech its Appl* (2012) 391:1855–64. doi:10.1016/j.physa.2011.10.015

65. Yang J, McAuley J, and Leskovec J. Community Detection in Networks with Node Attributes. In: 2013 IEEE 13th International Conference on Data Mining, 7-10 December 2013, Dallas, TX (2013). 1151–6. doi:10.1109/ICDM.2013.167

66. Aletras N, Tsarapatsanis D, Preoţiuc-Pietro D, and Lampos V. Predicting Judicial Decisions of the European Court of Human Rights: A Natural Language Processing Perspective. *PeerJ Comput Sci* (2016) 2:e93. doi:10.7717/peerj-cs.93

67. Medvedeva M, Vols M, and Wieling M. Using Machine Learning to Predict Decisions of the European Court of Human Rights. *Artif Intell L* (2019) 28:237–66. doi:10.1007/s10506-019-09255-y