Getting inside the letter: Using visualisation tools to explore the language of emigrant correspondence

Workshop on visualizing letter collections, Omagh 14-15 March, 2014

Chris Culy

Why visualizations?

Visualizations put ideas into our heads …

DoubleTreeJS + KWIC

DoubleTreeJS + KWIC

Input file: Browse... LOUGH_plain.all.txt Word to use: can

Sort branches by: token POS branching

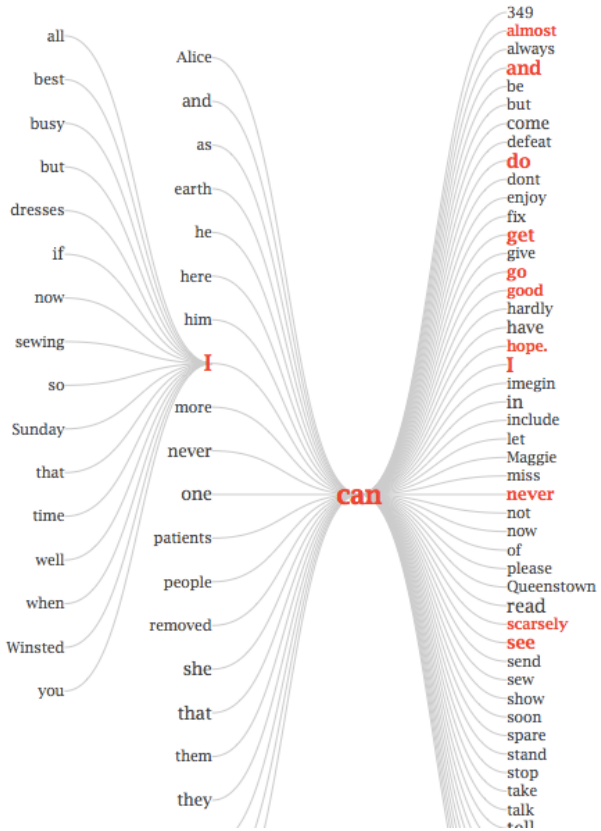
Limit first left POS to:

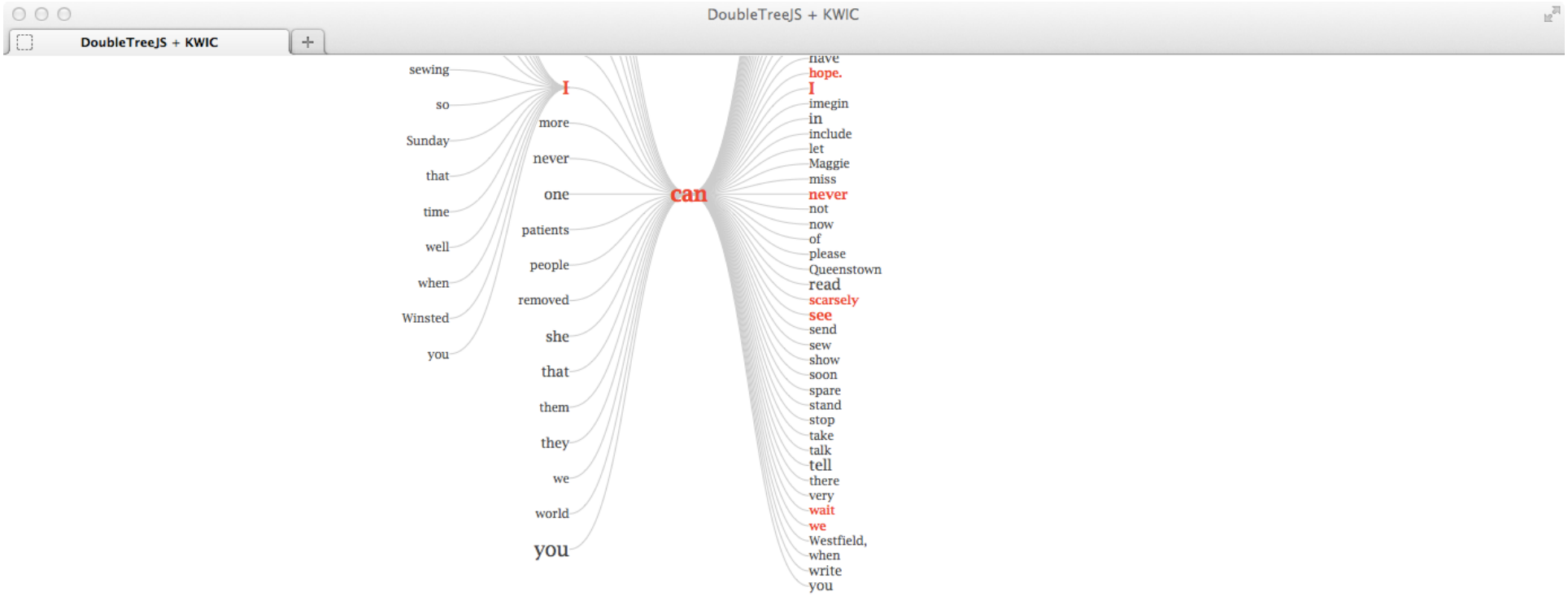
Limit first right POS to: Update filters

Find in tree: [aA] Clear Show POS

Make image

DoubleTreeJS + KWIC, © 2012-2013 Chris Culy



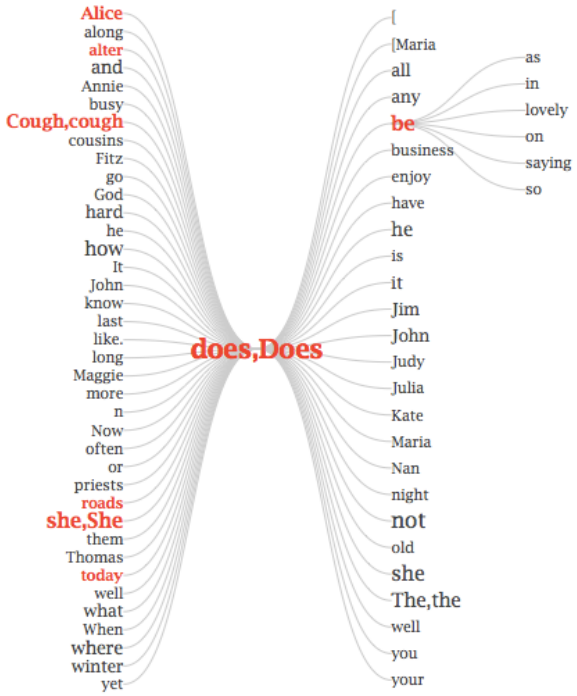


my picture in my next letter if I	**can**	I Wish you would try an have some
have a picture taken from yours if I	**can**	I am goeing to have a large one
make Mary and yourself as happy as you	**can**	I am sending the order in this letter
that she will write to me when she	**can**	I am very glad to hear you will
you are trying to do the best you	**can**	I think of you so often I hope
take very good care of yourself if you	**can**	I hope the girls are all well and
you are trying to do the best you	**can**	I did not write to you at the
you must try and Keep well if you	**can**	I Know you never can stop thinking of

DoubleTreeJS + KWIC

DoubleTreeJS + KWIC

Make image



with her she told me to day she **does** be in Ireland every night in her dreams

I do have good fun with them Alice **does** be saying my dear Auntie she calls Annie

very well this winter and if your Cough **does** be as bad as usual and are you

have them all through the year the alter **does** be lovely today and Easter Sunday also evryone

cannot very come out in winter the roads **does** be so bad her two girls Jennie and

about [OBreens?] marriage what happens in Ireland today **does** be on the papers tomorrow. It [says?] the

Some other visualizations: letters

- Letters
  - **Mapping the Republic of Letters**
    – NB: metadata only
  - **Siirtola et al 2011** on Early English letters
    – NB: sociolinguistic variables, but letters do not form a *correspondence*
  - The *Vistola* Project — Visualization and analysis of person-oriented correspondence: **http://www.sfs.uni-tuebingen.de/~cculy/vistola/**
  - Others?

Some other visualizations: email

Email is a type of correspondence!

- **U. Maryland workshop** 2005, (just paper titles)
- **Enron** (Jeffrey Heer, presented at workshop above)
- By Fernanda Viégas
  - **themail**: Words by person
  - **PostHistory**: Metadata only
  - **Mountain**: Contacts only

Source: **Viégas on email**

Thinking about the data

- What data do we have?
- How can we analyse that data?
- What kinds of questions do we want to be able to answer? – How should we *model* that data?

Thinking about the visualization

- What are we trying to do with the visualization?
  - Explore
  - Comprehend (analyze)
  - Communicate (present)
- How should we visualize the data and its analyses?

Data: Lough Letters

- 99 Letters, from 6 authors (4 sisters), plus "Unknown"
- Taken as transcribed
- Additional automatic analysis
  – Lemmas, part of speech tags using TreeTagger
  – Phrases in common using my own program
  – "Topics" using Mallet

Sources: Moreton 2012,
TreeTagger: **http://www.cis.uni-muenchen.de/~schmid/tools/TreeTagger/**
Mallet: **http://mallet.cs.umass.edu/**

Data: types of questions

– Language usage in general

DoubleTreeJS + KWIC

DoubleTreeJS + KWIC

Input file: Browse... LOUG01_tagged_from.all.txt Word to use: am_VBP

Sort branches by: token POS branching

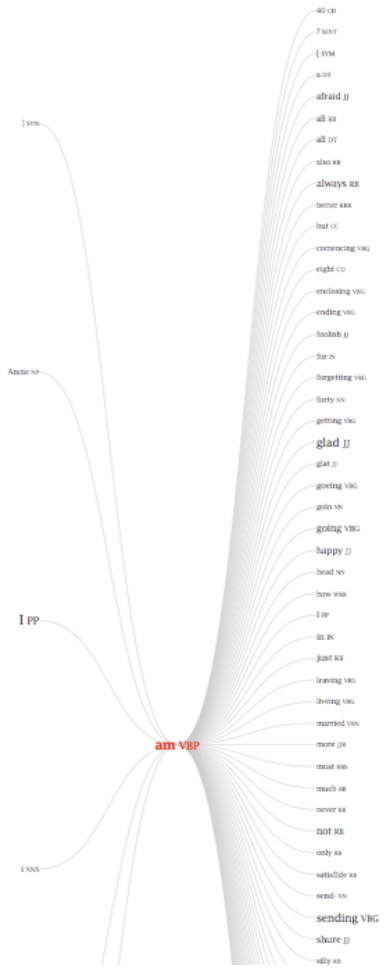
Limit first left POS to:

Limit first right POS to: Update filters

Find in tree: [aA] Clear Show POS

Make image

DoubleTreeJS + KWIC, © 2012-2013 Chris Culy



DoubleTreeJS + KWIC

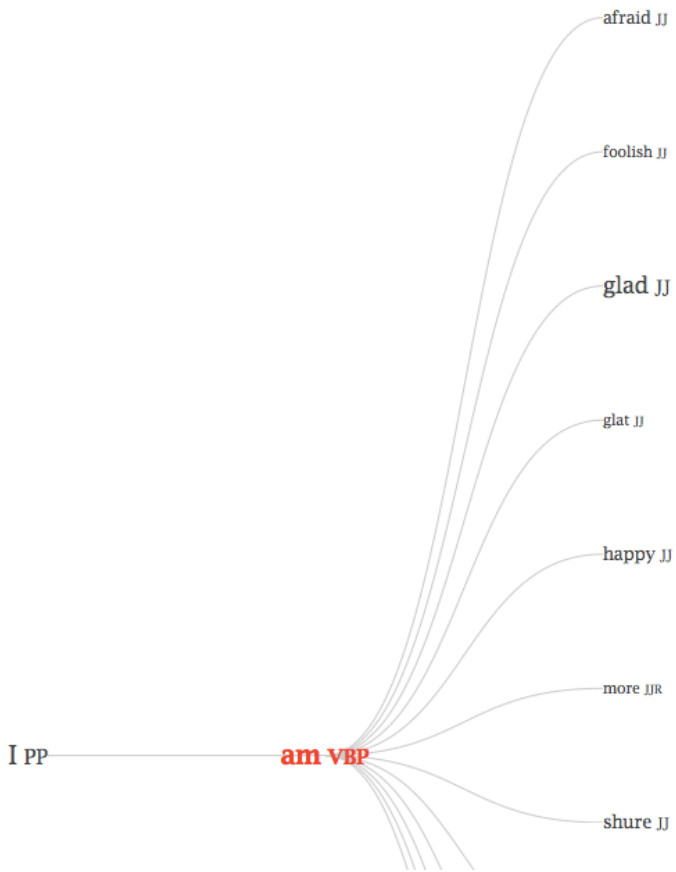
DoubleTreeJS + KWIC

Limit first left POS to: PP

Limit first right POS to: JJ Update filters

Find in tree: [aA] Clear Show POS

Make image



DoubleTreeJS + Filtered KWIC

DoubleTreeJS + Filtered KWIC

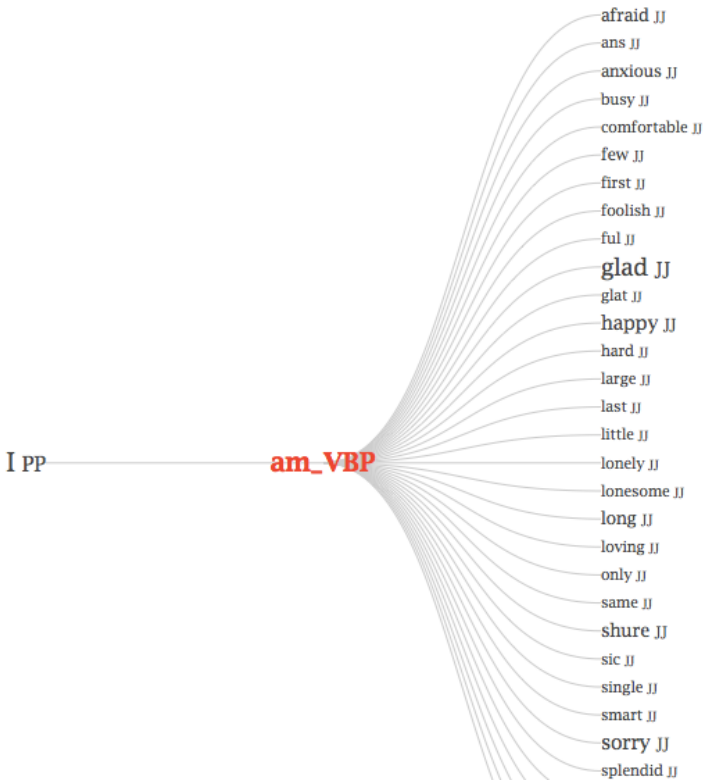
Context: ○ Full context
◉ Include only POS: PP,JJ
○ Exclude all POS:
Update context

Limit first left POS to: PP

Limit first right POS to: JJ Update filters

Find in tree: [aA] Clear ☑ Show POS

Make image



DoubleTreeJS comparison demo

DoubleTreeJS comparison demo

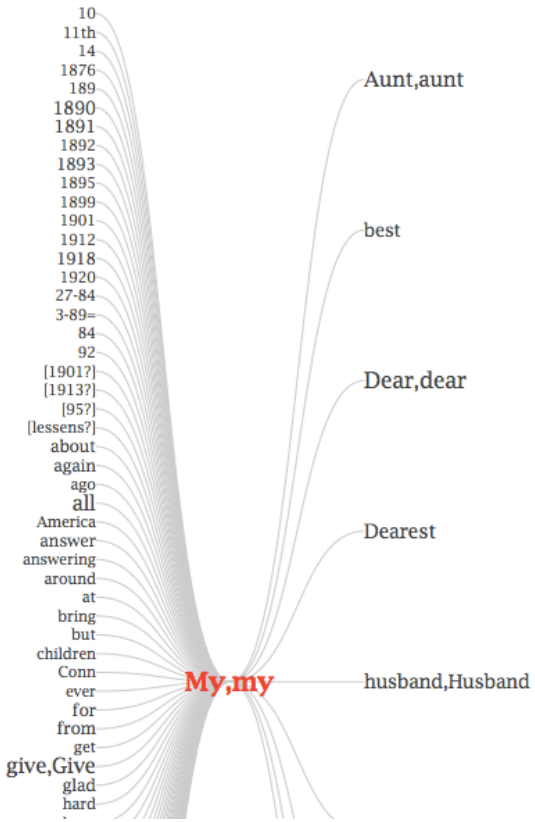
DoubleTreeJS comparison, © 2012-2013 Chris Culy

Input file: Browse... LOUGH_plain.all.txt

Word: my Find in tree: [aA] Find! Clear

Sort branches by: token

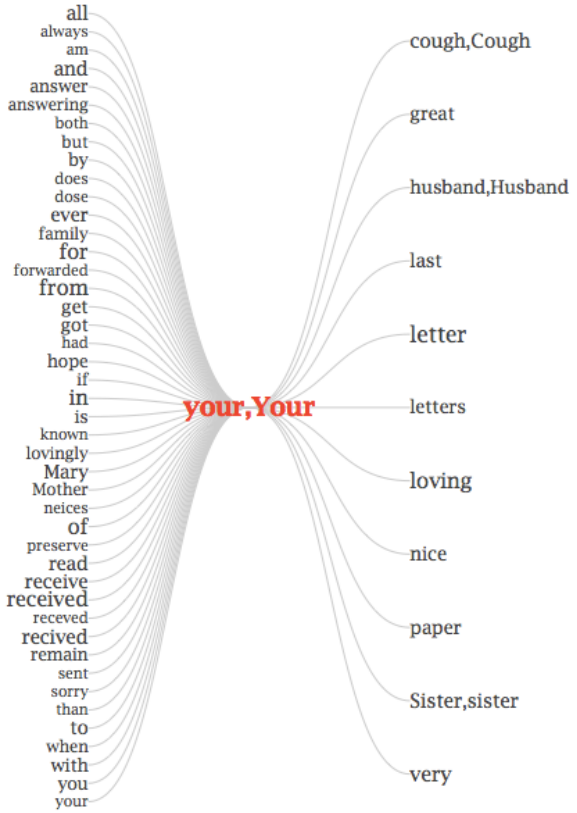
First left count ≥: First right count ≥: 5 Update filters



Word: your Find in tree: [aA] Find! Clear

Sort branches by: token

First left count ≥: First right count ≥: 5 Update filters



Comparative Contexts

Comparative Contexts

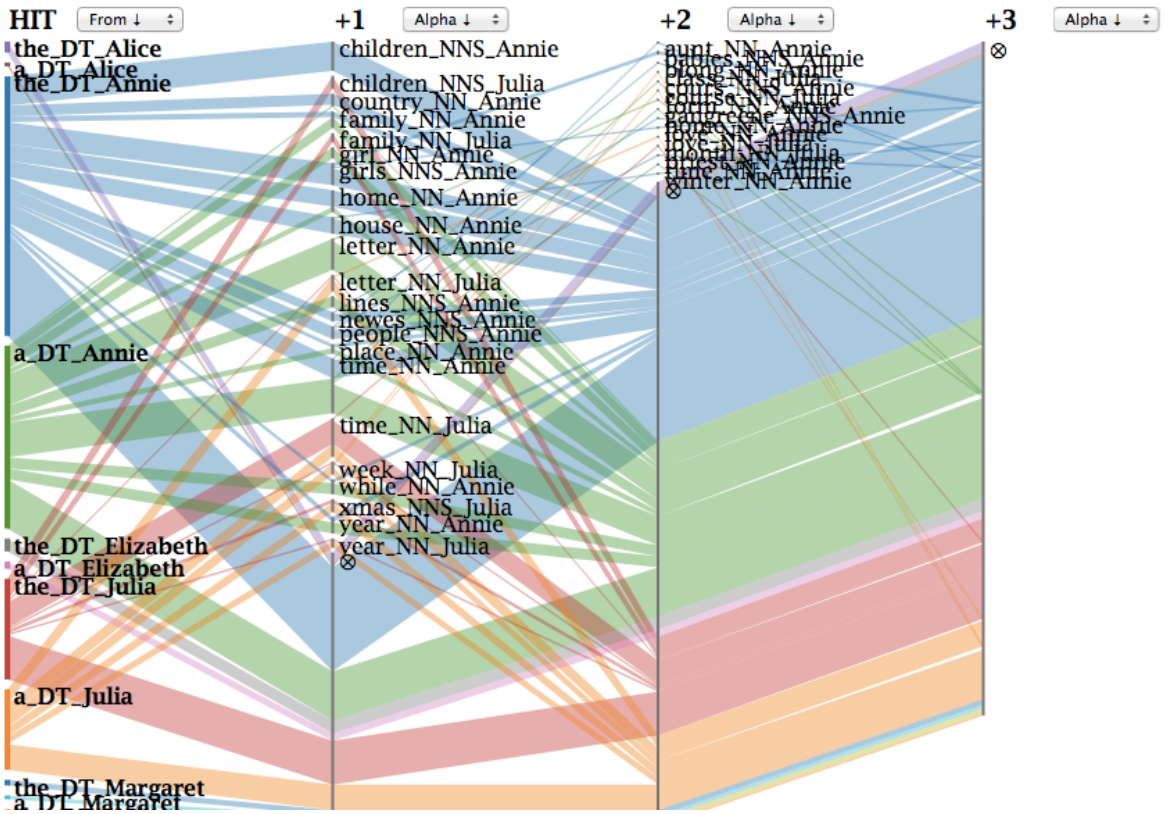
Comparative Contexts

A visualisation technique for *comparing textual contexts*. Works best with up to c. 1000 results. By Chris Culy

Use this example to search (including using regular expressions) for items in a text (e.g. the Robin Hood example in the texts/ folder) and visualize the results. More details are below.

Configure | Showing all, including only: NN,NNS in (SENT...SENT) using at most 1000 random hits.

Input file: Browse... LOUGH_tagged_from.all.txt Search for something: ((a)|(the))_DT Showing 369 results

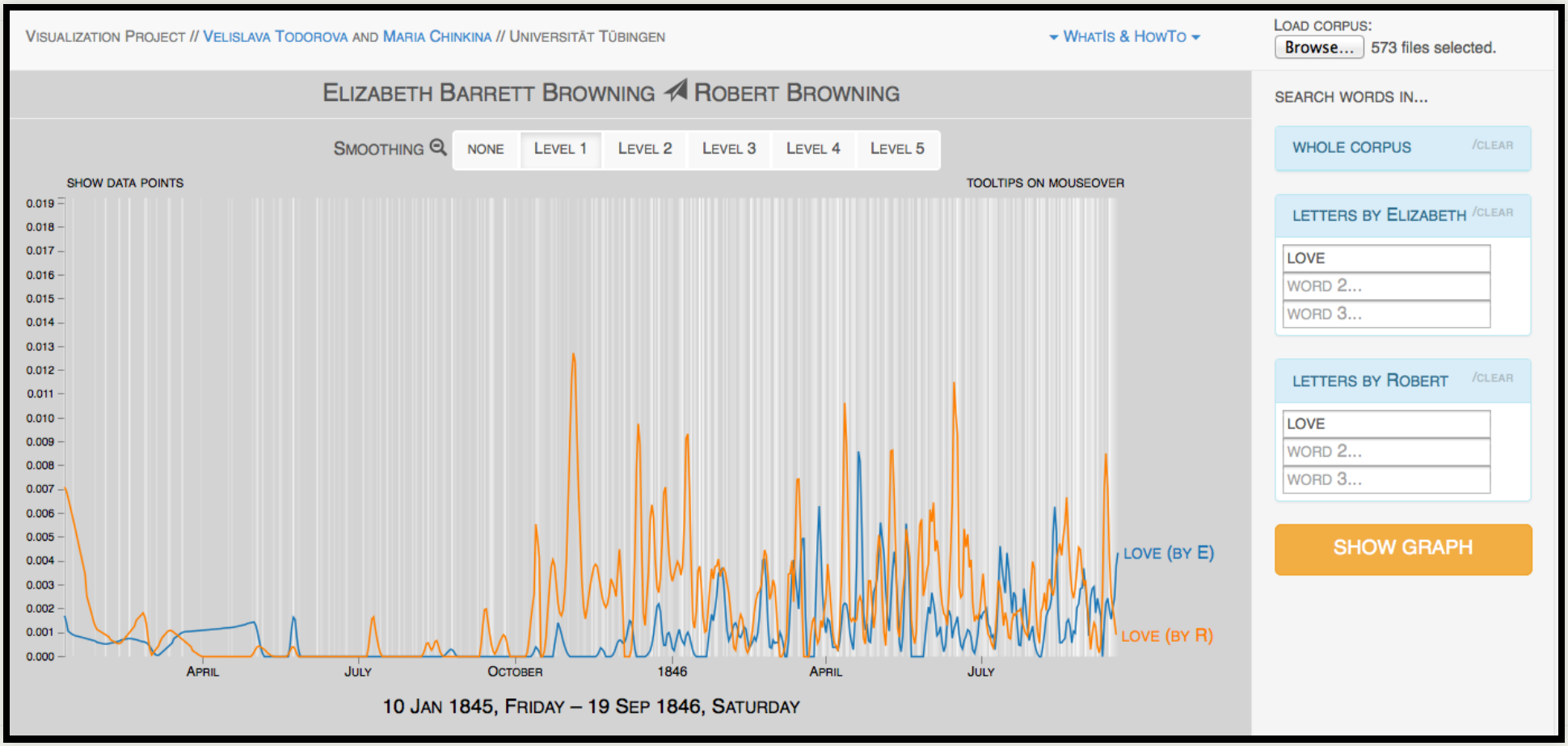


Data: types of questions

– Connections among the letters, across time
e.g., language usage, themes, etc.

Time Student Project: ngramViewer

Maria Chinkina and Velislava Todorova

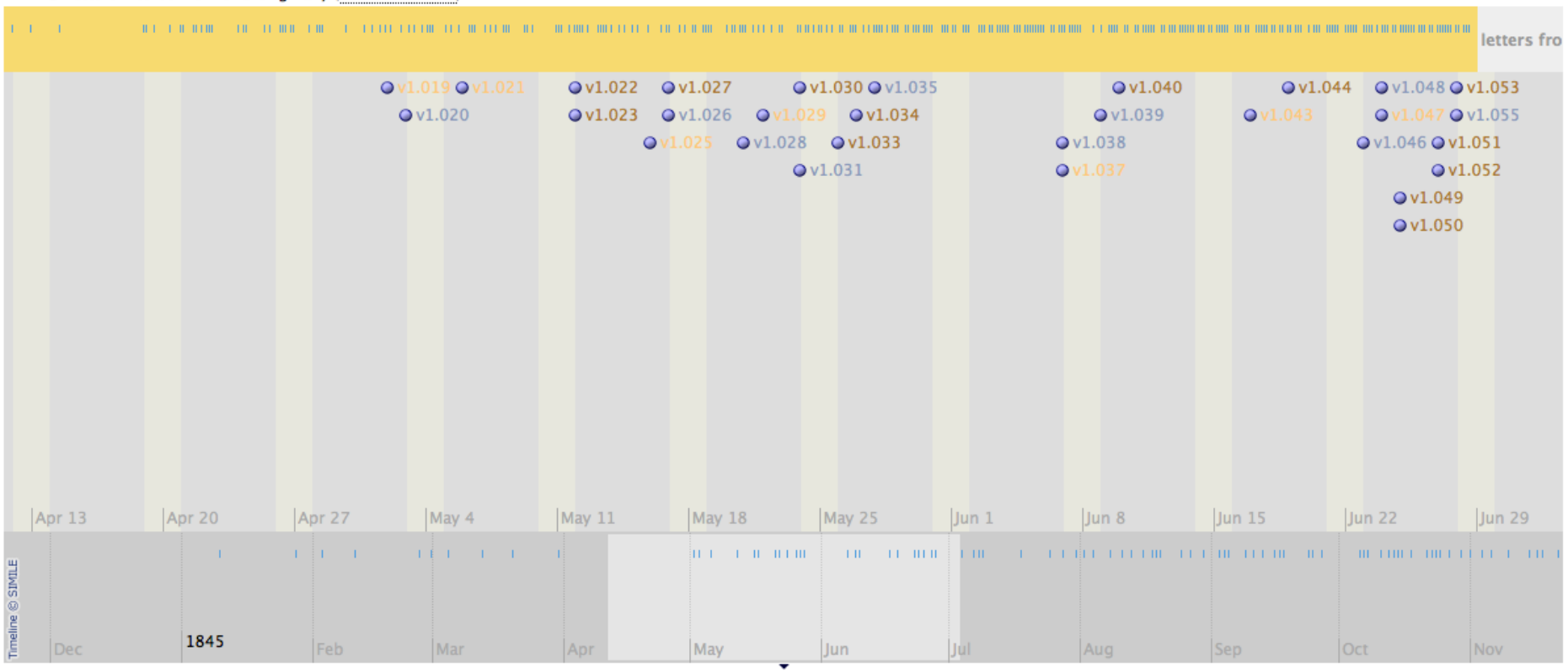


Time Student Project: Topic Timeline

Mike Burkhardt and Lars Horber

OVERVIEW • TIMELINE WITH COLOR CODED SENDERS • **TIMELINE WITH COLOR CODED TOPICS**

440 letters filtered from 573 originally (Reset All Filters)



kenyon tuesday week wednesday evening house thought note promise question

love heart kiss hope matter feeling note thank care suppose

love thought woman heart wrong side full sort light stand

Data: types of questions

– Connections with the world outside of the letters
e.g. local, national, world events;
cultural phenomena, etc.

Terms across time and people in the LOUGH letters

Terms across time and people i...

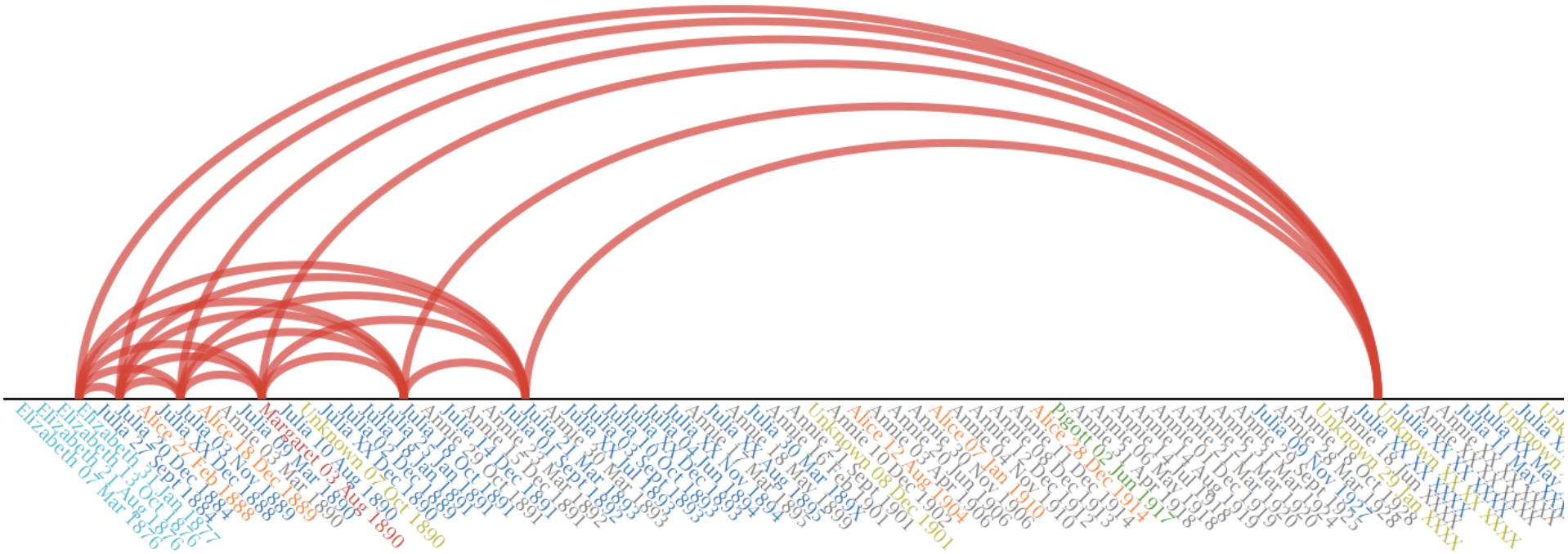
Terms across time and people in the LOUGH letters

Chris Culy, University of Tübingen

ye Add item Remove all #ye_ Error: None

Results: Found 14 new results

Order by Rough Date


Elizabeth 07 Mar 1876
Elizabeth 21 Aug 1876
Elizabeth 13 Oct 1876
Elizabeth 31 Jan 1877
Julia 27 Sep 1884
Alice 27 Feb 1888
Julia 03 Nov 1889
Annie 18 Dec 1888
Julia 09 Mar 1889
Margaret 03 Aug 1890
Julia 10 Aug 1890
Unknown 07 Oct 1890
Julia 04 Dec 1890
Julia 18 Jan 1891
Annie 29 Oct 1891
Julia 14 Dec 1891
Annie 15 Dec 1891
Julia 23 Mar 1892
Julia 01 Sept 1892
Annie 30 Mar 1893
Julia 03 Jul 1893
Julia 10 Oct 1893
Julia 04 Dec 1893
Julia 17 Jun 1894
Julia 17 Nov 1894
Julia 18 Aug 1895
Julia 30 Mar 1897
Annie 16 Feb 1899
Unknown 08 Sept 1901
Alice 12 Aug 1904
Annie 10 Dec 1901
Annie 03 Apr 1906
Annie 30 Nov 1906
Alice 07 Jan 1910
Annie 04 Nov 1906
Annie 08 Dec 1910
Annie 11 Dec 1912
Annie 01 Dec 1913
Pigott 03 Jun 1914
Alice 28 Dec 1914
Annie 06 May 1917
Annie 14 Jul 1918
Annie 21 Aug 1918
Annie 21 Dec 1918
Annie 31 Mar 1926
Julia 09 Nov 1925
Annie 28 Sept 1925
Unknown 18 Oct 1927
Annie 18 Mar 1928
Julia 18 Jun 1928
Unknown 29 Jan XXXX
Julia XX XX XXXX
Julia XX XX XXXX
Julia XX XX

Terms across time and people in the LOUGH letters

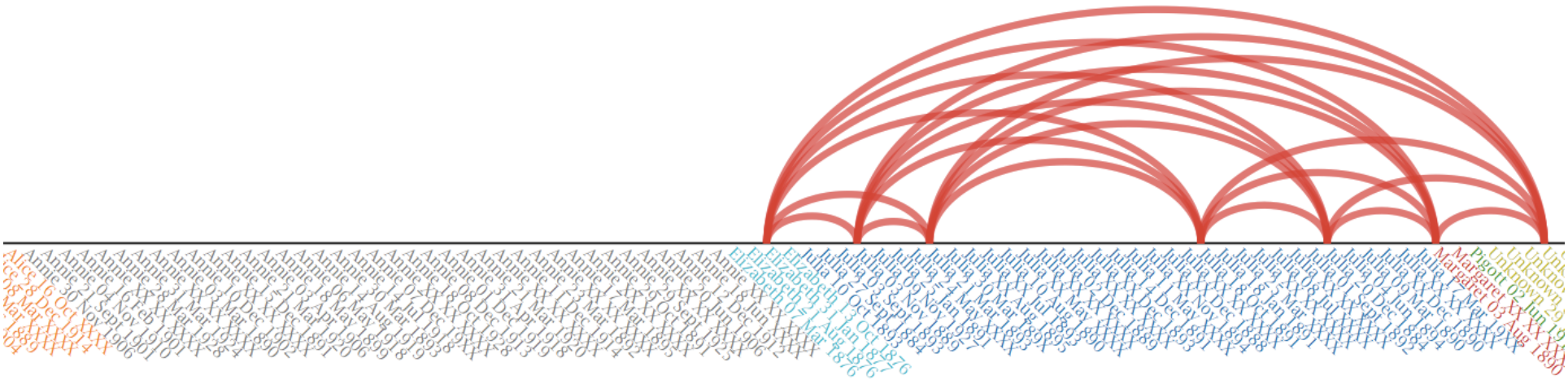
Terms across time and people i...

Terms across time and people in the LOUGH letters

Chris Culy, University of Tübingen

m Remove all #ye_ Error: None

results



Terms across time and people in the LOUGH letters

Terms across time and people i...

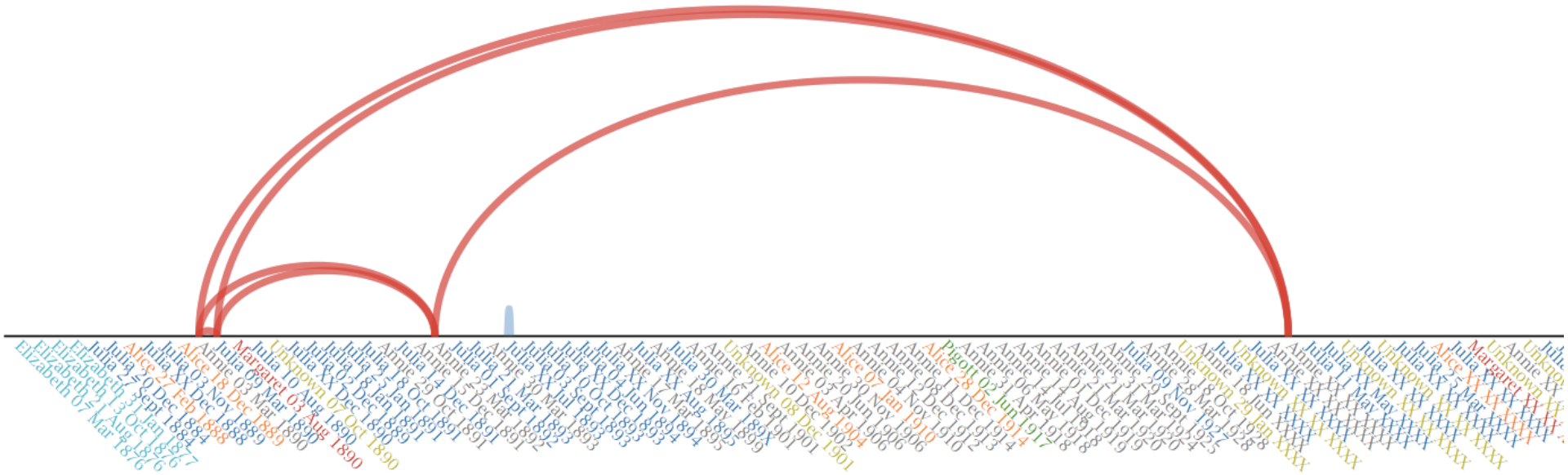
Terms across time and people in the LOUGH letters

Chris Culy, University of Tübingen

Gripp | Add item | Remove all | #Gripp_ #grip_ #ye_ | Error: None

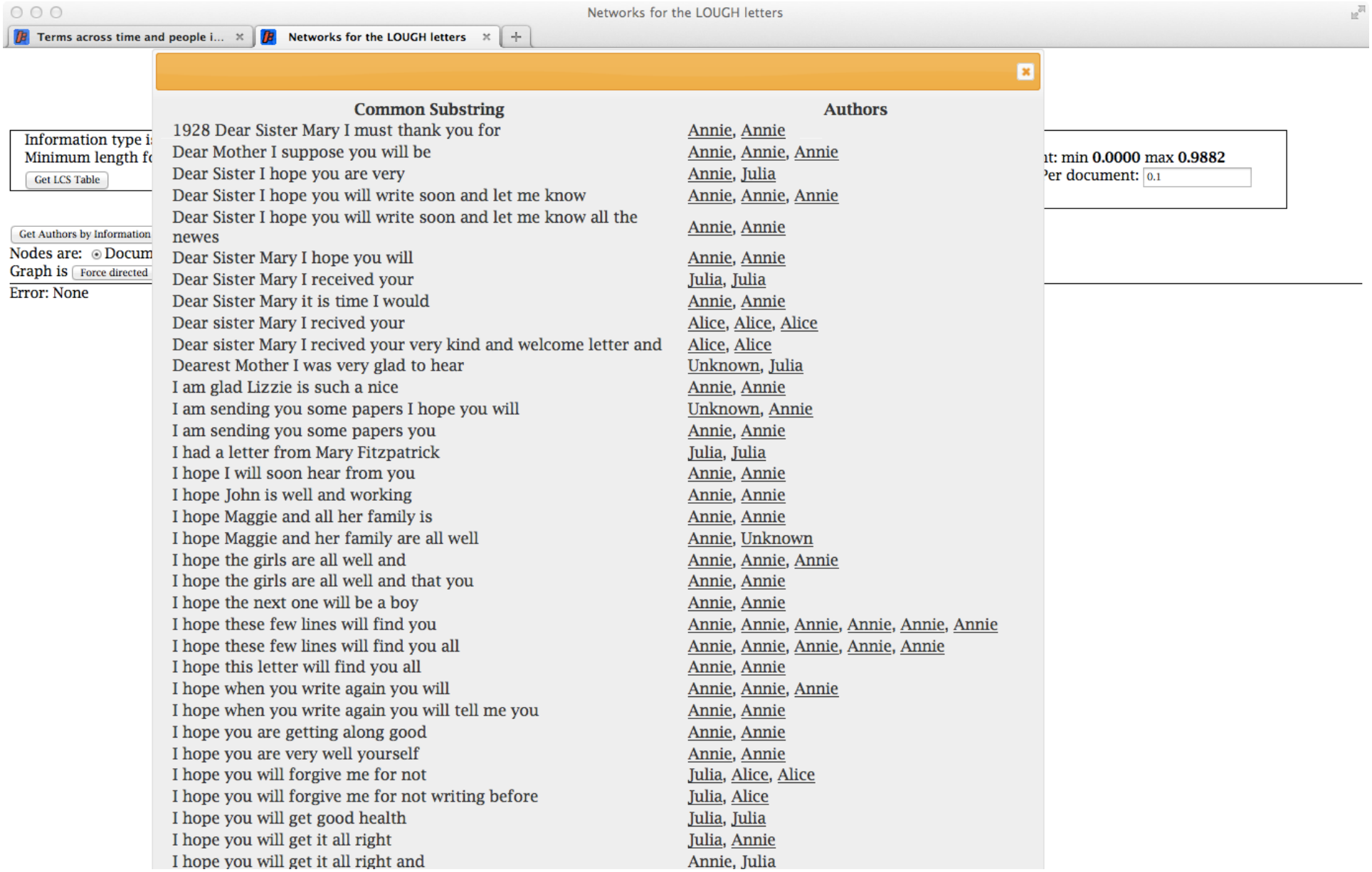
Results: Found 1 new results

Order by Rough Date


Julia 27 Sept 1884
Alice 22 Dec 1884
Julia XX Dec 1888
Alice 18 Feb 1888
Annie 03 Nov 1888
Julia 09 Dec 1889
Margaret 03 Mar 1889
Unknown 07 Aug 1890
Julia 10 Oct 1890
Julia XX Dec 1890
Julia 07 Dec 1890
Julia 18 Jan 1891
Annie 29 Oct 1891
Julia 24 Dec 1891
Julia 01 Sept 1891
Annie 15 Dec 1891
Julia 21 Mar 1892
Annie 30 Mar 1893
Julia XX Sept 1893
Julia 10 Oct 1893
Julia 04 Dec 1893
Julia XX Jun 1894
Julia XX Nov 1894
Julia 17 Aug 1894
Julia 18 May 1895
Julia 30 Mar 1895
Annie 15 Feb 1899
Unknown 08 Sept 1901
Alice 12 Dec 1901
Annie 16 Aug 1904
Annie 03 April 1906
Alice 07 Jan 1910
Annie 30 Nov 1906
Annie 04 Nov 1906
Annie 08 Dec 1910
Alice 28 Dec 1913
Pigott 03 Jun 1914
Annie 06 April 1917
Annie 14 May 1918
Annie 01 Aug 1918
Annie 27 Dec 1916
Annie 21 Mar 1916
Julia 09 Sept 1920
Julia 09 Nov 1924
Unknown 18 Oct 1925
Annie 18 Mar 1928
Julia 18 Jun 1928
Unknown 29 Jan XXXX
Julia XX XXXX
Julia XX XXXX
Julia XX XXXX
Unknown XXXX
Julia XX XXXX
Julia XX XXXX
Unknown XXXX
Julia XX XXXX
Alice XXXX
Margaret XXXX
Unknown XXXX
Julia XX XXXX
Julia XX XXXX
Julia XX

Data: types of questions

– Connections among the participants
e.g. who wrote to whom when and where, but also *how*



Common Substring	Authors
1928 Dear Sister Mary I must thank you for	Annie, Annie
Dear Mother I suppose you will be	Annie, Annie, Annie
Dear Sister I hope you are very	Annie, Julia
Dear Sister I hope you will write soon and let me know	Annie, Annie, Annie
Dear Sister I hope you will write soon and let me know all the newes	Annie, Annie
Dear Sister Mary I hope you will	Annie, Annie
Dear Sister Mary I received your	Julia, Julia
Dear Sister Mary it is time I would	Annie, Annie
Dear sister Mary I recived your	Alice, Alice, Alice
Dear sister Mary I recived your very kind and welcome letter and	Alice, Alice
Dearest Mother I was very glad to hear	Unknown, Julia
I am glad Lizzie is such a nice	Annie, Annie
I am sending you some papers I hope you will	Unknown, Annie
I am sending you some papers you	Annie, Annie
I had a letter from Mary Fitzpatrick	Julia, Julia
I hope I will soon hear from you	Annie, Annie
I hope John is well and working	Annie, Annie
I hope Maggie and all her family is	Annie, Annie
I hope Maggie and her family are all well	Annie, Unknown
I hope the girls are all well and	Annie, Annie, Annie
I hope the girls are all well and that you	Annie, Annie
I hope the next one will be a boy	Annie, Annie
I hope these few lines will find you	Annie, Annie, Annie, Annie, Annie, Annie
I hope these few lines will find you all	Annie, Annie, Annie, Annie, Annie
I hope this letter will find you all	Annie, Annie
I hope when you write again you will	Annie, Annie, Annie
I hope when you write again you will tell me you	Annie, Annie
I hope you are getting along good	Annie, Annie
I hope you are very well yourself	Annie, Annie
I hope you will forgive me for not	Julia, Alice, Alice
I hope you will forgive me for not writing before	Julia, Alice
I hope you will get good health	Julia, Julia
I hope you will get it all right	Julia, Annie
I hope you will get it all right and	Annie, Julia

Networks for the LOUGH letters

Terms across time and people i... | Networks for the LOUGH letters

Networks for the LOUGH letters

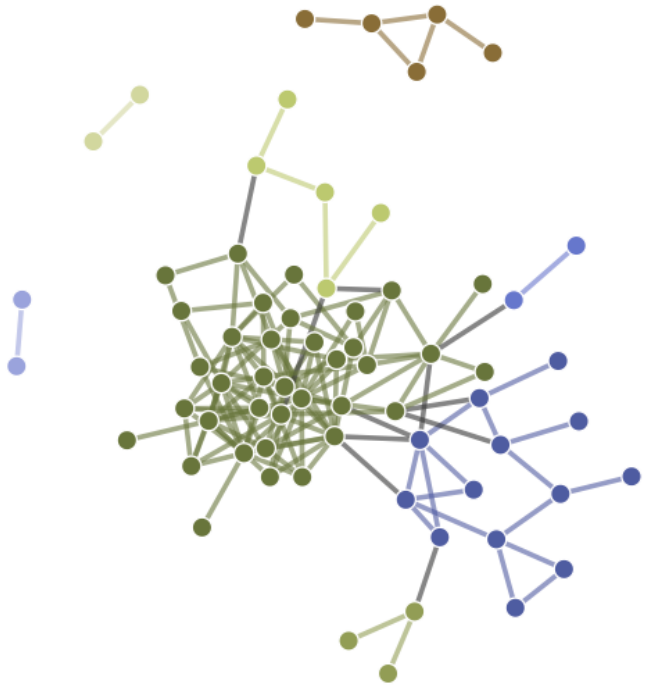
Chris Culy: University of Tübingen

Information type is ◉ Long Common Substrings
Minimum length for common substrings (including spaces): 32
Get LCS Table

Information type is ○ Automatic topics
Values — Global: min **0.0122** max **2.7567**, Per Document: min **0.0000** max **0.9882**
Minimal thresholds for topics. Global: 0.05 Per document: 0.1
Get Topic Table

Get Authors by Information
Nodes are: ◉ Documents ○ People
Graph is Force directed Advanced: Proportion of edges to omit: 0.1 Get Graph

Error: None Results: Number of groups 8



Networks for the LOUGH letters

Terms across time and people i... | Networks for the LOUGH letters

Networks for the LOUGH letters

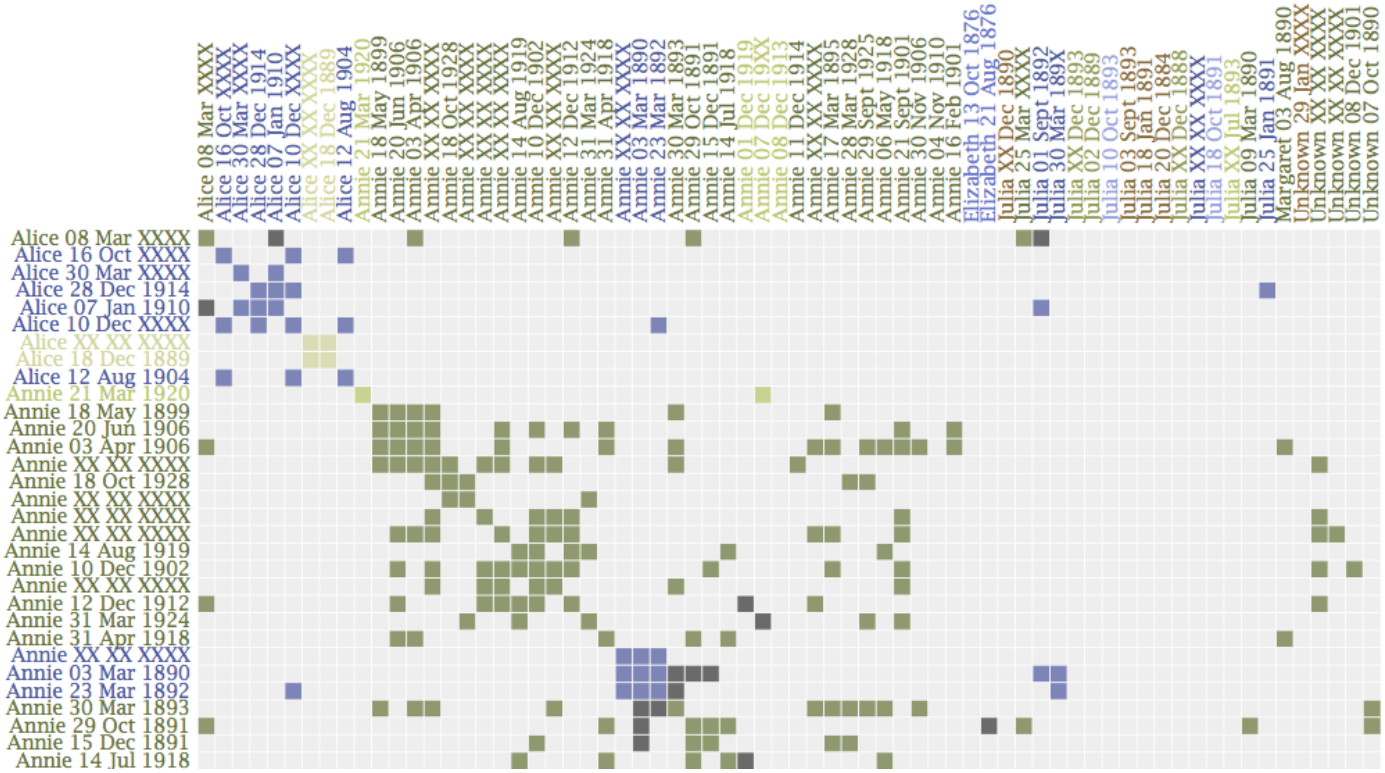
Chris Culy: University of Tübingen

Information type is ◉ Long Common Substrings
Minimum length for common substrings (including spaces): 32
Get LCS Table

Information type is ○ Automatic topics
Values — Global: min **0.0122** max **2.7567**, Per Document: min **0.0000** max **0.9882**
Minimal thresholds for topics. Global: 0.05 Per document: 0.1
Get Topic Table

Get Authors by Information
Nodes are: ◉ Documents ○ People
Graph is Matrix Advanced: Proportion of edges to omit: 0.1 Get Graph

Error: None Results: Number of groups 8
Order by Person



Networks for the LOUGH letters

Terms across time and people i... | Networks for the LOUGH letters

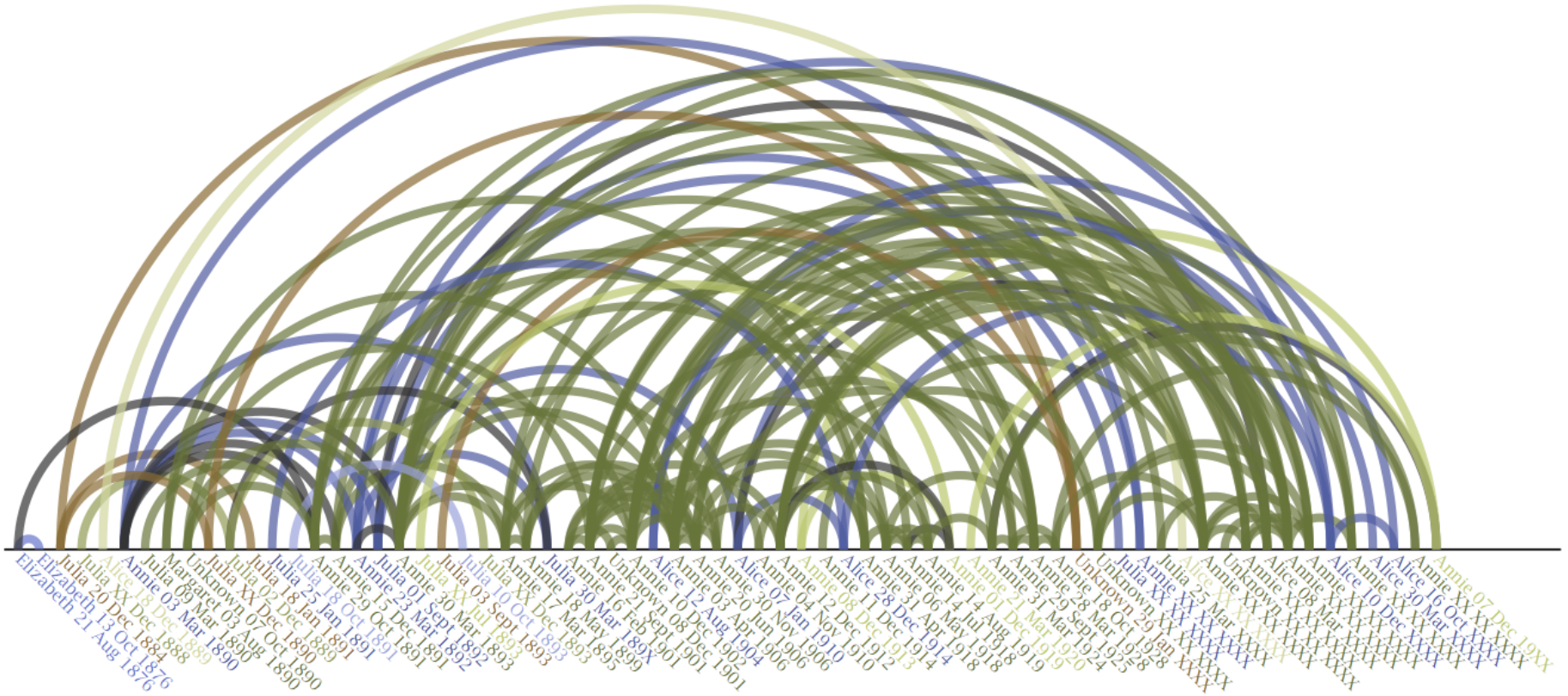
Get Authors by Information

Nodes are: ◉ Documents ○ People

Graph is Arcs Advanced: Proportion of edges to omit: 0.1 Get Graph

Error: None Results: Number of groups 8

Order by Group



Hypothesizing about the *Unknowns*

Can we use visualizations to help fill in the metadata gaps?

016: Unknown_07Oct1890.txt
038: Unknown_08Dec1901.txt
046: Unknown_29JanXXXX.txt
063: Unknown_XX_XX_XXXX.txt
071: Unknown_XX_XX_XXXX.txt
073: Unknown_XX_XX_XXXX.txt
082: Unknown_XX_XX_XXXX.txt
084: Unknown_XX_XX_XXXX.txt

Networks for the LOUGH letters

Terms across time and people i... | Networks for the LOUGH letters

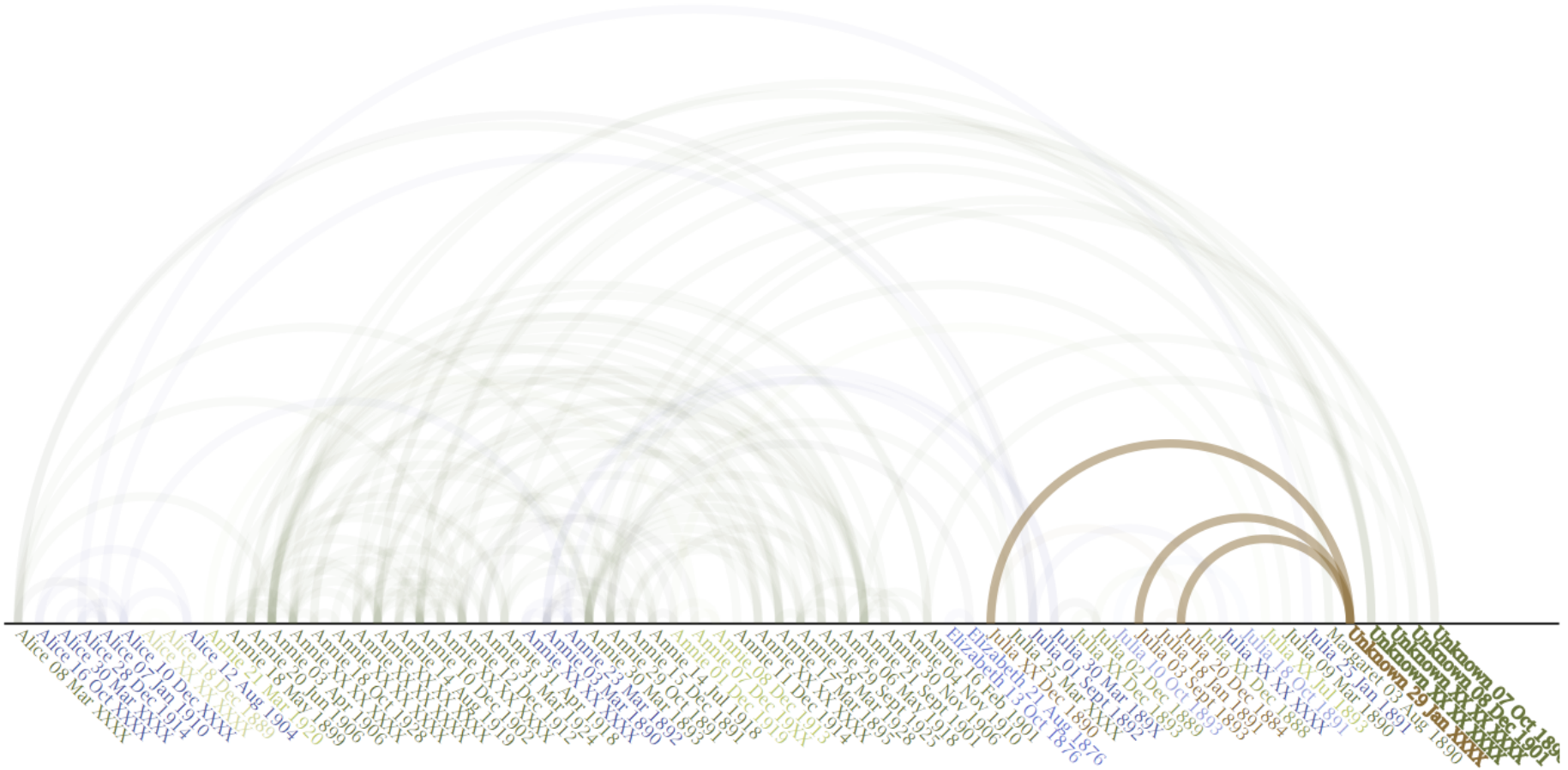
Get Authors by Information

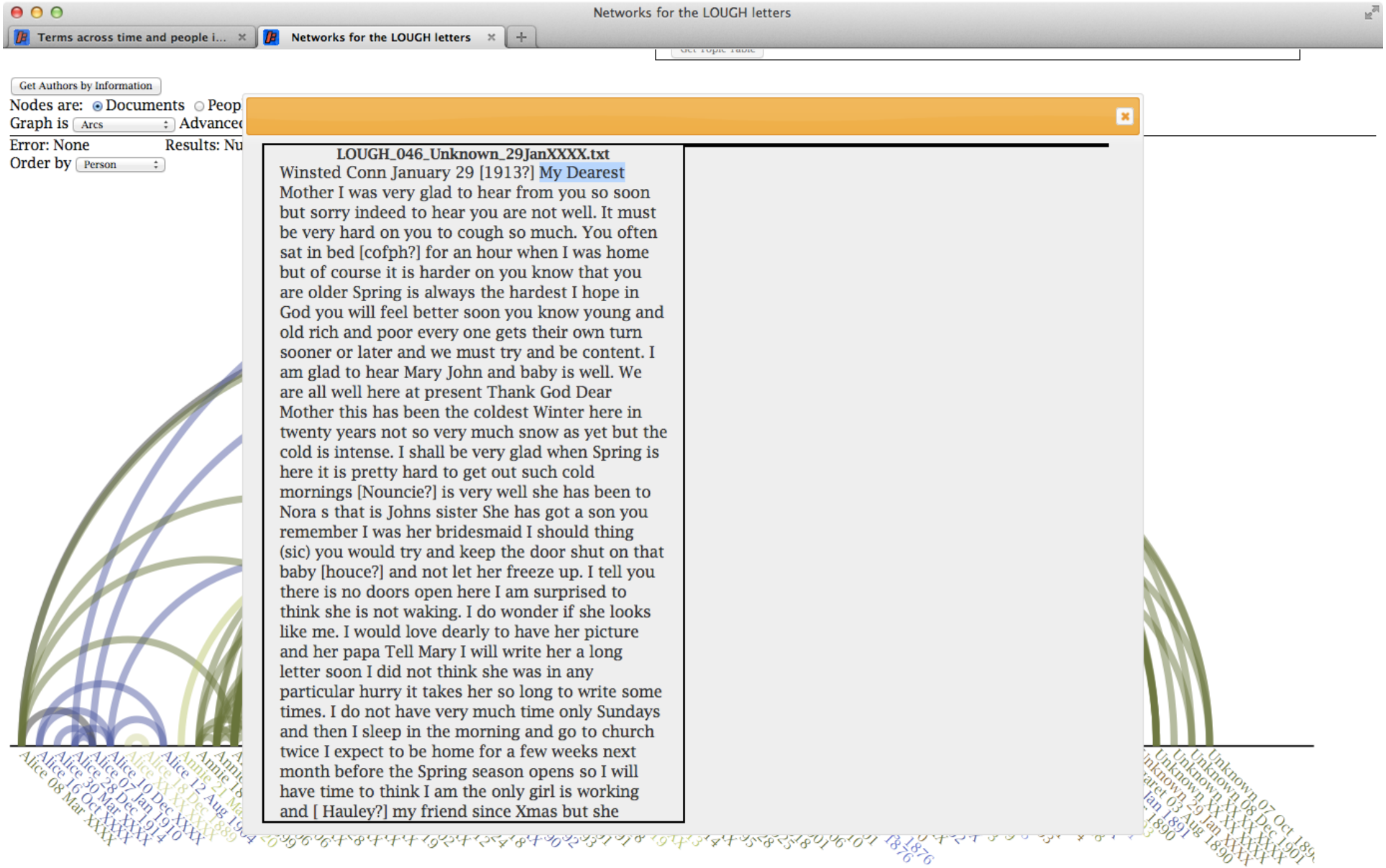
Nodes are: Documents People

Graph is Arcs Advanced: Proportion of edges to omit: 0.1 Get Graph

Error: None Results: Number of groups 8

Order by Person




Networks for the LOUGH letters
Terms across time and people i...
Networks for the LOUGH letters
Get Authors by Information
Nodes are: Documents Peop
Graph is Arcs Advanced
Error: None Results: Nu
Order by Person
LOUGH_046_Unknown_29JanXXXX.txt
Winsted Conn January 29 [1913?] My Dearest Mother I was very glad to hear from you so soon but sorry indeed to hear you are not well. It must be very hard on you to cough so much. You often sat in bed [cofph?] for an hour when I was home but of course it is harder on you know that you are older Spring is always the hardest I hope in God you will feel better soon you know young and old rich and poor every one gets their own turn sooner or later and we must try and be content. I am glad to hear Mary John and baby is well. We are all well here at present Thank God Dear Mother this has been the coldest Winter here in twenty years not so very much snow as yet but the cold is intense. I shall be very glad when Spring is here it is pretty hard to get out such cold mornings [Nouncie?] is very well she has been to Nora s that is Johns sister She has got a son you remember I was her bridesmaid I should thing (sic) you would try and keep the door shut on that baby [houce?] and not let her freeze up. I tell you there is no doors open here I am surprised to think she is not waking. I do wonder if she looks like me. I would love dearly to have her picture and her papa Tell Mary I will write her a long letter soon I did not think she was in any particular hurry it takes her so long to write some times. I do not have very much time only Sundays and then I sleep in the morning and go to church twice I expect to be home for a few weeks next month before the Spring season opens so I will have time to think I am the only girl is working and [Hauley?] my friend since Xmas but she

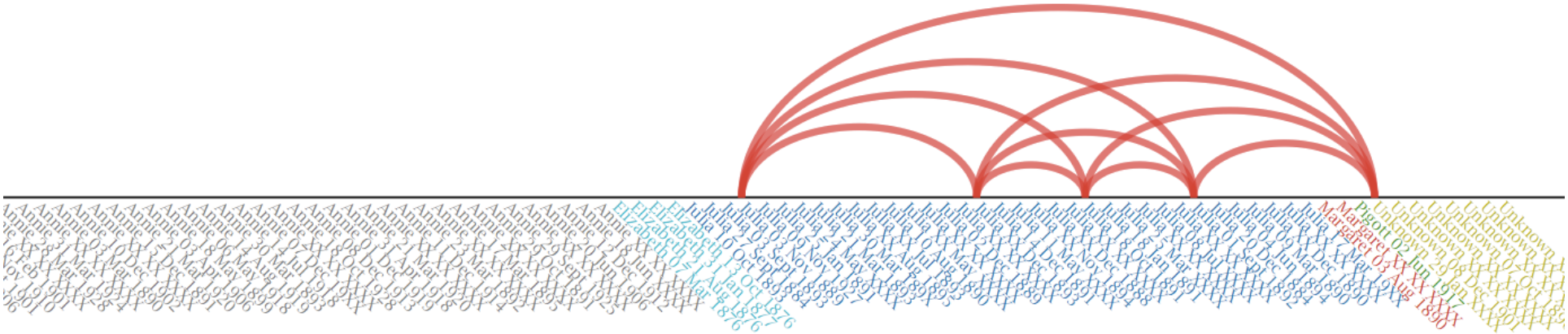
Terms across time and people in the LOUGH letters

Terms across time and people i... | Networks for the LOUGH letters

Terms across time and people in the LOUGH letters

Chris Culy, University of Tübingen

My# #Dearest Error: None



Hypothesizing about the *Unknowns*

We *can* use visualizations to help fill in the metadata gaps

016: Unknown_07Oct1890.txt : **Annie**
038: Unknown_08Dec1901.txt : **Annie**
046: Unknown_29JanXXXX.txt : **Julia**
063: Unknown_XX_XX_XXXX.txt : Julia ? or Annie
071: Unknown_XX_XX_XXXX.txt : TBD
073: Unknown_XX_XX_XXXX.txt : TBD
082: Unknown_XX_XX_XXXX.txt : TBD
084: Unknown_XX_XX_XXXX.txt : TBD

Visualizations put ideas into our heads!

Thank you!

Contact: **http://www.sfs.uni-tuebingen.de/~cculy/**

My free visualization software:
http://www.sfs.uni-tuebingen.de/~cculy/index_research.html

The *Vistola* Project — Visualization and analysis of person-oriented correspondence:
http://www.sfs.uni-tuebingen.de/~cculy/vistola/