CrossMark

Bending the law: geometric tools for quantifying influence in the multinetwork of legal opinions

Greg Leibon[1,2] · Michael Livermore[3] · Reed Harder[4] ·
Allen Riddell[5,6,7] · Dan Rockmore[8,9,10]

Abstract Legal reasoning requires identification through search of authoritative legal texts (such as statutes, constitutions, or prior judicial opinions) that apply to a given legal question. In this paper, using a network representation of US Supreme Court opinions that integrates citation connectivity and topical similarity, we model the activity of law search as an organizing principle in the evolution of the corpus of legal texts. The network model and (parametrized) probabilistic search behavior generates a Pagerank-style ranking of the texts that in turn gives rise to a natural

✉ Dan Rockmore
  rockmore@math.dartmouth.edu

  Greg Leibon
  gleibon@gmail.com

  Michael Livermore
  mal5un@virginia.edu

  Reed Harder
  reedharder@gmail.com

  Allen Riddell
  abr@riddell.org

[1] Coherent Path, Inc., Boston, MA 02114, USA

[2] Department of Mathematics, Dartmouth College, Hanover, NH 03755, USA

[3] School of Law, University of Virginia, Charlottesville, VA 22904, USA

[4] Thayer School of Engineering, Dartmouth College, Hanover, NH 03755, USA

[5] Department of Mathematics, Dartmouth College, Hanover, NH 03755, USA

[6] Department of English, Dartmouth College, Hanover, NH 03755, USA

[7] Present Address: School of Information Sciences, University of Indiana, Bloomington, IN, USA

[8] Department of Mathematics, Dartmouth College, Hanover, NH 03755, USA

[9] Department of Computer Science, Dartmouth College, Hanover, NH 03755, USA

geometry of the opinion corpus. This enables us to then measure the ways in which new judicial opinions affect the topography of the network and its future evolution. While we deploy it here on the US Supreme Court opinion corpus, there are obvious extensions to large evolving bodies of legal text (or text corpora in general). The model is a proxy for the way in which new opinions influence the search behavior of litigants and judges and thus affect the law. This type of "legal search effect" is a new legal consequence of research practice that has not been previously identified in jurisprudential thought and has never before been subject to empirical analysis. We quantitatively estimate the extent of this effect and find significant relationships between search-related network structures and propensity of future citation. This finding indicates that "search influence" is a pathway through which judicial opinions can affect future legal development.

1 Introduction

Judicial decision-making is characterized by the application by courts of author-itative rules to the stylized presentation of disputed claims between competing litigants. These authoritative rules are set forth in legal source materials such as constitutions, statutes, and written opinions supporting prior decisions. For a legal source to have bearing on a current dispute, it must be retrievable by the relevant legal actors. The problem of organizing legal texts into a comprehensible whole has been recognized since Justinian I's Corpus Juris Civilis issued in 529–534. The acute problems of identifying relevant legal sources (i.e., legal precedent) presented by the common law tradition has spurred codification and classification efforts that have ranged from Blackstone's "Commentaries on the Laws of England (1765–1769)" to the codification movement in the late nineteenth century (Garoupa and Morriss 2012), to the development and spread of the West American Digest System in the twentieth century (West 1909). Most recently, the effect of digitization on the evolution of the law, primarily in its impact on legal research, has become a subject of inquiry (see e.g., Berring 1986, 1987; Fronk 2010; Hanson and Allan 2002; Hellyer 2005; Katsh 1993; McGinnis and Wasick 2015; Schauer and Wise 2000).

In this paper we consider the textual corpus of legal sources as an evolving *landscape* that carries a natural geometry and comprises *regions* of the law whose development and shifting boundaries are influenced by the dynamics and feedback of *law search*. Everything devolves from a model of the process of legal research carried out in the corpus in which "actors" start from a case or opinion and then build out an understanding of the relevant issues by (1) following citations, (2) searching for cases that cite the initial case of interest, and (3) identifying textually

---

[10]   The Santa Fe Institute, Santa Fe, NM 87501, USA

similar cases. These actions have a natural network—more precisely, a *multinet-work*—formulation, in which legal sources are connected to each other based on citation information and textual similarity as described by a *topic model* representation of their textual content. Topic models represent texts (embodied as word-frequency distributions or "bag-of-words" representations) as mixtures of *topics*. "Topic" as used in this sense has a technical meaning and is defined as a probability distribution over the vocabulary in the corpus. Topics are uncovered and discovered according to a well-known and by now widely deployed methodology (see e.g., Blei 2012) that we briefly describe below. Our use of three kinds of connectivity (as opposed to one) in the text corpus structures the corpus in a multinetwork representation, a combinatorial structure that has proved useful in a number of different contexts, such as biology and economics (e.g., Barigozzi et al. 2011; Blinov et al. 2012; Kivelä et al. 2014). In this work we introduce for the first time the multinetwork concept to the novel contexts of text-mining and text search, with a specific application to judicial texts.

We use the multinetwork framework to define a notion of search generalizing the Markov model (discrete time random walk) that encodes Google's famous "websurfer" webpage search model (Brin and Page 1998). The webpage ranking system Pagerank is simply the stationary vector of this model (Bryan and Leise 2006). Rankings are of course useful (and of course profitable), but the random walk also will give rise to a natural notion of distance on the underlying state space, roughly defined in terms of the expected time (number of steps) needed to go from one state to another and it is this metric point of view that we explore herein. In our setting, distance reflects the ease with which a human user of the legal corpus could navigate from one legal source to another, based on a weighted combination of searches along the underlying citation and topical similarity networks. The latter is usually reduced to a keyword search in standard resources (e.g., through a commercial database such as Lexis-Nexis). The derived inter-opinion distances support the discovery of well-defined regions (in this case, groups of legal sources) that are relatively close to each other, but relatively distant from other regions. Distance is also a proxy for relevance. When new judicial decisions are issued and the supporting opinions are incorporated into the legal corpus, they interact with search technology to change the legal sources that will be discovered during the next search. For example, some new opinions can link together previously distant opinions, making them more easily discoverable. In turn, these new connections can foster new arguments. This is a new kind of legal effect that, as far as we know, has never been identified as a theoretical possibility, much less formalized and subjected to an empirical test.

The random walk setting also enables the creation/definition of a notion of *curvature* for the underlying state space (think of a state space as the cities and towns in a landscape of rolling hills and valleys). As per the usual interpretation of this geometric notion, the more negative the curvature of a region[1] of the legal

---

[1] The standard example of a point of negative curvature is the *saddle point*—so named for the curvature of the center of a riding saddle. A marble placed there would rapidly move away from the point, if in an indeterminate direction.

landscape, the easier it is to navigate to legal sources outside that region from legal sources that are inside of the region. Curvature may change over time as new legal sources are added to the corpus. An increase in curvature in a given region[2] indicates increasing difficulty in navigating from the interior of the region to legal sources outside it. This has the interpretation that the region has become more isolated from the rest of the legal corpus and thus is less relevant to new opinions outside of the region. We refer to this effect as *puddling*. The opposite effect wherein curvature decreases is referred to as *drainage*. Drainage is characterized by ease of navigation from points (legal sources) inside the region to those that are outside. Notions of network curvature have only just begun to make their way into applied literature. Some early work has adapted the idea of Ricci curvature to the network setting, mainly for its relation to various isoperimetric inequalities (see e.g., Chung and Yau 1996; Lin and Yau 2010). More recent work approaches the idea from the point of view of *optimal transport* (Ollivier 2009). This in turn makes strong connections to discrete Markov chains—as does ours—but this other work is quite different from the approach taken herein.

Use of the citation network to measure the influence of judicial opinions is now well-studied (see e.g., Bommarito et al. 2009; Fowler and Jeon 2008; Fowler et al. 2007), although interesting potential avenues of this kind of investigation in the judicial context remain underexplored (see e.g. Uzzi et al. 2013 for a citation network analysis in the context of scientific articles). Topic models, however, have only just very recently entered legal studies and have already showed great promise as a foundation for new quantitative avenues of analysis (George et al. 2014; Livermore et al. 2017; Nardi and Moe 2014; Rice 2012).

Citation networks and topic modeling are examples of computational methods useful to legal studies. Early conversations concerning law and digitization focused on distinction in "context" between digital and physical forms, for example, whether digitization enhanced or reduced reading comprehension or facilitated or undermined serendipity in conducting searches. In particular, the legal significance of the effects of various search modalities (citation-based, keyword, unstructured text) are only just becoming apparent (see e.g. McGinnis and Wasick 2015). Our work may suggest ways to begin to quantify some of these effects and empirical studies comparing our search model with actual human search results is in preparation. In this paper we focus on the collection of all U.S. Supreme Court cases from 1951 to 2002. A project to extend our work to include the Circuit courts is already underway.

In the next section we explain in a bit more detail the mathematical background and framework. Section 3 presents our results, showing that the precise notions of puddling and drainage correspond to a measurable waning and waxing respectively of relevance over time. We also briefly introduce the publicly accessible database and user interface (www.bendingthelaw.org) that we have constructed for the engagement with and visualization of the multinetwork of opinions. We then conclude with some thoughts about next steps and extensions of this work. Two technical appendices provides a more detailed mathematical justification (based on

---

[2] A well is a standard example of a point of positive curvature.

Riemannian geometry) for our definition of multinetwork curvature as well as motivation for a certain parameter choice in the analysis. The paper can be read without these sections, but we include them for the sake of completeness.

2 The mathematical framework

2.1 A random walk model for legal research

The geometry we construct for the legal corpus is based on an encoding of the corpus as a multinetwork that supports the *legal search process*. We frame legal search in this setting as a probabilistic process of "local" exploration of the opinion corpus modeling the way in which a user of the legal corpus might navigate from opinion to opinion while researching an issue. This navigation is naturally viewed as a *Markov chain* (see e.g., Grinstead and Snell 1997), formulated as a matrix T of *transition probabilities* where the states are indexed by the opinions: given opinions a and b the value of the entry $T(a, b)$ is the probability of "moving to" opinion b "from" opinion a in an exploration of the legal corpus.[3] More precisely, framing this as a "random walk" in "opinion space" $T(a, b)$ is the probability of moving at the next step to case b, given that you are currently at case a, i.e., the *conditional probability*

$$T(a, b) = P(b|a),$$

in standard notation.

The transition probabilities are constructed as a combination of a several terms, reflecting our stylized model of navigation of the space of legal opinions.[4] We assume the possibility of three basic types of local exploration from an initial opinion a: (1) consideration of opinions cited by a; (2) consideration of opinions that cite to a, and (3) consideration of opinions that are *textually similar* to a. Our Markov chain (transition matrix) is thus represented as a linear combination of the individual chains, $T_{\text{cited-by}}, T_{\text{cited}}$, and T_{sim}.

We allow for the possibility that an exploratory mode (i.e., the weights given to the three forms of connection in the network) can vary for any given search. It may depend on the searcher, where he/she is in the overall research process, and the current opinion. The last of these is the easiest to embody and in this case the overall chain can be written as

---

[3] T varies over time as new opinions are introduced, but very slowly in comparison with the legal search process. Our use of the chain is with respect to the search that is accomplished at some instant in time, so we can assume the process is time homogenous and represented by a matrix.

[4] Other legal sources, including statutes and constitutions, have other types of internal ordering (such as organization by chapter or article) that may be relevant for law search. For purposes of this analysis, we restrict our application to the body of U.S. Supreme Court opinions and do not incorporate other sources of law. The framework of search that we develop, however, is generalizable to these other legal sources.

$$T(a,b) = p_{\text{cited}}(a)T_{\text{cited}}(a,b) + p_{\text{cited-by}}(a)T_{\text{cited-by}}(a,b)$$
$$+ \ p_{\text{sim}}(a)T_{\text{sim}}(a,b) \qquad (1)$$

with the proviso that $p_{\text{cited}}(a) \geq 0$, $p_{\text{cited-by}}(a) \geq 0$, $p_{\text{sim}}(a) \geq 0$, and

$$p_{\text{cited}}(a) + p_{\text{cited-by}}(a) + p_{\text{sim}}(a) = 1$$

at each state a. As per the notation, the weights may vary by initial state (a), though in what follows we will typically have them globally constant. In fact, for the sake of analysis we will assume these weights are uniform (each equal to $\frac{1}{3}$). Our implementation allows the weights to vary (cf. Sect. 3.2). In general, throughout this paper, we typically choose our parameters to be simple natural choices, reflecting the initiatory nature of this paper and the early stages of this project. Any particular parameter of groups of parameters could be optimized with more data and an appropriate training paradigm. Ideally, the weights would be determined by training them with respect to an appropriate objective function, and the ideal objective function would be related to the effectiveness of the exploration. This would require feedback from users, and in Sect. 3.2 we discuss an implementation which could eventually allow for such a training paradigm to be implemented.

2.2 Construction of the components $T_{\text{cited}}, T_{\text{cited-by}}$, and T_{sim}

The transition matrices T_{cited} and $T_{\text{cited-by}}$, based on the citation network are straightforward to construct. A natural and standard choice is to weight equally all opinions cited by a given opinion, and similarly for all opinions that cite the given opinion. Thus, if opinion a cites opinions $b_1, \ldots, b_k$ then $T_{\text{cited}}(a, b_i) = \frac{1}{k}$. Similarly, if a is cited by opinions $b_1, \ldots, b_k$, then $T_{\text{cited-by}}(a, b_i) = \frac{1}{k}$. While we choose to work with equal weights, this weighting could be modified in some way, perhaps accounting for some notion of the importance of an opinion. To find the citation network we make use of the excellent "Supreme Court Citation Network Data" database created by Fowler and Jeon (cf. Supreme Court 2015).

Navigation via textual similarity using something deeper than keywords is a novel contribution of this work and for this we make use of a *topic model*. A detailed description of topic modeling is beyond the scope of this paper, but a short description will suffice for the purposes of exposition. Very briefly, a topic—in the technical sense—is a probability distribution over a vocabulary. Topic modeling is the unsupervised derivation of a set of such distributions that represents a text corpus of *documents* (technically defined as a roughly contiguous set of words in the corpus, that is usually itself composed of larger portions of text—e.g., full opinions as opposed to the word blocks it comprises). Topics are defined according to a simple generative *bag-of-words model*[5] for the documents in the corpus: given a document, first a topic is chosen at random and then a word is chosen at random within the topic. The topics are then the best fit solution to the actual bag-of-words

---

[5] "Bag-of-words" means that the document is summarized as the probability (frequency) distribution of the words comprising it.

representation of the documents. Recalling that bag-of-words is essentially a representation of each document as a word distribution, the topic model derives the "atomic" probability distributions that express each document in the corpus as a mixture of such atoms. The wide applicability of topic models in many disciplines has made for a broad community of topic modelers and the topic modeling technology has quickly become an "off-the-shelf" technology ready for deployment (see e.g., MALLET 2015) with a minimum of start-up cost. See Blei (2012) for one of the many friendly explanations of topic modeling.

The only supervision in the basic topic modeling algorithm is the choice of number of topics to be computed. We choose to use 100 topics, which for our corpus of 21,893 opinions (documents) is adequate. The most widely discussed method for choosing the number of topics involves treating the number of topics as a model parameter and inferring it from the data (Griffiths and Steyvers 2004). This method requires, however, more computational resources than are typically available as resources needed increase rapidly with the number of topics allowed. With such a large corpus of (long) documents, for example, fitting a corpus with 1000 topics is not possible in a reasonable amount of time. The approach we adopt—and we think it reflects the current best practice—is to choose a maximum number of topics based on time and computational resources available. Picking a larger number of topics than the data supports is not a risk because the widely used specifications of the topic model [used by MALLET (2015) and in the software we use Buntine and Mishra (2014)] will simply leave them empty. For example, if the data suggest that 50 topic distributions is sufficient to account for the data, fitting a model with a maximum of 100 topics will recover the same model as fitting the model with a maximum of 50 topics.

When the topic modeling is completed we therefore have a set of topics $\text{Topic}_1, \ldots, \text{Topic}_{100}$, where each word w in the vocabulary has a weight in each topic $\text{Topic}_k(w) \geq 0$ and any given opinion a is represented as a distribution over topics, $\sum_k \alpha_k(a) \text{Topic}_k \left(\sum_k \alpha_k(a) = 1; \ \alpha_k(a) \geq 0 \right)$. Table 1 shows the most highly weighted words in five of the topics. The indexing of the topics in the table is not relevant. The labels (in parentheses) are assigned by the user (in this case the authors of this paper). The full set of topics for our SCOTUS dataset is available online.[6]

While there are a number of different kinds of topic models (see e.g., Blei 2007; Blei and Lafferty 2006; Roberts et al. 2013), the "latent Dirichlet allocation" (LDA) model (the "Dirichlet" refers to an underlying assumption of a Dirichlet distribution in the model) is perhaps the best known and most widely used (Blei et al. 2003). This is the topic model that we use here.

With the topic modeling accomplished, we are now in a position to construct T_{sim}. For this, we only consider as relevant to a given opinion the "top" topics and similarly for a given topic, only consider as relevant to our exploration those opinions who express it most strongly. More precisely, we fix integer parameters $N_{\mathcal{T}}$ and $N_{\mathcal{O}}$ and for a given opinion a identify the $N_{\mathcal{T}}$ most heavily weighted topics expressed in opinion a (using the $\alpha_k(a)$ to define the weight) and for a given topic

---

[6] http://www.bendingthelaw.com/topicdata.

Table 1 Some representative topics derived from the SCOTUS corpus

Topic$_1$ (jury process)	Topic$_2$ (housing)	Topic$_8$ (evidence)	Topic$_{58}$ (abortion)	Topic$_{59}$ (search)
Jury	Housing	Court	Abortion	Search
Trial	Lease	Case	State	Warrant
Evidence	Property	Evidence	Woman	Fourth
Defendant	Rent	Record	Medical	Amendment
Error	credit	Fact	Physician	Evidence
Verdict	Building	Question	Life	Arrest
Reasonable	Bond	Facts	Health	Police
Instruction	Tenant	Did	Roe	Cause
Doubt	real	Issue	Consent	Probable
Instructions	Rental	Findings	Statute	Seizure

Topic$_k$ identify the N_O opinions in which Topic$_k$ was most strongly expressed (using the α_k here as well).[7] Intuitively we view this as the process of a search returning the top N_T topics related to the initial opinion a followed by a search of the top N_O opinions associated to each of these top topics. To weight the final results of the search, for the given opinion a we create an $N_T \times N_O$ matrix in which the i, j entry is the index of the jth most significant opinion in the corpus for the ith most significant topic in opinion a. If we define $W_{a,b}$ to be the number of times opinion b occurs in this matrix, then T_{sim} is the random walk produced by normalizing according to these weights. More precisely, for any b with $W_{a,b} > 0$,

$$T_{\text{sim}}(a, b) = \frac{W_{a,b}}{\sum_c W_{a,c}}.$$

With this we have now defined each component random walk for our cumulative walk T.

2.3 The exploration geometry

The cumulative Markov chain

$$T = p_{\text{cited}} T_{\text{cited}} + p_{\text{cited-by}} T_{\text{cited-by}} + p_{\text{sim}} T_{\text{sim}}$$

is a natural generalization of the random walk (the "random surfer") whose equilibrium distribution is the source of the original PageRank algorithm underlying the early implementation of the Google search engine (Brin and Page 1998). Of interest to us is the geometry that this search model (or any random walk) produces. In

---

[7] The use of α_k can be justified for N_T by the interpretation $P(\text{Topic}_k|a) = \alpha_k$. While assuming that cases are equally relevant a priori, we have for a fixed Topic$_k$ that $P(a|\text{Topic}_k) = \frac{P(a)}{P(\text{Topic}_k)} P(\text{Topic}_k|a) \propto P(\text{Topic}_k|a) = \alpha_k$; so we can use α_k to order N_O as well.

particular, this kind of Markov-based search produces a metric on the network space that we call *PageDist*.[8] We call the induced geometry an *exploration geometry*.[9]

To define *PageDist* we attach one last parameter r to the random walk of (1): at each step assume a probability $r > 0$ of continuing the exploration. Then given r and starting at an opinion a, the expected number of visits to opinion b is

$$R(a,b) = \sum_{k=0}^{\infty} r^k T^k(a,b) \tag{2}$$

where as per usual, $T^k(a,b)$ is the probability of transition from a to b in k steps. Intuitively, $R(a,\cdot)$ forms an exploration neighborhood of opinion a in the sense that the higher the value of $R(a,b)$ the more opinion b is considered to be in a neighborhood of a. Notice, r governs the size of this neighborhood as a sort of radius. If $r = 0$ then the neighborhood consist of only the opinion a, while if $r = 1$ (and the chain is irreducible) then the series diverges everywhere and the whole space is a's exploration neighborhood. So we need a value between 0 and 1 and in what follows we chose $r = \frac{1}{2}$ to keep it simple. As discussed above, with a fixed objective function and enough training data one could could optimize this choice of r (perhaps even locally).

By comparing the overlap of the neighborhoods defined by different opinions we can form a metric *PageDist*, given by

$$\text{PageDist } (a,b) = ||R(a,\cdot) - R(b,\cdot)||_p \tag{3}$$

where p denotes the p-norm.[10] Notice that if the neighborhood description of a and b nearly agree then this will be near zero, and if they are very distant $R(a, x)$ will be nearly zero when $R(b, x)$ is large and vice versa, resulting in a large value of PageDist (a, b) (in other words, a large distance between the opinions). So the *PageDist* metric will capture a notion of distance within the landscape. Figure 1 shows the distribution of distances among our corpus of Supreme Court opinions. In what follows, we chose the Euclidean norm ($p = 2$) to keep it simple. Again, with a fixed objective function and enough training data the choice of p could also be optimized.

The random walk setting also makes possible a definition of *curvature* that encodes a level of difficulty for escape from a given point in the execution of a random walk. If the degree of difficulty is large, a walk will have a tendency to get "stuck" in the neighborhood of the state. This can be interpreted as an opinion that doesn't connect usefully with its surrounding or nearby opinions. Conversely, a more "fluid" area around an opinion suggests that it engages usefully with the

---

[8] We are indebted to Peter Doyle for early conversations regarding the geometrization of Markov chains and PageDist.

[9] It is worth noting that another natural candidate for a textual geometry is given in Leibon and Rockmore (2013) wherein the concept of a *network with directions* is introduced. Therein, "directions" function as "points at infinity", producing a hyperbolic metric on the network. For this—and any text corpus—the pure topics provide an obvious choice of direction.

[10] Recall that this notation means $\left(\sum_x |R(a,x) - R(b,x)|^p\right)^{1/p}$.
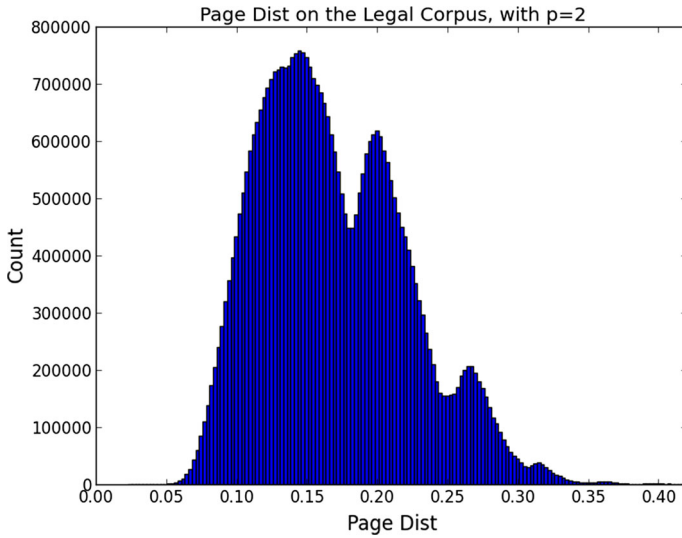
Fig. 1 Here we see a histogram of the PageDist values when computed on the legal corpus. Each sample is a pair of points (opinions) in the corpus and this histogram includes all distinct pairs of points. The modes in the histogram correspond to pairs in distinct regions. We choose $p = 2$, $r = \frac{1}{2}$, and $N_T = N_O = 10$

broader opinion landscape. This kind of idea will be key to understanding the *relevance* of an opinion.

This ability to "escape" from a region while random walking is a problem studied widely in a variety of mathematical contexts. We take inspiration from the study of Brownian motion on manifolds and from this define for the random walk on the network a local notion of *curvature* as

$$\kappa(a) = \log(R(a, a) - 1). \tag{4}$$

In "Appendix A" we explain from a technical point of view why this corresponds to scalar curvature from Riemannian geometry and thus provides a natural definition.

For us the key is that as the network evolves a measure of change in the local connectivity of the opinions can be expressed in terms of changing κ. We think of this change as a measure of how the network is *bending*. Let us make this precise. Given the node set N of a network with a transition matrix T reflecting a Markov process on the nodes, let $S \subset N$, be some subset of nodes. A Markov chain on N induces a chain on the subset S by using the weights

$$W_S(a, b) = P(a, b) + \sum_{k \in N \setminus S, \ a \neq b} P(a, k) P(k, b),$$

for $a, b \in S$. Note that we are simply lumping together into one term all transitions a to b that go outside of S. We form a new transition matrix $P(a, b; S, N)$ normalizing $W_S(a, b)$ so that the weights sum to one at each vertex. We call this the *induced local exploration*. This induces a corresponding exploration geometry and a

curvature κ (defined as in (3,4)) for S relative to N which we denote as $\kappa(a; S, N)$. Bending will encode the change in curvature as S grows.

Consider the network at two different time points $t_0 < t_1$ with corresponding node sets N_0 and N_1. Since the opinion corpus only grows in time, $N_0 \subseteq N_1$. Then we can quantify a change in the induced exploration geometry as

$$\text{Bending } (N_1, N_0)(a) = \kappa(a; N_0, N_1) - \kappa(a; N_0, N_0) \qquad (5)$$

where $\kappa(a; N_0, N_0) = \kappa(a)$ in the network at time t_0. Identifying the network with the timestamp we might also write

$$\text{Bending } (a; t_1 > t_0) = \kappa(a; t_1 > t_0) - \kappa(a; t_0). \qquad (6)$$

Bending is easy to interpret, it indicates whether the induced geometry at a point evolves in such a way that it became easier or more difficult to escape from the point. Regions where it becomes more difficult to make such transitions we call *puddling regions* and regions where it becomes easier are called *drainage regions*. A precise definition works with the distribution of bending values: we call the subset corresponding to the bottom quartile of Bending $(*; t_1, t_0)$ the *Drainage* region (relative to the defining era)—or *Drainage* (t_1, t_0). Similarly, we call the subset corresponding to the top quartile of Bending $(*; t_1, t_0)$ the *Puddling* region (relative to the defining era)—or *Puddling* (t_1, t_0). Figure 2 shows the distribution of $\kappa(*; 1990)$ as well as the bending of 1995 relative to 1990 in the Supreme Court opinion corpus (*Bending*$(*; 1995 > 1990)$).

3 Results

The metrics we have developed enable us to determine the "relevance" of an opinion, as defined by its proximity to new opinions that are added to the corpus.
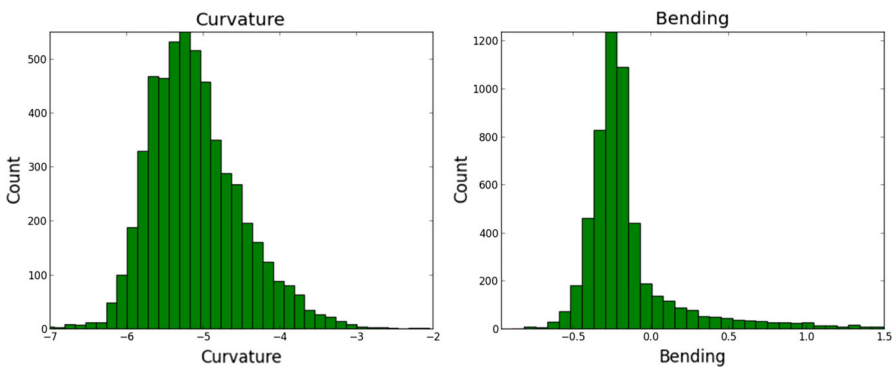
Fig. 2 On the left we see a histogram of the the curvature $\kappa(*; 1990)$ computed on the corpus at 1990, and on the right we see the bending Bending $(*; 1995 > 1990)$. This gives a sense of the variation of the curvature over time. Notice, the curvature histogram on the left is far from uniform. The bending histogram on the right is very telling. The right tail is the Puddling region and the left tail the Drainage region

3.1 Metrics for relevance

To exhibit the utility of our various definitions we first quantify what it means for a case to be "relevant". Our proxy is that it is nearby (in terms of PageDist). Thus, let N_t denote the set of nodes (opinions) in the network (corpus) at time t. Given $t_2 \geq t_1 \geq t_0$, define the *set of relevant cases* (at some threshold d) as

$$\text{Rel}_{t_2,t_1,t_0;d} = \{a \in N_{t_0} \mid \text{PageDist}(a,b) < d, \text{ for some } b \in N_{t_2} - N_{t_1}\}. \quad (7)$$

This set (with these parameter values) comprises the "early" opinions a at time t_0 (i.e., those that could serve as precedent) that find themselves close to newly arrived (later) opinions (those issued in the period between t_1 and t_2). This means that the opinions in $\text{Rel}_{t_2,t_1,t_0;d}$ are those opinions published no later than t_0 that are close to the new opinions published between times t_1 and t_2.

The threshold d can be set based on various criteria. A natural way to set d is by taking into account the PageDist distribution. A guiding principle is to set d according to the percentage of cases that we want to declare as "relevant" over a given initial or baseline period. For fixed time periods $t_0 < t_1 < t_2$, as the threshold d increases, so does the fraction of opinions in the corpus at time t_0 that are considered relevant. Conversely, as the fraction of cases that will be viewed as relevant grows, this implicitly corresponds to an increased threshold d.

We further define the *Initial Relevance Probability* (IRP) (for $t_1 > t_0$ and a given threshold d) as the fraction of opinions present at time t_0 that are in $\text{Rel}_{t_1,t_0,t_0;d}$— i.e., the fraction of opinions that remain relevant at time t_1 according to a threshold d. Our goal is to understand how to predict which cases remain relevant as time goes on. Figure 3 shows how IRP varies with relevance to future cases $P(\text{Rel}_{t_2,t_1,t_0;d} \mid \text{Rel}_{t_1,t_0,t_0;d})$.[11] Therein we plot (using $t_0 = 1990$, $t_1 = 1995$, and $t_2 = 2000$)

$$\text{Momentum} = P(\text{Rel}_{t_2,t_1,t_0;d} \mid \text{Rel}_{t_1,t_0,t_0;d}) - \text{IRP}$$

against IRP (recall that since d increases monotonically with IRP, we can view both axes as functions of d). Thus, "*Momentum*" measures the fraction of opinions that continue to be relevant. This behaves as might be expected, with an increasing percentage of opinions remaining relevant, until such a time as too many initial cases are tossed in, some of which will be opinions that have become vestigial.

Our goal is to identify the region R which contains the recent legal action. If we imagine that we have constructed a random region with each of our independent samples, then $P(\text{Rel}_{t_2,t_1,t_0;d} \mid \text{Rel}_{t_1,t_0,t_0;d}) \approx IRP$. So the *Momentum* measures how far beyond random our construction is, and we define the optimally "relevant" region as the one that's furthest beyond random. Let us now fix $d = d_{max}$ so as to correspond to the $IRP = 0.2$ in Fig. 3. With the choice of d set, we now have fixed the parameter by which we identify opinions as relevant. A mathematical justification for this choice can be found in "Appendix B".

[11] Note that the conditional notation has the usual interpretation of $P(A \mid B) = \#(A \cap B)/\#B$.
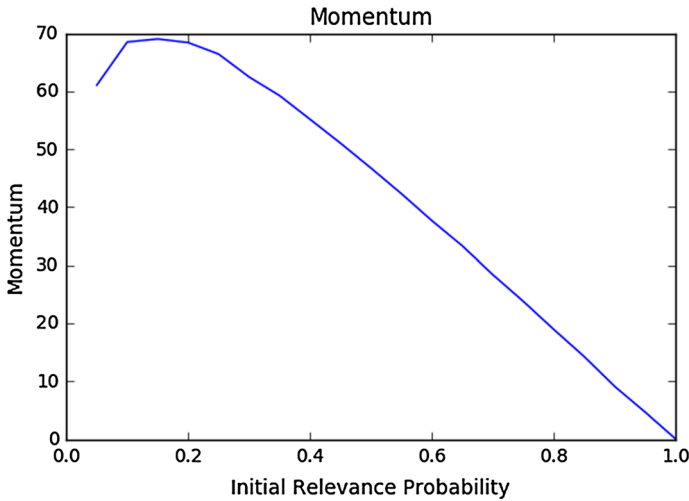
Fig. 3 Here the x-axis is Initial Relevance Probability relative to $t_0 = 1990$, $t_1 = 1995$, so the fraction of cases before 1990 that are within a distance d of cases that come in after 1990 and before 1995. As d increases so does IRP, so that the x-axis reflects a steady increase in d. Similarly, for the y-axis, *Momentum* (expressed as a percent) with $t_0 = 1990$, $t_1 = 1995$, and $t_2 = 2000$, which is also a function of d. So, the curve we see here is effectively a parametrized plot of *Momemtum* against IRP as d increases. Recall that *Momentum* (with these parameters) is the difference between the proportion of early (pre-1990) opinions that continue to be relevant in the 1995-2000 period, given that they were relevant in the 1990-1995 period, and the fraction of opinions that initially were relevant to opinions written between 1990 and 1995. Thus, we are subtracting out some baseline guess of how many of these early cases you would expect to be relevant in this time based on earlier information. This measures how much larger than random the future relevance is given recent relevance. This is all a function of d or equivalently, IRP. We see that IRP $= 0.2$ is roughly an optimal value

Having fixed d we can now examine the interaction between curvature and relevance, and in particular, the effect of being in either the drainage or puddling groups as respects the relevance of future cases. Let us start by defining our *Future Relevance Probability relative to a condition A* as

$$FRP(A) = P(\text{ Rel }_{t_2,t_1,t_0;d_{max}} \mid \text{ Rel }_{t_1,t_0,t_0;d_{max}} \& A). \tag{8}$$

This measures how much knowing a condition A helps to predict future relevance. And our goal is to see whether knowing something about the dynamic geometry, namely if we are in a drainage or puddling region, helps us predict whether that regions is more or less likely to be relevant in the near future. This entails the comparison of $FRP(\text{ Drainage })$, $FRP(\text{ Puddling })$, and $FRP(\text{ All })$.

This comparison is shown in Fig. 4. We see the relevance of future cases (the blue line - in the online - and solid line in the paper copy) compared to the relevance of future cases in the drainage and puddling regions. Therein we see that indeed, drainage regions (low bending) have roughly a greater than 10% chance more of being relevant for future cases than do puddling regions (high bending). That is, the drainage regions that are connecting up the space are more associated to future relevance.
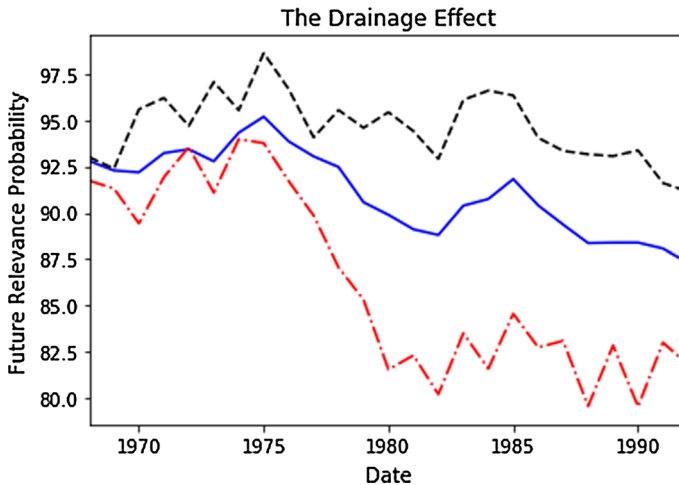
Fig. 4 Here the *x*-axis is the year the case was decided, and the *y*-axis is a probability expressed as a percent. The blue/solid curve is *FRP*(All) with $t_0 = date$, $t_1 = date + 5$, and $t_2 = date + 10$. In black/dashed we see *FRP*(Drainage) and in red dot-dashed we see *FRP*(Puddling) with the same timing parameter values. Notice that indeed, the bending is (negatively) correlated with long term relevance as predicted, and that after around 1978 we see a fairly stable 10% difference. (Color figure online)

To confirm that this relationship is not driven by random chance, let the null hypothesis be that there is nothing but a random difference between the drainage and puddling regions. So for a fixed measurement, under the null hypothesis there would be a fifty-fifty chance that we confirm our suspicion (technically, bounded by 50% when allowing for ties). Furthermore, for events that differ by at least 5 years, the $N_{t_2} \backslash N_{t_1}$ populations are distinct, so that the measurements are suitably independent. Thus, we have 6 independent measurements with a perfect track record which would be expected by chance with a likelihood of $\frac{1}{2^6}$. The null hypothesis that there is nothing but a random difference between drainage and puddling regions is thus highly unlikely.

3.2 Implementation

The ideas presented in this paper form the foundation of new web-based search tool for exploring a space of legal opinions using the exploration geometry introduced in the body of this paper. Specifically, we have built a prototype website and user interface (UI) that will enable the exploration according to PageDist of an opinion database, that ultimately will encompass all Federal Court and Supreme Court cases. At present it is running on a small subset (SC cases 1950–2001). This prototype can be found at www.bendingthelaw.org.

Currently, our UI introduces users to cases in the "vicinity" (in the sense of our exploration geometry) of a pre-identified case specified by the user. The anticipation is that these cases will be strong candidates for precedent-based reasoning. As per

([1](#)) the search returns the "neighborhood" of the case that depends on the database of cases as well as the individual weights assigned to the three-component random walk process encoding the exploration geometry—that is, a choice of weights $p_{\text{cited}}, p_{\text{cited-by}}$, and p_{sim}. As a first step we allow a choice of weights from $\{0, 1, 2\}$ with at least one positive weight, so that $W = w_{\text{cited}} + w_{\text{cited-by}} + w_{\text{sim}}$, $p_{\text{cited}} = w_{\text{cited}}/W$, $p_{\text{cited-by}} = w_{\text{cited-by}}/W$, and $p_{\text{sim}} = w_{\text{sim}}/W$.

Recall that the similarity piece of the random walk, T_{sim} requires that we construct the "topic by opinion" matrix of a given size. We choose that to be 10×10—i.e., that for any given topic we consider the 10 opinions that make the most use of it and conversely, for any opinion, we consider the 10 topics that make the strongest contribution to it.

Given an initial query, the UI provides two complementary representations: (1) a ranked list of geometrically closest (in terms of PageDist) cases and (2) a map of the space, centered on a case of origin (the original input). As a "map", this representation shows not only the relation of cases to the initial query, but also the relations of the closest cases to each other. The associated visual integrates a network representation wherein cases are linked if the overall weight between them exceeds a threshold. The map is generated by clicking on "View Case Network" (after executing the query). The opinion map produced from the query "329 US 187: Ballard v. United States" is shown in Fig. [5](#).
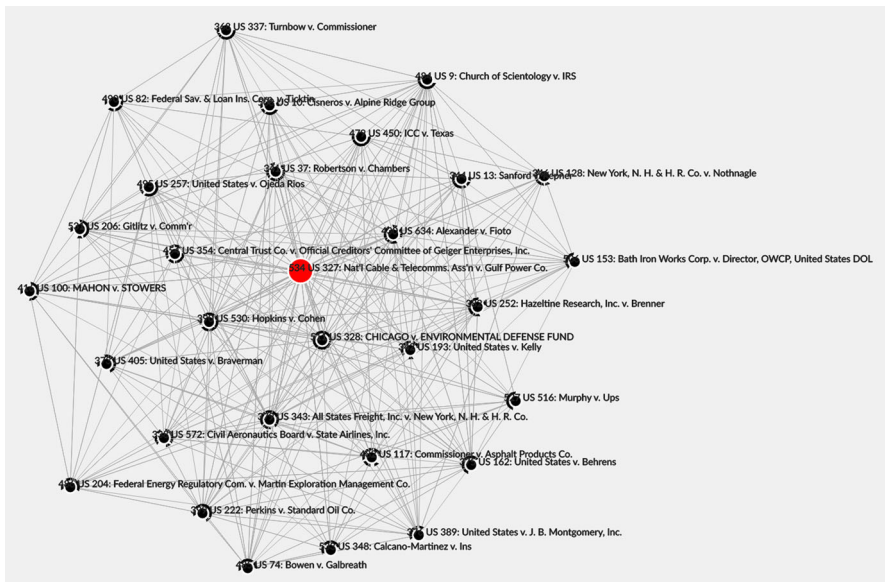
Fig. 5 Here is a snapshot from our alpha version UI for exploring the space of legal opinions. The current UI is built on the database of Supreme Court opinions over the time period 1950–2001. What we see here is the 2-d MDS visualization of the PageDist neighborhood of 30 cases closest to "329 US 187: Ballard v. United States". Cases are linked if the overall weight between them exceeds some threshold. The exploration weights have been set to 2 ("cited"), 1 ("cited by"), and 2 ("topic similarity")

4 Closing thoughts

In this paper we introduce a new multinetwork framework integrating citation and textual information for encoding relationships between a large set of Supreme Court opinions. The citation component derives from the underlying citation network of opinions. The textual piece derives from an LDA topic model computed from the text corpus. A metric on the opinion space is the reification of a basic model of legal search as would be executed by a prototypical legal researcher ("homo legalus") looking for cases relevant to some initial case through textual similarity and citation. The model of search is articulated as a Markov chain on the network, built as a linear combination of the individual chains on the citation and topic networks. The Markov process produces a notion of *distance* between opinions which can also be thought of as a proxy for relevance. Along with distance, the Markov chain gives rise to a notion of curvature, and with this an implicit framing of the opinion corpus as a "landscape" which we call "the legal landscape". We have implemented a first generation website that will allow users to explore a smallish subset of Supreme Court opinions using this search tool (www.bendingthelaw.org).

The text corpus evolves in the sense that cases enter the corpus regularly and in so doing continually transform the associated text landscape, changing interpoint distances and local curvatures. Of particular interest are those cases that remain relevant over long periods of time. Some regions of the legal landscape have the property that they serve as nexuses of connection for regions of the landscape. We show that those regions which over time become significantly more negatively curved are such connective areas. With the analogy of flow in mind, we call such areas, regions of "drainage". Areas which experience a significant increase in curvature we call "puddling regions". We show that drainage areas are more likely to contain continually relevant cases than the puddling regions. We further show that opinions that start off relevant, in the sense of entering the landscape highly relevant to many cases over a short period of time tend to remain relevant, thereby suggesting a property of (legal) *momentum*.

There are natural next steps to take with this idea. In one direction we will expand the text corpus to include all Supreme Court and Appellate Court Opinions. We also plan to validate and compare our model by asking users to compare the results of our search algorithm (under a range of parameter choices) with their own usual research approaches. Our newly introduced opinion distance function gives a new variable to explore the relations of opinions to all kinds of social and economic variables. It is also natural to export this model to other court systems that produce English language opinions. In this regard it would be interesting to see the ways in which the "bending" of the courts systems vary, and try to understand what might account for such (possible) variation. Ultimately, it would also be of interest to effect the integration of distinct corpora via this model. In a related, but different direction, we will deploy this new navigation and search model on other corpora. To this end, the *Bending the Law* website includes navigable access to the United States Code (USC), Code of Federal Regulations (FCR), and Internal Revenue Code (IRC). In these corpora, sections and subsections are linked and referenced, and the

topic modeling takes place on the level of sections. Future work will describe our findings in analyzing these newly multinetworked corpora, but for now, they exist as domains for new explorations for the public.

Appendix A: Scalar curvature, Riemannian geometry and the ease of escape

In Sect. 2.3, we suggest that a definition of curvature at a node in a Markov Chain should reflect the ease of escape from a point (specifically, as values range from small to large it should become more difficult to escape). In this appendix we justify this claim. The inspiration for our definition comes from differential geometry and for completeness, we give here a summary of the relevant technical facts and analogies. A standard reference for the various basic notions of differential geometry is the classic textbook of Helgason (2001).

Let (M, g) denote a Riemannian manifold of dimension d (and inner product g), and let $\tau_g(x)$ denote it's scalar curvature at a point $x \in M$. To describe the relationship, we first define a notion of *escape curvature* that is clearly the ease of escape from point on (M, g) and demonstrate its relationship to $\tau_g(x)$. We then examine *resolvent curvature*, which is the analog of our Markov chain definition of the ease of escape κ on (M, g), and prove it is related to $\tau_g(x)$ in the same way as the escape curvature.

Escape curvature. Let let X_t be the Brownian motion process on (M, g). And define the *escape time* of a path from a subset $A \subset M$ as

$$T_A = \inf\{t > 0 \mid X_t \notin A\}.$$

It is useful to think of the escape time of leaving a ball of radius $\sqrt{\delta}$ at x, and we denote this ball as $B_{\sqrt{\delta}}(x)$. In the analogy with a Markov chain, we are going to think of δ as the distance of a typical single ("infinitesimal") step in the discrete chain. We have the following theorem about the expected time for X_t to leave this ball:

Theorem 1 (see Pinsky 1984) *For small δ,*

$$E\left(T_{B_{\sqrt{\delta}}(x)}\right) = a_d\delta + b_d\tau_g(x)\delta^2 + O(\delta^3)$$

where a_d and b_d are constants that depend only on M's dimension d.

We define the *escape curvature* as

$$K(x, g, \delta) = \log\left(E\left(T_{B_{\sqrt{\delta}}(x)}\right)\right),$$

and using the fact that $\log(1 + \epsilon) = \epsilon + O(\epsilon^2)$ and $\log(ab) = \log(a) + \log(b)$, Theorem 1 implies:

Corollary 1 (Escape Bending) *For small δ,*

$$\frac{K(x, g_1, \delta) - K(x, g_0, \delta)}{\delta} = c_d(\tau_{g_1}(x) - \tau_{g_0}(x)) + O(\delta)$$

where c_d is a constant that depend only on M's dimension d.

It is worth pointing out the significance of taking the difference. For a Riemannian manifold we have a canonical flat space (Euclidean space) which we can use use to set a baseline for what it means to be curved. For a Markov chain there is no such obvious choice, but the bending, namely the difference in curvature, makes good sense without such a baseline. Perhaps more importantly there is not one Euclidean space but one for each dimension d, and dimension is a concept that also resists a canonical definition on a Markov chain. In fact, this formula encourages us to view a chain's dimension as variable (but locally stable at under reasonable changes in the metric); what we end up calling curvature is really a mixture of curvature and a factor that depends on dimension that is rather subtle to decouple.

Resolvent curvature. Our definition of a Markov chain's curvature κ defines an ease of escape from a point for a discrete chain. In this section, we see that κ has an analog on a Riemannian manifold that satisfies the relationship to sectional curvature described in Corollary 1. To do so, we acknowledge the dependence of $R(a, b)$ (see Equation (2)) with the notation and note that $R_r(a, b)$ satisfies

$$R_r(a, b) = R(a, b) = \sum_{k=0}^{\infty} r^k T^k(a, b) = (I - rT)^{-1}. \qquad (9)$$

Consider the last term, $(I - rT)^{-1}$. The operator $I - T$ is well known as the analog of the Laplacian or Laplace operator for a Markov chain (very generally, a "Laplacian" is the operator that takes the difference of function with the average values of its neighbors). Let $\Delta = (I - T)$. Plugging this into Eq. (9) we arrive at

$$R_r(a, b) = \frac{1}{r}\left(\left(\frac{r}{1 - r}\right) - \Delta\right)^{-1},$$

which now has both a Riemannian and Markov chain interpretation. In fact, it is a rescaled version of a very well studied operator, the *resolvent operator*. We denote its kernel in the Riemannian setting as $R_{r,g}(x, y)$ (which is continuous in dimensions 2 and 3, see Polterovich 2000). Furthermore r is governing the small distance from a point x, (i.e., a smoothed out ball) and in the Riemannian setting we express this as $r = r_0\delta$. We have

Theorem 2 (see Polterovich 2000) *In dimensions $d = 2$ and $d = 3$ (and in a formal sense for all dimensions) we have, for small δ,*

$$R_{r_0\delta,g}(x,x) = \frac{1}{r}\left(\frac{r}{1-r}\right)^{1-d/2}\left(A_d + \alpha_d r_0 \tau_g(x)\delta + O(\delta^2)\right)$$

where A_d and α_d are constants that depend only on M's dimension d.

So we define the *resolvent curvature* as

$$\kappa(x,g,r_0,\delta) = \log\left(R_{r_0\delta,g}(x,x) - 1\right)$$

on (M, g) as a direct analog of our κ, and we have:

Corollary 2 (Resolvent Bending) *In dimensions $d = 2$ and $d = 3$ (and in a formal sense in all dimensions) we have:*

$$\frac{\kappa(x,g_1,r_0,\delta) - \kappa(x,g_0,r_0,\delta)}{\delta} = r_0 C_d(\tau_1(x) - \tau_0(x)) + O(\delta)$$

were C_d is a constant that depend only on M's dimension d.

So it is indeed reasonable to call κ the Markov chain's curvature. As such, it is important that in Corollaries 1 and 2 dimension does not show up in the re-scaling of $\kappa(x,g_1,\delta) - \kappa(x,g_0,\delta)$ to be a finite value. In this appendix we see that the log is in fact required to do this. Notice, Corollary 2 is still true using any constants $A > 0$ and $B \geq -1$ for a definition of $\kappa = \log(AR(x,x) + B)$. Our choice of $A = 1$ and $B = -1$ is based on the notion that $R(a,a) \geq 1$ and when $R(a,a,) = 1$ the walker starting at x cannot return to x, and we could reasonably view x as having infinite negative curvature. Using $\kappa = \log(R(x,x,) - 1)$ makes this true and is particularly simple and easy to interpret.

Appendix B: Hunting for the relevant region with momentum

In this Appendix we justify why finding the argmax of the *Momentum* corresponds to finding the best approximation of the Relevant Region as implemented in Sect. 3.1. To do so, it's useful, as in Appendix A, to imagine there is a "true" (Platonic) legal space at time t which for simplicity we view as a compact Riemannian manifold (M, g) (where M is the underlying manifold and g is the metric). We view our historic cases as a sample of this M with our *PageDist* metric computed at time t as approximating the Riemannian metric at this time. We assume there is a relevant region in M around time t and denote this region as R. We view R as an open subset of M with smooth boundary. Furthermore, we view $N_{t_1} - N_{t_0}$ an $N_{t_2} - N_{t_1}$ as independent random samples of R with respect to the measure determined by the metric's volume form. Implicitly, this assumes the timescale for creation of the samples was small in comparison to the timescale in which the true, unknown, relevant region is changing. We will denote these independent samples as S_1 and S_2 respectively in what follows. Lastly, we normalize the the metric so that the total

volume of M is 1. As such, the volume of a region is its probability of being sampled when uniformly sampling the space.

Recall, given a choice of d we found all the points within d of a sample and denoted this region as $\text{Rel}_{t_2,t_1,t_0;d}$ for S_2 and $\text{Rel}_{t_1,t_0,t_0;d}$ for S_1. Then we used the fact that d determines the size of the region (the *IRP*) and vice-versa, to view this region as function of *IRP*. It is useful to express this implicit relationship explicitly and let the neighborhood of a sample of size *IRP* be denoted as $Rel(S, IRP)$, and similarly for the momentum we let

$$p(IRP) = P(Rel(S_2, IRP) \mid Rel(S_1, IRP)) - IRP.$$

to denote the momentum as a function of *IRP*.

Now our goal is to find an *IRP* so that $Rel(S, IRP)$ forms good approximations of the true, unknown R. So we first need to decide on what constitutes a "good" approximation. We choose to maximize the well known *Jaccard index* $J(Rel(S, IRP), R)$ where

$$J(A, B) = \frac{P(A \cap B)}{P(A \cup B)}.$$

We now justify the construction in Sect. 3.1 by observing:

Theorem 3 Given two independent dense samples S_k for $k \in \{1, 2\}$,

$$\operatorname{argmax}_{IRP}\{J(Rel(S_k, IRP), R)\} \approx \operatorname{argmax}_{IRP}\{p(IRP)\}.$$

Proof Our first order of business is to define what it means for a sample to be dense. We say a sample S is ϵ-dense for $\epsilon > 0$ if for every point in R the ball of radius ϵ around it contains a point in S. For a fixed ϵ, if we increase the size of the sample, then the probability that a random sample is ϵ-dense tends to one. So a sample being ϵ-dense is morally equivalent to being a large sample.

Viewing *IRP* as a function of d, we see by the triangle inequality that if S is ϵ-dense then $R \subset Rel(S, IRP(2\epsilon))$. Furthermore, we can estimate $P(Rel(S, IRP(2\epsilon)) \setminus R)$ by noticing that every point in $Rel(S, IRP(2\epsilon))$ is within 2ϵ of the of the closest point in the normal direction away from the (assumed smooth) boundary of our region, which we call the region's *collar*. Letting the A denote the surface area of the smooth boundary, we have $P(collar) = 2\epsilon A + O(\epsilon^2)$, so $P(IRP(2\epsilon)) = P(R) + 2\epsilon A + O(\epsilon^2)$. In particular, $P(IRP(2\epsilon))$ is arbitrarily close to $P(R)$ for an ϵ-dense set and small enough ϵ.

Armed with this observation about estimating the collar we find:

Lemma 1 *For a sufficiently dense sample,*

$$\operatorname{argmax}_{IRP}\{J(Rel(S, IRP), R)\} \approx P(R).$$

Proof For $IRP < P(R)$

$$J(Rel(S, IRP), R) \leq \frac{IRP}{P(R)} < 1,$$

and for $IRP > P(R)$ we have

$$J(Rel(S, IRP), R) \leq \frac{P(R)}{IRP} < 1.$$

So when the IRP is sufficiently far away from $P(R)$ we have $J(Rel(S, IRP), R) < 1$, and for a dense enough sample we have $R \subset Rel(S, IRP(\epsilon))$ and so by our collar estimate

$$J(Rel(S, P(R)), R) = \frac{P(R)}{P(R) + 2\epsilon S + O(\epsilon^2)}$$

assuring us that $J(Rel(S, P(R)), R) \approx 1$, as needed in order to identify $P(R)$ as the argmax of $J(Rel(S, IRP))$. □

From this Lemma 1 if we knew R, then we would set $IRP = P(R)$. But we do not know R. We do however have two independent samples and the following lemma:

Lemma 2 *For a dense sample,*

$$\mathrm{argmax}_{IRP}\{p(IRP)\} \approx P(R).$$

Proof By the above observation regarding collars, if $IRP < P(R)$ then $Rel(S, IRP)$ does not cover R and so $d < 2\epsilon$. Hence $P(Rel(S, IRP) \setminus R) < 2\epsilon S + O(\epsilon^2)$ and the region outside R is small for small ϵ. So $Rel(S, IRP)$ can be viewed as a random subset of R taking up $\approx \frac{IRP}{P(R)}$ worth of R. This is true of any independently specified subset of R, so for our independent samples S_1 and S_2 we have

$$P(Rel(S_2, IRP)|Rel(S_1, IRP)) \approx \frac{IRP}{P(R)}$$

and

$$p(IRP) \approx IRP\left(\frac{1}{P(R)} - 1\right).$$

Now for a dense sample S and $IRP > P(R)$ if we let d_k be such that $IRP = IRP(d_k)$ for S_k, then by the triangle inequality $Rel(S_2, IRP(d_1 - \epsilon)) \subset Rel(S_1, IRP(d_1)) \subset Rel(S_2, IRP(d_1 + \epsilon))$ assuring us that $(d_1 - 2\epsilon) < d_2 < (d_1 + 2\epsilon)$. So, as $P(Rel(S, IRP(d)))$ is continuous in d, we have $P(Rel(S_2, IRP(d_2)) \setminus Rel(S_1, IRP(d_1)) = O(\epsilon)$; and, as both approximations contain R,

$$P(Rel(S_2, IRP)|Rel(S_1, IRP)) = 1 + O(\epsilon)$$

telling us that

$$p(IRP) \approx 1 - IRP.$$

Together these two approximations tell us that, up to $O(\epsilon)$, $p(IRP)$ increases as IRP increases from 0 to $P(R)$, and $p(IRP)$ decreases as IRP increases from $P(R)$ to 1; so $\operatorname{argmax}_{IRP}\{p(IRP)\} \approx P(R)$ as required. □

Taken together Lemmas 1 and 2 imply Theorem 3. □

References

Barigozzi M, Fagiolo G, Mangioni G (2011) Identifying the community structure of the international-trade multi-network. Phys A 390(11):2051–2066

Berring RC (1986) Full-text databases and legal research: backing into the future. Berkeley Technol Law J 1:27

Berring RC (1987) Legal research and legal concepts: where form molds substance. Cal Law Rev 75:15

Blei DM, Lafferty JD (2006) Dynamic topic models. In: Proceedings of the 23rd international conference on machine learning, ICML '06. ACM, New York, pp 113–120

Blei DM (2012) Probabilistic topic models. Commun ACM 55(4):77–84

Blei D, Lafferty J (2007) A correlated topic model of Science. Ann Appl Stat 1(1):17–35

Blei D, Ng A, Jordan M (2003) Latent dirichlet allocation. J Mach Learn Res 3:993–1022

Blinov ML, Udyavar A, Yarbrough W, Wang J, Estrada L, Quaranta V (2012) Multi-network modeling of cancer cell states. Biophys J 102(3):22a

Bommarito MJ, Katz DM, Zelner J (2009) Law as a seamless web? Comparison of various network representations of the United States Supreme Court corpus (1791–2005). In: Proceedings of the 12th international conference on artificial intelligence and law (ICAIL 2009), pp 234–235

Brin S, Page L (1998) The anatomy of a large-scale hypertextual web search engine. In: Crouch M, Lindsey T (eds) Computer networks and ISDN systems. Elsevier, Amsterdam, pp 107–117

Bryan K, Leise T (2006) The $25,000,000,000 eigenvector: the linear algebra behind Google. SIAM Rev 48(3):569–581

Buntine WL, Mishra S (2014) Experiments with non-parametric topic models. In: Proceedings of the 20th ACM SIGKDD international conference on knowledge discovery and data mining. ACM, pp 881–890

Chung F, Yau ST (1996) Logarithmic Harnack inequalities. Math Res Lett 3:793–812

Fowler JH, Jeon S (2008) The authority of Supreme Court precedent. Soc Netw 30:16–30

Fowler JH, Johnson TR, Spriggs FJ, Jeon S, Wahlbeck P (2007) Network analysis and the law: measuring the legal importance of Supreme Court precedents. Polit Anal 15(3):324–346

Fronk CR (2010) The cost of judicial citation: an empirical investigation of citation practices in the federal appellate courts. Univ Ill J Law Technol Policy 2010(1):5825–5829

Garoupa N, Morriss AP (2012) The fable of the codes: the efficiency of the common law, legal origins and codification movements. Univ Ill Law Rev 5:1443

George CP, Puri S, Wang DZ, Wilson J, Hamilton W (2014) Smart electronic legal discovery via topic modeling. In: Proceedings of the 27th international FLAIRS conference, pp 327–332

Griffiths TL, Steyvers M (2004) Finding scientific topics. Proc Natl Acad Sci 101(Suppl. 1):5228–5235

Grinstead CM, Snell JL (1997) Introduction to probability. American Mathematical Society, Providence

Hanson FA, Allan F (2002) From key numbers to keywords: how automation has transformed the law. Law Libr J 94:563

Helgason S (2001) Differential geometry, lie groups, and symmetric spaces (graduate studies in mathematics). American Mathematical Society, Providence

Hellyer P (2005) Assessing the influence of computer-assisted legal research: a study of California Supreme Court opinions. Law Libr J 97:285

Katsh E (1993) Law in a digital world: computer networks and cyberspace. Vill Law Rev 38:403

Kivelä M, Arenas A, Barthelemy M, Gleeson JP, Moreno Y, Porter MA (2014) Multilayer networks. J Complex Netw 2(3):203–271

Leibon G, Rockmore DN (2013) Orienteering in knowledge spaces: the hyperbolic geometry of wikipedia mathematics. PLoS ONE. https://doi.org/10.1371/journal.pone.0067508

Lin Y, Yau ST (2010) Ricci curvature and eigenvalue estimate on locally finite graphs. Math Res Lett 17:345–358

Livermore M, Riddell A, Rockmore D (2017) The Supreme Court and the judicial genre. Arizona Law Rev 59:837

MALLET. http://mallet.cs.umass.edu/topics.php. Accessed Jan 2015

McGinnis JO, Wasick S (2015) Law's algorithm. Fla Law Rev 66:991

Nardi DJ, Moe L (2014) Understanding the Myanmar Supreme Court's docket. In: Crouch M, Lindsey T (eds) Law, Society and Transition in Myanmar. Hart Publishing

Ollivier Y (2009) Ricci curvature of Markov chains on metric spaces. J Funct Anal 256:810–864

Pinsky MA (1984) Brownian motion, exit times and stochastic Riemannian geometry. Math Comput Simul 26(4):357–360

Polterovich I (2000) A commutator method for computation of heat invariants. Indag Math 11:139–149

Rice D (2012) Measuring the issue content of Supreme Court opinions through probabilistic topic models. In: Presentation at the 2012 Midwest Political Science Association Conference. Illinois, Chicago

Roberts M, Stewart B, Tingley D, Airoldi EM (2013) The structural topic model and applied social science. In: Advances in neural information processing systems workshop on topic models: computation, application, and evaluation

Schauer F, Wise VJ (2000) Nonlegal information and the delegalization of law. J Legal Stud 29:495–515

Supreme Court Citation Network Data. http://jhfowler.ucsd.edu/judicial.htm. Accessed Jan 2015

Uzzi B, Mukherjee S, Stringer M, Jones B (2013) Atypical combinations and scientific impact. Science 342(6157):468–472

West JB (1909) Multiplicity of reports 2. Law Libr J 4