

RESEARCH ARTICLE

Anatomy of elite and mass polarization in social networks

Ali Salloum¹ , Ted Hsuan Yun Chen²  and Mikko Kivelä¹

¹Department of Computer Science, Aalto University, Espoo, Finland and ²Department of Environmental Science and Policy, George Mason University, Fairfax, VA, USA

Corresponding author: Ali Salloum; Email: ali.salloum@aalto.fi

Abstract

Political polarization is a group phenomenon in which opposing factions, often of unequal size, exhibit asymmetrical influence and behavioral patterns. Within these groups, elites and masses operate under different motivations and levels of influence, challenging simplistic views of polarization. Yet, existing methods for measuring polarization in social networks typically reduce it to a single value, assuming homogeneity in polarization across the entire system. While such approaches confirm the rise of political polarization in many social contexts, they overlook structural complexities that could explain its underlying mechanisms. We propose a method that decomposes existing polarization and alignment measures into distinct components. These components separately capture polarization processes involving elites and masses from opposing groups. Applying this method to Twitter discussions surrounding the 2019 and 2023 Finnish parliamentary elections, we find that (1) opposing groups rarely have a balanced contribution to observed polarization, and (2) while elites strongly contribute to structural polarization and consistently display greater alignment across various topics, the masses, too, have recently experienced a surge in alignment. Our method provides an improved analytical lens through which to view polarization, explicitly recognizing the complexity of and need to account for elite-mass dynamics in polarized environments.

Keywords: elections; polarization; elite polarization; mass polarization; social media; issue alignment; climate; immigration

1. Introduction

In an era of increasing political polarization (Böttcher and Gersbach, 2020; Karjus and Cuskley, 2024), the ability to dissect and understand the fractured structures within social systems is crucial. The proliferation of online platforms has transformed the landscape of political discourse, creating vast digital arenas where polarization can be observed, measured, and analyzed in unprecedented detail.

Current approaches to measuring polarization often reduce complex social dynamics to a single dimension (Salloum et al., 2022), thereby failing to expose hidden patterns behind the observed societal divisions. While these methods have been successfully applied to report polarization trends online, they fall short of capturing the full picture, obscuring the multifaceted nature of polarization, where different groups and their members contribute to and experience division in distinct ways.

Political polarization manifests in multiple modalities. This paper focuses on *ideological polarization*, which can be analytically disaggregated into two constituent dimensions: *issue divergence* (how far apart groups are on policy questions) and *issue alignment* (how tightly aligned individuals' views are across issues) (Lelkes, 2016). Our contribution is to show how the magnitude of these dimensions can be estimated separately for elites and masses within a single social system.

A common strategy is to model political interaction as a network, which then provides a foundation for detecting polarized groups and quantifying their separation. To build such networks, researchers often represent relevant interactions, such as friendships, agreements, or endorsements, as ties between individuals or groups (Adamic and Glance, 2005; Akoglu, 2014; Conover *et al.*, 2021; Garimella *et al.*, 2018). Once the network is constructed, community detection techniques can be applied to reveal group structure, and structural polarization scores, such as Random Walk Controversy (Garimella *et al.*, 2018) or the Adaptive EI-index (Chen *et al.*, 2021), are used to evaluate separation and cohesion.

We propose a method that utilizes these existing polarization scores and network representations of social systems to measure polarization separately for elites and masses. We draw on the well-established notion of *core-periphery* structure (Borgatti and Everett, 2000; Rombach *et al.*, 2017) to identify hierarchical groups in the network, as it closely matches the observation of elites being central and well-connected in social and political systems (Castells, 2011; Rahman Khan, 2012). Our approach not only unveils notable imbalances between polarized groups but also highlights the disproportionate role of elites in shaping polarized environments.

The distinction between *elite polarization* and *mass polarization* is crucial for understanding political division in modern democracies. While headlines and public discourse often paint a picture of a deeply divided populace, this perception may be skewed by the outsized influence of political elites in media narratives. By conflating elite and mass polarization, we risk misdiagnosing the stability of democratic systems and implementing misguided solutions. Separating these measurements allows us to discern whether the apparent chasm in political attitudes truly reflects widespread societal division or if it predominantly exists among a small, albeit influential, segment of the population.

Traditionally, elites have been viewed as more ideologically coherent and predictable than the general public, with some theories (e.g., Kinder and Kalmoe, 2017; Layman and Carsey, 2002) even suggesting that the masses remain resistant to elite polarization while opposing elite groups grow increasingly distant and internally homogeneous (Robison and Mullinix, 2016). However, recent research complicates this narrative, revealing potential ideological realignment among the masses (Kozlowski and Murphy, 2021; Levendusky, 2010).

Equally important is the substantial asymmetry in the relationship between political elites and the public. Elites often serve as influential cue-givers (Diermeier and Li, 2019; Green *et al.*, 2020; Skytte, 2021; Van Boven and Sherman, 2021) who shape mass attitudes and behaviors (Alley, 2023; Berinsky, 2007; Kousser and Tranter, 2018). This *elite signaling* phenomenon, in which followers adjust their stances and behavior in response to elite ideological divergence (Banda and Cluverius, 2018; Fine and Hunt, 2023), has been linked to political polarization (Bäck *et al.*, 2023; Skytte, 2021).

Because elites and masses play distinct roles in the process of polarization, it is critical to analyze the contributions of each separately. Elites cue public positions and, when aligned, can produce policy gridlock; the masses are slower to consolidate their partisan and ideological identities, but when they do, they harden in- and out-group divisions. Our method provides a systematic way to detect these dynamics and to assess whether and how elite-mass polarization patterns are evolving in social systems. We demonstrate this by examining five topics that have become significantly more polarized online in Finland between 2019 and 2023, a period marked by major global events including a pandemic and a war in Europe. By analyzing polarization trends across two snapshots of the Finnish Twittersphere—centered around the parliamentary elections of 2019 and 2023—we reveal distinct polarization dynamics across different networks. Our findings show that polarized groups, and the elites and masses within them, shape polarization trends in markedly different ways. Our analysis not only provides insight into Finland's evolving political landscape but also offers a method for understanding polarization dynamics in other contexts.

Our paper is structured as follows: Section 2 briefly reviews current polarization measures. Section 3 outlines our method for inferring hierarchical structures in polarized networks, and

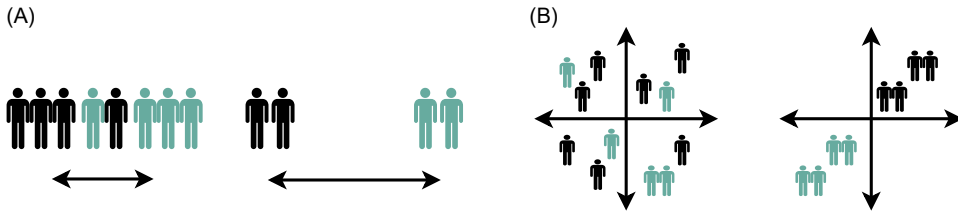


Figure 1. Political polarization can be assessed either by the degree of divergence between groups' opinions on a single issue (A) or by the extent to which their positions align across multiple political topics (B). In illustration A, the scenario on the right would show a higher degree of structural polarization than the one on the left, as the level of agreement between groups is lower, resulting in a deeper divide. In illustration B, different quadrants represent distinct pairs of stances an individual may hold on two separate topics (*for-for*, *for-against*, *against-for*, and *against-against*). The scenario on the left displays more mixing, as stances do not appear to be linked to each other, whereas the situation on the right shows strong alignment, with an individual's stance on the first topic entirely determining their stance on the second topic.

decomposes polarization using these hierarchies. Section 4 presents our main findings, analyzing hierarchical polarization trends on Twitter during the 2019 and 2023 Finnish parliamentary elections. We conclude with a discussion in Section 5.

2. Polarization measures

Quantifying political polarization in society is a complex task and can be approached from different perspectives (see Figure 1). Traditional approaches have relied on survey-based data to measure opinion bimodalities and issue alignment, estimating ideological shifts and relationships between distinct issue positions (e.g., Kozłowski and Murphy, 2021). However, the research landscape has evolved with the emergence of social media platforms, which now serve as critical arenas for daily political discourse and debate. This digital transformation has prompted researchers to increasingly focus on these platforms as rich data sources for understanding contemporary political dynamics (Bright, 2018; Conover et al., 2021; Cossard et al., 2020; Green et al., 2020).

Polarization is often assumed to impact social network structures, giving rise to two internally tightly-knit groups with sparse connections between them. This makes it possible to estimate opinion distribution bimodality by measuring communication limitations or reachability between opposing groups in a network (Garimella et al., 2018; Hohmann et al., 2023; Morales et al., 2015). The underlying intuition is that sharper, more distant peaks in opinion distribution correspond to more isolated communities in topic-specific discussion networks (see Figure 1A). This isolation is often termed *structural* (Salloum et al., 2022) or *interactional* (Falkenberg et al., 2024) polarization. Most methods designed to capture this phenomenon assess the separation of interaction patterns between two distinct groups (Garimella et al., 2018), although recent work has expanded to handle multiple groups (Martin-Gutierrez et al., 2023; Nair and Iamnitchi, 2024).

When measuring structural polarization in social systems, it is crucial to first represent the system as a network that clearly distinguishes polarized group structures from other kinds of interactions. This means defining the network carefully by determining which units (nodes) and relationships (links) to include. Specifically, networks used to measure polarization with these methods should represent positive relationships – such as friendships, similarity of preferences, endorsements, or agreements (Akoglu, 2014; Conover et al., 2021; Garimella et al., 2018) – between individuals, organizations, or even blogs (Adamic and Glance, 2005).

Measuring structural polarization in political communication networks typically involves three steps (Salloum et al., 2022): First, a network is constructed from collected data. Second, functional groups are identified using community detection techniques. Finally, the strength of the division is evaluated using structural polarization scores. Examples of metrics for this final step include Random Walk Controversy (Garimella et al., 2018), Dipole Index (Morales et al., 2015), Boundary

Index (Guerra *et al.*, 2013), and Adaptive EI-index (Chen *et al.*, 2021), which assess the degree of inter-group isolation or intra-group cohesion. The common intuition behind most of these scores is that they capture the extent to which individuals are “trapped” within their own ideological communities or avoid engaging with opposing viewpoints.

Many polarization quantification methods require identifying key network figures, often using endorsement counts (e.g., likes or reposts) as a proxy for importance or “eliteness” (Garimella *et al.*, 2018; Morales *et al.*, 2015; Salloum *et al.*, 2022). While some studies use known node labels to mark politicians in a network (Falkenberg *et al.*, 2022; Xia *et al.*, 2024), this approach may overlook influential users who lack explicit political affiliations, such as opinion leaders outside formal politics, who often outpace actual politicians in social media popularity (Mukerjee *et al.*, 2022). Such influential figures might include media personalities, activists, cultural figures, or academics. Although our method was not specifically designed to address this limitation, inferring key figures solely from structural properties can be advantageous in certain scenarios, particularly when explicit labeling is unavailable or insufficient.

Structural polarization scores, while insightful, often have limitations. They typically focus on two-group systems and single-issue interaction patterns, overlooking the critical aspects of issue alignment and multi-group environments. Issue alignment – the degree of agreement across multiple topics – is increasingly recognized as a more pernicious form of polarization (Chen *et al.*, 2021; Mason, 2015; Törnberg, 2022). It occurs when individuals or groups adopt collective stances on a range of issues based on shared values or ideologies (see Figure 1B), potentially exacerbating societal divisions (Mason, 2015; Törnberg, 2022). Researchers have applied more classic statistical methods, such as correlation (Kozłowski and Murphy, 2021), and information theory-based scores for quantifying issue alignment in polarized systems (Chen *et al.*, 2021; Iannucci *et al.*, 2024). These methods, while mostly used in two-group settings, can be extended naturally to multi-group systems.

3. Groups and hierarchies in polarized networks

Measuring polarization separately for elites and the masses requires the inference of their group memberships and hierarchy. We focus on two polarized groups for this study, but this method can be readily adapted to scenarios involving more than two opposing groups or hierarchies.

We begin by presenting the steps for identifying polarized groups in a network. Then, we partition these groups into smaller hierarchical subgroups based on predefined connectivity patterns, and conceptualize their possible interactions in a polarized environment. Finally, we connect the opposing groups, their hierarchies, and actions to existing polarization measures.

3.1 Identifying the polarized groups

The conventional approach for finding the assortative groups in polarized networks is to use clustering algorithms (Salloum *et al.*, 2022). We use the stochastic block model (SBM), which is a probabilistic model used to represent and analyze community structure in networks. It groups nodes into blocks, where the probability of connections between nodes depends solely on their block memberships. We specifically use a constrained version of this model, the planted-partition model (Zhang and Peixoto, 2020), which has been previously applied to find polarized groups (Peralta *et al.*, 2024).

The generative nature of the SBM enables rigorous model selection, as one can apply the Occam’s razor principle to select the configuration with the lowest description length (Grünwald, 2007). The property is particularly important as an observed community structure that could be explained due to randomness can be flagged, as the description length for an SBM with no community structure would be lower than that of a model attempting to overfit by dividing

the network into two communities (Peixoto, 2014b, 2019; Yan, 2016; Zhang and Peixoto, 2020). Therefore, when combined with model selection, SBM-based approaches would find zero polarization in entirely random networks, which is in contrast to most commonly used polarization pipelines introduced in the literature (Guimerà et al., 2004; Salloum et al., 2022). We sweep the number of blocks (groups) between one and two, and select the configuration leading to the lowest description length.

3.2 Unraveling the hierarchical structure

Hierarchies naturally emerge within social groups where members exhibit varying levels of influence, expertise, and dominance (Koski et al., 2015; Ureña-Carrión et al., 2023). In political contexts, one such common structure is the division between elites and masses. Elites are typically defined as individuals or groups that hold a disproportionate amount of power, influence, and resources in a given society (Rahman Khan, 2012). In contrast, the masses form the broader population with less direct control over political and economic decisions. This group often includes ordinary citizens and voters, whose political engagement may vary considerably.

The central role of elites in society and their ability to maintain power through interconnected networks (Castells, 2011) is mirrored in the structural characteristics of social networks. This has been observed, for instance, in social media studies suggesting that the most-followed accounts form cohesive clusters online, driven by social, cultural, and business forces (Motamedi et al., 2020). To effectively analyze these dynamics, we employ the concept of a core-periphery structure, which consists of a densely interconnected core surrounded by a sparsely connected periphery (Rombach et al., 2017). This structural pattern is commonly observed in various contexts, including social networks (Borgatti and Everett, 2000; Yang et al., 2018), economic systems (Csérmely et al., 2013; Wang, 2016), and many other fields (Csérmely et al., 2013; Kojaku and Masuda, 2017).

In our operationalization, the core of each network is treated as the elites and the periphery as the masses. These positions can be viewed as *structural elites* and *structural masses*, since status is determined by network position rather than fixed attributes. In digital spaces with many ongoing and distinct discussions that are open in principle to anyone, these categories are fluid and vary across platforms and topics. An influential opinion leader in climate discussions, for instance, may belong to the elite core in that context, while appearing as part of the masses in health policy debates. This dynamic structural definition has the advantage of being flexible enough to capture differences across domains and accurate in grounding elite–mass distinctions in observed network structures rather than external assumptions – although externally defined elite and mass groups could also be supplied if desired.

Several methods exist for inferring core-periphery structures in networks (Borgatti and Everett, 2000; Jeude et al., 2019; Kojaku and Masuda, 2017). To identify these structural hierarchies embedded in previously detected polarized groups, we follow a method that infers core-periphery structures using the SBM framework (Gallagher et al., 2021). Each group is treated independently, allowing for separate comparisons between their observed hierarchies and a null model. We use the common Erdős–Rényi model as the baseline, as it assumes no inherent hierarchical structure, including hierarchies potentially explained by the network's degree sequence (Kojaku and Masuda, 2018). The strengths and limitations of this approach are discussed further in Appendix B.

3.3 Interactions in polarized systems

To analyze polarization in greater detail, we distinguish among the types of interactions that can occur within and between elite and mass members of polarized groups (Figure 2A). Current approaches do not characterize these patterns separately, missing the conceptualization needed for decomposing observed polarization.

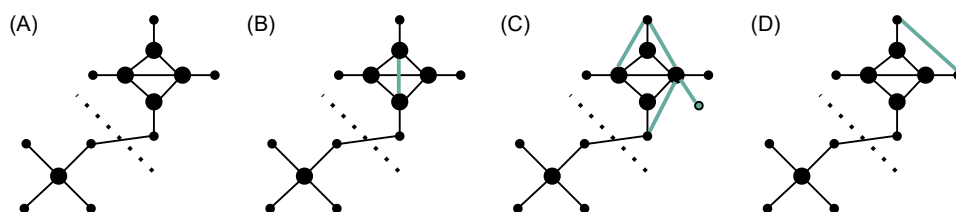


Figure 2. Our conceptualization enables categorizing the actions that lead to an increase in the structural polarization score. (A) is an example of a polarized network that has been partitioned into two groups representing communities with distinct stances on a specific political topic. Large nodes correspond to the elite members, while smaller ones indicate the mass. In (B), a new connection is formed between two elite members, making the elite group more cohesive. Another polarization-increasing action is when a new connection is formed towards the elite group either from an existing node or a new node. This is depicted in (C) and the overall degree of this type of actions is called mass amplification. Lastly, a connection between two members belonging to the mass, is shown in (D), illustrating the interpretation of mass cohesion. Lower mass cohesion can be seen corresponding to higher centralized opinion leadership among the elites, as most connections are directed towards them.

Elite-elite

Each new connection between two elites within their own group can be deemed as making the core more connected, and therefore, increasing the *elite cohesion* of that group (Figure 2B). Structurally, elite cohesion reaches its maximum when core members form a fully connected subgroup, or clique. Conceptually, elite cohesion refers to the extent of consolidation among elite individuals, where dense internal ties facilitate coordination, strengthen shared identity, and reinforce mutual power (Knoke, 1993).

Mass-elite

Elites can receive endorsements from the surrounding masses, increasing the total *mass amplification* of the corresponding group in the system (Figure 2C). This interaction may arise when an individual chooses to support an elite member, either due to shared beliefs or personal appeal (Fishbein and Coombs, 1974). Mass amplification reaches its maximum when each mass member has endorsed all elites within the group's core.

Mass-mass

A polarized group may have a very centralized opinion leadership, where the masses primarily amplify elites, or it may have a more evenly distributed pattern, where members of the masses in the periphery also endorse one another. This is captured by the group's *mass cohesion* (Figure 2D), which increases as connections are formed between two mass members.

Bridge

While the majority of interactions in polarized systems tend to concentrate within one's own group, this sharp divide is often softened by the presence of a *bridge*. It refers to cross-group connections, in which a member of one polarized group interacts with a member of the opposing group. Depending on the context, this type of interaction is, in certain cases, seen to reduce echo chambers or facilitate dialog within the system (Interian *et al.*, 2023).

3.4 Defining structural polarization and issue alignment for groups and hierarchies

We define structural polarization for a system of two groups with hierarchies in Section 3.4.1. Then, we derive the marginal polarization for the given structural score in Section 3.4.2. Finally,

in Section 3.4.3, we show how alignment is computed for the overall system, separately for elites and masses.

3.4.1 Structural polarization

A wide range of methods exist for quantifying structural polarization in networks. A comparative assessment showed that the best-performing method in detecting polarized networks was Adaptive EI-index (AEI; Salloum et al., 2022). Intuitively, the AEI produces a score based on the distribution of connections among nodes within and between groups, where a greater density of links within groups relative to those between groups indicates a more polarized system.

Moving to a mathematical formulation, let $G = (V, M)$ represent a network, where V is the set of nodes and M is the set of links. We define a partition of G into two groups, denoted as A and B , such that: $V = V_A \cup V_B$ and $V_A \cap V_B = \emptyset$, where V_A and V_B represent the sets of nodes belonging to groups A and B respectively. The AEI is defined for such a system as

$$P_{AEI} = \frac{i_A + i_B - 2e_{AB}}{i_A + i_B + 2e_{AB}}, \quad (1)$$

where i_X denotes the *internal links density* (observed links divided by the possible number of links) within group X and e_{AB} denotes the *external links density* between groups A and B .

In conventional applications of the AEI, all links are treated homogeneously within each group, which prevents it from distinguishing between the activity of the elites and the masses. However, as discussed in previous sections, these links are qualitatively distinct. Therefore, we want to distinguish the link densities according to the hierarchies in the network. We do this by decomposing the elements in P_{AEI} further, such that

$$i_X = \frac{I_{cX} + I_{cpX} + I_{pX}}{\frac{1}{2}n_X(n_X - 1)}, \quad (2)$$

where I_{cX} represents the interaction between group X 's core nodes, while I_{cpX} denotes the interaction between its core and periphery, I_{pX} denotes the interaction among its periphery nodes, and n_X denotes the number of nodes in group X . We can now substitute the corresponding terms of Eq. (2) into Eq. (1), which gives us

$$P_{AEI} = \left(\overbrace{\underbrace{\hat{i}_{cA}}_{\text{elite cohesion}} + \underbrace{\hat{i}_{cpA}}_{\text{mass amplification}} + \underbrace{\hat{i}_{pA}}_{\text{mass cohesion}}}^{\text{Group A's contribution}} \right) + \left(\underbrace{\hat{i}_{cB} + \hat{i}_{cpB} + \hat{i}_{pB}}_{\text{Group B's contribution}} - \underbrace{2\hat{e}_{AB}}_{\text{bridge}} \right), \quad (3)$$

where we use the hat symbol to denote the normalized form, where each component is divided by the denominator in Eq. (1). Note that there is a one-to-one correspondence between the polarization score and the bridge contribution: $2\hat{e}_{AB} = (1 - P_{AEI})/2$. For example, an increase in polarization corresponds to a proportional decline in the bridge.

We refer to Eq. (3) as *polarization decomposition* as it allows us to further study the direct effects of different groups and hierarchies on the polarization score. By splitting the observed polarization score into components, we can detect the differences between groups and their hierarchies on the polarization measure and determine whether their contributions have changed over time. When we say ‘contribution’ or ‘impact’ of a certain group on observed polarization, we specifically refer to these components. This is valuable information in the context of polarization, where we are often interested in monitoring the evolution of the opposing groups and their power dynamics.

3.4.2 Marginal polarization

Our polarization decomposition measures the contributions of elites and the masses as groups, and it cannot distinguish between the case where a large number of users each contribute small amounts to the polarization score and the case with a small number of users that have a high contribution. To distinguish between these cases we need to ask that what is the contribution of an *average* user in the group in addition to asking what is the contribution of the whole group.

We operationalize this by defining marginal polarization. It measures how much the polarization score shifts when an average node of a given type, either an elite or a mass participant from a particular group, is added to the system. By “average node” we mean a hypothetical node whose attributes are drawn from the observed distribution for that type and whose edges are placed according to the attachment patterns observed in the data. That is, marginal polarization measures sensitivity, i.e., how much additional polarization each type would generate if more of them entered. Together with the polarization decomposition, it tells us whether a group’s contribution comes from its number of members or the average effect of each member.

Formally, we define marginal polarization as the change in structural polarization when adding a representative node of each type. Consider a core node in group *A*: it has k_{cA} links to other core nodes in the same group, it receives k_{cpA} links from the periphery and forms k_{out} external links to group *B*. For a periphery node belonging to group *A*, it can have k_{pA} links to other periphery nodes. It amplifies k_{cpA} core nodes and connects to k_{out} links outside its own group. By substituting the average of observed values into these decomposed degrees, we can approximate the marginal polarization for a mean core node as

$$\Delta_{cA} P_{AEI} \approx \frac{2}{\alpha} \left(\frac{\langle k_{cA} \rangle + \langle k_{cpA} \rangle}{n_A^2} - \frac{\langle k_{outB} \rangle}{n_A n_B} \right), \quad (4)$$

where α denotes the denominator of Eq. (1). The formula is equivalent for the situation where an average periphery node is added to the polarized system ($\Delta_{pA} P_{AEI}$), but the $k_{\bullet}$ values must be adjusted to reflect the link distribution of peripheral nodes. Interestingly, the formula shows that having an equal number of links to the in-group and out-group is not enough to have zero net impact on polarization when group *A* is smaller than group *B*. See Appendix A for full derivation of this marginal polarization formula.

3.4.3 Issue alignment

In addition to structural polarization, we aim to quantify the degree of issue alignment separately for the elites and masses. Issue alignment contributes to political polarization by creating a situation where individuals or groups align themselves along specific issues. When individuals strongly identify with a particular issue or group and use it as a defining factor for their political stance towards other topics, it can lead to several detrimental consequences (Chen et al., 2021). We measure alignment using the normalized mutual information (NMI) method (Chen et al., 2021) for cores and peripheries separately. In essence, it quantifies the extent to which knowing the group of a user in one topic informs us about the user’s group in another topic.

This can be exemplified as follows. Consider three individuals, *i*, *j* and *k*, who have participated in discussions on both climate and immigration politics. In the climate discussion, individuals *i* and *j* hold pro-climate stances ($x(i) = 1, x(j) = 1$), while *k* holds an anti-climate stance ($x(k) = 0$). This results in the stance vector $x = [1, 1, 0]$.

When analyzing the immigration network, two scenarios are possible. First, if *i* and *j* share the same stance on immigration (pro-immigration) and *k* maintains an opposite view (anti-immigration), the immigration stance vector is $y_1 = [1, 1, 0]$. In this scenario, there is complete alignment between these individuals’ stances on climate and immigration, yielding $NMI(x, y_1) = 1$. This means we can perfectly predict individuals’ immigration stances based on their climate positions. NMI would also yields maximum alignment when $y_1 = [0, 0, 1]$, which

is a desirable property for measuring alignment because what matters is the alignment of cleavages. Alternatively, if i remains pro-immigration while j and k are anti-immigration, the stance vector becomes $y_2 = [1, 0, 0]$. This scenario results in a lower alignment score, as $\text{NMI}(x, y_2) < 1$, indicating reduced predictability regarding individuals' stances on these issues.

To estimate user j 's stance on a specific issue, we use its group membership, $j \in V_{A_1}$ or $j \in V_{B_1}$ for the first network, and $j \in V_{A_2}$ or $j \in V_{B_2}$, for the second network. Since this assessment requires user's presence in both networks, we add the stances for each node in $V_1 \cap V_2$ in stance vectors, s_1 and s_2 , corresponding to the respective issues, i.e., these stance vectors encode each user's group membership in the networks. Finally, we compute the issue alignment separately for the distinct hierarchies: $\text{NMI}(s_1^c, s_2^c)$ for the cores (elite alignment) and $\text{NMI}(s_1^p, s_2^p)$ for the peripheries (mass alignment). This measure is constrained between 0 and 1, where 0 indicates no alignment, and 1 indicates complete alignment between the issues in question.

4. Hierarchical polarization in Finnish parliamentary elections 2019 and 2023

In this section, we first present the data for our case study, which investigates polarization trends during the 2019 and 2023 Finnish parliamentary elections. We demonstrate how the overall structural polarization and issue alignment have increased in four years, and then decompose the polarization scores separately for the elites and the masses.

For both years, the networks undergo the exact same partitioning pipeline, as our goal is to understand the dynamics between elite and mass user groups at the system level rather than analyzing specific sets of users. In other words, the comparison is between two like-for-like systems (i.e., the same system at different time points) rather than between specific users at different time points. This enables us to assess the current state of polarization based on the network's present structural hierarchy, ensuring that our analysis reflects the system's current status rather than being influenced by individuals who were relevant in the past.

4.1 Data

We use data collected from Twitter (Public API v1 and v2) during the Finnish parliamentary elections in 2019 and 2023. To ensure consistency when comparing these two snapshots, we consider the 12-week period leading up to the election day: January 21–April 14, 2019 for the first election and January 9–April 2, 2023 for the second election.

For each election, we constructed five networks from sets of keywords related to larger topics, such as immigration and climate change (for more details, see Chen et al., 2021). Before constructing the networks, we removed retweets that included keywords from multiple topics, as they could inflate the observed alignment between topic-specific networks. This filtering step yields a more conservative estimate of actual alignment, ensuring that a user does not appear in multiple debates simply due to a single tweet referencing distinct topics. In addition, we excluded quote tweets, as these can carry disagreement or irony rather than endorsement.

Each network can be viewed as a graph where each user contributing to the topic is assigned to one node. In this graph, an edge between two nodes represents a retweet, which can be deemed as an agreement or a shared point of view on a selected issue between the corresponding users (Metaxas et al., 2015). We removed self-loops and parallel edges. An average graph consists of approximately 10,000 nodes and 30,000 edges (see Table in Appendix C for more details).

We confirm that our fitted models to this specific dataset provides sensible representations, as political figures are 3–4 times more likely to appear in the core than in the periphery (see Appendix D), and retweets predominantly flow from the periphery toward the core, supporting the concept of mass amplification (see Appendix E).

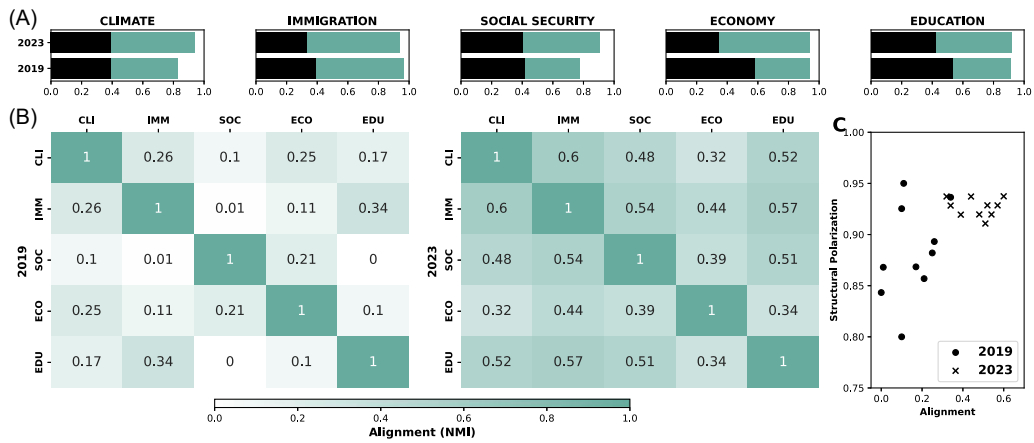


Figure 3. (A) demonstrates the increase in overall structural polarization across all networks, as measured by the AEI. Black bar corresponds to the proportion explained by the null model. The largest increase in the portion not explained by the null model was observed in the network representing economy-related discussions online. (B) The heatmaps depict the evolution of issue alignment over the four years, with 2019 on the left and 2023 on the right. Every pair of topics has experienced a substantial increase in the degree of alignment, as measured by the adjusted NMI. Climate and immigration were already reasonably aligned in 2019, however, the alignment doubled after four years. (C) illustrates the relationship between observed alignment and the average structural polarization scores for all topic pairs in both years. Note that in 2019, although some networks had high structural polarization scores, issue alignment remained relatively low. In contrast, by 2023, networks showed both high structural polarization and high issue alignment.

4.2 Drastic increase in polarization

Before breaking down polarization by groups and hierarchies, we first report the overall structural polarization and issue alignment across the networks in each snapshot (see Figure 3). In all systems analyzed, polarization is computed between two groups, labeled as either left-leaning (A) or right-leaning (B), based on the distribution of political parties within these groups (see Appendix F for labeling details). We also confirm that the inferred group structure is more plausible than a no-group structure using the assortativity test described in Section 3.

Structural polarization has increased in all the networks studied here, as shown in Figure 3. Over these four years, the topics *<climate>* and *<economy>* experienced the highest increase. Not only have the opposing groups moved further apart in some individual topics, they have also become more aligned as the observed alignment values experienced an upward swing in each pair of topics studied here. In other words, the tendency of the same individuals to be divided on other topics as well has become more pronounced. In 2023, knowing an individual's stance on immigration politics provided the most information about their potential views on other topics, where the strongest alignment was detected between *<immigration>* and *<climate>*, with an NMI score of 0.6, twice as high as in 2019. The least aligned topics were *<climate>* and *<economy>*, with an NMI value of 0.32, a slight increase from the 2019 alignment level of 0.25.

4.3 Groups have unequal impact on the overall structural polarization

Political polarization has clearly increased online in Finnish Twitter, a trend observed across various offline contexts as well (Kestilä-Kekkonen *et al.*, 2024). But how do these patterns vary across subgroups and different political processes? The decompositions in Figure 4A illustrate the dynamic nature of structural polarization, specifically how the dominant polarized group (i.e., the side accounting for the majority of the observed polarization) shifts over time and across issue domains. The portions of observed polarization attributed to the left-leaning group are shaded in red and correspond to the components $\hat{i}_{c_A}$ (elite cohesion), $\hat{i}_{p_A}$ (mass amplification), and $\hat{i}_{p_B}$ (mass

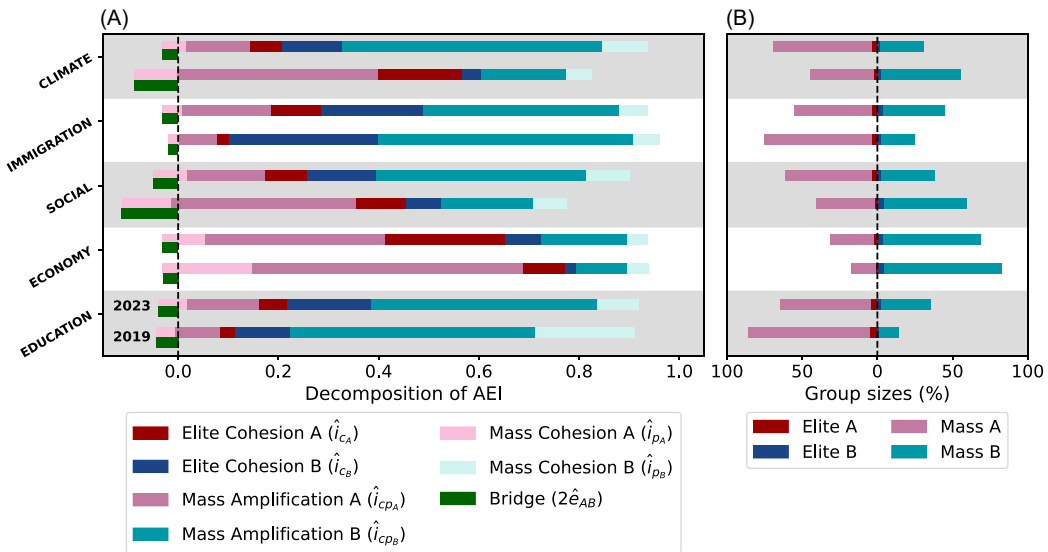


Figure 4. (A) Polarization decomposition for structural contributions of different groups and their hierarchies to AEI-score. Groups represent polarized communities on Finnish Twitter, with group A (red) being left-leaning and group B (blue) right-leaning. The figure illustrates the predominant influence of elite cohesion ($\hat{i}_{c_A}$ & $\hat{i}_{c_B}$), mass amplification ($\hat{i}_{cp_A}$ & $\hat{i}_{cp_B}$), and mass cohesion ($\hat{i}_{p_A}$ & $\hat{i}_{p_B}$) on the overall score. The green part of spectrum represents the impact of the bridge between the opposing entities ($2 \times \hat{e}_{AB}$). The contributions of different hierarchical members not only vary within individual networks but also across the distinct networks. The part of the spectrum that corresponds to the internal structures is shifted to the left by an amount equal to the cross-interactions. This enables us to read the unadjusted AEI score for each network directly from the figure. (B) Groups vary in their sizes, and mostly consists of the masses. Smaller groups can have a great impact on the observed polarization. The group sizes are normalized by dividing the number of nodes in each group by the total number of nodes in the graph. The pink and turquoise bars represent the proportion of mass members in groups A and B respectively (number of mass members in each group divided by total graph size). The same normalization applies to elites' group sizes.

cohesion). Similarly, the right-leaning group's parts are represented by blue-shaded components $\hat{i}_{c_B}$, $\hat{i}_{cp_B}$, and $\hat{i}_{p_B}$. The structurally depolarizing cross-interactions are captured by the green-shaded bridge-component.

From these decompositions, we identified number of patterns indicating that the increase in polarization from 2019 to 2023 involves distinct processes for elites and masses. Additionally, these polarization dynamics are unevenly distributed between the two opposing groups, each contributing to the overall observed scores to varying degrees.

First, focusing on the difference between elites and the masses, we observe that elites play a substantially larger role than the masses in the polarization observed in the system. While the public's interaction patterns tend to explain large parts of the observed polarization in absolute terms across all networks, the relative contribution (when adjusted to the sizes of these hierarchical groups) of the elite members is, on average, much greater, as is apparent from Figures 4B and 5B. This was particularly striking in the immigration context, where right-leaning elites made up less than 3% of users in 2019 but were responsible for over 30% of polarization. By 2023, they still accounted for only about 4% of users but contributed roughly 20% of polarization. Similarly, for education-related discussions in 2023, the right-leaning elites made up around 2% of participants yet contributed roughly 18% to polarization. We report disproportionately concentrated polarization for all groups and networks in Appendix G.

Second, we find that polarization patterns vary across opposing groups depending on context and time. A particularly pronounced shift occurred in climate-related discussions. The left-leaning group, dominant in 2019 (72% of polarization), saw its contributed share decrease to only 25% by 2023 despite growing in size. Meanwhile, the right-leaning elites became substantially more

consolidated, as their cohesion increased from 4 to 12%. There was also a notable rise in mass polarization on the right, where amplification and cohesion combined rose from 24 to 63%. We find similar patterns between the left and right across other issues. In fact, economic discussions remained the only one where the left-leaning group was more polarized than the right across time, with left-leaning elites nearly tripling their cohesion, from 9 to 25%, in 2023, whereas the overall right group became only minimally more polarized.

Finally, our results also show signs that the gap between groups is solidifying. Communication between polarized groups has mostly dropped or remained at the same level across all topics. This type of communication, labeled as the bridge, and discussed in Section 3.3, is the only interaction that structurally decreases polarization. Its share of the decomposition has declined markedly in the *<climate>* and *<social>* networks, while remaining broadly stable in *<education>*, *<economy>*, and *<immigration>* (see Figure 4A). Generally, in polarized networks, cross-group interactions have a limited impact on polarization measures compared to intra-group interactions, because they are often numerically and structurally outweighed by the internal organization of each group. This further emphasizes the importance of analyzing these interactional patterns separately.

The present discussion not only highlights the need to consider disaggregated sources of polarization, it demonstrates how our method can be a tool for studying how these distinct mechanisms drive polarization. Specifically, if a particular mechanism were indeed responsible for the observed asymmetries, we would expect that mechanism to leave a corresponding pattern in the decomposition results. For instance, previous studies have demonstrated that right-wing populist parties often act as polarizing forces within political systems (Loxbo, 2024; Silva, 2018). Such polarization can occur either through the mobilization of masses by deepening the sociocultural divide (Loxbo, 2024), or by reinforcing opposition among the adversaries (Nicholson, 2012). In our case, the right-leaning group of both the *<climate>* and *<immigration>* networks are predominantly composed of candidates from the right-wing populist party. In the decompositions of these networks, we observed that the right side of the climate debate gained a 39 percentage point increase in mass amplification, suggesting that the increase in polarization primarily occurred through mobilizing the masses. Meanwhile, the slight growth in the left group's overall impact on immigration reduced the earlier imbalance, suggesting that the intensified polarization in this case may have occurred through reinforcing adversarial opposition.

Additionally, our method allows exploration of how group composition, such as the prevalence of specific political parties within groups, shapes polarization patterns (see Appendix F). For instance, the notably low cohesion among right-leaning elites in the *<economy>* network could reflect the more fragmented composition of this overall group, as it was the only network where centrist parties were predominantly in the right-leaning cluster. Here, although centrist and right-wing parties appeared to hold sufficiently similar views to form a single cluster, relatively few endorsements between their elites were observed, demonstrating there are barriers to elite cooperation across more diverse preferences in a polarized environment.

Figure 5B shows how marginal polarization varies across domains. In 2019, the entry of an elite actor produced a particularly strong increase in polarization in the *<economy>* network, while in *<education>* the effect of an additional elite was much closer to that of a mass participant. Networks of *<climate>*, *<immigration>*, and *<social>* fell in between, where the entry of elites shifted polarization somewhat more than masses but not to the same degree as in *<economy>*. By 2023, the same broad pattern held, with elite entry still producing stronger shifts in *<economy>* and weaker shifts in *<education>*.

4.4 Elites are substantially more aligned than the masses

Given the crucial role elites play in shaping public discourse, we next analyze how issue alignment differs between elite and mass populations. To do this, we apply the method described

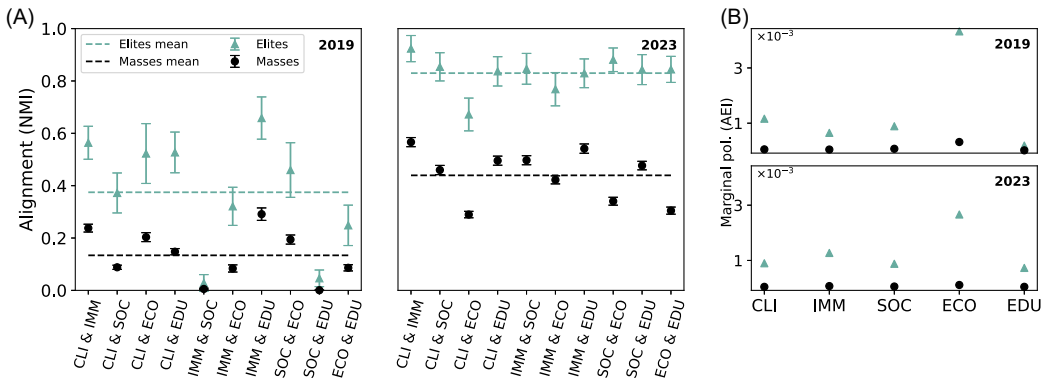


Figure 5. (A) Elites are consistently more aligned than masses across all topic pairs. Elites became more aligned in 2023, together with a smaller increase in the alignment of mass opinion on various issues. To capture the uncertainty around the observed values, we bootstrapped 500 pairs of networks for each topic pair. Each bootstrap sample represents a subgraph of the original network, where the sizes of cores and peripheries are subject to random fluctuations. We do this by sampling the nodes into groups according to their original group probabilities. (B) Elites tend to have higher marginal polarization as well compared to the masses. In both years, introducing a new elite member to the <economy> network had the greatest impact on its observed score. A weighted average of both groups' marginal values is applied to obtain a single value representing the hierarchy's mean effect on AEI.

in Section 3.4.3, which measures alignment pairwise across all the topics studied. We find that elites are consistently more aligned than the masses in all networks and across both elections, as shown in Figure 5A, which indicates that elites regularly maintain more interdependent stances on various political issues. This reflects the understanding that elites are generally more ideologically coherent and entrenched in their positions than the general public tends to be (Layman and Carsey, 2002). Elites became even more aligned by 2023, with eight out of ten pairs of topics having an NMI value over 0.8, which can be considered very high – especially when compared to previous findings in the literature (Chen et al., 2021; Iannucci et al., 2024). As for the masses, they showed relatively weak issue alignment in 2019, and the gap in mean alignment levels between them and the elites not only persisted but even widened by 2023. Despite this, it is important to highlight that the masses experienced a much sharper increase in issue alignment over the four-year period compared to the elites. The mean alignment for the masses rose from 0.13 to 0.44, representing an increase of approximately 228%, which was notably more pronounced compared to the elites' increase of 121%.

Looking deeper into specific areas of alignment, our findings show that immigration and climate politics are more strongly aligned than immigration and social security issues, despite social security and immigration often intersecting in policy debates (e.g., discussions about welfare benefits for immigrants). In 2019, the topic pairs <immigration>–<education> and <immigration>–<climate> showed the highest alignment, while pairs involving <social> showed the lowest alignment. By 2023, <immigration>–<climate> had become the most aligned pair, with <economy>–<climate> now showing the least alignment across both the elites and masses.

This pattern of immigration and climate politics being strongly aligned is consistent with prior research that links this specific issue alignment to mechanisms of partisan sorting and the broader universalist-communitarian divide (Chen et al., 2021). Our decomposition provides even more details on this, as it reveals that by 2023, elites were about 1.6 times more aligned than the masses on climate and immigration issues. Over time, alignment among elites had increased by approximately 64%, while for the masses, it rose by 138%. Regarding the relatively low alignment observed between climate and economic issues, one might have expected a stronger connection given that economic growth, consumption patterns, and regulation are central to climate politics.

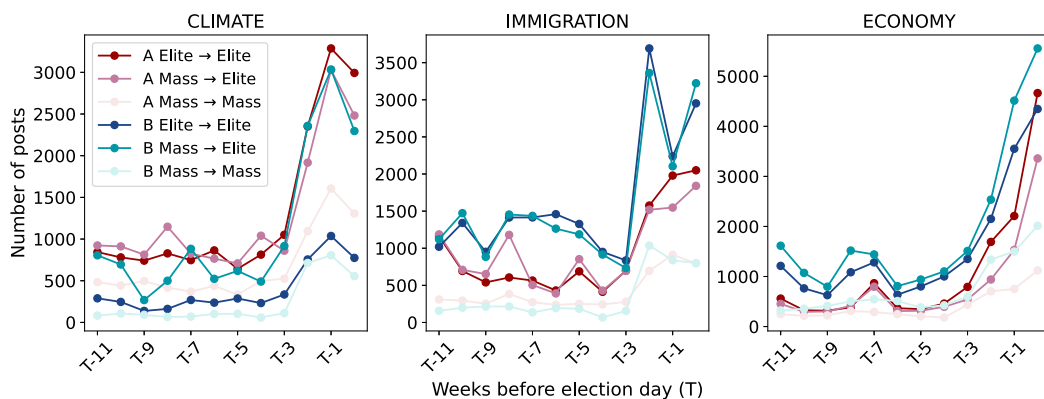


Figure 6. Activity patterns in the most polarized networks in 2023 separated at group and hierarchy level. In all networks, the largest jump in activity takes place approximately three weeks before the election day, particularly within the right-leaning group in *<immigration>* network. Which opposing group is more active depends on the issue. For instance, activity within the right-leaning elites is substantially higher in *<immigration>* and *<economy>*, whereas in *<climate>*, *<social>* and *<education>*, left-leaning elites appear to be more active. The extent of activity of a specific group does not translate into the observed polarization. Figures for the remaining topics and for 2019 can be found in Appendix H.

Nevertheless, our analysis indicates that elites' climate views were in fact 38% more closely aligned with their positions on immigration than with economic politics.

Our method detects critical differences in issue alignment between elites and the masses. It makes it possible to track, for example, whether a certain campaign, event, or shock, framed by elites via multiple issues, has an aligning impact on the surrounding masses in the system. As for now, the gap in issue alignment between elites and masses remains wide in the Finnish context.

4.5 Group sizes and activity patterns are inadequate indicators of polarization contribution

It may seem intuitive that larger or more active groups would contribute more to polarization. However, our analysis reveals a more complex relationship. As shown in Figure 4B, larger groups do not necessarily contribute more to polarization. For instance, in the immigration- or economy-related discussions, smaller groups accounted for a substantially larger share of polarization. Specifically, only 17% (in 2019) and 32% (in 2023) of participants belonged to the left-leaning group in *<economy>*, yet their contribution to the observed polarization exceeded 70% in both years.

Beyond group size, polarized groups also differ in activity patterns. From temporal activity plots in Figure 6, we see that higher levels of activity do not directly dictate a group's structural contribution to observed polarization. For instance, in *<climate>*, the elite cohesion of the right-leaning group in 2023 was approximately twice that of left-leaning elites. Yet, left-leaning elites consistently endorsed each other more frequently throughout the period. In contrast, the *<immigration>* network showed correspondingly greater elite and mass activity for the right-leaning group, aligning closely with their dominance in elite cohesion and mass amplification. Additionally, differences in elite and mass activity clearly varies by issue. In *<immigration>* and *<economy>*, mass-elite and elite-elite interactions saw similar activity levels, whereas in *<climate>*, the former substantially exceeded the latter, especially within the right-leaning group. Taken together, these results support our claim that activity volume cannot be used as a proxy for polarization contribution; organizational structure within groups better explains the observed polarization. We further substantiate this claim in Appendix I.

5. Discussions and conclusions

A system becomes polarized when groups exhibit strong internal cohesion and sharp external division. We already know how to measure the intensity of this division, but less is known about the underlying structures contributing to it. This paper aimed to express the observed polarization in terms of groups, as well as the elites and masses within these groups.

Decomposing polarization revealed (1) asymmetric contributions, highlighting that one opposing group typically dominates in structural polarization; (2) a disproportionate role of elites, who consistently demonstrate greater internal cohesion and issue alignment compared to the masses; and (3) the dynamic nature of polarization, illustrated through significant shifts observed in the Finnish political landscape between the 2019 and 2023 parliamentary elections. Instead of just reporting continuously heightened polarization scores for different political or social systems, we pinpointed the exact structures and the extent to which they contribute to polarization.

Our findings point to the critical and disproportionate role of elites in polarized systems. Elite groups not only appear to be structurally more polarized but also exhibit stronger alignment across different political issues. Given their central positions, elites serve as influential cue-givers whose polarization can directly shape mass attitudes and behaviors. This dynamic holds potential risks; elites propagating extreme or dangerous frames may significantly influence public discourse, possibly intensifying polarization at the mass level. On the other hand, the increased clarity in elite positioning may engage the public more actively in political discussions, reducing the perceived gap between elites and masses that reduces overall political participation (Levendusky, 2010).

Importantly, our findings also show that mass groups have become increasingly aligned over time, indicating that polarization is not confined solely to elite circles. This rise in mass alignment suggests that some ideological realignment could be extending into broader segments of the electorate, consistent with recent research (e.g., Kozłowski and Murphy, 2021) that challenges some earlier theories positing that polarization primarily occurs among elites (e.g., Kinder and Kalmoe, 2017).

We believe our method offers additional advantages beyond the insights already discussed. In particular, the polarization decompositions provide a framework for testing hypotheses about the origins and dynamics of emergent patterns. One notable strength of the approach is its capacity to function as an analytical detector: any dominant underlying mechanism should leave discernible traces, such as shifts, correlations, or other quantifiable features, within the decompositions. By systematically comparing these decompositions for different scenarios, one can assess whether a proposed mechanism is indeed a credible driver.

Our work helps make sense of today's polarized environments. We demonstrate the importance of measures capturing layered group dynamics in these contexts. By distinguishing between elite and mass polarization, we uncover patterns that are important to understanding how societal divides emerge and evolve.

Acknowledgments. The calculations presented above were performed using computer resources within the Aalto University School of Science "Science-IT" project.

Funding statement. Salloum and Kivelä were funded by the Research Council of Finland. Salloum received funding under grant number 353799, and Kivelä's work was supported by grant numbers 349366, 353799, 352561, and 357743.

Data availability statement. The data is published on Zenodo at <https://zenodo.org/records/17446451>, and the code is available on GitHub at <https://github.com/alesalloum/elite-mass-polarization>.

Competing interests. The authors declare that they have no competing interests.

References

- Adamic, L. A., & Glance, N. (2005). The political blogosphere and the 2004 us election: divided they blog. In: *Proceedings of the 3rd International Workshop on Link Discovery* (pp. 36–43).

- Akoglu, L. (2014). Quantifying political polarity based on bipartite opinion networks. In: *Proceedings of the International AAAI Conference on Web and Social Media* (vol. 8, pp. 2–11).
- Alley, J. (2023). Elite cues and public attitudes towards military alliances. *Journal of Conflict Resolution*, 67(7–8), 1537–1563.
- Bäck, H., Carroll, R., Renström, E., & Ryan, A. (2023). Elite communication and affective polarization among voters. *Electoral Studies*, 84, 102639.
- Banda, K. K., & Cluverius, J. (2018). Elite polarization, party extremity, and affective polarization. *Electoral Studies*, 56, 90–101.
- Berinsky, A. J. (2007). Assuming the costs of war: events, elites, and american public support for military conflict. *The Journal of Politics*, 69(4), 975–997.
- Borgatti, S. P., & Everett, M. G. (2000). Models of core/periphery structures. *Social Networks*, 21(4), 375–395.
- Böttcher, L., & Gersbach, H. (2020). The great divide: drivers of polarization in the us public. *EPJ Data Science*, 9(1), 32.
- Bright, J. (2018). Explaining the emergence of political fragmentation on social media: the role of ideology and extremism. *Journal of Computer-Mediated Communication*, 23(1), 17–33.
- Castells, M. (2011). Network theory – a network theory of power. *International Journal of Communication*, 5, 15.
- Chen, T. H. Y., Salloum, A., Gronow, A., Ylä-Anttila, T., & Kivelä, M. (2021). Polarization of climate politics results from partisan sorting: evidence from finnish twittersphere. *Global Environmental Change*, 71, 102348.
- Conover, M., Ratkiewicz, J., Francisco, M., Goncalves, B., Menczer, F., & Flammini, A. (2021). Political polarization on twitter. *Proceedings of the International AAAI Conference on Web and Social Media*, 5(1), 89–96. doi: [10.1609/icwsml.v5i1.14126](https://ojs.aaai.org/index.php/ICWSM/article/view/14126). <https://ojs.aaai.org/index.php/ICWSM/article/view/14126>.
- Cossard, A., De Francisci Morales, G., Kalimeri, K., Mejova, Y., Paolotti, D., & Starnini, M. (2020). Falling into the echo chamber: the italian vaccination debate on twitter. *Proceedings of the International AAAI Conference on Web and Social Media*, 14(1), 130–140. doi: [10.1609/icwsml.v14i1.7285](https://ojs.aaai.org/index.php/ICWSM/article/view/7285). <https://ojs.aaai.org/index.php/ICWSM/article/view/7285>.
- Csermely, P., London, A., Wu, L.-Y., Uzzi, B. (2013). Structure and dynamics of core/periphery networks. *Journal of Complex Networks*, 1(2), 93–123.
- Diermeier, D., & Li, C. (2019). Partisan affect and elite polarization. *American Political Science Review*, 113(1), 277–281.
- Falkenberg, M., Galeazzi, A., Torricelli, M., Di Marco, Nò, Larosa, F., Sas, M., Mekacher, A., Pearce, W., Zollo, F., Quattrociochi, W., Baronchelli, A. (2022). Growing polarization around climate change on social media. *Nature Climate Change*, 12(12), 1114–1121.
- Falkenberg, M., Zollo, F., Quattrociochi, W., Pfeffer, J., & Baronchelli, A. (2024). Patterns of partisan toxicity and engagement reveal the common structure of online political communication across countries. *Nature Communications*, 15(1), 9560.
- Fine, J. A., & Hunt, M. F. (2023). Negativity and elite message diffusion on social media. *Political Behavior*, 45(3), 955–973.
- Fishbein, M., & Coombs, F. S. (1974). Basis for decision: an attitudinal analysis of voting behavior 1. *Journal of Applied Social Psychology*, 4(2), 95–124.
- Gallagher, R. J., Young, J.-G., & Foucault Welles, B. (2021). A clarified typology of core-periphery structure in networks. *Science Advances*, 7(12), eabc9800.
- Garimella, K., De Francisci Morales, G., Gionis, A., & Mathioudakis, M. (2018). Quantifying controversy on social media. *ACM Transactions on Social Computing* 1(1), 27.1. doi: [10.1145/3140565](https://doi.org/10.1145/3140565).
- Green, J., Edgerton, J., Naftel, D., Shoub, K., & Cranmer, S. J. (2020). Elusive consensus: polarization in elite communication on the covid-19 pandemic. *Science Advances*, 6(28), eabc2717. doi: [10.1126/sciadv.abc2717](https://doi.org/10.1126/sciadv.abc2717).
- Grünwald, P. D. (2007). The Minimum Description Length Principle. Adaptive Computation and Machine Learning series. Cambridge: MIT Press. isbn: 9780262072816.
- Guerra, P., Meira, W., Jr, 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, pp. 215–224).
- Guimerà, R., Sales-Pardo, M., & Amaral, L. I. A. N. (2004). Modularity from fluctuations in random graphs and complex networks. *Physical Review E*, 70(2), 025101. doi: [10.1103/PhysRevE.70.025101](https://doi.org/10.1103/PhysRevE.70.025101).
- Hohmann, M., Devriendt, K., & Coscia, M. (2023). Quantifying ideological polarization on a network using generalized euclidean distance. *Science Advances*, 9(9), eabq2044.
- Iannucci, L., Faqeeh, A., Salloum, A., Chen, T. H. Y., & Kivelä, M. (2024). Multiway alignment of political attitudes. arXiv: 2408.00139.
- Interian, R., Marzo, R. A. G., Mendoza, I., & Ribeiro, C. C. (2023). Network polarization, filter bubbles, and echo chambers: an annotated review of measures and reduction methods. *International Transactions in Operational Research*, 30(6), 3122–3158.
- Jeude, J. VAN L. DE, Caldarelli, G., & Squartini, T. (2019). Detecting core-periphery structures by surprise. *Europhysics Letters*, 125(6), 68001.
- Karjane, A., & Cuskley, C. (2024). Evolving linguistic divergence on polarizing social media. *Humanities and Social Sciences Communications*, 11(1), 1–14.

- Kestilä-Kekkonen, E., Rapeli, L., & Söderlund, P. (2024). *Pääministerivaalit polarisaation aikakaudella: eduskuntavaalitutkimus 2023 [in Suomi] Justitieministeriets publikationer: Utredningar och anvisningar 10*. Finland: Oikeusministeriö. isbn: 978-952-400-891-4.
- Kinder, D. R., & Kalmoe, N. P. (2017). *Neither Liberal Nor Conservative: Ideological Innocence in the American Public* (pp. 2–3). University of Chicago Press.
- Knoke, D. (1993). Networks of elite structure and decision making. *Sociological Methods & Research*, 22(1), 23–45.
- Kojaku, S., & Masuda, N. (2017). Finding multiple core-periphery pairs in networks. *Physical Review E*, 96(5), 052313. doi: [10.1103/PhysRevE.96.052313](https://link.aps.org/doi/10.1103/PhysRevE.96.052313). <https://link.aps.org/doi/10.1103/PhysRevE.96.052313>
- Kojaku, S., & Masuda, N. (2018). Core-periphery structure requires something else in the network. *New Journal of Physics*, 20(4), 043012.
- Koski, J. E., Xie, H., & Olson, I. R. (2015). Understanding social hierarchies: the neural and psychological foundations of status perception. *Social Neuroscience*, 10(5), 527–550.
- Kousser, T., & Tranter, B. (2018). The influence of political leaders on climate change attitudes. *Global Environmental Change*, 50, 100–109.
- Kozłowski, A. C., & Murphy, J. P. (eds.) (2021). Issue alignment and partisanship in the american public: revisiting the partisans without constraint thesis. *Social Science Research*, 94, 102498.
- Layman, G. C., & Carsey, T. M. (2002). Party polarization and "conflict extension" in the american electorate. *American Journal of Political Science*, 46(4), 786–802. issn: 00925853, 15405907, accessed August 23, 2024. <http://www.jstor.org/stable/3088434>
- Lelkes, Y. (2016). Mass polarization: manifestations and measurements. *Public Opinion Quarterly*, 80(S1), 392–410.
- Levendusky, M. S. (2010). Clearer cues, more consistent voters: a benefit of elite polarization. *Political Behavior*, 32, 111–131.
- Loxbo, K. (2024). How the radical right reshapes public opinion: the sweden democrats' local mobilisation, 2002–2020. *West European Politics*, 48(7), 1653–1680. <https://doi.org/10.1080/01402382.2024.2396775>.
- Martin-Gutierrez, S., Losada, J. C., & Benito, R. M. (2023). Multipolar social systems: measuring polarization beyond dichotomous contexts. *Chaos, Solitons & Fractals*, 169, 113244.
- Mason, L. (2015). I disrespectfully agree": the differential effects of partisan sorting on social and issue polarization. *American Journal of Political Science*, 59(1), 128–145.
- Metaxas, P., Mustafaraj, E., Kily Wong, L. Z., O'Keefe, M., & Finn, S. (2015). What do retweets indicate? Results from user survey and meta-review of research. In: *Proceedings of the International AAAI Conference on Web and Social Media* (vol. 9, pp. 658–661).
- 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. doi: [10.1063/1.4913758](https://doi.org/10.1063/1.4913758). issn: 1054-1500.
- Motamedi, R., Jamshidi, S., Rejaie, R., & Willinger, W. (2020). Examining the evolution of the twitter elite network. *Social Network Analysis and Mining*, 10, 1–18.
- Mukerjee, S., Jaidka, K., & Lelkes, Y. (2022). The political landscape of the us twitterverse. *Political Communication*, 39(5), 565–588.
- Nair, S., & Iamnitchi, A. (2024). Cross-community affinity: a polarization measure for multi-community networks. *Online Social Networks and Media*, 43, 100280.
- Nicholson, S. P. (2012). Polarizing cues. *American Journal of Political Science*, 56(1), 52–66.
- Peixoto, T. P. (2014a). Hierarchical block structures and high-resolution model selection in large networks. *Physical Review X*, 4(1), 011047. doi: [10.1103/PhysRevX.4.011047](https://doi.org/10.1103/PhysRevX.4.011047).
- Peixoto, T. P. (2014b). *The graph-tool python library*. figshare. doi: [10.6084/m9.figshare.1164194](https://doi.org/10.6084/m9.figshare.1164194).
- Peixoto, T. P. (2019). Bayesian stochastic blockmodeling. In *Chap. 11 in Advances in Network Clustering and Blockmodeling* (vol. 11, pp. 289–332). Chichester: John Wiley & Sons, Ltd. doi: [10.1002/9781119483298.ch11](https://doi.org/10.1002/9781119483298.ch11). isbn: 9781119483298.
- Peralta, A. F., Ramaciotti, P., Kertész, J., & Iñiguez, G. (2024). Multidimensional political polarization in online social networks. *Physical Review Research*, 6(1), 013170.
- Khan, R., & Shamus (2012). The sociology of elites. *Annual Review of Sociology*, 38(1), 361–377.
- Robison, J., & Mullinix, K. J. (2016). Elite polarization and public opinion: how polarization is communicated and its effects. *Political Communication*, 33(2), 261–282.
- Rombach, P., Porter, M. A., Fowler, J. H., & Mucha, P. J. (2017). Core-periphery structure in networks (revisited). *SIAM Review*, 59(3), 619–646.
- Rousseeuw, P. J. (1984). Least median of squares regression. *Journal of the American Statistical Association*, 79(388), 871–880.
- Salloum, A., Chen, T. H. Y., & Kivelä, M. (2022). Separating polarization from noise: comparison and normalization of structural polarization measures. *Proceedings of the ACM on Human-Computer Interaction*, 6(CSCW1). doi: [10.1145/3512962](https://doi.org/10.1145/3512962).
- Silva, B. C. (2018). Populist radical right parties and mass polarization in the netherlands. *European Political Science Review*, 10(2), 219–244.
- Skytte, R. (2021). Dimensions of elite partisan polarization: disentangling the effects of incivility and issue polarization. *British Journal of Political Science*, 51(4), 1457–1475.

- Törnberg, P. (2022). How digital media drive affective polarization through partisan sorting. *Proceedings of the National Academy of Sciences*, 119(42), e2207159119.
- Ureña-Carrión, J., Karimi, F., Íñiguez, G., & Kivelä, M. (2023). Assortative and preferential attachment lead to core-periphery networks. *Physical Review Research*, 5(4), 043287.
- Van Boven, L., & David, K. S. (2021). Elite influence on public attitudes about climate policy. *Current Opinion in Behavioral Sciences*, 42, 83–88.
- Wang, C. (2016). Core-periphery trading networks. doi: [10.2139/ssrn.2747117](https://doi.org/10.2139/ssrn.2747117). <https://ssrn.com/abstract=2747117>.
- Xia, Y., Gronow, A., & Arttu, M. Ä., et al. (2024). The russian invasion of ukraine selectively depolarized the finnish nato discussion on twitter. *EPJ Data Science*, 13, 1. doi: [10.1140/epjds/s13688-023-00441-2](https://doi.org/10.1140/epjds/s13688-023-00441-2).
- Yan, X. (2016). Bayesian model selection of stochastic block models. In: *Proceedings of the 2016 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining*. ASONAM (vol. 16, pp. 323–328). Davis: IEEE Press. isbn: 9781509028467.
- Yanchenko, E., & Sengupta, S. (2023). Core-periphery structure in networks: a statistical exposition. *Statistics Surveys*, 17, 42–74. doi: [10.1214/23-SS141](https://doi.org/10.1214/23-SS141).
- Yang, J., Zhang, M., Shen, K. N., Ju, X., & Guo, X. (2018). Structural correlation between communities and core-periphery structures in social networks: evidence from twitter data. *Big Data Analytics for Business Intelligence, Expert Systems with Applications*, 111, 91–99. doi: [10.1016/j.eswa.2017.12.042](https://doi.org/10.1016/j.eswa.2017.12.042). issn: 0957-4174. <https://www.sciencedirect.com/science/article/pii/S0957417417308643>.
- Zhang, L., & Peixoto, T. P. (2020). Statistical inference of assortative community structures. *Physical Review Research*, 2(4), 043271.

Appendix A. Marginal polarization

Marginal polarization quantifies the incremental increase in overall structural polarization resulting from the addition of an “average” core (or periphery) node to group A (or B). Assume that an average core node has k_{c_A} connections to other core nodes within the same group, receives k_{cp_A} connections from the periphery, and establishes k_{out} connections with the opposing group B. The marginal polarization can then be expressed as follows:

$$\begin{aligned} \Delta_{c_A} P_{AEI} = & \frac{\frac{(I_{c_A} + k_{c_A} + I_{cp_A} + k_{cp_A} + I_{p_A})}{\binom{n_A+1}{2}} + \frac{(I_{c_B} + I_{cp_B} + I_{p_B})}{\binom{n_B}{2}}}{\alpha'} \\ & - 2 \times \frac{E + k_{out_B}}{(n_A + 1)n_B} \\ & - \frac{\frac{(I_{c_A} + I_{cp_A} + I_{p_A})}{\binom{n_A}{2}} + \frac{(I_{c_B} + I_{cp_B} + I_{p_B})}{\binom{n_B}{2}}}{\alpha} \\ & - 2 \times \frac{E}{n_A n_B}, \end{aligned} \quad (A1)$$

where $I_{\bullet}$ is the observed number of links within the corresponding structure, and E denotes the observed number of links between groups A and B. Each term is adjusted for density, as described in the main text. We approximate $\alpha' = \alpha$. By substitution, we get:

$$\begin{aligned} \Delta_{c_A} P_{AEI} = & \frac{\frac{(I_{c_A} + k_{c_A} + I_{cp_A} + k_{cp_A} + I_{p_A})}{\frac{1}{2}n_A(n_A + 1)} - 2 \times \frac{E + k_{out_B}}{(n_A + 1)n_B}}{\alpha} \\ & - \frac{\frac{(I_{c_A} + I_{cp_A} + I_{p_A})}{\frac{1}{2}n_A(n_A - 1)} + 2 \times \frac{E}{n_A n_B}}{\alpha} \end{aligned}$$

$$\begin{aligned}
& \frac{(I_{c_A} + k_{c_A} + I_{cp_A} + k_{cp_A} + I_{p_A})}{\frac{1}{2}n_A(n_A)} - 2 \times \frac{E + k_{out_B}}{n_A n_B} \\
& \approx \frac{\alpha}{\alpha} \\
& \frac{(I_{c_A} + I_{cp_A} + I_{p_A})}{\frac{1}{2}n_A(n_A)} + 2 \times \frac{E}{n_A n_B} \\
& - \frac{\alpha}{\alpha} \\
& = \frac{(k_{c_A} + k_{cp_A})}{\frac{1}{2}n_A(n_A)} - 2 \times \frac{k_{out_B}}{n_A n_B} \\
& \frac{\alpha}{\alpha} \\
& = 2 \times \frac{(k_{c_A} + k_{cp_A})}{(n_A)^2} - \frac{k_{out_B}}{n_A n_B} \\
& \frac{\alpha}{\alpha} \\
& = \frac{2}{\alpha} \times \left(\frac{(k_{c_A} + k_{cp_A})}{n_A^2} - \frac{k_{out_B}}{n_A n_B} \right) \tag{A2}
\end{aligned}$$

Similarly, for the remaining node types:

$$\Delta_{c_B} P_{AEI} \approx \frac{2}{\alpha} \times \left(\frac{(k_{c_B} + k_{cp_B})}{n_B^2} - \frac{k_{out_A}}{n_B n_A} \right) \tag{A3}$$

$$\Delta_{p_A} P_{AEI} \approx \frac{2}{\alpha} \times \left(\frac{(k_{p_A} + k_{cp_A})}{n_A^2} - \frac{k_{out_B}}{n_A n_B} \right) \tag{A4}$$

$$\Delta_{p_B} P_{AEI} \approx \frac{2}{\alpha} \times \left(\frac{(k_{p_B} + k_{cp_B})}{n_B^2} - \frac{k_{out_A}}{n_B n_A} \right) \tag{A5}$$

Appendix B. Model selection

Our method requires the network to be partitioned into k polarized groups, based on the assortativity assumption. Each of these polarized groups is then further decomposed into hierarchical subgroups, following the hierarchy assumption. For example, when $k = 2$, the resulting decomposition yields four distinct groups, as each larger group has their own elites and masses.

Many existing methods for uncovering the modular structure of networks struggle to differentiate meaningful patterns from noise, often due to the lack of statistical validation. Our goal is to address this limitation by asking two specific questions:

1. Is there enough evidence that the network can be decomposed into two polarized groups?
2. If so, is there sufficient evidence to further partition each group into hierarchical (core-periphery) substructures?

To address these questions, we implement the following two-step partitioning pipeline:

1. The network is initially divided into two blocks using the non-uniform planted-partition model.
2. Each group is then examined independently, applying the core-periphery detection algorithm to identify hierarchical structure.

We adopt the traditional hub-and-spoke interpretation of core-periphery organization, as it has been shown to better account for patterns of online amplification compared to the layered

variant (Gallagher *et al.*, 2021). This two-step decomposition process is repeated 100 times, and each resulting partition is stored for model selection based on the minimum description length principle (Grünwald, 2007; Peixoto, 2014b, 2019).

If the network does not exhibit significant assortative grouping, we conclude that it lacks structural polarization. In our experiments, however, all networks were more effectively described by two groups, based on reductions in description length.

We then turn our attention to evaluating hierarchical structure within each group. Detecting hierarchy is inherently more challenging than detecting its absence. A natural baseline for statistical comparison is the Erdős–Rényi model, in which each possible edge (i, j) appears independently with fixed probability p . Any apparent heterogeneity under this model can be attributed to random fluctuations, making it a reasonable null model for our purposes.

A stricter alternative is the configuration model, which preserves the degree sequence of the original network. While more constrained, this model presents limitations when applied to core-periphery detection. Previous work has shown that, in some cases, the degree distribution alone can account for the observed core-periphery structure (Kojaku and Masuda, 2018). We chose the ER model over the configuration model for three key reasons:

1. Our objective is to identify structurally dominant users, and it is acceptable for this dominance to correlate with higher degree (e.g., Garimella *et al.*, 2018).
2. Prior studies (e.g., Yanchenko and Sengupta, 2023) have shown that the configuration model can suffer from low statistical power in core-periphery detection. For the ideal structure, it may allow only the original network to satisfy the fixed degree sequence, making rewiring impossible and resulting in a p-value of 1. Other work suggests that an additional structural block—such as a community or another core-periphery—is needed to establish significance (Kojaku and Masuda, 2018).
3. The Erdős–Rényi model is simple, intuitive, and avoids complications such as self-loops or parallel edges.

Ideally, we would apply a constrained version of the nested stochastic block model (Peixoto, 2014a) to reflect the hierarchical nature of our approach. However, developing such a method is beyond the scope of this study. Instead, we rely on the methodology outlined above for model selection.

Appendix C. Summary of networks

See Appendix Table C1.

Table C1. Retweet networks were constructed for both election years, covering five distinct networks each. These networks were constructed based on Twitter data obtained over a span of 12 weeks leading up to the respective election day, which were 14.4. for 2019 and 2.4. for 2023. $|N|$ represents the count of unique nodes (users), and $|E|$ denotes the count of unique edges (retweets) in the network after the preprocessing

Topic	Label	$ N _{19}$	$ E _{19}$	$ N _{23}$	$ E _{23}$
Climate politics	<climate>	16,639	50,605	11,264	34,509
Immigration politics	<immigration>	6739	20,923	9519	36,889
Social security	<social>	13,926	39,778	11,395	34,173
Economic policy	<economy>	5880	10,962	10,031	37,111
Education policy	<education>	9843	20,314	12,814	35,619

Table D1. Values of γ demonstrate the higher chance of finding a politician from the cores versus peripheries

Topic	γ
<climate>	3.3
<immigration>	2.8
<social>	3.1
<economy>	4.0
<education >	3.6

Appendix D. Political actors in cores versus peripheries

To verify that the presence of political elite is higher in the cores than in peripheries, we computed the following for each network: How much more likely it was for a randomly selected node from the cores to belong to the list of political candidates compared to a node sampled from the peripheries, i.e.

$$\gamma = \frac{\text{Probability of core node being a candidate}}{\text{Probability of periphery node being a candidate}}$$

We report the values of γ for each network in Table D1.

Appendix E. Public amplification

See Appendix Figure E1.

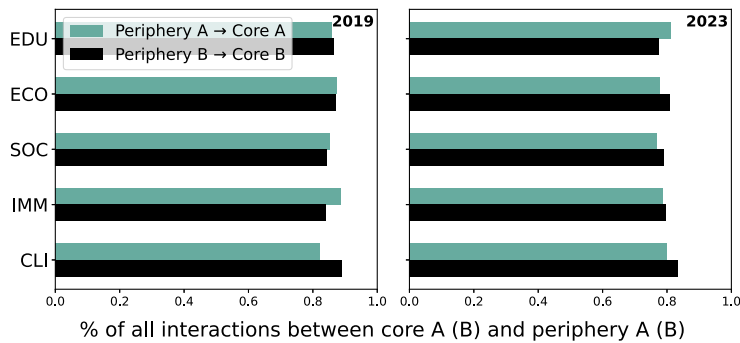


Figure E1. To view the core-periphery interactions as mass amplification, we confirm that the majority of connections are directed towards the core. In all networks examined, most of the links between the core and periphery consist of retweets originating from the outer periphery. This proportion remains relatively consistent across different polarized groups. Therefore, it is reasonable to characterize this dynamic as mass amplification.

Appendix F. Political parties in inferred polarized groups

We verified that the inferred groups can be called either left-leaning or right-leaning based on the real political affiliations of the political actors in the groups in Figure F1.

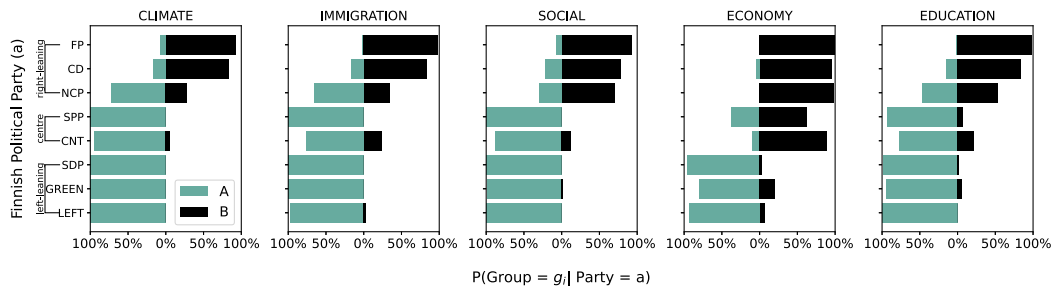


Figure F1. Majority of the political candidates from left-leaning parties are grouped together in the inferred polarized groups, as are those from right-leaning parties. The abbreviations used in the figure are expanded as follows: **FP**: finns party, **CD**: christian democrats, **NCP**: national coalition party, **SPP**: swedish people's party, **CNT**: center party, **SDP**: social democratic party, **GREEN**: green league & **LEFT**: left alliance.

Appendix G. Concentration of polarization in divided networks

See Appendix Figures G1 and G2.

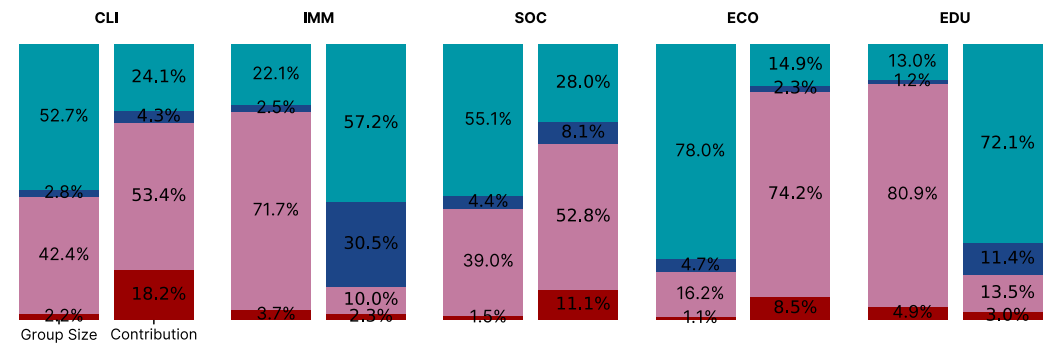


Figure G1. Concentration of polarization in 2019. Bars compare group sizes (left bar in each topic) with their respective contributions to polarization (right bar), revealing that smaller groups often exert disproportionately larger influence on observed polarization. For example, within the climate issue (CLI), the left elites constitute only 2.2% of the network size yet account for 18.2% of the polarization contribution. Note also how the right elites, despite representing only 2.5% of the immigration network, contribute over 12 times more to polarization relative to their size.

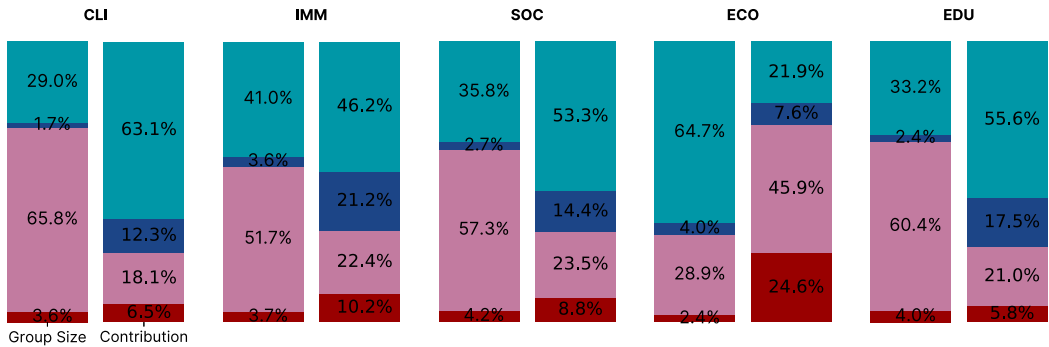


Figure G2. Concentration of polarization in 2023. Bars compare group sizes (left bar in each topic) with their respective contributions to polarization (right bar), revealing that smaller groups often exert disproportionately larger influence on observed polarization. For example, within the immigration issue (IMM), the right elites constitute 3.6% of the network size yet account for 21.2% of the polarization contribution. Left masses across topics tend to represent the largest subgroup but do not necessarily have the largest proportional contribution.

Appendix H. Activity patterns

See Appendix Figures H1, H2 and H3.

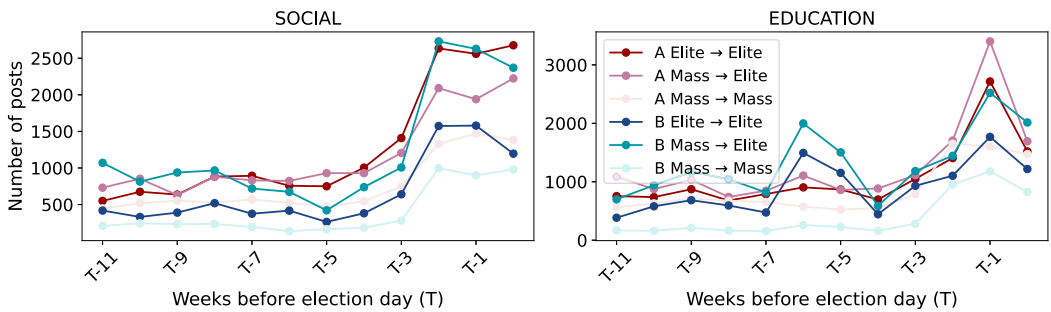


Figure H1. Activity patterns in <social> and <education> networks 2023. See caption of Figure 6 for more details.

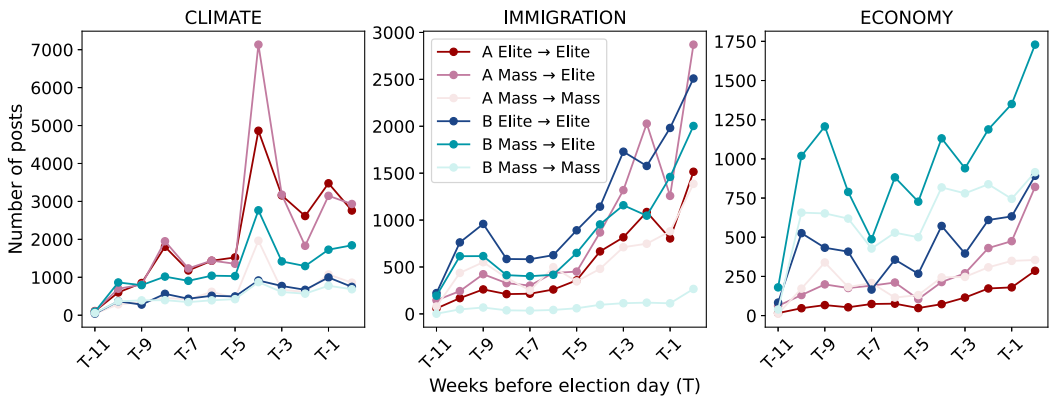


Figure H2. Activity patterns in <climate>, <immigration> and <economy> networks 2019. See caption of Figure 6 for more details.

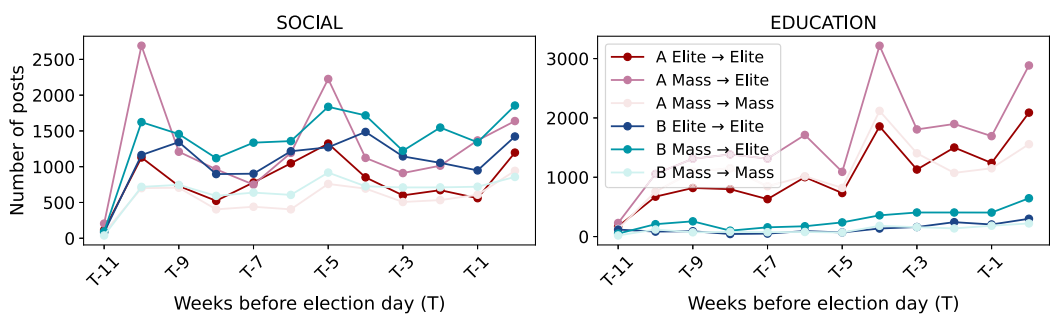


Figure H3. Activity patterns in <social> and <education> networks 2019. See caption of Figure 6 for more details.

Appendix I. Activity versus Marginal polarization

To evaluate the relationship between activity and marginal polarization, we conducted two complementary analyses. In both cases, we first applied *least trimmed squares regression* (Rousseeuw, 1984) in order to identify a robust set of points, and computed correlations only on this subset. In the first analysis, we measured the correlation between the absolute group-level values of activity and marginal polarization. In the second, we examined the correlation between their temporal differences, using relative changes for activity. Both analyses (Figures I1, I2) consistently indicate that neither activity itself, nor its change over time, is a reliable indicator of how polarized a group is.

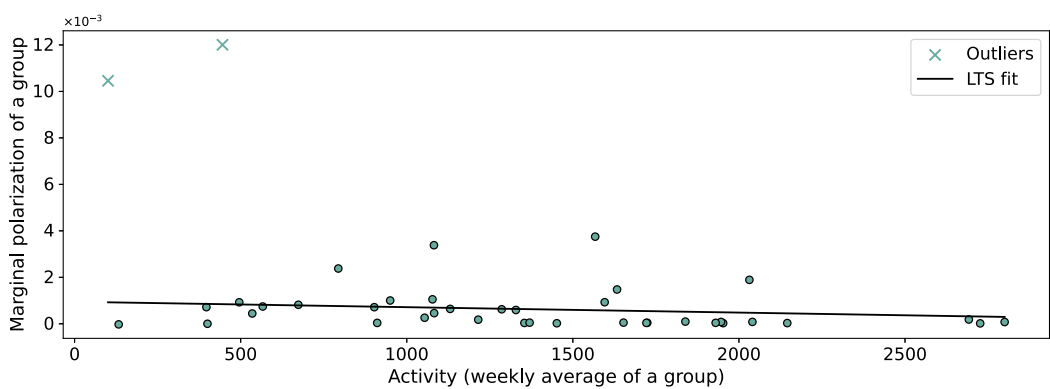


Figure I1. Each point corresponds to a hierarchical group (elite or mass within groups A and B) in a given topic and year. With five topics, four groups per topic, and two time points, the plot shows a total of 40 observations. For each observation, the horizontal axis shows the group's weekly mean activity, and the vertical axis shows the group's marginal polarization. The scatterplot shows that most groups cluster at low levels of marginal polarization, with just a few outliers reaching much higher values. To reduce the impact of these outliers, the fitted line is obtained with *least trimmed squares regression* (LTS) (Rousseeuw, 1984), which estimates the relationship using the 38 best-fitting points. Correlation coefficients, computed on this set, are low and negative (Pearson $r = -0.17$, $p = 0.30$; Kendall $\tau = -0.20$, $p = 0.08$), indicating at best weak evidence. Overall, there is no statistically reliable association between activity and marginal polarization.

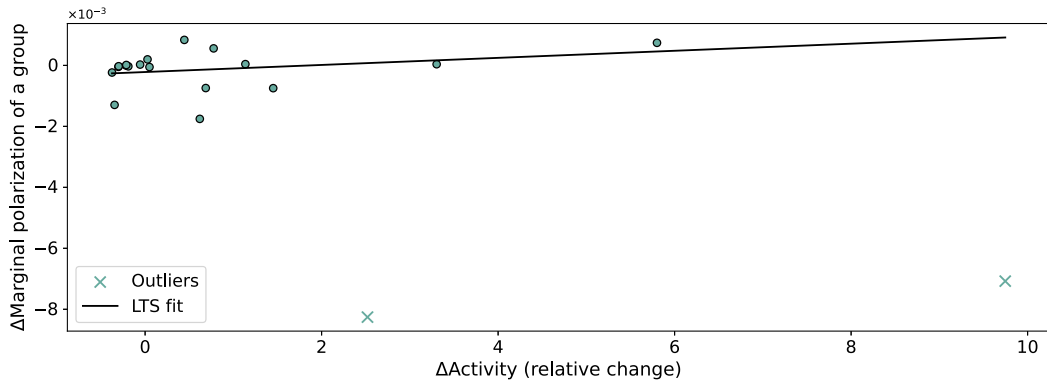


Figure I2. Each point corresponds to the change in activity and marginal polarization of a hierarchical group (elite or mass within groups A and B) between the two observed years. With five topics and four groups per topic, the plot shows a total of 20 paired observations. The horizontal axis shows the relative change in the group's weekly mean activity, and the vertical axis shows the corresponding change in marginal polarization. The results indicate weak associations (Pearson $r = 0.28$, $p = 0.26$; Kendall $\tau = 0.29$, $p = 0.10$). Overall, the evidence does not support a reliable link between changes in activity and changes in marginal polarization.