HAL open science

ERIS: an Approach Based on Community Boundaries to Assess Polarization in Online Social Networks

Alexis Guyot, Annabelle Gillet, Éric Leclercq, Nadine Cullot

To cite this version:

Alexis Guyot, Annabelle Gillet, Éric Leclercq, Nadine Cullot. ERIS: an Approach Based on Community Boundaries to Assess Polarization in Online Social Networks. Research Challenges in Information Science. RCIS 2022, May 2022, Barcelone, Spain. hal-03889719

HAL Id: hal-03889719

https://hal.science/hal-03889719v1

Submitted on 8 Dec 2022

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

ERIS: an Approach Based on Community Boundaries to Assess Polarization in Online Social Networks

Alexis Guyot, Annabelle Gillet, Éric Leclercq, and Nadine Cullot

Laboratoire d'Informatique de Bourgogne - EA 7534
University of Bourgogne-Franche-Comté
Dijon, France
alexis.guyot@u-bourgogne.fr
annabelle.gillet@u-bourgogne.fr
eric.leclercq@u-bourgogne.fr
nadine.cullot@u-bourgogne.fr

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. Our experiments show a significant improvement in terms of efficiency in comparison to existing solutions. Finally, we apply our proposal on real data harvested from Twitter with a case study about the vaccines and the COVID-19.

Keywords: social networks, polarization, community boundaries, community structure, graph mining

1 Introduction

Online Social Networks (OSN) are large scale environments of exchanges and debates. Social Network Analysis (SNA) has diverse and numerous applications in domains such as sociology, politics, marketing, health, etc. The intrinsic characteristics of the large volume of data generated in OSN [3], such as the power law distribution, entail analysts to use algorithmic approaches to extract value.

SNA can benefit from the graph theory since graph structures are natural representations for OSN, where users can be represented by vertices and their interactions by edges. Communities of individuals form dense areas of nodes and can therefore be detected by algorithms like Louvain, Walktrap or Infomap

for non-overlapping communities and SLPA, OSLOM or Game for overlapping ones [11].

Discussions about hot topics can lead to the creation of mutually antagonistic communities, with few individuals remaining neutral or holding an intermediate position. In social sciences, this phenomenon is called polarization [21]. Polarized communities negatively impact OSN by fostering social division, ideological isolation and misinformation spreading [4]. Detecting such communities is of major interest to proactively assist moderation and therefore avoid further escalation between users. Journalists could also benefit from this feature to identify areas in the network where fact checking could be needed. Moreover, detecting polarization allows a more precise understanding of individuals through their relationships. Domains such as politics or business intelligence could benefit from it to adapt their decision making and communication strategies.

In the literature, echo chambers are usually considered as the consequence of polarization [7]. However, only showing that a community is an echo chamber does not allow to conclude about its polarization. An echo chamber is a configuration in which one is exposed only to opinions that agree with their own [12]. So this phenomenon describes a global behavior within a community whereas polarization is also about relationships between communities [4]. Thus, the polarization is also carried by community members exposed to other communities and exposing the main topic of their community to the outside through their interactions. These members form the community boundaries. More formally, we can define a community boundary as the set of nodes having edges directed toward both the inside and the outside of the community [8]. In the literature, boundaries are fairly unexplored parts of communities. Nevertheless, the behaviors of boundary users have a significant impact on the strength of the polarization but also on the fragility of the echo chamber [10] as they contribute to the porosity of the boundary.

The major contributions of our work are: 1) a formal graph approach relying on community boundaries to unveil the polarization of networks created from the interactions between individuals; 2) two metrics to characterize the level of polarization and the porosity of boundaries; 3) an efficient algorithm based on matrix computations suitable for large volumes of data, and; 4) a case study on real data extracted from Twitter to experimentally validate our proposal.

The remainder of this paper is structured as follows. First, our method is positioned in relation to the state of the art in section 2. In section 3 we formally define the ERIS approach and propose an efficient algorithm to compute polarization metrics. The case study led on real data and validated by domain experts is described in section 4. Finally, we draw conclusions of our work and open up perspectives for the future in section 6.

2 Related Work

The problem of polarization in OSN was addressed back in 2011 [9]. The authors consider that echo chambers and polarized communities are the same.

But with this assumption, interactions and relationships between communities, carried by community boundaries, are ignored. Moreover, the approach is an applied methodology which cannot be included in an automatic analytical workflow ready-to-use for domain experts.

Many works on social polarization use exploratory analyzes combining metrics from the graph theory with interpretations provided by Natural Language Processing (NLP) tools like sentiment analysis based on Naïves Bayes [2,20] or sentence embedding models based on Retweet-BERT [22]. These approaches best capture the semantics of discussions but require heavy involvement from the analyst, especially during the preprocessing step. Indeed, to set the stage for the NLP algorithms, many difficulties must be manually addressed like spelling approximations, abbreviations, slang, or ambiguities caused by humor, sarcasm or irony as discussed in [16,23,24]. Thus, they leave a large area for subjectivity and lack automatism.

Other approaches focus on the network structure with weakly-supervised strategies where just a limited amount of extra knowledge is used to initialize algorithms. In [25] an opinion score is manually assigned to seed users (*elites*) and then propagated to the other nodes of the network (*listeners*) in order to create two opposing groups and to assess their polarization degree. In [1], a similarity measure must be wisely chosen to create clusters of tweets (*assertions*) with the aim of unveiling polarized groups inside a network. To do so, they use a matrix factorization and an ensemble based gradient descent algorithm applied on the adjacency matrices of a bipartite source-assertion graph and a social influence graph. In any case the relevance of the results depends a lot on the extra knowledge brought, which must be revised for each dataset studied. Therefore, their automatism is limited. Moreover, they do not consider the relationships between polarized communities, meaning that two communities behaving as echo chambers but never interacting because they do not know each other could be considered as polarized.

Boundaries have a major impact on the polarization of their community by defining both how the community is exposed to the outside and how the outside is exposed to the community. A first non-supervised approach based on community boundaries was described by Guerra et al. in [17] with the aim of computing complementary metrics to be used besides cohesion and homophily metrics such as modularity. Antagonism between communities is assessed by measuring the involvement of users interacting with both the inside and the outside of their community. This approach does not need any *a priori* knowledge on the graph or on the individuals represented and can therefore be included in automatic analytical workflows designed for domain experts. However, the main limitation of this method is its specification on undirected and unweighted graphs whereas social interaction graphs usually are directed and weighted. Furthermore, only non-overlapping communities are handled whereas users of social networks more naturally belong to multiple communities [26].

As a conclusion, fully automatic methods are the best option to detect polarization. This property is indeed very important in SNA to allow domain experts

like sociologists or decision-makers to use a method and to permit comparisons between datasets. Furthermore, the polarization of a community can be misinterpreted when interactions between communities are not considered, which can be avoided by examining the behavior of community boundaries.

3 The ERIS Method

In this section, we formally introduce the ERIS method and its metrics, *i.e.*, the *community antagonism* and the *porosity of boundaries*. Our approach relies on edge weighting and direction and handles overlapping communities. We also propose an efficient algorithm based on matrix computations to assess the metrics.

3.1 Formal Definitions

In the following definitions, a graph $G = (V, E)$ is composed of a set of vertices V and of a set of directed edges $E \subseteq V \times V$. An edge $e_{a,b} \in E$ connects a source $a \in V$ and a destination $b \in V$ with a weight $w(e_{a,b}) \in \mathbb{R}$. Communities are locally dense connected subsets of V.

Two communities (C_i, C_j) are polarized if they are mutually antagonistic. According to [17] and [4], a strong involvement from a boundary individual within the community, especially expressed by numerous interactions with the internal members, reveals a substantial emotional attachment to the community and its main topics. This attachment could easily lead to the expression of antagonism in response to a criticism, an attack or the broadcast of a negative opinion or information about these topics by another community.

The ERIS method consists in identifying, for each pair of communities (C_i, C_j): 1) the internal area $I_{i,j}$ of C_i, that is the set of vertices in C_i without any edge directed toward C_j, and; 2) the boundary area $B_{i,j}$ of C_i, that is the set of vertices in C_i with at least one edge directed toward $I_{i,j}$ and another one toward C_j. The method assesses the average antagonism expressed by the community C_i to the community C_j by measuring the involvement of the vertices in B_{ij}.

From the previous intuitive descriptions, we have established the following formal definitions:

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

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

For each boundary, we consider the set of outgoing edges directed toward the other community (*external edges* or $EE_{i,j}$) as well as the set of edges directed toward the internal area $I_{i,j}$ of C_i (*internal edges* or $IE_{i,j}$):

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

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

We also consider $EE^v_{i,j}$ the external and $IE^v_{i,j}$ the internal edges of a vertex v as the subsets of edges, respectively included in $EE_{i,j}$ and $EI_{i,j}$, where v is the source of the edge:

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

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

The antagonism $A^v_{i,j}$ expressed by a vertex v is assessed as the weighted ratio of its internal edges' weights with the sum of its internal and external edges' weights. This value is compared to a null hypothesis, *i.e.*, each node spreads its edges equally between internal nodes and nodes from the other community [17]:

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

Finally, the antagonism $A_{i,j}$ expressed by a boundary $B_{i,j}$ is the average antagonism expressed by its members:

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

By assessing the antagonism values for each possible pair of communities in a graph, we obtain an asymmetrical matrix called the *antagonism matrix*, containing values ranging from -0.5 to 0.5. A community boundary with a value close to 0.5 should be considered as likely to be antagonistic toward the other community of the pair. Values on the lines of the antagonism matrix express how much the community heading the line is likely to express antagonism toward the communities heading the columns. Conversely, values on the columns indicate how much the community heading the column is likely to receive antagonism from the communities heading the lines.

Boundary vertices with negative antagonism values weaken the polarization of their community. Indeed, by interacting more with the outside than with the inside, they reduce the isolation of the community that leads to the creation of an echo chamber. As fellow members, they also seem more credible in the eyes of the others when they share more nuanced opinions about the main topics of the community [10]. Based on these ascertainments, we propose a novel metrics $P_{i,j}$ called the *porosity* of the boundary $B_{i,j}$, measuring the fragility of the boundary of C_i with C_j:

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

with $NB_{i,j} = \left\{v : v \in B_{i,j}, A^v_{i,j} < 0\right\}$ the subset of $B_{i,j}$ including all the vertices having negative antagonism values. Porosity values also can be represented inside an asymmetrical matrix called the *porosity matrix*.

We now illustrate the different sets and values presented in this subsection by applying the definitions on the toy example of figure 1. We focus our explanation

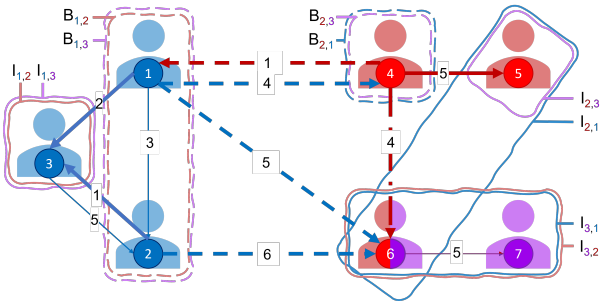
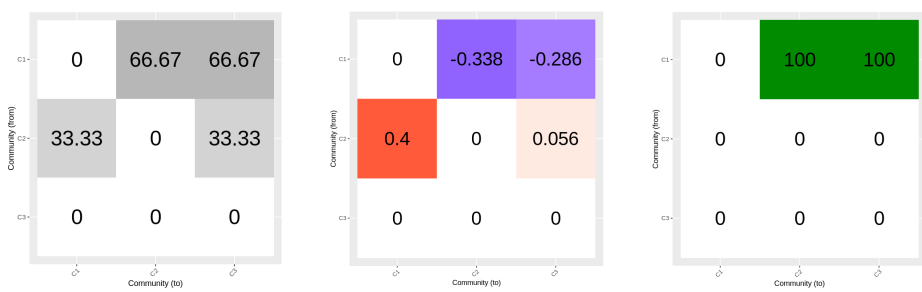


Fig. 1: Toy example with 3 communities C_1 (blue), C_2 (red) and C_3 (purple). Communities C_2 and C_3 overlap on vertex 6. For edges, solid lines are internal edges, dotted lines are external edges, thin lines are edges neither internal nor external. Note that $e_{4,6}$ is both internal and external. For areas, internal areas are surrounded by solid lines, boundary areas by dotted lines.

on the community C_1. Both internal areas of C_1 with C_2 and C_3 include the same vertices, that is $I_{1,2} = I_{1,3} = \{3\}$. The same observation can be made with its boundary areas, $B_{1,2} = B_{1,3} = \{1, 2\}$. External and internal edges of the pair (C_1, C_2) are $EE_{1,2} = \{e_{1,4}, e_{1,6}, e_{2,6}\}$ and $IE_{1,2} = \{e_{1,3}, e_{2,3}\}$.



(a) Size of boundaries (in %)

(b) Antagonism matrix

(c) Porosity matrix (in %)

Fig. 2: Values calculated from the toy example

Figure 2 shows the antagonism and the porosity matrices obtained on the graph of the toy example. Figure 2a expresses the sizes of the different boundary areas as a percentage of community members belonging to the boundary. In figure 2b, the antagonism value $A_{1,2}$ expressed by the boundary of the community C_1 toward the community C_2 is equal to -0.338 and therefore does not reveal an antagonistic behavior. However, the boundary of the community C_2 is pretty likely to be antagonistic toward the community C_1 since its antagonism value $A_{2,1}$ is equal to 0.4. The matrix of figure 2a reveals that all the values equal to 0 in the antagonism matrix are default values resulting from empty boundaries. In figure 2c, the porosity value $P_{1,2}$ is equal to 100, meaning that $B_{1,2}$ is very porous as its members interact more with C_2 than with $I_{1,2}$.

3.2 Algorithm

The adjacency matrix of a graph is at the core of the algorithm designed for ERIS, which uses a series of matrix computations to get the antagonism (M_{ANT}) and the porosity (M_{POR}) matrices. Naming conventions used in following paragraphs are listed in table 1. We define ✦ as an operator computing element-wise multiplication between a vector of size N and each column of a matrix of size $N \times M$, resulting in a new matrix of size $N \times M$.

Symbol	Definition	Symbol	Definition
G	The graph to analyze	C	The set of communities in G
V	The set of vertices in G	$\|V\|$	The number of vertices in G
E	The set of edges in G	$\|C\|$	The number of communities in G

Table 1: Naming Conventions

The inputs of the algorithm are the adjacency matrix M_A of G and a community membership matrix called M_C. M_A is a square matrix of size $|V| \times |V|$ containing in each cell the weight of the edge whose source is the vertex heading the row and the destination is the vertex heading the column. M_C is a binary matrix of size $|V| \times |C|$ in which the value 1 means that the vertex heading the row belongs to the community heading the column, and 0 if not.

Symbol	Size	Type	Name
M_A	$\|V\| \times \|V\|$	Int	Adjacency Matrix
M_C	$\|V\| \times \|C\|$	Bin	Community Membership Matrix
M_{EE}	$\|V\| \times \|C\|$	Int	External Edges Weight Matrix
M_I	$\|V\| \times \|C\|$	Bin	Internals Matrix
M_{IC}	$\|V\| \times \|C\|$	Bin	Current Internals Matrix
M_{IE}	$\|V\| \times \|C\|$	Int	Internal Edges Weight Matrix
M_{BIE}	$\|V\| \times \|C\|$	Bin	Binary Internal Edges Matrix
M_{VANT}	$\|V\| \times \|C\|$	Real	Vertices Antagonism Matrix
M_{ANT}	$\|C\| \times \|C\|$	Real	Antagonism Matrix
M_{POR}	$\|C\| \times \|C\|$	Real	Porosity Matrix

Table 2: Matrices used in Algorithm 1

The initialization part of the algorithm consists in computing M_{EE}, an aggregated version of M_A grouped by community (line 1). The matrix contains the sum of the edges' weights whose source is the vertex heading the row and the destination is a vertex belonging to the community heading the column. This matrix is then used to extract M_I, a binary mask of M_{EE} in which the vertices belonging to at least one internal area of their communities are identified (line 2).

From these two common matrices, the main part of the algorithm computes for each community the antagonism and porosity values of its boundaries through four main steps:

- the detection of the internal areas of the current community c for each pair involving c (line 4) ;

Algorithm 1 Matrix computations to assess the metrics of 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**

- the aggregation of M_A to sum the weights of the edges directed toward the internal areas of c (line 5) ;
- the computation of the antagonism values for the vertices belonging to the boundaries of c (line 6) ;
- the computation of the antagonism and porosity values for the boundaries of c (lines 8-9).

An open source implementation of this algorithm in R is available on GitHub[1] to allow the use of ERIS on graphs built from real datasets.

4 Experimentations

We want to experimentally show the suitability of our method on real data from OSN. First, we verify its applicability on large graphs through an analysis of the algorithmic complexity in time achieved by our matrix computation based algorithm. Then, we explore the validity of our polarization metrics through a case study led on real data harvested from Twitter with the help of domain experts validating our results and interpretations.

4.1 Execution on Large Graphs

We compare the execution times of 3 algorithms aiming to measure the polarization of communities in graphs built from interactions between individuals:

- the matrix computation based algorithm of ERIS presented in the previous section (implemented in R) ;
- an iterative algorithm of ERIS proposed in a previous work [19] (implemented in R) ;

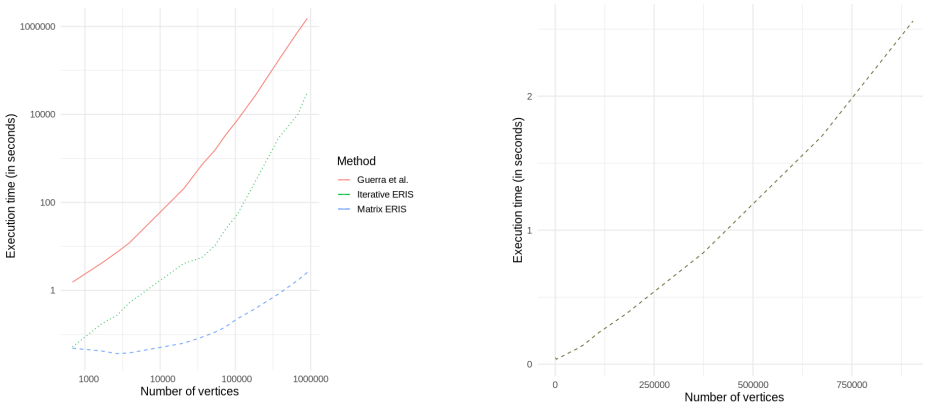
[1] https://github.com/AlexisGuyot/ERIS

– the only algorithm of Guerra et al.'s method available online[2], not developed by the authors (implemented in Python).

We chose to compare ERIS with Guerra et al.'s method because both share a lot of common characteristics (no supervision needed, based on graph mining, etc.).

We have generated artificial graphs with decreasing sizes, ranging from 1 million to 500 vertices, based on a real graph extracted from a dataset harvested from Twitter. In this subsection, we do not take into consideration the semantics of the computed metrics, only the impact of the graph structure on the execution time.

For each algorithm, we have measured the elapsed time between the call of the function computing the antagonism matrices (the common metrics) and the return of a result[3]. For the algorithm 1, this interval corresponds to lines 1 to 10. The three algorithms were run on a Dell PowerEdge R440 server with the following characteristics: Intel(R) Xeon(R) Bronze 3204 CPU @ 1.92GHz, 6 cores, 128Go RAM.



(a) Comparison between the execution times of the 3 methods (log-log scale)

(b) Focus on the execution times of the matrix computation based algorithm

Fig. 3: Execution times of the algorithms

Execution times of the three algorithms are compared on figure 3a. Figure 3b focuses on the execution times of the algorithm 1 described in the last section. We can see that the matrix computation based algorithm of ERIS outperforms all the other implementations. For our biggest graph, the one with 1 million vertices directly extracted from the real corpus that we have harvested from Twitter, the matrix computation based version of ERIS took 2.5 seconds to compute the metrics. It is 12,828 times faster than the iterative version (32,070

[2] https://github.com/rachel-bastos/boundaries-polarization

[3] See https://github.com/AlexisGuyot/ERIS/tree/main/experiment_complexity for more detailed explanations on the experiment.

seconds or almost 9 hours) and 592,399 times faster than the algorithm of Guerra et al.'s method (1,528,389 seconds or more than 17 days).

Theoretically, the computational complexity for assessing both polarization metrics with the algorithm based on matrix computations is $\mathcal{O}(|V|^2|C|^2)$, as long as $|C| < \sqrt{\frac{|V|}{3}}$. Beyond this value, the order reaches $\mathcal{O}(|V||C|^3)$. However, in most of practical analyzes, the number of significant communities in a graph remains relatively small due to the resolution limit. Furthermore, domain experts also only require a small number of communities to left results open to interpretation. In these cases, $|C| \ll |V|$ and thus the computational complexity can be considered as $\mathcal{O}(|V|^2)$.

According to the previous theoretical and practical analyzes of the algorithmic complexity of our proposal, we can conclude that the matrix computation based algorithm of ERIS achieves our goals of applicability on large graphs and outperforms of several orders of magnitude the other algorithms available online to automatically assess polarization on graphs extracted from OSN.

4.2 Case Study on Real Data

We experimentally illustrate the interest of our approach on a real dataset about COVID-19 vaccines, which includes more than 18 millions tweets harvested in the context of the interdisciplinary project Cocktail[4] by the architecture Hydre [14] from December 1, 2020 to March 31, 2021 (120 days).

From this dataset, a directed graph of quotes[5] G_Q is extracted, in which the vertex representing an individual u has an outgoing edge directed toward the vertex representing the user n if u has already quoted at least twice the tweets of n. The weight w of an edge indicates the exact number of times u quoted n. Following [13], we do not consider isolated quotes as they can be random noise. The characteristics of G_Q are presented in table 3.

Vertices count	24,591
Edges count	55,703
Average Strength	4.46
Diameter	338
Power Law Exponent γ	2.27
Resolution Limit	333
Significant Community count	8
Modularity	0.59

Table 3: Characteristics of G_Q

We chose the quote as type of interaction to be consistent with the previous works led on polarization on Twitter. Indeed, the literature mainly agrees that retweets often imply endorsement [6] and thus not antagonism, and that the

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

[5] Quotes are retweets with additional comments.

mention network is usually not polarized [9]. However, quotes are often used to twist a message out of its original context for humor and criticism purposes, leading to antagonistic responses [18]. Thus, quotes are the best type of interactions for community boundary approaches like ERIS to assess polarization.

G_Q is a scale-free network as the degree distribution of its vertices follows a power law of exponent $2 < 2.27 < 3$ [3]. As a result, modularity can be computed and therefore community detection algorithms based on the optimization of modularity, like Louvain [5], can be run. On G_Q, the previous algorithm revealed 8 significant communities, *i.e.*, having a largest size than the resolution limit of the graph [15] (table 3). The overall modularity of G_Q is 0.59.

To better understand the communities and their relationships, the domain experts of the interdisciplinary project have manually assigned to each community a label related to its main topics by analyzing, for each, its 30 most used hashtags (top-hashtags). This labeling step revealed that the two biggest communities of G_Q gather respectively pro and anti-vaccine individuals. Table 4 lists the elements among the 30 top-hashtags of these last two communities used to infer the labels.

Community	**Top-hashtags (translated from French)**
Pro-vaccines	Mutation, Lockdown3, Curfew, Schools, DigitalGreenCertificate, HealthDictatorship, Israel, IGetVaccinated, Pasteur
Anti-vaccines	Ivermectine, HealthDictatorship, IWillNotConfineMyself, Raoult, Hydroxychloroquine, AndTheTreatment, Plandemic, VeranResignation, TheStonesWillCryOut, GreatReset, Ethics, BeBraveWHO, IWillNotGetVaccinated

Table 4: Top-hashtags highlighting the main topics of the pro and anti-vaccine communities

Since these two main topics are opposite, we expect polarization between the communities. Thus, the anti and pro-vaccine communities should be cohesive and closed communities and their mutual relationship should be antagonistic. To experimentally confirm this expectation, we apply our implementation of the algorithm 1 on G_Q. The computed results are shown in figures 5 and 6. Figure 4 gives supplementary information about the sizes of the boundaries.

Values on the lines of the antagonism matrix (figure 5) express how much the community heading the line is likely to express antagonism toward the communities heading the columns. Conversely, values on the columns indicate how much the community heading the column is likely to receive antagonism from the communities heading the lines.

Columns related to the pro and anti-vaccine communities show that both do not receive much antagonism from the other communities. The community the most likely to be antagonistic with the pro-vaccine community is the anti-vaccine community (0.278) and *vice versa* (0.152). Lines related to these two communities show however that both are pretty likely to have antagonistic behaviors with all the communities. Two hypotheses might explain the lower values between

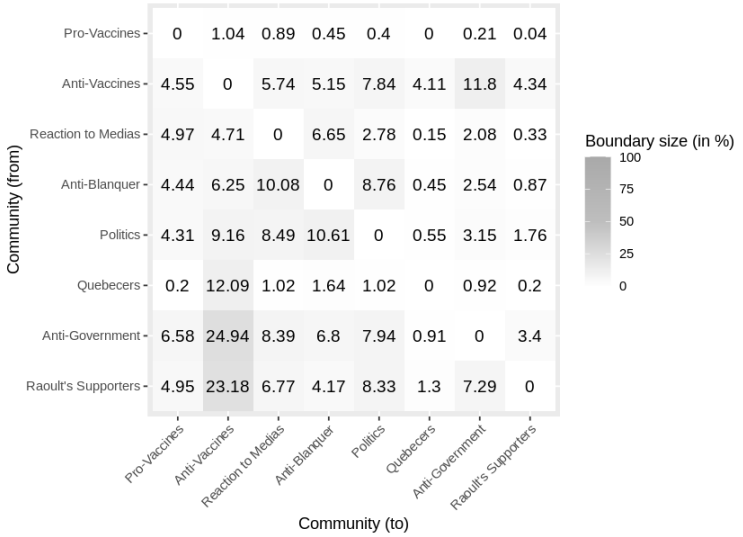


Fig. 4: Size of boundaries of G_Q

the two communities in comparison with the others: 1) the anti and pro-vaccine boundaries are not very antagonistic with each other ; 2) the lively debates between these communities lead some boundary members to communicate more with the outside than with the inside.

The matrix representing the porosity of boundaries (figure 6) allows to decide between the two previous hypotheses and shows that the second one is the more likely. Indeed, we see that 10% of the boundary members of the pro-vaccine community interact more with the anti-vaccine community than with the core members of their own community. For the anti-vaccine community, the equivalent value is around 5%. Thus, these contributions to the debates cause the decrease of the antagonism values for both boundaries. A possible interpretation for this observation is a need for these boundary users to convince their opponents to change their mind.

A deeper understanding of the behavior and roles of these two communities can also be achieved through the lines and columns of the porosity matrix. First, from a broader perspective, we can see that the values on the lines related to both communities are pretty low in comparison with the other ones, meaning that their boundaries do not interact much with the outside. So, anti and pro-vaccine communities are fairly closed communities. Furthermore, on the line related to the anti-vaccine community only, we can see that, even if the values are low, all the boundaries are nearly as porous. Therefore, the anti-vaccine community is almost equally exposed everywhere, which could reveal an additional need to control the debate and the image of the community.

Columns related to these two communities show that they both have a significant impact on the porosity of the other communities, pointing out the general interest of the individuals forming our corpus for the vaccination topic. Higher

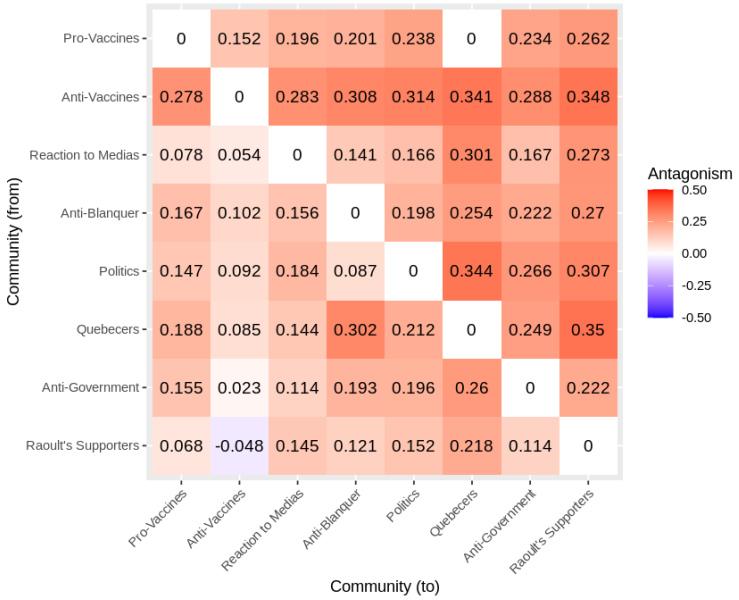


Fig. 5: Antagonism Matrix of G_Q

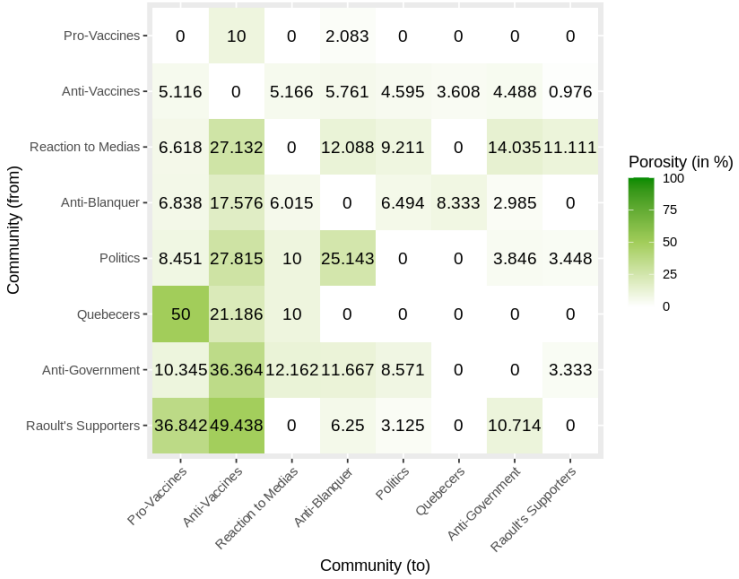


Fig. 6: Porosity Matrix of G_Q

values in the column related to the anti-vaccine community reveal a trend for this community to trigger a lot of reactions from the boundaries of the other communities. Because we are working with quotes, these reactions could be sarcastic, ironic or humorous, and therefore rather negative.

In brief, the metrics of the ERIS method describe a relationship likely to be antagonistic between two fairly closed communities. From this observation, we can conclude that, as expected, the pro and anti-vaccine communities are polarized in our corpus. The ERIS method successfully highlights the traces of polarization within a graph built from interactions between individuals in OSN.

5 Discussions

In this section, we discuss about threats to validity for our method and ideas to improve the interpretability of our metrics.

Before using ERIS, a well informed user should be careful about some points that could threaten its interpretations. First, one should make sure of the objectivity of the data harvesting process, to avoid bias in the data. The longer the time period covered by the dataset is, the more the information is diluted because of higher probabilities to find random or one-time interactions between individuals. There also may be a shift of attention to subjects. ERIS needs well defined communities, so the user should pay a close attention to the chosen algorithm. For example, the modularity of the graph should be high enough to ensure the realness of detected communities with methods like Louvain. Some communities also could be too small to be significant, so the resolution limit of the graph should be respected. Finally, some relationships do not carry antagonism. For example, sharing features like retweets usually imply endorsement. The user should therefore be well aware of the usual meaning of the chosen interaction before drawing any conclusion with the computed metrics.

If the results are not threatened by the previous points, further analytics can be led to achieve a better understanding of the detected polarization. First of all, keywords or hashtags can be used to characterize the source of the disagreement, for example by looking at the main topics of the boundaries. The impact of the different boundary users on the overall polarization of their community can be further investigated by computing their centrality inside the community and inside the whole graph. Finally, the detected polarization could be contextualized in time to gain insights on its first appearance and its evolution.

6 Conclusion

Social Network Analysis allows the extraction of value from interactions between individuals and communities of individuals. Discussions and debates about controversial topics can lead to the polarization of the communities of individuals, *i.e.*, their isolation inside closed and mutually antagonistic groups.

In this article, we propose ERIS, an automatic approach to assess polarization between pairs of communities inside graphs built from social interactions. The

method analyzes the behavior of community boundaries, individuals acting as intermediaries between the inside and the outside of their community, to compute two metrics called the *community antagonism* and the *porosity of boundaries*.

Our formal definition of ERIS takes into consideration three major characteristics of graphs built from social interactions: the weighting, the edge direction and the possible presence of overlapping communities. We also propose an efficient algorithm based on matrix computations as well as an open source implementation in R freely available online.

By allowing a more precise description of the roles inside communities through the concepts of internal and boundary areas, the method could also be used to achieve several other objectives in future works. For example, a possible evolution of ERIS could improve discourse analyzes by exploiting these areas to comment the diffusion of topics from and toward the outside of a community.

Finally, boundary members of communities were identified as key elements to get rid of ideological echo chambers, created from the rejection of contradictory opinions [10]. As the ERIS method allows to detect both polarized pairs of communities and the individuals leading to the porosity of boundaries, we would like to investigate how ERIS could be used to build a tool to favor the depolarization of some targeted OSN communities.

Acknowledgment

This work is supported by ISITE-BFC (ANR-15-IDEX-0003) coordinated by G. Brachotte, CIMEOS Laboratory (EA 4177), University of Burgundy.

References

1. Al Amin, M.T., Aggarwal, C., Yao, S., Abdelzaher, T., Kaplan, L.: Unveiling polarization in social networks: A matrix factorization approach. In: IEEE INFOCOM 2017-IEEE Conference on Computer Communications. pp. 1–9. IEEE (2017)
2. Alamsyah, A., Adityawarman, F.: Hybrid sentiment and network analysis of social opinion polarization. In: 2017 5th International Conference on Information and Communication Technology (ICoIC7). pp. 1–6. IEEE (2017)
3. Barabási, A.L., Pósfai, M.: Network science. Cambridge University Press, Cambridge (2016)
4. Baumann, F., Lorenz-Spreen, P., Sokolov, I.M., Starnini, M.: Modeling echo chambers and polarization dynamics in social networks. Physical Review Letters **124**(4), 048301 (2020)
5. Blondel, V.D., Guillaume, J.L., Lambiotte, R., Lefebvre, E.: Fast unfolding of communities in large networks. Journal of statistical mechanics: theory and experiment **2008**(10), P10008 (2008)
6. Boyd, D., Golder, S., Lotan, G.: Tweet, tweet, retweet: Conversational aspects of retweeting on twitter. In: 2010 43rd Hawaii international conference on system sciences. pp. 1–10. IEEE (2010)
7. Cinelli, M., Morales, G.D.F., Galeazzi, A., Quattrociocchi, W., Starnini, M.: The echo chamber effect on social media. Proceedings of the National Academy of Sciences **118**(9) (2021)

8. Clauset, A.: Finding local community structure in networks. Physical review E **72**(2), 026132 (2005)
9. Conover, M., Ratkiewicz, J., Francisco, M., Gonçalves, B., Menczer, F., Flammini, A.: Political polarization on twitter. In: Proceedings of the International AAAI Conference on Web and Social Media. vol. 5 (2011)
10. Donkers, T., Ziegler, J.: The Dual Echo Chamber: Modeling Social Media Polarization for Interventional Recommending. In: Fifteenth ACM Conference on Recommender Systems. pp. 12–22. ACM, Amsterdam Netherlands (Sep 2021). https://doi.org/10.1145/3460231.3474261, https://dl.acm.org/doi/10.1145/3460231.3474261
11. Fortunato, S.: Community detection in graphs. Physics reports **486**(3-5), 75–174 (2010)
12. Garimella, K., De Francisci Morales, G., Gionis, A., Mathioudakis, M.: 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 (2018)
13. Garimella, K., Morales, G.D.F., Gionis, A., Mathioudakis, M.: Quantifying controversy on social media. ACM Transactions on Social Computing **1**(1), 1–27 (2018)
14. Gillet, A., Leclercq, É., Cullot, N.: Évolution et formalisation de la Lambda Architecture pour des analyses à hautes performances — Application aux données de Twitter. Revue ouverte d'ingénierie des systèmes d'information (ISI) (Numéro 1), 1–26 (2021)
15. Goldstein, M.L., Morris, S.A., Yen, G.G.: Problems with fitting to the power-law distribution. The European Physical Journal B-Condensed Matter and Complex Systems **41**(2), 255–258 (2004)
16. González-Ibánez, R., Muresan, S., Wacholder, N.: 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 (2011)
17. Guerra, P., Meira Jr, W., Cardie, C., Kleinberg, R.: 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 (2013)
18. Guerra, P., Nalon, R., Assunçao, R., Meira Jr, W.: 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 (2017)
19. Guyot, A., Gillet, A., Leclercq, É.: Frontières des communautés polarisées: application à l'étude des théories complotistes autour des vaccins
20. Habibi, M.N., Sunjana: 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) (2019)
21. Isenberg, D.J.: Group polarization: A critical review and meta-analysis. Journal of personality and social psychology **50**(6), 1141 (1986)
22. Jiang, J., Ren, X., Ferrara, E.: Social media polarization and echo chambers: A case study of covid-19. arXiv preprint arXiv:2103.10979 (2021)
23. Joshi, A., Bhattacharyya, P., Carman, M.J.: Automatic sarcasm detection: A survey. ACM Computing Surveys (CSUR) **50**(5), 1–22 (2017)
24. McGlone, M.S.: Contextomy: the art of quoting out of context. Media, Culture & Society **27**(4), 511–522 (2005)
25. Morales, A.J., Borondo, J., Losada, J.C., Benito, R.M.: Measuring political polarization: Twitter shows the two sides of venezuela. Chaos: An Interdisciplinary Journal of Nonlinear Science **25**(3), 033114 (2015)

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