POLITECNICO DI TORINO
Repository ISTITUZIONALE

Learning Hidden Influences in Large-Scale Dynamical Social Networks: A Data-Driven Sparsity-Based Approach, in Memory of Roberto Tempo

Original
Learning Hidden Influences in Large-Scale Dynamical Social Networks: A Data-Driven Sparsity-Based Approach, in Memory of Roberto Tempo / Ravazzi, Chiara; Dabbene, Fabrizio; Lagoa, Constantino; Proskurnikov, Anton V.. - In: IEEE CONTROL SYSTEMS. - ISSN 1066-033X. - ELETTRONICO. - 41:5(2021), pp. 61-103. [10.1109/MCS.2021.3092810]

Availability:
This version is available at: 11583/2924340 since: 2021-09-16T15:32:20Z

Publisher:
IEEE

Published
DOI:10.1109/MCS.2021.3092810

Terms of use:
openAccess

This article is made available under terms and conditions as specified in the corresponding bibliographic description in the repository

Publisher copyright
IEEE postprint/Author's Accepted Manuscript

©2021 IEEE. Personal use of this material is permitted. Permission from IEEE must be obtained for all other uses, in any current or future media, including reprinting/republishing this material for advertising or promotional purposes, creating new collecting works, for resale or lists, or reuse of any copyrighted component of this work in other works.

(Article begins on next page)

18 October 2022

Learning hidden influences in large-scale dynamical social networks

A data-driven sparsity-based approach

"In memory of Roberto Tempo"

Chiara Ravazzi, *Member, IEEE*, Fabrizio Dabbene, *Senior Member, IEEE*, Constantino Lagoa, *Member, IEEE*, Anton V. Proskurnikov, *Senior Member, IEEE*

Towards a unified system theory of opinion formation and social influence

Processes of information diffusion over social networks (for example, opinions spread and beliefs formation) are attracting substantial interest to various disciplines ranging from behavioral sciences to mathematics and engineering. Since the opinions and behaviors of each individual are influenced by interactions with others, understanding the structure of interpersonal influences is a key ingredient to predict, analyze and, possibly, control information and decisions [1]. With the rapid increase of social media platforms that provide instant messaging, blogging, and other social networking services – see **"Online Social Networks"** (OSN) – people can easily share news, opinions, and preferences. Information can reach a broad audience much faster than before, and opinion mining and sentiment analysis is now becoming a key challenge in modern society [2]. The first anecdotal evidence of this fact is probably the use that the Obama campaign made of social networks during the 2008 US presidential elections [3]. More recently, several news outlets stated that Facebook users played a major role in feeding into fake news that might have influenced the outcome of the 2016 US presidential election [4]. This can be explained by the phenomena of homophily and biased assimilation [5]–[7] in social networks,

C. Ravazzi and F. Dabbene are with the National Research Council of Italy, CNR-IEIIT, c/o Politecnico di Torino, Corso Duca degli Abruzzi 14, 10129, Turin, Italy.

C. Lagoa is with The Pennsylvania State University, University Park, PA 16802.

A.V. Proskurnikov is with Politecnico di Torino, Corso Duca degli Abruzzi 14, 10129, Turin, Italy, and also with the Institute for Problems in Mechanical Engineering of the Russian Academy of Sciences, St. Petersburg, Bolshoy pr. V.O. 61, Russia.

which corresponds to the tendency of people to follow the behaviors of their friends and establish friendship with like-minded individuals. The inference of social ties from empirical data becomes of central interest in political organizations and business firms due to its potential impact in decision making and action planning. According to the report published by McKinsey & Company [8], *"Marketing-induced consumer-to-consumer word of mouth generates more than twice the sales of paid advertising."* The influence analysis is becoming a key input into sophisticated recommendation engines that identify potential customers, exploiting similarities between several users to predict preferences. The same report [8] estimates that 35% of Amazons revenue and 75% of what users watch on Netflix came from product recommendations. The study of structures in networks (such as, for example, community detection and computing the node's centralities) has been the main concern of *social network analysis* (SNA) [9], embraced now by the multidisciplinary field of network science [10]–[12]. On a parallel line of research, many works have been published in physical, mathematical, and engineering that focus on dynamical models of opinion diffusion (see [13]–[19] and references therein).

Figure 1: Data-driven analysis of social systems.

There are numerous gaps between SNA and opinion dynamics modeling, and the relations between structures of social influence and information spread mechanisms are far from being well studied. This article takes a step towards filling these gaps (and in the direction of deriving a unified theory) by describing the intricate relations between structural and dynamical properties of social systems. In this new area, the methods of systems and control should play a key role. The aim, as explained in "Summary," is to provide a general overview of the main concepts, algorithmic tools, results, and open problems in the systematic study of learning interpersonal influence in networked systems.

Online Social Networks

Online social networks (OSN) or techno-social networks refer to a group of individuals or organizations that use new communication technologies (that is, the Internet and mobile devices) as a communication medium, forming a social structure described by particular relations [20]. The study of OSN has increasingly attracted the attention of the scientific community. In fact, online services, such as Facebook, Twitter, or Instagram play an increasingly important role in the dissemination of opinions and in the emergence of certain behaviors. They facilitate social interactions, helping individuals to find other people with common interests, establish a forum for discussion, and exchange information [21], [22]. The Special Digital Report 2020 [23] states that digital, mobile, and social media are a fundamental part of people's daily lives around the world. According to Statista.com, the social penetration rate of OSN in 2019 reached 70% in East Asia and North America, followed by North Europe at 67%, leading to a global social penetration rate of 45%. Moreover, since the COVID-19 outbreak was declared a public health emergency of international concern on January 30, 2020, social media usage has reportedly increased significantly. On March 24, 2020, Facebook recorded a 50% increase in total messages in many of the countries most affected by the virus, with a 70% increase in the time users spent on social media since the beginning of the pandemic [24]. Figure 2 shows the total number of active users of the most popular social media networks [25].

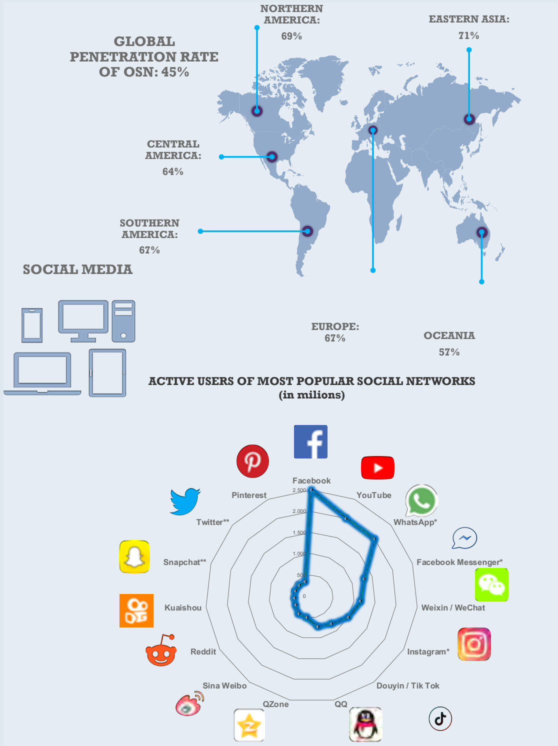


Figure 2: Global penetration of online social networks.

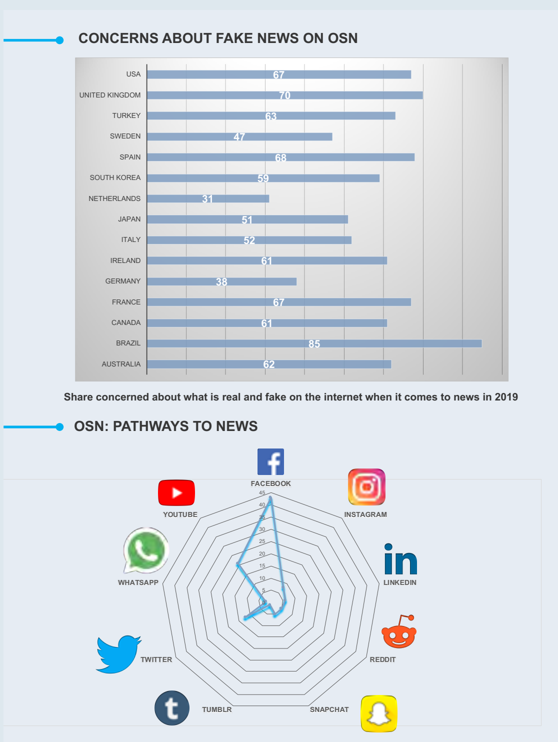


Figure 3: Online social networks and fake news.

A downside of this high social penetration rate is the spread of *fake news* whose detection is becoming a serious problem [26], [27]. Fig. 3 shows the share concerned about what is real and fake on the Internet regarding news. The data are updated until 2019 [28]. Moreover, approximately 43% adults in USA get news from Facebook (according to a survey conducted in July and August of 2018 [29]). This proportion is much higher than the percentage of adults who get news through YouTube (21%), Twitter (only 12%), and other platforms. Recent contributions in systems and control focus on modeling the dynamics of the spread of misinformation [30]. It has been shown [31] that in Twitter diffusion networks, misleading content spreads deeper than mainstream news with a small number of followers (and communities sharing fake news are more connected and clustered). Structural properties of Twitter diffusion networks (such as the number and size of weakly connected components, average clustering coefficient, diameter of the largest weakly connected components) can effectively be used to identify misleading and harmful information [31]. Inferring the networks structure and learning global properties from partial information becomes a central question [32], [33].

Summary

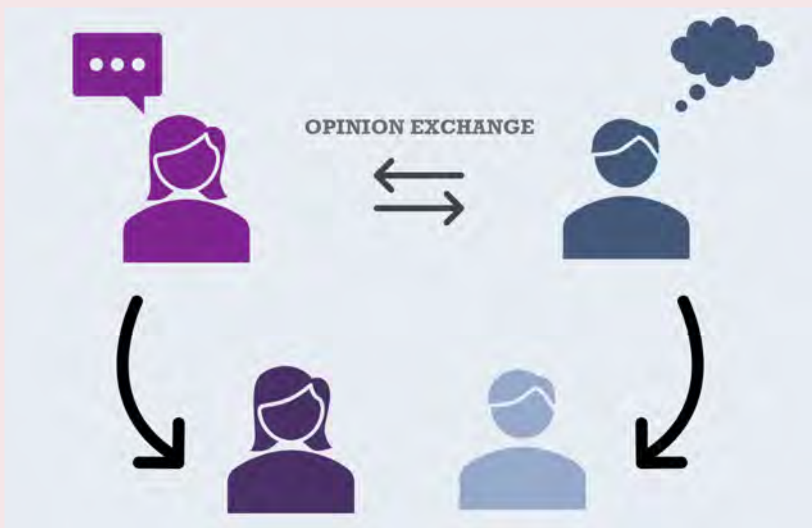
Interpersonal influence estimation from empirical data is a central challenge in the study of social structures and dynamics. *Opinion dynamics* theory is a young interdisciplinary science that studies opinion formation in social networks and has huge potential in applications such as marketing, advertisement, and recommendations. The term *social influence* refers to the behavioral change of individuals due to the interactions with others in a social system (for example, organization, community, or society in general). The advent of the Internet has made a huge volume of data easily available that can be used to measure social influence over large populations. The aim of this work is qualitatively and quantitatively inferring social influence from data using a *systems and control viewpoint*. First, some definitions and models of opinions dynamics are introduced, then some structural constraints of online social networks are considered based on the notion of sparsity. Then, the main approaches to infer the network's structure from a set of observed data are reviewed. Finally, algorithms that exploit the introduced models and structural constraints are presented, focusing on the sample complexity and computational requirements.

As summarized in **"Opinion dynamics in a nutshell,"** the main research lines in this field can be grouped into three broad categories: modeling, analysis, and control. *Modeling* aims to find a coherent mathematical description of the social interactions. To build a mathematical model, one must define (a) the interaction protocol (for example, the times of interactions that can be discrete or continuous), the contact modes (deterministic or random), and the frequency of interactions among social network members; and (b) the dynamical mechanism of social interactions (or ties), which can be described by linear or nonlinear functions [34]–[36]. In the simplest situation, each social tie is described by a single scalar, treated as the "influence weight" one individual assigns to the other [14]. *Analysis* of social networks usually focuses on the study of the qualitative and quantitative properties of the opinion dynamics (such as asymptotic convergence or oscillations, eventual consensus, or disagreement). It is also important to extract low-dimensional features of the network, for example, to identify communities or the most influential leaders. A long-standing goal in the study of social networks is to *control* the final distribution of opinions [37]–[40].

Opinion dynamics over networks in a nutshell

Assumptions:

- Population of individuals (or actors)
- Individuals interact and exchange their opinions
- As a result, their opinions evolve



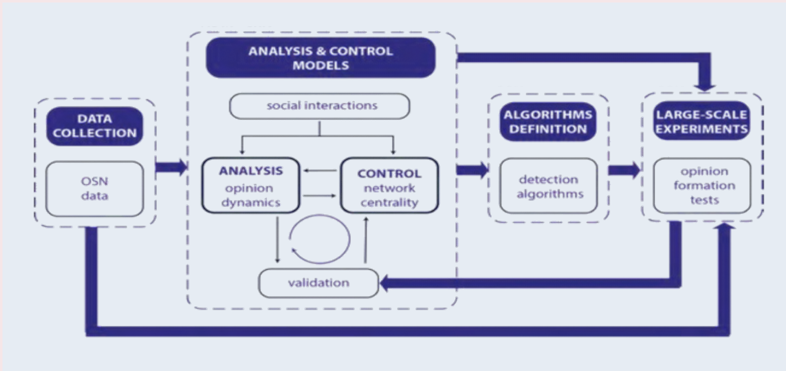
Three research directions:

1) **Mathematical modeling** of social behaviors
   - *interaction protocol*: discrete/continuous-time, deterministic/random contacts
   - *social ties*: linear (parameterized by scalar "weights") or nonlinear functions
2) **Analysis**: sociology + computer science
   - the *evolution of opinions (convergence/oscillations, consensus/disagreement)*
   - low-dimensional features (communities, opinion leaders)
3) **Control**: design mechanisms to provide the desired behavior of the opinion profile
   - induce qualitative changes (for example, consensus)
   - induce quantitative changes (for example, drive the individual's opinions to a desired value)

The role of the systems and control community in the area of social networks has increased with the introduction of *dynamical models*, which capture phenomena observed in sociology and thus enable understanding of social processes and interactions. The workflow envisaged for a systematic study of opinion formation and the networks' structural properties is shown in the diagram in **"Data-driven systems and control approach."** All research [including data collection, design of a mathematical framework (modeling, analysis, and control), algorithms development, and design of large-scale experiments] must proceed in parallel with continuous interactions among the blocks. Data collection and processing constitute the backbone of computational social science [41] and require a careful systematization. More precisely, one must define what kind of information can be acquired, the frequency of the samples, and subsequently, how much information is contained in each sample. In this context, the first issue is to encode people's opinions, sentiments, and preferences from written language into a formal language or numerical representation that can be processed with numerical techniques. To this regard, several methods for sensing opinions based on sentiment analysis have been proposed in recent years [42]. Efficient sampling procedures of graph signals require only a few nodes in a network to be directly observed and sensed [43], [44] and remove irrelevant information to improve the performance of processing. The analysis must identify the best models that accurately characterize the social system and select the evaluation metrics to quantify the interpersonal influence in the network. Efficient algorithms and new control mechanisms of centrality measures (see "Centrality measures") must be designed to improve social network interconnectivity and resilience. Note that the control models should be *data-driven* (that is, simulations based on data collected from real social networks must be performed to validate and refine the dynamic models, then predict and control the opinion diffusion over the network). Since parameters of system dynamics models are subject to uncertainty, a sensitivity analysis is crucial to explore the effects of parameter uncertainty on the behavior patterns.

Data-driven systems and control approach

The ultimate goal is the development of a theoretical framework that is based on systems theory and data-driven control, able to predict the processes of opinion formation and information spread in large-scale social networks, and provides well-grounded tools to quantify and control the impact of specific actions. The workflow for a systematic study of network structures and dynamics is shown in the diagram below.



Friedkin and Johnsen experiment [45], [46]

A seminal example of coupling the theory with empirical research can be found in [46], where The Friedkin and Johnsen model [45] for single-issue opinion dynamics is validated for small and medium-sized groups of individuals (social actors).

Experiment design:

- 30 groups
- 4 individuals
- 15 issues (Risk Choice Dilemmas)

The influence is measured experimentally by asking the participants to distribute "chips" to measure the influence weights each participant has assigned to him/herself and to the others during the decision making process.

"The issues involve opinions on the minimum level of confidence (which is a value in the [0, 1] interval) required to accept a risky option with a high payoff over a less risky option with a low payoff."

Example:

"Investment Choice: Imagine you want to invest some money you recently inherited. You may invest in secure, low-return securities (small risk), or alternately, in more risky securities that offer the possibility of large gains (great risk)."

- Initial opinions of actors recorded
- Fifteen minutes of discussion
- Actors distribute "chips" to quantify the influence
- Final actors' opinions recorded

2

30 groups consisting of four people participate in the experiment. The participants are asked to express their opinions on 15 different issues selected uniformly at random without replacement from a set of risk choice dilemmas. Risk choice dilemmas are hypothetical life decision situations that are used to measure the willingness to assume a risk. More precisely, the agents are asked to express their minimum level of confidence, (that is, a scalar value in the range $[0, 1]$) required to accept a risky option with a high payoff over a less risky option with a low payoff. Individuals in the group record their initial opinions on the issue, then a 15-min discussion is opened and the final opinions are recorded.

Influence determination: To estimate social influence, people are asked in the post discussion to distribute "chips" between the actors they interacted with as a subjective measure of influence exercised by other group members. The model validation is then performed by showing that the opinions predicted through the model are close to those recorded.

Social ties between the individuals have to be quantified to validate and examine the models. In a small-sized group participating in a roundtable discussion, individuals can estimate the influence of themselves and the others on the formation of their opinions (**"Friedkin-Johnsen experiment"**). However, this approach is inapplicable to large-scale groups and online social networks whose structures of influence relations can only be inferred from data. The rapid development of the Internet, on one hand, makes a large volume of data easily available for analysis. On the other hand, it poses new challenges. Data size is getting larger, and information collected becomes heterogeneous and more complex. The massive data in OSN consist of linked data, mainly in the form of graphic structures, describing the communications between any two entities, text, images, audio, and video that must be processed. Hence, efficient analytic tools and algorithms to reconstruct social influence mechanisms are required. These considerations motivate the present work, which aims to present a unified overview onto the two main aspects of interpersonal influence estimation: i) the social network sensing problem and ii) network reconstruction algorithms, with a particular focus on sample complexity and computational requirements. The main challenge is to guarantee efficiency and scalability of the algorithms in the face of big data produced by OSN. It is shown that the interpersonal influence estimation problem can leverage a mature technical background and strong mathematical foundations, and it can be addressed efficiently using modern techniques. The main studies performed on this subject are highly innovative, blending learning tools with high-dimensional data analysis, including principal component analysis [47], compressed sensing [48] and graph analytics [49], and encompassing various fields of research, for example,

1) graph theory and linear algebra;
2) control theory techniques: stability, controllability, system identification, and optimal and robust control;

3) signal processing, statistics, and machine learning for big data analysis;
4) efficient optimization-based algorithms for sampling and reconstruction of graph signals.

The main focus of the article is based on previous works [50]–[57]. An interested reader is referred to the additional literature for more insight. This survey has been partly presented in the tutorial section "Control and Learning for Social Sciences: Dynamical Networks of Social Influence" of IFAC World Congress 2020. The remainder of this article is organized as follows. The main body of the text are the"Summary boxes" (in light pink) and "Focus boxes" (in light blue). The "Summary boxes" contain information and discussions introduced in the main text and aim to improve understanding with concise and schematic blocks. The "Focus boxes" address specific topics and are intended as further in-depth study. They contain technical results and additional discussions. The reader can skip these parts without losing understanding.

Defining influence in social networks

As defined in [58], [59], the interpersonal influence is a *"causal effect of one actor on another,"* such as a change in opinions and behaviors of the influenced actor [60]. Quantification and measurement of social ties are long-standing problems that have been studied since the 1950s [61]–[64]. One principal difficulty is to separate direct and indirect influence: *"if the opinion change has occurred within a system of influences involving other actors, then these other actors may have induced the observed opinion difference or change"* [58]. Another problem is the co-evolution of social ties and the individuals' behaviors. On one hand, people modify their behaviors to align them with the behaviors of their friends (social influence). However, people tend to form friendship with others like themselves (social selection). Opinions and other mutable characteristics of people are thus formed by the *interplay* between social selection and influence [12]. Research on interpersonal influence exist in the literature, among which three main directions prevail. The first direction of research develops the seminal ideas of Granovetter [65], defining the strength of a social tie between two individuals as a function of their *positions* in the social group: For instance, the more common friends actors A and B have, the stronger is the tie among them [60], [65]. Social influence introduced in this way depends only on the structure of a social network. A large amount of available data from real-world social networks and the existence of efficient tools for their analysis make this approach very attractive for both behavioral and computer sciences. The second line of research relates the social influence to temporal (dynamical) mechanisms, modifying some numerical attributes of social actors such as (for example) opinions or quantities related to them. The influence (or power) of actor B over actor A is a parameter of the corresponding mechanism, measuring A's sensitivity to the opinion of B or the level of trust in B's opinions. This idea has been elaborated in the Friedkin-Johnsen theory of social influence [14], [45], [46], [58]. The fundamental results reported in [58] establish interrelations between the structural and dynamical approaches to social influence. Namely, in networks of scientific collaborations, social positions (“opinions”) of individual researchers can be encoded by multidimensional vectors. Two opinions are close if the researchers have similar (in some sense) sets of collaborators. The evolution of these opinions is predicted by the Friedkin-Johnsen model of opinion formation (see Section IV) whose parameters can be constructed via a structural analysis. The third direction of research on influence in complex networks (not necessarily social) is *statistical* (learning-based) methods of network reconstruction. Like the second approach, it assumes that the actors at a network’s nodes are endowed with numerical values that are supposed to be random. Unlike the second approach, the existence of a temporal mechanism modifying the values is not stipulated. A tie between two nodes corresponds to *statistical correlation* between their values, and the strength of this tie is naturally measured by the correlation coefficient. In other words, a network is considered a probabilistic *graphical model* [66], [67] and is analyzed by methods of statistics and statistical learning theory.

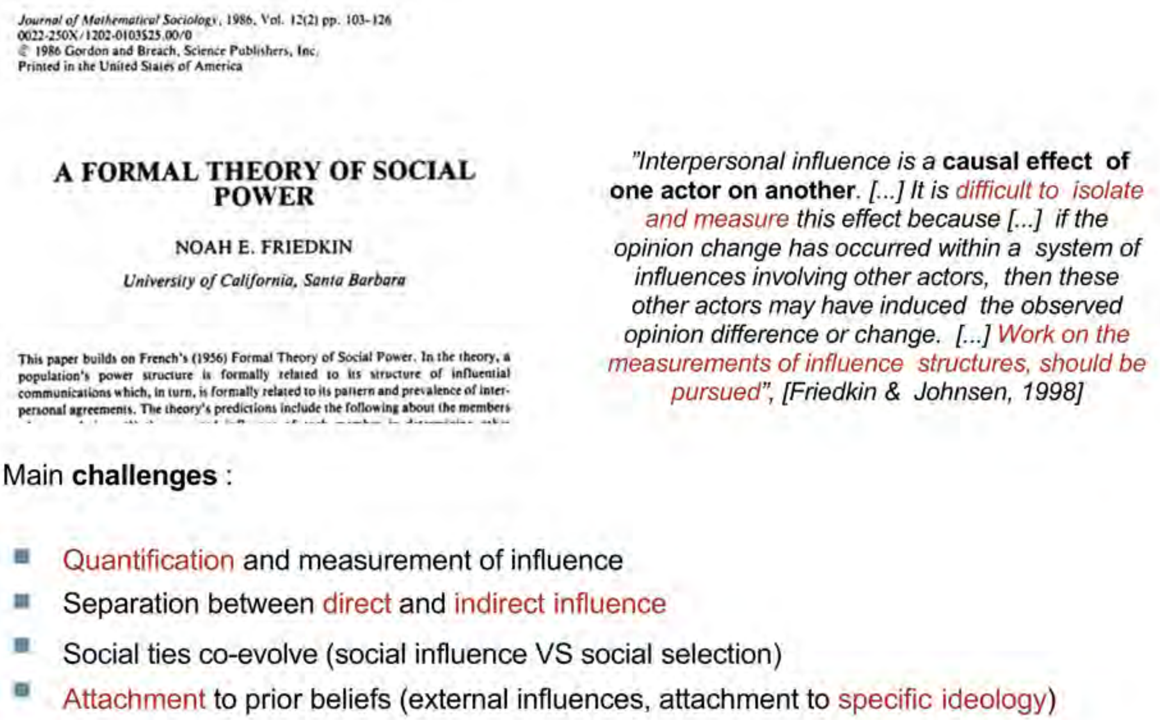


Figure 4: Measuring influence, affinities, and social ties.

This article focuses on the second and third lines of research. Section **??** considers statistical estimation of social influence (the third direction). Sections through address identification of dynamic mechanisms of opinion formation, namely, the Friedkin-Johnsen model. Both approaches use the reconstruction of a *weighted directed graph* whose nodes have numerical attributes (considered as opinions of social actors), where the arcs represent social ties whose strengths are described by weights. A natural question is how the estimates of these weights can be used to study the structure of a social network (for example, exploring communities)? The remainder of this section is devoted to this problem and introduces important characteristics of a weighted graph.

Influence-related measures

A social network consists of two main components: i) social actors (individuals or organizations), and ii) the dependency, influence ,or similarity relations. Each actor has a numerical attribute (representing an opinion). The social network can be mathematically described by a directed weighted graph (see box **"A Glossary on Graphs"** for graph-related definitions). At the local level, the social influence is a *directional effect from node i to node j, which is related to the edge strength* $(i, j) \in \mathcal{E}$ [60]. To comply with previously published works on opinion dynamics [14], [56] the arc (i, j) is associated with the influence of j on i. Social influence relations can thus be encoded in the *social influence matrix* $\mathbf{W} = [w_{ij}]$, which is adapted to the graph (if $(i, j) \notin \mathcal{E}$, then the corresponding entry w_{ij} is zero). In the dynamic models of social influence [14], this matrix is often normalized to be row stochastic. At the global level, some nodes can be more influential than others due to the network interconnections. Several global measures have been introduced to identify the most relevant entities in the network. These global measures can refer to nodes or edges and can be defined in several ways according to the specific context and application, leading to different notions of *centrality* (a node's/edge's importance) measure. Various measures on centrality are defined in **"Centrality measures in weighted graphs."**

A Glossary on Graphs

An **unweighted graph** $\mathcal{G}$ is represented by the couple $(\mathcal{V}, \mathcal{E})$, where

- $\mathcal{V}$ is the set of **nodes** (corresponding for example, to agents in the network), indexed as $1, \ldots, n$.
- $\mathcal{E} \subseteq \mathcal{V} \times \mathcal{V}$ is a set of ordered pairs of nodes describing the relationships if $(i, j) \in \mathcal{E}$ then j is influenced by i. The couples (i, j) are referred to as the **edges** of the graph.

Given an unweighted graph, **adjacency matrix** $\mathbf{Adj}$is defined, with ij entry $[\mathbf{Adj}]_{ij} = 1$ if $(i, j) \in \mathcal{E}$ and zero otherwise.

A **weighted graph** $\mathcal{G}$ is represented by a triple $(\mathcal{V}, \mathcal{E}, \mathbf{W})$, where $\mathcal{V}$ and $\mathcal{E}$ are the nodes and edges of the graph, and $\mathbf{W} = [w_{ij}]$ is the weighted adjacency matrix (known as the **influence matrix**) whose entry w_{ij} defines the **weight** of the edge (i, j) [$w_{ij} = 0$ if $(i, j) \notin \mathcal{E}$, that is, i and j are not connected]. Each square matrix $\mathbf{W} = (w_{ij})_{i,j \in \mathcal{V}}$ can be associated with a graph $\mathcal{G}[\mathbf{W}] = (\mathcal{V}, \mathcal{E}, \mathbf{W})$, where $\mathcal{E} = \{(i, j) : w_{ij} \neq 0\}$. Figure 5 shows the pictorial representation of a graph.

A matrix $\mathbf{M}$ is said to be **adapted to the graph** $\mathcal{G}$ if $\mathcal{G}[\mathbf{M}] = \mathcal{G}$. By construction, the adjacency matrix $\mathbf{Adj}$ and every influence matrix $\mathbf{W}$ of a graph $\mathcal{G}$ are adapted to $\mathcal{G}$.

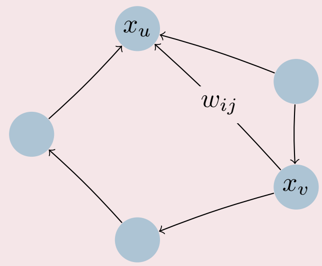


Figure 5: Graph: pictorial representation.

The matrix $\mathbf{W}$ is said to be **row stochastic** if its rows sum to one, that is, $\sum_i w_{ij} = 1$. In a compact form, $\mathbf{W}\mathbb{1} = \mathbb{1}$, with $\mathbb{1} \doteq [1 \cdots 1]^\top$. Similarly, matrix $\mathbf{W}$ is **column stochastic** if $\sum_j w_{ij} = 1$, that is, $\mathbb{1}^\top \mathbf{W} = \mathbb{1}^\top$. For unweighted graphs, $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ is an **undirected graph** if $(i, j) \in \mathcal{E}$ implies that (j, i) is also an edge in $\mathcal{E}$. For a weighted graph $\mathcal{G} = (\mathcal{V}, \mathcal{E}, \mathbf{W})$, it is also required that the weights of edges (i, j) and (j, i) coincide: $\mathbf{W} = \mathbf{W}^\top$.
The **Laplacian** matrix of a weighted graph (possibly, directed) is defined as

$$\mathbf{L} \doteq \mathbf{D} - \mathbf{W},$$

where $\mathbf{D} \doteq \mathrm{diag}(d_1, \ldots, d_n)$ is the weighted degree matrix, and $d_i = \sum_j w_{ij}$. For each node $i \in \mathcal{V}$, its **neighborhood** is denoted by $\mathcal{N}_i \doteq \{j \in \mathcal{V} : (i, j) \in \mathcal{E}\}$.
A sequence of edges $(i, i_1), (i_1, i_2), \ldots, (i_{m-1}, j)$ without repeated vertices forms a **path** from i to j.
A graph is said to be **strongly connected** if there exists a path between any pair of nodes.

The simplest and most popular definition of centrality is the *degree centrality*, that is, the number of neighbors of a node. This measure can be interpreted as a measure of the immediate risk of a node catching (in-degree) or spreading (out-degree) information. A more general concept is *K-path* centrality [68], defined as the number of paths of length K starting from a node. Both degree and K-path centrality definitions are local. Other alternative centrality measures have been considered to measure the importance of a node for the graph as a whole. Among them, the closeness, betweenness, and eigenvector centrality are briefly discussed. The *closeness centrality* is a measure of how much a node is close to most of the other nodes [69] and provides insight on how long it will take to spread information from i to all other nodes in the network.

Centrality measures in weighted graphs

Consider a weighted graph $\mathcal{G} = (\mathcal{V}, \mathcal{E}, \mathbf{W})$, where $\mathcal{V}$ is the set of agents in the network, $\mathcal{E} \subseteq \mathcal{V} \times \mathcal{V}$ is a set of links describing the interpersonal influences, and $\mathbf{W} \in [0,1]^{\mathcal{V}\times\mathcal{V}}$ is the *social influence matrix* (which is adapted to the graph).

A **centrality measure** is a nonnegative scalar measuring the importance of a node or an arc in the graph. Alternative definitions of centrality are illustratedon a simple directed network known as the football dataset [70]. The network records 35 soccer teams that participated in the 1998 World Championship in Paris. Every edge records the number of national team players of one country who play in the league of another country.

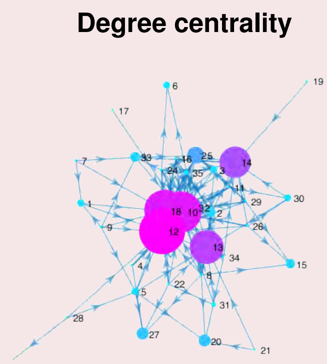


Figure 6: Nodes colored according to degree centrality.

The in/out-degree of $i \in \mathcal{V}$ are defined as

$$\text{in-deg}(i) = |\{j \in \mathcal{V} : w_{ij} \neq 0\}|,$$
$$\text{out-deg}(i) = |\{j \in \mathcal{V} : w_{ji} \neq 0\}|,$$

respectively, where $|\mathcal{X}|$ denotes the cardinality of the set $\mathcal{X}$. In social systems, the degree corresponds to the number of paths of length 1 starting from a node. The *weighted* in/out-degree are defined as the sum of weights when analyzing weighted networks:

$$\text{in-deg}_{\mathbf{W}}(i) = \sum_{j\in\mathcal{V}} w_{ij}, \quad \text{out-deg}_{\mathbf{W}}(i) \quad = \sum_{j\in\mathcal{V}} w_{ji}.$$

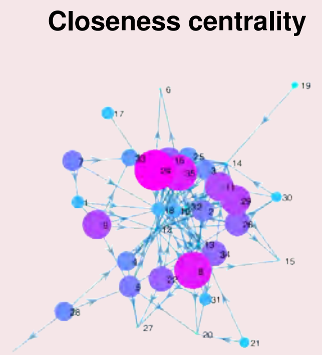


Figure 7: Nodes colored according to closeness centrality.

The *closeness centrality* of node i is defined as

$$c_i = \frac{1}{\sum_{j \in \mathcal{V} \setminus \{i\}} d_{ij}},$$

where d_{ij} denotes the length of the shortest path between i and j. This notion can be modified using other definitions of distances, as considered in [71], [72]. Closeness centrality for weighted graphs can be defined by introducing "weighted distance d_{ij}," that is, the minimal weight of all paths that connect i to j. The weight of a path is naturally defined as the sum of the weights on the traversed edges.

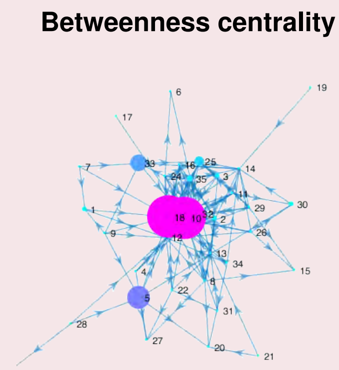


Figure 8: Nodes colored according to betweenness centrality.

The *betweenness* of node i is defined as

$$b_i = \sum_{j,k \in \mathcal{V}, j \neq k \neq i} \frac{|\mathcal{S}_i(j,k)|}{|\mathcal{S}(j,k)|},$$

where $\mathcal{S}(j,k)$ denotes the set of shortest paths from j to k, and $\mathcal{S}_i(j,k)$ IS the set of shortest paths from j to k that contain the node i. For weighted graphs, the length of each edge forming the paths in $\mathcal{S}(j,k)$ and $\mathcal{S}_i(j,k)$ is measured through the entries of the influence matrix $\mathbf{W}$.

Figure 9: Nodes colored according to eigenvalue centrality.

The idea of *eigenvector centrality* is based on a simple principle: A node is important if it is connected to other important nodes. This centrality measure is determined by the dominating (Perron-Frobenius) eigenvector $x^\star$ of some properly defined nonnegative matrix $\mathbf{A}$ that is compatible with the graph. Formally,

$$\mathbf{A}x^\star = \lambda x^\star, \quad \mathbb{1}_n^\top x^\star = 1, \quad x_i^* \geq 0 \, \forall i.$$

where $\lambda = \rho(\mathbf{A})$ is the maximal positive eigenvalue (being also the spectral radius) of $\mathbf{A}$. In the standard definition of eigenvector centrality [11], $\mathbf{A} = \mathbf{Adj}$ is the standard adjacency matrix. A more general construction

$$\mathbf{A}(\mathbf{M}) = (1-m)\mathbf{M} + \frac{m}{n}\mathbb{1}_n\mathbb{1}_n^\top,$$

where $\mathbf{M}$ is a *column stochastic* matrix and $m \in (0,1)$, arises in the definition of PageRank centrality [73].

Another relevant measure is represented by the node *betweenness* [71], [72]. Nodes with a high betweenness occupy critical positions in the network and are bridges between two groups of vertices within the network (since many paths in different groups must pass through this node). The *eigenvector centrality* of a node is a function of its neighbors, and the relevance is assigned according to the entries of the leading eigenvector $x^\star$ of a suitable weighted adjacency matrix of the network. Contrary to the degree centrality, this notion does not depend on the number of neighbors, but considers the relevance of its neighbors. In this way, a node with a few influent neighbors has larger eigenvector centrality than a node with various neighbors of limited influence. The most famous eigenvalue centrality measure is the PageRank centrality [74], which was introduced in the context of ranking of webpages. Many other centrality measures (such as Katz centrality [75], Bonacich centrality [76], [77], and harmonic influence centrality [14], [15],

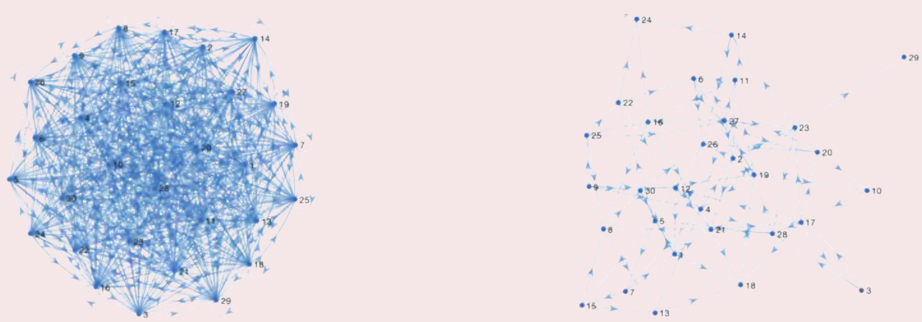
[78]) naturally arise as extensions of the eigenvector centrality and PageRank.

Sparsity structure of social network

A systematic study of the structure of social networks offers several metrics and algorithms for extracting low-dimensional network's features. Metrics can quantify global or local structural properties. The *network density* is an aggregate network metric defined as the ratio $|\mathcal{E}|/n^2$ of the number of social relationships observed in the network to the total number of possible relationships that could exist among nodes (that is, the proportion of ties within the network). A collection of large social network datasets is made available by Stanford Network Analysis Platform (SNAP, [79]) and can be visualized using the software GraphViz [80]. In **"Sparsity structure in OSN,"** the table reports the type, number of nodes, and network density of some social networks. Note that OSN have some common features:

- They are massive networks with a number of nodes $n = |\mathcal{V}|$ ranging from tens of thousands to millions;
- They are not dense, in the sense that the number of edges is not close to n^2 (that is, the maximal number of possible edges) and is, at most, of order of the network size.

Sparsity structure in OSN



Dense network Sparse network

"Most of real social networks are sparse" [Stanford Large Network Dataset Collection]

Name	Type	Nodes	Edges	Network density
ego-Facebook	Undirected	4039	88234	$5,408 \cdot 10^{-3}$
ego-Gplus	Directed	107614	13673453	$1,180 \cdot 10^{-3}$
ego-Twitter	Directed	81306	1768149	$2,674 \cdot 10^{-4}$
soc-Epinions1	Directed	75879	508837	$8,837 \cdot 10^{-5}$
soc-LiveJournal1	Directed	4847571	68993773	$2,936 \cdot 10^{-6}$
soc-Pokec	Directed	1632803	30622564	$1,148 \cdot 10^{-5}$
soc-Slashdot0922	Directed	82168	948464	$1,404 \cdot 10^{-4}$
wiki-Vote	Directed	7115	103689	$2,048 \cdot 10^{-3}$
gemsec-deezer	Undirected	143884	846915	$4,090 \cdot 10^{-5}$
gemsec-facebook	Undirected	134833	1380293	$7,592 \cdot 10^{-5}$

The distinction of dense and sparse graphs depends on the context. The index with which sparsity is commonly measured in network graphs is edge density [81]. Consider the following asymptotic definition of sparsity. Assume that a graph is *sparse* if the number of edges is not larger than a quantity that scales linearly in the number of nodes, that is, $|\mathcal{E}| \leq \alpha n$ with $\alpha \in (0, 1)$. There are other metrics to define sparsity, for example, the generalization of the Gini Index [82] for networks. Refer to [83] for an overview of the sparsity definitions adapted for networks. Another important statistical characteristics is the in- and out-degree distribution (see **"Centrality measures in weighted graphs"**). If the in-degree of a node is small compared to the network size, then the corresponding row in the influence matrix $\mathbf{W}$ is *sparse* and contains few nonzero entries (see **"Sparse models"**). Many real-world networks exhibit power-law degree distributions [11]. Remarkably, such a distribution has been discovered in the early works on sociometry [84]. The fraction of nodes with degree k is distributed as

$$p_{\text{deg}}(k) \sim k^{-\gamma} \tag{1}$$

for some exponent $\gamma > 1$ and minimum degree k_{min}. Networks with power-law distributions are called "scale-free" because power laws have the same functional form at all scales, that is, the power law $p_{\text{deg}}(k)$ remains unchanged (other than a multiplicative factor) when rescaling the independent variable k [as it satisfies $p_{\text{deg}}(\alpha k) = \alpha^{-\gamma} p_{\text{deg}}(k)$]. In [85], the structural properties of Facebook Ego-networks are analyzed. Ego-networks are well studied, as they capture local information about network structure from the perspective of a vertex. The Ego-network of a focal node (called Ego) is defined as the subgraph induced over nodes that are directly connected to it, but excluding the ego itself. Note that since the ego node is removed from the network, an Ego-network can be disconnected. **"Degree distribution in Facebook Ego-Networks"** shows the normalized degree distribution of three Ego-Networks [79]. As shown, some are more "concentrated" around a mean value, while others show a power-law decay with smaller exponent γ. As discussed in the next section, this concentration property plays a crucial role in the inference of trust network from few data.

Degree distribution in Facebook Ego-Networks

Ego-networks analysis represents a common tool for the investigation of the relationships between individuals and their peers in online social networks [86]. Moreover, the structural properties of Ego-networks are shown to be correlated to many aspects of human social behavior, such as willingness to cooperate and share resources [87]. The Ego-network of a focal node is defined as the subgraph induced over nodes that are directly connected to it, but excluding the ego itself. Figure 10 depicts the empirical degree distribution of three Facebook Ego-Networks retrieved from the Stanford Network Database [79]. Note that some degree distributions are more concentrated around a mean value, while others show a power-law decay. The tails of the distribution are well approximated by (1) with $\gamma \in [1.2, 3]$.

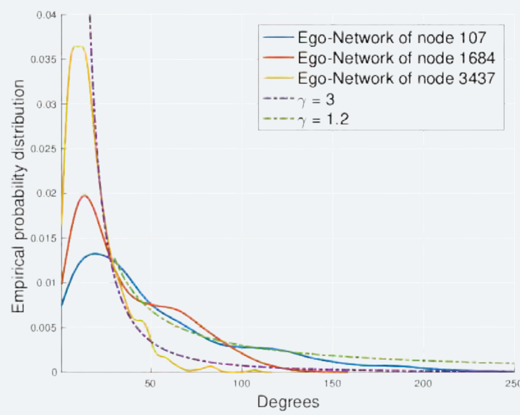
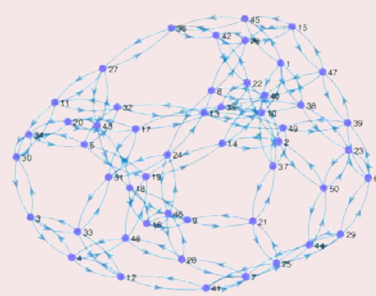


Figure 10: Degree distribution of Facebook Ego-networks.

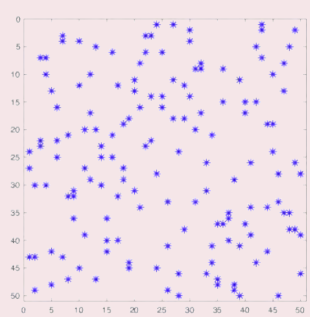
Examining the Twitter networks of 14 destination marketing organizations, Finally, other networks exhibit the presence of few clusters [85], that is, a community of individuals with dense friendship patterns internally and sparse friendships externally. This inherent tendency to cluster is measured by the *average clustering* coefficient [89]. These types of networks are described by an influence matrix that can be decomposed as a sum of a low-rank and sparse matrix. To better address some ideas, **"Sparse models"** provides different examples that summarize how sparsity can be exploited for social networks analysis.

Sparse models

Sparse models to represent high-dimensional data have been used in several areas, such as in statistics, signal and image processing, machine learning, coding, and control theory [90]. Intuitively, data are considered sparse or compressible if they are so highly correlated that only a few degrees of freedom compared to their ambient dimension are significant. This general definition leads many possible interpretations and alternative measures of sparsity can be defined according to the data and applications. The simplest definition is the sparsity in the elements. A signal is sparse if the number of nonzeros (or significantly different from zero) are few compared to the signal dimension. Moreover, a signal $x \in \mathbb{R}^n$ is k-sparse if $\|x\|_0 \doteq |\{i \in \{1,\ldots,n\} : x_i \neq 0\}| \leq k$ with $k \ll n$. In social networks analysis, sparsity can be exploited in several ways. This article provides examples to easily address ideas. FEW FRIENDS. If, from a sociological perspective, an agent is influenced by few friends, then the in-degree is low compared to the size of the network. As a consequence, the corresponding adjacency matrix is sparse, that is, it contains few nonzero elements. In the following figures, a typical sparse adjacency matrix is depicted together with the signal obtained by stacking the matrix by columns. Note that only a small portion of elements is different from zero.



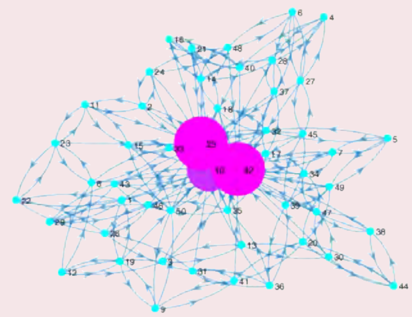
Few friends.



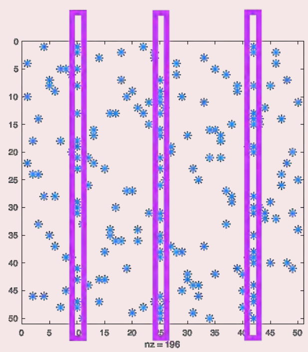
Sparsity in the elements of adjacency matrix.

FEW INFLUENCERS. If the networks contains few leaders (that is, few individuals influencing many people in the network), then the adjacency matrix will exhibit a sparse structure with few dense columns. The adjacency matrix of a network with five influencers is shown in the following figure on the left. On the right, the elements of the matrix are stacked by columns. Note that the signal is sparse with few dense patterns.

In literature, this feature is also known as block-sparsity.

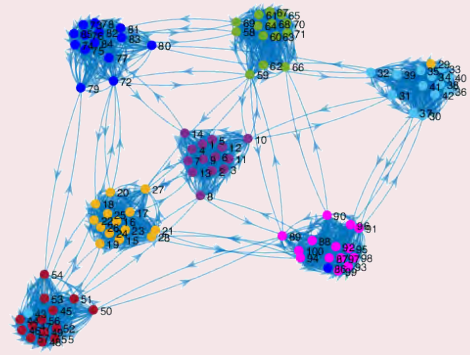


Few influencers.

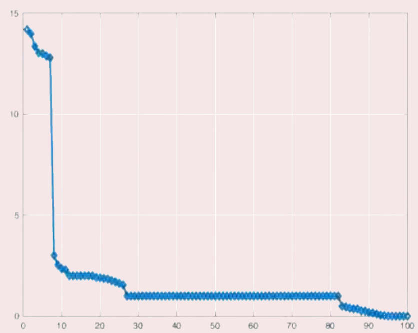


Block-sparsity in the adjacency matrix.

Few communities. Other networks show the presence of few communities, that is, a set of individuals with dense friendship patterns internally and sparse friendships externally. For these type of networks, the adjacency matrix that can be decomposed as $A = L + S$, where L is a low-rank matrix and S is a sparse matrix. In the following figures, a typical adjacency matrix of a network with few communities (left) and the corresponding eigenvalues in absolute values (right). Note that the eigenvalues are highly compressible and only four eigenvalues (even to the number of communities) contain the most energy of the signal.



Few communities.



Sparsity in the eigenvalues of the adjacency matrix.

Sparse models for multidimensional metworks

Multidimensional networks describe different types of relations among nodes. For example, friendships in a social network may arise for various reasons (for example, because users are colleagues, teammates in sports, or share some other hobbies). In this case, consider multiplex networks, where different layers of interconnections can be distinguished that correspond to different types of relations. Another real-world example is when a social group discusses several issues in parallel. For example, Twitter is comprised of microblogs, and users express opinions on *different topics*. The influence between the users is *topic-dependent* (for example, networks are sometimes called *heterogeneous* [91]). Each layer in the multilayer network considers the influence relations among people when they discuss a certain subject. The analysis of multiplex networks is an active field of research (see [92] and references therein). If the underlying social network is composed of the same individuals, then it is expected that the social systems share a common feature. The above intuition entails that the networks describing the micro-level mechanisms of social influence with respect to topics are not completely independent. It follows that (besides the sparsity model describing the degrees of freedom of each network), the model must be augmented by considering the correlations of the networks relative to the different topics. In this sense, [93] introduces correlated models. The first model, *the common component model* ($\mathcal{M}_{\text{cc}}$), considers the cases where the networks relative to the different topics $\ell = 1, \ldots, n_\ell$ only differ for a few edges. In this case, all influence matrices share a common base and contain an innovation. Formally, the influence matrix $\mathbf{W}^{(\ell)}$ describing the interaction network relative to the ℓ-th topic is decomposed as

$$\mathbf{W}^{(\ell)} = \bar{\mathbf{W}} + \boldsymbol{\delta}\mathbf{W}^{(\ell)}, \tag{2}$$

where the matrices $\bar{\mathbf{W}}$ and $\boldsymbol{\delta}\mathbf{W}^{(\ell)}$ (representing respectively the common part and the innovation part) are both sparse (see examples in **"Multidimensional networks"**). The second model, the *common support model* ($\mathcal{M}_{\text{cs}}$) describes situations where the topology is equal for all the different topics but the weights are different. This is captured by a model in which all transition matrices share a common support $\Omega \subseteq \{1, \ldots, n\} \times \{1, \ldots, n\}$, that is,

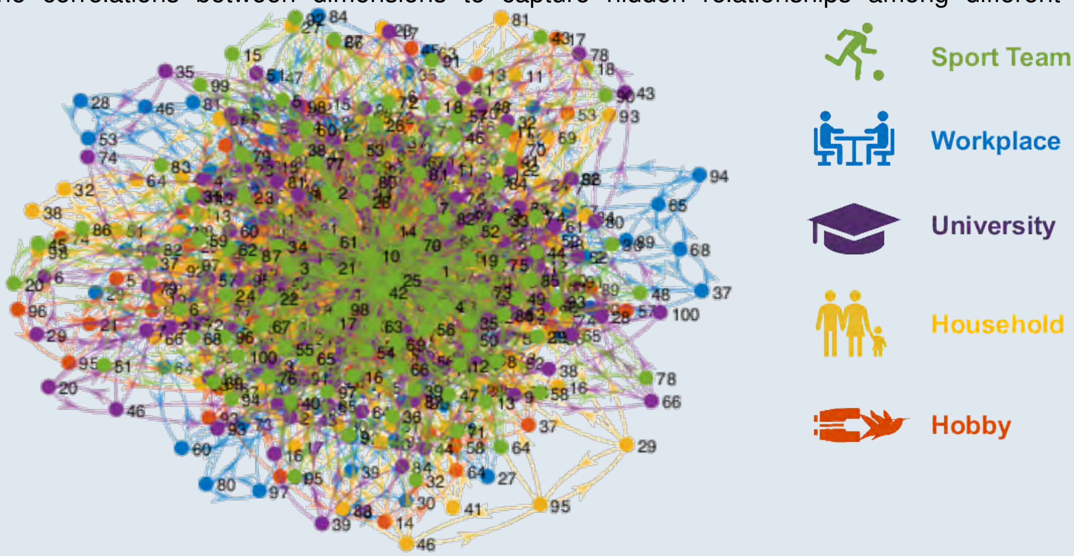
$$\mathbf{W}_{ij}^{(\ell)} \neq 0 \iff (i, j) \in \Omega, \forall \ell \in \{1, \ldots, m\}. \tag{3}$$

An example of this model can be found in deliberative groups that address a sequence of issues, such as department faculty in universities and boards of directors in large organizations. Empirical findings show [94] that the weights evolve according to a natural social process known as reflected appraisal [95], [96]. The model of this process proposed in [94] is squarely based on the Friedkin-Johnsen model of opinion formation and will be considered in Section . Summarizing,

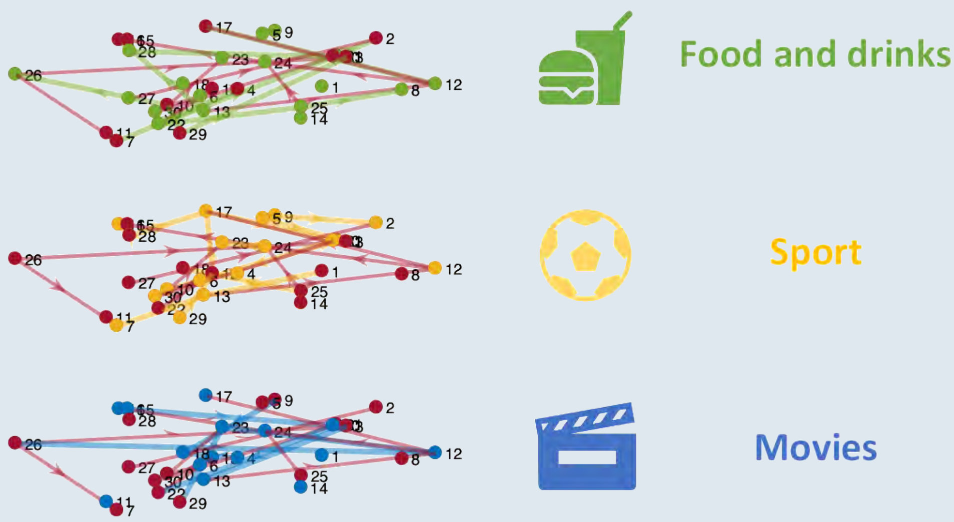
any efficient technique for social media modeling, analysis, and optimization must consider the large size of the networks and exploit the notion of sparsity as a structural constraint. From the previous discussion, note that the key ingredient for performing a social influence analysis is the knowledge of influence matrix $\mathbf{W}$. The next sections focus on algorithms inferring matrix $\mathbf{W}$ and related computational aspects and consider two different approaches. The "static" method addresses the inference of matrix $\mathbf{W}$ from samples of some observables $\{x_j\}_{j \in \mathcal{V}}$, whereas the "dynamic" approach addresses identification of opinion formation models.

Multidimensional networks

Multidimensional (multiplex, multilayer) networks allow for distinguishing between different kinds of links between the nodes and naturally arise in social sciences [92], economics and finance [97], transportation [98], and biology [99]. There are multiple ways to define a multidimensional network. For simplicity, consider networks denoted by a triple $\mathcal{G} = (\mathcal{V}, \mathcal{E}, \mathcal{L})$, where $\mathcal{V}$ is a set of nodes, $\mathcal{L}$ is the set of layers, $\mathcal{E} = \bigcup_{d \in \mathcal{L}} \mathcal{E}_d$ is the set of edges, and $\mathcal{E}_d$ the set of edges at layer d. Temporal networks are a special type of multiplex networks with explicit dimensions and can be represented as a sequence of graphs, where a single dimension is represented as a separate layer. Therefore, it is mandatory in this new framework to: (a) generalize the centrality measures defined for classical monodimensional networks; and (b) study the correlations between dimensions to capture hidden relationships among different layers.



This picture represents the relations between different individuals. Different colors correspond to different origins of the friendship (for example, friendships arisen in sport teams, at the workplace, and at the university). Consider the single type of networks as separate graphs and the multiplex networks given by the union of all graphs. In this case, each layer has been generated independently as an Erdős-Rényi graph.



In this example, consider a three-dimensional networks. The network is constituted by different layers and represent the influence network of a community based on the topic under discussion (for example, food and drinks, sport, and movies). Each layer is constituted by a sparse common component (the subgraph in red) and a sparse innovation component (the subgraph in orange, purple, and light blue, respectively).

In [100], a measure is defined to quantify how much two dimensions are similar. These measures can be seen as an extension of the classical Jaccard correlation coefficient to address more than two sets. Consider an $\mathcal{L}$-dimensional network. Let $\mathcal{D} \subseteq \mathcal{L}$ be a set dimensions of a network $\mathcal{G} = (\mathcal{V}, \mathcal{E}, \mathcal{L})$. The pair $\mathcal{D}$-correlation is defined as

$$\rho_{\mathcal{D}} = \frac{|\bigcap_{d\in\mathcal{D}} \mathcal{E}_d|}{|\bigcup_{d\in\mathcal{D}} \mathcal{E}_d|}.$$

In the example, $\rho_{\{\mathsf{M,S}\}} = 0.3636$, $\rho_{\{\mathsf{M,F\&D}\}} = 0.3871$ and $\rho_{\{\mathsf{S,FD}\}} = 0.3333$.

Learning graphs from data

The influence network estimation problem discussed in this article represents a special instance of the general problem of reconstructing the graph topology from data measured on the nodes. This problem, known under the name of *graph learning* or *network inference*, has seen an increasing interest in the past years. Interested readers are referred to [101], on which this section is largely based. The literature distinguishes between several approaches for the influence network estimation. These mainly depend on the assumptions on the networks under observation and the available data. The methods are categorized into three classes: i) statistical models, ii) learning models for social similarity and influence, and iii) model-based approach. Most methods address undirected graphs and nondynamical (static) variables, and the extensions to directed and/or dynamically varying topologies are usually rather complex. For this reason, this section mostly focuses on the simpler case of static undirected graphs.

Statistical models

Statistical models presume the availability of N measurements (usually scalar) at each node $i \in \mathcal{V}$,

$$x_i(1), \ldots, x_i(N), \quad i \in \mathcal{V}.$$

The main idea behind statistical models is to interpret the observed data as independent realizations of random variables $\{x_i\}, i \in \mathcal{V}$ whose joint probability distribution is determined by the topology of the graph $\mathcal{G}$. Hence, a connection between two nodes translates into a statistical correlation between the signals at those nodes. In particular, one can introduce the so-called *probabilistic graphical models* [66], [102], in which data are interpreted as multiple outcomes of random experiments. A graphical model is introduced to capture the conditional dependence between random variables. When applied to social opinion analysis, these representations are sometimes referred to as “model-free,” in the sense that they do not exploit any analytical model of the dynamical evolution of the opinion (but only assume statistical correlation between opinions). In the simple case of undirected graphs and continuous variables, the most popular models proposed in the literature are *Markov random fields* (MRFs) [103]. Given a graph $\mathcal{G} = \{\mathcal{V}, \mathcal{E}\}$, MRFs are postulated by requiring that the random variables at the different nodes satisfy a series of local Markov properties. Of particular interest is the so-called *pairwise Markov property*, which states that two variables are conditionally independent given all other variables if and only if they are not connected by an edge. That is,

$$x_i \perp\!\!\!\perp x_j \mid \{x_w\}_{w \in \mathcal{V} \setminus \{i,j\}} \iff (i,j) \notin \mathcal{E},$$

where $\perp\!\!\!\perp$ denotes statistical independence. It is shown in [104] that the above conditional independence property holds if the probability mass function (if the beliefs are discrete) or probability density function (if the beliefs are continuous) belongs to the family of exponential distributions. That is,, it is of the form

$$p(\mathbf{x}|\mathbf{\Theta}) = \frac{1}{Z(\mathbf{\Theta})} \exp\left[-\frac{1}{2} \left(\sum_{v \in \mathcal{V}} \theta_{ii} x_j^2 + \sum_{(i,j) \in \mathcal{E}} \theta_{ij} x_i x_j \right) \right],$$

where $\mathbf{x} \doteq [x_1, \ldots, x_n]^\top$ is a collection of the random variables across nodes, $\mathbf{\Theta} = [\theta_{ij}]$ is a parameter matrix, and $Z(\mathbf{\Theta})$ is a normalization constant. Conditional independence between x_i

and x_j translates into $\theta_{ij} = 0$. In other words, the parameter matrix $\boldsymbol{\Theta}$ is adapted to the graph. This class is named *exponential random graphs* or p^* models. In the literature, estimation schemes for such graphs based on Monte Carlo maximum likelihood estimation have been proposed [105]. As observed in [106], these methods naturally extend the classical statistical approach based on estimating the *partial* (Pearson) correlation coefficient starting from the observations $\mathbf{x}$. A commonly adopted assumption in exponential random graphs stipulates that the observations are realizations of the multivariate Gaussian distribution

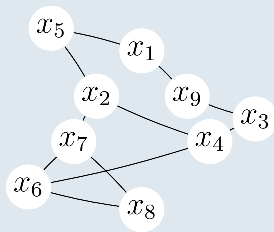
$$p(\mathbf{x}|\boldsymbol{\Theta}) = \frac{\det(\boldsymbol{\Theta})^{1/2}}{(2\pi)^{N/2}} \exp\left(-\frac{1}{2}\mathbf{x}^\top \boldsymbol{\Theta}\mathbf{x}\right),$$

where $\boldsymbol{\Theta}$ represents the so-called precision matrix, that is, the inverse of the covariance matrix $\boldsymbol{\Sigma}$ $\boldsymbol{\Theta} = \boldsymbol{\Sigma}^{-1}$. This leads to the family of *Gaussian graphical models* [104]. It can be observed that (in this case) the existence of a nonzero entry in the precision matrix, immediately implies a partial correlation between the corresponding random variables. The goal then becomes to estimate the precision matrix from the observed data $\{x_j\}_{j\in\mathcal{V}}$. To this end, several procedures have been proposed for computing the maximum-likelihood (ML) estimator via a log-determinant program. In this class of algorithms, the so-called graphical Lasso (G-Lasso) [107] method has become extremely popular **"Gaussian graphical models and G-Lasso."** Note that, although the convergence of G-Lasso is guaranteed under suitable conditions, this method has some drawbacks. First, the whole procedure only works in the case of undirected networks. Second, in many contexts (for example, the opinion formation processes discussed in this article), data are the result of a dynamic process. This situation is not well captured by the G-Lasso framework, since data can be highly correlated for these problems, leading to a dense precision matrix. Finally, the sample covariance matrix may fail to have full rank due to the lack of observed data, giving rise to numerical problems in the identification of the network. It is worth emphasizing that the estimation of the precision matrix via graphical models does not allow a direct interpretation of social influence but is able to reflect pairwise correlation between opinions in the social system. The estimation of social influence, however, is primarily aimed at predicting a direct causal effect of this influence.

Gaussian graphical models and G-Lasso

Graphical models are graphs capturing the relationships between many variables, providing a compact representation of joint probability distributions. In these models, the nodes correspond to random variables and edges represent statistical dependencies between node pairs. In GGM , the variables at each node are normally distributed, $\mathbf{x} \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma}_\mathbf{x})$, and for any i and $j \in \mathcal{V}$, a zero in the i, j entry of the precision matrix means conditional independence (given all other variables):

$$x_i \perp\!\!\!\perp x_j \mid \{x_w\}_{w\in\mathcal{V}\setminus\{i,j\}} \iff \theta_{ij} = 0 \quad \boldsymbol{\Theta} = {\boldsymbol{\Sigma}_\mathbf{x}}^{-1}.$$



$$\Theta = \begin{bmatrix} 0 & 0 & 0 & 0 & \star & 0 & 0 & 0 & \star \\ 0 & 0 & 0 & \star & \star & 0 & \star & 0 & 0 \\ 0 & 0 & 0 & \star & 0 & 0 & 0 & 0 & \star \\ 0 & \star & \star & 0 & 0 & \star & 0 & 0 & 0 \\ \star & \star & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & \star & 0 & 0 & \star & \star & 0 \\ 0 & \star & 0 & 0 & 0 & \star & 0 & \star & 0 \\ 0 & 0 & 0 & 0 & 0 & \star & \star & 0 & 0 \\ \star & 0 & \star & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

$$(i,j) \notin \mathcal{E} \Longrightarrow \theta_{ij} = 0$$

Consider N observations $\{\mathbf{x}(1), \mathbf{x}(2), \ldots, \mathbf{x}(N)\}$ from the multivariate Gaussian distribution, This work is interested in estimating the precision matrix $\Theta = \mathbf{\Sigma}^{-1}$. The classical maximum likelihood (ML) estimator is obtained by solving the optimization problem

$$\hat{\mathbf{\Theta}}_{\text{ML}} = \max_{\Theta \succeq 0} \log \det(\mathbf{\Theta}) - \operatorname{tr}(\mathbf{S}\mathbf{\Theta}) \qquad \text{with} \quad \mathbf{S} = \frac{1}{n} \sum_{k=1}^{n} \mathbf{x}(k)\mathbf{x}(k)^\top,$$

where $\operatorname{tr}(\cdot)$ denotes the trace of a matrix. Classical theory guarantees that in the high-dimensional regime, $\hat{\mathbf{\Theta}}_{\text{ML}}$ converges to the truth as sample size $N \to \infty$.

In practice, we are often in the regime where sample size N is small compared to the dimension n. Therefore, S is not full rank and the ML estimation problem does not admit a unique solution. The main approach in these cases is to assume that many pairs of variables are conditionally independent, that is, many links are missing in the graphical model or, equivalently, $\mathbf{\Theta}$ *is sparse*. The key idea in G-Lasso [108] is to treat each node as a response variable and solve the convex program

$$\hat{\mathbf{\Theta}}_{\text{G-lasso}} = \max_{\mathbf{\Theta} \succeq 0} \log \det(\mathbf{\Theta}) - \operatorname{tr}(\mathbf{S}\mathbf{\Theta}) - \rho \|\mathbf{\Theta}\|_1$$

where ρ tunes the number of zero entries in $\mathbf{\Theta}$.

Graph Signal Processing

Recent years have witnessed a growing interest from the signal processing community in analysis of signals that are supported on the vertex set of weighted graphs, leading to the field of graph signal processing (GSP), [109]. By generalizing classical signal processing concepts and tools, GSP enables the processing and analysis of signals that lie on structured but irregular domains. In particular, GSP allows for redefining concepts as such as Fourier transform, filtering, and frequency response for data residing on graphs. Note that the signals in the graph are not time-dependent. They instead vary spatially, and their spatial dynamics are governed by the underlying graph. A brief overview of GSP is provided in **"Graph signal processing."**

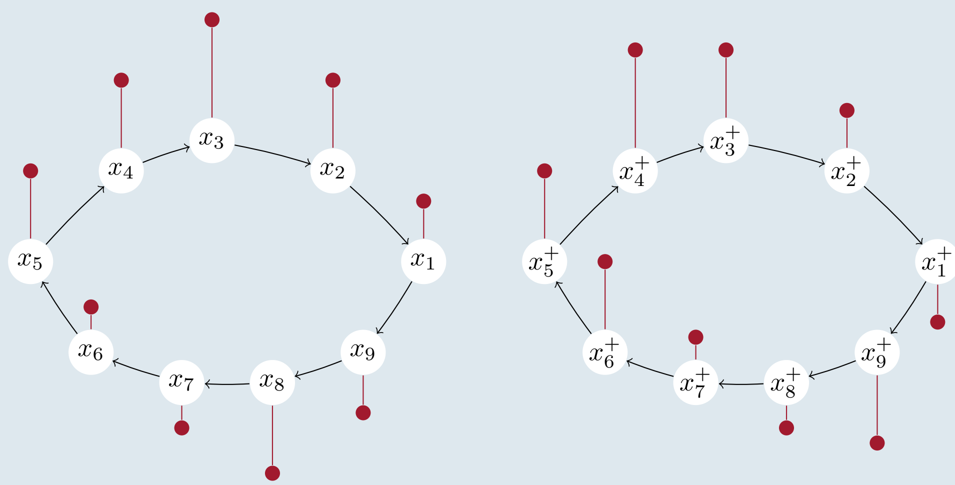
Graph signal processing

The rapidly growing field of graph signal processing (GSP) provides tools to represent signals that are supported on the vertices of a graph. A *graph signal* is defined as a function $\mathbf{x} : \mathcal{V} \to \mathbb{R}^n$ that assigns a scalar value to the vertices of a graph. It can be represented as a vector $\mathbf{x} \in \mathbb{R}^n$, where x_i stores the value of the signal on the i-th vertex.

A simple way to understand the basics of GSP is to consider how the classical concept of shift operator is extended to graph signals. First, observe that a periodic discrete-time signal can be represented by a circular directed unweighted graph, in which the k-th node represents the value of the signal x at the discrete-time instant k. The next figure represents a signal $\mathbf{x}$ (left) and its shifted version $\mathbf{x}^+$ (right), which follow the classical relationship $\mathbf{x}^+ = \textsc{Shift}(\mathbf{x}) \doteq \mathbf{S}\mathbf{x}$, defined as follows

$$\begin{aligned} x_i^+ &= x_{i-1}, \quad i = 2, \ldots, N \\ x_1^+ &= x_N, \end{aligned}$$

where the last equation follows from the *circular shift* assumption. Note that, in this case, the shift operator $\mathbf{S}$ coincides with the adjacency matrix $\mathbf{Adj}$ of the directed graph.



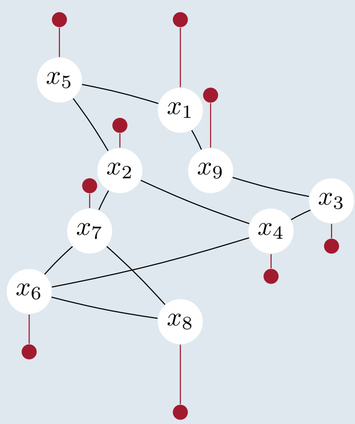
a) A periodic signal on a directed graph $\mathbf{x} = x_1, x_2, \ldots, x_8, x_9$ and b) its shifted version.

$$\mathbf{x}^+ = \mathbf{S}\mathbf{x}, \qquad \mathbf{S} = \begin{bmatrix} 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

This parallelism between shift operators and graphs may be extended to general graphs. In GSP, given a graph $\mathcal{G}$, a **graph shift operator** is defined as a matrix $\mathbf{S} \in \mathbb{R}^{n,n}$ *adapted to the graph*, and the shift operation is given by $\mathbf{S}\mathbf{x}$. Different choices of $\mathbf{S}$ define different shifts. For undirected graphs, the most typical choice of graph shift operator is the Laplacian $\mathbf{L}$: for any graph signal $\mathbf{x}$, define the new signal $\mathbf{x}^+ = \mathbf{S}\mathbf{x} = \mathbf{L}\mathbf{x}$, whose element x_u^+ is given by

$$x_u^+ = [\mathbf{L}\mathbf{x}]_u = \sum_{v \in \mathbb{N}_u} w_{ij}(x_i - x_j).$$

From the above formulation, it can be easily observed that the Laplacian acts as a difference operator on graph signals.



$$\mathbf{S} = \mathbf{L} = \begin{bmatrix} 2 & 0 & 0 & 0 & -\frac{1}{2} & 0 & 0 & 0 & -\frac{1}{2} \\ 0 & 3 & 0 & -\frac{1}{3} & -\frac{1}{3} & 0 & -\frac{1}{3} & 0 & 0 \\ 0 & 0 & 2 & -\frac{1}{2} & 0 & 0 & 0 & 0 & -\frac{1}{2} \\ 0 & -\frac{1}{3} & -\frac{1}{3} & 3 & 0 & -\frac{1}{3} & 0 & 0 & 0 \\ -\frac{1}{2} & -\frac{1}{2} & 0 & 0 & 2 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & -\frac{1}{3} & 0 & 3 & -\frac{1}{3} & -\frac{1}{3} & 0 \\ 0 & -\frac{1}{3} & 0 & 0 & 0 & -\frac{1}{3} & 3 & -\frac{1}{3} & 0 \\ 0 & 0 & 0 & 0 & 0 & -\frac{1}{2} & -\frac{1}{2} & 2 & 0 \\ -\frac{1}{2} & 0 & -\frac{1}{2} & 0 & 0 & 0 & 0 & 0 & 2 \end{bmatrix}$$

b) A signal defined on a generic *undirected* graph, and the corresponding graph shift operator, defined in terms of the Laplacian.

From the definition of the graph shift operator, the extension of the concept of Fourier transform to graph signal descends almost immediately [110]. For $\mathbf{S} = \mathbf{L}$ (under the assumption of connectivity of the network), if one considers the eigenvalue decomposition of $\mathbf{L} = \mathbf{L}^\top$,

$$\mathbf{L} = \mathbf{U}\boldsymbol{\Lambda}\mathbf{U}^\top, \quad \boldsymbol{\Lambda} = \mathrm{diag}(\lambda_1, \ldots, \lambda_n), \quad \mathbf{U} = [\mathbf{u}_1 \cdots \mathbf{u}_n],$$

where $\mathbf{U}$ is the eigenvector matrix, that is, the matrix containing the eigenvectors of $\mathbf{L}$ as columns (which are orthonormal, being $\mathbf{L}$ symmetric) and λ_i-s are the eigenvalues (which are real and ordered), with $0 < \lambda_2 \leq \cdots \leq \lambda_n$. The **graph Fourier transform** (GFT) associated with the Laplacian may then be defined as

$$\tilde{x}_k \doteq \mathbf{u}_k^\top \mathbf{x} = \sum_{j=1}^{n} x_j [\mathbf{u}_k]_j.$$

Note that the Laplacian-based GFT only works for undirected graphs. Extensions to directed graphs are nontrivial, since the GFT definition does not cover situations where $\mathbf{L}$ has complex eigenvalues or is not diagonalizable.

While the main directions of research in GSP focus on the development of methods for analyzing signals defined over given *known* graphs, the inverse problem concerned with *learning* the graph topology from measurements of the signals on the graph has also been considered. The existing mathematical results adopt specific assumptions about the characteristics of the graph

Fourier transform. The most common approach for GSP-based graph topology reconstruction is based on the assumption that the underlying graph signal is *smooth* on the graph. That is, the links in the graph should be chosen in such a way that signals on neighboring nodes are close to each other. As a measure of smoothness of the signal $\mathbf{x}$ on the graph $\mathcal{G}$, the so-called Laplacian quadratic form is usually adopted [111],

$$\mathbf{x}^\top \mathbf{L}\mathbf{x} = \frac{1}{2}\sum_{i,j} w_{ij}(x_i - x_j)^2. \tag{4}$$

Several approaches have been proposed in the literature for learning a graph (or, in this case, its Laplacian matrix $\mathbf{L}$), such that the Laplacian quadratic cost (4) is small (that is the signal variations on the resulting graph is small). The reader is referred to [101] for a detailed overview of this approach whose central step is to solve the optimization problem

$$\min_{\mathbf{L},\mathbf{y}} \|\mathbf{x} - \mathbf{y}\|_2^2 + \alpha \mathbf{y}^\top \mathbf{L}\mathbf{y}.$$

The first term enforces data fidelity, and the second one enforces smoothness of the signal. This approach is extended in subsequent works [111], [112], by adding additional constraints on the Laplacian $\mathbf{L}$, thus allowing for the volume of the graph to be fixed and imposing specific structures on the graph. Other GSP-based approaches for deriving topological characteristics of the graph are based on graph signal measurement and assume that the graph signals are generated by applying a graph-filtering operation to a latent signal. In particular, the graph signal $\mathbf{x}$ is assumed to be generated by a *diffusion process* of the form

$$\mathbf{x} = \sum_{k=0}^{K} \alpha_k \mathbf{S}^k \mathbf{u},$$

where $\mathbf{S}$ is a given graph operator (again, usually the Laplacian matrix $\mathbf{L}$) capturing the graph connectivity. The ensuing algorithm is well-suited for learning graph topologies when the observations are the result of a diffusion process on a graph. This is the case for many diffusion dynamics in social systems. The existing methods for reconstruction stem from the observation that, when the "input" signal $\mathbf{u}$ is uncorrelated (white noise) and the graph is undirected, then the eigenvalues of $\mathbf{S}$ coincide with the eigenvalues of the covariance matrix $\mathbf{\Sigma}_\mathbf{x}$ of $\mathbf{x}$. This, in turn, may be approximated via the sample covariance. Finally, note the approaches using spectral graph dictionaries for efficient signal representation [113]. In this case, a graph signal diffusion model is envisioned, which represents data as (sparse) combinations of overlapping local patterns that reside on the graph.

Model based learning of directed and dynamical graphs

As discussed, the large majority of the graph learning approaches available in the literature address *undirected* and *stationary* graphs, whereas their extensions to directed graphs meet serious difficulties. In the case of probabilistic graphical models, for instance, directed graphical models [also called Bayesian Networks or Belief Networks (BNs)] require the introduction of a more complicated notion of independence, which considers the asymmetry of interconnections. In GSP-based techniques, directionality of the graph destroys the symmetry of its operator $\mathbf{S}$, thus Complicating the mere definition of the graph Fourier transform (GFT). In many contexts (as in the case of social interactions reconstruction, which represents the main focus of this work), learning directed graphs is more desirable, especially for those cases where the edges directions translate to causal dependencies between the variables that the vertices represent. In this case, model-based approaches appear more naturally. The main assumption is that data are results of a dynamical process, and the problem is an inverse optimization problem exploiting prior information on the model. This research is related to sparse vector autoregressive (SVAR) estimation [106], [114], [115], inverse optimization from partial samples [116], and models from opinion dynamics [53], [117]. The next section provides an overview of the main models introduced in the literature.

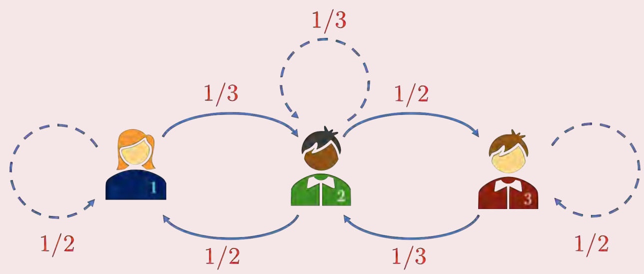
Social influence in opinion dynamics

The approaches to social influence discussed up to now represent a social network as either a weighted graph or a probabilistic graphical model. An alternative approach, leading to the so-called *social influence network theory* (SINT) [58], [118], considers a social network as a *dynamical system*. The relevant mathematical models describe diffusion of some information over the network, which can manifest itself as evolution of individual opinions, attitudes, or beliefs. The individuals interact (during face-to-face meetings or via social media) and display their opinions on issues to each other. Based on the opinions displayed to them, each individual updates their own opinion on an issue. The within-individual mechanisms of opinion assimilation are related to psychological studies on information integration [119] and cognitive dissonance [120]. Their mathematical models are currently limited to simple opinion update rules, such as iterative averaging. In such simplified models, social influence is naturally represented by *influence weights* an individual assigns to their own and others' opinions. Models are considered that stem from the French-Harary-DeGroot's model of iterative opinion pooling.

Original French's model [63]

- a group of n individuals are associated with nodes of a directed graph;
- individual i holds an *opinion* x_i, assumed to be a scalar real value;
- individual j discloses their opinion to individual i if the graph has a directed arc (j, i);
- individuals know their own opinions, and thus each node in the graph has a self-arc;
- at each period $k = 0, 1, \ldots$, an individual updates their opinion to the *mean values* of all opinions

displayed to them.



For instance, the graph shown in the picture induces the opinion formation process

$$\begin{bmatrix} x_1(k+1) \\ x_2(k+1) \\ x_3(k+1) \end{bmatrix} = \begin{pmatrix} 1/2 & 1/2 & 0 \\ 1/3 & 1/3 & 1/3 \\ 0 & 1/2 & 1/2 \end{pmatrix} \begin{bmatrix} x_1(k) \\ x_2(k) \\ x_3(k) \end{bmatrix}$$

The most typical behavior of the model is eventual consensus (unanimity) of the opinions. Due to the positivity of self-weights, consensus (for an arbitrary initial condition) is achieved if and only if [15], [121], [122] the graph has a node from which all other nodes are reachable (that is, there is an agent who influences, directly or indirectly, all other agents).

The French-Harary-DeGroot model

One of the simplest models of opinion formation has been proposed by French in his seminal work on social power [63] and later examined by the renowned graph theorist F. Harary [121], [123] (and discussed in **"Original French's model"**). The most known, however is its generalized version proposed by DeGroot [124] (and, independently, by Lehrer [125], [126]),

$$\mathbf{x}_i(k+1) = \sum_{j=1}^{n} w_{ij}\mathbf{x}_j(k), \quad i = 1, \ldots, n. \tag{5}$$

Here $\mathbf{x}_i(k)$ stands for the opinion of agent i at the kth stage of the opinion evolution, and $\mathbf{W} = [w_{ij}]$ is a row-stochastic matrix (a nonnegative matrix whose rows sum to 1). It is remarkable that the work [124] (unlike the pioneer works [63], [121]) introduced *multidimensional* opinions. Such opinions can represent individual's positions on several issues, for instance, the optimal distribution of resources between several entities [127] or subjective probability distribution of outcomes in some random experiment [124], [128]. Unless otherwise stated, assume the opinions to be *row* vectors:

$$\mathbf{x}_i(k) = [x_i^{(1)}(k), \ldots, x_i^{(m)}(k)].$$

It is convenient to stack theses rows on top of each other, thus obtaining a $n \times m$ *matrix* of opinions:

$$\mathbf{X}(k) = \begin{bmatrix} \mathbf{x}_1(k) \\ \vdots \\ \mathbf{x}_n(k) \end{bmatrix} = [\mathbf{x}^{(1)}(k), \dots, \mathbf{x}^{(m)}(k)] \in \mathbb{R}^{n \times m}. \tag{6}$$

The ℓ-th column of this matrix $\mathbf{x}^{(\ell)}(k) = (x_1^{(\ell)}, \dots, x_n^{(\ell)})^\top$ contains the actors' positions on issue $\ell = 1, \dots, m$. DeGroot's model is then rewritten in the matrix form

$$\mathbf{X}(k+1) = \mathbf{W}\mathbf{X}(k), \quad k = 0, 1, \dots. \tag{7}$$

According to the DeGroot model, at each stage of the opinion iteration, individuals simultaneously update their opinions to convex combinations of all opinions disclosed to them. The weights w_{ij} of this convex combination serve as natural measures of mutual *influences* among individuals [14], [129]. Social influence can be thought of as a finite resource individuals distribute between themselves and their peers (this is modeled as a distribution of chips in the Friedkin-Johnsen experiment, see p. 8), The weight $w_{ij} \geq 0$ assigned by agent i to another agent j measures the importance of j's opinion for i. If $w_{ij} = 1$ (maximal value), agent i fully relies on agent j's opinion on the issue and is insensitive to the opinions of the others,

$$w_{ij} = 1 \Longleftrightarrow \mathbf{x}_i(k+1) = \mathbf{x}_j(k).$$

An individual assigning the maximal weight $w_{ii} = 1$ to self is often called *stubborn* (radical), being completely closed to social influence and keeping their opinion unchanged:

$$\mathbf{x}_i(k+1) = \mathbf{x}_i(k) = \ldots = \mathbf{x}_i(0). \tag{8}$$

If $w_{ij} = 0$, the opinion of agent j is either not disclosed to agent i or not taken into account by them. Mathematically, agent j's opinion at step k does not influence the opinion of agent i at the consecutive step $k+1$, however, it can *indirectly* influence i's opinions at the subsequent steps $k+2, k+3, \dots$ through the opinions of other individuals (a chain of influence $j \to j' \to j'' \to \ldots \to i$). Along with French's model, the DeGroot model (5) also predicts *consensus* of opinions, that is, the convergence of all opinion vectors $\mathbf{x}_i(k)$ to the same vector as $k \to \infty$. Equivalently, consensus means that the matrices $\mathbf{W}^k$ converge to a stochastic matrix of rank 1, that is,

$$\mathbf{W}^k \xrightarrow[k \to \infty]{} \mathbf{1}_n \mathbf{p}^\top, \quad \mathbf{p}^\top \mathbf{1} = 1. \tag{9}$$

Here, $\mathbf{p}$ is a nonnegative vector, being a left eigenvector of $\mathbf{W}$ so that $\mathbf{p}^\top \mathbf{W} = \mathbf{p}^\top$. This vector can be considered as a centrality measure on the social network (similar in spirit to the eigenvector centrality) and characterizing *social powers* of individuals [63], [129]. The vector of consensus opinions of the group is given by

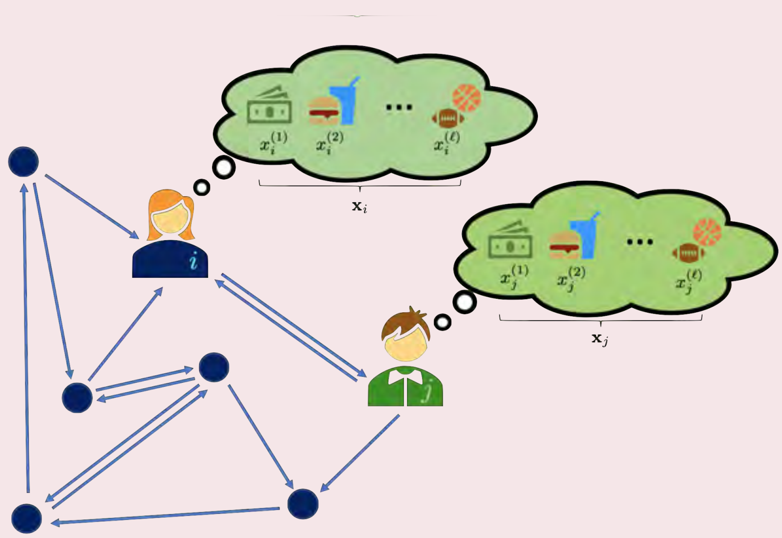
$$\lim_{k\to\infty} \mathbf{x}(k) = \sum_i p_i \mathbf{x}_i(0),$$

and hence the element p_i quantifies the influence of the initial opinion of individual i on the ultimate group's opinions. Matrices satisfying (9) are known as stochastic, indecomposable, aperiodic (SIA) matrices [130]. The SIA property can be proven for primitive (irreducible and aperiodic) [131], [132] matrices, for which $\mathbf{W}^m$ has strictly positive entries for sufficiently large exponent m [124]. Another standard criterion guarantees consensus if all diagonal entries w_{ii} are positive and the graph corresponding to $\mathbf{W}$ has a globally reachable node (some individual influences all others directly or indirectly) [133], [134]. A necessary and sufficient graph-theoretic condition for consensus can be found in [15], [122].

DeGroot's model as a dynamics over a graph

Social network $\leftrightarrow$ Weighted graph $\mathcal{G} = (\mathcal{V}, \mathcal{E}, \mathbf{W})$

- agents $\leftrightarrow v \in \mathcal{V}$
- interactions $\leftrightarrow \mathcal{E} \subseteq \mathcal{V} \times \mathcal{V}$
- influences $\leftrightarrow \mathbf{W} \in \mathbb{R}^{\mathcal{V}\times\mathcal{V}}$
- $w_{ij} = 0$ if $(i,j) \notin \mathcal{E}$
- opinions on issue $\ell \leftrightarrow x_v^{(\ell)}(k) \in \mathbb{R}$
- the *row* vectors of multidimensional opinions $\mathbf{x}_i(k) = (x_i^1(k), \ldots, x_i^m(k))$ obey (5);
- the vectors of positions on each issue $\mathbf{x}^{(\ell)}(k) = (x_1^{(\ell)}(k), \ldots, x_n^{(\ell)}(k))^\top$ evolve as $\mathbf{x}^{(\ell)}(k+1) = \mathbf{W}\mathbf{x}^{(\ell)}(k)$;
- the matrix of opinions (6) evolves in accordance with (7).



From consensus to disagreement

Since real social groups often fail to reach consensus, realistic models of opinion formation should be able to explain not only "regular" consensus behavior, but also various "disordered" behaviors featured by disagreement of opinions. To find such a model is a problem that has been studied since the 1960s [34], [135] and is known as *Abelson's diversity puzzle* (or the problem of *community cleavage* [14]). Most of the models portraying community cleavage (for instance the convergence of the opinions to several clusters) replace the DeGroot equation by nonlinear dynamics, taking into account various effects of information assimilation and integration within individual and communication between individuals [6], [35], [136]–[144]. The most studied are *bounded confidence* models [16], [128], [136], [137], [143] capturing the effect of homophily in social groups and assuming that individuals tend to assimilate opinions of like-minded individuals and meet dissimilar opinions with discretion or even ignore them. The simplest Hegselmann-Krause model [136] may be considered as an extension of the DeGroot dynamics (5) with opinion-dependent influence weights $w_{ij}(k) = w_{ij}(\mathbf{X}(k))$. Identifiability properties of nonlinear models are, however, almost unexplored. The models' dynamics are very sensitive to the structures of nonlinear couplings (for example, the lengths of confidence intervals in bounded confidence models), noise, and numerical errors [37]. Hence, in spite of some recent progress in identification of nonlinear networks [145]–[147], they are not considered in this survey. In linear models of opinion formation, the disagreement of opinions is typically explained by two factors: antagonistic interactions between individuals and their stubborness (reluctance to change the initial opinion). Models of the first type revise the basic assumption on the convex combination mechanism of opinion evolution and allow not only attraction of the agents' opinions, but also their *repulsion* [148]–[155]. The presence of negative influence is typically explained by "boomerang," reactance, and anticonformity effects [34], [150] (that is, the *resistance of some individuals to social influence*). The theory of signed (or "coopetitive") dynamical networks developed in the aforementioned literature is extremely important due to various applications in economics, physics and biology [156]. However, its applicability to social influence systems is still disputable for several reasons. The evidence of ubiquity of negative influence has not been secured experimentally. Since the first definition of boomerang effects [157], the empirical literature has concentrated on the special conditions under which these effects might arise in dyadic interpersonal interactions. Whereas positive and negative *relations* (friendship/enmity, trust/distrust) between individuals are ubiquitous, it is still unclear whether such relations really correspond to positive and negative influences [158]. Under natural assumption of *strong connectivity*, clustering of opinions in the presence of antagonistic interactions usually requires various forms of *structural balance* of positive and negative ties [149], [152], [154], [159], [160]. Clustering of opinions without structural balance is usually guaranteed by special hierarchical ("extended leader-following") structures in the graph [161], [162]. If the graph of social influence is time-varying, the absence of strong connectivity (understood in some uniform sense [163]) may lead to the existence of oscillatory solutions. At the same time, the usual DeGroot model is able to explain disagreement of opinions, assuming the existence of several stubborn individuals that are closed to social influence and keep their opinions unchanged (equivalently, their self-weights are maximal $w_{ii} = 1$) [164]. Further development of the DeGroot model with stubborn individuals has naturally led to the Friedkin-Johnsen (FJ) model (considered in the next subsection). Unlike many other models proposed in physical and engineering literature, the FJ model has been experimentally assessed on small- and medium-size groups [45], [46], [127], [165]. An essential part of these experiments is the empirical procedure of matrix $\mathbf{W}$ reconstruction, see **"Friedkin-Johnsen experiment."**

The Friedkin-Johnsen model

Whereas the DeGroot model allows stubborn individuals that are completely closed to social influence, the FJ model allows "partial" stubbornness, which is measured by a *susceptibility* coefficient $\lambda_i \in [0, 1]$. An agent with minimal susceptibility is the stubborn individual retaining its initial opinion (8), whereas the agent with maximal susceptibility assimilates to the others' opinions in accordance with the conventional DeGroot mechanism (5). In general, the individual opinion at each iteration is influenced by both the others' opinions and their initial opinion,

$$\mathbf{x}_i(k+1) = \lambda_i \sum_{j=1}^{n} w_{ij}\mathbf{x}_j(k) + (1-\lambda_i)\mathbf{x}_i(0). \tag{10}$$

The matrix $\mathbf{W}$ is stochastic and has the same meaning as in DeGroot's model, namely, w_{ij} represents the influence weight individual i accords to individual j. Without loss of generality, it can be assumed that $\lambda_i = 0$ for the agents with the maximal self-weights $w_{ii} = 1$, as both conditions imply the full stubbornness in the sense of (8) (see **"Simple properties of the FJ model"**). As discussed in [45], individuals' anchorage at their initial opinions can be explained by an ongoing effect of some exogenous factors that previously influenced the social group in the past. An initial opinion can also be considered as an individual's *prejudice* [50], [56], [166] that influences their opinion on subsequent steps.

Simple properties of the Friedkin- Johnsen model

Using induction on $k = 0, 1, \ldots$, a number of properties of the Friedkin- Johnsen (FJ) model can be proven.

1) *(self-weight and stubbornness)* An agent with maximal self-weight $w_{ii} = 1$ is stubborn independent of the susceptibility value, that is, $\mathbf{x}_i(k) = \mathbf{x}_i(0)$. For this reason, it is convenient to assume that $\lambda_i = 0$ whenever $w_{ii} = 1$.

2) *(consensus preservation)* If the initial opinions are in consensus $\mathbf{x}_1(0) = \ldots = \mathbf{x}_n(0) = \mathbf{x}_0^*$, this consensus is not deteriorated:

$$\mathbf{x}_1(k) = \ldots = \mathbf{x}_n(k) = \mathbf{x}_0^* \quad \forall k;$$

3) *(containment property)* More generally, at each stage of the opinion iteration, the opinions are contained by the *convex hull* of their initial values, that is, $\mathbf{x}_i(k) \in \mathfrak{X}_0$, where

$$\mathfrak{X}_0 = \left\{ \sum_{i=1}^{n} a_i \mathbf{x}_i(0) : a_i \geq 0, \sum_{i=1}^{n} a_i = 1 \right\}.$$

Whereas the containment property is very intuitive in the case of scalar opinions (where the set $\mathfrak{X}_0$ is the interval $[\min_i x_i(0), \max_i x_i(0)]$), its validity in higher dimensions is a nontrivial property of a social influence network, predicted by the FJ theory. Even for three-dimensional opinions, it is difficult to visualize the convex hull $\mathfrak{X}_0$ (being a convex polyhedron) without special software. Nevertheless, experiments on rational decision making on resource allocation [127] illustrate that multidimensional decisions of individuals typically stay in the convex polyhedron $\mathfrak{X}_0$.

Similar to DeGroot's model, the opinions $\mathbf{x}_i(k)$ may be scalar or multidimensional. Stacking them on top of each other to obtain the opinion matrix $\mathbf{X}(k)$, and the FJ system (10) can be rewritten in the matrix form as

$$\mathbf{X}(k+1) = \mathbf{\Lambda} \mathbf{W} \mathbf{X}(k) + (\mathbf{I}_n - \mathbf{\Lambda}) \mathbf{X}(0). \tag{11}$$

Here, $\mathbf{\Lambda} = \mathrm{diag}(\lambda_1, \ldots, \lambda_n)$ represents the diagonal matrix composed of the susceptibility coefficients. DeGroot's model arises as a special case of (11) with $\mathbf{\Lambda} = \mathbf{I}_n$.

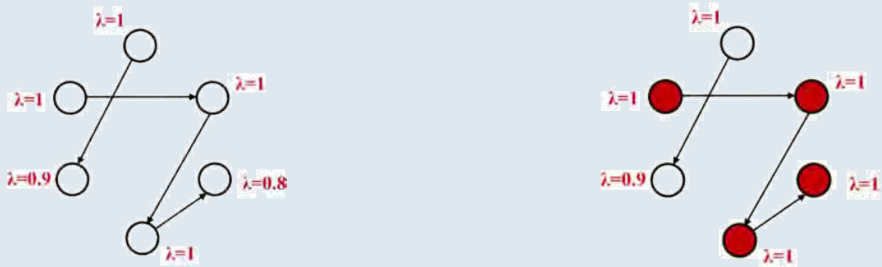
Schur stability criteria

Consider the graph of social influence $\mathcal{G}[\mathbf{W}]$ associated with the matrix $\mathbf{W}$ and let $\mathcal{S} \subseteq \{1, \ldots, n\}$ represent the set of individuals that are fully or partially stubborn (anchored at their initial opinions):

$$\mathcal{S} = \{i : \lambda_i < 1\}.$$

As discussed in [45], an individual's attachment to their initial opinion may be explained as a direct ongoing effect of some previous experience or other external factors that previously influenced the group. The FJ model is Schur stable if and only if the opinions of the remaining individuals (with $\lambda_i = 1$) also remain influenced by these factors via paths of influence (that is, any node $l \in \mathcal{V} \setminus \mathcal{S}$ is connected by a walk to a node from $\mathcal{S}$).

Theorem. [50], [56] The matrix $\mathbf{\Lambda W}$ is Schur stable if and only if every node of $\mathcal{G}[\mathbf{W}]$ either belongs to $\mathcal{S}$ or is connected to a node from $\mathcal{S}$ by a walk. This holds if $\mathcal{S} \neq \emptyset$ and $G[\mathbf{W}]$ is a strongly connected graph.



The social group shown on the left corresponds to a Schur stable matrix $\mathbf{\Lambda W}$ (each node with $\lambda = 1$ is connected to one of the nodes with $\lambda < 1$). For the group shown on the right, the matrix $\mathbf{\Lambda W}$ is not Schur stable: the group of red nodes is not connected to the unique node with $\lambda < 1$.

Due to presence of fully or partially stubborn agents, the FJ dynamics usually do not lead to consensus of opinions (except for special situations, where the FJ system reduces to DeGroot's model). In generic situations, the opinions converge. The most interesting case where such a convergence can be established is where the matrix $\mathbf{\Lambda W}$ is *Schur stable*, that is, all its eigenvalues $\mu_1, \ldots, \mu_n$ belong to the open unit disk $|\mu_j| < 1$. A graph-theoretical criterion of Schur [15], [50], [56] is summarized in **"Schur stability criteria."** If $\mathbf{\Lambda W}$ is a Schur stable matrix, then the matrix of opinions converges [50], [56]:

$$\begin{aligned} \mathbf{X}(\infty) &= \lim_{k\to\infty} \mathbf{X}(k) = \mathbf{V}\mathbf{X}(0), \\ \mathbf{V} &= (\mathbf{I}_n - \mathbf{\Lambda W})^{-1}(\mathbf{I}_n - \mathbf{\Lambda}). \end{aligned} \tag{12}$$

The matrix $\mathbf{V} = [v_{ij}]$ appears to be *row-stochastic* [14], [15] and is referred to as the *control matrix,* as it determines the ability of individuals to control the final opinion of others (see **"Control matrix and Friedkin's centrality"**).

Control matrix and influence centrality

When agents' opinions converge, the final opinion of agent i can be represented as

$$\mathbf{x}_i(\infty) = \sum_{j=1}^{n} v_{ij}\mathbf{x}_j(0).$$

In this sense, the entry v_{ij} serves as a measure of *social power* [129]) of individual j over individual i, that is, j's ability to influence i's terminal opinion. The *average* power of individual j over the group

$$c_j = \frac{1}{n}\sum_{i=1}^{n} v_{ij}$$

serves as a natural *measure of centrality* for the nodes of the social network. Choosing different matrices $\mathbf{\Lambda}$, a whole family of centrality measures is obtained for the weighted graph $G[\mathbf{W}]$ that were first introduced by Friedkin [78] for the case where $\mathbf{\Lambda} = \alpha\mathbf{I}_n$ with a scalar $\alpha \in (0,1)$. In this situation,

$$\mathbf{V} = (1-\alpha)(I - \alpha\mathbf{W})^{-1},$$

and the vector of influence centralities $\mathbf{c} = (c_1, \dots, c_n)^\top$ can be found as

$$\mathbf{c} = \frac{1}{n}\mathbf{V}^\top \mathbb{1}_n = (1-\alpha)\left(I - \alpha \mathbf{W}^\top\right)^{-1} \mathbb{1}_n.$$

For a specially chosen matrix [167] $\mathbf{W}$ and $\alpha = 1 - m$ [where $m \in (0,1)$], the latter vector coincides with the PageRank centrality measure, which appeared in [78] seven years earlier than the seminal work by Brin and Page [74]. Relations between the influence centrality and PageRank are discussed in more detail in [15], [168], [169].

Dynamics of reflected appraisal

The concept of influence centrality (see **"Control matrix and Friedkin's centrality"**) serves as a basis for the dynamical models describing the evolution of influence matrix $\mathbf{W}$ and is known as dynamics of *reflected appraisals*. As argued in [94], in deliberative groups (such as standing policy bodies and committees, boards of directors, juries and panels of judges), *an individual's influence centrality on an issue alters his or her expectation of future group-specific influence on issues*. In other words, the influence matrix may evolve as the social group discusses a sequence of different issues, see **"Reflected appraisal model."** Notice that models of appraisal dynamics do not portray the evolution of opinions and operate with quantities that are hard to measure experimentally, such as the centrality vectors. This, along with highly nonlinear dynamics of the centrality vectors, makes such models extremely difficult for identification. The problem of network reconstruction from appraisal dynamics is beyond the scope of this survey.

Extensions of the FJ model

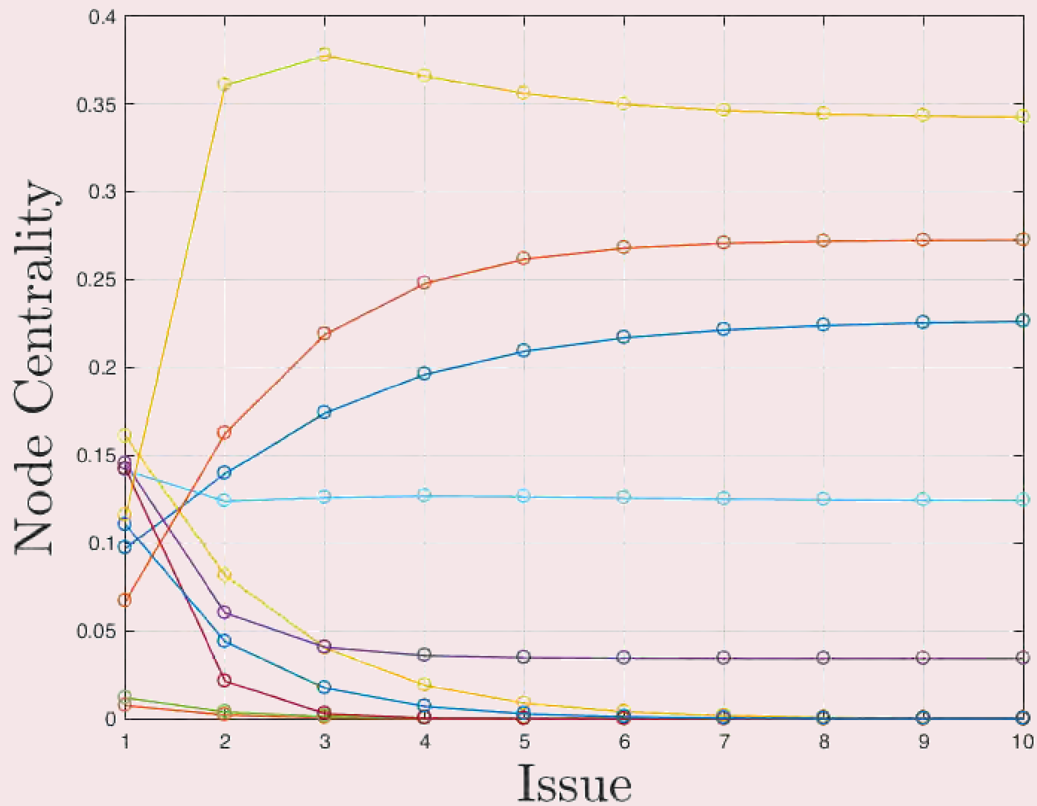
The seminal FJ model can be extended in many directions, among which only three are considered. The first extension is concerned with the dynamics of multidimensional opinions, which represent the agents' positions on several logically related issue. Such an opinion may be considered a special case of a *belief system*, defined as "a configuration of ideas and attitudes in which the elements are bound together by some form of constraint or functional interdependence" [170]. Contradictions and other inconsistencies between beliefs, attitudes, and ideas may trigger tensions and discomfort ("cognitive dissonance") that can be resolved by a within-individual process. This process, studied in cognitive dissonance and cognitive consistency theory, is thought to be an automatic process of the brain, enabling humans to develop coherent systems of beliefs [120], [171]. Modeling the dynamics of opinions on interrelated issues is a challenging problem, and only a few models have been proposed in the literature (most of them are featured by nonlinear dynamics [172]–[174]). The box **"A model of a belief system's dynamics"** is devoted to a simple linear model proposed in [56], [175], [176]. In general, the presence of the logical relations between the issues can affect the recoverability of the influence network from partial observations [53]. Another extension of the FJ model revises the restrictive assumption on simultaneous communication. As stated in [45], *the interpersonal influences do not occur simultaneously* and the assumption of synchronous rounds of interactions is too simplistic. In other words, individuals in real social groups are featured by asynchronous *ad-hoc* interactions. More realistic models that assumes only a couple of individuals can interact at each step were introduced in [50], [51], [56]. Such multiagent communication protocols are known as *gossiping* [177]. The model is summarized in **"Asynchronous gossip-based FJ model."** In [56], a gossip-based version of the extended FJ model (16) is considered. Another potential "culprit" of randomness is a noise, representing the effects of individuals' free will and unpredictability of their decisions (one model with noise is discussed in **"Dynamics on multiplex networks"**).

Reflected appraisal model

In psychology, the theory of reflected appraisal states that people's perceptions are influenced by the evaluation of others [96]. In [94], the evolution of power across a series of issues over time is explained as the result of direct and indirect interpersonal influences on group members. More formally, the phenomenon is described by the following dynamical system:

$$\begin{aligned}\mathbf{W}^{(s)} &= \mathbf{I} - \mathbf{\Lambda}^{(s)} + \mathbf{\Lambda}^{(s)}\mathbf{C}\\ (\mathbf{c}^s)^\top &= \frac{\mathbb{1}^\top}{n}(\mathbf{I} - \mathbf{\Lambda}^{(s-1)}\mathbf{W}^{(s-1)})^{-1}(\mathbf{I} - \mathbf{\Lambda}^{(s-1)}),\\ \mathbf{\Lambda}^{(s)} &= I - \mathrm{diag}(\mathbf{c}^{(s)}) = I - \mathrm{diag}(\mathbf{W}^{(s)}).\end{aligned} \tag{13}$$

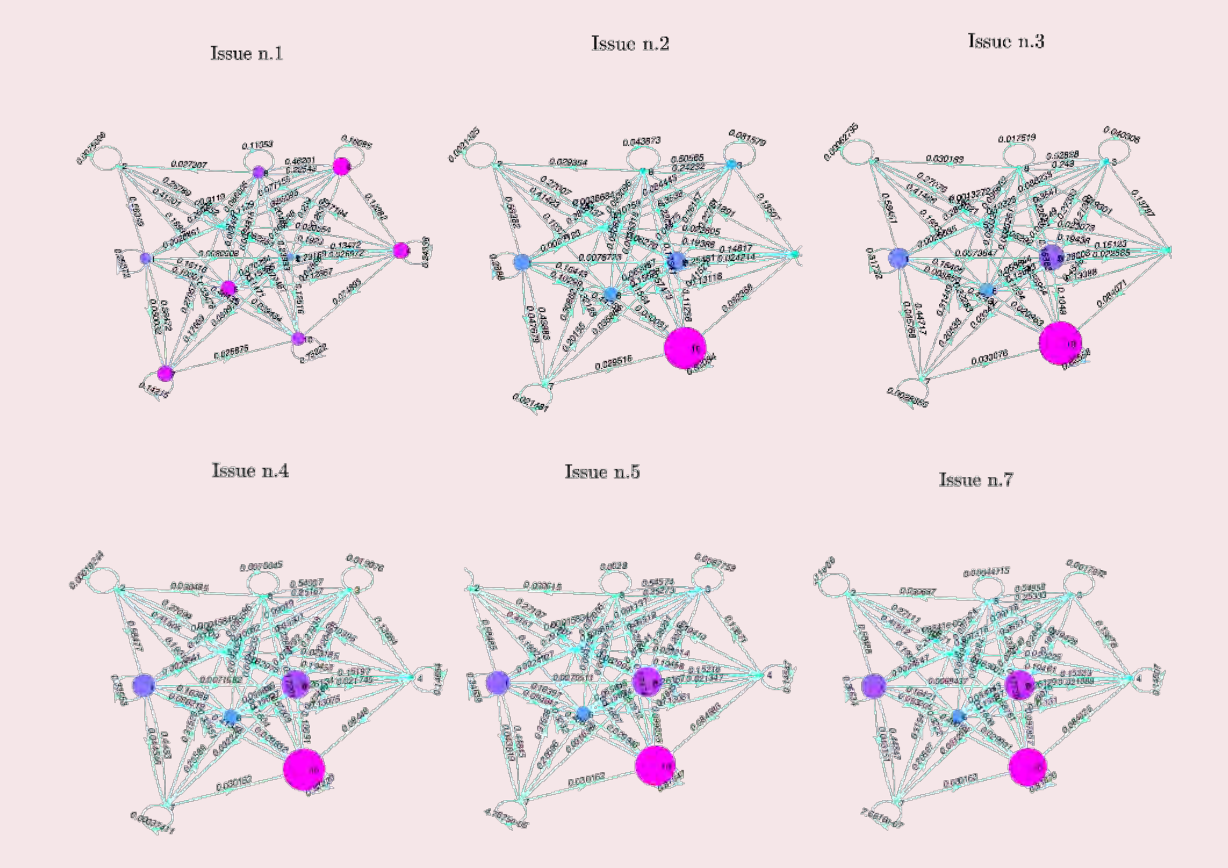
where $\mathbf{c}^s$ is influence centrality vector during the discussion on issue s and $\mathbf{C}$ is a constant matrix with zero diagonal entries. From control-theoretic viewpoint, this mechanism can be interpreted as a nonlinear feedback. The social power $c_i(s)$ individual has acquired in the discussion on issue $s-1$ influences their self-weight (and thus, also the weights assigned to the others) during the discussion on issue s. Notice that the structure of the graph remains unchanged (being encoded in the matrix $\mathbf{C}$), as the mechanism alters only the influence weights.



The evolution of the influence network and the social power is depicted as a function of issue sequence, respectively. It should be noted that the topology of the networks is the same at each layer. However, the strength of influence changes across issue sequence. This is captured by a model in which all transition matrices share a common support $\Omega \subseteq \{1, \ldots, n\} \times \{1, \ldots, n\}$, that is,

$$\mathbf{W}_{ij}^{(s)} \neq 0 \qquad \forall (i,j) \in \Omega, \forall \ell \in \{1, \ldots, m\}. \tag{14}$$

A simpler model of appraisal dynamics, which is now referred to as the DeGroot-Friedkin model [57], was proposed in [178]. Unlike (13), it is based on the DeGroot model and describes the evolution of social power vector $\mathbf{p}$ from (9) instead of the influence centrality vector $\mathbf{c}$.



A model of a belief system's dynamics

Adjusting their position on one of the interdependent issues, an individual might have to adjust positions on several related issues simultaneously to maintain the belief system's consistency. Such an adjustment can be thought of as an operator $\mathbf{x_i} \mapsto \mathcal{C}_i(\mathbf{x_i})$ that preserves the vector's dimension. Whereas the actual mathematical representation of introspective tension-resolving processes is unknown, it was conjectured in [176] that (in some situations) the operator $\mathcal{C}_i$ may be linear (and is represented by a matrix $\mathbf{C}_i$), so that $\mathbf{x}_i \mapsto \mathbf{x}_i\mathbf{C}_i^\top$ (recall that, according to our conventions, the multidimensional opinion is represented by a row m-dimensional vector, so that $\mathbf{C}_i \in \mathbb{R}^{m\times m}$). Assuming the tension-resolving process follows the integration of opinions from the neighbors, the FJ model (10) is replaced by the dynamics

$$\mathbf{x_i}(k+1) = \lambda_i \left(\sum_{j=1}^{n} w_{ij}\mathbf{x_j}(k)\right) \mathbf{C_i}^\top + (1-\lambda_i)\mathbf{x}_i(0). \tag{15}$$

It has been shown (see the supplementary material to [176]) that if the matrix $\mathbf{\Lambda W}$ is Schur stable and all matrices $\mathbf{C}_i$ are row-stochastic or (more generally) $\mathbf{C}_i = (c_{lm}^{(i)})$, where $\sum_m |c_{lm}^{(i)}| \leq 1$ for each

2

l and i, then the linear operator

$$\mathbf{X} \mapsto \mathbf{\Lambda W} \left(\begin{array}{c} \mathbf{x}_1 \mathbf{C_1}^\top \\ \hline \mathbf{x}_2 \mathbf{C_2}^\top \\ \hline \vdots \\ \hline \mathbf{x}_n \mathbf{C_n}^\top \end{array} \right)$$

is Schur stable. Specifically, the opinion matrix $\mathbf{X}(k)$ determined by (15) converges as $k \to \infty$. Equation (15) becomes more elegant in the case of homogeneous agents $\mathbf{C}_1 = \ldots = \mathbf{C}_n = \mathbf{C}$. In this situation, (15) may be rewritten in the matrix form very similar to (16):

$$\mathbf{X}(k+1) = \mathbf{\Lambda W X}(k)\mathbf{C}^\top + (\mathbf{I}_n - \mathbf{\Lambda})\mathbf{X}(0). \tag{16}$$

The FJ model is a special case of (16), corresponding to $\mathbf{C} = \mathbf{I}_n$ (if the issues are not logically related, it is natural to assume that the different dimensions of the opinion evolve independently). In [56], $\mathbf{C}$ is referred to as the multi-issue dependency structure(*MiDS*) matrix. An example of the system (16) with three-dimensional opinions has been considered in [176]. It was conjectured that the speech of Colin Powell, the highly respected US Secretary of State, in the UN Security Council presented a logic structure on three truth statements:

i) Saddam Hussein has a stockpile of weapons of mass destruction;
ii) Hussein's weapons of mass destruction are real and present dangers to the region and the world;
iii) An invasion of Iraq would be a just war.

It was a logic structure, as high certainty of belief on statement i) implies high certainty of belief on statements ii) and iii). On the other hand, if statement i) is false, then statements ii) and iii) are also false. This corresponds to the MiDS matrix

$$\mathbf{C} = \begin{pmatrix} 1 & 0 & 0 \\ 1 & 0 & 0 \\ 1 & 0 & 0 \end{pmatrix}.$$

A numerical example considered in [176] shows that if the population initially has a high certainty on statement i), then the belief system dynamics over a random graph generates a consensus that a preemptive invasion is a just war. At the same time, if statement i) is considered to be false, then the the population's certainty belief on all three statements is dramatically lowered. Hence, the model can explain the fluctuation of the public opinion on the Iraq invasion.

Asynchronous gossip-based FJ model

The FJ model [179] can be extended to the case where the interactions follow a model more consistent with the "usual" social network interactions where only a few agents interact at a time. In this case, the opinions evolve as follows [50]:

- each agent $i \in \mathcal{V}$ starts from an initial belief $x_i(0) \in \mathbb{R}$;
- at each period $k \in \mathbb{Z}_{\geq 0}$, a subset of *active* nodes $\mathcal{V}_k$ is randomly selected from a uniform

2

distribution over $\mathcal{V}$;

- the opinions of inactive agents remain unchanged, where each active agent $i \in \mathcal{V}_k$ interacts with a randomly chosen neighbor j and updates its belief according to a rule that resembles the FJ mechanism, which results in the equations

$$\begin{aligned} x_i(k+1) &= \lambda_i\big((1-w_{ij})x_i(k) + w_{ij}x_j(k)\big) + (1-\lambda_i)x_i(0) \qquad \forall i \in \mathcal{V}_k \\ x_\ell(k+1) &= x_\ell(k) \qquad \forall \ell \in \mathcal{V} \setminus \mathcal{V}_k, \end{aligned} \tag{17}$$

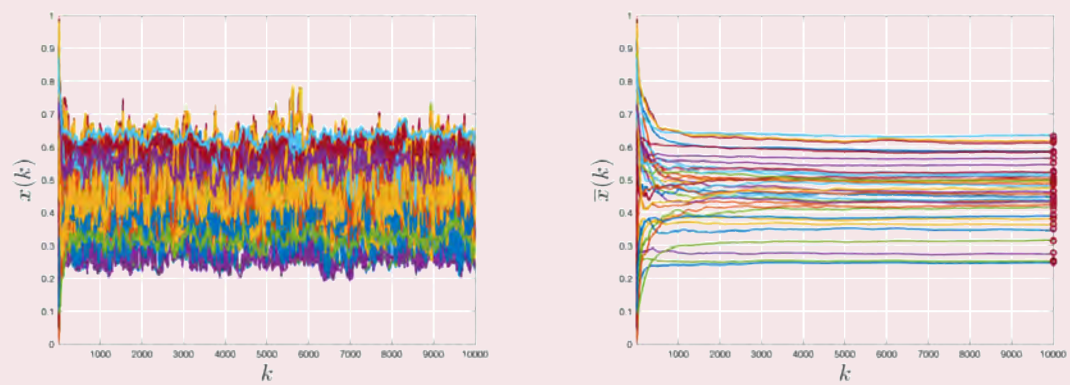
By denoting the set of neighbors of node $i \in \mathcal{V}$ by $\mathcal{N}_i \doteq \{j \in \mathcal{V} : (i,j) \in \mathcal{E}\}$, and introducing the out-degree $d_i \doteq |\mathcal{N}_i|$, the dynamics (17) can be formally rewritten in the following form: Given $\mathcal{V}_k$ and letting $\theta(k) \doteq \{\theta_i\}_{i \in \mathcal{V}_k}$,

$$\mathbf{x}(k+1) = \mathbf{\Gamma}(k)\mathbf{x}(k) + \mathbf{B}(k)\mathbf{x}(0), \tag{18}$$

where the coefficients are defined as

$$\mathbf{\Gamma}(k) \doteq \left(\mathbf{I}_n - \sum_{i \in \mathcal{V}_k} e_i e_i^\top (\mathbf{I}_n - \mathbf{\Lambda})\right)\left(\mathbf{I}_n + \sum_{i \in \mathcal{V}_k} \mathbf{W}_{i\theta_i}(\mathbf{e}_i \mathbf{e}_{\theta_i}^\top - \mathbf{e}_i \mathbf{e}_i^\top)\right), \qquad \mathbf{B}(k) \doteq \sum_{i \in \mathcal{V}_k} e_i e_i^\top (\mathbf{I}_n - \mathbf{\Lambda})$$

and θ_i is a uniformly distributed random element of $\mathcal{N}_i$ (that is, $\theta_i = j \in \mathcal{N}_i$ with probability $1/d_i$). It can be shown that the sequence $\{\mathbf{x}(k)\}_{k \in \mathbb{Z}_{\geq 0}}$ is a Markov process [180], which fails to converge in a deterministic sense and shows persistent oscillations.



However, if the matrix $\mathbf{\Lambda W}$ is Schur stable (see **"Schur stability criteria"**), the convergence of the *expectations* and the ergodicity of the oscillations can be ensured. Namely, it was shown in [52] that the opinions' expected values obey the equation

$$\mathbb{E}[\mathbf{x}(k+1)] = \overline{\mathbf{\Gamma}}\mathbb{E}[\mathbf{x}(k)] + \overline{\mathbf{b}}$$

where

$$\overline{\mathbf{\Gamma}} \doteq \mathbb{E}[\mathbf{\Gamma}(k)] = (1-\beta)\,\mathbf{I}_n + \beta\mathbf{\Lambda}(\mathbf{I}_n - \mathbf{D}^{-1}(\mathbf{I} - \mathbf{W})), \qquad \overline{\mathbf{b}} \doteq \beta(\mathbf{I}_n - \mathbf{\Lambda})\mathbf{x}(0), \tag{19}$$

$\beta = |\mathcal{V}_k|/|\mathcal{V}|$, and D is the degree matrix of the network (a diagonal matrix whose diagonal entry is equal to the degree $d_i = |\mathcal{N}_i|$). Moreover, the sequence $\mathbb{E}[\mathbf{x}(k)]$ converges to

$$\mathbb{E}[\mathbf{x}(\infty)] = (\mathbf{I}_n - \overline{\mathbf{\Gamma}})^{-1}\overline{\mathbf{b}}.$$

The opinion sequence has a few more interesting ergodicity properties that can be exploited in estimation algorithms. Namely, i) $\mathbf{x}(k)$ converges in distribution to a random variable $\mathbf{x}_\infty$, and the distribution is the unique invariant distribution for (17); ii) the process is ergodic, iii) the limit random variable satisfies $\mathbb{E}[\mathbf{x}_\infty] = (\mathbf{I}_n - \overline{\mathbf{\Gamma}})^{-1}\overline{\mathbf{b}}$, and iv) the Cesáro averages converge almost surely (and in the sense of p-th moment for each $p \geq 1$):

$$\bar{\mathbf{x}}(k) = \frac{1}{k+1}\sum_{\ell=0}^{k}\mathbf{x}(\ell) \xrightarrow[k\to\infty]{} \mathbf{x}_\infty$$

To identify influences in a social networks, the opinions cross-correlation matrix are useful, which are defined as

$$\mathbf{\Sigma}^{[\ell]}(k) \doteq \mathbb{E}\left[\mathbf{x}(k)\mathbf{x}(k+\ell)^\top\right].$$

As shown in [181], these correlation matrices satisfy

$$\mathbf{\Sigma}^{[\ell+1]}(k) = \mathbf{\Sigma}^{[\ell]}(k)\overline{\mathbf{\Gamma}}^\top + \mathbb{E}[\mathbf{x}(k)]\overline{\mathbf{b}}^\top. \tag{20}$$

Moreover $\mathbf{\Sigma}^{[\ell]}(k)$ converges (as $k \to \infty$) to the limit $\mathbf{\Sigma}^{[\ell]}(\infty)$ for all $\ell \in \mathbb{Z}_{\geq 0}$, satisfying

$$\mathbf{\Sigma}^{[\ell+1]}(\infty) = \mathbf{\Sigma}^{[\ell]}(\infty)\overline{\mathbf{\Gamma}}^\top + \mathbb{E}[\mathbf{x}(\infty)]\overline{\mathbf{b}}^\top. \tag{21}$$

Notice that the relation in (21) is a sort of Yule-Walker equation [182] used for estimation in autoregressive processes.

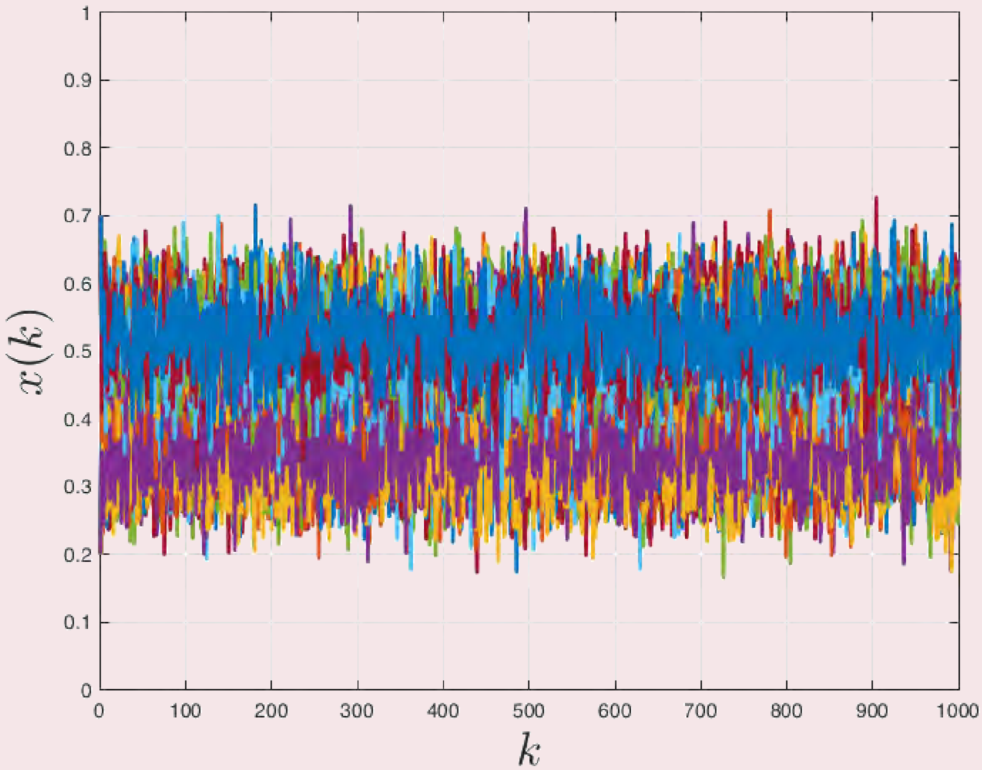
Dynamics on multiplex networks

The FJ model previously described can be extended to the cases where the social network discusses several issues and the influence network matrix is different depending on the topic. Since the underlying social network is essentially the same, it is expected that the social systems will share some common feature. In this sense, $\mathcal{M}_{cc}$ (the common component model) considers the cases where the networks differ in few components. $\mathcal{M}_{cs}$ (the common support model), instead describes situations where the topology is equal for all systems but the weights are different (see and **"Multidimensional networks"**). More precisely, consider the following set of dynamical equations:

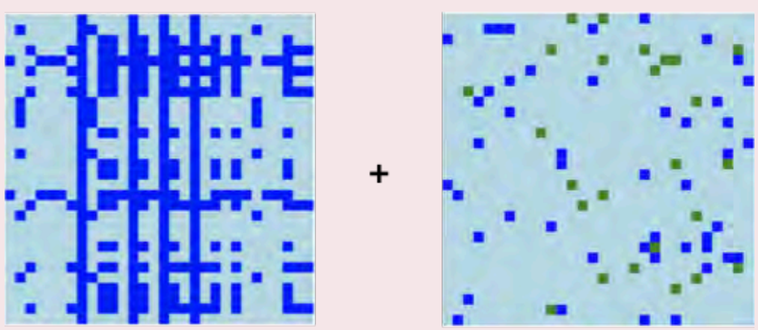
$$\begin{aligned}
\mathbf{x}^{(s)}(k+1) &= \mathbf{\Lambda}^{(s)}\mathbf{W}^{(s)}\mathbf{x}(k) + (\mathbf{I} - \mathbf{\Lambda}^{(s)})\mathbf{u}^{(s)} + \boldsymbol{\eta}^{(s)},\\
\mathbf{x}^{s}(0) &= \mathbf{u}^{(s)},
\end{aligned}$$

where $\mathbf{x}^{(s)}$ represent the agents' opinions on a specific subject s and $\boldsymbol{\eta}^{(s)}(t) \sim \mathcal{N}(\mathbf{0}, \mathbf{Q}_\eta)$ is an additive noise. The Markov process exhibit persistent fluctuations due to the random uncertainty in the dynamical system. It can also be shown that the expected opinions and cross-correlations matrices converge to a final pattern of values.

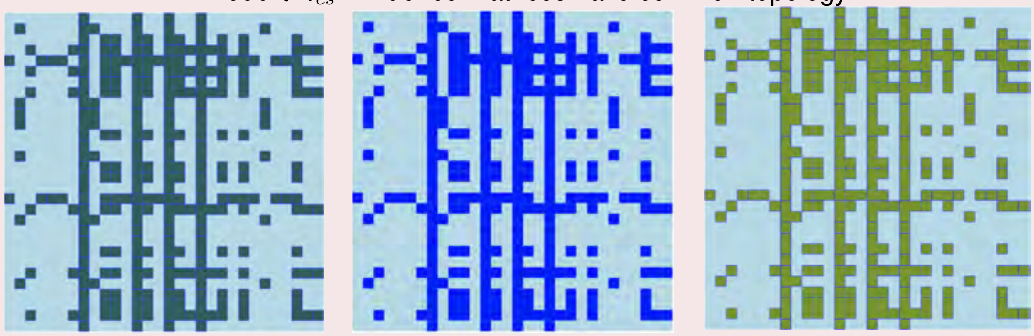
2



Model $\mathcal{M}_{cc}$: Influence matrices differ in few components.



Model $\mathcal{M}_{cs}$: Influence matrices have common topology.



However, if the matrices $\mathbf{\Lambda}^{(s)}\mathbf{W}^{(s)}$ are Schur stable (see **"Schur stability criteria"**), the convergence of the *expectations* and the ergodicity of the oscillations can be ensured. The sequence $\mathbb{E}[\mathbf{x}(k)]$ converges to

$$\mathbb{E}[\mathbf{x}^{(s)}(\infty)] = (\mathbf{I}_n - \mathbf{\Lambda}^{(s)}\mathbf{W}^{(s)})^{-1}(\mathbf{I}_n - \mathbf{\Lambda}^{(s)})\mathbf{u}^{(s)},$$

and the opinions' cross-correlation matrices satisfy the following relations (see [55]):

$$\mathbf{\Sigma}_{(s)}^{[\ell+1]}(k) = \mathbf{\Sigma}_{(s)}^{[\ell]}(k)(\overline{\mathbf{\Gamma}}^{(s)})^\top + \mathbb{E}[\mathbf{x}^{(s)}(k)](\overline{\mathbf{b}}^{(s)})^\top, \tag{22}$$

with $\overline{\mathbf{\Gamma}}^{(s)} = \mathbf{\Lambda}^{(s)}\mathbf{W}^{(s)}$ and $\overline{\mathbf{b}} = (\mathbf{I}_n - \mathbf{\Lambda}^{(s)})\mathbf{u}$ and

$$\begin{aligned}\mathbf{\Sigma}_{(s)}^{[0]}(\infty) = {} & \mathbf{\Lambda}^{(s)}\mathbf{W}^{(s)}\mathbf{\Sigma}_{(s)}^{[0]}(\infty)(\mathbf{W}^{(s)})^\top\mathbf{\Lambda}^{(s)} + \mathbf{\Lambda}^{(s)}\mathbf{W}^{(s)}\mathbb{E}[\mathbf{x}(\infty)]\mathbf{u}^\top(\mathbf{I}_n - \mathbf{\Lambda}^{(s)}) \\ & + (\mathbf{I}_n - \mathbf{\Lambda}^{(s)})\mathbf{u}\mathbb{E}[\mathbf{x}(\infty)]^\top(\mathbf{W}^{(s)})^\top\mathbf{\Lambda}^{(s)} + (\mathbf{I}_n - \mathbf{\Lambda}^{(s)})\mathbf{u}\mathbf{u}^\top(\mathbf{I}_n - \mathbf{\Lambda}^{(s)}) + \mathbf{Q}_\eta.\end{aligned} \tag{23}$$

Inference over networks: model-based approach

This section provides the classification of different approaches for inference over networks, while subsequent sections introduce two classes of them. The models presented in the previous sections have proven to be powerful tools for the analysis of interactions in social networks. However, in general, the general structure of the network is not available. Hence, the following question arises: *Given measurements of the evolution of the opinions and a model of the opinion evolution, how can one estimate the interaction graph and the strength of the connections?* With this in mind, in the second part of this article describes recent approaches in the literature to infer the social influences in a given a group of individuals whose opinions on m independent issues are supposed to evolve according to a prespecified model. This study is limited to the FJ model. The approaches described can be adapted to the DeGroot model and other models of opinion formation. The inference methods can be categorized according to the available information. Two main research areas are distinguished: The first one considers uncontrolled experiments, assuming that we can not intervene with the social system and only individual opinion updates can be tracked. This is a passive approach taken in previous works [183]. The latter considers controlled experiments, developing a social radar by exploiting the special role of stubborn agents. As shown in [117], the stubborn agents can help expose the network structure through a set of steady-state equations. This strategy generally assumes partial knowledge of the support of the social graph and considers an optimized placement of stubborn agents injecting inputs that affect the natural behavior of the opinion dynamics. This section reviews the first approach, while the reader is referred to [117], [184]–[186] for the inference networks estimation via controlled experiments. Consider two strategies to estimate the interactions in the network that are referred to as *persistent measurement* and *sporadic measurement* identification procedures. In the experiments of the first kind (persistent measurement), the opinions are observed during T rounds of conversation, and the influence matrix is estimated as the matrix best fitting the dynamics for $0 \leq k < T$. In such cases, the available results on parsimonious systems identification can be used to determine the unknown parameters [106], [187]. To exemplify how such an approach can be used in the context of network inference estimation, assume that the opinions evolve according to the multidimensional FJ model (11), which are recalled here for readability:

$$\mathbf{X}(k+1) = \mathbf{\Lambda}\mathbf{W}\mathbf{X}(k) + (\mathbf{I}_n - \mathbf{\Lambda})\mathbf{X}(0). \tag{24}$$

Assume that measurements of $\mathbf{X}(k)$ are available for $k = 0, 1, 2, \ldots, T$ This assumption can be relaxed by assuming that a sufficient number of measurement pairs, $(\mathbf{X}(k+1), \mathbf{X}(k))$, are available. To simplify the development to follow, first rewrite this model in a standard system identification form as

$$\mathbf{X}(k+1) = \mathbf{A}\mathbf{X}(k) + \mathbf{B}\mathbf{X}(0),$$

with $\mathbf{A} \doteq \mathbf{\Lambda}\mathbf{W}$, $\mathbf{B} = \mathrm{diag}(b_1, b_2, \ldots, b_n) \doteq \mathbf{I}_n - \mathbf{\Lambda}$. Denote by $\mathrm{off}(\mathbf{A}) \doteq \mathbf{A} - \mathrm{diag}(\mathbf{A})$ the matrix composed of the off-diagonal elements of $\mathbf{A}$ and having a zero diagonal. Then, given measurement error ϵ, the problem of estimating the sparsest interaction graph that is compatible with the measurements collected can be formulated as the following optimization problem:

$$\begin{aligned}
&\min_{\mathbf{A},\mathbf{B}} \; \|\mathrm{off}(\mathbf{A})\|_0 \\
&\text{subject to } \|\mathbf{X}(k+1) - \mathbf{A}\mathbf{X}(k) - \mathbf{B}\mathbf{X}(0)\|_\infty \leq \epsilon \\
&\qquad\qquad\qquad\qquad\qquad k = 0, 1, 2 \ldots, T-1 \\
&\qquad \mathbf{B} = \text{diag}(b_1, b_2, \ldots, b_n) \\
&\qquad \sum_{j=1}^{n} \mathbf{A}_{i,j} = 1 - b_i \text{ for } i = 1, 2, \ldots n \\
&\qquad \mathbf{A}_{i,j} \geq 0 \text{ and } 0 \leq b_i \leq 1 \text{ for } i, j = 1, 2, \ldots, n,
\end{aligned}$$

where $\|\mathbf{A}\|_0$ denotes the number of nonzero entries of the matrix $\mathbf{A}$. The optimization problem above is a nonconvex combinatorial problem, due to the presence of the zero-norm cost. To approximate the solution, a commonly used convex relaxation is to relax this norm to the ℓ_1-norm (see Compressed Sensing for additional details).

$$
\begin{aligned}
\min_{\mathbf{A},\mathbf{B}} \; & \|\mathrm{off}(\mathbf{A})\|_1 \\
\text{subject to } & \|\mathbf{X}(k+1) - \mathbf{A}\mathbf{X}(k) - \mathbf{B}\mathbf{X}(0)\|_\infty \leq \epsilon \\
& \qquad\qquad\qquad\qquad k = 0, 1, 2 \ldots, T-1 \\
& \mathbf{B} = \mathrm{diag}(b_1, b_2, \ldots, b_n) \\
& \sum_{j=1}^{n} A_{i,j} = 1 - b_i \text{ for } i = 1, 2, \ldots n \\
& A_{i,j} \geq 0 \text{ and } 0 \leq b_i \leq 1 \text{ for } i, j = 1, 2, \ldots, n
\end{aligned}
$$

Since $\|\mathrm{off}(\mathbf{A})\|_1 \doteq \sum_{i=1}^{n} \sum_{j \neq i} A_{ij}$, the latter problem can be decomposed into n independent problems. This is especially useful in very large networks. The drawback of these methods is that they require knowing the discrete-time indices for the observations made and storing a sufficiently long subsequence of opinions $\mathbf{x}(k),\ \mathbf{x}(k+1), .., \ \mathbf{x}(k+M-1)$. This knowledge may be difficult to obtain in general, and the collection may involve a large amount of data. The loss of data from one of the agents requires restarting the experiment. Moreover, the system could be updated with an unknown interaction rate, and the interaction timing between agents can be unobservable in practice [188], [189]. These considerations make the persistent measurement approach inapplicable in many practical situations, as also discussed in [117]. To circumvent these issues, this article describes two approaches that fall in the second class of methods; that is,, they only use sporadic data and (therefore) a complete history of agents' opinions is not required and the interactions are not limited to any prescribed number of rounds. In the first one, similar to the experiments from [117], the agents interact until their opinions stabilize and the identification problem considers only the initial and the final opinions. In the second one, it is only assumed that one has access to random measurements of the agents' opinions, and statistics of the measurement process are used to estimate the structure of the social network that generated the measurements. Figure 11 and Table 1 summarize the main differences and requirements of the reviewed methods.

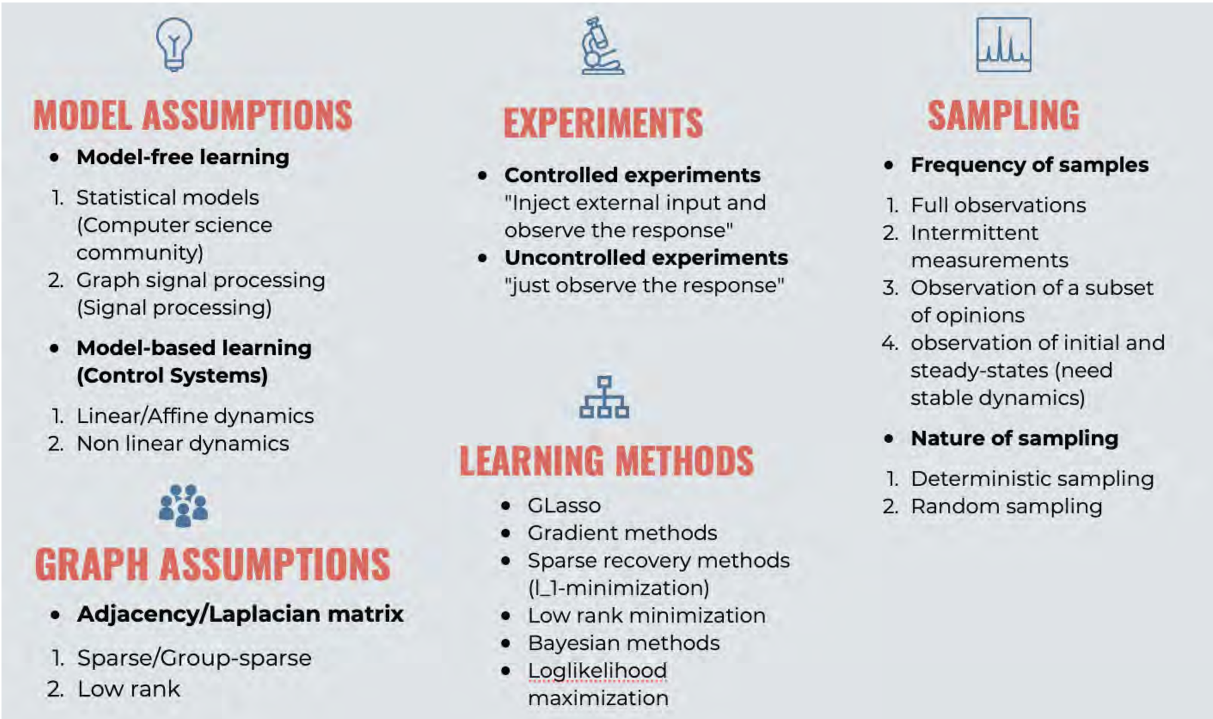


Figure 11: Learning graphs from data: main features.

The influence estimation problem: infinite horizon approach

As the first approach to the problem of estimating the structure of social network from infrequent data, consider the FJ model in (24) and assume there is knowledge of n the prejudices $\mathbf{X}(0)$ and final opinions $\mathbf{X}(\infty) = \lim_{k\to\infty} \mathbf{X}(k)$. The goal is to estimate $\mathbf{W}$ from this data only, under the assumption of network sparsity. To this end, some model identifiability considerations must be made. This is accomplished in the next subsection.

Model identifiability

First, note that due to the *consensus preservation* property discussed in **"Simple properties of the FJ model,"** whenever the initial opinions are at consensus, the final opinions are also at consensus. In this case, the problem is not well posed, since any stochastic matrix $\mathbf{W}$ will be consistent with the data. Motivated by this consideration, assume that for all $\ell = 1, \ldots, m$, there exists $i, j \in \mathcal{V}$ such that $x_i^{(\ell)}(0) \neq x_j^{(\ell)}(0)$. Similarly, when all agents are completely susceptible (that is, $\mathbf{\Lambda} = \mathbf{I}_n$), the FJ model reduces to DeGroot's model, typically leading to consensus of opinions. Clearly, the problem is not well posed in this case, since there are infinitely many matrices leading the dynamics in (24) to the same value of consensus. This fact is illustrated in the following example, borrowed from [53].

Example 1: Let $\mathbf{\Lambda} = \mathbf{I}_n$, and let $\mathbf{W} \in \mathbb{R}^{n\times n}$ be any doubly stochastic matrix that is irreducible (the graph $\mathcal{G}[\mathbf{W}]$ is strongly connected) and aperiodic. Then, the Perron Frobenius theorem [192] guarantees that

$$\mathbf{X}(\infty) = \mathbb{1}\mathbb{1}^\top \mathbf{X}(0)/n.$$

Similarly, if $\mathbf{\Lambda} = 0$, then $\mathbf{X}(\infty) = \mathbf{X}(0)$ and all stochastic matrices $\mathbf{W}$ are consistent with the data.

Note that an agent i with susceptibility $\lambda_i = 0$ is totally stubborn, that is, is not influenced by any other agent. Hence, to avoid ambiguities, the remainder of this article assume that $\lambda_i \neq 0$ for all $i \in \mathcal{V}$, $\mathbf{\Lambda} \neq \mathbf{I}_n$, and for any node $i \in \mathcal{V}$, there exists a path from i to a node j such that $\lambda_j < 1$ (each agent is influenced by at least one partially stubborn agent). With these assumptions, for any initial profile, the opinion dynamics leads asymptotically to an equilibrium point that can be computed from the weights, obstinacy levels, and initial opinions. It follows that recovering $\mathbf{W}$ amounts at solving the system of equations

$$: \begin{cases} (\mathbf{I}_n - \mathbf{\Lambda}\mathbf{W})\mathbf{x}^{(\ell)}(\infty) = (\mathbf{I}_n - \mathbf{\Lambda})\mathbf{x}^{(\ell)}(0), \\ \mathbf{W}\mathbb{1} = \mathbb{1}, \\ \mathbf{W} \geq 0, \mathbf{\Lambda} \geq 0 \end{cases} . \tag{25}$$

However, as shown in [53], this system contains an implicit ambiguity: If $(\mathbf{\Lambda}, \mathbf{W})$ is a solution of (25), then it is possible to construct a different solution pair $(\mathbf{\Lambda}', \mathbf{W}')$ as

$$\mathbf{\Lambda}' = \mathbf{I}_n - \mathbf{D}(\mathbf{I}_n - \mathbf{\Lambda})$$
$$\mathrm{off}(\mathbf{\Lambda}'\mathbf{W}') = \mathbf{D}\ \mathrm{off}(\mathbf{\Lambda}\mathbf{W})$$
$$\mathrm{diag}(\mathbf{\Lambda}'\mathbf{W}') = \mathbb{1} - \mathbf{D}((\mathbf{I}_n - \mathbf{\Lambda})\mathbb{1} + \mathrm{off}(\mathbf{\Lambda}\mathbf{W})\mathbb{1}),$$

for any nonnegative diagonal matrix $\mathbf{D}$ with $[\mathbf{D}]_{ii} \in [0, 1]$. The ambiguity described, which was already noted in [117] in the setting of DeGroot models with stubborn agents, is since the information about the rate of social interactions is missing (and it cannot be removed without making the additional assumption that the susceptibilities $\mathbf{\Lambda}$ are known). For $m \geq n$, if the system in (25) is full rank, then the problem in (25) admits a unique solution and may be easily solved (for example, using linear programming or any solver for convex optimization [193]). Following these considerations, assume that $\mathbf{\Lambda}$ is known and focus on the more interesting case when $m \ll n$. Note that if the matrix $\mathbf{\Lambda}$ is part of the learning, an invariant quantity must be defined among the ambiguous solutions [for instance by defining equivalence classes and resolve the ambiguity by imposing constraints on $\mathrm{diag}(\mathbf{W})$]. This is further discussed at the end of Section .

Sparse identification problem

Motivated by the discussion in Section , the identification approach is based on the observation that a social network is typically sparse, in the sense that the interactions among the agents are few when compared to the network dimension. For given $\mathbf{\Lambda}$, $\mathbf{X}(0)$, and $\mathbf{X}(\infty)$, this leads to estimating the social influence networks by solving a *sparsity problem*. Formally, determining the sparsest network that is compatible with the available information can be expressed as the following ℓ_0-minimization problem [194]:

$$\min_{\mathbf{W}\in\mathbb{R}^{n\times n}} \|\mathbf{W}\|_0, \qquad \text{s.t.} \begin{cases} \mathbf{\Phi}\mathbf{W}^\top = \mathbf{\Psi}^\top, \\ \mathbf{W}\mathbb{1} = \mathbb{1}, \\ \mathbf{W} \geq 0, \end{cases} \tag{26}$$

where $\|\mathbf{W}\|_0$ is the number of nonzeros of the matrix $\mathbf{W}$, $\mathbf{\Phi} \doteq \mathbf{X}(\infty)^\top$, $\mathbf{\Psi} \doteq \mathbf{\Lambda}^{-1}[\mathbf{X}(\infty) - (\mathbf{I}_n - \mathbf{\Lambda})\mathbf{X}(0)]$. This problem is decomposable into n subproblems, since each row of $\mathbf{W} = [\mathbf{w}_1^\top, \ldots, \mathbf{w}_n^\top]^\top$ can be learned independently from the others. More precisely,

$$\min_{w_j\in\mathbb{R}^{n}} \|\mathbf{w}_j\|_0, \qquad \text{s.t.} \begin{cases} \mathbf{\Phi}\mathbf{w}_j = \boldsymbol{\psi}_j, \\ \mathbb{1}^\top\mathbf{w}_j = 1, \\ \mathbf{w}_j \geq 0, \end{cases} \tag{27}$$

where $\boldsymbol{\psi}_j$ is the j-th row of $\mathbf{\Psi}$ for every $j \in [n]$. As discussed in **"Schur stability criteria,"** the reachability of each node from a partially stubborn node is an assumption that the true network must satisfy to guarantee the stability of the affine dynamics in (24) and the existence of the final opinion profile. This property is automatically ensured if $\mathbf{\Lambda} < \mathbf{I}$. If $\lambda_i < 1$ for at least one i, the stability property is *generic*: the set of matrices $\mathbf{W}$ such that $\mathbf{\Lambda}\mathbf{W}$ has eigenvalue at 1 has zero Lebesgue measure. In practice (taking into account inaccuracies in the opinion measurement), it is impossible to satisfy the first constraint in (25) unless the matrix $(\mathbf{I} - \mathbf{\Lambda}\mathbf{W})$ is invertible. However, as it should be noticed in the optimization problem (27), this constraint is not imposed in the recovery problem.

Compressed sensing

The optimization problems in (27) are a particular case of the so-called sparse recovery problem starting from compressed measurements [194], a problem also known as compressed sensing (CS). More precisely, sparse recovery problems are of the form

$$\min_{\mathbf{z}\in\mathbb{R}^{n}} \|\mathbf{z}\|_0$$

$$\text{s.t. } \mathbf{\Phi}\mathbf{z} = \boldsymbol{\psi},$$

where $\mathbf{\Phi} \in \mathbb{R}^{m \times n}$ with $m < n$, and $\|\mathbf{z}\|_0$ defines the ℓ_0 quasi-norm (which corresponds to the number of nonzero elements of $\mathbf{z}$). Note that the linear system of equations in the optimization problem above is underdetermined and admits infinitely many solutions. However, a sufficient condition for determining a solution can be derived by exploiting the sparsity of the desired solution and using the notion of *spark* of a matrix [48].

Spark of a matrix. The spark of a given matrix $\mathbf{\Phi}$, denoted with $\mathrm{spark}(\mathbf{\Phi})$, is the smallest number of columns from $\mathbf{\Phi}$ that are linearly dependent.

When dealing with sparse vectors, the spark concept provides a complete characterization of when sparse recovery is possible. The interested reader can refer to [48, Theorem 1.1] for a proof.

Proposition 1. For any vector $\boldsymbol{\psi}$, there exists at most one vector $\mathbf{z}$, such that $\boldsymbol{\psi} = \mathbf{\Phi}\mathbf{z}$ if and only if

$$\mathrm{spark}(\mathbf{\Phi}) > 2\|\mathbf{z}\|_0.$$

Computing the spark of a matrix involves checking the dependence of combinations of columns. Testing the condition in Proposition 1 is computationally expensive for practical purposes, as it requires a combinatorial search. Moreover the CS problem described above is known to be , in general, NP-hard. For this reason ℓ_1-based relaxations are often used to approximate the solution of CS problems. More precisely, this relaxation has the form

$$\min_{\mathbf{z} \in \mathbb{R}^n} \|\mathbf{z}\|_1$$
$$\text{s.t. } \mathbf{\Phi}\mathbf{z} = \boldsymbol{\psi}.$$

Much of the theory concerning explicit performance bounds for the relaxation described above relies on the concept of restricted isometry property (RIP). RIP characterizes matrices that are nearly orthonormal, at least when acting on sparse vectors [195].

Restricted Isometry Property. Let $\mathbf{\Phi} \in \mathbb{R}^{m \times n}$. Suppose that there exists a constant $\delta_s \in (0, 1)$, such that

$$(1 - \delta_s)\|\mathbf{z}\|_2^2 \leq \|\mathbf{\Phi}\mathbf{z}\|_2^2 \leq (1 + \delta_s)\|\mathbf{z}\|_2^2,$$

for all $\mathbf{z} \in \mathcal{Z}_s = \{\mathbf{z} \in \mathbb{R}^n : \|\mathbf{z}\|_0 \leq s\}$. Then, the matrix $\mathbf{\Phi}$ is said to satisfy the s-restricted isometry property with restricted isometry constant δ_s.

Denote with $\mathbf{\Phi}_S$ the matrix with columns indexed by $S \subseteq [n]$. It can be shown (see [195]) that, if a given matrix satisfies the RIP of order $2s$ with a constant $\delta_{2s} \in (0, 1/(\sqrt{2}+1))$, then one can uniquely recover a s-sparse vector using the ℓ_1 relaxation described above. It is straightforward to see that

$$(1 - \delta_s) \leq \mu_{\min}(\mathbf{\Phi}_S^\top \mathbf{\Phi}_S) \leq \mu_{\max}(\mathbf{\Phi}_S^\top \mathbf{\Phi}_S) \leq (1 + \delta_s),$$

where $\mu_{\min}$ and $\mu_{\max}$ denote the smallest and largest eigenvalue. Consequently, it must hold that

$$1 \approx \frac{\mu_{\max}(\mathbf{\Phi}_S^\top \mathbf{\Phi}_S)}{\mu_{\min}(\mathbf{\Phi}_S^\top \mathbf{\Phi}_S)} \leq 2,$$

for all $S \subseteq [n]$ with $|S| \leq s$.

It is well known [195] that sub-Gaussian random matrices with i.i.d. entries satisfy the RIP of order $2s$ with constant δ_{2s} (with probability close to 1) if

$$m \geq \frac{cs}{\delta_{2s}^2} \log\left(\frac{n}{\delta_{2s} s}\right)$$

, where $c > 0$ is a positive constant. Note that in the classical framework of CS, the sensing matrix is generally chosen by the user and is independent of the signal to be recovered.

2 **Recovery via convex optimization**

The convex relaxation of (27) (where the constraint $\mathbf{w}_j \geq 0$ is removed)

$$\min_{\mathbf{w}_j \in \mathbb{R}^n} \|\mathbf{w}_j\|_1, \qquad \text{s.t.} \begin{cases} \mathbf{\Phi}\mathbf{w}_j = \boldsymbol{\psi}_j, \\ \mathbb{1}^\top \mathbf{w}_j = 1, \end{cases} \tag{28}$$

can be formulated as a linear program and has been extensively studied (see **"Compressed Sensing"**). A large amount of algorithms have been proposed aiming at solving it efficiently, especially for the case where the dimension of the vector to be sparsified is high [48]. It is well known that under certain conditions on the matrix $\mathbf{\Phi}$, the number of measurements m, and the sparsity of $\mathbf{w}_j$, both (27) and (28) have the same unique solution [194]. However, in the case of influence estimation in SDNs considered in this article and (contrary to other problems in compressed sensing) the sensing matrix $\mathbf{\Phi}$ cannot be designed to satisfy the recovery properties mentioned above because it depends on the model's parameters,

$$\mathbf{\Phi} = \mathbf{X}(\infty)^\top = \mathbf{X}(0)^\top(\mathbf{I}_n - \mathbf{\Lambda})(\mathbf{I}_n - \mathbf{\Lambda}\mathbf{W})^{-1\top}.$$

In the case where the initial opinions are independent and identically distributed Gaussian random variables and the agents are all "very stubborn" ($\mathbf{\Lambda}$ has small diagonal entries), one could, in principle, consider $\mathbf{\Phi} \approx \mathbf{X}(0)^\top$ which would satisfy the RIP recovery condition with high probability. However, this is a very special case that does not cover most of the SDN recovery problems of interest. If $\mathbf{\Lambda}$ is not close to zero, then $\mathbf{\Phi}$ is a random variable whose entries are coupled and available results on compressed sensing do not apply. The remainder of this section reviews the results in [53], where recovery conditions specific to SDNs are derived. Assume that the initial opinions $\mathbf{x}^{(\ell)}(0)$ on topic ℓ are independent and identically distributed random variables having a Gaussian distribution with zero mean and unit variance. The hypothesis on the Gaussian distribution of the initial condition is a common assumption in opinion dynamics literature [196]. It can also be explained by the fact that initial opinions can be preaveraged opinions or several criteria that can be treated as independent random variables. Therefore, the distribution of initial opinions has a Gaussian-like shape due to the central limit theorem [197]. Moreover, the assumption on the zero mean and identity covariance matrix is not a restrictive one. Given any

Gaussian distribution of the initial opinions, one can always perform a linear transformation and obtain an equivalent problem that satisfies the assumptions. If $x^{(\ell)}(0)$ have a nonzero expected value, then consider $x^{(\ell)}(0)-\overline{x}^{(\ell)}(0)\mathbb{1}$ and $z^{(\ell)}(\infty)=\mathbf{V}(x^{(\ell)}(0)-\overline{x}^{(\ell)}(0)\mathbb{1})$, where $\mathbf{V}$ is the total effects matrix. Since the total effects matrix is stochastic, then $z^{(\ell)}(\infty)=x^{(\ell)}(\infty)-\overline{x}(0)^{(\ell)}\mathbb{1}$. Note that for SDNs influence estimation problem formulated in this section, the probability of violating the RIP condition can be very close to one (even for very simple graphs). More precisely, in [53] it is proven that for Gaussian initial conditions $x(0)^{\ell}\sim\mathcal{N}(0,I)$ for all $\ell\in[m]$ and for an arbitrary set $S\subseteq[n]$ of size $|S|=s$, for the defined $\widehat{\mathbf{\Sigma}}_{SS}=\mathbf{\Phi}_S^{\top}\mathbf{\Phi}_S/m$, with probability greater than $1-2\mathrm{e}^{-m/32}$ we have

$$\frac{\mu_{\max}(\widehat{\mathbf{\Sigma}}_{SS})}{\mu_{\min}(\widehat{\mathbf{\Sigma}}_{SS})}\geq\frac{1}{3}\frac{\mu_{\max}(\mathbf{\Sigma}_{SS})}{\mu_{\min}(\mathbf{\Sigma}_{SS})}$$

where $\mathbf{\Sigma}=(I-\mathbf{\Lambda}\mathbf{W})^{-1}(I-\mathbf{\Lambda})^{2}(I-\mathbf{\Lambda}\mathbf{W})^{-\top}$. For this reason, more powerful tools are needed for the performance analysis of the problem. More precisely, the so-called "nullspace property" [194] is needed, which provides a necessary and sufficient condition for recovery. This property is summarized in **"Necessary and Sufficient Conditions for Recovery."**

Necessary and sufficient conditions for recovery

The concept of null space property [194] is needed to derive more general conditions for sparse recovery.

Null Space Property (NSP). The matrix $\mathbf{\Phi}\in\mathbb{R}^{m\times n}$ satisfies the NSP of order s if, given

$$\mathcal{C}(\ell)=\{\eta\in\mathbb{R}^n:\|\eta_{S^c}\|_1\leq\|\eta_S\|_1\},$$

$$\mathcal{C}(\ell)\cap\mathsf{Ker}(\mathbf{\Phi})=\{0\},$$

for all index set S with $|S|\leq s$, where Ker denotes the kernel of a matrix.

Using this definition, Theorem 1 in [194] provides additional results on when one can recover sparse solutions from systems of linear equations. More precisely, consider matrix $\mathbf{\Phi}\in\mathbb{R}^{m\times n}$. Then, the optimization problem

$$\{\min\|\mathbf{z}\|_1:\mathbf{\Phi}\mathbf{z}=\boldsymbol{\psi}\}$$

uniquely recovers all s-sparse vectors $\mathbf{z}^{\star}$ from measurements

$$\boldsymbol{\psi}=\mathbf{\Phi}\mathbf{z}^{\star}$$

if and only if $\mathbf{\Phi}$ satisfies the nullspace property with order $2s$. To further analyze when can sparse solutions can be recovered, introduce the concept of restricted eigenvalue criterion.

Definition 1 (Restricted Eigenvalue Condition (REC)): A matrix $\mathbf{\Phi}$ satisfies the REC of order s if there exists a $\delta_s>0$, such that

$$\frac{1}{m}\|\mathbf{\Phi}\mathbf{z}\|_2^2\geq\delta_s^2\|\mathbf{z}\|_2^2$$

for all $\mathbf{z} \in \mathcal{C}(\ell)$, uniformly for all index sets $S \subseteq [n]$ with $|S| \leq s$.
It is straightforward to see that REC is equivalent to NSP. For random matrices $\mathbf{\Phi}$ with i.i.d. entries drawn from particular distributions or for unitary matrices, it can be shown that, if enough measurements are available, then REC condition is satisfied with the prescribed δ_s with probability close to 1 [198].

These necessary and sufficient conditions on sparse recovery enable one to study when it is possible to recover sparse models for the SDN. It is shown in [53] that the initial condition $\mathbf{x}^{(\ell)}(0) \sim \mathcal{N}(0, \mathbf{I}_n)$ and if the number of considered topics satisfies

$$m \geq 4c \frac{(1+\lambda_{\max})^2(1-\lambda_{\min})^2}{(1-\lambda_{\max})^4} d_{\max} \log n, \tag{29}$$

then the solution to (28) is unique and coincides with that of (27) with probability at least $1 - c' \mathrm{e}^{-c'' m}$, where c, c' and c'' are positive constants, $d_{\max} = \max_{v \in \mathcal{V}} |\mathcal{N}(v)|$, $\lambda_{\max} = \max_j \lambda_j$, and $\lambda_{\min} = \min_j \lambda_j$. Equation (29) shows that to recover a sparse influence model, the sensitivity to other opinions cannot be high. More precisely, if $\lambda_{\max} \to 1$, then the number of measurements needed for recovery diverges to infinity. This is reasonable, since the final opinions are a function of preconceived opinions and the network sensing performance should depend on the strength of the influencing power of the prejudices. Moreover (as conjectured in [117]), another important issue that affects the reconstruction performance is the degree distribution in the social network. More precisely, for a fixed total number of edges, it is easier to recover a network with a concentrated degree distribution (for example,, the Watts-Strogatz network [199]) while a network with power law degree distribution (for example,, the Barabasi-Albert network [200]) is more difficult to recover. To finalize the discussion in this section, recall that if $\mathbf{\Lambda}$ is not known, then the identification problem is not well posed. The ambiguity is due to the missing information about the rate of social interactions. This ambiguity cannot be removed without making additional assumptions. However, an invariant quantity can be determined among the ambiguous solutions by defining equivalence classes and resolve the ambiguity by imposing constraints on $\mathrm{diag}(\mathbf{W})$. More precisely, in [53], it is shown that the problem of learning sensitivity matrix $\mathbf{\Lambda}$ can be cast as in (28) with $\mathbf{\Phi} \doteq [\mathbf{X}(\infty)^\top, \mathbf{x}_j(0) - \mathbf{x}_j(\infty)]$ and $\mathbf{B} = \mathbf{X}(0)$, where $\mathbf{x}_j(\infty), \mathbf{x}_j(0)$ are the column vectors corresponding to j-th row of $\mathbf{X}(\infty)$ and $\mathbf{X}(0)$, respectively, with the additional constraint that $w_{jj} = 0$.

Influence estimation from random opinion measurements

While the infinite horizon approach is surely innovative under various aspects, it suffers the clear drawback of being static. Indeed, the identification does not exploit the dynamical nature of the system, and it requires knowledge of initial and final opinions on several different topics to build the necessary information to render the problem identifiable. Even if the number of topics that is necessary to correctly identify the network is strictly smaller than the size of the graph (and in many cases, scales logarithmically with it), this information may sometimes be hard to collect. This section reviews an alternative approach to the social network estimation problem that exploits the dynamical evolution of the opinions. At the same time, it does not require the observations of opinions on different topics or perfect knowledge of the interaction times, and it can be adapted to cases when some information is missing or partial. More precisely, the availability of "intermittent" measurements of the opinions is assumed to identify the dynamics of the evolution of the opinions and, as a consequence, the influence matrix [182], [201]. Such an approach is especially useful in the case where not all opinions are updated at the same time and random sampling of the opinions might be a less onerous way of estimating the behavior of the network. Hence, this section focuses on the asynchronous gossip-based FJ model and assumes that random measurements of opinions are available. For simplicity (as in [181]), consider the case when a single topic is discussed. However, the reasoning can be easily extended to cases involving multiple topics.

Observation models

Consider the gossip opinion dynamics in (17), where the influence matrix $\mathbf{W}$ is unknown. Assume that each time k does not have complete knowledge of the opinion vector $\mathbf{x}(k)$, only partial information is available. More precisely, assume the following random model for the observations:

$$\mathbf{z}(k) = \mathbf{P}(k)\mathbf{x}(k), \tag{30}$$

where the diagonal matrix $\mathbf{P}(k)$ is a random measurement matrix defined by

$$\mathbf{P}(k) = \mathrm{diag}(\mathbf{p}(k)),$$

and $\mathbf{p}(k) \in \{0,1\}^n$ is a random selection vector with known distribution representing which opinions are measured at time k. Different probability distributions of the matrix $\mathbf{P}(k)$ lead to very different observation models. For example, if

$$\mathbf{p}(k) = \begin{cases} \mathbb{1} & \text{w.p. } \rho \\ 0 & \text{otherwise,} \end{cases}$$

then the so-called *intermittent observation model* is present, where at $k \in \mathbb{Z}_{\geq 0}$, all observations are available with probability ρ or no observations at all are observed. This model allows for capturing the typical situation in which the actual rates at which the interactions occur is not perfectly known (thus, sampling time is different from interaction time). Moreover, if at each time $k \in \mathbb{Z}_{\geq 0}$ the selection vector is $p_i(k) \sim \mathrm{Ber}(\rho_i)$ for all $i \in \mathcal{V}$, then the so-called *independent* *random sampling model* [182], [201] is present, where the opinions are observed independently with probability $\rho_i \in [0,1]$. In the case where the observations are made with equal probability $\rho_i = \rho$ for all $i \in \mathcal{V}$, this model is referred to as *independent and homogeneous sampling*. If $\rho = 1$, then there are full observations; if $\rho \neq 1$, then there is partial information. This model has a clear interpretation for SDNs, describing the situation where only a subset of individuals can be contacted at each time k (for example, random interviews). This section reviews the approach described in [181], where the objective is as follows. Given the sequence of observation $\{\mathbf{z}(k)\}_{k=1}^{t}$ estimate of the matrix $\mathbf{W}$ is referred to as $\widehat{\mathbf{W}}_t$. In [181], theoretical conditions are also provided on the number of samples that are sufficient to have an error not larger than a fixed tolerance ϵ with high probability. For clarity of exposition, these theoretical results are not reviewed here.

Overview of the proposed approach to influence estimation

The main stream of the methodology is summarized in Figure 12. To reconstruct the influence matrix, recall the definition of then opinions' cross-correlation matrix:

$$\mathbf{\Sigma}^{[\ell]}(k) := \mathbb{E}\left[\mathbf{x}(k)\mathbf{x}(k+\ell)^\top\right].$$

It has been shown that the evolution of the covariance matrix $\mathbf{\Sigma}^{[\ell]}(k)$ is described by

$$\mathbf{\Sigma}^{[\ell+1]}(k) = \mathbf{\Sigma}^{[\ell]}(k)\overline{\mathbf{\Gamma}}^\top + \mathbb{E}[\mathbf{x}(k)]\overline{\boldsymbol{b}}^\top, \tag{31}$$

where $\overline{\mathbf{\Gamma}} = \overline{\mathbf{\Gamma}}(\mathbf{\Lambda}, \mathbf{W})$ and $\overline{\boldsymbol{b}} = \overline{\boldsymbol{b}}(\mathbf{\Lambda}, \mathbf{x}(0))$ are defined in (19). Moreover, $\mathbf{\Sigma}^{[\ell]}(k)$ converges to $\mathbf{\Sigma}^{[\ell]}(\infty)$ for all nonnegative integer ℓ that satisfy

$$\mathbf{\Sigma}^{[\ell+1]}(\infty) = \mathbf{\Sigma}^{[\ell]}(\infty)\overline{\mathbf{\Gamma}}^\top + \mathbb{E}[\mathbf{x}(\infty)]\overline{\boldsymbol{b}}^\top, \tag{32}$$

where $\mathbf{\Sigma}^{[\ell]}(\infty) := \lim_{k\to\infty} \mathbf{\Sigma}^{[\ell]}(k)$. The simple linear equation above provides the motivation for the approach described in this section. This approach can be summarized as follows: First, from the partial random measurements $\{\mathbf{z}(k)\}_{k=1}^{t}$, estimate the expected terminal state $\mathbb{E}[\mathbf{x}(\infty)]$ and the terminal covariance matrices $\mathbf{\Sigma}^{[0]}(\infty)$ through $\mathbf{\Sigma}^{[\ell]}(\infty)$ for some ℓ. Given these estimates and using (32), estimate the matrix $\overline{\mathbf{\Gamma}}$. Finally, to estimate the influence matrix $\mathbf{W}$ by exploiting the relation between $\overline{\mathbf{\Gamma}}$ and $\mathbf{W}$.

Estimating the expected opinion profile and the cross-correlation matrices

It is now shown how to exploit the model of the observations and the data collected to estimate the opinion's expectation an covariance. To estimate the expected opinion profile $\mathbb{E}[\mathbf{x}(\infty)]$, start with time averages of the observations $\mathbf{z}(k)$. It can be shown that

$$\mathbb{E}[\mathbf{z}(k)] = \pi \circ \mathbb{E}[\mathbf{x}(k)],$$

where $\pi = \mathbb{E}[\mathbf{p}(k)]$ and $\circ$ denotes the entrywise product. This allows for estimating the expectation of the opinions from available data. More precisely, begin by estimating $\mathbb{E}[\mathbf{z}(k)]$ using time averages

$$\overline{\mathbf{z}}(t) = \frac{1}{t}\sum_{k=1}^{t} \mathbf{z}(k),$$

and obtain

$$\widehat{x}_i(t) = \frac{\overline{z}_i(t)}{\pi_i}. \tag{33}$$

Estimating covariance matrices can be done in a similar way. The cross-correlation matrices $\mathbf{\Sigma}^{[\ell]}(\infty)$ are estimated from the empirical covariance matrix of the observations $\mathbf{z}(k)$. Denote

$$\mathbf{S}^{[\ell]}(k) := \mathbb{E}[\mathbf{z}(k)\mathbf{z}(k+\ell)^\top].$$

Then,

$$\mathbf{S}^{[\ell]}(k) = \mathbf{\Pi}^{[\ell]}(k) \circ \mathbf{\Sigma}^{[\ell]}(k)$$

, where $\mathbf{\Pi}^{[\ell]} = \mathbb{E}[\mathbf{p}(k)\mathbf{p}(k+\ell)^\top]$ and $\circ$ denotes the Hadamard product. Since $\mathbf{S}^{[\ell]}(k)$ is unknown, estimate $\mathbf{S}^{[\ell]}(k)$ using time averages

$$\widehat{\mathbf{S}}^{[\ell]}(t) = \frac{1}{t-\ell}\sum_{k=1}^{t-\ell} \mathbf{z}(k)\mathbf{z}(k+\ell)^\top,$$

from which

$$\widehat{\mathbf{\Sigma}}_{ij}^{[\ell]}(t) = \widehat{\mathbf{S}}_{ij}^{[\ell]}(t)/\mathbf{\Pi}_{ij}^{[\ell]}. \tag{34}$$

Although these seem rather ad-hoc estimates of the needed quantities, it can be shown that they converge to the desired values as the number of measurements tend to infinity. More precisely, in [181], a careful analysis of the procedures developed above shows that the estimates converge to the true values at a rate of $O(1/\sqrt{t})$, where t is the number of measurements used.

As an example of the estimation procedure above, consider first the case of *independent homogeneous random sampling*. In this case, $\pi = \rho$,

$$\begin{aligned}
\mathbf{\Pi}^{[0]} &= \mathbb{P}(i,j \in \mathcal{V}_k) = \rho\mathbf{I}_n + \rho^2(\mathbb{1}\mathbb{1}^\top - \mathbf{I}_n)\\
\mathbf{\Pi}^{[\ell]} &= \mathbb{P}(i \in \mathcal{V}_k, j \in \mathcal{V}_{k+\ell}) = \rho^2\mathbb{1}\mathbb{1}^\top \quad \text{if } \ell \neq 0,
\end{aligned}$$

from which $\widehat{x}(t) = \overline{z}(t)/\rho$ and

$$\widehat{\mathbf{\Sigma}}_{ij}^{[\ell]}(t) = \frac{1}{\rho^2}\widehat{\mathbf{S}}^{[\ell]}(t) - \left(\frac{1-\rho}{\rho^2}\widehat{\mathbf{S}}^{[\ell]}(t) \circ \mathbf{I}_n\right)\mathbf{1}(\ell = 0). \tag{35}$$

As a second example, consider the case of *intermittent observations*. In this case, $\pi = \rho$,

$$\mathbf{\Pi}^{[0]} = \rho\mathbb{1}\mathbb{1}^\top \quad \text{and} \quad \mathbf{\Pi}^{[\ell]} = \rho^2\mathbb{1}\mathbb{1}^\top \quad \text{if } \ell \neq 0,$$

from which $\widehat{\mathbf{x}}(t) = \overline{\mathbf{z}}(t)/\rho$ and $\widehat{\mathbf{\Sigma}}_{ij}^{[\ell]}(t) = \widehat{\mathbf{S}}^{[\ell]}(t)/\rho^2$.

Estimating the influence matrix

In principle, the estimators $\widehat{\boldsymbol{\Sigma}}^{[1]}(t)$ and $\widehat{\boldsymbol{\Sigma}}^{[0]}(t)$ of $\boldsymbol{\Sigma}^{[1]}(t)$ and $\boldsymbol{\Sigma}^{[0]}(t)$, together with (32), can be used to estimate the dynamics matrix $\overline{\boldsymbol{\Gamma}}$. However, there is a significant obstacle to address when using such a "naive" approach. Given the fact that one has random observations, it is likely that the procedure described above produces "poor" estimates of $\boldsymbol{\Sigma}^{[1]}(t)$ since, for several k, many of the entries of $\mathbf{z}(k)\mathbf{z}(k+\ell)^\top$ might be zero. To circumvent this, start by choosing a number N_Σ of covariance matrices that are to be considered in the estimation of dynamics and use a combination of these covariance matrices. More precisely, given estimates $\widehat{\boldsymbol{\Sigma}}^{[\ell]}(t)$, compute

$$\widehat{\boldsymbol{\Sigma}}_-(t) \doteq \frac{1}{N_{\boldsymbol{\Sigma}}} \sum_{\ell=0}^{N_\Sigma - 1} \widehat{\boldsymbol{\Sigma}}^{[\ell]}(t); \qquad \widehat{\boldsymbol{\Sigma}}_+(t) \doteq \frac{1}{N_{\boldsymbol{\Sigma}}} \sum_{\ell=1}^{N_\Sigma} \widehat{\boldsymbol{\Sigma}}^{[\ell]}(t), \tag{36}$$

and note that these matrices (approximately) satisfy (32). Hence, they can be used to estimate the structure of the network. Consider two types of networks. First, assume that one knows in advance that the network is dense. In this case, a possible estimator of $\overline{\boldsymbol{\Gamma}}$ can be obtained by directly solving the set of linear equations (32). In other words, the estimator is

$$\widehat{\widehat{\overline{\boldsymbol{\Gamma}}}}(t)^\top = \widehat{\boldsymbol{\Sigma}}_-(t)^\dagger (\widehat{\boldsymbol{\Sigma}}_+(t) - \overline{\mathbf{x}}(t)\overline{\boldsymbol{b}}^\top). \tag{37}$$

In the case of networks that are known to be sparse, one can solve a sparsity inducing optimization problem aimed at finding the sparsest graph that is compatible with available information. The estimator can be obtained by solving

$$\begin{aligned} \widehat{\widehat{\overline{\boldsymbol{\Gamma}}}}(t)^\top &= \underset{\mathbf{M} \in \mathbb{R}^{\mathcal{V}\times\mathcal{V}}}{\operatorname{argmin}} \sum_{i,j,i\neq j} |\mathbf{M}_{ij}| \\ \text{s.t.} \quad & \|\widehat{\boldsymbol{\Sigma}}_-(t)\mathbf{M} - (\widehat{\boldsymbol{\Sigma}}_+(t) - \overline{\mathbf{x}}(t)\overline{\boldsymbol{b}}^\top)]\|_{\max} \leq \eta. \end{aligned}$$

Performance of influence estimation: asynchronous gossip-based FJ model

The estimation error on matrix $\overline{\boldsymbol{\Gamma}}(t)$ is based on the previous estimation of the cross-correlation matrices. Using (37) $\hat{\Sigma}_+$ must be inverted, and the estimation error depends on the singular values of $\hat{\Sigma}_\pm$. It can be shown that, with probability at least $1-\delta$,

$$\begin{aligned} &\|\overline{\boldsymbol{\Gamma}}(t) - \widehat{\widehat{\overline{\boldsymbol{\Gamma}}}}(t)\|_2 \\ &= O\left(\frac{n(\sigma_{\max}^+ + n)}{(\underline{\sigma}_{\min}^-)^2 \Pi^\star \sqrt{\delta(t+1)\beta(1-\lambda_{\max})}} \right), \end{aligned} \tag{38}$$

where $\sigma_{\max}^+ = \|\Sigma_+\|_2$ and $\underline{\sigma}_{\min}^- \doteq \min(\sigma_{\min}^-, \widehat{\sigma}_{\min}^-)$ (where $\sigma_{\min}^-$, $\widehat{\sigma}_{\min}^-$ are the minimum singular value of Σ_- and $\widehat{\Sigma}_-$, respectively).

Estimating the network topology and the influence matrix

Once an estimate of the average transition matrix $\overline{\mathbf{\Gamma}}(t)$ has been obtained, the topology of the influence network can be retrieved in a straightforward manner by noticing that $\mathrm{supp}(\overline{\mathbf{\Gamma}}) = \mathrm{supp}(\mathbf{W})$. Hence, we can reconstruct the support of $\mathbf{W}$ using the elements of the estimated matrix $\widehat{\overline{\mathbf{\Gamma}}}$ that are significantly larger than zero. The estimation of the intensity of the influence can be done by exploiting previously developed results. More precisely, the following equality holds:

$$\widehat{\mathbf{W}}(t) = \widehat{\mathbf{D}}\mathbf{\Lambda}^{-1}\left[\overline{\mathbf{\Gamma}}(t) - (1-\beta)\mathbf{I}_n - \beta\mathbf{\Lambda}\left(\mathbf{I}_n - \widehat{\mathbf{D}}^{-1}\right)\right],$$

where $\widehat{\mathbf{D}}$ represents an estimate of the degree matrix $\mathbf{D}$ obtained from the reconstructed support. That is, $\widehat{\mathbf{D}}$ is the diagonal matrix with elements

$$\widehat{\mathbf{D}}_{i,i} = \|\mathrm{supp}(\boldsymbol{\gamma}_i)\|_0,$$

with $\boldsymbol{\gamma}_i^\top$ being the i-th row of matrix $\widehat{\overline{\mathbf{\Gamma}}}$.

Influence estimation in multiplex networks

For the model described in **"F&J model on multiplex networks,"** one can estimate the cross-correlation matrices and then use relations (22) and (23) for each dynamical system replacing the theoretical covariances $\mathbf{\Sigma}^{(s)}_{[\ell]}(t)$ with estimated value $\widehat{\mathbf{\Sigma}}^{(s)}_{[\ell]}(t)$ (see the methodology summarized in Figure 12). Leveraging estimation of VAR processes [182] and on the ergodicity of the dynamical systems, it can be shown that with probability at least $1-\delta$,

$$\|\mathbf{W}^{(s)} - \widehat{\mathbf{W}}^{(s)}(t)\|_F \le \frac{C(n, \|\mathbf{Q}_\eta\|)}{(1-\sigma_{\max})^4\sqrt{t}\rho},$$

where $C(n, \|\mathbf{Q}_\eta\|)$ is a constant independent of t. This bound can be improved by imposing new constraints in the recovery by exploiting correlations among different dynamical systems (see models $\mathcal{M}_c c$ and $\mathcal{M}_{cs}$ in **"F&J model on multiplex networks"**). If the correlations are not known among influence matrices, the idea proposed in [55] is to leverage global properties of the local processes to correct the local estimates of $\mathbf{S}^{(s)}_{[0]}(\infty)$. Moreover, the reconstruction performance suffers when the sample size is not large, as the number of observed data must be larger than the number of unknowns to have a full rank estimation of $\hat{\mathbf{\Sigma}}^{(s)}_{[0]}(\infty)$. The Bayesian approach is a powerful estimation framework, since it combines prior probabilistic information (parametrized by some unknown hyperparameters) and gathered observations.

Bayesian estimation of $\mathrm{S}_{[0]}^{(s)}(\infty)$

In the absence of additional information on the model, the selection of the prior distribution is quite delicate. A commonly used approach is to consider the conjugate prior of the multivariate normal distribution. More precisely, consider the inverse-Wishart with matrix $\mathbf{\Psi}$ and $\nu > n+1$ degrees or, equivalently,

$$\hat{\mathbf{S}}_{[0]}^{(s)}(\infty) = \gamma^{(s)}\bar{\mathbf{S}} + (1-\gamma^{(s)})\hat{\mathbf{S}}_{\mathsf{SCM}}^{(s)}(\infty),$$

where

- $\hat{\mathbf{S}}_{\mathsf{SCM}}^{(s)}(\infty)$ is the sample covariance matrix
- $\bar{\mathbf{S}} = \frac{\mathbf{\Psi}}{\nu-(n+1)}$ is the prior mean/mode
- $\gamma^{(s)} = \frac{\nu-(n+1)}{\nu+T^{(s)}-(n+1)} \in (0,1)$is a term balancing the two contributions according to the sample size $T^{(s)}$ and informative level of the prior (degrees of freedom ν).

The Inverse-Wishart parameters estimations are obtained via **alternating minimization:**

$$\begin{aligned}&(\hat{\mathbf{\Psi}}, \hat{\nu}) \\ &= \underset{\mathbf{\Psi}>0,\, \nu>n+1}{\operatorname{argmin}} \; -\sum_{s=1}^{m} \log \frac{\det^{\frac{\nu}{2}}(\Psi)\,\Gamma_n\!\left(\frac{\nu+T^{(s)}+n}{2}\right)}{\pi^{\frac{nT^{(s)}}{2}}\det^{\frac{\nu+T^{(s)}}{2}}(\Psi+\mathbf{Z}^{(s)}(\mathbf{Z}^{(s)})^{\top})}\end{aligned}$$

In [55], the performance of the proposed estimators are tested within $\mathcal{M}_{cc}$ and $\mathcal{M}_{cs}$. The simulations show that the approach based on the Bayesian method achieves better performance in the estimation. For both considered models, the variance of the reconstruction error is much lower for the proposed approach compared to the conventional ML estimator. It is worth remarking that the recovery of the transition matrices depends significantly on the conditioning of the estimated covariance matrices. Although the matrices are invertible, the reconstruction performance suffer when the sample size is not large. In this sense, the Bayesian method acts as a regularizer of the covariance estimation in an adaptive fashion, that is, with automatic selection of the regularization parameter. By putting a prior distribution on the covariance matrix, the reconstruction formula will be a combination of a sample statistic (computed from the observed data) and a function of the hyperparameters (prior information). The latter can indeed help in the case of scarce data, while its effect vanishes asymptotically as posterior estimates converge to the ML counterparts for large samples (Bernstein-von Mises theorem), thus converging to the classical SCM. This is a "natural weighting" mechanism, which automatically regulates (through the parameters $\gamma^{(s)}$) the relative importance of the prior model and data according to the sample size, automatically switching to a noninformative prior (retrieved for limit values of the hyperparameters) if conversely the sample size is large.

Concluding remarks

Although the phenomenon of social influence has long been studied in social and behavioral sciences, mathematical characterization of influence between individuals is not a trivial task. How to understand which connections between people are most essential and who are the genuine leaders of the group is a difficult problem. Granovetter [65], [202] proposed the theory of “strong” and “weak” ties connecting close friends and acquaintances, respectively. Strong ties build densely connected subgraphs (communities) in a network, whereas weak ties build bridges between these densely knit communities. This principle led to a number of mathematical characteristics [60] measuring influence between two individuals as a function of their positions in a network. At the same time, Granovetter argued that some “weak” ties not only have a strong impact on an individual, but *are actually vital for an individual’s integration into modern society* [202]. “Weak” ties facilitate exchange of information between closed communities, enabling the mobility of labor and integration of individuals into political movements. Hence, “static” characteristics considering only links between individuals and ignoring the specific features of their interactions can be misleading. Alternative methods are needed that consider a social network as a dynamical system. This survey focuses on two novel directions of research concerned with dynamical networks of social influence. The statistical approach adopted in machine learning considers a social network as a probabilistic graphical model treating social influence as a measure of statistical correlation between some data produced by individuals (for example, information on which events they attend and which goods they consume). The approach in social influence network theory [46] considers social influence as a process altering opinions of the individuals; to find the parameters of these models, methods of identification theory should be used. Even for a parsimonious opinion formation model, proposed by Friedkin and Johnson, the problem of parameter identification appears to be nontrivial and is closely related to compressed sensing and other rapidly growing branches of signal processing theory. Many problems related to recovery of influence networks’ structure remain beyond the scope of this survey and are still waiting for solutions. Identification problems become quite challenging when a dynamical model nonlinearly depends on unknown parameters as, for example, bounded confidence models surveyed in [16]. Along with continuous (real-valued) measurements, models can address discrete (finite valued) data as for example, cellular automata considered in physical literature [13] or continuous opinion - discrete action (CODA) models [203]. Even more complicated for analysis is the case of *temporal* social networks where both nodes and arcs can emerge and disappear. Such models are vital to understanding online social networks dynamics where individuals can easily create and delete user profiles. Systems theory lacks tools to address such temporal models, and a promising framework of *open* multiagent systems was recently proposed in [204], [205]. The most challenging problem at the frontier of computer science, social sciences, and systems theory is to extract the structure of an online social network from big data produced by users. Unlike simplified mathematical models, people do not broadcast numbers and communicate via web forums, microblogs, mobile apps, and other social media. The numbers must be extracted from textual and multimedia information, which requires advanced tools for video and language processing, big data analytics, and efficient numerical methods that are able to address large-scale dynamical systems. We hope that this survey will help to recruit young talented researchers to the vibrant and fascinating area of dynamical social network analysis.

Acknowledgment

The authors would like to thank F. Bullo, G. Como, F. Fagnani, P. Frasca, and N. E. Friedkin for interesting conversations on topics related to this article. This article is dedicated to the memory of Roberto Tempo, who has been an outstanding scientist in control and coauthor of many articles on this topic. He has been promoting interdisciplinary research with his positive and encouraging attitude towards spreading and sharing scientific results. Anton Proskurnikov acknowledges financial support of the Russian Foundation for Basic Research (RFBR) under grant 20-01-00619.

References

[1] J. Tang, J. Sun, C. Wang, and Z. Yang, "Social influence analysis in large-scale networks," in *Proceedings of the 15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, ser. KDD '09. New York, NY, USA: ACM, 2009, pp. 807–816. [Online]. Available: http://doi.acm.org/10.1145/1557019.1557108

[2] P. T. Metaxas and E. Mustafaraj, "Social media and the elections," *Science*, vol. 338, no. 6106, pp. 472–473, 2012. [Online]. Available: http://science.sciencemag.org/content/338/6106/472

[3] S. L. Stirland, "The Obama campaign: A great campaign or the greatest," *Wired*, Nov 2008.

[4] H. Allcott and M. Gentzkow, "Social media and fake news in the 2016 election," *Journal of Economic Perspectives*, vol. 31, no. 2, pp. 211–36, May 2017. [Online]. Available: http://www.aeaweb.org/articles?id=10.1257/jep.31.2.211

[5] M. McPherson, L. Smith-Lovin, and J. M. Cook, "Birds of a feather: Homophily in social networks," *Annual Review of Sociology*, vol. 27, no. 1, pp. 415–444, 2001. [Online]. Available: http://arjournals.annualreviews.org/doi/abs/10.1146/annurev.soc.27.1.415

[6] P. Dandekar, A. Goel, and D. T. Lee, "Biased assimilation, homophily, and the dynamics of polarization," *Proceedings of the National Academy of Sciences*, vol. 110, no. 15, pp.

5791–5796, 2013. [Online]. Available: https://www.pnas.org/content/110/15/5791

[7] P. Singla and M. Richardson, "Yes, there is a correlation: - from social networks to personal behavior on the web," in *Proceedings of the 17th International Conference on World Wide Web*, ser. WWW '08. New York, NY, USA: ACM, 2008, pp. 655–664. [Online]. Available: http://doi.acm.org/10.1145/1367497.1367586

[8] J. Bughin, J. Doogan, and O. J. Vetvik, "A new way to measure word-of-mouth marketing," *McKinsey Quarterly*, 01 2010.

[9] S. Wasserman and K. Faust, *Social Network Analysis in the Social and Behavioral Sciences*, ser. Structural Analysis in the Social Sciences. Cambridge University Press, 1994, p. 3?27.

[10] A.-L. Barabasi, *Linked: The New Science of Networks*. Cambridge, MA: Perseus, 2002.

[11] M. Newman, "The structure and function of complex networks," *SIAM Review*, vol. 45, no. 2, pp. 167–256, 2003.

[12] D. Easley and J. Kleinberg, *Networks, Crowds and Markets. Reasoning about a Highly Connected World*. Cambridge: Cambridge Univ. Press, 2010.

[13] C. Castellano, S. Fortunato, and V. Loreto, "Statistical physics of social dynamics," *Rev. Mod. Phys.*, vol. 81, pp. 591–646, 2009.

[14] 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.*, vol. 35, no. 3, pp. 40–51, 2015.

[15] A. Proskurnikov and R. Tempo, "A tutorial on modeling and analysis of dynamic social networks. Part I," *Annual Reviews in Control*, vol. 43, pp. 65–79, 2017.

[16] ——, "A tutorial on modeling and analysis of dynamic social networks. Part II," *Annual Reviews in Control*, vol. 45, 2018.

[17] Y. Dong, M. Zhan, G. Kou, Z. Ding, and H. Liang, "A survey on the fusion process in opinion dynamics," *Information Fusion*, vol. 43, pp. 57–65, 2018.

[18] O. Abid, S. Jamoussi, and Y. B. Ayed, "Deterministic models for opinion formation through communication: A survey," *Online Social Networks and Media*, vol. 6, pp. 1 – 17, 2018. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S246869641730085X

[19] L. Mastroeni, P. Vellucci, and M. Naldi, "Agent-based models for opinion formation: A bibliographic survey," *IEEE Access*, vol. 7, pp. 58 836–58 848, 2019.

[20] M. Cruz-Cunha and F. Moreira, *Handbook of Research on Mobility and Computing: Evolving Technologies and Ubiquitous Impacts*. Information Science Reference, 2011. [Online]. Available: https://books.google.com.ag/books?id=2koyMAEACAAJ

[21] C. Cheung, P.-Y. Chiu, and M. Lee, "Online social networks: Why do students use Facebook?" *Computers in Human Behavior*, vol. 27, pp. 1337–1343, 07 2011.

[22] A. Helmond, “Lifetracing. the traces of a networked life,” in *Networked: a (networked book) about (networked art)*, H. Thorington, E. Navas, and J. Green, Eds. MA: Turbulence: Roslindale, 2009.

[23] S. Kemp, “Digital 2020: 3.8 billion people use social media,” *We are social*, Jan 2020. [Online]. Available: https://wearesocial.com/blog/2020/01/digital-2020-3-8-billion-people-use-social-media

[24] K. Majewska, “Stats on social media use in COVID-19 hotspots: February vs. March 2020,” *napoleoncat.com/blog*, July 2019.

[25] A. Rasool, “Facebook, Instagram, TikTok. Which apps are the most downloaded in July by Apple and Android users? [Report],” *Digitalinformationworld.com*, July 2019.

[26] S. Vosoughi, D. Roy, and S. Aral, “The spread of true and false news online,” *Science*, vol. 359, pp. 1146–1151, 03 2018.

[27] A. Bovet and H. A. Makse, “Influence of fake news in Twitter during the 2016 US presidential election,” *Nature Communications*, vol. 10, no. 1, p. 7, 2019.

[28] N. McCarthy, “Where concern is highest about fake news on the Internet,” *Statista.com*, June 2019.

[29] J. Gramlich, “10 facts about Americans and Facebook,” *Report Pew Research Center*, May 2019.

[30] M. Del Vicario, A. Bessi, F. Zollo, F. Petroni, A. Scala, G. Caldarelli, H. E. Stanley, and W. Quattrociocchi, “The spreading of misinformation online,” *Proceedings of the National Academy of Sciences*, 2016. [Online]. Available: https://www.pnas.org/content/early/2016/01/02/1517441113

[31] F. Pierri, C. Piccardi, and S. Ceri, “Topology comparison of Twitter diffusion networks effectively reveals misleading information,” *Scientific Reports*, vol. 10, 2020.

[32] M. Gomez Rodriguez, J. Leskovec, and A. Krause, “Inferring networks of diffusion and influence,” in *Proceedings of the 16th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, ser. KDD ’10. New York, NY, USA: ACM, 2010, pp. 1019–1028. [Online]. Available: http://doi.acm.org/10.1145/1835804.1835933

[33] M. Gomez Rodriguez, J. Leskovec, D. Balduzzi, and B. Schölkopf, “Uncovering the structure and temporal dynamics of information propagation,” *Network Science*, vol. 2, no. 1, pp. 26–65, 2014.

[34] R. Abelson, “Mathematical models in social psychology,” in *Advances in Experimental Social Psychology, v. 3*, L. Berkowitz, Ed. New York: Acad. Press, 1967, pp. 1–49.

[35] J. Hunter, J. Danes, and S. Cohen, *Mathematical Models of Attitude Change, vol.1*. Acad. Press, Inc., 1984.

[36] P. Bolzern, P. Colaneri, and G. Nicolao, “Opinion influence and evolution in social networks: a markovian agents model,” *Autom.*, vol. 100, pp. 219–230, 2019.

[37] R. Hegselmann, S. König, S. Kurz, C. Niemann, and J. Rambau, "Optimal opinion control: The campaign problem," *Journal of Artificial Societies and Social Simulation*, vol. 18, no. 3, p. 18, 2015. [Online]. Available: http://jasss.soc.surrey.ac.uk/18/3/18.html

[38] N. Masuda, "Opinion control in complex networks," *New Journal of Physics*, vol. 17, no. 3, p. 033031, mar 2015.

[39] J. Zhao, Q. Liu, L. Wang, and X. Wang, "Competitiveness maximization on complex networks," *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, vol. 48, no. 7, pp. 1054–1064, 2018.

[40] M. Grabisch, A. Mandel, A. Rusinowska, and E. Tanimura, "Strategic influence in social networks," *Mathematics of Operations Research*, vol. 43, no. 1, pp. 29–50, 2018.

[41] D. Lazer, A. Pentland, L. Adamic, S. Aral, A.-L. Barabási, D. Brewer, N. Christakis, N. Contractor, J. Fowler, M. Gutmann, T. Jebara, G. King, M. Macy, D. Roy, and M. Van Alstyne, "Computational social science," *Science*, vol. 323, no. 5915, pp. 721–723, 2009. [Online]. Available: http://science.sciencemag.org/content/323/5915/721

[42] B. Liu, *Sentiment Analysis and Opinion Mining*. Morgan & Claypool Publishers, 2012.

[43] M. Tsitsvero, S. Barbarossa, and P. D. Lorenzo, "Signals on graphs: Uncertainty principle and sampling," *IEEE Transactions on Signal Processing*, vol. 64, no. 18, pp. 4845–4860, Sep. 2016.

[44] G. Puy, N. Tremblay, R. Gribonval, and P. Vandergheynst, "Random sampling of bandlimited signals on graphs," *Applied and Computational Harmonic Analysis*, 11 2015.

[45] N. E. Friedkin and E. C. Johnsen, "Social influence networks and opinion change," *Advances in Group Processes*, vol. 16, pp. 1–29, 1999.

[46] ——, *Social Influence Network Theory: A Sociological Examination of Small Group Dynamics*. Cambridge, U.K.: Cambridge Univ. Press,, 2011.

[47] H. Abdi and L. Williams, "Principal component analysis," *Wiley Interdisciplinary Reviews: Computational Statistics*, vol. 2, no. 4, pp. 433–459, 2010. [Online]. Available: http://dx.doi.org/10.1002/wics.101

[48] Y. Eldar and G. Kutyniok, *Compressed Sensing: Theory and Applications*, ser. Compressed Sensing: Theory and Applications. Cambridge University Press, 2012.

[49] D. Yan, Y. Tian, and J. Cheng, *Systems for Big Graph Analytics*, ser. Springer Briefs in Computer Science. Springer, 2017.

[50] P. Frasca, C. Ravazzi, R. Tempo, and H. Ishii, "Gossips and prejudices: Ergodic randomized dynamics in social networks," in *Proc. of IFAC NecSys 2013 Workshop*, Koblenz, Germany, 2013, pp. 212–219.

[51] P. Frasca, H. Ishii, C. Ravazzi, and R. Tempo, "Distributed randomized algorithms for opinion formation, centrality computation and power systems estimation: A tutorial overview," *Europ. J. Control*, vol. 24, no. 7, pp. 2–13, 2015.

[52] C. Ravazzi, P. Frasca, R. Tempo, and H. Ishii, “Ergodic randomized algorithms and dynamics over networks,” *IEEE Trans. Control of Network Syst.*, vol. 2, no. 1, pp. 78–87, 2015.

[53] C. Ravazzi, R. Tempo, and F. Dabbene, “Learning influence structure in sparse social networks,” *IEEE Transactions on Control of Network Systems*, vol. 5, no. 4, pp. 1976–1986, Dec 2018.

[54] C. Ravazzi, S. Hojjatinia, C. M. Lagoa, and F. Dabbene, “Randomized opinion dynamics over networks: influence estimation from partial observations,” in *2018 IEEE Conference on Decision and Control (CDC)*, 2018, pp. 2452–2457.

[55] A. Coluccia, F. Dabbene, and C. Ravazzi, “Bayesian identification of distributed vector autoregressive processes,” in *2019 18th European Control Conference (ECC)*, 2019, pp. 842–847.

[56] S. Parsegov, A. Proskurnikov, R. Tempo, and N. Friedkin, “Novel multidimensional models of opinion dynamics in social networks,” *IEEE Trans. Autom. Control*, vol. 62, no. 5, pp. 2270–2285, 2017.

[57] B. D. Anderson, F. Dabbene, A. V. Proskurnikov, C. Ravazzi, and M. Ye, “Dynamical networks of social influence: Modern trends and perspectives,” *IFAC-PapersOnLine*, 2020, in Proceedings of IFAC World Congress 2020.

[58] N. E. Friedkin, *A Structural Theory of Social Influence*, 1998.

[59] P. V. Marsden and N. E. Friedkin, “Network studies of social influence,” *Sociological Methods & Research*, vol. 22, no. 1, pp. 127–151, 1993. [Online]. Available: https://doi.org/10.1177/0049124193022001006

[60] J. Sun and J. Tang, “A survey of models and algorithms for social influence analysis,” in *Social Network Data Analytics*, C. C. Aggarwal, Ed. Boston, MA: Springer US, 2011, pp. 177–214.

[61] J. G. March, “Measurement concepts in the theory of influence,” *The Journal of Politics*, vol. 19, no. 2, pp. 202–226, 1957. [Online]. Available: https://doi.org/10.2307/2127196

[62] H. A. Simon, “Notes on the observation and measurement of political power,” *The Journal of Politics*, vol. 15, no. 4, pp. 500–516, 1953. [Online]. Available: http://www.jstor.org/stable/2126538

[63] J. French Jr., “A formal theory of social power,” *Physchol. Rev.*, vol. 63, pp. 181–194, 1956.

[64] J. French and B. Raven, “The bases of social power,” in *Studies in Social Power*, D. Cartwright, Ed. Ann Arbor, MI: Univ. of Michigan Press, 1959, pp. 251–260.

[65] M. S. Granovetter, “The strength of weak ties,” *American Journal of Sociology*, vol. 78, no. 6, pp. 1360–1380, 1973.

[66] M. I. Jordan, "Graphical models," *Statist. Sci.*, vol. 19, no. 1, pp. 140–155, 02 2004. [Online]. Available: https://doi.org/10.1214/088342304000000026

[67] E. M. Airoldi, "Getting started in probabilistic graphical models," *PLOS Computational Biology*, vol. 3, no. 12, pp. 1–5, 12 2007. [Online]. Available: https://doi.org/10.1371/journal.pcbi.0030252

[68] D. S. Sade, "Sociometrics of macaca mulatta iii: n-path centrality in grooming networks," *Social Networks*, vol. 11, no. 3, pp. 273 – 292, 1989, special Issue on Non-Human Primate Networks.

[69] L. C. Freeman, "Centrality in social networks: Conceptual clarification," *Social networks*, vol. 1, no. 3, pp. 215–239, 1979.

[70] L. Krempel, "Dagstuhl seminar: Link analysis and visualization," July 2001.

[71] L. C. Freeman, "A set of measures of centrality based upon betweenness," *Sociometry*, vol. 40, pp. 35–41, 1977.

[72] D. R. White and S. P. Borgatti, "Betweenness centrality measures for directed graphs," *Social Networks*, vol. 16, no. 4, pp. 335–346, 1994.

[73] A. Z. Broder, R. Lempel, F. Maghoul, and J. Pedersen, "Efficient pagerank approximation via graph aggregation," *Information Retrieval*, vol. 9, no. 2, pp. 123–138, Mar 2006.

[74] S. Brin and L. Page, "The anatomy of a large-scale hypertextual web search engine," *Computer Networks and ISDN Systems*, vol. 30, no. 1, pp. 107–117, 1998.

[75] L. Katz, "A new status index derived from sociometric analysis," *Psychometrika*, vol. 18, no. 1, pp. 39–43, Mar 1953.

[76] P. Bonacich, "Power and centrality: A family of measures," *American Journal of Sociology*, vol. 92, no. 5, pp. 1170–1182, 1987.

[77] P. Bonacich and P. Lloyd, "Eigenvector-like measures of centrality for asymmetric relations," *Social Networks*, vol. 23, pp. 191–201, 2001.

[78] N. Friedkin, "Theoretical foundations for centrality measures," *Amer. J. Sociology*, vol. 96, no. 6, pp. 1478–1504, 1991.

[79] J. Leskovec and R. Sosič, "Snap: A general-purpose network analysis and graph-mining library," *ACM Transactions on Intelligent Systems and Technology (TIST)*, vol. 8, no. 1, p. 1, 2016.

[80] J. Ellson, E. Gansner, L. Koutsofios, S. North, G. Woodhull, S. Description, and L. Technologies, "Graphviz ? open source graph drawing tools," in *Lecture Notes in Computer Science*. Springer-Verlag, 2001, pp. 483–484.

[81] M. Scholz, "Node similarity as a basic principle behind connectivity in complex networks," *Journal of Data Mining & Digital Humanities*, vol. 2015, Jan 2015. [Online]. Available: https://jdmdh.episciences.org/77

[82] F. Farris, "The Gini index and measures of inequality," *American Mathematical Monthly*, vol. 117, pp. 851–864, 12 2010.

[83] S. Goswami, C. A. Murthy, and A. K. Das, "Sparsity measure of a network graph: Gini index," *Inf. Sci.*, vol. 462, pp. 16–39, 2018. [Online]. Available: https://doi.org/10.1016/j.ins.2018.05.044

[84] J. Moreno and H. Jennings, "Statistics of social configurations," *Sociometry*, vol. 1, no. 3/4, pp. 324–374, 1938.

[85] V. Arnaboldi, M. Conti, M. L. Gala, A. Passarella, and F. Pezzoni, "Ego network structure in online social networks and its impact on information diffusion," *Computer Communications*, vol. 76, pp. 26 – 41, 2016. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S014036641500465X

[86] C. Boldrini, M. Toprak, M. Conti, and A. Passarella, "Twitter and the press: An ego-centred analysis," in *Companion Proceedings of the The Web Conference 2018*, ser. WWW '18. Republic and Canton of Geneva, Switzerland: International World Wide Web Conferences Steering Committee, 2018, pp. 1471–1478. [Online]. Available: https://doi.org/10.1145/3184558.3191596

[87] V. Arnaboldi, M. Conti, A. Passarella, and F. Pezzoni, "Analysis of Ego-network structure in online social networks," in *2012 International Conference on Privacy, Security, Risk and Trust and 2012 International Confernece on Social Computing*, Sep. 2012, pp. 31–40.

[88] J. F. Bokunewicz and J. Shulman, "Influencer identification in Twitter networks of destination marketing organizations," *Journal of Hospitality and Tourism Technology*, vol. 8, no. 2, pp. 205–219, 2017. [Online]. Available: https://doi.org/10.1108/JHTT-09-2016-0057

[89] T. Opsahl and P. Panzarasa, "Clustering in weighted networks." *Social Networks*, vol. 31, no. 2, pp. 155–163, 2009. [Online]. Available: http://dblp.uni-trier.de/db/journals/socnet/socnet31.html#OpsahlP09

[90] I. Rish, G. A. Cecchi, A. Lozano, and A. Niculescu-Mizil, *Practical Applications of Sparse Modeling*. The MIT Press, 2014.

[91] L. Liu, J. Tang, J. Han, and S. Yang, "Learning influence from heterogeneous social networks," *Data Mining and Knowledge Discovery*, vol. 25, no. 3, pp. 511–544, Nov 2012. [Online]. Available: https://doi.org/10.1007/s10618-012-0252-3

[92] M. T. Thai, W. Wu, and H. Xiong, *Big Data in Complex and Social Networks*, 1st ed. Chapman & Hall/CRC, 2016.

[93] A. Coluccia, "Regularized covariance matrix estimation via empirical bayes," *IEEE Signal Processing Letters*, vol. 22, pp. 2127–2131, 2015.

[94] N. E. Friedkin, P. Jia, and F. Bullo, "A theory of the evolution of social power: Natural trajectories of interpersonal influence systems along issue sequences," *Sociological*

Science, vol. 3, pp. 444–472, 2016.

[95] C. Cooley, *Human nature and the social order*. Scribner, 1922. [Online]. Available: https://books.google.mw/books?id=cVDxKsGmxggC

[96] N. E. Friedkin, "A formal theory of reflected appraisals in the evolution of power," *Administrative Science Quarterly*, vol. 56, no. 4, pp. 501–529, 2020/06/28 2011. [Online]. Available: https://doi.org/10.1177/0001839212441349

[97] C. Perillo and S. Battiston, "A multiplex financial network approach to policy evaluation: the case of euro area quantitative easing," *Applied Network Science*, vol. 3, 12 2018.

[98] J. Wu, C. Pu, L. Li, and G. Cao, "Traffic dynamics on multilayer networks," *Digital Communications and Networks*, vol. 6, no. 1, pp. 58 – 63, 2020. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S2352864818301135

[99] A. Halu, M. De Domenico, A. Arenas, and A. Sharma, "The multiplex network of human diseases," *bioRxiv*, 2017. [Online]. Available: https://www.biorxiv.org/content/early/2017/01/18/100370

[100] M. Berlingerio, M. Coscia, F. Giannotti, A. Monreale, and D. Pedreschi, "Multidimensional networks: Foundations of structural analysis," *World Wide Web*, vol. 16, 11 2013.

[101] X. Dong, D. Thanou, M. Rabbat, and P. Frossard, "Learning graphs from data: A signal representation perspective," *IEEE Signal Processing Magazine*, vol. 36, no. 3, pp. 44–63, May 2019.

[102] D. Koller and N. Friedman, *Probabilistic Graphical Models: Principles and Techniques - Adaptive Computation and Machine Learning*. The MIT Press, 2009.

[103] H. Rue and L. Held, *Gaussian Markov Random Fields: Theory And Applications (Monographs on Statistics and Applied Probability)*. Chapman & Hall/CRC, 2005.

[104] M. J. Wainwright and M. I. Jordan, "Graphical models, exponential families, and variational inference," *Found. Trends Mach. Learn.*, vol. 1, no. 1-2, pp. 1–305, Jan. 2008. [Online]. Available: http://dx.doi.org/10.1561/2200000001

[105] G. Robins, P. Pattison, Y. Kalish, and D. Lusher, "An introduction to exponential random graph (p^*) models for social networks," *Social networks*, vol. 29, pp. 173–191, 2007.

[106] G. B. Giannakis, Y. Shen, and G. V. Karanikolas, "Topology identification and learning over graphs: Accounting for nonlinearities and dynamics," *Proceedings of the IEEE*, vol. 106, no. 5, pp. 787–807, May 2018.

[107] R. Mazumder and T. Hastie, "The graphical lasso: New insights and alternatives," *Electron. J. Statist.*, vol. 6, pp. 2125–2149, 2012. [Online]. Available: https://doi.org/10.1214/12-EJS740

[108] J. Friedman, T. Hastie, and R. Tibshirani, "Sparse inverse covariance estimation with the graphical lasso," *Biostatistics*, vol. 9, no. 3, pp. 432–441, Jul. 2008. [Online]. Available: http://biostatistics.oxfordjournals.org/content/9/3/432

[109] D. I. Shuman, S. K. Narang, P. Frossard, A. Ortega, and P. Vandergheynst, "The emerging field of signal processing on graphs: Extending high-dimensional data analysis to networks and other irregular domains." *IEEE Signal Process. Mag.*, vol. 30, no. 3, pp. 83–98, 2013. [Online]. Available: http://dblp.uni-trier.de/db/journals/spm/spm30.html#ShumanNFOV13

[110] A. Sandryhaila and J. M. F. Moura, "Discrete signal processing on graphs: Frequency analysis," *IEEE Transactions on Signal Processing*, vol. 62, no. 12, pp. 3042–3054, June 2014.

[111] X. Dong, D. Thanou, P. Frossard, and P. Vandergheynst, "Learning Laplacian matrix in smooth graph signal representations," *IEEE Transactions on Signal Processing*, vol. 64, no. 23, pp. 6160–6173, Dec 2016.

[112] S. P. Chepuri, S. Liu, G. Leus, and A. O. Hero, "Learning sparse graphs under smoothness prior," in *Proc. IEEE Int. Conf. Acoustics, Speech and Signal Processing*, 2017, pp. 6508–6512.

[113] X. Zhang, X. Dong, and P. Frossard, "Learning of structured graph dictionaries," in *2012 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, March 2012, pp. 3373–3376.

[114] J. Mei and J. M. F. Moura, "Signal processing on graphs: Causal modeling of unstructured data," *IEEE Transactions on Signal Processing*, vol. 65, no. 8, pp. 2077–2092, 2017.

[115] A. Bolstad, B. D. V. Veen, and R. Nowak, "Causal network inference via group sparse regularization," *IEEE Transactions on Signal Processing*, vol. 59, no. 6, pp. 2628–2641, 2011.

[116] D. Bertsimas, V. Gupta, and I. C. Paschalidis, "Data-driven estimation in equilibrium using inverse optimization," *Mathematical Programming*, vol. 153, no. 2, pp. 595–633, Nov 2015. [Online]. Available: https://doi.org/10.1007/s10107-014-0819-4

[117] H. Wai, A. Scaglione, and A. Leshem, "Active sensing of social networks," *IEEE Transactions on Signal and Information Processing over Networks*, vol. 2, pp. 406–419, 2016.

[118] N. Friedkin and E. Johnsen, *Social Influence Network Theory*. New York: Cambridge Univ. Press, 2011.

[119] N. H. Anderson, *Foundations of Information Integration Theory*. Academic Press, 1981.

[120] L. Festinger, *A Theory of Cognitive Dissonance*. Stanford, CA: Stanford Univ. Press, 1957.

[121] F. Harary, "A criterion for unanimity in French's theory of social power," in *Studies in Social Power*, D. Cartwright, Ed. Ann Arbor, MI: Univ. of Michigan Press, 1959, pp. 168–182.

[122] M. Jackson, Ed., *Social and Economic Networks*. Princeton and Oxford: Princeton Univ. Press, 2008.

[123] F. Harary, R. Norman, and D. Cartwright, *Structural Models. An Introduction to the Theory of Directed Graphs*. New York, London, Sydney: Wiley & Sons, 1965.

[124] M. H. DeGroot, "Reaching a Consensus," *Journal of the American Statistical Association*, vol. 69, no. 345, pp. 118–121, 1974.

[125] K. Lehrer, "When rational disagreement is impossible," *Noûs*, vol. 10, no. 3, pp. 327–332, 1976.

[126] K. Lehrer and C. Wagner, *Rational Consensus in Science and Society*. Dordrecht, Boston, London: D. Reidel Publ., 1981.

[127] N. E. Friedkin, A. V. Proskurnikov, W. Mei, and F. Bullo, "Mathematical structures in group decision-making on resource allocation distributions," *Scientific Reports*, vol. 9, no. 1, p. 1377, 2019.

[128] L. Li, A. Scaglione, A. Swami, and Q. Zhao, "Consensus, polarization and clustering of opinions in social networks," *IEEE J. on Selected Areas in Communications*, vol. 31, no. 6, pp. 1072–1083, 2013.

[129] N. Friedkin, "A formal theory of social power," *J. of Math. Sociology*, vol. 12, no. 2, pp. 103–126, 1986.

[130] J. Wolfowitz, "Products of indecomposable, aperiodic, stochastic matrices," *Proceedings of Amer. Math. Soc.*, vol. 15, pp. 733–737, 1963.

[131] F. Gantmacher, *The Theory of Matrices*. AMS Chelsea Publishing, 2000, vol. 2.

[132] R. Horn and C. Johnson, *Topics in Matrix Analysis*. New York: Cambridge Univ. Press, 1991.

[133] W. Ren and R. Beard, *Distributed Consensus in Multi-Vehicle Cooperative Control: Theory and Applications*. London: Springer-Verlag, 2008.

[134] F. Bullo, *Lectures on Network Systems*, 1st ed. Kindle Direct Publishing, 2019, with contributions by J. Cortes, F. Dorfler, and S. Martinez. [Online]. Available: http://motion.me.ucsb.edu/book-lns

[135] R. Abelson, "Mathematical models of the distribution of attitudes under controversy," in *Contributions to Mathematical Psychology*, N. Frederiksen and H. Gulliksen, Eds. New York: Holt, Rinehart & Winston, Inc, 1964, pp. 142–160.

[136] R. Hegselmann and U. Krause, "Opinion dynamics and bounded confidence, models, analysis and simulation," *Journal of Artificial Societies and Social Simulation*, vol. 5, no. 3, p. 2, 2002.

[137] G. Deffuant, D. Neau, F. Amblard, and G. Weisbuch, "Mixing beliefs among interacting agents," *Advances in Complex Systems*, vol. 3, pp. 87–98, 2000.

[138] M. W. Macy, J. Kitts, A. Flache, and S. Benard, "Polarization in dynamic networks: a Hopfield model of emergent structure," in *Dynamic Social Network Modeling and*

Analysis: Workshop Summary and Papers, R. Breiger, K. Carley, and P. Pattison, Eds., 2003, pp. 162–173.

[139] D. Urbig, "Attitude dynamics with limited verbalisation capabilities," *Journal of Artifical Societies and Social Simulation (JASSS)*, vol. 6, no. 1, p. 2, 2003.

[140] A. Fläche and M. Macy, "Small worlds and cultural polarization," *Journal of Math. Sociology*, vol. 35, no. 1–3, pp. 146–176, 2011.

[141] M. Mäs, A. Flache, and D. Helbing, "Individualization as driving force of clustering phenomena in humans," *PLoS Comput. Biol.*, vol. 6, no. 10, p. e1000959, 2010.

[142] M. Mäs and A. Flache, "Differentiation without distancing. explaining bi-polarization of opinions without negative influence," *PLoS One*, vol. 8, no. 11, p. e74516, 2013.

[143] S. Motsch and E. Tadmor, "Heterophilious dynamics enhances consensus," *SIAM Review*, vol. 56, no. 4, pp. 577–621, 2013.

[144] F. Ceragioli and P. Frasca, "Consensus and disagreement: The role of quantized behaviors in opinion dynamics," *SIAM J. Control Optim.*, vol. 56, no. 2, pp. 1058–1080, 2018.

[145] J. Jin, Y. Yuan, W. Pan, C. Tomlin, A. A. Webb, and J. Gonçalves, "Identification of nonlinear sparse networks using sparse Bayesian learning," in *2017 IEEE 56th Annual Conference on Decision and Control (CDC)*, 2017, pp. 6481–6486.

[146] N. M. Mangan, S. L. Brunton, J. L. Proctor, and J. N. Kutz, "Inferring biological networks by sparse identification of nonlinear dynamics," *IEEE Transactions on Molecular, Biological and Multi-Scale Communications*, vol. 2, no. 1, pp. 52–63, 2016.

[147] E. Kaiser, J. N. Kutz, and S. L. Brunton, "Sparse identification of nonlinear dynamics for model predictive control in the low-data limit," *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*, vol. 474, no. 2219, p. 20180335, 2018.

[148] C. Altafini, "Dynamics of opinion forming in structurally balanced social networks," *PLoS ONE*, vol. 7, no. 6, p. e38135, 2012.

[149] ——, "Consensus problems on networks with antagonistic interactions," *IEEE Trans. Autom. Control*, vol. 58, no. 4, pp. 935–946, 2013.

[150] Z. Cao, F. Jiao, X. Qu, W.-X. Wang, M. Yang, X. Yang, and B. Zhang, "Rebels lead to the doctrine of the mean: A heterogeneous DeGroot model," *Journal of Systems Science and Complexity*, vol. 31, no. 6, pp. 1498–1509, 2018.

[151] W. Xia, M. Cao, and K. Johansson, "Structural balance and opinion separation in trust–mistrust social networks," *IEEE Trans. Control Netw. Syst.*, vol. 3, no. 1, pp. 46–56, 2016.

[152] J. Liu, X. Chen, T. Basar, and M. Belabbas, "Exponential convergence of the discrete- and continuous-time Altafini models," *IEEE Trans. Autom. Control*, vol. 62, no. 12, pp. 6168–6182, 2017.

[153] W. Xia and M. Cao, "Clustering in diffusively coupled networks," *Automatica*, vol. 47, no. 11, pp. 2395–2405, 2011.

[154] G. Shi, A. Proutiere, M. Johansson, J. Baras, and K. Johansson, "Emergent behaviors over signed random dynamical networks: State-flipping model," *IEEE Trans. on Control of Networked Systems*, vol. 2, no. 2, pp. 142–153, 2015.

[155] ——, "The evolution of beliefs over signed social networks," *Oper. Res.*, vol. 64, no. 3, pp. 585–604, 2016.

[156] G. Shi, C. Altafini, and J. Baras, "Dynamics over signed networks," *SIAM Review*, vol. 61, no. 2, pp. 229–257, 2019.

[157] C. Hovland, I. Janis, and H. Kelley, *Communication and Persuasion*. New Haven: Yale Univ. Press, 1953.

[158] K. Takács, A. Fläche, and M. Mäs, "Discrepancy and disliking do not induce negative opinion shifts," *PLoS One*, vol. 11, no. 6, p. e0157948, 2016.

[159] A. Proskurnikov, A. Matveev, and M. Cao, "Consensus and polarization in Altafini's model with bidirectional time-varying network topologies," in *Proceedings of IEEE CDC 2014*, Los Angeles, CA, 2014, pp. 2112–2117.

[160] A. Proskurnikov and M. Cao, "Opinion dynamics using Altafini's model with a time-varying directed graph," in *Proceedings of IEEE ISIC 2014 (Part of IEEE MSC 2014)*, Antibes, 2014, pp. 849–854.

[161] D. Meng, "Bipartite containment tracking of signed networks," *Automatica*, vol. 79, pp. 282 – 289, 2017.

[162] D. Meng, M. Du, and Y. Wu, "Extended structural balance theory and method for cooperative-antagonistic networks," *IEEE Transactions on Automatic Control*, vol. 65, no. 5, pp. 2147–2154, 2020.

[163] A. Proskurnikov, A. Matveev, and M. Cao, "Opinion dynamics in social networks with hostile camps: Consensus vs. polarization," *IEEE Trans. Autom. Control*, vol. 61, no. 6, pp. 1524–1536, 2016.

[164] D. Acemoğlu, G. Como, F. Fagnani, and A. Ozdaglar, "Opinion fluctuations and disagreement in social networks," *Math. Oper. Res.*, vol. 38, no. 1, pp. 1–27, Feb. 2013. [Online]. Available: http://dx.doi.org/10.1287/moor.1120.0570

[165] C. Childress and N. Friedkin, "Cultural reception and production," *Amer. Soc. Rev.*, vol. 77, pp. 45–68, 2012.

[166] A. V. Proskurnikov, R. Tempo, M. Cao, and N. E. Friedkin, "Opinion evolution in time-varying social influence networks with prejudiced agents," *IFAC-PapersOnLine*, vol. 50, no. 1, pp. 11 896–11 901, 2017.

[167] H. Ishii and R. Tempo, "The PageRank problem, multiagent consensus, and web aggregation: A systems and control viewpoint," *IEEE Control Syst. Mag.*, vol. 34, no. 3, pp. 34–53, 2014.

[168] N. Friedkin and E. Johnsen, "Two steps to obfuscation," *Social Networks*, vol. 39, pp. 12–13, 2014.

[169] A. Proskurnikov, R. Tempo, and M. Cao, "PageRank and opinion dynamics: Missing links and extensions," in *Proc. of IEEE Conf. Norbert Wiener in the 21st Century*, Melbourne, 2016, pp. 12–17.

[170] P. Converse, "The nature of belief systems in mass publics," in *Ideology and Discontent*, D. Apter, Ed. New York: Free Press, 1964, pp. 206–261.

[171] B. Gawronski and F. Strack, Eds., *Cognitive consistency: A fundamental principle in social cognition*. New York, NY: Gulford Press, 2012.

[172] H. Noorazar, M. J. Sottile, and K. R. Vixie, "An energy-based interaction model for population opinion dynamics with topic coupling," *International Journal of Modern Physics C*, vol. 29, no. 11, p. 1850115, 2018.

[173] F. Xiong, Y. Liu, L. Wang, and X. Wang, "Analysis and application of opinion model with multiple topic interactions," *Chaos: An Interdisciplinary Journal of Nonlinear Science*, vol. 27, no. 8, p. 083113, 2017.

[174] H. Noorazar, "Recent advances in opinion propagation dynamics: a 2020 survey," *Eur. Phys. J. Plus*, vol. 135, p. 521, 2020.

[175] S. Parsegov, A. Proskurnikov, R. Tempo, and N. Friedkin, "A new model of opinion dynamics for social actors with multiple interdependent attitudes and prejudices," in *Proc. of IEEE Conference on Decision and Control (CDC)*, 2015, pp. 3475 – 3480.

[176] N. Friedkin, A. Proskurnikov, R. Tempo, and S. Parsegov, "Network science on belief system dynamics under logic constraints," *Science*, vol. 354, no. 6310, pp. 321–326, 2016.

[177] S. Boyd, A. Ghosh, B. Prabhakar, and D. Shah, "Randomized gossip algorithms," *IEEE Trans. Inform. Theory*, vol. 52, no. 6, pp. 2508–2530, 2006.

[178] P. Jia, A. MirTabatabaei, N. E. Friedkin, and F. Bullo, "Opinion dynamics and the evolution of social power in influence networks," *SIAM Review*, vol. 57, no. 3, pp. 367–397, 2015.

[179] N. Friedkin and E. Johnsen, "Social influence and opinions," *J. Math. Sociol.*, vol. 15, no. 3-4, pp. 193–205, 1990.

[180] G. Grimmett and D. Stirzaker, *Probability and Random Processes*, ser. Probability and Random Processes. OUP Oxford, 2001.

[181] C. Ravazzi, S. Hojjatinia, F. Dabbene, and C. M. Lagoa, "Randomized opinion dynamics over networks: influence estimation from partial observations," *IEEE Trans. on Automatic Control, submitted*, 2020.

[182] M. Rao, T. Javidi, Y. C. Eldar, and A. Goldsmith, "Estimation in autoregressive processes with partial observations," in *2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2017, pp. 4212–4216.

[183] X. Wu, H. Wai, and A. Scaglione, "Estimating social opinion dynamics models from voting records," *IEEE Transactions on Signal Processing*, 4 2018.

[184] H.-T. Wai, A. Scaglione, and A. Leshem, "Chapter 22 - active sensing of social networks: Network identification from low-rank data," in *Cooperative and Graph Signal Processing*. Academic Press, 2018, pp. 601 – 622. [Online]. Available: http://www.sciencedirect.com/science/article/pii/B9780128136775000225

[185] H.-T. Wai, S. Segarra, A. Ozdaglar, A. Scaglione, and A. Jadbabaie, "Blind community detection from low-rank excitations of a graph filter," *IEEE Transactions on Signal Processing*, vol. 68, pp. 436–451, 2020, cited By 0. [Online]. Available: https://www.scopus.com/inward/record.uri?eid=2-s2.0-85079694589&doi=10.1109%2fTSP.2019.2961296&partnerID=40&md5=bd84f463435bfab686633d96fa6fae88

[186] H. Wai, A. Scaglione, B. Barzel, and A. Leshem, "Joint network topology and dynamics recovery from perturbed stationary points," *IEEE Transactions on Signal Processing*, vol. 67, no. 17, pp. 4582–4596, Sep. 2019.

[187] B. Yilmaz, K. Bekiroglu, C. Lagoa, and M. Sznaier, "A randomized algorithm for parsimonious model identification," *IEEE Transactions on Automatic Control*, vol. 63, no. 2, pp. 532–539, Feb. 2018.

[188] M. Timme, "Revealing Network Connectivity from Response Dynamics," *Physical Review Letters*, vol. 98, no. 22, pp. 224 101+, May 2007.

[189] X. Wang, X. Chen, and L. Wang, "Evolutionary dynamics of fairness on graphs with migration," *Journal of Theoretical Biology*, vol. 380, pp. 103 – 114, 2015.

[190] A. De, S. Bhattacharya, P. Bhattacharya, N. Ganguly, and S. Chakrabarti, "Learning a linear influence model from transient opinion dynamics," in *Proceedings of the 23rd ACM International Conference on Conference on Information and Knowledge Management*, ser. CIKM '14. New York, NY, USA: Association for Computing Machinery, 2014, pp. 401–410.

[191] W.-X. Wang, Y.-C. Lai, C. Grebogi, and J. Ye, "Network reconstruction based on evolutionary-game data via compressive sensing," *Phys. Rev. X*, vol. 1, p. 021021, Dec 2011. [Online]. Available: https://link.aps.org/doi/10.1103/PhysRevX.1.021021

[192] G. Strang, *Introduction to Linear Algebra*. Wellesley-Cambridge Press, 2009.

[193] M. Grant and S. Boyd, "CVX: Matlab software for disciplined convex programming, version 2.1," http://cvxr.com/cvx, Mar. 2014.

[194] E. J. Candès, J. K. Romberg, and T. Tao, "Stable signal recovery from incomplete and inaccurate measurements," *Comm. Pure Appl. Math.*, vol. 59, no. 8, pp. 1207–1223, Aug. 2006.

[195] E. Candes, "The restricted isometry property and its implications for compressed sensing," *Comptes Rendus Mathematique*, vol. 346, no. 9-10, pp. 589–592, May 2008.

[196] A. Mirtabatabaei, P. Jia, and F. Bullo, "Eulerian opinion dynamics with bounded confidence and exogenous inputs," *Siam Journal on Applied Dynamical Systems*, vol. 13, no. 1, p. 425–446, 2014.

[197] J. Lorenz, "Universality in movie rating distributions," Jul. 2009.

[198] E. J. Candes and T. Tao, "Decoding by linear programming," *IEEE Trans. Inf. Theor.*, vol. 51, no. 12, pp. 4203–4215, Dec. 2005.

[199] D. J. Watts and S. H. Strogatz, "Collective dynamics of'small-world'networks." *Nature*, vol. 393, no. 6684, pp. 409–10, 1998.

[200] R. Z. Albert, "Statistical mechanics of complex networks," Ph.D. dissertation, University of Notre Dame, Notre Dame, IN, USA, 2001, aAI3000268.

[201] M. Rao, A. Kipnis, T. Javidi, Y. C. Eldar, and A. Goldsmith, "System identification from partial samples: Non-asymptotic analysis," in *2016 IEEE 55th Conference on Decision and Control (CDC)*, 2016, pp. 2938–2944.

[202] M. S. Granovetter, "The strength of weak ties: A network theory revisited," *Sociological Theory*, vol. 1, pp. 201–233, 1983.

[203] A. C.R. Martins, "Discrete opinion models as a limit case of the coda model," *Physica A: Statistical Mechanics and its Applications*, vol. 395, pp. 352 – 357, 2014.

[204] J. M. Hendrickx and S. Martin, "Open multi-agent systems: Gossiping with random arrivals and departures," in *2017 IEEE 56th Annual Conference on Decision and Control (CDC)*, 2017, pp. 763–768.

[205] M. Franceschelli and P. Frasca, "Proportional dynamic consensus in open multi-agent systems," in *2018 IEEE Conference on Decision and Control (CDC)*, 2018, pp. 900–905.

Author Biography

Chiara Ravazzi (M'13) received the Ph.D. degree in mathematics for engineering sciences from Politecnico di Torino in 2011. In 2010, she held a visiting position at the Laboratory for Information and Decision Systems, Massachusetts Institute of Technology, Cambridge, an, from 2011 to 2016, she held postdoctoral positions in the Departments of Mathematics and Electronics and Telecommunications, Politecnico di Torino. Since 2017, she has been a tenured researcher with the CNR-IEIIT, working with the Systems Modeling and Control Group. She has served as associate editor of the IEEE Transactions on Signal Processing and IEEE Control Systems Letters. Her research interests include signal processing, optimization, and control of network systems.

Fabrizio Dabbene (SM'09) is a director of research at the IEIIT of the National Research Council of Italy (CNR), where he is the coordinator of the Information & Systems Technology Group. He has held visiting and research positions at The University of Iowa, Penn State University, and the Russian Academy of Sciences, Institute of Control Science, Moscow. He has published more than 100 research articles and two books. He served as associate editor for Automatica (2008-2014) and IEEE Transactions on Automatic Control (2008-2012), and he is currently senior editor of the IEEE Control Systems Society Letters. He was elected a member of the Board of Governors (2014-2016), and he served as vice-president for publications (2015-2016). He is currently chairs the IEEE-CSS Italy Chapter. **Constantino M. Lagoa** received the Ph.D. degree from the University of Wisconsin at Madison in 1998. He joined the Electrical Engineering Department of Pennsylvania State University in August 1998, where he currently holds the position of professor. He has a wide range of research interests, including robust optimization and control, chance constrained optimization, controller design under risk specifications, system identification, and control of computer networks. Dr. Lagoa has served as associate editor of IEEE Transactions on Automatic Control (2012-2017) and IEEE Transactions on Control Systems Technology (2009-2013), and he is currently associate editor of Automatica. **Anton Proskurnikov** (M'13, SM'18) is an assistant professor in the Department of Electronics and Telecommunications, Politecnico di Torino (Italy) and a part-time senior researcher at the Institute for Problems in Mechanical Engineering of the Russian Academy of Sciences. He received the M.Sc. and Ph.D. degrees in applied mathematics from St. Petersburg State University in 2003 and 2005, respectively. He previously was an assistant professor at St. Petersburg State University (2003-2010), and a postdoctoral researcher at the University of Groningen (2014-2016) and Delft University of Technology (2016-2018). Anton Proskurnikov is an associate editor in IEEE Transactions on Automatic Control and Journal of Mathematical Sociology. His research interests include network science, multiagent systems and technologies, nonlinear dynamics, control applications in social and biological sciences.

TABLE 1: Comparison between selected model-based learning methods.

	Model assumptions	**Graph assumptions**	**Experiments**	**Sampling**
Abir & al. [190]	DeGroot	Sparse networks	Passive	Full observations
Wang & al. [191]	DeGroot	Sparse models	Passive	Full observations
H. T. Wai & al. [117]	DeGroot	Sparse models	Controlled	Infinite horizon
H. T. Wai & al. [184]	DeGroot	Sparse/Low rank models	Controlled	Infinite horizon
H. T. Wai & al. [185]	DeGroot	Low rank models	Controlled	Infinite horizon
H. T. Wai & al. [186]	Nonlinear dynamics	Sparse	Controlled	Infinite horizon
Ravazzi & al. [53]	F&J dynamics	Sparse models	Passive	Infinite horizon
Ravazzi & al. [54], [181]	F&J dynamics	Sparse models	Passive	Random intermittent measurements
Coluccia & al. [55]	F&J dynamics	Distributed sparse models	Passive	Random intermittent measurements
Anderson & al. [57]	F&J dynamics	Sparse models	Passive	Finite/Infinite horizon

TABLE 2: Main notation symbols.

Symbol	First appearance	Meaning
$\mathcal{G} = (\mathcal{V}, \mathcal{E})$	p. 12	A (directed) graph with set of nodes $\mathcal{V}$ and set of arcs $\mathcal{E}$. Elements $\mathcal{V}$ are also in one-to-one correspondence with individuals of a social network
n	p. 12	The number of individuals in a social network ($n = \lvert\mathcal{V}\rvert$).
$\mathcal{G}[\mathbf{M}]$	p. 13	The graph corresponding to matrix $\mathbf{M}$.
$\mathbf{Adj}$	p. 12	Adjacency binary (0/1) matrix of a given graph.
$\mathbf{W} = (w_{ij})$	p. 12	Weighted adjacency matrix of a graph. Also, the stochastic matrix of influence weights in the DeGroot and Friedkin-Johnsen model. Entry w_{ij} is a strength of individual j's influence on individual i.
$\mathbf{L}$	p. 13	Laplacian matrix of the weighted graph.
$\mathbb{1}$	p. 13	The column vector of n ones.
$\mathcal{N}_i$	p. 13	Neighborhood of node i (the set of nodes to which j is connected).
in-deg(i), out-deg(i)	p. 13	In- and out-degrees of node i.
$\rho(\mathbf{A})$	p. 16	The spectral radius of matrix $\mathbf{A}$ (for non-negative matrices, also the largest in modulus eigenvalue).
$\lVert\mathbf{x}\rVert_0$	p. 20	The number of nonzero elements in vector $\mathbf{x}$ (same notation applies also to matrices).
$\mathbf{W}^{(\ell)}$	p. 23	In multidimensional (multiplex) networks and models of reflected appraisal dynamics, the matrix of influence weights during the discussion on ℓ-th topic.
$\perp\!\!\!\perp$	p. 26	The symbol of statistical independence between random variables
$p(x\lvert\mathbf{\Theta})$	p. 26	Family of distributions with a parameter matrix $\mathbf{\Theta}$
$\mathbf{\Sigma}$	p. 27	Covariance matrix
$\mathbf{\Theta}$	p. 26	Inverse covariance or Precision matrix
$\mathbf{x}_i(k) = (x_i^{(1)}, \ldots, x_i^{(m)})$	p. 33	The multidimensional opinion of individual i, consisting of the individual's positions on m different topics (issues). Evolves over discrete time $k = 0, 1, \ldots$.
$\mathbf{X}(k) = \begin{bmatrix} \mathbf{x}_1(k) \\ \vdots \\ \mathbf{x}_n(k) \end{bmatrix}$	p. 34	The matrix of individual opinions at time k.
$\mathbf{p}$	p. 34	The vector of social power in the French-DeGroot model
$\mathbf{\Lambda} = \mathrm{diag}(\lambda_1, \ldots, \lambda_n)$	p. 38	In the Friedkin-Johnsen model, the diagonal matrix of individual susceptibilities to social influence
$\boldsymbol{I}_n$	p. 38	Identity $n \times n$ matrix
$\mathbf{V}$	p. 39	The row-stochastic "control matrix", which determines the outcome of opinion formation process in the Friedkin-Johnsen model
$\mathbf{c}$	p. 40	The vector of Friedkin's influence centrality
$\mathbf{\Gamma}(k)$, $\mathbf{B}(k)$	p. 45	Random matrices, describing the randomized gossip-based opinion dynamics
$\bar{\mathbf{\Gamma}}$, $\bar{\mathbf{b}}$	p. 45	In the gossip-based opinion dynamics, the expectations of matrix $\mathbf{\Gamma}(k)$ and vector $\mathbf{B}(k)\mathbf{x}(0)$.
$\mathbf{P}(k) = \mathrm{diag}(\mathbf{p}(k))$	p. 58	In the models with random opinion measurement, the (random) measurement matrix. The vector $\mathbf{p}(k) \in \{0,1\}^n$ is the random selection vector.
$\lVert\mathbf{x}\rVert_1$	p. 54	ℓ_1 norm of vector $\mathbf{x}$.
$\boldsymbol{S}$	p. 63	Sample covariance matrix

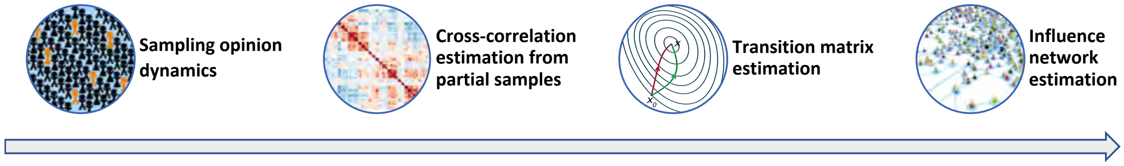


Figure 12: Main stream of the methodology.

.