Politecnico di Torino

Sony CSL

Politecnico di Torino

Master of Science Program in Physics of Complex Systems

A.a. 2024/2025

Opinion Dynamics and Network Balance in Social Systems under External Perturbations

Supervisors:
Luca Dall'Asta
Vittorio Loreto
Emanuele Brugnoli
Ruggiero Lo Sardo

Candidate: Alice Nappa

Acknowledgements

Completing this master's thesis has been a challenging but incredibly rewarding journey that would not have been possible without the guidance and support of many people. I would like to thank Vittorio Loreto, Emanuele Brugnoli and Ruggiero Lo Sardo for having me in Sony CSL. I am deeply grateful to my supervisors for their advice, patience, and constant encouragement throughout this work. I could not have asked for a better environment, the atmosphere at Sony CSL was free, stimulating, and scientifically high level. I always felt encouraged to explore ideas, challenge assumptions, and grow as a researcher. I am profoundly grateful and hope to continue collaborating in the future. My sincere thanks go to my internal supervisor Luca Dall'Asta for his invaluable guidance and encouragement throughout this work.

Abstract

In modern societies social networks are gaining a growing importance for political debate, which makes opinion dynamics models especially valuable for their descriptive and predictive power. Understanding how opinions form and change online can help keep public discourse healthy and protect the democratic foundations of our increasingly digital societies. This master's thesis explores opinion dynamics in complex social systems, with a particular focus on the role of leaders and their interaction with external events. To this end, we based our approach on Heider social balance theory. This latter is well documented in the literature, but the connection between microlevel balancing processes and broader network dynamics has received limited attention in existing research. To address this gap, we start from an opinion dynamics model that incorporates agreement dynamics and external information effects and we extend the model by introducing two new elements: Heider reputation heuristics, a structural property that leads to social balance, and leader–follower dynamics, essential to reproduce real interaction patterns. To connect theoretical modelling with empirical analysis, we examined whether real-world networks follow Heider reputation heuristics and how their features respond to external perturbations. We analysed interactions between normal users and Italian politicians on Twitter/X over five years, focusing on the Covid-19 pandemic as an external perturbation. The empirical analysis reveals significant pandemic-related changes: network balancedness decreased, cross-community retweet flows increased, and average triadic coherence diminished. We characterized triadic configurations based on leader-follower composition, identifying different stability patterns across different triadic types. Once the observables affected by the external event had been identified in the data, we sought to study these same observables within the model to validate its robustness. These findings establish that structural properties coming from triadic relations can be connected to emergent collective phenomena such as opinion polarization and leader cooperation dynamics. The extended model effectively captures how opinion formation and network stability evolve under external perturbations, linking micro-level structural network configurations with macro level collective responses to crisis events.

Contents

Chapter 1

Introduction

1.1 Generalities

Over the last two decades, the quantitative study of social sciences has attracted increasing attention because the complexity of modern societies requires new approaches. Digital technologies, online communication platforms and the global scale movement of people and goods generate enormous amounts of data: from social media interactions and mobility traces to economic transactions. The scientific community has used the analysis of these digital traces to shed light on patterns of human behaviour that were previously invisible and unquantifiable.

At the same time, many of the most pressing challenges of today are the result of collective phenomena that emerge from the interaction of millions of individuals: as simple examples, one can think of climate change, financial crises, pandemics, and political polarization.

In the field of opinion dynamics, everything we observe can be understood as functioning like a true physical model: interactions at the microscopic level give rise to observable macroscopic phenomena. For example, in physics, the electrical force keeps electrons bound to their atoms, and atoms together form molecules. Similarly, in social systems, the "forces" at play are psychological and relational affinity, hostility, friendship, and influence, which govern how individuals interact at the micro level among each other. These interactions give rise to large scale collective phenomena such as polarization, the formation of communities, and other emergent patterns studied in computational social science. Understanding such phenomena through models can capture how decisions at the micro scale give rise to large-scale regularities, just as statistical physics describes how microscopic interactions produce

macroscopic laws. This analogy between social interaction and physical forces has attracted the interest of physicists and mathematicians, inspiring them to apply quantitative and modelling approaches to understand human behaviour. Governments, policymakers, and private organizations increasingly depend on this kind of evidence to design interventions, anticipate crises, and evaluate the impact of their actions.

In short, the need for predictive and explanatory power in the face of complex, data-rich and interdependent societies is what drives the global interest in the complexity science.

1.2 Historical trajectory

Here, we aim to illustrate a historical trajectory that begins with the qualitative descriptions of social harmony developed within psychology and evolves toward the precise mathematical formulation of structural balance theory, ultimately leading to quantitative models of opinion dynamics. Before the mid-20th century, there was little common ground between the social sciences and the formal disciplines of physics and mathematics. However, following the development of the first social psychological theories, this began to change. These conceptual advances, opened the way for subsequent formalizations that translated such qualitative ideas into mathematical language, establishing the foundations for the quantitative study of social systems.

One of the earliest systematic frameworks is Fritz Heider's social balance theory [17].

Structural balance theory, proposed by Heider [17], was the first attempt to explain the structure and origin of human tensions in terms of friendship and hostility relationships. In particular, it postulates that social systems with simultaneous friendly/hostile interactions tend to evolve to reduce stress[26].

A decade later, Cartwright and Harary[6] gave a mathematical interpretation to Heider's theory, exploiting the language of graph theory. This was a turning point for computational social science and the study of group dynamics on graphs. In particular, representing individuals as nodes of a signed graph and their friendly or hostile relations as positive or negative edges, they proved that a network is structurally balanced when its vertices can be partitioned into two mutually antagonistic groups with positive ties inside each group and negative ties across groups [6].

Beginning in the 1970, the focus gradually shifted from static descriptions

of balanced configurations to models of opinion formation and belief updating. A major contribution towards this shift was given by the DeGroot consensus model [10], where each agent updates its opinion by taking a weighted average of its neighbours' opinions. This iterative scheme became the prototype for modern opinion dynamics, and is still used to describe many social processes, inspiring all the following research on consensus, polarization, and the spread of influence in complex social networks.

The pioneer study in the early formalization of opinion dynamics from a probabilistic and statistical-physics perspective was made by Holley and Liggett (1975) [18], who introduced and analysed the voter model. In their work they established a rigorous mathematical framework for describing how local interactions among agents, modelled as stochastic processes on a lattice, can lead to consensus in the population. The voter model became a cornerstone for later developments in computational social science and statistical physics, inspiring numerous extensions that incorporated features such as biased interactions, social influence, and network structure. Among these, it is worth mentioning the Deffuant model (2000) [9] that is the first of dealing with continuos opinions as opposed to binary ones, introducing the concept of bounded confidence, where individuals only interact with others whose opinions differ by less than a threshold. This model is the first one able to reproduce the homophily as a mathematical feature. These early model offered valuable insights about how local interactions can drive consensus or polarization, but they typically assumed simplified interaction structures, often regular lattices or fully mixed population, while real social systems are characterized of more structural complexity. This exigence, together with the first developments of network theory from Watts and Strogatz (1998)[30] and Barabasi (1999) [2], revolutionized the newborn field of opinion dynamics. The introduction of complex network models made it possible to move beyond the oversimplified representations of social interactions like regular lattices or mixed populations. Network theory provided a framework to describe the intricate web of social relationships with a much higher degree of realism, capturing heterogeneity in connectivity, clustering, and community structure.

All this, combined with the later development of network science and the advent of big data, led to the consolidation of a new interdisciplinary paradigm. The seminal review by Castellano et al. (2009) [7] on the Statistical Physics of Social Dynamics and the manifesto by Lazer et al. (2009) [20] on Computational Social Science clearly articulated this shift. Together, they established the theoretical and methodological foundations of a new field

aimed at understanding social phenomena through the joint use of mathematical modelling, computational tools, and large-scale empirical data. These works marked a turning point, framing the study of collective human behaviour as a true complex system grounded in both physics and data-driven social science.

Chapter 2

System Model: Opinion dynamics with disagreement and modulated information

This chapter is devoted to the description of an opinion dynamics model originally introduced by Sîrbu et al. in [27]. The key innovation of this model lies in the simultaneous inclusion of disagreement and external information: it integrates both attractive and repulsive social interactions in a self-consistent manner, without the need to introduce additional parameters. Furthermore, it incorporates modulated external information, allowing for the simultaneous promotion of multiple options rather than a single dominant one. In this framework, each individual is represented by a probability distribution over several possible choices, reflecting the likelihood of adopting a given option. Subsequent studies [29] have analysed the emergent macroscopic properties of the system, such as cohesion, consensus, and the conditions under which they arise.

We chose this model as the basis for our master's thesis because its characteristics align well with the exigence to describe opinion shaping in presence of an external perturbation. In our case, we want to describe the behaviour of real data from X, and to do so we want to exploit the framework of Social Balance Theory. This model, after some modifications and reinterpretations of parameters, presents all the features that we need:

- The model's multiple-choice setting reflects the diversity of topics or political clusters present in online debates. Each choice can represent a political issue, a stance on a specific topic, or the affiliation of an individual to a particular party/ideological group.
- The agreement/disagreement mechanism mirrors the interaction dynamics typical of social media platforms, especially X (formerly Twitter), where users express agreement through actions such as retweets or likes, and disagreement through critical comments or opposing posts.
- The inclusion of external information is particularly suited to capture the hierarchical nature of online systems, where influential users or opinion leaders act as information sources. These actors control and diffuse content to their followers, shaping collective discussions and opinion trends.
- The binary nature of pairwise interactions allows the system to be represented as a signed network, where positive and negative links correspond to agreement and disagreement, respectively. This makes it possible to study the resulting structure through the lens of structural balance theory.
- Compared to many others opinion dynamics models, our model does not assume bounded confidence. This allows for greater generality and enables the representation of a wider range of social scenarios, for example consider *The Strength of Weak Ties* from Granovetter [16].

2.1 Mathematical formulation

Each agent is represented as a probability vector of K components. The K components represent the different opinions that the agents can express on a given topic and the i_{th} component of the agent probability vector $\mathbf{x} = [p_1, .., p_K]$ is the probability of the agent making the i_{th} opinion choice, with $\sum_{i=1}^{K} p_i = 1$. Let us consider a simple example. Suppose the topic concerns low-carbon energy sources, and the $K = 3$ options are nuclear, solar, and wind energy. An individual represented by $\mathbf{x} = [0.3, 0.3, 0.4]$ can be interpreted as supporting nuclear and solar energy with equal probability (0.3 each), and showing a slightly stronger preference for wind energy (0.4). Geometrically speaking, given a population of N agents, each of them can be represented as a point of the $(K-1)-$simplex. Specifically, a $(K-1)-$simplex

is a $(K-1)-$dimensional polytope that is the convex hull of its K vertices.

2.2 Population generation

To assess the similarity between two agent in the opinion space, we compute the *cosine similarity* o_{ij} between their respective opinion vectors $\mathbf{x}_i$ and $\mathbf{x}_j$:

$$o_{ij} = \frac{\mathbf{x}_i \cdot \mathbf{x}_j}{\|\mathbf{x}_i\| \, \|\mathbf{x}_j\|} = \frac{\sum_{k=1}^{K} p_k^i \, p_k^j}{\sqrt{\sum_{k=1}^{K} (p_k^i)^2} \sqrt{\sum_{k=1}^{K} (p_k^j)^2}}. \tag{2.1}$$

The components of the opinion vectors are non-negative, so that $o^{ij} \in [0,1]$. If $o_{ij} = 1$, agents i and j occupy the same point on the simplex, whereas if $o_{ij} = 0$, they lie on two distinct corners of the $(K-1)-$simplex. This is why we will exploit this quantity to generate the initial population.

The main parameter regulating the generation of the population is the *initial overlap* (or *cohesion*), defined as:

$$\bar{o} = \frac{2 \sum_{i,j} o^{ij}}{N(N-1)}, \tag{2.2}$$

which represents the probability that a randomly selected pair of individuals will follow agreement dynamics. This parameters is associated with the initial distribution of the individuals on the simplex. To build a population of N individuals, we generate one by one its components. First, we generate a $K-$vector $\mathbf{x} = [p_1, ..., p_K]$ according to Dirichlet distribution on the $(K-1)-$simplex so that $\sum_{i=1}^{K} p_i = 1$ and $p_i \geq 0 \, , \forall i$.

Once the vector is created, we calculate its entropy S through:

$$S = -\sum_{i=1}^{K} p_i \, log_2(p_i) \tag{2.3}$$

To obtain a population with a desired total overlap, we set a threshold $S_{threshold}$ on the entropy value and discard a generated vector with probability 0.9 whenever its entropy S exceeds this threshold.

As the generation process is controlled by $S_{threshold}$, whereas the system dynamics depends on the initial overlap, Figure 2.1 reports the relationship

between the two quantities, allowing us to select the appropriate threshold to obtain a population in K dimensions with the desired overlap.

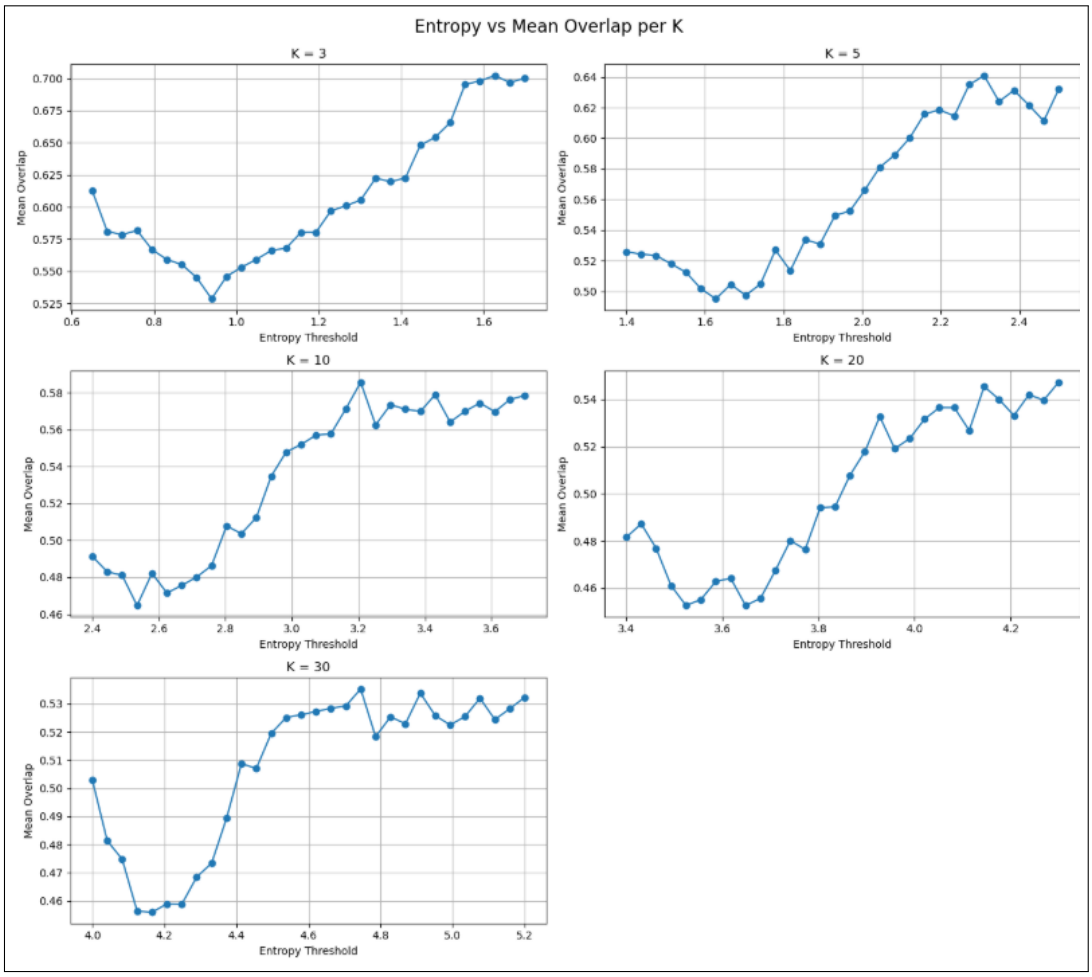


Figure 2.1: Entropy - Overlap trends for $K = 3, 5, 10, 20, 30$.

From Figure 2.1, we extract the ranges in which the two variables exhibit a monotonic increasing relationship, enabling direct control of the overlap parameter through the entropy threshold. These intervals, reported below, are adopted throughout the remainder of this work.

$$\text{entropy\_ranges} = \begin{cases} 3 & \mapsto (0.8,\ 1.6) \\ 5 & \mapsto (1.6,\ 2.4) \\ 10 & \mapsto (2.5,\ 3.7) \\ 20 & \mapsto (3.6,\ 4.3) \\ 30 & \mapsto (4.2,\ 5.5) \end{cases} \tag{2.4}$$

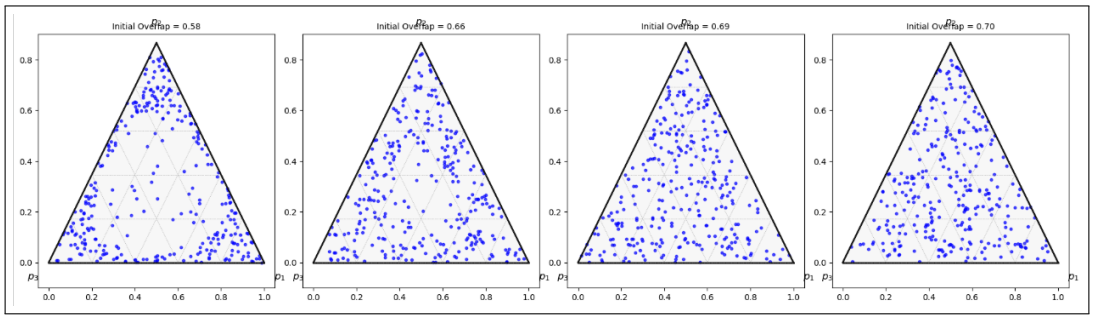


Figure 2.2: Population generation of $N = 300$ agents with $K = 3$. By varying the entropy threshold, we obtain corresponding overlap values of $\bar{o} = 0.58, 0.66, 0.69, 0.70$ for $S_{\text{threshold}} = 1.2$,1.50,1.60,1.80, respectively (Values reported in [29] are verified).

It will become clearer later in this chapter how the initial overlap, as an initial condition, determines the system's evolutionary outcome. For now, let us focus on the physical meaning of this parameter. To better illustrate its role, Figure 2.2 shows the distribution of a population of $N = 300$ agents in $K = 3$ dimensions on the simplex, for increasing values of the total overlap. For lower overlap values, as shown in the leftmost simplex, the centre is sparsely populated compared to the corners. Conversely, higher entropy values correspond to a more uniform coverage of the available two-dimensional surface. It is worth noting that once the simplex becomes fully covered, the system reaches a plateau, and only small oscillations around the plateau value of the entropy can be observed, as shown in Figure 2.1.

2.3 Peer interaction with disagreement

The evolution of the generated population proceeds by **randomly selecting**, at each time step, a pair of individuals i and j to interact (assuming a fully connected network). In this framework, we refer to i as the listener and j as the speaker. The outcome of the interaction is calculated as:

$$\begin{cases} p^{ij}_{\text{agree}} = \min(1, \max(0, o_{ij} \pm \epsilon)) \,, \\ p^{ij}_{\text{disagree}} = 1 - p^{ij}_{\text{agree}}. \end{cases} \tag{2.5}$$

p_{agree} represents the probability that, when i and j interact, the outcome of their interaction will be agreement. Conversely, p_{disagree} denotes the probability that the interaction will result in disagreement. ϵ is a noise term which

avoids lack of interaction due to null vectors similarity and the ± choice is made at random at each time step. After the interaction, the state of the speaker j remains unchanged and the state of the listener i is updated according to the following rule:

$$p_l^i(t+1) = \begin{cases} p_l^i(t) \pm \alpha \cdot \text{sign}(p_l^j - p_l^i), & \text{if } |p_l^j - p_l^i| > \alpha \\ p_l^i(t) \pm \frac{1}{2}(p_l^j - p_l^i), & \text{otherwise} \end{cases} \tag{2.6}$$

Here, l is a randomly selected component of the opinion vector. It is updated by $+\alpha$ (or $-\alpha$) if agent i agree (or disagree) with j, unless the difference between their opinions is smaller than α, in which case the change is set to half of the difference. The parameter α is fixed and determines the agents' flexibility, as it sets the time scale for local agreement or disagreement. The larger α, the faster the two individuals will agree or get separated.

When a component (p_i^l) of the vector is modified, the others must be adjusted so that the total sum remains equal to 1. This is done by uniformly redistributing, across the other $K-1$ components, the amount by which component l was changed. Since 0 and 1 are absorbing states, the adjustment must be performed iteratively.

For example, if position l is increased by α, each of the other components should decrease by $-\alpha/(K-1)$. However, some of them may become negative in the process. In such cases, those components are set to zero, and their deficit is collected into a new amount α' to be redistributed among the remaining non-zero components (excluding l). The procedure is repeated until no negative values appear. This method ensures that the absolute value of the change in position l is the same for both agreement and disagreement updates. The outcome of this interaction scheme can be summarized as follows: when the listener agrees with the speaker, their opinion moves closer to that of the speaker on the simplex, and vice versa. An illustrative example of this process is shown in Figure 4.1 and the pseudo-code for the update is summarized in Algorithm 1.

Algorithm 1 Peer Interaction

1: Select two random individuals i and j
2: Compute their similarity o_{ij} and the agreement probability:

$$p^{ij}_{\text{agree}} = \min\left(1, \max\left(0, o_{ij} \pm \epsilon\right)\right)$$

where the sign $\pm$ is randomly chosen at each time step.
3: Select a random topic l among $l \in \{1, \ldots, K\}$
4: Update the listener's value p^i_l according to:

$$p^i_l(t+1) = \begin{cases} p^i_l(t) \pm \alpha \cdot \text{sign}(p^j_l - p^i_l) & \text{if } |p^j_l - p^i_l| > \alpha \\ p^i_l(t) \pm 0.5(p^j_l - p^i_l) & \text{otherwise} \end{cases}$$

where the sign $\pm$ depends on the case of agreement/disagreement. The change always has the same sign as the speaker-listener difference: it tends to bring p^i_l closer to p^j_l in case of agreement, and to move it away in case of disagreement.
5: Iteratively redistribute the amount added/subtracted across the other vector components, avoiding negative values.

Figure 2.3: Example of the update procedure of the opinion of the listener (A) after an interaction with the speaker (B) who doesn't change state. Agreement and disagreement cases are shown in a $K = 3$ case. Picture taken from [28].

2.3.1 The effect of initial overlap on the evolution

The initial overlap, defined by Equation (2.2), represents the initial degree of similarity among individuals in the population. In this subsection, our aim is to characterize the role of the initial overlap in a system evolving through peer interaction. We first provide a phenomenological visualization of the population's initial and final states on the simplex, and subsequently a quantitative parameter to describe the observed dynamics. As a case study, consider a population of $N = 100$ individuals that evolve through peer interactions over $t = 10{,}000$ steps. Two distinct entropy thresholds are set: $S_{low}0.9$, corresponding to a low initial overlap ($\bar{o}_{low} = 0.56$), and $S_{high} = 1.6$, corresponding to a high initial overlap ($\bar{o}_{high} = 0.71$). The other parameters in the model, namely α and ϵ, do not affect the collective dynamics, as discussed in the Section 2.3.2, but only determine the time-steps required to reach equilibrium. Accordingly, we adopt values of $\alpha = 0.0167, \epsilon = 0.1$, consistent with the settings used in [28].

Figure 2.4 clearly shows two distinct equilibrium configurations. In the low-overlap case, peer interactions are insufficient to promote cooperation among agents, and the population remains segregated near the corners of the simplex (a minimum-entropy configuration). Conversely, in the high-overlap case, the opinion vectors converge toward the centre of the simplex, corresponding to a maximum-entropy configuration.

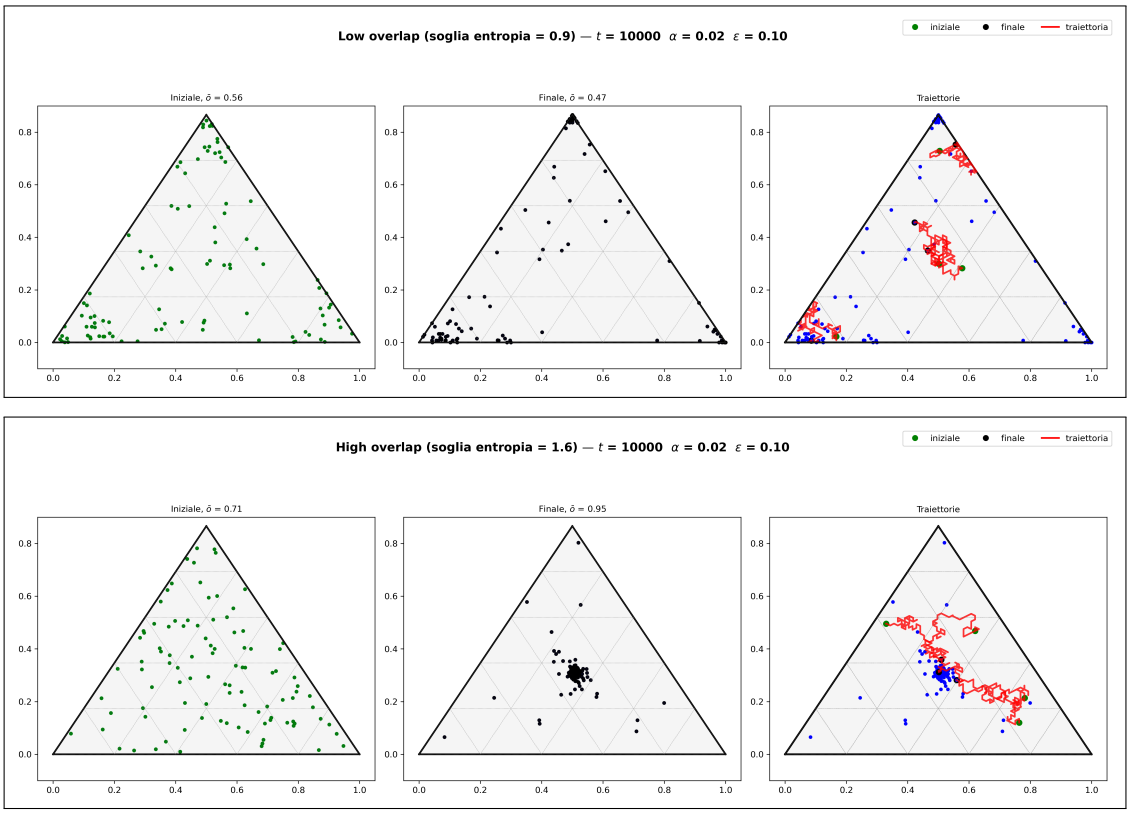


Figure 2.4: Example of the evolution of the population in the simplex with only peer interaction.

Inspired by this approach and following [27], we proceed to provide a quantitative analysis. Once the population has evolved, we perform complete-linkage hierarchical clustering on the final state [21], using a threshold of 0.8 to ensure that agents within the same cluster exhibit an overlap greater than 0.8. We denote by C the number of clusters obtained at the end of the opinions formation process, i.e., when the number of clusters stabilizes. Since the number of clusters alone is not sufficiently informative, as it does not capture how agents are distributed across clusters, we define the *Partecipation Ratio* (PR) as:

$$PR = \frac{(\sum_{i=1}^{C} c_i)^2}{\sum_{i=1}^{C} c_i^2} \tag{2.7}$$

where c_i represents the size of cluster i. In the case of a population organized into two clusters, $PR = 2$ if the clusters are of equal size, whereas $PR \approx 1$ if one cluster is much smaller than the other. More generally, for a population that can form up to K clusters, $PR \approx 1$ indicates the presence of one

dominant cluster, while $PR \approx K$ corresponds to a division of the population into K clusters of roughly equal size. Figure 2.5 shows how the transition point increases with K, meaning that cohesion is facilitated by the presence of more opinion choices. On the other side, the case of agreement in the population ($PR = 1$) also implies a general state of *indecision*. In addition, we find that the balance between local agreement and repulsion across distant groups promotes the emergence of segregated communities located at the corners of the simplex. Importantly, the overall dynamics do not depend on the value of K, so this parameter can be freely chosen without affecting the results.

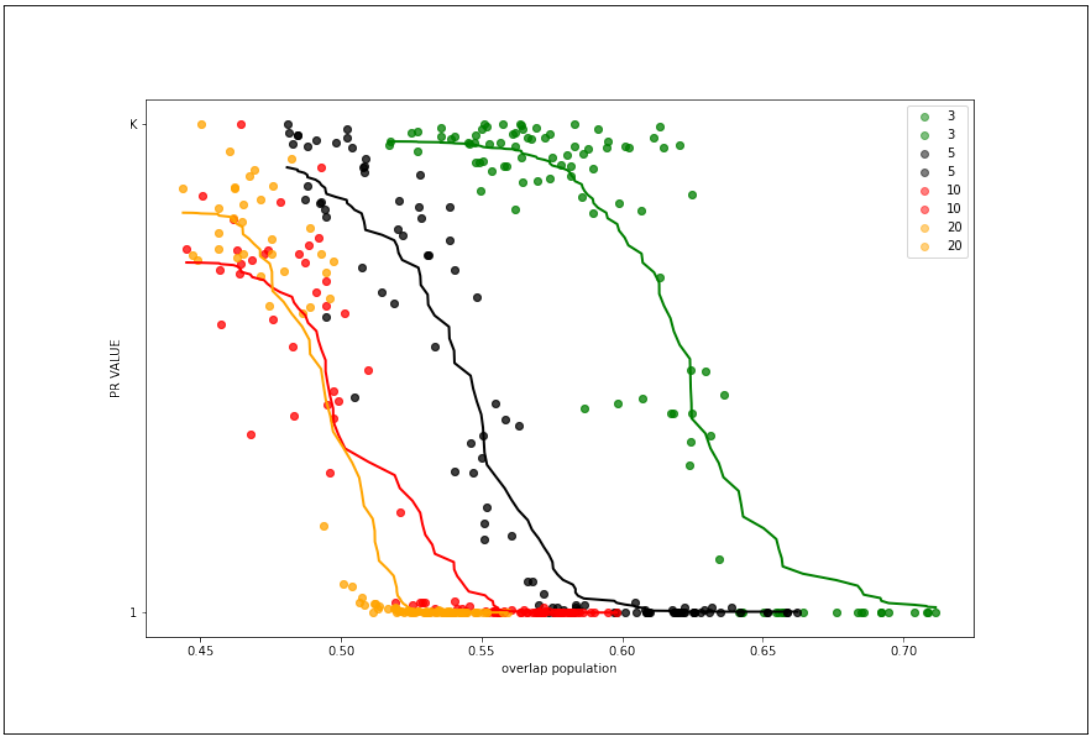


Figure 2.5: PR values for different initial conditions $\bar{o}$. Dots are the PR values across the simulations and lines are averages.

2.3.2 Parameters tuning

Our model involves three key parameters: the noise parameter ϵ, the number of agents N, and the convergence distance α. Throughout all simulations the noise parameter is conventionally set to 0.1, while the other two parameters require a more detailed analysis. The convergence study with respect to N is shown in Figure 2.7 and α in Figure 2.6.

The results demonstrate that the system's behaviour is robust with respect to both parameters. Figure 2.6 shows that the choice of α produces predictable effects:

- If α is too large, agents make very big opinion updates, which drives the system to full consensus for any initial overlap because agreement is reached too easily.
- Reducing α does not qualitatively alter the final state since the phase transition is still present, but the equilibrium time gets longer.
- Increasing N makes the phase transition sharper but also significantly increases the computational time to reach equilibrium. Hence, it is possible to use populations as small as $N = 300$ to keep simulation times reasonable without affecting the final outcomes.

For this reason, in the following we fix α as a K-dependent value reported in Table 2.1.

K	α
3	0.0167
5	0.0100
10	0.0050
20	0.0025

Table 2.1: Values of α for each value of K.

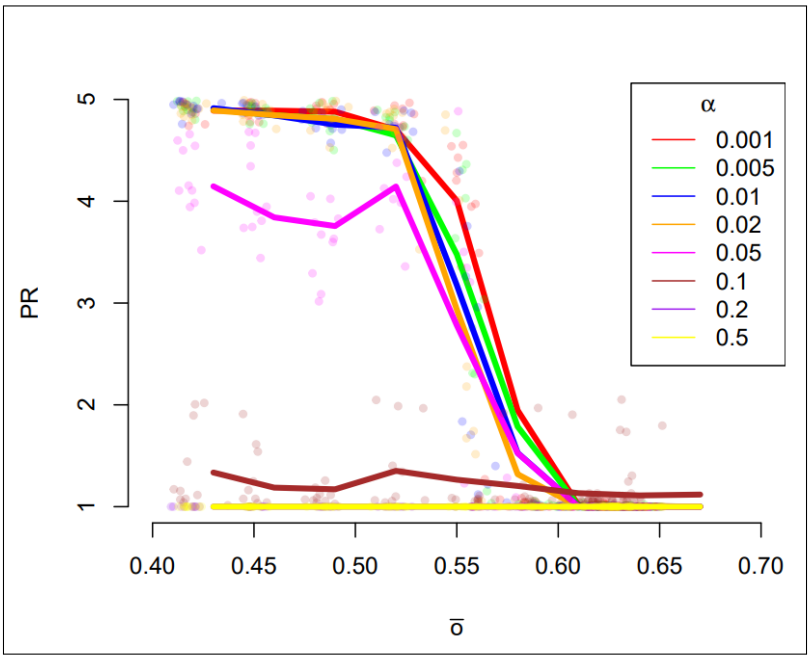


Figure 2.6: PR values for different values of α to perform calibration. Picture taken from [28].

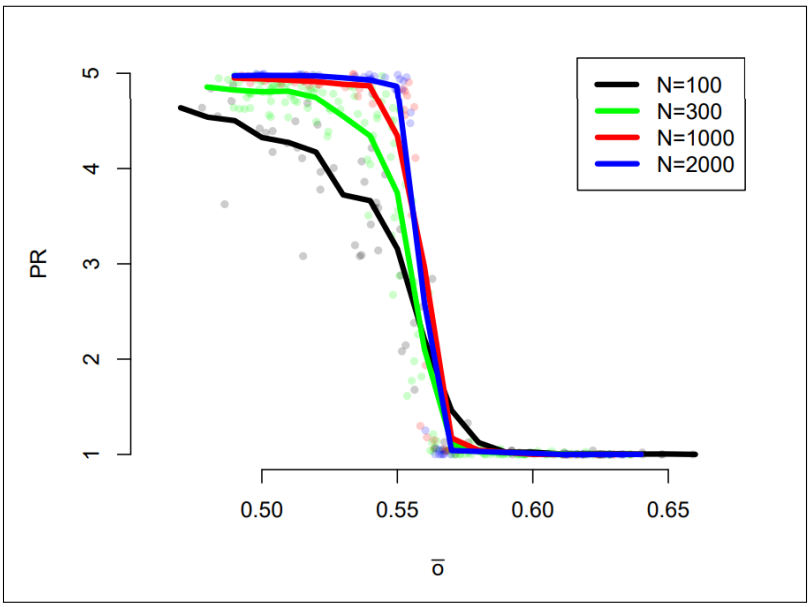


Figure 2.7: PR values for different values of N to perform calibration. Picture taken from [28].

2.4 Modulated sources of external information

Again following the work of [27] and [28] we add to the system another kind of interaction under the name of external information with modulated sources. Taking inspiration from real-life process of shaping opinion, where each individual interacts with peers and with "stable sources" like TV and radio, we introduce K information sources. The external information source is represented as a symmetric $K \times K$ matrix $\mathbf{I}^*$ where the i^{th} row vector $\mathbf{I}^i$ corresponds to the information source promoting the i^{th} opinion *more strongly*. By denoting with $a \in [0,1]$ the parameter describing the source polarization towards its favourite opinion, $\mathbf{I}^*$ can be written as:

$$\mathbf{I}^* = \begin{pmatrix} \mathbf{I}^1 \\ \vdots \\ \mathbf{I}^K \end{pmatrix} = \begin{pmatrix} a & b & \cdots & b \\ b & a & & \vdots \\ \vdots & & \ddots & \\ b & \cdots & b & a \end{pmatrix}, \tag{2.8}$$

where each row corresponds to a different information source, $b = \frac{1-a}{K-1}$, and $a \geq b$.

This means that the information source $\mathbf{I}^i$ encourages adoption of opinion i with probability a, and promotes all the other opinions with lower probability b. When a is close to 1, the sources are highly biased ("extreme"); when a is close to b, they are more balanced ("mild"). This setup assumes that all opinions receive the same overall level of promotion, but can easily be adapted to allow different opinions to be promoted to different degrees.

To evolve the system with the new interaction at each time step, every individual chooses among the sources according to their current opinion vector $\mathbf{x}_i$: the component of option i represents the probability of interacting with the source $\mathbf{I}^i$. Once the source is selected, the interaction will happen with probability P_I with the same update rules already described, where the individual is always the listener, and the information source the speaker. Of course, any adaptation is possible, but for now we will consider fixed sources who cannot interact among each other. The pseudo-code is reported as Algorithm 2.

Algorithm 2 External Information Interaction

1: At each time step, with probability P_I, select a random individual i.
2: Choose the information source $\mathbf{I}^k$ according to the individual's current opinion vector:

$$P(\mathbf{I}^k \text{ is selected}) = \mathbf{x}_{i,k}(t), \qquad k \in \{1, \dots, K\}.$$

3: Compute the similarity o_{ik} between the individual and the chosen source and the agreement probability:

$$p^{ik}_{\text{agree}} = \min\left(1,\ \max\left(0,\ o_{ik} \pm \epsilon\right)\right),$$

where the sign $\pm$ is chosen randomly at each time step.
4: Select a random topic $l \in \{1, \dots, K\}$.
5: Update the listener's value p^i_l according to Algorithm 1.

In other words, our choice corresponds to people tending to consult information that aligns with their own views. For example, in a political analogy, right-leaning voters are more likely to read right-leaning newspapers and engage with content promoted by their favourite leaders.

An individual with very polarized opinions will interact with alternative sources only rarely, whereas a moderate individual, with a mild vector, will engage with a variety of sources over time. This mechanism well represents the dynamics of promotion of content adopted from recommender systems.

Figure 2.8 shows a visualization of the effect of external information interaction on the simplex. We set $K = 3, N = 300, \alpha = 0.0167, \epsilon = 0.1, PI = 0.5,$, and $a = 0.75$ as parameters.

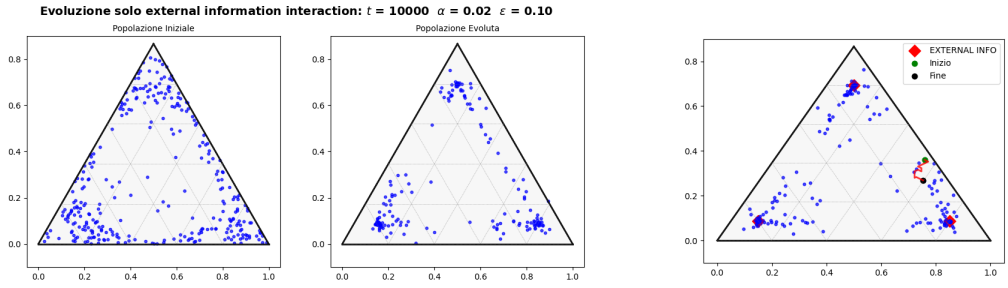


Figure 2.8: Example of population evolution under interaction with external information only.

As expected, the points of the simplex corresponding to the K vectors of the external information are "attractors" for the individuals. In other words, whereas in the absence of external information the system evolution was fully determined by the initial overlap, here the polarization of the information source plays a dominant role.

To fully understand the phenomenon and the global effect of combining these two types of interactions, we study again the Participation Ratio PR as a function of a. We fix $K = 5$, as this behaviour does not depend on the number of opinions. We generate an initial population in two configurations: one corresponding to high initial overlap ($S_{treshold,high} = 2.2$) and the other corresponding to low initial overlap ($S_{treshold,low} = 1.65$). The population then evolves under both peer and external information interactions with $P_I = 0.5$. The resulting PR values for the different a values, $a = [0.2,0.3,0.4,0.5,0.6,0.7,0.8,0.9]$, are reported in Figure 2.9.

In the system with peer interaction only, $PR = 1$ for high initial overlap and $PR = K = 5$ for the low case. When we introduce leader interaction, the leaders act as attractors for the followers, as shown in Figure 2.8. As a consequence, in the low a case, where the leaders are located in the centre of the simplex, $PR = 1$ also for low-overlap, where the peer dynamics would push the individuals apart but mild leaders win over repulsion and bring the system in general agreement (cohesion) . Conversely, in the high-overlap case, for highly polarized sources ($a > 0.8$) the leaders are separated, each one is located in one corner of the simplex. Although the evolution of the peer interaction system would lead to cohesion, the leaders' polarization breaks the cohesion and polarize even high overlap population, so $PR = K$.

This means that the presence of extremely mild or highly polarized sources of information can steer the evolution of the system in their direction. In this scenario, the dynamics are no longer governed solely by the external information source, but are also influenced by the degree of polarization of the leaders. The overall evolution thus results from a complex interplay between these two driving forces.

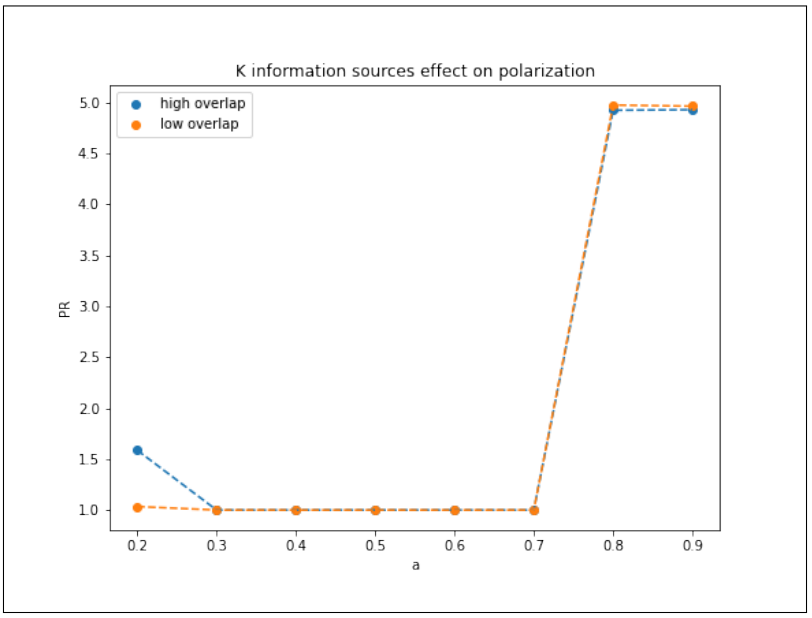


Figure 2.9: Trend of the final population Partecipation Ratio vs. external source mildness. The two lines refer to low and high initial overlap configurations.

Chapter 3

Introducing Social Balance Theory

This chapter introduces *social* (or *structural*) *balance theory* as a framework for understanding how patterns of agreement and disagreement shape the stability of social systems. In essence, social balance theory models social relations as a *signed network*, where positive ties encode affinity or agreement and negative ties encode hostility or disagreement, and studies how local configurations relate to global stability and polarization. We begin with a brief state of the art, highlighting key results on balanced and partially balanced structures and their implications for cohesion, community formation, and conflict dynamics.

Our aim in this thesis is to use the *Sîrbu–Loreto* opinion-dynamics model together with real data as empirical evidence to characterize how a social system responds to an external shock. To this purpose, we chose to adopt social balance theory framework, it lets us cast both the model's interaction outcomes and the data-derived interactions as a signed network, whose properties are richer than those of an unsigned graph and directly tied to notions of tension, equilibrium, and the reconfiguration of alliances. In this view, an external shock perturbs the system's equilibria by altering the pattern of signs (agreements/disagreements) among actors; social balance provides the language and metrics to track how these patterns are and how they reorganize after the perturbation.

3.1 Generalities and state of the art

The roots of social balance theory lie in Heider framework [17] and its network interpretation by Cartwright and Harary [6]. According to Structural Balance Theory (SBT), social balance can be achieved through balanced cycles of relationships. Among all closed loops possibly present in the network, triads (or triangles) are the most widely studied structures in literature. Triads are cycles of length three, between any three members of a social network, and they are characterized as balanced or unbalanced according to these heuristics:

Triad	Configuration	Common saying
	Balanced configurations	
$(+++)$	three mutual friends	*"Friend of a friend is a friend"*
$(+--)$	two mutual enemies, both friends with the same person	*"Friend of an enemy is an enemy"*
$(-+-)$	symmetric to $(+--)$	*"Enemy of a friend is an enemy"*
$(--+)$	two enemies share a common enemy	*"Enemy of an enemy is a friend"*
	Unbalanced configurations	
$(++-)$	two friends disagree about a third person	*"Friend of my friend is my enemy"*
$(---)$	everyone is enemy with everyone	*"All against all"*

Table 3.1: Triads in a signed network: sign patterns and their interpretations according to structural balance theory.

This classification belong to the so called *Structural Strong Balance Theory* (SSBT). Later developments [8] broadened the classic framework of SBT by introducing the idea of K-balanced networks. In this view, a signed graph is considered balanced when its nodes can be divided into K disjoint groups such that the links within each group are positive while the links between different groups are negative. This more general notion of balance forms the basis of *Structural Weak Balance Theory* (SWBT). Under SWBT, even triads consisting entirely of negative edges $(---)$ are regarded as balanced, because each node can be treated as belonging to its own separate group if needed and thus, intuitively, social frustration in this case is lower than in other traditionally unbalanced configurations. A summary of possible configurations is reported in Figure 3.1.

When referring to structural balance, several possible measures of balancedness can be defined. In this work, inspired by [13] and [14], we adopt the Structural Strong Balance measure, quantified as the average fraction of balanced triads at each time-step, or, equivalently, through its frustration $f = \frac{n_+ - n_-}{n_+ + n_-}$ (the two are actually a scaled version of each other so we can use them independently).

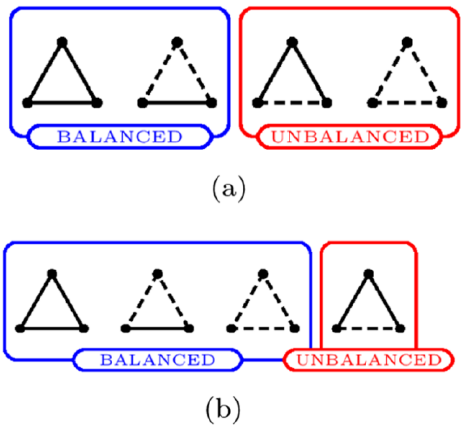


Figure 3.1: (a) Structurally strong balanced configurations are reported. (b) Structurally weak balanced ones. Picture taken from [14].

Beyond these classical definitions, the notion of social balance has inspired a broad range of formal models and empirical investigations. Early work primarily focused on the static detection of balanced structures: given a fully observed signed network, the task was to test whether its configuration of positive and negative links conforms to the predictions of strong or weak balance theory. Over time, however, researchers have moved toward a more dynamic perspective, asking not only whether a network is balanced but also how balance emerges and is maintained.

A first line of dynamic models treats the signs of edges as the outcome of local adjustment processes. For example, "spin" models of structural balance [1] represent each edge as a binary variable that flips when it reduces the number of unbalanced triads, eventually driving the system toward a globally balanced or metastable state. Such models highlight how triadic interactions can create large scale patterns of polarization, even when individual agents follow only simple local rules.

A complementary direction has examined whether higher-order effects are truly necessary to explain the prevalence of balanced triads. For example, Pham et al. [25] demonstrated that the abundance of balanced motifs across diverse social networks can be reproduced by models relying solely on dyadic homophily, that is, pairwise interaction. More recently, Galesic et al. [13] extended this finding by empirical evidence in support of the theoretical claim. Specifically, by randomly assigning groups to receive either triadic or dyadic information, they tested whether individuals must perceive triads for structural balance to emerge—and found that this is not the case.

These contrasting perspectives illustrate an ongoing debate: is the characteristic pattern of balanced triads a genuine emergent property of triadic interactions, or merely a reflection of underlying pairwise preferences? In chapter 4 we give our original contribution for testing structural balance and for disentangling the roles of dyadic non-homophilic interactions and neighbours' opinions in shaping the topology of signed networks. But , in the following sectinos, we will explore the properties of balancedness of the Sîrbu–Loreto model without modifying it just re-framing its formalism.

3.2 Mathematical formulation

Given a System Model time evolution, to build the corresponding signed network at time t, we take a snapshot of the system network history, defining some rules:

- If agent i and agent j interacted and the outcome of the *last* interaction they had is $y_{ij}(t_{last}) \in \{-1,1\}$.
- If the interaction happens again, we simply substitute the old value with the new one (arbitrary, averages and other update rules can be defined).
- The signed network of the system is a time evolving matrix $A(t)$ with entries $A_{ij}(t) = y_{ij}(t_{last})$.

Let $A \in \{-1,0,+1\}^{N\times N}$ be the symmetric signed adjacency matrix (with $A_{ii} = 0$), where if two never interacted their entry is 0. A *triad* is any 3–cycle (i,j,k) for which $A_{ij}A_{jk}A_{ki} \neq 0$. It is *balanced* if $A_{ij}A_{jk}A_{ki} = +1$, *unbalanced* if $A_{ij}A_{jk}A_{ki} = -1$.
Denote by

$$\begin{aligned} n_+ &= \sum_{1\leq i<j<k\leq n} \mathbb{1}\{A_{ij}A_{jk}A_{ki} = 1\}, \\ n_- &= \sum_{1\leq i<j<k\leq n} \mathbb{1}\{A_{ij}A_{jk}A_{ki} = -1\}. \end{aligned} \tag{3.1}$$

Clearly $n_+ + n_-$ is the total number of triads. The *triadic balancedness* of the network is the fraction of balanced triads:

$$\boxed{\mathcal{B} = \frac{n_+}{n_+ + n_-}.}$$

Since $n_+ - n_- = \sum_{1\le i<j<k\le n} A_{ij}A_{jk}A_{ki}$, one can also write

$$\mathcal{B} = \frac{1}{2}\left(1 + \frac{n_+ - n_-}{n_+ + n_-}\right) = \frac{1}{2}(1+f). \tag{3.2}$$

Finally, using the matrix identity $\mathrm{tr}(A^3) = 6\sum_{1\le i<j<k\le n} A_{ij}A_{jk}A_{ki}$, the same quantity can be expressed as

$$\mathcal{B} = \frac{\mathrm{tr}(|A|^3) + \mathrm{tr}(A^3)}{2\ \mathrm{tr}(|A|^3)},$$

where $|A|$ is obtained from A by taking the absolute value of each entry.

3.3 Exploring balancedness properties of the System Model

3.3.1 Peer interaction effect on triadic balancedness

Consider a population of individuals generated as described in Section 2.2, which evolves solely through random peer interactions until reaching equilibrium in triadic balancedness $\mathcal{B}$. This equilibrium value is determined empirically, based on when the observed balancedness curves cease to oscillate and stabilize around a constant value. Since the behaviour of the individuals on the simplex under peer interaction alone is governed by the population's initial overlap (see Section 2.3.1), we compute the average balancedness for $K = 3, 5, 10, 20$ under two conditions: high initial overlap and low initial overlap. This allows us to assess both the influence of K on social balancedness and the effect of the population's initial configuration. Table 3.2 reports the simulation parameters and the resulting final PR values. All simulations were run for $t = 5 \cdot 10^5$ steps, which is sufficient for ever the largest K to reach equilibrium.

The curves for the different values of K are shown in Figure 3.2. From the plots, we can observe that:

- For all values of K, the population converges to a stable value of $\mathcal{B}$, which appears to be independent of K.
- The equilibrium value of $\mathcal{B}$ depends on whether the system starts from a low- or high-overlap configuration.

K	S_{low}	S_{high}	α	PR_{low}	PR_{high}
3	0.9	1.6	0.0167	2.996	1.000
5	1.5	2.3	0.0100	4.630	1.000
10	2.45	3.6	0.0050	7.246	1.000
20	3.6	4.2	0.0025	12.658	1.041

Table 3.2: Initial values chosen for the simulations. The values of PR are useful to understand whether the expected behaviour is confirmed.

- The number of iterations required to reach equilibrium increases with K, as expected.

The first observation is that, when the initial condition allows it, the system reaches a maximum balancedness level of approximately 85%. This result is consistent with the findings of [12], which suggest that social systems tend to self-organize in order to minimize internal tension, thereby exhibiting high, though never complete, levels of global balancedness. Conversely, to interpret the lower balancedness observed in the low-overlap case, it is instructive to examine the distribution of triad signs. To this end, we take a snapshot of the system at two points in time—at the early stage of the simulation ($t = 2000$) and at its end—and compare the triad composition in both the low- and high-overlap configurations. Figure 3.3 reports the triad counts for the case $K = 10$, as the qualitative behaviour does not depend on the number of opinion options. The histogram clearly shows that, in the low–overlap case, As the system evolves toward a fully connected state, the distribution remains relatively stable over time: triads of type $(++-)$ and $(--+)$ dominate. These configurations are typical of polarized systems, characterized by positive intra–community links and negative inter–community ones.

The behaviour reflects the two main outcomes of the peer–interaction dynamics: cohesion and polarization. Moreover, in the low–overlap regime the distribution of triad types preserves its overall shape during the evolution, since the system evolves toward a configuration qualitatively similar to the initial one. Conversely, in the high–overlap regime the initial state covers the entire 2D simplex surface, while in the final state it collapses toward the centre of the simplex.

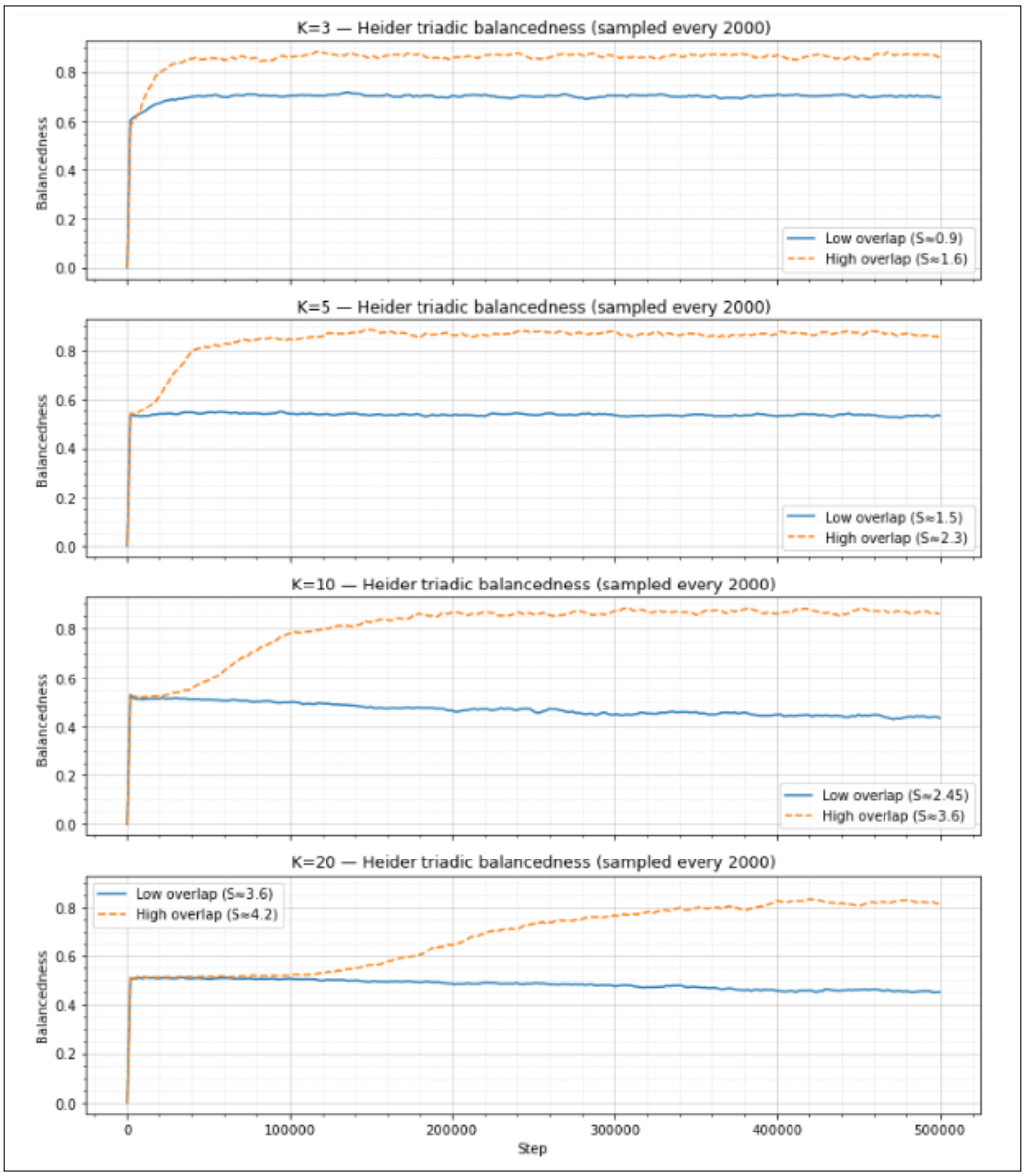
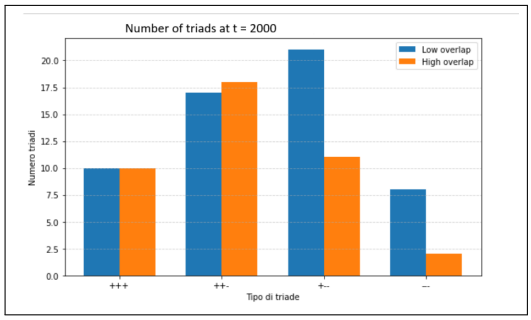


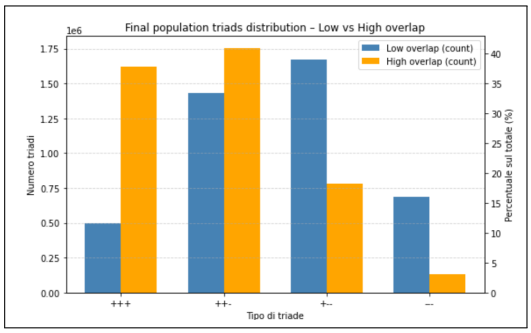
Figure 3.2: Balancedness trends for $K = 3,5,10,20$. Blue curve indicates the low overlap trend, orange curves the high overlap trend.

3.3.2 Leader interaction effect on triadic balancedness

Figure 3.4 shows the evolution of a population of $N = 300$ individuals with both peer and leader interactions, for $K = 3, 5, 10$, and 20. A few observations can be made:



(a) Number of triads per triad type at $t = 2{,}000$ timesteps.



(b) Number of triads per triad type at $t = 200{,}000$ timesteps (final time).

Figure 3.3: Analysis of the triad composition during time for $K = 10$

- In the *high-overlap* case, for $K = 5, 10$, and 20, the presence of leaders—regardless of their level of polarization—tends to homogenize the equilibrium value of the balancedness. This occurs even though, as shown in Figure 2.9, highly polarized information sources can still drive the population toward equilibrium states concentrated near the corners of the simplex, despite the initially high overlap.
- This effect, however, is not observed for $K = 3$, where the system dynamics appear to behave differently.

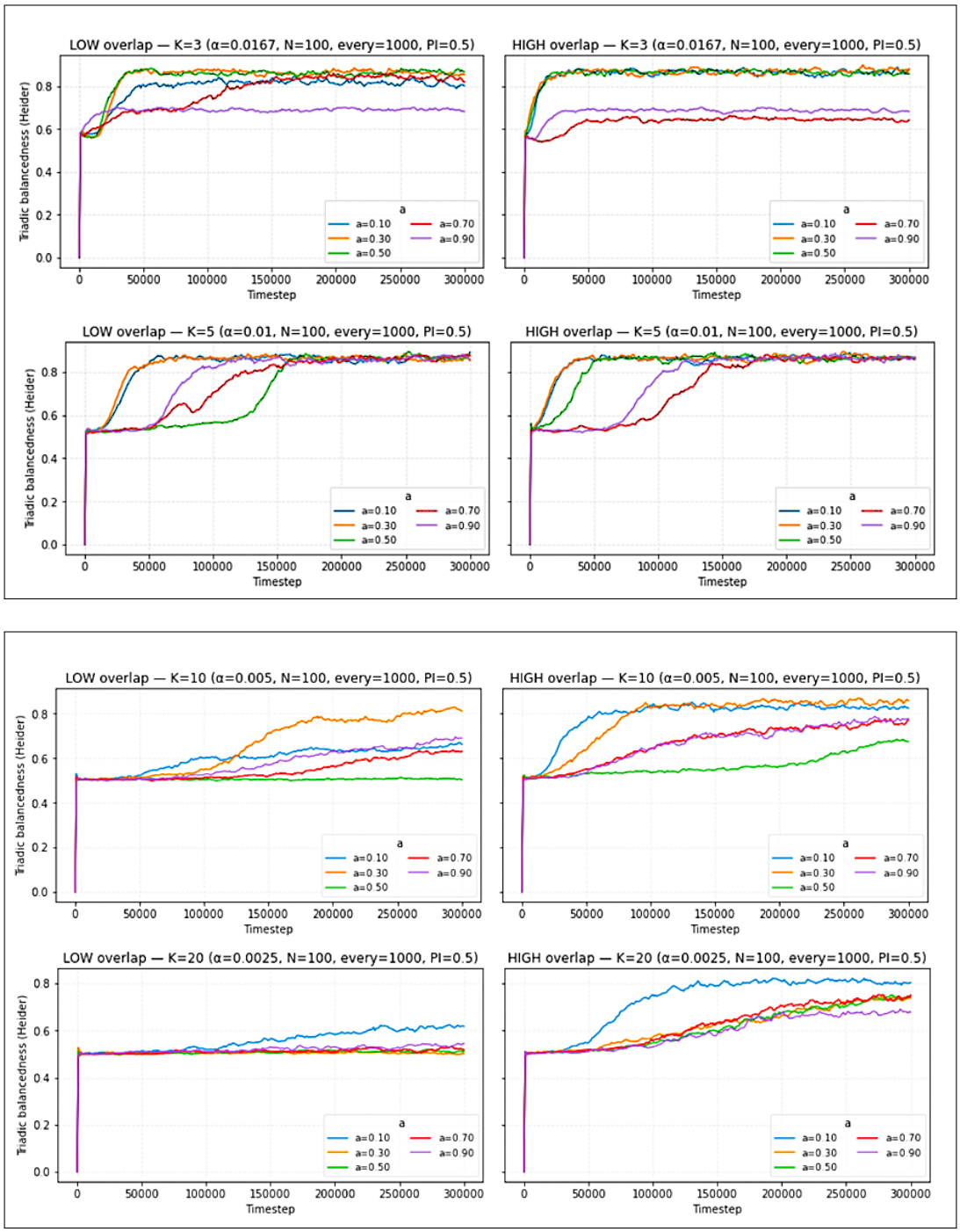


Figure 3.4: Balancedness trends for $K = 3,5,10,20$ in the case of low initial overlap (left) and high initial overlap (right). Colors correspond to different values of leader polarization $a \in [0,1]$.

Chapter 4

A Signed-Network Extension of the System Model Framework via Social Pressure

In this chapter, we extend the Sîrbu–Loreto model by introducing a mechanism that accounts for social pressure as an additional driving force toward triadic balance. The original framework captures the dynamics of opinion formation through pairwise interactions only, without accounting for group dynamics of social pressure when deciding on agreement or disagreement. However, under normal conditions, social systems often display persistent polarization, where opposing groups maintain unbalanced or conflicting relationships over time, according to the Heider heuristics dynamics. Conversely, under exceptional circumstances, such as crises or emergencies, cooperative behaviours can suddenly emerge, driven by a collective urge to reduce social tension and align toward a shared goal.

To reproduce this phenomenon, we modify the model by incorporating a many-body interaction term that represents social pressure. Conceptually, this term acts as an emergent collective influence, reinforcing system's tendency to evolve toward balanced triadic configurations. While in the original model balance emerges solely from local, dyadic interactions, the inclusion of social pressure introduces a global coupling effect: the state of the overall social network now influences the stability and evolution of each triad.

By linking social pressure to the system's response to external stressors,

such as the urgency, emotional intensity, or societal crisis, we capture how macroscopic social conditions can alter microscopic interaction rules. The resulting framework provides a richer description of how collective adaptation unfolds under stress, showing how external pressure can accelerate, inhibit, or qualitatively transform the path toward social balance.

4.1 Settings

The evolution of the population's opinions within the Sîrbu–Loreto framework is governed by Algorithm 1 (peer interaction) and Algorithm 2 (external information), which updates the opinion vector of the listener i at each time step with probability P_I. The outcome of the interaction is controlled by $p_{agree} \propto o_{ij}$ where o_{ij} is the cosine similarity between the opinion vectors of i and j (see Equation (2.2)). It is important to underline again that in our model the Peer Interaction is not homophilic, whereas the interaction with leaders is probabilistically homophilic. Each interaction has a stochastic binary outcome of agreement or disagreement, thus our definition of the signed network starts from here.

Accordingly, for every interaction between individuals i and j at time step t, we introduce a binary random variable

$$X_{ij}^{(t)} = \begin{cases} +1, & \text{if the interaction results in } \textit{agreement}, \\ -1, & \text{if the interaction results in } \textit{disagreement}. \end{cases}$$

This variable encodes the signed outcome of the peer (or leader–follower) encounter at that specific time.

Since interactions take place sequentially, the signed network is itself a time–evolving object. Therefore, we consider the *signed adjacency matrix* at time t,

$$A(t) = \left[w_{ij}(t)\right]_{i,j=1}^{N}.$$

Since the entries of $A(t)$ are the running empirical means of the outcomes of the interactions between individuals i and j up to time t, we can write:

$$w_{ij}(t) = \frac{1}{n_{ij}(t)} \sum_{\tau \in \mathcal{T}_{ij}(t)} X_{ij}^{(\tau)}, \tag{4.1}$$

where $\mathcal{T}_{ij}(t)$ is the set of time steps $\tau \leq t$ at which i and j have interacted and $n_{ij}(t) = |\mathcal{T}_{ij}(t)|$ is the total number of such interactions. Hence, it is clear that $w_{ij} \in [-1,1]$.

Positive weights $w_{ij}(t) > 0$ indicate a predominance of agreements up to time t, negative weights $w_{ij}(t) < 0$ indicate a predominance of disagreements, and $w_{ij}(t) = 0$ corresponds to a perfectly balanced (neutral) history of interactions. Because each edge carries both a sign and a magnitude $|w_{ij}(t)|$, the structure $(V, E(t), w(t))$ forms a *time–evolving weighted signed network*. Note that, as time progresses, the interaction network becomes fully connected.

4.2 Triadic interaction modelling

A natural question arises when extending the original model: how can we introduce a contribution that accounts for the social pressure exerted by the surrounding network? In other words, how can we represent the tendency of neighbouring nodes to promote triadic structural balance? In the standard formulation presented in Chapter 2, whenever two individuals are selected to interact, the sign of the outcome depends solely on the overlap of their opinions. However, real social behaviour suggests that this is only part of the picture.

To better capture these dynamics, we aim to incorporate Social Balance Theory into the model's "decision rule" by introducing a term that reflects triadic social pressure—the collective influence exerted by the immediate neighbourhood of the interacting pair. Consider, for instance, the case of sharing content on social media. Before reposting an opinion, a user typically considers not only their personal agreement with the content, but also the social identity of the original poster. Sharing material associated with an opposing faction can generate tension: friends or family may question the choice, and individuals tend to minimize such social friction. A similar mechanism occurs in everyday life—when introducing a new acquaintance to a close group of friends, one's perception of the newcomer is often influenced by the group's collective attitude. This illustrates a fundamental tension between personal opinion and the social environment: an individual's stance can be reinforced or challenged by the views of their neighbours.

To model this effect, we introduce a parameter of social pressure, denoted by η, which represents the relative weight of the neighbourhood's influence compared to the individual, unbiased reputation based on opinion overlap. Accordingly, we redefine the pairwise overlap o_{ij} (see Equation 2.2) into an effective overlap $\tilde{o}_{ij}$, modified to account for the pressure toward triadic balancedness within the local neighbourhood. To gain intuition about the behaviour of this effective overlap, let us examine the two illustrative cases (a)

and (b) presented in Figure 4.1.

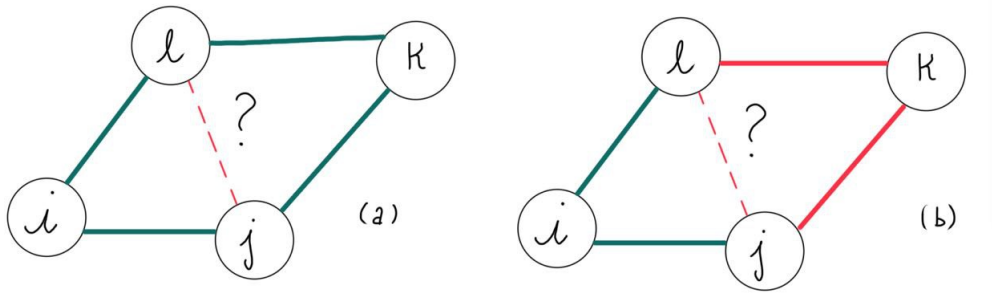


Figure 4.1: (a) Example of social pressure that pushes a disagreement link to be more "agreeing". (b) Example where the two social circles are contradictory. Red links are disagreement/negative opinion.

In case (a), following the "friend of a friend" heuristic, the influence of the neighbours tends to make the negative link (jl) more positive. Similarly, in case (b), although the two links in the triad (ljk) are negative, the "enemy of my enemy is a friend" heuristic promotes greater agreement on link (jl). However, a natural question arises: what happens when the various triads involving (jl) are not concordant, that is, when the local configurations suggest conflicting influences on the same link? This situation calls for the introduction of a weighted average. Accordingly, we define the signed triadic contribution on the link (ij) as:

$$\tau_{ij}^{\text{signed}} = \frac{\sum_{k \in N(i) \cap N(j)} w_{ik}\, w_{kj}\, w_{ikj}}{\sum_{k \in N(i) \cap N(j)} w_{ikj}} \tag{4.2}$$

where $N(i) \cap N(j)$ denotes the set of common neighbours of nodes i and j and $w_{ikj} = |w_{ik}||w_{kj}|$. This definition corresponds to a weighted average over all triads involving the two nodes, with weights defined in Equation (4.1). This triadic contribution is then incorporated into the original pairwise overlap to obtain an effective overlap:

$$\tilde{o_{ij}} = (1 - \eta) o_{ij} \; + \eta \, \tau_{ij} \tag{4.3}$$

where η is the *social pressure* parameter, controlling the relative importance of neighbors' influence compared to the original overlap o_{ij}.

Let us calculate the overlaps o_{ij} and $\tilde{o}_{ij}$ in the two cases (a) and (b) shown in Figure 4.2, with η fixed at 0.3.

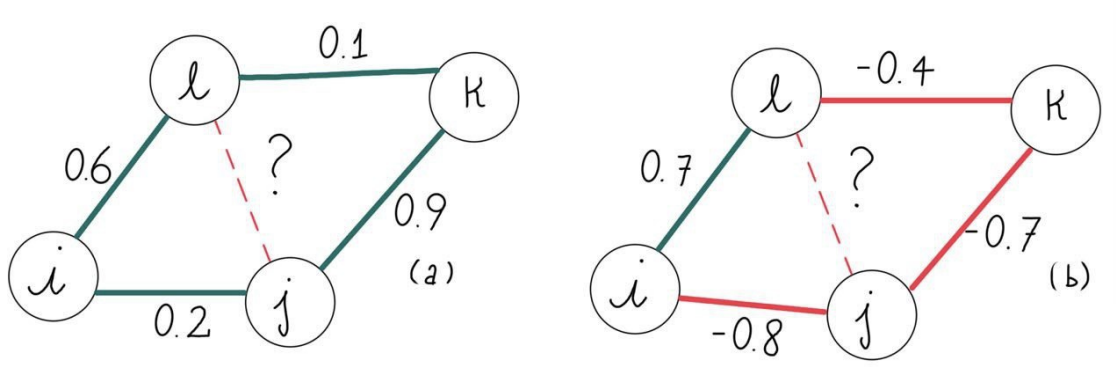


Figure 4.2: Numerical examples of two cases of social pressure influence on the final $\tilde{o}$. In case (a) the neighbours' opinion are concordant and push the overlap to be bigger. In case (b) they are in disagreement, and the final outcomes depends on the links' weights.

In case (a), we obtain $\tau_{ij,signed} = 0.107$ and $\tau_{ij} = 0.5535$. Computing the overlaps, we have $o_{ij} = 0.4$ and $\tilde{o}_{ij} = 0.45$ which is consistent with the expected outcome, since both triads in the example act to push the link toward a more positive value. In case (b), we obtain $\tau_{ij,signed} = -0.28$ and $\tau_{ij} = 0.36$. Computing the overlaps, we have $o_{ij} = 0.4$ and $\tilde{o}_{ij} = 0.388$. Here, the right-hand triad tends to push the link toward a positive value, whereas the left-hand triad favours a negative one. As a result, we expected the triad with the stronger influence to determine the final effective overlap.

4.3 Influence of triadic interactions on social balance

4.3.1 Effect of social pressure on peer interactions

Figure 3.2 shows that the value of K does not qualitatively alter the system dynamics; it primarily affects the convergence time and the minimum level of balancedness reached. This property allows us to fix K for subsequent analyses. Accordingly, from this point onward, we consider $K = 5$, for reasons that will become clear in the following chapter. To investigate the effect of η on the system's social balance, we first consider a standard scenario: a population of $N = 300$ individuals evolving solely through peer interaction, without any leaders. The system is allowed to evolve for $t = 2 \cdot 10^5$ time steps,

while the parameter η is swept over the interval [0,1] in increments of 0.1. The simulation has been run in the two usual configurations of low and high overlap. Results are illustrated in Figure 4.3.

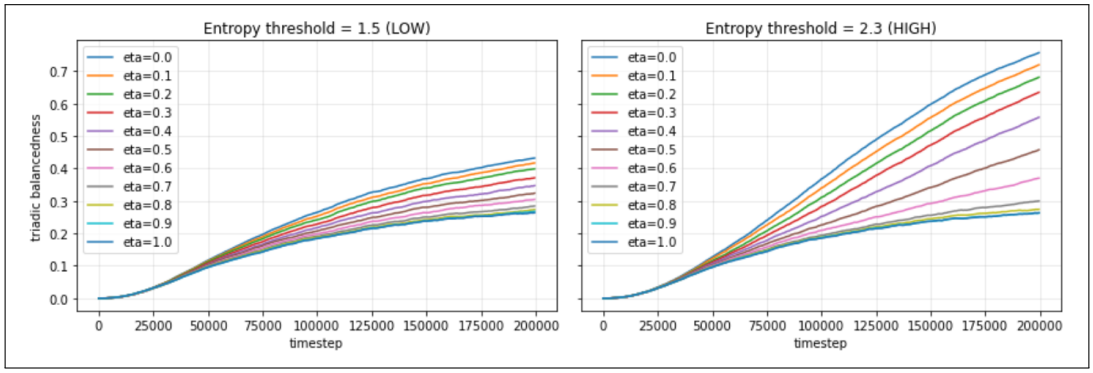


Figure 4.3: Results of sweeping of η parameter on a $K = 5$ population with peer interactions only.

As we can see, η has a direct impact on social balancedness, even if the outcome of its influence is not obvious: indeed, in both configurations higher values of η cause the balancedness of the system to diminish instead of getting higher, which is highly counter-intuitive since the parameter η has been explicitly created to respect social balancedness rules on triads. This is a first taste of the important result that we are going to prove.

4.3.2 Effect of social pressure on peer and leader interactions

We perform the same analysis as in the previous subsection, but now with η fixed at 0.3, a reasonable value for social systems, while sweeping the polarization parameter a over the interval [0,1]. The result is illustrated in Figure 4.4.

As we can see in this case as well the level of balancedness is determined by a, although not in a proportional way: indeed, to the highest a value we don't get the highest or lowest balancedness value accordingly. We get a pretty intuitive result from the simulation. In the low overlap case, we have that in the evolution the population will collapse in the 3 corners of the simplex, this is why a highly polarized source of information will help the social balance of the system as we can see from left plot in Figure 4.4. On the other side, in the high overlap case, the population will evolve towards a $PR = 1$ situation where all the individuals are in the center: that is why mild

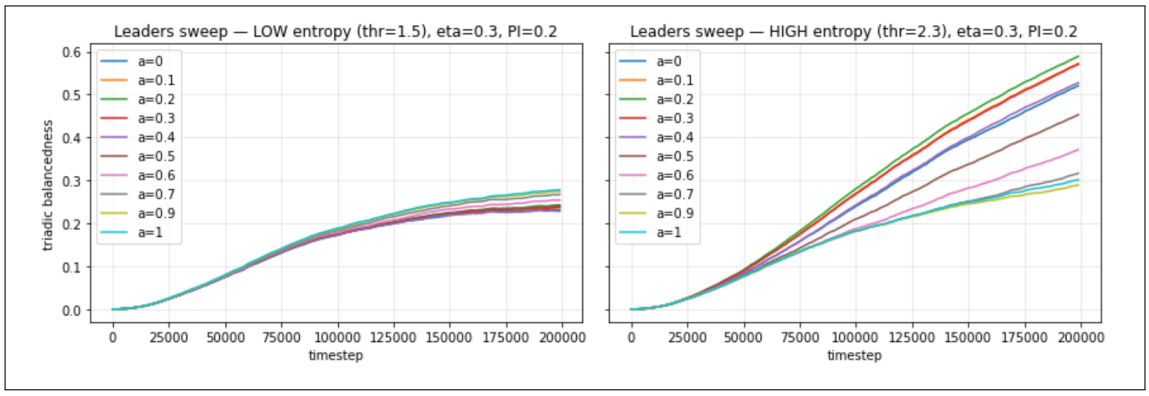


Figure 4.4: Results of sweeping a parameter on a $K = 5$ population with leaders and peer interactions. $\eta = 0.3$ fixed.

polarization of a are the ones that favour the most social balance. The take home message of this is that having sources of information that represent the opinion of the population is crucial to obtain social balance in a system.

4.3.3 How social pressure, polarization, and noise shape social balance dynamics

To complete the characterization of our model, we investigate the interplay of social pressure (η) and noise (ϵ). These parameters are of particular interest as they capture the main effects of external perturbations on the system. As before, we consider peer interaction only, sweeping η and ϵ in the intervals [0,1] and $[0, 0.5]$, respectively.

As we can see, the low and high overlap configurations behave in the same way, differing only by the value of the balancedness. Raising too much the value of ϵ break the spontaneous high level of social balance caused by the low value of η. This is due to the fact that adding a high noise in the overlap completely breaks the dependency of p_{agree} from o_{ij} that the dyadic interaction that guarantees the balancedness in the system.

Let's do the same thing but with leaders interactions as well, with fixed a. Just to prove that again the behaviour doesn't change but the value of balancedness is lower. We proved before that if the external information sources are polarized they have a high level of balancedness with a polarized population, that is why if we fix $a = 0.8$, as we can see below, the level of balancedness is low in both cases.

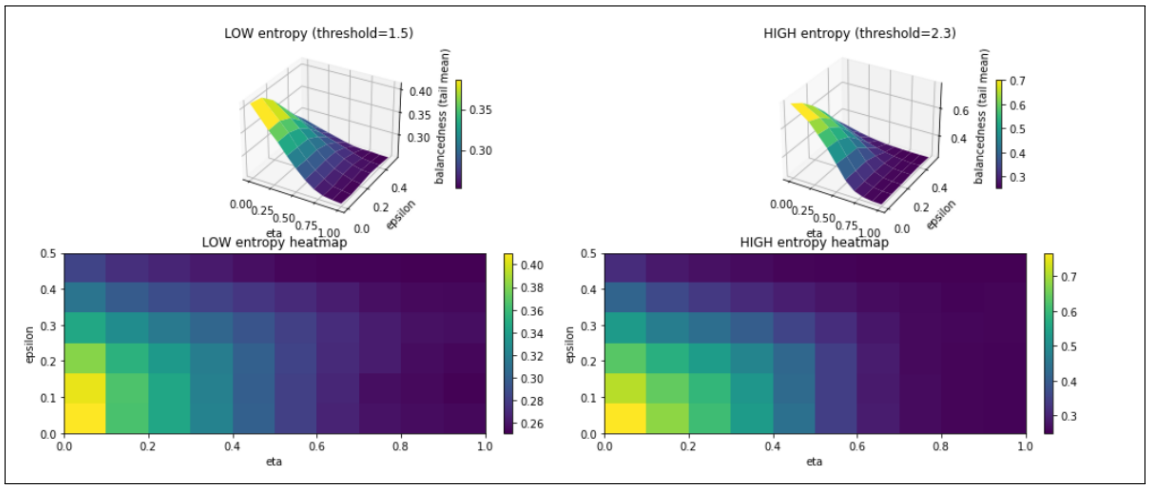


Figure 4.5: Double sweeping of parameters of social pressure and noise for a $K = 5$ population of 300 individuals, evolution is subject only to peer interaction.

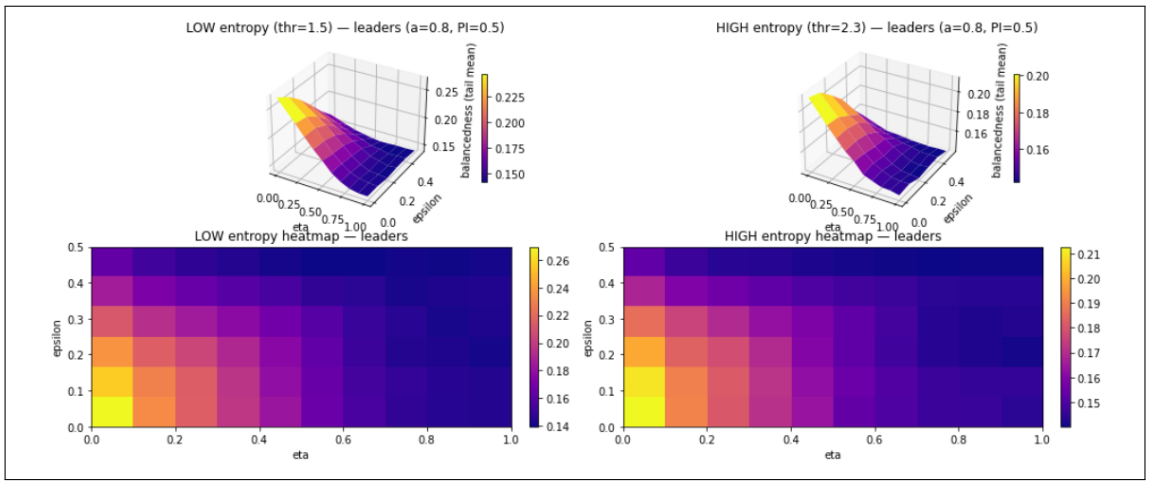


Figure 4.6: Double sweeping of parameters of social pressure and noise for a $K = 5$ population of 300 individuals, with peer leaders interaction.

Note that the values are way lower than if we assigned completely randomly the signs to a fully connected network, where in that case it would be of $\sim 50\%$.

To conclude the characterization of balancedness of the evolution according to the parameters, let's study the double parameters sweep of η and a, keeping $\epsilon = 0.1$ fixed, the others parameters stay unchanged.

As we can see, in both overlap cases there is a change in the balancedness

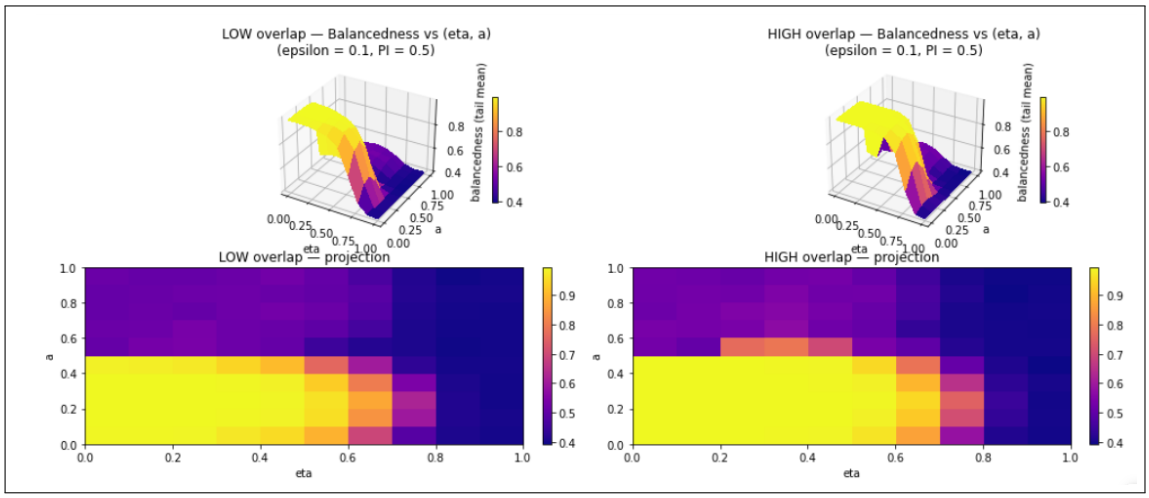


Figure 4.7: Double sweeping of parameter of social pressure and leaders polarization

when $a = 0.5$. Furthermore, we can confirm that when η, the social pressure parameter is high, we obtain low social balance.

4.4 Emergent balance and the limits of social pressure

The results obtained in the previous section are far from trivial. We introduced a parameter representing social pressure, which was expected to drive the system toward a higher level of balancedness by fostering coordination among agents. Surprisingly, the opposite behaviour emerges: increasing social pressure systematically reduces the level of triadic balance across all tested scenarios.

This finding suggests that triadic balancedness does not originate because of social pressure, but rather emerges spontaneously from the underlying dyadic interactions themselves [24]. In other words, the tendency toward structural balance appears to be an intrinsic property of local pairwise dynamics [13], rather than a consequence of externally imposed collective constraints. Conversely, when social pressure becomes strong and dominates over individual opinions, the system exhibits lower overall balancedness, indicating that excessive conformity can hinder the natural self-organizing process through which balance emerges.

This is a striking and counter-intuitive result: it challenges the intuitive

expectation that social pressure necessary stabilizes or reinforces social harmony. Instead, it reveals the subtle and potentially destabilizing role that collective influence can play in shaping social dynamic, suggesting that when individuals align their behaviour too closely with others' opinions rather than their own, the group may become less, rather than more, socially coherent.

Chapter 5

Opinion Dynamics on Real-World Data: The X Case Study

As often happens in theoretical physics, once a theory is firmly grounded, the next step regards testing theoretical models against real-world data. This practice is well established in traditional branches of physics but it is often more challenging within the study of complex systems. In particular, modelling social dynamics presents significant difficulties, as these phenomena are hard to formalize mathematically and even harder to measure accurately. As a result, many studies in social physics remain at a theoretical level, proposing a variety of models for the same processes but with limited empirical validation or connection to real data. In this chapter, we aim to bridge this gap by confronting the model developed in the previous sections with empirical observations from a real opinion system derived from Twitter data. To do so, we base our work on two tools which are recently becoming fundamental in social science research to measure people's opinions dynamics: social network's data and natural language processing. First, we will give a detailed description of the dataset in use, apply some statistics and describe intresting observables to see how the dataset behaves during COVID-19 emergency period. Then, the data will be manipulated to build the correct framework to be used in our simulations, and lately a comparison between model and real data is made.

5.1 Data collection

X (formerly Twitter) is a social networking and microblogging platform that allows users to share and engage with short messages—known as tweets—which may include text, images, links, or emojis. For the purposes of this study, we focus exclusively on the textual content of tweets. Our analysis focuses on the social debate that took place on X between 2018 and 2022, driven by Italian information leaders from both the political and media spheres. To capture a comprehensive picture of this debate, the dataset includes tweets produced by a curated set of Italian news outlets and political actors, compiled from external authoritative sources. With the exception of political accounts, the information leaders included in the dataset are accompanied by a *Reliability Rating* assigned by Newsguard [23], and are therefore classified as trustworthy (T), non-trustworthy (N), or satirical (S). This categorization is particularly valuable for our purposes, as it closely aligns with the leaders' interaction type present in our theoretical framework under the label of external information interaction (see Algorithm 2). By first examining the structure and dynamics of the leaders-only network, we gain crucial insights into the broader network organization. Given the inherent complexity of the full model, this stepwise approach—analyzing its main components separately—provides a clearer understanding of how external information sources influence the overall opinion dynamics.

For each tweet published by the information leaders selected, all corresponding retweets and quote tweets were collected. In addition, for any retweets, quote tweets, or replies produced by these sources, all referenced tweets were also retrieved. The dataset used in this study corresponds to (or partially overlaps with) the one previously analysed in [3, 4, 5]. For comprehensive details on its collection and analysis, we refer the reader to the cited works. For the purpose of this analysis, we organized the dataset as represented in Table 5.1.

The dataset is composed of $\sim 50\,million$ interactions made by $\sim 2\,million$ users, and the distribution of the `reference_type` (one of *quoted, retweeted, commented, replied_to*) is as follows:

The variable `agreement` represents a score in [0,1] derived from stance detection, as described in [4]. Retweets as treated as instances of full endorsement, and therefore assigned an agreement score of 1. Assuming that all retweets represent endorsement imposes a positive bias on interaction data. However, this assumption aligns with socio-psychological research literature in which it is recognized that social systems tend to over represent

Table 5.1: Description of the dataset columns.

Column	Description
`reference_author`	Identifier or username of the author who made the reference.
`reference_type`	Category of the reference.
`referenced_id`	Identifier of the content being referenced.
`referenced_author`	Author or source of the referenced item.
`reference_time`	Timestamp indicating when the original post was made.
`agreement`	Boolean or categorical indicator showing whether the reference expresses agreement, disagreement, or neutrality.

Note. Every time an interaction (retweet, comment, reply, or quote) occurs, it is assigned a unique identifier that is recorded. In this case, `reference_id` denotes the identifier of the interaction, while `referenced_id` refers to the post on which the interaction takes place.

Reference Type	Number of Tweets
quoted	9,402,418
retweeted	30,410,513
replied_to	50,311
commented	23,047

Table 5.2: Number of tweets by `reference _type`.

agreement and similarity among connected individuals [15]. In highly polarized or homophilic networks, where echo chambers prevail, such bias is not only expected but often reflective of real underlying dynamics.

In terms of network description, our dataset describes a tripartite temporal network consisting on the three layers: `reference_author`, `referenced_id`, and `referenced_author`, as represented in Figure 5.1, where green links indicate agreement references, while the red ones stand for disagreement.

We average on the `referenced_id` layer, month by month, to obtain a weighted bipartite network `reference_author` $\leftrightarrow$ `referenced_author`. In this way, for each month, we construct a directed weighted edge from one layer to another. The weight of this edge corresponds to the average interaction score computed over all the contents produced by the `referenced_author`

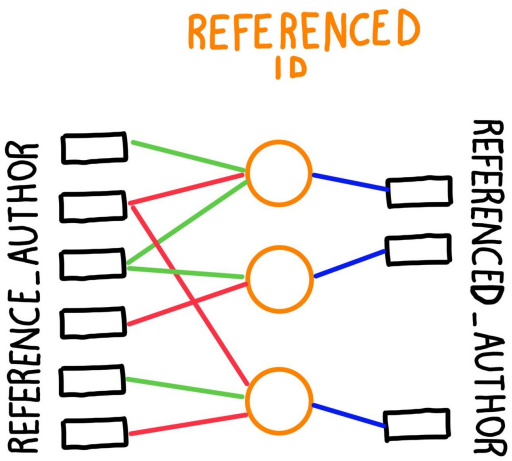


Figure 5.1: Network representation of the dataset structure. The green links in the first layer indicate agreement references, while the red ones stand for disagreement.

during that month. For example, if user A interacts with four different posts from user B in April 2019, the bipartite network will contain an undirected link $A \leftrightarrow B$, whose weight is given by the average of all the corresponding red and green links shown in Figure 5.1.

5.2 Behaviour of leaders network

5.2.1 Grouping political accounts

Starting from the leaders classification and filtering the complete dataset to retain only the ids corresponding to political figures, we can obtain a `leader` $\rightarrow$ `leader` interaction network, where both layers consist exclusively of leaders—specifically, 39 profiles. Morevover, we retain only the *retweeted* `reference_type`, in order to examine more closely how endorsement dynamics between political communities were influenced by an external perturbation—namely, the COVID-19 pandemic. In words, we focus exclusively on agreement interactions occurring from leader to leader. By averaging all the retweet interactions between each pair of leaders, we obtain a weighted leader–leader network. Applying the Louvain community detection algorithm [22] to this network yields a partition into five distinct communities (with modularity $Q = 0.5$), as reported in Table 5.3.

This is particularly valuable for network analysis, as it allows us to further

Table 5.3: Detected political communities and their members.

Community A (4 members)
gparagone, mov5stelle, giuseppeconteit, luigidimaio

Community B (6 members)
unione_popolare, demagistris, direzioneprc, manifesta_it, movimentodema, potere_alpopolo

Community C (7 members)
nfratoianni, si_sinistra, articolounomdp, europaverde_it, angelobonelli1, robersperanza, ellyesse

Community D (9 members)
fratelliditalia, giorgiameloni, legasalvini, coraggio_italia, luigibrugnaro, forza_italia, giovannitoti, matteosalvinimi, berlusconi

Community E (9 members)
matteorenzi, carlocalenda, emmabonino, piu_europa, enricoletta, pdnetwork, azione_it, sbonaccini, italiaviva

enrich our dataset by introducing a `cluster` column associated with the `referenced_author` values. In this way, we complement the endorsement information captured by retweets with additional insight into the community affiliation of the endorsed user, offering a deeper understanding of political interactions on the platform.

5.2.2 Temporal evolution of politically aligned leader communities

How did the leaders distribute their endorsement onto the five political groups division? To capture the temporal evolution of political orientation, we built monthly opinion vectors representing how each leader's retweet activity was distributed across the identified clusters. This process involved a systematic sequence of data filtering, aggregation, and normalization steps, allowing us to track how leaders' alignment with different political communities changed over time. Namely, we first filtered the dataset so that both `reference_author` and `referenced_author` corresponded to leader identifiers. Then, for each month, we computed a probability vector for every leader over the five detected political communities. This representation allows us to map leaders onto an opinion simplex, as illustrated in Figure 5.2, providing a compact visualization of their temporal positioning within the

political landscape (Data are displayed at 5-month intervals).

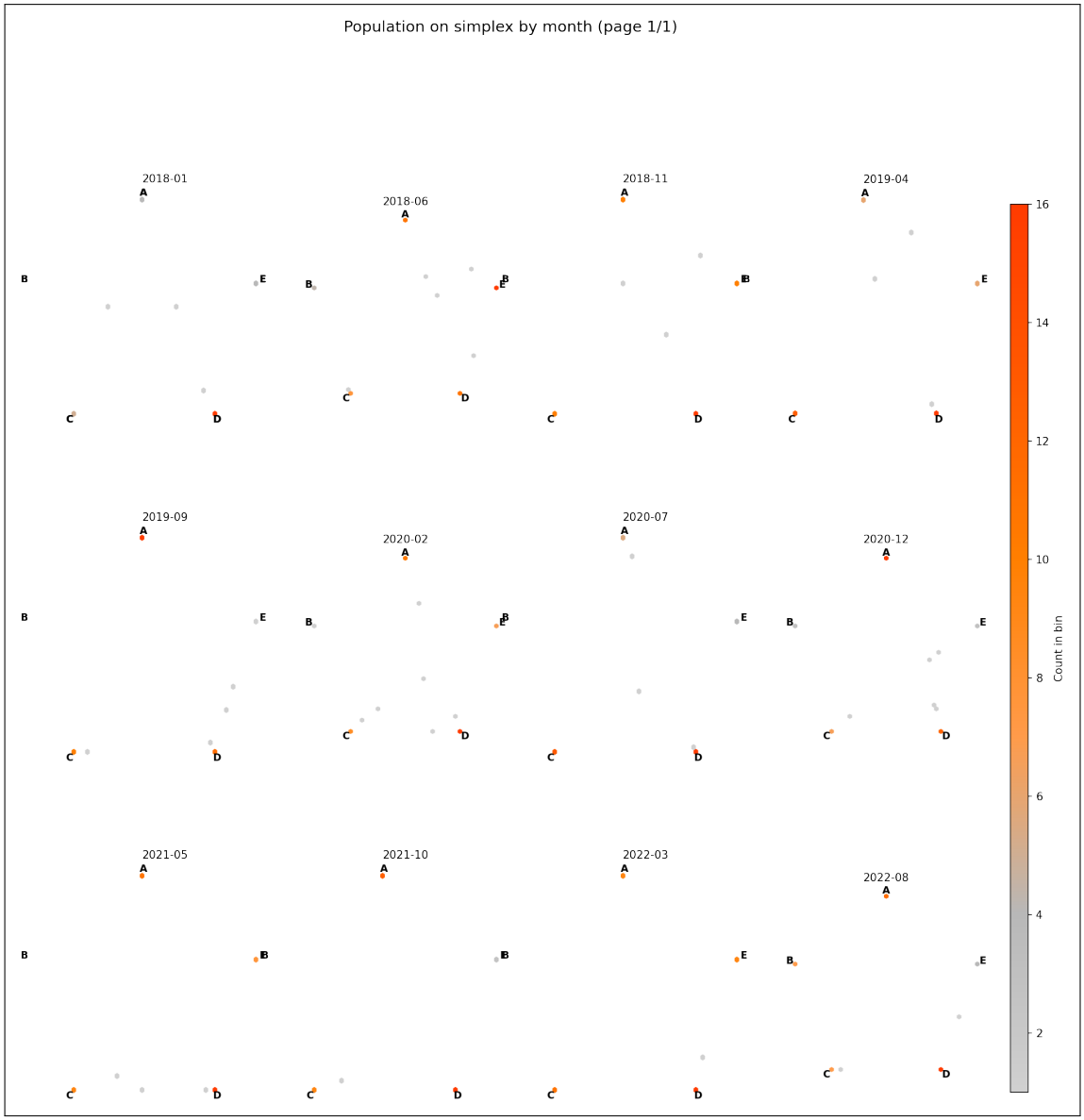


Figure 5.2: Italian information leaders represented on an opinion simplex. Data are displayed at 5-month intervals.

In contrast, constructing opinion vectors for followers is challenging: followers spread their retweets across many other followers who are not clearly embedded in a stable community, and community detection on the full retweet network performs poorly being the network structure too large and too dense.

For these reasons, we defer the estimation of follower opinion vectors to future work and, in this study, restrict our analysis to the projection over leaders who mainly interact with other leaders, whose opinion cluster can be detected.

5.2.3 Defining the COVID-19 period for analysis

In the temporal analyses presented in the following, three main annotations will be considered to indicate the key phases of the COVID-19 pandemic in Italy:

- March 2020 - Outbreak of the COVID-19 in Italy [31];
- December 2020 - Start of the national vaccination campaign [19];
- March 2022 - Beginning of the reopening phase and end of the state of emergency [11].

The interval between March 2020 and March 2022 is therefore identified as the COVID-19 period in the following analyses. This temporal window is highlighted in the plots to facilitate the comparison of network and opinion dynamics before, during, and after the pandemic.

5.2.4 From polarization to cooperation: Opinion leaders during the COVID-19 Crisis

As shown in Figure 5.2, leaders appear highly polarized, with mass concentrated near the corners of the simplex. However, based on visual inspection alone, it is not immediately clear whether significant changes occur during exceptional periods of crisis or uncertainty. Over the period analyzed (2018–2022), the event that most strongly impacted the stability of the information system and political equilibria was the COVID-19 pandemic. For this reason, we aim to characterize leaders' behaviour during this critical period. Specifically, within the System Model framework—where the control parameter is entropy or overlap—we ask: Does a strong external perturbation foster cooperation and cohesion among opinion leaders, or, in the face of a heightened threat, do communities withdraw further into their echo chambers? What is the most appropriate metrics to understand the opinion dynamics of leaders during the emergency?

To answer these questions, we compute several monthly metrics: the mean entropy (Figure 5.4, top), the mean distance of points from the centre of

the simplex (Figure 5.4, centre), the percentage of cross-community retweets (Figure 5.4, bottom), and the average monthly overlap of the opinion vectors, as defined by Equation (2.2). In addition, on a monthly basis, we compute the percentage of cross-community retweets between political accounts (Figure 5.3). These quantities reveal a very unequivocal behaviour discussed below.

- **Cross-community retweets spike.** We observe an increase in the percentage of cross-community retweets. Although the overall retweet volume grows during the COVID-19 pandemic, normalization ensures that this measure remains informative. The spike in cross-community interactions indicates that previously separated ideological groups engaged more with each other. Leaders appear to interact across community boundaries, likely reflecting a shared focus on the common crisis or heightened public attention. Interestingly, this behaviour was largely confined to the COVID-19 period, as the metric returns to pre-pandemic levels once the emergency subsides.

- **Average overlap.** The average overlap of the population, as defined in Equation 2.2 within the System Model framework, is not shown in detail because it exhibits no notable spike. The relative alignment between opinion vectors remains largely unchanged. This indicates that, although the content of opinions becomes somewhat more varied and less extreme, the underlying ideological structure and group affiliations among leaders remain largely stable over all period.

- **Entropy spike.** The entropy of the opinion vectors grows during the COVID-19 period, indicating that more leaders distribute their retweets beyond their own community, producing fewer vectors of the form [1,0,0,0,0] and more with multiple non-zero components. This points to a broader exploration of topics or positions and a less uniform discourse, probably due to a topic shift, though investigating topical dynamics lies beyond the scope of this thesis. In terms of metrics, it is worth noting that we proved that in real-world systems, overlap and entropy are not necessarily linked. The direct proportionality observed in the System Model arises from the homogeneous distribution of opinions on the simplex; however, this condition is far removed from that of real system communities. Even so, entropy appears to be meaningful for analyzing leader cooperation, as it exhibits a clear spike. In contrast, overlap does not,

since a simple rigid shift of the population on the simplex does not affect the average overlap, even though it significantly alters the dynamics.

- **Component-level trends.** Inspecting the time evolution of individual components, we detect a mild increase in communities E and C, together with a slight decrease in component A. While these shifts are not visually striking, they are consistent with the entropy and cross-community evidence, suggesting a gradual redistribution of attention across communities rather than a wholesale realignment. These communities are the one defined in Table 5.3.

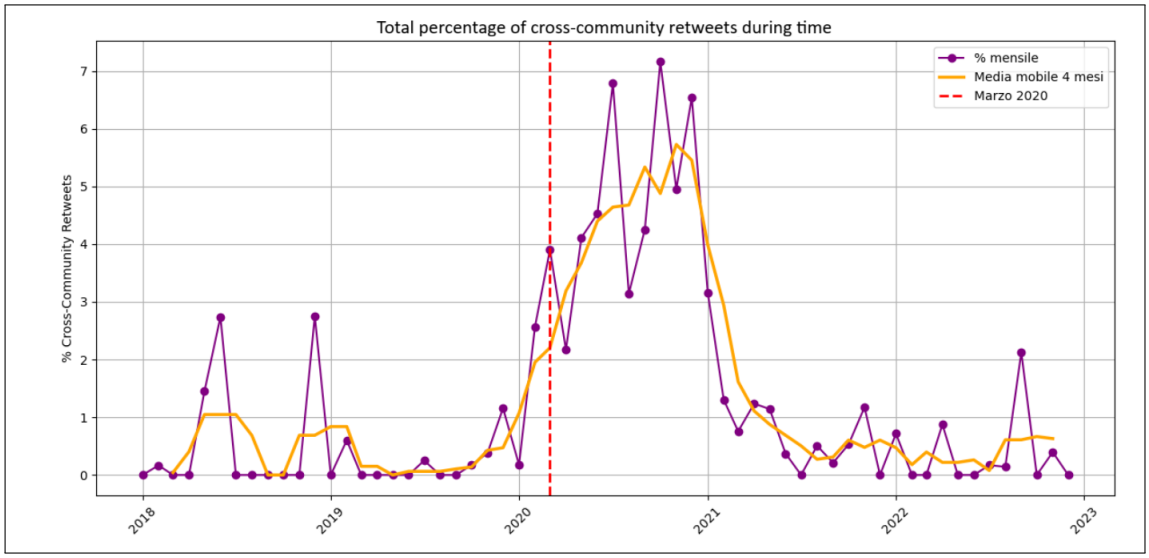


Figure 5.3: Monthly averaged percentage of cross-community retweets in the politicians network. In orange the three-month moving average.

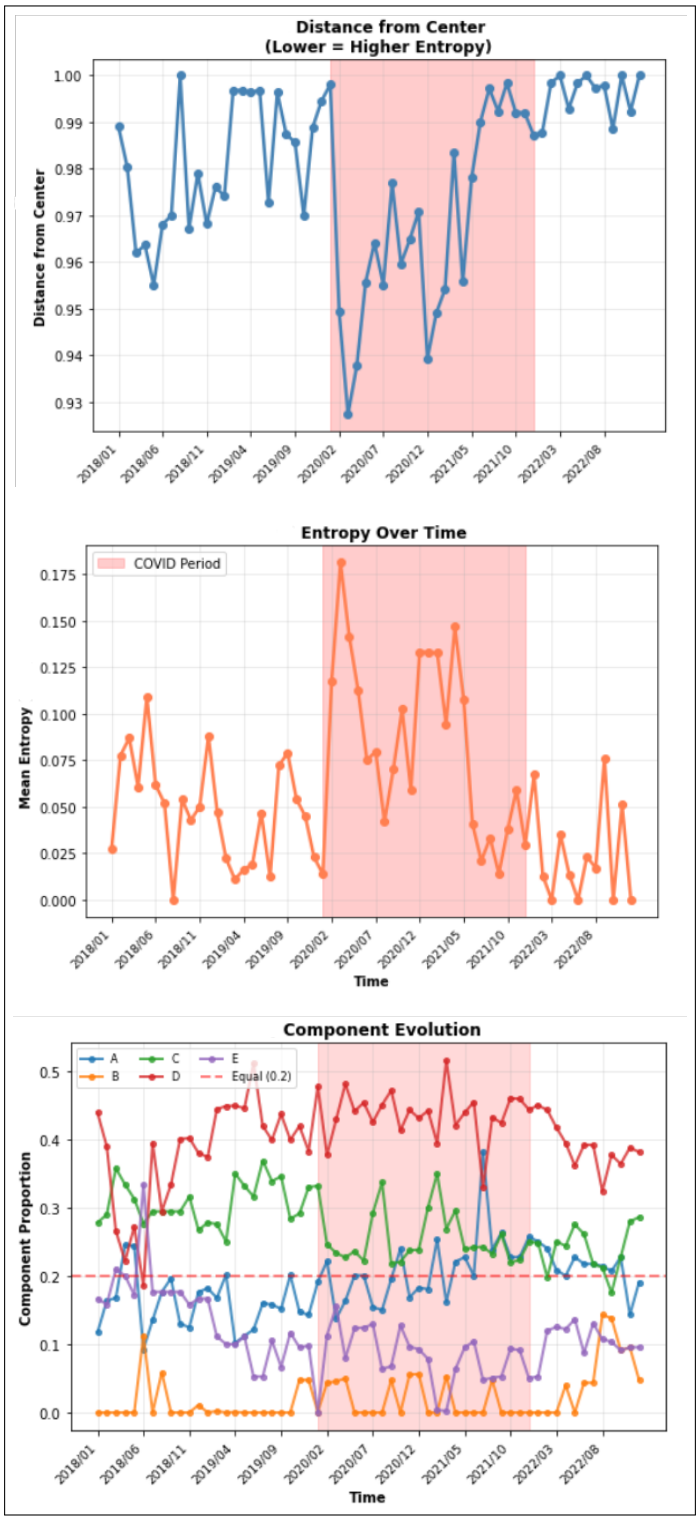


Figure 5.4: In the top, centre, and bottom panels, we show, respectively, the distance from the centre, entropy over time, and the evolution of a single component over time. The red shaded area highlights the COVID-19 period, as defined in Subsection 5.2.3.

Chapter 6

Balancedness and Stability in the X network

In this chapter, we extend our analysis of X by examining patterns of triadic balancedness and stability within the interaction data, with the aim of assessing how these properties are affected by external perturbations.

A central motivation for this study is to understand how network stability evolves during major societal disruptions, with particular attention to the COVID-19 pandemic. This period represents a natural experiment in collective behaviour, offering a rare opportunity to observe how social structures adapt under stress.

Analysing stability in real data is essential both to validate theoretical predictions and to uncover the mechanisms that drive social resilience and structural change. By studying the temporal evolution of triadic relations on Twitter, we aim to detect patterns that reflect the stability or fragility of social alliances in digital environments, and to clarify the role played by opinion leaders within these dynamics.

In Chapter 5, the System Model already provided a characterization of political leaders in terms of the formalism underlying opinion dynamics. However, that framework could not be directly extended to ordinary users. We therefore argue that social balance theory offers a sufficiently general lens to investigate emergent coordination patterns in the full network — encompassing both leaders and followers — particularly during periods of crisis.

The following sections detail the data preparation, methodological design, and empirical results, offering insights into how balance theory manifests in large-scale, real-world social systems.

6.1 Triad extraction and categorization procedure

The dataset described in Table 5.1 represents a weighted and signed (`agreement`) interaction edgelist between pairs of users (`reference_author-referenced_author`), each associated with a timestamp (`referenced_time`). By aggregating these interactions on a monthly basis for each pair, we can construct a triads dataset, where each entry represents a set of three interconnected nodes and their corresponding signed relationships, structured as follows.

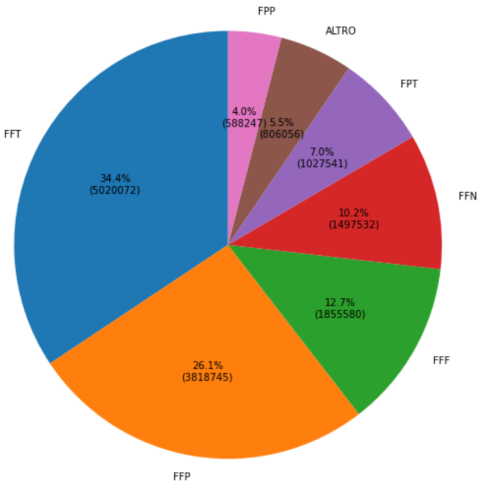
- Each triad entry includes identifiers for the three vertices (`V1_name`, `V2_name`, `V3_name`) along with their corresponding ratings (`V1_Rating`, `V2_Rating`, `V3_Rating`), which distinguish whether each vertex is a leader—classified according to one of the Newsguard categories (T for trustworthy, N for non-trustworthy, S for satirical) or P for political actors—or a follower. The dataset consists of 14,613,773 triads.
- A categorical field `rating_combination` is created to encode the triad type (e.g., `FFN`, `FFT`, `FFP`, `PPP`, etc.).
- Since the analysis is conducted on a monthly basis, a new character field `YYYY_mon` (e.g., `2019_jan`, `2020_mar`, ...) is created to identify the reference month within the period under study. Each triad is associated with a column for each month, containing a Boolean value that is `True` if the triad is balanced in that month-year, `False` if it is unbalanced, and `NaN` if the triad does not exist in that month-year.

We computed the distribution of triad types (`rating_combination` in the dataset. Approximately 60% of the triads consist of two followers and one leader labeled as P or T. About 13% of the triads are composed solely of followers (FFF), while the remaining 27% correspond to the other possible combinations of leaders and followers (See Figure 6.1a).

6.2 Monthly triad stability metrics

All metrics are computed on a monthly basis, considering only the triads with non-missing data in the given month.

Percentage Balanced (balancedness). For month t, let B_t be the number of triads marked as balanced and N_t the number of triads with



(a) Distribution of triad types.

Triad type	Condition
Only Followers	`FFF`
Only Politicians	`PPP`
Mixed (P+F)	contains P and F
Satire included	contains at least one S
Mixed content	all other combinations

(b) Triad types.

observed values. The percentage of balanced triads of is then

$$\text{pct\_balanced}(t) = 100 \times \frac{B_t}{N_t}.$$

Number of Changes. For each triad, `changes` counts how many times it transitions between balanced and unbalanced states over the entire observation period (2018–2022).

Coherence. Let `non_na` denote the number of months a triad is observed. Coherence measures the stability of a triad over time:

$$\text{coherence} = \begin{cases} 1 - \dfrac{\text{changes}}{\text{non\_na} - 1}, & \text{if non\_na} > 1, \\ \text{NA}, & \text{otherwise.} \end{cases}$$

When aggregated monthly, the reported metric is the average coherence of all observed triads in that month, expressed as a percentage.

6.3 Temporal trends in triadic balancedness and stability

In this section, we present the temporal evolution of the monthly averaged quantities defined in Section 6.2.

The results reported in Figure 6.2 indicate a clear decline in triadic balancedness during the COVID-19 period, indicating that the network became structurally less stable.

Figure 6.2: Metrics for the total triads dataset with COVID-19 Event study.

Several qualitative factors may explain this pattern. First, during the pandemic, previously unpopular actors, such as virologists and epidemiologists, gained sudden visibility and influence. Being largely apolitical, these figures were accepted across different ideological communities, temporarily softening group-based biases and resistance to cross-community sharing. Second, alliances appear to have shifted in response to an overwhelming external challenge, fostering momentary cooperation within the system. As shown quantitatively in Section 5.2.4, the leaders' network dynamics experienced a temporary convergence, reflected in reduced distances on the opinion simplex. This small-scale alignment among leaders likely contributed to a broader disruption of social balance across the wider network.

The metric *average changes* measures the mean number of state transitions (balanced↔unbalanced) per triad within each month. An increase in this indicator during the COVID-19 period suggests a surge in dynamically active

triads, i.e. configurations that changed sign more frequently. This reflects a phase of rapid opinion realignment and a decrease in social balance, which can be interpreted as an immediate collective response to the crisis. This pattern reflects a phase of rapid opinion realignment and reduced social balance, which can be interpreted as an immediate collective response to the crisis. However, these newly formed connections appear to be *weak*: the higher rate of change and lower coherence observed during the emergency period imply that these ties were unstable and unlikely to persist. In this sense, triadic volatility serves as a proxy for the fragility of the system's adaptive response. This behaviour also aligns with the increased overall retweet activity observed in that period, driven partly by the engagement of previously inactive, and likely less polarized, users.

The *coherence* parameter exhibits a complementary pattern. Defined as the temporal stability of a triad's configuration, its decline during the pandemic period indicates that active triads were more prone to changing sign, further supporting the interpretation of weaker and more transient relational structures.

6.4 Triadic dynamics by category

To investigate this instability in greater detail, triads were further aggregated into three main categories, and the same set of metrics was computed and plotted for each group. References to these plots are provided in the *Image Label* column of Table 6.1.

Category	**Percentage**	**Image Label**
Other	48.3%	6.3
Mixed P + F	38.4%	6.4
Only Followers	12.7%	6.5

Table 6.1: Distribution of follower categories with corresponding image labels.

These plots retain substantial descriptive information:

- **"Other" and "P+F" categories.** Both categories (Figure 6.3 and Figure 6.4, respectively) exhibit patterns consistent with those observed in the aggregate analysis (Figure 6.2). Specifically, we observe a concurrent decline in Balancedness and Coherence, accompanied by a rise

in Changes. This behaviour is expected, given that these two categories represent the majority of triads in the dataset. Their alignment with the general trend confirms that the overall destabilization observed during the pandemic primarily originated from triads involving at least one leader.

- **"FFF" category.** In contrast, triads composed exclusively of followers show a remarkably stable pattern over time. As depicted in Figure 6.5, their social balancedness remains consistently at 100% throughout the entire observation period, showing no response to the COVID-19 outbreak. This suggests that interactions among ordinary users are highly cohesive and structurally resilient. Since these users are not directly involved in the more volatile processes of influence or opinion leadership, their relational dynamics appear largely insulated from large-scale external shocks such as the pandemic.

Figure 6.3: Metrics for the "*other*" triads category with COVID-19 Event study.

Figure 6.4: Metrics for the "$P + F$" triads category with COVID-19 Event study. Triad category is any combination of "P" and "F".

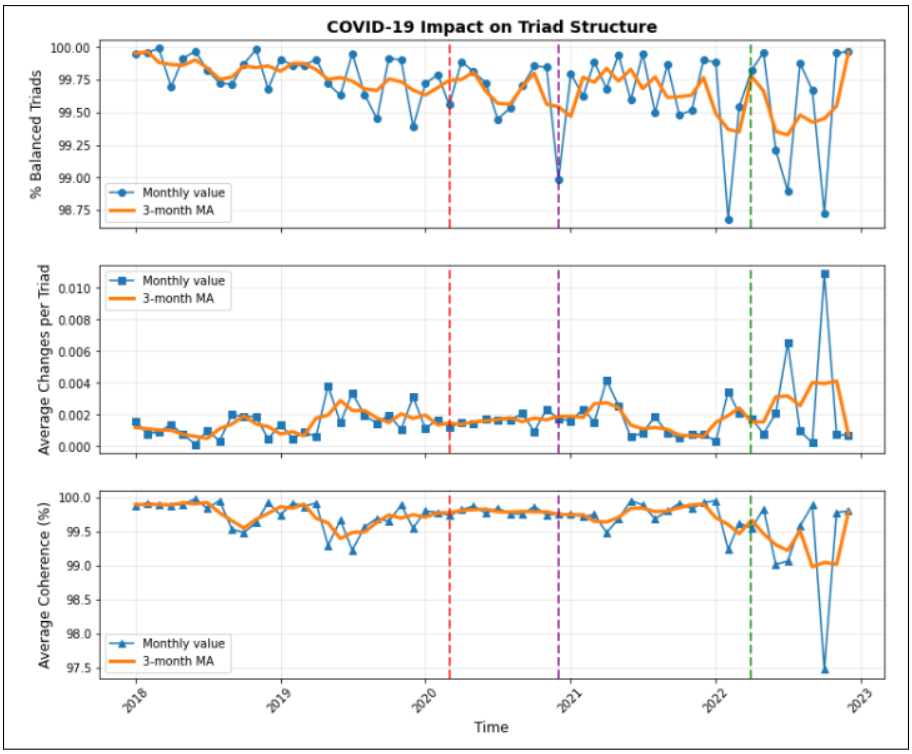


Figure 6.5: Metrics for the "FFF" triads category with COVID-19 Event study. We refer here to only followers triads.

6.5 Differences in political and information triads with followers

Once the general behaviour for each triad category has been analysed, we now provide a final characterization to complete the description.

In Figure 6.6, we report the analysis of the same metrics specified in Section 6.2, using a 3-month moving average and further distinguishing between triads "FFN" and "FFT" to assess whether the presence of a trustworthy or non-trustworthy source alters the behaviour. One can immediately observe that triads "FFT" exhibit lower balancedness, while "$P + F$" and "FFN" share similar average balancedness and coherence, and also show a comparable number of changes. Triads involving a trustworthy source of information consistently exhibit a higher number of changes, which results in lower coherence, though this does not directly explain the reduced balancedness.

To quantify this process, Figure 6.7 shows the percentages of each triad category over the total number of triads, tracked over time. It is very clear that during the pandemic crisis period, the category "FFT" experiences a striking increase, along with a milder rise in the percentage of "FFN" triads. Correspondingly, the proportion of "FFP" triads decreases. This indicates that activity and interactions involving information leaders, rather than political figures, increased significantly during the COVID period.

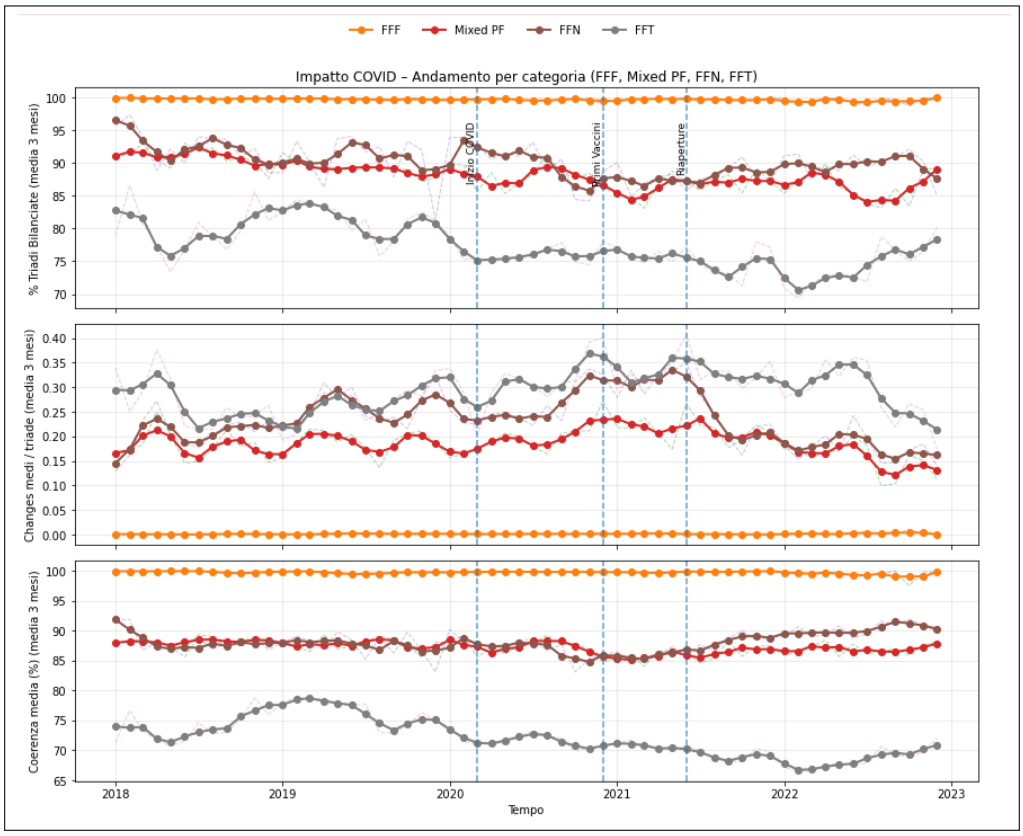


Figure 6.6: 3 months averaged key-metrics for triads division in "FFF","FFN","FFT", "$P+F$".

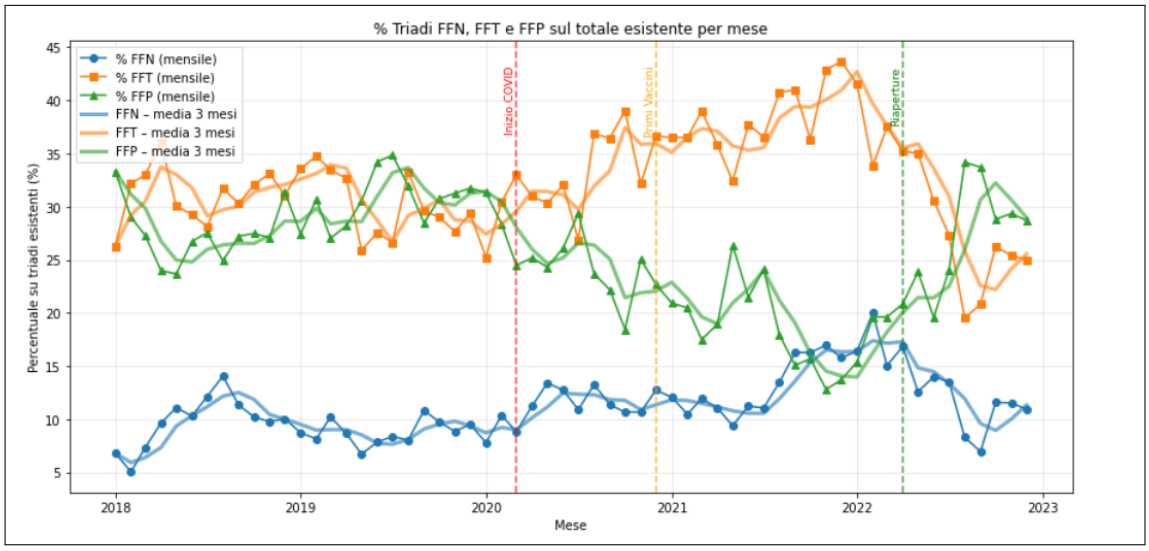


Figure 6.7: Percentage of triads types over total number of triads present in the dataset, for triads "FFN", "FFT" and "FFP".

6.6 Summary of triadic dynamics and actor roles during COVID-19

To conclude, we summarize the main findings on triadic balancedness and actor-type interactions during the COVID-19 crisis, highlighting how different behaviours relate and influence each other:

- **Global instability and structural imbalance.** The total network exhibits a clear decrease in balancedness and coherence, coupled with a rise in the number of changes during the pandemic (Figure 6.2). This reflects a systemic instability, where opinion realignments and temporary shifts in alliances disrupted previously stable configurations. The decrease in coherence and growth in changes is associated to a rise of *volatility* of relationships in the system.

- **Dominant role of mixed and heterogeneous triads.** The categories "$P + F$" and *"Other"*, which make up the majority of the dataset (Table 6.1), replicate the general trend of instability (Figures 6.3 and 6.4). This indicates that the overall network behaviour is largely driven by these heterogeneous triads involving both political and follower actors, whose dynamics were more sensitive to the crisis.

- **Stability of purely follower-based interactions.** In contrast, the "FFF" category (Figure 6.5) remains entirely stable, showing 100% balancedness throughout. This suggests that ordinary users, when interacting only among themselves, form a stable and insulated subnetwork, less exposed to the influence-driven volatility that affects leader-related triads, and are characterized of higher coherence.

- **Information triads and trustworthiness effects.** A finer-grained analysis (Figure 6.6) shows that triads involving non-trustworthy information sources ("FFT") experience the lowest balancedness and coherence. Meanwhile, trustworthy information actors ("FFN") lead to more dynamic interactions (more changes), but without improving balance. During the pandemic, such triads surged in proportion (Figure 6.7), while political ones decreased, indicating a shift from political to informational engagement. This behavioural shift in influence dynamics is probably a key driver of the broader network destabilization observed.

Chapter 7

Conclusions

Understanding how societies respond to crises enables us to anticipate political instability during future disruptions and, crucially, to defend democracy: by modelling opinion dynamics, we can detect anomalous patterns indicative of manipulation and distinguish organic consensus building from coordinated influence operations. This thesis investigates how opinions evolve under large external perturbations, with a focus on the COVID-19 pandemic as a shock to social systems. We study whether cooperative behaviour among leaders can emerge during crises, what system-level consequences follow, and which modelling frameworks best describe and forecast these dynamics.

Summary of the work

We built the work on the Sîrbu–Loreto framework to model agreement/disagreement dynamics with modulated external information and reframed the process within Social Balance Theory (SBT), so that individuals interactions outcomes could be represented as a signed, time-evolving network. The project combined three components:

1. **Theory and modelling.** We reviewed the literature from Heider's social balance to modern opinion dynamics and chose the Sîrbu–Loreto model as a baseline. The original model was reproduced and discussed.

2. **SBT re-framing and extension.** We chose to adopt social balance theory to opinion dynamics, mapping interaction outcomes to a signed network and measured triadic balancedness in the original model, to study it under a different perspective. We then proposed an extension introducing a *social pressure* term η that modifies the effective overlap

via triadic cues, operationalized through a weighted triadic signal τ_{ij}. We found out that social balance originates from dyadic interaction and social pressure, contrarily, damages social balance.

3. **Empirical analysis on Twitter/X.** We constructed monthly opinion vectors for users and leaders in the Italian political sphere (2018–2022), creating a system model kind of population. From the data, we derived monthly signed triads of real users and tracked balancedness, changes, and coherence across the pandemic window and recorded spikes under the effect of the external perturbation.

Main findings

Triadic balancedness in the model.

- In peer-only dynamics, triadic balancedness $\mathcal{B}$ settles to stable values that *do not depend on* K but *do reflect* initial cohesion. Consistent with the literature, the system organises at high (but not perfect) balancedness in favourable conditions.
- Introducing social pressure η (triad-informed effective overlap) yields a counter-intuitive result: increasing η *reduces* balancedness, both with and without leaders. This suggests that, in this setting, triadic balance emerges *spontaneously from dyadic dynamics*, while strong conformance to local triadic pressure can disrupt the system's natural route to balance.
- Noise ϵ erodes balancedness when large, by decoupling p_{agree} from opinion similarity. With leaders present, the absolute level of $\mathcal{B}$ depends on the match between source polarization a and the emergent opinion geometry (e.g., high a aligns with polarized follower populations but depresses balancedness when populations are in a cohesion state).

Empirical patterns in Italian political Twitter/X (2018–2022).

- The leaders' retweet network partitions into five communities with substantial modularity. During COVID-19, we proved quantitatively that the leaders' network undergoes a transition that can be characterized in terms of our model: leaders got closer on the simplex, indicating a temporary relaxation of echo-chamber boundaries and the emergence of interactions across different ideological blocs.

- In the triad time series Leaders and followers are behaving in a complete different way. Social stability among normal users is not affected by the perturbation, but when leaders take part in the triads with users, the balancedness decreases. This, in addition to the leaders cooperation that emerged from data, means that the real change, on opinion dynamics level, happens on leaders. The way the changing in leaders dynamics propagate to followers needs further study.

Conceptual implications

- **Emergence of collective phenomena** The thesis shows that structural properties derived from triadic relations (balancedness, coherence) can be linked to macro phenomena (cooperation) under external perturbations.
- **Leaders as control parameters.** In the dynamics we came across leaders and followers don't play absolutely the same role. Leader's behaviour shapes follower's one and follower's stability is not affected by the external perturbation.
- **More social pressure does not imply more social stability.** The finding that larger η *reduces* balancedness challenges the intuition that explicit triadic interaction improves stability. In our setting, balance is an emergent outcome of dyadic dynamics; imposing strong social pressure may introduce conflicting constraints that destabilize local adaptation.

Limitations

- **Signed-network construction from data.** Month-level aggregation and last-interaction (or averaging) rules for signs inevitably simplify temporal causality. We got rid of the post-id layer by averaging but that is an extreme semplification.
- **Model idealizations.** The model chooses randomly the interacting people, as an extreme and useful simplification.

Future directions

- **Performing statistical validation on the signed network links.** To solve the simplification made when filtering out the post-id layer, a statistical validation of links should be performed, to both decrease the number of links without filtering and to avoid simple averages. This could lead to more statistically valid results.
- **Heterogeneous triadic pressure.** Replace global η with node- or community-specific η_i, or context-dependent $\eta(t)$ that rises during emergencies and relaxes afterward. This could reproduce the different group-dynamics observed in the Italian communities.
- **Richer exposure models.** Introduce network topology and platform mediation (e.g., algorithmic recommender biases) to move beyond fully mixed peer selection; test how structural features interact with a, ϵ, and η.
- **Partial/weak balance.** Track both strong and weak balance metrics and motif-resolved pathways (e.g., $++-$ vs. $---$) to highlight which motif transitions dominate instability during shocks and what changes to the balancedness this could lead to.

Bibliography

[1] T. Antal, P. L. Krapivsky, and S. Redner. "Dynamics of social balance on networks". In: *Physical Review E* 72.3 (Sept. 2005). ISSN: 1550-2376. DOI: 10.1103/physreve.72.036121.

[2] Albert-László Barabási and Réka Albert. "Emergence of Scaling in Random Networks". In: *Science* 286.5439 (Oct. 1999), pp. 509–512. ISSN: 1095-9203. DOI: 10.1126/science.286.5439.509.

[3] Alessandro Bellina et al. *Language bubbles in online social networks*. 2025. arXiv: 2507.13068 [physics.soc-ph]. URL: https://arxiv.org/abs/2507.13068.

[4] Emanuele Brugnoli and Donald Ruggiero Lo Sardo. "Community-based Stance Detection". In: *Proceedings of the 10th Italian Conference on Computational Linguistics (CLiC-it 2024)*. Ed. by Felice Dell'Orletta et al. Pisa, Italy: CEUR Workshop Proceedings, Dec. 2024, pp. 98–105. ISBN: 979-12-210-7060-6. URL: https://aclanthology.org/2024.clicit-1.13/.

[5] Emanuele Brugnoli et al. "Fine-Grained Clustering of Social Media: How Moral Triggers Drive Preferences and Consensus". In: *Proceedings of the 16th International Conference on Agents and Artificial Intelligence, ICAART 2024, Volume 3, Rome, Italy, February 24-26, 2024*. Ed. by Ana Paula Rocha, Luc Steels, and H. Jaap van den Herik. SCITEPRESS, 2024, pp. 1405–1412. DOI: 10.5220/0012595000003636.

[6] Dorwin Cartwright and Frank Harary. "Structural Balance: A Generalization of Heider's Theory". In: *Psychological Review* 63.5 (1956), pp. 277–293. DOI: 10.1037/h0046049.

[7] Claudio Castellano, Santo Fortunato, and Vittorio Loreto. "Statistical physics of social dynamics". In: *Reviews of Modern Physics* 81.2 (May 2009), pp. 591–646. ISSN: 1539-0756. DOI: 10.1103/revmodphys.81.591.

[8] James A. Davis. "Clustering and Structural Balance in Graphs". In: *Human Relations* 20.2 (1967), pp. 181–187. DOI: 10.1177/001872676702000206.

[9] Guillaume Deffuant et al. "Mixing beliefs among interacting agents". In: *Advances in Complex Systems* 3.1-4 (2000), pp. 87–98. DOI: 10.1142/S0219525900000078.

[10] Morris H. DeGroot. "Reaching a Consensus". In: *Journal of the American Statistical Association* 69.345 (1974), pp. 118–121. DOI: 10.1080/01621459.1974.10480137.

[11] Dipartimento della Protezione Civile. *Coronavirus: il 31 marzo si chiude lo stato di emergenza*. URL: https://www.protezionecivile.gov.it/it/notizia/coronavirus-il-31-marzo-si-chiude-lo-stato-di-emergenza/ (visited on 10/09/2025).

[12] Giuseppe Facchetti, Giovanni Iacono, and Claudio Altafini. "Computing global structural balance in large-scale signed social networks". In: *Proceedings of the National Academy of Sciences* 108.52 (2011), pp. 20953–20958. DOI: 10.1073/pnas.1109521108.

[13] Mirta Galesic, Henrik Olsson, T. M. Pham, et al. "Experimental evidence confirms that triadic social balance can be achieved through dyadic interactions". In: *npj Complex* 2 (2025), p. 1. DOI: 10.1038/s44260-024-00022-y.

[14] A. Gallo, D. Garlaschelli, R. Lambiotte, et al. "Testing structural balance theories in heterogeneous signed networks". In: *Communications Physics* 7 (2024), p. 154. DOI: 10.1038/s42005-024-01640-7.

[15] Sharad Goel, Winter Mason, and Duncan J. Watts. "Real and perceived attitude agreement in social networks". In: *Journal of Personality and Social Psychology* 99.4 (2010), pp. 611–621.

[16] Mark S. Granovetter. "The Strength of Weak Ties". In: *American Journal of Sociology* 78.6 (1973). Accessed: 8 Oct. 2025, pp. 1360–1380. URL: http://www.jstor.org/stable/2776392.

[17] Fritz Heider. "Attitudes and Cognitive Organization". In: *Journal of Psychology* 21.1 (1946), pp. 107–112. DOI: 10.1080/00223980.1946.9917275.

[18] Richard A. Holley and Thomas M. Liggett. "Ergodic Theorems for Weakly Interacting Infinite Systems and the Voter Model". In: *The Annals of Probability* 3.4 (Aug. 1975), pp. 643–663. DOI: 10.1214/aop/1176996306.

[19] Istituto Superiore di Sanità. *Piano strategico nazionale di vaccinazione COVID-19*. URL: https://www.epicentro.iss.it/vaccini/covid-19-piano-vaccinazione (visited on 10/09/2025).

[20] David Lazer et al. "Computational Social Science". In: *Science* 323.5915 (2009), pp. 721–723. DOI: 10.1126/science.1167742.

[21] Christopher D. Manning and Hinrich Schütze. *Foundations of Statistical Natural Language Processing*. Cambridge, MA: MIT Press, 1999. ISBN: 0-262-13360-1.

[22] NetworkX Developers. *louvain_communities — NetworkX Documentation*. https://networkx.org/documentation/stable/reference/algorithms/generated/networkx.algorithms.community.louvain.louvain_communities.html. Accessed: 2025-10-06. 2025.

[23] NewsGuard Technologies, Inc. *News Reliability Ratings*. https://www.newsguardtech.com/solutions/news-reliability-ratings/. Accessed: 2025-10-06. 2025.

[24] T. M. Pham et al. "Balance and fragmentation in societies with homophily and social balance". In: *Scientific Reports* 11 (2021), p. 17188. DOI: 10.1038/s41598-021-96357-9.

[25] T. M. Pham et al. "Empirical social triad statistics can be explained with dyadic homophylic interactions". In: *Proceedings of the National Academy of Sciences* 119.21 (2022), e2121103119. DOI: 10.1073/pnas.2121103119.

[26] Somaye Sheykhali, Amir Hossein Darooneh, and Gholam Reza Jafari. "Partial balance in social networks with stubborn links". In: *Physica A: Statistical Mechanics and its Applications* 548 (2020), p. 123882. ISSN: 0378-4371. DOI: 10.1016/j.physa.2019.123882.

[27] ALINA SÎRBU et al. "COHESION, CONSENSUS AND EXTREME INFORMATION IN OPINION DYNAMICS". In: *Advances in Complex Systems* 16.06 (Aug. 2013), p. 1350035. ISSN: 1793-6802. DOI: 10.1142/s0219525913500355.

[28] Alina Sîrbu et al. "Opinion Dynamics with Disagreement and Modulated Information". In: *Journal of Statistical Physics* 151.1–2 (Feb. 2013), pp. 218–237. ISSN: 1572-9613. DOI: 10.1007/s10955-013-0724-x.

[29] Alina Sîrbu et al. "Opinion Dynamics: Models, Extensions and External Effects". In: (May 2016), pp. 363–401. ISSN: 1860-0840. DOI: 10.1007/978-3-319-25658-0_17.

[30] Duncan J. Watts and Steven H. Strogatz. "Collective dynamics of 'small-world' networks". In: *Nature* 393.6684 (1998), pp. 440–442. DOI: 10.1038/30918.

[31] Wikipedia contributors. *Pandemia di COVID-19 in Italia*. URL: https://it.wikipedia.org/wiki/Pandemia_di_COVID-19_in_Italia (visited on 10/09/2025).