Modeling the Influence of Legal Opinions

Ravi Parikh
Email: rparikh@stanford.edu

George Tang
Email: gtang7@stanford.edu

Keegan Poppen
Email: kpoppen@stanford.edu

***Abstract*—Although legal case law has been recorded for over 200 years, and we can easily see the socially and politically influencial cases, there is no current definitive measure of a legally influencial case other than citation count. In this paper we explore different models for measuring the influence of disruptive legal documents over history using graph analysis techniques such as PageRank. We show that naive pagerank is inadequate in accounting for the temporally significant nature of case law, and propose two separate decay models in order to more accurately reflect the characteristics of our dataset. Lastly, we will attempt to verify our results for the most influential cases with cluster analysis.**

I. Introduction

A "disruptive" legal document or opinion is one that drastically changes the future legal landscape, by overturning or establishing precedent in an important area. The impact of disruptive legal documents throughout history cannot be understated. Our goal is to construct a model to identify the changing influence of such documents: how did the influence of document X change over time? Has it become more or less relevant to the US legal landscape? Furthermore, can we compare its changing influence to that of other related cases? Examples of when this might be relevant are when landmark law overturns old precedent, as *Brown vs. Board of Education* did to *Plessy vs. Ferguson*.

Our dataset is the Stanford Law dataset, which consists of 6.5 million legal opinions from the US judiciary from 1800 to present at state and federal levels. In addition, our data includes links from any given legal opinion to any other legal opinions that it cites, allowing us to generate a citation graph. The comprehensiveness and size of the dataset makes it ideal for our goals. Our starting point is the PageRank algorithm to naively identify influence in the citation graph, with the understanding that a high Pagerank indicates a strongly influential case; however, several characteristics of our dataset necessitate a more refined model. Furthermore, older cases may generate many citations in their era but lose relevance later, a characteristic that naive PageRank does not address. Our model to correct for these issues is discussed in this paper.

We present our process and our results below. Section 2 is a discussion of previous work in the field; section 3 discusses our data and the various transformations we applied; section 4 concerns our use of PageRank to do before-after analysis; section 5 is our results; section 6 is a discussion of the implications of our research; section 7 is a discussion of future work; and section 8 concludes this paper.

II. Previous Work

There are many previous and related works. We consulted the original PageRank paper (Brin, Page, Motwani, and Winograd; 1998) for a primer on the algorithm. Their algorithm was of course developed for web hyperlinks. It iteratively determines a numerical weighting for each document in a network of linked documents. Ding and Yan leveraged PageRank for ranking authors in co-citation network, that is a network of where researchers were linked if they co-authored a paper. One of their key insights was to use damping factors, which informed our own model.

III. Data Set and Transformations

We only considered United States Supreme Court cases for our model. This was because the entire citation graph includes a lot of decisions made at the state rather than federal level. Many of these cases are unimportant in the national sense because they concern state-specific law. Furthermore, a landmark ruling in California will receive citations from other California cases, of which there are many, while such a ruling from Alabama will receive many fewer citations because of the smaller population of Alabama. Thus it is very difficult to compare cases across state boundaries. Instead, we chose to focus only on national Supreme Court cases. Another reason for limiting our data set was that there were only approximately 80,000 Supreme Court cases during the time period concerned, while there were over 6.5 million total cases. The smaller number of cases allowed us to focus more closely on qualitative analysis, which was necessary for this project due to the lack of a gold standard data set or a standard heuristic against which to evaluate the performance of our model. Instead, we were limited by only being able to qualitatively analyze the success of our model and develop rough but imperfect metrics for quantitative analysis. We also did not consider any of the natural language features of our dataset; while we did have access to the legal briefs for each one of these cases, we only considered the citation graph primarily. Nodes were also labeled with the date on which the legal opinion was rendered, allowing us to do things like create time slices of our data (e.g. only consider cases before 1950). After generating the citation graph and running our models, we were then able to look at them qualitatively by matching case IDs to case titles and leveraging online legal resources.

IV. PageRank

We will present a brief review of the PageRank algorithm. The PageRank algorithm can be presented as a state space

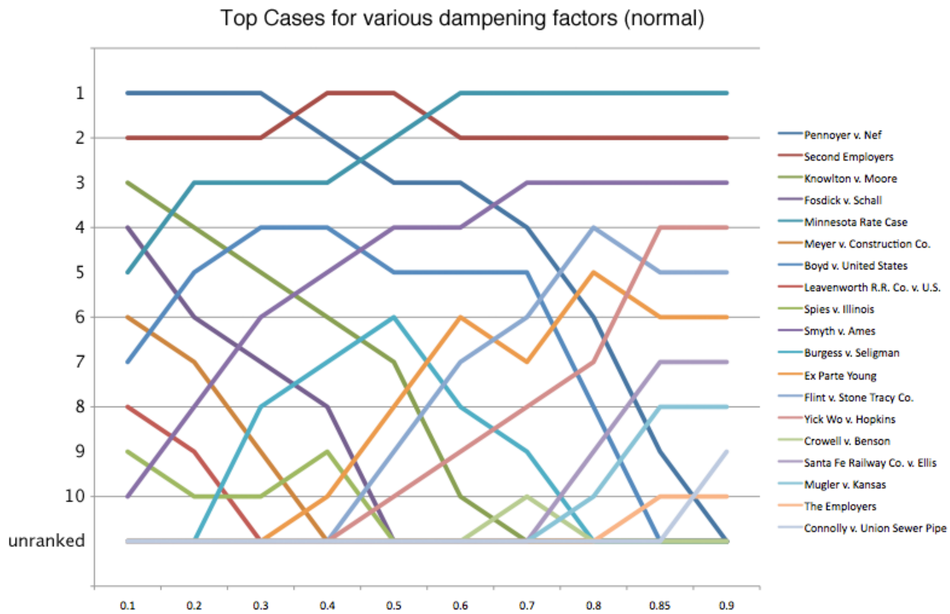

Fig. 1. Graph of top 10 cases for normal pagerank as we vary the dampening factor.

system of form $x(n+1) = dTx(n) + b$ (Ding and Yan, 09), where the state $x(n)$ is a column vector of length v and the sum of all components of $x = 1, T$ is a $v \times v$ matrix on R^v whose entries are all positive and each column sums to one, where v is the number of nodes in consideration. The components T_{jk} of T represent the probability that a random surfer will follow a link from Webpage k to Webpage j. The jth component $x_j(n)$ of $x(n)$ is the probability that the surfer is at jth node at time n. d is the dampening factor, or the probability that the surfer will follow a link on the page to another webpage. $1-d$ is the probability that the surfer will go to a new webpage not through a link on the current page. The jth component b_j of b is the probability that the surfer will choose jth webpage without following any links. To a certain extend, Pagerank works particularly well for our goal. The measure of influence of a case is directly correlated with how many cases cite it, as well as the strength of the influence of the case that cites it. The cases with high pagerank are very likely going to be high influence cases.

V. Modified PageRank Model

The dampening factor, which estimates probability of jumping to a random webpage without a link, is more or less irrelevant for our dataset (there is no meaning to a random jump in citation). Thus, our qualitative hypothesis is that the higher the dampening factor (less probability of random jump), the more accurately the pagerank values will reflect influence. We will attempt to show this quantitatively as well, by experimenting with pagerank models that vary the dampening factor d from 0 ¡ d ¡ 1.

One problem with the original PageRank is that it places the same weight on every edge in the dataset. The even distribution of weights to links does not always accurately reflect real world circumstances, specifically in a time-sensitive dataset like ours. Often, older cases that generated many citations in their era are no longer relevant today, and may even be incorrectly identified as more "influential" than modern cases that overturned their precedent. One such example is the 1877 Supreme Court ruling *Pennoyer v. Ness*, in which it was ruled that "A court may enter a judgment against a non-resident only if the party 1) is personally served with process while within the state, or 2) has property within the state, and that property is attached before litigation begins" (Lawnix). This case was considered very influential at the time of its ruling, and generated 191 citations in our graph. However, subsequent rulings over the years greatly modified the prevailing precedent in this area, and the case has lost relevance today, though is still a popular legal academic subject (Borchers).

We will explore mathematically how to add weights to our PageRank implementation that will reflect our case relationships more accurately.

First of all, we segmented our data into 20 different sections, each one containing the cases and the citations before 1810, 1820, 1830 ... 2010. With these segments of data, we can run pagerank over the segments and watch the development and changes in influence of a case over time.

We propose two separate models to test on the segmented data; a Linear decay weight and an Exponential decay weight modification on Pagerank.

Linear Decay Weighting of Edges

$$PR(A) = \frac{1-d}{N} + d\frac{w(a)-\min(w_i)}{\max(w_i)-\min(w_i)}\left(\frac{PR(B)}{L(B)} + \frac{PR(C)}{L(C)} + \cdots\right)$$

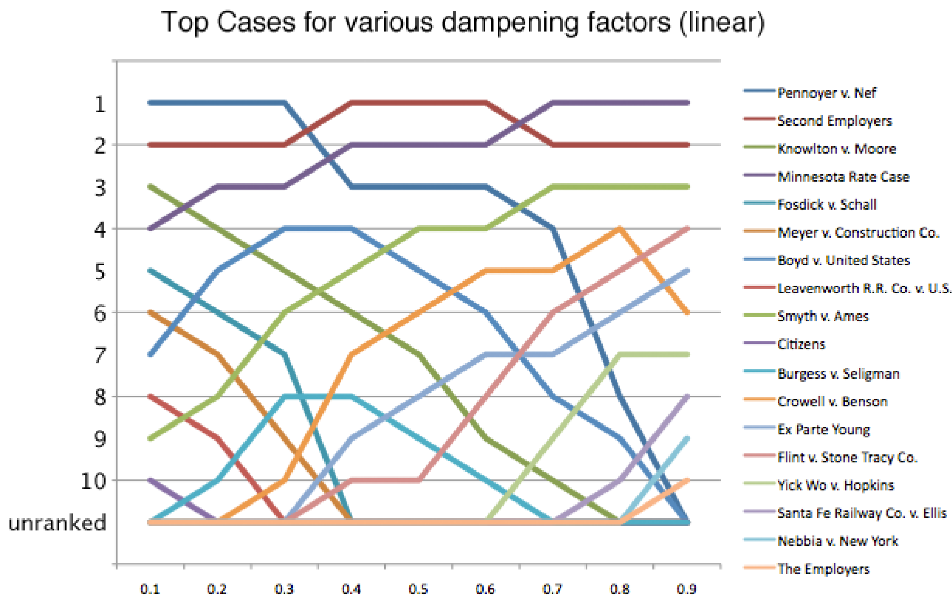

Fig. 2. Graph of top 10 cases for linear decay pagerank as we vary the dampening factor.

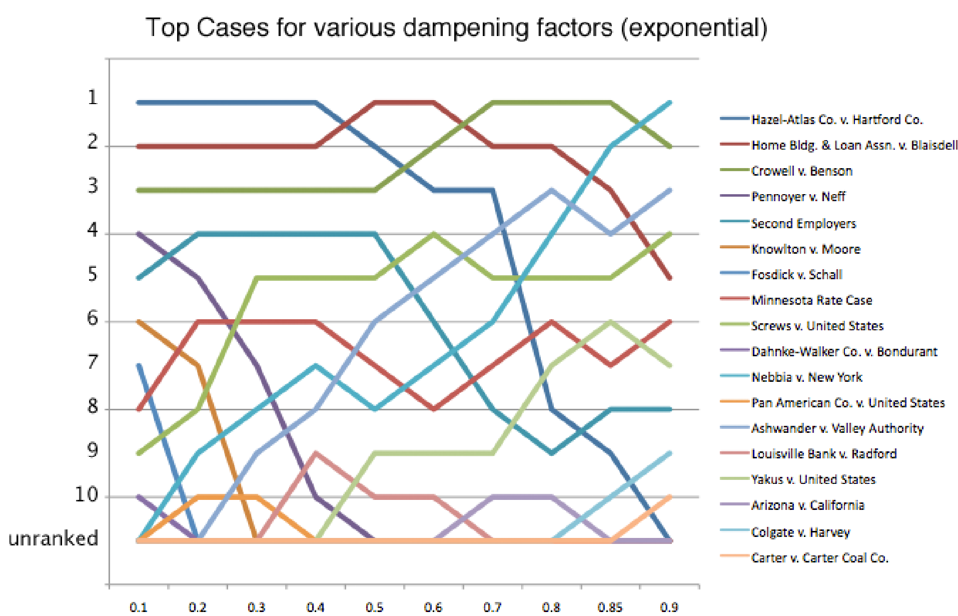

Fig. 3. Graph of top 10 cases for exponential decay pagerank as we vary the dampening factor.

Exponential Decay Weighting of Edges

$$PR(A) = \frac{1-d}{N} + d\left(\frac{PR(B)}{L(B)} + \frac{PR(C)}{L(C)} + \cdots\right)(e^{-\lambda t})$$

Each node will start out with the same weight whenever it is introduced to the graph. This weight however, will decay over time linearly or exponentially based on our model. Each edge introduced will take the weight of the directing node at that time instance. This will create a temporally-weighted graph that we can run pagerank on. The results are below.

VI. Results

Figure 1, 2 and 3 show our results testing different dampening factors for the entire graph for normal pagerank, linear decay weighting, and exponential decay weighting respectively. The figures show the top ten cases ordered by Pagerank for dampening values 0.1, 0.2, up to 0.9. From a high level, we see that there is little overlap between the normal pagerank top 10 cases and the exponential top 10 cases, though the normal and linear have very strong overlap. This suggests that the linear decay model is not necessarily sufficient to make a significant

Case name	Year	PR Rank	PR Rank - Linear	PR Rank Exp
Minnesota Rate Case	1913	1	1	6
Second Employers' Liability Cases	1912	2	2	8
Smyth v. Ames Same	1898	3	3	n/a
Yick Wo v. Hopkins	1886	4	7	n/a
Flint v. Stone Tracy Co.	1911	5	4	n/a
Ex Parte Young	1908	6	5	n/a
Gulf Ry Co v. Ellis	1897	7	8	n/a
Mugler v. State of Kansas State of Kansas Tufts	1887	8	n/a	n/a
Pennoyer v. Neff,	1877	9	9	n/a
Nebbia v. New York	1934	n/a	n/a	1
Crowell v. Benson	1932	n/a	n/a	2
Ashwander v. Tennessee Valley Authority	1936	n/a	n/a	3
Screws v. United States	1945	n/a	n/a	4
Home Building and Loan Assn. v. Blaisdell	1934	n/a	n/a	5
Yakus v. United States	1944	n/a	n/a	7
Colgate v. Harvey	1935	n/a	n/a	9
Carter v. Carter Coal Company	1936	n/a	n/a	10

Fig. 4. Top cases for each of our simulations. The rank represents their PageRank ranking out of all of our cases. Models run with damping factor of 0.9.

difference in penalizing old cases. We will take a closer look at this hypothesis later. Because the linear and normal dampening graphs are similar, we will only be comparing the normal and the exponential graph.

As we can see, the normal graph has a far more stable top 10 as the dampening factor increases. The top 10 cases dont change significantly from 0.7 - 0.9 for the normal pagerank, while the top 10 for the exponential graph change all the way through. Given that the larger the dampening factor, the import important the inbound links, and less important the random link jumping, this suggests that the exponential version of pagerank is more affected by inbound links to a node. This fits our natural hypothesis, given that the weights on the edges make real connections more important than a random jump in citation. This is also supported by the fact that most of the top cases for the exponential graph have risen from not being ranked at all at d = 0.1, to being in the top 10 by d = 0.9, again suggesting that the increasing importance of links as d increases changes the pagerank of cases significantly.

Looking at figure 4, we again see how the normal and linear pageranks are very similar, with very similar top 10 cases. We also see that the exponential model does a better job in ranking later cases. The top 10 cases for the normal pagerank all hover around the late 1800s or the early 1900s, while the top 10 for the exponential pagerank are mostly around the 1930s and 40s. Although this may be seen as a natural bias towards the later dates due to exponential decay, we see that the 6th and the 8th highest pagerank for the exponential model are around 1900 and are also in the top 10 with the normal and linear models. This implies that these two cases were likely continuing to be cited over time, and thus retained their high pageranks, while the other ones fell off due to lack of relevancy, which is the exact behavior that we are looking for in determining high influence cases.

In order to compare the characteristics of the three models, and the general behavior of the pageranks of nodes over time, we computed several heuristics for each model. First, we computed the total number of nodes that were ever in the 99th percentile in pagerank (i.e. had a pagerank greater than at least 99We found that nave pagerank had slightly more nodes in this set (1148) than the linear model (932) or the exponential model (947). While these numbers are somewhat close compared to the total number of nodes in the original dataset, it indicates that, regardless of the operation of the three models, there was overall less changing of the guard in terms of what cases were deemed important by the linear and exponential models. While this is desirable in an influence model– ideally the number of cases that are considered to be important should be as small as possible with respect to the total number of important cases– this heuristic is not in itself conclusive in determining which model is more accurate. Second, we calculated the change in pagerank for each node between successive decades for all 20 decades of data we considered in this paper. For each decade, we identified all nodes that had an increase in pagerank within a factor of 5 of the largest increase of that decade (gainers), and all nodes that had a decrease in pagerank within a factor of 5 of the largest decrease (losers). We then counted the number of nodes that were gainers or losers at least once, the number of nodes that were gainers or losers at least twice, and the size of the intersections of gainers and losers at both of these thresholds. We observed that the pagerank of low-rank nodes fluctuated wildly decade by decade in all three models, so we decided to restrict the sets of gainers and losers to nodes that belonged to the sets we calculated in the first heuristic in order to cut down on meaningless noise in the results, which were as illustrated in Figure 5.

There are several things of note about this chart. First, it is interesting to note that by and large there are more gainers

	Naive			Linear			Exponential		
	drop	gain	int.	drop	gain	int.	drop	gain	int.
Once	264	545	194	849	860	840	729	900	691
Twice	165	155	48	571	759	539	461	4	4

Fig. 5. Case drop and gain table for our three experiments.

than losers, which is an indication that pageranks decrease more slowly than they increase in all three graphs. This is a desirable characteristic, because it is much more common that a decision becomes immediately important to judicial precedent than that it would become instantly invalidated by another decision (although this also does happen in the data, as we discuss later). Second, the number of droppers and gainers is much larger in the linear and exponential models than in the nave model. This is a reflection of what we were trying to accomplish with the linear and exponential models– we wanted the measured importance of cases that were no longer important to fall more rapidly than they did in the nave model, so that only cases that have staying power are considered important. Third, the size of the intersection as a fraction of the number of gainers and droppers is lower in the exponential and naive models than in the linear model. In other words, nearly all of the nodes that the linear model deemed important it also deemed unimportant at some point. This suggests that the linear model is quite volatile– many nodes jump in pagerank enough to be considered important, but the overwhelming percentage of these gainers fall as quickly as they rose. Empirically, this is unbecoming of a desirable model for node importance because the set of nodes considered important is in constant flux, making it difficult to disambiguate nodes that are temporarily important versus nodes that are actually important.

Results from linear decay weight:

Results from exponential decay weight:

Also show in results what happened to the example cases that we cited as treated incorrectly by PageRank.

A. Case Study: *Pennoyer v. Ness*

Due to the lack of a gold standard of legal influence against which to compare our data, we decided to do some qualitative analysis to gauge the accuracy of our models. We looked at cases which were treated very differently between the different models, namely cases that unmodified PageRank ranked highly but exponential damping did not rank as highly. We noticed that this tended to happen when we predicted it would, namely in cases that used to be influential but no longer were, which unmodified PageRank as we have discussed is ill-equipped to deal with. One such case was *Pennoyer v. Ness*, in which Neffs land seized for outstanding debts while he was out of state, put up for auction and bought by Pennoyer.

Neff sued to recover land and won. Pennoyer appealed to the supreme court, at which point the court ruled that the state could not seize property from an out of state resident and sell it without notice, and Neff won the appeal. It was a key judgment at the time for personal jurisdiction, and was one of the landmark cases of its era. As a result, this case generated many citations in the coming decades. However, many facets of the case have since been overturned throughout the 20th century, and it has gradually lost much of its significance. Today, the case is an interesting academic study but is no longer relevant in the modern legal landscape. Thus we would expect it to have a low influence on cases today.

We ran PageRank at different decade intervals, and this case maintains its high PageRank up until modern day despite receiving very few citations in recent yeras. In unmodified PageRank, this case is ranked as the 9th most influential case in the entire dataset for the year 2010. In the exponential model, by contrast, the case is not ranked nearly as highly and its ranking drops steadily throughout the latter half of the 20th century. The linear model also performed well in accurately evaluating the influence of *Pennoyer v. Ness* over time, performing in line with the exponential damping model.

VII. Discussion

Our results were very compelling. The original PageRank on the entire Supreme Court dataset did a mediocre job of identifying important cases; while very influential cases such as *Second Employers* and *Minnesota Rate Case* were ranked quite highly in the modern era despite their age, there were also other old cases that were ranked highly which should not have been due to their being overturned by precedent. These included many cases from the late 19th century that may have been influential in that era but no longer are. Conversely, the exponential damping factor ranked many cases from the 1930s - 50s quite highly, which upon qualitative analysis proved to be influential and important cases to this day. The exponential model also treated the aforementioned two cases well, ranking them both in the top ten most influential cases despite being old, because of their enduring influence. As a further (non-rigorous) sanity check, we looked up the top ten cases in both the unmodified PageRank and the exponential model on Wikipedia, and found that while only 2 of the top 10 in the first model had their own Wikipedia article, there were 8 of 10 in the exponential model that had their own Wikipedia article. With merits a Wikipedia article as a very informal measure of importance this may illustrate that the exponential damping model did better at identifying these cases. Our case study of the *Pennoyer v. Ness* case allowed us greater insight into the subtleties of our model vs unmodified PageRank. It showed that over time, PageRank fails to penalize the case for not generating many new citations after a certain date, but in the exponential model the PageRank of that case drops steadily over the latter half of the 20th century.

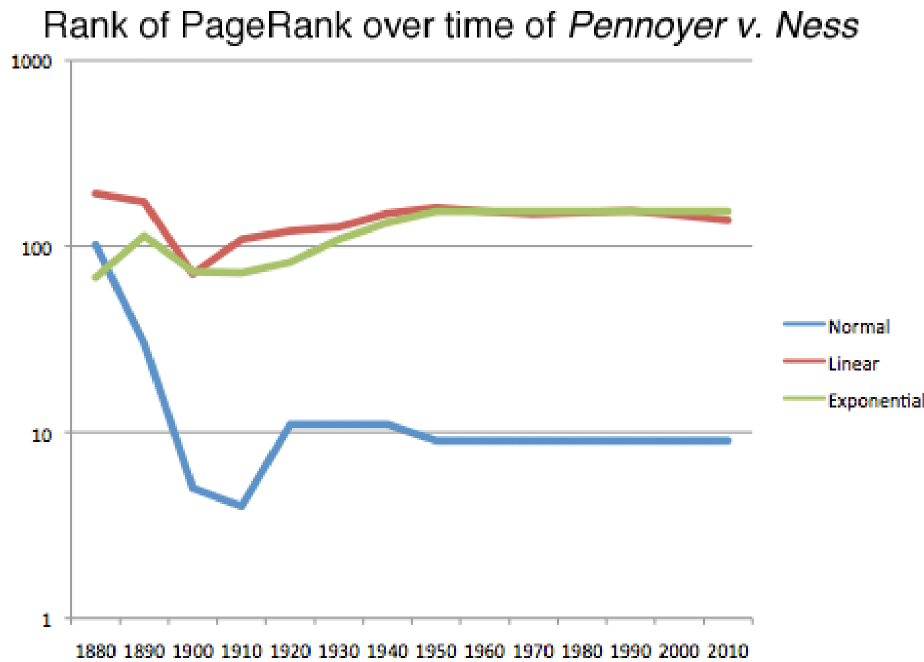

Fig. 6. Graph of the Rank with respect to Pagerank of Pennoyer vs Ness over time. Linear and Exponential graph has a significantly lower ranking than the normal graph

A. *Verification through Cluster Analysis*

Despite our good results, we looked for a further unbiased verification of the influence of the cases that we found. We propose using cluster analysis to identify the strongest clusters in the dataset, and looking for our high influence cases in those clusters. Because our dataset was very large, we could not use typical methods of graph and cluster analysis. After some research, we found a specific method called the Louvain method (Louvain, Fast unfolding of communities in large networks). At a high level, the Louvain method is a specific technique for analyzing sets of very large sizes and detecting communities. The method uses a modified hierarchical algorithm consisting of two phases. First, it looks for "small" communities by optimizing modularity in a local way. Second, it aggregates nodes of the same community and builds a new network whose nodes are the communities. These steps are repeated iteratively until a maximum of modularity is attained. The actual algorithm is implemented as follows

Assume N nodes in the Graph

Phase 1 Assign a different community to each node in network For each node i, consider neighbors j of i Evaluate gain in modularity (defined as a scalar between -1 and 1 that measures density of links inside communities / links between communities) place node i in community for which modularity is maximized (only if there is positive gain) Repeat steps 1-2 for all nodes .

Phase 2 Create a new network whose nodes are the communities found in first phase weights of links between new nodes = sum of weight of links between nodes between the two communities

The algorithm is fast because the gain in modularity that needs to be calculated each time when comparing communities can be done extremely efficiently.

We created a python implementation of the Louvain method, and combined with networkx techniques, created a graphical representation of the clusters in a random subset of the citation graph. The Louvain method assumes every node is part of its own community to start, and then builds communities by optimizing for modularity. Thus, we will only consider communities that are larger than 40 cases to remove the noise.

As we can see from figure 7, there are only a few (about 20) clusters in the large dataset. In the largest cluster, we see that the number 1 pagerank case (Minnesota Rate Case) is contained in the largest cluster, and the number 2 pagerank case (Second Employees Liability Case) is also in a very large cluster. This verifies that our pagerank analysis has validity, and that pagerank is a good measure of importance of a case.

VIII. Future Work

This is a rich field that has much potential for future research. While we considered our dataset mostly as a whole, law has many subsets (i.e. fields) in which it might be more possible to accurately assess the influence of the case. Considering subsets of our legal data might yield more applicable results in certain fields of law. Another approach to this might be to instead use a machine learning clustering algorithm, so that instead of manually partitioning the set, we see if these fields are emergent from the citation graph itself.

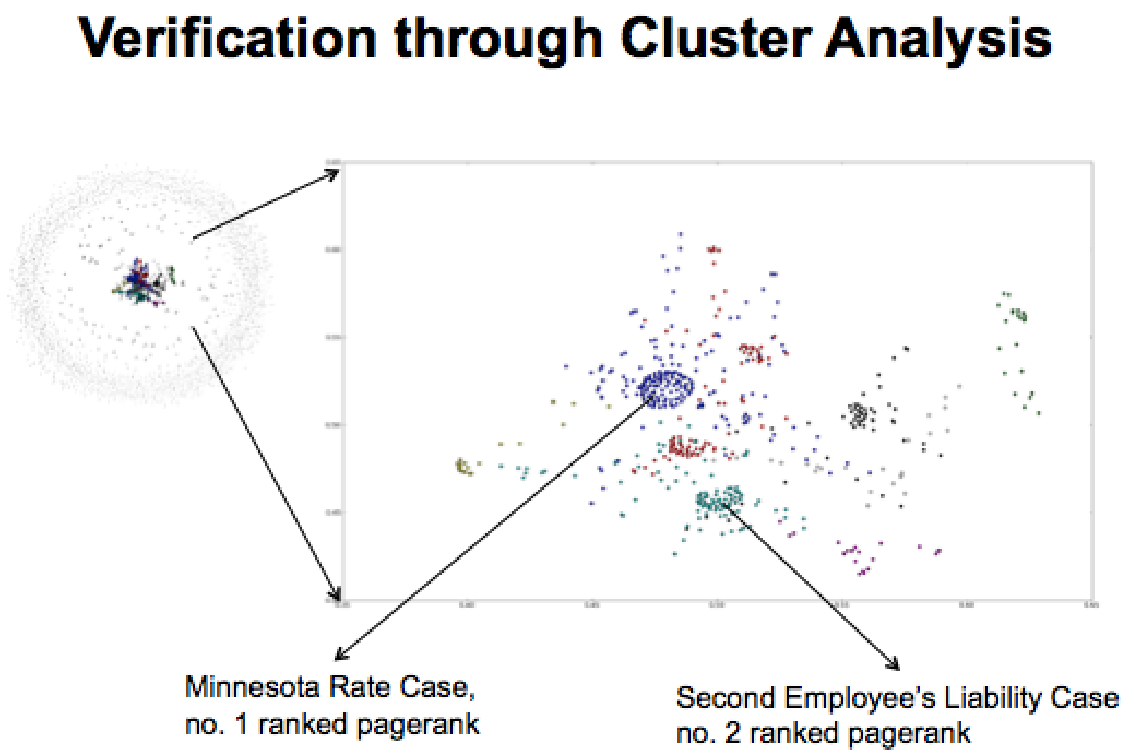

Fig. 7. Clustering of cases based on citation graph. Cases with high interconnectivity of citations tend to cluster together.

We only considered U.S. Supreme Court cases for this model for a variety of reasons enumerated in previous sections. However, we might get different results if we considered our entire legal dataset. Often, the influence of a supreme court case is determined by other supreme court cases that cite it; however, it is also determined by lower-level cases that cite it as well. The supreme court tends not to take on cases unless they concern ambiguities in the U.S. Constitutional Law, and thus a case might not get cited by other supreme court cases even if it is influential, if it establishes very clear and unambiguous precedent. Thus an interesting future research project would be to consider the rest of the dataset.

Another promising future direction is attempting to predict what the future of a case will become based on early returns. Can we effectively determine the expected influence of a case that is only a few years old? It would be interesting to identify the mathematical features that are common to eventually influential cases throughout the years.

IX. Conclusion

We first explored PageRank in its unmodified form as a means to determine the influence and importance of cases throughout history. These early results showed that PageRank often gave different results than what might be expected from a qualitative analysis and our legal intuition, and this we sought to construct a more comprehensive model. We thus varied the dampening factor, and added two different weighing models to pagerank; linear and exponential scale of decay. Our subsequent model for determining the influence of a legal opinion is superior to unmodified PageRank. We evaluated our model on the basis of certain cases that we knew qualitatively to be either influential or not in our era. *Pennoyer v. Ness* is an example of a case that was treated significantly differently with the introduction of the different models. Our ultimate conclusion is that the exponential decay is a far better model for considering temporal factors when using pagerank to judge influence.

We conclude that our novel approach can be used to identify influential legal cases at any point in history, more fairly taking into account the recentness of the case's influence. There are still many compelling directions that our work can be taken in the future as outlined above.

Acknowledgment

The authors would like to thank Professor Lescovic and the teaching staff of CS 224W for their help throughout the quarter. We would also like to thank Andrew Baine for generously providing the Stanford Law dataset and himself as a resource, without which this project would not have been possible.

REFERENCES

[1] Page, Larry; Sergey Brin, R. Motwani, and T. Winograd. “The pagerank citation ranking: Bringing order to the web.” 1998.

[2] Ding, Ying and Erjia Yan. “PageRank for ranking authors in co-citation networks.” Journal of the American Society for Information Science and Technology. 60.11 (2009): 2229-2243.

[3] ”Pennoyer v. Neff - Case Brief Summary.” Lawnix - Law Resources and Legal Information. Web. 07 Dec. 2010. ¡http://www.lawnix.com/cases/pennoyer-neff.html¿.

[4] Borchers, Patrick J. The Death of the Constitutional Law of Personal Jurisdiction: From Pennoyer to Burnham and Back Again 24 U.C. Davis L. Rev. 19 (1990).