

Analyse thématique du pétitionnement en ligne

THOMAS SOUBIRAN

CERAPS (UMR 8026 CNRS - Université de Lille)

Colloque du projet ANR APPEL

INTERNET ET LES NOUVELLES FORMES DE PARTICIPATION POLITIQUE

Lille, 29 mars 2019

- ▶ analyse de données issues du site de pétitionnement en ligne **lapetition.be**
- ▶ qui illustre la tendance qui conduit à disposer de collections de plus en plus **volumineuses** de **toute nature**
- ▶ en l'occurrence, un volume conséquent de
 - ▶ **données textuelles** (les pétitions)
 - ▶ et **relationnelles** (les signatures aux pétitions)
- ▶ **objet de la communication** : utilisation **des modèles thématiques** pour analyser
 - ▶ **les textes** des pétitions
 - ▶ **les signatures multiples** aux pétitions

- ▶ les modèles thématiques désignent une classe de modèles statistiques qui vise à **organiser l'information**
- ▶ pour faire ressortir **des thèmes** dans des collections de documents
- ▶ l'application au corpus du texte des pétitions montre qu'ils permettent aussi d'éclairer **les relations entre pétitions**
- ▶ au travers des signataires qu'elles ont **en commun**
- ▶ en les regroupant **par thèmes**

- ▶ La base lapetition.be
- ▶ Analyse thématique des textes des pétitions
- ▶ Analyse thématique des signatures aux pétitions

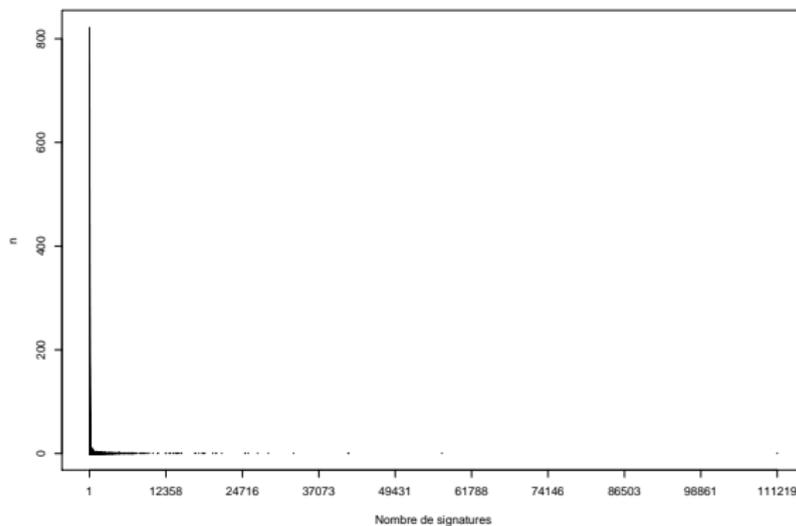
La base lapetition.be

- ▶ site lancé **fin 2006**
- ▶ extraction de la base SQL servant de *backend* au site en **février 2015**
- ▶ la base renseigne
 - ▶ plus de **15 000 pétitions**
en fait, plutôt 12 4000, le nombre variant en fonction de la table considérée. . .
 - ▶ près de **3,8 millions** de signatures
- ▶ la base renseigne aussi (avec un nombre de données manquantes variables) :
 - ▶ hash des adresses électroniques utilisées pour valider les signatures
ces hashes pouvant correspondre à plusieurs personnes utilisant l'adresse
 - ▶ prénom du signataire
 - ▶ horodatage de la signature et de sa validation
 - ▶ la commune de résidence du signataire
 - ▶ . . .

Le graphe des signatures

- ▶ la base ne contient que **peu de renseignements** sur les signataires pour analyser la dynamique pétitionnaire
- ▶ mais elle permet de générer **un graphe**
- ▶ toutefois, le graphes (biparti) des signatures aux pétitions présente différentes **propriétés**
- ▶ qui rendent difficile son analyse comme **un réseau**
*ces difficultés ne se réduisant pas à la seule **taille** du graphe*
- ▶ et donc la détection, p. ex., de **communautés** se mobilisant autour de causes spécifiques
- ▶ ce qui constitue **un problème structurel** pour les grands graphes
- ▶ mais qui se trouve **renforcé** ici par certaines des caractéristiques du site

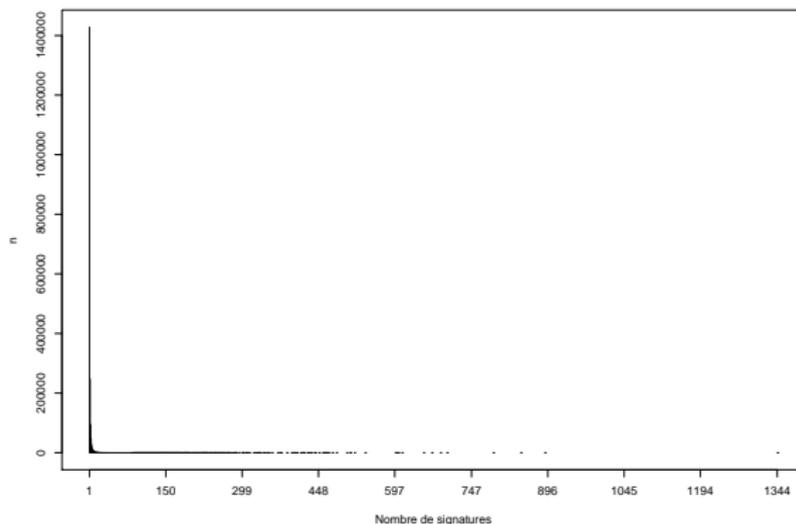
Distribution du nombre de signatures par pétitions



Déciles du nombre de signatures aux pétitions :

	0%	10%	20%	30%	40%	50%	60%	70%	80%	90%	100%
N	1	2	4	7	13	23	45	89	192	544.20	111219

Distribution du nombre de signatures par signataires

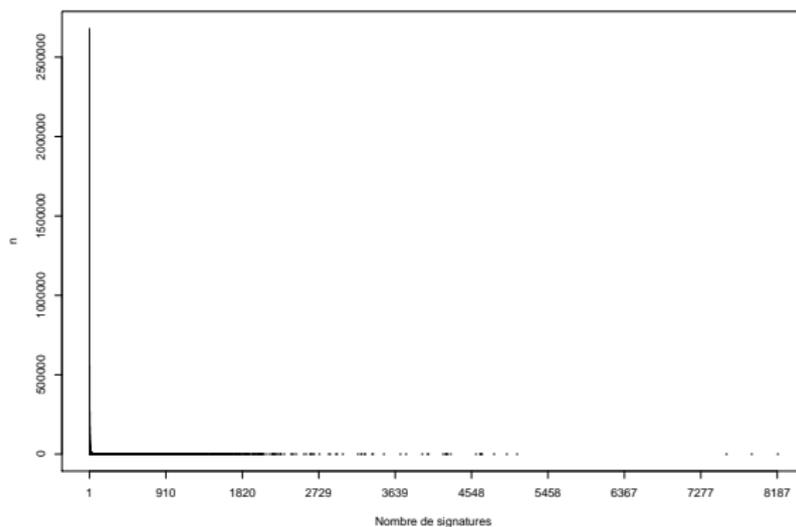


Déciles du nombre de signatures par hash d'adresse électronique :

	0%	10%	20%	30%	40%	50%	60%	70%	80%	90%	100%
N	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	2.00	3.00	1344.00

Note : Le nombre maximal est celui d'un botnet publicitaire. Le nombre maximal de pétitions signées par un humain est de 890.

Distribution du nombre de signatures partagées par les pétitions

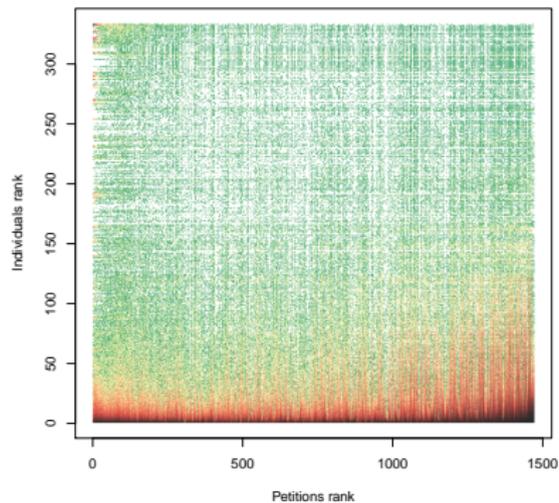


Déciles du nombre de signatures partagées par les pétitions (pour $w_{ij} > 0$) :

	0%	10%	20%	30%	40%	50%	60%	70%	80%	90%	100%
N	1.00	1.00	1.00	1.00	1.00	1.00	2.00	2.00	4.00	12.00	8187.00

Note : pour 6% des arêtes (4 603 180), $w_{ij} > 0$ pour un total de 31 786 165 liens

Un site peu « communautaire »



Cf. (DYER, 2010)

- ▶ difficultés pour **l'analyse**
 - ▶ données volumineuses mais avec peu de renseignements (bruités), corrélées et déséquilibrées
 - ▶ textes comportant beaucoup de fautes (et parfois difficilement intelligibles)
 - ▶ propriétés même du graphe
 - ▶ extrême asymétrie des distributions
 - ▶ comporte beaucoup de liens mais épars
 - ▶ cf. (CONTAMIN, LÉONARD et SOUBIRAN, 2017)
- ▶ par voie de conséquence, difficultés à faire ressortir **des réseaux** ou **des grappes** de signataires
 - ▶ notamment de par un effet de plate-forme (cf. *infra* p. 34)
 - ▶ même si on peut trouver de nombreux indices de leur présence

- ▶ **la dynamique** des signatures apparaît largement **exogène** au site
et que la dynamique interne ne produit pas « communautés » d'utilisateurs
- ▶ le graphe est plutôt **une manifestation** de réseaux plutôt que le réseau lui-même
- ▶ ces réseaux n'expliquant **qu'une partie** des liens
 - ▶ les réseaux n'étant **qu'une manière** parmi d'autres de diffuser une pétition
 - ▶ à commencer par **le site lui-même**
on verra que les liens entre pétitions, et donc entre individus, est très largement le produit de **l'interaction avec le site**
- ▶ l'analyse du graphe des pétitions a toutefois conduit à formuler l'hypothèses de relations **thématiques** entre les signatures
- ▶ pour appréhender la dynamique **propre au site**

L'hypothèse a conduit à adopter la démarche suivante :

1. **regrouper** les pétitions **par thèmes**

au moyen de modèles thématiques

2. puis,

- 2.a générer **le graphe** des co-signatures aux pétitions

*soit le graphe dont les arrêtes comportent le nombre de signatures que deux pétitions ont **en commun***

- 2.b **regrouper** par thème **les co-signatures** aux pétitions

3. analyser **le graphe des pétitions** ainsi regroupées

au moyen d'une analyse spectrale du graphe

Analyse thématique des textes des pétitions

- ▶ approche issue de **l'informatique**
- ▶ visant à **faire ressortir** des thèmes, des sujets ou encore des concepts de collections de documents
- ▶ pour **décrire, organiser et requérir ces collections**
*ils ont en effet développé dans une perspective de **récupération de l'information** et de **fouille textuelle***
- ▶ ils reposent sur le postulat que les documents ont une **structure inobservée** (latente)
- ▶ et que cette structure peut être inférée à partir des **co-occurrences** des mots contenus dans les documents
- ▶ le but est alors d'obtenir p. ex. **des classes ou des dimensions** permettant de faire ressortir la composition latente des documents

- ▶ les modèles thématiques (et d'autres) reposent sur l'analyse de matrice **termes–documents** :

$$\begin{pmatrix} n_{1,1} & \dots & n_{1,V} \\ \vdots & \ddots & \vdots \\ n_{1,D} & \dots & n_{D,V} \end{pmatrix}$$

où $n_{d,v}$ compte le nombre d'occurrences du terme v dans le document d

- ▶ et cherchent à factoriser la matrice de départ en **deux matrices** de dimensions réduites
- ▶ qui donnent, respectivement
 - ▶ la position **des documents** dans l'espace latent
 - ▶ la position **des termes** dans l'espace latent
- ▶ **différents** modèles thématiques ont été proposés
- ▶ la méthode utilisée ici a pour nom **l'allocation de Dirichlet latente** (LDA) (BLEI, NG et JORDAN, 2003)

- ▶ pour la LDA :

- ▶ proportion de **chaque classe** dans chaque document

$$\theta = \begin{pmatrix} \theta_1 \\ \vdots \\ \theta_D \end{pmatrix} = \begin{pmatrix} \theta_{1,1} & \cdots & \theta_{1,K} \\ \vdots & \ddots & \vdots \\ \theta_{1,D} & \cdots & \theta_{D,K} \end{pmatrix} \text{ avec } \sum_{k=1}^K \phi_{d,k} = 1$$

- ▶ proportion de **chaque terme** dans chaque classe

$$\phi = \begin{pmatrix} \phi_1 \\ \vdots \\ \phi_K \end{pmatrix} = \begin{pmatrix} \phi_{1,1} & \cdots & \phi_{1,V} \\ \vdots & \ddots & \vdots \\ \phi_{1,K} & \cdots & \phi_{K,V} \end{pmatrix} \text{ avec } \sum_{v=1}^V \phi_{k,v} = 1$$

- ▶ **Note** : l'interprétation de ces deux matrices varient selon la méthodes

- ▶ analyse **morpho-syntaxique** avec udpipe
 - avec beaucoup de déchets et d'erreurs
- ▶ **suppression** de la ponctuation, des conjonctions, des déterminants, . . .
 - avec le problème que tous les mots n'ont pas été correctement étiquetés
- ▶ ont été **retenus** :
 - ▶ les mots apparaissant dans au moins quatre documents
 - ▶ les pétitions comportant au moins sept mots du vocabulaire
- ▶ **Note** : les textes **des pétitions non signées** n'ont pas été inclus dans l'analyse

▶ test sur **les pétitions libellées** :

- ▶ « Droits de l'Homme »
- ▶ « Environnement, nature et écologie »
- ▶ « Politique »
- ▶ « Social »
- ▶ soit un total de 8 159 pétitions

n'ont pas été incluses les catégories de pétitions : « Art et culture », « Protection animale », « Loisirs », « Humour/Insolite », « Autres »

▶ modèle à **130 classes**

▶ validation croisée de **la perplexité du modèle**

$$b^{H(\tilde{p}, q)} = b^{-\sum \tilde{p}(x) \log_b q(x)}$$

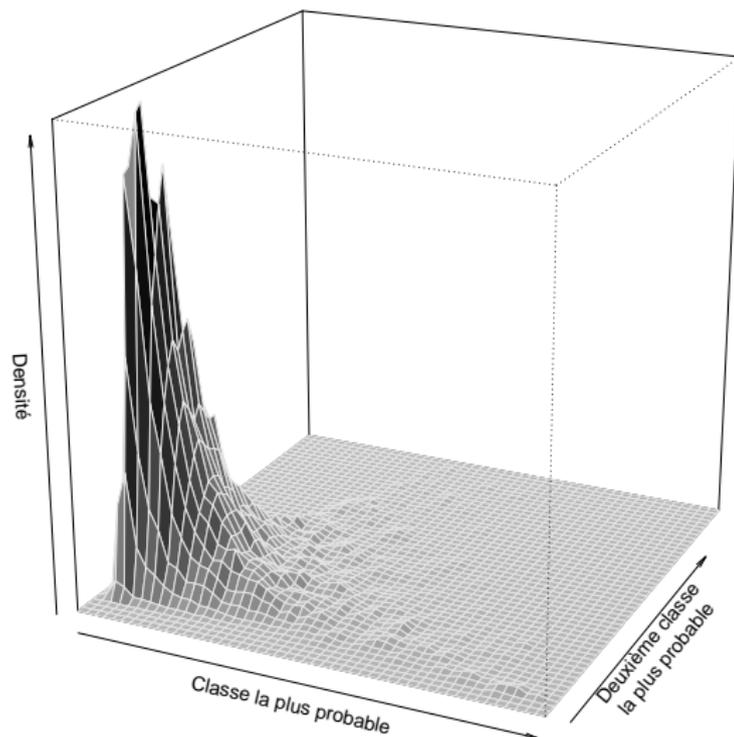


FIGURE 1: Densité des probabilités θ des deux classes les plus probables des documents

- ▶ densité des probabilités $\theta_{d,k}$ d'apparition **des thèmes les plus probables** dans chaque document
- ▶ la distribution indique **une grande dispersion**
- ▶ un nombre conséquent de textes de pétition ne sont **pas bien caractérisés** par le modèle

cf. les résultats de la classification des textes infra p. 32

- ▶ les classes résultant de l'analyse ne sont **pas toujours interprétables** comme des thèmes
- ▶ (ALSUMAIT et al., 2009) ont proposé un score pour distinguer **les classes insignifiantes** (« *junk topic* ») des autres
- ▶ ce score est une combinaison pondérée de **trois mesures de dissimilarité** :

$$S^{(k)} = \sum_u^U \Psi_u S_u^{(k)}$$

avec $S_u^{(k)}$, la mesure, Ψ_u , son poids total et U , le nombre de mesures

- ▶ une mesure **d'uniformité** qui mesure la dissimilarités entre $\phi^{(k)}$ et la distribution

$$p(w_n|\Omega^{\mathcal{U}}) = \frac{1}{V}, \forall n \in \{1, \dots, V\}$$

où tous les mots du vocabulaire sont considérés comme équiprobables.

- ▶ une mesure **de vacuité** qui mesure la dissimilarité entre $\phi^{(k)}$ et la distribution marginale des mots prédite par le modèle

$$p(w_n|\Omega^{\mathcal{V}}) = \sum_{k=1}^K p(w_n|k) p(k)$$

- ▶ une mesure **de communalité** des classes qui mesure la dissimilarité cette fois entre $\theta^{(k)}$ et la distribution

$$p(d_m|\Omega^{\mathcal{B}}) = \frac{1}{M}, \forall d \in \{1, \dots, D\}$$

qui donne une mesure de la dispersion des thèmes dans le corpus

- ▶ ces dissimilarités peuvent, entre autres possibilités, être calculées avec la divergence de Kullback-Leibler.

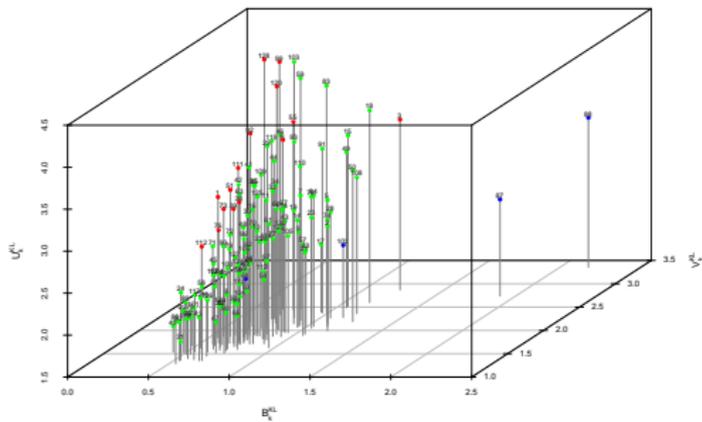


FIGURE 2: Dissimilarités U_k^{KL} , V_k^{KL} et B_k^{KL}

- ▶ classes spécifiques au pétitionnement :

- ▶ petition, signature, mail, envoyer, signer, adresse, faire, recevoir, lien, oublier, maximum, valider, connaissance courrier confirmer, recolter
- ▶ petition , signer, accord, lancer, but, merci, avance, inviter, partager, montrer, signataire, hesiter, demander, possible, circuler, soutenir

- ▶ classes pétitions rédigées comme des lettres ouvertes

monsieur , madame, ministre, messieurs, attention, president, mesdames, agreer, demander, remercier, part, expression, salutation, esperer, lettre

Registres des pétitions :

▶ **problème**

autre, avoir, etre, probleme, cas, faire, part, fait, point, solution, effet, question, poser, grand, nombre, agir

▶ **laisser-faire**

laisser, savoir, faire, jour, etre, voir, agir, peur, dire, monde, avoir, devenir, rendre, accepter, vivre, parler

▶ **écoute**

compte, tenir, choix, demander, avoir, prendre, moyen, avis, faire, choisir, permettre, possibilite, critere

▶ **décision**

decision, prendre, mesure, decider, concerner, consequence, agir, indispensable, voir, otage, titre, danger viser

▶ **changement**

grand, temps, changer, chose, faire

Note : toutes les pétitions ne relèvent pas d'un registre ou d'un autre

▶ registre et thème

▶ démocratie

citoyen, referendum, population, demander, democratie, populaire, donner, nouveau, élu, consultation, moyen, pouvoir, reforme, avis, initiative, souhaiter

▶ liberté

droit, liberte, homme, respect, humain, libre, respecter, fondamental, article, egalite, avoir, expression, declaration, dignite, universel, garantir, citoyen

▶ classes linguistiques

▶ les deux autres langues officielles de la Belgique

flamand et (très rarement) allemand

▶ mais aussi : anglais, castillan (+ català) et italien

▶ temporel

▶ 2009, 2010, 2011, 2008, mars, 2012, novembre, octobre, juin, mai, jour, janvier, septembre, 10, avril

▶ an, deux, mois, trois, annee, age, dernier, premier, fois, fin

▶ ainsi que qu'une classe de verbes, de numéraux,...

Classification des pétitions

- ▶ le modèle thématique ne donne que des probabilités **d'observer une thème** dans un documents
- ▶ et non pas **des classes** de documents
- ▶ les textes des pétitions ont été **classifiés** à partir de ces probabilités
- ▶ avec **des cartes auto-organisatrices** de Kohonen

en excluant les classes non-thématiques

Classification des pétitions

- ▶ **108 classes** de documents (grille de 9 × 12)
- ▶ dont **un blob** concernant 34 % des documents
- ▶ qui contient les pétitions qui n'ont **pas pu être caractérisées par** la classifications thématique
- ▶ les types de pétitions **les plus fréquentes** sont :

- ▶ Belgique

- ▶ et plus particulièrement la politique belge
- ▶ principalement, les crises politiques qu'à connu le pays au tournant des années 2000
 - constitution d'un gouvernement, unité de la Belgique, . . .*
- ▶ ou encore l'Islam, l'immigration
- ▶ et le sauvetage des fritkots

- ▶ fiscalité et le coût de la vie

- ▶ scolaire

principalement pour la création ou contre la fermeture de classe la situation d'un établissement, . . .

- ▶ différents type de pétitions faisant référence à la liberté, aux droits de l'homme

Classification des pétitions

- ▶ les type de pétitions **les plus fréquentes** (suite) :

- ▶ justice et, plus particulièrement la prison
- ▶ pétitions en flamand
- ▶ transport ferroviaire suppression de ligne, prix d'un trajet spécifique, horaire
- ▶ les expulsions
- ▶ opposition à des projets d'urbanisme
- ▶ enfance protection maltraitance
- ▶ ...

- ▶ **Note** : le regroupement peut conduire à obtenir des classes thématiquement **cohérentes** mais rassemblant des pétitions **antagonistes** :

Exemples :

- ▶ **justice** :

- ▶ laxisme
- ▶ emprisonnements arbitraires

- ▶ **éolien** : pour ou contre l'installation d'éoliennes

Pourquoi le blob ?

- ▶ **syntaxe** et **orthographe** (parfois très –très–) approximative des textes des pétitions
- ▶ déséquilibre dans **la taille** des textes

10% des textes sont constitués 11 mots ou moins (après sélection)

- ▶ **contenu thématique faible** (et pas toujours intelligible) × un vocabulaire très diffus

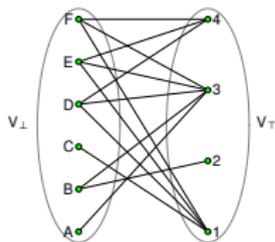
- ▶ le thème est parfois manifesté par quelques mots, le reste renvoyant à d'autres registres (indignation, déni de justice, appel à l'action, . . .)
- ▶ vocabulaire d'arrière-fond lie parfois plus les pétitions que le vocabulaire thématique

Analyse des signatures aux pétitions

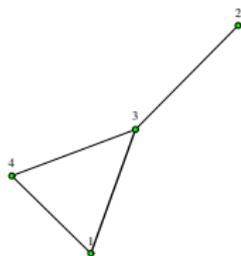
- ▶ les pétitions et les signataires sont liés par **un graphe biparti** (SOUBIRAN, 2017)
- ▶ un graphe biparti peut être défini par **un triplet** $G = (V_{\perp}, V_{\top}, E)$
- ▶ les sommets ne sont reliés **qu'aux sommets de l'autre ensemble**
- ▶ les deux extrémités de chaque arête $E \subseteq V_{\perp} \times V_{\top}$ appartiennent nécessairement à **deux groupes** différents
 - ▶ les signataires sont liés par les pétitions qu'ils ont signées
 - ▶ les pétitions sont liées par les signataires qui les ont signées

Les graphes bipartis

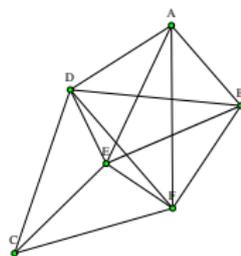
(a) Graphe biparti



(b) Graphe de V_T



(c) Graphe de V_{\perp}



(d) Matrice d'affiliation

	1	2	3	4
A	0	0	1	0
B	0	1	1	0
C	1	0	0	0
D	1	0	1	1
E	1	0	1	1
F	1	0	1	1

(e) Matrice d'adjacence de V_T

	1	2	3	4
1	4	0	3	3
2	0	1	1	0
3	3	1	5	3
4	3	0	3	3

(f) Matrice d'adjacence de V_{\perp}

	A	B	C	D	E	F
A	1	1	0	1	1	1
B	1	2	0	1	1	1
C	0	0	1	1	1	1
D	1	1	1	3	3	3
E	1	1	1	3	3	3
F	1	1	1	3	3	3

Les graphes bipartis

- ▶ un graphe biparti peut être représenté de façon compacte par une matrice \mathbf{B} de dimensions $|V_{\perp}| \times |V_{\top}|$ (**matrice d'affiliation**)
- ▶ avec, en l'occurrence, les signataires en ligne et les pétitions en colonne

	pétition ₁	...	pétition _j	...	pétition _{V_⊥}
signataire ₁	$\mathbb{1}_{1,1}$...	$\mathbb{1}_{1,j}$...	$\mathbb{1}_{1, V_{\top} }$
⋮	⋮	⋱	⋮	⋱	⋮
signataire _i	$\mathbb{1}_{i,1}$...	$\mathbb{1}_{i,j}$...	$\mathbb{1}_{i, V_{\top} }$
⋮	⋮	⋱	⋮	⋱	⋮
signataire _{V_⊥}	$\mathbb{1}_{ V_{\perp} ,1}$...	$\mathbb{1}_{ V_{\perp} ,j}$...	$\mathbb{1}_{ V_{\perp} , V_{\top} }$

- ▶ la matrice **d'adjacence** a quant à elle pour forme :

$$\mathbf{A} = \begin{bmatrix} \mathbf{0} & \mathbf{B} \\ \mathbf{B}^{\top} & \mathbf{0} \end{bmatrix}$$

- ▶ un graphe biparti peut donc être vu comme un cas particulier de graphe uniparti où les sommets sont répartis en deux ensembles distincts V_{\perp} et V_{\top}

Les graphes bipartis

- ▶ tout graphe biparti peut être transformé en deux graphes unipartis par **projection** :

- ▶ graphe des signataires

$$G_{\perp} = (V_{\perp}, \{\{u, v\} \mid \exists w \in V_{\top} : u \sim w, v \sim w\})$$

- ▶ graphes des pétitions

$$G_{\top} = (V_{\top}, \{\{u, v\} \mid \exists w \in V_{\perp} : u \sim w, v \sim w\})$$

- ▶ **les arêtes du graphe** renseignent, respectivement :

- ▶ le nombre de fois où deux \sim personnes ont signé la même pétition
- ▶ le nombre de personnes ayant signé la même pétition

- ▶ les projections peuvent être obtenues grâce aux **multiplications** suivantes de **B** :

$$\mathbf{A}_{\perp} = \mathbf{B} \mathbf{B}^{\top} \text{ projection dans } \perp$$

$$\mathbf{A}_{\top} = \mathbf{B}^{\top} \mathbf{B} \text{ projection dans } \top$$

- ▶ la projection permet de traiter les données **comme un graphe uniparti** pondéré mais avec une perte d'information (la dualité du graphe)
- ▶ et donne au graphe un caractère parfois **artificiel**

Le graphe biparti du pétitionnement

- ▶ problème de **la dualité** du graphe biparti
- ▶ les deux ensembles de sommets ne sont **pas directement** reliés
 - tout graphe biparti ne peut contenir que des cycles **impairs**
- ▶ la signature à une même pétition peut créer un lien entre des personnes n'ayant **jamais eu de contact** (et qui n'en auront jamais)
- ▶ en l'espèce, une partie des liens entre signataires sont le produit d'un effet de **la plate-forme**
 - la dynamique interne du site provient essentiellement de la navigation sur le site
- ▶ et semble expliquer une large part **des signatures multiples**
 - et masquent les mobilisations autour de pétitions*

Diagonalisation des signatures

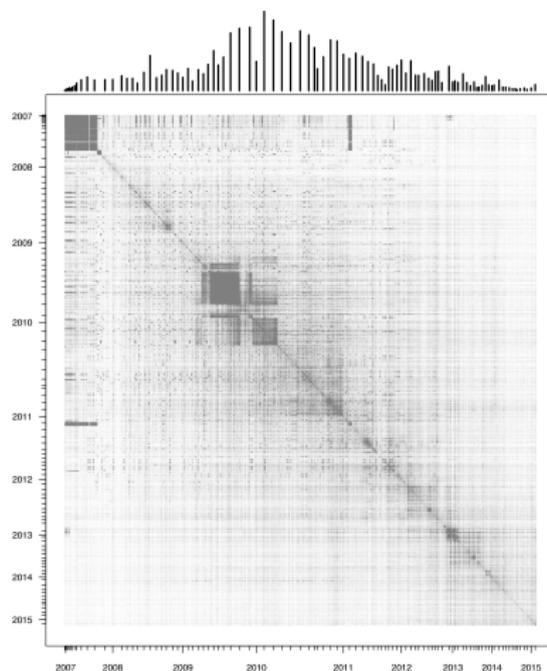


FIGURE 3: Relations binaires des pétitions dans le temps ($n_{ij} > 0$)

L'effet de plate-forme

- ▶ la diagonale montre que les personnes signent des pétitions **proches dans le temps**
- ▶ ainsi, 388 836 des 1 865 246 (.21) signatures multiples ont lieu dans **l'heure** suivant la précédente signature

tendance sans doute renforcée par la mise en avant sur le site des pétitions les plus signées du moment

- ▶ une large partie des liens entre les pétitions (et donc les signataires) provient très probablement de **la navigation sur le site**

la signature de pétitions en ligne conduit donc à la signature de pétitions que les personnes n'auraient peut-être jamais signées autrement

- ▶ à ce titre, il semble de plus que les comportements changent avec **la pratique du site**

plus le nombre de signatures augmente, plus le temps entre deux signatures se réduit

Temps écoulé depuis la signature précédente

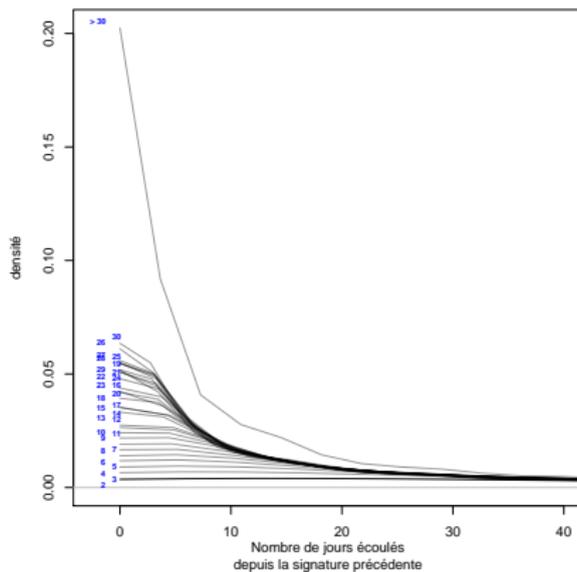


FIGURE 4: Temps écoulé depuis la signature précédente en fonction de l'ordre de signature

Analyse thématique des signatures aux pétitions

- ▶ l'analyse de \mathbf{A}_T fait avant tout ressortir la relation **temporelle** entre les signatures
- ▶ elle a aussi conduit à formuler l'hypothèse que les signatures multiples sur le site répondaient au moins partiellement à **une logique thématique**
 - ▶ les signatures multiples portaient parfois sur des pétitions aux thèmes similaires
 - ▶ certains thèmes de pétitions semblaient plus centraux, quel que soit le thème des pétitions auxquels elles étaient associées
- ▶ c'est pourquoi les entrées de \mathbf{A}_T ont été **regroupées** par thèmes
- ▶ et analysée par une décomposition **spectrale** du graphe pour faire ressortir les liens thématiques entre les signatures

- ▶ l'analyse spectrale vise à minimiser (entre autres possibilités)

$$N\text{cut}(S, G) = \frac{\text{cut}(S, G)}{\text{vol}(S, G)}$$

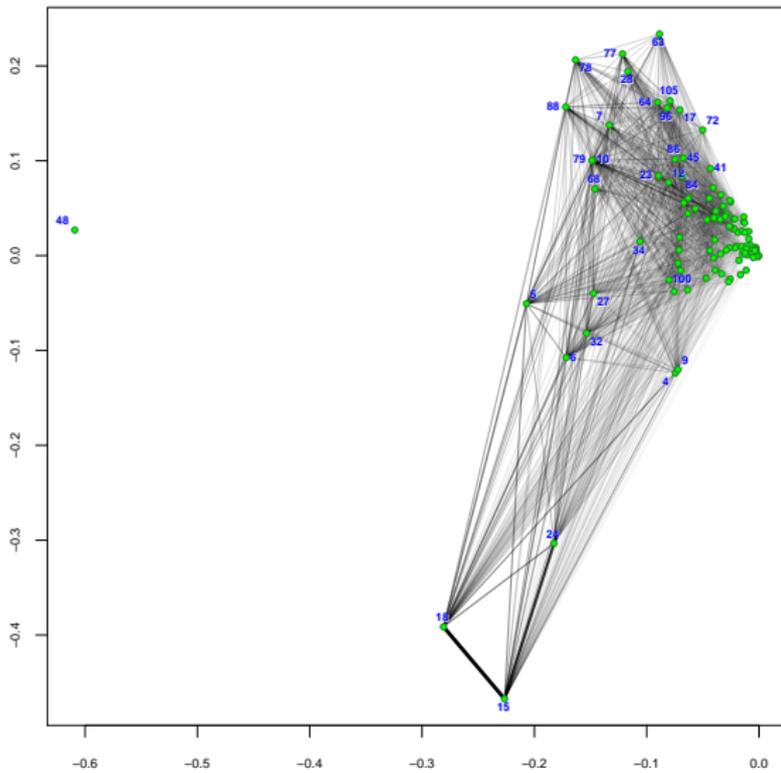
où, pour un sous-ensemble $S \subseteq V$ $\text{cut}(S, G) = \sum_{v_i \in S, v_j \in S^c} a_{ij}$

- ▶ soit, à minimiser le nombre de liens entre chaque grappes
- ▶ critère alternatif (QIN et ROHE, 2013) qui permet de corriger l'effet des différences entre les degrés des sommets

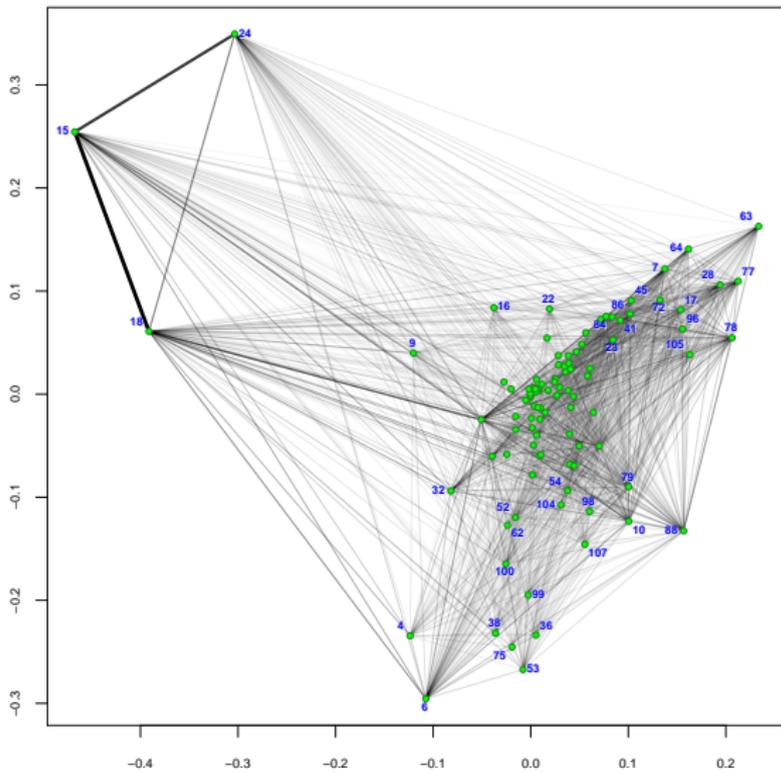
$$\text{cut}_\tau(A_1, \dots, A_K) = \frac{\text{cut}(S, G) + \frac{\tau}{N}|S||S^c|}{\text{vol}(S, G) + \tau|S|}$$

- ▶ avec $N = |V|$

1-2



2-3



- ▶ la première dimension oppose **le blob** aux autres pétition
- ▶ **la deuxième dimension** oppose d'une part :

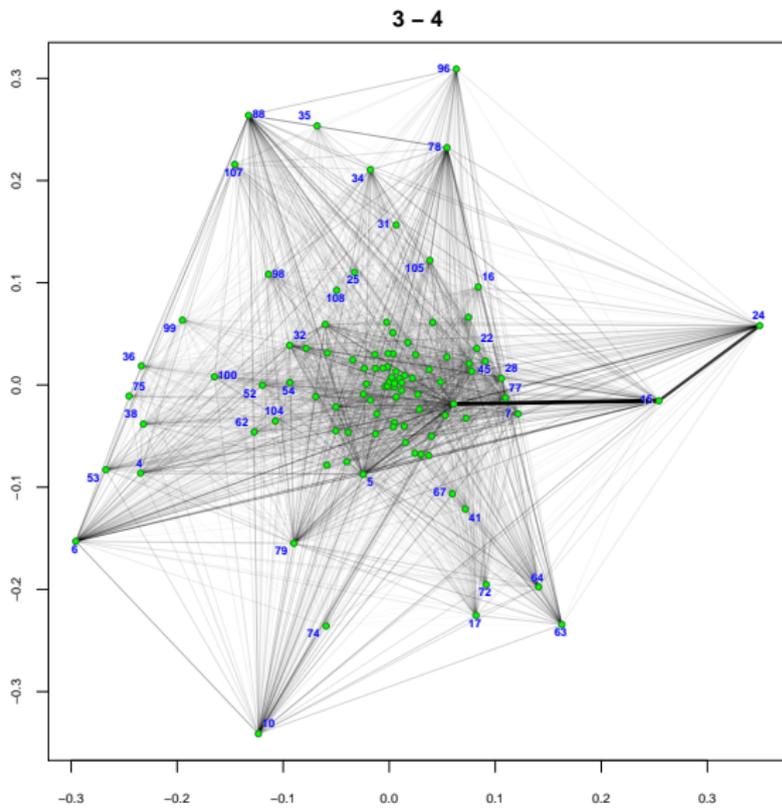
- ▶ les pétitions sur la fiscalité

avec une classe spécifique sur la TVA

- ▶ associées à la classe de pétitions relatives à la Belgique
- ▶ et, d'autre part, un ensemble de pétitions relatives aux

- ▶ conditions de travail, licenciement
- ▶ la Palestine
- ▶ la protection des espaces naturels
- ▶ les expulsions et le droit au logement
- ▶ la liberté d'expression dans les médias
- ▶ pollution de sites (exploitations minière, déchets, . . .
- ▶ les arrestations arbitraires et les violences policières

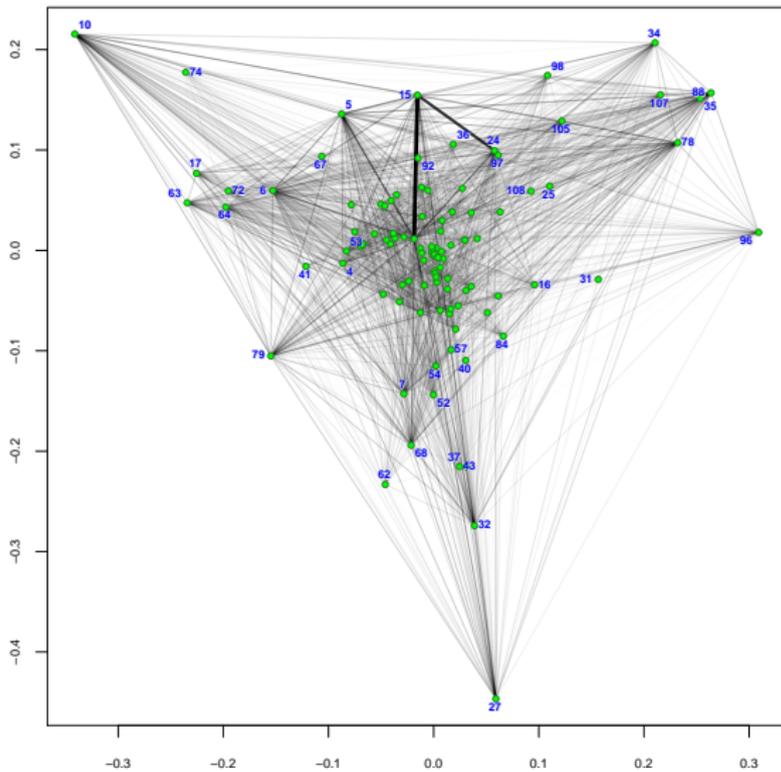
cette classes regroupe plus généralement les pétitions sur les questions de sécurité et de police



▶ **la troisième dimension** oppose

- ▶ les pétitions sur la fiscalité et le coût de la vie, d'une part
- ▶ et, d'autre part, un ensemble de pétitions relatives aux
 - ▶ aménagement urbains
 - ▶ protection des espaces naturels
 - ▶ les pistes cyclables
 - ▶ nuisances aériennes
- ▶ ainsi que des pétitions
 - ▶ lieux, quartiers
 - ▶ l'université
 - ▶ Bruxelles

4 - 5



▶ **la quatrième dimension** oppose

▶ des pétitions sur

- ▶ les expulsions et le droit d'asile
- ▶ régularisation de sans papier

▶ et des pétitions sur

- ▶ l'éolien
- ▶ l'énergie, particulièrement la production d'électricité
- ▶ l'urbanisation
- ▶ la protection d'espace naturels
- ▶ les OGM

▶ **la cinquième dimension** oppose

- ▶ les pétitions précédentes

avec les pétitions sur la TVA

- ▶ et des pétitions relatives aux religions

- ▶ principalement sur l'Islam
- ▶ adoptant des points de vue contrastés

« contre le port de la burka » (sic)
« oui pour le port de la burka »

- ▶ mais aussi sur la laïcité
- ▶ ou l'Église catholique

Conclusion

- ▶ difficulté à détacher des réseaux ou des communautés dans le graphe
- ▶ l'analyse thématique des signatures permet toutefois de faire ressortir différents **patrons** dans les signatures multiples
- ▶ mais elle est fortement **contrainte** par les limites de l'analyse du texte des pétitions
- ▶ différentes pistes sont toutefois envisageables pour améliorer **les performances** de la classification thématique
- ▶ **question (ouverte)** : l'analyse thématique peut-elle aider à détecter **les structures locales** du graphe ?

Merci pour votre attention

Bibliographie

- ALSUMAIT, Loulwah, Daniel BARBARÁ, James GENTLE et Carlotta DOMENICONI (2009), « Topic Significance Ranking of LDA Generative Models », *Machine Learning and Knowledge Discovery in Databases*. Sous la dir. de Wray BUNTINE, Marko GROBELNIK, Dunja MLADENIĆ et John SHAWE-TAYLOR, Berlin, Heidelberg, Springer Berlin Heidelberg, p. 67–82.
- BLEI, David M., Andrew Y. NG et Michael I. JORDAN (2003), « Latent Dirichlet Allocation », *J. Mach. Learn. Res.* 3, p. 993–1022.
- CONTAMIN, Jean-Gabriel., Thomas LÉONARD et Thomas SOUBIRAN (2017), « Les transformations des comportements politiques au prisme de l'e-pétitionnement. Potentialités et limites d'un dispositif d'étude pluridisciplinaire », *Réseaux*, 204, 4, p. 97–131.
- DYER, Justin (2010), "Visualizing and modeling joint behavior of categorical variables with a larg number of levels". Thèse de doct. Stanford University.
- QIN, Tai et Karl ROHE (2013), « Regularized Spectral Clustering under the Degree-Corrected Stochastic Blockmodel », *arXiv e-prints*. arXiv :1309.4111.
- SOUBIRAN, Thomas (2017), « La détection de réseaux sociaux à partir de leurs traces : l'exemple d'un site de pétitions en ligne », *Journée MADICS*. Big Data Mining and Visualisation. (Lille, 6 juin 2017),