# Dynamics of Opinion Polarization

Elisabetta Biondi, Chiara Boldrini, Andrea Passarella, and Marco Conti

*Abstract*—For decades, researchers have been trying to understand how people form their opinions. This quest has become even more pressing with the widespread usage of online social networks and social media, which seem to amplify the already existing phenomenon of polarization. In this work, we study the problem of polarization assuming that opinions evolve according to the popular Friedkin–Johnsen (FJ) model. The FJ model is one of the few existing opinion dynamics models that has been validated on small/medium-sized social groups. First, we carry out a comprehensive survey of the FJ model in the literature (distinguishing its main variants) and of the many polarization metrics available, deriving an invariant relation among them. Second, we derive the conditions under which the FJ variants are able to induce opinion polarization in a social network, as a function of the social ties between the nodes and their individual susceptibility to the opinion of others. Third, we discuss a methodology for finding concrete opinion vectors that are able to bring the network to a polarized state. Finally, our analytical results are applied to two real social network graphs, showing how our theoretical findings can be used to identify polarizing conditions under various configurations.

*Index Terms*—Friedkin–Johnsen (FJ) model, opinion dynamics, polarization.

## I. INTRODUCTION

**W**ITH the rise of social media and online social networks, online interactions have started playing an increasingly important role in how people form their opinions, to the point that news consumption itself is now often mediated by social interactions [1], [2]. Social networks, though, do not merely provide a transparent technological substrate that facilitates interactions in the online dimension. Their algorithmic personalization, aimed at highlighting content that

is more interesting to each of us, effectively reinforces our cognitive biases, reducing the cognitive discomfort we experience when exposed to opinions challenging our beliefs but at the same time reducing the diversity and range of opinions we are exposed to. By reinforcing consonant opinions and downplaying, or even removing, discordant ones, social networks cradle us into curated filter bubbles and comfortable echo chambers. However, whether this leads to actual polarization [3], [4], [5] is still debated. Some argue that the very nature of social networks, i.e., the socialization of information consumption, may counteract the above effects [6], others that individual choices (to bond with similar others and to prefer concordant information) are more predominant than algorithmic filtering [7], others again that exposure to opposing views is more likely to actually backfire than to widen our perspectives [8]. To make matter worse, information may not only be partisan but it could also be blatantly fake [9].

This quest toward a better understanding of the impact of the social algorithm [9] and misinformation on our societies is ingrained with a more general question that, even when removing the cyber-dimension, still remains unsolved: how do people form their opinions? This question has fascinated sociologists and economists alike since much before the advent of the Internet, but it has recently gained new momentum, with computational sociologists and control theorists now weighing in. The literature on opinion dynamics is vast, with many models being proposed that aim at capturing a variety of cognitive and social mechanisms that lead to forming an opinion, such as social influence (which determines whose opinion you are affected by), cognitive dissonance (which triggers your willingness to adapt), anchoring to one's own opinion (which captures our prejudices). For an in-depth discussion, we refer the interested reader to recent surveys, such as [10], [11], and [12].

So-called *averaging models* are one of the most popular classes of such opinion dynamics models [13], [14], [15]. In these models, the final opinions (also known as expressed opinions) are a function of a repeated weighted averaging of the opinions of neighboring (in the influence graph) nodes. The strengths of averaging models lie in their mathematical tractability [16], ability to capture strong[1] opinion diversity [18], and their general flexibility (e.g., they can capture the *wisdom of the crowd* phenomenon [19] or include prominent agents [20], such as media sources and politicians that may be systematically biased and not willing to change their opinion at all).

The Friedkin–Johnsen (FJ) model [15] is the most popular averaging model in the related literature. It is the only

---

[1]This is in contrast with the weak opinion diversity generated by models like the Hegselmann–Krause model [17], where the final opinions form clusters in which every opinion is the same.

model that has been validated on small and medium-sized groups [21], [22], and even in human-AI group experiments [23]. Focusing on it, our first contribution is to provide a comprehensive review of all the major variants of the FJ model and of the polarization metrics described in the related literature. For them, we will highlight their key features and the differences between each other. We found that polarization metrics are linked together through an invariant relationship. As a second contribution, we derive the conditions under which the FJ model yields polarization, for each of the polarization metrics identified before. In addition, we also prove that the polarizing opinion vectors can be found analytically in most cases. All the results obtained are exploited to identify polarizing conditions, under different configurations, with two popular datasets of real social networks.

### A. Background and Motivation

The simplest averaging model is DeGroot's model [13], whereby the opinion of a node is simply the average opinion of its neighbors, weighted by the strength of their social influence. This model, however, is not considered realistic, since, when it converges (i.e., if the nodes' opinions stabilize), it always leads to *consensus*, i.e., to a final state in which all nodes have exactly the same opinion [24]. To overcome this problem, Friedkin and Johnsen [15] proposed a variation on Groot's model that introduces a certain degree of stubbornness in nodes. Their hypothesis is that a personal opinion always remains at least partly anchored to the initial opinion (or *prejudice*), more or less so depending on the individual's attitude to be influenced by others. The FJ model does not lead to consensus (except in very particular cases [25]) and has been widely popular in the related literature [16], [22], [25], [26], [27], [28], [29], [30], [31]. The FJ model has enjoyed two main avenues of research: on the one hand, the derivation of the conditions for convergence or consensus has been the main focus of the research efforts from the control theory domain [16], [31], [32], [33]. On the other hand, the graph-theoretical efforts [26], [27], [29], [30], [34], [35], [36] have been focused on understanding the effects of the underlying influence graph on opinion formation, polarization, and on how to interfere with the opinion formation process in order to obtain a desired outcome, (e.g., shifting the opinion in a specific direction, minimizing polarization and/or disagreement).

While all the above works refer to the opinion dynamics model they leverage as FJ, they are often relying on a simplified version of it. Specifically, they use the more mathematically tractable version [which we refer to, later on, as restricted FJ (rFJ)], which, however, is not able to capture polarization (we discuss this point later in this article). This has resulted in great confusion regarding which finding holds under which hypothesis. The second gap in the related literature, and a direct consequence of the above confusion, lies in whether the FJ model is actually able to capture polarization or not. Indeed, despite being opinion polarization a fundamental feature of a realistic opinion formation process, only Gionis et al. [26] and Dandekar et al. [35] have explicitly tackled this problem. Analyzing the problem on undirected

social networks, they have proved that two *variants* of the FJ model are neither capable of changing the average opinion of the social network nor of increasing the weighted difference of opinions among nodes of the same neighborhood. However, what happens with the general FJ model and with other polarization metrics is yet unknown.

## II. MODELING FRAMEWORK

We explicitly differentiate between the social graph and the influence graph. They both comprise the same set $\mathcal{V}$ of $n$ vertices and the same set of edges $\mathcal{E}$, but the weights of the edges are different and have different meanings. The social graph, denoted with $\mathcal{S}$, represents people (vertices) and the *social* relationships between them (through the edge weights $\hat{w}_{ij}$). The strength of a social relationship is typically measured in terms of the number of interactions that two people have [37] and for this reason the few results on polarization in the related works assume that the social graph is undirected [26], [35]. In this article, we will consider the general case of a directed social graph, specifying how results change in the specific case of an undirected one. The *influence graph* $\mathcal{I}$ describes how a node's opinion is influenced by that of its neighbors. The existence of an edge from node $i$ to node $j$ in $\mathcal{I}$ implies that node $j$ exerts an influence on the opinion of node $i$, and the strength of this influence is expressed by the edge weight $w_{ij}$. Lacking additional information, the influence graph can be derived from the social graph, leveraging the intuition that stronger social relationships will influence more than weak ones. Specifically, starting from the social weights $\hat{w}_{ij}$, the influence $w_{ij}$ can be computed as $w_{ij} = (\hat{w}_{ij}/[\sum_{j=1}^{n} \hat{w}_{ij}])$. Please note that this definition is the only one that allows a unique correspondence between all the variants of the FJ model. The matrix $W = (w_{ij})$ is called *influence matrix* and is assumed to be row stochastic (because it captures how the influence a node is subject to is split among its neighbors). The influence matrix is in general asymmetric (corresponding to a directed influence graph), even starting from a symmetric social matrix $\hat{W} = (\hat{w}_{ij})$, because the influence weight $w_{ij}$ expresses the relative importance of $j$ with respect to all $i$'s social relationships. Hence, the same social relationship intensity can weigh very differently depending on the strength of other relationships.

### A. Friedkin–Johnsen Family of Opinion Dynamics Models

A discrete-time opinion dynamics model tracks the evolution of $z_i(k)$, the opinion expressed by a node $i$ at time $k$. Opinions are generally assumed to be real valued, i.e., continuous in a certain reference interval. Similarly to the related literature [21], [26], [27], here we assume that opinions belong to $[-1, 1]$. Thus, extremes $-1$ and $1$ represent opposing viewpoints on an issue. For a given configuration of its input parameters, the model is said to be *convergent* if $z_i(k+1) \rightarrow z_i$ for all $i$ as $k$ grows to infinity. A convergent model is said to *reach consensus* if $z_i(k+1) \rightarrow z$ for all $i$ as $k$ grows to infinity. In the FJ model family, before the opinion formation process starts, each node $i$ has an initial opinion $s_i$, often referred to as *internal* or *fixed* opinion (or *prejudice*). In contrast, the opinion $z_i(k)$ is often referred to as the *expressed opinion* at time $k$.

TABLE I
FJ FAMILY OF MODELS

| | |
|---|---|
| gFJ: | $z_i(k+1) = (1-\lambda_i)s_i + \lambda_i \sum_{j \in \{i\} \cup N(i)} w_{ij} z_j(k)$   (1) |
| vFJ: | $z_i(k+1) = \dfrac{\hat{w}_{ii} s_i + \sum_{j \in N(i)} \hat{w}_{ij} z_j(k)}{\hat{w}_{ii} + \sum_{j \in N(i)} \hat{w}_{ij}}$   (2) |
| rFJ: | $z_i(k+1) = \dfrac{s_i + \sum_{j \in N(i)} \hat{w}_{ij} z_j(k)}{1 + \sum_{j \in N(i)} \hat{w}_{ij}}$   (3) |

In Table I we summarize the variants of the FJ models that can be found in the literature and we discuss them separately hereafter. We denote with $N(i)$ the neighborhood set of node $i$.

*1) Generalized Friedkin–Johnsen Model:* Equation (1) in Table I corresponds to the more general version of the model originally proposed by Friedkin and Johnsen [15]. The outermost weighted average depends on parameter $\lambda_i$, corresponding to the *susceptibility* of node $i$ to the opinions of other nodes. The innermost weighted average depends on the *influence* $w_{ij}$ that node $j$ exerts on node $i$. Two main mechanisms are at play here: *anchoring*, to the node $i$'s internal opinion $s_i$, and variable *susceptibility* $\lambda_i$, to other nodes' opinions. Nodes with zero susceptibility value are *stubborn* nodes and they never change their opinion. A common matrix formulation of the model is the following:

$$z(k+1) = (I-\Lambda)s + \Lambda W z(k) \quad (4)$$

where $\Lambda$ is a diagonal matrix containing the susceptibility values $\lambda_i$, while $W$ is the influence matrix. Note that the opinion of a node $i$ depends both on its initial prejudice $s_i$ (by a weight $1-\lambda_i$) and on its current opinion (by a weight $\lambda_i w_{ii}$). The only case in which this does not happen is when the node is stubborn ($\lambda_i = 0$) or when $w_{ii} = 0$. $\Lambda$ and $W$ are sometimes coupled via the condition $1-\lambda_i = w_{ii}$ [16], however we do not make this assumption here. The conditions under which the generalized FJ (gFJ) model achieves convergence and consensus have been thoroughly studied in the related literature [16], [25], [32]. A sufficient condition for convergence [31] is reported below, which we will use often in the rest of this article.

*Theorem 1 (Sufficient Condition for the gFJ):* If $\Lambda W$ is stable (i.e., has eigenvalues inside the open unit circle $\{z \in \mathbf{C} : |z| < 1\}$), the gFJ model is convergent and its only stationary point $z$ (i.e., steady-state solution) is given by the following:

$$z = (I-\Lambda W)^{-1}(I-\Lambda)s. \quad (5)$$

We refer the reader to SI Appendix for a brief summary of the main findings on the topic of opinion convergence.

*2) Variational Friedkin–Johnsen Model:* Dandekar et al. [35] and Matakos et al. [34] used a variant of FJ that we call the variational FJ model (vFJ), whose update function can be found in (2) of Table I. According to this model, the current opinion of a node is the weighted average between its prejudice and the current opinion of the other nodes. Thus, in this variant of the FJ

model, the current opinion of the node itself is not taken into account. We can formulate the expressed opinion in matrix form in the following way:

$$z = \left(D + \tilde{A} - A\right)^{-1} \tilde{A} s \quad (6)$$

where $D$ is the diagonal degree matrix ($\sum_j \hat{w}_{ij}$ for the $i$th diagonal element), $A$ is the adjacency matrix (whose $i,j$ element is $\hat{w}_{ij}$ and the diagonal is null), and $\tilde{A}$ is a diagonal matrix whose $i$th diagonal entry is equal to $\hat{w}_{ii}$. To model stubborn nodes, we can admit $\hat{w}_{ii}$ to be equal to $\infty$. In this case, matrix $\tilde{A}$ contains infinite values and (6) should be treated as discussed in SI Appendix. The relation between vFJ and gFJ has never been explicitly discussed in the related literature, where the two are implicitly treated as interchangeable and generically referred to as FJ model. However, the two models are not mathematically equivalent: the vFJ does not include node $i$'s current opinion $z_i$ in the averaging process, while gFJ pools both the initial opinion $s_i$ and the current opinion $z_i$.[2] The different flexibility of the two models becomes clear when observing that while the vFJ only features the matrix $\hat{W} = (\hat{w}_{ij})$ as parameters of the model (leading to a maximum $n^2$ degrees of freedom, with $n = |\mathcal{V}|$), the gFJ includes also matrix $\Lambda$, thus in total its degrees of freedom are $n^2 + n$. From a practical point of view, however, the only difference between the two models is the parameter $w_{ii}$, which, in gFJ, takes into account node $i$'s opinion $z_i$ in the averaging process, as we will see in the proof of Corollary 8.

*3) Restricted Friedkin–Johnsen Model:* The vFJ model with $\hat{w}_{ii}$ set to 1 as in (3) of Table I is very popular in the related literature, mainly due to its mathematical tractability. The model has been used in [26], [27], [29], [30], and [38]. The main difference between the rFJ and the vFJ model is the absence of the weight for $s_i$, so the parameters are only $\hat{w}_{ij}$ for all $i \neq j$ thus implying $n^2 - n$ degrees of freedom. Note that, since the weights $\hat{w}_{ij}$ are free to vary ($\hat{w}_{ij} \geq 0$), it is impossible to control the susceptibility (i.e., the importance of one's own initial opinion), even indirectly. A common matrix-formulation of the rFJ model is the following:

$$(D+I)z(k+1) = s + Az(k) \quad (7)$$

where $D$ and $A$ are defined as described for vFJ. The solution to the above problem can be written as $z = (L+I)^{-1}s$, where $L = D - A$ is the Laplacian matrix. The formulation of the rFJ model is particularly convenient from a mathematical standpoint (since $L+I$ is symmetric and many useful matrix formulas leverage symmetry), and this is the reason why it has been so often used in the related literature.

*4) Matrix Representation of the FJ Model:* In the whole set of FJ models, the final opinion $z$ of the opinion formation process can be expressed as $z = Hs$, where $H$ is a matrix that varies depending on the specific FJ version considered, whose formulas are summarized in Table II. In the remaining of this article, we will see that those matrices will be the key to the analysis of FJ polarization.

---

[2]Note that the coupling condition $\lambda_i = 1 - w_{ii}$ makes no sense for vFJ since the weight of node $i$'s current opinion is zero, so there is nothing to couple.

TABLE II
MATRIX $H$ FOR THE DIFFERENT FJ MODELS

| | | |
|---|---|---|
| gFJ: | $H_g = (I - \Lambda W)^{-1}(I - \Lambda)$ | (8) |
| vFJ: | $H_v = (D + \tilde{A} - A)^{-1}\tilde{A}$ | (9) |
| rFJ: | $H_r = (D + I - A)^{-1}$ | (10) |

### B. Polarization Metrics

In an opinion formation process, polarization is observed when there is a variation in a target index $\Phi$ (any, e.g., of the indices in Definition 2) between the initial opinion and the final opinion of nodes. A rigorous definition is provided in the following. Note that Definition 1 below is basically an abstraction of the polarization definitions in the related literature. In fact, while related works typically focus on a specific polarization metric and define polarization based on it, here we abstract the metric into the variable $\Phi$ and we provide a general definition that holds for all the polarization metrics discussed later on in Definition 2.

*Definition 1 (Polarization):* For a polarization index $\Phi$, we say that the opinion formation model $\mathcal{M}$ is $\Phi$-*polarizing* or *polarizing for* $\Phi$ if it exists at least an initial opinion vector $s$ such that the corresponding final opinion vector $z$ satisfies the following inequality:

$$\Phi(z) > \Phi(s). \tag{11}$$

In this case, we say that $s$ yields to $\Phi$-polarization and we call it *polarizing vector* or *polarizing prejudice*; its induced polarization is measured in terms of the polarization shift, i.e., by the function $\Delta_\Phi$ defined as follows:

$$\Delta_\Phi(s) = \Phi(z) - \Phi(s). \tag{12}$$

If the model $\mathcal{M}$ is not polarizing, we say that it is $\Phi$-*depolarizing* or *depolarizing for* $\Phi$.

Please observe that the definitions of polarizing and depolarizing model $\mathcal{M}$ are not symmetric: to depolarize, a model $\mathcal{M}$ should let opinions evolve in such a way that, at the end of the process, $\Phi$ is always decreasing for all possible choices of internal opinions ($s_i$); instead, $\mathcal{M}$ is polarizing if $\Phi$ does not decrease for at least one initial opinion vector $s$. The justification of the asymmetry lies in the importance of determining whether a model *can* capture the polarization phenomenon, which means that it does it in at least one case. Please note that, for the sake of brevity, in the following we may simply refer to the opinion vector as *opinion*, omitting the word "vector."

For the polarization index $\Phi$, the related literature has explored several different metrics, each capturing a different property of an opinion vector. Below we have collected the most popular definitions, for which we provide a short discussion.

*Definition 2:* For an opinion $x = (x_i) \in [-1, 1]^n$ the following polarization indices are defined:

$$\text{NDI}(x) = \sum_{(i,j)\in\mathcal{E}} w_{ij}(x_i - x_j)^2 \tag{13}$$

$$\text{GDI}(x) = \sum_{i,j\in\mathcal{V}:i<j} (x_i - x_j)^2 \tag{14}$$

$$P_1(x) = \sum_{i\in\mathcal{V}}(x_i - \bar{x})^2 = \|x - \bar{x}\|_2^2 \tag{15}$$

$$P_2(x) = \frac{1}{|\mathcal{V}|}\sum_{i\in\mathcal{V}} x_i^2 = \frac{1}{|\mathcal{V}|}\|x\|_2^2 \tag{16}$$

$$P_3(x) = \sum_{i\in\mathcal{V}} x_i^2 = \|x\|_2^2 \tag{17}$$

$$P_4(x) = \sum_{i\in\mathcal{V}} |x_i| = \|x\|_1. \tag{18}$$

The network-disagreement index (NDI) [27], [29], [30], [34], [35] is the sum, over all nodes, of the weighted disagreement in each node pair, which represents (except for the division by $n$) the average disagreement in the network as a whole. NDI is the only topology-dependent metric, in the sense that the same opinions may give rise to a completely different NDI depending on how the vertices are connected. The global disagreement index [35] (GDI) measures the conflict between all the users in the network, regardless of whether they share a social link or not. $P_1$ [29], corresponding to the mean-centered 2-norm of opinions, measures the polarization as a deviation of the opinions from the average. The definitions of $P_2$ [34] and $P_3$ [30], instead, intend the polarization as the deviation from the complete neutrality, represented with the value 0 (the middle ground between the two extremes - 1 and 1). Finally, $P_4$ is referred to as *total absolute opinion* and has been introduced by Friedkin and Johnsen [21]. While all previous indices were related to 2-norms, the total opinion is equivalent to the 1-norm. Similarly to $P_2$ and $P_3$, the index $P_4$ measures the "absolute total" opinion in the network and has the same semantic: it measures the deviation from the neutrality (represented by 0). While not directly a measure of polarization, the concept of choice shift caused by the opinion formation process (see definition below) is sometimes used in the related literature as an intermediate step in gauging the direction toward which opinion moves.

*Definition 3 (Choice Shift):* A choice shift occurs when the mean attitude of the group at the end is different from the mean attitude at the beginning

$$\sum_i z_i \neq \sum_i s_i. \tag{19}$$

The choice shift has been analyzed, for rFJ, by Gionis et al. in [26], where it is found that, if the social graph is undirected ($w_{ij} = w_{ji}$), changing the graph topology will not determine a choice shift. In the following, we will discuss if this finding carries over to gFJ and under which conditions.

*1) Polarization Invariants:* The above polarization indices have been introduced in the literature mostly as standalone metrics. In the remaining of the section, we establish equivalence relationships among them (Lemmas 1 and 2) and we derive a polarization invariant (Lemma 3).

*Lemma 1:* It holds that $\text{GDI}(x) = |\mathcal{V}| \cdot P_1(x)$, thus the two metrics GDI and $P_1$ are equivalent.

TABLE III
CLASSES OF POLARIZATION

| Type | What is captured | Indices |
|------|------------------|---------|
| **Local** | opinion spread among neighboring nodes | $NDI$ |
| **Dispersion** | opinion spread among all nodes | $GDI$, $P_1$ |
| **Absolute** | quadratic closeness to the extremes | $P_2$, $P_3$ |
| **Total** | linear closeness to extremes | $P_4$ |

*Proof:* See SI Appendix. ∎

*Lemma 2:* It holds that $P_3(x) = |\mathcal{V}| \cdot P_2(x)$, thus the two metrics $P_2$ and $P_3$ are equivalent.

*Proof:* Differently from Lemma 1, the proof is trivial and the thesis can be derived straightforwardly from Definition 2. ∎

Leveraging the results above, we can classify the polarization indices into four main classes of equivalence (Table III), in the sense that the behavior of a model is invariant in each class.

The four classes capture four different concepts of polarization. However, they are correlated by the following important invariant that will be used in the next section and whose proof is given in SI Appendix.

*Lemma 3 (Polarization Invariant):* For all opinion vectors $x$, the following inequality holds:

$$P_1(x) \geq P_3(x) - \frac{P_4(x)^2}{|\mathcal{V}|}. \tag{20}$$

From this relation, the following corollary follow, whose proof is provided in SI Appendix.

*Corollary 1:* When there is no choice shift the polarization moves in the same direction for both $P_1$ and $P_3$.

Corollary 1 says that, when there is no choice shift, the polarization with $P_1$ implies the polarization with $P_3$ and vice versa. A straightforward remark is that, in all the cases when the choice shift is null (for example when opinions are positive and $W$ is symmetric, as shown by Gionis et al. [26]), the Dispersion and Absolute classes of polarization are identical and represent the only global class of polarization.

## III. GFJ IS GLOBALLY POLARIZING BUT LOCALLY DEPOLARIZING

We start by focusing on the most general FJ model, the gFJ, and we investigate whether in this case the dynamics of the process lead to polarization or not.

### A. Polarization Under NDI

The first result is about the local polarization captured by the NDI index.

*Theorem 2 [gFJ (Local Polarization With NDI)]:* The gFJ model is always depolarizing with respect to NDI, in the sense that, for every prejudice $s$, we have that $\mathrm{NDI}(z) \leq \mathrm{NDI}(s)$.

*Proof:* As stated in Theorem 1, the gFJ model converges to the vector $z$ obtained from $z = (I - \Lambda W)^{-1}(I - \Lambda)s$. For each node $i$, consider the following cost function:

$$f_i(z_i) = (1 - \lambda_i)(s_i - z_i)^2 + \lambda_i \sum_{j=1}^n w_{ij}(z_i - z_j)^2 \tag{21}$$

which penalizes opinion $z_i$ if far from $s_i$ ($i$'s initial prejudice) and from $\sum_{j=1}^n w_{ij}z_j$ (the mean opinion of $i$'s neighborhood). We can prove that the expressed opinion $(z_i)_i$ of gFJ provided by (5) is the Nash Equilibrium of cost function (21) (for details, please refer to the SI), i.e., $z_i$ minimizes $f_i$ for all $i$, so that $f_i(z_i) \leq f_i(s_i)$ for all $i$. Since $\mathrm{NDI}(z)\lambda_i \leq \sum_i f_i(z_i)$ and $\mathrm{NDI}(s)\lambda_i = \sum_i f_i(s_i)$, we obtain that $\mathrm{NDI}(z) \leq \mathrm{NDI}(s)$ and, as a consequence, it follows that the gFJ is NDI-depolarizing. ∎

The result described above is intuitive: by definition, gFJ captures the willingness of each node to reduce the conflict (weighted by the matrix $W$) caused by the discordance of opinions with its neighbors, which is exactly what NDI measures. For this reason, the gFJ model is depolarizing in a local sense, but this however does not imply anything about global polarization. On the contrary, we will prove that gFJ can be polarizing at the global level depending on the interplay between the social network weights and nodes' susceptibility to the opinion of others. This is a key result, since it proves that gFJ does capture the polarization phenomenon in social networks.

### B. Polarization Under $P_2$, $P_3$, and $P_4$

We start by deriving the conditions under which gFJ is polarizing for the global metrics $P_2$, $P_3$, and $P_4$ (the proof is provided in SI Appendix).

*Theorem 3 [gFJ (Global Polarization With $P_2, P_3$, and $P_4$)]:* gFJ is *polarizing* with $P_2, P_3$, and $P_4$ if and only if matrix $H_g$ defined in Table I is not doubly stochastic (i.e., a square non-negative matrix, each of whose rows and columns sums to 1). Furthermore, we can distinguish the following two cases.

1) If there are naive nodes (i.e., $\exists i \in \mathcal{V} : \lambda_i = 1$), matrix $H_g$ is never doubly stochastic and thus gFJ is polarizing.
2) If there are no naive nodes (i.e., $\forall i \in \mathcal{V}, \lambda_i < 1$), matrix $H_g$ is not doubly stochastic, and equivalently gFJ is polarizing with $P_2, P_3$, and $P_4$, if and only if the following condition holds true for at least one node $i \in \mathcal{V}$:

$$\sum_{j \in \mathcal{V}} \frac{\lambda_j w_{ji}}{1 - \lambda_j} \neq \frac{\lambda_i}{1 - \lambda_i}. \tag{22}$$

Intuitively, the fact that $H_g$ is not double stochastic is a measure of the presence of nodes that are more influential than others. This is straightforward to see in the case of naive nodes [Theorem 3 1)], where all the non-naive nodes play the role of influencers (because they are always able to sway the naive nodes' opinions toward theirs), potentially increasing the polarization. When there are no naive nodes, the intuition behind Theorem 3 is more difficult to grasp. Let us split the effect of social influence and individual susceptibility. To isolate the former, let all nodes have the same susceptibility $\lambda$. Since (22) is reduced to $\sum_{j \in \mathcal{V}} w_{ji} \neq 1$, $W$ not being double stochastic becomes the condition for polarization, which corresponds to the case where the social influence out of any node $i$ is equivalent to the incoming social influence. However, in the general case, pure social influence is

dampened by individual susceptibility: stubborn nodes are not swayed, regardless of the social influence they are subject to. The condition in (22) exactly captures this interplay between susceptibility and social influence.

When gFJ is depolarizing, it is also unable to produce choice shift, as the following corollary states.

*Corollary 2:* When $H_g$ is doubly stochastic and thus gFJ is depolarizing with $P_2$, $P_3$, and $P_4$, for all opinion vector $s$, it holds that $P_4(z) = P_4(s)$ and $\sum_i z_i = \sum_i s_i$.

*1) How to Find Polarizing Opinion Vectors:* In Theorem 3, we have derived the sufficient and necessary condition for gFJ to be polarizing. We can give a first characterization of the polarizing vectors with $P_2$, $P_3$, and $P_4$: they can always be chosen with concordant entries (i.e., $\text{sgn}(s_i) = \text{sgn}(s_j) \ \forall i, j$). For a vector $x = (x_i)_i$, we will indicate with $x^{\text{abs}}$ the vector with positive entries given by $x^{\text{abs}} = (|x_i|)_i$. This is what is affirmed by the following theorem.

*Theorem 4 ($P_2$, $P_3$, and $P_4$ Polarizing Vectors Can be Obtained With Concordant Entries):* Whenever the model is polarizing with $P_i$, $i = 2, 3, 4$, and $s$ is a polarizing opinion vector, $\pm s^{\text{abs}}$ (that have concordant entries) are polarizing vectors inducing greater or equal (than that of $s$) polarization. Furthermore, if the network has no naive nodes and $\Lambda W$ is *irreducible* (i.e., the graph induced by $\Lambda W$ is strongly connected), the polarizing opinion vector that *maximises* the polarization has concordant entries.

This result is important and also pretty intuitive: since the $P_{2,3,4}$ polarization captures the shift from a neutral state (close to 0) to an extreme state (close to 1 or $-1$), the polarization calculated on the same vector with all the entries concordant must be greater or equal, because it is easier for nodes to cooperatively move toward the corresponding extreme. Instead, when entries are discordant, nodes have to mitigate between discordant opinions and thus are less free to vary in one of the two directions. This always occurs if the nodes are susceptible and nonstubborn, otherwise there would be disconnected communities and the cooperation would be impossible (this is what the conditions of the second part guarantee).

We can go one step further and provide (Theorems 5 and 6 below) concrete cases of initial opinion vectors under which gFJ is polarizing with $P_2$, $P_3$, and $P_4$. In the specific case of $P_2$ and $P_3$, we prove that finding the prejudice vector that yields maximum polarization, i.e., the maximum of function $\Delta_{P_{2,3}}$ defined in (12), is NP-hard, so we also discuss a possible approximation algorithm (Corollary 3).

*Theorem 5 [gFJ (Polarizing Initial Opinions for $P_2$ and $P_3$)]:* Whenever the model is polarizing for $P_2$ and $P_3$ (i.e., according to the conditions of Theorem 3), the polarizing prejudices $s_{B_2(1)}, s_{B_2(t)}, s_{\max P_{2,3}}$ can be derived as follows.

1) Two polarizing prejudices $\pm s_{B_2(1)}$ correspond to the unitary eigenvectors associated with the largest eigenvalue of matrix $H_g^T H_g$ and they correspond to the point of local maximum for the $P_2$, $P_3$-polarization on the $L_2$-ball of radius 1 $B_2(1) = \{x \in [0, 1]^n : \|x\|_2 \leq 1\}$. In particular, it holds exactly $\Delta_{P_3}(\pm s_{B_2(1)}) = \sigma_1^2 - 1 = \|H_g\|_2^2 - 1 = |\mathcal{V}|\Delta_{P_2}(\pm s_{B_2(1)})$, where $\sigma_1$ is the greatest singular value of the matrix $H_g$. Both these vectors have concordant entries.

2) The opinion vectors $\pm s_{B_2(t)}$ that yield the local maximum for $P_2$, $P_3$-polarization on the $L_2$-ball of radius $t$ $B_2(t) = \{x \in [0, 1]^n : \|x\|_2 \leq t\}$ are given by $\pm s_{B_2(t)} = \pm t \cdot s_{B(1)}$, where $t = 1/s_{B_2(1)}^{(k)}$, with $s_{B(1)}^{(k)}$ denoting the largest entry of $s_{B_2(1)}$. In particular, its polarization is exactly $t^2$ times the polarization of $s_{B_2(1)}$. Both these vectors have concordant entries.

3) The global maximum for $P_2$, $P_3$-polarization is achieved for the initial opinion vectors $\pm s_{\max P_{2,3}} = \pm \sum_i \alpha_i v_i$, whose components $\alpha = (\alpha_i)$ can be obtained as the solution to the following optimization problem:

$$\max \ \sum_i \alpha_i^2 \left( \sigma_i^2 - 1 \right)$$
$$\text{s.t.} \ \ 0 \leq B\alpha \leq 1 \qquad (23)$$

where $\sigma_1, \ldots, \sigma_n$ are the singular values of $H_g$, $\alpha = (\alpha_1, \ldots, \alpha_n)^T$ is the vector of the coefficients that express $s_{\max}$ with respect to the basis $\mathcal{B} = \{v_1,{}^3 \ldots, v_n\}$ composed of the unitary eigenvectors of $H_g^T H_g$, and $B$ is the matrix whose columns are the vectors of $\mathcal{B}$. The constraint guarantees that the solution $s_{\max P_{2,3}}$ has positive (and $-s_{\max P_{2,3}}$ has, respectively, negative) is a proper opinion vector in $[-1, 1]$ with concordant entries. This optimization problem, being quadratic nonconvex programming, is NP-hard.

Corollary 3 below tells us that, in case matrix $H_g$ has more than one singular value greater than one, it is possible to design subproblems of the optimal problem described in (23) over spaces larger than $B_2(t)$ but smaller than the entire domain. These subproblems are convex-quadratic programming and can be solved in polynomial time. Depending on the dimension of the network, numerical solutions may still not be found. Thus, we have designed a heuristic that always finds a solution $\pm s_{V_{>1}}^{heu}$ whose polarization is greater than that of $\pm s_{B_2(t)}$. The corresponding derivations can be found in SI Appendix.

*Corollary 3 [gFJ (Polarizing Initial Opinions for $P_2$ and $P_3$ on the Subspaces $V_{>1}$ and $V_{\geq 1}$)]:* When matrix $H_g$ has more than one singular value greater than one, it is possible to design subproblems of the optimal problem in (23) over $V_{>1}$ (vector space generated by the eigenvectors associated with the singular values strictly greater than 1) and over $V_{\geq 1}$ (vector space generated by the eigenvectors associated with the singular values greater or equal to 1). These subproblems yield polarizing vectors $s_{V_{>1}}, s_{V_{\geq 1}}$, respectively, and they are convex-quadratic programming (with polynomial time complexity). A heuristic that always finds a solution $\pm s_{V_{>1}}^{heu}$ is proposed.

With Theorem 5 and Corollary 3, we are able to identify the initial opinion vectors $\pm s_{\max P_{2,3}}, \pm s_{B_2(1)}, \pm s_{B_2(t)}, \pm s_{V>1}, \pm s_{V\geq 1}$ leading to polarization maxima on the corresponding subspaces. While computing the opinion $\pm s_{\max P_{2,3}}$ yielding the global maximum is an NP-hard problem [Theorem 5 3)], an approximate solution could be obtained using standard numerical solvers (not in all cases, as we discuss in the Experimental Evaluation section). The *local* polarization maxima are found

---

[3]Please observe that $v_1$ is the vector $s_{B_2(1)}$ because it is the unitary eigenvector corresponding to the largest singular value.

reducing the problem on the subspaces corresponding to eigenvectors of $H_g^T H_g$ associated with singular values strictly greater or weakly greater than one. In particular, the vectors $\pm s_{B_2(t)}$ of Theorem 5 2) [which are a scalar multiple of $\pm s_{B_2(1)}$ in Theorem 5 1)] are the vector that maximize the $P_2, P_3$-polarization on the space generated by the eigenvectors $\pm s_{B_2(1)}$ of $H_g^T H_g$ [also denoted with $v_1$ in Theorem 5 3)] that correspond to the singular value $\sigma_1 > 1$. The vectors $\pm s_{V_{>1}}$ in Corollary 3 maximize the polarization on the larger subspace $V_{>1}$ generated by all the eigenvectors that correspond to the singular values strictly greater than one. Finally, the vectors $\pm s_{V_{\geq 1}}$ in Corollary 3 maximize $P_2, P_3$-polarization on the even larger subspace $V_{\geq 1}$ generated by all the eigenvectors that corresponds to the singular values weakly greater than one. Since these vectors correspond to the maximum of polarization over subspaces that are subset of each other, it is trivial to derive the following inequality:

$$\Delta_\Phi\left(\pm s_{B_2(1)}\right) \leq \Delta_\Phi\left(\pm s_{B_2(t)}\right) \leq \Delta_\Phi\left(\pm s_{V_{>1}}\right) \leq$$
$$\leq \Delta_\Phi\left(\pm s_{V_{\geq 1}}\right) \leq \Delta_\Phi\left(\pm s_{\max_{P_{2,3}}}\right) \quad (24)$$

for $\Phi = P_2, P_3$.

While the results in Theorem 5 and Corollary 3 do not have an immediate practical interpretation, we can get the gist of them with a simple numerical example. Consider a network composed of three nodes—a naive node A, a node B with a susceptibility value equal to 0.5, and a stubborn node C—with mutual weights equal to 0.5. Applying Theorem 5, we obtain that $s_{B_2(1)} = (0, 0.30, 0.95)$ and, dividing by 0.95 as in Theorem 5 2), we obtain $s_{B_2(t)} = (0, 0.31, 1)$, which leads to a final opinion vector $(0.8, 0.61, 1)$. The prejudice of the naive node A is opposite to that of stubborn node C, and A's opinion shifts significantly (from 0 to 0.8). The opinion of the intermediate node B is approximately doubled. The opinion vector achieving maximum polarization $s_{\max_{P_{2,3}}}$ is instead $(0, 0.75, 1)$, whose corresponding final opinion is $(0.95, 0.89, 1)$. In this case, the combined effect of non-naive nodes' strong prejudices pushes A's final opinion to the opposite extreme. In some way, it is as if $s_{B_2(t)}$ (which only takes into account one singular value of $H$) selected the prejudice that maximizes the shift leveraging only to the most influential node (node C). Instead, the $s_{\max_{P_{2,3}}}$ (which yields the global maximum) is able to enforce a synergy between non-naive nodes. In this simple case since $H$ has only one singular value greater than 1, we cannot obtain the vectors $s_{V_{>1}}$ and $s_{V_{\geq 1}}$.

*Theorem 6 [gFJ (Polarizing Vectors for $P_4$)]:* Whenever the model is polarizing for $P_4$ (i.e., according to the conditions of Theorem 3), the following hold true.
  1) Two prejudice vectors $\pm s_{B_1(1)}$ that yields to $P_4$-polarization are the $j$th vector of the standard basis in $\mathbb{R}^n$ (i.e., a vector whose components are all zero, except the $j$th that equals 1) and its opposite, where $j = \arg\max_j \sum_i h_{ij}$ (i.e., $j$ corresponds to the index of the column of $H_g = \{h_{ij}\}_{ij}$ with the greatest column-sum). This prejudice vector is also the point of maximum of $P_4$-polarization on the 1-norm ball $B_1(1) = \{x \in [0, 1]^n : \|x\|_1 \leq 1\}$ and its polarization is exactly given by $\Delta_{P_4}(\pm s_{B_1(1)}) = \|H_g\|_1 - 1$.

2) With the same notations of Theorem 5, the global maximum for $P_4$-polarization is achieved for the initial opinion vectors $\pm s_{\max_{P_4}} = \pm \sum_i \alpha_i v_i$ with concordant entries, whose components $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_n)^T$ can be obtained as the solution to the following optimization problem:

$$\max \quad \sum_i \alpha_i\left(\sigma_i^2 - 1\right)\langle v_i, \mathbf{1}\rangle$$
$$\text{s.t.} \quad \mathbf{0} \leq B\boldsymbol{\alpha} \leq \mathbf{1}. \quad (25)$$

This optimization problem is a linear programming problem that can be numerically solved.

As observed for $P_2, P_3$-polarization, it is holds that

$$P_4\left(\pm s_{B_1(1)}\right) \leq P_4\left(\pm s_{\max_{P_4}}\right). \quad (26)$$

### C. Polarization Under $P_1$ and GDI

We conclude the analysis of gFJ by studying the polarization under $P_1$ and GDI. For this case, Theorem 7 asserts that whenever gFJ does not polarize in $P_2, P_3$, and $P_4$, it does not polarize in $P_1$, GDI either. Instead, when gFJ is polarizing in $P_2, P_3$, and $P_4$, we can guarantee that it also polarizes in $P_1$, GDI only if the sufficient condition in Theorem 7 is satisfied. Again, the proof of the theorem below can be found in SI Appendix.

*Theorem 7 [gFJ (Global Polarization With $P_1$, GDI)]:* For polarization indices $P_1$ and GDI, the following results hold.
  1) If gFJ is depolarizing for $P_2, P_3$, and $P_4$, then it is also depolarising for $P_1$, GDI.
  2) gFJ is polarizing in $s$ if the following condition holds true:

$$\sum_i \alpha_i^2\left(\sigma_i^2 - 1\right) \geq \frac{1}{n}\left[\sum_i |\alpha_i|\left(\sigma_i^2 - 1\right)\langle |v_i|, \mathbf{1}\rangle\right]$$
$$\cdot \left[\sum_i |\alpha_i|\left(\sigma_i^2 + 1\right)\langle |v_i|, \mathbf{1}\rangle\right] \quad (27)$$

where $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_n)^T$ is the expression of $s$ in terms of the basis $\mathcal{B}$ of the unitary eigenvectors of $H_g^T H_g$.

### D. Role of Stubborn and Naive Nodes

We now show (Corollary 4 below, proof in SI Appendix) a general result regarding stubborn nodes (i.e., nodes whose opinion is not at all swayed by that of their peers, which translates into $\lambda_i = 0$), whose role has not a direct impact on polarization. In fact, we will see that even if their strong *anchoring* attitude would intuitively suggest that they always have an effect on the final opinion, the network structure can instead invalidate it.

*Corollary 4:* While naive nodes tend to make polarization easier, stubborn nodes do not have a clear directional effect on the polarization with $P_2, P_3$, and $P_4$.

Leveraging Theorems 3 and 7, we can also study a special case involving naive nodes. This result, whose proof can be found in SI Appendix of this article, emphasizes the role of naive nodes (the ones with $\lambda_i = 1$), which essentially forget

their prejudice and move their opinion toward the opinion of the other nodes.

*Corollary 5:* Let us assume that the set of nodes $\mathcal{V}$ is composed of two disjoint groups, $\mathcal{I}$ and $\mathcal{J}$, such that all non-naive nodes have the same opinion $\tau$, while the naive nodes' opinions are free to vary in $[-1, 1]$, or equivalently

$$\forall i \in \mathcal{I} \quad \lambda_i = 1, s_i \in [-1, 1] \quad \forall j \in \mathcal{J} \quad \lambda_j < 1, s_j = \tau. \quad (28)$$

Then, the final opinion $z$ is exactly the vector $z = \tau \mathbf{1}$. In addition, this configuration is never polarizing for $P_1$ and GDI, while, as long as $|s_i| < 1$ for at least one node $i$, it always exists a $\tau$ value such that $P_2$, $P_3$, and $P_4$ are polarizing.

## IV. vFJ POLARIZES WHEN gFJ DOES

As already observed, in vFJ the opinion of a node $i$ at step $k$ does not take into account its own opinion at step $k - 1$ (as it happens, instead, with gFJ, which weighs it with $w_{ii}$). Thus, from a mathematical standpoint, the two models are different. However, apart from this specific contribution (i.e., in the case $w_{ii}$ is null), vFJ can be manipulated to exactly yield the same polarization as gFJ, if in an indirect and less intuitive way. In fact, in gFJ the susceptibility parameter directly captures the innate tendency of a node to be influenced (and to which degree) by others. In vFJ, instead, the rate at which a node is influenced by its peers is captured by: 1) the social strength of the node with all its neighbors $\hat{w}_{ij}$ and 2) the anchoring-degree of the node itself $\hat{w}_{ii}$, i.e., the importance it assigns to its initial prejudice.

Theorem 8 below establishes a complete equivalence, in terms of polarization properties, between gFJ and vFJ.

*Theorem 8 [vFJ (Local and Global Polarization)]:* For all polarization metrics, the vFJ model yields polarization under exactly the same conditions as gFJ. Specifically, if we replace matrix $H_g$ with the vFJ matrix $H_v$ and we set $\hat{w}_{ii} = 0$ for naive nodes (if present), the results of Theorems 3–7 and Corollaries 5 and 6 hold true. In particular, the condition for $H_v$ not being doubly stochastic reduces from (22) to the following one:

$$\frac{\sum_{j \neq i} \hat{w}_{ij}}{\hat{w}_{ii}} - \sum_{j \neq i} \frac{\hat{w}_{ji}}{\hat{w}_{jj}} \neq 0. \quad (29)$$

*Proof:* The proof consists in the derivation of vFJ from gFJ. This can be done using the following mapping:

$$\hat{w}_{ii} \neq \infty \rightarrow \begin{cases} \lambda_i = \frac{\sum_{k \in \mathcal{N}(i)} \hat{w}_{ik}}{\hat{w}_{ii} + \sum_{k \in \mathcal{N}(i)} \hat{w}_{ik}} \\ w_{ij} = \frac{\hat{w}_{ij}}{\sum_{k \in \mathcal{N}(i)} \hat{w}_{ik}} \end{cases} \quad (30)$$

$$\hat{w}_{ii} = \infty \rightarrow \begin{cases} \lambda_i = 0 \\ w_{ij} = 0. \end{cases} \quad (31)$$

Thus, the thesis follows from the results obtained for gFJ. ∎

Equation (29) simplifies when the social graph is undirected (which corresponds to the matrix $\hat{W}$ being symmetric). As stated in Corollary 6 below, in that case, when the self-weights are identical for all nodes (i.e., $\hat{w}_{ii} = \hat{w} \ \forall i$) vFJ is *never* polarizing in any metric and the average opinion is invariant to the opinion formation process.

*Corollary 6 (vFJ on Undirected Social Graphs):* When the social graph is undirected (i.e., matrix $\hat{W}$ is symmetric), vFJ is polarizing with $P_2$, $P_3$ and $P_4$ if and only if $\hat{w}_{ii}$ are not identical for all $i$. When $\hat{w}_{ii} = \hat{w} \ \forall i$, vFJ is *never* polarizing in any metric and it holds that $\sum_i z_i = \sum_i s_i$, i.e., there is never a choice shift in the network and the average final opinion is the same as the average initial opinion.

## V. rFJ MODEL IS NEVER POLARIZING IN UNDIRECT NETWORKS

In this section, we derive the results of polarization on the rFJ model. We already know from Bindel et al. [27] that rFJ does not polarize according to the local definition NDI, and, from Gionis et al. [26], that in the specific case of the undirected social graph it does not polarize according to the global definition $P_4$. Here, we generalize these findings. To this aim, note that rFJ is equivalent to vFJ after setting $w_{ii} = 1$. Thus, Theorem 8 also applies in this case. The condition for $H_r$ (the equivalent of $H_g$ but for rFJ) not being doubly stochastic, simply reduces from (29) to the following one:

$$\sum_{j \neq i} \hat{w}_{ij} - \sum_{j \neq i} \hat{w}_{ji} \neq 0. \quad (32)$$

And when the social graph is undirected, we obtain an even stronger result, summarized in Corollary 7 below.

*Corollary 7 (rFJ on Undirected Social Graphs):* The rFJ model is never polarizing, in any polarization metrics, for any initial opinion vector. In addition, it holds that $\sum_i z_i = \sum_i s_i$, i.e., there is never a choice shift in the network and the average final opinion is the same as the average initial opinion.

*Remark:* While polarization was still possible under vFJ on undirected social graph, with rFJ polarization never happens. The practical implication of this result for undirected social graphs is that polarization (in all its variations) can never be induced "naturally" by an opinion formation process following rFJ. Even more interestingly, polarization cannot be induced by altering the social graph (as long as it stays symmetric). Thus, when an initial state $s$ is given, the final state $z$ with rFJ can only naturally evolve toward nonpolarization. Vice versa, when the social graph is directed, the above result does not hold since, in a directed graph, the opinion of nodes with stronger social power tends to steer the opinion of the others. While relationship-oriented online social networks, like Facebook, tend to feature undirected graphs, directed social graphs are common in information-driven online social networks like Twitter.

## VI. EXPERIMENTAL EVALUATION

In this section we analyze the theoretical results on two real social network graphs: 1) the Karate Club graph [37] and 2) a Facebook graph [39]. The Karate Club dataset corresponds to an unweighted graph composed of 34 members. The Facebook dataset is a Facebook snapshot comprising 4039 users. Also in this case the graph is unweighted. After discarding isolated nodes (since they do not contribute at all to the opinion formation process), we end up with a network of 1519 nodes.

With these datasets we obtain the values $\hat{w}_{ij}$ that describe the social links between different users.

For both graphs, we obtain the influence matrix from the social matrix $\hat{W}$ normalizing by rows, i.e., $w_{ij} = (\hat{w}_{ij}/[\sum_k \hat{w}_{ik}])$. To proceed with the analysis we should set the susceptibility values of nodes, which are not fixed by the social network. To this aim, since both networks have a few very central nodes, as displayed in SI Appendix, we decided to use a centrality measure to set them. In the following, we will show the results obtained considering the PageRank centrality, which is the centrality measure that better captures the influence among nodes [16], but similar results hold with other centrality measures (betweenness, degree, eigenvector, and $k$-shell centrality). In our experiments, if the Pagerank centrality of a node $i$ is $C_i$, we assign $\lambda_i$ the value of $C_i$ (and $C_i^{-1}$) rescaled to $(0, 1)$, so that the more central the nodes (and, respectively, the less central), the higher their susceptibility values. Furthermore, we will also show the case in which all nodes have the same susceptibility, set to 0.8. In SI Appendix, we provide a visualization of the social networks we consider and of the susceptibility values obtained in this way for both datasets.

We can now search for the initial opinion vectors that yield polarization in the social network, by applying Theorems 5 and 6 and Corollary 3. For the sake of brevity, in the following we will consider only the positive polarizing vectors but analogous results can be obtained for negative ones, as stated in the corollaries. We compute the polarizing vectors $s_{B_2(1)}$, $s_{B_2(t)}$, $s_{\max P_{2,3}}$, $s_{V>1}$, $s_{V>1}^{heu}$, $s_{B_1(1)}$, $s_{\max P_4}$ as described in Theorems 5 and 6 and Corollary 3, and we compare their polarization with the one of an opinion vector $s_{\text{unif}}$ with entries randomly drawn from a uniform distribution in [0, 1]. Table IV shows the polarization induced by the above vectors on the Karate social network, for the three susceptibility configurations we are considering. Recall that the polarization shift $\Delta_\Phi(s)$ for a given polarization metric $\Phi$ (with $\Phi = P_1, \ldots, P_4, \text{NDI}, \text{GDI}$) and initial opinion $s$ is derived as $\Phi(Hs) - \Phi(s)$. When $\Delta_\Phi(s)$ is positive, then gFJ polarizes in $s$. We can see in Table IV that the theoretical results are confirmed (this is not surprising, since our theorems are obtained without any approximation). A random prejudice vector $s_{\text{unif}}$ leads to depolarization for all the polarization metrics. Instead, the prejudices from Theorem 5 and Corollary 3 yield to $P_2, P_3$-polarization. As expected, according to (24), their corresponding $P_2, P_3$-polarization shifts are progressively increasing moving from $s_{B_2(1)}$ to $s_{\max P_{2,3}}$ (because the solution is searched for into a larger domain). Note that, since the network is small, the numerical solver is able to find the solutions $s_{\max P_{2,3}}$ and $s_{V>1}$ (the latter is not applicable to the case $\lambda_i = 0.8$, because its $H_g$ has only one singular value greater than 1). It is interesting to observe that the solution $s_{V>1}^{heu}$ found with the heuristic is, in one case, exactly equal to the one obtained numerically ($s_{V>1}$) and, in the other case, extremely close to it, which confirms the heuristic validity. The prejudice vectors found according to Theorem 6, instead, yield to $P_4$-polarization, and satisfy the inequality in (26). With respect to $P_1$, GDI-polarization, while Theorem 7 cannot tell us whether polarization is achieved in general, we can use it to predict

### TABLE IV
GFJ IN THE KARATE NETWORK: VALUES OF $\Delta_\Phi$ FOR ALL POLARIZATION METRICS, FOR THE THREE $\lambda_i$ CONFIGURATIONS. THE SHADOWED AREA HIGHLIGHTS WERE THE CORRESPONDING OPINION VECTORS $s$ ARE EXPECTED TO YIELD POLARIZATION

| | $\Delta_{P_1}$ | $\Delta_{P_2}$ | $\Delta_{P_3}$ | $\Delta_{P_4}$ | $\Delta_{NDI}$ | $\Delta_{GDI}$ |
|---|---|---|---|---|---|---|
| $\lambda_i \propto P_i$ | | | | | | |
| $s_{unif}$ | $-8.8e{-1}$ | $-1.7e{-2}$ | $-5.7e{-1}$ | $3.1e{-1}$ | $-1.03$ | $-29.93$ |
| $s_{B_2(1)}$ | $-6.4e{-2}$ | $2.2e{-3}$ | $7.6e{-2}$ | $4.2e{-1}$ | $-3.1e{-1}$ | $-2.18$ |
| $s_{B_2(t)}$ | $-1.15$ | $4.0e{-2}$ | $1.36$ | $1.77$ | $-5.61$ | $-38.94$ |
| $s_{V>1}$ | $-1.18$ | $4.1e{-2}$ | $1.40$ | $1.80$ | $-5.67$ | $-40.28$ |
| $s_{V>1}^{heu}$ | $-1.18$ | $4.1e{-2}$ | $1.40$ | $1.80$ | $-5.67$ | $-40.28$ |
| $s_{\max P_{2,3}}$ | $-1.81$ | $6.0e{-2}$ | $2.05$ | $2.05$ | $-6.94$ | $-61.48$ |
| $s_{B_1(1)}$ | $-2.0e{-1}$ | $-5.4e{-3}$ | $-1.9e{-1}$ | $1.6e{-1}$ | $-1.3e{-1}$ | $-6.63$ |
| $s_{\max P_4}$ | $-3.57$ | $3.4e{-2}$ | $1.16$ | $2.65$ | $-10.65$ | $-121.421$ |
| $\lambda_i \propto P_i^{-1}$ | | | | | | |
| $s_{unif}$ | $-1.62$ | $-4.8e{-2}$ | $-1.63$ | $-1.1e{-2}$ | $-1.41$ | $-55.22$ |
| $s_{B_2(1)}$ | $6.8e{-1}$ | $4.3e{-2}$ | $1.45$ | $3.32$ | $-3.1e{-1}$ | $23.19$ |
| $s_{B_2(t)}$ | $7.6e{-1}$ | $4.3e{-2}$ | $1.63$ | $3.52$ | $-3.5e{-1}$ | $25.97$ |
| $s_{V>1}$ | $3.0e{-1}$ | $1.3e{-1}$ | $4.48$ | $7.50$ | $-5.08$ | $10.07$ |
| $s_{V>1}^{heu}$ | $3.1e{-1}$ | $1.1e{-1}$ | $3.61$ | $6.74$ | $-3.72$ | $10.45$ |
| $s_{\max P_{2,3}}$ | $-2.30$ | $2.01e{-1}$ | $6.86$ | $6.86$ | $-3.98$ | $-78.17$ |
| $s_{B_1(1)}$ | $3.1e{-1}$ | $2.2e{-2}$ | $7.5e{-1}$ | $2.98$ | $-1.95$ | $10.67$ |
| $s_{\max P_4}$ | $-4.27$ | $1.3e{-1}$ | $4.59$ | $10.04$ | $-9.36$ | $-145.22$ |
| $\lambda_i = 0.8$ | | | | | | |
| $s_{unif}$ | $-2.78$ | $-7.7e{-2}$ | $-2.62$ | $1.7e{-1}$ | $-2.45$ | $-94.67$ |
| $s_{B_2(1)}$ | $-3.3e{-1}$ | $1.5e{-1}$ | $5.2e{-1}$ | $2.44$ | $-1.22$ | $-11.30$ |
| $s_{B_2(t)}$ | $-1.43$ | $6.6e{-2}$ | $2.23$ | $5.06$ | $-5.22$ | $-48.49$ |
| $s_{\max P_{2,3}}$ | $-1.73$ | $1.3e{-1}$ | $4.57$ | $4.57$ | $-3.33$ | $-58.73$ |
| $s_{B_1(1)}$ | $-8.2e{-1}$ | $-1.5e{-2}$ | $-5.2e{-1}$ | $2.34$ | $-5.60$ | $-27.90$ |
| $s_{\max P_4}$ | $-6.86$ | $-1.7e{-1}$ | $-1.7e{-1}$ | $8.10$ | $-13.32$ | $-233.38$ |

whether $P_1$, GDI-polarization is achieved with the same prejudices that yield $P_2, P_3$, or $P_4$ polarization. We find that the condition (sufficient for polarization) of Theorem 7 is verified only for the $P_2, P_3$-polarizing prejudices and $\lambda_i \propto C_i^{-1}$. The columns $\Delta_{P_1}$ and $\Delta_{\text{GDI}}$ of Table IV confirm polarization in these cases. Finally, as expected from Theorem 2, gFJ is always depolarizing in NDI.

Similar results are obtained with the Facebook network (Table V). Two points are worth emphasizing. First, the centrality of nodes in the Facebook graph is extremely skewed, with one very central node dominating the graph. Thus, when $\lambda_i \propto C_i$, there are very few susceptible nodes and polarization is harder to achieve. The opposite effect is observed when $\lambda_i \propto C_i^{-1}$, and the polarization shifts are higher. Second, note that since the Facebook network size is large, the global solutions ($s_{\max P_{2,3}}$ and $s_{\max P_4}$) could not be found numerically and $s_{V>1}$ could only be obtained for $\lambda_i \propto C_i$. This example showcases the importance of the heuristics derived in the previous section, which can always return a polarizing vector.

We conclude this section by having a closer look at how polarizing prejudices are structured. In Fig. 1(a), each arrow corresponds to one node in the Karate graph, and it starts at its prejudice and ends at its final opinion. For $P_2, P_3$, and $P_4$, an increase in polarization is linked, intuitively, to some opinions moving from more neutral states (close to 0) to more extreme states (close to 1). Indeed, this is what happens in all the cases presented in the figure. In particular, the vectors $s_{\max P_{2,3}}$ and $s_{\max P_4}$ that maximize the polarization feature the maximum number of components with initial opinion equal to 1 (with
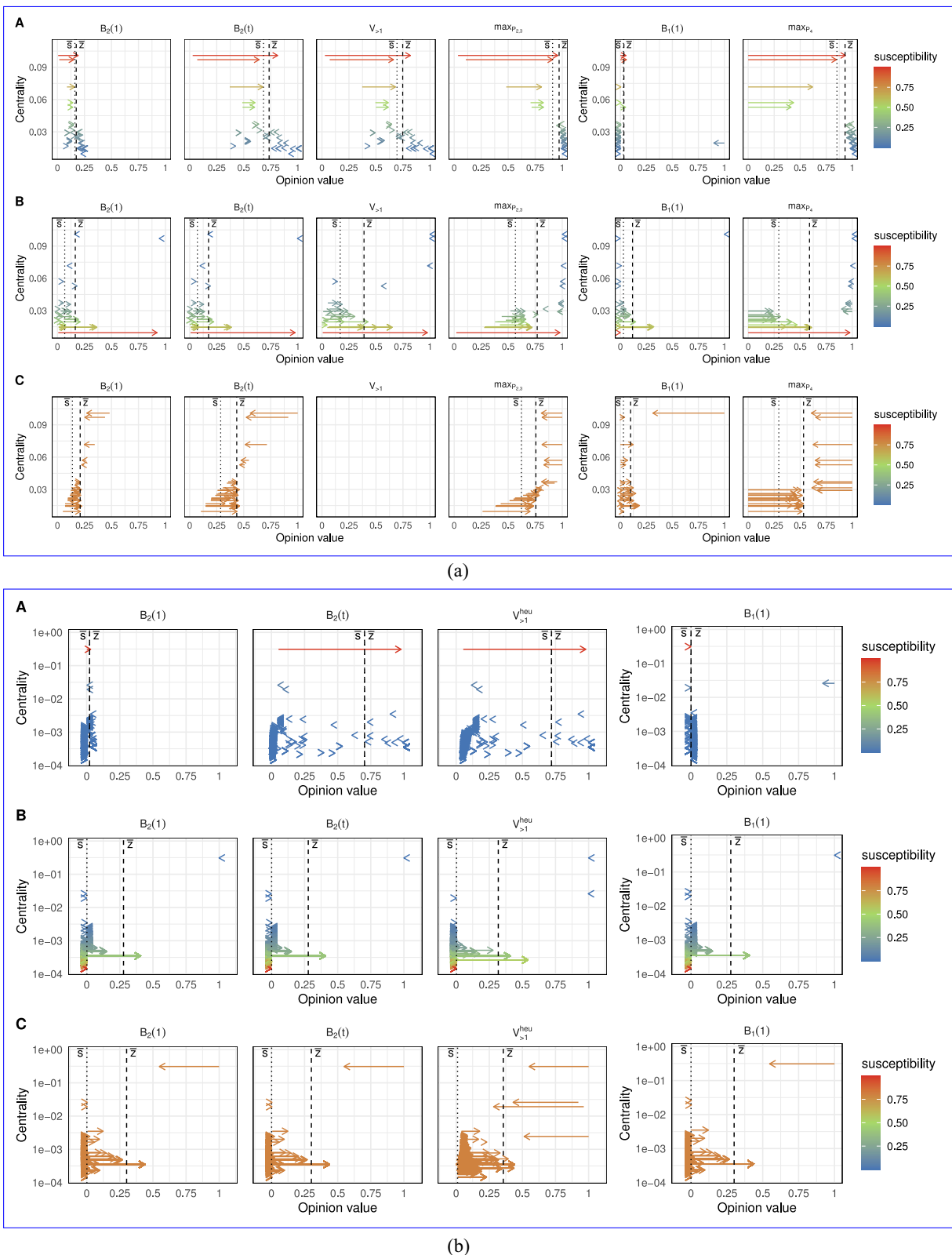
Fig. 1.　Polarizing opinions versus nodes centrality. Each arrow corresponds to one node in the graph. An arrow starts at the initial prejudice and ends at the final opinion. The color of the arrow corresponds to the susceptibility assigned to the node. (A) $\lambda_i \propto C_i$, (B) $\lambda_i \propto C_i^{-1}$, and (C) $\lambda_i = 0.8$. In each panel, on the left the opinions yielding $P_2$, $P_3$-polarization, on the right the opinions yielding $P_4$-polarization. With the dotted and dashed line we denote the average initial and final opinion, respectively. (a) gFJ on the Karate network. (b) gFJ on the Facebook network.

respect to the other opinion vectors): in this way, the nodes with more extreme opinions work synergistically to push the others' opinions closer to theirs. For selecting such an optimal

"cooperative" group of extreme nodes, one should be able to search for a solution to the optimization problem within the entire domain of opinions. When this is not the case, only

TABLE V
GFJ IN THE FACEBOOK NETWORK: VALUES OF $\Delta_\Phi$ FOR ALL
POLARIZATION METRICS, FOR THE THREE $\lambda_i$ CONFIGURATIONS.
THE SHADOWED AREA HIGHLIGHTS WERE THE CORRESPONDING
OPINION VECTORS ARE EXPECTED TO YIELD POLARIZATION

| | $\Delta_{P_1}$ | $\Delta_{P_2}$ | $\Delta_{P_3}$ | $\Delta_{P_4}$ | $\Delta_{NDI}$ | $\Delta_{GDI}$ |
|---|---|---|---|---|---|---|
| | | | $\lambda_i \propto C_i$ | | | |
| $s_{unif}$ | $-6.2e{-}1$ | $-1.7e{-}4$ | $-2.6e{-}1$ | $3.6e{-}1$ | $-0.31$ | $-940$ |
| $s_{B_2(1)}$ | $-3.6e{-}4$ | $5.5e{-}7$ | $8.4e{-}4$ | $2.8e{-}2$ | $-8.8e{-}4$ | $-5.5e{-}1$ |
| $s_{B_2(t)}$ | $-3.8e{-}1$ | $5.8e{-}4$ | $8.7e{-}1$ | $8.9e{-}1$ | $-9.2e{-}1$ | $-557$ |
| $s_{V>1}$ | $-4.1e{-}1$ | $5.8e{-}4$ | $8.8e{-}1$ | $9.0e{-}1$ | $-9.3e{-}1$ | $-630$ |
| $s_{V>1}^{heu}$ | $-4.1e{-}1$ | $5.8e{-}4$ | $8.8e{-}1$ | $9.0e{-}1$ | $-9.3e{-}1$ | $-630$ |
| $s_{B_1(1)}$ | $-1.6e{-}1$ | $-1.0e{-}4$ | $-1.6e{-}1$ | $6.1e{-}2$ | $-2.7e{-}1$ | $-240$ |
| | | | $\lambda_i \propto C_i^{-1}$ | | | |
| $s_{unif}$ | $-64.51$ | $-1.2e{-}1$ | $-180.64$ | $-129.81$ | $-9.45$ | $-97985$ |
| $s_{B_2(1)}$ | $53.83$ | $1.1e{-}1$ | $171.10$ | $419.61$ | $-6.3e{-}2$ | $81774$ |
| $s_{B_2(t)}$ | $53.94$ | $1.1e{-}1$ | $171.45$ | $420.04$ | $-6.3e{-}1$ | $81941.1$ |
| $s_{V>1}^{heu}$ | $49.80$ | $1.34e{-}1$ | $203.59$ | $497.04$ | $-1.35$ | $75646$ |
| $s_{B_1(1)}$ | $53.73$ | $1.1e{-}1$ | $170.75$ | $420.62$ | $-6.3e{-}1$ | $81615$ |
| | | | $\lambda_i = 0.8$ | | | |
| $s_{unif}$ | $-109.77$ | $-1.6e{-}1$ | $-243.81$ | $152.36$ | $-22.71$ | $-166737$ |
| $s_{B_2(1)}$ | $60.34$ | $1.3e{-}1$ | $198.13$ | $455.2$ | $-1.02$ | $91656$ |
| $s_{B_2(t)}$ | $60.44$ | $1.3e{-}1$ | $198.44$ | $455.56$ | $-1.02$ | $91693$ |
| $s_{V>1}^{heu}$ | $25.47$ | $1.4e{-}1$ | $217.38$ | $523.82$ | $-4.03$ | $38687$ |
| $s_{B_1(1)}$ | $60.36$ | $1.3e{-}1$ | $197.81$ | $455.93$ | $-1.03$ | $91693$ |

suboptimal polarization is achieved. For example, the vectors $s_{B_2(1)}, s_{B_2(t)}, s_{B_1(1)}$, only manage to select one single extreme node responsible for pushing the more neutral opinions of others, while $s_{V>1}$ is in an intermediate position, being able to select more extreme nodes than $s_{B_2(t)}$ and fewer than $s_{\max P_{2,3}}$. We can also observe that in panels A and B of Fig. 1(a), where the susceptibility varies across nodes, the nodes with initial opinion 1 are always the most stubborn, so that they create a field of attraction for more susceptible nodes. Effectively, the susceptibility assigned to nodes overrides their centrality in the network, hence very central nodes can become attractors or attractees depending on how stubborn they are. Vice versa, when the susceptibility of all nodes is the same [panel C of Fig. 1(a)], we observe the unfiltered effect of centrality: the most polarizing prejudices are those in which the most central nodes have initial opinions close to 1, and their final opinion changes much less than the others' opinions. This also confirms that the PageRank centrality is able to capture the ability of nodes to convince the others, and thus it identifies the most influential nodes.

In Fig. 1(b) we can see the results obtained with the Facebook network. In this case, since the network is large, it is not possible to find the global solutions $s_{\max P_{2,3}}$ and $s_{\max P_4}$. However, the considerations we made for the Karate graph hold also in this case. In particular, the polarizing vectors assign to more stubborn nodes initial opinions closer to 1, so that they can influence susceptible nodes to which they are connected. In the Facebook network, though, due to the scale-free topology with just a few hubs and many poorly connected nodes, we also observe very susceptible nodes that do not change their opinions (Fig. 1(b), panel B). These nodes have typically a single edge toward a stubborn node sharing its opinion.

## VII. CONCLUSION

In this work, we have investigated under which conditions the popular FJ model yields polarized opinions. The first contribution of the work has been to systematize the variety of FJ models used in the literature, and the many definitions of polarization. Then, as the main contribution of the work, we have derived the conditions under which the FJ models yield to polarization, for each of the polarization classes identified from the related literature. Moreover, we have identified a methodology for obtaining polarizing prejudices in most cases. When exact solutions could not be found (because the corresponding problem was NP-hard), we have defined heuristics to find a suboptimal solution. Our theoretical results have then been tested on two real-life social networks. We have seen that both the centrality of nodes in the social network as well as their individual susceptibility to the opinions of other nodes play a key role in defining their influence power, hence their ability to polarize.

The results presented in this work can be used to understand under which conditions polarization of opinions will emerge for a given social network. While the application to online social networks immediately comes to mind (as showcased in Section VI, the social graph can be collected from online social network platforms, such as Twitter, Facebook, Reddit, etc.), other applications can be foreseen, such as failure mode and effect analysis in reliability engineering [40]. In addition, the results presented in this article can be exploited to design interventions to bring polarization under control. More in general, since opinions in the FJ model are actually abstracted as values in the $[0, 1]$ or $[-1, 1]$ domain, the FJ model could be used to study information propagation, the evolution of decision processes, and consensus/polarization on networks, as long as the mapping in the same unidimensional domain remains appropriate.

## REFERENCES

[1] A. Hermida, F. Fletcher, D. Korell, and D. Logan, "Share, like, recommend: Decoding the social media news consumer," *Journalism Stud.*, vol. 13, nos. 5–6, pp. 815–824, 2012.

[2] J. Gottfried and E. Shearer, *News Use Across Social Medial Platforms 2016*. Washington, DC, USA: Pew Res. Center, 2016.

[3] D. Spohr, "Fake news and ideological polarization: Filter bubbles and selective exposure on social media," *Bus. Inf. Rev.*, vol. 34, no. 3, pp. 150–160, 2017.

[4] M. P. Fiorina and S. J. Abrams, "Political polarization in the American public," *Annu. Rev. Polit. Sci.*, vol. 11, pp. 563–588, 2008.

[5] M. D. Conover, J. Ratkiewicz, M. Francisco, B. Gonçalves, F. Menczer, and A. Flammini, "Political polarization on Twitter," in *Proc. 5th Int. AAAI Conf. Weblogs Soc. Media*, 2011, pp. 89–96.

[6] S. Messing and S. J. Westwood, "Selective exposure in the age of social media: Endorsements trump partisan source affiliation when selecting news online," *Commun. Res.*, vol. 41, no. 8, pp. 1042–1063, 2014.

[7] E. Bakshy, S. Messing, and L. A. Adamic, "Exposure to ideologically diverse news and opinion on Facebook," *Science*, vol. 348, no. 6239, pp. 1130–1132, Jun. 2015.

[8] C. A. Bail et al., "Exposure to opposing views on social media can increase political polarization," *Proc. Nat. Acad. Sci. U S A*, vol. 115, no. 37, pp. 9216–9221, 2018.

[9] D. Lazer, "The rise of the social algorithm," *Science*, vol. 348, no. 6239, pp. 1090–1091, 2015.

[10] Y. Dong, M. Zhan, G. Kou, Z. Ding, and H. Liang, "A survey on the fusion process in opinion dynamics," *Inf. Fusion*, vol. 43, pp. 57–65, Sep. 2018.

[11] B. D. O. Anderson and M. Ye, "Recent advances in the modelling and analysis of opinion dynamics on influence networks," *Int. J. Autom. Comput.*, vol. 16, no. 2, pp. 129–149, 2019.

[12] H. Hassani, R. Razavi-Far, M. Saif, F. Chiclana, O. Krejcar, and E. Herrera-Viedma, "Classical dynamic consensus and opinion dynamics models: A survey of recent trends and methodologies," *Inf. Fusion*, vol. 88, pp. 22–40, Dec. 2022.

[13] M. H. DeGroot, "Reaching a consensus," *J. Amer. Stat. Assoc.*, vol. 69, no. 345, pp. 118–121, 1974.

[14] J. R. French Jr., "A formal theory of social power," *Psychol. Rev.*, vol. 63, no. 3, p. 181, 1956.

[15] N. E. Friedkin and E. C. Johnsen, "Social influence and opinions," *J. Math. Sociol.*, vol. 15, nos. 3–4, pp. 193–206, 1990.

[16] A. V. Proskurnikov and R. Tempo, "A tutorial on modeling and analysis of dynamic social networks. Part I," *Annu. Rev. Control*, vol. 43, pp. 65–79, May 2017.

[17] R. Hegselmann and U. Krause, "Opinion dynamics and bounded confidence: Models, analysis and simulation," *J. Artif. Soc. Soc. Simul.*, vol. 5, pp. 1–33, Jun. 2002.

[18] M. Mäs, A. Flache, and J. A. Kitts, "Cultural integration and differentiation in groups and organizations," in *Perspectives on Culture and Agent-Based Simulations: Integrating Cultures*. Cham, Switzerland: Springer, 2014, pp. 71–90.

[19] A. Das, S. Gollapudi, R. Panigrahy, and M. Salek, "Debiasing social wisdom," in *Proc. 19th ACM KDD*, 2013, pp. 500–508.

[20] D. Acemoglu and A. Ozdaglar, "Opinion dynamics and learning in social networks," *Dyn. Games Appl.*, vol. 1, no. 1, pp. 3–49, Mar. 2011.

[21] N. E. Friedkin and E. C. Johnsen, *Social Influence Network Theory: A Sociological Examination of Small Group Dynamics*, vol. 33. Cambridge, U.K.: Cambridge Univ. Press, 2011.

[22] N. E. Friedkin and F. Bullo, "How truth wins in opinion dynamics along issue sequences," *Proc. Nat. Acad. Sci. U S A*, vol. 114, no. 43, pp. 11380–11385, Oct. 2017.

[23] O. Askarisichani, F. Bullo, N. E. Friedkin, and A. K. Singh, "Predictive models for human–AI nexus in group decision making," *Ann. New York Acad. Sci.*, vol. 1514, no. 1, pp. 70–81, 2022.

[24] D. Krackhardt, "A plunge into networks," *Science*, vol. 326, no. 5949, pp. 47–48, 2009.

[25] N. E. Friedkin, "The problem of social control and coordination of complex systems in sociology: A look at the community cleavage problem," *IEEE Control Syst. Mag.*, vol. 35, no. 3, pp. 40–51, Jun. 2015.

[26] A. Gionis, E. Terzi, and P. Tsaparas, "Opinion maximization in social networks," in *Proc. SIAM Data Min. Conf.*, 2013, pp. 387–395.

[27] D. Bindel, J. Kleinberg, and S. Oren, "How bad is forming your own opinion?" *Games Econ. Behav.*, vol. 92, pp. 248–265, Jul. 2015.

[28] A. Matakos and A. Gionis, "Tell me something my friends do not know: Diversity maximization in social networks," in *Proc. ICDM*, 2018, pp. 327–336.

[29] C. Musco, C. Musco, and C. E. Tsourakakis, "Minimizing polarization and disagreement in social networks," in *Proc. WWW*, 2018, pp. 369–378.

[30] X. Chen, J. Lijffijt, and T. De Bie, "Quantifying and minimizing risk of conflict in social networks," in *Proc. 24th ACM KDD*, 2018, pp. 1197–1205.

[31] A. V. Proskurnikov, C. Ravazzi, and F. Dabbene, "Dynamics and structure of social networks from a systems and control viewpoint: A survey of Roberto Tempo's contributions," *Online Soc. Netw. Media*, vol. 7, pp. 45–59, Sep. 2018.

[32] S. E. Parsegov, A. V. Proskurnikov, R. Tempo, and N. E. Friedkin, "Novel multidimensional models of opinion dynamics in social networks," *IEEE Trans. Autom. Control*, vol. 62, no. 5, pp. 2270–2285, May 2017.

[33] A. V. Proskurnikov and R. Tempo, "A tutorial on modeling and analysis of dynamic social networks. Part II," *Annu. Rev. Control*, vol. 45, pp. 166–190, Jun. 2018.

[34] A. Matakos, E. Terzi, and P. Tsaparas, "Measuring and moderating opinion polarization in social networks," *Data Min. Knowl. Discov.*, vol. 31, no. 5, pp. 1480–1505, Sep. 2017.

[35] P. Dandekar, A. Goel, and D. T. Lee, "Biased assimilation, homophily, and the dynamics of polarization," *Proc. Nat. Acad. Sci. U S A*, vol. 110, no. 15, pp. 5791–5796, 2013.

[36] R. Abebe, J. Kleinberg, D. Parkes, and C. E. Tsourakakis, "Opinion dynamics with varying susceptibility to persuasion," in *Proc. 24th ACM KDD*, 2018, pp. 1089–1098.

[37] W. W. Zachary, "An information flow model for conflict and fission in small groups," *J. Anthropol. Res.*, vol. 33, no. 4, pp. 452–473, 1977.

[38] U. Chitra and C. Musco, "Analyzing the impact of filter bubbles on social network polarization," in *Proc. WSDM*, 2020, pp. 115–123.

[39] J. Leskovec and J. J. McAuley, "Learning to discover social circles in ego networks," in *Proc. Int. Conf. Adv. Neural Inf. Process. Syst.*, 2012, pp. 539–547.

[40] H. Zhang, Y. Dong, J. Xiao, F. Chiclana, and E. Herrera-Viedma, "Consensus and opinion evolution-based failure mode and effect analysis approach for reliability management in social network and uncertainty contexts," *Rel. Eng. Syst. Saf.*, vol. 208, Apr. 2021, Art. no. 107425.

**Elisabetta Biondi** received the Ph.D. degree in information engineering from the University of Pisa, Pisa, Italy, in 2017.

She is a Researcher with the Institute for Informatics and Telematics, National Research Council of Italy (CNR), Pisa. She is currently working in the field of opinion diffusion. She is a member of the CNR research unit for the H2020 SoBigData++ and HumanE-AI-Net projects and was involved in other H2020 and FP7 projects too. The general focus of her research activity is the mathematical modeling of human behaviors in social networks, with a special focus on mobility and information diffusion.

**Chiara Boldrini** received the Ph.D. degree in information engineering from the University of Pisa, Pisa, Italy, in 2010.

She is a Senior Researcher with IIT-CNR, Pisa. She is the IIT-CNR Co-PI for H2020 SoBigData++ and HumanE-AI-Net projects, and was involved in several EC projects since FP7. Her research interests are in decentralized AI, computational social sciences, and mobile and ubiquitous systems. She has published more than 50 papers on these topics.

Dr. Boldrini is in the Editorial Board of *Pervasive and Mobile Computing* (Elsevier) and *Computer Communications* (Elsevier). She was the Lead Guest Editor for the PMC Special Issue on IoT for Fighting COVID-19. She has served as the TPC Vice-Chair of IEEE PerCom'21 and over the years, has been in the organizing committee of several IEEE and ACM conferences/workshops, including IEEE PerCom and ACM MobiHoc.

**Andrea Passarella** received the Ph.D. degree in information engineering from the University of Pisa, Pisa, Italy, in 2005.

He is the Research Director with the Institute for Informatics and Telematics (IIT) of CNR, Pisa. Prior to joining IIT, he was with the Computer Laboratory of the University of Cambridge, Cambridge, U.K. He has coauthored the book *Online Social Networks: Human Cognitive Constraints in Facebook and Twitter Personal Graphs* (Elsevier, 2015). He has published more than 200 papers on human-centric mobile networks, online and mobile social networks, opportunistic, ad hoc, and sensor networks.

Dr. Passarella received four best paper awards, including the IFIP Networking 2011 and IEEE WoWMoM 2013. He was the General Co-Chair of IEEE PerCom 2022 and WoWMoM 2019, and the Workshops Co-Chair for IEEE INFOCOM 2019. He was the PC Co-Chair of IEEE WoWMoM 2011, the Workshops Co-Chair of several IEEE and ACM conferences. He is the founding Associate EiC of the *Online Social Networks and Media* (Elsevier). He is the Chair of the IFIP WG 6.3 "Performance of Communication Systems."

**Marco Conti** received the degree in information science with the University of Pisa, Pisa, Italy, in 1987.

He is a CNR Research Director and he is currently the Director of IIT-CNR Institute, Pisa. He has published more than 400 scientific articles related to design, modeling, and experimentation of computer and communication networks, pervasive systems, and online social networks.

Dr. Conti has received several awards, including the Best Paper Award at IFIP Networking 2011, IEEE ISCC 2012, and IEEE WoWMoM 2013. He was included in the "2017 Highly Cited Researchers" list compiled by Web of Science for the most cited articles in Computer Science. He served as the General/Program Chair of several major conferences, including IFIP Networking 2002, IEEE WoWMoM 2005 and 2006, IEEE PerCom 2006 and 2010, ACM MobiHoc 2006, IEEE MASS 2007, and IEEE SmartComp 2021. He is the founding EiC of *Online Social Networks and Media*, the EiC for Special Issues of *Pervasive and Mobile Computing*, and from 2009 to 2018, and the EiC of *Computer Communications*.