BOSTON UNIVERSITY GRADUATE SCHOOL
COLLEGE OF ART AND SCIENCE & COLLEGE OF ENGINEERING

# CS506 Deliverable 2 & 3
# WGBH Reliability of Informant Cases

**Changhao Liang U16843909**
**Xiang Liu U46406505**
**Qitong Wang U44926429**
**Chen Xin U72406425**

Department of Computer Science, Boston University
Department of Systems Engineering, Boston University

**April 21, 2020**

# Contents

# List of Figures

# 1 Introduction

The WGBH Informant Project is an effort to create a comprehensive analysis of criminal cases which have been overturned due to the use of an informant in New Hampshire and Rhode Island over the past decade (2008-2018). The project aims to

- Uncover never before-seen patterns of judicial behavior and decision-making in criminal cases
- Recast and spur more aggressive news coverage of the court system and the use of informants
- Create a valuable understanding of the effectiveness of informant testimony
- Find relationship between overturned cases and presence of informant

In a criminal case, only the defendant has a right to an appeal in most states. Lower court cases can be challenged for any erroneous rulings by the judge on the evidence or on the law, the prosecutor's conduct during the case, or even the trial attorney's handling of the defense.

# 2 Data Collection

Criminal cases from lower courts can be reversed in full or reversed in parts. We want to collect both full and partial reversed decisions. To do this, we need to scrape for all cases that were sent to appellate courts of New Hampshire and Rhode Island.

## 2.1 Data Source

Our data source are the appellate cases posted on the supreme court websites of New Hampshire and Rhode Island. Each case is stored in pdf format.

- Cases from New Hampshire
  https://www.courts.state.nh.us/supreme/opinions/
- Cases from Rhode Island
  https://www.courts.ri.gov/Courts/SupremeCourt/Pages/Opinions%20and%20Orders%20Issued%20in%20Supreme%20Court%20Cases.aspx

Based on our research, we found that 9 of the 50 states in U.S. do not have an intermediate appellate court, and New Hampshire and Rhode Island are among these states. (https://en.wikipedia.org/wiki/List_of_state_intermediate_appellate_courts) Their appeals from lower courts go directly to their supreme courts.

## 2.2 Data Downloading

**Selenium Webdriver** is a collection of open source APIs which are used to automate the testing of a web application. To collect all case records from 2008 to 2018 in New Hampshire and Rhode Island, we used Selenium Webdrive to automate the download process of all the case files from their supreme court websites.

### 2.2.1 New Hampshire

The website of New Hampshire cases has a link for each year's cases from 2008 to 2018. Clicking on one of the links directs us to the web page that lists all the downloading links of case records in the respective year. Inside each of these web pages, clicking on one of the downloading links downloads the clicked case as a pdf file.

To automate the downloading process, we created and configured a Selenium Webdriver using a Python script. It first opened up the website of New Hampshire cases with a Chrome browser. Then we looked at the website's html and found the XPath of all links of years from 2008 to 2018. The web driver located all the links using the XPath and then clicked each of the links from 2008 to 2018 in a loop.

| Year | 2008 | 2009 | 2010 | 2011 | 2012 | 2013 | 2014 | 2015 | 2016 | 2017 | 2018 | total |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| # of cases | 140 | 146 | 150 | 146 | 136 | 94 | 113 | 100 | 92 | 79 | 86 | 1282 |

Table 1: Number of files downloaded including both civil and criminal cases from New Hampshire supreme court

| Year Period | 2008 ∼ 2009 | 2009 ∼ 2010 | 2010 ∼ 2011 | 2011 ∼ 2012 | 2012 ∼ 2013 | 2013 ∼ 2014 | 2014 ∼ 2015 | 2015 ∼ 2016 | 2016 ∼ 2017 | 2017 ∼ 2018 | total |
|---|---|---|---|---|---|---|---|---|---|---|---|
| # of criminal cases | 31 | 31 | 37 | 7 | 46 | 40 | 21 | 10 | 11 | 7 | 241 |

Table 2: Number of files downloaded including only criminal cases from Rhode Island supreme court

Each time after clicking on a specific year, the web driver located all the downloading links based on the XPath we found for case titles and downloading links in the website's html, and then it clicked through all the downloading links to download all the pdf case files in the clicked year. The pdf files were downloaded into a directory configured when creating the web driver. The pdfs of different years were automatically saved inside different folders inside the download directory.

The files contained both civil cases and criminal cases, and we only needed to look at criminal cases. We filtered all cases to extract only criminal cases using the text information inside the pdf files in the following data pre-processing step for cases in New Hampshire. Table 1

Downloaded pdf files: *data/pdf/pdf_nh*
Python script: *src/scraper/download_nh_cases.py*

#### 2.2.2 Rhode Island

The website of Rhode Island cases also has a link for each year's cases from 2008 to 2018, which can be located by XPath from the website's html. Clicking on one of the links directs us to the web page that lists all the cases in the clicked year. For each case, it shows the case title, the case head notes and the downloading link for the case text as a pdf file.

The website has two lists of links, one lists the court opinions of each year and one lists the court orders of each year. To automate the downloading process, we again created and configured a Selenium Webdriver to open up the website. Then we extracted the XPath of the court opinions' year list, and the web driver used the XPath to locate all the years of case opinions and clicked each year from 2008 to 2018.

Each time after clicking on a specific year, we extracted the XPath containing the case title and the case's pdf downloading link from html and then let the web driver click the downloading links to download the pdfs using the XPaths. The XPath structures were different for 2008 ∼ 2011, 2011 ∼ 2017 and 2017 ∼ 2018, and we modified our code into 3 sections to adapt and download using different XPath structures. The pdfs of different year periods were downloaded into the separated folders.

Since text information of cases in Rhode Island does not contain clear indicator of whether a case is criminal or civil, we only downloaded cases that had case titles starting with "State v", which fit in the profile of a criminal case because all criminal cases had state as the prosecutor. And we skipped cases where the defendant was an organization instead of an individual to filter out lawsuits between state and non-human entities. In this way, we only downloaded criminal cases from all cases on the website for 2008 to 2018. Table 2

Python script: *src/scraper/download_ri_criminal_cases.ipynb*
Downloaded pdf files: *data/pdf/pdf_ri_crminal*

| pdf file name | case title | case type (criminal or civil) | case decision | case text |
|---|---|---|---|---|
| string | string | criminal<br><br>non-criminal | affirmed<br>affirmed in part<br>not affirmed | list of strings |

Table 3: contents and format of a tuple

### 2.3 Data Pre-processing

Our goal of data pre-processing is to remove invalid data entries and then arrange data of each case into a tuple containing pdf file name, case title, case type, case decision and case text. Table 3

The list of tuple could be stored inside a csv file or a json file for further data analysis.

For both New Hampshire and Rhode Island, the pdf file name and case title were already recorded when downloading the cases. The case text of each case was extracted from the pdf file using **pdftotext** library.

#### 2.3.1 New Hampshire

We extracted the case type of New Hampshire case by looking for the RSA number in the third paragraph of the case text from pdf. By the legal document RSA of RI, if the labeled RSA number of a case is between 625 and 652, then the case is criminal and we label the case as "criminal". If RSA number was out of that range or we do not find a RSA, we label the case as "non-criminal".

The case decision for New Hampshire was written at the end of each pdf file. We wrote a script to split each text into paragraphs which helped us with locating and extracting the case decision. Our case decisions were recorded as "affirmed", "affirmed in part" or "not affirmed" based on our extracted information.

Python script: *src/scraper/get_nh_cases.py*
Processed data files: *data/*

#### 2.3.2 Rhode Island

The case type of Rhode Island cases were all "criminal", because we already only collected cases that were criminal when we downloaded the pdf files of the cases.

The case decision for New Hampshire was written in a section called conclusion in each pdf file. We again wrote a script to split each text into paragraphs to find a paragraph called "conclusion" and the case decision followed right after it. Same as New Hampshire, our case decisions were recorded as "affirmed", "affirmed in part" or "not affirmed" based on our extracted information.

Python script: *src/scraper/get_ri_cases.ipynb*
Processed data files: *data/*

## 3 Data Analysis

Our data representation is

| | file name | title | type | decision | text |
|---|---|---|---|---|---|
| 0 | 06-290.pdf | State v. Michael Tetreault, No. 06-290 (June 1... | criminal | affirmed | ['', 'Supreme Court', 'No. 2006-290-C.A.', '(P... |
| 1 | 08-27.pdf | State v. Thomas P. Byrne, No. 08-27 (June 19, ... | criminal | not affirmed | ['', 'Supreme Court', 'No. 2008-27-C.A.', '(P2... |
| 2 | 07-108.pdf | State v. Robert Collazo, No. 07-108 (April 3, ... | criminal | affirmed | ['', 'Supreme Court', 'No. 2007-108-C.A.', '(P... |
| 3 | 07-334.pdf | State v. Samuel Adewumi, No. 07-334 (March 17,... | criminal | affirmed | ['', 'Supreme Court', 'No. 2007-334-C.A.', '(W... |
| 4 | 07-123.pdf | State v. Phillip Jackson, No. 07-123 (March 20... | criminal | not affirmed | ['', 'Supreme Court', 'No. 2007-123-C.A.', '(P... |

| State | New Hampshire | | | Rhode Island | | |
|---|---|---|---|---|---|---|
| Decision | affirmed | partly affirmed | reversed | affirmed | partly affirmed | reversed |
| C.I. cases | 11 | 1 | 3 | 10 | 0 | 3 |
| non-C.I. cases | 183 | 8 | 63 | 196 | 5 | 26 |

Table 4: The number of all kinds of cases. In this table, "confidential informant" is abbreviated as "C.I.".

### 3.1 Data Visualization

To find the relevance between confidential informant cases and the final decision of cases, we need to find the distribution of decision in confidential informant cases and cases which are not confidential informant cases.

There are three kinds of final decision: "affirmed", "partly affirmed" and "reversed". By extracting keywords that we need, the total number of all kinds of cases are shown in Table 4.

Moreover, we draw some pie graphs to show the distribution of cases. Details are shown in Figure 1 and Figure 2.

According to these pic pictures, we find:

• In New Hampshire State, the final decision of cases is not relevant to the confidential informant in cases.

• However, in Rhode Island State, confidential informant cases are more likely to be reversed than that of non-confidential informant cases. So when there is more effective evidences from confidential informant, case appeal is less likely to succeed.

As is stated above, the circumstances of the case may differ greatly in different states, which may be caused by different laws in different states; or the confidential informant in different states provides different information for the case; perhaps the information provided by confidential informant in RI state is more useful than confidential informant in NH state. This is what we think causing the above phenomenon (see Figure 1 and Figure 2).

### 3.2 Data Cleaning

Use Tf-idf to delete common words that appear many times in every case. For example, in the New Hampshire's data, 'New', 'Hampshire', 'court', 'cases' will appear many times. But these words are not important. To solve this, we input all cases into Tf-idf model to get the most frequent 20 words. Then, we delete this words from our cases' text.

Remove the number, punctuation and stop words. Stop words refer to meaningless words, like 'I','a','be'. These text are not important to use, so we delete them.

### 3.3 Tf-idf

We use Tf-idf to get the most frequent features in criminal cases, affirmed cases, reversed cases and partial affirmed cases. We found in NH's data, reversed cases have two frequent features: sexual and child. These two features are not appearing in affirmed frequent features. Figure 3. We found in RI's data, there is no clear pattern. Figure 4.

We also use Tf-idf to get the most frequent features in criminal cases with informants and without informants. There is no clear pattern in New Hampshire data with informants and without. Figure 5. For Rhode Island, we found 'murder' feature in criminal cases with informants. But 'murder' does not appeal in non-informant frequent features. We may get conclusion that cases with informants are more likely related to 'murder'. Figure 6.

For reversed cases, we select the most frequent 10 words in cases with informant and without. In New Hampshire, we found 'sexual' feature in reversed cases without informants. Figure 7. We may get conclusion that the reversed cases related to sex don't have informants. In Rhode Island, we found 'child' in reversed cases with informants. Figure 8. We may get conclusion that the reversed cases related to child may have informants.

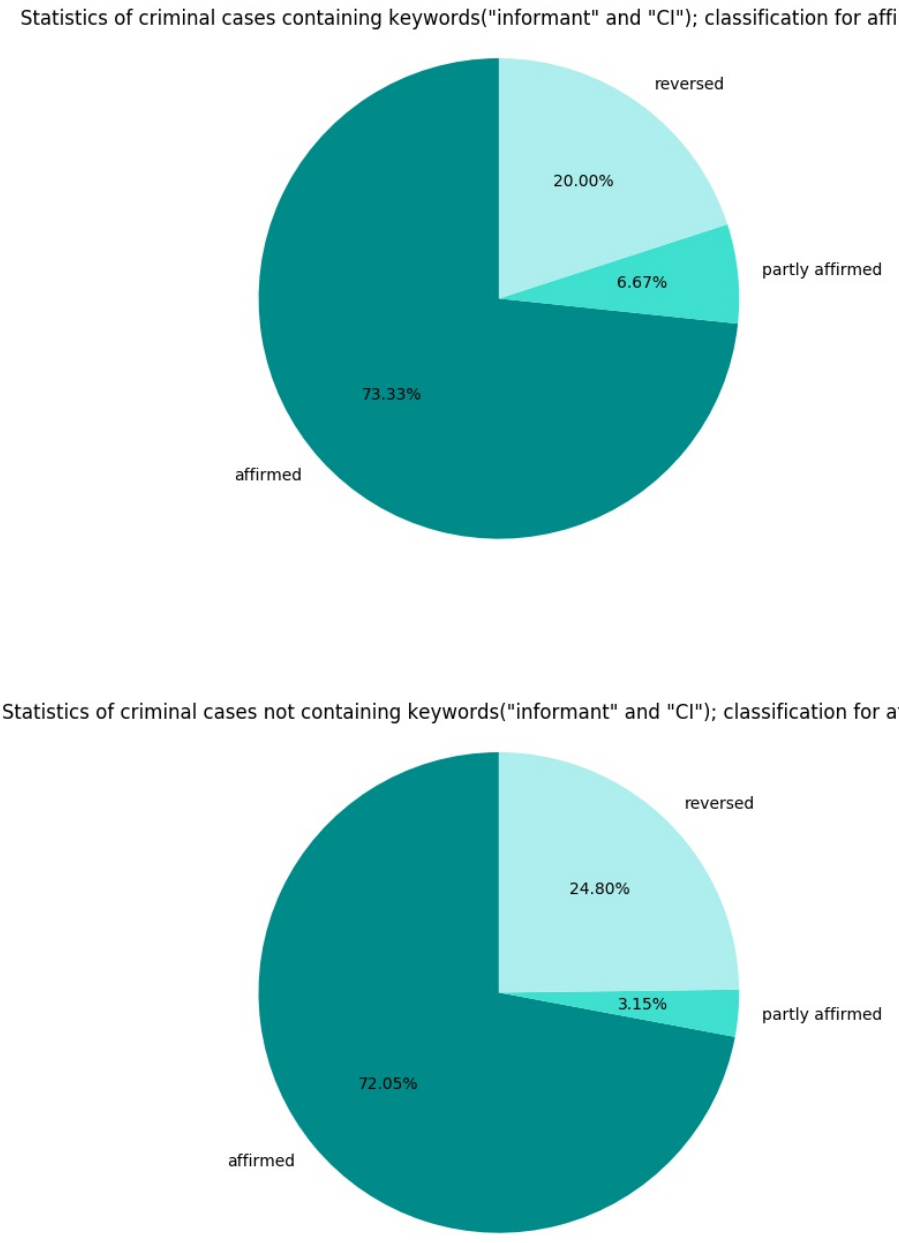


Figure 1: Distribution of cases in New Hampshire State

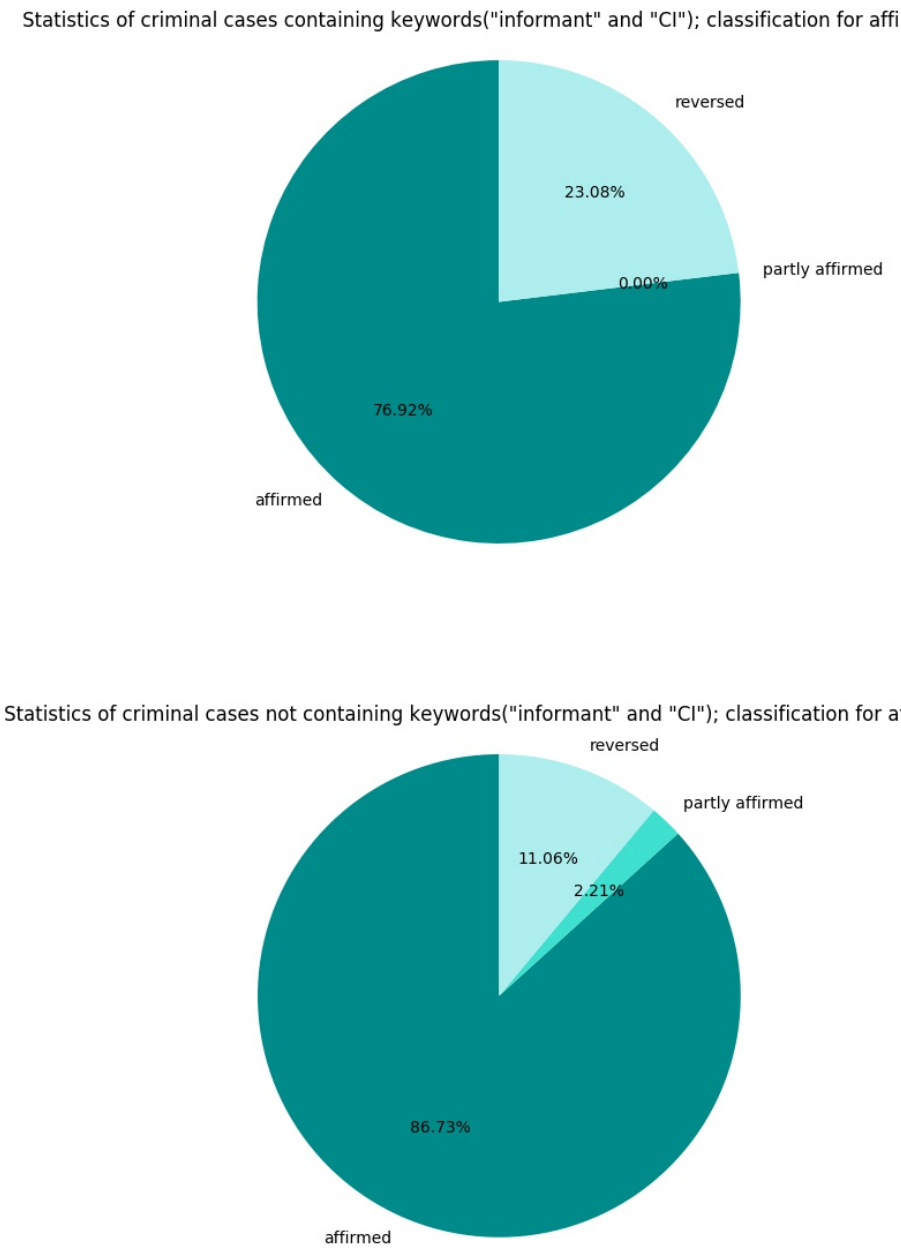


Figure 2: Distribution of cases in Rhode Island State

```
criminal_word
['assault',
 'conduct',
 'crime',
 'criminal',
 'defense',
 'error',
 'law',
 'motion',
 'opinion',
 'person',
 'quotation',
 'reasonable',
 'sexual',
 'testified',
 'time']

affirmed_word
['assault',
 'conduct',
 'crime',
 'defense',
 'error',
 'following',
 'law',
 'motion',
 'opinion',
 'person',
 'quotation',
 'reasonable',
 'testified',
 'time',
 'witness']

reversed_word
['assault',
 'child',
 'conduct',
 'criminal',
 'defense',
 'issue',
 'law',
 'motion',
 'officers',
 'opinion',
 'person',
 'quotation',
 'reasonable',
 'sexual',
 'testified']

part_word
['based',
 'charge',
 'conclude',
 'criminal',
 'deadly',
 'doubt',
 'error',
 'fact',
 'harmless',
 'juror',
 'law',
 'nixon',
 'reasonable',
 'statements',
 'weapon']
```

Figure 3: Affirmed and Reversed Cases in New Hampshire

```
criminal_word
['criminal',
 'defense',
 'new',
 'officer',
 'quoting',
 'stated',
 'superior',
 'supreme',
 'time',
 'witness']

affirmed_word
['criminal',
 'defense',
 'new',
 'quoting',
 'stated',
 'statement',
 'superior',
 'supreme',
 'time',
 'witness']

reversed_word
['criminal',
 'defense',
 'fact',
 'law',
 'officer',
 'review',
 'right',
 'superior',
 'time',
 'witness']

part_word
['apartment',
 'cell',
 'conditions',
 'phone',
 'probation',
 'sentence',
 'stated',
 'text',
 'time',
 'violation']
```

Figure 4: Affirmed and Reversed Cases in Rhode Island

```
informant_word
['constitution',
 'home',
 'informant',
 'information',
 'reasonable',
 'search',
 'statements',
 'told',
 'warrant',
 'witness']
```

```
non_informant_word
['assault',
 'conduct',
 'defense',
 'error',
 'law',
 'person',
 'quotation',
 'reasonable',
 'sexual',
 'time']
```

Figure 5: Informant in New Hampshire

```
informant_word
['ciresi',
 'criminal',
 'defendants',
 'detective',
 'information',
 'murder',
 'quoting',
 'review',
 'statement',
 'time']
```

```
non_informant_word
['criminal',
 'defense',
 'new',
 'officer',
 'quoting',
 'stated',
 'superior',
 'supreme',
 'time',
 'witness']
```

Figure 6: Informant in Rhode Island

```
info_r_word
['circumstances',
 'confidentiality',
 'detective',
 'home',
 'officer',
 'officers',
 'promise',
 'requirement',
 'search',
 'warrant']
```

```
non_info_r_word
['assault',
 'conduct',
 'criminal',
 'defense',
 'issue',
 'motion',
 'person',
 'quotation',
 'reasonable',
 'sexual']
```

Figure 7: Informant in Reversed Cases For New Hampshire

```
info_r_word
['baccaire',
 'child',
 'defense',
 'det',
 'error',
 'information',
 'laforest',
 'phillip',
 'right',
 'said']
```

```
non_info_r_word
['criminal',
 'defense',
 'fact',
 'issue',
 'law',
 'officer',
 'review',
 'superior',
 'time',
 'witness']
```

Figure 8: Informant in Reversed Cases For Rhode Island

### 3.4 Word2Vec

In this section, we want to find the similarity between the 'reversed' and other words in reversed cases. First, we put all reversed cases together, that is, put all words appeared in reversed cases into a list. Then we use Word2vec to find the similarity. We found that the word with the highest similarity to 'reversed' is 'remanded' in New Hampshire. Figure 9. We found that the word with the highest similarity to 'reversed' is 'remand' in Rhode Island. Figure 10. We can not get any useful conclusion from the 10 most similar words.

```
word2vec_model.wv.most_similar('reversed')
```

```
[('remanded', 0.9986464977264404),
 ('dissented', 0.9959361553192139),
 ('superior', 0.9943211674690247),
 ('broderick', 0.9942866563796997),
 ('conflict', 0.9933634996414185),
 ('appointed', 0.993209958076477),
 ('johnson', 0.9921435713768005),
 ('dugas', 0.9921237826347351),
 ('green', 0.9920369386672974),
 ('respondent', 0.9920048117637634)]
```

Figure 9: the Similarity Between 'reversed' and Other Words in Reversed Cases for NH

```
word2vec_model.wv.most_similar('reversed')
```

```
[('remand', 0.9991968274116516),
 ('district', 0.9991153478622437),
 ('instant', 0.99909508228302),
 ('specifically', 0.999093770980835),
 ('new', 0.9990875720977783),
 ('pepper', 0.9990874528884888),
 ('conclusion', 0.9990872144699097),
 ('shower', 0.9990850687026978),
 ('including', 0.9990758895874023),
 ('close', 0.9990744590759277)]
```

Figure 10: the Similarity Between 'reversed' and Other Words in Reversed Cases for RI

### 3.5 N-gram

Because using word2vec method cannot offer useful results of words similarity, we try the N-gram method to analysis text data. An N-gram model is a type of probabilistic language model for predicting the next item in such a sequence in the form of a $(n-1)$ order Markov model.

Using N-gram, we can extract important language contexts which has high weights from 'text' data. For example, if some language sequence often appears in many cases, we can get those sequence and try to find the relationship between these sequences and other things such as the case decision or if the case has informant.

### 3.6 Visualization of Ngram results

To make the model better fits our data, we visualize the n-gram results into a graph by utilizing *networkx* package. to find the best parameters. We found that if we give a high weight factor for each sequence, the plot will be more obvious, there will be some Aggregation and sparse place in the graph, it can help us find the relationship among sequence more quickly.

In the graph, each node is a word, we add edges by n-gram results. All the nodes are sorted by weights, we only add top 100 nodes into graph, this can avoid analysing useless text information.

Besides, we give different color for graph nodes by node degree. If a node degree is high, it means that it might be more important because it connects with many other nodes. If a node degree is low, it means that it

might be less important or it has its own context and far away from other aggregation parts.

## 3.7 New Hampshire

We analysis New Hampshire relationship between cases text and criminal cases decision, relationship between cases text and cases decision if having informants.

### 3.7.1 Relationship between Case Text and Criminal Cases Decision

Figure 11 and Figure 12, we found that in reversed cases, there is more likely to be cases that related to some serious crime such as:

- sexual crime
- assault
- murder
- felony

We think the decision of these serious cases should make decision carefully, which means the probability of reversion of serious cases is higher than that of normal cases.

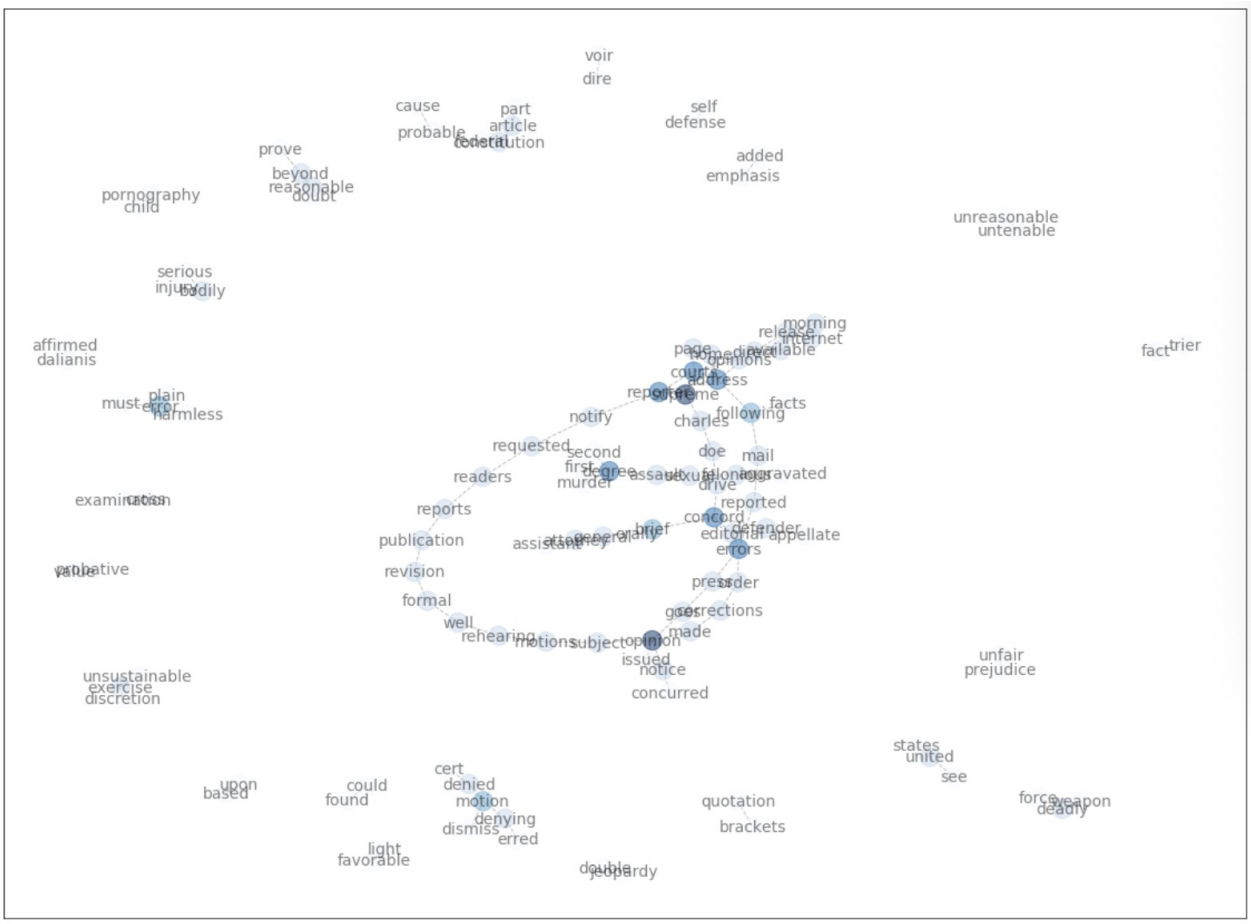


Figure 11: New Hampshire criminal cases which has affirm decision

### 3.7.2 Relationship between Case Text and Informant/Non-informant Cases

Figure 13 and Figure 14, we think that there is no obvious difference between the information networks with confidential informant and that without confidential informant.

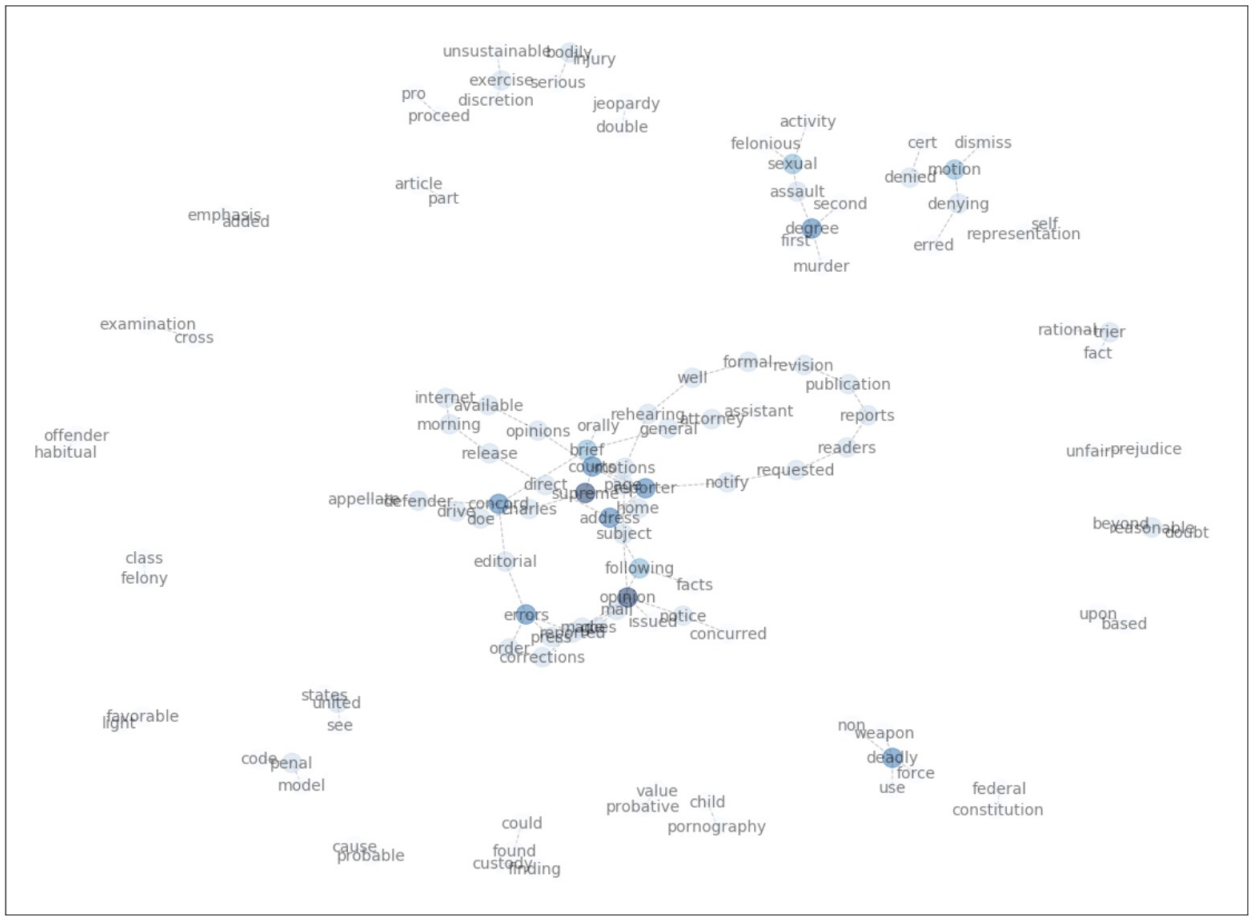


Figure 12: New Hampshire criminal cases which has reverse decision

We think it is caused by that the existence at confidential informant at NH state do not cause obvious effect the the final decision results of cases. (See 3.1 for more details.)

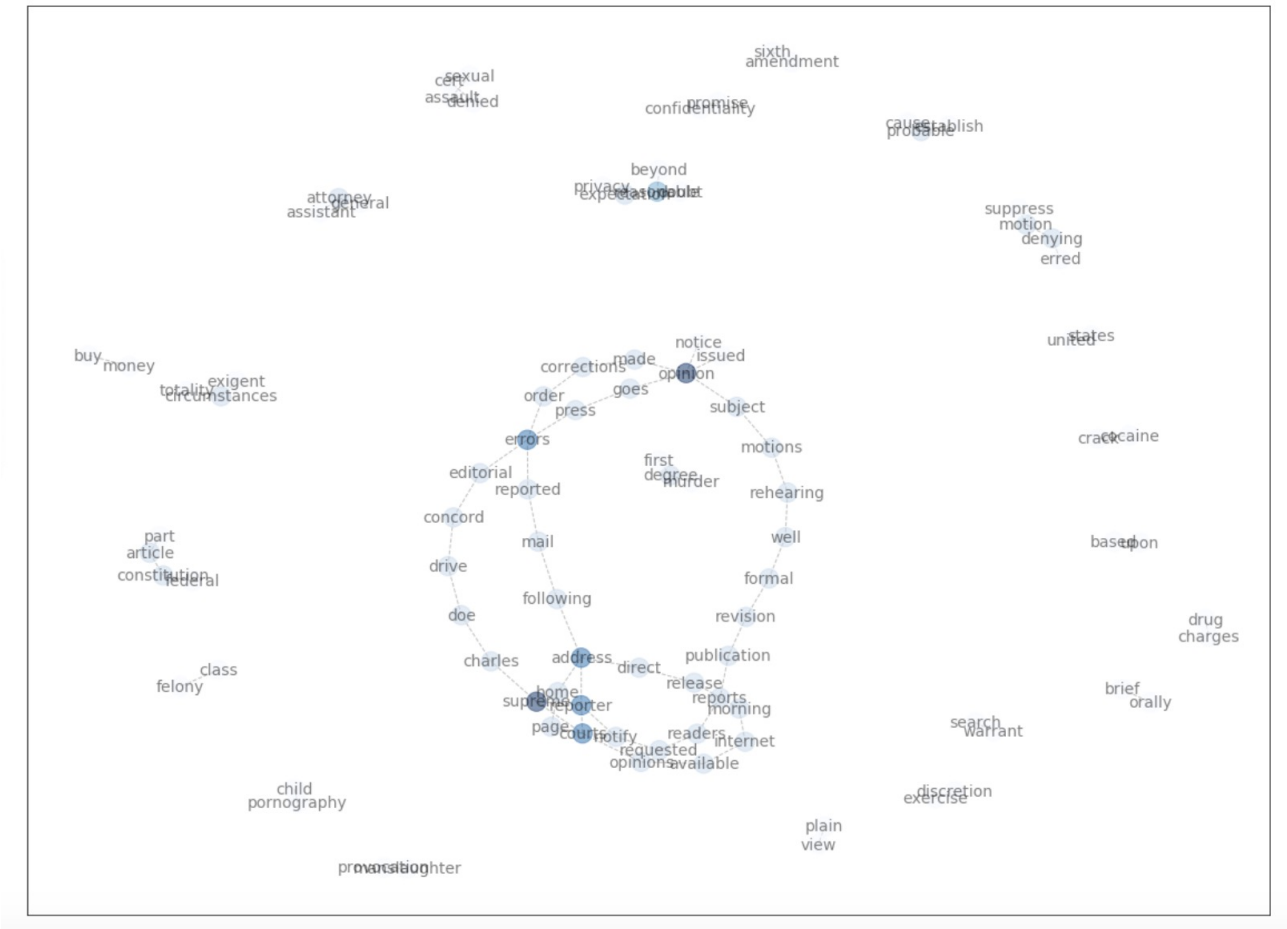


Figure 13: New Hampshire cases which has informant

### 3.7.3 Relationship between Affirmed Case Text and Informant Cases Decision

In the Figure 15, there are some informative sequence including

- bodily injury
- sexual assault
- self defense

In the Figure 16, there are some informative sequence including

- child pornography
- drug charges
- provocation manslaughter

In affirmed case with informants, we notice that some of the cases are related to drugs. However, some use of drugs are illegal. That means the cases with drug may need more informants. Because informants may have important meaning in such cases.

In this part, we get conclusion that in NH's data, the informants are more credible in drug-related cases.

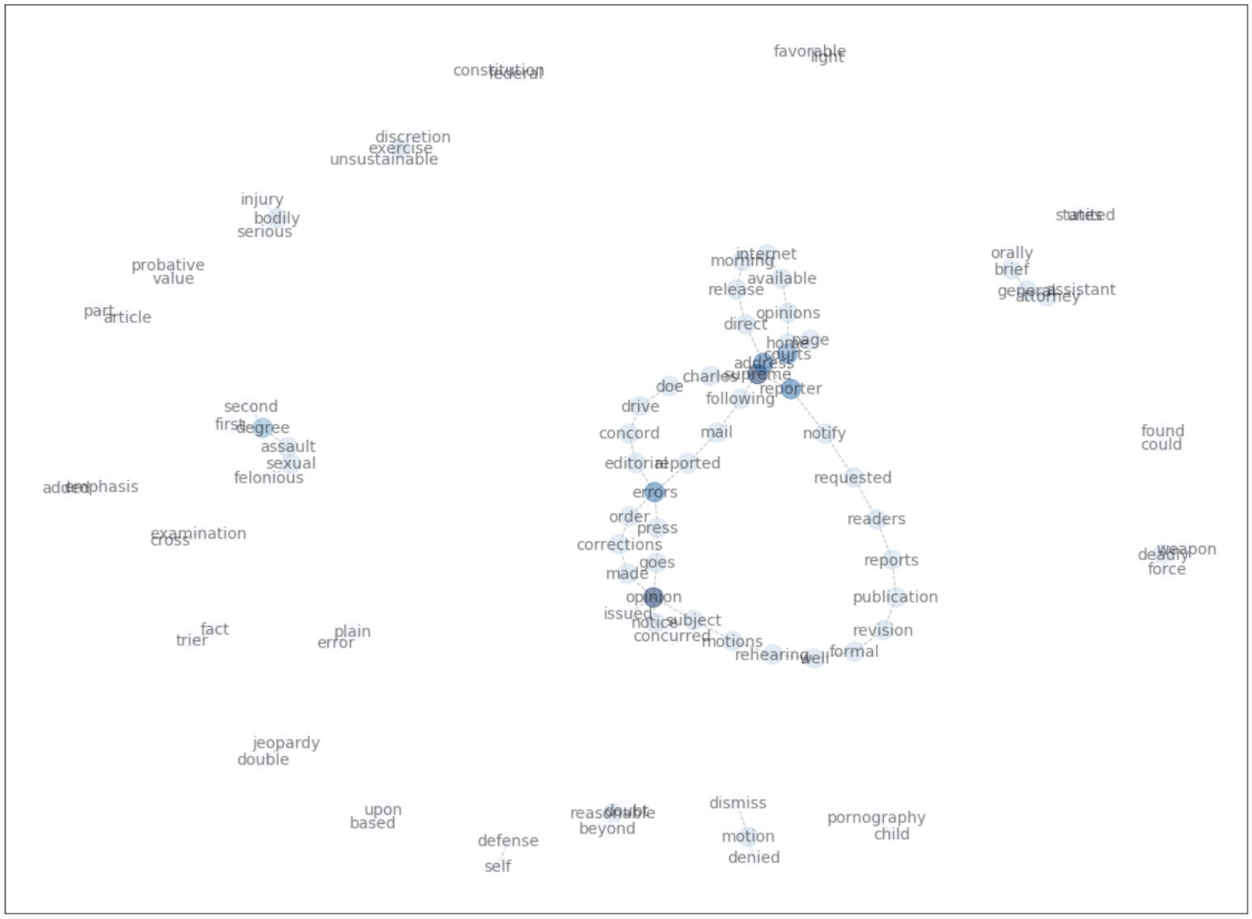


Figure 14: New Hampshire cases which has no informant

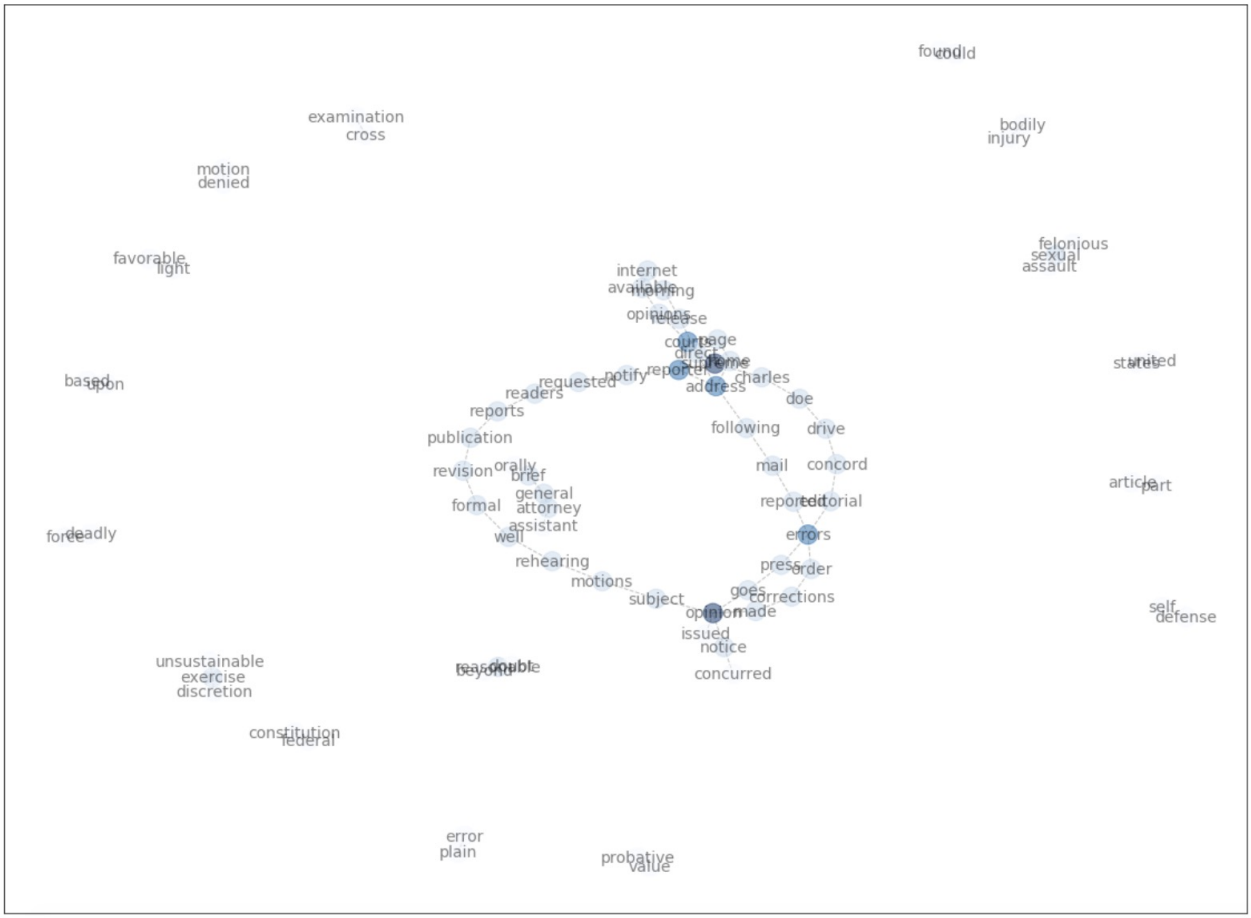


Figure 15: New Hampshire cases which do not have informant and decision is affirmed

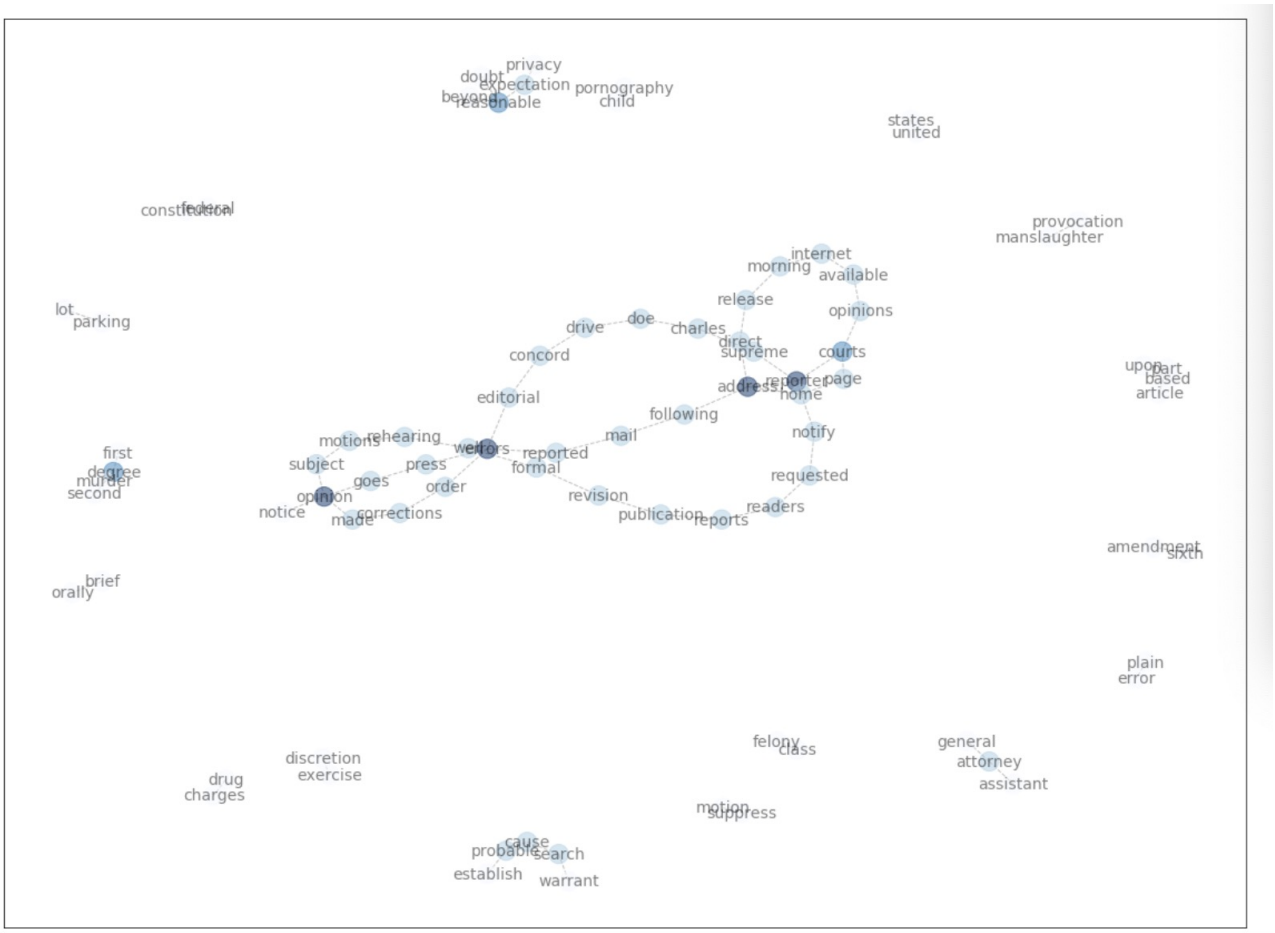


Figure 16: New Hampshire cases which has informant and decision is affirm

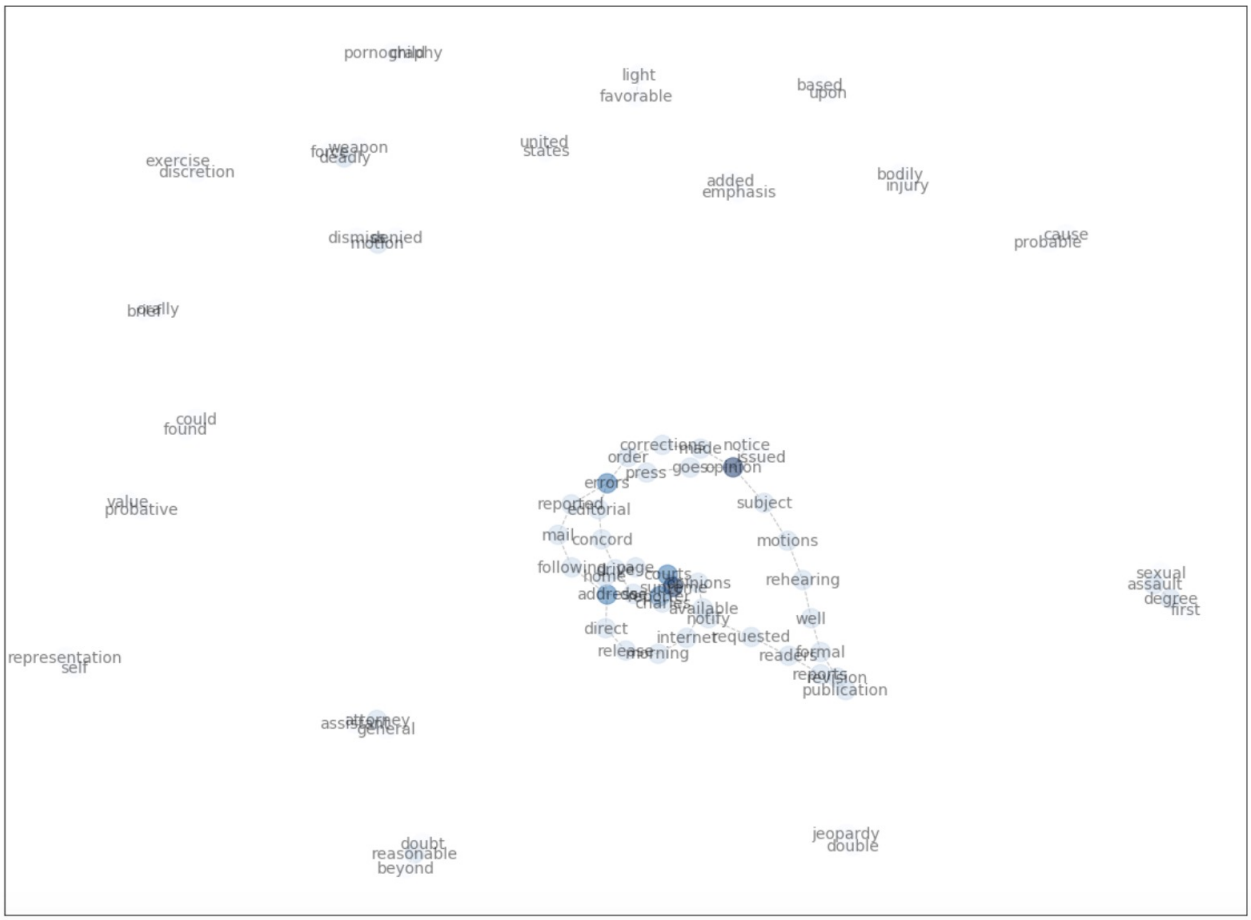


Figure 17: New Hampshire cases which do not have informant and decision is reversed

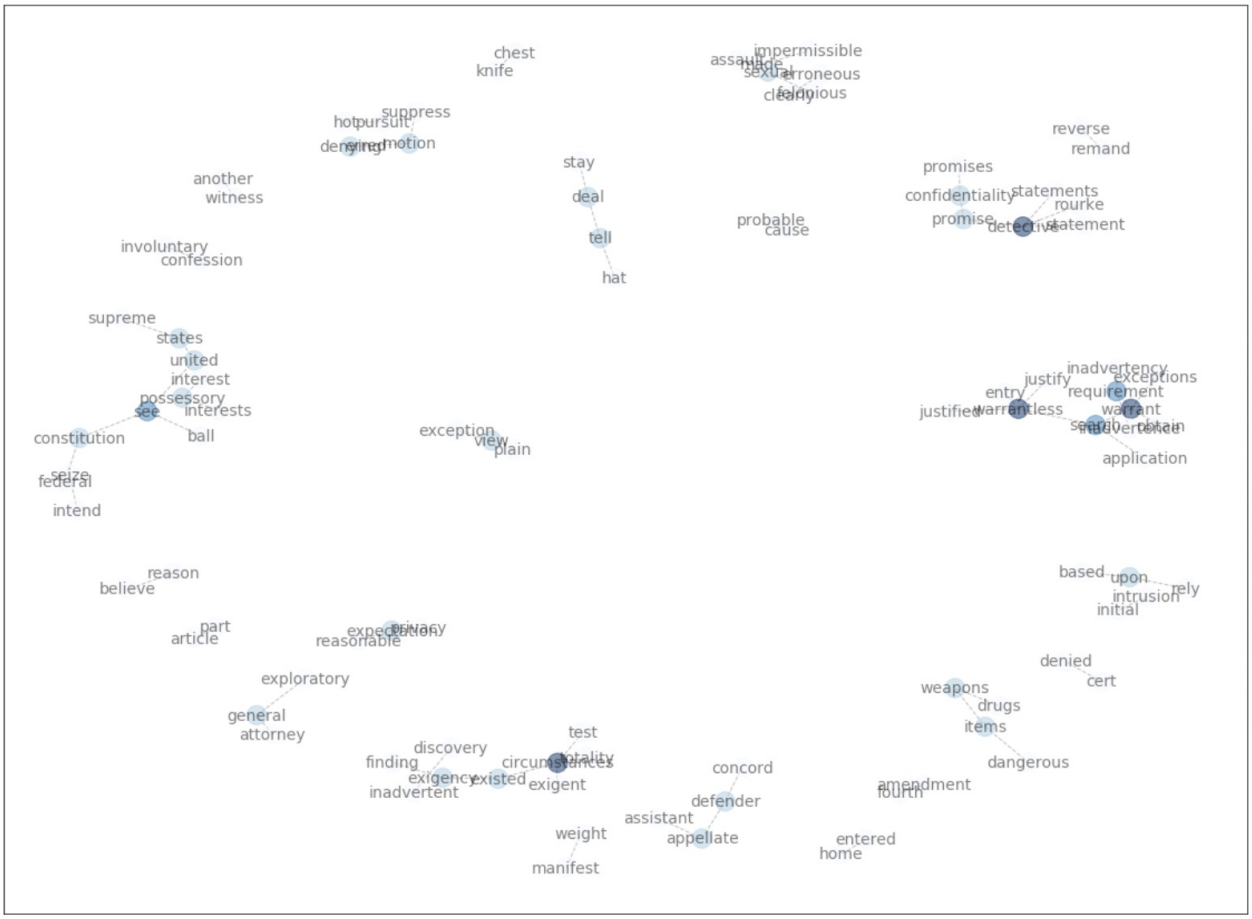


Figure 18: New Hampshire cases which has informant and decision is reversed

### 3.7.4 Relationship between Reversed Case Text and Informant Cases Decision

In the Figure 17, there are some informative sequence including

- assault sexual
- bodily injury
- child pornography
- weapon

In the Figure 18, there are some informative sequence including

- dangerous items weapon, drugs
- assault sexual

In reversed cases with informants, we found these cases are more likely to related to 'drug' and 'weapon'. According to subsection 3.7.3, the affirmed cases with informants are also related to 'drug'. We get the conclusion that the cases related to 'drug' are more likely to have informants. But we can not know whether the informants have positive effect on the decisions.

## 3.8 Rhode Island

We analysis Rhode Island relationship between cases text and criminal cases decision, relationship between cases text and cases decision if having informants.

### 3.8.1 Relationship between Case Text and Criminal Cases Decision

In the graph Figure 19, there are some informative sequence including

- murder, child, sexual, molestation
- witness credibility,
- dangerous weapon
- abuse

For affirmed criminal cases, it seems that words are not very serious, which might means the decision is fair and reasonable. Defendants do not appeal to court.

Witness credibility might guarantee the decision reasonable, this can be from confidential informant. We will compare results about informant in next part.

In the graph Figure 20, there are some informative sequence including

- sex, sexual offender
- felony
- corporal punishment
- self defense
- serious physically injury
- abuse child
- forth, sixth amendment
- exigent circumstances

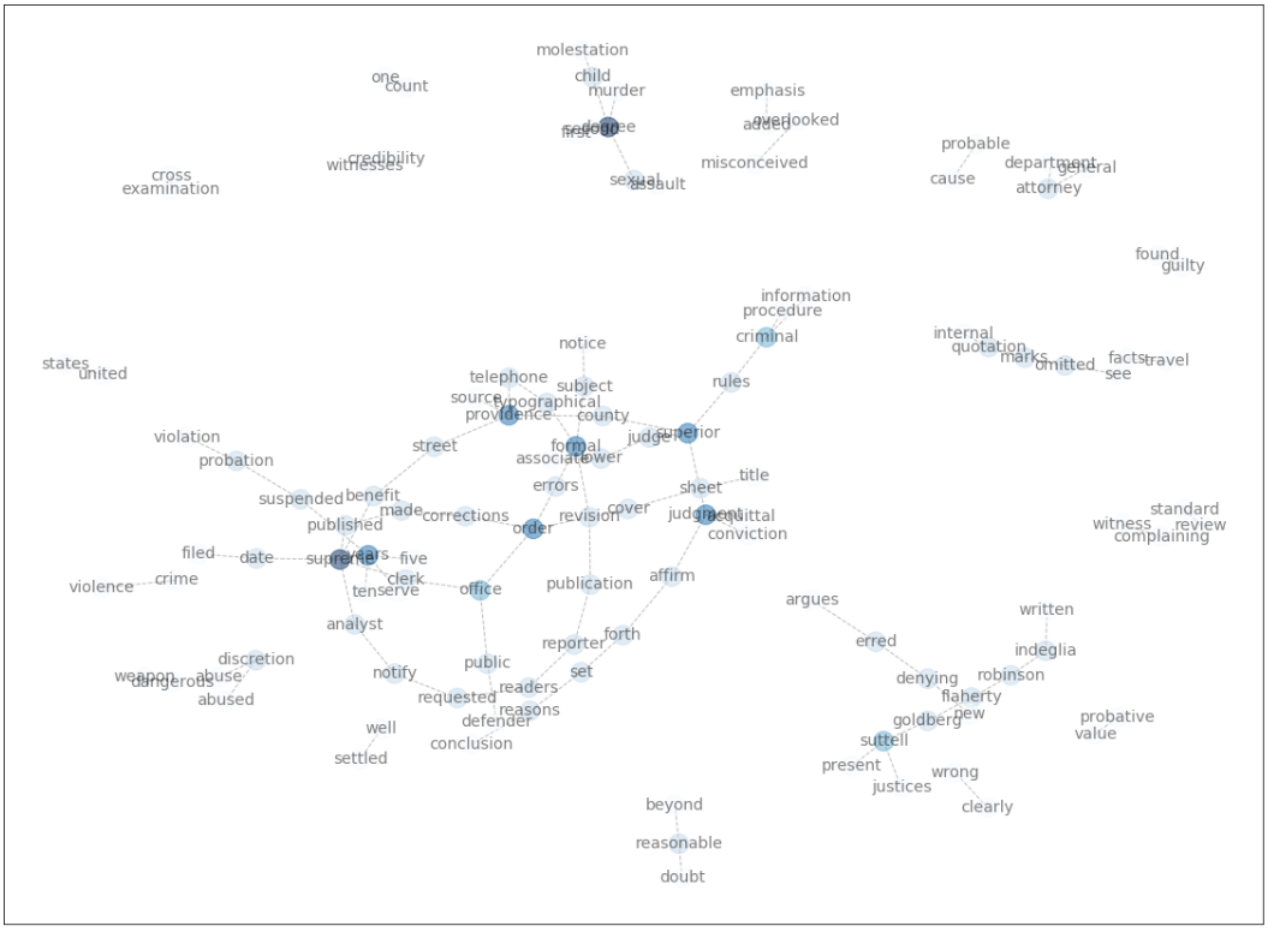


Figure 19: Rhode Island criminal cases which has informants and affirm decision

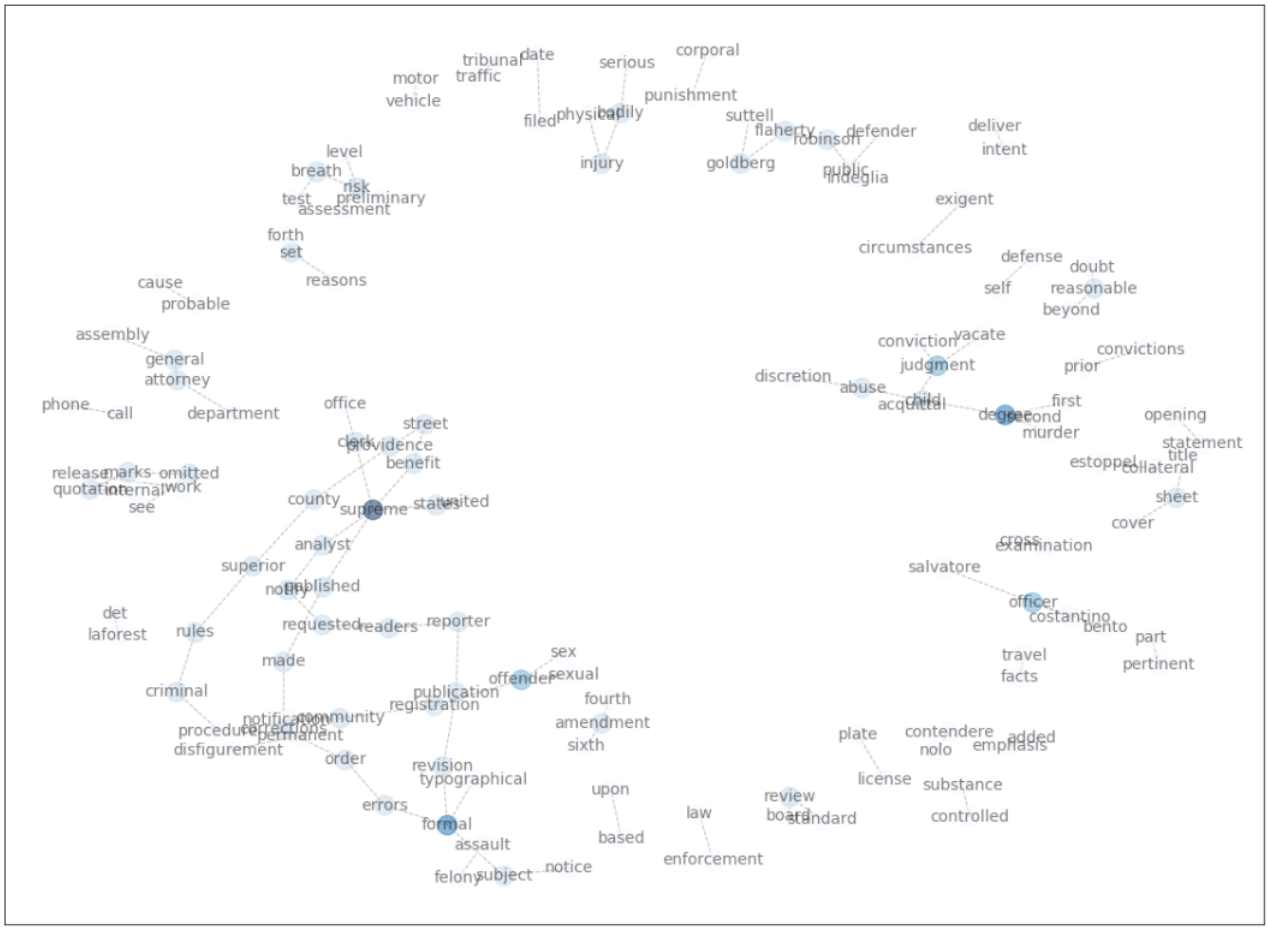


Figure 20: Rhode Island criminal cases which has reverse decision

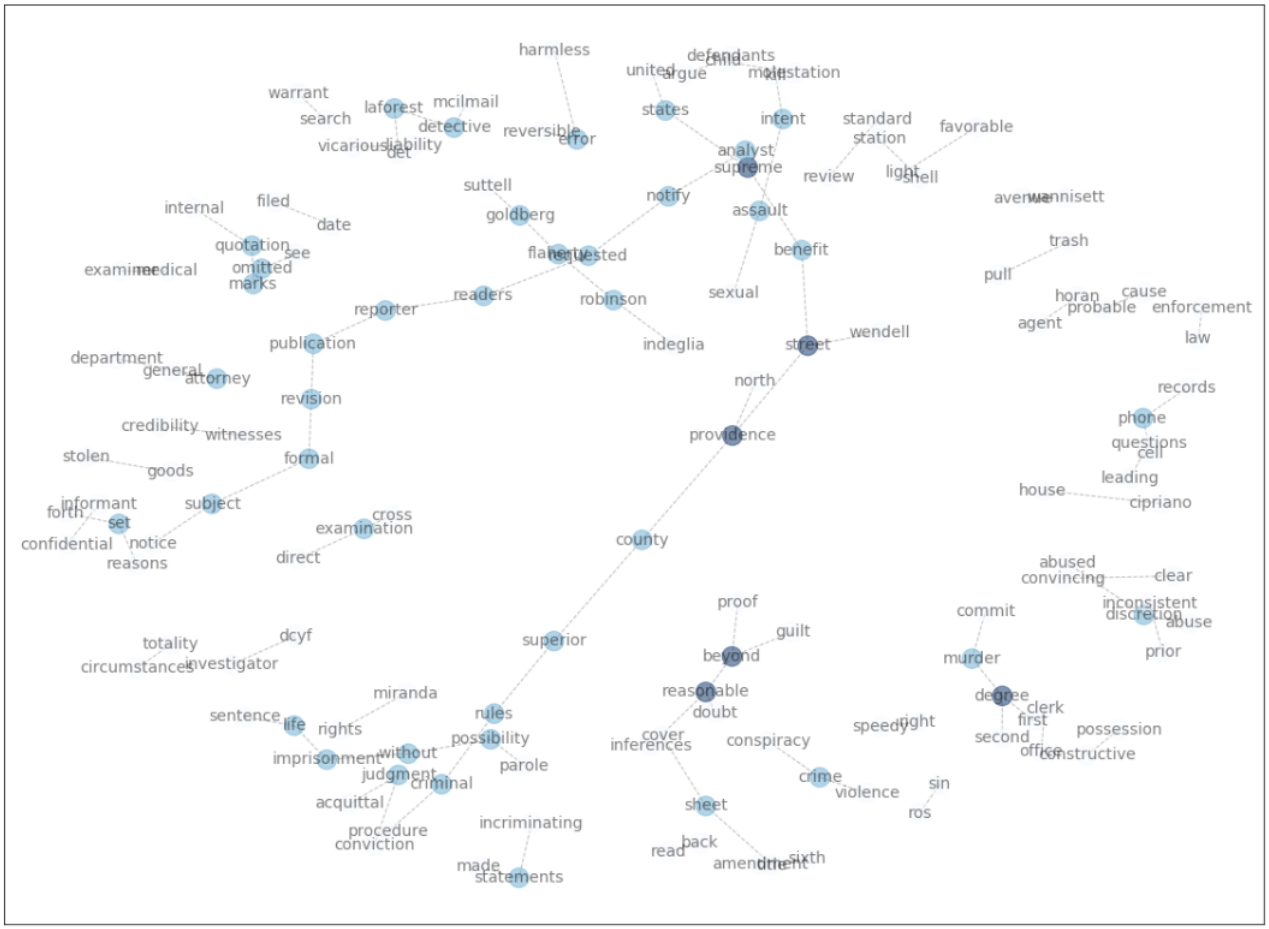


Figure 21: Rhode Island criminal cases which has informant

We found that for reversed criminal cases, the cases often contains serious criminal information such as abusing child, corporal punishment, sexual problem and so on. Self defense might be because some one want to kill him or something else, it must be very serious. For serious criminal cases, the court must give a serious decision, but it might be controversial to some extent. Cases about sex or child are also very serious and controversial since it might be hard to get confidential proof.

Besides, we can found that Fourth and Sixth Amendment often appears in reversed criminal cases. These two amendments are used to protect defendant right and help them to appeal. Therefore, cases including these words, defendants will be more likely to appeal to the court to get a more light conviction.

### 3.8.2 Relationship between Case Text and Informant/Non-informant Cases

In the graph Figure 21, there are some informative sequence including

- stolen goods
- commit murder
- witnesses credibility
- confidential informant
- life imprisonment without possibility parole

In cases with informants, we can find 'confidential informant' and 'witness credibility'. We think if a cases have these keywords and these features appears many time, it means informants' words are important to this case. The cases with informants may be more often related to 'stolen goods' and 'murder' and are not

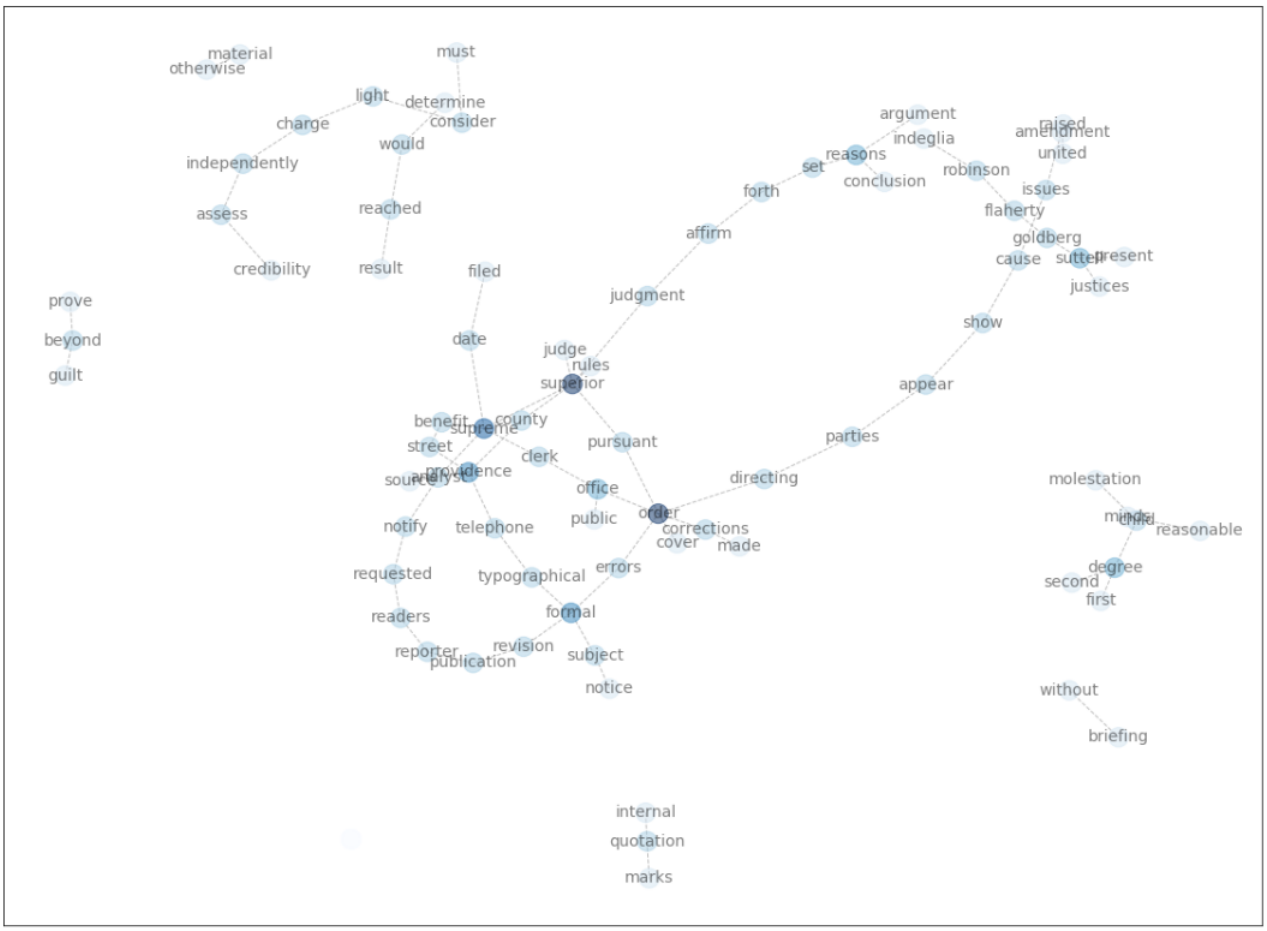


Figure 22: Rhode Island criminal cases which do not have informant

constrained to a single type of crime. Meanwhile, we found 'murder' appears many times in Tf-idf, which means criminal cases with informants have more chance related to 'murder' than 'stolen goods'. From the sequence 'life imprisonment without possibility parole', we can also know these cases are serious.

In the graph Figure 22, there are some informative sequence including

- child molestation

Criminal cases without informants do not have a lot of valuable information. There are many such cases, these cases have less differences with the patterns in all criminal cases. That's why we have few valuable sequences. The criminal cases without informants have frequent sequences: 'child molestation'. This pattern is same as the criminal cases.

### 3.8.3 Relationship between affirmed Cases Text and Informant Cases Decision

In the graph Figure 23, which is reversed cases containing informants, there are some informative sequence including

- abuse
- dangerous weapon
- sexual assult
- first / second degree murder
- ten years serve

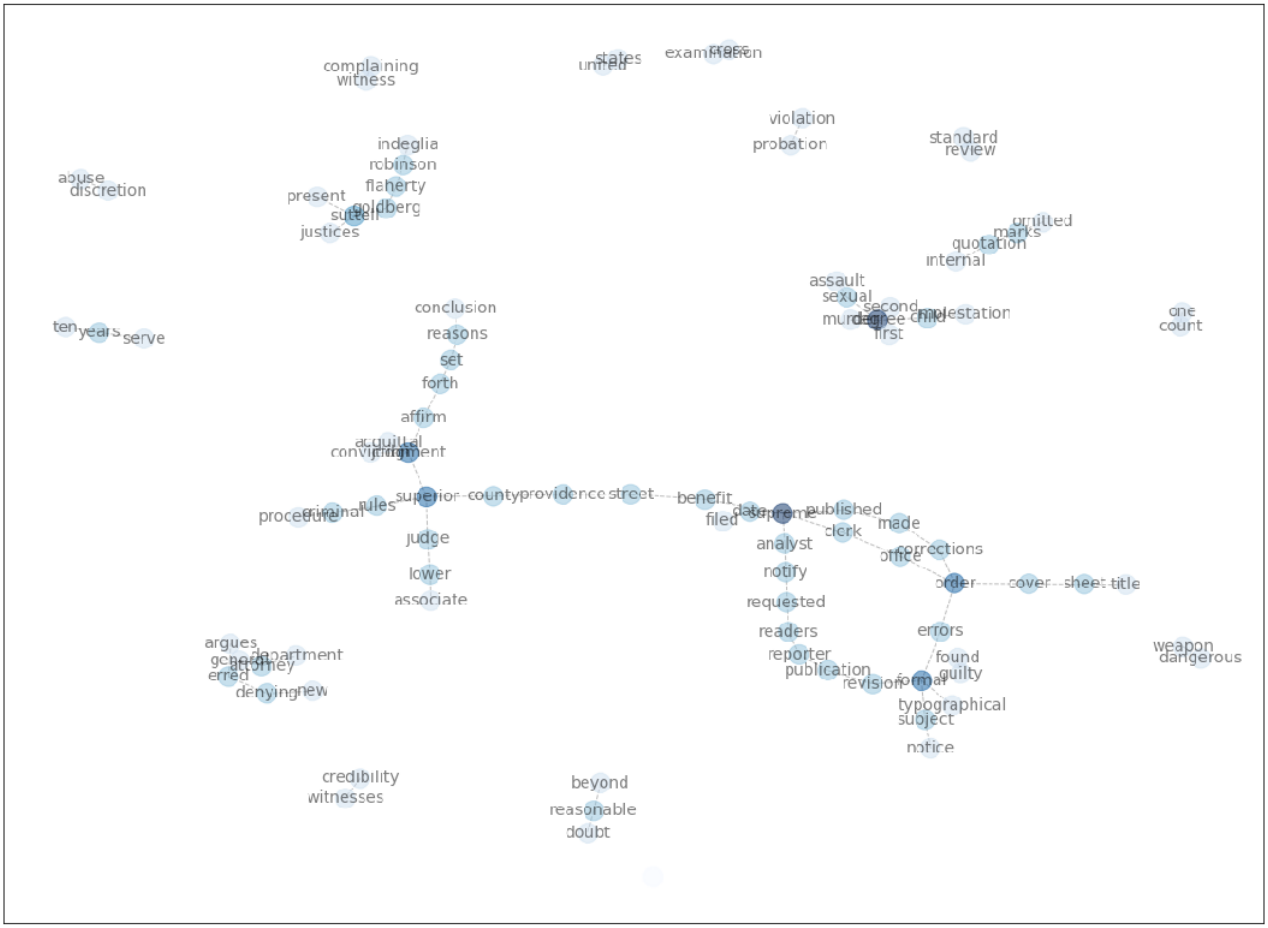


Figure 23: Rhode Island criminal cases which don't have informant and decision is affirm

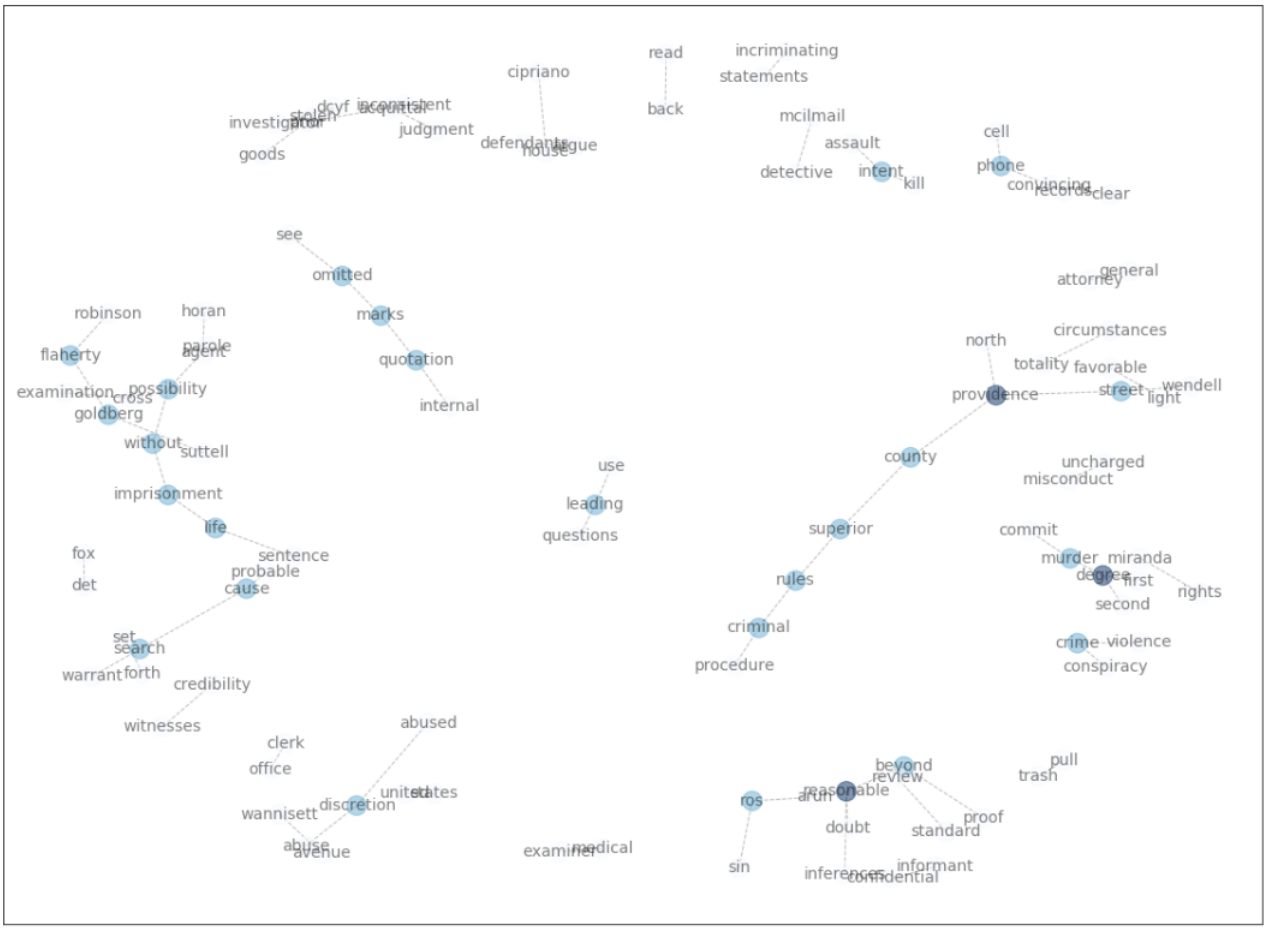


Figure 24: Rhode Island criminal cases which has informant and decision is affirm

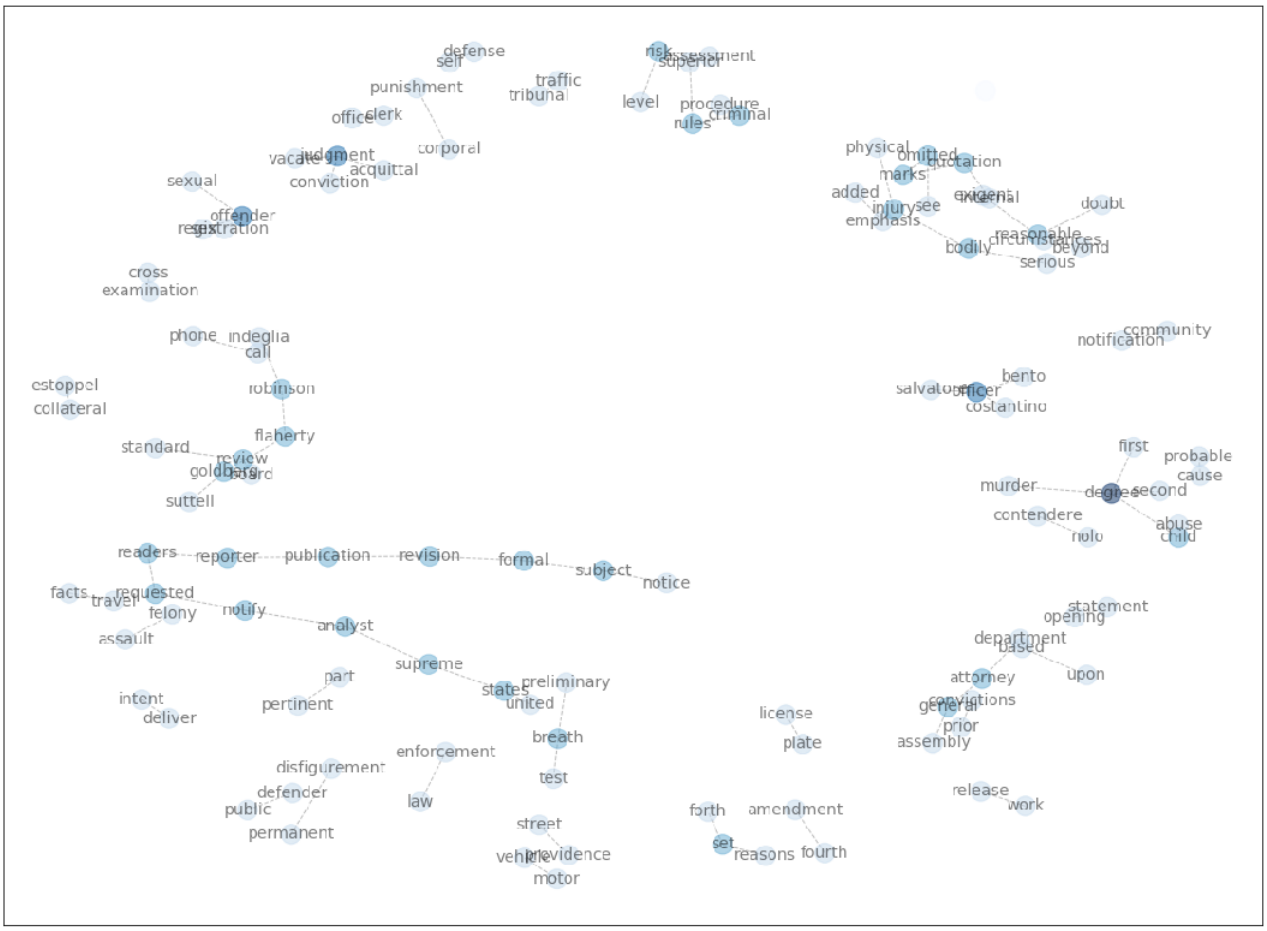


Figure 25: Rhode Island criminal cases which don't have informant and decision is reversed

In the graph Figure 24, which shows affirmed cases containing informants, there are some informative sequence including

- witness credibility
- life imprisonment without possibility parole
- first, second degree murder
- confidential informant
- intent kill
- abuse

We found that in this cases, "intent kill" might need more information from confidential informant to make final decision. Because "intent kill" is more specific and more severe than "murder", which are included in many kinds of crimes.

### 3.8.4 Relationship between Reversed Cases Text and Informant Cases Decision

Comparing Figure 25 and Figure 26, we found that informant cannot offer strong proof on self-defence cases. In graph Figure 25, which shows reversed cases containing informants.

In both graphs, "sex", "child", "abuse" appears. We can find that informant cannot have much influence on cases about them.

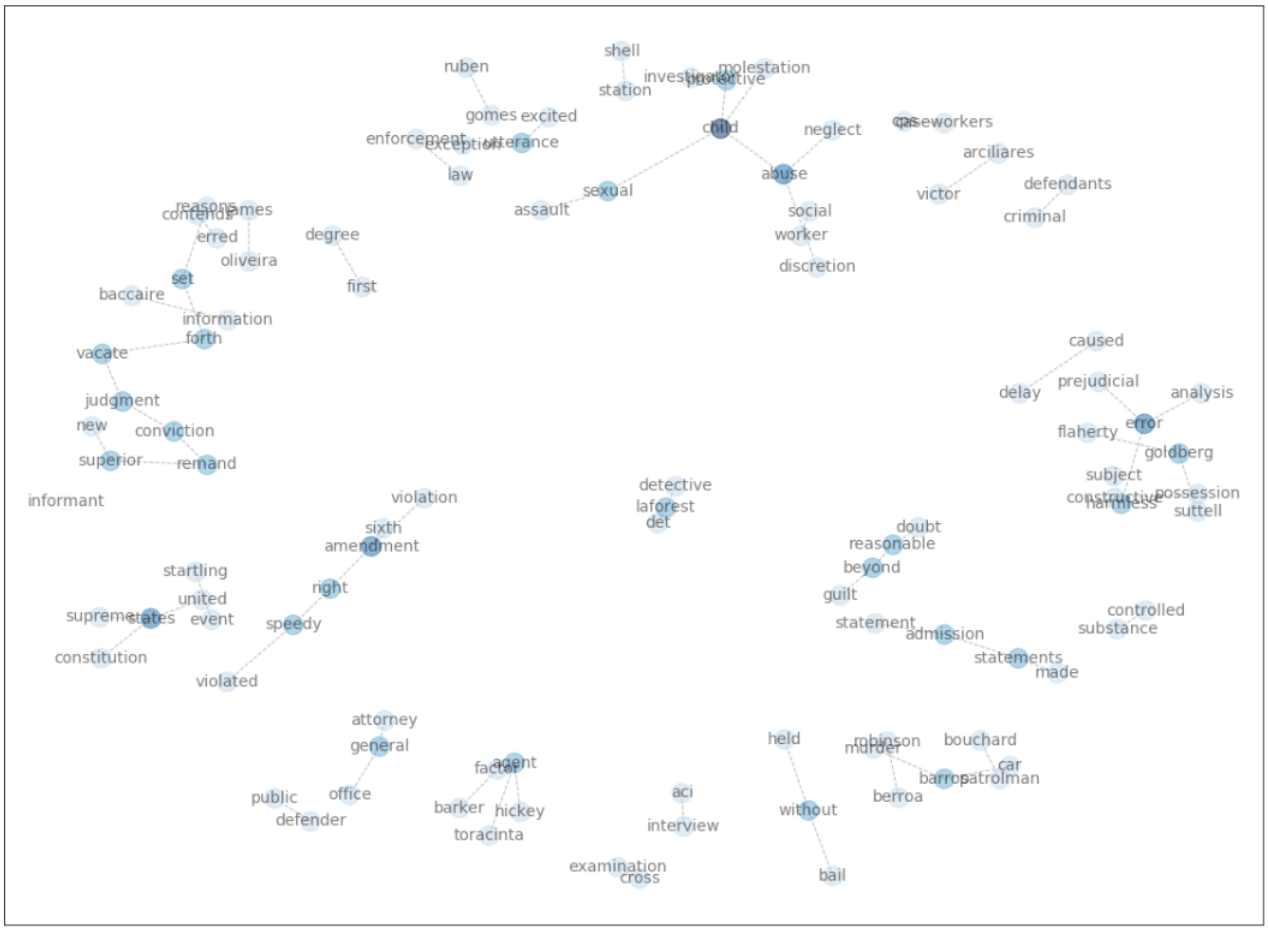


Figure 26: Rhode Island criminal cases which don't have informant and decision is reversed

# 4 Conclusion

In this project, we worked with the data of criminal cases from appellate courts in both New Hampshire and Rhode Island with steps of data collection, pre-processing and detailed analysis using Tfidf and N-gram.

Overall, we found that in criminal cases, cases that involved more severe charges such as murder and child sexual abuse were more likely to be overturned by comparing affirmed and reversed cases. In the cases that involved informants, informants usually played an important role in making the judgement and the case were more involved with murder. When we eventually compared reversed cases with informants and without informants, we were able to find that cases with informant were more likely to be affirmed. As we were working on the project, we hoped that there could be more data to be fed into our analysis methods. Using more data of criminal cases to continuously analyse the relation between confidential informant and final decision results can be our future work.

We hope our analysis for criminal cases in New Hampshire and Rhode Island can help courts and reporters rethinking what can cause obvious effect to the final decision for criminal cases and help courts and confidential informants make better contribution to the criminal cases' decisions in the future.