



Digital Humanities and Gender History

Mandell, Laura C. (2021): Statistics versus Machine Learning: Gender, Intersectionality, and Big Data. In: Mettele, Gisela/Prell, Martin/Marzell, Pia (Hg.), Digital Humanities and Gender History. Jena. Online unter: https://www.db-thueringen.de/receive/dbt_mods_00049174.

Statistics versus Machine Learning:

Gender, Intersectionality, and Big Data

Laura Mandell
Director, CoDHR.tamu.edu
@mandellc

19 February 2021

Sample versus Case

Matthew L. Jockers

29
Thursday
SEP 2011

The LDA Buffet is Now Open; or, Latent Dirichlet Allocation for English Majors

POSTED BY MATTHEW JOCKERS IN TEXT-MINING

≈ COMMENTS OFF ON
THE LDA BUFFET IS
NOW OPEN; OR,
LATENT DIRICHLET
ALLOCATION FOR
ENGLISH MAJORS

For my forthcoming book, which includes a chapter on the uses of topic modeling in literary studies, I wrote the following vignette. It is my imperfect attempt at making the mathematical magic of LDA palatable to the average humanist. Imperfect, but hopefully more fun than plate notation. . .

. . . *imagine a quaint town, somewhere in New England perhaps. The town is a writer's retreat, a place they come in the summer months to seek inspiration. Melville is there, Hemingway, Joyce, and Jane Austen just fresh from across the pond. In this mythical town there is spot popular among the inhabitants; it is a little place called the "LDA Buffet." Sooner or later all the writers go there to find themes for their novels. . .*

◆ CONTACT



Matthew L. Jockers
Susan J. Rosowski Associate
Professor of English
Department of English
University of Nebraska-Lincoln
Lincoln, NE 68588
402/472-1896
email: mjockers{at}unl[dot]edu
Twitter: @mjockers

◆ RECENT COMMENTS

- Introduction to Digital Humanities | Intro to DH on 500 Themes

<http://www.smithsonianmag.com/smart-news/data-mining-the-classics-clusters-women-authors-together-puts-mellville-out-on-a-raft-16028354/>


SMARTNEWS *Keeping you current*

Data Mining the Classics Clusters Women Authors Together, Puts Mellville Out On a Raft

A new macroanalysis method compares thousands of books in order to identify systems of influence, school of thought or other groupings that human scholars might have missed

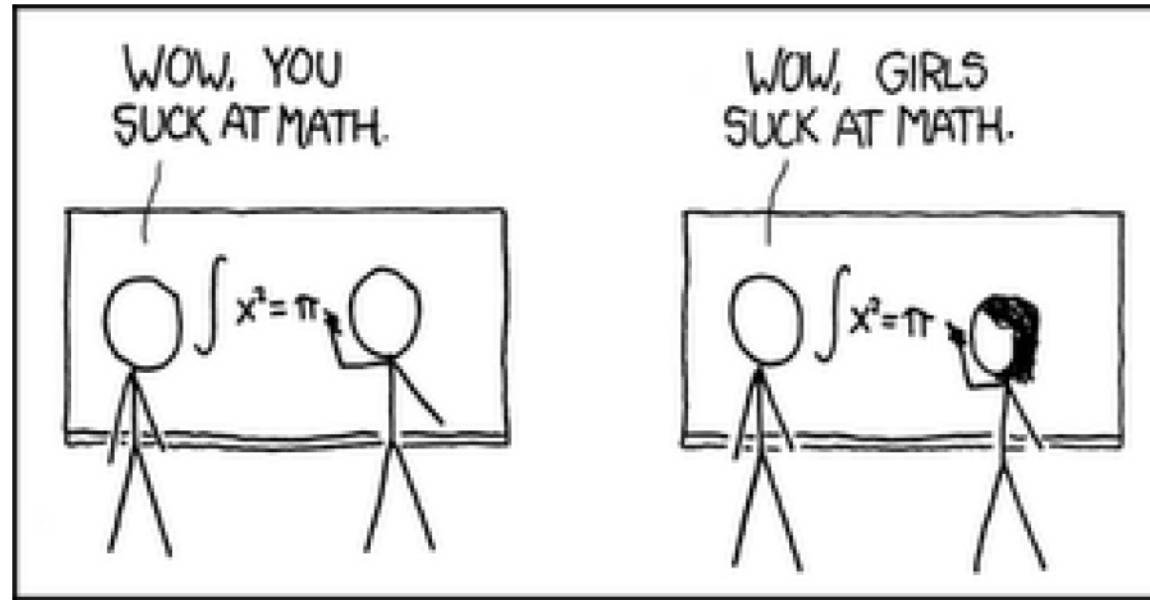
By [Rachel Nuwer](#)
SMITHSONIAN.COM
AUGUST 27, 2012

News release:

The novels of the period by gender. Female authors were grouped closely at one end of the book map, though their gender was not part of how they were evaluated. The purplish areas represent groups of women authors; indicating female authorship is objectively detectable. 

*““whale sea men ship whales penfon air side life bounty night
oil natives shark seas beard sailors hands harpoon mast top feet
arms teeth length voyage eye heart leviathan islanders flask soul
ships fishery sailor sharks company. . . “”*

*“marriage happiness daughter union fortune heart wife
consent affection wishes life attachment lover family promise
choice proposal hopes duty alliance affections feelings
engagement conduct sacrifice passion parents bride misery
reason fate letter mind resolution rank suit event object time
wealth ceremony opposition age refusal result determination
proposals. . .”*



<https://www.bitchmedia.org/post/are-girls-bad-at-math-are-we-still-having-this-discussion>

Gender and Cultural Analytics: Finding or Making Stereotypes?

Laura Mandell

<https://dhdebates.gc.cuny.edu/projects/debates-in-the-digital-humanities-2019>

Vive la différence: Tracing the (Authorial) Gender Signal by Multivariate Analysis of Word Frequencies

Jan Rybicki

Institute of English Studies, Jagiellonian University, Kraków, Poland

Abstract

Multivariate analysis of word frequencies is used to identify the gender of authors in a corpus of 18th- and early 19th-century English sentimentalist and Gothic fiction. Results obtained with most frequent words are compared to those produced with medium-frequency Burrows's Zeta words characteristic for both genders. Gender-sensitive words from two periods (18th/19th c. and 19th/20th c.) are compared in terms of their usefulness for gender identification in literary texts.

Correspondence:

Jan Rybicki, Institute of English Studies, Jagiellonian University, Al. Mickiewicza 9, 33-332, Kraków, Poland.

E-mail:

jkrybicki@gmail.com

Why are Jockers's findings not reproducible on a similar dataset?

1. Overfitting
2. Sample versus Case



The view that quantitative researchers look at many cases, while qualitative researchers look at only one or a small number of cases, can be maintained only by allowing considerable slippage in what is meant by “case.” . . . [M]uch of what is considered large-*N* research must also be seen as case-study research, and the tendency to conflate *qualitative study* and *case study* should be resisted. (3-4)

[D]igital humanities method . . . Consists in repeatedly coadjusting human concepts and machine technologies until the two stabilize each other in temporary postures of truth that neither by itself could sustain. (416)

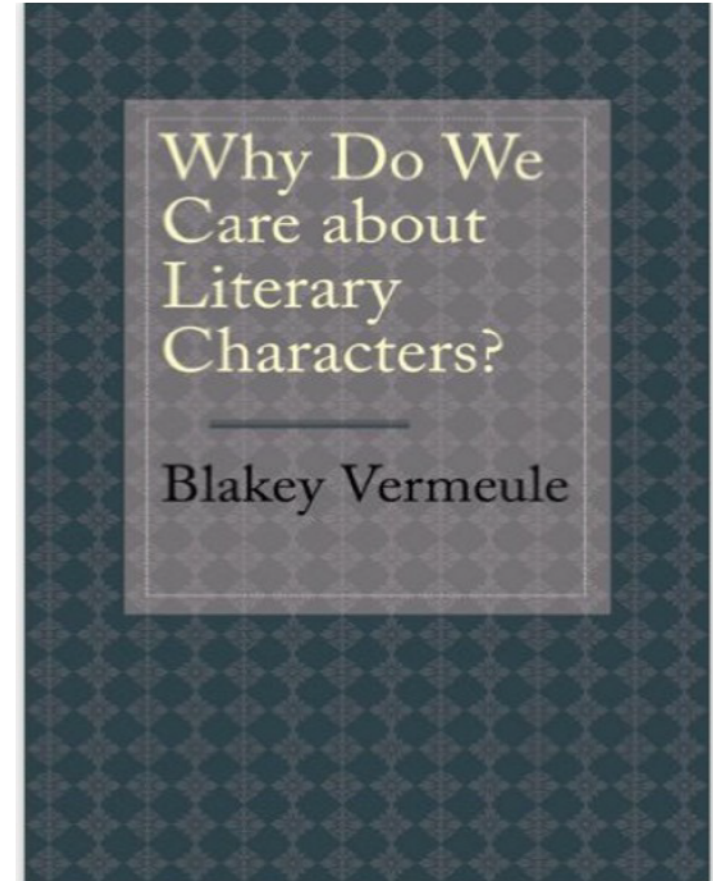
Alan Liu, "What is the Meaning of Digital Humanities to the Humanities?"
PMLA 2013

CHAPTER EIGHT

What's the Matter with Miss Bates?



Re: Jane Austen's *Emma*



In return, Miss Bates allows herself to be a comic butt, and what's worse, she has to thank everybody constantly for their favors—"so very obliged" is her mantra. How painful it must be to be "so very obliged" all the time—or at least so the narrator observes when she is talking about a different character: it is "a great deal better to choose than to be chosen, to excite gratitude than to feel it." Miss Bates has none of these choices—not only no "power of removal" and no "power of refusal" (two abilities Emma herself longs for at different points). She has no power to choose or to excite gratitude. Let me fancifully suggest that underneath all that babble, Miss Bates is secretly, silently suffering. After all, to be utterly dependent, to be always aware that one is a grade below one's patrons—in a society where such things are constantly measured and tacitly acknowledged—is the very antithesis of power and self-esteem. Mr. Knightley, the dominant male on the scene, is very nice about making people not feel dependent on him, but then he is an especially decorous and tactful person. This seems to be just what is required by social norms; Highbury is a place that frowns on excessive boasting and excessive displays. Like Miss Bates, Knightley has carefully adjusted his character to the niche of his circumstances. Had he been born in a different social organization, he might have been more overtly competitive—if, for example, he were a Northwest Coast T. Indian Chieftain, he would have been forced into ever more aggressive displays of gift-giving (e.g., at a potlatch, where, to mark his high status, an Indian chieftain throws food and goods on a fire).

apples came home, Mrs. Wallis sent them by her boy; they are extremely civil and **obliging** to us, the Wallises, always—I have heard some people say that Mrs. Wallis can be uncivil and give a very rude answer, but we have never known any thing but the greatest attention from them. And it cannot be for the value of our custom now, for what is our consumption of bread, you know? (Ibid., 220)

Miss Bates's voice is an aggressive screen, which blocks as much as it shows.

VOYANT

see through your text

Add Texts 🔍 🌙 ?

Type in one or more URLs on separate lines or paste in a full text.

Voyant Tools is a web-based reading and analysis environment for digital texts.

@mandellc

*“marriage happiness daughter union fortune heart wife
consent affection wishes life attachment lover family promise
choice proposal hopes duty alliance affections feelings
engagement conduct sacrifice passion parents bride misery
reason fate letter mind resolution rank suit event object time
wealth ceremony opposition age refusal result determination
proposals. . .”*
obliging”

Katherine Bode, 2021

Shifting to a conception of measurements as effects of material arrangements [as in Physics, reveals that] our knowledge-making practices contribute to and are part of the phenomena we describe.

Bode argues for a performative Computational Literary Studies

in which

“entities are not pre-existing but generated in an ongoing and emergent way by interactions, and those interactions are always historical, symbolic, and material simultaneously”

“Computational Modeling: From Data Representation to Performative Materiality”

Laura Nelson

10/02/2020; TAMIDS-CoDHR Seminar: [Leveraging the Alignment between Machine Learning and Intersectionality: Using Word Embeddings to Visualize Intersectional Experiences of the Nineteenth-Century U.S. South](#), [Laura Nelson](#), Northeastern University, [Presentation Slides](#), [Video](#).

Intersectionality and ML share epistemologies –

It can be used to uncover categories and patterns that we have not yet discovered.

Emerging Genders: The Feminist Controversy Project

<http://emerginggenders.dh.tamu.edu/>

Corpus Background

Feminist Controversy in England, 1788-1810

Garland Publishing, 1974

Editor: Gina Luria Walker

44 titles in 89 volumes

← → ↻ ⓘ www.18thconnect.org/typewright/documents 🔑 ☆ ⋮

[My18th](#)

[Search](#)

[TypeWright](#)

[Exhibits](#)

[News](#)

[HOME](#)

signed in as MeganNPearson | [log out](#) [What is 18thConnect?](#) | [Peer Review](#)

Welcome to TypeWright

TypeWright¹ is a tool for correcting the text-version of a document made up of page images. These text-versions are crucially necessary: they are what enables full-text searching, datamining, preserving, and curating² texts of historical importance. Right now, the text running behind the page images of these texts has been mechanically typed,³ leaving behind errors that need to be corrected by human eyes and hands.

Search for TypeWright enabled texts

[Advanced Search](#)

Featured Text

T H E	The winter's tale, a play. Alter'd from Shakespear, by Charles Marsh.
--------------	-----------------------------------------------------------------------

"Letters on education. With obse..."

abc Insert Above

abc Insert Below

MeganNPearson

OCR Source: gale

Document Home



first



prev

Page

93



next



last



53

our name, and so fallen on dull content, whose gay and splendid appearances promise a richer

possession.

How can happiness exist without pleasure, says Sensibility, and if pleasure is the ground work of happiness, says Reason, why certainly the more pleasure, the more happiness, and thus by a rational deduction, we come to the conclusion, that pleasure and the means to obtain it, is the only wish of man. Your argument would be very just, replies Experience, were our opportunity always within our reach, and were the powers of the human frame and all its faculties, as capable of creating and lengthening enjoyment, as are the beautiful desires of man. But as this is not so, in Nature, as there are limits to which the intense pleasures soon arrive, and from which they ever afterwards decline, they are by necessity of their duration; and if we endeavour to suspend for such limitations by the frequency of repetition, we shall lose more than we gain by the fatigue of our faculties, and the diminution of sensibility.

Now it is the foolish voluptuary only who has no remorse in the depth of his faculties, the impetuosity of desires which can never be gratified, and the necessity of pleasures which will return no more. No; the mannikin whose enjoyment depends more on the delights, which are adapted to soothe his imagination, on the variety of his amusements

possession.

How can happiness exist without pleasure, says Sensibility; and if pleasure is the ground work of

2 gay and splendid appearances promise a richer

3 possession.



4 How can happiness exist without pleasure, says

report this page

resize red box

Document Home



first



prev

Page

93



next



last





Table of Contents

1. Global Cross-Reference

2. Terms and Details

[owl:ObjectProperty \(173\)](#)

[owl:Class \(55\)](#)

[owl:AnnotationProperty \(1\)](#)

[owl:Class Instances](#)

[cwrc:Certainty \(4\)](#)

[cwrc:EducationalAward \(2\)](#)

[cwrc:Ethnicity \(29\)](#)

[cwrc:Gender \(11\)](#)

[cwrc:GeographicHeritage \(2\)](#)

[cwrc:NationalHeritage \(14\)](#)

The CWRC Ontology Specification

0.99.75



THE CANADIAN WRITING RESEARCH COLLABORATORY

CWRC / CSÉC

LE COLLABORATOIRE SCIENTIFIQUE DES ÉCRITS DU CANADA

The CWRC Ontology is the ontology of the Canadian Writing Research Collaboratory.

gender terms

androgynou
S

cisgender

cisman

ciswoman

man

transgender

transman

transwoman

woman



CWRC Ontologies

gender

A subclass of cultural form for indicating a person's gender, whether attributed or self-defined, with accompanying context, where present, provided by Gender Context annotations. Although in popular culture gender and biological sex are conflated and understood to be binary, the concept of gender stresses the relationality, constructedness, and performativity of gendered identities and gendered behaviour, whose categories are historically contingent and shifting, and the boundaries between them blurry. Simone de Beauvoir (1973) stated "One is not born, but rather becomes, a woman," a belief that has been taken up by many other theorists of gender including Judith Butler (1990). Gender studies to date has investigated the social construction of femininity more than that of masculinity. The multi-layered constructedness of gender is also underlined by Donna Haraway who understands it as a "socially, historically, and semiotically positioned difference" (Donna Haraway, 1988). Gender is understood as fluid, situational, and sometimes plural, and it is related to, though not commensurate with, sexual identity and orientation, just as it is related to but not defined by specific forms of embodiment. Rather than seeing biological sex as a pre-social or natural given, the body is understood as a site of inscription (cf. (Elizabeth Grosz, 1994)) which is also socially constructed and indeed



CWRC Ontologies

woman

Indicates a subject identification with or labelling as the gender woman, and who is often but not necessarily understood to be sexed female. "One is not born, but rather becomes, woman" (Simone de Beauvoir, 1973)

URI:

<http://sparql.cwrc.ca/testing/cwrc#woman>

Tier 1 Tagging Taxonomy

Personality

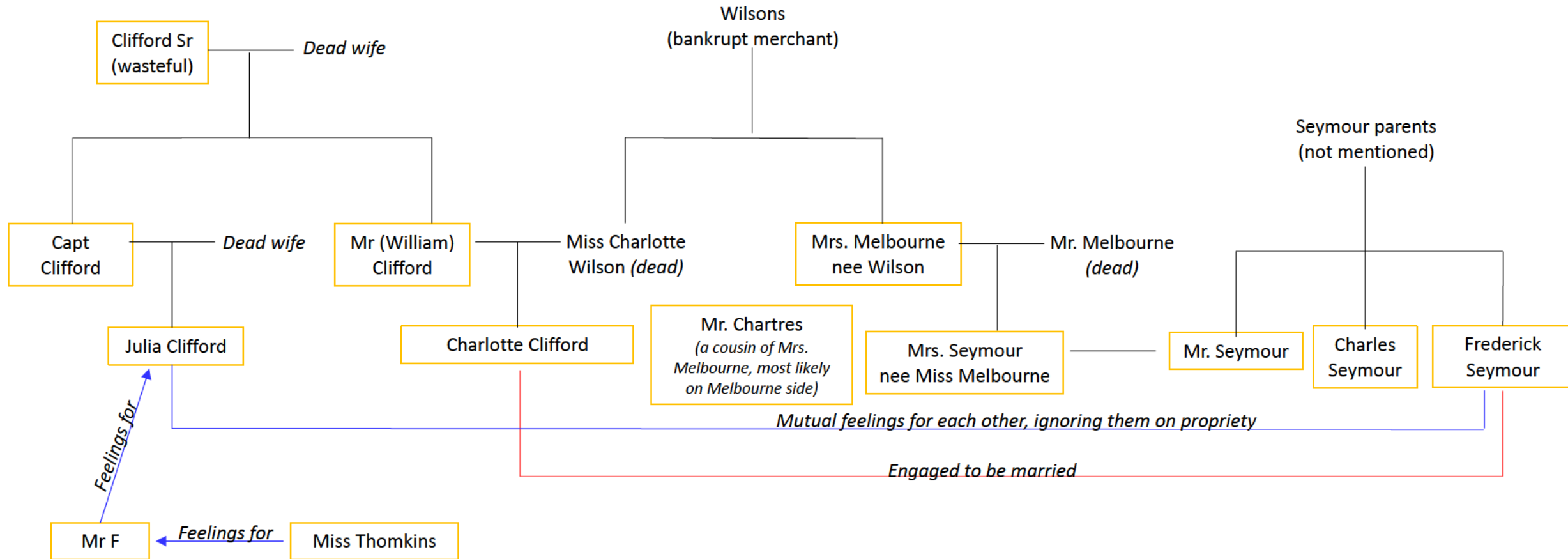
financial status
friendship
talents
social status
education
occupation
intellect
habitual manner
temperament
morality
physical characteristics
partnered (legal)
emotion/sensibility
religious
race (non-white or
ambiguous)
fortune
responsibility

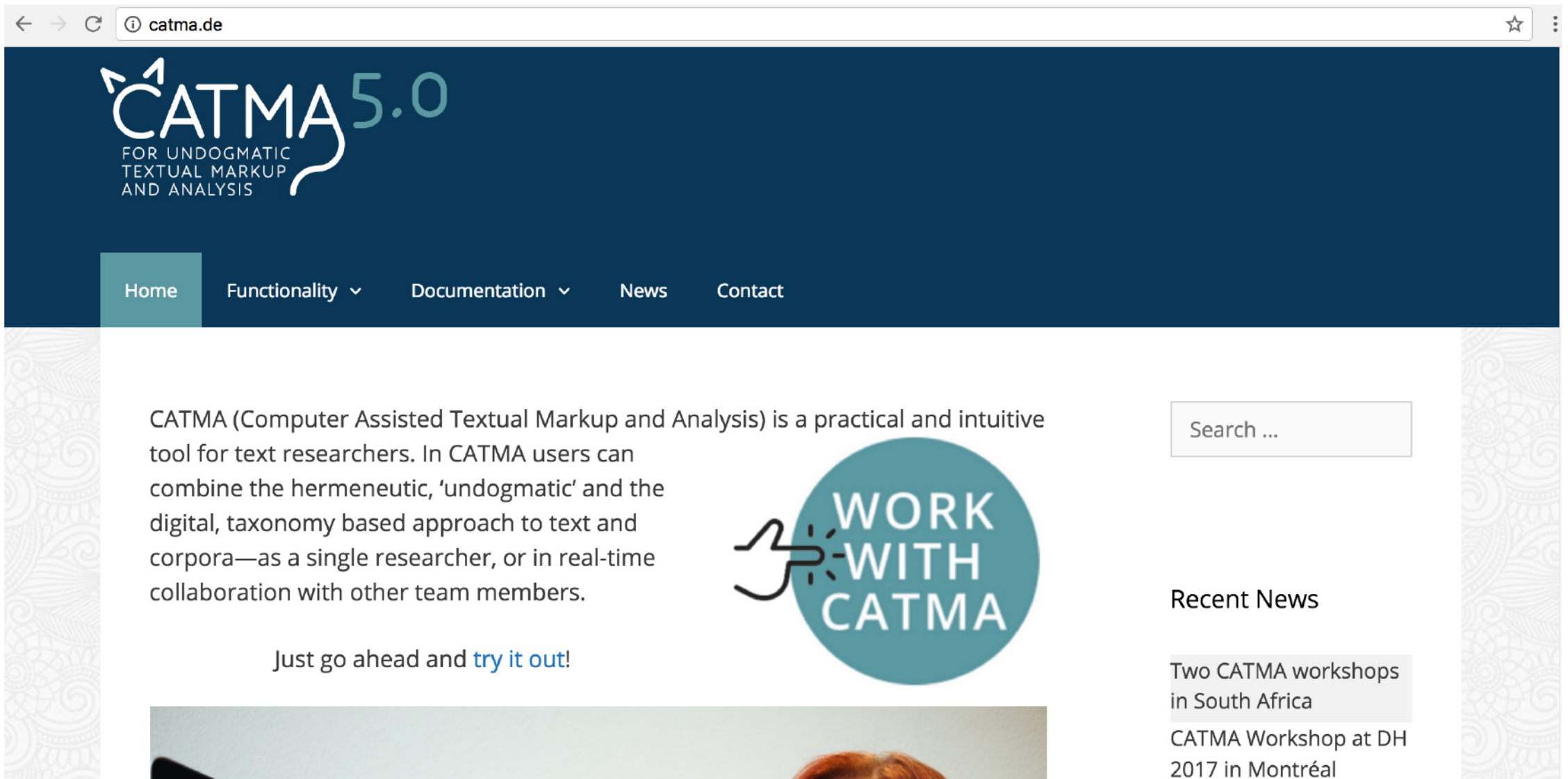
Activities

purchasing or not
playing (leisure)
arguing or not
being alone or not
outdoor activity
educating or not
insulting or not
reading or not
threatening or not
committing a crime
drinking/smoking/etc (vices) or not (virtues)
recognizing or not
injuring or not
creating or destroying
illness or not
seeing or not
moving or not
inheriting or not
direct address
singing/playing an instrument
indoor activity
possessing (material) or not

loving or hating
suffering or not
being deprived or gaining (a person)
learning or not
traveling or not
giving or taking
working or not
writing or not
communicating or not
caretaking or not
finding or losing (possessions)
exhibiting or hiding
knowing or not
emoting or not
fantasizing/imagining or not
thinking/contemplating or not
competing or not
feeling (internal)
enjoying or not
dressing (toiletry)
listening/hearing or not
rebellious or conforming

Julia, A Novel Vol. 1 Character Diagram for Tier 2 Tagging Taxonomy





The screenshot shows a web browser window with the address bar displaying "catma.de". The page has a dark blue header with the CATMA 5.0 logo and a navigation menu. The main content area is white with a decorative pattern on the sides. It features a paragraph about CATMA, a "WORK WITH CATMA" graphic, a search bar, and a "Recent News" section with a link to a workshop in Montréal.

← → ↻ catma.de ☆ ☰


CATMA 5.0

FOR UNDOGMATIC
TEXTUAL MARKUP
AND ANALYSIS

Home Functionality ▾ Documentation ▾ News Contact

CATMA (Computer Assisted Textual Markup and Analysis) is a practical and intuitive tool for text researchers. In CATMA users can combine the hermeneutic, 'undogmatic' and the digital, taxonomy based approach to text and corpora—as a single researcher, or in real-time collaboration with other team members.

Just go ahead and [try it out!](#)



WORK
WITH
CATMA

Search ...

Recent News

Two CATMA workshops in South Africa

CATMA Workshop at DH 2017 in Montréal

Example of Tagged Text in CATMA: Julia a Novel



Manage Resources

Manage Tags

Annotate

Analyze

Visualize

[About](#) [Terms Of Use](#)

[Imprint](#) [Privacy Statement](#)

[Manual](#) [?](#)

[mpearson42@tam](#)
[- Sign out](#)

Julia, a Novel Vol 1 x

Captain Clifford's younger brother, after the paternal estate was disposed of, went in pursuit of fortune

to the East Indies. ~~he was a man of a plain understanding and an excellent heart. Just in his principles,~~

and generous in his disposition, he acquired wealth slowly, but honourably. Mr. Clifford married at

Bengal, and his only daughter, Charlotte, was sent when a child to England for education, and committed

to the care of her aunt Mrs. Melbourne, the sister of Charlotte's mother.—At eighteen Charlotte was taken

from school at Queen Square, to live with her aunt, till the return of her father from the East Indies.

Charlotte was one of those sweet lively characters, whose unaffected manners and invariable good-humour

strongly engage the affections, and with whom one would wish to pass thro' life. The gay powers of wit

Active Tagsets

Active Annotations

Open Annotations

Close



Select all visible

Deselect all visible

Annotations	Tag color	Visible	Writable
▼ Julia a Novel Vol 1		<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
▶ ↕ characters		<input checked="" type="checkbox"/>	
▶ ↕ personality		<input checked="" type="checkbox"/>	
▶ ↕ activities		<input checked="" type="checkbox"/>	

Writable Annotation Collection: Julia a Novel Vol 1

Annotation	Colr
↕ MrClifford	■
↕ financial status	■
↕ traveling or not	■

Remove Annotation

Edit Property values

Annotation Info



1 / 1



Analyze Document

1



100



100 % page size

Mary, a Fiction x

temperament

reading

fantasizing/imagining

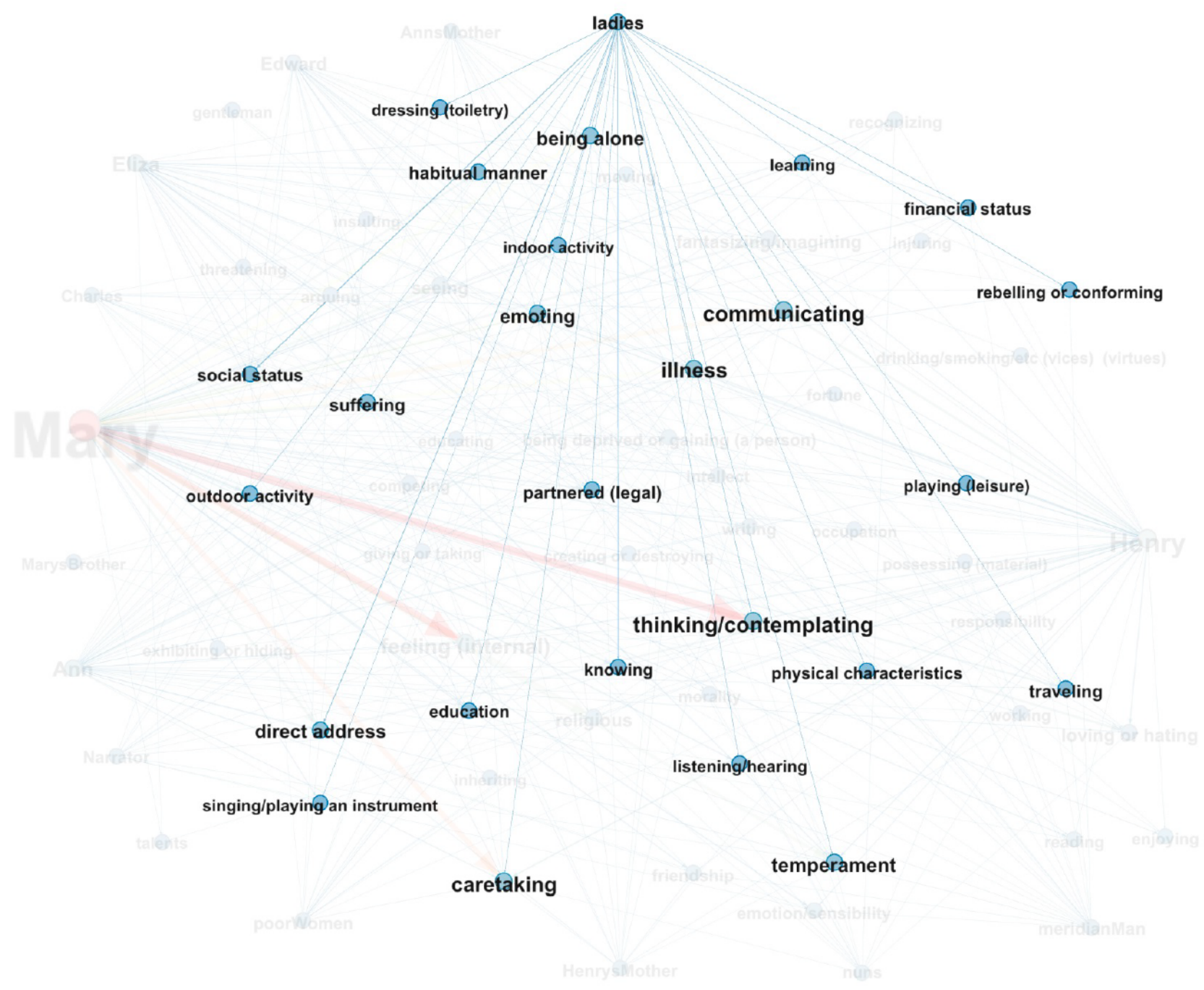
thinking/contemplating

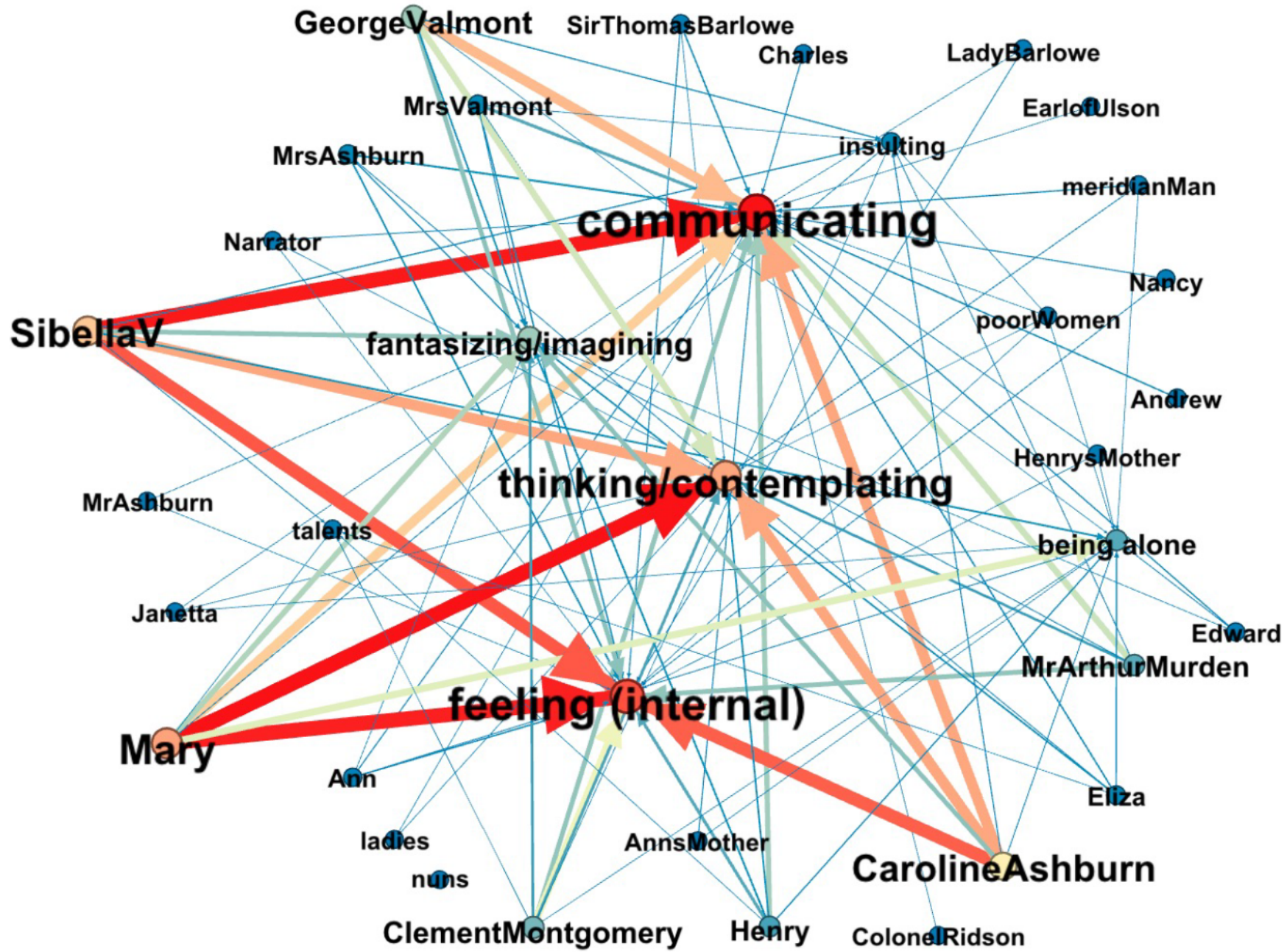
narrator

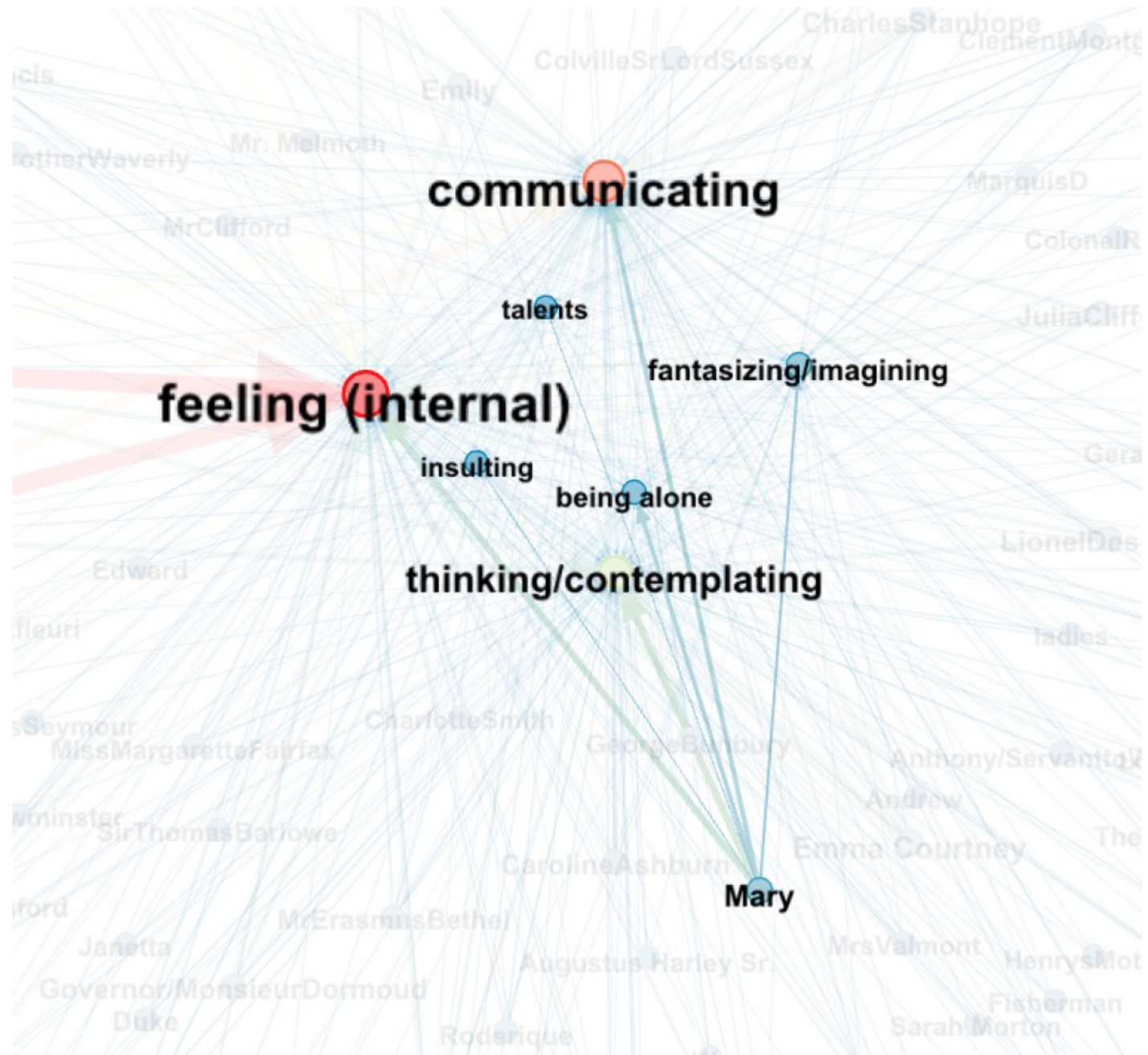
failing to communicate

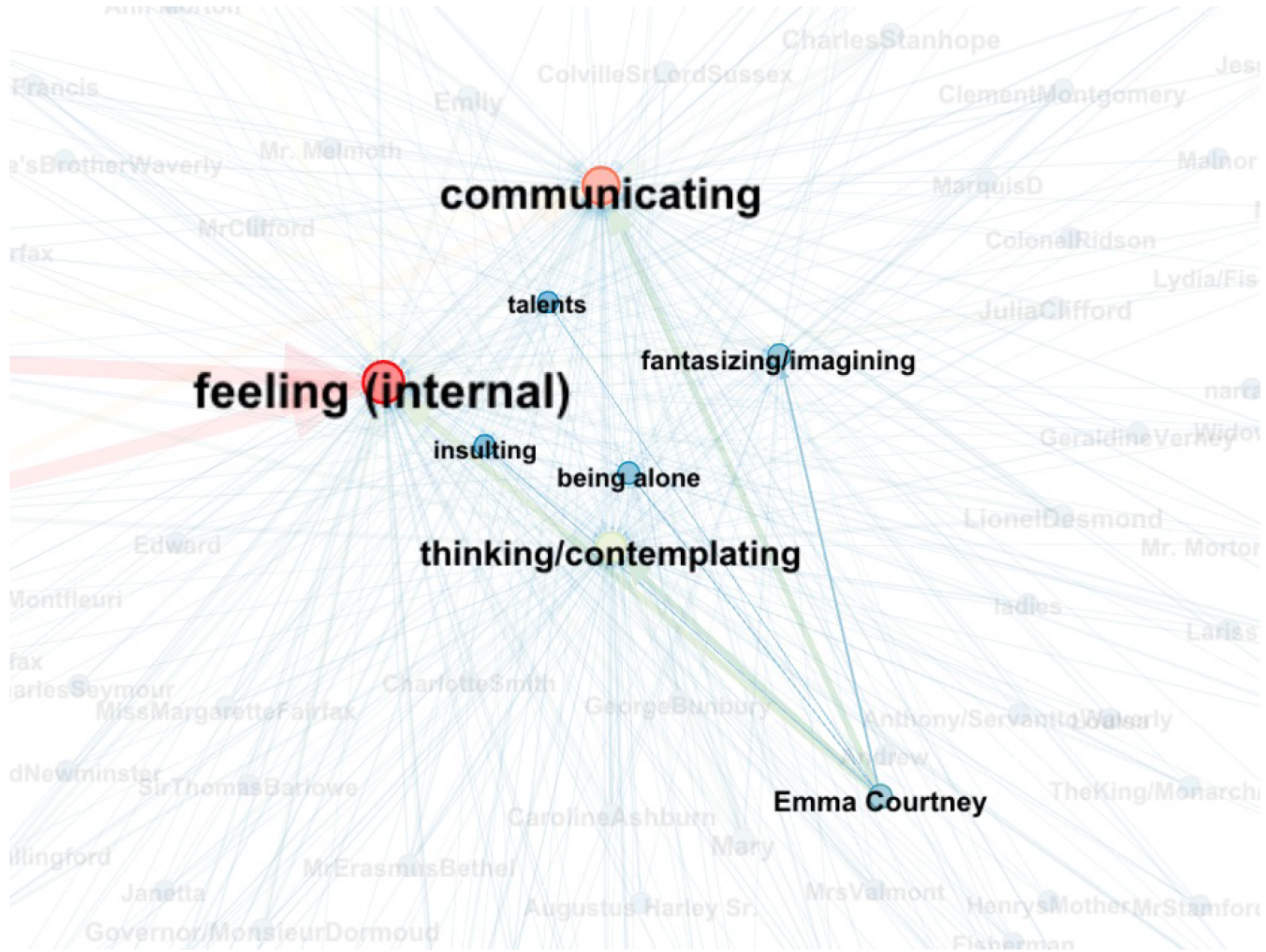
communicating (gestures)

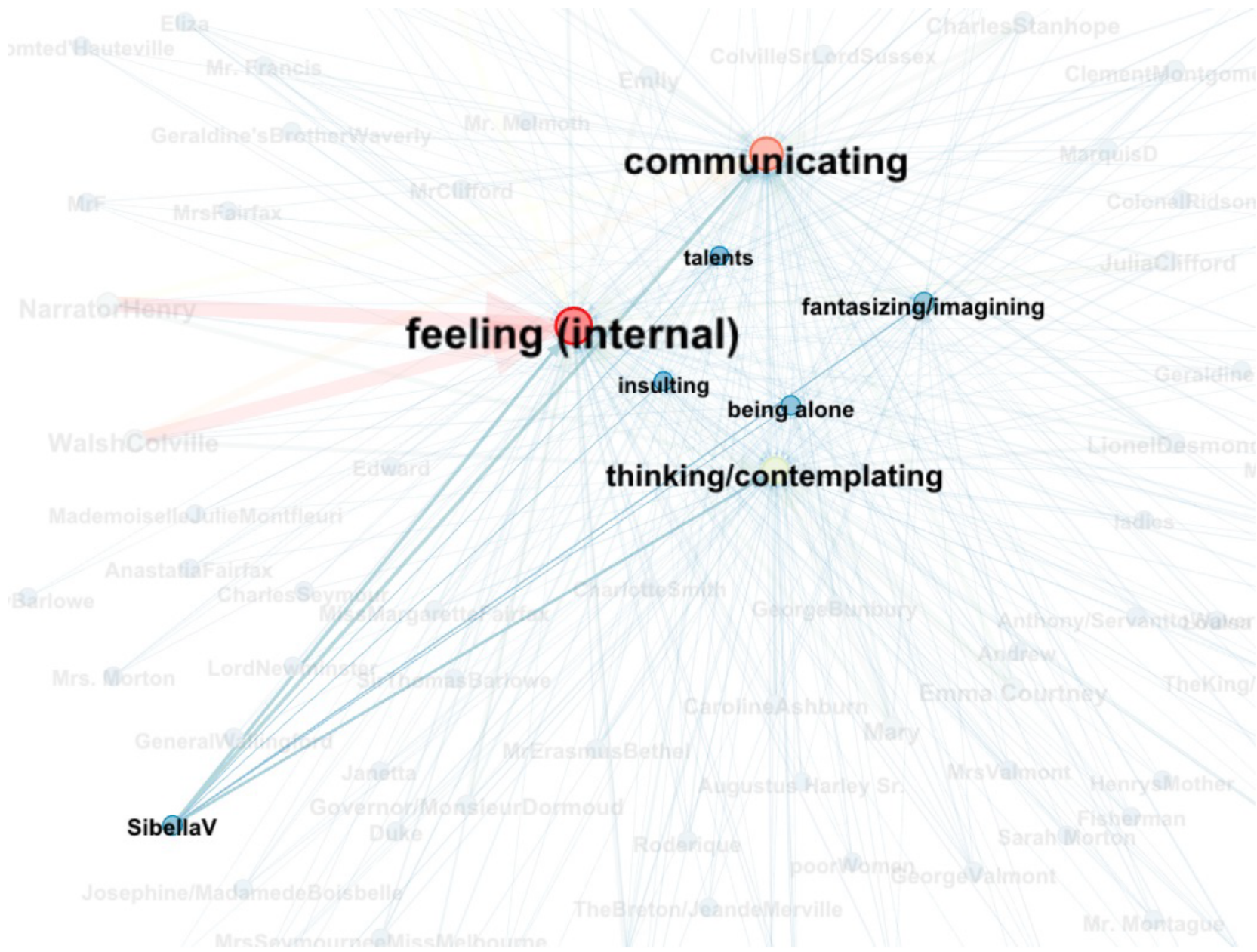
She was chaste, according to the vulgar acceptation of the word, that is, she did not make any actual faux pas; she feared the world, and was indolent; but then, to make amends for this seeming self-denial, she read all the sentimental novels, dwelt on the love-scenes, and, had she thought while she read, her mind would have been contaminated; as she accompanied the lovers to the lonely arbors, and would walk with them by the clear light of the moon. She wondered her husband did not stay at home. She was jealous—why did he not love her, sit by her side, squeeze her hand, and look unutterable things? Gentle reader, I will tell thee; they neither of them felt what they could not utter. I will not pretend to say that they always annexed an idea to a word; but they had none of those feelings which are not easily analyzed.

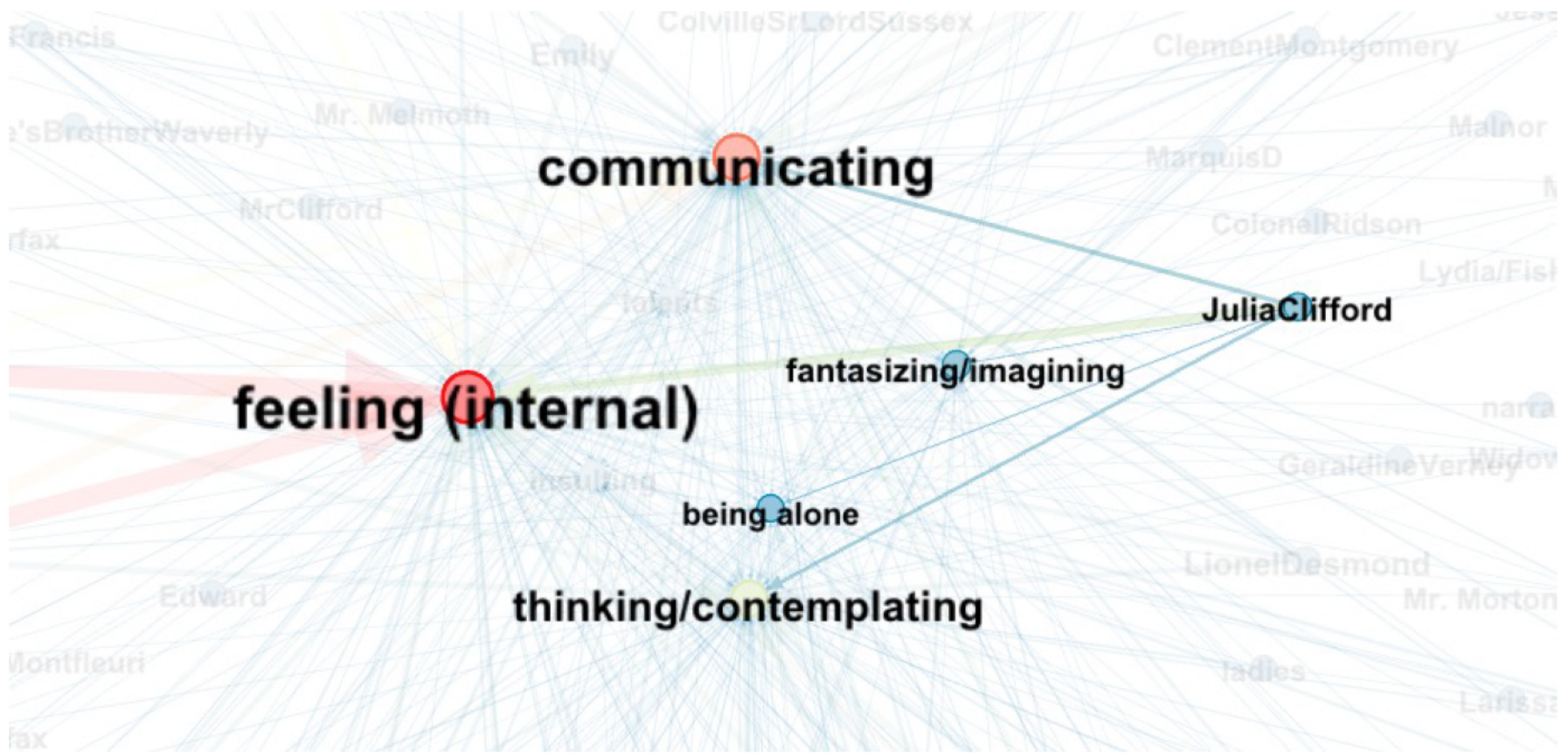












feeling (internal)

communicating

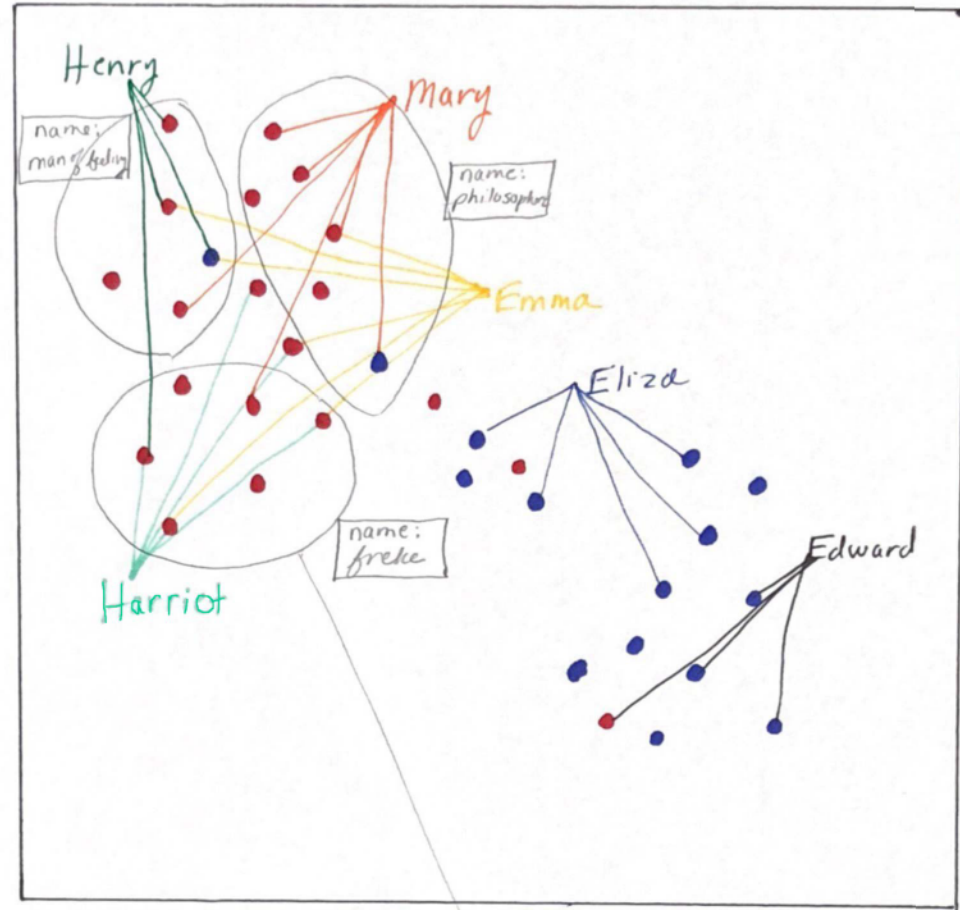
Julia Clifford

fantasizing/imagining

being alone

thinking/contemplating

Network View



- Gender Normative
- Different

user groups traits together and names "genders"

NOVELS

Secresy

Walsh Colville

Julia Volume I (selected)

Mary

**Created by Bryan
Tarpley
@halcyonflies**

- friendship
- intellect
- friendship
- recognizing
- thinking/contempla

CLASSIFY ✕

Not classified
 Gender Normative
 Different

New label:

thinking/contemplating

Detail:

None

Assigned to character:

Augustus Harley Jr. ▾

Tagged text:

ong train of patient and laborious experiments must precede our deductions and conclusions. The science of mind is not less demonstrative, and far more

Close Save changes

CHARACTERS

- Augustus Harley Jr.
- Mr. Francis
- Mr. Melmoth
- Ann Morton
- Emma Courtney
- Mr. Courtney (Emma's dad)
- Mr. Montague
- Mr. Morton (Emma's uncle)
- Widow Harley
- Mrs. Denbeigh (cousin)
- Mrs. Morton
- Sarah Morton
- Joanna
- Augustus Harley Sr.
- Mrs. Melmoth

MR. FRANCIS

- direct address

NLP processing

Machine Learning

What is the relationship between literature and reality?

what will we “say” about the research? What does it mean?