

Computing and Visualizing the 19th Century Literary Genome

Matthew Jockers
Stanford University

Talents imitate, geniuses steal
(Oscar Wilde)

All ideas are second hand, consciously and unconsciously drawn from a million outside sources.

(Mark Twain, 1903)

The historical sense compels a man to write not merely with his own generation in his bones, but with a feeling that the whole of literature has a simultaneous existence
(T. S. Eliot, 1920)

Every work of art is the result of a complex interrelation of individual features. . . .

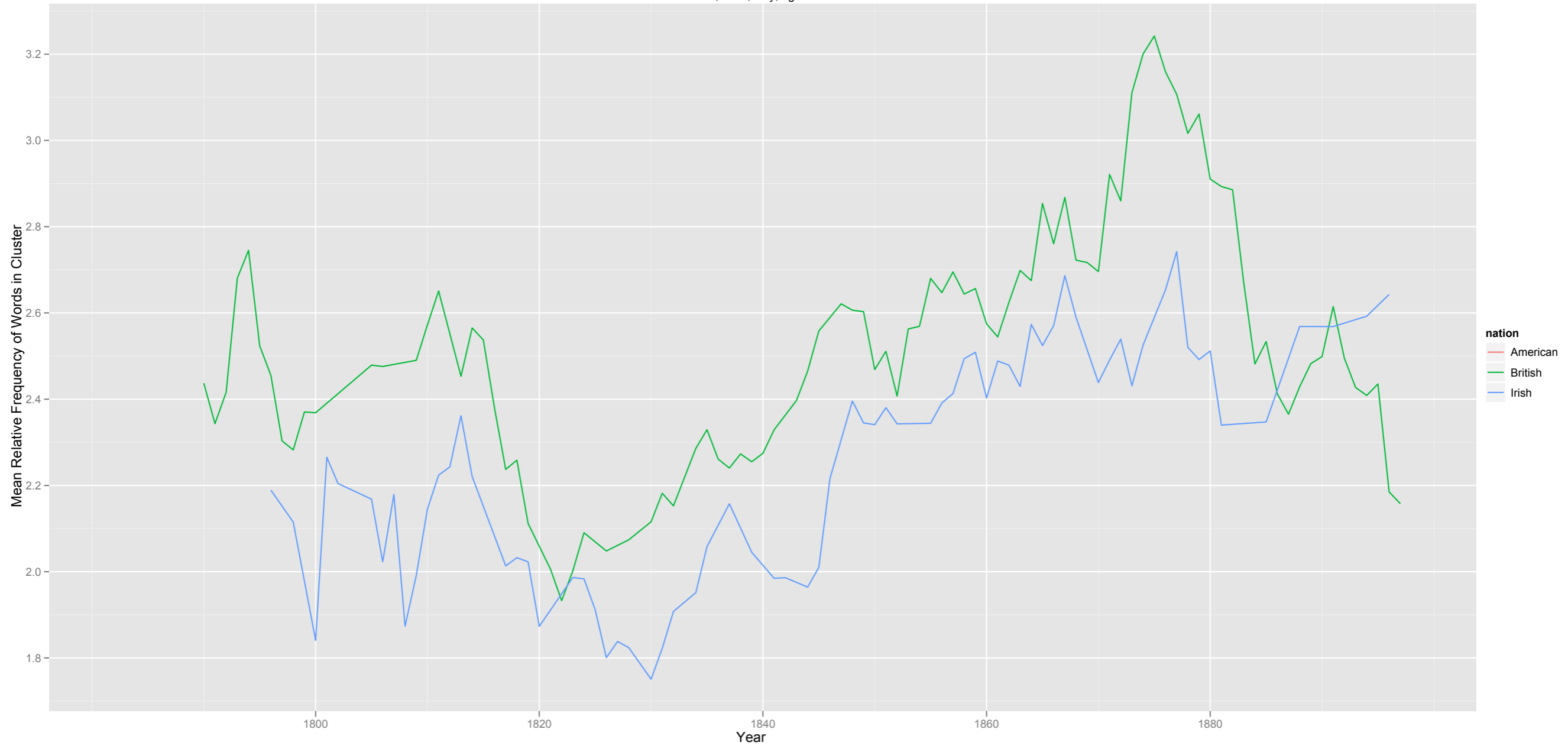
The elements of which the artwork is created are external to the author and independent of him.

(Osip Brik, 1929)

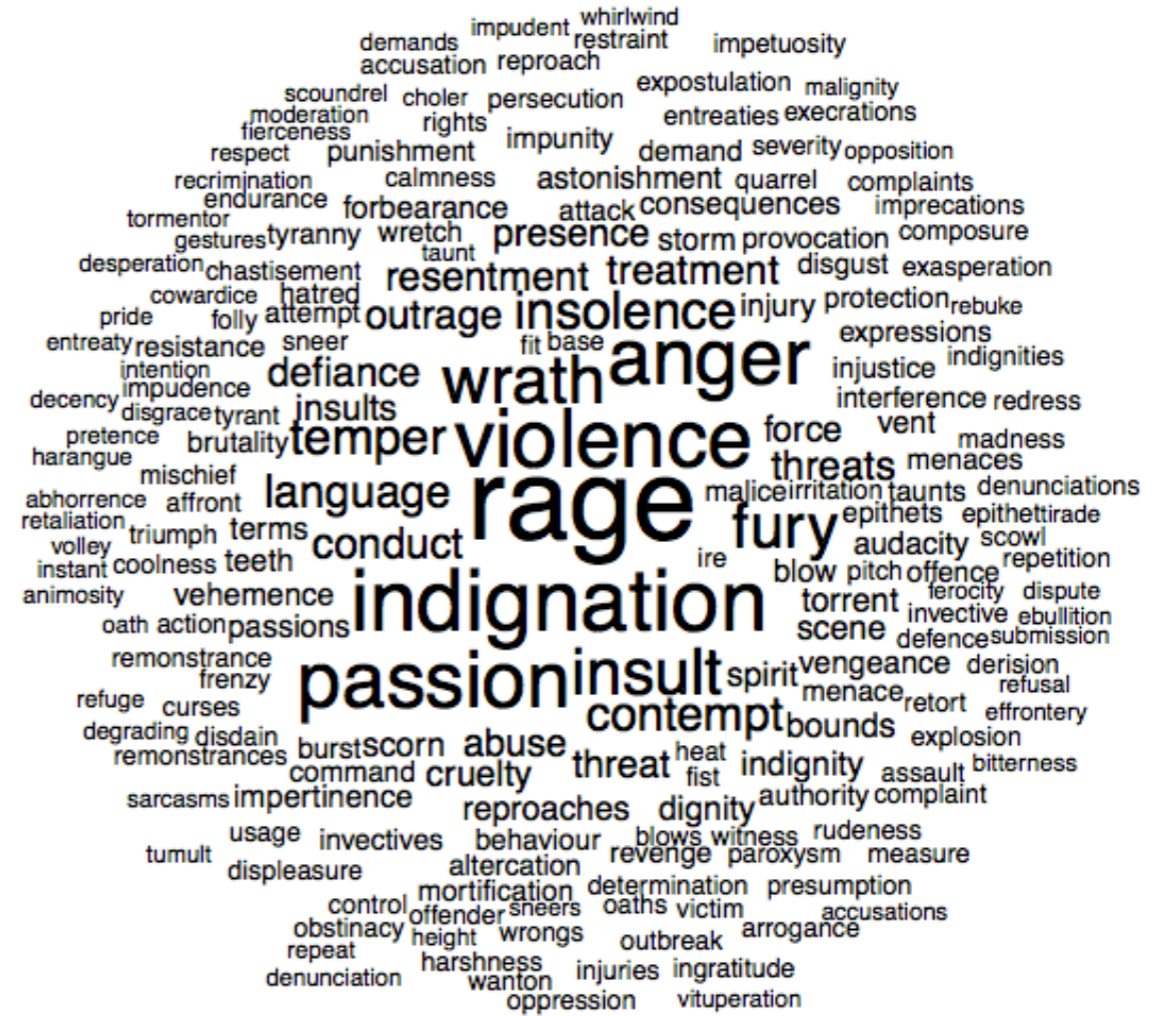
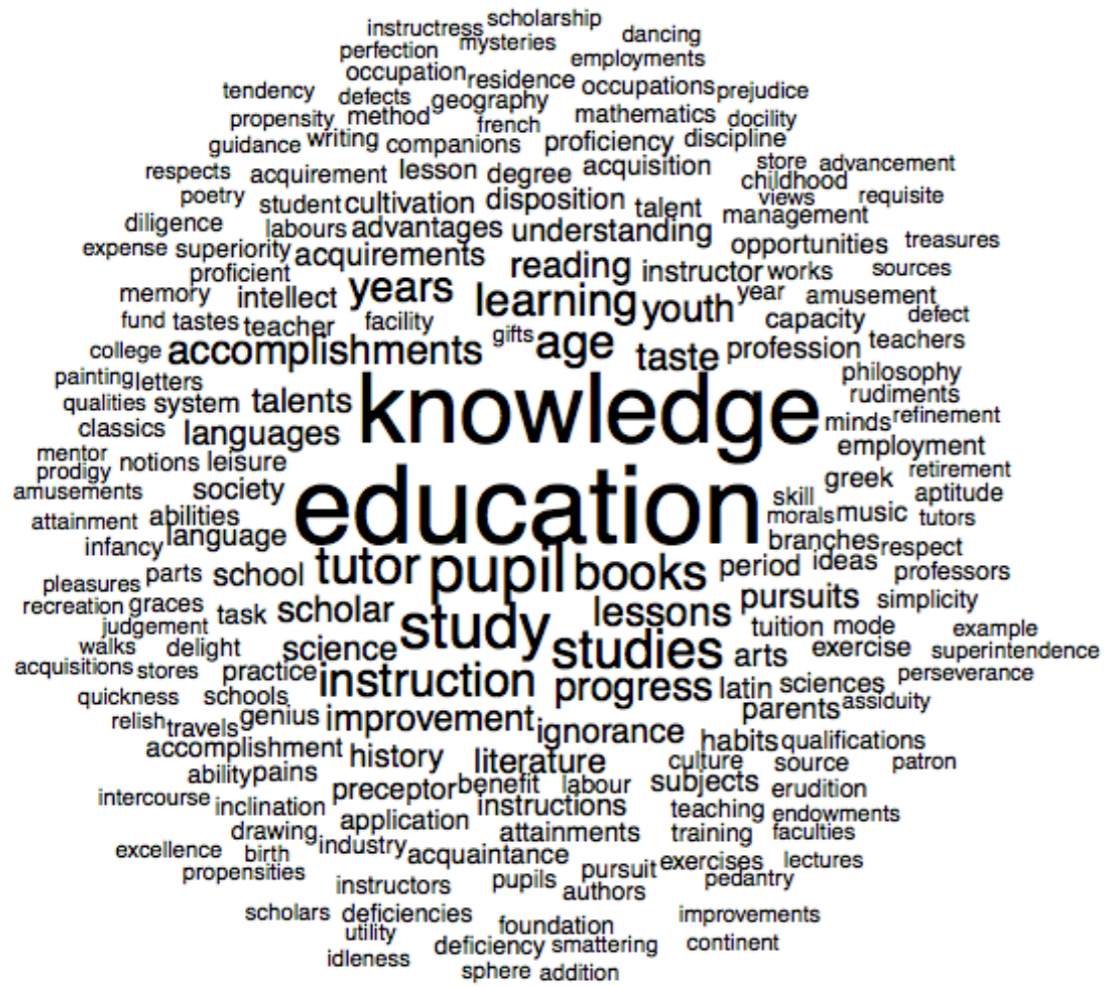
The main concern of the literary historian is not with assessing the unique contributions of individual writers, but with spotting the migratory poetic formulae; accounting for their appearance in various ethnic milieus. . . and tracing them through all vicissitudes back to the starting point.

(Alexander Veselovsky)

Distinguishing British Features: always, should, never, sure, not, must, do, dont, no, always, nothing, certain, therefore, because, can, cannot, knew, know,
last, once, only, right

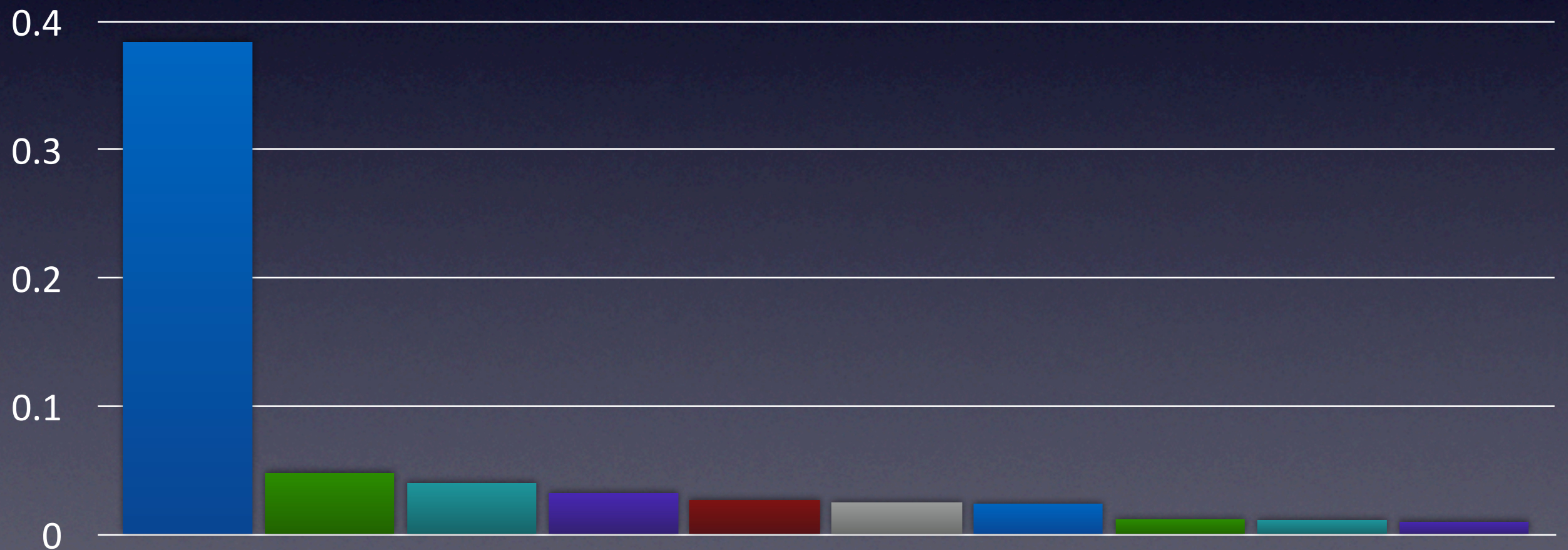


nation
American
British
Irish



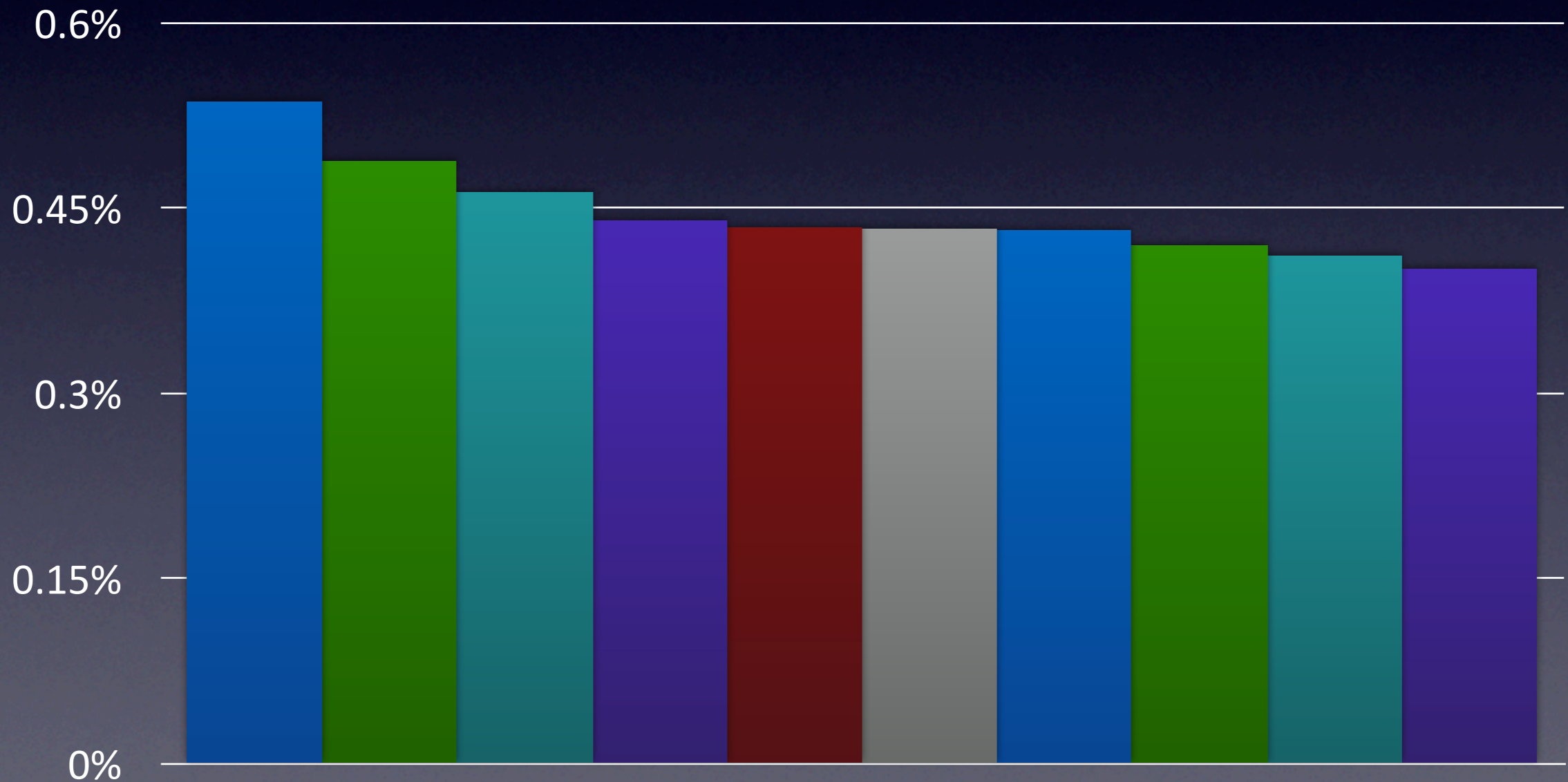
Top Ten Topics in Melville's Moby Dick

- Whaling
- Sailing and Sailors
- Steam Ships and Ocean Travel
- Heaven and Creation
- Physical Features
- Guns Ships and Seamen
- Boats
- Physical Features
- Animals and Beasts
- Death despair hell demons monsters

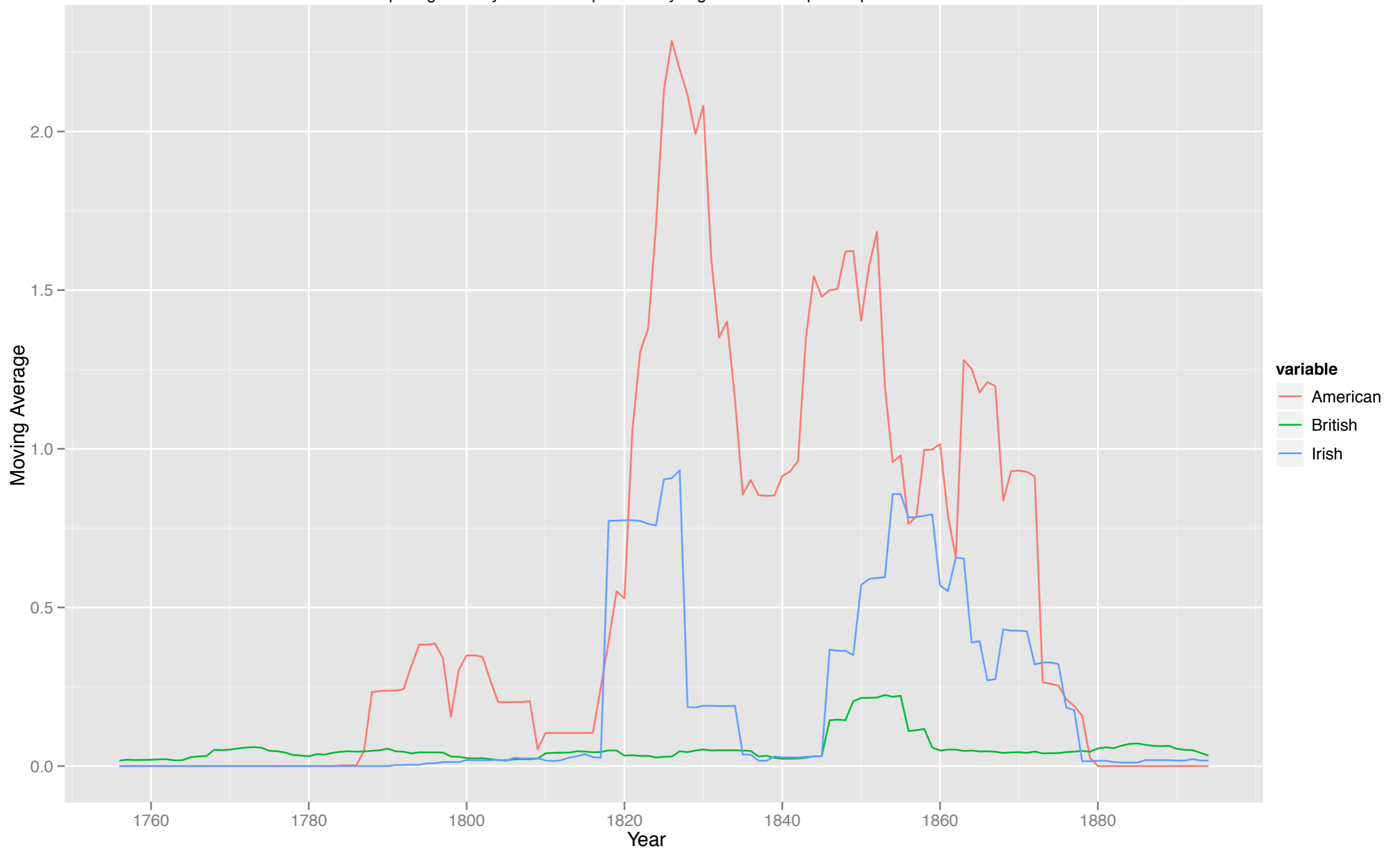


Top Ten Topics In Corpus

- Women
- Female Features and Female Emotion
- Physical Features
- Difficult Emotional Situations
- Marriage Love and Happiness
- Culture and Society
- Facial Features
- Pursuit
- Happiness Affection feelings
- Night Darkness Outdoors



ID: Topic_392 Rank: 195 Spread Rank: 50 Label: INDIANS Words: indians chief warriors war savages warrior party prairie chiefs men spirit tribes knife council enemies friends fire face tomahawk nation scalp death whites skin river hands faces race women medicine skins band feet children prisoners path buffalo arrows squaw ground eyes arms fort pale enemy wigwam rest scalps companions



Some Sample Data

	A	B	C	D	E	F	G	H	I	J	
1	filename	year	author	title	gender	nation	X0	X1	X2	X3	X4
2	lilaBookId_0	1886	Norris, William Edward	A bachelor's blunder	M	British	0.00144036	0.00379127	5.07E-05	0.00051523	0.00
3	lilaBookId_0	1894	Marryat, Florence	A bankrupt heart	F	British	0.00094324	0.00019608	0	0	0.00
4	lilaBookId_0	1883	Sherard, Robert Harborough	A bartered honour a novel	M	British	0.00604787	0.00033374	0.00030248	0.00025493	0.0
5	lilaBookId_0	1887	Croker, Bithia Mary	A bird of passage	F	British	0.00039955	0.00129833	0.00011703	0	0.00
6	lilaBookId_0	1890	Duchess	A born coquette A novel	U	British	0.00062631	0.00030337	0.00029895	7.17E-05	0.00
7	lilaBookId_0	1864	Clarke, Charles	A box for the season a sporting	M	British	0.00098075	0.15005655	0.00013004	0	0.00
8	lilaBookId_0	1888	Wall, Caroline	A Breton maiden	F	British	0.00107204	0.00050388	0	0	0.00
9	lilaBookId_0	1892	Chetwynd, Henry Wayland Mr	A brilliant woman	F	British	6.71E-05	0	0	0	
10	lilaBookId_0	1879	Marryat, Florence	A broken blossom a novel	F	British	0.00039274	7.95E-05	0	0	0.00
11	lilaBookId_0	1879	Hardy, Iza Duffus	A broken faith In three volume	F	British	0.0003172	0	0.00023165	0.00015709	0.00
12	lilaBookId_0	1882	Collins, Frances Cotton	A broken lily	M	British	0	0.00069119	0.00013984	0	0.00
13	lilaBookId_0	1883	Dillwyn, Elizabeth Amy	A burglary or Unconscious influ	F	British	0.01672055	0.00141945	0.00050679	0	0.00
14	lilaBookId_00	1873	Howells, William Dean	A Chance Acquaintance 1873	M	American	0.0006256	0.00062938	0	0	0.00
15	lilaBookId_0	1882	Clare, Austin	A child of the menhir a novel	M	British	5.10E-05	0.00056025	0	0.0001924	0.00
16	lilaBookId_0	1891	Williamson, Emma Sara	A child widow	F	British	0.00097765	0.00015175	0	0.00013141	0.00
17	lilaBookId_0	1894	Alexander Mrs	A choice of evils a novel	F	British	0.00066444	0.00018749	0	0.00028248	0.00
18	lilaBookId_00	1843	Dickens, Charles	A Christmas Carol In Prose Bein	M	British	0.00024272	0.00023042	0.00047962	0.0004065	0.00

The Euclidean Metric

Book	f1	f2	f3	f4
b1	10	5	3	5
b2	11	6	5	7
b3	4	13	2	6

$$\text{distance}(p, q) = \sqrt{(q_1-p_1)^2 + (q_2-p_2)^2 + (q_3-p_3)^2 + (q_4-p_4)^2}$$

$$d(b1, b2) = \sqrt{(10-11)^2 + (5-6)^2 + (3-5)^2 + (5-7)^2} = \underline{3.162278}$$

$$d(b1, b3) = \sqrt{(10-4)^2 + (5-13)^2 + (3-2)^2 + (5-6)^2} = \underline{10.09950}$$

$$d(b2, b3) = \sqrt{(11-4)^2 + (6-13)^2 + (5-2)^2 + (7-6)^2} = \underline{10.39230}$$

Example Distance Matrix

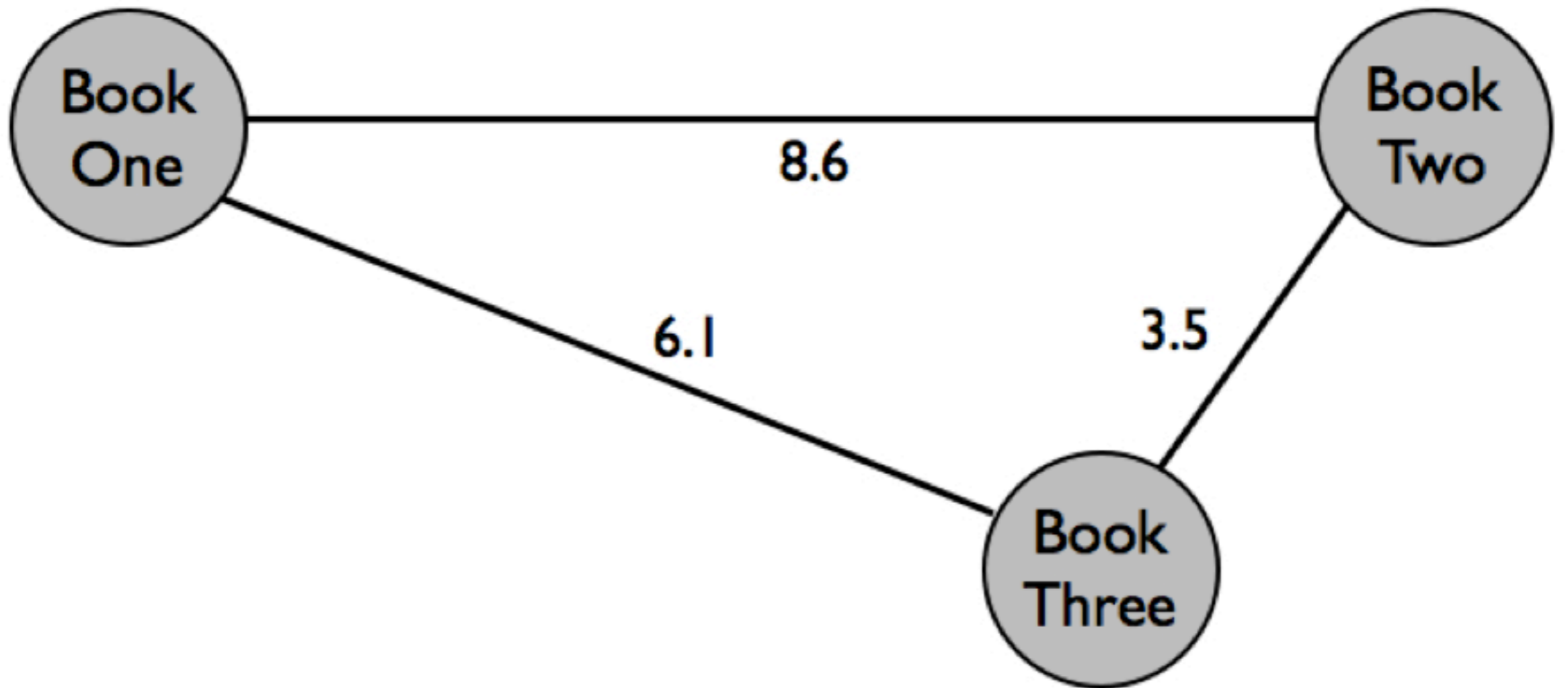
	Carleton1.xml	Carleton10.xml	Carleton11.xml	Carleton12.xml	Carleton13.xml	Carleton14.xml	Carleton2.xml	Carleton3.xml
Carleton1.xml	0	1.679938521	0.960117038	1.185782588	1.312755814	1.2661047	3.131012543	1.142872814
Carleton10.xml	1.679938521	0	2.175933776	1.357208861	1.134930518	0.823822018	2.645069283	1.713419024
Carleton11.xml	0.960117038	2.175933776	0	1.457827099	1.72201182	1.693232655	3.236680611	1.436069718
Carleton12.xml	1.185782588	1.357208861	1.457827099	0	0.831772568	0.841700015	2.891452698	1.534835047
Carleton13.xml	1.312755814	1.134930518	1.72201182	0.831772568	0	0.811858474	3.078884535	1.607330953
Carleton14.xml	1.2661047	0.823822018	1.693232655	0.841700015	0.811858474	0	2.787815743	1.617856943
Carleton2.xml	3.131012543	2.645069283	3.236680611	2.891452698	3.078884535	2.787815743	0	3.029465699
Carleton3.xml	1.142872814	1.713419024	1.436069718	1.534835047	1.607330953	1.617856943	3.029465699	0
Carleton4.xml	2.927245416	1.805539326	3.422179554	2.775434562	2.377607243	2.342042548	2.93852519	2.645065541
Carleton5.xml	2.344395221	1.887380429	2.360881422	2.010835208	1.834111514	1.902635776	2.393144922	2.287838026
Carleton6.xml	2.413222638	2.862184391	2.18772205	2.621739633	2.741889301	2.72221064	3.70078259	2.553058604
Carleton7.xml	1.608091626	1.536605275	2.05384419	1.617786625	1.755194251	1.674775789	2.712071375	1.08944657
Carleton8.xml	1.61832525	1.93979794	1.733104665	2.00054681	2.196088694	1.886388078	2.662655378	1.173533852
Carleton9.xml	1.346445917	1.226978519	1.900104227	1.282116398	1.195418349	1.285518299	3.102348535	1.028470208
donovan1.xml	3.208722136	3.002231538	3.041143161	2.922275555	3.19201154	2.937319697	1.942345688	3.250528872
donovan2.xml	3.013258787	2.65610827	2.988078986	2.813349016	3.011374774	2.748147558	1.869706668	2.893820481
driscoll1.xml	2.800494097	2.946290878	2.833370652	2.559098173	2.928932144	2.683019106	3.018568911	3.326596092
driscoll2.xml	3.26804655	3.320896374	3.16994615	2.856864482	3.248027396	3.044317737	2.947012631	3.709384024
driscoll3.xml	3.215380659	3.351376172	3.022400453	2.877551482	3.307455743	3.092626771	2.837238017	3.575723611
Edgeworth1.xml	3.739932732	3.003865843	4.267383564	3.596325528	3.470395118	3.32814493	3.821859579	3.366754462
jessop1.xml	3.058174318	3.391432064	2.649488456	2.839119863	3.238303372	3.134563699	3.149106527	3.279860156
jessop2.xml	2.539824108	2.973229784	2.230047953	2.52919168	2.980204209	2.737161038	2.740866396	2.755374969
jessop3.xml	3.435887575	3.525877353	3.099797727	3.107884257	3.581301612	3.284872185	2.750448318	3.644718969
kyne1.xml	1.593219719	2.006946272	1.562601528	1.638422826	1.868172349	1.577675146	3.086472948	1.988732026
kyne2.xml	1.616915309	2.071493428	1.464882487	1.509869997	1.836933736	1.563089453	2.833702222	2.111666024
LeFanu1.xml	2.984346226	2.372679987	3.514904299	2.824324284	2.772668825	2.495606024	3.226656378	3.352776275
LeFanu2.xml	2.444718647	2.209096447	2.599204712	2.205244924	2.293600828	2.086120298	3.198023254	3.014117507
LeFanu3.xml	4.080713935	3.720972789	4.094475464	3.75301751	3.996250941	3.659111478	3.018636654	4.483533431

Distances from *Pride and Prejudice*

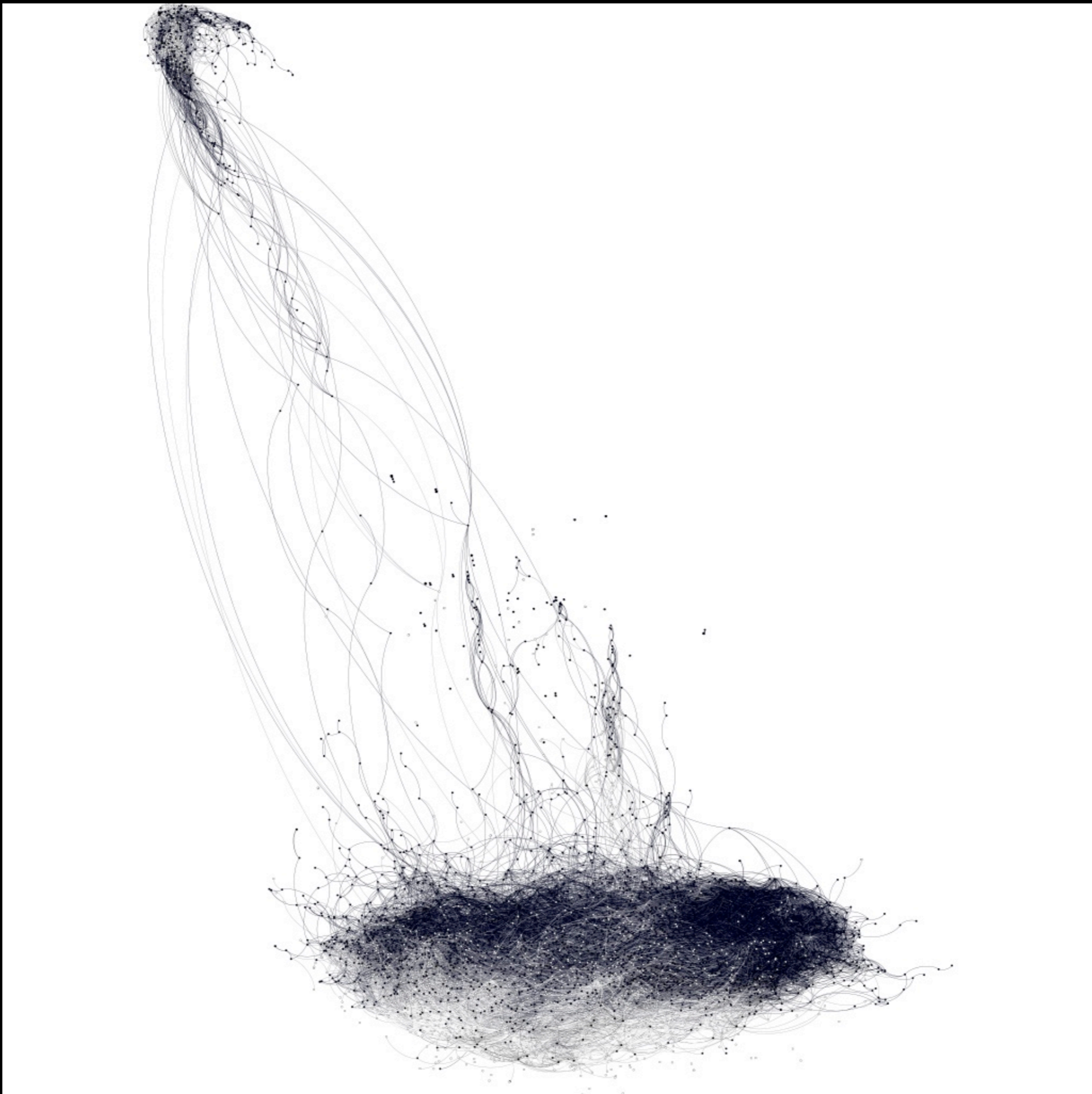
Author	Title	Distance
Austen, Jane	<i>Pride and Prejudice</i>	0.000000
Austen, Jane	<i>Emma</i>	1.260236
Austen, Jane	<i>Sense and Sensibility</i>	1.268725
Austen, Jane	<i>Mansfield Park</i>	1.421373
Austen, Jane	<i>Northanger Abbey</i>	1.600394
Austen, Jane	<i>Persuasion</i>	1.673071
Gaskell, Elizabeth	<i>Ruth</i>	1.716687
Craik, Dinah Maria	<i>Olive</i>	1.745832
Church A. B. Mrs.	<i>Greymore a Story of Country Life</i>	1.747513
Grant, Louisa	<i>Charles Stanley</i>	1.765758
Tainsh, Edward Campbell	<i>One Maiden Only</i>	1.767951

Distances from a *Tale of Two Cities*

Author	Title	Distance
Dickens, Charles	<i>A Tale of Two Cities</i>	0.000000
Kirkland, Caroline Matilda	<i>The Fountain and the Bottle</i>	1.361071
Milman, Edward Augustus	<i>Arthur Conway; or, Scenes in the Tropics</i>	1.385395
Liddell Charles Francis	<i>Hidden Links; or, The Schoolfellows</i>	1.466322
Dickens, Charles	<i>The Old Curiosity Shop</i>	1.492650
Armstrong, Francis Claudius	<i>The Pirates of The Foam</i>	1.500570
Spofford, Harriet Elizabeth Prescott	<i>Sir Rohan's Ghost</i>	1.508781
Fay, Theodore Sedgwick	<i>Norman Leslie; A Tale of the Present Times</i>	1.509204
Shillaber Benjamin Penhallow	<i>Knitting Work; A Web of Many Textures</i>	1.534282
Dickens, Charles	<i>Barnaby Rudge</i>	1.544074
Paulding, James Kirke	<i>Chronicles of the City of Gotham</i>	1.548381







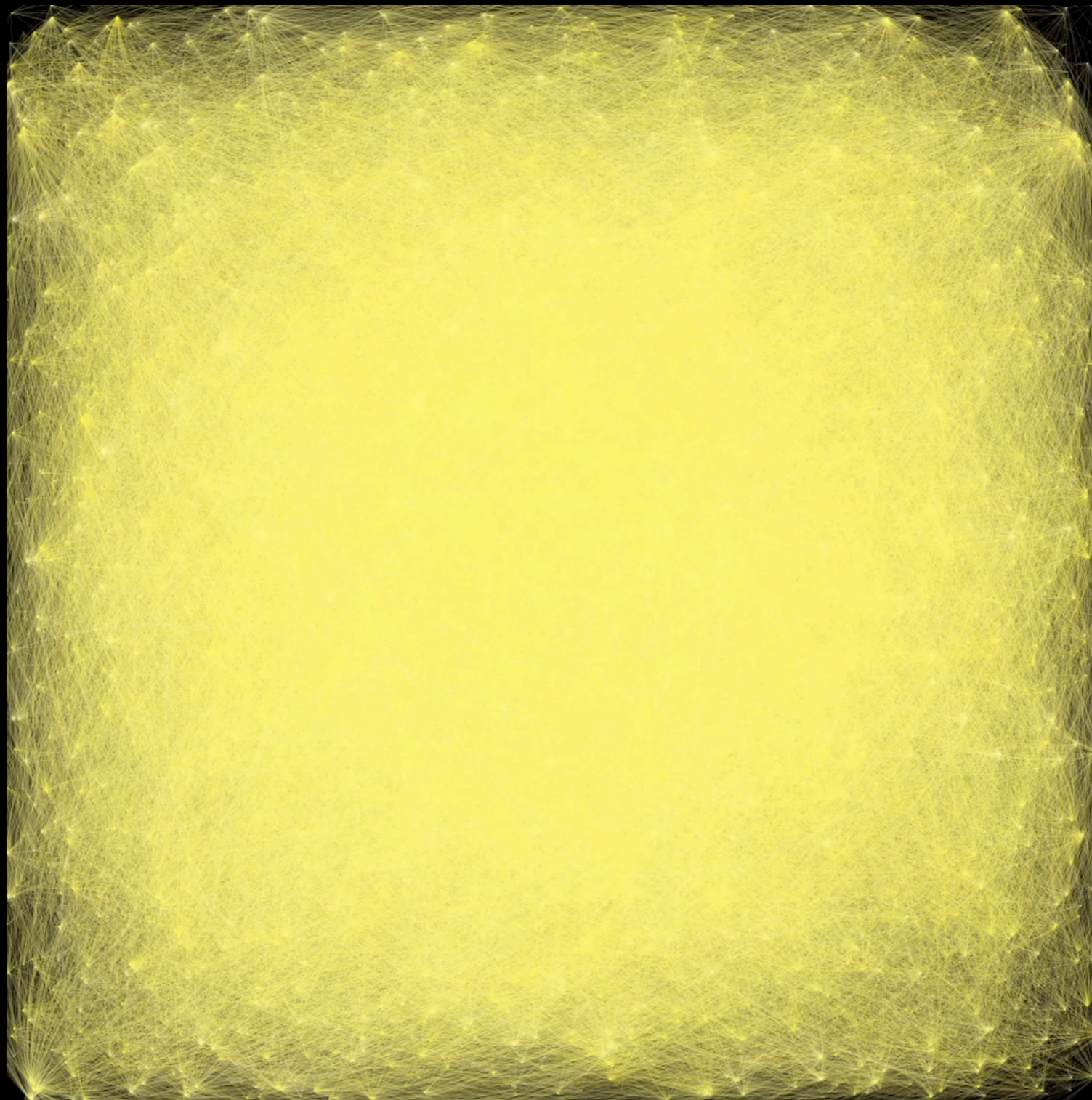


Image Credit: Elijah Meeks