Social Networks and Collective Action in Large Populations: An application to the Egyptian Arab Spring*

July 2024

Lachlan Deer[a], Chih-Sheng Hsieh[b], Michael D. König[c,d,e], Fernando Vega-Redondo[f,g]

[a] *Tilburg University, Tilburg, The Netherlands.*
[b] *Department of Economics, National Taiwan University, Taipei, Taiwan.*
[c] *Tinbergen Institute and Department of Spatial Economics, VU Amsterdam, Amsterdam, The Netherlands.*
[d] *ETH Zurich, Swiss Economic Institute (KOF), Zurich, Switzerland.*
[e] *Centre for Economic Policy Research (CEPR), London, United Kingdom.*
[f] *Department of Decision Sciences, Bocconi University, Milan, Italy.*
[g] *Department of Economics, Universidad Carlos III, Madrid, Spain.*

Abstract

We study a dynamic model of collective action in which agents are connected by a social network. Our approach highlights the importance of communication in this problem and conceives that network – which is continuously evolving – as providing the channel through which agents not only interact but also communicate. We consider two alternative scenarios that differ only on how agents form their expectations: while in the "benchmark" context agents are completely informed, in the alternative one their expectations are formed through a combination of local observation and social learning à la DeGroot. We completely characterize the long-run behavior of the system in both cases and show that only in the latter scenario (arguably the most realistic) there is a significant long-run probability that agents eventually achieve collective action within a meaningful time scale. We suggest that this sheds light on the puzzle of how large populations can coordinate on globally desired outcomes. Finally, we illustrate the empirical potential of the model by showing that it can be efficiently estimated for the Egyptian Arab Spring using large-scale cross-sectional data from Twitter. This estimation exercise also suggests that, in this instance, network-based social learning played a leading role in the process underlying collective action.

Keywords: collective action, networks, coordination, social protests, DeGroot, social learning
JEL: D74, D72, D71, D83, C72

---

*We would like to thank Ozan Candogan, Ruben Enikolopov, James Fenske, Sanjeev Goyal, Bryan Graham, Angelo Mele, Kei Ikegami, Michihiro Kandori, Dominic Rohner, Mathias Thoenig, Yves Zenou, Fabrizio Zilibotti and seminar participants at the HEC Lausanne, Tinbergen Institute, Bocconi, Academia Sinica, Cambridge, Chicago, Columbia, Princeton and Northwestern University for the helpful comments.

Email addresses: `lachlan.deer@gmail.com` (Lachlan Deer), `cshsieh@ntu.edu.tw` (Chih-Sheng Hsieh), `m.d.konig@vu.nl` (Michael D. König), `fernando.vega@unibocconi.it` (Fernando Vega-Redondo)

1. Introduction

Much of the interaction taking place in socio-economic environments is mediated by the social networks that connect their agents for a wide range of different purposes. Their connections not only underlie some of the payoffs these agents obtain but also serve them to communicate when facing problems of collective action such as addressing social emergencies, staging peaceful demonstrations, or even igniting violent riots. In modern societies, a large fraction of that social interaction is carried out virtually through online social media such as Facebook, Twitter/X, VK, or Weibo (cf. Acemoglu et al., 2018; Enikolopov et al., 2020; Qin et al., 2021). This renders it feasible for researchers to collect massive data on the operation of large-scale social networks, in turn opening up rich possibilities for an *integrated* theoretical and empirical study of collective-action problems in large populations. Developing theoretical and empirical methods that can be used to handle and better understand this type of data is, in a nutshell, the main objective of this paper.

The existing literature, both theoretical and empirical (see, e.g., Chwe (2000) and Cardoso et al. (2019)), has highlighted a key consideration underlying whether collective action succeeds or fails within different environments is how information, communication, and coordination interplay in each of them. To shed light on this interplay, we shall argue that it is useful to account explicitly for the ways in which individuals gather information and communicate with others. And when the population is large, the assumption that such information and communication are locally constrained by the social network is not only more realistic than the alternative common assumption of complete (or global) information, it also leads, somewhat paradoxically, to a more plausible model of how sizable populations tackle the severe coordination challenge posed by large-scale collective action.[1]

Our approach to the question combines, and integrates, a theoretical and an empirical analysis of the problem. On the theoretical front, we model the situation dynamically as a *population game* played on an *evolving network* in which each agent adjusts over time both her action (say, whether to join or not collective action) and her connections/links in the social network. Those adjustments are taken to depend on what payoff-relevant features of the state of the system the agent observes and how she forms beliefs about those features she does *not* observe. More specifically, the assumptions we make in our model are:

(a) the agents accurately *know* the *actions* chosen by their neighbors in the social network;
(b) they hold *beliefs*, possibly inaccurate, on the *average action* chosen by the population.

In our model, (a) and (b) encompass all the considerations that are payoff relevant in the two different scenarios we consider. For both of them, point (a) applies, i.e., the precise local observation holds, as mediated by the currently prevailing network. However, the two scenarios significantly differ in terms of how accurately agents form their beliefs about the "global" characteristics of the prevailing situation – most importantly, about the overall support given to collective action. The following two alternative assumptions capture, in a stylized manner, the contrast between them:

- The first assumption posits that every agent has perfect information (i.e., accurate beliefs) on the average action of all others; we describe the context as one of *global (or complete) information* (GI).
- The second assumption supposes that, given the prevailing social network, each agent forms individual (possibly inaccurate) *beliefs* by combining the information she observes locally (her neighbors' actions) and the beliefs she is locally exposed to (her neighbors' beliefs); this context is therefore described as one with *local (thus limited) information and learning* (LIL).

[1] A reflection of this challenge is what Shadmehr (2021) labels Tullock's Paradox of Revolution (cf. Tullock (1971)).

There is a large literature that, starting from the seminal papers by Kandori et al. (1993) and Young (1993), has modeled equilibrium selection in coordination games as the outcome of a noisy evolutionary process of learning under complete/global information.[2] A conceptual problem with this approach is that, if the population is large, the expected rate of convergence to the long-run theoretical prediction is generally very slow, which makes the expected time required to reach it very long. This renders it difficult to interpret the induced results as reasonable predictions for a time scale that is relevant to many real-world phenomena. A way that has been proposed to tackle this issue is to suppose that agents' interactions – and therefore the payoffs following from it – are local-based (see Ellison (1993) for an early proponent of this idea). For our purposes, however, the problem with this formulation is that it defines an interaction setup that is strictly local, hence failing to capture the fact that in many social contexts (in particular, in some important problems of collective action), the coordination required also embodies a substantial population-wide (i.e., global) component.

Motivated by the conceptual limitations faced by strictly local or global frameworks, in this paper we introduce a theoretical framework that blends local and global interaction and allows us to compare the implications of the alternative learning/belief-formation scenarios induced by the GI and LIL assumptions. First, we find that the GI scenario inherits the qualitative properties commonly displayed in the received evolutionary literature on coordination games under global interaction but also its conceptual problems – that is, the GI scenario delivers a sharp equilibrium-selection prediction, but the convergence to it is exceedingly slow. In contrast, the alternative learning framework studied in the LIL scenario exhibits, *jointly*, the desirable properties separately induced by the global- and local-interaction models of social coordination. More specifically, we show that such a scenario enjoys the following two features: (a) *local* interaction and (also locally based) social learning play an important role in speeding up the convergence process; (b) *global* interaction is the key force underlying agents' payoffs and therefore also their incentives to join or shun collective action.

From a methodological viewpoint, the first important step in our analysis of the model is to obtain a full characterization of the long-run behavior of the evolutionary process, as captured by the induced *limiting probability distribution* over possible states. This characterization, undertaken for both the GI and the LIL scenarios, has two important implications – one for the development of the theory, and another for its empirical application.

On the theoretical side, it allows us to conduct a detailed comparative analysis of the different forces at work in the evolutionary process, as well as an exhaustive exploration of the effects induced on it by the different model parameters. This helps us shed some light on the central question of how our different assumptions on the interplay of information, communication, and coordination impinge on the plausibility of collective action in large populations. The main insight we gather from the exercise is that the arguably more "realistic" LIL scenario provides much stronger support for the attainment of collective action in the following two respects. First, it expands substantially the range of conditions (i.e., primitives of the model) where long-run behavior is consistent with collective action. Second, it induces a much faster convergence to such a state of affairs. In addition, our (closed-form) determination of the long-run behavior of the process permits us to answer the following types of questions. What is the comparative importance of observation and learning in shaping agents' beliefs? What is the effect of individual heterogeneity – and, possibly, homophily – in forming the social network? Does easier/cheaper connectivity – and hence a more dense

[2]See also the monographs by Samuelson (1998), Vega-Redondo (1996), and Young (1998) that provide early overviews of this literature.

network – help coordination on collective action? What is the impact of behavioral conformity (i.e., the agents' desire to coordinate actions with others), either locally or globally, have on the evolution of the process? Overall, our analysis shows that all of the above phenomena – observation and learning, heterogeneity and homophily, network density, and behavioral conformity – have, in interplay, a significant effect on whether collective action eventually materializes or not.

On the empirical side, the fact that our theory provides a characterization of the limiting distribution of the process in closed form (i.e., as an explicit function of the parameters) allows us to use that same distribution as a likelihood function for the structural estimation of the model, using data obtained from the massive social protests taking place in Egypt during the so-called Arab Spring. Specifically, we rely on Twitter data of tweets and connections of 225,578 users during the military backlash that, in June 2013, toppled the Morsi government. This information is used for two purposes. First, we construct the underlying social network, identifying inter-agent links with the bilateral relationships that show reciprocal influence. Second, we rely on machine-learning techniques used in the field of Natural Language Processing (NLP) to infer the agents' characteristics (e.g., their gender, political bias, or religious affiliation) as well as ongoing beliefs and behavior (in particular, their support of, or opposition to, the social movement protesting Mosi's removal by the Egyptian military.).

Because of the large size of our data set, conventional maximum-likelihood methods are not feasible, nor are the simulation-based Bayesian approaches proposed by the recent literature (Badev, 2021; Hsieh et al., 2020; Hsieh et al., 2022; Mele, 2017). Although we also utilize the Bayesian approach in this paper, we opt for the composite likelihood approach approach (Lindsay, 1988; Varin et al., 2011) and implement a case-control procedure on a sparse matrix to alleviate the computational burden. The estimates obtained are significant and have the expected signs, in line with the theory. They also provide an intuitive understanding of the context being studied. For example, we find that both social learning and local observation are significant components of belief formation (the former having a higher weight than the latter). Such endogenously generated expectations are one of the important forces driving behavior. The other force derives from the inherent costs or risks, as well as benefits and ambitions, that agents anticipate when deciding whether or not to contribute to collective action. In our model, their *net* effect is captured by a single (fixed) parameter. As part of our econometric exercise, this parameter is estimated from our data and it turns out to have a positive sign. This suggests that, in the Egyptian revolt against the military, the population perceived, on average, that the intrinsic costs and risks of joining the protests were more than offset by the corresponding benefits.

Our estimated model allows us to analyze counterfactual scenarios where we can study the impact of changes in the fundamental parameters of the model on rioting behavior. As an illustration, we focus on the following two cases: (i) we examine the role of linking costs in affecting rioting behavior; (ii) we analyze how biasing the beliefs towards a specific action can influence the rioting outcome. In case (i), we observe that a reduction of the linking cost by 20% yields an increase in the fraction of rioting agents by 15%. This shows that as linking and exchanging information via the network becomes more costly (e.g., by interrupting or blocking social media), fewer links are being formed, coordination among agents becomes more difficult, and fewer agents participate in the protest as a consequence. This finding illustrates and quantifies the importance of online social networks in the emergence of instances of collective action such as protest movements or riots. In case (ii), we contemplate the possibility of a government influencing the belief updating equation of the agents, biasing it towards the preferred action of the government (status quo). Our results show that while such a belief manipulation (say, "propaganda") does not affect the network density it has a drastic effect on rioting, reducing the fraction of rioting agents by up to 30%. These findings highlight that the manipulation of information can mitigate the formation of collective action but cannot suppress it entirely.

We conclude this introduction with a brief discussion of the relationship between our research and existing literature. On the theoretical front, we may highlight three different strands of work. One is the extensive research that has been conducted on coordination games in networks. For fixed networks, the problem has been studied (besides the already mentioned paper of Ellison (1993)) by, for example, Blume (1993), Brock and Durlauf (2001) or Morris (2000), while the analysis was extended to co-evolving endogenous networks by Jackson and Watts (2002) and Goyal and Vega-Redondo (2005).

A second branch of related literature includes the booming recent research on learning in networks (cf. Golub and Sadler (2016) for a survey). As a small sample, we may refer to the influential contributions by Acemoglu et al. (2014), DeMarzo et al. (2003), and Jackson and Golub (2010). While the first paper adopts a Bayesian approach to the problem, the latter two build upon the bounded-rationality framework proposed by DeGroot (1974). The latter approach – which posits that agents update their beliefs by linearly combining their own with those of their network neighbors – has received some experimental support (c.f. Chandrasekhar et al. (2015)) and heavily inspires the learning dynamics contemplated by our model.

Third, our work connects to models of collective action that, in line with the seminal work of Granovetter (1978), frame it as a threshold coordination problem. In this class of models, agents assess the cost and benefits of joining the collective action in terms of the fraction of other agents who may also join. Many papers in this literature stress, as we do, the crucial role played by information – in particular, they focus on the question of how agents gather such information, on the basis of what each of them can individually observe (see e.g. the aforementioned paper by Chwe (2000) or those of Barberà and Jackson (2020) and Lohmann (2000)). Further, Bueno de Mesquita (2010) and Shadmehr (2021) show that in this class of models "revolutionary entrepreneurs" may act as informational vanguards that lead the rest of the population. A variation of such a triggering vanguard-led mechanism for collective action also arises in our framework when we model the problem as one where agents learn from the limited information provided by their connections.

The empirical component of our paper builds upon the recent body of work that has developed econometric methods designed to study the co-determination of networks and actions in social contexts, addressing the difficult identification/endogeneity issues entailed (see, for example, the recent papers by Goldsmith et al (2013), Hsieh et al. (2016), and Johnsson and Moon (2021)). A few papers, such as those by Boucher (2016) and the aforementioned (Badev, 2021; Hsieh et al., 2020; Hsieh et al., 2022; Mele, 2017), apply these methods to carry out, as in our case, structural estimation of an underlying theoretical model. Their econometric methods, however, are computationally unfeasible in dealing with large social networks such as those common on social media platforms.

Our work also relates to the recent literature studying the role of social media in facilitating collective action – in particular, in supporting massive events of social protest. González-Bailón et al. (2011) study the role that Twitter had in the surge of the anti-austerity mobilizations that took place in Spain in May 2011. They show that the induced online network played an important role in the recruitment process by means of local "contagion." Acemoglu et al. (2018) focus on the same instance of social protests as we do – the Egyptian Arab Spring – and find support for the conclusion that a rise in Twitter activity preceded the triggering of social protests. Steinert-Threlkeld et al. (2015) and Steinert-Threlkeld (2017) further document the association between Twitter posts and protests throughout the wider Arab Spring, emphasizing the role of individuals on the periphery of the network in influencing the levels of protest.

In a similar vein but with a different methodological perspective, Enikolopov et al. (2020) study the wave of social protests that took place in Russia in 2011. Their main contribution is to identify a causal positive relationship between differences in the degree of social media penetration and the extent of social protest. Interestingly, they also show that the main basis for this effect is *not* the wider access to information the social media provide; instead, they highlight "the importance of horizontal information exchange on people's ability to overcome the collective action problem." This is in line with our own analysis, where the social network acts as a channel through which neighbors obtain information. More specifically, our approach is grounded on the idea that agents gather that information not only by observing their neighbors' actions but also by learning/exchanging their beliefs. The key role played by the network in shaping people's beliefs is also highlighted by the experimental evidence studied by Cantoni et al. (2019). They run a field experiment during the Hong Kong protests of 2016, manipulating citizens' beliefs of others' protest participation in an effort to identify how these beliefs affect their own participation in the protests. Their results suggest that individual's decision to protest are strategic substitutes. In our case, we base our analysis on Twitter (non-experimental) data and a structural model, and our approach to identifying agents' attitudes and beliefs is based on fine-tuned pretrained language models from the field of Natural Language Processing (NLP). We find that protest behavior displays strategic complementarities, locally among network neighbors and also globally.

The rest of the paper is organized as follows. Section 2 presents the theoretical framework (the game-theoretic setup, the law of motion for actions and links, and the different belief formation scenarios). Section 3 undertakes the formal analysis of the model, characterizing the long-run behavior of the system for each of the two scenarios considered (GI and LIL) and comparing their implications. Section 4 summarizes the data used in our application of the model to the Egyptian Arab Spring, while in Section 5 we discuss the estimation method and discuss the estimation results. Section 6 carries out counterfactual analysis that illustrates the role of network connections and information manipulation on the extent of collective action. Section 7 concludes the main body of the paper, while in Appendix A we include the formal proofs of all our theoretical results. We also include various supplementary appendices: Supplementary Appendix B discusses some extensions of the model; Supplementary Appendix C provides the details of the equilibrium characterization for vanishing noise in the LIL scenario; in Supplementary Appendix D we derive predictions of the theoretical model for finite noise and complement them with numerical simulations; in Online Appendix E we provide a complete characterization of the stochastically stable states for finite populations; in Supplementary Appendix F we describe the historical conditions that help contextualize our data; Supplementary Appendix G provides additional details for the data construction; Supplementary Appendix H summarizes the specifications of prior distributions and the implementation of the Bayesian estimation procedure; Supplementary Appendix I describes a Monte Carlo simulation study that examines the performance of the Bayesian estimation algorithm (based on the composite likelihood function and the case-control approach); and Supplementary Appendix J performs a robustness check of our estimation results.

2. The Model

We divide the presentation of the model into three parts. First, in Subsection 2.1 we introduce the basic interaction setup, i.e., we describe the primitives that define the interaction, induce the payoffs, and characterize the state of the system. Then, in Subsection 2.2 we specify the dynamics, i.e., the law of motion that determines how actions and links change over time, as a function of the prevailing beliefs. Finally, Subsection 2.3 introduces alternative formulations on how such beliefs are formed, depending on the information agents have access to.

2.1. Basic Setup

Consider a population $\mathcal{N} = \{1, \ldots, n\}$, conceived as large, which is involved in a problem of *collection action.* For concreteness, we interpret it to represent some instance of social protest and call it a "riot." Each individual $i \in \mathcal{N}$ must choose an action s_i, which is a dichotomous decision of whether to join the riot or not. Formally,[3] it is convenient to identify joining the riot with $+1$ and not doing so with -1. Thus, an action profile for the whole population is given by a vector $\mathbf{s} = (s_1, \ldots, s_n)^\top \in \mathbf{S} = \{-1, +1\}^n$ whose cardinality is given by $\#(\{-1, +1\}^n) = 2^n$.

The population is also connected through bilateral links as given by the current *social network* G. Any such network can be represented by its adjacency (binary) matrix $\mathbf{A} = (a_{ij})_{i,j=1}^n$, where each entry a_{ij} $(i, j = 1, 2, \ldots, n)$ is either 1 or 0 if i is either connected to j or not (with the convention that $a_{ii} = 0$). For simplicity, we shall consider undirected networks, which means that the matrix $\mathbf{A}$ is symmetric, i.e., $a_{ij} = a_{ji}$ for all $i, j \in \mathcal{N}$. The set of all undirected networks of size n is denoted by $\mathcal{G}^n$.[4]

Given any action profile $\mathbf{s}$ and an adjacency matrix $\mathbf{A}$, we expand on the classical interaction models studied by Brock and Durlauf (2001) and Blume et al. (2011) and posit that each agent $i \in \mathcal{N}$ holds some point belief $\psi_i \in [-1, +1]$ on the average action chosen by every other agent in the population. Then, assuming that agent i observes *perfectly* the action s_j chosen by each of her (immediate) network neighbors j (i.e. those j with $a_{ij} = 1$), her expected payoff is defined as follows:

$$\pi_i(\mathbf{s}, G; \psi_i) = \underbrace{\sum_{j=1, j\neq i}^{n} (\rho s_i \psi_i + \theta a_{ij} s_i s_j)}_{\text{interaction effect}} + \underbrace{\gamma_i s_i}_{\substack{\text{idiosyn.}\\ \text{bias}}} - \underbrace{\kappa s_i}_{\substack{\text{action}\\ \text{cost}}} - \underbrace{\sum_{j=1, j\neq i}^{n} a_{ij} \zeta_{ij}}_{\text{linking costs}}, \tag{1}$$

where

- $\rho \in (0, 1)$ is the parameter modulating a force towards *global conformity* with her *expectation* ψ_i of the average (population-wide) action;
- $\theta \in (0, 1)$ is a parameter capturing a force towards *local conformity* with the *accurately perceived* actions s_j of her (direct) network neighbors (i.e. those j s.t. $a_{ij} = 1$);
- $\gamma_i \in \{-1, +1\}$ is i's *idiosyncratic characteristic* shaping her bias for either action;
- $\kappa \geq 0$ is a *common cost* for choosing action $s_i = +1$ (e.g., the effort/risk of rioting);
- $\zeta_{ij} \geq 0$ is the *linking cost* between agents i and j.

We propose the following interpretation of the payoff structure (1). Its first term, the *interaction effect*, captures the genuinely strategic part of the model and has two components: a global and a local one. The global component, $\sum_{j\neq i}^n \rho s_i \psi_i$, whose relative weight is parametrized by ρ, embodies the essential coordination problem involved in collective action as it is perceived by agent i, given her beliefs ψ_i on the average action in the population. More concretely, one can think of such average action as shaping the probability that collective action is successful, with the expected benefit (or penalty) from being (mis)aligned with the more likely outcome depending on how concentrated those beliefs are (i.e., how close they are to either $+1$ or -1).[5]

[3] This is also the customary convention adopted in the analysis of the classical Ising model (cf. Grimmett, 2010).
[4] To understand how the analysis could be adapted if links are taken to be directed, see Supplementary Appendix B.1. However, here we consider undirected links, as it is standard in the social networks literature on peer effects, and we leave the detailed analysis of directed networks to future work.
[5] To fix ideas, a simple formalization could be as follows. Suppose that the probability of the collective effort on

The second local component of the interaction effect, $\sum_{j\neq i}^{n} a_{ij}s_is_j$, whose weight is parameterized by θ, reflects the classical assumption commonly made in the literature on games on networks (cf. Jackson and Zenou, 2015): agents like to have their behavior well aligned with that of their network partners/friends. As we shall see, the main role played by this local conformity component in our model is to provide a simple basis to guide agents' networking (linking) behavior. The next two terms in (1) – i.e., the idiosyncratic bias and the action cost– introduce different types of asymmetries between the two actions. The first one is associated with the fixed type $\gamma_i \in \{+1, -1\}$ of agent i and provides a clear meaning to the notion of type in our context: every agent has a bias in favor of the action that matches her type. In contrast, the second source of action asymmetry induced by κ is intrinsically associated with the action choice of the agent, independently of her type. The fact that we speak of it as a "cost" implicitly suggests that κ is positive and therefore presumes that the support for collective action (+1) bears an *intrinsic* disutility. Our model, however, allows for the possibility that agents may have a substantial preference for rioting that exceeds any costs (termed "entertainment value of participation" by Tullock (1971)). This would be captured by a $\kappa < 0$.[6] Finally, the last term in (1) includes linking costs ζ_{ij} for the bilateral link that may formed or maintained between any given pair of agents, i and j. Type heterogeneity is taken to have an impact on these costs as follows:[7]

$$\zeta_{ij} = \zeta_1 - \frac{\zeta_1 - \zeta_2}{2}(1 - \gamma_i\gamma_j) = \begin{cases} \zeta_1, & \text{if } \gamma_i = \gamma_j, \\ \zeta_2, & \text{if } \gamma_i \neq \gamma_j, \end{cases} \tag{2}$$

with $0 \leq \zeta_1 < \zeta_2$. The above formulation entails that agents with the same idiosyncratic preferences enjoy a lower linking cost, hence inducing a bias/preference for connections between individuals of the same type. This phenomenon, known as homophily, has been shown to be a common feature in human nature, long highlighted by sociologists (cf. Lazarsfeld and Merton, 2014; McPherson et al., 2001) and recently studied by economists as well (see, e.g., Currarini et al., 2009).

2.2. Dynamics

In our model, both action and linking choices are endogenous variables and define the state of the system, $\boldsymbol{\omega} = (\mathbf{s}, G) \in \Omega$, as it changes over time. For technical tractability, we model time continuously and denote it by $t \in [0, \infty)$. The dynamics consist of three components: action adjustment, link creation, and link removal, which will be separately defined below. These adjustments will be assumed to depend on the expected payoffs perceived by the agents at the time of their adjustment. This requires specifying how each agent i forms her beliefs $p_{it} \in [-1, +1]$ on the average action of others, $\frac{1}{n-1}\sum_{j\neq i}^{n} s_j$. For the moment, we formulate this in abstract terms and simply postulate that, for each i, her beliefs are related to the prevailing state through a function $\psi_i : \Omega \to [-1, +1]$. Different concrete possibilities for the functions $\psi_i(\cdot)$ are considered below, in Subsection 2.3.

action +1 being successful is given by the following "contest success function" (Tullock, 1980): $p^+ = \frac{n^+}{n^+ + n^-} = \frac{n^+}{n}$, where $n^{\pm} = |\{i \in \mathcal{N} : s_i = \pm 1\}|$ counts the number of agents choosing action +1 or −1, respectively. Then note that $n^+ = \frac{1}{2}\sum_{i=1}^{n}(1 + s_i)$ and therefore we can write $p^+ = \frac{1}{2n}\sum_{i=1}^{n}(1 + s_i) = \frac{1}{2}(1 + \bar{s})$, where $\bar{s} = \frac{1}{n}\sum_{i=1}^{n} s_i$ is the average action in the population. Knowing this, agent i' s expectations on $\bar{s}$ would give rise to (point) expectation of the probability on p^+ would be given by $\tilde{\psi}_i = \frac{1}{2}(1 + \psi_i)$, which yields global payoffs as specified in (1) with $\tilde{\psi}_i$ in lieu of $\frac{1}{2}(1 + \psi_i)$. It is easy to see that such affine transformation does not alter in any significant way our analysis.

[6] Note that since the actions are symmetric except for the effect of the parameter κ, the results for $\kappa < 0$ can be obtained from the results for $\kappa \geq 0$ by switching the labels from +1 to −1.

[7] For an extension of the model that allows for endogenous (i.e., action-dependent) linking costs, see Online Appendix B.2. There, we show that this extension induces the same functional form as in (1), up to a shift in parameter θ.

As customary in the evolutionary literature, expected payoffs will be assumed to be subject to persistent random noise. This noise can be motivated as the result of a number of different (non-exclusive) factors. One possibility proposed e.g., by Brock and Durlauf (2001), is that the game is subject to shocks, which are observed by the agents but not by the modeler. Another option is to suppose that the noise captures agents' uncertainty about the payoff (and hence behavior) of others. Finally, a third motivation that has been highlighted by evolutionary game theory (cf. Blume, 1993; Kandori et al., 1993; Young, 1993) is that agents make mistakes or simply experiment with some exogenous probability.

Mathematically, the evolutionary adjustment of actions and links defines a stochastic process that induces a probability measure over the set of state paths of the form $(\boldsymbol{\omega}_t)_{t\in\mathbb{R}_+}$, $\boldsymbol{\omega}_t \in \Omega$, where each state $\boldsymbol{\omega}_t = (\mathbf{s}_t, G_t)$ consists of a vector of agents' actions $\mathbf{s}_t \in \{-1,+1\}^n$ and a network $G_t \in \mathcal{G}^n$. Its law of motion can be described as follows.[8]

In every time interval of infinitesimal length, $[t, t+\Delta t)$, $t \in \mathbb{R}_+$, the following subprocesses simultaneously operate:

Action adjustment: *At rate $\chi > 0$, every agent $i \in \mathcal{N}$ is randomly given an independent opportunity to change her current action $s_{it} \in \{-1,+1\}$ to the alternative s_i'. Upon receiving this opportunity, the action change is implemented if, and only if, the agent perceives it beneficial in terms of the expected payoffs specified in (1) and an additive random shock ε_{it}. Thus, the probability that any given agent i switches from action s_{it} to s_i' is given by:*

$$\begin{aligned}\mathbb{P}\left(\boldsymbol{\omega}_{t+\Delta t} = (s_i', \mathbf{s}_{-it}, G_t) | \boldsymbol{\omega}_t = (s_{it}, \mathbf{s}_{-it}, G_t)\right) = \\ \chi\, \mathbb{P}\left(\pi_i(s_i', \mathbf{s}_{-it}, G_t; \psi_i(\mathbf{s}_t, G_t)) - \pi_i(s_{it}, \mathbf{s}_{-it}, G_t; \psi_i(\mathbf{s}_t, G_t)) + \varepsilon_{it} > 0\right) \Delta t + o(\Delta t)\end{aligned} \tag{3}$$

Link adjustment: *At rate $\lambda > 0$, every pair of agents, i and j, are randomly given an independent opportunity to either form a link ij if they are not currently connected, or remove a link if connected. Upon receiving this opportunity, the link is established/maintained if, and only if, both agents perceive it as beneficial in terms of the expected payoffs specified in (1) and an additive random shock $\varepsilon_{ij,t}$. Thus, the probability that any such link ij is formed is given by:*

$$\begin{aligned}&\mathbb{P}\left(\boldsymbol{\omega}_{t+\Delta t} = (\mathbf{s}_t, G_t \pm ij) | \boldsymbol{\omega}_{t-1} = (\mathbf{s}, G_t)\right) = \\ &\quad \lambda\, \mathbb{P}\big(\{\pi_i\left(\mathbf{s}_t, G_t \pm ij; \psi_i(\mathbf{s}_t, G_t)\right) - \pi_i(\mathbf{s}_t, G_t; \psi_i(\mathbf{s}_t, G_t)) + \varepsilon_{ij,t} > 0\} \;\cap \\ &\quad \{\pi_j(\mathbf{s}_t, G_t \pm ij; \psi_i(\mathbf{s}_t, G_t)) - \pi_j(\mathbf{s}_t, G_t; \psi_i(\mathbf{s}_t, G_t)) + \varepsilon_{ij,t} > 0\}\big)\Delta t + o(\Delta t),\end{aligned} \tag{4}$$

where $G_t \pm ij$ denotes the network G_t with the link ij added $(+)$ or removed $(-)$.

Throughout, we shall make the assumption that all random shocks are independently and logistically distributed with mean zero and the same scale parameter $\eta \geq 0$. Therefore, its cumulative distribution function $F_\varepsilon(x)$ is given by $\frac{e^{\eta x}}{1+e^{\eta x}}$ and we can write the action-adjustment rule (3) in

[8] The adjustment process has some similarity to that of Hsieh et al. (2022), time being measured continuously and revision opportunities arriving as a Poisson process (cf. Sandholm, 2010).

the following explicit form:[9]

$$\begin{aligned}\mathbb{P}(\boldsymbol{\omega}_{t+\Delta t} = (s_i',\mathbf{s}_{-it}, G_t) | \boldsymbol{\omega}_t = (s_i, \mathbf{s}_{-it}, G_t)) \\ &= \chi\, \mathbb{P}\left(-\varepsilon_{it} < \pi_i(s_i', \mathbf{s}_{-it}, G_t; \psi_i(\mathbf{s}_t, G_t)) - \pi_i(s_{it}, \mathbf{s}_{-it}, G_t; \psi_i(\mathbf{s}_t, G_t))\right) \Delta t + o(\Delta t) \\ &= \chi\, \frac{e^{\eta\, \pi_i(s_i', \mathbf{s}_{-it}, G_t;\, \psi_i(\mathbf{s}_t, G_t))}}{e^{\eta\, \pi_i(s_i', \mathbf{s}_{-it}, G_t;\, \psi_i(\mathbf{s}_t, G_t))} + e^{\eta\, \pi_i(s_i, \mathbf{s}_{-it}, G_t;\, \psi_i(\mathbf{s}_t, G_t))}} \Delta t + o(\Delta t).\end{aligned}$$

And, proceeding analogously for the link adjustment rule (4), we arrive at the following corresponding expressions for link adjustment

$$\begin{aligned}\mathbb{P}\left(\boldsymbol{\omega}_{t+\Delta t} = (\mathbf{s}_t, G_t \pm ij) | \boldsymbol{\omega}_t = (\mathbf{s}_t, G_t)\right) &= \lambda \frac{e^{\eta\, \pi_i(\mathbf{s}_t, G_t \pm ij;\, \psi_i(\mathbf{s}_t, G_t))}}{e^{\eta\, \pi_i(\mathbf{s}_t, G_t \pm ij;\, \psi_i(\mathbf{s}_t, G_t))} + e^{\eta\, \pi_i(\mathbf{s}_t, G_t;\, \psi_i(\mathbf{s}_t, G_t))}} \Delta t + o(\Delta t) \\ &= \lambda \frac{e^{\eta\, \pi_j(\mathbf{s}_t, G_t \pm ij;\, \psi_i(\mathbf{s}_t, G_t))}}{e^{\eta\, \pi_j(\mathbf{s}_t, G_t \pm ij;\, \psi_i(\mathbf{s}_t, G_t))} + e^{\eta\, \pi_j(\mathbf{s}_t, G_t;\, \psi_i(\mathbf{s}_t, G_t))}} \Delta t + o(\Delta t),\end{aligned}$$

where note that the link-adjustment probabilities are identical for the two agents, i and j, involved in any link change (be it creation or removal) because, given the logistic noise formulation, and the symmetry of the payoff function, the corresponding change in payoffs induced by it is the same for both of them.

2.3. Beliefs

To complete the description of the model, we now introduce the two different belief-formation scenarios that we shall consider and contrast. One embodies the classical formulation considered by much of the evolutionary literature of learning in games: at each point in the process, agents are completely informed of all the payoff-relevant features of the current state of the process. These features include, specifically, the average support for collective action provided by the rest of the population (i.e., the average action). For conciseness, this first scenario is labeled *Global Information* (GI), and is simply captured by the belief-formation mapping $\boldsymbol{\psi}^{GI} = \left(\psi_i^{GI}\right)_{i \in \mathcal{N}} : \Omega \to [-1, 1]^n$ that, for each $\boldsymbol{\omega} = (\mathbf{s}, G) \in \Omega$, is defined as follows:

$$\psi_i^{GI}(\boldsymbol{\omega}) = \frac{1}{n-1} \sum_{j=1, j \neq i}^{n} s_j \qquad (i = 1, 2, \ldots, n). \tag{5}$$

Thus, for every t, the beliefs $p_{it} = \psi_i(\omega_t)$ held by each agent $i \in \mathcal{N}$ coincide with the "true" average action chosen by the rest of the population.

In the alternative scenario, which we label *Local Information and Learning* (LIL), we suppose that agents gather information *locally* on the overall average support for collective from a combination of:

(a) the *observation* of that support among their network neighbors;
(b) the *learning* derived from interacting with those same neighbors.

The observation of the local average action posited in (a) derives from our assumption that every agent directly observes the actions $\{s_j : a_{ij} = 1\}$ of her network neighbors and, naturally, know their degree $d_i = \sum_{j \neq i}^{n} a_{ij}$. On the other hand, the local interaction/learning in (b) is modeled along the lines of the well-known framework proposed by DeGroot (1974).[10] More specifically, we posit

[9]Note that if z is logistically distributed with mean 0 and scale parameter η, then the random variable $\varepsilon = -z$ has a distribution function $F_\varepsilon(\cdot)$ given by $F_\varepsilon(y) = 1 - F_z(-y) = \frac{e^{\eta y}}{1+e^{\eta y}}$.

[10]See also Berger (1981), Jackson and Golub (2010), Golub and Jackson (2012) and DeMarzo et al. (2003). Chan-

that given the state $\boldsymbol{\omega_t} = (\mathbf{s}_t, G_t) = [(s_{1t}, s_{2t}, \ldots, s_{nt})^\top, G_t]$ prevailing at any given time t in the evolutionary adjustment process, there is a sequence of learning rounds, indexed by $u = 0, 1, 2, \ldots$, where the point beliefs $\mathbf{p}_t^u = (p_{1t}^u, p_{2t}^u, \ldots, p_{nt}^u)$ are updated as follows:

$$p_{it}^{u+1} = \varphi \underbrace{\frac{1}{d_{it}} \sum_{j=1}^{n} a_{ij,t} s_{jt}}_{\substack{\text{local average}\\ \text{actions}}} + (1-\varphi) \underbrace{\frac{1}{d_{it}+1} \left[p_{it}^u + \sum_{j=1}^{n} a_{ij,t} p_{jt}^u \right]}_{\substack{\text{local average}\\ \text{beliefs}}} \qquad (i = 1, 2, ..., n) \tag{6}$$

where the first term in (6) is assumed to be zero if $d_{it} = 0$, and $\varphi \in (0, 1)$ is the updating weight given to local observation while the complementary value $(1 - \varphi)$ is the weight given to social learning. Such social learning reflects the simple idea that agents update their beliefs by mixing uniformly their own previous beliefs and those of their network neighbors.

In line with the assumption made for the GI scenario, let us postulate that, also for the LIL scenario, the belief-updating adjustment formalized above occurs very fast and reaches a stationary point $\mathbf{p}_t^\star$. It can be easily confirmed that such a stationary point always exists and is unique. To see it, let us write (6) in compact matrix form as follows:

$$\mathbf{p}_t^{u+1} = \varphi \mathbf{D}_t^{-1} \mathbf{A}_t \mathbf{s}_t + (1-\varphi) \widehat{\mathbf{D}}_t^{-1} \widehat{\mathbf{A}}_t \, \mathbf{p}_t^u \tag{7}$$

where $\mathbf{D}_t \equiv \text{diag}(d_{1t}, \ldots, d_{nt})$ is the diagonal matrix of agents' degrees at t, $\widehat{\mathbf{D}}_t \equiv \mathbf{I}_n + \mathbf{D}_t$ with $\mathbf{I}_n$ being the identity matrix, $\widehat{\mathbf{A}}_t \equiv \mathbf{I}_n + \mathbf{A}_t$, and $\mathbf{p}_t$ and $\mathbf{s}_t$ are interpreted as column vectors of agents' beliefs and actions. Then, the induced stationary beliefs are given by:

$$\mathbf{p}_t^* = \varphi \left[\mathbf{I}_n - (1-\varphi) \widehat{\mathbf{D}}_t^{-1} \widehat{\mathbf{A}}_t \right]^{-1} \mathbf{D}_t^{-1} \mathbf{A}_t \mathbf{s}_t \tag{8}$$

which is a well-defined expression since the matrix $\widehat{\mathbf{D}}_t^{-1} \widehat{\mathbf{A}}_t$ is row-stochastic and $\varphi > 0$.[11]

Thus, in sum, the LIL scenario is characterized by the belief-formation mapping $\boldsymbol{\psi}^{LIL} = \left(\psi_i^{LIL}\right)_{i \in \mathcal{N}} : \Omega \to [-1, 1]^n$ that, for each $\boldsymbol{\omega} = (\mathbf{s}, G) \in \Omega$, is defined as follows:

$$\boldsymbol{\psi}^{LIL}(\boldsymbol{\omega}) = \varphi \left[\mathbf{I}_n - (1-\varphi) \widehat{\mathbf{D}}^{-1} \widehat{\mathbf{A}} \right]^{-1} \mathbf{D}^{-1} \mathbf{A} \mathbf{s}, \tag{9}$$

where $\mathbf{A}$ and $\mathbf{D}$ specify, respectively, the adjacency matrix of the network G and its corresponding diagonal matrix of agents' degrees while, as before, $\widehat{\mathbf{A}} \equiv \mathbf{I}_n + \mathbf{A}$ and $\widehat{\mathbf{D}} \equiv \mathbf{I}_n + \mathbf{D}$.

3. Theoretical Analysis

In this section we conduct the theoretical analysis of the model, studying in turn the two belief-formation scenarios considered: GI in Subsection 3.1, and LIL in Subsection 3.2. Then, in Subsection 3.3 we compare the conclusions and insights derived from the two contexts.

drasekhar et al. (2015) and Grimm and Mengel (2015) provide empirical evidence that individuals who attempt to learn the underlying state of the world in a network are well described by DeGroot-type models.

[11] Of course, this would not be the case for the extreme value of $\varphi = 0$, for which we would arrive at the customary DeGroot model. In it, actions play no role and stationary beliefs – if they exist – depend on the starting ones $\mathbf{p}_0$ in a way that reflects the architecture of the social network.

3.1. Global Information

In order to characterize the long-run behavior of the system under GI, the key point to note is that, given the belief-formation mapping $\boldsymbol{\psi}^{GI} = (\psi_i^{GI})_{i\in\mathcal{N}}$ defined in (5), the function $\Phi : \Omega \to \mathbb{R}$ given by:

$$\Phi(\boldsymbol{\omega}) = \sum_{i=1}^{n} \left\{ (\gamma_i - \kappa)s_i + \sum_{j=1, j\neq i}^{n} \tfrac{1}{2} \left[\rho s_i \psi_i^{GI}(\boldsymbol{\omega}) + a_{ij}(\theta s_i s_j - \zeta_{ij}) \right] \right\} \tag{10}$$

which can be rewritten as

$$\Phi(\boldsymbol{\omega}) = \sum_{i=1}^{n} \left\{ (\gamma_i - \kappa)s_i + \tfrac{1}{2} \left[\rho s_i (n-1) \psi_i^{GI}(\boldsymbol{\omega}) + \sum_{j\neq i} a_{ij}(\theta s_i s_j - \zeta_{ij}) \right] \right\} \tag{11}$$

is a *potential* for the expected payoff function given in (1) under belief-formation rule $\boldsymbol{\psi}^{GI}$ defined in (5). This means that, for any change in a *single* component of the state (an action or a link) involving any particular agent i, the change on the expected payoffs $\pi_i(\cdot; \psi_i^{GI}(\cdot))$ experienced by this agent matches exactly the corresponding change displayed by the function $\Phi(\cdot)$. Formally, the following two conditions must hold $\forall i, j \in \mathcal{N}$, $G \in \mathcal{G}^n$, $\mathbf{s} \in \mathcal{S}$ and $s_i' \in \mathcal{S}_i = \{-1, +1\}$:

$$\begin{gathered} \text{Let } \boldsymbol{\omega} = (s_i, \mathbf{s}_{-i}, G),\ \boldsymbol{\omega}' = (s_i', \mathbf{s}_{-i}, G) \in \Omega. \text{ Then,} \\ \Phi(\boldsymbol{\omega}') - \Phi(\boldsymbol{\omega}) = \pi_i(\boldsymbol{\omega}'; \psi_i^{GI}(\boldsymbol{\omega})) - \pi_i(\boldsymbol{\omega}; \psi_i^{GI}(\boldsymbol{\omega})). \end{gathered} \tag{12}$$

$$\begin{aligned} &\text{Let } \boldsymbol{\omega} = (\mathbf{s}, G),\ \boldsymbol{\omega}' = (\mathbf{s}, G \pm ij) \in \Omega. \text{ Then,} \\ \Phi(\boldsymbol{\omega}') - \Phi(\boldsymbol{\omega}) &= \pi_i(\boldsymbol{\omega}'; \psi_i^{GI}(\boldsymbol{\omega})) - \pi_i(\boldsymbol{\omega}; \psi_i^{GI}(\boldsymbol{\omega})) \\ &= \pi_j(\boldsymbol{\omega}'; \psi_j^{GI}(\boldsymbol{\omega})) - \pi_j(\boldsymbol{\omega}; \psi_j^{GI}(\boldsymbol{\omega})). \end{aligned} \tag{13}$$

In the second condition above, $G \pm ij$ stands for the network given by G with the link ij added $(+)$ or deleted $(-)$. A compact way of rewriting the previous two conditions that will be useful hereafter is as follows. Given any agent i and two states $\boldsymbol{\omega}$ and $\boldsymbol{\omega}'$, let us say that $\boldsymbol{\omega}'$ is adjacent to $\boldsymbol{\omega}$ for agent i and write $\boldsymbol{\omega}' \in \mathcal{A}_i(\boldsymbol{\omega})$ if the former state differs from the latter either in action chosen by i or in one of her links (either present or absent in state $\boldsymbol{\omega}$). Then, it is easy to confirm that an equivalent way of rewriting (12)-(13) as a single condition is:

$$\begin{gathered} \text{Let } \boldsymbol{\omega}, \boldsymbol{\omega}' \in \Omega \text{ be such that } \boldsymbol{\omega}' \in \mathcal{A}_i(\boldsymbol{\omega}) \text{ for some } i \in \mathcal{N}. \text{ Then,} \\ \Phi(\boldsymbol{\omega}') - \Phi(\boldsymbol{\omega}) = \pi_i(\boldsymbol{\omega}'; \psi_i^{GI}(\boldsymbol{\omega})) - \pi_i(\boldsymbol{\omega}; \psi_i^{GI}(\boldsymbol{\omega})). \end{gathered} \tag{14}$$

To understand intuitively why the function $\Phi(\cdot)$ defined in (11) satisfies (14), it is useful to think of it as adding the payoffs attained by all individuals in any given state $\boldsymbol{\omega}$, with the important caveat that all those payoffs derived from interaction (i.e., from coordination or link formation) are split equally between the two agents involved. It is precisely because of this feature, and the additivity plus agent symmetry displayed across the different components of payoff interaction, that the function $\Phi(\cdot)$ yields the desired match between: (a) the change in individual payoffs experienced by the agent(s) involved in any action or link revision, (b) the corresponding potential change. For the sake of completeness, this straightforward conclusion is stated in the following result, whose formal proof can be found in Appendix A.

Proposition 1. *Given beliefs $\psi_i^{GI}(\cdot)$, the function $\Phi(\cdot)$ given in (10) defines a potential for the agents' expected payoffs $\pi_i(\cdot; \psi_i^{GI}(\cdot))$ for each $i \in \mathcal{N}$, as specified in (1); that is, condition (14) holds.*

The previous result leads to a couple of interesting consequences. The first one is standard: *noiseless* best-response adjustment of actions and links converge, almost surely, to an equilibrium (stationary point) where the potential is locally maximized.[12] A second consequence is that, even for possibly large noise (as captured by a low η), the existence of such a potential leads to a probabilistic prediction about the long-run behavior of the induced stochastic process that has a sharp exponential form. For, as stated by our next result (whose proof is also included in Appendix A), an adaptation of arguments used in traditional models of statistical physics delivers the following proposition (cf. Blume, 1993; Sandholm, 2010).

Proposition 2. *Consider the stochastic process $(\boldsymbol{\omega}_t)_{t\in\mathbb{R}_+}$ defined by (3)-(5), where the additive shocks perturbing agents' payoffs are i.i.d. logistically distributed with parameter $\eta > 0$ and the belief-formation rules of the global-information scenario apply, i.e., they are given by the function $\psi_i^{GI}(\cdot)$. Then, this process induces an ergodic Markov chain whose unique invariant distribution μ_η^{GI}, defined on the measurable space $(\Omega, \mathcal{F})$, is determined for every $\omega = (\mathbf{s}, G) \in \Omega$ as follows:*

$$\mu_\eta^{GI}(\boldsymbol{\omega}) = \frac{e^{\eta\Phi(\boldsymbol{\omega})}}{\sum_{\boldsymbol{\omega}'\in\Omega} e^{\eta\Phi(\boldsymbol{\omega}')}} = \frac{e^{\eta\Phi(\mathbf{s},G)}}{\sum_{G'\in\mathcal{G}^n}\sum_{\mathbf{s}'\in\{-1,+1\}^n} e^{\eta\Phi(\mathbf{s}',G')}}. \tag{15}$$

The above proposition provides an explicit solution of the model in the GI scenario by specifying, in closed form, how the probability distribution μ_η^{GI} that characterizes the long-run behavior of the process depends on the noise level (modulated by η) and all other parameters of the model. This theoretical prediction, captured by (15), is illustrated with numerical simulations in Supplementary Appendix D.1.

A straightforward implication of Proposition 2 is that the changes across adjacent states in the potential $\Phi(\cdot)$ and the distribution μ_η^{GI} are of the same sign, i.e., for all $\eta > 0$, $\boldsymbol{\omega}, \boldsymbol{\omega}' \in \Omega$, $\mu_\eta^{GI}(\boldsymbol{\omega}') \geq \mu_\eta^{GI}(\boldsymbol{\omega}) \Leftrightarrow \Phi(\boldsymbol{\omega}') \geq \Phi(\boldsymbol{\omega})$. Furthermore, it also follows that the set of potential-maximizing states:

$$\Omega^* \equiv \{\omega \in \Omega : \ \Phi(\omega) \geq \Phi(\omega') \ \ \forall \omega' \in \Omega\} = \{\omega \in \Omega : \ \mu_\eta^{GI}(\boldsymbol{\omega}) \geq \mu_\eta^{GI}(\boldsymbol{\omega}') \ \ \forall \omega' \in \Omega\}. \tag{16}$$

singles out those states that are individually attributed the highest probability by the uniquely invariant (or stationary) distribution of the process. Or, if the noise of the process becomes vanishingly small, it also identifies the support of the induced distribution, thus singling out the states that are observed with significant probability under infinitesimal noise (see Subsection 3.3 below).

Motivated by these considerations, in what follows we characterize the set Ω^*, as a function of the parameters of the model. For expositional simplicity, our focus below is on the most interesting case where the population is large and the lowest linking cost is not prohibitive (i.e., $\zeta_1 < \theta$).[13] Also, it is useful to divide the analysis into two subcases, depending on whether the fraction of individuals of type $+1$ in the population (a parameter of the model) is higher or lower than $1/2$. Thus, denoting such fraction by ν_+ we arrive at our main result for the GI scenario.

Proposition 3. *Assume all agents form their beliefs as prescribed by the function $\boldsymbol{\psi}^{GI}(\cdot)$ defined by (5). Let $n_+ = \#(\{\gamma_i = +1 : i \in \mathcal{N}\})$ and $\nu_+ = n_+/n$. Then there exists some $\hat{n} \in \mathbb{N}$ such that if $n \geq \hat{n}$, the set Ω^* is generically a singleton and its unique state $\boldsymbol{\omega}^* = (\boldsymbol{s}^*, G^*) \in \Omega^*$ can be characterized as follows:*

[12] This equilibrium can be viewed as a Nash equilibrium of the complete-information game where actions and linking proposals are chosen independently by the agents, the links being created and maintained only by consensus of the agents involved.

[13] See Supplementary Appendix E.1 for a general characterization for any finite population size n.

(a) The action profile $\boldsymbol{s}^* = (s_i^*)_{i\in\mathcal{N}}$ *satisfies:*

- *if* $\kappa > 2\nu_+ - 1$, $s_i^* = -1$ *for all* $i \in \mathcal{N}$.
- *if* $\kappa < 2\nu_+ - 1$, $s_i^* = +1$ *for all* $i \in \mathcal{N}$.

(b) The network G^* *satisfies:*

- *if* $\theta > \zeta_2$ *the network is complete, whereas if* $\theta < \zeta_1$ *it is empty.*
- *if* $\zeta_1 < \theta < \zeta_2$ *the network is segmented into two (disjoint and completely connected) cliques,* K_{n_+} *and* K_{n-n_+}*, with no links across them and each including all agents of types* $+1$ *and* -1*, respectively.*

The characterization provided by Proposition 3 delivers some intuitive conclusions. The first two items (a)-(b), which pertain to action choice, indicate that the generically unique state $\boldsymbol{\omega}^* = (\boldsymbol{s}^*, G^*) \in \Omega^*$ that maximizes the potential $\Phi(\cdot)$ (and therefore the long-run probability given by $\mu_\eta^{GI}(\cdot)$) involves a uniform action profile where every agent in the population plays either action $+1$ or -1 depending on whether or not the inherent negative bias against action $+1$ given by its cost (assuming $\kappa > 0$) is compensated by the strength of the coordination-based incentives derived from choosing that action if $\nu_+ > 1/2$ (and therefore there is majority of agents in the population that have a preference for action $+1$). The last two items (c)-(d) concern the network component of the state $\boldsymbol{\omega}^*$ and simply specify that the associated network G^* is segmented along type-homogenous cliques if the linking costs across different types exceed the best possible payoff to be expected from local interaction and it is completely connected if the opposite inequality holds. A prominent feature of these conclusions is that homophily (as reflected by a gap in linking costs that has network implications) has no eventual bearing on the long-run choice of actions. They are independent in the sense that while the latter only depends on κ and ν_+, the former only responds to the values of θ and ζ_2. As we shall see, such independence does not arise in the LIL scenario, and indeed it is the key feature that explains the contrast between the two scenarios.

3.2. Local Information and Learning

Now we turn to the study of the LIL scenario where, as explained in Subsection 2.3, beliefs are governed by the mapping $\boldsymbol{\psi}^{LIL}(\cdot)$ given by (9). By analogy with the GI scenario, it may be speculated that the expression in (11) that defines a potential in that case can be adapted to the LIL context by simply relying on the corresponding belief-formation mapping $\boldsymbol{\psi}^{LIL}(\cdot)$ in lieu of $\boldsymbol{\psi}^{GI}(\cdot)$. This, however, does not work because, in contrast with the GI scenario, the aggregate "externality" imposed on the ***expected*** payoffs of the overall population by any agent revising her behavior (action or link) is no longer equal to the effect induced on this agent's own payoff. In fact, such an externality will tend to decay fast along the social network (and therefore become relatively small for large populations) if social learning – instead of local observation – is the main driving force of belief formation. This in turn suggests that an adaptation of (11) that simply dispenses with the "correction factor" of 1/2 on expected global payoffs may behave as a good approximate potential for the LIL scenario. More precisely, the conjecture is that the function $\tilde{\Phi} : \Omega \to \mathbb{R}$ defined, for each $\boldsymbol{\omega} = (\mathbf{s}\,, G) \in \Omega$, as follows

$$\tilde{\Phi}(\boldsymbol{\omega}) = \sum_{i=1}^{n} \left\{ (\gamma_i - \kappa)s_i + \rho s_i(n-1)\psi_i^{LIL}(\boldsymbol{\omega}) + \tfrac{1}{2}\left[\sum_{j=1, j\neq i}^{n} a_{ij}(\theta s_i s_j - \zeta_{ij}) \right] \right\}, \tag{17}$$

may capture the payoff incentives underlying agents' adjustments quite well in the LIL scenario, provided that the population is large and social learning plays a dominant role in belief formation. Indeed, this is the gist of Proposition 4 below. Specifically, it shows that, under the aforementioned circumstances, such a modified function behaves essentially as an ordinal potential for LIL-based adjustment.

Before proceeding to the formal statement of this result, it may be useful to understand heuristically the key ideas underlying its proof. To this end, let us start with the simple observation that, for any given pair of adjacent states for an agent i, $\boldsymbol{\omega}$ and $\boldsymbol{\omega}'$, the induced change in the corresponding values of the function $\tilde{\Phi}$ when transitioning from the former state to the latter satisfies:

$$\tilde{\Phi}(\boldsymbol{\omega}')-\tilde{\Phi}(\boldsymbol{\omega}) = \pi_i(\boldsymbol{\omega}', \psi_i^{LIL}(\boldsymbol{\omega})) - \pi_i(\boldsymbol{\omega}, \psi_i^{LIL}(\boldsymbol{\omega})) + \rho(n-1)\sum_{j=1}^{n} s_j'(\psi_j^{LIL}(\boldsymbol{\omega}') - \psi_j^{LIL}(\boldsymbol{\omega})). \quad (18)$$

Thus, the change induced on the function $\tilde{\Phi}(\cdot)$ from such a transition is equal to the change in payoffs of the agent(s) involved in it plus a remainder given by $\rho(n-1)\sum_{j=1}^{n} s_j'(\psi_j^{LIL}(\boldsymbol{\omega}') - \psi_j^{LIL}(\boldsymbol{\omega}))$. The interpretation of this remainder should be clear: it measures the aggregate impact of the change on the expected payoffs that the population at large derives from global interaction. In this light, therefore, in order to show that the function $\tilde{\Phi}(\cdot)$ behaves approximately as a potential in the LIL context, it needs to be shown that any such transition has a small impact on population beliefs. Our approach to proving this conclusion relies on the following assumption: as the population grows large (its size n rises) and social learning becomes more important (i.e., $(1-\varphi)$ approaches 1), the long-run evolution of the process (as summarized by its limit distribution) yields social networks with the properties typically found in large random networks with homophily – in particular, sparseness and global connectivity (see below for the formal details).

Mathematically, our approach considers the *sequence of state spaces* associated to progressively larger population sizes n, $\{\Omega_n = \{\mathbf{S}_n, \mathcal{G}_n\}\}_{n\in\mathbb{N}}$, and the corresponding sequence $\{\mu_n^{LIL}\}_{n\in\mathbb{N}}$ of (unique) invariant distributions induced by the respective adjustment process in the LIL scenario.[14] Denote by $d_i(\boldsymbol{\omega})$ the degree of agent i in state $\boldsymbol{\omega} = (G, \mathbf{s})$ and by $\bar{d}(\boldsymbol{\omega})$, $\bar{d}^-(\boldsymbol{\omega})$, and $\bar{d}^+(\boldsymbol{\omega})$ the average degrees over all nodes or restricted to those choosing actions $+1$ or -1, respectively. Then, along the sequence $\{\Omega_n\}_{n\in\mathbb{N}}$, our focus turns to state subsets $\hat{\Omega}_n \subset \Omega_n$ satisfying the following properties:

(P1) *Growing but sparse connectivity*: $\exists\{(b(n), c(n))\}_{n\in\mathbb{N}}$ satisfying $\lim_{n\to\infty} b(n) = \infty$ and $\lim_{n\to\infty} c(n)/n = 0$ such that if $\boldsymbol{\omega} = (G, \mathbf{s}) \in \hat{\Omega}_n$, $\forall i \in \mathcal{N}$, $b(n) \leq d_i(\boldsymbol{\omega}) \leq c(n)$.

(P2) *Network connectedness*: $\forall \boldsymbol{\omega} = (G, \mathbf{s}) \in \hat{\Omega}_n$, $\forall \boldsymbol{\omega}' \in \mathcal{A}(\boldsymbol{\omega}) \equiv \cup_{i\in\mathcal{N}}\mathcal{A}_i(\boldsymbol{\omega})$, networks G and G' display a single component, i.e., all nodes are path-connected.

(P3) *Homophily.* $\forall \boldsymbol{\omega} = (G, \mathbf{s}) \in \hat{\Omega}_n$, $\sum_{i\in\mathcal{N}} s_i \geq 0 \Leftrightarrow \bar{d}^+(\boldsymbol{\omega}) \geq \bar{d}^-(\boldsymbol{\omega})$.

Verbally, the former properties can be described as follows. P1 asserts that, as the population size grows, even though individual degrees also grow without bound, their growth is much slower than that of the population – consequently, overall network connectivity remains sparse. Next, P2 switches attention to global/indirect connectivity (rather than local/direct) and posits that, for every pair of nodes, there is a path on the network indirectly connecting one to the other. Finally, P3 reflects the idea that, because of the action-based homophily displayed by our model, those agents taking the action chosen by the majority of the population enjoy a higher average degree than the ones choosing the alternative action.

[14]The uniqueness of the invariant distribution follows from the fact that, due the contemplated noise impinging on

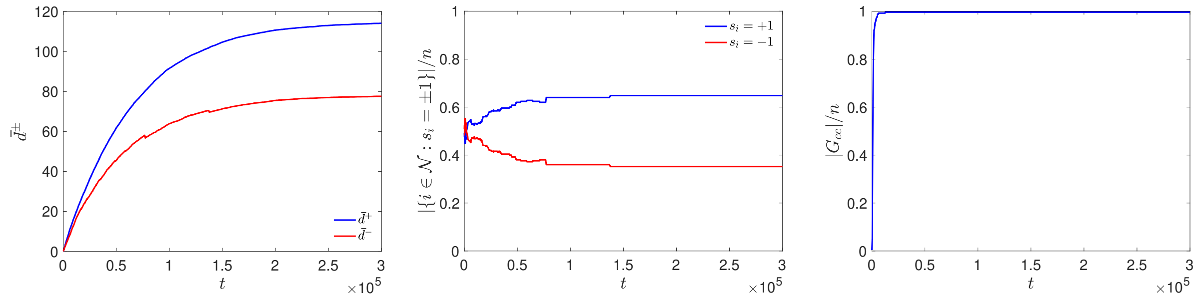

Figure 1: The left panel shows the average degree of agents choosing the majority action, $\bar{d}^{-}$, and the average degree of the agents choosing the minority action, $\bar{d}^{+}$, the middle panel shows the fraction of agents choosing the action plus one, $|\{i \in \mathcal{N} : s_i = +1\}|/n$, or minus one, $|\{i \in \mathcal{N} : s_i = -1\}|/n$ and the right panel shows the fraction of agents in the largest component $|G_{cc}|/n$. The parameters used are $n = 200$, $\zeta_1 = 0.1$, $\zeta_2 = 0.8$, $\theta = 0.15$, $\rho = 0.01$, $\beta = 0.5$, $\varphi = 0.1$, $\kappa = -1.5$, and $\eta = 1$.

The features embodied by P1-P3 are commonly displayed by dynamic network models that, as in our case, combine random link formation and homophily – see, e.g., Currarini et al. (2009). The forces underlying such conclusions in those more stylized setups operate in our context as well. Thus, since the complexity of our model renders it difficult to derive those properties from more primitive conditions, we suppose instead that they are held with high probability in the long run and build upon this assumption in the ensuing analysis. In support of this approach, we have conducted extensive numerical simulations for our LIL scenario whose results, illustrated in Figure 1, are well aligned with P1-P3. They show that, after a transitory phase, the process reaches a stationary state in which the fraction of agents choosing the action $+1$, n^{+}/n, or -1, n^{-}/n, stabilize, the average degree of agents choosing the majority action, $\bar{d}^{-}$, is higher than the average degree of the agents choosing the minority action, $\bar{d}^{+}$, and the size of the largest component encompasses all agents in the network (i.e., $|G_{cc}|/n$ approaches 1).

Building upon Properties P1-P3, the main assumption of our analysis is that, as the population size n gets large, the states that satisfy them are observed in the long run with a probability that approaches one, as captured by the corresponding limit invariant distribution μ_n^{LIL} (where note that, for notational simplicity, we abstract from the dependence on η, which is taken to be fixed in the present discussion). Thus, we assume:

A.1: There exists a sequence $\left\{\hat{\Omega}_n\right\}_{n=1,2,\ldots}$, with $\hat{\Omega}_n$ satisfying P1-P3 for all n, such that $\lim_{n\to\infty} \mu_n^{LIL}(\hat{\Omega}_n) = 1$.

To simplify the analysis, it is convenient to rely as well on the ancillary assumption that, if the population size n is large enough, the long-run probability mass that the process associates to a "thin band" around the states with an average action exactly equal to zero can be made arbitrarily small. Our formal statement of this assumption uses the following notation. Given any $n \in \mathbb{N}$ and $\delta > 0$, define $H_n^{\delta} \equiv \{\boldsymbol{\omega} = (\mathbf{s}, G)) \in \Omega_n : |\frac{1}{n}\sum_{i\in\mathcal{N}} s_i| \leq \delta\}$ as the set of all those states that, for a population of the size of n, have their average action no farther from zero than δ. Then, we posit:

A.2: For any $\epsilon > 0$ there exists some $\delta > 0$ and $\hat{n} \in \mathbb{N}$ such that, if $n \geq \hat{n}$, $\mu_n^{LIL}(H_n^{\delta}) \leq \epsilon$.

The rationale for this assumption is derived from the idea that the set of states whose action profile $\mathbf{s} \in \{-1, +1\}^n$ lies in a small neighborhood around the "fragile" point of exactly equal balance between the two actions must be quite ephemeral – hence it should be observable in the long run only with very small probability if the population is large.

agents' adjustments, the induced stochastic process is clearly ergodic.

Under (A.1)-(A-2), the following proposition establishes conditions under which the function $\tilde{\Phi}$ behaves approximately as an ordinal potential. This notion, a relaxation of the formerly used one of (cardinal) potential, only requires that the payoffs of a revising agent and the corresponding changes in the potential display the same sign. That is, instead of the condition (14) considered for the GI scenario, we now rely on the following condition:

$$\begin{aligned}&\text{Let } \boldsymbol{\omega}, \boldsymbol{\omega}' \in \Omega \text{ be such that } \boldsymbol{\omega}' \in \mathcal{A}_i(\boldsymbol{\omega}) \text{ for some } i \in \mathcal{N}. \text{ Then,}\\ &\tilde{\Phi}(\boldsymbol{\omega}') \geq \tilde{\Phi}(\boldsymbol{\omega}) \Leftrightarrow \pi_i(\boldsymbol{\omega}'; \psi_i^{LIL}(\boldsymbol{\omega})) \geq \pi_i(\boldsymbol{\omega}; \psi_i^{LIL}(\boldsymbol{\omega})).\end{aligned} \tag{19}$$

As advanced, the following result establishes that the function $\tilde{\Phi}(\cdot)$ satisfies the above requirement with arbitrarily high probability in the long run if the population is large and the weight of social learning is high.

Proposition 4. *Let the adjustment dynamics be as in the LIL scenario and consider a sequence of growing populations of size n, assume A.1 and A.2, and let $\zeta_1 \neq \theta \neq \zeta_2$. Consider a sequence of environments with a growing population size n and let $\{\tilde{\Omega}_n\}_{n=1,2,...}$ be a correspondent sequence of state subsets such that, for each n, $\tilde{\Omega}_n$ is the set of states for that population size that satisfy (19). Then, given any $\epsilon > 0$, there exist $\hat{n} \in \mathbb{N}$ and $\hat{\varphi} > 0$ such that if $n \geq \hat{n}$ and $\varphi \leq \hat{\varphi}$, $\mu_\eta^{LIL}(\tilde{\Omega}_n) \geq 1-\epsilon$.*

As explained, compared to the result established by Proposition 1 for the GI scenario, the conclusion claimed for the LIL scenario by Proposition 4 is weaker in two respects: the function $\tilde{\Phi}(\cdot)$ is claimed to behave as a potential (a) only in an ordinal sense and (b) only with arbitrarily high probability, provided the population size and the relative weight of social learning are large enough. While the latter caveat is in line with our desire to focus on large-scale problems of collective action where the possibilities for direct observation are severely limited by the size of the population, the ordinal relaxation of the potential-alignment property is more problematic. In particular, it prevents us from claiming, in full mathematical rigor, that the long-run distribution of the process has the precise exponential form established for the GI scenario.

The previous considerations notwithstanding, we can still rely on the main idea underlying Proposition 4 – namely, that under the specified conditions the deviations from strict potential alignment should be typically small – to suggest a close parallelism between the theoretical approach pursued in the GI and LIL scenarios. More concretely, we conjecture that the same exponential function of the potential $\Phi(\cdot)$ used under global GI should approximate well, when applied to $\tilde{\Phi}(\cdot)$, the long-run behavior of the LIL dynamics. This leads us to positing that the probability distribution $\mu_\eta^{LIL}(\cdot)$ defined, for every $\omega = (\mathbf{s}, G) \in \Omega$, as follows:

$$\mu_\eta^{LIL}(\boldsymbol{\omega}) = \frac{e^{\eta\tilde{\Phi}(\boldsymbol{\omega})}}{\sum_{\boldsymbol{\omega}'\in\Omega} e^{\eta\tilde{\Phi}(\boldsymbol{\omega}')}} = \frac{e^{\eta\tilde{\Phi}(\mathbf{s},G)}}{\sum_{G'\in\mathcal{G}^n}\sum_{\mathbf{s}'\in\{-1,+1\}^n} e^{\eta\tilde{\Phi}(\mathbf{s}',G')}}, \tag{20}$$

approximates well the invariant distribution of the adjustment process under LIL.

To explore the validity of the former claim, we have conducted a wide range of numerical exercises that approach the issues from a variety of different complementary angles. First, at the most basic (but also only indirect) level, we have obtained the results displayed in Figure 2, which show a strong correlation between the changes in payoffs and the function $\tilde{\Phi}(\cdot)$. Second, proceeding in a more direct manner, we have generated a wide range of simulations of the LIL-based adjustment process that – as a counterpart of those shown in Figure D.1 for the GI scenario in Supplementary Appendix D.1 – trace empirically the average values for key population variables as one changes the model parameters. These numerical results, found in Figure D.2 in Supplementary Appendix D.2, are closely aligned with the theoretical predictions induced by the distribution $\mu_\eta^{LIL}(\cdot)$ defined

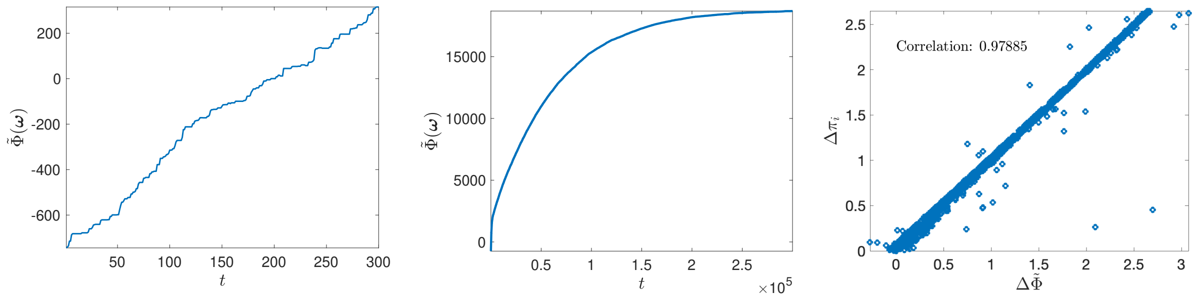

Figure 2: The left and middle panels show a typical trajectory of the quasi-potential $\tilde{\Phi}$, while the right panel shows the correlation between the change in the quasi-potential and in payoffs, $\Delta\tilde{\Phi}$ and $\Delta\pi_i$. The parameters used are the same as in Figure 1.

by (20). And third, we have also carried out a parameter-recovery exercise that we believe delivers the strongest support for our overall approach. For, in an integrated manner, not only does it support the theory but also the econometric methodology we have pursued to bring that theory to the data. This exercise is explained in detail in Supplementary Appendix I, once the econometric methodology has been presented formally. Here we just provide a brief and informal advance. It starts with the generation of artificial data obtained by simulating the dynamics of our process according to the rules posited in our theoretical model. And, as a cross-check on the validity of our general approach, we do this not only for the LIL scenario but also for the GI scenario. We then apply our composite-likelihood econometric approach – based on either the distribution $\mu_\eta^{GI}(\cdot)$ or $\mu_\eta^{LIL}(\cdot)$, depending on the scenario considered – to such artificially generated data and obtain (highly significant) estimates for all of the model parameters involved in each case. Finally, we contrast the estimated parameters with the ones actually used to generate the data and find that they are very close in both contexts. In view of form displayed by $\mu_\eta^{LIL}(\cdot)$ in (20) we can again claim for the LIL scenario that the long-run probability across adjacent states imputed by this distribution changes in the same direction as their respective values of the function $\tilde{\Phi}(\cdot)$ and that, as a counterpart of (16), the set of states that maximize such probability, denoted by Ω^{**}, maximize such probability, i.e.,

$$\Omega^{**} \equiv \{\omega \in \Omega : \ \tilde{\Phi}(\omega) \geq \tilde{\Phi}(\omega') \ \ \forall \omega' \in \Omega\} = \{\omega \in \Omega : \ \mu_\eta^{LIL}(\boldsymbol{\omega}) \geq \mu_\eta^{LIL}(\boldsymbol{\omega}') \ \ \forall \omega' \in \Omega\}. \tag{21}$$

The following proposition characterizes the set Ω^{**}, as a function of the parameters of the model, under the same conditions characterizing Ω^* in Propositions 3 for the GI scenario.[15]

Proposition 5. *Assume all agents form their beliefs as prescribed by the function $\psi^{LIL}(\cdot)$ given by (9) and suppose $\theta > \zeta_1$. Let $n_+ = \#(\{\gamma_i = +1 : i \in \mathcal{N}\})$ and $\nu_+ = n_+/n$. Then there exists some $\hat{n} \in \mathbb{N}$ such that if $n \geq \hat{n}$, the set Ω^{**} is generically a singleton and its unique state $\boldsymbol{\omega}^{**} = (\boldsymbol{s}^{**}, G^{**}) \in \Omega^{**}$ can be characterized as follows:*

(a) If $\kappa > 1$, the action profile in $\boldsymbol{s}^{**} = (s_i^{**})_{i\in\mathcal{N}}$ has $s_i^{**} = -1$ for all $i \in \mathcal{N}$.

(b) If $\kappa < 1$, the action profile in $\boldsymbol{s}^{**} = (s_i^{**})_{i\in\mathcal{N}}$ is such that:

(i) when $\theta < \zeta_2$, for all $i \in \mathcal{N}$ we have $s_i^{**} = \gamma_i$;

(ii) when $\theta > \zeta_2$, for all $i \in \mathcal{N}$ we have $s_i^{**} = \begin{cases} +1 & \text{if } \kappa < 2\nu_+ - 1, \\ -1 & \text{if } \kappa > 2\nu_+ - 1. \end{cases}$

(c) If $\theta < \zeta_2$, G^{**} is segmented into two (completely connected) cliques, K_{n_+} and K_{n-n_+}, with

[15] In Supplementary Appendix E.2 we provide a characterization for any finite population size n.

no cross-links and including all agents of types $+1$ *and* -1*, respectively.*

(d) *If* $\theta > \zeta_2$*,* G^{**} *coincides with the complete network.*

The previous proposition provides an interesting contrast with the counterpart result (Proposition 3) we obtained for the GI scenario. As advanced, the main difference derives from the fact that, in the present LIL scenario, the action profiles displayed by the state $\boldsymbol{\omega}^{**} \in \Omega^{**}$ as parameters change are *not* independent of the conditions that shape the network. Indeed, it is still true that the properties of the (endogenous) network (that is, whether it is complete or segmented) exclusively depend on whether or not the linking cost across types, ζ_2, exceeds the local-interaction parameter θ. However, the important difference to note is that, in the LIL scenario, what network architecture then arises has a key bearing on how the cost κ affects the action choices. Thus, if the network is fragmented – because $\theta < \zeta_2$ and (c) applies – then (a) and (b.i) indicate that the action choice depends solely on whether the cost κ is lower or higher than 1, independently on what is the fraction ν_+ of agents of type $+1$. Instead, the interplay between κ and ν_+ becomes again important, just as it was the case in the GI scenario, when $\theta > \zeta_2$. For, in this case, (d) implies that the network is complete and (b.ii) applies. In the following subsection we further elaborate on this interplay between network architecture and action choice, discussing the insights it contributes to the question and how even large populations are sometimes able to tackle demanding collective action problems.

3.3. Achieving Collective Action under Small Noise

The analysis conducted in the previous subsection leads to some useful insights on how the alternative informational assumptions contemplated by our two scenarios, GI and LIL, bear on the process through which a large population may attain collective action. This can be assessed by comparing the action profiles prevailing in each context in the states $\boldsymbol{\omega}^{*}$ and $\boldsymbol{\omega}^{**}$ characterized, respectively, in Propositions 3 and 5. As explained before, the contrast between these two contexts arises when their different belief-formation rules interplay with homophily in network formation, resulting from a gap in the linking costs incurred by agents of the same or different types. Focusing for the moment on the simple question of whether the induced states induce some positive fraction of agents choosing action $+1$, a graphical summary of the situation is provided in Figure 3. There we see that while the condition $\kappa < 1$ is always necessary for both scenarios (i.e., action $+1$ cannot be too costly), only under LIL is it also sufficient. For, under GI, in order to attain some degree of collective action two additional conditions are needed. One is that the individuals of type $+1$ are a majority of the population (i.e., represent more than 50% of it). A second condition is that $\kappa < 2\nu_+ - 1$. This inequality is obviously more stringent than simply requiring that $\kappa < 1$, except for the extreme case where $\nu_+ = 1$, i.e., when essentially all individuals in the population are of type $+1$. Overall, therefore, we conclude that, in the presence of consequential homophily, the LIL scenario provides a much wider set of conditions for which, in the stationary configuration of the adjustment process, there is some significant extent of collective action. Those conditions not only pertain to the magnitude of the cost incurred by joining it but, most importantly, to the fraction of the population whose type favors doing so (which is subject to no lower bound).

Another complementary perspective to understand the contrast between the GI and LIL scenarios is to compare how the fraction of the population choosing action $+1$ in states $\boldsymbol{\omega}^{*}$ and $\boldsymbol{\omega}^{**}$, which we respectively denote by $\varsigma^{*}_{+} = |\{i \in \mathcal{N} : s^{*}_{i} = +1\}|/n$ and $\varsigma^{**}_{+} = |\{i \in \mathcal{N} : s^{**}_{i} = +1\}|/n$, depend on the parameters ν_+ and κ. This is graphically represented in Figure 4, which considers separately the case where either one of these parameters unilaterally changes while the other remains fixed at values $\hat{\nu}_+$ and $\hat{\kappa}$ such that $\hat{\nu}_+ = \frac{1}{2}(1 - \hat{\kappa}) = \frac{3}{2}$. Note that this relationship defines a point in the corresponding parameter space where, according to Proposition 3, there is an abrupt transition

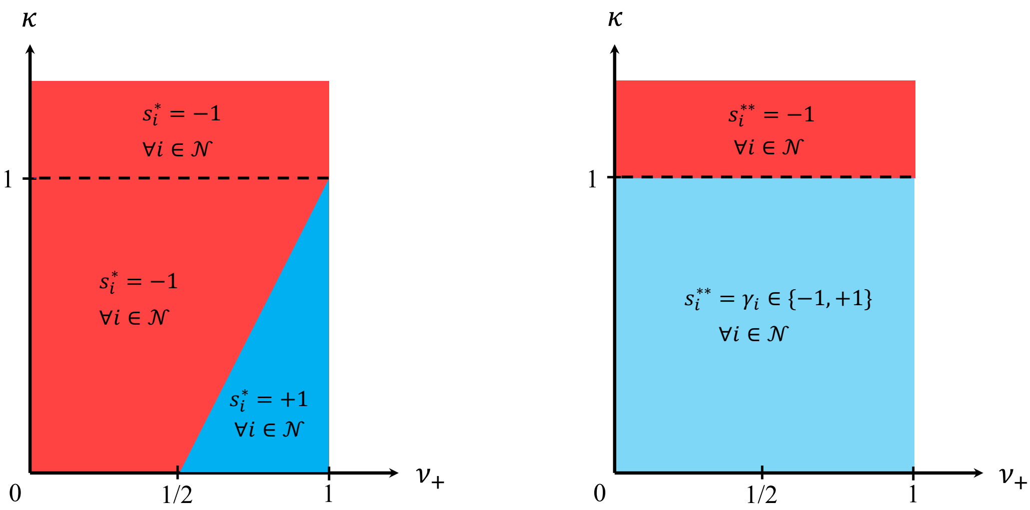

Figure 3: Comparing the action profiles of the states $\boldsymbol{\omega}^*$ and $\boldsymbol{\omega}^{**}$ when $\theta < \zeta_2$, for the GI model (left panel) and the LIL model (right panel), respectively.

under GI across the two extreme population fractions, $\varsigma^* = 0$ and $\varsigma^* = 1$). This transition is sharply seen in the upper two panels of Figure 4, and is also found in the lower panels for the LIL scenario when $\theta > \zeta_2$, and therefore the network G^{**} prevailing in $\boldsymbol{\omega}^{**}$ is complete. Instead, when the opposite inequality $\theta < \zeta_2$ holds and the induced network G^{**} is segmented along agent-type lines, we find that the fraction ς^{**} of the population choosing action $+1$ grows linearly for all values of ν_+, even very small ones. This shows in an especially stark manner the interplay between links and actions that, as already discussed, can arise in the LIL context. As we shall explain shortly, this gives the process a plasticity that has important dynamic implications.

So far we have interpreted the singleton sets Ω^* and Ω^{**} as identifying the unique states that enjoy the highest probability in the unique invariant (or stationary) distribution of the process. But, in line with what is common in the modern evolutionary game-theoretic literature (see, e.g., the seminal papers by Kandori et al. (1993) and Young (1993)), we can also consider the limiting case of vanishing noise (i.e. an arbitrarily large η) and arrive at the following dynamic selection result: in the long run, as the noise becomes progressively small, the GI- and LIL-induced processes will remain almost surely (i.e., with an arbitrarily high probability) in the vicinity of the corresponding aforementioned sets – Ω^* or Ω^{**} respectively – independently of initial conditions. The states included in those sets are customarily labeled the *stochastically stable states* (SSS) of the processes. Naturally, the practical relevance of this notion hinges upon the question of "how long is the long run." We explain next that the answer to this question turns out to be very different behavior in the two scenarios we considered.

On the one hand, for the GI scenario one can readily adapt well-known results in the literature (see, e.g., Ellison (1993)) to claim that the convergence to its long-run predictions is very slow for large populations, with an expected delay that grows exponentially with population size. This suggests that such predictions are problematic descriptions of what can be expected to happen within realistic time scales. For, in particular, it indicates that even when the unique SSS involves all agents joining collective action (i.e., all playing action $+1$, as it happens when $\kappa < 2\nu_+ - 1$), that prediction can be expected to be eventually a good approximation of the situation only within a time frame that is practically irrelevant if the population is moderately large and, say, most agents start choosing -1.

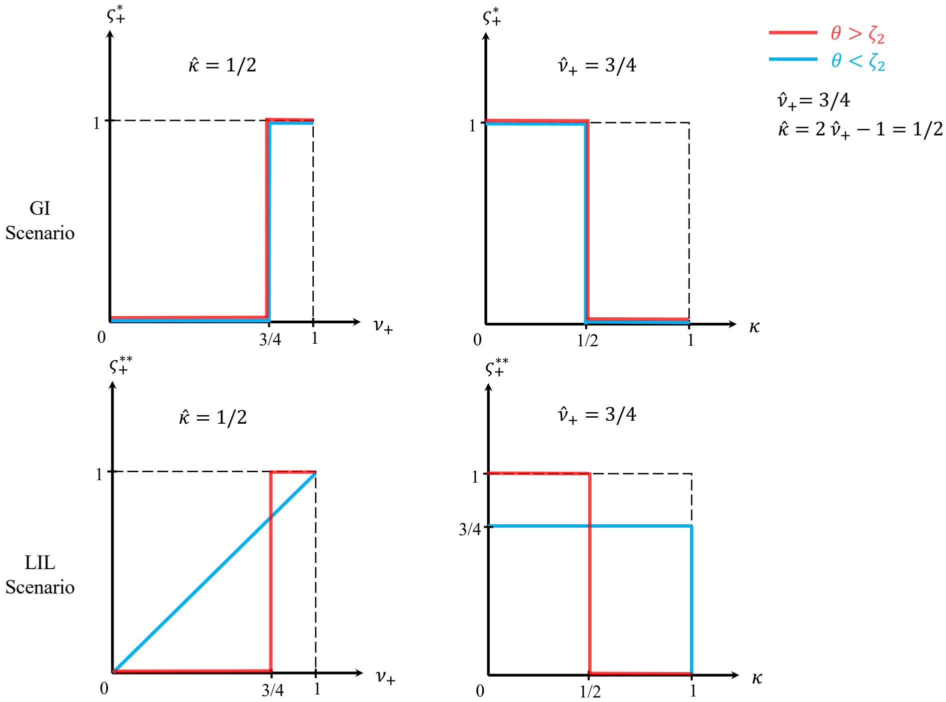

Figure 4: The fractions of the population ς_+^* and ς_+^{**} choosing action +1 in the states $\boldsymbol{\omega}^*$ and $\boldsymbol{\omega}^{**}$, respectively, as a function the parameters ν_+ and κ.

In contrast, the situation is different for the LIL context. The reason for this is well illustrated in Figure 4 – refer, more concretely, to its lower-left panel. There we see that, provided that the cost κ incurred by action +1 is not too high (lower than 1), the SSS of the corresponding process always involves a fraction of agents choosing action +1 if there is that same fraction of agents of the homonymous type. Indeed, this applies no matter how small such a fraction might be. In essence, the reason why those heteromorphic configurations can arise and persist in the long run should be clear: since agents' expectations are formed through local observation and learning, a clique of agents with the required bias/type can continue choosing action +1 – and robustly so even if they are in the minority – if they continue holding expectations that the population as a whole is aligned with them. Their expectation-formation process feeds on what they do and they expect, thus making the configuration stable and quite resistant to noise. This then makes a transition to such a configuration relatively "easy" (or likely) to implement from other states, by just adding *gradually* to some initial (possibly very small) "seed" of individuals choosing action +1 other agents with a bias towards that action.

Mathematically, the logic underlying the aforementioned considerations has been systematically articulated by Ellison (2000) through what he labels "step-by-step evolution" and the notions of *radius* and *modified co-radius*. Intuitively, while the radius captures how much the process needs to rely on noise to abandon the basin of attraction of an otherwise stable state, the modified co-radius reflects an analogous idea but concerning the opposite transition, i.e., a move into the state's basin of attraction. Stochastically stable states are then those for which the modified co-radius (which allows for a concatenation of robust intermediate transitions) is lower than their radius. Two further interesting properties of the modified co-radius of any given stochastically stable state are: (a) it determines the expected waiting time required by the process to arrive at that state from arbitrary initial conditions; (b) when such a transition relies on intermediate steps that are

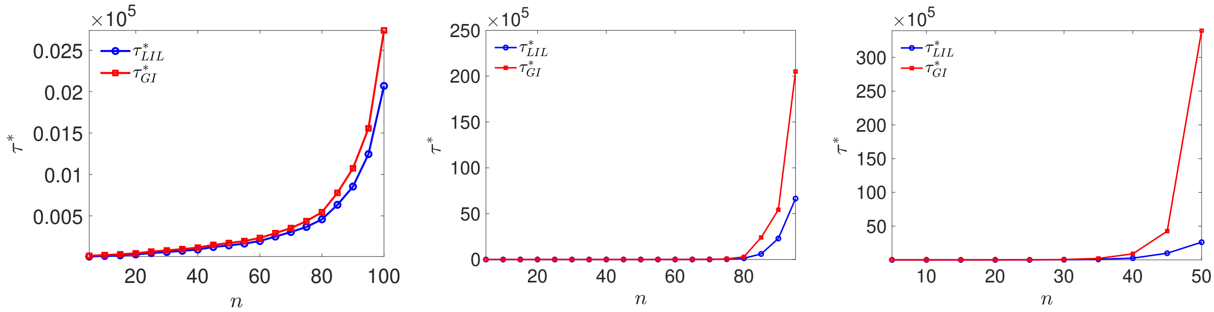

Figure 5: The convergence times, τ^*, to a configuration with at least one-quarter of the population choosing action +1 for different population sizes (n) starting from an empty network, $\overline{K}_n$, where all agents choose action $s_i = -1$ and have idiosyncratic preferences $\gamma_i = +1$ for the GI and the LIL models. The markers represent averages obtained across 1000 Monte Carlo runs under the following parameters: $n = 5, \ldots, 100$, $\eta = 1$, $\lambda = \chi = \xi = 1$, $\kappa = 0.1$, $\varphi = 0.5$, $\theta = 0.02$, $\zeta_1 = 0.01$, $\zeta_2 = 1$, $\rho = 1.5/n$ (left panel), $\rho = 2/n$ (middle panel), and $\rho = 2.5/n$ (right panel).

relatively likely – in particular, enjoy a probability that is *independent of the population size* – the full path is comparably likely as well. This line of reasoning can also be applied to our model and, for the LIL scenario, exploit the fact its network-based procedure of belief formation allows transitions to be implemented through robust intermediate steps, whose probability is independent of population size. This, in the end, implies that its SSS-prediction for the LIL scenario is unaffected by the "large-population curse" and, therefore, the speed at which the process converges to it as the population grows large should be much faster than in the GI context.

Due to space limitations, a formal rigorous analysis of the speed-of-convergence issue just discussed is beyond what is feasible to do in this paper. Instead, next we present some simple numerical simulations that may help illustrate our main point. The exercise involves a large set of simulation runs separately conducted for the adjustment processes operating in either the GI or LIL scenario process. The underlying parameter conditions used in all cases are such that they yield the same (unique) SSS-induced prediction in both scenarios: a state where all agents choose action +1 and they are segmented in type-homogenous cliques. Every simulation run is also conducted from the same initial state where all agents in the population choose action −1 and the social network is empty (i.e., there are no prevailing links). We then ask the question of what is, on average over 1000 independent runs, the delay incurred in reaching a state where one-quarter of the population chooses action +1.

The different panels displayed in Figure 5 show how the answer to the previous question depends on two of the parameters that best embody the tension lying at the core of the collective action problem:

- the weight ρ that modulates the payoff importance of coordinating with the whole population;
- the population size n bearing on the information/communication complexity of coordination.

A marked contrast between the two scenarios arises as the population grows. For, already at a population size as small as $n = 100$, the divergence on their induced delays becomes quite wide when ρ increases. This provides further support to our point that a more realistic model of how expectations are formed in large populations (i.e., through local observation and learning, as in the LIL context) can also shed light on how large populations can address the coordination challenge involved in collective action.

4. Data

The empirical application of our model uses online social network data from Twitter focusing on riots and demonstrations in Egypt during the Arab Spring in 2013 (cf. Borge-Holthoefer et al., 2015). We use Twitter data to study protest behavior in Egypt for two reasons. First, social media data provides fine-grained, micro-level data on individuals' decisions to gather and disseminate political information in countries with authoritarian regimes where traditional media is partly controlled via the state (Steinert-Threlkeld, 2017). As a result, information contained in social media closely reflected the views in the offline world. Second, among social media platforms, Twitter was the main social networking platform for opinion exchange over Egypt's Arab Spring (cf. Clarke and Kocak, 2020). The empirical analysis leverages protest-related tweets surrounding the Second Egyptian Revolution, where the incumbent Islamist president, Mohamed Morsi, is ousted and the military returns to power.[16]

Sample. We collected around 6 million Arabic language tweets based on data from Borge-Holthoefer et al. (2015) and focus on the period between July 4, 2013, when the military overthrew the regime of President Morsi, and August 19, 2013, when the scales of demonstrations and the volumes of tweets reached a saturation point.[17] Our analysis focuses on users who have posted more than two tweets that are either retweets or include mentions of other users, and have fewer than the 90th percentile of followers (2,918 followers). It provides us with a sufficient sample of tweets for each user, enabling us to classify them by their political affiliation and gender using their tweet text. Restricting users based on the number of followers is designed to remove accounts of politicians, celebrities, and the accounts of news and other media sources. The resulting estimation sample consists of 225,578 users.

Network construction. The theoretical model constructed in Section 2 employs the notion of bilateral, undirected, and unweighted links to represent a social network. We construct our counterpart empirical social network to match these features. Bidirectional and undirected links between Twitter users are constructed using retweets of original tweets and the use of @-mentions within tweets. In particular, we define a connection between two users A and B to exist when (i) A has either retweeted or @-mentioned user B, *and* (ii) B has either retweeted or @-mentioned user A. This notion of bidirectional links between users on social media can be thought of as "strong ties" (Shi, 2014).

4.1. From Text to Quantitative Measurement

The remainder of our empirical data is constructed using tools from Natural Language Processing (Ash and Hansen, 2023; Gentzkow et al, 2017). We create three variables from the text of each user's tweets: (i) whether a user is "rioting" (i.e., is the pro- or anti-military intervention), (ii) the political affiliation of each user (Secular vs Islamist), and (ii) the gender of each user. The construction of each variable is treated as an independent binary classification problem where we infer each user's trait from the text of their tweets. The classification models are designed to predict stance, affiliation, and gender at the tweet level. We use majority voting within tweets from an individual to aggregate predictions from multiple tweets to a unique prediction for each individual.

Our approach leverages the latest generation of pre-trained BERT language models (Devlin

[16]A more detailed account of the historical context can be found in Supplementary Appendix F. In particular, our analysis covers Phase IV of the Egyptian Arab Spring discussed in Supplementary Appendix F.2.

[17]Supplementary Appendix G.1 describes the data collection process in more detail. In addition to the date window featuring the largest scale of demonstrations over this phase of Egypt's Arab Spring driving our decision on the time horizon to focus our study rests on when the tweets originally collected by Borge-Holthoefer et al. (2015) identify protest-relevant tweets. Borge-Holthoefer et al. (2015)'s data collection process stopped updating the search terms used to identify and collect protest-related tweets in mid-August 2013. As the hashtags and keywords used by protesters continue to evolve, their collection process no longer includes all relevant tweets. Using information on the number of protests from the Armed Conflict Location and Event Data Project (ACLED, 2019) over the same sampling period, we find a significant concurrence of Twitter message volumes and real-world protests The Pearson correlation coefficient between the two indices, ACLED and Twitter, is positive (0.32). Note that we only have information on the number of protests from ACLED but not their size. This will likely lead to an underestimate of the correlation between the actual protest participation and the volume of protest-related Twitter messages

et al., 2018). The pre-trained model is then "fine-tuned" for each of the three specific classification tasks mentioned above. Since pre-trained language models are built to achieve a general understanding from a diverse set of text inputs, the fine-tuning approach can achieve good performance even when only a small number of training samples are provided (Ash and Hansen, 2023). We create three fine-tuned models using a pre-trained BERT language model developed for Arabic, "AraBERT" (Antoun et al., 2020). AraBERT is the dominant pre-trained language model for the Arabic language and has state-of-the-art performance across a large range of Arabic language tasks (Abdul-Mageed et al, 2021).[18] For each fine-tuning task, we add a feed-forward Softmax classification layer to the top of the AraBERT encoder output. The classifier and the pre-trained model weights from AraBERT are trained jointly during fine-tuning to maximize the log-probability of correct class assignment. Tweets are cleaned through a two-step procedure before being passed across AraBERT: (i) words are segmented using the "Farasa segmenter" (Abdelali et al., 2016) and (ii) fed through a "SentencePiece" tokenizer (Kudo and Richardson, 2018). Each fine-tuned AraBERT model uses a probability threshold of 0.5 when predicting classes. After fine-tuning, we use the model to predict all tweets written by users included in the estimation sample.

Pro- vs. anti-military intervention. We want to identify which users are rioting. However, tweets don't explicitly tell us a user's stance. To achieve this, we build a classifier that identifies a user's pro- vs. anti-military intervention disposition from the text of their tweets. We take this as an indication of an individual's willingness to riot or protest. Throughout our empirical analysis, the action $+1$ is defined as an anti-military intervention stance, and -1 denotes a pro-military intervention stance. As training data, we use 4,150 tweets that are hand-classified into the pro- or anti-military intervention stances by two human coders proficient in Arabic.[19] We then build a binary classifier of tweets (pro- or anti-military intervention) using a fine-tuned AraBERT model. We use 80% of the annotated tweets as training data for fine-tuning and the remaining 20% to validate model performance on unseen data. The out-of-sample performance is shown in the bottom left panel in Figure 6. The action classifier has a high accuracy (representing the number of correctly classified data instances over the total number of instances) with the corresponding confusion matrix used to compute the accuracy shown in the top right panel in Figure 6. Also, the F1-score (the harmonic mean of precision and recall) is high. This shows that our classifier performs well in classifying the users' actions from the tweets' texts.

[18] We use the 'v2' AraBERT model using a maximum sequence size of 128 tokens and a batch size of 16. We set the learning rate to 2e-5 and fine-tune each model for 4 epochs (Tamkin et al., 2020; Zhang et al., 2020). The model from the fourth epoch is used for prediction purposes in all tasks.

[19] To check for cross-coder consistency, 1,000 tweets are the same for both coders. We obtained a consistency measure of Cohen's $\kappa = 0.67$. This indicates a sufficiently strong agreement between the two coders (whereby agreement due to chance is factored out).

a) Predicted

Actual	Anti-MI	Pro-MI
Anti-MI	0.74	0.26
Pro-MI	0.18	0.82

	Anti-MI	Pro-MI	Weighted Avg.
F1-score	0.77	0.79	0.78
Accuracy			0.78

b) Predicted

Actual	Islamist	Secular
Islamist	0.79	0.21
Secular	0.17	0.83

	Islamist	Secular	Weighted Avg.
F1-score	0.76	0.85	0.81
Accuracy			0.81

c) Predicted

Actual	Male	Female
Male	0.83	0.17
Female	0.27	0.73

	Male	Female	Weighted Avg.
F1-score	0.88	0.58	0.83
Accuracy			0.81

Figure 6: The top panels show the classifier confusion matrices for action and political affiliation, while the bottom panels show the corresponding classifier performances (F1-scores and accuracies). The left column a) shows the action classifier (anti-MI vs. pro-MI), the middle column b) the political affiliation classifier (Islamist vs. secular), and the right column c) the gender classifier (male vs. female).

Political affiliation. We also leverage the text of tweets to infer the political ideology of Twitter users. We utilize an already annotated sample of 20,886 Egyptian Twitter users and their political affiliations provided by Weber et al. (2013) to fine-tune an AraBERT model aimed at predicting whether tweets are written by a Twitter user with either Islamist or Secular political views.[20] The matched users are randomly split 80/20 into training/test samples for the AraBERT fine-tuning. All tweets by a user allocated to the test (training) sample are used in the parameter fine-tuning (evaluation). Out-of-sample performance of the political affiliation classifier at the Twitter user level is reported in the middle panel of Figure 6. The high F1-score and accuracy, together with the corresponding confusion matrix shown in the top middle panel in Figure 6 indicate that the classifier succeeds at identifying political affiliation from user tweets. Supplementary Appendix Figure G.1 shows the fraction of Islamist versus secular users over the sample period. Note that in our sample, the majority of agents are Islamists, and that is why the action +1, which is protesting against the military intervention, dominates (around 73%, see also Table 1). An illustration of the users' political affiliations and links between them can be seen in Supplementary Appendix Figure G.2. The figure shows that users are mainly connected with other users with the same political views.

Gender. To classify the gender of Twitter users, we proceed analogously to our approach on political affiliation, fine-tuning a third AraBERT model at the tweet level before aggregating them to a user-level prediction. For training data, we source the gender classification of a subset of Egyptian Twitter users from Weber et al. (2013), which provides gender classifications alongside the political affiliations utilized above.[21] Users are again split 80/20 to form training and test samples. The out-of-sample performance of our gender classifier at the user level is shown in the bottom right panel in Figure 6. High F1-score and accuracy score, along with the confusion matrix (shown in the top right panel in Figure 6) indicate that the classifier performs well at predicting

[20] In practice, we build a training sample by matching the classified users in Weber et al. (2013) to users in our sample Further details about the construction of the training data are available in Supplementary Appendix G.2.

[21] Because tweets about riots might not be sufficiently informative about a user's gender, we augment the protest-relevant tweets with a larger sample of tweets written by these users for fine-tuning. To do so, we extract the most recent 3200 tweets from each of the users in Weber et al.'s data that match our data from the public Twitter API. The API pull was implemented in December 2020. 3200 tweets is the maximum allowed via Twitter's Public API.

Table 1: Summary statistics.

	Mean	Std.	Max	Min
Action	0.48	0.87	1	-1
Female	0.10	0.30	1	0
Islamist	0.62	0.49	1	0
Number of followers	472.71	569.10	2918	1
Network degree	0.83	2.13	47	0
Number of nodes		225,578		
Number of links		93,762		

gender from text.

Table 1 provides summary statistics of the variables used in the empirical analysis. In our sample, 74.24% of users have the action identified as one, and the majority are male and Islamist, with an average of 0.83 connections and 472 followers.

5. Empirical Analysis

5.1. Identification and Estimation Method

The stationary distributions corresponding to the GI and LIL scenarios in (15) and (20) are known as Gibbs measures (or the Gibbs random field; see cf. Wainwright and Jordan, 2008), which provide us with a probability (likelihood) measure for estimating unknown structural parameters in the potential functions (11) and (17) from the empirical data described in Section 4. Before proceeding with estimation, we first reduce the dimensionality of unknown parameters in the model by specifying the idiosyncratic preference γ_i and the linking cost ζ_{ij} in (11) and (17) as functions of observed and unobserved individual characteristics:

$$\gamma_i = \mathbf{x}_i^\top \boldsymbol{\beta} + \tau z_i, \tag{22}$$

and

$$\zeta_{ij} = \phi_0 + \sum_{k=1}^{K} h_k(x_{ik}, x_{jk})\phi_k - z_i - z_j, \tag{23}$$

where $\mathbf{x}_i = (x_{i1}, \ldots, x_{iK})^\top$ is a $K \times 1$ vector of exogenous regressors including individual gender, political affiliation, and log number of Twitter followers, while $\boldsymbol{\beta}$ is the corresponding vector of unknown parameters. The variable z_i represents an individual random effect which captures unobserved individual heterogeneity and is assumed to be identically and independently normally distributed with mean zero and variance σ_z^2, i.e., $z_i \sim \mathcal{N}(0, \sigma_z^2)$. The function $h_k(x_{ik}, x_{jk})$ in ζ_{ij} can be either an indicator function ($\mathbb{1}(x_{ik} = x_{jk})$) when x_{ik} is a dummy variable or a distance function ($|x_{ik} - x_{jk}|$) when x_{ik} is continuous, reflecting homophily (or heterophily). The higher values of random effects z_i and z_j in ζ_{ij} lead to a lower linking cost, capturing the extent of inter-agent heterogeneity due to unobservables (Dzemski, 2019; Graham, 2017). We denote the unknown parameters in ζ_{ij} by $\boldsymbol{\phi} = (\phi_0, \phi_1, \ldots, \phi_K)$.

There are two main identification issues regarding our model specification. First, the noise parameter η in the logistic disturbance ε of (15) and (20) is not separately identifiable from other parameters in the (quasi) potential function, a commonly known problem in discrete choice models. Therefore, following the usual practice in this literature, we set η to one during estimation. Second, in the GI potential given by (11), we capture the global conformity effect ρ through the "leave-one-out" sum of actions, $\sum_{j \neq i}^{n} s_j$. When the sample size n is large, the leave-one-out sum has only a negligible variation at the individual level and therefore the coefficient ρ is hardly disentangled from the rioting cost κ. To deal with this identification problem, we replace the "leave-one-out" sum with the constant $(n-1)\bar{s}$ in (11), where $\bar{s} = \frac{1}{n}\sum_{j=1}^{n} s_j$ is the sample mean of actions, and drop κ for the GI scenario. To mark this modification, we denote the coefficient in front of $(n-1)\bar{s}$ in this modified model by $\check{\rho}$. A similar problem does not appear in the LIL scenario of (17) because the global conformity effect can be identified through variations on the individual belief ψ_i^{LIL}. It is also worth mentioning that identification of individual random effect z_i is made possible by exploiting variations on link decisions, $a_{i1}, a_{i2}, a_{i3}, \cdots$ of agent i in a similar manner as an individual fixed effect in panel data. More specifically, as z_i appears in all link decisions of agent i, we can identify z_i from the conditional probability,

$$\mu(a_{ij} = 1|\mathbf{s}, G_{-ij}) = \frac{\exp(\theta s_i s_j - \phi_0 - \sum_{k=1}^{K} h_k(x_{ik}, x_{jk})\phi_k + z_i + z_j)}{1 + \exp(\theta s_i s_j - \phi_0 - \sum_{k=1}^{K} h_k(x_{ik}, x_{jk})\phi_k + z_i + z_j)}.$$

Given z_i is identified, we can also identify the coefficient τ in γ_i in (22).

The main challenge to the estimation of the Gibbs measures in (15) and (20) comes from the appearance of a normalizing constant in the denominator, which involves the evaluation of the (quasi) potential function over all possible networks $G \in \mathcal{G}^n$ and action profiles $\mathbf{s} \in \{-1, +1\}^n$. When the network size is large, this normalizing constant becomes intractable, and thus directly calculating the likelihood for conventional frequentist or Bayesian estimation is not possible.[22] The most commonly used methods to tackle such estimation challenges include the composite likelihood approach (Lindsay, 1988; Varin et al., 2011),[23] the Monte Carlo simulated likelihood approach (Geyer and Thompson, 1992), the Bayesian exchange algorithm (Møller et al., 2006; Murray et al., 2006) with exact sampling, or the Bayesian double Metropolis-Hastings algorithm (Liang, 2010; Mele, 2017). Given the enormous sample size faced in the present study, the latter simulation-based methods are not feasible, and therefore we adopt the composite likelihood method to estimate our model.[24] The composite likelihood for our model is defined as

$$\mu(\mathbf{s}|G)\mu(G|\mathbf{s}), \tag{24}$$

where $\mu(\mathbf{s}|G)$ and $\mu(G|\mathbf{s})$ represent the conditional likelihood function of the action given the network and the network given the actions, respectively. We first look at $\mu(\mathbf{s}|G)$. Under the GI scenario, it is

$$\mu(\mathbf{s}|G) = \frac{\exp\left(\sum_{i=1}^{n}(\mathbf{x}_i^\top \boldsymbol{\beta} + \tau z_i + \check{\rho}(n-1)\bar{s})s_i + \frac{\theta}{2}\sum_{i=1}^{n}\sum_{j\neq i}^{n} a_{ij} s_i s_j\right)}{\mathcal{Z}(G)}, \tag{25}$$

where $\mathcal{Z}(G) = \sum_{\mathbf{s}\in\{-1,+1\}^n} \exp(\sum_{i=1}^{n}(\mathbf{x}_i^\top \boldsymbol{\beta} + \tau z_i + \check{\rho}(n-1)\bar{s})s_i + \frac{\theta}{2}\sum_{i=1}^{n}\sum_{j\neq i}^{n} a_{ij} s_i s_j)$. Similarly,

---

[22] Note that there are 2^n possible action profiles $\mathbf{s} \in \{-1, +1\}^n$ and $2^{\binom{n}{2}}$ possible networks $G \in \mathcal{G}^n$ with n agents.

[23] In general, the composite likelihood for $y = (y_1, \dots, y_n)$ can be written as $f(y) = \prod_{r=1}^{R} p(y_{A_r}|y_{B_r})$ where observations are divided into R blocks and the block $B_r = \mathcal{N} \setminus A_r$ with $\mathcal{N} = \{1, 2, \cdots, n\}$ denotes the index set of y. The well-known pseudo likelihood proposed by Besag (1975) and Strauss and Ikeda (1990) for the spatial processes refers to a special case where the block A_r contains just a singleton.

[24] Several theoretical results on the asymptotic consistency of the composite likelihood estimation for the Gibbs measure are available in the literature (see e.g., Bhattacharya and Mukherjee, 2018; Chatterjee, 2007; Comets, 1992). Moreover, the simulation results in Varin et al. (2011) and Friel (2012) show that the composite likelihood method permits reliable inference when the sample size is not too small, and the network dependence is moderate. Since our sample size is indeed very large, and the estimated network effects shown in Subsection 5.2 are not particularly large, the composite likelihood method is arguably an adequate estimation method to be used for our analysis.

under the LIL scenario,

$$\tilde{\mu}(\mathbf{s}|G,\boldsymbol{\psi}^{LIL}) = \frac{\exp\left(\sum_{i=1}^{n}(\mathbf{x}_i^\top\boldsymbol{\beta} + \tau z_i + \kappa + \rho(n-1)\psi_i^{LIL})s_i + \frac{\theta}{2}\sum_{i=1}^{n}\sum_{j\neq i}^{n} a_{ij}s_is_j\right)}{\tilde{\mathcal{Z}}(G)}, \tag{26}$$

where $\tilde{\mathcal{Z}}(G) = \sum_{\mathbf{s}\in\{-1,+1\}^n}\exp(\sum_{i=1}^{n}(\mathbf{x}_i^\top\boldsymbol{\beta}+\tau z_i+\kappa+\rho(n-1)\psi_i^{LIL})s_i+\frac{\theta}{2}\sum_{i=1}^{n}\sum_{j\neq i}^{n} a_{ij}s_is_j)$. Since the conditional likelihood functions in (25) and (26) still contain intractable normalizing constants, we also replace them with the composite likelihood function. Considering first the GI instance in (25), the conditional probability of agent i choosing action $s_i = 1$, given all other agents' actions $\mathbf{s}_{-i}$ and network G, is given by

$$\mu(s_i = +1|\mathbf{s}_{-i}, G) = \frac{\exp(\mathbf{x}_i^\top\boldsymbol{\beta} + \tau z_i + \check{\rho}(n-1)\bar{s} + \theta\sum_{j\neq i}^{n} a_{ij}s_j)}{2\cosh(\mathbf{x}_i^\top\boldsymbol{\beta} + \tau z_i + \check{\rho}(n-1)\bar{s} + \theta\sum_{j\neq i}^{n} a_{ij}s_j)}. \tag{27}$$

The conditional probability of agent i choosing action $s_i = -1$ can be similarly calculated. Therefore, the composite likelihood of action profile $\mathbf{s}$, conditional on the network G, is given by

$$\prod_{i=1}^{n}\mu(s_i|\mathbf{s}_{-i}, G) = \prod_{i=1}^{n}\mu(s_i = 1|\mathbf{s}_{-i}, G)^{\frac{1+s_i}{2}}\mu(s_i = -1|\mathbf{s}_{-i}, G)^{\frac{1-s_i}{2}}. \tag{28}$$

In the LIL scenario, a similar approach can be used to compute the composite likelihood $\tilde{\mu}(\mathbf{s}|G,\boldsymbol{\psi}^{LIL})$ in (26):

$$\prod_{i=1}^{n}\tilde{\mu}(s_i|\mathbf{s}_{-i}, G,\boldsymbol{\psi}^{LIL}) = \prod_{i=1}^{n}\tilde{\mu}(s_i = 1|\mathbf{s}_{-i}, G,\boldsymbol{\psi}^{LIL})^{\frac{1+s_i}{2}}\tilde{\mu}(s_i = -1|\mathbf{s}_{-i}, G,\boldsymbol{\psi}^{LIL})^{\frac{1-s_i}{2}}, \tag{29}$$

where

$$\tilde{\mu}(s_i = +1|\mathbf{s}_{-i}, G,\boldsymbol{\psi}^{LIL}) = \frac{\exp(\mathbf{x}_i^\top\boldsymbol{\beta} + \tau z_i + \kappa + \rho(n-1)\psi_i^{LIL} + \theta\sum_{j\neq i}^{n} a_{ij}s_j)}{2\cosh(\mathbf{x}_i^\top\boldsymbol{\beta} + \tau z_i + \kappa + \rho(n-1)\psi_i^{LIL} + \theta\sum_{j\neq i}^{n} a_{ij}s_j)}.$$

Next, we consider the conditional likelihood of network G on a given action profile $\mathbf{s}$:

$$\mu(G|\mathbf{s}) = \prod_{i=1}^{n}\prod_{j>i}^{n}\mu(a_{ij}|\mathbf{s}, G_{-ij}) = \prod_{i=1}^{n}\prod_{j>i}^{n}\frac{\exp(a_{ij}(\theta s_is_j - \phi_0 - \sum_{k=1}^{K}h_k(x_{ik}, x_{jk})\phi_k + z_i + z_j))}{1+\exp(\theta s_is_j - \phi_0 - \sum_{k=1}^{K}h_k(x_{ik}, x_{jk})\phi_k + z_i + z_j)}. \tag{30}$$

Since network links are conditionally (pairwise) independent, (30) shows that the conditional likelihood of the network G can be represented by the product of conditional probabilities for each link (with indicators a_{ij}). Note that even though the network links are conditionally independent given the actions, they are not unconditionally independent due to the interdependence of actions in the presence of the peer effect ($\theta \neq 0$). This is an important feature of our model that distinguishes it from an *inhomogeneous random graph* model (Bollobas et al., 2007). Although computing (30) is feasible, in the case of a large network it can still be very demanding. To further alleviate the computational burden, we adopt the so-called *case-control* approach of Raftery et al. (2012), which allows us to reduce the computational cost associated with (30) from $O(n^2)$ to $O(n)$. To explain the idea, we consider the log-likelihood function based on (30)

$$\ell(G|\mathbf{s}) = \sum_{i=1}^{n} \ell_i(a_{i.}|\mathbf{s}, G_{-i}), \tag{31}$$

where $a_{i.}$ denotes the i^{th} row of matrix $\mathbf{A}$ and $\ell_i(a_{i.}|\mathbf{s}, G_{-i}) \equiv \sum_{j>i}^{n} \ln \mu(a_{ij}|\mathbf{s}, G_{-ij})$. To calculate $\ell_i(a_{i.}|\mathbf{s}, G_{-i})$, it is useful to divide the observations in $a_{i.}$ into the groups of edges ("cases") and non-edges ("control") and perform the following decomposition

$$\begin{aligned}
\ell_i(a_{i,.}|\mathbf{s}, G_{-i}) &= \sum_{j>i}^{n} a_{ij}(\theta s_i s_j - \zeta_{ij}) - \ln(1 + \exp(\theta s_i s_j - \zeta_{ij})) \\
&= \sum_{j>i, a_{ij}=1}^{n} (\theta s_i s_j - \zeta_{ij} - \ln(1 + \exp(\theta s_i s_j - \zeta_{ij}))) + \sum_{j>i, a_{ij}=0}^{n} (-\ln(1 + \exp(\theta s_i s_j - \zeta_{ij}))) \\
&= \ell_{i,1} + \ell_{i,0}.
\end{aligned} \tag{32}$$

In (32), $\ell_{i,1}$ and $\ell_{i,0}$ stand for the log-likelihood from edges and non-edges respectively. When the network links are sparse, the quantity $\ell_{i,0}$ can be viewed as a population total statistics. This population total can be estimated by a random sample of the population,

$$\widetilde{\ell}_{i,0} = \frac{n_{i,0}}{m_{i,0}} \sum_{r=1}^{m_{i,0}} (-\ln(1 + \exp(\theta s_i s_r - \zeta_{ir}))), \tag{33}$$

where $n_{i,0}$ is the total number of zeros in the i^{th} row of the upper triangular part of matrix $\mathbf{A}$, and $m_{i,0}$ is the number of samples selected from zero entries in the i^{th} row of the upper triangular part of the matrix $\mathbf{A}$. $\widetilde{\ell}_{i,0}$ is an unbiased estimator of $\ell_{i,0}$ given the random samples. When the network size is large, we can choose a small $m_{i,0}$ to compute $\widetilde{\ell}_{i,0}$ and thus reduce the amount of computation needed.[25]

An additional computational issue concerning the composite likelihood function of (24) is that one needs to integrate over the random effects $\mathbf{z} = (z_1, \ldots, z_n)$ in order to obtain the likelihood for observed data, i.e., $\mu(\mathbf{s}|G)\mu(G|\mathbf{s}) = \int \mu(\mathbf{s}|G, \mathbf{z})\mu(G|\mathbf{s}, \mathbf{z})f(\mathbf{z})d\mathbf{z}$. The frequentist approach typically uses Gaussian quadratures or Monte Carlo integration to evaluate such a likelihood function. However, given the high-dimensional integration, performing these methods can still be cumbersome. As an alternative approach, Bayesian Markov Chain Monte Carlo (MCMC) estimation has shown to be effective for estimating nonlinear models with random effects (Zeger and Karim, 1991). Thus, in this paper, we apply the Bayesian MCMC approach to estimate the unknown model parameters $\Theta = (\theta, \rho, \boldsymbol{\beta}^\top, \tau, \kappa, \boldsymbol{\phi}, \sigma_z^2)$ and unobserved random effects $\mathbf{z}$ with the posterior distribution $p(\Theta, z|\mathbf{s}, G)$ derived based on the composite conditional likelihoods in (24). The specific details regarding the specifications of prior distributions and the MCMC procedure can be found in Supplementary Appendix H. To assess the finite sample performance of the proposed estimation approach, we employed a Monte Carlo simulation exercise on parameter recovery. Detailed information on the simulation design and the outcomes can be found in Supplementary Appendix I. The results affirm the effectiveness of the proposed estimation approach.

[25] In this study, we set $m_{i,o} = 100 + 5\sum_{j\neq i} a_{ij}$ to obtain the empirical results in Subsection 5.2 and simulation results in Supplementary Appendix I. To check the robustness of the empirical results with respect to the choice of $m_{i,o}$, we have also tried $m_{i,o} = 1000 + 5\sum_{j\neq i} a_{ij}$ and found the results, as shown in Tables J.1 and J.2 of Supplementary Appendix J, are quantitatively unchanged.

Table 2: Estimation results for the global information (GI) scenario.

		with random effects (1)	w/o random effects (2)
Local spillover	(θ)	0.1732***	0.2287***
		(0.0031)	(0.0020)
Global conformity	($\check{\rho}$)	3.09e-6***	3.07e-6***
		(9.16e-8)	(7.76e-8)
Individual preference			
Female	(β_1)	-0.0633***	-0.0636***
		(0.0103)	(0.0072)
Islamist	(β_2)	0.1026***	0.1054***
		(0.0054)	(0.0046)
(Log) Number of followers	(β_3)	0.0117***	0.0088***
		(0.0016)	(0.0015)
Random effect	(τ)	0.0057***	–
		(0.0006)	
Linking cost			
Constant	(ϕ_0)	14.7484***	12.5889***
		(0.0195)	(0.0077)
Same gender	(ϕ_1)	-0.1661***	-0.1908***
		(0.0134)	(0.0082)
Same religiousness	(ϕ_2)	-0.0762***	-0.0033
		(0.0080)	(0.0058)
Diff. in followers count	(ϕ_3)	0.0871***	0.0986***
		(0.0033)	(0.0022)
Variance of random effect	(σ_z^2)	2.1568***	–
		(0.0178)	
Sample size (# of nodes)		225,578	

Notes: For the purpose of identification, we replace $\rho \sum_{j \neq i}^{n} s_j$ with $\check{\rho}(n-1)\bar{s}$ and drop κ in the GI scenario. The parameter estimates reported in this table are the posterior mean and the posterior standard deviation from the Bayesian MCMC sampling procedure. The asterisks ***(**,*) indicate that the 99% (95%, 90%) highest posterior density interval (HDI) of the corresponding draws does not cover zero.

5.2. Estimation Results

We conducted MCMC sampling over 30,000 iterations.[26] After discarding the initial 5,000 iterations for burn-in, we derived the posterior mean and posterior standard deviation from the remaining converged draws, which we use to derive our estimation results. The estimation results of the GI and LIL scenarios are reported in Tables 2 and 3, respectively. In each table, column (1) presents the estimation results with individual random effects – capturing unobserved heterogeneity – and column (2) presents the results without individual random effects. Our first finding is that both the local and global interaction effects measured by θ and ρ are, as expected, positive and significant in both columns. Comparing, however, the two columns we can see that the estimates of the local spillover effect (θ) and other coefficients are biased when failing to control for individual unobserved heterogeneity through the inclusion of random effects. In particular, the estimate of the local spillover effect (θ) in the GI scenario is upward biased by 32%; and in the LIL scenario it is upward biased by 126%; together with a 27% downward bias on the estimate of the global conformity effect in the LIL scenario. This bias stems from individual-specific factors that affect both, rioting behavior and network formation (correlated effects) that cannot be accounted for in the model without random effects (cf. Hsieh et al., 2016). We further analyze the direction of this bias in Supplementary Appendix I with artificial data by comparing the models with and without random effects. The fact that the global conformity effect (ρ) is significant in Table 3 provides a strong motivation for the belief-based formulation process under local information. We therefore use the LIL scenario as the benchmark model for the different counterfactuals analyzed in Section

[26] For the implementation details see Supplementary Appendix H.

6.

Moreover, we obtain an estimate of the weighting parameter φ for the belief updating in (6) equal to 0.0961, which suggests that, in general, the weight that agents put on local average beliefs is more than nine times larger than the weight put on local average actions in updating their own beliefs. Such an importance given to beliefs by our estimates motivates us, in Subsection 6.2, to study a counterfactual in which we explore how the manipulation of beliefs towards a specific action impacts overall rioting behavior.

The results also confirm that various sources of heterogeneity, as captured by the idiosyncratic preference γ_i, play a prominent and intuitive role in rioting decisions. Specifically, we find that females are less likely to support (or possibly attend) riots. On the contrary, popular individuals (who have more followers on Twitter) and Islamists (who are major supporters of Morsi) are more likely to support or attend riots. The individual random effects also show a positive effect, as captured by the estimate of τ. In the LIL scenario, we obtain a negative estimate of κ. This suggests that, in the Egyptian revolt against the military that we study, the population perceived, on average, that the intrinsic costs and risks entailed were more than offset by the corresponding benefits of joining in. Finally, in terms of the linking costs, our estimation results show a high constant cost and a clear homophily pattern in which similar characteristics (e.g., same gender or similar religiousness) lower those costs.

In Table 3, columns (3) and (4) further report estimation results obtained when we omit the local spillover effect and the global conformity effect, respectively. These results show that when one of the two aforementioned effects is omitted, the other effect is confounded, leading to an upward estimation bias. This finding illustrates the importance of controlling for both local and global interaction effects simultaneously, as they both play an indispensable role in determining people's collective actions.

6. Counterfactual Analyses

Building upon the estimation results obtained for the LIL scenario in Subsection 5.2, we can perform various counterfactual analyses in order to investigate the effect of changing specific single parameters of the model, while keeping the remaining ones at their estimated values in column (1) of Table 3. The impact of these changes on the different outcome variables is assessed through the corresponding invariant distribution $\mu_\eta^{LIL}(\cdot)$ in (20).[27] In what follows, we describe our results for two specific implementations. In the first one, we examine the role of linking costs in affecting rioting behavior. In the second, we study how biasing beliefs towards a specific action can influence the extent of rioting.

6.1. Linking Costs and Rioting Behavior

During the protests against the Egyptian government, mobile phone operators were instructed to suspend services in selected areas, with internet access being blocked and mobile phone and text messaging services disabled or working only sporadically (Kravets, 2011). These interventions by the government were aimed at suppressing protests by making it harder for people to communicate and coordinate via online social media. Similar interventions have also been undertaken in other countries (e.g., China), where social media platforms such as Facebook, YouTube, and Twitter were domestically blocked, so that users could use them only indirectly via a VPN service (Willnat

[27] To simulate artificial network and outcome data from $\mu_\eta^{LIL}(\cdot)$, we follow the procedure in Supplementary Appendix I.

Table 3: Estimation results for the local information and learning (LIL) scenario.

		with random effects (1)	w/o random effects (2)	w/o local spillovers (3)	w/o global conformity (4)
Local spillover	(θ)	0.0705***	0.1596***	–	0.2289***
		(0.0035)	(0.0028)		(0.0026)
Global conformity	(ρ)	2.37e-6***	1.72e-6***	3.03e-6***	–
		(5.03e-8)	(3.86e-8)	(3.22e-8)	
Weight of local observation	(φ)	0.0961***	0.0805***	0.1066***	–
		(0.0051)	(0.0041)	(0.0035)	
Individual preference					
Female	(β_1)	-0.0553***	-0.0542***	-0.0564***	-0.0613***
		(0.0096)	(0.0078)	(0.0074)	(0.0081)
Islamist	(β_2)	0.1136***	0.1178***	0.1104***	0.1060***
		(0.0061)	(0.0048)	(0.0053)	(0.0049)
(Log) Number of followers	(β_3)	0.0057***	0.0034**	0.0089***	0.0087***
		(0.0020)	(0.0015)	(0.0016)	(0.0015)
Random effect	(τ)	0.0054***	–	–	–
		(0.0003)			
Rioting cost	(κ)	-0.2965***	-0.3054***	-0.2805***	-0.3348***
		(0.0110)	(0.0089)	(0.0091)	(0.0092)
Linking cost					
Constant	(ϕ_0)	14.7964***	12.5633***	12.5194***	12.5921***
		(0.0229)	(0.0093)	(0.0096)	(0.0103)
Same gender	(ϕ_1)	-0.1708***	-0.1930***	-0.2012***	-0.1934***
		(0.0146)	(0.0094)	(0.0092)	(0.0094)
Same religiousness	(ϕ_2)	-0.0784***	-0.0048	-0.0065	-0.0024
		(0.0095)	(0.0062)	(0.0063)	(0.0066)
Diff. in followers	(ϕ_3)	0.0876***	0.0992***	0.1003***	0.0984***
		(0.0031)	(0.0025)	(0.0026)	(0.0028)
Variance of random effect	(σ_z^2)	2.2477***	–	–	–
		(0.0211)			
Sample size (# of nodes)		225,578			

Notes: The parameter estimates reported in this table are the posterior mean and the posterior standard deviation from the Bayesian MCMC sampling procedure. The asterisks ***(**,*) indicate that the 99% (95%, 90%) highest posterior density interval (HDI) of the corresponding draws does not cover zero.

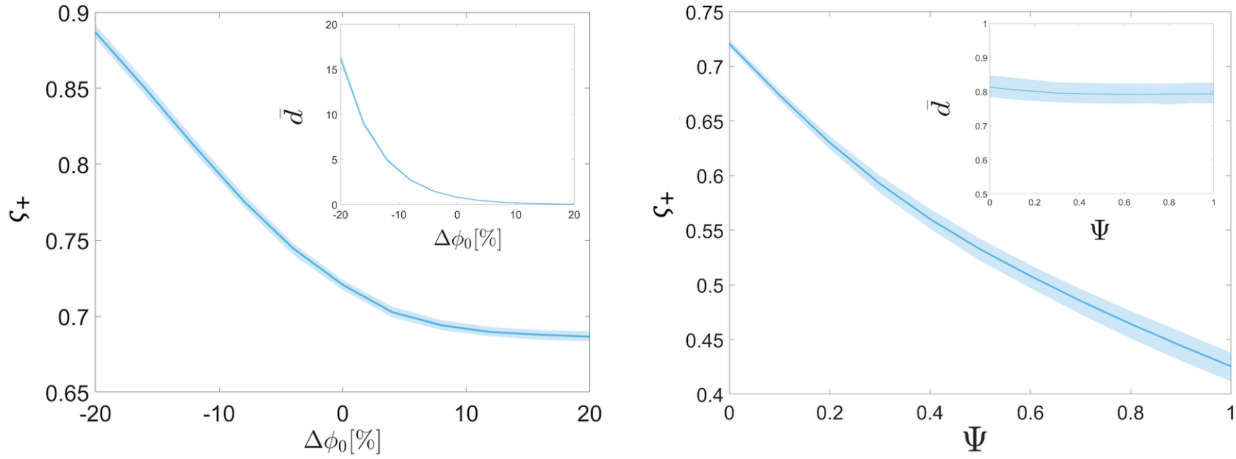

Figure 7: The left panel shows the change in the fraction of rioting agents, $\varsigma_+ = |\{i \in \mathcal{N} : s_i = +1\}|/n$, over percentage changes in the linking cost parameter ϕ_0. The corresponding change in the average degree, $\bar{d}$, is shown in the top-right inset. The right panel shows the change in the fraction of rioting agents, ς_+, over the strength of propaganda, Ψ, with the corresponding change in the average degree, $\bar{d}$, shown in the top-right inset. In all figures we plot the mean and 95% confidence interval over 300 simulation runs.

et al., 2015). We can operationalize such attempts at hampering communication as an increase in the linking cost and explore its effect on rioting behavior, according to our estimated model.

For our counterfactual exercise, we change the value of the estimate of cost parameter ϕ_0, which is the constant term in the linking cost in (23), over the range of [-20%, 20%] with 11 evenly distributed grid points. For each grid point, we simulate the network and action profiles 300 times and compute the average fraction of rioting agents together with the average network degree in the new equilibrium. The inset in the left panel in Figure 7 shows the average degree, $\bar{d}$, over varying ϕ_0. As expected, we can see that the linking cost ϕ_0 has a substantial effect on the average degree so lowering the linking cost gives rise to a much denser network structure. The left panel in Figure 7 shows the fraction of rioting agents, $\varsigma_+ = |\{i \in \mathcal{N} : s_i = +1\}|/n$, over varying ϕ_0. We observe that a reduction of ϕ_0 by 20% yields an increase in the fraction of rioting agents by 15%. Conversely, this indicates that as linking and exchanging information via the network becomes more costly (e.g., by interrupting or blocking social media), fewer links are being formed, coordination among agents becomes more difficult, and fewer agents participate in the protest as a consequence. This finding illustrates and quantifies the importance of the role of online social networks in the formation of protest movements or riots and the emergence of collective action.

6.2. Belief Manipulation and Rioting Behavior

In our second counterfactual analysis, we examine the effectiveness of manipulating the beliefs of the agents to mitigate rioting behavior. Governments often use manipulation of the information available on social or other media to distort the users' view (cf. Edmond, 2013; Zhuravskaya et al., 2020). For example, King et al. (2017) document the massive effort of the Chinese government to post content on social media that is mainly devoted to supporting positive views about the state. Similar efforts have been documented in Egypt (El-Khalili, 2013).

We introduce a government influencing the belief updating in (6) as follows:

$$p_{it}^{u+1} = (1-\Psi)\left\{\varphi \frac{1}{d_{it}}\sum_{j=1}^{n} a_{ij,t}s_{jt} + (1-\varphi)\frac{1}{d_{it}+1}\left[p_{it}^{u} + \sum_{j=1}^{n} a_{ij,t}p_{jt}^{u}\right]\right\} + \Psi g, \tag{34}$$

where $\Psi \in [0, 1]$ and $g = -1$ is supposed to be the preferred action of the government (no rioting). As in the previous section, for each of the points in an evenly distributed grid for Ψ in the range of [0,1], we simulate repeatedly the evolutionary process and then compute the average fraction of rioting agents, $\varsigma_+ = |\{i \in \mathcal{N} : s_i = +1\}|/n$, and the average network degree, $\bar{d}$, resulting from the induced invariant distributions. The corresponding changes in those magnitudes for varying levels of the influence of propaganda, Ψ, are plotted in the right panel in Figure 7. Our results show that while propaganda (belief manipulation) does not affect the network density – which remains roughly stable – it has a significant effect on rioting, reducing the fraction of rioting agents by up to 30%. It does not succeed, however, in reducing rioting below 40% of the population, even when propaganda is the only source of belief updating (i.e., when $\Psi = 1$). In line with the recent literature on the phenomenon (cf. Azzimonti and Fernandes, 2022; Gu et al., 2017), these findings illustrate the effect that the manipulation of information may have on the formation of collective action, but also point to its limitations.

7. Conclusion

In this paper we have developed a dynamic model of collective action in a large population where agents interact and learn through a co-evolving social network. The model describes a large population of agents on an evolving network. Agents choose between the status quo or joining a collective action – such as a protest – based on their idiosyncratic preferences, the observation of actions chosen by their network neighbors, and their beliefs over the average decision chosen by the population as a whole. We provide a complete characterization of the equilibrium action choices, beliefs, and networks, and identify the conditions under which a significant degree of collective action arises in the long-run equilibrium. We also show such conditions are substantially affected by the assumptions made on the information the agents have about the state of the system. In particular, we consider two cases: one where agents have perfect information on the average action of all others; and an alternative, more realistic situation, where agents form beliefs over the average action via local information, whereby they combine information on the action choices and beliefs of their neighbors in the network. Intriguingly, our findings reveal that the conditions required for substantial collective action are stronger and less plausible when we assume perfect information.

We then structurally estimate the model's parameters using a large-scale Twitter dataset. The dataset contains information about the network connections and the text of posts written by Egyptian Twitter users during a period of social unrest in the Arab Spring. Our empirical results show that: (i) both local peer effects and the global preference towards conformity have significant effects on agents' decisions; (ii) ignoring endogeneity in the network formation process driven by unobserved individual heterogeneity causes substantial bias in the estimates of these two effects; and (iii) local information and learning via neighbors in the network play a significant role in belief formation.

Finally, we use counterfactual simulations that utilize the structural estimates to quantify the role of linking costs between agents and the role of biased beliefs on the extent of rioting. First, we consider an increase in the fixed cost of forming links. This can originate from real world interventions such as a government interrupting and/or blocking social media. Our results show that a 20% increase in the cost of forming links decreases the fraction of rioting agents by 15%. The mechanism at work here is that as linking costs rise, agents form fewer links which, in turn, makes coordination among agents more difficult. Second, we consider attempts to manipulate the beliefs of the agents. We find that this can have a significant effect on rioting, with a reduction in rioting behavior of up to 40%. However, the effort to manipulate beliefs falls short of completely eliminating rioting, indicating that government attempts to use information manipulation to suppress protests against their rule may only have limited impact.

References

Abdelali, A., Darwish, K., Durrani, N., and Mubarak, H. (2016). Farasa: A fast and furious segmenter for Arabic. In *Proceedings of the 2016 conference of the North American chapter of the association for computational linguistics: Demonstrations*, 11-16.

Acemoglu, D., K. Bimpikis,, and A. Ozdaglar (2014). Dynamics of information exchange in endogenous social networks. *Theoretical Economics*, 9(1):41–97.

Acemoglu, D., S. Johnson, A. Kermani, J. Kwak, and T. Mitton (2016). The value of connections in turbulent times: Evidence from the United States. *Journal of Financial Economics* 121(2):368 – 91.

Acemoglu, D., T. Hassan, A. Tahoun (2018). The power of the street: Evidence from Egypt's Arab Spring. *Review of Financial Studies* 31(1):1 – 42.

ACLED (2019). Armed conflict location and event dataset (ACLED). *Center for the Study of Civil War, International Peace Research Institute, Oslo (PRIO).*

Antoun, W., F. Baly, and H. Hajj (2020). AraBERT: Transformer-based model for Arabic language understanding. arXiv preprint arXiv:2003.00104.

Ash, E. and Hansen, S. Text algorithms in economics. *Annual Review of Economics* 15:659–688.

Azzimonti, M. and Fernandes, M. (2022). Social media networks, fake news, and polarization. *European Journal of Political Economy*, 102256.

Badev, A. (2021). Nash equilibria on (un)stable networks. *Econometrica*, 89(3):1179-1206.

Barberà, S. and Jackson, M. O. (2020). A model of protests, revolution, and information. *Quarterly Journal of Political Science*, 15(3):297–335.

Berger, R. L. (1981). A necessary and sufficient condition for reaching a consensus using DeGroot's method. *Journal of the American Statistical Association*, 76:415–418.

Besag, J. (1975). Statistical analysis of non-lattice data. *Journal of the Royal Statistical Society: Series D (The Statistician)*, 24(3):179–195.

Bhattacharya, B. and Mukherjee, S. (2018) Inference in Ising models. *Bernoulli*, 24(1):493–525.

Blume, L. (1993). The statistical mechanics of strategic interaction. *Games and Economic Behavior*, 5(3):387–424.

Blume, L., Brock, W., Durlauf, S., and Ioannides, Y. (2011). Identification of social interactions, Volume 1B of *Handbook of Social Economics*, Chapter 18, 853–964. Elsevier BV, North-Holland.

Bollobás, B. and Janson, S. and Riordan, O. (2007). The phase transition in inhomogeneous random graphs. *Random Structures & Algorithms.*, 31(1):3–122.

Borge-Holthoefer, J., Magdy, W., Darwish, K., and Weber, I. (2015). Content and network dynamics behind Egyptian political polarization on Twitter. In *Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing*, 700–711. ACM.

Boucher, V. (2016). Conformism and self-selection in social networks. *Journal of Public Economics*, 136: 30–44.

Brock, W. and Durlauf, S. (2001). Discrete choice with social interactions. *The Review of Economic Studies*, 68(2):235–260.

Bueno de Mesquita, E. (2010). Regime change and revolutionary entrepreneurs. *American Political Science Review*, 104(3):446–466.

Cantoni D., D. Y. Yang, and N. Yuchtman. Are protests games of strategic complements or substitutes? Experimental evidence from Hong Kong's democracy movement. *Quarterly Journal of Economics*, 134(2):1021–1077.

Cardoso, A., M.-C. Boudreau, J. Á. Carvalho. Organizing collective action: Does information and communication technology matter? *Information and Organization*, 29(3):100256.

Chandrasekhar, A. G., H. Larreguy, J. P. and Xandri (2020). Testing models of social learning on networks: Evidence from a lab experiment in the field. *Econometrica*, 88(1):1–32.

Chatterjee, S. (2007). Estimation in spin glasses: A first step. *The Annals of Statistics*, 35(5):1931–1946.

Chwe, M. S.-Y. (2000). Communication and coordination in social networks. *The Review of Economic Studies*, 67(1):1–16.

Clarke, K., and Kocak, K. (2020). Launching revolution: Social media and the Egyptian uprising's first movers. *British Journal of Political Science*, 50(3):1025–1045.

Comets, F. (1992). On consistency of a class of estimators for exponential families of Markov random fields on the lattice. *The Annals of Statistics*, 20(1):455–468.

Currarini, S., Jackson, M., and Pin, P. (2009). An economic model of friendship: Homophily,

minorities and segregation. *Econometrica*, 77(4):1003–1045.
DeGroot, M. H. (1974). Reaching a consensus. *Journal of the American Statistical Association*, 69(345):118–121.
DeMarzo, P., Vayanos, D., and Zwiebel, J. (2003). Persuasion bias, social influence, and unidimensional opinions. *Quarterly Journal of Economics*, 118(3):909–968.
Devlin, J., Chang, M.W., Lee, K., and Toutanova, K. (2018). BERT: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint*, arXiv:1810.04805.
Dzemski, A. (2019). An empirical model of dyadic link formation in a network with unobserved heterogeneity. *The Review of Economics and Statistics*, 101(5):763 – 776.
Edmond, C. (2013). Information manipulation, coordination, and regime change. *The Review of Economic Studies*, 80(4):1422–1458.
Ellison, G. (1993). Learning, local interaction and coordination. *Econometrica* 61(5):1047–1071.
Ellison, G. (2000). Basins of attraction, long-run stochastic stability, and the speed of step-by-step evolution. *Review of Economic Studies* 67(1):17–45.
Enikolopov, R., Makarin, A., Petrova, M., et al. (2020). Social media and protest participation: Evidence from Russia. *Econometrica*, 88(4):1479–1514.
Evans, J.A. and P. Aceves (2016). Machine translation: Mining text for social theory. *Annual Review of Sociology*, 42:21–50.
Friel, N. (2012). Bayesian inference for Gibbs random fields using composite likelihoods. in *Proceedings of the 2012 Winter Simulation Conference (WSC)*, 1–8.
Gentzkow, M., B. T. Kelly, and M. Taddy (2017). Text as data. *Journal of Economic Literature*, 57(3):535–74.
Geyer, C. J. and Thompson, E. A. (1992). Constrained Monte Carlo maximum likelihood for dependent data *Journal of the Royal Statistical Society: Series B (Methodological)*, 54(3):657–683.
Goldsmith-Pinkham, P. and G.W. Imbens (2013). Social networks and the identification of peer effects. *Journal of Business & Economic Statistics*, 31(3):253–264.
Golub, B. and Jackson, M. (2012). How homophily affects the speed of learning and best-response dynamics. *The Quarterly Journal of Economics*, 127(3):1287–1338.
Golub, B. and E. Sadler (2016). Learning in social networks. In Bramoullé, Y., Galeotti, A., and Rogers, B. (eds.), *Oxford Handbook of Economic Networks*, Oxford University Press.
González-Bailón, S., Borge-Holthoefer, J., Rivero, A., and Moreno, Y. (2011). The dynamics of protest recruitment through an online network. *Scientific Reports*, 1(1):1–7.
Goyal, S. and Vega-Redondo, F. (2005). Network formation and social coordination. *Games and Economic Behavior*, 50(2):178–207.
Graham, B. S. (2017). An econometric model of network formation with degree heterogeneity. *Econometrica*, 85(4):1033–1063.
Granovetter, M. (1978). Threshold models of collective behavior. *American Journal of Sociology*, 83(6):1420–1443
Grimm, V. and Mengel, F. (2015). An experiment on belief formation in networks. *University of Essex, Working Paper.*
Grimmett, G. (2010). *Probability on graphs.* Cambridge University Press.
Gu, L., Kropotov, V., and Yarochkin, F. (2017). The fake news machine. How propagandists abuse the Internet and manipulate the public. Pobrane. Trend Micro Incorporated.
Hsieh, C.-S. and L.F. Lee (2016). A social interactions model with endogenous friendship formation and selectivity. *Journal of Applied Econometrics*, 31(2):301–319.
Hsieh, C.-S., Lee, L.-f., and Vincent, B. (2020). Specification and estimation of network formation and network interaction models with the exponential probability distribution. *Quantitative Economics*, 11(4):1349–1390.
Hsieh, C.-S., König, M. D. and Liu, X. (2022) A structural model for the coevolution of networks and behavior. *The Review of Economics and Statistics*, 104(2):355–367.
Jackson, M. and Golub, B. (2010). Naive learning in social networks: Convergence, influence and wisdom of crowds. *American Economic Journal: Microeconomics*, 2(1):112–149.
Jackson, M. and A. Watts (2002). On the formation of interaction networks in social coordination games. *Games and Economic Behavior*, 41(2):265–291.
Jackson, M. and Rogers, B. (2007). Meeting strangers and friends of friends: How random are social networks? *American Economic Review*, 97:890–915.

Jackson, M. and Zenou, Y. (2015). Games on networks. *Handbook of Game Theory with Economic Applications*, Elsevier, 4:95–163.
Johnsson, I. and H. Moon (2021). Estimation of peer effects in endogenous social networks: Control function approach. *The Review of Economics and Statistics*, 103(2):328–345.
Kandori, M., Mailath, G., and Rob, R. (1993). Learning, mutation, and long run equilibria in games. *Econometrica*, 61(1):29–56.
El-Khalili, S. (2013). *Social media as a government propaganda tool in post-revolutionary Egypt.* First Monday.
King, G. and Pan, J. and Roberts, M. E. (2017). How the Chinese government fabricates social media posts for strategic distraction, not engaged argument. *American Political Science Review*, 111(3):484–501.
Kravets, D. (2011). Internet down in Egypt, tens of thousands protest in "Friday of Wrath". Wired.com ThreatLevel Blog, URL: https://www.wired.com/2011/01/egypt-internet-down/.
Kudo, K., and Richardson J. (2018). Sentencepiece: A simple and language independent subword tokenizer and detokenizer for neural text processing. *arXiv preprint*, arXiv:1808.06226.
Lazarsfeld, P.F. and R.K. Merton (1954). Friendship as a social process: A substantive and methodological analysis. In M. Berger (ed.), *Freedom and Control in Modern Society*, New York: Van Nostrand.
Liang, F. (2010). A double Metropolis-Hastings sampler for spatial models with intractable normalizing constants. *Journal of Statistical Computation and Simulation*, 80(9):1007–1022.
Lindsay, B. (1988). Composite likelihood method. *Contemporary Mathematics*, 80(1):221–239.
Lohmann, S. (2000). Collective action cascades: An informational rationale for the power in numbers. *Journal of Economic Surveys*, 14(5):655-684.
Abdul-Mageed, M. and Elmadany, A, and Nagoudi, E.M.B., (2022). ARBERT & MARBERT: Deep bidirectional transformers for Arabic. *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing* , 7088–7105.
McPherson, M., Smith-Lovin, L., and Cook, J. M. (2001). Birds of a feather: Homophily in social networks. *Annual Review of Sociology*, 27:415–444
Mele, A. (2017). A structural model of dense network formation. *Econometrica*, 85(3):825–850.
Møller, J and Pettitt, A. N. and Reeves, R. and Berthelsen, K. K. (2006). An efficient Markov chain Monte Carlo method for distributions with intractable normalising constants. *Biometrika*, 93(2):451–458.
Morris, S. (2000). Contagion. *The Review of Economic Studies*, 67(1):57–78.
Murray, I. and Ghahramani, Z. and MacKay, D. J. C. (2006). MCMC for doubly-intractable distributions. In *Proceedings of the Twenty-Second Conference on Uncertainty in Artificial Intelligence*, 359–366. AUAI Press.
Norris, J. R. (1998). *Markov chains.* Cambridge University Press.
Pauli, F., Racugno, W., and Ventura, L. (2011). Bayesian composite marginal likelihoods. *Statistica Sinica*, 149–164.
Qin, B., Strömberg, D., and Wu, Y. (2021). Social media and collective action in China. *CEPR discussion paper no. dp16731*
Qiu, M., Sim., Y., Smith, N.A., Jiang, J. (2015). Modelling user arguments, interactions and attributes for stance prediction in online debate forums. *Proceedings of the 2015 SIAM International Conference on Data Mining*, 855–863.
Raftery, A. E., Niu, X., Hoff, P. D., and Yeung, K. Y. (2012). Fast inference for the latent space network model using a case-control approximate likelihood. *Journal of Computational and Graphical Statistics*, 21(4):901–919.
Sandholm, W. (2010). *Population games and evolutionary dynamics.* MIT Press.
Samuelson, L. (1998). *Evolutionary games and equilibrium selection.* MIT Press.
Shadmehr, M. (2021). Tullock's Paradox, Hong Kong experiment, and the strength of weak states. *Quarterly Journal of Political Science*, 16(3):245–264.
Shi, Z., Rui, H. and Whinston, A.B. (2014). Content sharing in a social broadcasting environment. *MIS Quarterly*, 38:123–142.
Steinert-Threlkel, Z., Mocanu, D., Vespignani, A., and Fowler, J. (2015). Online social networks and offline protest *EPJ Data Science*, 4(1):1–9.
Steinert-Threlkel, Z. (2017). Spontaneous collective action: Peripheral mobilization during the Arab Spring *American Political Science Review*, 111(2):379–403.

Strauss, D. and Ikeda, M. (1990). Pseudolikelihood estimation for social networks. *Journal of the American Statistical Association*, 85(409):204–212.
Tamkin, A, Singh, T., Giovanardim D., and Goodman., N. (2020). *Investigating transferability in pretrained language models.* arXiv preprint arXiv:2004.14975.
Tullock, G. (1971). The paradox of evolution. *Public Choice* 11, 89–99.
Tullock, G. (1980) Efficient rent seeking. In J. M. Buchanan, R. D. Tollison, and G. Tullock (Eds.), *Toward a Theory of Rent Seeking Society*, Texas A&M University Press, 3–15.
Varin, C., Reid, N. and Firth, D. (2011). An overview of composite likelihood methods. *Statistica Sinica*, 21(1):5–42.
Vega-Redondo, F. (1996). *Evolution, games, and economic behaviour.* Oxford University Press.
Wainwright, M. J. and Jordan, M. I. (2008). Graphical models, exponential families, and variational inference. *Foundations and Trends in Machine Learning*, 1(1-2):1–305.
Weber, I., Garimella, V. R. K., amd Batayneh, A. (2013). Secular vs. islamist polarization in Egypt on Twitter. In *Proceedings of the 2013 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining*, 290–97.
Willnat, L., Wei, L., and Jason A. M. (2015). *Politics and social media in China*, In *Routledge Handbook of Chinese Media*, 199–220.
Young, H. P. (1993). The evolution of conventions. *Econometrica*, 61(1):57–84.
Young, H. P. (1998). *Individual strategy and social structure: An evolutionary theory of institutions.* Princeton University Press.
Zhang, T., Wu, F., Katiyar, A., Weinberger, K.Q., and Artzi, Y. (2020). Revisiting few-sample BERT fine-tuning. arXiv preprint arXiv:2006.05987.
Zeger, S. L. and Karim, M. R. (1991). Generalized linear models with random effects: A Gibbs sampling approach. *Journal of the American Statistical Association*, 86(413):79–86.
Zhuravskaya, E., Petrova, M. and Enikolopov, R. (2020). Political effects of the internet and social media. *Annual Review of Economics*, 12:415–38.

Appendix

A. Proofs

Proof of Proposition 1. First note that, for any vector of beliefs $\boldsymbol{\psi}^{GI} \in [-1,1]^n$, any state $\boldsymbol{\omega} = (\mathbf{s}, G)$, any pair of agents $i, j \in \mathcal{N}$, and action choice $s_i' \in S_i$ by agent i, the following equalities hold: $\Phi(s_i', \mathbf{s}_{-i}, G, \boldsymbol{\psi}^{GI}(s_i', \mathbf{s}_{-i}, G)) - \Phi(\mathbf{s}, G, \boldsymbol{\psi}^{GI}(\mathbf{s}, G)) = \gamma_i(s_i' - s_i) + \theta(s_i' - s_i)\sum_{j=1}^n a_{ij}s_j + \rho(s_i' - s_i)\sum_{j\neq i}^n p_i - \kappa(s_i' - s_i) = \pi_i(s_i', \mathbf{s}_{-i}, G; \psi_i^{GI}(\mathbf{s}, G)) - \pi_i(\mathbf{s}, G; \psi_i^{GI}(\mathbf{s}, G))$ and $\Phi(\mathbf{s}, G \pm ij, \boldsymbol{\psi}^{GI}(\mathbf{s}, G \pm ij)) - \Phi(\mathbf{s}, G, \boldsymbol{\psi}^{GI}(\mathbf{s}, G)) = \pm(\theta s_i s_j - \zeta_{ij}) = \pi_i(\mathbf{s}, G \pm ij; \psi_i^{GI}(\mathbf{s}, G)) - \pi_i(\mathbf{s}, G; \psi_i^{GI}(\mathbf{s}, G)) = \pi_j(\mathbf{s}, G \pm ij; \psi_j^{GI}(\mathbf{s}, G)) - \pi_j(\mathbf{s}, G; \psi_j^{GI}(\mathbf{s}, G))$, which confirms (14), as desired. □

Proof of Proposition 2. Define the triple $(\Omega, \mathcal{F}, \mathbb{P})$ to be the probability space over sample paths representing our process (where Ω is the state space and $\mathcal{F}$ the suitable smallest σ-algebra). Since, in our case, the process is Markov, we start by introducing the one-step transition matrix $\mathbf{P}(t) : \Omega^2 \to [0,1]$ specifying the probability of a transition from a state $\boldsymbol{\omega} \in \Omega$ prevailing at t to a state $\boldsymbol{\omega}' \in \Omega$ after some small time interval of length Δt. If $\boldsymbol{\omega}' \neq \boldsymbol{\omega}$, this probability is given by $\mathbb{P}(\boldsymbol{\omega}_{t+\Delta t} = \boldsymbol{\omega}' | \boldsymbol{\omega}_t = \boldsymbol{\omega}) = q(\boldsymbol{\omega}, \boldsymbol{\omega}')\Delta t + o(\Delta t)$, where $q(\boldsymbol{\omega}, \boldsymbol{\omega}')$ is the transition rate from state $\boldsymbol{\omega}$ to state $\boldsymbol{\omega}'$. In our case, since the Markov process is time-homogeneous, the transition-rate matrix (or infinitesimal generator) $\mathbf{Q} = (q(\boldsymbol{\omega}, \boldsymbol{\omega}'))_{\boldsymbol{\omega}, \boldsymbol{\omega}' \in \Omega}$ is independent of time. Given the postulated adjustment rules, it has the following form:

$$q(\boldsymbol{\omega},\boldsymbol{\omega}')=\begin{cases}\chi\frac{e^{\eta\Phi(s_i,\mathbf{s}_{-i},G)}}{e^{\eta\Phi(s_i,\mathbf{s}_{-i},G)}+e^{\eta\Phi(s_i',\mathbf{s}_{-i},G)}} & \text{if } \boldsymbol{\omega}'=(s_i',\mathbf{s}_{-i},G) \text{ and } \boldsymbol{\omega}=(\mathbf{s},G),\\ \lambda\frac{e^{\eta\Phi(\mathbf{s},G+ij)}}{e^{\eta\Phi(\mathbf{s},G+ij)}+e^{\eta\Phi(\mathbf{s},G)}} & \text{if } \boldsymbol{\omega}'=(\mathbf{s},G+ij) \text{ and } \boldsymbol{\omega}=(\mathbf{s},G),\\ \lambda\frac{e^{\eta\Phi(\mathbf{s},G-ij)}}{e^{\eta\Phi(\mathbf{s},G-ij)}+e^{\eta\Phi(\mathbf{s},G)}} & \text{if } \boldsymbol{\omega}'=(\mathbf{s},G-ij) \text{ and } \boldsymbol{\omega}=(\mathbf{s},G),\\ -\sum_{\boldsymbol{\omega}'\neq\boldsymbol{\omega}}q(\boldsymbol{\omega},\boldsymbol{\omega}') & \text{if } \boldsymbol{\omega}'=\boldsymbol{\omega},\\ 0 & \text{otherwise.}\end{cases} \tag{A.1}$$

where with have denoted by $\Phi(\boldsymbol{\omega})$ for $\Phi(\boldsymbol{\omega},\boldsymbol{\psi}^{GI}(\boldsymbol{\omega}))$ to simplify the notation. The matrix $\mathbf{Q}$ satisfies the Chapman-Kolmogorov forward equation $\frac{d}{dt}\mathbf{P}(t)=\mathbf{P}(t)\mathbf{Q}$ and therefore we can write $\mathbf{P}(t)=\mathbf{I}+\mathbf{Q}\Delta t+o(\Delta t)$. Furthermore, the stationary distribution $\mu_\eta^{GI}:\Omega\to[0,1]$ is then the solution to $\boldsymbol{\mu}^\eta\mathbf{P}=\boldsymbol{\mu}^\eta$ and can be equivalently computed as $\boldsymbol{\mu}^\eta\mathbf{Q}=\mathbf{0}$ (cf. Norris, 1998).

Note that the embedded discrete-time Markov chain is irreducible and aperiodic, and thus is ergodic and has a unique stationary distribution. Hence, also the continuous-time Markov chain is ergodic and has a unique stationary distribution. The stationary distribution solves $\boldsymbol{\mu}^\eta\mathbf{Q}=\mathbf{0}$ with the transition rates matrix $\mathbf{Q}=(q(\boldsymbol{\omega},\boldsymbol{\omega}'))_{\boldsymbol{\omega},\boldsymbol{\omega}'\in\Omega}$ of (A.1). This equation is satisfied when the probability distribution $\mu_\eta^{GI}(\boldsymbol{\omega})$ satisfies the detailed balance condition (cf. Norris, 1998)

$$\mu_\eta^{GI}(\boldsymbol{\omega})q(\boldsymbol{\omega},\boldsymbol{\omega}')=\mu_\eta^{GI}(\boldsymbol{\omega}')q(\boldsymbol{\omega}',\boldsymbol{\omega}), \tag{A.2}$$

for all $\boldsymbol{\omega},\boldsymbol{\omega}'\in\Omega$. Observe that the detailed balance condition is trivially satisfied if $\boldsymbol{\omega}'$ and $\boldsymbol{\omega}$ differ in more than one link or more than one action level. Hence, we consider only the case of link creation $G'=G+ij$ (and removal $G'=G-ij$) or an adjustment in action $s_i'\neq s_i$ for some $i\in\mathcal{N}$. For the case of link creation with a transition from $\boldsymbol{\omega}=(\mathbf{s},G)$ to $\boldsymbol{\omega}'=(\mathbf{s},G+ij)$ we can write the detailed balance condition as follows

$$\frac{1}{\mathcal{Z}^\eta}e^{\eta\Phi(\mathbf{s},G)}\frac{e^{\eta\Phi(\mathbf{s},G+ij)}}{e^{\eta\Phi(\mathbf{s},G+ij)}+e^{\eta\Phi(\mathbf{s},G)}}\lambda=\frac{1}{\mathcal{Z}^\eta}e^{\eta\Phi(\mathbf{s},G+ij)}\frac{e^{\eta\Phi(\mathbf{s},G)}}{e^{\eta\Phi(\mathbf{s},G)}+e^{\eta\Phi(\mathbf{s},G+ij)}}\lambda.$$

This equality is trivially satisfied. A similar argument holds for the removal of a link with a transition from $\boldsymbol{\omega}=(\mathbf{s},G)$ to $\boldsymbol{\omega}'=(\mathbf{s},G-ij)$ where the detailed balance condition reads

$$\frac{1}{\mathcal{Z}^\eta}e^{\eta\Phi(\mathbf{s},G)}\frac{e^{\eta\Phi(\mathbf{s},G-ij)}}{e^{\eta\Phi(\mathbf{s},G-ij)}+e^{\eta\Phi(\mathbf{s},G)}}\lambda=\frac{1}{\mathcal{Z}^\eta}e^{\eta\Phi(\mathbf{s},G-ij)}\frac{e^{\eta\Phi(\mathbf{s},G)}}{e^{\eta\Phi(\mathbf{s},G)}+e^{\eta\Phi(\mathbf{s},G-ij)}}\lambda.$$

For a change in the agents' actions with a transition from $\boldsymbol{\omega}=(s_i,\mathbf{s}_{-i},G)$ to $\boldsymbol{\omega}'=(s_i',\mathbf{s}_{-i},G)$ we get the following detailed balance condition

$$\frac{1}{\mathcal{Z}^\eta}e^{\eta\Phi(s_i,\mathbf{s}_{-i},G)}\frac{e^{\eta\Phi(s_i',\mathbf{s}_{-i},G)}}{e^{\eta\Phi(s_i,\mathbf{s}_{-i},G)}+e^{\eta\Phi(s_i',\mathbf{s}_{-i},G)}}\chi \quad = \quad \frac{1}{\mathcal{Z}^\eta}e^{\eta\Phi(s_i',\mathbf{s}_{-i},G)}\frac{e^{\eta\Phi(s_i,\mathbf{s}_{-i},G)}}{e^{\eta\Phi(s_i,\mathbf{s}_{-i},G)}+e^{\eta\Phi(s_i',\mathbf{s}_{-i},G)}}\chi.$$

Hence, the probability measure $\mu_\eta^{GI}(\boldsymbol{\omega})$ satisfies a detailed balance condition of (A.2) and therefore is the stationary distribution of the Markov chain with transition rates $q(\boldsymbol{\omega},\boldsymbol{\omega}')$. □

Proof of Proposition 3. The potential function can be written as

$$\Phi(\mathbf{s},G)=\sum_{i=1}^n\left(\gamma_i+\frac{\rho}{2}\sum_{j\neq i}^n s_j-\kappa\right)s_i+\frac{1}{2}\sum_{i=1}^n\sum_{j=1}^n a_{ij}\left(\theta s_is_j-\zeta_1+\frac{\zeta_1-\zeta_2}{2}(1-\gamma_i\gamma_j)\right). \tag{A.3}$$

Note that only the last summatory in (A.3) depends on the network G, through the entries of the elements a_{ij} of its adjacency matrix $\mathbf{A} = [a_{ij}]$. In particular, the term $\sum_{i=1}^n \sum_{j=1}^n a_{ij} s_i s_j$ is maximized over $s_i, s_j \in \{-1, +1\}$ for $a_{ij} = 1$ iff $s_i = s_j$. The term $\sum_{i=1}^n \sum_{j=1}^n a_{ij}(\theta s_i s_j - \zeta_1 + (\zeta_1 - \zeta_2)(1 - \gamma_i\gamma_j)/2$ is maximized over $s_i, s_j \in \{-1, +1\}$ for $a_{ij} = 1$ iff $s_i = s_j = \gamma_i = \gamma_j$ if $\zeta_1 < \theta < \zeta_2$ and $s_i = s_j$ if $\zeta_2 < \theta$. If $\theta < \zeta_1$ then $a_{ij} = 0$ and we obtain the empty network, $\overline{K}_n$.

To summarize, the candidate networks and action profiles that maximize the potential must be either complete, K_n, empty, $\overline{K}_n$, or composed of two disconnected cliques, $K_{n_1} \cup K_{n-n_1}$, in which all agents in the same clique chose the same action and have the same idiosyncratic preferences.

In the following we denote by $n_+ = \#(\{\gamma_i = 1 : i = 1, \dots, n\})$, $\nu_+ = n_+/n$ and define

$$\begin{aligned}
\theta^* &= \zeta_2 + \frac{2(1-\kappa)}{n - n_+} - 2\rho, \\
\theta^{**} &= \zeta_2 + \frac{2(1+\kappa)}{n_+} - 2\rho, \\
\kappa^* &= 2\nu_+ - 1, \\
\rho^* &= \frac{1-\kappa}{n - n_+}, \\
\rho^{**} &= \frac{1+\kappa}{n_+}.
\end{aligned}$$

Further, w.l.o.g. κ is assumed to be non-negative. Note also that $\kappa^* \leq 1$ and $\kappa^* \leq 0$ if $\nu_+ \leq 1/2$.

We first consider the case of $\theta < \zeta_1$. In this case the potential maximizing network will be the empty network $\overline{K}_n$. When all agents choose the action $s_i = -1$ the potential is given by

$$\Phi((-1, \dots, -1), \overline{K}_n) = n - 2n_+ + \frac{\rho n(n-1)}{2} + \kappa n.$$

When all agents choose the action $s_i = +1$ the potential is given by

$$\Phi((+1, \dots, +1), \overline{K}_n) = n_+ - (n - n_+) + \frac{\rho n(n-1)}{2} - \kappa n.$$

When all agents choose the action $s_i = \gamma_i$ the potential is given by

$$\Phi(\boldsymbol{\gamma}, \overline{K}_n) = n + \frac{\rho}{2}\left((n - 2n_+)^2 - n\right) - \kappa(2n_+ - n).$$

The difference in the potential between $s_i = -1$ and $s_i = \gamma_i$ is given by

$$\Phi((-1, \dots, -1), \overline{K}_n) - \Phi(\boldsymbol{\gamma}, \overline{K}_n) = 2n_+(\kappa + \rho(n - n_+) - 1).$$

Setting this to zero gives us the threshold

$$\rho^* = \frac{1-\kappa}{n - n_+}.$$

The difference in the potential between $s_i = +1$ and $s_i = \gamma_i$ is given by

$$\Phi((+1, \dots, +1), \overline{K}_n) - \Phi(\boldsymbol{\gamma}, \overline{K}_n) = -2(n - n_+)(\kappa + 1 - n_+\rho).$$

Setting this to zero gives us the threshold

$$\rho^{**} = \frac{1+\kappa}{n_+}.$$

Moreover, the difference in the potential between $s_i = +1$ and $s_i = -1$ is given by

$$\tilde{\Phi}((+1,\ldots,+1),\overline{K}_n) - \tilde{\Phi}((-1,\ldots,-1),\overline{K}_n) = -2((\kappa+1)n - 2n_+),$$

which is positive if

$$\kappa < \kappa^* = 2\nu^* - 1 \leq 1.$$

Thus if $\theta < \zeta_1$, the potential maximizing state is given by the empty network $\overline{K}_n$ where

1. If $\nu_+ < 1/2$
   (a) and $\rho > \rho^*$, then all agents choose the action $s_i = -1$.
   (b) and $\rho < \rho^*$, then all agents choose the action $s_i = \gamma_i$.
2. If $\nu_+ > 1/2$,
   (a) $\kappa > \kappa^*$
      i. and $\rho > \rho^*$, then all agents choose the action $s_i = -1$.
      ii. and $\rho < \rho^*$, then all agents choose the action $s_i = \gamma_i$.
   (b) $\kappa < \kappa^*$
      i. and $\rho > \rho^{**}$, then all agents choose the action $s_i = +1$.
      ii. and $\rho < \rho^{**}$, then all agents choose the action $s_i = \gamma_i$.

Note that for $n \to \infty$ we get that

$$\rho^* = \frac{1-\kappa}{n-n_+} \to 0,$$
$$\rho^{**} = \frac{1+\kappa}{n_+} \to 0.$$

Hence, when $n \to \infty$ then if $\theta < \zeta_1$, the potential maximizing state is given by the empty network $\overline{K}_n$:

1. If $\nu_+ < 1/2$ then all agents choose the action $s_i = -1$.
2. If $\nu_+ > 1/2$ and
   (a) $\kappa > \kappa^*$ then all agents choose the action $s_i = -1$.
   (b) $\kappa < \kappa^*$ then all agents choose the action $s_i = +1$.

Next we consider the case of $\theta > \zeta_1$. The potential maximizing network is either complete, K_n, or composed of two cliques, $K_{n_+} \cup K_{n-n_+}$. In the following we consider separately the cases of $\nu_+ < 1/2$ and $\nu_+ > 1/2$.

Case 1: $\nu_+ < 1/2$. The potential for the complete Graph K_n with all agents choosing action $s_i = -1$ is given by

$$\Phi((-1,\dots,-1),K_n) = (n-2n_+) + \frac{1}{2}[n(n-1) - 2n_+(n-n_+)](\theta-\zeta_1) + n_+(n-n_+)(\theta-\zeta_2) + \frac{\rho n(n-1)}{2} + \kappa n.$$

The potential for the union of cliques $K_{n_+} \cup K_{n-n_+}$ with all agents choosing action $s_i = \gamma_i$ is given by

$$\Phi(\boldsymbol{\gamma}, K_{n_+} \cup K_{n-n_+}) = n + \frac{1}{2}[n(n-1) - 2n_+(n-n_+)](\theta-\zeta_1) + \frac{\rho}{2}[(n-2n_+)^2 - n] - \kappa(2n_+ - n).$$

The potential difference between the complete network K_n in which all agents choose action $s_i = -1$) and the union of cliques $K_{n_+} \cup K_{n-n_+}$ in which the agents choose their idiosyncratic preferences $s_i = \gamma_i$) is given by

$$\Phi((-1,\dots,-1),K_n) - \Phi(\boldsymbol{\gamma}, K_{n_+} \cup K_{n-n_+}) = n_+[2(\kappa-1) - (n-n_+)(\zeta_2 - \theta - 2\rho)].$$

Setting this to zero we find the threshold:

$$\theta^* = \zeta_2 + \frac{2(1-\kappa)}{n-n_+} - 2\rho = \zeta_2 + 2(\rho^* - \rho).$$

Moreover

$$\Phi((-1,\dots,-1),K_n) - \Phi((-1,\dots,-1),K_{n_+} \cup K_{n-n_+}) = n_+(\theta-\zeta_2)(n-n_+),$$

which is positive if

$$\theta > \zeta_2.$$

Further

$$\Phi((-1,\dots,-1),K_{n_+} \cup K_{n-n_+}) - \Phi(\boldsymbol{\gamma}, K_{n_+} \cup K_{n-n_+}) = 2n_+(\rho(n-n_+) - (1-\kappa)).$$

which is positive for

$$\rho > \rho^* = \frac{1-\kappa}{n-n_+}.$$

We can then conclude that:

1. If $\theta > \theta^*$ and
   (a) $\theta > \zeta_2$ then the potential function of the complete graph K_n with $s_i = -1$ is higher than that of the union of cliques $K_{n_+} \cup K_{n-n_+}$ with $s_i = -1$ or $s_i = \gamma_i$. Therefore, the stochastically stable state is the complete graph K_n with $s_i = -1$.
   (b) $\theta < \zeta_2$ then the potential function of the union of cliques $K_{n_+} \cup K_{n-n_+}$ with $s_i = -1$ is higher than that of the complete graph K_n with $s_i = -1$. Therefore, the stochastically stable state is the union of cliques $K_{n_+} \cup K_{n-n_+}$ with $s_i = -1$.
2. If $\theta < \theta^*$ and
   (a) $\rho < \rho^*$ then the potential function of the union of cliques $K_{n_+} \cup K_{n-n_+}$ with $s_i = \gamma_i$ is higher than that of the complete graph K_n with $s_i = -1$ or the union of cliques

$K_{n_+} \cup K_{n-n_+}$ with $s_i = -1$. Therefore, the stochastically stable state is the union of cliques $K_{n_+} \cup K_{n-n_+}$ with $s_i = \gamma_i$.

(b) $\rho > \rho^*$ then the potential function of the union of cliques $K_{n_+} \cup K_{n-n_+}$ with $s_i = -1$ is higher than that of the complete graph K_n with $s_i = -1$ or the union of cliques $K_{n_+} \cup K_{n-n_+}$ with $s_i = \gamma_i$. Therefore, the stochastically stable state is the union of cliques $K_{n_+} \cup K_{n-n_+}$ with $s_i = -1$.

Case 2: $\nu_+ > 1/2$. We first note that $\Phi((-1,\dots,-1), K_n) - \Phi((+1,\dots,+1), K_n) = 2(n(\kappa + 1) - 2n_+)$, which is increasing in κ. For $\nu_+ < 1/2$ this equation is strictly positive for any value of $\kappa \geq 0$. In contrast, for $\nu_+ > 1/2$ we have that $\Phi((-1,\dots,-1), K_n) < \Phi((+1,\dots,+1), K_n)$ if $\kappa < \kappa^* = \frac{2n_+}{n} - 1$, and κ^* being positive only if $n_+ > n/2$. Moreover, $\Phi((-1,\dots,-1), K_{n_+} \cup K_{n-n_+})$ $-\Phi((+1,\dots,+1), K_{n_+} \cup K_{n-n_+}) = 2n(\kappa - \kappa^*)$, which is positive for $\kappa > \kappa^*$ and negative for $\kappa < \kappa^*$. We thus need to distinguish further the cases $\kappa > \kappa^*$ and $\kappa < \kappa^*$.

1. **Subcase 2.1:** $\kappa > \kappa^*$. The potential for the complete graph K_n in which all agents choose action $s_i = -1$ is given by

$$\Phi((-1,\dots,-1), K_n) = (n - 2n_+) + \frac{1}{2}[n(n-1) - 2n_+(n - n_+)](\theta - \zeta_1) + n_+(n - n_+)(\theta - \zeta_2) + \frac{\rho n(n-1)}{2} + \kappa n.$$

The potential for the union of cliques $K_{n_+} \cup K_{n-n_+}$ in which all agents choose action $s_i = \gamma_i$ is given by

$$\Phi(\boldsymbol{\gamma}, K_{n_+} \cup K_{n-n_+}) = n + \frac{1}{2}[n(n-1) - 2n_+(n - n_+)](\theta - \zeta_1) + \frac{\rho}{2}[(n - 2n_+)^2 - n] - \kappa(2n_+ - n).$$

The difference between the potentials for the complete graph K_n (all $s_i = -1$) and the union of cliques $K_{n_+} \cup K_{n-n_+}$ ($s_i = \gamma_i$) is given by

$$\Phi((-1,\dots,-1), K_n) - \Phi(\boldsymbol{\gamma}, K_{n_+} \cup K_{n-n_+}) = n_+[2(\kappa - 1) - (n - n_+)(\zeta_2 - \theta - 2\rho)].$$

Setting this to zero we find the threshold

$$\theta^* = \zeta_2 + \frac{2(1-\kappa)}{n - n_+} - 2\rho = \zeta_2 + 2(\rho^* - \rho).$$

Moreover,

$$\Phi((-1,\dots,-1), K_n) - \Phi((-1,\dots,-1), K_{n_+} \cup K_{n-n_+}) = n_+(\theta - \zeta_2)(n - n_+),$$

which is zero if $\theta = \zeta_2$. Further,

$$\Phi((-1,\dots,-1), K_{n_+} \cup K_{n-n_+}) - \Phi(\boldsymbol{\gamma}, K_{n_+} \cup K_{n-n_+}) = 2n_+(\rho(n - n_+) - (1 - \kappa)),$$

which is positive for

$$\rho > \rho^* = \frac{1-\kappa}{n - n_+}.$$

We then can conclude the following:

(a) If $\theta > \theta^*$ and

i. $\theta > \zeta_2$ then the potential function of the complete graph K_n with $s_i = -1$ is higher than that of the union of cliques $K_{n_+} \cup K_{n-n_+}$ with $s_i = -1$ or $s_i = \gamma_i$. Therefore, the stochastically stable state is the complete graph K_n with $s_i = -1$.

ii. $\theta < \zeta_2$ then the potential function of the union of cliques $K_{n_+} \cup K_{n-n_+}$ with $s_i = -1$ is higher than that of the complete graph K_n with $s_i = -1$ or the union of cliques $K_{n_+} \cup K_{n-n_+}$ with $s_i = \gamma_i$. Therefore, the stochastically stable state is the union of cliques $K_{n_+} \cup K_{n-n_+}$ with $s_i = -1$.

(b) If $\theta < \theta^*$ and

i. $\rho < \rho^*$ then the potential function of the union of cliques $K_{n_+} \cup K_{n-n_+}$ with $s_i = \gamma_i$ is higher than that of the complete graph K_n with $s_i = -1$ or the union of cliques $K_{n_+} \cup K_{n-n_+}$ with $s_i = -1$. Therefore, the stochastically stable state is the union of cliques $K_{n_+} \cup K_{n-n_+}$ with $s_i = \gamma_i$.

ii. $\rho > \rho^*$ then the potential function of the union of cliques $K_{n_+} \cup K_{n-n_+}$ with $s_i = -1$ is higher than that of the complete graph K_n with $s_i = -1$ or the union of cliques $K_{n_+} \cup K_{n-n_+}$ with $s_i = \gamma_i$. Therefore, the stochastically stable state is the union of cliques $K_{n_+} \cup K_{n-n_+}$ with $s_i = -1$.

2. **Subcase 2.2:** $\kappa < \kappa^*$. The potential for the complete graph K_n with all agents choosing action $s_i = +1$ is given by

$$\Phi((+1,\dots,+1),K_n) = (2n_+ - n) + \frac{1}{2}[n(n-1) - 2n_+(n-n_+)](\theta - \zeta_1) + n_+(n-n_+)(\theta - \zeta_2) + \frac{\rho n(n-1)}{2} - \kappa n.$$

The potential in the union of cliques $K_{n_+} \cup K_{n-n_+}$ with all agents choosing action $s_i = \gamma_i$ is given by

$$\Phi(\boldsymbol{\gamma}, K_{n_+} \cup K_{n-n_+}) = n + \frac{1}{2}[n(n-1) - 2n_+(n-n_+)](\theta - \zeta_1) + \frac{\rho}{2}[(n-2n_+)^2 - n] - \kappa(2n_+ - n).$$

The difference between the potentials of the complete graph K_n (all $s_i = +1$) and the union of cliques $K_{n_+} \cup K_{n-n_+}$ (all $s_i = \gamma_i$) is then

$$\Phi((+1,\dots,+1),K_n) - \Phi(\boldsymbol{\gamma}, K_{n_+} \cup K_{n-n_+}) = -(n-n_+)[2(\kappa+1) + n_+(\zeta_2 - \theta - 2\rho)].$$

Setting this to zero we obtain the threshold

$$\theta^{**} = \zeta_2 + \frac{2(1+\kappa)}{n_+} - 2\rho = \zeta_2 + 2(\rho^{**} - \rho).$$

Moreover,

$$\Phi((+1,\dots,+1),K_n) - \Phi((+1,\dots,+1),K_{n_+} \cup K_{n-n_+}) = n_+(\theta - \zeta_2)(n-n_+),$$

which is zero if $\theta = \zeta_2$. Further,

$$\Phi((+1,\dots,+1),K_{n_+} \cup K_{n-n_+}) - \Phi(\boldsymbol{\gamma}, K_{n_+} \cup K_{n-n_+}) = -2(n-n_+)((\kappa+1) - n_+\rho),$$

which is negative for

$$\rho < \rho^{**} = \frac{1+\kappa}{n_+}.$$

We then can summarize the above cases as follows:

(a) If $\theta > \theta^{**}$ and

   i. $\theta > \zeta_2$ then the potential function of the complete graph K_n with $s_i = +1$ is higher than that of the union of cliques $K_{n_+} \cup K_{n-n_+}$ with $s_i = \gamma_i$ or $s_i = +1$. Therefore, the stochastically stable state is the complete graph K_n with $s_i = +1$.

   ii. $\theta < \zeta_2$ then the potential function of the union of cliques $K_{n_+} \cup K_{n-n_+}$ with $s_i = +1$ is higher than that of the complete graph K_n with $s_i = +1$ or the union of cliques $K_{n_+} \cup K_{n-n_+}$ with $s_i = \gamma_i$. Therefore, the stochastically stable state is the union of cliques $K_{n_+} \cup K_{n-n_+}$ with $s_i = +1$.

(b) If $\theta < \theta^{**}$ and

   i. $\rho < \rho^{**}$ then the potential function of the union of cliques $K_{n_+} \cup K_{n-n_+}$ with $s_i = \gamma_i$ is higher than that of the complete graph K_n with $s_i = +1$ or the union of cliques $K_{n_+} \cup K_{n-n_+}$ with $s_i = +1$. Therefore, the stochastically stable state is the union of cliques $K_{n_+} \cup K_{n-n_+}$ with $s_i = \gamma_i$.

   ii. $\rho > \rho^{**}$ then the potential function of the union of cliques $K_{n_+} \cup K_{n-n_+}$ with $s_i = +1$ is higher than that of the complete graph K_n with $s_i = +1$ or the union of cliques $K_{n_+} \cup K_{n-n_+}$ with $s_i = \gamma_i$. Therefore, the potential maximizing state is the union of cliques $K_{n_+} \cup K_{n-n_+}$ with $s_i = +1$.

Thus, if $\theta > \zeta_1$, the stochastically stable state is either complete K_n or composed of two cliques $K_{n_+} \cup K_{n-n_+}$. More precisely:

1. If $\nu_+ < 1/2$,

   (a) $\theta > \theta^*$ and

      i. $\theta > \zeta_2$, then the stochastically stable state is the complete graph K_n with $s_i = -1$.

      ii. $\theta < \zeta_2$, then the stochastically stable state is the union of cliques $K_{n_+} \cup K_{n-n_+}$ with $s_i = -1$.

   (b) $\theta < \theta^*$ and

      i. $\rho < \rho^*$ then the stochastically stable state is the union of two cliques $K_{n_+} \cup K_{n-n_+}$ with $s_i = \gamma_i$.

      ii. $\rho > \rho^*$ then the stochastically stable state is the union of two cliques $K_{n_+} \cup K_{n-n_+}$ with $s_i = -1$.

2. If $\nu_+ > 1/2$,

   (a) $\kappa > \kappa^*$,

      i. $\theta > \theta^*$ and

         - $\theta > \zeta_2$, then the stochastically stable state is the complete graph K_n with $s_i = -1$.
         - $\theta < \zeta_2$, then the stochastically stable state is the union of cliques $K_{n_+} \cup K_{n-n_+}$ with $s_i = -1$.

      ii. $\theta < \theta^*$ and

         - $\rho < \rho^*$ then the stochastically stable state is the union of two cliques $K_{n_+} \cup K_{n-n_+}$ with $s_i = \gamma_i$.
         - $\rho > \rho^*$ then the stochastically stable state is the union of two cliques $K_{n_+} \cup K_{n-n_+}$ with $s_i = -1$.

   (b) $\kappa < \kappa^*$,

i. $\theta > \theta^{**}$ and
   - $\theta > \zeta_2$, then the stochastically stable state is the complete graph K_n with $s_i = +1$.
   - $\theta < \zeta_2$, then the stochastically stable state is the union of cliques $K_{n_+} \cup K_{n-n_+}$ with $s_i = +1$.

ii. $\theta < \theta^{**}$ and
   - $\rho < \rho^{**}$ then the stochastically stable state is the union of two cliques $K_{n_+} \cup K_{n-n_+}$ with $s_i = \gamma_i$
   - $\rho > \rho^{**}$ then the stochastically stable state is the union of two cliques $K_{n_+} \cup K_{n-n_+}$ with $s_i = +1$.

This proves Proposition E.1. Further, note that for $n \to \infty$ we get that

$$\rho^* = \frac{1-\kappa}{n-n_+} \xrightarrow[n\to\infty]{} 0,$$
$$\rho^{**} = \frac{1+\kappa}{n_+} \xrightarrow[n\to\infty]{} 0,$$
$$\lim_{n\to\infty} \theta^* = \lim_{n\to\infty} \theta^{**} = \zeta_2 - 2\rho.$$

Hence, when $n \to \infty$ then if $\theta > \zeta_1$, the stochastically stable state maximizing the potential is either complete K_n or composed of two cliques $K_{n_+} \cup K_{n-n_+}$ where:

1. If $\nu_+ < 1/2$,
   (a) $\theta > \zeta_2 - 2\rho$ and
      i. $\theta > \zeta_2$, then the stochastically stable state is the complete graph K_n with $s_i = -1$.
      ii. $\theta < \zeta_2$ then the stochastically stable state is the union of two cliques $K_{n_+} \cup K_{n-n_+}$ with $s_i = -1$.
   (b) If $\theta < \zeta_2 - 2\rho$, then the stochastically stable state is the union of two cliques $K_{n_+} \cup K_{n-n_+}$ with $s_i = -1$.
2. If $\nu_+ > 1/2$,
   (a) $\kappa > \kappa^*$,
      i. $\theta > \zeta_2 - 2\rho$ and
         - $\theta > \zeta_2$, then the stochastically stable state is the complete graph K_n with $s_i = -1$.
         - $\theta < \zeta_2$, then the stochastically stable state is the union of two cliques $K_{n_+} \cup K_{n-n_+}$ with $s_i = -1$.
      ii. $\theta < \zeta_2 - 2\rho$, then the stochastically stable state is the union of two cliques $K_{n_+} \cup K_{n-n_+}$ with $s_i = -1$.
   (b) If $\kappa < \kappa^*$,
      i. $\theta > \zeta_2 - 2\rho$ and
         - $\theta > \zeta_2$, then the stochastically stable state is the complete graph K_n with $s_i = +1$.
         - $\theta < \zeta_2$, then the stochastically stable state is the union of cliques $K_{n_+} \cup K_{n-n_+}$ with $s_i = +1$.

ii. $\theta < \zeta_2 - 2\rho$, then the stochastically stable state is the union of two cliques $K_{n_+} \cup K_{n-n_+}$ with $s_i = +1$.

A concise summary of the above cases in the large population limit ($n \to \infty$) is stated in Proposition 3. This concludes the proof. □

Proof of Proposition 4. As a preliminary step in our analysis of the LIL scenario, we prove the following lemma. Its purpose is to provide a useful closed-form expression that specifies how, depending on the prevailing state of the system, the beliefs are formed in a large population when belief updating is mainly driven by social learning.

Lemma 1. *Let* $\boldsymbol{\omega} = (\mathbf{s}, G) \in \hat{\Omega}_n (\subset \Omega_n)$*. Then,* $\lim_{\varphi \to 0^+} \psi_i^{LIL}(\boldsymbol{\omega}) = \mathbf{p}^*$ *where* $\mathbf{p}^* = (p^*, p^*, \ldots, p^*) \in [0,1]^n$ *with* $p^* \propto \sum_{i=1}^n \frac{1+d_i}{d_i} \sum_{j=1}^n a_{ij} s_j$.

Proof. As explained in Subsection 2.3 (cf. Equation (8)), the limit beliefs of the learning process taking place at any state $\boldsymbol{\omega} = (\mathbf{s}, G)$ are characterized by the following stationarity condition: $\mathbf{p}^* = \varphi \mathbf{D}^{-1}\mathbf{A}\mathbf{s} + (1-\varphi)\widehat{\mathbf{D}}^{-1}\widehat{\mathbf{A}}\mathbf{p}^*$, where $\mathbf{D} \equiv \text{diag}(d_1, \ldots, d_n)$ is the diagonal matrix of agents' degrees, $\widehat{\mathbf{D}} \equiv \mathbf{I}_n + \mathbf{D}$ with $\mathbf{I}_n$ being the identity matrix, $\mathbf{A}$ is the adjacency matrix of the prevailing network, and $\widehat{\mathbf{A}} \equiv \mathbf{I}_n + \mathbf{A}$.

Since the largest eigenvalue of the row-stochastic matrix $\widehat{\mathbf{D}}^{-1}\widehat{\mathbf{A}}$ is one and $(1-\varphi) \in (0,1)$, the matrix $\mathbf{I}_n - (1-\varphi)\widehat{\mathbf{D}}^{-1}\widehat{\mathbf{A}}$ is invertible and therefore we can write:

$$\mathbf{p}^* = \varphi \left[\mathbf{I}_n - (1-\varphi)\widehat{\mathbf{D}}^{-1}\widehat{\mathbf{A}}\right]^{-1} \mathbf{D}^{-1}\mathbf{A}\mathbf{s}. \tag{A.4}$$

Consider now any $\boldsymbol{\omega} = (\mathbf{s}, G) \in \hat{\Omega}_n$. By Property P2, its corresponding network G is path-connected. Thus, the matrix denoted by $\widehat{\mathbf{W}} \equiv \widehat{\mathbf{D}}^{-1}\widehat{\mathbf{A}}$ is row-stochastic matrix (each of its rows summing to one) and has a unique largest eigenvalue $\lambda_1 = 1$ with corresponding right eigenvector $\mathbf{u} = (1, \ldots, 1)^\top$ and left eigenvector $\boldsymbol{\pi}$ (thus satisfying the equation $\boldsymbol{\pi}\widehat{\mathbf{W}} = \boldsymbol{\pi}$) given by:

$$\boldsymbol{\pi} = \frac{1}{2m+n}(d_1+1, \ldots, d_n+1), \tag{A.5}$$

where m is the number of links in the network. Next, we introduce the similarity transform $\mathbf{S} = \widehat{\mathbf{D}}^{1/2}\widehat{\mathbf{W}}\widehat{\mathbf{D}}^{-1/2}$ and, conversely, $\widehat{\mathbf{W}} = \widehat{\mathbf{D}}^{-1/2}\mathbf{S}\widehat{\mathbf{D}}^{1/2}$. The matrix $\mathbf{S}$ is a symmetric matrix with the same eigenvalues λ_i as $\widehat{\mathbf{W}}$, $1 = \lambda_1 > \lambda_2 \geq \ldots \geq \lambda_n \geq -1$. Therefore, by the Perron-Frobenius theorem, the following spectral representation holds (Seneta, 2006): $\mathbf{S} = \sum_{i=1}^n \lambda_i \mathbf{v}_i \mathbf{v}_i^\top$, where $\mathbf{v}_i$ is the eigenvector to the i-th eigenvalue λ_i of the real and symmetric matrix $\mathbf{S}$, i.e. $\mathbf{S}\mathbf{v}_i = \lambda_i \mathbf{v}_i$ for each $i = 1, 2, ..., n$. Next, given any $\alpha < 1$, we can apply the following Neumann series expansion (Meyer, 2000) to write: $(\mathbf{I}_n - \alpha\widehat{\mathbf{W}})^{-1} = \sum_{k=1}^\infty \alpha^k \widehat{\mathbf{W}}^k$. And from the spectral representation $\mathbf{S} = \sum_{i=1}^n \lambda_i \mathbf{v}_i \mathbf{v}_i^\top$, combined with the fact that the eigenvectors form an orthonormal basis in $\mathbb{R}^n$ (i.e. $\mathbf{v}_i^\top \mathbf{v}_i = 1$ while $\mathbf{v}_i^\top \mathbf{v}_j = 0$ for $j \neq i$) we have: $\widehat{\mathbf{W}}^k = (\widehat{\mathbf{D}}^{-1/2}\mathbf{S}\widehat{\mathbf{D}}^{1/2})^k = (\widehat{\mathbf{D}}^{-1/2}\sum_{i=1}^n \lambda_i \mathbf{v}_i \mathbf{v}_i^\top \widehat{\mathbf{D}}^{1/2})^k$ $= (\widehat{\mathbf{D}}^{-1/2}\sum_{i=1}^n \lambda_i \mathbf{v}_i \mathbf{v}_i^\top \widehat{\mathbf{D}}^{1/2}) \cdots (\widehat{\mathbf{D}}^{-1/2}\sum_{i=1}^n \lambda_i \mathbf{v}_i \mathbf{v}_i^\top \widehat{\mathbf{D}}^{1/2}) = \widehat{\mathbf{D}}^{-1/2}\sum_{i=1}^n \lambda_i^k \mathbf{v}_i \mathbf{v}_i^\top \widehat{\mathbf{D}}^{1/2}$. Hence the Neumann series expansion from above can be developed as follows: $(\mathbf{I}_n - \alpha\widehat{\mathbf{W}})^{-1} = \sum_{i=1}^n \widehat{\mathbf{D}}^{-1/2}\mathbf{v}_i$ $\mathbf{v}_i^\top \widehat{\mathbf{D}}^{1/2} \sum_{k=1}^\infty \alpha^k \lambda_i^k = \sum_{i=1}^n \frac{1}{1-\alpha\lambda_i}\widehat{\mathbf{D}}^{-1/2}\mathbf{v}_i\ \mathbf{v}_i^\top \widehat{\mathbf{D}}^{1/2} = \frac{1}{1-\alpha\lambda_1}\ \widehat{\mathbf{D}}^{-1/2}\mathbf{v}_1\ \mathbf{v}_1^\top\ \widehat{\mathbf{D}}^{1/2} + \sum_{i=2}^n \frac{1}{1-\alpha\lambda_i}\widehat{\mathbf{D}}^{-1/2}$ $\mathbf{v}_i\ \mathbf{v}_i^\top \widehat{\mathbf{D}}^{1/2} = \frac{1}{1-\alpha}\ \widehat{\mathbf{D}}^{-1/2}\mathbf{v}_1\ \mathbf{v}_1^\top\ \widehat{\mathbf{D}}^{1/2} + \sum_{i=2}^n \frac{1}{1-\alpha\lambda_i}\ \widehat{\mathbf{D}}^{-1/2}\ \mathbf{v}_i \mathbf{v}_i^\top \widehat{\mathbf{D}}^{1/2}$. Thus noting that $|\lambda_i| < 1$ for all $i = 2, \ldots, n$, when α converges to 1 from below we can write: $\lim_{\alpha \to 1^-}(1-\alpha)(\mathbf{I}_n - \alpha\widehat{\mathbf{W}})^{-1}$ $= \widehat{\mathbf{D}}^{-1/2}\mathbf{v}_1\mathbf{v}_1^\top \widehat{\mathbf{D}}^{1/2} = \mathbf{u}(\mathbf{v}_1^\top)^2 = \frac{1}{2m+n}(1, \ldots, 1)^\top \times (d_1+1, \ldots, d_n+1)$ where we have used the fact that the similarity transform maps the left-eigenvector $\boldsymbol{\pi}$ into $\mathbf{v}_1 = \widehat{\mathbf{D}}^{-1/2}\boldsymbol{\pi}$ with $v_{1i} = \sqrt{\frac{d_i+1}{2m+n}}$ (cf. Van Mieghem, 2011, Sec. 3.7).

Finally, applying the above derivations to Equation (A.4) when $\varphi \to 0^+$ we obtain, as claimed, that $\mathbf{p}^* \sim \varphi \frac{1}{1-(1-\varphi)} \mathbf{u}(\mathbf{v}_1^\top)^2 \mathbf{W}\mathbf{s} = \frac{1}{2m+n}(\sum_{i=1}^n \frac{1+d_i}{d_i} \sum_{j=1}^n a_{ij}s_j, \ldots, \sum_{i=1}^n \frac{1+d_i}{d_i} \sum_{j=1}^n a_{ij}s_j)^\top$ □

When the population is large, we can establish the following corollary that refines in a useful way the conclusion obtained from Lemma 1. To state formally the result, we rely on the following notation. Given any $\boldsymbol{\omega} \in \Omega$, denote by $\zeta_d^+(\boldsymbol{\omega})$ and $\zeta_d^-(\boldsymbol{\omega})$ the fraction of individuals with degree d that choose, respectively, actions $+1$ and -1 in state $\boldsymbol{\omega}$. Note that, obviously, the overall degree distribution $\{\zeta_d(\boldsymbol{\omega})\}_{d\in\mathbb{N}}$ at $\boldsymbol{\omega}$ satisfies $\zeta_d(\boldsymbol{\omega}) = \zeta_d^+(\boldsymbol{\omega}) + \zeta_d^-(\boldsymbol{\omega})$ and therefore $\sum_{d\in\mathbb{N}}[\zeta_d^+(\boldsymbol{\omega}) + \zeta_d^-(\boldsymbol{\omega})] = 1$.

Corollary 1. *For any $\eta > 0$ there exists some $\hat{n} \in \mathbb{N}$ and $\hat{\varphi} > 0$ such that if $n \geq \hat{n}$, $\varphi \leq \hat{\varphi}$, then for all $\boldsymbol{\omega} \in \hat{\Omega}_n$ and all $i \in \mathcal{N}$, we have $\left|\psi_i^{LIL}(\boldsymbol{\omega}) - \left[\sum_{d\in\mathbb{N}} \left(\zeta_d^+(\boldsymbol{\omega}) - \zeta_d^-(\boldsymbol{\omega})\right) d\right]\right| \leq \eta$.*

Proof. For states $\boldsymbol{\omega} \in \hat{\Omega}_n$ we know from Property P1 that, as the population size n grows, the degree d_i of each agent i can be made arbitrarily large as the population size increases. Thus, if n is sufficiently large, we can rely on Lemma 1 to arrive at the following further approximation for small φ:

$$\mathbf{p}^* = \boldsymbol{\psi}^{LIL}(\boldsymbol{\omega}) \approx \frac{1}{2m} \begin{bmatrix} \sum_{i=1}^n \sum_{j=1}^n a_{ij}s_j \\ \vdots \\ \sum_{i=1}^n \sum_{j=1}^n a_{ij}s_j \end{bmatrix} = \frac{1}{2m} \begin{bmatrix} \sum_{j=1}^n d_j s_j \\ \vdots \\ \sum_{j=1}^n d_j s_j \end{bmatrix} = \sum_{d\in\mathbb{N}} \left(\zeta_d^+(\boldsymbol{\omega}) - \zeta_d^-(\boldsymbol{\omega})\right) d \begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix}. \tag{A.6}$$

which implies the desired conclusion. □

Next, we build on the preceding results to establish a series of additional lemmas that, under assumptions A.1 and A.2, eventually allow us to characterize the long run behavior of the adjustment process with an arbitrarily high probability. To this end, it will be useful to center our attention on the *complement* of the sets we have denoted by H_n^δ. That is, on the sets $\bar{H}_n^\delta \equiv \{\boldsymbol{\omega} = (\mathbf{s}, G)) \in \Omega_n : |\frac{1}{n}\sum_{i=1}^n s_i| > \delta\}$ that include the states that display an average action differing from zero by more than δ. In fact, the focus of our analysis will be narrower than that: it will consider the sets $\hat{\Omega}_n^\delta \equiv \hat{\Omega}_n \cap \bar{H}_n^\delta$ consisting of the states that not only satisfy the aforementioned property but also the properties we have labeled as P1-P3.

For any such state $\boldsymbol{\omega} \in \hat{\Omega}_n^\delta$, the objective is to show that every transition to an adjacent $\boldsymbol{\omega}'$ leads to changes in the function $\tilde{\Phi}(\cdot)$ given by (cf. (18)):

$$\tilde{\Phi}(\boldsymbol{\omega}') - \tilde{\Phi}(\boldsymbol{\omega}) = \pi_i(\boldsymbol{\omega}', \psi_i^{LIL}(\boldsymbol{\omega})) - \pi_i(\boldsymbol{\omega}, \psi_i^{LIL}(\boldsymbol{\omega})) + \rho(n-1)\sum_{j=1}^n s_j'(\psi_j^{LIL}(\boldsymbol{\omega}') - \psi_j^{LIL}(\boldsymbol{\omega})) \tag{A.7}$$

that display the a sign (ordinal) alignment with the change in the payoffs of the agent(s) implementing the transition. As informally discussed in Section 3.2, the focus will be on relating/comparing the change in payoffs of those agents as given by the first two terms (payoff difference) in the RHS of (A.7) with the third term of it, which can be conceived as a remainder embodying the effect that the change has on the rest of the population. The approach will be different in the case of action adjustment (which is unilateral) and the case where the adjustment involves a link (and is therefore bilateral). In the first case, what we shall show is that the conditions that determine the sign of both terms depends in the same way on the same condition – hence the desired alignment readily follows. Instead, in the second case, the argument will hinge upon showing that, even when these terms are of opposite sign, the payoff changes of the agents involved in the link revision always dominate in magnitude to the effect induced on others. This, of course, also leads to the desired alignment.

Starting by the analysis of transitions induced by action revision, the next result characterizes the sign of the payoff change expected by the revising agent when the current state is in $\hat{\Omega}_n^\delta$, the population is large, and belief updating is mostly driven by social learning.

Lemma 2. *Given any $\delta > 0$, there exist some $\hat{n} \in \mathbb{N}$ and $\hat{\varphi} > 0$ such that, if $n \geq \hat{n}$ and $\varphi \leq \hat{\varphi}$, any adjustment taking place at a state $\boldsymbol{\omega} = (\mathbf{s}, G)$ in $\hat{\Omega}_n^\delta$ toward an adjacent state $\boldsymbol{\omega}' = (\mathbf{s}', G') \in \mathcal{A}_i(\boldsymbol{\omega})$ for some $i \in \mathcal{N}$ and $G' = G$ satisfies:*[28]

$$\pi_i(\boldsymbol{\omega}', \psi(\boldsymbol{\omega})) - \pi_i(\boldsymbol{\omega}, \psi(\boldsymbol{\omega})) > 0 \;\Leftrightarrow\; s_i' \sum_{k=1}^{n} s_k > 0. \tag{A.8}$$

Proof. The argument proceeds in three stages, which we number consecutively.

1. Consider any state $\boldsymbol{\omega} \in \hat{\Omega}_n^\delta$. Then, given δ, we can rely on Lemma 1 and Corollary 1 to assert that, if the population size n is large and φ is small, the common beliefs $\psi_i^{LIL}(\boldsymbol{\omega})$ shared by all $i \in \mathcal{N}$ at the end of the learning process taking place at $\boldsymbol{\omega}$ approximate those given in (A.6). More precisely, the claim here is that for any $\eta > 0$ there must exist $\hat{n}_1 \in \mathbb{N}$ and $\hat{\varphi} > 0$ such that if $n \geq \hat{n}_1$ and $\varphi \leq \hat{\varphi}$, then the uniform "consensus" profile of individual beliefs $\boldsymbol{\psi}(\boldsymbol{\omega}) = (p^*(\boldsymbol{\omega}), \ldots, p^*(\boldsymbol{\omega}))$ satisfies $\left| p^*(\boldsymbol{\omega}) - \left[\sum_{d\in\mathbb{N}} \left(\zeta_d^+(\boldsymbol{\omega}) - \zeta_d^-(\boldsymbol{\omega})\right) d\right] \right| \leq \eta$.

2. Let us choose η to satisfy $\eta < \delta/2$. Then we argue that, under the conditions specified in stage 1, we have that

$$\left[\sum_{d\in\mathbb{N}} \left(\zeta_d^+(\boldsymbol{\omega}) - \zeta_d^-(\boldsymbol{\omega})\right) d\right] \cdot p^*(\boldsymbol{\omega}) > 0, \tag{A.9}$$

i.e. both factors in the above inequality are of the same (non-zero) sign. The reason is two-fold. First we note that, since $\boldsymbol{\omega} \in \hat{\Omega}_n$ and therefore it satisfies Property P3, the entailed action-based homophily guarantees the required signed of the second factor of the claimed inequality. To confirm this claim, suppose for concreteness that $\boldsymbol{\omega}$ is such that $\sum_{i=1}^n s_i > 0$ and hence (by Property P3) $\bar{d}^+(\boldsymbol{\omega}) > \bar{d}^-(\boldsymbol{\omega})$. Then, denoting by $\zeta^+(\boldsymbol{\omega}) \equiv \sum_{d\in\mathbb{N}} \zeta_d^+(\boldsymbol{\omega})$ and $\zeta^-(\boldsymbol{\omega}) \equiv \sum_{d\in\mathbb{N}} \zeta_d^-(\boldsymbol{\omega})$ the aggregate frequencies of the population respectively playing actions $+1$ and -1, the fact that $\bar{d}^+(\boldsymbol{\omega}) \geq 1$ and $\bar{d}^+(\boldsymbol{\omega}) - \bar{d}^-(\boldsymbol{\omega}) > 0$ implies:

$$\begin{aligned}\sum_{d\in\mathbb{N}} \zeta_d^+(\boldsymbol{\omega})d - \zeta_d^-(\boldsymbol{\omega})d &= \zeta^+(\boldsymbol{\omega}) \cdot \bar{d}^+(\boldsymbol{\omega}) + \zeta^-(\boldsymbol{\omega}) \cdot \bar{d}^-(\boldsymbol{\omega}) \\ &= \bar{d}^+(\boldsymbol{\omega}) \left[\zeta^+(\boldsymbol{\omega}) - \zeta^-(\boldsymbol{\omega})\right] + \zeta^-(\boldsymbol{\omega}) \left[\bar{d}^+(\boldsymbol{\omega}) - \bar{d}^-(\boldsymbol{\omega})\right] \\ &\geq \zeta^+(\boldsymbol{\omega}) - \zeta^-(\boldsymbol{\omega}) = \frac{1}{n}\sum_{i=1}^{n} s_i > \delta > 0.\end{aligned} \tag{A.10}$$

To check now that $p^*(\boldsymbol{\omega})$, the second factor in the LHS of the inequality in (A.9), is also positive we start by noting that, since $\boldsymbol{\omega}$ is in $\hat{\Omega}_n^\delta$, we have that $\boldsymbol{\omega} \in \bar{H}_n^\delta$ and therefore $\frac{1}{n}\sum_{i=1}^n s_i > \delta > 0$. Thus, since we have chosen η to satisfy $\eta < \delta/2$, by virtue of (A.10) and the argument spelled out in stage 1 we have that $p^*(\boldsymbol{\omega}) > \delta/2$ and is therefore positive, as desired. The argument for the case where $\frac{1}{n}\sum_{i=1}^n s_i < 0$ is symmetric to the one just considered.

[28] Note that, being state $\boldsymbol{\omega}'$ adjacent to $\boldsymbol{\omega}$ and $G' = G$, their respective action profiles must have $s_j' = s_j$ for all $j \neq i$ while $s_i' = -s_i$.

3. Under the conditions explained in stage 2, we know that $|p^*(\boldsymbol{\omega})| > \delta/2$, with δ being chosen positive, independently of n. Furthermore, since $\boldsymbol{\omega}$ satisfies Property P1, as the population gets large the network degrees of the agents grow at an infinitesimally smaller rate than the population size n. Hence, given the (positive) parameters θ and ρ, if the population is large enough the expected payoffs of agent i must become dominated by global conformity-inducing payoffs (whose *relative* weight is modulated by $p^*(\boldsymbol{\omega}) \cdot \rho/\theta$), rather than local ones. This, in sum, implies that there exists some $\hat{n} \geq \hat{n}_1$ (as chosen in stage 1) such that, if $n \geq \hat{n}$, the expected change in agent i's payoffs from switching to action s_i' is positive if, and only if, $s_i' \cdot p^*(\boldsymbol{\omega}) > 0$ – or, equivalently (in view of stage 2), if, and only if, the new action s_i' is of the same sign as that of $\frac{1}{n}\sum_{j=1}^n s_j$, the average population action, as claimed is the statement in the Lemma. □

As advanced, the following counterpart of the preceding Lemma 2 can be established for the remainder term in the RHS of (A.7).

Lemma 3. *Given any $\delta > 0$, there exist some $\hat{n} \in \mathbb{N}$ and $\hat{\varphi} > 0$ such that, if $n \geq \hat{n}$ and $\varphi \leq \hat{\varphi}$, any adjustment taking place at a state $\boldsymbol{\omega} = (\mathbf{s}, G)$ in $\hat{\Omega}_n^\delta$ toward an adjacent state $\boldsymbol{\omega}' = (\mathbf{s}', G') \in \mathcal{A}_i(\boldsymbol{\omega})$ for some $i \in \mathcal{N}$ and $G' = G$ satisfies:*

$$\sum_{j=1}^n s_j' \left[\psi_j^{LIL}(\boldsymbol{\omega'}) - \psi_j^{LIL}(\boldsymbol{\omega})\right] > 0 \;\Leftrightarrow\; s_i' \sum_{j=1}^n s_j > 0. \tag{A.11}$$

Proof. Let $\boldsymbol{\omega} \in \hat{\Omega}_n^\delta$ be any prevailing state and suppose that agent i switches to action $s_i' \neq s_i$. Then, rewriting the second term in the RHS of (A.7) as follows: $\sum_{j=1}^n s_j' \left[\psi_j^{LIL}(\boldsymbol{\omega'}) - \psi_j^{LIL}(\boldsymbol{\omega})\right] = (p^*(\boldsymbol{\omega}') - p^*(\boldsymbol{\omega}))(s_i' + \sum_{j\neq i}^n s_j)$, we claim that there exists some $\hat{n}$ such that if $n \geq \hat{n}$,

$$\text{sign}\left[(p^*(\boldsymbol{\omega}') - p^*(\boldsymbol{\omega}))\left(s_i' + \sum_{j\neq i} s_j\right)\right] = \text{sign}\left[s_i'\left(s_i' + \sum_{j\neq i} s_j\right)\right] \tag{A.12}$$

$$= \text{sign}\left[s_i' \sum_{j=1}^n s_j\right] \tag{A.13}$$

where (A.12) follows from the fact that agent i's action adjustment toward action s_i' shifts the modified consensus beliefs $p^*(\boldsymbol{\omega}')$ in the direction of s_i', i.e. $s_i' \cdot (p^*(\boldsymbol{\omega}') - p^*(\boldsymbol{\omega})) > 0$, and (A.13) holds because, provided $\hat{n}$ above is chosen large enough, we must have: $s_i' + \sum_{j\neq i} s_j > 0 \Leftrightarrow \sum_{j=1}^n s_j > 0$. To see this, suppose first for concreteness that $\sum_{j=1}^n s_j > 0$. Then, since $\boldsymbol{\omega}$ is in $\hat{\Omega}_n^\delta$, it is in $\bar{H}_n^\delta$ as well. Hence we know that $\frac{1}{n}\sum_{j=1}^n s_j > \delta$ and therefore:

$$s_i' + \sum_{j\neq i}^n s_j = n\left(\frac{1}{n}\sum_{j=1}^n s_j\right) + s_i' - s_i \geq n\delta - 2 > 0, \tag{A.14}$$

provided $\delta > 2/n$, which obviously holds if $n \geq \hat{n}$ for a sufficiently large $\hat{n}$. A symmetric argument applies if $\sum_{j=1}^n s_j < 0$, thus completing the proof of the result. □

Lemmas 2 and 3 jointly yield the following straightforward corollary.

Corollary 2. *Let $\tilde{\Phi} : \Omega \to \mathbb{R}$ be the function defined by (17) and consider any $\boldsymbol{\omega} = (\mathbf{s}, G) \in \hat{\Omega}_n^\delta$ for some given δ. Then, there exist some $\hat{n} \in \mathbb{N}$ and $\hat{\varphi} > 0$ such that, if $n \geq \hat{n}$ and $\varphi \leq \hat{\varphi}$, any* *adjustment taking place at state* $\boldsymbol{\omega}$ *toward an adjacent state* $\boldsymbol{\omega}' = (\mathbf{s}', G') \in \mathcal{A}_i(\boldsymbol{\omega})$ *for some* $i \in \mathcal{N}$ *and* $G' = G$ *satisfies:*

$$\left[\tilde{\Phi}(\boldsymbol{\omega}') - \tilde{\Phi}(\boldsymbol{\omega})\right] \left[\pi_i(s_i', \mathbf{s}_{-i}, G; \psi_i^{LIL}(\boldsymbol{\omega})) - \pi_i(s_i, \mathbf{s}_{-i}, G; \psi_i^{LIL}(\boldsymbol{\omega}))\right] \geq 0. \tag{A.15}$$

Proof. Consider any given transition from some $\boldsymbol{\omega} \in \hat{\Omega}_n^\delta$ to an adjacent state $\boldsymbol{\omega}' \in \mathcal{A}_i(\boldsymbol{\omega})$ for some $i \in \mathcal{N}$. In combination, the characterizations provided by (A.8) and (A.11) imply that the induced change on the function $\tilde{\Phi}(\cdot)$, as specified in (A.7), has the first two (payoff difference) and the last term in the RHS of this latter equation both being of the same sign. Since its first two terms capture the change in the payoffs of agent i, we conclude that the payoffs of this agent and the function $\tilde{\Phi}(\cdot)$ experience a change of the same sign, as claimed. □

The previous corollary identifies conditions under which, for an arbitrary δ and corresponding states in $\hat{\Omega}_n^\delta$, the function $\tilde{\Phi}(\cdot)$ acts as an ordinal potential when the adjustment implementing the transition of an action revision. Next, we address a similar concern pertaining to link revision.

Lemma 4. *Given any* $\vartheta > 0$*, there exist some* $\hat{n} \in \mathbb{N}$ *such that, if* $n \geq \hat{n}$*, any adjustment taking place at a state* $\boldsymbol{\omega} = (\mathbf{s}, G)$ *in* $\hat{\Omega}_n^\delta$ *toward an adjacent state* $\boldsymbol{\omega}' = (\mathbf{s}', G') \in \mathcal{A}(\boldsymbol{\omega})$ *with* $\mathbf{s}' = \mathbf{s}$ *satisfies*[29] $\Xi = \left|\rho(n-1)\sum_{i=1}^n s_i(\psi(\boldsymbol{\omega}') - \psi(\boldsymbol{\omega}))\right| \leq \vartheta$.

Proof. From Lemma 1 we have that, for states $\boldsymbol{\omega}$ and $\boldsymbol{\omega}'$ as specified there, the beliefs $\psi_i(\boldsymbol{\omega})$ and $\psi_i(\boldsymbol{\omega}')$ commonly held by all $i \in \mathcal{N}$ satisfy: $\psi_i(\boldsymbol{\omega}') - \psi_i(\boldsymbol{\omega}) = \pm\frac{1}{2m+n}\left[\frac{1+d_k}{d_k}\, s_\ell + \frac{1+d_\ell}{d_\ell}\, s_k\right]$, where k and ℓ above are the single pair of agents whose link is changed, i.e. formed (+) or removed (−), in switching from $\boldsymbol{\omega}$ to $\boldsymbol{\omega}'$. This implies that $\Xi = \frac{\rho(n-1)}{2m+n}\left|\frac{1+d_k}{d_k}\, s_\ell + \frac{1+d_\ell}{d_\ell}\, s_k\right|$, where recall that $m = \frac{1}{2}\, n\, \bar{d}(\boldsymbol{\omega})$ is the number of links in the network G prevailing in state $\boldsymbol{\omega}$. Since this state is in $\hat{\Omega}_n$, by P1 we have that $\lim_{n\to\infty} \frac{\rho(n-1)}{2m+n} = 0$ while $\lim_{n\to\infty}\left|\frac{1+d_k}{d_k}\, s_\ell + \frac{1+d_\ell}{d_\ell}\, s_k\right| \leq 2$. Hence, for large enough n, Ξ can be made arbitrarily small, which yields the desired conclusion. □

As a counterpart of Lemma 2 (which applies to action adjustments), Lemma 4 above has some sharp implications on the changes on the value of the function $\tilde{\Phi}(\cdot)$ induced by link adjustments. They are stated in the following corollary.

Corollary 3. *Assume* $\zeta_1 \neq \theta \neq \zeta_2$ *and let* $\tilde{\Phi} : \Omega \to \mathbb{R}$ *be the function defined by (17). Then, there exist some* $\hat{n} \in \mathbb{N}$ *such that, if* $n \geq \hat{n}$*, any adjustment taking place at a state* $\boldsymbol{\omega} = (\mathbf{s}, G)$ *in* $\hat{\Omega}_n^\delta$ *toward an adjacent state* $\boldsymbol{\omega}' = (\mathbf{s}', G') \in \mathcal{A}(\boldsymbol{\omega})$ *with* $\mathbf{s}' = \mathbf{s}$ *satisfies:*

$$\left[\tilde{\Phi}(\boldsymbol{\omega}') - \tilde{\Phi}(\boldsymbol{\omega})\right] \left[\pi_i(\mathbf{s}, G'; \psi_i^{LIL}(\boldsymbol{\omega})) - \pi_i(\mathbf{s}, G; \psi_i^{LIL}(\boldsymbol{\omega}))\right] > 0. \tag{A.16}$$

for the single pair of agents i *and* j *for whom* $a_{ij} = a_{ij} \neq a'_{ij} = a'_{ij}$.

Proof. Denote $\varpi \equiv \min\{|\zeta_1|, |\zeta_2|\}$ to be the smallest impact that a link adjustment can have on the local-interaction payoffs of the two agents involved. Since we assume that $\zeta_1 \neq \theta \neq \zeta_2$, it follows that $\varpi > 0$, independently of the population size n. Thus, provided $G' \neq G$, we have that $|\pi_k(\mathbf{s}, G'; \psi_k^{LIL}(\boldsymbol{\omega})) - \pi_k(\mathbf{s}, G; \psi_k^{LIL}(\boldsymbol{\omega}))| > \varpi$ for the two agents $k = i, j$ whose link is affected by the adjustment considered. Now note that, by virtue of Lemma 4, we can guarantee that there exist some large enough $\hat{n}$ such that, if $n \geq \hat{n}$, then $\Xi < \varpi$. This implies that the last term in the RHS of Equation (A.7) is dominated (in absolute value) by the first two terms, the latter being

[29] In analogy with Footnote 28, again note that, being state $\boldsymbol{\omega}'$ adjacent to $\boldsymbol{\omega}$ with $\mathbf{s}' = \mathbf{s}$, their respective adjacency matrices must have $a'_{ij} = a_{ij}$ for all i and j except for a single pair of agents, k and ℓ, for whom $a'_{k\ell} \neq a_{k\ell}$.

precisely equal to the payoff change experienced by agents i and j. This readily establishes the second, thus establishes (A.16), as desired. □

Finally, to complete the ***proof of Proposition 4***, we combine the results established by Lemmas 1-4 and their Corollaries 1-3. We start by claiming that, by virtue of Assumption A.1, for any arbitrarily small ϵ_1 there exists some $\hat{n}_1$ such that if the population size $n \geq \hat{n}_1$, then in the long run (i.e. for large enough t) the randomly visited state $\boldsymbol{\omega}$ belongs to $\hat{\Omega}_n$ with a probability (given by μ_n^{LIL}, the invariant distribution of the process) that is no smaller than $1 - \epsilon_1$. Then we invoke Assumption A.2 to assert that, for any ϵ_2, again chosen arbitrarily small, there must exist some δ and some $\hat{n}_2$ such that if $n \geq \hat{n}_2$, $\mu_n^{LIL}(\bar{H}_n^{\delta}) \geq 1 - \epsilon_2$. It then follows that if we define $\epsilon = \epsilon_1 + \epsilon_2$ and $n \geq \max\{\hat{n}_1, \hat{n}_2\}$ there is probability no lower than $1 - \epsilon$ that a state $\boldsymbol{\omega}$ randomly visited in the long run belongs to the set $\hat{\Omega}_n^{\delta}$. Next, we invoke Lemmata 1-3 and their Corollaries 1-2 to argue that there exists some $\hat{n}_3 \geq \max\{\hat{n}_1, \hat{n}_2\}$ such that, if $n \geq \hat{n}_3$ and an action-revision opportunity arises at the aforementioned long-run state $\boldsymbol{\omega}$, then the entailed adjustment towards some adjacent state $\boldsymbol{\omega}' = (\mathbf{s}', G) \in \mathcal{A}(\boldsymbol{\omega})$ satisfies (A.15). Similarly, by further relying on Lemma 4 and its Corollary 3 we claim that that there exists some $\hat{n}_4 \geq \max\{\hat{n}_1, \hat{n}_2\}$ such that, if $n \geq \hat{n}_4$, (A.16) applies whenever the adjustment taking place at the state $\boldsymbol{\omega}$ involves a link. Thus, if we choose $\hat{n} = \max\{\hat{n}_3, \hat{n}_4\}$ the desired conclusion follows, i.e. the subset of states $\tilde{\Omega}_n$ that satisfy the payoff-potential alignment condition (19) has long-run probability $\mu_\eta^{LIL}(\tilde{\Omega}_n) \geq 1 - \epsilon$. This completes the proof of the proposition. □

Proof of Proposition 5. The proof of Proposition 5 is a straightforward extension of the proof of Proposition 3 with the additional requirement that beliefs are stationary. The details of the proof can be found in Supplementary Appendix C. □

Supplementary Appendices for "Social Networks and Collective Action in Large Populations: An application to the Egyptian Arab Spring"

Lachlan Deer[a], Chih-Sheng Hsieh[b], Michael D. König[c,d,e] and Fernando Vega-Redondo[f,g]

[a] *Tilburg University, Tilburg, The Netherlands.*
[b] *Department of Economics, National Taiwan University, Taipei, Taiwan.*
[c] *Tinbergen Institute and Department of Spatial Economics, VU Amsterdam, Amsterdam, The Netherlands.*
[d] *ETH Zurich, Swiss Economic Institute (KOF), Zurich, Switzerland.*
[e] *Centre for Economic Policy Research (CEPR), London, United Kingdom.*
[f] *Department of Decision Sciences, Bocconi University, Milan, Italy.*
[g] *Department of Economics, Universidad Carlos III, Madrid, Spain.*

B. Extensions

In the following we briefly illustrate how the simple theoretical model discussed in Subsection 2.1 could be extended along two important dimensions: In Subsection B.1 we allow for directed links, and in Subsection B.2 we consider heterogeneous linking costs that depend on the actions of the agents.

B.1. Directed Links

It is possible to consider a directed network. Assume for simplicity a constant linking cost. In this case the potential function for the GI environment needs to be modified as follows

$$\Phi(\mathbf{s}, G) = \sum_{i=1}^{n} \gamma_i s_i + \theta \sum_{i=1}^{n} \sum_{j=1}^{n} \left(\frac{1}{2} a_{ij} a_{ji} + a_{ij}(1 - a_{ji}) \right) s_i s_j + \frac{\rho}{2} \sum_{i=1}^{n} \sum_{j \neq i}^{n} s_i s_j - \kappa \sum_{i=1}^{n} s_i - m\zeta. \quad \text{(B.1)}$$

Here we consider undirected links, as it is standard in the social networks literature on peer effects, and we leave the detailed analysis of directed networks to future work.

B.2. Action-Specific Heterogeneous Linking Costs

Consider a linking cost between agents i and j given by $\zeta_{ij} = \zeta_1 d_i - \zeta_2 \sum_{j=1}^{n} a_{ij}(1 - s_i s_j)$ that allows for linking costs to be lower between agents choosing the same strategy. The corresponding payoff function in the GI environment is given by $\pi_i(\mathbf{s}, G) = \gamma_i s_i + (\theta + \zeta_2) \sum_{j=1}^{n} a_{ij} s_i s_j + \rho \sum_{j=1}^{n} s_j s_i - \kappa s_i - \zeta_1 d_i$. This is the same functional form as in (1) up to a shift of the parameter θ.

C. Proof of Proposition 5

In the following we compute an absorbing state of the Markov process formalizing the LIL scenario in the limit of $\eta \to \infty$ characterizing the stochastically stable states. In such an absorbing state $(\mathbf{s}, G, \mathbf{p})$, given the beliefs $\mathbf{p}$ agents do not have an incentive to change their actions, $\mathbf{s}$, or links, G. Because differences in the potential correspond to differences in payoffs, this is satisfied if the potential is maximized for such $(\mathbf{s}, G)$ given the beliefs $\mathbf{p}$. Conversely, given $(\mathbf{s}, G)$, the stationary belief updating in (8) must be satiesfied, that is, $p_i = f_i(\mathbf{s}, \mathbf{p}, G) = \varphi \frac{1}{d_i} \sum_{j=1}^{n} a_{ij} s_j + (1 - \varphi) \frac{1}{d_i+1}(p_i + \sum_{j=1}^{n} a_{ij} p_j)$ for all $i = 1, \ldots, n$. We then proceed by a guess and verify approach to check that the conditions for such a fixed point are satisfied.

We first consider the potential maximizing states $(\mathbf{s}, G)$ given the beliefs $\mathbf{p}$. The potential function can be written as:

$$\tilde{\Phi}(\mathbf{s}, G, \mathbf{p}) = \sum_{i=1}^{n} \gamma_i s_i + \frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{n} a_{ij}(\theta s_i s_j - \zeta_{ij}) + \rho(n-1) \sum_{i=1}^{n} p_i s_i - \kappa \sum_{i=1}^{n} s_i.$$

This can be rewritten as:

$$\tilde{\Phi}(\mathbf{s}, G, \mathbf{p}) = \tilde{\boldsymbol{\gamma}}^\top \mathbf{s} + \frac{\theta}{2} \mathbf{s}^\top \mathbf{A} \mathbf{s} - \frac{1}{2} \mathbf{u}^\top (\boldsymbol{\zeta} \odot \mathbf{A}) \mathbf{u},$$

where we have denoted by

$$\tilde{\gamma}_i = \gamma_i + \rho(n-1) p_i - \kappa.$$

For a given vector of beliefs, $\mathbf{p}$, the scalar product $\langle \tilde{\boldsymbol{\gamma}}, \mathbf{s} \rangle = \tilde{\boldsymbol{\gamma}}^\top \mathbf{s}$ is maximized for $s_i = \text{sign}(\tilde{\gamma}_i)$, and the quadratic form $\mathbf{s}^\top \mathbf{A} \mathbf{s}$ is maximized for $a_{ij} = 1$ iff $\text{sign}(s_i) = \text{sign}(s_j)$, or equivalently $\text{sign}(\tilde{\gamma}_i) = \text{sign}(\tilde{\gamma}_j)$ in the case of $\zeta_{ij} < \theta$. This implies that the stochastically stable network must be either complete, K_n, empty, $\overline{K}_n$, or composed of two cliques, $K_{n^+} \cup K_{n-n^+}$, where in each clique the agents choose the same actions.

Next, recall that the stationary beliefs Equation (8) can be written as:

$$p_i = \varphi \frac{1}{d_i} \sum_{j=1}^{n} a_{ij} s_j + (1-\varphi) \frac{1}{d_i + 1} (p_i + \sum_{j=1}^{n} a_{ij} p_j),$$

In a network where all connected agents choose the same action and have the same beliefs, this simplifies to:

$$p_i = \varphi s_i + (1-\varphi) p_i,$$

which is satisfied for

$$p_i = s_i.$$

When $p_i = s_i$ for all $i = 1, \ldots, n$ then the potential function can be written as

$$\tilde{\Phi}(\mathbf{s}, G) = \sum_{i=1}^{n} \left(\gamma_i + \rho(n-1) s_i - \kappa\right) s_i + \frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{n} a_{ij} \left(\theta s_i s_j - \zeta_1 + \frac{\zeta_1 - \zeta_2}{2} (1 - \gamma_i \gamma_j) \right).$$

Note that only the last term in this equation depends on the network (through the entries of the adjacency matrix elements a_{ij}). In particular, the term $\sum_{i=1}^{n} \sum_{j=1}^{n} a_{ij} s_i s_j$ is maximized over $s_i, s_j \in \{-1, +1\}$ for $a_{ij} = 1$ iff $s_i = s_j$. The term $\sum_{i=1}^{n} \sum_{j=1}^{n} a_{ij}(\theta s_i s_j - \zeta_1 + (\zeta_1 - \zeta_2)(1 - \gamma_i \gamma_j)/2$ is maximized over $s_i, s_j \in \{-1, +1\}$ for $a_{ij} = 1$ iff $s_i = s_j = \gamma_i = \gamma_j$ if $\zeta_1 < \theta < \zeta_2$ and $s_i = s_j$ if $\zeta_2 < \theta$. If $\theta < \zeta_1$ then $a_{ij} = 0$ and we obtain the empty network, $\overline{K}_n$.

To summarize, the candidate networks and action profiles that maximize the potential must be either complete, K_n, empty, $\overline{K}_n$, or composed of two disconnected cliques, $K_{n_1} \cup K_{n-n_1}$, in which all agents in the same clique chose the same action and have the same idiosyncratic preferences.

In the following we denote by $n_+ = \#(\{\gamma_i = 1 : i = 1, \dots, n\})$, $\nu_+ = n_+/n$ and define

$$\begin{aligned}\tilde{\theta}^* &= \zeta_2 + \frac{2(1-\kappa)}{n-n_+},\\ \tilde{\theta}^{**} &= \zeta_2 + \frac{2(1+\kappa)}{n_+},\\ \kappa^* &= 2\nu_+ - 1.\end{aligned}$$

W.l.o.g. we will further assume that $\kappa \geq 0$. Note also that $\kappa^* \leq 1$ and $\kappa^* \leq 0$ if $\nu_+ \leq 1/2$.

We first consider the case of $\theta < \zeta_1$. In this case the quasi potential maximizing network will be the empty network $\overline{K}_n$. When all agents choose the action $s_i = -1$ the quasi potential is given by

$$\tilde{\Phi}((-1, \dots, -1), \overline{K}_n) = n - 2n_+ + \rho n(n-1) + \kappa n.$$

When all agents choose the action $s_i = +1$ the quasi potential is given by

$$\tilde{\Phi}((+1, \dots, +1), \overline{K}_n) = n_+ - (n - n_+) + \rho n(n-1) - \kappa n.$$

When all agents choose the action $s_i = \gamma_i$ the quasi potential is given by

$$\tilde{\Phi}(\boldsymbol{\gamma}, \overline{K}_n) = n + \rho(n-1)n - \kappa(2n_+ - n).$$

The difference in the quasi potential between $s_i = -1$ and $s_i = \gamma_i$ is given by

$$\tilde{\Phi}((-1, \dots, -1), \overline{K}_n) - \tilde{\Phi}(\boldsymbol{\gamma}, \overline{K}_n) = 2(\kappa - 1)n_+,$$

which is positive for $\kappa > 1$ and negative for $\kappa < 1$. The difference in the quasi potentials between $s_i = +1$ and $s_i = \gamma_i$ is given by

$$\tilde{\Phi}((+1, \dots, +1), \overline{K}_n) - \tilde{\Phi}(\boldsymbol{\gamma}, \overline{K}_n) = -2(n - n_+)(1 + \kappa)$$

which is negative for all parameter choices. Moreover, the difference in the quasi potentials between $s_i = +1$ and $s_i = -1$:

$$\tilde{\Phi}((+1, \dots, +1), \overline{K}_n) - \tilde{\Phi}((-1, \dots, -1), \overline{K}_n) = -2((\kappa + 1)n - 2n_+),$$

which is positive if

$$\kappa < \kappa^* = 2\nu_+ - 1 \leq 1.$$

Thus if $\theta < \zeta_1$, the quasi potential maximizing state is given by the empty network $\overline{K}_n$ where

1. if $\nu_+ < 1/2$ and
   (a) $\kappa > 1$, then all agents choose the action $s_i = -1$.
   (b) $\kappa < 1$, then all agents choose the action $s_i = \gamma_i$.
2. if $\nu_+ > 1/2$,
   (a) $\kappa > \kappa^*$ and
      i. $\kappa > 1$, then all agents choose the action $s_i = -1$.
      ii. $\kappa < 1$, then all agents choose the action $s_i = \gamma_i$.
   (b) $\kappa < \kappa^*$, then all agents choose the action $s_i = +1$.

Next we consider the case of $\theta > \zeta_1$. The quasi potential maximizing network is either complete, K_n, or composed of two cliques, $K_{n_+} \cup K_{n-n_+}$. Further, note that

$$\tilde{\Phi}((-1,\dots,-1),K_n) - \tilde{\Phi}((+1,\dots,+1),K_n) = 2((\kappa+1)n - 2n_+)$$

which is increasing in κ and negative if

$$\kappa < \kappa^* = 2\nu^* - 1,$$

where, $\kappa^* \leq 1$ with $\kappa^* = 1$ if $\nu_+ = 1$ and $\kappa^* < 0$ if $\nu_+ < 1/2$. The latter implies that for $\nu_+ < 1/2$ we have that $\tilde{\Phi}((-1,\dots,-1),K_n) > \tilde{\Phi}((+1,\dots,+1),K_n)$ for any value of $\kappa \geq 0$. We thus need to distinguish between the cases $\nu_+ < 1/2$ and $\nu_+ > 1/2$.

Case 1: $\nu_+ < 1/2$. The potential for the complete Graph K_n with all agents choosing action $s_i = -1$ is given by

$$\begin{aligned}\tilde{\Phi}((-1,\dots,-1),K_n) = (n-2n_+) + \frac{1}{2}(n(n-1) - 2n_+(n-n_+))(\theta-\zeta_1) \\ + n_+(n-n_+)(\theta-\zeta_2) + \rho n(n-1) + \kappa n.\end{aligned}$$

The potential for the union of cliques $K_{n_+} \cup K_{n-n_+}$ with all agents choosing action $s_i = \gamma_i$ is given by

$$\tilde{\Phi}(\boldsymbol{\gamma}, K_{n_+} \cup K_{n-n_+}) = n + \frac{1}{2}(n(n-1) - 2n_+(n-n_+))(\theta-\zeta_1) + \rho(n-1)n - \kappa(2n_+ - n).$$

The potential difference between the complete network K_n in which all agents choose action $s_i = -1$) and the union of cliques $K_{n_+} \cup K_{n-n_+}$ in which the agents choose their idiosyncratic preferences $s_i = \gamma_i$) is given by

$$\tilde{\Phi}((-1,\dots,-1),K_n) - \tilde{\Phi}(\boldsymbol{\gamma}, K_{n_+} \cup K_{n-n_+}) = n_+(2(\kappa-1) + n(\theta-\zeta_2) + n_+(\zeta_2-\theta)).$$

Setting this to zero to find $\tilde{\theta}^*$:

$$\tilde{\theta}^* = \zeta_2 + \frac{2(1-\kappa)}{n-n_+}.$$

Moreover,

$$\tilde{\Phi}((-1,\dots,-1),K_n) - \tilde{\Phi}((-1,\dots,-1),K_{n_+} \cup K_{n-n_+}) = n_+(\theta-\zeta_2)(n-n_+),$$

which is positive for $\theta > \zeta_2$. Further

$$\tilde{\Phi}((-1,\dots,-1),K_{n_+} \cup K_{n-n_+}) - \tilde{\Phi}(\boldsymbol{\gamma}, K_{n_+} \cup K_{n-n_+}) = -2n_+(1-\kappa),$$

which is negative for $\kappa < 1$. We can then conclude that:

1. If $\theta > \tilde{\theta}^*$ and
   (a) $\theta > \zeta_2$ then the potential function of the complete graph K_n with $s_i = -1$ is higher than that of the union of cliques $K_{n_+} \cup K_{n-n_+}$ with $s_i = -1$ or $s_i = \gamma_i$. Therefore, the stochastically stable state is the complete graph K_n with $s_i = -1$.
   (b) $\theta < \zeta_2$ then the potential function of the union of cliques $K_{n_+} \cup K_{n-n_+}$ with $s_i = -1$ is higher than that of the complete graph K_n with $s_i = -1$. Therefore, the stochastically

stable state is the union of cliques $K_{n_+} \cup K_{n-n_+}$ with $s_i = -1$.

2. If $\theta < \tilde{\theta}^*$ and

   (a) $\kappa < 1$, then the potential function of the union of cliques $K_{n_+} \cup K_{n-n_+}$ with $s_i = \gamma_i$ is higher than that of the complete graph K_n with $s_i = -1$ or the union of cliques $K_{n_+} \cup K_{n-n_+}$ with $s_i = -1$. Therefore, the stochastically stable state is the union of cliques $K_{n_+} \cup K_{n-n_+}$ with $s_i = \gamma_i$.

   (b) $\kappa > 1$, then the potential function of the union of cliques $K_{n_+} \cup K_{n-n_+}$ with $s_i = -1$ is higher than that of the complete graph K_n with $s_i = -1$ or the union of cliques $K_{n_+} \cup K_{n-n_+}$ with $s_i = \gamma_i$. Therefore, the stochastically stable state is the union of cliques $K_{n_+} \cup K_{n-n_+}$ with $s_i = -1$.

Case 2: $\nu_+ > 1/2$. We first note that

$$\tilde{\Phi}((-1,\dots,-1), K_{n_+} \cup K_{n-n_+}) - \tilde{\Phi}((+1,\dots,+1), K_{n_+} \cup K_{n-n_+}) = 2n(\kappa - \kappa^*),$$

which is positive for $\kappa > \kappa^*$ and negative for $\kappa < \kappa^*$. We thus need to further distinguish the cases $\kappa > \kappa^*$ and $\kappa < \kappa^*$.

1. **Subcase 2.1:** $\kappa > \kappa^*$. The quasi potential for the complete graph K_n in which all agents choose action $s_i = -1$ is given by

$$\begin{aligned}\tilde{\Phi}((-1,\dots,-1), K_n) = (n - 2n_+) + \frac{1}{2}(n(n-1) - 2n_+(n-n_+))(\theta - \zeta_1) \\ + n_+(n-n_+)(\theta - \zeta_2) + \rho(n-1)n + \kappa n.\end{aligned}$$

   The quasi potential for the union of cliques $K_{n_+} \cup K_{n-n_+}$ in which all agents choose action $s_i = \gamma_i$ is given by

$$\tilde{\Phi}(\boldsymbol{\gamma}, K_{n_+} \cup K_{n-n_+}) = n + \frac{1}{2}(n(n-1) - 2n_+(n-n_+))(\theta - \zeta_1) + \rho(n-1)n - \kappa(2n_+ - n).$$

   The difference between the quasi potentials for the complete graph K_n (all $s_i = -1$) and the union of cliques $K_{n_+} \cup K_{n-n_+}$ ($s_i = \gamma_i$) is given by

$$\tilde{\Phi}((-1,\dots,-1), K_n) - \tilde{\Phi}(\boldsymbol{\gamma}, K_{n_+} \cup K_{n-n_+}) = n_+(2(\kappa - 1) + n(\theta - \zeta_2) + n_+(\zeta_2 - \theta)).$$

   Setting this to zero to find $\tilde{\theta}^*$:

$$\tilde{\theta}^* = \zeta_2 + \frac{2(1-\kappa)}{n - n_+}.$$

   Moreover,

$$\tilde{\Phi}((-1,\dots,-1), K_n) - \tilde{\Phi}((-1,\dots,-1), K_{n_+} \cup K_{n-n_+}) = n_+(\theta - \zeta_2)(n - n_+),$$

   which is zero if $\theta = \zeta_2$. Further,

$$\tilde{\Phi}((-1,\dots,-1), K_{n_+} \cup K_{n-n_+}) - \tilde{\Phi}(\boldsymbol{\gamma}, K_{n_+} \cup K_{n-n_+}) = -2n_+(1-\kappa),$$

   which is negative for $\kappa < 1$. We then can conclude the following:

   (a) If $\theta > \tilde{\theta}^*$ and

i. $\theta > \zeta_2$ then the quasi potential function of the complete graph K_n with $s_i = -1$ is higher than that of the union of cliques $K_{n_+} \cup K_{n-n_+}$ with $s_i = -1$ or $s_i = \gamma_i$. Therefore, the stochastically stable state is the complete graph K_n with $s_i = -1$.

ii. $\theta < \zeta_2$ then the quasi potential function of the union of cliques $K_{n_+} \cup K_{n-n_+}$ with $s_i = -1$ is higher than that of the complete graph K_n with $s_i = -1$ or the union of cliques $K_{n_+} \cup K_{n-n_+}$ with $s_i = \gamma_i$. Therefore, the stochastically stable state is the union of cliques $K_{n_+} \cup K_{n-n_+}$ with $s_i = -1$.

(b) If $\theta < \tilde{\theta}^*$ and

i. $\kappa < 1$, then the quasi potential function of the union of cliques $K_{n_+} \cup K_{n-n_+}$ with $s_i = \gamma_i$ is higher than that of the complete graph K_n with $s_i = -1$ or the union of cliques $K_{n_+} \cup K_{n-n_+}$ with $s_i = -1$. Therefore, the stochastically stable state is the union of cliques $K_{n_+} \cup K_{n-n_+}$ with $s_i = \gamma_i$.

ii. $\kappa > 1$, then the quasi potential function of the union of cliques $K_{n_+} \cup K_{n-n_+}$ with $s_i = -1$ is higher than that of the complete graph K_n with $s_i = -1$ or the union of cliques $K_{n_+} \cup K_{n-n_+}$ with $s_i = \gamma_i$. Therefore, the stochastically stable state is the union of cliques $K_{n_+} \cup K_{n-n_+}$ with $s_i = -1$.

2. **Subcase 2.2:** $\kappa < \kappa^*$. The quasi potential for the complete graph K_n with all agents choosing action $s_i = +1$ is given by

$$\tilde{\Phi}((+1,\ldots,+1),K_n) = (2n_+ - n) + \frac{1}{2}(n(n-1) - 2n_+(n-n_+))(\theta - \zeta_1) + n_+(n-n_+)(\theta-\zeta_2) + \rho(n-1)n - \kappa n.$$

The quasi potential in the union of cliques $K_{n_+} \cup K_{n-n_+}$ with all agents choosing action $s_i = \gamma_i$ is given by

$$\tilde{\Phi}(\boldsymbol{\gamma}, K_{n_+} \cup K_{n-n_+}) = n + \frac{1}{2}(n(n-1) - 2n_+(n-n_+))(\theta-\zeta_1) + \rho(n-1)n - \kappa(2n_+ - n).$$

The quasi potential in the union of cliques $K_{n_+} \cup K_{n-n_+}$ with all agents choosing action $s_i = +1$ is

$$\tilde{\Phi}((+1,\ldots,+1), K_{n_+} \cup K_{n-n_+}) = (2n_+ - n) + \frac{1}{2}(n(n-1) - 2n_+(n-n_+))(\theta-\zeta_1) + \rho(n-1)n - \kappa n.$$

The difference between the quasi potentials of the complete graph K_n (all $s_i = +1$) and the union of cliques $K_{n_+} \cup K_{n-n_+}$ (all $s_i = \gamma_i$) is then

$$\tilde{\Phi}((+1,\ldots,+1),K_n) - \tilde{\Phi}(\boldsymbol{\gamma}, K_{n_+} \cup K_{n-n_+}) = -(n-n_+)(2(\kappa+1) + n_+(\zeta_2 - \theta)).$$

Setting this to zero we obtain the threshold:

$$\tilde{\theta}^{**} = \zeta_2 + \frac{2(\kappa+1)}{n_+}.$$

Moreover,

$$\tilde{\Phi}((+1,\ldots,+1),K_n) - \tilde{\Phi}((+1,\ldots,+1), K_{n_+} \cup K_{n-n_+}) = n_+(\theta - \zeta_2)(n - n_+).$$

which is zero if $\theta = \zeta_2$. Further,

$$\tilde{\Phi}((+1,\ldots,+1), K_{n_+} \cup K_{n-n_+}) - \tilde{\Phi}(\boldsymbol{\gamma}, K_{n_+} \cup K_{n-n_+}) = -2(n-n_+)(\kappa+1),$$

which is negative for all parameter values. We can summarize the above cases as follows:

(a) If $\theta > \tilde{\theta}^{**}$ (which also implies that $\theta > \zeta_2$) then the potential function of the complete graph K_n with $s_i = +1$ is higher than that of the union of cliques $K_{n_+} \cup K_{n-n_+}$ with $s_i = \gamma_i$ or $s_i = +1$. Therefore, the stochastically stable state is the complete graph K_n with $s_i = +1$.

(b) If $\theta < \tilde{\theta}^{**}$ then the potential function of the union of cliques $K_{n_+} \cup K_{n-n_+}$ with $s_i = \gamma_i$ is higher than that of the complete graph K_n with $s_i = +1$ or the union of cliques $K_{n_+} \cup K_{n-n_+}$ with $s_i = +1$. Therefore, the stochastically stable state is the union of cliques $K_{n_+} \cup K_{n-n_+}$ with $s_i = \gamma_i$.

Thus, if $\theta > \zeta_1$, the stochastically stable state is either complete K_n or composed of two cliques $K_{n_+} \cup K_{n-n_+}$. More precisely,

1. if $\nu_+ < 1/2$,
   (a) $\theta > \tilde{\theta}^*$ and
      i. $\theta > \zeta_2$, then the stochastically stable state is the complete graph K_n with $s_i = -1$.
      ii. $\theta < \zeta_2$, then the stochastically stable state is the union of cliques $K_{n_+} \cup K_{n-n_+}$ with $s_i = -1$.
   (b) $\theta < \tilde{\theta}^*$ and
      i. $\kappa < 1$, then the stochastically stable state is the union of two cliques $K_{n_+} \cup K_{n-n_+}$ with $s_i = \gamma_i$.
      ii. $\kappa > 1$, then the stochastically stable state is the union of two cliques $K_{n_+} \cup K_{n-n_+}$ with $s_i = \gamma_i$.
2. if $\nu_+ > 1/2$,
   (a) $\kappa > \kappa^*$,
      i. $\theta > \tilde{\theta}^*$ and
         - $\theta > \zeta_2$, then the stochastically stable state is the complete graph K_n with $s_i = -1$.
         - $\theta < \zeta_2$, then the stochastically stable state is the union of cliques $K_{n_+} \cup K_{n-n_+}$ with $s_i = -1$.
      ii. $\theta < \tilde{\theta}^*$ and
         - $\kappa < 1$, then the stochastically stable state is the union of two cliques $K_{n_+} \cup K_{n-n_+}$ with $s_i = \gamma_i$.
         - $\kappa > 1$, then the stochastically stable state is the union of two cliques $K_{n_+} \cup K_{n-n_+}$ with $s_i = -1$.
   (b) $\kappa < \kappa^*$ and
      i. $\theta > \tilde{\theta}^{**}$ then the stochastically stable state is the complete graph K_n with $s_i = +1$.
      ii. $\theta < \tilde{\theta}^{**}$, then the stochastically stable state is the union of two cliques $K_{n_+} \cup K_{n-n_+}$ with $s_i = \gamma_i$.

This proves Proposition E.2. Next, note that for $n \to \infty$ we get

$$\lim_{n\to\infty} \theta^* = \lim_{n\to\infty} \theta^{**} = \zeta_2.$$

Hence, when $n \to \infty$ then if $\theta > \zeta_1$, the stochastically stable state is either complete K_n or composed of two cliques $K_{n_+} \cup K_{n-n_+}$. More precisely,

1. if $\nu_+ < 1/2$,
   (a) $\theta > \zeta_2$, then the stochastically stable state is the complete graph K_n with $s_i = -1$.
   (b) $\theta < \zeta_2$ and
      i. $\kappa < 1$, then the stochastically stable state is the union of two cliques $K_{n_+} \cup K_{n-n_+}$ with $s_i = \gamma_i$.
      ii. $\kappa > 1$, then the stochastically stable state is the union of two cliques $K_{n_+} \cup K_{n-n_+}$ with $s_i = -1$.
2. if $\nu_+ > 1/2$,
   (a) $\kappa > \kappa^*$,
      i. $\theta > \zeta_2$, then the stochastically stable state is the complete graph K_n with $s_i = -1$.
      ii. $\theta < \zeta_2$ and
         - $\kappa < 1$, then the stochastically stable state is the union of two cliques $K_{n_+} \cup K_{n-n_+}$ with $s_i = \gamma_i$.
         - $\kappa > 1$, then the stochastically stable state is the union of two cliques $K_{n_+} \cup K_{n-n_+}$ with $s_i = -1$.
   (b) $\kappa < \kappa^*$ and
      i. $\theta > \zeta_2$, then the stochastically stable state is the complete graph K_n with $s_i = +1$.
      ii. $\theta < \zeta_2$, then the stochastically stable state is the union of two cliques $K_{n_+} \cup K_{n-n_+}$ with $s_i = \gamma_i$.

A concise summary of the above cases in the large population limit ($n \to \infty$) is stated in Proposition 5. This concludes the proof.

D. Finite Noise Equilibrium Characterization

In this section we analyze the stationary states in the case of finite noise (while the stochastically stable state characterizations in Propositions 3 and 5 cover only the case of vanishing noise, i.e., the limit of $\eta \to \infty$). For concreteness, we focus on two especially relevant statistics, the average connectivity (network degree) and the average action, as they depend on the linking costs ζ ($= \zeta_1 = \zeta_2$) and the noise parameter η.

D.1. Global Information

We start by characterizing the expected number of links induced by the distribution $\mu_\eta^{GI}(\cdot)$.

Proposition D.1. *Assume homogeneous linking costs, $\zeta_1 = \zeta_2 = \zeta$. Then the expected number of links in the stationary state is given by*

$$\mathbb{E}^\eta(m) = \frac{1}{\mathcal{Z}^\eta} \sum_{k=0}^{n} \sum_{j=0}^{\min\{k,n_+\}} \binom{n_+}{j} \binom{n-n_+}{k-j} e^{\eta(2k-n)} e^{\eta\left(\rho\left(2l(k,j)-\binom{n}{2}\right)\right)-\kappa(n-2(n_++k-2j))}$$
$$\times \left(1+e^{\eta(\theta-\zeta)}\right)^{l(k,j)} \left(1+e^{-\eta(\theta+\zeta)}\right)^{\binom{n}{2}-l(k,j)} \left(\frac{l(k,j)}{1+e^{-\eta(\theta-\zeta)}} + \frac{\binom{n}{2}-l(k,j)}{1+e^{\eta(\theta+\zeta)}}\right), \tag{D.1}$$

where $l(k,j)$ is given by

$$l(k,j) = \frac{n^2 + (2(2j-k)-1)n + 2(2j-k)^2 - 2(n+2(2j-k)-n_+)n_+}{2}, \tag{D.2}$$

$n_+ = \#(\{\gamma_i = 1 : i = 1, \ldots, n\})$, *and we have that* $\lim_{\zeta \to \infty} \mathbb{E}^\eta(m) = 0$.[30]

Before proceeding with the proof of Proposition D.1 we state three useful lemmas that will be needed later. In these lemmas we assume that $\zeta_1 = \zeta_2 = \zeta$.

Lemma D.1. *Assume that $\zeta_1 = \zeta_2 = \zeta$. The marginal distribution of the action levels, $\mathbf{s} \in \mathbf{S} = \{-1,+1\}^n$, is given by*

$$\mu_\eta^{GI}(\mathbf{s}) = \frac{1}{\mathcal{Z}^\eta} e^{\eta \mathcal{H}^\eta(\mathbf{s})}, \tag{D.3}$$

where we have denoted by

$$\mathcal{H}^\eta(\mathbf{s}) \equiv \sum_{i=1}^n \left(\left(\gamma_i - \kappa + \frac{\rho}{2} \sum_{j \neq i}^n s_j \right) s_i + \sum_{j=i+1}^n \left(\frac{1}{\eta} \ln \left(1 + e^{\eta(\theta s_i s_j - \zeta)} \right) \right) \right), \tag{D.4}$$

and the normalizing constant is given by

$$\mathcal{Z}^\eta = \sum_{\mathbf{s} \in \{-1,+1\}^n} e^{\eta \mathcal{H}^\eta(\mathbf{s})}. \tag{D.5}$$

Proof of Lemma D.1. We first compute the *partition function* (cf. Grimmett, 2010; Wainwright and Jordan, 2008), which appears as the denominator in (15), explicitly. We have that

$$\mathcal{Z}^\eta \equiv \sum_{G \in \mathcal{G}^n} \sum_{\mathbf{s} \in \{-1,+1\}^n} e^{\eta \Phi(\mathbf{s},G)} = \sum_{\mathbf{s} \in \{-1,+1\}^n} e^{\eta \sum_{i=1}^n \left(\gamma_i - \kappa + \frac{\rho}{2} \sum_{j \neq i}^n s_j \right) s_i} \prod_{i=1}^n \prod_{j=i+1}^n \left(1 + e^{\eta(\theta s_i s_j - \zeta)} \right), \tag{D.6}$$

where we have used the fact that

$$\sum_{G \in \mathcal{G}^n} e^{\sum_{i<j}^n a_{ij} \sigma_{ij}} = \prod_{i=1}^n \prod_{j=i+1}^n \left(1 + e^{\sigma_{ij}} \right), \tag{D.7}$$

for any real and symmetric $\sigma_{ij} = \sigma_{ji}$. Introducing the *Hamiltonian* (cf. Grimmett, 2010)

$$\mathcal{H}^\eta(\mathbf{s}) \equiv \sum_{i=1}^n \left(\left(\gamma_i - \kappa + \frac{\rho}{2} \sum_{j \neq i}^n s_j \right) s_i + \sum_{j=i+1}^n \left(\frac{1}{\eta} \ln \left(1 + e^{\eta(\theta s_i s_j - \zeta)} \right) \right) \right), \tag{D.8}$$

we can write the partition function as follows $\mathcal{Z}^\eta = \sum_{\mathbf{s} \in \{-1,+1\}^n} e^{\eta \mathcal{H}^\eta(\mathbf{s})}$. With the Hamiltonian we

[30] An explicit expression for the partition function $\mathcal{Z}^\eta$ can be found in Lemma D.3 in Appendix A.

can write the marginal distribution as follows

$$\mu_\eta^{GI}(\mathbf{s}) = \frac{1}{\mathcal{Z}^\eta} \sum_{G \in \mathcal{G}^n} e^{\eta \Phi(\mathbf{s},G)} = \frac{1}{\mathcal{Z}_n^\eta} e^{\eta \sum_{i=1}^n \left(\gamma_i - \kappa + \frac{\rho}{2} \sum_{j \neq i}^n s_j\right) s_i} \prod_{i=1}^n \prod_{j=i+1}^n \left(1 + e^{\eta(\theta s_i s_j - \zeta)}\right) = \frac{1}{\mathcal{Z}^\eta} e^{\eta \mathcal{H}^\eta(\mathbf{s})}, \tag{D.9}$$

where $\mathcal{H}^\eta(\mathbf{s})$ has been defined in (D.8). □

Lemma D.2. *Assume that* $\zeta_1 = \zeta_2 = \zeta$. *Conditional on the action profile,* $\mathbf{s} \in \mathbf{S} \in \{-1,+1\}^n$, *the probability of observing the network* G *is given by* $\mu_\eta^{GI}(G|\mathbf{s}) = \prod_{i=1}^n \prod_{j=i+1}^n p_{ij}(s_i,s_j)^{a_{ij}} (1 - p_{ij}(s_i,s_j))^{1-a_{ij}}$, *where*

$$p_{ij}(s_i,s_j) = \frac{e^{\eta(\theta s_i s_j - \zeta)}}{1 + e^{\eta(\theta s_i s_j - \zeta)}}. \tag{D.10}$$

Proof of Lemma D.2. With the marginal distribution from (D.3) we can write the conditional distribution as

$$\begin{aligned}
\mu_\eta^{GI}(G|\mathbf{s}) = \frac{\mu_\eta^{GI}(\mathbf{s},G)}{\mu_\eta^{GI}(\mathbf{s})} &= \frac{e^{\eta\left(\sum_{i=1}^n \left(\gamma_i - \kappa + \frac{\rho}{2}\sum_{j\neq i}^n s_j\right)s_i + \frac{\theta}{2}\sum_{i=1}^n \sum_{j=1}^n a_{ij} s_i s_j - m\zeta\right)}}{e^{\eta \sum_{i=1}^n \left(\gamma_i - \kappa + \frac{\rho}{2}\sum_{j\neq i}^n s_j\right)s_i} \prod_{i=1}^n \prod_{j=i+1}^n \left(1 + e^{\eta(\theta s_i s_j - \zeta)}\right)} \\
&= \frac{e^{\eta \sum_{i<j}^n a_{ij}(\theta s_i s_j - \zeta)}}{\prod_{i=1}^n \prod_{j=i+1}^n \left(1 + e^{\eta(\theta s_i s_j - \zeta)}\right)} = \prod_{i<j}^n \frac{e^{\eta a_{ij}(\theta s_i s_j - \zeta)}}{1 + e^{\eta(\theta s_i s_j - \zeta)}} \\
&= \prod_{i<j}^n \left(\frac{e^{\eta(\theta s_i s_j - \zeta)}}{1 + e^{\eta(\theta s_i s_j - \zeta)}}\right)^{a_{ij}} \left(1 - \frac{e^{\eta(\theta s_i s_j - \zeta)}}{1 + e^{\eta(\theta s_i s_j - \zeta)}}\right)^{1-a_{ij}} \\
&= \prod_{i<j}^n p_{ij}(s_i,s_j)^{a_{ij}} (1 - p_{ij}(s_i,s_j))^{1-a_{ij}}.
\end{aligned} \tag{D.11}$$

Hence, conditional on the action choices $\mathbf{s}$, we obtain the likelihood of an *inhomogeneous random graph* with link probability (Bollobas et al., 2007): $p_{ij}(s_i,s_j) = e^{\eta(\theta s_i s_j - \zeta)}/(1 + e^{\eta(\theta s_i s_j - \zeta)})$. □

In the following we provide an explicit computation of the partition function introduced in (D.6).

Lemma D.3. *Assume that* $\zeta_1 = \zeta_2 = \zeta$. *Then the partition function,* $\mathcal{Z}^\eta = \sum_{G \in \mathcal{G}^n} \sum_{\mathbf{s} \in \{-1,+1\}^n} e^{\eta \Phi(\mathbf{s},G)}$, *is given by*

$$\begin{aligned}
\mathcal{Z}^\eta = \sum_{k=0}^n \sum_{j=0}^{\min\{k,n_+\}} \binom{n_+}{j} \binom{n-n_+}{k-j} e^{\eta(2k-n)} e^{\eta\left(\rho\left(2l(k,j) - \binom{n}{2}\right)\right) - \kappa(n - 2(n_+ + k - 2j))} \\
\times \left(1 + e^{\eta(\theta - \zeta)}\right)^{l(k,j)} \left(1 + e^{-\eta(\theta + \zeta)}\right)^{\binom{n}{2} - l(k,j)},
\end{aligned} \tag{D.12}$$

where $l(k,j) = \frac{1}{2}(n^2 + (2(2j-k)-1)n + 2(2j-k)^2 - 2(n + 2(2j-k) - n_+)n_+)$, *and* $n_+ = \#(\{\gamma_i = 1 : i = 1, \ldots, n\})$.

Note that, while the evaluation of the partition function in (D.6) requires the computation of a sum with 2^n terms, the partition function in (D.12) requires the evaluation of only $\frac{1}{2}(n_+ +$

$1)(2(n+1)-n_+) = O(n)$ terms. With (D.12) the marginal distribution $\mu_\eta^{GI}(\mathbf{s})$ in (D.3) can then be efficiently computed.

Proof of Lemma D.3. Assume w.l.o.g. that the agents are ordered such that $\gamma_1 = \dots \gamma_{n_+} = +1$ and $\gamma_{n_++1} = \dots \gamma_n = -1$, with $0 \leq n_+ \leq n$. Let us consider all configurations $\mathbf{s} \in \{-1,+1\}^n$ for which there $k = 0, \dots, n$ agents with $s_i = \gamma_i$. For a given k, there are $\binom{n_+}{j}$ ways to select j agents from n_+ choosing $s_i = \gamma_i = +1$, and there are $\binom{n-n_+}{k-j}$ ways to select $k-j$ agents from n_- choosing $s_i = \gamma_i = -1$, for each $j = 0, \dots, \min\{k, n_+\}$. Hence, there are $\sum_{j=0}^{\min\{k,n_+\}} \binom{n_+}{j}\binom{n-n_+}{k-j}$ ways to obtain alignments of $\boldsymbol{\gamma}$ and $\mathbf{s}$ such that $\sum_{i=1}^n s_i\gamma_i = k-(n-k) = 2k-n$.

Next, we consider the products $s_i s_j$. Since all the j agents in n_+ with $s_i = +1$ choose the same action $+1$, and all the $k-j$ agents in n_- with $s_i = -1$ choose the same action -1 we obtain $l(k,j) = \binom{j}{2} + \binom{k-j}{2} + \binom{n_+-j}{2} + \binom{n-n_+-(k-j)}{2} + (n_+-j)(k-j) + j(n-n_+-(k-j))$ pairs whose product of actions gives $s_i s_j = +1$. The first term in the equation above counts all pairs of agents with action $+1$ in the first set (with all $\gamma_i = +1$), the second all pairs of agents with action -1 in the second set (with all $\gamma_i = -1$), the third term the pairs of agents with action -1 in the first set (with all $\gamma_i = +1$), the fourth term the pairs of agents with action $+1$ in the second set (with all $\gamma_i = -1$), the fifth term counts the pairs with agents in the first set who choose action -1 and the agents in the second set who chose action -1, while the last term counts the pairs with agents in the first set who choose action $+1$ and agents in the second set who also choose action $+1$.

We can further simplify $l(k,j)$ to $l(k,j) = \frac{1}{2}(n^2 + (2(2j-k)-1)n + 2(2j-k)^2 - 2(n+2(2j-k)-n_+)n_+)$. Then we can write

$$
\begin{aligned}
\mathcal{Z}^\eta = \sum_{\mathbf{s}\in\{-1,+1\}^n} e^{\eta\mathcal{H}^\eta(\mathbf{s})} &= \sum_{k=0}^{n}\sum_{j=0}^{\min\{k,n_+\}} \binom{n_+}{j}\binom{n-n_+}{k-j} \exp\left\{ \eta \left[(2k-n) \right.\right. \\
&\quad -\kappa\left(n - 2(n_+ + k - 2j)\right) + \rho\left(2l(k,j) - \binom{n}{2}\right) \\
&\quad \left.\left. + \frac{l(k,j)}{\eta}\ln\left(1+e^{\eta(\theta-\zeta)}\right) + \frac{\binom{n}{2} - l(k,j)}{\eta}\ln\left(1+e^{-\eta(\theta+\zeta)}\right)\right]\right\} \\
&= \sum_{k=0}^{n}\sum_{j=0}^{\min\{k,n_+\}} \binom{n_+}{j}\binom{n-n_+}{k-j} e^{\eta(2k-n)} e^{\eta\left(\rho\left(2l(k,j)-\binom{n}{2}\right)\right)-\kappa(n-2(n_++k-2j))} \\
&\quad \times \left(1+e^{\eta(\theta-\zeta)}\right)^{l(k,j)} \left(1+e^{-\eta(\theta+\zeta)}\right)^{\binom{n}{2}-l(k,j)},
\end{aligned}
$$

where $n_+ = \#(\{\gamma_i = 1 : i = 1, \dots, n\})$. □

Proof of Proposition D.1. With the partition function in Lemma D.3 we can compute the expected number of links, m, as follows

$$
\mathbb{E}^\eta(m) = \sum_{G\in\mathcal{G}^n}\sum_{\mathbf{s}\in\{-1,+1\}^n} m\mu_\eta^{GI}(\mathbf{s},G) = \frac{1}{\mathcal{Z}^\eta}\sum_{G\in\mathcal{G}^n}\sum_{\mathbf{s}\in\{-1,+1\}^n} \underbrace{me^{\eta\Phi(\mathbf{s},G)}}_{-\frac{1}{\eta}\frac{\partial}{\partial\zeta}e^{\eta\Phi(\mathbf{s},G)}} = -\frac{1}{\eta}\frac{1}{\mathcal{Z}^\eta}\frac{\partial\mathcal{Z}^\eta}{\partial\zeta}. \tag{D.13}
$$

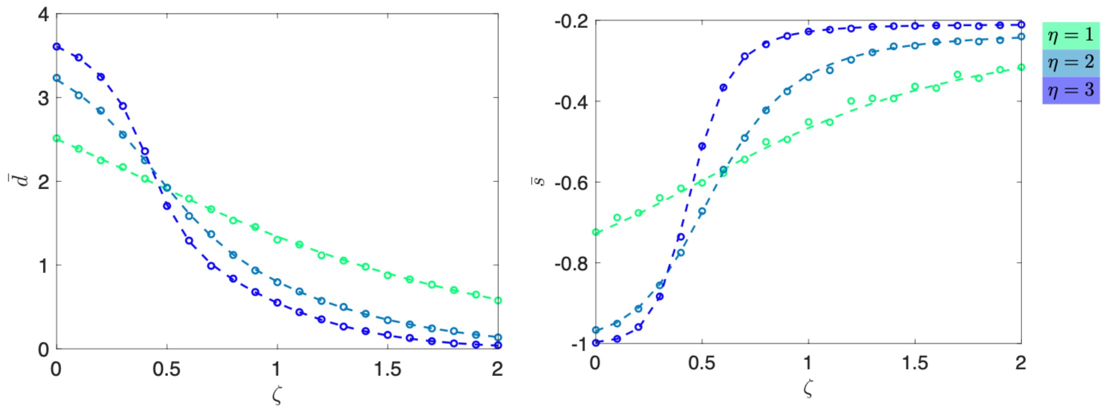

Figure D.1: The average degree $\bar{d} = 2m/n$ (left panel) and the average action level $\bar{s}$ (right panel) across different values of the linking cost $\zeta \in [0, 2]$ and varying noise $\eta \in \{1, 2, 3\}$. The parameters used are $n = 5$, $n_+ = 2$, $\kappa = 0.1$, $\rho = 0.1$, $\lambda = \chi = \xi = 1$ and $\theta = 0.75$. Dashed lines indicate the theoretical prediction of (D.1) in Proposition D.1 and (D.15) in Proposition D.2, respectively, while circles indicate averages across 1000 numerical Monte Carlo simulations of the model using the "next reaction method" for simulating a continuous time Markov chain (cf. Anderson, 2012; Gibson and Bruck, 2000).

With (D.6) and (D.13) we then can compute the expected number of links as

$$\mathbb{E}^\eta(m) = \frac{1}{\mathcal{Z}^\eta} \sum_{k=0}^{n} \sum_{j=0}^{\min\{k,n_+\}} \binom{n_+}{j} \binom{n-n_+}{k-j} e^{\eta(2k-n)} e^{\eta\left(\rho\left(2l(k,j)-\binom{n}{2}\right)\right)-\kappa(n-2(n_++k-2j))} \times \left(1+e^{\eta(\theta-\zeta)}\right)^{l(k,j)} \left(1+e^{-\eta(\theta+\zeta)}\right)^{\binom{n}{2}-l(k,j)} \left(\frac{l(k,j)}{1+e^{-\eta(\theta-\zeta)}} + \frac{\binom{n}{2}-l(k,j)}{1+e^{\eta(\theta+\zeta)}}\right), \tag{D.14}$$

and one can show that $\lim_{\zeta\to\infty} \mathbb{E}^\eta(m) = 0$. □

Note further that for $\theta = \rho = 0$ (D.14) simplifies to $\mathbb{E}^\eta(m) = \frac{1}{\mathcal{Z}^\eta} \binom{n}{2} \left(1+e^{-\eta\zeta}\right)^{\binom{n}{2}} \frac{1}{1+e^{\eta\zeta}} \sum_{k=0}^{n} \sum_{j=0}^{\min\{k,n_+\}} \binom{n_+}{j} \binom{n-n_+}{k-j} e^{\eta(2k-n)} = \frac{1}{\mathcal{Z}^\eta} \frac{e^{-\eta n}}{\pi(1+e^{\eta\zeta})} \binom{n}{2} \left(1+e^{-\eta\zeta}\right)^{\binom{n}{2}} \left[\pi \left(1+e^{2\eta}\right)^n - e^{2(n+1)\eta} \sin(n\pi)\, \Gamma(n+1)\, {}_2F_1\left(1,1;n+2;-e^{2\eta}\right)\right]$.

In the left panel of Figure D.1 we compare the average degree $\bar{d}$ obtained by averaging across simulations with the expected value $2\mathbb{E}^\eta(m)/n$ from Proposition D.1 for different values of the linking cost $\zeta \in [0, 2]$ and noise parameter $\eta \in \{1, 2, 3\}$. The theoretical result predicts well the simulated average degree, which naturally decreases with increasing linking costs ζ.

Now we turn to the average action level, which leads to the following counterpart of Proposition D.1.

Proposition D.2. *Assume homogeneous linking costs, $\zeta_1 = \zeta_2 = \zeta$. Then the expected average action level, $\bar{s} = \frac{1}{n}\sum_{i=1}^{n} s_i$, in the stationary state is given by*

$$\mathbb{E}^\eta(\bar{s}) = \frac{1}{\mathcal{Z}^\eta} \sum_{k=0}^{n} \sum_{j=0}^{\min\{k,n_+\}} \binom{n_+}{j} \binom{n-n_+}{k-j} \frac{n+4j-2(n_++k)}{n} e^{\eta(2k-n)} \times e^{\eta\left(\rho\left(2l(k,j)-\binom{n}{2}\right)\right)-\kappa(n-2(n_++k-2j))} \left(1+e^{\eta(\theta-\zeta)}\right)^{l(k,j)} \left(1+e^{-\eta(\theta+\zeta)}\right)^{\binom{n}{2}-l(k,j)}, \tag{D.15}$$

where $l(k, j)$ is defined in (D.2) and $n_+ = \#(\{\gamma_i = 1 : i = 1, \ldots, n\})$.

Proof of Proposition D.2. Assume that $\zeta_1 = \zeta_2 = \zeta$. Then the average action level $\bar{s} = \frac{1}{n}\sum_{i=1}^n s_i = \frac{1}{n}\langle \mathbf{u}, \mathbf{s}\rangle$ is given by

$$\begin{aligned}
\mathbb{E}^\eta(\bar{s}) &= \sum_{\mathbf{s}\in\{-1,+1\}^n} \bar{s}\mu_\eta^{GI}(\mathbf{s}) = \frac{1}{\mathcal{Z}^\eta}\sum_{\mathbf{s}\in\{-1,+1\}^n} \frac{1}{n}\langle \mathbf{u}, \mathbf{s}\rangle e^{\eta\mathcal{H}^\eta(\mathbf{s})} \\
&= \frac{1}{\mathcal{Z}^\eta}\sum_{k=0}^{n}\sum_{j=0}^{\min\{k,n_+\}} \binom{n_+}{j}\binom{n-n_+}{k-j}\frac{n+4j-2(n_+ + k)}{n} \\
&\quad \times e^{\eta(2k-n)} e^{\eta\left(\rho\left(2l(k,j)-\binom{n}{2}\right)\right)-\kappa(n-2(n_+ + k-2j))}\left(1+e^{\eta(\theta-\zeta)}\right)^{l(k,j)}\left(1+e^{-\eta(\theta+\zeta)}\right)^{\binom{n}{2}-l(k,j)}.
\end{aligned}$$

□

The average action level $\bar{s}$ in Proposition D.2 is illustrated in the right panel of Figure D.1 across different values of the linking cost ζ and for varying levels of noise, as parameterized by η. The average action is increasing with ζ and more sharply so as the level of noise is decreasing (respectively, η is increasing).

D.2. Local Information and Learning

The analysis of the LIL model is more complicated because its belief-formation $\boldsymbol{\psi}^{LIL}(\cdot)$ mapping given in (9) depends in an intricate manner on the current network structure. To make this characterization tractable, we rely on a *mean field approximation* that is commonly used in analyzing stochastic network formation models (see e.g., Jackson and Rogers, 2007). By making this approximation, in the stationary beliefs equation derived from (9):

$$\mathbf{p} = \varphi\left[\mathbf{I} - (1-\varphi)\widehat{\mathbf{D}}^{-1}\widehat{\mathbf{A}}\right]^{-1}\mathbf{D}^{-1}\mathbf{A}\mathbf{s}, \tag{D.16}$$

we replace the entries of the adjacency matrix, $\mathbf{A} = (a_{ij})_{i,j=1}^n$ with their expected values: $a_{ij} = e^{\eta(\theta s_i s_j - \zeta)}/(1+e^{\eta(\theta s_i s_j-\zeta)})$ for all $1 \leq i, j, \leq n$. Similarly, $\mathbf{D}$, $\widehat{\mathbf{A}}$ and $\widehat{\mathbf{D}}$ are computed. Under this approximation, the beliefs, $\mathbf{p}$, become a function of the actions, $\mathbf{s}$, only. This will allow us to compute the partition function ($\mathcal{Z}^\eta$) and other statistics of interest – such as the average degree or the average action level – for an arbitrary level of noise (η).

The following proposition characterizes the expected number of links for an arbitrary level of noise under a mean field approximation.

Proposition D.3. *Consider homogeneous linking costs, $\zeta_1 = \zeta_2 = \zeta$. Then, under a mean field approximation, the expected number of links is given by*

$$\mathbb{E}^\eta(m) \simeq \frac{1}{\mathcal{Z}^\eta}\sum_{\mathbf{s}\in\{-1,+1\}^n} e^{\eta\langle\tilde{\gamma},\mathbf{s}\rangle} h^\eta(\mathbf{s}), \tag{D.17}$$

where $\tilde{\gamma}_i = \gamma_i + \rho(n-1)p_i - \kappa$, beliefs $\mathbf{p}$ are given by Eq. (D.16), the adjacency matrix $\mathbf{A} = (a_{ij})_{i,j=1}^n$ has elements $a_{ij} = e^{\eta(\theta s_i s_j-\zeta)}/(1+e^{\eta(\theta s_i s_j-\zeta)})$, $\mathbf{D} = \mathrm{diag}(d_1,\ldots,d_n)$ is the diagonal matrix of the degrees $d_i = \sum_{j=1}^n a_{ij}$, the partition function is

$$\mathcal{Z}^\eta = \sum_{\mathbf{s}\in\{-1,+1\}^n} e^{\eta\langle\tilde{\gamma},\mathbf{s}\rangle} f^\eta(\mathbf{s}), \tag{D.18}$$

and

$$h^{\eta}(\mathbf{s}) = \frac{\left(e^{-\eta(\zeta+\theta)}+1\right)^{\alpha(\mathbf{s})}\left(e^{\eta(\theta-\zeta)}+1\right)^{\beta(\mathbf{s})-1}\left((\alpha(\mathbf{s})+\beta(\mathbf{s}))e^{\eta(\theta-\zeta)}+\alpha(\mathbf{s})+\beta(\mathbf{s})e^{2\eta\theta}\right)}{1+e^{\eta(\theta+\zeta)}},$$

$$f^{\eta}(\mathbf{s}) = \left(1+e^{-\eta(\zeta+\theta)}\right)^{\alpha(\mathbf{s})}\left(1+e^{\eta(\theta-\zeta)}\right)^{\beta(\mathbf{s})}, \tag{D.19}$$

with $\alpha(\mathbf{s}) = n_+(\mathbf{s})(n-n_+(\mathbf{s}))$, $\beta(\mathbf{s}) = \frac{1}{2}(n(n-1)-2n_+(\mathbf{s})(n-n_+(\mathbf{s})))$, $n_+(\mathbf{s}) = \#(\{s_i = 1 : i = 1,\dots,n\})$ *and* $\langle\cdot,\cdot\rangle$ *is the usual scalar product in* $\mathbb{R}^N$.

Before proceeding with the proof of Proposition D.3 we state the following lemma which will be useful later.

Lemma D.4. *For any* $\mathbf{s} \in \mathbf{S} = \{-1,+1\}^n$ *we have that*

$$\prod_{i=1}^{n}\prod_{j=i+1}^{n}\left(1+e^{\eta(\theta s_i s_j-\zeta)}\right) = \left(1+e^{-\eta(\theta+\zeta)}\right)^{n_+(n-n_+)}\left(1+e^{\eta(\theta-\zeta)}\right)^{\frac{n(n-1)-2n_+(n-n_+)}{2}},$$

where $n_+ = \#(\{\gamma_i = 1 : i = 1,\dots,n\})$.

Proof of Lemma D.4. In the following we denote by $f(\mathbf{s}) \equiv \prod_{i=1}^{n}\prod_{j=i+1}^{n}\left(1+e^{\eta(\theta s_i s_j-\zeta)}\right)$ and $g(s_i,s_j) \equiv 1+e^{\eta(\theta-\zeta)}$. Then we can write

$$f(\mathbf{s}) = \prod_{i=1}^{n_+-1}\left(\prod_{j=i+1}^{n_+} g(s_i,s_j)\prod_{j=n_++1}^{n} g(s_i,s_j)\right)\prod_{j=n_++1}^{n} g(s_{n_+},s_j)\prod_{i=n_++1}^{n}\prod_{j=i+1}^{n} g(s_i,s_j)$$

$$= g(+1,-1)^{n_+(n-n_+)}g(+1,+1)^{\frac{n(n-1)-2n_+(n-n_+)}{2}} = \left(1+e^{-\eta(\theta+\zeta)}\right)^{n_+(n-n_+)}\left(1+e^{\eta(\theta-\zeta)}\right)^{\frac{n(n-1)-2n_+(n-n_+)}{2}}$$

This concludes the proof. □

Proof of Proposition D.3. Assume that $\zeta_1 = \zeta_2 = \zeta$. Then, in the belief-based model, the quasi partition function is given by

$$\mathcal{Z}^{\eta} \equiv \sum_{G\in\mathcal{G}^n}\sum_{\mathbf{s}\in\{-1,+1\}^n} e^{\eta\tilde{\Phi}(\mathbf{s},\mathbf{p},G)} = \sum_{\mathbf{s}\in\{-1,+1\}^n} e^{\eta\sum_{i=1}^{n}\tilde{\gamma}_i s_i}\prod_{i=1}^{n}\prod_{j=i+1}^{n}\left(1+e^{\eta(\theta s_i s_j-\zeta)}\right),$$

where we have denoted by $\tilde{\gamma}_i = \gamma_i + \rho(n-1)p_i - \kappa$. The expected number of links is given by

$$\mathbb{E}^{\eta}(m) = -\frac{1}{\eta}\frac{1}{\mathcal{Z}^{\eta}}\frac{\partial\mathcal{Z}^{\eta}}{\partial\zeta} = -\frac{1}{\eta}\frac{1}{\mathcal{Z}^{\eta}}\sum_{\mathbf{s}\in\{-1,+1\}^n} e^{\eta\sum_{i=1}^{n}\tilde{\gamma}_i s_i}\frac{\partial}{\partial\zeta}\prod_{i=1}^{n}\prod_{j=i+1}^{n}\left(1+e^{\eta(\theta s_i s_j-\zeta)}\right).$$

Denoting by $f^{\eta}(\mathbf{s}) \equiv \prod_{i=1}^{n}\prod_{j=i+1}^{n}\left(1+e^{\eta(\theta s_i s_j-\zeta)}\right)$ from Lemma D.4 it follows that $f^{\eta}(\mathbf{s}) = (1+e^{-\eta(\theta+\zeta)})^{\alpha(\mathbf{s})}(1+e^{\eta(\theta-\zeta)})^{\beta(\mathbf{s})}$, where $\alpha(\mathbf{s}) = n_+(\mathbf{s})(n-n_+(\mathbf{s}))$, $\beta(\mathbf{s}) = \frac{1}{2}(n(n-1)-2n_+(\mathbf{s})(n-n_+(\mathbf{s})))$ and $n_+(\mathbf{s}) = \#(\{s_i = 1 : i = 1,\dots,n\})$. Moreover one can show that

$$h^{\eta}(\mathbf{s}) \equiv \frac{\partial f^{\eta}(\mathbf{s})}{\partial\zeta} = \frac{\left(e^{-\eta(\zeta+\theta)}+1\right)^{\alpha(\mathbf{s})}\left(e^{\eta(\theta-\zeta)}+1\right)^{\beta(\mathbf{s})-1}\left((\alpha(\mathbf{s})+\beta(\mathbf{s}))e^{\eta(\theta-\zeta)}+\alpha(\mathbf{s})+\beta(\mathbf{s})e^{2\eta\theta}\right)}{1+e^{\eta(\zeta+\theta)}},$$

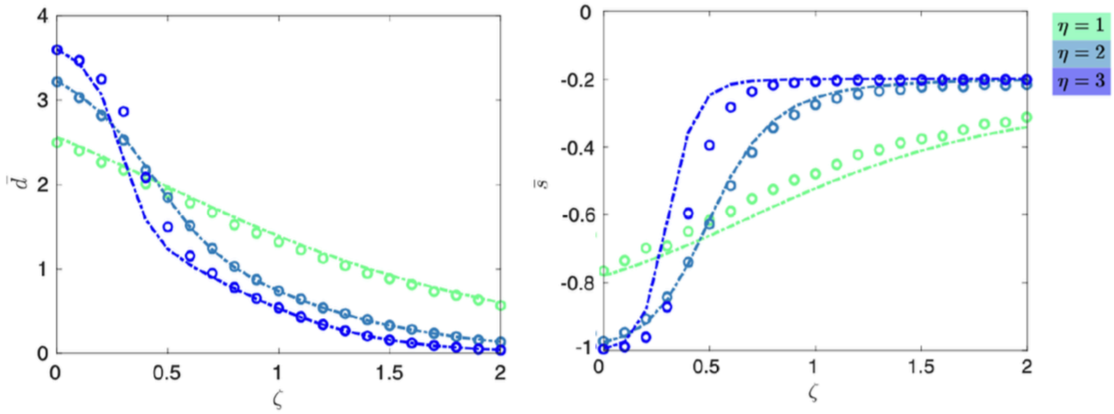

Figure D.2: The average degree $\bar{d} = 2m/n$ (left panel) and the average action level $\bar{s}$ (right panel) across different values of the linking cost $\zeta \in [0, 2]$, $\eta \in \{1, 2, 3\}$, $n = 5$, $n_+ = 2$, $\kappa = 0.1$, $\rho = 0.1$, $\lambda = \chi = \xi = 1$, $\varphi = 0.5$ and $\theta = 0.75$. Dashed-dotted lines indicate the theoretical predictions of $\bar{d} = 2m/n$ in (D.17) in Proposition D.3 and of $\bar{s}$ in (D.21) of Proposition D.4, respectively, while circles indicate averages across 1000 numerical Monte Carlo simulations of the model using the "next reaction method" for simulating a continuous time Markov chain (cf. Anderson, 2012; Gibson and Bruck, 2000).

and we can write $\mathbb{E}^\eta(m) = \frac{1}{\mathcal{Z}^\eta} \sum_{\mathbf{s}\in\{-1,+1\}^n} e^{\eta\langle\tilde{\gamma},\mathbf{s}\rangle} h^\eta(\mathbf{s})$, where $\mathcal{Z}^\eta = \sum_{\mathbf{s}\in\{-1,+1\}^n} e^{\eta\langle\tilde{\gamma},\mathbf{s}\rangle} f^\eta(\mathbf{s})$. Finally, stationary beliefs are given by (A.4) so that we can write them as a function of the actions and network as $\mathbf{p} = \varphi \left[\mathbf{I} - (1-\varphi)\widehat{\mathbf{D}}^{-1}\widehat{\mathbf{A}}\right]^{-1} \mathbf{D}^{-1}\mathbf{A}\mathbf{s}$. Moreover, from Lemma D.2 we know that the linking probability conditional on actions $\mathbf{s}$ is given by

$$p_{ij} = \frac{e^{\eta(\theta s_i s_j - \zeta)}}{1 + e^{\eta(\theta s_i s_j - \zeta)}}, \tag{D.20}$$

and the expected value of a_{ij} of the (i,j)-th element of $\mathbf{A}$ is given by p_{ij}. This concludes the proof. $\square$

The left panel in Figure D.2 shows the average degree $\bar{d} = 2\mathbb{E}^\eta(m)/n$ across different values of the linking cost $\zeta \in [0, 2]$ and $\eta \in \{1, 2, 3\}$. The average degree is decreasing with the linking cost ζ. The decrease is becoming sharper as the level of noise is decreasing (respectively, η is increasing).

The next proposition characterizes the average action level for an arbitrary level of noise under a mean field approximation.

Proposition D.4. *Consider homogeneous linking costs, $\zeta_1 = \zeta_2 = \zeta$. Then, under a mean field approximation, the expected average action level, $\bar{s}$, is given by*

$$\mathbb{E}^\eta(\bar{s}) \simeq \frac{1}{\mathcal{Z}^\eta}\frac{1}{n} \sum_{\mathbf{s}\in\{-1,+1\}^n} \langle \mathbf{u}, \mathbf{s}\rangle e^{\eta\langle\tilde{\gamma},\mathbf{s}\rangle} f^\eta(\mathbf{s}), \tag{D.21}$$

with $f^\eta(\cdot)$ given by (D.19), the partition function is given by (D.18), $\tilde{\gamma}_i = \gamma_i + \rho(n-1)p_i - \kappa$, beliefs $\mathbf{p}$ are given by Eq. (D.16), the adjacency matrix $\mathbf{A} = (a_{ij})_{i,j=1}^n$ has elements $a_{ij} = e^{\eta(\theta s_i s_j - \zeta)}/(1 + e^{\eta(\theta s_i s_j - \zeta)})$, $\mathbf{D} = \mathrm{diag}(\delta_1, \ldots, \delta_n)$ is the diagonal matrix of the degrees $\delta_i = \sum_{j=1}^n a_{ij}$, and $\langle \cdot, \cdot \rangle$ is the usual scalar product in $\mathbb{R}^N$.

Proof of Proposition D.4. Assume that $\zeta_1 = \zeta_2 = \zeta$. Then the average action level $\bar{s} = \frac{1}{n}\sum_{i=1}^n s_i = \frac{1}{n}\mathbf{u}^\top\mathbf{s} = \frac{1}{n}\langle \mathbf{u}, \mathbf{s}\rangle$ is given by $\mathbb{E}^\eta(\bar{s}) = \sum_{\mathbf{s}\in\{-1,+1\}^n} \bar{s}\mu_\eta^{GI}(\mathbf{s}) = \frac{1}{\mathcal{Z}^\eta}\sum_{\mathbf{s}\in\{-1,+1\}^n} \frac{1}{n}\langle \mathbf{u}, \mathbf{s}\rangle\ e^{\eta\sum_{i=1}^n \tilde{\gamma}_i s_i} \prod_{i=1}^n \prod_{j=i+1}^n \left(1 + e^{\eta(\theta s_i s_j - \zeta)}\right)$ where we have denoted by $\tilde{\gamma}_i = \gamma_i + \rho(n-1)p_i - \kappa$. Denoting by $f^\eta(\mathbf{s}) \equiv \prod_{i=1}^n \prod_{j=i+1}^n \left(1 + e^{\eta(\theta s_i s_j - \zeta)}\right)$ from Lemma D.4 it follows that $f^\eta(\mathbf{s}) = \left(1 + e^{-\eta(\theta+\zeta)}\right)^{\alpha(\mathbf{s})}$

$\left(1+e^{\eta(\theta-\zeta)}\right)^{\beta(\mathbf{s})}$, where $\alpha(\mathbf{s}) = n_+(\mathbf{s})(n-n_+(\mathbf{s}))$, $\beta(\mathbf{s}) = \frac{1}{2}(n(n-1)-2n_+(\mathbf{s})(n-n_+(\mathbf{s})))$ and $n_+(\mathbf{s}) = \#(\{s_i = 1 : i = 1,\ldots,n\})$, and hence $\mathbb{E}^\eta(\bar{s}) = \frac{1}{\mathcal{Z}^\eta}\frac{1}{n}\sum_{\mathbf{s}\in\{-1,+1\}^n} \langle \mathbf{u},\mathbf{s}\rangle e^{\eta\sum_{i=1}^n \tilde{\gamma}_i s_i}$ $(1+ e^{-\eta(\theta+\zeta)})^{\alpha(\mathbf{s})}(1+ e^{\eta(\theta-\zeta)})^{\beta(\mathbf{s})}$. Finally, the stationary beliefs $\mathbf{p}(\mathbf{s})$ as a function of the actions $\mathbf{s}$ can be computed as in the proof of Proposition D.3. □

The right panel in Figure D.2 shows the average action level $\bar{s}$ across different values of the linking cost $\zeta \in [0,2]$ and noise $\eta \in \{1,2,3\}$. The average action level is increasing with ζ. The increase is becoming sharper as the level of noise is decreasing (respectively, η is increasing). Figure D.2 also illustrates a good match between the theory and simulations for different values of ζ and η.

E. Equilibrium Characterization for Finite Populations & Small Noise

E.1. Complete Information

The following proposition provides a complete characterization of the Stochastically Stable States (SSS) for finite populations in the complete information environment (generalizing Proposition 3).

Proposition E.1. *Let $n_+ = \#(\{\gamma_i = 1 : i = 1,\ldots,n\})$, $\nu_+ = n_+/n$ and denote by*

$$\begin{aligned}
\theta^* &= \zeta_2 + \frac{2(1-\kappa)}{n-n_+} - 2\rho,\\
\theta^{**} &= \zeta_2 + \frac{2(1+\kappa)}{n_+} - 2\rho,\\
\kappa^* &= 2\nu_+ - 1,\\
\rho^* &= \frac{1-\kappa}{n-n_+},\\
\rho^{**} &= \frac{1+\kappa}{n_+}.
\end{aligned}$$

Moreover, w.l.o.g. assume that $\kappa \geq 0$.

(i) If $\theta < \zeta_1$, the stochastically stable state in the limit of $\eta \to \infty$ is given by the ***empty network****, $\overline{K}_n$ with the following action profiles:*

1. *If $\nu_+ < 1/2$ and*
   - *(a) $\rho > \rho^*$, then all agents choose the action $s_i = -1$.*
   - *(b) $\rho < \rho^*$, then all agents choose the action $s_i = \gamma_i$.*
2. *If $\nu_+ > 1/2$,*
   - *(a) $\kappa > \kappa^*$ and*
     - *i. $\rho > \rho^*$, then all agents choose the action $s_i = -1$.*
     - *ii. $\rho < \rho^*$, then all agents choose the action $s_i = \gamma_i$.*
   - *(b) $\kappa < \kappa^*$ and*
     - *i. $\rho > \rho^{**}$, then all agents choose the action $s_i = +1$.*

ii. $\rho < \rho^{**}$, *then all agents choose the action* $s_i = \gamma_i$.

(ii) For the case of $\theta > \zeta_1$, *the stochastically stable state is either a* ***complete network***, K_n, *or composed of the* ***union of two cliques***, $K_{n_+} \cup K_{n-n_+}$, *where all agents in the same clique have the same preference* γ_i *and choose the same action. More precisely:*

1. *If* $\nu_+ < 1/2$,
   (a) $\theta > \theta^*$ *and*
      i. $\theta > \zeta_2$, *then the stochastically stable state is the* ***complete network*** K_n *with* $s_i = -1$.
      ii. $\theta < \zeta_2$, *then the stochastically stable state is the* ***union of two cliques*** $K_{n_+} \cup K_{n-n_+}$ *with* $s_i = -1$.
   (b) $\theta < \theta^*$ *and*
      i. $\rho < \rho^*$ *then the stochastically stable state is the* ***union of two cliques*** $K_{n_+} \cup K_{n-n_+}$ *with* $s_i = \gamma_i$.
      ii. $\rho > \rho^*$ *then the stochastically stable state is the* ***union of two cliques*** $K_{n_+} \cup K_{n-n_+}$ *with* $s_i = -1$.
2. *If* $\nu_+ > 1/2$,
   (a) $\kappa > \kappa^*$,
      i. $\theta > \theta^*$ *and*
         - $\theta > \zeta_2$, *then the stochastically stable state is the* ***complete network*** K_n *with* $s_i = -1$.
         - $\theta < \zeta_2$, *then the stochastically stable state is the* ***union of two cliques*** $K_{n_+} \cup K_{n-n_+}$ *with* $s_i = -1$.
      ii. $\theta < \theta^*$ *and*
         - $\rho < \rho^*$ *then the stochastically stable state is the* ***union of two cliques*** $K_{n_+} \cup K_{n-n_+}$ *with* $s_i = \gamma_i$.
         - $\rho > \rho^*$ *then the stochastically stable state is the* ***union of two cliques*** $K_{n_+} \cup K_{n-n_+}$ *with* $s_i = -1$.
   (b) *If* $\kappa < \kappa^*$,
      i. $\theta > \theta^{**}$ *and*
         - $\theta > \zeta_2$, *then the stochastically stable state is the* ***complete network*** K_n *with* $s_i = +1$.
         - $\theta < \zeta_2$, *then the stochastically stable state is the* ***union of two cliques*** $K_{n_+} \cup K_{n-n_+}$ *with* $s_i = +1$.
      ii. $\theta < \theta^{**}$ *and*
         - $\rho < \rho^{**}$ *then the stochastically stable state is the* ***union of two cliques*** $K_{n_+} \cup K_{n-n_+}$ *with* $s_i = \gamma_i$
         - $\rho > \rho^{**}$ *then the stochastically stable state is the* ***union of two cliques*** $K_{n_+} \cup K_{n-n_+}$ *with* $s_i = +1$.

Proof of Proposition E.1. The statement follows from the proof of Proposition 3 for finite n. □

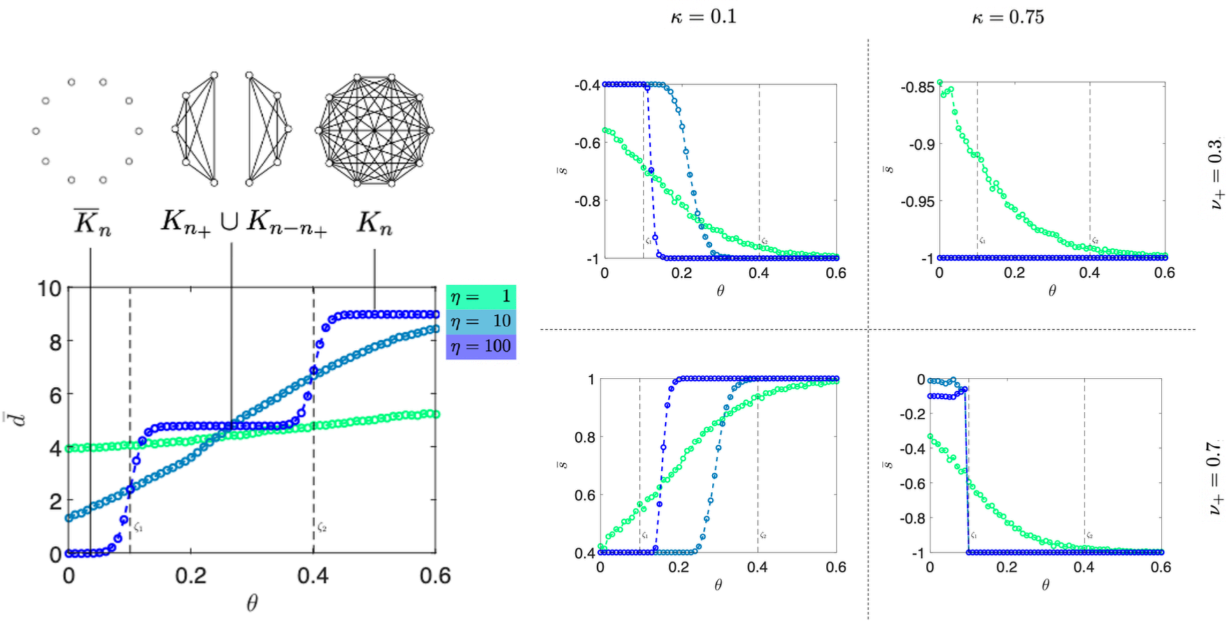

Figure E.1: Simulation results for the ***GI scenario*** on the average degree of the network $\bar{d}$ (left panel) and the average action $\bar{s}$ (right panel) for varying values of θ and η, using the "next reaction method" for simulating a continuous time Markov chain (cf. Anderson, 2012; Gibson and Bruck, 2000). The circles represent averages obtained across 1000 Monte Carlo runs under the following parameters: $n = 10$, $\eta \in \{1, 10, 100\}$, $\lambda = \chi = \xi = 1$, and $\rho = 0.1$. The thresholds ζ_1 and ζ_2 are indicated with a vertical dashed line. As η becomes large, for $\theta < \zeta_1$ the network is empty ($\overline{K}_n$), for $\zeta_1 < \theta < \zeta_2$ the network is partitioned into two type-homogeneous cliques (completely connected subnetworks, denoted by K_{n^+} and K_{n-n^+}), and for $\zeta_2 < \theta$ it is complete (K_n).

Proposition E.1 shows that when the idiosyncratic preference is large enough (i.e., θ is small enough) in the payoff function of (1) then the society is segregated in disconnected communities in which each agent is choosing the action in accordance with her idiosyncratic preference ($\gamma_i = s_i$ for all $i = 1, \ldots, n$), while if the peer effect is strong enough (the conformity parameter θ is large enough) then the society becomes completely connected and all agents choose the same action (homogeneous society). The action chosen in the latter case is determined by the idiosyncratic preference of the majority. That is, if more agents have an idiosyncratic preference $\gamma_i = +1$ (and $\nu_+ < 1/2$) then all agents will choose $s_i = +1$, and vice versa. Finally, if linking is too costly ($\zeta_2 > \theta$), then all agents are isolated and choose their idiosyncratic preference if the global conformity parameter ρ is not too high ($\rho < \rho^*$).

E.2. Local Information and Learning

The following proposition provides a characterization of the Stochastically Quasi-stable States (QSS) for finite populations in the belief formation environment (generalizing Proposition 5).

Proposition E.2. *Let $n_+ = \#(\{\gamma_i = 1 : i = 1, \ldots, n\})$, $\nu_+ = n_+/n$ and denote by*

$$\tilde{\theta}^* = \zeta_2 + \frac{2(1-\kappa)}{n - n_+},$$

$$\tilde{\theta}^{**} = \zeta_2 + \frac{2(1+\kappa)}{n_+},$$

$$\kappa^* = 2\nu_+ - 1.$$

Moreover, w.l.o.g. assume that $\kappa \geq 0$. Then, in the stochastically stable state in the limit of $\eta \to \infty$, we have that beliefs are consistent with actions, $p_i = s_i$, for all $i = 1, \ldots, n$, where:

(i) If $\theta < \zeta_1$ then the stochastically stable state is given by the ***empty network****, $\overline{K}_n$. Further,*

1. *if $\nu_+ < 1/2$ and*
   - *(a) $\kappa > 1$, then all agents choose the action $s_i = -1$.*
   - *(b) $\kappa < 1$, then all agents choose the action $s_i = \gamma_i$.*
2. *if $\nu_+ > 1/2$ and*
   - *(a) $\kappa > \kappa^*$:*
     - *i. $\kappa > 1$, then all agents choose the action $s_i = -1$.*
     - *ii. $\kappa < 1$, then all agents choose the action $s_i = \gamma_i$.*
   - *(b) $\kappa < \kappa^*$, then all agents choose the action $s_i = +1$.*

(ii) If $\theta > \zeta_1$ then the stochastically stable is either the ***complete network****, K_n, or composed of a* ***union of two cliques****, $K_{n_+} \cup K_{n-n_+}$, where all agents in the same clique have the same preference γ_i and choose the same action. More precisely,*

1. *if $\nu_+ < 1/2$,*
   - *(a) $\theta > \tilde{\theta}^*$ and*
     - *i. $\theta > \zeta_2$, then the stochastically stable state is the* ***complete network*** *K_n with $s_i = -1$.*
     - *ii. $\theta < \zeta_2$, then the stochastically stable state is the* ***union of two cliques*** *$K_{n_+} \cup K_{n-n_+}$ with $s_i = -1$.*
   - *(b) $\theta < \tilde{\theta}^*$ and*
     - *i. $\kappa < 1$, then the stochastically stable state is the* ***union of two cliques*** *$K_{n_+} \cup K_{n-n_+}$ with $s_i = \gamma_i$.*
     - *ii. $\kappa > 1$, then the stochastically stable state is the* ***union of two cliques*** *$K_{n_+} \cup K_{n-n_+}$ with $s_i = -1$.*
2. *if $\nu_+ > 1/2$,*
   - *(a) $\kappa > \kappa^*$,*
     - *i. $\theta > \tilde{\theta}^*$ and*
       - *$\theta > \zeta_2$, then the stochastically stable state is the* ***complete network*** *K_n with $s_i = -1$.*
       - *$\theta < \zeta_2$, then the stochastically stable state is the* ***union of two cliques*** *$K_{n_+} \cup K_{n-n_+}$ with $s_i = -1$.*
     - *ii. $\theta < \tilde{\theta}^*$ and*
       - *$\kappa < 1$, then the stochastically stable state is the* ***union of two cliques*** *$K_{n_+} \cup K_{n-n_+}$ with $s_i = \gamma_i$.*
       - *$\kappa > 1$, then the stochastically stable state is the* ***union of two cliques*** *$K_{n_+} \cup K_{n-n_+}$ with $s_i = -1$.*
   - *(b) $\kappa < \kappa^*$,*
     - *i. $\theta > \tilde{\theta}^{**}$ and*
       - *$\theta > \zeta_2$, then the stochastically stable state is the* ***complete network*** *K_n with $s_i = +1$.*

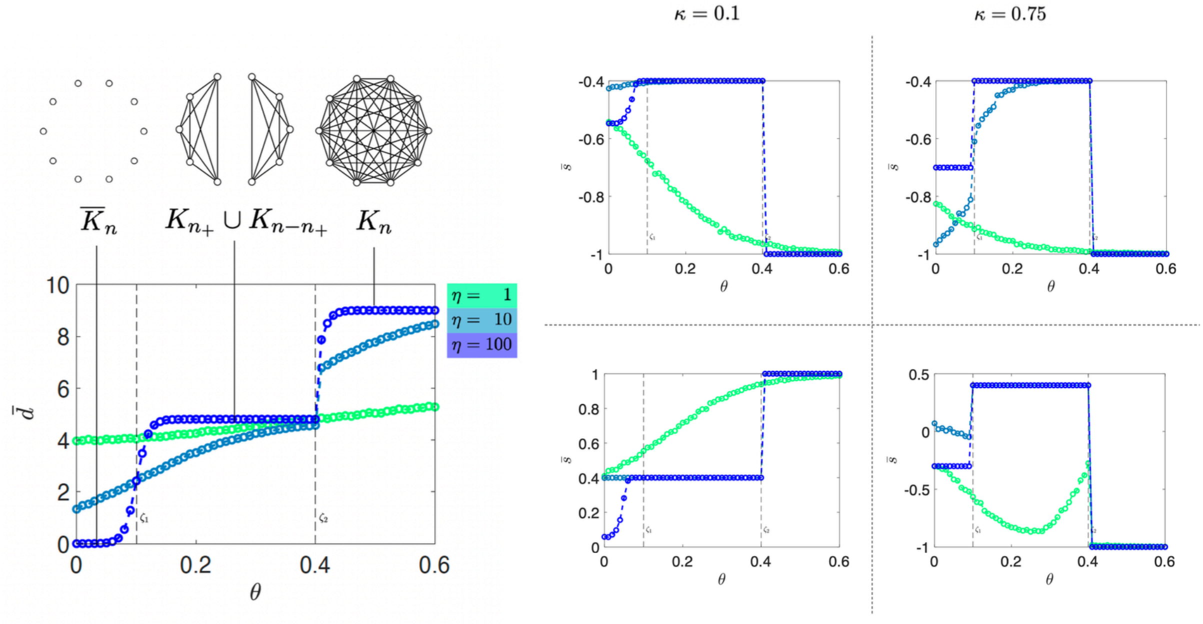

Figure E.2: Simulation results for the ***LIL scenario*** on the average degree of the network $\bar{d}$ (left panel) and the average action $\bar{s}$ (right panel) for varying values of θ and η, using the "next reaction method" for simulating a continuous time Markov chain (cf. Anderson, 2012; Gibson and Bruck, 2000). The circles represent averages obtained across 1000 Monte Carlo runs under the following parameters: $n = 10$, $\eta \in \{1, 10, 100\}$, $\lambda = \chi = \xi = 1$, $\varphi = 0.5$, and $\rho = 0.1$. The stochastically stable states in the limit of large η corresponds to Proposition 5. The thresholds ζ_1 and ζ_2 are indicated with a vertical dashed line. As η becomes large, for $\theta < \zeta_1$ the network is empty ($\overline{K}_n$), for $\zeta_1 < \theta < \zeta_2$ the network is partitioned into two type-homogeneous cliques (K_{n^+} and K_{n-n^+}), and for $\zeta_2 < \theta$ it is complete (K_n).

- *$\theta < \zeta_2$, then the stochastically stable state is the* ***union of two cliques*** *$K_{n_+} \cup K_{n-n_+}$ with $s_i = +1$.*

ii. $\theta < \tilde{\theta}^{**}$, then the stochastically stable state is the ***union of two cliques*** *$K_{n_+} \cup K_{n-n_+}$ with $s_i = +1$.*

Proof of Proposition E.2. The statement follows from the proof of Proposition 5 for finite n. □

From Proposition E.2 we observe that the possible stochastically stable actions and networks are the same as in Proposition E.1, and the beliefs are identical to the actions. This implies that when the stochastically stable network is complete, then the beliefs (about the average action chosen in the entire population) are correct. But when the stochastically stable network is a union of two cliques, $K_{n^+} \cup K_{n-n_+}$, then the beliefs do not correspond to the average action chosen in the entire population, but represent only the average action chosen in the local clique.

F. Context and Historical View of the Egyptian Arab Spring

In the following we provide a brief historical overview of Egyptian politics and the civil unrest in Egypt that began as part of the Arab Spring.

F.1. Historical and Political Background

Egypt had been under what was effectively one-party rule since the 1952 coup that remove King Farouk from power. The ruling political party, originally named Liberation Rally, transitioned to the politically centrist National Democratic Party (NDP). The NDP ideology centered around modernist and anti-Islamist, and members were secular elite, bureaucrats, and regime cronies. Hosni Mubarak rose to the head of the NDP movement in 1981. In addition to the NDP, the Egyptian military sustains power and influence in the political arena. The military also has a vast presence in civilian industry, making it wealthy and opposes Islamist rule.

Egypt's main opposition to the NDP's rule was in the Islamist movement, whose main social organization is the Muslim Brotherhood (MB). The MB's ideology centers around a literal interpretation of scriptures and advocates a return to idealized Islamic society. MB's followers are urbanized middle and lower classes. The MB was outlawed in 1954 connected to the assassination attempt of then-president Nasser with many influential leaders of the movement jailed. From the 1970s, leaders were freed, and MB moved towards an official political party as many leaders were released from prisons. By the 2005 elections, MB gained approximately 20 percent of the seats in the Egyptian parliament by running candidates as independents, making them a force within Egyptian society and politics, despite the state denying that their existance. Furthermore, Mubarak's regime took a vigorous position against Islamic investment companies. This served as a severe attack on the MB and its largest source of finance; more than 40% of the owners of the Islamic investment companies were MB members and supporters.

F.2. The Egyptian Arab Spring

We can divide Egypt's Arab Spring into four stages: (I) the lead up to and then fall of Mubarak, (II) a period of rule by the Egyptian Military, (III) rule by Islamist President Mohamed Morsi, and (IV) the fall of President Morsi and the return to power by the military, the latter of which is the focal point of our empirical exercise.

Phase I: Fall of Mubarak. Under Mubarak's rule, and particularly in the latter stages, NDP members acquired vast wealth while the civilian population stagnated Following the removal of Tunisian President Bin Ali in early 2011, the fervor against privileged elites and ruling parties in North Africa and the Middle East grew. This led to thousands of protesters congregating in Cairo's Tahir Square in a public demonstration against the Mubarak regime organized by young middle-class Egyptians, rather than an Islamist opposition. The MB later encouraged its members to participate without invoking the MB's Islamist slogans or ideology. After the initial protest, demonstrations continued, growing in size to to 50K on January 28, and by February 1 to over 500K protesters in Tahir Square alone. In reponse, on the evening of February 11, Mubarak resigns and hands over power to the military. After this handover, protests continued until relative stability in mid-March. The first phase of Egypt's Arab Spring ended on April 16, 2011, when an administrative court dissolved the NDP on charges of corruption and seized its assets.

Phase II: First phase of military rule. Directly after the uprising, the Supreme Council of the Armed Forces (SCAF) of the military faced a dilemma. The SCAF had to decide either to proceed to elections to end their post-revolutionary rule, or slow down the electoral timetable and prioritize the writing of a new constitution. Ultimately, the SCAF decided to hold parliamentary elections before drafting a new constitution. The decision to hold election first was in part a response to continuing protests by Egyptian citizens. The MB also rallied behind plan to hold parliamentary elections prior to drafting a new constitution. In the ensuing elections, the MB fields brands it's political organization and the Freedom and Justice Party (FJP), and it's presidential candidate is Mohammed Morsi. Results of the first round of the election announced on May 28 and the results of the runoff election announced on June 24.

Phase III: Rule of Mohammed Morsi. Mohammed Morsi narrowly won the parliamentary elections against former General Ahmed Shafiq with 51.7% of the vote. However, the constitution imposed by the SCAF during their time in power, left Morsi with limited power. On August 12, 2012, Morsi revokes the interim rules imposed upon him, transferring power back to the president, giving himself absolute legislative authority. Over the early stage of Morsi's rule he continually struggled to assert power over the military. In response, and to further provide himself power, he removes 5 key military figures over the month of August 2012 and tensions between the two continue to rise.

Public opposition to Morsi began building in November 2012 when, wishing to ensure that the Islamist-dominated constituent assembly could finish drafting a new constitution, the president issued a decree granting himself further power. At the same time, critics claimed he had mishandled the economy and failed to deal with the very issues that led to the uprising that brought him to power. Continuing calls for rights and social justice led to an accelerated decrease in the popularity of Morsi over the next few months. As his popularity continued to dwindle, a referendum passed a new constitution on December 23, promoting political Islam Citizens responded to the new constitution with alternating protests in Tahir Square, rotating between pro- and anti-Islamist movements.

Phase IV: Fall of Morsi and return to military rule. On June 26, 2013, Morsi delivers a divisive address to defuse growing defiance to his rule. This leads to larger protests in the following days involving two sets of protesters, pro- versus anti-Morsigroups, in different locations across Egypt. On July 1, in response to the growing tension and civilian unrest the military issues an ultimatum to Morsi to call an early election. On July 2, Morsi refuses to step down. On July 3, the Egyptian Military overthrows Morsi's regime in coup, and anti-military intervention protests grow. By July 24, the military encourages pro-military intervention protests to counteract the opposition protest to their takeover. From July 27 to mid-August, large demonstrations take place from both sides.

G. Constructing the Data on the Egyptian Protests

G.1. Reconstructing the Dataset of Borge-Holthoefer et al. (2015)

The data used in our empirical analysis is a reconstruction of the data originally used in Borge-Holthoefer et al. (2015) who document opinion dynamics about the Egyptian protests over the Summer of 2013. They collect a sample of all Arabic language tweets from the Twitter API over the June 21, 2013 to September 30, 2013. Then tweets relevant to Egypt and the protest movement were identified by constructing over 100 queries that included keywords and hashtags covering aspects of Egyptian politics, government and the protest movement.

Twitter's terms of use prevent researchers sharing Twitter data either directly or by posting the raw data online. However, authors can provide the 'Tweet-IDs' - a numerical identifier - and numeric 'User-IDs' - which uniquely identify Twitter user profiles - that were used in their research. After obtaining the IDs, we then performed a process known as 'Tweet Hydration' to pull the tweets and user-info with any associated meta-data using software created by the Documenting the Now project (`https://www.docnow.io/`).[31] This process involves a process of repeatedly querying

[31] The original Tweet IDs and User-IDs along with the codes that perform the tweet hydration process are available in our replication package.

the public facing Twitter API and requesting the associated meta-data for either a Tweet-ID or User-ID.

The API returns the complete data about each tweet or user from the query provided that the tweet or user profile has not been deleted or taken down from the platform. Importantly for our application, the returned data includes the textual content of the tweet, date and time posted and the user-ID. When returning information about each user, the API returns the information at the time the API query was made - thus the username, friends and follower counts as measured in December 2020.[32]

G.2. Inferring Political Affiliation of Twitter Users

We briefly describe the process undertaken by Weber et al. (2013) to label 20,886 Egyptian Twitter users by their political affiliation, which we use as training data to classify each of Twitter users in our dataset. Starting with a set of manually labelled Twitter users, the authors collect data on other users who interact with these accounts. Table G.2 contains the list of the seed users. For each of the seed users, their most recent 3,200 tweets extracted from the Twitter API and then the user meta-data of up to 200 retweeters of each tweet were downloaded.[33] From this set of retweeters, only those who's location could be identified as being located in Egypt were retained in the sample.[34] The remaning Egyptian residing users were classified as Islamist or Secularist according to their retweeting behaviour: A user who retweeted n_I distinct Islamist seed users and n_S distinct Secular seed users over two time periods (January and March 2013) is classified as a political Islamist if $n_I/(n_I + n_S) > 0.5$.

Table G.1: Seed users and their political affiliation.

Secularists		Islamists	
Screen Name	Twitter Handle	Screen Name	Twitter Handle
Mohamed El Baradei	@ElBaradei	Muhammad Morsi	@MuhammadMorsi
Alaa Al-Aswany	@alaaaswany	Fadel Soliman	@FadelSoliman
Ayman Nour	@AymanNour	Essam Al Erian	@EssamAlErian
Wael Abbas	@waelabbas	Almogheer	@almogheer
Belal Fadl	@belalfadl	Hazem Salah	@HazemSalahTW
Dr. Hazem Abdelazim	@Hazem__Azim	Khaleed Abdallah	@KhaleedAbdallah
MohamedAbuHamed	@MohamedAbuHamed	Melhamy	@melhamy
HamzawyAmr	@HamzawyAmr	Dr Mohamed Aly	@dr__mohamed__aly
E3adet Nazar	@E3adet__Nazar	Mustafa Hosny	@MustafaHosny
GameelaIsmail	@GameelaIsmail	El Awa	@El__Awa
shabab6april	@shabab6april		
waelabbas	@waelabbas		

Figure G.2 shows the users' political affiliations and links between them based on a "snowball" sample of a randomly drawn user and her neighbors, neighbors' neighbors, etc. of the original network of users (Goodman, 1961). A clear separation between densely connected clusters of users with the same political affiliation can be seen with only a few links across these clusters.[35] This indicates that users are mainly connected with other users with the same political views.

[32]To the best of our knowledge there is no way to get the meta-information for each user retroactively (i.e. their 2013 values).

[33]These are the quantity limits are imposed by the Twitter API in early 2013.

[34]User locations where determined by their self reported location in their Twitter meta-data or by references to place names in these users own tweets - which Weber et. al. collected and passed through Yahoo Placemaker.

[35]The hierarchical structure illustrated in the network is mainly due to the nature of the "snowball" sampling

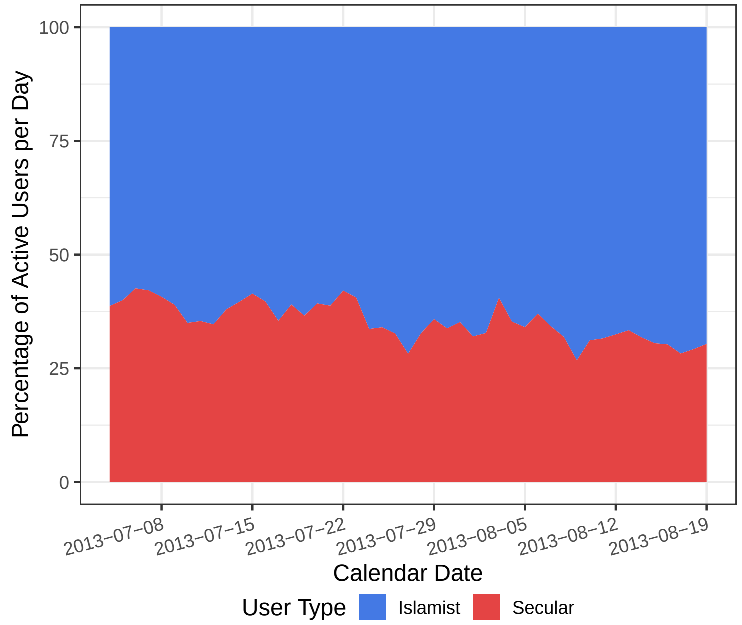

Figure G.1: The fraction of Islamist versus secular users over the sample period.

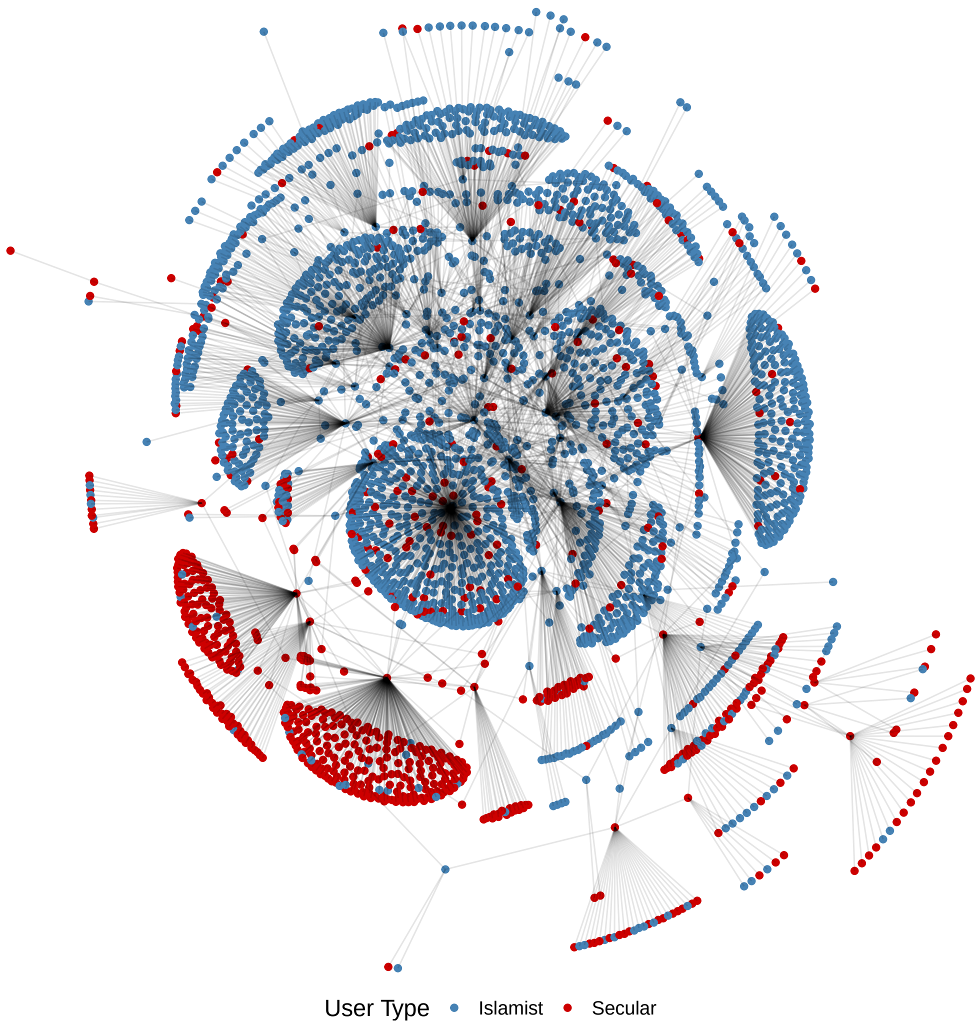

Figure G.2: Users' political affiliations and links based on a "snowball" sample of a randomly drawn user and her neighbors, neighbors' neighbors, etc. of the original network of users (Goodman, 1961).

H. Bayesian Estimation: Prior distributions and MCMC procedure

We specify the prior distributions as follows: $\theta \sim \mathcal{N}(0, \sigma_\theta^2)$, $\rho \sim \mathcal{N}(0, \sigma_\rho^2)$, $\boldsymbol{\beta} \sim \mathcal{N}(0, \Sigma_{\boldsymbol{\beta}})$, $\tau \sim \mathcal{N}(0, \sigma_\tau^2)$, $\kappa \sim \mathcal{N}(0, \sigma_\kappa^2)$, $\boldsymbol{\phi} \sim \mathcal{N}(0, \Sigma_{\boldsymbol{\phi}})$, $z_i \sim \mathcal{N}(0, \sigma_z^2)$, and $\sigma_z^2 \sim \mathcal{IG}\left(\frac{\nu_0}{2}, \frac{\chi_0}{2}\right)$, where $\mathcal{N}$ and $\mathcal{IG}$ are normal and inverse gamma conjugate priors. We choose the hyperparameters to make the prior distributions relatively flat and cover a wide range of parameter space. Specifically, we set $\sigma_\theta^2 = \sigma_\rho^2 = \sigma_\tau^2 = \sigma_\kappa^2 = 10$, $\Sigma_\beta = 10 \cdot I_{|\boldsymbol{\beta}|}$, $\Sigma_{\boldsymbol{\phi}} = 10 \cdot I_{|\boldsymbol{\phi}|}$, $\nu_0 = 2.2$ and $\chi_0 = 0.1$. Given the above prior distributions and the composite likelihood function of (24), we can derive the joint posterior distribution of Θ and $\mathbf{z}$. Since it is difficult to simulate draws from this high-dimensional joint posterior distribution, we implement the Metropolis-Hastings-within-Gibbs algorithm to simulate draws sequentially from the conditional posterior densities in the following steps:

1. simulate the random variable z_i using the Metropolis-Hastings (M-H) algorithm based on $p(z_i|\mathbf{s}, G, \Theta)$ for $i = 1, \ldots, n$.
2. simulate σ_z^2 using the conjugate inverse gamma conditional posterior distribution.
3. simulate θ using the M-H algorithm based on $p(\theta|\mathbf{s}, G, z, \Theta \setminus \theta)$.
4. simulate ρ using the M-H algorithm based on $p(\rho|\mathbf{s}, G, z, \Theta \setminus \rho)$.
5. simulate $\boldsymbol{\beta}$ using the M-H algorithm based on $p(\boldsymbol{\beta}|\mathbf{s}, G, z, \Theta \setminus \boldsymbol{\beta})$.
6. simulate τ using the M-H algorithm based on $p(\tau|\mathbf{s}, G, z, \Theta \setminus \tau)$.
7. simulate κ using the M-H algorithm based on $p(\kappa|\mathbf{s}, G, z, \Theta \setminus \kappa)$.
8. simulate $\boldsymbol{\phi}$ using the M-H algorithm based on $p(\boldsymbol{\phi}|\mathbf{s}, G, z, \Theta \setminus \boldsymbol{\phi})$.

I. Parameter Recovery Analysis

In this section, we conduct a Monte Carlo simulation study that attains two complementary objectives. On the one hand, it demonstrates that the proposed Bayesian MCMC estimation based on the composite likelihood function of (24) and the case-control approach described in Subsection 5.1 can indeed recover true parameter values. As an interesting side benefit of this exercise, we also confirm the direction of estimation bias observed in Tables 2 and 3 when individual unobserved heterogeneity (random effect) is ignored. On the other hand, the Monte Carlo simulations also provide support to the conjecture that underlay our theoretical analysis of the LIL scenario. More specifically, it shows that limit distribution $\tilde{\mu}_\eta(\cdot)$ defined in (20) provides a good approximation of the invariant behavior of the process under limited information and therefore can be suitably used as the likelihood for estimation in this context.

We set the number of Monte Carlo repetitions to 300. In each repetition, we generate the artificial networks G_t and the action profiles $\mathbf{s}_t$ through a data generating process (DGP) that mimics the dynamic process introduced in Subsection 2.2 for a setup with a number of nodes $n = 3,000$. In such a DGP, we generate the individual types γ_i from an expression of the form $\beta x_i + \tau z_i$, where the variables x_i represent observed individual characteristics that are generated from a mixture of normal distributions (specifically, two-fifths of the values are generated from a

procedure.

$\mathcal{N}(-4, 36)$ and three-fifths from $\mathcal{N}(4, 36)$), while the variables z_i represent unobserved individual random effects generated from a standard normal distribution, i.e. $z_i \sim \mathcal{N}(0, 1)$. The coefficients β and τ are set to 0.5. Then, the linking costs are determined through the expression $\zeta_{ij} = \phi_0 + h(x_i, x_j)\phi_1 - z_i - z_j$, where we set $h(x_i, x_j) = |x_i - x_j|$ and the coefficients ϕ_0 and ϕ_1 are fixed equal to 2 and 1, respectively. The true values of the local spillover effect θ and the global conformity effect ρ are set to 0.05 and 0.001 respectively, and the parameter η of the logistic disturbance is normalized to 1. Finally, for the LIL scenario, we fix the true rioting cost parameter κ to 1.5 and the belief weight φ to 0.4.

To generate the artificial data, we implement an iterative process that simulates the choice of individual network links and actions with the corresponding conditional probabilities. That is, in each iteration an individual is chosen at random for the option to update her network links or action – and, in the LIL scenario, as well as her beliefs – conditionally on the network links and actions chosen by others in the previous iteration. We run this iterative process sufficiently long and treat the last realization of the network and action profiles from this iterative process as our artificial data.

Next, we describe the implementation of this iterative process in more detail. On the one hand, to update the networking decisions of any given individual i, we use the following conditional probability for her network links ij:

$$\mu(a_{ij} = 1|\mathbf{s}, G_{-ij}) = \frac{\exp(a_{ij}(\theta s_i s_j - \phi_0 - |x_i - x_j|\phi_1 + z_i + z_j)}{1 + \exp(\theta s_i s_j - \phi_0 - |x_i - x_j|\phi_1 + z_i + z_j)}.$$

And, in order to speed up the simulation, we take advantage of the conditional independence of network links,[36] to update the choice of all i's links $\{a_{ij}\}_{j=1}^n$ synchronously. On the other hand, to update the action choice s_i of any given individual i, the details of course depend on the scenario under consideration. For the GI scenario, we use the following conditional probability for each action s_i:

$$\mu(s_i = 1|\mathbf{s}_{-i}, G) = \frac{\exp(\beta x_i + \tau z_i + \rho \sum_{j \neq i} s_j + \theta \sum_{j \neq i}^n a_{ij} s_j)}{2\cosh(\beta x_i + \tau z_i + \rho \sum_{j \neq i} s_j + \theta \sum_{j \neq i}^n a_{ij} s_j)},$$

where note that, in order to avoid the identification problem discussed in Subsection 5.1, we do not include the rioting cost κ in the GI scenario. Instead, to simulate the choice of action s_i for the LIL scenario, we use the following conditional probability for action choice:

$$\tilde{\mu}(s_i = 1|\mathbf{s}_{-i}, G, \boldsymbol{\psi}^{LIL}) = \frac{\exp(\beta x_i + \tau z_i - \kappa + \rho(n-1)\psi_i^{LIL} + \theta \sum_{j \neq i}^n a_{ij} s_j)}{2\cosh(\beta x_i + \tau z_i - \kappa + \rho(n-1)\psi_i^{LIL} + \theta \sum_{j \neq i}^n a_{ij} s_j)},$$

where ψ_i^{LIL} is generated from the induced stationary beliefs given in (9). Overall, the iteration procedure described executes the dynamic process formulated in Section 2.2 for rates of adjustment that lead individuals to update their network links and actions with equal frequency.

For estimation, we implement the Bayesian MCMC sampling with 30,000 iterations and drop the first 5,000 iterations for burn-in. The results for the GI scenario are summarized in Table I.1 and the results for the LIL scenario in Table I.2. The values reported in the tables are the mean and standard deviation of parameter estimates calculated across repetitions. In both tables, we see

[36] As explained in Subsection 5.1, it is important to bear in mind that the network links are *not* unconditionally independent due to the interdependence of actions in the presence of a positive spillover effect θ.

that the estimation exercise can successfully recover the true parameter values when considering the full econometric model with random effects (which is the true DGP). Moreover, in line with the pattern observed in our empirical study, when random effects are ignored the estimate of the local spillover θ is upward biased whereas that of the global-conformity parameter ρ is downward biased.

Table I.1: Simulation results of the global information (GI) scenario.

			with random effects		w/o random effects	
		DGP	Mean	Std.	Mean	Std.
Local spillover	(θ)	0.0500	0.0593	0.0108	0.1451	0.0072
Global conformity	(ρ)	0.0010	0.0009	0.0004	-0.0003	0.0002
Individual preference						
Individual characteristic x	(β)	0.5000	0.5059	0.1285	0.4155	0.0677
Random effect	(τ)	0.5000	0.4510	0.1911	–	–
Linking cost						
Constant	(ϕ_0)	2.0000	2.0145	0.0666	1.5291	0.0091
Diff. in characteristics $\lvert x_i - x_j\rvert$	(ϕ_1)	1.0000	1.0145	0.0064	0.8808	0.0034
Variance of random effect	(σ_z^2)	1.0000	1.1045	0.0353	–	–
Number of nodes				3000		
MC repetitions				300		

Table I.2: Simulation results of the local information and learning (LIL) scenario.

			with random effects		w/o random effects	
		DGP	Mean	Std.	Mean	Std.
Local spillover	(θ)	0.0500	0.0543	0.0124	0.0755	0.0067
Global conformity	(ρ)	0.0010	0.0011	0.0005	0.0006	0.0003
Weight of local observation	(φ)	0.4000	0.4580	0.1328	0.6255	0.1705
Individual preference						
Individual characteristic x	(β)	0.5000	0.5303	0.1770	0.5576	0.2040
Random effect	(τ)	0.5000	0.5451	0.2394	–	–
Rioting cost	(κ)	1.5000	1.6169	0.6518	1.9963	0.7965
Linking cost						
Constant	(ζ_0)	2.0000	1.9418	0.0633	1.4647	0.0084
Diff. in characteristics $\lvert x_i - x_j\rvert$	(ζ_1)	1.0000	1.0143	0.0056	0.8848	0.0032
Variance of random effect	(σ_z^2)	1.0000	1.1013	0.0350	–	–
Number of nodes				3000		
MC repetitions				300		

J. Additional Empirical Results

In the following we provide additional estimation results when using a different sampling rate in the case-control design discussed in Section 5.1. We find that by changing $m_{i,o}$ from $100+5\sum_{j\neq i} a_{ij}$ to $1000+5\sum_{j\neq i} a_{ij}$, the results in Tables J.1 and J.2 are qualitatively similar to those in Tables 2 and 3 in the main text. This illustrates that the results in Tables 2 and 3 are robust against using alternative sampling rates in our case-control design.

Table J.1: Robustness check for estimation results of the global information (GI) scenario – by setting $m_{i,o} = 1000 + 5\sum_{j\neq i} a_{ij}$.

		with random effects (1)	w/o random effects (2)
Local spillover	(θ)	0.1716***	0.2292***
		(0.0029)	(0.0026)
Global conformity	($\check{\rho}$)	3.10e-6***	3.02e-6***
		(9.18e-8)	(7.90e-8)
Individual preference			
Female	(β_1)	-0.0637***	-0.0562***
		(0.0090)	(0.0062)
Islamist	(β_2)	0.1007***	0.1103***
		(0.0055)	(0.0037)
(Log) followers	(β_3)	0.0119***	0.0090***
		(0.0017)	(0.0015)
Random effect	(τ)	0.0059***	–
		(0.0009)	
Linking cost			
Constant	(ϕ_0)	14.6860***	12.5847***
		(0.0226)	(0.0088)
Same gender	(ϕ_1)	-0.1629***	-0.1873***
		(0.0161)	(0.0073)
Same religiousness	(ϕ_2)	-0.0781***	-0.0011
		(0.0082)	(0.0061)
Diff. in followers count	(ϕ_3)	0.0858***	0.0987***
		(0.0032)	(0.0027)
Variance of random effect	(σ_z^2)	2.0719***	–
		(0.0175)	
Sample size (# of nodes)		225,578	

Notes: For the purpose of identification, we replace $\rho\sum_{j\neq i}^{n} s_j$ with $\check{\rho}(n-1)\bar{s}$ and drop κ in the GI scenario. The parameter estimates reported in this table are the posterior mean and the posterior standard deviation from the Bayesian MCMC sampling procedure. The asterisks ***(**,*) indicate that the 99% (95%, 90%) highest posterior density interval (HDI) of the corresponding draws does not cover zero.

Table J.2: Robustness check for estimation results of the local information and learning (LIL) scenario – by setting $m_{i,o} = 1000 + 5\sum_{j\neq i} a_{ij}$.

		with random effects (1)	w/o random effects (1)
Local spillover	(θ)	0.0709***	0.1601***
		(0.0036)	(0.0025)
Global conformity	(ρ)	2.36e-6***	1.71e-6***
		(5.15e-8)	(3.88e-8)
Weight of local observation	(φ)	0.0961***	0.0802***
		(0.0050)	(0.0040)
Individual preference			
Female	(β_1)	-0.0565***	-0.0523***
		(0.0112)	(0.0087)
Islamist	(β_2)	0.1144***	0.1181***
		(0.0063)	(0.0047)
(Log) followers	(β_3)	0.0059**	0.0032**
		(0.0020)	(0.0015)
Random effect	(τ)	0.0055***	–
		(0.0004)	
Rioting cost	(κ)	-0.2951***	-0.3065***
		(0.0113)	(0.0085)
Linking cost			
Constant	(ϕ_0)	14.6812***	12.5650***
		(0.0226)	(0.0094)
Same gender	(ϕ_1)	-0.1655***	-0.1947***
		(0.0148)	(0.0087)
Same religiousness	(ϕ_2)	-0.0776***	-0.0038
		(0.0071)	(0.0065)
Diff. in followers	(ϕ_3)	0.0871***	0.0990***
		(0.0033)	(0.0025)
Variance of random effect	(σ_z^2)	2.1127***	–
		(0.0205)	
Sample size (# of nodes)		225,578	

Notes: The parameter estimates reported in this table are the posterior mean and the posterior standard deviation from the Bayesian MCMC sampling procedure. The asterisks ***(**,*) indicate that the 99% (95%, 90%) highest posterior density interval (HDI) of the corresponding draws does not cover zero.

References

Anderson, D. F. (2012). An efficient finite difference method for parameter sensitivities of continuous time Markov chains. *SIAM Journal on Numerical Analysis*, 50(5):2237–2258.

Bollobás, B. and Janson, S. and Riordan, O. (2007). The phase transition in inhomogeneous random graphs. *Random Structures & Algorithms.*, 31(1):3–122.

Borge-Holthoefer, J., Magdy, W., Darwish, K., and Weber, I. (2015). Content and network dynamics behind Egyptian political polarization on Twitter. In *Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing*, 700–711. ACM.

Gibson, M. A. and Bruck, J. (2000). Efficient exact stochastic simulation of chemical systems with many species and many channels. *The Journal of Physical Chemistry A*, 104(9):1876–1889.

Goodman, Leo A. (1961). Snowball sampling. *The Annals of Mathematical Statistics*, 148-170.

Grimmett, G. (2010). *Probability on graphs.* Cambridge University Press.

Jackson, M. and Rogers, B. (2007). Meeting strangers and friends of friends: How random are social networks? *American Economic Review*, 97:890–915.

Meyer, C. D. (2000). *Matrix analysis and applied linear algebra.* Society for Industrial and Applied Mathematics.

Seneta, E. (2006). *Non-negative matrices and Markov Chains.* Springer.

Van Mieghem, P. (2011). *Graph spectra for complex networks.* Cambridge University Press.

Weber, I., Garimella, V. R. K., amd Batayneh, A. (2013). Secular vs. islamist polarization in Egypt on Twitter. In *Proceedings of the 2013 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining*, 290–97.

Wainwright, M. J. and Jordan, M. I. (2008). Graphical models, exponential families, and variational inference. *Foundations and Trends in Machine Learning*, 1(1-2):1–305.