Examining the Evolution of Legal Precedent Through Citation Network Analysis

Iain Carmichael

James Wudel

Michael Kim

James Jushchuk

Follow this and additional works at: http://scholarship.law.unc.edu/nclr

Part of the Law Commons

Recommended Citation

EXAMINING THE EVOLUTION OF LEGAL PRECEDENT THROUGH CITATION NETWORK ANALYSIS[*]

INTRODUCTION

Precedent—the rule that judges rely on past decisions in adjudicating the disputes in front of them—is a fundamental part of the American legal system. Precedent plays a role in understanding how the law evolves, determining how the courts behave, and helping legal search engines recommend relevant cases.[1] Studying the network of citations between judicial opinions can provide an empirical perspective on precedent. This Comment examines the Supreme Court citation network and the Federal Appellate citation network.[2] Building upon previous empirical legal work, this Comment uses network analysis to provide insights into the role of precedent in the courts and to identify important cases in various contexts. Vertex centrality metrics, which measure how *important* a vertex is in a network in different ways, provide a way of quantifying the notion of importance of a case in a citation network. There are many kinds of vertex centrality metrics. This Comment further develops a methodology to evaluate vertex centrality metrics in an evolving network based on how predictive a metric is of future citations. This methodology is able to identify several possibly surprising results regarding court behavior and behavior of the metrics themselves. In particular, it unexpectedly shows that the number of cases cited in an opinion is a stronger predictor of whether that opinion will be cited in the future than the number of times that opinion has already been cited by other opinions.

The broader aim of this research is to understand the factors driving the evolution of the law. The law evolves incrementally by building on precedent as judges answer novel questions based on principles set out in prior cases.[3] Understanding the precedential weight of a case is a challenging but productive task for several reasons. Scholars, for instance, use precedent to examine what factors influence the evolution of the law, identify which issues are currently most relevant, and predict what issues might become active in the future.[4] Practitioners are often required to identify relevant cases that

1. *See* William M. Landes & Richard A. Posner, *Legal Precedent: A Theoretical and Empirical Analysis*, 19 J.L. & ECON. 249, 250 (1976).

2. The citation network here means the network of cases and the citations between them and is discussed more in Section II.A.

3. *See* Landes & Posner, *supra* note 1, at 250.

4. *See, e.g.*, Frank B. Cross & James F. Spriggs II, *The Most Important (and Best) Supreme Court Opinions and Justices*, 60 EMORY L.J. 407, 442 (2010).

are most likely to convince a judge.[5] Additionally, nonprofit organizations, researchers, and companies might want to build legal research tools.[6]

One way that precedent can be examined is through citations in written judicial opinions. This method of analysis operates on the assumption that "[e]ach judicial citation contained in an opinion is essentially a latent judgment about the case cited."[7] In other words, a citation is a good indication that a cited case is precedent for the case at hand.[8] Based on this assumption, scholars have begun using empirical methods based on the network of legal citations to rank the value of cases.[9]

Formally, a *network* is a collection of objects (called *vertices* or *nodes*) and connections between them (called *edges*).[10] The study of networks has become popular in recent decades the internet is a network of computers,[11] Facebook and Twitter capture human social networks,[12] and neuroscientists study the brain as a network of neurons connected by white matter fiber tracts.[13] Another popular subject of study is the citation networks of academic papers.[14] This

---

5. *See, e.g.*, Kevin Bennardo, *Testing the Geographical Proximity Hypothesis: An Empirical Study of Citations to Nonbinding Precedents by Indiana Appellate Courts*, 90 NOTRE DAME L. REV. ONLINE 125, 149 (2015).

6. For an example of a legal research tool and how they may operate, see *Natural Language Searches*, THOMSON REUTERS WESTLAW, https://lawschool.westlaw.com /marketing/display/RE/151 [http://perma.cc/JQ4C-XQGT] (describing Westlaw's Natural Language search tool).

7. James H. Fowler & Sangick Jeon, *The Authority of Supreme Court Precedent*, 30 SOC. NETWORKS 16, 17 (2008) (emphasis omitted). Some scholars have resisted this context-neutral approach to citations and prefer to use data from LexisNexis's citator tool, Shepard's, or Westlaw's citator tool, KeyCite, to only capture citations with a positive valence. *See, e.g.*, Matthew P. Hitt, *Measuring Precedent in a Judicial Hierarchy*, 50 L. & SOC'Y REV. 57, 63 (2016).

8. *See* Fowler & Jeon, *supra* note 7, at 17.

9. *See, e.g.*, James H. Fowler et al., *Network Analysis and the Law: Measuring the Legal Importance of Precedents at the U.S. Supreme Court*, 15 POL. ANALYSIS 324, 325 (2007).

10. ERIC D. KOLACZYK, STATISTICAL ANALYSIS OF NETWORK DATA: METHODS AND MODELS 15–16 (2009).

11. *Id.* at 168.

12. JULIAN MCAULEY & JURE LESKOVEC, LEARNING TO DISCOVER SOCIAL CIRCLES IN EGO NETWORKS 1 (2012), http://cs.stanford.edu/people/jure/pubs/circles-nips12.pdf [perma.cc/3UFB-UUP9].

13. *See* R. Cameron Craddock et al., *Imaging Human Connectomes at the Macroscale*, 10 NATURE METHODS 524, 526 (2013).

14. *See, e.g.*, JURE LESKOVEC, JON KLEINBERG & CHRISTOS FALOUTSOS, GRAPHS OVER TIME: DENSIFICATION LAWS, SHRINKING DIAMETERS AND POSSIBLE EXPLANATIONS 3 (2005), https://www.cs.cmu.edu/~jure/pubs/powergrowth-kdd05.pdf [http://perma.cc/YV2K-SGKH].

Comment examines legal citation networks, defined as networks of judicial opinions and the citations going between them.[15] Each case is a *vertex*, and each citation is an *edge*. Figure 1 shows an example of a network. In particular, the vertices (dots) represent *Roe v. Wade*[16] and cases that either cite to or are cited by *Roe v. Wade*. The edges (lines) represent citations between this set of cases.

Figure 1: A visual of a network; this plot shows *Roe v. Wade* and its neighboring cases. Each dot represents a case and each gray arrow represents a citation from one case to another.[17]

There are many different ways to quantify the importance of a vertex in a network, called *vertex centrality metrics*.[18] Two of the simplest vertex centrality metrics are *in-degree* and *out-degree*. In-degree is the count of citations a case has received, while out-degree is the count of cases cited in an opinion.[19] Citations are directed backwards in time, which means that a citation would go *out* from a case in 2017 and *in* to a case in 1990. Figure 2 illustrates a simple

---

15. For an example of a legal citation network that includes statutory provisions, see Michael J. Bommarito II, Daniel Martin Katz & Jillian Isaacs-See, *An Empirical Survey of the Population of U.S. Tax Court Written Decisions*, 30 VA. TAX REV. 523, 524 (2011).

16. 410 U.S. 113 (1973).

17. This graphic is reproduced from Fowler et al., *supra* note 9, at 329.

18. *See infra* Part II.

19. *See* KOLACZYK, *supra* note 10, at 16. In-degree equates to the number of cases listed in the "Citing References" tab of Westlaw or "Citing Decisions" tab of a Shepard's report in LexisNexis. Similarly, out-degree corresponds to the number of cases in a Table of Authorities in Westlaw or a Shepard's report in LexisNexis.

example of a hypothetical citation network with six nodes/cases (dots) and eight directed edges/citations (arrows). The highlighted case A is cited by three other cases meaning it has an in-degree of three. Similarly, case A cites two other cases meaning it has an out-degree of two.
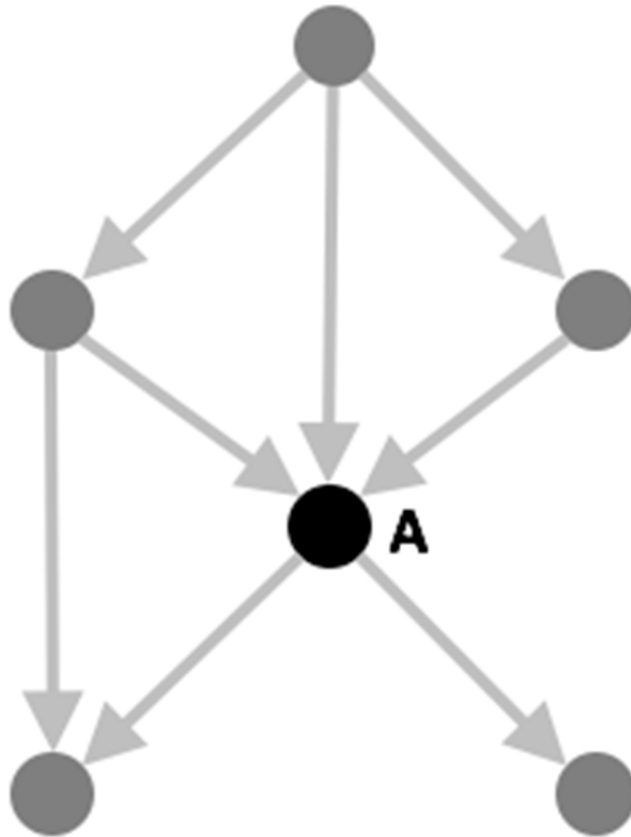

Figure 2: A simple example of a citation network with six cases. The highlighted case A has been cited three times. Therefore, case A has an in-degree equal to three. Similarly, case A cites two cases and therefore has an out-degree equal to two.

In addition to the in-degree and out-degree metrics, there are many more sophisticated measures of vertex centrality that are built on other assumptions about how the structure of a network reflects

importance.[20] However, despite the large number of vertex centrality metrics, there is little empirical work that evaluates which centrality metric is the right choice for a given task.[21]

This Comment develops a novel methodology to evaluate vertex centrality metrics in a citation network. The core assumption of this methodology is that a good vertex centrality metric should be able to predict future citations. Because of the theoretical foundation established by legal scholars connecting citations to precedent, the results of this methodology may have important implications for the study of precedent.[22]

Furthermore, this Comment models methods of reasoning that lawyers will have to engage in more frequently in the future. The number of areas of legal practice in which lawyers would benefit from a baseline understanding of the principles underlying legal tools is growing.[23] Therefore lawyers will have to make informed decisions about technical statistical issues. Legal research, electronic discovery, and transactional practice are all increasingly being shaped by complicated statistical methods,[24] and statistical methods are creating new categories of tools and services, such as outcome prediction in litigation.[25] Thus, lawyers in a growing number of fields stand to gain from understanding statistical concepts and increasingly risk liability if they do not make themselves aware of how their tools work.[26]

This Comment further develops the use of citation network analysis to study legal precedent. Our findings are based on a novel statistical methodology, which we developed to empirically compare vertex centrality metrics in a citation network. We find novel

---

20. *Id.* at 88–93. The particular vertex centrality metrics used in this Comment are discussed in Section II.A.

21. For a survey of prior legal work that has attempt to compare vertex centrality metrics, see *infra* notes 60–62 and accompanying text.

22. This is not to say that legal citation network analysis is a priori an analysis of precedent, nor is it to say that precedent is the only area of legal study that could benefit from network analysis.

23. *See* William Henderson, *What the Jobs Are: New Tech, New Client Needs Create a New Field of Legal Operations*, A.B.A. J., Oct. 2015, at 36, 41.

24. *See id.*; *See also* Steve Lohr, *A.I. Is Doing Legal Work. But It Won't Replace Lawyers, Yet.*, N.Y. TIMES (Mar. 19, 2017), https://www.nytimes.com/2017/03/19 /technology/lawyers-artificial-intelligence.html [http://perma.cc/9GXW-F9RC] (discussing the use of artificial intelligence in different legal tasks and practices).

25. *See, e.g.*, *What We Do*, LEX MACHINA, https://lexmachina.com/what-we-do/ [perma.cc/TXV4-R58Z].

26. *See, e.g.*, Debra Cassens Weiss, *Lawyer's E-Discovery Error Led to Release of Confidential Info on Thousands of Wells Fargo Clients*, A.B.A. J. (July 27, 2017) http://www.abajournal.com/news/article/lawyers_e_discovery_error_led_to_release_of _confidential_wells_fargo_client [http://perma.cc/9HQC-6WTB].

evidence to support previous claims made about the nature of precedent and argue that how well-grounded an opinion is in existing precedent, the presence of multiple opinions in a case, and the age of an opinion all contribute to how likely a case is to be cited in the future.

This Comment proceeds in four parts. Part I surveys prior relevant work on American case law citation networks. Part II gives additional background information on vertex centrality metrics and presents a detailed description of the data and the methodology. Part III presents the results of the statistical analyses of both the United States Supreme Court network and the Federal Appellate network. The methodology demonstrates that out-degree is more predictive of future citations than in-degree, which was an unexpected result. Less surprisingly, the methodology also demonstrates strong aging effects in the network, or that newer cases are more likely to be cited than older cases. The results reaffirm and build upon existing research in both assessments of vertex centrality metrics and of the effect of time on citation rates. Part IV discusses the legal and conceptual importance of these results. Specifically, it speculates about which qualities of cases will drive out-degree performance, particularly how well supported a case is and the presence of multiple opinions within a case. The presence of aging effects in the network is explained from both a legal and statistical perspective. Part IV concludes by arguing that one vertex centrality metric, which did not perform well at predicting future citations, is biased towards cases of first impression.

I. PRECEDENT AND CASE LAW CITATION NETWORK RESEARCH

Landes and Posner define precedent as "something done in the past that is appealed to as a reason for doing the same thing again."[27] Precedent in case law develops based on analogical reasoning where one decision can "be an authority for another is that the facts are alike, or, if the facts are different, that the principle which governed the first case is applicable to the variant facts."[28] The principle that allows precedent to control is stare decisis. Stare decisis is "[t]he

---

27. Landes & Posner, *supra* note 1, at 250. Black's Law Dictionary defines precedent as "[a]n action or official decision that can be used as support for later actions or decisions" or "a decided case that furnishes a basis for determining later cases involving similar facts or issues." *Precedent*, BLACK'S LAW DICTIONARY (10th ed. 2014).

28. Landes & Posner, *supra* note 1, at 250 (quoting WILLIAM M. LILE ET AL., BRIEF MAKING AND THE USE OF LAW BOOKS 288 (Roger W. Cooley & Charles Lesley Ames eds., 3d ed. 1914)).

doctrine of precedent, under which a court must follow earlier judicial decisions when the same points arise again in litigation."[29]

Precedent and stare decisis operate in a variety of ways.[30] The concept that a lower court must follow the decisions of a higher court in the same jurisdiction is referred to as "vertical" stare decisis.[31] "Horizontal" stare decisis, on the other hand, is the doctrine that a court, generally an appellate court, "must adhere to its own prior decisions, unless it finds compelling reasons to overrule itself."[32] Other authorities are merely persuasive, meaning that they are "not binding on a court, but . . . [are] entitled to respect and careful consideration."[33]

Past empirical research on legal citation networks has been largely concerned with precedent. This research operates on the principle that "[e]ach judicial citation contained in an opinion is essentially a latent judgment about the case cited."[34] Or, rather, the fact that a judge has taken the time to include a citation to a particular case is a judgment on the quality of that case.[35] Studying patterns of citations enables a better understanding about the evolution, growth, and state of the law.

Early empirical studies of precedent through citations involved counting and collating the citations a court made in a given time period.[36] Or, in network terms, early empirical studies of precedent relied on in-degree and out-degree.[37] Scholars have used and continue to use this method to examine the writings of a range of courts, including the Supreme Court of California,[38] federal circuit courts,[39]

---

29. *Stare Decisis*, BLACK'S LAW DICTIONARY (10th ed. 2014).

30. In addition to the descriptive discussion of precedent and stare decisis, there is an active normative discussion of the power of precedent, or rather, how much deference courts *should* give to past decisions. *See* MICHAEL J. GERHARDT, THE POWER OF PRECEDENT 3–5 (2008).

31. *Stare Decisis*, *supra* note 30.

32. *Id*.

33. *Precedent*, *supra* note 28. Persuasive authorities include cases decided in a neighboring jurisdiction, which a court might evaluate "without being bound to decide the same way." *Id*.

34. Fowler & Jeon, *supra* note 7, at 17.

35. *Id*.

36. *See, e.g.*, John Henry Merryman, *The Authority of Authority: What the California Supreme Court Cited in 1950*, 6 STAN. L. REV. 613, 652–54 (1954).

37. *See supra* note 20 and accompanying text. For an example, see Merryman, *supra* note 37, at 653 (tabulating and categorizing every citation by the California Supreme Court in 1950).

38. *See* Merryman, *supra* note 37, at 613, 617–18.

39. Landes & Posner, *supra* note 1, at 252.

and the Supreme Court of the United States.[40] With the dramatic increase of computing power and storage of the last twenty years and the corresponding increased electronic availability of legal information, especially case law, scholars have been able to assess larger numbers of cases and citations.[41] At the same time, scholars have also begun using more sophisticated and computationally intensive network methods to measure the positions of cases and patterns of citation within bodies of law.[42] In fact, the Supreme Court citation network is used with some frequency to demonstrate novel vertex centrality measures and other network methods.[43] Moreover, it is used as an introduction to the subject of networks and vertex centrality in at least one introductory quantitative research book.[44]

One common application of vertex centrality measures is to use them as a proxy for overall importance and, in turn, use them to rank cases.[45] Perhaps the most forward example of this ranking genre is presented by Cross and Spriggs in their article, *The Most Important (and Best) Supreme Court Opinions and Justices*.[46] The authors counted citations to Supreme Court opinions by the Supreme Court, the circuit courts, and the district courts and incorporated a more sophisticated network centrality into their rankings.[47] The authors found that a wide range of factors influence how strong a precedent is, including the issue area of the case, the age of the case, and the

---

40. *Id.* A closely parallel, but conceptually distinct, line of study using the same method of citation counting is the study of the influence of individual judges. *See, e.g.*, William M. Landes, Lawrence Lessig & Michael E. Solimine, *Judicial Influence: A Citation Analysis of Federal Courts of Appeals Judges*, 27 J. LEGAL STUD. 271, 271–72 (1998).

41. *Compare* Landes & Posner, *supra* note 1, at 252–53 (studying the citations present in roughly 1,000 cases in 1976) *and* Merryman, *supra* note 37, at 652–53 (studying the citations present in roughly 300 cases in 1954), *with* Fowler & Jeon, *supra* note 7, at 17 (studying the citations between roughly 30,000 cases in 2008).

42. *See, e.g.*, Fowler et al., *supra* note 9, at 328–30.

43. *See, e.g.*, John W. Patty, Elizabeth Maggie Penn & Keith E. Schnakenberg, *Measuring the Latent Quality of Precedent: Scoring Vertices in a Network*, *in* ADVANCES IN POLITICAL ECONOMY: INSTITUTIONS, MODELLING AND EMPIRICAL ANALYSIS 249, 252 (Norman Schofield, Gonzalo Caballero & Daniel Kselman eds., 2013); Dane Taylor et al., *Eigenvector-Based Centrality Measures for Temporal Networks*, 15 MULTISCALE MODELING & SIMULATION 537, 556, 564 (2017).

44. TAYLOR ARNOLD & LAUREN TILTON, HUMANITIES DATA IN R: EXPLORING NETWORKS, GEOSPATIAL DATA, IMAGES, AND TEXT 85–87 (2015).

45. *See, e.g.*, Cross & Spriggs, *supra* note 4, at 410–11.

46. *Id*. at 407, 410–11.

47. *Id*. at 431–42. More specifically, the authors use a "legal relevance score," which is related to hubs and authorities. *Id*.; *see also id*. at 416 (discussing use of links between cases as the measure of legal importance). See Appendix A for a discussion of hubs and authorities.

NORTH CAROLINA LAW REVIEW [Vol. 96

length of the opinion.[48] Somewhat surprisingly, they found that unanimous decisions are less influential at the Supreme Court level, and that decisions by minimum-winning coalitions are no less influential than other decisions.[49]

Another example of the ranking genre, and an early example of a study using a network centrality measure other than in-degree, is found in Fowler and Jeon's *The Authority of Supreme Court Precedent*.[50] The authors analyzed the Supreme Court network using a more sophisticated vertex centrality metric and compared the results to lists of important Supreme Court cases compiled by legal experts and published by Congressional Quarterly, the Legal Information Institute, and the *Oxford Guide*.[51] Fowler and Jeon found that all ten of their most highly ranked authorities "are considered to be important by either Congressional Quarterly, the Legal Information Institute, or the *Oxford Guide*."[52] They further parsed their results by issue area and found similarly strong correlations between opinions of legal experts and the top five opinions identified as authorities within the broad categories of civil rights, criminal law, First Amendment, and privacy.[53]

Other scholars employ these measures to study various qualities of case law.[54] One study examined the speed at which the probability of a case to be cited changed over time and found that a case's chances of being cited "depreciate[s] about 81 percent and 85 percent between [its] first and 20th years of age at the Supreme Court and courts of appeals, respectively."[55] Multiple studies have examined differences in citation patterns across levels of the judicial hierarchy.[56] One demonstrated that the Supreme Court frequently cited "doctrinal paradoxes . . . , opinions of the Court for which every rationale for the Court's judgment is rejected by a majority," in an iterative process of working through complicated legal questions.[57]

---

48. *Id*. at 476–80.

49. *Id*. at 479–80.

50. Fowler & Jeon, *supra* note 7, at 16–17.

51. *Id*. at 17–24. More specifically, the authors use hubs and authorities, which is discussed at more length in Appendix A.

52. *Id*. at 21.

53. *Id*. at 22–23.

54. *See, e.g.*, Ryan C. Black & James F. Spriggs II, *The Citation and Depreciation of U.S. Supreme Court Precedent*, 10 J. EMPIRICAL LEGAL STUD. 325, 325 (2013).

55. *Id.* at 325; *see also* Thom Neale, *Citation Analysis of Canadian Case Law*, 1 J. OPEN ACCESS L., 1, 51–52 (finding that Canadian cases, apart from those of the Supreme Court, are rarely cited more than 15 years after publication).

56. Cross & Spriggs, *supra* note 4, at 431; Hitt, *supra* note 7, at 57.

57. Hitt, *supra* note 7, at 67–68.

Conversely, doctrinal paradoxes were no more or less likely to be cited at the circuit court level and were less likely to be cited at the district court level.[58]

At a more sophisticated level, some papers have evaluated rankings produced by vertex centrality metrics. These papers evaluate vertex centrality metrics rankings based on a comparison to some external factor, such as expert opinion[59] or page views.[60] One article, *Network Analysis and the Law: Measuring the Legal Importance of Precedents at the U.S. Supreme Court*, evaluates vertex metrics based on their ability to predict if a case will be cited in an upcoming year.[61] This Comment's methodology is similar to that of *Network Analysis and the Law*, but it uses richer data and examines potential citations at the case level as opposed to aggregating all cases in each year.

II. METHODS

A. Vertex Centrality Metrics

There are a number of different kinds of vertex centrality metrics. The most popular ones can be grouped into three categories: degree-based, eigenvector-based, and positional.[62] Degree-based metrics include in-degree and out-degree; these metrics simply count raw numbers of citations.[63] Degree-based measures include in-degree and out-degree.[64] While the various centrality metrics are often related, they are driven by different structural properties of the network.

The class of eigenvector centrality metrics is based on the idea that a case is important if it is cited by a lot of cases that are themselves important.[65] Eigenvector centrality metrics judge a case to be more important if it is cited by many cases that are themselves cited by many other cases. Figure 3 demonstrates this idea in a small, hypothetical citation network; the circles represent cases and the

58. *Id*. at 59.

59. Fowler & Jeon, *supra* note 7, at 21–23. Other papers have used multiple vertex centrality metrics without a method to evaluate those rankings. *See, e.g.*, Cross & Spriggs, *supra* note 4, at 420.

60. Neale, *supra* note 56, at 22–23 (evaluating vertex centrality metrics on their ability to predict internet page views of Canadian case law).

61. Fowler et al., *supra* note 9, at 342.

62. *See* KOLACZYK, *supra* note 10, at 88–93.

63. *See id*. at 80 (noting the degree of a vertex "provides a basic quantification of the extent to which *v* is connected to other vertices within the graph").

64. *See supra* note 20 and accompanying text.

65. *See* KOLACZYK, *supra* note 10, at 90.

arrows represent citations. The highlighted cases A and B both receive the same number of citations (i.e., two) meaning they have the same in-degree. An eigenvector centrality metric would rank case A better than case B since A is cited by case C, which has a large number of citations.

Eigenvector centrality measures include: PageRank,[66] Eigenvector centrality,[67] and hubs and authorities.[68] PageRank is one of the key mathematical components of Google's search algorithm.[69] While PageRank works very well for networks of web pages, Section IV.D discusses why it is less appropriate for citation networks.
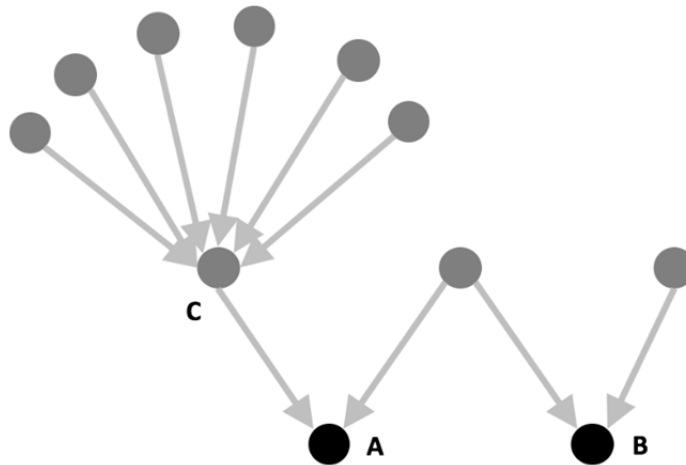
Figure 3: Figure 3 shows a small citation network. Cases A and B would be ranked equally by in-degree, but case A would be ranked better by eigenvector centrality metrics as described in text above.

---

66. *See* Kurt Bryan & Tanya Leise, *The $25,000,000,000 Eigenvector: The Linear Algebra Behind Google*, 48 Soc'y. for Indus. & Applied Mathematics Rev. 569, 569 (2006).

67. *See* Kolaczyk, *supra* note 10, at 90. Confusingly, eigenvector centrality refers to the broader category and a particular member of this category.

68. *See* Jon M. Kleinberg, *Authoritative Sources in a Hyperlinked Environment*, 46 J. Assoc. for Computing Machinery 604, 696 (1999).

69. *See* Bryan & Leise, *supra* note 67, at 569.

The class of metrics that are called positional are based on the idea that important cases are "close" to other cases in the sense that "distance" is measured by the number of citations it would take to go from one case to another.[70] Figure 4 shows a hypothetical network with seven nodes. The highlighted case A would be ranked highest by positional metrics since it is "closest" to all other nodes on average. Positional metrics include betweenness centrality and closeness centrality.[71]
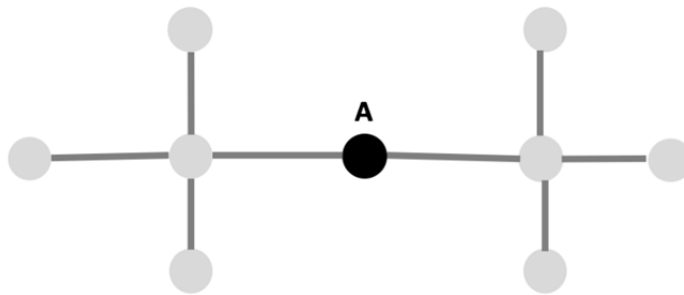

Figure 4: Figure 4 shows a simple network. The highlighted node A would be ranked highest by most positional vertex centrality metrics since it is "closest" to all other nodes. The network is undirected here for simplicity.

Time plays an important role in the evolution of the legal citation network in that more recent cases are often cited over older cases.[72] In some circumstances, taking time into account may be desirable. For example, a legal search engine may want to favor newer cases in order to give attorneys quicker access to the most current understanding of the law.[73] Although none of the standard vertex centrality metrics discussed above in this Section incorporate time,[74] it is possible to construct time aware vertex centrality metrics that take the date of each case into account—typically by decreasing the weight

---

70. *See* KOLACZYK, *supra* note 10, at 88–90.
71. *See id.*
72. *See* Black & Spriggs, *supra* note 55, at 327.
73. *See infra* text accompanying note 140.
74. *See* KOLACZYK, *supra* note 10, at 88–93.

of older cases.[75] Consequently, Section III.A.2 also examines two time-aware vertex centrality metrics: CiteRank and the number of citations in recent years.[76]

B. The Methodology

This Comment develops a methodology to compare vertex centrality metrics, which measure some notion of how important a case is in a citation network, based on the evolution of the citation network. The core assumption underlying this methodology is that a better vertex centrality metric will better predict future citations. In particular, an experiment is run on the citation network, which compares how well each vertex centrality metric can predict future citations.

1. Sort Experiment

The experiment attempts to predict which existing cases a new opinion will cite based only on vertex centrality metrics of the citation network at that time. Vertex centrality metrics do not by themselves provide enough information to accurately predict citations; therefore, we slightly modify the problem.[77] Instead of making binary predictions about whether an existing case will be cited by the new case, all existing cases are ranked by vertex centrality metrics. Then, the cases that were actually cited by the new opinion are examined in order to quantify how well these cases were ranked by each vertex centrality metric.

One thousand test cases between 1900 and 2016 were randomly selected to evaluate. For each test case, the citation network just before the test case enters is considered[78] and all vertex centrality metrics of interest (e.g., in-degree, PageRank, etc.) are computed. Each vertex centrality metric gives a ranking of the cases (e.g., in-degree ranks the case with the most citations at this time as the top case). Then, the cases that were actually cited by the test case were observed[79] and the *mean rank score*[80] of these cases was computed for

75. *See, e.g.*, Walker et al., *Ranking Scientific Publications Using a Model of Network Traffic*, 2007 J. Stat. Mechanics 1, 3 (2007).

76. There are a number of technical details discussed in Appendix A such as the differences between directed and undirected centrality metrics and the details of the time aware centrality metrics.

77. A model that took the topic of the opinion or the judge authoring the opinion into account would make better predictions.

78. For computational reasons, the network and vertex centrality metrics are computed once each year from 1900 to 2016.

79. Opinions that cite zero cases are ignored.

each vertex centrality metric ranking. Mean rank score is related to the average position of the citations in a ranking: *smaller values of mean rank score indicate more predictive ranking*.[81] For each vertex centrality metric the mean rank score was averaged for all one thousand test cases to get an aggregate measure of how predictive of future citations a given vertex centrality metric is. A more detailed discussion of this experiment is provided in Appendix A.

The sort experiment[82] described above was run twice, once on the Supreme Court network, and once on the entire Federal Appellate network. For the Supreme Court network, the experiment only looks at citations between Supreme Court cases. In the full Federal Appellate network, citations between Supreme Court cases, between the Supreme Court and the circuit courts, and between the circuit courts are included.

2. Statistical Interpretation of the Methodology

This methodology has two statistical interpretations. The first interpretation is based on approximating the *link prediction* problem, and the second is based on evaluating the rankings of a *recommender system*.

Link prediction is a problem in network science where one builds a statistical model that attempts to predict future links in network based on current information (e.g., the "people you may know" feature on Facebook).[83] Typically, this problem tries to accurately predict new links based on all available information. The sort experiment approximates this problem in two ways. First, unlike the link prediction model, the sort experiment is based only on vertex centrality metrics and ignores other sources of information such as the topic of the case. Second, link prediction models are typically evaluated by accuracy (e.g., what percent of true future links did the model predict), whereas the sort experiment is evaluated by ranking. This ranking methodology is preferable in this circumstance because one does not expect these predictions to be very accurate[84] and it

---

80. Massimiliano Zanin et al., *Preferential Attachment, Aging and Weights in Recommendation Systems*, 2007 Proceedings of the Workshop Net-Works 1, 10–11 (2007).

81. There are other *ranking metrics* such as *reciprocal rank*. See Appendix B for a discussion of these choices.

82. The word sort comes from the fact that the experiment sorts cases by vertex centrality metrics.

83. *Network Analysis and the Law* uses a variant of this link prediction methodology where they use vertex centrality metrics to predict whether or not a case will be cited by a given court or in a given year. Fowler et al., *supra* note 9, at 326.

84. Any classification accuracy rate would likely be very low and therefore noisier.

makes fewer statistical assumptions.[85] Mathematically, one can view a ranking as a relaxation of a probabilistic prediction.

Recommender systems, such as search engines like Google or product recommendations like Netflix or Amazon, rank things to display them in response to a query (e.g., a Google search displays the top ten most relevant results on the first page). The ranking provided by a search engine can be evaluated with some external feedback such as clicks on search results. The sort experiment pretends a new opinion is a "search query" and ranks cases by vertex centrality metrics. The actual citations of the new opinion are then used as feedback to evaluate this ranking.

C. Data

This Comment would not have been possible without the freely available data from CourtListener and the Supreme Court Database ("SCDB").[86] CourtListener describes itself as "a free legal research website containing millions of legal opinions from federal and state courts" that allows "lawyers, journalists, academics, and the public" to "research an important case, stay up to date with new opinions as they are filed, or do deep analysis using our raw data."[87] It contains over three million court opinions from more than 400 jurisdictions[88] and has identified over twenty-five million citations between these opinions.[89]

The SCDB is a freely-accessible database that codifies qualities of Supreme Court cases, including dates of argument and decision, descriptions of litigants, lower court qualities and actions, and issue areas.[90] Widely used by legal scholars,[91] it provides a list of Supreme

---

85. The rankings are non-parametric in the sense that they do not rely on a specific probabilistic model such as logistic regression.

86. We would also like to thank Mike Lissner, lead developer and co-founder of the Free Law Project, for his willingness to troubleshoot and his clear enthusiasm for quantitative legal research.

87. COURTLISTENER, http://www.courtlistener.com [http://perma.cc/38BK-RKMU]. CourtListener is an initiative of the Free Law Project, a non-profit "providing free access to primary legal materials, developing legal research tools, and supporting academic research on legal corpora." FREE LAW PROJECT, https://free.law/ [http://perma.cc/FV9V-W87Y].

88. FREE LAW PROJECT, *supra* note 88.

89. *See* Mike Lissner, *Some Citation Parsing Statistics*, FREE LAW PROJECT (Feb. 17, 2016), https://free.law/2016/02/17/some-citation-parsing-statistics/ [http://perma.cc/3H9J-3HNQ].

90. SUPREME COURT DATABASE, http://scdb.wustl.edu/ [http://perma.cc/S9MN-BY5J]. Where CourtListener is truly a research tool with a Google-esque search box and the ability to read texts and follow links between them, *see* COURTLISTENER, *supra* note 88, the Supreme Court Database is a lower level tool, which is generally presented as a

Court cases and a number of pieces of metadata about each case.[92] The SCDB helps mitigate data quality issues from the CourtListener database.[93]

This Comment applies the methodology to two networks: the Supreme Court citation network and the Federal Appellate citation network.[94] Both networks contain cases from 1791 through part of 2016. The Supreme Court network as analyzed contains 27,885 cases and 235,881 citations. For the Supreme Court network, the nodes are Supreme Court opinions and the edges are citations between two Supreme Court cases.

The Federal Appellate network includes cases from the thirteen federal appellate jurisdictions and citations among these opinions. The Federal Appellate network contains 959,985 cases and 6,649,916 citations. The Supreme Court network is a subnetwork of the full Federal Appellate network.

There are several overall limitations to the network data. First, CourtListener only identified whether there is at least one citation between two cases, rather than counting the number of citations between them.[95] Second, CourtListener did not quantify the quality of a citation: whether the citing case follows, distinguishes, or has another relationship to the cited case.[96] Finally, CourtListener grouped all opinions in a case together, so citations in or to a dissent are not distinguished from citations in or to a majority opinion or a concurrence.[97] If these data become available in the future, it would be interesting to conduct additional experiments to see if the results change significantly.

---

spreadsheet or other structured data file. *See Data*, SUPREME COURT DATABASE, http://scdb.wustl.edu/data.php [http://perma.cc/XCJ7-6KDC].

91. *See, e.g.*, Black & Spriggs, *supra* note 55, at 337; Cross & Spriggs, *supra* note 4, at 451; Yonatan Lupu & James H. Fowler, *Strategic Citations to Precedent on the U.S. Supreme Court*, 42 J. LEGAL STUD. 151, 165 (2013).

92. SUPREME COURT DATABASE, *supra* note 91.

93. *Issue Area*, SUPREME COURT DATABASE: ONLINE CODEBOOK, http://scdb.wustl.edu/documentation.php?var=issueArea [http://perma.cc/JJU7-9W8G].

94. The numbers of opinions and citations in the following paragraphs reflect the data available from CourtListener as of June 27, 2016, slightly reduced by data processing. *See* COURTLISTENER, *supra* note 88. The code used to process the data is available at the GitHub site referenced at the beginning of the Appendix.

95. *Id.* Most legal citation network analysis shares this limitation. *See, e.g.*, Fowler & Jeon, *supra* note 7, at 18.

96. *See* COURTLISTENER, *supra* note 88; *see also supra* notes 7–8 and accompanying text.

97. For a discussion of how this limitation may have impacted our results, see *infra* Section IV.A.2.

III. RESULTS

This Part first discusses the results of the sort experiment on the Supreme Court network. It first examines in-degree driven metrics, out-degree driven metrics, and then time-aware metrics. It also briefly discusses the results of the sort experiment on the full Federal Appellate network, which are largely similar to the Supreme Court results.

Overall, the results agree with prior work that showed that authorities (an eigenvector-based vertex centrality metric) predicted future citations better than in-degree. The results further show that time-aware metrics predict future citations better than time-agnostic metrics, which also corresponds to prior work. Finally, most surprisingly, the results show that out-degree and related metrics predict future citations more accurately than in-degree and related metrics.[98]

A. *Supreme Court Results*

This Section considers the result of the sort experiment run on only the Supreme Court network. That is, the vertices of the network are Supreme Court cases, and only citations between Supreme Court cases are considered. The sort experiment in this context compares vertex centrality metrics by how well they predict Supreme Court to Supreme Court citations.

1. Time Agnostic Metrics

a. *In-Degree Driven Metrics*

Figure 5 shows the comparison between some of the most commonly used vertex centrality metrics: PageRank, in-degree, authorities, and betweenness centrality.[99] Smaller values of mean rank score indicate better prediction of future citations.

---

98. A note about statistical significance for the result: paired t-tests were used to confirm that the differences in mean rank score between vertex centrality metrics are in fact statistically significant. All comparisons that are discussed in the text of this paper were confirmed to be statistically significant with a significance level *alpha* = 0.05. This Comment did not control for multiple testing; however, most p-values were very small (order 10^{-10} or smaller) so any reasonable multiple testing procedure would likely not have changed the conclusions appreciably.

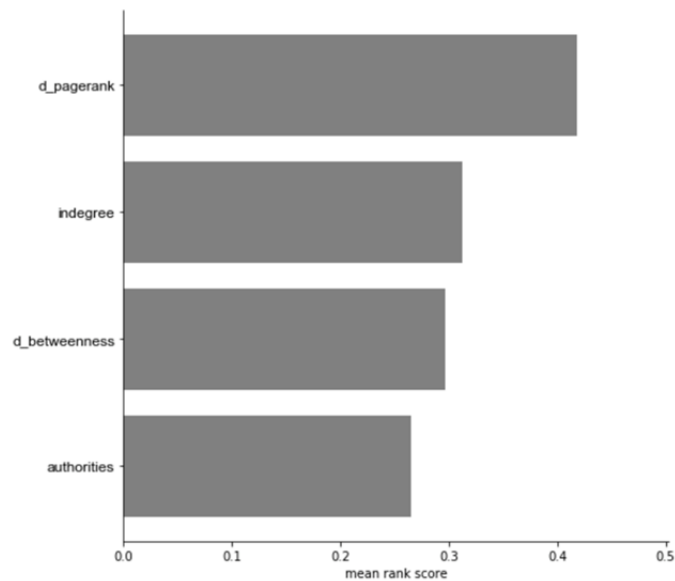99. KOLACZYK, *supra* note 10, at 16, 80, 88, 92–93.

Figure 5: Authorities performs better than the three other in-degree based centrality metrics.

Fowler's *Network Analysis and the Law* compared rankings of Supreme Court cases produced by hubs, authorities, in-degree, and eigenvector based on their ability to predict future citations.[100] Pursuant to their methodology, the authors concluded that the authorities score is the vertex centrality metric that best predicts future citations.[101] Consistent with this conclusion, as shown in Figure 5, authorities beat PageRank, in-degree, and betweenness, which indicates that authorities is more predictive of future citations than these other centrality measures.[102]

b. Out-Degree Driven Metrics

Figure 6 shows the results of the sort experiment for the four in-degree based metrics discussed above and out-degree metrics. Surprisingly, the sort experiment results indicate that out-degree (the number of cases an opinion cites to) is more predictive of future

---

100.   *See* Fowler et al., *supra* note 9, at 338–39.
101.   *Id*. at 343.
102.   *Id*. at 338–39.

citations than in-degree (the number of citations an opinion has received).[103]
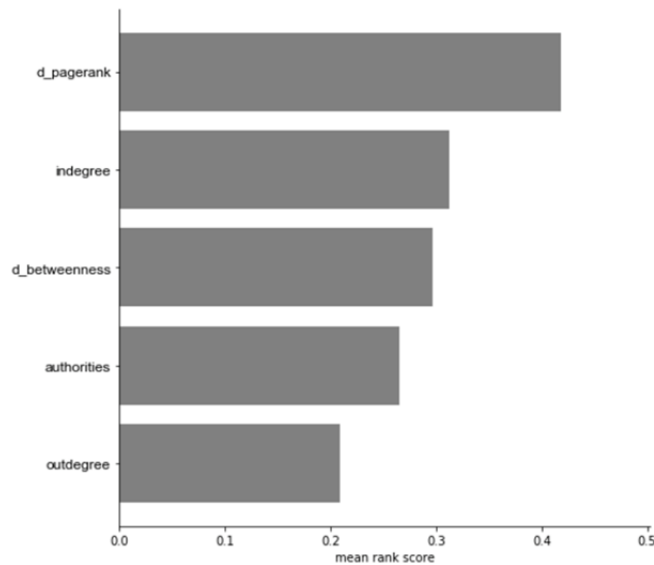

Figure 6: Figure 6 is the same as Figure 5 with the addition of out-degree. Out-degree outperforms in-degree and other more sophisticated vertex centrality metrics.

It seems counterintuitive that the number of cases a judge decided to cite in her opinion would be more predictive of future citations than the number of times other judges have found an opinion worth citing. There are also theoretical mathematical reasons that make this result unexpected, which are discussed in Section IV.A.1.

One possible explanation for out-degree's success is that out-degree is a proxy for case length. This explanation can be at least partially investigated with the current data. This Comment measures the length of an opinion by the number of words appearing in the opinion text. A linear regression, shown below in Figure 7, of number of words versus out-degree resulted in an R^2 value of 36%. Thus, out-

---

103. Additional results relating to vertex centrality metrics, which are driven in part by out-degree, can be found in Appendix A.

degree is related to opinion text length. However, opinion text length does not appear to be the sole driver of out-degree.[104]

Figure 7: This figure shows a scatter plot of opinion text length and out-degree for all Supreme Court cases. The plot includes the linear model fit of out-degree versus number of words (R^2=0.36, p-value <10^{-3}). Note that outliers were first removed by removing the top 1% longest cases and top 1% highest out-degree cases. The linear model found a significant relationship at an alpha level of 0.05. The conclusion is that opinion text length and out-degree are related.

Opinion text length was also included in the sort experiment (i.e., cases were ranked by their text lengths). As shown in Figure 8, opinion text length beats in-degree but does not beat out-degree. To make sure this result is not an artifact of randomness, a pairwise difference t-test found the difference between text length and out-

---

104. Other possible explanations for out-degree's success are discussed in Section IV.A.2.

degree to be statistically significant at alpha level = 0.05. Looking at Figure 8, this difference does appear to be moderate. This result suggests that while opinion text length is important in predicting future citations, out-degree is capturing something beyond just opinion length.
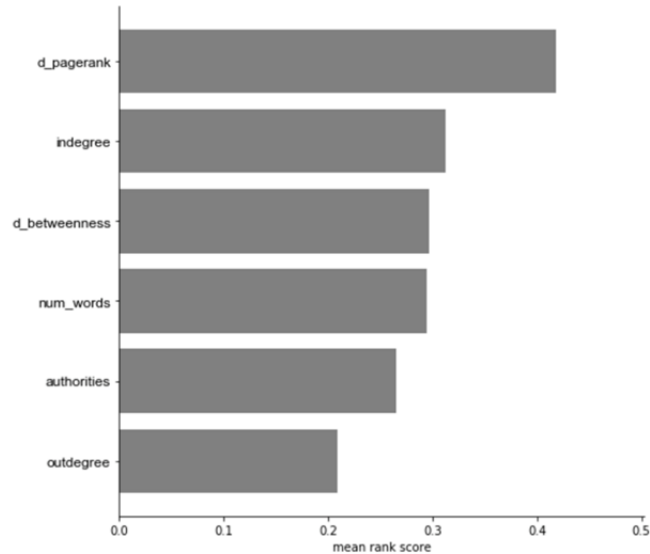

Figure 8: Text length, measured by the number of words contained in the court opinion, does better than in-degree in the sort experiment by a small but statistically significant amount.

2. Time Aware Metrics

None of the metrics considered so far have taken the age of the case into account. A case that received a large number of citations in the 1920s may not be considered relevant today, but that case will still have a large in-degree value. The data from this experiment and others have shown that cases tend to favor citing recent cases.[105] For example, Figure 9 shows a histogram of Supreme Court citation ages (citation age=year of citing case–year of cited case). The distribution of citation ages is strongly skewed to the left with a median citation

---

105.   *See*, *e.g.*, Black & Spriggs, *supra* note 55, at 327.

age of fourteen years; in other words, most citations are to recent cases while only some citations are to older cases.
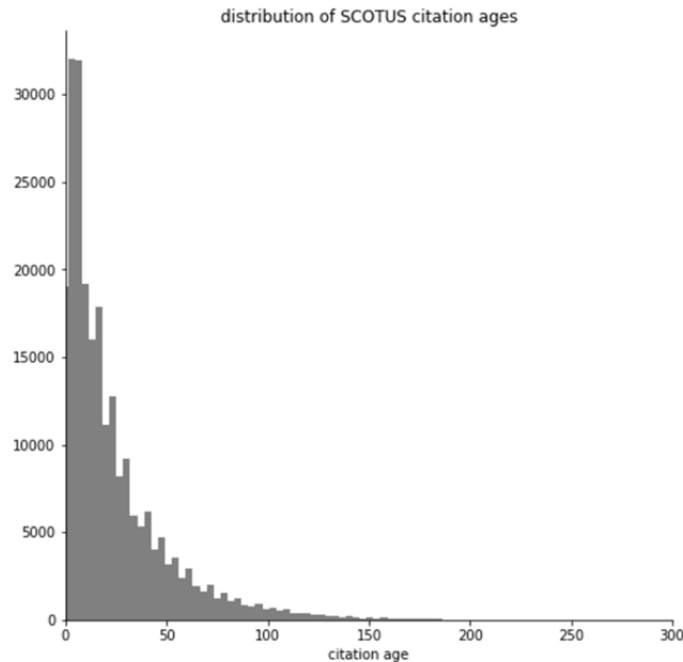

distribution of SCOTUS citation ages

Figure 9: Histogram of citation ages; the Supreme Court favors citing recent cases over older cases.

There is a growing literature on time-aware vertex centrality metrics; however, much of it is beyond the scope of this Comment.[106] This Comment evaluates two time-aware vertex centrality metrics: number of citations in the past several years (referred to as RecentCite) and CiteRank.[107] RecentCite is not a standard name in the networks literature. However, it is a simple way of measuring how important a case is in recent years. CiteRank, which appears in the networks literature, is a modified version of PageRank that takes the age of a vertex into account and decreases the score for older cases.[108]

Both CiteRank and RecentCite are each really a family of vertex centrality metrics because both have a parameter that controls how

---

106. *See* Taylor et al., *supra* note 44, at 538.
107. Walker et al., *supra* note 76, at 3.
108. *Id*.

250   *NORTH CAROLINA LAW REVIEW*   [Vol. 96

heavily older cases are penalized, which can make a significant difference in the case rankings. For example, RecentCite's parameter is simply the cutoff age of whether a citation is counted or excluded. Selecting the value of one of these parameters is beyond the scope of this Comment and is an active statistics research question. Therefore, a range of values is considered for each parameter.

Figure 10 shows the results of the sort experiment including several time-aware centrality metrics. The age of a case is included as a baseline (i.e., cases are ranked by the date of decision).
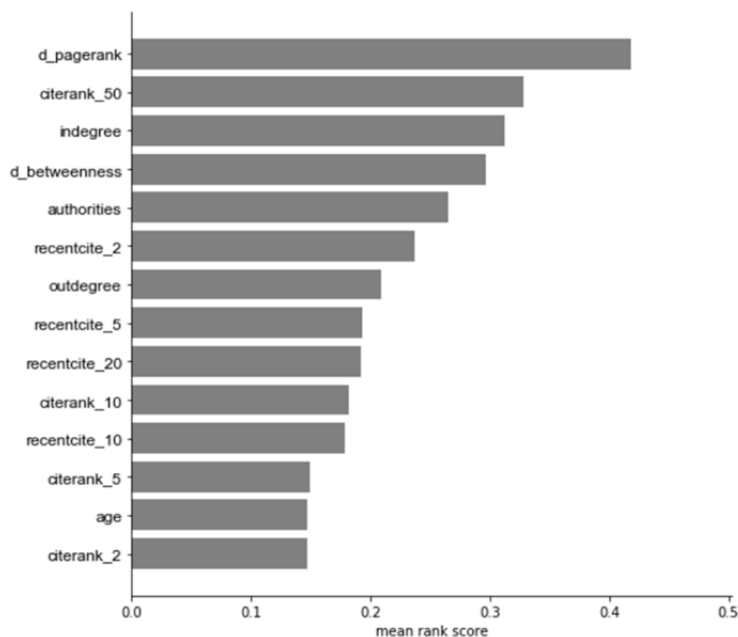

Figure 10: Results of the sort experiment for both time-aware and time-agnostic metrics. Various centrality metrics are on the y-axis and the mean rank score is on the x-axis. Generally, time-aware metrics perform better than the time-agnostic metrics.

The first notable feature of Figure 10 is that most time-aware metrics are better at predicting future citations than the time-agnostic metrics. Based on the citation age distribution in Figure 9, this may not be surprising: opinions favor citing more recent cases. Therefore, explicitly including a recent time bias will make for better predictions.

However, it is surprising that age beats each time-agnostic metric and most time-aware metrics because age ignores all citation information.

Age and RecentCite's performance suggest that many court cases are cited frequently for a period of time soon after they are written and are then cited less frequently. This result is supported by other empirical legal research.[109] Figure 10 gives some information about the time period in which a case is most likely to be cited. RecentCite ten, which only counts citations from the past ten years, beats both RecentCite two and twenty. The rough conclusion to draw from this is that a case is more likely to be cited when it is more than two and less than twenty years old. This corresponds to the findings of other scholars, one of whom has found that the rate at which a case is cited "depreciate[s] about 81 percent and 85 percent between their first and 20th years of age at the Supreme Court and courts of appeals, respectively."[110]

B. Federal Appellate Results

Many of the results discussed above remained broadly the same in the Federal Appellate network sort experiment, with a few exceptions. Figure 11 shows the results of the sort experiment run on the entire Federal Appellate network.[111] The results for in-degree, out-degree, and authorities are all very close.

---

109. *See* Black & Spriggs, *supra* note 55, at 327–28; *see, e.g.*, Neale, *supra* note 56, at 47–48.

110. Black & Spriggs, *supra* note 55, at 325; *cf.* Neale, *supra* note 56, at 2, 47–48 ("[S]tatistical and functional analysis of network rankings of each case over time suggest that [Canadian] cases typically cease to be cited in 3 to 15 years[.]").

111. The results for time-aware metrics were not appreciably different than those for the Supreme Court.
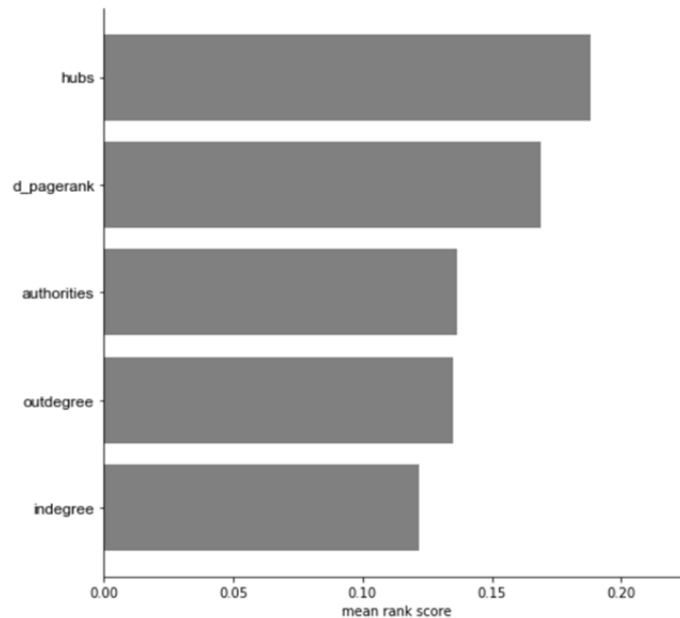
Figure 11: Y-axis is vertex centrality metrics and x-axis is mean rank score. Results of the sort experiment for metrics performed on the Federal Appellate courts: in-degree now does better than both out-degree and authorities. Out-degree beats authorities. Hubs now does worse than in-degree.

In contrast to the results for the Supreme Court, in-degree's predictive power is essentially tied with authorities. Given that in-degree is a simpler metric, this might lead one to prefer in-degree over authorities in the larger network. Furthermore, out-degree is tied with in-degree while hubs score is the worst performing metric. The takeaway is that out-degree still matters, but somewhat less than in the Supreme Court only network.

IV. DISCUSSION

Understanding the notion of legal precedent is necessary to understand how the law evolves.[112] But the notion of precedent can be difficult to quantify—and, therefore, to study—using empirical

---

112. Fowler & Jeon, *supra* note 7, at 16–17.

methods.[113] The legal citation network, however, provides a natural way of studying precedent because "[e]ach judicial citation contained in an opinion is essentially a latent judgment about the case cited."[114] Building on this theoretical foundation, it is a reasonable assumption that identifying case qualities that are predictive of future citation could yield insights into precedent.

A. Out-Degree Beats In-Degree

The most surprising result of the sort experiment is that out-degree beats in-degree: the number of citations *in* a case is more predictive of future citations than the number of citations *to* a case.[115] Furthermore, other out-citation centrality metrics (e.g., hubs, reversed PageRank) beat in-citation metrics (e.g., authorities, PageRank).

Because it seems intuitive that cases that have received a lot of citations in the past are likely to receive more citations in the future, in-degree was expected to do well in the sort experiment. In addition to this intuition, there are theoretical reasons discussed below related to why one might expect in-degree to be a good predictor of future citations. Conversely, it seems counterintuitive that the number of cases a judge decided to cite in his opinion would be more predictive of future citations than the number times other judges have found an opinion worth citing, and there is no obvious explanation for why court opinions that cite more cases might be more influential.

Previous scholarship has suggested an association between out-degree and legal relevance.[116] The results of the sort experiment provide additional evidence that the number of citations is strongly associated with future legal relevance. It is not clear why this association exists or which way causation goes.

1. Preferential Attachment

One statistical theoretical reason why it is surprising that out-degree is more predictive of future citations than in-degree stems from the concept of preferential attachment. In recent decades, researchers have studied time-evolving networks, such as citation networks, both empirically and theoretically.[117] Researchers construct

---

113. *Id.* at 17.
114. *Id.*
115. *See supra* Section III.A.1.b.
116. *See, e.g.*, Cross & Spriggs, *supra* note 4, at 469–71.
117. *See* REMCO VAN DER HOFSTAD, RANDOM GRAPHS AND COMPLEX NETWORKS 1, 159 (Z. Ghahramani et al. eds., 2017).

mathematical models of how a network evolves that explain features of observed networks.[118] One popular class of models is called preferential attachment.[119] The signature feature of a preferential attachment model is that new vertices tend to favor citing cases that already have a lot of citations.[120] Preferential attachment is also referred to as "the rich get richer" phenomenon since cases with a lot of citations will tend to accumulate more citations at a higher rate than cases with fewer citations.[121] Preferential attachment models exhibit topological features[122] that real world networks typically have, such as power law degree distribution.[123]

Preferential attachment models are favored by the networks community because they are one of the few types of simple models that exhibit many of the topological features that real world networks, such as the Supreme Court citation network, tend to exhibit.[124] If one assumes the evolutionary dynamics the Supreme Court network obey some kind of preferential attachment model, then one would expect in-degree to be a very strong predictor of future citations. The sort experiment shows that in-degree is not as predictive of future citations as other quantities such as out-degree or case length. This fact suggests that there is possibly some unobserved or latent quantity that is driving the growth of the citation network in a significant way. Understanding what factors are driving the growth of these legal citation networks is an interesting question from both a statistical and legal standpoint.

2. Case Qualities Possibly Driving Out-Degree

The sort experiment looks at case qualities that are related to the citation network. It is likely that many of these network features are being driven by other case qualities, such as the subject matter of the case, the author of an opinion, or whether a case includes a dissent.

---

118. *See, e.g.*, G. Udny Yule, *A Mathematical Theory of Evolution*, 213 Phil. Transactions Royal Soc'y London 21, 21–22 (1925).

119. Albert-László Barabási & Réka Albert, *Emergence of Scaling in Random Networks*, 286 Sci. 509, 509–11 (1999).

120. *See id*. at 509.

121. *See id.* at 511.

122. A topological feature is a global structural property of a network.

123. *See* Van Der Hofstad, *supra* note 118, at 6, 256, 266. In a power law distribution, a small number of vertices have a large proportion of the total number of edges. *See id.* at 6–7.

124. Both the Supreme Court and Federal Appellate networks exhibit a power law distribution of citations.

Some of these qualities may be related to out-degree.[125] Or, in other words, opinion writers probably do not choose to cite cases on the basis of their out-degree but do chose to cite cases on the basis of qualities that are correlated to out-degree, and it would deepen understanding of precedent to uncover what those qualities are.

One possible quality that out-degree could reflect is that a case with a higher out-degree is better grounded in existing law. A number of legal scholars have hypothesized that judges prefer citing cases that are better grounded in precedent.[126] In particular, Fowler and Jeon provide evidence of the relation between how well-grounded a case is and future citations by looking at citations to and from cases during the Warren Court, the period in which Earl Warren was Chief Justice of the Supreme Court.[127] This period was marked by novel, highly progressive decisions including *Brown v. Board of Education*,[128] *Miranda v. Arizona*,[129] and *Griswold v. Connecticut*.[130] Relatedly, the Warren Court overruled more precedents than any other Court.[131] The fact that the Warren Court broke with existing precedent would mean that their opinions were not grounded in existing law may and potentially reflect that by having lower average out-degree.

As expected, Fowler and Jeon observe that the Warren Court shows a drop in out-degree. "Since the process of creating new law frequently involves breaking with existing precedent, it is no surprise that the Warren Court cited fewer cases in their opinions."[132] However, they also observe that the Warren Court shows a drop in in-degree.[133] In other words, the Warren Court cited fewer Supreme Court opinions and is cited less frequently by future Supreme Court opinions. Fowler and Jeon suggest that the Warren Court's tendency to break from precedent meant its opinions had "weak legal basis," which is reflected in the drop in out-degree, and which is, in turn, the

---

125. This is not to say that the qualities discussed in this Comment are the sole possible drivers of out-degree's performance. For example, the number of legal topics a case addresses could be driving out-degree's performance. It is also not to say that any one of these qualities is exclusively responsible for out-degree's performance.

126. *See* Cross & Spriggs, *supra* note 4, at 467–68, 480; Lupu & Fowler, *supra* note 92, at 152–53.

127. *See* Fowler & Jeon, *supra* note 7, at 19.

128. 347 U.S. 483 (1954).

129. 384 U.S. 436 (1966)

130. 381 U.S. 479 (1965); *see also* Fowler & Jeon, *supra* note 7, at 19.

131. *Id.*

132. *Id.*

133. *Id.*

driving force behind the Warren Court's lack of citation by subsequent Courts.[134]

The data here show the same patterns. Figure 12 shows the median in-degree and median out-degree by year for Supreme Court cases.[135] The vertical bars show the timeframe of the Warren Court. During the Warren Court the typical out-degree and the typical in-degree both dip (i.e., future Courts appear to avoid citing Warren Court cases to some extent).
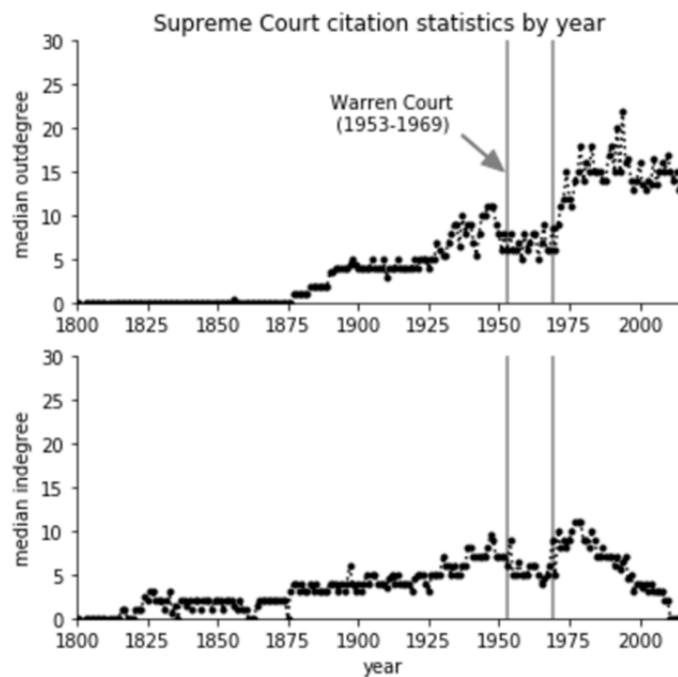

Figure 12: This figure shows the median in-degree and out-degree of Supreme Court cases by year.[136] The Warren Court, which lasted from 1953 to 1969, is visible in the dip in in-degree, out-degree, and case length.

---

134. *Id.* Fowler and Jeon also speculate briefly that the lack of citation to the Warren Court may instead reflect the conservative policy preferences of the Burger and Rehnquist Courts that followed. *Id.* at 19–20.

135. *Id.* at 19.

136. Median was selected instead of mean because median is more robust to outliers.

The results of the sort experiment also arguably support the claim at a broader level. If out-degree corresponds generally to how well-grounded an opinion is, then out-degree's predictive power can be understood as demonstrating a preference in the Supreme Court for citing opinions that are well-grounded in existing law. The Warren Court correlation between in-degree and out-degree would be part of a larger trend rather than an end in and of itself.

Another quality that may be driving the sort experiment's out-degree results is the presence of dissents and concurrences in a case. Because the data used in the sort experiment groups all opinions in a case together, citations in a dissent or to a dissent are not distinguished from citations in or to a majority opinion or a concurrence.[137] Therefore, a case that contains multiple opinions might have a higher out-degree than a unanimous opinion of equivalent length because opinions coming to a different conclusion, or to the same conclusion but for different reasons, would likely cite to different bodies of cases to support their reasoning.

This observation still leaves unanswered the question of why cases with multiple opinions would be cited more frequently than unanimous opinions. One scholar has demonstrated that the Supreme Court tends to cite "paradoxes"—decisions with a controlling majority as to result—but not as to the grounds of that result.[138] This tendency reflects the way the Court will incrementally arrive at a rule in a contested area of the law.[139] In contrast, the circuit courts are more likely to cite stable Supreme Court precedent, as opposed to multi-opinion, fractured decisions.[140]

This difference is possibly reflected in the sort experiment. Out-degree was more predictive of future citations than in-degree and more sophisticated metrics in the Supreme Court network. However, in the full Federal Appellate network, out-degree was only as predictive of future citations as in-degree. This could reflect the difference in citation preference between the Supreme Court and the circuit courts. A next step to further explore this possibility would be to re-run the sort experiment on the network of the circuit courts (or

---

137. *See supra* Section II.B.

138. *See* Hitt, *supra* note 7, at 59. Hitt describes *National Mutual Insurance v. Tidewater Transfer*, 337 U.S. 582 (1949), as a quintessential paradox: a majority of justices held that citizens of D.C. could sue in federal court under diversity jurisdiction. But no majority existed on the underlying logic: two justices found in favor of diversity jurisdiction on constitutional grounds, three justices found in favor of jurisdiction on the ability of Congress to grant it. Hitt, *supra* note 7, at 67–68.

139. *See id.* at 61.

140. *See id.* at 59–61.

258   *NORTH CAROLINA LAW REVIEW*   [Vol. 96

to remove citations by the Supreme Court in the full Appellate network), to see if out-degree becomes even less predictive.

Alternatively, it could be that the influence of out-degree is being diluted in the Federal Appellate network by long opinions produced by the circuit courts with very low precedential value, in particular appeals of right in criminal cases with multiple defendants. These cases require very long opinions but are less likely to advance, clarify, or re-shape the law in a substantive way.[141] Table 1 lists the ten cases with the highest out-degree in the Supreme Court network and the full Federal Appellate network. Nine of the ten cases with the highest out-degree in the appellate network are multi-defendant criminal cases, which could indicate that whatever quality is driving the performance of out-degree is being diluted by necessarily long circuit court opinions.

---

141. *See* FED. R. APP P. 3 (outlining the procedures for appeals of right to the circuit courts).

Highest Out-Degree Supreme Court		Highest Out-Degree Federal Appellate		
Case Name	Year	Case Name	Year	Court
Miller Brothers Co. v. Maryland	1954	United States v. Haldeman	1977	D.C. Cir.
Commissioner v. Estate of Church	1949	United States v. Decoster	1979	D.C. Cir.
Baker v. Carr	1962	United States v. Alvarez	1987	9th Cir.
Nebbia v. New York	1934	United States v. Byers	1984	D.C. Cir.
McGautha v. California	1971	United States v. Phillips	1886	5th Cir.
Crowell v. Benson	1932	United States v. Mitchell	2007	9th Cir.
Communist Party of United States v. Subversive Activities Control Board	1961	United States v. Moore	2011	D.C. Cir.
The Minnesota Rate Cases	1890	United States v. Mitchell	2007	9th Cir.
Fay v. Noia	1963	United States v. Rigoberto	1988	7th Cir.
Oregon v. Mitchell	1970	Ruiz v. Estelle	1982	5th Cir.

Table 1: This table shows the top ten cases by out-degree for the Supreme Court and in the Federal Appellate network. Nine of the top ten cases as ranked by out-degree in the appellate network are multi-defendant criminal cases.

B. Time Awareness Improves Prediction of Future Citations

The sort experiment also indicated that time-aware centrality metrics are more predictive of future citation than time-agnostic centrality metrics. In other words, incorporating information about

how recently a case was decided into a metric improves the ability to predict whether a case will be cited in the future. This result is not surprising but does benefit from further explanation. From a doctrinal perspective, it makes sense that more recent decisions would bear more strongly on current disputes.[142] Given the principle of stare decisis and the analogical processes by which precedent is applied to current disputes, it stands to reason that a more recent decision would capture the nuances of past decisions, providing the most useful "basis for determining later cases involving similar facts or issues."[143] This principle is often articulated in first-year legal-writing courses, which instruct students to prefer newer cases when selecting authorities to use in an argument.[144]

Time plays a large role in the evolution of the citation network (e.g., Figure 9). For legal scholars interested in network analysis, it is worth looking into the growing literature about temporal vertex centrality metrics.[145]

C. PageRank and Questions of First Impression

Given PageRank's success in ranking web-pages,[146] one might expect PageRank to do well in the sort experiment. However, the network topology of the Internet is different from the network topology of a citation network. In a citation network, unlike the Internet, edges can only go in one direction: backwards in time.[147] In other words, while two web pages may link to each other, two cases will only very rarely cite each other.[148] In a network like a citation network, PageRank is known to be biased in favor of older vertices.[149] For an explanation of why this bias occurs, see Appendix C.

---

142. *See* Landes & Posner, *supra* note 1, at 250.

143. *See Precedent*, *supra* note 28 (quoting WILLIAM M. LILE ET AL., BRIEF MAKING AND THE USE OF LAW BOOKS 288 (Roger W. Cooley & Charles Lesley Ames eds., 3d ed. 1914)).

144. *See, e.g.*, ALEXA Z. CHEW & KATIE ROSE GUEST PRYAL, THE COMPLETE LEGAL WRITER 63–64 (2016).

145. Taylor et al., *supra* note 44, at 538.

146. *See* Bryan & Leise, *supra* note 67, at 569–70.

147. Formally, this quality of a citation network makes it a *directed acyclic graph*. *See* ERIC SINK, VERSION CONTROL BY EXAMPLE 47–51 (Brody Finney ed., 2011) (defining the data structure of directed acyclic graphs).

148. The primary exception to this rule is when the Supreme Court releases two opinions that reference each other on the same day. *See, e.g.*, Washington v. Glucksberg, 521 U.S. 702, 751 (1997) (citing Vacco v. Quill, 521 U.S. 793 (1997) (noting explicitly that two cases were decided on the same day)).

149. *See* Manuel Sebastian Mariani et al., *Identification of Milestone Papers through Time-Balanced Network Centrality*, 10 J. INFORMETRICS 1207, 1208 (2016), https://arxiv.org/pdf/1608.08414.pdf [http://perma.cc/YX7F-ZBD3].
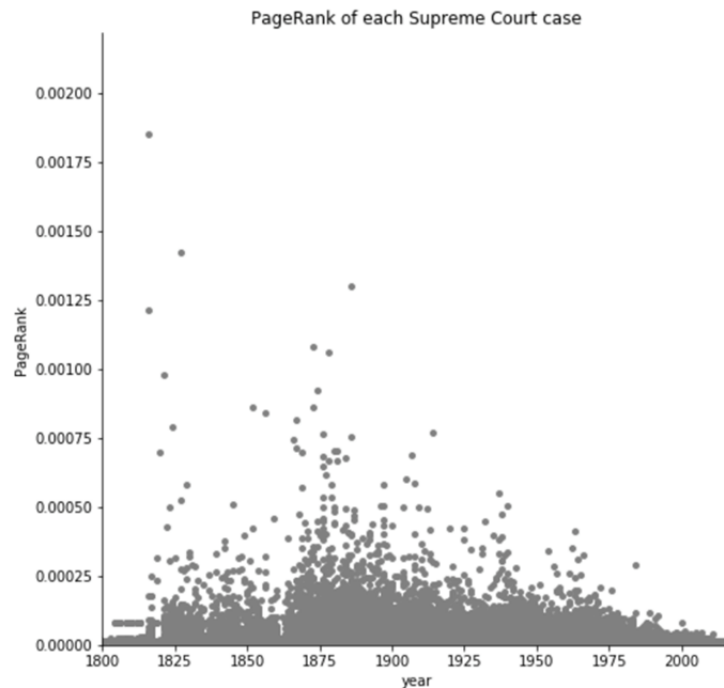
Figure 13: PageRank is biased to favor older cases. This is a plot of each case's PageRank value versus the year that case occurs.

Figure 13 shows that PageRank favors older cases. The time bias makes PageRank a particularly bad metric for predicting future citations, since, as discussed in Section III.A.2, the Supreme Court prefers citing recent cases. PageRank does, however, seem to pick up on an important quality of case law: questions of first impression. A question of first impression is a legal issue that has not been addressed by the court before. It often involves the first interpretation of a statute or constitutional provision.

Top Ten Supreme Court Cases Per PageRank	
Case Name	Date Decided
Gibbons v. Ogden	1816-03-2
Martin v. Hunter's Lessee	1816-03-20
McCulloch v Maryland	1819-03-18
Brown v. Maryland	1827-03-12
Boyd v. United States	1886-02-01
Slaughter-House Cases	1873-04-14
Davidson v. New Orleans	1878-01-18
Cohens v. Virginia	1821-03-18
Ex Parte Lange	1874-01-30
Cooley v. Board of Wardens of Port of Philadelphia	1852-03-18

Table 2: Top ten cases by PageRank.

Consider the top ten cases ranked by PageRank, listed above. At least four of the top ten presented questions of first impression to the Supreme Court or established fundamental principles of constitutional law. *Gibbons v. Ogden*[150] is the foundational case for interpretation of the Commerce Clause,[151] and *McCulloch v. Maryland*[152] established that the federal government may exercise powers not specifically enumerated in the Constitution.[153] Similarly, *Martin v. Hunter's Lessee*[154] established the Supreme Court's ability to review the decisions of state supreme courts.[155] *The Slaughter-House*

---

150. 22 U.S. (9 Wheat.) 1 (1824).
151. *Id.* at 235–40.
152. 17 U.S. 316 (4 Wheat.) (1819).
153. *Id.* at 436–37.
154. 14 U.S. 304 (1 Wheat.) (1816).
155. *Id.* at 324–25.

Cases[156] contain the Supreme Court's first interpretation of the Fourteenth Amendment,[157] which is at the heart of many of the most famous twentieth and twenty-first century cases regarding individual rights, including *Brown v. Board of Education*,[158] *Roe v. Wade*,[159] and *Obergefell v. Hodges*.[160]

D. Deciding Which Vertex Centrality Metric to Use

As mentioned above, there are a growing number of legal practice areas in which lawyers can benefit from a baseline understanding of the principles underlying legal tools.[161] In some practice areas, such as e-discovery, lawyers may even have to make decisions on technical statistical issues. However, rather than attempting to identify the "best" vertex centrality metric, this Section discusses a range of considerations should be used when deciding which vertex centrality metric to employ in a given circumstance.

As a starting point, an algorithm (such as a vertex centrality metric) will always produce some kind of answer (such as a ranked list of cases). But the fact that the answer has been generated does not indicate that the answer is meaningful or trustworthy. At one level, this Comment interrogates the answers given by a set of algorithms, vertex centrality metrics, and attempts to determine whether the answers given by them are meaningful, and in turn what those answers mean.

But more broadly, the answers produced by the methodology are themselves subject to question. The central assumption of this methodology—that a good vertex centrality metric has the ability to predict future citations—may not be the most appropriate starting assumption in picking a vertex centrality metric. A legal historical study, for instance, may seek to identify factors driving why cases are consistently cited over time, which this methodology would not help with.[162] Awareness of starting assumptions and qualities of statistical tools help ground the scope of questions that can be asked, and in turn the situations in which a tool might be appropriately used.

---

156. 83 U.S. 36 (16 Wall.) (1872).
157. *Id.* at 37–38.
158. 347 U.S. 483, 490 (1954).
159. 410 U.S. 113, 152–53 (1973).
160. 135 S. Ct. 2584, 2593 (2015).
161. *See supra* notes 24–27 and accompanying text.
162. *See supra* Section IV.C (discussing an important quality of cases that this methodology does not do well in identifying).

To give a more concrete example from the present study, the authorities score is more predictive of future citations than in-degree according to the methodology. However, Occam's razor—the proposition that the simplest solution to a problem is most likely the best one[163]—suggests one should prefer a degree-based metric such as in-degree since they are the simplest metrics unless one of the other metrics performs substantially better. In-degree is less conceptually complicated than authorities; authorities is harder to interpret and requires an understanding of higher math. Additionally, simple algorithms tend to be preferred over complex algorithms, as the more complex an algorithm is, the more things can go wrong because they can be more sensitive to noise and more easily statistically biased.[164]

To give another example from this Comment, simply because the methodology shows that out-degree is more predictive of future citations than other measures does not necessarily mean it should be the centrality measure of first choice for a search engine or other predictive tool. Out-degree is simple, but the connection between out-degree and future citations is relatively opaque. For example, if out-degree's predictive performance is relatively unique to the Supreme Court, or if out-degree is a proxy for opinion length or number of topics discussed in an opinion, implementing out-degree as a ranking tool would privilege case qualities that are generally not considered important in the context of legal research.

Outside the context of this Comment, a growing number of e-discovery proceedings employ statistical machine learning processes.[165] Attorneys managing these processes would be better equipped to advise their client and direct employees and contractors with an understanding of the assumptions these statistical processes are built on. And as the number of areas of practice which rely on machine learning and other statistical processes grows, lawyers will be asked to reason about topics like vertex centrality more frequently.[166]

---

163. *See Occam's razor*, OXFORD ENGLISH DICTIONARY ONLINE, http://www.oed.com /view/Entry/234636?redirectedFrom=occam%27s+razor& [http://perma.cc/5DJ5-HYY2].

164. *See* GARETH JAMES ET AL., AN INTRODUCTION TO STATISTICAL LEARNING WITH APPLICATIONS IN R 24–36 (2013).

165. *See* Henderson, *supra* note 24, at 41.

166. *Cf.* Lohr, *supra* note 25 ("[T]he law firm partner of the future will be the leader of a team, 'and more than one of the players will be a machine.'" (quoting Michael Mills, a layer and the Chief Strategy Officer of a legal technology start-up)).

CONCLUSION AND POSSIBLE FUTURE WORK

This Comment introduces a methodology that evaluates the predictive power of vertex centrality metrics. It hypothesizes latent qualities that could be driving the performance of vertex centrality metrics, such as case length and questions of first impression. A number of potential future research questions, both technical and legal, are discussed in the Appendix.

The sort experiment methodology compares vertex centrality metrics by how well they predict future citations of a case. For a given test case, all previous cases are ranked by a given centrality metric. The methodology then looks at the actual citations of the test cases and where they land in this ranking. The idea is that better centrality metrics will tend to put the cited cases closer to the top of the ranking. The sort experiment is one way of using data to compare vertex centrality metrics. It can be interpreted as evaluating each metric's ability to predict future citations. It can also be interpreted as evaluating a metric's ability to rank cases for a search engine.

The most surprising finding of the sort methodology is that out-degree is more predictive of future citations than in-degree. This result may be evidence for the importance of precedent. It is possible that the number of citations in an opinion (out-degree) is a proxy for how well-grounded in precedent that opinion is. It is also possibly a proxy for whether an opinion contains a dissent or concurrences. Significant additional questions remain in this line of research, including whether the performance of out-degree is unique to the Supreme Court. The methodology could also be improved by use of more nuanced data, such as counting the number of citations between opinions, rather than just the fact of one opinion citing another. Most broadly, citation analysis is a powerful tool that has the potential to illuminate a great deal about the structure and evolution of the law.

IAIN CARMICHAEL, JAMES WUDEL, MICHAEL KIM & JAMES JUSHCHUK[**]

APPENDIX

The code which deals with the Legal Citation network and runs the experiments can be found at: https://github.com/idc9/law-net/tree/master/examining_evolution_code.

A. Vertex Centrality Metrics

This Comment focuses on the directed version of the citation network but also considers the undirected version of the network. Most vertex centrality metrics considered are defined for both directed and undirected networks. Ignoring the edge direction means treating citations going into a case the same as citations going out of a case. For example, the degree (undirected) is equal to in-degree plus out-degree. Given the results about out-degree, undirected metrics may be a reasonable choice over directed metrics. This Comment briefly considers the "reversed" network in which the direction of the citations is reversed. This is done primarily to look at a reversed version of PageRank. Some vertex centrality metrics are driven by in-degree (e.g., PageRank, authorities) and other centrality metrics are driven by out-degree (e.g., reversed PageRank, hubs).

Undirected and reversed metrics are considered because of the surprising performance of out-degree. The undirected and reversed metrics tend to perform well, which is further evidence for the importance of out-degree in the evolution of the citation network.

Two time-aware vertex centrality metrics are used: CiteRank[167] and the number of recent citations. CiteRank is similar to PageRank but down weights older cases. In particular, instead of a uniform "jump" distribution, when CiteRank makes a random jump it selects a new vertex C with probability proportional to $2^{-\frac{age(C)}{H}}$ (i.e. exponentially decaying based on case age with *half-life = H*). For the latter metric, the in-degree is computed for each case, but only counting citations that occurred in the most recent K years. The latter is a simple measure of how popular a case is at a given moment in time.

B. Sort Experiment

This Section discusses some details of the sort experiment. One thousand test cases are selected uniformly at random from all cases between 1900 and 2016, excluding cases that cite zero other cases (i.e., that have zero out-degree).

---

167. Walker et al., *supra* note 76, at 3.

For each test case, we extract the subnetwork snapshot just before the test case occurs. For example, for a test case on May 15, 1990, we look at the citation network of all cases that occur before May 14, 1990 and call snapshot of the network just before the arrival of the test case. For every case in this network snapshot, we compute each vertex centrality metric we are interested in. Some of the vertex centrality metrics are very computationally intensive, so computing them over and over again takes a long time. We reduce the computational burden by looking at network snapshots once each year from 1900 to 2016 (i.e., look at 116 subnetworks instead of 1000 subnetworks). We then use these annual values to approximate the true values of the centrality metrics at the time of each test case.

The sort experiment compares a ranking of each case with the cases that were actually cited. To compute how well this ranking performed given the actual citations, we use a ranking metric. There are a number of standard ranking metrics we considered: precision, recall, precision at K, and reciprocal rank.[168]

We selected *mean rank score*, which is defined as follows.[169] Suppose we have a ranking of N cases. Suppose K cases are selected and are ranked $R_1, \ldots, R_k$. The mean rank score is then $\frac{1}{K}\sum_{i=1}^{K}\frac{R_i}{N}$. The smaller the typical rank, the lower the mean rank score. A random ranking would give a mean rank score value of around 0.5.

Most of the above ranking metrics are used for search engines where one expects the selected results to be near the top of the list. We do not expect a simple vertex centrality metric to place the cited cases near the top of the list. However, we do hope a centrality metrics captures some signal, making the mean rank score more appropriate. We computed all of the above ranking metrics to make sure our results were not sensitive to the particular evaluation choices we made. The results were not qualitatively different for different metrics.

C.   PageRank Time Bias

A citation network is a directed acyclic graph (DAG). This Section explains why PageRank is biased to favor older cases in a DAG.

This bias is true because of the way PageRank is defined. One way to describe PageRank is using a random walk around a network.

---

168. KEVIN P. MURPHY, MACHINE LEARNING: A PROBABILISTIC PERSPECTIVE 303–04 (2012) (explaining how mean reciprocal rank can be used as a ranking metric).
169. *See* Zanin et al., *supra* note 81, at 10.

The World Wide Web is a collection of web pages and the links between them. Consider surfing the web for a very long time and jumping from one web page to the next in the following random way. Say you are on webpage X; with probability 0.85, follow one of the links coming from X to one of the webpages X links to; otherwise, with probability 0.15, pick any web page online at uniformly at random.

The PageRank value of a given web page is the proportion of the time the random walk spent at that web page. The intuition is that the more a page is linked to, the more likely the random walk will land on that web page. Furthermore, the more a page X is linked to by pages that are themselves linked to by many pages, the more likely the random walk will land on page X.

For the citation network, PageRank follows citations with a similar random walk. Most random steps follow a citation and go backward in time. This means the random walk will spend more time on older cases.

There are number of other vertex centrality metrics that are driven, at least in part, by out-degree such as hubs or undirected versions of any directed vertex centrality metrics.[170] By ignoring the direction of citations, a citation network can be viewed as an undirected network. In this case metrics such as degree are driven by a combination of both out- and in-degree. Figure 14 shows that undirected metrics out-perform directed metrics. It is likely this boost in performance comes from the addition of out-degree.

---

170. Up until now, the citation network is considered to be a directed graph (i.e., edges go from one case to another case). The citation network can be viewed as an undirected network by ignoring the citation direction.
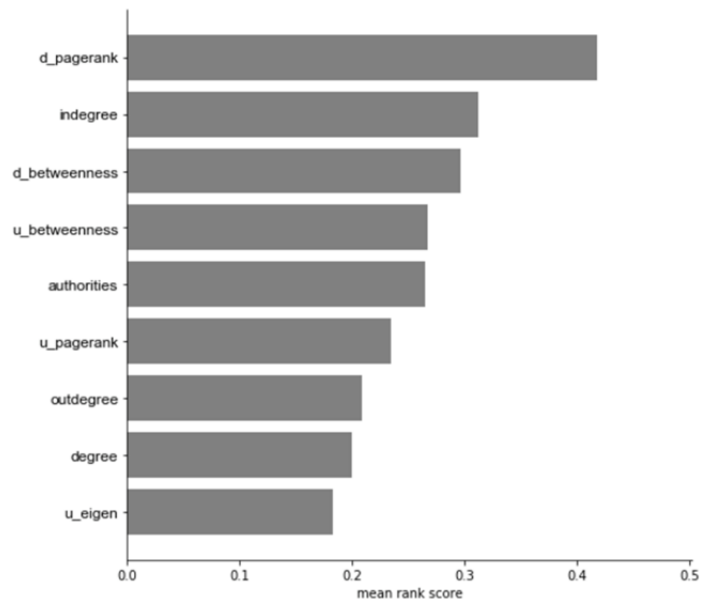
Figure 14: Results of sort experiment for PageRank and Hubs on a reversed graph compared to previous metrics. Hubs performed the best among these metrics and reversed PageRank performed better than all but out-degree and Hubs.

Figure 14 shows that Hubs and reversed PageRank all beat in-degree and the other in-citation driven metrics. Hubs and reversed PageRank are associated with out-degree and it is likely that their success in the sort experiment is driven by out-degree.