Évaluer la polarisation des Réseaux Sociaux Numériques par l'analyse des frontières de ses communautés : l'approche ERIS

Assessing polarization in Online Social Networks through community boundaries : the ERIS approach

Alexis Guyot, Annabelle Gillet, Éric Leclercq, Nadine Cullot

Laboratoire d'Informatique de Bourgogne - EA 7534, Université de Bourgogne, Dijon, France
{alexis.guyot, annabelle.gillet, eric.leclercq, nadine.cullot}@u-bourgogne.fr

RÉSUMÉ. La détection et la caractérisation de la polarisation d'un réseau sont des problématiques majeures en analyse des réseaux sociaux. Par ce biais, les sujets conflictuels qui animent les interactions entre les utilisateurs peuvent être mis en évidence et mieux compris. En tant qu'intermédiaires avec l'extérieur de leur communauté, les individus situés aux frontières contribuent de manière significative à sa polarisation. Nous proposons ERIS, une approche formelle basée sur les graphes, qui exploite les frontières des communautés et les interactions entre les individus pour évaluer deux indicateurs : l'antagonisme des communautés et la porosité de leurs frontières. Ces valeurs représentent respectivement le degré d'opposition entre les communautés et leur tendance à s'exposer hors de la communauté. Conjointement, elles décrivent les comportements et rôles des différentes communautés et permettent ainsi de mieux comprendre la polarisation du réseau. Nous présentons également un algorithme pour calculer ces indicateurs, basé sur des opérations matricielles et dont le code source est librement accessible en ligne. Nous proposons une comparaison de notre méthode par rapport aux solutions existantes. Pour finir, nous appliquons notre proposition sur des données réelles, collectées à partir de Twitter, au travers d'une étude de cas portant sur les vaccins et la COVID-19.

ABSTRACT. Detection and characterization of polarization are of major interest in Social Network Analysis, especially to identify conflictual topics that animate the interactions between users. As gatekeepers of their community, users in the boundaries significantly contribute to its polarization. We propose ERIS, a formal graph approach relying on community boundaries and users' interactions to compute two metrics : the community antagonism and the porosity of boundaries. These values assess the degree of opposition between communities and their aversion to external exposure, allowing an understanding of the overall polarization through the behaviors of the different communities. We also present an implementation based on matrix computations, freely available online. We compare our method with existing solutions. Finally, we apply our proposal on real data harvested from Twitter with a case study about the vaccines and the COVID-19.

MOTS-CLÉS. Réseaux Sociaux, Polarisation, Frontière de communauté, Structure de communauté, Fouille de graphe

KEYWORDS. Social Networks, Polarization, Community Boundaries, Community Structure, Graph Mining

1. Introduction

Les Réseaux Sociaux Numériques (RSN) sont des outils de communication, d'échanges et de débats et les données qu'ils produisent peuvent être utilisées dans de nombreux domaines tels que le marketing, la politique ou la sociologie. L'analyse algorithmique des données des réseaux sociaux a pour objectif d'extraire de la valeur de ces données massives et propose des outils pour comprendre les interactions entre les individus et les groupes (Barabási & Pósfai 2016).

L'analyse des données des réseaux sociaux exploite souvent la théorie des graphes. Lorsque c'est le cas, les individus sont représentés par des sommets et leurs interactions par des arêtes. Les communautés d'individus apparaissent alors en tant que zones denses en arêtes et peuvent ainsi être identifiées par des algorithmes comme Louvain, Walktrap ou Informap dans le cas des communautés non-recouvrantes, et SLPA, OSLOM ou Game dans le cas des communautés recouvrantes (Fortunato 2010).

Les discussions autour de sujets sensibles peuvent mener à la création de communautés mutuellement antagonistes, au sein desquelles peu d'individus restent neutres ou dans une position intermédiaire. En sciences sociales, ce phénomène est qualifié de polarisation (Isenberg 1986). Son étude présente l'opportunité, entre autres choses, d'identifier les sujets délicats et de pouvoir adapter en conséquence une stratégie marketing, ou bien encore de lutter contre la diffusion de désinformation (Baumann et al. 2020). Dans la littérature, les chambres d'écho sont généralement considérées comme des conséquences de la polarisation (Cinelli et al. 2021). Seulement, montrer qu'une communauté est une chambre d'écho ne suffit pas pour déterminer si elle est polarisée. En effet, une chambre d'écho est une configuration dans laquelle un individu est seulement exposé à des idées ou opinions qui sont en accord avec les siennes (Garimella, De Francisci Morales, Gionis & Mathioudakis 2018). Ce concept décrit alors un comportement interne partagé par les membres d'une seule communauté, là où la polarisation s'intéresse aux relations entre les communautés (Baumann et al. 2020).

Par conséquent, la polarisation est un phénomène grandement influencé par les individus exposés aux opinions des autres communautés et qui, par le biais de leurs interactions, exposent eux-mêmes l'opinion de leur communauté aux autres. Ces membres forment les frontières de leur communauté. De manière plus formelle, on peut définir une frontière d'une communauté comme l'ensemble des sommets qui possèdent des arêtes dirigées à la fois vers l'intérieur et vers l'extérieur de cette communauté (Clauset 2005). Dans la littérature, les frontières sont des zones peu explorées. Néanmoins, les comportements des individus qui composent les frontières ont un impact significatif sur l'intensité de la polarisation de la communauté, mais également sur la fragilité de la chambre d'écho formée (Donkers & Ziegler 2021) puisqu'ils agissent sur la porosité de ses frontières.

Les principales contributions de notre travail sont : 1) une approche formelle, basée sur la théorie des graphes, qui exploite les frontières des communautés pour détecter des signes de polarisation au sein de réseaux sociaux ; 2) deux indicateurs pour caractériser le niveau d'antagonisme et la porosité des frontières des communautés ; 3) un algorithme basé sur des opérations matricielles, applicable sur de larges volumes de données ; 4) une étude de cas menée sur des données réelles en provenance de Twitter pour illustrer expérimentalement la validité de notre proposition.

Le reste de l'article est structuré de la manière suivante. D'abord, notre méthode est positionnée par rapport aux travaux connexes dans la section 2. Dans la section 3, nous définissons formellement l'approche ERIS et nous proposons un algorithme de faible complexité pour calculer les indicateurs de polarisation. L'étude de cas menée sur les données réelles et validée par des experts du domaine est présentée dans la section 4. Enfin, nous tirons des conclusions de notre travail et ouvrons de nouvelles perspectives pour le futur dans la section 5.

2. Travaux connexes

La problématique de la polarisation au sein des RSN a été abordée dès 2011 (Conover et al. 2011). Les auteurs considèrent alors les chambres d'écho et les communautés polarisées comme des concepts équivalents. Par conséquent, les interactions et relations entre communautés, portées par leurs frontières, sont ignorées. L'approche proposée est une méthodologie appliquée, qui nécessite l'intervention d'experts des données.

De nombreux travaux sur la polarisation sociale utilisent des approches supervisées combinant des mesures de la théorie des graphes et des interprétations produites par des outils de Traitement du Langage Naturel (TLN), tels que l'analyse de sentiments par classification naïve bayésienne (Alamsyah & Adityawarman 2017, Habibi & Sunjana 2019) ou encore l'extraction de modèles par plongement de phrases avec Retweet-BERT (Jiang et al. 2021). Un réseau de neurones profond est également utilisé dans (Mønsted & Lehmann 2022) pour classifier comme pro ou comme anti-vaccins les membres d'un réseau d'interactions extrait de Twitter. Les auteurs utilisent ensuite ces résultats pour établir une typologie des médias partagés par les différents groupes d'utilisateurs et pour mesurer le niveau de renfermement de ces groupes. Ces différentes approches parviennent à capturer précisément la sémantique des discussions mais nécessitent un investissement important de la part d'un analyste des données, notamment lors de la phase de pré-traitement. En effet, de nombreuses difficultés doivent être traitées manuellement pour permettre une exécution correcte des algorithmes de TLN, telles que la gestion des approximations orthographiques, des abréviations, de l'argot ou encore des ambiguïtés provoquées par l'utilisation d'humour, de sarcasme ou d'ironie, comme discuté dans (González-Ibánez et al. 2011, Joshi et al. 2017, McGlone 2005). Cet investissement important peut mener à la prise de décisions subjectives qui biaisent les analyses et rend difficile une utilisation complètement automatique de ces méthodes.

D'autres approches se concentrent uniquement sur la structure du réseau en appliquant des stratégies faiblement supervisées, dans lesquelles seule une quantité limitée de connaissances *a priori* est nécessaire pour initialiser les algorithmes. Dans (Morales et al. 2015), des scores d'opinion sont manuellement assignés à un ensemble d'individus graines (*elites*) puis propagés aux autres sommets du réseau (*listeners*) afin de créer deux groupes opposés et d'évaluer leur degré de polarisation. Dans (Al Amin et al. 2017), une mesure de similarité doit correctement être choisie pour créer des groupes de tweets (*assertions*), exploités par la suite pour révéler les communautés d'utilisateurs polarisées. La méthode proposée utilise une factorisation de matrices et un algorithme de descente de gradient ensembliste, appliqués sur la matrice d'adjacence d'un graphe biparti source-assertion et sur celle d'un graphe d'influence sociale. Une approche structurelle non-supervisée basée sur certaines mesures de la théorie des graphes comme la densité et l'indice I/E est proposée dans (Ertan et al. 2022) pour évaluer la polarisation dans un système multi-partis. Les mesures sont faites sur un graphe particulier qualifié de "réseau cognitif politique", qui met en relation les réponses collectées par les auteurs dans le cadre d'un sondage sur les dernières élections présidentielles en Turquie. Dans ces trois dernières approches, la pertinence des résultats dépend donc de connaissances *a priori* sur les données et sur leur traitement, qui doivent être repensées à chaque nouveau jeu de données étudié. Ainsi, la possibilité de les automatiser pour les appliquer à d'autres cas d'utilisation et/ou à plus grande échelle est limitée. De plus, elles ne considèrent pas non plus les relations entre les communautés, ce qui ne permet pas de différencier les communautés polarisées des chambres d'écho.

Les frontières ont un impact majeur sur la polarisation de leur communauté en définissant à la fois comment elle est exposée aux autres et comment les autres lui sont exposés. Une première approche non-supervisée, basée sur les frontières des communautés, a été décrite par Guerra et al. dans (Guerra et al. 2013) avec comme objectif de proposer une valeur complémentaire aux indicateurs de cohésion et d'homophilie de la théorie des graphes comme la modularité (Newman 2006). L'antagonisme entre les communautés est évalué selon l'investissement des utilisateurs qui interagissent à la fois avec l'intérieur et l'extérieur de leur communauté. Cette approche ne nécessite aucune connaissance *a priori* sur le graphe étudié ou sur les individus représentés et peut donc être incluse dans un processus d'analyse

complètement automatique utilisable par des experts du domaine. Cependant, sa spécification est limitée aux graphes non-dirigés et non-pondérés, ce qui est éloigné des graphes réels construits à partir d'interactions sociales, qui sont au contraire généralement dirigés et pondérés. De plus, cette approche, comme les précédentes, ne traite pas du cas des communautés recouvrantes (Devi & Poovammal 2016, Saini & Mangat 2023, Xie et al. 2013). Or, les utilisateurs des réseaux sociaux appartiennent plus naturellement à plusieurs communautés (Xie et al. 2013).

Pour conclure, les méthodes non-supervisées constituent la meilleure alternative pour détecter la polarisation. Elles permettent notamment aux experts du domaine, tels que des sociologues ou des décideurs, d'utiliser l'approche choisie en toute autonomie. Elles permettent également la comparaison de résultats issus de jeux de données différents. Ces deux propriétés sont très importantes pour l'analyse des réseaux sociaux. Il est également nécessaire de prendre en considération les interactions entre communautés pour éviter une mauvaise interprétation de leur polarisation. Pour cela, il est essentiel d'examiner le comportement des frontières des communautés.

3. La méthode ERIS

Dans cette section, nous définissons formellement la méthode ERIS et ses deux indicateurs : l'antagonisme exprimé par une communauté envers une autre et la porosité de ses frontières. Notre approche prend en considération la direction et la pondération des arêtes et est compatible avec des structures communautaires recouvrantes. Nous proposons également un algorithme basé sur des opérations matricielles pour calculer rapidement les indicateurs de polarisation.

3.1. *Définitions formelles*

Symbole	Définition	Symbole	Définition
C_i	Communauté d'indice i	$v \in C_i$	Sommet membre de C_i
$\mathrm{e}_{v,n}$	Arête dont la source est le sommet v et la destination le sommet n	$\mathrm{w}(\mathrm{e}_{v,n})$	Poids de l'arête $\mathrm{e}_{v,n}$
$I_{i,j}$	Zone interne de la communauté C_i formée par l'étude de la paire (C_i, C_j)	$B_{i,j}$	Zone frontière de la communauté C_i formée par l'étude de la paire (C_i, C_j)
$IE_{i,j}$	Arêtes internes pour la paire (C_i, C_j)	$EE_{i,j}$	Arêtes externes pour la paire (C_i, C_j)
$IE^v_{i,j}$	Arêtes internes pour la paire (C_i, C_j) dont la source est le sommet v	$EE^v_{i,j}$	Arêtes externes pour la paire (C_i, C_j) dont la source est le sommet v
$A^v_{i,j}$	Antagonisme exprimé par le sommet v, membre de C_i, envers la communauté C_j	$A_{i,j}$	Antagonisme exprimé par les sommets frontières de C_i envers C_j
$NB_{i,j}$	Membres de la frontière $B_{i,j}$ qui ont une valeur d'antagonisme négative	$\lvert B_{i,j}\rvert$	Nombre de sommets dans $B_{i,j}$
$P_{i,j}$	Porosité de la zone frontière $B_{i,j}$		

TABLEAU 1. *Conventions de nommage pour les définitions. Une zone est un ensemble de sommets.*

Dans les définitions suivantes, un graphe $G = (V, E)$ est composé d'un ensemble de sommets V et d'un ensemble d'arêtes dirigées $E \subseteq V \times V$. Une arête $e_{v,n} \in E$ connecte une source $v \in V$ et une destination $n \in V$ avec un poids $w(e_{v,n}) \in \mathbb{R}$. Les communautés sont des sous-ensembles de V densément connectés. Un sommet représente un individu et une arête une interaction entre individus. Les poids des arêtes sont des entiers positifs représentant le nombre d'interactions entre deux individus.

La figure 1 présente un exemple de graphe. Les conventions de nommage utilisées dans les paragraphes suivants sont synthétisées dans le tableau 1.

Deux communautés (C_i, C_j) sont polarisées si elles sont mutuellement antagonistes. Selon (Guerra et al. 2013) et (Baumann et al. 2020), un investissement important d'un individu frontière au sein de sa communauté, notamment traduit par de nombreuses interactions avec les membres internes, révèle un attachement émotionnel à la communauté et à ses sujets d'intérêt principaux. Cet attachement favorise l'expression d'antagonisme en réponse à une critique, à une attaque ou au partage d'une opinion ou information négative sur la communauté.

La méthode ERIS consiste à identifier deux zones pour chaque paire de communautés distinctes (C_i, C_j) :

— la **zone interne** $I_{i,j}$ de C_i, c'est-à-dire l'ensemble des sommets de C_i qui ne sont pas sources d'une arête en direction de C_j ;

— la **zone frontière** $B_{i,j}$ de C_i, c'est-à-dire l'ensemble des sommets de C_i qui sont sources d'au moins une arête en direction de $I_{i,j}$ et d'une autre en direction de C_j.

La méthode évalue l'antagonisme moyen exprimé par la communauté C_i envers la communauté C_j en mesurant l'investissement des sommets de B_{ij}.

Nous proposons des définitions formelles pour ces différents concepts. Une application de ces définitions sur le graphe exemple de la figure 1 est proposée en guise d'illustration à la fin de cette sous-section.

Dans les définitions suivantes, nous considérons $i \neq j$, c'est-à-dire deux communautés C_i et C_j distinctes. Pour chaque paire de communautés (C_i, C_j), on peut former une zone interne $I_{i,j}$ et une zone frontière $B_{i,j}$ (qui peuvent dans certains cas être vides) :

$$I_{i,j} = \{v \in C_i \mid \forall n \in C_j, \nexists e_{v,n}\} \tag{1}$$

$$B_{i,j} = \{v \in C_i \mid (\exists n_1 \in C_j, \exists e_{v,n_1}) \wedge (\exists n_2 \in I_{i,j}, \exists e_{v,n_2})\} \tag{2}$$

Pour chaque frontière non vide, nous considérons l'ensemble des arêtes dont la destination se situe dans l'autre communauté (**arêtes externes** ou $EE_{i,j}$) ainsi que l'ensemble des arêtes dont la destination se situe dans la zone interne $I_{i,j}$ de C_i (**arêtes internes** ou $IE_{i,j}$) :

$$EE_{i,j} = \{e_{s,d} \mid s \in B_{i,j} \wedge d \in C_j\} \tag{3}$$

$$IE_{i,j} = \{e_{s,d} \mid s \in B_{i,j} \wedge d \in I_{i,j}\} \tag{4}$$

Nous considérons également les arêtes externes ($EE^v_{i,j}$) et internes ($IE^v_{i,j}$) associées à un sommet v en tant que sous-ensembles d'arêtes, respectivement inclus dans $EE_{i,j}$ et dans $IE_{i,j}$, dont v est la source :

$$EE^v_{i,j} = \{e_{v,d} | e_{v,d} \in EE_{i,j}\} \tag{5}$$

$$IE^v_{i,j} = \{e_{v,d} | e_{v,d} \in IE_{i,j}\} \tag{6}$$

L'**antagonisme** $A^v_{i,j}$ exprimé par un sommet v correspond au ratio entre la somme des poids de ses arêtes internes et la somme des poids de ses arêtes internes ou externes. Cette valeur est comparée à l'hypothèse nulle suivante : chaque sommet possède autant d'arêtes en direction de la zone interne de sa communauté qu'en direction de l'autre communauté (Guerra et al. 2013). La formule de $A^v_{i,j}$ est donc donnée par :

$$A^v_{i,j} = \frac{\sum_{e \in IE^v_{i,j}} w(e)}{\sum_{e \in IE^v_{i,j}} w(e) + \sum_{e \in EE^v_{i,j}} w(e)} - 0.5 \tag{7}$$

Pour finir, l'antagonisme $A_{i,j}$ exprimé par la frontière $B_{i,j}$ correspond à la valeur moyenne d'antagonisme exprimé par ses membres :

$$A_{i,j} = \frac{1}{|B_{i,j}|} \sum_{v \in B_{i,j}} A^v_{i,j} \tag{8}$$

En mesurant les valeurs d'antagonisme pour chaque paire possible de communautés dans un graphe, on obtient une matrice asymétrique appelée **matrice d'antagonisme**, composée de réels compris entre -0.5 et 0.5. Une frontière de communauté dont la valeur avoisine 0.5 est susceptible d'exprimer de l'antagonisme envers l'autre communauté de la paire. Les valeurs sur les lignes de la matrice d'antagonisme indiquent à quel point la communauté associée à une ligne est susceptible d'exprimer de l'antagonisme envers les communautés représentées dans les colonnes. Inversement, les valeurs sur les colonnes indiquent à quel point la communauté associée à une colonne est susceptible de recevoir de l'antagonisme de la part des communautés représentées dans les lignes.

Les sommets des frontières qui possèdent une valeur d'antagonisme négative fragilisent la polarisation de leur communauté. En effet, en interagissant plus avec l'extérieur qu'avec l'intérieur, ils réduisent l'isolement de leur communauté et, de ce fait, le risque de devenir une chambre d'écho. De plus, puisqu'ils appartiennent à la communauté, ils paraissent également plus crédibles aux yeux des autres membres lorsqu'ils partagent des opinions plus nuancées à propos des sujets d'intérêts principaux de leur communauté (Donkers & Ziegler 2021). Pour mesurer cette fragilité, nous proposons un nouvel indicateur $P_{i,j}$, appelé **porosité** de la frontière $B_{i,j}$:

$$P_{i,j} = \frac{|NB_{i,j}|}{|B_{i,j}|} \times 100 \tag{9}$$

avec $NB_{i,j} = \left\{ v \in B_{i,j} \mid A^v_{i,j} < 0 \right\}$ le sous-ensemble de $B_{i,j}$ qui inclut tous les sommets possédant une valeur d'antagonisme négative. Les valeurs de porosité peuvent aussi être représentées au sein d'une matrice asymétrique appelée **matrice de porosité**.

Nous illustrons maintenant les différents ensembles et valeurs présentés dans cette sous-section en les appliquant au graphe exemple de la figure 1. Nous concentrons nos explications sur la paire de communautés (C_1, C_2).

Pour cette paire, la zone interne $I_{1,2}$ correspond à l'ensemble des sommets de la communauté C_1 qui ne sont la source d'aucune arête dirigée vers un membre de la communauté C_2. On remarque que les sommets 1 et 2 ne remplissent pas ce critère puisqu'ils sont notamment sources des arêtes $e_{1,4}$ et $e_{2,6}$ qui ont pour destinations des sommets qui appartiennent à la communauté C_2 (sommets 4 et 6). Seul le sommet 3 n'est source d'aucune arête dirigée vers C_2. Pour la paire de communautés (C_1, C_2), on a donc $I_{1,2} = \{3\}$. Sur la figure 1, les zones internes sont représentées par des cadres avec un trait plein. La couleur du cadre varie en fonction de la seconde communauté de la paire. Ici, le cadre qui entoure $I_{1,2}$ prend donc la couleur de C_2, soit la couleur rouge.

La zone frontière $B_{1,2}$ contient tous les sommets de C_1 qui sont sources à la fois d'une arête dirigée vers un membre de C_2 et d'une arête dirigée vers un membre de la zone interne $I_{1,2}$. Comme dit dans le paragraphe précédent, les sommets 1 et 2 sont tous les deux sources d'une arête dirigée vers la communauté C_2. Mais ils sont également sources des arêtes $e_{1,3}$ et $e_{2,3}$, qui ont pour destination un sommet membre de $I_{1,2}$ (sommet 3). Pour la paire de communautés (C_1, C_2), on a donc $B_{1,2} = \{1, 2\}$. Sur la

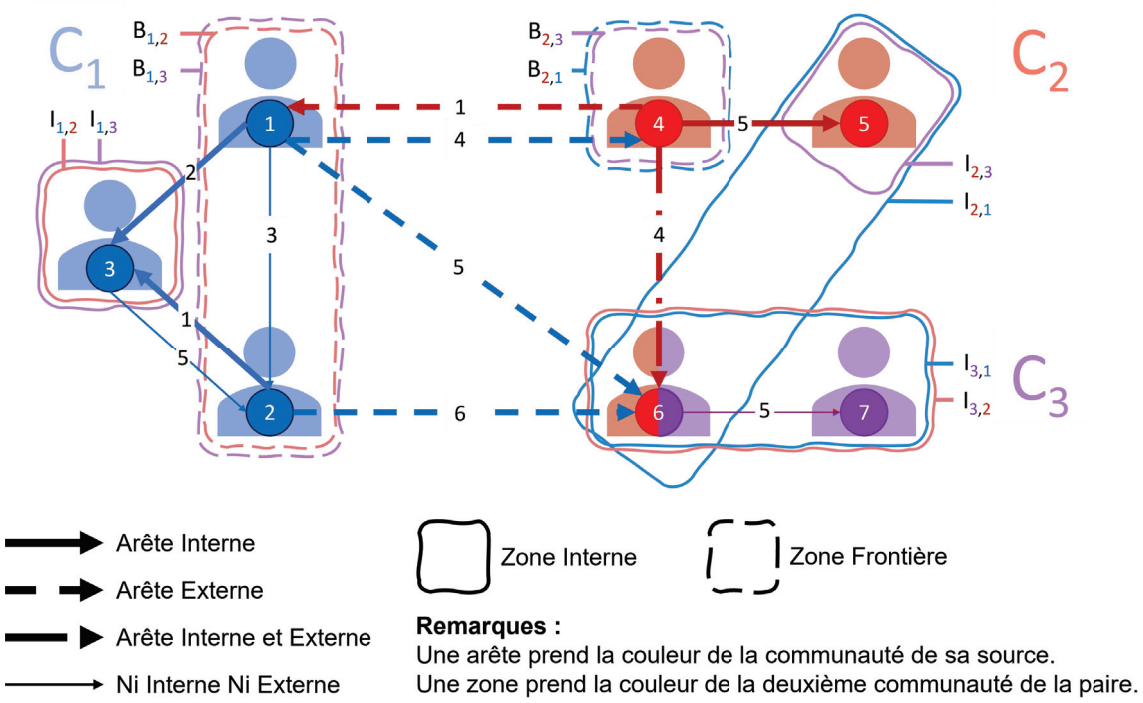
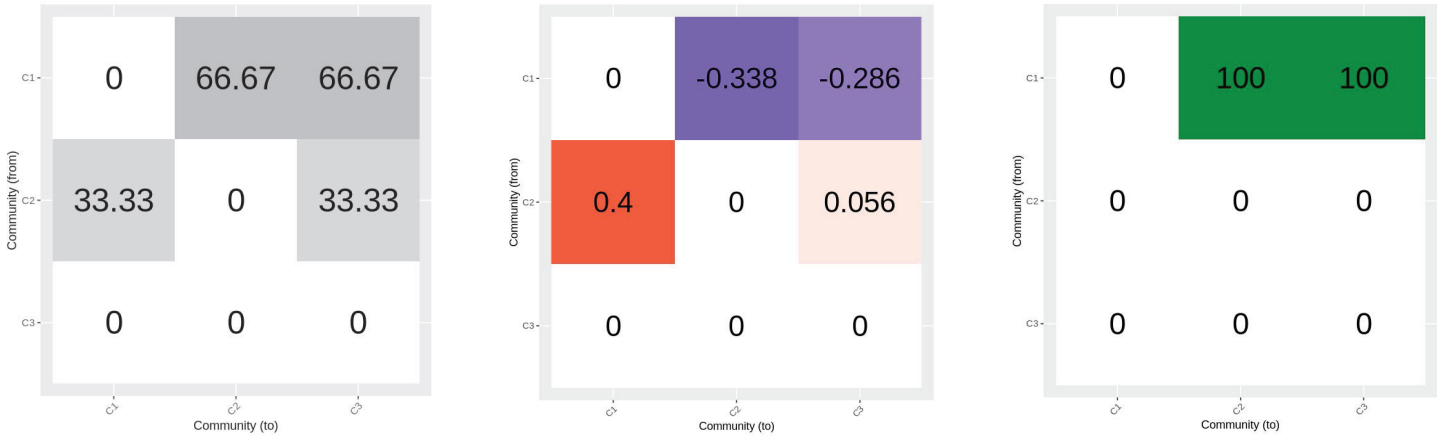


FIGURE 1. *Graphe exemple avec 3 communautés C_1 (bleu), C_2 (rouge) and C_3 (violet). Les communautés C_2 et C_3 se recouvrent sur le sommet 6.*

figure 1, cette zone frontière est représentée par un cadre rouge en trait pointillé autour des sommets 1 et 2.

Pour la paire de communautés (C_1, C_2), les arêtes internes $IE_{1,2}$ sont les arêtes du graphe qui ont comme source le sommet 1 ou le sommet 2 (membres de $B_{1,2}$) et comme destination le sommet 3 (membre de $I_{1,2}$). Les arêtes externes $EE_{1,2}$ sont celles qui ont pour source le sommet 1 ou le sommet 2 (membres de $B_{1,2}$) et comme destination un sommet membre de C_2. On identifie alors les ensembles d'arêtes suivants : $IE_{1,2} = \{e_{1,3}, e_{2,3}\}$ et $EE_{1,2} = \{e_{1,4}, e_{1,6}, e_{2,6}\}$. Sur la figure 1, les arêtes internes sont représentées par des flèches pleines épaisses et les arêtes externes par des flèches pointillées.



(a) *Taille des frontières (en %)* (b) *Matrice d'antagonisme* (c) *Matrice de porosité (en %)*

FIGURE 2. *Indicateurs calculés sur le graphe de la figure 1*

En appliquant les formules de calcul de l'antagonisme et de la porosité aux ensembles identifiés précédemment, on peut calculer les valeurs des matrices correspondantes en position (C_1, C_2). En répétant ce calcul pour chaque paire de communautés possible où $i \neq j$, on peut obtenir les matrices complètes.

La figure 2 présente les matrices d'antagonisme et de porosité complètes obtenues à partir du graphe exemple. La figure 2a renseigne sur la taille de chaque frontière, exprimée en tant que pourcentage de membres de la communauté C_i qui font partie de la zone frontière $B_{i,j}$. Pour la paire (C_1, C_2), deux sommets de C_1 sur trois (sommets 1 et 2) font partie de la zone frontière $B_{1,2}$, soit environ 66.67% de ses membres. Cette matrice permet notamment de différencier les valeurs nulles des deux autres matrices qui ont été calculées de celles qui ont été mises par défaut car la zone frontière concernée est vide.

Sur la figure 2b, on remarque que la valeur $A_{1,2}$ d'antagonisme exprimé par la frontière de la communauté C_1 envers la communauté C_2 est égale à -0.338. On obtient ce résultat de la manière suivante :

$$A^1_{1,2} = \frac{w(e_{1,3})}{w(e_{1,3}) + w(e_{1,4}) + w(e_{1,6})} - 0.5 = \frac{2}{2+4+5} - 0.5 = -\frac{7}{22} \tag{10}$$

$$A^2_{1,2} = \frac{w(e_{2,3})}{w(e_{2,3}) + w(e_{2,6})} - 0.5 = \frac{1}{1+6} - 0.5 = -\frac{5}{14} \tag{11}$$

$$A_{1,2} = \frac{-\frac{7}{22} - \frac{5}{14}}{2} \approx -0.338 \tag{12}$$

Cette valeur d'antagonisme négative ne révèle pas de comportement susceptible d'engendrer de l'antagonisme. *A contrario*, on peut remarquer que la frontière de la communauté C_2 est susceptible d'exprimer de l'antagonisme envers la communauté C_1 puisque sa valeur d'antagonisme $A_{2,1}$ est égale à 0.4. Cette valeur reflète directement le comportement de la frontière concernée. En effet, on peut remarquer sur la figure 1 que le membre de la zone frontière $B_{2,1}$ (sommet 4) ne s'expose que très peu aux membres de C_1, avec seulement une interaction avec le sommet 1, alors qu'il interagit beaucoup avec les membres de la zone interne $I_{2,1}$ (sommets 5 et 6). Si on se réfère aux observations de (Guerra et al. 2013) et (Baumann et al. 2020), on peut considérer ce comportement comme susceptible d'être vecteur d'antagonisme, d'où une valeur proche de 0.5.

Sur la figure 2c, la valeur de porosité $P_{1,2}$ est égale à 100%, ce qui signifie que $B_{1,2}$ est très poreuse. Cela traduit le fait que tous ses membres possèdent une valeur d'antagonisme négative et donc interagissent plus avec C_2 qu'avec la zone interne $I_{1,2}$.

3.2. *Algorithme*

La matrice d'adjacence du graphe est le point de départ de notre algorithme, qui utilise ensuite une série d'opérations matricielles pour obtenir les matrices d'antagonisme (M_{ANT}) et de porosité (M_{POR}). Les conventions de nommage utilisées dans les paragraphes suivants sont répertoriées dans le tableau 2. Nous définissons ✦ en tant qu'opérateur réalisant le produit terme à terme entre un vecteur de taille N et chaque colonne d'une matrice de taille $N \times M$. Le résultat de cette opération est une nouvelle matrice de taille $N \times M$.

Symbole	Définition	Symbole	Définition
G	Le graphe à analyser	C	L'ensemble des communautés de G
V	L'ensemble des sommets de G	$\|V\|$	Le nombre de sommets dans G
E	L'ensemble des arêtes de G	$\|C\|$	Le nombre de communautés dans G

TABLEAU 2. *Conventions de nommage*

L'algorithme prend en entrée la matrice d'adjacence M_A de G et une matrice d'appartenance aux communautés M_C. M_A est une matrice carrée de taille $|V| \times |V|$ qui contient dans chaque cellule le poids de l'arête dont la source est le sommet associé à la ligne et la destination le sommet associé à la colonne. M_C est une matrice binaire de taille $|V| \times |C|$ dans laquelle la valeur 1 indique que le sommet associé à la ligne appartient à la communauté associée à la colonne. Si ce n'est pas le cas, la valeur est 0.

Symbole	Taille	Type	Nom
M_A	$\|V\| \times \|V\|$	Int	Matrice d'adjacence
M_C	$\|V\| \times \|C\|$	Bin	Matrice d'appartenance aux communautés
M_{EE}	$\|V\| \times \|C\|$	Int	Matrice des poids des arêtes externes
M_I	$\|V\| \times \|C\|$	Bin	Matrice d'appartenance aux zones internes
M_{IC}	$\|V\| \times \|C\|$	Bin	Matrice courante d'appartenance aux zones internes
M_{IE}	$\|V\| \times \|C\|$	Int	Matrice des poids des arêtes internes
M_{BIE}	$\|V\| \times \|C\|$	Bin	Masque binaire des poids des arêtes internes
M_{VANT}	$\|V\| \times \|C\|$	Real	Matrice d'antagonisme par sommet
M_{ANT}	$\|C\| \times \|C\|$	Real	Matrice d'antagonisme
M_{POR}	$\|C\| \times \|C\|$	Real	Matrice de porosité

TABLEAU 3. *Matrices utilisées dans l'algorithme 1*

Algorithme 1 Algorithme basé sur des produits de matrices pour estimer les indicateurs d'ERIS

Require: M_A, M_C

Ensure: M_{ANT}, M_{POR}

1: $M_{EE} \leftarrow M_A \times M_C$
2: $M_I \leftarrow (M_{EE} == 0)$
3: **for** $c = 1, \ldots, |C|$ **do**
4: $\quad M_{IC} \leftarrow M_C[, c] \blacklozenge (M_I \cdot \neg M_C)$
5: $\quad M_{IE} \leftarrow (M_C[, c] \blacklozenge (M_A \times M_{IC})) \cdot \neg M_I$
6: $\quad M_{BIE} \leftarrow (M_{IE}! = 0)$
7: $\quad M_{VANT} \leftarrow ((M_{IE}/(M_{IE} + M_{EE})) - 0.5) \cdot M_{BIE}$
8: $\quad M_{ANT}[c,] \leftarrow (M_C^T \times M_{VANT})/(M_C^T \times M_{BIE})$
9: $\quad M_{POR}[c,] \leftarrow 100 * (M_C^T \times (M_{VANT} < 0))/(\sum_{i=1}^{|V|} M_{BIE}[i,])$
10: **end for**

La phase d'initialisation de l'algorithme consiste à calculer M_{EE}, une version agrégée de M_A où les valeurs sont regroupées par communauté (ligne 1). La matrice contient la somme des poids des arêtes dont la source est le sommet associé à la ligne et la destination un sommet appartenant à la communauté associée à la colonne. Cette matrice est ensuite utilisée pour extraire M_I, un masque binaire de M_{EE} au sein duquel les sommets appartenant à au moins une zone interne de leur(s) communauté(s) sont identifiés (ligne 2).

À partir de ces deux matrices, la partie principale de l'algorithme calcule, pour chaque communauté, les valeurs d'antagonisme et de porosité de ses frontières en suivant quatre étapes :

— la détection des zones internes de la communauté courante c pour chaque paire impliquant c (ligne 4) ;

— l'agrégation de M_A pour sommer les poids des arêtes dont la destination se situe dans une des zones internes de c (ligne 5) ;
— le calcul des valeurs d'antagonisme des sommets appartenant aux frontières de c (lignes 6-7) ;
— le calcul des valeurs d'antagonisme et de porosité des frontières de c (lignes 8-9).

Cette approche permet de manipuler tous les graphes de la même manière, qu'ils soient dirigés, pondérés, ou non. De même, le fait que les communautés soient recouvrantes n'entraîne pas de traitement particulier. Par conséquent, prendre en considération ces paramètres n'impacte pas la complexité de l'algorithme. Une implémentation en R est disponible en libre accès sur GitHub [1].

4. Expérimentations

Dans cette section, nous expérimentons notre proposition sur de grands graphes au travers d'une étude de la complexité de l'algorithme principal de la méthode ERIS. Nous présentons également une étude de cas dont les résultats ont été validés par des experts du domaine. La sous-section 4.3 discute de certaines limites actuelles de la méthode ERIS et de possibles améliorations pour de futurs travaux.

4.1. *Exécution sur des grands graphes*

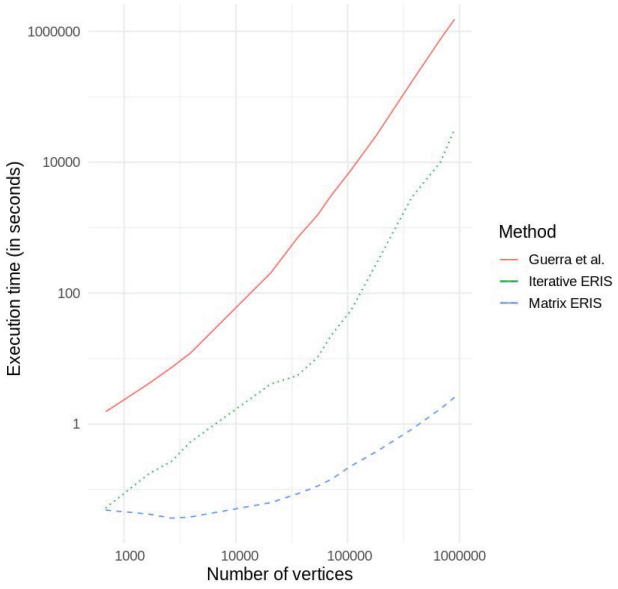
Nous comparons les temps d'exécution de 3 algorithmes qui mesurent la polarisation des communautés au sein de graphes d'interactions entre individus :
— l'algorithme de la méthode ERIS basé sur des opérations matricielles, présenté dans la section précédente (implémenté en R) ;
— un algorithme itératif de la méthode ERIS proposé dans un article précédent (Guyot et al. 2021) (implémenté en R) ;
— le seul algorithme issu des travaux de Guerra et al. (Guerra et al. 2013) disponible en ligne [2], non développé par les auteurs (implémenté en Python).

Nous avons choisi de comparer ERIS avec la méthode de Guerra et al. (Guerra et al. 2013) puisque les deux approches partagent un certain nombre de caractéristiques communes (aucune supervision nécessaire, calcul d'antagonisme, basé sur des graphes, etc.).

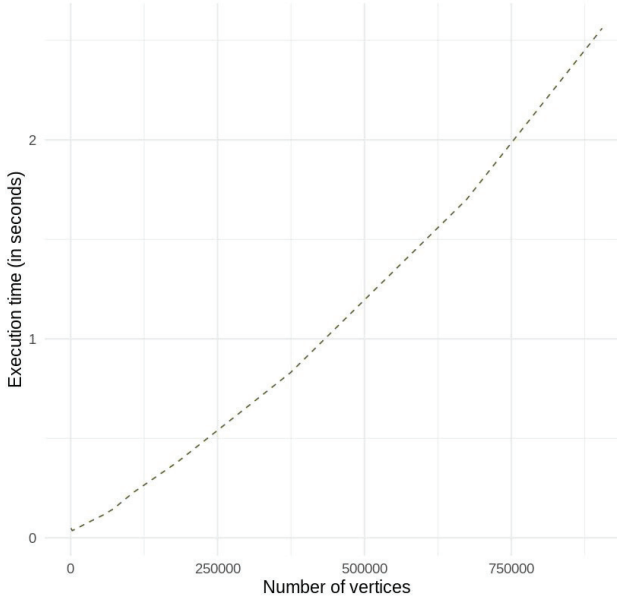
Nous avons généré des graphes artificiels de tailles décroissantes, en faisant varier le nombre de sommets de 1 million à 500, à partir d'un graphe de départ issu de données réelles collectées à partir de Twitter. Dans cette sous-section, nous ne prenons pas en considération la sémantique des indicateurs calculés. Nous nous intéressons seulement à l'impact de la structure du graphe sur le temps d'exécution de l'algorithme.

Pour chaque algorithme, nous avons mesuré le temps écoulé entre l'appel à la fonction calculant les matrices d'antagonisme (l'indicateur commun aux trois méthodes) et le retour d'un résultat [3]. Pour l'algorithme 1, cela correspond au temps d'exécution des lignes 1 à 10. Les trois algorithmes ont été exécutés sur un serveur Dell PowerEdge R440 avec les caractéristiques suivantes : Intel(R) Xeon(R) Bronze 3204 CPU @ 1.92GHz, 6 cores, 128Go RAM.

1. https://github.com/AlexisGuyot/ERIS
2. https://github.com/rachel-bastos/boundaries-polarization
3. Voir https://github.com/AlexisGuyot/ERIS/tree/main/experiment_complexity pour des explications plus détaillées sur l'expérimentation mise en place.



(a) *Comparaison entre les temps d'exécution des 3 méthodes (échelle log-log)*



(b) *Focalisation sur les temps d'exécution de l'algorithme basé sur les opérations matricielles*

FIGURE 3. *Temps d'exécution des algorithmes*

Les temps d'exécution des trois algorithmes sont comparés dans la figure 3a. La figure 3b se concentre sur les temps d'exécution de l'algorithme 1 décrit dans la section précédente. Nous constatons que l'algorithme de la méthode ERIS basé sur les opérations matricielles surpasse toutes les autres implémentations. Sur notre plus grand graphe, celui avec 1 million de sommets, cet algorithme a pris 2.5 secondes pour calculer la matrice d'antagonisme. C'est 12 828 fois plus rapide que l'algorithme itératif de la méthode ERIS (32 070 secondes soit presque 9 heures) et 595 399 fois plus rapide que l'algorithme inspiré de la méthode de Guerra et al. (1 528 389 secondes soit plus de 17 jours).

En théorie, la complexité algorithmique pour calculer les deux indicateurs de polarisation avec l'algorithme de la méthode ERIS basé sur les opérations matricielles est $\mathcal{O}(|V|^2|C|^2)$, tant que $|C| < \sqrt{\frac{|V|}{3}}$. Au-delà de cette valeur, l'ordre atteint $\mathcal{O}(|V||C|^3)$. Cependant, en pratique, le nombre de communautés significatives dans un graphe reste relativement restreint à cause de la limite de résolution des algorithmes utilisant une mesure de modularité (Fortunato & Barthelemy 2007). De plus, les experts du domaine ont généralement besoin de conserver un nombre limité de communautés pour que les résultats restent humainement interprétables. De ce fait, $|C| \ll |V|$ et la complexité algorithmique peut donc être réduite à $\mathcal{O}(|V|^2)$.

À partir des résultats des études théoriques et pratiques de la complexité algorithmique de notre proposition, nous pouvons conclure que l'algorithme de la méthode ERIS basé sur les opérations matricielles atteint nos objectifs d'applicabilité sur de grands graphes et surpasse de plusieurs ordres de grandeur les autres algorithmes équivalents actuellement disponibles.

4.2. *Étude de cas sur des données réelles*

Nous illustrons expérimentalement l'intérêt de notre approche au travers d'une étude de cas menée sur des données réelles dans le cadre du projet interdisciplinaire Cocktail[4], dont le but est d'étudier la

4. https://projet-cocktail.fr/

circulation des discours sur Twitter dans les domaines de l'alimentaire et de la santé. Notre corpus est constitué de plus de 18 millions de tweets en français portant sur la thématique des vaccins contre la COVID-19, collectés par l'architecture Hydre (Gillet et al. 2021) entre le 1^{er} décembre 2020 et le 31 mars 2021 (120 jours).

À partir de ce jeu de données, nous avons extrait un graphe dirigé G_Q de citations [5], dans lequel le sommet représentant l'individu u est source d'une arête dont la destination est le sommet représentant l'individu n si u a déjà cité plus de deux fois les tweets de n. Le poids w d'une arête indique la fréquence exacte à laquelle u a cité n. De la même manière que (Garimella, Morales, Gionis & Mathioudakis 2018), nous ne considérons pas les citations isolées, qui peuvent constituer du bruit aléatoire. Les caractéristiques de G_Q sont présentées dans le tableau 4.

Nombre de sommets	24,591
Nombre d'arêtes	55,703
Force moyenne	4.46
Diamètre	338
Exposant de la loi de puissance γ	2.27
Limite de résolution	333
Nombre de communautés significatives	8
Modularité	0.59

TABLEAU 4. *Caractéristiques de G_Q*

Nous avons choisi la citation comme type d'interaction pour rester cohérents avec les travaux précédents menés sur la polarisation sur Twitter. En effet, la littérature considère généralement que les retweets impliquent une approbation du contenu partagé (Boyd et al. 2010), ce qui est incompatible avec une recherche d'antagonisme. De même, un réseau construit à partir de mentions est généralement considéré comme non polarisé (Conover et al. 2011). Les citations, quant à elles, sont souvent utilisées pour sortir un message de son contexte original à des fins d'humour ou pour le critiquer, ce qui peut mener à des réponses antagonistes (Guerra et al. 2017). Ainsi, les citations sont la meilleure alternative pour évaluer la polarisation d'un réseau à partir d'approches comme ERIS, basées sur les frontières de communautés.

G_Q est un réseau sans échelle puisque la distribution des degrés de ses sommets suit une loi de puissance d'exposant $2 < 2.27 < 3$ (Barabási & Pósfai 2016). Par conséquent, sa modularité peut être calculée et des algorithmes de détection de communautés basées sur cette mesure, comme Louvain (Blondel et al. 2008), peuvent être appliqués. Sur G_Q, cet algorithme identifie 8 communautés significatives, dont la taille dépasse la limite de résolution du graphe (Goldstein et al. 2004) (tableau 4).

Pour mieux appréhender les communautés et leurs relations, les experts du domaine du projet interdisciplinaire ont manuellement assigné à chaque communauté une étiquette correspondant à sa thématique principale. Pour cela, ils ont inspecté les 30 hashtags les plus utilisés dans chaque communauté (top-hashtags). Cette étape d'étiquetage a révélé que les deux plus grosses communautés de G_Q rassemblent respectivement des individus plutôt pro et plutôt anti-vaccins. Le tableau 5 liste les éléments qui ont permis d'inférer ces étiquettes parmi les hashtags du top-30 de ces deux communautés. En suivant le

5. Les citations sont une fonctionnalité de partage de Twitter avec ajout d'un commentaire supplémentaire.

même procédé, les six autres communautés ont reçu les étiquettes suivantes : Réactions aux Médias, Anti-Blanquer, Politique, Québecois, Anti-Gouvernement, Soutiens Raoult.

Communauté	**Top-hashtags**
Pro-vaccins	Mutation, Confinement3, CouvreFeu, Ecoles, PasseportVert, Dictature-Sanitaire, Israel, JeMeFaisVacciner, Pasteur
Anti-vaccins	Ivermectine, DictatureSanitaire, JeNeMeConfineraiPas, Raoult, Hydroxychloroquine, EtLesSoins, Plandemie, VeranDemission, LesPierres-Crieront, GreatReset, Ethique, BeBraveWHO, JeNeMeFeraiPasVacciner

TABLEAU 5. *Top-hashtags révélant la thématique principale des communautés pro et anti-vaccins*

Puisque ces deux thématiques sont opposées, nous nous attendons à détecter une polarisation entre ces communautés. Cette intuition repose également sur ce qui a été observé dans d'autres études sur le sujet, telles que (Jain et al. 2022) ou (Mønsted & Lehmann 2022). Si tel est le cas, les communautés pro et anti-vaccins devraient être des communautés renfermées interagissant mutuellement de manière antagoniste. Pour confirmer cette hypothèses, nous appliquons notre implémentation en R de l'algorithme 1 sur G_Q. Les résultats obtenus sont présentés dans les figures 5 et 6. La figure 4 présente des informations complémentaires à propos des tailles des frontières.

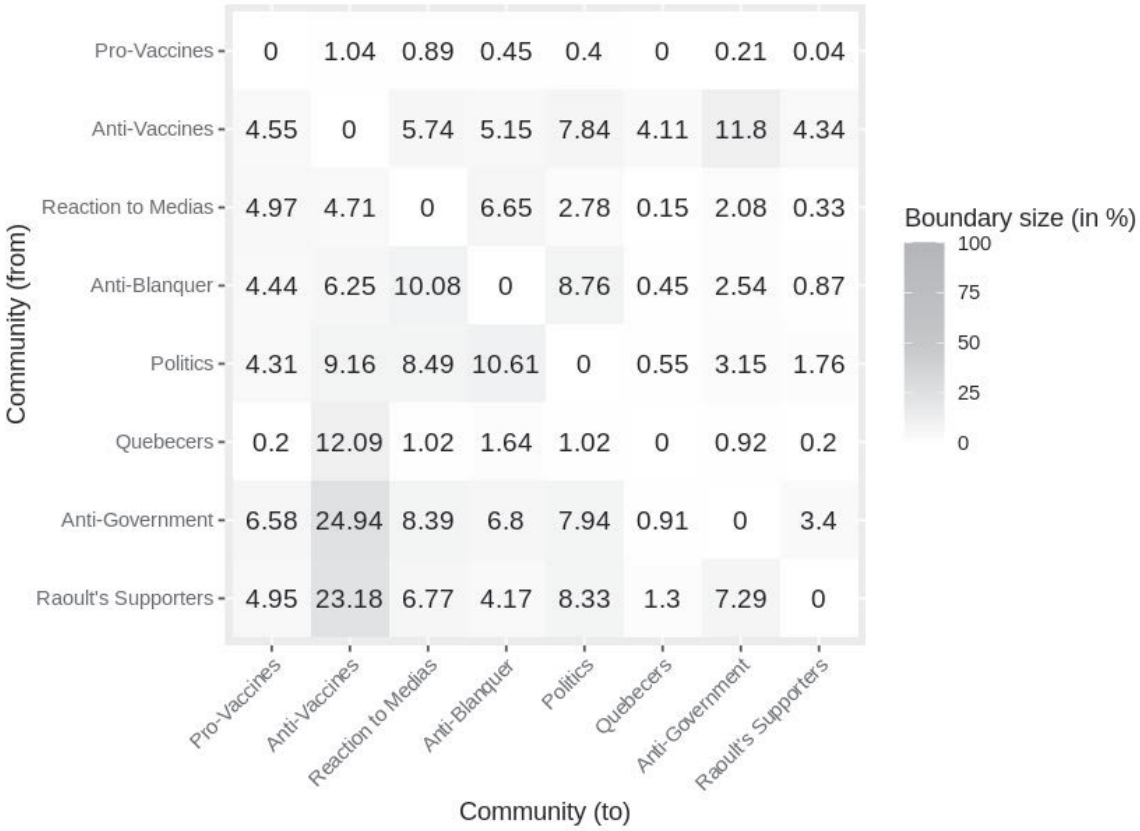


FIGURE 4. *Taille des frontières de G_Q*

Pour rappel, les valeurs sur les lignes de la matrice d'antagonisme (figure 5) révèlent à quel point la communauté associée à une ligne est susceptible d'exprimer de l'antagonisme envers les communautés représentées dans les colonnes. Inversement, les valeurs sur les colonnes indiquent à quel point la communauté associée à une colonne est susceptible de recevoir de l'antagonisme de la part des communautés représentées dans les lignes.

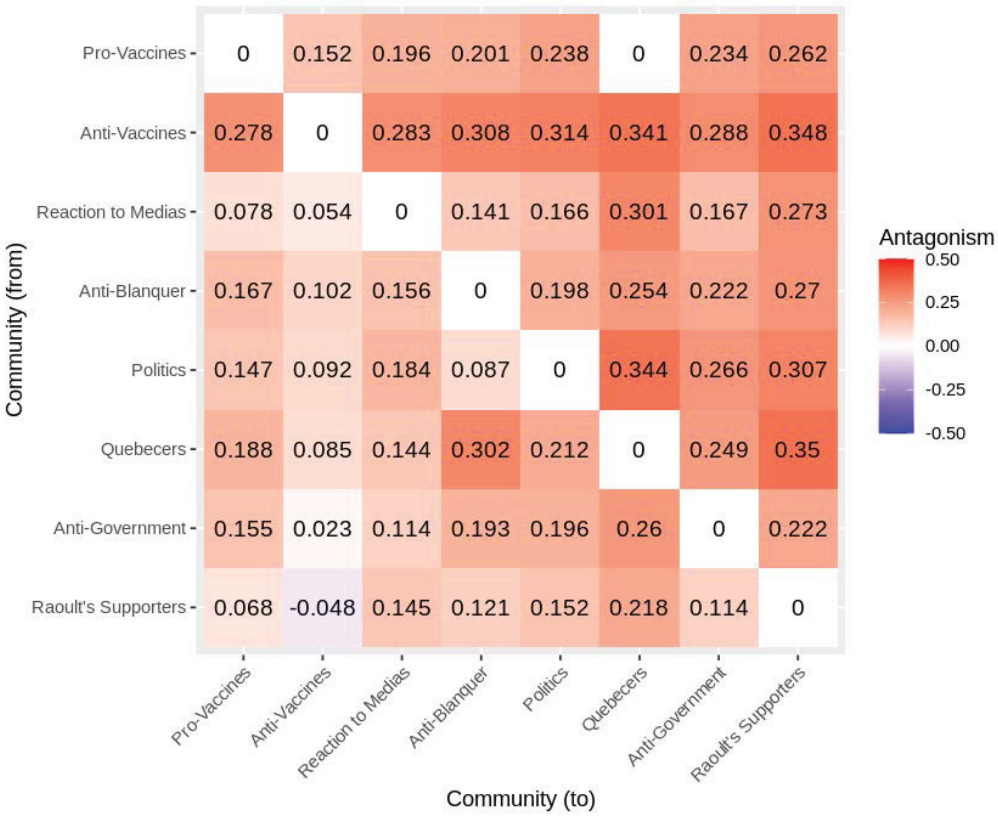


FIGURE 5. *Matrice d'antagonisme de* G_Q

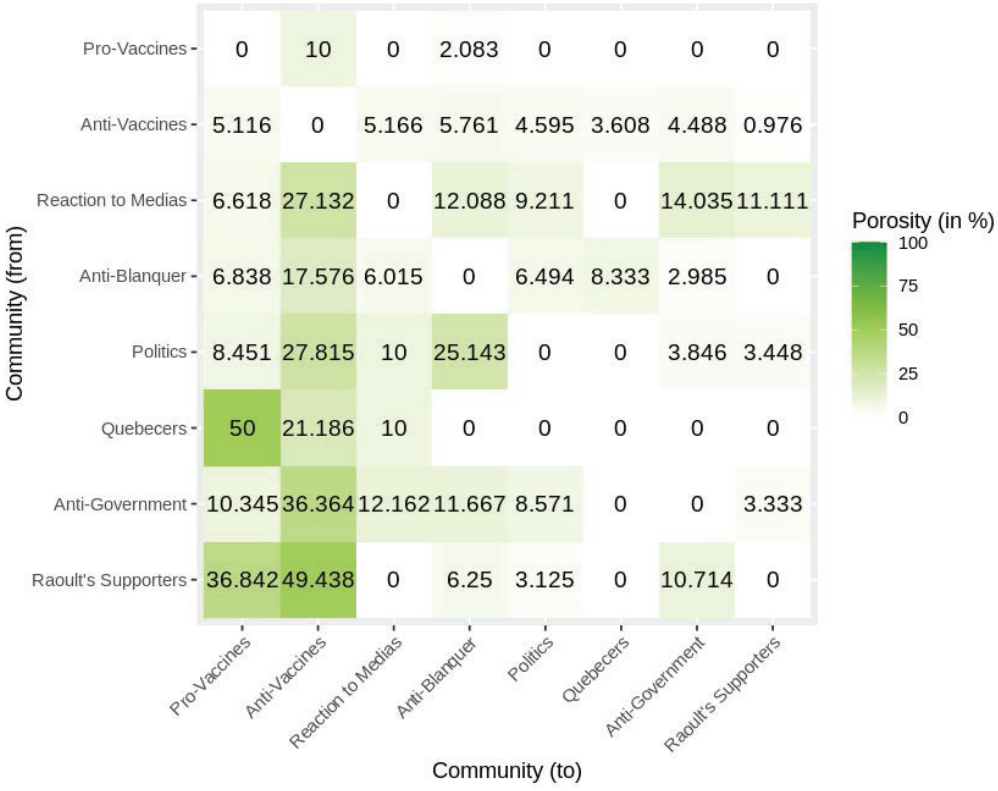


FIGURE 6. *Matrice de porosité de* G_Q

Les colonnes associées aux communautés pro et anti-vaccins montrent qu'aucune des deux ne reçoit beaucoup d'antagonisme des autres communautés. La communauté la plus susceptible d'être antagoniste avec la communauté pro-vaccins est la communauté anti-vaccins (0.278) et *vice versa* (0.152).
Les lignes associées à ces deux communautés montrent toutefois qu'elles sont toutes les deux assez susceptibles d'exprimer de l'antagonisme envers les autres communautés. Deux hypothèses peuvent expliquer pourquoi les valeurs entre ces deux communautés sont plus faibles que celles avec les autres :

— les frontières pro et anti-vaccins ne sont pas très antagonistes l'une avec l'autre ;
— les débats animés entre ces deux communautés mènent un nombre important de membres frontières à communiquer plus avec l'autre communauté qu'avec l'intérieur de la leur.

La matrice représentant la porosité des frontières (figure 6) permet de trancher entre ces deux hypothèses en identifiant la seconde comme plus probable. En effet, on remarque que 10% des membres frontières de la communauté pro-vaccins interagissent plus avec la communauté anti-vaccins qu'avec les membres de la zone interne de leur propre communauté. Cette valeur est beaucoup plus importante qu'avec toutes les autres communautés significatives du corpus (environ 5 fois plus que la seconde plus grande valeur). Ceci implique ainsi une présence importante de membres frontières avec une valeur d'antagonisme négative, qui viennent réduire la valeur d'antagonisme générale de la frontière. Pour la communauté anti-vaccins, cette valeur avoisine 5% et fait aussi partie des valeurs les plus élevées de la ligne. Ces valeurs importantes de porosité traduisent un plus fort investissement des frontières de ces deux communautés l'une envers l'autre qu'avec le reste des communautés. Les deux thématiques défendues par ces communautés étant totalement opposées, une interprétation possible à ce phénomène est l'existence d'un besoin plus important pour ces frontières de se contredire, qui intensifie davantage les débats qu'avec les autres communautés.

Il est possible d'affiner la compréhension des comportements et des rôles de ces deux communautés au sein de leur environnement en étudiant plus précisément les lignes et les colonnes de la matrice de porosité. D'abord, on peut constater que les valeurs sur les lignes associées à ces deux communautés sont plutôt faibles par rapport à celles associées aux autres communautés. Cela signifie qu'elles s'ouvrent moins aux interactions avec les autres. On peut alors considérer les communautés pro et anti-vaccins comme plutôt renfermées. Ce renfermement a également été observé dans (Mønsted & Lehmann 2022) sur un réseau d'interactions entre pro et anti-vaccins sur Twitter construit à partir de mentions plutôt que de citations.

Sur la ligne associée à la communauté anti-vaccins seulement, on peut également remarquer que, même si les valeurs sont assez faibles, toutes les frontières sont globalement autant poreuses les unes que les autres. On observe ainsi deux comportements différents de la part des communautés pro et anti-vaccins. D'un côté, les valeurs de porosité des frontières pro-vaccins montrent que ces dernières s'investissent peu auprès des autres communautés significatives du corpus, excepté la communauté anti-vaccins comme discuté précédemment. Cette communauté s'expose donc globalement peu sur les autres thématiques liées à la vaccination et à la COVID-19 mais annexes à la sienne comme par exemple le traitement médiatique ou la politique de gestion de crise. Et à l'inverse, les valeurs de porosité globalement constantes des frontières anti-vaccins montrent que cette communauté s'investit de façon équivalente auprès des autres communautés. Ses frontières sont donc dans l'ensemble plus exposées dans les différents débats. Ce comportement peut traduire un besoin plus important pour cette communauté d'avoir le contrôle sur le débat général autour de la vaccination et de la COVID-19 en s'exposant de manière plus systématique dans les discussions autour de différentes sous-thématiques.

	G_{RT}	G_{Q2}	G_{REP}	G_{RTC}
Nombre de sommets	1,189,541	239,488	151,469	1,145,159
Nombre d'arêtes	7,590,285	850,010	570,655	7,493,021
Exposant de la loi de puissance γ	2.12	2.17	3.61	2.12
Modularité	0.57	0.39	0.26	0.57
Limite de résolution	3896	1303	1068	3871
Nombre de communautés significatives	11	8	8	11

TABLEAU 6. *Caractéristiques des différents graphes manipulés dans la section 4.3.*

Finalement, les colonnes associées à ces deux communautés montrent qu'elles ont toutes les deux un impact assez important sur la porosité des frontières des autres communautés. Cela souligne un intérêt général de tous les individus qui forment notre corpus aux thématiques liées à la vaccination. Les plus grandes valeurs dans les colonnes associées à la communauté anti-vaccins révèlent une tendance à provoquer un plus grand nombre de réactions de la part des frontières des autres communautés. Puisque nous travaillons avec des citations, ces réactions pourraient être en grande partie du sarcasme, de l'ironie ou de l'humour, soit dans l'ensemble des réactions plutôt négatives.

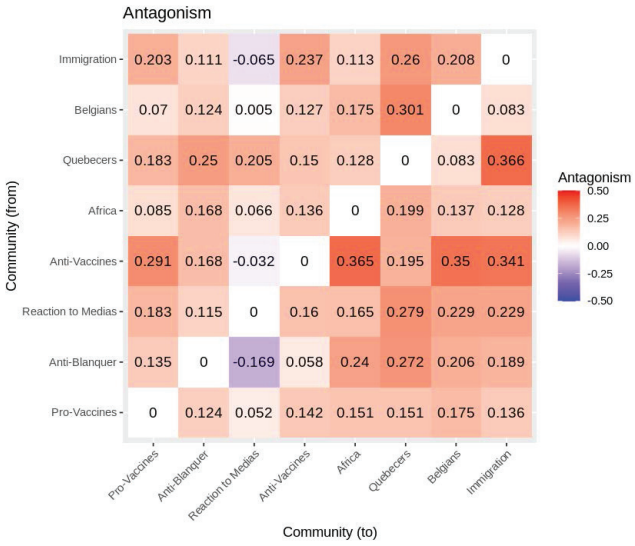
En résumé, les indicateurs de la méthode ERIS décrivent une relation susceptible d'être mutuellement antagoniste entre deux communautés plutôt renfermées. À partir de cette observation, nous pouvons conclure que, comme attendu et comme observé dans d'autres travaux similaires (Jain et al. 2022, Mønsted & Lehmann 2022), les communautés pro et anti-vaccins de notre corpus sont polarisées. La méthode ERIS a donc réussi avec succès à mettre en évidence des traces de polarisation au sein d'un graphe d'interactions entre individus sur un RSN. Ces déductions ont été rendues possibles par l'étude des frontières des communautés dans le contexte de graphes dirigés et pondérés, qui a permis une analyse et une compréhension plus fines de la structure des communautés et des rôles de leurs membres grâce aux matrices d'antagonisme et de porosité.

4.3. *Discussion sur les résultats d'ERIS*

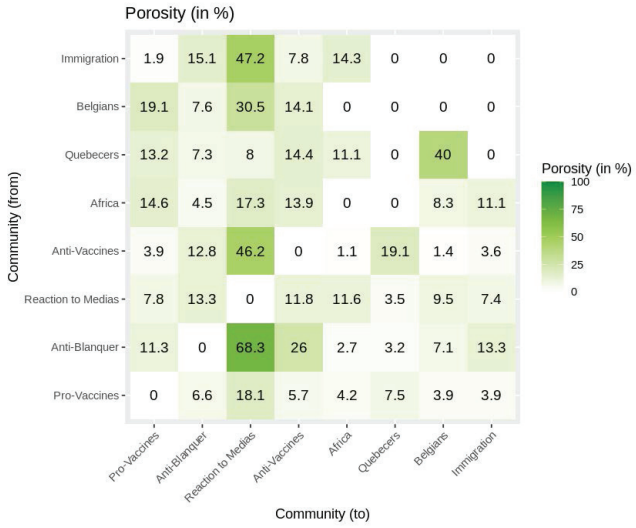
Dans cette sous-partie, nous discutons l'impact de différents facteurs sur la pertinence et la sémantique des résultats retournés par ERIS et de leur analyse. En particulier, nous mettons en avant l'importance de la méthode utilisée pour former les communautés et de celle de l'interaction choisie. Ces facteurs font apparaître certaines limites de la méthode. Nous discutons donc également de possibles pistes de résolution pour de futurs travaux à la fin de la sous-section.

Impact de la nature du graphe sur les résultats

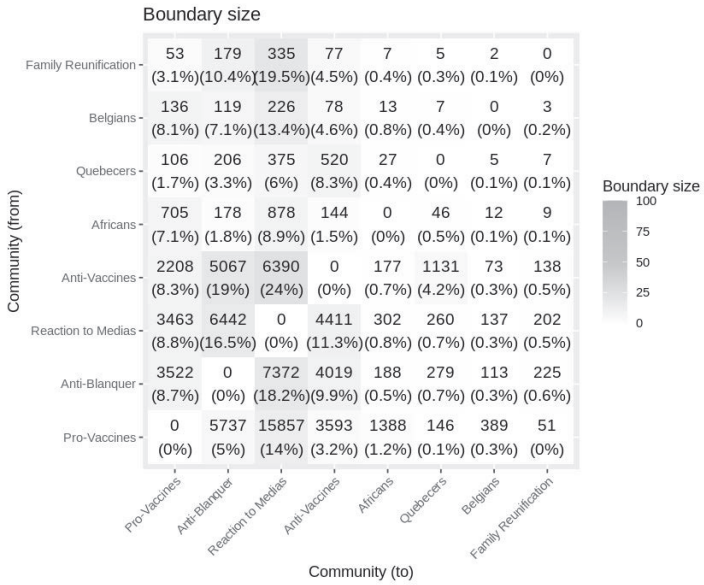
Dans un premier temps, nous proposons de ré-appliquer la méthodologie présentée dans la section 4.2 sur un graphe de citations issu du même corpus et construit de la même manière que G_Q mais en faisant varier la méthode utilisée pour identifier les communautés. Plutôt que de calculer directement les communautés sur le graphe de citations G_Q, nous commençons par construire un graphe de retweets G_{RT} (ses caractéristiques sont présentées dans le tableau 6) sur lequel nous appliquons l'algorithme de détection de communautés Louvain (Blondel et al. 2008). Nous retirons ensuite les sommets du graphe de citations G_Q qui ne sont pas présents dans G_{RT} pour obtenir un nouveau graphe de citations G_{Q2} sur



(a) *Matrice d'antagonisme de* G_{Q2}.



(b) *Matrice de porosité de* G_{Q2}.



(c) *Tailles des frontières des communautés de* G_{Q2}.

FIGURE 7. *Résultats retournés par ERIS sur le graphe de citations* G_{Q2}.

lequel nous pouvons attribuer à chaque sommet l'étiquette de la communauté à laquelle il appartient dans G_{RT}. L'objectif de cette manipulation est d'obtenir des communautés davantage cohésives en se basant sur le fait que le retweet implique une forme d'approbation du contenu partagé (Boyd et al. 2010).

L'application de la méthode ERIS sur le nouveau graphe de citations fournit les résultats présentés dans la figure 7. Le changement de méthode de détection de communautés entraîne l'émergence de nouvelles communautés. Après étiquetage par les experts du domaine, les thématiques principales suivantes leur ont été attribuées : Pro-vaccins, Anti-Blanquer, Réactions aux Médias, Anti-vaccins, Francophones d'Afrique, Francophones du Québec, Francophones de Belgique, Regroupement Familial. On remarque

certaines thématiques communes avec les résultats de la section 4.2, notamment l'existence d'une communauté explicitement pro-vaccins et d'une autre explicitement anti-vaccins. Il est toutefois important de noter que, même si les communautés pro ou anti-vaccins des résultats de la section 4.2 et des résultats de la figure 7 partagent probablement un certain nombre de membres en commun, il s'agit en réalité de communautés bien distinctes.

Communauté	**Top-hashtags**
Francophones d'Afrique	RDC, FreeSenegal, Gabon, Chinois, Covid19Sn, VariantCambodgien, Mali, Cameroun, Burundi, Tchad, CombodianMutant, Togo
Francophones du Québec	Polqc, Polcan, Covid19Qc, JDQ, PolMtl, CndPoli, LaPressePlus, Québec, EduQc
Francophones de Belgique	BeGov, Covid19Be, Belgique, Europe, BelgiumFailedState, ResetBelgium, Bruxelles, Wallonie
Regroupement Familial	Maroc, Algérie, Regroupement_Familial, LoveIsNotTourism, Visas_Regroupement_Familial, Vie_De_Famille, Tunisie

TABLEAU 7. *Top-hashtags des nouvelles thématiques qui apparaissent dans* G_{Q2}.

La matrice 7a fait apparaître de plus faibles valeurs d'antagonisme entre les frontières pro et anti-vaccins (0.291 / 0.142) qu'avec les communautés francophones d'Afrique, du Québec et de Belgique ainsi qu'avec celle liée au regroupement familial [6]. Quand on s'intéresse de plus près aux problématiques qui unissent les membres de ces quatre dernières communautés (voir tableau 7), on identifie des sujets très spécifiques, propres aux différentes régions concernées. Par exemple, les hashtags "Covid19Sn", "Covid19Qc" et "Covid19Be" sont utilisés pour parler spécifiquement de la gestion de la crise sanitaire au Sénégal, au Québec et en Belgique. La procédure de regroupement familial, bien que ouverte à tout étranger qui possède un titre de séjour en France, est ici plutôt associée à la région du Maghreb. La spécificité de ces communautés régionales par rapport aux autres est que dans la majeure partie des cas, seuls les membres de ces communautés vont interagir sur ces sujets. Par conséquent, elles se retrouvent moins exposées aux frontières des autres communautés associées à des thématiques plus générales comme les pro et anti-vaccins. Cette isolation thématique est alors confondue avec un comportement antagoniste par la méthode ERIS, qui n'a accès qu'à la structure des interactions, ce qui explique les fortes valeurs d'antagonisme envers ces communautés. Face aux résultats fournis par la méthode, il est donc recommandé d'être à la fois attentif : 1) sur la réelle existence des communautés étudiées d'un point de vue structurel, en vérifiant la modularité et la limite de résolution du graphe (comme présenté dans la section 4.2), et ; 2) sur le fait qu'étudier des communautés très cohésives peut provoquer une confusion entre des comportements antagonistes et une isolation thématique, c'est-à-dire des frontières qui interagissent peu ou pas avec d'autres communautés parce que leurs thématiques sont trop éloignées.

Faire varier la méthode de construction des communautés peut tout de même être un bon moyen pour s'assurer de la présence d'antagonisme et pour faire apparaître de nouveaux comportements à analyser. Si on laisse de côté les communautés régionales en analysant les matrices 7a et 7b, on remarque une nouvelle fois que la communauté la plus susceptible d'être antagoniste avec les anti-vaccins est la com-

6. Pendant la crise sanitaire, la fermeture des frontières en France a mené de nombreuses demandes de visas déposées dans le cadre de la procédure de regroupement familial à être annulées (https ://www.lepoint.fr/politique/regroupement-familial-en-raison-du-covid-le-conseil-d-etat-suspend-la-decision-de-l-executif-22-01-2021-2410696_20.php)

munauté pro-vaccins et *vice versa*. De même, on observe de faibles valeurs de porosité sur les lignes associées à ces communautés, traduisant un renfermement assez important des frontières. On retrouve donc ici les mêmes signes de polarisation que dans la section 4.2. On identifie toutefois aussi un nouveau comportement particulier des communautés vis-à-vis de celle liée aux médias traditionnels. Les frontières des différentes communautés avec celle-ci ont une taille beaucoup plus importante et sont beaucoup plus poreuses, comme le révèlent les colonnes "Reaction to Medias" des matrices 7b et 7c. On remarque ainsi que la communauté liée aux médias joue un rôle assez central au sein du graphe et provoque un déplacement des membres internes des communautés vers les frontières. C'est une information qui n'apparaît pourtant pas clairement sur les résultats de la section 4.2, où les communautés sont formées en se basant uniquement sur des informations structurelles, là où les communautés de G_{Q2} intègrent aussi la sémantique liée au retweet. Le choix de la méthode de détection des communautés a donc un impact important sur les résultats d'ERIS et ne doit donc pas être négligé lors de leur analyse.

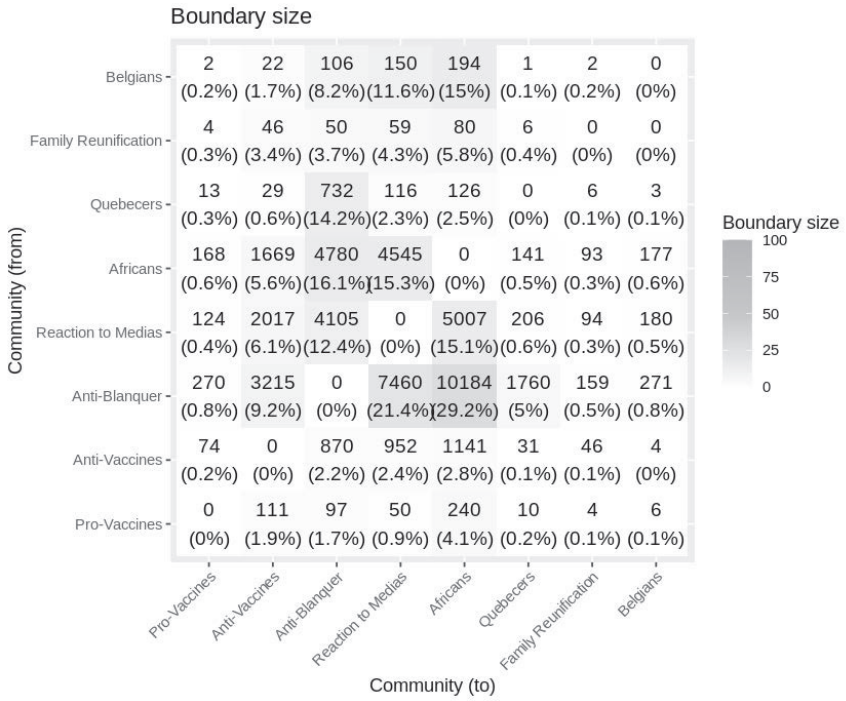
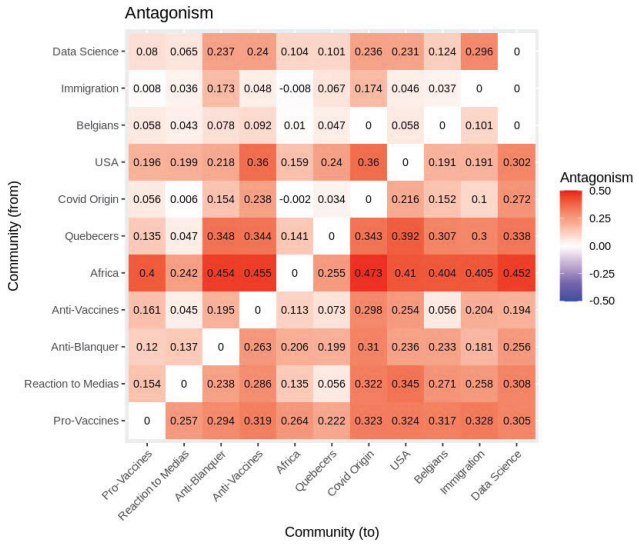


FIGURE 8. *Tailles des frontières des communautés de G_{REP}.*

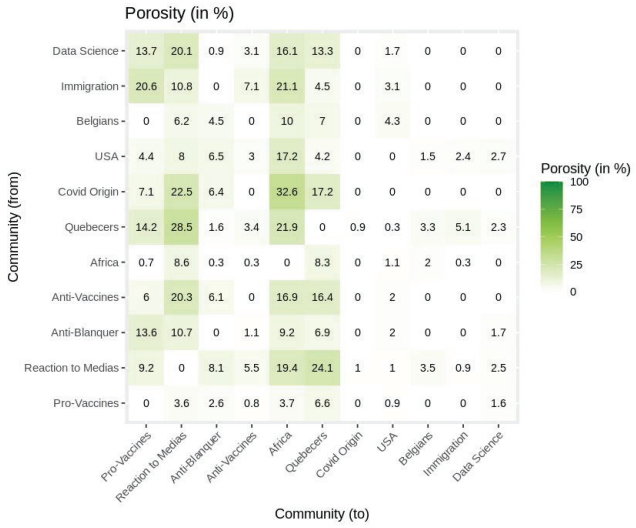
Un autre aspect important à prendre en considération est le type d'interaction choisie. Nous avons construit le graphe G_{REP} présenté dans le tableau 6 en suivant la même méthodologie que pour G_{Q2} mais en choisissant la réponse à un tweet comme méthode d'interaction plutôt que la citation. La réponse à un tweet est une interaction qui peut être vecteur d'antagonisme, mais les réseaux construits à partir de ce type d'interaction sont généralement considérés comme non polarisés (Conover et al. 2011). La modularité peu élevée du graphe montre d'abord que sa structure communautaire n'est pas très bien établie. Si on essaye de tout de même appliquer ERIS sur ce graphe sans calculer les mesures d'antagonisme et de porosité, on obtient des frontières quasi inexistantes qui reflètent le fait que les zones internes des communautés ne sont pas très bien établies (voir figure 8).

Sur le graphe G_{RTC}, construit de la même manière que G_{Q2} et G_{REP} mais en choisissant le retweet comme méthode d'interaction, on obtient une modularité bien plus élevée et donc une structure communautaire plus forte. La matrice 9c fait alors apparaître de plus grosses frontières. Toutefois, le retweet

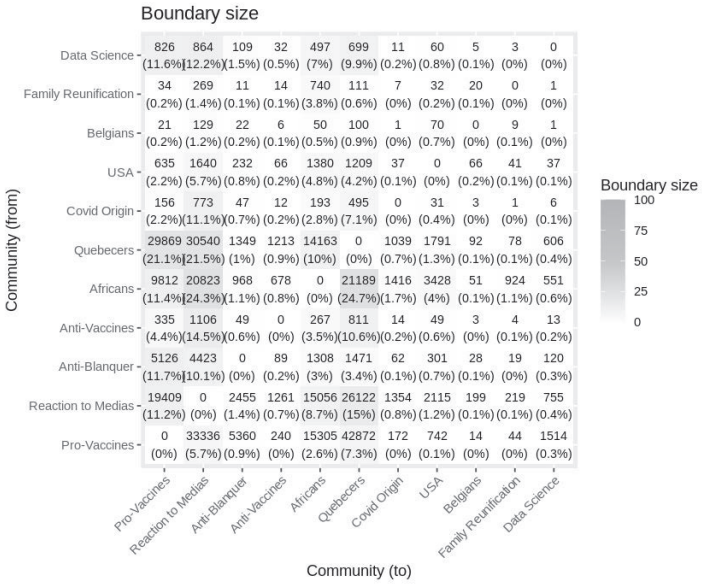
étant une interaction qui implique une forme d'approbation, il serait incohérent de réellement considérer les valeurs de la matrice 9a comme de l'antagonisme. La signification exacte de ces mesures d'un point de vue social sort du cadre de ce travail. D'un point de vue structurel, une forte valeur "d'antagonisme" dans la matrice 9a indique que la frontière partage très peu de contenu en provenance de l'autre communauté de la paire en comparaison du contenu en provenance de la zone interne. La porosité sur ce type de réseau évalue le pourcentage d'individus frontières qui partagent plus de contenu en provenance de l'extérieur que de l'intérieur de la communauté. On peut ainsi remarquer de très fortes valeurs sur les lignes de la matrice 9a associée aux communautés francophones d'Afrique ou du Québec, cohérentes avec les hypothèses précédentes relatives à la spécificité des thématiques. Les valeurs plus faibles associées à la communauté francophone de Belgique peuvent être causées par la très petite taille de ses frontières.



(a) *Matrice d'antagonisme de* G_{RTC}.



(b) *Matrice de porosité de* G_{RTC}.



(c) *Tailles des frontières des communautés de* G_{RTC}.

FIGURE 9. *Résultats retournés par ERIS sur le graphe de retweets nettoyé* G_{RTC}.

Cas limites et pistes de résolution

La méthode ERIS ouvre la voie à une grande variété d'analyses des comportements des communautés sur les réseaux sociaux numériques. Toutefois, l'interprétation de ses mesures doit considérer certains facteurs tels que la méthode de détection des communautés utilisée et la sémantique liée aux différents types d'interactions, qui peuvent avoir un impact sur les résultats à ne pas négliger lors des analyses. L'étude de ces facteurs nous a permis d'identifier certaines limites de la méthode ERIS. Par exemple, certains comportements des frontières structurent le réseau de la même manière qu'un comportement antagoniste et peuvent alors être faussement identifiés comme tel. De même, tous les types d'interactions ne sont pas autant propices à la formation de frontières et ainsi certaines informations peuvent plus ou moins clairement apparaître dans les résultats en fonction de comment sont formées les communautés et du type de réseau.

Pour répondre à ces problématiques, nous prévoyons d'affiner les mesures calculées par ERIS en agissant sur plusieurs leviers dans de futurs travaux. Tout d'abord, nous souhaitons déterminer si notre définition des frontières n'est pas trop stricte et ne passe pas à côté de certains éléments. Il est notamment important de noter que dans la version actuelle d'ERIS, les interactions entre membres frontières ainsi qu'entre les frontières et les utilisateurs qui ne sont pas membres d'une zone interne sont ignorées. Cela exclut des frontières certains utilisateurs qui interagissent avec l'autre communauté de la paire mais pas directement avec la zone interne de leur communauté. L'exclusion de ces utilisateurs des frontières pourrait être à l'origine des tailles très variables des frontières de communautés dans certains cas.

La mesure d'antagonisme pourrait également être affinée. Actuellement, tous les membres des frontières ont le même impact sur la valeur finale d'antagonisme. Or, tous les individus n'ont pas la même influence dans le réseau ou au sein de leur communauté. Cet aspect pourrait être pris en considération dans la méthode en intégrant des mesures de centralité dans le calcul de l'antagonisme, en plus du poids des arêtes et de leur direction.

Nous souhaitons également déterminer le type de réseau le plus propice à l'utilisation de méthodes comme ERIS basées sur l'étude des frontières. Puisque cette dernière notion nécessite l'existence de zones internes fortes à l'intérieur des communautés, il est peut-être plus pertinent de d'abord calculer les communautés sur un graphe auxiliaire où les interactions traduisent une forme d'approbation, puis de transférer la classification sur un graphe où les interactions peuvent être vecteurs d'antagonisme (comme proposé dans cette section) plutôt que de tout faire sur le même graphe (comme proposé dans la section 4.2). Une autre alternative pourrait également être de mélanger les types d'interactions au sein d'un même graphe, en considérant les interactions qui traduisent de l'approbation entre les membres d'une même communauté et celles vecteurs d'antagonisme entre les frontières et les autres communautés.

5. Conclusion

L'analyse des données de réseaux sociaux permet l'extraction de valeur de la masse des données à partir de l'étude des interactions entre individus et entre communautés d'individus. Les discussions et débats autour de thématiques controversées peuvent mener à la polarisation des communautés d'individus, c'est-à-dire leur isolement au sein de groupes renfermés et mutuellement antagonistes.

Dans cet article, nous avons proposé ERIS, une approche non-supervisée pour évaluer la polarisation entre paires de communautés au sein de graphes d'interactions sociales. La méthode analyse le comportement des individus frontières, qui agissent comme intermédiaires entre l'intérieur et l'extérieur de leur communauté, pour calculer deux indicateurs appelés antagonisme des communautés et porosité des frontières.

Notre définition formelle d'ERIS prend en considération trois caractéristiques majeures des graphes construits à partir d'interactions sociales : la pondération, la direction des arêtes et la possible présence de communautés recouvrantes. Nous proposons également un algorithme rapide basé sur des opérations matricielles ainsi qu'une implémentation en R librement accessible en ligne.

Nous avons étudié expérimentalement et théoriquement la complexité en temps de cet algorithme afin de démontrer son applicabilité sur des données massives. Sur un jeu de données réelles lié à la thématique de la vaccination contre la COVID-19, nous avons également détaillé une manière d'exploiter notre méthode pour interpréter les mesures présentées (antagonisme, porosité) et confirmer des hypothèses.

L'évaluation de l'antagonisme et de la porosité dans les réseaux sociaux numériques ouvre la voie à de nombreuses applications, notamment en termes de modération. Les zones antagonistes peuvent par exemple être considérées comme des zones à risques, qui méritent une attention particulière pour éviter les débordements. L'identification des utilisateurs aux frontières permet de mettre en évidence des points d'entrée plus pertinents pour partager de l'information à l'intérieur des communautés, notamment au travers des utilisateurs qui contribuent à la porosité des frontières. Cette information peut ensuite être utilisée pour tenter de diversifier la nature des opinions et des sources d'information qui circulent au sein des communautés avec comme objectif d'affaiblir les "chambres d'écho épistémiques" décrites dans (Donkers & Ziegler 2021, Mønsted & Lehmann 2022), et ainsi par effet de bord la polarisation des différentes communautés. Cette utilisation des mesures pose toutefois quelques questions d'éthique, notamment en cas de détournement à des fins plus malveillantes comme le partage de désinformation ou d'autres comportements de manipulation de l'information tels que décrits dans (Couturier 2022). La trop forte intrusion d'outils de dépolarisation au sein des communautés pourrait également avoir l'effet inverse de celui recherché en provoquant une plus grande aversion des utilisateurs envers la plate-forme. Cela pourrait ultimement mener au déplacement de certaines communautés vers d'autres réseaux sociaux numériques moins modérés et ainsi créer des silos d'utilisateurs qui ne communiquent plus entre eux (Couturier 2022).

Pour mieux cerner le rôle que pourrait jouer la méthode ERIS dans les applications précédemment décrites, nous prévoyons donc d'étudier le lien entre ses mesures et la diffusion de l'information au sein des différentes zones des communautés (internes et frontières) plus en détails dans de futurs travaux. Nous envisageons par exemple d'effectuer des analyses de discours à l'échelle des zones pour pouvoir identifier l'impact des mesures d'antagonisme et de porosité sur l'homogénéité des sujets abordés au sein d'une même communauté. De tels travaux nous permettrons d'obtenir une meilleure perspective sur l'influence des frontières sur la formation et le maintien de la cohésion de leur communauté autour de thématiques communes et donc par conséquent sur le possible rôle qu'elles pourraient jouer à des fins de dépolarisation.

Remerciements

Ce travail est soutenu par le programme « Investissements d'Avenir », projet ISITE-BFC (contrat ANR-15-IDEX-0003). Le projet Cocktail est piloté scientifiquement par Gilles Brachotte, laboratoire CIMEOS EA-4177, Université de Bourgogne.

Bibliographie

Al Amin, M. T., Aggarwal, C., Yao, S., Abdelzaher, T. & Kaplan, L. (2017), Unveiling polarization in social networks : A matrix factorization approach, *in* 'IEEE INFOCOM 2017-IEEE Conference on Computer Communications', IEEE, pp. 1–9.

Alamsyah, A. & Adityawarman, F. (2017), Hybrid sentiment and network analysis of social opinion polarization, *in* '2017 5th International Conference on Information and Communication Technology (ICoIC7)', IEEE, pp. 1–6.

Barabási, A.-L. & Pósfai, M. (2016), *Network science*, Cambridge University Press, Cambridge.
URL: *http ://barabasi.com/networksciencebook/*

Baumann, F., Lorenz-Spreen, P., Sokolov, I. M. & Starnini, M. (2020), 'Modeling echo chambers and polarization dynamics in social networks', *Physical Review Letters* **124**(4), 048301.

Blondel, V. D., Guillaume, J.-L., Lambiotte, R. & Lefebvre, E. (2008), 'Fast unfolding of communities in large networks', *Journal of statistical mechanics : theory and experiment* **2008**(10), P10008.

Boyd, D., Golder, S. & Lotan, G. (2010), Tweet, tweet, retweet : Conversational aspects of retweeting on twitter, *in* '2010 43rd Hawaii international conference on system sciences', IEEE, pp. 1–10.

Cinelli, M., Morales, G. D. F., Galeazzi, A., Quattrociocchi, W. & Starnini, M. (2021), 'The echo chamber effect on social media', *Proceedings of the National Academy of Sciences* **118**(9).

Clauset, A. (2005), 'Finding local community structure in networks', *Physical review E* **72**(2), 026132.

Conover, M., Ratkiewicz, J., Francisco, M., Gonçalves, B., Menczer, F. & Flammini, A. (2011), Political polarization on twitter, *in* 'Proceedings of the International AAAI Conference on Web and Social Media', Vol. 5.

Couturier, B. (2022), 'La démocratie malade des réseaux sociaux', *Constructif* **61**(1), 37–40.

Devi, J. C. & Poovammal, E. (2016), 'An analysis of overlapping community detection algorithms in social networks', *Procedia Computer Science* **89**, 349–358.

Donkers, T. & Ziegler, J. (2021), The Dual Echo Chamber : Modeling Social Media Polarization for Interventional Recommending, *in* 'Fifteenth ACM Conference on Recommender Systems', ACM, Amsterdam Netherlands, pp. 12–22.
URL: *https ://dl.acm.org/doi/10.1145/3460231.3474261*

Ertan, G., Çarkoğlu, A. & Aytaç, S. E. (2022), 'Cognitive political networks : A structural approach to measure political polarization in multiparty systems', *Social Networks* **68**, 118–126.

Fortunato, S. (2010), 'Community detection in graphs', *Physics reports* **486**(3-5), 75–174.

Fortunato, S. & Barthelemy, M. (2007), 'Resolution limit in community detection', *Proceedings of the national academy of sciences* **104**(1), 36–41.

Garimella, K., De Francisci Morales, G., Gionis, A. & Mathioudakis, M. (2018), Political discourse on social media : Echo chambers, gatekeepers, and the price of bipartisanship, *in* 'Proceedings of the 2018 World Wide Web Conference', pp. 913–922.

Garimella, K., Morales, G. D. F., Gionis, A. & Mathioudakis, M. (2018), 'Quantifying controversy on social media', *ACM Transactions on Social Computing* **1**(1), 1–27.

Gillet, A., Leclercq, É. & Cullot, N. (2021), Lambda+, the renewal of the lambda architecture : Category theory to the rescue, *in* 'International Conference on Advanced Information Systems Engineering', Springer, pp. 381–396.

Goldstein, M. L., Morris, S. A. & Yen, G. G. (2004), 'Problems with fitting to the power-law distribution', *The European Physical Journal B-Condensed Matter and Complex Systems* **41**(2), 255–258.

González-Ibánez, R., Muresan, S. & Wacholder, N. (2011), Identifying sarcasm in twitter : a closer look, *in* 'Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics : Human Language Technologies', pp. 581–586.

Guerra, P., Meira Jr, W., Cardie, C. & Kleinberg, R. (2013), A measure of polarization on social media networks based on community boundaries, *in* 'Proceedings of the International AAAI Conference on Web and Social Media', Vol. 7.

Guerra, P., Nalon, R., Assunçao, R. & Meira Jr, W. (2017), Antagonism also flows through retweets : The impact of out-of-context quotes in opinion polarization analysis, *in* 'Proceedings of the International AAAI Conference on Web and Social Media', Vol. 11.

Guyot, A., Gillet, A. & Leclercq, É. (2021), Frontières des communautés polarisées : application à l'étude des théories complotistes autour des vaccins, *in* 'INFormatique des ORganisations et Systèmes d'Information et de Décision'.

Habibi, M. N. & Sunjana (2019), 'Analysis of indonesia politics polarization before 2019 president election using sentiment analysis and social network analysis.', *International Journal of Modern Education & Computer Science* **11**(11).

Isenberg, D. J. (1986), 'Group polarization : A critical review and meta-analysis.', *Journal of personality and social psychology* **50**(6), 1141.

Jain, G., Sreenivas, A. B., Gupta, S. & Tiwari, A. A. (2022), The dynamics of online opinion formation : Polarization around the vaccine development for covid-19, *in* 'Causes and Symptoms of Socio-Cultural Polarization', Springer, pp. 51–72.

Jiang, J., Ren, X. & Ferrara, E. (2021), 'Social media polarization and echo chambers : A case study of covid-19', *arXiv preprint arXiv :2103.10979* .

Joshi, A., Bhattacharyya, P. & Carman, M. J. (2017), 'Automatic sarcasm detection : A survey', *ACM Computing Surveys (CSUR)* **50**(5), 1–22.

McGlone, M. S. (2005), ‘Contextomy : the art of quoting out of context’, *Media, Culture & Society* **27**(4), 511–522.

Mønsted, B. & Lehmann, S. (2022), ‘Characterizing polarization in online vaccine discourse—a large-scale study’, *PloS one* **17**(2), e0263746.

Morales, A. J., Borondo, J., Losada, J. C. & Benito, R. M. (2015), ‘Measuring political polarization : Twitter shows the two sides of venezuela’, *Chaos : An Interdisciplinary Journal of Nonlinear Science* **25**(3), 033114.

Newman, M. E. (2006), ‘Modularity and community structure in networks’, *Proceedings of the national academy of sciences* **103**(23), 8577–8582.

Saini, M. & Mangat, V. (2023), ‘Multidimensional empirical analysis of overlapping community detection methods in social networks’, *Multimedia Tools and Applications* pp. 1–17.

Xie, J., Kelley, S. & Szymanski, B. K. (2013), ‘Overlapping community detection in networks : The state-of-the-art and comparative study’, *Acm computing surveys (csur)* **45**(4), 1–35.

