

Perspectiva



Data 4 Good

Utiliser les outils LLM pour analyser et fournir un retour plus rapide et fidèle aux contributeurs d'une consultation publique.

Enjeux sur la communication des résultats

Porteur du projet : Cyril François

Louis Boulanger

Mathieu Gervais-Dubé

Jonathan Legrand

Alexandre Wendling

Cas d'usage pour la consultation citoyenne

Objectif :

Analyser et comprendre ce qu'un large groupe de personnes pense en s'appuyant sur leurs propres mots, grâce aux outils intelligents du traitement du langage.

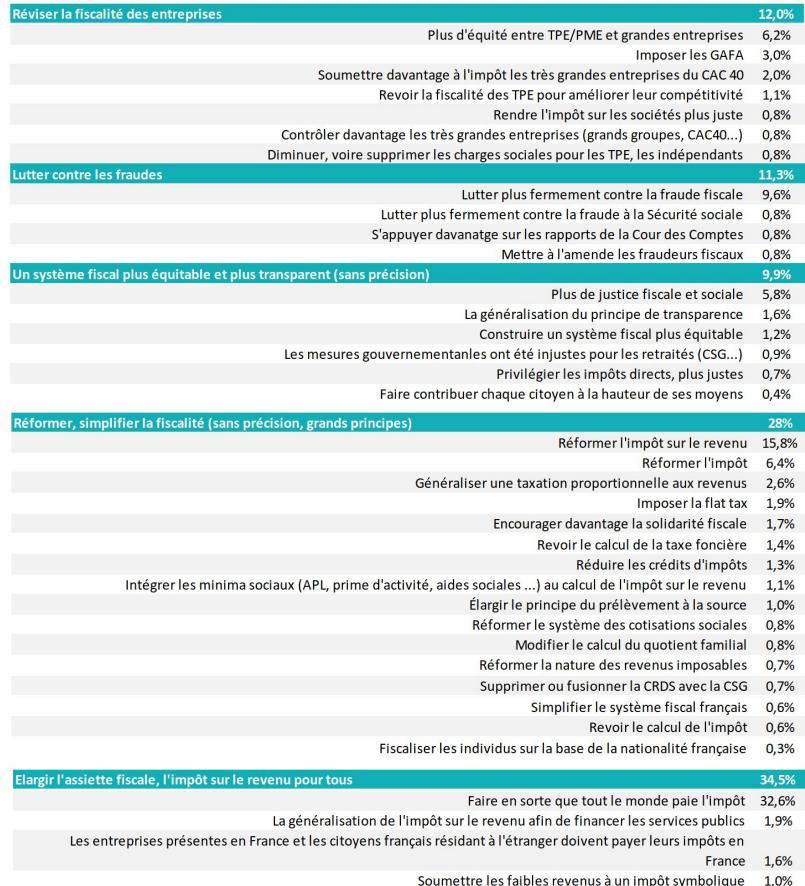
Exemple d'application : question 163 du GRAND DEBAT.

"Que faudrait-il faire pour rendre la fiscalité plus juste et plus efficace ?"

- 52 000 contributions répondant à une question ouverte
- Traité par une société privée Qwam (interview de Christian Langevin par Maxence Fabrion, journaliste chez Decode Media)
 - 2 mois de travail
 - 5 personnes à temps complet
 - plusieurs millions €
 - outils IA (2019, NER, Sentence détection, similarité sémantique)

Exemple de donnée et analyse actuelle

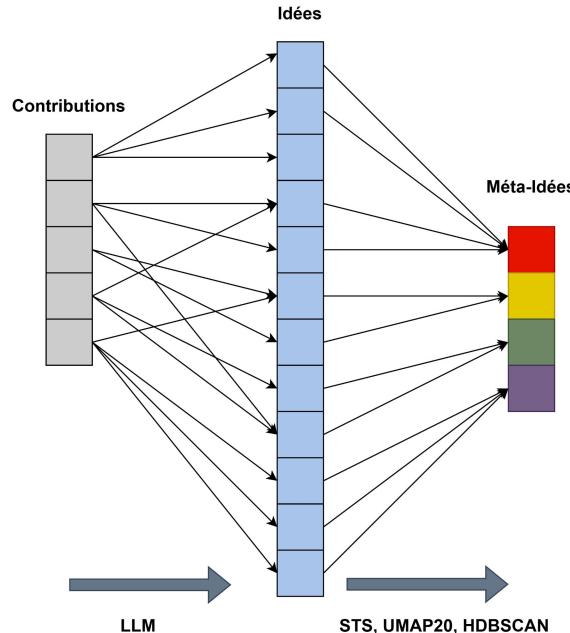
contribution_id	content
	int64
0 - 5.37k	10.1%
1 - 2.33k	99%
0	Repartir les richesses. suppression de la taxe d habitation pour tous les français et de la csg reindexé les retraites . Ne plus diviser les français (les patrons ont besoin des plus modestes et vice versa. Les patrons ne réussissent que si les plus modestes sont aussi avec eux) . Les français ne veulent plus de l assistanat mais veulent vivre dignement de leur salaire
1	Les droits soient automatiques, comme nos devoirs de payer les impôts
2	réduire drastiquement la fraude fiscale. Imposer les grands groupes (GAFA) qui ne le sont pas suffisamment Renforcer la taxe sur les transactions ...
3	diminuer le taux de prélevement pour les retraités percevant moins de 2550 euros
4	TOUT FRANÇAIS DEVRA PAYER L'IMPÔT QU'IL SOIT DOMICILIÉ OU NON EN FRANCE LES FRANÇAIS NE VOULANT PLUS PAYER L'IMPÔT EN FRANCE DEVONT CHANGER DE...
5	Rétablissement l'ISF sans délai avec incorporation dans son assiette de tous les éléments de fortune (objets d'art inclus + yachts etc) sans aucune autre...



Problématiques :

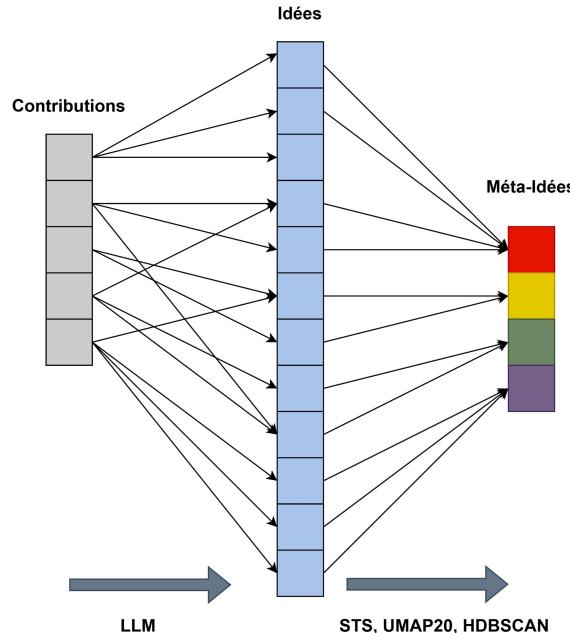
- Reproductibilité de l'étude
- Représentation de la synthèse du traitement
 - Schéma d'importance des idées
 - Association entre idée
- Création de persona
 - Profils types des contributeurs
 - Importance de leurs idées
 - Liens entre les persona

Pipeline proposé par Cyril



Tout le monde doit participer à l'effort. La lutte contre l'évasion fiscale doit être une priorité car il est facile de faire la chasse aux petits revenus qui tentent de contourner la législation alors que les plus riches bénéficient de tout un système qui leur est dédié. Ramener dans les comptes tout l'argent "évacué" solutionnerait beaucoup de problème de budget (santé, social, public, etc.) Il est également inadmissible que de grands chefs d'entreprises du CAC 40, représentant la France soit résident fiscal ailleurs qu'en France

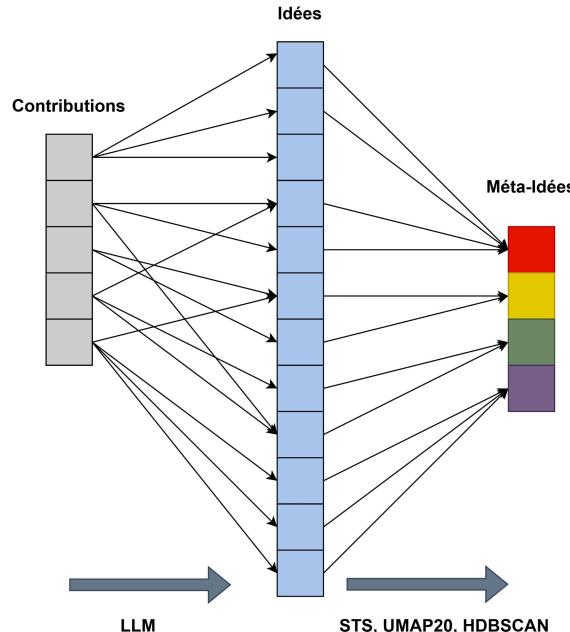
Pipeline proposé par Cyril



Tout le monde doit participer à l'effort. La lutte contre l'évasion fiscale doit être une priorité car il est facile de faire la chasse aux petits revenus qui tentent de contourner la législation alors que les plus riches bénéficient de tout un système qui leur est dédié. Ramener dans les comptes tout l'argent "évacué" solutionnerait beaucoup de problème de budget (santé, social, public, etc.) Il est également inadmissible que de grands chefs d'entreprises du CAC 40, représentant la France soit résident fiscal ailleurs qu'en France

- participation de tous à l'effort
- lutte contre l'évasion fiscale
- inadmissible que de grands chefs d'entreprises du CAC 40, représentant la France soit résident fiscal ailleurs qu'en France
- ramener dans les comptes tout l'argent "évacué"

Pipeline proposé par Cyril



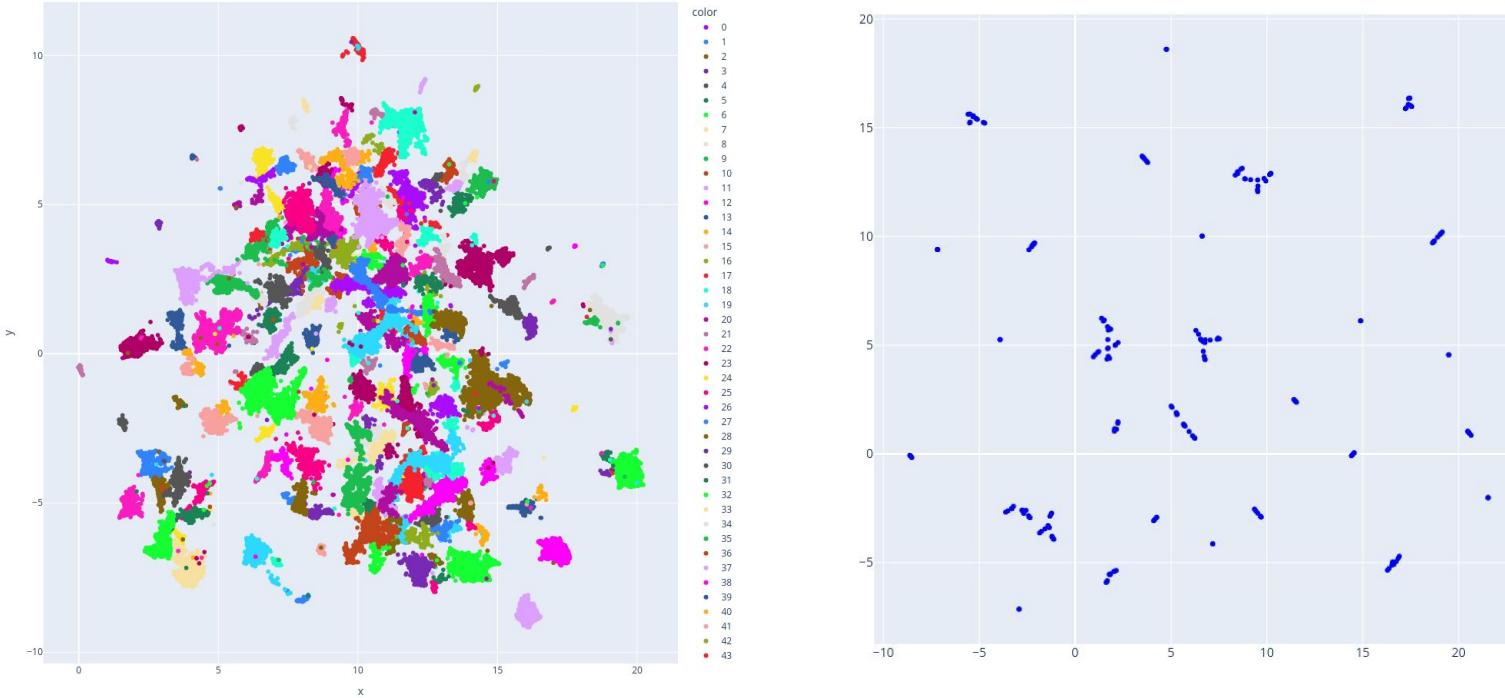
Tout le monde doit participer à l'effort. La lutte contre l'évasion fiscale doit être une priorité car il est facile de faire la chasse aux petits revenus qui tentent de contourner la législation alors que les plus riches bénéficient de tout un système qui leur est dédié. Ramener dans les comptes tout l'argent "évadé" solutionnerait beaucoup de problème de budget (santé, social, public, etc.) Il est également inadmissible que de grands chefs d'entreprises du CAC 40, représentant la France soit résident fiscal ailleurs qu'en France

- participation de tous à l'effort
- lutte contre l'évasion fiscale
- inadmissible que de grands chefs d'entreprises du CAC 40, représentant la France soit résident fiscal ailleurs qu'en France
- ramener dans les comptes tout l'argent "évadé"

- L'appel à la participation citoyenne de 5%
- Lutter contre l'évasion fiscale
- Réformer la fiscalité des entreprises en France

Analyse du jeu de donnée

Visualisation actuelle et analyse des méta-idée



- title:Meilleure répartition plus équitable.
- title:Redistribution des richesses plus juste.
- title:Obligation de partager les bénéfices.

- title:Répartition plus équitable de l'impôt.

- title:Impôt proportionnel: rendre l'imposition plus équitable.

- title:Égalité fiscale: la réforme de l'impôt pour une société plus juste.

- title:Taxe sur les luxes de transport.

- title:Taxe exponentielle sur les produits de luxe.

- title:Taxe sur les produits de luxe et les œuvres d'art.

- title:Surtaxe sur tabac et alcool, TVA sur luxe et risques.

- title:Supprimer la redérence audiovisuelle.
- title:Suppression des avantages fiscaux des journalistes.

- title:Suppression des niches fiscales.

- title:Supprimer les niches fiscales pour les riches.

- title:Efficacité

- title:Renforcement

- title:Renforcement d
- title:Compétitivité française: lutter contre l'obésité et la distorsion de concurrence.

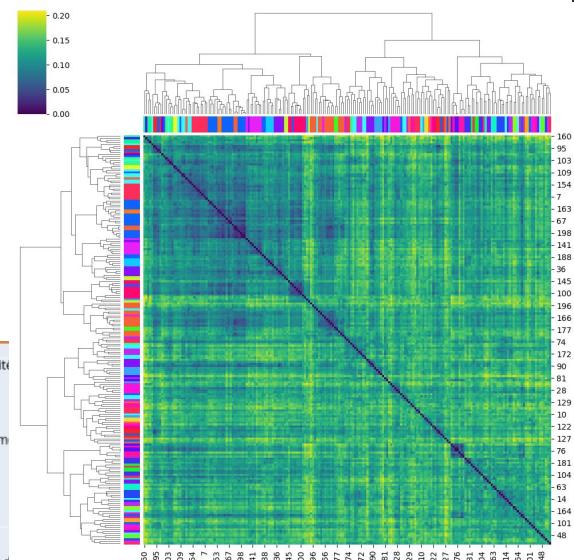
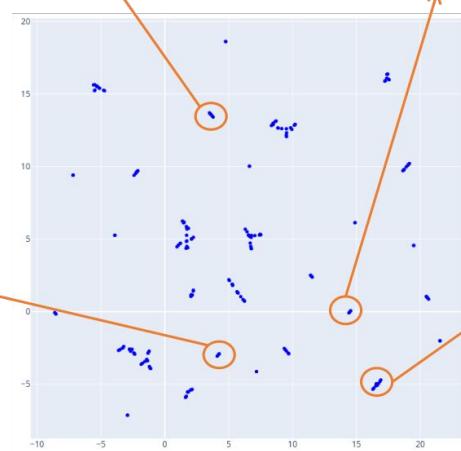
- title:Lutter contre le dumping fiscal et social en Europe.

- title:Lutter contre la fraude fiscale et sociale.

- title:Lutter contre les paradis fiscaux et la fraude fiscale.

- title:Lutter contre la fraude sociale et les abus.

- title:Lutter contre la fraude et améliorer le contrôle.





Défis 1 : représentation de la synthèse du traitement

Cooccurrence des métas-idées

		métas-idée <i>i</i>		Total
métas-idée <i>j</i>	Présence	Absence		
Présence	nb cooccurrence (\mathbf{n}_{ij})	nb <i>j</i> seul		total <i>j</i> (\mathbf{n}_j)
Absence	nb <i>i</i> seul	nb contributeur sans <i>i</i> et <i>j</i>		total contributeur sans <i>j</i>
Total	total <i>i</i> (\mathbf{n}_i)	total contributeur sans <i>i</i>		total contributeur (\mathbf{N})

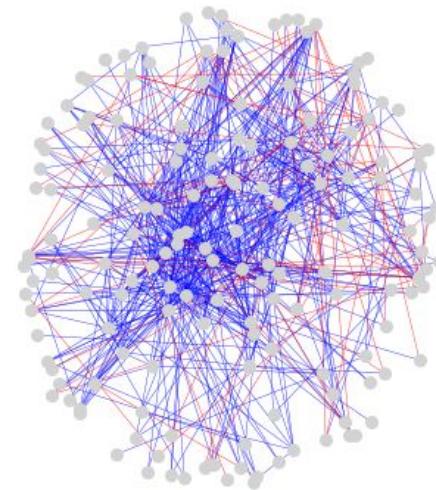
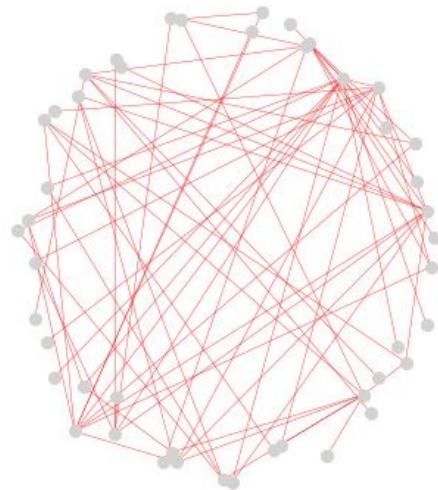
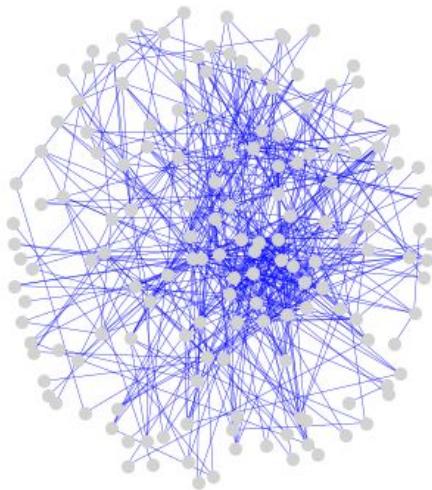
H_0 : les métas-idées *i* et *j* sont évoqués indépendamment par les contributeurs

H_1 : les métas-idées *i* et *j* sont associés

$$X \sim \mathcal{HG}(\mathbf{N}, \mathbf{n}_i, \mathbf{n}_j) \quad p^+ = P(X \geq \mathbf{n}_{ij}) \quad p^- = P(X \leq \mathbf{n}_{ij})$$

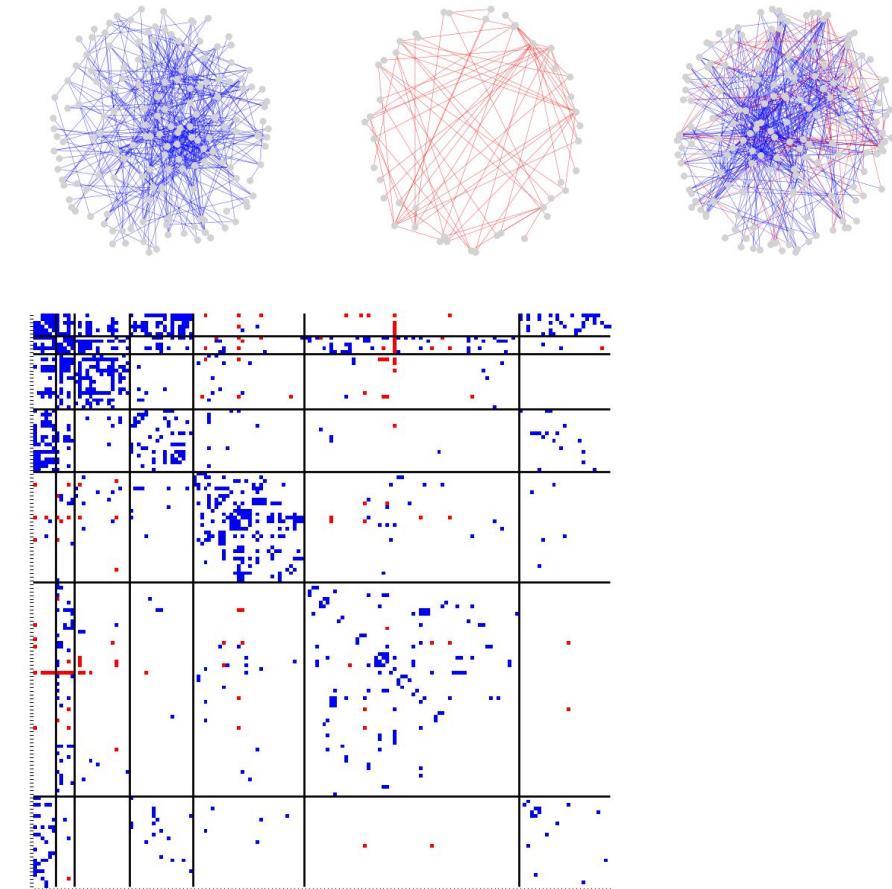
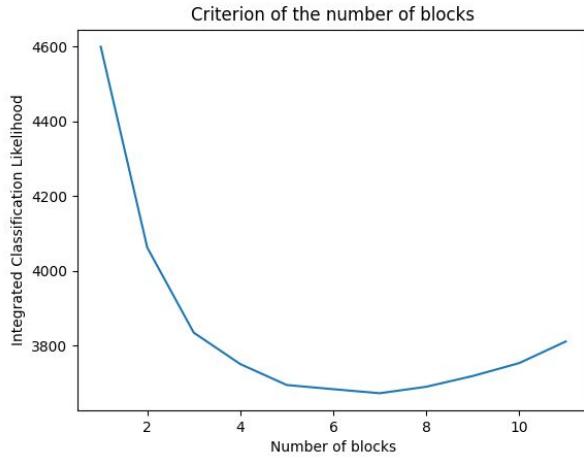
Test unilatéral de Fisher + Correction de bonferroni

Graphe d'association

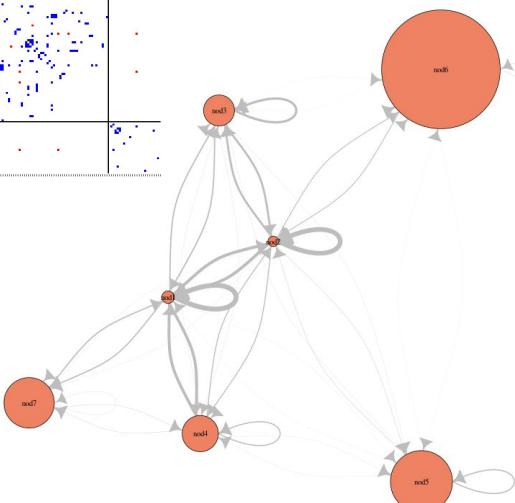
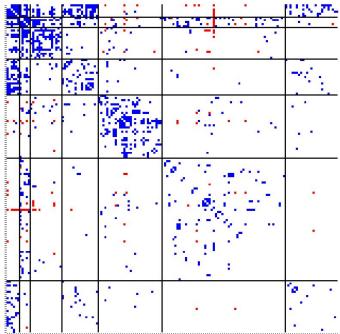


Graphe d'association

Node clustering : Stochastic block model



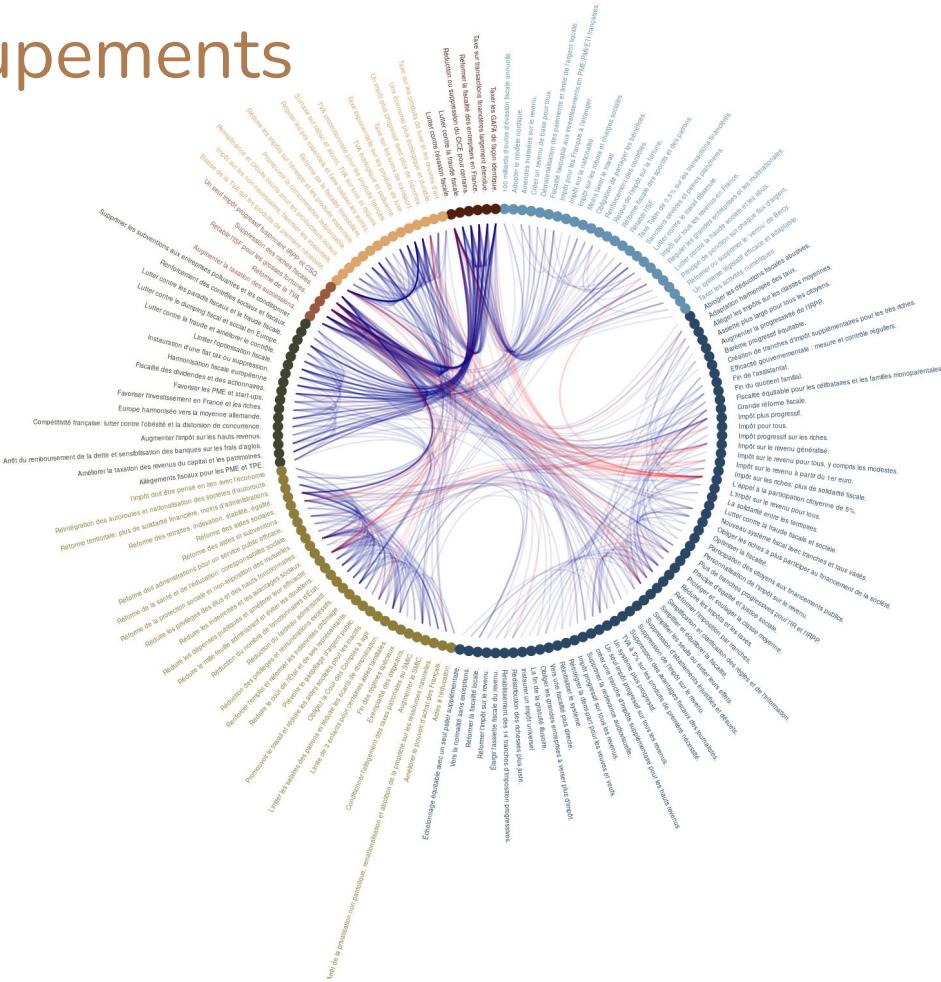
Cluster de métta-idée



Résumé (objectif principal)

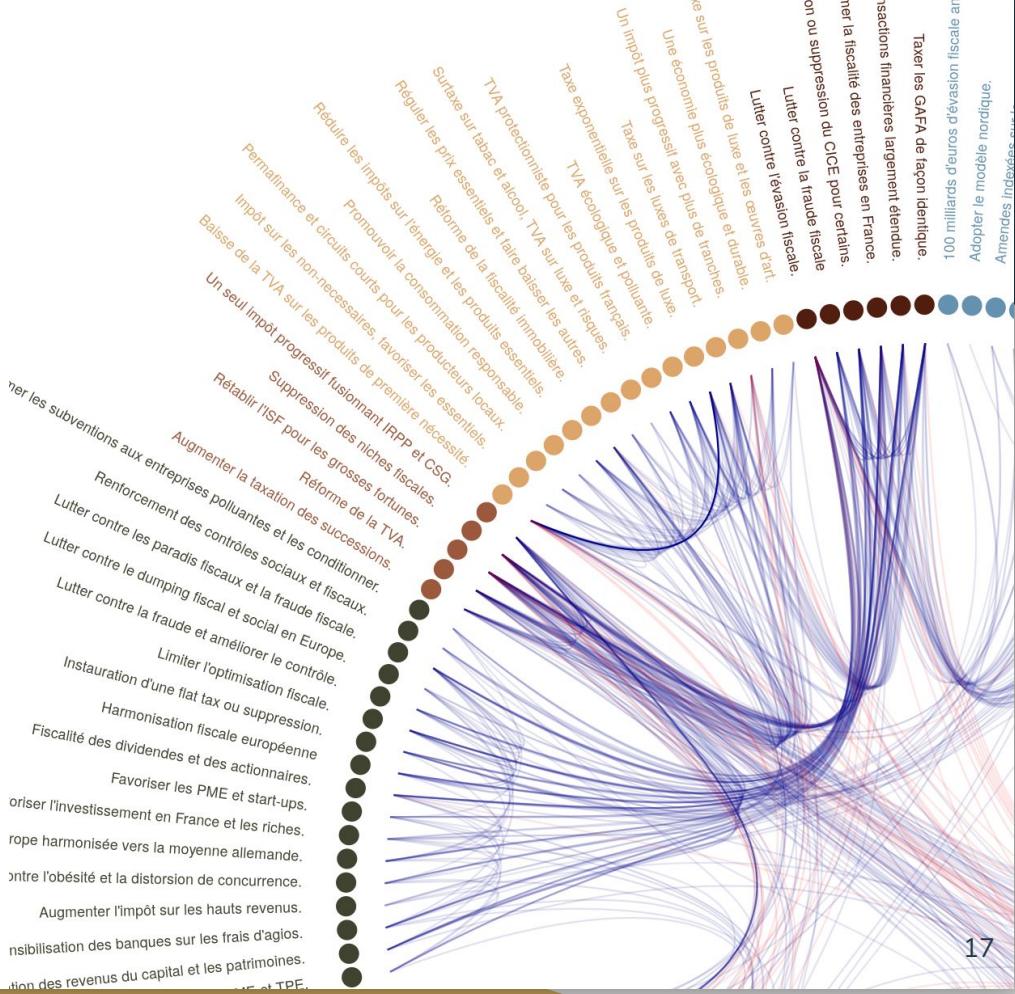
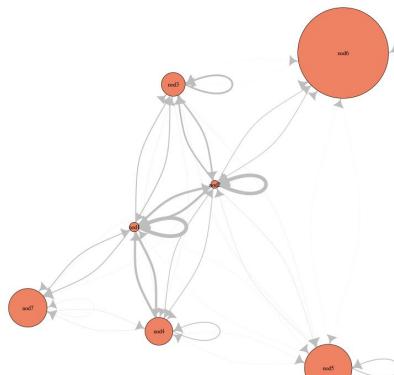
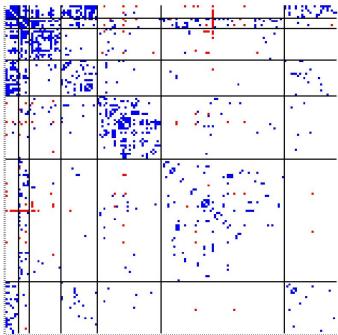
- 1 Lutter contre la fraude et l'évasion fiscales des grandes entreprises
Évasion fiscale, fraude fiscale, réforme fiscalité des entreprises, taxation du numérique
- 2 Réduire les niches et augmenter la fiscalité sur les plus riches
Suppression des niches, rétablissement de l'ISF, réforme TVA, impôt progressif unique
- 3 Favoriser une fiscalité écologique et sociale
Plus de progressivité IR, baisse TVA produits essentiels, taxe luxe, fiscalité écologique
- 4 Contrôler mieux et harmoniser à l'échelle européenne
Contrôle fiscal, optimisation, fiscalité européenne, taxation des dividendes et du capital
- 5 Mieux gérer l'argent public et réduire les priviléges
Réduction dépenses publiques, réforme retraites/aides sociales, limitation privilégiés
- 6 Rendre l'impôt plus simple, lisible et équitable
Impôt progressif, nouvelles tranches IR, impôt universel, justice fiscale pour les classes moyennes
- 7 Étendre l'impôt à tous et renforcer les contrôles, notamment des riches
Taxation des expatriés, revenus des très riches, sanctions, lutte contre travail dissimulé

Visualisation des regroupements de méta-idées et leurs liens



Regroupements de métas- idées et leurs liens

Création d'une représentation des idées et de leur liens via les **graphes d'association** et les **modèles à bloc stochastiques**.



Comparaison avec l'analyse de Qwam

Analyse Qwam

- Elargir l'assiette fiscale, l'impôt sur le revenu pour tous
- Réformer, simplifier la fiscalité (sans précision, grands principes)
- Réviser la fiscalité vers plus de progressivité, taxer davantage les hauts revenus
- Baisser la fiscalité
- Réviser la fiscalité des entreprises
- Lutter contre les fraudes
- Un système fiscal plus équitable et plus transparent (sans précision)
- Mieux gérer l'argent public
- Taxer ou créer de nouvelles taxes
- Mentions générales sur le pouvoir d'achat
- Revoir la fiscalité au niveau local (plus de moyens, plus de justice)

Notre analyse

- Lutter contre la fraude et l'évasion fiscales des grandes entreprises
- Réduire les niches et augmenter la fiscalité sur les plus riches
- Favoriser une fiscalité écologique et sociale
- Contrôler mieux et harmoniser à l'échelle européenne
- Mieux gérer l'argent public et réduire les priviléges
- Rendre l'impôt plus simple, lisible et équitable
- Étendre l'impôt à tous et renforcer les contrôles, notamment des riches

Défi 2 : création de persona

p_i



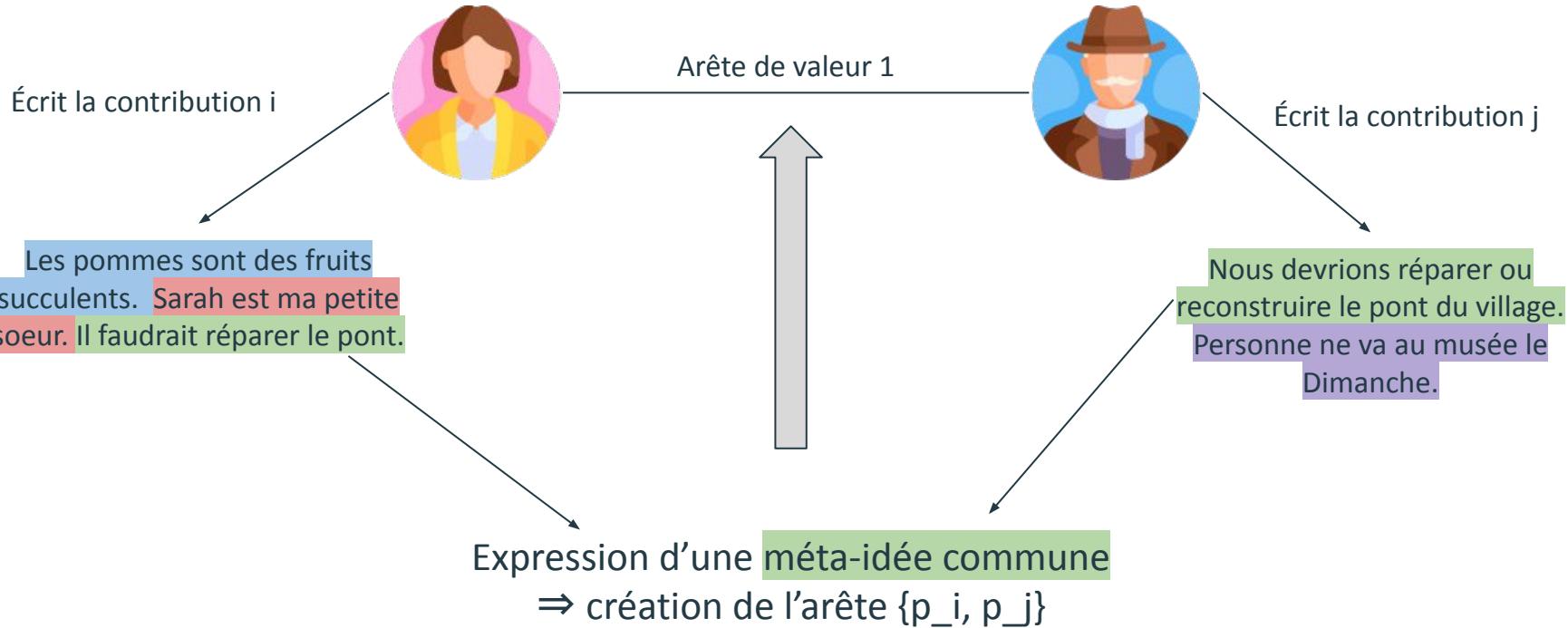
p_j



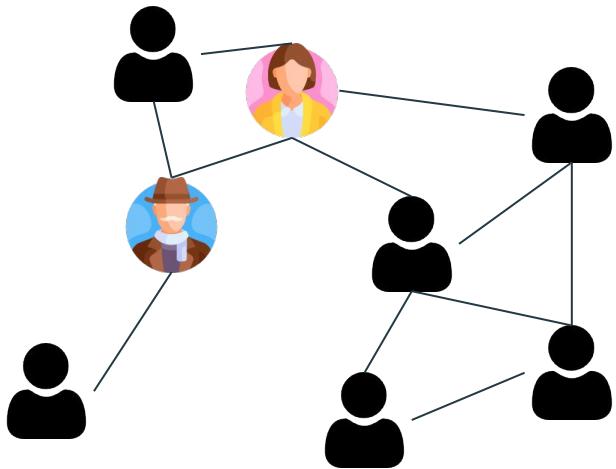
Étape 1 : Création d'un graphe des contributions



Étape 1 : Création d'un graphe des contributions

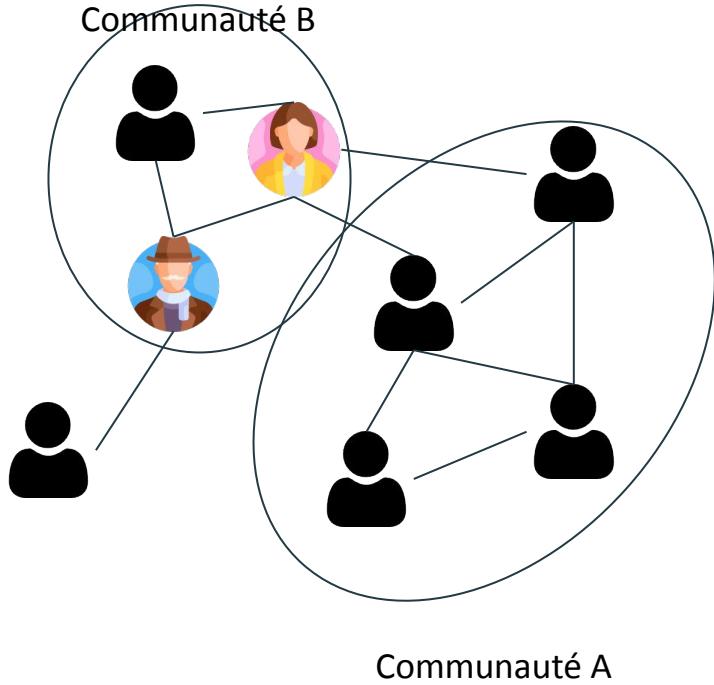


Étape 1 : Création d'un graphe des contributions



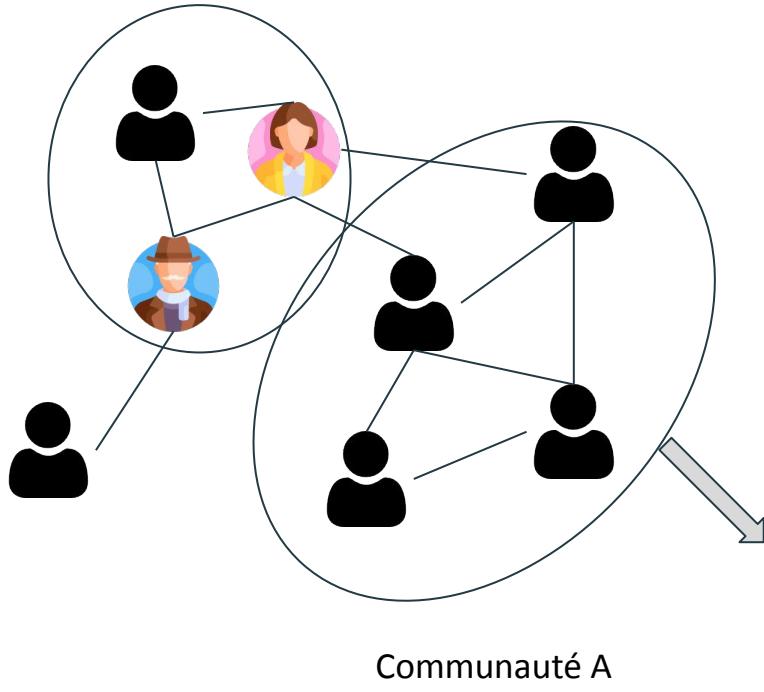
~50000 noeuds, ~ 10^7 arêtes : visualisation directe impossible

Recherche de communautés dans le graphe : Algorithme de Louvain



~50000 noeuds, ~ 10^7 arêtes : visualisation directe impossible

Algorithme de Louvain : Trouver un nombre réduit de communautés de contributeurs fortement connectés

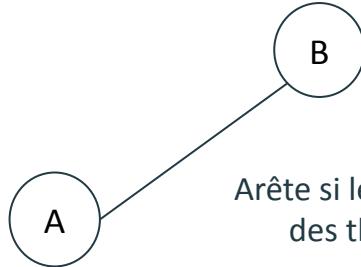
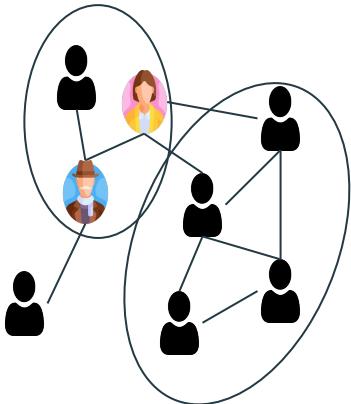


~ 50000 noeuds, $\sim 10^7$ arêtes : visualisation directe impossible

Algorithme de Louvain : Trouver un nombre réduit de communautés de contributeurs fortement connectés

$$[F_A^A, \dots, F_A^{103}]$$

F_A^i : Fréquence de la métta-idée i dans la communauté A.



Arrête si les communautés ont
des thèmes communs

Épaisseur de l'arête

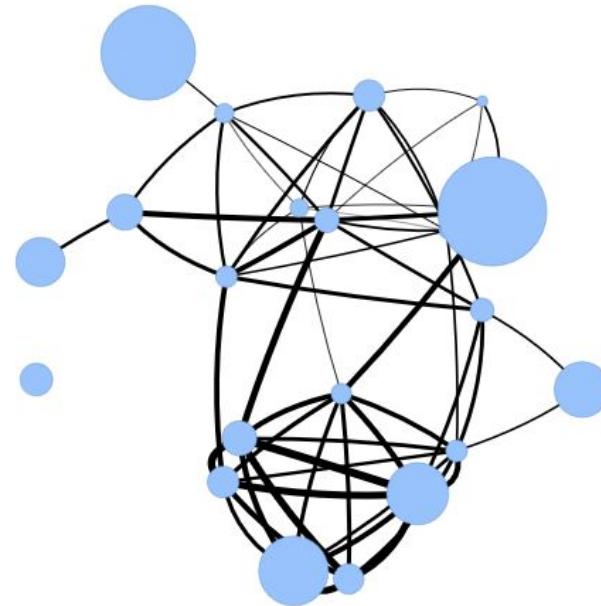
$$s = \frac{1}{1 + dist(F_A, F_B)}$$

Construction du graphe des communautés d'utilisateurs

Construction du graphe : networkx
Rendu visuel : Pyviz



Graphe interactif



Limites et perspectives

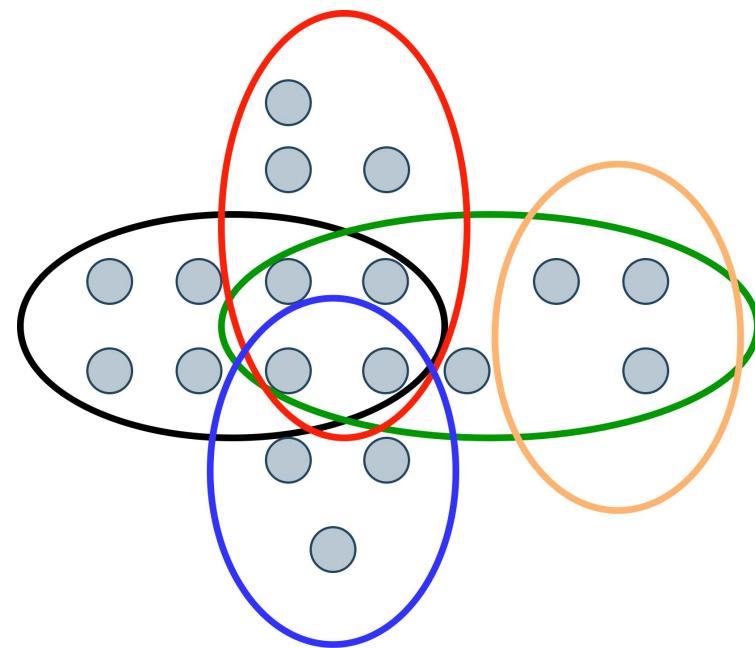
- Le calcul de Louvain nécessite un sous échantillonnage des réponses
- Nommer les communautés avec un modèle de langage (cf annexes)
- Permettre à l'utilisateur d'explorer les différentes hiérarchies générées par l'algorithme de Louvain.
- Les métas-idendées se recoupent trop, utiliser moins de clusters

Tentative

Problème des personas sous forme de programme linéaire.

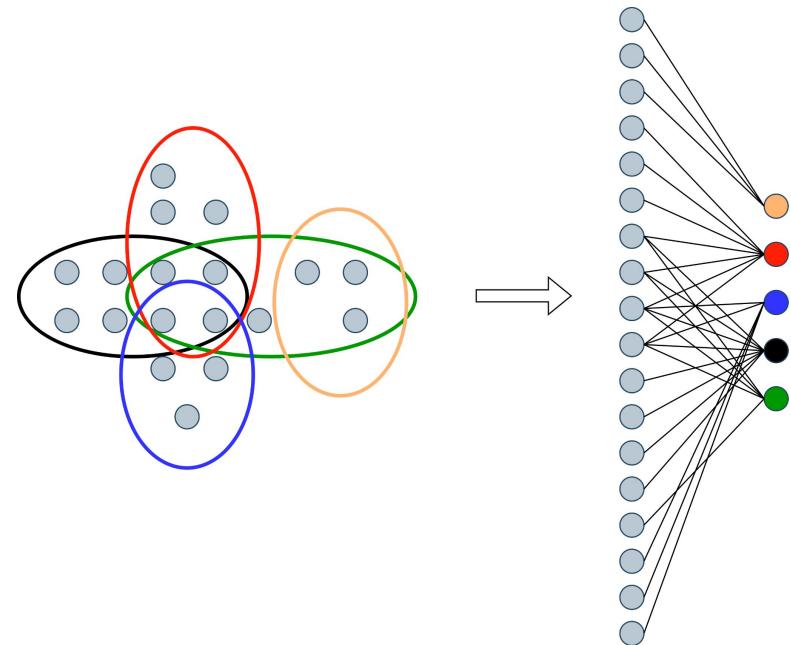
Hypergraphe contributeurs métidoées

- Peut-on agglomérer les contributeurs pour former des personas?
- Hypergraphe où les sommets sont des contributeurs et les arêtes les clusters d'idées.
- **Hypothèse:** Les personas doivent posséder des métidoées en commun.



Formulation

- Les sommets sont reliés entre eux ssi la contribution c est associée à une méta-idée m.
- On énumère les bicliques maximales pour former des personas candidats.
- **Propriété:** Les contributeurs expriment un nombre limité d'idée.
- **Conséquence:** le nombre de bicliques maximales dans H est limité.
- 59 605 bicliques trouvées sur 53 724 contributions et 203 méta-idées.



Formulation

1. $N \in \mathbb{N}$ est le nombre de personas désiré.
2. M est l'ensemble des méta-idées.
3. C est l'ensemble des contributeurs.
4. G est un hypergraphe.
5. $\mathcal{B}(G)$ est l'ensemble des bicliques maximales dans le graphe biparti H dont les sommets sont formés par les méta-idées M à gauche et les contributeurs C à droite ainsi que des arêtes présentes entre $c \in C$ et $m \in M$ si le contributeur c possède une idée faisant partie de la méta-idée m . Ces bicliques forment des personas potentiels.

Formulation

6. Soit $b = (c, m) \in \mathcal{B}(G)$ où $c \subseteq C$ et $m \subseteq M$. On définit $n_1(b) := |c|$ et $n_2(b) := |m|$.
7. $\alpha(b) : \mathcal{B}(G) \rightarrow \mathbb{R}^+$ est une fonction croissante selon le nombre de cluster dans b . On pose

$$\alpha(b) := n_2(b).$$

8. $c(b) : \mathcal{B}(G) \rightarrow \mathbb{R}^+$ est un bonus positif associé au choix du persona b . On pose

$$c(b) := \alpha(b)n_1(b) = n_2(b)n_1(b) \quad \forall b \in \mathcal{B}(G)$$

9.

$$x_b := \begin{cases} 1, & \text{si le groupe } b \text{ forme un persona et} \\ & \\ 0 & \text{sinon.} \end{cases}$$

Formulation

$$\max_{\substack{x \in \{0,1\}^{|\mathcal{B}(G)|} \\ y \in \mathbb{R}}} \sum_{b \in \mathcal{B}(G)} c_b x_b - y,$$

sujet à

$$\sum_{b \in \mathcal{B}^c} x_b \leq 1, \quad \forall c \in C$$

Chaque contributeur est assigné à au plus un persona.

$$\sum_{b \in \mathcal{B}} x_b = N$$

Il y a exactement N persona.

$$\sum_{b \in \mathcal{B}} n_1(b) x_b + y \geq 0.85|C|$$

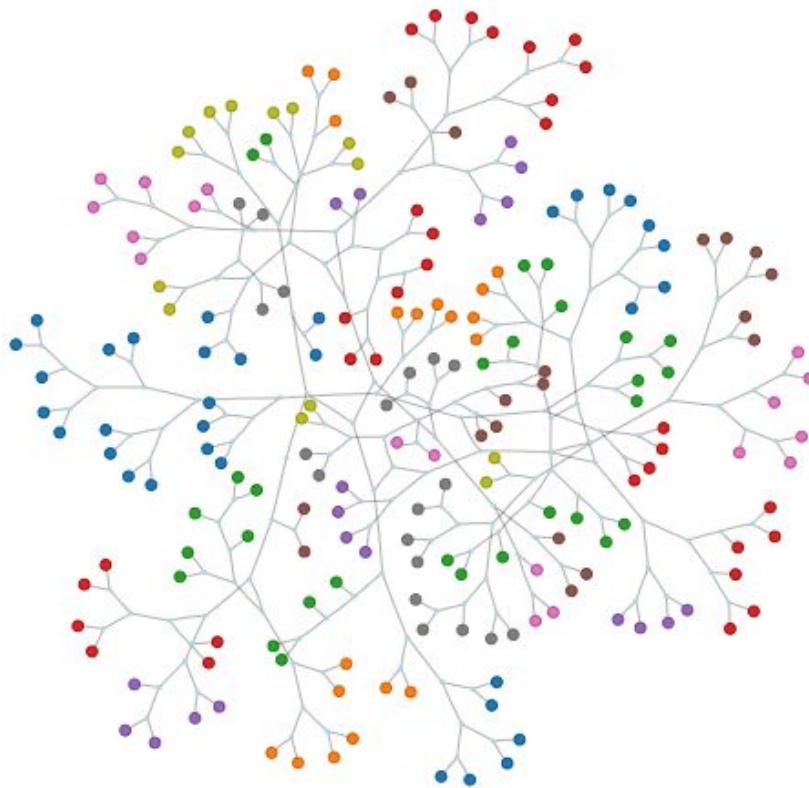
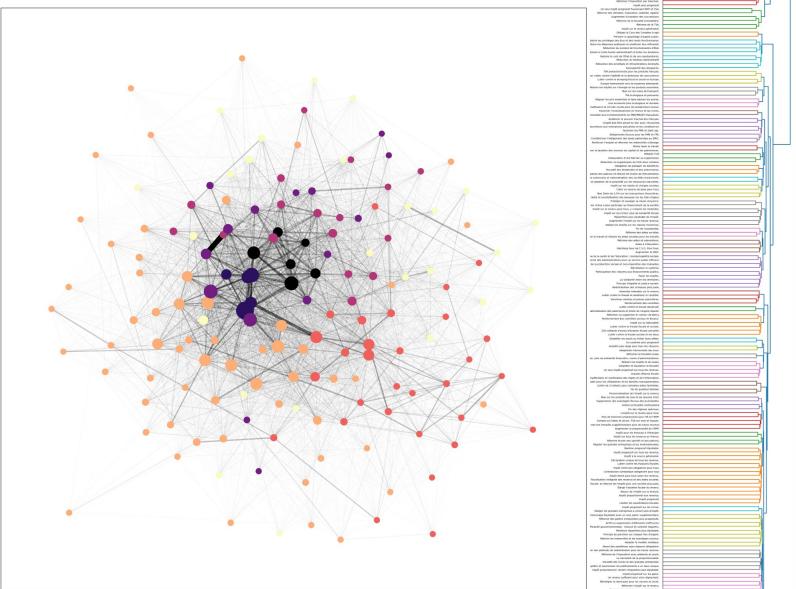
Pénalité si moins de 85% sont associés à des personas.

Résultats

- Modélisation à l'aide de Python-MIP.
- Presque partition en bicliques maximales est une condition trop forte.
- Il n'y a pas de presque partition de C en un petit nombre de bicliques maximales.
- 6 288 contributeurs choisis avec $N = 11$.
- 20 961 contributeurs choisis avec $N = 60$.
- Certains paquets de contributeurs intéressants ne sont pas captés par le modèle.
- Ceci justifie les approches statistiques employées.

Conclusion

Annexes



ID	Nom du Persona	Mesure phare
1	Le Progressiste Fiscal	Créer plus de tranches d'imposition pour rendre l'impôt plus progressif
2	L'Égalitariste Intransigeant	Imposer le revenu dès le 1er euro
3	Le Contribuable Universel	Instaurer un impôt pour tous
4	Le Réformateur de la Consommation	Réformer la TVA et baisser celle sur les produits de première nécessité
5	Le Rationalisateur Fiscal	Simplifier et équilibrer le système fiscal
6	Le Solidaire Volontaire	Appeler à une plus forte contribution des plus riches
7	Le Démantèle-Niches	Supprimer les niches fiscales
8	Le Contrôleur Éthique	Lutter contre la fraude fiscale et améliorer les contrôles
9	Le Minimaliste Fiscal	Réduire les impôts et les taxes
10	Le Réformiste Immobilier	Réformer la fiscalité immobilière
11	Le Capitaliste	Réduire les barrières réglementaires et fiscales pour les entreprises