Reference Networks and Civil Codes

Adam B. Badawi and Giuseppe Dari-Mattiacci*

Abstract

A primary goal in the study of comparative law is to understand the extent to which legal systems borrow from other legal systems. Existing methodologies to determine these legal origins have substantial limitations. One approach has been to study the texts, practices, and institutions of a select number of countries in a qualitative way. While this technique can determine the similarities and differences of legal systems in a thorough manner, its deep and qualitative nature limits the number of comparisons that can be done. An alternative approach is to use a small number of blunt indicators to determine legal origin. Common options include whether a country uses a civil law or common law system and variables for the colonizing country. This approach increases the number of countries that can be compared, but does so in a way that loses information about the differences between individual civil and common law systems. In this chapter, we describe a computational approach that

*Badawi: UC Berkeley, School of Law, abadawi@berkeley.edu. Dari-Mattiacci: University of Amsterdam, gdarimat@uva.nl. We are grateful for the support of the National Science Foundation and the Netherlands Organisation for Scientific Research (NWO) through the Digging into Data initiative and for helpful comments from participants at the 2017 Computational Study of Law working group at the Santa Fe Institute. The results presented here are part of a larger project with Rens Bod, James Daily, Tom Ginsburg, Bart Karstens, and Marijn Koolen.

uses the network structure of legal codes to assess the shared features of legal systems. This technique can increase the amount of detail used in comparisons in a way that can scale. We apply this approach to the civil codes of several major European countries to show the promise–as well as the potential limitations–of using computational techniques to assess legal commonality.

1 Introduction

How can scholars assess the differences between legal systems? A primary way for legal comparativists to answer this question has been to sort legal systems into different "families." Some distinctions are quite clear and uncontroversial, such as the fundamental difference between common law systems (which are largely made up of former English colonies) and civil law systems. And there is general agreement that the French and German traditions are the primary influences in civil law. But things get murky quickly. While colonization played a central role in shaping the legal systems of many countries, many of these nations have sought to distinguish themselves after independence. To further confuse matters, some countries that were not subject to colonization made independent choices about their legal systems by borrowing from multiple traditions.

The diversity of legal systems has made it difficult for researchers to reach consensus on the appropriate grouping of legal families. For traditional legal comparativists, who generally use qualitative methods to categorize legal systems, there are debates about whether comparisons should be made at a macro-level or micro-level. Moreover, the time-consuming nature of the qualitative inquiries means that it is difficult to compare a large number of legal systems. Economists have also relied on groupings of legal systems in the "legal origins" literature. This line of research seeks to attribute development to the use of certain legal systems. Doing so requires categorizing legal systems, which has generally been done by the inclusion of relatively blunt variables such as the use of common law or civil law and indicators for the colonizing country. While

these simple distinctions allow for a large number of comparisons, there have significant challenges to this work based on what the drivers of this relationship are, what the essential differences are, and how to properly classify countries.

In this chapter we propose a method that can capture some of the subtlety and detail that distinguishes legal systems in a way that can scale. This approach uses the digitization and machine reading of legal codes to determine the network structure of these bodies of text. The network structure depends on the use of cross-references contained within the code. We then propose to compare legal systems by using a series of metrics that analyze the similarity of network structures between codes. These comparisons can be used to validate existing categorizations–such as the proposition that the legal systems of countries subject to the same colonizer are more similar than those of countries subject to a different colonizer. This method can also allow more nuanced comparisons between countries. Rather than the discrete measures used in the legal origins literature, comparisons that use network structure can use a continuous measure to assess the similarity between legal systems. Measures of this sort should allow more refined analysis of the legal origins hypotheses and also hold the promise of reorienting the conventional conceptions of which legal systems are related to each other.

At the end of the chapter, we apply these methods to several major European countries that use the civil law system. We show substantial evidence that the network structures map on to existing understandings of legal similarity. We also uncover some additional detail in the network structures that suggests some previously unidentified ways that these civil codes differ.

This chapter proceeds as follows. Part II briefly reviews the traditional and empirical approaches to grouping legal families and outlines the limitations that each faces. Part III provides an overview of the general structure of legal codes and constitutions, the use of cross references, and the preservation of these cross references as legal systems evolve

and get transmitted to new environments. Part IV describes how cross-references can be used to characterize the structure of legal codes and discusses different metrics that can be used to compare those structures between codes and constitutions. Part V concludes.

2 Existing Comparative Approaches

This section provides a brief overview of the techniques used by comparative scholars and the limitations of those techniques. The two primary approaches are orthodox comparative law and empirical comparative law. The latter group includes empirical study of comparative constitutions and scholars of law and finance who focus on the association between legal origins and economic outcomes.

Traditional Comparative Law

Traditional approaches to comparative law have a long and methodologically varied history. A significant amount of this research attempts to describe different legal systems and categorize them. The leading work in this area was performed by René David, who sought to classify legal systems into families (Gordley 2006). David divided these systems on the basis of the ability of a lawyer to move between common systems and the existence of shared philosophical and political principles (David and Brierley 1978). David saw three major classifications: Romano-Germanic, common law, and socialist. Among the most persistent classifications in this vein were the seven families identified by Arminjon, Nolde, and Wolff (1951): French, German, Scandinavian, English, Russian, Islamic, and Hindu.

Because this branch of orthodox comparative law bears most directly on this project, we discuss these efforts in some detail. Before doing so, however, it worth noting that–as with many areas of comparative law–there is little agreement on the appropriate methodology for classifying legal systems. As one scholar explains, "It has been said that there

are as many classifications as there are comparatists, and the number and variety of classifications is itself an indication of the failure of the enterprise."(Glenn 2006, p. 437). Nevertheless, there are some commonalities to classification attempts. One is the primacy of the distinction between common law and civil law, which differ substantially in their methodologies. The common law evolves through case-by-case distinctions made by judges while civil law relies largely on existing codifications. Within the civil law tradition, the French Napoleonic Code and the German Civil Code have had broad ranging historical influence. The Napoleonic Code came out of the French Revolution and was drafted in a relatively short time. While the drafters were legal technocrats rather than revolutionary firebrands, a defining characteristic of the Code is its clarity. Legal historians attribute this clarity to the conscious goal of making the code intelligible to lay people. Commentators describe the Napoleonic Code as relatively sparse–especially in comparison to the German Civil Code–and also note its "clear and memorable phrases" and "the absence of cross-references and jargon." (Kötz and Zweigert 1998, p. 91) The German Civil Code, BGB is a stark contrast in style to the French Code. Drafted at a time of relative political stability, there was no apparent attempt to make the BGB intelligible to lay people. The audience was instead trained lawyers. Stylistically, it has been described as "legal filigree work of extraordinary precision." (Isele 1949, p. 6) And one in which "repetitions are avoided by means of cross-references to amplifying sections." (Kötz and Zweigert 1998, p. 145).

Both of these codes have had significant influence in Europe.[1] The Italian Civil Code, which was drafted prior to the BGB, took significant inspiration from the French Civil Code. But as time progressed, Italian lawyers began to favor the German approach to legal reasoning and that choice was expressed itself as the Italian Code developed in . Accord-

1. The remaining review summarizes the analysis of Kötz and Zweigert (1998).

ing to comparativists, The French imprint on Spanish law has been more lasting. Like the Italian Code, the Spanish Civil Code was developed in the period between the enactment of the Napoleonic Code and the BGB. Perhaps for that reason, Spanish law drew substantially from the French model and many of those portions have remained in force.

Some European countries have developed legal systems that are quite distinct from the French and German traditions. Austria enacted its own civil code around the same time period as the promulgation of the Napoleonic Code. Perhaps because it was not borne of a revolution, the Austrian Code does not reflect Enlightenment principles as much of the Napoleonic Code. Some comparativists have suggested that this relative lack of modernity in the Austrian Code is one reason why it has not had has wide reaching the influence that the German and French Codes have had.

This brief review is far from comprehensive, but it does give a sense of how comparativists have grouped Continental legal systems. The qualitative nature of the comparisons generally means that there is little technical precision to these categorizations. It also makes it difficult to provide a ranked list of how similar one legal system is to all the other European civil law systems. To the degree this is even possible, disagreements about methodology and the importance of different features would make it difficult to harmonize lists across countries.

Empirical Comparative Law

The genesis of the recent work in empirical comparative law is a series of papers in law and finance that developed the "legal origins" hypothesis. These papers focused on the presence of investor and creditor-protections across countries and found positive correlations with these laws and economic outcomes (La Porta et al. 1998, 2000; La Porta et al. 2002). Subsequent work in the area moved beyond the corporate sphere to hypothesize the better economic outcomes in common law countries may be a

product the common law's lack of reliance on statutes and stronger protections for an independent judiciary relative to civil law systems Glaeser and Shleifer (2002). Much empirical work followed and that research sought to examine the connection between legal systems and economic measures.

Many of these studies purport to show an association between formalist and less interventionist legal systems and better economic outcomes. To pick just a few examples, Djankov et al. (2003) uses thorough survey evidence from law firms around the world to develop evidence on the speed and efficiency of evictions. The authors find that civil law countries tend to be slower with these processes than common law countries. In Djankov et al. (2008) the authors conduct a similar study that focuses on the legal restrictions on self-dealing. The study again finds that common law countries appear to have systematically better outcomes than civil law countries. There is a relatively wide ranging literature that uses these techniques and much of the literature finds similar associations with positive outcomes and the use of common law (Porta, Lopez-de-Silanes, and Shleifer 2008).

Studies of this sort have been subject to quite intense criticism. Some of these critiques focus on the difficulty of distinguishing between a country's use of a particular legal system and the identity of the colonizing country (Klerman et al. 2011). Others argue that modern political economy concerns are a more effective explanation for different economic outcomes than legal origins are (Roe 2006). But the critique that is most relevant to our effort to identify legal similarity through network structures is that the measures of legal systems and families are too blunt and inaccurate to be effective.

Spamann is perhaps the most persistent and effective methodological critic of empirical comparative law. In a series of articles he has pointed out both the theoretical and applied problems posed by associating economic outcomes with legal systems. A complete review of these criticisms is beyond the scope of this chapter, but he has taken particular

issue with the difficulties of coding comparative legal systems and with survey methodology. With respect to the former, he has recoded the data in prominent law and finance studies and has found quite different results (Spamann 2009b). With regard to the latter, he has argued that survey evidence can be unreliable when there are not detailed protocols for eliciting and coding information (Spamann 2015). The method that we propose would avoid these issues by using objective metrics to measure the differences between codes.

3 Cross-References and Networks

Both the traditional and empirical approaches to comparative law have drawbacks when it comes to the quality of the comparisons. The traditional approach lacks consensus on the appropriate way to compare legal systems and the comparisons that exist tended to be qualitative ones based on the history of each country's legal system. The empirical work tends to use a small number of binary measures to assess how legal systems vary. In what follows we describe a method that can surmount some of these difficulties. This method was developed in Badawi et al. (2018). This technique uses cross-references within legal codes to characterize the network structure of these legal codes. We can then draw on techniques in the analysis of networks to describe the similarities and differences between these different legal codes. This section begins with a discussion of the role of cross references in legal codes and shows how these references can be depicted as a network of directed links. These networks can then be summarized based on a variety of network statistics. The section concludes with an applied discussion of how to compare networks and, specifically, how to compare network representations of legal codes.

Legal Cross-References and Networks

The use of cross-references in legal codes has a very long history. The Code of Hammurabi, the oldest known written legal code, makes use of a

cross-reference in Law 182, which expressly incorporates a concept from Law 181. This use of a cross-reference–to refer to a definition or idea from a separate section of the legal code–is perhaps the most common use of a cross reference. Codes will also use cross references to point readers to similar concepts that appear elsewhere in the code or to carve out exceptions to a rule. As the discussion on traditional comparative law implied, legal systems have a different degree of affinity for the use of internal links. Legal systems that write codes for attorneys and others with technical legal proficiency are likely to make more use of cross references than those that strive to make legal codes more accessible to non-lawyers. Cross references may also be appealing where a legal system seeks to have cohesion across parts or even the entire code. For example, the definition of a particular word or concept could be used at one point and incorporated by cross-reference throughout other sections of the code.

The cross-references in legal codes can be characterized as a network. This network is composed of the cross-referenced and cross referencing sections, called nodes, and links between these nodes, which are the cross-references themselves. These codes can be depicted as directed networks. The networks are directed because each link in the network has a direction attached to it. In a legal code that uses cross-references, we characterize the direction of the link as going from the cross-referencing section to the cross-referenced section.[2] The complete representation of the code is then all the nodes that either contain a cross reference or are cross referenced and the links that represent the direction of the cross-references.

Network science has developed a wide-ranging suite of tools to analyze the features of networks. This section will begin by looking at some quite basic statistics and then explain some of the more complex tools

2. In network notation, a directed network can be represented as an $n \times n$ adjacency matrix where n is the number of nodes, a 1 indicates that there is a link that goes from node i to node j, and a zero indicates that there is no such link.

that researchers can use to characterize networks. The first basic statistic is network density, which is the number of actual links in a network divided by the number of potential nodes. For a directed graph, the network density is $d = \frac{m}{n(n-1)}$, where m is the number of actual links and n is the number of nodes (the term $n(n-1)$ provides the number of potential links in a network). If every section of a code provided a cross-reference to every other section of the code, the network density would be one, while a code with no cross references would have density zero. In practice, even legal codes that make wide ranging use of cross references will only use a few per section (or node). That means that even those codes that make relatively substantial use of cross references will have quite low densities. The second measure of network complexity, average degree, has a close relationship with network density. A degree is either an incoming or outgoing link from a node. The average degree captures the mean degrees per node in the network and is calculated as $\frac{2m}{n}$. Because it uses the same variables as network density, it conveys similar information in a different way. For legal codes it gives a sense of how often cross references tend to be used across the code.

We characterize legal codes as directed networks within incoming and outgoing links. This structure allows us to capture the amount of influence that specific nodes have. Depending on the context of the network, nodes that have many incoming links or outgoing links may be considered particularly influential. Indegree is the term for incoming links to a node and outdegree is the term for outgoing links to a node. Average indegree and outdegree are not informative because they will be equal to each other, but the variance of these measures does tell us something about the relative concentration of these types of links. If indegree (outdegree) has low variance that means that incoming (outgoing) links are relatively evenly distributed across the network. But if the variance is large, that indicates that certain nodes have substantially more incoming (outgoing) links relative to the others. In the context of legal codes, it is of particular interest when indegree and outdegree have high vari-

ance. A high indegree variance indicates that there are certain concepts and definitions in the code that are being cross-referenced at a high rate. That may mean that certain concepts are used in a relatively uniform way throughout the code. When outdegree has a high variance that suggests that some sections of the code referring to a large number of other sections. This may mean that these sections are particularly well-integrated into other sections of the code.

The size of hubs and authorities is another way to understand the degree of connectedness in a directed network. A hub is a node that has a substantial number of outgoing links and an authority is a node that has a sizable number of incoming links. Statistics such as hub5 and hub10 indicate the number of nodes in the network have at least 5 and 10 outgoing links, respectively. Authority5 and Authority10 capture how many nodes have at least 5 and 10 incoming links respectively.

The concept of a connected component provides an indication of how cohesive a network is. In a directed network, one tends to focus on weakly connected components, which ignore the direction of any links. A component is weakly connected if there is at least one undirected path from each distinct node to all other distinct nodes. Imagine, for example, that a code has five sections numbered one through five. Further imagine that there are links from section 1 to section 2, from section 2 to section 3, and from section 4 to section 5. This network has two weakly connected components (the first is formed by section 1, 2, and 3 and the second is made up of section 4 and 5). There is a path from 1 to 2, from 2 to 3, and from 1 to 3, but there is no path from sections 1, 2, and 3 to sections 4 and 5. A network with a small number of weakly connected components will appear more cohesive than one with a large number of weakly connected components.

Transplants, Transmission, and Network Similarity

Much of the work of comparative law is tracing the origin of different legal systems. For all but a few legal systems, those origins can be traced

back to another legal system. There are at least two potential ways that legal influence of this sort can be traced. The first is replication through express copying. Empirical scholars have been able to track this phenomenon by identifying express citations to other legal systems as well as the copying of statutes from other legal systems (Spamann 2009a). A second way that influence might manifest itself is through legal style. Instead of, or perhaps in addition to, outright copying, legal systems might find the philosophical and rhetorical approaches of another legal system appealing. Examining the network structures of a code based on cross-references should be able to capture both of these types of influence. With respect to the first type of copying, if a legal system were to copy that of another one verbatim, including the cross-references, the network structure of the two should be identical. But even if the copying were not verbatim, one would expect to see substantial similarity. For example, the BGB begins with a set of general principles and refers back to those principles throughout the code. A code wishing to follow this model could be expected to have a substantial number of cross references even if the translation were far from perfect. Specifically, one would expect to see many cross-references to the statement of general principles in the transplanted code.

Some of the empirical work that seeks to document copying has identified transmission of this sort.[3] The analysis of network structure can add to this analysis and capture some previously unavailable nuance. But it is in understanding common drafting and philosophical styles where the network structure can capture much more detail than has previously been accessible. Legal codes differ in their technical complexity and de-

3. Though it is not the focus of this chapter, understanding the potential channels of legal transmission may help validate comparisons of network structure. Beyond the obvious channel of colonization, Spamann identifies the ways that legal content and culture can continue to be transmitted. These include development aid and advice, shared commercial trade, and the migration of graduate students between countries.

sire for cohesion. As discussed above, the French Code sought to be accessible to non-lawyers and as a consequence includes relatively few cross-references. The BGB, alternatively, is a technical lawyer's project and consequently makes wide ranging use of cross references. This preference for linking to other sections of the code should not only be apparent in the BGB, it should also be evident in codes that draw substantial inspiration from the BGB. This sort of influence may not manifest itself in detectable copying but it may be evident in comparing the network structures of two codes. Codes that draw on the French model, which tends to minimize the use of cross-references should share a similar lack of network connectivity.

The next section will examine the networks created by cross references in the BGB and the Napoleonic Codes and compare them to codes that comparativists claim have been influenced by the German or French model. But before doing so it is worth identifying some of the potential problems with detecting legal transplants through network metrics. The first is that it is not always clear when the influence process begins, when it stops, and when it comes from different sources. In some cases–such as the distinct and deliberate imprint that the BGB left on the drafting of the Japanese Civil Code in the late 19th century–the initial influence is obvious. But in others it is not. The influence could come from a gradual shift in the views legal scholars and practitioners about which codes should be used as models. An example here would be Italy, which was initially heavily influenced by the French Code before the German model gained more cachet in the Italian legal community. To compound the problem, countries may begin to develop their own legal styles and those approaches may deviate from earlier influences. These potentially confounding effects warrant caution about drawing too strong of an inference from similarities in network statistics.

A second caveat involves the evolution of codes over time. An ideal dataset would include original versions of a civil code along with every update to that code. Unfortunately, there is very little in the way of dig-

itized histories of civil codes. In most cases, countries have only recently digitized their codes and there are a few systems in place that can accurately track amendments to a code. As a consequence, we are only able to compile network data for current or nearly current versions of the code. This feature can complicate comparisons and especially so in cases where the influence occurred a long time ago. Imagine, for example, that the French code exerted a fair degree of influence on another country's code during the late nineteenth century. If we have the modern day codes for both of these countries, it is fair to assume that both the French code and the code of the influenced country have evolved substantially in one hundred plus years since the influence was exerted. To the degree that stylistic and structural approaches persist, the influence of the French code may still be reflected in the modern codes. But if there has been a shift–especially one that relates to cross-referencing practice–it may be difficult to detect those changes. We discuss the results of our analysis with both of these concerns in mind.

4 A View of Civil Codes

This section will apply the method set out above to a selection of European civil codes that includes Austria, France, Germany, Italy, and Spain. The first section will explain the data gathering process and the section will discuss and compare the network structures of these different codes.

Methodology

Determining the network structure of a civil code first requires a parsable digital copy of the code. Many modern nations provide these documents on government web sites, although the formatting of some of them can complicate parsing the codes. The first step in parsing is a Python program that converts the civil code into a single XML (eXtensible Markup

Language) file. This XML file preserves the hierarchical information embedded in the digital copies of the code (i.e. divisions into Article, Chapters, Sections, etc.). To extract this information we develop dictionaries of the words that each code uses to refer to divisions in the code. We construct these dictionaries by reviewing the hierarchy of the code and, in most instances, this review is conducted by speakers of the relevant language. The XML files produced by this process also preserve the names, numbering, and associated statutory content of each section.

As part of this process, we must decide the level at which the hierarchy should terminate (i.e. do we treat subsections or numbered lists within subsections as nodes). Following previous work in this area, we identify the lowest level at which the code can terminate and use this level as the bottom of the hierarchy (Katz and Bommarito 2014). To make this concept more concrete, imagine a code that is divided into articles, sections, and subsections. Imagine further that all the articles have sections underneath them, but that not all of the sections have subsections below them. In this example, the lowest level at which the hierarchy can terminate is the section because some sections have no subsections underneath them. While taking this approach loses some amount of detail, it makes it easier to compare and interpret links within the code. This choice does so by ensuring that each node is attached to a complete unit.

The next step is to convert the XML files to JSON (JavaScript Object Notation) files using Python. For nearly all countries, this process is where we identify cross-references within the code.[4] As before, we develop dictionaries that contain the language used to make cross-references. The program then identifies all of the distinct nodes within the legal code and they identify all of the links between these nodes. This information allows us to use the networkx package in Python to compute network

4. The electronic version of the French Code has cross-reference information embedded in the XML files, which allowed us to identify cross-references when processing the XML file.

statistics and to produce visualizations of each network.

The Network Structures of Civil Codes

This section provides a brief overview of the network structure of five major European countries that use the civil law system. The goal of this exercise is to show that the methodology we use produces results that are consistent with traditional understandings of the differences between legal families. We focus on France, Germany, Italy, Spain, and Austria because there is substantial traditional scholarship on the historical development of these codes.

We begin with visualizations of the reference networks. Figures 1 through 5 provide the visualizations for each of the five countries. To make the visualizations manageable we filter the nodes displayed by minimum degree (i.e. the minimum number of incoming and outgoing links from the node). For four out of the five figures we use a minimum degree of five, but for Italy, which has an exceptionally dense network, we use a minimum degree of 9. The visualizations confirm some of the insights of the traditional comparativists. The German network appears quite dense, while the French network is substantially more sparse. The Italian code, which comparativists claim has been strongly influenced by the German code is even more networked than the BGB. The Spanish Civil Code, which drew substantial inspiration from the Napoleonic Code is, like the French code, quite sparse. The Austrian Civil Code, which draws on a more independent tradition, falls somewhere in the middle of the German-influence and French-influenced codes.

Table 1, which supplies some basic network statistics, further confirms what the visualizations suggest. The BGB, which has about 450 fewer nodes than the French code, has almost double the number of cross-references (links). The French and Spanish codes, which comparativists claim are strongly related to each other, have very similar densities. The two codes with the highest densities are the related German and

Figure 1. *Internal Citation Network of the French Civil Code*

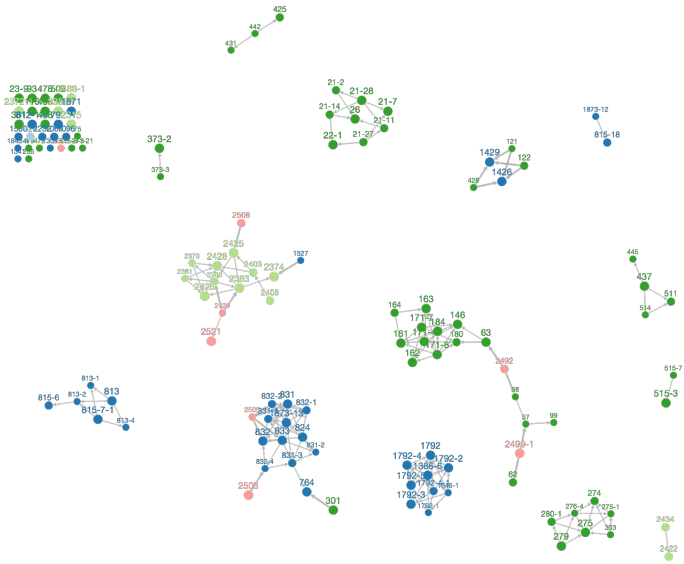


Figure 2. *Internal Citation Network of the German Civil Code*

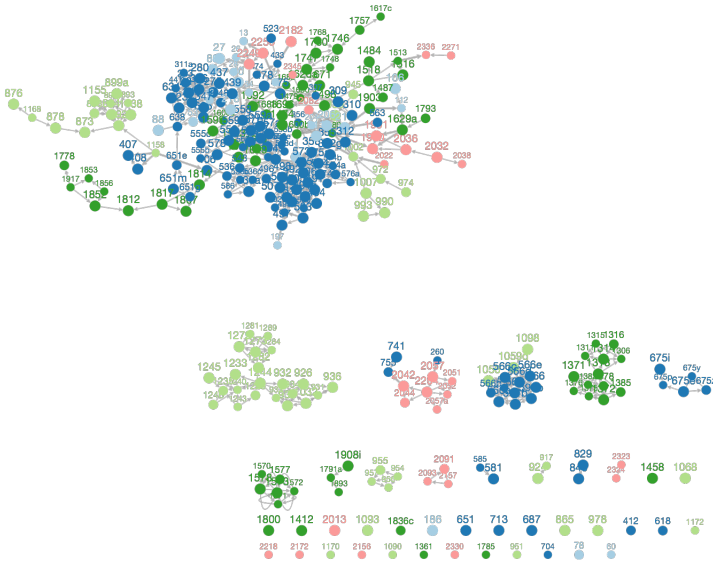


Figure 3. *Internal Citation Network of the Italian Civil Code*

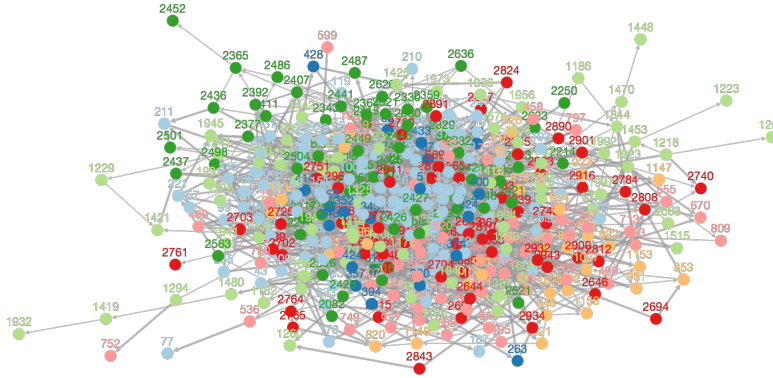


Figure 4. *Internal Citation Network of the Spanish Civl Code*

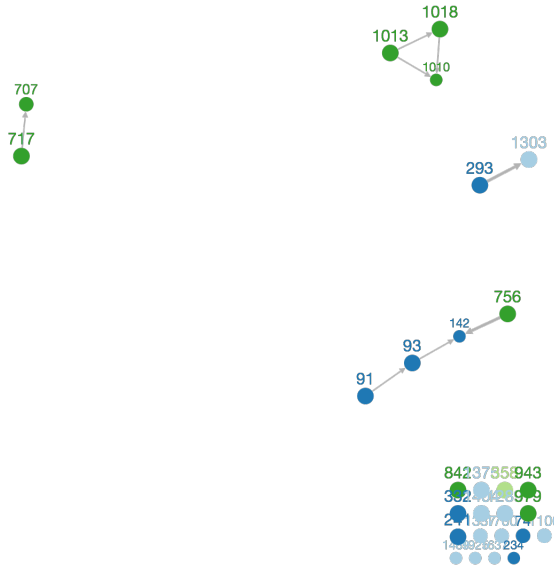
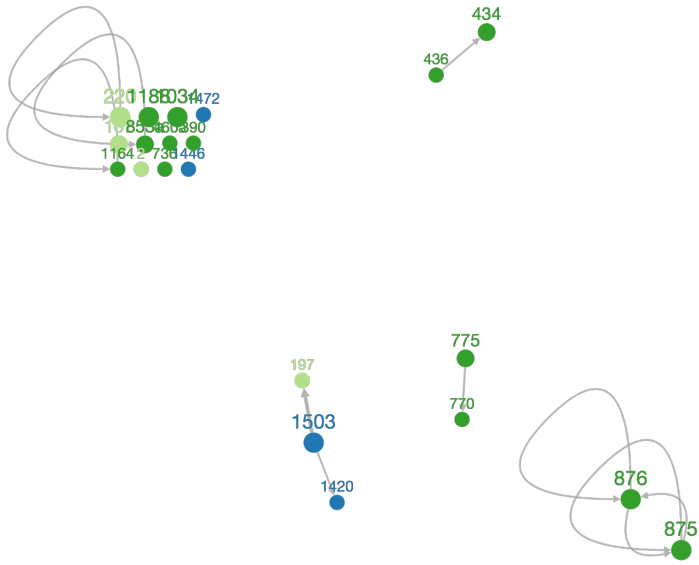


Figure 5. *Internal Citation Network of the Austrian Civil Code*



Italian codes. The density of the Austrian code falls in between these two other groups.

Table 2 displays some more refined measures of network structure. The number of weakly connected components provides an indication of how networked the code is. A low number of these components suggests a high degree of connectivity (if every node is reachable from every other node, the number of weakly connected components is one). From lowest to highest, the rank order for this statistic is: Italy, Germany, Austria, Spain, France. This suggests the same grouping as the basic density statistics. It is worth noting, however, that Spain has a smaller number of weakly connected components than France even though Spain's code is less dense. This means that Spain's code has more clustering of cross-references relative to France.

The indegree and outdegree variances provide information about the distribution of incoming and outgoing links within the network. There are several items worth noting. The first is that Italy dwarfs all the other countries in terms of the magnitude of indegree and outdegree variance. This means that Italy has a small number of nodes with extremely high levels of incoming and outgoing links. The second is that for all the countries except Italy, the outdegree variance is larger than the indegree variance. This means that, relative to the baseline for each code's network, there more nodes that have a large number of references out to other sections than there are nodes that have incoming references. This phenomenon may indicate that there are sections that are well integrated into the rest of the code and that–other than the Italian code–there are relatively few sections that provide definitions that are used throughout the code.

5 Conclusion

Cross-references are a simple and widely-spread characteristic of texts. They serve the purpose of facilitating reading by directing the reader to earlier or later parts of the text. In legal texts, cross-references have the

Table 1. *Basic Network Statistics*

Country	Nodes	Links	Density	Average Degree
Austria	1327	371	0.000211	0.56
France	2847	1301	0.000161	0.91
Germany	2392	2517	0.000440	2.10
Italy	3051	6697	0.000720	4.39
Spain	1992	578	0.000146	0.58

Table 2. *Additional Network Statistics*

Country	# of Weakly Connected Comp.	Indeg. Variance	Outdeg. Variance	Hub5	Hub10	Auth.5	Auth.10
Austria	1009	0.391	1.544	7	1	0	0
France	1881	1.180	1.879	38	8	22	2
Germany	862	2.412	7.233	99	27	45	4
Italy	556	62.499	11.170	293	85	260	98
Spain	1480	0.397	1.645	19	11	0	0

additional feature of supplementing or qualifying the normative import of a certain rule by referring the reader, for instance, to general definitions provided at the outset, to exceptions to the rule or to further applications of a general principle. Cross-references have been a feature of legal texts since the earliest documented written rules and continue to do so today.

Because they determine the way in which a text delivers its normative meaning, cross-references embed important aspects of the philosophy that guided the drafters and hence can be used to uncover differences and similarities among codification efforts and the resulting codes. Doing so has some advantages over both traditional and empirical methods of comparative legal analysis. Compared to traditional methods, network analysis of internal cross-references allows the analyst to scale up the analysis to large numbers of countries, which is virtually impossible in the traditional setup. Compared to recently-developed methods of empirical comparative analysis, the networks of internal cross-references provide a continuous, rather than discrete or even binary, measure of similarities and differences among legal systems which do not rely on assumption as to how different systems should be grouped into "families". The methodology illustrated in the previous sections can thus be used to yield novel insights as to how legal systems develop, influence each other and diverge over time.

References

Arminjon, Pierre, Boris Nolde, and Martin Wolff. 1951. *Traité de droit comparé*. Paris: Librairie générale.

Badawi, Adam B, Rens Bod, James Daily, Giuseppe Dari-Mattiacci, Tom Ginsburg, Bart Karstens, and Marijn Koolen. 2018. "Citation Networks and Legal Origins."

David, Rene, and John EC Brierley. 1978. *Major legal systems in the world today: an introduction to the comparative study of law*. Simon / Schuster.

Djankov, Simeon, Rafael La Porta, Florencio Lopez-de-Silanes, and Andrei Shleifer. 2003. "Courts." *The Quarterly Journal of Economics* 118 (2): 453–517.

———. 2008. "The law and economics of self-dealing." *Journal of financial economics* 88 (3): 430–465.

Glaeser, Edward L, and Andrei Shleifer. 2002. "Legal origins." *The Quarterly Journal of Economics* 117 (4): 1193–1229.

Glenn, H. Patrick. 2006. "The Oxford Handbook of Comparative Law." Chap. Comparative Legal Families and Comparative Legal Traditions, edited by Mathias Reimann and Reinhard Zimmermann, 422–439. Oxford University Press. http://www.oxfordhandbooks.com/view/10.1093/oxfordhb/9780199296064.001.0001/oxfordhb-9780199296064-e-013.

Gordley, James. 2006. "Comparative Law and Legal History," edited by Mathias Reimann and Reinhard Zimmermann. http://www.oxfordhandbooks.com/view/10.1093/oxfordhb/9780199296064.001.0001/oxfordhb-9780199296064-e-024.

Isele, Hellmut Georg. 1949. "Ein halbes Jahrhundert deutsches Bürgerliches Gesetzbuch." *Archiv für die civilistische Praxis* 150 (H. 1): 1–27.

Katz, Daniel Martin, and MJ Bommarito. 2014. "Measuring the complexity of the law: the United States Code." *Artificial intelligence and law* 22 (4): 337–374.

Klerman, Daniel M, Paul G Mahoney, Holger Spamann, and Mark I Weinstein. 2011. "Legal origin or colonial history?" *Journal of Legal Analysis* 3 (2): 379–409.

Kötz, Hein, and Konrad Zweigert. 1998. "An Introduction to Comparative Law." *Vol. II (T. Weir trans. 2ed 1987)*.

La Porta, Rafael, Florencio Lopez-de-Silanes, Andrei Shleifer, and Robert Vishny. 2002. "Investor protection and corporate valuation." *The journal of finance* 57 (3): 1147–1170.

La Porta, Rafael, Florencio Lopez-de-Silanes, Andrei Shleifer, and Robert W Vishny. 1998. "Law and finance." *Journal of political economy* 106 (6): 1113–1155.

———. 2000. "Agency problems and dividend policies around the world." *The journal of finance* 55 (1): 1–33.

Porta, Rafael La, Florencio Lopez-de-Silanes, and Andrei Shleifer. 2008. "The Economic Consequences of Legal Origins." *Journal of Economic Literature* 46 (2): 285–332. http://www.jstor.org/stable/27646991.

Roe, Mark J. 2006. "Legal origins, politics, and modern stock markets." *Harvard Law Review:* 460–527.

Spamann, Holger. 2009a. "Contemporary legal transplants: legal families and the diffusion of (Corporate) law." *BYU L. Rev.* 1813.

———. 2009b. "The "antidirector rights index" revisited." *The Review of Financial Studies* 23 (2): 467–486.

———. 2015. "Empirical comparative law." *Annual Review of Law and Social Science* 11:131–153.