

Connecting the Dots in a Topic Model

A Network Representation of Topics and Periodicals

Sanja Sarić
Centre for Information Modelling, University of Graz

Workshop on Information Visualization in the (Digital) Humanities
27.10.2022

01

INTRODUCTION

The corpus and method

02

VISUALIZING TOPICS

Visualization of topic modeling results

03

NETWORKS OF TOPICS

How can networks be used to visualize
topic modeling results

04

CONCLUSION AND DISCUSSION

Lessons learned and future work

introduction

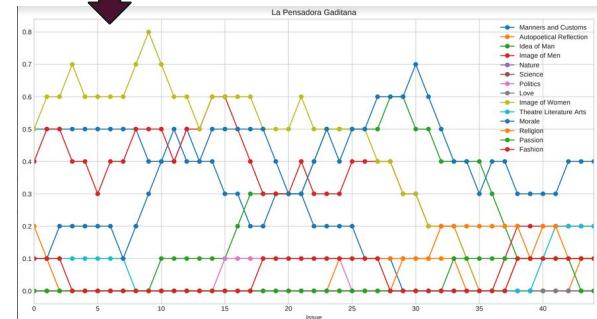
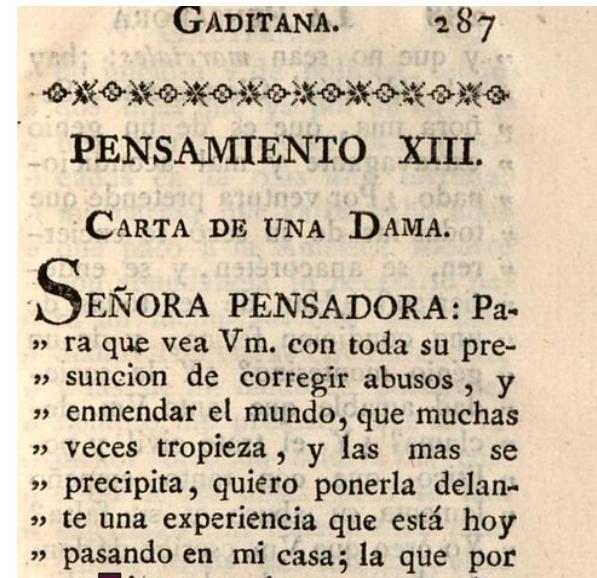
From Spectators ...

- digital edition *The Spectators in the international context*
<http://gams.uni-graz.at/spectators> (Ertler et al. 2011-2018)
- journalistic genre in the 18th century
- created in England
- imitations and translations in Europe

... to Distant Spectators

- text mining to extend close reading expertise
- explorative approach
- project team: Martina Scholger, Bernhard Geiger, Elisabeth Hobisch, Philipp Koncar, Sanja Sarić, Yvonne Völkl, Christina Glatz

<http://gams.uni-graz.at/dispecs> (Scholger et al. 2018-2021)



Topic Modeling

- Python library *dariah_topics* (DARIAH-DE 2019) with MALLET (McCallum 2002-2018) based on Latent Dirichlet Allocation (LDA)
- adapted and expanded Jupyter Notebooks for DiSpecs: <https://github.com/distspectators/DiSpecs>
- individual analysis of the French, Spanish and Italian sub-corpora

Topic Modeling with LDA

→ unsupervised probabilistic method that aims to identify hidden semantic structures in large text collections

Principle

→ frequently co-occurring words are most likely thematically related

Topics

→ lists of thematically or structurally related document tokens

Topics are not always subjects!

All topics contain all corpus tokens!

01 INTRODUCTION

|Topic | Probability | Topic Keys |

0	0,23769	auteur lettre ouvrage livre écrire public
1	0,19942	crime mort coup ennemi honneur malheureux
2	0,16349	nature idée ame corps imagination objet ph
3	0,10047	manière savoir espece propre caractere mer
4	0,13858	fille jeune pere mariage fils mere mari fa
5	0,14065	madame ami disje estce étois oui aime pour
6	0,40803	penser vérité mal vrai honneur ami réflexi
7	0,1569	vin ville boire table passer savoir heure
8	0,15418	manière savoir idée imagination pourtant s
9	0,62005	vertu mérite vie nature ame hommes propre
10	0,1401	argent riche payer fortune pauvre marchand
11	0,10668	terre eau mer bois arbre montagne air coul
12	0,13441	poète auteur théâtre art goût comédie trag
13	0,09296	paris ville avanture madame maître part te
14	0,1687	manière égard discours vie ami lieu servir
15	0,09396	savoir temps peuple liberté révolution lor
16	0,14837	vie mort mourir mal maladie malade douleur
17	0,04122	réflexion pére jeu médisance exemple estil
18	0,13054	heureux dieu doux oeil tendre main ciel am
19	0,07902	religion dieu eglise saint divinité foi ch
20	0,14865	roi prince peuple france guerre etat polit
21	0,26827	aimer sentir bonheur passion lettre amant
22	0,10836	dame quoiqu égard manière passion tem sexe
23	0,42768	maison demander heure chambre paraître jeu
24	0,2317	air bel sexe dame beauté jeune visage oeil

|Document

Van-Effen_Vol-1_Nr-001_2948_0000.txt	0.5245948920304114	0.003459821565001878
Van-Effen_Vol-1_Nr-002_2949_0000.txt	0.2561009284068167	0.029435135491672523
Van-Effen_Vol-1_Nr-002_2949_0001.txt	0.007006822238497824	0.09736989171939585
Van-Effen_Vol-1_Nr-002_2949_0002.txt	0.26793749798185273	0.0012650613140997428
Van-Effen_Vol-1_Nr-002_2949_0003.txt	0.32278118362822367	0.0012813175561526499
Van-Effen_Vol-1_Nr-003_2950_0000.txt	0.01000574528035826	0.09926407415213746
Van-Effen_Vol-1_Nr-003_2950_0001.txt	0.004051546563875589	0.05709483002660381
Van-Effen_Vol-1_Nr-003_2950_0002.txt	0.1537533046332643	0.0012650613140997428
Van-Effen_Vol-1_Nr-004_2951_0000.txt	0.00780184711090586	0.013864242317971957
Van-Effen_Vol-1_Nr-004_2951_0001.txt	0.051025129888380254	0.01096203947566564
Van-Effen_Vol-1_Nr-004_2951_0002.txt	0.001208734122883809	0.09255196299733717
Van-Effen_Vol-1_Nr-005_2952_0000.txt	0.06394934052406501	0.06377748822116623
Van-Effen_Vol-1_Nr-005_2952_0001.txt	0.0013610024477733495	0.0011419163651037634
Van-Effen_Vol-1_Nr-005_2952_0002.txt	0.001044128840339357	8.760511871312107E-4
Van-Effen_Vol-1_Nr-005_2952_0003.txt	0.0033177740543805082	0.0027386951179334
Van-Effen_Vol-1_Nr-006_2953_0000.txt	0.21444079541369898	0.08633243674280665
Van-Effen_Vol-1_Nr-006_2953_0001.txt	0.13308203138668756	0.06432936057638246
Van-Effen_Vol-1_Nr-006_2953_0002.txt	0.03502556696807462	0.005804413250973456
Van-Effen_Vol-1_Nr-006_2953_0003.txt	0.04363144633467837	8.499160161578237E-4
Van-Effen_Vol-1_Nr-007_2954_0000.txt	0.018615727832998863	8.760511871312107E-4
Van-Effen_Vol-1_Nr-007_2954_0001.txt	0.023459277854650615	0.25575402352231136
Van-Effen_Vol-1_Nr-007_2954_0002.txt	0.22119526156151273	0.022446135437119603
Van-Effen_Vol-1_Nr-008_2955_0000.txt	0.001155832919434894	0.054461294575947984
Van-Effen_Vol-1_Nr-008_2955_0001.txt	0.010523348565625952	0.005640631891533191
Van-Effen_Vol-1_Nr-008_2955_0002.txt	0.05544671662429108	0.017327898417114997
Van-Effen_Vol-1_Nr-009_2956_0000.txt	0.0014524899853557624	0.0012186767827939164
Van-Effen_Vol-1_Nr-009_2956_0001.txt	0.04505529426233271	0.038592017586774156
Van-Effen_Vol-1_Nr-009_2956_0002.txt	0.0014796156582604037	0.0012414359261407476
Van-Effen_Vol-1_Nr-009_2956_0003.txt	0.0016900273386654993	0.001417976785164665
Van-Effen_Vol-1_Nr-010_2961_0000.txt	0.00132312075916181	0.2683109968674494
Van-Effen_Vol-1_Nr-010_2961_0001.txt	0.001196564029191859	0.15203087960070263
Van-Effen_Vol-1_Nr-010_2961_0002.txt	0.008685316585441557	0.1637921639385455
Van-Effen_Vol-1_Nr-010_2961_0003.txt	0.001537024414155714	0.040089379219018194
Van-Effen_Vol-1_Nr-011_2962_0000.txt	0.0012667100508796407	0.04902697180431327
Van-Effen_Vol-1_Nr-011_2962_0001.txt	0.006168685893638386	0.055818483478407514
Van-Effen_Vol-1_Nr-011_2962_0002.txt	0.006358842605096306	0.07808990441744103
Van-Effen_Vol-1_Nr-011_2962_0003.txt	0.0010871088834004143	0.17928746823196942
Van-Effen_Vol-1_Nr-012_2963_0000.txt	0.06221037031422289	0.001103987359407848

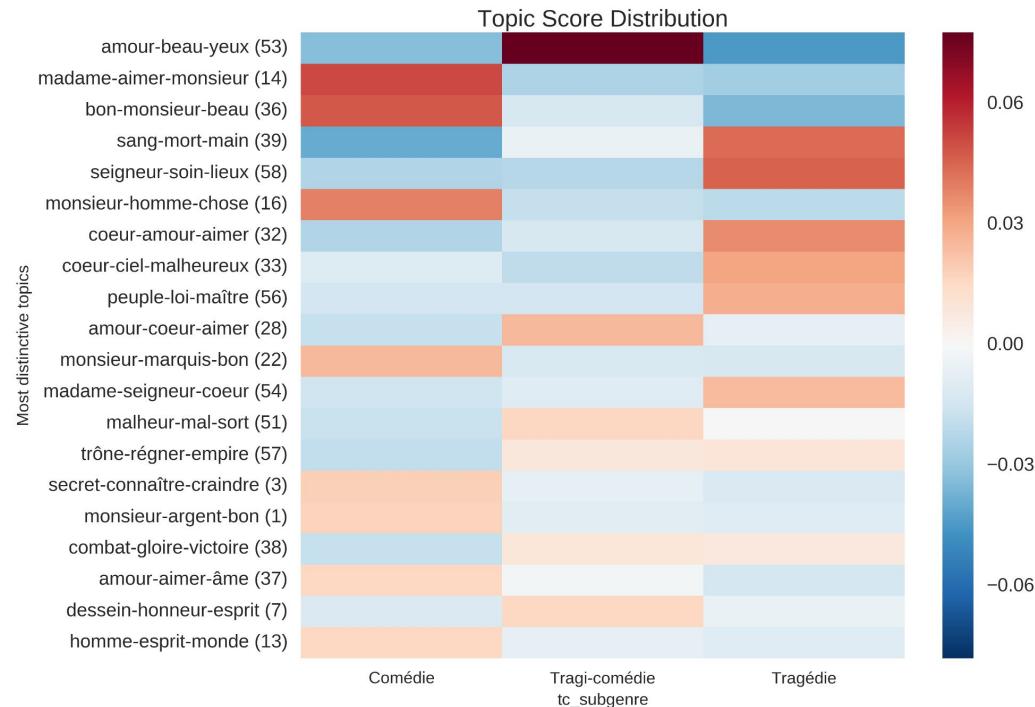
| Topic | Probability | Topic Keys |

0	0,23769	auteur lettre ouvrage livre écrire public	
1	0,19942	crime mort coup ennemi honneur malheureux	
2	0,16349	nature idée ame corps imagination objet ph	→ philosophical ideas
3	0,10047	manière savoir espece propre caractere mer	
4	0,13858	fille jeune pere mariage fils mere mari fa	→ marriage and family
5	0,14065	madame ami disje estce étois oui aime pour	
6	0,40803	penser vérité mal vrai honneur ami réflexi	→ (self-)reflection
7	0,1569	vin ville boire table passer savoir heure	
8	0,15418	manière savoir idée imagination pourtant s	
9	0,62005	vertu mérite vie nature ame hommes propre	→ virtuous behavior and practices
10	0,1401	argent riche payer fortune pauvre marchand	→ economy
11	0,10668	terre eau mer bois arbre montagne air coul	
12	0,13441	poète auteur théâtre art goût comédie trag	→ theatre and opera
13	0,09296	paris ville avanture madame maître part te	
14	0,1687	manière égard discours vie ami lieu servir	
15	0,09396	savoir temps peuple liberté révolution lor	
16	0,14837	vie mort mourir mal maladie malade douleur	→ life vs death
17	0,04122	réflexion père jeu médisance exemple estil	
18	0,13054	heureux dieu doux oeil tendre main ciel am	
19	0,07902	religion dieu eglise saint divinité foi ch	→ church and religion
20	0,14865	roi prince peuple france guerre etat polit	→ politics
21	0,26827	aimer sentir bonheur passion lettre amant	→ love discourse
22	0,10836	dame quoiqu égard manière passion tem sexe	→ narrative vocabulary and dialogue
23	0,42768	maison demander heure chambre paraître jeu	
24	0,2317	air bel sexe dame beauté jeune visage oeil	→ appearance of woman

visualizing topics

Common ways to represent Topics

Heat maps



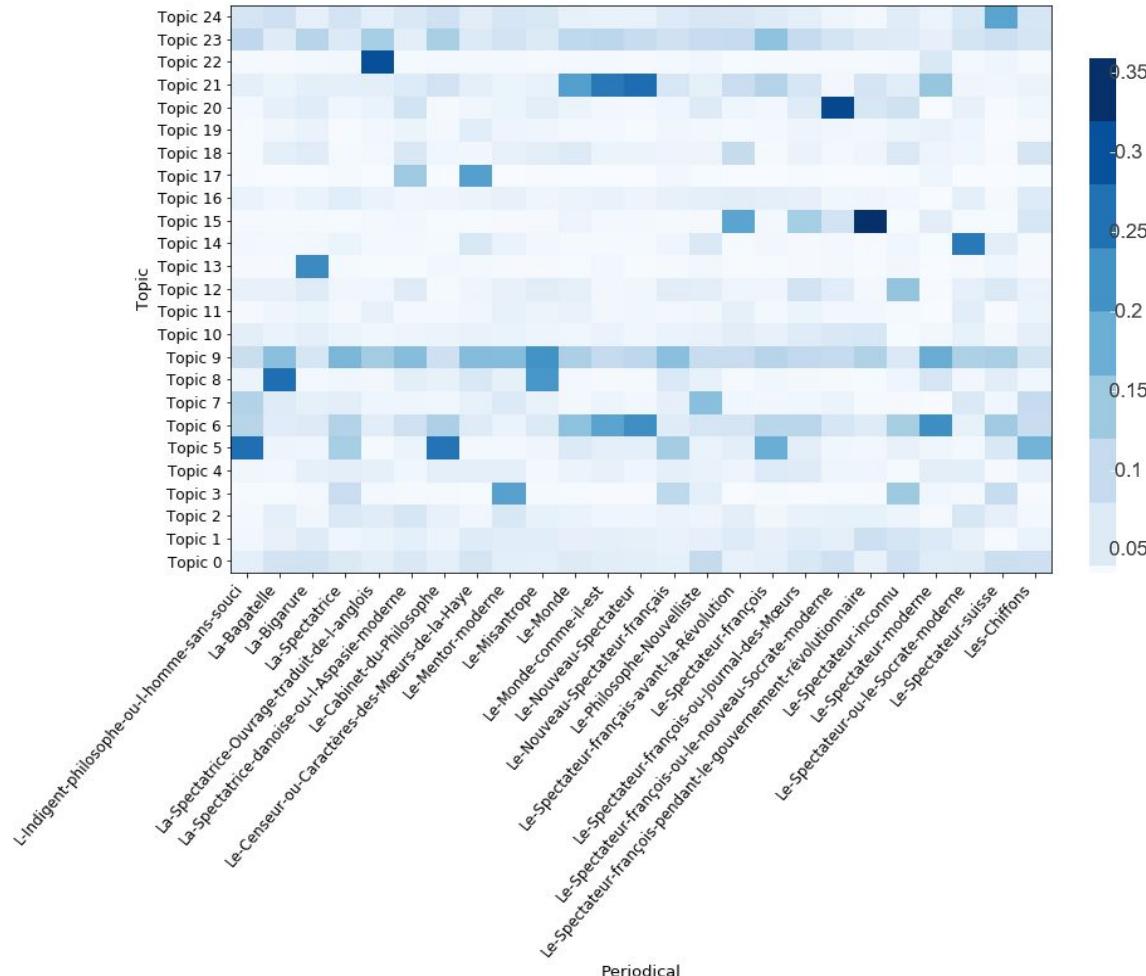
[Schöch 2017, Fig. 8](#)

Common ways to represent Topics

Heat maps

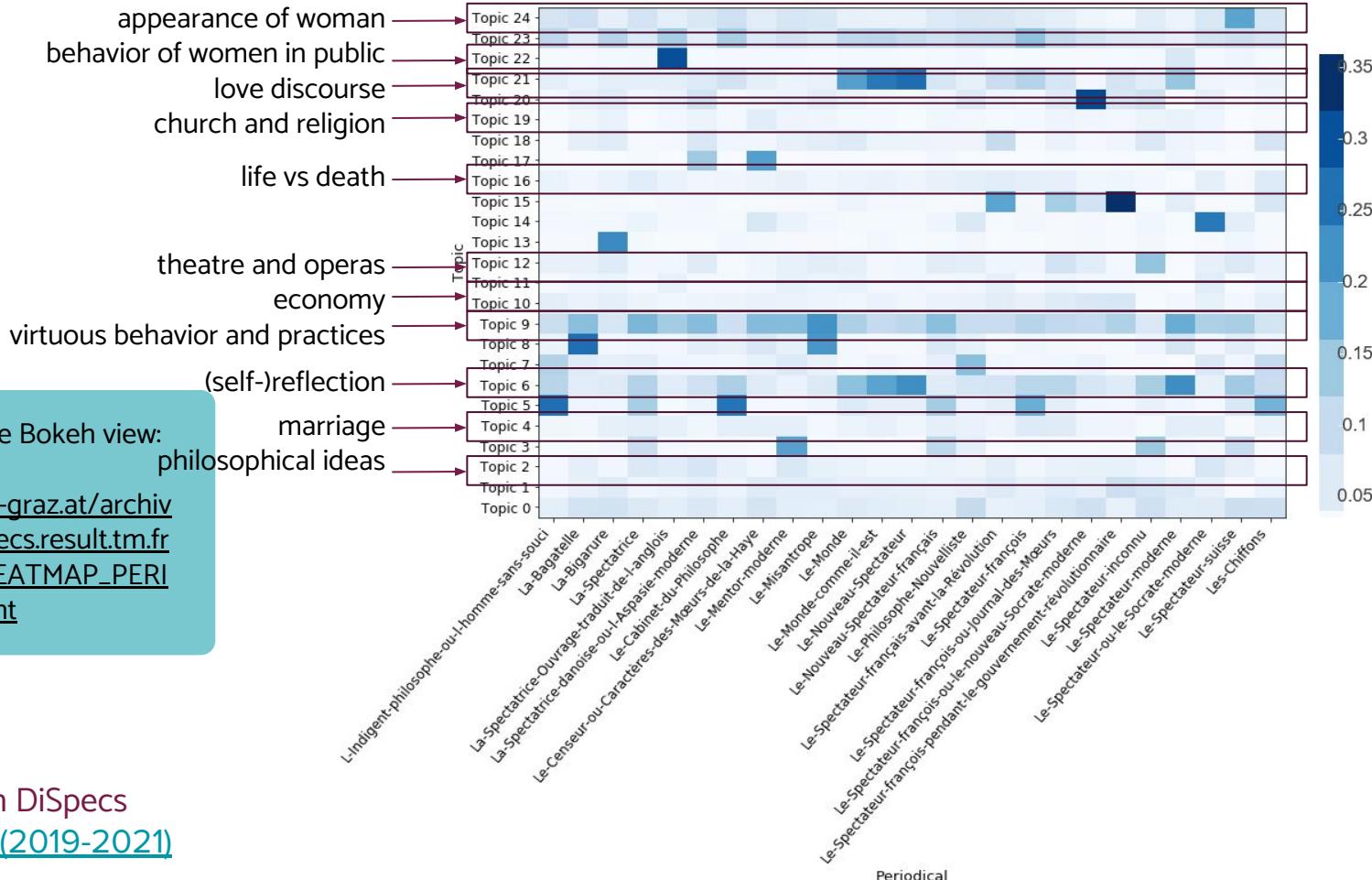
For the hoverable Bokeh view:

https://gams.uni-graz.at/archiv/e/objects/o:dispecs.result.tm.fr/datastreams/HEATMAP_PERIODICALS/content



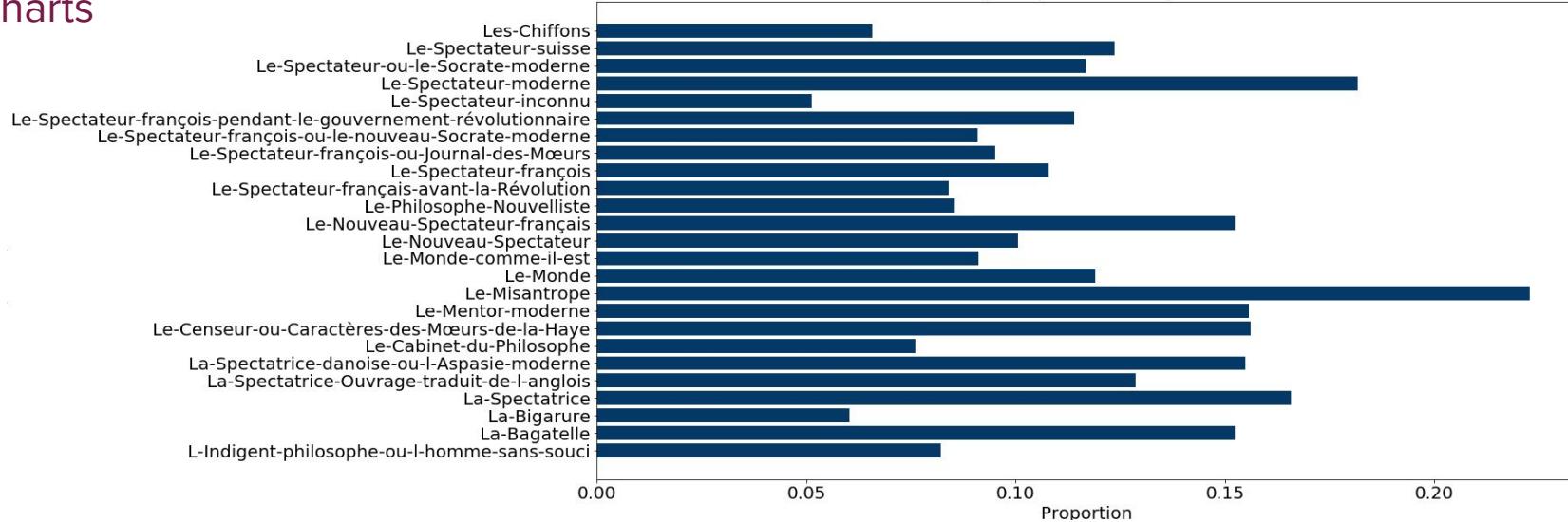
02 VISUALIZING TOPICS

For the hoverable Bokeh view:
https://gams.uni-graz.at/archiv/e/objects/o:dispecs.result.tm.fr/datastreams/HEATMAP_PERIODICALS/content



Common ways to represent Topics

Bar charts

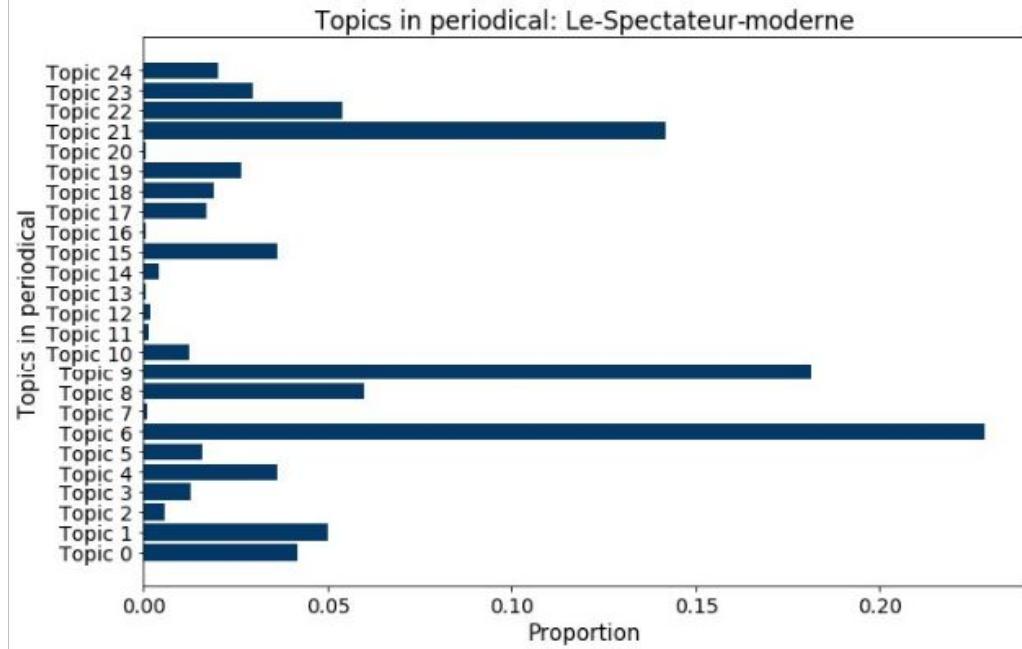


Topic 9 in French periodicals
virtuous behavior and practices
[Scholger et al. \(2019-2021\)](#)

Common ways to represent Topics

Bar charts

Topics in Le Spectateur moderne
Scholger et al. (2019-2021)



Common ways to represent Topics

Word clouds



Jockers 2016, Fig. 2

Topics:

- character, civilization, memphis, attributes, scythians, mercury, sphinx, symbols, customs, babylon, nymph, maidens, chariots, marble, idol, sacrifice, names, aticosins, giants, philosopher, festival, nymphs, death, orpheus, traditions, river, zeus, athenians, ancients, regions, beauty, order, odin, persians, mortals, chorus, supremacy, number, oracles, mythology, deities, goddesses, pan, lands, empire, war, oracle, spartans, statue, city, goddess, sacrifices, palace, sages, vata, form, wisdom, century, spartan, deity, greek, race, cities, persia, archon, incense, fountain, people, feet, throne, nation, arts, sea, priests, men, grecian, island, history, idols, citizens, conquest, gods, jupiter, divinities, tyrant, athens, egypt, greeks, earth, days, romans, earth, days, gold, delphi, egypt, druids, egyprians, sappho, plato, statues, papyrus, serpent, image, shores, chariot, tyre, heathen, ages, nile, herodotus, offerings, agamemnon, reverence, fathers, olympus, mysteries, poets, perditas, sons, treasures, acropolis, images, antiquity, pyramid, shrinetomb, samos, cambyses, religion, birth, idolatry, expedition, ceremonies, geniuses, worshippers, story, worshipper, tribes, legends, isles, poet, form, persian, races, pharaohs, hades, reign, perseus, mankind, emblem, pisistratus, garlands, custom, marathon, pythagoras.

Common ways to represent Topics

Word clouds



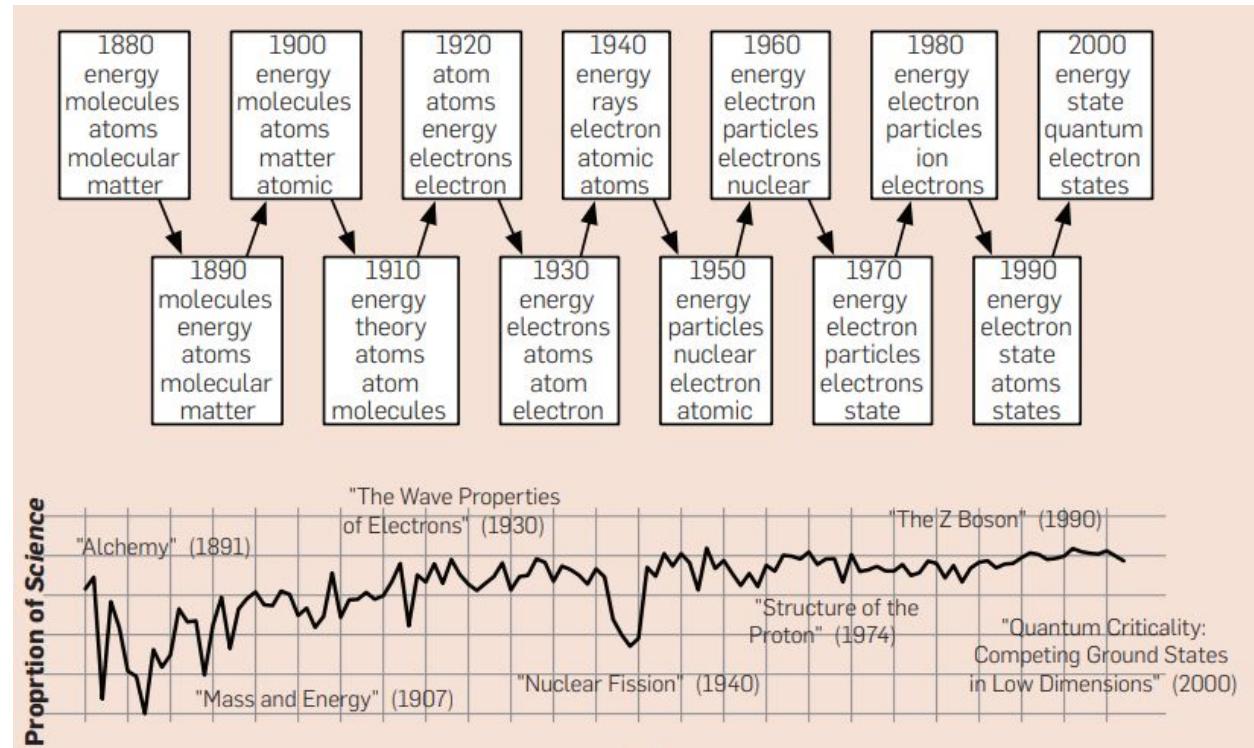
French Topic 0

elements of the spectatorial communication process

Scholger et al. (2019-2021)

Common ways to represent Topics

Line diagrams



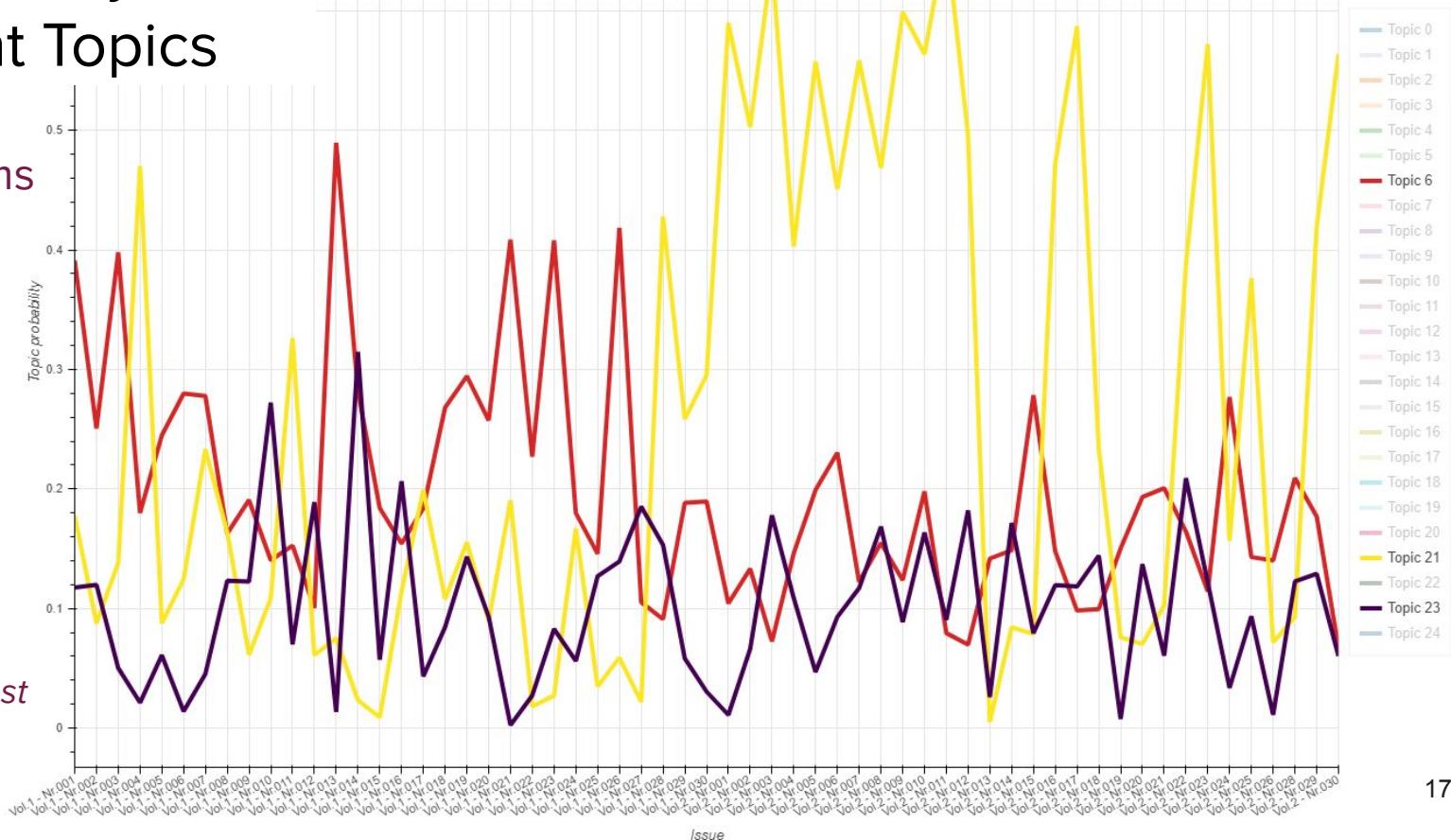
[Blei 2012, Fig. 5](#)

Common ways to represent Topics

Line diagrams

6: (self-)reflection
 21: love discourse
 23: narrative vocabulary and dialogue

Topic prevalence over time in *Le Monde comme il est*
 Scholger et al.
 (2019-2021)



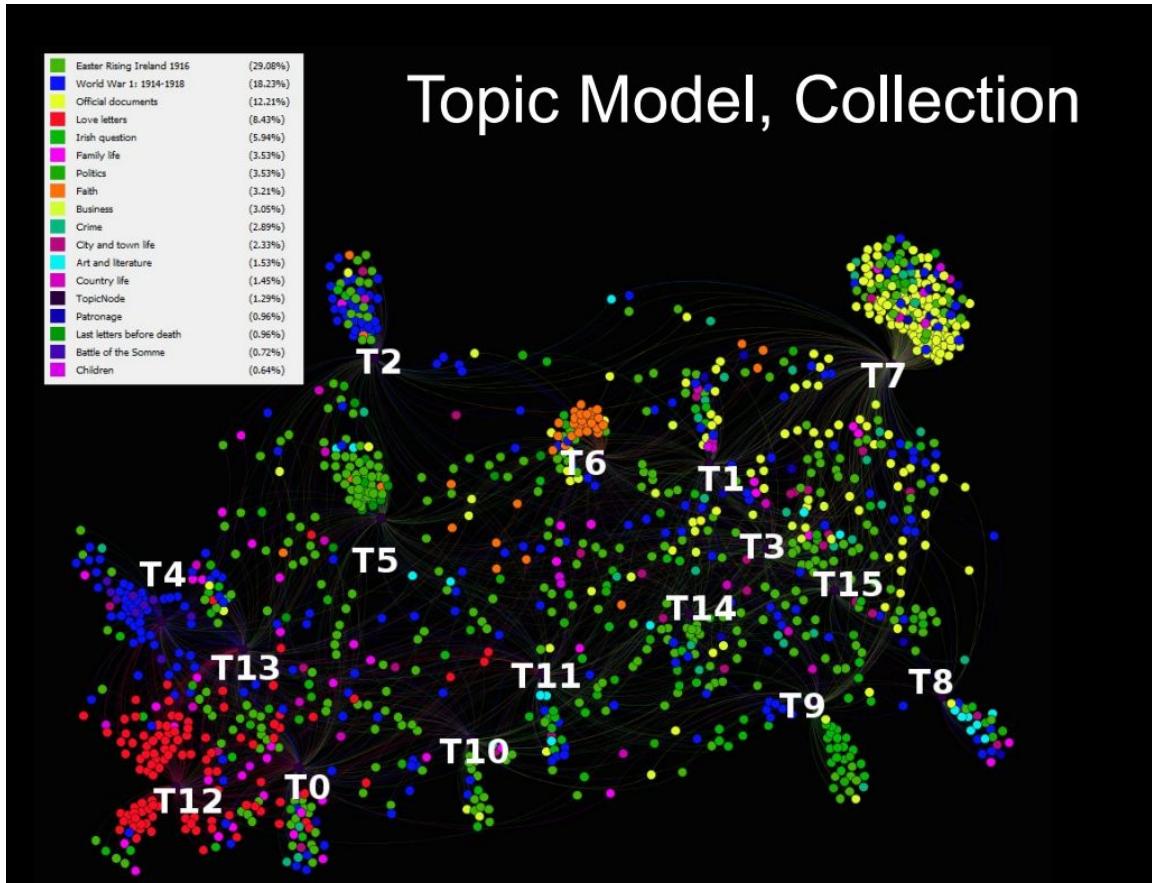
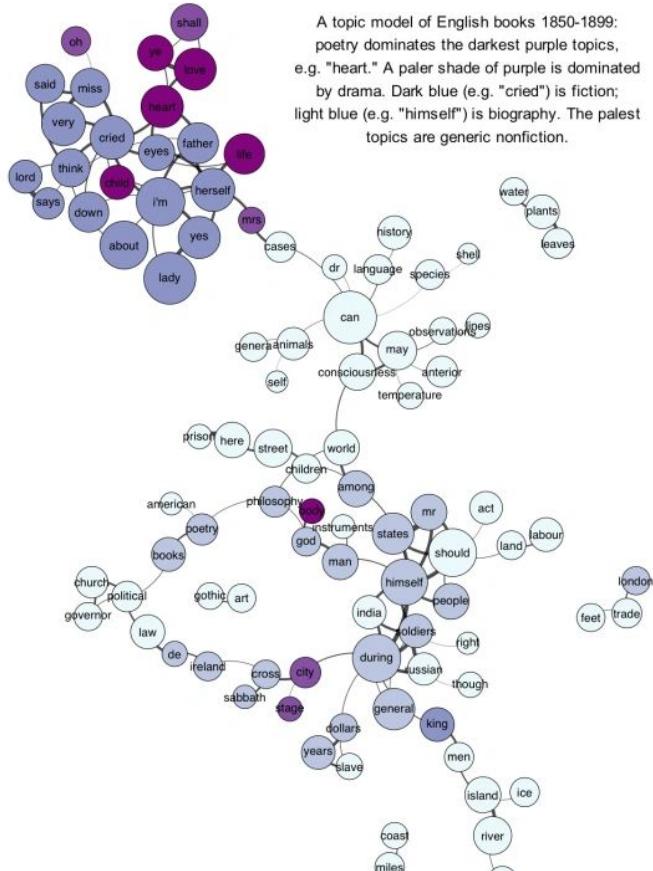
Common ways to represent Topics

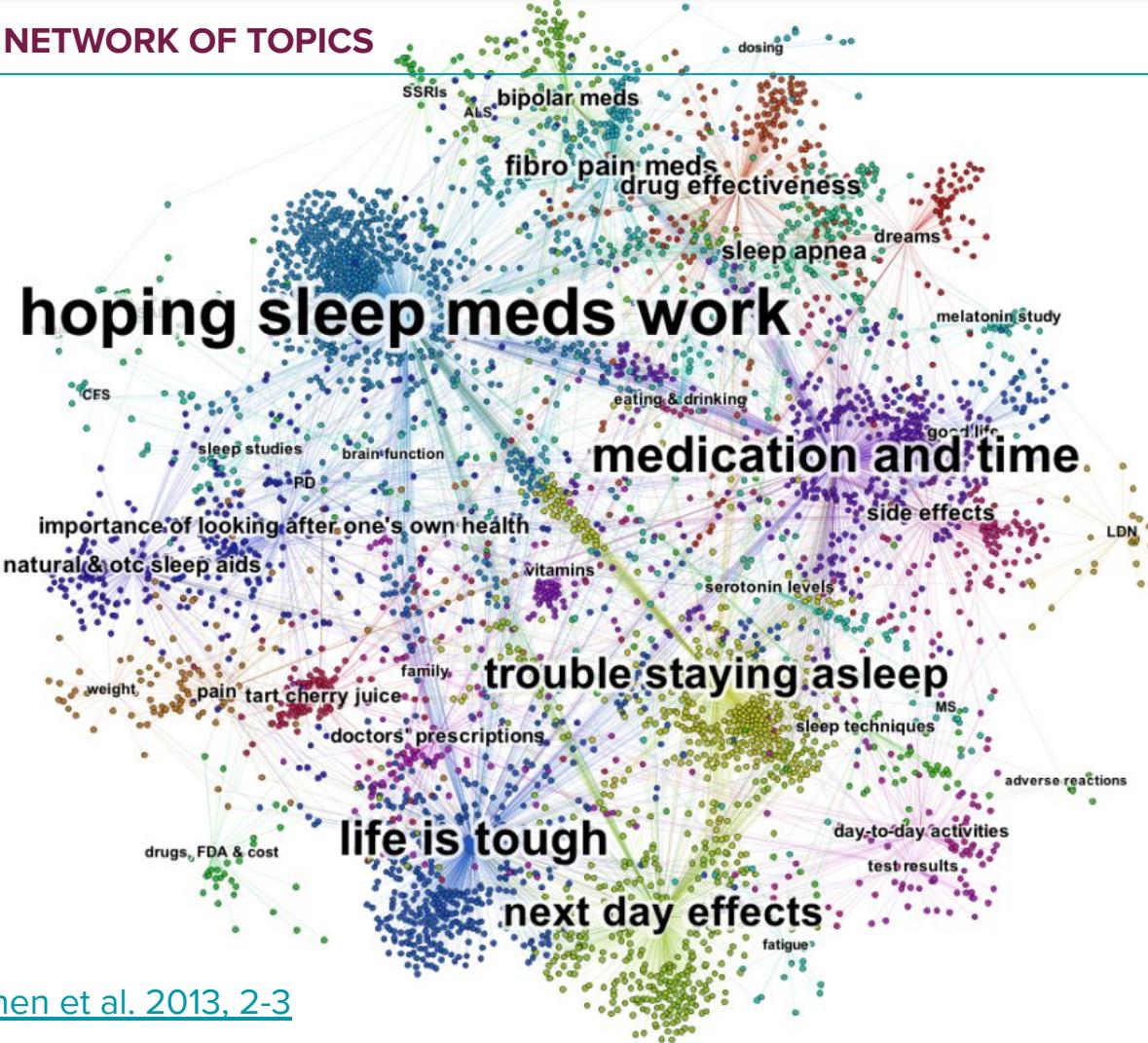
Data table

TopicID	Percentage	Topic (first 20 terms in rank order)	Topic Label
30	10.952%	天大皇四帝海萬太功夷方王三命元聖乾業宗武	empire and sovereign power
16	5.952%	不如食生足骨口死飢肉眼齒腹為土耳力飲可破	mortality, eating and drinking
139	5.238%	塵人春新身親勤鄰殷頻津為巾輪濱四辰辛貧巡	tsyen 真 rhyme category
7	3.81%	我不為此言爾者何有亦如生人與得苦無所願令	pronouns and particles
135	3.095%	將軍兵戰旗馬旌功劍弓戎營騎鼓戈羽角虜射箭	military, armies
110	3.095%	未自慚知豈雖薄終非已顧愧負辭寧猶難甘心敢	shame, regret
32	3.095%	十三年二五四六七一八百千九月今歲前載老第	numbers, calendrical time
15	3.095%	王漢秦國宮吳陵帝楚武臺陽長子安亡苑作蘇梁	Han and pre-Han kingdoms
49	2.381%	南北東西山風望國城日向斗河闕起海長直復從	territorial space, cardinal directions
42	2.381%	庭中滿洞月山風樹裏長高入空雲水煙起日上陽	courtyard / landscape scene?
125	1.667%	門深開日闊高戶滿入客出堂朱外巷院掩牆庭靜	gates, halls, and built spaces

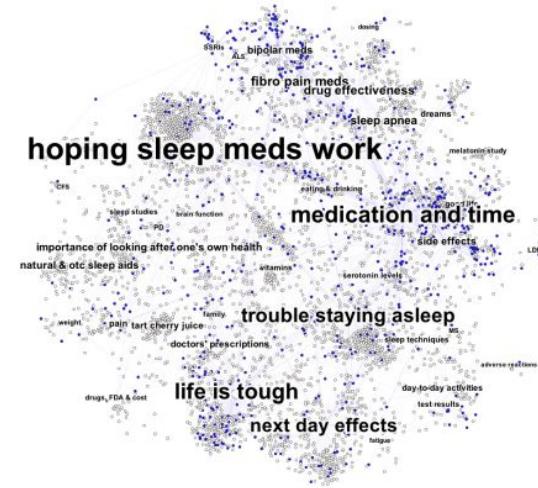
visualizing topics using networks

03 NETWORK OF TOPICS





Ambien (top); trazodone (bottom)



Aim of the network visualization in DiSpecs

- project with an **explorative** and **experimental** approach
- representing topics and periodicals
- ... to answer questions such as:
 - Which periodicals share similar topics?
 - What can a network reveal about individual topics?
 - Can we compare topics with manually annotated keywords using a network?

Input data for *Gephi* (V. 0.9.2)

Nodes

Label	category	size	image
Topic 0	topic	23.768999	NaN
L-Indigent-philosophe-ou-l-homme-sans-souci	periodical	7.000000	L'Indigent philosophe ou l'homme sans souci.png
La-Bagatelle	periodical	100.000000	La Bagatelle.png
La-Bigarure	periodical	197.000000	La Bigarure.png
La-Spectatrice	periodical	15.000000	La Spectatrice.png
La-Spectatrice-Ouvrage-traduit-de-l-anglois	periodical	25.000000	La Spectatrice. Ouvrage traduit de l'anglois.png
La-Spectatrice-danoise-ou-l-Aspasie-moderne	periodical	83.000000	La Spectatrice danoise, ou l'Aspasie moderne.png
...

Edges

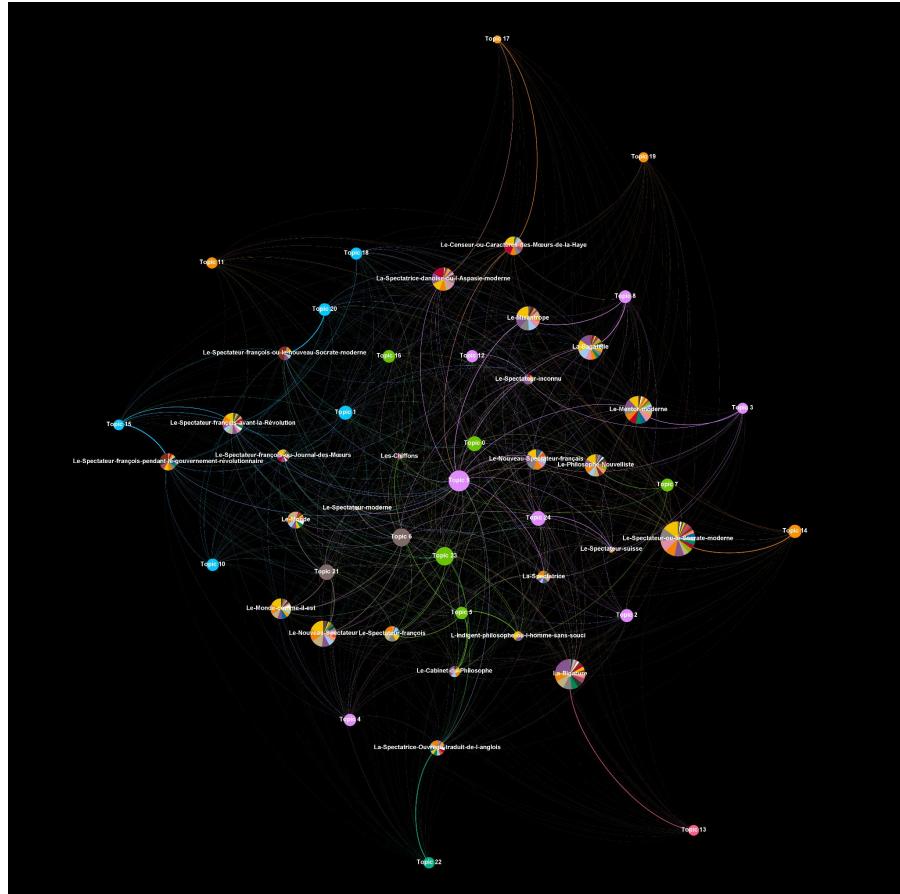
Source	Target	Type	Weight
Topic 0	L-Indigent-philosophe-ou-l-homme-sans-souci	Directed	0.032563
Topic 0	La-Bagatelle	Directed	0.064769
Topic 0	La-Bigarure	Directed	0.068592
Topic 0	La-Spectatrice	Directed	0.048331
Topic 0	La-Spectatrice-Ouvrage-traduit-de-l-anglois	Directed	0.036678
...
Topic 24	Le-Spectateur-inconnu	Directed	0.041099
Topic 24	Le-Spectateur-moderne	Directed	0.020350
Topic 24	Le-Spectateur-ou-le-Socrate-moderne	Directed	0.053887
Topic 24	Le-Spectateur-suisse	Directed	0.196733
Topic 24	Les-Chiffons	Directed	0.058543

Network of topics and periodicals in DiSpecs

View all topic modeling results on the DiSpecs website:

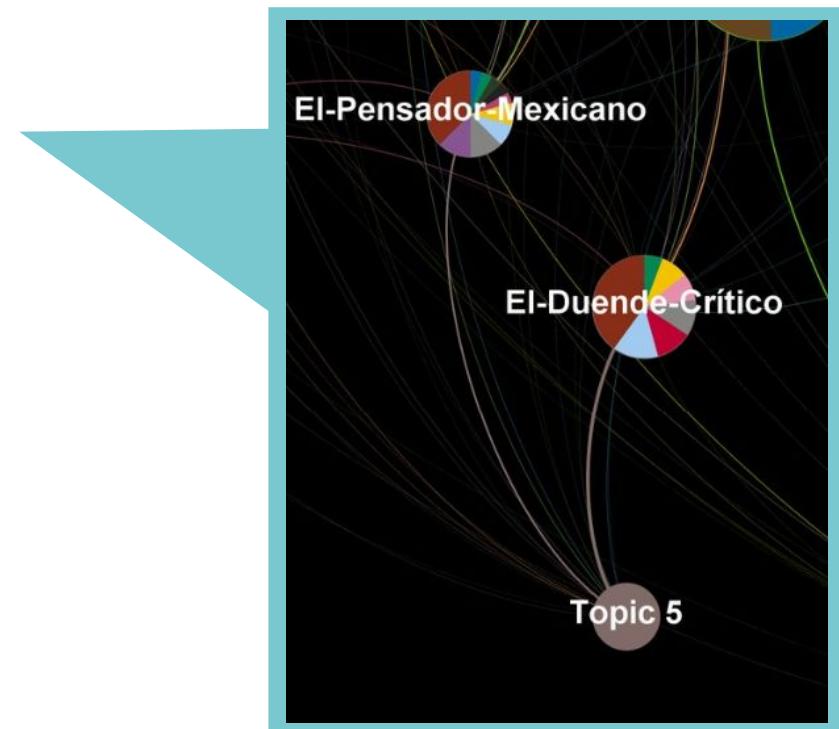
<https://qams.uni-graz.at/archive/objects/context:dispecs/methods/sdef:Context/get?mode=tm-results>

Topic network in French periodicals
[Scholger et al. \(2019-2021\)](#)



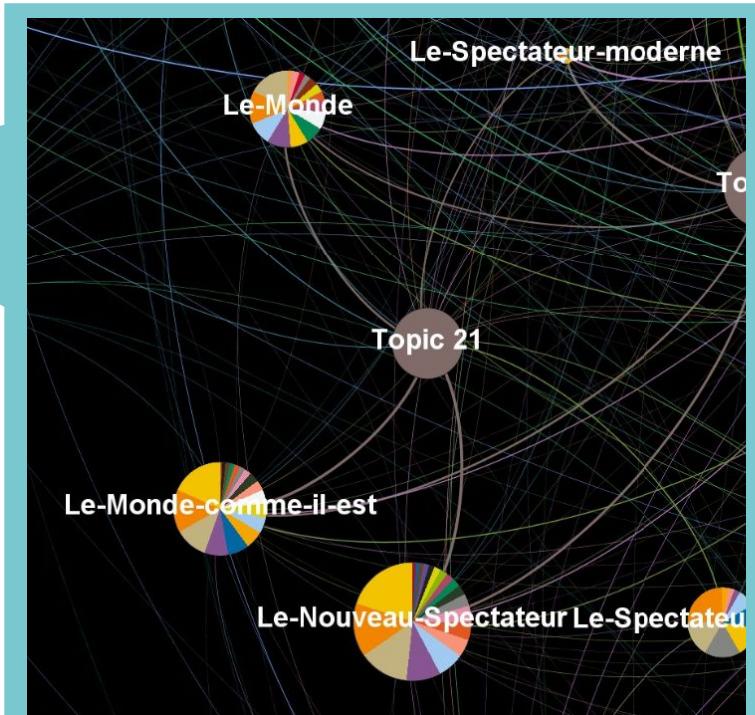
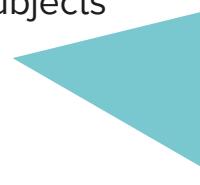
What can we read from the network visualization?

- Which periodicals share similar topics?
 - periodicals who write about similar subjects



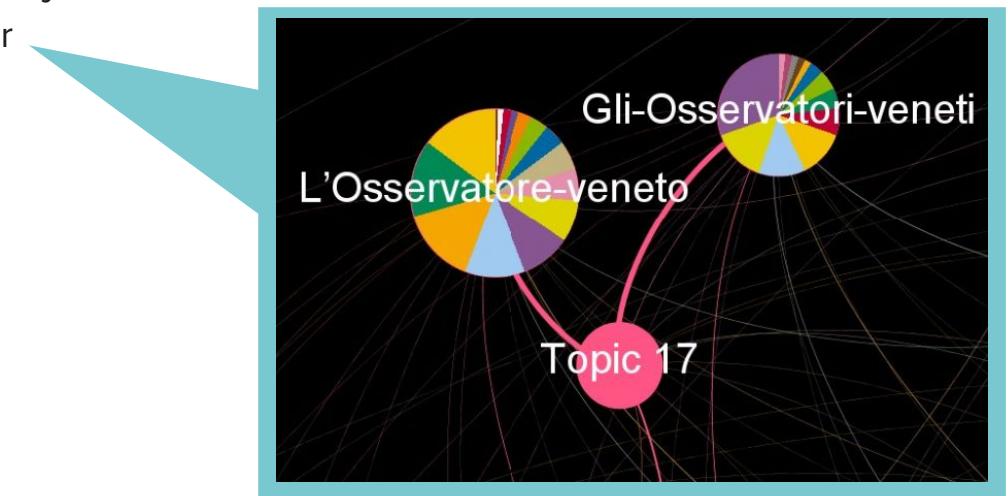
What can we read from the network visualization?

- Which periodicals share similar topics?
 - periodicals who write about similar subjects
 - periodicals of the same author/editor



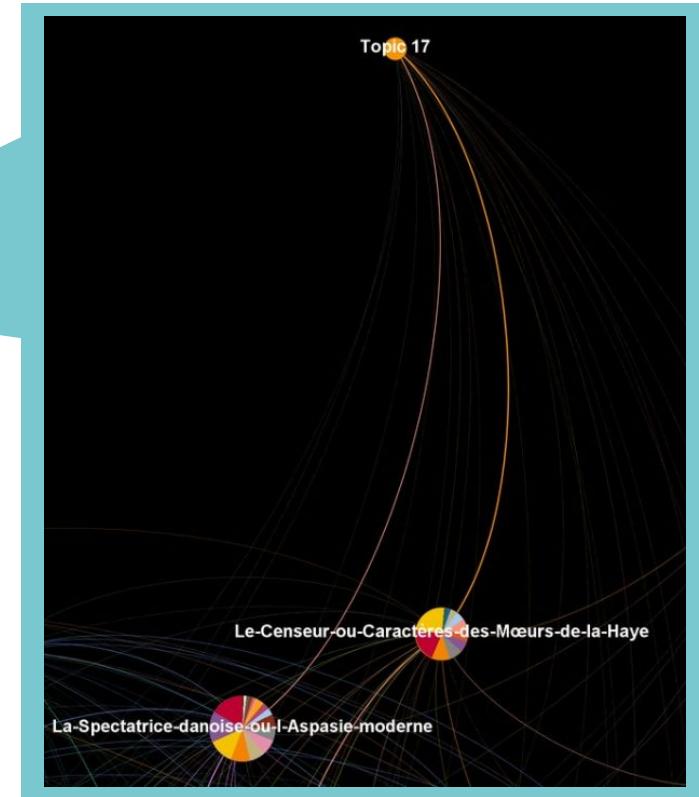
What can we read from the network visualization?

- Which periodicals share similar topics?
 - periodicals who write about similar subjects
 - periodicals of the same author/editor



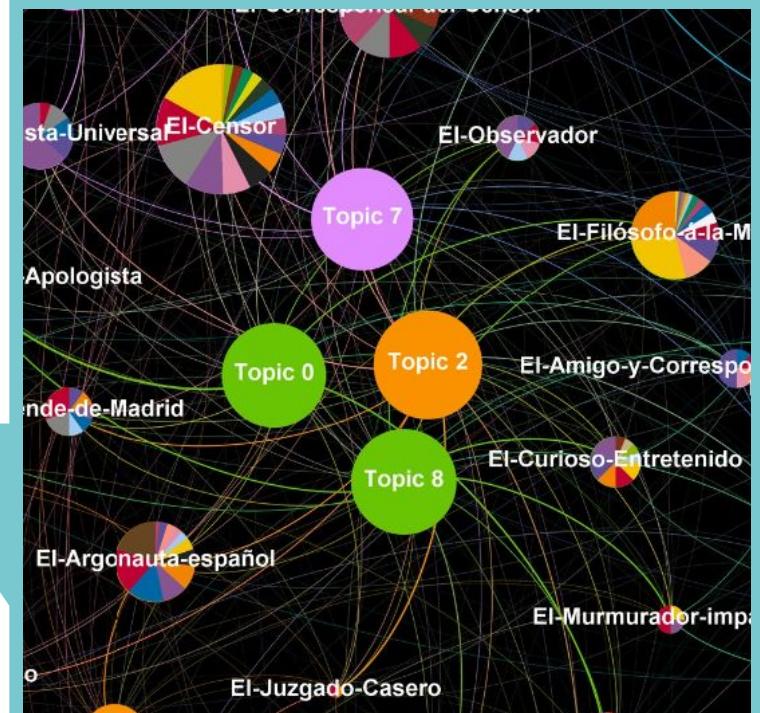
What can we read from the network visualization?

- Which periodicals share similar topics?
 - periodicals who write about similar subjects
 - periodicals of the same author/editor
 - periodicals with similar writing style



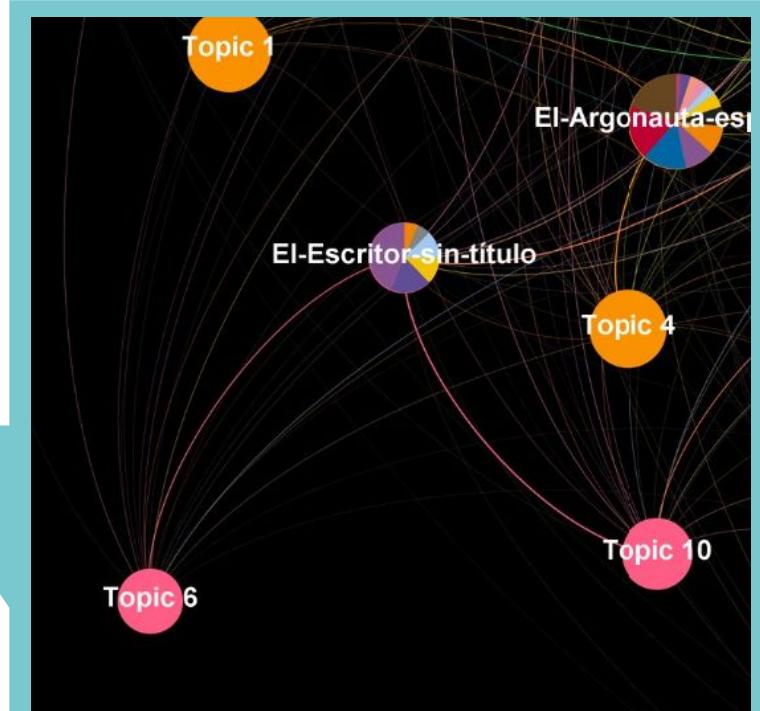
What can we read from the network visualization?

- Which periodicals share similar topics?
 - periodicals who write about similar subjects
 - periodicals of the same author/editor
 - periodicals with similar writing style
- What can a network reveal about individual topics?
 - central (genre typical) vs. outlier (periodical specific) topics



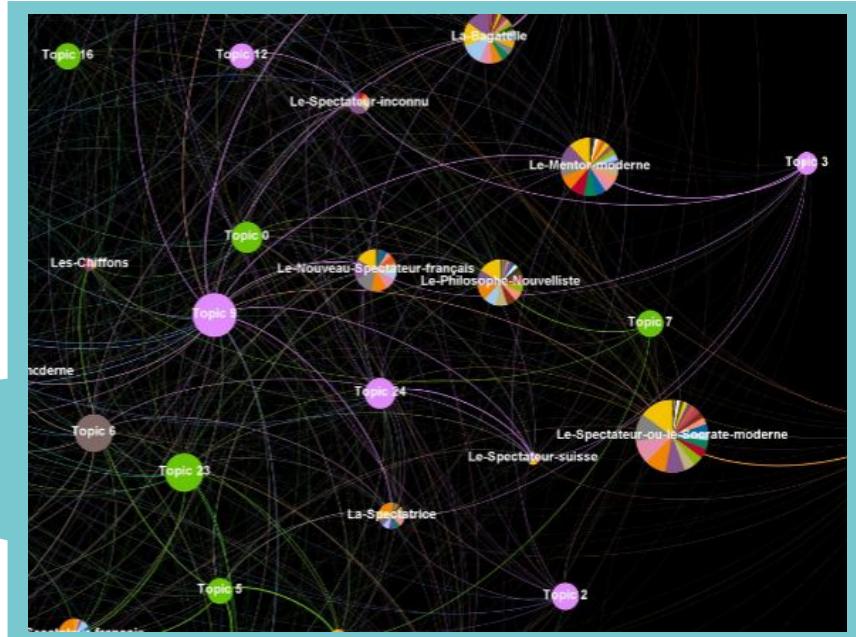
What can we read from the network visualization?

- Which periodicals share similar topics?
 - periodicals who write about similar subjects
 - periodicals of the same author/editor
 - periodicals with similar writing style
- What can a network reveal about individual topics?
 - central (genre typical) vs. outlier (periodical specific) topics



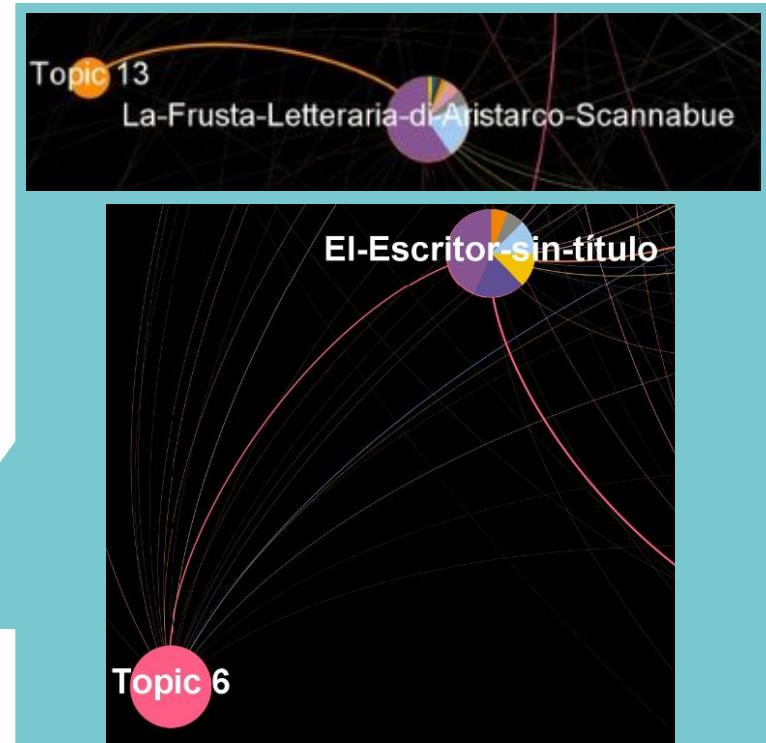
What can we read from the network visualization?

- Which periodicals share similar topics?
 - periodicals who write about similar subjects
 - periodicals of the same author/editor
 - periodicals with similar writing style
- What can a network reveal about individual topics?
 - central (genre typical) vs. outlier (periodical specific) topics
 - modularity group?



What can we read from the network visualization?

- Which periodicals share similar topics?
 - periodicals who write about similar subjects
 - periodicals of the same author/editor
 - periodicals with similar writing style
- What can a network reveal about individual topics?
 - central (genre typical) vs. outlier (periodical specific) topics
 - modularity group?
- Can we compare topics with manually annotated keywords using a network?



conclusion and discussion

Some lessons learned from theory and practice

- avoid information reduction
- the danger of over-interpretation
- a topic model is not really a network
- for categorical data, choose colors of maximum contrast
- descriptive labeling might be better for usability
- document the chosen parameters and steps
- networks have the potential to visualize various aspects of topics and documents

“ ”

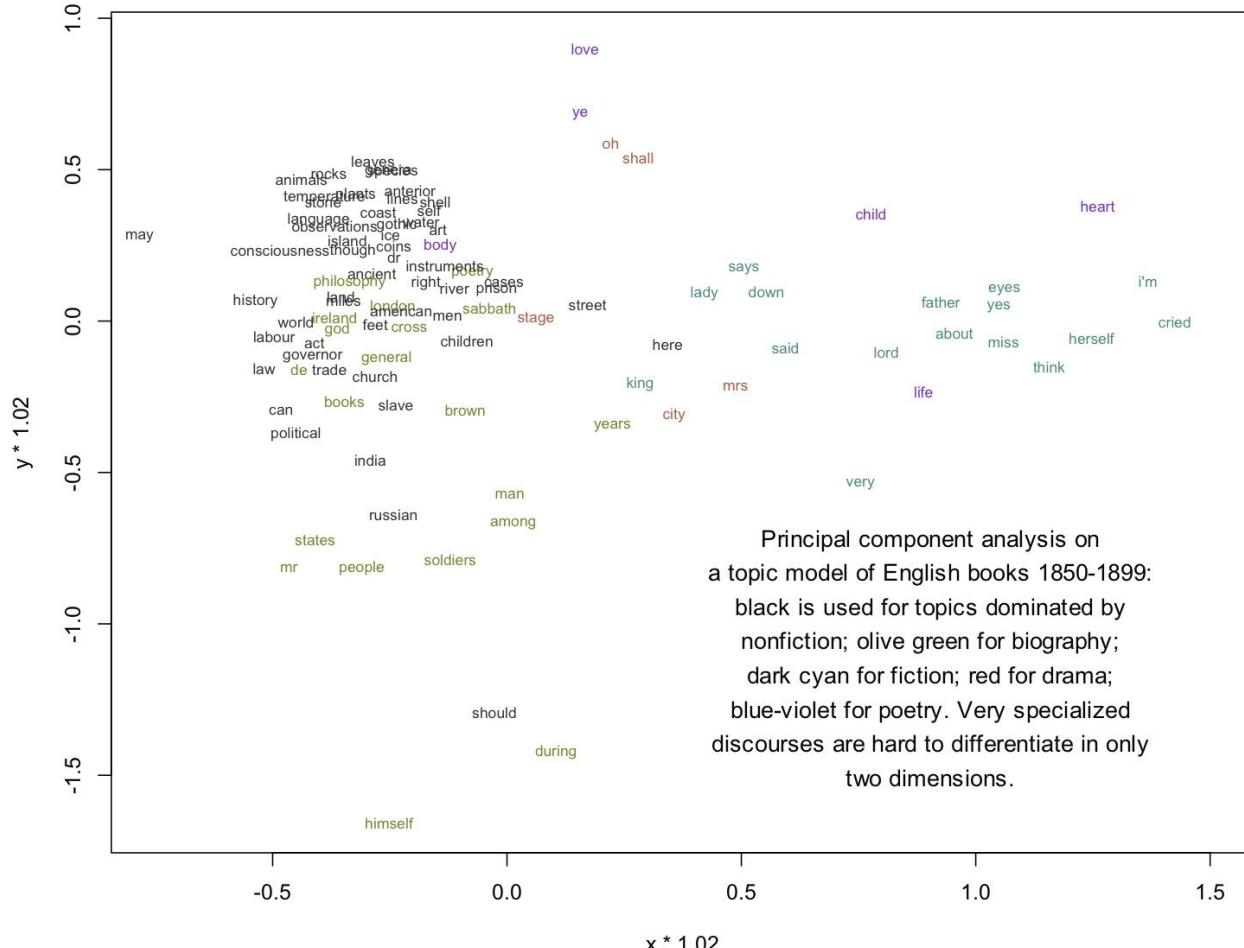
the network scientist / statistician in me twitches uncontrollably whenever he sees someone creating a network out of a topic model by picking the top- n topics associated with each document and using those as edges in a topic-document network

Scott Weingart

Alternatives

Principal component analysis scatter plot

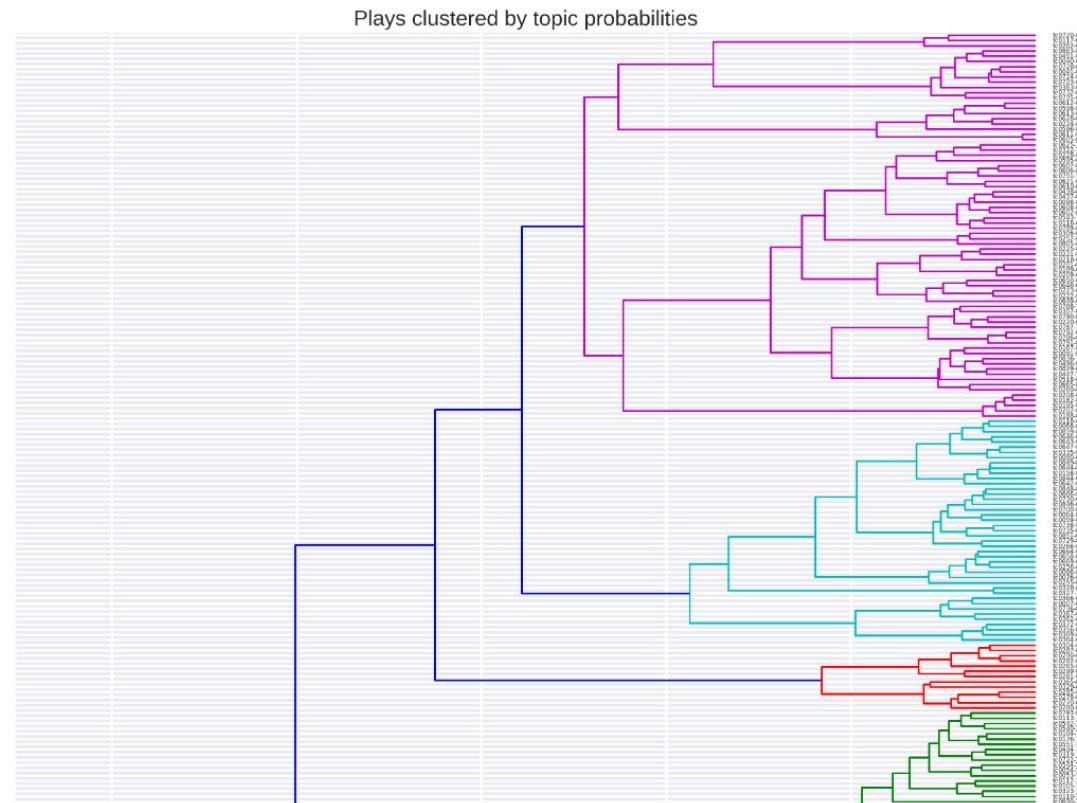
Underwood 2012



Alternatives

Cluster dendrogram

Schöch 2017, Fig. 11



Bibliography

- Blei, David M. (2012): "Probabilistic Topic Models". URL: <https://www.cs.columbia.edu/~blei/papers/Blei2012.pdf>, DOI: 10.1145/2133806.2133826.
- Bleier, Roman (2015): "Topic Modelling des Letters of 1916 Briefkorpus". URL: <http://gams.uni-graz.at/o:dhd2015.v.005>.
- Blondel, Vincent D; Guillaume, Jean-Loup; Lambiotte, Renaud; Lefebvre, Etienne (2008): "Fast unfolding of communities in large networks". In: Journal of Statistical Mechanics: Theory and Experiment, 2008 (10). DOI: 10.1088/1742-5468/2008/10/P10008.
- Chen, Annie; Sheble, Laura; Eichler, Gabriel (2013): "Topic Modeling and Network Visualization to Explore Patient Experiences". DOI: 10.13140/2.1.4268.4481.
- DARIAH-DE (2019): "DARIAH Topics. Easy Topic Modeling in Python". V. 2.0.1. URL: <https://github.com/DARIAH-DE/Topics>.
- Ertler, Klaus-Dieter; Fuchs, Alexandra; Fischer-Pernkopf, Michaela; Hobisch, Elisabeth; Scholger, Martina; Völkl, Yvonne (2011-2018). "The Spectators in the International Context". URL: <https://gams.uni-graz.at/spectators>.
- McCallum, Andrew Kachites (2002-2020): "MALLET. A Machine Learning for Language Toolkit". V. 2.0.8. URL: <http://mallet.cs.umass.edu>.
- Underwood, Ted (2011): "Visualizing topic models". URL: <https://tedunderwood.com/2012/11/11/visualizing-topic-models/>.
- Scholger, Martina; Geiger, Bernhard; Hobisch, Elisabeth; Koncar, Philipp; Sarić, Sanja; Völkl, Yvonne; Glatz, Christina (2019-2021): "Distant Spectators (DiSpecs). Distant Reading for Periodicals of the Enlightenment". URL: <https://gams.uni-graz.at/dispecs>.
- Schöch, Christof (2017): "Topic Modeling Genre: An Exploration of French Classical and Enlightenment Drama". In: Digital Humanities Quarterly 011, H. 2. URL: <http://digitalhumanities.org:8081/dhq/vol/11/2/000291/000291.html>.

THANKS

Does anyone have any thoughts on this?

<https://gams.uni-graz.at/dispecs>

Twitter #DiSpecs
@snjsrc

dispecs@uni-graz.at
sanja.saric@uni-graz.at

