



3.2	Data organization . . . . .	2
<b>4</b>	<b>Analysis</b>	<b>7</b>
4.1	Macro topics . . . . .	7
4.2	CSV files and validation . . . . .	7
4.3	HTTP(S) server behavior . . . . .	8
4.4	Construction of network graph from CSV files . . . . .	8
<b>5</b>	<b>Results</b>	<b>13</b>
5.1	Macro metrics . . . . .	13
5.2	Degree distributions . . . . .	14
5.3	Diameter (or longest path) . . . . .	18
5.4	Title interconnectiveness . . . . .	22
5.5	Centrality measurements . . . . .	22
<b>6</b>	<b>Conclusion</b>	<b>24</b>
<b>A</b>	<b>List of Code of Virginia Titles</b>	<b>25</b>
<b>B</b>	<b>Code of Virginia sections that refers to itself</b>	<b>29</b>
<b>C</b>	<b>Centrality</b>	<b>31</b>
<b>D</b>	<b>Miscellaneous files</b>	<b>33</b>
<b>E</b>	<b>References</b>	<b>34</b>

## List of Tables

1	Code of Virginia hierarchical organization with examples. . . . .	5
2	Different types of URLs. . . . .	9
3	Count of different types of URLs in CSV data. . . . .	9
4	The six sections with highest OUT links. . . . .	17
5	The six sections with highest IN links. . . . .	17
6	The longest augmented path in the CSV data. . . . .	18
7	A list of Code of Virginia Titles. . . . .	25

## List of Figures

1	The Code of Virginia start page. . . . .	3
2	The Code of Virginia download page. . . . .	4

3	Reference to the Editor in a section title. . . . .	6
4	Sample web page resolvable with multiple URLs. . . . .	10
5	An example of a self-referential page. . . . .	12
6	Wordcloud based on words from Code of Virginia titles. . . . .	13
7	OUT degree distribution in three resolutions. . . . .	15
8	IN degree distribution in three resolutions. . . . .	16
9	Wordcloud based on words from Code of Virginia titles longest path. . . . .	21
10	Heatmap of Code of Virginia linkages. . . . .	23

# 1 Introduction

We explore the Code of Virginia as a network graph, using citation analysis techniques. During the exploration, we report the size and connectiveness of the graph, how many of the graph nodes are self referential, and the graph's longest path. We uncover some "hidden" problems with the data (including references to notes that do not appear to be part of the publicly available record), and inconsistencies in format that challenges machine driven exploration.

This report looks at the Code of Virginia as a HTML linked structure. We do not make any statements about the relative merits, or relationships implied by the HTML linkages, just that the HTML pages are connected in some way via HTML link tags.

## 2 Purpose

Almost all laws do not exist in a vacuum. The one exception being the foundational set of laws (sometimes called a constitution), usually brought into being by a change in government. Because laws are built upon, and refer to other laws, they can be viewed as a "tree" where each law is a branch or leaf, whose origin can be traced back to a larger branch, or trunk. Sometimes a law will refer to another law for some reason. When this happens the law "tree" structure can look as if it has vines growing on it, whereby it is possible to go from one "tree" leaf to another without going all the way back to the trunk. When these "vine" link pathways from one leaf to another exist, then the "tree" can be logically transformed into a graph where tree leaves are called nodes (or vertices), and the connections between the leaves are called arcs (or edges). An edge exists between one or more nodes (an edge cannot exist without a node). Nodes can exist without any edges.

One example of this graph structure is the United States Code (USC)[4].

*"The US Code represents the codification of the laws of the United States. While it is a well-organized and curated corpus of documents, the legal text remains nearly impenetrable for non-lawyers. In this paper, we treat the US Code as a citation network and explore its complexity using traditional network metrics. We find interesting topical patterns emerge from the citation structure, and begin to interpret network metrics in the context of the legal corpus. This approach has potential for determining policy dependency and robustness, as well as modeling of future policies."*

Lyte, et al. [4]

The purpose of this report is to see if a graph structure exists for the Code of Virginia.

*"The Code of Virginia contains the laws passed by the General Assembly and signed by the Governor for the Commonwealth of Virginia."*



This report looks at the Code of Virginia as a HTML linked structure. We do not make any statements about the relative merits, or relationships implied by the HTML linkages, just that the HTML pages are connected in some way via HTML link tags.

## 3 Sources of data

We discuss where and how the data is available, and some interesting aspects of the data.

### 3.1 Data availability

The Code of Virginia is available from several different sources within the `law.lis.virginia.gov` domain. Virginia contracts out the creation of the domain content to LexisNexis, and can make minor changes to domain content through the Division of Legislative Automated Systems (DLAS) at `http://dlas.virginia.gov/`. Because there is one organization that presents a complete, and consistent domain, and another that can make minor changes to the domain, it is possible for differences to appear between the original domain, and the maintained domain.

A complete and current Code of Virginia can be downloaded via web-page scraping starting from the root page (see Figure 1). Web-scraping can take a long time, and be error-prone when the pages have different cascading style sheets (CSS), or are inconsistent. The LIS makes the Code of Virginia also available in other formats, including electronic publication format (EPUB), MOBI for electronic book readers, portable document format (PDF), and comma separated value (CSV). Presumably, these different file formats are consistent with the LexisNexis deliverables, prior to any modifications by DLAS staff (see Figure 2).

We elected to download the CSV files.

### 3.2 Data organization

The Code of Virginia is organized into a hierarchical structure, and depending on where you look at the data, the names of the layers differ. Parsing a sectional web page (either via scraping, or from a CSV file), can be problematic because the Body of the page may have a section that appears to be links to historical data, but the body does not have a “marker” as to where the historical data begins. Also, the historical section may contain links to external programs (as indicated by the text `cgi-bin`), intermingled with indistinguishable links to other pages (see Table 1).

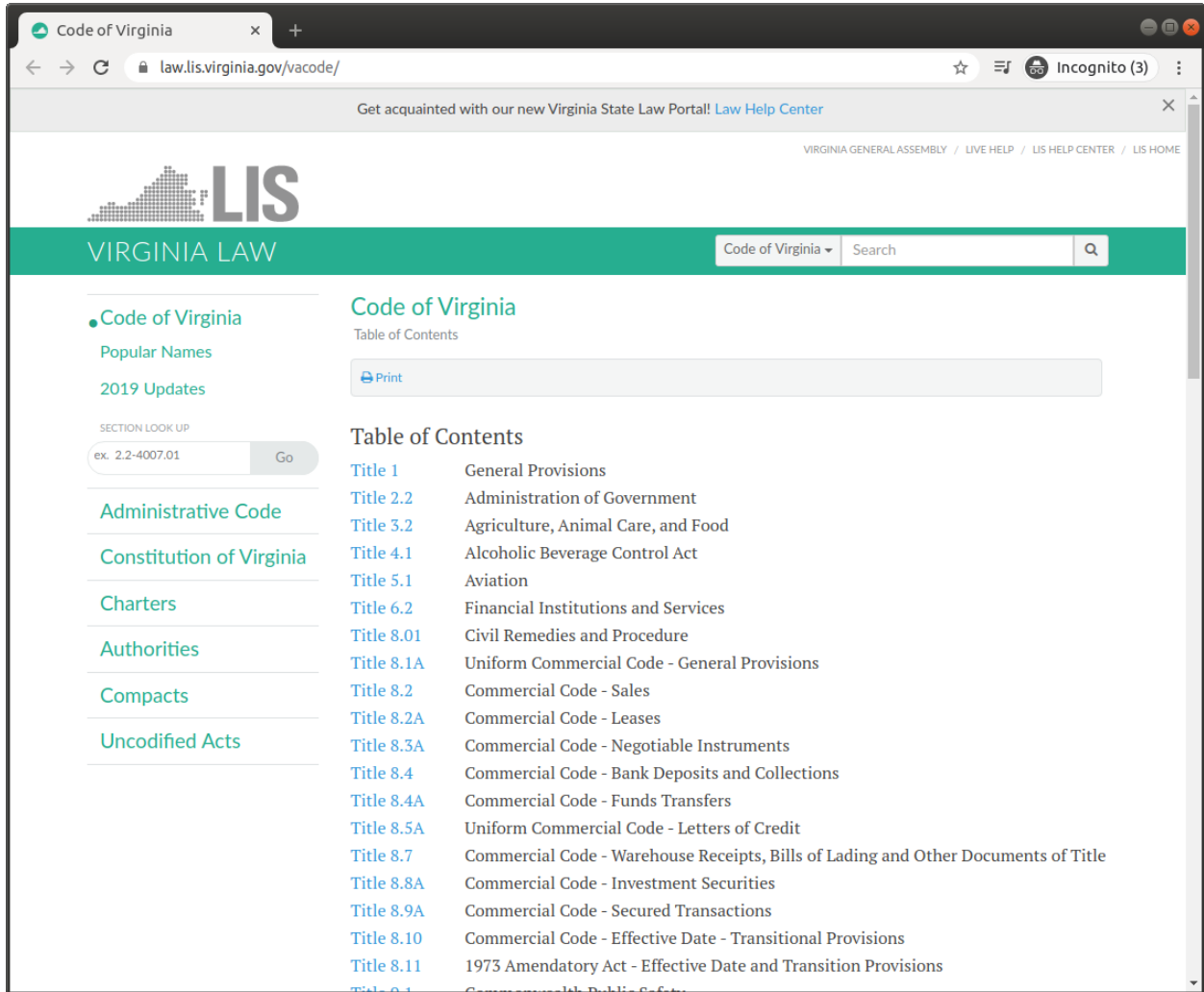


Figure 1: The Code of Virginia start page. The 76 different titles have their “root” on this page <https://law.lis.virginia.gov/vacode/>. Pages linked to from this page can be modified by DLAS.

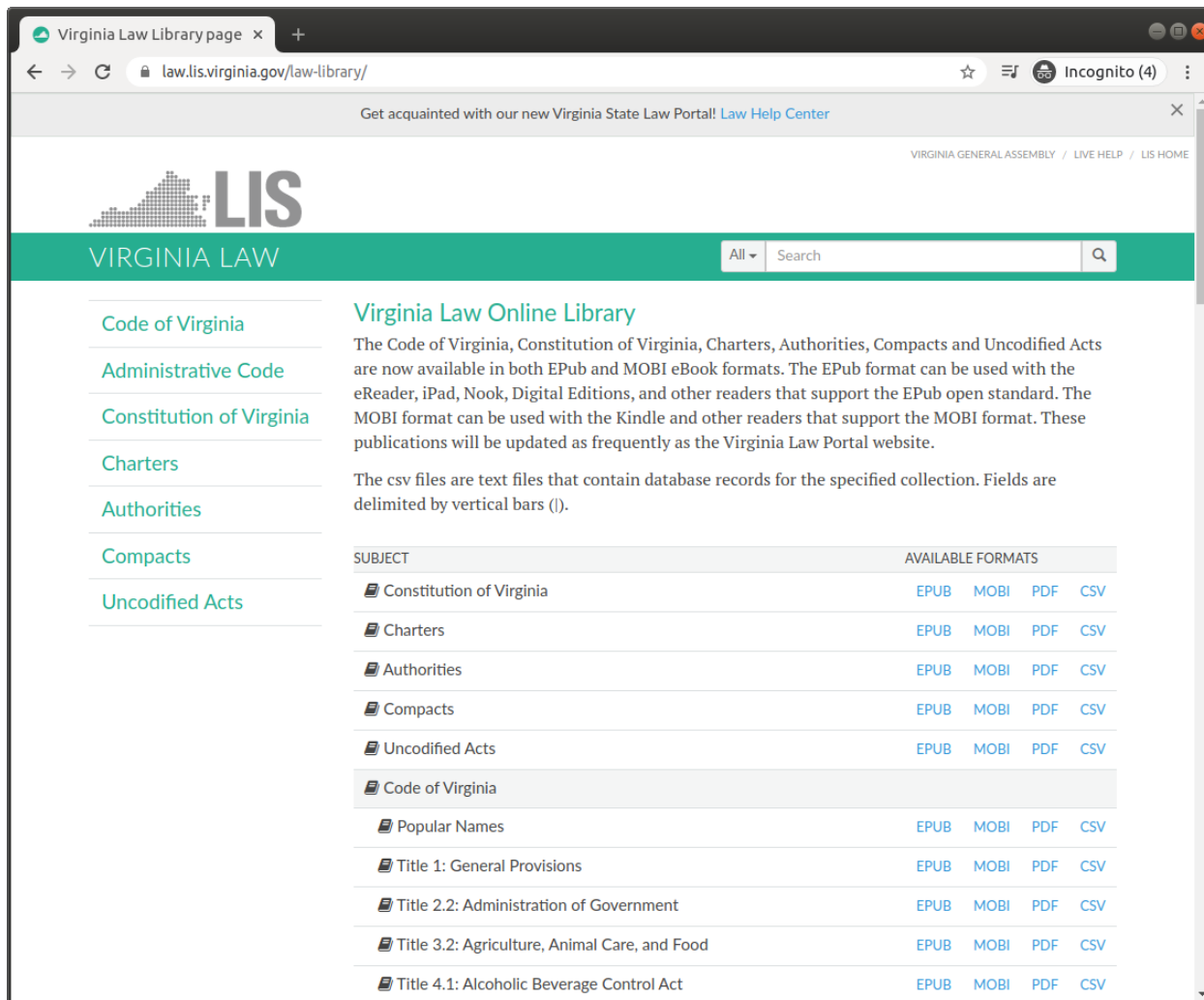


Figure 2: The Code of Virginia download page. Code of Virginia (as well as other parts of Virginia law) can be downloaded in different formats <https://law.lis.virginia.gov/law-library/>.

Table 1: Code of Virginia hierarchical organization with examples.

<b>From CSV files</b>	<b>From web pages</b>	<b>Example web page</b>
TitleNum–TitleName	Title	<a href="https://law.lis.virginia.gov/vacode/">https://law.lis.virginia.gov/vacode/</a>
SubtitleNum–SubtitleName	(sometimes subdivide chapters)	<a href="https://law.lis.virginia.gov/vacode/title8.9A/part6/part1/">https://law.lis.virginia.gov/vacode/title8.9A/part6/part1/</a>
PartNum–PartName	(sometimes take the place of chapters)	<a href="https://law.lis.virginia.gov/vacode/title8.7/">https://law.lis.virginia.gov/vacode/title8.7/</a>
ChapterNum–ChapterName	Chapter number – Chapter name	<a href="https://law.lis.virginia.gov/vacode/title1/">https://law.lis.virginia.gov/vacode/title1/</a>
ArticleNum–ArticleName	(sometimes subdivide chapters)	<a href="https://law.lis.virginia.gov/vacode/title53.1/chapter2/">https://law.lis.virginia.gov/vacode/title53.1/chapter2/</a>
SubPartNum–SubPartName	(sometimes subdivide parts)	<a href="https://law.lis.virginia.gov/vacode/title8.9A/part6/part1/">https://law.lis.virginia.gov/vacode/title8.9A/part6/part1/</a>
Section–Title–Body	Section – Title – Body	<a href="https://law.lis.virginia.gov/vacode/1-200/">https://law.lis.virginia.gov/vacode/1-200/</a>

An additional problem with the data, is the inclusion of directions to “see Editor’s note” in a section title (see Figure 3). It isn’t clear from the page how to interpret, or locate the note.

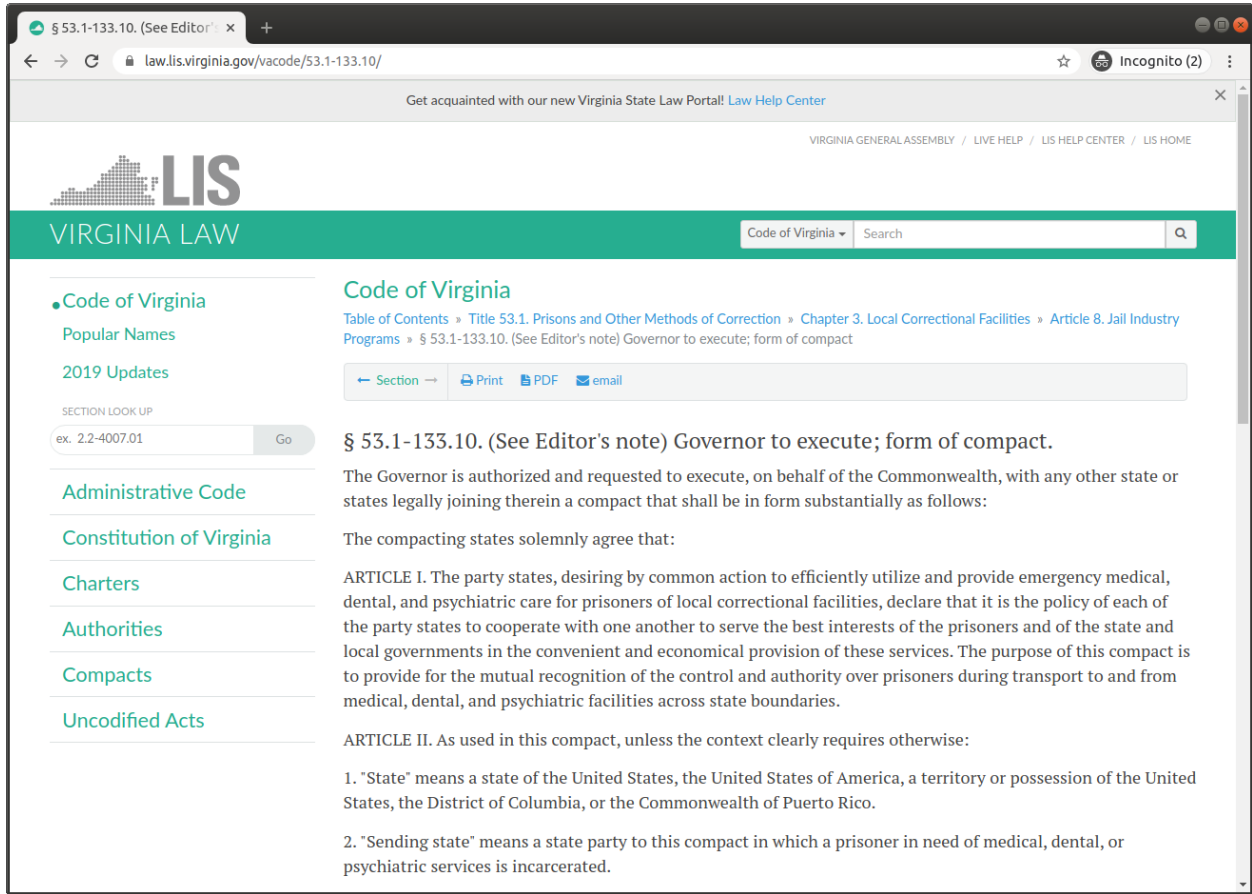


Figure 3: Reference to the Editor in a section title. From this page <https://law.lis.virginia.gov/vacode/53.1-133.10/>. There are six such statements in the data.

## 4 Analysis

Analysis of the network structure will start at the macro level, and then proceed to more detailed items.

### 4.1 Macro topics

The CSV files appear to be one large file that is segmented based on the Code of Virginia title number. The first file in this collection is “CoVTitle\_1.csv” has a header line listing all the fields. The header identifies these fields:

1. TitleNum	6. PartName	11. SubPartNum
2. TitleName	7. ChapterNum	12. SubPartName
3. SubtitleNum	8. ChapterName	13. Section
4. SubtitleName	9. ArticleNum	14. Title
5. PartNum	10. ArticleName	15. Body

### 4.2 CSV files and validation

While the files are labeled as CSV, they are not. The delimiter is actually the pipe symbol (|). Comparison of the raw web pages, and CSV files showed that fields 13 (Section) and 15 (Body) were the ones that could be used to “link” the pages together. The Body field contains the HTML links to other pages based on the Section. The raw CSV files were “validated” looking at a number of things including that each:

1. line in each file had 15 fields, no more and no less, and
2. Section field had some sort of data, and
3. Body field contained valid HTML. The Body may, or may not contain links to other pages, but that it must contain valid HTML. HTML is a specialized subset of the extensible markup language (XML). So HTML has to pass an XML checker.

Here are some sample results:

```
[1] "File ../Data/CSVData-Orig/CoVTitle_1.csv, line 1, field 15, Body"
<XMLInputError: XML content does not seem to be XML: ''>
[1] "File ../Data/CSVData-Orig/CoVTitle_15.2.csv, line 1569, field 15, Repealed by Acts
<XMLInputError: XML content does not seem to be XML: '5</a>, cl. 11, effective October 1
[1] "File ../Data/CSVData-Orig/CoVTitle_15.2.csv, line 1623, field 15, Repealed by Acts
```

```

<XMLInputError: XML content does not seem to be XML: '5</a>, cl. 11, effective October 1
...
[4] "file ../Data/CSVData-Orig/CoVTitle_15.2.csv line 2166 field 13 has 0 characters"
[5] "file ../Data/CSVData-Orig/CoVTitle_15.2.csv line 2167 field 13 has 0 characters"
[6] "file ../Data/CSVData-Orig/CoVTitle_19.2.csv line 713 has 14 vice 15"
[7] "See this URL https://law.lis.virginia.gov/vacode/title19.2/chapter25/section19.2-
[8] "file ../Data/CSVData-Orig/CoVTitle_19.2.csv line 713 has 14 fields unable to chec
[9] "file ../Data/CSVData-Orig/CoVTitle_2.2.csv line 1144 field 13 has 0 characters"
...

```

A complete list of validation results is included (see Section D). The files that failed validation were edited manually to correct the fields that failed validation.

### 4.3 HTTP(S) server behavior

The Body field contains the links to other pages via HTML Anchor tags. An Anchor tags can contain any one of several different types of URLs (Uniform Resource Locator). A URL is a special case of a uniform resource identifier (URI)[1]. For our purposes, URLs can be categorized into four different types (see Table 2).

There appears to be an process on the LIS HTTP servers that interprets URLs (links) in an interesting way. The following URLs will resolve to the same web page (see Figure 4):

- <https://law.lis.virginia.gov/vacode/title3.2/chapter3.1/section3.2-306/>
- <https://law.lis.virginia.gov/vacode/title3.2/AAA/section3.2-306/>
- <https://law.lis.virginia.gov/vacode/title3.2/section3.2-306/>
- <https://law.lis.virginia.gov/vacode/AAA/section3.2-306/>

It appears that the server keys off the last part of the URL, as long as there is a next-to-last part that does not equal the characters “vacode”.

### 4.4 Construction of network graph from CSV files

Links were extracted from the Body field from each line in CSV file. These links were “connected” from the source page to a destination page, possibly in another CSV file. In this manner a connected graph was created from the CSV files. During the creation process, additional “artificial” nodes were added to the graph to replicate the way a user might access the Code of Virginia graph via the internet. These artificial nodes ensured that all web pages from the CSV files were reachable. A web page might be unreachable, if there are not any pages that link to it.

Table 2: Different types of URLs. There are several different types of URLs. This is a notional list based on the canonical source[1].

<b>Notional name</b>	<b>Example</b>	<b>Explanation</b>
Absolute URI	<code>http://www.example.com/images/icons.png</code>	Contains the scheme, a colon (“:”), double slashes (“//”), and an authority path.
Relative path reference	<code>http://../images/icons.png</code>	Contains the scheme, a colon (“:”), double slashes (“//”), and a document path relative to the current document.
Absolute path reference	<code>http:///images/icons.png</code>	Contains the scheme, a colon (“:”), double slashes (“//”), and an absolute document path.
Network path reference	<code>//www.example.com/images/icons.png</code>	Inherits the scheme from the current scheme, double slashes (“//”), and an absolute document path.

Table 3: Count of different types of URLs in CSV data. The number of links between pages, is also the number of edges (or arcs) in the graph of the CSV pages.

<b>Notional name</b>	<b>Protocol</b>	<b>Count</b>
Absolute URI	<code>http:</code>	101,039
	<code>https:</code>	19
Absolute path reference	<code>admincode</code>	1
	<code>vacode</code>	15,065
<b>Total</b>		116,124





Figure 4: Sample web page resolvable with multiple URLs. See the text for URLs that will resolve to this page.

After the graph was created, the pages that “linked” back to themselves (self referential) were identified. One example of a self-referential page is §56-235.8. Retail supply choice for natural gas customers (see Figure 5). The page’s URL is <https://law.lis.virginia.gov/vacode/ttt/section56-235.8/>. Searching the text of the page, will locate two places where the text refers back to itself. A complete list of self referential pages can be found elsewhere (see Section B).

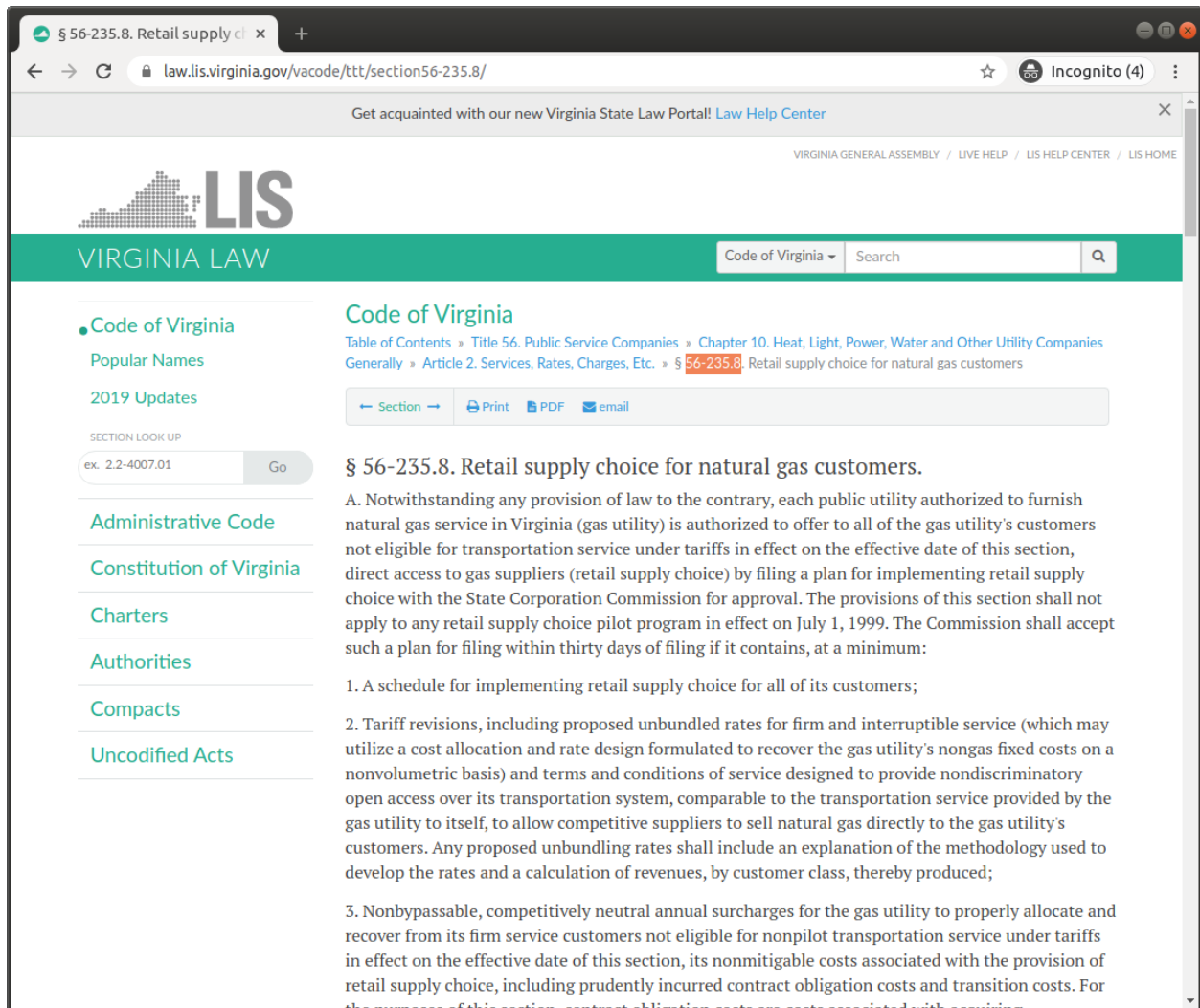


Figure 5: An example of a self-referential page. The page's URL is <https://law.lis.virginia.gov/vacode/ttt/section56-235.8/>. Searching the text of the page, will locate two places where the text refers back to itself.



## 5.2 Degree distributions

One way to “evaluate” the importance of a node, is how many nodes is it connected to. Because the web pages make up a directed graph (in that web page A links to web page B, but page B does not necessarily know that page A connects to it), a web page can have “OUT” going links, and “IN” coming links. A page can also have no links of any type. A page with lots of OUT going links could be considered to be very dependent on those other links, and a change in one of the linked page could have a profound impact, without the source page knowing about it. Conversely, a page that has many IN links, could have a profound effect on many pages without knowing about those other pages. A page with lots of OUT links could be considered “weak” because it is dependent on so many others. A page with many IN links could be considered “strong” because so many other pages depend on it. These weak and strong pages can also be called hubs and authorities[3].

We can look at the distribution of the OUT and IN links to try and identify the weak and the strong nodes (see Figures 7, and 8). The nodes with the fewest number of OUT or IN links overwhelms data from the other nodes, so three different resolutions are provided. The “weakest” and “strongest” nodes are identified (see Table 4 and Table 5)

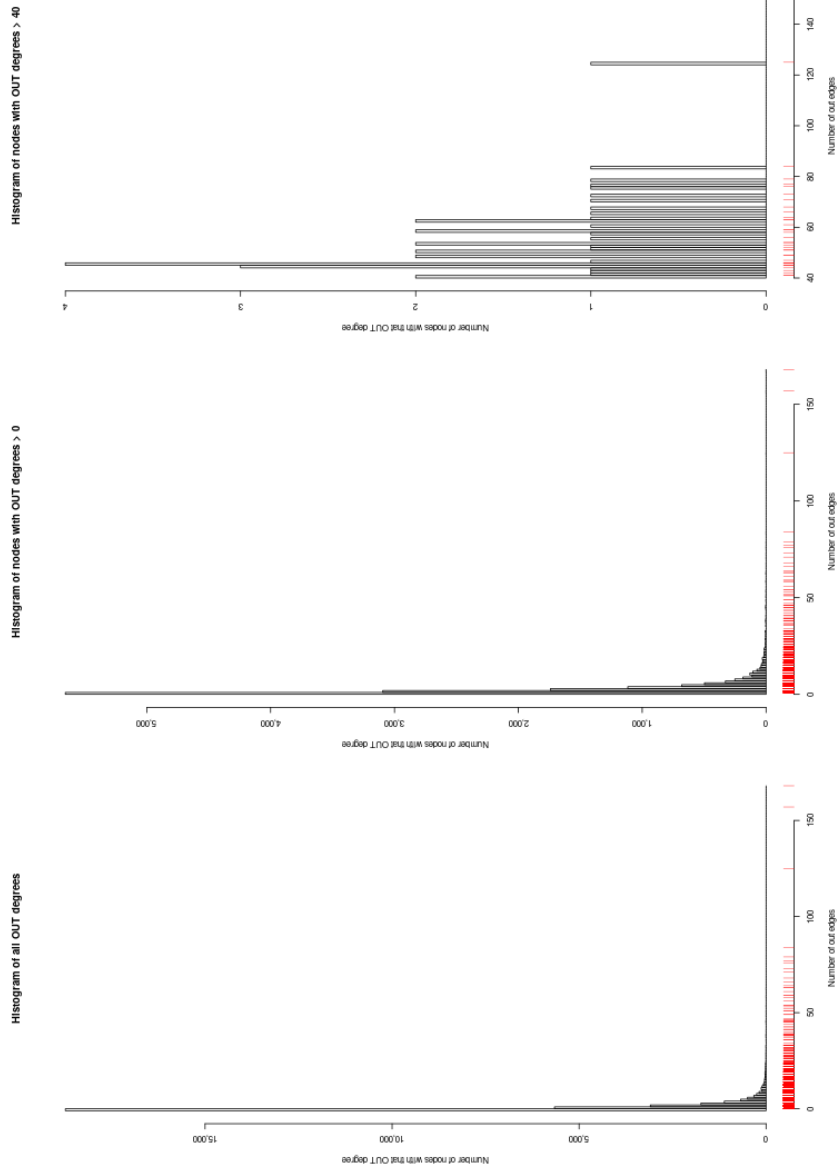


Figure 7: OUT degree distribution in three resolutions. In the figures, the Y-axis is the number of nodes per number of edges (either OUT or IN). The distribution of the edges is shown at three different resolutions because of the range of data on the Y-axis. In the leftmost image, all data is plotted, and the Y-axis ranges from 0 to a large number. This large number is driven by the number of pages that have 0 edges. The middle image shows the same data, only the most frequent value (0 or 1) is removed, and the range of the Y-axis is reduced considerably. The rightmost image shows the distribution of nodes with at least 40 edges. Beneath each image is a “rug” indicating where there is data along the X-axis, even when that data is too small to be displayed on the Y-axis.

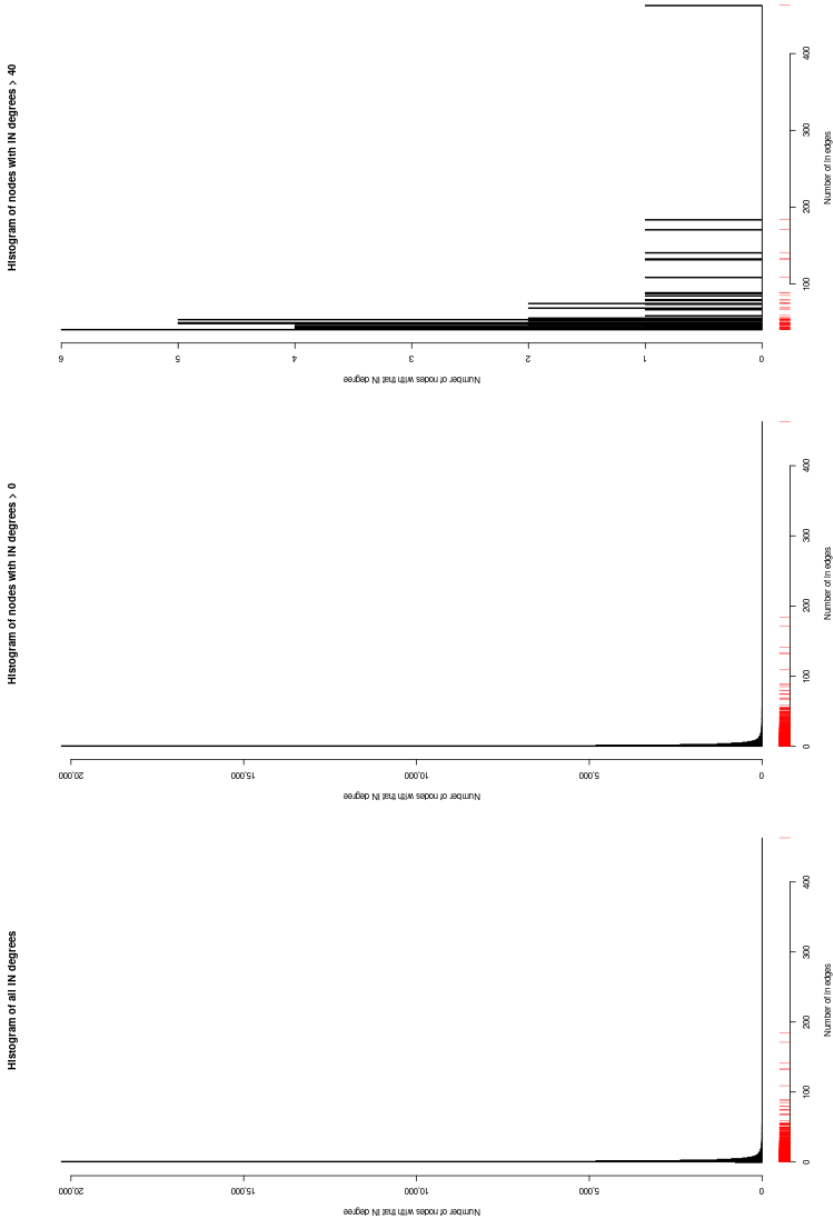


Figure 8: IN degree distribution in three resolutions. In the figures, the Y-axis is the number of nodes per number of edges (either OUT or IN). The distribution of the edges is shown at three different resolutions because of the range of data on the Y-axis. In the leftmost image, all data is plotted, and the Y-axis ranges from 0 to a large number. This large number is driven by the number of pages that have 0 edges. The middle image shows the same data, only the most frequent value (0 or 1) is removed, and the range of the Y-axis is reduced considerably. The rightmost image shows the distribution of nodes with at least 40 edges. Beneath each image is a “rug” indicating where there is data along the X-axis, even when that data is too small to be displayed on the Y-axis.

Table 4: The six sections with highest OUT links. This report looks at the Code of Virginia as a HTML linked structure. We do not make any statements about the relative merits, or relationships implied by the HTML linkages, just that the HTML pages are connected in some way via HTML link tags.

<b>Num.</b>	<b>Section</b>	<b>Title</b>	<b>Out Count</b>
1	§19.2-392.02	National criminal background checks by businesses and organizations regarding employees or volunteers providing care to children or the elderly or disabled.	168
2	§38.2-4319	Statutory construction and relationship to other laws.	157
3	§17.1-805	Adoption of initial discretionary sentencing guideline midpoints.	125
4	§38.2-4214	Application of certain provisions of law.	84
5	§32.1-127.1:03	Health records privacy.	79
6	§19.2-389	(Effective January 1, 2021) Dissemination of criminal history record information.	77

Table 5: The six sections with highest IN links. This report looks at the Code of Virginia as a HTML linked structure. We do not make any statements about the relative merits, or relationships implied by the HTML linkages, just that the HTML pages are connected in some way via HTML link tags.

<b>Num.</b>	<b>Section</b>	<b>Title</b>	<b>In Count</b>
1	§2.2-4000	TRANSACTION OF PUBLIC BUSINESS	463
2	§2.2-3700	TRANSACTION OF PUBLIC BUSINESS	184
3	§2.2-2813	Definitions; compensation and expense payments from state funds for service on collegial bodies.	171
4	§2.2-2825	Reimbursement for certain travel expenditures; restrictions on reimbursement.	141
5	§18.2-266	Driving motor vehicle, engine, etc., while intoxicated, etc.	133
6	§46.2-100	Definitions.	132



### 5.3 Diameter (or longest path)

Traditionally, the longest path is the number of edges along the geodesic path between the two most remote nodes. In our case, we are interested in the number of web pages that a visitor would have to visit in order to get from the universal starting point for the Code of Virginia to the most remote web page. We arrive at that path by using classical graph exploration techniques to determine the path based on the CSV data, and then prepend that path with the web pages necessary to get to the path’s starting web page. When we apply these algorithm to the CSV data, we create a path that is 36 nodes long (see Table 6). Based on the longest path, we can create a wordcloud of the titles along the path (see Figure 9).

Table 6: The longest augmented path in the CSV data. Classical longest path from the CSV data was prepended with the necessary nodes to get from the Code of Virginia start page to the first node.

Node	Section	Title
1	Code of Virginia	Code of Virginia
2	Title 64.2	Title 64.2
3	§64.2-1702	Appointment of guardians.
4	§64.2-1701	Testamentary guardians.
5	§64.2-1704	Guardian’s bond.
6	§64.2-1411	When fiduciary may qualify without security; requirements for issuance of certificates of qualification; payments.
7	§6.2-1003	When security not required; payment of probate taxes and fees.
8	§19.2-386.6	Bond to secure possession.
9	§19.2-386.12	Sale of forfeited property.
10	§19.2-386.10	Forfeiture; default judgment; remission; trial.
11	§19.2-386.9	Appearance by owner or lien holder.
12	§8.01-317	What order of publication to state; how published; when publication in newspaper dispensed with.
13	§8.01-316	Service by publication; when available.
14	§64.2-1426	Nonresident fiduciaries.

(Continued on the next page.)

Table 6. (Continued from the previous page.)

<b>Node</b>	<b>Section</b>	<b>Title</b>
15	§64.2-2011	Qualification of guardian or conservator; clerk to record order and issue certificate; reliance on certificate.
16	§64.2-2014	Clerk to index findings of incapacity or restoration; notice of findings.
17	§64.2-2023	Estate planning.
18	§32.1-325	Board to submit plan for medical assistance services to U.S. Secretary of Health and Human Services pursuant to federal law; administration of plan; contracts with health care providers.
19	§2.2-4002	Exemptions from chapter generally.
20	§28.2-1000.2	Annual closure of the Chesapeake Bay purse seine fishery for Atlantic menhaden.
21	§28.2-402	License fee to take menhaden with purse nets.
22	§28.2-201	Authority of Commission to make regulations, establish licenses, and prepare fishery management plans; accept federal grants; enforcement; penalty for violation of regulation.
23	§28.2-302.2:1	Special combined individual sportfishing licenses.
24	§29.1-101	Game Protection Fund.
25	§58.1-344.3	Voluntary contributions of refunds requirements.
26	§22.1-212.2:2	Educational technology foundations and public school foundations.
27	§2.2-4343	Exemption from operation of chapter for certain transactions.
28	§23.1-2213	Medical center management; capital projects; leases of property; procurement.
29	§23.1-2212	Operations of Medical Center.
30	§23.1-1116	Commonwealth not to limit revenues of institutions.

(Continued on the next page.)

Table 6. (Continued from the previous page.)

<b>Node</b>	<b>Section</b>	<b>Title</b>
31	§23.1-1106	Bonds generally.
32	§23.1-1112	Bonds; interest.
33	§6.2-303	Contracts for more than legal rate of interest.
34	§36-55.31	(Effective October 1, 2019) Powers relative to making mortgage loans and temporary construction loans to housing sponsors and persons and families of low and moderate income.
35	§36-55.33:1	Mortgage loan terms and conditions.
36	§36-55.34:1	Power to supervise housing sponsors.

(Last page.)

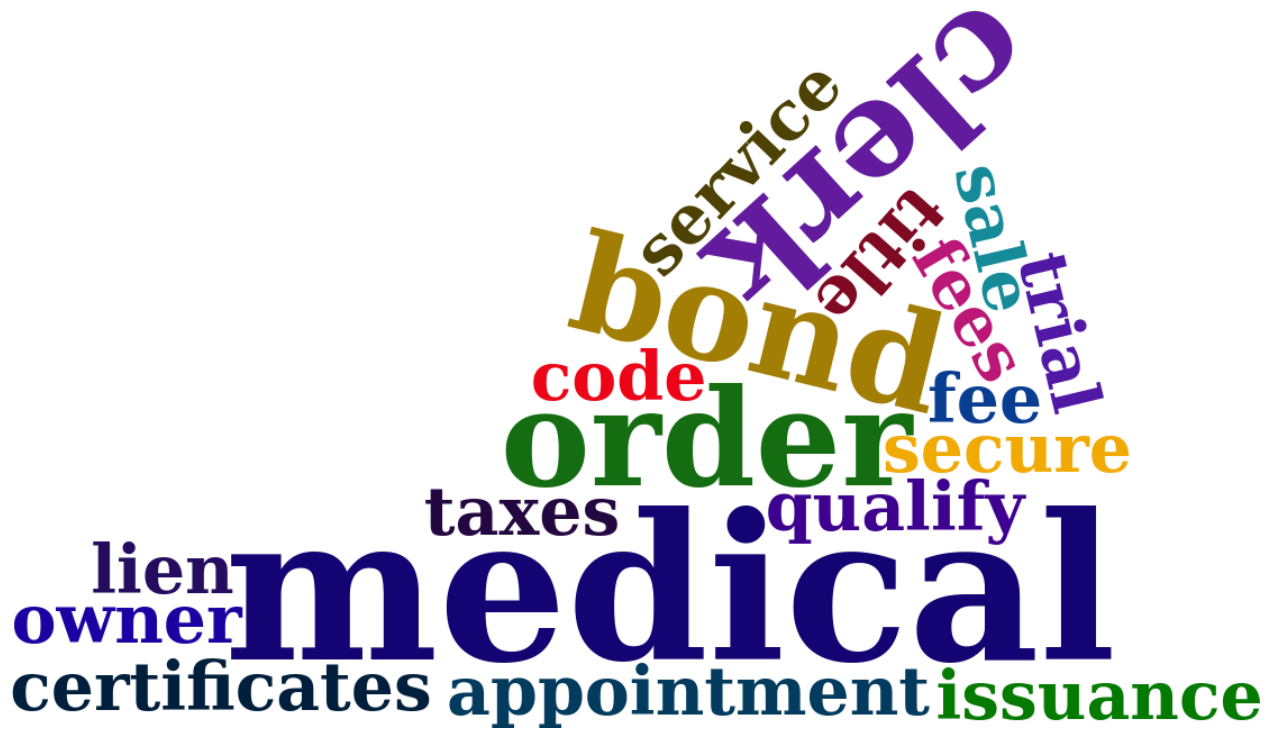


Figure 9: Wordcloud based on words from Code of Virginia titles longest path. Words in a wordcloud are present in some body of text. The size of the word is indicative of the number of times that word was present in the text. The color, location, and orientation of the word is random, and will change each time the wordcloud is “printed.”

## 5.4 Title interconnectiveness

The Code of Virginia network contains 33,469 web pages, and 80,887 links. While these numbers aren't too large for small computers to process, they are too large for reasonable representation on the "printed" page. To reduce the scale of these values, while maintaining the essence of what they represent, we will use a variation of dimensional reduction to reduce the number of nodes to 77 and the number of edges to 5,929.

One way to illustrate the "connectiveness" of the web pages is to create a "heatmap" of their linkages. Typically a heatmap uses color to represent data values. A spectrum of colors is created, subdivided into some number of "bands", each band is assigned to a portion of the portion of the data, and finally a color for a specific piece of data is selected based on which band the data falls into. A heatmap was created for the linkages between different Code of Virginia Titles (see Figure 10). The heatmap shows that most often, indices (Titles) link to themselves, and seldom to other titles. It wasn't practical to out readable labels on the X and Y axes, so sequence numbers were used. The numbers are the same as those listed in the list of Code of Virginia Titles (see Section A).

## 5.5 Centrality measurements

Centrality measurements are a way to quantify the "importance" of an edge or vertex in a graph. The size of our graph precluded computing these metrics (see Section C). One attempt to compute the edge centrality was terminated after 36 hours of 100% core run time.

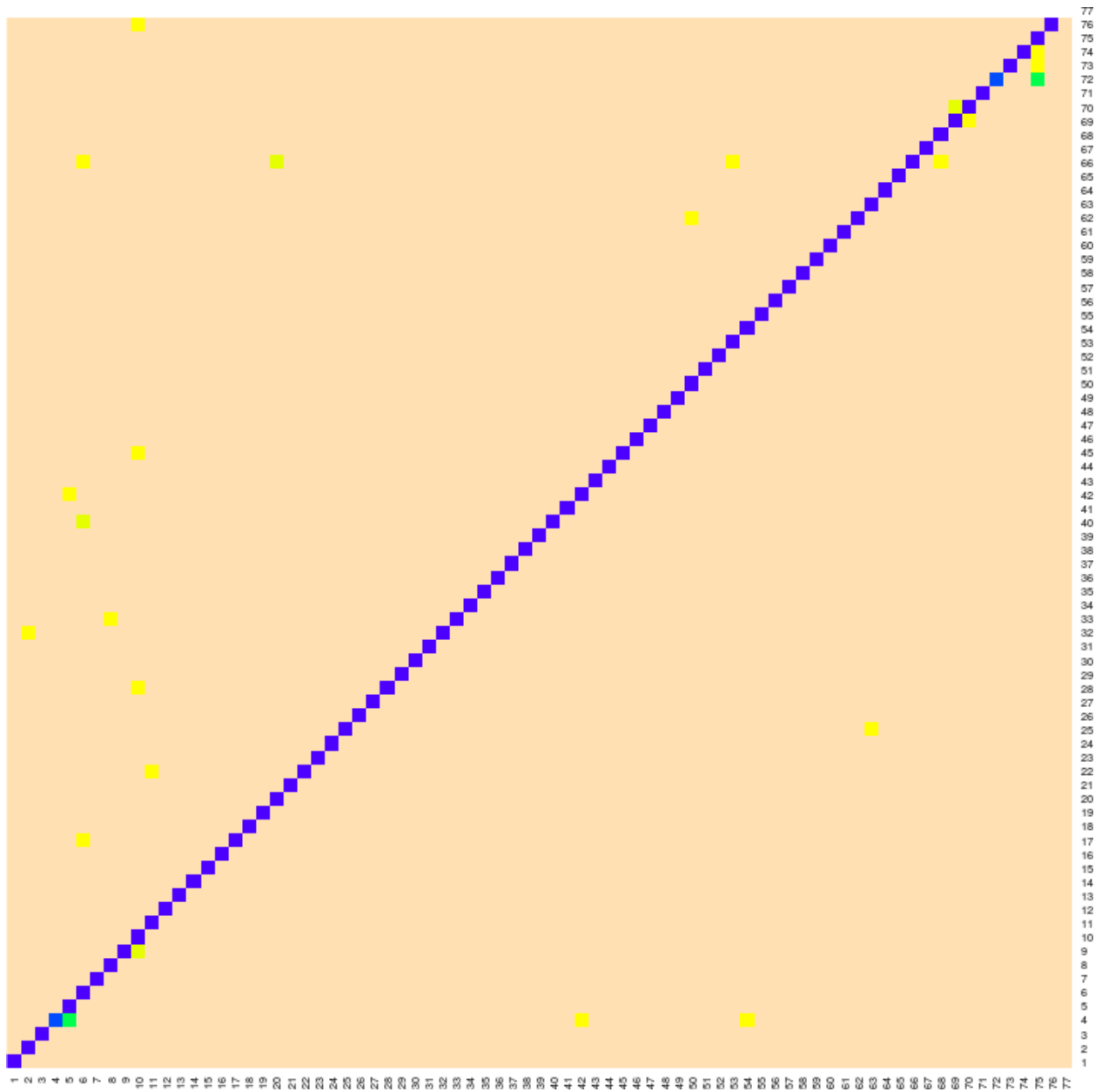


Figure 10: Heatmap of Code of Virginia linkages. The numbers along the X and Y axes correspond to the entries in (see Table 7). The X axis numbers are the ones linking to the Y axis numbers. Numbers 1 through 76 represent actual Code of Virginia Titles. 77 is an artificial node representing the “root” of the graph. The root links to all titles, but is not reached by any others. This is shown in the image by looking up the image from 77 on the X axis, seeing a color for every index along the Y axis, except for 77. Looking across from the Y axis, the darker the color, the more links there are from the X axis to that value on the Y axis. The darker the color, the more links there are. It is evident that most titles link to themselves. A few titles link to others. For example X index 75 links mostly to 75, lesser to 72, lesser still to 73 and 74, and very seldom to the other indices.

## 6 Conclusion

This report looks at the Code of Virginia as a HTML linked structure. We do not make any statements about the relative merits, or relationships implied by the HTML linkages, just that the HTML pages are connected in some way via HTML link tags. We have looked at the Code of Virginia downloaded as a series of CSV files converted to a network graph, and conducted a citation analysis of the graph looking at how the sections (nodes) were connected using basic citation analysis. Artificial nodes were added to the graph to more closely mimic how the graph could be accessed from the Internet. The resulting graph has 33,469 nodes (web pages), 80,887 of unique links (some web pages had multiple links to the same destination page), and the diameter of the graph was 35 edges, 36 nodes including the artificial nodes. 78 of the nodes were self-referential, meaning they contained links directly back to themselves. Visual examination of a heatmap showing intra- and inter- Title linking, showed that the vast majority of the Titles linked within themselves, and only somewhat to other Titles.

It wasn't possible to compute edge and vertex centrality values with the available hardware. The software attempted to compute these values for 36 hours without results.

## A List of Code of Virginia Titles

A list of Code of Virginia Titles as contained in the CSV files (see Table 7).

Table 7: A list of Code of Virginia Titles. The sequence number is used as an index by the heatmap image.

Seq. Num.	Title Num.	Title Name
1	1	GENERAL PROVISIONS
2	10.1	Conservation
3	11	CONTRACTS
4	12.1	STATE CORPORATION COMMISSION
5	13.1	CORPORATIONS
6	15.2	COUNTIES, CITIES AND TOWNS
7	16.1	COURTS NOT OF RECORD
8	17.1	COURTS OF RECORD
9	18.2	CRIMES AND OFFENSES GENERALLY
10	19.2	Criminal Procedure
11	2.2	ADMINISTRATION OF GOVERNMENT
12	20	DOMESTIC RELATIONS
13	21	DRAINAGE, SOIL CONSERVATION, SAN- ITATION AND PUBLIC FACILITIES DIS- TRICTS
14	22.1	EDUCATION
15	23.1	INSTITUTIONS OF HIGHER EDUCA- TION; OTHER EDUCATIONAL AND CUL- TURAL INSTITUTIONS
16	24.2	Elections
17	25.1	EMINENT DOMAIN
18	27	Fire Protection
19	28.2	FISHERIES AND HABITAT OF THE TIDAL WATERS
20	29.1	GAME, INLAND FISHERIES AND BOAT- ING

(Continued on the next page.)



Table 7. (Continued from the previous page.)

Seq. Num.	Title Num.	Title Name
21	3.2	AGRICULTURE, ANIMAL CARE, AND FOOD
22	30	GENERAL ASSEMBLY
23	32.1	HEALTH
24	33.2	HIGHWAYS AND OTHER SURFACE TRANSPORTATION SYSTEMS
25	34	HOMESTEAD AND OTHER EXEMPTIONS
26	35.1	Hotels, Restaurants, Summer Camps, and Campgrounds
27	36	HOUSING
28	37.2	BEHAVIORAL HEALTH AND DEVELOPMENTAL SERVICES
29	38.2	INSURANCE
30	4.1	ALCOHOLIC BEVERAGE CONTROL ACT
31	40.1	Labor and Employment
32	41.1	LAND OFFICE
33	42.1	LIBRARIES
34	43	MECHANICS' AND CERTAIN OTHER LIENS
35	44	MILITARY AND EMERGENCY LAWS
36	45.1	MINES AND MINING
37	46.2	MOTOR VEHICLES
38	47.1	NOTARIES AND OUT-OF-STATE COMMISSIONERS
39	48	NUISANCES
40	49	OATHS, AFFIRMATIONS AND BONDS
41	5.1	AVIATION
42	50	PARTNERSHIPS
43	51.1	PENSIONS, BENEFITS, AND RETIREMENT
44	51.5	PERSONS WITH DISABILITIES

(Continued on the next page.)

Table 7. (Continued from the previous page.)

Seq. Num.	Title Num.	Title Name
45	52	POLICE (STATE)
46	53.1	Prisons and Other Methods of Correction
47	54.1	Professions and Occupations
48	55.1	Property and Conveyances [Effective October 1, 2019]
49	55	Property and Conveyances [Repealed effective October 1, 2019]
50	56	PUBLIC SERVICE COMPANIES
51	57	Religious and Charitable Matters; Cemeteries
52	58.1	TAXATION
53	59.1	TRADE AND COMMERCE
54	6.2	FINANCIAL INSTITUTIONS AND SERVICES
55	60.2	UNEMPLOYMENT COMPENSATION
56	61.1	WAREHOUSES, COLD STORAGE AND REFRIGERATED LOCKER PLANTS
57	62.1	WATERS OF THE STATE, PORTS AND HARBORS
58	63.2	WELFARE (SOCIAL SERVICES)
59	64.2	Wills, Trusts, and Fiduciaries
60	65.2	WORKERS' COMPENSATION
61	66	JUVENILE JUSTICE
62	67	VIRGINIA ENERGY PLAN
63	8.01	CIVIL REMEDIES AND PROCEDURE
64	8.10	COMMERCIAL CODE – EFFECTIVE DATE – TRANSITIONAL PROVISIONS
65	8.11	1973 AMENDATORY ACT – EFFECTIVE DATE AND TRANSITION PROVISIONS
66	8.1A	UNIFORM COMMERCIAL CODE – GENERAL PROVISIONS
67	8.2	COMMERCIAL CODE - SALES
68	8.2A	COMMERCIAL CODE – LEASES

(Continued on the next page.)

Table 7. (Continued from the previous page.)

Seq. Num.	Title Num.	Title Name
69	8.3A	COMMERCIAL CODE – NEGOTIABLE INSTRUMENTS
70	8.4	COMMERCIAL CODE – BANK DEPOSITS AND COLLECTIONS
71	8.4A	COMMERCIAL CODE – FUNDS TRANSFERS
72	8.5A	UNIFORM COMMERCIAL CODE – LETTERS OF CREDIT
73	8.7	COMMERCIAL CODE – WAREHOUSE RECEIPTS, BILLS OF LADING AND OTHER DOCUMENTS OF TITLE
74	8.8A	COMMERCIAL CODE – INVESTMENT SECURITIES
75	8.9A	COMMERCIAL CODE – SECURED TRANSACTIONS
76	9.1	COMMONWEALTH PUBLIC SAFETY

(Last page.)

## B Code of Virginia sections that refers to itself

Analysis based on the Code of Virginia CSV files identified web pages that linked back to themselves. The pages (sections), along with the number of self-referential links, are listed below.

1. §10.1-1197.6 – 1	27. §38.2-233 – 1	53. §55-225.01 – 1
2. §13.1-1049.2 – 1	28. §38.2-236 – 1	54. §55-361.1 – 1
3. §13.1-774 – 1	29. §38.2-3229 – 1	55. §55-66.3 – 3
4. §15.2-1310 – 1	30. §38.2-3300 – 1	56. §55-79.29 – 1
5. §15.2-1644 – 1	31. §4.1-348 – 1	57. §55.1-129 – 1
6. §15.2-209 – 1	32. §46.2-1101 – 1	58. §55.1-2201 – 1
7. §15.2-2141 – 1	33. §46.2-1176 – 1	59. §55.1-2500 – 1
8. §15.2-2258 – 1	34. §46.2-1569 – 1	60. §55.1-339 – 3
9. §15.2-2296 – 2	35. §46.2-1569.1 – 1	61. §55.1-609 – 4
10. §15.2-2309 – 1	36. §46.2-2001.1 – 2	62. §56-235.8 – 2
11. §15.2-2403 – 1	37. §46.2-391.2 – 2	63. §57-39.3 – 1
12. §15.2-7500 – 1	38. §46.2-626.1 – 1	64. §57-39.4 – 1
13. §18.2-384 – 1	39. §46.2-706 – 1	65. §57-39.6 – 1
14. §18.2-421 – 1	40. §50-73.136 – 1	66. §57-39.7 – 1
15. §2.2-2340 – 1	41. §50-73.137:3 – 1	67. §58.1-1021.04:2 – 1
16. §2.2-3705.1 – 1	42. §50-73.52:2 – 1	68. §58.1-3965 – 1
17. §20-111.1 – 1	43. §51.1-169 – 1	69. §58.1-3980 – 1
18. §20-159 – 1	44. §51.1-217 – 2	70. §59.1-406 – 1
19. §21-118.8 – 1	45. §51.1-513.2 – 1	71. §62.1-140 – 4
20. §29.1-300.2 – 1	46. §51.5-68 – 1	72. §62.1-44.34:19 – 1
21. §30-112 – 1	47. §51.5-72 – 2	73. §62.1-44.43 – 1
22. §32.1-291.1 – 1	48. §53.1-84 – 1	74. §64.2-1003 – 1
23. §36-19.5 – 1	49. §54.1-306 – 1	75. §64.2-106 – 1
24. §36-96.18 – 1	50. §54.1-402 – 1	76. §66-13 – 1
25. §38.2-1301 – 1	51. §55-109.2 – 4	77. §8.01-407.1 – 1
26. §38.2-1415 – 1	52. §55-12.6 – 1	78. §8.2A-209 – 1

Use these steps to validate the self references:

1. Identify a section of interest.
2. Create a URL that looks like this:  
`https://law.lis.virginia.gov/vacode/ttt/sectionSectionIdentifier/`
3. Load the section into a browser.
4. Search within that page for the SectionIdentifier. (Usually you will find it in the plain text of the page, but sometimes you will have to look at the page's source code.)

For example, find the self references for §56-235.8.

1. Section of interest is: 56-235.8
2. URL is: `https://law.lis.virginia.gov/vacode/ttt/section56-235.8/`
3. Load the page.
4. Search for: 56-235.8  
Find four strings that match the section number.
  - (a) One in the “navigation” bar at the top of the page.
  - (b) One in the title of the page.
  - (c) Two in paragraph I, section 3.

## C Centrality

We let the following quote be our guidelines for computing and talking about graph based centrality metrics:

*“The intuition about centrality is that it denotes an order of importance on the vertices or edges of a graph by assigning real values to them. . . . as a minimal requirement we demand that the result of a centrality index is only depending on the structure of the graph.”*

Brandes and Erlebach [2]

Our vertex and edge centrality values are dependent on finding the shortest path between any two arbitrary vertices ( $u$  and  $v$ ) in the graph.

- A graph is composed of vertices ( $V$ ) and edges ( $E$ ).
- A path is a series of edges between vertices in the graph connects the two vertices.
- The length of the path is the distance between the two vertices, and is designated:

$$d(u, v)$$

- The shortest path is the path with the fewest edges between the two vertices. There may be multiple paths of the same length. Only one path will be designated as the shortest, and is designated:

$$\delta(u, v)$$

- The shortest path between vertices  $s$  and  $t$  that uses vertex  $v$  is designated:

$$\delta_{st}(v)$$

- The shortest path between vertices  $s$  and  $t$  that uses edge  $e$  is designated:

$$\delta_{st}(e)$$

- The centrality of vertex  $v$  is:

$$c(v) = \sum_{s \neq v \in V} \sum_{t \neq v \in V} \delta_{st}(v)$$







- The centrality of edge  $e$  is:

$$c(e) = \sum_{s \in V} \sum_{t \in V} \delta_{st}(e)$$

Computing either vertex or edge centrality requires computing all possible paths between all nodes. Depending on implementation, execution time can range from:  $\mathcal{O}(nm + n^2 \log n)$  to  $\mathcal{O}(n^3)$  operations[2]; where  $n = |V|$  and  $m = |E|$ . In our case that means between 14,377,599,300 and 3,749,110,000,000 operations.

## D Miscellaneous files

A collection of miscellaneous files mentioned in the report.

- rawCSV.zip – Raw CSV files downloaded from <https://law.lis.virginia.gov/law-library/> on 2 March 2020. 
- cleanedCSV.zip – CSV files “cleansed” and validated. 
- processData.R – R script file to read data files, analyze the data, and create images. 
- validationResults.txt – Results of validation checks on “raw” CSV files. 
- wordCloud.R – R script to print wordclouds. 
- savedData.rdat – Wordcloud data file. 



## E References

- [1] Tim Berners-Lee, Roy Fielding, and Larry Masinter, *Uniform resource identifier (uri): Generic syntax (rfc 3986)*, Network Working Group (2005).
- [2] Ulrik Brandes and Thomas Erlebach, *Network Analysis: Methodological Foundations*, Springer Verlag, 2005.
- [3] Jon M Kleinberg, *Hubs, authorities, and communities*, ACM computing surveys (CSUR) **31** (1999), no. 4es, 5–es.
- [4] Alexander Lyte, David Slater, and Shaun Michel, *Network measures of the united states code*, Tech. report, MITRE: McLean, Va, 2015.
- [5] Virginia’s Legislative Information System Staff, *Virginia law*, <https://law.lis.virginia.gov/>, 2020.