Politecnico di Torino

ScuDo
Scuola di Dottorato ~ Doctoral School
WHAT YOU ARE, TAKES YOU FAR

UNIVERSITÀ DI TORINO

Doctoral Dissertation
Doctoral Program in Pure and Applied Mathematics (37^{th}cycle)

Subset Selection and Detection Problems in Opinion Dynamics Models

By

Roberta Raineri

******

Supervisor:
Prof. Fabio Fagnani, Supervisor
Prof. Giacomo Como, Co-Supervisor

Doctoral Examination Committee:
Prof. Sophie Fosson, Politecnico di Torino
Prof. Barbara Franci, Politecnico di Torino
Prof. Paolo Frasca, CNRS & Université Grenoble Alpes
Prof. Sekhar Tatikonda, Yale University
Prof. Mattia Zorzi, Università di Padova

Politecnico di Torino - Università di Torino
2025

Declaration

I hereby declare that, the contents and organization of this dissertation constitute my own original work and does not compromise in any way the rights of third parties, including those relating to the security of personal data.

Roberta Raineri
2025

* This dissertation is presented in partial fulfillment of the requirements for the degree of ***Philosophiæ Diploma*** **(PhD degree)** in **Pure and Applied Mathematics**.

Abstract

In the last decades, the study of Opinion Dynamics in Social Networks has gained significant attention, as it provides essential tools to understand, predict, and possibly influence the dissemination of information and collective decision-making processes. This growing relevance is not limited to academic domains, as confirmed by a McKinsey & Company report which states that “marketing-induced consumer-to-consumer word of mouth generates more than twice the sales of paid advertising”, highlighting the critical economic value of understanding how opinions spread and evolve. Different examples further emphasize the tangible consequences of these dynamics. The strategic use of Twitter bots during presidential elections and during the first impeachment of U.S. President Donald Trump illustrates how automated agents can play a central role in manipulating public opinion. Similarly, for companies conducting surveys or opinion polls, it is crucial to identify which individuals should be asked in order to obtain reliable insights on the overall society with minimal cost.

This dissertation investigates mathematical models of Opinion Dynamics on Social Network, focusing on two central problems: a Subset Selection Problem and a Detection Problem. The first problem focuses on identifying the "most informative" agents, meaning the subset of individuals that, if questioned, can most accurately represent the average opinion within the entire network. We formalize a selection criterion based on variance reduction estimation and apply it across various

settings built on the widely studied French-DeGroot and Friedkin-Johnsen opinion dynamics models, also accounting for phenomena such as stubborn agents and noisy interactions. Although the problem is NP-Hard, we demonstrate that, under certain reasonable assumptions, the objective function is submodular, thereby enabling the use of efficient greedy algorithms. Additionally, by working within a graph-based framework, we identify connections between the optimal selection and key network properties and topology. The second problem aims at detecting stubborn agents, such as bots or persistent opinion holders, without prior knowledge of the underlying social network. We formulate the stubborn nodes identification and their influence estimation problems as a low-rank approximation problem. We then propose a novel detection method grounded in matrix factorization techniques, specifically interpolative decomposition, that infers latent social structure and identifies stubborn behaviors directly from opinion observations. We determine sufficient conditions on the model parameters to ensure the algorithm's resilience even in the presence of noisy data or finite time observations.

Overall, this thesis bridges theoretical modeling with practical applications of opinion dynamics, offering scalable and robust solutions suited for real-world environments. The proposed methods contribute to advancing the understanding of information spread in modern interconnected societies.

Contents

List of Symbols and Abbreviations

The next list describes several symbols and abbreviations that will be later used within the body of the dissertation.

Graphs

$\Lambda = [\lambda]$ In the Friedkin-Johsen model indicates the individual openness to interactions

$\mathcal{G} = (\mathcal{V}, \mathcal{E}, W)$ Directed graph with set of nodes $\mathcal{V}$, set of edges $\mathcal{E}$ and weighted matrix W

FJ Abbreviation for Friedkin-Johnsen model

P Normalized weight matrix, i.e. $P = D^{-1}W$ where $D = \text{diag}(w)$ and $w = W\mathbb{1}$.

W Weight matrix of the graph

x Equilibrium or limit opinion of the model

$x(t)$ Vector of opinions of network agents at time t

Matrices

$[\lambda]$ Given $\lambda \in \mathbb{R}^n$, it refers to a $n \times n$ diagonal matrix Λ with $\Lambda_{ii} = \lambda_i$

$\mathbb{1}$ All-1 column vector

$\mathcal{A}$ Calligrafic letters are used to denote sets

Π Permutation matrix

$\sigma_i(A)$ i-th largest singular value of matrix A. It holds $\sigma_i^2(A) = \lambda_i(AA')$

Aa Uppercase letters are used to denote matrices and random variables

aa Lowercase letters are used to denote vectors

$\text{diag}(A)$ Indicates the diagonal elements of matrix A

$\text{rank}(A)$ Rank of matrix A

$\text{tr}(A)$ Trace of $n \times n$ matrix A, i.e. $\text{tr}(A) = \sum_{i=1}^{n} A_{ii}$

I Identity matrix

$X_{-\mathcal{A}}$ Given $X \in \mathbb{R}^{\mathcal{V}\times\mathcal{V}}$ where $\mathcal{V}$ is a finite set and given $\mathcal{A}, \mathcal{B} \subseteq \mathcal{V}$, it indicates $X_{\mathcal{V}\setminus\mathcal{A}}$

$X_{\mathcal{A}\mathcal{B}}$ Given $X \in \mathbb{R}^{\mathcal{V}\times\mathcal{V}}$ where $\mathcal{V}$ is a finite set and given $\mathcal{A}, \mathcal{B} \subseteq \mathcal{V}$, it indicates the submatrix of X having rows in $\mathcal{A}$ and columns in $\mathcal{B}$

$X_{\mathcal{B}}$ Given $X \in \mathbb{R}^{\mathcal{V}\times\mathcal{V}}$ where $\mathcal{V}$ is a finite set and given $\mathcal{A}, \mathcal{B} \subseteq \mathcal{V}$, it indicates $X_{\mathcal{V}\mathcal{B}}$

Number Sets

$\mathbb{N}$ Natural Numbers

$\mathbb{R}$ Real Numbers

$\mathbb{R}_+$ Positive Real Numbers

CHAPTER 1

Introduction

1.1 Opinion Dynamics on Social Networks

Last decades have seen an increasing attention on the study of Opinion Dynamics on Social Networks, a field which focuses on the formation and evolution of agents' opinions about a given topic over time as a consequence of social relationships and connections. Agents may in general refer to single individuals or organizations which take part of the considered society. Understanding the structure and influence mechanisms within such networks is crucial to predict, analyze and possible control the flow of information and decision-making processes [1]. The proliferation of social media platforms has significantly amplified this phenomenon, allowing people to share news, opinions and preferences in a simple and immediate way. Some key examples that highlight the central role of social network dynamics in real-world scenarios include the strategic use of social media by the Obama campaign in the U.S. presidential elections of 2008 [1]; the widespread dissemination of fake news by Facebook users on topics ranging from U.S. elections [2] to public health issues like vaccines [3, 4] and the COVID-19 pandemic [5]; and the role of Twitter bots during the first impeachment of U.S. President Donald Trump [6, 7].

The mathematical modeling of opinion dynamics naturally complements the classical approaches of social network analysis [8], which primarily focused on structural properties of networks, and computational social science [9], which aims to understand collective human behavior through the analysis of available social datasets. Various opinion dynamics models have been proposed in mathematical sociology and sociophysics in an endeavor to portray the rich and diverse behaviors of social groups, capturing the most important effects of the temporal evolution of behaviors and attitudes of the social agents caused by their interactions. One of the most popular approaches is to consider them as a particular **multi-agent system**, a rapidly advancing field that attracts researchers across disciplines, including social sciences, economics, physics and engineering (see [10–14] for comprehensive reviews of its history and recent developments).

In engineering and mathematical literature, the most extensively studied models rely on **iterative opinion averaging** as the key driving force of opinion formation. This idea traces back to early work on social power dynamics [15] and rational consensus formation [16, 17]. Recent empirical studies have confirmed the prevalence of averaging behavior in small groups [18], medium-size groups [19], and large-scale online communities [20]. In an agent-based model each individual is represented by an agent with some mathematical model describing its dynamics, i.e. how individuals process new sources of information but with a confirmation bias towards the individual's established views. A central element of averaging-based models is the weighted directed graph of social influence, which may be static or may co-evolve with opinions. This graph depicts social ties among individuals and quantifies the weights each agent assigns to those they are connected with.

As discussed extensively in the following chapter, the most prominent model of opinion formation is arguably the **French-DeGroot** dynamics [15, 16], according to which agents simply update simultaneously their opinions taking the weighted average of the opinions of adjacent nodes [12]. Many different generalizations of this model have been developed, from asynchronous updates for gossip models [21] to more general time-varying networks, where the topology evolves over time [22, 23] or where the system is coupled with other social dynamics [24, 25], in order to investigate how different phenomena may arise. A particular case is where one or more agents are stubborn, i.e. individuals whose opinions remain unchanged independent of the others'. This scenario was first introduced by [26] and then widely studied to analyse phenomena such as opinion fluctuations and disagreement

[27], adopted to model opinion formation in deliberative groups [28] and validated with real experimental data [29]. A seminal and extensively studied model of opinion formation that will be fundamental in the following, naturally extending the French-DeGroot iterative averaging dynamics, is the **Friedkin-Johnsen** model [30, 31]. In addition to the weighted influence graph, this model assigns each agent a constant innate opinion, factored into each opinion update iteration. Originally defined as the agent's initial opinion [31], the innate opinion can also be shaped by the agent's prejudices or some other sources of information, such as social media. The strength of an agent's anchorage to their innate opinion is regulated by an additional parameter, interpreted as the agent's susceptibility to social influence [10]; some studies interpret this constant as a measure of conformity under group pressure [32]. While remaining linear, the Friedkin-Johnsen model can result in diverse distributions of final opinions, ranging from consensus to multimodal polarized states [10].

1.2 Focus of the Thesis

The above analysis provides a general overview of opinion dynamics models, aiming to contextualize the field of application of this dissertation. Additional references and a more detailed literature review will be presented at the beginning of the following chapters to support further in-depth exploration. We now proceed to outline the two principal research problems investigated in this dissertation, together with the underlying motivations that have shaped the direction and relevance of this study.

1.2.1 Subset Selection Problem

Research Problem and Motivation. The first question we address is how to identify the subset of agents whose opinions, if observed, would enable us to accurately infer the opinions of the entire network. Namely, we will identify the so called *most informative* individuals of the social network. Accurately estimating the average opinion of a society is a crucial challenge given its implications for policy-making, strategic business decisions and study of sociological trends. Seminal works, such as [33], show how public opinion plays a critical role in creating policies that guide decisions and actions in areas such as government, business or society. Moreover, accurate forecasting may also be helpful in trends and misinformation spread mon-

itoring. Typically, opinion polls are a valuable tool commonly used to effectively capture public opinion. However, since it is not feasible to survey the entire population, the main challenge lies in selecting a group of individuals that best represents the population of interest.

Based on preliminary works by [34–36] we formalize the optimization problem considering as target function the Variance Reduction. Specifically, we let the optimal subset of variable to observe be the one that maximizes the reduction of the variance on the variable to estimate. The context of application considered is a social network where the observable variables are the opinions of the individuals and the quantity to be estimated is the average opinion of the community. Individuals of the society update their opinions according to a linear averaging dynamics. For a more comprehensive analysis, we addressed different possible models. We distinguished the nodes as stubborn individuals, that keep their opinion unchanged, potentially representing influential agents like politicians or authorities, and regular ones, open to social interactions [1]. The opinions of regular individuals are assumed to be determined by network interactions plus an additional random noise, according to a DeGroot model [16] with stubborn agents [27, 37]. Referring to [38–40], we outline multiple real-world scenarios that could explain the role of the noise term. It may arise from transmission errors, e.g. an agent receiving the information via a third party acting as a distorting trasmission channel, such as newspapers or TV news reports. Alternatively, it could result from ambiguity introduced by stubborn agent themselves, e.g. a politician deliberately exposing a deceiving version of their opinion during a public speech. Technically, in this framework we will assume the network structure to be known and to have no prior information about the network nodes' opinions. This represents a standard inferential problem, and the analysis, widely presented in Chapter 3 and 4, will heavily rely on a statistical approach.

This problem can be considered more in general as an instance of a subset selection problem. This is a key question in multiple data-driven applications, spanning a wide range of problems, including for instance feature selection in machine learning [34, 35, 41], sensor placement for environmental monitoring [36], or even smart testing problem, which seeks to develop effective strategies for controlling the spread of epidemics [42]. Far from being exhausted, research on this topic continues to evolve, as evidenced by the recent contributions of [43, 44], which confirm the persistent relevance of this field.

Literature Review. This problem falls within the well-known class of subset selection problems, which are NP-complete [34] . Referring to previous literature, we highlight how many criteria have been proposed for characterizing the quality of placements, such as entropy or mutual information [36]. A typical sensor placement technique is to greedily add sensors where uncertainty about the phenomena is highest, that is, the highest entropy location [45]. Unfortunately, as analysed by [36], this criterion suffers from a significant flaw: entropy is an indirect criterion, not considering the prediction quality of the selected placements. The highest entropy set, that is, the sensors that are most uncertain about each other's measurements, is usually characterized by sensor locations that are as far as possible from each other. Thus, the entropy criterion tends to place sensors along the borders of the area of interest. An alternative criterion, proposed by [46], mutual information, seeks to find sensor placements that are most informative about unsensed locations. A main limitation of this optimization criterion is that it directly measures the effect of sensor placements on the posterior uncertainty, requiring an accurate model for the sensor distributions and a high computational cost [36]. To address these limitations, an alternative approach proposed by [34–36] involves adopting the variance reduction on the target variable as the objective function. Given the computational complexity of the resulting optimization problem, the use of greedy algorithms has been explored as a tractable solution. In particular, [34] introduces the notion of suppressor variables as a sufficient condition to ensure the effectiveness of the greedy algorithm. However, this condition is generally difficult to verify and, as will be extensively discussed in the following chapters, is not well-suited to the opinion dynamics framework. More broadly, to the best of our knowledge, this problem has not been explicitly studied within the context of Opinion Dynamics. As a result, the available results in the literature often rely on assumptions tailored to other domains and cannot be directly applied to the setting considered in this dissertation.

Main Contribution. We succeed in deriving an explicit formulation for the variance reduction function, both in the specific context of opinion dynamics and in a more general framework involving an arbitrary set of variables of interest. In the social network setting, we establish explicit connections with well-known centrality measures and demonstrate that their performance is outperformed by the proposed method. Furthermore, we prove that the Variance Reduction function, used as the objective in the subset selection problem, exhibits the property of submodularity

within this specific setting under some non restrictive assuptions. This property enables the design of a greedy algorithm that identifies high-quality suboptimal solutions while significantly reducing the computational complexity associated with the problem. From an applicative perspective, we conducte a detailed analysis of the two main models in opinion dynamics. For the French-DeGroot model, we provide network-related insights to clarify the conditions under which the greedy algorithm is applicable. Conversely, for the Friedkin-Johnsen model, we succeed in establishing a set of selection criteria for the observable set in both the high and low network influence regimes, revealing a strong connection with hitting times in Markov chains.

A preliminary version of these results was published in [47]. The specific analysis of the problem in the case of French-DeGroot model with stubborn agents and the formulation of the greedy algorithm have been published in [48], which extension is available in [49]. A more comprehensive paper is actually in progress.

1.2.2 Detection Problem

Research Problem and Motivation. The second key question we address concerns the detection of stubborn behavior in social networks. Online social networks have reached a large pervasiveness and relevance in opinion formation and content dissemination. Understanding the dynamics of information propagation within these networks is crucial for predicting inclinations and preferences in order to design targeting actions or to prevent critical issues. However, individuals often exhibit stubborn behavior [27] and tend to resist change, leading to the emergence of phenomena like opinion polarization. Furthermore, stubbornness may be associated with the behavior of bots, which share polarized and often misleading contents, without being effected by other network agents, thereby amplifying for instance the diffusion of fake news [7]. For this reason, a main research problem in opinion dynamics is the detection of stubborn individuals [7, 50], extremists and automatic bots [7, 51, 6], providing tools to intervene and take specific countermeasures.

Literature Review. Most state-of-the-art inference methods are based on the extensive collection and processing of data, often requiring preliminary network knowledge [7, 6, 52] and the computation of centrality measures [12]. This approach implies multiple limitations as observing and analyzing the complete structure of

social networks can be difficult or totally impossible. Examples can be deliberative groups or forum discussions, where interpersonal influences can be inferred only a posteriori after deliberation [28], or platforms like *Truth* or *4chan*, for which there is no clear definition of social structure due to user anonymity and with no concept of "friends" or "followers". Classical system identification techniques used in control theory (see [53] and references therein) also find limited applicability in this context. Indeed, without additional assumptions about the network structure, they typically require a number of observations proportional to the number of links, leading to a too high computational cost [54].

Main Contribution. We propose a novel algorithm for bot detection without requiring prior knowledge of the network structure or its explicit reconstruction. Due to their nature, bots are designed to spread predetermined content without participating in the network's opinion dynamics; we therefore model them as stubborn agents within a French–DeGroot opinion framework. By leveraging linear algebraic techniques, particularly matrix decomposition methods, we demonstrate that it is possible to infer essential structural information directly from opinion data. Our approach relies solely on the observation of agents' opinions across multiple topics, a realistic and practical assumption in the context of Big Data. Furthermore, the proposed method remains effective in the presence of noise and when only finite-time observations are available.

The results of this work have been published in [55].

1.3 List of Publications

Journal Papers

- [55] Raineri R., Ravazzi C., Como G., Fagnani F., “Detecting Stubborn Behaviors in Influence Networks: A Model-Based Approach for Resilient Analysis”, in IEEE Control Systems Letters, vol. 8, pp. 2343-2348, 2024 (**Chapter 5**)
- [56] Raineri R., Zino L., Proskurnikov A., “FJ-MM: Friedkin-Johnsen Opinion Dynamics Model with Memory and Multi-Hop Social Influence”, in European Journal of Control ECC 2025 Special Issue, 2025
- Raineri R., Como G., Fagnani F., “Optimal Subset Selection in Opinion Dynamics on Social Networks”, in progress (**Chapter 3-4**)
- [25] Raineri R., Ye M., Zino L., "Controlling a Social Network of Individuals with Coevolving Actions and Opinions", under review, on arXiv

Conference Papers

- [57] Licciardi A., Raineri R., Proskurnikov A., Rondoni L., Zino L., "Sociodynamics-inspired Adaptive Coalition and Client Selection in Federated Learning", under review, on arXiv
- [48, 49] Raineri R., Como G., Fagnani F., “Optimal selection of the most informative nodes for a noisy DeGroot model with stubborn agents”, 23rd European Control Conference, 2025 (**Chapter 3-4**)
- [24] Raineri R., Como G., Fagnani F., Ye M., Zino L., “On Controlling a Coevolutionary Model of Actions and Opinions”. IEEE 63rd Conference on Decision and Control, 2024, pp. 4550-4555
- [47] Raineri R., Como G., Fagnani F., “Optimal selection of the most informative nodes in Opinion Dynamics on Networks”, IFAC-PapersOnLine 56 (2), 4192-4197, 2023 (**Chapter 4**)

Extended Abstract

- Raineri R., Zino L., Proskurnikov A., “Opinion Dynamics: Effects of Memory and Higher-Order Neighbours”, NetSci 2025, Maastricht

1.4 Organization of the Dissertation

The thesis is organized as follows.

In Chapter 2 we recall some preliminary theoretical elements and introduce the general framework. Specifically, in Section 2.1 and 2.2 we recall some basic notions on non-negative matrices and graph theory that will be essential in the subsequent formalization of the problem. In Section 2.3-2.4 we propose a general overview of the opinion dynamics models that will be used in the research problems of interest. More precisely, in Section 2.3 we focus on some instances of deterministic linear positive Opinion Dynamics models. We formally present the French-DeGroot model and its two main generalization of interest, i.e. the French-DeGroot model with stubborn agents and the Friedkin-Johnsen model. For all the models we present also the main convergence result and a proper formalization of the corresponding equilibrium vector. In Section 2.4 we generalize the French-DeGroot model adding randomness sources and thus moving the attention on stochastic linear positive Opinion Dynamics models. We formally analyse the case of random stubborn agents and of external random noise in the communication channel.

Part I consists of Chapter 3 and 4 and it is focused on the Subset Selection problem. In Chapter 3 the problem is first analysed from a general perspective without a specific application to network analysis. Here, indeed, in a general statistical setting, we consider an arbitary random vector of observable attributes and a predictor variable of interest to estimate. In Section 3.1 we first formalize the subset selection problem in the setting of interest. In Section 3.2 we find an explicit formulation for the objective function chosen, i.e. the variance reduction function, and we analyse the problem computational complexity. In Section 3.3 we present a Greedy Algorithm as a way to approximate the problem solution with a notably reduced computational cost. Finally, in Section 3.4. and 3.5 we focus on the specific performances of this approach under specific properties of the model parameters related to submodularity property of the objective function. In Chapter 4 the focus moves on the specific application context, i.e. opinion dynamics application, and the results of the previous chapter are here recasted according to the hypothesis on the models of interest. In Section 4.1 the French-DeGroot model is analysed, with two possible sources of randomness, i.e. random noise or random external input. In Section 4.4.1 an explicit formulation of the variance reduction is presented, leadying in Section 4.1.2 to a wider discussion about its analogies with known centrality mea-

sures. Finally, conditions which guarantee submodularity are analysed to provide some insights related to the specific application. At the end of the section some examples and simulations are provided in order to corroborate the previous results. Section 4.2 is centred on the Friedkin-Johnsen model with a specific focus on the interaction rate parameter. Specifically, it is proposed a set of optimal criteria for the two limit cases, respectively of high and low interactions. Finally, some examples and simulations are provided in order to corroborate the previous results.

Part II consists of Chapter 5 and it is focused on the second research problem, i.e. the Detection problem. First, in Section 5.1, a proper mathematical formalization of the problem is provided. Then, in Section 5.2 and 5.3 are solved respectively the problem for the noise-free scenario and for the more general noisy scenario. Specifically, we present the conditions on the model parameters under which the subset of stubborn nodes is correctly detected and the performance guarantees bounds on the influence matrix estimation. At the end of the chapter, in Section 5.4 some examples are provided in order to corroborate the results in case of both noisy observations and observations at finite time, far from the steady state. Finally, in section 5.4.3 through some empitical simulations some preliminary observations are presented for the partial observation case, when the full state matrix is not observable.

CHAPTER 2

Introduction to Opinion Dynamics

This chapter is organized into three main sections. First, we review fundamental mathematical concepts from non-negative matrix theory and graph theory that serve as the foundation for the analyses presented in this dissertation. Next, we provide an overview of the principal models in Opinion Dynamics, with a particular focus on the well-established French-DeGroot and Friedkin-Johnsen models, which will play a central role throughout this work. Finally, we introduce the core research problems that will be addressed in the subsequent chapters.

2.1 Notions on Non-Negative Matrices

Let us now recall some terminology that will be useful in the rest of the dissertation.

- A matrix $A \in \mathbb{R}^{\mathcal{V}\times\mathcal{V}}$ is said to be *non-negative* (respectively *positive*) if all of its entries A_{ij} are non-negative (respectively positive) and it is denoted respectively by $A \geq 0$, or $A > 0$. Similarly, for a vector x in $\mathbb{R}^{\mathcal{V}}$, $x \geq 0$ and $x > 0$ indicate component-wise inequalities, i.e. for all i in $\mathcal{V}$, $x_i \geq 0$ and $x_i > 0$.

- A matrix $A \in \mathbb{R}^{\mathcal{V}\times\mathcal{V}}$ is said to be *Schur stable* if its spectral radius is less then 1, where the spectral radius of a matrix A is the absolute value of its maximum eigenvalue and it is indicated with $\rho(A)$.
- A matrix $A \in \mathbb{R}^{\mathcal{V}\times\mathcal{V}}$ is said to be *positive definite* if, for all nonzero vectors x, $x'Ax$ is real and positive. We will use the notation $A \succ 0$ to indicate a positive definite matrix.
- A matrix $A \in \mathbb{R}^{\mathcal{V}\times\mathcal{V}}$ is said to be *irreducible* if there exists no permutation to make it similar to a block upper triangular matrix. Formally, it does not exist a permutation matrix P such that

$$P'AP = \begin{bmatrix} B & C \\ 0_{n-r,r} & D \end{bmatrix}$$

  with $n = |\mathcal{V}|$, $1 \leq r \leq n-1$ for a proper $B \in \mathbb{R}^{r\times r}, C \in \mathbb{R}^{r\times n-r}, D \in \mathbb{R}^{n-r\times n-r}$.
- A matrix $A \in \mathbb{R}^{\mathcal{V}\times\mathcal{V}}$ is referred to as a *stochastic* matrix if is non negative and it is such that $A\mathbb{1} = \mathbb{1}$, i.e. all its rows sum up to 1. A matrix is *sub-stochastic* if the sum of each row is not greater than 1, i.e. $A\mathbb{1} < \mathbb{1}$.
- A matrix $A \in \mathbb{R}^{\mathcal{V}\times\mathcal{V}}$ is a *Metzler* matrix if its extradiagonal entries are all non-negative, i.e. $A_{ij} \geq 0$ for any $i \neq j$ in $\mathcal{V}$.
- A stochastic matrix $A \in \mathbb{R}^{\mathcal{V}\times\mathcal{V}}$ is said to be *reversible* with respect to a non negative vector $y \in \mathbb{R}^{\mathcal{V}}$ if for all i, j in $\mathcal{V}$ it holds $y_i A_{ij} = y_j A_{ji}$.

Let us now recall here one of the main results on non-negative matrices [58].

Theorem 2.1 (Perron-Frobenius)**.** *Let A in $\mathbb{R}_+^{n\times n}$ be a non-negative square matrix. Then, the spectral radius $\rho(A)$ is an eigenvalue of A and there exist non-negative vectors $x \neq 0$ and $y \neq 0$ such that:*

(i) $Ax = \rho(A)x$, $A'y = \rho(A)y$;

(ii) *every eigenvalue λ of A is such that $|\lambda| \leq \rho(A)$;*

(iii) $\max\{w_{min}, w_{min}^-\} \leq \rho(A) \leq \min\{w_{max}, w_{max}^-\}$ *where*

$$w_{min} = \min_i \sum_j A_{ij}, \quad w_{min}^- = \min_j \sum_i A_{ij}$$

are the minimum row and column sum of A, while

$$w_{max} = \max_i \sum_j A_{ij}, \quad w^-_{max} = \max_j \sum_i A_{ij}$$

are the maximum row and column sum of A.

Moreover, if A is a stochastic irreducible matrix then

(iv) $\rho(A) = 1$ *is simple and there exist a unique invariant distribution* $\pi \in \mathbb{R}^n$ *s.t.* $A'\pi = \pi$ *and* $\pi > 0$*;*

(v) any eigenvalue $\lambda \neq 1$ *of A is such that* $|\lambda| < 1$.

2.2 Elements on Graph Theory

In this section, we introduce fundamental concepts of graph theory that will be essential for the rest of this dissertation. Graphs provide a powerful framework for representing complex systems, where a large number of agents (nodes) interact through interconnections (edges). These structures effectively capture the relationships and dynamics within such systems. Throughout this work, the terms "network" and "graph" will be used interchangeably whenever there is no risk of confusion.

A finite *directed weighted graph* $\mathcal{G}$ is defined by the triple $(\mathcal{V}, \mathcal{E}, W)$ where:

- $\mathcal{V} = \{1, \ldots, N\}$ is the finite set of $N \in \mathbb{N}$ *nodes*, usually labeled by positive integer numbers.
- $\mathcal{E} \subseteq \mathcal{V} \times \mathcal{V}$ refers to the set of ordered *edges* or *links*, i.e. the connections among the nodes. A link is represented using the ordered pair of nodes (i, j) with $i, j \in \mathcal{V}$. A link for which the tail node and the head node coincide is referred to as a self-loop. Two nodes i and j which are connected by a link are *neighbours*. The number of neighbours of a node represents its *degree*.
- $W \in \mathbb{R}_+^{\mathcal{V}\times\mathcal{V}}$ is the *weight matrix* associated to the graph which is such that $W_{ij} \neq 0$ if and only if $(i, j) \in \mathcal{E}$. In this thesis we will only consider graphs with positive edge weights, i.e. $W_{ij} > 0$. However, other works may consider negative edge weights to represent antagonistic behaviours among agents.

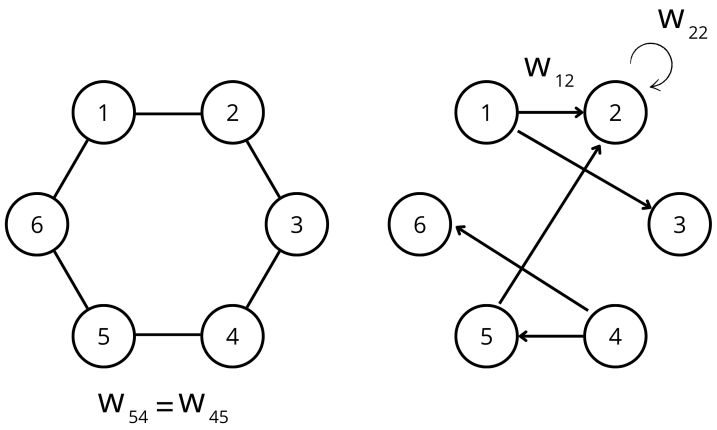


Fig. 2.1 An example of network with 6 nodes. On the left a simple undirected graph. On the right a directed graph. The edge which connects node 2 with itself is an example of self-loop.

Analogously, given any square matrix $W \in \mathbb{R}_+^{n\times n}$, we can define the *associated directed weighted graph* $\mathcal{G} = (\mathcal{V}, \mathcal{E}, W)$, such that $\mathcal{V} = \{1, \dots, n\}$ and $\mathcal{E} = \{(i, j) \in \mathcal{V} \times \mathcal{V} : W_{ij} \neq 0\}$.

An example of undirected and directed graph is reported in Fig.2.1.

A *subgraph* of $\mathcal{G} = (\mathcal{V}, \mathcal{E}, W)$ is any graph $\mathcal{H} = (\mathcal{U}, \mathcal{F}, Z)$ with node set $\mathcal{U} \subseteq \mathcal{V}$, link set $\mathcal{F} \subseteq \mathcal{E}$ and link weights $Z_{ij} \leq W_{ij}$ for every $i, j \in \mathcal{U}$. Given $\mathcal{U} \subseteq \mathcal{V}$, the subgraph of $\mathcal{G}$ *induced* by $\mathcal{U}$ is given by $\mathcal{G}|_{\mathcal{U}} = (\mathcal{U}, \mathcal{F}, W_{\mathcal{U}\times\mathcal{U}})$ where $\mathcal{F}$ is obtained removing from $\mathcal{E}$ all the links that involve at least one node in the complementary set $\mathcal{V}\backslash\mathcal{U}$.

Let us now recall some extra terminology that will be usefull in the following. A graph $\mathcal{G} = (\mathcal{V}, \mathcal{E}, W)$ is referred to as:

- *unweighted* if $W_{ij} \in \{0, 1\}$ for all nodes $i, j \in \mathcal{V}$. In this case W is also defined as the *adjacency matrix* of $\mathcal{G}$ and it can be unambiguously deduced from the link set $\mathcal{E}$.
- *undirected* if the weight matrix W is symmetric, i.e. if a link (i, j) exist if and only if it exists also the link in the reverse direction (j, i) and they have the same weight $W_{ij} = W_{ji}$.
- *simple* if it is undirected, unweighted, and the weight matrix W has zero diagonal, i.e. it has no self-loops.

- *regular* if each node has the same degree, where the degree indicates the number of neighbours.

Finally, let us define some extra elements related to graphs that will be essential in the rest of this dissertation.

Let us define the *normalized weight matrix* P associated to the graph $\mathcal{G}$ as

$$P := D^{-1}W$$

where $D = \text{diag}(w)$, with $w = \sum_{j\in\mathcal{V}} W_{ij}$.
Then, the non-negative vectors y such that

$$P'y = y$$

are referred to as *invariant distributions* of the graph $\mathcal{G}$. Moreover, invariant distributions $\pi = P'\pi$ satisfying $\mathbb{1}'\pi = 1$ are referred to as *invariant probability distributions* or briefly as *invariant probabilities* of the graph $\mathcal{G}$.

2.2.1 Reachability

Let us now introduce some notions about nodes reachability in a graph.

- A *walk* from a node i to a node j over $\mathcal{G}$ is a sequence of nodes $\gamma = (\gamma_0, \gamma_1, \ldots, \gamma_l)$ such that $\gamma_0 = i$, $\gamma_l = j$ and (γ_{h-1}, γ_h) is a link in $\mathcal{E}$ for all $h = 1, \ldots, l$. l is called the *length* of the walk.

  Given W the graph weight matrix, for each l-length walk we can define its *weight* as the product of its l link weights. Precisely,

  $$W_\gamma = \prod_{\leq h\leq l} W_{\gamma_{h-1}\gamma_h}.$$

  Trivially, a walk of length 1 from i to j coincides with the link between the two nodes, i.e. W_{ij}. In wider terms, it holds that the (i,j)-th entry of W^l coincides with the sum of weights of length-l walks from i to j, i.e. for any i, j in $\mathcal{V}$

  $$(W^l)_{ij} = \sum_{\gamma \text{ length-l walk from } i \text{ to } j} W_\gamma$$

- A *cycle* is a closed walk $\gamma = (\gamma_0, \gamma_1, \ldots, \gamma_l)$ starting and ending in the same node, i.e. $\gamma_0 = \gamma_l$. The length of a cycle is the number of edges in the cyclic path.

- A node j is said to be *reachable* from a node i if there exists a walk from i to j.

- A graph $\mathcal{G}$ is *strongly connected* if given any two nodes i and j, we have that i is reachable from j.

- Given a graph $\mathcal{G} = (\mathcal{V}, \mathcal{E}, W)$, the subset of nodes $\mathcal{U} \subseteq \mathcal{V}$ is said to be *globally reachable* if for every node i in $\mathcal{V} \backslash \mathcal{U}$ there exists a walk from i to some node j in $\mathcal{U}$. In particular, $\mathcal{G}$ is strongly connected if and only if every subset of nodes is globally reachable.

- Another important concept is that of *periodicity*. Given a graph $\mathcal{G} = (\mathcal{V}, \mathcal{E}, W)$, the *period* $\text{per}_{\mathcal{G}}(i)$ of a node i in $\mathcal{V}$ is defined as the greatest common divisor of the lengths of all walks starting and ending in i, with the convention that $\text{per}_{\mathcal{G}}(i) = 1$ if there are no circuits starting and ending in i.

  The common period of all nodes in a strongly connected graph $\mathcal{G}$ is denoted by $\text{per}_{\mathcal{G}}$ and referred to as the period of $\mathcal{G}$. A strongly connected graph $\mathcal{G}$ is called *aperiodic* if $\text{per}_{\mathcal{G}} = 1$.

2.2.2 Network Centrality Measures

In this section, we focus on those measures that capture the importance of a node in a graph $\mathcal{G} = (\mathcal{V}, \mathcal{E}, W)$. These are commonly referred to as *centrality measures* [59]. There is extensive literature on these topics. Here, we will limit our analysis to the measures that will be useful later in the dissertation.

- **Degree Centrality.** It is a measure of the popolarity of node I in the network, based on the number of neighbours it is linked to. In directed network we should distinguish between *in-degree* centrality w_i^- which counts the number of links pointing to node i, and *out-degree* centrality w_i^+ which counts the number of links pointing out from node i.

- **Eigenvector Centrality.** It assigns scores to nodes such that a node's score is proportional to the sum of the scores of its neighbors. This means nodes

connected to highly central nodes receive a higher score themselves, capturing both direct and indirect influence in the network. Formally, if $\mathcal{G}$ is a strongly connected graph, the vector of cebntralities $z \in \mathbb{R}^n$ is s.t.

$$z = \rho(W)^{-1} W' z .$$

- **Katz-Bonacich Centrality <[60].** It modifies the notion of eigenvector centrality by allowing nodes to get some centrality, independenlty of their in-neighbours. A parameter $\beta \in (0,1]$ and a nonnegative vector μ are chosen which can be thought as a sort of some intrinsic centrality. The standard choice is $\mu = \mathbb{1}$. Then, the Kats centrality vector is defined as

$$z^{(\beta)} = (I - \rho(W)^{-1}(1-\beta)W')^{-1}\beta\mu .$$

If the normalized weight matrix P is used then the solution of

$$z^{(\beta)} = (1-\beta)P' z^{(\beta)} + \beta\mu$$

is referred to as the Bonacich centrality.

- **Cycle Centrality [61].** It measures the importance of a node in a network based on its participation in cycles. Formally,

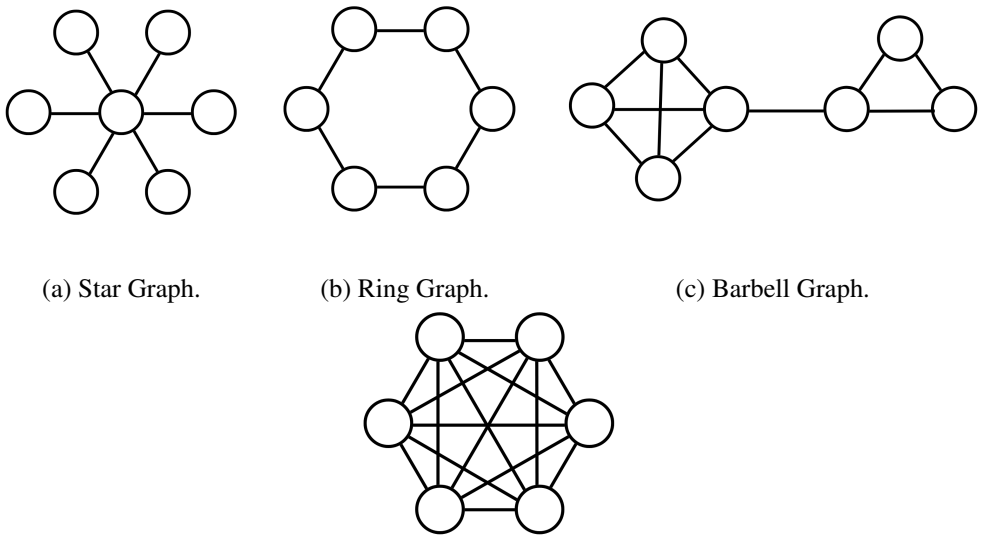
$$c_i = \sum_{\text{cycles } \gamma \ni i} f(\gamma)$$

where $f(\gamma)$ is a function that quantifies the significance of cycle γ.

- **Key Player Centrality or Intercentrality [62].** Given a network $\mathcal{G}$ with weight matrix W and a scalar α s.t. $M = (I - \alpha W)^{-1}$ is well-defined and nonnegative, the intercentrality measure is defined as

$$c_i = \frac{((M\mathbb{1})_i)^2}{M_{ii}} .$$

In the context of a game played over the consider metwork the players with highest intercentrality are the key players, whose removal results in the maximal decrease in overall network activity.



(a) Star Graph. (b) Ring Graph. (c) Barbell Graph.

(d) Complete Graph.

Fig. 2.2 Some relevant graph topologies presented in the examples of this section.

2.2.3 Standard Families of Graphs

Recall now some standard families of graph we will use in the rest of this dissertation.

Example 2.1 (Complete graph)**.** *A complete graph with N nodes is a graph where each node is connected to itself and to all the other nodes. Formally,*

$$(i,j) \in \mathcal{E} \qquad \forall i,j \in \mathcal{V} \text{ with } i \neq j\,.$$

Here, each node has maximum degree, i.e. $d_i = |\mathcal{V}|$*. See Fig. 2.2-(d) for an example.*

Example 2.2 (Line Graph)**.** *A line graph with N nodes is a graph where the edge set is defined as follows:*

$$\begin{cases} (i,i+1) \in \mathcal{E} & 1 \leq i < n \\ (i-1,i) \in \mathcal{E} & 1 < i \leq n \end{cases}$$

Example 2.3 (Ring Graph)**.** *A ring or cycle graph, denoted with* C_N*, is obtained from a line graph adding the edges* $(1,N)$ *and* $(N,1)$*. Formally,*

$$(i,i+1),(i-1,i) \in \mathcal{E} \qquad \forall i \in \mathcal{V}\,.$$

Here, each node has degree 2. An example is presented in Fig. 2.2-(b).

Example 2.4 (Star Graph)**.** *A star graph S_N is a graph with N nodes where, indicated with 1 the central node, the edge set is defined as*

$$\mathcal{E} = \{(1,i),(i,1),i = 2,\dots,N-1\}\,.$$

An example is presented in Fig. 2.2-(a).

Example 2.5 (Barbell Graph)**.** *A barbell graph is a graph obtained by connecting two independent complete graphs with a link between a node in each side. An example is presented in Fig. 2.2-(c).*

Example 2.6 (Erdos-Renyi Random Graph)**.** *It is the first random graph model introduced in 1959 [63]. $\mathcal{G}(N,p)$ is a random undirected simple graph woth N nodes where each couple of edges (i,j) and (j,i) is present with a probability $p \in (0,1)$, independently from the other couple of edges.*

Example 2.7 (Watts-Strogatz Random Graph)**.** *The Watts-Strogatz model is a generalization of the Erdos-Renyi model, introduced in 1998 [64]. It is a random graph generated according to the following rewiring process:*

- *Build a regular graph with degree d*
- *With probability p rewire each edge. This means that given edge (i,j), node i is instead connected to random node node k, where k is chosen unifromly at random. Notice that if $p = 1$ it coincides with an Erdos-Renyi graph.*

2.3 Deterministic Positive Linear Opinion Dynamics Models

Social networks can be effectively represented using graph theory, where individuals (or agents) are modeled as *nodes*, their interactions or relationships are depicted as *edges* connecting these nodes, and the strength of such interactions by the elements of a weight matrix. Thus, all models presented below have as a starting point a weighted graph $\mathcal{G} = (\mathcal{V}, \mathcal{E}, W)$. The corresponding normalized weight matrix P is known in this context as the *influence matrix*: each entry P_{ij} quantifies the strength of influence that agent j exerts on agent i. In the sequel, we assume that $\mathcal{V} = \{1, \dots, n\}$.

All the models considered in this thesis consist of positive linear[1] dynamical system through which the opinions of the various agents evolve over time and eventually reach an equilibrium value. Moreover, the agents are assumed to update their opinions synchronously[2]. Specifically, we consider a discrete-time dynamical system of type

$$x(t+1) = Ax(t) + u(t) \tag{2.1}$$

where A is a real non negative matrix of dimension $n \times n$. The vector $x(t) \in \mathbb{R}^n$ represents the vector of opinions of the various agents at time t, while $u(t)$ is a vector of dimension n that can model endogenous or exogenous inputs to the system (e.g. disturbances, opinion of external agents, primitive beliefs). The network structure of system (2.1) comes from the assumption that A is adapted to the graph $\mathcal{G}$ in the sense that for $i \neq j$, $A_{ij} = 0$ whenever $(i, j) \notin \mathcal{E}$, i.e. A and W share the same sparsity pattern. Typically, A is directly derived from the normalized weight matrix of the underlying graph $\mathcal{G}$. In the model described by (2.1), at every time instant t each agent i updates its opinion as

$$x_i(t+1) = \sum_{j \in \mathcal{V}} A_{ij} x_j(t) + u_i(t)$$

namely it takes a linear combination of its own opinion and of their out-neighbors and of some external input. Recursive relations (2.1) lead to the following explicit

[1] These hypothesis have been largely generalized in existing literature like considering signed graphs [65] or non linear models [12, 66]. Their analysis is out of the scope of this dissertation but of possible interest for future generalizations of the proposed results.

[2] Asynchronous models [21], widely studied in literature, are out of the scope of this dissertation.

form of the opinion vector at time t:

$$x(t) = A^t x(0) + \sum_{k=0}^{t-1} A^k u(t-k-1) \tag{2.2}$$

where $x(0) \in \mathbb{R}^n$ is the initial opinion vector.

In the following lemma (refer to [37] for details) we highlight the limit behaviour of the dynamics in case of constant exogenous vector u.

Lemma 2.1. *Consider the dynamical system in* (2.1) *with A Schur stable matrix and* $u(t) = u$ *for every* $t = 0, 1, \ldots$ *Then for every initial condition* $x(0)$ *it exists a steady state x such that*

$$x := \lim_{t \to +\infty} x(t) = (I - A)^{-1} u\,. \tag{2.3}$$

2.3.1 French-DeGroot Model

One of the first opinion formation models is the one proposed by the social psychologist French in [15] and then further generalized by DeGroot in [16]. In what follows, we present an overview of the main literature on the topic [67, 21, 68, 69], which provides the foundation for the subsequent analysis.

The French-DeGroot model describes a discrete time process of opinion formation in a set $\mathcal{V}$ of n agents interacting in a social network. As more in general introduced before, each agent i starts with an initial opinion on a given topic of interest, which we denote by $x_i(0) \in \mathbb{R}$, then agents exchange information about their beliefs with their neighbours.

A stochastic $n \times n$ matrix P collects the influence weights of the network, with $P_{ij} \geq 0$ expressing the influence of agent j on agent i. A greater weight P_{ij} shows a stronger influence of agent j on the opinion of agent i. As previously introduced P coincides with the normalized weight matrix of the network. At each time t, agent i opinion updates according to the following relation

$$x_i(t+1) = \sum_{j=1}^{n} P_{ij} x_j(t)\,. \tag{2.4}$$

Let us notice that this system is a special case of (2.1) with $A = P$ and $u(t) = 0$ for all t. Moreover, since we assume that $\sum_{j \in \mathcal{V}} P_{ij} = 1$, then $x_i(t+1)$ is a convex

combination, or weighted average, of his neighbours' opinions. This kind of models are indeed also known as **weighted averaging models**.

Let $x(t) = [x_1(t), \dots, x_n(t)]'$ be the vector of network opinions at time t, (2.4) can be rewritten as

$$x(t+1) = Px(t), \quad \forall\, t \geq 0\,. \tag{2.5}$$

A remarkable problem of interest is the study of the convergence of the model in (2.5) and the definition of the final opinion distribution. Let us recall here some standard notions and results [69, 70].

Proposition 2.1. *Consider a graph $\mathcal{G} = (\mathcal{V}, \mathcal{E}, W)$ and let P be the corresponding normalized weight matrix. If $\mathcal{G}$ admits a globally reachable and aperiodic vertex, then, for every initial condition $x(0)$ in $\mathbb{R}^n$, the system* (2.5) *converges to the steady state x such that*

$$x := \lim_{t \to +\infty} x(t) = \alpha \mathbb{1}, \quad \textit{with } \alpha = \nu' x(0),$$

where ν is the dominant left eigenvector of P with $\nu \mathbb{1} = 1$.

Formally, x is an equilibrium for the system (2.5) *since $x = Ax$, i.e. repeated application of the update map leaves the state exactly unchanged. Morever, since $x_1 = x_2 = \cdots = x_n = \alpha$, then x is called consensus.*

For a detailed proof of this result see Theorem 2.2 in [70].

Let us notice, as a particular case of interest, that strongly connected aperiodic graphs are a special case of the one described in the previous proposition.

Corollary 2.1. *If $\mathcal{G}$ is a strongly connected and aperiodic graph, then the system* (2.5) *is convergent and the equilibrium is given by*

$$x = \alpha \mathbb{1},$$

with $\alpha = \nu' x(0)$, where ν is the dominant left eigenvector of the stochastic influence matrix P, such that $\nu \mathbb{1} = 1$.

It is worth noting that the DeGroot model has been extensively studied in literature. Significant attention has been devoted to identifying conditions on the network topology that ensure convergence to consensus, even in more general settings such as

time-varying networks, where the topology evolves over time [22, 23]. Another key research direction, particularly relevant in the context of control strategies, concerns the rate of convergence to equilibrium. In this dissertation, we focus on a different, yet equally important, question: how to predict the consensus outcome based on partial observations of the initial opinions of a subset of nodes.

2.3.2 French-DeGroot Model with Stubborn Agents

In this subsection, we extend the classical French–DeGroot framework by introducing a subset of agents whose behavior is endogenously determined: we refer to such agents as **stubborn agents** as in our models they will be characterized by keeping their opinion constant over time, not influenced by the remaining agents. The behaviour of such agents could model the influence of mass media, newspapers or other informational sources on individuals' opinions.

The French-DeGroot model with stubborn agents is a model where the set of agents $\mathcal{V}$ is partitioned into two disjoint sets, i.e. $\mathcal{V} = \mathcal{R} \cup \mathcal{S}$. The nodes in $\mathcal{S}$, denoted *stubborn agents*, maintain their opinion fixed at all times, i.e.

$$x_i(t) = x_i(0) \quad \forall i \in \mathcal{S}, \forall t \geq 0\,.$$

Conversely, the nodes in $\mathcal{R}$, called *regular agents*, update their opinion as a weighted average of their neighbours, as in the classical French-DeGroot model, i.e.

$$x_i(t+1) = \sum_{j=1}^{n} P_{ij} x_j(t) \quad \forall i \in \mathcal{R}, \forall t \geq 0\,.$$

If we denote by $x(t)$ the vector of opinion of the regular agents only and with u the vector of initial opinion of stubborn agents, the evolution of the opinion of regular agents can be written as

$$x(t+1) = Ax(t) + Bu \tag{2.6}$$

where $A = P_{\mathcal{RR}}$ and $B = P_{\mathcal{RS}}$. From now on, we will stick to the case where all regular agents are influenced directly or indirectly by some stubborn agents. We notice that when this is not the case, it means that there is a subset of the regular nodes that form a subsystem not influenced by the stubborn nodes and thus evolving

according to a classical French-DeGroot dynamics previously analyzed. We have the following result.

Proposition 2.2. *Consider a graph $\mathcal{G} = (\mathcal{V}, \mathcal{E}, P)$ and a partition $\mathcal{R} \cup \mathcal{S} = \mathcal{V}$ such that the set of stubborn agents $\mathcal{S}$ is globally reachable in $\mathcal{G}$. Let P be the normalized weight matrix of $\mathcal{G}$ and put $A = P_{\mathcal{R}\mathcal{R}}$ and $B = P_{\mathcal{R}\mathcal{S}}$. Then, A is Schur stable and for every values $u \in \mathbb{R}^{\mathcal{S}}$ of the opinion of the stubborn agents, system* (2.6) *is convergent and the equilibrium is given by*

$$x = (I - A)^{-1} B u \,. \tag{2.7}$$

for every initial condition $x(0) \in \mathbb{R}^{\mathcal{R}}$ of the regular agents.

Proof. Let us first prove that, under the assumption of global reachability, the matrix $A = P_{\mathcal{R}\mathcal{R}}$ is Schur Stable. From [69] it is known that a substochastic matrix A either is Schur stable or has a stochastic submatrix $\tilde{A} = (A_{ij})_{i,j \in \tilde{\mathcal{V}}}$, where $\tilde{\mathcal{V}} \subseteq \mathcal{V}$. Assume by contradiction that A is not Schur Stable. Then, it exists a stochastic submatrix $\tilde{A}$ which implies the existence of at least one non-empty subset of regular agents that is not influenced, either directly or indirectly, by any stubborn agent. However, this contradicts the assumption of global reachability of the set $\mathcal{S}$. Therefore, the matrix A must be Schur stable. From Lemma 2.1 the convergence result follows. □

2.3.3 Friedkin-Johnsen Model

The Friedkin-Johnsen (FJ) model [30, 31] is an extensively studied model of opinion formation, naturally extending the French-DeGroot iterative averaging dynamics. In addition to the weighted influence graph, the FJ model assigns each agent a constant innate opinion, factored into each opinion update iteration. Originally defined as the agent's initial opinion [31], the innate opinion can also be shaped by the agent's prejudices or some external sources of information, such as social media. An agent's degree of "anchorage" to their innate opinion is regulated by an additional parameter, which determines their susceptibility to social influence [10]. Some studies interpret this parameter as a measure of conformity under group pressure [32]. Despite its linearity, the FJ model can lead to highly diverse distributions of final opinions, ranging from consensus to polarization, including bimodal and multimodal states [10].

Formally, the classical Friedkin-Johnsen model describes the evolution of opinions according to the following equation:

$$x_i(t+1) = \lambda_i \sum_{j \in \mathcal{V}} A_{ij} x_j(t) + (1-\lambda_i) u_i, \tag{2.8}$$

where $\lambda_i \in [0,1]$ and $A = (A_{ij})_{i,j \in \mathcal{V}}$ is a non-negative stochastic matrix. As mentioned earlier, the innate opinion u_i may correspond to the initial opinion $x_i(0)$, a certain predisposition or prejudice on the given topic, or to an exogenous factor.

At each iteration, an agent's updated opinion $x_i(t+1)$ is determined by two components: its innate opinion u_i and the weighted average of its own and others' opinions, i.e. $\sum_{j=1}^{n} A_{ij} x_j(t)$. Thus, matrix A collects the influence weights of network nodes. The parameter $\lambda_i \in [0,1]$ represents the agent's openness to assimilating others' opinions, or its susceptibility to social influence. An agent with $\lambda_i = 0$ is a fully stubborn node, fully anchored to its innate opinion u_i, while $\lambda_i = 1$ corresponds to the classical French-DeGroot iterative opinion averaging, as discussed in Section 2.3.1.

Remark 2.1. *Let us recall that the previously introduced French-DeGroot model is a particular case of the Friedkin-Johnsen model. Specifically, if $\lambda_i = 1$ for all $i \in \mathcal{V}$ then* (2.8) *coincides in particular with* (2.4)*. Moreover, if the nodes in $\mathcal{V}$ are splitted into two disjoint subsets $\mathcal{S}$ and $\mathcal{R}$ such that $\lambda_i = 0$ for all $i \in \mathcal{S}$ and $\lambda_i = 1$ for all $i \in \mathcal{R}$, then* (2.8) *coincides with the DeGroot with stubborn model in* (2.6)*.*

Let $\Lambda = [\lambda]$ be a diagonal matrix collecting the susceptibility parameters and u be the vector of innate opinions, then the dynamics can be rewritten in a more compact form as follows:

$$x(t+1) = \Lambda A x(t) + (I - \Lambda) u\,. \tag{2.9}$$

The limit behaviour of the opinions is described in the following results from [21, 69]. The statements are rephrased for clarity in the context of the thesis.

Proposition 2.3. *Consider a graph $\mathcal{G} = (\mathcal{V}, \mathcal{E}, A)$. Let A be the normalized weight matrix of $\mathcal{G}$. For every constant innate opinion u in $\mathbb{R}^n$, if it exists at least one agent $i \in \mathcal{V}$ such that $\lambda_i < 1$, then the dynamics in* (2.9) *is asymptotically stable and converges to*

$$x := \lim_{t \to +\infty} x(t) = (I - \Lambda A)^{-1} (I - \Lambda) u\,. \tag{2.10}$$

Proof. Given that A is row-stochastic by assumption, the imposed bounds on the susceptibility parameters λ_i ensure that the spectral radius $\rho(\Lambda A) < 1$. The Schur stability of matrix ΛA thus guarantees that the conditions of Lemma 2.1 are satisfied, and the convergence result follows directly. □

2.4 Stochastic Linear Opinion Dynamics Models

In this section, we generalize the French-De Groot model presented above adding two sources of randomness in the dynamical system (2.6). First, we will assume that the **stubborn agents**' opinion, introduced in Section 2.3.2, may be unknown. For instance, if we model politicians as stubborn agents within the social network, their true policy positions, often kept confidential during coalition negotiations, must be treated as unkown random variables rather then observable input. Additionally, in real-world networks, communication among agents is rarely faithful. The transmission of information may be affected by distortions, delays, or misinterpretations arising from factors such as misinformation, biased reporting, or limitations in communication channels. This phenomenon, referred to as **noise**, can significantly influence the evolution of opinions within the network. It may occur at each time instant and it can accommodate biases that result from both "hearing" and "sending" errors [71].

Let us preliminary specify the notation used in this section. As it is customary, all random variables will be denoted using capital letters, in particular $X(t)$ will be used for the evolving opinion of regular agents. Formally, we are interested the following linear system

$$X(t+1) = AX(t) + BU + V(t) \tag{2.11}$$

where U is a random vector describing the stubborn agents' opinion and $V(t)$ is a sequence of i.i.d. random variables which are assumed to be independent from random vector U and that represent the extra noise term.

To improve readability, we will first examine each source of randomness independently; we then conclude the section by deriving a unified result for the system in (2.11).

2.4.1 French-DeGroot Model with Stochastic Stubborn Agents

First, let us focus on the case in which the opinion of the stubborn agents form a random vector U with bounded second moments, with mean μ and covariance matrix C_U, and no extra noise term occurs, i.e. $V(t) = 0$ for all $t = 0, 1, \ldots$. In this case, starting from Proposition 2.2, we are able to state the following convergence result.

Corollary 2.2. *Consider a graph $\mathcal{G} = (\mathcal{V}, \mathcal{E}, P)$ and a partition $\mathcal{R} \cup \mathcal{S} = \mathcal{V}$ such that the set of stubborn agents $\mathcal{S}$ is globally reachable in $\mathcal{G}$. Let P be the normalized weight matrix of $\mathcal{G}$ and put $A = P_{\mathcal{R}\mathcal{R}}$ and $B = P_{\mathcal{R}\mathcal{S}}$. Consider system* (2.11) *with U random vector with bounded second moments, with mean μ and covariance matrix C_U, and $V(t) = 0$. Then, the regular agents' opinion converges almost surely to the equilibrium opinion vector X with mean*

$$\mathbb{E}[X] = (I - A)^{-1} B\mu,$$

and covariance matrix

$$C_X = \mathbb{E}[(X - \mathbb{E}[X])(X - \mathbb{E}[X])'] = (I - A)^{-1} B C_U B' (I - A')^{-1}.$$

Proof. Based on Proposition 2.2, we derive that $X(t)$ converges almost surely to a random vector X such that

$$X = (I - A)^{-1} BU.$$

Then, based on mean and covariance matrix definition, it holds

$$\begin{aligned} \mathbb{E}[X] &= (I - A)^{-1} B\mathbb{E}[U] = (I - A)^{-1} B\mu \\ C_X &= \mathbb{E}[(X - \mathbb{E}[X])(X - \mathbb{E}[X])'] = (I - A)^{-1} B C_U B' (I - A')^{-1}. \end{aligned}$$

□

2.4.2 Noisy French-DeGroot Model with Deterministic Stubborn Agents

Let us now focus on the scenario in which an extra random noise $V(t)$ is added at each time instant t, while the stubborn agents $\mathcal{S}$ are assumed to be deterministic, i.e. $U = u \in \mathbb{R}^{|\mathcal{S}|}$. In this case, to analyse the limit behaviour of the dynamics in

(2.11), we can no longer apply the convergence Lemma 2.1. However, under mild conditions on the i.i.d process $V(t)$, as stated in Theorem 1 in [21], we can obtain a convergence result in distribution[3]. A standard proof from literature on stochastic linear systems [72], is here presented for completeness.

Theorem 2.2. *Consider a graph $\mathcal{G} = (\mathcal{V}, \mathcal{E}, P)$ and a partition $\mathcal{R} \cup \mathcal{S} = \mathcal{V}$ such that the set of stubborn agents $\mathcal{S}$ is globally reachable in $\mathcal{G}$. Let P be the normalized weight matrix of $\mathcal{G}$ and put $A = P_{\mathcal{R}\mathcal{R}}$ and $B = P_{\mathcal{R}\mathcal{S}}$. Assume that $V(t)$ is a sequence of i.i.d. random variables with bounded second moments, with zero mean and covariance matrix C_V, while U is a deterministic vector $u \in \mathbb{R}^{|\mathcal{S}|}$. Then, the random process $X(t)$ recursively defined by system* (2.11) *converges in distribution, for every deterministic initial condition $x(0)$, to an equilibrium opinion vector X having mean*

$$\mathbb{E}[X] = (I - A)^{-1} B u$$

and covariance matrix C_X that is the unique solution of the discrete-time Lyapunov equation

$$C_X = A C_X A' + C_V \,. \tag{2.12}$$

Proof. Preliminary notice that from Proposition 2.2, under global reachability of stubborn nodes, A is a Schur stable matrix. Recursive relation (2.11) leads to the following explicit form of the opinion vector at time t:

$$X(t) = A^t x(0) + \sum_{k=0}^{t-1} A^k (Bu + V(t-k-1))$$

where $x(0) \in \mathbb{R}^n$ is the initial opinion vector, assumed deterministic.

Let us define a new process $\bar{X}(t)$ such that

$$\bar{X}(t) = A^t X(0) + \sum_{k=0}^{t-1} A^k (Bu + V(k))$$

which has the same distribution as $X(t)$ for every $t \geq 0$.

Notice that, from Schur stability of A and since $V(t)$ has bounded second order moment by assumption, $\bar{X}(t)$ converges deterministically to a random vector $\bar{X}$.

[3]Convergence in distribution. $X_n \xrightarrow{D} X$ if for every continuous bounded function $\psi : \mathbb{R}^n \to \mathbb{R}$ it holds $\mathbb{E}[\psi(X_n)] \to \mathbb{E}[\psi(X)]$ fo $n \to \infty$

Consequently, $X(t)$ converges in distribution to a random vector X which has the same distribution of $\bar{X}$.

Let us now compute explicitely the mean and the covariance matrix of the equilibrium vector X. First, compute the mean value, i.e.

$$\mathbb{E}[X(t)] = \mathbb{E}[AX(t-1) + Bu + V(t)] = A\mathbb{E}[X(t-1)] + Bu \tag{2.13}$$

where the last equality follows since $V(t)$ is zero mean by definition and from standard properties of expected value. Since the update equation in (2.13) takes the same structure as the dynamics analyzed in Lemma 2.1 and given that $\rho(A) < 1$, it follows that, in the limit as $t \to +\infty$, the system converges to

$$\lim_{t \to +\infty} \mathbb{E}[X(t)] = (I - A)^{-1} Bu\,. \tag{2.14}$$

Second, compute the covariance matrix. In the following we will avoid the term Bu since it is assumed to be deterministic. Using the covariance definition it holds

$$\begin{aligned} C_X(t) &= \mathbb{E}[(X(t) - \mathbb{E}[X(t)])(X(t) - \mathbb{E}[X(t)])'] \\ &= \mathbb{E}[(A(X(t-1) - \mathbb{E}[X(t-1)]) + V(t))(A(X(t-1) - \mathbb{E}[X(t-1)]) + V(t))'] \end{aligned}$$

From explicit computations for $C_X(t)$ it holds that

$$\begin{aligned} C_X(t) &= \mathbb{E}[A(X(t-1) - \mathbb{E}[X(t-1)])(X(t-1) - \mathbb{E}[X(t-1)])'A'] \\ &\quad + \mathbb{E}[V(t)V'(t)] \\ &\quad + 2\mathbb{E}[A(X(t-1) - \mathbb{E}[X(t-1)])V'(t)] \end{aligned}$$

Observing now that $\mathbb{E}[V(t)V'(t)] = C_V$ and given that the third term is equal to 0 since $V(t)$ is assumed to be independent from $X(t)$, it holds

$$C_X(t) = AC_X(t-1)A' + C_V\,. \tag{2.15}$$

From standard results on stochastic linear systems [72], Schur stability of A guarantees that the system in (2.15) is asymptotically stable and it exists a unique solution C_X for the discrete-time Lyapunov equation

$$C_X = AC_XA' + C_V\,. \tag{2.16}$$

The thesis follows. □

Under some extra assumptions on the weight matrix A, we can find a more explicit formulation for the covariance matrix C_X of the limit opinion X.

Let us preliminary recall some definitions that will be useful in the following.

Definition 2.1. *A matrix A is diagonally similar to a symmetric matrix (DSS) if it exists a nonsingular diagonal matrix D such that $D^{-1}AD$ is a symmetric matrix.*

We notice that a huge class of networks satisfies this condition as specified in the following lemma [70].

Lemma 2.2. *If the weight matrix A is reversible, irreducible and aperiodic, then A is a DSS matrix with $D = diag(\pi)^{1/2}$ and π invariant distribution, i.e. π corresponding to the left dominant eigenvector of A'. Precisely, a sufficient condition is to require that A is reversible and Schur-stable, i.e. $\rho(A) < 1$.*

Based on this preliminary lemma, the following result can be proven.

Proposition 2.4. *Consider a graph $\mathcal{G} = (\mathcal{V}, \mathcal{E}, P)$ and a partition $\mathcal{R} \cup \mathcal{S} = \mathcal{V}$ such that the set of stubborn agents $\mathcal{S}$ is globally reachable in $\mathcal{G}$. Let P be the normalized weight matrix of $\mathcal{G}$ and put $A = P_{\mathcal{R}\mathcal{R}}$ and $B = P_{\mathcal{R}\mathcal{S}}$. Assume that $U = u$ is a deterministic vector and $V(t)$ is a sequence of i.i.d. random variables with bounded second moments, with zero mean and invertible covariance matrix C_V. Then, if A is reversible, given $X(t)$ the random process recursively defined by system* (2.11)*,*

(i) $X(t)$ converges in the limit to a random vector X with

$$\mathbb{E}[X] = (I-A)^{-1}Bu, \quad C_X = \mathbb{E}[(X-\mathbb{E}[X])(X-\mathbb{E}[X])'] = C_V(I-A^2)^{-1}$$

(ii) the covariance matrix C_X is positive definite, i.e. $C_X \succ 0$

(iii) the precision matrix $H := C_X^{-1}$ is symmetric positive definite and such that $-H$ is a Metzler matrix, i.e. $H_{ij} \leq 0$ for any $i \neq j$.

Proof. (i) It follows from Theorem 2.2. To compute the covariance matrix C_X recall that, from Lemma 2.2, since A is a DSS matrix, then $A = D^{-1}SD$. Substituting the

factorization of A matrix in (2.12), we obtain

$$C_X = DSD^{-1}C_XD^{-1}SD + C_V\,.$$

Properly collecting terms, we get

$$\begin{gathered} D(D^{-1}C_XD^{-1} - SD^{-1}C_XD^{-1}S)D = C_V \\ (I-S^2)D^{-1}C_XD^{-1} = D^{-1}C_VD^{-1} \end{gathered}$$

and consequently, by some extra computations,

$$C_X = C_VD^{-1}(I-S^2)^{-1}D\,.$$

Substituting $S{=}D^{-1}AD$ and $I{=}D^{-1}ID$ in the above, then

$$\begin{aligned} C_X &= C_VD^{-1}(D^{-1}ID - D^{-1}A^2D)^{-1}D \\ &= C_VD^{-1}(D^{-1}(I-A^2)D)^{-1}D\,. \end{aligned}$$

The thesis follows.
(ii) Given the covariance matrix definition in (2.12), we retrieve that, for any $x \in \mathbb{R}^{|\mathcal{R}|}\backslash\{0\}$ it holds

$$x'C_Xx = (A'x)'C_X(A'x) + x'C_Vx\,.$$

Here, $(A'x)'C_X(A'x) \geq 0$, since $A \geq 0$ by model assumptions on A and $C_X \geq 0$ as, using Neumann series expansion for covariance matrix definition in (i), it is the limit of a series of non negative terms. In addition, $x'C_Vx > 0$ since C_V it the covariance matrix of $V(t)$ whose variables $V_i(t)$ are independent by definition (see Chap. 7 in [73] for reference). Consequently, $x'C_Xx > 0$ for any $x \in \mathbb{R}^{|\mathcal{R}|}\backslash\{0\}$ and thus, $C_X \succ 0$ by definition.
(iii) As the precision matrix is defined as the inverse of the covariance matrix, it holds

$$H = C_X^{-1} = (I-A^2)C_V^{-1}\,, \tag{2.17}$$

where all extradiagonal entries of H are non-positive since $A \geq 0$. □

2.4.3 Noisy French-DeGroot Model with Stochastic Stubborn Agents

Finally, we can gather the two scenarios above and consider the case where randomness comes from both the two sources, i.e. the opinion of the stubborn agents and the noise in the opinion formation process. Builing on Corollary 2.2 and Theorem 2.2 and using linearity properties of the system, we have the following result.

Corollary 2.3. *Consider a graph $\mathcal{G} = (\mathcal{V}, \mathcal{E}, P)$ and a partition $\mathcal{R} \cup \mathcal{S} = \mathcal{V}$ such that the set of stubborn agents $\mathcal{S}$ is globally reachable in $\mathcal{G}$. Let P be the normalized weight matrix of $\mathcal{G}$ and put $A = P_{\mathcal{R}\mathcal{R}}$ and $B = P_{\mathcal{R}\mathcal{S}}$. Let U be a random vector with mean μ and covariance matrix C_U. Assume that $V(t)$ is a sequence of i.i.d. random variables with bounded second moments, with zero mean and covariance matrix C_V and that $V_i(t)$ are independent from U. Then, the random process $X(t)$ recursively defined by system* (2.11) *converges in distribution, for every deterministic initial condition $x(0)$, to the equilibrium opinion vector X with mean*

$$\mathbb{E}[X] = (I - A)^{-1} B\mu \tag{2.18}$$

and covariance matrix

$$C_X = \mathbb{E}[(X - \mathbb{E}[X])(X - \mathbb{E}[X])'] = C_1 + C_2, \tag{2.19}$$

where

$$C_1 = (I - A)^{-1} B C_U B' (I - A')^{-1}$$

and C_2 unique solution of discrete time Lyapunov equation

$$C_2 = A C_2 A' + C_V \,.$$

Proof. In order to prove this Corollary we may define $X(t)$ as the combination of two independent processes $X(t) = X_1(t) + X_2(t)$ where

$$X_1(t+1) = AX_1(t) + BU, \qquad X_2(t+1) = AX_2(t) + V(t)\,.$$

The corollary follows by combining the results of Corollary 2.2 and Theorem 2.2. The convergence of $X(t)$ to the limiting vector X holds only in distribution, as the convergence established in Theorem 2.2 is weaker than the amost sure convergence

guaranteed in Corollary 2.2. This combination is valid since, by construction, the two processes are independent. □

2.5 Research Objectives

Opinion dynamics is a long-standing research area that has attracted interest from the sociology, computer science, and control engineering communities. This dissertation focuses on two complementary research problems within this domain. As highlighted in Sections 2.3-2.4, the models considered can be viewed as specific instances of linear dynamical systems. This broader perspective enables a generalization of the results, offering potential applications in a wider range of contexts.

Let us now formally present our two core research questions, preliminary introduced in Section 1.2.

(I) **Subset Selection Problem.**
The problem consists in selecting a subset of $k < n$ variables $\{X_i\}_{i\in\mathcal{V}}$ from a larger set of n variables, with the objective of best inferring a target variable Y, defined as a linear combination of the $\{X_i\}_{i\in\mathcal{V}}$. Given prior knowledge of the covariances among the variables $\{X_i\}_{i\in\mathcal{V}}$ and Y, the aim is to identify the subset that most effectively predicts Y.

In the context of opinion dynamics, the variables $\{X_i\}_{i\in\mathcal{V}}$ represent the opinions of the network nodes, while the covariances, as will be extensively discussed in the following chapters, are intrinsically linked to the network's influence matrix.

A natural formulation of the optimization problem is to maximize the variance reduction on Y, that is, the difference between the variance of Y and the expected conditional variance of Y given the selected subset of variables. Thus, the optimization problem of interest, as it will be largely analysed in Chapters 3-4, can be expressed as

$$\arg \max_{\mathcal{K}\subseteq\mathcal{V}:|\mathcal{K}|=k} \mathrm{Var}(Y) - \mathbb{E}[\mathrm{Var}(Y|X_{\mathcal{K}})]\,.$$

(II) **Stubborn Behaviour Detection.**
Here we assume that the network structure is unknown, while the opinions of

the network nodes at equilibrium are available across a set of different topics. We model the underlying opinion dynamics through a French-DeGroot system with stubborn agents. The objective is to identify the set of stubborn agents $\mathcal{S}$ and estimate the network's influence matrix Γ.

From the opinion dynamics theory discussed in the previous section, it follows that the equilibrium opinions of regular agents can be expressed as a linear combination of the opinions of the stubborn agents, namely $X_{\mathcal{R}} = \Gamma X_{\mathcal{S}}$.

Building on matrix decomposition techniques, the problem can thus be formalized as the search for the minimal subset of agents whose opinions form a basis capable of generating the entire network opinion matrix. Detailed explanation will follow in Chapter 5.

PART I

Subset Selection Problem

CHAPTER 3

Subset Selection Problem - A General Formulation

In this Chapter, as well as in Chapter 4, we focus on the first problem of interest: the optimal subset selection problem. The preliminary results of this work were published in [47], while a subsequent analysis of a special case, related to a DeGroot opinion dynamics model with stubborn agents and noise, was presented in [48] and in its extension [49]. Although our primary focus is on applications in Opinion Dynamics, in this Chapter we introduce the problem in a more general framework. This broader perspective highlights the versatility of our approach, demonstrating that the results obtained are applicable to a wide range of domains and thus reinforcing the significance of our analysis. Notable examples include feature selection in machine learning, sparse approximation and compressed sensing in signal processing, sensor placement for environmental monitoring, risk assessment in medical studies, and active surveying in social sciences [36]. Most recent applications deal with the smart testing problem aimed to find an efficient strategy to control epidemic spread [42].

Main Contributions. Referring to the state-of-the-art, the analysis of the subset selection problem has attracted significant attention from both the statistics and computer science communities, leading to extensive research in the field [34, 35, 74, 75]. Building on existing results, this chapter's main contributions include first an

explicit formulation of the objective function for the optimization problem, without imposing additional assumptions on the probability distribution of the variable to be estimated, often assumed to be Gaussian in previous works. Second, the development of a greedy algorithm aimed at significantly reducing computational complexity of the proposed optimization problem, which is known to be NP-Hard in general. This ensures the scalability of the approach, making it feasible for large-scale systems. Finally, by extending the results of [34, 35] and appropriately adapting them to our formulation, we provide a detailed analysis of the performance guarantees associated with the greedy algorithm. Particular emphasis is placed on the model assumptions necessary to ensure the submodularity of the objective function, a key property for deriving approximation bounds.

3.1 Problem Formulation

Consider the problem of estimating a predictor variable of interest Y from a large set of observable attributes or parameters X_i. Sample all variables X_i each time a prediction of Y is needed is impractical. Therefore, the objective is to find the optimal subset of s variables X_i to best predict Y.

First, let us formally describe the problem.
All the observable variables, for simplicity, are collected in a random vector X with values in $\mathbb{R}^n$ with zero mean and bounded second moment, i.e. $\mathbb{E}[X_i^2] < \infty$ for every i in $\{1,\dots,n\}$. We denote with

$$C = \mathbb{E}[XX']$$

its covariance matrix, that we assume to be invertible, with inverse $H = C^{-1}$ called *precision matrix* .
Let Y be a predictor variable of interest. In many practical applications, such as disease risk prediction (e.g., heart disease) based on observable quantities (e.g., blood pressure, cholesterol levels, etc.), the prediction of Y is often performed using linear regression. Specifically, Y is modeled as a linear combination of the observed variables:

$$Y = \sum_{i=1}^{n} \alpha_i X_i$$

where the coefficients $\alpha_i \in \mathbb{R}$ are appropriately estimated to best fit the data.

Suppose we are allowed to observe only a subset $\mathcal{K} \subseteq \mathcal{V} = \{1, \ldots, n\}$ of cardinality s of the entries of X to estimate Y. To determine the optimal subset $\mathcal{K}$, it is essential to first establish an appropriate evaluation metric that quantifies the quality of the estimate. This metric will serve as the objective criterion for selecting the subset that best predicts Y. In particular, we focus on two closely related metrics, widely used in statistical modeling and machine learning for assessing subset quality in predicting a target variable Y:

- **Mean Squared Error** (MSE): It evaluates the expected squared difference between the predicted and actual values, i.e.,

  $$\mathbb{E}[(Y - g(X_{\mathcal{K}}))^2]$$

  where $g(X_{\mathcal{K}})$ is an estimator of Y using the observed set $\mathcal{K}$. It is well known (see Section 1.5.1 in [76]) that the best estimator $g(X_{\mathcal{K}})$ of Y that minimizes the MSE, i.e.

  $$g^*(\mathcal{K}) = \arg\min_{g(X_{\mathcal{K}})} \mathbb{E}[(Y - g(X_{\mathcal{K}}))^2],$$

  is

  $$g^*(X_{\mathcal{K}}) = \mathbb{E}[Y|X_{\mathcal{K}}]\,.$$

  Thus, using this result, we can find an explicit formulation for the mean squared error as function of $\mathcal{K}$ observation, i.e.

  $$\begin{aligned} G(\mathcal{K}) &= \mathbb{E}[(Y - g^*(X_{\mathcal{K}}))^2] \\ &= \mathbb{E}[\mathbb{E}[(Y - \mathbb{E}[Y|X_{\mathcal{K}}])^2|X_{\mathcal{K}}]] \\ &= \mathbb{E}[\mathrm{Var}(Y|X_{\mathcal{K}})] \end{aligned} \tag{3.1}$$

  Statistically, we retrieve that the mean squared error coincides with the residual variance on Y given observation $X_{\mathcal{K}}$.

- **Variance Reduction**: It measures the reduction in uncertainty about Y when conditioning on the selected subset $\mathcal{K}$, i.e.,

  $$F(\mathcal{K}) = \mathrm{Var}(Y) - \mathbb{E}[\mathrm{Var}(Y|X_{\mathcal{K}})].$$

Analogously, using $G(\mathcal{K})$ definition introduced in (3.1), we can define it as

$$F(\mathcal{K}) = \mathrm{Var}(Y) - G(\mathcal{K}). \tag{3.2}$$

The preferred choice between them depends on the specific application context. While MSE is more intuitive and directly measures prediction accuracy, variance reduction provides a more structured approach for deriving theoretical guarantees and is often more convenient for optimization in certain settings.

We now aim to formulate a formal definition of our optimization problem. The objective is to select the subset $\mathcal{K}$ that provides the best estimation of Y. However, in practical scenarios, not all variables in $\mathcal{V}$ may be observable due to external constraints, such as budget limitations, time restrictions, or environmental factors. To account for these limitations, let $\mathcal{O} \subseteq 2^{\mathcal{V}}$ denote the set of feasible observable subsets, constrained by the problem's external conditions.

Formally, our problem can be formalized both a MSE minimization or as a Variance Reduction maximization. The two optimization problems lead to the same result but through different explicit characterizations of the objective functions. In the following we will specifically focus on the maximization one, i.e.

$$\widehat{\mathcal{K}} = \underset{\mathcal{K} \in \mathcal{O}}{\arg\max}\, F(\mathcal{K}). \tag{3.3}$$

This choice is motivated by the insight it provides into various applications, as we will discuss in detail in the rest of this dissertation.

Applicability of the Proposed Method

The proposed problem addresses a wide range of real-world scenarios and the assumption that Y is a linear combination of the variables X_i does not appear to be restrictive. For instance, in a medical context, one may be interested in predicting the risk of a specific disease. A notable example is the CHA_2DS_2-VASc score, which is used to estimate the risk of stroke in people with non-rheumatic atrial fibrillation. Similar to our case, this score, associated with the predictor Y in our model, is a linear combination of various risk factors, represented by the X_i variables. The importance of the proposed algorithm lies in its ability to estimate the score by identifying a subset of relevant conditions, thereby reducing the number of necessary analyses. Moving to a different field of application, the Common Weakness Scoring System [77] evaluates software vulnerabilities by assigning a score based on a linear combination of technical factors. This similarity allows us to apply our method in the domain of cybersecurity, facilitating the assessment of software weaknesses through a more focused and efficient analysis.

3.2 Explicit Formulation of Variance Reduction

In this section, we propose an explicit formulation for the variance reduction function $F(\mathcal{K})$, without further specific assumptions on the context of application. This is a key step since it will help us find out useful intepretations of the optimization problem and discuss about its properties. Moreover, a similar analysis is performed here for the mean squared error $G(\mathcal{K})$ since having its explicit formulation will be helpful later on in deriving additional results.

Firstly, let us introduce a preliminary lemma that will further be useful in the following.

Lemma 3.1. *Consider a random vector X with zero mean and invertible covariance C, and let $Y = \alpha' X$, with $\alpha \in \mathbb{R}^n$. Given $\mathcal{K} \subseteq \mathcal{V} = \{1,\dots,n\}$, it holds*

$$\mathbb{E}[Y|X_{\mathcal{K}}] = \sum_{i\in\mathcal{K}} \hat{\alpha}_i^{\mathcal{K}} X_i \tag{3.4}$$

where $\hat{\alpha}^{\mathcal{K}}$ in $\mathbb{R}^{\mathcal{V}}$ is the vector of optimal regression coefficients for the set of predictor variables $X_{\mathcal{K}}$ and it is defined such that

$$\hat{\alpha}^{\mathcal{K}}_{\mathcal{K}} = (C_{\mathcal{K}\mathcal{K}})^{-1}(C\alpha)_{\mathcal{K}} \quad ; \quad \hat{\alpha}^{\mathcal{K}}_{-\mathcal{K}} = 0\,.$$

Proof. From Result 7.7 in [78] it is well known that, given $Y = \alpha' X$, then $\mathbb{E}[Y|X_{\mathcal{K}}] = (\hat{\alpha}^{\mathcal{K}})' X$ with

$$\hat{\alpha}^{\mathcal{K}} = \underset{\beta:\mathrm{supp}(\beta)\subseteq\mathcal{K}}{\arg\min}\ \mathbb{E}[(Y - \sum_{i\in\mathcal{K}} \beta_i X_i)^2]\,. \tag{3.5}$$

Using (3.5), the following series of equations holds true:

$$\begin{aligned}
\hat{\alpha}^{\mathcal{K}} &= \underset{\beta:\mathrm{supp}(\beta)\subseteq\mathcal{K}}{\arg\min}\ \mathbb{E}[(Y-\beta' X)^2] \\
&= \underset{\beta:\mathrm{supp}(\beta)\subseteq\mathcal{K}}{\arg\min}\ \mathbb{E}[((\alpha'-\beta')X)^2] \\
&= \underset{\beta:\mathrm{supp}(\beta)\subseteq\mathcal{K}}{\arg\min}\ \mathbb{E}[(\alpha-\beta)' X X' (\alpha-\beta)] \\
&= \underset{\beta:\mathrm{supp}(\beta)\subseteq\mathcal{K}}{\arg\min}\ (\alpha-\beta)' C (\alpha-\beta) \\
&= \underset{\beta:\mathrm{supp}(\beta)\subseteq\mathcal{K}}{\arg\min}\ -2\beta' C\alpha + \beta' C\beta
\end{aligned}$$

The minimum can then be found imposing that for any $k \in \mathcal{K}$

$$\frac{\partial\left(-2(\hat{\alpha}^{\mathcal{K}})' C\alpha + (\hat{\alpha}^{\mathcal{K}})' C\hat{\alpha}^{\mathcal{K}}\right)}{\partial \hat{\alpha}^{\mathcal{K}}_k} = 0$$

which is equivalent to

$$(C\alpha)_{\mathcal{K}} - (C\hat{\alpha}^{\mathcal{K}})_{\mathcal{K}} = 0 \tag{3.6}$$

and thus

$$C_{\mathcal{K}\mathcal{K}}(\alpha_{\mathcal{K}} - \hat{\alpha}^{\mathcal{K}}_{\mathcal{K}}) + C_{\mathcal{K}-\mathcal{K}}\alpha_{-\mathcal{K}} = 0\,.$$

Given that by assumption $C \succ 0$, then its submatrix $C_{\mathcal{K}\mathcal{K}}$ is invertible and the thesis follows.

□

Theorem 3.1. *Consider a random vector X with zero mean and invertible covariance matrix C, and let $Y = \alpha' X$, with $\alpha \in \mathbb{R}^n$. Given $\mathcal{K} \subseteq \mathcal{V} = \{1,\ldots,n\}$,*

(i) the variance reduction is computed as

$$F(\mathcal{K}) = (C\alpha)'_{\mathcal{K}}(C_{\mathcal{K}\mathcal{K}})^{-1}(C\alpha)_{\mathcal{K}}\,. \tag{3.7}$$

(ii) the mean squared error is computed as

$$G(\mathcal{K}) = \alpha'_{-\mathcal{K}}(H_{-\mathcal{K}-\mathcal{K}})^{-1}\alpha_{-\mathcal{K}}\,. \tag{3.8}$$

Proof. (i) Preliminary recall that the law of total variances states that

$$\mathrm{Var}(Y) = \mathrm{Var}(\mathbb{E}[Y|X_{\mathcal{K}}]) + \mathbb{E}[\mathrm{Var}(Y|X_{\mathcal{K}})]\,.$$

We can now prove that the following series of equalities holds:

$$\begin{aligned} F(\mathcal{K}) &= \mathrm{Var}(Y) - \mathbb{E}[\mathrm{Var}(Y|X_{\mathcal{K}})] = \\ &= \mathrm{Var}(\mathbb{E}[Y|X_{\mathcal{K}}]) = \\ &= (\hat{\alpha}^{\mathcal{K}})'C\hat{\alpha}^{\mathcal{K}} = \\ &= (\hat{\alpha}^{\mathcal{K}}_{\mathcal{K}})'(C\hat{\alpha}^{\mathcal{K}})_{\mathcal{K}}\,. \end{aligned}$$

More specifically, the first equality comes from the law of total variances, the second from Lemma 3.1 and the last one by $\hat{\alpha}^{\mathcal{K}}_{-\mathcal{K}} = 0$. Referring now to the relation in (3.6), we know that

$$(C\hat{\alpha}^{\mathcal{K}})_{\mathcal{K}} = (C\alpha)_{\mathcal{K}}\,.$$

Combining this with the definition of $\hat{\alpha}^{\mathcal{K}}$ in Lemma 3.1, the last equation of the previous series of equalities takes the following form

$$(\hat{\alpha}^{\mathcal{K}}_{\mathcal{K}})'(C\hat{\alpha}^{\mathcal{K}})_{\mathcal{K}} = (\hat{\alpha}^{\mathcal{K}}_{\mathcal{K}})'(C\alpha)_{\mathcal{K}} = (C\alpha)'_{\mathcal{K}}(C_{\mathcal{K}\mathcal{K}})^{-1}(C\alpha)_{\mathcal{K}}\,.$$

The thesis follows, i.e.

$$F(\mathcal{K}) = (C\alpha)'_{\mathcal{K}}(C_{\mathcal{K}\mathcal{K}})^{-1}(C\alpha)_{\mathcal{K}}\,.$$

(ii) Using $F(\mathcal{K})$ definition in (3.7), we know that

$$G(\mathcal{K}) = \mathrm{Var}(Y) - F(\mathcal{K})\,.$$

Let us now write an extended expression for $F(\mathcal{K})$ using the result in (i):

$$\begin{aligned} F(\mathcal{K}) = \; & \alpha'_{\mathcal{K}} C_{\mathcal{K}\mathcal{K}} \alpha_{\mathcal{K}} + \alpha'_{\mathcal{K}} C_{\mathcal{K}-\mathcal{K}} \alpha_{-\mathcal{K}} + \\ & \alpha'_{-\mathcal{K}} C_{-\mathcal{K}\mathcal{K}} \alpha_{\mathcal{K}} + \alpha'_{-\mathcal{K}} C_{-\mathcal{K}\mathcal{K}} (C_{\mathcal{K}\mathcal{K}})^{-1} C_{\mathcal{K}-\mathcal{K}} \alpha_{-\mathcal{K}} \,. \end{aligned}$$

Analogously, let us write an extended expression for $\mathrm{V}ar(Y)$:

$$\begin{aligned} \mathrm{Var}(Y) \; & = \alpha' C \alpha = \\ & = \alpha'_{\mathcal{K}} C_{\mathcal{K}\mathcal{K}} \alpha_{\mathcal{K}} + \alpha'_{-\mathcal{K}} C_{-\mathcal{K}\mathcal{K}} \alpha_{\mathcal{K}} + \alpha'_{\mathcal{K}} C_{\mathcal{K}-\mathcal{K}} \alpha_{-\mathcal{K}} + \alpha'_{-\mathcal{K}} C_{-\mathcal{K}-\mathcal{K}} \alpha_{-\mathcal{K}} \,. \end{aligned}$$

Thus, from some basic computations,

$$\begin{aligned} G(\mathcal{K}) \; & = \; \mathrm{Var}(Y) - F(\mathcal{K}) = \\ & = \; \alpha'_{-\mathcal{K}} (C_{-\mathcal{K}-\mathcal{K}} - C_{-\mathcal{K}\mathcal{K}} (C_{\mathcal{K}\mathcal{K}})^{-1} C_{\mathcal{K}-\mathcal{K}}) \alpha_{-\mathcal{K}} \,. \end{aligned}$$

Based on theory about inverse of a partitioned nonsingular matrix (see Appendix A.1 for reference), it is well known that, given $H = C^{-1}$,

$$(H_{-\mathcal{K}-\mathcal{K}})^{-1} = C_{-\mathcal{K}-\mathcal{K}} - C_{-\mathcal{K}\mathcal{K}} (C_{\mathcal{K}\mathcal{K}})^{-1} C_{\mathcal{K}-\mathcal{K}} \,.$$

The thesis follows. □

As a direct consequence of Theorem 3.1, when $|\mathcal{K}| = 1$ we can introduce a simplified expression for variance reduction.

Corollary 3.1. *Consider a random vector X with zero mean and invertible covariance C, and let $Y = \alpha' X$, with $\alpha \in \mathbb{R}^n$. If $|\mathcal{K}| = 1$, i.e. for $\mathcal{K} = \{k\}$, variance reduction $F(\{k\})$ is computed as*

$$F(\{k\}) = \frac{(C\alpha)_k^2}{C_{kk}} \,. \tag{3.9}$$

Example 3.1. *A notable example of application of Corollary 3.1 is the case in which we are interested in the estimate of the state of a particular variable i, formally, we want to predict $Y = \alpha' X$ with $\alpha = \delta_i$. Assume we can make one single observation but we are not allowed to directly observe X_i, i.e., the observable set $\mathcal{O}$ is defined as follows:*

$$\mathcal{O} = \{\{j\} : j \in \mathcal{V} \backslash \{i\}\} \,.$$

Under this specific setting, Eq. (3.9) *can be rewritten using* (3.10) *as*

$$F(\{k\}) = Cor^2(X_i, X_k) Var(X_i)\,.$$

where $Cor(X_i, X_k)$ *is the Correlation Coefficient between variables i and k, i.e. from 3.18 in [79], given* i, k *in* $\mathcal{V}$*,*

$$Cor(X_i, X_k) = \frac{Cov(X_k, X_i)}{\sqrt{Var(X_k)}\sqrt{Var(X_i)}}\,. \tag{3.10}$$

3.2.1 Computational Complexity Analysis

An important challenge related to the problem of interest is its computational complexity, since the Subset Selection problem is known to be computationally expensive. More precisely, let us notice that

- The evaluation of the objective function $F(\mathcal{K})$, given subset $\mathcal{K}$, requires

  (i) $O(s^3)$ operations due to matrix inversion, which can be reduced to $O(s^{2.376})$ using fast matrix multiplication methods, like Coppersmith-Winograd algorithm;

  (ii) $O(s^2)$ operations due to matrix-vector multiplication.

  The overall complexity for a given subset $\mathcal{K}$ is thus equal to $O(s^{2.376})$, since it is dominated by matrix inversion.

- The optimization problem requires the evaluation of $F(\mathcal{K})$ for every possible subsets of s elements resulting in a number of operations equal to

$$\binom{n}{s} = \frac{n!}{(n-s)!s!} \approx O(n^s)$$

Thus, the **exact brute-force solution** of the optimization problem of interest has complexity

$$O(n^s s^{2.376}) \tag{3.11}$$

which is exponential in s and makes exact optimization infeasible for large n.

3.3 Greedy Approximation

In this section, we address the method computational cost issue by proposing a greedy algorithm that reduces overall complexity. We will then analyse some performance guarantees of the approximation based on established findings in the literature [34, 35, 80]. For a general overview in Greedy Algorithms refer to Appendix A.2.2.

First, let us define the main steps of the greedy algorithm adapted to our specific problem of interest.

1. Denote with $\mathcal{K}_t$ the subset of nodes selected at t-th algorithm iteration and initialize $\mathcal{K}_0 = \emptyset$.

2. For each iteration $t = 1, \dots, s$, with s cardinality of the observed set to select, set $\mathcal{K}_t = \mathcal{K}_{t-1} \cup \{i_t\}$ with

$$i_t \in \arg \max_{i \in \mathcal{R} \setminus \mathcal{K}_{t-1}} F(\mathcal{K}_{t-1} \cup \{i_t\}) .$$

3. At the end of s-th iteration, the algorithm will propose $\mathcal{K}^G = \mathcal{K}_s$ as an approximated solution to the optimization problem

Notably, applying the greedy algorithm reduces the number of required evaluations of the objective function to $O(ns)$. However, we now focus on optimizing the evaluation of $F(\mathcal{K})$ to further reduce computational cost by avoiding matrix inversion at each iteration.

Proposition 3.1. *Consider a random vector X with zero mean and invertible covariance C, and let $Y = \alpha' X$, with $\alpha \in \mathbb{R}^n$. Let F be the variance reduction function as in* (3.7)*, $\mathcal{K}$ be the observed set at step $t-1$ and i be the candidate extra node to observe at t-iteration. The iterative step of the greedy algorithm is defined as*

$$F(\mathcal{K} \cup \{i\}) = \delta + \frac{(C\alpha)_i^2}{C_{ii}} \left[1 + C_{i\mathcal{K}} (C_{\mathcal{K}\mathcal{K}})^{-1} \left(1 - 2 \frac{(C\alpha)_{\mathcal{K}}}{C_{ii}} \right) \right]$$

where $\delta = (C\alpha)'_{\mathcal{K}} (C_{\mathcal{K}\mathcal{K}})^{-1} (C\alpha)_{\mathcal{K}}$ is a constant which depends only on $\mathcal{K}$.

Proof. Recall that

$$F(\mathcal{K} \cup \{i\}) = (C\alpha)'_{\mathcal{K}\cup\{i\}} (C_{\mathcal{K}\cup\{i\}\mathcal{K}\cup\{i\}})^{-1} (C\alpha)_{\mathcal{K}\cup\{i\}}$$

Step 1.
Note that, using the Schur Complement (see Appendix A.1 for reference), it holds:

$$(C_{\mathcal{K}\cup\{i\}\mathcal{K}\cup\{i\}})^{-1} = \begin{bmatrix} C_{\mathcal{K}\mathcal{K}} & C_{\mathcal{K}i} \\ C_{i\mathcal{K}} & C_{ii} \end{bmatrix}^{-1} = \begin{bmatrix} K_{11} & K_{12} \\ K_{21} & K_{22} \end{bmatrix}$$

where

$$\begin{aligned} K_{11} &= \left(C_{\mathcal{K}\mathcal{K}} - \tfrac{C_{\mathcal{K}i}C_{i\mathcal{K}}}{C_{ii}}\right)^{-1} \\ K_{12} &= -K_{11}\tfrac{C_{\mathcal{K}i}}{C_{ii}} \\ K_{21} &= -\tfrac{C_{\mathcal{K}i}}{C_{ii}}K_{11} \\ K_{22} &= \tfrac{1}{C_{ii}} + \tfrac{C_{i\mathcal{K}}}{C_{ii}^2}K_{11}C_{\mathcal{K}i} \end{aligned}$$

Step 2.
Note that, using Sherman-Morrison formula (see Appendix A.1 for reference), it holds:

$$K_{11} = \left(C_{\mathcal{K}\mathcal{K}} - \tfrac{C_{\mathcal{K}i}C_{i\mathcal{K}}}{C_{ii}}\right)^{-1} = C_{\mathcal{K}\mathcal{K}}^{-1} + \tfrac{C_{\mathcal{K}\mathcal{K}}^{-1}C_{\mathcal{K}i}C_{i\mathcal{K}}C_{\mathcal{K}\mathcal{K}}^{-1}}{C_{ii}-C_{i\mathcal{K}}C_{\mathcal{K}\mathcal{K}}^{-1}C_{\mathcal{K}i}}$$

Step 3.
We thus obtain:

$$\begin{aligned} F(\mathcal{K}\cup\{i\}) = \; & (C\alpha)'_{\mathcal{K}}(C_{\mathcal{K}\mathcal{K}})^{-1}(C\alpha)_{\mathcal{K}} - (C\alpha)_i\tfrac{C_{\mathcal{K}i}}{C_{ii}}(C_{\mathcal{K}\mathcal{K}})^{-1}(C\alpha)_{\mathcal{K}} \\ & -(C\alpha)'_{\mathcal{K}}(C_{\mathcal{K}\mathcal{K}})^{-1}\tfrac{C_{\mathcal{K}i}}{C_{ii}}(C\alpha)_i + \tfrac{(C\alpha)_i^2}{C_{ii}} + \tfrac{(C\alpha)_i^2}{C_{ii}}C_{i\mathcal{K}}(C_{\mathcal{K}\mathcal{K}})^{-1}C_{\mathcal{K}i} \end{aligned}$$

The thesis follows. □

Remark 3.1. *Notice that at the first step, i.e. $t = 0$, $\mathcal{K}_0 = \emptyset$ and the best node to observe is computed as*

$$F(\{i\}) = \frac{(C\alpha)_i^2}{C_{ii}}$$

as expected from Corollary 3.1.

Proposition 3.2. *The inverse matrix $(C_{\mathcal{K}\mathcal{K}})^{-1}$ can be computed iteratively without requiring any matrix inversion. More precisely,*

- *for $t = 1$ since $\mathcal{K}_1 = \{i\}$ then $(C_{\mathcal{K}_1\mathcal{K}_1})^{-1} = \frac{1}{C_{ii}}$*

- *In the subsequent steps, indicated with I the inverse matrix at time $t-1$, i.e. $(C_{\mathcal{K}_{t-1}\mathcal{K}_{t-1}})^{-1}$, then $I = (C_{\mathcal{K}_t\mathcal{K}_t})^{-1}$ is computed as*

$$I = \begin{bmatrix} \bar{I} & -\bar{I}\frac{C_{\mathcal{K}i^*}}{C_{i^*i^*}} \\ -\frac{C_{\mathcal{K}i^*}}{C_{i^*i^*}}\bar{I} & \frac{1}{C_{i^*i^*}} + \frac{C_{i^*\mathcal{K}}}{C_{i^*i^*}} IC_{\mathcal{K}i^*} \end{bmatrix}$$

where $\bar{I} = I + \frac{IC_{\mathcal{K}i^*}C_{i^*\mathcal{K}}I}{C_{i^*i^*} - C_{i^*\mathcal{K}}IC_{\mathcal{K}i^*}}$

Proof. We apply here the same strategy proposed in Step 1-2 the proof of Proposition 3.1. No matrix inversion is required since $(C_{\mathcal{K}\mathcal{K}})^{-1}$ can be written as function of the inverse matrix at previous step. In addition, the first inverse matrix of the iterative procedure is computed for a single node observation, hence $(C_{ii})^{-1} = \frac{1}{C_{ii}}$. □

Based on Proposition 3.1 - 3.2, we formally define the pseudo-code for the greedy algorithm, as presented in Algorithm 1.

Algorithm 1: Greedy algorithm for optimal subset selection

Data: C, s, $\mathcal{R}$
Result: $\mathcal{K}$
$\mathcal{K} \leftarrow \emptyset$;
$F_{max} \leftarrow 0$;
$I \leftarrow 0$;
for $t = 1, \ldots, s$ **do**
  $F_{max} \leftarrow 0$;
  for $i \in \mathcal{R} \backslash \mathcal{K}$ **do**
    $F_i \leftarrow \frac{(C\alpha)_i^2}{C_{ii}} \left[1 + C_{i\mathcal{K}} I \left(1 - 2\frac{(C\alpha)_{\mathcal{K}}}{C_{ii}}\right)\right]$;
    if $F_i > F_{max}$ **then**
      $F_{max} \leftarrow F_i$;
      $i^* \leftarrow i$;
    end
  end
  $\mathcal{K} \leftarrow \mathcal{K} \cup \{i^*\}$;
  $\bar{I} \leftarrow I + \frac{IC_{\mathcal{K}i^*}C_{i^*\mathcal{K}}I}{C_{i^*i^*} - C_{i^*\mathcal{K}}IC_{\mathcal{K}i^*}}$;
  $I \leftarrow \begin{bmatrix} \bar{I} & -\bar{I}\frac{C_{\mathcal{K}i^*}}{C_{i^*i^*}} \\ -\frac{C_{\mathcal{K}i^*}}{C_{i^*i^*}}\bar{I} & \frac{1}{C_{i^*i^*}} + \frac{C_{i^*\mathcal{K}}}{C_{i^*i^*}} IC_{\mathcal{K}i^*} \end{bmatrix}$;
end

We now analyze the computational complexity of the proposed greedy algorithm

in relation to the exact brute-force approach presented in the previous section, to remark its scalability also for large-scale networks.

Proposition 3.3 (Computational Cost)**.** *The greedy approximation proposed in Algorithm 1 achieves an overall computational cost of $O(ns^3)$ for the subset selection problem, significantly reducing the complexity compared to the exact brute-force algorithm as highlighted in* (3.11)*.*

Proof. Let us analyse the computational cost of the two main steps:

(i) Since one element is added at each step the total number of iterations of the algorithm is equal to s and the total number of evaluation of the objective function $F(\mathcal{K})$ are thus equal to $O(ns)$;

(ii) Applying the iterative computations introduced in Proposition 3.2 the computational cost of matrix inversion is reduced to $O(s^2)$ and this coincides with the cost associated to the single evaluation of the objective function $F(\mathcal{K})$.

Collecting these results the thesis follows.

□

Although the greedy algorithm yields a sub-optimal solution, previous studies [34] have demonstrated that it retains performance guarantees under certain conditions. Notably, even in the absence of additional assumptions on the involved variables, if the covariance terms are sufficiently small, the greedy algorithm ensures strong approximation guarantees for $F(\mathcal{K})$. More precisely, from Theorem 3.5 in [34], the following bound holds.

Proposition 3.4 (Performance guarantees)**.** *If $C_{ij} \leq \delta < \frac{1}{4s}$ for all i, j, with $s = |\mathcal{K}|$, the set $\mathcal{K}^G$ returned by the greedy algorithm guarantees that*

$$F(\mathcal{K}^G) \geq (1 - 4\delta s)F(\mathcal{K}^*),$$

where $\mathcal{K}^*$ is the set maximizing $F(\mathcal{K})$ among all the possible size s sets $\mathcal{K} \subseteq \mathcal{Z}$.

3.4 Submodular Variance Reduction

The performance of the greedy algorithm improves when additional structural assumptions are imposed on the objective function. The most well-established result in the literature involves the concept of submodularity, as stated by [80]. In this section, we first recall the definition of submodularity and its main performance guarantees to ensure the self-consistency of this dissertation. We then delve deeper into our specific problem, analyzing the conditions under which the variance reduction function exhibits submodularity.

3.4.1 Background on Submodularity and Algorithm Performances

Let us first introduce a particular class of functions known as *submodular*. The analysis of these functions is significant because it enables us to achieve strong approximation results in several challenging combinatorial optimization problems. In particular, this is relevant for the greedy approximation algorithm proposed in our dissertation, as outlined in [80].

Definition 3.1. *Given a set $\mathcal{V}$, a function $f : 2^{\mathcal{V}} \to \mathbb{R}$ is*

(i) sub-modular if

$$f(\mathcal{A} \cup \{k\}) - f(\mathcal{A}) \geq f(\mathcal{B} \cup \{k\}) - f(\mathcal{B}), \tag{3.12}$$

for every $\mathcal{A} \subseteq \mathcal{B} \subseteq \mathcal{V}$ and $k \in \mathcal{V} \backslash \mathcal{B}$;

(ii) super-modular if $-f$ is sub-modular, i.e., if

$$f(\mathcal{A} \cup \{k\}) - f(\mathcal{A}) \leq f(\mathcal{B} \cup \{k\}) - f(\mathcal{B}), \tag{3.13}$$

for every $\mathcal{A} \subseteq \mathcal{B} \subseteq \mathcal{V}$ and $k \in \mathcal{V} \backslash \mathcal{B}$.

One of the most widely used properties about submodular functions is that a simple greedy algorithm approximately optimizes the function subject to a cardinality constraint. In addition, despite the sub-optimality of the proposed solution, from [80] if function f is submodular, then we have some guarantees on method performances, as properly stated in the following Theorem.

Theorem 3.2 (Nemhauser et al. [80]). *Given $f : 2^{\mathcal{V}} \to \mathbb{R}$ a submodular function, the set $\mathcal{K}^G$ of size s returned by the greedy algorithm guarantees that*

$$f(\mathcal{K}^G) \geq \left(1 - \frac{1}{e}\right) f(\mathcal{K}^*),$$

where $\mathcal{K}^*$ is the set maximizing $f(\mathcal{K})$ among all the possible size s sets $\mathcal{K} \subseteq \mathcal{Z}$.

Proof. Indicate with $\mathcal{K}_l$ the selected subset at iteration l. Since $\mathcal{K}^*$ is the optimal subset, the following inequality yields:

$$f(\mathcal{K}^*) \leq f(\mathcal{K}^* \cup \mathcal{K}^l) = f(\mathcal{K}^l) + f(\mathcal{K}^* \backslash \mathcal{K}^l \cup \mathcal{K}^l) - f(\mathcal{K}^l)$$

Let $\rho_i(\mathcal{K}) = f(\mathcal{K} \cup \{i\}) - f(\mathcal{K})$, from submodularity definition it holds

$$f(\mathcal{K}^*) \leq f(\mathcal{K}^l) + \sum_{i=1}^{l} \rho_i(\mathcal{K}^l) \leq f(\mathcal{K}^l) + l \max \rho_i(\mathcal{K}^l)$$

Thus,

$$f(\mathcal{K}^{l+1}) - f(\mathcal{K}^l) \geq \frac{1}{l}[f(\mathcal{K}^*) - f(\mathcal{K}^l)]$$

$$f(\mathcal{K}^{l+1}) \geq \frac{1}{l} f(\mathcal{K}^*) + \left(1 - \frac{1}{l}\right) f(\mathcal{K}^l)$$

$$\begin{aligned} f(\mathcal{K}^{l+1}) - f(\emptyset) &\geq \tfrac{1}{l}\left(1 + \cdots + \left(1 - \tfrac{1}{l}\right)^{l+1}\right)(f(\mathcal{K}^*) - f(\emptyset)) \\ &= \left(1 - \left(1 - \tfrac{1}{l}\right)^{l}\right)(f(\mathcal{K}^*) - f(\emptyset)) \\ &\geq \left(1 - \tfrac{1}{l}\right) f(\mathcal{K}^*) - f(\emptyset) \end{aligned}$$

The thesis follows. □

3.4.2 Variance Reduction Submodularity Conditions

In Section 3.4.1, we observed that the greedy algorithm yields stronger approximation guarantees when the objective function is submodular, as stated in Theorem 3.2. We now analyze the conditions under which the problem of interest satisfies submodularity. More precisely, we generalize the submodularity result for variance reduction function, stated in Theorem 8 in [81], in case of arbitrary variables, not necessarily multivariate Gaussian.

Theorem 3.3. *Let X be a random vector with invertible covariance matrix C and $Y = \alpha' X$. If $\alpha \in \mathbb{R}^n_+$ and the precision matrix $H := C^{-1}$ is such that $H_{ij} \leq 0$ for any $i \neq j$ in $\mathcal{V} = \{1, \dots, n\}$, i.e. $-H$ is a Metzler matrix, then:*

(i) the function $G : 2^{\mathcal{V}} \to \mathbb{R}$ defined by

$$G(\mathcal{K}) = \alpha'_{-\mathcal{K}} (H_{-\mathcal{K}-\mathcal{K}})^{-1} \alpha_{-\mathcal{K}}, \quad \mathcal{K} \subseteq \mathcal{V} \tag{3.14}$$

is super-modular;

(ii) the function $F : 2^{\mathcal{V}} \to \mathbb{R}$ defined by

$$F(\mathcal{K}) = (C\alpha)'_{\mathcal{K}} (C_{\mathcal{K}\mathcal{K}})^{-1} (C\alpha)_{\mathcal{K}}, \quad \mathcal{K} \subseteq \mathcal{V} \tag{3.15}$$

is sub-modular.

Proof. (i) We first derive a representation of the function G. To this aim, let D in $\mathbb{R}^{n \times n}$ have entries

$$D_{ii} = 1/\sqrt{H_{ii}}, \qquad D_{ij} = 0, \qquad \forall i \neq j \in \mathcal{V}.$$

Define

$$\overline{H} = DHD, \qquad \Theta = I - \overline{H}.$$

Notice that, by construction,

$$\Theta_{ii} = 1 - \overline{H}_{ii} = 1 - D_{ii} H_{ii} D_{ii} = 1 - \frac{H_{ii}}{\sqrt{H_{ii} H_{ii}}} = 0, \tag{3.16}$$

for every i in $\mathcal{V}$. Moreover, using the assumption on H,

$$\Theta_{ij} = -\overline{H}_{ij} = -D_{ii} H_{ij} D_{jj} = \frac{-H_{ij}}{\sqrt{H_{ii} H_{jj}}} \geq 0, \tag{3.17}$$

for every $i \neq j$ in $\mathcal{V}$.

We now fix a subset $\mathcal{K} \subseteq \mathcal{V}$ and we notice that the matrix $\Theta_{-\mathcal{K},-\mathcal{K}}$ is non-negative thanks to (3.16) and (3.17). Perron-Frobenius Theorem then implies that $\Theta_{-\mathcal{K},-\mathcal{K}}$ has a real positive dominant eigenvalue $\lambda_{\mathcal{K}} \geq 0$ such that $\rho(\Theta_{-\mathcal{K}-\mathcal{K}}) = \lambda_{\mathcal{K}}$.

Moreover, since H is symmetric positive definite, so is $\overline{H} = DHD$. By the Sylvester criterium, also $\overline{H}_{-\mathcal{K}-\mathcal{K}} = I - \Theta_{-\mathcal{K}-\mathcal{K}}$ is symmetric positive definite, so that necessarily $\lambda_{\mathcal{K}} < 1$. It follows that $\rho(\Theta_{-\mathcal{K}-\mathcal{K}}) < 1$ and we can write

$$(\overline{H}_{-\mathcal{K}-\mathcal{K}})^{-1} = (I - \Theta_{-\mathcal{K}-\mathcal{K}})^{-1} = \sum_{l=0}^{+\infty} (\Theta_{-\mathcal{K}-\mathcal{K}})^l . \tag{3.18}$$

For every i and j in $\mathcal{V} \setminus \mathcal{K}$, and $l \geq 0$, let

$$\Gamma_{ij}^l(\mathcal{K}) = \left\{ \gamma = (\gamma_0, \gamma_1, \ldots, \gamma_l) \in (\mathcal{V} \setminus \mathcal{K})^{l+1} \colon \gamma_0 = i, \gamma_l = j \right\},$$

and define the sequence weights

$$\theta_\gamma = \prod_{1 \leq h \leq l} \Theta_{\gamma_{h-1}\gamma_h}, \qquad \forall \gamma \in \Gamma_{ij}^l(\mathcal{K}),$$

(with the convention that $\theta_\gamma = 1$ when $l = 0$ and $\gamma = (i)$). Then, (3.18) reads

$$((\overline{H}_{-\mathcal{K}-\mathcal{K}})^{-1})_{ij} = \sum_{l=0}^{+\infty} \sum_{\gamma \in \Gamma_{ij}^l(\mathcal{K})} \theta_\gamma . \tag{3.19}$$

Now, observe that (3.14) and (3.19) imply that

$$\begin{aligned}
G(\mathcal{K}) &= \alpha'_{-\mathcal{K}} (H_{-\mathcal{K}-\mathcal{K}})^{-1} \alpha_{-\mathcal{K}} \\
&= \alpha'_{-\mathcal{K}} ((D^{-1}\overline{H}D^{-1})_{-\mathcal{K}-\mathcal{K}})^{-1} \alpha_{-\mathcal{K}} \\
&= \alpha'_{-\mathcal{K}} (D^{-1}_{-\mathcal{K}-\mathcal{K}} \overline{H}_{-\mathcal{K}-\mathcal{K}} D^{-1}_{-\mathcal{K}-\mathcal{K}})^{-1} \alpha_{-\mathcal{K}} \\
&= \alpha'_{-\mathcal{K}} D_{-\mathcal{K}-\mathcal{K}} (\overline{H}_{-\mathcal{K}-\mathcal{K}})^{-1} D_{-\mathcal{K}-\mathcal{K}} \alpha_{-\mathcal{K}} \\
&= \sum_{i,j \in \mathcal{V} \setminus \mathcal{K}} \frac{\alpha_i ((\overline{H}_{-\mathcal{K}-\mathcal{K}})^{-1})_{ij} \alpha_j}{\sqrt{H_{ii} H_{jj}}} \\
&= \sum_{i,j \in \mathcal{V} \setminus \mathcal{K}} \sum_{l=0}^{+\infty} \sum_{\gamma \in \Gamma_{ij}^l(\mathcal{K})} \frac{\alpha_i \theta_\gamma \alpha_j}{\sqrt{H_{ii} H_{jj}}} .
\end{aligned} \tag{3.20}$$

Now, consider an arbitrary k in $\mathcal{V} \setminus \mathcal{K}$ and let $\overline{\mathcal{K}} = \mathcal{K} \cup \{k\}$. Using representation (3.20), we express $G(\mathcal{K}) - G(\overline{\mathcal{K}})$ as a sum of three terms: the first one is a sum over sequences starting and ending in k, the second one in terms of sequences starting or ending in k but not both, and the third sum is over sequences not starting and not

ending in k. We further define, for every i and j in $\mathcal{V} \setminus \overline{\mathcal{K}}$,

$$\Gamma_{ij}^{l}(\mathcal{K},k) = \left\{ \gamma \in \Gamma_{ij}^{l}(\mathcal{K}) \,|\, k \in \{\gamma_1 \ldots, \gamma_{n-1}\} \right\},$$

We now represent, using (3.20),

$$G(\mathcal{K}) - G(\overline{\mathcal{K}}) = E_1(\mathcal{K}) + E_2(\mathcal{K}) + E_3(\mathcal{K}) \tag{3.21}$$

where

$$\begin{aligned}
E_1(\mathcal{K}) &= \sum_{l=0}^{+\infty} \sum_{\gamma \in \Gamma_{kk}^{l}(\mathcal{K})} \frac{\alpha_k \theta_\gamma \alpha_k}{H_{kk}} \\
E_2(\mathcal{K}) &= 2 \sum_{i \in \mathcal{V} \setminus \overline{\mathcal{K}}} \sum_{l=0}^{+\infty} \sum_{\gamma \in \Gamma_{ik}^{l}(\mathcal{K})} \frac{\alpha_i \theta_\gamma \alpha_k}{\sqrt{H_{ii} H_{kk}}} \\
E_3(\mathcal{K}) &= \sum_{i,j \in \mathcal{V} \setminus \overline{\mathcal{K}}} \sum_{l=0}^{+\infty} \sum_{\gamma \in \Gamma_{ij}^{l}(\mathcal{K},k)} \frac{\alpha_i \theta_\gamma \alpha_j}{\sqrt{H_{ii} H_{jj}}}.
\end{aligned}$$

Notice that, since $\theta_\gamma \geq 0$, $\alpha \in \mathbb{R}_+^n$, and the subsets $\Gamma_{ij}^{l}(\mathcal{K})$ and $\Gamma_{ij}^{l}(\mathcal{K},k)$ are monotonically decreasing in $\mathcal{K}$, for every $\mathcal{A} \subseteq \mathcal{B} \subseteq \mathcal{V}$, we have that

$$E_h(\mathcal{A}) \geq E_h(\mathcal{B}), \qquad \forall h = 1,2,3,$$

so that

$$\begin{aligned}
G(\mathcal{A}) - G(\mathcal{A} \cup \{k\}) &= E_1(\mathcal{A}) + E_2(\mathcal{A}) + E_3(\mathcal{A}) \\
&\geq E_1(\mathcal{B}) + E_2(\mathcal{B}) + E_3(\mathcal{B}) \\
&= G(\mathcal{B}) - G(\mathcal{B} \cup \{k\}),
\end{aligned}$$

thus showing that G is a super-modular function.
(ii) Since

$$G(\mathcal{K}) = \mathrm{Var}(Y) - F(\mathcal{K}),$$

it follows that G super-modular implies F sub-modular. □

As a direct consequence of Theorem 3.3 due to Theorem 3.2, the following result holds true.

Corollary 3.2. *Let X be a random vector with precision matrix H and $Y = \alpha' X$, with $\alpha \in \mathbb{R}_+^n$. If $-H$ is a Metzler matrix, i.e. $H_{ij} \leq 0$ for all $i \neq j$ in $\mathcal{V} = \{1,\ldots,n\}$, then Algorithm 1 is a $(1-1/e)$-approximation for Problem 3.3. More precisely, fixed the cardinality of the observable set to s and given $\mathcal{K}^G$ the solution proposed by the*

greedy algorithm, it holds

$$F(\mathcal{K}^G) \geq \left(1 - \frac{1}{e}\right) \max_{\mathcal{K} \subseteq \mathcal{V}: |\mathcal{K}| = s} F(\mathcal{K}),$$

where F is the variance reduction fucntion introduced in (3.15)*.*

3.4.3 Comparison with State-of-the-Art Conditions

The submodularity of the variance reduction function in the context of the subset selection problem has been previously analyzed by Das and Kempe in [34]. However, our analysis appears to be more general, particularly in relation to the specific context of application that we focus on in this dissertation. Before delving into the conditions proposed in [34] let us preliminary introduce some definitions, taken from the statistics community [82], that will be fundamental in the following.

Definition 3.2. *Consider a random variable $Y \in \mathbb{R}$ to estimate and a random vector $X \in \mathbb{R}^n$ of observable variables with invertible covariance C. The **residual** $Res(Y, X_\mathcal{A})$ is defined as the part of Y that is not correlated with the variables X_i for all $i \in \mathcal{A} \subseteq \mathcal{V} = \{1, \ldots, n\}$, i.e. the part of Y that cannot be explained by $X_\mathcal{A}$. Formally,*

$$Res(Y, X_\mathcal{A}) = Y - \sum_{i \in \mathcal{A}} \hat{\alpha}_i^\mathcal{A} X_i \tag{3.22}$$

where $\hat{\alpha}_i^\mathcal{A}$ are the best regression coefficients to estimate Y given $X_\mathcal{A}$, i.e., as in (3.5)*,*

$$\hat{\alpha}^\mathcal{A} = (C_{\mathcal{A}\mathcal{A}})^{-1} b_\mathcal{A},$$

with $b_\mathcal{A}$ defined s.t. $(b_\mathcal{A})_i = Cov(Y, X_i)$.

Definition 3.3 (Suppressor variable)**.** *Consider a random variable $Y \in \mathbb{R}$ to estimate and a random vector $X \in \mathbb{R}^n$ of observable variables with invertible covariance C. A variable X_j is said to be a **suppressor variable** if X_i is not correlated (or only slightly) with Y but it appears to be much more correlated with Y once X_j is sampled. Formally, X_j is a suppressor variable if for some variable X_i*

$$|Cor(Y, Res(X_i, X_j)| > |Cor(Y, X_i)|,$$

where $Res(X_i, X_j)$ is the residual of X_i with respect to X_j, i.e. the part of X_i not captured by X_j and $Cor(\cdot,\cdot)$ is the correlation coefficient introduce in (3.10).

Definition 3.4 (Conditional suppressor variable)**.** *Consider a random variable $Y \in \mathbb{R}$ to estimate and a random vector $X \in \mathbb{R}^n$ of observable variables with invertible covariance C. Given a set S of random variables, X_j is referred to as a suppressor conditioned on S if and only if*

$$|Cor(Y, Res(Res(X_i, S), Res(X_j, S)))| > |Cor(Y, Res(X_i, S))|.$$

Let us now shift our focus to the submodularity conditions for the variance reduction function, with proper reference to the state-of-the-art. The authors in [34] highlight that a key condition for ensuring submodularity and consequenlty a good performance for Algorithm 1 is the absence of "*conditional suppressor variables*", i.e. the absence of suppressor variables conditioned on any possible subset $S \subseteq \mathcal{V}$. However, requiring the absence of suppressor variables presents two potential challenges. First, it is a condition that is not straightforward to verify. Furthermore, from a statistical perspective, suppressor variables are valuable for enhancing predictive performance [82], making their exclusion potentially restrictive. Therefore, our result has broader potential applications.

3.5 Approximate Submodular Variance Reduction

Referring to [35], it is well known that the applicability of the greedy algorithm, while still preserving strong performance guarantees, extends even further than what has been presented so far. Specifically, even if the function is not submodular, it may still be close to be submodular.

In this Section, we will recall the concept of approximate submodularity introduced in [35] and the performance of the greedy algorithm guaranteed by this weaker condition. The results in this section are a reformulation of those presented by Kempe, but our main contribution lies in thoroughly analyzing the problem in the context of our specific application and in reconstructing, in detail, the proofs that are either partially omitted or not immediately accessible in the existing literature.

3.5.1 Background on Approximate Submodularity and Algorithm Performances

In order to indicate how close a certain function $f : 2^{\mathcal{V}} \to \mathbb{R}$ is to submodular, the notion of *submodularity ratio* is introduced.

Definition 3.5. *The submodularity ratio of a monotone function $f : 2^{\mathcal{V}} \to \mathbb{R}$ with respect to a set $\mathcal{K}$ and a parameter $s \geq 1$ is*

$$\gamma_{\mathcal{K},s} = \min_{\mathcal{A}\subseteq\mathcal{K},\mathcal{B}:|\mathcal{B}|\leq s,\mathcal{A}\cap\mathcal{B}=\emptyset} \frac{\sum_{k\in\mathcal{B}} f(\mathcal{A}\cup\{k\}) - f(\mathcal{A})}{f(\mathcal{A}\cup\mathcal{B}) - f(\mathcal{A})} \tag{3.23}$$

This definition generalizes submodularity notion in Definition 3.1, as properly stated in the following Proposition.

Proposition 3.5 (Proposition 3 from Das and Kempe [35])**.** *The monotonic function $f : 2^{\mathcal{V}} \to \mathbb{R}$ is submodular if and only if $\gamma_{\mathcal{K},s} \geq 1$ for all $\mathcal{K}$ and s.*

Analogously to previous section, we recall here the generalization of Theorem 3.2 in case of approximately submodular functions.

Theorem 3.4 (Theorem 6 from Das and Kempe [35])**.** *Given $f : 2^{\mathcal{V}} \to \mathbb{R}$ a monotone function and a parameter $s \geq 1$, the set $\mathcal{K}^G$ returned by the greedy algorithm guarantees that*

$$f(\mathcal{K}^G) \geq \left(1 - e^{-\gamma_{\mathcal{K}^G,s}}\right) f(\mathcal{K}^*),$$

where $\mathcal{K}^*$ is the set maximizing $f(\mathcal{K})$ among all the possible size s sets $\mathcal{K} \subseteq \mathcal{V}$, and $\gamma_{\mathcal{K}^G,s}$ is the submodularity ratio of f defined as in (3.23)*.*

However, the exact computation of the ratio in (3.23) is not easy since it is defined as a minimum over an exponential number values. Thus, depending on the specific problem, it is usually of interest to compute a proper lower bound.

3.5.2 Analysis of Approximate Submodularity Conditions

The applicability of the introduced Greedy Algorithm 1 becomes even broader when considering conditions under which the problem is approximately submodular, as outlined in Theorem 3.4.

First, we consider the submodularity ratio in the context of our specific problem. In particular, we propose an alternative formulation based on Definition 3.5, tailored to the application of interest.

Proposition 3.6. *Consider X a random vector with invertible covariance matrix C such that $C_{ii} \leq 1$ for every $i \in \mathcal{V} = \{1,\dots,n\}$, and a random variable $Y = \alpha' X$ with $\alpha \in \mathbb{R}^n$. Given $F : 2^{\mathcal{V}} \to \mathbb{R}$ the variance reduction function in* (3.15)*, it holds:*

$$\gamma_{\mathcal{K},s} = \min_{\mathcal{A}\subseteq\mathcal{K},\mathcal{B}:|\mathcal{B}|\leq s,\mathcal{A}\cap\mathcal{B}=\emptyset} \frac{(b_{\mathcal{B}}^{\mathcal{A}})'(b_{\mathcal{B}}^{\mathcal{A}})}{(b_{\mathcal{B}}^{\mathcal{A}})'(C_{\mathcal{B}\mathcal{B}}^{\mathcal{A}})^{-1}(b_{\mathcal{B}}^{\mathcal{A}})} \tag{3.24}$$

where

$$b^{\mathcal{A}} := D^{\frac{1}{2}}\mathbb{E}[YRes(X_{\mathcal{B}},X_{\mathcal{A}})]\ ,\ C^{\mathcal{A}} := D^{\frac{1}{2}}Cov(Res(X_{\mathcal{B}},X_{\mathcal{A}}),Res(X_{\mathcal{B}},X_{\mathcal{A}}))D^{\frac{1}{2}}$$

with $D := [Cov(Res(X_i,X_{\mathcal{A}}),Res(X_i,X_{\mathcal{A}}))]_{i\in\mathcal{B}}$ normalization diagonal matrix.

Proof. Let us preliminary rewrite the definition of $\gamma_{\mathcal{K},s}$ introduced in (3.23), using the notation specific to our problem of interest:

$$\gamma_{\mathcal{K},s} = \min_{\mathcal{A}\subseteq\mathcal{K},\mathcal{B}:|\mathcal{B}|\leq s,\mathcal{A}\cap\mathcal{B}=\emptyset} \frac{\sum_{i\in\mathcal{B}} F(\mathcal{A}\cup\{i\}) - F(\mathcal{A})}{F(\mathcal{A}\cup\mathcal{B}) - F(\mathcal{A})} \tag{3.25}$$

Our aim is now to prove that it is equivalent to (3.24). Preliminary, let us observe that, using Lemma B.2 in Appendix B, i.e.

$$\begin{array}{ll} F(\mathcal{A}\cup\{i\}) - F(\mathcal{A}) & = F(\mathrm{Res}(X_i,X_{\mathcal{A}})) \\ F(\mathcal{A}\cup\mathcal{B}) - F(\mathcal{A}) & = F(\mathrm{Res}(X_{\mathcal{B}},X_{\mathcal{A}})) \end{array}\ ,$$

the ratio (3.25) can be rewritten as

$$\frac{\sum_{i\in\mathcal{B}} F(\mathcal{A}\cup\{i\}) - F(\mathcal{A})}{F(\mathcal{A}\cup\mathcal{B}) - F(\mathcal{A})} = \frac{\sum_{i\in\mathcal{B}} F(\mathrm{Res}(X_i,X_{\mathcal{A}}))}{F(\mathrm{Res}(X_{\mathcal{B}},X_{\mathcal{A}}))}\,. \tag{3.26}$$

First, focus on the numerator of the ratio. The following series of equalities hold:

$$\begin{aligned} F(\mathrm{Res}(X_i,X_{\mathcal{A}})) &= \mathrm{Var}(Y) - G(\mathrm{Res}(X_i,X_{\mathcal{A}})) = \\ &= \mathrm{Var}(Y) - \mathbb{E}[(Y - \mathbb{E}[Y|\mathrm{Res}(X_i,X_{\mathcal{A}})])^2] = \\ &= \mathrm{Var}(Y) - \mathbb{E}[(Y - \hat{\alpha}\mathrm{Res}(X_i,X_{\mathcal{A}}))^2] = \\ &= 2\hat{\alpha}\mathbb{E}[Y\mathrm{Res}(X_i,X_{\mathcal{A}})] - \hat{\alpha}^2\mathbb{E}[(\mathrm{Res}(X_i,X_{\mathcal{A}}))^2] \end{aligned} \tag{3.27}$$

where the first equality comes from F definition in (3.2), the second follows from (3.1) and the third from (3.4). Using (3.5) we can also find an explcit formulation for $\hat{\alpha}$, i.e.

$$\hat{\alpha} = (C_{\text{Res}(X_i,X_\mathcal{A})})^{-1} b_{\text{Res}(X_i,X_\mathcal{A})}$$

where

$$C_{\text{Res}(X_i,X_\mathcal{A})} = \text{Cov}(\text{Res}(X_i,X_\mathcal{A}),\text{Res}(X_i,X_\mathcal{A}))\ ,\ b_{\text{Res}(X_i,X_\mathcal{A})} = \mathbb{E}[Y\text{Res}(X_i,X_\mathcal{A})]\,.$$

Substituting the proper value of $\hat{\alpha}$ in (3.27) and noticing that $\mathbb{E}[(\text{Res}(X_i,X_\mathcal{A}))^2] = C_{\text{Res}(X_i,X_\mathcal{A})}$, we retrieve that

$$F(\text{Res}(X_i,X_\mathcal{A})) = (C_{\text{Res}(X_i,X_\mathcal{A})})^{-1} b^2_{\text{Res}(X_i,X_\mathcal{A})}\,.$$

Set $b_i^\mathcal{A} := (C_{\text{Res}(X_i,X_\mathcal{A})})^{-\frac{1}{2}} b_{\text{Res}(X_i,X_\mathcal{A})}$, the right hand numerator in (3.26) coincides with the one in the thesis, indeed

$$\sum_{i\in\mathcal{B}} F(\text{Res}(X_i,X_\mathcal{A})) = \sum_{i\in\mathcal{B}} (b_i^\mathcal{A})^2 = (b_\mathcal{B}^\mathcal{A})'(b_\mathcal{B}^\mathcal{A})\,.$$

Moreover, from previous computations we retrieve that the following series of equalities hold:

$$\begin{aligned} F(\text{Res}(X_\mathcal{B},X_\mathcal{A})) &= (b_{\text{Res}(X_\mathcal{B},X_\mathcal{A})})'(C_{\text{Res}(X_\mathcal{B},X_\mathcal{A})})^{-1} b_{\text{Res}(X_\mathcal{B},X_\mathcal{A})} = \\ &= (b_\mathcal{B}^\mathcal{A})' D^{-\frac{1}{2}} (C_{\text{Res}(X_\mathcal{B},X_\mathcal{A})})^{-1} D^{-\frac{1}{2}} b_\mathcal{B}^\mathcal{A} = \\ &= (b_\mathcal{B}^\mathcal{A})' (C_{\mathcal{B}\mathcal{B}}^\mathcal{A})^{-1} (b_\mathcal{B}^\mathcal{A}) \end{aligned} \quad .$$

where the second equality comes from $b_i^\mathcal{A}$ definition, D is a diagonal matrix defined s.t. $D_{ii} = C_{\text{Res}(X_i,X_\mathcal{A})}$ for all $i \in \mathcal{B}$ and the final equation follows once we define $C_{\mathcal{B}\mathcal{B}}^\mathcal{A} = D^{\frac{1}{2}} C_{\text{Res}(X_\mathcal{B},X_\mathcal{A})} D^{\frac{1}{2}}$.

Combining these results, the thesis follows. □

Hence, given $\gamma_{\mathcal{K},s}$ defined as above, $\mathcal{K}^G$ the optimal subset solution of the Greedy Algorithm, and $\mathcal{K}^*$ the global optimum, from Theorem 3.4 the following performance guarantee holds

$$F(\mathcal{K}^G) \geq \left(1 - e^{-\gamma_{\mathcal{K}^G,s}}\right) F(\mathcal{K}^*)\,.$$

However, computing $\gamma_{\mathcal{K}^G,s}$ is often computationally prohibitive. A common strategy, therefore, is to derive a suitable lower bound. For this reason, we now present a key result from the established literature [35] that applies to our setting.

Preliminary, let us introduce a technical Lemma that will be useful for the subsequent proof. This comes from Lemma 15 in [34], but an alternative and simpler proof, based on Weyl's Theorem (see Proposition A.2 in Appendix A.1), is here proposed.

Lemma 3.2. *Consider a random vector $X \in \mathbb{R}^n$ with invertible covariance matrix C. Given $\tilde{C} \in \mathbb{R}^{(n-1)\times(n-1)}$ the covariance matrix corresponding to the $n-1$ random variables $Res(X_1,X_n),\dots,Res(X_{n-1},X_n)$, then*

$$\lambda_{min}(C) \leq \lambda_{min}(\tilde{C})\,.$$

Proof. First, let us write explicitely the relation between the two matrices C and $\tilde{C}$. Focusing on the arbitrary (i,j) entry of the $\tilde{C}$ matrix, it holds

$$\begin{aligned}\tilde{C}_{ij} &= \mathrm{Cov}(\mathrm{Res}(X_i,X_n),\mathrm{Res}(X_j,X_n)) = \\ &= \mathbb{E}[\mathrm{Res}(X_i,X_n)\mathrm{Res}(X_j,X_n)] = \\ &= \mathbb{E}[(X_i - \tfrac{C_{in}}{C_{nn}}X_n)(X_j - \tfrac{C_{jn}}{C_{nn}}X_n)] = \\ &= C_{ij} - \tfrac{C_{in}C_{jn}}{C_{nn}}\end{aligned} \tag{3.28}$$

Let us indicate with $\mathcal{Z} = \{1,\dots,n-1\}$ and define

$$T := \frac{C_{\mathcal{Z}n}C_{n\mathcal{Z}}}{C_{nn}}$$

which is a symmetric matrix by construction. Thus, from (3.28) it holds

$$\tilde{C} = C_{\mathcal{Z}\mathcal{Z}} + T\,.$$

From Weyl's Theorem, then

$$\lambda_{min}(C_{\mathcal{Z}\mathcal{Z}}) \leq \lambda_{min}(\tilde{C})\,.$$

Finally, applying the Cauchy's interlacing theorem (see Proposition A.3 in Appendix A.1), since $C_{\mathcal{Z}\mathcal{Z}}$ is a submatrix of C, it holds

$$\lambda_{min}(C) \leq \lambda_{min}(C_{\mathcal{Z}\mathcal{Z}})\,.$$

The thesis follows. □

Proposition 3.7. *Consider a random vector $X \in \mathbb{R}^n$ with invertible covariance matrix C such that $C_{ii} \leq 1$ for every $i = 1, \dots, n$, and a random variable $Y = \alpha' X$ with $\alpha \in \mathbb{R}^n$. Given arbitrary subset $\mathcal{K} \subseteq \mathcal{V} = \{1, \dots, n\}$ of cardinality s, the submodularity ratio associated to variance reduction function $F : 2^{\mathcal{V}} \to \mathbb{R}$ is bounded by*

$$\gamma_{\mathcal{K},s} \geq \lambda_{min}(C)$$

with $\lambda_{min}(C)$ the minimum eigenvalue of covariance matrix C.

Proof. First, let us notice that, applying the Rayleigh characterization of eigenvalues (see Proposition A.1 in Appendix A.1), it holds

$$\frac{(b_{\mathcal{B}}^{\mathcal{A}})'(C_{\mathcal{B}\mathcal{B}}^{\mathcal{A}})^{-1}(b_{\mathcal{B}}^{\mathcal{A}})}{(b_{\mathcal{B}}^{\mathcal{A}})'(b_{\mathcal{B}}^{\mathcal{A}})} \leq \max_x \frac{x'(C_{\mathcal{B}\mathcal{B}}^{\mathcal{A}})^{-1}x}{x'x} = \lambda_{max}((C_{\mathcal{B}\mathcal{B}}^{\mathcal{A}})^{-1}) = \frac{1}{\lambda_{min}(C_{\mathcal{B}\mathcal{B}}^{\mathcal{A}})}.$$

Thus, combining this result with submodular ratio definition in (3.24), we retrieve

$$\gamma_{\mathcal{K},s} = \min_{\mathcal{A}\subseteq\mathcal{K},\mathcal{B}:|\mathcal{B}|\leq s,\mathcal{A}\cap\mathcal{B}=\emptyset} \lambda_{min}(C_{\mathcal{B}\mathcal{B}}^{\mathcal{A}})$$

Using now repeated applications of Lemma 3.2 in Appendix B, we relate $\lambda_{min}(C_{\mathcal{B}\mathcal{B}}^{\mathcal{A}})$ with $\lambda_{min}(C_{\mathcal{A}\cup\mathcal{B},\mathcal{A}\cup\mathcal{B}})$ and consequenlty with $\lambda_{min}(C)$. Specifically, for each $i \in \mathcal{A}$, let $C^{(i)}$ be the covariance of the random variables $\mathrm{Res}(X_{\mathcal{B}\cup\{i\}}, X_{\mathcal{A}\setminus\{i\}})$. Then, from Lemma 3.2, it holds

$$\lambda_{min}(C^{(i)}) \leq \lambda_{min}(C^{(i-1)}).$$

Using iteratively this inequality we obtain

$$\lambda_{min}(C_{\mathcal{B}\mathcal{B}}^{\mathcal{A}}) = \lambda_{min}(C^{(0)}) \geq \lambda_{min}(C^{(1)}) \geq \cdots \geq \lambda_{min}(C^{(|\mathcal{A}|)}) = \lambda_{min}(C_{\mathcal{A}\cup\mathcal{B},\mathcal{A}\cup\mathcal{B}}).$$

Finally, since C is a symmetric matrix by definition we can apply the Cauchy's interlacing theorem (see Proposition A.3 in Appendix A.1), and thus

$$\lambda_{min}(C_{\mathcal{A}\cup\mathcal{B},\mathcal{A}\cup\mathcal{B}}) \geq \lambda_{min}(C).$$

□

Based on Proposition 3.7, applied specifically to the optimal set $\mathcal{K}^G$ returned by the greedy algorithm, we can now state a performance bound for the greedy approximation that does not assume submodularity and depends solely on the minimum eigenvalue of the covariance matrix C. Specifically, we can restate Theorem 3.4 as follows.

Proposition 3.8. *Consider a random vector $X \in \mathbb{R}^n$ with invertible covariance matrix C such that $C_{ii} \leq 1$ for every $i = 1, \ldots, n$, and a random variable $Y = \alpha' X$ with $\alpha \in \mathbb{R}^n$. The set $\mathcal{K}^G$ returned by the greedy algorithm guarantees that*

$$F(\mathcal{K}^G) \geq \left(1 - e^{-\lambda_{min}(C)}\right) F(\mathcal{K}^*),$$

where K^* is the set maximizing $F(\mathcal{K})$ among all the possible size s sets $\mathcal{K} \subseteq \mathcal{V}$.

Remark 3.2. *It is remarkable to notice that the bound proposed in Proposition 3.8 for the performance guaranteed by Algorithm 1 for an approximate submodular function cannot overcome the one of submodularity case.*

This observation follows from some standard computations. First, since the covariance matrix C is such that $diag(C) \leq I$, its trace satisfies

$$tr(C) \leq n\,.$$

Second, by standard property of matrix eigenvalues, i.e.

$$\sum_i \lambda_i = tr(C)\,.$$

Collecting these two results the following relation holds:

$$\lambda_{min}(C) \leq \frac{\sum_i \lambda_i}{n} = \frac{tr(C)}{n} \leq 1\,.$$

This implies that the performance guaranteed using the greedy approximation for an approximate submodular function is equal to the submodular case if $\lambda_{min}(C) = 1$, otherwise the bound is strictly lower.

Note that the value of the minimum eigenvalue depends on the specific case under consideration. Particular scenarios of interest will be further analyzed in the subsequent sections.

Moreover, let us observe that if $\gamma_{\mathcal{K},s} > 1$, the corresponding performance guarantees strictly improve upon those obtained under the assumption of submodularity. This suggests that alternative approximations of the ratio $\gamma_{\mathcal{K},s}$ may lead to stronger guarantees. Although approximating $\gamma_{\mathcal{K},s}$ via $\lambda_{\min}(C)$is computationally more tractable, as highlighted in the previous remark, it may yield suboptimal performance bounds depending on the specific application. Therefore, it might be beneficial to explore alternative approximations tailored to the problem structure, potentially leading to tighter bounds and improved practical performance

CHAPTER 4

Most Informative Nodes in Opinion Dynamics Models

The optimal subset selection problem, introduced in Chapter 3, has a broad range of applications. The results presented so far are generally applicable to any scenario in which the vector X is a random variable following a given probability distribution. This includes machine learning problems such as feature selection [83], dimensionality reduction and active learning [75].

In this Chapter, we refine our focus to the specific case of Opinion Dynamics on Networks. Here, the vector X represents the opinions of a set of agents who interact over a social network according to a linear dynamical system, specifically following one of the update dynamics introduced in Section 2.3. Within this framework, the objective is to identify the individuals to survey in a poll to best estimate the overall societal opinion on a given topic. Due to this interpretation we will refer to these nodes as **"most informative"** ones.

Formally, recall that in the following sections we will consider a weighted directed graph $\mathcal{G} = (\mathcal{V}, \mathcal{E}, A)$, where $\mathcal{V}$ is the set of agents, $\mathcal{E}$ is the set of edges or interactions among agents and $A \in \mathbb{R}_+^{\mathcal{V}\times\mathcal{V}}$ is the associated influence matrix. From now on, we will assume A to be substochastic and such that $\rho(A) < 1$.

Main Contributions. We formally define the problem of interest introduced in Chapter 3 within the particular framework of Opinion Dynamics on Networks. Specifically, we focus on the role of network structure in the selection of the most informative network nodes and on a comprehensive comparison with well-known network centrality measures.

4.1 French-DeGroot Model

First, we focus on the French-DeGroot Model introduced in Section 2.3. Consider $X(t) \in \mathbb{R}^{\mathcal{V}}$ the vector of agents opinions on a certain topic at time t which updates according to the following equation

$$X(t+1) = AX(t) + BU + V(t+1), \tag{4.1}$$

where BU refers to the influence of exogeneous factors in the opinion formation process and $V(t)$ is a random noise that may stem from exogenous factors themselves or from the communication channel, e.g., a misinterpretation or distortion of information during transmission, which alters otherwise unbiased opinion updates.

More precisely, in the following discussion, we will focus on two specific cases:

- **Noisy French-DeGroot with deterministic input**. Here, U is assumed to be a deterministic vector u which takes values in $\mathbb{R}^{\mathcal{V}}$ and $V(t)$ a random vector with zero mean and covariance matrix C_V.
- **Noiseless French-DeGroot with stochastic input**. Here, $V(t) = 0$ and U is assumed to be a random vector with mean μ and covariance matrix C_U.

Shifting now the focus on the optimization problem of interest, firstly introduced in (3.3), we can formalize it in the specific context of Opinion Dynamics on Networks.

Problem 4.1. *Let X be the equilibrium opinion of the network agents, whose update dynamics equation is introduced in* (4.1)*, and consider a weighted average of networks opinions $Y = \alpha' X$ with $\alpha \in \mathbb{R}^{\mathcal{V}}_{+}$. Our goal is to select the optimal subset of agents $\widehat{\mathcal{K}} \subseteq \mathcal{V}$ to observe, among the available observable sets $\mathcal{O} \subseteq 2^{\mathcal{V}}$, to best estimate Y. Formally, we are interested in the solution of the following maximization*

problem

$$\widehat{\mathcal{K}} = \underset{\mathcal{K} \in \mathcal{O}}{\arg\max}\, F(\mathcal{K}), \tag{4.2}$$

where the variance reduction function $F(\mathcal{K})$ satisfies (3.7)*.*

4.1.1 Explicit Formulation of Variance Reduction

Building on the results from Section 2.3.1 regarding the steady state of the French-DeGroot model, as well as the theoretical findings established in Chapter 3, we can now derive an explicit formulation for the Variance Reduction function within the specific context of interest.

Noisy French-DeGroot with deterministic input

Let us start considering the noisy model with deterministic external input, formally,

$$X(t+1) = AX(t) + u + V(t+1)\,. \tag{4.3}$$

The formulation of the Variance Reduction resulting from the observation of a general subset of nodes $\mathcal{K} \subseteq \mathcal{V}$ follows directly from the combination of Theorem 3.1 and Proposition 2.4. For completeness, the detailed statement is presented in the following proposition.

Proposition 4.1. *Consider $\mathcal{G} = (\mathcal{V}, \mathcal{E}, A)$ a weighted directed network with reversible Schur-stable influence matrix A and a vector X, equilibrium vector for the dynamics in* (4.3)*. Let $u \in \mathbb{R}^{\mathcal{V}}$ be a deterministic vector and $V(t)$ a random vector with invertible covariance matrix C_V. Given arbitrary $\mathcal{K} \subseteq \mathcal{R}$, the variance reduction is computed as*

$$F(\mathcal{K}) = (C\mathbb{1})'_{\mathcal{K}} (C_{\mathcal{K}\mathcal{K}})^{-1} (C\mathbb{1})_{\mathcal{K}}\,, \tag{4.4}$$

where $C = C_V (I - A^2)^{-1}$.

We now narrow our analysis to the specific case of a single observation, i.e. $\mathcal{K} = \{k\}$. This result is remarkable, laying the foundation for a comparison with well-known centrality measures, whose comprehensive analysis is presented in Section 4.1.2.

Corollary 4.1. *Let the same assumptions of Proposition 4.1 hold. If $\mathcal{K} = \{k\}$, the variance reduction $F(\{k\})$ is computed as*

$$F(\{k\}) = \sigma_k \eta_k,$$

where $\sigma_k = \sqrt{(C_V)_{kk}}$ *and* $\eta_k := \frac{((I-A^2)^{-1}\mathbb{1})_k^2}{((I-A^2)^{-1})_{kk}}$.

Noiseless French-DeGroot with stochastic input

We now focus on the noiseless scenario with stochastic input vector, formally,

$$X(t+1) = AX(t) + BU\,. \tag{4.5}$$

Here, the formulation of the Variance Reduction resulting from the observation of a general subset of nodes $\mathcal{K} \subseteq \mathcal{V}$ follows from the combination of Theorem 3.1 and Corollary 2.2. For completeness, the detailed statement is presented in the following proposition.

Proposition 4.2. *Consider a weighted directed network $\mathcal{G} = (\mathcal{V}, \mathcal{E}, A)$ with A a Schur-stable influence matrix. Let X be the equilibrium opinion of network agents, steady state of the dynamics in* (4.5)*. Consider a matrix B taking values in $\mathbb{R}^{n\times s}$ for a certain $s \leq n = |\mathcal{V}|$ and a random vector $U \in \mathbb{R}^s$ with invertible covariance matrix C_U. Given arbitrary $\mathcal{K} \subseteq \mathcal{R}$, the variance reduction is computed as*

$$F(\mathcal{K}) = (C\mathbb{1})'_{\mathcal{K}} (C_{\mathcal{K}\mathcal{K}})^{-1} (C\mathbb{1})_{\mathcal{K}}\,, \tag{4.6}$$

where $C = (I-A)^{-1}BC_UB'(I-A')^{-1}$.

In this specific setting some extra considerations on the cardinality of the observation set $\mathcal{K}$ can be done.

Remark 4.1. *Let X be the steady state of the dynamics in* (4.5)*. If the random vector U takes values in $\mathbb{R}^s$ with $s < n$, from Corollary 2.2, the equilibrium X lies in a vector space of dimension at most s since it is computed as $X = (I-A)^{-1}BU$. Thus, if we observe a subset $\mathcal{K}$ of s linear independent components of X, it is sufficient to exactly recover the entire vector X.*

As in the previously presented case, a specific formulation arises for the single observation case and it will be useful for analyzing the connections with well-known centrality measures, as extensively discussed in Section 4.1.2.

Corollary 4.2. *Let the same assumptions of Proposition 4.2 hold. If $\mathcal{K} = \{k\}$, then given $M = (I - A)^{-1}B$, the variance reduction $F(\{k\})$ is computed as:*

$$F(\{k\}) = \frac{(MC_U v)_k^2}{c_k}, \tag{4.7}$$

where

$$v_k = \sum_i M_{ik} \quad \textit{and} \quad c_k = \sum_i M_{ki}^2 (C_U)_{ik}^2 .$$

4.1.2 Comparison with Known Centrality Measures

Let us note that computing the variance reduction function conditioned on the observation of a single node, i.e. $F(\{k\})$, is equivalent to determining an appropriate node ranking, which we refer to as a **variance reduction centrality measure**. Comparing this newly introduced centrality measure with established metrics in the literature is essential for benchmarking purposes. Such a comparison enables us to evaluate whether the new measure provides added value in terms of analysis or prediction, and to position it within the broader context of existing methodologies highlighting both its strengths and potential limitations

Bonacich and cycle centrality measures. Focusing on the formulation presented in Corollary 4.2, we highlight that variables v and c_k correspond in particular to two known centrality measures, that are respectively the Bonacich centrality defined by [60] and the cycle centrality defined by [61].

The Bonacich centrality v_k for node $k \in \mathcal{V}$ coincides with the number of all the paths in $\mathcal{G}$ that start at k, weighted down by their length: both cycles from k to k and outer paths from k to general node $i \in \mathcal{V}$.

The cycle centrality c_k of node $k \in \mathcal{V}$ coincides instead with the weighted sum of the number of network cycles the agent is in. Specifically, if we rewrite M through

the power series expansion, i.e.

$$M = \left[\sum_{l=0}^{\infty} A^l\right] B,$$

then for $B = I$,

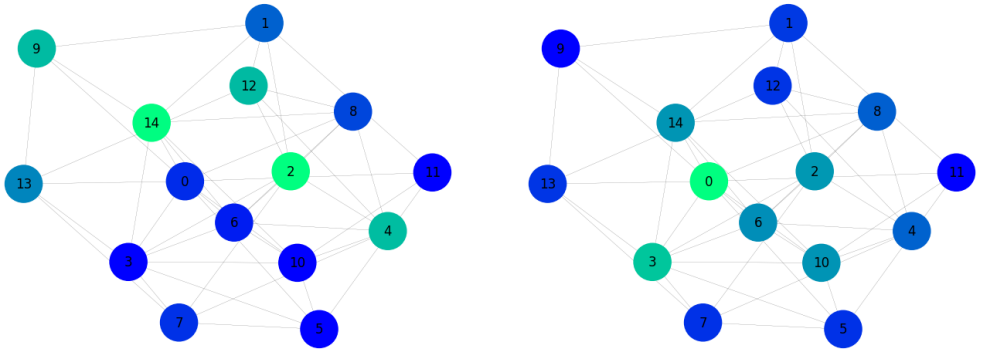
$$\begin{aligned} M_{ki}^2 &= \left(\sum_{l=0}^{\infty}\sum_{m=0}^{\infty} (A^l)_{ki}(A^m)_{ki}\right) \\ &= \left(\sum_{r=0}^{\infty}\sum_{s=0}^{r} (A^s)_{ki}(A^{r-s})_{ki}\right). \end{aligned}$$

Thus focusing on the dependence from network structure, we observe how the proposed measure $F(\mathcal{K})$, from (4.7), is directly proportional to the sum of all the paths in $\mathcal{G}$, but inverse proportional to the number of cycles.

However, our analysis highlights also how the proposed centrality measure deviates from commonly known ones. More precisely, referring to the variance reduction function $F(\{k\})$, we emphasize the distinction between our concept of the **"most informative"** node and the **"most influential"** node, which is typically defined in the literature as the node that maximizes Bonacich centrality. A preliminary proof of this distinction was first presented in our earlier work [47] in the specific case of Gaussian-distributed stubborn nodes and in the absence of additional noise. The following example further corroborates this key difference in a more general setting of interest.

Example 4.1. *Consider the Watts-Strogatz graph with $|\mathcal{V}| = 15$ and 3 stubborn nodes, represented in Fig.4.1. We compare the results of the proposed Variance Reduction measure $F(\{k\})$ with the Bonacich centrality. The colormap reflects the centrality measure for each node in the network (notice that in Figure the values are normalized). It is worth noting that our method diverges from existing approaches: the node identified as the most informative (green node in Fig.4.1a) does not coincide with the one with highest Bonacich centrality (green node in Fig.4.1b).*

Intercentrality measure. Another interesting connection of the proposed formula with well-known centrality measure can be highlighed comparing $F(\{k\})$ with the intercentrality measure, introduced by [62] for the Key Player problem (see Section



(a) Variance Reduction (b) Bonacich

Fig. 4.1 Comparison of centrality measures over a Watts-Strogatz graph. The colormap indicates the normalized Variance Reduction and the normalized Bonacich Centrality, respectively.

2.2.2 for details). For the sake of clarity, let us recall the general definition of intercentrality measure, i.e.

$$c_k = \frac{((M\mathbb{1})_k)^2}{M_{kk}}.$$

Analysing now the $F(\{k\})$ definition in Corollary 4.1 we observe that η_k coincides with c_k where $M = (I - A^2)^{-1}$.

In order to provide some more insights linked to this comparison, let us preliminary define a new graph $\hat{\mathcal{G}} = (\hat{\mathcal{V}}, \hat{\mathcal{E}})$ s.t. $\hat{\mathcal{V}} = \mathcal{V}$ and $\hat{\mathcal{E}}$ defined s.t. $(i, j) \in \hat{\mathcal{E}}$ iff it exists a path of length 2 connecting i and j in $\mathcal{G}$. $\hat{\mathcal{G}}$ is known as the the **2-hop neighbour graph** [84] associated to $\mathcal{G}$. As an example, in Fig.4.2 the 2-hop neighbour graph associated to a cycle graph with 7 nodes is represented, i.e., the links in $\hat{\mathcal{E}}$ are the one highlighted in red in Figure and connect those nodes which are at distance 2 in the original graph $\mathcal{G}$. The function η_k thus coincides with the intercentrality [62] of agent k, defined on the 2-hop neighbour graph $\hat{\mathcal{G}}$.

Remark 4.2. *The comparison between our proposed centrality measure and intercentrality is of particular interest, as it reveals a meaningful connection with 2-hop neighbors, a specific instance of higher-order networks. These structures have recently attracted growing attention within the scientific community due to their ability to capture more complex interaction patterns. This connection is especially relevant to our current research focus, out-of-the-scope of this dissertion, which explores the role of higher-order networks in modeling opinion dynamics [56]. In*

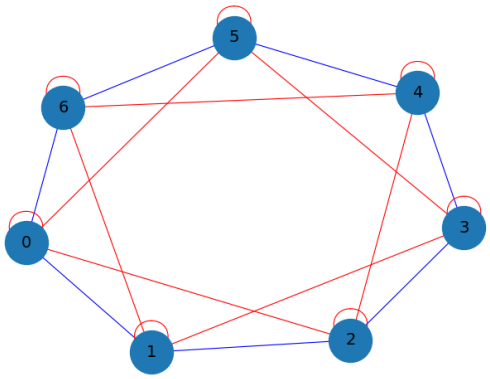


Fig. 4.2 Cycle graph with 7 nodes. The blue links represent $\mathcal{E}$, while the red one are associated to $\hat{\mathcal{E}}$.

particular, in [56] we provide an extension of classical opinion dynamics framework, showing how higher-order interactions introduce distinctive features in terms of equilibrium behavior and convergence speed. Building on this foundation, identifying and analyzing the most informative nodes, such as those highlighted by our centrality measurem, represents a natural and promising next step. The preliminary link observed in this thesis between our approach and intercentrality thus opens up several directions for future investigation.

4.1.3 Submodular Variance Reduction

From Chapter 3, we know that solving the optimization problem in (4.2) is computationally expensive, but that, under certain additional assumptions on the model parameters, we can mitigate this computational cost. Specifically, the submodularity of the function ensures good performance when using the greedy algorithm.

Thus, in this section, we will analyse the conditions that guarantee this property of the variance reduction function in the specific scenario of interest, aiming to provide meaningful interpretations in the context of network topology.

Building on the results previously proven in Section 2.3, a formal characterization can be obtained in case of Noisy French-DeGroot model with deterministic input.

Proposition 4.3. *Let X be the steady state of the dynamics in* (4.1)*, with u deterministic external input vector, A Schur-stable matrix and V random noise with zero*

mean and invertible covariance matrix C_V , i.e. it is the equilibrium of

$$X(t+1) = AX(t) + Bu + V(t+1)\,.$$

For every subset of nodes $\mathcal{K} \subseteq \mathcal{V}$, let $F(\mathcal{K})$ be the variance reduction defined in (3.7) *for the estimation of $Y = \alpha' X$ from $X_{\mathcal{K}}$, with $\alpha \in \mathbb{R}_+^{\mathcal{V}}$. If the influence matrix A is reversible, then, $F : 2^{\mathcal{V}} \to \mathbb{R}$ is a submodular function.*

Proof. The thesis follows combining the results from previous chapters:

- First, from Proposition 2.4 it is known that given A a reversible and Schur-stable matrix, u a deterministic input vector and C_V an invertible matrix, then the associated precision matrix H is such that $-H$ is a Metzler matrix;
- Second, from Theorem 3.3 if the precision matrix H satisfes $-H$ being a Metzler matrix then the variance reduction F is submodular.

□

This result is particularly significant as it ensures that greedy approximation methods can be effectively applied across a wide range of application domains. Notably, the additional assumptions required on the network topology are relatively mild, making the approach broadly applicable.

Corollary 4.3. *Let X be the steady state of the opinion dynamics in* (4.1) *with A Schur-stable matrix and V random noise with zero mean and invertible covariance C_V. If the set of external input agents, whose state is collected in the deterministic vector u, is globally reachable and the influence matrix A is reversible, then F is submodular.*

Shifting our attention to the Noiseless French-DeGroot model with stochastic input, establishing general conditions that guarantee the submodularity of function F becomes more challenging, particularly when aiming for conditions that admit a clear and interpretable formulation in terms of the network's structural properties.

Proposition 4.4. *Let X be the steady state of the opinion dynamics in* (4.1) *with stochastic input vector U, s.t. $C_U = I$, $B = I$, A Schur-stable matrix and $V(t) = 0$, i.e.*

$$X = AX + U\,.$$

If for every $i, j \in \mathcal{V}$ s.t. $i \neq j$ it holds

$$\sum_{k \in \mathcal{V}} A_{ki} A_{kj} \leq A_{ij} + A_{ji}, \tag{4.8}$$

then the variance reduction function F is submodular.

This implies that if two nodes i and j in $\mathcal{V}$ do not influence each other, then they cannot share a common ancestor k in $\mathcal{V}$.

Proof. From Section 2.3 we know that the equilibrium opinion takes the following form

$$X = (I - A)^{-1} U,$$

and that the corresponding precision matrix H is computed as

$$H = (I - A')(I - A) = I - (A + A') + A'A\,.$$

From Theorem 3.3, if the precision matrix H has non-positive off-diagonal entries, then F is submodular. Thus, it coincides with requiring that, for every $i \neq j$ in $\mathcal{V}$,

$$\begin{aligned} &(A'A)_{ij} - A_{ij} - A_{ji} \leq 0 \\ &(A'A)_{ij} \leq A_{ij} + A_{ji} \\ &\textstyle\sum_{k \in \mathcal{V}} A_{ki} A_{kj} \leq A_{ij} + A_{ji} \end{aligned}$$

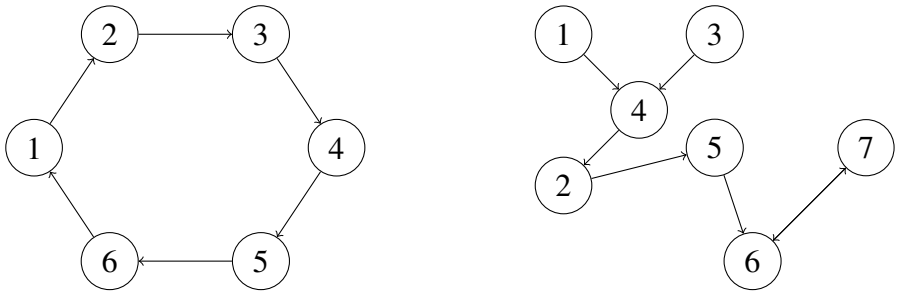
The thesis follows. □

Example 4.2. *For the sake of clarity, we present here a simple example in which condition* (4.8) *holds. Let us consider a directed weighted cycle graph like the one in Fig.4.3a. The proof follows immediately noticing that*

$$0 = \sum_{k \in \mathcal{V}} A_{ki} A_{kj} < A_{ij} + A_{ji} = A_{ij}\,.$$

Analogously, we can observe that the condition also holds for an arbitrary directed graph in which each node has out-degree at most 1, i.e. $w_i^+ \leq 1$ for all i in $\mathcal{V}$, and in-degree at least 1, i.e. $w_i^- \geq 1$ for all i in $\mathcal{V}$. An example of such a graph is shown in Fig. 4.3b. A notable class of networks satisfying these properties is

represented by anti-arborescence graphs, namely those whose edge-reversed graph is an arborescence[1].



(a) Directed cycle graph with 6 nodes. (b) Directed acyclic graph with 7 nodes.

Fig. 4.3 Examples of graph structures which satisfy the submodularity condition in (4.8).

In the general case, if the condition (4.8) on the influence matrix A stated in the previous proposition is not satisfied, we cannot draw general conclusions about the submodularity of the function; while each scenario may need to be analyzed individually. We now present a simple yet insightful example in which explicit computations are tractable. In this setting, if the assumption on A is violated, the function F fails to be submodular. Therefore, within this specific framework, the previously stated condition on A is not only sufficient but also necessary for submodularity.

Example 4.3. *Let us consider a network $\mathcal{G}$ where there exists a couple of nodes which are not directly connected, i.e. $A_{ij} = 0$, but sharing a common neighbour $k \in \mathcal{V}\backslash\{i,j\}$, i.e. $A_{ki} \neq 0$ and $A_{kj} \neq 0$. In this specific setting, inequality* (4.8) *is violated, thus Proposition 4.4 does not provide us any additional information about the submodularity of the function.*

Building on precision matrix H definition, we derive that

$$H_{ij} = \sum_{k \in \mathcal{V}} A_{ki} A_{kj} > 0\,.$$

An example is shown in Fig.4.4 where the positive edge is highlighted in red. Moreover, explicitely computing the variance reduction conditioned to the observation of

[1] A directed graph $\mathcal{G}$ is an arborescence if it is a rooted directed tree in which all edges point away from the root, and there exists a unique directed path from the root to any other node.

set $\mathcal{K}$ and of set $\mathcal{K} \cup \{k\}$ we retrieve that

$$Var(X_i|X_{\mathcal{K}\cup\{k\}}) - Var(X_i|X_{\mathcal{K}}) = \frac{H_{kj}}{2H_{ki}H_{ij}} > 0$$

This implies that it exists at least one subset $\mathcal{K} \subseteq \mathcal{V}$ such that the the incremental differences are increasing. From submodularity definition (see Definition 3.1 for details) this implies that F fails to be a submodular function.

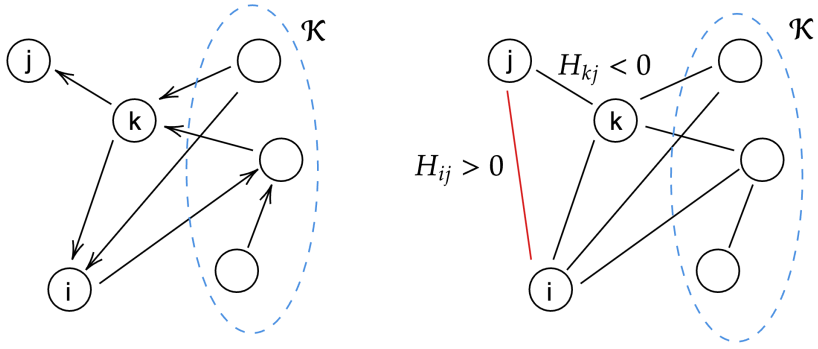


Fig. 4.4 Arbitrary graph where H fails to have non-positive off-diagonal entries. The corresponding positive edge is highlighted in red. In the left figure is represented the graph associated to A, on the right the one associated to the precision matrix H.

Finally, in terms of performance guarantees, as known from previous chapter, the best bound that we can achieve, when F is submodular, is the one proposed by Nemhauser, i.e.

$$F(\mathcal{K}^G) \geq \left(1 - \frac{1}{e}\right) \max_{\mathcal{K}\subseteq\mathcal{R}:|\mathcal{K}|=s} F(\mathcal{K}),$$

where $\mathcal{K}^G$ is the solution proposed by the greedy approximation.

4.1.4 Approximate Submodular Variance Reduction

As widely explained in Chapter 3, even when the problem fails to be submodular, approximate submodularity remains applicable and the following performance for the greedy approximation is guaranteed:

$$F(\mathcal{K}^G) \geq \left(1 - e^{-\lambda_{min}(\bar{C})}\right) \max_{\mathcal{K}\subseteq\mathcal{R}:|\mathcal{K}|=s} F(\mathcal{K}), \tag{4.9}$$

where $\bar{C}$ is the covariance matrix rescaled such that $\bar{C}_{ii} \leq 1$ for all $i \in \mathcal{V}$, i.e.

$$\bar{C} = \left[\frac{1}{\max_i C_{ii}}\right] C\,.$$

Let us now focus on the Noiseless French-DeGroot with Stochastic Input, under the assumptions of Proposition 4.4, i.e. $B = I$, $C_U = I$. Within this setting we can find a lower-bound for the performance guarantees which depends on the network influence matrix A.

Proposition 4.5. *Let X be the steady state of the opinion dynamics in* (4.1) *with stochastic input vector U, s.t. $C_U = I$, $B = I$, A Schur-stable and $V(t) = 0$, i.e.*

$$X = AX + U\,.$$

The set $\mathcal{K}^G$ returned by the greedy algorithm guarantees that

$$F(\mathcal{K}^G) \geq \left(1 - e^{-\frac{(1-\lambda_{min}(A))^2}{\alpha}}\right) F(\mathcal{K}^*),$$

where K^* is the set maximizing $F(\mathcal{K})$ among all the possible size s sets $\mathcal{K} \subseteq \mathcal{Z}$ and $\alpha = \max_{i\in\mathcal{V}} C_{ii}$, with C covariance matrix of C defined by $C = (I-A)^{-1}(I-A')^{-1}$.

Proof. Based on (4.9) we will look now for an explicit formulation for $\lambda_{min}(\bar{C})$. For the clarity of notation, let us define $\alpha = \max_i C_{ii}$.

$$\begin{aligned}
\lambda_{min}(\bar{C}) &= \lambda_{min}\left(\left[\tfrac{1}{\alpha}\right] C\right) \\
&= \tfrac{1}{\alpha}\lambda_{min}(C) \\
&= \tfrac{1}{\alpha}\lambda_{min}(\left((I-A)^{-1}(I-A')^{-1}\right)) \\
&\geq \tfrac{1}{\alpha}\lambda_{min}((I-A)^{-1})\lambda_{min}((I-A')^{-1}) \\
&= \tfrac{1}{\alpha}\lambda_{max}(I-A)\lambda_{max}(I-A') \\
&= \tfrac{1}{\alpha}(1-\lambda_{min}(A))(1-\lambda_{min}(A')) \\
&= \tfrac{1}{\alpha}(1-\lambda_{min}(A))^2
\end{aligned}$$

The thesis follows substituting this lower bound for $\lambda_{min}(\bar{C})$ in (4.9). □

Building on this result, we can provide additional insights related to the network structure.

Remark 4.3. *The performance of the greedy algorithm improves for lower values of minimum eigenvalue of influence matrix A, i.e. as $\lambda_{min}(A)$ approaches to 0. To prove this let us analyse when $\lambda_{min}(\bar{C})$ approaches 1 given that, from Remark 3.2, we know that we cannot overcome the performance achieved in case of function submodularity since $\lambda_{min}(\bar{C}) \leq 1$.*

First, notice that $\alpha > 1$ since

$$\alpha = \max_i C_{ii} = \max_i \left((I-A)^{-1}(I-A')^{-1}\right)_{ii} = \max_i \left(\sum_{k\geq 0}(A^k)\sum_{k\geq 0}(A')^k\right)_{ii} > 1$$

Consequently, $\lambda_{min}(\bar{C})$ takes higher values as $(1-\lambda_{min}(A))^2$ does. This coincides with requiring that $\lambda_{min}(A)$ approaches 0 since $\lambda_{min}(A)$ cannot take negative values due to the assumptions of the model on the influence matrix A.

In a social network this condition on $\lambda_{min}(A)$ is satisfied, for instance, in case of multiple communities that are weakly connected. This reflects the practical relevance of the proposed algorithm, which can be effectively applied across a variety of real-world scenarios.

4.1.5 Examples and Simulations

We now present some Python simulations designed to support the theoretical results discussed above. Specifically, we consider the following setting. Let $\mathcal{G} = (\mathcal{V}, \mathcal{E}, P)$ be a graph with P the normalized weight matrix. Assume the individuals to be partitioned in two classes: regular agents $\mathcal{R}$ and stubborn agents $\mathcal{S}$, such that $\mathcal{V} = \mathcal{R} \cup \mathcal{S}$, and assume that $\mathcal{S}$ forms a globally reachable set in $\mathcal{G}$. Let $X(t)$ indicates the opinion vector of the regular agents at time t, which evolves according to the following dynamics, as introduced in previous sections:

$$X(t+1) = AX(t) + BU + V(t+1),$$

where U is the vector of stubborn agents opinions, $A := P_{\mathcal{RR}}$ captures the interactions among regular agents and $B := P_{\mathcal{RS}}$ represents the influence of stubborn nodes on regular ones.

We will now explore a set of illustrative examples, each capturing one of the key scenarios previously analyzed from a theoretical perspective.

Noisy French-DeGroot with Deterministic Input

Let us consider a noisy French-DeGroot opinion dynamics model with deterministic input, i.e.

$$X(t+1) = AX(t) + Bu + V(t+1),$$

where $V(t)$ is a random noise with covariance matrix C_V. The deterministic input vector u coincides in particular with the deterministic opinion of stubborn agents.

In our simulations, we assume $C_V = [\sigma]$, with $\sigma = 0.3$. Moreover, t he influence matrix A is assumed to satisfy the assumptions of Proposition 4.1, i.e. A reversible and Schur-stable matrix.

Within this framework, we assess the performance of the greedy solution against the exact one, considering both a randomly generated Watts-Strogatz network and a real-world case study. The results of these simulations were presented in [48].

Example 4.4 (Watts-Strogatz Network)**.** *Let us consider a randomly generated Watts-Strogatz network with 15 nodes, 3 of which are stubborn. Suppose that we are interested in observing up to 4 nodes and compare the results obtained both using the exact computation from equation* (3.7) *and applying the Greedy Algorithm. In Fig.4.5 is represented the percentage of residual variance conditioned to the observation of subset $\mathcal{K}$. We highlight how the Greedy Algorithm achieves comparable performances in terms of residual variance while significantly reducing the computational cost. To be thorough, in Fig.4.6 it is reported the proposed selection for a 4 nodes observation set both in case of exact computation and Greedy Algorithm application.*

Example 4.5 (Real-case scenario)**.** *Consider the real- world case study of a network associated with social contacts in a village in rural Malawi, whose dataset is available on Sociopatterns [85]. Our analysis focuses on the network's largest connected component, which consists of 84 individuals and 346 weighted, undirected edges, as illustrated in Fig. 4.7. Among them we assume that 10 individuals, randomly selected, are stubborn agents. While the greedy algorithm exhibits lower performance, as shown in Fig.4.8, its lower computational cost makes it feasible for tackling problem's dimensionality.*

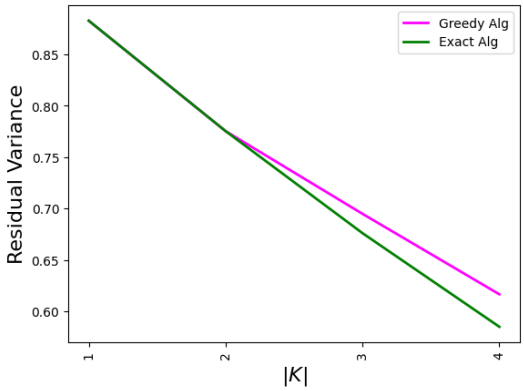
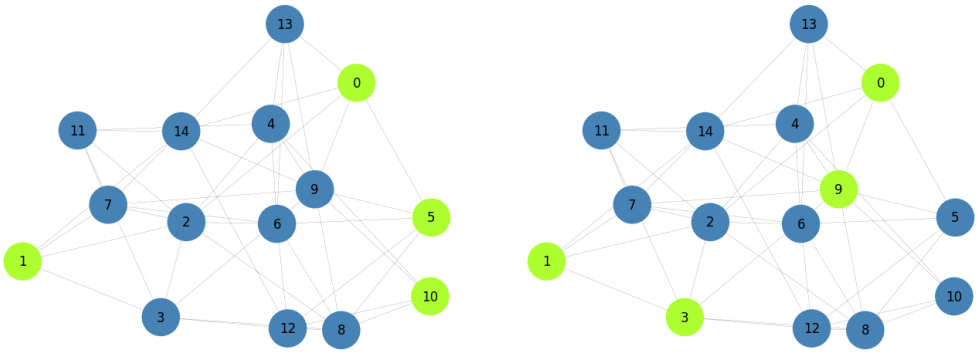


Fig. 4.5 Comparison between the percentage of residual variance as function of the cardinality of the observed set $\mathcal{K}$, chosen through exact method vs Greedy Algorithm, for a randomly generated Watts-Strogatz network with 15 nodes.



(a) Exact Selection (b) Greedy Selection

Fig. 4.6 Comparison between the candidate set $\mathcal{K}$ chosen through exact method vs Greedy Algorithm for a randomly generated Watts-Strogatz network with 15 nodes, for $|\mathcal{K}| = 4$.

Noiseless French-DeGroot with Stochastic Input

We consider a noiseless French-DeGroot opinion dynamics with stochastic input, i.e.

$$X(t+1) = AX(t) + BU,$$

where U is a random variable with mean μ and covariance matrix C_U. The input vector U represents the opinions of the stubborn agents, which are here assumed to be stochastic. This assumption reflects the idea that the exact opinions of these agents on the topic of interest are not known with certainty.

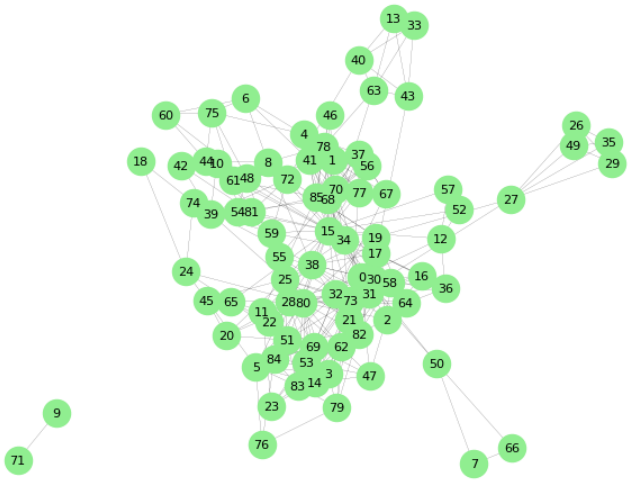


Fig. 4.7 Graphical representation of the Malawi real-world network.

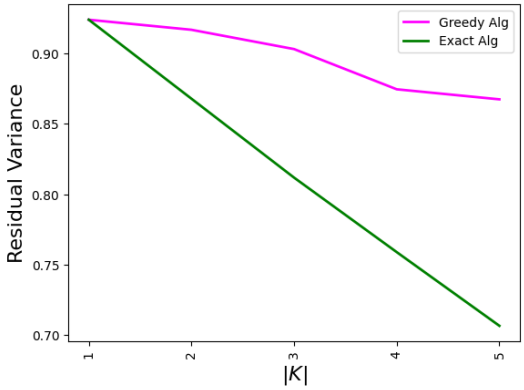


Fig. 4.8 Comparison between the percentage of residual variance as function of the cardinality of the observed set $\mathcal{K}$, chosen through exact method vs Greedy Algorithm, for the real world Malawi network.

The goal remains to estimate the average societal opinion, which, under the assumptions of Proposition 4.2, is expressed as a linear combination of U. The following example aims to show that the choice between observing a regular agent or a stubborn one depends on the specific context of application.

Example 4.6. *Take a simple graph of 3 nodes with two stubborn nodes* $S = \{1,3\}$ *and one regular node* $R = \{2\}$. *Let the matrix of interactions between stubborn and regular nodes be defined by* $B = [\alpha\ \beta]$. *The opinion of the stubborn nodes is*

unknown and it is so described by the following random variable

$$U \sim \mathcal{N}(\mu, C_U) \quad \textit{with} \quad C_U = \begin{bmatrix} \sigma_1^2 & 0 \\ 0 & \sigma_3^2 \end{bmatrix}.$$

Suppose we can observe only one node. Under those assumptions we can explicitely compute the variance reduction function.

$$\begin{aligned} F(1) &= \sigma_1^2(\alpha+1) \\ F(3) &= \sigma_3^2(\beta+1) \\ F(2) &= \alpha(\alpha+2)\sigma_1^2 + \beta(\beta+2)\sigma_3^2 + \frac{(\alpha\sigma_1^2+\beta\sigma_3^2)^2}{\alpha^2\sigma_1^2+\beta^2\sigma_3^2} \end{aligned}$$

We retrieve so that the optimal choice depends on the value of the model parameters. In particular, surely if $\alpha > \frac{-1+\sqrt{5}}{2}$ *it is more convinient to observe the regular node rather than the stubborn ones.*

4.2 Friedkin-Johnsen Model

Let us now turn our attention to the second key opinion dynamics model, the Friedkin-Johnsen model. First, let us recall the setting we are working in. As in the previous section, we consider a network of agents $\mathcal{V}$ with interactions matrix A in $\mathbb{R}_+^{\mathcal{V}\times\mathcal{V}}$. Now, every agent i is equipped with an original belief U_i and a parameter λ_i in $[0,1]$ describing its openness to social interaction. The original belief U is unknown and it is modeled as a random variable with mean μ and covariance matrix C_U. According to the Friedkin-Johnsen model, the equilibrium opinion of the social system is represented by a vector X in $\mathbb{R}^{\mathcal{V}}$ such that

$$X = (I - [\lambda]A)^{-1}(I - [\lambda])U\,. \tag{4.10}$$

Interpretatively, higher values of λ indicate a dominance of network interactions in shaping opinions, whereas lower values reflect a stronger influence of individual prior beliefs or prejudices. Let us notice that it can be seen as a generalization of the Noiseless French-DeGroot model with Stochastic input and previous results apply also in this particular case study.

As done for the French-DeGroot model, for the sake of completeness, let us define the optimization problem in this particular setting of interest.

Problem 4.2. *Let X be the equilibrium opinion defined by* (4.10). *Let $Y = \alpha' X$ be a weighted average of networks opinions, with $\alpha \in \mathbb{R}_+^{\mathcal{V}}$. Our goal is to select the optimal subset $\widehat{\mathcal{K}} \subseteq \mathcal{V}$ to observe, among the available observable sets $\mathcal{O} \subseteq 2^{\mathcal{V}}$, to best estimate Y. Formally, we are interested in the solution of the following maximization problem*

$$\widehat{\mathcal{K}} = \underset{\mathcal{K} \in \mathcal{O}}{\arg\max}\, F(\mathcal{K}), \tag{4.11}$$

where the variance reduction function $F(\mathcal{K})$ satisfies (3.7).

In this section we will more precisely focus on the role of model parameter λ in the optimal set selection. As will be thoroughly discussed in the following, a key finding of our analysis is that the optimal node selection depends explicitly on this parameter and this crucial insight differentiates our approach from traditional centrality measures, such as Bonacich centrality, where no such dependency arises. This fundamental distinction underscores the significant advantage of our proposed centrality measure, demonstrating its ability to capture important structural effects that existing measures overlook.

Extension to different context of application

Let us observe that, in wider terms, this analysis can be extended to quadratics games on networks [86]. Specifically, consider the following utility function:

$$u_i(x) = a_i x_i - \frac{1}{2}x_i^2 + \lambda_i x_i \sum_j A_{ij} x_j$$

where $a_i x_i - \frac{1}{2}x_i^2$ represents the utility of agent i without social interactions, with a_i the marginal reward and $-\frac{1}{2}x_i^2$ corresponding cost. In the Nash Equilibrium each player chooses an opinion in order to minimize his cost, which in terms of cost function derivatives coincides with requiring $u_i'(x) = 0$ for all i in $\mathcal{V}$. Thus, in order to find the opinions in the Nash equiibrium we should solve the following system:

$$a_i - x_i + \lambda_i \sum_j A_{ij} x_j = 0 \quad \forall i \in \mathcal{V}\,. \tag{4.12}$$

As in Friedkin-Johnsen model, the parameter λ_i captures the strenght of the peer effect and A is the adjacency matrix of the graph. Assuming also in this case that $[\lambda]A$ is a Schur-stable matrix, then the unique Nash equilibrium solution of (4.12) is defined by

$$x = (I - [\lambda]A)^{-1}U \quad \text{where } U_i = a_i\,.$$

4.2.1 Limit Cases. No Interaction vs Full Interaction

First, we analyse two extreme cases which occur when the network doesn't influence at all the dynamics ($\lambda = 0$) or conversely when the agents are not conditioned by their prior beliefs ($\lambda = \mathbb{1}$).

No interaction ($\lambda = 0$). In this scenario, agents rely entirely on their initial opinions, and no information exchange occurs within the network. The system reduces to a trivial case where each agent's final opinion is solely determined by their intrinsic bias, independent of the network structure.

Given the measure of interest $Y = \alpha' X = \alpha' U$, let C_U be the covariance matrix of random vector U defined s.t. $(C_U)_{ii} = \sigma_i^2$. Thus, the variance reduction function

takes the following form:

$$F(\mathcal{K}) = \sum_{i \in \mathcal{K}} \alpha_i^2 \sigma_i^2$$

In that particular case the greedy solution coincides with solving the exact optimization problem and the optimal strategy is to select the set of s nodes with highest weighted variance.

Full interaction ($\lambda = \mathbb{1}$). Here, the dynamics depend only on the network structure and the problem coincides with a classical French-DeGroot model

$$X_i(t+1) = \sum_{j \in \mathcal{V}} A_{ij} X_j(t)\,.$$

Here, as described in Section 2.3, if the graph is strongly connected, then the model converges to a consensus. Thus, the observation of one single arbitrary node provides complete knowledge over the network.

4.2.2 Low-Interaction Rate Scenario

In this section, we focus on the case $\lambda \to 0$, which refers to those scenarios where the interactions among agents have minimal influence on the opinion formation process. This includes situations where individuals strongly adhere to their initial beliefs or where the network structure exhibits weak connectivity. For this analysis we will assume all the agents to have the same openness to interaction, i.e. $\lambda_i = \lambda$ for all i in $\mathcal{V}$, and we will indicate $\Lambda := [\lambda]$.

Since the variance reduction function is computationally expensive to evaluate, our goal is to determine whether, under the assumption of low interactions, i.e. λ approaches to 0, we can identify analytical criteria that facilitate the selection of an optimal subset of nodes to observe. To this end, we employ a second-order Maclaurin expansion of the function of interest $F(\mathcal{K})$ to derive criteria that guide the optimal node selection, improving computational efficiency without compromising result accuracy. In this analysis we truncate the expansion at second order, under the assumption that the probability of two candidate sets exhibiting identical Taylor expansions up to this order is negligible. However, the framework can be systemati-

cally extended to include higher-order terms, thereby enabling the discrimination between such degenerate cases when necessary.

Theorem 4.1. *Consider X equilibrium opinion of a Friedkin-Johnsen model, defined as in* (4.10)*, with U random vector with diagonal covariance matrix C_U such that $(C_U)_{ii} = \sigma_i^2$. Let $Y = \alpha' X$ with $\alpha \in \mathbb{R}^n$. Given $F(\mathcal{K})$ the variance reduction subjected to observation $\mathcal{K}$, in case of low interaction, i.e. when λ approaches to* 0*, the following second-order McLaurin expansion holds:*

$$F(\mathcal{K}) = F_0^0(\mathcal{K}) + \lambda F_0^1(\mathcal{K}) + \lambda^2 F_0^2(\mathcal{K}) + o(\lambda^2)$$

where

$$\begin{aligned}
F_0^0(\mathcal{K}) &= \textstyle\sum_{k\in\mathcal{K}} \alpha_k^2 \sigma_k^2 \\
F_0^1(\mathcal{K}) &= \textstyle\sum_{i\in\mathcal{K}} \sum_{j\in\mathcal{V}} \alpha_i \alpha_j \sigma_j^2 A_{ij} + \sum_{i\in\mathcal{K}} \sum_{j\in\mathcal{V}\setminus\mathcal{K}} \alpha_i \alpha_j \sigma_i^2 A_{ji} \\
F_0^2(\mathcal{K}) = \; & 3(\alpha'_{\mathcal{K}}[A_{\mathcal{K}\mathcal{K}}][C_U\alpha]_{\mathcal{K}} - \alpha'_{\mathcal{K}}[C_U A'\alpha]_{\mathcal{K}}) + 4\alpha'_{\mathcal{K}}[A^2 C_U\alpha]_{\mathcal{K}} + [C_U\alpha]'_{\mathcal{K}}[(A')^2 \\
& + (A^2)' + (A' {C_U}^{-1} A C_U)]_{\mathcal{K}\mathcal{K}}\alpha_{\mathcal{K}} + 2[AC_U A'\alpha]'_{\mathcal{K}}\alpha_{\mathcal{K}} \\
& + ([\alpha' A]_{\mathcal{K}} - ((C_U^{-1} A C_U)_{\mathcal{K}\mathcal{K}}\alpha_{\mathcal{K}})')([AC_U\alpha]_{\mathcal{K}} + [(AC_U)'\alpha]_{\mathcal{K}}) \\
& - 2([AC_U\alpha]'_{\mathcal{K}} + [C_U A'\alpha]'_{\mathcal{K}})[A']_{\mathcal{K}\mathcal{K}}\alpha_{\mathcal{K}}
\end{aligned}$$

Proof. Given X defined as in (4.10), let $M := (I - \Lambda A)^{-1}(I - \Lambda)$.

First, note that using power series expansion it holds

$$M = \sum_{k=0}^{+\infty} (\Lambda A)^k (I - \Lambda) = (I - \Lambda)(1 + \Lambda A + \Lambda^2 A^2 + o(\Lambda^2)) .$$

From (3.7), recall that the optimization function of interest $F(\mathcal{K})$ is computed as

$$F(\mathcal{K}) = (C\alpha)'_{\mathcal{K}} (C_{\mathcal{K}\mathcal{K}})^{-1} (C\alpha)_{\mathcal{K}} .$$

Thus, let's compute the McLaurin expansion of all these factors.

$$\begin{aligned}(C\alpha)_{\mathcal{K}} &= (MC_UM'\alpha)_{\mathcal{K}}\\ &= (1-\lambda)^2[(I+\Lambda A+\Lambda^2A^2+o(\lambda^2))C_U(I+\Lambda A+\Lambda^2A^2+o(\lambda^2))'\alpha]_{\mathcal{K}}\\ &= (C_U)_{\mathcal{K}\mathcal{K}}[\alpha_{\mathcal{K}}+\lambda((C_U{}^{-1}AC_U-2I+A')\alpha)_{\mathcal{K}}+\\ &\quad+\lambda^2((C_U{}^{-1}A^2C_U-2C_U{}^{-1}AC_U+(A^2)'-2A'+C_U{}^{-1}AC_UA'+I)\alpha)_{\mathcal{K}}\\ &\quad+o(\lambda^2)]\end{aligned}$$

$$\begin{aligned}C_{\mathcal{K}\mathcal{K}} &= (MC_UM')_{\mathcal{K}\mathcal{K}}\\ &= (C_U)_{\mathcal{K}\mathcal{K}}[I+\lambda(C_U{}^{-1}AC_U-2I+A')+\\ &\quad+\lambda^2(C_U{}^{-1}A^2C_U-2C_U{}^{-1}AC_U+(A^2)'-2A'+C_U{}^{-1}AC_UA'+I)\\ &\quad+o(\lambda^2)]_{\mathcal{K}\mathcal{K}}\end{aligned}$$

$$\begin{aligned}(C_{\mathcal{K}\mathcal{K}})^{-1} &= [I_{\mathcal{K}\mathcal{K}}-\lambda(C_U{}^{-1}AC_U-2I+A')_{\mathcal{K}\mathcal{K}}\\ &\quad+\lambda^2(3I+(A')^2+A'C_U{}^{-1}AC_U-2A'-(A^2)')_{\mathcal{K}\mathcal{K}}\\ &\quad+o(\lambda^2)]((C_U)_{\mathcal{K}\mathcal{K}})^{-1}\end{aligned}$$

Collecting all these results, we can write the expansion of the overall function:

$$\begin{aligned}F(\mathcal{K}) &= (C\alpha)'_{\mathcal{K}}(C_{\mathcal{K}\mathcal{K}})^{-1}(C\alpha)_{\mathcal{K}}\\ &= \alpha'_{\mathcal{K}}(C_U)_{\mathcal{K}\mathcal{K}}\alpha_{\mathcal{K}}\\ &\quad+\lambda[((C_U{}^{-1}AC_U-2I+A')\alpha)'_{\mathcal{K}}(C_U)_{\mathcal{K}\mathcal{K}}\alpha_{\mathcal{K}}\\ &\qquad+\alpha'_{\mathcal{K}}(C_U)_{\mathcal{K}\mathcal{K}}((C_U{}^{-1}AC_U-2I+A')\alpha)_{\mathcal{K}}\\ &\qquad-\alpha'_{\mathcal{K}}(C_U)_{\mathcal{K}\mathcal{K}}(C_U{}^{-1}AC_U-2I+A')_{\mathcal{K}\mathcal{K}}\alpha_{\mathcal{K}}]\\ &\quad+\lambda^2[((C_U{}^{-1}A^2C_U-2C_U{}^{-1}AC_U+(A^2)'-2A'+C_U{}^{-1}AC_UA'+I)\alpha)'_{\mathcal{K}}\cdot\\ &\qquad\cdot(C_U)_{\mathcal{K}\mathcal{K}}\alpha_{\mathcal{K}}+\alpha'_{\mathcal{K}}(C_U)_{\mathcal{K}\mathcal{K}}(3I+(A')^2+A'C_U{}^{-1}AC_U-2A'-(A^2)')_{\mathcal{K}\mathcal{K}}\alpha_{\mathcal{K}}\\ &\qquad+\alpha'_{\mathcal{K}}(C_U)_{\mathcal{K}\mathcal{K}}((C_U{}^{-1}A^2C_U-2C_U{}^{-1}AC_U+(A^2)'-2A'+C_U{}^{-1}AC_UA'+I)\alpha)_{\mathcal{K}}\\ &\qquad-((C_U{}^{-1}AC_U-2I+A')\alpha)'_{\mathcal{K}}(C_U)_{\mathcal{K}\mathcal{K}}(C_U{}^{-1}AC_U-2I+A')_{\mathcal{K}\mathcal{K}}\alpha_{\mathcal{K}}\\ &\qquad+((C_U{}^{-1}AC_U-2I+A')\alpha_{\mathcal{K}})'(C_U)_{\mathcal{K}\mathcal{K}}((C_U{}^{-1}AC_U-2I+A')\alpha)_{\mathcal{K}}\\ &\qquad-\alpha'_{\mathcal{K}}(C_U)_{\mathcal{K}\mathcal{K}}(C_U{}^{-1}AC_U-2I+A')_{\mathcal{K}\mathcal{K}}((C_U{}^{-1}AC_U-2I+A')\alpha)_{\mathcal{K}}]\\ &\quad+o(\lambda^2)\,.\end{aligned}$$

The thesis follows. □

Building on the coefficients of the McLaurin expansion presented in Theorem 4.1, we now present a methodology to identify the optimal subset of nodes to observe in the Friedkin–Johnsen model (4.10) under low-interaction regime.

Algorithm 4.1 (**General $\mathcal{K}$ Selection for Low-Interaction**). *Let X be the equilibrium opinion vector of the Friedkin-Johnsen model* (4.10)*, where U is a random vector with diagonal covariance matrix C_U such that $(C_U)_{ii} = \sigma_i^2$. Let $Y = \alpha' X$ with $\alpha \in \mathbb{R}^n$. In the low-interaction regime, for any cardinality of the observation set $\mathcal{K}$, we proceed as follows:*

1. ***First criterion***. *Choose the set of nodes $\mathcal{K}$ with highest weighted total variance:*

$$F_0^0(\mathcal{K}) = \sum_{k\in\mathcal{K}} \alpha_k^2 \sigma_k^2$$

2. ***Second criterion***. *If two or more candidate optimal sets have the same total variance, then choose the one which maximizes:*

$$F_0^1(\mathcal{K}) = \sum_{i\in\mathcal{K}} \sum_{j\in\mathcal{V}} \alpha_i \alpha_j \sigma_j^2 A_{ij} + \sum_{i\in\mathcal{K}} \sum_{j\in\mathcal{V}\setminus\mathcal{K}} \alpha_i \alpha_j \sigma_i^2 A_{ji}$$

   where the first term measures the weighted out-degree from $\mathcal{K}$ to $\mathcal{V}$ and the second one the weighted in-degree from $\mathcal{V}\setminus\mathcal{K}$ to $\mathcal{K}$.

3. ***Third criterion***. *If there are two candidate solutions $\mathcal{K}_1$ and $\mathcal{K}_2$ such that $F_0^0(\mathcal{K}_1) = F_0^0(\mathcal{K}_2)$ and $F_0^1(\mathcal{K}_1) = F_0^1(\mathcal{K}_2)$, then we choose the set which maximizes the second order coefficient $F_0^2(\mathcal{K})$ that is a combination of node degrees and paths of length 2. Refer to Theorem 4.1 for the complete expression.*

Note that when Y is the network average opinion, i.e. $\alpha = \frac{\mathbb{1}}{n}$, and all the nodes have indentical variance, the previously presented criteria can be simplified yielding some extra insights into which agents are most critical to observed.

Corollary 4.4. *Let $Y = \frac{X\mathbb{1}}{n}$ and assume that $\sigma_i^2 = \sigma^2$ for all the nodes in $\mathcal{V}$.*

1. *$F_0^0(\mathcal{K})$ is a constant. The first criterion is not informative.*

2. *The first-order expansion coefficient is simplified as*

$$F_0^1(\mathcal{K}) = \sum_{i\in\mathcal{K}} \sum_{j\in\mathcal{V}\setminus\mathcal{K}} A_{ji}\,.$$

   It corresponds to the total in-degree of the set $\mathcal{K}$ from the nodes in $\mathcal{V}\setminus\mathcal{K}$, i.e.the sum of the weights of all incoming edges to $\mathcal{K}$ originating from outside the set.

3. *The second-order expansion coefficient is simplified as*

$$F_0^2(\mathcal{K}) = 2((A^2)'\mathbb{1})'_{\mathcal{K}}\mathbb{1}_{\mathcal{K}} + 2(AA'\mathbb{1})'_{\mathcal{K}}\mathbb{1}_{\mathcal{K}} + \mathbb{1}'_{\mathcal{K}}(A'A)_{\mathcal{K}\mathcal{K}}\mathbb{1}_{\mathcal{K}} - 2(A'\mathbb{1})'_{\mathcal{K}}(A_{\mathcal{K}\mathcal{K}} + A'_{\mathcal{K}\mathcal{K}})\mathbb{1}_{\mathcal{K}}$$

where:

- $2((A^2)'\mathbb{1})'_{\mathcal{K}}\mathbb{1}_{\mathcal{K}}$ *is the sum of the* $\mathcal{K}$ *columns of* $(A^2)'$
- $2(AA'\mathbb{1})'_{\mathcal{K}}\mathbb{1}_{\mathcal{K}} = 2\sum_{i\in\mathcal{K}}\sum_{j,t} A_{it}A_{jt}$ *is the sum of the* $\mathcal{K}$ *rows of* AA'
- $\mathbb{1}'_{\mathcal{K}}(A'A)_{\mathcal{K}\mathcal{K}}\mathbb{1}_{\mathcal{K}} = \sum_{i,j\in\mathcal{K}}\sum_t A_{ti}A_{tj}$
- $-2(A'\mathbb{1})'_{\mathcal{K}}(A_{\mathcal{K}\mathcal{K}} + A'_{\mathcal{K}\mathcal{K}})\mathbb{1}_{\mathcal{K}} = -2(\sum_{i,j\in\mathcal{K}}(A_{ij} + A_{ji})(\sum_t A_{tj}))$

Let us now present a simplified version of Algorithm 4.1 is case of one single observation, i.e. $\mathcal{K} = \{k\}$ when Y is defined as the network average opinion.

Algorithm 4.2 (**Single Node Selection for Low-Interaction**). *Let* $Y = \frac{X\mathbb{1}}{n}$. *If* $\mathcal{K} = \{k\}$, *the selection method is simplified as follows:*

1. ***First criterion***. *Choose the node k with highest variance:*

$$F_0^0(k) = \sigma_k^2$$

2. ***Second criterion***. *If two or more candidate optimal nodes have the same variance, then choose the one with highest rescaled degree:*

$$F_0^1(k) = 2\sigma_k^2\sum_i A_{ik} + 2\sum_i \sigma_i^2 A_{ki}$$

If all the nodes have the same variance, it coincides with choosing the node with highest in-degree:

$$F_0^1(k) = \sum_i A_{ik}$$

3. ***Third criterion***. *If there are two candidate solutions* k_1 *and* k_2 *such that* $F_0^0(k_1) = F_0^0(k_2)$ *and* $F_0^1(k_1) = F_0^1(k_2)$, *then we choose the one which maximizes the second order coefficient that is a combination of node degrees and paths of length 2:*

$$F_0^2(k) = 2\sigma_k^2\sum_{i,j\neq k} A_{ij}A_{jk} + 2\sum_{i,j}\sigma_i^2 A_{kj}A_{ji} + 2\sum_{i,j\neq k}\sigma_i^2 A_{ki}A_{ji} + \sum_i \sigma_i^2 (A_{ki})^2$$

Where

- $A_{ij}A_{jk}$ *represents the paths of length 2 which start in i and end in k, i.e. it indicates those nodes which are influenced by node k through another intermediate node;*
- $A_{kj}A_{ji}$ *represents the paths of length 2 from k to i, i.e. the nodes which influence node k through another intermediate node;*
- $A_{ki}A_{ji}$ *represents those nodes which influence node k and another node j at the same time.*

If all the nodes have the same variance, the criterion is reduced to:

$$F_0^2(k) = 2\sum_{i,j\neq k} A_{ij}(A_{jk}+A_{kj}) + \frac{1}{d(k)}$$

with $d(k)$ out-degree of node k on the unweighted graph, i.e. the number of children of node k.

Computational Complexity Analysis. Building on Eq. (3.11) in Chapter 3 the computational cost associated with the exact computation of the optimal solution is $O(n^s s^{2.376})$. In this expression, the exponential term n^s arises from the combinatorial nature of the problem, i.e. the need to evaluate all possible subsets of size s, while the factor $s^{2.376}$ corresponds to the cost of inverting a $s \times s$ matrix. We highlight how the proposed selection criteria are significantly less computationally expensive compared to the matrix inversion step. For clarity, let us specify that the following analysis is carried out under the worst-case assumption, where no additional structural properties of the system matrices can be exploited to reduce computational effort.

We now proceed to analyze the individual improvements enabled by each of the proposed criteria.

- The **first criterion** involves selecting the s elements with the highest weighted variance $\alpha' C_U$. Computing the weighted variance requires $O(n^2)$ operations, while selecting the s highest values can be achieved by sorting the resulting vector, which requires $O(n\log(n))$ operations.

  The overall computational cost is thus $O(n^2)$.

- The **second criterion**, given a fixed subset $\mathcal{K}$, requires $O(ns)$ operations, associated to the matrix multiplications required. Moreover, the first term of the sum would require $O(n\log(n))$ operations to select the best subset of s nodes since it is enough to apply a sorting algorithm on the weighted row sum. In contrast, the second term requires to compute the sum for all the possible subsets $\mathcal{K} \subseteq \mathcal{V}$ of cardinality s . Thus, in the worst case scenario the overall computational cost is $O(sn^{s+1})$. However, let us notice that not all the $\binom{n}{s}$ possible subsets must be checked, but only those with the same value for first criterion $F_0^0(\mathcal{K})$.
- The **third criterion** is the most expensive. Given a subset $\mathcal{K}$ the computational cost to evaluate $F_0^2(\mathcal{K})$ is of the same order of the matrix inversion for the brute force algorithm, i.e. $O(s^{2.376})$. Thus as for the brute force method in the worst case scenario the overall computational cost is $O(n^2 s^{2.376})$ However, also in this case not all the $\binom{n}{s}$ subsets need to be checked, but just those with equal first and second criterion.

Function Submodularity Analysis. Let us now analyse function submodularity to discuss when the greedy algorithm represents a good heuristic to apply to reduce the overall computational complexity.

Corollary 4.5. *Under the assumptions of Theorem 4.1, in low-interaction regime, i.e.* $\lambda \to 0$*, submodularity property of the* $F(\mathcal{K})$ *expansion coefficients holds up to the first-order coefficient* $F_0^1(\mathcal{K})$.

Proof. First, observe that $F_0^0(\mathcal{K})$ coincides with the total variance of the set $\mathcal{K}$, thus its submodularity follows immediately. Let us now focus on $F_0^1(\mathcal{K})$.
Consider two subsets $\mathcal{A},\mathcal{B}$ such that $\mathcal{A} \subseteq \mathcal{B} \subseteq \mathcal{V}$ and a node $k \in \mathcal{V}\backslash\mathcal{B}$, then, from submodularity definition in (3.12), we are looking for

$$F_0^1(\mathcal{A}\cup\{k\}) - F_0^1(\mathcal{A}) \geq F_0^1(\mathcal{B}\cup\{k\}) - F_0^1(\mathcal{B}).$$

Indeed, given

$$\begin{aligned} F_0^1(\mathcal{A}\cup\{k\}) - F_0^1(\mathcal{A}) &= \textstyle\sum_{i\in\mathcal{V}\backslash\mathcal{A}\cup\{k\}} \sum_{j\in\mathcal{A}\cup\{k\}} A_{ij} - \sum_{i\in\mathcal{V}\backslash\mathcal{A}} \sum_{j\in\mathcal{A}} A_{ij} = \\ &= \textstyle\sum_{i\in\mathcal{V}\backslash\mathcal{A}\cup\{k\}} A_{ik} - \sum_{j\in\mathcal{A}} A_{kj} \end{aligned}$$

$$\begin{aligned} F_0^1(\mathcal{B}\cup\{k\}) - F_0^1(\mathcal{B}) &= \textstyle\sum_{i\in\mathcal{V}\setminus\mathcal{B}\cup\{k\}}\sum_{j\in\mathcal{B}\cup\{k\}} A_{ij} - \sum_{i\in\mathcal{V}\setminus\mathcal{B}}\sum_{j\in\mathcal{B}} A_{ij} = \\ &= \textstyle\sum_{i\in\mathcal{V}\setminus\mathcal{B}\cup\{k\}} A_{ik} - \sum_{j\in\mathcal{B}} A_{kj} \end{aligned}$$

Thus,

$$\begin{array}{ll} \sum_{i\in\mathcal{V}\setminus\mathcal{A}\cup\{k\}} A_{ik} - \sum_{j\in\mathcal{A}} A_{kj} & \geq \sum_{i\in\mathcal{V}\setminus\mathcal{B}\cup\{k\}} A_{ik} - \sum_{j\in\mathcal{B}} A_{kj} \\ \sum_{i\in\mathcal{V}\setminus\mathcal{B}\cup\{k\}} A_{ik} + \sum_{i\in\mathcal{B}\setminus\mathcal{A}\cup\{k\}} A_{ik} - \sum_{j\in\mathcal{A}} A_{kj} & \geq \sum_{i\in\mathcal{V}\setminus\mathcal{B}\cup\{k\}} A_{ik} - \sum_{j\in\mathcal{A}} A_{kj} - \sum_{j\in\mathcal{B}\setminus\mathcal{A}} A_{kj} \\ \sum_{i\in\mathcal{B}\setminus\mathcal{A}\cup\{k\}} A_{ik} + \sum_{j\in\mathcal{B}\setminus\mathcal{A}} A_{kj} & \geq 0 \end{array}$$

The inequality is satisfied for every choice of subsets, by the definition of matrix A. On the contrary, submodularity property fails to be satisfied for $F_0^2(\mathcal{K})$ due to the last two terms in its formulation. Precisely, the submodularity inequality

$$F_0^2\mathcal{A}\cup\{k\}) - F_0^2(\mathcal{A}) \geq F_0^2(\mathcal{B}\cup\{k\}) - F_0^2(\mathcal{B})$$

coincides with the following one which is not satisfied in the general case:

$$\sum_{i\in\mathcal{B}\setminus A}\sum_{t}[A_{ti}A_{tk} - 2(A_{ik}+A_{ki})(A_{tk}+A_{ti})] \leq 0\,.$$

□

4.2.3 High-Interaction Rate Scenario

In this section we will focus on the case of predominance of interactions in the opinion dynamic model, i.e. λ approaches to 1. Similarly to the previous section, the adopted strategy is to consider the Taylor expansion of $M(\lambda) = (I-[\lambda]A)^{-1}(I-[\lambda])$ for $\lambda \to 1$. However, since the analysis of this case study is more difficult, we will provide some extra assumptions on the model to succeed in highlighting some interesting insights on the selection criteria. Specifically, we will go through the case of one single observation allowed, i.e. $\mathcal{K} = \{k\}$.

Assumption 4.1. *Let X be defined as in* (4.10)*, with U randomly distributed with covariance matrix C_U. Assume C_U to be a multiple of the identity matrix, i.e. $C_U = \sigma^2 I$ with $\sigma^2 \in (0,1)$, and $\Lambda = [\lambda]$ to be a diagonal matrix with diagonal elements all equal to λ. Let $Y = \alpha' X$ with $\alpha = \mathbb{1}/n$, which coincides with the average network status.*

Theorem 4.2. *Let X be the equilibrium opinion vector defined as in* (4.10) *s.t. Assumption 4.1 are satisfied. Denote π the stationary distribution, s.t. $\pi = A\pi, \pi = A'\pi$, and define*

$$Z := \sum_{t\geq 0} (A^t - \mathbb{1}\pi') \,. \tag{4.13}$$

Given $F(\{k\})$ the variance reduction subjected to observation $\mathcal{K} = \{k\}$, in case of high interaction rate, i.e. λ approaches to 1*, the following second-order Taylor expansion holds:*

$$F(\{k\}) = F_1^0(\{k\}) + (1-\lambda)F_1^1(\{k\}) + (1-\lambda)^2 F_1^2(\{k\}) + o((1-\lambda)^2)$$

where $F_1^0(\{k\})$ and $F_1^1(\{k\})$ are constant terms which do not depend on $\{k\}$, and

$$\begin{aligned} F_1^2(\{k\}) = \; & (4n - \tfrac{3n}{\sum_i \pi_i^2})(\textstyle\sum_i \pi_i Z_{ki})^2 \\ & - \left[2n(\textstyle\sum_i \pi_i^2 - 1) + \tfrac{2}{\sum_i \pi_i^2}(\textstyle\sum_{i,j} \pi_i Z_{ji})\right] \textstyle\sum_i \pi_i Z_{ki} \\ & + (2n \textstyle\sum_i \pi_i^2 - 2n)(\textstyle\sum_{i,j} \pi_j Z_{ki} Z_{ij}) \\ & - n \textstyle\sum_i \pi_i^2 \sum_i (Z_{ki})^2 \\ & + 2 \textstyle\sum_{i,j} Z_{ki} Z_{ji} \end{aligned} \tag{4.14}$$

Proof. Building on Assumption 4.1, let the equilibrium vector X be defined as $X = (I - \lambda A)^{-1}(1-\lambda)U$. Let $M(\lambda) = (I - \lambda A)^{-1}(I - \lambda)$. For simplicity of notation we will refer to $M(\lambda)$ simply using M.

Step 1. Based on Corollary 3.1 and noticing that $C = \mathbb{E}[XX'] = MM'$, let us rewrite the variance reduction function subjected to one single observation k as

$$F(\{k\}) = \frac{(MM'\mathbb{1})_k^2}{(MM')_{kk}}$$

and compute the Taylor expansion of M for $\lambda \to 1$:

$$M(\lambda) = M(1) + (1-\lambda)M^{(1)}(1) + \frac{(1-\lambda)^2}{2} M^{(2)}(1) + o((1-\lambda)^2)$$

where $M^{(i)}$ represents the i-th order derivative of M.

First, let us focus on a way to compute $M(1)$. Preliminary notice that building on M definition, the following equations hold

$$M(\lambda)(I-\lambda A) = (1-\lambda)I \qquad (I-\lambda A)M(\lambda) = (1-\lambda)I\,. \tag{4.15}$$

Consequently, for $\lambda = 1$,

$$M(1)(I-A) = 0 \qquad (I-A)M(1) = 0$$

from which we deduce that

$$M(1) = \mathbb{1}\pi'$$

with π stationary distribution.

Let us now focus on $M^{(1)}(1)$. If we substitute in (4.15) the Taylor expansion of $M(\lambda)$ we retrieve

$$\Big(M(1)+(1-\lambda)M^{(1)}(1)+\frac{(1-\lambda)^2}{2}M^{(2)}(1)\Big)(I-A+(1-\lambda)A) = (1-\lambda)I$$

and equivalently

$$(I-A+(1-\lambda)A)\Big(M(1)+(1-\lambda)M^{(1)}(1)+\frac{(1-\lambda)^2}{2}M^{(2)}(1)\Big) = (1-\lambda)I$$

Comparing now the two equation coefficients we get that

$$M(1)A+M^{(1)}(1)(I-A) = I \qquad (I-A)M^{(1)}(1)+AM(1) = I\,.$$

Thus, known that $M(1) = \mathbb{1}\pi'$, the equations can be rewritten as

$$M^{(1)}(1)(I-A) = I-\mathbb{1}\pi' \qquad (I-A)M^{(1)}(1) = I-\mathbb{1}\pi'\,.$$

leading to

$$M^{(1)}(1) = Z := \sum_{t\geq 0}(A^t-\mathbb{1}\pi')\,.$$

The above relation can be properly verified by the following series of equalities:

$$\begin{aligned} Z(I-P) &= (\textstyle\sum_{t\geq 0}(A^t - \mathbb{1}\pi'))(I-A) \\ &= \textstyle\sum_{t\geq 0}(A^t - \mathbb{1}\pi') - \sum_{t\geq 0}(A^{t+1} - \mathbb{1}\pi') \\ &= \textstyle\sum_{t\geq 0}(A^t - \mathbb{1}\pi') - \sum_{t\geq 1}(A^t - \mathbb{1}\pi') \\ &= I - \mathbb{1}\pi' \end{aligned}$$

and equivalently for $(I-A)Z = I - \mathbb{1}\pi'$.

Finally, let us compute an explicit expression for $M^{(2)}(1)$. Analysing the coefficients for $(1-\lambda)^2$, we retrieve that

$$\frac{1}{2}M^{(2)}(1)(I-A) + M^{(1)}A = 0 \qquad \frac{1}{2}(I-A)M^{(2)}(1) + AM^{(1)}(1) = 0$$

and so, building on previous results, the following series of equality holds:

$$\begin{aligned} M^{(2)}(1)(I-A) &= -2ZA \\ M^{(2)}(1)(I-A) &= -2Z + 2(I - \mathbb{1}\pi') \\ M^{(2)}(1)(I-A)Z &= -2Z^2 + 2Z - 2\mathbb{1}\pi'Z \\ M^{(2)}(1)(I - \mathbb{1}\pi') &= -2Z^2 + 2Z - 2\mathbb{1}\pi'Z \end{aligned}$$

Thus, we retrieve that

$$M^{(2)}(1) = -2Z^2 + 2Z = -2\sum_{t\geq 1} t(A^t - \mathbb{1}\pi').$$

Step 2. The definitions of i-th order derivatives of M can be now used to write the second order expansion of function $F(\{k\})$, i.e.

$$F(\{k\}) = \frac{(MM'\mathbb{1})_k^2}{(MM')_{kk}}$$

To improve the readability of the subsequent analysis we will set $1-\lambda = \tau$. First, let's compute an explicit espression for the product MM', fundamental element both for the numerator and the denominator of the formula of interest.

$$MM' = \mathbb{1}\pi'\pi\mathbb{1}' + \tau(Z\pi\mathbb{1}' + \mathbb{1}\pi'Z') + \tau^2(\mathbb{1}\pi'(Z-Z^2)' + (Z-Z^2)\pi\mathbb{1}' + ZZ') + o(\tau^2)$$

Let us now focus on the numerator, i.e. $(MM'\mathbb{1})_k^2$.

$$\begin{aligned}(MM'\mathbb{1})_k = \;& (\mathbb{1}\pi'\pi\mathbb{1}'\mathbb{1})_k + \tau((Z\pi\mathbb{1}' + \mathbb{1}\pi'Z')\mathbb{1})_k + \\ & + \tau^2((\mathbb{1}\pi'(Z-Z^2)' + (Z-Z^2)\pi\mathbb{1}' + ZZ')\mathbb{1})_k + o(\tau^2) \\ (MM'\mathbb{1})_k^2 = \;& \left(n\textstyle\sum_i \pi_i^2\right)^2 + \tau\Big[2(n\textstyle\sum_i \pi_i^2)((Z\pi\mathbb{1}' + \mathbb{1}\pi'Z')\mathbb{1})\Big]_k + \\ & + \tau^2\Big[(((Z\pi\mathbb{1}' + \mathbb{1}\pi'Z')\mathbb{1})_k)^2 + 2(n\textstyle\sum_i \pi_i^2)((\mathbb{1}\pi'(Z-Z^2)' + (Z-Z^2)\pi\mathbb{1}' + \\ & + ZZ')\mathbb{1})_k\Big] + o(\tau^2)\end{aligned}$$

Analogously, we now focus on the denominator, i.e $(MM')_{kk}$, and we compute its mutual.

$$(MM')_{kk} = \sum_i \pi_i^2 + \tau\left[Z\pi\mathbb{1}' + \mathbb{1}\pi'Z'\right]_{kk} + \tau^2\left[\mathbb{1}\pi'(Z-Z^2)' + (Z-Z^2)\pi\mathbb{1}' + ZZ'\right]_{kk} + o(\tau^2)$$

$$\begin{aligned}\tfrac{1}{(MM')_{kk}} = \;& \tfrac{1}{(\sum_i \pi_i^2)^2}\Big[\textstyle\sum_i \pi_i^2 - \tau\,(Z\pi\mathbb{1}' + \mathbb{1}\pi'Z')_{kk} + \\ & + \tau^2\Big(([Z\pi\mathbb{1}' + \mathbb{1}\pi'Z']_{kk})^2 - (\textstyle\sum_i \pi_i^2)\Big[\mathbb{1}\pi'(Z-Z^2)' + (Z-Z^2)\pi\mathbb{1}' + ZZ'\Big]_{kk}\Big) + \\ & + o(\tau^2)\Big]\end{aligned}$$

Step 3. Collecting these expressions, we can now properly compute the second order expansion for $F(\{k\})$, i.e.

$$F(\{k\}) = F_1^0(\{k\}) + \tau F_1^1(\{k\}) + \frac{\tau^2}{2}F_1^2(\{k\}) + o(\tau^2)\,.$$

Specifically, we retrieve:

- Constant term.
$$F_1^0(\{k\}) = n^2\sum_i \pi_i^2$$
This term is a constant that does not depend on the chosen node k.

- First order coefficient.
$$F_1^1(\{k\}) = 2n((Z\pi\mathbb{1}' + \mathbb{1}\pi'Z')\mathbb{1})_k - n^2\Big(Z\pi\mathbb{1}' + \mathbb{1}\pi'Z'\Big)_{kk} = 2n\sum_{i,j}\pi_i Z_{ji}$$

This term is a constant that does not depend on the chosen node k. Moreover, we observe that if π is uniform then it is in particular equal to 0.

- Second order coefficient.

$$\begin{aligned}
F_1^2(\{k\}) = n^2\Big(([Z\pi\mathbb{1}' + \mathbb{1}\pi'Z']_{kk})^2 - (\textstyle\sum_i \pi_i^2)\left[\mathbb{1}\pi'(Z-Z^2)' + (Z-Z^2)\pi\mathbb{1}' + ZZ'\right]_{kk}\Big) + \\
- [2n((Z\pi\mathbb{1}' + \mathbb{1}\pi'Z')\mathbb{1})]_k \tfrac{(Z\pi\mathbb{1}'+\mathbb{1}\pi'Z')_{kk}}{\sum_i \pi_i^2} + \tfrac{1}{\sum_i \pi_i^2}\left[(((Z\pi\mathbb{1}' + \mathbb{1}\pi'Z')\mathbb{1})_k)^2\right] + \\
+2n((\mathbb{1}\pi'(Z-Z^2)' + (Z-Z^2)\pi\mathbb{1}' + ZZ')\mathbb{1})_k \\
\sim (4n - \tfrac{3n}{\sum_i \pi_i^2})(\textstyle\sum_i \pi_i Z_{ki})^2 - \left[2n(\sum_i \pi_i^2 - 1) + \tfrac{2}{\sum_i \pi_i^2}(\sum_{i,j} \pi_i Z_{ji})\right]\sum_i \pi_i Z_{ki} + \\
+(2n\textstyle\sum_i \pi_i^2 - 2n)(\sum_{i,j} \pi_j Z_{ki} Z_{ij}) - n\sum_i \pi_i^2 \sum_i (Z_{ki})^2 + 2\sum_{i,j} Z_{ki} Z_{ji}
\end{aligned}$$

The thesis follows. □

Building on the results of Theorem 4.2 we can now define a proper strategy to select the best node to observe. Notice that the constant coefficients of the Taylor expansion do not provide any useful information for the node selection.

Algorithm 4.3 (**Single Node Selection for High-Interaction**). *Assume to be under the Assumptions of Theorem 4.2. The best node $k^* \in \mathcal{O} \subseteq \mathcal{V}$ to observe, among the observable ones in $\mathcal{O}$ in case oh high interaction, i.e. when λ approaches to* 1*, is the one which maximizes $F_1^2(\{k\})$. Formally,*

$$k^* = \arg\max_{k \in \mathcal{O}} F_1^2(\{k\}).$$

Remark 4.4. *Referring to the Taylor expansion of $F(\{k\})$ for $\lambda \to 1$ we highlight that $F_1^0(\{k\})$ to be a constant matches what is expected for the extreme case $\lambda = 1$ where all the nodes provide the same information once the consensus has been reached. It coincides in particular with the variance of the stationary distribution.*

The expression retrieved takes a well interpretable form in particular in the case of uniform stationary distribution. More precisely, in the following we will interpret the model of interest as a discrete time Markov Chain with transition matrix A.

Corollary 4.6. *If π is uniform, the optimization problem can be rewritten as follows*

$$k^* = \arg\min_{k \in \mathcal{O}} \left|\left| Z_{k,:} - \sum_j Z_{j,:} \right|\right|_2^2,$$

with Z defined as in (4.13) *and $\mathcal{O} \subseteq \mathcal{V}$ observable set.*

Proof. Assumed π to be uniform, i.e. $\pi = \frac{\mathbb{1}}{n}$, and known that $\sum_i Z_{ki} = 0$, the function $F_1^2(\{k\})$ introduced in (4.14) can be simplified, avoiding constant terms, as

$$F_1^2(\{k\}) = -\sum_i (Z_{ki})^2 + 2\sum_{i,j} Z_{ki} Z_{ji},$$

which can be refactored as follows

$$F_1^2(\{k\}) = -\sum_i \left[\left(Z_{ki} - \sum_j Z_{ji} \right)^2 \right] + \sum_i \left(\sum_j Z_{ji} \right)^2 .$$

Thus, neglecting the last term which does not depend on k, the optimization problem in (3.3), whose goal is to find that $k \in \mathcal{O}$ which maximizes the variance reduction $F(\{k\})$, coincides with maximizing $F_1^2(\{k\})$. Substituting the explicit expression for $F_1^2(\{k\})$, the thesis follows. □

Corollary 4.7. *When π is uniform, the following alternative formulation for the optimization problem, in terms of hitting times, holds:*

$$\arg\min_{k \in \mathcal{O}} \left|\left| \frac{\sum_j E_j T_i}{n} - E_k T_i \right|\right|_2^2 \tag{4.16}$$

where $E_k T_i$ is the mean hitting time starting from node k to hit node i and with $E_\pi T_i$ the expected time to hit node i starting from an arbitrary node with uniform probability distribution π. The retriven expression corresponds to select that node whose expected hitting times vector minimize the distance from the average of expected network hitting times.

Proof. First, recall that, from [87], the function Z introduced in (4.13) corresponds to the fundamental matrix of a Discrete Time Markov Chain with transition matrix A. Thus, from [87], assumed π uniform, it follows that

$$Z_{ii} - Z_{ki} = \frac{1}{n} E_k T_i$$

and

$$Z_{ii} = \frac{1}{n} E_\pi T_i = \frac{1}{n^2} \sum_j E_j T_i .$$

Building on the results of Corollary 4.6, the thesis follows. □

4.2.4 Examples and Simulations

In this section we provide some examples and simulations in the setting of Friedkin-Johnsen opinion dynamics model, aimed at two main analysis: first, the comparison of our centrality measure against commonly known ones; on the other hand, the analysis of the viability of the stated criteria as simplification of the exact function computation. Let us point out that the centrality measures will be compared in terms of average performance error, defined as:

$$\text{err} = \frac{F(\{\hat{k}\}) - F(\{k^*\})}{F(\{k^*\})}$$

with $\hat{k}$ best node to observe according to the chosen centrality measure for the comparison and k^* optimal selection according to the proposed Variance Reduction function. Even though some of the proposed examples may appear trivial, they are included here because they help to highlight certain emerging behaviors that would be less evident when working with higher-dimensional, more complex networks.

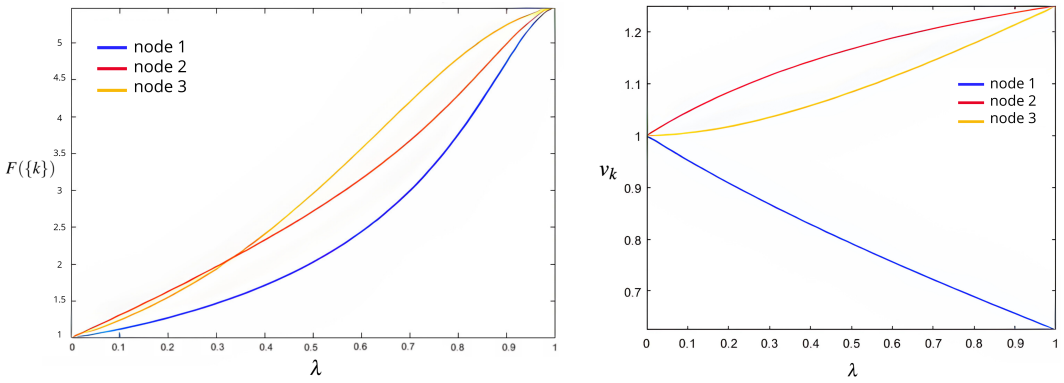
Line Graph

Fig. 4.9 Line graph with 5 nodes.

Let's start considering the case of a line graph. Specifically, we assume to work with an undirected graph of 5 nodes, represented in Fig. 4.9, with weight matrix A defined as follows

$$A = \begin{bmatrix} 0 & 1 & 0 & 0 & 0 \\ 1/2 & 0 & 1/2 & 0 & 0 \\ 0 & 1/2 & 0 & 1/2 & 0 \\ 0 & 0 & 1/2 & 0 & 1/2 \\ 0 & 0 & 0 & 1 & 0 \end{bmatrix}.$$

Assume that we can choose only one node to observe. As widely explained in the previous chapter, the goal is to find that node k in $\mathcal{V}$ which maximizes the variance reduction $F(\{k\})$. Due to the graph simmetry we can just focus our attention on 3 candidate solutions, i.e. nodes 1, 2 or 3.



(a) Variance Reduction as function of λ

(b) Bonacich Centrality as function of λ

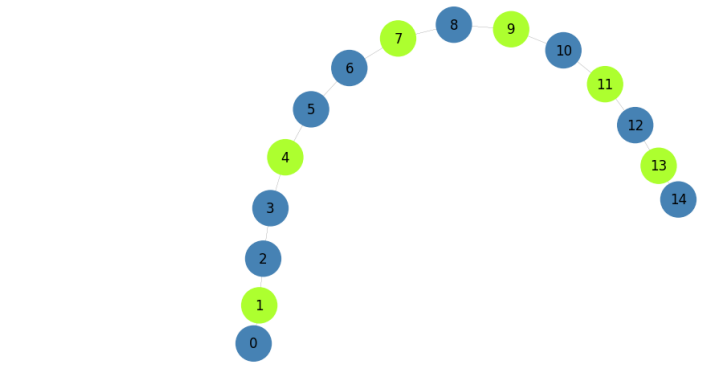
Fig. 4.10 Best node to observe over a line graph with 5 nodes as function of parameter λ according to Variance Reduction function (a) vs Bonacich Centrality measure (b).

In Fig.4.10 the three candidate solutions are compared. Precisely, the graph in Fig.4.10-(a) represents the variance reduction achieved by the observation of one out of the three candidate cases, as function of parameter λ. From the results achieved we highlight how the model parameters clearly influence the final choice, indeed for small values of λ the best node to observe is node 2, while for greater values of the parameter the central node is the preferred one. Comparing it with Fig.4.10-(b), which focuses on the corresponding Bonacich centrality values, we observe that the latter does not depend on the model parameter λ and the node with highest centrality is node 2 which coincides with the optimal node selected by our algorithm only for small values of λ.

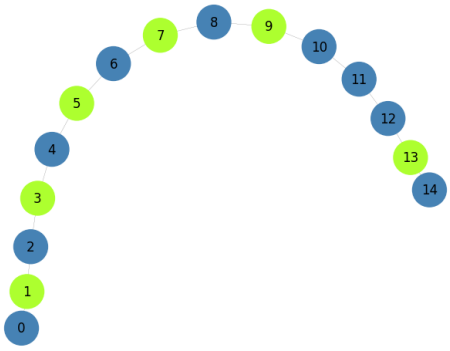
A well known problem of the proposed algorithm is its computational complexity. Building on the approximations proposed in Section 4.2.2 for $\lambda \to 0$, we will corroborate now our criteria in the simple case of study. If we let $\sigma_i^2 = \sigma^2$ for every $i \in \mathcal{V}$, the optimal selection criterion to apply is Criterion 2, i.e. the first order McLaurin expansion coefficient of $F(\{k\})$. In this trivial examples we can make explicit computations from which we get:

$$F_0^1(1) = 1 + \frac{1}{2} \qquad F_0^1(2) = 1 + \frac{3}{2} \qquad F_0^1(3) = 1 + 1\,.$$

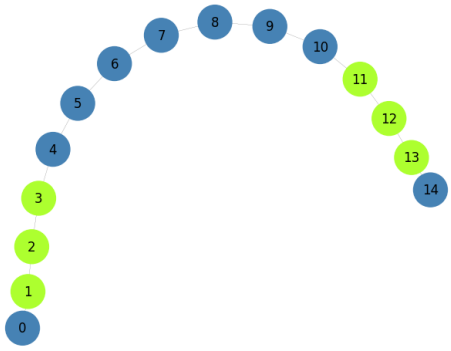
This is in line with what expected, i.e. the best node to observe is node 2. Clearly, a precise characterization of how small λ must be for the simplified criteria to remain valid is non trivial and depends on the specific case of interest.



(a) Exact Optimal Selection



(b) Approximated Criterion Selection



(c) Bonacich Centrality Selection

Fig. 4.11 Comparison of node selections for $|\mathcal{K}| = 6$ on a line graph with 15 nodes using respectively (a) exact functional (b) approximated criteria and (c) Bonacich centrality.

We can generalize the analysis considering $|\mathcal{K}| > 1$. Let's take for example a line graph with 15 nodes among which we want to select the optimal subset of 6 nodes for $\lambda \to 0$. In Fig. 4.11 the different selections depending on the chosen criterion are shown over the line graph. We highlight that in this case the approximated criteria return a relative error of about 0.7% while using Bonacich centrality measure (i.e., select the subset of s nodes with highest centrality) the error lines up to 18%.

To analyse how much the function approximation is robust to changes of parameter λ we compare the results obtained for a line of 50 nodes under the assumption of being able to observe up to 5 nodes, for λ in the range $[0.01, 0.4]$. As shown in Fig.4.12 the resultant error gets a maximum value of 12% for maximum value of λ equal to 0.4 which is still promising.

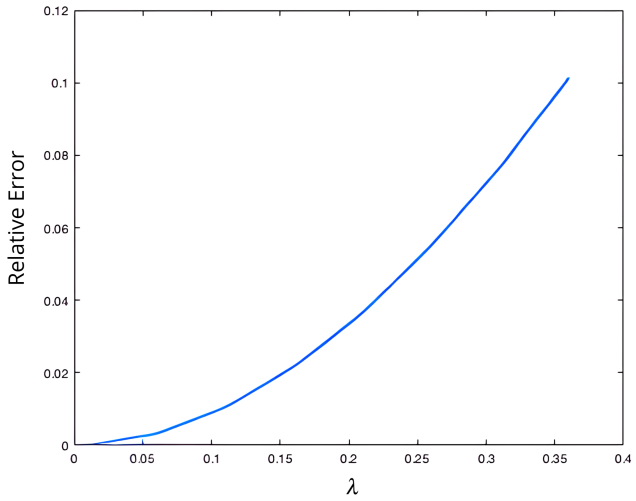


Fig. 4.12 Relative error of function approximation (for $\lambda \to 0$) with respect to exact solution for a line of 50 nodes, $|\mathcal{K}| = 5$ and λ in $[0.01, 0.4]$.

Let's study now the case $\lambda \to 1$. In the simple case of 5 nodes we have already highlighted that the widely known Bonacich centrality doesn't provide the optimal node selection and its behaviour is not influenced at all by the value of model parameter λ. Consider now a line graph with 30 nodes and parameter $\lambda = 0.9$. Despite the defined criterion doesn't identify the optimal node (node 8 optimal choice for the exact function vs node 16 optimal choice for the approximation criterion), in this case the relative error amounts to 1.33%. On the contrary, adopting the Bonacich centrality measure to choose the best node to observe the relative error detected would have been equal to 34.5% (optimal choice node 2).

Further level of generalization takes into consideration the variance reduction as function of the cardinality of the observed set $\mathcal{K}$ given that an higher number of observations implies also an higher cost. Thus, we focus here not on the selection set but on its cardinality. The pattern is highlighted in the Fig.4.13. We may observe that vhanges to the value of parameter λ reflect to the impact of cardinality of $\mathcal{K}$ on the variance reduction. Higher values of λ lower the relative improvement in variance reduction.

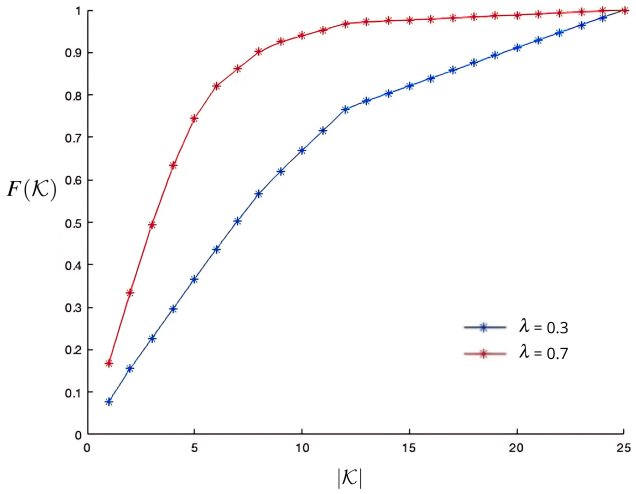


Fig. 4.13 Variance reduction as function of the number of observations $|\mathcal{K}|$ with different values of parameter λ for a line graph with 30 nodes.

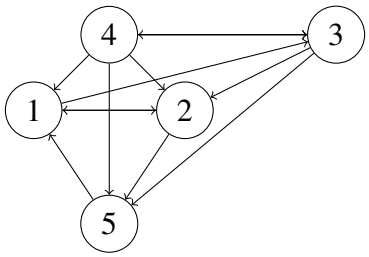


Fig. 4.14 Example of graph with 5 nodes with two nodes with equally optimal in-degree.

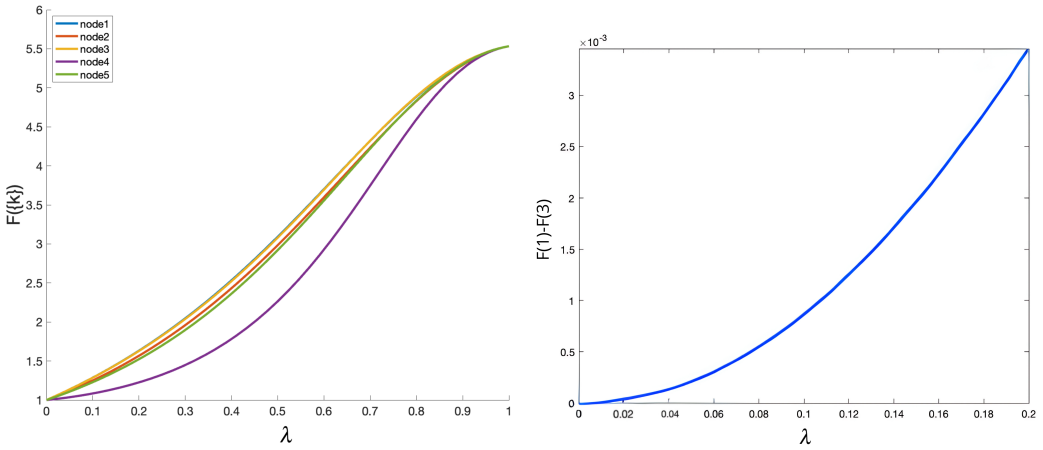
Ad hoc Example Graph

Let's consider now the particular graph in Fig.4.14 described by the following adiacency matrix

$$P = \begin{bmatrix} 0 & 0.5 & 0.5 & 0 & 0 \\ 0.5 & 0 & 0 & 0 & 0.5 \\ 0 & 0.4 & 0 & 0.3 & 0.3 \\ 0.5 & 0 & 0.3 & 0 & 0.2 \\ 0.3 & 0.2 & 0.5 & 0 & 0 \end{bmatrix}$$

and assume $\sigma_i^2 = \sigma^2$ for all the nodes $i \in \mathcal{V}$.

Let's first study the case $\lambda \to 0$. Consider as benchmark the result of the exact optimization problem according to which the best node to observe is node 1 as highlighted in Fig.4.15. By applying the proposed criteria we observe that the second



(a) Variance Reduction as function of λ (b) Variance Reduction differences

Fig. 4.15 (a) Best node to observe over the graph chosen as example as function of parameter λ according to our objective function (b) Comparison of objective functions for nodes 1 and 3 with equal in-degree.

criterior is not exaustive given that nodes 1 and 3 have both the highest in-degree equal to $F_0^1(1) = F_0^1(3) = 1.3$. Thus, we need to consider the third criterion which involves the path of length 2 in the considered graph. From the computations we therefore deduce that the optimal node to observe is node 1 given that $F_0^2(1) = 4.99 > F_0^2(3) = 4.73$. This is in line with what was expected compared with the result of the application of the exact variance reduction function as shown in Fig.4.15, while not matching the maximum point of Bonacich centrality, i.e. node 3.

Let's consider now the case $\lambda \to 1$. The optimal node detected through the exact function is not the same identified by the approximated criteria(optimal choice node 3 vs criterion choice node 2), but the error is about 0.44%. The optimal choice in this case matches the Bonacich centrality one.

Finally, we analyse the variance reduction as function of the cardinality of the observed set $\mathcal{K}$ for the ad hoc example built for the previous section. As expected, increasing the value of parameter λ the nodes will decrease their stubborness, making the observation of few nodes sufficient to significantly reduce the residual variance.

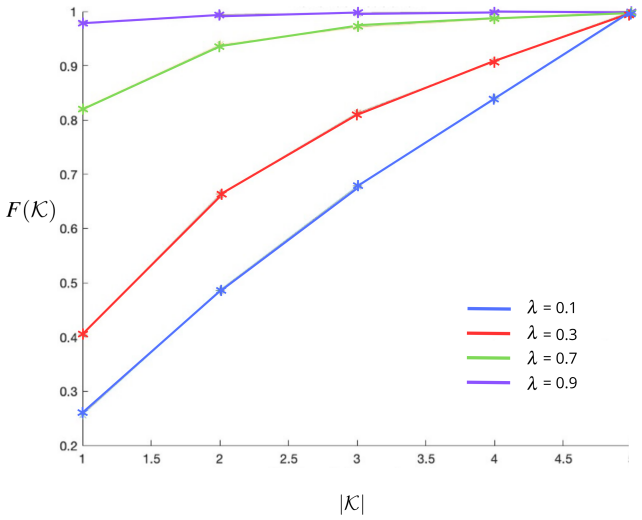


Fig. 4.16 Variance reduction as function of the number of observations $|\mathcal{K}|$ for a line graph with 30 nodes as function of parameter λ

Small-World Network

We will focus now on a more complex graph, a Small-World network defined according to the Watts-Strogatz's Model, which better describes real world networks. The parameters used will be the following:

- N: number of nodes of the graph;
- k: mean degree of each node before rewiring;
- β: probability of edge rewiring.

We first study the case $\lambda \to 0$. In the example shown in Fig.4.17 we took as parameters $N = 500$, $k = 30$ and $\beta = 0.0006$. 1000 simulations were carried out in order to account for the randomness in the graph structure and get statistically relevant results. The approximation criteria perform well, reaching an average relative error of about $\sim 10^{-3}$. The result outperforms the random choice which instead retrieves an error of about 1.2% given the same parameters.

Generalizing to the case $|\mathcal{K}| > 1$, we consider here Small-World graphs with $N = 50$, $k = 5$, $\beta = 0.1$ and $\lambda = 0.2$. In our simulations, targeted at selecting the best 7 nodes out of 50, the approximated criteria provide the exact solution, while using the Bonacich centrality the error lines up to 5.7%.

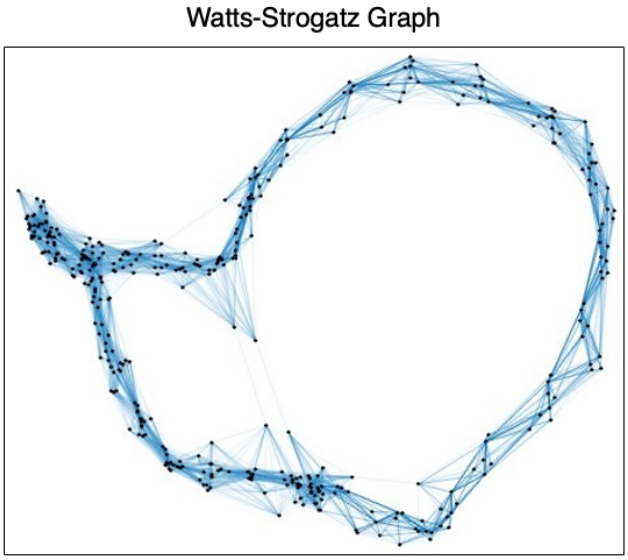


Fig. 4.17 Watts-Strogatz model with 500 nodes, mean node degree 30 and $\beta = 0.0006$.

Let's consider now the case $\lambda \to 1$ and take into account 1000 Watts-Strogatz's models with parameters $N = 500$, $k = 30$, $\beta = 0.3$ and $\lambda = 0.9$. The average error obtained using approximated criteria is equal to 0.31%.

The behaviour of the variance reduction in Fig.4.18 as function of the cardinality of the observed set $\mathcal{K}$ and of parameter λ is in line with what was shown in previous examples.

Real Network - Facebook dataset

The last simulation uses SNAP network available in Standford Large Network Dataset Collection [88], looking in particular to Facebook Social Network. These data are real anonymized data collected by Facebook app and they consist of 4039 nodes, 88234 edges, average clustering coefficient 0.6055 and diameter equal to 8. From our analysis we observe that for small λ values the optimal selections provided by the exact variance reduction function and by the approximated criteria coincide. On the contrary, for higher values of λ the relative error is low only for values which tend to $\lambda = 1$ (e.g. for $\lambda = .9$ the error detected is about 68%, for $\lambda = .999$ the error is about 33%, while for $\lambda = .9999$ we retrieve a lower error of about 1.5%). This suggests that the error due to approximation is not negligible for higher λ.

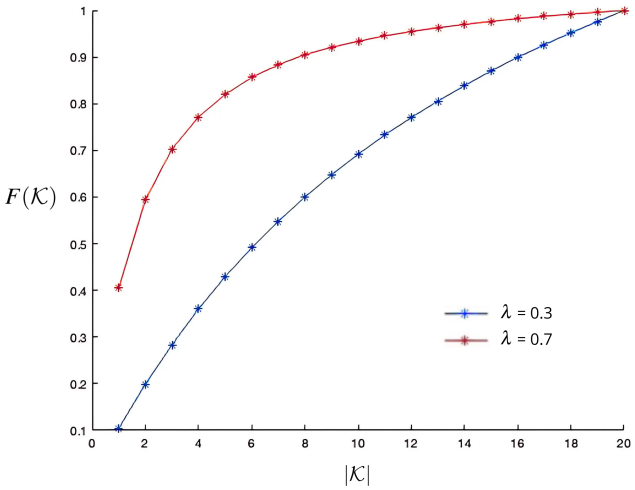


Fig. 4.18 Variance reduction as function of the number of observations $|\mathcal{K}|$ in a Watts-Strogatz model with 20 nodes, mean node degree 9 and $\beta = 0.4$

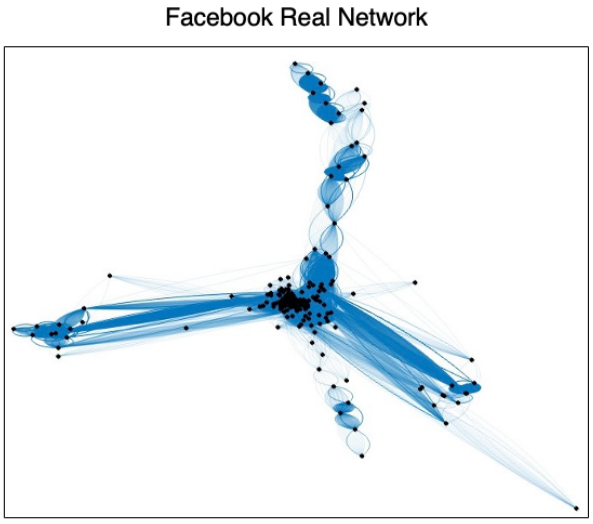


Fig. 4.19 Facebook Social Network representation.

PART II

Stubborn Nodes Detection

CHAPTER 5

Detecting Stubborn Behaviours

In this Chapter, we analyse the second main problem of interest of this dissertation, focused on detecting stubborn agents without network knowledge and without the need to reconstruct it. Specifically, we consider the classical French-DeGroot opinion dynamics model where some of the agents are stubborn and persistently express a fixed opinion. As introduced in Chapter 2.3, in the asymptotic equilibrium, each non stubborn agent reaches an opinion that is a convex combination of the stubborn agents' ones. Weights of such combinations depend on the topology of interactions and form the so called influence matrix. Here, we assume our data to be a noisy observation of the complete set of such equilibrium opinions for a number of different discussions (e.g. initial conditions) and our goal is to estimate the set of stubborn nodes and the corresponding influence matrix.

Main Contributions. Our contribution is two-fold. First, we formulate the estimation problem as an optimization problem tailored to be computationally efficiently solved by Interpolative Decomposition techniques. Second, we derive sufficient conditions on the model parameters and on the noise's size guaranteeing that any solution of the optimization problem will correctly determine the set of stubborn nodes

and well approximate the influence matrix. Finally, we corroborate our analysis through numerical results.

The results of this work have been published in [55].

5.1 Problem Formulation

Notation. Let us preliminary recall some notation that will be used in the following. For a rectangular matrix X, $\sigma_k(X)$ denotes its k-th largest singular value and $\text{rank}(X)$ its rank. The spectral norm of a matrix is denoted by $\|X\|_2$ and it coincides with the maximum singular value $\sigma_1(X)$. The Frobenius or Hilbert-Schmidt norm is denoted by $\|X\|_F$. The maximum entry of X (in absolute value) is denoted by $\|X\|_{\max} = \max_{(i,j)} |X_{ij}|$. In the following the notation with a superscript $*$ is meant to denote the corresponding true reference value, e.g. $\mathcal{S}^*$ indicates the true set of stubborn nodes we want to detect.

We consider an influence network described by a weighted graph $\mathcal{G} = (\mathcal{V}, \mathcal{E}, W)$ where the set of nodes $\mathcal{V}$ represents the agents, the set of (directed) edges $\mathcal{E}$ indicates the pairwise interactions and the weight matrix $W \in \mathbb{R}_+^{\mathcal{V}\times\mathcal{V}}$ indicates the strength of the interactions. Given agents $i, j \in \mathcal{V}$, we have that $(i, j) \in \mathcal{E}$ if and only if $W_{ij} > 0$.

We assume that each agent i is endowed with a state $x_i(t) \in \mathbb{R}$ representing its opinion/belief at time t. Opinions vary in time according to a classical French-DeGroot model with stubborn agents (See Chapter 2.3 for more details). Precisely, we assume that the set of nodes is partitioned into two disjoint sets: $\mathcal{V} = \mathcal{R}^* \cup \mathcal{S}^*$. Nodes in $\mathcal{S}^*$ (called stubborn agents) maintain their opinion fixed at all times, while nodes in $\mathcal{R}^*$ (the regular agents) update it according to the linear dynamics below:

$$x_i(t+1) = \begin{cases} \sum_{j\in\mathcal{V}} P_{ij} x_j(t), & \forall i \in \mathcal{R}^* \\ x_i(t), & \forall i \in \mathcal{S}^* \end{cases} \tag{5.1}$$

where P is the normalized weight matrix, i.e. $P_{ij} = W_{ij}/(\sum_h W_{ih})$, which will thus be row-stochastic.

Differing from previous chapters, to make the notation more readable, it would be convenient to represent the agents' opinion on a given topic using row vectors, respectively, $x_{\mathcal{R}^*}(t)$ for regular agents and $x_{\mathcal{S}^*}(t)$ for stubborn ones. From Proposition

2.2, recall that, for every initial condition $x(0) = (x_{\mathcal{R}^*}(0), x_{\mathcal{S}^*})$, the dynamics in (5.1) converges to the limit opinion

$$x'_{\mathcal{R}^*} = (I - A)^{-1} B x'_{\mathcal{S}^*},$$

with $A = P_{\mathcal{R}^*\mathcal{R}^*}$ and $B = P_{\mathcal{R}^*\mathcal{S}^*}$. From now on, for notational simplicity, we denote

$$\Gamma^* := ((I - A)^{-1} B)'.$$

The column-stochastic matrix Γ^* is referred to as the *influence matrix*: its entries Γ^*_{ij} measure the relative influence of stubborn agent i on the final opinion of regular agent j.

In the problem setting of interest, for each agent i we consider the opinions on multiple discussion topics $\mathcal{T}$ and collect them in a column vector $X_i \in \mathbb{R}^{\mathcal{T}}$. We assemble the row vectors of the asymptotic opinions of all agents into a matrix X in $\mathbb{R}^{\mathcal{T}\times\mathcal{V}}$ so that $X_{\ell i}$ represents the asymptotic opinion of agent i in $\mathcal{V}$ under the discussion on topic ℓ in $\mathcal{T}$. It follows from (2.7) that

$$X_{\mathcal{R}^*} = X_{\mathcal{S}^*}\Gamma^*, \tag{5.2}$$

where $X_{\mathcal{R}^*}$ and $X_{\mathcal{S}^*}$ indicate the sub-matrices of X consisting of the columns in $\mathcal{R}^*$ and $\mathcal{S}^*$, respectively.

We assume to have access to the observations

$$Y = X + \Xi, \tag{5.3}$$

where Ξ in $\mathbb{R}^{\mathcal{T}\times\mathcal{V}}$ is a noise matrix. Our goal is to identify the true stubborn set $\mathcal{S}^*$ and estimate the influence matrix Γ^* starting from the observation of Y, without any prior knowledge of the network. The problem can be properly mathematically formalized as follows.

Problem 5.1 (Detection and estimation)**.** *For a tolerance $\varepsilon \geq 0$, consider the following low-rank approximation problem*

$$\begin{array}{ll} \min_{\mathcal{S}\subseteq\mathcal{V}} |\mathcal{S}| & \\ \exists \Gamma \in \mathbb{R}_+^{\mathcal{S}\times\mathcal{R}} & \text{s.t. } \|Y_{\mathcal{R}} - Y_{\mathcal{S}}\Gamma\|_2 \leq \varepsilon,\ \|\Gamma\|_{\max} \leq 1 \end{array} \tag{5.4}$$

We denote by $\widehat{\mathcal{S}}_\varepsilon$ a solution of the problem and with $\widehat{\Gamma}_\varepsilon$ any corresponding matrix Γ of minimum 2-norm, satisfying the constraints.

To guarantee correct identification of the stubborn agents in a situation where the network is unknown and observations are only asymptotic we assume the following assumptions hold.

Assumption 5.1. *The matrix $X_{\mathcal{S}^*}$ has rank $k^* = |\mathcal{S}^*|$.*

Assumption 5.2. *There exists $\delta > 0$ such that for every v in $\mathcal{R}^*$, there exist $s_1 \neq s_2$ in $\mathcal{S}^*$ with $\min\{\Gamma^*_{s_1 v}, \Gamma^*_{s_2 v}\} \geq \delta$.*

Assumption 5.1 requires that $|\mathcal{T}| \geq |\mathcal{S}^*|$, i.e., the number of topics is at least equal to the number of stubborn agents, and that the k^* column vectors that gather the initial conditions of each stubborn agent on the different topics are linearly independent. Assumption 5.2 requires instead that every regular agent is influenced by at least two stubborn agents and is thus stronger than the assumption in Proposition 2.2, i.e. global reachability of stubborn nodes set.

Both assumptions are necessary to correctly satisfy our goal. Indeed, if the opinions of one stubborn agent s were a linear combination of the opinions of other stubborn agents, s would be undistinguishable from a regular agent influenced by such stubborn agents through the same linear combination. Similarly, if a regular node v was influenced by just one stubborn node s, then the role of v and s could be interchanged with no possibility to distinguish their roles.

Notice that Assumption 5.1 holds true (with high probability) if stubborn nodes' opinions are considered exogenous inputs modeled as i.i.d. random variables, which is a common assumption in the opinion dynamics literature [89, 90]. Finally, observe that in many applicative scenarios for the detection of trolls and bots [52, 6, 28, 7] the number of stubborn nodes is a small percentage of the total number of nodes so that Assumption 5.1 is computationally feasible.

5.2 Noise-free Scenario Analysis

First, let us focus on the solutions of Problem 5.1 in the noise-free case. The following preliminary result ensures uniqueness of the partition $\mathcal{V} = \mathcal{R}^* \cup \mathcal{S}^*$ when a linear relation as (5.2) holds with Γ^* in $[0,1]^{\mathcal{S}\times\mathcal{R}}$.

Lemma 5.1. *Let Assumptions 5.1 and 5.2 hold true. Let $\mathcal{V} = \mathcal{R} \cup \mathcal{S}$ be a binary partition such that $k^* = |\mathcal{S}| = rank(X_{\mathcal{S}})$, $\mathcal{S} \neq \mathcal{S}^*$, where $\mathcal{S}^*$ is the true subset of stubborn nodes, and let Γ in $\mathbb{R}^{\mathcal{S}\times\mathcal{R}}$ be such that*

$$X_{\mathcal{R}} = X_{\mathcal{S}}\Gamma. \tag{5.5}$$

Then, there exists some (i, j) in $\mathcal{V} \times \mathcal{V}$ such that

$$\Gamma_{ij} \notin \left(-\frac{\delta}{2(1-\delta)}, 1 + \frac{\delta}{2(1-\delta)}\right). \tag{5.6}$$

Proof. Define $\mathcal{A} = \mathcal{R}^* \cap \mathcal{R}$, $\mathcal{B} = \mathcal{R}^* \cap \mathcal{S}$, $\mathcal{C} = \mathcal{S}^* \cap \mathcal{S}$, $\mathcal{D} = \mathcal{S}^* \cap \mathcal{R}$. Since $\mathcal{S} \neq \mathcal{S}^*$, $\mathcal{B}$ and $\mathcal{D}$ are nonempty. Since $|\mathcal{S}| = |\mathcal{S}^*|$, we have $|\mathcal{B}| = |\mathcal{D}|$. From (5.2) and (5.5), respectively, we get

$$X_{\mathcal{B}} = X_{\mathcal{C}}\Gamma^*_{\mathcal{C}\mathcal{B}} + X_{\mathcal{D}}\Gamma^*_{\mathcal{D}\mathcal{B}}$$

and

$$X_{\mathcal{D}} = X_{\mathcal{B}}\Gamma_{\mathcal{B}\mathcal{D}} + X_{\mathcal{C}}\Gamma_{\mathcal{C}\mathcal{D}}.$$

So,

$$X_{\mathcal{B}} - X_{\mathcal{C}}\Gamma^*_{\mathcal{C}\mathcal{B}} = X_{\mathcal{D}}\Gamma^*_{\mathcal{D}\mathcal{B}} = X_{\mathcal{B}}\Gamma_{\mathcal{B}\mathcal{D}}\Gamma^*_{\mathcal{D}\mathcal{B}} + X_{\mathcal{C}}\Gamma_{\mathcal{C}\mathcal{D}}\Gamma^*_{\mathcal{D}\mathcal{B}},$$

that can be rearranged as

$$X_{\mathcal{B}}(I - \Gamma_{\mathcal{B}\mathcal{D}}\Gamma^*_{\mathcal{D}\mathcal{B}}) = X_{\mathcal{C}}(\Gamma^*_{\mathcal{C}\mathcal{B}} + \Gamma_{\mathcal{C}\mathcal{D}}\Gamma^*_{\mathcal{D}\mathcal{B}}).$$

Since $\text{rank}(X_{\mathcal{S}}) = |\mathcal{S}|$ and $\mathcal{B} \cup \mathcal{C} = \mathcal{S}$, the columns of $X_{\mathcal{B}}$ and $X_{\mathcal{C}}$ are linearly independent, so that the above implies

$$\Gamma_{\mathcal{B}\mathcal{D}}\Gamma^*_{\mathcal{D}\mathcal{B}} = I, \qquad \Gamma^*_{\mathcal{C}\mathcal{B}} = -\Gamma_{\mathcal{C}\mathcal{D}}\Gamma^*_{\mathcal{D}\mathcal{B}}. \tag{5.7}$$

Now, for b in $\mathcal{B}$ and d in $\mathcal{D}$, let

$$\gamma^+_b = \max_{i\in\mathcal{D}} \Gamma_{bi}, \quad \gamma^+_{bd} = \max_{i\in\mathcal{D}\setminus\{d\}} \Gamma_{bi}, \quad \gamma^-_{bd} = \min_{i\in\mathcal{D}\setminus\{d\}} \Gamma_{bi}.$$

Now, we distinguish two cases. On the one hand, if $\Gamma^*_{cb} \geq \delta$, for some c in $\mathcal{C}$ and b in $\mathcal{B}$, then $\sum_{d\in\mathcal{D}} \Gamma^*_{db} \leq 1 - \delta$, as Γ^* is column stochastic. Then, by the first equation

in (5.7) we get

$$1 = \sum_{d\in\mathcal{D}} \Gamma_{bd}\Gamma^*_{db} \leq \gamma_b^+ \sum_{d\in\mathcal{D}} \Gamma^*_{db} \leq \gamma_b^+(1-\delta)\,,$$

which implies that

$$\gamma_b^+ \geq \frac{1}{1-\delta} = 1 + \frac{\delta}{1-\delta}\,.$$

On the other hand, if $\Gamma^*_{cb} < \delta$, for every c in $\mathcal{C}$ and b in $\mathcal{B}$, then, by Assumption 5.2, every column of $\Gamma^*_{\mathcal{DB}}$ contains two entries not smaller than δ, and, since $|\mathcal{B}| = |\mathcal{D}|$, so does at least one of its rows, i.e., there exists d in $\mathcal{D}$, and $b \neq i$ in $\mathcal{B}$ such that $\Gamma^*_{db} \geq \delta$ and $\Gamma^*_{di} \geq \delta$.

Since Γ^* is column stochastic, $\Gamma^*_{-db} = \sum_{j\neq d} \Gamma^*_{jb}$ and $\Gamma^*_{-di} = \sum_{j\neq d} \Gamma^*_{ji}$ satisfy

$$\Gamma^*_{-db} \leq 1-\delta, \quad \Gamma^*_{-di} \leq 1-\delta.$$

Moreover, the first equation in (5.7) yields

$$\Gamma_{bd}\Gamma^*_{db} + \gamma^+_{bd}\Gamma^*_{-db} \geq 1\,, \qquad \Gamma_{bd}\Gamma^*_{di} + \gamma^-_{bd}\Gamma^*_{-di} \leq 0\,. \tag{5.8}$$

Now, if $\Gamma_{bd} \leq 0$, then

$$1 \leq \Gamma_{bd}\Gamma^*_{db} + \gamma^+_{bd}\Gamma^*_{-db} \leq \gamma^+_{bd}\Gamma^*_{-db} \leq \gamma^+_{bd}(1-\delta),$$

so that

$$\gamma^+_{bd} \geq \frac{1}{1-\delta} = 1 + \frac{\delta}{1+\delta}\,.$$

Similarly, if $\gamma^+_{bd} \leq 0$, then

$$\Gamma_{bd} \geq 1 + \frac{\delta}{1+\delta}\,.$$

Conversely, if $\Gamma_{bd} > 0$ and $\gamma^+_{bd} > 0$, then the second inequality in (5.8) implies that $\gamma^-_{bd} < 0$. Substituting now in (5.8)

$$\Gamma^*_{-db} = 1 - \Gamma^*_{db} \quad \text{and} \quad \Gamma^*_{-di} = 1 - \Gamma^*_{di},$$

we get

$$\frac{1-\gamma^+_{bd}(1-\Gamma^*_{db})}{\Gamma^*_{db}} \leq -\gamma^-_{bd}\frac{1-\Gamma^*_{di}}{\Gamma^*_{di}}\,.$$

Here we can distinguish two cases. First, if $\gamma_{bd}^{+} \leq 1$ we retrieve

$$-\gamma_{bd}^{-} \frac{1-\Gamma_{di}^{*}}{\Gamma_{di}^{*}} \geq 1$$

which implies

$$-\gamma_{bd}^{-} \geq \frac{\delta}{1-\delta}\,.$$

Second, if $\gamma_{bd}^{+} > 1$ then necessarily, given that $\gamma_{bd}^{-} < 0$, the following inequality must hold

$$-\gamma_{bd}^{-} \geq \frac{\Gamma_{di}^{*}}{1-\Gamma_{di}^{*}} \frac{(1-\gamma_{bd}^{+}(1-\Gamma_{db}^{*}))_{+}}{\Gamma_{db}^{*}}\,.$$

Since $1 < \gamma_{bd}^{+} \leq \frac{1}{1-\Gamma_{db}^{*}}$, the above implies that $-\gamma_{bd}^{-} \geq \frac{\delta}{1-\delta}$.

□

Based on Lemma 5.1, the following result for the solution of Problem 5.1, in absence of noise, holds.

Proposition 5.1 (Noise free scenario)**.** *Let Assumptions 5.1 and 5.2 hold true. If* $\Xi = 0$ *and* $\varepsilon = 0$*, then Problem 5.1 admits a unique optimal solution*

$$(\widehat{\mathcal{S}}_0, \widehat{\Gamma}_0) = (\mathcal{S}^*, \Gamma^*)\,.$$

Proof. First, note that, for $\varepsilon = 0$, the first constraint in (5.4) is equivalent to (5.5). Since Assumption 5.1 implies that (5.5) cannot be satisfied by any Γ if $|\mathcal{S}| < k^*$, there cannot be any feasible pair $(\mathcal{S}, \Gamma)$ for Problem 5.1 with $|\mathcal{S}| < k^*$ and the $\mathcal{S}$ columns of X must be independent, i.e. $\text{rank}(X_{\mathcal{S}}) = k^*$. On the other hand, it follows from (5.2) that $(\mathcal{S}^*, \Gamma^*)$ is feasible for Problem 5.1 when $\Xi = 0$ and $\varepsilon = 0$. Therefore, $(\mathcal{S}^*, \Gamma^*)$ is an optimal solution for Problem 5.1.

Moreover, Lemma 5.1 implies that, if $|\mathcal{S}| = k^*$, $\mathcal{S} \neq \mathcal{S}^*$, then there is no Γ in $\mathbb{R}_{+}^{\mathcal{S}\times\mathcal{R}}$ s.t. $\|\Gamma\|_{\max} \leq 1$ and (5.5) is satisfied. Hence, every optimal solution $(\widehat{\mathcal{S}}_0, \widehat{\Gamma}_0)$ of Problem 5.1 is such that $\widehat{\mathcal{S}}_0 = \mathcal{S}^*$. Finally, observe that if $\widehat{\Gamma}_0$ in $\mathbb{R}_{+}^{\mathcal{S}^*\times\mathcal{R}^*}$ is such that $X_{\mathcal{R}^*} = X_{\mathcal{S}^*}\widehat{\Gamma}_0$, then $X_{\mathcal{S}^*}(\Gamma^* - \widehat{\Gamma}_0) = 0$, so that Assumption 5.1 implies that $\widehat{\Gamma}_0 = \Gamma^*$. This proves that $(\mathcal{S}^*, \Gamma^*)$ is the unique optimal solution for Problem 5.1. □

5.3 Noisy Scenario Analysis

Let us now move our focus on the noisy scenario. The following result provides sufficient conditions for every solution of Problem 5.1 to detect the true set of stubborn nodes $\mathcal{S}^*$ and accurately estimate the true influence matrix Γ^*.

To simplify the notation, in the following we will denote by

$$\Delta := \Delta(X, \mathcal{S}^*) = \frac{\|X_{\mathcal{R}^*}\|_2}{\sigma_{k^*}(X_{\mathcal{S}^*})},$$

i.e. the ratio between spectral norm of $X_{\mathcal{R}^*}$ and the minimum positive singular value of $X_{\mathcal{S}^*}$. Before presenting the main general result about the conditions which guarantee the exact detection of $\mathcal{S}^*$ and the performance guarantees on the estimate of Γ^* for Problem 5.1, we will explain two preliminary results focusing on the two problems separately.

5.3.1 Estimation of Number of Stubborn Nodes

Let's start analysing the conditions which guarantee the recovery of the correct number of stubborn nodes, i.e. the cardinality of subset $\mathcal{S}^*$. Preliminary, let us introduce a Lemma that will be fundamental for the subsequent Proposition.

Lemma 5.2. *If Assumption 5.1 holds true, then, given Y defined as in* (5.3)*,*

$$\|Y_{\mathcal{R}^*} - Y_{\mathcal{S}^*}\Gamma^*\|_2 \leq \|\Xi\|_2(1 + \|\Gamma^*\|_2), \qquad \|\Gamma^*\|_2 \leq \Delta. \tag{5.9}$$

Proof. It follows from (5.2) and (5.3) that

$$\begin{aligned}\|Y_{\mathcal{R}^*} - Y_{\mathcal{S}^*}\Gamma^*\|_2 &\leq \|Y_{\mathcal{R}^*} - X_{\mathcal{R}^*}\|_2 + \|X_{\mathcal{R}^*} - X_{\mathcal{S}^*}\Gamma^*\|_2 + \|X_{\mathcal{S}^*}\Gamma^* - Y_{\mathcal{S}^*}\Gamma^*\|_2 \\ &\leq \|\Xi\|_2(1 + \|\Gamma^*\|_2).\end{aligned}$$

Moreover,

$$\|X_{\mathcal{R}^*}\|_2 = \|X_{\mathcal{S}^*}\Gamma^*\|_2 \geq \sigma_{k^*}(X_{\mathcal{S}^*})\|\Gamma^*\|_2,$$

where the last inequality follows from the fact that $X_{\mathcal{S}^*}$ is left invertible and $\sigma_{k^*}(X_{\mathcal{S}^*}) = \|X_{\mathcal{S}^*}^{-1}\|_2^{-1}$. Thus,

$$\|\Gamma^*\|_2 \leq \Delta.$$

□

The following result determines conditions under which, for any solution of Problem 5.1, the number of stubborn nodes is neither over nor under-estimated.

Proposition 5.2. *Suppose Assumption 5.1 holds true and let* $(\widehat{\mathcal{S}}_\varepsilon, \widehat{\Gamma}_\varepsilon)$ *be a solution of Problem 5.1. Then*

(i) if $\varepsilon > \|\Xi\|_2(1+\Delta)$*, then* $|\widehat{\mathcal{S}}_\varepsilon| \leq k^*$*;*

(ii) if $\varepsilon < \sigma_{k^*}(X) - \|\Xi\|_2$*, then* $|\widehat{\mathcal{S}}_\varepsilon| \geq k^*$*.*

Proof. (i): It follows from (5.9) and the assumption on ε in Problem 5.1 that $\|Y_{\mathcal{R}^*} - Y_{\mathcal{S}^*}\Gamma^*\|_2 \leq \varepsilon$. Consequently, optimality of $\widehat{\mathcal{S}}_\varepsilon$ yields $|\widehat{\mathcal{S}}_\varepsilon| \leq |\mathcal{S}^*| = k^*$.

(ii): By contradiction, assume that $|\widehat{\mathcal{S}}_\varepsilon| < k^*$. We have that

$$\begin{aligned} \|Y_{\widehat{\mathcal{R}}_\varepsilon} - Y_{\widehat{\mathcal{S}}_\varepsilon}\widehat{\Gamma}_\varepsilon\|_2 &\geq \sigma_{|\widehat{\mathcal{S}}_\varepsilon|+1}(Y) \\ &\geq \sigma_{k^*}(Y) \\ &\geq \sigma_{k^*}(X) - \|\Xi\|_2 \\ &> \varepsilon\,. \end{aligned}$$

where the first inequality comes from low rank approximation in Lemma A.1 (Appendix A.1), the second since $|\widehat{\mathcal{S}}_\varepsilon| + 1 \leq k^*$ by the hypothesis on $|\widehat{\mathcal{S}}_\varepsilon|$, the third from Lemma A.2-(i) (Appendix A.1), and the last one by the hypothesis of the proposition.

This contradicts the assumption made on ε in Problem 5.1. □

We will now show under which conditions the number of stubborn is correctly detected and we will present some extra properties of the involved matrices.

Proposition 5.3. *Let Assumption 5.1 and 5.2 be satisfied. If*

$$\|\Xi\|_2 \leq \frac{\sigma_{k^*}(X)}{2(1+\Delta)} \tag{5.10}$$

$$\varepsilon \in [\|\Xi\|_2(1+\Delta), \sigma_{k^*}(X) - \|\Xi\|_2(1+\Delta)] \tag{5.11}$$

then every solution $(\widehat{\mathcal{S}}_\varepsilon, \widehat{\Gamma}_\varepsilon)$ *of Problem 5.1 is such that*

(i) $|\widehat{\mathcal{S}}_\varepsilon| = k^*$.

(ii) $\|X_{\widehat{\mathcal{R}}_\varepsilon} - X_{\widehat{\mathcal{S}}_\varepsilon}\widehat{\Gamma}_\varepsilon\|_2 \leq \varepsilon + \|\Xi\|_2(1+\Delta)$

(iii) The columns of $X_{\widehat{\mathcal{S}}_\varepsilon}$ *are linearly independent.*

Proof. (i) follows directly from Proposition 5.2.

(ii) Applying (5.3) and triangular inequality, we estimate:

$$\begin{aligned}\|X_{\widehat{\mathcal{R}}_\varepsilon} - X_{\widehat{\mathcal{S}}_\varepsilon}\widehat{\Gamma}_\varepsilon\|_2 &\leq \|Y_{\widehat{\mathcal{R}}_\varepsilon} - Y_{\widehat{\mathcal{S}}_\varepsilon}\widehat{\Gamma}_\varepsilon\|_2 + \|\Xi_{\widehat{\mathcal{R}}_\varepsilon} - \Xi_{\widehat{\mathcal{S}}_\varepsilon}\widehat{\Gamma}_\varepsilon\|_2 \\ &\leq \varepsilon + \|\Xi_{\widehat{\mathcal{R}}_\varepsilon}\|_2 + \|\Xi_{\widehat{\mathcal{S}}_\varepsilon}\|_2\|\widehat{\Gamma}_\varepsilon\|_2 \\ &\leq \varepsilon + \|\Xi\|_2(1+\|\widehat{\Gamma}_\varepsilon\|_2) \\ &\leq \varepsilon + \|\Xi\|_2(1+\Delta)\,,\end{aligned}$$

where the third inequality comes from Lemma A.2-(iii) in Appendix A.1 recalling that the spectral norm coincides with the largest spectral value, and the last inequality follows from the fact that since $|\widehat{\mathcal{S}}_\varepsilon| = |\mathcal{S}^*|$ and given that from Lemma 5.2, given ε as in (5.11), $(\mathcal{S}^*, \Gamma^*)$ is a feasible solution of Problem 5.1, then by optimality it must hold

$$\|\widehat{\Gamma}_\varepsilon\|_2 \leq \|\Gamma^*\|_2 \leq \Delta\,.$$

(iii) If the columns of $X_{\widehat{\mathcal{S}}_\varepsilon}$ were linearly dependent, then,

$$\|X_{\widehat{\mathcal{R}}_\varepsilon} - X_{\widehat{\mathcal{S}}_\varepsilon}\widehat{\Gamma}_\varepsilon\|_2 \geq \sigma_{k^\star}(X)\,.$$

This coupled with the inequality in (ii) contradicts the assumptions on ε. Finally, condition in (5.10) guarantees that interval (5.11) is non-empty. □

5.3.2 Estimation of Influence Matrix

Let us now present some results related to the estimation of the influence matrix Γ^*. Preliminary, let us introduce a Lemma that will be fundamental in the following.

Lemma 5.3. *Suppose to be under assumptions of Proposition 5.3, i.e. let Assumption 5.1 and 5.2 be satisfied and assume that*

$$\begin{aligned}&\|\Xi\|_2 \leq \tfrac{\sigma_{k^*}(X)}{2(1+\Delta)}, \\ &\varepsilon \in [\|\Xi\|_2(1+\Delta), \sigma_{k^*}(X) - \|\Xi\|_2(1+\Delta)]\,.\end{aligned}$$

Let $(\widehat{\mathcal{S}}_\varepsilon, \widehat{\Gamma}_\varepsilon)$ *be a solution of Problem 5.1, then*

$$\sigma_{k^*}(X_{\widehat{\mathcal{S}}_\varepsilon}) \geq \frac{\sigma_{k^*}(X) - \|\Xi\|_2(1+\Delta) - \varepsilon}{1+\Delta}. \tag{5.12}$$

Proof. Given Z be the best rank k^*-1 approximation of $X_{\widehat{\mathcal{S}}_\varepsilon}$, low rank approximation in Lemma A.1 - Appendix A.1 yields

$$\|X_{\widehat{\mathcal{S}}_\varepsilon} - Z\|_2 = \sigma_{k^*}(X_{\widehat{\mathcal{S}}_\varepsilon}).$$

Then,

$$\begin{aligned}
\sigma_{k^*}(X) &\leq \|[X_{\widehat{\mathcal{R}}_\varepsilon}\ X_{\widehat{\mathcal{S}}_\varepsilon}] - Z[\widehat{\Gamma}_\varepsilon\ I]\|_2 \\
&\leq \|X_{\widehat{\mathcal{R}}_\varepsilon} - X_{\widehat{\mathcal{S}}_\varepsilon}\widehat{\Gamma}_\varepsilon\|_2 + \|(X_{\widehat{\mathcal{S}}_\varepsilon} - Z)\|_2\|[\widehat{\Gamma}_\varepsilon\ I]\|_2 \\
&\leq \varepsilon + \|\Xi\|_2(1+\Delta) + \sigma_{k^*}(X_{\widehat{\mathcal{S}}_\varepsilon})(1+\Delta).
\end{aligned}$$

where the second inequality comes from triangular inequality and the last one from Proposition 5.3 - (ii) and from previously proven equality. □

Proposition 5.4. *Suppose to be under assumptions of Proposition 5.3, i.e. let Assumption 5.1 and 5.2 be satisfied and assume that*

$$\begin{aligned}
&\|\Xi\|_2 \leq \tfrac{\sigma_{k^*}(X)}{2(1+\Delta)}, \\
&\varepsilon \in [\|\Xi\|_2(1+\Delta), \sigma_{k^*}(X) - \|\Xi\|_2(1+\Delta)].
\end{aligned}$$

Let $(\widehat{\mathcal{S}}_\varepsilon, \widehat{\Gamma}_\varepsilon)$ *be a solution of Problem 5.1, then there exists a matrix* $\bar{\Gamma}$ *such that* $X_{\widehat{\mathcal{R}}_\varepsilon} = X_{\widehat{\mathcal{S}}_\varepsilon}\bar{\Gamma}$*, which satisfies*

$$\|\widehat{\Gamma}_\varepsilon - \bar{\Gamma}\|_2 \leq \frac{(1+\Delta)(\varepsilon + \|\Xi\|_2(1+\Delta))}{\sigma_{k^*}(X) - \|\Xi\|_2(1+\Delta) - \varepsilon}. \tag{5.13}$$

Proof. Existence of $\bar{\Gamma}$ follows from the fact that, thanks to Proposition 5.3, the rank of X is $|\widehat{\mathcal{S}}_\varepsilon| = k^*$ and the columns of $X_{\widehat{\mathcal{S}}_\varepsilon}$ are linearly independent.

Moreover, the following series of equations holds true

$$\begin{aligned}
Y_{\widehat{\mathcal{R}}_\varepsilon} - Y_{\widehat{\mathcal{S}}_\varepsilon}\widehat{\Gamma}_\varepsilon &= X_{\widehat{\mathcal{R}}_\varepsilon} + \Xi_{\widehat{\mathcal{R}}_\varepsilon} - (X_{\widehat{\mathcal{S}}_\varepsilon} + \Xi_{\widehat{\mathcal{S}}_\varepsilon})\widehat{\Gamma}_\varepsilon \\
&= X_{\widehat{\mathcal{S}}_\varepsilon}\bar{\Gamma} + \Xi_{\widehat{\mathcal{R}}_\varepsilon} - (X_{\widehat{\mathcal{S}}_\varepsilon} + \Xi_{\widehat{\mathcal{S}}_\varepsilon})\widehat{\Gamma}_\varepsilon,
\end{aligned}$$

from which, using triangular inequality and recalling that $\|Y_{\mathcal{R}^*} - Y_{\mathcal{S}^*}\Gamma^*\|_2 \leq \varepsilon$, we obtain

$$\|X_{\widehat{\mathcal{S}}_\varepsilon}(\bar{\Gamma} - \widehat{\Gamma}_\varepsilon)\|_2 \leq \varepsilon + \|\Xi_{\widehat{\mathcal{R}}_\varepsilon}\|_2 + \|\Xi_{\widehat{\mathcal{S}}_\varepsilon}\|_2 \|\widehat{\Gamma}_\varepsilon\|_2.$$

Since $X_{\widehat{\mathcal{S}}_\varepsilon}$ is left invertible and recalling that $\|\widehat{\Gamma}_\varepsilon\|_2 \leq \Delta$, it follows

$$\|\widehat{\Gamma}_\varepsilon - \bar{\Gamma}\|_2 \leq \frac{\varepsilon + \|\Xi\|_2(1+\Delta)}{\sigma_{k^*}(X_{\widehat{\mathcal{S}}_\varepsilon})}.$$

Finally, using (5.12), we conclude the proof. □

5.3.3 Analysis of the Optimal Solution

Collecting the results from Sections 5.3.1–5.3.2, we can now state a comprehensive theorem that guarantees the exact identification of the number of stubborn nodes and provides a quantification of the estimation error for the matrix Γ^*.

Theorem 5.1. *Let Assumption 5.1 and 5.2 hold true. If*

$$\|\Xi\|_2 \leq \frac{\gamma \sigma_{k^*}(X)}{2(1+\Delta)}, \tag{5.14}$$

where $\gamma = \frac{\delta}{(4-\delta)(1+\Delta)-3\delta\Delta}$, *and*

$$\|\Xi\|_2(1+\Delta) \leq \varepsilon \leq \gamma \sigma_{k^*}(X) - \|\Xi\|_2(1+\Delta), \tag{5.15}$$

then every optimal solution $(\widehat{\mathcal{S}}_\varepsilon, \widehat{\Gamma}_\varepsilon)$ *of Problem 5.1 is such that*

$$\widehat{\mathcal{S}}_\varepsilon = \mathcal{S}^*, \qquad \|\widehat{\Gamma}_\varepsilon - \Gamma^*\|_2 \leq \frac{\varepsilon + \|\Xi\|_2(1+\Delta)}{\sigma_{k^*}(X_{\mathcal{S}^*})}, \tag{5.16}$$

where $\mathcal{S}^*$ *is the true stubborn set and* Γ^* *is the true influence matrix.*

Proof. Notice that all assumptions of Proposition 5.3 hold true due to (5.14), (5.15) and the fact that $\gamma \leq 1$. In particular, there exists a matrix $\bar{\Gamma}$ as in Proposition 5.4 satisfying (5.13). Using the upper bound on ε in (5.15) inside (5.13), it holds

$$\|\widehat{\Gamma}_\varepsilon - \bar{\Gamma}\|_2 \leq \frac{\gamma(1+\Delta)\sigma_{k^*}(X)}{(1-\gamma)\sigma_{k^*}(X)} < \frac{\delta}{2(1-\delta)} \tag{5.17}$$

where last inequality follows from the way γ has been defined.

If, by contradiction, $\widehat{\mathcal{S}}_\varepsilon \neq \mathcal{S}^*$, given that from Proposition 5.3 $X_{\widehat{\mathcal{S}}_\varepsilon}$ has linearly independent columns, then from Lemma 5.1 and the fact that $\widehat{\Gamma}_\varepsilon$ has all entries in $[0,1]$, we would have

$$\|\widehat{\Gamma}_\varepsilon - \bar{\Gamma}\|_2 \geq \|\widehat{\Gamma}_\varepsilon - \bar{\Gamma}\|_{\max} \geq \frac{\delta}{2(1-\delta)}$$

contradicting (5.17).

Then, $\widehat{\mathcal{S}}_\varepsilon = \mathcal{S}^*$. It follows that

$$\begin{aligned} Y_{\mathcal{R}^*} - Y_{\mathcal{S}^*}\widehat{\Gamma}_\varepsilon &= X_{\mathcal{R}^*} + \Xi_{\mathcal{R}^*} - (X_{\mathcal{S}^*} + \Xi_{\mathcal{S}^*})\widehat{\Gamma}_\varepsilon \\ &= X_{\mathcal{S}^*}\Gamma^* + \Xi_{\mathcal{R}^*} - (X_{\mathcal{S}^*} + \Xi_{\mathcal{S}^*})\widehat{\Gamma}_\varepsilon\,, \end{aligned}$$

so that

$$\|X_{\mathcal{S}^*}(\Gamma^* - \widehat{\Gamma}_\varepsilon)\|_2 \leq \varepsilon + \|\Xi_{\mathcal{R}^*}\|_2 + \|\Xi_{\mathcal{S}^*}\|_2\|\widehat{\Gamma}_\varepsilon\|_2\,.$$

The fact that $X_{\mathcal{S}^*}$ is left-invertible and $\|\widehat{\Gamma}_\varepsilon\|_2 \leq \|\Gamma^*\|_2 \leq \Delta$ now yield (5.16). □

Notice that (5.14) implies that the range of values of the admissible tolerance ε in (5.15) is nonempty. Thus, Theorem 5.1 guarantees that for sufficiently small noise, we can find tolerance values ε for which every optimal solution of Problem 1 correctly identifies the subset $\mathcal{S}^*$ and yields an influence matrix $\widehat{\Gamma}_\varepsilon$ close to Γ^*.

Clearly, a careful choice for the tolerance ε is crucial for the applicability of the stated Theorem. Focusing on the admissible range in (5.15), we note that its selection generally requires additional information about the stubborn set, as reflected in both k^* and Δ. However, as illustrated in the following example, by leveraging suitable results from linear algebra, alternative bounds can be derived based on the specific application context, thereby extending the applicability of the proposed Theorem.

Example 5.1. *Let Ξ have i.i.d. Gaussian random entries with zero mean and variance ξ^2. Then, by Corollary A.2 in Appendix A.1, $\|\Xi\|_2 \leq 3\xi\sqrt{n}$ with high probability as the number of nodes n grows large. Let also $X_{\mathcal{S}^*}$ have i.i.d. sub-Gaussian random entries with zero mean and unitary variance and assume that $|\mathcal{S}^*|/n \to \theta \in (0,1)$ and $|\mathcal{T}|/n \to \beta > \theta$. Then, Theorem A.1 in Appendix A.1 implies that*

$$(\sqrt{\beta}-\sqrt{\theta})\sqrt{n} + o(\sqrt{n}) \leq \sigma_{k^*}(X_{\mathcal{S}^*}) \leq \|X_{\mathcal{S}^*}\|_2 \leq (\sqrt{\beta}+\sqrt{\theta})\sqrt{n} + o(\sqrt{n})\,,$$

as $n \to +\infty$, so that with high probability

$$\Delta = \frac{\|X_{\mathcal{R}^*}\|_2}{\sigma_{k^*}(X_{\mathcal{S}^*})} \leq \frac{\|X_{\mathcal{S}^*}\|_2 \|\Gamma^*\|_2}{\sigma_{k^*}(X_{\mathcal{S}^*})} \leq \frac{\sqrt{\beta}+\sqrt{\theta}}{\sqrt{\beta}-\sqrt{\theta}} \|\Gamma^*\|_2 (1+o(1)) .$$

It follows that the right hand side of (5.14) *satisfies*

$$\frac{\gamma \sigma_{k^*}(X)}{2(1+\Delta)} \geq \frac{\delta \sigma_{k^*}(X_{\mathcal{S}^*})}{2(4-\delta)(1+\Delta)^2} \geq \frac{\delta(\sqrt{\beta}-\sqrt{\theta})\sqrt{n}}{2(4-\delta)(1+\Delta)^2}(1+o(1)) .$$

Therefore, (5.14) *is satisfied with high probability if*

$$\xi \leq \frac{\delta(\sqrt{\beta}-\sqrt{\theta})^3}{6(4-\delta)(\sqrt{\beta}-\sqrt{\theta}+(\sqrt{\beta}+\sqrt{\theta})\|\Gamma^*\|_2)^2} ,$$

subjected to $\|\Gamma^*\|_2$ being bounded. Due to Γ^* stochasticity,

$$\|\Gamma^*\|_2^2 \leq \|\Gamma^*\|_1 \|\Gamma^*\|_\infty = \max_{s \in \mathcal{S}^*} \sum_{j \in \mathcal{R}^*} \Gamma^*_{js} ,$$

so that $\|\Gamma^*\|_2$ is guaranteed to remain bounded whenever the total influence of every stubborn agent remains bounded.

Remark 5.1 (Observation at finite time)**.** *The case when the opinions are observed at finite time $t < +\infty$ can be cast into our framework by letting $\Xi = Y - X$ be the difference between the opinion matrix at time t and the equilibrium opinion matrix. As it is known [91] that $\|\Xi\|_2 \leq \lambda_{\max}^{t_{\text{obs}}} \|X(0)-X\|_2$, where $\lambda_{\max} < 1$ denotes the dominant eigenvalue of the substochastic matrix $P_{\mathcal{R}^*\mathcal{R}^*}$, one can readily derive sufficient conditions on t to ensure that* (5.14) *holds true, thus enabling the application of Theorem 5.1. Specifically, Theorem 5.1 is guaranteed if*

$$\lambda_{\max}^{t_{\text{obs}}} \|X(0)-X\|_2 \leq \frac{\gamma \sigma_{k^*}(X)}{2(1+\Delta)},$$

which implies

$$t_{obs} \geq \frac{1}{\ln(\lambda_{max})} \ln\left(\frac{\gamma \sigma_{k^*}(X)}{2(1+\Delta)\|X(0)-X\|_2}\right) .$$

5.3.4 Algorithmic Analysis: Interpolative Decomposition

It is remarkable to notice how the Problem 5.1 is a computational expensive problem due to an external combinatorial optimization over a domain of exponential size $2^{|\mathcal{V}|}$. Consequently, selecting an efficient algorithm to solve it is essential in order to ensure good performance with a limited computational cost.

In particular, drawing from existing literature [92–94], we address the problem of matrix decomposition using the Interpolative Decomposition (ID), a method used to obtain approximate factorizations of low-rank matrices. For the sake of completeness, let us here briefly recall some theory related to this matrix decomposition technique.

Definition 5.1 (Interpolative Decomposition [92])**.** *Given an $m \times n$ matrix X, an $m \times k$ matrix A whose columns are a subset of those of X, and a $k \times n$ matrix C, s.t.*

- *some size-k subset of those of C form the $k \times K$ identity matrix, and*
- *no entry of C has absolute value greater than 2,*

*AC constitutes an **interpolative decomposition** of X.*

The parallelism with the problem of interest is clear simply referring to the definition in (5.2). Indeed, given Π a proper permutation matrix which reorders the columns to place on the left the ones corresponding to $\mathcal{S}^*$, it holds

$$X\Pi = [X_{\mathcal{S}^*} \quad X_{\mathcal{R}^*}] = [X_{\mathcal{S}^*} \quad X_{\mathcal{S}^*}\Gamma^*] = X_{\mathcal{S}^*}[I \quad \Gamma^*].$$

Look at Fig.5.1 for a sketch of the decomposition idea.

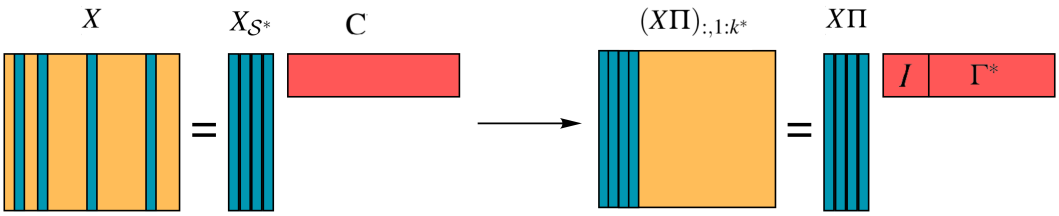


Fig. 5.1 Sketch of the idea behind Interpolative Decomposition

Therefore, the problem reduces to identifying a suitable interpolative decomposition of the final opinion matrix X. Specifically, we adopt the ID algorithm proposed in [92] which enables optimization over the domain defined in Eq. (5.4) imposing

non-negativity constraints on the matrix Γ and succeed in solving the optimization problem requiring only $O(|\mathcal{S}||\mathcal{R}||\mathcal{T}|)$ floating-point operations [93].

5.4 Examples and Simulations

Here we illustrate some numerical simulations to corroborate our theoretical results. The method was implemented using the *interpolative* module from Python's *scipy.linalg* library [94–96]. The metrics employed to assess the accuracy of the prediction include:

- relative error on Γ^* estimate, defined as

$$\mathsf{err} := \frac{||\Pi[I \quad \widehat{\Gamma}_\varepsilon] - [I \quad \Gamma^*]||_2}{||[I \quad \Gamma^*||_2}$$

  where Π is a permutation matrix which reorders the columns to place on the left the ones corresponding to $\mathcal{S}^*$.

- true positive rate (or sensitivity) and false positive rates

$$\mathsf{TPR} = \frac{|\mathcal{S}^* \cap \widehat{\mathcal{S}}_\varepsilon|}{|\mathcal{S}^*|} \quad \mathsf{FPR} = \frac{|\mathcal{R}^* \cap \widehat{\mathcal{S}}_\varepsilon|}{|\mathcal{R}^*|}.$$

We consider two scenarios for our simulations. In the first one we assume the observations to be taken at steady state, while in the second observations are taken at finite time. Without loss of generality, in all simulations initial opinions are randomly generated uniformly in the range $[0,1]$. We indicate with n the number of nodes, k^* the number of stubborn agents, and m the number of observations. In alle the simulations we consider a minimum number of observations equal to the cardinality of the stubborn nodes set.

5.4.1 Watts-Strogatz Network with Noisy Observations

We consider a French-DeGroot influence system over a Watts-Strogatz random graph, where observations are taken at steady state and are corrupted by Gaussian noise with zero mean and variance ξ^2.

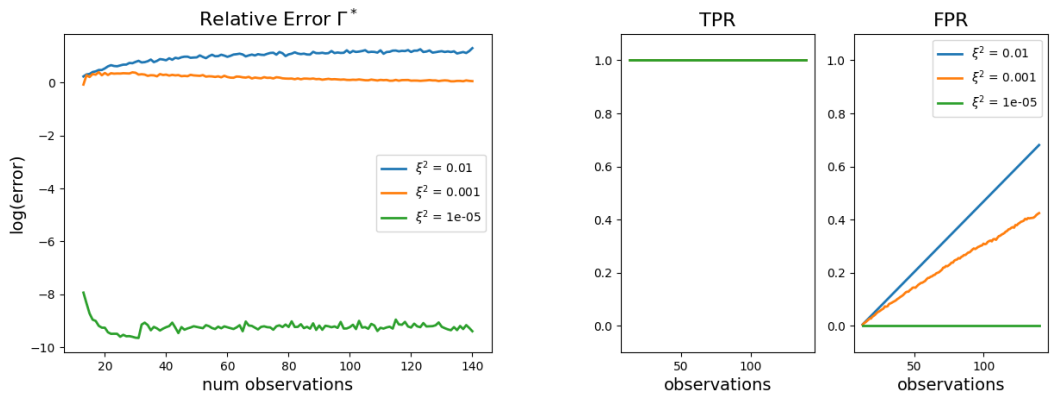


Fig. 5.2 Watts-Strogatz graph with $n = 200$, $k^* = 12$. Additional noise $\Xi \sim N(0, \xi^2)$ with $\xi^2 \in \{10^{-2}, 10^{-3}, 10^{-5}\}$. On the left is represented the relative error on Γ^*. On the right the two plots represents respectively the TPR, i.e. true positive rate, and the FPR, i.e. false positive rate, for stubborn nodes detection. In the TPR plot, all curves coincide, indicating that the true positives are consistently detected regardless of the value of ξ^2.

In the first set of simulations, we have chosen the size equal to $n = 200$ nodes with $k^* = 12$ stubborn nodes. Figure 5.2 displays the indices err, TPR, and FPR, as function both of variance ξ^2 and number of observations m, for a fixed tolerance ε satisfying (5.15). As expected from Theorem 5.1, we notice the existence of a threshold $\bar{\xi} > 0$ such that if $\xi < \bar{\xi}$ the algorithm succeeds both in the estimation of Γ^* and in the correct detection of stubborn nodes $\mathcal{S}^*$. Conversely, the stubborn nodes are overestimated and the relative error on Γ^* prediction is higher (see orange and blue lines in Figure 5.2). Despite this, since the $\log(\text{err})$ remains close to zero without significant growth, the simulations highlight the robustness of the proposed method even under larger perturbations. Finally, we observe that in case of high noise on data the stubborn nodes are overestimated (see the two indices TPR and FPR in Figure 5.2): all stubborn nodes are indeed correctly detected (no false negatives) while some regular nodes have been misclassified (false positive).

Figure 5.3 focuses on the dependance from the choice of tolerance ε. Fixed $\xi = 10^{-5}$ we compute numerically the bounds for ε in (5.15) which reveals to be equal to $[10^{-3}, 10^{-2}]$. In order to analyse the performance as function of ε we let it varying in $\{10^{-5}, 10^{-3}, 10^{-2}, 0.5\}$. As expected from Theorem 5.1, if we choose a non admissible ε (either too small or too high) the prediction error on Γ^* is non negligible. If we focus on stubborn detection when the chosen tolerance ε is too small (red line in Figure), the method overstimates the number of stubborn nodes;

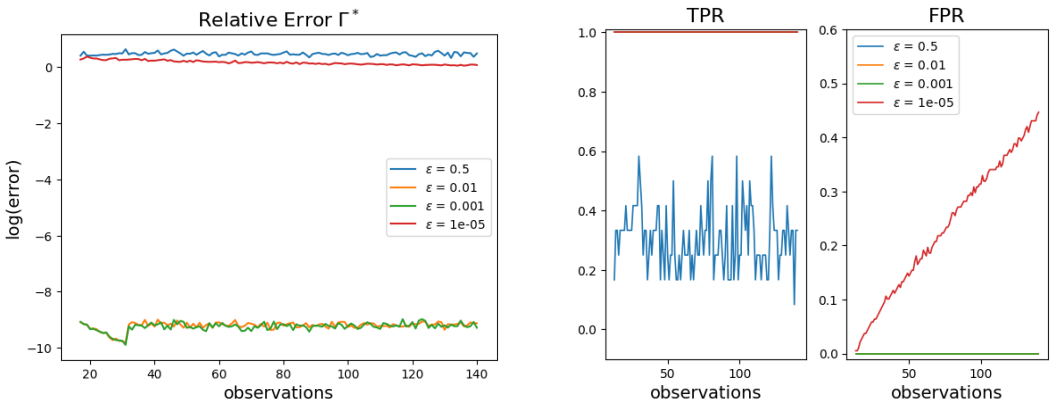


Fig. 5.3 Watts-Strogatz graph with $n = 200$, $k^* = 12$. Additional noise $\Xi \sim N(0, \xi^2)$. Comparison of the performances as function of for $\varepsilon \in \{10^{-5}, 10^{-3}, 10^{-2}, 0.5\}$. On the left the relative error on Γ^*. On the right the True Positive Rate TPR and the False Positive Rate FPR for stubborn nodes detection. Notice that in the TPR plot the orange, green and red lines coincides , while in FPR plot the green, orange and blue lines coincides.

conversely, if ε is too high (blue line in Figure), stubborn nodes are underestimated. Consequently, these simulations corroborate our theoretical findings emphasizing the critical importance of precision parameter ε calibration in relation to the variance ξ^2 to ensure optimal performance and accurate node detection within the network.

Finally, the plot in Figure 5.4 shows the performance of the method with respect to network size n. For a fixed fraction of stubborn nodes equal to 20% of the nodes, the plot shows the behavior of the relative error as a function of the number of nodes for different percentages of observations. The simulation corroborates our approach highlighting its scalability with respect to network size.

5.4.2 Finite-Time Observations Scenario

Here we collect opinions over a certain time interval for a Watts-Strogatz random graph with $n = 100$ nodes and $k^* = 12$ stubborn nodes. For the sake of simplicity, we assume no observation noise. Figure 5.5 illustrates the proposed method's performance, i.e. the average relative error on the estimation Γ^* and the average false positive rate as function of different parameters. The color intensity of each cell represents the magnitude of the error, where blue color indicates success and red color highlights error. We limit the time window shown to the interval $15 \leq t \leq 25$

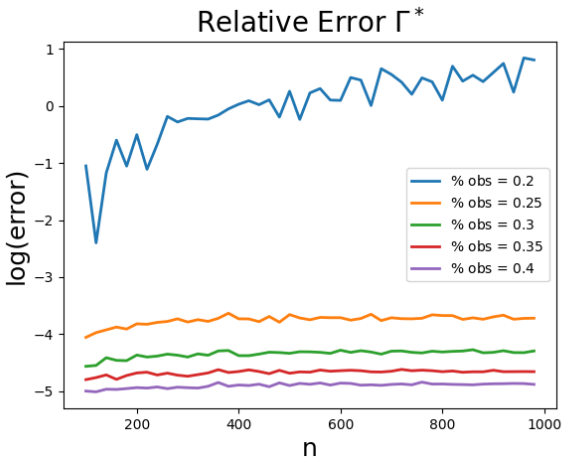


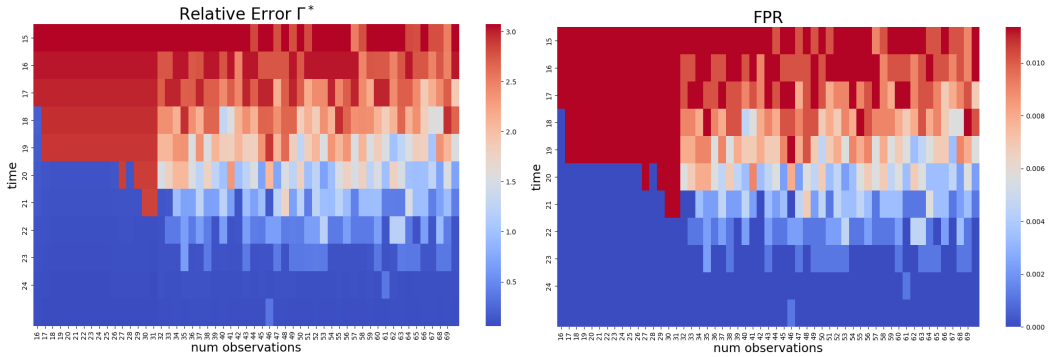
Fig. 5.4 Relative error as a function of size of the network n with fixed fraction of stubborn nodes $k^*/n = 0.2$ and percentage of observation $m/n \in \{0.2, 0.25, 0.3, 0.35, 0.4\}$. For representative reasons on the y-axis is represented the logarithm of the corresponding error.

observing that for $t > 24$ the behavior is comparable with the equilibrium one. Consistently with Remark 5.1, as the observation time increases, the method's performance improves. The TPR is not displayed in figure as all stubborn nodes are correctly detected. The FPR error shown in Fig. 5.5-(b) can instead be interpreted in terms of residual influence among regular agents far from the equilibrium.

5.4.3 Partial Observations Scenario

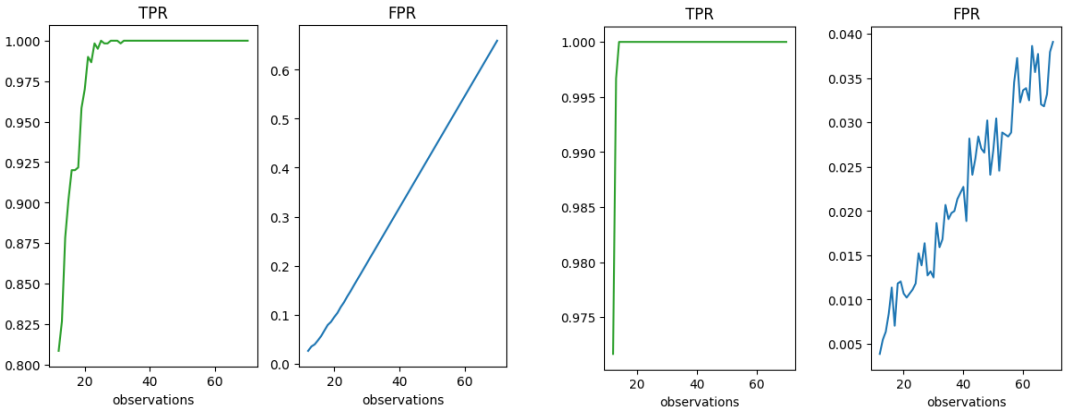
Another problem of interest concerns partial opinion observation, which reflects many real-world scenarios in which the full state matrix X is not observable for all nodes. This lack of information can be addressed through various matrix completion strategies. In this section, we present a series of illustrative examples. All simulations are conducted on a Watts–Strogatz graph with $n = 100$ nodes, among which 12 are designated as stubborn agents. We compare the results under varying levels of missing information. Specifically, we denote by p the probability that a given observation is missing.

As a first approach, given that each $X_{li} \in [0,1]$, for every $i \in \mathcal{V}$ and $l \in \mathcal{T}$, a simple approximation consists in assigning the missing values of X to the mean value, i.e., 0.5. The results of the simulations, shown in Figure 5.6, highlight a high rate of false positives in the detection of stubborn agents.



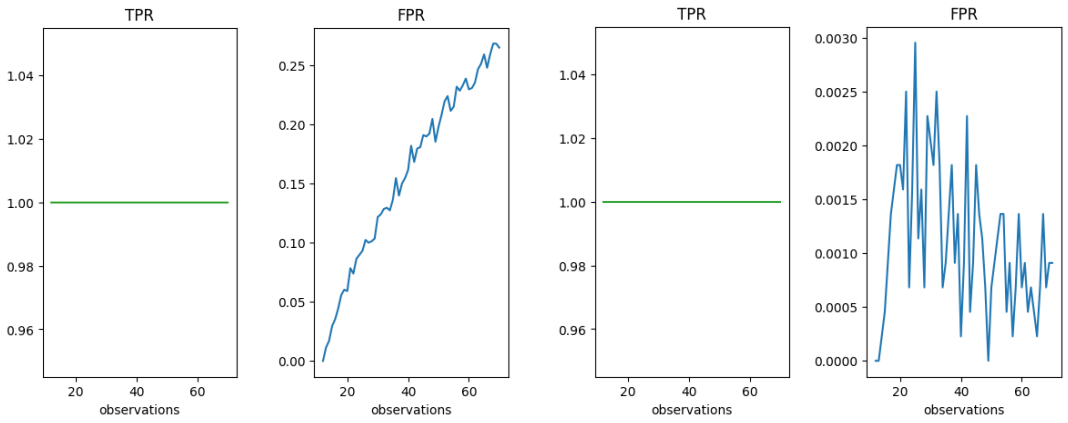
(a) (b)

Fig. 5.5 Watts-Strogatz graph with $n = 100$, $k^* = 12$. (a) Relative error on Γ^* estimation; (b) FPR for stubborn nodes detection as function of number of observations m, chosen such that $m > k^*$, and of the time of observation t_{obs}.



(a) $p = 0.05$ (b) $p = 0.0005$

Fig. 5.6 Stubborn detection for different values of probability p of missing information. Adjustment criteria: substitute missing values with average opinion, i.e. $X = 0.5$.



(a) $p = 0.05$ (b) $p = 0.0005$

Fig. 5.7 Stubborn detection for different values of probability p of missing information. Adjustment criteria: substitute missing values with network average opinion on the topic.

Alternatively, the unknown values can be imputed by assigning each missing entry the empirical average of the opinions of the remaining observable nodes on the same topic, thus preserving the network's average opinion. Formally, for a missing X_{li} we approximate it by

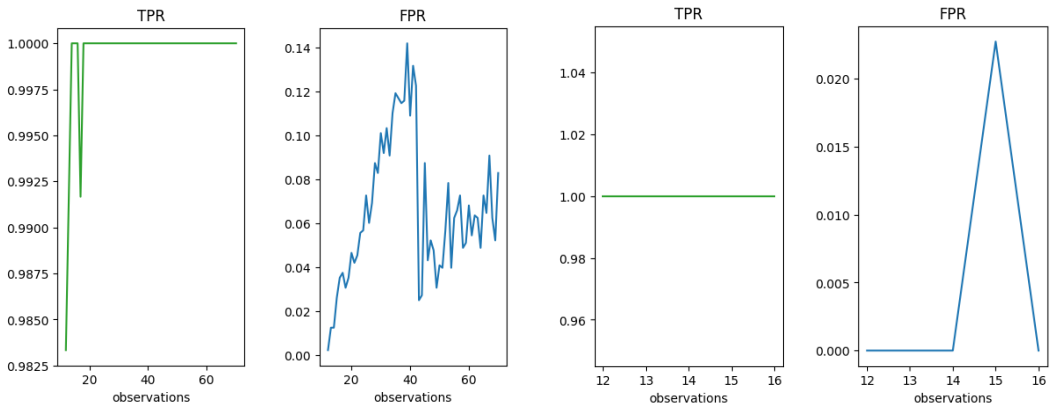
$$\tilde{X}_{li} = \sum_{j \in \mathcal{O}_l \subseteq \mathcal{V}} \frac{X_{lj}}{|\mathcal{O}_l|}$$

where $\mathcal{O}_l \subseteq \mathcal{V}$ is the set of available opinions on topic $l \in \mathcal{T}$. The results obtained with this strategy show a clear reduction in the false positive rate (FPR), as highlighted in Figure 5.7, thus outperforming the naive mean imputation baseline presented earlier.

A refined variant of mean-based imputation can be considered by incorporating topic-specific deviations from the mean. Specifically, this method accounts for how much a node's opinion typically deviates from the average across different topics. Formally, for a missing X_{li} we approximate it by

$$\tilde{X}_{li} = \frac{1}{|\mathcal{O}_l|} \sum_{j \in \mathcal{O}_l \subseteq \mathcal{V}} X_{lj} \left(1 + \max_{k \in \mathcal{T}} \frac{X_{kj} - \text{avg}(X_{k,:})}{X_{kj}} \right).$$

This correction factor introduces a node-specific amplification based on historical bias across topics, allowing the approximation to reflect more realistic heterogeneity in node behavior. As shown in Figure 5.8a, this method yields improved performance compared to the classical mean-based strategy.



(a) Adjusted mean value (b) Nuclear norm optimization

Fig. 5.8 Stubborn detection for $p = 0.05$ for more complex matrix completion methods.

Finally, more advanced matrix completion techniques can be employed, whose performance strongly depend on the structure of the specific application scenario. A widely adopted approach for recovering low-rank matrices, originally motivated by problems such as the Netflix recommendation system, is based on rank minimization, formulated as follows [97]:

$$\min_{\tilde{X} \in \mathbb{R}^{\mathcal{T} \times \mathcal{V}}} \text{rank}(\tilde{X}) \quad \text{s.t.} \quad \mathcal{A}_{\mathcal{O}}(\tilde{X}) = \mathcal{A}_{\mathcal{O}}(X)$$

where $\mathcal{A}_{\mathcal{O}}(X) = (X_{ij \in \mathcal{O}})$ denotes the observed elements of true matrix X.

Since the rank minimization problem is NP-hard, a common strategy is to solve its convex relaxation by minimizing the nuclear norm (also known as the trace norm), defined as:

$$||X||_* = tr((XX')^{1/2}).$$

In our experiments, this method was implemented using the *NuclearNormMinimization* function from the *fancyimpute* Python library. The results, summarized in Figure 5.8b, demonstrate a clear improvement in detection performance over the previously discussed heuristics. However, this gain in accuracy comes at the expense of significantly increased computational cost and runtime, which may limit its applicability in large-scale or real-time settings.

CHAPTER 6

Conclusions and Future Research

6.1 Conclusions

In this thesis, we focused on two relevant problems in the field of opinion dynamics, although, as discussed in the previous chapters, many of the presented results can be readily generalized to other application contexts.

In Part I, we analyzed a subset selection problem both from a general perspective and within the specific framework of opinion dynamics models. We derived an explicit formulation for the variance reduction function $F(\mathcal{K})$ under an arbitrary distribution of the observable variables. Furthermore, due to the computational complexity of the problem, we developed a suitable greedy algorithm that significantly reduces the computational cost required for optimal subset identification. We proved that, under specific assumptions on the model precision matrix, the variance reduction function is submodular, thus satisfying the hypotheses of Nemhauser's Theorem [80] for the performance guarantee of greedy solutions. In comparison with existing literature [34], we identified sufficient conditions that not only allow easier verification but also offer clear interpretations from a graph-theoretic perspective. Moreover, we showed that even when the function is not strictly submodular, the

approximate greedy solution still provides a close estimation of the optimal one, due to the function exhibiting approximate submodularity, a concept introduced in [35]. As additional contribution, we provided more detailed and rigorous proofs of the theoretical results originally outlined in [35], adapting them to our specific context of application. In the case of the French-DeGroot model with random noise and deterministic external input, we proved that the assumptions required to guarantee submodularity are extremely mild, requiring only the influence matrix to be reversible, condition satisfied by a wide variety of real-world networks. By comparing the obtained results with well-known centrality measures, we highlighted how the notion of "most informative" nodes introduced for the problem of interest diverges from that of "most influential" nodes associated with classical Bonacich centrality, thereby underscoring the relevance of the newly proposed measure. Finally, as an alternative strategy to reduce computational costs even when submodularity does not hold, in the specific context of the Friedkin–Johnsen model we proposed a set of applicable criteria for two limiting cases, namely, networks characterized by high and low openness to interactions. The results, derived through Taylor expansions of the functions of interest, also allow for insightful interpretations in terms of graph structure, and their effectiveness is supported by experimental evidence.

In Part II, the focus shifts to a detection problem. In the specific context of opinion dynamics, as extensively discussed in the introduction, a crucial role is played by stubborn agents, making their accurate detection a fundamental task. Formulating the problem as a low-rank factorization and exploiting Interpolative Decomposition techniques from [94, 95], we developed a method capable of precisely detecting stubborn nodes and reliably estimating the influence matrix based on noisy network observations at steady state across different topics. Moreover, we provided a detailed theoretical analysis of the method's performance and robustness. We also demonstrated that the results remain valid when observations are collected at finite time, where the deviation from the steady state can be incorporated as an error term. The applicability of the proposed approach extends beyond the French-DeGroot model, encompassing a broader class of opinion dynamics models where a small number of degrees of freedom are sufficient to characterize steady-state opinions and observed opinions are sufficiently close to these steady states. Finally, even nonlinear relationship between initial and final opinions can be interpreted through a linear approximation, with nonlinear effects captured within the error term.

Both presented problems lay the groundwork for a broader and more general problem formulation and analysis that extends beyond the specific context of application considered here.

6.2 Future Research

Still within the context of opinion dynamics, Chapter 4 opens the door to several potential extensions. On one hand, the model can be generalized to incorporate the joint presence of stochastic behavior in stubborn agents along with additional random noise in the communication process. On the other hand, attention can be shifted toward scenarios with limited prior knowledge of the network, examining how noisy or partial information about the network structure influences the objective function of interest. Furthermore, an additional line of investigation could focus on alternative opinion dynamics models and their impact on the selection process. In particular, models involving higher-order networks are currently attracting increasing attention within both complex systems and control communities [98–100]. Since a link to such structures has already emerged in this thesis, specifically in the interpretation of single-node selection for the French-DeGroot model, it would be of significant interest to further investigate this problem applying it to our newly proposed Friedkin-Johnsen model with higher-order neighbours [56].

In a more general framework, as highlighted in Chapter 3, the problem formulation naturally extends to a wide range of application contexts. For example, further studies could explore cybersecurity applications, helping with positioning of network sensors to evaluate the overall security state, to avoid the risk of full network compromise. In addition, another area of growing interest, particularly within the control and game theory communities, is Federated Learning [101], a decentralized machine learning technique where multiple clients collaboratively train a shared model while keeping their data local and only exchanging model updates, not raw data. However, since local training on each client can be computationally expensive, updating all clients at each iteration becomes unfeasible. Therefore, our proposed methods could be effectively applied in this domain to selectively choose a subset of clients for model updates at each step, optimizing the training process, minimizing the communication costs while still guaranteeing that data privacy is preserved.

Considering the detection problem discussed in Chapter 5, we recall that, for simplicity of exposition, the scenario was restricted to agents communicating synchronously over a static communication network. However, the proposed model can be naturally extended to accommodate asynchronous communications [51], as well as influence networks characterized by continuous transitions toward oligarchic structures, where social power increasingly concentrates within a small group [50]. Further natural extensions within the opinion dynamics framework include the exploration of how detection bounds depend on the underlying network topology, as well as the analysis of how partial information impacts the overall method performance. Finally, given the potential relevance of the proposed method in practical scenarios, it would be of significant interest to apply the framework to real-world datasets, and to compare its performance against state-of-the-art approaches, which typically rely on Artificial Intelligence techniques designed to handle large-scale data analysis.

References

[1] Chiara Ravazzi, Fabrizio Dabbene, Constantino Lagoa, and Anton V Proskurnikov. Learning hidden influences in large-scale dynamical social networks: A data-driven sparsity-based approach, in memory of Roberto Tempo. *IEEE Control Syst. Mag.*, 41(5):61–103, 2021.

[2] Hunt Allcott and Matthew Gentzkow. Social media and fake news in the 2016 election. *Journal of Economic Perspectives*, 31(2):211–36, May 2017.

[3] Dror Walter, Yotam Ophir, and Kathleen Hall Jamieson. Russian twitter accounts and the partisan polarization of vaccine discourse, 2015–2017. *American Journal of Public Health*, 110(5):718–724, 2020. PMID: 32191516.

[4] David A. Broniatowski, Amelia M. Jamison, SiHua Qi, Lulwah AlKulaib, Tao Chen, Adrian Benton, Sandra C. Quinn, and Mark Dredze. Weaponized health communication: Twitter bots and russian trolls amplify the vaccine debate. *American Journal of Public Health*, 108(10):1378–1384, 2018. PMID: 30138075.

[5] Emilio Ferrara. What types of covid-19 conspiracies are populated by twitter bots? *First Monday*, 25(6), May 2020.

[6] Michael Rossetti and Tauhid Zaman. Bots, disinformation, and the first impeachment of U.S. President Donald Trump. *Plos One*, 18(5), May 2023.

[7] N. Guenon des Mesnards, David Scott Hunter, Zakaria el Hjouji, and Tauhid Zaman. Detecting bots and assessing their impact in social networks. *Operations Research*, 70(1):1–22, 2021.

[8] L.C. Freeman. *The Development of Social Network Analysis: A Study in the Sociology of Science*. Empirical Press, 2004.

[9] David Lazer, Alex Pentland, Lada Adamic, Sinan Aral, Albert-László Barabási, Devon Brewer, Nicholas Christakis, Noshir Contractor, James Fowler, Myron Gutmann, Tony Jebara, Gary King, Michael Macy, Deb Roy, and Marshall Van Alstyne. Computational social science. *Science*, 323(5915):721–723, 2009.

[10] N. Friedkin. The problem of social control and coordination of complex systems in sociology: A look at the community cleavage problem. *IEEE Control Syst. Mag.*, 35(3):40–51, 2015.

[11] Loretta Mastroeni, Pierluigi Vellucci, and Maurizio Naldi. Agent-based models for opinion formation: A bibliographic survey. *IEEE Access*, 7:58836–58848, 2019.

[12] Anton V. Proskurnikov and Roberto Tempo. A tutorial on modeling and analysis of dynamic social networks. part ii. *Annual Reviews in Control*, 45:166–190, 2018.

[13] Michel Grabisch and Agnieszka Rusinowska. A survey on nonstrategic models of opinion dynamics. *Games*, 11(4):65, 2020.

[14] B. D. O. Anderson and M. Ye. Recent advances in the modelling and analysis of opinion dynamics on influence networks. *Int. J. Autom. Comput.*, 16:129–149, 2019.

[15] John R. P. French. A formal theory of social power. *Psychological Review*, 63(3):181–194, 1956.

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

[17] Keith Lehrer. When rational disagreement is impossible. *Noûs*, 10(3):327–332, 1976.

[18] Noah E. Friedkin, Anton Proskurnikov, and Francesco Bullo. Group dynamics on multidimensional object threat appraisals. *Soc. Netw.*, 65:157–167, 2021.

[19] Károly Takács, Andreas Flache, and Michael Mäs. Discrepancy and disliking do not induce negative opinion shifts. *PLOS ONE*, 11(6):e0157948, 2016.

[20] Ivan V. Kozitsin. Formal models of opinion formation and their application to real data: Evidence from online social networks. *J. Math. Sociol.*, 46(2):120–147, 2022.

[21] Chiara Ravazzi, Paolo Frasca, Roberto Tempo, and Hideaki Ishii. Ergodic randomized algorithms and dynamics over networks. *IEEE Transactions on Control of Network Systems*, 2(1):78–87, 2015.

[22] Ming Cao, A. Stephen Morse, and Brian D. O. Anderson. Reaching a consensus in a dynamically changing environment: A graphical approach. *SIAM Journal on Control and Optimization*, 47(2):575–600, 2008.

[23] Wei Ren and R.W. Beard. Consensus seeking in multiagent systems under dynamically changing interaction topologies. *IEEE Transactions on Automatic Control*, 50(5):655–661, 2005.

[24] Roberta Raineri, Giacomo Como, Fabio Fagnani, Mengbin Ye, and Lorenzo Zino. On controlling a coevolutionary model of actions and opinions. *Proc. 63rd IEEE Conf. Decis. Control*, pages 4550–4555, 2024.

[25] Roberta Raineri, Giacomo Como, Fabio Fagnani, Mengbin Ye, and Lorenzo Zino. Controlling a social network of individuals with coevolving actions and opinions, 2025.

[26] Mauro Mobilia. Does a single zealot affect an infinite group of voters? *Phys. Rev. Lett.*, 91:028701, Jul 2003.

[27] Daron Acemoğlu, Giacomo Como, Fabio Fagnani, and Asuman Ozdaglar. Opinion fluctuations and disagreement in social networks. *Mathematics of Operations Research*, 38(1):1–27, 2013.

[28] Sissi Xiaoxiao Wu, Hoi-To Wai, and Anna Scaglione. Estimating social opinion dynamics models from voting records. *IEEE Transactions on Signal Processing*, 66(16):4193–4206, 2018.

[29] Noah Friedkin, Peng Jia, and Francesco Bullo. A theory of the evolution of social power: Natural trajectories of interpersonal influence systems along issue sequences. *Sociological Science*, 3:444–472, 2016.

[30] N.E. Friedkin and E.C. Johnsen. Social influence and opinions. *Journal of mathematical Sociology*, 15(3-4):193–205, 1990.

[31] N. E. Friedkin and E. C. Johnsen. Social influence networks and opinion change. *Adv. Group Processes*, 16:1–29, 1999.

[32] Justin Semonsen, Christopher Griffin, Anna Squicciarini, and Sarah Rajtmajer. Opinion dynamics in the presence of increasing agreement pressure. *IEEE Trans. Cybern.*, 49(4):1270–1278, 2019.

[33] Benjamin I. Page and Robert Y. Shapiro. Effects of Public Opinion on Policy. *American Political Science Review*, 77(1):175–190, 1983.

[34] Abhimanyu Das and David Kempe. Algorithms for subset selection in linear regression. In *Proc. of the Fortieth Annual ACM Symposium on Theory of Computing*, STOC '08, pages 45–54. Ass. for Comp. Machinery, 2008.

[35] Abhimanyu Das and David Kempe. Approximate submodularity and its applications: Subset selection, sparse approximation and dictionary selection. *Journal of Machine Learning Research*, 19(3):1–34, 2018.

[36] Andreas Krause, Ajit Singh, and Carlos Guestrin. Near-optimal sensor placements in Gaussian Processes: Theory, efficient algorithms and empirical studies. *J. Mach. Learn. Res.*, 9:235–284, jun 2008.

[37] G. Como and F. Fagnani. From local averaging to emergent global behaviors: The fundamental role of network interconnections. *Systems and Control Letters*, 95:70–76, 2016.

[38] Minyi Huang and Jonathan H. Manton. Opinion dynamics with noisy information. In *52nd IEEE Conference on Decision and Control*, pages 3445–3450, 2013.

[39] Abhijit Banerjee and Olivier Compte. Consensus and disagreement: Information aggregation under (not so) naive learning. *Journal of Political Economy*, 132, 01 2024.

[40] Matthew O. Jackson, Suraj Malladi, and David McAdams. Learning through the grapevine and the impact of the breadth and depth of social networks. *Proceedings of the National Academy of Sciences*, 119(34):e2205549119, 2022.

[41] Yue Han and Lei Yu. A Variance Reduction framework for stable feature selection. In *2010 IEEE International Conference on Data Mining*, pages 206–215, 2010.

[42] Pau Batlle, Joan Bruna, Carlos Fernandez-Granda, and Victor M. Preciado. Adaptive test allocation for outbreak detection and tracking in social contact networks. *SIAM Journal on Control and Optimization*, 60(2):S274–S293, 2022.

[43] Zhihan Zhu, Yanhao Zhang, and Yong Xia. Optimal criteria for best subset selection, 2025.

[44] Ehsan Tohidi, Mario Coutino, and David Gesbert. Revisiting matching pursuit: Beyond approximate submodularity. *Signal Processing*, 226:109638, 2025.

[45] N. A. C. Cressie. *Statistics for Spatial Data*. John Wiley & Sons, 1991.

[46] W.F. Caselton and J.V. Zidek. Optimal monitoring network designs. *Statistics & Probability Letters*, 2(4):223–227, 1984.

[47] Roberta Raineri, Giacomo Como, and Fabio Fagnani. Optimal selection of the most informative nodes in opinion dynamics on networks. *IFAC-PapersOnLine*, 56(2):4192–4197, 2023. 22nd IFAC World Congress.

[48] Roberta Raineri, Giacomo Como, and Fabio Fagnani. Optimal selection of the most informative nodes for a noisy degroot model with stubborn agents. 2025. 23rd European Control Conference.

[49] Roberta Raineri, Giacomo Como, and Fabio Fagnani. Optimal selection of the most informative nodes for a noisy degroot model with stubborn agents, 2025.

[50] Chiara Ravazzi, Francesco Bullo, and Fabrizio Dabbene. Unveiling oligarchy in influence networks from partial information. *IEEE Transactions on Control of Network Systems*, 10(3):1279–1290, 2023.

[51] Chiara Ravazzi, Francesco Malandrino, and Fabrizio Dabbene. Towards proactive moderation of malicious content via bot detection in fringe social networks. *IEEE Control Systems Letters*, 6:2960–2965, 2022.

[52] M. H. Saeed, S. Ali, J. Blackburn, E. D. Cristofaro, S. Zannettou, and G. Stringhini. Trollmagnifier: Detecting state-sponsored troll accounts on reddit. In *IEEE Symposium on Security and Privacy*, pages 2161–2175, San Francisco, CA, USA, 2022.

[53] Torsten Söderström. Errors-in-variables methods in system identification. *Automatica*, 43(6):939–958, 2007.

[54] Chiara Ravazzi, Sina Hojjatinia, Constantino M Lagoa, and Fabrizio Dabbene. Ergodic opinion dynamics over networks: Learning influences from partial observations. *IEEE Trans. Autom. Contr.*, 66:2709–2723, 2021.

[55] Roberta Raineri, Chiara Ravazzi, Giacomo Como, and Fabio Fagnani. Detecting stubborn behaviors in influence networks: A model-based approach for resilient analysis. *IEEE Control Systems Letters*, 8:2343–2348, 2024.

[56] Roberta Raineri, Lorenzo Zino, and Anton Proskurnikov. FJ-MM: The Friedkin-Johnsen opinion dynamics model with memory and higher-order neighbors. *online as ArXiv:2504.06731*, 2025.

[57] Alessandro Licciardi, Roberta Raineri, Anton Proskurnikov, Lamberto Rondoni, and Lorenzo Zino. Sociodynamics-inspired adaptive coalition and client selection in federated learning, 2025.

[58] Abraham Berman and Robert J. Plemmons. *Nonnegative Matrices in the Mathematical Sciences*. Society for Industrial and Applied Mathematics, 1994.

[59] Francis Bloch, Matthew O. Jackson, and Pietro Tebaldi. Centrality measures in networks, 2021.

[60] P. Bonacich. Power and centrality: a family of measures. *American Journal of Sociology*, 5:1170–1182, 1987.

[61] Eduard Talamas and Omer Tamuz. Network cycles and welfare. *SSRN Electronic Journal*, 01 2017.

[62] Coralio Ballester, Antoni Calvó-Armengol, and Yves Zenou. Who's who in networks. wanted: The key player. *Econometrica*, 74(5):1403–1417, 2006.

[63] P. Erdös and A. Rényi. On random graphs i. *Publicationes Mathematicae Debrecen*, 6:290, 1959.

[64] Duncan J. Watts and Steven H. Strogatz. Collective dynamics of 'small-world' networks. *Nature*, 393(6684):440–442, 1998.

[65] Guodong Shi, Claudio Altafini, and John S. Baras. Dynamics over signed networks. *SIAM Review*, 61(2):229–257, 2019.

[66] Rainer Hegselmann and Ulrich Krause. Opinion dynamics under the influence of radical groups, charismatic leaders, and other constant signals: A simple unifying model, 2015.

[67] Daron Acemoglu and Asuman Ozdaglar. Opinion dynamics and learning in social networks. *Dynamic Games and Applications*, 1(1):3–49, 2011.

[68] Benjamin Golub and Matthew O. Jackson. Naïve Learning in Social Networks and the Wisdom of Crowds. *American Economic Journal: Microeconomics*, 2(1):112–149, February 2010.

[69] Anton V. Proskurnikov and Roberto Tempo. A tutorial on modeling and analysis of dynamic social networks. part i. *Annual Reviews in Control*, 43:65–79, 2017.

[70] Fabio Fagnani and Paolo Frasca. *Introduction to Averaging Dynamics over Networks*. Springer, 01 2017.

[71] Abhijit Banerjee and Olivier Compte. Consensus and disagreement: Information aggregation under (not so) naive learning. *SSRN Electronic Journal*, 2022.

[72] B. D. O. Anderson and J. B. Moore. Optimal filtering. *Prentice-Hall*, 2:2387–2399, 1979.

[73] Roger A. Horn and Charles R. Johnson. *Matrix Analysis*. Cambridge University Press, 1990.

[74] Yifei Ma, Roman Garnett, and Jeff Schneider. Σ -optimality for active learning on gaussian random fields. In C.J. Burges, L. Bottou, M. Welling, Z. Ghahramani, and K.Q. Weinberger, editors, *Advances in Neural Information Processing Systems*, volume 26. Curran Associates, Inc., 2013.

[75] George P. Kontoudis and Michael Otte. Closed-form active learning using expected variance reduction of gaussian process surrogates for adaptive sampling. In *2023 American Control Conference (ACC)*, pages 4626–4632, 2023.

[76] S.M. Ross. *Stochastic processes*. Wiley, 1996.

[77] The Mitre Corporation. Common weakness scoring system. September 5, 2014.

[78] Richard Arnold Johnson and Dean W. Wichern. *Applied multivariate statistical analysis*. Prentice Hall, Upper Saddle River, NJ, 5. ed edition, 2002.

[79] Gareth James, Daniela Witten, Trevor Hastie, and Robert Tibshirani. *An Introduction to Statistical Learning: with Applications in R*. Springer, 2013.

[80] George Nemhauser, Laurence Wolsey, and M. Fisher. An analysis of approximations for maximizing submodular set functions—i. *Mathematical Programming*, 14:265–294, 12 1978.

[81] Jonathan Kelner, Frederic Koehler, Raghu Meka, and Ankur Moitra. Learning some popular gaussian graphical models without condition number bounds, 2020.

[82] Shanta Pandey and William Elliott. Suppressor variables in social work research: Ways to identify in multiple regression models. *Journal of the Society for Social Work and Research*, 1(1):28–40, 2010.

[83] Yue Han and Lei Yu. A variance reduction framework for stable feature selection. *Statistical Analysis and Data Mining*, 5:206–215, 12 2010.

[84] Zhipu Jin and Richard M. Murray. Multi-hop relay protocols for fast consensus seeking. In *Proceedings of the 45th IEEE Conference on Decision and Control*, pages 1001–1006, 2006.

[85] Laura Ozella, Daniela Paolotti, Guilherme Lichand, Jorge P. Rodríguez, Simon Haenni, John Phuka, Onicio B. Leal-Neto, and Ciro Cattuto. Using wearable proximity sensors to characterize social contact patterns in a village of rural Malawi. *EPJ Data Sci.*, 10(1), 2021.

[86] David Bindel, Jon Kleinberg, and Sigal Oren. How bad is forming your own opinion? *Games Econ. Behav.*, 92:248–265, 2015.

[87] David Aldous and James Allen Fill. Reversible markov chains and random walks on graphs, 2002.

[88] Jure Leskovec and Andrej Krevl. SNAP Datasets: Stanford large network dataset collection, June 2014.

[89] Jan Lorenz. Universality in movie rating distributions. *The European Physical Journal B*, 71(2):251 – 258, 2009.

[90] A. Mirtabatabaei, P Jia, and F. Bullo. Eulerian opinion dynamics with bounded confidence and exogenous inputs. *Siam Journal on Applied Dynamical Systems*, 13(1):425–446, 2014.

[91] Javad Ghaderi and R. Srikant. Opinion dynamics in social networks with stubborn agents: Equilibrium and convergence rate. *Automatica*, 50(12):3209–3215, 2014.

[92] Rishi Advani and Sean O'Hagan. Efficient algorithms for constructing an interpolative decomposition, 2022.

[93] Edo Liberty, Franco Woolfe, Per-Gunnar Martinsson, Vladimir Rokhlin, and Mark Tygert. Randomized algorithms for the low-rank approximation of matrices. *Proceedings of the National Academy of Sciences of the United States of America*, 104(51):20167–20172, Dec. 2007.

[94] Per-Gunnar Martinsson, Vladimir Rokhlin, Yoel Shkolnisky, and Mark Tygert. Id: A software package for low-rank approximation of matrices via interpolative decompositions, version 0.2. 10 2008.

[95] H. Cheng, Z. Gimbutas, P. G. Martinsson, and V. Rokhlin. On the compression of low rank matrices. *SIAM Journal on Scientific Computing*, 26(4):1389–1404, 2005.

[96] Per-Gunnar Martinsson, Vladimir Rokhlin, and Mark Tygert. A randomized algorithm for the decomposition of matrices. *Applied and Computational Harmonic Analysis*, 30(1):47–68, 2011.

[97] Raghunandan H. Keshavan, Andrea Montanari, and Sewoong Oh. Matrix completion from noisy entries. *CoRR*, abs/0906.2027, 2009.

[98] C. Chen, B. Zhang, Q. Lu, Q. Wu, J. Wang, and S. Liu. A consensus algorithm based on nearest second-order neighbors' information. *IFAC-PapersOnLine*, 48(11):545–550, 2015.

[99] Zuobai Zhang, Wanyue Xu, Zhongzhi Zhang, and Guanrong Chen. Opinion dynamics incorporating higher-order interactions. In *2020 IEEE Int. Conf. Data Min.*, pages 1430–1435, 2020.

[100] Z. Zhang, W. Xu, Z. Zhang, and Guanrong Chen. Opinion dynamics in social networks incorporating higher-order interactions. *Data Min. Knowl. Discov.*, 38:4001–4023, 2024.

[101] Kate Donahue and Jon M. Kleinberg. Optimality and stability in federated learning: A game-theoretic approach. *CoRR*, abs/2106.09580, 2021.

[102] R.A. Horn and C.R. Johnson. *Topics in Matrix Analysis*. Cambridge Univ. Press, New York, 1991.

[103] G.W. Stewart. A note on the perturbation of singular values. *Linear Algebra and its Applications*, 28:213–216, 1979.

[104] Roman Vershynin. Introduction to the non-asymptotic analysis of random matrices, 2011.

[105] G.C. Calafiore and L.E. Ghaoui. *Optimization Models*. Cambridge University Press, 2014.

[106] Abhimanyu Das. *Subset selection algorithms for prediction*. PhD thesis, 2011.

APPENDIX A

Technical Notions

A.1 Matrices and Linear Algebra

A.1.1 Notions on Matrix Eigenvalues

Let us here recall for completeness some well-known notions related to matrix eigenvalues [73].

First, we focus on a characterization of the minimum and maximum eigenvalues of a general symmetric square matrix A as the solutions to minimum and maximum problems involving the Rayleigh quotient $\frac{x'Ax}{x'x}$.

Proposition A.1 (Rayleigh-Ritz characterization)**.** *Given A be a symmetric square matrix taking values in $\mathbb{R}^{n\times n}$ and*

$$\lambda_{min} = \lambda_1 \leq \lambda_2 \leq \cdots \leq \lambda_{n-1} \leq \lambda_n = \lambda_{max}$$

the eigenvalues of A, it holds

$$\lambda_{max} = \max_{x\in\mathbb{R}^n, x\neq 0} \frac{x'Ax}{x'x} \qquad \textit{and} \qquad \lambda_{min} = \min_{x\in\mathbb{R}^n, x\neq 0} \frac{x'Ax}{x'x}.$$

Let us now recall some important eigenvalue inequalities involving either the sum of two matrices ot the extraction of a submatrix (see Section 4.3 in [73] for reference).

Proposition A.2 (Weyl's Theorem)**.** *Let A and B be two symmetric matrices in $\mathbb{R}^{n\times n}$ and assume B positive definite. Then*

$$\lambda_i(A) \leq \lambda_i(A+B), \quad i = \ldots, n$$

Proposition A.3 (Cauchy's interlacing theorem)**.** *Let B a symmetric matrix in $\mathbb{R}^{n\times n}$, $y \in \mathbb{R}^n$, $a \in \mathbb{R}$, and*

$$A = \begin{bmatrix} B & y \\ y' & a \end{bmatrix} \in \mathbb{R}^{(n+1)\times(n+1)}.$$

Then,

$$\lambda_1(A) \leq \lambda_1(B) \leq \lambda_2(A) \leq \cdots \leq \lambda_n(A) \leq \lambda_n(B) \leq \lambda_{n+1}(A).$$

A.1.2 Notions on Matrix Singular Values

Let us here recall for completeness some well-known notions related to matrix singular values [73, 102, 103].

Lemma A.1 (Low rank approximation - 7.4.2 in [102])**.** *For every matrix $A \in \mathbb{R}^{m\times n}$ and positive integer k, we have the following characterization of singular values:*

$$\begin{aligned} &\min_{\substack{X \in \mathbb{R}^{m\times n} \\ rank(X) \leq k}} \quad ||A - X||_2 = \sigma_{k+1}(A) \end{aligned}$$

Lemma A.2. *Let $m \leq n$. The following relations properties are true*

(i) Let A, B in $\mathbb{R}^{m\times n}$, then for any $i = 1, 2, \ldots, m$

$$|\sigma_i(A+B) - \sigma_i(A)| \leq \|B\|_2 = \sigma_1(B)$$

(ii) Let $A \in \mathbb{R}^{m\times n}$ be left invertible and $B \in \mathbb{R}^{n\times p}$, then

$$\|AB\|_2 \geq \sigma_m(A)\|B\|_2$$

(iii) Let $A_{\mathcal{J}}$ denote a submatrix of A formed by selecting a subset of columns indexed by $\mathcal{J}$. For any index set $\mathcal{J}$, the singular values of A satisfy

$$\sigma_i(A) \geq \sigma_i(A_{\mathcal{J}}),\ i \geq 1\,.$$

Proof. (i): See formula (3) in [103]. (ii): The inequality follows from

$$\|B\|_2 = \|A^{-1}AB\|_2 \leq \|A^{-1}\|_2 \|AB\|_2$$

and the fact that $\sigma_{\min}(A) = \|A^{-1}\|_2^{-1}$.

(iii): This follows from interlacing inequalities for eigenvalues applied to singular values, recalling that $\sigma_i(A) = \sqrt{\lambda_i(A'A)}$. □

A.1.3 Matrix Inversion Tools

Let us recall here some key tools related to matrix inversion, that have been used in the dissertation.

First, let us consider the case of a general block matrix M.

Proposition A.4. *Let M be a block matrix*

$$M = \begin{bmatrix} A & B \\ C & D \end{bmatrix}$$

*such that its blocks A and D are square matrices. The matrix $A - BD^{-1}C$ is called the **Schur Complement** of D in M. If D and $A - BD^{-1}C$ are invertible, then M is invertible and*

$$M^{-1} = \begin{bmatrix} (A - BD^{-1}C)^{-1} & -(A - BD^{-1}C)^{-1}BD^{-1} \\ -D^{-1}C(A - BD^{-1}C)^{-1} & D^{-1} + D^{-1}C(A - BD^{-1}C)^{-1}BD^{-1} \end{bmatrix}.$$

Another important class of problems is the one of inverse computation for a rank-k correction of an arbitrary matrix.

Proposition A.5 (Woodbury formula)**.** *Let E and G be square, invertible matrices of size $n \times n$ and $m \times m$, respectively. Let F and H of size $n \times m$ and $m \times n$, respectively.*

Then, the following identity holds:

$$(E+FGH)^{-1} = E^{-1} - E^{-1}F(G^{-1}+HE^{-1}F)^{-1}HE^{-1}\,.$$

A simplified expression can be found in the case of rank-1 perturbation.

Corollary A.1 (Shermann-Morrison formula)**.** *Let $G = 1$, $F = u$ and $H = v$ with $u, v \in \mathbb{R}^n$. Then, from Woodbury formula, it holds:*

$$(E+uv')^{-1} = E^{-1} - \frac{E^{-1}uv'E^{-1}}{1+v'E^{-1}u}\,.$$

A.1.4 Random Matrix Theory

Let us here focus on the extreme singular values of random matrices, whose entries are assumed to be independent and centered random variables. For more details on the following result refers to [104].

Let us preliminary introduce a definition on random variable that will we useful in the subsequent Lemma.

Definition A.1 (Sub-gaussian random variables - Lemma 5.5 in [104])**.** *A random variable X is **sub-gaussian** if it satisfies one of the following equivalent properties, with $K_i > 0$ parameters differing from each other by at most an absolute constant factor,*

- *Tails*
$$\mathbb{P}(|X| > t) \leq exp\left(1 - \frac{t^2}{K_1^2}\right), \text{ for all } t \geq 0$$
- *Moments*
$$(\mathbb{E}[|X|^p])^{\frac{1}{p}} \leq K_2\sqrt{p}, \text{ for all } p \geq 1$$
- *Super exponential moment*
$$\mathbb{E}[exp(X^2/K_3^2)] \leq e$$

Example A.1. *Classical examples of sub-gaussian random variables are Gaussian, Bernoulli and all bounded random variables.*

Theorem A.1 (Theorem 5.31 in [104]). *Let X be an $\mathbb{R}^{m\times n}$ random matrix with $m \leq n$, where the entries are independent, identically distributed with zero mean and unit variance. Suppose that the dimensions m and n grow to infinity while the aspect ratio m/n converges to a constant in $[0,1]$. Then,*

$$\sigma_m(X) = \sqrt{n} - \sqrt{m} + o(\sqrt{n}), \qquad \sigma_1(X) = \sqrt{n} + \sqrt{m} + o(\sqrt{n})$$

almost surely.

Corollary A.2 (Corollary 5.35 in [104]). *Let X be an $\mathbb{R}^{m\times n}$ random matrix with $m \leq n$, where the entries are independent, identically distributed subgaussian random variables with zero mean and variance ξ^2. Then, for every $\eta > 0$, it holds that*

$$\xi(\sqrt{n} - \sqrt{m} - \eta) \leq \sigma_m(X) \leq \sigma_1(X) \leq \xi(\sqrt{n} + \sqrt{m} + \eta)\,.$$

with probability larger than $1 - 2\exp(-\eta^2/2)$.

A.2 Notions of Optimization and Complexity

In this Section, we review fundamental concepts related to optimization problems and their computational complexity, providing the necessary background for the dissertation. We introduce the Greedy Algorithm as an iterative approach for approximating solutions, discussing its performance guarantees under specific assumptions on the objective function.

The results presented in this section are well established in the literature [105, 80]; however, they are reported here for completeness, as well as to ensure clarity and self-containment of the exposition.

A.2.1 Optimization Problem and Computational Complexity

A standard form to define a classical optimization problem is the following [105]:

$$x^* = \max_{x\in\mathcal{X}} f(x)$$

where

- vector $x \in \mathbb{R}^n$ is the *decision variable*;
- $f : \mathbb{R}^n \to \mathbb{R}$ is the *objective function* of interest;
- $\mathcal{X}$ is the *feasible set* of the optimization problem, defined as

$$\mathcal{X} = \{x \in \mathbb{R}^n : g_i(x) \leq 0, i = 1, \ldots, m\}$$

  with $g_i : \mathbb{R}^n \to \mathbb{R}$ problem's *constraints*;
- x^* is the *optimal value*, i.e. the value of the decision variable, among the admissible ones, which maximize the objective function.

To solve an optimization problem, we aim to define an appropriate algorithm and analyze the computational cost required to obtain the solution.

- An optimization problem is in the **class P** if it can be solved through a polynomial-time $O(n^k)$ algorithm. This are also referred to as easy or tractable problems.
- An optimization problem is in the **class NP** if it can be solved through a nondeterministic polynomial algorithm, i.e. it can be solved in polynomial time by a nondeterministic machine. Let us notice that $P \subseteq NP$ since if the problem can be solved by a polynomial algorithm then it can also be solved through a nondeterministic polynomial one.
- An optimization problem is **NP-complete** if it belongs to NP and all other NP-complete problems can be polynomial-time reduced to it. To prove that a problem belongs to this class, one must show that it can be reduced to the SAT (Boolean Satisfiability) problem, which has been proven to be the hardest problem in NP. Alternatively, it suffices to reduce it to another known NP-complete problem, establishing its computational complexity within this class.

A.2.2 Greedy Algorithm

The Greedy Algorithm is a widely used optimization technique for combinatorial optimization problems, such as graph traversal, scheduling, and subset selection. It

iteratively makes locally optimal choices to approximate a global solution, selecting the best immediate option at each step without reconsidering previous decisions. The main advantage of the Greedy Algorithm is its computational efficiency, typically running in polynomial time, making it suitable for large-scale problems. Moreover, as we will explore later, for specific problem classes, it offers strong approximation guarantees, often achieving near-optimal solutions at a significantly lower computational cost than exact methods.

Let us now delve deeper into this class of algorithms, exploring their fundamental principles and performance guarantees. Let's assume that, given a real valued function $f(\mathcal{K})$ defined on the set of subsets of $\mathcal{Z}$, the problem of interest is to find a subset $\mathcal{K}$ of cardinality less than or equal to a specific integer s such that $f(\mathcal{K})$ is maximum, i.e.

$$\max_{\mathcal{K}\subseteq\mathcal{Z}} \{f(\mathcal{K}) : |\mathcal{K}| \leq s\}. \tag{A.1}$$

The introduced problem is an example of a combinatorial optimization problem since in order to find the optimal subset $\mathcal{K}$, solution of (A.1), we have to compute $f(\mathcal{K})$ for all the possible subsets of cardinality s from $\mathcal{Z}$, i.e. we need $\binom{n}{s}$ evaluations with $n = |\mathcal{Z}|$. Given that this computations may become prohibitive, it is useful to look for a proper heuristic.

Let us define a Greedy Algorithm adapted to our problem:

1. Denote with $\mathcal{K}_t$ the subset of nodes selected at t-th algorithm iteration and initialize $\mathcal{K}_0 = \emptyset$.

2. For each iteration $t = 1, \ldots, s$, with s cardinality of the set to select, set $\mathcal{K}_t = \mathcal{K}_{t-1} \cup \{i_t\}$ with
$$i_t \in \arg\max_{i\in\mathcal{R}\setminus\mathcal{K}_{t-1}} f(\mathcal{K}_{t-1} \cup \{i_t\}).$$

3. At the end of s-th iteration, the algorithm will propose $\mathcal{K}^G = \mathcal{K}_s$ as an approximated solution to problem (A.1).

Since at each step is required only one evaluation of function $f(\mathcal{K})$, the computational complexity is here reduced to s times the cost of computing $f(\mathcal{K})$. It is important to note that the solution obtained is sub-optimal; consequently, the performance guarantees of this heuristic must be evaluated on a case-by-case basis.

APPENDIX B

Technical Proofs of Approximate Submodularity

B.1 Residuals Properties

The results presented in this section are based on Das's PhD thesis [106] and are reformulated here using our specific notation for the sake of completeness.

Lemma B.1. *Given the random variable Y, X_i and the set of random variables $X_{\mathcal{A}}$, s.t $C_{\mathcal{A}\cup\{i\},\mathcal{A}\cup\{i\}}$ is the invertible covariance matrix among X_i and $X_{\mathcal{A}}$, it holds*

$$Res(Y, X_{\mathcal{A}\cup\{i\}}) = Res(Res(Y, X_{\mathcal{A}}), Res(X_i, X_{\mathcal{A}}))$$

Proof. Let us define

$$\hat{Y} := \mathrm{Res}(Y, X_{\mathcal{A}}),\ \hat{X}_i := \mathrm{Res}(X_i, X_{\mathcal{A}}) \text{ and } T := \mathrm{Res}(Y, X_{\mathcal{A}\cup\{i\}})\,.$$

Based on residuals definition (3.22), the following equalities hold:

$$\begin{aligned} \hat{Y} &= Y - \textstyle\sum_{j\in\mathcal{A}} \alpha_j X_j \\ \hat{X}_i &= X_i - \textstyle\sum_{j\in\mathcal{A}} \beta_j X_j \\ \hat{Y} &= \mathrm{Res}(\hat{Y}, \hat{X}_i) + \gamma \hat{X}_i\,, \end{aligned}$$

where $\{\alpha_j\}_{j\in\mathcal{A}}, \{\beta_j\}_{j\in\mathcal{A}}$ and γ are properly chosen coefficients. Then, combinig these relations we obtain:

$$\begin{aligned} Y &= \hat{Y} + \textstyle\sum_{j\in\mathcal{A}} \alpha_j X_j = \\ &= \gamma\hat{X}_i + \mathrm{Res}(\hat{Y},\hat{X}_i) + \textstyle\sum_{j\in\mathcal{A}} \alpha_j X_j = \\ &= \gamma\hat{X}_i + \mathrm{Res}(\hat{Y},\hat{X}_i) + \textstyle\sum_{j\in\mathcal{A}} (\alpha_j - \gamma\beta_j) X_j \,. \end{aligned} \tag{B.1}$$

Analogously, we can rewrite T as

$$T = Y - \eta X_i + \sum_{j\in\mathcal{A}} \xi_j X_j \tag{B.2}$$

for properly chosen coefficients η and $\{\xi_j\}_{j\in\mathcal{A}}$.

Combining (B.1) and (B.2) , it holds

$$T + \eta X_i + \sum_{j\in\mathcal{A}} \xi_j X_j = \gamma X_i + \mathrm{Res}(\hat{Y},\hat{X}_i) + \sum_{j\in\mathcal{A}} (\alpha_j - \gamma\beta_j) X_j$$

which is equivalent, simply rearranging the equations term, to:

$$(\eta - \gamma) X_i + \sum_{j\in\mathcal{A}} (\xi_j + \gamma\beta_j - \alpha_j) X_j = \mathrm{Res}(\hat{Y},\hat{X}_i) - T \,.$$

If we compute now the covariance with respect to X_i and $X_\mathcal{A}$ for both the left and the right side of the equation we obtain:

$$C_{\mathcal{A}\cup\{i\},\mathcal{A}\cup\{i\}} \begin{bmatrix} \xi_\mathcal{A} + \gamma\beta_\mathcal{A} - \alpha_\mathcal{A} \\ \eta - \gamma \end{bmatrix} = 0$$

Finally, since the covariance matrix $C_{\mathcal{A}\cup\{i\},\mathcal{A}\cup\{i\}}$ is assumed to be invertible then it is a full rank matrix and thus it holds

$$\eta - \gamma = 0 \,,\; \xi_j + \gamma\beta_j - \alpha_j = 0, j \in \mathcal{A} \,.$$

Observing so that $\mathrm{Res}(\hat{Y},\hat{X}_i) - T = 0$, the thesis follows.

□

Lemma B.2. *For any random variable Y, X_i and a set of random variables $X_\mathcal{A}$, it holds*

$$F(\mathcal{A}\cup\{i\}) - F(\mathcal{A}) = F(Res(X_i, X_\mathcal{A}))$$

Proof. Let us define

$$\hat{Y} := \mathrm{Res}(Y, X_{\mathcal{A}}),\ \hat{X}_i := \mathrm{Res}(C_i, X_{\mathcal{A}}.)$$

Based on Lemma B.1 we know that

$$\mathrm{Res}(Y, A_{\mathcal{A}\cup\{i\}}) = \mathrm{Res}(\hat{Y}, \hat{X}_i)\,. \tag{B.3}$$

Recalling that $\mathrm{Res}(Y, X_{\mathcal{A}}) = Y - \mathbb{E}[Y|X_{\mathcal{A}}]$, we can rewrite (B.3) as follows:

$$Y - \mathbb{E}[Y|X_{\mathcal{A}\cup\{i\}}] = \hat{Y} - \mathbb{E}[\hat{Y}|\hat{X}_i]\,.$$

Properly substituting $\hat{Y}$ based on its definition, then

$$Y - \mathbb{E}[Y|X_{\mathcal{A}\cup\{i\}}] = Y - \mathbb{E}[Y|X_{\mathcal{A}}] - \mathbb{E}[\hat{Y}|\hat{X}_i]\,.$$

Avoiding Y and computing the variance for both sides of the equation we obtain

$$\mathrm{Var}(\mathbb{E}[Y|X_{\mathcal{A}\cup\{i\}}]) = \mathrm{Var}(\mathbb{E}[Y|X_{\mathcal{A}}]) + \mathrm{Var}(\mathbb{E}[\hat{Y}|\hat{X}_i])$$

and observing that, conditioned to $\hat{X}_i$, it holds $\mathrm{Var}(\mathbb{E}[\hat{Y}|\hat{X}_i]) = \mathrm{Var}(\mathbb{E}[Y|\hat{X}_i])$, then

$$\mathrm{Var}(\mathbb{E}[Y|X_{\mathcal{A}\cup\{i\}}]) = \mathrm{Var}(\mathbb{E}[Y|X_{\mathcal{A}}]) + \mathrm{Var}(\mathbb{E}[Y|\hat{X}_i])\,.$$

Finally, since by definition, given an arbitrary subset $\mathcal{K}$:

$$F(\mathcal{K}) = \mathrm{Var}(Y) - \mathbb{E}[\mathrm{Var}(Y|X_{\mathcal{K}})] = \mathrm{Var}(\mathbb{E}[Y|X_{\mathcal{K}}])$$

then the thesis follows.

□