DISCUSSION PAPER SERIES

DP18093

# RIOT NETWORKS AND THE TULLOCK PARADOX: AN APPLICATION TO THE EGYPTIAN ARAB SPRING

Chih-Sheng Hsieh, Lachlan Deer, Michael Koenig and Fernando Vega-Redondo

PREVENTING CONFLICT: POLICIES FOR PEACE

CEPR

*ISSN 0265-8003*

# RIOT NETWORKS AND THE TULLOCK PARADOX: AN APPLICATION TO THE EGYPTIAN ARAB SPRING

*Chih-Sheng Hsieh, Lachlan Deer, Michael Koenig and Fernando Vega-Redondo*

Discussion Paper DP18093
Published 20 April 2023
Submitted 19 April 2023

Centre for Economic Policy Research
33 Great Sutton Street, London EC1V 0DX, UK
Tel: +44 (0)20 7183 8801
www.cepr.org

This Discussion Paper is issued under the auspices of the Centre's research programmes:

- Preventing Conflict: Policies for Peace

# RIOT NETWORKS AND THE TULLOCK PARADOX: AN APPLICATION TO THE EGYPTIAN ARAB SPRING


## Abstract

We study a dynamic model of collective action in which agents interact and learn through a co-evolving social network. We consider two alternative scenarios that differ on how agents form their expectations: while in a "benchmark" agents are assumed completely informed of the prevailing state, in the other context agents shape their expectations through a combination of local observation and social learning a la DeGroot. We completely characterize the long-run behavior of the system in both cases and show that only in the latter scenario (arguably the most realistic) there is a significant long-run probability of successful collective action within a meaningful time scale. This, we argue, sheds light on the puzzle of how large populations can "achieve" 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 using large-scale cross-sectional data from Twitter.



JEL Classification: D74, D72, D71, D83, C72

Keywords: N/A



Chih-Sheng Hsieh - cshsieh@ntu.edu.tw
*National Taiwan University*

Lachlan Deer - lachlan.deer@gmail.com
*Tilburg University*

Michael Koenig - m.d.konig@vu.nl
*Department of Economics, Vrije Universiteit Amsterdam and CEPR*

Fernando Vega-Redondo - fernando.vega@unibocconi.it
*Bocconi University, CEPR & NBER*



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

# 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,e], Fernando Vega-Redondo[f]

[a] *Department of Economics, National Taiwan University, Taipei, Taiwan.*
[b] *Tilburg University, Warandelaan 2, 5037 AB Tilburg, The Netherlands.*
[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] *Centre for Economic Policy Research (CEPR), London, United Kingdom.*
[f] *Department of Decision Sciences, Bocconi University, Via Roentgen, 1, 20136 Milan, Italy.*

**Abstract**

We study a dynamic model of collective action in which agents interact and learn through a co-evolving social network. We consider two alternative scenarios that differ on how agents form their expectations: while in a "benchmark" agents are assumed completely informed of the prevailing state, in the other context agents shape their expectations through a combination of local observation and social learning à la DeGroot. We completely characterize the long-run behavior of the system in both cases and show that only in the latter scenario (arguably the most realistic) there is a significant long-run probability of successful collective action within a meaningful time scale. This, we argue, sheds light on the puzzle of how large populations can "achieve" 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 using large-scale cross-sectional data from Twitter.

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

[☆] We would like to thank Ozan Candogan, Ruben Enikolopov, James Fenske, Sanjeev Goyal, Kei Ikegami, Michihiro Kandori, Dominic Rohner, Mathias Thoenig, Yves Zenou, Fabrizio Zilibotti, and seminar participants at the Tinbergen Institute, Cambridge, Chicago, Princeton, 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 handle and better understand this type of data on collective-action phenomena is, in a nutshell, the main objective of this 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 *belief* she holds on the average action chosen by the rest of the population.

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

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

Alternative (b.1) is the classical formulation considered by the evolutionary learning literature in games, which may be described as of 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*.

In this context, a primary aim of our theory 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 with Gordon Tullock (see Tullock, 1971) – one that derives from the "puzzle" of how *large* populations achieve 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 issue, we shall argue that, from a modeling viewpoint, the received assumption of complete information embodied by (b.1) above is not as useful as (b.2) for our purposes. One of the reasons for this claim is of a general nature: if the population is large, the limited (local) information embodied by (b.2) is so much more plausible. For, in a large context, the assumption that individuals have an accurate perception (even if only approximate) of the current state of affairs can hardly be judged "realistic".

There are, however, two additional reasons supporting (b.2) that bear more specifically on our objective of shedding light on how collective action can arise in large populations. The first one derives from the fact that, under (b.1), our theoretical analysis of the model will show that,

in order for ***any*** significant degree of collective action to arise, there has to be a large fraction of the population (more than half) that displays a high propensity/bias for it. Thus, heuristically, we could describe the situation as one in which collective action requires that the population include a majority of "militants." Instead, under (b.2), we find that, even when the set of such biased militants is small, clusters of local collective action can arise and persist in the long run.

In contrast, the second reason is explicitly dynamic in that, instead of comparing the long-run (stationary) predictions of the model, it focuses on the expected time required by the process to approximate such long-run predictions. Under (b.1), we find that even when the model predicts that collective action can succeed in the long run, the expected time required for it to materialize to any significant extent is inordinately long. Specifically, it grows exponentially in population size, hence yielding a mostly irrelevant long-run prediction in large contexts. Instead, (b.2) induces much shorter and thus meaningful expected times, even when the population is large. Thus, somewhat paradoxically, our theoretical framework leads to the conclusion that more local (network-bound) information greatly facilitates collective action. Intuitively, this happens because, under such limited information, it is much easier for local cliques to form as robust seeds for collective action, upon which sizable collective action can gradually unfold in a time scale that grows only slowly with populations size.

From a methodological viewpoint, the first key step in our analysis of the model is to obtain a full characterization the long-run behavior of the evolutionary process, as captured by the induced *limiting 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 theoretical side, 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 importance of observation and learning (recall (b.2)) in shaping agents' beliefs? What is the effect of individual heterogeneity – and, possibly, homophily – in the formation of the social network? Does easier/cheaper connectivity – and hence a more dense network – help coordination on collective action? What is the impact of the agents' tendency towards local conformity (the desire to have their behavior aligned with that of their neighbors)? And of global conformity (i.e., behavioral alignment with the population at large)?

On the empirical side, our characterization of the limiting distribution of the process in closed form (i.e. as an explicit function of the parameters) is also particularly fruitful. In essence, it means that the theory directly provides a likelihood function that can be used for the structural estimation of the model. And then, of course, such an 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 second part of the paper, where we bring our theory to the analysis of a "big" data set on the massive social protests sparked by the so-called Arab Spring in Egypt.

Specifically, we rely on Twitter data on tweeting and networking by 225,578 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, 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 anti-military 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., 2022; Mele, 2017]. Therefore, we rely on the so-called maximum composite likelihood estimation approach [Lindsay, 1988; Varin et al., 2011], applying suitable computational techniques that combine sparse-matrix and case-control 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 of belief formation (the former having a higher weight than the latter). Such endogenously generated expectations are one of the important forces driving behavior. The other force derives from the inherent costs or risks, as well as benefits and ambitions, that agents anticipate when deciding whether or not to contribute to collective action. In our model their ***net*** effect is captured by a single (fixed) parameter and is therefore exogenous. As part of our econometric exercise, this parameter is estimated from our data and it turns out to have a positive sign. This suggests that, in the Egyptian revolt against the military, the population perceived, on average, that the intrinsic costs and risks entailed were more than offset by the corresponding benefits of joining in.

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

We conclude this Introduction with a brief discussion of the relationship between our research and existing literature, theoretical 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 was 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] for a survey). As a small sample, we may refer to the influential contributions by Acemoglu et al. [2014], DeMarzo et al. [2003], and Jackson and Golub [2010]. While the former paper adopts a Bayesian approach to the problem, the latter two build upon the bounded-rationality framework proposed by DeGroot [1974]. The latter approach – which posits that agents update their beliefs by linearly combining their own with those of their network neighbors – heavily inspires our learning formulation and has received some experimental support

(c.f. Chandrasekhar et al. [2015]). Thirdly, we refer to 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 theoretical contributions whose strategic (equilibrium) analysis of social protests and "revolutions" display such a threshold behavior can be found in the papers by Chwe [2000] and Barberà and Jackson [2020]. Our approach to modeling collective action incorporates many of the signature features of the three aforementioned strands of literature – coordination, (DeGroot) learning, and threshold behavior. As already advanced, it is their integration into a unified framework that turns out to yield new 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, 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 the aforementioned [Badev, 2021; Hsieh et al., 2020; Hsieh et al., 2022; Mele, 2017] that apply these methods to carry out, as in our case, structural estimation of an underlying theoretical model. As already explained, however, their econometric methods are computationally unfeasible in dealing with the large data sets that are our primary concern.

Next, we refer to the recent literature that shares with this paper its focus on the role of social media in facilitating collective action – in particular, in 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 support for the conclusion that a rise in Twitter activity preceded 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. Their main contribution is to identify a causal positive relationship between differences in the degree of social media penetration and the extent of social protest. Interestingly, they also show that the main basis for this effect is *not* the wider access to information the social media provide; instead, they highlight "... the importance of horizontal information exchange on people's ability to overcome the collective action problem." This is in line with the importance that our own analysis attributes to the social network as a channel for the 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 highlighted by the experimental evidence studied in the interesting paper by Cantoni et al. [2019]. 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 on 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. Section 2 presents the theoretical framework (the game-theoretic setup, the law of motion for actions and links, and the different belief formation scenarios). Section 3 undertakes the formal analysis of the model, characterizing the long-run behavior of the system for each of two scenarios considered (global and local information) and comparing their implications. Section 4 summarizes the data used in our application of the model to the Egyptian Arab Spring, while in Section 5 we conduct the econometric analysis, discuss the estimation results, and perform a parameter-recovery exercise that provides support to our econometric approach. Section 6 carries out various types of counterfactual analysis that illustrate the role of some of the forces at work and also serve to explore the effect of possible interventions. Section 7 concludes the main body of the paper, while in Appendix A we include the formal proofs of all our theoretical results. Then, in various Supplementary Appendices, we include the following material: in Supplementary Appendix B we discuss some extensions of the model; in Supplementary Appendix C we rely on the predictions of the model for finite noise to check its consistency with numerical simulations; in Supplementary Appendix D we provide a complete characterization of the stochastically stable states for finite populations; in Supplementary Appendix E we describe the historical conditions that help contextualize our data; Supplementary Appendix F provides additional details for the data construction; and Supplementary Appendix G performs a robustness check for our estimation results.

## 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 the state of the system. Then, in Subsection 2.2 we specify the dynamic, i.e., the law of motion of actions and links that changes the state over time. Naturally, this dynamic 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.

### 2.1. Basic Setup

Consider a population $\mathcal{N} = \{1, \ldots, n\}$, conceived as large, which is involved in a problem of *collection action*. For concreteness, we interpret it to represent some instance of social protest and call it a "riot." Each individual $i \in \mathcal{N}$ must choose an action $s_i$, which is a dichotomous decision of whether to join the riot or not. Formally,[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 $\#(\{-1, +1\}^n) = 2^n$.

[1] Natural Language Processing is a branch of the Artificial Intelligence literature that applies machine-learning methods to text and is becoming widely used in social sciences. Among the growing number of overview articles that can be checked as suitable references, we can list Gentzkow et al. (2017) for its application to economics, or Grimmer and Brandon [2013] who focus 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] This is also the customary convention adopted in the analysis of the classical Ising model [cf. Grimmett, 2010].

The population is also connected through bilateral links as given by the current *social network* $G$. Any such network can be represented by its adjacency (binary) matrix $\mathbf{A} = (a_{ij})_{i,j=1}^{n}$, where each entry $a_{ij}$ $(i, j = 1, 2, \ldots, n)$ is either 1 or 0 if $i$ is either connected to $j$ or not (with the convention that $a_{ii} = 0$). For simplicity, we shall consider undirected networks, which means that the matrix $\mathbf{A}$ is symmetric, i.e., $a_{ij} = a_{ji}$ for all $i, j \in \mathcal{N}$. The set of all undirected networks of size $n$ is denoted by $\mathcal{G}^n$.[3]

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

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

where:

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

We propose the following interpretation of the payoff structure (1). Its first term, the *interaction effect*, captures the genuinely strategic part of the model and has two components: a global and a local one. The global component, whose weight on payoffs is parametrized by $\rho$, embodies the essence of the collective-action problem faced by the population, as perceived by agent $i$. Thus, identifying the aggregate action of all other agents $j \neq i$ with the support for collective action provided by the rest of the population, that aggregate magnitude can also be conceived (when suitably normalized)[4] as the probability that player $i$ attributes to the collective effort being successful. Of course, the probability of failure is then symmetrically determined, which in turn allows a quantification of the relative strength of agent $i$'s incentives to be aligned with the more likely outcome.[5] The second local component of the interaction effect, whose weight is parametrized by $\theta$, reflects the classical assumption commonly made in network setups: agents like to have their behavior well aligned with that of their network partners/friends. As we shall see, the main role played by this local-conformity component in our model is to provide a simple basis to guide agents' networking (linking) behavior.

[3] To understand how the analysis could be adapted if links are taken to be directed, see Supplementary Appendix B.1. However, here we consider undirected links, as it is standard in the social networks literature on peer effects, and we leave the detailed analysis of directed networks to future work.

[4] Specifically, consider the following affine transformation: $\frac{1}{2} + \frac{1}{2(n-1)} \sum_{j\neq i}^{n} s_{ij}$. Alternatively, one could also use other monotone non-affine transformations, but this would render the model significantly less tractable.

[5] In principle, one could also introduce an additional term in (1) that *only* depends on the probability of success for collective action *per se*, i.e. on the aggregate support it receives. But then, since the relative impact on this magnitude of any single individual would be negligible when the population is large, so would be the effect it would have on payoffs and thus on behavior. In view of these considerations, we choose to dispense with such a component for the sake of modeling simplicity.

The next two terms in (1) – i.e. the idiosyncratic bias and the action cost– introduce different types of asymmetries between the two actions. The first one is associated to the fixed type $\gamma_i \in \{+1, -1\}$ of agent $i$ and provides a clear meaning to the notion of type in our context: every agent has a bias in favor of the action that matches her type. In contrast, the second source of action asymmetry induced by $\kappa$ is intrinsically associated to the action choice of the agent, independently of her type. The fact that we speak of it as a "cost" implicitly suggests that $\kappa$ is positive and therefore presumes that the support for collective action (+1) bears an *intrinsic* disutility. Our model, however, allows for the possibility that agents may have a substantial preference for rioting that exceeds any costs. This would be captured by some $\kappa < 0$. In fact, as we shall find in Section 5, the Egyptian Arab-Spring evidence considered by our empirical application of the model yields such a negative estimate for this parameter.

Finally, the last term in (1) includes linking costs $\zeta_{ij}$ for the bilateral link that may formed or maintained between any given pair of agents, $i$ and $j$. Type heterogenenity is taken to have an impact of these costs as follows:[6]

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

with $0 \leq \zeta_1 < \zeta_2$. The above formulation entails that agents with the same idiosyncratic preferences enjoy a lower linking cost, hence inducing a bias/preference for connections between individuals of the same type. This phenomenon, known as homophily, has been shown to be a 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 consist 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 requires specifying how each agent $i$ forms her beliefs $\psi_i$ on the average action of others, $\frac{1}{n-1}\sum_{j\neq i}^{n} a_j$. 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 be assumed to be subject to persistent random noise. This noise can be motivated as the result of a number of different (non-exclusive) factors. One possibility, proposed e.g., by Brock and Durlauf [2001], is that the game is subject to shocks, which are observed by the agents but not by the modeler. Another option is to suppose that the noise captures agents' uncertainty about the payoff (and hence behavior) of others. Finally, a third motivation that has been highlighted by evolutionary game theory [cf. Blume, 1993; Kandori et al., 1993; Young, 1993] is that agents make mistakes or

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

simply experiment with some exogenous probability.

Mathematically, the evolutionary adjustment of actions and links defines a stochastic process that induces a probability measure over the set of state paths of the form $(\boldsymbol{\omega}_t)_{t\in\mathbb{R}_+}$, $\boldsymbol{\omega}_t \in \Omega$, where each state $\boldsymbol{\omega}_t = (\mathbf{s}_t, G_t)$ consists of a vector of agents' actions $\mathbf{s}_t \in \{-1,+1\}^n$ and a network $G_t \in \mathcal{G}^n$. Its law of motion can be described as follows.[7]

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

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

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

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

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

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

Throughout we shall make the assumption that all random shocks are independently and logistically distributed with mean zero and the same scale parameter $\eta \geq 0$. Therefore, its cumulative distribution function $F(\cdot)$ 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:[8]

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

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

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

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

sponding expressions for link adjustment

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

where note that the link-adjustment probabilities are identical for the two agents 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 the process, agents are completely informed of all the payoff-relevant features of the current state of the process. These features include, specifically, the average support for collective action provided by the rest of the population (i.e., the average action). For conciseness, this first scenario is labeled *Global Information* (GI), and is simply captured by the belief-formation mapping $\boldsymbol{\psi}^{GI}=\left(\psi_i^{GI}\right)_{i\in\mathcal{N}}:\Omega\to[-1,1]^n$ that, for each $\boldsymbol{\omega}=(\mathbf{s},G)\in\Omega$, is defined as follows:

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

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

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

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

The observation of the local average action posited in (a) derives from our assumption that every agent directly observes the actions $\{s_j:a_{ij}=1\}$ of her network neighbors and, naturally, know their degree $d_i=\sum_{j\neq i}^{n}a_{ij}$. On the other hand, the local interaction/learning in (b) is modeled along the lines of the well-known framework proposed by DeGroot [1974].[9] More specifically, we posit that given the state $\boldsymbol{\omega_t}=(\mathbf{s}_t,G_t)=[(s_{1t},s_{2t},...,s_{nt})^\top,G_t]$ prevailing at any given time $t$ in the evolutionary adjustment process, there is a sequence of learning rounds, indexed by

[9]See also Berger [1981], Jackson and Golub [2010], Golub and Jackson [2012] and DeMarzo et al. [2003]. Chandrasekhar et al. [2015] and 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.

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

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

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

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

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

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

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

which is a well-defined expression since the matrix $\widehat{\mathbf{D}}_t^{-1} \widehat{\mathbf{A}}_t$ is row-stochastic and $\varphi > 0$.[10]

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

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

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

## 3. Theoretical Analysis

In this section we conduct the theoretical analysis of the model, proceeding in parallel with 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

[10] 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 (5), the function $\Phi : \Omega \to \mathbb{R}$ given by:

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

which can be rewritten as

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

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

$$\begin{aligned} &\text{let } \boldsymbol{\omega} = (s_i, \mathbf{s}_{-i}, G),\ \boldsymbol{\omega}' = (s_i', \mathbf{s}_{-i}, G) \in \Omega; \\ &\text{then, } \Phi(\boldsymbol{\omega}') - \Phi(\boldsymbol{\omega}) = \pi_i(\boldsymbol{\omega}'; \psi_i^{GI}(\boldsymbol{\omega})) - \pi_i(\boldsymbol{\omega}; \psi_i^{GI}(\boldsymbol{\omega})), \end{aligned} \tag{12}$$

and

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

where $G \pm ij$ stands for the network given by $G$ with the link $ij$ added (+) or deleted (−).

To understand intuitively why the function $\Phi(\cdot)$ defined in (11) satisfies (12)-(13), it is useful to think of it as adding the payoffs attained by all individuals in any given state $\boldsymbol{\omega}$, with the important caveat that all those payoffs derived from interaction (i.e. from coordination or link formation) are split equally between the two agents involved. In view of this feature, the additivity and agent symmetry displayed across the different payoff components readily yields the desired match between: (a) the change in individual payoffs experienced by the agent(s) involved in any action or link revision, (b) the corresponding potential change. For the sake of completeness, this conclusion is stated in the following result, whose formal proof is included in Appendix A.

**Proposition 1.** *Given beliefs $\psi_i^{GI}(\cdot)$, the function $\Phi(\cdot)$ given in (10) defines a potential for the agents' expected payoffs $\pi_i(\cdot; \psi_i^{GI}(\cdot))$ for each $i \in \mathcal{N}$, as specified in (1); that is, conditions (12)-(13) hold.*

The previous result leads to 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.[11] In terms of our model, such an adjustment process corresponds to the limit situation where the persistent noise introduced in our adjustment subprocesses (3) becomes vanishingly small, i.e., $\eta \to \infty$. But even for any arbitrary $\eta$ (i.e., possibly small, inducing large noise), the existence of a potential allows one to obtain a sharp (probabilistic) prediction about the long-run behavior of the induced stochastic process. For, as stated by our next result, an adaptation of arguments used in standard models of statistical physics (spelled out in Appendix A) delivers the following result.

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

$$\mu^\eta(\boldsymbol{\omega}) = \frac{e^{\eta\Phi(\boldsymbol{\omega})}}{\sum_{\boldsymbol{\omega}'\in\Omega} e^{\eta\Phi(\boldsymbol{\omega}')}} = \frac{e^{\eta\Phi(\mathbf{s},G)}}{\sum_{G'\in\mathcal{G}^n}\sum_{\mathbf{s}'\in\{-1,+1\}^n} e^{\eta\Phi(\mathbf{s}',G')}}. \tag{14}$$

The previous proposition provides an explicit solution of the model in the GI scenario by specifying, in closed form, how the probability distribution $\mu^\eta$ that characterizes the long-run behavior of the process depends on the noise level modulated by $\eta$ and all other parameters of the model. The empirical validity of the theoretical prediction embodied by (14) is illustrated in Supplementary Appendix C.1 through numerical simulations. Specifically, Figure C.1 focuses on the average degree and average action, tracing how they are affected by changes in the linking cost $\zeta$, a key parameter of the model. We find that the dependence on $\zeta$ exhibited by the simulations is well aligned with the theory, closely approximating the mean degree and mean action prescribed by the corresponding distribution $\mu^\eta(\cdot\,;\zeta)$.

Next, we turn to conducting a parallel analysis for the LIL scenario, where beliefs are governed by the mapping $\boldsymbol{\psi}^{LIL}$ defined in (9). To this end, as a counterpart of the potential posited in (11), we propose a function $\tilde{\Phi} : \Omega \to \mathbb{R}$ that, for all $\boldsymbol{\omega} = (\mathbf{s}\,, G) \in \Omega$, is given by:[12]

$$\tilde{\Phi}(\boldsymbol{\omega}) = \sum_{i=1}^{n}\left\{(\gamma_i - \kappa)s_i + \rho s_i(n-1)\psi_i^{LIL}(\boldsymbol{\omega}) + \tfrac{1}{2}\left[\sum_{j\neq i} a_{ij}(\theta s_i s_j - \zeta_{ij})\right]\right\}. \tag{15}$$

The function $\tilde{\Phi}$ is, of course, much more complicated than $\Phi$ in that, in contrast with the previous case, the belief-formation mapping $\boldsymbol{\psi}^{LIL}$ embedded in it is *not* network-free – that is, it relies on information about the *whole network architecture* prevailing at each possible state $\boldsymbol{\omega}$. Because of this, $\tilde{\Phi}$ fails to be separately linear in the different components of the state (i.e. actions and links), which prevents us from relying on the methodological approach used for the GI scenario to claim that it is strict potential. In what follows, however, we formulate (and

[11] 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.

[12] Note that, in contrast with $\Phi$, 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.

later support, numerically and econometrically) the conjecture that it works well as a useful approximation for a strict potential function in the LIL scenario. Thus, for want of 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 construction of the potential for the GI scenario suggests that the equal-splitting factor of 1/2 correcting for double counting should be applied only to the payoffs resulting from local interaction. Instead, for the payoffs agents *expect* from global interaction, the mostly limited and local nature of agents' belief adjustments render it unnecessary, as an approximation, to account for the aforementioned correction.

The combination of all three observations above suggests that, if the population is large and the noise small, the changes over time of the function $\tilde{\Phi}(\boldsymbol{\omega}_t)$ must be largely dominated by the payoff changes experienced by those agents who revise their decisions (on 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 the left panel of Figure 1 a typical trajectory for small noise. Furthermore, in the right panel of Figure 1 we show that the relative differences between the changes in the quasi-potential and the changes in payoffs are small and get smaller the closer the process is to the stationary state. This indicates that the quasi-potential $\tilde{\Phi}$ exhibits a behavior similar to that established for the GI scenario in terms of the potential $\Phi$ (cf. Proposition 1).

Next, motivated by the previous considerations, we formulate the following conjecture: under LIL, the long-run behavior of the process is well described (approximately) by a probability distribution with the same format arising for the GI scenario in terms of the potential $\Phi$ (cf. Proposition 2). That is, we posit that the entailed stochastic process induces a limiting distribution that is well approximated by the distribution $\tilde{\mu}^\eta$ defined, for every $\omega = (\mathbf{s}, G) \in \Omega$, as follows:

$$\tilde{\mu}^\eta(\boldsymbol{\omega}) = \frac{e^{\eta\tilde{\Phi}(\boldsymbol{\omega})}}{\sum_{\boldsymbol{\omega'}\in\Omega} e^{\eta\tilde{\Phi}(\boldsymbol{\omega'})}} = \frac{e^{\eta\tilde{\Phi}(\mathbf{s},G)}}{\sum_{G'\in\mathcal{G}^n}\sum_{\mathbf{s}'\in\{-1,+1\}^n} e^{\eta\tilde{\Phi}(\mathbf{s}',G')}}. \tag{16}$$

We provide support for (16) 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 (16). As an illustration, in Figure C.2 in Supplementary Appendix C.2, we describe results that reproduce, for the LIL scenario, the very good match between theory and simulations that are

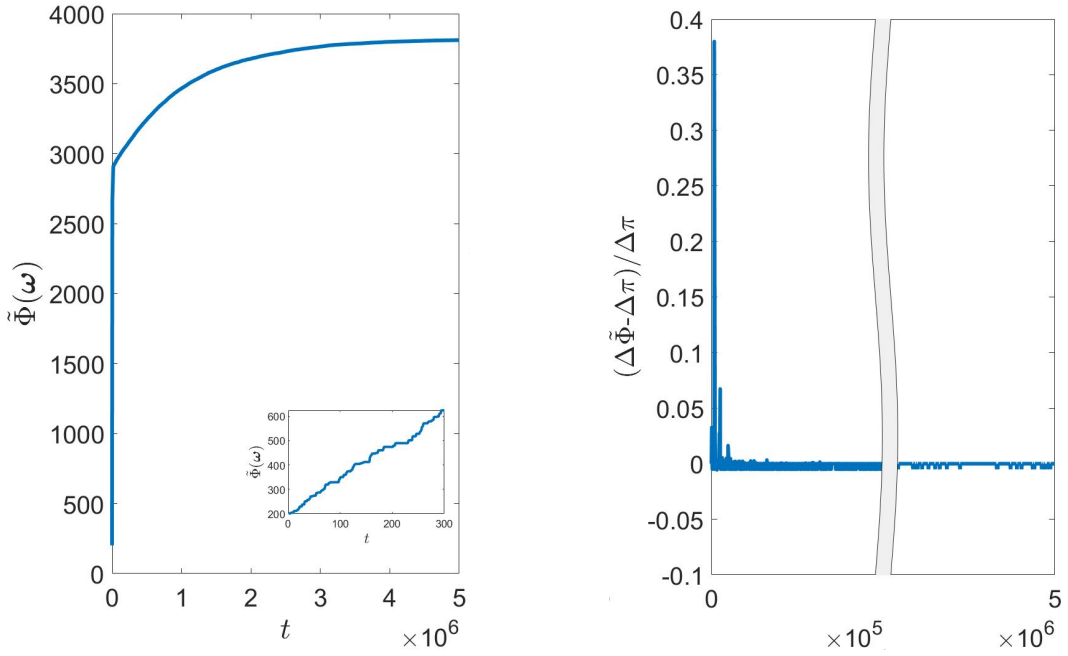



Figure 1: The left panel shows a typical trajectory of the quasi-potential, $\tilde{\Phi}(\omega)$, under the dynamics outlined in Subsection 2.2. The quasi-potential is increasing and reaches a plateau when the process converges to the stationary state, with the inset showing the initial transition. 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 quasi-potential and the changes in payoffs are small and get smaller the closer the process is to the stationary state.

obtained for GI context (shown in Figure C.1 in Supplementary Appendix C.1). Specifically, they show that, as the linking cost $\zeta$ 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 provides an interesting complementary perspective on the issue. For, by involving our econometric methodology, it is also reassuring for our empirical strategy (not only in the LIL scenario but in the GI one as well). This approach will be explained in detail in Subsection 5.3 once the econometric analysis has been presented and applied to our data in Subsections 5.1 and 5.2. Here, we only provide a brief advance of the 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. And, as mentioned, we do it 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 indicates that the composite-likelihood approach underlying the econometric analysis is an effective estimation procedure for both informational scenarios (of course, as long as we

choose the the same scenario for the generation of the data and the estimation exercise). Second, it provides additional support to the claim that not only the (strict) potential approach used for the GI scenario, but also the (approximate) one used for the LIL context, both capture the essential features of the long-run behavior of the system. Only if this is indeed the case can one reasonably justify that they represent useful bases to conduct the econometric estimation in both cases.

### 3.2. Achieving Collective Action

In this subsection we compare the two informational scenarios, GI and LIL, from a standpoint 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 sometimes 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 modeling the context as one where individuals' information about the current state is not global but local, and such local information is updated through a process of social learning mediated through the network.

More concretely, our analysis of the problem approaches it from two different 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 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 really is). As we show in this section, for each of these alternative approaches to the question, the LIL-based framework provides a substantially stronger basis than that based on the GI assumption for understanding the rise of collective action in large populations.

To study the problem, we pursue a methodology, commonly used in evolutionary theory (see e.g., the seminal work of Kandori et al. [1993] and 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 $\mu^\eta$ is given by (14), 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 (16). Since in this case this distribution is based on what we have called 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).

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:

$$\lim_{\eta\to\infty} \mu^\eta(\mathbf{s}, G) \begin{cases} > 0, & \text{if } \Phi(\mathbf{s}, G) \geq \Phi(\mathbf{s}', G'), \quad \forall \mathbf{s}' \in \{-1, +1\}^n, \quad G' \in \mathcal{G}^n, \\ = 0, & \text{otherwise} \end{cases} \tag{17}$$

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 by, respectively, Propositions D.1 and D.2 in Supplementary Appendix D. Figures 2 and 3 illustrate

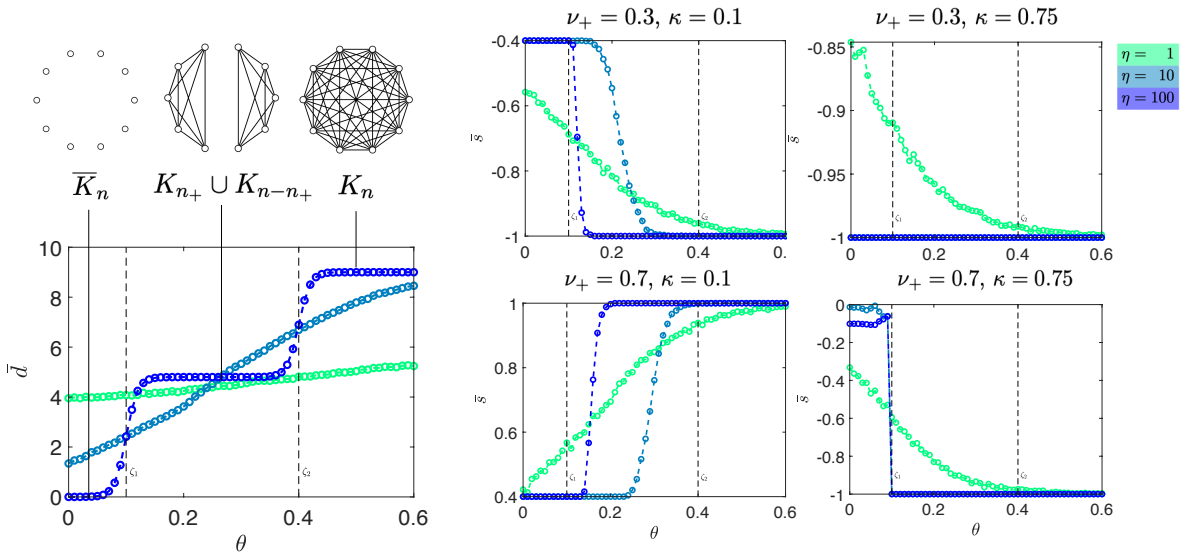



Figure 2: Simulation results for the ***GI scenario*** on the average degree of the network $\bar{d}$ (left panel) and the average action $\bar{s}$ (right panel) for varying values of $\theta$ and $\eta$, using the "next reaction method" for simulating a continuous time Markov chain [cf. Anderson, 2012; 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 $\zeta_1$ and $\zeta_2$ are indicated with a vertical dashed line. As $\eta$ 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 ($K_n$).

graphically how the long-run average degree $\bar{d}$ and average action $\bar{s}$ change as $\eta$ grows, i.e., as the noise becomes progressively smaller, hence approximating stochastically (quasi-)stable values. We observe that, despite the fact that the diagrams pertain to contexts with a relatively small population ($n = 10$), the displayed functions approach a step-like form as $\eta$ grows. This points to the sharp selection patterns (on actions and network) that is predicted in extreme form by our theory when the noise is small and population large (cf. Propositions 3-6).

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

(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 and the network becomes empty. 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, explained in the Introduction, the Tullock Paradox is most acute. Formally, the way in which we capture these large-population considerations is to formulate our stability notions on the limiting 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:

$$\Omega^* = \{\omega = (\mathbf{s}, G) : \lim_{n\to\infty} \lim_{\eta\to\infty} \mu^\eta(\omega) > 0\}, \tag{18}$$

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

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

As it turns out, the characterizations of the two sets defined in (18) and (19) depend on 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 $\nu_+$ is lower or higher than 1/2. It is helpful, therefore, to state separate results characterizing the sets given in of the for each of these two cases. Considering first the GI scenario, the set of LSSS in the set $\Omega^*$ 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 (5) and suppose $\theta > \zeta_1$. Let $n \to \infty$ and assume $0 < \lim_{n\to\infty} \nu_+ < 1/2$. 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 (5) and suppose $\theta > \zeta_1$. Let $n \to \infty$ and assume $1/2 < \lim_{n\to\infty} \nu_+ < 1$. 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. 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 with such a complete network. If the number of $+1$ types is less than half (as in Proposition 3), there is

[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

[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.

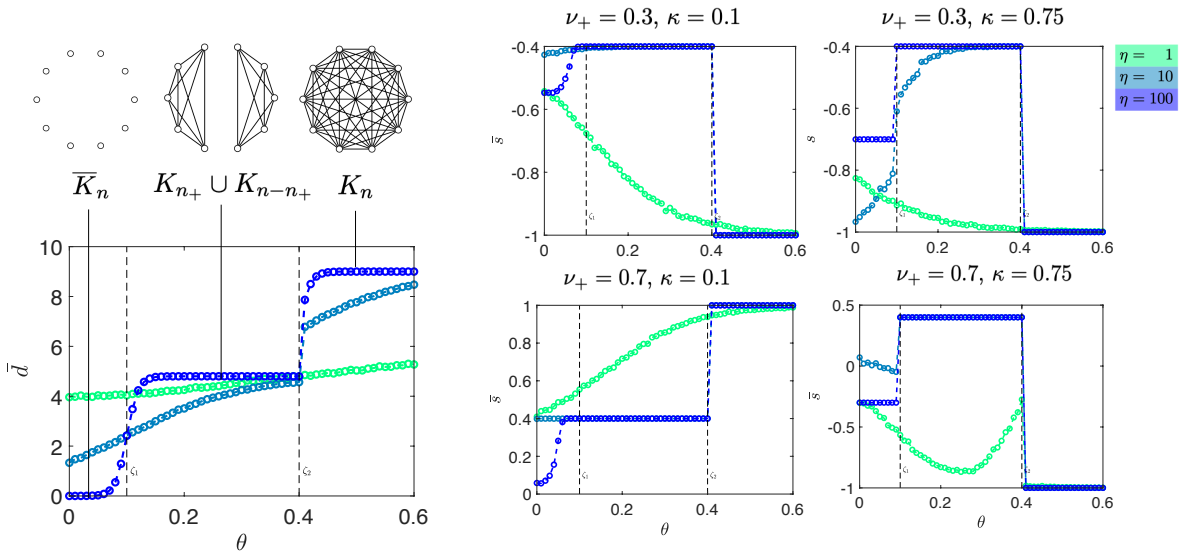



Figure 3: Simulation results for the ***LIL scenario*** on the average degree of the network $\bar{d}$ (left panel) and the average action $\bar{s}$ (right panel) for varying values of $\theta$ and $\eta$, using the "next reaction method" for simulating a continuous time Markov chain [cf. Anderson, 2012; Gibson and Bruck, 2000]. The circles represent averages obtained across 1000 Monte Carlo runs under the following parameters: $n = 10$, $\eta \in \{1, 10, 100\}$, $\lambda = \chi = \xi = 1$, $\varphi = 0.5$, and $\rho = 0.1$. The stochastically stable states in the limit of large $\eta$ correspond to Propositions 5 and 6, respectively. The thresholds $\zeta_1$ and $\zeta_2$ are indicated with a vertical dashed line. As $\eta$ 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 ($K_n$).

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 $\kappa$ 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 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 $\kappa$ of action $+1$, the fraction of agents who display the corresponding type $+1$, and the strength of the global conformity parameter $\rho$. 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 LSSS.

Returning to Figure 2, we find an illustration of the fact that the pattern of network ar-

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

chitecture and action choice predicted by Propositions 3 and 4 matches the pattern displayed by a numerical simulation of the process for a finite population. Specifically, we observe that, as $\theta$ varies across the three regions marked by the linking costs $\zeta_1$ and $\zeta_2$, 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 previous results for the LIL scenario and the set $\Omega^{**}$ of LSQS. Again, we find it useful to separate the conclusions obtained when the limit fraction $\nu_+$ of agents of type $+1$ is either lower or higher than $1/2$. type $+1$.

**Proposition 5.** *Assume all agents form their beliefs as prescribed by the function $\psi^{LIL}(\cdot)$ given by (9) and suppose $\theta > \zeta_1$. Let $n \to \infty$ and assume $0 < \lim_{n\to\infty} \nu_+ < 1/2$. Then we have:*

1. *If $\theta < \zeta_2$ the unique LSQS 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 LSQS 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 LSQS 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 (9) and suppose $\theta > \zeta_1$. Let $n \to \infty$ and assume $1/2 < \lim_{n\to\infty} \nu_+ < 1$. Then we have:*

1. *Suppose $\kappa > 2\nu_+ - 1$.*
   (a) *If $\theta < \zeta_2$ the unique LSQS 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 LSQS 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 LSQS 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 LSQS 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 LSQS has all agents choosing the action $\boldsymbol{s_i = \gamma_i}$.*
   (b) *If $\theta > \zeta_2$ then the unique LSQS is given by the **complete network** $K_n$ with all agents choosing the **action** $\boldsymbol{s_i = +1}$.*

In analogy with what was shown in Figure 2 for the GI scenario, Figure 3 illustrates that the simulation results for finite populations and small noise in the LIL scenario are well aligned with the theoretical predictions stated in Propositions 5-6. The issue then arises as to how, in general, these latter predictions for the LIL scenario contrast with those obtained by Propositions 3-4 for the GI scenario. To address this question, 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 for both scenarios (i.e., action $+1$ cannot be too costly), under LIL it is also sufficient. Instead, in order to attain the indicated target for 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

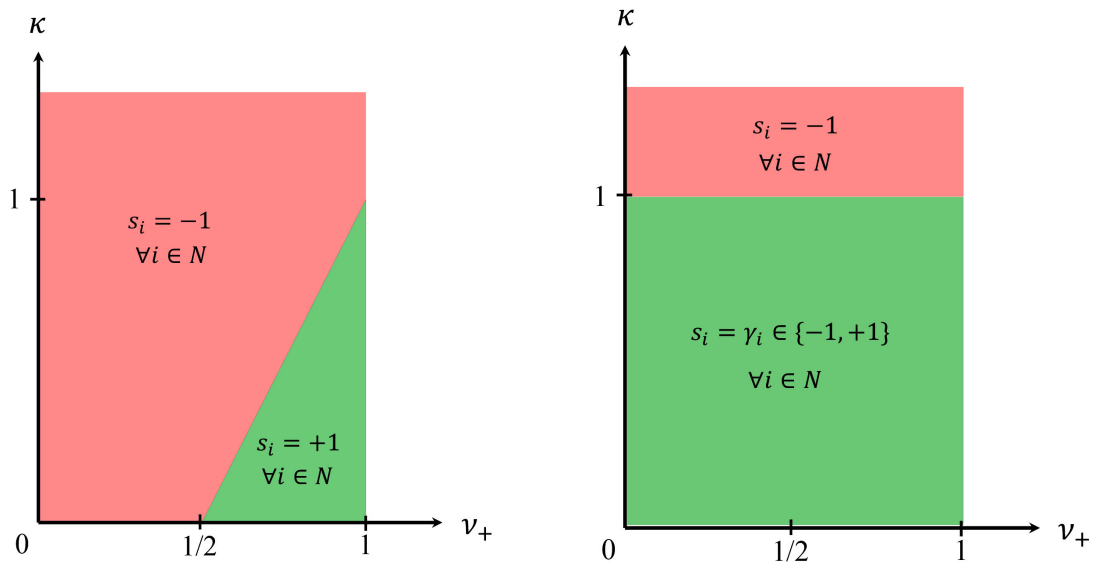



Figure 4: Comparing the two limit stochastically stable and quasi-stable states predictions with $\zeta_1 < \theta < \zeta_2$ for the GI model (left panel) and the LIL model (right panel).

condition is that $\kappa < 2\nu_+ - 1$. This inequality is obviously more stringent than simply requiring that $\kappa < 1$, except for the extreme case where $\nu_+ = 1$, i.e., when essentially all individuals in the population are of type $+1$. For a graphical description of the situation, the reader can refer to Figure 4.

The previous discussion highlights the point that a model of collective action where agents have limited information and learn from their peers is not only more realistic but also expands significantly the range of circumstances (parameters) for which such an outcome is predicted to eventually materialize. This seems a useful first step towards addressing, for example, what we have labeled the Tullock Paradox – i.e. the puzzle that, despite apparently considerable difficulties, revolutions (a form of collective action) sometimes succeed even when the population involved is large. But the problem can also be studied from a genuinely dynamic viewpoint that is complementary to such long-run analysis. This approach shifts the focus from the limiting behavior of the process to the speed at which such long-run outcome is reached. Even though a formal analysis of the question is beyond the scope of this paper, in what follows we provide some illustrative simulations that convey the main gist of the following idea, namely, that the LIL framework also provides some basis to understand why collection can arise within a reasonably short time scale.

The 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 involve 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* – the social network is empty, and action $-1$ is chosen by everyone in the population – we ask the following questions: What is the expected time required to reach such a sizable level of collective action? How does this expected time change as the population size grows?

Figure 5 describes the simulation results that explore the former questions for a typical configuration of parameters that meet the conditions described above. (Specifically, they satisfy

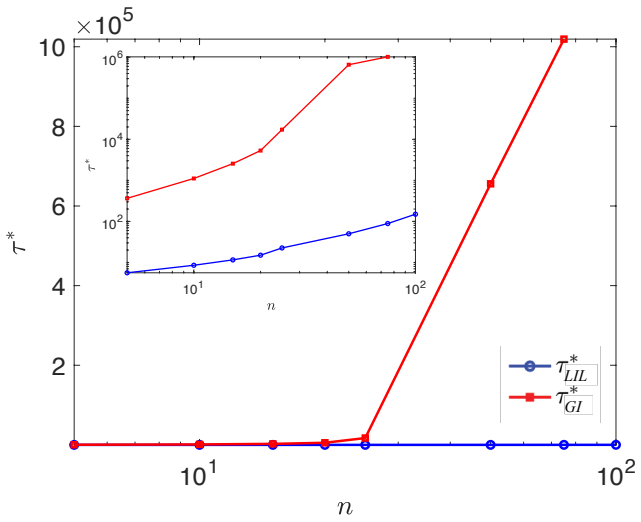



Figure 5: The convergence times, $\tau^*$, to a configuration with at least half of the population choosing action $+1$ for different population sizes ($n$) starting from an empty network, $\overline{K}_n$, where all agents choose action $s_i = -1$ for the GI and the LIL models. The circles represent 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 in the log scale.

that $\zeta_1 < \theta < \zeta_2$, $\nu_+ > 1/2$ and $\kappa < 2\nu_+ - 1$.) The plot 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. It shows a sharp contrast between these two alternative contexts. Specifically, we find that the average delay in the former case is much higher than in the latter, even if the population size $n$ is relatively small – thus, for $n = 100$, the difference spans more than 5 orders of magnitude. This indicates that, even when the GI scenario also yields the prediction that a substantial level for collective action will be achieved in the long run, actually reaching this level takes much more "time" (adjustment instances) than in the LIL context. This is yet an additional reason why the theoretical framework that posits local information and local learning appears as a better one to understand how collective action arises in large-population contexts. For it is not only that the LIL scenario supports collective action in the long run under a significantly wider set of circumstances, and it does so by relying on a more realistic behavioral assumption – it also provides a much more limited time scale (hence a more plausible dynamic basis) for how the population reaches that situation.

## 4. Data

The empirical application of our model uses online social network data from Twitter focussing on riots and demonstrations in Egypt during the Arab Spring in 2013 [cf. Borge-Holthoefer et al., 2015; Magdy et al., 2016]. Our decision to use Twitter stems from it serving as the the main social networking platform for opinion exchange over Egypt's Arab Spring [cf. Clarke and Kocak, 2020]. The empirical analysis leverages protest-related tweets surrounding the so-called Second Egyptian Revolution, where the incumbent Islamist president, Mohamed Morsi,

is ousted the military returns to power.[16]

We collected around 6 million Arabic language tweets based on data from Magdy et al. [2016] and Borge-Holthoefer et al. [2015] and focus on the period between July 4, 2013, when the military overthrew the regime of president Morsi, and August 19, 2013, when the scales of demonstrations and the volumes of tweets reached a saturation point.[17] Using information on the number of protests from the Armed Conflict Location and Event Data Project [ACLED, 2019] over the same sampling period, we find a significant concurrence of Twitter message volumes and real-world protests.[18]

### 4.1. Sample Selection & Network Construction

Our analysis focusses on users that have posted more than two tweets that are either retweets or include mentions of other users, and have less than the 90th percentile of followers (2,918 followers). Focussing on accounts with at least two interactions with other users is designed to allow the network connections to play a role in the content a user chooses to post. It also provides us with a large enough sample of tweets for each user to use their tweet text classify them by their political affiliation and gender. Restricting users based on the number of followers is designed to remove accounts of politicians, celebrities, and the accounts of news and other media sources. The resulting estimation sample consists of 225,578 users.

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

### 4.2. From Text to Quantitative Measurement

The remainder of our empirical data is constructed using tools from Natural Language Processing [Ash and Hansen, 2023; Gentzkow et al, 2017; Grimmer, Roberts and Stewart]. In particular, we create three variables from the text of each user's tweets: (i) is a user "rioting" (i.e. are the pro- or anti miliatry intervention)? (ii) the political affiliation of each user (Secular

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

[17] Supplementary Appendix F.1 describes the data collection process in more detail. In addition to the date window featuring the largest scale of demonstrations over this phase of Egypt's Arab Spring driving our decision on the time horizon to focus our study rests on when the tweets originally collected by Borge-Holthoefer et al. [2015] identify protest relevant tweets. Borge-Holthoefer et al. [2015]'s data collection process stops updating the search terms used to identify and collect protest related tweets in mid-August 2013. As the hashtags and keywords used by protesters continue to evolve, their collection process no longer includes all relevant tweets.

[18] The Pearson correlation coefficient between the two indices, ACLED and Twitter, is positive and given by 0.32. Note that we only have information on the number of protests from ACLED but not their size. This will likely lead to an underestimate of the correlation between the actual protest participation and the protest-related Twitter messages.

vs Islamist), and (ii) the gender of each user.[19] The construction of each variable is treated as binary classification problem where we infer each user's trait from the text of each user's tweets [Rao et al., 2010]. The underlying idea is that the tweets written by a Twitter user contain words, phrases and ways of writing are predictive of their stance on the military intervention, political affiliation and gender.

Our approach leverages the latest generation of pre-trained BERT language models [Devlin et al., 2018]. The pre-trained model is then "fine-tuned" for the each of the three specific classification tasks mentioned above.[20] Since pre-trained language models are built from building a general understanding from a diverse set of text inputs, the fine-tuning apprach can achieve good performance even when only a small number of training samples are provided [Ash and Hansen, 2023].

We create three fine-tuned models using a pre-trained BERT language model developed for Arabic, "AraBERT" [Antoun et al., 2020].[21] AraBERT is the dominant pre-trained language models for the Arabic language and has state of the art performance across a large range of Arabic language tasks [Abdul-Mageed et al, 2021].[22] The base AraBERT model is trained on over 200 million lines of Arabic text using a transformer architecture identical to the "baseline" English language BERT model. For each fine-tuning task, we add a feed-forward Softmax classification layer the top of the AraBERT encoder output. Classifier and the pre-trained model weights from AraBERT are trained jointly during fine-tuning to maximize the log-probability of correct class assignment. Tweets are cleaned through a two step procedure before being passed across AraBERT: (i) words are segmented using the "Farasa segmenter" [Abdelali et al., 2016] and (ii) training a "SentencePiece" tokenzier [Kudo and Richardson, 2018].[23] Each fine-tuned AraBERT model uses a probability threshold of 0.5 when predicting classes. After fine-tuning we use the model to predict on all tweets in the estimation sample.

### 4.3. 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.[24] 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 of the willingness

[19] See Hinds and Joinson. [2018] for a systematic review of using digital footprints to reveal demographic attributes.

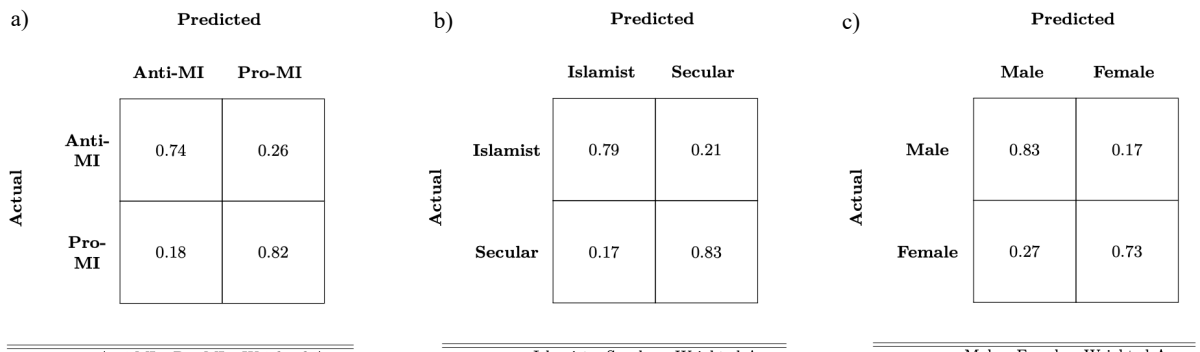
[20] See Bana [2022]for another recent example of fine-tuning a BERT model in ecomomics, albeit in a different setting. He fine-tunes a BERT model to predict salaries from the text of job postings.

[21] The decision to use an Arabic specific language model like AraBERT as opposed to a cross-lingual model such multilingual BERT, 'mBERT' [Devlin et al., 2018] and those developed by Conneau et al [2020] is due to the different morphologic structure of Arabic when compared to Latin based languages where the multi-lingual approach is more common.

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

[23] Segmenting Arabic words is a building block of Arabic NLP that helps recover meaning by reducing lexical sparsity and simplifying syntactic analysis.

[24] *Stance* detection differs from sentiment analysis. Stance refers to the expression of someone's judgement (expressed in text) towards a given idea, in our case the military intervention, thus stance detection aims to predict a user's position on the idea. In contrast, sentiment analysis outputs whether the text uses sentiment laden words that are positive, negative or neutral in tone to classify a piece of text as such. See Aldayel and Magdy [2021] for a detailed review of the stance detection literature.




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

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. We use 80% of the and annotated tweets as training data for fine-tuning and the remaining 20% to validate model performance on unseen data. The out-of-sample performance is shown in the bottom left panel in Figure 6. The action classifier has a high accuracy (representing the number of correctly classified data instances over the total number of instances) with the corresponding confusion matrix used to compute the accuracy shown in the top right panel in Figure 6. Also, the F1-score (the harmonic mean of precision and recall) is high. This shows that our classifier performs well in classifying the users' actions from the tweets' texts. The classifier is then applied to all tweets in the estimation sample. Our stance detection algorithm's out of sample performance is similar to those in the existing literature [Darwish et al, 2018; Qiu et al., 2015; Rajadesingan and Liu, 2014].

### 4.4. User Characteristics: Political affiliation

Egyptian society is politically polarized between secularists and Islamists since well before the onset of the Arab Spring. The two dominant sides are an 'Islamist' group that supports the Muslim Brotherhood and then-incumbent president Morsi and a 'Secularist' group that endorsed religious tolerance.[25] The Islamist camp is known to support the Muslim Brotherhood whilst

[25]Despite the associations made between political Islamism and religious Islamism by western media over the Arab spring the connection between the two, although postively correlated, is more subtle. See Weber et al. [2013] for a further discussion.

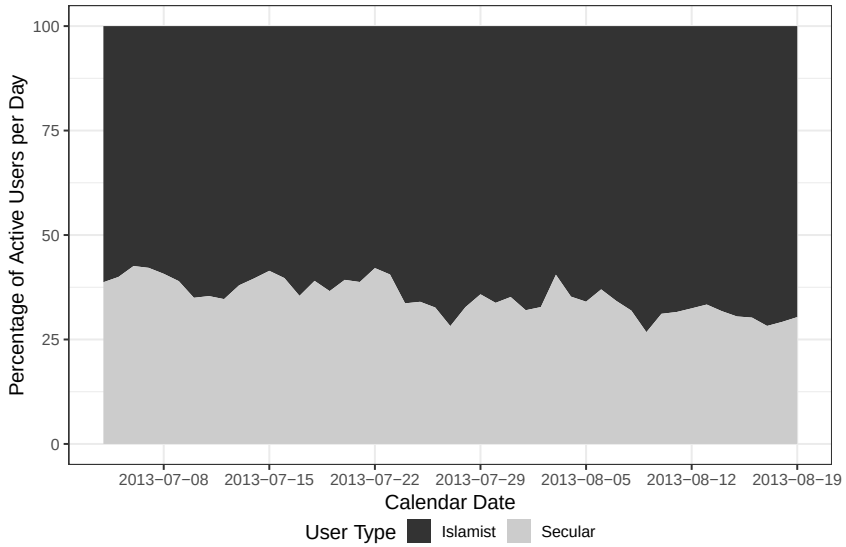



Figure 7: The fraction of Islamist versus secular users over the sample period.

the Secular 'opposition' stands against the Brotherhood and their ideology. Which of these groups a Twitter user identifies with is likely to influence their stance on Morsi's removal from power via the military intervention, i.e whether they are rioting.

We leverage the text of tweets to infer the political ideology of Twitter users using a supervised classification algorithm. We utilize an already annotated sample of 20,886 Egyptian Twitter users and their political affiliations provided by Weber et al. [2013] to finetune an AraBERT model aimed at predicting whether tweets are written by an Twitter user with either Islamist or Secular views. In practice, we build a training sample by matching the classified users in Weber et al. [2013] to users in our sample.[26] The matched users are randomly split 80/20 into training/test samples for the AraBERT finetuning. All tweets by a user allocated to the test (training) sample are used in the parameter fine-tuning (evaluation). The fine-tuned AraBERT model predicts whether an individual tweet is written by an Islamist or Secularist. To aggregate these predictions from individual tweets up to the user level we use a "majority vote" over their tweets to predict their political affiliation. For example, a user with three secular tweets and five Islamist tweets is classified as a political Islamist.

Out of sample performance of the political affiliation classifier at the Twitter user level is reported in the middle panel of Figure 6. The high F1-score and accuracy, together with the corresponding confusion matrix shown in the top middle panel in Figure 6 indicate that the classifier succeeds at identifying political affiliation from user tweets.

Figure 7 shows the fraction of Islamist versus secular users over the sample period. Note that in our sample, the majority of agents are Islamists, and that is why the action +1, which is protesting again the military intervention, dominates (around 73%, see also Table 1). Figure 8 shows the users' political affiliations and links between them based on a "snowball" sample of a randomly drawn user and her neighbors, neighbors' neighbors, etc. of the original network of users [Goodman, 1961]. A clear separation between densely connected clusters of users with the

[26]Further details about the construction of the training data are available in Supplementary Appendix F.2.

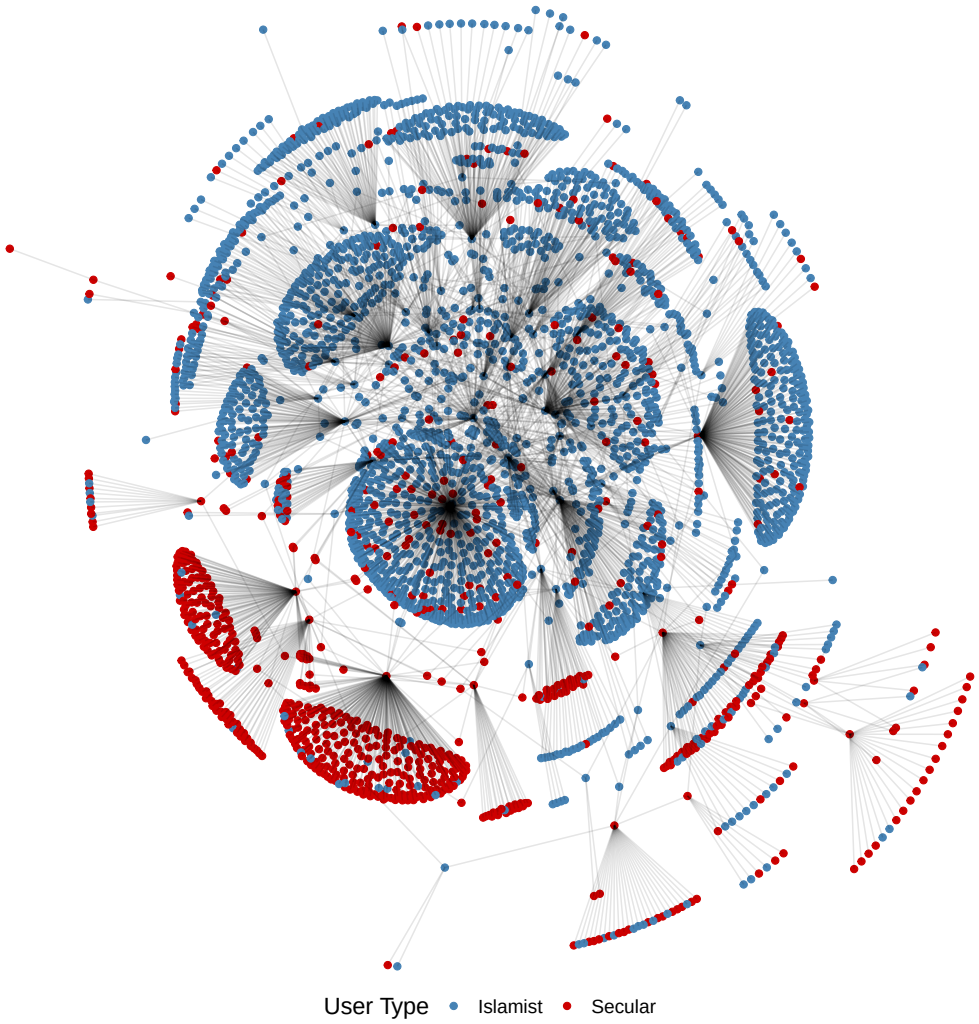



Figure 8: Users' political affiliations and links based on a "snowball" sample of a randomly drawn user and her neighbors, neighbors' neighbors, etc. of the original network of users [Goodman, 1961].

same political affiliation can be seen with only a few links across these clusters.[27] This indicates that users are mainly connected with other users with the same political views.

## 4.5. User Characteristics: Gender

Like many countries in the Middle East, Egyptian society features prevalent gender roles rooted in of cultural norms stemming from religious beliefs and historical traditions. These norms may, in our context, lead to differences in the propensity to protest between male and female users and disparities in the structure and dynamics of networks between genders. As a result, we build a classification model to identify protester's gender from their tweets.

To classify the gender of Twitter users we proceed analogously to our approach on political affiliation, finetuning a third AraBERT model at the tweet level before aggregating up to a

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

user level prediction. For training data we source the gender classification of subset of Egyptian Twitter users from Weber et al. [2013], who provides gender classifications alongside the political affiliations utilized above. Because tweets about riots might not be sufficiently informative about a user's gender, we augment the protest relevant tweets with a larger sample of tweets written by these users. To do so, we extract the most recent 3200 tweets from each of the users in Weber et al's data that match our data from the public Twitter API.[28]

Users are split 80/20 into a traning/test sample split with all tweets of each user being used in the the finetuning process. AraBERT predictions are again at the tweet level and we aggregate to predict a user's gender using majority vote. For example, if a user has three "male" tweets and five "female" tweets classify a user as female. The out-of-sample performance of our gender classifier at the user level is shown in the bottom right panel in Figure 6. High F1-score and accuracy score, along with the confusion matrix (shown in the top right panel in Figure 6) indicate that the classifier does peforms well at predicting gender from text. Our model's performance a achieves similar level to other recently developed Arabic-based BERT models that have been fine-tuned for predicting gender from Arabic text [Abdul-Mageed et al, 2021; Zhang and Abdul-Mageed, 2019].

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

Table 1: Summary statistics.

| | mean | s.d. | max | min |
|---|---|---|---|---|
| Action | 0.4848 | 0.8746 | 1 | -1 |
| Female | 0.0974 | 0.2965 | 1 | 0 |
| Islamist | 0.6157 | 0.4864 | 1 | 0 |
| Number of followers | 472.7100 | 569.1000 | 2918 | 1 |
| Degree | 0.8313 | 2.1345 | 47 | 0 |
| Number of nodes | | 225,578 | | |
| Number of links | | 93,762 | | |

# 5. Empirical Analysis

## 5.1. Identification and Estimation Method

The stationary distributions corresponding to the GI and LIL scenarios in (14) and (16) are known as Gibbs measures (or the Gibbs random field; see cf. Grimmett, 2010; Wainwright and Jordan, 2008), which provide us with a probability (likelihood) measure for estimating unknown structural parameters in the potential functions (11) and (15) from 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 $\gamma_i$ and the linking cost $\zeta_{ij}$ in (11) and (15) as functions of observed and unobserved individual

[28]The API pull was implemented in December 2020. 3200 tweets is the maximum allowed via Twitter's Public API.

characteristics:

$$\gamma_i = \mathbf{x}_i^\top \boldsymbol{\beta} + \tau z_i, \tag{20}$$

and

$$\zeta_{ij} = \phi_0 + \sum_{k=1}^{K} h_k(x_{ik}, x_{jk})\phi_k - z_i - z_j, \tag{21}$$

where $\mathbf{x}_i$ is a $K \times 1$ vector of exogenous regressors including individual gender, political affiliation, and log number of Twitter followers, while $\boldsymbol{\beta}$ are the corresponding unknown parameters. The variable $z_i$ represents the individual random effect which captures unobserved individual heterogeneity and is assumed to be identically and independently normally distributed with mean zero and variance $\sigma_z^2$, i.e., $z_i \sim \mathcal{N}(0, \sigma_z^2)$. The function $h_k(x_{ik}, x_{jk})$ in $\zeta_{ij}$ can be either an indicator function ($\mathbb{1}(x_{ik} = x_{jk})$) when $x_{ik}$ is a dummy variable or a distance function ($|x_{ik} - x_{jk}|$) when $x_{ik}$ is continuous, reflecting homophily (or heterophily). The higher values of random effects $z_i$ and $z_j$ in $\zeta_{ij}$ lead to a lower linking cost, capturing the extent of inter-agent heterogeneity due to unobservables [Dzemski, 2019; Graham, 2017; Jochmans, 2018]. We denote the unknown parameters in $\zeta_{ij}$ by $\boldsymbol{\phi} = (\phi_0, \phi_1, \cdots, \phi_K)$.

There are two main identification issues regarding our model specification. First, the noise parameter $\eta$ in the logistic disturbance $\varepsilon$ of (14) and (16) is not separately identifiable from other parameters in the (quasi) potential function, a commonly known problem in discrete choice models. Therefore, as common in this literature, we set $\eta$ to one during estimation. Second, in the GI potential given by (11), we capture the global conformity effect $\rho$ through the "leave-one-out" sum of actions, $\sum_{j \neq i}^{n} s_j$. When the sample size $n$ is large, the leave-one-out sum has only a negligible variation at the individual level and therefore the coefficient $\rho$ is hardly disentangled from the rioting cost $\kappa$. 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 $\kappa$ for the GI scenario. To mark this modification, we denote the coefficient in front of $(n-1)\bar{s}$ in this modified model by $\widetilde{\rho}$. A similar problem does not appear in the LIL scenario of (15) because the global conformity effect can be identified through variations on the individual belief $\psi_i^{LIL}$. It is also worth mentioning that identification of individual random effect $z_i$ is made possible by exploiting variations on link decisions, $a_{i1}, a_{i2}, a_{i3}, \cdots$ of agent $i$ in a similar manner as an individual fixed effect in panel data. More specifically, as $z_i$ appears in all link decisions of agent $i$, we can identify $z_i$ from the conditional probability,

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

Given $z_i$ is identified, we can also identify the coefficient $\tau$ in $\gamma_i$ in (20).

The main challenge to the estimation of the Gibbs measures in (14) and (16) comes from the appearance of a normalizing constant in the denominator, which involves the evaluation of the (quasi) potential function over all possible networks $G \in \mathcal{G}^n$ and action profiles $\mathbf{s} \in \{-1, +1\}^n$. When the network size is large, this normalizing constant becomes intractable, and thus directly calculating the likelihood for conventional frequentist or Bayesian estimation is not possible.[29] The most commonly used methods to tackle such estimation challenges include the composite likelihood approach [Lindsay, 1988; Varin et al., 2011],[30] the Monte Carlo simulated likelihood

[29] Note that there are $2^n$ possible action profiles $\mathbf{s}$ and $2^{\binom{n}{2}}$ possible networks $G$ with $n$ agents.

[30] In general, the composite likelihood for $y = (y_1, \cdots, y_n)$ can be written as $f(y) = \prod_{r=1}^{R} p(y_{A_r} | y_{B_r})$ where observations are divided into $R$ blocks and the block $B_r = \mathcal{N} \setminus A_r$ with $\mathcal{N} = \{1, 2, \cdots, n\}$ denotes the index

approach [Geyer and Thompson, 1992], the Bayesian exchange algorithm [Møller et al., 2006; Murray et al., 2006] with exact sampling, or the Bayesian double Metropolis-Hastings algorithm [Badev, 2021; Hsieh et al., 2020; Hsieh et al., 2022; Liang, 2010; Mele, 2017]. Given the enormous sample size faced in the present study, the latter simulation-based methods are not feasible, and therefore we adopt the composite likelihood method to estimate our model.[31] The composite likelihood for our model is defined as

$$\mu(\mathbf{s}|G)\mu(G|\mathbf{s}), \tag{22}$$

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

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

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

$$\tilde{\mu}(\mathbf{s}|G,\boldsymbol{\psi}^{LIL}) = \frac{\exp\left(\sum_{i=1}^{n}(\mathbf{x}_i^\top\boldsymbol{\beta} + \tau z_i + \kappa + \rho(n-1)\psi_i^{LIL})s_i + \frac{\theta}{2}\sum_{i=1}^{n}\sum_{j\neq i}^{n} a_{ij}s_is_j\right)}{\mathcal{Z}(G)}, \tag{24}$$

where $\mathcal{Z}(G) = \sum_{\mathbf{s}\in\{-1,+1\}^n}\exp(\sum_{i=1}^{n}(\mathbf{x}_i^\top\boldsymbol{\beta}+\tau z_i+\kappa+\rho(n-1)\psi_i^{LIL})s_i+\frac{\theta}{2}\sum_{i=1}^{n}\sum_{j\neq i}^{n} a_{ij}s_is_j)$. Since the conditional likelihood functions in (23) and (24) still contain intractable normalizing constants, we also replace them with the composite likelihood function. Considering first the GI instance in (23), the conditional probability of agent $i$ choosing action $s_i = 1$, given all other agents' actions $\mathbf{s}_{-i}$ and network $G$, is given by

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

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

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

set of $y$. The well-known pseudo likelihood proposed by Besag [1974, 1975] and Strauss and Ikeda [1990] for the spatial processes refers to a special case where the block $A_r$ contains just a singleton.

[31] Several theoretical results on the asymptotic consistency of the composite likelihood estimation for the Gibbs measure are available in the literature [see e.g., Bhattacharya and Mukherjee, 2018; Chatterjee, 2007; Comets, 1992; Ghosal and Mukherjee, 2020]. Moreover, the simulation results in Varin et al. [2011], Zhou and Schmidler [2009], Hughes et al. [2011], and Friel [2012] show that the composite likelihood method permits reliable inference when the sample size is not too small, and the network dependence is moderate. Since our sample size is indeed very large, and the estimated network effects shown in Subsection 5.2 are not particularly large, the composite likelihood method is arguably an adequate estimation method to be used for our analysis.

Therefore, the composite likelihood of action profile $\mathbf{s}$, conditional on network $G$, is given by

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

In the LIL scenario, the composite likelihood of replacing $\tilde{\mu}(\mathbf{s}|G, \boldsymbol{\psi}^{LIL})$ is given by

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

where

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

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

$$\mu(G|\mathbf{s}) = \prod_{i=1}^{n} \prod_{j>i}^{n} \mu(a_{ij}|\mathbf{s}, G_{-ij}) = \prod_{i=1}^{n} \prod_{j>i}^{n} \frac{\exp(a_{ij}(\theta s_i s_j - \phi_0 - \sum_{k=1}^{K} h_k(x_{ik}, x_{jk})\phi_k + z_i + z_j))}{1 + \exp(\theta s_i s_j - \phi_0 - \sum_{k=1}^{K} h_k(x_{ik}, x_{jk})\phi_k + z_i + z_j)}. \tag{29}$$

Since network links are conditionally (pairwise) independent, (29) shows that the conditional likelihood of $G$ can be represented by the product of conditional probabilities of $a_{ij}$'s. Note that even though the network links are conditionally independent given the actions, they are not unconditionally independent due to the interdependence of actions in the presence of the peer effect ($\theta \neq 0$). This is an important feature of our model that distinguishes it from an *inhomogeneous random graph* model [Bollobas et al., 2007]. Although computing (29) is feasible, in the case of a large network it can still be very demanding. To further alleviate the computational burden, we adopt the so-called *case-control* approach of Raftery et al. [2012], which allows us to reduce the computational cost associated with (29) from $O(n^2)$ to $O(n)$. To explain the idea, we consider the log-likelihood function based on (29)

$$\ell(G|\mathbf{s}) = \sum_{i=1}^{n} \ell_i(a_{i\cdot}|\mathbf{s}, G_{-i}), \tag{30}$$

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

$$\begin{aligned}
\ell_i(a_{i,\cdot}|\mathbf{s}, G_{-i}) &= \sum_{j>i}^{n} a_{ij}(\theta s_i s_j - \zeta_{ij}) - \ln(1 + \exp(\theta s_i s_j - \zeta_{ij})) \\
&= \sum_{j>i, a_{ij}=1} (\theta s_i s_j - \zeta_{ij} - \ln(1 + \exp(\theta s_i s_j - \zeta_{ij}))) + \sum_{j>i, a_{ij}=0} (-\ln(1 + \exp(\theta s_i s_j - \zeta_{ij}))) \\
&= \ell_{i,1} + \ell_{i,0}.
\end{aligned} \tag{31}$$

In (31), $\ell_{i,1}$ and $\ell_{i,0}$ stand for the log-likelihood from edges and non-edges respectively. When the network links are sparse, the quantity $\ell_{i,0}$ can be viewed as a population total statistics.

This population total can be estimated by a random sample of the population,

$$\widetilde{\ell}_{i,0} = \frac{n_{i,0}}{m_{i,0}} \sum_{r=1}^{m_{i,0}} (-\ln(1+\exp(\theta s_i s_r - \zeta_{ir}))), \tag{32}$$

where $n_{i,0}$ is the total number of zero's in the $i^{th}$ row of the upper triangular part of matrix $\mathbf{A}$, and $m_{i,0}$ is the number of samples selected from zero entries in the $i^{th}$ row of the upper triangular part of the matrix $\mathbf{A}$. $\widetilde{\ell}_{i,0}$ is an unbiased estimator of $\ell_{i,0}$ given the random samples. When the network size is large, we can choose a small $m_{i,0}$ to compute $\widetilde{\ell}_{i,0}$ and thus reduce the amount of computation needed.[32]

An additional computational issue concerning the composite likelihood function of (22) is that one needs to integrate over the random effects $\mathbf{z} = (z_1, \cdots, z_n)$ in order to obtain the likelihood for observed data, i.e., $\mu(\mathbf{s}|G)\mu(G|\mathbf{s}) = \int \mu(\mathbf{s}|G,\mathbf{z})\mu(G|\mathbf{s},\mathbf{z})f(\mathbf{z})d\mathbf{z}$. The frequentist approach typically uses Gaussian quadratures or Monte Carlo integration to evaluate such a likelihood function. However, given 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 $\Theta = (\theta, \rho, \boldsymbol{\beta}^\top, \tau, \kappa, \boldsymbol{\phi}, \sigma_z^2)$ and unobserved random effects $\mathbf{z}$ with the posterior distribution $p(\Theta, z|\mathbf{s}, G)$ derived based on the composite conditional likelihoods in (22). We specify the prior distributions as follows: $\theta \sim \mathcal{N}(0, \sigma_\theta^2)$, $\rho \sim \mathcal{N}(0, \sigma_\rho^2)$, $\boldsymbol{\beta} \sim \mathcal{N}(0, \Sigma_{\boldsymbol{\beta}})$, $\tau \sim \mathcal{N}(0, \sigma_\tau^2)$, $\kappa \sim \mathcal{N}(0, \sigma_\kappa^2)$, $\boldsymbol{\phi} \sim \mathcal{N}(0, \Sigma_{\boldsymbol{\phi}})$, $z_i \sim \mathcal{N}(0, \sigma_z^2)$, and $\sigma_z^2 \sim \mathcal{IG}\left(\frac{\nu_0}{2}, \frac{\chi_0}{2}\right)$, where $\mathcal{N}$ and $\mathcal{IG}$ are normal and inverse gamma conjugate priors. We choose the hyperparameters to make the prior distributions relatively flat and cover a wide range of parameter space. Specifically, we set $\sigma_\theta^2 = \sigma_\rho^2 = \sigma_\tau^2 = \sigma_\kappa^2 = 10$, $\Sigma_\beta = 10 \cdot I_{|\boldsymbol{\beta}|}$, $\Sigma_{\boldsymbol{\phi}} = 10 \cdot I_{|\boldsymbol{\phi}|}$, $\nu_0 = 2.2$ and $\chi_0 = 0.1$. Given the above prior distributions and the composite likelihood function of (22), we can derive the joint posterior distribution of $\Theta$ and $\mathbf{z}$. Since it is difficult to simulate draws from this high-dimensional joint posterior distribution, we implement the Metropolis-Hastings-within-Gibbs algorithm to simulate draws sequentially from the conditional posterior densities in the following steps:

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

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 which will be discussed in the following section.

[32] In this study, we set $m_{i,o} = 100 + 5\sum_{j \neq i} a_{ij}$ to obtain the empirical results in Subsection 5.2 and simulation results in Subsection 5.3. To check the robustness of the empirical results with respect to the choice of $m_{i,o}$, we have also tried $m_{i,o} = 1000 + 5\sum_{j \neq i} a_{ij}$ and found the results in Supplementary Appendix Tables G.1 and G.2 are qualitatively unchanged.

### 5.2. Estimation Results

The estimation results of the GI and LIL scenarios are reported in Tables 2 and 3, respectively. In each table, column (1) presents the estimation results with individual random effects – capturing unobserved heterogeneity – and column (2) presents the results without individual random effects. Our first finding is that both the local and global interaction effects measured by $\theta$ and $\rho$ are, as expected, positive and significant in both columns. Comparing, however, the two columns we can see that the estimates of the local spillover effect ($\theta$) and other coefficients are biased when failing to control for individual unobserved heterogeneity through the inclusion of random effects. In particular, the estimate of the local spillover effect ($\theta$) in the GI scenario is upward biased by 32%; and in the LIL scenario it is upward biased by 126%; together with a 27% downward bias on the estimate of the global conformity effect in the LIL scenario. This bias stems from individual-specific factors that affect both, rioting behavior and network formation (correlated effects) that cannot be accounted for in the model without random effects [cf. Hsieh et al., 2016]. We further analyze the direction of this bias in Subsection 5.3 with artificial data by comparing the models with and without random effects. The fact that the global conformity effect ($\rho$) is significant in Table 3 provides a strong motivation for the belief-based formulation process under local information. We therefore use the LIL scenario as the benchmark model for the different counterfactuals analyzed in Section 6.

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

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

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

### 5.3. Parameter Recovery Analysis

In this section, we describe a Monte Carlo simulation study that attains a two complementary objectives. On the one hand, it demonstrates that the proposed Bayesian MCMC estimation

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

| | | with random effects (1) | w/o random effects (2) |
|---|---|---|---|
| Local spillover | ($\theta$) | 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 | ($\beta_1$) | -0.0633*** | -0.0636*** |
| | | (0.0103) | (0.0072) |
| Islamist | ($\beta_2$) | 0.1026*** | 0.1054*** |
| | | (0.0054) | (0.0046) |
| (Log) followers | ($\beta_3$) | 0.0117*** | 0.0088*** |
| | | (0.0016) | (0.0015) |
| Random effect | ($\tau$) | 0.0057*** | – |
| | | (0.0006) | |
| **Linking cost** | | | |
| Constant | ($\phi_0$) | 14.7484*** | 12.5889*** |
| | | (0.0195) | (0.0077) |
| Same gender | ($\phi_1$) | -0.1661*** | -0.1908*** |
| | | (0.0134) | (0.0082) |
| Same religiousness | ($\phi_2$) | -0.0762*** | -0.0033 |
| | | (0.0080) | (0.0058) |
| Diff. in followers count | ($\phi_3$) | 0.0871*** | 0.0986*** |
| | | (0.0033) | (0.0022) |
| Variance of random effect | ($\sigma_z^2$) | 2.1568*** | – |
| | | (0.0178) | |
| Sample size (# of nodes) | | 225,578 | |

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

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

| | | with random effects (1) | w/o random effects (2) | w/o local spillover (3) | w/o global conformity (4) |
|---|---|---|---|---|---|
| Local spillover | ($\theta$) | 0.0705*** | 0.1596*** | _ | 0.2289*** |
| | | (0.0035) | (0.0028) | | (0.0026) |
| Global conformity | ($\rho$) | 2.37e-6*** | 1.72e-6*** | 3.03e-6*** | _ |
| | | (5.03e-8) | (3.86e-8) | (3.22e-8) | |
| Weight of local observation | ($\varphi$) | 0.0961*** | 0.0805*** | 0.1066*** | _ |
| | | (0.0051) | (0.0041) | (0.0035) | |
| **Individual preference** | | | | | |
| Female | ($\beta_1$) | -0.0553*** | -0.0542*** | -0.0564*** | -0.0613*** |
| | | (0.0096) | (0.0078) | (0.0074) | (0.0081) |
| Islamist | ($\beta_2$) | 0.1136*** | 0.1178*** | 0.1104*** | 0.1060*** |
| | | (0.0061) | (0.0048) | (0.0053) | (0.0049) |
| (Log) followers | ($\beta_3$) | 0.0057*** | 0.0034** | 0.0089*** | 0.0087*** |
| | | (0.0020) | (0.0015) | (0.0016) | (0.0015) |
| Random effect | ($\tau$) | 0.0054*** | _ | _ | _ |
| | | (0.0003) | | | |
| Rioting cost | ($\kappa$) | -0.2965*** | -0.3054*** | -0.2805*** | 0.3348*** |
| | | (0.0110) | (0.0089) | (0.0091) | (0.0092) |
| **Linking cost** | | | | | |
| Constant | ($\phi_0$) | 14.7964*** | 12.5633*** | 12.5194*** | 12.5921*** |
| | | (0.0229) | (0.0093) | (0.0096) | (0.0103) |
| Same gender | ($\phi_1$) | -0.1708*** | -0.1930*** | -0.2012*** | -0.1934*** |
| | | (0.0146) | (0.0094) | (0.0092) | (0.0094) |
| Same religiousness | ($\phi_2$) | -0.0784*** | -0.0048 | 0.0065 | 0.0024 |
| | | (0.0095) | (0.0062) | (0.0063) | (0.0066) |
| Diff. in followers | ($\phi_3$) | 0.0876*** | 0.0992*** | 0.1003*** | 0.0984*** |
| | | (0.0031) | (0.0025) | (0.0026) | (0.0028) |
| Variance of random effect | ($\sigma_z^2$) | 2.2477*** | _ | _ | _ |
| | | (0.0211) | | | |
| Sample size (# of nodes) | | 225,578 | | | |

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

based on the composite likelihood function of (22) and the case-control approach described in Subsection 5.1 can indeed recover true parameter values. As an interesting side effect of this exercise, we also confirm the direction of estimation bias observed in Tables 2 and 3 when individual unobserved heterogeneity (random effect) is ignored. On the other hand, the Monte Carlo simulations also provide support to the conjecture that underlay our theoretical analysis of the LIL scenario. More specifically, it shows that limit distribution $\tilde{\mu}\eta(\cdot)$ defined in (16) provides a good approximation of the invariant behavior of the process under limited information and therefore can be suitably use as likelihood for estimation in this context.

We set the number of Monte Carlo repetitions to 300. In each repetition, we generate the artificial networks $G_t$ and the action profiles $\mathbf{s}_t$ through a data generating process (DGP) that mimics the dynamic process introduced in Subsection 2.2 for a setup with a number of nodes $n = 3,000$. In such a DGP, we generate the individual types $\gamma_i$ from a expression of the form $\beta x_i + \tau z_i$, where the variables $x_i$ represent observed individual characteristics that are generated from a mixture of normal distributions (specifically, two-fifths of the values are generated from a $\mathcal{N}(-4, 36)$ and three-fifths from $\mathcal{N}(4, 36)$), while the variables $z_i$ represent unobserved individual random effects generated from a standard normal distribution, i.e. $z_i \sim \mathcal{N}(0, 1)$. The coefficients $\beta$ and $\tau$ are set to 0.5. Then, the linking costs are determined through the expression $\zeta_{ij} = \phi_0 + h(x_i, x_j)\phi_1 - z_i - z_j$, where we set $h(x_i, x_j) = |x_i - x_j|$ and the coefficients $\phi_0$ and $\phi_1$ are fixed equal to 2 and 1, respectively. The true values of the local spillover effect $\theta$ and the global conformity effect $\rho$ are set to 0.05 and 0.001 respectively, and the parameter $\eta$ of the logistic disturbance is normalized to 1. Finally, for the LIL scenario, we fix the true rioting cost parameter $\kappa$ to 1.5 and the belief weight $\varphi$ to 0.5.

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

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

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

And, in order to speed up the simulation, we rely on the conditional independence of network links,[33] to update the choice of all $i$'s links $\{a_{ij}\}_{j=1}^n$ synchronously. On the other hand, to update the action choice $s_i$ of any given individual $i$, the details of course depend on the scenario under consideration. For the GI scenario, we use the following conditional probability for each action $s_i$:

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

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

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

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

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

For estimation, we implement the Bayesian MCMC sampling with 30,000 iterations and drop the first 5,000 iterations in the burn-in phase. The results for the GI scenario are summarized in Table 4 and the results for the LIL scenario 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 the estimation exercise can successfully recover the true parameter values when considering the full econometric model with random effects (which is the true DGP). Moreover, in line with the pattern observed in our empirical study, when random effects are ignored the estimate of the local spillover $\theta$ is upward biased whereas that of the global-conformity parameter $\rho$ is downward biased.

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

| | | | with random effects | | w/o random effects | |
|---|---|---|---|---|---|---|
| | | DGP | Mean | Std. | Mean | Std. |
| Local spillover | ($\theta$) | 0.0500 | 0.0593 | 0.0108 | 0.1451 | 0.0072 |
| Global conformity | ($\rho$) | 0.0010 | 0.0009 | 0.0004 | -0.0003 | 0.0002 |
| **Individual preference** | | | | | | |
| Individual characteristic $x$ | ($\beta$) | 0.5000 | 0.5059 | 0.1285 | 0.4155 | 0.0677 |
| Random effect | ($\tau$) | 0.5000 | 0.4510 | 0.1911 | – | – |
| **Linking cost** | | | | | | |
| Constant | ($\phi_0$) | 2.0000 | 2.0145 | 0.0666 | 1.5291 | 0.0091 |
| Diff. in characteristics $\lvert x_i - x_j \rvert$ | ($\phi_1$) | 1.0000 | 1.0145 | 0.0064 | 0.8808 | 0.0034 |
| Variance of random effect | ($\sigma_z^2$) | 1.0000 | 1.1045 | 0.0353 | – | – |
| Number of nodes | | | | 3000 | | |
| MC repetitions | | | | 300 | | |

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

| | | | with random effects | | w/o random effects | |
|---|---|---|---|---|---|---|
| | | DGP | Mean | Std. | Mean | Std. |
| Local spillover | $(\theta)$ | 0.0500 | 0.0543 | 0.0127 | 0.0740 | 0.0067 |
| Global conformity | $(\rho)$ | 0.0010 | 0.0013 | 0.0005 | 0.0007 | 0.0003 |
| Weight of local obsev. | $(\varphi)$ | 0.5000 | 0.6002 | 0.1833 | 0.6197 | 0.1685 |
| **Individual preference** | | | | | | |
| Individual characteristic $x$ | $(\beta)$ | 0.5000 | 0.5439 | 0.2022 | 0.5441 | 0.2020 |
| Random effect | $(\tau)$ | 0.5000 | 0.5596 | 0.2631 | – | – |
| Rioting cost | $(\kappa)$ | 1.5000 | 1.6224 | 0.7728 | 1.8748 | 0.7777 |
| **Linking cost** | | | | | | |
| Constant | $(\phi_0)$ | 2.0000 | 1.9493 | 0.0660 | 1.4656 | 0.0084 |
| Diff. in characteristics $\lvert x_i - x_j \rvert$ | $(\phi_1)$ | 1.0000 | 1.0138 | 0.0057 | 0.8845 | 0.0032 |
| Variance of random effect | $(\sigma_z^2)$ | 1.0000 | 1.0998 | 0.0351 | – | – |
| Number of nodes | | | | 3000 | | |
| MC repetitions | | | | 300 | | |

# 6. Counterfactual Analyses

Building upon the estimation results obtained for the LIL scenario in Subsection 5.2, we have performed various counterfactual analyses in order to investigate the effect of changing specific single parameters of the model, while keeping the remaining ones at their estimated values in column (1) of Table 3. The impact of these changes on the different outcome variables is assessed through the corresponding invariant distribution $\mu(\cdot)$ that are generated from the same iterative process described in Subsection 5.3. In what follows, we describe our results for two specific implementations. In the first one, we examine the role of linking costs in affecting rioting behavior. In the second, we study how biasing beliefs towards a specific action can influence the extent of rioting.

## 6.1. Linking Costs and Rioting Behavior

During the protests against the Egyptian government, mobile phone operators were instructed to suspend services in selected areas, with internet access being blocked and mobile phone and text messaging services disabled or working only sporadically [Kravets, 2011]. These interventions by the government were aimed at suppressing protests by making it harder for people to communicate and coordinate via online social media. Similar interventions have also been undertaken in other countries (e.g. China), where social media platforms such as Facebook, YouTube, and Twitter were domestically blocked, so that users could use them only indirectly via a VPN service [Willnat et al., 2015]. We can operationalize such attempts at hampering communication as an increase in the linking cost and explore what is its effect on rioting behavior, according to our estimated model.

For our counterfactual exercise we change the value of the estimate of cost parameter $\phi_0$, which is the constant term in the linking cost in (21), over the range of [-20%, 20%] with 11 evenly distributed grid points. For each grid point, we simulate the network and action profiles 300 times and compute the average fraction of rioting agents together with the average network degree in the new equilibrium. The right panel in Figure 9 shows the average degree over varying

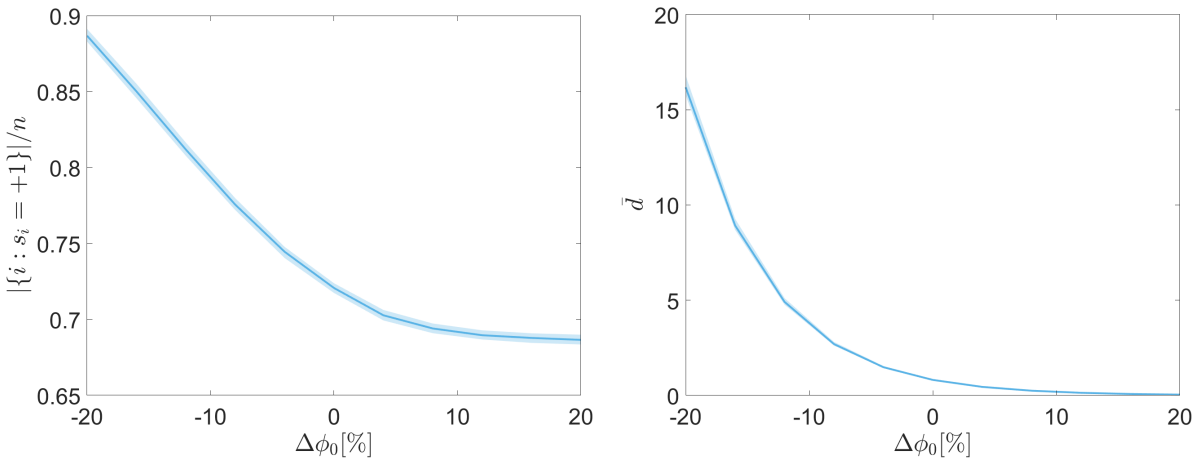



Figure 9: Changes in the fraction of rioting agents, $|\{i : s_i = +1\}|/n$ (left panel), and the average degree, $\bar{d}$ (right panel), over percentage changes in the linking cost $\phi_0$. We plot the mean and 95% interval from 300 simulation repetitions.

$\phi_0$. As expected, we can see that the linking cost $\phi_0$ has a substantial effect on the average degree so lowering the linking cost gives rise to a much denser network structure. The left panel in Figure 9 shows the fraction of rioting agents over varying $\phi_0$. We observe that a reduction of $\phi_0$ by 20% yields an increase in the fraction of rioting agents by 15%. Conversely, this indicates that as linking and exchanging information via the network becomes more costly (e.g., by interrupting or blocking social media), fewer links are being formed, coordination among agents becomes more difficult, and fewer agents participate in the protest as a consequence. This finding illustrates and quantifies the importance of the role of online social networks in the formation of protest movements or riots and the emergence of collective action.

### 6.2. Belief Manipulation and Rioting Behavior

Governments often use manipulation of the information available on social or other media to distort the users' view [cf. Edmond, 2013; Zhuravskaya et al., 2020]. For example, King et al. [2017] document the massive effort of the Chinese government to post content on social media that is mainly devoted to supporting positive views about the state. Similar efforts have been documented in Egypt [El-Khalili, 2013].

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 in (6) as follows:

$$p_{it}^{u+1} = (1-\Psi)\left\{\varphi\frac{1}{d_{it}}\sum_{j=1}^{n} a_{ij,t}s_{jt} + (1-\varphi)\frac{1}{d_{it}+1}\left[p_{it}^{u} + \sum_{j=1}^{n} a_{ij,t}p_{jt}^{u}\right]\right\} + \Psi g, \tag{33}$$

where $\Psi \in [0,1]$ and $g = -1$ is supposed to be the preferred action of the government (no rioting). As in the previous section, for each of the points in an evenly distributed grid for $\Psi$ in the range of [0,1], we simulate repeatedly the evolutionary process and then compute the average fraction of rioting agents and the average network degree resulting from the induced invariant distributions. The corresponding changes in those magnitudes for varying levels of the

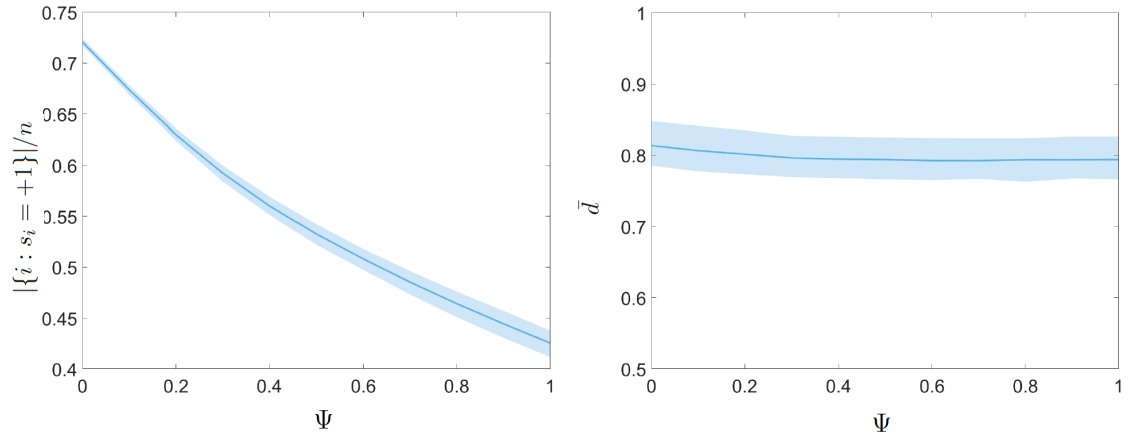



Figure 10: Changes in the fraction of rioting agents, $|\{i : s_i = +1\}|/n$ (left panel), and the average degree, $\bar{d}$ (right panel), over the strength of propaganda $\Psi$. We plot the mean and 95% interval from 300 simulation repetitions.

influence of propaganda, $\Psi$, are plotted in Figure 10. Our results show that while propaganda (belief manipulation) does not affect the network density – which remains roughly stable – it has a very significant effect on rioting, reducing the fraction of rioting agents by up to 30%. It does not succeed, however, to reduce rioting below 40% of the population, even when propaganda is the only source of belief updating (i.e. when $\Psi = 1$). In line with the recent literature on the phenomenon [cf. Azzimonti and Fernandes, 2022; Gu et al., 2017], these findings illustrate the effect that the manipulation of information may have on the formation of collective action, but also point to its limitations.

## 7. Conclusion

In this paper we have introduced a model of collective action (e.g. protest participation) in a large population where the social network co-evolves with actions and beliefs. We provide a complete characterization of the equilibrium action choices, beliefs, and networks, and show that there are conditions in the environment (in particular, on action costs and the profile of individual characteristics) under which a significant degree of collection arises in the long-run equilibrium of the model. We also show that those conditions are substantially affected by the assumptions being made on the extent of information the agents have about the state of the system. Somewhat paradoxically, they turn out to be considerably stronger (and hence less plausible) if one assumes that the population is fully informed about the current state than under the more realistic assumption that the agents have only local information and also learn from their network neighbors. In the second part of the paper, we bring our model to the data by relying on large-scale Twitter data during the Arab Spring on social unrest in Egypt. We perform a structural estimation of the model parameters and use random effects to capture the unobserved component in the idiosyncratic preferences of the agents for retaining the *status quo*. We jointly estimate (and disentangle) the local peer effect and the global force towards conformity effect in the agents' decisions, and show that both are significant. Moreover, we find that ignoring endogeneity in the network formation process may bias the estimates of these two effects. It must be acknowledged, however, that while our empirical analysis is made possible due to the availability of massive data gathered from social media, it is also limited by the fact that the information gathered about online behavior on these platforms translate

only imperfectly into the actions materializing (e.g. in actual protest participation) in the real world.

## 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.

Abdul-Mageed, M. and Elmadany, A, and Nagoudi, E.M.B., (2022). ARBERT & MARBERT: Deep Bidirectional Transformers for Arabic *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing* , 7088–7105.

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.

ACLED (2019). Armed Conflict Location and Event Dataset (ACLED). *Center for the Study of Civil War, International Peace Research Institute, Oslo (PRIO)*.

Aldayel, A., and Magdy, W. (2021). Stance detection on social media: State of the art and trends *Information Processing & Management*, 58(4).

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. 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.

Ash, E. and Hansen, S. (forthcoming). Text Algorithms in Economics. *Annual Review of Economics*.

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

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

Bana, S.H. (2022). work2vec: Using Language Models to Understand Wage Premia. *Working Paper*.

Barbera, P. (2015). Birds of the Same Feather Tweet Together: Bayesian Ideal Point Estimation Using Twitter Data. *Political Analysis* 23, 76-91.

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

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

Besag, J. (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. *The 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.

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.

Conneau, A, Khandelwal, K., Goyal, N., Chaudhary, V., Wenzek, G., Guzman, F., Grave, E., Ott, M., Zettlemoyer., and Stoyanov (2020). Unsupervised Cross-lingual Representation Learning at Scale. *Arxiv print*, arXiv:1911.02116.

Currarini, S., Jackson, M., and Pin, P. (2009). An Economic Model of Friendship: Homophily, Minorities and Segregation. *Econometrica*, 77(4):1003–1045.

Darwish, K., Magdy, W., Rahimi, A., Baldwin, Y., and Abokhodair (2018). Predicting Online Islamaphobic Behaviour After the Paris Attacks *The Journal of Web Science*.

DeGroot, M. H. (1974). Reaching a consensus. *Journal of the American Statistical Association*, 69(345):118–121.

DeMarzo, P., Vayanos, D., and Zwiebel, J. (2003). Persuasion Bias, Social Influence, and Unidimensional Opinions. *Quarterly Journal of Economics*, 118(3):909–968.

Devlin, J., Chang, M.W., Lee, K., and Toutanova, K. (2018). BERT: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint*, arXiv:1810.04805.

Dzemski, A. (2019). An empirical model of dyadic link formation in a network with unobserved heterogeneity. *The Review of Economics and Statistics*, 101(5):763 – 776.

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. *The 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).

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 Bramoulle, Y., A. Galeotti, and B. Rogers (eds.), *Oxford Handbook of Economic Networks*, Oxford: Oxford University Press.

González-Bailón, S., Borge-Holthoefer, J., Rivero, A., and Moreno, Y. (2011). The dynamics of

protest recruitment through an online network. *Scientific Reports*, 1(1):1–7.
Goodman, Leo A. (1961). Snowball sampling. *The Annals of Mathematical Statistics*, 148-170.
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.S. (2013). Text as data: The promise and pitfalls of automatic content analysis methods for political texts. *Political Analysis*, 21(3):26–297.
Grimmer J., Roberts, M., and Stewart, B. (2022). Text as data: A New Framework for Machine Learning and the Social Sciences. *Princeton University Press*.
Grimmett, G. (2010). *Probability on Graphs*. Cambridge University Press.
Gu, L., Kropotov, V., and Yarochkin, F. (2017). The fake news machine. How propagandists abuse the Internet and manipulate the public. Pobrane. Trend Micro Incorporated.
Hinds, J., and Joinson, A.N (2018). What demographic attributes do our digital footprints reveal? A systematic review. *PLOS One*, 13(11).
Hsieh, C.-S. and L.F. Lee (2016). A social interactions model with endogenous friendship formation and selectivity. *Journal of Applied Econometrics*, 31(2):301–319.
Hsieh, C.-S., Lee, L.-f., and Vincent, B. (2020). Specification and estimation of network formation and network interaction models with the exponential probability distribution. *Quantitative Economics*, 11(4):1349–1390.
Hsieh, C.-S., König, M. D. and Liu, X. (2022) A Structural Model for the Coevolution of Networks and Behavior *The Review of Economics and Statistics*, 104(2):355–367.
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.
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.
Kudo, K., and Richardson J. (2018). Sentencepiece: A simple and language independent subword tokenizer and detokenizer for neural text processing. *arXiv preprint*, arXiv:1808.06226.
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/.
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). #ISISisNotIslam 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.
Pauli, F., Racugno, W., and Ventura, L. (2011). Bayesian composite marginal likelihoods. *Statistica Sinica*, 149–164.
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.
Qiu, M., Sim., Y., Smith, N.A., Jiang, J. (2015). Modelling User Arguments, Interations and Attributes for Stance Prediction in Online Debate Forums. *Proceeedings of the 2015 SIAM International Conference on Data Mining*, 855–863.
Raftery, A. E., Niu, X., Hoff, P. D., and Yeung, K. Y. (2012). Fast inference for the latent space network model using a case-control approximate likelihood. *Journal of Computational and Graphical Statistics*, 21(4):901–919.
Rajadesingan, A., Liu, H. (2014). Identifying Users with Opposing Opinions in Twitter Debates. *International Conferece on Social Computing, Behavioural-Cultural Modelling and Prediction*, 153–160.
Rao, D., Yaraowsky, D., Shreevats, A., and Gupta, M. (2010). Classifying Latent User Attributes in Twitter. *Proceedings of the 2nd International Workshop on Search and Mining User Generated Contents*, 37–44.
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.
Shi, Z., Rui, H. and Whinston, A.B. (2014). Content Sharing in A Social Broadcasting Environment *MIS Quarterly*, 38:123–142.
Strauss, D. and Ikeda, M. (1990). Pseudolikelihood estimation for social networks *Journal of the American Statistical Association*, 85(409):204–212.
Tamkin, A, Singh, T., Giovanardim D., and Goodman., N. (2020). *Investigating transferability in pretrained language models*,. arXiv preprint arXiv:2004.14975.
Tullock, G. (1971). The paradox of evolution *Public Choice* 11, 89–99.
Varin, C., Reid, N. and Firth, D. (2011). An overview of composite likelihood methods. *Statistica Sinica*, 21(1):5–42.
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.

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.
Zhang, T., Wu, F., Katiyar, A., Weinberger, K.Q., and Artzi, Y. (2020). *Revisiting few-sample BERT fine-tuning*,. arXiv preprint arXiv:2006.05987.
Zhang, C., and Abdul-Mageed, M. (2019). *BERT-Based Social Media Author Profiling*,. arXiv preprint arXiv:1909.04181.
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. 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 (12)-(13), 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 $\Omega$ is the state space and $\mathcal{F}$ the suitable smallest $\sigma$-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 $\Delta t$. 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:

$$q(\boldsymbol{\omega}, \boldsymbol{\omega}') = \begin{cases} \chi \frac{e^{\eta\Phi(s_i, \mathbf{s}_{-i}, G)}}{e^{\eta\Phi(s_i, \mathbf{s}_{-i}, G)} + e^{\eta\Phi(s_i', \mathbf{s}_{-i}, G)}} & \text{if } \boldsymbol{\omega}' = (s_i', \mathbf{s}_{-i}, G) \text{ and } \boldsymbol{\omega} = (\mathbf{s}, G), \\ \lambda \frac{e^{\eta\Phi(\mathbf{s}, G+ij)}}{e^{\eta\Phi(\mathbf{s}, G+ij)} + e^{\eta\Phi(\mathbf{s}, G)}} & \text{if } \boldsymbol{\omega}' = (\mathbf{s}, G + ij) \text{ and } \boldsymbol{\omega} = (\mathbf{s}, G), \\ \lambda \frac{e^{\eta\Phi(\mathbf{s}, G-ij)}}{e^{\eta\Phi(\mathbf{s}, G-ij)} + e^{\eta\Phi(\mathbf{s}, G)}} & \text{if } \boldsymbol{\omega}' = (\mathbf{s}, G - ij) \text{ and } \boldsymbol{\omega} = (\mathbf{s}, G), \\ -\sum_{\boldsymbol{\omega}' \neq \boldsymbol{\omega}} q(\boldsymbol{\omega}, \boldsymbol{\omega}') & \text{if } \boldsymbol{\omega}' = \boldsymbol{\omega}, \\ 0 & \text{otherwise.} \end{cases} \tag{34}$$

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. 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 (34). This equation is satisfied when the probability distribution $\mu^\eta(\boldsymbol{\omega})$ satisfies the detailed balance condition [cf. Norris, 1998]

$$\mu^\eta(\boldsymbol{\omega}) q(\boldsymbol{\omega}, \boldsymbol{\omega}') = \mu^\eta(\boldsymbol{\omega}') q(\boldsymbol{\omega}', \boldsymbol{\omega}), \tag{35}$$

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

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

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

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

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

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

Hence, the probability measure $\mu^\eta(\boldsymbol{\omega})$ satisfies a detailed balance condition of (35) and therefore is the stationary distribution of the Markov chain with transition rates $q(\boldsymbol{\omega}, \boldsymbol{\omega}')$. □

*Proof of Propositions 3 and 4.* 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 (2) the potential function can be written as

$$\Phi(\mathbf{s}, G) = \sum_{i=1}^n \left( \gamma_i + \frac{\rho}{2} \sum_{j \neq i}^n s_j - \kappa \right) s_i + \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n a_{ij} \left( \theta s_i s_j - \zeta_1 + \frac{\zeta_1 - \zeta_2}{2} (1 - \gamma_i \gamma_j) \right). \tag{36}$$

Note that only the last term in (36) 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, $\overline{K}_n$. To summarize, the candidate networks and action profiles that maximize the potential must be either complete, $K_n$, empty, $\overline{K}_n$, or composed of two disconnected cliques, $K_{n_1} \cup K_{n-n_1}$, in which all agents in

the same clique chose the same action and have the same idiosyncratic preferences.

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

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

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

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

Moreover,

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

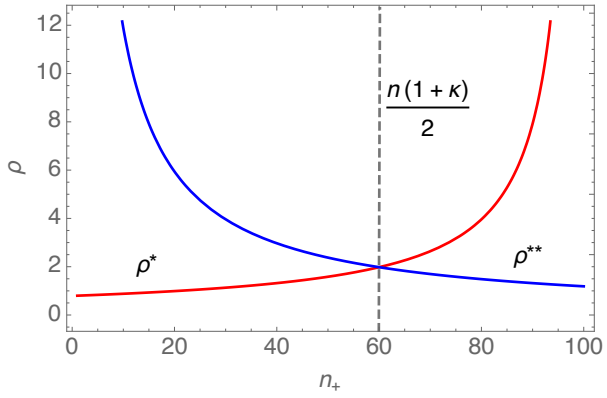



Figure 11: The two thresholds $\rho^*$ and $\rho^{**}$ as a function of $n_+$. The dashed line indicates where they coincide, i.e., $\rho^* = \rho^{**}$, which happens for $\nu_+ = \frac{n_+}{n} = \frac{1+\kappa}{2}$.

which is positive if $\rho$ is larger than

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

Similarly, solving for $\kappa$ 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 $\kappa$ is smaller than

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

Note that $\kappa^*$ 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 12 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

$$\begin{aligned}
&\Phi(\mathbf{s}', K_{n_+-k} \cup K_{n-n_++k}) - \Phi(\boldsymbol{\gamma}, K_{n_+} \cup K_{n-n_+}) \\
&\quad = \frac{k}{2}\left(2\zeta_1 n_+ + \zeta_2 - \theta(1 + 2n_+) - 4(1-\kappa) - k(2\zeta_1 + \zeta_2 - 3\theta) + 4\rho(k + n - 2n_+)\right). \quad (37)
\end{aligned}$$

Note that this is (under some regularity conditions) a convex function of $k$ (see Figure 12).[34] 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 $\theta$. Solving $\Phi((-1,\ldots,-1), K_n) = \Phi(\boldsymbol{\gamma}, K_{n_+} \cup K_{n-n_+})$ for $\theta$ yields the 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

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

We then have that

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

which is negative for

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

[34] 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.

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 increasing in $\theta$. 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 $\rho>\rho^*$.

Next we analyze 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

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

Further, we have that

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

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

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

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 $\theta$. 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, $\overline{K}_n$.

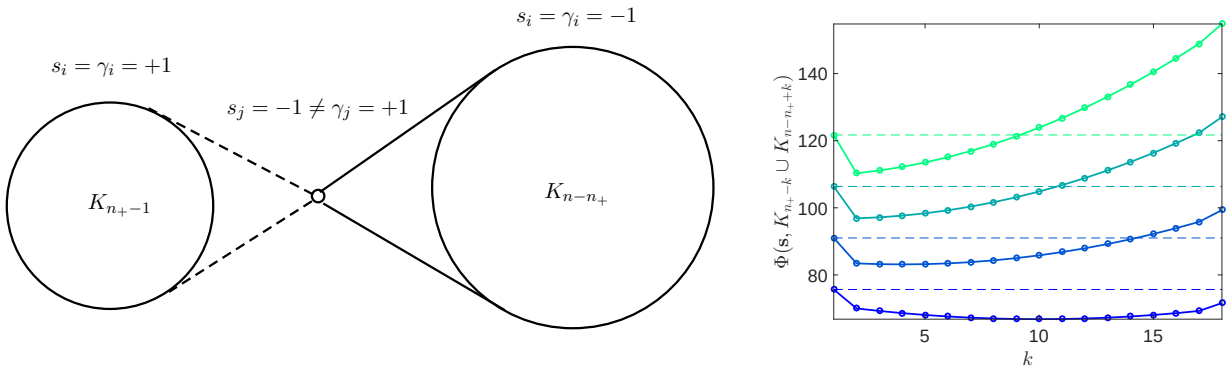



Figure 12: (Left panel) Illustration of two cliques, $K_{n+}$ and $K_{n-n_+}$ and the relocation of one node $j$ from $K_{n+}$ to $K_{n-n_+}$. (Right panel) The resulting potential for relocating node $j$ from the clique $K_{n_+}$ to the clique $K_{n-n_+}$ for $\theta \in \{0.05, 0.075, 0.1, 0.125\}$, $n_+ = 17$, $n = 50$, $\rho = 0$ 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 $\theta$ 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 $K_{n_+} \cup K_{n-n_+}$.

Next, note that

$$\Phi((-1, \ldots, -1), K_n) - \Phi((+1, \ldots, +1), K_n) = 2(n(\kappa + 1) - 2n_+), \tag{38}$$

which is increasing in $\kappa$. For $n_+ < \frac{n}{2}$ (38) is strictly positive for any value of $\kappa$. 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 $\kappa^*$ 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 and 6.* In the following we compute an absorbing state of the Markov process formalizing the LIL scenario in the limit of $\eta \to \infty$ characterizing the stochastically stable states. In such an absorbing state $(\mathbf{s}, G, \mathbf{p})$, given the beliefs $\mathbf{p}$ agents do not have an incentive to change their actions, $\mathbf{s}$, or links, $G$. Because differences in the potential correspond to differences in payoffs, this is satisfied if the potential is maximized for such $(\mathbf{s}, G)$ given the beliefs $\mathbf{p}$. Conversely, given $(\mathbf{s}, G)$, the stationary belief updating in (8) must be satiesfied, that is, $p_i = f_i(\mathbf{s}, \mathbf{p}, G) = \varphi \frac{1}{d_i} \sum_{j=1}^n a_{ij} s_j + (1 - \varphi) \frac{1}{d_i + 1} (p_i + \sum_{j=1}^n a_{ij} p_j)$ for all $i = 1, \ldots, n$. We then proceed by a guess and verify approach to check that the conditions for such a fixed point are satisfied.

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

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

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 (9) we know that the stationary beliefs satisfy the following relationship

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

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

$$s_i = \begin{cases} +1, & \text{if } \rho > (1+\kappa)/(n-1) \xrightarrow[n\to\infty]{} 0, \\ -1, & \text{for any vales of } \rho \text{ and } \kappa. \end{cases} \tag{39}$$

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

$$s_i = \begin{cases} +1, & \text{if } \rho > (\kappa-1)/(n-1) \xrightarrow[n\to\infty]{} 0, \\ -1, & \text{if } \rho > (1-\kappa)/(n-1) \xrightarrow[n\to\infty]{} 0. \end{cases} \tag{40}$$

Further, from (7) we know that the stationary beliefs must satisfy $p_i = \varphi\frac{1}{d_i}\sum_{j=1}^{n} a_{ij}s_j + (1-\varphi)\frac{1}{d_i+1}(p_i + \sum_{j=1}^{n} a_{ij}p_j)$. In a network where all connected agents choose the same action and have the same beliefs, this simplifies to $p_i = \varphi s_i + (1-\varphi)p_i$, 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

$$\zeta_{ij} = \zeta_1 - \frac{\zeta_1 - \zeta_2}{2}(1 - \gamma_i\gamma_j) = \begin{cases} \zeta_1, & \text{if } \gamma_i = \gamma_j, \\ \zeta_2, & \text{if } \gamma_i \neq \gamma_j, \end{cases} \tag{41}$$

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

$$\tilde{\Phi}(\mathbf{s}, G) = \sum_{i=1}^{n} (\gamma_i + \rho(n-1)s_i - \kappa)\, s_i + \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n} a_{ij}\left(\theta s_i s_j - \zeta_1 + \frac{\zeta_1 - \zeta_2}{2}(1 - \gamma_i\gamma_j)\right). \quad (42)$$

Note that only the last term in (42) depends on the network (through the entries of the adjacency matrix elements $a_{ij}$). In particular, the term $\sum_{i=1}^{n}\sum_{j=1}^{n} a_{ij}s_i s_j$ is maximized over $s_i, s_j \in \{-1, +1\}$ for $a_{ij} = 1$ iff $s_i = s_j$. The term $\sum_{i=1}^{n}\sum_{j=1}^{n} a_{ij}(\theta s_i s_j - \zeta_1 + (\zeta_1 - \zeta_2)(1 - \gamma_i\gamma_j)/2$ is maximized over $s_i, s_j \in \{-1, +1\}$ for $a_{ij} = 1$ iff $s_i = s_j = \gamma_i = \gamma_j$ if $\zeta_1 < \theta < \zeta_2$ and $s_i = s_j$ if $\zeta_2 < \theta$. If $\theta < \zeta_1$ then $a_{ij} = 0$ and we obtain the empty network, $\overline{K}_n$. To summarize, the candidate networks and action profiles that maximize the potential must be either complete, $K_n$, empty, $\overline{K}_n$, or composed of two disconnected cliques, $K_{n_1} \cup K_{n-n_1}$, in which all agents in the same clique chose the same action and have the same idiosyncratic preferences.

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 agents 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 (40) is trivially satisfied with $\kappa > 1$ as $\rho \geq 0$. The same holds 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 $\theta$. Solving $\tilde{\Phi}((-1, \ldots, -1), K_n) = \tilde{\Phi}(\boldsymbol{\gamma}, K_{n_+} \cup K_{n-n_+})$ for $\theta$ 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 $\theta$. 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 (40) 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 $\theta$. Solving $\Phi((+1,\ldots,+1),K_n) = \tilde{\Phi}(\boldsymbol{\gamma}, K_{n_+} \cup K_{n-n_+})$ for $\theta$ 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 $\theta$. 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

$$\tilde{\Phi}((-1,\ldots,-1),K_n) - \tilde{\Phi}((+1,\ldots,+1),K_n) = 2(n(\kappa+1) - 2n_+), \tag{43}$$

which is increasing in $\kappa$. For $n_+ < \frac{n}{2}$ (43) is strictly positive for any value of $\kappa$. In contrast, for $n_+ > \frac{n}{2}$ we have that

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

if

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

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

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

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. □

# Supplementary Appendix for "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,e] and Fernando Vega-Redondo[f]

[a] *Department of Economics, National Taiwan University, Taipei, Taiwan.*
[b] *Tilburg University, Warandelaan 2, 5037 AB Tilburg, The Netherlands.*
[c] *Department of Spatial Economics, VU Amsterdam, De Boelelaan 1105, 1081HV Amsterdam, The Netherlands.*
[d] *ETH Zurich, Swiss Economic Institute (KOF), Zurich, Switzerland.*
[e] *Centre for Economic Policy Research (CEPR), London, United Kingdom.*
[f] *Department of Decision Sciences, Bocconi University, Via Roentgen, 1, 20136 Milan, Italy.*

## Contents

# B. Extensions

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

## B.1. Directed Links

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

$$\Phi(\mathbf{s}, G) = \sum_{i=1}^{n} \gamma_i s_i + \theta \sum_{i=1}^{n} \sum_{j=1}^{n} \left( \frac{1}{2} a_{ij} a_{ji} + a_{ij}(1 - a_{ji}) \right) s_i s_j + \frac{\rho}{2} \sum_{i=1}^{n} \sum_{j \neq i}^{n} s_i s_j - \kappa \sum_{i=1}^{n} s_i - m\zeta. \tag{B.44}$$

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

## B.2. Action-Specific Heterogeneous Linking Costs

Consider a linking cost between agents $i$ and $j$ given by

$$\zeta_{ij} = \zeta_1 d_i - \zeta_2 \sum_{j=1}^{n} a_{ij}(1 - s_i s_j)$$

that allows for linking costs to be lower between agents choosing the same strategy. The corresponding payoff function in the GI environment is given by

$$\pi_i(\mathbf{s}, G) = \gamma_i s_i + (\theta + \zeta_2) \sum_{j=1}^{n} a_{ij} s_i s_j + \rho \sum_{j=1}^{n} s_j s_i - \kappa s_i - \zeta_1 d_i.$$

This is the same functional form as in (1) up to a shift of the parameter $\theta$.

# C. 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, the average connectivity (network degree) and the average action, as they depend on the linking costs $\zeta$ ($= \zeta_1 = \zeta_2$) and the noise parameter $\eta$.

## C.1. Global Information

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

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

$$\mathbb{E}^\eta(m) = \frac{1}{\mathcal{Z}^\eta} \sum_{k=0}^{n} \sum_{j=0}^{\min\{k,n_+\}} \binom{n_+}{j} \binom{n-n_+}{k-j} e^{\eta(2k-n)} e^{\eta\left(\rho\left(2l(k,j)-\binom{n}{2}\right)\right)-\kappa(n-2(n_++k-2j))} \\ \times \left(1+e^{\eta(\theta-\zeta)}\right)^{l(k,j)} \left(1+e^{-\eta(\theta+\zeta)}\right)^{\binom{n}{2}-l(k,j)} \left(\frac{l(k,j)}{1+e^{-\eta(\theta-\zeta)}} + \frac{\binom{n}{2}-l(k,j)}{1+e^{\eta(\theta+\zeta)}}\right), \quad \text{(C.45)}$$

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

$$l(k,j) = \frac{n^2 + (2(2j-k)-1)n + 2(2j-k)^2 - 2(n+2(2j-k)-n_+)n_+}{2}, \quad \text{(C.46)}$$

*$n_+ = \#(\{\gamma_i = 1 : i = 1, \ldots, n\})$, and we have that $\lim_{\zeta\to\infty} \mathbb{E}^\eta(m) = 0$.*[35]

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

**Lemma C.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*

$$\mu^\eta(\mathbf{s}) = \frac{1}{\mathcal{Z}^\eta} e^{\eta \mathcal{H}^\eta(\mathbf{s})}, \quad \text{(C.47)}$$

*where we have denoted by*

$$\mathcal{H}^\eta(\mathbf{s}) \equiv \sum_{i=1}^{n} \left( \left( \gamma_i - \kappa + \frac{\rho}{2} \sum_{j\neq i}^{n} s_j \right) s_i + \sum_{j=i+1}^{n} \left( \frac{1}{\eta} \ln\left(1 + e^{\eta(\theta s_i s_j - \zeta)}\right) \right) \right), \quad \text{(C.48)}$$

*and the normalizing constant is given by*

$$\mathcal{Z}^\eta = \sum_{\mathbf{s}\in\{-1,+1\}^n} e^{\eta \mathcal{H}^\eta(\mathbf{s})}. \quad \text{(C.49)}$$

*Proof of Lemma C.1.* We first compute the *partition function* [cf. Grimmett, 2010; Wainwright

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

and Jordan, 2008], which appears as the denominator in (14), explicitly. We have that[36]

$$\begin{aligned}
\mathcal{Z}^{\eta} &\equiv \sum_{G\in\mathcal{G}^n}\sum_{\mathbf{s}\in\{-1,+1\}^n} e^{\eta\Phi(\mathbf{s},G)} \\
&= \sum_{\mathbf{s}\in\{-1,+1\}^n}\sum_{G\in\mathcal{G}^n} e^{\eta\left(\sum_{i=1}^n(\gamma_i-\kappa)s_i+\frac{\rho}{2}\sum_{i=1}^n\sum_{j\neq i}^n s_is_j+\sum_{i=1}^n\sum_{j=i+1}^n a_{ij}(\theta s_is_j-\zeta)\right)} \\
&= \sum_{\mathbf{s}\in\{-1,+1\}^n} e^{\eta\sum_{i=1}^n\left(\gamma_i-\kappa+\frac{\rho}{2}\sum_{j\neq i}^n s_j\right)s_i}\sum_{G\in\mathcal{G}^n} e^{\eta\sum_{i=1}^n\sum_{j=i+1}^n a_{ij}(\theta s_is_j-\zeta)} \\
&= \sum_{\mathbf{s}\in\{-1,+1\}^n} e^{\eta\sum_{i=1}^n\left(\gamma_i-\kappa+\frac{\rho}{2}\sum_{j\neq i}^n s_j\right)s_i}\prod_{i=1}^n\prod_{j=i+1}^n\left(1+e^{\eta(\theta s_is_j-\zeta)}\right),
\end{aligned} \tag{C.50}$$

where we have used the fact that

$$\sum_{G\in\mathcal{G}^n} e^{\sum_{i<j}^n a_{ij}\sigma_{ij}} = \prod_{i=1}^n\prod_{j=i+1}^n\left(1+e^{\sigma_{ij}}\right), \tag{C.51}$$

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

$$\mathcal{H}^{\eta}(\mathbf{s}) \equiv \sum_{i=1}^n\left(\left(\gamma_i-\kappa+\frac{\rho}{2}\sum_{j\neq i}^n s_j\right)s_i+\sum_{j=i+1}^n\left(\frac{1}{\eta}\ln\left(1+e^{\eta(\theta s_is_j-\zeta)}\right)\right)\right), \tag{C.52}$$

we can write the partition function as follows

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

With the Hamiltonian we can write the marginal distribution as follows

$$\begin{aligned}
\mu^{\eta}(\mathbf{s}) &= \frac{1}{\mathcal{Z}^{\eta}}\sum_{G\in\mathcal{G}^n} e^{\eta\Phi(\mathbf{s},G)} \\
&= \frac{1}{\mathcal{Z}_n^{\eta}} e^{\eta\sum_{i=1}^n\left(\gamma_i-\kappa+\frac{\rho}{2}\sum_{j\neq i}^n s_j\right)s_i}\prod_{i=1}^n\prod_{j=i+1}^n\left(1+e^{\eta(\theta s_is_j-\zeta)}\right) \\
&= \frac{1}{\mathcal{Z}^{\eta}} e^{\eta\mathcal{H}^{\eta}(\mathbf{s})},
\end{aligned} \tag{C.53}$$

where $\mathcal{H}^{\eta}(\mathbf{s})$ has been defined in (C.52). □

**Lemma C.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*

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

[36] Note that when the network is exogenous (i.e., when $\lambda=0$) 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*

$$p_{ij}(s_i, s_j) = \frac{e^{\eta(\theta s_i s_j - \zeta)}}{1 + e^{\eta(\theta s_i s_j - \zeta)}}. \tag{C.54}$$

*Proof of Lemma C.2.* With the marginal distribution from (C.47) we can write the conditional distribution as

$$\begin{aligned}
\mu^\eta(G|\mathbf{s}) = \frac{\mu^\eta(\mathbf{s}, G)}{\mu^\eta(\mathbf{s})} &= \frac{e^{\eta\left(\sum_{i=1}^n \left(\gamma_i - \kappa + \frac{\rho}{2}\sum_{j\neq i}^n s_j\right)s_i + \frac{\theta}{2}\sum_{i=1}^n\sum_{j=1}^n a_{ij}s_i s_j - m\zeta\right)}}{e^{\eta\sum_{i=1}^n\left(\gamma_i - \kappa + \frac{\rho}{2}\sum_{j\neq i}^n s_j\right)s_i}\prod_{i=1}^n\prod_{j=i+1}^n\left(1 + e^{\eta(\theta s_i s_j - \zeta)}\right)} \\
&= \frac{e^{\eta\sum_{i<j}^n a_{ij}(\theta s_i s_j - \zeta)}}{\prod_{i=1}^n\prod_{j=i+1}^n\left(1 + e^{\eta(\theta s_i s_j - \zeta)}\right)} \\
&= \prod_{i<j}\frac{e^{\eta a_{ij}(\theta s_i s_j - \zeta)}}{1 + e^{\eta(\theta s_i s_j - \zeta)}} \\
&= \prod_{i<j}\left(\frac{e^{\eta(\theta s_i s_j - \zeta)}}{1 + e^{\eta(\theta s_i s_j - \zeta)}}\right)^{a_{ij}}\left(1 - \frac{e^{\eta(\theta s_i s_j - \zeta)}}{1 + e^{\eta(\theta s_i s_j - \zeta)}}\right)^{1-a_{ij}} \\
&= \prod_{i<j} p_{ij}(s_i, s_j)^{a_{ij}}\left(1 - p_{ij}(s_i, s_j)\right)^{1-a_{ij}}.
\end{aligned} \tag{C.55}$$

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 (C.50).

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

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

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

Note that, while the evaluation of the partition function in (C.50) requires the computation of a sum with $2^n$ terms, the partition function in (C.56) requires the evaluation of only $\frac{1}{2}(n_+ + 1)(2(n+1) - n_+) = O(n)$ terms. With (C.56) the marginal distribution $\mu^\eta(\mathbf{s})$ in (C.47) can then be efficiently computed.

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

$$\sum_{j=0}^{\min\{k,n_+\}} \binom{n_+}{j}\binom{n-n_+}{k-j}$$

ways to obtain alignments of $\boldsymbol{\gamma}$ and $\mathbf{s}$ such that $\sum_{i=1}^n s_i\gamma_i = k - (n-k) = 2k - n$.

Next, we consider the products $s_i s_j$. Since all the $j$ agents in $n_+$ with $s_i = +1$ choose the same action $+1$, and all the $k - j$ agents in $n_-$ with $s_i = -1$ choose the same action $-1$ we obtain

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

pairs whose product of actions gives $s_i s_j = +1$. The first term in the equation above counts all pairs of agents with action $+1$ in the first set (with all $\gamma_i = +1$), the second all pairs of agents with action $-1$ in the second set (with all $\gamma_i = -1$), the third term the pairs of agents with action $-1$ in the first set (with all $\gamma_i = +1$), the fourth term the pairs of agents with action $+1$ in the second set (with all $\gamma_i = -1$), the fifth term counts the pairs with agents in the first set who choose action $-1$ and the agents in the second set who chose action $-1$, while the last term counts the pairs with agents in the first set who choose action $+1$ and agents in the second set who also choose action $+1$.

We can further simplify $l(k,j)$ to

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

Then we can write

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

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

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

$$\mathbb{E}^\eta(m) = \sum_{G\in\mathcal{G}^n}\sum_{\mathbf{s}\in\{-1,+1\}^n} m\mu^\eta(\mathbf{s},G) = \frac{1}{\mathcal{Z}^\eta}\sum_{G\in\mathcal{G}^n}\sum_{\mathbf{s}\in\{-1,+1\}^n}\underbrace{me^{\eta\Phi(\mathbf{s},G)}}_{-\frac{1}{\eta}\frac{\partial}{\partial\zeta}e^{\eta\Phi(\mathbf{s},G)}} = -\frac{1}{\eta}\frac{1}{\mathcal{Z}^\eta}\frac{\partial\mathcal{Z}^\eta}{\partial\zeta}. \quad \text{(C.57)}$$

With (C.50) and (C.57) we then can compute the expected number of links as

$$\begin{aligned}\mathbb{E}^\eta(m) = \frac{1}{\mathcal{Z}^\eta}\sum_{k=0}^{n}\sum_{j=0}^{\min\{k,n_+\}}\binom{n_+}{j}\binom{n-n_+}{k-j}e^{\eta(2k-n)}e^{\eta\left(\rho\left(2l(k,j)-\binom{n}{2}\right)\right)-\kappa(n-2(n_++k-2j))}\\ \times\left(1+e^{\eta(\theta-\zeta)}\right)^{l(k,j)}\left(1+e^{-\eta(\theta+\zeta)}\right)^{\binom{n}{2}-l(k,j)}\left(\frac{l(k,j)}{1+e^{-\eta(\theta-\zeta)}}+\frac{\binom{n}{2}-l(k,j)}{1+e^{\eta(\theta+\zeta)}}\right), \quad \text{(C.58)}\end{aligned}$$

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

Note further that for $\theta=\rho=0$ (C.58) simplifies to

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

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

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

**Proposition C.2.** *Assume homogeneous linking costs, $\zeta_1=\zeta_2=\zeta$. Then the expected average*

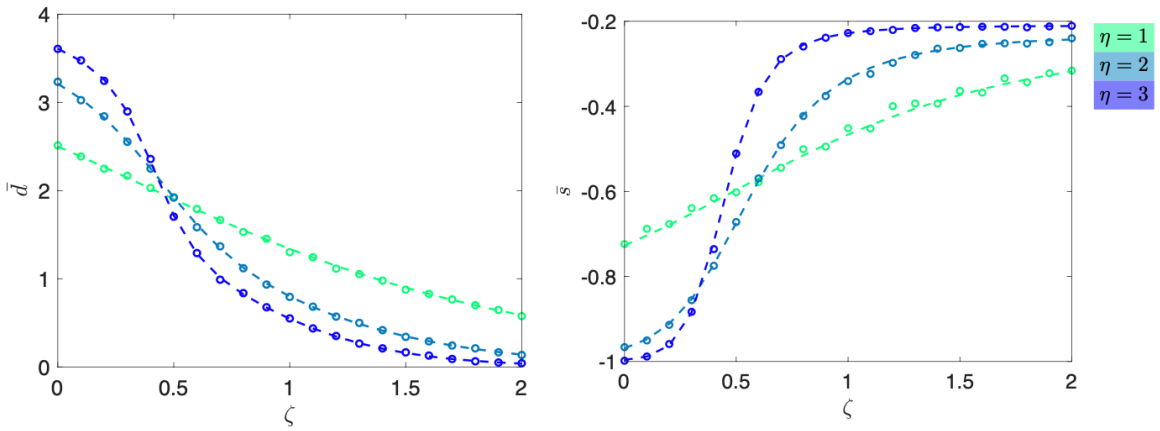



Figure C.1: The average degree $\bar{d} = 2m/n$ (left panel) and the average action level $\bar{s}$ (right panel) across different values of the linking cost $\zeta \in [0, 2]$ and varying noise $\eta \in \{1, 2, 3\}$. The parameters used are $n = 5$, $n_+ = 2$, $\kappa = 0.1$, $\rho = 0.1$, $\lambda = \chi = \xi = 1$ and $\theta = 0.75$. Dashed lines indicate the theoretical prediction of (C.45) in Proposition C.1 and (C.59) in Proposition C.2, respectively, while circles indicate averages across 1000 numerical Monte Carlo simulations of the model using the "next reaction method" for simulating a continuous time Markov chain [cf. Anderson, 2012; Gibson and Bruck, 2000].

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

$$\mathbb{E}^{\eta}(\bar{s}) = \frac{1}{\mathcal{Z}^{\eta}} \sum_{k=0}^{n} \sum_{j=0}^{\min\{k,n_+\}} \binom{n_+}{j}\binom{n-n_+}{k-j} \frac{n+4j-2(n_++k)}{n} e^{\eta(2k-n)}$$
$$\times e^{\eta\left(\rho\left(2l(k,j)-\binom{n}{2}\right)\right)-\kappa(n-2(n_++k-2j))} \left(1+e^{\eta(\theta-\zeta)}\right)^{l(k,j)} \left(1+e^{-\eta(\theta+\zeta)}\right)^{\binom{n}{2}-l(k,j)}, \quad \text{(C.59)}$$

*where* $l(k, j)$ *is defined in (C.46) and* $n_+ = \#(\{\gamma_i = 1 : i = 1, \ldots, n\})$.

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

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

□

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

## C.2. Local Information and Learning

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

$$\mathbf{p} = \varphi \left[ \mathbf{I} - (1-\varphi)\widehat{\mathbf{D}}^{-1}\widehat{\mathbf{A}} \right]^{-1} \mathbf{D}^{-1}\mathbf{A}\mathbf{s}, \tag{C.60}$$

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

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

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

$$\mathbb{E}^{\eta}(m) \simeq \frac{1}{\mathcal{Z}^{\eta}} \sum_{\mathbf{s}\in\{-1,+1\}^n} e^{\eta\langle\tilde{\boldsymbol{\gamma}},\mathbf{s}\rangle} h^{\eta}(\mathbf{s}), \tag{C.61}$$

*where $\tilde{\gamma}_i = \gamma_i + \rho(n-1)p_i - \kappa$, beliefs $\mathbf{p}$ are given by Eq. (C.60), the adjacency matrix $\mathbf{A} = (a_{ij})_{i,j=1}^{n}$ has elements $a_{ij} = e^{\eta(\theta s_i s_j - \zeta)}/(1 + e^{\eta(\theta s_i s_j - \zeta)})$, $\mathbf{D} = \mathrm{diag}(d_1, \ldots, d_n)$ is the diagonal matrix of the degrees $d_i = \sum_{j=1}^{n} a_{ij}$, the partition function is*

$$\mathcal{Z}^{\eta} = \sum_{\mathbf{s}\in\{-1,+1\}^n} e^{\eta\langle\tilde{\boldsymbol{\gamma}},\mathbf{s}\rangle} f^{\eta}(\mathbf{s}), \tag{C.62}$$

*and*

$$\begin{aligned} h^{\eta}(\mathbf{s}) &= \frac{\left(e^{-\eta(\zeta+\theta)}+1\right)^{\alpha(\mathbf{s})}\left(e^{\eta(\theta-\zeta)}+1\right)^{\beta(\mathbf{s})-1}\left((\alpha(\mathbf{s})+\beta(\mathbf{s}))e^{\eta(\theta-\zeta)}+\alpha(\mathbf{s})+\beta(\mathbf{s})e^{2\eta\theta}\right)}{1+e^{\eta(\theta+\zeta)}}, \\ f^{\eta}(\mathbf{s}) &= \left(1+e^{-\eta(\zeta+\theta)}\right)^{\alpha(\mathbf{s})}\left(1+e^{\eta(\theta-\zeta)}\right)^{\beta(\mathbf{s})}, \end{aligned} \tag{C.63}$$

*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 $\mathbb{R}^N$.*

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

**Lemma C.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 C.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 C.3.* Assume that $\zeta_1 = \zeta_2 = \zeta$. Then, in the belief-based model, the quasi partition function is given by

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

where we have denoted by

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

The expected number of links is given by

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

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

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

and we can write

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

where

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

Finally, stationary beliefs are given by (8) so that we can write them as a function of the actions and network as $\mathbf{p} = \varphi\left[\mathbf{I}-(1-\varphi)\widehat{\mathbf{D}}^{-1}\widehat{\mathbf{A}}\right]^{-1}\mathbf{D}^{-1}\mathbf{A}\mathbf{s}$. Moreover, from Lemma C.2 we know that the linking probability conditional on actions $\mathbf{s}$ is given by

$$p_{ij} = \frac{e^{\eta(\theta s_i s_j-\zeta)}}{1+e^{\eta(\theta s_i s_j-\zeta)}}, \tag{C.64}$$

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. □

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

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

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

$$\mathbb{E}^{\eta}(\bar{s}) \simeq \frac{1}{\mathcal{Z}^{\eta}}\frac{1}{n}\sum_{\mathbf{s}\in\{-1,+1\}^n}\langle\mathbf{u},\mathbf{s}\rangle e^{\eta\langle\tilde{\boldsymbol{\gamma}},\mathbf{s}\rangle}f^{\eta}(\mathbf{s}), \tag{C.65}$$

*with $f^{\eta}(\cdot)$ given by (C.63), the partition function is given by (C.62), $\tilde{\gamma}_i = \gamma_i + \rho(n-1)p_i - \kappa$,*

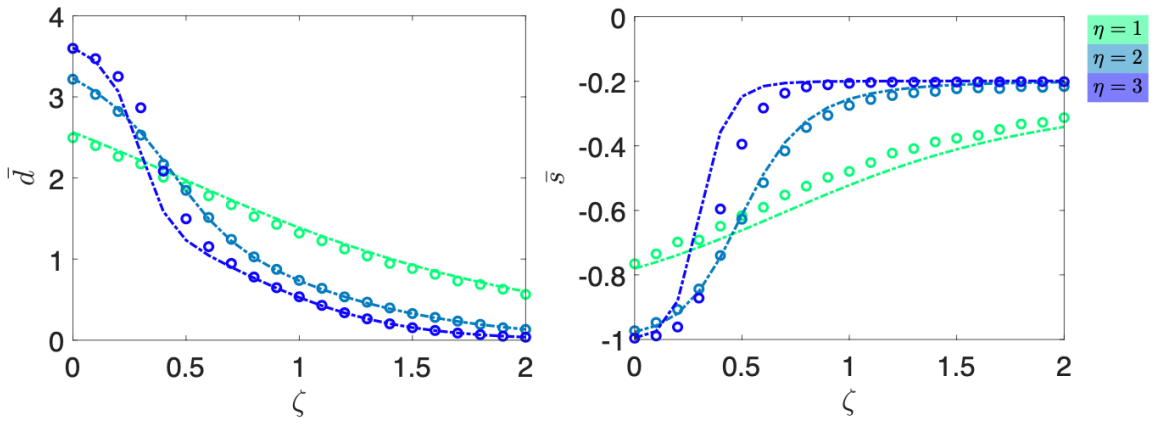



Figure C.2: The average degree $\bar{d} = 2m/n$ (left panel) and the average action level $\bar{s}$ (right panel) across different values of the linking cost $\zeta \in [0, 2]$, $\eta \in \{1, 2, 3\}$, $n = 5$, $n_+ = 2$, $\kappa = 0.1$, $\rho = 0.1$, $\lambda = \chi = \xi = 1$, $\varphi = 0.5$ and $\theta = 0.75$. Dashed-dotted lines indicate the theoretical predictions of $\bar{d} = 2m/n$ in (C.61) in Proposition C.3 and of $\bar{s}$ in (C.65) of Proposition C.4, respectively, while circles indicate averages across 1000 numerical Monte Carlo simulations of the model using the "next reaction method" for simulating a continuous time Markov chain [cf. Anderson, 2012; Gibson and Bruck, 2000].

*beliefs* $\mathbf{p}$ *are given by Eq. (C.60), the adjacency matrix* $\mathbf{A} = (a_{ij})_{i,j=1}^n$ *has elements* $a_{ij} = e^{\eta(\theta s_i s_j - \zeta)}/(1 + e^{\eta(\theta s_i s_j - \zeta)})$, $\mathbf{D} = \mathrm{diag}(\delta_1, \ldots, \delta_n)$ *is the diagonal matrix of the degrees* $\delta_i = \sum_{j=1}^n a_{ij}$, *and* $\langle \cdot, \cdot \rangle$ *is the usual scalar product in* $\mathbb{R}^N$.

*Proof of Proposition C.4.* Assume that $\zeta_1 = \zeta_2 = \zeta$. Then the average action level $\bar{s} = \frac{1}{n}\sum_{i=1}^n s_i = \frac{1}{n}\mathbf{u}^\top \mathbf{s} = \frac{1}{n}\langle \mathbf{u}, \mathbf{s} \rangle$ is given by

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

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 C.4 it follows that

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

where $\alpha(\mathbf{s}) = n_+(\mathbf{s})(n - n_+(\mathbf{s}))$, $\beta(\mathbf{s}) = \frac{1}{2}(n(n-1) - 2n_+(\mathbf{s})(n - n_+(\mathbf{s})))$ and $n_+(\mathbf{s}) = \#(\{s_i = 1 : i = 1, \ldots, n\})$, and hence

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

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 C.3. □

The right panel in Figure C.2 shows the average action level $\bar{s}$ across different values of the

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

# D. Finite Population Equilibrium Characterization

## D.1. Complete Information

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

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

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

*(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* $\gamma_i$ *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* $s_i = +1$*.*

*Proof of Proposition D.1.* This is a direct consequence of the proof of Propositions 3 and 4. □

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

## D.2. Local Information and Learning

The following proposition provides a characterization of the Stochastically Quasi-stable States (QSS) for finite populations in the belief formation environment (generalizing Propositions 5 and 6).

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

$$\begin{aligned} \tilde{\theta}^* &= \zeta_2 + \frac{2(1-\kappa)}{n-n_+}, \\ \tilde{\theta}^{**} &= \zeta_2 + \frac{2(1+\kappa)}{n_+}, \\ \kappa^* &= 2\nu_+ - 1 \ (\leq 1). \end{aligned} \tag{D.66}$$

*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 $\gamma_i$ 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* $s_i = \gamma_i$*.*

*Proof of Proposition D.2.* This is a direct consequence of the proof of Propositions 5 and 6 . □

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

# E. Context and Historical View of the Egyptian Arab Spring

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

## E.1. Historical and Political Background

Egypt had been under what was effectively one-party rule since the 1952 coup that remove King Farouk from power. The ruling political party, originally named Liberation Rally, transitioned to the politically centrist National Democratic Party (NDP). The NDP ideology centered around modernist and anti-Islamist, and members were secular elite, bureaucrats, and regime cronies. Hosni Mubarak rose to the head of the NDP movement in 1981.

In addition to the NDP, the Egyptian military sustains power and influence in the political arena. The military also has a vast presence in civilian industry, making it wealthy and opposes Islamist rule.

Egypt's main opposition to the NDP's rule was in the Islamist movement, whose main social organization is the Muslim Brotherhood (MB). The MB's ideology centers around a literal interpretation of scriptures and advocates a return to idealized Islamic society. MB's followers are urbanized middle and lower classes. Outlawed in 1954 connected to the assassination attempt of president Nasser. From the 1970s, leaders were freed, and MB moved towards an official political party as many leaders were released from prisons (tolerated but not liberated). In the 2005 elections, MB gained approximately 20 percent of the seats in the Egyptian parliament by 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 supporters. Egyptian society was also characterized by various protest movements over time, some with pro- and some with anti-government orientation.

## E.2. The Egyptian Arab Spring

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

**Phase I: Fall of Mubarak.** Under Mubarak's rule, and particularly in the latter stages, NDP members acquired vast wealth while the civilian population stagnated Following the removal of Tunisian President Bin Ali in early 2011, the fervor against privileged elites and ruling parties in North Africa and the Middle East grew. This led to thousands (5K) of protesters congregating in Cairo's Tahir Square in a public demonstration against the Mubarak regime organized by young middle-class Egyptians, not Islamist opposition. MB later encouraged its members to participate without invoking the MB's Islamist slogans or ideology. Note protests are illegal in Egypt. After the initial protest, demonstrations continued, growing to 50K on Jan 28, and by Feb 1 over 500K protesters. On 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 is 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 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 the military, culminating in the 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 passed a new constitution promoting political Islam and expanded military power passes despite the 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 to defuse growing defiance to his rule. This leads to larger protests in the following days involving two sets of protesters, pro- versus anti-Morsigroups, in different locations across Egypt. On July 1, the military issues an ultimatum to Morsi to call an early election. On July 2, Morsi refuses to step down. On July 3, the Egyptian Military overthrows Morsi's regime in coup, and anti-military intervention protests grow. 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 the military for anti-intervention protesters.

# F. Constructing the Data on the Egyptian Protests

## F.1. Reconstructing the Dataset of Borge-Holthoefer et al. [2015]

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

Twitter's terms of use prevent researchers sharing Twitter data either directly or by posting the raw data online. However, authors can provide the 'Tweet-IDs' - a numerical identifier - and numeric 'User-IDs' - which uniquely identify Twitter user profiles - that were used in their research. After obtaining the IDs, we then performed a process known as 'Tweet Hydration' to pull the tweets and user-info with any associated meta-data using software created by the Documenting the Now project (`https://www.docnow.io/`).[37] This process involves a process of repeatedly querying the public facing Twitter API and requesting the associated meta-data for either a Tweet-ID or User-ID.

The API returns the complete data about each tweet or user from the query provided that the tweet or user profile has not been deleted or taken down from the platform. Importantly for our application, the returned data includes the textual content of the tweet, date and time posted and the user-ID. When returning information about each user, the API returns the information at the time the API query was made - thus the username, friends and follower counts as measured in December 2020.[38]

## F.2. Inferring Political Affiliation of Twitter Users

We briefly describe the process undertaken by Weber et al. [2013] to label 20,886 Egyptian Twitter users by their political affiliation, which we use as training data to classify each of Twitter users in our dataset. Starting with a set of manually labelled Twitter users, the authors collect data on twitter users who interact with these accounts. Table F.2 contains the list of seed

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

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

users. For each of the seed users, their most recent 3,200 tweets extracted from the Twitter API and then the user meta-data of up to 200 retweeters of each tweet were downloaded.[39] From this set of retweeters, only those who's location could be identified as being located in Egypt were retained in the sample.[40] The remaning Egyptian residing users were classified as Islamist or Secularist according to their retweeting behaviour: A user who retweeted $n_I$ distinct Islamist seed users and $n_S$ distinct Secular seed users over two time periods (January and March 2013) is classified as a political Islamist if $n_I/(n_I + n_S) > 0.5$.

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

| Secularists | | Islamists | |
|---|---|---|---|
| Screen Name | Twitter Handle | Screen Name | Twitter Handle |
| Mohamed El Baradei | @ElBaradei | Muhammad Morsi | @MuhammadMorsi |
| Alaa Al-Aswany | @alaaaswany | Fadel Soliman | @FadelSoliman |
| Ayman Nour | @AymanNour | Essam Al Erian | @EssamAlErian |
| Wael Abbas | @waelabbas | Almogheer | @almogheer |
| Belal Fadl | @belalfadl | Hazem Salah | @HazemSalahTW |
| Dr. Hazem Abdelazim | @Hazem_Azim | Khaleed Abdallah | @KhaleedAbdallah |
| MohamedAbuHamed | @MohamedAbuHamed | Melhamy | @melhamy |
| HamzawyAmr | @HamzawyAmr | Dr Mohamed Aly | @dr_mohamed_aly |
| E3adet Nazar | @E3adet_Nazar | Mustafa Hosny | @MustafaHosny |
| GameelaIsmail | @GameelaIsmail | El Awa | @El_Awa |
| shabab6april | @shabab6april | | |
| waelabbas | @waelabbas | | |

## G. Additional Empirical Results

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

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

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

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

| | | with random effects (1) | w/o random effects (2) |
|---|---|---|---|
| Local spillover | ($\theta$) | 0.1716*** | 0.2292*** |
| | | (0.0029) | (0.0026) |
| Global conformity | ($\widetilde{\rho}$) | 3.10e-6*** | 3.02e-6*** |
| | | (9.18e-8) | (7.90e-8) |
| **Individual preference** | | | |
| Female | ($\beta_1$) | -0.0637*** | -0.0562*** |
| | | (0.0090) | (0.0062) |
| Islamist | ($\beta_2$) | 0.1007*** | 0.1103*** |
| | | (0.0055) | (0.0037) |
| (Log) followers | ($\beta_3$) | 0.0119*** | 0.0090*** |
| | | (0.0017) | (0.0015) |
| Random effect | ($\tau$) | 0.0059*** | – |
| | | (0.0009) | |
| **Linking cost** | | | |
| Constant | ($\phi_0$) | 14.6860*** | 12.5847*** |
| | | (0.0226) | (0.0088) |
| Same gender | ($\phi_1$) | -0.1629*** | -0.1873*** |
| | | (0.0161) | (0.0073) |
| Same religiousness | ($\phi_2$) | -0.0781*** | -0.0011 |
| | | (0.0082) | (0.0061) |
| Diff. in followers count | ($\phi_3$) | 0.0858*** | 0.0987*** |
| | | (0.0032) | (0.0027) |
| Variance of random effect | ($\sigma_z^2$) | 2.0719*** | – |
| | | (0.0175) | |
| Sample size (# of nodes) | | 225,578 | |

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

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

| | | with random effects (1) | w/o random effects (1) |
|---|---|---|---|
| Local spillover | ($\theta$) | 0.0709*** | 0.1601*** |
| | | (0.0036) | (0.0025) |
| Global conformity | ($\rho$) | 2.36e-6*** | 1.71e-6*** |
| | | (5.15e-8) | (3.88e-8) |
| Weight of local observation | ($\varphi$) | 0.0961*** | 0.0802*** |
| | | (0.0050) | (0.0040) |
| **Individual preference** | | | |
| Female | ($\beta_1$) | -0.0565*** | -0.0523*** |
| | | (0.0112) | (0.0087) |
| Islamist | ($\beta_2$) | 0.1144*** | 0.1181*** |
| | | (0.0063) | (0.0047) |
| (Log) followers | ($\beta_3$) | 0.0059** | 0.0032** |
| | | (0.0020) | (0.0015) |
| Random effect | ($\tau$) | 0.0055*** | – |
| | | (0.0004) | |
| Rioting cost | ($\kappa$) | -0.2951*** | -0.3065*** |
| | | (0.0113) | (0.0085) |
| **Linking cost** | | | |
| Constant | ($\phi_0$) | 14.6812*** | 12.5650*** |
| | | (0.0226) | (0.0094) |
| Same gender | ($\phi_1$) | -0.1655*** | -0.1947*** |
| | | (0.0148) | (0.0087) |
| Same religiousness | ($\phi_2$) | -0.0776*** | -0.0038 |
| | | (0.0071) | (0.0065) |
| Diff. in followers | ($\phi_3$) | 0.0871*** | 0.0990*** |
| | | (0.0033) | (0.0025) |
| Variance of random effect | ($\sigma_z^2$) | 2.1127*** | – |
| | | (0.0205) | |
| Sample size (# of nodes) | | 225,578 | |

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

## References


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

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

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

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

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

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

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

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