Hubs and authorities for network analysis

Authors: Anupama Chandroth, Rakesh Santha Kumaran

Email: anuc@iu.edu, rsantha@iu.edu

Introduction and objectives

Recognizing that legitimacy is essential to achieve the policy objectives, the members of the Court justify their substantive rulings through court opinions, it is based on legal rules and previous decisions made by the court. The work by Fowler and Jeon (2008) is an attempt to utilize the quantity and quality of judicial citations in Supreme Court majority opinions to understand how legal policies are formulated in the judiciary. This was achieved through network analysis, by computing the Hub and Authority scores of the decisions/ Cases. The authority score of a case depends on the number of times it is cited and the quality of the cases that cite it. Symmetrically, the hub score of a case depends on the number of cases it cites, and the quality of the cases cited.

Here we aim to use the complete network of citations in all 30,288 majority opinions contained in the U.S. Reports from 1754 to 2002 to demonstrate how network data can aid in the study of precedent and its influence in judicial decision-making and aim to replicate the results from Fowler Jeon. We aim to compute "hub" and "authority" scores for nodes(cases) in a network and apply them to a network of Supreme Court opinions based of "The Authority of Supreme Court Precedent" Social Networks (2008), Fowler and Jeon. The primary aim is to visualize the rate at which the authority score varies over a span of 30 years for 2 important cases. Moreover, in this project we try to replicate results for a network of 92 cases that is associated with Roe Vs Wade, i.e., the 91 cases that cited it and the case itself. This would also help in evaluating the legitimacy of the program that we created as part of this project.

Hub and Authority score

Initially the approach to compute the importance of a case was based on the inward citations only , i.e., the number of cases that cites a particular case contributes to how important it is. This is based on the theory of eigen vector centrality. However, the advancement in social network theory helped including both inward and outward citations (links) to establish the importance of a case (Kleinberg, 1998).

The concept of hub scores and authority scores could be seen as a ranking system. In the context of Supreme Court Precedents, a hub is a case that cites many cases, which would help in identifying legally relevant cases and authority is that case which is cited by many others, there by being "an important" decision. But here many decision acts as both hubs and authorities.

A good hub will point towards a good authority so if x is a vector of the authority score it would be proportional to the sum of all the hubs (let that be y) that point towards it, we can represent it as $x_i \propto a_{1i}y_1 + a_{2i}y_2 + \cdots + a_{ni}y_n$ and similarly since good authorities point are pointed to by good hubs, the hub score y would be proportional to sum of the authority scores and thus $y_i \propto a_{1i}x_1 + a_{2i}x + \cdots + a_{ni}x$. We could convert this to matrix form attain λx = ATy and λy = Ax [1].

Where A is the network matrix, where each element a_{ij} is determined by the relation, of i and j, if case i cites case j it would be 1 else 0. We start with x and y for all elements to be equal to 1 and use iterative multiplication to find the solution. Kleinberg (1998) has obtained the solution for the above equation [1], and it converges to λx* = ATAx* and λy* = AATy*, where λ is the principal eigenvalue and x* and y* are the principal eigenvectors of the matrices ATA and AAT, respectively. (AT is A transpose).

Design concepts

Our Aim is to visualize how, the speed at which precedents become legally influential and we aim to reproduce the results from Fowler and Jeon (2008) specifically Figure 6, which explore the speed at which the authority score of Roe Vs Wade (1973) and Brown Vs Board of Education (1954) changes over time. The data for replication was extracted from (Fowler and Jeon (2007)).

As mentioned above we require a network matrix A, which indicates the relationships between the two cases and the column elements indicates the cases that are cited by the elements in the row. We are provided a file (allcites.txt) which has the list of cases and the cases that it cites to make the final decision from the year 1754 to 2002. We convert this data into matrix form to create the complete network matrix, where element a_{ij} is 1 if the case i cites case j and else 0. We use the package Matrix in R and create a sparse matrix, since we are dealing with a matrix with numerous elements that are 0. This helped us reduce the memory used and the computational complexity involved.

We are required to compute the authority score for each year, and to attain this we use a top-down approach. We create subsets of the A matrix, by looking at the case ID (a numerical value) for the last case for which the decision was made in a given year, called terminal year (this corresponds to the largest case number for that year). Here we also use a supplementary file with case numbers and year in which those cases were passed. Then extracting the matrix with a dimension that corresponds to the row and column for that given case (Figure 1. A). Our initial approach was to create a matrix for each terminal year using the file (allcites.txt), however this was more computationally expensive, because it involved going each row of the file and finding the cases that are cited in each terminal year and creating a new matrix from scratch.

Once we have the matrix, the next step is to compute the hub and authority score. We start with computing the transpose of the A matrix (or the subset for each terminal year) and then extract the principal eigen vector for A*AT and AT*A, which would eventually give the hub and authority score. We used the package "RSpectra," since we are dealing with large sparse matrices. This package helps in extracting the eigenvectors and eigenvalues for large matrices. The computed scores are normalized to the sum of the squares of the scores equal to one (Figure 1.B). Initially we had a design flaw as we used iterative method to compute the eigen vector, by first extracting the principal eigenvalues and using iterative multiplication with the previously computed scores until it converges (normalized at each step). This turned out to be more computationally expensive and redundant since we could directly extract the eigenvector as it is the solution.

The ultimate step is to execute the hub and authority for a given case. In this function, we create subsets of the A matrix for each terminal year that we are interested in and then compute the scores for each terminal year. Then the scores for the case that we are interested in are extracted and given as the final output of the function.

A

```r
library(readr)
library(Matrix)
library(RSpectra)

#Once we have the complete network matrix,  we extract the matrix for each
#terminal year using the function below (year_determined)

#Extracting the A matrix for each year

#Parameters
#Case_year: Data with case number and year the decision was made
#Final_A_mat: A network matrix that includes the relation with each case until
#2002. Element A[i.j] is 1 if case i has cited case j, else 0.
#year_need: Year for which we want to subset the A matrix for

year_determined<-function(Final_A_mat,case_year,year_need) {
  list_cases_idx<- which(case_year[,2]==year_need)
  final_mat_idx <- max(list_cases_idx)
  return(Final_A_mat[1:final_mat_idx,1:final_mat_idx])
}
```

B

```r
#The following function computes the hub and authority score
#Parameters:
#A_mat_subset: Subset of the final A matrix, which has the relationships till
# 2002, we subset it for each year we are interested in

Aut_hub_score<-function(A_mat_subset){
  aut_hub_score<-data.frame()
  a<-rep(1,times=ncol(A_mat_subset))
  h<-rep(1, times = ncol(A_mat_subset))
  AtA_aut <-t(A_mat_subset)%*%(A_mat_subset)
  AAt_hub <-A_mat_subset%*%t(A_mat_subset)
  #gives us the largest eigen values and
  #eigen vectors in the case of a sparse matrix
  ei_aut<-eigs(AtA_aut,k=1,which = "LM")
  ei_hub<-eigs(AAt_hub,k=1,which = "LM")
  aut_score<-abs(round((ei_aut$vector),digits=3))
  hub_score<-abs(round((ei_hub$vector),digits=3))

  aut_hub_score<-rbind(aut_hub_score,data.frame(aut_score,hub_score))
  return(aut_hub_score)
}
```

Figure 1. A) Function to subset network matrix for terminal year, B) to compute hub and authority score

Verification of the design

We wrote 3 programs, one to extract the subset of the network matrix A, for each terminal year, second to compute the hub and authority scores and final one to extract the hub and authority scores of as given case for a given number of years. We tested our functions using “testthat” package available in R as well as by executing the function for a subset of cases.

In total we performed 12 tests using “testthat”. We first checked whether the subset created has the required number of rows and columns, and whether the relationship between the cases is accurately captured by comparing whether the value for the element is 0 or 1 based on the cases that are related and not related, respectively. For both execution function and the function that computes the scores we checked whether the number of outcomes corresponds to the number of years and cases, respectively. The scores were also verified by comparing it with the outcomes in the paper.

As a form of informal testing, we extracted the cases that have cited Roe Vs Wade, this formed a network of 92 cases. We computed the hub and authority scores of these cases and extracted the results for the 5 most cited cases in the network and compared our results with the ones in the publication (Table 1 in Fowler and Jeon (2008). Our result matched perfectly with the publication except for the hub score of Roe Vs Wade.

Discussion

With the programs that we wrote, we were able to successfully replicate the results from the publication (Figure 2). Authority scores help in understanding the rise and fall of a precedents’ importance with the ever-evolving legal network. Here we have explored the same for 2 cases Roe Vs Wade and Brown Vs Board of Education. We are not only able to see the importance of a case but also the speed at which these becomes legally influential. In Figure 2 we observe that the speed at which each of these cases achieved a level of significance is quite different. The authority score for Brown was initially low, indicating that this precedent was legally weak. We see a spike after a decade, the case strengthened as a result of the Civil Rights Act of 1964.

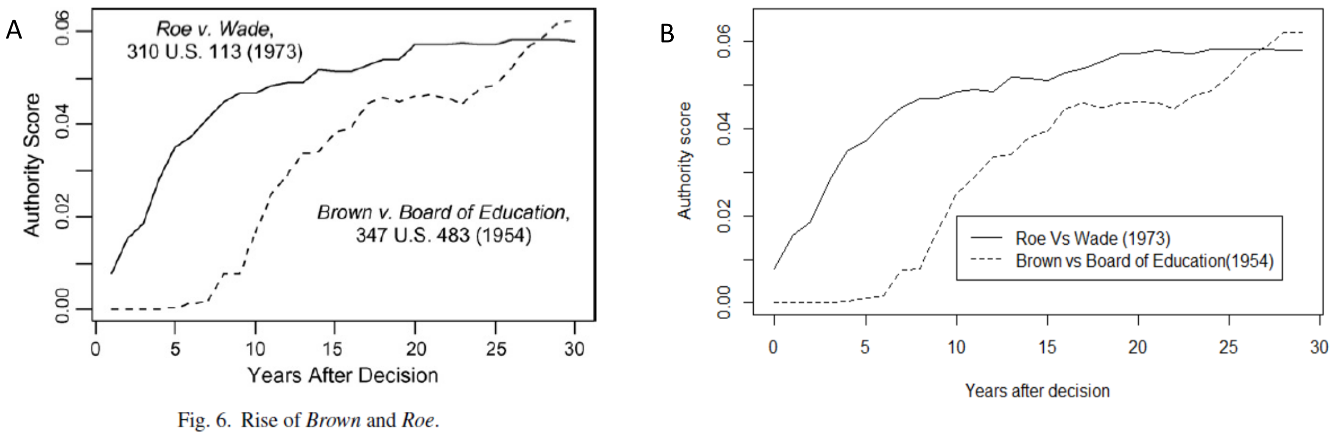

Figure 2. Side by side comparison of the Rise of Roe and Brown from A) Fowler and Jeon (2008) and B) current project

Brown as case alone did not set a legal standard, however it being cited by the Civil Rights act to establish as an example of basic right by desegregation of schools, increased its influence. The case states that segregated schools are unconstitutional, but it was not established as a legal rule. Whereas Roe Vs Wade let alone created a new path and established the restriction of state regulation and abortion and created a new standard for future cases by voiding the laws in every state that limits abortion. Even after a decade of Brown, only a small percent of the schools was desegregated whereas within in few months of Roe, 181140 abortions were performed. When we compare the authority scores after a decade of each case, we observe that the authority score for Roe was twice that of Brown. This indicates that the legal influence of Roe increased at a much higher rate than Brown.

When we look at the network of 92 cases associated with Roe, we see that Roe has a lower hub score than the 4 other cases with the next highest authority scores, this is because it only cites one other case in the network. However, when we consider the overall network Roe has a higher hub score than Thornburgh and Webster, this could be because good hubs cite good authorities. We could infer that Roe cites better authorities.

Conclusion

Through this project we are able to create a program that computes the hub and authority scores that corresponds to the rankings that are provided by legal experts for each precedent. We are able to identify the important precedents and the decisions that are firmly rooted in prior opinions. In future studies it would be interesting to see how the recent overturn of Roe Vs Wade might have affected the authority score of the case. We could also expand the current program to other cases to understand how they have progressed over time

Repository name: Stat610_anuc_rsantha

Link to git hub: https://github.iu.edu/anuc/Stat610_anuc_rsantha

References

Fowler, James H., and Sangick Jeon. "The authority of Supreme Court precedent." Social networks 30.1 (2008): 16-30.

James H. Fowler; Sangick Jeon, 2007, "Replication data for: The Authority of Supreme Court Precedent", https://doi.org/10.7910/DVN/XMBQL6, Harvard Dataverse, V1, UNF:3:Suen5yCvzTQO0zq0nySNCg== [fileUNF]

Kleinberg, Jon M. "Authoritative sources in a hyperlinked environment." SODA. Vol. 98. 1998.

Notes for Hub and Authority score: https://nlp.stanford.edu/IR-book/html/htmledition/hubs-and-authorities-1.html