Riot Networks and the Tullock Paradox: An application to the Egyptian Arab Spring[☆]

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

[a] *Department of Economics, National Taiwan University, Taipei, Taiwan.*
[b] *Tilburg University, Warandelaan 2, Tilburg, Netherlands 5037 AB.*
[c] *Tinbergen Institute and Department of Spatial Economics, VU Amsterdam, De Boelelaan 1105, 1081 HV Amsterdam, The Netherlands.*
[d] *ETH Zurich, Swiss Economic Institute (KOF), Leonhardstrasse 21, 8092 Zurich, Switzerland.*
[e] *Department of Decision Sciences, Bocconi University, Via Roentgen, 1, 20136 Milan, Italy.*

Abstract

We study a dynamic model of collective action – for concreteness, we speak of a riot – in which agents interact through, and learn from, a co-evolving social network. We consider two different scenarios on how agents form their expectations when changing their behavior. In one of them, conceived as a "benchmark", they are assumed to be completely informed of the prevailing state (action profile and network). Instead, in the alternative scenario, agents are assumed to shape their expectations about the state from a combination of local observation and social learning (modeled à la DeGroot). In both cases we provide a complete characterization of the long-run behavior of the system. While the first assumption of complete information is common, the second one is arguably more realistic. Furthermore, we show that only the latter assumption yields the following twin conclusion: a significant long-run probability of successful collective action *and* a meaningful time scale of convergence to this state of affairs. This, we argue, suggests a plausible route to understanding what otherwise seems a puzzle, i.e. how do very large populations attain ("coordinate on") collective action. Finally, we illustrate the empirical potential of the model by showing that it can be efficiently estimated for the so-called Egyptian Arab Spring by relying on large-scale cross sectional data on agents' choices and their network of interactions. The estimation results are fully in line with the predictions of the theory.

Key words: collective action, networks, riots, protests, DeGroot, social learning
JEL: D74, D72, D71, D83, C72

[☆] We would like to thank Fabrizio Zilibotti, Yves Zenou, James Fenske, Michihiro Kandori, Kei Ikegami, Ruben Enikolopov and seminar participants at the Tinbergen Institute, Northwestern, Bocconi and Columbia University for the helpful comments.

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

1. Introduction

Social networks provide the structure through which people interact and communicate when facing problems of collective action, e.g. addressing social emergencies, staging peaceful protests, or igniting violent riots. Nowadays, a large fraction of such social interaction is carried out virtually, through online social media such as Facebook, Twitter, or Instagram. This, in principle, makes it possible to collect massive data on the operation of large-scale social networks, which in turn opens up rich possibilities for an integrated theoretical and empirical study of collective-action problems in large populations. To develop theoretical and empirical methods that can be used to to handle and better understand this type of data on collective-action phenomena is, in a nutshell, the main objective of the present paper.

The approach we pursue to study the problem has, therefore, a theoretical and an empirical side to it. On the theoretical front, we model the situation as a *population game* played on an *evolving network* in which each agent decides whether or not to join the collective action (e.g. a social riot) on the basis of:

(a) the *observation* of the *actions* chosen by her network neighbors;
(b) the *beliefs* she holds on the average action chosen by the rest of the population.

While, in general, the beliefs in (b) could be determined by some arbitrary expectation-formation function of the current state of the system, only two concrete alternatives will be formally studied and compared:

(b.1) agents have perfect information on the average action of all others;
(b.2) they form *beliefs* by combining the information they gather locally from their neighbors on the actions they choose (see (a)) and the the beliefs they hold.

Alternative (b.1) is the classical formulation considered by the evolutionary learning literature in games, which may be described as complete information. In contrast, alternative (b.2) presumes that not just observation but also social learning (i.e. the process by which the beliefs of others can influence one's own) is *constrained by the current social network*.

A primary aim of our model is to shed light on the following fundamental issue: How does a large population, connected through a dynamic social network, manage to achieve collective action? In particular, this question speaks to one of the several paradoxes associated to Gordon Tullock (see Tullock [1971]), which concerns the form of collective action that underlies massive social protests (or "revolutions"). It has been succinctly explained by Shadmehr [2021] as follows:

> "... societies are large and one person's effect on the success of the revolution is negligible, while participating in a revolution is costly, so that revolutions should not occur."

In addressing this central issue, we shall argue that, at least from a modeling viewpoint, the usual assumption of complete information embodied by (b.1) above is not as useful as (b.2). One reason for this is that, if the population is quite large, the constrained local information reflected by the latter assumption seems so much more plausible. There are, however, two additional considerations that are equally important. First, we find that, under (b.1), collective action arises only when a large fraction of the population (more than half) displays a high propensity/bias for it i.e. are "militants." Instead, under (b.2), we have clusters of collection action that display a relatively small size arising even when the fraction of militants is small. This is and interesting and intuitive feature in itself. It is also a key factor underlying the second consideration that favors a model based on (b.2): while the expected time scale required

for collective action to materialize under (b.1) is inordinately long, hence making it a largely irrelevant long-run prediction, (b.2) delivers much shorter (and hence relevant) expected times, even if the population is large. Thus, somewhat paradoxically, our theoretical framework leads to the somewhat paradoxical conclusion that more limited and local (network-bound) information greatly facilitates collective action. And, as we shall see, social learning, also network-bound, plays a a key role in the process.

Methodologically, the essential step in our analysis of the model is arriving at a full characterization of the long-run behavior of the induced evolutionary process, as captured by the corresponding *limit probability distribution* over possible states. This characterization has two important implications: one on the development of the theory, and another on its empirical application. On the theory, it allows us to conduct an exhaustive comparative analysis of the effect of the different forces (parameters) at work. For example, we can answer the following type of questions. What is the comparative impact of observation and learning (cf. (b.2)) in shaping agents' beliefs? What is the effect of individual heterogeneity (and, possibly, linking homophily) in the formation of the social network? Does easier/cheaper connectivity – and hence a more dense network – help coordination and collective action? What is the influence on behavior exerted by the force toward local conformity (i.e. toward alignment with one's neighbors)? And the impact of global conformity (i.e. alignment with the population at large)?

On the empirical side, our characterization of the limit distribution of the process as an explicit function of the parameters is also especially fruitful. For, in essence, it means that the theory directly provides a likelihood function that can be used for a structural estimation of the model. And then, of course, such a structural estimation may serve both to test the theory and to apply it to the study of specific real-world instances of collective action. This is what we do in the final part of the paper, where we bring our theory to the analysis of a "big" data set on one of the massive social protests sparked by the so-called Arab Spring.

Specifically, we focus on the Egyptian context, and use Twitter data that include tweeting and networking information of 194,760 users during the military backlash that, in June 2013, toppled the Morsi government. This information is used for two important purposes. First, we trace the underlying social network by identifying inter-agent links with those bilateral relationships that show reciprocal influence. Second, we rely on state-of-the-art machine-learning techniques 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 of protest).

Given the large size of our data set, conventional maximum-likelihood methods are not feasible, and neither are the simulation-based approaches proposed in the literature [Badev, 2021; Hsieh et al, 2020; Hsieh et al., 2021; Mele, 2017]. Therefore, we consider the maximum composite likelihood estimation approach [Lindsay, 1988; Varin et al., 2011] and apply to it suitable computational techniques that combine sparse-matrix and parallel-implementation procedures. The estimates thus obtained are fully in line with the theory and also provide an intuitive understanding of the context being studied. For example, we find that both social learning and local observation are significant components (the former having a higher weight than the latter) in the process through which agents form their expectations on the prevailing state. These endogenous expectations are one of the forces that drive their behavior. The other force – which in our model is state-independent and hence exogenous – is given by the inherent costs, risks, and benefits that agents assess when they consider if contributing to collective action. Their net effect is also estimated as part of the econometric exercise, and it turns out to have a positive sign. 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.

We conclude this Introduction with a brief discussion of the relationship between our research and existing literature, theoretical or/and empirical. 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 was studied, for example, by Blume [1993], Brock and Durlauf [2001] and Morris [2000], while the analysis has been extended to co-evolving endogenous networks by Jackson and Watts [2002] and Goyal and Vega-Redondo [2005]. A second strand is the booming recent research on learning in networks (cf. Golub and Sadler [2016]). As a small sample, we may refer to the influential contributions by DeMarzo et al. [2003], Jackson and Golub [2010] and Acemoglu et al. [2014]. While the latter paper adopts a Bayesian approach to the problem, the former 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 (see Chandrasekhar et al. [2015]) and is a key component of our model. Thirdly, we mention the literature that, in line with the seminal work of Granovetter [1978], models the behavior of individuals facing a problem of collective action as a threshold phenomenon. Two interesting contributions displaying this feature are the papers by Chwe [2000] and Barberà and Jackson [2020]. To sum up, our approach in this paper aims at integrating the signature features of the three aforementioned strands of literature – coordination, learning, and threshold behavior – into a common framework. As anticipated, such an integration yields valuable theoretical insights into the problem of collective action.

Turning now to the empirical literature, 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, thus 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]. The are also a few papers, such as those by Boucher [2016] and Badev [2021], which apply these methods to carry out, as in our case, a structural estimation of an underlying theoretical model. In contrast with our paper, however, their approach relies on numerical methods that are computationally unfeasible in dealing with the large data sets that are our primary concern. As we shall explain in detail, our closed-form determination of the predictions of the model permits an (exact) analytical application of the so-called "maximum composite likelihood" method to estimate the model parameters, despite our large sample size.

Next, we refer to the recent literature that shares with this paper its concern with the effect of social media on facilitating collective action – in particular, on supporting massive events of social protest. An early study of the phenomenon was undertaken by González-Bailón et al. [2011], who 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." Another good example is the paper by Acemoglu et al. [2016], which focuses on the same instance of social protests as we do – the Egyptian Arab Spring – and finds that a rise in Twitter activity tends to precede the triggering of social protests.

In a similar vein but with a different methodological perspective, the recent paper by Enikolopov et al. [2020] studies the wave of social protests that took place in Russia in 2011. Its 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, in essence, is in line with the importance that our own analysis attributes to the social network as a channel for exchange of information across neighbors. For, as already explained (recall items (a)-(b) above), our approach is grounded on the idea that agents gather information "horizontally," not only observing the neighbors' actions but also learning/exchanging their beliefs.

The key role played by the network in shaping people's beliefs is also supported by the experimental evidence obtained by Cantoni et al. [2019] in another interesting paper along these lines. The primary contribution of their paper is, in their words, "... to isolate the causal effect of variation in beliefs regarding others' protest participation on one's own protest participation." In their experimental context, this conclusion is reached through targeted interventions that selectively affect agents' beliefs. In our case, where we base our analysis in Twitter (non-experimental) data, our approach to identifying agents' attitudes and beliefs is based on an extensive application of the state-of-the art techniques developed by the field of Natural Language Processing (NLP), used in combination with the Arabic pre-trained language model *AraBERT*.[1]

The rest of the paper is organized as follows. In Section 2 we introduce the benchmark model where agents are assumed to enjoy global (thus complete) information. In Subsection 2.1, we describe the game-theoretic setup, in Subsection 2.2 the law of motion for actions and links, and in Subsection 2.3 we introduce the different belief formation scenarios. Section 3 provides a theoretical analysis of the model, with Subsection 3.1characterizing the long-run behavior of the system and Subsection 3.2 comparing the two informational scenarios that we consider. In Section 4 we describe the data used for the econometric analysis of the model. The estimation exercise itself is included in Section 5: Subsection 5.2 explains the estimation procedure, Subsection 5.3 reports the results and Subsection 5.4 discusses further properties of the estimation algorithm. Section 6 conducts counterfactual analyses. Finally, Section 7 concludes. Extensions of the model can be found in Appendix A, a finite equilibrium characterization is provided in Appendix C, Appendix D describes the historical context of our data and all proofs of the theoretical results are relegated to Appendix E.

2. The Model

For the sake of clarity, 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 what are the primitives that define the interaction, induce the payoffs, and characterize a state of the system. Then, in Subsection 2.2 we specify the dynamics, i.e. the law of motion of actions and links that changes the state over time. Naturally, this dynamics is crucially dependent on the beliefs agents hold. Subsection 2.3 explains alternative formulations on how such beliefs are formed, depending on the information agents have access to.

[1]Natural language processing is the branch of the wide area studying Artificial Intelligence that applies machine learning methods to text and is becoming quite widely used in social sciences. Among the growing number of overview articles and other sources that can be checked as suitable references, we can list Gentzkow et al. (2017) for a focus on economics, while Grimmer and Brandon [2013] focuses on political science, Evans and Aceves [2016] on sociology, or Humphrey and Wang [2017] on marketing. On the other hand, specifically concerning *AraBERT*, see Antoun et al. [2020].

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 and action s_i, which is a dichotomous decision of whether to join the riot or not. Formally,[2] 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. For simplicity, we shall consider undirected networks, which means that the matrix G is symmetric, i.e. $a_{ij} = a_{ji}$ for all $i, j \in \mathcal{N}$. The set of all such symmetric square matrices of size n is denoted by $\mathcal{G}^n$.[3]

Given any given 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 every agent $i \in \mathcal{N}$ holds some point belief $\psi_i \in [-1, +1]$ on the average action chosen by all other agents in the population. Then, assuming that agents observe perfectly the actions chosen by their (immediate) network neighbors, the expected payoff of any given agent i is defined as follows:

$$\pi_i(\mathbf{s}, G; \psi_i) = (\gamma_i - \kappa)s_i + \sum_{j=1, j\neq i}^{n} \left[\rho s_i \psi_i + a_{ij}(\theta s_i s_j - \zeta_{ij})\right], \tag{1}$$

where

- $\gamma_i \in \{-1, +1\}$ 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 cost/risk of rioting);
- $\rho \in (0, 1)$ is the parameter modulating the strength of the global (population-wide) conformity effect, which applies to the agent i's expectation on the aggregate action of all other agents induced by ψ_i], i.e. $\sum_{j\neq i} \psi_i = (n-1)\psi_i$;
- $\theta \in (0, 1)$ is a parameter capturing the force of local conformity, which applies to the *accurate perception* of the aggregate action $\sum_{j=1}^n a_{ij}s_j$ of i's neighbors;
- $\zeta_{ij} \geq 0$ is the linking cost between agents i and j.

For presentational convenience, our presentation and discussion of the theory will be carried out under the presumption that κ is non-negative. The model, however, allows for a general (type-independent) preference for rioting, as captured by some $\kappa < 0$. Since the two actions are symmetric relative to κ, all the results derived in the following sections apply after switching $\gamma_i = +1$ and $\gamma_i = -1$ in the equilibrium characterization below.[4]

For simplicity, the model attributes any possible differences in linking costs to whether or

[2] This is also the customary convention adopted in the analysis of the classical Ising model [cf. Grimmett, 2010].

[3] To understand how the analysis could be adapted if links are taken to be directed, see Appendix A.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.

[4] In fact, as we shall find in Section 5, the Egyptian Arab-Spring context considered by our empirical application of the model displays a negative estimate for κ.

not the two agents involved, i and j, are of the same type, i.e. have the same idiosyncratic action bias, γ_i and γ_j:[5]

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

with $0 \leq \zeta_1 < \zeta_2$. This specification implies that the linking cost between agents with the same idiosyncratic preferences are smaller. Its implication on network formation is that agents must display a bias/preference for connecting to those of the same type. This phenomenon, known as homophily, has been shown to be a quite 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 consists of three components: action adjustment, link creation, and link removal, which will be separately defined below. Naturally, these adjustments will be assumed to depend on the expected payoffs perceived by the agents at the time of their adjustment. This, of course, requires specifying how each agent i forms her beliefs ψ_i on the aggregate action of others, $\sum_{j \neq i} a_j$. 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.

As customary in the evolutionary literature, expected payoffs will also 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 route to motivate it is to suppose that the noise captures agents' uncertainty about the behavior of others or the consequences (say costs) of their actions. Finally, a third motivation that has been highlighted by evolutionary game theory [cf. e.g. 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 $G_t \in \mathcal{G}^n$. Formally, its law of motion can be described as follows.[6]

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

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

[6] The adjustment process has some similarity to that of Hsieh et al. [2021], time being measured continuously and revision opportunities arriving as a Poisson process [cf. Sandholm, 2010].

Action adjustment: *At rate $\chi \geq 0$, every agent $i \in \mathcal{N}$ is randomly given an independent opportunity to change his current action $s_{it} \in \{-1, +1\}$ to the alternative s_i'. Upon receiving this opportunity, the action change is implemented if the agent perceives it profitable 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 formation: *At rate $\lambda \geq 0$, every pair of agents ij who are not already connected are randomly given an independent opportunity to form a link. Upon receiving this opportunity, the link is established 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:*

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

Link removal: *At rate $\xi \geq 0$, each pair of connected agents i, j receives an opportunity to remove their link. The link is eliminated if at least one agent finds it profitable in terms of the expected payoffs specified in (1) and an additive perturbation given by an i.i.d. logistically-distributed (unbiased) shock $\varepsilon_{ij,t}$. Upon receiving this opportunity, the marginal payoffs from removing the link ij are perturbed by an additive random shock $\varepsilon_{ij,t}$. Thus, the probability that the link ij is removed is given by:*

$$\begin{aligned}&\mathbb{P}\left(\boldsymbol{\omega}_{t+\Delta t} = (\mathbf{s}_t, G_t - ij) | \boldsymbol{\omega}_{t-1} = (\mathbf{s}, G_t)\right) = \\ &\quad \lambda\, \mathbb{P}\big[\{\pi_i\left(\mathbf{s}_t, G_t - ij; \psi_i(\mathbf{s}_t, G_t)\right) - \pi_i(\mathbf{s}_t, G_t; \psi_i(\mathbf{s}_t, G_t)) + \varepsilon_{ij,t} > 0\} \cup \\ &\quad \{\pi_j(\mathbf{s}_t, G_t - ij; \psi_i(\mathbf{s}_t, G_t)) - \pi_j(\mathbf{s}_t, G_t; \psi_i(\mathbf{s}_t, G_t)) + \varepsilon_{ij,t} > 0\}\big] \Delta t + o(\Delta t).\end{aligned} \tag{5}$$

Throughout we will 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 cdf $F(\cdot)$ is given by $\frac{e^{\eta x}}{1+e^{\eta x}}$ and we can write the action-adjustment rule (3) in the following explicit form:[7]

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

And, proceeding analogously for the link adjustment rules (4) and (5), we arrive at the following corresponding expressions for link formation

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

[7] 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 $F_\varepsilon(y) = 1 - F_z(-y) = \frac{e^{\eta y}}{1+e^{\eta y}}$.

and link removal

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

where note that the link-adjustment probabilities are identical for the two agents involved in any link change (be it creation or removal) because, given the logistic noise formulation, 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 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. their average action). For conciseness, this first scenario is labeled as one of *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:

$$\psi_i^{GI}(\boldsymbol{\omega}) = \frac{1}{n-1} \sum_{j=1, j \neq i}^{n} s_j \qquad (i = 1, 2, \ldots, n). \tag{6}$$

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

In the alternative scenario, which we label as one of *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, and
(b) the interaction with (learning from) 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, their degree $d_i = \sum_{j \neq i} a_{ij}$. On the other hand, the local interaction/learning in (b) is modeled along the lines of the well-known framework proposed by DeGroot [1974].[8] More specifically, we posit that, given the state $\boldsymbol{\omega_t} = (\mathbf{s}_t, G_t) = [(s_{1t}, s_{2t}, ..., s_{nt}), (a_{jk,t})_{j,k \in \mathcal{N}}]$ prevailing at any given time t in the evolutionary adjustment process, there is a sequence of learning rounds, indexed

[8] See also Berger [1981]; DeMarzo et al. [2003]; Golub and Jackson [2012]; Jackson and Golub [2010]. Chandrasekhar et al. [2015]; Grimm and Mengel [2015] provide empirical evidence that individuals that attempt to learn the underlying state of the world in a network are well described by DeGroot-type models.

by $u = 0, 1, 2, ...$, where the point beliefs $\mathbf{p}_t^u = (p_{1t}^u, p_{2t}^u, ..., p_{nt}^u)$ are updated as follows:

$$p_{it}^{u+1} = \varphi \underbrace{\frac{1}{d_{it}} \sum_{j=1}^{n} a_{ij,t} s_{jt}}_{\substack{\text{local average} \\ \text{actions}}} + (1-\varphi) \underbrace{\frac{1}{d_{it}+1} \left[p_{it}^u + \sum_{j=1}^{n} a_{ij,t} p_{jt}^u \right]}_{\substack{\text{local average} \\ \text{beliefs}}} \qquad (i = 1, 2, ..., n) \tag{7}$$

where $\varphi \in (0, 1)$ is the updating weight given to local observation, while the complementary value $(1 - \varphi)$ is the weight given to the social-learning component of the updating rule. 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 exist and is unique. To see it, let us write (7) in compact matrix form as follows:

$$\mathbf{p}_t^{u+1} = \varphi \mathbf{D}_t^{-1} \mathbf{A}_t \mathbf{s}_t + (1-\varphi) \widehat{\mathbf{D}}_t^{-1} \widehat{\mathbf{A}}_t \, \mathbf{p}_t^u \tag{8}$$

where $\mathbf{D}_t \equiv \text{diag}(d_{1t}, \ldots, d_{nt})$ is the diagonal matrix of agents' degrees at t, $\widehat{\mathbf{D}}_t \equiv \mathbf{D}_t + \mathbf{I}_n$ with $\mathbf{I}_n$ being the identity matrix, $\widehat{\mathbf{A}}_t \equiv \mathbf{A}_t + \mathbf{I}_n$, 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

$$\mathbf{p}_t^* = \varphi \left[\mathbf{I}_n - (1-\varphi) \widehat{\mathbf{D}}_t^{-1} \widehat{\mathbf{A}}_t \right]^{-1} \mathbf{D}_t^{-1} \mathbf{A}_t \mathbf{s}_t \tag{9}$$

which is a well-defined expression since the matrix $\widehat{\mathbf{D}}_t^{-1} \widehat{\mathbf{A}}_t$ is row-stochastic and $\varphi > 0$.[9]

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:

$$\boldsymbol{\psi}^{LIL}(\boldsymbol{\omega}) = \varphi \left[\mathbf{I}_n - (1-\varphi) \widehat{\mathbf{D}}^{-1} \widehat{\mathbf{A}} \right]^{-1} \mathbf{D}^{-1} \mathbf{A} \mathbf{s}, \tag{10}$$

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, proceeding in parallel for the two belief-formation scenarios considered, GI and LIL. The overall analysis is also carried out in two steps. First, in Subsection 3.1, we describe the invariant probability distributions that summarize the long-run behavior of the process in each alternative scenario. Then, in Subsection 3.2 we compare the long-run predictions (for small noise) of the two contexts, paying special attention to the following key questions: (i) when does collective action arise with significant

[9] 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.

probability, even when the population is large; (ii) what is, in expected terms, the delay involved in arriving at such a state of affairs.

3.1. Long-run Behavior

In order to characterize the long-run behavior of the system, we shall show that its adjustment process (on both actions and links) can be described in terms of a suitable potential – a strict potential function in the case of GI, and an approximate one for LIL.

Starting with the GI scenario, the key point to note is that, given the belief-formation mapping $\boldsymbol{\psi}^{GI} = \left(\psi_i^{GI}\right)_{i\in\mathcal{N}}$ defined in (6), the function $\Phi : \Omega \to \mathbb{R}$ given by:

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

which can be rewritten as

$$\Phi(\boldsymbol{\omega}) = \sum_{i=1}^{n} \left\{ (\gamma_i - \kappa)s_i + \tfrac{1}{2}\left[\rho s_i (n-1)\psi_i^{GI}(\boldsymbol{\omega}) + \sum_{j\neq i} a_{ij}(\theta s_i s_j - \zeta_{ij})\right] \right\} \tag{12}$$

is a *potential* for the expected payoff functions given in (1). This means that, for any change in a *single* component of the state (an action or a link) that involves an agent i, the change on the expected payoffs $\pi_i(\cdot; \psi_i^{GI}(\cdot))$ anticipated by this agent must match exactly the change induced on the function $\Phi(\cdot)$. Note that, in ensuring this requirement, this function crucially relies on the additivity of individual payoffs and on the symmetry of any payoff obtained from interaction, be it local or global. Thus, for any revision of action or link altering the interaction between two agents, the desired property obtains if the function $\Phi(\cdot)$ divides equally the *individual* payoff change expected by the agent "initiating" the revision between the pair of agents affected by it.

Formally, we declare a function Φ to be a potential if the following two conditions hold:

$$\begin{aligned} \forall i \in \mathcal{N},\, \forall \mathbf{s} \in \mathcal{S},\, \forall s_i' \in \mathcal{S}_i,\ \text{let } \boldsymbol{\omega} = (s_i, \mathbf{s}_{-i}, G),\, \boldsymbol{\omega}' = (s_i', \mathbf{s}_{-i}, G) \in \Omega. \\ \text{then, } \Phi(\boldsymbol{\omega}') - \Phi(\boldsymbol{\omega}) = \pi_i(\boldsymbol{\omega}'; \psi_i^{GI}(\boldsymbol{\omega})) - \pi_i(\boldsymbol{\omega}; \psi_i^{GI}(\boldsymbol{\omega})), \end{aligned} \tag{13}$$

and

$$\begin{aligned} \forall i, j \in \mathcal{N},\, \forall G \in \mathcal{G}^n,\, \forall \mathbf{s} \in \mathcal{S},\ \text{let } \boldsymbol{\omega} = (\mathbf{s}, G),\, \boldsymbol{\omega}' = (\mathbf{s}, G \pm ij) \in \Omega; \\ \text{then, } \Phi(\boldsymbol{\omega}') - \Phi(\boldsymbol{\omega}) = \pi_i(\boldsymbol{\omega}'; \psi_i^{GI}(\boldsymbol{\omega})) - \pi_i(\boldsymbol{\omega}; \psi_i^{GI}(\boldsymbol{\omega})) \\ = \pi_j(\boldsymbol{\omega}'; \psi_j^{GI}(\boldsymbol{\omega})) - \pi_j(\boldsymbol{\omega}; \psi_j^{GI}(\boldsymbol{\omega})), \end{aligned} \tag{14}$$

where $G \pm ij$ stands for the network given by G with the link ij added $(+)$ or deleted $(-)$. For the sake of completeness, the conclusion derived from the previous informal discussion is stated in the following result, whose proof is included in Appendix E.

Proposition 1. *Given beliefs $\boldsymbol{\psi}_i^{GI}(\cdot)$, the function $\Phi(\cdot)$ given in (11) 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, conditions (13)-(14) hold.*

The previous result yields a number of interesting consequences. A standard one is that the noiseless best-response ("myopic") adjustment of actions and links converges, almost surely, to

the unique equilibrium (stationary point) where the potential is maximized.[10] In terms of our model, such an adjustment process corresponds to the limit situation where the persistent noise introduced in our adjustment subprocesses (3)-(5) becomes vanishingly small, i.e. $\eta \to \infty$. But even for any arbitrary η (i.e. possibly large noise), the fact that the utilities of our theoretical framework define a potential yields a sharp (probabilistic) prediction on the long-run behavior of the stochastic process induced in the GI informational scenario. For, as stated by our next result, an adaptation of arguments used in standard models of statistical physics (spelled out in Appendix E) lead to the following result.

Proposition 2. *Consider the stochastic process $(\boldsymbol{\omega}_t)_{t\in\mathbb{R}_+}$ defined by (3)-(6), 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 $\psi_i^{GI}(\cdot)$ for each $i \in \mathcal{N}$, as specified in (6). Then, this process induces an ergodic Markov chain whose unique invariant distribution μ^η, defined on the measurable space $(\Omega, \mathcal{F})$, is determined for every $\omega = (\mathbf{s}, G) \in \Omega$ as follows:*

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

The previous proposition provides an explicit solution of the model in the GI scenario by specifying, in closed form, how the probability distribution μ^η that characterizes the long-run behavior of the process depends on the noise level modulated by η and all other parameters of the model. The empirical validity of such a theoretical prediction, embodied by (15), is illustrated in Appendix B.1 through numerical simulations. Specifically, Figure 12 focuses on the average degree and average action, tracing how they are affected by changes on the linking cost ζ, a key parameter of the model. We find that the dependence on ζ exhibited by the simulations is well aligned with the theory, as it is closely approximated by the mean degree and mean action induced by the corresponding distribution μ^η.

Next, we turn to conducting a parallel analysis for the LIL scenario, where beliefs are governed by the mapping $\boldsymbol{\psi}^{LIL}$ defined in (10). Specifically, we propose a counterpart of the potential posited in (12)in the form of a function $\tilde{\Phi} : \Omega \to \mathbb{R}$ that, for all $\boldsymbol{\omega} = (\mathbf{s}\,, G) \in \Omega$, is defined as follows:[11]

$$\tilde{\Phi}(\boldsymbol{\omega}) = \sum_{i=1}^{n}\left\{(\gamma_i - \kappa)s_i + \rho s_i(n-1)\psi_i^{LIL}(\boldsymbol{\omega}) + \tfrac{1}{2}\left[\sum_{j\neq i} a_{ij}(\theta s_i s_j - \zeta_{ij})\right]\right\}. \tag{16}$$

The function $\tilde{\Phi}$ is, of course, much more complicated than Φ. For, in contrast with the previous case, it embeds the belief-formation mapping $\boldsymbol{\psi}^{LIL}$ that relies on information about the whole architecture of the prevailing network. This means, in particular, that it fails to be separately linear in the different components of the prevailing state, actions and links. In what follows, however, we formulate (and later support, numerically and econometrically) the conjecture that the function $\tilde{\Phi}$ works well as a useful approximation for a strict potential function. For want of

[10] Alternatively, this equilibrium can be viewed as a Nash equilibrium of a corresponding 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.

[11] Note that, in contrast with Φ, the function $\tilde{\Phi}$ does *not* include the "equal-splitting factor" of 1/2 affecting the individual expected payoffs anticipated from global interaction. The reason for it will be discussed when we explain in more detail – see item (c) below – our approach to studying the LIL scenario.

a better term, we call it a *quasi-potential*.

A preliminary step in motivating the aforementioned conjecture is based on the following observations.

(a) Consider any single-component change from a state $\boldsymbol{\omega}$ to another state $\boldsymbol{\omega}'$, after which the corresponding beliefs remain unaltered (i.e. $\boldsymbol{\psi}^{LIL}(\boldsymbol{\omega}) = \boldsymbol{\psi}^{LIL}(\boldsymbol{\omega}')$). Then, the payoff change perceived by an agent $i \in \mathcal{N}$ involved in the adjustment exactly matches the change displayed by $\tilde{\Phi}$ – that is, $\pi_i(\boldsymbol{\omega}', \psi_i^{LIL}(\boldsymbol{\omega})) - \pi_i(\boldsymbol{\omega}, \psi_i^{LIL}(\boldsymbol{\omega})) = \tilde{\Phi}(\boldsymbol{\omega}') - \tilde{\Phi}(\boldsymbol{\omega})$.

(b) Consider a state $\boldsymbol{\omega}$ where the social network is sparse (roughly, most of the possible links are absent) but agents typically have a sizable degree (still much lower than the population size). Then, one may expect that any single-component revision toward some other state $\boldsymbol{\omega}'$ should induce individual belief changes $\psi_j^{LIL}(\boldsymbol{\omega}') - \psi_j^{LIL}(\boldsymbol{\omega})$ that are small and, furthermore, should affect significantly only those agents j who are close in the network to an agent i involved in the change.

(c) In view of (a)-(b), an adaptation of the considerations underlying the potential Φ for the GI scenario suggests that the equal-splitting factor of $1/2$ should be applied only to the payoffs resulting from local interaction but are best ignored, as an approximation, for its counterpart quasi-potential $\tilde{\Phi}$ to be used under LIL.

The combination of all three observations above suggest that, provided the behavioral noise is small, the changes over time of the function $\tilde{\Phi}(\boldsymbol{\omega}_t)$ must be largely dominated by the corresponding payoff changes expected by those agents who revise their actions or links at each point in time. Thus, since our model prescribes that such payoff changes are always perceived (with some noise) as non-negative, the overall adjustment process should display a monotonically increasing trend for the quasi-potential $\tilde{\Phi}$ over time. To illustrate the point, we show in Figure 1 a typical trajectory for small noise.

Then, exploiting further the parallelism with the methodological approach pursued for the GI scenario, we are led to formulating the following conjecture: under LIL, the long-run behavior of the process may be well described (approximated) in the same invariant-distribution format arising in the previous case (c.f. Proposition 2). That is, to be more precise, we posit that the entailed stochastic process induces a limit distribution that is well approximated by the distribution $\tilde{\mu}^{\eta}$ defined, for every $\omega = (\mathbf{s}, G) \in \Omega$ as follows:

$$\tilde{\mu}^{\eta}(\boldsymbol{\omega}) = \frac{e^{\eta\tilde{\Phi}(\boldsymbol{\omega})}}{\sum_{\boldsymbol{\omega}'\in\Omega} e^{\eta\tilde{\Phi}(\boldsymbol{\omega}')}} = \frac{e^{\eta\tilde{\Phi}(\mathbf{s},G)}}{\sum_{G'\in\mathcal{G}^n}\sum_{\mathbf{s}'\in\{-1,+1\}^n} e^{\eta\tilde{\Phi}(\mathbf{s}',G')}}. \tag{17}$$

We provide support for (15) through two different routes. The first one is the most direct. It involves comparing the long-run average values of key variables obtained from extensive numerical simulations with the corresponding mean values derived from the distribution $\tilde{\mu}(\cdot)$ given in (17). As an illustration, in Figure 13 in Appendix B.2, we describe results that reproduce, for the LIL scenario, the very good match between theory and simulations that are obtained for GI context (recall Figure 12). Specifically, they show that, as the linking cost ζ changes over a wide range, the mean degree and mean action predicted by $\tilde{\mu}(\cdot)$ trace closely the average values of these variables that result from the simulations.

The second route explored in support of our theory for the LIL scenario is more indirect, but is complementary to the first one in an interesting way. For, by crucially involving our econometric approach, it is also reassuring for our empirical strategy (not only the LIL scenario

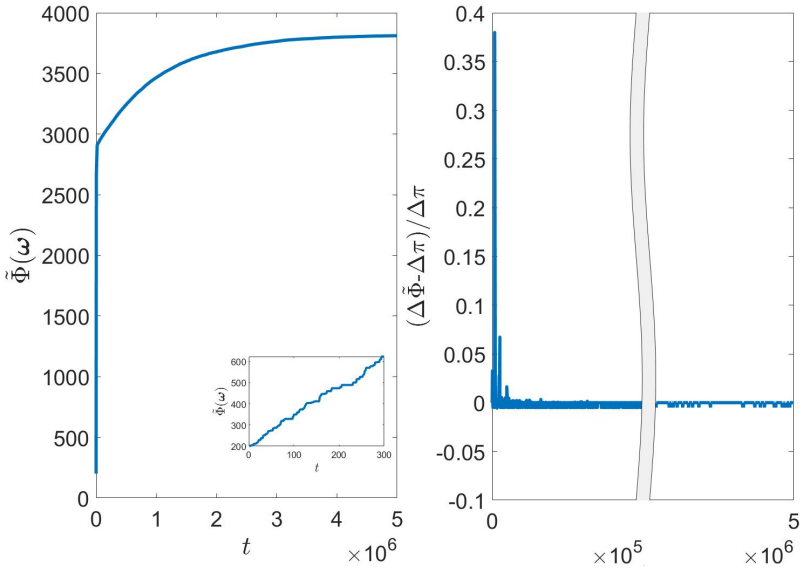


Figure 1: The left panel shows a typical trajectory of the quasi-potential, $\tilde{\Phi}(\omega)$, under the dynamics outlined in Section 2.2. The quasi-potential is increasing and reaches a plateau when the process converges to the stationary state, with the inset showing the initial transition. The right panel shows the difference between changes in the quasi-potential when a link changes ($\Delta\tilde{\Phi} = \tilde{\Phi}(\mathbf{s}, G \pm ij, \boldsymbol{\psi}^{LIL}(\mathbf{s}, G \pm ij)) - \tilde{\Phi}(\mathbf{s}, G, \boldsymbol{\psi}^{LIL}(\mathbf{s}, G))$) and marginal payoffs ($\Delta\pi = \pi_i(\mathbf{s}, G \pm ij, \psi_i^{LIL}(\mathbf{s}, G)) - \pi_i(\mathbf{s}, G, \psi_i^{LIL}(\mathbf{s}, G))$) divided by marginal payoffs, or when an action is adjusted (comparing changes in the quasi-potential, $\Delta\tilde{\Phi} = \tilde{\Phi}(s_i', \mathbf{s}_{-i}, G, \boldsymbol{\psi}^{LIL}(s_i', \mathbf{s}_{-i}, G)) - \tilde{\Phi}(\mathbf{s}, G, \boldsymbol{\psi}^{LIL}(\mathbf{s}, G))$, with changes in payoffs, $\Delta\pi = \pi_i(s_i', \mathbf{s}_{-i}, G, \psi_i^{LIL}(\mathbf{s}, G)) - \pi_i(\mathbf{s}, G, \psi_i^{LIL}(\mathbf{s}, G))$), over the time evolution of the stochastic process. We observe that the relative differences of the changes in the potential and the changes in payoffs are small and get smaller the closer the process is to the stationary state.

but for the GI one as well). This analysis is explained in detail in Section 5.4 once the econometric approach has been presented and applied to our data in Subsections 5.2 and 5.3. Here, we only provide a brief advance of its main conclusions.

The analysis starts with the generation of artificial data obtained by simulating the dynamics of our process according to the rules posited in our theoretical model of Section 2. And, as indicated, we do this for both informational scenarios, i.e. either assuming that the agents have global or local information. Then, we apply our composite-likelihood econometric approach to such artificially generated data and obtain (highly significant) estimates for all of the parameters involved in the model. Finally, we contrast the estimated parameters with the ones actually used to generate the data and find that they are very close in both scenarios. This suggests two related points. First, it suggests that the composite-likelihood approach underlying the econometric analysis is an effective estimation procedure, independently of the informational assumption we make (of course, as long as we choose the "right" informational scenario under which the dat have been generated). Second, it provides further support to the claim that, not only the (strict) potential approach used for the GI scenario, but also the one based on the quasi-potential used for the LIL context, capture the essential features of the long-run behavior of the system. For, only if this is indeed the case can one reasonably justify that they represent useful bases to conduct the econometric estimation in both scenarios.

3.2. Achieving Collective Action

In this subsection we compare the two informational scenarios, GI and LIL, from a perspective that speaks directly to what we have labeled "The Tullock paradox". How can we understand that collective action (e.g. a massive social protest) does sometime arise, even when the population is very large and, therefore, the coordination problem they face is hard and risky? As we have advanced, our contribution to understanding this issue involves understanding/modeling the context as one where individuals' information of the current situation is not global but local, and such information is updated through a process of social learning mediated through the network.

More concretely, our analysis to the problem approaches it from two different, but complementary, angles. In one of these we ask how likely it is that, if the noise is small, collective action may materialize in the long run within each informational scenario. In contrast, the alternative viewpoint focuses instead on comparing not just the long-run predictions in both scenarios but asks "how long is the long run" in each case (and therefore how relevant it is). As we shall show in this section, for each of these alternative perspectives on the problem, the LIL-based framework provides a substantially wider basis than that of the GI assumption for understanding the rise collective action, in particular for large populations.

To study the problem, we pursue a methodology, commonly used in evolutionary theory (see e.g. the seminal work of [cf. e.g. Kandori et al., 1993; Young, 1993]), that starts by identifying the *support* of the long-run distribution of the process when the noise level becomes *vanishing small*– in our context, this amounts to studying the limit $\eta \to \infty$. Under GI, whose invariant distribution μ^η is given by (2) that support includes the states $\omega = (\mathbf{s}, G) \in \Omega$ such that $\lim_{\eta\to\infty} \mu^\eta(\mathbf{s}, G) > 0$. Usually, these are called the *Stochastically Stable States* (SSS) of the process. On the other hand, for the LIL scenario, we carry out a similar exercise on the counterpart distribution $\tilde{\mu}^\eta$ given by (17). However, since in this case this distribution is based on a quasi-potential, the states $\omega = (\mathbf{s}, G) \in \Omega$ such that $\lim_{\eta\to\infty} \tilde{\mu}^\eta(\mathbf{s}, G) > 0$ are called the *Stochastically Quasi-stable States* (SQS).

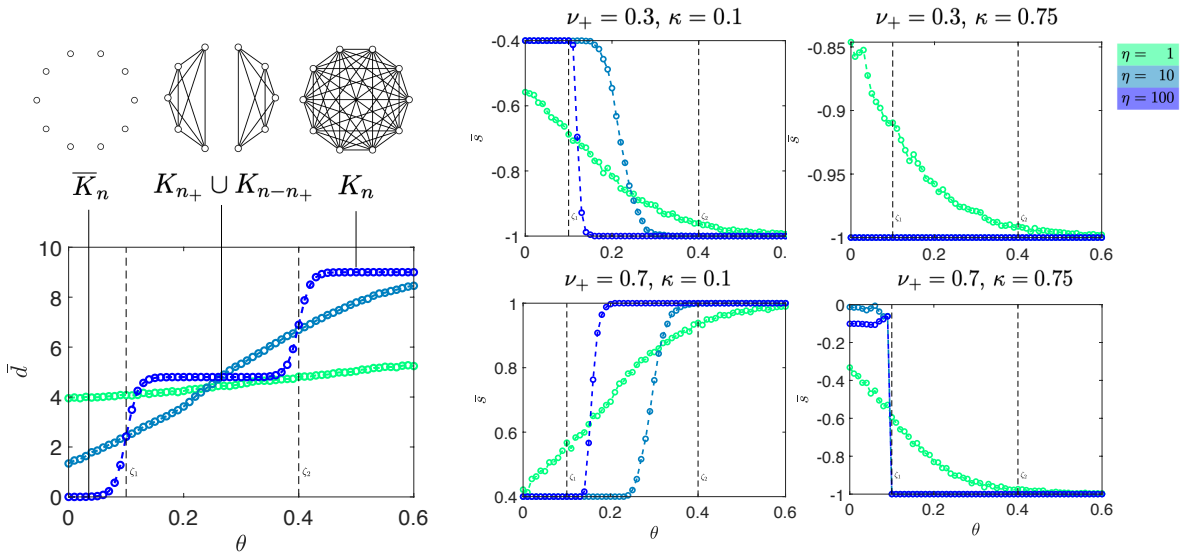


Figure 2: Simulation results for the ***GI scenario*** on the average degree of the network and the average action $\bar{s}$ (right panel) for varying values of θ and η, using the "next reaction method" for simulating a continuous time Markov chain [cf. Anderson, 2012; Gibson and Bruck, 2000]. The circles 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_1 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).

In our context, where the long-run distributions $\mu^\eta(\cdot)$ and $\tilde{\mu}^\eta(\cdot)$ have an exponential form, it is clear that the SSS and SQS are those that maximize the corresponding potential $\Phi(\cdot)$ and quasi-potential $\tilde{\Phi}(\cdot)$, respectively. That is, for the GI scenario, and writing $\Phi(\boldsymbol{\omega})$ for $\Phi(\boldsymbol{\omega}, \boldsymbol{\psi}^{GI}(\boldsymbol{\omega}))$ to simplify the notation, we have:[12]

$$\lim_{\eta\to\infty} \mu^\eta(\mathbf{s}, G) \begin{cases} > 0, & \text{if } \Phi(\mathbf{s}, G) \geq \Phi(\mathbf{s}', G'), \quad \forall \mathbf{s}' \in \{-1, +1\}^n, \quad G' \in \mathcal{G}^n, \\ = 0, & \text{otherwise} \end{cases} \tag{18}$$

and the counterpart conditions for the LIL scenario in terms of $\lim_{\eta\to\infty} \tilde{\mu}^\eta(\cdot)$ and the quasi-potential $\tilde{\Phi}(\cdot)$.

The SSS and SQS are completely characterized for all parameter configurations in, respectively, Propositions 11 and 12 in Appendix C. Figures 2 and 3 illustrate graphically how the long-run distribution values for the average degree $\bar{d}$ and average action $\bar{s}$ changes as η grows, i.e. as the noise becomes progressively smaller. We observe that, despite involving a small population, the displayed functions approach a step-like form as η grows. As we shall see below (cf. Propositions 3-6), this reflects the pattern of networking and behavior that is predicted by our theory, in both scenarios, when the population is large.

For the sake of focus, rather that explaining in detail the aforementioned general results, here in the main text we restrict our discussion to the most interesting context where:

[12]Note that the potential function given in (**??**) satisfies the "single crossing differences" property introduced in Arkolakis and Eckert [2017], and this property is sufficient to guarantee that an iterative best response algorithm can find the global maximum of the potential.

(a) the linking costs are not "prohibitive," i.e. $\zeta_1 < \theta$;
(b) the population is large, i.e. we make $n \to \infty$.

Note that if (a) does not hold, then linking costs are so large that they deter agents from forming links. This, in effect, brings us back to the classical setup in the study of collective action, where the problem is *not* embedded in a social network and hence renders our approach essentially redundant. On the other hand, the reason for focusing on the limit context given by condition (b) is that it directs attention to a context where, being the population large, the collective-action problem is truly challenging. This is indeed the context where the Tullock Paradox explained in the Introduction is most relevant. Formally, the way in which we capture these large-population considerations is to formulate our stability notions on the the limit distribution obtained when the population grows unboundedly. More precisely, in the GI scenario we shall focus on the set of what we call the Limit Stochastically Stable States (LSSS), defined as follows:

$$\Omega^* = \{\omega = (\mathbf{s}, G) : \lim_{n\to\infty} \lim_{\eta\to\infty} \mu^\eta(\omega) > 0\}, \tag{19}$$

while for the LIL scenario we shall consider what we label the Limit Stochastically Quasi-stable States (LQSS) in the following set:

$$\Omega^{**} = \{\omega = (\mathbf{s}, G) : \lim_{n\to\infty} \lim_{\eta\to\infty} \tilde{\mu}^\eta(\omega) > 0\}. \tag{20}$$

As it turns out, the characterizations of the former two sets depend on what is the fraction of individuals of type $+1$ in the population, which we denote by $\nu_+ \equiv n_+/n$. Specifically, we find that they are qualitatively different if the limit fraction ν_+ is lower or higher than $1/2$. It is helpful, therefore, to state separate results for each of these two cases. Considering first the GI scenario, the set of LSSS in the set Ω^* are as stated in the following two propositions.

Proposition 3. *Assume all agents form their beliefs as prescribed by the function $\boldsymbol{\psi}^{GI}(\cdot)$ defined by (6) and suppose $\theta > \zeta_1$. Let $n \to \infty$ and assume $1/2 > \lim_{n\to\infty} \nu_+ > 0$. Then we have:*

*(a) If $\theta < \zeta_2$ the unique LSSS has the network segmented into **two cliques**,[13] K_{n_+} and K_{n-n_+}, with no cross-links and including all agents of types $+1$ and -1, respectively. The action profile in the LSSS has all agents $i \in \mathcal{N}$ choosing the **action** $\boldsymbol{s_i = -1}$.*

*(b) If $\theta > \zeta_2$ the unique LSSS is given by the **complete graph** K_n with all agents choosing the **action** $\boldsymbol{s_i = -1}$.*

Proposition 4. *Assume all agents form their beliefs as prescribed by the function $\boldsymbol{\psi}^{GI}(\cdot)$ defined by (6) and suppose $\theta > \zeta_1$. Let $n \to \infty$ and assume $1 > \lim_{n\to\infty} \nu_+ > 1/2$. Then we have:*

1. Suppose $\kappa > 2\nu_+ - 1$.

*(a) If $\theta < \zeta_2$ the unique LSSS has the network segmented into **two cliques**, K_{n_+} and K_{n-n_+}, with no cross-links and including all agents of types $+1$ and -1, respectively. The action profile in the LSSS has all agents $i \in \mathcal{N}$ choosing the **action** $\boldsymbol{s_i = -1}$.*

*(b) If $\theta > \zeta_2$ the unique LSSS is given by the **complete network** K_n with all agents choosing the **action** $\boldsymbol{s_i = -1}$.*

2. Suppose $\kappa < 2\nu_+ - 1$.

*(a) If $\theta < \zeta_2$ the unique LSSS has the network segmented into **two cliques**, K_{n_+} and K_{n-n_+}, with no cross-links and including all agents of types $+1$ and -1, respectively.*

[13] Recall that a clique is a completely connected subnetwork. In general, we use the notation K_m to denote a clique of size m, whereas $\overline{K}_m$ denotes an empty network with m nodes

The action profile has all agents $i \in \mathcal{N}$ choosing the ***action*** $\boldsymbol{s_i = +1}$.

(b) *If $\theta > \zeta_2$ the unique LSSS is given by the* ***complete network*** K_n *with all agents choosing the* ***action*** $\boldsymbol{s_i = +1}$.

Propositions 3 and 4 completely characterize the SSS for arbitrarily large n when the network is not empty (i.e. under the assumption that $\theta > \zeta_1$). The contrast between these two results is intuitive and instructive. A key consideration in both cases is whether or not $\theta > \zeta_2$. If this inequality holds, all links are profitable between any two agents if, independently of their type, they are action-coordinated. Naturally, this always leads to a LSSS with a complete network.[14] The differences in this case, therefore, can only pertain to the action profile associated to such a complete network. If the number of $+1$ types is less than half (as in Proposition 3), there is no "critical mass" for the costly action $+1$ and *every* player ends up choosing -1, irrespectively of the cost $\kappa \geq 0$ of action $+1$. (Note that, since the network is complete, there is no way to support any extent of action diversity in the population.) Instead, if such a critical mass exists (i.e. the number of $+1$ types is more than half – and still we have $\theta > \zeta_2$) – Proposition 4 establishes that whether action -1 or action $+1$ is chosen uniformly by the population depends on the cost κ of action $+1$. If high enough (i.e. $\kappa > 2\nu_+ - 1$), then the costless action -1 is selected in the vanishing-noise limit; otherwise it is the action $+1$.

A quite different situation arises if $\theta < \zeta_2$. In this case, no link between two agents is profitable whenever they are also of different types, even if they are choosing the same action. Then, the two-clique arrangement in type-homogeneous cliques is the most robust one, and therefore it is the configuration that prevails in *every* LSSS. Concerning actions, on the other hand, the population profile displayed at the LSSS depends on three parameters: the cost κ of action $+1$, the fraction of agents who display the corresponding type $+1$, and the strength of the global conformity parameter ρ. A *sufficient condition* to have all n_+ agents choose the action $+1$ is that $\kappa < 1 - \rho(1 - \nu_+)$. This condition, however, is not necessary. Specifically, if $\rho > 2$, it can be readily checked that there are $\kappa > 1 - \rho(1 - \nu_+)$ such that, if $\nu_+ > 1/2$, not only the agents of type $+1$ choose action $+1$ but also those of type -1.[15] Since no such a possibility exists when $\nu_+ < 1/2$, in this case the inequality $\kappa < 1 - \rho(1 - \nu_+)$ is both necessary and sufficient for action $+1$ to be played as the LSS.

Returning again to Figure 2, we find an illustration of the fact that the pattern of network architecture and action choice predicted by Propositions 3 and 4 matches the pattern displayed by a numerical simulation of the process for a finite population. More specifically, we observe that, as θ varies across the three regions marked by the linking costs ζ_1 and ζ_2, both the average action in the population and the network architecture (either complete or segmented into two cliques) are well approximated (not just qualitatively but also quantitatively) by the theoretical predictions of the model, if the noise is small enough ($\eta = 100$).

Next, we turn to the counterpart of the the previous results for the LIL scenario that in this case characterize the set Ω^{**} of LQSS. Again, we find it useful to separate the conclusions obtained when the limit fraction ν_+ of agents of type $+1$ is either lower or higher than $1/2$. type $+1$.

[14] In principle, one could conceive a "relatively" stable configuration where the population is partitioned into type-homogeneous components. These components, however, are not in this case robust against the creation of cross links, which leads to a single component and eventually action uniformity and link completeness.

[15] Note that $\rho > 2$ implies that $1 - \rho(1 - \nu_+) < 2\nu_+ - 1$ and therefore we can have κ that satisfies both $\kappa > 1 - \rho(1 - \nu_+)$ and $\kappa < 2\nu_+ - 1$.

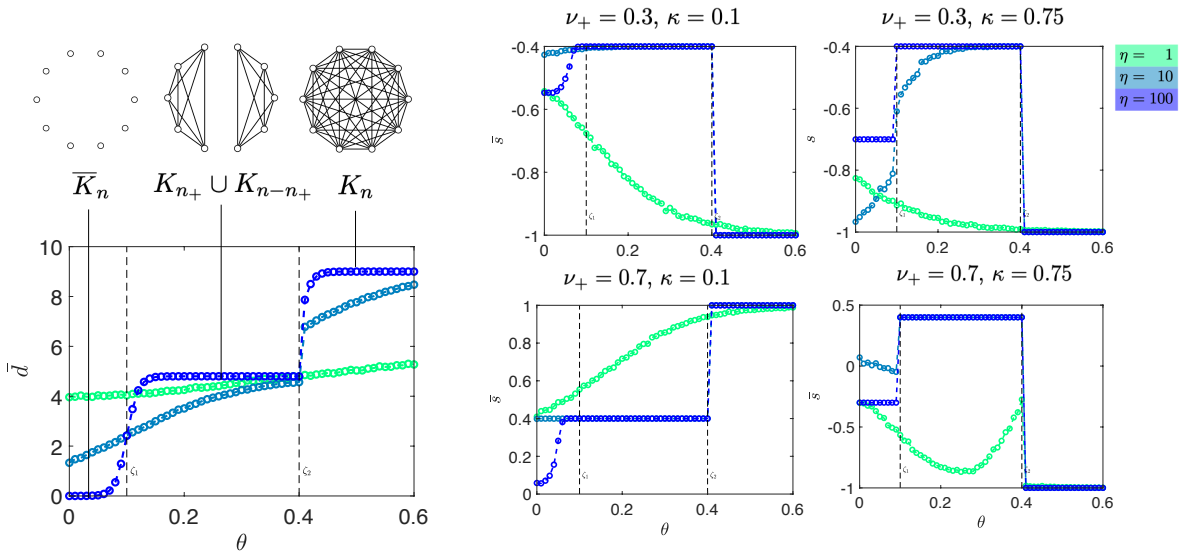


Figure 3: Simulation results for the ***LIL scenario*** on the average degree of the network and the average action $\bar{s}$ (right panel) for varying values of θ and η, using the "next reaction method" for simulating a continuous time Markov chain [cf. Anderson, 2012; Gibson and Bruck, 2000]. The circles 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 stochastically stable states in the limit of large η correspond to Propositions 5 and 6, respectively. The thresholds ζ_1 and ζ_1 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).

Proposition 5. *Assume all agents form their beliefs as prescribed by the function $\psi^{LIL}(\cdot)$ given by (10) and suppose $\theta > \zeta_1$. Let $n \to \infty$ and assume $1/2 > \lim_{n\to\infty} \nu_+ > 0$. Then we have:*

1. *If $\theta < \zeta_2$ the unique LSSS has the network segmented into **two cliques**, K_{n_+} and K_{n-n_+}, with no cross-links and including all agents of types $+1$ and -1, respectively. The action profile in the LSSS has all agents choosing the action $\boldsymbol{s_i = \gamma_i}$ if $\kappa < 1$ while if $\kappa > 1$ all agents in both cliques choose the **action** $\boldsymbol{s_i = -1}$*
2. *If $\theta > \zeta_2$ then the unique LSSS is given by the **complete network** K_n with all agents choosing the **action** $\boldsymbol{s_i = -1}$.*

Proposition 6. *Assume all agents form their beliefs as prescribed by the function $\psi^{LIL}(\cdot)$ given by (10) and suppose $\theta > \zeta_1$. Let $n \to \infty$ and assume $1 > \lim_{n\to\infty} \nu_+ > 1/2$. Then we have:*

1. *Suppose $\kappa > 2\nu_+ - 1$.*
   (a) *If $\theta < \zeta_2$ the unique LSSS has the network segmented into **two cliques**, K_{n_+} and K_{n-n_+}, with no cross-links and including all agents of types $+1$ and -1, respectively. The action profile in the LSSS has all agents choosing the action $\boldsymbol{s_i = \gamma_i}$ if $\kappa < 1$ while if $\kappa > 1$ all agents in both cliques choose the **action** $\boldsymbol{s_i = -1}$.*
   (b) *If $\theta > \zeta_2$ then the unique LSSS is given by the **complete network** K_n with all agents choosing the **action** $\boldsymbol{s_i = -1}$.*
2. *Suppose $\kappa < 2\nu_+ - 1$.*
   (a) *If $\theta < \zeta_2$ the unique LSSS has the network segmented into **two cliques**, K_{n_+} and K_{n-n_+}, with no cross-links and including all agents of types $+1$ and -1, respectively. The action profile in the LSSS has all agents choosing the action $\boldsymbol{s_i = \gamma_i}$.*
   (b) *If $\theta > \zeta_2$ then the unique LSSS is given by the **complete network** K_n with all agents*

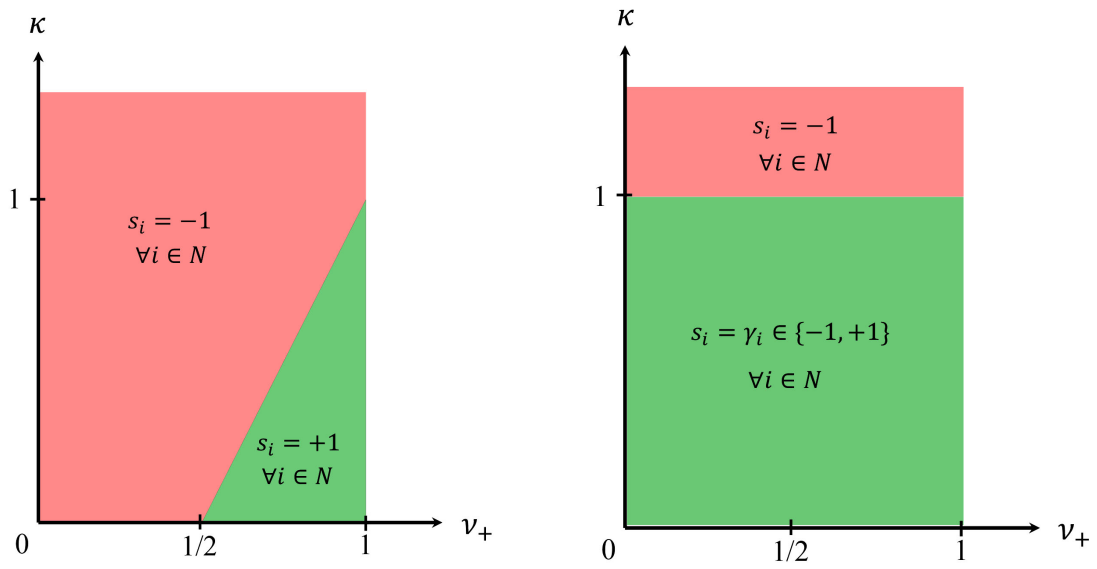


Figure 4: Comparing the two LSSS predictions with $\zeta_1 < \theta < \zeta_2$ for the GI model (left panel) and the LIL model (right panel).

choosing the ***action*** $\boldsymbol{s_i = +1}$.

In analogy with what was suggested for the GI scenario and Figure 2, Figure 3 illustrates that the simulation results for finite populations and small noise are well aligned with the theoretical predictions for the LIL scenario stated in Propositions 5-6. In view of this, an important question to address is how, in general, these latter predictions contrast with those obtained for the GI scenario. To address it, a useful route to take is to compare the parameter regions where, in each model, the corresponding condition of stochastic (quasi-)stability leads to states where there is a sizable level of collective action. Or, to be more specific, let us focus on the regions where the fraction of the population contributing to collective action includes at least the agents more inclined to such a contribution, i.e., those of type +1. Then we find that while the condition $\kappa < 1$ is always necessary in both scenarios (i.e. action +1 cannot be too costly), under LIL it is also a sufficient condition. Instead, in order to obtain the desired extent of collective action under GI we need two additional conditions. 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. essentially all individuals in the population are of type +1. For a graphical description of the situation, the reader can refer to Figure 4.

The previous discussion highlights that the model where agents are assumed to have limited information and learn from their current network neighbors provides a significantly wider range of circumstances where one may expect that large-scale collective action can materialize. This can be regarded as a useful step towards addressing the Tullock Paradox, which highlights the difficulty for revolutions (a form of collective action) to arise if the population involved is large. As advanced, we may also approach the problem in a (complementary) way by shifting the focus from a comparison centered on the long-run predictions captured by the limit distribution of the process to another one that compares how long does it take to reach/approximate such a limit behavior. Even though a formal analysis of this question is beyond the scope of this

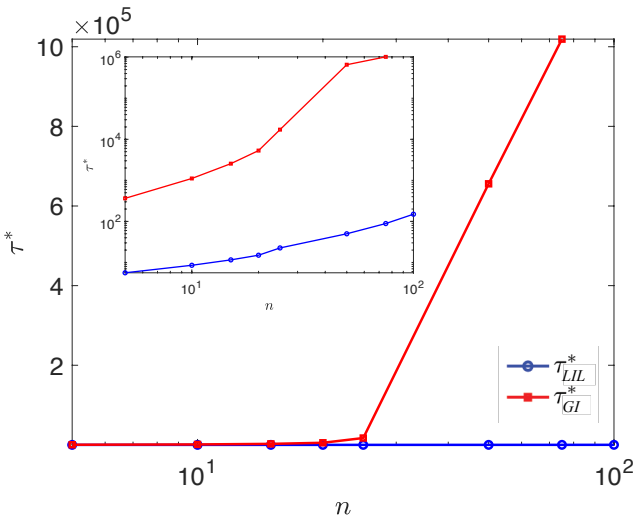


Figure 5: The convergence times, τ^*, to a configuration with at least half of the population choosing action $+1$ for different population sizes (n) starting from an empty network, $\overline{K}_n$, where all agents choose action $s_i = -1$ for the GI and the LIL models. The circles represent averages obtained across 1000 Monte Carlo runs under the following parameters: $n = 5, \ldots, 100$, $\eta = 1$, $\lambda = \chi = \xi = 1$, $\kappa = 0.1$, $\varphi = 0.5$, $\theta = 0.02$, $\zeta_1 = 4$, $\zeta_2 = 8$ and $\rho = 3/n$. The inset shows the same figure but with the y-axis is in log-scale.

paper, in what follows we provide some illustrative simulations that convey the main gist of the idea.

These simulations were conducted for parameter configurations that, both in the GI and LIL frameworks, deliver long-run predictions – as given by Propositions 4 and 6 – that embody substantial collective action (at least 50% of the population chooses action $+1$). Then, from initial conditions at which both the network as well as collective action start from *tabula rasa* – i.e. the social network is empty and everyone in the population chooses action -1 – we ask the following question: What is the expected time required to reach such a substantial level of collective action? How does this expected time change as the population size grows?

Figure 5 describes simulation results that answer the former questions for a typical configuration of parameters that meet the conditions described above. (Specifically, they satisfy that $\zeta_1 < \theta < \zeta_2$, $\nu_+ > 1/2$ and $\kappa < 2\nu_+ - 1$.) It traces how the average delay – measured by the number of individual adjustments involved – depends on the population size n for both the GI and LIL scenarios. The corresponding plots display a sharp contrast between these alternative contexts. Specifically, they show that the average delay in the former case is much higher that in the latter, even if the population size n is relatively small (for $n = 100$, the difference spans more that 5 orders of magnitude). This indicates that, still when the GI scenario also yields the long-run prediction of a substantial level for collective action, actually reaching this level takes much more "time" (adjustment instances) than in the LIL context. This is yet an additional reason why the assumption of local information and local learning appears as a better assumption to understand how collective action arises in large-population contexts. For it is not only that the LIL scenario supports collective action under a significantly wider set of circumstances (i.e. parameters), and through a more realistic assumption (agents are supposed to have an imperfect grasp of the overall state of a large system). It also provides a much shorter time scale (hence a more plausible dynamic basis) for how the population reaches that situation.

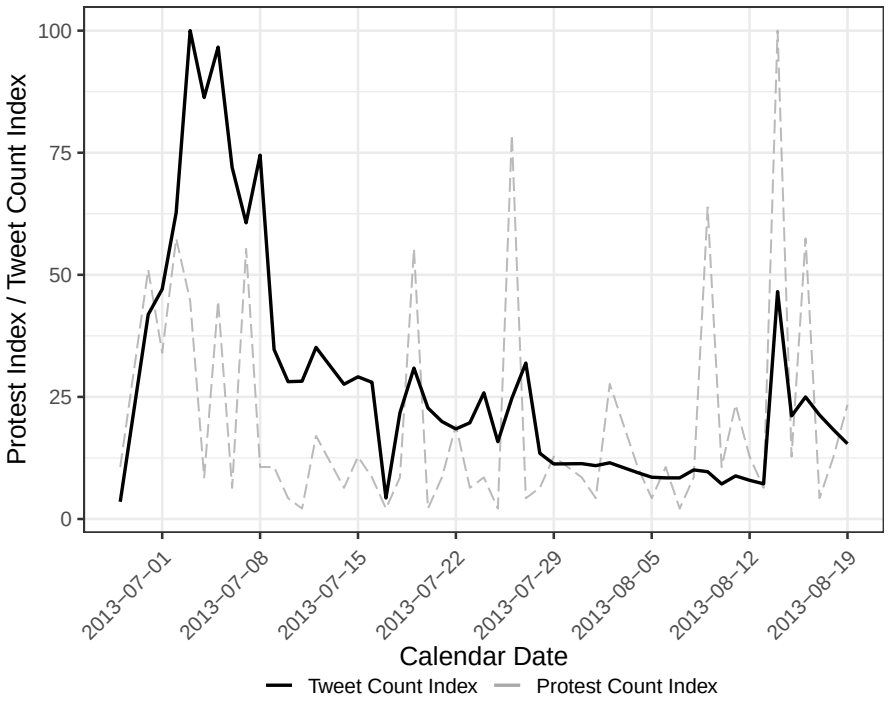


Figure 6: Twitter volume and protests.

4. Data

The empirical application of our model uses online social network data from the riots and demonstrations in Egypt during the Arab Spring in 2013 [cf. Borge-Holthoefer et al., 2015; Magdy et al., 2016]. Twitter can be regarded as one of the main social social networking platform for opinion exchange over Egypt's Arab Spring [cf. e.g. Clarke and Kocak, 2020]. Our empirical exercise uses protest related tweets during the fall of the incumbent Islamist president, Mohamed Morsi, and the return of military rule as a case study.[16] We collected around 6 million Arabic language tweets from approximately 700k Twitter user accounts based on data from Magdy et al. [2016] and Borge-Holthoefer et al. [2015]. Figure 6 provides an illustration of the concurrence of Twitter message volumes and protests.[17]

The original dataset have around 700k users (accounts), which are too large for estimating a structural network formation model. We therefore restrict our analysis to a subsample over the time period between July 4, 2013 when the military overthrew the regime of president Morsi, and August 19, 2013.[18] We drop users who had less than two tweets in the sample period, remove a user if she has a follower count above the 95^{th} percentile (likely to be media accounts or public figures) and drop users who have more than 78 connections (in the 99^{th} percentile) to exclude potential accounts of social bots. The resulting sample consists of 225,578 users.

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

[17] Information on protests has been gathered from the Armed Conflict Location and Event Data Project; https://acleddata.com) over the sampling period.

[18] As seen in Figure 6, the correlation of Twitter volume and protests declines after August 19,2013.

We construct the undirected links between users by using retweets and @ mentions. We assume that two users are connected if: User i has either retweeted or @ mentioned user j, and user j has either retweeted or @ mentioned user i. This construction of bidirectional links on Twitter aims at capturing "strong ties" between users necessary for identifying peer effects in individuals' behavior.

4.1. From Text to Quantitative Measurement

From the tweets' texts of a Twitter user we can infer crucial information about that user:

a) Is a user "rioting" (pro- or anti-military intervention)?
b) What is the user's political affiliation (secular vs. islamist)?
c) What is the user's gender (male vs. female)?

Each of these are binary and this can be analyzed using binary classifiers. What someone writes tells us if they are pro- or anti-military intervention (a). What and how someone writes is predictive of demographics (b and c).

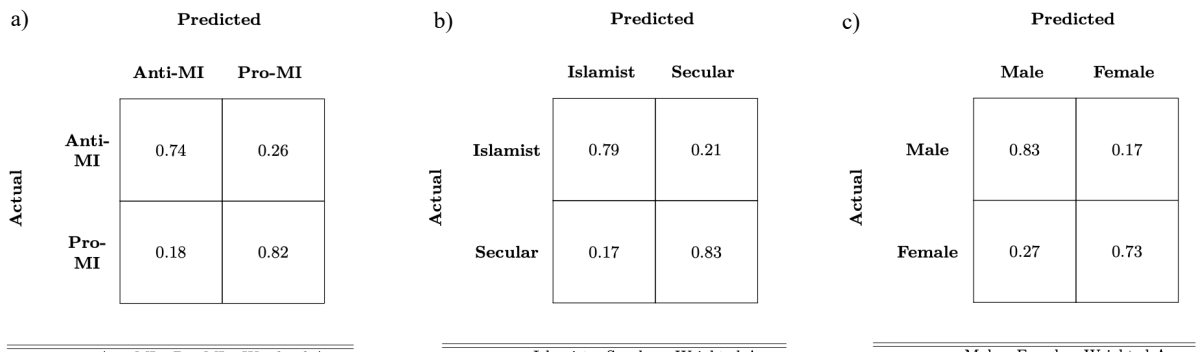
For this purpose of classification we take advantage of recent developments in Natural Language Processing (NLP) tools, so called "BERT" models, and in particular, a specifically pretrained model for Arabic language, the "AraBERT" model [Antoun et al., 2020]. Such pretrained language models can be seen as an algorithm which understands language and can do tasks in that language like a machine equivalent of a well read "human". The way this is achieved is by letting the algorithm "read" a large amount of unannotated data. In the case of AraBERT this consists of over 200 million lines of Arabic text from news sites and Wikipedia. The algorithm then "learns" the usage of words and how the language is written from a deep neural network trained model that has over 110 million parameters. We then fine-tune AraBERT's language "understanding" from a generalized context to task specific prediction of action classification or users' characteristics. Before applying our prediction algorithms we clean the text using the "Farasa segmenter" [Abdelali et al., 2016], which is a state of the art way to deal with stemming, dealing with prefixes, suffixes, etc.. This is important for BERT models when language is morphologically rich such as Arabic. To achieve a high predictive ability we then train (estimate) a language model on 80% of the data and test its predictive power on the unseen 20%.

4.2. Classifying Tweets: Pro- vs. anti military intervention

We want to identify which users are rioting. However, tweets don't explicitly tell us a user's stance. We therefore build a classifier that identifies the pro- vs- anti-military intervention disposition of a user from the user's tweets texts, and take this as an indication for the willingness of an individual to riot or protest. We follow the convention that action $+1$ identifies anti-military intervention and -1 as pro-military intervention.

As training data we use 4,150 hand classified tweets into pro- or anti-military intervention by two human coders proficient in Arabic. To check consistency 1,000 tweets are the same for both coders. We obtained as a consistency measure Cohen's $\kappa = 0.67$. This indicates a sufficiently strong agreement between the two coders (whereby agreement due to chance is factored out).

We then build a binary classifier of tweets (pro- or anti-military intervention) using fine-tuned AraBERT model. The out-of-sample performance is shown in the bottom left panel



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

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

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

Figure 7: The top panels show the classifier confusion matrices for 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).

in Figure 7. 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 7. 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. The classifier is then applied to all tweets in estimation sample.

4.3. User Characteristics: Political affiliation

Egyptian society is politically polarized between secularist and islamists since well before the onset of Arab Spring. We need to infer this in our sample since it is likely correlated with action (i.e rioting behavior). Our goal is to predict "political affiliation" – islamist vs. secular – from the tweets.

While riot related tweets are useful for inferring political affiliation, we need some pre-classified data. Here we can rely on an already annotated sample of 20,886 Egyptian Twitter users' and their political affiliations provided by Weber et al. [2013]. We identified the tweets of the users in Weber et al. [2013] in our own data sample to obtain a mapping between tweet content to political affiliation. We split users 80/20 into training/test data. We implement a binary classifier of tweets using another fine-tuned AraBERT model. We predict a user's political affiliation by a "majority vote" over their tweets where, for example, three secular tweets and five islamist tweets classifies a user to be islamist.

The out-of-sample performance of our classifier is shown in the bottom middle panel in Figure 7. The high F1-score and accuracy, together with the corresponding confusion matrix shown in the top middle panel in Figure 7 indicate that the classifier can well identify political affiliation from the tweets' text.

Figure 8 shows the fraction of islamist versus secular users over the sample period. Note that in our sample the majority of agents is islamists, and that is why the action +1, which is protesting again the military intervention, dominates (around 73%).

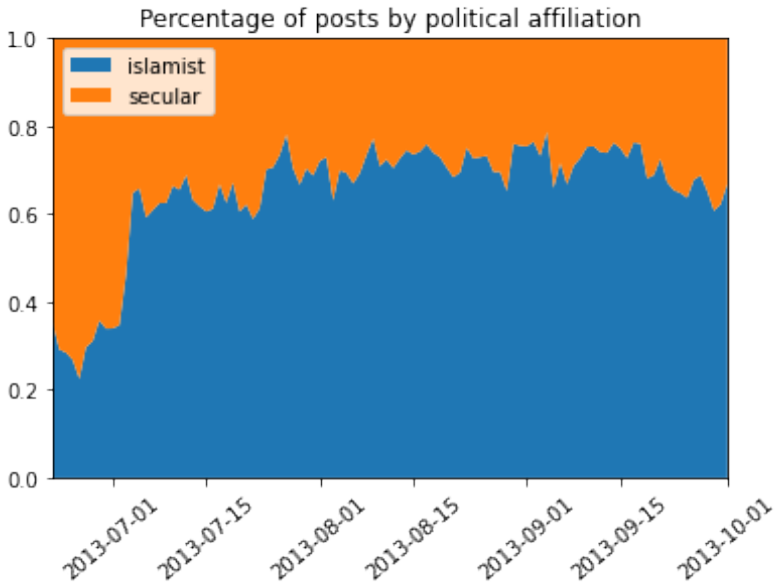


Figure 8: The fraction of islamist versus secular users over the sample period.

Figure 9 shows the users' political affiliations and links between them based on a "forest-fire" sample [Kolaczyk, 2009] of the original network of users. 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. This indicates that users mostly communicate with other users of the same political affiliation.

4.4. User Characteristics: Gender

We also want to predict gender from the tweets. However, tweets about riots might not be sufficiently informative about a user's gender. We therefore collected a wider variety of tweets from the users. We crawled the most recent tweets from each of the users in our data sample from the public Twitter API (as effective of Dec. 2020). We could obtain up to 3,200 tweets per user (Twitter API maximum) which yielded over 600 GB of tweets.

As training/test data we relied again on Weber et al. [2013]. These authors provide a classification of Egyptian Twitter users by gender. We split their users 80/20 into training/test data. We then build a binary classifier of tweets using another fine-tuned AraBERT model. We predict a user's gender by a majority vote over their tweets such that, for example, three "male" tweets and five "female" tweets classify a user as female. The out-of-sample performance of our gender classifier is shown in the bottom right panel in Figure 7. The high F1-score, accuracy and confusion matrix (shown in the top right panel in Figure 7) indicate that the classifier can well identify gender from the tweets' text.

Table 1 provides basic summary statistics of our data.

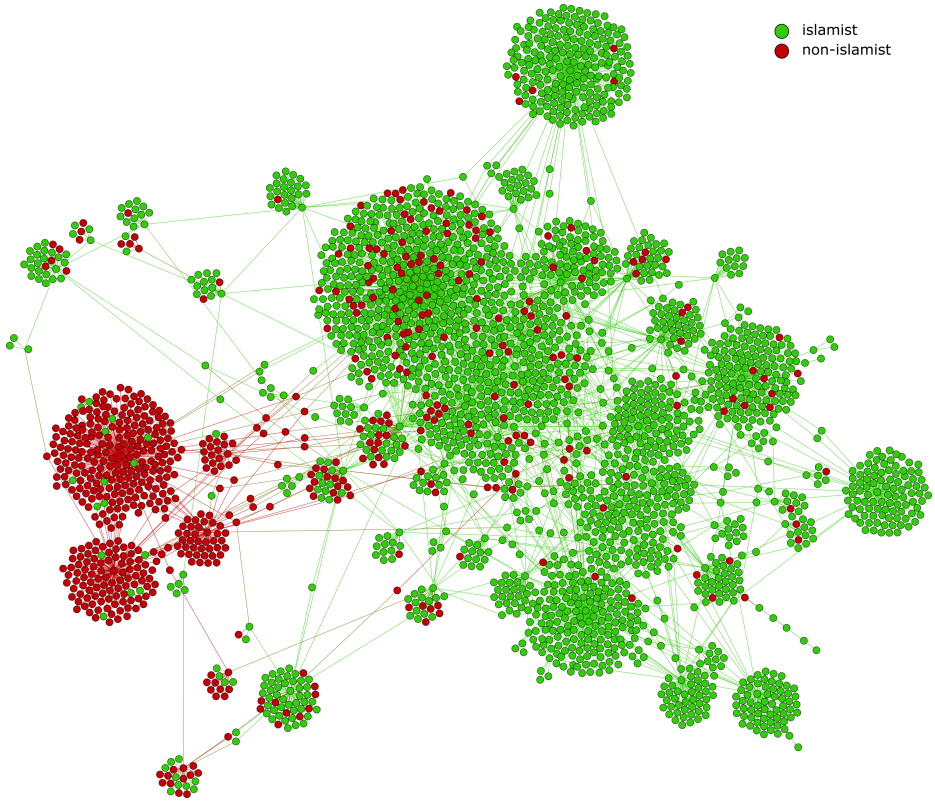


Figure 9: Users' political affiliations and links based on a "forest-fire" sample [Kolaczyk, 2009] of the original network of users.

5. Empirical Analysis

5.1. Parametric Specification and Identification

The stationary distributions corresponding to the GI and LIL scenarios in (15) and (17) are known as the Gibbs measure (or the Gibbs random field), which provides us a probability (likelihood) measure for estimating unknown structural parameters in the (quasi) potential functions of (11) and (16) with the empirical data described in Section 4. Before proceeding with the details of 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 (16) as functions of observed and unobserved individual characteristics,

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

and

$$\zeta_{ij} = \zeta_0 + \sum_{k=1}^{K} h_k(x_{ik}, x_{jk})\zeta_k + z_i + z_j, \tag{21}$$

where $\mathbf{x}_i$ is a $(K+1) \times 1$ vector of exogenous regressors including individual gender, political affiliation, number of Twitter followers, and a constant, and $\boldsymbol{\beta}$ is the corresponding vector of coefficients. Variable z_i represents unobserved individual heterogeneity (random effect) and is assumed following i.i.d. $(0, \sigma_z^2)$. The function $h_k(x_{ik}, x_{jk})$ in ζ_{ij} can be either an indicator function $(\mathbf{1}(x_{ik} = x_{jk}))$ when x_{ik} is a dummy variable or a distance function $(|x_{ik} - x_{jk}|)$ when x_{ik} is continuous, reflecting homophily. The random effects z_i and z_j in ζ_{ij} are used to capture

Table 1: Summary Statistics

	max	min	mean	s.d.
Action	1	-1	0.4848	0.8746
Female	1	0	0.0974	0.2965
Islamist	1	0	0.6157	0.4864
log number of followers	9.1722	0	5.3169	1.5569
Number of links	47	0	0.8313	2.1345
Sample size		225,578		

degree heterogeneity due to unobservables [Dzemski, 2019; Graham, 2017; Jochmans, 2018].

There are two identification issues regarding our model specification. First, the scale parameter η of the logistic disturbance ε in (15) and (17) is not separately identifiable from other parameters in the (quasi) potential function, a common problem in discrete choice models. Therefore, we fix η to one during estimation. Second, in the GI scenario of (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 only has a negligible variation at the individual level and therefore the coefficient ρ is hardly disentangled from both the constant term β_0 in γ_i and 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 β_0 and κ for the GI scenario. This modification allows us to gain the identification of the GI scenario at the cost of muffling true global conformity effect in (11). To mark this modification, we denote the coefficient of $(n-1)\bar{s}$ in this modified model by $\widetilde{\rho}$. A similar problem also appears in the LIL scenario of (16), where the global conformity effect can be identified through individual variations on the belief p_i, but the rioting cost κ and the constant term β_0 in γ_i are still entangled. Therefore, we drop β_0 in (16) and denote $\widetilde{\kappa} = \kappa - \beta_0$ in the LIL scenario.

5.2. Estimation

The main challenge to the estimation of the Gibbs measures in (15) and (17) comes from the appearance of an intractable normalizing constant in the denominator, which prevents the researcher from directly calculating the likelihood for conventional frequentist or Bayesian inference. The most commonly used methods to tackle such estimation challenge include the pseudo likelihood approach [Besag, 1975; Strauss and Ikeda, 1990], the Monte Carlo simulated likelihood approach [Geyer and Thompson, 1992], and the Bayesian exchange algorithm [Møller et al., 2006; Murray et al., 2006] with the exact sampling or the double Metropolis-Hastings algorithm with aid of simulated auxiliary variables [Badev, 2021; Hsieh et al, 2020; Hsieh et al., 2021; Liang, 2010; Mele, 2017]. Given the enormous sample size dealt in this study, the latter two simulation-based methods are not feasible. Therefore, we adopt the pseudo likelihood approach, or to be more generally, the composite likelihood approach [Lindsay, 1988; Varin et al., 2011] to estimate our models.[19] Several theoretical results on the asymptotic consistency of the pseudo (composite) likelihood estimation for the Gibbs measure are available in the literature [Bhattacharya and Mukherjee, 2018; Chatterjee, 2007; Comets, 1992; Ghosal and

[19] In general the composite likelihood can be written as $f(y) = \prod_{i=1}^{C} p(y_{A_i}|y_{B_i})$ where $B_i = A \setminus A_i$. The pseudo likelihood refers to the case where A_i is a singleton.

Mukherjee, 2020],[20] and thus a remaining practical issue is the empirical rate of convergence. The simulation results in van Duijn et al. [2009], Zhou and Schmidler [2009], Hughes et al. [2011], Friel [2012] show that the pseudo likelihood method permits reliable inference when the sample size is not small and the network dependence is moderate.[21] Since the sample size in this work is enormous, and the estimated network effects shown in Section 5.3 are not particularly large, the composite likelihood method is inarguably an adequate estimation method to be used.

The composite conditional likelihoods for our model is defined as

$$\mu_{\mathrm{c}}(\mathbf{s}, G) = \mu(\mathbf{s}|G)\mu(G|\mathbf{s}), \tag{22}$$

where $\mu(\mathbf{s}|G)$ and $\mu(G|\mathbf{s})$ separately represent the conditional probabilities of action and network given the other. We first look at the conditional probability of action profile on a given network. Under the GI scenario,

$$\begin{aligned}\mu(\mathbf{s}|G) &= \frac{\exp\left(\sum_{i=1}^{n}(\mathbf{x}_i^\top\boldsymbol{\beta} + z_i\tau + \widetilde{\rho}(n-1)\bar{s})s_i + \frac{\theta}{2}\sum_{i=1}^{n}\sum_{j\neq i}^{n} a_{ij}s_is_j\right)}{\mathscr{Z}(G)} \\ &= \frac{\exp\left(\mathbf{s}^\top(\mathbf{x}\boldsymbol{\beta} + z\tau + \widetilde{\rho}(n-1)\bar{s}\cdot\ell) + \frac{\theta}{2}\mathbf{s}^\top\mathbf{A}\mathbf{s}\right)}{\mathscr{Z}(G)},\end{aligned} \tag{23}$$

where $\mathbf{x} = (\mathbf{x}_1, \mathbf{x}_2, \cdots, \mathbf{x}_n)^\top$ and

$$\mathscr{Z}(G) = \sum_{\mathbf{s}'\in\{-1,+1\}^n} \exp\left(\mathbf{s}'^\top(\mathbf{x}\boldsymbol{\beta} + z\tau + \widetilde{\rho}(n-1)\bar{s}'\cdot\ell) + \frac{\theta}{2}\mathbf{s}'^\top\mathbf{A}\mathbf{s}'\right)$$

is an intractable normalizing term. Similarly, under the LIL scenario,

$$\begin{aligned}\tilde{\mu}(\mathbf{s}|G,\mathbf{p}) &= \frac{\exp\left(\sum_{i=1}^{n}(\mathbf{x}_i^\top\boldsymbol{\beta} + z_i\tau + \widetilde{\kappa} + \rho(n-1)p_i)s_i + \frac{\theta}{2}\sum_{i=1}^{n}\sum_{j\neq i}^{n} a_{ij}s_is_j\right)}{\mathscr{Z}(G)} \\ &= \frac{\exp\left(\mathbf{s}^\top(\mathbf{x}\boldsymbol{\beta} + z\tau + \widetilde{\kappa} + \rho(n-1)\mathbf{p}) + \frac{\theta}{2}\mathbf{s}^\top\mathbf{A}\mathbf{s}\right)}{\mathscr{Z}(G)},\end{aligned} \tag{24}$$

where $\mathscr{Z}(G) = \sum_{\mathbf{s}'\in\{-1,+1\}^n}\exp\left(\mathbf{s}'^\top(\mathbf{x}\boldsymbol{\beta} + z\tau + \widetilde{\kappa} + \rho(n-1)\mathbf{p}) + \frac{\theta}{2}\mathbf{s}'^\top\mathbf{A}\mathbf{s}'\right)$ is an intractable normalizing term. The conditional probabilities of action profile $\mathbf{s}$ on a given network G in (23) and (24) are related to the Ising model [Ising, 1925]. We apply the pseudo likelihood to approximate these conditional probabilities. Taking the GI scenario in (23) for illustration, the conditional probability of agent i choosing action $s_i = 1$ given the actions of all other agents

[20] A requirement for the consistency is that the adjacency matrix is uniformly bounded.

[21] Hughes et al. [2011] show that the pseudo likelihood approach permits a reliable inference when the sample size is beyond a 30× 30 network adjacency matrix. Zhou and Schmidler [2009] and Friel [2012] show that the pseudo likelihood estimation results are precise under 100×100 and 50×50 network samples, respectively.

$\mathbf{s}_{-i}$ and the network G is

$$\begin{aligned}
&\mu(s_i = 1|\mathbf{s}_{-i}, G) \\
&= \frac{\mu(s_i = 1, \mathbf{s}_{-i}|G)}{\mu(s_i = 1, \mathbf{s}_{-i}|G) + \mu(s_i = -1, \mathbf{s}_{-i}|G)} \\
&= \frac{\exp(\mathbf{x}_i^\top \boldsymbol{\beta} + z_i\tau + \widetilde{\rho}(n-1)\bar{s} + \theta \sum_{j\neq i}^n a_{ij}s_j)}{\exp(\mathbf{x}_i^\top \boldsymbol{\beta} + z_i\tau + \widetilde{\rho}(n-1)\bar{s} + \theta \sum_{j\neq i}^n a_{ij}s_j) + \exp(-\mathbf{x}_i^\top \boldsymbol{\beta} - z_i\tau - \widetilde{\rho}(n-1)\bar{s} - \theta \sum_{j\neq i}^n a_{ij}s_j)} \\
&= \frac{\exp(\mathbf{x}_i^\top \boldsymbol{\beta} + z_i\tau + \widetilde{\rho}(n-1)\bar{s} + \theta \sum_{j\neq i}^n a_{ij}s_j)}{2\cosh(\mathbf{x}_i^\top \boldsymbol{\beta} + z_i\tau + \widetilde{\rho}(n-1)\bar{s} + \theta \sum_{j\neq i}^n a_{ij}s_j)}.
\end{aligned} \tag{25}$$

Similarly, the conditional probability of agent i choosing action $s_i = -1$ is

$$\mu(s_i = -1|\mathbf{s}_{-i}, G) = \frac{\exp(-\mathbf{x}_i^\top \boldsymbol{\beta} - z_i\tau - \widetilde{\rho}(n-1)\bar{s} - \theta \sum_{j\neq i}^n a_{ij}s_j)}{2\cosh(\mathbf{x}_i^\top \boldsymbol{\beta} + z_i\tau + \widetilde{\rho}(n-1)\bar{s} + \theta \sum_{j\neq i}^n a_{ij}s_j)}. \tag{26}$$

Therefore, the pseudo likelihood of action profile $\mathbf{s}$, conditional on network G, is defined as

$$\mu_c(\mathbf{s}|G) = \prod_{i=1}^n \mu(s_i|\mathbf{s}_{-i}, G) = \prod_{i=1}^n \mu(s_i = 1|\mathbf{s}_{-i}, G)^{\frac{1+s_i}{2}} \mu(s_i = -1|\mathbf{s}_{-i}, G)^{\frac{1-s_i}{2}}, \tag{27}$$

which we will use to replace (23) in calculating (22).[22]

We next look at the conditional probability of network G on a given action profile $\mathbf{s}$,

$$\mu(G|\mathbf{s}) = \prod_{i=1}^n \prod_{j>i}^n \mu(a_{ij}|\mathbf{s}, G_{-ij}) = \prod_{i=1}^n \prod_{j>i}^n \frac{\exp(a_{ij}(\theta s_i s_j - \zeta_0 - \sum_{k=1}^K h_k(x_{ik}, x_{jk})\zeta_k - z_i - z_j))}{1 + \exp(\theta s_i s_j - \zeta_0 - \sum_{k=1}^K h_k(x_{ik}, x_{jk})\zeta_k - z_i - z_j)}. \tag{28}$$

Since network links are conditional pairwise independent, in (28) the joint probability of G can be represented as the product of conditional probability of a_{ij}, which largely simplifies the calculation of likelihood value. However, the computation required by (28) can still be heavy when the network size is large. To further alleviate the computational burden, we adopt the case-control approach of Raftery et al. [2012] which will reduce the computational cost from $O(n^2)$ to $O(n)$.[23] Based on (28), we consider the log likelihood function

$$l(G|\mathbf{s}) = \sum_{i=1}^n l_i(a_{i\cdot}|\mathbf{s}, G_{-i}), \tag{29}$$

where $a_{i\cdot}$ denotes the i^{th} row of matrix $\mathbf{A}$ and $l_i(a_{i\cdot}|\mathbf{s}, G_{-i}) := \sum_{j>i}^n \ln \mu(a_{ij}|\mathbf{s}, G_{-ij})$. To calcaulte $l_i(a_{i\cdot}|\mathbf{s}, G_{-i})$, it is useful to divide the observations $a_{i\cdot}$ into the groups of edges ("cases")

[22] Similarly, in the LIL scenario, we have $\tilde{\mu}_c(\mathbf{s}|G, \mathbf{p}) = \prod_{i=1}^n \tilde{\mu}(s_i = 1|\mathbf{s}_{-i}, G, \mathbf{p})^{\frac{1+s_i}{2}} \tilde{\mu}(s_i = -1|\mathbf{s}_{-i}, G, \mathbf{p})^{\frac{1-s_i}{2}}$.

[23] In medical case-control studies, available infection cases are collected and corresponding controls are sampled from the disease-free cohort.

and non-edges ("control") and perform the following decomposition:

$$\begin{aligned} l_i(a_{i,.}|\mathbf{s}, G_{-i}) &= \sum_{j>i}^{n} a_{ij}(\theta s_i s_j - \zeta_{ij}) - \ln(1+\exp(\theta s_i s_j - \zeta_{ij})) \\ &= \sum_{j>i, a_{ij}=1} (\theta s_i s_j - \zeta_{ij} - \ln(1+\exp(\theta s_i s_j - \zeta_{ij}))) + \sum_{j>i, a_{ij}=0} (-\ln(1+\exp(\theta s_i s_j - \zeta_{ij}))) \\ &= l_{i,1} + l_{i,0}. \end{aligned} \tag{30}$$

In (30), $l_{i,1}$ and $l_{i,0}$ separately stand for the log likelihoods from edges and non-edges. When the network links are sparse, the quantity $l_{i,0}$ can be viewed as a population total statistics. This population total can be estimated by a random sample of the population,

$$\widetilde{l}_{i,0} = \frac{n_{i,0}}{m_{i,0}} \sum_{r=1}^{m_{i,0}} (-\ln(1+\exp(\theta s_i s_r - \zeta_{ir}))), \tag{31}$$

where $n_{i,0}$ is the total number of zero's in the i^{th} row of the upper triangle 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 triangle of matrix $\mathbf{A}$. $\widetilde{l}_{i,0}$ is an unbiased estimator of $l_{i,0}$ given the random samples. When the network size is large, we can choose a small $m_{i,0}$ to compute $\widetilde{l}_{i,0}$ and reduce the amount of computation.[24]

Another computational issue concering the composite likelihood function of (22) is that one needs to integrate over the random effects z in order to obtain the likelihood for observed data, i.e., $\mu(\mathbf{s}|G)\mu(G|\mathbf{s}) = \int_z \mu(\mathbf{s}|G,z)\mu(G|\mathbf{s},z)f(z)dz$. The frequentist approach typically uses Gaussian quadratures or Monte Carlo integration to evaluate such likelihood values. However, given a 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 $\Theta = (\theta, \rho, \boldsymbol{\beta}^\top, \tau, \zeta, \sigma_z^2)$ and unobserved random effects z with the posterior distribution $p(\Theta, z|\mathbf{s}, G)$ derived based on the composite conditional likelihoods in (22). We specify the prior distributions as follows: $\theta \sim \mathcal{N}(0, \sigma_\theta^2)$, $\rho \sim \mathcal{N}(0, \sigma_\rho^2)$, $\boldsymbol{\beta} \sim \mathcal{N}(0, \Sigma_\beta)$, $\tau \sim \mathcal{N}(0, \sigma_\tau^2)$, $\zeta \sim \mathcal{N}(0, \Sigma_\zeta)$, $z_i \sim \mathcal{N}(0, \sigma_z^2)$, 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 the parameter space. Specifically, we set $\sigma_\theta^2 = \sigma_\rho^2 = \sigma_\tau^2 = 10$, $\Sigma_\beta = 10 \cdot I_{k+1}$, $\Sigma_\zeta = 10 \cdot I_{R+1}$, $\nu_0 = 2.2$ and $\xi_0 = 0.1$. We implement the Metropolis-Hastings-within-Gibbs algorithm to simulate draws sequentially from the following conditional posterior densities:

1. simulate the random variable z_i using the M-H algorithm based on $p(z_i|\mathbf{s}, G, \Theta)$ for $i = 1, \cdots, n$.
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 ζ using the M-H algorithm based on $p(\zeta|\mathbf{s}, G, z, \Theta \setminus \zeta)$.

[24] In this study, we set $m_{i,o} = 100 + 5\sum_{j \neq i} a_{ij}$ to obtain the empirical results in Section 5.3 and simulation results in Section 5.4.

We collect the draws from 30,000 iterations according to the above steps, drop the first 5,000 iterations for burn-in, and compute the posterior mean and the posterior standard deviation from the converged draws as our estimation results.

5.3. Estimation Results

The estimation results of the GI and LIL scenarios are reported in Tables 2 and 3 respectively. In each table, Column I presents the result of the full model and Column II presents the result of the model without individual random effects. Our first finding is that both local and global interaction effects (θ and ρ) are "important" as expected, i.e., positive and significant. Nonetheless, by comparing the two columns, we can see that the estimates of local spillover effect (θ) and other coefficients are biased when failing to control individual unobserved heterogeneity through random effects. In particular, the estimate of 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 global conformity effect $\tilde{\rho}$. The fact that the global conformity effect is significant in Table 3 provides a strong motivation to the belief-based formulation process under local information. Moreover, we obtain the estimate of φ in (10) which equals 0.0961, suggesting that in general the weights that agents put on local average beliefs are more than nine times larger than the weights put on local average actions in updating their own beliefs.

The results of the full model also show various sources of heterogeneity, as captured by the idiosyncratic preference γ_i, play a prominent and intuitive role in rioting decision. Specifically, we find that females are less likely to support or 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 captured by the estimate of τ on rioting. In the LIL scenario, we obtain a negative estimate of $\tilde{\kappa}$. This may reflect that both the rioting cost κ and the constant term β_0 in γ_i are positive but β_0 is larger than κ in magnitude. In terms of linking cost, our estimation results show a high constant cost and a clear homophily pattern in which similar characteristics (e.g., same gender or same religion) lower linking cost.

In Table 3 we further report results in columns III and IV where we omit the local spillover effect and the global conformity effect respectively. The results show that when one of these two effects is omitted, the other effect will be confounded which leads to an upward estimation bias. This finding reminds us the importance of controlling both local and global interaction effects simultaneously as they both play an indispensable role in determining people's collective actions.

5.4. Parameter Recovery Analysis

We conduct a Monte Carlo simulation study to demonstrate that the proposed Bayesian MCMC estimation based on the composite conditional likelihoods of (22) and the case-control approach described in Section 5.2 can indeed recover true parameter values. We also use this simulation to confirm the direction of estimation bias observed in Tables 2 and 3 when individual unobserved heterogeneity (random effect) is ignored.

We set the number of Monte Carlo repetitions to 300. In each repetition, we generate the artificial network (G) and the action profile ($\mathbf{s}$) from a data generating process (DGP) which

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

		(I) full model	(II) w/o random effects
Local spillover	(θ)	0.1732***	0.2287***
		(0.0031)	(0.0020)
Global conformity	($\widetilde{\rho}$)	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) followers count	(β_3)	0.0117***	0.0088***
		(0.0016)	(0.0015)
Random effect	(τ)	0.0057***	
		(0.0006)	
Linking cost:			
Constant	(ζ_0)	14.7484***	12.5889***
		(0.0195)	(0.0077)
Same gender	(ζ_1)	-0.1661***	-0.1908***
		(0.0134)	(0.0082)
Same religiousness	(ζ_2)	-0.0762***	-0.0033
		(0.0080)	(0.0058)
Diff. in followers count	(ζ_3)	0.0871***	0.0986***
		(0.0033)	(0.0022)
Variance of random effect	(σ_z^2)	2.1568***	
		(0.0178)	
Sample size		225,578	

Notes: The constant term β_0 in idiosyncratic preference, the rioting cost κ, and the global conformity effect $\rho \sum_{j \neq i}^{n} s_j$ are not distinguishable in the GI scenario. For the purpose of identification, we replace $\rho \sum_{j \neq i}^{n} s_j$ with $\widetilde{\rho}(n-1)\bar{s}$ in the model and drop β_0 and κ. Therefore, $\widetilde{\rho}$ reported in this table is a confounded estimate of the global conformity effect. The parameter estimates reported in this table are the posterior mean and the posterior standard deviation from the Bayesian MCMC sampling. The asterisks ***(**,*) indicate that the 99% (95%,90%) highest posterior density interval of the corresponding draws do not cover zero.

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

		(I) full model	(II) w/o random effects	(II) w/o local spillover	(II) w/o global conformity
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) 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:	($\widetilde{\kappa}$)	-0.2965***	-0.3054***	-0.2805***	0.3348***
		(0.0110)	(0.0089)	(0.0091)	(0.0092)
Linking cost:					
Constant	(ζ_0)	14.7964***	12.5633***	12.5194***	12.5921***
		(0.0229)	(0.0093)	(0.0096)	(0.0103)
Same gender	(ζ_1)	-0.1708***	-0.1930***	-0.2012***	-0.1934***
		(0.0146)	(0.0094)	(0.0092)	(0.0094)
Same religiousness	(ζ_2)	-0.0784***	-0.0048	0.0065	0.0024
		(0.0095)	(0.0062)	(0.0063)	(0.0066)
Diff. in followers	(ζ_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			225,578		

Notes: The constant β_0 in the individual preference and the rioting cost κ are not distinguishable in the LIL scenario. For the purpose of identification, we denote and estimate $\widetilde{\kappa} = \kappa - \beta_0$. The parameter estimates reported in this table are the posterior mean and the posterior standard deviation from the Bayesian MCMC sampling. The asterisks ***(**,*) indicate that the 99% (95%, 90%) highest posterior density interval of the corresponding draws do not cover zero.

mimics the dynamic process outlined in Section 2.2. We set the sample sizes (n) to 3,000. In the DGP, we capture individual idiosyncratic preference γ by $x\beta + z\tau$, where the variable x represents observed individual characteristic, which are generated from a mixture of normal distributions, i.e., two-fifths of values are generated from $\mathcal{N}(-4, 36)$ and three-fifths of values are generated from $\mathcal{N}(4, 36)$. The variable z represents unobserved individual random effects which are generated from a normal distribution, i.e., $\mathcal{N}(0, 1)$. The coefficients β and τ are both set to 0.5. For the linking cost function, $\zeta_{ij} = \xi_0 + h(x_i, x_j)\xi_1 + z_i + z_j$, we set $h(x_i, x_j) = |x_i - x_j|$ and the coefficients ζ_0 and ζ_1 are set 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. We normalize the scale parameter η of the logistic disturbance to 1. For the LIL scenario, we set the true rioting cost parameter κ to 1.5 and the true belief weight ψ to 0.5.

To generate artificial data, we implement an iterative process that simulates individual network links and actions from the corresponding conditional probabilities. In each iteration, an individual is chosen randomly to update either network links or action (and belief if it is the LIL scenario) conditional on other's network links and actions in the previous iteration. We run this iteration process sufficiently long and treat the final realization of network and action profile from this iteration process as our artificial data. Here are some remarks and details about this process. First, to simulate individual's network link, we use the following conditional probability of network link

$$\mu(a_{ij} = 1|\mathbf{s}, G_{-ij}) = \frac{\exp(a_{ij}(\theta s_i s_j - \zeta_0 - |x_i - x_j|\zeta_1 - z_i - z_j)}{1 + \exp(\theta s_i s_j - \zeta_0 - |x_i - x_j|\zeta_1 - z_i - z_j)}.$$

In fact, due to conditional independence of network links, we draw $\{a_{ij}\}_{j=1}^n$ synchronously to speed up the simulation. Second, to simulate individual action s_i from the GI scenario, we use the following conditional probability of action

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

To prevent the identification problem discussed in Section 5.1, we do not include an intercept term in $x_i\beta$ and the rioting cost κ in the GI scenario. Third, to simulate individual action s_i from the LIL scenario, we use the following conditional probability of action

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

where p_i is generated from the induced stationary beliefs given in (10). Again, to prevent the identification problem, we do not include an intercept term in $x_i\beta$. We use an iteration process that corresponds to the dynamic process in Section 2.2 for some particular rates of adjustment in which (i) individuals update network links and actions with equal frequency; (ii) beliefs adjustments occur very fast so that we can approximate an individual's belief using the induced stationary beliefs.

We implement the Bayesian MCMC estimation with 20,000 iterations and drop the first 5,000 iterations for burn-in. The simulation results of the GI scenario are summarized in Table 4 and the results of the LIL scenario are in Table 5. The values reported in the tables are the mean and standard deviation of parameter estimates calculated across repetitions. In both Tables we see that this estimation approach can successfully recover the true model parameters when considering the full model with random effects (the true DGP model). Moreover, similar to the

pattern that we observe in the empirical study, the estimate of local spillover θ is upward biased and the estimate of global conformity ρ is downward biased when unobserved heterogeneity (z_i) are ignored.

Let us reiterate that this simulation study serves two important purposes. First, it illustrates that the composite-likelihood approach underlying the econometric analysis appears to be an effective estimation procedure, independently of the informational assumption we make (of course, as long as we choose the "right" informational scenario). Second, it provides further support to the claim that not only the potential approach used for the GI scenario, but also the quasi-potential one applied to the LIL context, capture the essential long-run behavior of the system and hence represent useful bases to conduct the econometric estimation in either case.

Table 4: Simulation results of the global information (GI) scenario

			full model		w/o random effect	
		DGP	Est.	Std.	Est.	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						
	(β)	0.5000	0.5059	0.1285	0.4155	0.0677
	(τ)	0.5000	0.4510	0.1911		
Linking cost						
	(ζ_0)	2.0000	2.0145	0.0666	1.5291	0.0091
	(ζ_1)	1.0000	1.0145	0.0064	0.8808	0.0034
Variance of RE	(σ_z^2)	1.0000	1.1045	0.0353		
Sample size			3000			
MC repetitions			300			

Table 5: Simulation results of the local information and learning (LIL) scenario

			full model		w/o random effect	
		DGP	Est.	Std.	Est.	Std.
Local spillover	(θ)	0.0500	0.0543	0.0127	0.0740	0.0067
Global conformity	(ρ)	0.0010	0.0013	0.0005	0.0007	0.0003
Weight on local obsev.	(φ)	0.5000	0.6002	0.1833	0.6197	0.1685
Rioting cost	(κ)	1.5000	1.6224	0.7728	1.8748	0.7777
Individual preference						
	(β)	0.5000	0.5439	0.2022	0.5441	0.2020
	(τ)	0.5000	0.5596	0.2631		
Linking cost						
	(ζ_0)	2.0000	1.9493	0.0660	1.4656	0.0084
	(ζ_1)	1.0000	1.0138	0.0057	0.8845	0.0032
Variance of RE	(σ_z^2)	1.0000	1.0998	0.0351		
Sample size			3000			
MC repetitions			300			

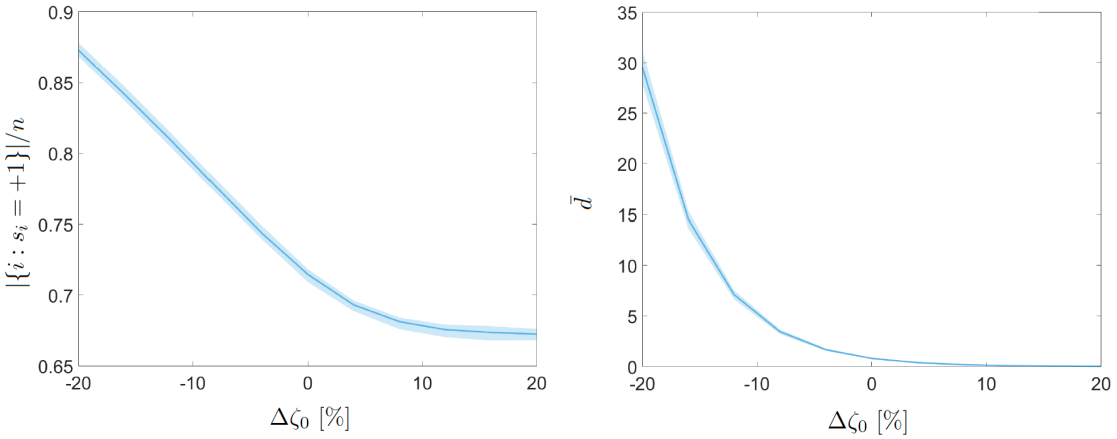


Figure 10: Changes in the fraction of rioting agents, $|\{i : s_i = +1\}|/n$ (left panel), and the average degree, $\bar{d}$ (right panel), over percentage changes in the linking cost ζ_0. We plot the mean and 95% interval from 300 simulation repetitions.

6. Counterfactual Analyses

Using the estimation result based on the LIL scenario in Section 5.3 we can perform some counterfactual analyses by changing specific parameters of the model (while keeping the remaining parameters at their estimated values in Table 3) and evaluating the impact on outcome variables through simulating network links and action profile from the invariant distribution $\mu(\mathbf{s}, G|\Theta)$ with the same iterative process described in Section 5.4. We will focus on two specific counterfactual scenarios in the following. In the first we examine the role of linking costs on rioting behavior. In the second we analyze how biasing the beliefs towards a specific action can influence the rioting outcome.

6.1. Impact of Linking Costs on Rioting Behavior

During the protests against the Egyptian government mobile phone operators have been instructed to suspend services in selected areas, internet access was blocked and mobile phone and text messaging services were 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 practiced by other countries such as China, where some social media platforms like Facebook, YouTube and Twitter have been blocked, and users can only indirectly use them via a VPN service [Willnat et al., 2015]. We can simulate such increased communication costs by analyzing the influence of the linking cost on rioting behavior in our estimated model.

For our counterfactual exercise, we change the value of the estimate of ζ_0 (which is the constant term in the linking cost in Equation (21)) over the following percentage grid points: [-20%, -16%, -12%, -8%, -4%, 0%, 4%, 8%, 12%, 16%, 20%]. For each case, we simulate the network and action profile 300 times and compute the average fraction of rioting agents together with the average network degree in the new equilibrium. The right panel in Figure 10 shows the average degree over varying ζ_0. As expected, we can see that the linking cost ζ_0 has a substantial effect on the average degree so that lowering the linking cost gives rise to a much denser network structure. The left panel in Figure 10 shows the fraction of rioting agents over

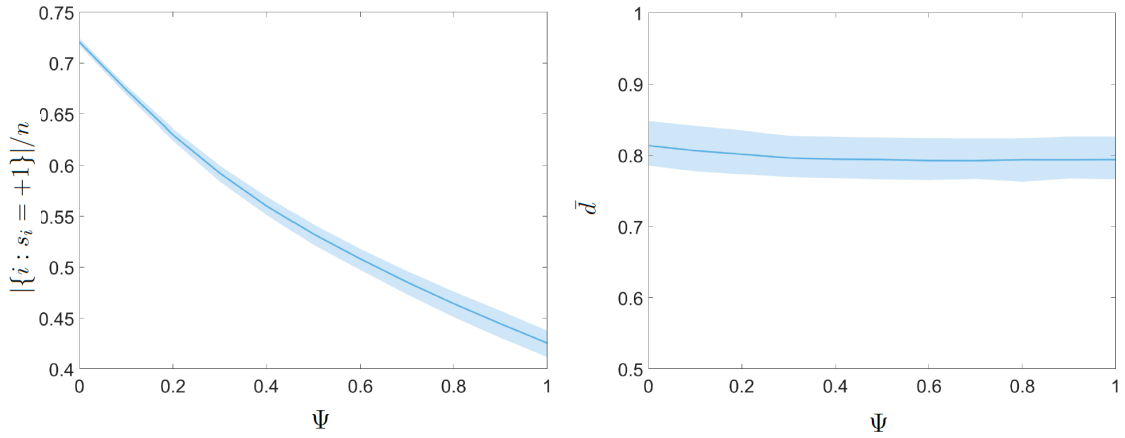


Figure 11: Changes in the fraction of rioting agents, $|\{i : s_i = +1\}|/n$ (left panel), and the average degree, $\bar{d}$ (right panel), over the strength of propaganda Ψ. We plot the mean and 95% interval from 300 simulation repetitions.

varying ζ_0. We observe that a reduction of ζ_0 by 20% yields an increase in the fraction of rioting agents by almost 15%. Conversely, this indicates that, as linking 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.

6.2. Impact of Beliefs Manipulation on Rioting Behavior

Governments often use manipulation of the information available on social or other media to distort the users' view [cf. e.g. 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].

In our second counterfactual analysis we examine the effectiveness of manipulating the beliefs of the agents to mitigate rioting behavior. We introduce a government influencing the belief updating Equation (7) as follows:

$$p_{it}^{u+1} = (1-\Psi)\left\{\varphi\frac{1}{d_{it}}\sum_{j=1}^{n}a_{ij,t}s_{jt} + (1-\varphi)\frac{1}{d_{it}+1}\left[p_{it}^{u} + \sum_{j=1}^{n}a_{ij,t}p_{jt}^{u}\right]\right\} + \Psi g, \tag{32}$$

where $\Psi \in [0,1]$ and $g = -1$ is the preferred action of the government (status quo). As in the previous section, for every value of Ψ, we simulate the network and action profile 300 times and compute the average fraction of rioting agents together with the average network degree in the new equilibrium. The changes of in the fraction of rioting agents and the average degree with varying levels of the influence of propaganda, Ψ, can be seen in Figure 11. Our results show that while propaganda (belief manipulation) does not affect the network density – which remains roughly stable – it has a drastic effect on rioting, reducing the fraction of rioting agents by 30%. Our findings highlight how the manipulation of information can mitigate the formation of collective action, but also that (within the simple model and context analyzed here) it cannot suppress it entirely.

7. Conclusion

In this paper we have introduced a model of protest participation in a connected population where the network co-evolves with actions and beliefs. We provide a complete characterization of the equilibrium action choices, beliefs and networks, and show that a threshold exists in the linking cost and the conformity parameter such that all agents coordinate on the same action. We further find that the introduction of incomplete information via beliefs lowers the threshold. We then bring our model to the data by relying on large-scale Twitter data during the Arab Spring on social unrests in Egypt. We perform a structural estimation of the model's parameters and use random/fixed effects to capture the unobserved component in the idiosyncratic preferences of the agents (for retaining the status quo vs. changing it). We jointly estimate (and disentangle) the local peer effect and the global conformity effect in the agents' participation decision, and show that both are significant. Moreover, we show that ignoring endogeneity from network formation may bias the estimates of these two effects.

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.

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.

Angeletos, G.-M., Hellwig, C., and Pavan, A. (2007). Dynamic global games of regime change: Learning, multiplicity, and the timing of attacks. *Econometrica*, 75(3):711–756.

Angeletos, G.-M. and Pavan, A. (2007). Socially optimal coordination: Characterization and policy implications. *Journal of the European Economic Association*, 5(2-3):585–593.

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

Arkolakis, C. and Eckert, F. (2017). Combinatorial discrete choice. *Yale University, Working Paper*.

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 Consensus Using DeGroot's Method. *Journal of the American Statistical Association*, 76:415–418.

Besag, J. (1974). Spatial interaction and the statistical analysis of lattice systems. *Journal of the Royal Statistical Society: Series B (Methodological)*, 36(2):192–225.

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, The Netherlands: 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. *Review of Economic Studies*, 68(2):235–260.
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.
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.
Chung, F. and Lu, L. (2007). *Complex Graphs and Networks*. American Mathematical Society.
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.
Dzemski, A. (2019). An empirical model of dyadic link formation in a network with unobserved heterogeneity. *Review of Economics and Statistics*, 101(5):763 – 776.
Earl, J. and Kimport, K. (2011). *Digitally enabled social change: Activism in the internet age*. MIT Press.
Edmond, C. (2013). Information manipulation, coordination, and regime change. *Review of Economic Studies*, 80(4):1422–1458.
Enikolopov, R., Makarin, A., Petrova, M., et al. (2016). 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.
Ghosal, P. and Mukherjee, S. (2020). Joint estimation of parameters in Ising model *The Annals of Statistics*, 48(2):785–810.
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.
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., A. Galeotti, and B. Rogers (eds.), *Oxford Handbook of Economic Networks*, Oxford: Oxford University Press.
González, F. (2016). Collective action in networks: Evidence from the chilean student movement. *Working Paper, Pontifical Catholic University of Chile – Institute of Economics*.

González-Bailón, S. (2017). *Decoding the social world: Data science and the unintended consequences of communication*. MIT 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.
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*.
Grimmer J. and Brandon M. Stewart, 2013. Text as data: The promise and pitfalls of automatic content analysis methods for political texts. *Political analysis*, 21(3):26–297.
Grimmett, G. (2010). *Probability on graphs*. Cambridge University Press.
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., König, M. D. and Liu, X. (2021) A Structural Model for the Coevolution of Networks and Behavior *Review of Economics and Statistics*, 104(2):355–367.
Humphrey, A. and R. J.-H. Wang (2017). Automated text analysis for consumer research. *Consumer Research*, 44(6):127–1306.
Hughes, J. and Haran, M. and Caragea, P. C. (2011). Autologistic models for binary data on a lattice. *Environmetrics*, 22(7):857–871.
Ising, E. (1925). Beitrag zur Theorie des Ferromagnetismus. *Zeitschrift für Physik A: Hadrons und Nuclei*, 31(1):253–258.
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.
Jochmans, K. (2018). Semiparametric analysis of network formation. *Journal of Business & Economic Statistics*, 36(4):705–713.
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.
Kolaczyk, E. D. (2009). *Statistical Analysis of Network Data: Methods and Models*. Springer.
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/.
Lee, L.-f., Li, J., and Lin, X. (2014). Binary choice models with social network under heterogeneous rational expectations. *Review of Economics and Statistics*, 96(3):402–417.
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.
Magdy, W., Darwish, K., Abokhodair, N., Rahimi, A., and Baldwin, T. (2016). #ISISisNo-

tIslam or #DeportAllMuslims?: Predicting unspoken views. In *Proceedings of the 8th ACM Conference on Web Science*, 95–106.
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.
Mesbahi, M. and Egerstedt, M. (2010). *Graph theoretic methods in multiagent networks*. Princeton University Press.
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.
Monderer, D. and Shapley, L. (1996). Potential Games. *Games and Economic Behavior*, 14(1):124–143.
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.
Phan, D. and Semeshenko, V. (2008). Equilibria in models of binary choice with heterogeneous agents and social influence. *European Journal of Economic and Social Systems*, 21(1):7–37.
Priante, A., Ehrenhard, M. L., van den Broek, T., and Need, A. (2018). Identity and collective action via computer-mediated communication: A review and agenda for future research. *New media & society*, 20(7):2647–2669.
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.
Shadmehr, M. (2021). Tullock's Paradox, Hong Kong experiment, and the strength of weak states *Quarterly Journal of Political Science*, 16(3):245–264.
Strauss, D. and Ikeda, M. (1990). Pseudolikelihood estimation for social networks *Journal of the American Statistical Association*, 85(409):204–212.
Tullock, G. (1971). The paradox of evolution *Public Choice* 11, 89–99.
Varin, C., Reid, N. and Firth, D. (2011). An overview of composite likelihood methods. *Statistica Sinica*, 5–42.
Van Mieghem, P. (2011). *Graph spectra for complex networks*. Cambridge University Press.
van Duijn, M. A. J., Gile, K. and Handcock, M. S. (2009). An overview of composite likelihood methods. *Social Networks*, 31(1):52–62.
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.
Wong, R. (2001). *Asymptotic approximations of integrals*, Society for Industrial Mathematics.
Yedidia, J. S., Freeman, W. T., and Weiss, Y. (2001). *Advances in Neural Information Processing Systems*, chapter Generalized belief propagation, 689–695. MIT Press, Cambridge, MA.
Young, H. P. (1993). The evolution of conventions. *Econometrica*, 61(1):57–84.
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.
Zhou, X. and Schmidler, S. C. (2009). Bayesian parameter estimation in Ising and Potts models: A comparative study with applications to protein modeling. Working paper, Department of Statistical Science, Duke University, Durham, NC.
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. Extensions

A.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

$$\Phi(\mathbf{s}, G) = \sum_{i=1}^{n} \gamma_i s_i + \theta \sum_{i=1}^{n} \sum_{j=1}^{n} \left(\frac{1}{2} a_{ij} a_{ji} + a_{ij}(1 - a_{ji}) \right) s_i s_j + \frac{\rho}{2} \sum_{i=1}^{n} \sum_{j \neq i}^{n} s_i s_j - \kappa \sum_{i=1}^{n} s_i - m\zeta. \quad (33)$$

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.

A.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 Equation (1) up to a shift of the parameter θ.

B. 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, 4, 5 and 6 cover only the case of vanishing noise, i.e. the limit of $\eta \to \infty$). For concreteness, we focus on two especially relevant statistics, their average connectivity (network degree) and their average action, as they depend on the linking costs θ $(= \zeta_1 = \zeta_2)$ and the noise parameter η.

B.1. Global Information

We start by characterizing the expected number of links induced by the distribution $\mu^\eta(\cdot)$.

Proposition 7. *Assume homogeneous linking costs, $\zeta_1 = \zeta_2 = \zeta$. Then the expected number of*

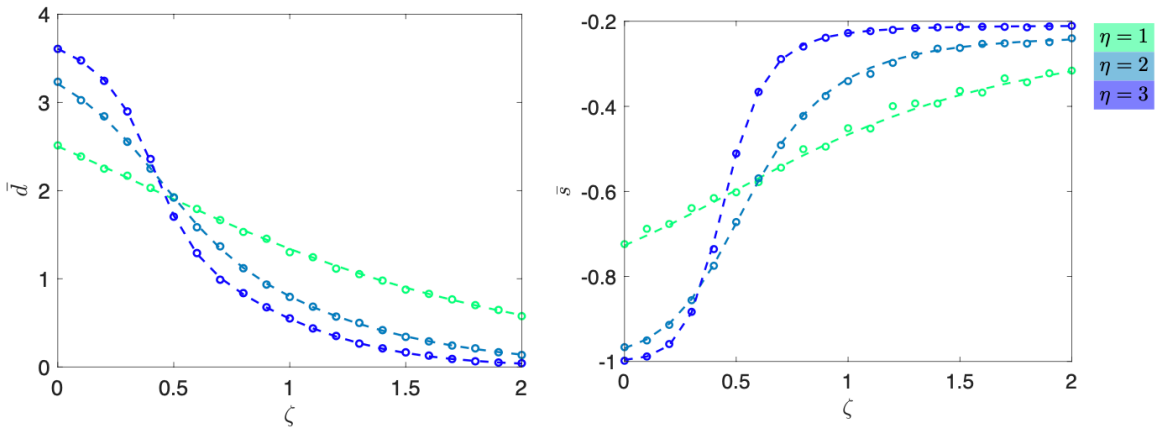


Figure 12: The average degree $\bar{d} = 2m/n$ (left panel) and the average action level $\bar{s}$ (right panel) across different values of the linking cost $\zeta \in [0, 2]$ and varying noise $\eta \in \{1, 2, 3\}$. The parameters used are $n = 5$, $n_+ = 2$, $\kappa = 0.1$, $\rho = 0.1$, $\lambda = \chi = \xi = 1$ and $\theta = 0.75$. Dashed lines indicate the theoretical prediction of Equations (34) in Proposition 7 and (36) in Proposition 8, 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].

links in the stationary state is given by

$$\mathbb{E}^{\eta}(m) = \frac{1}{\mathscr{Z}^{\eta}} \sum_{k=0}^{n} \sum_{j=0}^{\min\{k,n_+\}} \binom{n_+}{j} \binom{n-n_+}{k-j} e^{\eta(2k-n)} e^{\eta\left(\rho\left(2l(k,j)-\binom{n}{2}\right)\right)-\kappa(n-2(n_++k-2j))}$$
$$\times \left(1+e^{\eta(\theta-\zeta)}\right)^{l(k,j)} \left(1+e^{-\eta(\theta+\zeta)}\right)^{\binom{n}{2}-l(k,j)} \left(\frac{l(k,j)}{1+e^{-\eta(\theta-\zeta)}} + \frac{\binom{n}{2}-l(k,j)}{1+e^{\eta(\theta+\zeta)}}\right), \tag{34}$$

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

$$l(k,j) = \frac{n^2 + (2(2j-k)-1)n + 2(2j-k)^2 - 2(n+2(2j-k)-n_+)n_+}{2}, \tag{35}$$

$n_+ = \#(\{\gamma_i = 1 : i = 1, \ldots, n\})$, and we have that $\lim_{\zeta\to\infty} \mathbb{E}^{\eta}(m) = 0$.[25]

In the left panel of Figure 12 we compare the average degree $\bar{d}$ obtained by averaging across simulations with the expected value $2\mathbb{E}^{\eta}(m)/n$ from Proposition 7 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 7.

Proposition 8. *Assume homogeneous linking costs, $\zeta_1 = \zeta_2 = \zeta$. Then the expected average*

[25] An explicit expression for the partition function $\mathscr{Z}^{\eta}$ can be found in Lemma 3 in Appendix E.

action level, $\bar{s} = \frac{1}{n}\sum_{i=1}^{n} s_i$, in the stationary state is given by

$$\mathbb{E}^{\eta}(\bar{s}) = \frac{1}{\mathscr{Z}^{\eta}} \sum_{k=0}^{n} \sum_{j=0}^{\min\{k,n_+\}} \binom{n_+}{j}\binom{n-n_+}{k-j} \frac{n+4j-2(n_++k)}{n} e^{\eta(2k-n)} \times e^{\eta\left(\rho\left(2l(k,j)-\binom{n}{2}\right)\right)-\kappa(n-2(n_++k-2j))} \left(1+e^{\eta(\theta-\zeta)}\right)^{l(k,j)} \left(1+e^{-\eta(\theta+\zeta)}\right)^{\binom{n}{2}-l(k,j)}, \quad (36)$$

where $l(k,j)$ is defined in Equation (35) and $n_+ = \#(\{\gamma_i = 1 : i = 1,\ldots,n\})$.

The average action level $\bar{s}$ in Proposition 8 is illustrated in the right panel of Figure 12 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).

B.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 (10) 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 (10):

$$\mathbf{p} = \varphi\left[\mathbf{I} - (1-\varphi)\widehat{\mathbf{D}}^{-1}\widehat{\mathbf{A}}\right]^{-1}\mathbf{D}^{-1}\mathbf{A}\mathbf{s}, \quad (37)$$

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 ($\mathscr{Z}^{\eta}$) and other statistics of interest – such as the average degree or the average action level – for an arbitrary level of noise.

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

Proposition 9. *Consider homogeneous linking costs, $\zeta_1 = \zeta_2 = \zeta$. Then, under a mean field approximation, the expected number of links is given by*

$$\mathbb{E}^{\eta}(m) \simeq \frac{1}{\mathscr{Z}^{\eta}} \sum_{\mathbf{s}\in\{-1,+1\}^n} e^{\eta\langle\tilde{\boldsymbol{\gamma}},\mathbf{s}\rangle} h^{\eta}(\mathbf{s}), \quad (38)$$

where $\tilde{\gamma}_i = \gamma_i + \rho(n-1)p_i - \kappa$, beliefs $\mathbf{p}$ are given by Eq. (37), 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

$$\mathscr{Z}^{\eta} = \sum_{\mathbf{s}\in\{-1,+1\}^n} e^{\eta\langle\tilde{\boldsymbol{\gamma}},\mathbf{s}\rangle} f^{\eta}(\mathbf{s}), \quad (39)$$

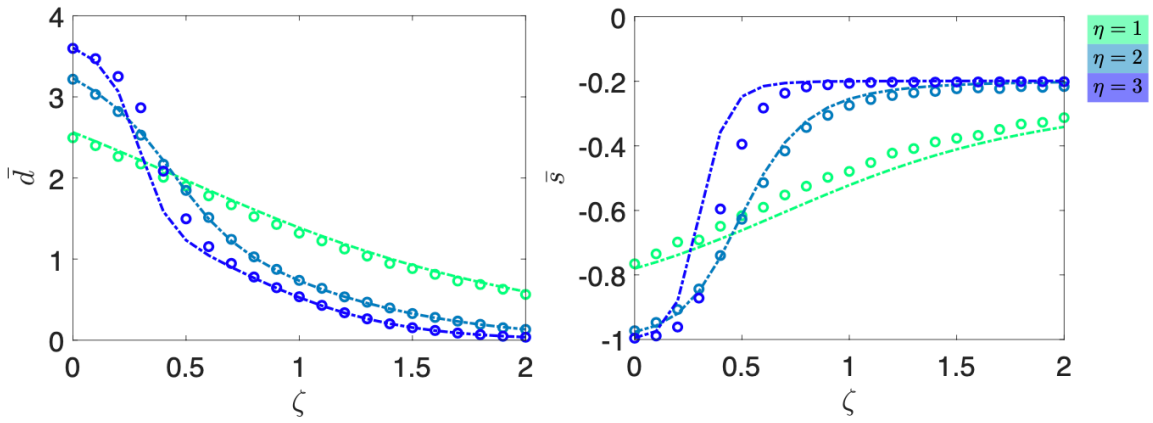


Figure 13: The average degree $\bar{d} = 2m/n$ (left panel) and the average action level $\bar{s}$ (right panel) across different values of the linking cost $\zeta \in [0, 2]$, $\eta \in \{1, 2, 3\}$, $n = 5$, $n_+ = 2$, $\kappa = 0.1$, $\rho = 0.1$, $\lambda = \chi = \xi = 1$, $\varphi = 0.5$ and $\theta = 0.75$. Dashed-dotted lines indicate the theoretical predictions of $\bar{d} = 2m/n$ in Equation (38) in Proposition 9 and of $\bar{s}$ in Equation (41) of Proposition 10, 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

$$h^{\eta}(\mathbf{s}) = \frac{\left(e^{-\eta(\zeta+\theta)}+1\right)^{\alpha(\mathbf{s})}\left(e^{\eta(\theta-\zeta)}+1\right)^{\beta(\mathbf{s})-1}\left((\alpha(\mathbf{s})+\beta(\mathbf{s}))e^{\eta(\theta-\zeta)}+\alpha(\mathbf{s})+\beta(\mathbf{s})e^{2\eta\theta}\right)}{1+e^{\eta(\theta+\zeta)}},$$
$$f^{\eta}(\mathbf{s}) = \left(1+e^{-\eta(\zeta+\theta)}\right)^{\alpha(\mathbf{s})}\left(1+e^{\eta(\theta-\zeta)}\right)^{\beta(\mathbf{s})}, \tag{40}$$

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

The left panel in Figure 13 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 10. *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*

$$\mathbb{E}^{\eta}(\bar{s}) \simeq \frac{1}{\mathscr{Z}^{\eta}} \frac{1}{n} \sum_{\mathbf{s} \in \{-1,+1\}^n} \langle \mathbf{u}, \mathbf{s} \rangle e^{\eta \langle \tilde{\gamma}, \mathbf{s} \rangle} f^{\eta}(\mathbf{s}), \tag{41}$$

with $f^{\eta}(\cdot)$ *given by Equation (40), the partition function is given by Equation (39),* $\tilde{\gamma}_i = \gamma_i + \rho(n-1)p_i - \kappa$, *beliefs* $\mathbf{p}$ *are given by Eq. (37), 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* $d_i = \sum_{j=1}^n a_{ij}$, *and* $\langle \cdot, \cdot \rangle$ *is the usual scalar product in* $\mathbb{R}^N$.

The right panel in Figure 13 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 13 also illustrates a good match between the theory and simulations for different values of ζ and η.

C. Finite Population Equilibrium Characterization

C.1. Complete Information

The following proposition provides a complete characterization of the SSS for finite populations in the complete information environment (generalizing Propositions 3 and 4).

Proposition 11. *Let $n_+ = \#(\{\gamma_i = 1 : i = 1, \ldots, n\})$, $\nu = n_+/n$ and denote by*

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

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

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

(ii) In the case of $\theta > \zeta_1$ the stochastically stable state is either complete, K_n, or composed 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 $\nu_+ < 1/2$ and*
   (a) *$\theta > \zeta_2$ and $\theta^* < \zeta_2$ or if $\theta^* > \zeta_2$ and $\theta > \theta^*$ then the stochastically stable state is the **complete graph** K_n in which all agents choose the action $s_i = -1$; if $\theta^* > \zeta_2$ and $\theta < \theta^*$ then the stochastically stable state is the **union of two cliques**, $K_{n_+} \cup K_{n-n_+}$, in which all agents choose the action $s_i = \gamma_i$;*
   (b) *$\theta < \zeta_2$ and $\theta^* < \zeta_2$ and $\theta > \theta^*$ then the stochastically stable state is the **union of***

two cliques, $K_{n_+} \cup K_{n-n_+}$, *in which all agents choose the action* $s_i = -1$ *while if* $\theta < \theta^*$ *then all agents choose the action* $s_i = -1$ *if* $\rho > \rho^*$ *and all agents choose the action* $s_i = \gamma_i$ *if* $\rho < \rho^*$*; if* $\theta^* > \zeta_2$ *then the stochastically stable state is the* ***union of two cliques***, $K_{n_+} \cup K_{n-n_+}$, *in which all agents choose the action* $s_i = \gamma_i$ *while if* $\rho > \rho^*$ *all agents in the cliques choose the action* $s_i = -1$*;*

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

Proposition 11 shows that when the idiosyncratic preference is large enough (i.e. θ is small enough) in the payoff function of Equation (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 chose $s_i = +1$, 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 ($\rho < \rho^*$).

C.2. Local Information and Learning

The following proposition provides a characterization of the SQS for finite populations in the belief formation environment (generalizing Propositions 5 and 6).

Proposition 12. *Let $n_+ = \#(\{\gamma_i = 1 : i = 1, \ldots, n\})$, denote by $\nu_+ = n/n_+$ and define*

$$\begin{aligned} \tilde{\theta}^* &= \zeta_2 + \frac{2(1-\kappa)}{n-n_+}, \\ \tilde{\theta}^{**} &= \zeta_2 + \frac{2(1+\kappa)}{n_+}, \\ \kappa^* &= 2\nu_+ - 1 \ (\leq 1). \end{aligned} \tag{42}$$

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

(ii) In the case of $\theta > \zeta_1$ the stochastically stable is either complete, K_n, or composed 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 $\nu_+ < 1/2$ and*
   (a) *$\theta > \zeta_2$ and $\kappa < 1$ (such that $\tilde{\theta}^* < \zeta_2$) and $\theta < \tilde{\theta}^*$ then the stochastically stable state is the **union of two cliques**, $K_{n_+} \cup K_{n-n_+}$ in which all agents choose the action $s_i = \gamma_i$, while if $\theta > \tilde{\theta}^*$ and $\rho > (1-\kappa)/(n-1)$ then the stochastically stable state is the **complete graph** K_n in which all agents choose the action $s_i = -1$; if $\kappa > 1$ then the stochastically stable state is the **complete graph** K_n in which all agents choose the action $s_i = -1$;*
   (b) *$\theta < \zeta_2$ and $\kappa < 1$ the stochastically stable state is the **union of two cliques**, $K_{n_+} \cup K_{n-n_+}$, in which all agents choose the action $s_i = \gamma_i$ while if $\kappa > 1$ (such that $\tilde{\theta}^* < \zeta_2$) and $\rho > (1-\kappa)/(n-1)$ then all agents in the cliques choose the action $s_i = -1$;*
2. *if $\nu_+ > 1/2$ and*
   (a) *$\kappa > \kappa^*$ and*
      i. *$\theta > \zeta_2$ and $\kappa < 1$ (such that $\tilde{\theta}^* < \zeta_2$) and $\theta < \tilde{\theta}^*$ then the stochastically stable state is the **union of two cliques**, $K_{n_+} \cup K_{n-n_+}$ in which all agents choose the action $s_i = \gamma_i$, while if $\theta > \tilde{\theta}^*$ and $\rho > (1-\kappa)/(n-1)$ then the stochastically stable state is the **complete graph** K_n in which all agents choose the action $s_i = -1$; if $\kappa > 1$ then the stochastically stable state is the **complete graph** K_n in which all agents choose the action $s_i = -1$;*
      ii. *$\theta < \zeta_2$ and $\kappa < 1$ the stochastically stable state is the **union of two cliques**, $K_{n_+} \cup K_{n-n_+}$, in which all agents choose the action $s_i = \gamma_i$ while if $\kappa > 1$ (such that $\tilde{\theta}^* < \zeta_2$) and $\rho > (1-\kappa)/(n-1)$ then all agents in the cliques choose the action $s_i = -1$;*
   (b) *$\kappa < \kappa^*$ and*
      i. *$\theta > \tilde{\theta}^{**}$ and $\rho > (1+\kappa)/(n-1)$ then the stochastically stable state is the **complete***

graph *K_n in which all agents choose the action $s_i = +1$,*

ii. $\theta < \tilde{\theta}^{**}$ then the stochastically stable state is the ***union of two cliques****, $K_{n_+} \cup K_{n-n_+}$, in which all agents choose the action $s_i = \gamma_i$.*

From Proposition 12 we observe that the possible stochastically stable actions and networks are the same as in Proposition 11, 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 the average action chosen in the entire population, but represent only the average action chosen in the local clique.

D. 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 begun as part of the Arab Spring.

D.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, members where 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 vast presence in civilian industry, making it wealthy opposes Islamist rule.

Egypt's main opposition of the NDP's rule was in the Islamist movement, whose main social organization is the Muslim Brotherhood (MB). The MB's ideology centres around a literal interpretation of scriptures and advocates a return to idealized Islamic society. MB's followers are urbanized middle and lower classes. Outlawed in 1954 connected to assassination attempt of then president Nasser. From 1970s leaders freed and MB moved towards an official political party as many leaders released from prisons (tolerated but not liberated). In the 2005 elections MB gained approx 20 percent of the seats in Egyptian parliament through running as independents, making them a new force within Egyptian society and politics, although the state was still officially denying that it existed.

Mubarak's regime took a vigorous position against Islamic investment companies. This was 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 supporter. The Egyptian society was also characterized by various protest movements over time, some with pro- and some with anti-government orientation.

D.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 civilian population stagnated Following the removal of Tunisian President Bin Ali in early 2011, the fervor against privileged elites and ruling parties in the North Africa and middle East grew. This lead to thousands (5K) of protesters congregating in Cairo's Tahir Square in a public demonstration against Mubarak regime organized by young middle class Egyptians, not Islamist opposition. MB later encouraged their members to participate without invoking the MB's Islamist slogans or ideology. Note protests are illegal in Egypt. After initial protest, demonstrations continued, growing to 50K on Jan 28, by Feb 1 over 500K protesters. In the evening of Feb 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 massive dilemma. The SCAF had to decide either to proceed to elections in order to end the post-revolutionary rule of the military, or slow down the electoral timetable and prioritize the writing of a new constitution. The SCAF decided to hold parliamentary elections before drafting a new constitution.

The demonstrations continued thereafter, pressuring the military finally to allow presidential elections to take place, with the results of the first round announced on May 28 and the results of the runoff election announced on June 24. The MB rallied behind the SCAF's plan to hold parliamentary elections prior to drafting a new constitution. MB run under the name Freedom and Justice Party (FJP).

Phase III: Rule of Mohammed Morsi. Islamist Mohammed Mursi narrowly won the parliamentary elections against the former general Ahmed Shafiq with 51.7% of the vote. However, the constitution imposed by the SCAF, left the Morsi with limited power. On August 12, 2012, Morsi revoked the interim declaration, thereby transferring power back to the president, including absolute legislative authority. The first stage of Morsi's rule was a struggle to assert power against military, culminating in removal of 5 key military figures (Comm-in-chief and 4 generals) in Aug 2012.

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 far-reaching powers. 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. Calls for rights and social justice led to decreasing popularity of Morsi. By Dec 23 2012 a referendum passing a new constitution promoting political islam and expanded military power passes despite secular boycott of the election. This was followed by alternating protests in Tahir Square, rotating between pro- and anti-Islamist movements.

Phase IV: Fall of Mohamed Morsi and return to military rule. On June 26, 2013, Morsi delivers a divisive address trying to defuse growing defiance to his rule. This leads

to larger protests in following days involving two sets of protesters, pro- versus anti-Morsi groups, in different locations across Egypt. On July 1 the military issues ultimatum to Morsi to call early election. On July 2 Morsi refuses to step down. On July 3 the Egyptian Military overthrows Morsi regime in coup, and anti-military intervention protests grow. On July 24 the military encourages pro-military intervention protests. From July 27 to mid August large demonstrations take place from both sides, leading to violent clashes with military for anti-intervention protesters.

E. 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:

$$\begin{aligned}
&\Phi(s_i', \mathbf{s}_{-i}, G, \boldsymbol{\psi}^{GI}(s_i', \mathbf{s}_{-i}, G)) - \Phi(\mathbf{s}, G, \boldsymbol{\psi}^{GI}(\mathbf{s}, G)) \\
&= \gamma_i(s_i' - s_i) + \theta(s_i' - s_i)\sum_{j=1}^{n} a_{ij}s_j + \rho(s_i' - s_i)\sum_{j\neq i}^{n} p_i - \kappa(s_i' - s_i) \\
&= \pi_i(s_i', \mathbf{s}_{-i}, G; \psi_i^{GI}(\mathbf{s}, G)) - \pi_i(\mathbf{s}, G; \psi_i^{GI}(\mathbf{s}, G))
\end{aligned}$$

and

$$\begin{aligned}
&\Phi(\mathbf{s}, G \pm ij, \boldsymbol{\psi}^{GI}(\mathbf{s}, G \pm ij)) - \Phi(\mathbf{s}, G, \boldsymbol{\psi}^{GI}(\mathbf{s}, G)) \\
&= \pm(\theta s_i s_j - \zeta_{ij}) \\
&= \pi_i(\mathbf{s}, G \pm ij; \psi_i^{GI}(\mathbf{s}, G)) - \pi_i(\mathbf{s}, G; \psi_i^{GI}(\mathbf{s}, G)) \\
&= \pi_j(\mathbf{s}, G \pm ij; \psi_j^{GI}(\mathbf{s}, G)) - \pi_j(\mathbf{s}, G; \psi_j^{GI}(\mathbf{s}, G))
\end{aligned}$$

which confirms (13)-(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 $\omega \in \Omega$ prevailing at t to a state $\omega' \in \Omega$ after some small time interval of length Δt. If $\omega' \neq \omega$, 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:

$$q(\boldsymbol{\omega}, \boldsymbol{\omega}') = \begin{cases} \chi \frac{e^{\eta\Phi(s_i, \mathbf{s}_{-i}, G)}}{e^{\eta\Phi(s_i, \mathbf{s}_{-i}, G)} + e^{\eta\Phi(s_i', \mathbf{s}_{-i}, G)}} & \text{if } \boldsymbol{\omega}' = (s_i', \mathbf{s}_{-i}, G) \text{ and } \boldsymbol{\omega} = (\mathbf{s}, G), \\ \lambda \frac{e^{\eta\Phi(\mathbf{s}, G+ij)}}{e^{\eta\Phi(\mathbf{s}, G+ij)} + e^{\eta\Phi(\mathbf{s}, G)}} & \text{if } \boldsymbol{\omega}' = (\mathbf{s}, G + ij) \text{ and } \boldsymbol{\omega} = (\mathbf{s}, G), \\ \lambda \frac{e^{\eta\Phi(\mathbf{s}, G-ij)}}{e^{\eta\Phi(\mathbf{s}, G-ij)} + e^{\eta\Phi(\mathbf{s}, G)}} & \text{if } \boldsymbol{\omega}' = (\mathbf{s}, G - ij) \text{ and } \boldsymbol{\omega} = (\mathbf{s}, G), \\ -\sum_{\boldsymbol{\omega}' \neq \boldsymbol{\omega}} q(\boldsymbol{\omega}, \boldsymbol{\omega}') & \text{if } \boldsymbol{\omega}' = \boldsymbol{\omega}, \\ 0 & \text{otherwise.} \end{cases} \tag{43}$$

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 $\mu^\eta : \Omega \to [0,1]$ 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. e.g. 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 Equation (43). This equation is satisfied when the probability distribution $\mu^\eta(\boldsymbol{\omega})$ satisfies the detailed balance condition [cf. e.g. Norris, 1998]

$$\mu^\eta(\boldsymbol{\omega}) q(\boldsymbol{\omega}, \boldsymbol{\omega}') = \mu^\eta(\boldsymbol{\omega}') q(\boldsymbol{\omega}', \boldsymbol{\omega}), \tag{44}$$

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

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

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

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

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

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

Hence, the probability measure $\mu^\eta(\boldsymbol{\omega})$ satisfies a detailed balance condition of Equation (44) and therefore is the stationary distribution of the Markov chain with transition rates $q(\boldsymbol{\omega}, \boldsymbol{\omega}')$. □

Before proceeding with the proof of Proposition 7 we state three useful lemmas that will be needed later. In these lemmas we assume that $\zeta_1 = \zeta_2 = \zeta$.

Lemma 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*

$$\mu^\eta(\mathbf{s}) = \frac{1}{\mathscr{Z}^\eta} e^{\eta \mathscr{H}^\eta(\mathbf{s})}, \tag{45}$$

where we have denoted by

$$\mathscr{H}^\eta(\mathbf{s}) \equiv \sum_{i=1}^{n} \left(\left(\gamma_i - \kappa + \frac{\rho}{2} \sum_{j \neq i}^{n} s_j \right) s_i + \sum_{j=i+1}^{n} \left(\frac{1}{\eta} \ln \left(1 + e^{\eta(\theta s_i s_j - \zeta)} \right) \right) \right), \tag{46}$$

and the normalizing constant is given by

$$\mathscr{Z}^{\eta} = \sum_{\mathbf{s}\in\{-1,+1\}^n} e^{\eta\mathscr{H}^{\eta}(\mathbf{s})}. \tag{47}$$

Proof of Lemma 1. We first compute the *partition function* [cf. e.g. Grimmett, 2010; Wainwright and Jordan, 2008], which appears as the denominator in Equation (15), explicitly. We have that[26]

$$\begin{aligned}
\mathscr{Z}^{\eta} &\equiv \sum_{G\in\mathcal{G}^n}\sum_{\mathbf{s}\in\{-1,+1\}^n} e^{\eta\Phi(\mathbf{s},G)} \\
&= \sum_{\mathbf{s}\in\{-1,+1\}^n}\sum_{G\in\mathcal{G}^n} e^{\eta\left(\sum_{i=1}^n(\gamma_i-\kappa)s_i+\frac{\rho}{2}\sum_{i=1}^n\sum_{j\neq i}^n s_is_j+\sum_{i=1}^n\sum_{j=i+1}^n a_{ij}(\theta s_is_j-\zeta)\right)} \\
&= \sum_{\mathbf{s}\in\{-1,+1\}^n} e^{\eta\sum_{i=1}^n\left(\gamma_i-\kappa+\frac{\rho}{2}\sum_{j\neq i}^n s_j\right)s_i}\sum_{G\in\mathcal{G}^n} e^{\eta\sum_{i=1}^n\sum_{j=i+1}^n a_{ij}(\theta s_is_j-\zeta)} \\
&= \sum_{\mathbf{s}\in\{-1,+1\}^n} e^{\eta\sum_{i=1}^n\left(\gamma_i-\kappa+\frac{\rho}{2}\sum_{j\neq i}^n s_j\right)s_i}\prod_{i=1}^n\prod_{j=i+1}^n\left(1+e^{\eta(\theta s_is_j-\zeta)}\right),
\end{aligned} \tag{48}$$

where we have used the fact that

$$\sum_{G\in\mathcal{G}^n} e^{\sum_{i<j}^n a_{ij}\sigma_{ij}} = \prod_{i=1}^n\prod_{j=i+1}^n\left(1+e^{\sigma_{ij}}\right), \tag{49}$$

for any real and symmetric $\sigma_{ij}=\sigma_{ji}$. Introducing the *Hamiltonian* [cf. e.g. Grimmett, 2010]

$$\mathscr{H}^{\eta}(\mathbf{s}) \equiv \sum_{i=1}^n\left(\left(\gamma_i-\kappa+\frac{\rho}{2}\sum_{j\neq i}^n s_j\right)s_i + \sum_{j=i+1}^n\left(\frac{1}{\eta}\ln\left(1+e^{\eta(\theta s_is_j-\zeta)}\right)\right)\right), \tag{50}$$

we can write the partition function as follows

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

With the Hamiltonian we can write the marginal distribution as follows

$$\begin{aligned}
\mu^{\eta}(\mathbf{s}) &= \frac{1}{\mathscr{Z}^{\eta}}\sum_{G\in\mathcal{G}^n} e^{\eta\Phi(\mathbf{s},G)} \\
&= \frac{1}{\mathscr{Z}_n^{\eta}} e^{\eta\sum_{i=1}^n\left(\gamma_i-\kappa+\frac{\rho}{2}\sum_{j\neq i}^n s_j\right)s_i}\prod_{i=1}^n\prod_{j=i+1}^n\left(1+e^{\eta(\theta s_is_j-\zeta)}\right) \\
&= \frac{1}{\mathscr{Z}^{\eta}} e^{\eta\mathscr{H}^{\eta}(\mathbf{s})},
\end{aligned} \tag{51}$$

[26] Note that when the network is exogenous (i.e. when $\xi=\lambda=0$) then in the limit of $\eta\to\infty$ the sum over all configurations $\mathbf{s}\in\{-1,+1\}^n$ is equivalent to summing over all max cuts of the underlying graph, whose enumeration is an NP hard problem (cf. A. Montanari, "Inference in Graphical Models", Stanford University, lecture notes, 2012).

where $\mathscr{H}^{\eta}(\mathbf{s})$ has been defined in Equation (50). □

Lemma 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*

$$\mu^{\eta}(G|\mathbf{s}) = \prod_{i=1}^{n} \prod_{j=i+1}^{n} p_{ij}(s_i, s_j)^{a_{ij}} \left(1 - p_{ij}(s_i, s_j)\right)^{1-a_{ij}},$$

where

$$p_{ij}(s_i, s_j) = \frac{e^{\eta(\theta s_i s_j - \zeta)}}{1 + e^{\eta(\theta s_i s_j - \zeta)}}. \tag{52}$$

Proof of Lemma 2. With the marginal distribution from Equation (45) we can write the conditional distribution as

$$\begin{aligned}
\mu^{\eta}(G|\mathbf{s}) = \frac{\mu^{\eta}(\mathbf{s}, G)}{\mu^{\eta}(\mathbf{s})} &= \frac{e^{\eta\left(\sum_{i=1}^{n}\left(\gamma_i - \kappa + \frac{\rho}{2}\sum_{j\neq i}^{n} s_j\right)s_i + \frac{\theta}{2}\sum_{i=1}^{n}\sum_{j=1}^{n} a_{ij} s_i s_j - m\zeta\right)}}{e^{\eta\sum_{i=1}^{n}\left(\gamma_i - \kappa + \frac{\rho}{2}\sum_{j\neq i}^{n} s_j\right)s_i} \prod_{i=1}^{n}\prod_{j=i+1}^{n}\left(1 + e^{\eta(\theta s_i s_j - \zeta)}\right)} \\
&= \frac{e^{\eta\sum_{i<j}^{n} a_{ij}(\theta s_i s_j - \zeta)}}{\prod_{i=1}^{n}\prod_{j=i+1}^{n}\left(1 + e^{\eta(\theta s_i s_j - \zeta)}\right)} \\
&= \prod_{i<j} \frac{e^{\eta a_{ij}(\theta s_i s_j - \zeta)}}{1 + e^{\eta(\theta s_i s_j - \zeta)}} \\
&= \prod_{i<j} \left(\frac{e^{\eta(\theta s_i s_j - \zeta)}}{1 + e^{\eta(\theta s_i s_j - \zeta)}}\right)^{a_{ij}} \left(1 - \frac{e^{\eta(\theta s_i s_j - \zeta)}}{1 + e^{\eta(\theta s_i s_j - \zeta)}}\right)^{1-a_{ij}} \\
&= \prod_{i<j} p_{ij}(s_i, s_j)^{a_{ij}} \left(1 - p_{ij}(s_i, s_j)\right)^{1-a_{ij}}.
\end{aligned} \tag{53}$$

Hence, conditional on the action choices $\mathbf{s}$, we obtain the likelihood of an *inhomogeneous random graph* with link probability [Bollobas et al., 2007]

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

□

In the following we provide an explicit computation of the partition function introduced in Equation (48).

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

$$\begin{aligned}
\mathscr{Z}^{\eta} = \sum_{k=0}^{n} \sum_{j=0}^{\min\{k, n_+\}} &\binom{n_+}{j}\binom{n - n_+}{k - j} e^{\eta(2k-n)} \\
&\times e^{\eta\left(\rho\left(2l(k,j) - \binom{n}{2}\right)\right) - \kappa(n - 2(n_+ + k - 2j))} \left(1 + e^{\eta(\theta-\zeta)}\right)^{l(k,j)} \left(1 + e^{-\eta(\theta+\zeta)}\right)^{\binom{n}{2} - l(k,j)},
\end{aligned} \tag{54}$$

where

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

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

Note that, while the evaluation of the partition function in Equation (48) requires the computation of a sum with 2^n terms, the partition function in Equation (54) requires the evaluation of only $\frac{1}{2}(n_+ + 1)(2(n+1) - n_+) = O(n)$ terms. With Equation (54) the marginal distribution $\mu^\eta(\mathbf{s})$ in Equation (45) can then be efficiently computed.

Proof of Lemma 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 $0 \le n_+ \le n$. 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

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

Then we can write

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

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

Proof of Proposition 7. With the partition function in Lemma 3 we can compute the expected number of links, m, as follows

$$
\mathbb{E}^{\eta}(m) = \sum_{G\in\mathcal{G}^n}\sum_{\mathbf{s}\in\{-1,+1\}^n} m\mu^{\eta}(\mathbf{s},G) = \frac{1}{\mathscr{Z}^{\eta}}\sum_{G\in\mathcal{G}^n}\sum_{\mathbf{s}\in\{-1,+1\}^n} \underbrace{m e^{\eta\Phi(\mathbf{s},G)}}_{-\frac{1}{\eta}\frac{\partial}{\partial\zeta}e^{\eta\Phi(\mathbf{s},G)}} = -\frac{1}{\eta}\frac{1}{\mathscr{Z}^{\eta}}\frac{\partial\mathscr{Z}^{\eta}}{\partial\zeta}. \tag{55}
$$

With Equations (48) and (55) we then can compute the expected number of links as

$$
\begin{aligned}
\mathbb{E}^{\eta}(m) = \frac{1}{\mathscr{Z}^{\eta}} \sum_{k=0}^{n} \sum_{j=0}^{\min\{k,n_+\}} &\binom{n_+}{j}\binom{n-n_+}{k-j} e^{\eta(2k-n)} e^{\eta\left(\rho\left(2l(k,j)-\binom{n}{2}\right)\right)-\kappa(n-2(n_+ +k-2j))} \\
&\times \left(1+e^{\eta(\theta-\zeta)}\right)^{l(k,j)} \left(1+e^{-\eta(\theta+\zeta)}\right)^{\binom{n}{2}-l(k,j)} \left(\frac{l(k,j)}{1+e^{-\eta(\theta-\zeta)}} + \frac{\binom{n}{2}-l(k,j)}{1+e^{\eta(\theta+\zeta)}}\right),
\end{aligned} \tag{56}
$$

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

For $\theta = \rho = 0$ Equation (56) simplifies to

$$\begin{aligned}
\mathbb{E}^{\eta}(m) &= \frac{1}{\mathscr{Z}^{\eta}} \sum_{k=0}^{n} \sum_{j=0}^{\min\{k,n_+\}} \binom{n_+}{j}\binom{n-n_+}{k-j} e^{\eta(2k-n)} \\
&\quad \times \left(1+e^{-\eta\zeta}\right)^{l(k,j)} \left(1+e^{-\eta\zeta}\right)^{\binom{n}{2}-l(k,j)} \left(\frac{l(k,j)}{1+e^{\eta\zeta}} + \frac{\binom{n}{2}-l(k,j)}{1+e^{\eta\zeta}}\right) \\
&= \frac{1}{\mathscr{Z}^{\eta}} \sum_{k=0}^{n} \sum_{j=0}^{\min\{k,n_+\}} \binom{n_+}{j}\binom{n-n_+}{k-j}\binom{n}{2} e^{\eta(2k-n)} \left(1+e^{-\eta\zeta}\right)^{\binom{n}{2}} \frac{1}{1+e^{\eta\zeta}} \\
&= \frac{1}{\mathscr{Z}^{\eta}} \binom{n}{2} \left(1+e^{-\eta\zeta}\right)^{\binom{n}{2}} \frac{1}{1+e^{\eta\zeta}} \sum_{k=0}^{n} \sum_{j=0}^{\min\{k,n_+\}} \binom{n_+}{j}\binom{n-n_+}{k-j} e^{\eta(2k-n)} \\
&= \frac{1}{\mathscr{Z}^{\eta}} \frac{e^{-\eta n}}{\pi(1+e^{\eta\zeta})} \binom{n}{2} \left(1+e^{-\eta\zeta}\right)^{\binom{n}{2}} \\
&\quad \times \left(\pi\left(1+e^{2\eta}\right)^{n} - e^{2(n+1)\eta}\sin(n\pi)\Gamma(n+1)\,{}_2F_1\left(1,1;n+2;-e^{2\eta}\right)\right).
\end{aligned}$$

Proof of Proposition 8. 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

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

□

Proof of Propositions 3, 4 and 11. The potential function is given by

$$\begin{aligned}
\Phi(\mathbf{s}, G) &= \sum_{i=1}^{n} \gamma_i s_i + \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n} a_{ij}(\theta s_i s_j - \zeta_{ij}) + \frac{\rho}{2}\sum_{i=1}^{n}\sum_{j\neq i}^{n} s_i s_j - \kappa\sum_{i=1}^{n} s_i \\
&= \sum_{i=1}^{n}\left(\gamma_i + \frac{\rho}{2}\sum_{j\neq i}^{n} s_j - \kappa\right) s_i + \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n} a_{ij}(\theta s_i s_j - \zeta_{ij}).
\end{aligned}$$

With the linking cost in Equation (62) the potential function can be written as

$$\Phi(\mathbf{s}, G) = \sum_{i=1}^{n} \left(\gamma_i + \frac{\rho}{2} \sum_{j \neq i}^{n} s_j - \kappa \right) s_i + \frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{n} a_{ij} \left(\theta s_i s_j - \zeta_1 + \frac{\zeta_1 - \zeta_2}{2} (1 - \gamma_i \gamma_j) \right). \quad (57)$$

Note that only the last term in Equation (63) depends on the network G (through the entries of the elements a_{ij} of its adjacency matrix A). 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, $\overline{K}_n$. Summarizing, 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.

Consider first the case of $\theta < \zeta_1$. Then the stochastically stable network is empty, $\overline{K}_n$ and the potential function simplifies to

$$\Phi(\mathbf{s}, \overline{K}_n) = \sum_{i=1}^{n} s_i \gamma_i + \frac{\rho}{2} \sum_{i=1}^{n} \sum_{j \neq i}^{n} s_i s_j - \kappa \sum_{i=1}^{n} s_i.$$

Observe that the first term is maximized if $s_i = \gamma_i$, the second term is maximized if $s_i = s_j$ for all i and j, while the last term is maximized if $s_i = -1$ for all i. The second and third terms are jointly maximized if all agents choose $s_i = -1$. We thus need to consider only three possible cases for the action profiles. All agents i choose $s_i = -1$, all agents i choose $s_i = -1$ or all agent choose $s_i = \gamma_i$. We can ignore configurations different from the above in which some agent i with $\gamma_i = +1$ would choose an action $s_i = -1$. This is because if the potential would be higher in such a configuration, then it would be even higher in the case where all agents choose $s_i = -1$.

In the case of all agents choosing the action $s_i = -1$ the potential is given by

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

Conversely, in the case of all agents choosing the action $s_i = +1$ the potential is given by

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

In the case of all agents choosing the action $s_i = \gamma_i$ the potential is given by

$$\begin{aligned} \Phi(\boldsymbol{\gamma}, \overline{K}_n) &= n + \frac{\rho}{2} \left(n_+((n_+ - 1) - (n - n_+)) + (n - n_+)((n - n_+ - 1) - n_+) \right) \\ &\quad - \kappa(n_+ - (n - n_+)) \\ &= n + \frac{\rho}{2} \left((n - 2n_+)^2 - n \right) - \kappa(2n_+ - n). \end{aligned}$$

We then have that

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

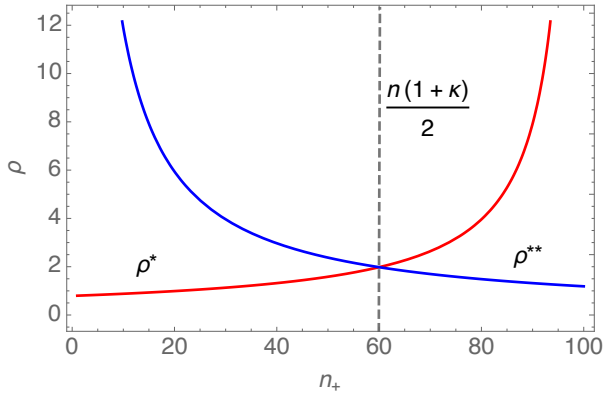


Figure 14: The two thresholds ρ^* and ρ^{**} as a function of n_+. The dashed line indicates where they coincide, i.e. $\rho^* = \rho^{**}$, which happens for $\nu_+ = \frac{n_+}{n} = \frac{1+\kappa}{2}$.

which is increasing in ρ. Solving $\Phi((-1,\ldots,-1),\overline{K}_n) = \Phi(\boldsymbol{\gamma},\overline{K}_n)$ for ρ yields the threshold

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

For $\rho > \rho^*$ the stochastically stable state will be the empty network $\overline{K}_n$ in which all agents choose the action $s_i = -1$, while for $\rho < \rho^*$ all agents choose the action $s_i = \gamma_i$. Similarly, solving for κ yields the threshold

$$\kappa^{**} = 1 - \rho(n-n_+) \xrightarrow[n\to\infty]{} -\infty \text{ (if } \rho > 0)$$

Moreover,

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

which is positive if ρ is larger than

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

Similarly, solving for κ yields the threshold

$$\kappa^{***} = \rho n_+ - 1 \xrightarrow[n\to\infty]{} +\infty.$$

Further,

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

which is positive if κ is smaller than

$$\kappa^* = \frac{2n_+}{n} - 1,$$

Note that κ^* is positive only if $n_+ > n/2$ and that $\kappa^* > \kappa^{***}$ only if $\rho < 2/n \xrightarrow[n\to\infty]{} 0$.

We next assume that $n_+ < \frac{n}{2}$ and $\theta > \zeta_1$. First, consider two cliques, K_{n_+} and K_{n-n_+} of sizes n_+ and $n-n_+$, respectively, where the agents in K_{n_+} choose $s_i = \gamma_i = +1$, and the agents

in K_{n-n_+} choose $s_i = \gamma_i = -1$. The potential function is then given by

$$\begin{aligned}
&\Phi(\boldsymbol{\gamma}, K_{n_+} \cup K_{n-n_+}) \\
&= n + \frac{1}{2}(n_+(n_+-1) + (n-n_+)(n-n_+-1))(\theta-\zeta_1) \\
&+ \frac{\rho}{2}\left(n_+((n_+-1)-(n-n_+)) + (n-n_+)((n-n_+-1)-n_+)\right) - \kappa(n_+-(n-n_+)) \\
&= n + \frac{1}{2}(n(n-1) - 2n_+(n-n_+))(\theta-\zeta_1) \\
&+ \frac{\rho}{2}\left(n_+(2n_+-n-1) + (n-n_+)(n-1-2n_+)\right) - \kappa(2n_+-n) \\
&= n + \frac{1}{2}(n(n-1) - 2n_+(n-n_+))(\theta-\zeta_1) + \frac{\rho}{2}\left((n-2n_+)^2 - n\right) - \kappa(2n_+-n).
\end{aligned}$$

We next consider the potential in a union of cliques $K_{n_+-k} \cup K_{n-n_++k}$, obtained from disconnecting k nodes j from the clique K_{n_+} and connecting them to all nodes in the clique K_{n-n_+}, while choosing the action $s_j = -1$ with $\gamma_j = +1$, with $k = 0, \ldots, n_+$. This is illustrated in the left panel in Figure 15 for $k = 1$. The corresponding potential function is given by

$$\begin{aligned}
&\Phi(\mathbf{s}', K_{n_+-k} \cup K_{n-n_++k}) \\
&= ((n_+-k) - k + n - n_+) - \kappa((n_+-k) - (n-n_++k)) \\
&+ \frac{1}{2}((n_+-k)(n_+-k-1) + (n-n_+)(n-n_+-1) + k(k-1))(\theta-\zeta_1) + k(n-n_+)(\theta-\zeta_2) \\
&+ \frac{\rho}{2}\left((n_+-k)((n_+-k-1)-(n-n_++k)) + (n-n_++k)((n-n_++k-1)-(n_+-k))\right) \\
&= (n-2k) - \kappa(2(n_+-k)-n) + \frac{\rho}{2}((n-2(n_+-k))^2 - n) \\
&+ \frac{1}{2}(2\left(k^2 - n_+(k+n) + n_+^2\right) + n(n-1))(\theta-\zeta_1) + \frac{1}{2}k(k-1)(\theta-\zeta_2).
\end{aligned}$$

It then follows that

$$\begin{aligned}
&\Phi(\mathbf{s}', K_{n_+-k} \cup K_{n-n_++k}) - \Phi(\boldsymbol{\gamma}, K_{n_+} \cup K_{n-n_+}) \\
&\quad = \frac{k}{2}\left(2\zeta_1 n_+ + \zeta_2 - \theta(1+2n_+) - 4(1-\kappa) - k(2\zeta_1 + \zeta_2 - 3\theta) + 4\rho(k+n-2n_+)\right). \quad (58)
\end{aligned}$$

Note that this is (under some regularity conditions) a convex function of k (see Figure 15).[27] A convex function attains its maximum at the boundaries, which is either the union of two cliques, $K_{n_+} \cup K_{n-n_+}$, or the complete graph, K_n, in which all agents choose $s_i = -1$. In the latter the potential is given by

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

We then have that

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

which is increasing in θ. Solving $\Phi((-1,\ldots,-1), K_n) = \Phi(\boldsymbol{\gamma}, K_{n_+} \cup K_{n-n_+})$ for θ yields the

[27] Denote by $\Delta\Phi(k) \equiv \Phi(\mathbf{s}', K_{n_+-k} \cup K_{n-n_++k}) - \Phi(\boldsymbol{\gamma}, K_{n_+} \cup K_{n-n_+})$. Then $\frac{d^2\Delta\Phi(k)}{dk^2} = 3\theta + 4\rho - 2\zeta_1 - \zeta_2 > 0$ if $\theta > \frac{2\zeta_1+\zeta_2-4\rho}{3}$. Further, note that if $\Delta\Phi(k)$ is convex, then also $\Phi(\mathbf{s}', K_{n_+-k} \cup K_{n-n_++k})$ is convex.

threshold

$$\theta^* = \zeta_2 + \frac{2(1-\kappa)}{n-n_+} - 2\rho \xrightarrow[n\to\infty]{} \zeta_2 - 2\rho$$

Further, consider the union of cliques $K_{n_+-k} \cup K_{n-n_++k}$ in which all agents i choose action $s_i = -1$. The potential is given by

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

We then have that

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

which is negative for

$$\rho < \rho^* = \frac{1-\kappa}{n-n_+} \xrightarrow[n\to\infty]{} 0$$

Similarly,

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

which is increasing in θ. We have that $\Phi((-1,\dots,-1), K_n) = \Phi((-1,\dots,-1), K_{n_+} \cup K_{n-n_+})$ for $\theta = \zeta_2$. Hence, for $\theta > \zeta_2 > \zeta_1$ the stochastically stable state is the complete graph K_n in which all agents choose the action $s_i = -1$, while for $\zeta_1 < \theta < \zeta_2$ it is the union of two cliques, $K_{n_+} \cup K_{n-n_+}$, in which all agents choose the action $s_i = \gamma_i$ if $\rho < \rho^*$ or all agents choosing action $s_i = -1$ if $\rho > \rho^*$.

Next we consider the case of $n_+ > \frac{n}{2}$ and $\theta > \zeta_1$. Consider the complete graph K_n in which all agents choose $s_i = +1$. Then

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

Further, we have that

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

which is increasing in θ. Solving $\Phi((+1,\dots,+1), K_n) = \Phi(\boldsymbol{\gamma}, K_{n_+} \cup K_{n-n_+})$ for θ yields the threshold

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

Next, consider the union of cliques $K_{n_+-k} \cup K_{n-n_++k}$ in which all agents i choose action $s_i = +1$. The potential is given by

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

We then have that

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

which is negative for

$$\rho<\rho^{**}=\frac{1+\kappa}{n_+}\xrightarrow[n\to\infty]{}0.$$

Similarly,

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

which is positive for $\theta>\zeta_2$, negative for $\theta<\zeta_2$ and increasing in θ. Further, we have that $\Phi((+1,\ldots,+1),K_n)=\Phi((+1,\ldots,+1),K_{n_+}\cup K_{n-n_+})$ for $\theta=\zeta_2$. Hence, for $\theta>\zeta_2>\zeta_1$ the stochastically stable state is the complete graph K_n in which all agents choose the action $s_i=+1$, while for $\zeta_1<\theta<\zeta_2$ it is the union of two cliques, $K_{n_+}\cup K_{n-n_+}$, in which all agents choose the action $s_i=\gamma_i$ if $\rho<\rho^{**}$ or all agents choosing action $s_i=+1$ if $\rho>\rho^{**}$. If $\theta<\zeta_1$ then we obtain the empty network, $\overline{K}_n$.

Next, note that

$$\Phi((-1,\ldots,-1),K_n)-\Phi((+1,\ldots,+1),K_n)=2(n(\kappa+1)-2n_+),\tag{59}$$

which is increasing in κ. For $n_+<\frac{n}{2}$ Equation (64) is strictly positive for any value of κ. In contrast, for $n_+>\frac{n}{2}$ we have that

$$\Phi((-1,\ldots,-1),K_n)<\Phi((+1,\ldots,+1),K_n)$$

if

$$\kappa<\kappa^*=\frac{2n_+}{n}-1,$$

and κ^* being positive only if $n_+>n/2$. Moreover,

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

which is positive for $\kappa>\kappa^*$ and negative for $\kappa<\kappa^*$. With the above discussion we have covered all possible partitions of agents into two cliques (including the complete and the empty graphs), and the actions they can choose. As these are the candidate potential maximizers, we have therefore identified the networks and action profiles that maximize the potential. This concludes the proof.

□

Proof of Propositions 5, 6 and 12. 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 holds if the potential is maximized for such $(\mathbf{s},G)$ given the beliefs $\mathbf{p}$. Conversely, the belief update Equation (**??**) must be stationary given $(\mathbf{s},G)$, that is, $p_i=f_i(\mathbf{s},\mathbf{p},G)=\varphi\frac{1}{d_i}\sum_{j=1}^n a_{ij}s_j+(1-\varphi)\frac{1}{d_i}\sum_{j=1}^n a_{ij}p_j$ 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

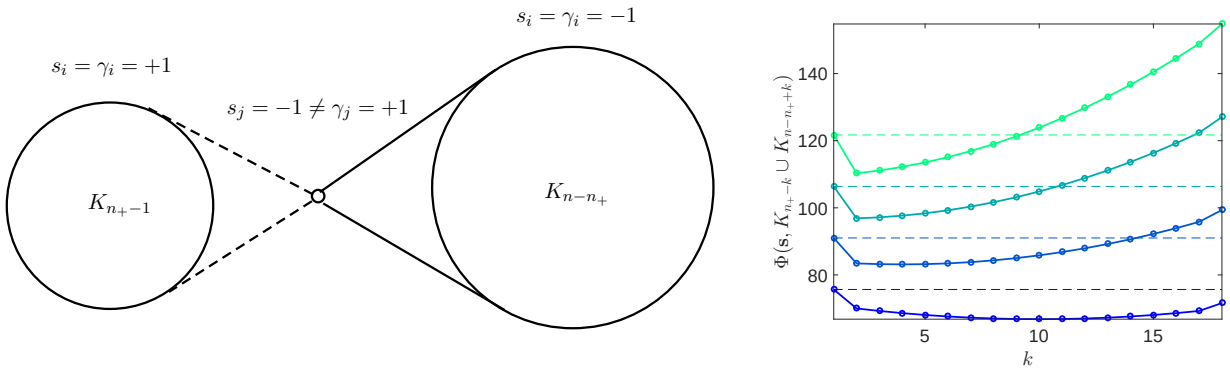


Figure 15: (Left panel) Illustration of two cliques, K_{n+} and K_{n-n_+} and the relocation of one node j from K_{n+} to K_{n-n_+}. (Right panel) The resulting potential for relocating node j from the clique K_{n_+} to the clique K_{n-n_+} for $\theta \in \{0.05, 0.075, 0.1, 0.125\}$, $n_+ = 17$, $n = 50$, $\rho = 0$ and $\zeta_1 = \zeta_2 = 0.01$. The threshold is given by $\theta^* = 0.061$. For small values of $\theta < \theta^*$ the union of cliques $K_{n_+} \cup K_{n-n_+}$ ($j = 0$) has the highest potential, while for increasing values of θ the potential is highest for the complete graph K_n ($j = n_+ = 17$). We also see that the potential in a union of cliques $K_{n_+-k} \cup K_{n-n_++k}$ for $k = 1, \ldots, n_+ - 1$ is always smaller than the potential in the complete graph K_n or in the union of cliques $K_{n_+} \cup K_{n-n_+}$.

function can be written as

$$\begin{aligned}\tilde{\Phi}(\mathbf{s}, G, \mathbf{p}) &= \sum_{i=1}^{n} \gamma_i s_i + \frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{n} a_{ij}(\theta s_i s_j - \zeta_{ij}) + \rho(n-1) \sum_{i=1}^{n} p_i s_i - \kappa \sum_{i=1}^{n} s_i \\ &= \tilde{\boldsymbol{\gamma}}^\top \mathbf{s} + \frac{\theta}{2} \mathbf{s}^\top \mathbf{A} \mathbf{s} - \frac{1}{2} \mathbf{u}^\top \boldsymbol{\zeta} \mathbf{u},\end{aligned}$$

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 $\zeta_{ij} < 1$. This implies that the stochastically stable network must be either complete, empty or composed of two cliques, where in each clique the agents choose the same actions.

From Equation (10) we know that the stationary beliefs satisfy the following relationship

$$\mathbf{p}(\boldsymbol{\omega}) = \varphi \left\{ \left[\mathbf{I} - (1-\varphi) \widehat{\mathbf{D}}^{-1}(G) \widehat{\mathbf{A}}(G) \right]^{-1} \mathbf{D}^{-1}(G) \mathbf{A}(G) \mathbf{s} \right\} \qquad (\boldsymbol{\omega} = (\mathbf{s}, G) \in \Omega).$$

Hence, we know that the absorbing state $(\mathbf{s}, G, \mathbf{p})$ must satisfy $s_i = \text{sign}(\gamma_i + \rho(n-1)p_i - \kappa)$ and $a_{ij} = 1$ iff $\text{sign}(s_i) = \text{sign}(s_j)$ in the case of $\zeta_{ij} < 1$, where the network must be either complete, empty or composed of two cliques, where in each clique the agents choose the same actions.

From the equation $s_i = \text{sign}(\gamma_i + \rho(n-1)s_i - \kappa)$ we see that for $\gamma_i = -1$ it must hold that

$$s_i = \begin{cases} +1, & \text{if } \rho > (1+\kappa)/(n-1) \xrightarrow[n\to\infty]{} 0, \\ -1, & \text{for any vales of } \rho \text{ and } \kappa. \end{cases} \tag{60}$$

Similarly, for $\gamma_i = +1$ it must hold that

$$s_i = \begin{cases} +1, & \text{if } \rho > (\kappa-1)/(n-1) \xrightarrow[n\to\infty]{} 0, \\ -1, & \text{if } \rho > (1-\kappa)/(n-1) \xrightarrow[n\to\infty]{} 0. \end{cases} \tag{61}$$

Further, from Equation (8) we know that the stationary beliefs must satisfy $p_i = \varphi \frac{1}{d_i} \sum_{j=1}^{n} a_{ij} s_j +$

$(1-\varphi)\frac{1}{d_i}\sum_{j=1}^{n} a_{ij}p_j$. In a network where all connected agents choose the same action and have the same beliefs, this simplifies to $p_i = \varphi s_i + (1-\varphi)p_i$, and this equation is satisfied for $p_i = s_i$.

When $p_i = s_i$ for all $i = 1, \ldots, n$ then the potential is given by

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

With the linking cost

$$\zeta_{ij} = \zeta_1 - \frac{\zeta_1 - \zeta_2}{2}(1 - \gamma_i\gamma_j) = \begin{cases} \zeta_1, & \text{if } \gamma_i = \gamma_j, \\ \zeta_2, & \text{if } \gamma_i \neq \gamma_j, \end{cases} \tag{62}$$

where $0 \leq \zeta_1 < \zeta_2$, the potential function can be written as

$$\tilde{\Phi}(\mathbf{s}, G) = \sum_{i=1}^{n} \left(\gamma_i + \rho(n-1)s_i - \kappa\right) s_i + \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n} a_{ij}\left(\theta s_i s_j - \zeta_1 + \frac{\zeta_1 - \zeta_2}{2}(1 - \gamma_i\gamma_j)\right). \tag{63}$$

Note that only the last term in Equation (63) 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, $\overline{K}_n$. Summarizing, 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.

Consider first the case of $\theta < \zeta_1$. Then the stochastically stable network is empty, $\overline{K}_n$ and the potential function simplifies to

$$\tilde{\Phi}(\mathbf{s}, \overline{K}_n) = \sum_{i=1}^{n} s_i\gamma_i + \rho(n-1)\sum_{i=1}^{n} s_i^2 - \kappa\sum_{i=1}^{n} s_i.$$

Observe that the first term is maximized if $s_i = \gamma_i$, while the last term is maximized if $s_i = -1$ for all i. The second and third terms are jointly maximized if all agents choose $s_i = -1$. We thus need to consider only two possible cases for the action profiles. All agents i choose $s_i = -1$ or all agent choose $s_i = \gamma_i$. We can ignore configurations different from the above in which some agent i with $\gamma_i = +1$ would choose an action $s_i = -1$. This is because if the potential would be higher in such a configuration, then it would be even higher in the case where all agents choose $s_i = -1$.

In the case of all agents choosing the action $s_i = -1$ the potential is given by

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

Conversely, in the case of all agents choosing the action $s_i = +1$ the potential is given by

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

In the case of all agents choosing the action $s_i = \gamma_i$ the potential is given by

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

We then have that

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

which is positive for $\kappa > 1$ and negative for $\kappa < 1$. Moreover,

$$\tilde{\Phi}((+1, \ldots, +1), \overline{K}_n) - \tilde{\Phi}(\boldsymbol{\gamma}, \overline{K}_n) = -2(n - n_+)(1 + \kappa) < 0,$$

which is negative for all parameter choices. Further, we have that

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

which is positive if

$$\kappa < \kappa^* = \frac{2n_+}{n} - 1 \leq 1,$$

where $\kappa^* = 1$ if $n_+ = n$ and $\kappa^* < 0$ if $n_+ < n/2$. Thus, for $\kappa > 1$ the stochastically stable state will be the empty network $\overline{K}_n$ in which all agents choose the action $s_i = -1$, while for $\kappa < 1$ all agents choose the action $s_i = \gamma_i$. Note that in the case that all agents choose action $s_i = -1$, the condition in Equation (61) is trivially satisfied with $\kappa > 1$ as $\rho \geq 0$. The same holds for the for the case that the agents choose their idiosyncratic preferences as actions when $\kappa < 1$.

We next assume that $n_+ < \frac{n}{2}$ and $\theta > \zeta_1$. First, consider two cliques, K_{n_+} and K_{n-n_+} of sizes n_+ and $n - n_+$, respectively, where the agents in K_{n_+} choose $s_i = \gamma_i = +1$, and the agents in K_{n-n_+} choose $s_i = \gamma_i = -1$. The potential function is then given by

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

In the complete graph, K_n, in which all agents choose $s_i = -1$, the potential is given by

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

We then have that

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

which is increasing in θ. Solving $\tilde{\Phi}((-1, \ldots, -1), K_n) = \tilde{\Phi}(\boldsymbol{\gamma}, K_{n_+} \cup K_{n-n_+})$ for θ yields the threshold

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

We therefore find that $\tilde{\theta}^* < \zeta_2$ if $\kappa > 1$ and $\tilde{\theta}^* > \zeta_2$ if $\kappa < 1$.

Further, consider the union of cliques $K_{n_+ - k} \cup K_{n - n_+ + k}$ in which all agents i choose action $s_i = -1$. The potential is given by

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

We then have that

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

which is negative for $\kappa<1$ and positive for $\kappa>1$. Similarly,

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

which is positive for $\theta>\zeta_2$, negative for $\theta<\zeta_2$ and increasing in θ. In particular, we have that $\tilde{\Phi}((-1,\ldots,-1),K_n)=\tilde{\Phi}((-1,\ldots,-1),K_{n_+}\cup K_{n-n_+})$ for $\theta=\zeta_2$. Hence, for $\theta>\zeta_2>\zeta_1$ the stochastically stable state is the complete graph K_n in which all agents choose the action $s_i=-1$, while for $\zeta_1<\theta<\zeta_2$ it is the union of two cliques, $K_{n_+}\cup K_{n-n_+}$, in which all agents choose the action $s_i=\gamma_i$ if $\kappa<1$ or all agents choosing action $s_i=-1$ if $\kappa>1$. Note that in the case that all agents choose action $s_i=-1$, the condition in Equation (61) is trivially satisfied with $\kappa>1$ as $\rho\geq 0$. The same holds for the idiosyncratic preferences fragmented cliques when $\kappa<1$.

Next we consider the case of $n_+>\frac{n}{2}$ and $\theta>\zeta_1$. Consider the complete graph K_n in which all agents choose $s_i=+1$. Then

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

Further, we have that

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

which is increasing in θ. Solving $\Phi((+1,\ldots,+1),K_n)=\tilde{\Phi}(\boldsymbol{\gamma},K_{n_+}\cup K_{n-n_+})$ for θ yields the threshold

$$\tilde{\theta}^{**}=\zeta_2+\frac{2(\kappa+1)}{n_+}\xrightarrow[n_+\to\infty]{}\zeta_2,$$

with $\tilde{\theta}^{**}>\zeta_2$.

Next, consider the union of cliques $K_{n_+}\cup K_{n-n_+}$ in which all agents i choose action $s_i=+1$. The potential is given by

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

We then have that

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

which is negative for all parameter values. Similarly,

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

which is positive for $\theta>\zeta_2$, negative for $\theta<\zeta_2$ and increasing in θ. In particular, we have that $\tilde{\Phi}((+1,\ldots,+1),K_n)=\tilde{\Phi}((+1,\ldots,+1),K_{n_+}\cup K_{n-n_+})$ for $\theta=\zeta_2$. Hence, for $\theta>\zeta_2>\zeta_1$ the stochastically stable state is the complete graph K_n in which all agents choose the action

$s_i = +1$, while for $\zeta_1 < \theta < \zeta_2$ it is the union of two cliques, $K_{n_+} \cup K_{n-n_+}$, in which all agents choose the action $s_i = \gamma_i$.

Finally, note that

$$\tilde{\Phi}((-1,\dots,-1),K_n) - \tilde{\Phi}((+1,\dots,+1),K_n) = 2(n(\kappa+1) - 2n_+), \tag{64}$$

which is increasing in κ. For $n_+ < \frac{n}{2}$ Equation (64) is strictly positive for any value of κ. In contrast, for $n_+ > \frac{n}{2}$ we have that

$$\tilde{\Phi}((-1,\dots,-1),K_n) < \tilde{\Phi}((+1,\dots,+1),K_n)$$

if

$$\kappa < \kappa^* = \frac{2n_+}{n} - 1 \leq 1,$$

where $\kappa^* = 1$ if $n_+ = n$. Moreover,

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

which is positive for $\kappa > \kappa^*$ and negative for $\kappa < \kappa^*$.

With the above discussion we have covered all possible partitions of agents into two cliques (including the complete and the empty graphs), and the actions they can choose. As these are the candidate potential maximizers, we have therefore identified the networks and action profiles that maximize the potential. This concludes the proof. □

Before proceeding with the proof of Proposition 9 we state the following lemma which will be useful later.

Lemma 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 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

$$
\begin{aligned}
f(\mathbf{s}) &= \prod_{i=1}^{n_+-1} \left(\prod_{j=i+1}^{n_+} g(s_i, s_j) \prod_{j=n_++1}^{n} g(s_i, s_j) \right) \prod_{j=n_++1}^{n} g(s_{n_+}, s_j) \prod_{i=n_++1}^{n} \prod_{j=i+1}^{n} g(s_i, s_j) \\
&= \prod_{i=1}^{n_+-1} \left(\prod_{j=i+1}^{n_+} g(+1,+1) \prod_{j=n_++1}^{n} g(+1,-1) \right) \prod_{j=n_++1}^{n} g(+1,-1) \prod_{i=n_++1}^{n} \prod_{j=i+1}^{n} g(-1,-1) \\
&= \prod_{i=1}^{n_+-1} g(+1,+1)^{n_+-i} g(+1,-1)^{n-n_+} g(+1,-1)^{n-n_+} \prod_{i=n_++1}^{n} g(-1,-1)^{n-i} \\
&= g(+1,-1)^{n-n_+} g(+1,-1)^{(n-n_+)(n_+-1)} \prod_{i=1}^{n_+-1} g(+1,+1)^{n_+-i} \prod_{i=n_++1}^{n} g(-1,-1)^{n-i} \\
&= g(+1,-1)^{n-n_+} g(+1,-1)^{(n-n_+)(n_+-1)} g(+1,+1)^{\frac{n_+(n_+-1)}{2}} g(+1,+1)^{\frac{(n-n_+)(n-n_+-1)}{2}} \\
&= g(+1,-1)^{n_+(n-n_+)} g(+1,+1)^{\frac{n(n-1)-2n_+(n-n_+)}{2}} \\
&= \left(1 + e^{-\eta(\theta+\zeta)}\right)^{n_+(n-n_+)} \left(1 + e^{\eta(\theta-\zeta)}\right)^{\frac{n(n-1)-2n_+(n-n_+)}{2}}.
\end{aligned}
$$

This concludes the proof. □

Proof of Proposition 9. Assume that $\zeta_1 = \zeta_2 = \zeta$. Then, in the belief-based model, the quasi partition function is given by

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

where we have denoted by

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

The expected number of links is given by

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

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 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 $n_+(\mathbf{s}) = \#(\{s_i = 1 : i = 1, \dots, n\})$. Morover one can show that

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

and we can write

$$\mathbb{E}^\eta(m) = \frac{1}{\mathscr{Z}^\eta} \sum_{\mathbf{s}\in\{-1,+1\}^n} e^{\eta\langle\tilde{\gamma},\mathbf{s}\rangle} h^\eta(\mathbf{s}),$$

where

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

Finally, stationary beliefs are given by Equation (9) so that we can write them as a function of the actions and network as $\mathbf{p} = \varphi\left\{\left[\mathbf{I} - (1-\varphi)\widehat{\mathbf{D}}^{-1}\widehat{\mathbf{A}}\right]^{-1}\mathbf{D}^{-1}\mathbf{A}\mathbf{s}\right\}$. Moreover, from Lemma 2 we know that the linking probability conditional on actions $\mathbf{s}$ is given by

$$p_{ij} = \frac{e^{\eta(\theta s_i s_j - \zeta)}}{1 + e^{\eta(\theta s_i s_j - \zeta)}}, \tag{65}$$

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. □

Proof of Proposition 10. 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

$$\begin{aligned}\mathbb{E}^\eta(\bar{s}) &= \sum_{\mathbf{s}\in\{-1,+1\}^n} \bar{s}\mu^\eta(\mathbf{s}) \\ &= \frac{1}{\mathscr{Z}^\eta} \sum_{\mathbf{s}\in\{-1,+1\}^n} \frac{1}{n}\langle\mathbf{u},\mathbf{s}\rangle e^{\eta\sum_{i=1}^n \tilde{\gamma}_i s_i} \prod_{i=1}^n \prod_{j=i+1}^n \left(1 + e^{\eta(\theta s_i s_j - \zeta)}\right)\end{aligned}$$

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 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 $n_+(\mathbf{s}) = \#(\{s_i = 1 : i = 1, \dots, n\})$, and hence

$$\mathbb{E}^\eta(\bar{s}) = \frac{1}{\mathscr{Z}^\eta}\frac{1}{n} \sum_{\mathbf{s}\in\{-1,+1\}^n} \langle\mathbf{u},\mathbf{s}\rangle e^{\eta\sum_{i=1}^n \tilde{\gamma}_i s_i} \left(1 + e^{-\eta(\theta+\zeta)}\right)^{\alpha(\mathbf{s})} \left(1 + e^{\eta(\theta-\zeta)}\right)^{\beta(\mathbf{s})}.$$

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 9. □