Network Analysis of Citations on Court Cases from the Supreme Court of the United States (SCOTUS)

Michael Kim, Chisung Cho

Department of Statistics and Operations Research
University of North Carolina, Chapel Hill
Submitted for Spring 2016 Review

Advisors: Professor Shankar Bhamidi and Ph.D. candidate Iain Carmichael

Abstract

In this report, the citations of court cases from the Supreme Court of the United States are analyzed through basic network analysis techniques covered by the igraph package in R. First, we assess some elementary information of the network, such as on the vertices, edges, and vertex attributes. Then, key network characteristics, such as centrality measures, transitivity, reciprocity, assortativity, clustering, and weakly-tied locations (cut vertices) are computed for the entire network. Finally, fundamental visualization and animation aspects are presented while working within the computational limitations of igraph and R, due to the bad complexity times of some of the algorithms in igraph's functions when handling large network data. The analysis results in few, significant findings on the relationship between some key network characteristics and the real-world context of the respective court cases.

1 INTRODUCTION

The aim of the report is to execute several network analysis techniques on the citations of the Supreme Court Cases of the United States ("SCOTUS"). The hope is that our analysis work yields interesting results, such as finding a specific court case that may have significant relationships with other court cases citing it. Or, finding court cases that may have resulted in significant, real-life changes on how United States society and government function.

All the code is executed through R, and many of them are inspired by the computing and theoretical knowledge from the book, *Statistical Analysis of Network Data with R* by Eric D. Kolaczyk and Gábor Csárdi. This report will cover ideas from chapter 3 ("Visualizing Network Data") and chapter 4 ("Descriptive Analysis of Network Graph Characteristics") of the book. The network data can be acquired from contacting UNC Chapel Hill's current Ph.D. candidate, Iain Carmichael or any of the members in UNC's Network Analysis Group under Professor Bhamidi.

That being said, huge thank you to Iain Carmichael for cleaning up the data acquired from https://www.courtlistener.com and making them into useable network objects. He is also responsible for a majority of the code under the next two sections: 2. BASIC NETWORK SUMMARY and 3. EXAMPLES OF USEFUL PLOTS.

Note: The data used are

- "scotus_net_EL_date" txt file, containing a loadable igraph object for R. It contains appropriate information on vertices, edges, and vertex attributes (timestamps of specific dates) of the SCOTUS network,
- "scotus_net_EL_year" txt file, containing a loadable igraph object for R. It contains appropriate information vertices, edges, and vertex attributes (timestamps of specific years) of the SCOTUS network.

2 BASIC NETWORK SUMMARY

2.1 Reader's Notes

The main package required for the network analysis and visualization within the report is "igraph." The codes and network data to acquire the outputs in the report are available in the GitHub site of the UNC Chapel Hill's Network Analysis Group: https://github.com/UNCscotus/scotus.

We will also be primarily working with the "scotus_net_EL_date" file, since the vertex attributes of dates are essentially inclusive of the vertex attributes of years.

2.2 Simple Overview of Network Data

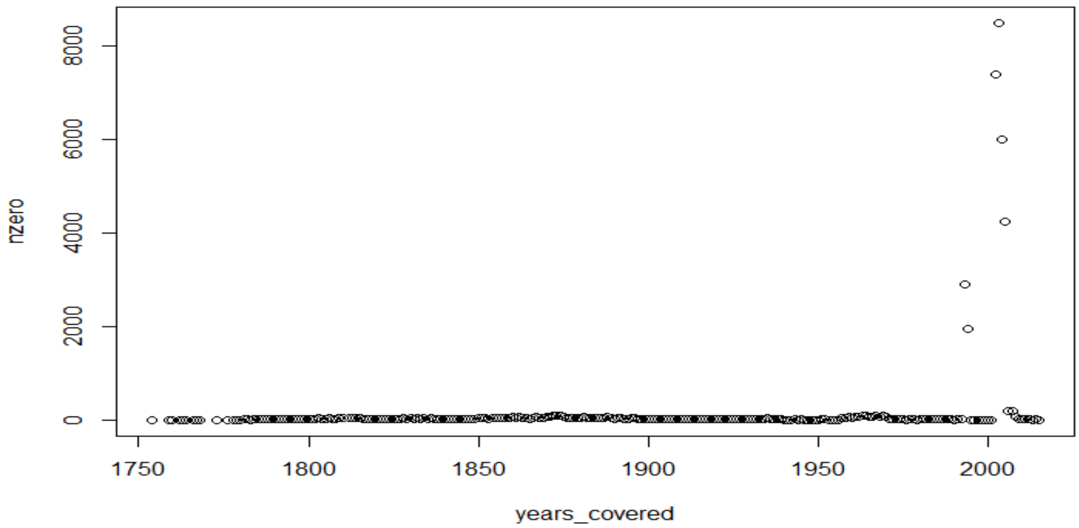
First, we must examine some basic information of the vertices and edges in the SCOTUS network. There are 63,744 vertices and 244,496 directed edges in the SCOTUS network. The network covers Supreme Court Cases within the time range of years from 1754 to 2015 with information missing in 1755-1758, 1761, 1765, 1769-1772, 1774, 1775, 1777, and 1811. As one can already tell, this is a big network, and some may even consider this as "big data," considering the computational difficulties R faces for visualizing large sections of the network, which is further explained in a later section of the report 5. VISUALIZING NETWORK.

We define the *popular years* within the network to be the years containing more than 1,000 vertices in the SCOTUS network. The *popular years* are 1993, 1994, 2002, 2003, 2004, and 2005. For the reader's information, the histograms of the in-degree, out-degree, and total-degree frequency of these years can be viewed under a later subsection of this report, "Shiny Application."

2.3 Isolated Vertices

Among the 63,744 vertices, 36,722 of them are completely isolated and are not attached to any other vertex or edge. Therefore, approximately 57.69% of our network data contain court cases that do not cite other court cases and are not cited by other court cases. Interestingly, a large portion of the isolated vertices come from the *popular years*. The *popular years* contain a total of 30,960 isolated vertices, so approximately 84.19% of the isolated vertices are actually from these *popular years*.

We can represent our finding of the 57.69% isolated vertices in the entire network through a plot of vertices with zero degrees for a given year. It's clear that the *popular years* contained many of these isolated vertices, as indicated by the jumps in the scatterplot:



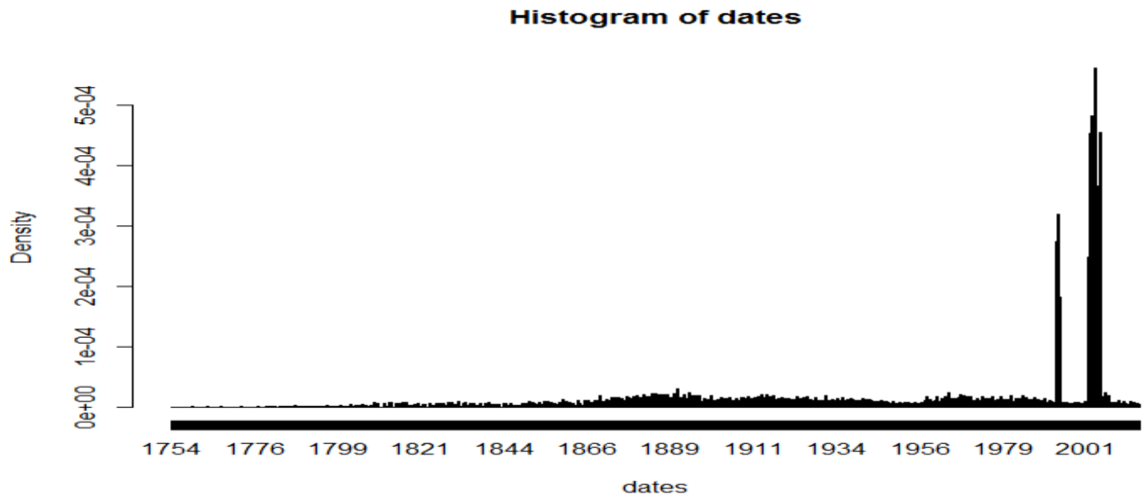
2.4 Connected Vertices

Additionally, we can see the number of connected vertices in the SCOTUS network, the vertices with at least one edge between themselves and the other vertices in the network. There are 26,792 vertices in the entire network that are connected (63,744 – 36,772 = 26,792). Thus, approximately 42.03% of the vertices in the entire network are connected.

Just as the *popular years* contained a significant 84.19% of the isolated vertices in the entire network, they contain only 897 of the connected vertices, or approximately 3.35% of the connected vertices. Therefore, it is safe to conclude that a majority of court cases from the *popular years* don't have much relation to each other.

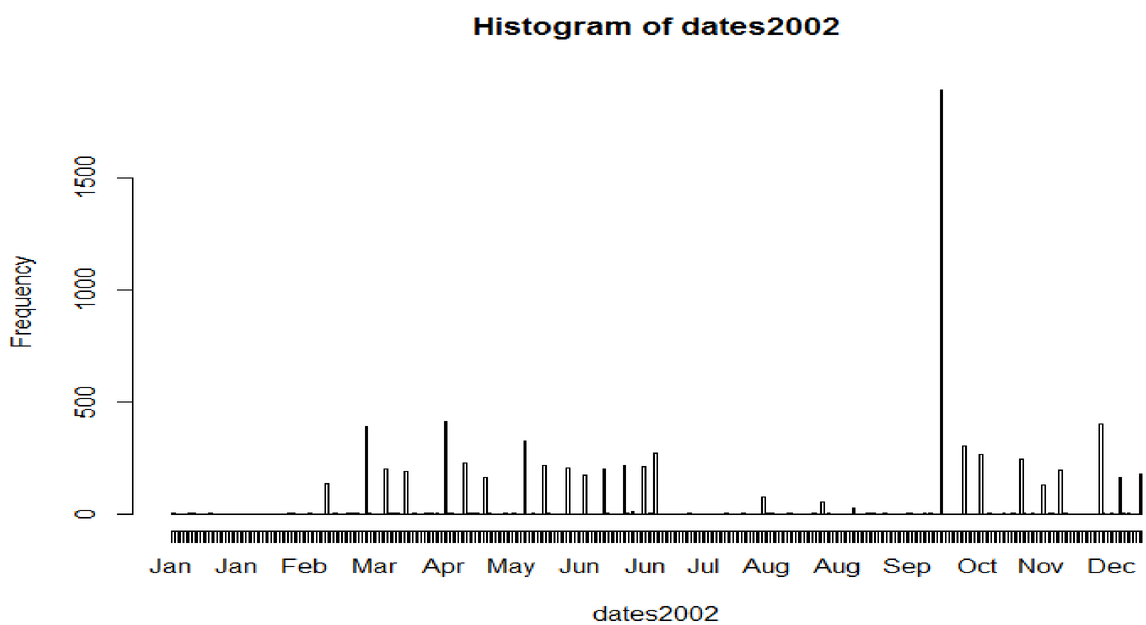
3 EXAMPLES OF USEFUL PLOTS

This is a histogram of the density of court cases by dates:



As predicted, the *popular years* are denser than the non-*popular years*. This is apparent by the spikes around the late 90's and early 2000's. Remember that the *popular years* are 1993, 1994, 2002, 2003, 2004, and 2005.

To verify the conclusions drawn from the above Density vs. Dates graph, we plot a frequency histogram of all the dates in 2002, one of the popular years:



It appears that many of the court cases in 2002 are from right before October 10.

By further investigation, on October 7th, 2002 there was a total of 1,890 court cases handled on that day. It seems this was a busy day for SCOTUS. The purpose of this section was to illustrate how one can extract the number of court cases around specific months or dates.

4 ANALYSIS OF NETWORK CHARACTERISTCS

To seek the influential court cases in the SCOTUS network, the standard procedure in network analysis is to get centrality measures, especially for the vertices. The common centrality measurements for vertices, suggested by *Kolaczyk and Csárdi*, are closeness (harmonic) centrality, betweenness centrality and eigenvector centrality (*47*). The history and math for these centrality measurements are well-explained by *Kolaczyk and Csárdi* in chapter 4. This section will mainly rely on the implementation of the igraph package to study these centrality values. Then we will see if vertices with high centrality values represent important court cases in the next section, 5. VISUALIZING NETWORK.

4.1 Vertex Closeness Centrality

Definition: assigns a value for how 'central' a certain vertex is through measurements of how 'close' that vertex is to other vertices, often dictated by the distance from the vertex of interest to all other vertices (*Kolaczyk and Csárdi, 47-48*); close centrality is also known as "harmonic centrality"

The formula is as follows (explanation in *47-48*):

$$c_{Cl}(v) = \frac{1}{\sum_{u \in V} dist(v,u)}$$

The vertex with the highest closeness centrality is 'id2959750' with a value of 41,238, according to igraph's implementation of above formula for closeness/harmonic centrality.

4.2 Vertex Betweenness Centrality

Definition: assigns a value to represent how much a vertex is 'between' other vertices, often dictated by the number of shortest paths that run through that vertex (*Kolaczyk and Csárdi, 48*)

The formula is as follows (explanation in *48*):

$$c_B(v) = \sum_{s \neq t \neq v \in V} \frac{\sigma(s,t|v)}{\sigma(s,t)}$$

The vertex with the highest betweenness centrality is 'id118365' with a value of 45,823, according to igraph's implementation of the above formula for betweenness centrality.

4.3 Vertex Eigenvector centrality

Definition: assigns a value to represent the 'status' or 'rank' of a vertex, often through the usage of eigenvector solutions of linear systems (*Kolaczyk and Csárdi, 48*)

The formula is as follows (explanation in *48*):

$$c_{Ei}(v) = \alpha \sum_{\{u,v\} \epsilon E} c_{Ei}(u)$$

The vertex with the highest eigenvector centrality is 'id106514' with a value of 2,293, according to igraph's implementation of the above formula for eigenvector centrality.

4.4 Maximum In-degree and Maximum Out-Degree

Other than the centrality measures, other interesting network characteristics are the maximum in-degree and maximum out-degree. In other words, one can find the most-cited court case (maximum in-degree), as well as finding the court case that cites other court cases the most (maximum out-degree).

The vertex with the maximum in-degree in the SCOTUS network is 'id96405' with an in-degree of 1,295, and the vertex with the maximum out-degree in the SCOTUS network is 'id104616' with an out-degree of 197.

4.5 Transitivity

The transitivity of the SCOTUS network is about 11.41%. This means that about 11.41% of the connected triples (three vertices defined on undirected ties) form triangles, which are closed and connected triples. Transitivity is synonymous to the "measure of global clustering," according to the implementation by igraph (*Kolaczyk and Csárdi, 56*).

4.6 Reciprocation

The reciprocation of the SCOTUS network is about 0.36% for dyads. This means that 0.36% of dyads in the network contained reciprocated/mutual, directed edges (*Kolaczyk and Csárdi, 56*)

The reciprocation of the SCOUS network was about 0.18% for directed edges. 0.18% of the directed edges in the network are reciprocating edges (*Kolaczyk and Csárdi, 56*)

This shows that reciprocation is a poor measurement to use to define relationships between vertices of SCOTUS, since most dyads or individual directed edges seem to have no reciprocation. Thus, basing results on reciprocation is not ideal in our situation, but it is good to know which methodologies should not be used or seem undesirable in further network analyses.

4.7 Assortativity

Kolaczyk and Csárdi explains assortativity coefficient as:

> *"Selective linking among vertices, according to certain characteristic(s), is termed assortative mixing in the social network literature. Measures that quantify the extent of assortative mixing in a given network have been referred to as assortativity coefficients" (56)*

The assortativity coefficient of the SCOTUS network was about 0.029.

It is hypothesized that the low assortativity coefficient may be relational to the low reciprocity for the SCOTUS network. Thus, assortativity is also deemed potentially undesirable for further network analyses. Further investigation is required.

4.8 Weakly-Connected Giant Component

As explained previously under 2. BASIC NETWORK SUMMARIES, 36,722 of the vertices are isolated and 26,792 of the vertices are connected. The R code for getting these results were based the number of vertices that had total-degree = in-degree + out-degree = 0, which represent the number of isolated vertices, and seeing the number of vertices that had total-degree > 0, which represent the number of connected vertices.

Although we initially hypothesized that the network may have 36,722 isolated vertices and one giant cluster of 26,792 vertices, the "decompose.graph" function provided by igraph, which return the number of vertices in the weakly-connected clusters (WCC's) of the network, proves otherwise:

1	2	3	4	26786
36799	69	3	3	1

The results show that there were 36,799 WCC's with 1 vertex (essentially 36,799 standalone vertices), 69 WCC's with 2 vertices each, 3 WCC's with 3 vertices each, 3 WCC's with 4 vertices each, and one giant WCC with 26,786 vertices.

It is interesting that the "decompose.graph" function returns 36,799 WCC's of 1 vertex, which is a higher number than 36,722, the number of isolated vertices. One might expect that there would be exactly 36,722 WCC's of 1 vertex as well, so it seems that the "decompose.graph" function took into account of 36,799 – 36,722 = 77 more vertices for its smallest WCC group. So even if these 77 vertices may be connected by definition, they did not form actual clusters among themselves or with other connected vertices, under the algorithm of igraph's "decompose.graph."

Furthermore, the "decompose.graph" function is able to detect that among the connected vertices, 69 + 3 + 3 = 75 vertices formed smaller WCC's than the giant WCC, each containing at least two vertices. Although we are unable to obtain the correct algorithm to extract these specific vertices, they are worthy to note for future studies, since they might have an interesting relationship among themselves, such as similar geographic location, or case-relevance by category.

The transitivity, reciprocation, and assortativity coefficients between the giant WCC and the entire SCOTUS network are very close, as shown through the GitHub link above. Thus, one may hypothesize that these network characteristics for the entire network with 63,744 vertices are mainly derived from the 26,786 vertices of the giant WCC. Or in other words, one may assume that these network characteristics are mainly derived from the connected vertices in the entire network. This implies that the isolated vertices have transitivity, reciprocation, and assortativity coefficients of zero.

Finally, we note that the giant WCC contains approximately 42.02% of the vertices in the entire SCOTUS network, which corresponds to the approximate 42.03% of the vertices in the entire network being connected.

4.9 Small World Property

According to *Kolaczyk and Csárdi*, giant clusters of real-world networks often share a network characteristic called the, *small world* property (*57*). The two conditions for this property are as follows:

(a) "the shortest path-distance between pairs of vertices is generally quite small" (*57*)
(b) "the clustering is relatively high" (*57*)

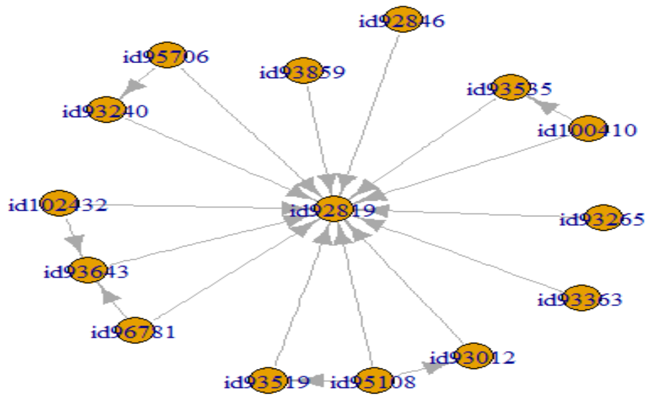
Unfortunately, it seems that the SCOTUS network already breaks condition (b), since only 42.02% of the vertices in the network account for the giant cluster. Therefore, although the average of the shortest-path distances within the SCOTUS network is approximately only 5.54, and the longest distance of the shortest-paths is only 25, the network does not satisfy *small world* property, unexpected of a real-world network.

4.10 Cut Vertices

There are 2,350 cut vertices in the SCOTUS network. A cut vertex is a "single vertex that disconnects the graph … Identification of such vertices can provide a sense of where the network is vulnerable" (*Kolaczyk and Csárdi, 58*). By definition of a cut vertex, these cut vertices would have either zero in-degree or zero out-degree, or else they could not disconnect the graph.

Cut vertices are potentially important to study because of their possible significance in American law. These court cases may have been filed early in American history, as this could potentially explain how some of the cut vertices have zero-out-degree but a high in-degree (many future court cases citing an early court case possibly).

For example, one of the cut vertices, 'id92819' yields this plot, showing that it has zero out-degree, while it is cited by many other cases:



This course case represents the *Union Bank of Chicago v. Kanasas City Bank, 136 U.S. 223 (1980)*: https://www.courtlistener.com/opinion/92819/union-bank-of-chicago-v-kansas-city-bank/. Advice from experts in law is preferred to accurately decide whether this court case had a major impact in U.S. law or society. However, our hypothesis is that this court case does not seem to be highly influential in real-world context, since it cited by only 22 opinions according to *CourtListener*.

5 VISUALIZING NETWORK

Due to the computational limits of R, computations exceeding 500 vertices causes a problem (slowing/crashing R or messy visualization) when plotting the network. *Kolaczyk and Csárdi* propose techniques to visualize large network data, such as with the Kamada-Kawai method (*37*), DrL method (*37-38*), and graph coarsening (*38*). When tested on a slightly smaller sub-network, such as the giant WCC with 26,786 vertices, these techniques proposed by *Kolaczyk and Csárdi* do work to some extent. However, due to the convoluted image produced when plotting large portions of the SCOTUS network, such as the giant WCC, we decided not focus on this aspect of visualization in our report. Due to this computational inefficiency of igraph and R in just visualizing a large, stationary network, an interesting software to consider for solving this dilemma is the "Graph Tools" package for Python.

Thus, we will visualize order-1 neighborhoods around vertices with high centrality measures, maximum in-degree, and maximum out-degree. All this information might give us more clue on how important these specific court cases are.

Remember from the previous section of the report:

- 'id118365' (vertex with highest betweenness centrality value)
- 'id2959750' (vertex with highest closeness centrality value)

- 'id106514' (vertex with highest eigenvector centrality value)
- 'id104616' (vertex with maximum out-degree of 197 in the entire network)
- 'id96405' (vertex with maximum in-degree of 1,295 in the entire network)

The organization of the neighborhood visualization for the centrality measures will be presented in the following order:

1. the neighborhood of order-1 *citing* the vertex of interest
2. the neighborhood of order-1 *cited by* the vertex of interest
3. a *combination* of the previous two graphs.

Furthermore, the bigger these neighborhoods become, GitHub shows that the appropriate code was executed to resize the vertices and edges for better visualization.

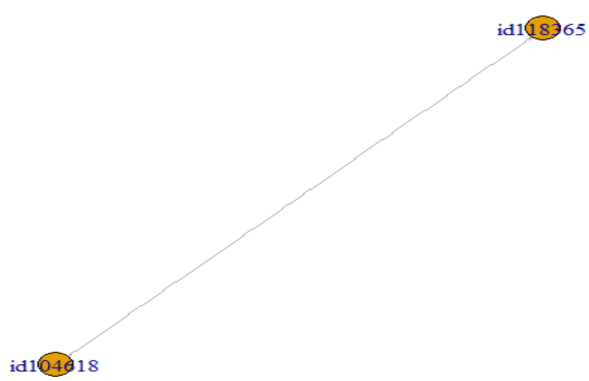
If the reader is interested, slight modification of the code in GitHub will allow for visualization of neighborhoods of higher orders (order-2, order-3, etc.). However, this carries the same risk of computational inefficiency as mentioned before, since the neighborhoods can become exponentially larger, containing thousands of vertices.

5.1 Betweenness Centrality Visualization ('id118365')

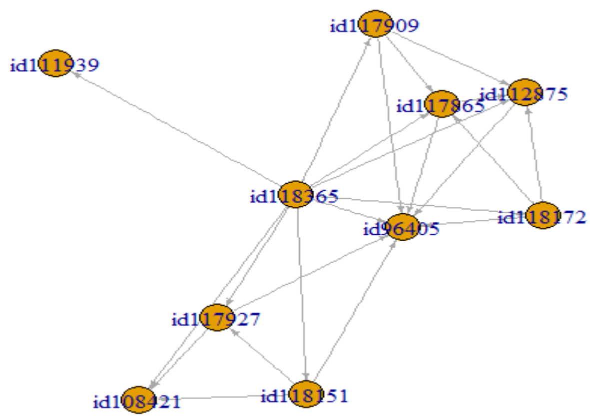
'id118365' was *Fischer v. United States, 529 U.S. 667 (2000)*:
https://www.courtlistener.com/opinion/118365/fischer-v-united-states/

Advice from experts in law is preferred to accurately decide whether this court case had a major impact in U.S. law or society. However, our hypothesis is that this court case does not seem to be highly influential in real-world context, since it is cited by only 53 opinions according to *CourtListener*.

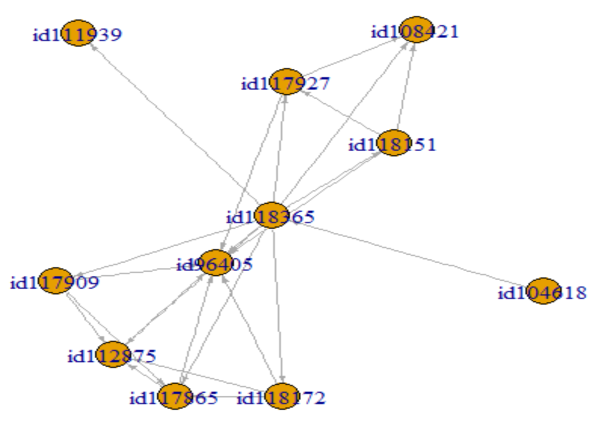
1. Neighborhood of vertices citing 'id118365' (only one):



2. Neighborhood of the nine vertices ‘id118365’ cites:



3. Neighborhood of vertices around ‘id118365’ (combination of previous two plots):



5.2 Closeness (Harmonic) Centrality Visualization ('id2959750')

'id2959750' was *Texas Dept. of Housing and Community Affairs v. Inclusive Communities Project, Inc., 2015 U.S. LEXIS 4249 (2015)*:
https://www.courtlistener.com/opinion/2959750/texas-dept-of-housing-and-community-affairs-v-inclusive-communities-project/

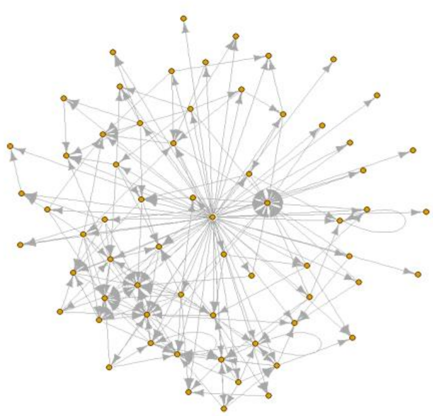
Advice from experts in law is preferred to accurately decide whether this court case had a major impact in U.S. law or society. Although this court case is not cited by any opinions according to *CourtListener*, this is most likely due it being a relatively new court case. Our hypothesis is that this court case does seem to be relatively significant, since a supreme court decision was made around neighborhood segregation and the famous, *Fair Housing Act*:

> *"policies that segregate minorities in poor neighborhoods, even if they do so unintentionally, violate the Fair Housing Act. In a 5-4 decision, the court ruled that so-called "disparate-impact claims"—claims that challenge practices that adversely affect minorities—can be brought under the Fair Housing Act. However, the court warned against remedies that impose outright racial quotas, a sign that disparate-impact claims must be brought cautiously" (Semuels, 2015).*

Interestingly, closeness/harmonic centrality measure produces one of the most recent court cases that is extremely relevant to today's U.S. society, touching on topics of racism and housing discrimination. Further studies are desired to decide whether the closeness/harmonic centrality has any relation to the recent and present nature of this court case.

1. Neighborhood of vertices citing 'id2959750' does not exist, which is expected due to 'id2959750' being a relatively new court case.

2. Neighborhood of 64 vertices 'id2959750' cites (node at center is 'id2959750'):



3. The combination of the above two plots is the same as the network graph above**.**

5.3 Eigenvector Centrality Visualization ('id106514')

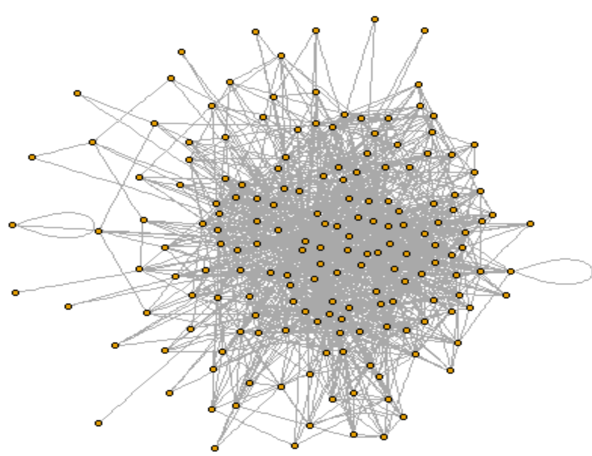
'id106514' was *NAACP v. Button, 371 U.S. 415 (1963)*:
https://www.courtlistener.com/opinion/106514/naacp-v-button/

Advice from experts in law is preferred to accurately decide whether this court case had a major impact in U.S. law or society. However, our hypothesis is that this court case does seem influential in real-world context, since it is cited by 1,674 opinions according to *CourtListener*. This court case seems significant, as there was a 6-3 ruling that:

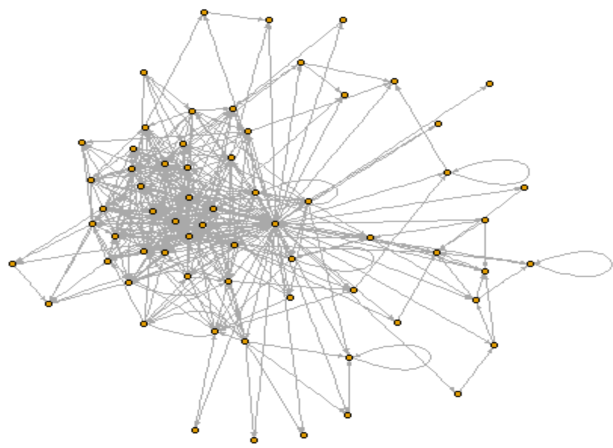
> *"Brief Fact Summary. A Virginia statute banning "improper solicitation of any legal or professional business" as applied to the NAACP was held unconstitutional because the NAACP uses litigation as a form of political expression." (Casebriefs LLC, 2016)*
>
> *"Synopsis of Rule of Law. The First Amendment of the United States Constitution (Constitution) protects vigorous advocacy of lawful ends against government intrusion." (Casebriefs LLC, 2016)*

Due to Virginia's statute to be deemed as a violation of the First Amendment and holding "racially discriminatory purpose" (*Casebriefs LLC, 2016*), this was an important case to classify NAACP's (National Association for the Advancement of Colored People) future litigations as constitutionally sound. This may explain the relatively high number of opinions this court case is cited by, according to *CourtListener* (174 citations). Further research is desired to concretely connect the relationship of this court case's historical and social significance to the eigenvector centrality. Examining future court cases surrounding the NAACP and seeing whether they cite this court case would be of interest.

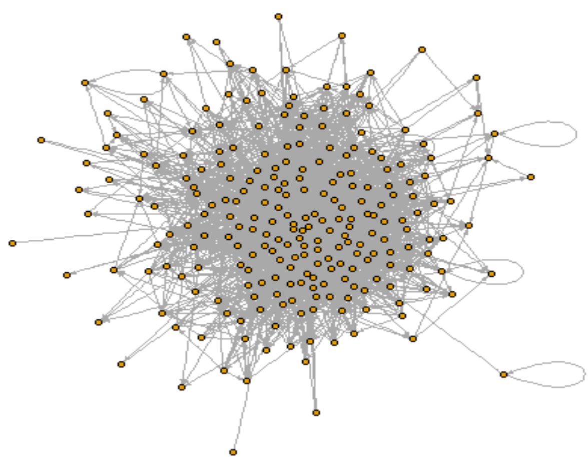
1. Neighborhood of 174 vertices citing 'id106514':



2. Neighborhood of 64 vertices ‘id106514’ cites:



3. Neighborhood of vertices around ‘id106514’ (combination of previous two plots):

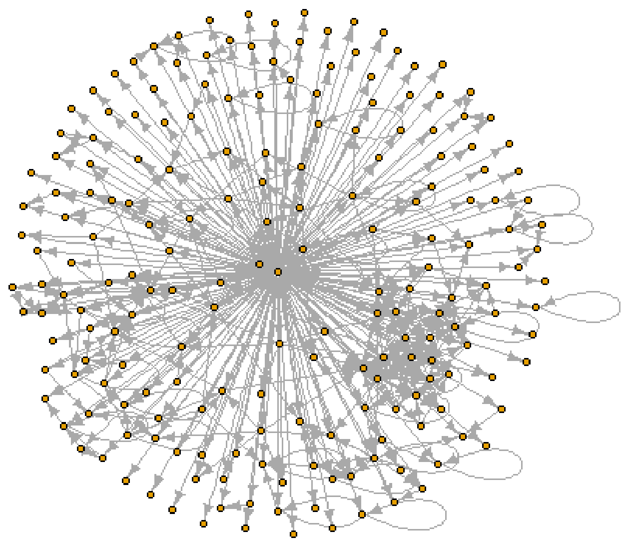


5.4 Maximum Out-Degree Visualization ('id104616')

'id104616' was *Commissioner v. Estate of Church, 335 U.S. 632 (1949)*:
https://www.courtlistener.com/opinion/104616/commissioner-v-estate-of-church/

Advice from experts in law is preferred to accurately decide whether this court case had a major impact in U.S. law or society. However, our hypothesis is that this court case does not seem to be highly influential in real-world context, since it is cited by only 81 opinions according to *CourtListener*. It seems that this court just has a rather high number of citations of other past court cases.

Neighborhood of 197 vertices 'id104616' cites (node at center is 'id104616'):



5.5 Maximum In-Degree Visualization ('id96405')

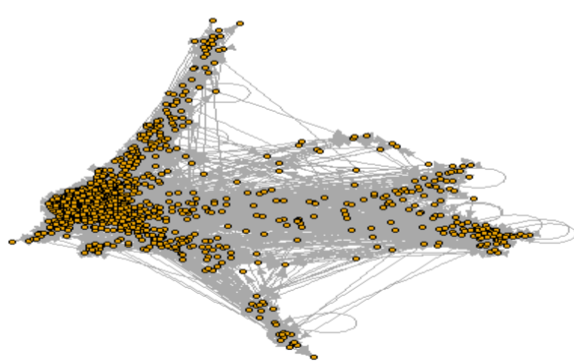
'id96405' was *United States v. Detroit Timber & Lumber Co., 200 U.S. 321 (1906)*:
https://www.courtlistener.com/opinion/96405/united-states-v-detroit-timber-lumber-co/

Advice from experts in law is preferred to accurately decide whether this court case had a major impact in U.S. law or society. Our hypothesis is that this court case does seem to be highly influential in real-world context, since it is cited by 693 opinions according to *CourtListener*. This court case seems significant, as noted by *337 of the court case:

> *"In the first place, the headnote is not the work of the court, nor does it state its decision though a different rule, it is true, is prescribed by statute in some States. It is simply the work of the reporter, gives his understanding of the decision, and is prepared for the convenience of the profession in the examination of the reports. In the second place, if the patent referred to in that headnote is a patent issued upon a wrongful entry, no such fact appeared in the case, because no patent was issued upon the entry charged to have been wrongful, but after that entry had been cancelled, a patent was issued to Diller on a new entry. If it refers to some other patent than one issued upon a wrongful entry, it has no pertinency, for the doctrine of relation never carries a patent back to the date of any other entry than that upon which it is issued. And finally the headnote is a misinterpretation of the scope of the decision."*

This text refers to the legal misusage by attorneys representing United States, as they relied on a reporter's headnote of a previous court case as legal evidence in their presentation of the argument, which the court ruled as a "*misinterpretation of the scope of the decision*" for that case. This court case warns attorneys to check the actual text of a court case to support their arguments. This court case may have become a reference point for future attorneys to not make the same mistake. This may explain why the in-degree for this court case (1295) is significantly higher than the second-most and third-most cited court cases, which have in-degrees of only 297 and 271, respectively.

Neighborhood of 1,295 vertices citing 'id96405':



5.6 Different Layouts for Visualization

Kolaczyk and Csárdi lists several layout options to represent network plots, which are essentially options to determine the "placements of vertices and edges in space" (*30*).

They mention:

- *circular layout,* where the vertices are placed to match the circumference of a circle and then edges are drawn (*31*)
- *Fruchterman and Reingold layout,* influenced by *spring-embedder* methods of network-plotting, which implements the idea of attractive and repulsive forces by "associating vertices with balls and edges with springs" (*31-32*)
- *Kamada and Kawai layout*, influenced by the *energy-placement* methods of network-plotting, which implements the idea of "energy as a function of vertex positions," by placing vertices to "minimize the total system energy (of a network or system)" (*32*)
- *Reingold and Tilford layout*, influenced by *tree data structure* representations (*33*)

The report does not make use of the different layout options offered by igraph, but the reader is free to implement these options in their code. However, we note that for visualizing the neighborhoods around vertices with key centrality measures, maximum in-degree, and maximum out-degree, the different representations of these neighborhoods did not seem particularly useful.

6 ANIMATION

6.1 Shiny Application

Earlier in the report, we mentioned that the histograms of the in-degree, out-degree, and total-degree frequency of the *popular years* can be viewed.

Now, this section of the report will present links to shiny applications with animation features, holding robust information about the in-degree, out-degree, and total-degree for the SCOTUS network. Thank you to Professor Bhamidi and his wife, Frances for aiding in the implementation of the initial shiny application.

in-degree information

https://unc-chapel-hill-scotus-research-bimc.shinyapps.io/shiny_michael5_in_deg/

histogram of in-degree in each year of SCOTUS network

https://unc-chapel-hill-scotus-research-bimc.shinyapps.io/shiny_michael5_cum_in_deg/

histogram of cumulative in-degree up to some year of SCOTUS network

https://unc-chapel-hill-scotus-research-bimc.shinyapps.io/shiny_michael5_log_in_deg/

log-log plot of in-degree distribution in each year of SCOTUS network

https://unc-chapel-hill-scotus-research-bimc.shinyapps.io/shiny_michael5_cum_log_in_deg/

log-log plot of cumulative in-degree distribution up to some year of SCOTUS network https://unc-chapel-hill-scotus-research-bimc.shinyapps.io/shiny_michael5_max_in_deg/

scatterplot of maximum in-degree in each year of SCOTUS network

out-degree information

https://unc-chapel-hill-scotus-research-bimc-continued.shinyapps.io/shiny_michael6_out_deg/

histogram of out-degree in each year of SCOTUS network

https://unc-chapel-hill-scotus-research-bimc-continued.shinyapps.io/shiny_michael6_cum_out_deg/

histogram of cumulative out-degree up to some year of SCOTUS network

https://unc-chapel-hill-scotus-research-bimc-continued.shinyapps.io/shiny_michael6_log_out_deg/

log-log plot of out-degree distribution in each year of SCOTUS network

https://unc-chapel-hill-scotus-research-bimc-continued.shinyapps.io/shiny_michael6_cum_log_out_deg/

log-log plot of cumulative out-degree distribution up to some year of SCOTUS network

https://unc-chapel-hill-scotus-research-bimc-continued.shinyapps.io/shiny_michael6_max_out_deg/

scatter plot of maximum out-degree in each year of SCOTUS network

total-degree information

https://unc-chapel-hill-scotus-research-bimc-continued2.shinyapps.io/shiny_michael7_total_deg/

histogram of total-degree in each year of SCOTUS network

https://unc-chapel-hill-scotus-research-bimc-continued2.shinyapps.io/shiny_michael7_cum_total_deg/

histogram of cumulative total-degree up to some year of SCOTUS network

https://unc-chapel-hill-scotus-research-bimc-continued2.shinyapps.io/shiny_michael7_log_total_deg/

log-log plot of total-degree distribution in each year of SCOTUS network

https://unc-chapel-hill-scotus-research-bimc-continued2.shinyapps.io/shiny_michael7_cum_log_total_deg/

log-log plot of cumulative total-degree distribution up to some year of SCOTUS network

https://unc-chapel-hill-scotus-research-bimc-continued2.shinyapps.io/shiny_michael7_max_total_deg/

scatter plot of maximum total-degree in each year of SCOTUS network

a) Notes on Shiny Applications

Many of the shiny apps are self-explanatory. Further inquiries can be asked to UNC Chapel Hill's Networks Analysis Group's member, Michael Kim.

Note: the cumulative histograms and cumulative log-log plots of the degree information up to the last year of 2015 represent the respective histograms and log-log plots of the degree information for the entire network.

Note: In each of the log-log plots, whether for total-degree, in-degree, or out-degree, there is an increasing rate of decay over time, hinting at an idea that a stronger method might be required to assess a linear decay trend for the degree distribution. A more in-depth look in this method is desired for future studies (*Kolaczyk and Csárdi, 46*).

6.2 Animation with GIFs

In addition to animation on Shiny, further animation can be completed through GIFs.

a) In-Degree GIF

Initially, the hope was to be able to graph the culmination of court cases citing the vertex with the highest in-degree ('id96405'). However, due to the high number of cases citing it (1,295 cases), the algorithm in igraph and R fails to efficiently produce the graphs.

Instead, a more manageable vertex was selected to test the algorithm. The vertex used to illustrate our algorithm is 'id93405, *Bardon v. Northern Pacific R. Co., 145 U.S. 535 (1892)* (https://www.courtlistener.com/opinion/93405/bardon-v-northern-pacific-r-co/).

The next link provides the GIF for the animation of the court cases citing 'id93405,' starting from its birth year, 1892, and up to 1983, the last year this court case was cited: http://imagizer.imageshack.us/a/img922/2521/a8cVRt.gif

b) Out-Degree GIF

A similar GIF was produced for the highest out-degree vertex. The highest out-degree vertex was 'id104616' with an out-degree of 197. Like mentioned above, 'id104616' was *Commissioner v. Estate of Church, 335 U.S. 632 (1949)* (https://www.courtlistener.com/opinion/104616/commissioner-v-estate-of-church/).

The next link provides the GIF for the animation of the court cases which 'id104616' cites, starting from the first court case cited by 'id104616' in 1884, and up to 1949, the last court case cited by 'id104616':
http://imagizer.imageshack.us/a/img921/2461/v8N9Sc.gif

c) Network Plot of Each Year GIF

It was also thought to be interesting to plot a culmination of all of SCOTUS's court cases throughout the years. However, due to computational inefficiency of igraph and R in just visualizing a large, stationary network, this animation process is deemed not possible currently.

Instead, the network of cases in each year was plotted instead: http://imagizer.imageshack.us/a/img921/5007/vYcMWI.gif

7 CONCLUSION

This report serves as an introductory exposure for those beginning in network analysis, showing important but simple ideas on how one may extract useful information out of big network data. We performed various network analysis techniques implemented in the igraph package of the R software, such as acquiring preliminary information on the global network structure, key network characteristics, and visualizations/animations of networks and histograms for the key network characteristics. Fortunately, after trials of numerous network analysis techniques, some of the court cases with key network characteristics produced meaningful patterns and relationships with the cases' real-world context. The case with the highest closeness/harmonic centrality was somehow one of the most recent court cases from SCOTUS involving the famous *Fair Housing Act*. Moreover, the case with the highest eigenvector centrality was responsible for allowing future NAACP litigations to not be subjected under statures violating the first amendment for their political expression. Finally, and most interestingly, the court case with the highest in-degree has approximately 1,000 more citations than the court case with the second highest in-degree, thus allowing us to hypothesize that this court case may have been referenced many times in future court cases to warn against attorneys to not solely depend on a reporter's interpreted summary of a case for evidence/argument.

8 REFERENCES

Kolaczyk, Eric D., and Gábor Csárdi. *Statistical Analysis of Network Data with R*. New York: Springer, 2014. Print.

"Union Bank of Chicago v. Kansas City Bank, 136 U.S. 223 (1890)." *CourtListener*. N.p., n.d. Web. 25 Apr. 2016.

"Fischer v. United States, 529 U.S. 667 (2000)." *CourtListener*. N.p., n.d. Web. 25 Apr. 2016.

"Texas Dept. of Housing and Community Affairs v. Inclusive Communities Project, Inc., 2015 U.S. LEXIS 4249 (2015)." *CourtListener*. N.p., n.d. Web. 25 Apr. 2016.

Semuels, Alana. "Supreme Court vs. Neighborhood Segregation." *The Atlantic*. Atlantic Media Company, 25 June 2015. Web. 25 Apr. 2016.

"NAACP v. Button, 371 U.S. 415 (1963)." *CourtListener*. N.p., n.d. Web. 25 Apr. 2016.

"NAACP v. Button." *CaseBriefs*. CaseBriefs LLC 2016, n.d. Web. 25 Apr. 2016.

"Commissioner v. Estate of Church, 335 U.S. 632 (1949)." *CourtListener*. N.p., n.d. Web. 25 Apr. 2016.

"United States v. Detroit Timber & Lumber Co., 200 U.S. 321 (1906)." *CourtListener*. N.p., n.d. Web. 25 Apr. 2016.

"Bardon v. Northern Pacific R. Co., 145 U.S. 535 (1892)." *CourtListener*. N.p., n.d. Web. 25 Apr. 2016.