

Thirty years of research into hate speech: topics of interest and their evolution

Supplementary Material

Alice Tontodimamma

Department of Neuroscience & Imaging, University G. d'Annunzio of Chieti-Pescara

Eugenia Nissi

Department of Economics, University G.d'Annunzio of Chieti-Pescara

Annalina Sarra

Department of Legal and Social Sciences, University G.d'Annunzio of Chieti-Pescara

Lara Fontanella

Department of Legal and Social Sciences, University G.d'Annunzio of Chieti-Pescara

Contents

Contents	1
1 Bibliographic data set	2
1.1 Search strategy	2
1.2 Bibliometric information	2
1.2.1 Research fields	2
1.2.2 Source	2
1.2.3 Countries' productivity and cooperation	4
1.2.4 Authors' productivity	7
2 Topic analysis	8
2.1 Word relevance for the identified topics	8
2.2 Topic temporal evolution	11
2.3 Topic co-occurrence networks	12

Bibliographic data set

1.1 Search strategy

Our search strategy was based on the following query: TITLE-ABS-KEY(“Hate speech” OR ((“hate” OR “hatred” OR “abusive language” OR “abusive discourse” OR “abusive speech” OR “offensive language” OR “offensive discourse” OR “offensive speech” OR “denigratory language” OR “denigratory discourse” OR “denigratory speech”) AND (“online” OR “social media” OR “web” OR “virtual” OR “cyber” OR “Orkut” OR “Twitter” OR “Facebook” OR “Reddit” OR “Instagram” OR “Snapchat” OR “Youtube” OR “Whatsapp” OR “Wechat” OR “QQ” OR “Tumblr” OR “Linkedin” OR “Pinterest ”)))

1.2 Bibliometric information

The data for this study were retrieved from the Scopus database on December 15, 2019 and the first document coherent with this query was published in 1992, so time span sets from 1992 to 2019. All types of publications were included in the search and 1614 documents related to hate speech were identified.

1.2.1 Research fields

Table 1.1 shows the research fields as classified by Scopus.

Table 1.1: Document distribution by research field	
Field of publication	Documents
Social Sciences	1028
Computer Science	460
Arts and Humanities	405
Psychology	120
Engineering	98
Medicine	84
Business, Management and Accounting	68
Mathematics	68
Decision Sciences	39
Economics, Econometrics and Finance	32

1.2.2 Source

Table 1.2 shows the top sources with more than seven documents.

Table 1.2: Document distribution by source

Source title	number of documents	Article	Book	Book Chapter	Conference Paper	Conference Review	Editorial	Note	Review
CEUR Workshop Proceedings	37				33	4			
Lecture Notes in Computer Science	36	1			32	3			
The Content and Context of Hate Speech: Rethinking Regulation and Responses	25		1	22			2		
ACM International Conference Proceeding Series	15				14	1			
Extreme Speech and Democracy	13		1	12					
Lodz Papers in Pragmatics	10	10							
Ethical Theory and Moral Practice	9	9							
Information and Communications Technology Law	9	9							
Communications in Computer and Information Science	8				6	2			
Computers in Human Behavior	8	7							1
Hate, Politics, Law: Critical Perspectives on Combating Hate	8		1	6			1		
Howard Journal of Communications	8	7							1
Religion and Human Rights	8	7							1
Social Media and Society	8	8							
First Monday	7	7							
Journal of Ethnic and Migration Studies	7	7							
Proceedings of the ACM on Human-Computer Interaction	7	7							
Res Publica	7	7							
Society	7	6						1	
Speech and Harm: Controversies Over Free Speech	7		1	5			1		
Teorija in Praksa	7	6							1

1.2.3 Countries' productivity and cooperation

Considering the country of affiliation of each author, Table 1.3 shows the countries with at least twenty documents, while Figure 1.1 represents all the countries.

Table 1.3: Document distribution by country

Country	Number of documents	Country	Number of documents
United States	709	Brazil	36
United Kingdom	302	Russian Federation	28
Italy	107	Hong Kong	28
India	87	Sweden	24
Australia	81	Ireland	24
Indonesia	75	China	24
Spain	74	Austria	23
Canada	66	Denmark	23
Germany	61	Nigeria	22
Finland	53	France	22
South Africa	47	Portugal	21
Malaysia	40	Singapore	21
Netherlands	39	Greece	20
Cyprus	37		

Figure 1.1: Document distribution by country



Figure 1.2 shows the network of collaboration between authors belonging to different countries and the network measures for each node are represented in Figure 1.3

Figure 1.2: Country cooperation network

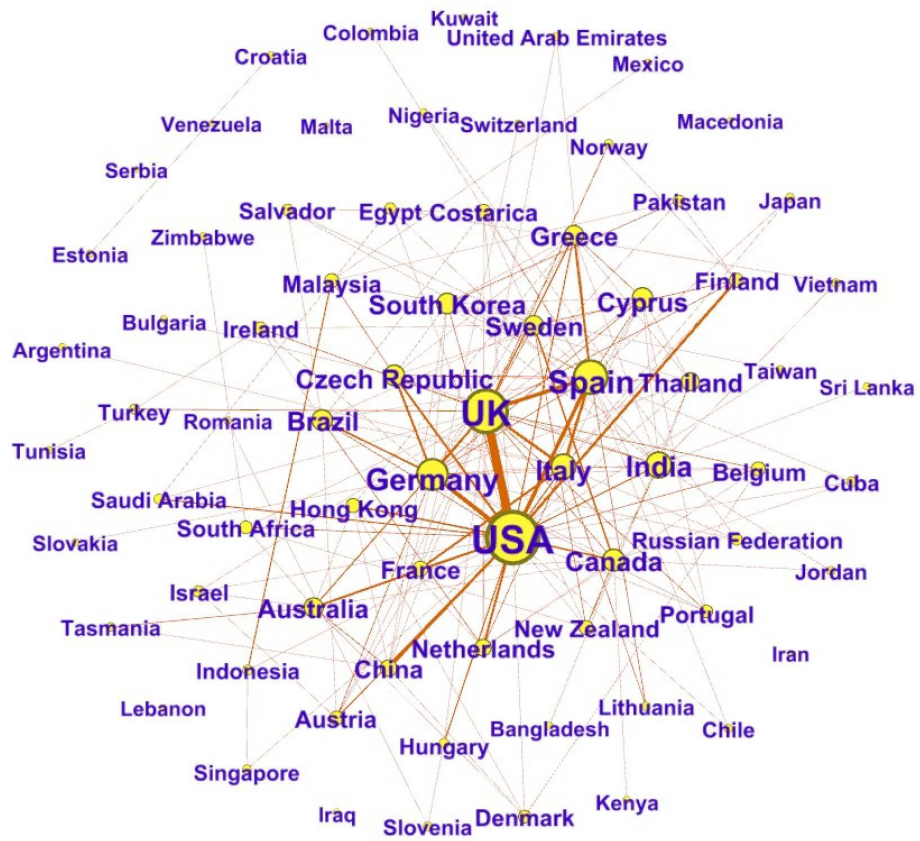
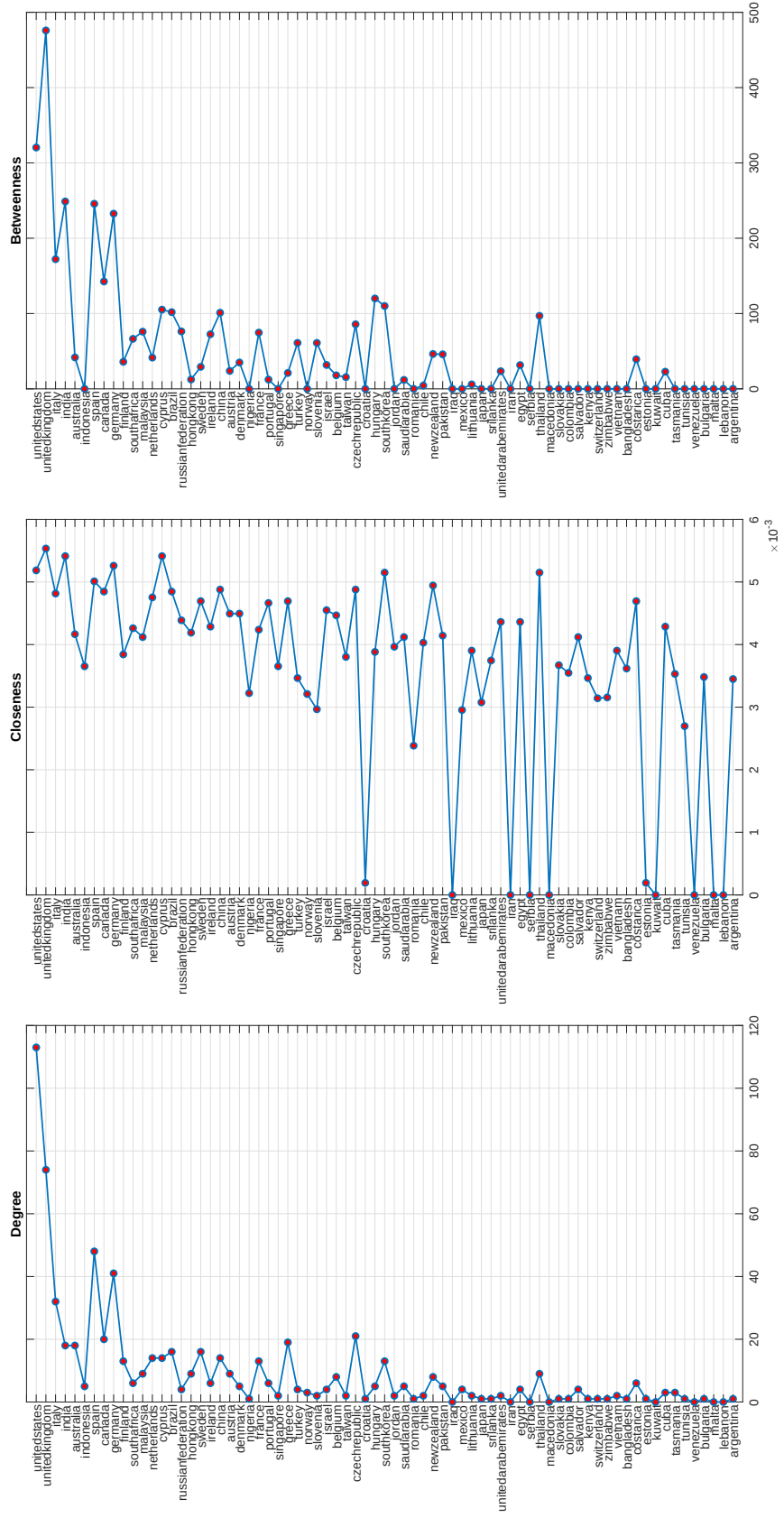


Figure 1.3: Centrality measures for the country cooperation network



1.2.4 Authors' productivity

The number of single authored document is equal to 836 and the average number of authors per document is 2.05 (Table 1.4).

Table 1.4: Document distribution by number of authors

number of authors	documents
1	836
2	339
3	163
4	108
5	74
6	31
7	13
8	6
9	1
10	2
12	2
13	1
15	1
19	1
Total	1578

Table 1.5 shows the authors with more than six publications.

Table 1.5: Most productive authors

Author	Affiliation	number of publications
Oksanen A.	Faculty of Social Sciences, University of Tampere, Finland	14
Räsänen P.	Faculty of Social Sciences, University of Turku, Finland	14
Hawdon J.	Department of Sociology, Virginia Tech University, United States	13
Cohen-Almagor R.	Department of Politics, The University of Hull, Hull, United Kingdom	11
Bosco C.	Computer Science Department, University of Turin, Italy	10
Patti V.	Computer Science Department, University of Turin, Italy	10
Awan I.	Centre for Applied Criminology Birmingham City University, United Kingdom	9
Gelber K.	School of Political Science and International Studies, University of Queensland, Australia	9
Sureka A.	ABB Corporate Research, Bangalore, India	8
Budi I.	Faculty of Computer Science, Universitas Indonesia, Indonesia	7
Jansen B.J.	Qatar Computing Research Institute, University of Turku, Finland	7
Salminen J.	Qatar Computing Research Institute, University of Turku, Finland	7

Topic analysis

2.1 Word relevance for the identified topics

Tables 2.1, 2.2 and 2.3 show the relevance of the 25 most coherent terms among the top 30 relevant words.

Table 2.1: Word relevance for topics 1-4

Topic 1		Topic 2		Topic 3		Topic 4	
<i>Word</i>	<i>Score</i>	<i>Word</i>	<i>Score</i>	<i>Word</i>	<i>Score</i>	<i>Word</i>	<i>Score</i>
speech	1.969	political	0.513	freedom	0.366	online	0.291
hate	0.432	politics	0.167	law	0.298	social	0.269
free	0.222	discourse	0.104	rights	0.276	users	0.211
harm	0.127	democracy	0.086	speech	0.254	network	0.190
freedom	0.095	election	0.075	expression	0.227	facebook	0.189
regulation	0.051	speech	0.070	legal	0.169	web	0.182
legal	0.048	democratic	0.063	court	0.147	networks	0.155
theory	0.039	party	0.063	religious	0.137	networking	0.140
amendment	0.038	populism	0.050	laws	0.120	youtube	0.136
harms	0.036	campaign	0.044	incitement	0.090	communities	0.107
liberal	0.028	power	0.043	constitutional	0.084	sites	0.103
authority	0.027	slurs	0.043	genocide	0.080	information	0.085
speakers	0.026	parties	0.043	amendment	0.075	videos	0.084
harmful	0.025	populist	0.042	human	0.071	user	0.076
values	0.023	media	0.039	denial	0.068	content	0.074
dignity	0.023	public	0.037	criminal	0.065	video	0.063
debate	0.022	rightwing	0.037	jurisprudence	0.060	groups	0.046
censorship	0.021	debate	0.036	international	0.060	community	0.044
equality	0.020	tolerance	0.036	protection	0.060	systems	0.043
hateful	0.020	politicians	0.034	courts	0.058	websites	0.042
counterspeech	0.018	presidential	0.034	legislation	0.057	internet	0.041
doctrine	0.016	elections	0.033	supreme	0.056	media	0.038
universities	0.015	nationalism	0.032	free	0.052	sharing	0.038
campus	0.014	citizenship	0.029	holocaust	0.048	mining	0.038
academic	0.014	rights	0.027	defamation	0.048	blogs	0.038

Table 2.2: Word relevance for topics 5-8

Topic 5		Topic 6		Topic 7		Topic 8	
<i>Word</i>	<i>Score</i>	<i>Word</i>	<i>Score</i>	<i>Word</i>	<i>Score</i>	<i>Word</i>	<i>Score</i>
racism	0.213	internet	0.389	terrorism	0.151	twitter	0.516
racist	0.145	information	0.189	religious	0.140	media	0.357
trolling	0.118	media	0.179	muslim	0.115	social	0.333
online	0.105	news	0.130	violence	0.087	tweets	0.229
women	0.096	digital	0.087	terrorist	0.075	sentiment	0.190
white	0.093	communication	0.069	extremist	0.075	content	0.178
race	0.081	privacy	0.067	muslims	0.060	analysis	0.132
digital	0.072	technology	0.066	global	0.059	platforms	0.094
cultural	0.068	brand	0.062	war	0.058	online	0.087
feminist	0.066	social	0.060	groups	0.058	networking	0.082
harassment	0.057	fake	0.050	extremism	0.055	data	0.072
gender	0.053	companies	0.042	religion	0.049	messages	0.050
racial	0.053	journalism	0.041	group	0.049	mining	0.047
identity	0.048	technologies	0.038	antimuslim	0.046	platform	0.042
critical	0.047	ethics	0.038	attacks	0.046	classification	0.041
media	0.045	team	0.037	islamic	0.045	microblogging	0.040
gendered	0.044	responsibility	0.034	islam	0.042	tweet	0.037
culture	0.042	hoax	0.033	virtual	0.040	information	0.035
games	0.039	public	0.032	hatred	0.039	posts	0.035
black	0.037	liability	0.032	islamophobia	0.034	bots	0.031
whiteness	0.037	illegal	0.031	fear	0.032	detection	0.029
movement	0.036	brands	0.028	migration	0.032	negative	0.029
practices	0.035	consumer	0.028	education	0.032	semantic	0.028
misogyny	0.034	consumers	0.026	ethnic	0.031	dataset	0.027
experiences	0.033	business	0.026	isis	0.025	offensive	0.027

Table 2.3: Word relevance for topics 9-12

Topic 9		Topic 10		Topic 11		Topic 12	
<i>Word</i>	<i>Score</i>	<i>Word</i>	<i>Score</i>	<i>Word</i>	<i>Score</i>	<i>Word</i>	<i>Score</i>
comments	0.378	online	0.281	sexual	0.290	learning	0.393
discourse	0.167	exposure	0.107	health	0.270	detection	0.355
language	0.154	crime	0.104	gender	0.189	classification	0.320
emotion	0.153	behavior	0.098	gay	0.114	language	0.202
analysis	0.145	young	0.081	transgender	0.101	machine	0.172
emotions	0.106	victimization	0.075	discrimination	0.096	text	0.170
linguistic	0.094	cyberbullying	0.069	lgbt	0.074	neural	0.125
attitudes	0.072	prejudice	0.069	lesbian	0.069	automatic	0.112
emotional	0.066	hate	0.069	sex	0.064	deep	0.080
corpus	0.061	adolescents	0.063	queer	0.058	classifier	0.072
readers	0.054	victims	0.057	violence	0.056	natural	0.072
negative	0.054	cyberhate	0.054	minorities	0.055	dataset	0.069
news	0.040	school	0.052	female	0.055	features	0.065
semantic	0.033	suicide	0.051	sexuality	0.053	vector	0.064
facebook	0.033	youth	0.046	homosexuality	0.041	feature	0.056
expressions	0.033	perceived	0.039	homophobia	0.040	word	0.055
instagram	0.033	coping	0.038	womens	0.040	detecting	0.051
image	0.032	aggression	0.037	orientation	0.037	network	0.050
speech	0.031	victim	0.037	bisexual	0.035	task	0.050
journalists	0.030	crimes	0.036	psychology	0.035	networks	0.049
nooj	0.028	psychology	0.034	minority	0.035	models	0.047
discursive	0.027	behaviors	0.033	women	0.035	performance	0.047
linguistics	0.024	schools	0.033	mental	0.033	recognition	0.045
rhetorical	0.024	risk	0.032	psychological	0.033	algorithms	0.044
texts	0.024	emotional	0.031	lgbtq	0.029	sentiment	0.043

2.2 Topic temporal evolution

Figures 2.1 and 2.2 show the temporal evolution of the research activity in the identified topics.

Figure 2.1: Heatmap of the number of documents assigned to the identified topics by year

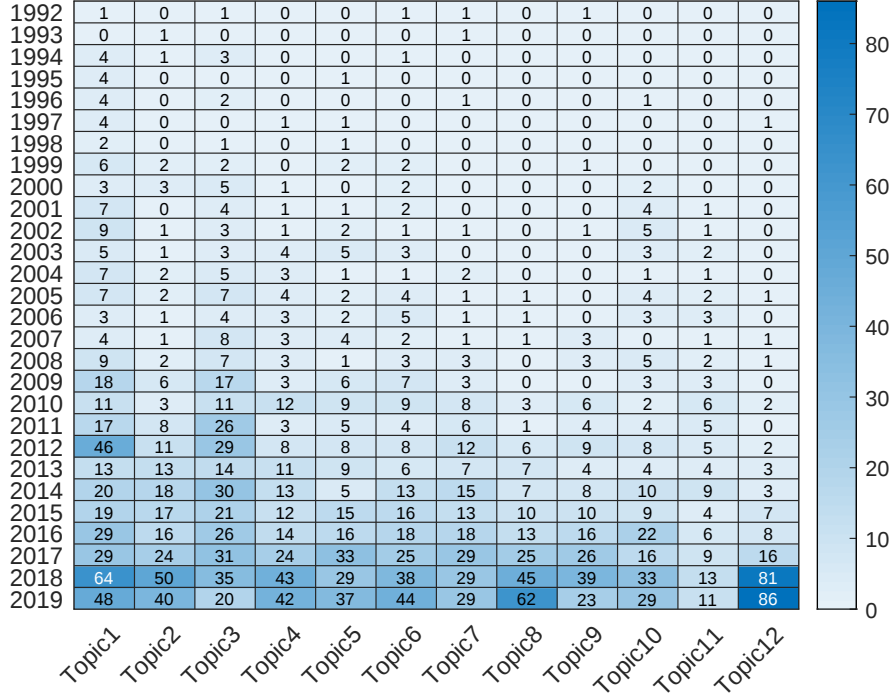
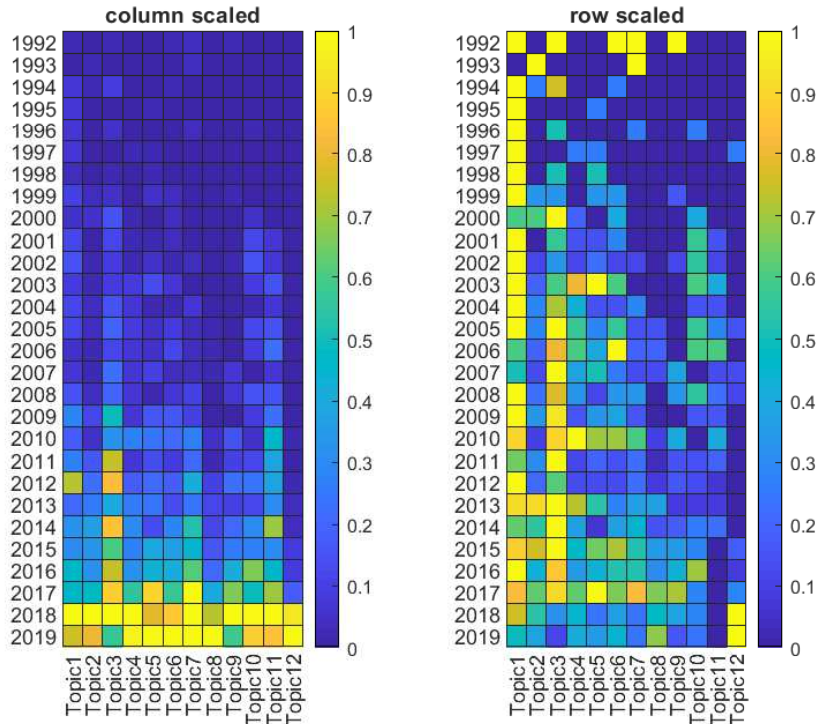


Figure 2.2: Heatmap of the number of documents assigned to the identified topics by year scaled by row and by column



2.3 Topic co-occurrence networks

Table 2.4 shows the links in the topic co-occurrence network for the publication on hate speech from 1992 to 2019, as well the links for the networks retrieved distinguishing three different stages in the historical development of online hate speech research.

Table 2.4: Co-occurrence networks for the publication on hate speech from 1992 to 2019

Topic	1992-2019		1992-2009		2010-2015		2016-2019	
	Connected Topics	# of connections	Connected Topics	# of connections	Connected Topics	# of connections	Connected Topics	# of connections
Topic 1	2, 3, 4, 5, 6, 7, 9, 10	8	2, 3, 4, 5, 6, 10	6	2, 3, 7, 9	4	2, 3, 4, 6, 7, 8, 910	8
Topic 2	1, 3, 5, 7, 9	5	1, 3, 5, 6	4	1, 3, 5, 7	4	1, 3, 4, 5, 6, 7, 9	7
Topic 3	1, 2, 6, 7	4	1, 2, 5, 6, 11	5	1, 2, 6, 7, 11	5	1, 2, 6	3
Topic 4	1, 5, 6, 7, 8, 9, 10, 12	8	1, 6, 7, 12	4	5, 6, 8, 12	4	1, 2, 5, 6, 8, 91012	8
Topic 5	1, 2, 4, 7	4	1, 2, 3	3	2, 4, 6, 7	4	2, 4, 8, 9	4
Topic 6	1, 3, 4, 10	4	1, 2, 3, 4, 10	5	3, 4, 5, 7, 9	5	1, 2, 3, 4, 812	6
Topic 7	1, 2, 3, 4, 5, 10	6	4	1	1, 2, 3, 5, 6, 10	6	1, 2, 8, 910	5
Topic 8	4, 12	2		0	4, 12	2	1, 4, 5, 6, 7, 91012	8
Topic 9	1, 2, 4	3		0	1, 6	2	1, 2, 4, 5, 7, 812	7
Topic 10	1, 4, 6, 7, 11	5	1, 6, 11	3	7, 11	2	1, 4, 7, 811	5
Topic 11	10	1	3, 10	2	3, 10	2	10	1
Topic 12	4, 8	2	4	1	4, 8	2	4, 6, 8, 9	4