

Le roman policier: approche du genre à travers ses „topics“

Christof Schöch

Université de Würzburg, Allemagne

Departement de philologie numérique

Groupe “Stylistique computationnelle des genres littéraires”

<http://www.christof-schoech.de/fr>

Atelier „Les interfaces numériques dans la recherche d'aujourd'hui“

Frankoromanistentag, Münster, 24.-27. septembre 2014

Plan

- 1. Contexte: le roman policier**
- 2. Questions et hypothèses**
- 3. Données: 200 romans**
- 4. Méthode: le topic modeling**
 - Le topic modeling
 - La chaîne de traitement
- 5. Résultats & interprétation**
 - Les topics à travers leurs mots
 - Les topics par décennie, auteurs, sous-genres
 - Groupements sur base de topics
- 6. Bilan**

1.
contexte :
le roman policier

Le roman policier: définition

Jacques Sadoul

„Le roman policier est le récit rationnel d'une enquête menée sur un problème dont le ressort dramatique principal est un crime.“
(Jacques Sadoul, cité par Lits 1993, 90)

Éléments de définition:

- prose narrative développée (≠ énigme; ≠ calcul)
- enquête rationnelle au centre de l'histoire (à voir)
- l'objet de l'enquête est un crime
- + personnage principal: l'enquêteur (≠ thriller, ≠ suspense)
- + situé dans un milieu urbain (sauf exception)

Le grandes tendances du roman policier

roman judiciaire

- roman policier en germe
- Boisgobey, Ponson du Terrail

roman policier archaïque

- roman d'aventure sur fond de crime et d'enquête
- Leblanc, Leroux, Gaboriau

roman de détection

- l'enquête rationnelle; détective; structure narrative double
- Gaboriau, Simenon

roman noir

- structure narrative simple, un monde en désordre, l'histoire
- Malet, Vargas, Daeninckx

néo-polar

- le crime et la violence;
- Manchette, Pouy, Japrisot

parodique / expérimental

- roman policier ironique, auto-réflexif, etc.
- Frédéric Dard

Le roman policier: sous-genres

Tzvetan Todorov

“La mise en évidence des genres à l'intérieur du roman policier promet donc d'être relativement simple.” (1971, 56-57)

Selon le personnage principal

Détective vs. criminel vs. victime

récit de détection vs. thriller vs. roman à suspense

Selon la structure narrative

progressive (simple) vs. régressive (double)

Selon la vision du monde

noir/blanc vs. gris

ordre vs. désordre

2. questions & hypothèses

structure ou thème ?

Sous-genres thématiques ?

“Le roman noir moderne s'est constitué non autour d'un procédé de présentation mais autour d'un milieu représenté, autour de personnages et de moeurs particuliers; autrement dit, sa caractéristique constitutive est dans ses thèmes.”

(Todorov 1971, 60-61)

Hypothèse

- L'arrivée du roman noir apporte un renouvellement thématique qui sera clairement visible dans les topics
- Les étapes intermédiaires seront visible également, mais confondues avec les auteurs.

Auteurs, genres, chronologie ?

Quel est le facteur déterminant?

Le facteur le plus déterminant pour la distribution des topics dans un roman policier donné, est-ce l'auteur qui l'a écrit, le sous-genre auquel il appartient, ou la date de sa publication?

Hypothèse

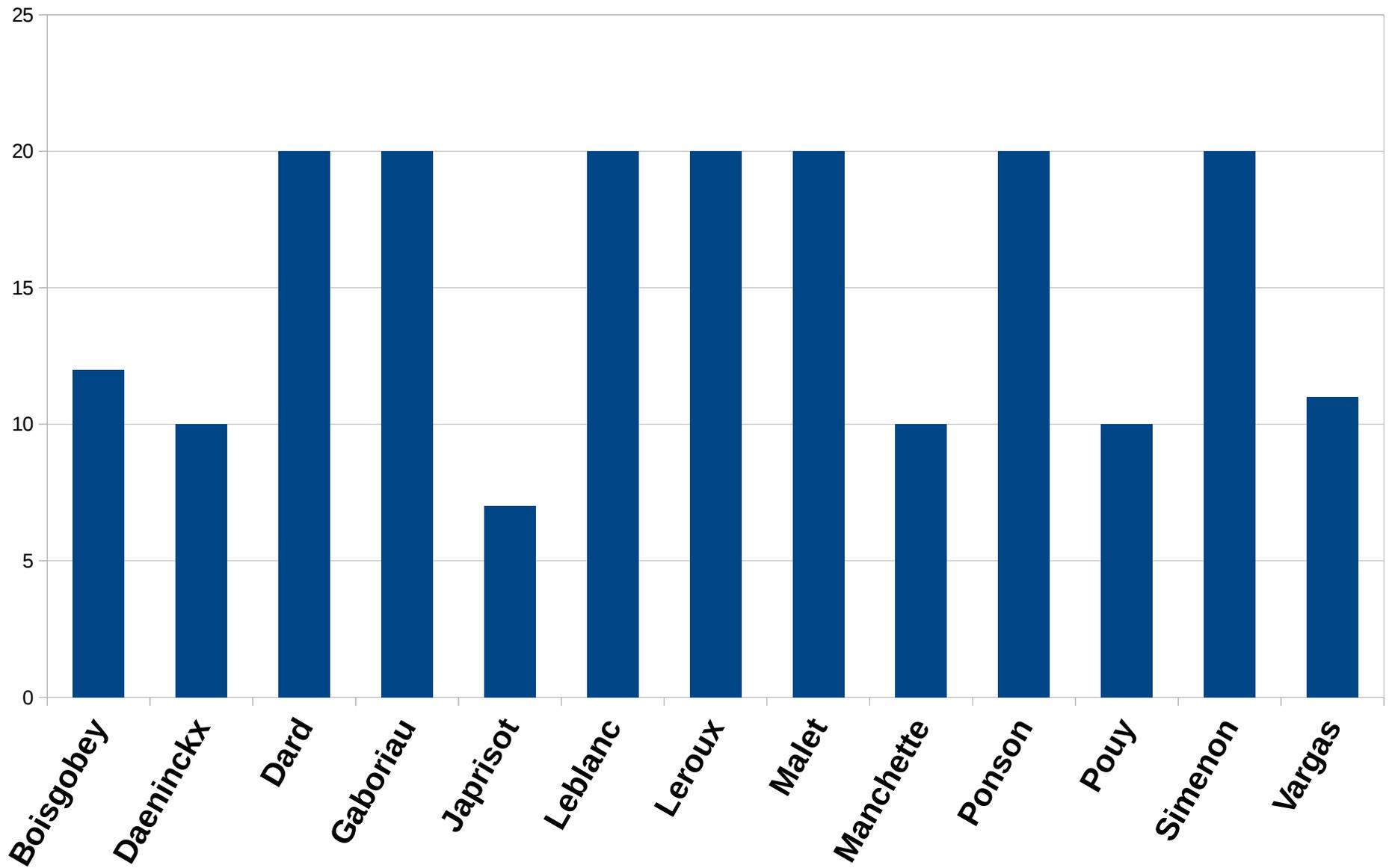
- Malgré des profils individuels par auteurs, plusieurs auteurs pratiquant le même sous-genre seront plus similaires les uns aux autres que des auteurs pratiquant des sous-genres différents.
- (Problème: ces trois facteurs étant fortement corrélées les uns avec les autres, il est difficile d'éprouver cette hypothèse.)

3.

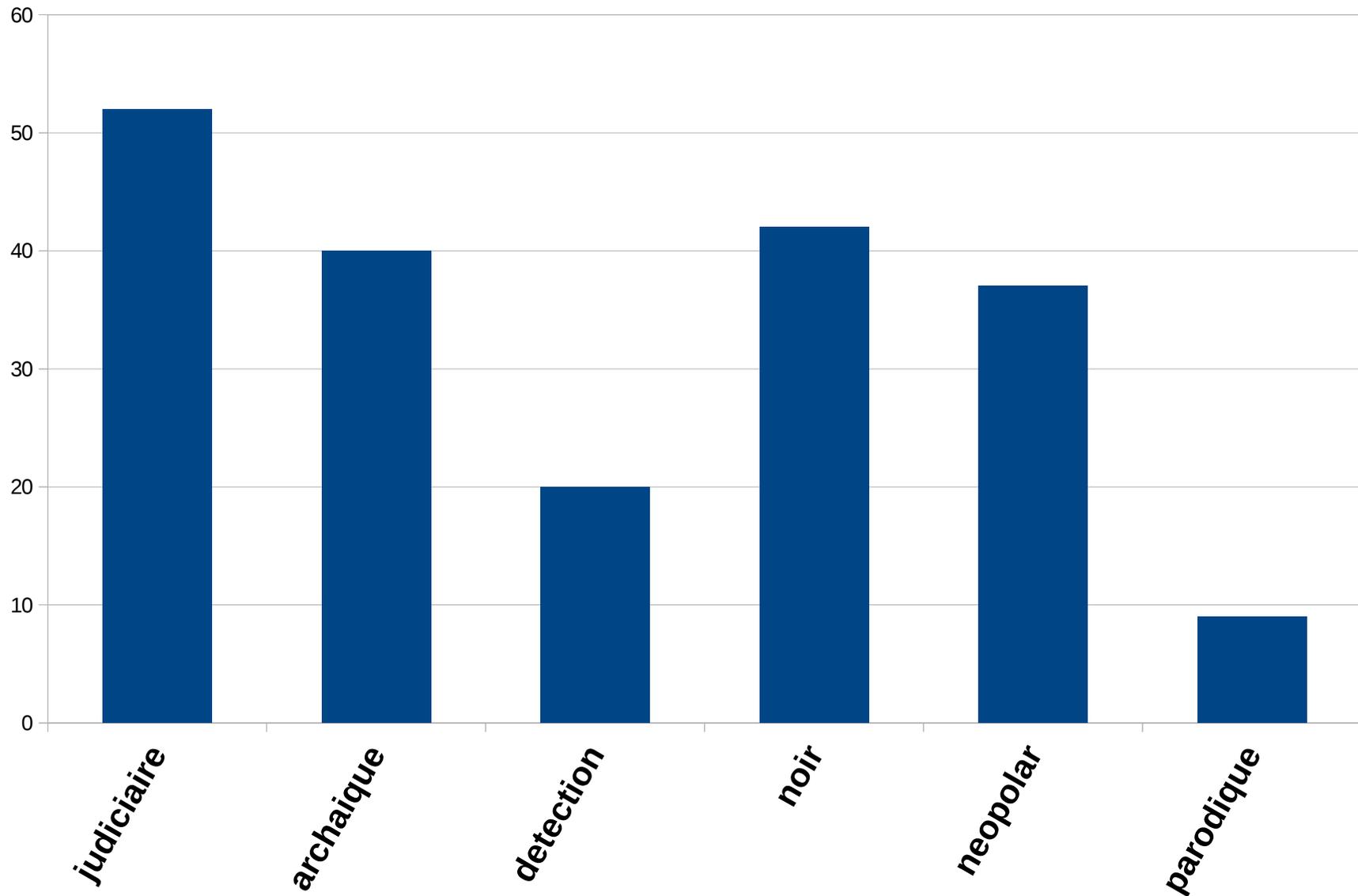
données :

**200 romans, 14 mio de mots,
balisage TEI**

Aperçu du corpus: romans par auteur

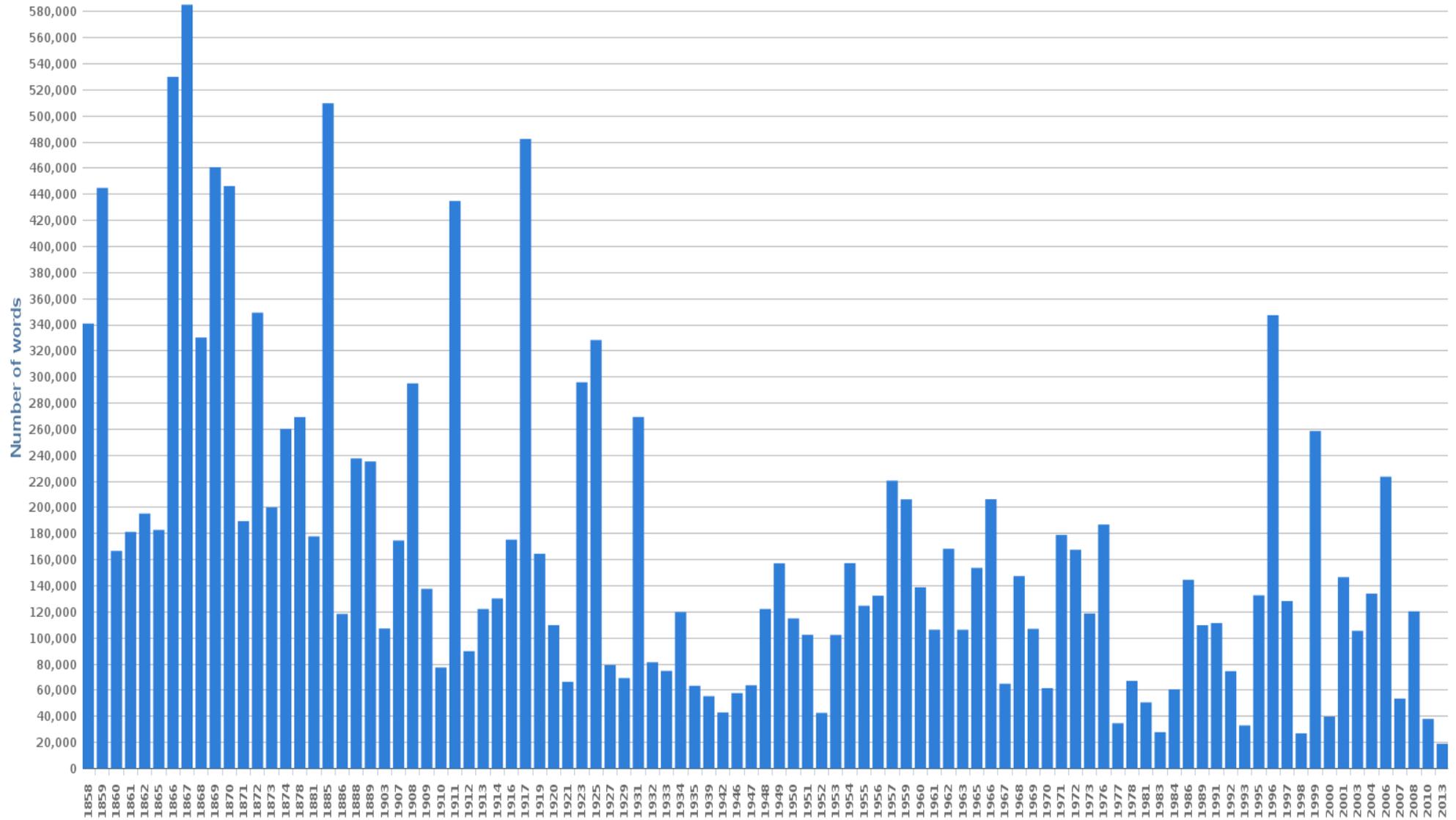


Aperçu du corpus: romans par sous-genres



Aperçu du corpus: mots par année

Dimensions of the annee partition in the RP200 corpus



Limitations du corpus

Lacunes du corpus

Auteurs ayant écrit moins de sept romans

Romans ayant été écrits entre 1890 et 1899

Romans à suspense, romans d'espionnage, r.p. historique

Paul Féval, Allain-Souvestre

Claude Aveline, Pierre Véry, St.-A. Steeman, Boileau-Narcejac

Frédéric Fajardie, Pierre Siniac

Jean-Claude Izzo, Maurice Attia, Yasmina Khadra, Pierre Magnan

Métadonnées problématiques

Propriété: “structure narrative”

Propriété: “personnage central”

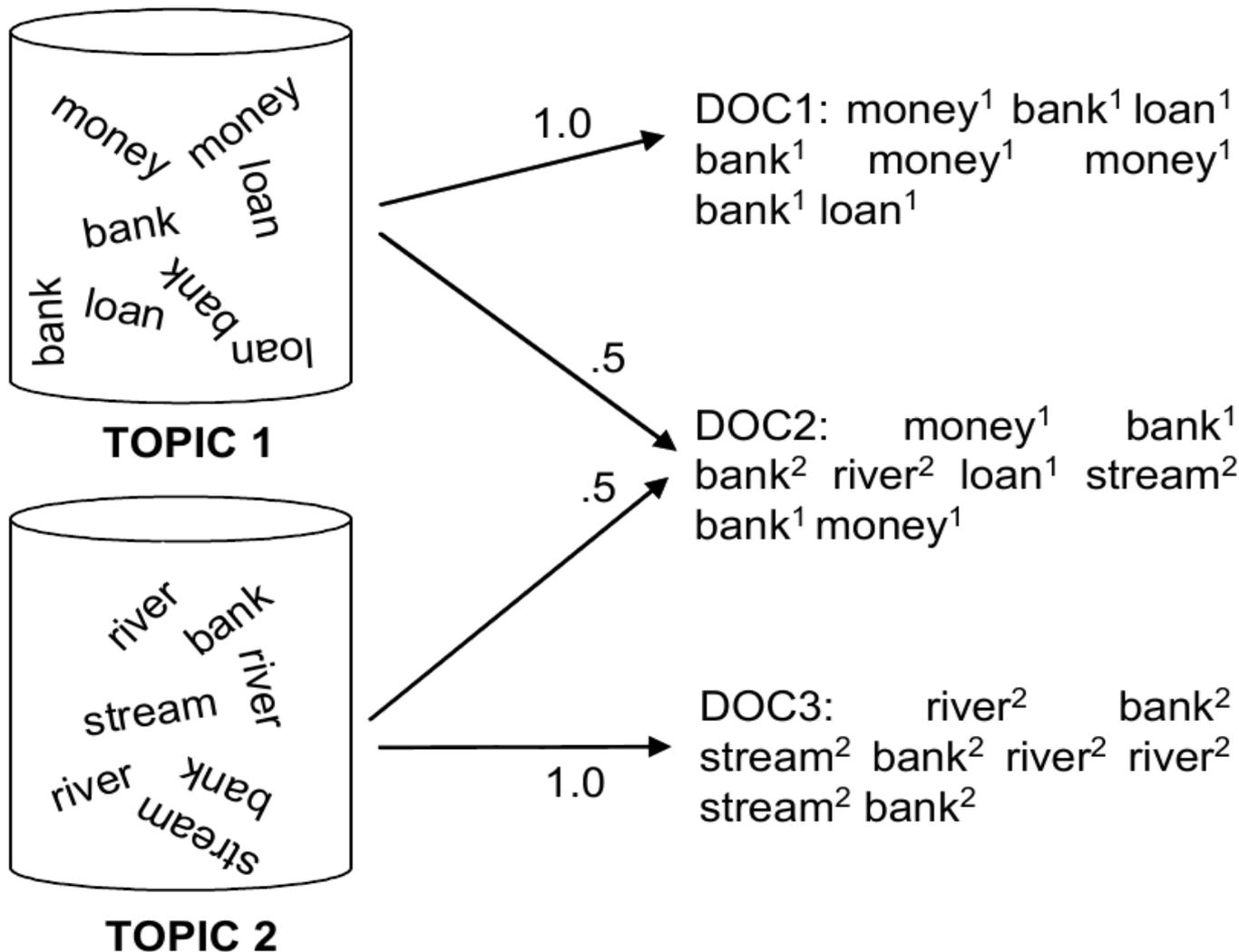
Propriété: “sous-genre” (!)

4.

**La méthode :
le „topic modeling“**

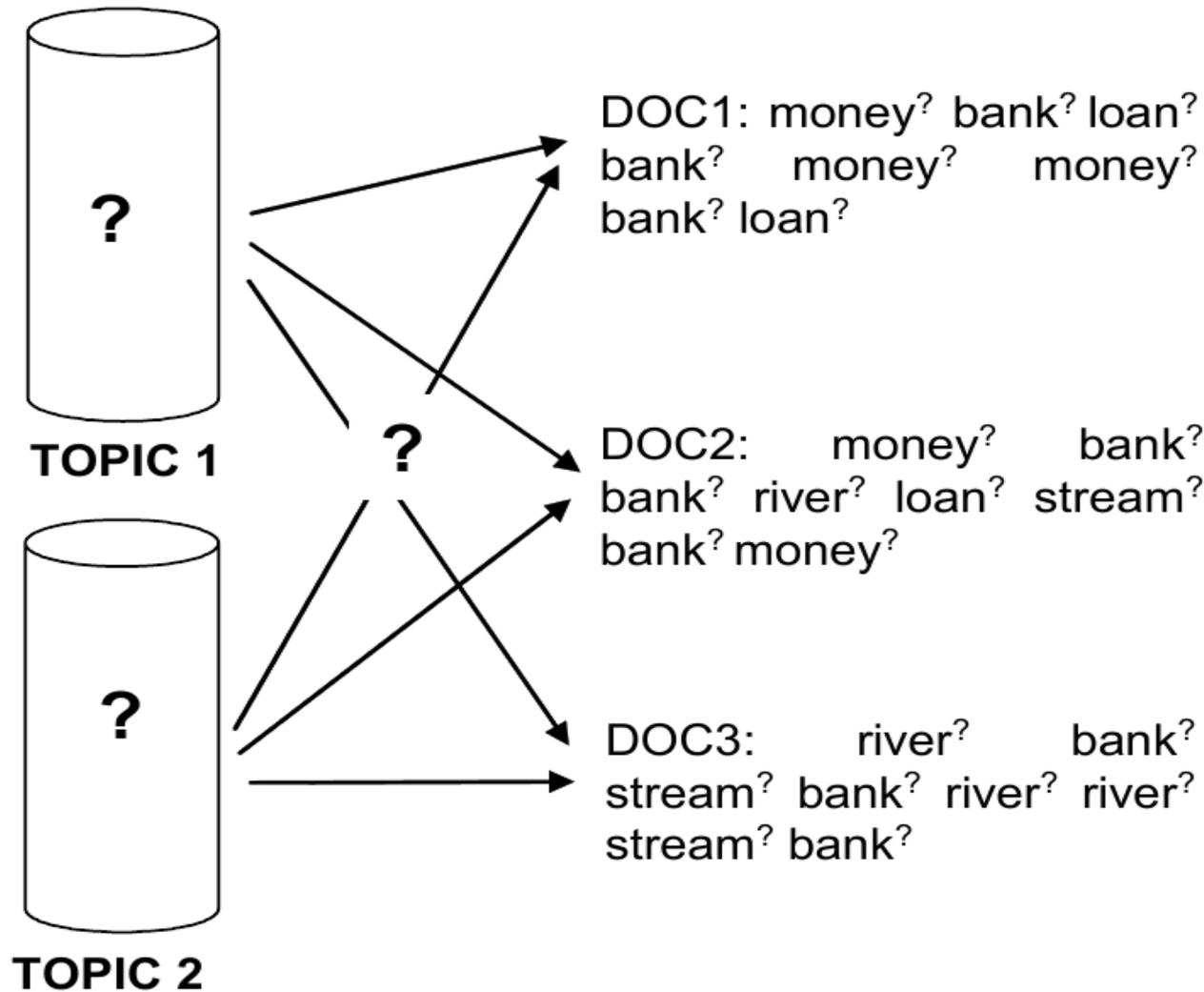
L'idée du topic modeling: word distributions

PROBABILISTIC GENERATIVE PROCESS



L'idée du topic modeling: word distributions

STATISTICAL INFERENCE



La chaîne de traitement

200
fichiers
TEI

extraction (lxml)

200
fichiers
TXT

tokenisation (re)

tokens

TreeTagger

tokens
balises
lemmes

fichiers
avec métad.

métadonnées

lemmes
uniquement

extraction (re)

35.000
fragments

chunking (py)

format
.mallet

import
(mots-vides)

topic
modeling

topics
avec mots

docs
avec topics

agrégation

topics
par auteur

topics
par décennie

topics
par genre

5. résultats & interprétation

topics sans spécificité générique

9 0.094 famille

père mère fille enfant jeune fils vieux jour femme famille nom parent vivre sœur mourir année mois appeler marier maison frère connaître veuf aimer parents parler pauvre mariage voir maman veuve beau madame fortune épouser mari apprendre habiter vie

18 0.079 argent

franc mille payer cent argent monsieur billet affaire somme million jour mois fortune acheter gagner vendre riche cinquante homme perdre banque livre prix cher or vivre poche maison besoin jouer gens posséder fond rente compter jeu valoir foi ruiner

36 0.040 maladie-guerison

docteur médecin malade mourir lit mort blessure chambre blessé hôpital jour infirmier soigner guérir voir poison heure revenir chirurgien demander madame corps maladie arriver malais malaise sang temps chercher soin vie effet accident fou croire pauvre bras transporter tomber

topics spécifiques au genre

0 0.068 enquête2-instruction

monsieur juge instruction homme police répondre justice agent magistrat affaire crime prison président avocat commissaire demander maître père coupable suivre témoin interroger policier arrêter prévenu innocent greffier condamner chef preuve voir cabinet paix brigadier cour reconnaître nom question reprendre

5 0.062 crime1-armeafeu

main tirer coup tête arme œil regarder homme sortir balle bras tomber feu pied poche revolver droit ouvrir essayer suivre visage épaule côté asseoir mur sang air terre lever fusil sol canon porte jeter bout pistolet mètre poser second

30 0.117 crime2-mort

mort croire tuer crime homme assassin voir police affaire penser cadavre connaître mourir retrouver suivre femme découvrir victime personne preuve nuit moment coup raison comprendre complice assassiner doute cacher commettre jour chercher agir trace certain histoire parler meurtre ignorer

35 0.077 enquête5-commissaire

commissaire heure voir femme bureau suivre parler homme rue foi temps demander inspecteur regarder connaître patron matin certain appeler croire attendre police moment maison arriver téléphone question entendre madame rester comprendre essayer penser téléphoner nouveau porte occuper gens place

topics un peu étranges

26 0.031 special1-verybritish

homme gris sir miss lord londres suivre irlandais révérend anglais répondre enfant voir abbé parler mistress regarder gentleman reprendre prêtre major milady jour heure prison attendre poursuivre croire nom angleterre entrer main maison demander sourire gouverneur français monsieur fils

11 0.155 heures-du-jour

heure jour soir matin attendre nuit partir arriver chambre revenir voir rentrer sortir quitter lendemain hôtel demander temps demain nouveau arrivé arrivée moment tard rester minute journée foi suivre femme maître dîner répondre coucher penser maison rue entrer conduire

32 0.024 enquête4-argot

voir type suivre penser monde regarder genre temps loin sentir foutre côté parler truc tête vrai mec partir coup jour foi tenter con groupe merde air comprendre moment sortir aimer rester blanc revenir vite longtemps train doute vie beau

quelques jolis topics

33 0.052 automobilité

voiture route train auto portier portière arrêter gare descendre rue droit volant loin
mètre chauffeur suivre côté arrière attendre voir arriver rouler traverser chemin
place sortir monter quai heure camion trottoir ouvrir rejoindre conduire direction
wagon temps marche véhicule

15 0.060 espaces1-souterrain

mur voir eau pied trou corde cave descendre souterrain sol tour terre fer bout main
échelle lanterne servir côté endroit mètre puits fond chemin planche bas corps
apercevoir tomber coup temps foi glisser creuser sorte marche homme lumière
escalier

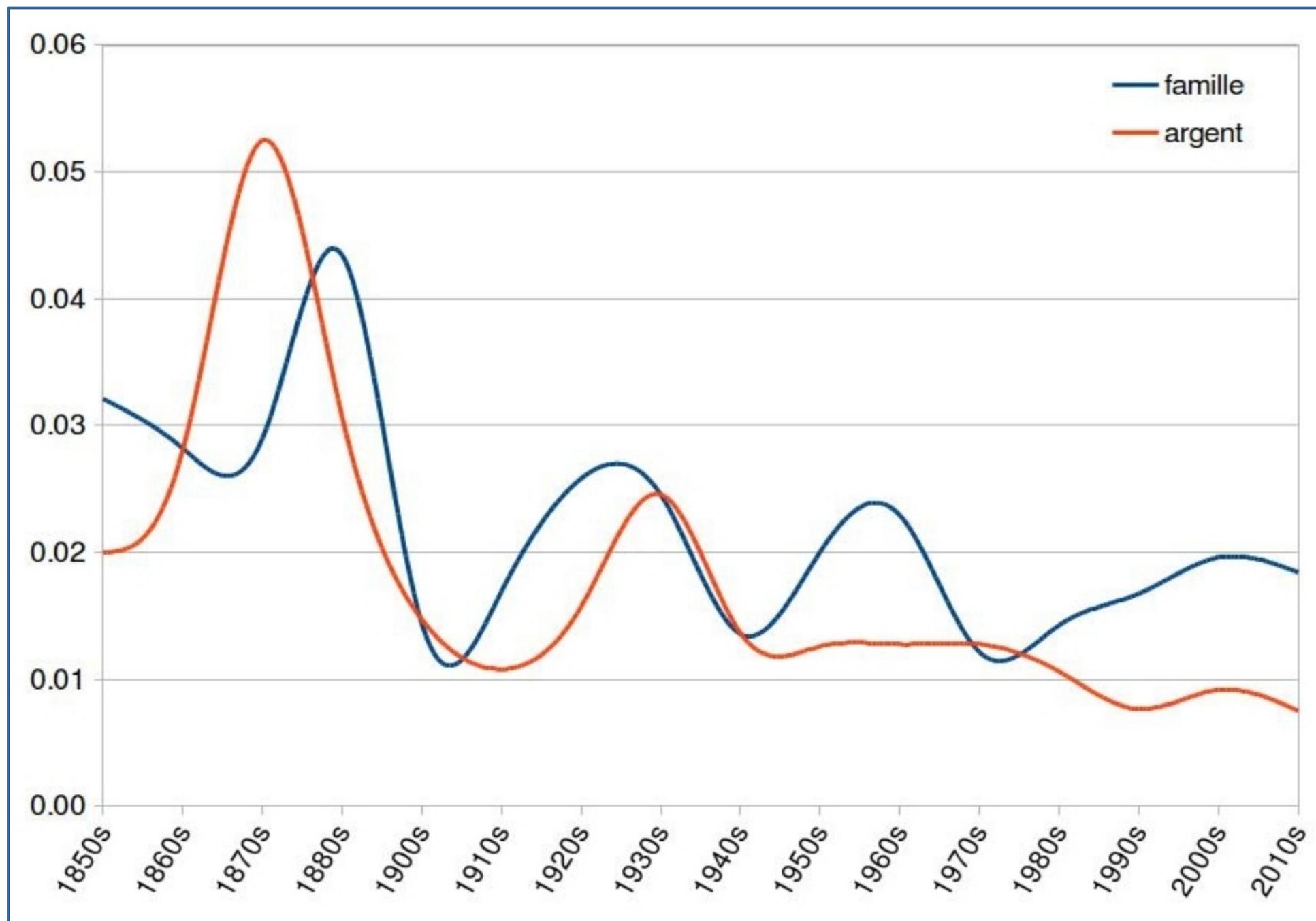
21 0.070 lecture-écriture

lettre écrire papier lire poche billet remettre ouvrir enveloppe main nom écriture
feuille bureau porter livre page signer adresse journal table tirer ligne contenir
tendre portefeuille demander monsieur jeter carte recevoir chercher apporter voir
 tiroir signature sortir paquet document

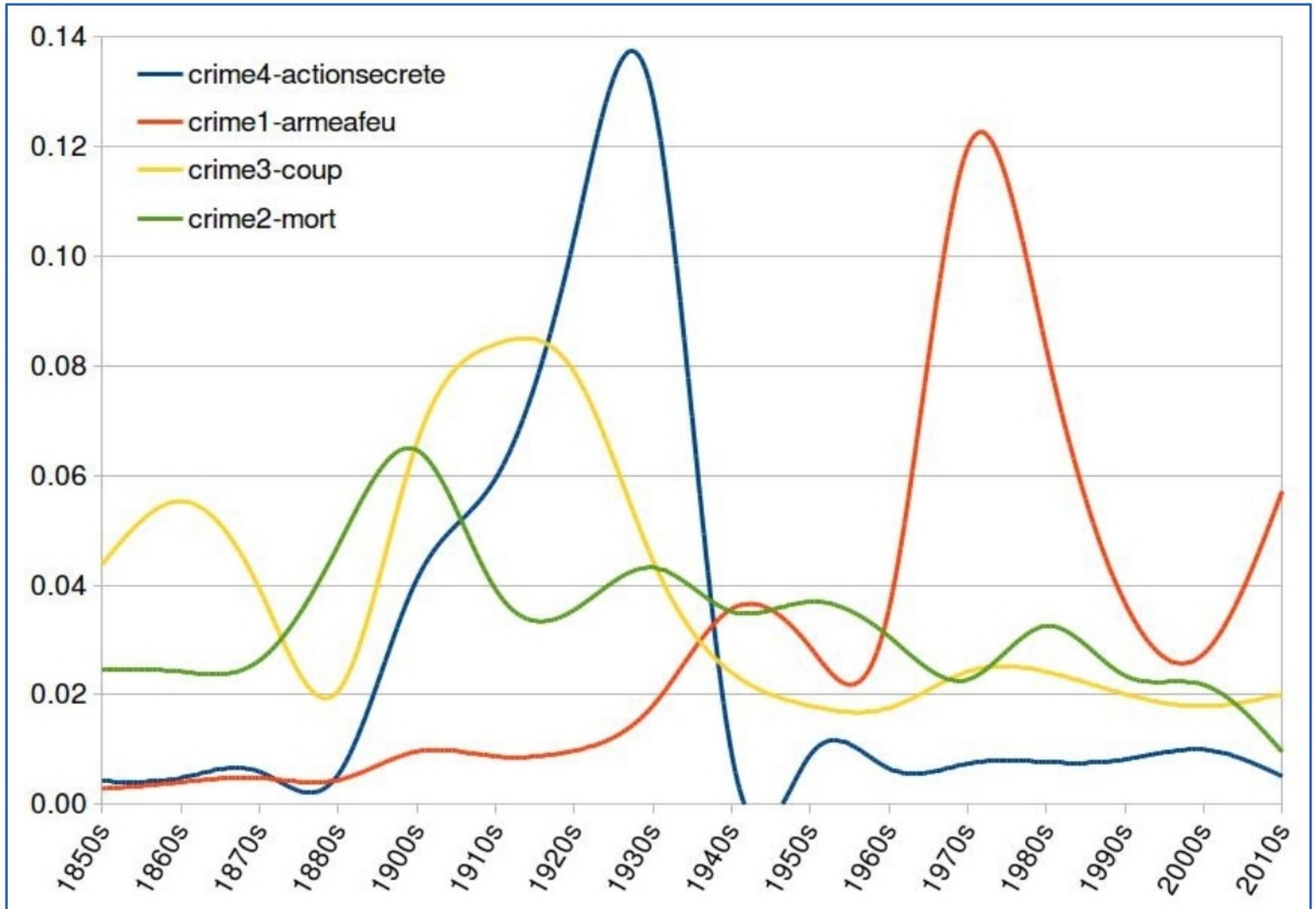
Les topics par auteurs: "signature topics"

topic	score	label	Boisgo	Daenin	Dard	Gabori	Japrisot	Leblanc	Leroux	Malet	Manch	Ponson	Pouy	Simen	Vargas
0	0.068	enquête2-instruction	0.03	0.01	0.01	0.05	0.00	0.03	0.02	0.01	0.01	0.01	0.01	0.02	0.01
1	0.068	crime4-actionsecrete	0.01	0.01	0.01	0.01	0.00	0.19	0.01	0.01	0.01	0.00	0.01	0.01	0.01
2	0.081	nocturne	0.01	0.02	0.02	0.01	0.02	0.02	0.03	0.03	0.02	0.03	0.02	0.02	0.02
3	0.119	existence-mondaine	0.02	0.04	0.03	0.03	0.01	0.04	0.06	0.03	0.04	0.01	0.04	0.03	0.04
4	0.132	société4-monsieur	0.03	0.01	0.01	0.27	0.01	0.02	0.02	0.01	0.01	0.02	0.01	0.01	0.01
5	0.062	crime1-armeafeu	0.01	0.05	0.03	0.01	0.07	0.01	0.01	0.03	0.24	0.00	0.04	0.02	0.03
6	0.040	spécial3-verbescognitives	0.00	0.02	0.01	0.00	0.40	0.01	0.01	0.01	0.03	0.00	0.02	0.01	0.02
7	0.147	spécial2-voir-tête	0.04	0.03	0.03	0.04	0.05	0.06	0.04	0.03	0.03	0.04	0.03	0.05	0.04
8	0.076	société6-madame	0.26	0.01	0.01	0.02	0.01	0.01	0.01	0.01	0.00	0.01	0.00	0.01	0.01
9	0.094	famille	0.04	0.02	0.01	0.03	0.02	0.02	0.02	0.02	0.01	0.03	0.01	0.03	0.02
10	0.070	société5-monsieur	0.01	0.00	0.01	0.02	0.00	0.02	0.08	0.01	0.01	0.02	0.00	0.01	0.00
11	0.155	heures-du-jour	0.03	0.02	0.02	0.04	0.03	0.05	0.05	0.03	0.03	0.05	0.02	0.03	0.03
12	0.052	enquête7-demander	0.00	0.02	0.01	0.00	0.01	0.00	0.00	0.01	0.02	0.00	0.02	0.01	0.41
13	0.098	spécial3-visage	0.02	0.03	0.03	0.02	0.03	0.02	0.02	0.03	0.06	0.02	0.02	0.04	0.02
14	0.138	amour1-famille	0.03	0.01	0.02	0.06	0.02	0.05	0.06	0.02	0.01	0.08	0.02	0.02	0.02
15	0.060	espaces1-souterrain	0.02	0.02	0.02	0.01	0.01	0.03	0.03	0.02	0.01	0.03	0.01	0.01	0.02
16	0.034	navigation	0.01	0.01	0.01	0.01	0.00	0.01	0.02	0.00	0.01	0.02	0.01	0.01	0.01
17	0.042	enquête6-inspecteur	0.00	0.32	0.02	0.00	0.01	0.01	0.00	0.01	0.06	0.00	0.04	0.01	0.02
18	0.079	argent	0.03	0.01	0.01	0.05	0.01	0.02	0.01	0.02	0.01	0.02	0.01	0.02	0.01
19	0.059	enquête1-suivre	0.00	0.01	0.03	0.00	0.01	0.01	0.00	0.35	0.03	0.00	0.01	0.01	0.01
20	0.214	enquête3-communiquer	0.07	0.03	0.05	0.05	0.05	0.05	0.10	0.05	0.05	0.04	0.04	0.06	0.06
21	0.070	lecture-ecriture	0.01	0.02	0.01	0.02	0.02	0.03	0.02	0.02	0.02	0.02	0.02	0.02	0.01
22	0.036	société2-empereur	0.00	0.00	0.01	0.00	0.00	0.01	0.07	0.00	0.00	0.01	0.00	0.00	0.00
23	0.137	crime3-coup	0.03	0.02	0.02	0.04	0.02	0.07	0.08	0.02	0.02	0.06	0.02	0.02	0.02
24	0.065	beauté-féminine	0.04	0.01	0.02	0.02	0.01	0.02	0.02	0.01	0.01	0.01	0.01	0.01	0.01
25	0.065	société3-monsieur	0.01	0.00	0.00	0.01	0.00	0.01	0.01	0.00	0.01	0.17	0.00	0.00	0.00
26	0.031	special1-verybritish	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.10	0.00	0.00	0.00
27	0.029	militaire1-officier	0.07	0.00	0.00	0.01	0.00	0.00	0.00	0.00	0.00	0.01	0.00	0.00	0.00
28	0.019	société1-roi	0.00	0.00	0.00	0.04	0.00	0.00	0.01	0.00	0.00	0.00	0.00	0.00	0.00
29	0.118	espaces2-maison	0.03	0.03	0.03	0.02	0.03	0.04	0.05	0.03	0.04	0.04	0.01	0.04	0.02
30	0.117	crime2-mort	0.05	0.02	0.02	0.03	0.02	0.04	0.05	0.05	0.01	0.02	0.02	0.03	0.03
31	0.073	boire-manger	0.01	0.04	0.02	0.01	0.02	0.01	0.02	0.02	0.04	0.01	0.02	0.03	0.02
32	0.024	enquête4-argot	0.00	0.02	0.01	0.00	0.00	0.00	0.00	0.01	0.03	0.00	0.41	0.00	0.01
33	0.052	automobilité	0.00	0.06	0.03	0.00	0.05	0.01	0.01	0.03	0.07	0.00	0.05	0.02	0.02
34	0.071	espaces2-paysagevillageois	0.02	0.02	0.01	0.02	0.01	0.04	0.03	0.01	0.03	0.03	0.02	0.01	0.02
35	0.077	enquête5-commissaire	0.01	0.02	0.03	0.01	0.03	0.02	0.01	0.02	0.03	0.00	0.02	0.39	0.02
36	0.040	maladie-guerison	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.00	0.01	0.01	0.01	0.01
37	0.049	vague1-suivre	0.00	0.01	0.36	0.00	0.00	0.00	0.00	0.01	0.02	0.00	0.02	0.00	0.01
38	0.076	voyager	0.05	0.01	0.01	0.03	0.01	0.01	0.02	0.01	0.01	0.05	0.01	0.01	0.01
39	0.035	militaire2-général	0.01	0.01	0.00	0.01	0.00	0.02	0.04	0.00	0.00	0.01	0.00	0.00	0.00

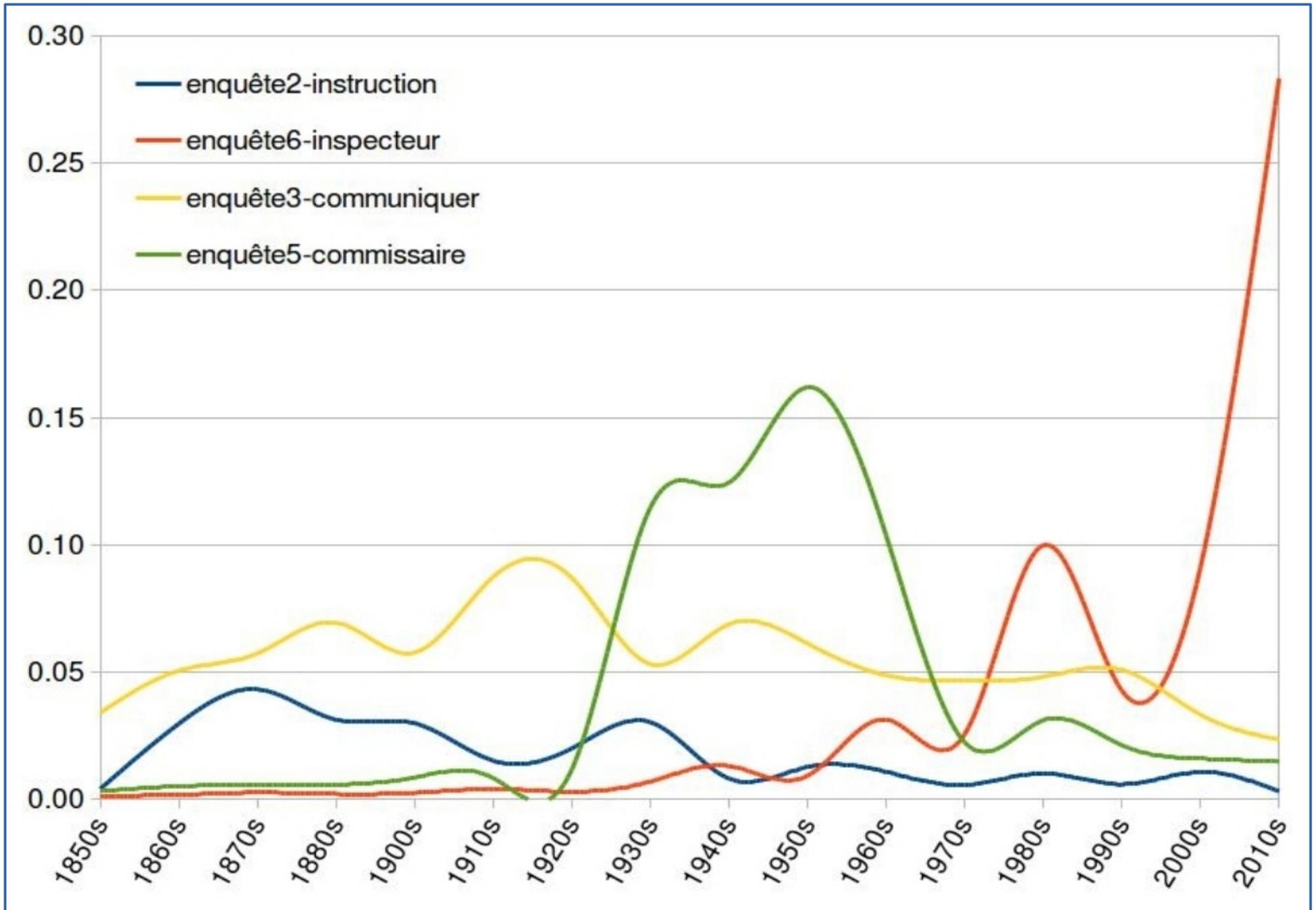
„famille“ et „argent“ par décennie



„crimes“ par décennie



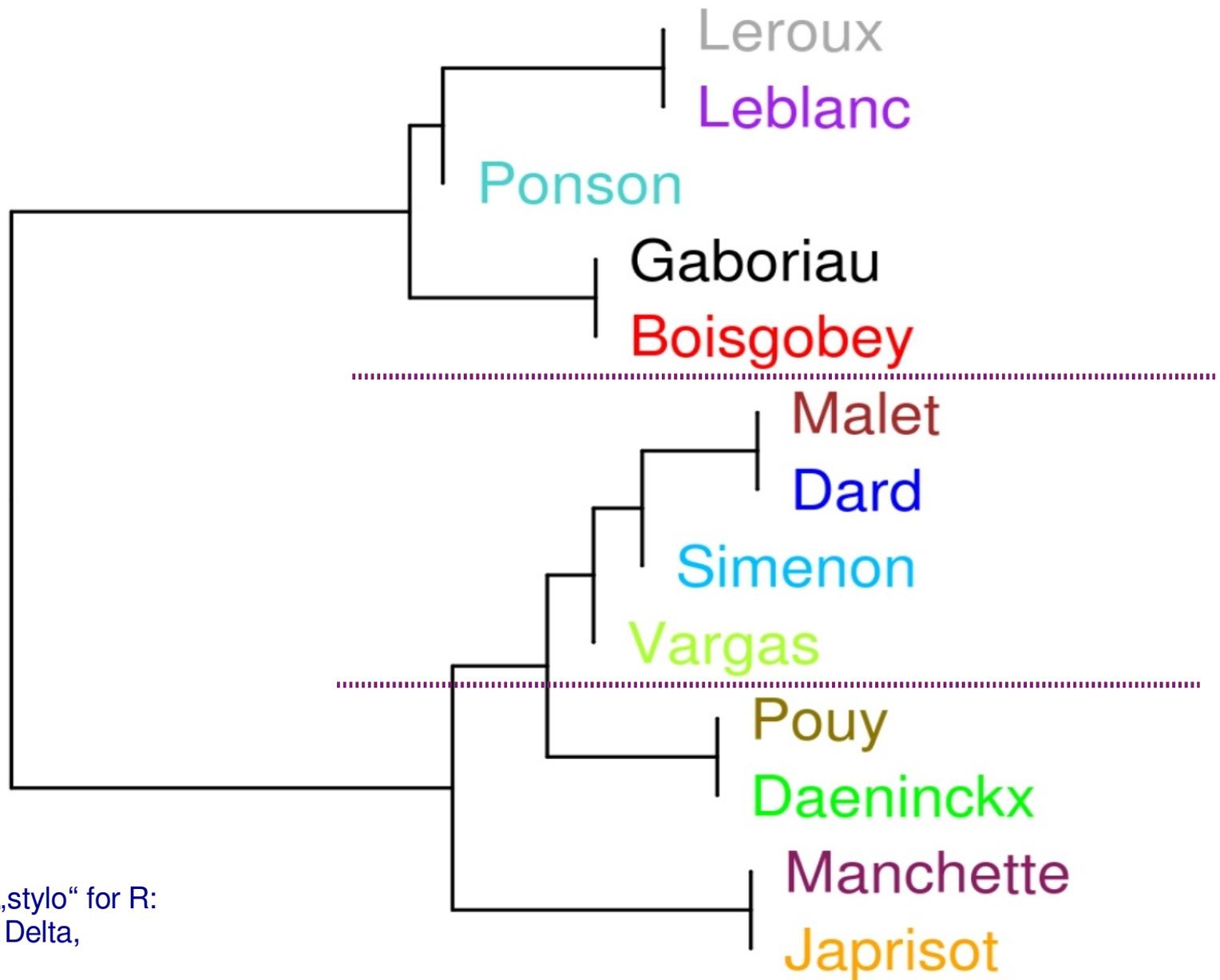
„enquêtes“ par décennie



Les topics par sous-genres / phase

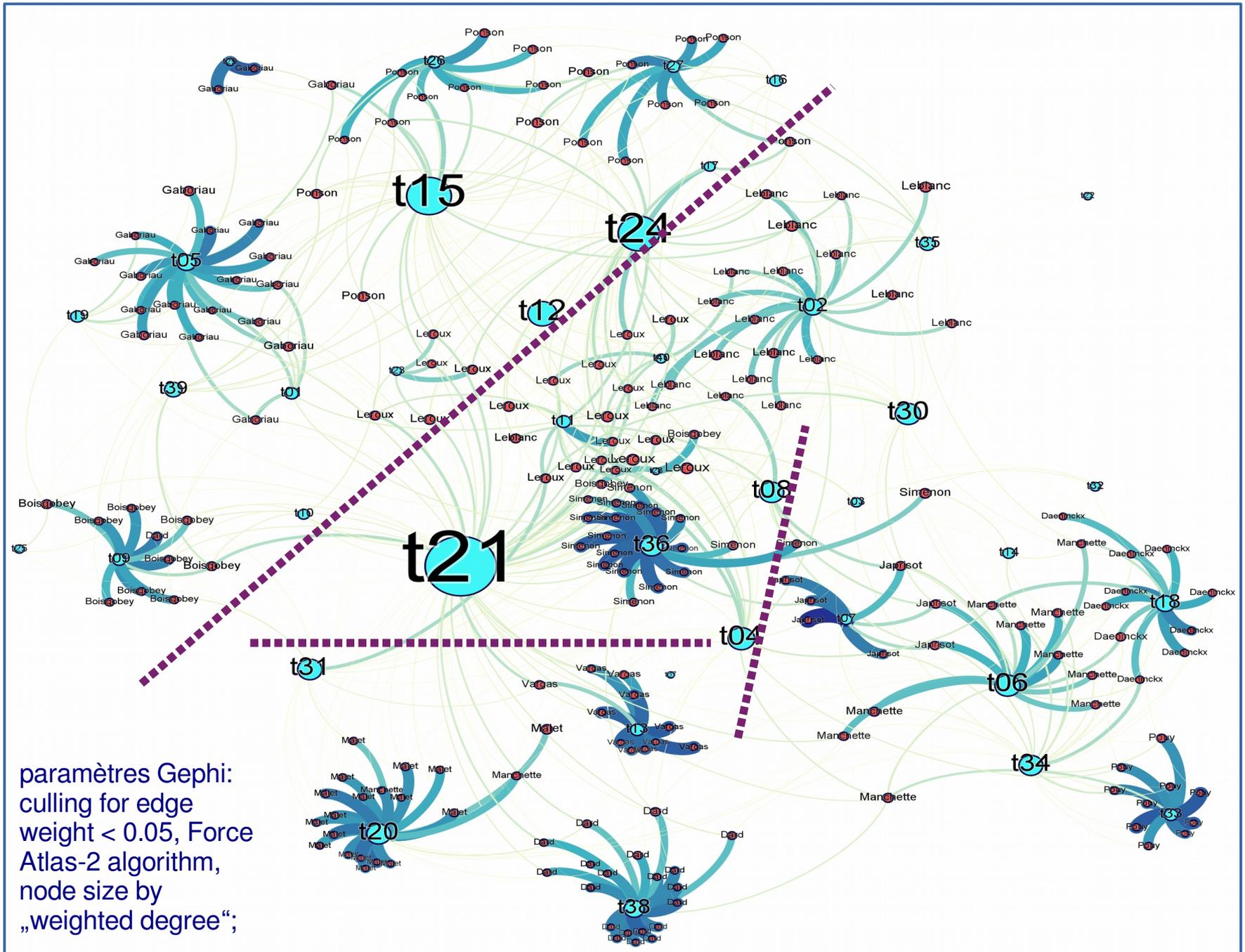
topic	score	label	judiciaire	archaïque	detection	noir	neopolar	parodique
0	0.068	enquête2-instruction	0.03	0.02	0.02	0.01	0.01	0.01
1	0.068	crime4-actionsecrete	0.01	0.08	0.01	0.01	0.01	0.01
2	0.081	nocturne	0.02	0.03	0.02	0.02	0.02	0.02
3	0.119	existence-mondaine	0.02	0.05	0.03	0.03	0.03	0.05
4	0.132	société4-monsieur	0.12	0.02	0.01	0.01	0.01	0.01
5	0.062	crime1-armeafeu	0.00	0.01	0.02	0.03	0.10	0.02
6	0.040	spécial3-verbescognitives	0.00	0.00	0.01	0.01	0.13	0.01
7	0.147	spécial2-voir-tête	0.04	0.05	0.05	0.03	0.03	0.02
8	0.076	société6-madame	0.07	0.01	0.01	0.01	0.00	0.01
9	0.094	famille	0.03	0.02	0.03	0.02	0.02	0.02
10	0.070	société5-monsieur	0.01	0.05	0.01	0.01	0.00	0.01
11	0.155	heures-du-jour	0.04	0.05	0.03	0.03	0.02	0.02
12	0.052	enquête7-demander	0.00	0.00	0.00	0.16	0.02	0.01
13	0.098	spécial3-visage	0.02	0.02	0.03	0.03	0.03	0.03
14	0.138	amour1-famille	0.06	0.06	0.02	0.02	0.01	0.03
15	0.060	espaces1-souterrain	0.02	0.03	0.01	0.02	0.01	0.01
16	0.034	navigation	0.01	0.02	0.01	0.01	0.01	0.01
17	0.042	enquête6-inspecteur	0.00	0.00	0.01	0.02	0.11	0.02
18	0.079	argent	0.03	0.01	0.02	0.01	0.01	0.01
19	0.059	enquête1-suivre	0.00	0.00	0.01	0.16	0.02	0.02
20	0.214	enquête3-communiquer	0.05	0.08	0.06	0.06	0.04	0.04
21	0.070	lecture-ecriture	0.02	0.02	0.02	0.01	0.02	0.01
22	0.036	société2-empereur	0.00	0.04	0.00	0.00	0.00	0.01
23	0.137	crime3-coup	0.04	0.08	0.02	0.02	0.02	0.02
24	0.065	beauté-féminine	0.02	0.02	0.01	0.01	0.01	0.02
25	0.065	société3-monsieur	0.07	0.01	0.00	0.00	0.00	0.00
26	0.031	special1-verybritish	0.04	0.00	0.00	0.00	0.00	0.00
27	0.029	militaire1-officier	0.02	0.00	0.00	0.00	0.00	0.00
28	0.019	société1-roi	0.02	0.00	0.00	0.00	0.00	0.00
29	0.118	espaces2-maison	0.03	0.04	0.04	0.03	0.03	0.02
30	0.117	crime2-mort	0.03	0.04	0.03	0.04	0.02	0.02
31	0.073	boire-manger	0.01	0.01	0.03	0.02	0.03	0.02
32	0.024	enquête4-argot	0.00	0.00	0.00	0.01	0.11	0.01
33	0.052	automobilité	0.00	0.01	0.02	0.03	0.06	0.02
34	0.071	espaces2-paysagevillageois	0.02	0.03	0.01	0.01	0.02	0.01
35	0.077	enquête5-commissaire	0.01	0.01	0.39	0.03	0.02	0.02
36	0.040	maladie-guerison	0.01	0.01	0.01	0.01	0.01	0.01
37	0.049	vague1-suivre	0.00	0.00	0.00	0.07	0.01	0.40
38	0.076	voyager	0.04	0.01	0.01	0.01	0.01	0.01
39	0.035	militaire2-général	0.01	0.03	0.00	0.00	0.00	0.00

Clustering selon le „taux des topics“

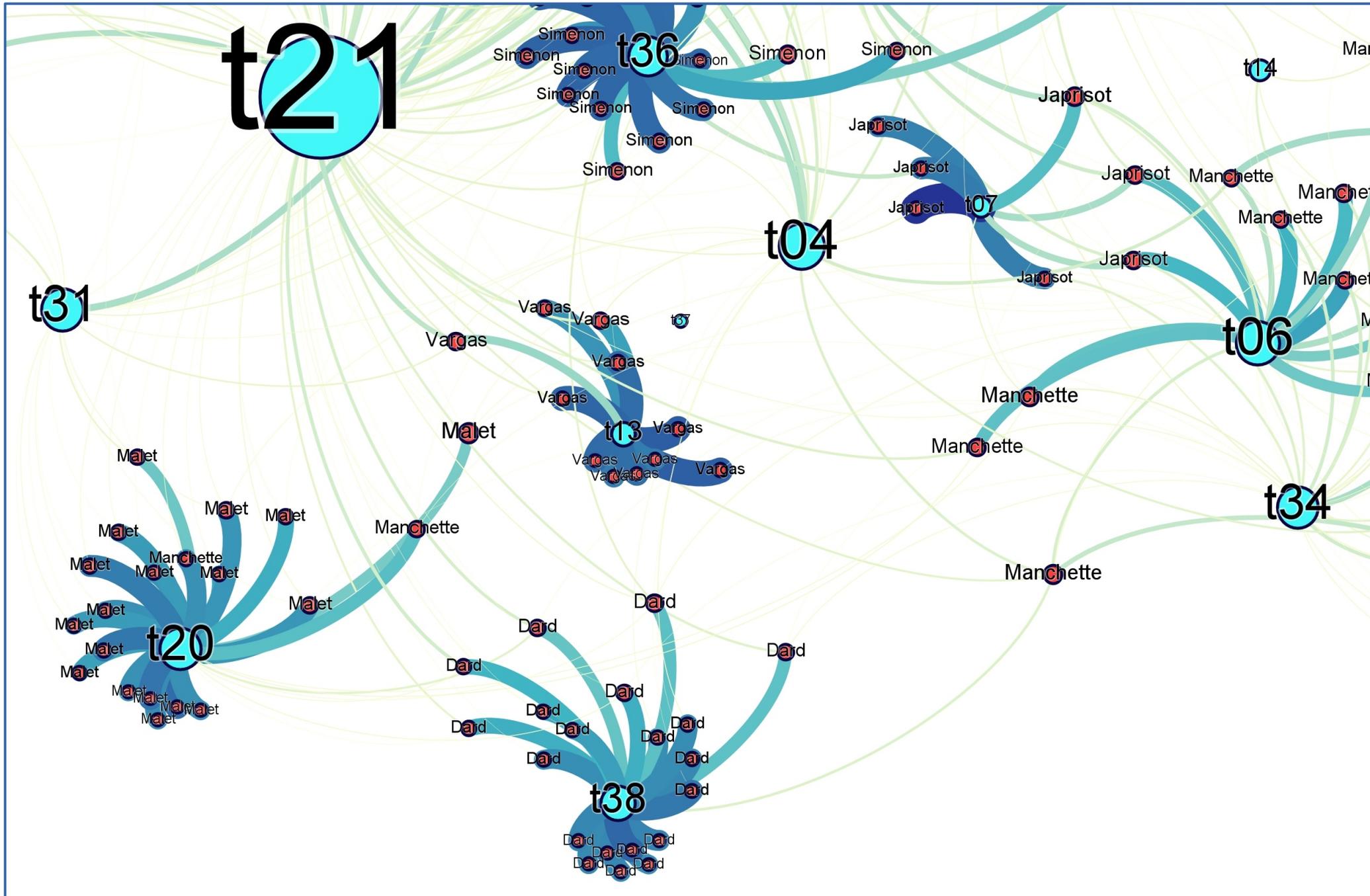


paramètres dans „stylo“ for R:
15 topics, Classic Delta,
Ward Clustering.

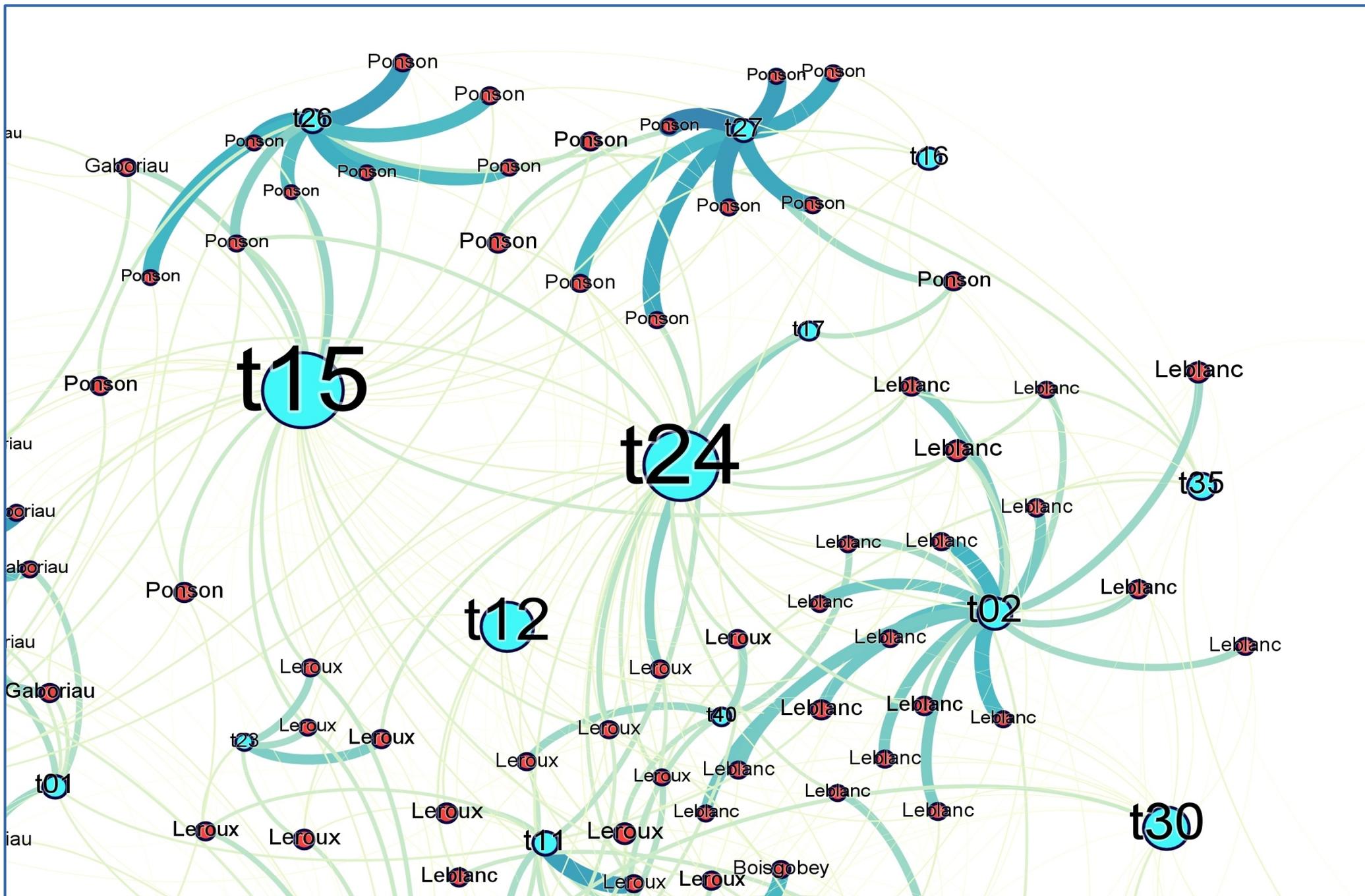
Un réseau topics-auteurs : genres?



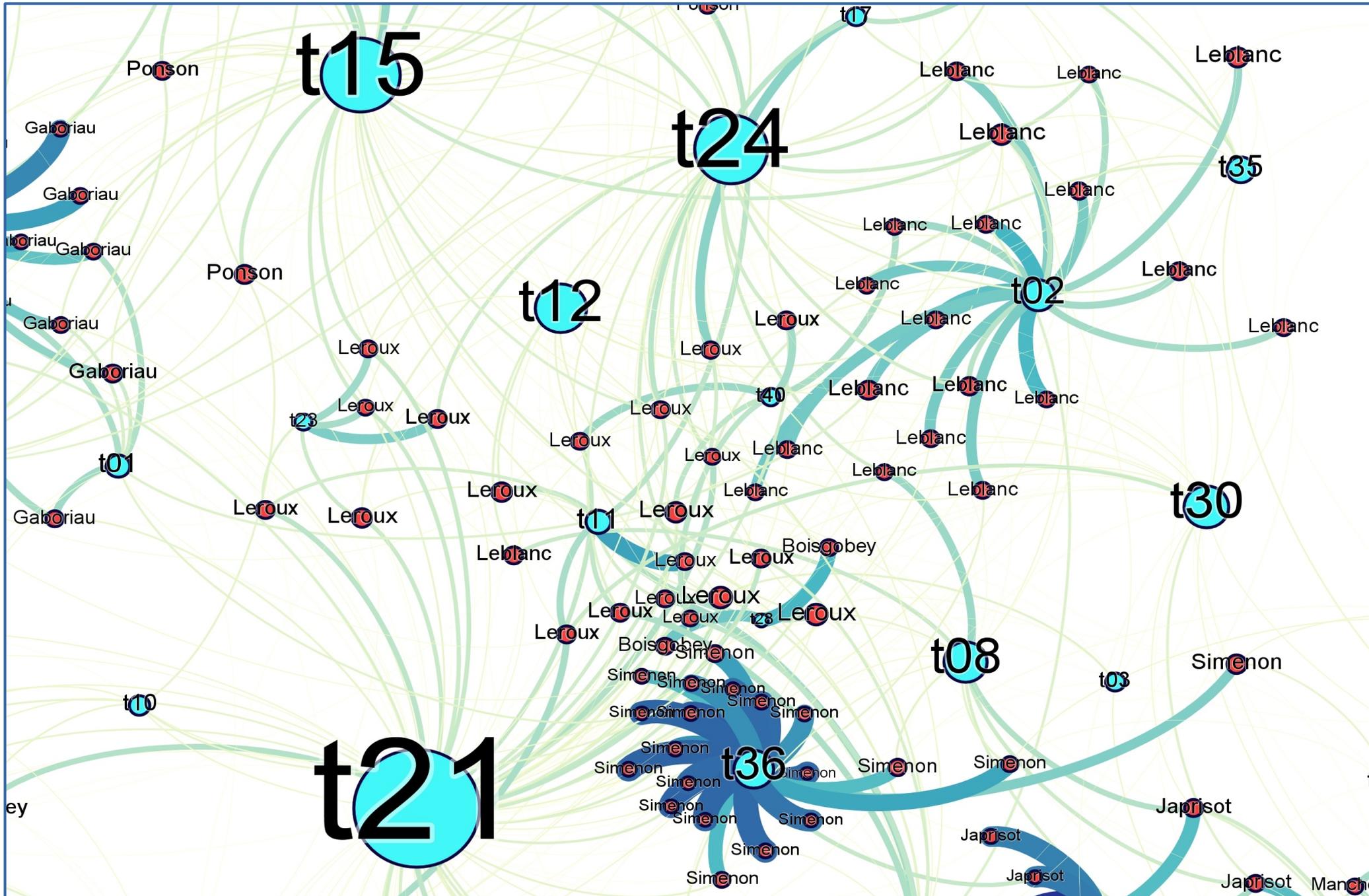
Les „auteurs-îles“: Dard, Malet



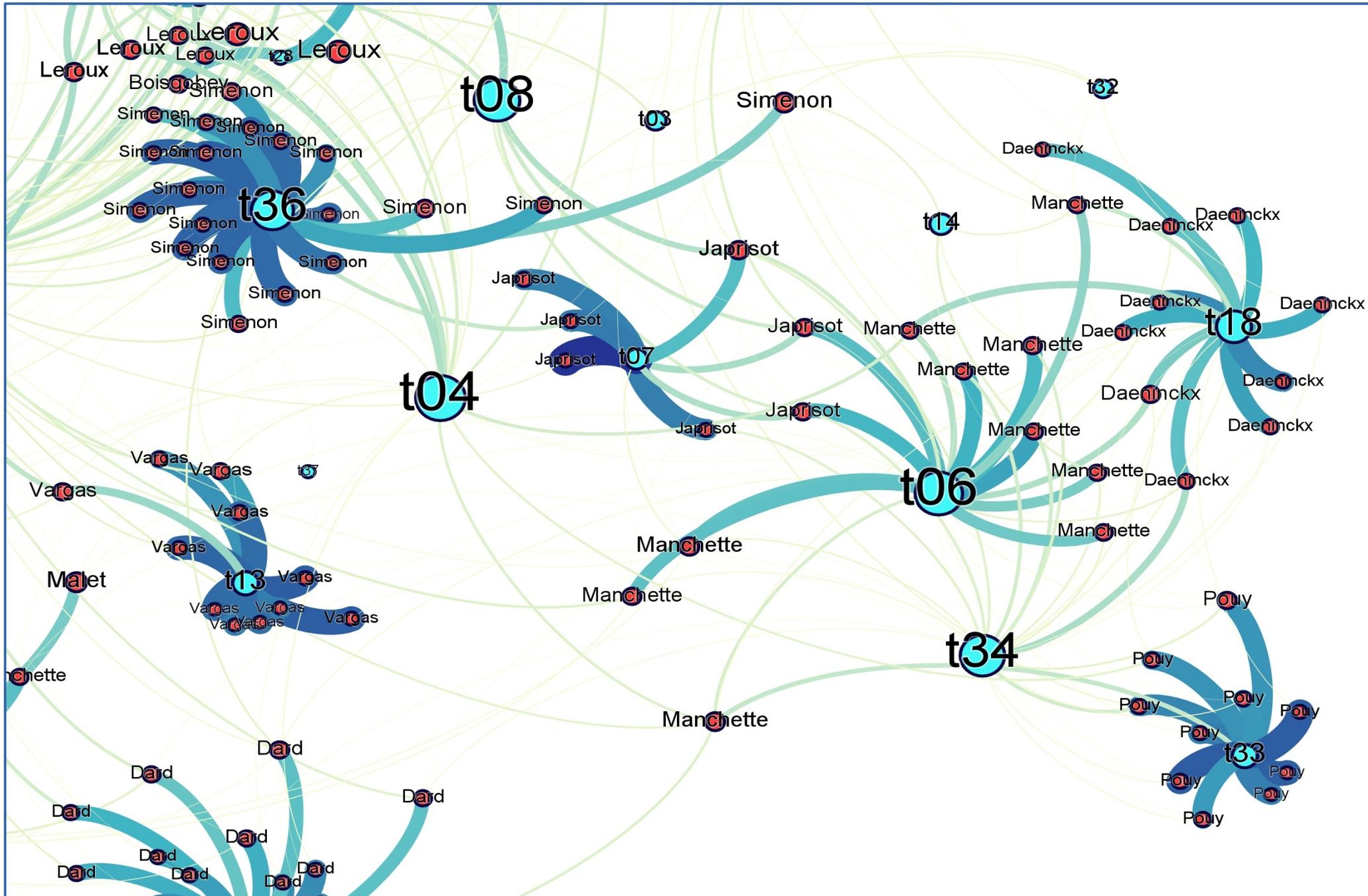
Les „topics-pont“



Leroux et Leblanc solidaires



Les auteurs de néopolar



6. bilan

Quelques résultats

Les topics obtenus

- les topics obtenus sont (presque tous) cohérents
- différents types de cohérence (thèmes, motifs, styles)
- topics et sous-topics (enquête, crime, société)

Approche supervisée (catégories *a priori*)

- presque tous les auteurs ont un „signature topic“
- tendances chronologiques assez claires (mais: auteurs)
- six types de romans: tendances moins convaincantes

Approches non supervisées (pas de catégories *a priori*)

- clustering: des groupements sensés
- réseau: cartographie intéressante

Réflexion méthodologiques

Améliorations futures

- plus de textes, meilleurs métadonnées
- hierarchical topic modeling? supervised topic modeling?

Le topic modeling comme interface

- la préparation des textes, les paramètres influent sur les résultats
- principe non-supervisée du tm est une contrainte forte
- flexibilité de Python, Mallet une „black box“

Références

- Blei, David M. 2011. "Introduction to Probabilistic Topic Models." Communication of the ACM.
- Boileau-Narcejac. 1975. *Le roman policier*. Paris: PUF.
- Doueïhi, Milad. 2012. Pour un humanisme numérique (2011). Paris: publie.net.
- McCallum, Andrew. 2002 "MALLET: A Machine Learning for Language Toolkit." <http://mallet.cs.umass.edu>. University of Massachusetts Amherst.
- Ramsay, Stephen. 2011. *Reading Machines. Toward an Algorithmic Criticism*. Urbana: Univ. of Illinois Press.
- Schulz-Buschhaus, Ulrich. 1975. *Formen und Ideologien des Kriminalromans : ein gattungsgeschichtlicher Essay*. Frankfurt am Main: Humanitas-Verlag.
- Todorov, Tzvetan. 1971. „Typologie du roman policier“. *Poétique de la prose*. Paris: Seuil, 55-65.
- Samuels, Lisa, and Jerome McGann. "Deformance and Interpretation." *New Literary History* 30, no. 1 (1999): 25–56.
- Steyvers, Mark, and Tom Griffiths. "Probabilistic Topic Models." *Latent Semantic Analysis: A Road to Meaning*, ed. T. Landauer et al. Laurence Erlbaum, 2006.

Christof Schöch

Universität de Würzburg, Allemagne

Departement de philologie numérique

Groupe “Stylistique computationnelle des genres littéraires”

<http://www.christof-schoech.de/fr>

Atelier „Les interfaces numériques dans la recherche d'aujourd'hui“

Frankoromanistentag, Münster, 24.-27. septembre 2014