On close and distant reading in digital humanities: a survey and future challenges

S. Jänicke[1], G. Franzini[2], M. F. Cheema[1] and G. Scheuermann[1]

[1]Image and Signal Processing Group, Department of Computer Science, Leipzig University, Germany
[2]Göttingen Centre for Digital Humanities, University of Göttingen, Germany

Introduction

- This is a survey about close and distant reading, and the different techniques and tools that can be used in each of them.
- It’s going to be fun.
- And it will be over before you even notice it.

Definition of close and distant reading

Close reading:

Definition of close and distant reading

Close reading:

Reading to uncovering layers of meaning that lead to deep comprehension.

Definition of close and distant reading

Close reading:

Reading to uncovering layers of meaning that lead to deep comprehension.

Close reading includes the analysis of:

Definition of close and distant reading

Close reading:

Reading to uncovering layers of meaning that lead to deep comprehension.

Close reading includes the analysis of:

- individuals, events, ideas and their development

Definition of close and distant reading

Close reading:

Reading to uncovering layers of meaning that lead to deep comprehension.

Close reading includes the analysis of:

- individuals, events, ideas and their development
- Words and phrases

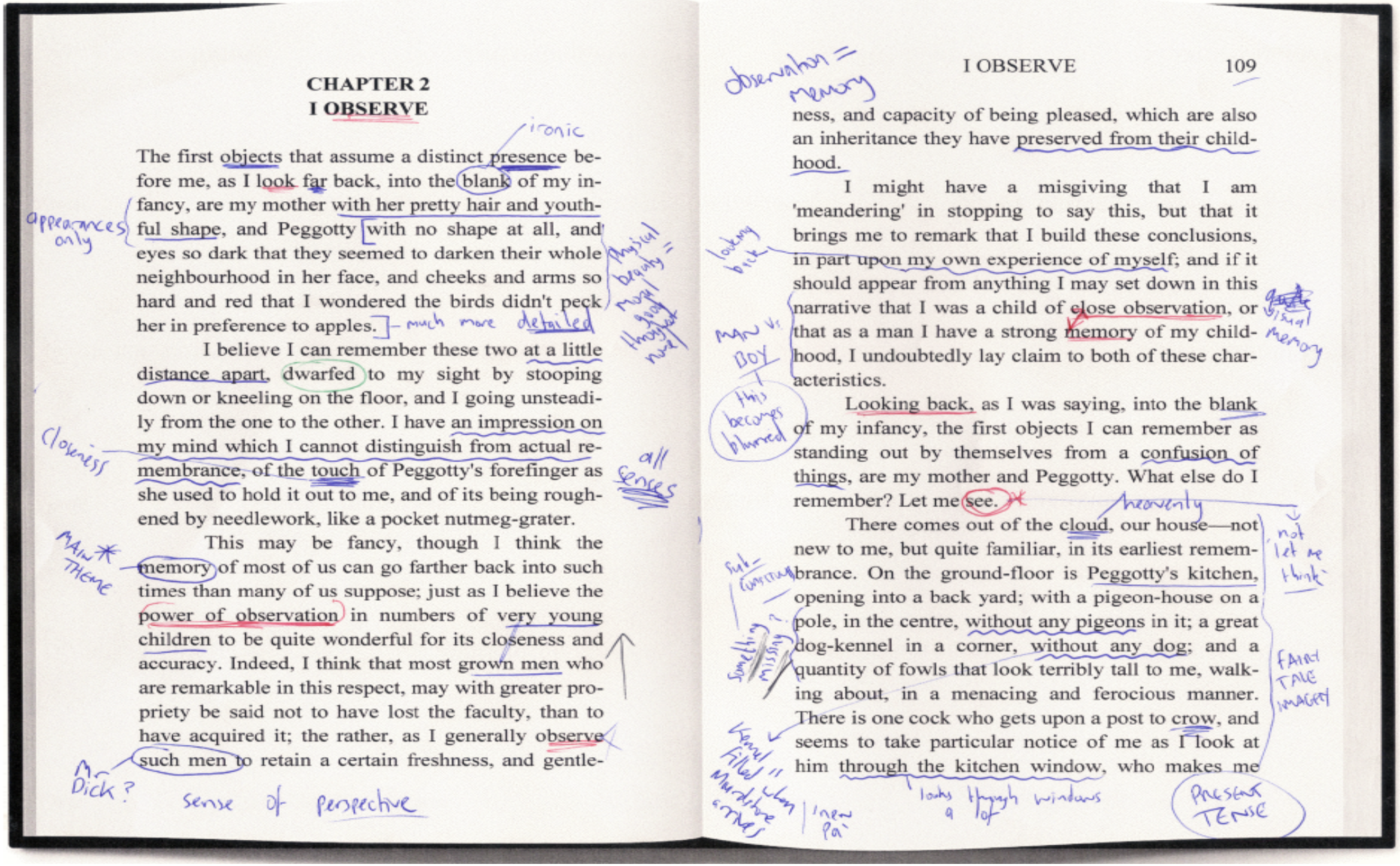
Definition of close and distant reading

Close reading:

Reading to uncovering layers of meaning that lead to deep comprehension.

Close reading includes the analysis of:

- individuals, events, ideas and their development
- Words and phrases
- Text structure and style



CHAPTER 2
I OBSERVE

The first objects that assume a distinct presence before me, as I look far back, into the blank of my infancy, are my mother with her pretty hair and youthful shape, and Peggotty with no shape at all, and eyes so dark that they seemed to darken their whole neighbourhood in her face, and cheeks and arms so hard and red that I wondered the birds didn't peck her in preference to apples.

I believe I can remember these two at a little distance apart, dwarfed to my sight by stooping down or kneeling on the floor, and I going unsteadily from the one to the other. I have an impression on my mind which I cannot distinguish from actual remembrance, of the touch of Peggotty's forefinger as she used to hold it out to me, and of its being roughened by needlework, like a pocket nutmeg-grater.

This may be fancy, though I think the memory of most of us can go farther back into such times than many of us suppose; just as I believe the power of observation in numbers of very young children to be quite wonderful for its closeness and accuracy. Indeed, I think that most grown men who are remarkable in this respect, may with greater propriety be said not to have lost the faculty, than to have acquired it; the rather, as I generally observe such men to retain a certain freshness, and gentle-

ness, and capacity of being pleased, which are also an inheritance they have preserved from their childhood.

I might have a misgiving that I am 'meandering' in stopping to say this, but that it brings me to remark that I build these conclusions, in part upon my own experience of myself; and if it should appear from anything I may set down in this narrative that I was a child of close observation, or that as a man I have a strong memory of my childhood, I undoubtedly lay claim to both of these characteristics.

Looking back, as I was saying, into the blank of my infancy, the first objects I can remember as standing out by themselves from a confusion of things, are my mother and Peggotty. What else do I remember? Let me see.

There comes out of the cloud, our house—not new to me, but quite familiar, in its earliest remembrance. On the ground-floor is Peggotty's kitchen, opening into a back yard; with a pigeon-house on a pole, in the centre, without any pigeons in it; a great dog-kennel in a corner, without any dog; and a quantity of fowls that look terribly tall to me, walking about, in a menacing and ferocious manner. There is one cock who gets upon a post to crow, and seems to take particular notice of me as I look at him through the kitchen window, who makes me

Close reading of the second chapter of Charles Dicken's "David Copperfield"

David Copperfield - Chapter 2 by Charles Dickens

Extract from http://www.gutenberg.org/ebooks/766

1869 Edition.

Section 1

1 CHAPTER 2. I OBSERVE

2 The first objects that assume a distinct presence before me, as
3 far back, into the blank of my infancy, are my mother with her pr
4 hair and youthful shape, and Peggotty with no shape at all, and
5 dark that they seemed to darken their whole neighbourhood in h
6 and cheeks and arms so hard and red that I wondered the birds
7 peck her in preference to apples.

8 I believe I can remember these two at a little distance apart, dwa
9 to my sight by stooping down or kneeling on the floor, and I goin
10 unsteadily from the one to the other. I have an impression on my
11 which I cannot distinguish from actual remembrance, of the touch of

I OBSERVE

John Thomas said:

All early chapter titles cast David in passive role: 'I am Born', 'I Observe', 'I have a Change', 'I fall into Disgrace', 'I am sent away from Home'. This changes after he is forced to leave home: 'I enlarge my Circle of Acquaintance'.

Peter Hughes said:

'I Observe' is not quite passive but I agree that David begins to play a more active role after he is forced to leave home.

Sarah Young said:

Observation and memory are closely linked throughout this chapter and throughout the novel.

Admin

Sections

15:15:16

All Groups

Switch

Tags

memory
senses
missing
childlike
mother
time
overprotected
humour
window
infancy
instant
new

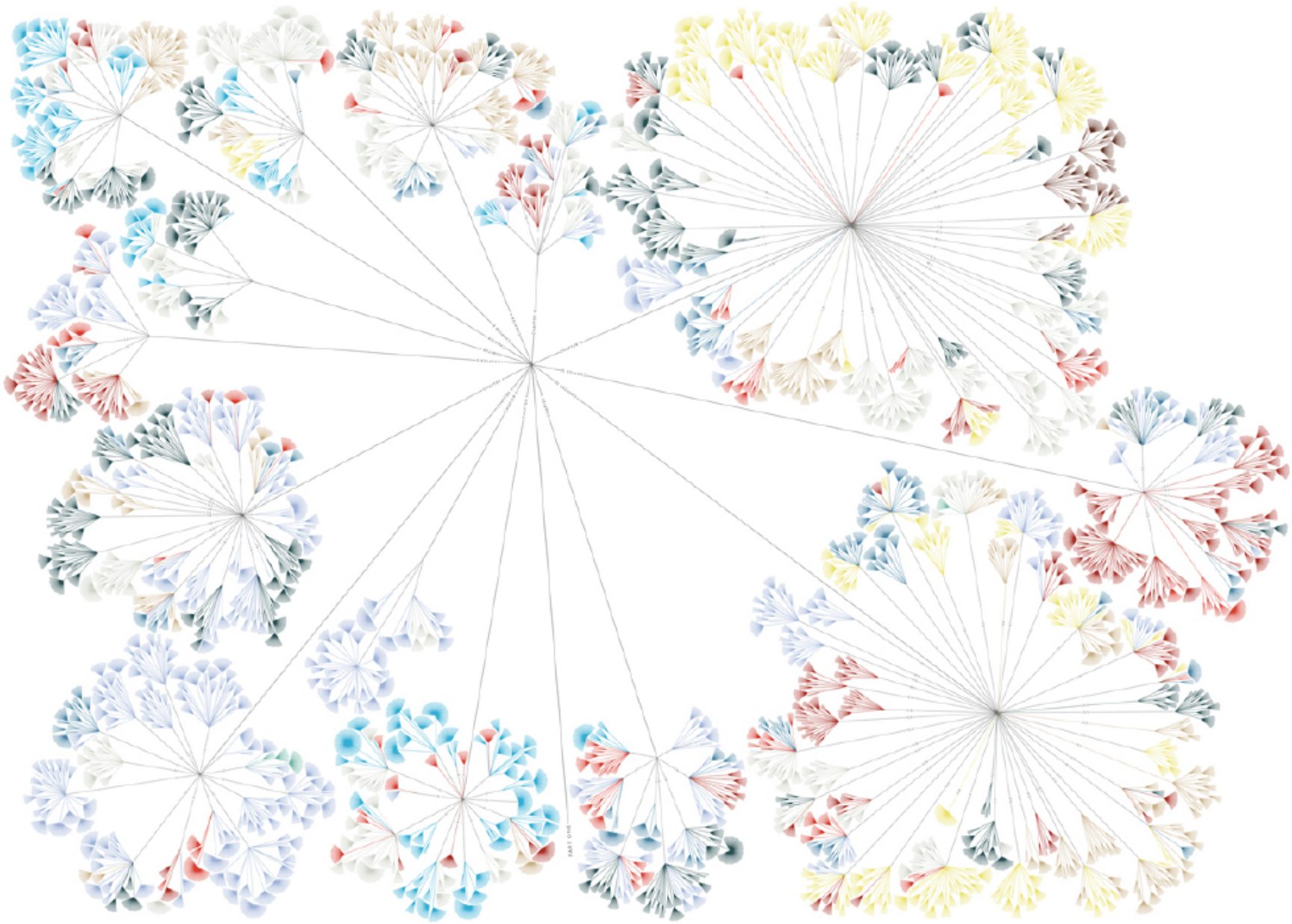
Definition of close and distant reading

Distant reading:

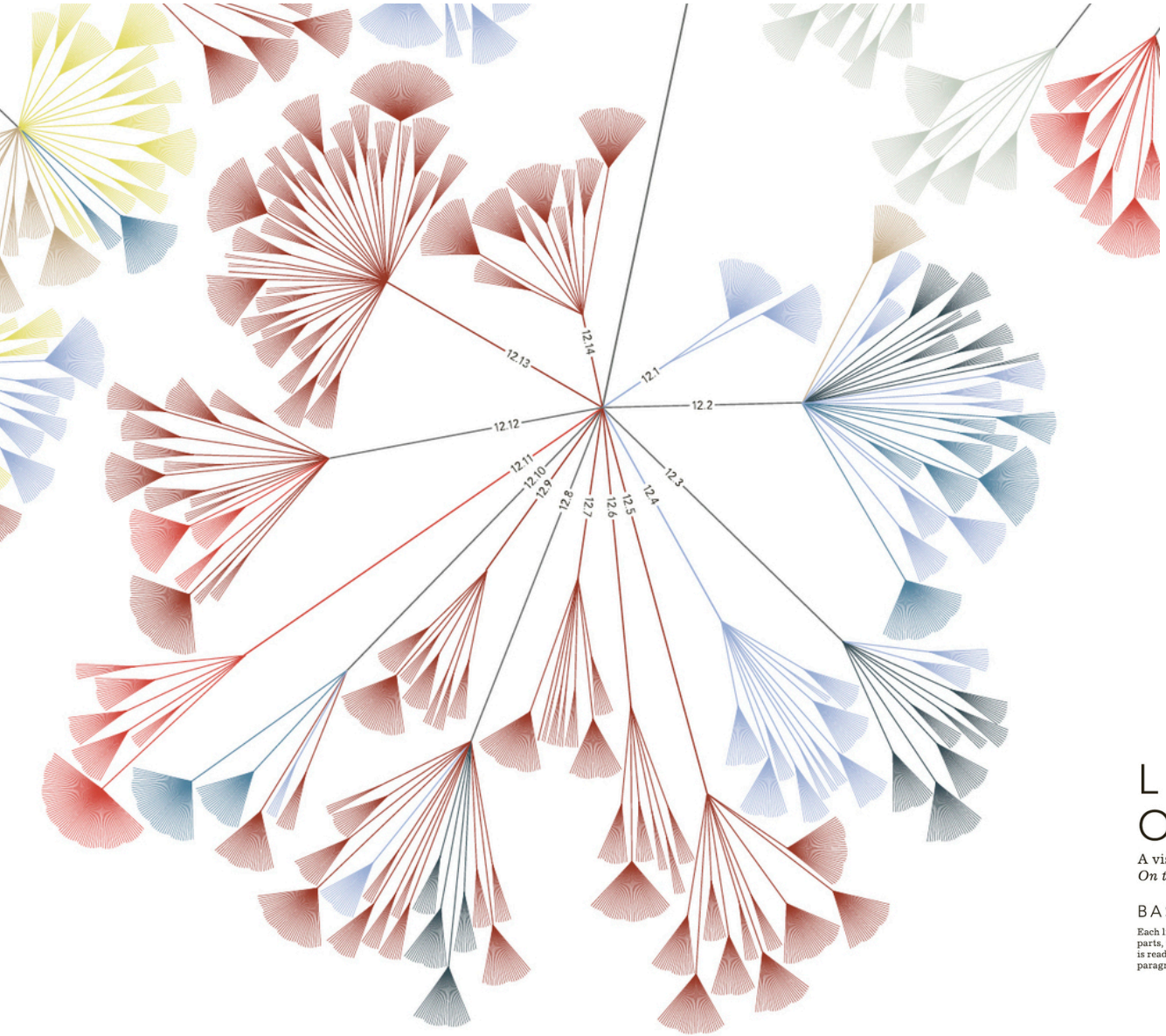
Definition of close and distant reading

Distant reading:

Distant reading aims to generate an abstract view by shifting from observing textual contact to observing global features of a single or multiple text(s).



This is a distant reading of “On the Road” by Jack Kerouac in the form of a tree

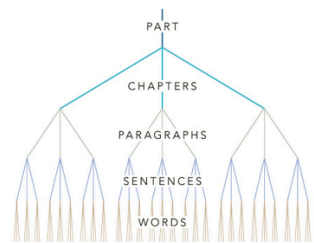


LITERARY ORGANISM

A visualization of Part One of *On the Road,* by Jack Kerouac

BASIC STRUCTURE

Each literary component can be divided into even smaller parts, the smallest in this diagram being words. The diagram is read clockwise, starting from the first chapter, paragraph, or sentence.



NOTATION

Each quotation can be referenced in *On the Road.*

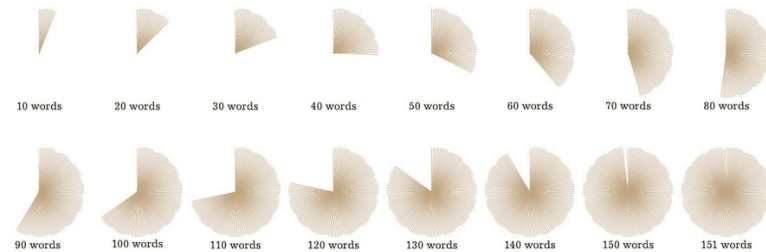
chapter
3.5
paragraph

COLORS

- Dean Moriarty (Protagonist)
- Bop & Jazz Music
- Social Events & Interaction
- Travel
- Sketches of Regional Life
- Parties, Drinking & Drugs
- Work & Survival
- Sal Paradise (Narrator)
- Women, Sex, & Relationships
- Illegal Activities & Police Encounters
- Character Sketches

WORD COUNT CHART

All sentences are in proportion to the largest sentence in Part One, which is 151 words long.



Do we still need close reading?

Do we still need close reading?

Yes.

Do we still need close reading?

Yes.

- Distant reading can’t replace close reading

Do we still need close reading?

Yes.

- Distant reading can’t replace close reading
- It can direct reader to a section that may deserve further investigation

Do we still need close reading?

Yes.

- Distant reading can’t replace close reading
- It can direct reader to a section that may deserve further investigation
- Highlights potentially interesting patterns

Close reading techniques

A visualization that need to retained the structure of the text, in order to allow close reading.

Close reading techniques

A visualization that need to retained the structure of the text, in order to allow close reading.

- Color

Close reading techniques

A visualization that need to retained the structure of the text, in order to allow close reading.

- Color
- Font size

Close reading techniques

A visualization that need to retained the structure of the text, in order to allow close reading.

- Color
- Font size
- Glyphs

Close reading techniques

A visualization that need to retained the structure of the text, in order to allow close reading.

- Color
- Font size
- Glyphs
- Connections

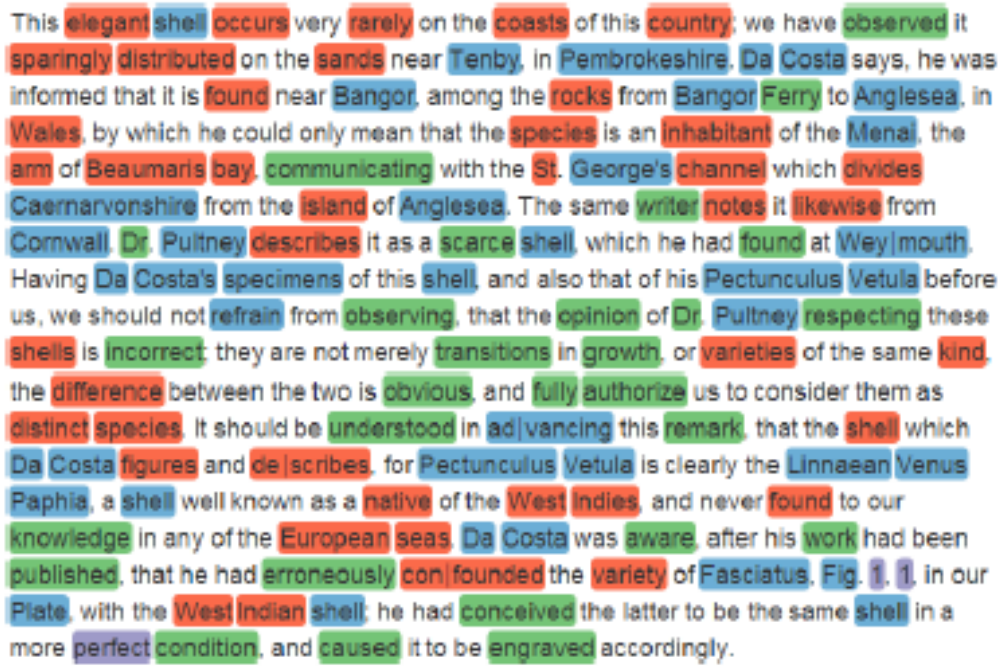
Close reading techniques - Color

Close reading techniques - Color

- Very common visual attribute to display features of text

Close reading techniques - Color

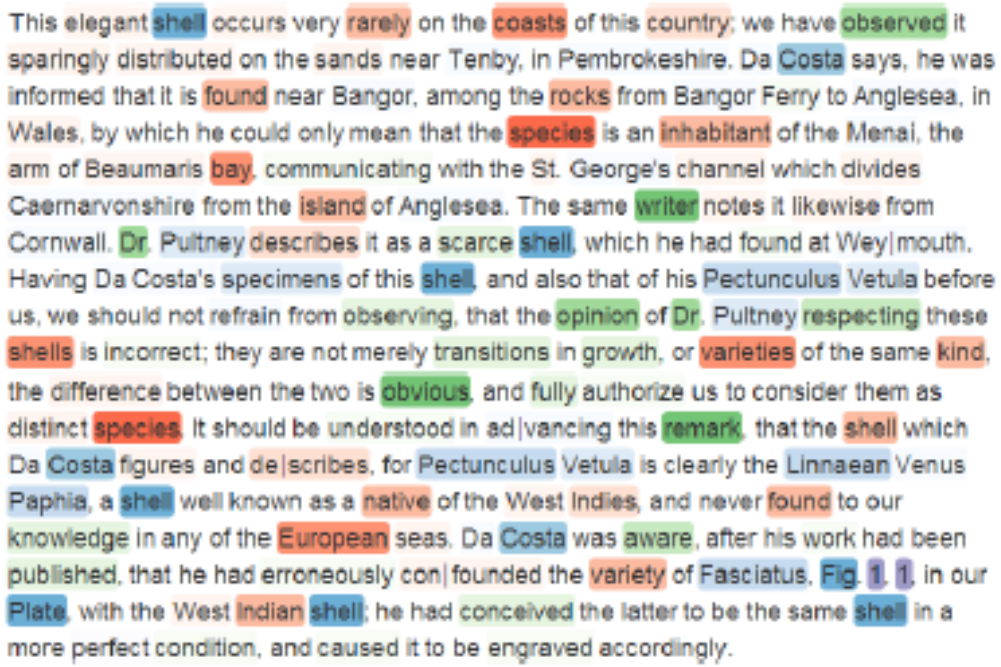
- Very common visual attribute to display features of text
- Color background



The Natural History of British Shell

Close reading techniques - Color

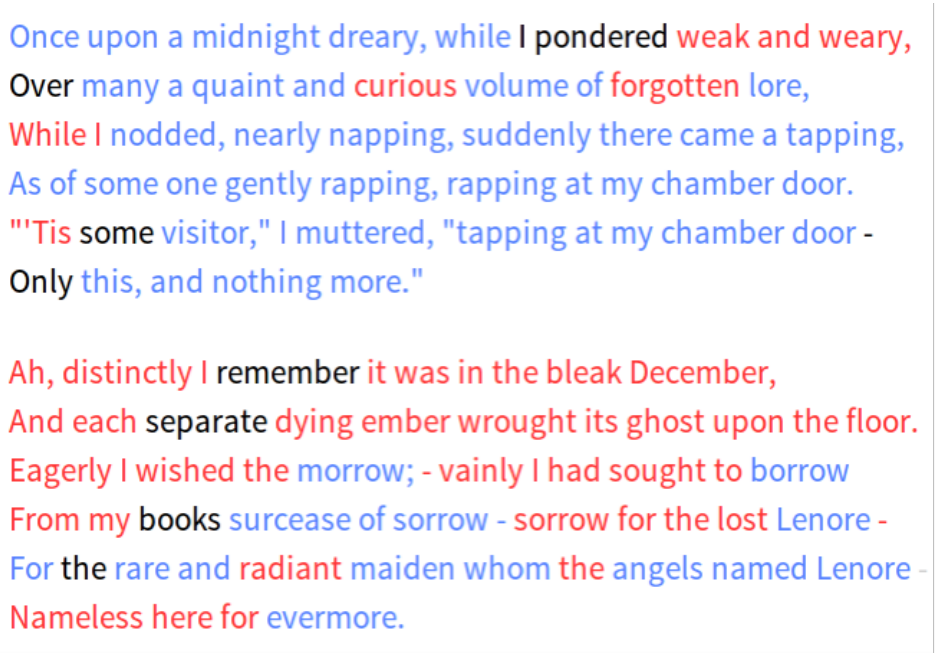
- Very common visual attribute to display features of text
- Color background
- Transparency of the background color for importance


This elegant shell occurs very rarely on the coasts of this country; we have observed it sparingly distributed on the sands near Tenby, in Pembrokeshire. Da Costa says, he was informed that it is found near Bangor, among the rocks from Bangor Ferry to Anglesea, in Wales, by which he could only mean that the species is an inhabitant of the Menai, the arm of Beaumaris bay, communicating with the St. George's channel which divides Caernarvonshire from the island of Anglesea. The same writer notes it likewise from Cornwall. Dr. Pultney describes it as a scarce shell, which he had found at Wey|mouth. Having Da Costa's specimens of this shell, and also that of his Pectunculus Vetula before us, we should not refrain from observing, that the opinion of Dr. Pultney respecting these shells is incorrect; they are not merely transitions in growth, or varieties of the same kind, the difference between the two is obvious, and fully authorize us to consider them as distinct species. It should be understood in ad|vancing this remark, that the shell which Da Costa figures and de|scribes, for Pectunculus Vetula is clearly the Linnaean Venus Paphia, a shell well known as a native of the West Indies, and never found to our knowledge in any of the European seas. Da Costa was aware, after his work had been published, that he had erroneously con|founded the variety of Fasciatus, Fig. 1. 1. in our Plate, with the West Indian shell; he had conceived the latter to be the same shell in a more perfect condition, and caused it to be engraved accordingly.

The Natural History of British Shell

Close reading techniques - Color

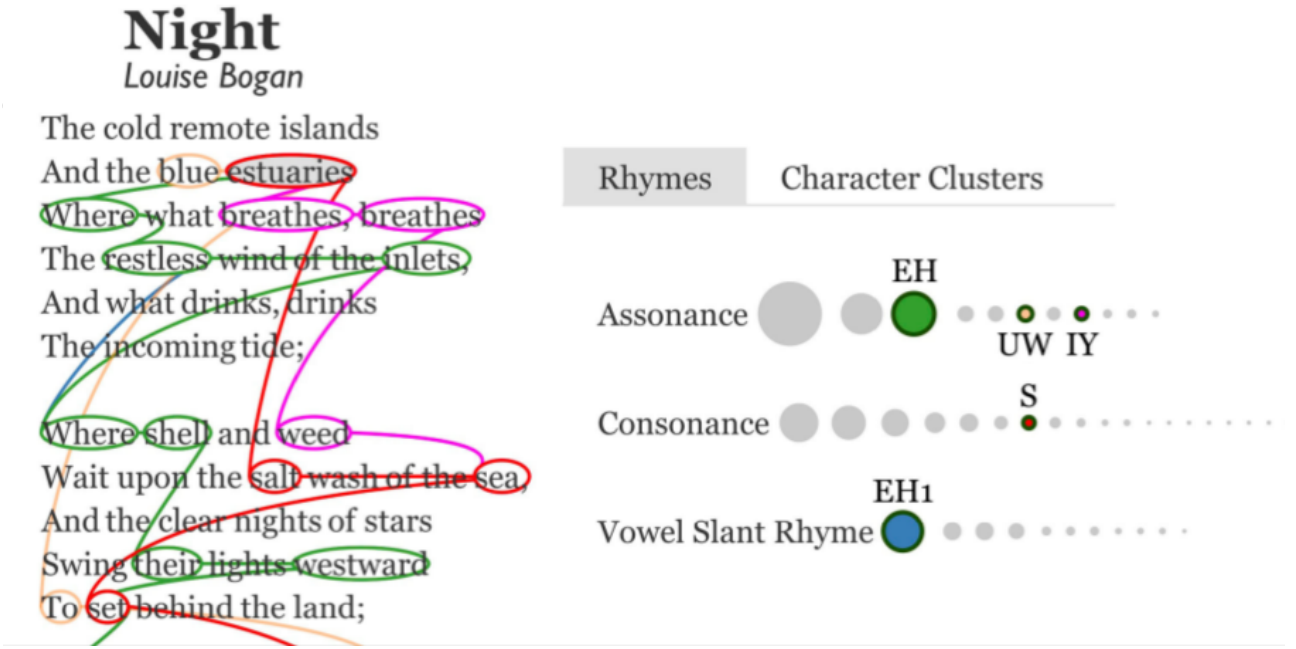
- Very common visual attribute to display features of text
- Color background
- Transparency of the background color for importance
- Font color



The Raven by Edgar Allan Poe

Close reading techniques - Color

- Very common visual attribute to display features of text
- Color background
- Transparency of the backgroun color for importance
- Font color
- Colored circles



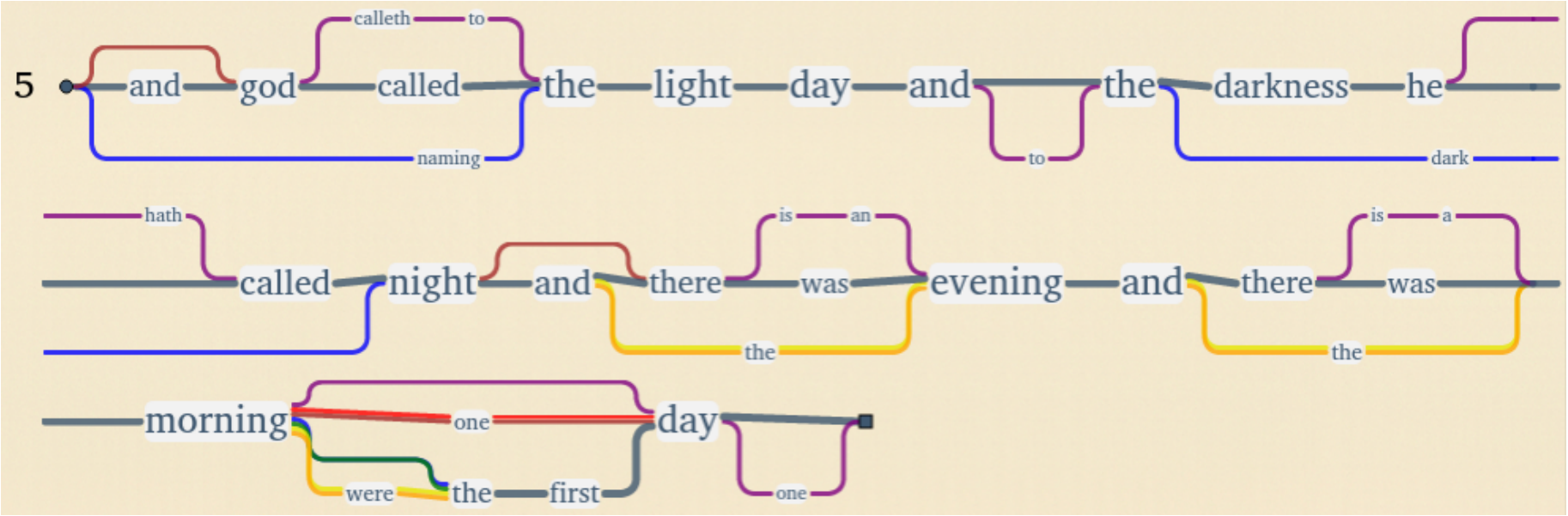
Close reading techniques – Font size

Close reading techniques – Font size

- Highlight significance or weight

Close reading techniques – Font size

- Highlight significance or weight
- Encodes the number of occurrences of a word among all editions.



Seven English translations of Genesis 1:5 connecting subsequent words displayed with variable font size

Close reading techniques – Font size

- Highlight significance or weight
- Encodes the number of occurrences of a word among all editions.
- Display the number of annotations of each word

The Raven by Edgar Allan Poe

Close reading techniques – Glyphs

Close reading techniques – Glyphs

- attaching glyphs to textual entities are to visualize abstract annotation that are hardly expressible with color of font size

Close reading techniques – Glyphs

- attaching glyphs to textual entities are to visualize abstract annotation that are hardly expressible with color of font size.
- Mostly used for poem.

Close reading techniques – Glyphs

- attaching glyphs to textual entities are to visualize abstract annotation that are hardly expressible with color of font size.
- Mostly used for poem.
- Use of glyphs in the form of rectangular and circular ma for the visualization of a poem's structure.

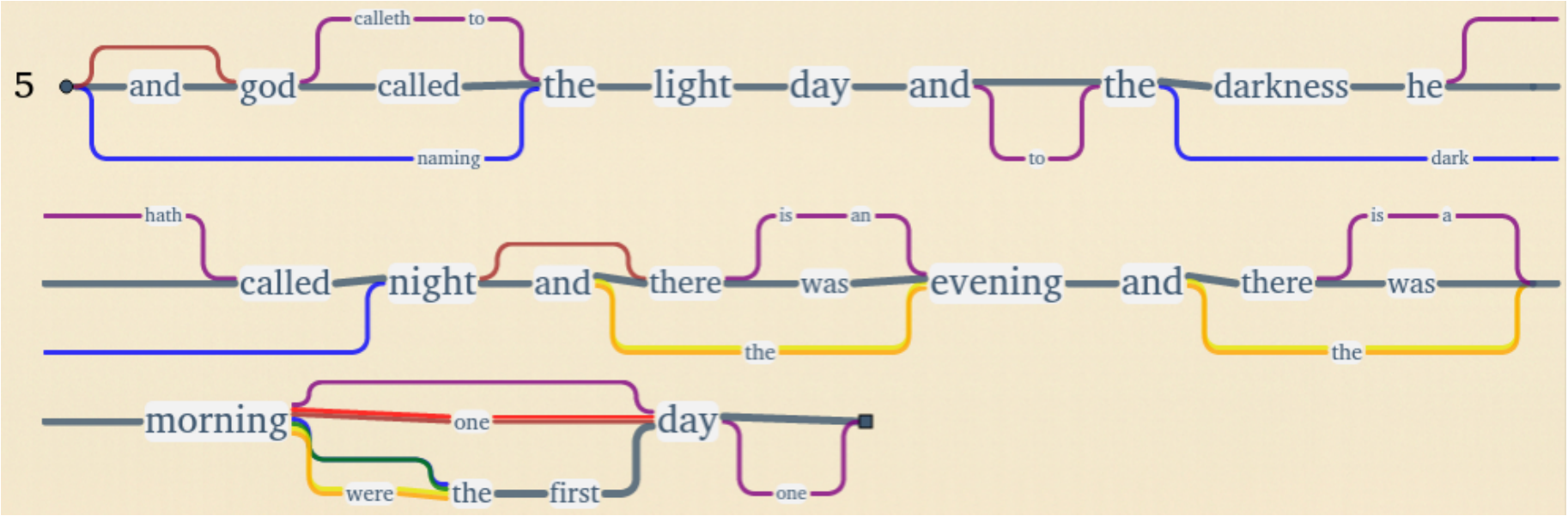
Close reading techniques - Connection

Close reading techniques - Connection

- Connections aid to illustrate the structure among textual entities.

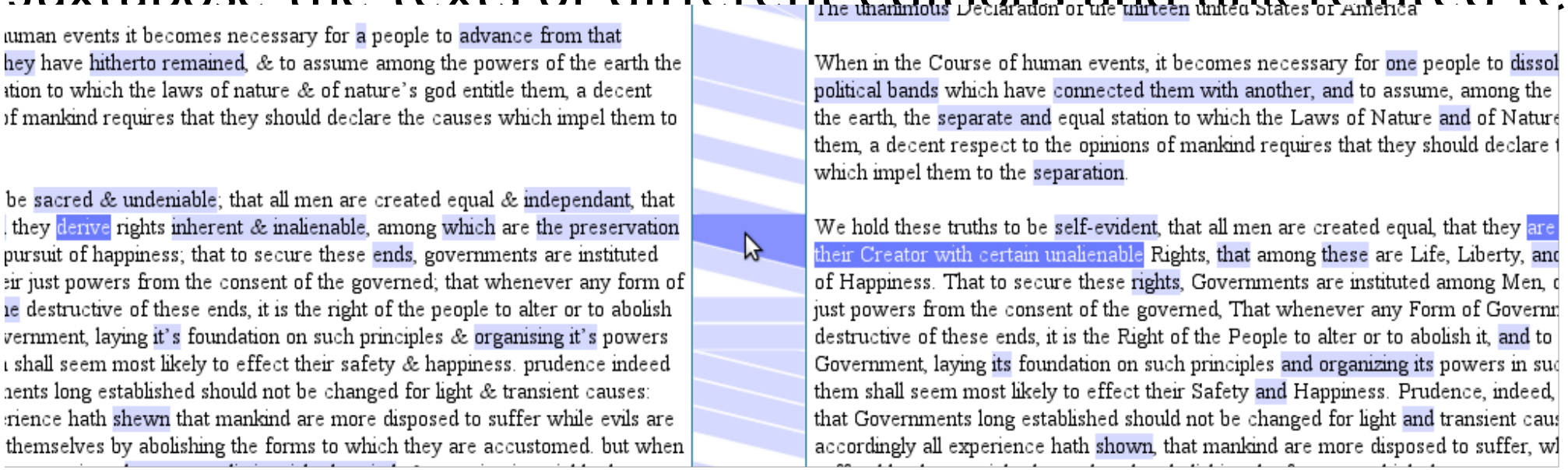
Close reading techniques - Connection

- Connections aid to illustrate the structure among textual entities.
- Track variation among various text editions.



Close reading techniques - Connection

- Connections aid to illustrate the structure among textual entities.
- Track variation among various text editions.
- Juxtapose the texts of different editions and link related text



United States Declaration of Independence

Close reading techniques - Connection

- Connections aid to illustrate the structure among textual entities.
- Track variation among various text editions.
- Juxtapose the texts of different editions and link related text passages
- Visualize the phonetic and semantic relations within poems.

You’re doing great.

Distant reading techniques

Distant reading techniques

A visualization that display summarized information of the given text.

Distant reading techniques

A visualization that display summarized information of the given text.

- Heat maps
- Tag clouds
- Maps
- Timelines
- Graphs
- Miscellaneous

Distant reading techniques – Heat maps

Distant reading techniques – Heat maps

- Highlight textual patterns

Distant reading techniques – Heat maps

- Highlight textual patterns
- Show relationships between texts

Distant reading techniques – Heat maps

- Highlight textual patterns
- Show relationships between texts
- Can be used to visualize the similarity or the flow of sound in pomes

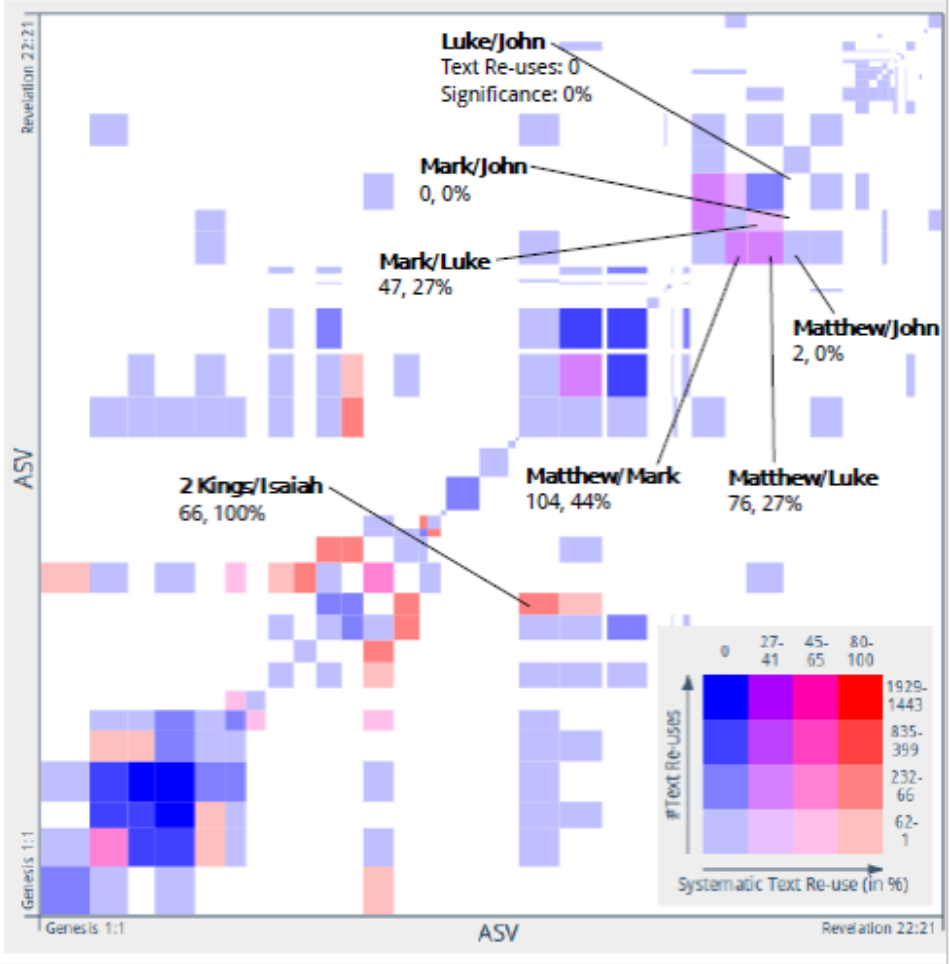
Distant reading techniques – Heat maps

- Highlight textual patterns
- Show relationships between texts
- Can be used to visualize the similarity or the flow of sound in pomes
- For example, to highlight the systematic text re-use among the books of the bible.

Distant reading techniques – Heat maps

- Highlight textual patterns
- Show relationships between text
- Can be used to visualize the similarity or the flow of sound in
- For example, to highlight the systematic text re-use among the books of the bible.

(texts share patterns of consecutive similar sentences)



Distant reading techniques – Tag clouds

Distant reading techniques – Tag clouds

- Intuitive visualizations for number of occurrences of words by using variable font size.

Distant reading techniques – Tag clouds

- Intuitive visualizations for number of occurrences of words by using variable font size.
- Can be limited to displaying only characteristic tags.

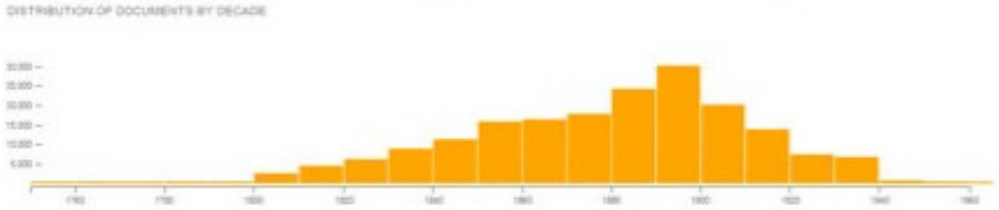
Distant reading techniques – Tag clouds

- Intuitive visualizations for number of occurrences of words by using variable font size.
- Can be limited to displaying only characteristic tags.
- Can be summarize to major tags for topics.

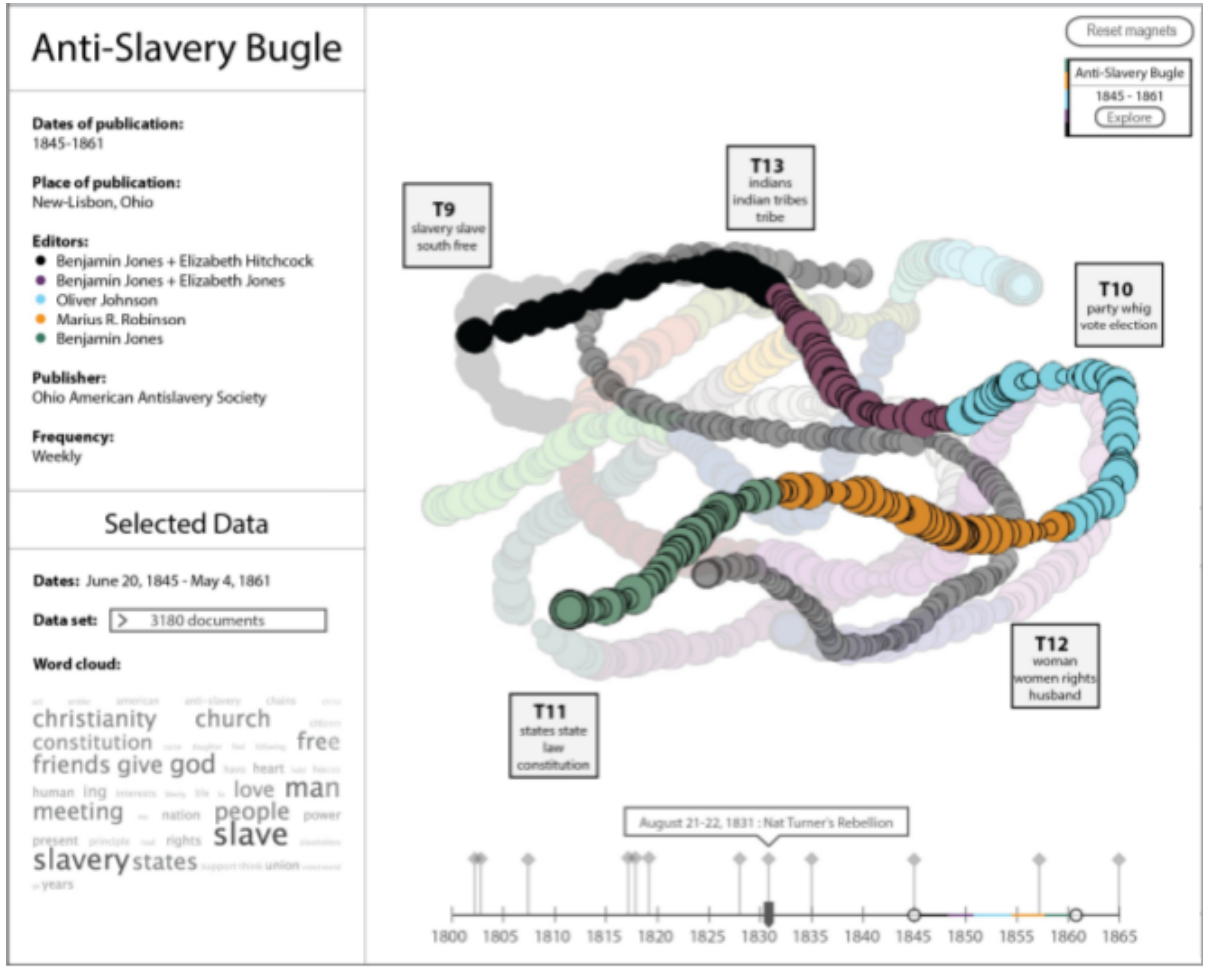
Distant reading techniques – Tag clouds

- Intuitive visualizations for number of occurrences of words by using variable font size.
- Can be limited to displaying only characteristic tags.
- Can be summarize to major tags for topics.
- Tag coloring is used to express additional information.

Gelände
Styled Map
27
41
37
168
575
290
488
3581
30450
14588
4150
46
2213
3519
2696
2607
507
2
1062
1590
323
13
Google


DISTRIBUTION OF DOCUMENTS BY DECADE

acid
animals
barley
bread
butter
cattle
coal
copper
corn
cotton
fish
flour
food
fruit
glass
gold
grain
horse
horses
ice
iron
lead
leather
lime
maps


Anti-Slavery Bugle
Dates of publication:
1845-1861
Place of publication:
New-Lisbon, Ohio
Editors:
Benjamin Jones + Elizabeth Hitchcock
Benjamin Jones + Elizabeth Jones
Oliver Johnson
Marius R. Robinson
Benjamin Jones
Publisher:
Ohio American Antislavery Society
Frequency:
Weekly
Selected Data
Dates: June 20, 1845 - May 4, 1861
Data set: > 3180 documents
Word cloud:
christianity church
constitution free
friends give god
human ing love man
meeting nation people power
present rights slave
slavery states
years
Reset magnets
Anti-Slavery Bugle
1845 - 1861
Explore
T9
slavery slave
south free
T13
indians
indian tribes
tribe
T10
party whig
vote election
T12
woman
women rights
husband
T11
states state
law
constitution
August 21-22, 1831 : Nat Turner's Rebellion
1800 1805 1810 1815 1820 1825 1830 1835 1840 1845 1850 1855 1860 1865

Distant reading techniques - Maps

Distant reading techniques - Maps

- Used to display geographic information contained in a text

Distant reading techniques - Maps

- Used to display geographic information contained in a text
- Some use thematic or density map but circles are more frequent

Combination of map, timeline and tag cloud for exploring commodity trading.

Distant reading techniques - Maps

- Used to display geographic information contained in a text
- Some use thematic or density ma but circles are more frequent
- But you can use more than just circles, to encode different types places.

circle - cities, square - regions, triangle - countries , diamond - others.

Distant reading techniques - Timeline

Distant reading techniques - Timeline

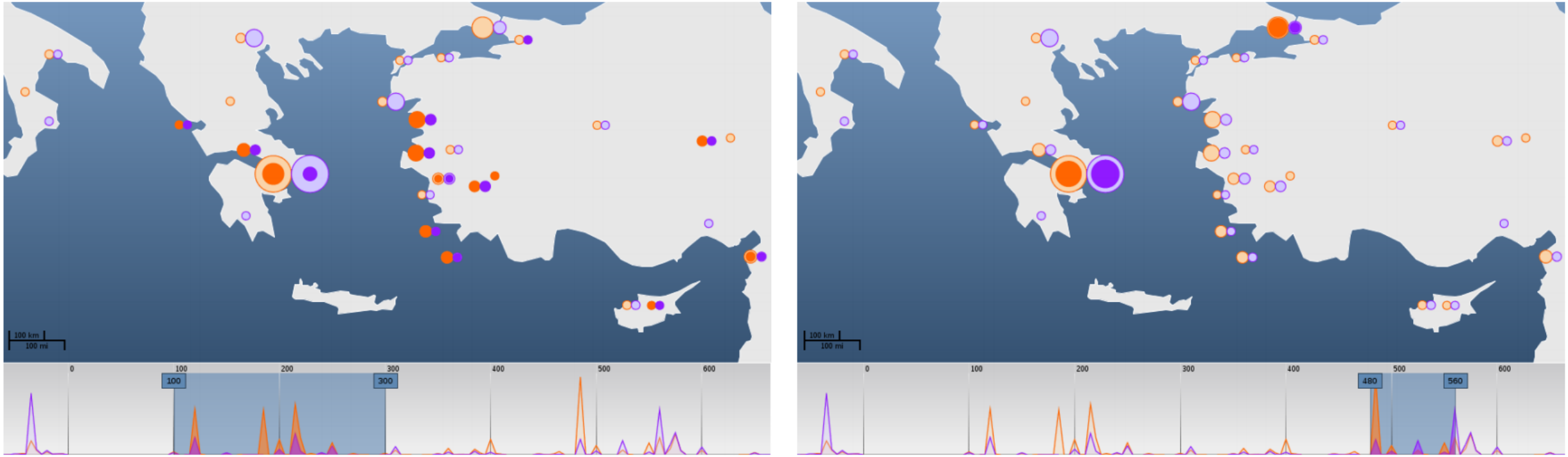
- This is used to visualize text corpus that has various types of temporal information.

Distant reading techniques - Timeline

- This is used to visualize text corpus that has various types of temporal information.
- One approach is the straightforward use of the text’s metadata

Distant reading techniques - Timeline

- Temporal analysis of a word's usage in ancient Greek texts



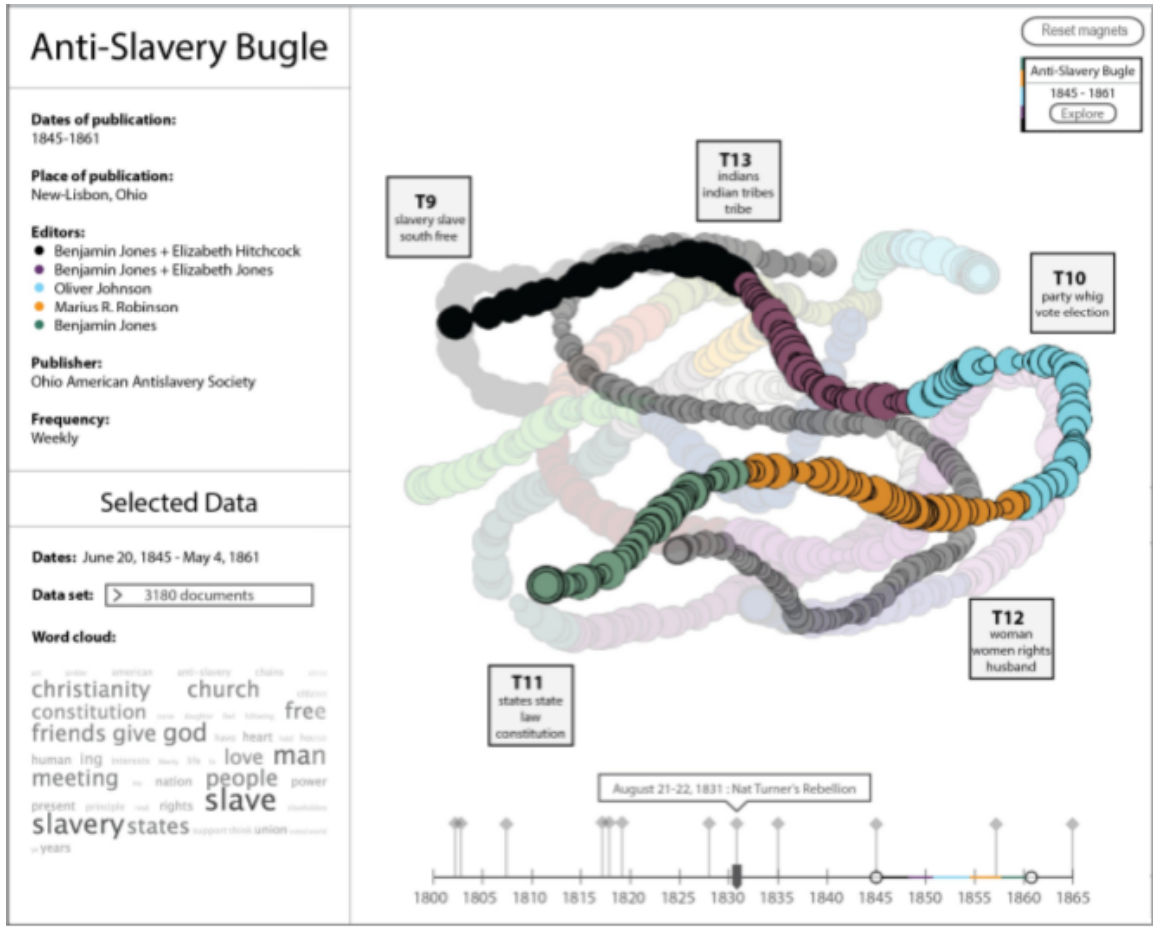
(a) Occurences during Middle Platonism

(b) Occurences during Neoplatonism

Occurrences of the words "Plato" (red) and "Aristotle" (blue) in ancient Greek texts

Distant reading techniques - Timeline

- The exploration of events in news articles



Distant reading techniques - Timeline

- Sometimes, the temporal information about events reported in a text
  needs to be extracted in order to visualize

Distant reading techniques – Graphs

Distant reading techniques – Graphs

- Used to visualize certain structural features of a texts

Distant reading techniques – Graphs

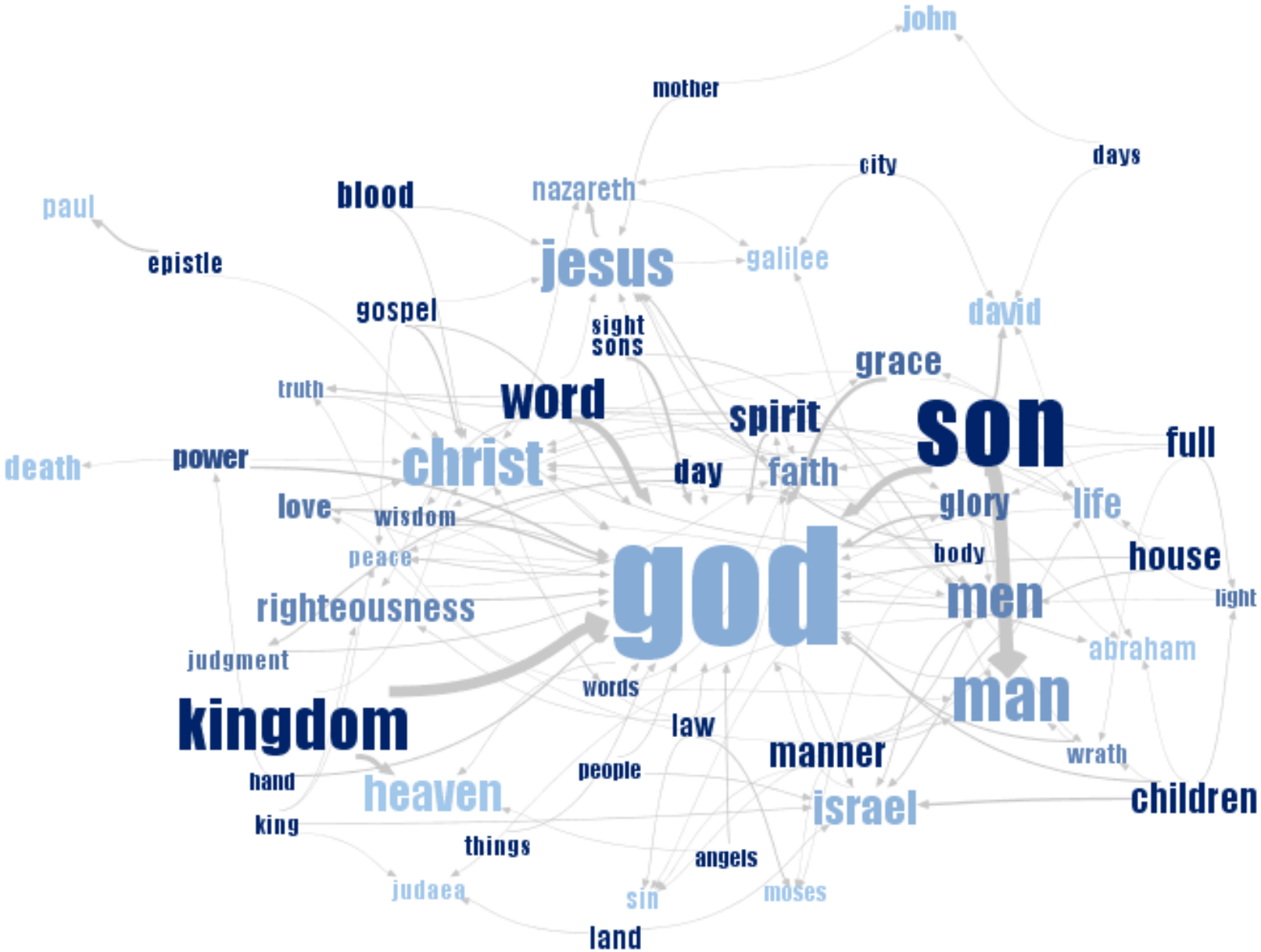
- Used to visualize certain structural features of a texts
- Relationships between the texts (represented by nodes) in the form of a network

Distant reading techniques – Graphs

- Used to visualize certain structural features of a texts
- Relationships between the texts (represented by nodes) in the form of a network
- Proximity is used to show similarity based on similar paragraphs or style

Distant reading techniques – Graphs

- Used to visualize certain structural features of a texts
- Relationships between the texts (represented by nodes) in the form of a network
- Proximity is used to show similarity based on similar paragraphs or style
- Phrase nets connect textual entities that appear in the certain form



Phrase nets for ’[word] of [word]’ within the Bible’s New Testament.

Distant reading techniques – Graphs

- Social networks are graphs visualizing the relationships between people.

Distant reading techniques – Graphs

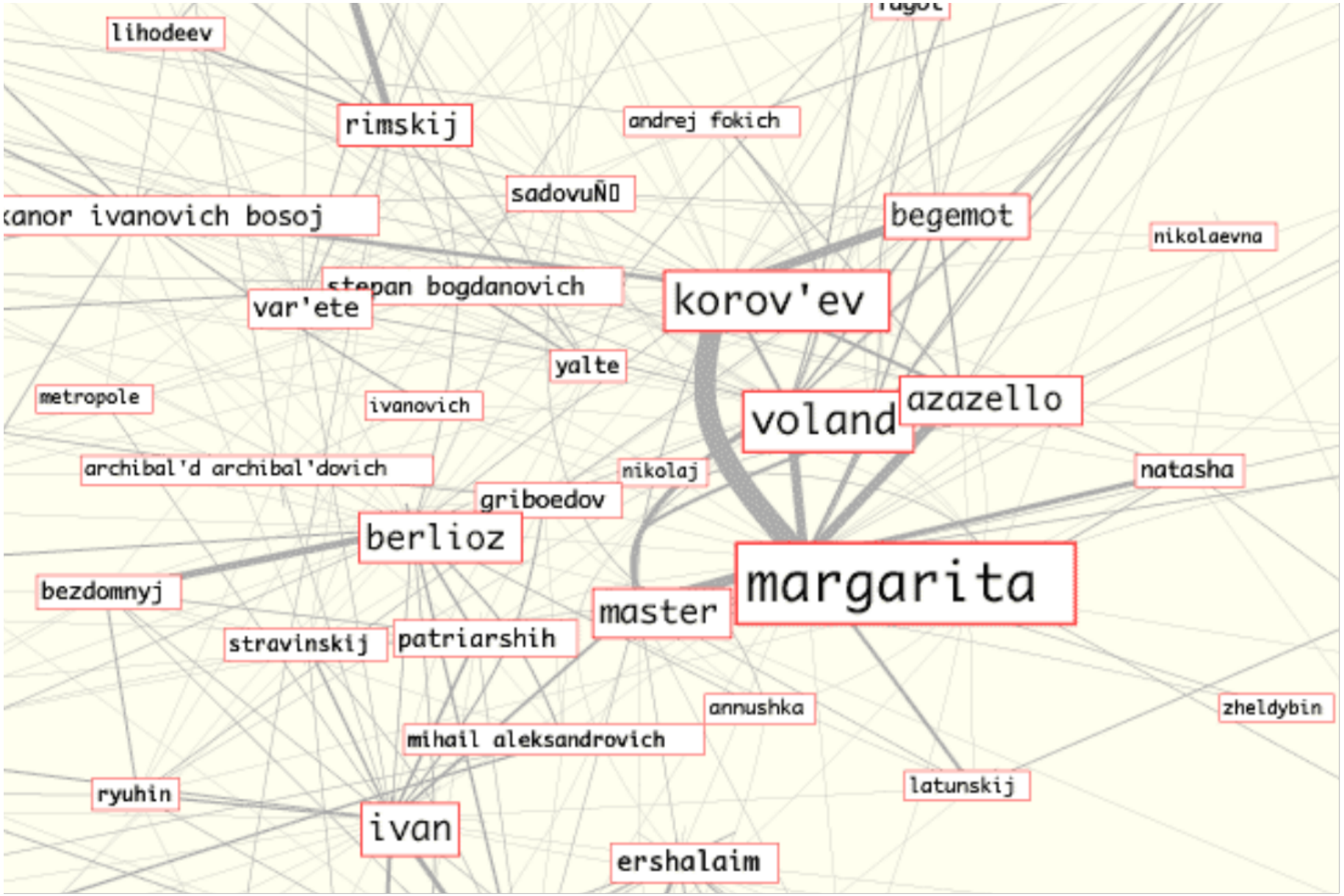
- Social networks are graphs visualizing the relationships between people.
- Used to illustrate the relationships between characters in literary texts.

Distant reading techniques – Graphs

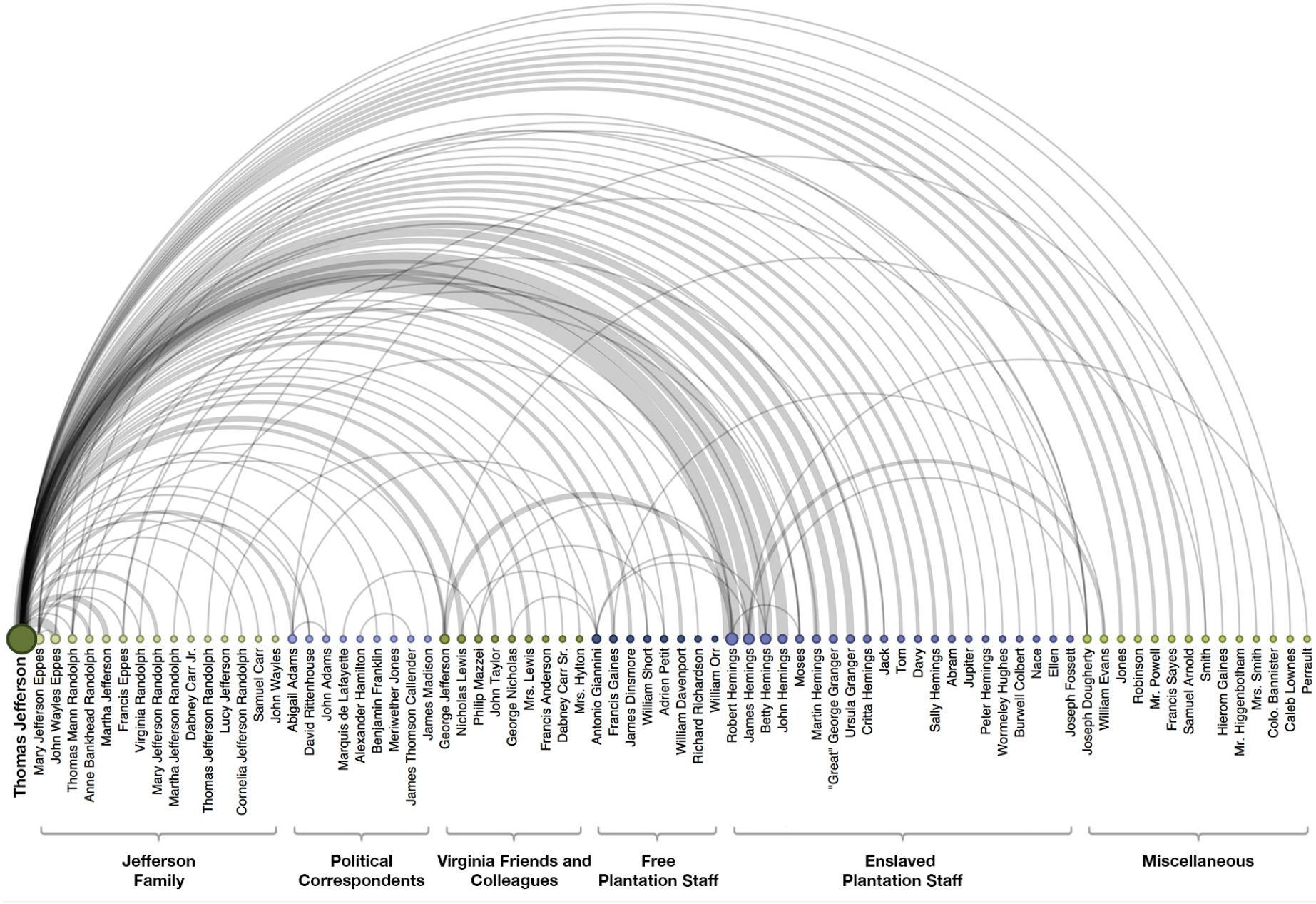
- Social networks are graphs visualizing the relationships between people.
- Used to illustrate the relationships between characters in literary texts.
- Size of a node can be used to encode the frequency of a character name in the text.

Distant reading techniques – Graphs

- Social networks are graphs visualizing the relationships between people.
- Used to illustrate the relationships between characters in literary texts.
- Size of a node can be used to encode the frequency of a character name in the text.
- the proximity of the nodes and the thickness of an edge can serve to reflect the strength of a relationship.



Excerpt from the social network in Mikhail Bulgakov's *Master and Margarita*

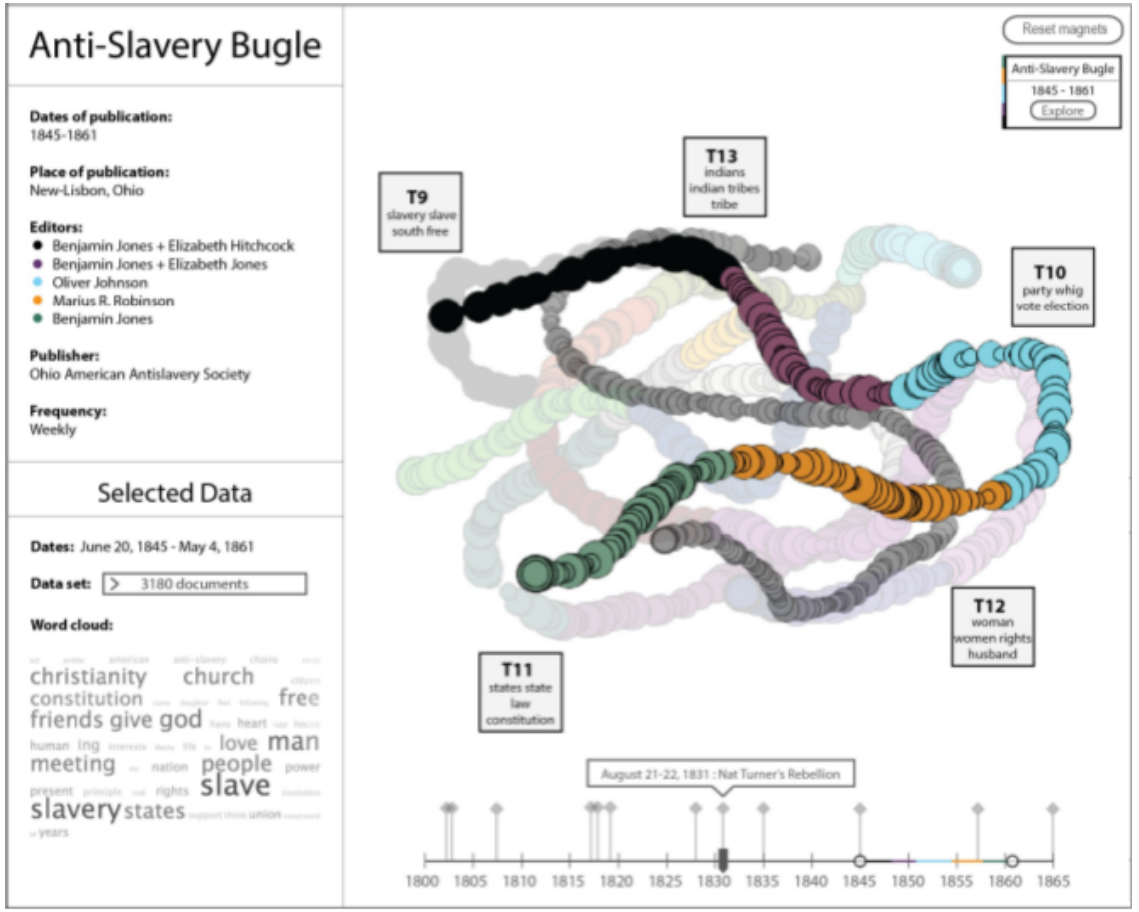


The visualization of Thomas Jefferson's social relationships, the nodes placed on a vertical axis are connected wit
arcs

Distant reading techniques – Miscellaneous

Distant reading techniques – Miscellaneous

- Dust-and-magnet for Thematic analysis of historical newspaper archives metaphor



You got this.

Combining close and distant reading

Combining close and distant reading

Top down -

Combining close and distant reading

Top down –

1. A distant view on the text is shown, and
   the user can manipulate the visualization by filtering or by zooming

Combining close and distant reading

Top down –

1. A distant view on the text is shown, and
   the user can manipulate the visualization by filtering or by zooming
2. Highlights certain patterns, can direct the reader

Combining close and distant reading

Top down –

1. A distant view on the text is shown, and
   the user can manipulate the visualization by filtering or by zooming
2. Highlights certain patterns, can direct the reader
3. Can retrieve details-on-demand by clicking on a potentially interesting data item.

Combining close and distant reading

Top down –

1. A distant view on the text is shown, and
   the user can manipulate the visualization by filtering or by zooming
2. Highlights certain patterns, can direct the reader
3. Can retrieve details-on-demand by clicking on a potentially interesting data item.

1) Heat map visualizes to show similarities between texts

Combining close and distant reading

Top down –

1. A distant view on the text is shown, and the user can manipulate the visualization by filtering or by zooming
2. Highlights certain patterns, can direct the reader
3. Can retrieve details-on-demand by clicking on a potentially interesting data item.

1) Heat map visualizes to show similarities between texts
2) Clicking on a grid cell opens a close reading view showing the corresponding two texts juxtaposed with connections between related text passages.

Combining close and distant reading

Bottom-up

Combining close and distant reading

Bottom-up

- Rarely applied.

Combining close and distant reading

Bottom-up

- Rarely applied.
- Major focus is the source text

Combining close and distant reading

Bottom-up

- Rarely applied.
- Major focus is the source text

1. The user selects a desired text passage in a text

Combining close and distant reading

Bottom-up

- Rarely applied.
- Major focus is the source text

1. The user selects a desired text passage in a text
2. Distant reading visualizations are processed based on that selection.

Combining close and distant reading

Bottom-up

- Rarely applied.
- Major focus is the source text

1. The user selects a desired text passage in a text
2. Distant reading visualizations are processed based on that selection.

1) The user selects a certain phrase in the text

Combining close and distant reading

Bottom-up

- Rarely applied.
- Major focus is the source text

1. The user selects a desired text passage in a text
2. Distant reading visualizations are processed based on that selection.

1) The user selects a certain phrase in the text
2) That phrase is searched within the text corpus and the phrase’s distribution is shown in the form of a heat map.

Combining close and distant reading

Top-down & bottom-up -

Combining close and distant reading

Top-down & bottom-up –

- Allow for switching between close and distant reading while taking into account manipulations of the preceding view.

Combining close and distant reading

Top-down & bottom-up –

- Allow for switching between close and distant reading while taking into account manipulations of the preceding view.
- Textual entities can be selected in both the graph (distant reading) and the text itself (close reading), triggering mutual updates.

Future challenges –

Future challenges –
Novel techniques for close reading

Future challenges – Novel techniques for close reading

- In most cases only simple visualization techniques, like color coding, are provided.

Future challenges – Novel techniques for close reading

- In most cases only simple visualization techniques, like color coding, are provided.
- Displaying information about textual features and structure alongside the source text.

Future challenges -
Visualizing transposition in parallel texts

Future challenges - Visualizing transposition in parallel texts

- Observing similarities and differences between diffrenet editions of a
text and detecting transpositions.

Future challenges - Visualizing transposition in parallel texts

- Observing similarities and differences between editions of a text and detecting transpositions.
- May occur on different text hierarchy levels.

Future challenges - Visualizing transposition in parallel texts

- Observing similarities and differences between editions of a text and detecting transpositions.
- May occur on different text hierarchy levels.
- A methods exist for the first two hierarchy levels (words, sentences)

Future challenges - Visualizing transposition in parallel texts

- Observing similarities and differences between editions of a text and detecting transpositions.
- May occur on different text hierarchy levels.
- A methods exist for the first two hierarchy levels (words, sentences)
- there’s a need for techniques to visualizing transpositions on all hierarchy levels by combining means of close and distant reading.

Future challenges – Geospatial uncertainty

Future challenges – Geospatial uncertainty

Many visualizations deal with extracting place names in texts to illustrate geographical knowledge.

Future challenges – Geospatial uncertainty

Many visualizations deal with extracting place names in texts to illustrate geographical knowledge.

- fictional place names which are hard to represent.

Future challenges – Geospatial uncertainty

Many visualizations deal with extracting place names in texts to illustrate geographical knowledge.

- fictional place names which are hard to represent.
- Place names can themselves carry uncertainty: “Sparta” and “Atlantis”

Future challenges – Geospatial uncertainty

Many visualizations deal with extracting place names in texts to illustrate geographical knowledge.

- fictional place names which are hard to represent.
- Place names can themselves carry uncertainty: “Sparta” and “Atlantis”
- Expressions like “in London” and “close to London”

That’s it.