Journal Pre-proof

Homophily and influence

Arnold Polanski and Fernando Vega-Redondo

PII: S0022-0531(22)00166-1
DOI: https://doi.org/10.1016/j.jet.2022.105576
Reference: YJETH 105576

To appear in: *Journal of Economic Theory*

Received date: 26 February 2019
Revised date: 7 November 2022
Accepted date: 12 November 2022

Please cite this article as: A. Polanski and F. Vega-Redondo, Homophily and influence, *Journal of Economic Theory*, 105576, doi: https://doi.org/10.1016/j.jet.2022.105576.

Homophily and Influence

Arnold Polanski
University of East Anglia

Fernando Vega-Redondo
Bocconi University & IGIER

October 28, 2022

Abstract

We study how learning and influence co-evolve in a social network by extending the classical model of DeGroot (1974) in two fundamental ways:
(a) opinions are multidimensional and the learning time-span is arbitrary;
(b) the effective social network is endogenously shaped by opinion-based homophily.

Our analysis starts by establishing the existence of an equilibrium where, following (a)-(b), the learning outcome and the social network are jointly determined. This is followed by its characterization in some simple contexts. Next, we show that, at equilibrium, the strength of the link between any two agents is always given by its "support" – roughly, the amount of third-party (indirect) influence impinging on both agents. This result leads to the key insight that distinct groups may fail to integrate if their (possibly many) cross-group links lack sufficient support. Building on this, we identify sets of conditions for which social fragmentation is robust (i.e. dynamically stable) or even the unique equilibrium.

Keywords: social learning; homophily; influence; echo chambers; integration.
JEL classif. codes: D83, D85.

1 Introduction

In this paper, we develop the idea that the network channeling social influence is shaped in conjunction with the learning process that unfolds on it – that is, we propose a model where the two dimensions, influence and learning, co-evolve. This perspective allows us to study the important issue of how social learning (e.g. the formation of individual opinions in a social context) is affected by the relaxation of the common, but unrealistic, assumption that the pattern of social influence stays fixed. Indeed, in the real world, "influence weights" are typically affected by the ongoing learning process and this may have a substantial and lasting effect on the final outcome. In particular, it may exacerbate the tendency of an initially cohesive society to become segregated into separate groups that hardly interact. This, in fact, has been highlighted as one of the distinct, and also worrying, features of modern societies because of the polarization and "echo chambers" it generates.[1]

At the end of this Introduction, we discuss some related papers that analyze the coevolution of networks and opinions. In different contexts, they show that polarization can be a stable outcome due to the endogenously determined network structure. In contrast to these papers, our model stresses how a suitable (topological) notion of "network embeddedness" – which we call *support* – shapes the process through which opinions are formed, possibly ending up in a situation where they become solidly polarized. In this case, the society becomes fragmented into *de facto* opinion-independent subpopulations, across which no information or influence effectively flows.

Within our model, such a state of affairs may arise even if those groups (subpopulations) of agents are heavily connected across (i.e. they have many links between them). Indeed, we show that polarization always happens to be an entrenched (i.e. dynamically robust) possible outcome whenever those cross-group links do *not* enjoy the aforementioned support. For, because of homophily, it turns out that such a support is crucially needed if any amount of influence is to flow across inter-group links, thus breaking polarization. A better understanding this issue, we shall argue, should help identifying some of the key features and trade-offs underlying this important problem, hence suggesting possible avenues to tackle it.

As indicated, our learning model builds upon the well-known setup proposed by DeGroot, which has been studied in the economic literature by, among others, DeMarzo, Vayanos and Zwiebel (2003) and Golub and Jackson (2010, 2012). In the DeGroot framework, the opinion (or belief)[2] held by an agent adjusts over time by combining linearly her immediately preceding beliefs and those of others. The vectors of weights specifying how each agent is affected by every other agent in the population define an influence network/matrix, which fully governs the overall social-learning process. Then, under the twin assumptions that the number of learning rounds is *unbounded* and the influence network is *connected*, a standard result in this literature is that, under mild regularity conditions, the population converges to consensus – i.e. all agents end up holding the same opinion.

The situation, however, is interestingly different if, as we posit here, learning involves a *finite* number of learning rounds and the influence network is *endogenous*. For, on the one

[1] A good illustration of the problem is documented by Adamic and Glance (2005), who study the deep divide between conservative and liberal blogs in the period preceding the U.S. Presidential Election of 2004. With a different focus, Bishop (2008) discusses the geographical basis for this phenomenon and its negative effect on social cohesion. For a more general approach to the problem, we refer to Boutyline and Willer (2017), who show that the interaction bias is more prevalent the more extreme are the political views held by individuals. Finally, another interesting illustration is provided by Golub and Jackson (2012), who describe how political prior alignment segmented the information (and hence ended up segmenting the opinion as well) of the American public on the question of weapons of mass destruction in Iraq.

[2] In this paper, we apply the term "belief" or "opinion" interchangeably. In general, our model could apply to any continuous behavioral trait with a compact range.

hand, given the pattern of interpersonal influence that eventually materializes, its evolving topology must obviously play a key role in modulating how much divergence of opinions may persist after social interaction. And on the other hand, if significant opinion heterogeneity still remains, the force of homophily[3]) may shape a social network that exacerbates divergence.

The importance of homophily in determining how individuals construct their social network has been widely documented in the sociological literature. For a good early account of its pervasiveness in social environments, the reader is referred to the seminal work by Lazarsfeld and Merton (1954), whereas a more recent survey can be found in McPearson, Smith-Lovin and Cook (2001).[4]

It is important to stress that, in this paper, homophily is to be regarded not as a normative postulate but as a *descriptive* one. For example, from the point of view of social learning, it may well yield a suboptimal outcome if information gathering is the main objective. For, in this case, when choosing with whom to connect, agents should target those who are different from them and hence likely to hold complementary information – see Golub and Jackson (2012) or Lobel and Sadler (2013) for an illustration of this point in different contexts (DeGroot learning in the first case, Bayesian in the second).[5]

In a nutshell, our homophily-based approach to endogenizing the network of social influence involves the postulate that the weight of each link should match the similarity of opinions/beliefs of the agents connected by it. Another important property that will also be assumed on the underlying environment is that it is rich enough to comprise opinions in multiple dimensions. Thus, for example, individuals may hold opinions on a number of different topics: economic, political, religious, etc. Or, even if restricted to just one such category, say the economic one, their concern may cover a wide range of different issues such as growth, unemployment, inflation, or income distribution. Such richness of the "topic space" is important in our context because it introduces the possibility of defining non-

[3]Homophily is sometimes defined in terms of some immutable characteristics. In this paper, it is based on mutable (endogenous) opinions/beliefs.

[4]Further interesting references providing a general perspective on the phenomenon are Cohen (1977), Kandel (1978) Marsden (1987, 1988), Alexander et al. (2001), McVicar and Polanski (2014), Moody (2001), and Knecht et al. (2010). Concerning the specific problem that concerns us here, namely, how homophily impinges on opinion and belief formation, Golman, Loewenstein, Moene, and Zarri (2016) carry out an insightful discussion of modern literature, covering a rich range of empirical evidence on what they call the innate human preference for "belief consonance." As they explain, in many cases this urge for consonance leads to belief clustering, i.e. the choice "to associate with – that is, become friends with, work with, and even have romantic relationships with – others who share their beliefs" (see op. cit., p. 177).

[5]One must bear in mind, however, that belief-based homophily will typically be only one of the forces at work and, therefore, may not always be the main driving force in social learning. For example, contrary to what is often presumed, Gentzkow and Shapiro (2011) and Boxell *et al.* (2017) have shown that Internet access need not reinforce ideological segregation. Specifically, they provide empirical evidence where it turns out to mitigate ideological polarization, at least as compared to alternative situations where face-to-face interaction is the primary source of news or inter-personal communication.

trivial correlations among the different opinion dimensions.[6] And, as we shall explain, the similarity measure that underlies our notion of homophily is assessed in terms of whether agents display correlated deviations from "benchmark opinions" in the different dimensions.

Based on such a notion of similarity, our model posits that, given a certain (exogenously given) network of inter-agent observation, social influence is adjusted in the following simple manner:

- *every agent adjust the influence she attributes to the agents she observes in proportion to the degree of (bilateral) similarity displayed with each of them.*

This is, in essence, the law of motion that defines the stylized dynamics that governs the evolution of inter-agent influence. The dynamical system is studied from different perspectives. First, we focus our analysis on its equilibria (i.e. stationary points). Each of these defines what we label an *Equilibrium Influence Matrix* (EIM). After establishing an existence result and characterizing the EIMs for a number of simple benchmark scenarios (e.g. when the number of learning rounds is unbounded or the observation network is complete), we turn to studying how, in general, their properties depend on the following two features of the environment:

(a) the ***learning span*** (how many adjustment rounds on agents' opinions are conducted);
(b) the **structure *of the observation network*** among the agents (for, naturally, an agent can be influenced only by those whom she observes).

A key result in this respect involves singling out a topological measure that characterizes the weighted pattern of influence that must prevail at ***any*** EIM. Specifically, it shows that, at equilibrium, the influence that any agent i exerts on another agent j must be proportional to the accumulated (indirect) influence that all third-party agents exert on *both* i and j. We call this magnitude the *support* of the relationship between i and j.

An important consequence of the aforementioned result is that fragmentation of the population into groups may be hard to overcome by the mere establishment of observation links that could "bridge" influence across them. To understand this issue dynamically, we focus on a canonical and especially transparent context where two originally disconnected groups come into contact, i.e. establish observational links between them. Then we ask: Will they be able to rely on such cross-group observation to build up enough cross-group influence as well, thus becoming a more integrated population? We find that, even if the new bridging links are numerous, this can fail to happen for a number of interestingly different reasons. More specifically, we characterize how factors such as asymmetries in the group sizes or the learning span may play a key role in reinforcing segmentation, making it a dynamically robust state of affairs.

[6]Even though the model proposed by DeGroot allows for multidimensionality of opinions – and so do, for example, DeMarzo, Vayanos and Zwiebel (2003) – their approach deals with each dimension independently.

We close this introduction with a brief discussion of related literature. The general idea that the co-evolution of links and behavior underlies the dynamics of many interesting processes in social environments is widely stressed in the network literature.[7] Concerning, more specifically, the phenomenon of social learning that is our focus here, early papers studying the network-influence interplay include Hegselmann and Krause (2002), Holme and Newman (2006), Cradall *et al.* (2008) and Pan (2010), while the paper by Flache *et al.* (2017) provides a useful discussion of the previous research in this area.

Within the more recent literature studying such an interplay, we illustrate its broad diversity by summarizing the papers by Melguizo (2019), Bolleta and Pin (2020), and Cerreia-Vioglio *et al.* (2021). All of of them are concerned with the problem of polarization, but each one explores it from a somewhat different angle. Meguizo proposes a model where inter-agent homophily is defined in terms of a suitable similarity measure defined on multi-dimensional space of attributes. Her main conclusion is that disagreement persists in the long run if, and only if, homophily in at least one attribute grows sufficiently fast. Instead, Bolleta and Pin study a strategic model of polarization where, along an adjustment process, heterogeneous agents choose their actions and links to reach a suitable "compromise" between abiding by their own type and minimizing "dissonance" with neighbors. If the initial distribution of types is sufficiently spread out, there is a convergence toward a polarized social structure where the population is partitioned into separate network components. Finally, Cerreia-Vioglio *et al.* study a general (but non-strategic) learning framework where, in contrast with the DeGroot model, agents use opinion aggregators that possibly display non-linear features. None of these papers, however, study the endogenous interplay of network embeddedness and opinion evolution that is our primary concern here.

To end our literature review, it is worth mentioning that polarization in social networks may arise even if the influence network remains fixed and therefore no influence-network interplay can be the driving force at work. This is well exemplified by papers of Levy and Razin (2018, 2020), which highlight the role played by the phenomenon often labeled *correlation neglect*. More specifically, they show that learning *per se* on a given network may also lead to the polarization of opinions/beliefs when agents neglect the correlation that results from being exposed, directly or indirectly, to non-disjoint sets of peers.[8]

The rest of the paper is divided into three more sections. Section 2 presents the basic framework, Section 3 carries out the analysis, Section 4 concludes with a summary. For the sake of a smooth presentation, all formal proofs are gathered in the Appendix.

[7] For a recent Handbook survey that discusses this issue for a broad range of socio-economic phenomena (e.g., bargaining, intermediation, conflict, public goods, or learning) see Vega-Redondo (2016).

[8] For an elegant axiomatic characterization of such behavior in social networks, see Molavi *et al.* (2018).

2 The model

2.1 The pattern of communication and influence

Consider a given population of agents, $N = \{1, 2, ..., n\}$, who are connected by an *exogenous observation network* on the set N. Formally, this directed network can be represented through an adjacency matrix $L \equiv (l_{ij})_{i,j\in N}$ with each $l_{ij} \in \{0, 1\}$ for all i and j. If $l_{ij} = 1$, this means that agent i observes (and hence can be influenced by) j. Naturally, we assume that every agent i can "observe"herself, so $l_{ii} = 1$ for every $i \in N$. The observation network is assumed exogenous, reflecting those features of the situation (say, geographical, ethnic, or linguistic) that determine whether an agent can directly influence another agent.

On the other hand, the pattern of *effective* inter-agent influence can also be formalized as a network – we shall call it the *influence network* – and represented through a corresponding matrix $A \equiv (a_{ij})_{i,j\in N}$. However, in contrast to the observation network (which is binary and exogenous), the influence network is *weighted* and *endogenous*. We elaborate on each of these two important characteristics in turn.

(i) A typical entry a_{ij} of the matrix A is interpreted as the (relative) intensity with which agent j influences i – i.e. the weight that i attributes to j's opinions in shaping her own. For convenience, every such measure of bilateral influence is taken to lie in the interval $[0, 1]$ and every row of A adds up to unity. Thus, A is a *row-stochastic matrix*.

(ii) The pattern of inter-agent influence is endogenized through the homophily-based dynamics that describes how agents adjust the influence they ascribe to each agent they observe. As formally defined below (Subsection 2.4), the stationary points of this dynamics display the feature that the weight an agent i attributes to any other j is proportional to their belief similarity, suitably quantified (c.f. Subsection 2.3).

To make the above considerations precise, we introduce three core components of the model: (a) the behavioral learning dynamics; (b) the notion of homophily; (c) the required consistency between the former two. To do so is the objective of the following three subsections.

2.2 Learning

The starting point of our model is the learning framework proposed by DeGroot (1974), which can be summarized as follows. Time proceeds discretely, $t = 0, 1, 2, ...$, and at every t each agent $i \in N$ holds some opinion $x_i(t) = (x_{i1}(t), x_{i2}(t), ..., x_{im}(t))$, identified with a point in a common and pre-specified compact and convex set, say an m-dimensional cube $[0, 1]^m$. Thus, given some prior beliefs (initial opinions), $x_i(0) \in [0, 1]^m$ for every $i \in N$, the subsequent opinions of individuals over each dimension evolve over time as follows:

$$x_{iq}(t) = \sum_{j=1}^{n} a_{ij} x_j(t-1) \qquad (i = 1, 2, ..., n;\ q = 1, 2, ..., m;\ t = 1, 2, ...),$$

or, in compact matrix form,

$$\boldsymbol{x}(t) = A\boldsymbol{x}(t-1) \qquad (t = 1, 2, ...), \tag{1}$$

where A is a square row-stochastic influence matrix (for the moment, fixed) of dimension n, and $\boldsymbol{x}(t) = (x_1(t), ..., x_n(t))$ is an $n \times m$-matrix consisting the m column vectors of dimension n listing the opinions of all agents at t for each of the m dimensions. Naturally, the learning process defined by (1) requires the specification of some initial conditions. These are given by a matrix of individual prior beliefs $\boldsymbol{x}(0) = \boldsymbol{\beta} = (\beta_{iq})_{i,q=1}^{n,m}$ specifying the initial opinion on every dimension from which each individual in the population starts the learning process.

Another important parameter of the learning process in our context is what we call the *learning span*. This refers to the number K of iterations for which learning proceeds before the influence matrix A is revised. In general, we want to think of K as finite, which reflects the idea that learning and influence adjustment proceed at comparable time scales. This contrasts with the usual approach pursued in the study of the DeGroot model, which typically focuses on the asymptotic outcome obtained when K grows unboundedly. The reason for a finite span is that, in general, we want to have the learning outcome depend on the topology of the network. For example, it is natural to expect that the opinion of an agent at the end of a learning spell – just at the point where her influence weights are adjusted – should depend on the relative network proximity to other agents and their different opinions. This, however, would not happen if learning spells are not bounded. For then, as explained in the Introduction, the population would typically reach a consensus if the influence network is connected.

2.3 Similarity

To measure similarity across different individuals' opinions, we shall rely on a versatile statistic that can accommodate interesting special cases. For a matrix of opinions $x = (x_{iq})_{i,q=1}^{n,m}$ and any pair of agents i and j, we define the similarity between them as follows:

$$\phi_{ij}^{x} = \textstyle\sum_{q=1}^{m} \omega_q (x_{iq} - g_{iq}(x))(x_{jq} - g_{jq}(x)), \tag{2}$$

where each ω_q $(q = 1, 2, ..., m)$ weighs the relative relevance (or normalizes the measurement units) of issue q, while $g_{iq}(x)$ acts as a benchmark of comparison for the opinion of i on that issue. The value of this benchmark can be any convex combination of the form

$$g_{iq}(x) = \lambda b_q + (1-\lambda) \textstyle\sum_{s=1}^{m} w_s x_{is}, \tag{3}$$

where b_q is an issue-specific constant opinion, $\lambda \in [0,1]$ and $\{w_s\}_{s=1}^m$ is a profile of weights on i's opinion about each issue $s = 1, 2, ..., m$.

At one extreme, when $g_{iq}(x) = b_q$, every agent evaluates her position on each issue q with respect to a common opinion b_q. This opinion can be viewed as representing an issue-specific canon in the society that is unchangeable under the learning dynamics. As a natural example, b_q could be the population-average of the initial opinions:[9]

$$b_q = \frac{1}{n}\sum_{i=1}^n \beta_{iq}. \tag{4}$$

At the other extreme, the benchmark is individual-specific in that it is given by a weighted sum of i's opinions on all issues and therefore depends only on the agent's opinion profile. In the special case where it coincides with i's average opinion:

$$g_{iq}(x) = \frac{1}{m}\sum_{q=1}^m x_{iq} \tag{5}$$

our similarity measure (2) boils down to a centered version of *cosine similarity* analogous to the standard Pearson correlation coefficient.[10]

Denote by $\Phi_x = (\phi_{ij}^x)_{i,j=1}^n$ the matrix of similarities induced by any given array x of agents' opinions. For simplicity, we shall refer to it as the covariation matrix and say that player i's and j's opinions are positively (negatively) correlated when $\phi_{ij}^x > 0$ ($\phi_{ij}^x < 0$), while saying that they are uncorrelated (or orthogonal) when $\phi_{ij}^x = 0$. When it is clear from the context, we shall also omit the index x and write simply ϕ_{ij} and Φ. As we shall see, this matrix plays a key role in our analysis. Even though its precise form naturally depends on the concrete value of the parameters in (2) that define the similarity measure under consideration, it is worth emphasizing that our results do not depend on them. That is, they hold for any similarity measure consistent with (2).

Remark 1 (Two routes to comparing agent opinions) *Our model relies on two different ways of comparing the opinions of pairs of agents, i and j. One of them is the traditional one in the DeGroot model. That is, for each time t within a given learning spell, we* **separately** *compare, on each opinion dimension $q = 1, 2, ..., m$, the* **distances** *of their opinions, $x_{iq}(t)$ and $x_{jq}(t)$. In contrast, the other comparison considered in the model involves a* **joint** *measurement of the similarity of the* **vectors** *of opinions of the agents, $(x_{ir}(t))_{r=1}^m$ and $(x_{jr}(t))_{r=1}^m$.*

[9] Any other population statistic of the population, such as the median, may also be used to capture a similar idea. Note that, in every case, agents are supposed to maintain the benchmark fixed throughout, as they move into a new round of learning.

[10] Both measures, cosine similarity and Pearson correlation, are measures of behavioral similarity prominently used by computer scientists in the design of *memory-based* recommender systems. These systems rely on what are called collaborative (or preference) filtering methods that identify those peers whose past behavior is similar to that of any given user, and hence can be used to predict the user's reaction to various behavioral stimuli. For a good overview of the early developments in this lively field of research in computer science, the reader may refer to the surveys by Adomavicius and Tuzhilin (2005) and Su and Khoshgoftaar (2009), or the monographs by Jannach *et al.* (2010) and Ricci *et al.* (2011).

Despite this distinction, however, the two measures – opinion agreement and agent similarity – are largely aligned at an equilibrium of the full-fledged dynamics. For, if two individuals hold close opinions in the various dimensions, they must also be similar according to our definition of agent similarity. And, conversely, if they are similar, they have to be exerting a significant influence on each other, which in turn implies that their opinions cannot be very different at equilibrium. This reciprocal effects between agent and opinion similarity will be at the basis of our equilibrium notion. By themselves, however, they do not characterize equilibria. For, in general, what equilibria may exist, whether they display polarization or not, and which of them are dynamically robust, depends on important characteristics of the environment. For example, it depends on the observation network, the distribution of prior beliefs, or the length of the learning spells in a non-trivial manner. Indeed, to gain a precise understanding of these matters is a primary objective of the analysis undertaken in Section 3.

2.4 Homophily and the evolution of social influence

As explained, our key modeling assumption is that the adjustment of inter-agent influence over time is shaped by opinion-based homophily. Or, to express it more concretely, we assume that agents reinforce the links with those who hold similar opinions. In this respect, an important point already made is worth repeating here: homophily is not conceived in this paper as a normative postulate but as a *positive* one – that is, as a manifestation of what, in essence, is a strong bias in human nature.[11]

Formally, the influence dynamics is taken to operate across (finite) learning spells indexed by $s \in \{1, 2, ...\}$. At every such s, there is a corresponding influence matrix $A(s)$ that governs social learning during that spell (cf. Subsection 2.2). For the moment we simplify matters (see Remark 3 below for various extensions) by supposing that agents start every learning spell by holding the same beliefs $\boldsymbol{\beta} = (\beta_{iq})_{i,q=1}^{n,m}$. Such a persistence of initial opinions may be understood as a reflection of some relatively *stable* and *generic* biases displayed by each individual agent when confronting a range of *concrete* issues. Thus, even though new such issues arise in each learning spell for every dimension $q = 1, ..., m$, individuals rely on a quite persistent set of overarching values to shape their initial positions on each of those dimensions.[12] This seems a reasonable modeling choice if those values change only slowly relative to the speed at which social influence adjusts.

[11] A possible explanation for this bias is that, even though in the modern world this bias may well have dysfunctional consequences (e.g. when strong complementarities are required), it may have had positive (fitness) implications early on in human history.

[12] For example, an individual may be a fiscal conservative and therefore will tend to have an *ex ante* bias against any particular proposal that entails an increase in government spending. Or she may be quite sensitive to environmental protection, or education, or gender balance and, therefore, tend to evaluate positively any specific proposal in those dimensions (e.g., limiting traffic in urban areas, a pay increase for teachers, or some new law that favors the hiring of women).

Starting from those initial conditions, and after updating their opinions K times during each learning spell s, the agents eventually reach the vector of opinion $\hat{\boldsymbol{x}}(s) = A(s)^K \beta$. Then, in adjusting their influence weights to arrive at the influence matrix $A(s+1)$ that will prevail in the next spell, agents rely on the matrix $\Phi_{\hat{\boldsymbol{x}}(s)}$ of bilateral covariations among the final opinions at s, $\hat{\boldsymbol{x}}(s)$. From Lemma 1 in the Appendix we know that this matrix can be directly computed as $\Phi_{\hat{x}(s)} = A(s)^K \, \Phi_\beta \, (A(s)^K)'$. This then allows to measure the bilateral similarity of the opinions held by agents i and j at the end of s through the following standardized measure of covariation:

$$\hat{\rho}_{ij}(A(s)) \equiv \frac{\max\{\phi_{ij}^{\hat{\boldsymbol{x}}(s)}, 0\}}{\sqrt{\phi_{ii}^{\hat{\boldsymbol{x}}(s)} \phi_{jj}^{\hat{\boldsymbol{x}}(s)}}}, \tag{6}$$

It is in terms of such similarity measures that we define the law of motion for inter-agent influences. Specifically, we simply posit that, at stage $s+1$, the influence on any given agent i exerted by every other other agent j (including i herself) is proportional to the corresponding standardized similarity between the two agents, $\hat{\rho}_{ij}(A(s))$. Thus, when we take into account that:

(a) the entries of $A(s+1)$ must be normalized so that this matrix is row-stochastic

(b) only those links in L that are part of the observation network can carry influence

we are led to the following expression for the influence weight on i exerted by j at $s+1$:

$$a_{ij}(s+1) = \frac{l_{ij}\, \hat{\rho}_{ij}(A(s))}{\sum_{k \in N} l_{ik}\, \hat{\rho}_{ik}(A(s))} \qquad (i, j = 1, 2, ..., n), \tag{7}$$

In compact form, it will be convenient to formalize the induced law of motion by the corresponding vector field $F : (\Delta^{n-1})^n \to (\Delta^{n-1})^n$ where, for any $A(s)$ and every $i, j = 1, 2, ..., n$, $F_{ij}(A(s))$ is given by the RHS of (7). Note that, in general, $F(\cdot)$ depends not only on the observation network L but also on the (covariation matrix of) prior beliefs $\boldsymbol{\beta}$ and the span K characterizing the learning spells.

Much of our analysis in this paper will focus on the stationary (or equilibrium) points induced by the postulated adjustment process, as described in the following definition.

Definition 1 *Given observation matrix $L \equiv (l_{ij})_{i,j \in N}$, the learning span K, and the set of initial opinions $\boldsymbol{x}(0) = \boldsymbol{\beta}$, a row-stochastic matrix $A^* \equiv \left(a_{ij}^*\right)_{i,j \in N}$ that satisfies:*

$$F(A^*) = A^* \tag{8}$$

is said to be homophily consistent *– or, equivalently, an* equilibrium influence matrix *(EIM).*

We end this section with an example and two remarks that bear on interesting aspects of the model. The example is intended to provide a stylized illustration that may help understand the nature of the opinion-influence adjustment process postulated here. Then, the first remark makes a methodological point concerning the role of opinion multidimensionality in the analysis, while the second one introduces a generalization of the assumption that prior beliefs are persistent across learning spells.

Example 1 (A twitter-motivated process of influence adjustment) *We consider three agents connected through the observation network L illustrated in Figure 1 (thus all entries in L are equal to 1 except $l_{13} = l_{32} = 0$). The interpretation of a directed link from i to j in this network is that i follows (and, hence, is influenced by) j on Twitter.*

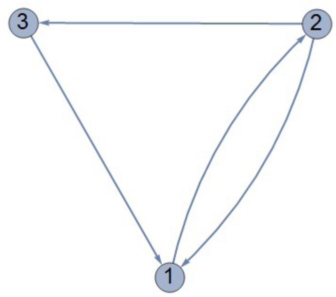

Figure 1: A directed link from i to j means that i follows (and, hence, is influenced by) j on Twitter.

Let $s = 1, 2, ...$ index the different consecutive weeks (learning spells). During the first part of each week s – which we conceive as the workweek – the agents write tweets expressing their opinions and also read the tweets of others, as restricted by the observation matrix L. After each round of tweets, they revise their opinions as posited by our model, in terms of the influence matrix $A(s)$ prevailing during that week s. Eventually, at the end of the workweek, this leads to an opinion vector $\hat{\boldsymbol{x}}(s) = A(s)^K \beta$ where K is the number of tweeting rounds during the week and β stands for the initial opinions that they held at the beginning of it.

Next, on Saturday, the agents reconsider the influence they attribute to (or the intensity with which they follow) each of the other two individuals. They do so by revising the influence weights among those whom they observe in proportion to the corresponding similarity of opinions displayed by the matrix $\Phi_{\hat{x}(s)}$. That is, the new influence matrix $A(s+1)$ that will be used next week is computed by (7), i.e., by setting each element $a_{ik}(s+1)$ equal to the row-normalized version of $l_{ik}\hat{\rho}_{ik}(\hat{x}(s))$.

Finally, on Sunday, they devote the day to gathering fresh information about issues relevant for the coming week by, e.g., reading several newspapers (not necessarily the same ones). These newspapers have a quite persistent position ("editorial lines") on the different general topics, so again they form the same opinions β as they held at the beginning of the previous week. These are, therefore, the starting beliefs of the interaction and opinion revision process unfolding in week $s+1$.

More concretely, suppose that on the very first week ($s = 1$) the agents assign their influence *weights as given by the following matrix:*

$$A(1) = \begin{pmatrix} 2/3 & 1/3 & 0 \\ 1/6 & 2/3 & 1/6 \\ 1/3 & 0 & 2/3 \end{pmatrix}.$$

Further suppose, for simplicity, that there is just one round of tweets every week ($K = 1$) *and that prior beliefs are uncorrelated for each pair of agents and all display the common individual variation* $\varpi > 0$*, i.e.* $\Phi_\beta = \varpi I$*, where* I *is the identity matrix. Then, the final beliefs at the end of the first week is* $\hat{\boldsymbol{x}}(1) = A(1)\beta$*. The corresponding similarity coefficients and the revised influence weights after this week can then be computed from the covariation matrix* $\Phi_{\hat{\boldsymbol{x}}(1)} = A(1)\Phi_\beta A(1)'$ *as follows:*

$$\hat{\rho}(\hat{x}(1)) \approx \begin{pmatrix} 1 & .63 & .4 \\ .63 & 1 & .32 \\ .4 & .32 & 1 \end{pmatrix}, \quad A(2) \approx \begin{pmatrix} .61 & .39 & 0 \\ .33 & .51 & .16 \\ .29 & 0 & .71 \end{pmatrix},$$

where $A(2)$ *is obtained by the adjustment rule (7) from* $A(1)$*. It can be readily verified that* $A(2)$ *is not an EIM, i.e. it does not satisfy the homophily consistency condition (8). However, if the process continues over several weeks, each of them starting with the influence weights inherited from the previous one, the influence matrix converges to the EIM,*

$$\lim_{s\to\infty} A(s) = A^* = \begin{pmatrix} 1/2 & 1/2 & 0 \\ 1/2 & 1/2 & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

Hence, in equilibrium, only players 1 *and* 2 *effectively communicate via Twitter.* $\Diamond$

Remark 2 (Opinion multidimensionality) *In our model, opinions are defined simultaneously on different issues (or dimensions). The question we address here is how much of such multidimensionality is necessary to render our approach meaningful. As in the previous example, let us suppose that agents' initial opinions are orthogonal and their individual variation is positive and identical, i.e.* $\Phi_\beta = \varpi I$*.*[13] *The common individual variation* $\varpi > 0$ *can be interpreted as a measure of how much, initially, the alternative issues can be viewed as genuinely different in agents' minds. If* $\varpi = 0$ *then* $x_{iq} = \bar{x}_i$ *for each agent* i *and issue* q *and all the opinion dimensions would be redundant. Instead, if* $\varpi > 0$*, one can easily verify that the similarity coefficient* $\hat{\rho}_{ij}(A^*)$ *in (8) does* not *depend on the magnitude of* $\varpi > 0$*. It follows, therefore, that the same EIM* A^* *satisfies (8) for any fixed value of* $\varpi > 0$*, and hence also in the limit* $\varpi \searrow 0$*. This allows one to view* A^* *as an EIM for* any *opinion multidimensionality no matter how small!*[14]

[13] Such orthogonality is not essential for the argument but it is assumed here just for the sake of formal simplicity.

[14] Methodologically, the approach is reminiscent of that pursued in other strands of literature (e.g. evolutionary game theory – cf. Kandori *et al.* (1993) or Young (1993)), where a similar approach (in this case, in terms of mutation rates) has been used as a powerful selection device.

Remark 3 (On the persistence of initial beliefs) *In our earlier presentation of the model, the assumption that the initial opinion/beliefs remain unchanged across learning spells was motivated on the grounds that a stable collection of general views/values underlie the particular positions agents take on concrete issues. This assumption, admittedly demanding, can be substantially generalized. Specifically, it can be readily checked that our analysis requires only the weaker condition that, at the beginning of each spell* s*, agents' beliefs* $\beta(s)$ *exhibit the same covariation matrix* $\Phi_{\beta(s)} = \Phi_{\beta}$*. By way of illustration, a stylized context where such condition would hold (despite agents changing their prior beliefs across spells) is as follows. Suppose that, at the beginning of each learning spell* s*, the* n*-dimensional vector of prior beliefs* $(\beta_{1q}(s), \beta_{2q}(s), ..., \beta_{nq}(s))$ *on any particular* $q \in \{1, 2, ..., m\}$ *is drawn at random according to a multivariate distribution defined over* $[0, 1]^n$ *with covariance matrix* Υ*. Then, if such random draws are stochastically independent across issues and the number* m *of them is very large (formally* $m \to \infty$*), the induced covariance matrix over agents' beliefs can be suitably approximated by* Υ*. This implies that, in the particular case where* $\lambda = 0$*, the covariation matrix* $\Phi_{\beta(s)}$ *can be viewed as roughly equal to* Υ*, even though the prior beliefs* $(\beta_{i1}(s), \beta_{i2}(s), ..., \beta_{im}(s))$ *of any given agent* i *will typically be quite different across learning spells* $s \in \{1, 2, ...\}$*.*

The previous illustrative case notwithstanding, it can still be reasonably argued that, for many interesting contexts, the assumption that $\Phi_{\beta(s)} = \Phi_{\beta}$ *for any spell* s *may be viewed as too stringent a description of the situation. This is why, in Subsection 3.5, we approach the problem in a different manner. Specifically, we extend the formulation of the model to allow for the possibility that prior beliefs in a certain spell* s *are partly updated on the basis of the outcome of social learning in previous spell* $s-1$*, i.e. the prior beliefs in a certain stage partly reflect the opinion-formation process conducted at earlier stages. There, we discuss how certain aspects of our basic analysis – in particular, those concerning the central issue of social integration – are affected by this extension. As we shall see, a natural insight that follows from this extension is that, when prior beliefs are gradually updated on the basis of previous learning, a faster rate of such a belief adjustment works towards social integration in a way analogous to how longer learning spells do.*

3 Analysis

In this section, we carry out the analysis of the model. First, Subsection 3.1 starts by addressing the basic issue of EIM existence and its characterization. Subsection 3.2 describes the unique connected EIMs in some benchmark cases. Subsection 3.3 discusses the relationship between homophily consistency (i.e. equilibrium) and a suitable notion of link support (or neighborhood overlap). Finally, Subsection 3.4 relies on the insights obtained in Subsection 3.3 to study the relationship between communication/observation bridges and integration in some illustrative cases.

3.1 Equilibrium existence

Our primary interest in this paper is to study the characteristics of endogenous patterns of influence as captured by the equilibrium notion in Definition 1. Thus a first basic question that must be addressed is the following: Does there always exist an equilibrium influence matrix that satisfies the required consistency between the learning outcome and the influence weights? A positive answer to this question and an EIM characterization is provided by the next result.

PROPOSITION 1 *Consider an arbitrary observation network with adjacency matrix $L \equiv (l_{ij})_{i,j \in N}$, any learning span $K \geq 1$, and any matrix β of agents' initial opinions with $\Phi_\beta \geq 0$. An influence matrix A^* that satisfies (8) always exists and it verifies,*

$$a^*_{ij} = a^*_{ii} l_{ij} \hat{\rho}_{ij}(A^*) \qquad (i, j = 1, 2, ..., n). \tag{9}$$

Proof. See Appendix.

The existence result in Proposition 1 follows from a standard fixed-point argument applied to the vector field $F(\cdot)$ defined by (7).

Figures 2 and 3 illustrate the EIMs for a randomly generated observation network with $n = 25$ nodes. They are obtained by iterating $F(\cdot)$ from a situation where every agent assigns the same influence weight to each of her neighbors in the observation network.

In Figure 2, we consider the case where the learning span is the shortest possible, $K = 1$. We observe that the resulting pattern of equilibrium influence displays a marked dichotomous structure, with small subsets of agents maintaining high-influence links among themselves while the links with all other agents carry much lower influence. This leads to end opinions in the population that are quite heterogeneous, with agents fragmented into opinion clusters that sustain only light cross-connections.

In contrast, Figure 3 displays an EIM that is spanned on the same observation network as in Figure 2 but with communication proceeding for $K = 100$ rounds. In this alternative case, the learning process eventually leads to a situation where agents hold essentially the same opinions. Such a perfect convergence among individual opinions induces in turn an influence matrix where every agent attributes the same influence weight to each of her neighbors (and herself). The induced EIM is therefore very different from the one resulting for $K = 1$ since it is determined by the exogenous observation structure alone. In the next subsection, we present some limit context where, as in Figure 3, the (unique) EIM approaches uniform influence weights.

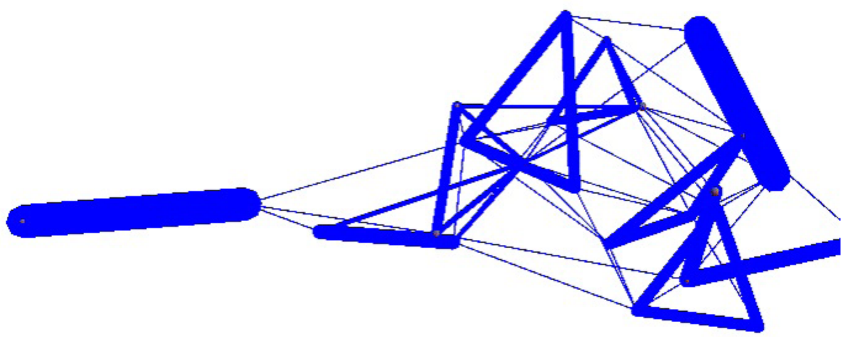

Figure 2: Graphical representation of an EIM $A^* \equiv (a^*_{ij})_{i,j \in N}$ for an underlying observation network with $n = 25$ nodes that is generated as a realization of an Erdös-Rényi random network. Initial opinions display identical individual variations and are orthogonal. The learning span is set to $K = 1$. The EIM is obtained through the iteration of (7) from initial conditions where every agent attributes a uniform weight to all her connections in the observation network. The thickness of each edge ij is proportional to $\frac{a^*_{ij}+a^*_{ji}}{2}$. Self links are not shown.

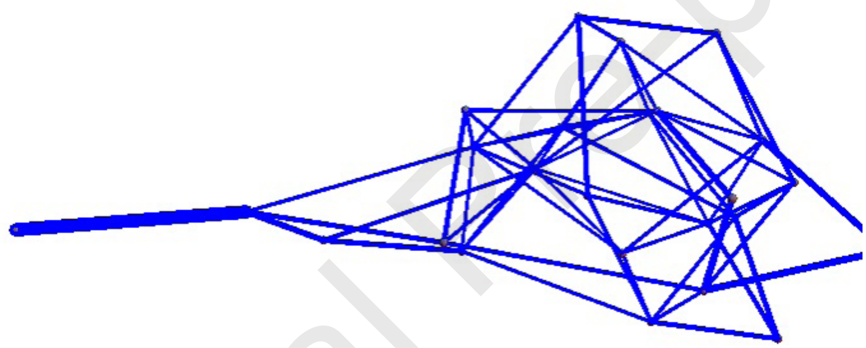

Figure 3: Graphical representation of an EIM $A^* \equiv (a^*_{ij})_{i,j \in N}$ under the same conditions as for Figure 1 except that the learning depth is $K = 100$.

3.2 Some benchmark scenarios

Here we discuss several benchmark scenarios that will prove useful in our subsequent analysis. First, we show that the outcome in Figure 3 can be extended to any (weakly) connected[15] observation matrix as long as the learning span K is unbounded.

Proposition 2 *Consider any (weakly) connected observation network with the observation matrix $L \equiv (l_{ij})_{i,j \in N}$ and any matrix β of agents' initial opinions with $\Phi_\beta \geq 0$. Then, as the learning span $K \to \infty$, the unique connected EIM converges to the matrix $A^* \equiv (a^*_{ij})_{i,j \in N}$*

[15]A directed graph is *weakly connected* if when considering it as an undirected graph it is connected, i.e., there is an undirected path between any pair of vertices, and *strongly connected* if there is a directed path between every pair of vertices. An analogous definition applies to weighted networks.

where

$$a^*_{ij} = \frac{l_{ij}}{\sum_{k \in N} l_{ik}} \qquad (i, j = 1, 2, ..., n),$$

and the induced bilateral similarity coefficients among end opinions satisfy $\hat{\rho}_{ij}(A^*) = 1$ *for all* $i, j \in N$*.*

Proof. See Appendix.

Thus, when the learning span is unbounded, the only connected influence matrix that meets the equilibrium requirement is the one that matches the observation network and assigns uniform weights across neighbors. This result makes the simple point that in order for equilibrium to be consistent with a connected influence pattern that displays a non-trivial structure, the learning process must fall short of delivering social consensus. If a consensus obtains, no feedback from the learning process will have any impact on the influence pattern. Then, the observation network fully imposes its structure on an EIM – i.e. the latter simply matches the former. These considerations suggest that, for our purposes, it is important to model social learning within a finite time frame and then study the learning span as an interesting parameter in the analysis.

Another useful benchmark case concerns situations where the population is segmented (endogenously) into separate influence components and, for each of these components, the underlying observation structure is complete. This is a stark setup but, as we shall see, it is a natural one to study the conditions under which fragmentation of an otherwise "observationally connected" population may persist. First, we establish the following result.

PROPOSITION 3 *Let the observation network be* completely *connected (i.e., the adjacency matrix* $L \equiv (l_{ij})_{i,j \in N}$ *satisfies* $l_{ij} = 1$ *for all* $i, j \in N$*). Then, for any given learning span* $K \geq 1$ *and any matrix* β *of agents' initial opinions with* $\Phi_\beta \geq 0$*, the unique weakly connected EIM* A^* *is* $Q(n) \equiv (q_{ij}(n))_{i,j \in N}$ *with* $q_{ij}(n) = \frac{1}{n}$ *for all* $i, j = 1, 2, .., n$*.*

Proof. See Appendix.

This result asserts that, if the observation network is complete, the only influence matrix that renders the population connected and meets the requirement of homophily consistency is the matrix $Q(n)$, where every agent influences *directly* each of the other agents with the same weight. On the one hand, the fact that this matrix is homophily-consistent is quite clear: under $Q(n)$, every agent is in effect exposed to the same convex combination of the initial opinions, which in turn supports the uniform influence matrix as an equilibrium. On the other, the fact that, even for a low K, there is *no other* connected EIM is less apparent. Intuitively, it follows from the cumulative effect in which indirect influence raises correlation and this, in turn, increases direct influence (and thus correlation) up to the point where only a fully symmetric pattern of influence can prevail at equilibrium.

Propositions 2 and 3 jointly imply that learning span and connection density can be seen as substitutes in reaching agreement: Under an equilibrium influence matrix, consensus can be reached either with sparse connections and many communication rounds (Proposition 2) or with the complete network and just one round of communication (Proposition 3).

Building upon Proposition 3, the following straightforward corollary follows.

Corollary 1 *Let the observation network L be* completely *connected and consider an r-element partition of the population given by $\{N_s\}_{s=1}^r$, with $n_s \equiv |N_s|$ for each $s = 1, 2, ..., r$, and $n_1 + ... + n_r = n$. Let $Q(n_r) \equiv (q_{ij}(n_r))_{i,j=1}^{n_r}$ be the square matrix of dimension n_r with $q_{ij}(n_r) = \frac{1}{n_r}$ for all $i, j = 1, 2, .., n_r$, and denote by $0_{n_r \times n_s}$ the matrix of dimension $n_r \times n_s$ consisting of all zeroes. Assume that $\Phi_\beta \equiv \left(\phi_{ij}^\beta\right)_{i,j \in N}$ satisfies $\phi_{k,\ell} = 0$ for $k \in N_s$ and $\ell \in N_{s'}$ with $s \neq s'$. Then, for all $K \geq 1$ and upon a suitable index labeling, the block diagonal matrix*

$$A = \begin{pmatrix} Q(n_1) & 0_{n_1 \times n_2} & \cdots & 0_{n_1 \times n_r} \\ 0_{n_2 \times n_1} & Q(n_2) & \cdots & 0_{n_2 \times n_r} \\ \vdots & \vdots & \ddots & \vdots \\ 0_{n_r \times n_1} & 0_{n_r \times n_2} & \cdots & Q(n_r) \end{pmatrix}$$

is the unique EIM where each set N_s is weakly connected and no agent in N_s influences an agent from a different set.

This corollary says that under the complete observation network and with uncorrelated initial opinions across groups of agents, there is a unique EIM that weakly connects all agents in the same group but separates them from agents in other groups. In this EIM, each agent has the same influence within their group but no influence on agents in other groups.

The relevance of Corollary 1 derives from the fact that, in studying the problem of segmentation among observationally connected agents, it allows us to restrict, without loss of generality, to subsets of agents who are internally connected through a uniform (and thus complete) pattern of influence. This will prove particularly useful when we study the problem of segmentation in Subsection 3.4. The setup in Corollary 1 allows for many other equilibria (where agents from different groups influence each other). This emphasizes an acute multiplicity problem in general settings. This is why, when we revisit the issue later on, our analysis will focus on scenarios where such a segmentation of the population into independent groups is either the only equilibrium or at least a dynamically robust one.

3.3 Influence and support

In this subsection, we identify a topological measure of the influence network that characterizes the equilibrium strengths of interpersonal influences at equilibrium. This characterization will then prove quite useful in our ensuing analysis. We start by defining our measure of *link support* for the link ik in a weighted network with the non-negative matrix of link weights $W \equiv (w_{ij})_{i,j \in N} \in R_+^{n \times n}$:

$$\varphi_{ik}(W) \equiv \frac{\sum_{s=1}^{n} w_{is} w_{ks}}{(\sum_{s=1}^{n} w_{is}^2)^{\frac{1}{2}} (\sum_{s=1}^{n} w_{ks}^2)^{\frac{1}{2}}}. \tag{10}$$

As $\varphi_{ik}(W)$ is simply the cosine of the angle between the rows i and k in W, it lies in the unit interval and it is symmetric ($\varphi_{ik}(W) = \varphi_{ki}(W)$) whenever these rows have at least one strictly positive entry each. In our context, the latter condition always holds as we assume that each node is connected to itself. We will apply $\varphi_{ik}(.)$ to the influence matrix A and also to the matrix $A^K \equiv B \equiv (b_{ij})_{ij \in N}$ that captures the indirect influence b_{ij} that an agent j has on agent i after K rounds of learning.

An instructive example of (10) derives from its application to the binary adjacency matrix $L \equiv (l_{ij})_{i,j \in N} \in \{0,1\}^{n \times n}$ that represents our observation network. It follows then directly from (10) that $\varphi_{ik}(L)$ is equal to the number of nodes to which both i and k are connected, divided by the geometric mean of i's and k's out-degrees. Or, in other words, it is a normalized measure of i's and j's shared neighborhood in L. We note that $\varphi_{ik}(L)$ is closely related to the notion of *neighborhood overlap* defined in Easley and Kleinberg (2010) as the ratio,

$$\frac{\text{number of nodes who are neighbors of both } i \text{ and } k}{\text{number of nodes who are neighbors of at least one of } i \text{ or } k}$$

where in the denominator they do not count i or k themselves (even though i and k are taken to be neighbors). This notion is also known as *link clustering coefficient* in the network literature (e.g., Pajevic and Plenz, 2012). It is, however, quite different from the usual notion of clustering, which is node-based. The measure $\varphi_{ik}(L)$ is also related to the notion of *link support* defined for undirected networks by Jackson, Rodríguez-Barraquer and Tan (2012). In their case, a link ik is said to be supported in an underlying network g if there exists *some* agent s, different from i and k, such that the links is and ks also belong to g. Thus, in contrast with our measure, theirs is binary and therefore does not quantify the extent of support.

As it turns out, the general measure of link support given by (10) allows for the following topological characterization of EIMs.

Proposition 4 *Given an observation network with adjacency matrix $L \equiv (l_{ij})_{i,j \in N}$, a learning span $K \geq 1$, and a covariation matrix of initial opinions that satisfies $\Phi_\beta \equiv \varpi I$ for some $\varpi > 0$, the EIM $A^* = (a^*_{ij})_{i,j \in N}$ verifies:*

$$a^*_{ik} = a^*_{ii}\, l_{ik}\, \varphi_{ik}((A^*)^K) \qquad (i,k = 1,2,...,n). \tag{11}$$

Proof. See Appendix.

Heuristically, the result in Proposition 4 reflects the idea that common K-order partners help "support" the relationship between i and j, and therefore the more such partners i and j share, the stronger their relationship. More specifically, in our context the support comes from the fact that, by being subject to common (indirect) influence, agents i and k will tend to strengthen the correlation of their behavior and hence, by homophily, their own link as well.

For the particular case of $K = 1$, Proposition 4 leads to the stark implication that link strength between two nodes, i and j, is *directly proportional* to their neighborhood overlap $\varphi_{ik}((A^*))$, i.e. to the (weighted) number of neighbors they have in common. Intuitively, this property suggests that, in endogenous networks, strong links will tend to be arranged in "triangles." The literature – see e.g. Kossinets and Watts, (2006, 2009) and Kumpula *et al.* (2007) – often rationalizes such configurations through the claim that, among strong links, the principle of triadic/transitive closure applies (i.e. the friend of a friend tends to become a friend). Our model, however, provides a quite different explanation for this phenomenon: it is not the strength of the links that brings about the transitivity in connections; rather, it is that only those links that are well supported (and hence form part of various triangles) tend to be strong at equilibrium.

The previous discussion pertains to the *endogenous* support conditions that must be enjoyed by strong links at equilibrium. However, still focusing on the sharpest setup where $K = 1$, analogous considerations also lead to counterpart conditions on the observational adjacency matrix L, an *exogenous* primitive of the model. To formulate it precisely, we use our topological measure of neighborhood overlap to adapt the definition of supported links in Jackson, Rodríguez-Barraquer and Tan (2012) to our context:

Definition 2 *Given the observation network* $L \equiv (l_{ij})_{i,j \in N}$*, an observational link* ik *(*$l_{ik} = 1$*) from* i *to* $k \neq i$ *is* supported *when* $\varphi_{ik}(L \backslash ik) > 0$.

Hence, in light of expression (10), a (directed) observational link ik is supported if after its removal from L both i and k are able to observe at least one agent $s \neq k$ (possibly, $s = i$), i.e., $l_{is} l_{ks} = 1$. The next proposition establishes that unsupported links must have zero weight at equilibrium when learning ends after the first round.

Proposition 5 *Consider any given observation network with adjacency matrix* $L \equiv (l_{ij})_{i,j \in N}$*, and assume that initial opinions satisfy* $\Phi_\beta \equiv \varpi I$ *for some* $\varpi > 0$*. Then, if* $K = 1$*, only supported links may have positive weights in an EIM* $A^* = (a^*_{k\ell})^n_{k,\ell=1}$*:*

$$\forall i, k \in N : l_{ik} = 1,\ a^*_{ik} > 0 \Rightarrow \varphi_{ik}(L \backslash ik) > 0.$$

The support-related considerations studied in this section bear on the important issue of *social fragmentation* that will be studied next in Section 3.4. More specifically, they concern the question of whether an originally segmented society – say, a population that is divided into two disjoint groups that neither influence, nor are influenced by, the other one – may persist in such a polarized state even when cross-group links are created to avert it. Or, reciprocally formulated, the issue is when cross-group "bridging" can be an effective route to integration.

3.4 Bridging

For clarity, we study the problem of bridging and fragmentation in a stylized context where the population N is partitioned into two groups, $G_1 = \{1, 2, ..., n_1\}$ and $G_2 = \{n_1 + 1, n_1 + 2, ..., n_1 + n_2\}$, with $n_1 + n_2 = n$. It is then useful to decompose the observation adjacency matrix L into a matrix that concerns the links between agents of the same group and another one that reflects the observation links between individuals of different groups. Formally, we write $L = L^0 + V$, where the within-group observation matrix $L^0 = (l^0_{ij})^n_{i,j=1}$ satisfies:

$$\left[l^0_{ij} = 1\right] \Rightarrow \left[i, j \in G_q,\ q \in \{1, 2\}\right], \tag{12}$$

and the cross-group counterpart $V = (v_{ij})^n_{i,j=1}$ has:

$$[v_{ij} = 1] \Rightarrow \left[i \in G_q,\ j \in G_{q'},\ q \neq q'\right]. \tag{13}$$

To fix ideas, we may think of L^0 as representing the original situation where the two groups had no contact whatsoever, while the addition of the matrix V embodies the pattern of cross-group observation supplementing the original one. In order to control the confounding effects of correlation in prior beliefs, we assume throughout this section that they are uncorrelated across agents and all display the common individual variation $\varpi > 0$, i.e. $\Phi_\beta = \varpi I$, where I is the identity matrix.

Our objective is to explore the conditions under which the ***possibility*** of cross-group influence created by new cross-group links leads to the ***materialization*** of some significant integration of their opinions. Two are our main results. The first is particularly strong as it applies across all equilibria and it is independent of the connection structure within each group. However, it is limited to situations with one round of learning and to large populations. In contrast, the second result applies to any number of learning rounds and population sizes but it assumes that the observation structure within each group is complete and focuses on stability of segmented configurations. In both cases, we find that segmentation can be a persistent phenomenon: either because it is a feature displayed by all equilibria, or because segmented equilibria are dynamically robust.

3.4.1 The persistence of social segmentation

To illustrate the persistence of social segmentation, we find it useful to start with a framework where the learning spells are short (for simplicity, $K = 1$) and the population is large

$(n \to \infty)$. We also expand on the benchmark model described above and postulate that the pattern of observational links across groups is stochastically generated. More precisely, it is determined by a random matrix $\tilde{V}$ (a random variable) that determines the realized matrix V defining the observation links connecting the two groups. We assume that such a random variable is constructed in a straightforward manner: each agent $i \in G_q$ $(q = 1, 2)$ observes a set of agents selected through a certain number of independent and uniformly distributed distinct draws from $G_{q'}$ $(q' \neq q)$. Thus, from an ex-ante perspective, when we combine within- and across-group observation links, the whole observation structure is given by the random matrix $\tilde{L} = L^0 + \tilde{V}$, where, as before, L^0 is some given adjacency matrix defining local observation.

As indicated, we are interested in studying the problem when the population is large. Formally, we study the limit case where $n \to \infty$ and all other parameters (in particular, node degrees $(d_i)_{i \in N}$ and group sizes $(n_q)_{q=1,2}$ are functions of n. Furthermore, we make the following assumptions:

(L.1) $\lim_{n\to\infty} \frac{n_q(n)}{n} \equiv \overline{n}_q > 0$ for each $q = 1, 2$, i.e. the limit size of each group is fractionally significant.

(L.2) The full observational degree of every agent i, $d_i(n) \equiv \sum_{j \neq i} \left[l^0_{ij}(n) + v_{ij}(n) \right]$, is uniformly bounded, i.e. $\lim_{n\to\infty} d_i(n) \leq D < \infty$ with probability one.

As $n \to \infty$, we trace some corresponding sequence of EIM $\{A^*(n)\}_{n=1,2,\ldots}$, which is arbitrary. Thus, since we know (see e.g. Corollary 1) that our theoretical framework allows for a wide multiplicity of equilibria, we want our result to be strong enough to apply across all of them.

Proposition 6 *Consider a large-population context as described above and suppose that* $K = 1$ *and that (L.1)-(L.2) hold. For each population size* $n \in \mathbb{N}$, *denote by* $p_{ij}(n)$ *the probability that, in some EIM* $A^*(n) = (a^*_{k\ell}(n))^n_{k,\ell=1}$, *agent* $i \in G_q$ *has* $a^*_{ij}(n) > 0$ *for some* $j \in G_{q'}$, $q' \neq q$. *Then,* $\lim_{n\to\infty} p_{ij}(n) = 0$.

Proof. See Appendix.

A simple illustration of the kind of observation network considered by Proposition 6 is provided in Figure 4. In a heuristic sense, it is similar in spirit to the well-known model of small worlds proposed by Watts and Strogatz (1998) where nodes are placed along a boundariless one-dimensional lattice, a "ring". It combines, as in the Watts-Strogatz model, a regular local structure with high clustering with some randomly drawn long-range links (so-called "shortcuts"). The difference with their model is that, in our case, there are two separate groups arranged in corresponding rings, links are directed, and the long range-links only connect nodes of different groups. By virtue of Proposition 6 we can assert that the probability that, at equilibrium, any given cross-group link conveys some significant influence converges to zero as the sizes of the two lattice subnetworks grows unboundedly.

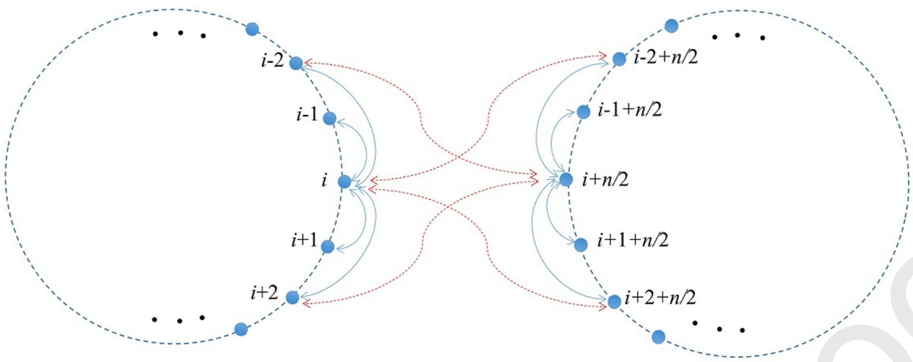

Figure 4: Diagrammatic illustration of the type of observation network studied in Proposition 6. The two groups, G_1 and G_2, are of equal size and their nodes are linearly arranged along corresponding rings. The neighborhoods of typical nodes are illustrated: blue links connect agents within a group, whereas red ones connect agents of different groups. The observation links within each group are taken to be two-sided, i.e. involve reciprocal observation. None of the cross-group links display this characteristic.

An interesting point that can be well explained in terms of the example described in Figure 4 is that the possibility of breaking group-aligned segmentation crucially depends, in general, on having an effective *interplay* between the pattern of cross-group links and the structure of internal connections within each group. In the example, segmentation prevails at any equilibrium because no cross-group observational link is supported in the sense of Definition 2. This can be easily changed, however, even without increasing the total number of cross-group links. What is needed is that the latter be formed by taking into account the internal connections within each group. Cross-group links, in other words, need to rely on the internal connection structure of each group to gather the support that is required to break the segmentation. For example, the link between i and $i-2+n/2$ would be supported if $i-1$ were also connected to the latter agent. And clearly, analogous changes could be implemented for every other cross-group link in a similar situation.

Another different way to tackle the problem would focus on the internal structure of observational links within each group. For example, if this structure were extended by doubling the radius of observation along the ring from 2 to 4, all cross-group links would be supported and again segmentation could be broken. Interestingly, this illustrates the important point that increasing the internal connectivity of groups does not necessarily render them more impervious to external influence. In some cases, as in the example discussed, it is precisely such an increase in the internal density of observational connections that renders the external links effective bridges.

3.4.2 The dynamic stability of social segmentation

Here we particularize the general bridging setup embodied by conditions (12)-(13) and posit that the observation network within each group is complete. That is, we assume that the within-group adjacency matrix for observation $L^0 = (l^0_{ij})^n_{i,j=1}$ is characterized by:

$$\forall i, j\ (i \neq j), \quad l^0_{ij} = 1 \Leftrightarrow [i, j \in G_q,\ q = 1, 2]. \tag{14}$$

Our analysis will focus on the implications of alternative patterns of cross-group observation as defined by the given adjacency matrix $V = (v_{ij})^n_{i,j=1}$. Furthermore, the approach will be dynamic, in that we shall identify bridging conditions that render segmentation dynamically robust, i.e. locally stable for influence-adjustment process implicitly underlying our equilibrium EIM notion. To address precisely this issue, we need to describe the initial conditions from which the adjustment process starts, as well as define formally such an adjustment process. Next, we address each of these in turn.

In view of Proposition 3, as the starting conditions of the adjustment process, it is natural to focus on the EIM that is completely connected within each group but disconnected across groups. This is the equilibrium given by the influence matrix

$$A^0 = \begin{pmatrix} Q(n_1) & 0_{n_1 \times n_2} \\ 0_{n_2 \times n_1} & Q(n_2) \end{pmatrix}, \tag{15}$$

where $0_{n_r \times n_s}$ stands for a matrix of dimension $n_r \times n_s$ consisting of all zeros and $Q(n_r)$ represents the uniform matrix with all entries equal to $1/n_r$.

Note that, as stated in Corollary 1, the influence matrix A^0 defines an EIM independently of the number of observation links *across* the two groups. So, in this light, the precise question we want to tackle is: When is such an equilibrium asymptotically (locally) stable? Or, instead, can a perturbation trigger enough cross-group influence such that segmentation is broken?

The relevant *dynamic adjustment process* has been described in Subsection 2.4 and it boils down to the synchronous dynamic system operating on $n \times n$ row-stochastic matrices,

$$A(s+1) = F(A(s)) \qquad (s = 1, 2, ...), \tag{16}$$

where $F(\cdot)$ has been defined by (7).

As advanced, our main result in this subsection specifies *sufficient* conditions for the segmented EIM A^0 in (15) to be asymptotically stable for the dynamical system (16) when the prevailing observation matrix is $L = L^0 + V$.

PROPOSITION 7 *Consider the observation matrix $L = L^0 + V$, where L^0 and V satisfy (12) and (13), respectively. Then, the segmented EIM A^0 in (15) is asymptotically stable for the* *dynamical system (16) when each cross-group link ij ($v_{ij} = 1$) satisfies the following condition, strictly for at least one link:*

$$n^i\, v^i + n^j\, v^j + (K-1)v \leq n^i \sqrt{n_1 n_2} \tag{17}$$

where, for each $k \in N$, $n^k = \sum_{\ell \in N} l^0_{k\ell}$ is the number of agents that player k observes in the group she belongs to, $v^k = \sum_{\ell \in N} v_{k\ell}$ is the number of agents observed by k in the other group, and $v = \sum_{k \in N} v^k$.

Proof. See Appendix.

To understand the intuition underlying Proposition 7, let us consider first the simple case where both groups are symmetric (i.e. $n_1 = n_2 = n/2$) and there is just one round of learning ($K = 1$). Then, (17) simply becomes $v^i + v^j \leq n/2$. This indicates that the stability of segmentation obtains if, for each cross-group link ij, agents i and j do not observe jointly too many individuals from the other group, with the upper bound being proportional to the population size. Then, the following two questions arise. Why is it that, in evaluating the destabilizing potential of a particular observational – and therefore *directional* – link from i to j, the volume of observations in *both* directions is relevant? Why is the upper bound related to the population size?

To answer the first question, note that i's and j's beliefs are highly correlated with those of the individuals in their own group; therefore, if i and j observe jointly only few individuals from the other group, their beliefs after one round of learning will be only weakly correlated and hence their influence on each other will tend to vanish after the equilibrium is perturbed. On the other hand, concerning the role played by population size in the stability condition, the point to understand here is that the destabilizing impact on segmentation resulting from cross-group observation has to be balanced against the pull towards restoring it that is induced by the internal observation links existing within each group. Therefore, the relative strength of the latter force compared to the former one depends on the size of the groups.

Finally, let us consider how the former considerations are affected if there is more than just one round of learning (i.e. $K \geq 2$) and/or the groups are asymmetric in size ($n_1 \neq n_2$). Mere inspection of (17) indicates that the higher K and the larger the size asymmetry the harder it is, *ceteris paribus*, to satisfy that stability condition. This is of course in line with intuition. First, multiple rounds of learning can only strengthen the effect of external influence, either directly (by the repeated observation of the other group through one's own links) or indirectly (through internal influence, which partly reflects the external observations conducted by other members of the same group). Second, size asymmetry favors the breakup of segmentation because, as one of the groups becomes relatively smaller, it is easier for external influence to have an impact on it.

From the proof of Proposition 7 it follows that, if all inequalities in (17) are reversed for each cross-group link and at least one of them holds strictly, then the segmented equilibrium will be destabilized by small perturbations. The intuition underlying this result is of course polar to that explained above for the stability of group-based segmentation.

Proposition 7 also allows for a useful and quite transparent analysis of the conditions under which segmentation is a robust phenomenon for large groups – for example, when the phenomenon is studied at a whole national level and the two groups consist of those that support a given political party, speak a different mother tongue, or live in a different region. This is the setup considered by the following corollary. In it we postulate, as in Subsection 3.4.1, that $n \to \infty$ and the relevant parameters change accordingly to a given function of n. Specifically, we assume:

(M.1) $\lim_{n\to\infty} \frac{n_q(n)}{n} \equiv \overline{n}_q > 0$ for each $q = 1, 2$, i.e. the limit size of each group is fractionally significant.

(M.2) There exists some function $b(n)$ – possibly growing unboundedly with n – such that for all $i \in N$ and $n \in \mathbb{N}$, the corresponding cross-group degree $v^i(n) \leq b(n)$ with $\lim_{n\to\infty} \frac{b(n)}{n} = 0$.

Condition (M.1) is exactly as the former (L.1), whereas (M.2) is weaker than (L.2) as the the cross-group degree of agents is allowed to grow unboundedly with n, but not comparably fast.[16]

Corollary 2 *Consider a large population as described above and suppose that K is constant*[17] *and (M.1)-(M.2) hold. Then, there exists some $\hat{n} \in \mathbb{N}$ such that for all $n \geq \hat{n}$ the corresponding segmented EIM $A^0(n)$ as given in (15) is asymptotically stable for the dynamical system (16).*

Proof. See Appendix.

This result indicates that segmentation is a robust phenomenon for sufficiently large groups with dense within-group connection structures even if the number of cross-group links grows unboundedly (but slower than $b(n)$).

We close our discussion in this section with a remark that, in line with what was explained in Subsection 3.4.1, illustrates that the robustness of segmented equilibrium not only depends on the number of cross-group links but also on their pattern of connection.

Remark 4 (Bridging and observation structure) *To bring our point across in a transparent manner, in this Remark we consider the simple case where the population is divided into*

[16] Thus, for example, the growth could be – as is often posited in the asymptotic theory of random networks – logarithmic in n, but it could also be faster, e.g. at the rate of $\sqrt{n}$.

[17] It is easy to see that this assumption could be generalized to allow for K to grow but at a rate lower than n

two groups of equal size (i.e. $n_1 = n_2 = \frac{n}{2}$), the within-group observation matrix L^0 satisfies (14), and there is just one round in every learning spell ($K = 1$). In this common context, we consider the following two scenarios.

First suppose that every agent in each group observes less than half of the individuals in the other group, i.e.

$$\forall i \in N, \quad v^i < n/4.$$

Then,

$$v^i + v^j < \frac{n}{2}, \tag{18}$$

and the particularization of condition (17) to this case is satisfied. The segmented equilibrium is thus locally stable even if the total number of cross-group links is of order n^2.

Second, we show that, in contrast to the previous case, the segmentation can be destabilized with a substantially smaller number of cross-group links if these are set in a more asymmetric manner. As a simple illustration, suppose that only one agent, say 1, in G_1, has bi-directional cross-group links to all $n/2$ agents in the other group, i.e., $v_{1k} = v_{k1} = 1$ and $v_{ik} = v_{ki} = 0$ for all $k \in G_2$ and $i \neq 1$. Then, we find that for each cross-group link $1k$ and $k1$,

$$v^1 + v^k = \frac{n}{2} + 1 > \frac{n}{2}.$$

Thus, the inequality in (18) is reversed, which means the segmented equilibrium is destabilized. Importantly, this is achieved with a total of n directed cross-group links (that is, $\frac{n}{2}$ bidirectional links), which for large n represents a much fewer number than in the previous construction, where their total number was allowed to grow at the order of n^2! Therefore, we conclude that what matters is not just the number of links across the two groups but the specific pattern in which they are arranged.

3.5 Adaptive prior beliefs

We close our discussion by exploring the implications of relaxing the assumption that the prior beliefs $\beta_{iq}(s)$ of agent i on issue q at the beginning of the learning spell s do not depend on her final beliefs in the previous spell. An alternative interesting scenario would be one where the final beliefs prevailing at the end of one spell fully determine the prior beliefs from which the following spell starts. Intuitively, such a formulation would suitably model a situation where, in essence, the same (or very similar) questions recurrently arise in each dimension and in every learning spell.

Arguably, the real world (and hence the ideal model) is somewhere between these two extremes. Here, we briefly explore this option and study a convex combination of the two extreme scenarios. Specifically, we postulate that each agent $i = 1, ..., n$ computes her prior beliefs β^s_{iq} on issue $q = 1, ..., m$ at the start of each spell $s = 1, 2, ...$ as follows:[18]

$$\beta^s_{iq} = \mu \cdot x^{s-1}_{iq} + (1-\mu)\beta_{iq}, \tag{19}$$

where $\mu \in [0,1]$ is the weight this agent attributes to the belief x^{s-1}_{iq} that she held at the end of the previous K-long learning spell $s-1$, and we set $x^0_{iq} = \beta_{iq}$ for all i and some corresponding fixed belief β_{iq}.

A steady state of the learning process is given by beliefs $\hat{x}$ that satisfy:

$$\hat{x} = A^K(\mu \cdot \hat{x} + (1-\mu)\beta), \tag{20}$$

where recall that $\beta = (\beta_{iq})^{n,m}_{i,q=1}$. If $\mu = 0$, we obtain a special case of our original model with $\beta^s_{iq} = \beta_{iq}$ and, hence, $\Phi_{\beta^s} = \Phi_\beta$ for all s. If $\mu = 1$ the model is formally identical to the case where $K \to \infty$. That is, for fixed A, opinions converge to a vector $\hat{\boldsymbol{x}}$ that satisfies:

$$\hat{x} = \lim_{s\to\infty} (A^K)^s \beta = \lim_{s\to\infty} A^s \beta, \tag{21}$$

which is uniquely determined if A is an aperiodic matrix. Clearly, it is a stationary point for every learning spell, i.e. $\hat{x} = A^K \hat{x}$.

On the other hand, when $\mu < 1$, the final beliefs $\hat{\boldsymbol{x}}$ in a steady state are found to be

$$\hat{x} = A^K(\mu \cdot \hat{x} + (1-\mu)\beta) = (1-\mu)(I - \mu A^K)^{-1} A^K \beta \equiv B_\mu \beta, \tag{22}$$

where B_μ is a row-stochastic matrix.[19]

In line with Definition 1, a row-stochastic matrix $A^* = (a^*_{ij})^n_{i,j=1}$ is said to be an EIM if it satisfies for each agent i and j the homophily condition $a^*_{ij} = a^*_{ii} l_{ij} \hat{\rho}_{ij}(\hat{x})$, given the steady state $\hat{x}$ it induces from either (21) or (22) (depending on whether $\mu = 1$ or $\mu < 1$ respectively) and the corresponding pattern of similarity coefficients $\hat{\rho}_{ij}(\hat{x})$. It can be readily verified that Propositions 1 and 3 directly generalize to the present scenario with $\mu \in [0,1]$, while former Proposition 2 can be extended as follows.

Proposition 8 *Consider a (weakly) connected observation network with adjacency matrix $L \equiv (l_{ij})_{i,j\in N}$ and any matrix β of agents' initial opinions with $\Phi_\beta \geq 0$. Then, if either $K \to \infty$ and $\mu \in [0,1]$, or $K \geq 1$ and $\mu = 1$, the unique connected EIM $A^* = (a^*_{ij})_{i,j\in N}$ is given by*

$$a^*_{ij} = \frac{l_{ij}}{\sum_{k\in N} l_{ik}} \qquad (i,j = 1,2,...,n),$$

and the induced similarity coefficients among end opinions satisfy $\hat{\rho}_{ij}(A^*) = 1$ for all $i, j \in N$.

[18] This formulation is reminiscent of that proposed by Friedkin and Johnsen (1999). In their model, the postulated belief adjustment rule prevailing over time places some fixed weight on a given vector of initial beliefs that play an "anchoring" role in the belief adjustment rule applying every period. Thus, it is a feature that concerns the process that, in our terminology, unfolds *within* a learning spell. Instead, in our case the given initial beliefs play an analogous anchoring role when the starting beliefs of a learning spell are adjusted *across* consecutive such spells.

[19] By the expansion $(I - \mu A^K)^{-1} = \sum_{s=0}^{+\infty} \mu^s (A^K)^s$ for $\mu \in (0,1)$ one obtains $B_\mu = (1-\mu)\sum_{s=0}^{\infty} \mu^s A^{K(s+1)}$.

Proof. See Appendix.

How does the generalization considered here bear on our former analysis of segmentation? In particular, how does the gradual adjustment of the prior beliefs in learning spells (i.e. a value of $\mu > 0$) affect the possibility of social integration in an originally segmented population? Proposition 8 and (21)-(22) suggest that μ plays a similar role to the learning span K – that is, increasing μ leads, *ceteris paribus*, to higher correlations among agents' beliefs and therefore favors integration.

The following numerical analysis illustrates the trade-off between K and μ in the following setup. A population consisting of 10 individuals is divided into two groups, the first group being the singleton $G_1 = \{1\}$, while the other group $G_2 = \{2, 3, ..., 10\}$ includes all other agents in the population. Suppose there is only one cross-group link, $l_{21} = 1$, i.e., agent 2 observes player 1, but within group G_2 the observation structure is complete. Consider now an equilibrium configuration where the corresponding EIM A^0 is as given in (15) for $n_1 = 1$ and $n_2 = 9$. Then, if $\mu = 0$, a sufficient condition ensuring that belief homophily destabilizes such an equilibrium is that the inequality in (17) is reversed for the cross-link 21,

$$9 + (K - 1) > 9\sqrt{9} \Rightarrow K > 19. \tag{23}$$

To check whether high enough values of μ can destabilize segmentation for a fixed value of K, we have relied on (20) to set up a generalized version of the influence-adjustment dynamics (7) for any given $\mu \in [0, 1]$. Applying then this law of motion to iterate influence matrices, we find an EIM A^* as its fixed point when starting from an influence matrix that puts a "small" weight on the single cross-group link 21. At the induced equilibrium, agent 1 will be able to influence directly player 2 and then, indirectly, all other followers $\{3, 4, ..., 10\}$ only if the corresponding equilibrium weight of that link, a^*_{21}, is strictly positive. In Figure 5, we report the results of this exercise for $K \in \{5, 10, 15, 20, 25\}$.

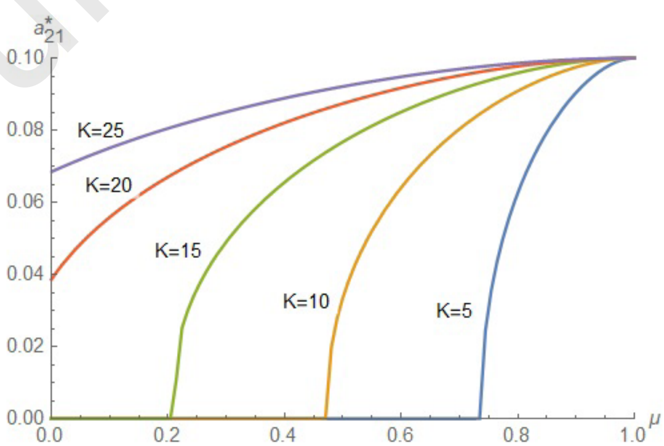

Figure 5: The impact of the parameter μ on the EIM element a^*_{21}.

We observe that agent 1 starts exercising influence after μ exceeds some threshold $\bar{\mu}(K)$, where $\bar{\mu}(K)$ decreases in K. In particular, $\bar{\mu}(20) = \bar{\mu}(25) = 0$, which is consistent with (23). We also notice that, independently of K, the equilibrium weight a^*_{21} converges to 0.1 as μ approaches one, i.e., it corresponds to the even distribution of attention among ten observation links of agent 2, which is in line with conclusion to this effect established by Proposition 8.

4 Conclusion

This paper studies a model of social learning on an endogenous social network where the learning framework extends the classical one proposed by DeGroot (1974) in two relevant dimensions:

- agents' opinions are multidimensional;
- the number of learning rounds is arbitrary.

In such a generalized framework, the learning outcome typically falls short of full consensus, which then allows us to identify the extent to which agents' final positions correlate, depending on the nature of their initial opinions, the architecture of the network, and the length of the learning process.

Our main objective in this paper has been to develop a theory that endogenizes the influence network through which social learning unfolds. This network is required to satisfy a two-fold requirement:

(a) it must respect the communication restrictions imposed by some exogenously given observation network;

(b) it has to be consistent with a notion of homophily whereby the strength of each bilateral relationship (influence) is proportional to the corresponding correlation of behavior (opinions).

We have shown that an important consequence of homophily-based consistency is that, in equilibrium, the strength of every link has to be proportional to what we have coined its "network support." This in turn has important implications for the problem of social integration, i.e. the question of whether groups that were originally forming their opinions independently may come to integrate their views through successful "bridges of influence." We have identified conditions that render group segmentation either a robust or a fragile state of affairs, relating those conditions to various features of the model, in particular to the parameters that determine the cross-group connectivity, the relative group sizes, or the span of the learning process.

Social integration is an important social problem because it often has serious welfare consequences. However, a proper study of its normative implications requires a proper assessment of its benefits, which in principle may be positive or negative. For, indeed, depending on how integration is defined, it may be the case that not always more of it is better – for example, along certain dimensions too much integration can be detrimental to the preservation of valuable diversity. To incorporate such considerations into the analysis should be one of the prominent objectives of future research.

In a related vein, another (largely complementary) focus of future research ought to be the introduction of some extent of payoff-guided behavior into the learning environment. The model we have studied here is purely behavioral. Other paradigms of social learning that may be considered include, e.g., observational learning (Bala and Goyal, 1998), Bayesian learning (Gale and Kariv, 2003), or a mixture of boundedly-rational and Bayesian learning (Mueller-Frank, 2014). It is conceivable that when such alternative forms of learning are combined with homophily-based network formation, interesting new perspectives on the problem may open up.

5 Appendix

Lemma 1 *The matrix* $\Phi_{\boldsymbol{x}} = \left(\phi^{x}_{ij}\right)_{i,j\in N}$ *of bilateral covariations (2) among end (stage-K) opinions* $x = A^K\beta$ *can be computed as*

$$\Phi_x = A^K\,\Phi_\beta\,(A^K)',$$

where $\Phi_\beta \geq 0$ *is the covariation matrix of the prior beliefs* β*,* A *is a row-stochastic influence matrix and* $K \geq 1$.

Proof: Let $B = A^K$ and note that A and B are row-stochastic. As posited in (3), given the end-beliefs $x = B\beta$, define for each issue q the corresponding benchmark $g_{iq}(x)$ as $g_{iq}(x) = \lambda b_q + (1-\lambda)\sum_{s=1}^{m} w_s x_{is}$. First, we prove that the benchmarks $G(x) = (g_{iq}(x))_{i,q=1}^{n,m}$ can be computed from the benchmarks $G(\beta) = (g_{iq}(\beta))_{i,q=1}^{n,m}$ of prior beliefs β,

$$g_{iq}(\beta) = \lambda b_q + (1-\lambda)\textstyle\sum_{s=1}^{m} w_s\beta_{is},$$

as $G(x) = BG(\beta)$. Below, we use the fact that $\sum_{k=1}^{n} B_{ik} = 1$ as B is row-stochastic:

$$\begin{aligned}(BG(\beta))_{iq} &= \textstyle\sum_{k=1}^{n} B_{ik}g_{kq}(\beta) = \lambda b_q\sum_{k=1}^{n} B_{ik} + (1-\lambda)\sum_{k=1}^{n}(B_{ik}\sum_{s=1}^{m} w_s\beta_{ks})\\ &= \lambda b_q + (1-\lambda)\textstyle\sum_{s=1}^{m} w_s x_{is} = g_{iq}(x) = (G(x))_{iq},\end{aligned}$$

It follows that for the "benchmarked" prior beliefs $\widetilde{\beta} = \beta - G(\beta)$ and end-beliefs $\widetilde{x} = x - G(x)$, it holds that:

$$\widetilde{x} = x - G(x) = B\beta - BG(\beta) = B(\beta - G(\beta)) = B\widetilde{\beta}.$$

Then, the similarity measure (2) is computed as:

$$\begin{aligned}\phi^x_{ij} = \textstyle\sum_{q=1}^{m}\omega_q\widetilde{x}_{iq}\widetilde{x}_{jq} = \sum_{q=1}^{m}\omega_q(\sum_{k=1}^{n} B_{ik}\widetilde{\beta}_{kq})(\sum_{k=1}^{n} B_{jk}\widetilde{\beta}_{kq}) =\\ \textstyle\sum_{q=1}^{m}\omega_q(B_{i1}B_{j1}\widetilde{\beta}^2_{1q} + B_{i1}B_{j2}\widetilde{\beta}_{1q}\widetilde{\beta}_{2q} + \ldots + B_{in}B_{jn}\widetilde{\beta}^2_{nq}) =\\ B_{i1}B_{j1}\phi^\beta_{11} + B_{i1}B_{j2}\phi^\beta_{12} + \ldots + B_{in}B_{jn}\phi^\beta_{nn} = (B\,\Phi_\beta\,B')_{ij},\end{aligned}$$

where in the last line we made the substitution

$$\phi^\beta_{ij} = \sum_{q=1}^{m}\omega_q\widetilde{\beta}_{iq}\widetilde{\beta}_{jq}. \qquad \boxtimes$$

Proof of Proposition 1: For a given (finite) learning span K, observation matrix $L \in \{0,1\}^{n\times n}$ and the initial covariation matrix Φ_β, the vector field $F(A; K, L, \Phi_\beta) : (\Delta^{n-1})^n \to (\Delta^{n-1})^n$, defined by (7), maps an $n-$dimensional stochastic matrix A into another $n-$dimensional stochastic matrix (Δ^{n-1} is an $n-$dimensional simplex). $F(\cdot)$ is continuous as it involves only a finite number of continuous matrix operations when $K < \infty$.

As $(\Delta^{n-1})^n$ is compact and convex, Brouwer fixed-point theorem implies that F has a fixed point A^*.

When $K \to \infty$, a possible issue is that A^K is not necessarily convergent and, then, $F(\cdot)$ is not well-defined. We restrict in case $K \to \infty$ the domain of $F(\cdot)$ to stochastic matrices $A \in (\Delta^{n-1})^n$ that satisfy $a_{ii} \geq a_{ik}$ for each $i, k \in N$ (by (7), $F(\cdot)$ returns only such matrices). Then, $\lim_{K\to\infty} A^K$ is convergent by Theorem 2 in the Mathematical Appendix A in Golub & Jackson (2010). As this restricted domain is compact and convex and $F(\cdot)$ defined on it is continuous also for $K \to \infty$, Brouwer fixed-point theorem implies that F has a fixed point A^*.

Finally, equation (9) for an EIM A^* follows from the condition (8) in the Definition 1 that implies,

$$\frac{a_{ik}^*}{a_{ii}^*} = \frac{l_{ik}\hat{\rho}_{ik}(A^*)}{l_{ii}\hat{\rho}_{ii}(A^*)} = l_{ik}\hat{\rho}_{ik}(A^*),$$

as $l_{ii} = \hat{\rho}_{ii}(A^*) = 1$ and, hence, $a_{ii}^* > 0$ for each $i \in N$. ⊠

Proof of Proposition 2: If A^* is an EIM for a fixed adjacency matrix $L \in \{0,1\}^{n\times n}$ and an initial covariation matrix $\Phi_\beta \geq 0$, then $a_{ii}^* > 0$ because $l_{ii} = 1$ and $\rho_{ii}(A^*) = 1$ for each $i \in N$. Hence, A^* must be aperiodic (e.g., Golub & Jackson, 2010). It is well known that for a connected (i.e., irreducible), aperiodic and stochastic matrix A^* each row of $B = \lim_{K\to\infty}(A^*)^K$ is equal to the left eigenvector of A^* associated to the eigenvalue 1. Then, for $\Phi_\beta \geq 0$, the matrix $\Phi_{\hat{x}} = B\Phi_\beta B'$ of bilateral covariations among end-stage opinions $\hat{x} = B\beta$ is such that $\phi_{ij}^{\hat{x}} = c > 0$. It follows that all correlations $\rho_{ik}(.)$ computed by (6) from $\Phi_{\hat{x}}$ for $i, k \in N$ are equal to one. Substituting unit correlations into the definition (8) of EIM yields the claim, $a_{ik}^* = l_{ik}/\sum_{s=1}^N l_{is}$ for all $i, k \in N$. The matrix A^* is then (weakly) connected as we assume this property for the observation network L. ⊠

Proof of Proposition 3: We start with the following Lemma.

Lemma 2 *For row-stochastic and strictly positive $n \times n$ matrices $A \equiv (a_{ik})_{i,k\in N}$ and $S \equiv (s_{ik})_{i,k\in N}$ such that $A \neq Q(n)$, $a_{ii} \geq a_{ik}$ and $a_{ik}/a_{ii} = a_{ki}/a_{kk}$ for all $i, k = 1, ..., n$, we have:*

$$\delta(A) < \delta(AS'), \quad where \quad \delta(X) \equiv \min_{i,k} \frac{x_{ik}x_{ki}}{x_{ii}x_{kk}}. \tag{24}$$

Proof: As the matrices A and S are strictly positive and row-stochastic, it holds that, for all $i, k = 1, ..., n$,

$$\begin{aligned}
&\min_s a_{is} \leq \textstyle\sum_{s=1}^n a_{is}s_{ks} = (AS')_{ik} \leq \max_s a_{is} = a_{ii} \Rightarrow \qquad (25)\\
&\min_s a_{is} \leq \min_s (AS')_{is} \leq (AS')_{ii} \leq \max_s (AS')_{is} \leq a_{ii} \Rightarrow\\
&\min_s \frac{a_{is}}{a_{ii}} \leq \min_s \frac{(AS')_{is}}{(AS')_{ii}}, \quad \forall i \Rightarrow \min_{i,s} \frac{a_{is}}{a_{ii}} \leq \min_{i,s} \frac{(AS')_{is}}{(AS')_{ii}}.
\end{aligned}$$

Note that $a_{ii} = \max_s a_{is} > \min_s a_{is}$ for at least one i. Otherwise, $a_{ii} = a_{is}$ for all i, s which implies $a_{is} = 1/n$ as A is row-stochastic. This, however, would contradict $A \neq Q(n)$. Therefore, the inequalities in (25) are strict for at least one i. Then, the last inequality in (25) is also strict, following from it and from the symmetry condition $a_{ik}/a_{ii} = a_{ki}/a_{kk}$ that

$$\delta(A) = \min_{i,k} \frac{a_{ik}a_{ki}}{a_{ii}a_{kk}} = \min_{i,k}(\frac{a_{ik}}{a_{ii}})^2 = (\min_{i,k} \frac{a_{ik}}{a_{ii}})^2$$
$$< (\min_{i,k} \frac{(AS')_{ik}}{(AS')_{ii}})^2 \leq \min_{i,k} \frac{(AS')_{ik}(AS')_{ki}}{(AS')_{ii}(AS')_{kk}} = \delta(AS').$$

⊠

To proceed with the proof of the proposition, first we note that $Q(n)$ is an EIM since $Q(n) = Q(n)' = Q(n)^K$ and $Q(n)XQ(n) = const \times Q(n)$ for any matrix X. For, in view of (6), we have that $\rho_{ik}(Q(n)) = 1$ for all $i, k \in N$. The substitution of unit correlations into the definition (8) shows that $Q(n)$ is an EIM for the completely connected observation network L.

To prove that $Q(n)$ is the unique connected EIM is somewhat more demanding. First, we show that any connected EIM $A \equiv (a_{ik})_{i,k \in N}$ on completely connected L must be strictly positive. To see this, consider nodes i, j and k such that $a_{ij}a_{jk} > 0$,

$$a_{ij}a_{jk} > 0 \Rightarrow a_{ij} > 0 \Rightarrow \rho_{ij}(A) = \rho_{ji}(A) > 0 \Rightarrow a_{ji} > 0,$$
$$a_{ij}a_{jk} > 0 \Rightarrow a_{jk} > 0 \Rightarrow \rho_{jk}(A) = \rho_{kj}(A) > 0 \Rightarrow a_{kj} > 0.$$

Hence, positive correlations $\rho_{ij}(A)$ and $\rho_{jk}(A)$ imply that $\rho_{ik}(A)$ is also positive,

$$\rho_{ij}(A)\rho_{jk}(A) > 0 \Rightarrow \rho_{ik}(A) > 0.$$

We obtain, therefore,

$$a_{ij}a_{jk} > 0 \Rightarrow a_{ik}a_{ki} > 0.$$

As A is connected, this argument propagates to all links in L. We conclude, therefore, that all elements in the connected EIM A are strictly positive.

On the other hand, condition (8) implies that, given the covariation matrix

$$C = (c_{ik})_{i,k \in N} \equiv A^K \Phi_\beta (A^K)',$$

the EIM A satisfies the following equalities:

$$\frac{a_{ik}}{a_{ii}} = \frac{a_{ki}}{a_{kk}} = \rho_{ik}(A) = \frac{c_{ik}}{(c_{ii}c_{kk})^{1/2}} \Rightarrow \frac{a_{ik}a_{ki}}{a_{ii}a_{kk}} = \frac{c_{ik}c_{ki}}{c_{ii}c_{kk}}, \quad \forall i, k \in N. \tag{26}$$

A necessary condition for (26) is

$$\delta(A) = \delta(C), \quad where \quad \delta(X) \equiv \min_{i,k} \frac{x_{ik}x_{ki}}{x_{ii}x_{kk}} \quad for \quad X \equiv (x_{ik})_{i,k \in N}. \tag{27}$$

where $\delta(\cdot)$ is the matrix operator defined in (24). The following properties of $\delta(\cdot)$ are easily verified:

$$\delta(X') = \delta(X), \quad \delta(cX) = \delta(X), \quad \delta(D_1 X D_2) = \delta(X),$$

where X is an arbitrary matrix with positive entries, D_1 and D_2 are diagonal matrices and c is a constant. By the last property, we obtain from (27),

$$\delta(A) = \delta(C) = \delta(CD_2) \equiv \delta(AS'), \tag{28}$$

where $S' \equiv A^{K-1}\Phi_\beta(A^K)'D_2$ and D_2 is a diagonal matrix that normalizes the sum of each column in $A^{K-1}\Phi_\beta(A^K)'$. Hence, S' is a column-stochastic strictly positive matrix for any $K \geq 1$. By Lemma 2, we know that, for strictly positive row-stochastic matrices S and $A \neq Q(n)$ such that $a_{ii} \geq a_{ik}$ and $a_{ik}/a_{ii} = a_{ki}/a_{kk}$ for all i, k, $\delta(A) < \delta(AS')$. As this contradicts (28), we conclude that only $A = Q(n)$ can be an EIM for the completely connected L. ⊠

Proof of Proposition 4: For the learning span $K \geq 1$, the covariation matrix $\Phi_\beta = (\phi^\beta_{ik})_{i,k\in N}$, the EIM A and the matrix $B = (b_{ik})_{i,k\in N} \equiv A^K$, the correlation (6) between final beliefs $\hat{x}_i(A)$ and $\hat{x}_k(A)$, i.e., rows i and k in the matrix $\hat{\boldsymbol{x}}(A) = B\beta$, is computed as

$$\hat{\rho}_{ik}(A) = \frac{\hat{\phi}^{\hat{x}}_{ik}}{(\hat{\phi}^{\hat{x}}_{ii})^{1/2}(\hat{\phi}^{\hat{x}}_{kk})^{1/2}} = \frac{(B\Phi_\beta B')_{ik}}{\sqrt{(B\Phi_\beta B')_{ii}}\sqrt{(B\Phi_\beta B')_{kk}}}, \tag{29}$$

$$(B\Phi_\beta B')_{ik} = \sum_{s=1}^{n}\sum_{t=1}^{n} b_{is}b_{kt}\phi^\beta_{st} = \sum_{s=1}^{n} b_{is}b_{ks}\phi^\beta_{ss} + \sum_{s=1}^{n}\sum_{t=1,t\neq s}^{n} b_{is}b_{kt}\phi^\beta_{st}.$$

When $\phi^\beta_{st} = 0$ for all $t \neq s$ and $\phi^\beta_{ss} = \varpi$ for all $s = 1, ..., n$, then $\hat{\rho}_{ik}(A)$ boils down to the neighborhood overlap $\varphi_{ik}(B)$. The claim follows then from (9). ⊠

Proof of Proposition 5: For the sake of contradiction, suppose the equilibrium influence $a^*_{ik} > 0$ of a link ik ($l_{ik} = 1$) that is not supported. Then, Definition 2 and (10) imply that:

$$\varphi_{ik}(L\backslash ik) = 0 \Rightarrow l_{is}l_{ks} = 0 \Rightarrow a^*_{is}a^*_{ks} = 0, \quad \forall s \neq k.$$

By Proposition 4 and the formula (10), we can calculate then a^*_{ik} as follows:

$$a^*_{ik} = a^*_{ii}\, l_{ik}\, \varphi_{ik}(A^*) = a^*_{ii}\frac{a^*_{ik}a^*_{kk}}{(\sum_{s=1}^n (a^*_{is})^2)^{\frac{1}{2}}(\sum_{s=1}^n (a^*_{ks})^2)^{\frac{1}{2}}} \leq a^*_{ik}, \tag{30}$$

where the last inequality follows because $\sum_{s=1}^n (a^*_{rs})^2 \geq (a^*_{rr})^2$ for $r \in \{i, k\}$. If the inequality is strict, we have reached a contradiction. Otherwise, it must hold that $\sum_{s=1}^n (a^*_{rs})^2 = (a^*_{rr})^2$ for $r \in \{i, k\}$. But this implies $a^*_{ik} = 0$. ⊠

Proof of Proposition 6: For any given population size n, let the *realized* observation network be represented by the adjacency matrix $L(n) = L^0(n) + V(n)$. Consider a pair of agents, $i, j \in N$, such that $i \in G_q$ and $j \in G_{q'}$ ($q' \neq q$), while $v_{ij}(n) = 1$; that is, agent

i observes j through a cross-group (directed) link ij. We know from Proposition 5 that $a^*_{ij}(n) > 0$ in an EIM $A^*(n)$ implies that the link ij is supported. In order to prove our claim, it suffices to show that the probability of the latter event - which we denote $E_{ij}(n)$ - vanishes as n grows without bounds. We note that $E_{ij}(n)$ can arise only if at least one the following two further constituent events occur:

- $E^1_{ij}(n)$: There is a cross-group link from i to some agent $s \in G_{q'} \backslash j$ whom j also observes, i.e. $l^0_{is}(n) l^0_{js}(n) = 1$.
- $E^2_{ij}(n)$: There is a cross-group link from j to some agent $s' \in G_q$ whom i also observes, i.e. $l^0_{js'}(n) l^0_{is'}(n) = 1$.

Denote by $r^1_{ij}(n)$ and $r^2_{ij}(n)$ the (conditional) probabilities of events $E^1_{ij}(n)$ and $E^2_{ij}(n)$, respectively, and by $q_{ij}(n)$ the probability of the event $E_{ij}(n) = E^1_{ij}(n) \cup E^2_{ij}(n)$. Then, noting that (L.2) implies that neither agent i nor j can: (i) have more that D draws to establish cross-group observational links, or (ii) observe more than D agents from her own group, the two former probabilities can be bounded as follows:

$$r^1_{ij}(n) \leq \left[\frac{1}{n_{q'} - D} D^2\right]^D$$
$$r^2_{ij}(n) \leq \left[\frac{1}{n_q - D} D^2\right]^D,$$

while for the latter one we can write:

$$q_{ij}(n) = 1 - (1 - r^1_{ij}(n)(1 - r^2_{ij}(n) \leq 1 - \left[1 - \frac{1}{n_{q'} - D} D^2\right]^D \left[1 - \frac{1}{n_q - D} D^2\right]^D.$$

Then, it follows that

$$\lim_{n\to\infty} q_{ij}(n) = 0 \tag{31}$$

since, by (L.1), we have:

$$\lim_{n\to\infty} n_q(n) = \lim_{n\to\infty} \overline{n}_q\, n = \infty \qquad (q = 1, 2). \qquad \boxtimes$$

Proof of Proposition 7: We start the proof by introducing some notation and establishing two auxiliary lemmas.

Let $\varpi = (\varpi_1, ..., \varpi_n)$ and define $I(\varpi)$ as an $n \times n$ diagonal covariation matrix of prior beliefs with the diagonal elements $(I(\varpi))_{ii} = \varpi_i > 0$. For the learning span $K \geq 1$, the $n \times n$ influence matrix $A \equiv (a_{ik})_{i,k\in N} \geq 0$ and the binary adjacency matrix $L \equiv (l_{ik})_{i,k\in N} \in \{0,1\}^{n\times n}$, we define

$$cov(A) \equiv A^K I(\varpi)(A^K)', \quad \rho_{ik}(A) \equiv \frac{cov_{ik}(A)}{\sqrt{cov_{ii}(A) cov_{kk}(A)}}, \tag{32}$$
$$\widetilde{F}_{ik}(A; L) \equiv \frac{l_{ik}\rho_{ik}(L \cdot A)}{\sum_{s=1}^n l_{is}\rho_{is}(L \cdot A)}, \qquad i, k = 1, ..., n,$$

where AB and $A^K = A...A$ are products of (compatible) matrices A and B, while $A \cdot B$ is the Hadamard product, i.e., $(A \cdot B)_{ik} = a_{ik} b_{ik}$. Note that $\widetilde{F}_{ik}(A; L)$ is identical to $F_{ik}(A; L)$, as defined by (7), when $A \leq L$.

Further, let $G_1 = \{1, ..., n_1\}$ and $G_2 = \{n_1 + 1, ..., n_1 + n_2\}$ be two groups with n_1 and n_2 nodes, $n_1 + n_2 = n$. For any $i = 1, ..., n$, let $G^i = G_q$, $n^i = n_q = \sum_{k \in G_q} l_{ik}$ and $\varpi^i = \varpi_q$ if $i \in G_q$ $(q = 1, 2)$. Let

$$A^0 \equiv \begin{pmatrix} Q(n_1) & 0_{n_1 \times n_2} \\ 0_{n_2 \times n_1} & Q(n_2) \end{pmatrix}, \quad L^0 \equiv sign(A^0), \quad I(\varpi) \equiv \begin{pmatrix} \varpi_1 I_{n_1 \times n_1} & 0_{n_1 \times n_2} \\ 0_{n_2 \times n_1} & \varpi_2 I_{n_1 \times n_2} \end{pmatrix},$$
$$where, \quad Q(n_r) \equiv (q_{ik}(n_r))_{i,k=1}^{n_r}, \quad q_{ik}(n_r) \equiv \frac{1}{n_r}, \quad \varpi \equiv (\varpi_1, ..., \varpi_1, \varpi_2, ..., \varpi_2).$$

LEMMA 3 *For any* $n \times n$ *matrix* $U \equiv \{u_{ij}\}_{i,j \in N}$ *and* $ik \in n \times n$,

$$\frac{d\rho_{ik}(A^0 + \omega U)}{d\omega}|_{\omega=0} = \begin{cases} \frac{n^i \varpi^k U_i^{G^k} + n^k \varpi^i U_k^{G^i} + (K-1)(\varpi^k U_{G^i}^{G^k} + \varpi^i U_{G^k}^{G^i})}{\sqrt{\varpi^i \varpi^k n^i n^k}}, & i, k : G^i \neq G^k, \\ 0, \quad i, k : G^i = G^k, \end{cases}, \quad (33)$$

where $U_M^{M'} \equiv \sum_{i \in M, j \in M'} u_{ij}$.

PROOF: We define the real matrix $A(\omega) \equiv A^0 + \omega U$ and the matrix-valued function

$$f^K(\omega) \equiv cov(A(\omega)) = A(\omega)^K I(\varpi)(A(\omega)^K)', \quad (34)$$

with a recursive structure:

$$\begin{aligned} f^K(\omega) &= A(\omega)^K I(\varpi)(A(\omega)^K)' = A(\omega)A(\omega)^{K-1} I(\varpi)(A(\omega)A(\omega)^{K-1})' \\ &= A(\omega)A(\omega)^{K-1} I(\varpi)(A(\omega)^{K-1})' A(\omega)' = A(\omega) f^{K-1}(\omega) A(\omega)', \\ f^0(\omega) &= I(\varpi). \end{aligned} \quad (35)$$

We use the product rule $(hg)' = h'g + hg'$ to compute the derivative of (35),

$$\begin{aligned} \frac{df^K(\omega)}{d\omega} &= U f^{K-1}(\omega) A(\omega)' + A(\omega)(\frac{df^{K-1}(\omega)}{d\omega} A(\omega)' + f^{K-1}(\omega) U') \\ &= A(\omega) \frac{df^{K-1}(\omega)}{d\omega} A(\omega)' + U f^{K-1}(\omega) A(\omega)' + A(\omega) f^{K-1}(\omega) U', \\ \frac{df^0(\omega)}{d\omega} &= 0. \end{aligned} \quad (36)$$

We solve (36) by successive substitution and evaluate $\frac{df^K(\omega)}{d\omega}$ at $\omega = 0$,

$$\frac{df^K(0)}{d\omega} = A^0 \frac{df^{K-1}(0)}{d\omega} A^0 + \Psi + \Psi' = ... = (K-1) A^0 (\Psi + \Psi') A^0 + \Psi + \Psi', \quad (37)$$

where $A^k(0) = A^0$ for any $k = 1, ..., K$, $\Psi \equiv Uf^K(0)$ and

$$f^K(0) = A^0 I(\varpi) A^0 \Rightarrow f^K_{ik}(0) \equiv \begin{cases} \varpi^i/n^i, & G^i = G^k, \\ 0, & G^i \neq G^k. \end{cases} \tag{38}$$

From (37) and (38), it can be verified directly that

$$\frac{df^K_{ik}(0)}{d\omega} = \frac{n^i \varpi^k U_i^{G^k} + n^k \varpi^i U_k^{G^i} + (K-1)(\varpi^k U_{G^i}^{G^k} + \varpi^i U_{G^k}^{G^i})}{n^i n^k}. \tag{39}$$

By applying the quotient rule $(h/g)' = (h'g - hg')/g^2$, we obtain the derivative:

$$\frac{d\rho_{ik}(A(\omega))}{d\omega}|_{\omega=0} = \frac{d(f^K_{ik}(\omega)/\sqrt{f^K_{ii}(\omega) f^K_{kk}(\omega)})}{d\omega}|_{\omega=0} \tag{40}$$

$$= \begin{cases} \sqrt{\frac{n^i n^k}{\varpi^i \varpi^k}} \frac{df^K_{ik}(0)}{d\omega}, & i,k : G^i \neq G^k, \\ \frac{n^i}{\varpi^i}(\frac{df^K_{ik}(0)}{d\omega} - \frac{1}{2}(\frac{df^K_{ii}(0)}{d\omega} + \frac{df^K_{kk}(0)}{d\omega})), & i,k : G^i = G^k, \end{cases}$$

where we used (38) to substitute for $f^K_{..}(0)$. The formula (33) obtains then by substituting $\frac{df^K_{..}(0)}{d\omega}$ from (39). In particular, $G^i = G^k$ implies $\varpi^i = \varpi^k$ and $n^i = n^k$ and all terms in (39) cancel out in this case. ⊠

Lemma 4 *For binary $n \times n$ matrices $V \equiv (v_{\tau\omega})_{\tau,\omega \in N}$ and $U^{st} \equiv (u^{st}_{\tau\omega})_{\tau,\omega \in N}$ such that $v_{\tau\omega} = 0$ if $G^\tau = G^\omega$ and $u^{st}_{st} = 1$, $u^{st}_{\tau\omega} = 0$ for $\tau\omega \neq st$,*

$$\frac{d}{d\omega} \widetilde{F}_{ik}(A^0 + \omega U^{st}; L^0 + V)|_{\omega=0} = \tag{41}$$

$$\begin{cases} \frac{n^i \varpi^k (\Upsilon^{st})_i^{G^k} + n^k \varpi^i (\Upsilon^{st})_k^{G^i} + (K-1)(\varpi^k (\Upsilon^{st})_{G^i}^{G^k} + \varpi^i (\Upsilon^{st})_{G^k}^{G^i})}{n^i \sqrt{\varpi^i \varpi^k n^i n^k}}, & if \quad v_{ik} = 1, \\ 0, \quad if \quad v_{st} = 0, \end{cases}$$

where $\Upsilon^{st} \equiv (L^0 + V) \cdot U^{st}$.

Proof: First note that the two cases in (41) are neither mutually exclusive nor collectively exhaustive, but they are the only relevant ones for the proof of Proposition 7. Then define $\Lambda \equiv (\lambda_{ik})_{i,k \in N} \equiv L^0 + V$, $A^{st}(\omega) \equiv \Lambda \cdot (A^0 + \omega U^{st}) = A^0 + \omega \Upsilon^{st}$ and the normalization factor

$$\eta_i(\omega, \Lambda) \equiv \textstyle\sum_{s=1}^n \lambda_{is} \rho_{is}(A^{st}(\omega)), \quad i = 1, .., n.$$

Then we observe that $\eta_i(0, \Lambda) = \sum_{s \in G^i} l^0_{is} = n^i$ due to the fact that $v_{is} = 0$ when $s \in G^i$ (i.e. $G^s = G^i$) and,

$$\rho(A^0) = L^0 = \begin{pmatrix} 1_{n_1 \times n_1} & 0_{n_1 \times n_2} \\ 0_{n_2 \times n_1} & 1_{n_2 \times n_2} \end{pmatrix}. \tag{42}$$

Applying the quotient rule, we calculate the following derivatives:

$$\frac{d\widetilde{F}_{ik}(A^0 + \omega U^{st}; \Lambda)}{d\omega}|_{\omega=0} = \frac{d}{d\omega} \frac{\lambda_{ik} \rho_{ik}(A^{st}(\omega))}{\eta_i(\omega, \Lambda)}|_{\omega=0} = \tag{43}$$

$$\frac{\lambda_{ik}}{(n^i)^2}(\frac{d\rho_{ik}(A^{st}(\omega))}{d\omega} n^i - \rho_{ik}(A^0) \frac{d\eta_i(\omega, \Lambda)}{d\omega})|_{\omega=0},$$

where

$$\frac{d\eta_i(\omega,\Lambda)}{d\omega} = \textstyle\sum_{s=1}^{n}\lambda_{is}\frac{d\rho_{is}(A^{st}(\omega))}{d\omega}. \quad (44)$$

For $v_{ik}=1$ the definition of V implies $G^i \neq G^k$ and, then, we have $\rho_{ik}(A^0)=l^0_{ik}=0$ and $\lambda_{ik}=v_{ik}$ by (42). Then, (43) takes the form:

$$\frac{d\widetilde{F}_{ik}(A^0+\omega U^{st};\Lambda)}{d\omega}|_{\omega=0} = \frac{v_{ik}}{n^i}(\frac{d\rho_{ik}(A^{st}(\omega))}{d\omega})|_{\omega=0}, \quad (45)$$

which after substitution from (33) specializes to the expression in (41). In order to prove (41) for $v_{st}=0$, we consider three mutually exclusive and collectively exhaustive cases.

1) $v_{st}=0$ and $G^s=G^t$ and $G^i\neq G^k$: This is a special case of the expression in (41) with $\Upsilon^{st}=U^{st}$ and $(\Upsilon^{st})_i^{G^k}=(\Upsilon^{st})_k^{G^i}=(\Upsilon^{st})_{G^i}^{G^k}=(\Upsilon^{st})_{G^k}^{G^i}=0$.

2) $v_{st}=0$ and $G^s=G^t$ and $G^i=G^k$: Then, $\frac{d\rho_{ik}(.)}{d\omega}|_{\omega=0}=0$ by (33) and $\rho_{ik}(A^0,\varpi,K)=1$ by (42). Hence, from (43), we obtain:

$$\frac{d\widetilde{F}_{ik}(A^0+\omega U^{st};\Lambda)}{d\omega}|_{\omega=0} = -\frac{\lambda_{ik}}{(n^i)^2}\frac{d\eta_i(\omega,\Lambda)}{d\omega})|_{\omega=0},$$

which, from (44), vanishes after substitution:

$$\frac{d\eta_i(\omega,\Lambda)}{d\omega} = \textstyle\sum_{s=1}^{n}\lambda_{is}\frac{d\rho_{is}(.)}{d\omega} = \textstyle\sum_{s:G^s=G^i}1\times 0+\textstyle\sum_{s:G^s\neq G^i}0\times\frac{d\rho_{is}(.)}{d\omega}=0.$$

3) $v_{st}=0$ and $G^s\neq G^t$: In this case, $\Upsilon^{st}=0$ and

$$\frac{d}{d\omega}\widetilde{F}_{ik}(A^0+\omega U^{st};\Lambda)|_{\omega=0} = \frac{d}{d\omega}\frac{\lambda_{ik}\rho_{ik}(\Lambda\cdot(A^0+\omega U^{st}))}{\sum_{s=1}^{n}\lambda_{is}\rho_{is}(\Lambda\cdot(A^0+\omega U^{st}))} =$$
$$\frac{d}{d\omega}\frac{\lambda_{ik}\rho_{ik}(A^0+\omega\Upsilon^{st})}{\sum_{s=1}^{n}\lambda_{is}\rho_{is}(A^0+\omega\Upsilon^{st})} = \frac{d}{d\omega}\widetilde{F}_{ik}(A^0+\omega\Upsilon^{st};\Lambda)|_{\omega=0}=0.$$

. ⊠

Now we turn to the proof of the proposition itself. In fact, we shall prove this result in a more general setting, where the initial opinions of all agents are uncorrelated and have a common variation ϖ_q within each group. Hence, $\phi^\beta_{ii}=\varpi_1$ for each $i\in G_1$, $\phi^\beta_{ii}=\varpi_2$ for each $i\in G_2$, and $\phi^\beta_{ij}=0$ when $i\neq j$.

For a vector valued function f, column vectors $\mathbf{x}$, $\mathbf{u}$ and a real number ω, the first order approximation of f at $\mathbf{x}$ is computed as,

$$f(\mathbf{x}+\omega\mathbf{u})\approx f(\mathbf{x})+\omega\frac{\partial f(\mathbf{x})}{\partial\mathbf{x}}\mathbf{u}\Rightarrow\frac{d}{d\omega}f(\mathbf{x}+\omega\mathbf{u})|_{\omega=0}=\frac{\partial f(\mathbf{x})}{\partial\mathbf{x}}\mathbf{u}, \quad (46)$$

where $\partial f(\mathbf{x})/\partial\mathbf{x}$ is the Jacobian of f at $\mathbf{x}$. In particular, for the vector $\mathbf{u}^t$ such that $\mathbf{u}^t_t=1$ and $\mathbf{u}^t_\nu=0$ for all $v\neq t$,

$$\frac{d}{d\omega}f_i(\mathbf{x}+\omega\mathbf{u}^t)|_{\omega=0}=(\frac{\partial f(\mathbf{x})}{\partial\mathbf{x}}\mathbf{u}^t)_i=\frac{\partial f_i(\mathbf{x})}{\partial x_t},$$

where $\frac{\partial f}{\partial \mathbf{x}}\mathbf{u}^t$ is the t^{th} column of the Jacobian $\partial f(\mathbf{x})/\partial \mathbf{x}$. By the same token, in our context we obtain for the $n \times n$ matrix $U^{st} \equiv (u^{st}_{\tau\omega})_{\tau,\omega \in N}$ such that $u^{st}_{st} = 1$ and $u^{st}_{ik} = 0$ for all $ik \neq st$,

$$\frac{d}{d\omega}\widetilde{F}_{ik}(A^0 + \omega U^{st}; L^0 + V)|_{\omega=0} = \frac{\partial \widetilde{F}_{ik}(A; L^0 + V)}{\partial a_{st}}|_{A=A^0} \equiv \mathcal{J}_{ik,st}, \tag{47}$$

where $ik \in n \times n$ indexes the row and $st \in n \times n$ indexes the column in the $n^2 \times n^2$ Jacobian matrix $\mathcal{J} \equiv \left(\frac{\partial \widetilde{F}_{ij}}{\partial a_{st}}\right)_{i,j,s,t \in N}$. From (47) and Lemma 4, we obtain the relevant entries in $\mathcal{J}$ and its transposed $\mathcal{J}'$ as illustrated in the tables below:

$$\mathcal{J}_{ik,st} = \left\{ \begin{array}{lll} & v_{st}=0 & v_{st}=1 \\ v_{ik}=0 & 0 & ? \\ v_{ik}=1 & 0 & \geq 0 \end{array} \right\} \Rightarrow \mathcal{J}'_{ik,st} = \left\{ \begin{array}{lll} & v_{st}=0 & v_{st}=1 \\ v_{ik}=0 & 0 & 0 \\ v_{ik}=1 & ? & \geq 0 \end{array} \right\} \tag{48}$$

Then, from the system of eigenvalue equations $\mathcal{J}'\mathbf{e} = \lambda \mathbf{e}$ for $\lambda \neq 0$ it follows that $e_{ik} = 0$ when $v_{ik} = 0$, where $\mathbf{e} = \{e_{ik}\}_{i,k \in N}$ is an $n^2 \times 1$ eigenvector of $\mathcal{J}'$ with entries indexed by $ik \in n \times n$. Hence, in light of (48), only elements of $\mathcal{J}'_{ik,st}$ with $v_{ik} = 1$ and $v_{st} = 1$ appear in the eigen equations for $\mathcal{J}'$. Therefore, for the computation of the eigenvalues and eigenvectors of $\mathcal{J}'$, we can think of all entries $\mathcal{J}'_{ik,st}$ as equal to zero except when $v_{ik} = 1$ and $v_{st} = 1$, in which case they are non-negative.

By the Perron-Frobenius Theorem, the largest eigenvalue of a nonnegative square matrix is real and positive and has an associated nonnegative eigenvector. Hence, for the system of eigen equations $\mathcal{J}'\mathbf{e} = \lambda \mathbf{e}$, we have that $\mathbf{e} > 0$ if $\lambda > 0$ is the Perron-Frobenius eigenvalue of $\mathcal{J}'$ (and, hence, of $\mathcal{J}$). Then, we can write the sum of the eigen equations as follows:

$$\textstyle\sum_{ik:v_{ik}=1} e_{ik} \sum_{st:v_{st}=1} \mathcal{J}'_{st,ik} = \lambda \sum_{ik:v_{ik}=1} e_{ik} > 0. \tag{49}$$

Dividing (49) by $\sum_{ik:v_{ik}=1} e_{ik}$ shows that λ is a convex combination of the values in the set

$$\{\textstyle\sum_{st:v_{st}=1} \mathcal{J}'_{st,ik}\}_{ik:v_{ik}=1} = \{\sum_{st:v_{st}=1} \mathcal{J}_{ik,st}\}_{ik:v_{ik}=1}.$$

For each ik such that $v_{ik} = 1$, we compute $\sum_{st:v_{st}=1} \mathcal{J}_{ik,st}$ by substituting for $\mathcal{J}_{ik,st}$ from (41) with $\Upsilon^{st} \equiv (L^0 + V) \cdot U^{st} = U^{st}$ as $v_{st} = 1$ (and, hence, $G^s \neq G^t$):

$$\sum_{st:v_{st}=1} \frac{n^i \varpi^k (U^{st})^{G^k}_i + n^k \varpi^i (U^{st})^{G^i}_k + (K-1)(\varpi^k (U^{st})^{G^k}_{G^i} + \varpi^i (U^{st})^{G^i}_{G^k})}{n^i \sqrt{\varpi^i \varpi^k n^i n^k}}$$
$$= \frac{n^i \varpi^k V^{G^k}_i + n^k \varpi^i V^{G^i}_k + (K-1)(\varpi^k V^{G^k}_{G^i} + \varpi^i V^{G^i}_{G^k})}{n^i \sqrt{\varpi^i \varpi^k n^i n^k}} \equiv z_{ik}.$$

We have shown, therefore, that the largest eigenvalue λ of the Jacobian matrix $\mathcal{J}$ is a convex combination of the values in the set $\{z_{ik}\}_{ik:v_{ik}=1}$. If all these values are smaller (greater) than one, then λ is smaller (greater) than one and the corresponding dynamic system is stable (unstable). The condition (17) follows then for the special case $\varpi_1 = \varpi_2$. ⊠

Proof of Corollary 2: It is enough to show that, for constant K, large n and under assumptions (M.1)-(M.2), condition (17) is satisfied for all cross-group links ij with $i \in G_q$ and $j \in G_{q'}$, $q' \neq q$. We first note that, for each $n \in \mathbb{N}$, the LHS of that condition can be bounded *above* for every such link ij as follows:

$$\begin{aligned} n^i\, v^i + n^j\, v^j + (K-1)v &\leq \overline{n}_q n\, b(n) + (1-\overline{n}_q) n\, b(n) + (K-1) n\, b(n) + n\,\mathcal{O}(n) \\ &= \alpha_0\, n\, [b(n) + \mathcal{O}(n)], \end{aligned}$$

for some $\alpha_0 > 0$ where $\mathcal{O}(n)$ is an infinitesimal in n. Combining this conclusion with the fact that, for large enough n, the RHS of condition (17) can be bounded *below* by a function of order $\mathcal{O}(n^2)$, i.e.

$$n^i\, \sqrt{n_1 n_2} \geq \alpha_1 n^2,$$

leads to the desired conclusion. ☒

Proof of Proposition 8: For $\mu = 1$, (19) implies that, given any $K \geq 1$ and any stochastic and aperiodic[20] matrix A, the steady state end-beliefs are equal to

$$\hat{x} = A^K \hat{x} \Rightarrow \hat{x} = (\lim_{K\to\infty} A^K)\beta.$$

The same steady state end-beliefs follow from (22) when $\mu \in [0, 1)$ and $K \to \infty$,

$$\hat{x} = (1-\mu)(I - \mu B)^{-1} B\beta = (1-\mu)(\textstyle\sum_{s=0}^{+\infty} \mu^s B) B\beta = B\beta,$$

where $B = \lim_{K\to\infty} A^K$. It is well known that for a connected (i.e., irreducible), aperiodic and stochastic matrix A each row of B is equal to the left eigenvector of A associated to the eigenvalue 1. Then, all correlations $\rho_{ik}(A)$, computed by (6) for $i, k \in N$, are equal to one. Substituting unit correlations into the definition (8) of EIM yields the claim, $a^*_{ik} = l_{ik}/\sum_{s=1}^{N} l_{is}$ for all $i, k \in N$. ☒

References

[1] Adamic, L. and N. Glance (2005): "The Political Blogosphere and the 2004 U.S. Election: Divided They Blog," *LinkKDD'05, Proceedings of the 3rd International Workshop on Link Discovery*, 36-43.

[2] Adomavicius, G. and A. Tuzhilin (2005): "Towards the next generation of recommender systems: A survey of the state-of-the-art and possible extensions," *IEEE Transactions on Knowledge and Data Engineering* **17**, 734-749.

[3] Alexander C, Piazza M, Mekos D, Valente T. (2001): "Peers, schools, and adolescent cigarette smoking," *Journal of Adolescent Health* **29**(1):22-30.

[20] Any EIM is aperiodic since all its main diagonal entries must be positive.

[4] Bala, V. and S. Goyal (1998): "Learning from Neighbors," *Review of Economic Studies* **65**, 595-621.

[5] Bishop, B. (2008): *The Big Sort: Why the Clustering of Like-Minded America is Tearing Us Apart*, New York: Mariner Books.

[6] Bolletta, U. and P. Pin (2022): "Polarization when people choose their peers", *Games & Political Behavior eJournal*.

[7] Boutyline A. and R. Willer (2017): "The Social Structure of Political Echo Chambers: Ideology and Political Homophily in Online Communication Networks", *Political Psychology* **38**, 551-569.

[8] Boxell, L., M. Gentzkow, and J. M. Shapiro (2017): "Is the internet causing political polarization? Evidence from demographics", Working Paper, No. 23258. National Bureau of Economic Research.

[9] Cerreia-Vioglio, S., R. Corrao, and G. Lanzani (2021): "Dynamic Opinion Aggregation: Long-run Stability and Disagreement," Bocconi University and MIT.

[10] Crandall, D., D. Cosley, D. Huttenlocher, J. Kleinberg, and S. Suri (2008): "Feedback effects between similarity and social influence in online communities," Proceedings of the 14th ACM SIGKDD international conference on Knowledge discovery and data mining, August 2008, 160-168.

[11] Cohen JM. (1977): "Sources of peer group homogeneity", *Sociology of Education* **50**(4):227-241.

[12] DeGroot, M. H. (1974): "Reaching a Consensus", *Journal of the American Statistical Association* **69**, 118-121.

[13] DeMarzo, P.M., D. Vayanos and J. Zwiebel (2003): "Persuasion Bias, Social Influence, and Unidimensional Opinions", *The Quarterly Journal of Economics*, **118**, 909-968.

[14] Easley, D. and J. Kleinberg (2010): "Networks, Crowds, and Markets: Reasoning About a Highly Connected World", Cambridge: Cambridge University Press.

[15] Flache, A., M. Mäs, T. Feliciani, E. Chattoe-Brown, G. Deffuant, S. Huet, and J. Lorenz (2017): "Models of Social Influence: Towards the Next Frontiers," *Journal of Artificial Societies and Social Simulation* **20**.

[16] Friedkin, N. E. and E.C. Johnsen (1999): "Social influence networks and opinion change," *Advances in Group Processes* **16**, pp. 1-29.

[17] Gale, D. and S. Kariv (2003): "Bayesian learning in social networks," *Games and Economic Behavior* **45**, 329-346.

[18] Gentzkow, M. and J. M. Shapiro (2011): "Ideological segregation online and offline", *The Quarterly Journal of Economics* **126**, 1799-1839.

[19] Golman, R., G. Loewenstein, K. O. Moene, and L. Zarri (2016): "The Preference for Belief Consonance," *Journal of Economic Perspectives* **30**, 165-188.

[20] Golub, B. and M. O. Jackson (2010): "Naïve Learning in Social Networks: Convergence, Influence, and the Wisdom of Crowds", *American Economic Journal: Microeconomics* **2**, 112-149.

[21] Golub, B. and M. O. Jackson (2012): "How Homophily Affects the Speed of Learning and Best Response Dynamics," *Quarterly Journal of Economics* **127**, 1287-1338.

[22] Hegselmann, R. and U. Krause (2002): "Opinion dynamics and bounded confidence models, analysis and simulations," *Journal of Artifical Societies and Social Simulation* **5**.

[23] Holme, P. and M. E. J Newman (2006): "Nonequilibrium phase transition in the coevolution of networks and opinions", *Physical Review E* **74**, 056108

[24] Jackson, M.O., T. Rodríguez-Barraquer and X. Tan (2012): "Social Capital and Social Quilts: Network Patterns of Favor Exchange", *American Economic Review* **102**, 1857-1897.

[25] Jackson, M.O. and E. C. Storms (2021): "Behavioral Communities and the Atomic Structure of Networks," Stanford University, *arXiv 1710.04656*.

[26] Jannach, D., M. Zanker, A. Felfernig, and G. Friedrich (2010): *Recommender Systems: An Introduction*, Cambridge: Cambridge University Press.

[27] Kandori, M., G. Mailath, and R. Rob (1993): "Learning, Mutation, and Long-run Equilibria in Games", *Econometrica* **61**, 29-56.

[28] Kandel DB (1978): "Homophily, selection, and socialization in adolescent friendships", *American Journal of Sociology* **84**(2):427-436.

[29] Knecht A, Snijders TAB, Baerveldt C, Steglich CEG, Raub W. (2010), "Friendship and Delinquency: Selection and Influence Processes in Early Adolescence", *Social Development* **19**(3):494-514.

[30] Kossinets, G., Watts, D.J., (2006): "Empirical Analysis of an Evolving Social Network", *Science* **311** (5757): 88-90.

[31] Kossinets, G., Watts, D.J., (2009): "Origins of Homophily in an Evolving Social Network," *American Journal of Sociology* **115**(2), 405-450

[32] Kumpula, J.M., Onnela, J.P., Saramaki, J., Kaski, K., Kertesz, J. (2007): "Emergence of communities in weighted networks", *Physical Review Letters* **99**, 228701.

[33] Lazarsfeld, P.F. and R.K. Merton (1954): "Friendship as a social process: a substanive and methodological analysis", in M. Berger (ed.), Freedom and Control in Modern Society, New York: Van Nostrand.

[34] Levy, G. and R. Razin (2018): "Information diffusion in networks with the Bayesian Peer Influence heuristic," *Games and Economic Behavior* **109**, 262-270.

[35] Levy, G. and R. Razin (2020): "Social media and political polarisation," *LSE Public Policy Review* **5**, 1-7.

[36] Lobel I. and E. Sadler (2013): "Preferences, Homophily, and Social Learning", mimeo.

[37] Marsden, P.V. (1987): "Core Discussion Networks of Americans", *American Sociological Review* **52**, 122-313.

[38] Marsden, P.V. (1988): "Homogeneity in Confiding Relations", *Social Networks* **10**, 57-76.

[39] McPherson, M., L. Smith-Lovin and J. M. Cook (2001): "Birds of a Feather: Homophily in Social Networks", *Annual Review Sociology* **27**, 415-44.

[40] McVicar, D. and A. Polanski (2014): "Peer effects in UK adolescent substance use: never mind the classmates?", *Oxford Bulletin of Economics and Statistics* **76**(4), 589-604.

[41] Melguizo, I. (2019): "Homophily and the persistence of disagreement", *The Economic Journal* **129**, 1400-1424,

[42] Molavi, P., A. Tahbaz-Salehi, and A. Jadbabaie (2018): "A theory of non-Bayesian social learning," *Econometrica* **86**, 445-90.

[43] Moody, J. (2001): "Race, School Integration, and Friendship Segregation in America", *The American Journal of Sociology* **107**(3), 679-716.

[44] Mueller-Frank, M. (2014): "Does one Bayesian make a difference?," *Journal of Economic Theory* **154**, 423-452.

[45] Pajevic, S. and D. Plenz (2012): "The organization of strong links in complex networks", *Nature Physics* **8**(5), 429-436.

[46] Pan, Z. (2010), "Trust, influence, and convergence of behavior in social networks," *Mathematical Social Sciences*, **60**, 69-78.

[47] Ricci, F., L. Rokach, and B. Shapira (2011): "Introduction to Recommender Systems Handbook", *Recommender Systems Handbook*, Springer, pp. 1-35.

[48] Su X. and T.M. Khoshgoftaar (2009): "A survey of collaborative filtering techniques," *Advances in Artificial Intelligence* article id. 421425, 19 pages.

[49] Young, P. (1993): "The Evolution of Conventions", *Econometrica* **61**, 29-56.

"Coevolution of dynamical states and interactions in dynamic networks", *Physical Review E* **69**, 065102.