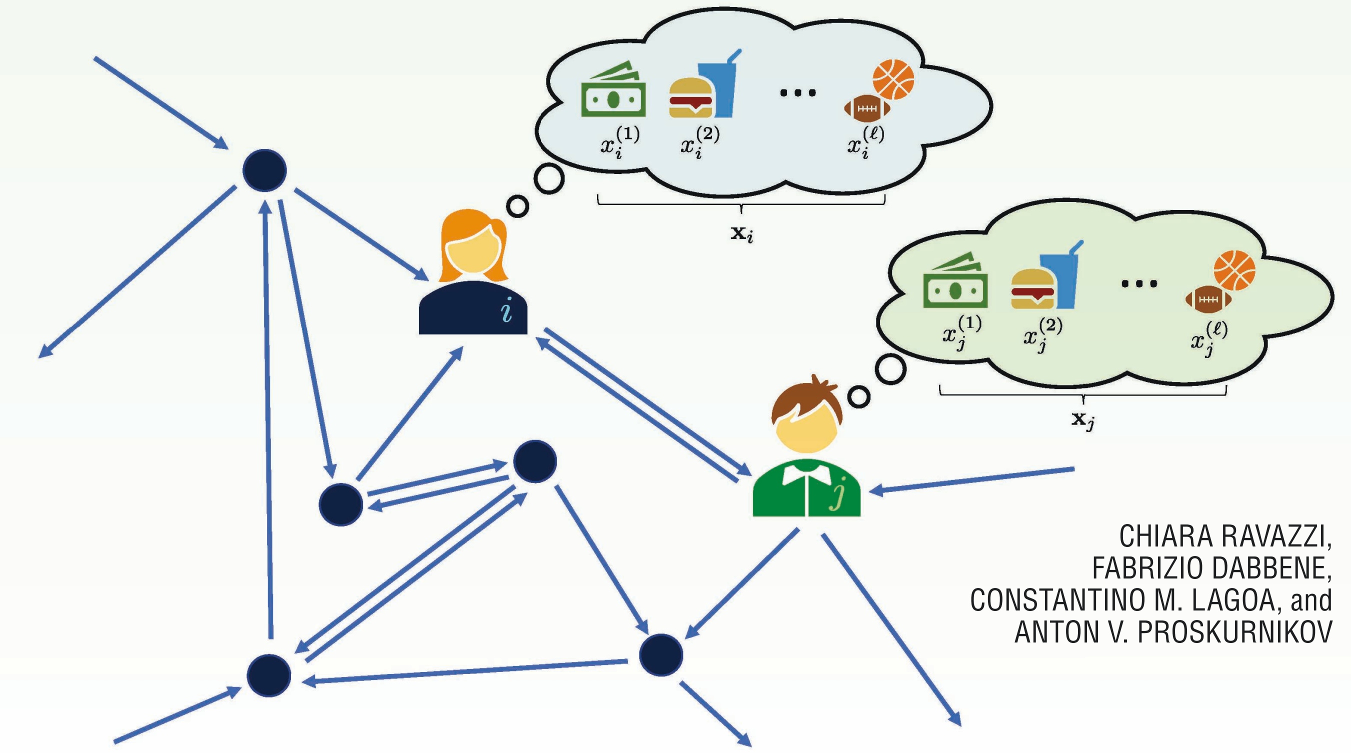
Learning Hidden Influences in Large-Scale Dynamical Social Networks

A DATA-DRIVEN SPARSITY-BASED APPROACH, *IN MEMORY OF ROBERTO TEMPO*



CHIARA RAVAZZI, FABRIZIO DABBENE, CONSTANTINO M. LAGOA, and ANTON V. PROSKURNIKOV

The processes of information diffusion across social networks (for example, the spread of opinions and the formation of beliefs) are attracting substantial interest in disciplines ranging from behavioral sciences to mathematics and engineering (see "Summary"). 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 proliferation of social media platforms that provide instant messaging, blogging, and other networking services (see "Online Social Networks") people can easily share news, opinions, and preferences. Information can reach a broad audience much faster than before, and opinion mining and sentiment analysis are becoming key challenges 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 U.S. presidential election [3]. More recently, several news outlets stated that Facebook users played a major role in spreading fake news that might have influenced the outcome of the 2016 U.S. presidential election [4]. This can be explained by the phenomena of homophily and biased assimilation [5]–[7] in social networks, which correspond to the tendency of people to follow the behaviors of their friends and establish relationships 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 on decision making and action planning (see Figure 1). According to a report by McKinsey & Company [8], "Marketing-induced consumer-to-consumer

Digital Object Identifier 10.1109/MCS.2021.3092810
Date of current version: 16 September 2021

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, advertising, 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, organizations, communities, and society in general). The advent of the Internet has made a huge volume of data easily available to measure social influence across large populations. The aim of this work is to qualitatively and quantitatively infer social influence from data, from a *systems and control viewpoint*. First, definitions and models of opinions dynamics are introduced, and structural constraints of online social networks are considered based on the notion of sparsity. Then, the main approaches to infer a 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 sample complexity and computational requirements.

word of mouth generates more than twice the sales of paid advertising." Influence analysis is becoming a key input for sophisticated recommendation engines that identify potential customers, exploiting similarities among several users to predict preferences. The same report [8] estimates that 35% of Amazon's revenue and 75% of what users watch on Netflix come from product recommendations. The study of structures in networks (such as community detection and computing a 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]. In 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 the references therein).

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 well studied. This article takes a step toward 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.

As summarized in "Opinion Dynamics Over Networks, 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 social interactions. To build a mathematical model, one must define 1) the interaction protocol (for example, the times of interactions, which can be discrete or continuous), the contact modes (deterministic and random), and the frequency of interactions among social network members and 2) the dynamical mechanism of social interactions (or ties), which can be described by linear and nonlinear functions [20]–[22]. In the simplest situation, each social tie is described by a single scalar, treated as the "influence weight" one individual assigns to another [14]. The *analysis* of social networks usually focuses on the study of the qualitative and quantitative properties of opinion dynamics (such as asymptotic convergence and oscillations, eventual consensus, and disagreement). It is also important to extract low-dimensional features of a network, for example, to identify communities and the most influential

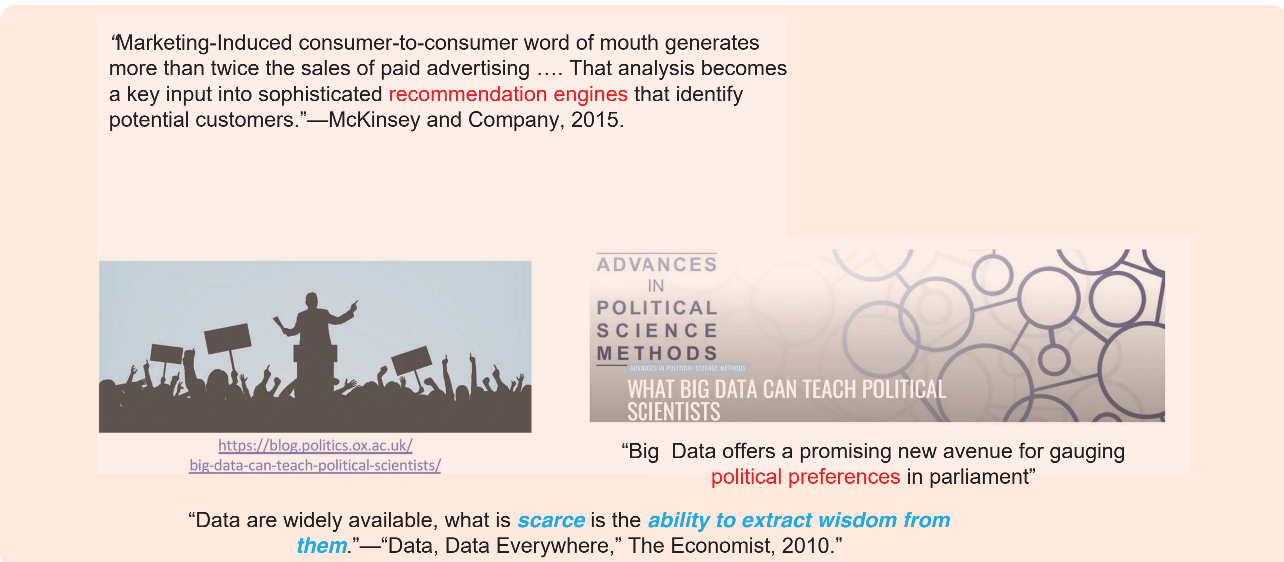


FIGURE 1 The data-driven analysis of social systems.

Online Social Networks

Online social networks (OSNs), or technosocial networks, are groups of individuals and 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 [S1]. The study of OSNs has increasingly attracted the attention of the scientific community. In fact, online services, such as Facebook, Twitter, and Instagram, play an increasingly important role in the dissemination of opinions and the emergence of certain behaviors. They facilitate

(*Continued*)

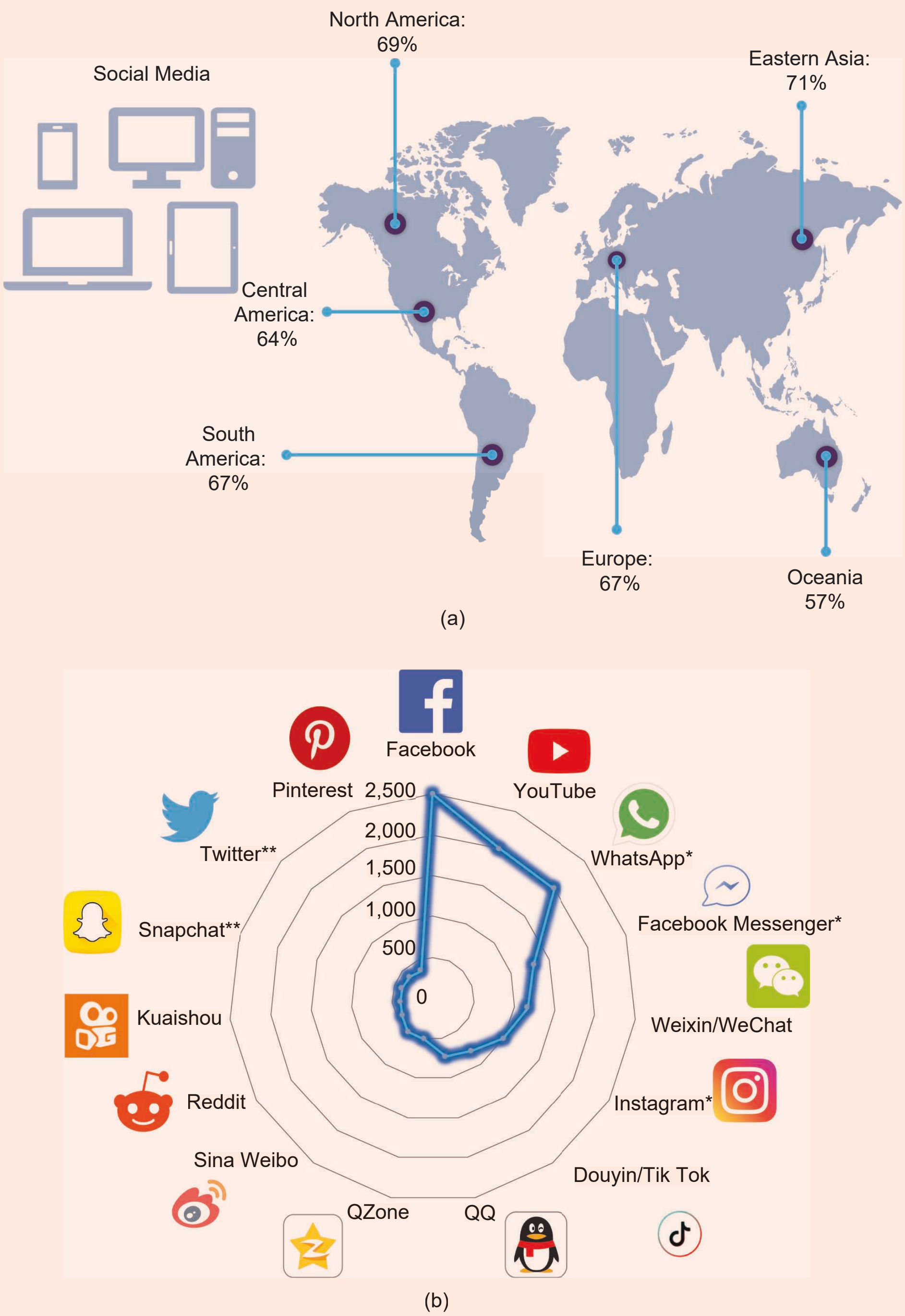


FIGURE S1 (a) The global penetration rate of online social networks (OSNs) (45%), broken down by region. (b) The number of active users of the most popular OSNs, in millions. (Source of logos: Freepik.com.)

Online Social Networks (*Continued*)

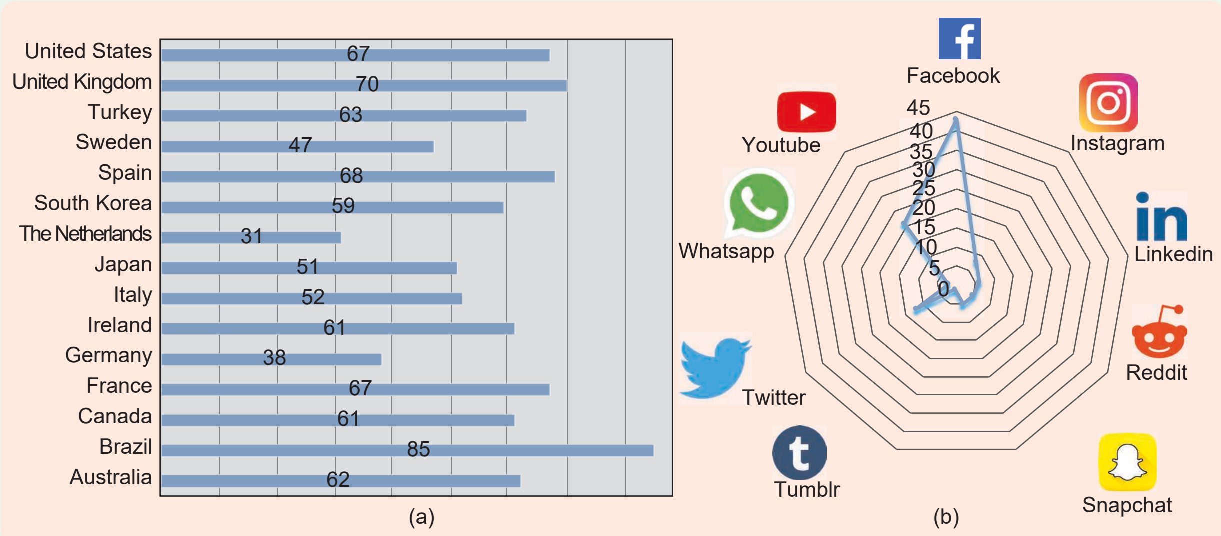


FIGURE S2 Online social networks (OSNs) and fake news. (a) The percentage of people concerned about what was real and fake on the Internet when it came to news in 2019. (b) OSN pathways to news. (Source of logos: Freepik.com.)

social interactions, helping individuals to find other people with common interests, establish a forum for discussion, and exchange information [S2], [S3]. A 2020 special digital report [S4] stated 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 OSNs in 2019 reached 70% in East Asia and North America, followed by Northern 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 use 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% jump in the time users spent on social media since the beginning of the pandemic [S5]. Figure S1 shows the total number of active users of the most popular social media networks [S6].

leaders. A long-standing goal in the study of social networks is to *control* the final distribution of opinions [23]–[26].

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 an understanding of social processes and interactions. The workflow envisaged for a systematic study of opinion formation and networks' structural properties is shown in "Data-Driven Systems and Control Approach." All research [including data collection, the design of a mathematical framework (modeling, analysis, and control), the development of algorithms, and design of large-scale experiments] must proceed in parallel with continuous interactions among blocks. Data collection and processing constitute the backbone of computational social science [27] 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 end, several methods for sensing opinions based on sentiment analysis have been proposed in recent years [28]. Efficient procedures for sampling graph signals require only a few nodes in a network to be directly observed and sensed [29], [30] and remove irrelevant information to improve the performance of processing. The analysis must identify the best models that accurately characterize a social system and select the evaluation metrics to quantify the interpersonal influence in a network. Efficient algorithms and new control mechanisms of centrality measures (see "Centrality Measures in Weighted Graphs") 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 dynamic models, then predict and control opinion diffusion across a 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.

Social ties among individuals have to be quantified to validate and examine the models. In a small group participating in a roundtable discussion, individuals can estimate the

A downside of this high social penetration rate is the spread of *fake news*, whose detection is becoming a serious problem [S7], [S8]. Figure S2 describes the percentage of people who are concerned about what is real and fake on the Internet regarding news. The data are current to 2019 [S9]. Moreover, approximately 43% of adults in the United States get news from Facebook (according to a survey conducted in July and August 2018 [S10]). 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 [S11]. It has been shown [S12] 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, the average clustering coefficient, and the diameter of the largest weakly connected components) can effectively be used to identify misleading and harmful information [S12]. Inferring the networks' structure and learning global properties from partial information become central tasks [S13], [S14].

REFERENCES

[S1] M. Cruz-Cunha and F. Moreira, *Handbook of Research on Mobility and Computing: Evolving Technologies and Ubiquitous Impacts*. Hershey, PA: IGI Global, 2011. [Online]. Available: https://books.google.com.ag/books?id=2koyMAEACAAJ

[S2] C. Cheung, P.-Y. Chiu, and M. Lee, "Online social networks: Why do students use Facebook?," *Comput. Human Behavior*, vol. 27, no. 4, pp. 1337–1343, July 2011. doi: 10.1016/j.chb.2010.07.028.

[S3] 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. Roslindale, MA: Turbulence, 2009.

[S4] 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

[S5] K. Majewska, "Stats on social media use in COVID-19 hotspots: February vs. March 2020," napoleoncat.com/blog, July 2019. [Online]. Available: https://napoleoncat.com/blog/stats-on-social-media-use-covid-19-hotspots/

[S6] A. Rasool, "Facebook, Instagram, TikTok. Which apps are the most downloaded in July by Apple and Android users?" Digitalinformationworld.com, July 2019. [Online]. Available: https://www.digitalinformationworld.com/2019/08/top-10-social-media-apps-worldwide-july-2019-downloads.html

[S7] S. Vosoughi, D. Roy, and S. Aral, "The spread of true and false news online," *Science*, vol. 359, no. 6380, pp. 1146–1151, Mar. 2018. doi: 10.1126/science.aap9559.

[S8] A. Bovet and H. A. Makse, "Influence of fake news in Twitter during the 2016 US presidential election," *Nature Commun.*, vol. 10, no. 1, p. 7, 2019. doi: 10.1038/s41467-018-07761-2.

[S9] N. McCarthy, "Where concern is highest about fake news on the Internet," *Statista.com*, June 2019.

[S10] J. Gramlich, "10 facts about Americans and Facebook," *Report Pew Research Center*, May 2019.

[S11] M. Del Vicario et al., "The spreading of misinformation online," in *Proc. Nat. Academy Sci.*, 2016. [Online]. Available: https://www.pnas.org/content/early/2016/01/02/1517441113

[S12] F. Pierri, C. Piccardi, and S. Ceri, "Topology comparison of Twitter diffusion networks effectively reveals misleading information," *Sci. Rep.*, vol. 10, no. 1, 2020. doi: 10.1038/s41598-020-58166-5.

[S13] M. Gomez Rodriguez, J. Leskovec, and A. Krause, "Inferring networks of diffusion and influence," in *Proc. 16th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining (KDD '10)*. New York: ACM, 2010, pp. 1019–1028. doi: 10.1145/1835804.1835933.

[S14] M. Gomez Rodriguez, J. Leskovec, D. Balduzzi, and B. Schölkopf, "Uncovering the structure and temporal dynamics of information propagation," *Netw. Sci.*, vol. 2, no. 1, pp. 26–65, 2014. doi: 10.1017/nws.2014.3.

influence they and the others have on the formation of their opinions (see "Friedkin–Johnsen Experiment"). However, this approach is inapplicable to large-scale groups and online social networks, whose structures of influence relations can be inferred only from data. The rapid development of the Internet, on the one hand, makes a large volume of data easily available for analysis. On the other hand, it poses new challenges. Data sizes are getting larger, and collected information becomes heterogeneous and more complex. The massive data in OSNs consist of linked information, mainly in the form of graphic structures, describing communications between any two entities (including 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 of the two main aspects of interpersonal influence estimation: 1) the social network sensing problem and 2) network reconstruction algorithms, with a particular focus on sample complexity and computational requirements. The main challenge is to guarantee the efficiency and scalability of the algorithms in the face of big data produced by OSNs. 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 [33], compressed sensing [34], and graph analytics [35], and encompassing various fields of research, including the following:

1) graph theory and linear algebra
2) control theory techniques, such as 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 the sampling and reconstruction of graph signals.

The main focus of the article is based on previous works [36]–[43]. The interested reader is referred to the literature for more insight. This survey was partly presented during the tutorial section "Control and Learning for Social Sciences: Dynamical Networks of Social Influence" at the 2020 IFAC World Congress. The remainder of this article is organized as follows. The main body of the text is complemented by

Opinion Dynamics Over Networks, in a Nutshell

Certain assumptions are involved, including the following:

- First, a population of individuals (or actors) forms.
- Individuals interact and exchange opinions.
- As a result, their opinions evolve.

There are three research directions, as described in the following:

1) *Mathematical modeling of social behaviors*:
   - *Interaction protocol*: discrete/continuous-time, deterministic/random contacts
   - *Social ties*: linear (parameterized by scalar "weights") and nonlinear functions
2) *Analysis (sociology plus computer science)*:
   - the evolution of opinions (convergence/oscillations and consensus/disagreement)
   - low-dimensional features (communities and 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 an individual's opinions to a desired value).

"summary" sidebars ("Opinion DynamicsOver Networks, in a Nutshell," "Data-Driven Systems and Control Approach," "A Glossary of Graphs," "Centrality Measures in Weighted Graphs," "Sparsity Structure in Online Social Networks," "Sparse Models," "French's Original Model," "DeGroot's Model as Dynamics Over aGraph," "Simple Properties of the Friedkin–Johnsen Model," "Control Matrix and Friedkin's Centrality," "Reflected Appraisal Model," "Asynchronous Gossip-Based Friedkin–Johnsen Model," "Dynamics of Multiplex Networks," "Necessary and Sufficient Conditions for Recovery," "Performance of Influence Estimation: Asynchronous Gossip-Based Friedkin–Johnsen Model," and "Bayesian Estimation of $\mathbf{S}_{[0]}^{(s)}(\infty)$") that aim to improve understanding through concise and schematic blocks and "focus" sidebars ("Online Social Networks," "Friedkin–Johnsen Experiment," "Degree Distribution in Facebook Ego Networks," "Multidimensional Networks," "Gaussian Graphical Models and the Graphical Lasso," "Graph Signal Processing," "Schur Stability Criteria," "A Model of a Belief System's Dynamics," and "Compressed Sensing") that address specific topics through further study, technical results, and additional discussions. The reader can skip these parts without a loss of comprehension.

DEFINING INFLUENCE IN SOCIAL NETWORKS

As defined in [44] and [45], interpersonal influence is a "causal effect of one actor on another," such as a change in opinions

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 and that is able to predict the processes of opinion formation and information spread in large-scale social networks, providing well-grounded tools to quantify and control the impact of specific actions. The workflow for a systematic study of network structures and dynamics appears in Figure S3.

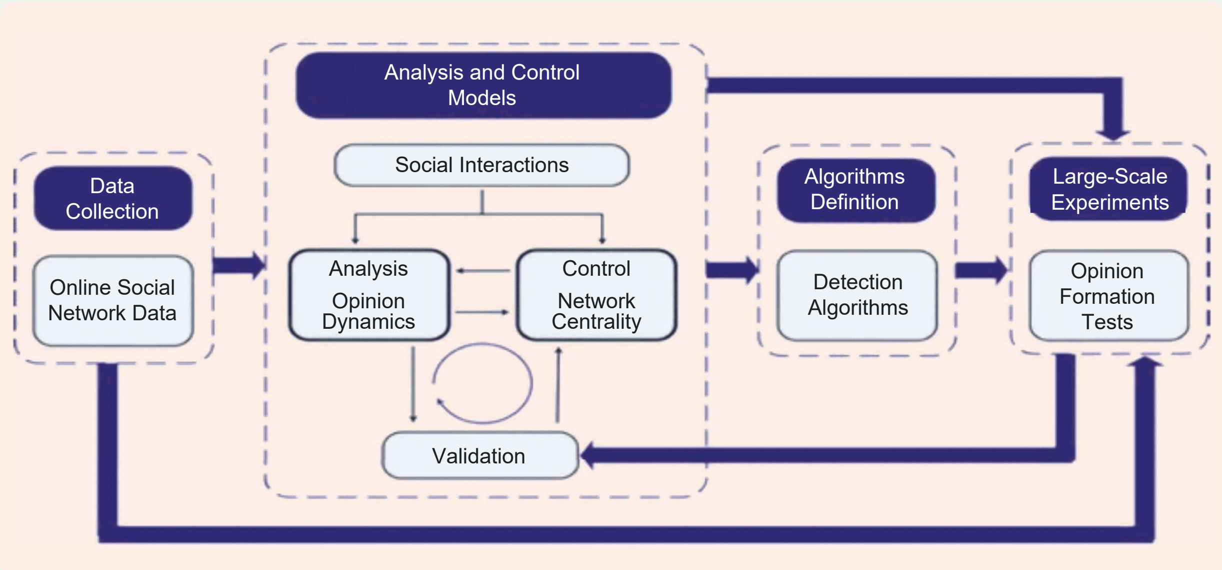


FIGURE S3 The workflow for a systematic study of network structures and dynamics.

and behaviors [46]. The quantification and measurement of social ties are long-standing problems that have been studied since the 1950s [47]–[50]. 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" [44]. Another problem is the coevolution of social ties and individuals' behaviors. On the one hand, people modify their behaviors to align with those of their friends (social influence). However, people tend to form friendships with others like themselves (social selection). Opinions and other mutable characteristics are thus formed by the *interplay* between social selection and influence [12].

Research into interpersonal influence exists in the literature, and three main directions prevail. The first develops the seminal ideas of Granovetter [51], defining the strength of a social tie between two individuals as a function of *positions* in a social group. For instance, the more friends that actors A and B have in common, the stronger is the tie among them [46], [51]. Social influence introduced in this way depends only on the structure of a network. A large amount of available data from real-world social networks and the existence of efficient tools for the information's analysis make this approach very attractive for both behavioral and computer sciences. The second line of research relates social influence to temporal (dynamical) mechanisms, modifying some numerical attributes of social actors, such as opinions and 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 and A's level of trust in B's opinions. This idea is elaborated in the Friedkin–Johnsen theory of social influence [14], [31], [32], [44].

The fundamental results reported in [44] establish interrelations between 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 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 Figure 2 and the section "Social Influence in Opinion Dynamics") whose parameters can be constructed via a structural analysis. The third research direction related to complex networks (not necessarily social ones) is *statistical* (learning-based) methods of network reconstruction. Like the second approach, it assumes that actors at a network's nodes are endowed with numerical values that are supposed to be

Friedkin–Johnsen Experiment

A seminal example of coupling theory with empirical research can be found in [32], where the Friedkin–Johnsen model [31] for single-issue opinion dynamics is validated for small and medium groups of individuals (social actors). An experiment was designed using 30 groups, four individuals, and 15 issues (risk choice dilemmas), and it included the following:

- recording actors' initial opinions
- discussions lasting 15 min
- actors distributing "chips" to quantify influence
- recording actors' final opinions.

FIGURE S4 Influence is experimentally measured by asking participants to distribute chips to measure the influence weights that each person has assigned to him/herself and others during the decision-making process. (Source: Freepik.com.)

The participants were asked to express their opinions about 15 issues selected uniformly at random without replacement from a set of risk choice dilemmas. Risk choice dilemmas are hypothetical life decisions that are used to measure willingness to assume a risk. More precisely, the agents were asked to express their minimum level of confidence (that is, a scalar value in the range [0, 1]) to accept a risky option with a high payoff over a less risky option with a low payoff. The individuals recorded their initial opinions on an issue; then, a 15-min discussion was opened, and the actors' final opinions were recorded.

- 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).

INFLUENCE DETERMINATION

To estimate social influence, participants were asked during the discussion to distribute chips (Figure S4) among themselves while they interacted, as a subjective measure of the influence exercised by other group members. Model validation was then performed by showing that predicted opinions were close to those recorded.

- 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.

random. Unlike the second technique, the existence of a temporal mechanism modifying values is not stipulated. A tie between two nodes corresponds to *statistical correlation* among 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* [52], [53] and analyzed by methods of statistics and statistical learning theory (see the section "Learning Graphs From Data").

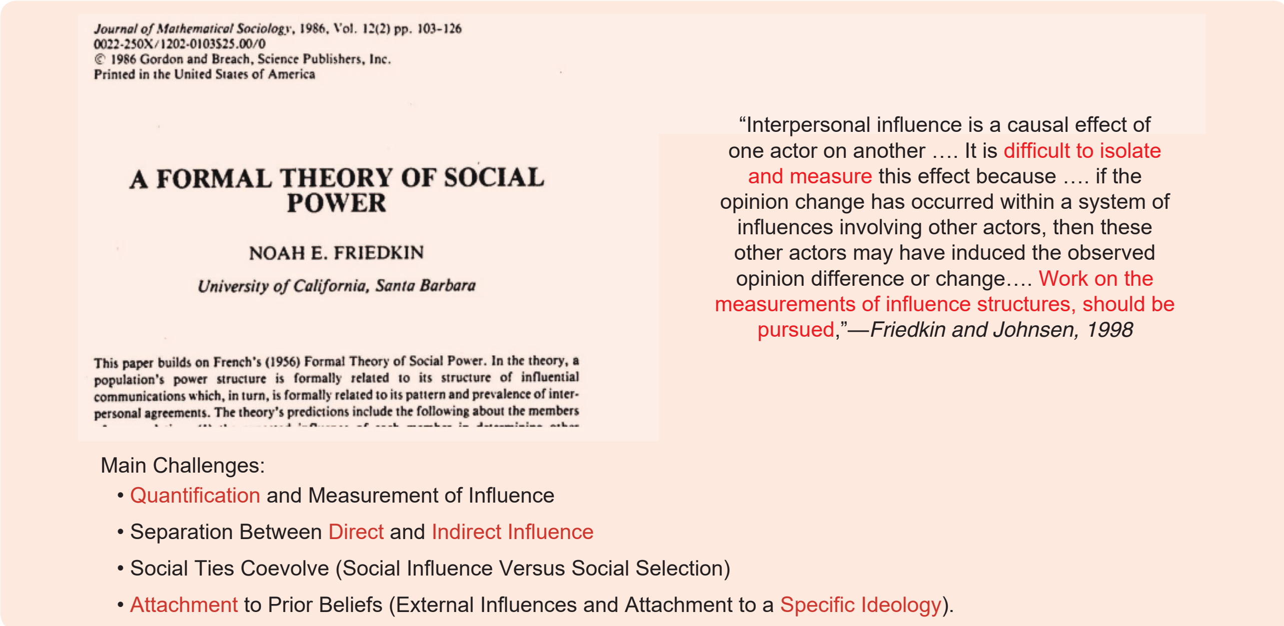


FIGURE 2 Measuring influence, affinities, and social ties.

A Glossary of Graphs

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

- $\mathcal{V}$ is the set of nodes (corresponding, for example, to agents in a network), indexed as $1, \ldots, n$.
- $\mathcal{E} \subseteq \mathcal{V} \times \mathcal{V}$ is a set of ordered pairs of nodes describing a relationship: 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 **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 S5 represents a graph.

A matrix **M** is said to be adapted to the graph $\mathcal{G}$ if $\mathcal{G}[\mathbf{M}] = \mathcal{G}$. By construction, the adjacency matrix **Adj** and every influence matrix **W** of a graph $\mathcal{G}$ are adapted to $\mathcal{G}$. The matrix **W** is said to be row stochastic if its rows sum to one; that is, $\Sigma_i w_{ij} = 1$. In compact form, $\mathbf{W}\mathbb{1} = \mathbb{1}$, with $\mathbb{1} \doteq [1 \cdots 1]^\top$. Similarly, matrix **W** is column stochastic if $\Sigma_j w_{ij} = 1$; 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: $\mathbf{W} = \mathbf{W}^\top$.

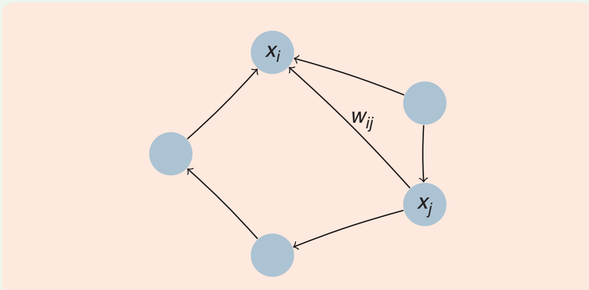


FIGURE S5 A representation of a graph.

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 $d_i = \Sigma_j w_{ij}$. 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 different nodes.

This article focuses on the second and third lines of research. Certain sections consider the statistical estimation of social influence (the third direction). Sections throughout address the identification of dynamic mechanisms of opinion formation, namely, the Friedkin–Johnsen model. Both approaches use the construction of a *weighted directed graph* whose nodes have numerical attributes (considered to be opinions of social actors), where 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 weighted graphs.

Influence-Related Measures

A social network consists of two main components: 1) social actors (individuals and organizations) and 2) dependency, influence, and similarity relations. Each actor has a numerical attribute (representing an opinion). A social network can be mathematically described by a directed weighted graph (see "A Glossary of Graphs" for definitions). At the local level, social influence is a *directional effect from node i to node j,*

TABLE 1 The main notation symbols.

Symbol	Meaning
$\mathcal{G} = (\mathcal{V}, \mathcal{E})$	A (directed) graph with a set of nodes $\mathcal{V}$ and a set of arcs $\mathcal{E}$; elements $\mathcal{V}$ are in one-to-one correspondence with individuals of a social network
n	The number of individuals in a social network ($n = \lvert \mathcal{V} \rvert$)
$\mathcal{G}[\mathbf{M}]$	The graph corresponding to matrix $\mathbf{M}$
Adj	Adjacency binary (0/1) matrix of a given graph
$\mathbf{W} = (w_{ij})$	Weighted adjacency matrix of a graph as well as the stochastic matrix of the influence weights in the DeGroot and Friedkin–Johnsen model; w_{ij} is a strength of individual j's influence on individual i.
$\mathbf{L}$	Laplacian matrix of the weighted graph
$\mathbb{1}$	The column vector of n ones
$\mathcal{N}_i$	Neighborhood of node i (the set of nodes to which j is connected)
in-deg(i), out-deg(i)	In- and out-degrees of node i
$\rho(\mathbf{A})$	Spectral radius of matrix $\mathbf{A}$ (for nonnegative matrices, also the largest in modulus eigenvalue)
$\lVert \mathbf{x} \rVert_0$	Number of nonzero elements in vector $\mathbf{x}$ (same notation applies also to matrices)
$\mathbf{W}^{(\ell)}$	In multidimensional (multiplex) networks and models of reflected appraisal dynamics, the matrix of influence weights during the discussion on ℓth topic
$\perp\!\!\!\perp$	Symbol of statistical independence among random variables
$p(x \mid \Theta)$	Family of distributions with a parameter matrix Θ
$\mathbf{\Sigma}$	Covariance matrix
$\mathbf{\Theta}$	Inverse covariance or precision matrix
$\mathbf{x}_i(k) = (x_i^{(1)}, \ldots, x_i^{(m)})$	Multidimensional opinion of individual i, consisting of the individual's positions on m different topics (issues); evolves in discrete time $k = 0, 1, \ldots$
$\mathbf{X}(k) = \begin{bmatrix} \mathbf{x}_1(k) \\ \vdots \\ \mathbf{x}_n(k) \end{bmatrix}$	Matrix of individual opinions at time k
$\mathbf{p}$	Vector of social power in the French–DeGroot model
$\mathbf{\Lambda} = \mathrm{diag}(\lambda_1, \ldots, \lambda_n)$	In the Friedkin–Johnsen model, the diagonal matrix of individual susceptibilities to social influence
$\mathbf{I}_n$	Identity $n \times n$ matrix
$\mathbf{V}$	Row-stochastic "control matrix," which determines the outcome of opinion formation process in the Friedkin–Johnsen model
$\mathbf{c}$	Vector of Friedkin's influence centrality
$\mathbf{\Gamma}(k), \mathbf{B}(k)$	Random matrices describing randomized gossip-based opinion dynamics
$\bar{\mathbf{\Gamma}}, \bar{\mathbf{b}}$	In gossip-based opinion dynamics, the expectations of matrix $\mathbf{\Gamma}(k)$ and vector $\mathbf{B}(k)\mathbf{x}(0)$
$\mathbf{P}(k) = \mathrm{diag}(\mathbf{p}(k))$	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.
$\lVert \mathbf{x} \rVert_1$	ℓ_1 norm of vector $\mathbf{x}$
$\mathbf{S}$	Sample covariance matrix

which is related to the edge strength $(i, j) \in \mathcal{E}$ [46]. To comply with previously published works on opinion dynamics [14], [42] 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] (see Table 1 for notations used in this article). In 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 network interconnections. Several global measures have been introduced to identify the most relevant entities in a network. These global measures can refer to nodes and 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 of centrality are defined in "Centrality Measures in Weighted Graphs."

The simplest and most popular definition of centrality is *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

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 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 illustrated using a simple directed network known as the *football data set* [S15]. The network recorded 35 soccer teams that participated in the 1998 World Championship, in Paris, France. Every edge recorded the number of national team players of one country who played in the league of another country.

DEGREE CENTRALITY

The in/out-degrees of $i \in \mathcal{V}$ (Figure S6) 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 one starting from a node. The weighted in/out-degree is defined as the sum of weights when analyzing weighted networks:

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

CLOSENESS CENTRALITY

The closeness centrality (Figure S7) 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 [56] and [57]. 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.

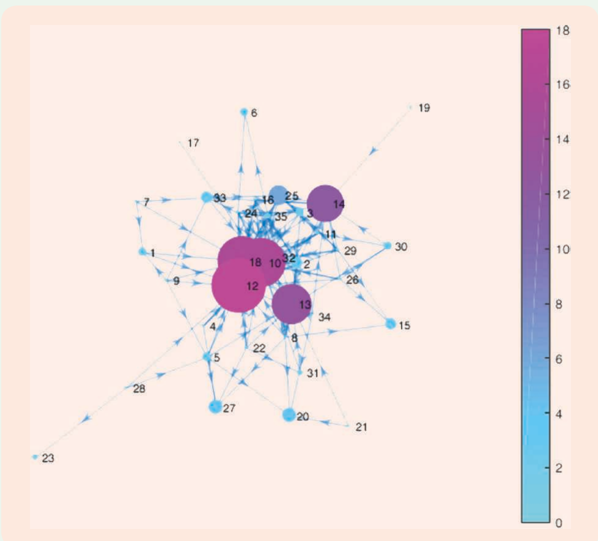


FIGURE S6 Nodes colored according to in-degree centrality.

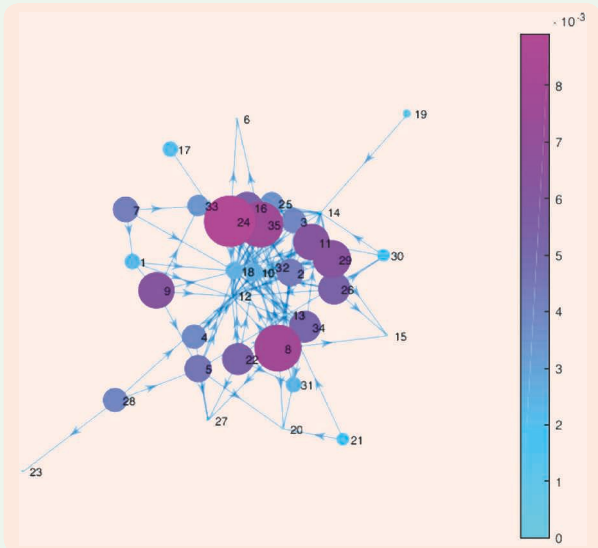


FIGURE S7 Nodes colored according to closeness centrality.

(out-degree) information. A more general concept is *K-path centrality* [54], defined as the number of paths of length K starting from a node. Both the degree and K-path centrality definitions are local. Alternative centrality measures have been considered to measure the importance of a node for a graph as a whole. Among them, closeness, betweenness, and eigenvector centrality are briefly discussed. Closeness centrality is a measure of how near a node is to most of the other ones [55] and provides insight into how long it will take to spread information from i to all other nodes in the network.

Another relevant measure is represented by node betweenness [56], [57]. Nodes with high betweenness occupy critical positions in a network and are bridges between two groups of vertices (since many paths in different groups must pass through them). The eigenvector centrality of a node is a function of the node's neighbors, and the relevance is assigned according to the entries of the leading eigenvector $x^\star$ of a suitable weighted adjacency matrix of a network. Contrary to degree centrality, this notion does not depend on the number of neighbors but considers the relevance of neighbors. In this way, a node with a few influential neighbors

BETWEENNESS CENTRALITY

The betweenness (Figure S8) 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 $\mathcal{S}(j,k)$ 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 $\mathcal{S}(j,k)$ is measured through the entries of the influence matrix $\mathbf{W}$.

EIGENVALUE CENTRALITY

The idea of *eigenvector centrality* (Figure S9) is based on a simple principle: a node is important if it is connected to other important ones. This centrality measure is determined by the dominating (Perron–Frobenius) eigenvector $\mathbf{x}^\star$ of some properly defined nonnegative matrix $\mathbf{A}$ that is compatible with the graph. Formally,

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

where $\lambda = \rho(\mathbf{A})$ is the maximal positive eigenvalue (being also the spectral radius) of $\mathbf{A}$. In the prevailing 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 [S16].

REFERENCES

[S15] "The football dataset." [Online] Available: http://www.casos.cs.cmu.edu/computational_tools/datasets/external/football/index2.html

[S16] A. Z. Broder, R. Lempel, F. Maghoul, and J. Pedersen, "Efficient pagerank approximation via graph aggregation," *Inform. Retrieval*, vol. 9, no. 2, pp. 123–138, Mar. 2006. doi: 10.1007/s10791-006-7146-1.

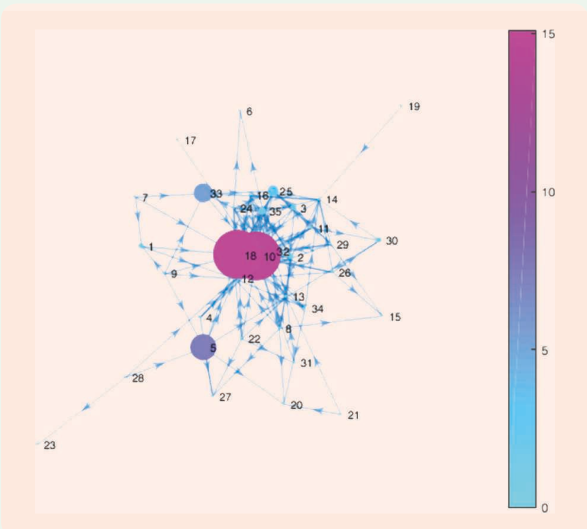


FIGURE S8 Nodes colored according to betweenness centrality.

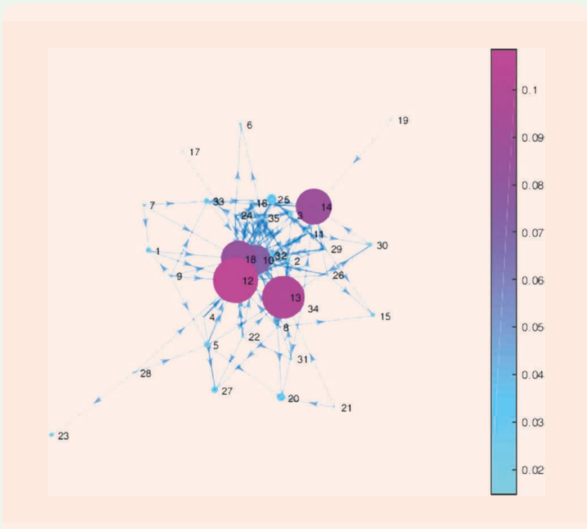


FIGURE S9 Nodes colored according to eigenvalue centrality.

has larger eigenvector centrality than a node with various neighbors of limited influence. The most famous eigenvalue centrality measure is PageRank centrality [58], which was introduced in the context of webpage ranking. Many other centrality measures (such as Katz centrality [59], Bonacich centrality [60], [61], and harmonic influence centrality [14], [15], [62]) naturally arise as extensions of eigenvector centrality and the PageRank.

SPARSITY STRUCTURE OF SOCIAL NETWORKS

A systematic study of the structure of social networks offers several metrics and algorithms for extracting low-dimensional features. Metrics can quantify global and local structural properties. Network density is an aggregate metric defined as the ratio $|\mathcal{E}|/n^2$ of the number of observed social relationships to the number of possible relationships among nodes (that is, the proportion of ties within a network). A collection of large social network data sets is made available by the Stanford Network Analysis Platform [63] and can be visualized using the software GraphViz [65]. In "Sparsity Structure in Online Social Networks," Table S1 reports the type, number of nodes, and density of some social networks. Note that OSNs have common features, including the following:

- 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 linearly depends on the network size.

The distinction between dense and sparse graphs depends on context. The index with which sparsity is commonly measured in network graphs is edge density [65]. 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 [66] for networks. Refer to [67] for an overview of sparsity definitions adapted for networks. Another important statistical characteristic 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 was discovered in early works about sociometry [68]. It is said that the network exhibits power law distribution if the fraction of nodes with degree k is distributed as

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

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 $p_{\deg}(k)$ remains unchanged (other than a multiplicative factor) when rescaling the independent variable k [as it satisfies $p_{\deg}(\alpha k) = \alpha^{-\gamma} p_{\deg}(k)$].

In [69], 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 the *ego*) is defined as a subgraph induced across nodes that are directly connected to it but exclude 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 [63]. As shown, some are more "concentrated" around a mean value, while others show power law decay with a smaller exponent γ. As discussed in the next section, this concentration property plays a crucial role in the inference of trust network from little data.

Examining the Twitter networks of 14 destination marketing organizations, Finally, other networks exhibit the presence of few clusters [69], 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 [71]. 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 network analysis.

SPARSE MODELS FOR MULTIDIMENSIONAL NETWORKS

Multidimensional networks describe different types of relations among nodes. For example, friendships in a social network may arise for various reasons (for instance, because users are colleagues, teammates in sports, or share hobbies). In this case, consider multiplex networks, where different layers of interconnections can be distinguished that correspond to various 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 about *different topics*. The influence between the users is *topic dependent* (for example, networks are sometimes called *heterogeneous* [72]). Each layer in a multilayer network considers influence relations among people when they discuss a certain subject. The analysis of multiplex networks is an active field of research (see [73] and the references therein). If an underlying social network is composed of the same individuals, then it is expected that social systems share a common feature.

The preceding intuition entails that networks describing the microlevel 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 networks relative to different topics. In this sense, [74] introduces correlated models. The first, *the common component model* ($\mathcal{M}_{cc}$), considers cases where networks relative to different topics $\ell = 1, \dots, m$ differ only 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 an interaction network relative to the ℓth topic is decomposed as

Sparsity Structure in Online Social Networks

Sparsity structure allows one to distinguish dense networks (Figure S10) and sparse networks (Figure S11). According to the Stanford Large Network Data Set Collection, "Most real social networks are sparse." Table S1 describes the type, number of nodes, and density of some social networks.

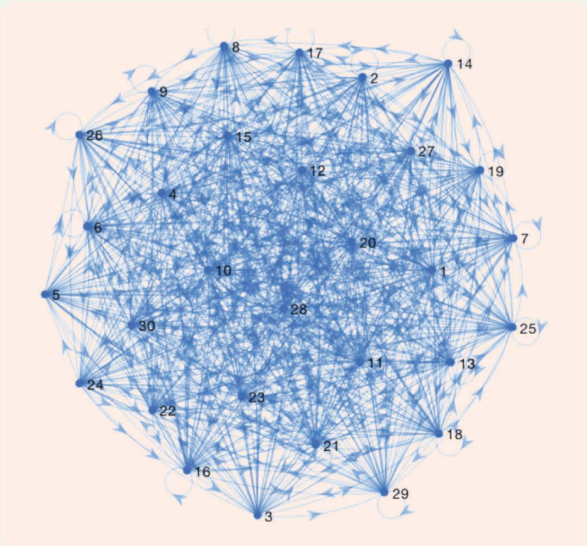


FIGURE S10 A dense network.

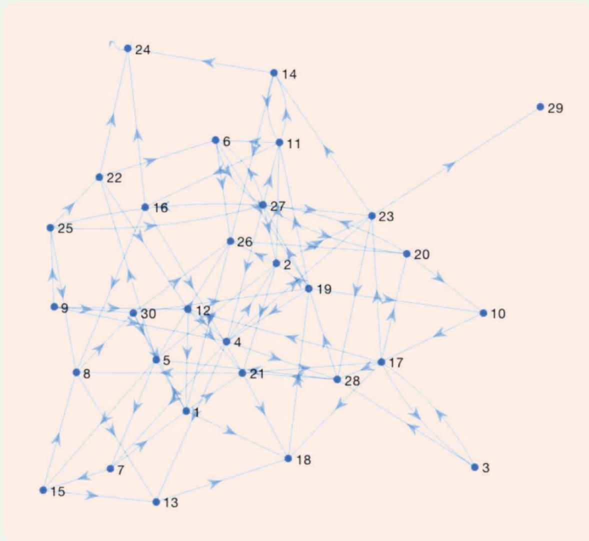


FIGURE S11 A sparse network.

TABLE S1 Social networks according to their type, number of nodes and edges, and density.

Name	Type	Nodes	Edges	Network Density
ego-Facebook	Undirected	4039	88,234	$5408 \cdot 10^{-3}$
ego-Gplus	Directed	107,614	13,673,453	$1180 \cdot 10^{-3}$
ego-Twitter	Directed	81,306	1,768,149	$2674 \cdot 10^{-4}$
soc-Epinions1	Directed	75,879	508,837	$8837 \cdot 10^{-5}$
soc-LiveJournal1	Directed	4,847,571	68,993,773	$2936 \cdot 10^{-6}$
soc-Pokec	Directed	1,632,803	30,622,564	$1148 \cdot 10^{-5}$
soc-Slashdot0922	Directed	82,168	948,464	$1404 \cdot 10^{-4}$
wiki-Vote	Directed	7115	103,689	$2048 \cdot 10^{-3}$
gemsec-deezer	Undirected	143,884	846,915	$4090 \cdot 10^{-5}$
gemsec-facebook	Undirected	134,833	1,380,293	$7592 \cdot 10^{-5}$

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

where the matrices $\overline{\mathbf{W}}$ and $\delta\mathbf{W}^{(\ell)}$ (representing, respectively, the common part and the innovation part) are both sparse (see the examples in "Multidimensional Networks"). The second model, the *common support model* ($\mathcal{M}_{cs}$), describes situations where the topology is equal for all the different topics but where the weights vary. 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,

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

An example of this model can be found in deliberative groups that address a sequence of issues, such as department faculty at universities and boards of directors in large organizations. Empirical findings show [75] that the weights evolve according to a natural social process known as *reflected appraisal* [76], [77]. The representation of this process, proposed in [75], is squarely based on the Friedkin–Johnsen model of opinion formation and will be considered in a following section.

Sparse Models

Sparse models to represent high-dimensional data have been used in several areas, such as statistics, signal and image processing, machine learning, coding, and control theory [S17]. Intuitively, data are considered sparse or compressible if they are so highly correlated that only a few degrees of freedom are significant compared to their ambient dimension. This general definition leads to many possible interpretations, and alternative measures of sparsity can be defined according to 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) is small compared to the signal dimension. Moreover, a signal $\mathbf{x} \in \mathbb{R}^n$ is k-sparse if $\|\mathbf{x}\|_0 \doteq |\{i \in \{1,\ldots,n\} \;:\; x_i \neq 0\}| \leq k$, with $k \ll n$.

In social networks analysis, sparsity can be exploited in several ways. This article provides examples to easily address ideas. If, from a sociological perspective, an agent is influenced by few friends (Figure S12), 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 (Figure S13). Here, a typical sparse adjacency matrix is depicted with the signal obtained by stacking

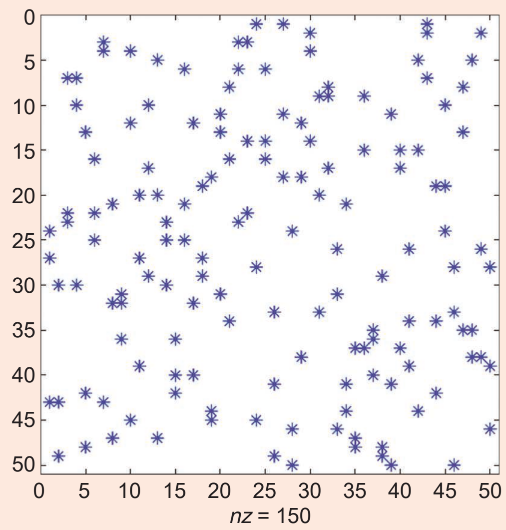


FIGURE S13 Sparsity in the elements of the adjacency matrix.

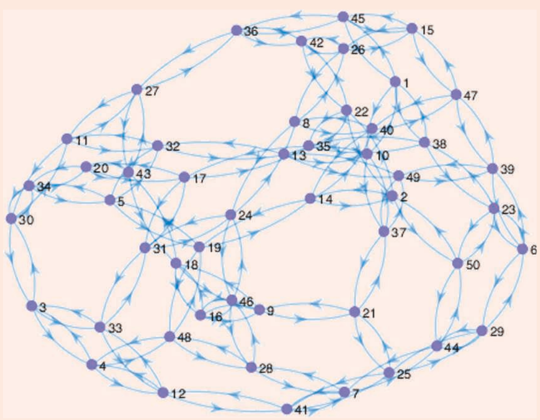


FIGURE S12 A network where each individual has few friends.

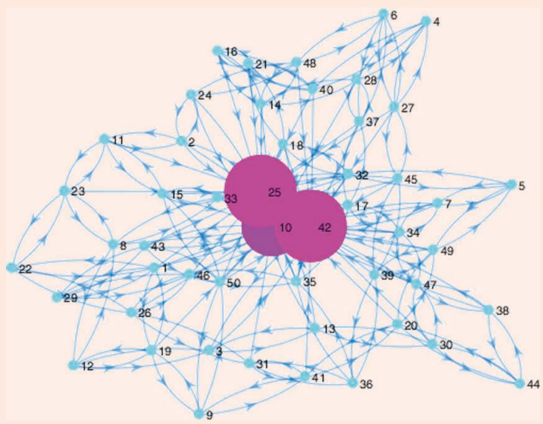


FIGURE S14 A network with few influencers.

To summarize, 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 social influence analysis is the knowledge of influence matrix **W**. The next sections focus on algorithms inferring matrix **W** and related computational aspects and consider two different approaches. The "static" method addresses the inference of matrix **W** from samples of some observables $\{x_j\}_{j \in \mathcal{V}}$, whereas the "dynamic" approach addresses the identification of opinion formation models.

LEARNING GRAPHS FROM DATA

The influence network estimation problem discussed in this article represents a special instance of the general problem of reconstructing a graph topology from data measured on nodes. This problem, known as *graph learning* or *network inference*, has attracted increasing interest in past years. Readers are referred to [78], on which this section is largely based. The literature distinguishes among several

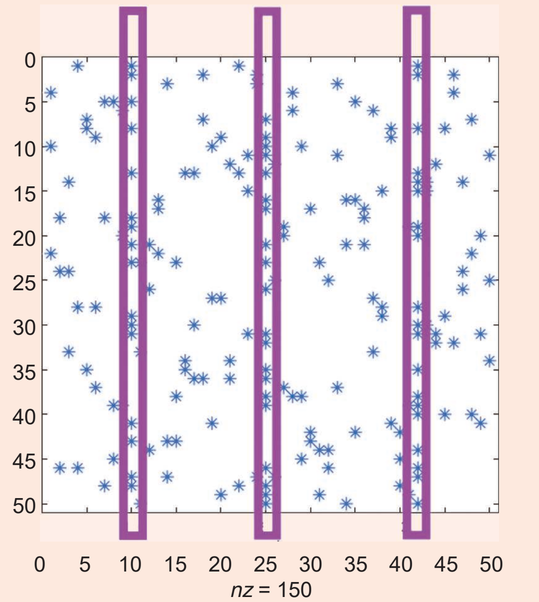


FIGURE S15 Block sparsity in the adjacency matrix.

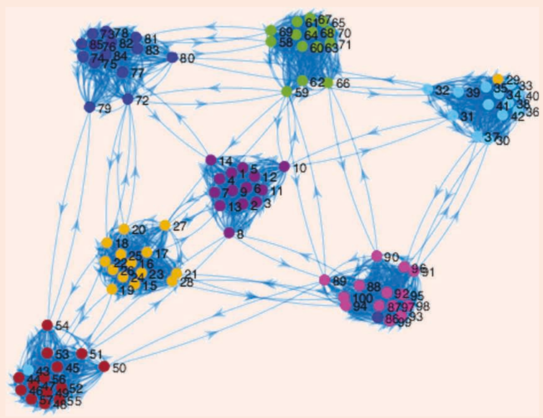


FIGURE S16 A network with few communities.

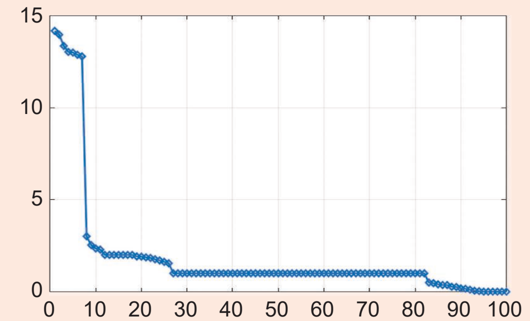


FIGURE S17 Sparsity in the eigenvalues of the adjacency matrix.

the matrix by columns. Note that only a small portion of the elements is different from zero.

If a network 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 (Figure S15). The adjacency matrix of a network with five influencers is shown in Figure S14. In Figure S15, the elements of the matrix are stacked by columns. Note that the signal is sparse with few dense patterns. In the literature, this feature is also known as *block sparsity*.

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 types of networks, the adjacency matrix can be decomposed as $\mathbf{A} = \mathbf{L} + \mathbf{S}$, where **L** is a low-rank matrix and **S** is a sparse one. Here, we show a typical adjacency matrix of a network with few communities (Figure S16) and the corresponding eigenvalues in absolute values (Figure S17). Note that the eigenvalues are highly compressible and that only four eigenvalues (even with the number of communities) contain the most energy of the signal.

REFERENCE

[S17] I. Rish, G. A. Cecchi, A. Lozano, and A. Niculescu-Mizil, *Practical Applications of Sparse Modeling*. Cambridge, MA: MIT Press, 2014.

Degree Distribution in Facebook Ego Networks

Ego network analysis represents a common tool for the investigation of relationships among individuals and their peers in online social networks [S18]. 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 [S19]. The ego network of a focal node is defined as a subgraph induced across nodes that are directly connected to it but exclude the ego itself. Figure S18 depicts the empirical degree distribution of three Facebook ego networks retrieved from the Stanford Network Database [63]. Note that some degree distributions are more concentrated around a mean value, while others show power law decay. The tails of the distribution are well approximated by (1), with $\gamma \in [1.2, 3]$.

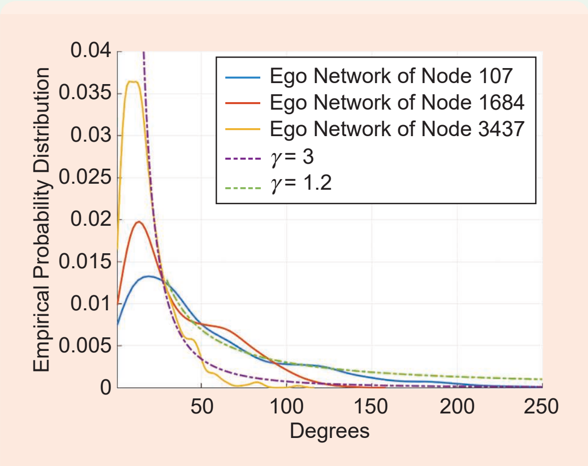


FIGURE S18 The degree distribution of Facebook ego networks.

REFERENCES

[S18] C. Boldrini, M. Toprak, M. Conti, and A. Passarella, "Twitter and the press: An ego-centred analysis," in *Companion Proc. Web Conf. (WWW '18)*, 2018, pp. 1471–1478. doi: 10.1145/3184558.3191596.

[S19] V. Arnaboldi, M. Conti, A. Passarella, and F. Pezzoni, "Analysis of ego-network structure in online social networks," in *Proc. Int. Conf. Privacy, Security, Risk Trust and Proc. Int. Conf. Soc. Comput.*, Sept. 2012, pp. 31–40.

Multidimensional Networks

Multidimensional (multiplex and multilayer) networks facilitate distinguishing different kinds of links among nodes and naturally arise in social sciences [73], economics and finance [S20], transportation [S21], and biology [S22]. 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 $\mathcal{E}_d$ is 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,

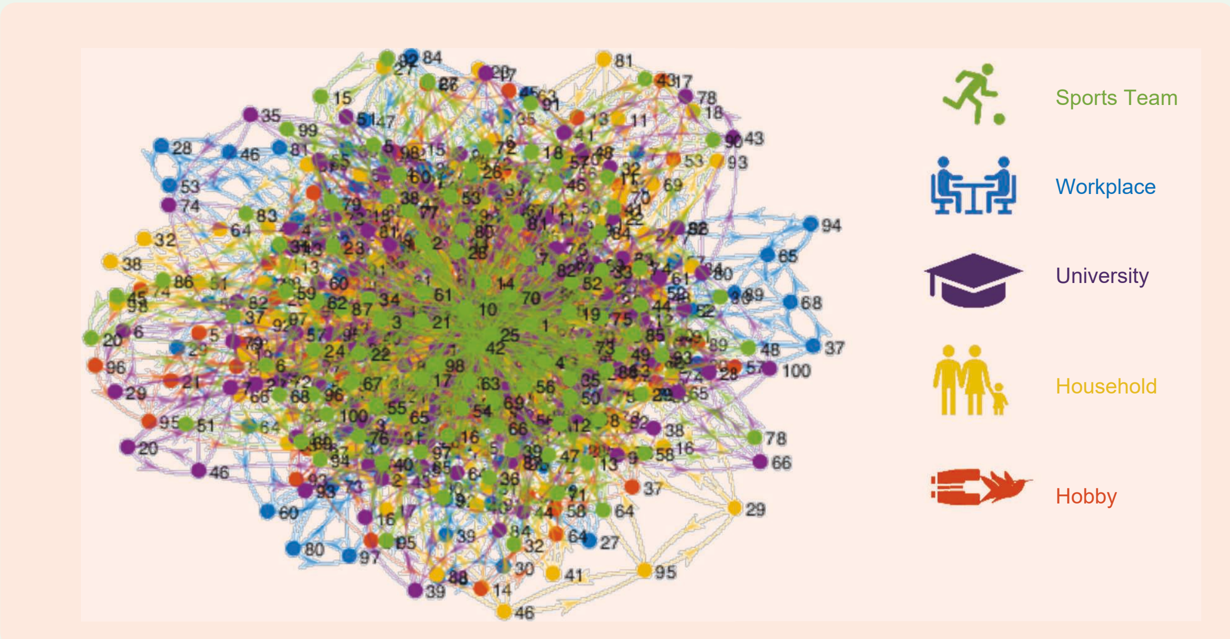


FIGURE S19 Relationships among various individuals: five different layers.

approaches for influence network estimation. These mainly depend on assumptions about networks under observation and available data. The methods are categorized into three classes: 1) statistical models, 2) learning models for social similarity and influence, and 3) model-based approaches. Most address undirected graphs and nondynamical (static) variables, and extensions to directed and 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 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 among signals at those nodes. In particular, one can introduce so-called probabilistic graphical models [52], [79], in which data are interpreted as multiple outcomes of random experiments.

A graphical model is introduced to capture conditional dependence among 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 an opinion (but assume only a 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*) [80]. Given a graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$, MRFs are postulated by requiring that random variables at 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,

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

where $\perp\!\!\!\perp$ denotes statistical independence.

It is shown in [81] that the preceding conditional independence property holds if the probability mass function

where a single dimension is represented as a separate layer. Therefore, it is mandatory in this new framework to 1) generalize the centrality measures defined for classical monodimensional networks and 2) study correlations among dimensions to capture hidden relationships among different layers.

Figure S19 represents relationships among different individuals. The colors correspond to various origins of friendships (for example, friendships arising in sports teams, at workplaces, and at universities). Consider single networks as separate graphs, while multiplex networks are given by the union of all graphs. In this case, each layer has been independently generated as an Erdős–Rényi graph. In Figure S20, 3D networks are considered. They are constituted by different layers and represent the influence network of a community based on a topic under discussion (for example, food and drink, sports, and movies). Each layer is constituted by a sparse common component (the subgraph in red) and a sparse innovation component (the subgraphs in orange, purple, and light blue, respectively).

In [S23], a measure is defined to quantify how similar two dimensions are. These measures can be seen as an extension of the classical Jaccard correlation coefficient to address more than two sets. Let $\mathcal{D} \subseteq \mathcal{L}$ be a subset of dimensions of a network $\mathcal{G} = (\mathcal{V}, \mathcal{E}, \mathcal{L})$. The pair $\mathcal{D}$ correlation is defined as

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

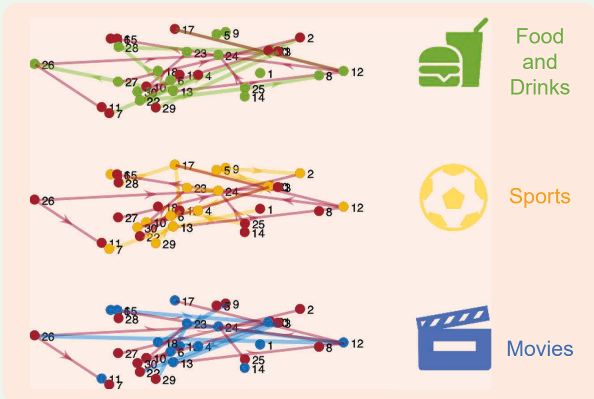


FIGURE S20 Relationships based on common interests.

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

REFERENCES

[S20] C. Perillo and S. Battiston, "A multiplex financial network approach to policy evaluation: The case of euro area quantitative easing," *Appl. Netw. Sci.*, vol. 3, no. 1, Dec. 2018. doi: 10.1007/s41109-018-0098-8.
[S21] J. Wu, C. Pu, L. Li, and G. Cao, "Traffic dynamics on multilayer networks," *Digital Commun. Netw.*, vol. 6, no. 1, pp. 58–63, 2020. doi: 10.1016/j.dcan.2018.10.011.
[S22] A. Halu, M. De Domenico, A. Arenas, and A. Sharma, "The multiplex network of human diseases," 2017. [Online]. Available: https://www.biorxiv.org/content/early/2017/01/18/100370
[S23] M. Berlingerio, M. Coscia, F. Giannotti, A. Monreale, and D. Pedreschi, "Multidimensional networks: Foundations of structural analysis," *World Wide Web*, vol. 16, nos. 5–6, Nov. 2013. doi: 10.1007/s11280-012-0190-4.

(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

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

where $\mathbf{x} \doteq [x_1,\ldots,x_n]^\top$ is a collection of the random variables across nodes, $\boldsymbol{\Theta} = [\theta_{ij}]$ is a parameter matrix, and $Z(\boldsymbol{\Theta})$ 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 (ML) estimation have been proposed [82]. As observed in [83], 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 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 $\boldsymbol{\Sigma\Theta} = I$. This leads to the family of *Gaussian graphical models* [81]. It can be observed that (in this case) the existence of a nonzero entry in the precision matrix immediately implies a partial correlation among 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 (ML) estimator via a log-determinant program. In this class of algorithms, the so-called graphical Lasso (G-Lasso) [84] method has become extremely popular (see "Gaussian Graphical Models and the Graphical Lasso").

Note that, although the convergence of the G-Lasso is guaranteed under suitable conditions, this method has some drawbacks. First, the procedure works only 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 a lack of observed data, giving rise to numerical problems in the identification of a network.

Gaussian Graphical Models and the Graphical Lasso

Graphical models are graphs capturing relationships among many variables, providing a compact representation of joint probability distributions. In these models, nodes correspond to random variables, and edges represent statistical dependencies between node pairs. In Gaussian graphical models, variables at each node are normally distributed, $\mathbf{x} \sim \mathcal{N}(\mathbf{0}, \Sigma_{\mathbf{x}})$, and for any i and $j \in \mathcal{V}$, a zero in the i, j entry of the precision matrix (Figure S21) means conditional independence (given all other variables):

$$x_j \perp\!\!\!\perp x_j \,\big|\, \{x_w\}_{w\in\mathcal{V}\setminus\{i,j\}} \Longleftrightarrow \theta_{ij} = 0 \quad \boldsymbol{\Theta} = \Sigma_{\mathbf{x}}^{-1}.$$

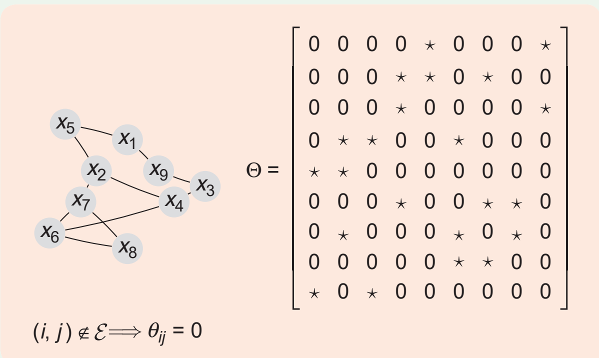


FIGURE S21 A Gaussian model with nine variables.

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 $\Theta = \boldsymbol{\Sigma}^{-1}$. The classical maximum likelihood (ML) estimator is obtained by solving the optimization problem

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

where $\mathrm{tr}(\cdot)$ denotes the trace of a matrix. Classical theory guarantees that in the high-dimensional regime, $\hat{\boldsymbol{\Theta}}_{\mathbf{ML}}$ 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, $\mathbf{S}$ 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, $\boldsymbol{\Theta}$ *is sparse*. The key idea in the graphical Lasso [S24] is to treat each node as a response variable and solve the convex program

$$\hat{\boldsymbol{\Theta}}_{\text{Glasso}} = \max_{\boldsymbol{\Theta}\succeq 0}\log\det(\boldsymbol{\Theta}) - \mathrm{tr}(\mathbf{S}\boldsymbol{\Theta}) - \rho\|\boldsymbol{\Theta}\|_1,$$

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

REFERENCE

[S24] J. Friedman, T. Hastie, and R. Tibshirani, "Sparse inverse covariance estimation with the graphical lasso," *Biostatistics*, vol. 9, no. 3, pp. 432–441, July 2008. doi: 10.1093/biostatistics/kxm045.

It is worth emphasizing that the estimation of the precision matrix via graphical models does not support a direct interpretation of social influence but is able to reflect a pairwise correlation among opinions in a social system. The estimation of social influence, however, is primarily aimed at predicting a direct causal effect of this influence.

Graph Signal Processing

Recent years have witnessed growing signal processing community interest in the analysis of signals that are supported on the vertex set of weighted graphs, leading to the field of graph signal processing (GSP), [85]. 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 facilitates redefining concepts such as the Fourier transform, filtering, and frequency response for data residing on graphs. Note that the signals in the graph are not time dependent. Instead, they vary spatially, and their spatial dynamics are governed by the underlying graph. A brief overview of GSP is provided in "Graph Signal Processing."

While the main directions of research in GSP focus on the development of methods for analyzing signals defined across *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 (GFT). 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 one another. As a measure of smoothness of the signal $\mathbf{x}$ on the graph $\mathcal{G}$, the so-called Laplacian quadratic form is usually adopted [86]:

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

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 are small). Readers are referred to [78] 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 the smoothness of the signal. This method is extended in subsequent works [86], [87] by adding additional constraints on the Laplacian $\mathbf{L}$, thus enabling the volume of the graph to be fixed and imposing specific structures on the graph. Other GSP-based approaches for deriving topological characteristics of a graph are based on graph signal measurement and assume that 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. 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 $\boldsymbol{\Sigma}_\mathbf{x}$ 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 [88]. In this case, a graph signal diffusion model is envisioned, which represents data as (sparse) combinations of overlapping local patterns that reside on a 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*) require the introduction of a more complicated notion of independence, which considers the asymmetry of interconnections. In GSP-based techniques, the directionality of a graph destroys the symmetry of its operator $\mathbf{S}$, thus complicating the mere definition of the GFT. In many contexts (as with social interaction reconstruction, which represents the main focus of this work), learning directed graphs is more desirable, especially for those cases where the edge directions translate to causal dependencies among 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 one exploiting prior information on the model. This research is related to sparse vector autoregressive estimation [83], [89], [90], inverse optimization from partial samples [91], and models from opinion dynamics [39], [92]. 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 so-called social influence network theory [44], [93], considers a social network to be a *dynamical system*. Relevant mathematical models describe the diffusion of some information across the network, which can manifest itself as the evolution of individual opinions, attitudes, and beliefs. Individuals interact (during face-to-face meetings and via social media) and display their opinions about issues to one another. Based on others' views, each individual updates his or her opinion. The within-individual mechanisms of opinion assimilation are related to psychological studies of information integration [94] and cognitive dissonance [95]. 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 his or her own and others' opinions. Models are considered that stem from the French–Harary–DeGroot model of iterative opinion pooling.

The French–Harary–DeGroot Model

One of the simplest models of opinion formation was proposed by French in his seminal work on social power [49] and later examined by the renowned graph theorist Harary [96], [98] (and discussed in "French's Original Model"). The most known version, however, is the generalized one proposed by DeGroot [99] (and, independently, by Lehrer [100], [101]):

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

Here $\mathbf{x}_i(k)$ stands for the opinion of agent i at the kth stage of opinion evolution, and $\mathbf{W} = [w_{ij}]$ is a row-stochastic matrix (a nonnegative matrix whose rows sum to one). It is remarkable that the work [99] (unlike the pioneer articles [49] and [96]) introduced *multidimensional* opinions, which can represent an individual's positions on several issues, for instance, the optimal distribution of resources

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 ith vertex. A simple way to understand the basics of GSP is to consider how the classical concept of the 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 kth node represents the value of the signal x at the discrete time instant k. In Figure S22, (a) represents a signal $\mathbf{x}$ and (b) shows the shifted version $\mathbf{x}^+$; they follow the classical relationship $\mathbf{x}^+ = \textsc{Shift}(\mathbf{x}) \doteq \mathbf{Sx}$, defined as follows:

$$x_i^+ = x_{i-1}, \quad i = 2, \ldots, n,$$
$$x_1^+ = x_n,$$

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 **Adj** of the directed graph:

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

This parallelism between shift operators and graphs may be extended to general graphs (Figure S23). In GSP,

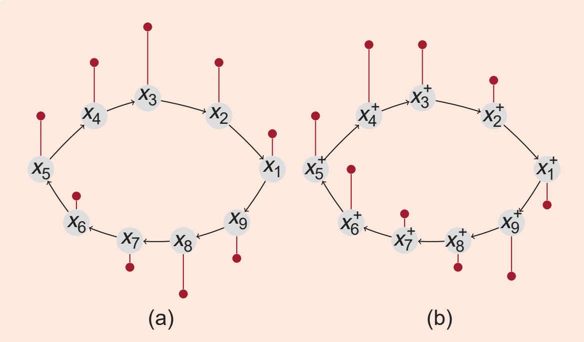


FIGURE S22 (a) A periodic signal on a directed graph $\mathbf{x} = x_1, x_2, \ldots, x_8, x_9$ and (b) its shifted version.

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 $\mathbf{Sx}$. Different choices of $\mathbf{S}$ define different shifts. For undirected graphs, the most typical choice of graph shift operator is the Laplacian L: for any graph signal x, define the new signal $\mathbf{x}^+ = \mathbf{Sx} = \mathbf{Lx}$, whose element x_u^+ is given by

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

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

From the definition of the graph shift operator, the extension of the concept of the Fourier transform to graph signals descends almost immediately [S25]. For $\mathbf{S} = \mathbf{L}$ (under the assumption of the connectivity of the between several entities [102], and the subjective probability distribution of outcomes in some random experiment [99], [103]. 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 one another, thus obtaining an $n \times m$ *matrix* of opinions:

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

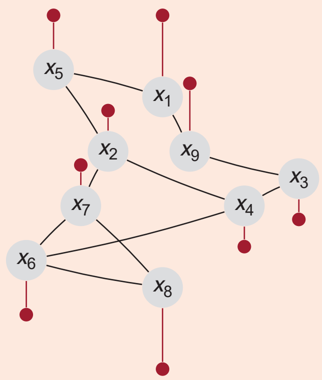
The ℓth column of this matrix $\mathbf{x}^{(\ell)}(k) = (x_1^{(\ell)}, \ldots, x_n^{(\ell)})^\top$ contains the actors' positions on issue $\ell = 1, \ldots, m$. DeGroot's model is then rewritten in the matrix form

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

According to the DeGroot model (see "Degroot's Model as Dynamics Over a Graph"), at each stage of opinion iteration, individuals simultaneously update their positions to convex combinations of all beliefs disclosed to them. The weights w_{ij} of this convex combination serve as natural measures of mutual *influences* among individuals [14], [104]. Social influence can be thought of as a finite resource that individuals distribute among themselves and their peers (this is modeled as a distribution of chips, as discussed in "Friedkin–Johnsen Experiment"). The weight $w_{ij} \geq 0$ assigned by agent i to another agent j measures the importance of j's opinion for i. If $w_{ij} = 1$ (the maximal value), agent i fully relies on agent j's opinion about an issue and is insensitive to the opinions of others:

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

An individual assigning the maximal weight $w_{ii} = 1$ to him- or herself is often called *stubborn* (radical): completely closed to social influence and keeping an opinion unchanged:



(a)

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

(b)

FIGURE S23 (a) A signal defined on a generic *undirected* graph and (b) the corresponding graph shift operator, defined in terms of the Laplacian.

network), if one considers the eigenvalue decomposition of $\mathbf{L} = \mathbf{L}^\top$, then

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

where U is the eigenvector matrix, that is, the matrix containing the eigenvectors of $\mathbf{L}$ as columns (which are orthonormal, being L symmetric), and λ_i 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 works only for undirected graphs. Extensions to directed graphs are nontrivial since the GFT definition does not cover situations where $\mathbf{L}$ has complex eigenvalues and when it is not diagonalizable.

REFERENCE

[S25] A. Sandryhaila and J. M. F. Moura, "Discrete signal processing on graphs: Frequency analysis," *IEEE Trans. Signal Process.*, vol. 62, no. 12, pp. 3042–3054, June 2014. doi: 10.1109/TSP.2014.2321121.

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

If $w_{ij} = 0$, the opinion of agent j is either not disclosed to agent i or not taken into account by agent i. 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 $k+2$, $k+3$, and so on through the opinions of other individuals (a chain of influence $j \longrightarrow j' \longrightarrow j'' \ldots \longrightarrow i$). Along with French's representation, the DeGroot model (5) predicts the *consensus* of opinions, that is, the convergence of all opinion vectors $x_i(k)$ to the same vector as $k \to \infty$. Equivalently, consensus means that the matrices $\mathbf{W}^k$ converge to a stochastic matrix of rank one; that is,

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

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 a centrality measure on a social network (similar in spirit to

French's Original Model

In this model [49], the following events occur:

- A group of n individuals is associated with nodes of a directed graph.
- Individual i holds an *opinion* x_i, assumed to be a scalar real value.
- Individual j discloses his or her 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 his or her opinion to the *mean values* of all opinions he or she has encountered.

For instance, Figure S24 describes the opinion formation process

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

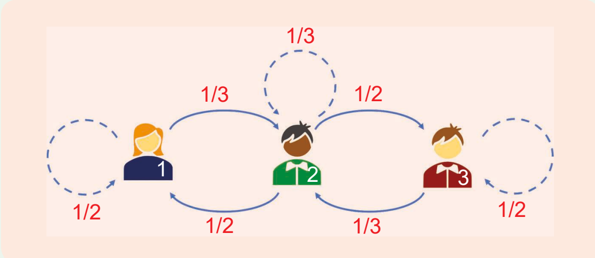


FIGURE S24 French's opinion formation model with $n = 3$ individuals.

The most typical behavior of the model is the 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], [96], [97] 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).

DeGroot's Model as Dynamics Over a Graph

Social network $\leftrightarrow$ weighted graph $\mathcal{G} = (\mathcal{V}, \mathcal{E}, \mathbf{W})$, according to the following (see Figure S25):

- Agents $\leftrightarrow v \in \mathcal{V}$.
- Interactions $\leftrightarrow \mathcal{E} \subseteq \mathcal{V} \times \mathcal{V}$.
- Influences $\leftrightarrow \mathbf{W} \in \mathbb{R}^{\mathcal{V}\times\mathcal{V}}$.
- Here, $w_{ij} = 0$ 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).

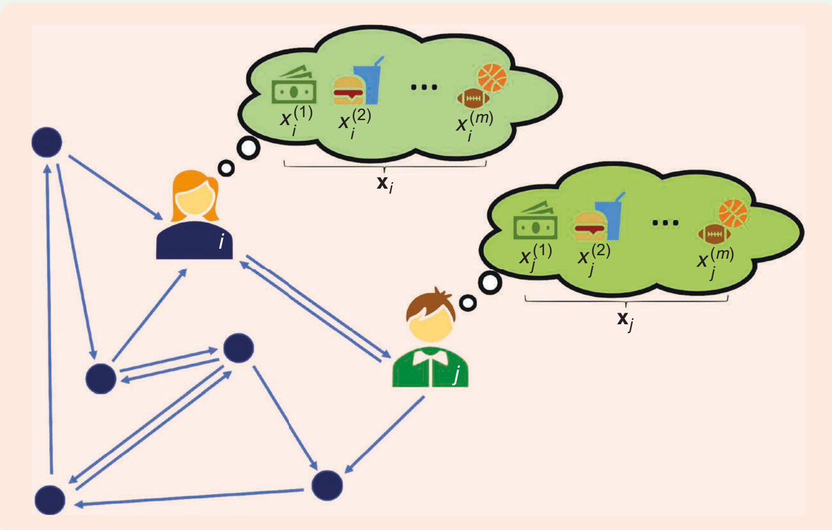


FIGURE S25 DeGroot's model: opinions on several issues.

eigenvector centrality) and characterizing *social powers* of individuals [49], [104]. The vector of the consensus opinions of a 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 [105]. The SIA property can be proved for primitive (irreducible and aperiodic) [106], [107] matrices, for which $\mathbf{W}^k$ has strictly positive entries for a sufficiently large exponent k [99]. Another standard criterion guarantees consensus if all diagonal entries w_{ii} are positive and if the graph corresponding to $\mathbf{W}$ has a globally reachable node (some individual influences all others directly or indirectly) [108], [109]. A necessary and sufficient graph-theoretic condition for consensus can be found in [15] and [97].

From Consensus to Disagreement

Since social groups often fail to reach a consensus, realistic models of opinion formation should be able to explain not only "regular" consensus behavior but various "disordered" actions featuring disagreement. Finding such a model is a problem that has been studied since the 1960s [20], [110], and the difficulty is known as *Abelson's diversity puzzle* (or the *problem of community cleavage* [14]). Most models portraying community cleavage (for instance, the convergence of opinions to several clusters) replace the DeGroot equation through nonlinear dynamics, taking into account various effects of information assimilation and integration within individuals and communication among individuals [6], [21], [111]–[119]. The most studied ones are *bounded confidence* models [16], [103], [111], [112], [118] capturing the effect of homophily in social groups and assuming that individuals tend to assimilate opinions of like-minded peers and meet dissimilar opinions with discretion and even ignore them. The simplest, the Hegselmann–Krause model [111], may be considered 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 [23]. Hence, despite some recent progress in the identification of nonlinear networks [120]–[122], 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 among individuals and their stubbornness (reluctance to change the initial opinion). Models of the first type revise the basic assumption about the convex combination mechanism of opinion evolution and allow not only the attraction of agents' opinions but their *repulsion* [123]–[130]. The presence of negative influence is typically explained by "boomerang," reactance, and anticonformity effects [20], [125] (that is, the *resistance of some individuals to social influence*). The theory of signed (or "coopetitive") dynamical networks developed in the literature is extremely important due to various applications in economics, physics, and biology [131]. However, its applicability to social influence systems is still disputable for several reasons. The evidence of the ubiquity of negative influence has not been experimentally secured. Since the first definition of boomerang effects [132], the empirical literature has concentrated on special conditions under which these effects might arise in dyadic interpersonal interactions. Whereas positive and negative *relations* (friendship/enmity and trust/distrust) between individuals are ubiquitous, it is still unclear whether such relations really correspond to positive and negative influences [133].

Under a natural assumption of *strong connectivity*, the clustering of opinions in the presence of antagonistic interactions usually requires various forms of a *structural balance* of positive and negative ties [124], [127], [129], [134], [135]. The clustering of opinions without structural balance is usually guaranteed by special hierarchical ("extended leader-following") structures in a graph [136], [137]. If a graph of social influence is time-varying, the absence of strong connectivity (understood in some uniform sense [138]) may lead to the existence of oscillatory solutions. At the same time, the usual DeGroot model is able to explain the 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 $w_{ii} = 1$) [139]. Further development of the DeGroot model with stubborn individuals has naturally led to the Friedkin–Johnsen model (considered in the next section). Unlike many other models proposed in physical and engineering literature, the Friedkin–Johnsen model has been experimentally assessed on small and medium groups [31], [32], [102], [140]. An essential part of these experiments is the empirical procedure of matrix $\mathbf{W}$ reconstruction (see "Friedkin–Johnsen Experiment").

Friedkin–Johnsen Model

Whereas the DeGroot representation accommodates stubborn individuals who are completely closed to social influence, the Friedkin–Johnsen model admits "partial" stubbornness, which is measured by a *susceptibility* coefficient $\lambda_i \in [0, 1]$. An agent with minimal susceptibility is the stubborn individual retaining his or her initial opinion (8), whereas the agent with maximal susceptibility assimilates to others' opinions in accordance with the conventional DeGroot mechanism (5). In general, an individual opinion at each iteration is influenced by both others' opinions and agents' initial opinions:

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

The matrix $\mathbf{W}$ is stochastic and has the same meaning as in DeGroot's model; namely, w_{ij} represents the influence weight that individual i accords to individual j. Without a 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 full stubbornness in the sense of (8) (see "Simple Properties of the Friedkin–Johnsen Model"). As discussed in [31], individuals' anchorage at their initial opinions can be explained by an ongoing effect of some exogenous factors that previously influenced a social group in the past. An initial opinion can also be considered an individual's *prejudice* [36], [42], [141] that influences his or her opinion in subsequent steps.

Similar to DeGroot's model, the opinions $\mathbf{x}_i(k)$ may be scalar or multidimensional. Stacking the opinions on top of one another to obtain the opinion matrix $\mathbf{X}(k)$, the Friedkin–Johnsen system (10) can be rewritten in 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 $\mathbf{\Lambda} = \mathbf{I}_n$.

Due to the presence of fully and partially stubborn agents, the Friedkin–Johnsen dynamics usually do not lead to a consensus of opinions (except for special situations where the Friedkin–Johnsen 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 $\mathbf{\Lambda}\mathbf{W}$ 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], [36], [42] is summarized in "Schur Stability Criteria." If $\mathbf{\Lambda}\mathbf{W}$ is a Schur-stable matrix, then the matrix of opinions converges [36], [42]:

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

The matrix $\mathbf{V} = [v_{ij}]$ is *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").

Dynamics of Reflected Appraisal

The concept of influence centrality (see "Control Matrix and Friedkin's Centrality") serves as a basis for dynamical models describing the evolution of influence matrix $\mathbf{W}$ and is known as the dynamics of *reflected appraisals*. As argued in [75], 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 a social group discusses a sequence of issues (see "Reflected Appraisal Model"). Notice that models of appraisal dynamics do not portray the evolution of opinions, and they operate with quantities that are hard to measure experimentally, such as centrality vectors. This, along with the highly nonlinear dynamics of 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 Friedkin–Johnsen Model

The seminal Friedkin–Johnsen model can be extended in many directions, among which only three are considered.

Simple Properties of the Friedkin–Johnsen Model

Using induction on $k = 0, 1, \ldots$, a number of properties of the Friedkin–Johnsen model can be proved, as in the following:

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 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 their iteration, 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 Friedkin–Johnsen theory. Even for 3D opinions, it is difficult to visualize the convex hull $\mathfrak{X}_0$ (being a convex polyhedron) without special software. Nevertheless, experiments concerned with rational decision making about resource allocation [102] illustrate that individuals' multidimensional decisions typically stay in the convex polyhedron $\mathfrak{X}_0$.

The first is concerned with the dynamics of multidimensional opinions, which represent an 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" [142]. Contradictions and other inconsistencies among beliefs, attitudes, and ideas may trigger

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 who are fully or partially stubborn (anchored at their initial opinions):

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

As discussed in [31], an individual's attachment to his or her initial opinion may be explained as a direct, ongoing effect of some previous experience or other external factors that previously influenced a group. The Friedkin–Johnsen 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

The matrix $\Lambda\mathbf{W}$ 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 [36], [42].

The social group in Figure S26(a) corresponds to a Schur-stable matrix $\Lambda\mathbf{W}$ (each node with $\lambda = 1$ is connected to one of the nodes with $\lambda < 1$). For the group in Figure S26(b), the matrix $\Lambda\mathbf{W}$ is not Schur stable: the group of red nodes is not connected to the unique node with $\lambda = 1$.

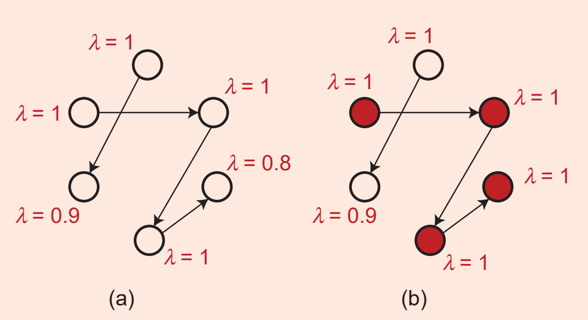


FIGURE S26 (a) A network corresponding to the Schur-stable matrix. (b) A network whose matrix is not Schur stable.

Control Matrix and Friedkin's 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 the *social power* [104]) 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 Λ, a whole family of centrality measures is obtained for the weighted graph $G[\mathbf{W}]$; it was introduced by Friedkin [62] for the case where $\Lambda = \alpha \mathbf{I}_n$ 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 $\mathbf{c} = (c_1,\ldots,c_n)^\top$ can be found as

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

For a specially chosen matrix [S26] $\mathbf{W}$ and $\alpha = 1 - m$ [where $m \in (0,1)$], the latter vector coincides with the PageRank centrality measure, which appeared in [62] seven years earlier than the seminal work by Brin and Page [58]. Relations between influence centrality and the PageRank are discussed in more detail in [15], [S27], and [S28].

REFERENCES

[S26] 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. doi: 10.1109/MCS.2014.2308672.

[S27] N. Friedkin and E. Johnsen, "Two steps to obfuscation," *Soc. Netw.*, vol. 39, pp. 12–13, Oct. 2014. doi: 10.1016/j.socnet.2014.03.008.

[S28] A. Proskurnikov, R. Tempo, and M. Cao, "PageRank and opinion dynamics: Missing links and extensions," in *Proc. IEEE Conf. Norbert Wiener 21st Century*, Melbourne, 2016, pp. 12–17. doi: 10.1109/NORBERT.2016.7547432.

Reflected Appraisal Model

In psychology, the theory of reflected appraisal states that people's perceptions are influenced by the evaluation of others [77]. In [75], 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:

$$\begin{aligned} \mathbf{W}^{(s)} &= \mathbf{I} - \boldsymbol{\Lambda}^{(s)} + \boldsymbol{\Lambda}^{(s)}\mathbf{C}, \\ (\mathbf{c}^{s})^{\top} &= \frac{\mathbb{1}^{\top}}{n}(\mathbf{I} - \boldsymbol{\Lambda}^{(s-1)}\mathbf{W}^{(s-1)})^{-1}(\mathbf{I} - \boldsymbol{\Lambda}^{(s-1)}), \\ \boldsymbol{\Lambda}^{(s)} &= I - \mathrm{diag}(\mathbf{c}^{(s)}) = I - \mathrm{diag}(\mathbf{W}^{(s)}), \end{aligned} \tag{S1}$$

where $\mathbf{c}^{s}$ is the influence centrality vector during a discussion about issue s and $\mathbf{C}$ is a constant matrix with zero diagonal entries. From a control-theoretic viewpoint, this mechanism can be interpreted as nonlinear feedback. The social power $c_i(s)$ that an individual has acquired in the discussion about issue $s-1$ influences his or her self-weight (and thus the weights assigned to others) during a discussion around issue s.

The evolution of the influence network and social power is depicted as a function of issue sequence, respectively (Figures S27 and S28). It should be noted that the network topology is the same at each layer, as the mechanism alters only the influence weights. However, the influence strength changes across the 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,

$$\mathbf{W}_{ij}^{(s)} \neq 0 \qquad \forall (i,j) \in \Omega, \forall \ell \in \{1,\ldots,m\}. \tag{S2}$$

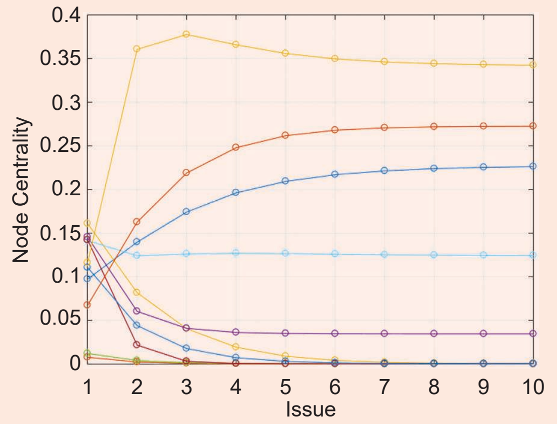


FIGURE S27 Reflected appraisal model (S1): the evolution of nodes' centralities.

A simpler model of appraisal dynamics, which is now referred to as the *DeGroot–Friedkin model* [43], was proposed in [S29]. Unlike (S1), it is based on the DeGroot model and describes the evolution of social power vector **p** from (9) instead of the influence centrality vector **c**.

REFERENCE

[S29] P. Jia, A. MirTabatabaei, N. E. Friedkin, and F. Bullo, "Opinion dynamics and the evolution of social power in influence networks," *SIAM Rev.*, vol. 57, no. 3, pp. 367–397, 2015. doi: 10.1137/130913250.

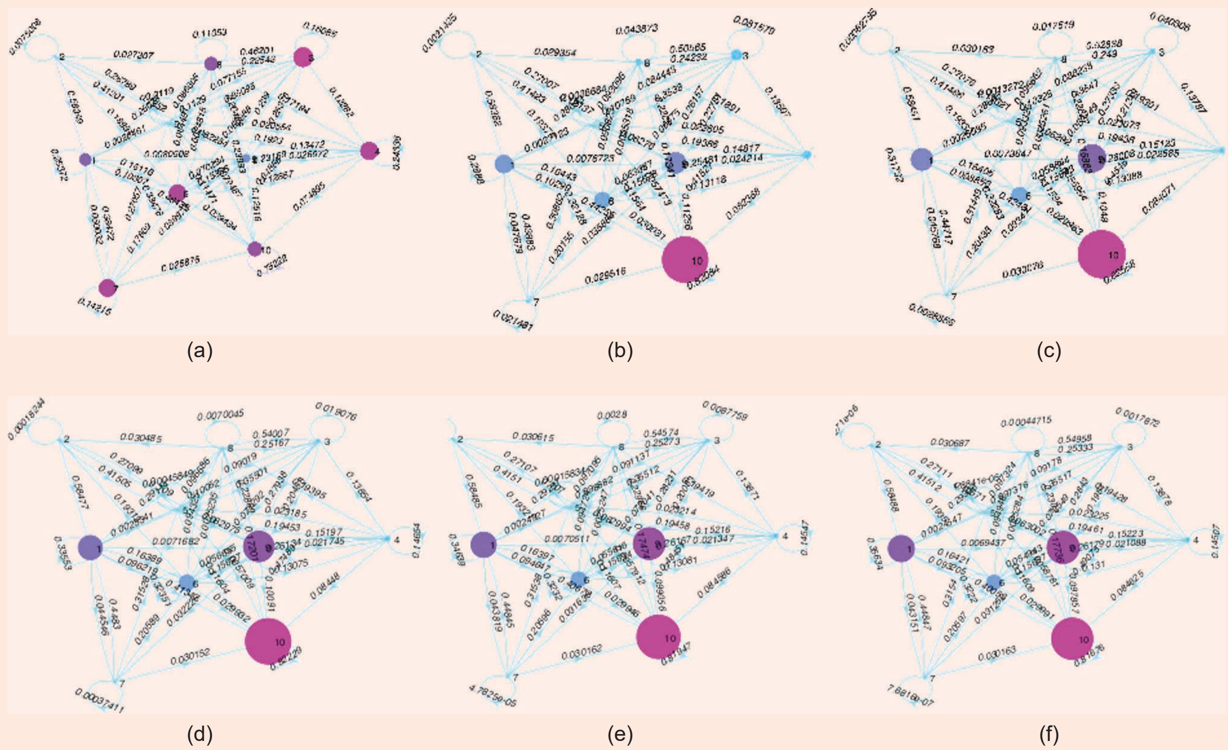


FIGURE S28 Reflected appraisal model (S1): the evolution of weights at periods $k = 0,\ldots,5$ [correspond to parts (a)–(f)].

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 function of the brain, enabling humans to develop coherent systems of beliefs [95], [143]. Modeling the dynamics of opinions about interrelated issues is a challenging problem, and only a few models have been proposed in the literature (most of them feature nonlinear dynamics [144]–[146]). "A Model of a Belief System's Dynamics" is devoted to a simple linear model proposed in [42], [147], and [148]. In general, the presence of logical relations among issues can affect the recoverability of an influence network from partial observations [39].

Another extension of the Friedkin–Johnsen model revises the restrictive assumption about simultaneous communication. As stated in [31], *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 assume that only a couple individuals can interact at each step were introduced in [36], [37], and [42]. Such multiagent communication protocols are known as *gossiping* [149]. The model is summarized in "Asynchronous Gossip-Based Friedkin–Johnsen Model." In [42], a gossip-based version of the extended Friedkin–Johnsen model (16) is considered. Another potential "culprit" of randomness is noise, representing the effects of individuals' free will and the unpredictability of their decisions (one model with noise is discussed in "Dynamics of Multiplex Networks").

INFERENCE ACROSS NETWORKS: MODEL-BASED APPROACH

This section provides a classification of different approaches for inference across networks, while subsequent sections introduce two classes of the methods. The models presented in the previous sections have proved to be

A Model of a Belief System's Dynamics

Adjusting his or her position on an interdependent issue, an individual might have to simultaneously change perspectives on several related issues to maintain a 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 [148] that (in some situations) the operator $\mathcal{C}_i$ may be linear (and 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, a multidimensional opinion is represented by a row m-dimensional vector so that $\mathbf{C}_i \in \mathbb{R}^{m\times m}$). Assuming that a tension-resolving process follows the integration of opinions from neighbors, the Friedkin–Johnsen model (10) is replaced by the dynamics

$$\mathbf{x}_i(k+1) = \lambda_i\left(\sum_{j=1}^{n} w_{ij}\mathbf{x}_j(k)\right)\mathbf{C}_i^\top + (1-\lambda_i)\mathbf{x}_i(0). \quad \text{(S3)}$$

It has been shown (see the supplementary material for [148]) that if the matrix $\Lambda\mathbf{W}$ is Schur stable and all matrices $\mathbf{C}_i$ are row stochastic or (more generally) $\mathbf{C}_i = (c_{lm}^{(i)})$, where $\Sigma_m \left|c_{lm}^{(i)}\right| \le 1$ for each l and i, then the linear operator

$$\mathbf{X} \mapsto \Lambda\mathbf{W}\begin{pmatrix}\mathbf{x}_1\mathbf{C}_1^\top\\ \mathbf{x}_2\mathbf{C}_2^\top\\ \vdots\\ \mathbf{x}_n\mathbf{C}_n^\top\end{pmatrix}$$

is Schur stable. Specifically, the opinion matrix $\mathbf{X}(k)$ determined by (S3) converges as $k \to \infty$. Equation (S3) becomes more elegant in the case of homogeneous agents $\mathbf{C}_1 = \ldots = \mathbf{C}_n = \mathbf{C}$. In this situation, (S3) may be rewritten in a matrix form very similar to (S4):

$$\mathbf{X}(k+1) = \Lambda\mathbf{W}\mathbf{X}(k)\mathbf{C}^\top + (\mathbf{I}_n - \Lambda)\mathbf{X}(0). \quad \text{(S4)}$$

The Friedkin–Johnsen model is a special case of (S4), 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 [42], $\mathbf{C}$ is referred to as the *multi-issue dependency structure* (*MiDS*) matrix. An example of the system (S4) with 3D opinions was considered in [148]. It was conjectured that a speech by Colin Powell, a highly respected U.S. Secretary of State, at the United Nations Security Council presented a logic structure through three truth statements:

1) Saddam Hussein had a stockpile of weapons of mass destruction.
2) Hussein's weapons of mass destruction were real and presented dangers to the region and the world.
3) An invasion of Iraq would be a just war.

It was a logic structure, as a high certainty of belief in statement 1 implies a high certainty of belief in statements 2 and 3. On the other hand, if statement 1 is false, then statements 2 and 3 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 [148] shows that if a population initially has a high certainty about statement 1, then the belief system dynamics across a random graph generate a consensus that a preemptive invasion is a just war. At the same time, if statement 1 is considered false, then the population's certainty belief in all three statements is dramatically lowered. Hence, the model can explain the fluctuation of public opinion about the Iraq invasion.

powerful tools for the analysis of interactions in social networks. However, the general structure of the network is typically not available. Hence, the following question arises: Given measurements of the evolution of 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, the second part of this article describes recent approaches in the literature to infer social influences in 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 Friedkin–Johnsen model. The approaches described can be adapted to the DeGroot model and other representations of opinion formation.

Inference methods can be categorized according to available information. Two main research areas are distinguished. The first considers uncontrolled experiments, assuming that we cannot intervene with a social system and that only individual opinion updates can be tracked. This is a passive approach taken in previous work [152]. The latter considers controlled experiments, developing a social radar by exploiting the special role of stubborn agents. As shown in [92], stubborn agents can help expose 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 [92] and [153]–[155] for inference network estimation via controlled experiments.

Consider two strategies to estimate interactions in a network that are referred to as *persistent measurement* and *sporadic measurement* identification procedures. In experiments of the first kind (persistent measurement), opinions are observed during T rounds of conversation, and the influence matrix is estimated as the matrix best fitting the dynamics for $0 \le k < T$. In such cases, the available results for parsimonious system identification can be used to determine the unknown parameters [83], [156]. To

Asynchronous Gossip-Based Friedkin–Johnsen Model

The Friedkin–Johnsen model [S30] can be extended to a case where interactions follow a model more consistent with "usual" social network interactions, during which only a few agents communicate at one time. In this case, the opinions evolve as follows [36]:

- Each agent $i \in \mathcal{V}$ starts from an initial belief $x_i(0) \in \mathbb{R}$.
- At each period $k \in \mathbb{Z}_{\ge 0}$, a subset of *active* nodes $\mathcal{V}_k$ is randomly selected from a uniform distribution across $\mathcal{V}$;
- 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 Friedkin–Johnsen mechanism, which results in the equations

$$\begin{aligned} x_i(k+1) &= \lambda_i((1-w_{ij})x_i(k) + w_{ij}x_j(k)) \\ &\quad + (1-\lambda_i)x_i(0), \qquad \forall i \in \mathcal{V}_k, \\ x_\ell(k+1) &= x_\ell(k), \qquad \forall \ell \in \mathcal{V} \setminus \mathcal{V}_k. \end{aligned} \tag{S5}$$

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 (S5) can be formally rewritten in the following form: given $\mathcal{V}_k$ and letting $\theta(k) \doteq \{\theta_i\}_{i \in \mathcal{V}_k}$,

$$\mathbf{x}(k+1) = \boldsymbol{\Gamma}(k)\mathbf{x}(k) + \mathbf{B}(k)\mathbf{x}(0), \tag{S6}$$

where the coefficients are defined as

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

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 $\{\mathbf{x}(k)\}_{k \in \mathbb{Z}_{\ge 0}}$ is a Markov process [S31], which fails to converge in a deterministic sense and shows persistent oscillations [Figure S29(a)].

However, if the matrix $\boldsymbol{\Lambda}\mathbf{W}$ 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 [38] that the opinions' expected values obey the equation

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

where

$$\begin{aligned} \bar{\boldsymbol{\Gamma}} &\doteq \mathbb{E}[\boldsymbol{\Gamma}(k)] = (1-\beta)\mathbf{I}_n + \beta\boldsymbol{\Lambda}(\mathbf{I}_n - \mathbf{D}^{-1}(\mathbf{I}_n - \mathbf{W})), \\ \bar{\mathbf{b}} &\doteq \beta(\mathbf{I}_n - \boldsymbol{\Lambda})\mathbf{x}(0), \end{aligned} \tag{S7}$$

where $\beta = |\mathcal{V}_k|/|\mathcal{V}|$ 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

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

The opinion sequence has a few more interesting ergodicity properties that can be exploited in estimation algorithms. Namely, 1) $\mathbf{x}(k)$ converges in a distribution to a random variable $\mathbf{x}_\infty$, and the distribution is the unique invariant distribution for (S5); 2) the process is ergodic; 3) the limit random variable satisfies $\mathbb{E}[\mathbf{x}_\infty] = (\mathbf{I}_n - \bar{\boldsymbol{\Gamma}})^{-1}\bar{\mathbf{b}}$; and 4) the Cesáro averages converge almost surely (and in the sense of pth moment for each $p \ge 1$):

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

exemplify how such an approach can be used in the context of network inference estimation, assume that opinions evolve according to the multidimensional Friedkin–Johnsen model (11), which is recalled here for readability:

$$\mathbf{X}(k+1) = \mathbf{\Lambda W X}(k) + (\mathbf{I}_n - \mathbf{\Lambda})\mathbf{X}(0). \tag{13}$$

Assume that measurements of $\mathbf{X}(k)$ are available for $k = 0, 1, 2, \dots, T$. This 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

$$\mathbf{X}(k+1) = \mathbf{AX}(k) + \mathbf{BX}(0),$$

with $\mathbf{A} \doteq \mathbf{\Lambda W}$, $\mathbf{B} = \operatorname{diag}(b_1, b_2, \dots, b_n) \doteq \mathbf{I}_n - \mathbf{\Lambda}$. Denote by $\operatorname{off}(\mathbf{A}) \doteq \mathbf{A} - \operatorname{diag}(\mathbf{A})$ 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 collected measurements can be formulated as the following optimization problem:

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

where $\|\mathbf{A}\|_0$ denotes the number of nonzero entries of the matrix $\mathbf{A}$. This optimization problem is a nonconvex combinatorial one, 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):

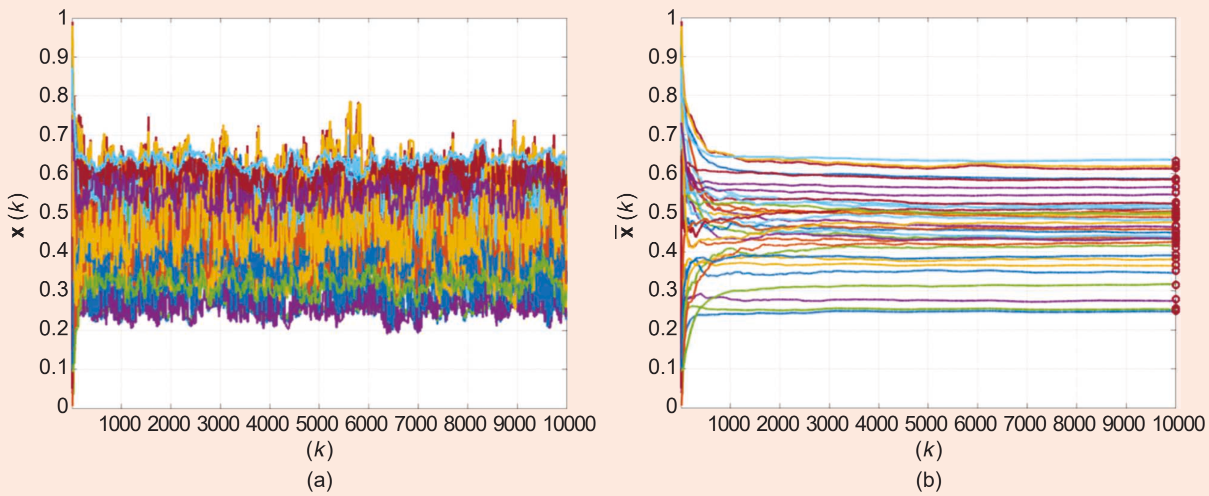


FIGURE S29 A gossip-based Friedkin–Johnsen model: (a) random opinions versus (b) the Cesáro averages.

To identify influences in a social network, the opinion cross correlation matrix is useful and defined as

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

As shown in [150], these correlation matrices satisfy

$$\mathbf{\Sigma}^{[\ell+1]}(k) = \mathbf{\Sigma}^{[\ell]}(k)\bar{\mathbf{\Gamma}}^\top + \mathbb{E}[\mathbf{x}(k)]\bar{\mathbf{b}}^\top. \tag{S8}$$

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

$$\mathbf{\Sigma}^{[\ell+1]}(\infty) = \mathbf{\Sigma}^{[\ell]}(\infty)\bar{\mathbf{\Gamma}}^\top + \mathbb{E}[\mathbf{x}(\infty)]\bar{\mathbf{b}}^\top. \tag{S9}$$

Notice that the relation in (S9) is a sort of Yule–Walker equation [151] used for estimation in autoregressive processes.

REFERENCES

[S30] N. Friedkin and E. Johnsen, "Social influence and opinions," *J. Math. Sociol.*, vol. 15, nos. 3–4, pp. 193–205, 1990. doi: 10.1080/0022250X.1990.9990069.

[S31] G. Grimmett and D. Stirzaker, *Probability and Random Processes*. Oxford University Press, Oxford, U.K., 2001.

Dynamics of Multiplex Networks

The Friedkin–Johnsen model can be extended to cases where a 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 social systems will share some common feature. In this sense, $\mathcal{M}_{cc}$ (the common component model; see Figure S30) considers cases where networks differ in few components. Instead, $\mathcal{M}_{cs}$ (the common support model; see Figure S31) 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:

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

where $\mathbf{x}^{(s)}$ represent agents' opinions about a specific subject s and $\boldsymbol{\eta}^{(s)}(t) \sim \mathcal{N}(\mathbf{0}, \mathbf{Q}_\eta)$ is additive noise. The Markov process (Figure S32) exhibits persistent fluctuations due to the random uncertainty in the dynamical system. It can also be shown that the expected opinions and cross correlation matrices converge to a final pattern of values.

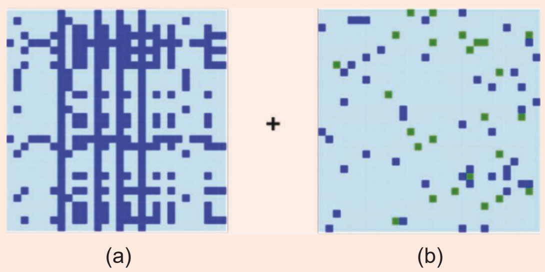


FIGURE S30 Model $\mathcal{M}_{cc}$: Influence matrices differ in few components. (a) A common component of an adjacency matrix of two influence networks. (b) Sparse innovation of two influence networks.

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 osscillations 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 [41]):

$$\mathbf{\Sigma}_{(s)}^{[\ell+1]}(k) = \mathbf{\Sigma}_{(s)}^{[\ell]}(k)(\bar{\mathbf{\Gamma}}^{(s)})^\top + \mathbb{E}[x^{(s)}(k)](\bar{\mathbf{b}}^{(s)})^\top, \tag{S10}$$

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

$$\begin{aligned}\mathbf{\Sigma}_{(s)}^{[0]}(\infty) = {} & \mathbf{\Lambda}^{(s)}\mathbf{W}^{(s)}\mathbf{\Sigma}_{(s)}^{[0]}(\infty)(\mathbf{W}^{(s)})^\top\mathbf{\Lambda}^{(s)} + \mathbf{\Lambda}^{(s)}\mathbf{W}^{(s)}\mathbb{E}[\mathbf{x}(\infty)]\mathbf{u}^\top(\mathbf{I}_n - \mathbf{\Lambda}^{(s)}) \\ & + (\mathbf{I}_n - \mathbf{\Lambda}^{(s)})\mathbf{u}\mathbb{E}[\mathbf{x}(\infty)]^\top(\mathbf{W}^{(s)})^\top\mathbf{\Lambda}^{(s)} \\ & + (\mathbf{I}_n - \mathbf{\Lambda}^{(s)})\mathbf{u}\mathbf{u}^\top(\mathbf{I}_n - \mathbf{\Lambda}^{(s)}) + \mathbf{Q}_\eta.\end{aligned} \tag{S11}$$

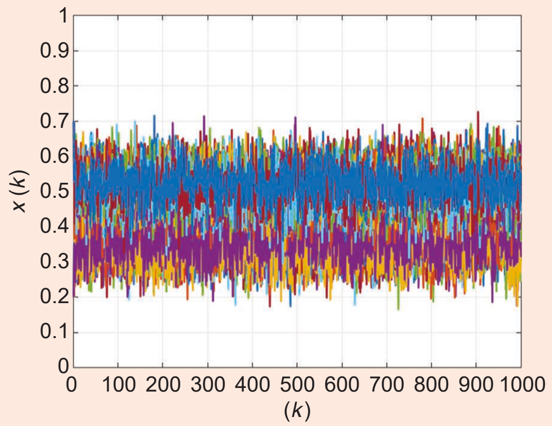


FIGURE S32 A sample trajectory of the evolution of believes.

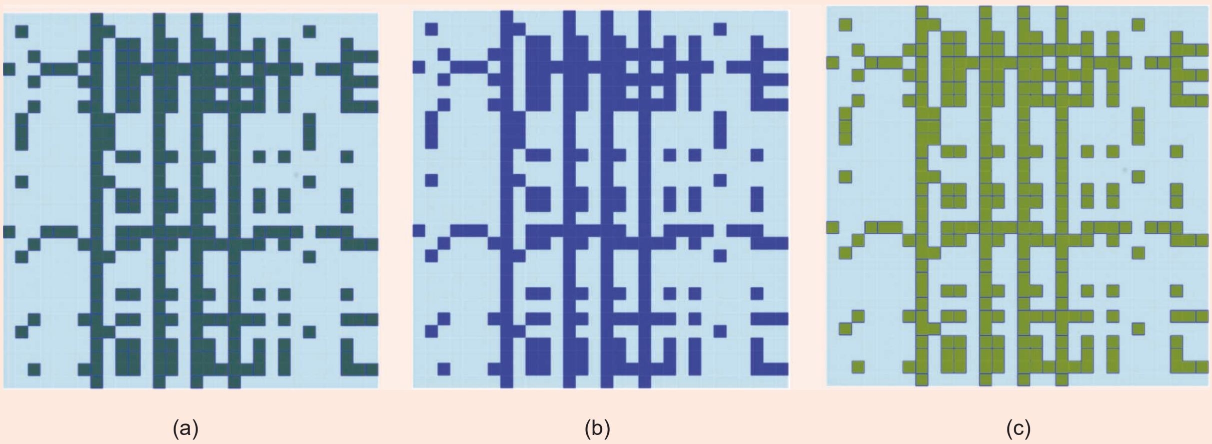


FIGURE S31 Model $\mathcal{M}_{cs}$: Influence matrices have common topology. (a)–(c) The adjacency matrices of three influence networks. They have common support but different entries.

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

Since $\| \mathrm{off}(\mathbf{A}) \|_1 \doteq \Sigma_{i=1}^{n} \Sigma_{j \ne i} A_{ij}$, the latter problem can be decomposed into n independent ones. 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 and storing a sufficiently long subsequence of opinions $\mathbf{x}(k)$, $\mathbf{x}(k+1), \ldots, \mathbf{x}(k+M-1)$. 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.

Compressed Sensing

The optimization issues in (16) are a particular case of the so-called sparse recovery problem starting from compressed measurements [163], a problem also known as *compressed sensing*. More precisely, sparse recovery problems are of the form

$$\begin{aligned}
&\min_{\mathbf{z} \in \mathbb{R}^n} \| \mathbf{z} \|_0 \\
&\text{s.t. } \mathbf{\Phi}\mathbf{z} = \boldsymbol{\psi},
\end{aligned}$$

where $\mathbf{\Phi} \in \mathbb{R}^{m \times n}$ with $m < n$, and $\| \mathbf{z} \|_0$ defines the ℓ_0 quasinorm (which corresponds to the number of nonzero elements of $\mathbf{z}$). Note that the linear system of equations in the preceding optimization problem 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 the *spark* of a matrix [34].

SPARK OF A MATRIX

The spark of a given matrix $\mathbf{\Phi}$, denoted with 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. Interested readers can refer to [34, 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 compressed sensing problem is known to be, in general, NP-hard. For this reason, ℓ_1-based relaxations are often used to approximate the solution of compressed sensing problems. More precisely, this relaxation has the form

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

Much of the theory concerning explicit performance bounds for the relaxation described previously relies on the concept of the restricted isometry property (RIP), which characterizes matrices that are nearly orthonormal, at least when acting on sparse vectors [S32].

Restricted Isometry Property

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

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

for all $\mathbf{z} \in \mathcal{Z}_s = \{\mathbf{z} \in \mathbb{R}^n : \| \mathbf{z} \|_0 \le s\}$. 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 [S32]) 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 an s-sparse vector using the ℓ_1 relaxation described previously. It is straightforward to see that

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

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

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

for all $S \subseteq [n]$ with $|S| \le s$. It is well known [S32] that sub-Gaussian random matrices with independent identically distributed entries satisfy the RIP of order $2s$ with constant δ_{2s} (with a probability close to one) if

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

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

REFERENCE

[S32] E. Candes, "The restricted isometry property and its implications for compressed sensing," *Comptes Rendus Mathematique*, vol. 346, nos. 9–10, pp. 589–592, May 2008.

Moreover, the system could be updated at an unknown interaction rate, and the interaction timing among agents can be unobservable, in practice [157], [158]. These considerations make the persistent measurement approach inapplicable in many practical situations, as discussed in [92]. To circumvent these issues, this article describes two approaches that fall in the second class of methods; that is, they use only 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 [92], agents interact until their opinions stabilize, and the identification problem considers only the initial and final opinions. In the second, it is assumed only that one has access to random measurements of agents' opinions, and statistics of the measurement process are used to estimate the structure of the social network that generated the measurements. Figure 3 and Table 2 summarize the main differences and requirements of the reviewed methods.

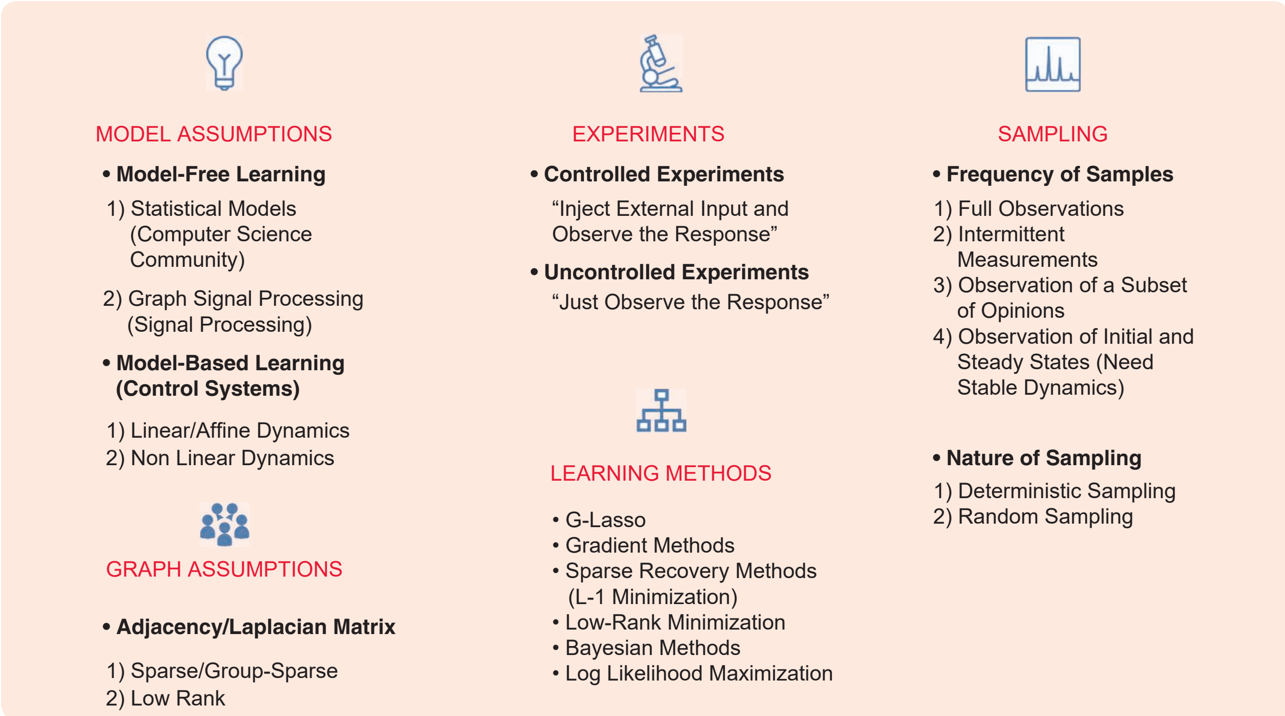


FIGURE 3 Learning graphs from data: main features.

TABLE 2 A comparison of selected model-based learning methods.

	Model Assumptions	Graph Assumptions	Experiments	Sampling
Abir et al. [159]	DeGroot	Sparse networks	Passive	Full observations
Wang et al. [160]	DeGroot	Sparse models	Passive	Full observations
Wai et al. [92]	DeGroot	Sparse models	Controlled	Infinite horizon
Wai et al. [153]	DeGroot	Sparse/low-rank models	Controlled	Infinite horizon
Wai et al. [154]	DeGroot	Low-rank models	Controlled	Infinite horizon
Wai et al. [155]	Nonlinear dynamics	Sparse	Controlled	Infinite horizon
Ravazzi et al. [39]	Friedkin–Johnsen dynamics	Sparse models	Passive	Infinite horizon
Ravazzi et al. [40], [150]	Friedkin–Johnsen dynamics	Sparse models	Passive	Random intermittent measurements
Coluccia et al. [41]	Friedkin–Johnsen dynamics	Distributed sparse models	Passive	Random intermittent measurements
Anderson et al. [43]	Friedkin–Johnsen dynamics	Sparse models	Passive	Finite/infinite horizon

THE INFLUENCE ESTIMATION PROBLEM: INFINITE HORIZON APPROACH

As the first approach to the problem of estimating the structure of a social network from infrequent data, consider the Friedkin–Johnsen model in (13) and assume there is knowledge of n 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 these data only, under the assumption of network sparsity. To this end, some model identifiability considerations must be made. This is accomplished in the next section.

Model Identifiability

First, note that due to the *consensus preservation* property discussed in "Simple Properties of the Friedkin–Johnsen Model," whenever initial opinions are at a consensus, final opinions are also in agreement. 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 Friedkin–Johnsen model reduces to DeGroot's representation, typically leading to a consensus of opinions. Clearly, the problem is not well posed in this case since there are infinitely many matrices leading the dynamics in (13) to the same value of consensus. This fact is illustrated in the following example, borrowed from [39].

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 [161] guarantees that

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

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, it is not influenced by any other agent. Hence, to avoid ambiguities, the remainder of this article assumes that $\lambda_i \neq 0$ for all $i \in \mathcal{V}$, $\mathbf{\Lambda} \neq \mathbf{I}_n$, and that 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 lead asymptotically to an equilibrium point that can be computed from weights, obstinacy levels, and initial opinions. It follows that recovering $\mathbf{W}$ amounts to solving the system of equations

$$\begin{cases} (\mathbf{I}_n - \mathbf{\Lambda}\mathbf{W})\mathbf{x}^{(\ell)}(\infty) = (\mathbf{I}_n - \mathbf{\Lambda})\mathbf{x}^{(\ell)}(0), \\ \mathbf{W}\mathbb{1} = \mathbb{1}, \\ \mathbf{W} \geq 0, \mathbf{\Lambda} \geq 0. \end{cases} \tag{14}$$

However, as shown in [39], this system contains an implicit ambiguity: if $(\mathbf{\Lambda}, \mathbf{W})$ is a solution of (14), then it is possible to construct a different solution pair $(\mathbf{\Lambda}', \mathbf{W})$ as

$$\begin{aligned} \mathbf{\Lambda}' &= \mathbf{I}_n - \mathbf{D}(\mathbf{I}_n - \mathbf{\Lambda}), \\ \operatorname{off}(\mathbf{\Lambda}'\mathbf{W}') &= \mathbf{D}\operatorname{off}(\mathbf{\Lambda}\mathbf{W}), \\ \operatorname{diag}(\mathbf{\Lambda}'\mathbf{W}') &= \mathbb{1} - \mathbf{D}((\mathbf{I}_n - \mathbf{\Lambda})\mathbb{1} + \operatorname{off}(\mathbf{\Lambda}\mathbf{W})\mathbb{1}), \end{aligned}$$

for any nonnegative diagonal matrix $\mathbf{D}$ with $[\mathbf{D}]_{ii} \in [0, 1]$. The ambiguity, which was noted in [92] in the setting of DeGroot models with stubborn agents, arises because 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 (14) is full rank, then the problem in (14) admits a unique solution and may be easily solved (for example, using linear programming and any solver for convex optimization [162]). 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 resolving the ambiguity by imposing constraints on $\operatorname{diag}(\mathbf{W})$].

Sparse Identification Problem

Motivated by the discussion in the previous section, the identification approach is based on the observation that a social network is typically sparse, in the sense that interactions among agents are few when compared to the network dimension. For a given $\mathbf{\Lambda}$, $\mathbf{X}(0)$ and $\mathbf{X}(\infty)$, this leads to estimating 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 [163]:

$$\min_{\mathbf{W}\in\mathbb{R}^{n\times n}} \|\mathbf{W}\|_0, \quad \text{s.t.} \begin{cases} \mathbf{\Phi}\mathbf{W}^\top = \mathbf{\Psi}^\top, \\ \mathbf{W}\mathbb{1} = \mathbb{1}, \\ \mathbf{W} \geq 0, \end{cases} \tag{15}$$

where $\|\mathbf{W}\|_0$ is the number of nonzeros of the matrix $\mathbf{W}$, $\mathbf{\Phi} \doteq \mathbf{X}(\infty)^\top$, and $\mathbf{\Psi} \doteq \mathbf{\Lambda}^{-1}[\mathbf{X}(\infty) - (\mathbf{I}_n - \mathbf{\Lambda})\mathbf{X}(0)]$. 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 independently learned from the others. More precisely,

$$\min_{w_j\in\mathbb{R}^n} \|w_j\|_0, \quad s.t. \begin{cases} \mathbf{\Phi}\mathbf{w}_j = \psi_j, \\ \mathbb{1}^\top \mathbf{w}_j = 1, \\ \mathbf{w}_j \geq 0, \end{cases} \tag{16}$$

where ψ_j is the jth 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 (13) and the existence of the final opinion profile. This property is automatically ensured if $\Lambda < \mathbf{I}$. If $\lambda_i < 1$ for at least one i, the stability property is *generic*: the set of matrices $\mathbf{W}$ such that $\Lambda\mathbf{W}$ has an eigenvalue at one has zero Lebesgue measure. In practice (taking into account inaccuracies in the opinion measurement), it is impossible to satisfy the first constraint in (14) unless the matrix $(\mathbf{I} - \Lambda\mathbf{W})$ is invertible. However, as should be noticed in the optimization problem (16), this constraint is not imposed in the recovery problem.

Recovery via Convex Optimization

The convex relaxation of (16) (where the constraint $\mathbf{w}_j \geq 0$ is removed)

$$\min_{\mathbf{w}_j \in \mathbb{R}^n} \|\mathbf{w}_j\|_1, \quad \text{s.t.} \begin{cases} \Phi \mathbf{w}_j = \psi_j, \\ \mathbb{1}^\top \mathbf{w}_j = 1, \end{cases} \tag{17}$$

can be formulated as a linear program and has been extensively studied (see "Compressed Sensing"). A large number of algorithms has been proposed to solve it efficiently, especially for the case where the dimension of the vector to be sparsified is high [34]. It is well known that under certain conditions on the matrix Φ, the number of measurements m, and the sparsity of $\mathbf{w}_j$, (16) and (17) have the same unique solution [163]. However, in the case of influence estimation in the social dynamical networks (SDNs) considered in this article and (contrary to other problems in compressed sensing) the sensing matrix Φ cannot be designed to satisfy the recovery properties mentioned in the preceding because it depends on the model's parameters:

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

Where the initial opinions are independent identically distributed Gaussian random variables and the agents are all "very stubborn" (Λ has small diagonal entries), one could (in principle) consider $\Phi \approx \mathbf{X}(0)^\top$, which would satisfy the restricted isometry (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 Λ is not close to zero, then Φ is a random variable whose entries are coupled, and available results through compressed sensing do not apply. The remainder of this section reviews the results in [39], where recovery conditions specific to SDNs are derived. Assume that the initial opinions $\mathbf{x}^{(\ell)}(0)$ on topic ℓ are independent identically distributed random variables having a Gaussian distribution with zero mean and unit variance. The hypothesis about the Gaussian distribution of the initial condition is a common assumption in opinion dynamics literature [164]. It can also be explained by the fact that initial opinions can be preaveraged 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 [165].

Moreover, the assumption about 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 $x^{(\ell)}(0) - \bar{x}^{(\ell)}(0)\mathbb{1}$ and $z^{(\ell)}(\infty) = \mathbf{V}(x^{(\ell)}(0) - \bar{x}^{(\ell)}(0)\mathbb{1})$, where $\mathbf{V}$ is the total effects matrix. Since the total effects matrix is stochastic, then $z^{(\ell)}(\infty) = x^{(\ell)}(\infty) - \bar{x}(0)^{(\ell)}\mathbb{1}$. Note that for the SDN 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 [39] it is proved 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 $\hat{\Sigma}_{SS} = \Phi_S^\top \Phi_S / m$, with a probability greater than $1 - 2e^{-m/32}$, we have

$$\frac{\mu_{\max}(\hat{\Sigma}_{SS})}{\mu_{\min}(\hat{\Sigma}_{SS})} \geq \frac{1}{3} \frac{\mu_{\max}(\Sigma_{SS})}{\mu_{\min}(\Sigma_{SS})},$$

where $\Sigma = (I - \Lambda\mathbf{W})^{-1}(I -)^2(I - \Lambda\mathbf{W})^\top$. For this reason, more powerful tools are required for the performance analysis of the problem. More precisely, the so-called null space property [163] is needed, which provides a necessary and sufficient condition for recovery. This property is summarized in "Necessary and Sufficient Conditions for Recovery."

These necessary and sufficient conditions on sparse recovery enable one to study when it is possible to recover sparse models for an SDN. It is shown in [39] that the initial condition $\mathbf{x}^{(\ell)}(0) \sim \mathcal{N}(0, \mathbf{I}_n)$, and if the number of considered topics satisfies

$$m \geq 4c \frac{(1 + \lambda_{\max})^2 (1 - \lambda_{\min})^2}{(1 - \lambda_{\max})^4} d_{\max} \log n, \tag{18}$$

then the solution to (17) is unique and coincides with that of (16) with a probability of at least $1 - c' e^{-c'' m}$, 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 (18) 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 ones, and network sensing performance should depend on the strength of the influencing power of prejudices.

Moreover (as conjectured in [92]), another important issue that affects reconstruction performance is the degree distribution in a 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, a Watts–Strogatz network [166]), while a network with a power law degree distribution (for example, a Barabasi–Albert network [167]) is more difficult to recover. To finalize the discussion in this section, recall that if Λ is not known, then the identification problem is not well posed. The ambiguity is due to missing information about the rate of social interactions. It cannot be removed without making additional assumptions. However, an invariant quantity can be determined among the ambiguous solutions by defining equivalence classes and resolving the ambiguity by imposing constraints on $\mathrm{diag}(\mathbf{W})$. More precisely, in [39], it is shown that the problem of learning sensitivity matrix $\boldsymbol{\Lambda}$ can be cast as in (17), with $\boldsymbol{\Phi} \doteq [\mathbf{X}(\infty)^\top, \mathbf{x}_j(0) - \mathbf{x}_j(\infty)]$ and $\mathbf{B} = \mathbf{X}(0)$, where $\mathbf{x}_j(\infty)$, $\mathbf{x}_j(0)$ are the column vectors corresponding to jth 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 in 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 about several topics to build the necessary information to render the problem identifiable. Even if the number of topics that is necessary to correctly identify a 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 opinions. At the same time, the technique does not require observations of opinions about different topics or perfect knowledge of interaction times, and it can be adapted to cases when some information is missing or partial.

More precisely, the availability of "intermittent" measurements of opinions is assumed to identify the dynamics of the evolution of opinions and, as a consequence, the influence matrix [151], [168]. Such an approach is especially useful in the case where not all opinions are updated at the same time and where a random sampling of opinions might be a less onerous way of estimating the behavior of the network. Hence, this section focuses on the asynchronous gossip-based Friedkin–Johnsen model and assumes that random measurements of opinions are available. For simplicity (as in [150]), consider a 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 (S5), 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:

$$\mathbf{z}(k) = \mathbf{P}(k)\mathbf{x}(k), \tag{19}$$

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 a known distribution representing which opinions are measured at time k.

Various probability distributions of the matrix $\mathbf{P}(k)$ lead to very different observation models. For example, if

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

Necessary and Sufficient Conditions for Recovery

The concept of the null space property (NSP) [163] is needed to derive more general conditions for sparse recovery.

NULL SPACE PROPERTY

The matrix $\boldsymbol{\Phi} \in \mathbb{R}^{m\times n}$ satisfies the NSP of order s if given

$$\mathcal{C}(\ell) = \{\eta \in \mathbb{R}^n : \| \,|\eta_{S^c}\|_1 \le \| \,|\eta_S\|_1\},$$
$$\mathcal{C}(\ell) \cap \mathrm{Ker}(\boldsymbol{\Phi}) = \{0\},$$

for all index sets S with $|S| \le s$, where Ker denotes the kernel of a matrix. Using this definition, Theorem 1 in [163] provides additional results for when one can recover sparse solutions from systems of linear equations. More precisely, consider matrix $\boldsymbol{\Phi} \in \mathbb{R}^{m\times n}$. Then, the optimization problem

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

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

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

if and only if $\boldsymbol{\Phi}$ satisfies the NSP with order $2s$. To further analyze when can sparse solutions can be recovered, introduce the concept of restricted eigenvalue criterion (REC).

Definition 1: Restricted Eigenvalue Criterion

A matrix $\boldsymbol{\Phi}$ satisfies the REC of order s if there exists a $\delta_s > 0$ such that

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

for all $\mathbf{z} \in \mathcal{C}(\ell)$ uniformly for all index sets $S \subseteq [n]$ with $|S| \le s$. It is straightforward to see that the REC is equivalent to the NSP. For random matrices $\boldsymbol{\Phi}$ with independent identically distributed entries drawn from particular distributions and for unitary matrices, it can be shown that, if enough measurements are available, then the REC condition is satisfied with the prescribed δ_s, with a probability close to one [S33].

REFERENCE

[S33] E. J. Candes and T. Tao, "Decoding by linear programming," *IEEE Trans. Inf. Theory*, vol. 51, no. 12, pp. 4203–4215, Dec. 2005. doi: 10.1109/TIT.2005.858979.

“Weak” ties facilitate the exchange of information among closed communities, enabling the mobility of labor and integration of individuals into political movements.

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 where no observations are observed. This model facilitates capturing a typical situation in which the actual rates at which interactions occur are not perfectly known (thus, the sampling time is different from the 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 [151], [168] is present, where opinions are independently observed with a probability of $\rho_i \in [0,1]$. In a case where observations are made with an equal probability of $\rho_i = \rho$ 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 $\rho \neq 1$, there is partial information.

This model has a clear interpretation for SDNs, describing a 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 [150], where the objective is as follows. Given the sequence of observation $\{\mathbf{z}(k)\}_{k=1}^{t}$, the estimate of the matrix $\mathbf{W}$ is referred to as $\widehat{\mathbf{W}}_t$. In [150], theoretical conditions are also provided for the number of samples that is sufficient to have an error not larger than a fixed tolerance ϵ with high probability. For the 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 4. To reconstruct the influence matrix, recall the definition of opinions' cross correlation matrix:

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

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

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

where $\bar{\boldsymbol{\Gamma}} = \bar{\boldsymbol{\Gamma}}(\boldsymbol{\Lambda},\mathbf{W})$ and $\bar{\boldsymbol{b}} = \bar{\boldsymbol{b}}(\boldsymbol{\Lambda},\mathbf{x}(0))$ are defined in (S7). Moreover, $\boldsymbol{\Sigma}^{[\ell]}(k)$ converges to $\boldsymbol{\Sigma}^{[\ell]}(\infty)$ for all nonnegative integers ℓ that satisfy

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

where $\boldsymbol{\Sigma}^{[\ell]}(\infty) := \lim_{k\to\infty}\boldsymbol{\Sigma}^{[\ell]}(k)$. The preceding simple linear equation provides the motivation for the approach described in this section, which 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 $\boldsymbol{\Sigma}^{[0]}(\infty)$ through $\boldsymbol{\Sigma}^{[\ell]}(\infty)$ for some ℓ. Given these estimates and using (21), estimate the matrix $\bar{\boldsymbol{\Gamma}}$. Finally, estimate the influence matrix $\mathbf{W}$ by exploiting the relation between $\bar{\boldsymbol{\Gamma}}$ and $\mathbf{W}$.

Estimating the Expected Opinion Profile and Cross Correlation Matrices

It is now shown how to exploit the model of observations and collected data to estimate an opinion's expectation 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 facilitates estimating the expectation of the opinions from available data. More precisely, begin by estimating $\mathbb{E}[\mathbf{z}(k)]$ using time averages

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

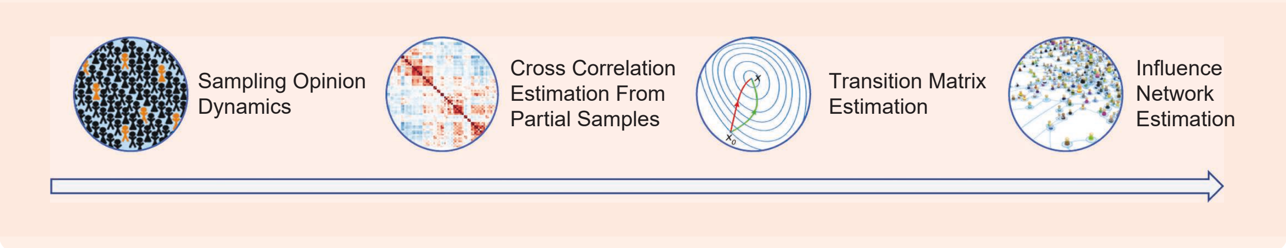


FIGURE 4 The main stream of the methodology.

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.

and obtain

$$\hat{x}_i(t) = \frac{\bar{z}_i(t)}{\pi_i}. \tag{22}$$

Estimating covariance matrices can be done in a similar way. The cross correlation matrices $\boldsymbol{\Sigma}^{[\ell]}(\infty)$ 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,

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

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

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

from which

$$\hat{\boldsymbol{\Sigma}}_{ij}^{[\ell]}(t) = \hat{\mathbf{S}}_{ij}^{[\ell]}(t)/\boldsymbol{\Pi}_{ij}^{[\ell]}. \tag{23}$$

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 tends to infinity. More precisely, in [150], a careful analysis of the procedures developed in the preceding shows that the estimates converge to the true values at a rate of $O(1/\sqrt{t})$, where t is the number of measurements. As an example of the estimation procedure, consider first the case of *independent homogeneous random sampling*. In it, $\pi = \rho$,

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

from which $\hat{x}(t) = \bar{z}(t)/\rho$, and

$$\hat{\boldsymbol{\Sigma}}_{ij}^{[\ell]}(t) = \frac{1}{\rho^2}\hat{\mathbf{S}}^{[\ell]}(t) - \left(\frac{1-\rho}{\rho^2}\hat{\mathbf{S}}^{[\ell]}(t) \circ \mathbf{I}_n\right)1(\ell = 0). \tag{24}$$

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

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

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

Estimating the Influence Matrix

In principle, the estimators $\hat{\boldsymbol{\Sigma}}^{[1]}(t)$ and $\hat{\boldsymbol{\Sigma}}^{[0]}(t)$ of $\boldsymbol{\Sigma}^{[1]}(t)$ and $\boldsymbol{\Sigma}^{[0]}(t)$, together with (21), can be used to estimate the dynamics matrix $\bar{\boldsymbol{\Gamma}}$. 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 previously 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 to be considered in the estimation of dynamics, and use a combination of these covariance matrices. More precisely, given estimates $\hat{\boldsymbol{\Sigma}}^{[\ell]}(t)$, compute

$$\hat{\boldsymbol{\Sigma}}_-(t) \doteq \frac{1}{N_\Sigma}\sum_{\ell=0}^{N_\Sigma - 1}\hat{\boldsymbol{\Sigma}}^{[\ell]}(t), \quad \hat{\boldsymbol{\Sigma}}_+(t) \doteq \frac{1}{N_\Sigma}\sum_{\ell=1}^{N_\Sigma}\hat{\boldsymbol{\Sigma}}^{[\ell]}(t), \tag{25}$$

and note that these matrices (approximately) satisfy (21). Hence, they can be used to estimate the structure of a network. Consider two types of networks. First, assume that one knows in advance that a network is dense. In this case, a possible estimator of $\bar{\boldsymbol{\Gamma}}$ can be obtained by directly solving the set of linear equations (21). In other words, the estimator is

$$\hat{\bar{\boldsymbol{\Gamma}}}(t)^\top = \hat{\boldsymbol{\Sigma}}_-(t)^\dagger(\hat{\boldsymbol{\Sigma}}_+(t) - \bar{\mathbf{x}}(t)\bar{\boldsymbol{b}}^\top). \tag{26}$$

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 the available information. The estimator can be obtained by solving

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

Estimating the Network Topology and Influence Matrix

Once an estimate of the average transition matrix $\bar{\boldsymbol{\Gamma}}(t)$ has been obtained, the topology of the influence network can be retrieved in a straightforward manner by noticing that $\operatorname{supp}(\bar{\boldsymbol{\Gamma}}) = \operatorname{supp}(\mathbf{W})$. Hence, we can reconstruct the support of $\mathbf{W}$ using the elements of the estimated matrix $\hat{\bar{\boldsymbol{\Gamma}}}$ 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:

$$\hat{\mathbf{W}}(t) = \hat{\mathbf{D}}\boldsymbol{\Lambda}^{-1}[\bar{\boldsymbol{\Gamma}}(t) - (1-\beta)\mathbf{I}_n - \beta\boldsymbol{\Lambda}(\mathbf{I}_n - \hat{\mathbf{D}}^{-1})],$$

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.

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

$$\hat{\mathbf{D}}_{i,i} = \| \mathrm{supp}(\gamma_i) \|_0,$$

with $\gamma_i^\top$ being the ith row of matrix $\hat{\bar{\Gamma}}$.

Influence Estimation in Multiplex Networks

For the model described in "Dynamics of Multiplex Networks," one can estimate the cross correlation matrices and then use relations (S10) and (S11) for each dynamical system, replacing the theoretical covariances $\Sigma_{[\ell]}^{(s)}(t)$ with estimated value $\hat{\Sigma}_{[\ell]}^{(s)}(t)$ (see the methodology summarized in Figure 4). Leveraging the estimation of autoregressive processes [151] and on the ergodicity of dynamical systems, it can be shown that with a probability of at least $1-\delta$,

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

where $C(n, \|\mathbf{Q}_\eta\|)$ is a constant independent of t (see "Performance of Influence Estimation: Asynchronous Gossip-Based Friedkin–Johnsen Model). This bound can be improved by imposing new constraints in the recovery by exploiting correlations among different dynamical systems (see models $\mathcal{M}_{cc}$ and $\mathcal{M}_{cs}$ in "Dynamics of Multiplex Networks"). If the correlations are not known among influence matrices, the idea proposed in [41] is to leverage global properties of the local processes to correct the local estimates of $\mathbf{S}_{[0]}^{(s)}(\infty)$. Moreover, the reconstruction performance suffers when the sample size is not large, as the quantity of observed data must be larger than the number of unknowns to have a full-rank estimation of $\hat{\Sigma}_{[0]}^{(s)}(\infty)$. The Bayesian approach is a powerful estimation framework since it combines prior probabilistic information (parametrized by some unknown hyperparameters) and gathered observations (see "Bayesian Estimation of $\mathbf{S}_{[0]}^{(s)}(\infty)$).

In [41], the performance of the proposed estimators is tested within $\mathcal{M}_{cc}$ and $\mathcal{M}_{cs}$. 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 suffers when the sample size is not large. In this sense, the Bayesian method acts as a regularizer of

Performance of Influence Estimation: Asynchronous Gossip-Based Friedkin–Johnsen Model

The estimation error on matrix $\bar{\Gamma}(t)$ is based on the previous estimation of the cross correlation matrices. Using (26), $\hat{\Sigma}_+$ must be inverted, and the estimation error depends on the singular values of $\hat{\Sigma}_\pm$. It can be shown that, with a probability of at least $1-\delta$,

$$\left\| \bar{\Gamma}(t) - \hat{\bar{\Gamma}}(t) \right\|_2 = O\left(\frac{n(\sigma_{\max}^+ + n)}{(\underline{\sigma}_{\min}^-)^2 \Pi^* \sqrt{\delta(t+1)\beta(1-\lambda_{\max})}} \right), \quad \text{(S12)}$$

where $\sigma_{\max}^+ = \|\Sigma_+\|_2$ and $\underline{\sigma}_{\min}^- \doteq \min(\sigma_{\min}^-, \hat{\sigma}_{\min}^-)$ (where $\sigma_{\min}^-$ and $\hat{\sigma}_{\min}^-$ are the minimum singular value of Σ_- and $\hat{\Sigma}_-$, respectively). We refer the reader to [150] for details.

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

In the absence of additional information about a 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 Ψ and $\nu > n+1$ degrees or, equivalently,

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

where the following are true:

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

The inverse Wishart parameter estimations are obtained via alternating minimization:

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

We refer the reader to [41] for details.

Unlike simplified mathematical models, people do not broadcast numbers; they communicate via web forums, microblogs, mobile apps, and other social media.

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 (the Bernstein–von Mises theorem), thus converging to the classical sample covariance matrix. 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 among individuals is not a trivial task. How to understand which connections between people are most essential and who are the genuine leaders of a group is a difficult problem. Granovetter [51], [169] 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 ones build bridges among densely knit communities. This principle led to a number of mathematical characteristics [46] 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* [169]. "Weak" ties facilitate the exchange of information among closed communities, enabling the mobility of labor and integration of individuals into political movements. Hence, "static" characteristics considering only links among individuals and ignoring specific features of their interactions can be misleading.

Alternative methods are needed that consider a social network as a dynamical system. This survey focused on two novel directions of research concerned with dynamical networks of social influence. The statistical approach adopted in machine learning considers a social network a probabilistic graphical model treating social influence as a measure of statistical correlation among data produced by individuals (for example, information about events they attend and goods they consume). The approach in social influence network theory [32] considers social influence as a process that alters opinions of 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 the 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, the 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] and continuous-opinion–discrete-action models [170]. 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 network 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 [171] and [172].

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; they communicate via web forums, microblogs, mobile apps, and other social media. 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 this survey will help to recruit young, talented researchers to the vibrant and fascinating area of dynamical SNA.

ACKNOWLEDGMENTS

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 was an outstanding scientist in control and a coauthor of many articles on this topic. He promoted interdisciplinary research through his positive and encouraging attitude toward spreading and sharing scientific results. Anton Proskurnikov acknowledges financial support from the Russian Foundation for Basic Research, under grant 20-01-00619.

AUTHOR INFORMATION

Chiara Ravazzi 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, and 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 National Research Council of Italy Institute of Electronics, Information Engineering, and Telecommunications, Turin, 10129, Italy, working with the Systems Modeling and Control Group. She has served as associate editor of *IEEE Transactions on Signal Processing* and *IEEE Control Systems Letters*. Her research interests include signal processing, optimization, and control of network systems. She is a Member of IEEE.

Fabrizio Dabbene (fabrizio.dabbene@ieiit.cnr.it) is a director of research at the National Research Council of Italy Institute of Electronics, Information Engineering, and Telecommunications, Turin, 10129, Italy, where he is the coordinator of the Information and Systems Technology Group. He has held visiting and research positions at the University of Iowa, Pennsylvania State University, and the Russian Academy of Sciences, Institute of Control Science. He has published more than 100 research articles and two books. He served as associate editor of *Automatica* (2008–2014) and *IEEE Transactions on Automatic Control* (2008–2012), and he is currently senior editor of *IEEE Control Systems Society Letters*. He was a member of the IEEE Control Systems Society (CSS) Board of Governors (2014–2016), served as CSS vice president for Publications (2015–2016), and currently chairs the CSS Italy Chapter. He is a Senior Member of IEEE.

Constantino M. Lagoa received the Ph.D. degree from the University of Wisconsin at Madison in 1998. He joined the Department of Electrical Engineering, Pennsylvania State University, State College, Pennsylvania, 16801, USA, 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 the control of computer networks. He 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*. He is a Member of IEEE.

Anton V. Proskurnikov is an assistant professor in the Department of Electronics and Telecommunications, Politecnico di Torino, Turin, 10129, Italy, and a part-time senior researcher at the Institute for Problems in Mechanical Engineering, Russian Academy of Sciences, St. Petersburg, Russia. 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). He is an associate editor of *IEEE Transactions on Automatic Control* and *Journal of Mathematical Sociology*. His research interests include network science, multiagent systems and technologies, nonlinear dynamics, and control applications in social and biological sciences. He is a Senior Member of IEEE.

REFERENCES

[1] J. Tang, J. Sun, C. Wang, and Z. Yang, "Social influence analysis in large-scale networks," in *Proc. 15th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining (KDD '09)*, 2009, pp. 807–816. doi: 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. doi: 10.1126/science.1230456.

[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," *J. Economic Perspectives*, vol. 31, no. 2, pp. 211–236, May 2017. doi: 10.1257/jep.31.2.211.

[5] M. McPherson, L. Smith-Lovin, and J. M. Cook, "Birds of a feather: Homophily in social networks," *Annu. Rev. Sociol.*, vol. 27, no. 1, pp. 415–444, 2001. doi: 10.1146/annurev.soc.27.1.415.

[6] P. Dandekar, A. Goel, and D. T. Lee, "Biased assimilation, homophily, and the dynamics of polarization," *Proc. Nat. Academy Sci.*, vol. 110, no. 15, pp. 5791–5796, 2013. doi: 10.1073/pnas.1217220110.

[7] P. Singla and M. Richardson, "Yes, there is a correlation: From social networks to personal behavior on the web," in *Proc. 17th Int. Conf. World Wide Web (WWW '08)*, 2008, pp. 655–664. doi: 10.1145/1367497.1367586.

[8] J. Bughin, J. Doogan, and O. J. Vetvik, "A new way to measure word-of-mouth marketing," *McKinsey Quart.*, vol. 2, pp. 113–116, Apr. 2010.

[9] S. Wasserman and K. Faust, *Social Network Analysis in the Social and Behavioral Sciences (Structural Analysis in the Social Sciences)*. New York: Cambridge Univ. Press, 1994, pp. 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 Rev.*, vol. 45, no. 2, pp. 167–256, 2003. doi: 10.1137/S003614450342480.

[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, no. 2, pp. 591–646, 2009. doi: 10.1103/RevModPhys.81.591.

[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," *Annu. Rev. Control*, vol. 43, pp. 65–79, 2017. doi: 10.1016/j.arcontrol.2017.03.002.

[16] A. Proskurnikov and R. Tempo, "A tutorial on modeling and analysis of dynamic social networks. Part II," *Annu. Rev. Control*, vol. 45, 2018. doi: 10.1016/j.arcontrol.2018.03.005.

[17] Y. Dong, M. Zhan, G. Kou, Z. Ding, and H. Liang, "A survey on the fusion process in opinion dynamics," *Inform. Fusion*, vol. 43, pp. 57–65, Sept. 2018. doi: 10.1016/j.inffus.2017.11.009.

[18] O. Abid, S. Jamoussi, and Y. B. Ayed, "Deterministic models for opinion formation through communication: A survey," *Online Soc. Netw. Media*, vol. 6, pp. 1–17, June 2018. doi: 10.1016/j.osnem.2018.02.002.

[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. doi: 10.1109/ACCESS.2019.2913787.

[20] R. Abelson, "Mathematical models in social psychology," in *Advances in Experimental Social Psychology*, vol. 3, L. Berkowitz, Ed. New York: Academic Press, 1967, pp. 1–49.

[21] J. Hunter, J. Danes, and S. Cohen, *Mathematical Models of Attitude Change*, vol. 1. New York: Academic Press, 1984.

[22] P. Bolzern, P. Colaneri, and G. Nicolao, "Opinion influence and evolution in social networks: A Markovian agents model," *Automatica*, vol. 100, pp. 219–230, Feb. 2019. doi: 10.1016/j.automatica.2018.11.023.

[23] R. Hegselmann, S. König, S. Kurz, C. Niemann, and J. Rambau, "Optimal opinion control: The campaign problem," *J. Artif. Soc. Soc. Simul.*, vol. 18, no. 3, p. 18, 2015. doi: 10.18564/jasss.2847.
[24] N. Masuda, "Opinion control in complex networks," *New J. Phys.*, vol. 17, no. 3, p. 033031, Mar. 2015. doi: 10.1088/1367-2630/17/3/033031.
[25] J. Zhao, Q. Liu, L. Wang, and X. Wang, "Competitiveness maximization on complex networks," *IEEE Trans. Syst., Man, Cybern., Syst.*, vol. 48, no. 7, pp. 1054–1064, 2018. doi: 10.1109/TSMC.2016.2636240.
[26] M. Grabisch, A. Mandel, A. Rusinowska, and E. Tanimura, "Strategic influence in social networks," *Math. Oper. Res.*, vol. 43, no. 1, pp. 29–50, 2018. doi: 10.1287/moor.2017.0853.
[27] D. Lazer et al., "Computational social science," *Science*, vol. 323, no. 5915, pp. 721–723, 2009. doi: 10.1126/science.1167742.
[28] B. Liu, *Sentiment Analysis and Opinion Mining*. San Rafael, CA: Morgan & Claypool, 2012.
[29] M. Tsitsvero, S. Barbarossa, and P. D. Lorenzo, "Signals on graphs: Uncertainty principle and sampling," *IEEE Trans. Signal Process.*, vol. 64, no. 18, pp. 4845–4860, Sept. 2016. doi: 10.1109/TSP.2016.2573748.
[30] G. Puy, N. Tremblay, R. Gribonval, and P. Vandergheynst, "Random sampling of bandlimited signals on graphs," *Appl. Comput. Harmon. Anal.*, vol. 44, no. 2, pp. 446–475, Mar. 2018.
[31] N. E. Friedkin and E. C. Johnsen, "Social influence networks and opinion change,"in *Advances in Group Processes*, vol. 16, S. R. Thye, E. J. Lawler, M. W. Macy, and H. A. Walker, Eds. Bingley, U.K.: Emerald Publishing Limited, 1999, pp. 1–29.
[32] N. E. Friedkin and E. C. Johnsen, *Social Influence Network Theory: A Sociological Examination of Small Group Dynamics*. Cambridge, U.K.: Cambridge Univ. Press, 2011.
[33] H. Abdi and L. Williams, "Principal component analysis," *Wiley Interdiscipl. Rev., Comput. Statist.*, vol. 2, no. 4, pp. 433–459, 2010. doi: 10.1002/wics.101.
[34] Y. Eldar and G. Kutyniok, *Compressed Sensing: Theory and Applications, ser. Compressed Sensing: Theory and Applications*. New York: Cambridge Univ. Press, 2012.
[35] D. Yan, Y. Tian, and J. Cheng, *Systems for Big Graph Analytics* (Springer Briefs in Computer Science). Berlin: Springer-Verlag, 2017.
[36] P. Frasca, C. Ravazzi, R. Tempo, and H. Ishii, "Gossips and prejudices: Ergodic randomized dynamics in social networks," in *Proc. IFAC NecSys 2013 Workshop*, Koblenz, Germany, 2013, pp. 212–219. doi: 10.3182/20130925-2-DE-4044.00024.
[37] 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. doi: 10.1016/j.ejcon.2015.04.002.
[38] C. Ravazzi, P. Frasca, R. Tempo, and H. Ishii, "Ergodic randomized algorithms and dynamics over networks," *IEEE Trans. Control Netw. Syst.*, vol. 2, no. 1, pp. 78–87, 2015. doi: 10.1109/TCNS.2014.2367571.
[39] C. Ravazzi, R. Tempo, and F. Dabbene, "Learning influence structure in sparse social networks," *IEEE Trans. Control Netw. Syst.*, vol. 5, no. 4, pp. 1976–1986, Dec. 2018. doi: 10.1109/TCNS.2017.2781367.
[40] C. Ravazzi, S. Hojjatinia, C. M. Lagoa, and F. Dabbene, "Randomized opinion dynamics over networks: Influence estimation from partial observations," in *Proc. IEEE Conf. Decision Control (CDC)*, 2018, pp. 2452–2457. doi: 10.1109/CDC.2018.8619770.
[41] A. Coluccia, F. Dabbene, and C. Ravazzi, "Bayesian identification of distributed vector autoregressive processes," in *Proc. 18th European Control Conf. (ECC)*, 2019, pp. 842–847. doi: 10.23919/ECC.2019.8796302.
[42] 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. doi: 10.1109/TAC.2016.2613905.
[43] B. D. Anderson, F. Dabbene, A. V. Proskurnikov, C. Ravazzi, and M. Ye, "Dynamical networks of social influence: Modern trends and perspectives," in *Proc. IFAC World Congr. 2020*, 2020, vol. 53, no. 2, pp. 17,616–17,627. doi: 10.1016/j.ifacol.2020.12.2684.
[44] N. E. Friedkin, *A Structural Theory of Social Influence*. New York: Cambridge Univ. Press, 1998.
[45] P. V. Marsden and N. E. Friedkin, "Network studies of social influence," *Sociol. Meth. Res.*, vol. 22, no. 1, pp. 127–151, 1993. doi: 10.1177/0049124193022001006.
[46] 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: Springer U.S., 2011, pp. 177–214.
[47] J. G. March, "Measurement concepts in the theory of influence," *J. Politics*, vol. 19, no. 2, pp. 202–226, 1957. doi: 10.2307/2127196.
[48] H. A. Simon, "Notes on the observation and measurement of political power," *J. Politics*, vol. 15, no. 4, pp. 500–516, 1953. doi: 10.2307/2126538.
[49] J. French, Jr., "A formal theory of social power," *Psychol. Rev.*, vol. 63, no. 3, pp. 181–194, 1956. doi: 10.1037/h0046123.
[50] 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.
[51] M. S. Granovetter, "The strength of weak ties," *Amer. J. Sociol.*, vol. 78, no. 6, pp. 1360–1380, 1973. doi: 10.1086/225469.
[52] M. I. Jordan, "Graphical models," *Statist. Sci.*, vol. 19, no. 1, pp. 140–155, Feb. 2004. doi: 10.1214/088342304000000026.
[53] E. M. Airoldi, "Getting started in probabilistic graphical models," *PLOS Comput. Biol.*, vol. 3, no. 12, pp. 1–5, Dec. 2007. doi: 10.1371/journal.pcbi.0030252.
[54] D. S. Sade, "Sociometrics of Macaca Mulatta III: N-path centrality in grooming networks," *Soc. Netw.*, vol. 11, no. 3, pp. 273–292, 1989. doi: 10.1016/0378-8733(89)90006-3.
[55] L. C. Freeman, "Centrality in social networks: Conceptual clarification," *Soc. Netw.*, vol. 1, no. 3, pp. 215–239, 1979. doi: 10.1016/0378-8733(78)90021-7.
[56] L. C. Freeman, "A set of measures of centrality based upon betweenness," *Sociometry*, vol. 40, no. 1, pp. 35–41, 1977. doi: 10.2307/3033543.
[57] D. R. White and S. P. Borgatti, "Betweenness centrality measures for directed graphs," *Soc. Netw.*, vol. 16, no. 4, pp. 335–346, 1994. doi: 10.1016/0378-8733(94)90015-9.
[58] S. Brin and L. Page, "The anatomy of a large-scale hypertextual web search engine," *Comput. Netw. ISDN Syst.*, vol. 30, nos. 1–7, pp. 107–117, 1998. doi: 10.1016/S0169-7552(98)00110-X.
[59] L. Katz, "A new status index derived from sociometric analysis," *Psychometrika*, vol. 18, no. 1, pp. 39–43, Mar. 1953. doi: 10.1007/BF02289026.
[60] P. Bonacich, "Power and centrality: A family of measures," *Amer. J. Sociol.*, vol. 92, no. 5, pp. 1170–1182, 1987. doi: 10.1086/228631.
[61] P. Bonacich and P. Lloyd, "Eigenvector-like measures of centrality for asymmetric relations," *Soc. Netw.*, vol. 23, no. 3, pp. 191–201, 2001. doi: 10.1016/S0378-8733(01)00038-7.
[62] N. Friedkin, "Theoretical foundations for centrality measures," *Amer. J. Sociol.*, vol. 96, no. 6, pp. 1478–1504, 1991. doi: 10.1086/229694.
[63] J. Leskovec and R. Sosič, "Snap: A general-purpose network analysis and graph-mining library," *ACM Trans. Intell. Syst. Technol. (TIST)*, vol. 8, no. 1, p. 1, 2016. doi: 10.1145/2898361.
[64] J. Ellson et al., "Graphviz? open source graph drawing tools," in *Lecture Notes in Computer Science*. Berlin: Springer-Verlag, 2001, pp. 483–484.
[65] M. Scholz, "Node similarity as a basic principle behind connectivity in complex networks," *J. Data Mining Digital Humanities*, vol. 2015, Jan. 2015. doi: 10.46298/jdmdh.33.
[66] F. Farris, "The Gini index and measures of inequality," *Amer. Math. Monthly*, vol. 117, pp. 851–864, Dec. 2010.
[67] 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. doi: 10.1016/j.ins.2018.05.044.
[68] J. Moreno and H. Jennings, "Statistics of social configurations," *Sociometry*, vol. 1, no. 3/4, pp. 324–374, 1938. doi: 10.2307/2785588.
[69] 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," *Comput. Commun.*, vol. 76, pp. 26–41, Feb. 2016. doi: 10.1016/j.comcom.2015.09.028.
[70] J. F. Bokunewicz and J. Shulman, "Influencer identification in Twitter networks of destination marketing organizations," *J. Hospitality Tourism Technol.*, vol. 8, no. 2, pp. 205–219, 2017. doi: 10.1108/JHTT-09-2016-0057.
[71] T. Opsahl and P. Panzarasa, "Clustering in weighted networks," *Soc. Netw.*, vol. 31, no. 2, pp. 155–163, 2009. doi: 10.1016/j.socnet.2009.02.002.
[72] L. Liu, J. Tang, J. Han, and S. Yang, "Learning influence from heterogeneous social networks," *Data Mining Knowl. Discovery*, vol. 25, no. 3, pp. 511–544, Nov. 2012. doi: 10.1007/s10618-012-0252-3.
[73] M. T. Thai, W. Wu, and H. Xiong, Eds., *Big Data in Complex and Social Networks*, 1st ed. Boca Raton, FL: Chapman & Hall/CRC, 2016.
[74] A. Coluccia, "Regularized covariance matrix estimation via empirical bayes," *IEEE Signal Process. Lett.*, vol. 22, no. 11, pp. 2127–2131, 2015. doi: 10.1109/LSP.2015.2462724.
[75] 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," *Sociol. Sci.*, vol. 3, pp. 444–472, June 2016. doi: 10.15195/v3.a20.
[76] C. Cooley, *Human Nature and the Social Order*. New York: Scribner, 1922. [Online]. Available: https://books.google.mw/books?id=cVDxKsGmxggC
[77] N. E. Friedkin, "A formal theory of reflected appraisals in the evolution of power," *Administ. Sci. Quart.*, vol. 56, no. 4, pp. 501–529, 2011. doi: 10.1177/0001839212441349.
[78] X. Dong, D. Thanou, M. Rabbat, and P. Frossard, "Learning graphs from data: A signal representation perspective," *IEEE Signal Process. Mag.*, vol. 36, no. 3, pp. 44–63, May 2019. doi: 10.1109/MSP.2018.2887284.
[79] D. Koller and N. Friedman, *Probabilistic Graphical Models: Principles and Techniques – Adaptive Computation and Machine Learning*. Cambridge, MA: MIT Press, 2009.
[80] H. Rue and L. Held, *Gaussian Markov Random Fields: Theory and Applications (Monographs on Statistics and Applied Probability*. Boca Raton, FL: Chapman & Hall/CRC, 2005.
[81] M. J. Wainwright and M. I. Jordan, "Graphical models, exponential families, and variational inference," *Found. Trends Mach. Learn.*, vol. 1, nos. 1–2, pp. 1–305, Jan. 2008. doi: 10.1561/2200000001.
[82] G. Robins, P. Pattison, Y. Kalish, and D. Lusher, "An introduction to exponential random graph (ρ*) models for social networks," *Soc. Netw.*, vol. 29, no. 2, pp. 173–191, May 2007.
[83] G. B. Giannakis, Y. Shen, and G. V. Karanikolas, "Topology identification and learning over graphs: Accounting for nonlinearities and dynamics," *Proc. IEEE*, vol. 106, no. 5, pp. 787–807, May 2018. doi: 10.1109/JPROC.2018.2804318.
[84] R. Mazumder and T. Hastie, "The graphical lasso: New insights and alternatives," *Electron. J. Statist.*, vol. 6, pp. 2125–2149, Aug. 2012. doi: 10.1214/12-EJS740.
[85] 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. doi: 10.1109/MSP.2012.2235192.
[86] X. Dong, D. Thanou, P. Frossard, and P. Vandergheynst, "Learning Laplacian matrix in smooth graph signal representations," *IEEE Trans. Signal Process.*, vol. 64, no. 23, pp. 6160–6173, Dec. 2016. doi: 10.1109/TSP.2016.2602809.
[87] S. P. Chepuri, S. Liu, G. Leus, and A. O. Hero, "Learning sparse graphs under smoothness prior," in *Proc. IEEE Int. Conf. Acoustics, Speech Signal Process.*, 2017, pp. 6508–6512.
[88] X. Zhang, X. Dong, and P. Frossard, "Learning of structured graph dictionaries," in *Proc. IEEE Int. Conf. Acoustics, Speech Signal Process. (ICASSP)*, Mar. 2012, pp. 3373–3376.
[89] J. Mei and J. M. F. Moura, "Signal processing on graphs: Causal modeling of unstructured data," *IEEE Trans. Signal Process.*, vol. 65, no. 8, pp. 2077–2092, 2017. doi: 10.1109/TSP.2016.2634543.
[90] A. Bolstad, B. D. V. Veen, and R. Nowak, "Causal network inference via group sparse regularization," *IEEE Trans. Signal Process.*, vol. 59, no. 6, pp. 2628–2641, 2011. doi: 10.1109/TSP.2011.2129515.
[91] D. Bertsimas, V. Gupta, and I. C. Paschalidis, "Data-driven estimation in equilibrium using inverse optimization," *Math. Programming*, vol. 153, no. 2, pp. 595–633, Nov. 2015. doi: 10.1007/s10107-014-0819-4.
[92] H. Wai, A. Scaglione, and A. Leshem, "Active sensing of social networks," *IEEE Trans. Signal Inform. Process. Netw.*, vol. 2, no. 3, pp. 406–419, 2016. doi: 10.1109/TSIPN.2016.2555785.
[93] N. Friedkin and E. Johnsen, *Social Influence Network Theory*. New York: Cambridge Univ. Press, 2011.
[94] N. H. Anderson, *Foundations of Information Integration Theory*. New York: Academic Press, 1981.
[95] L. Festinger, *A Theory of Cognitive Dissonance*. Stanford, CA: Stanford Univ. Press, 1957.
[96] 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.
[97] M. Jackson, Ed., *Social and Economic Networks*. Princeton and Oxford: Princeton Univ. Press, 2008.
[98] F. Harary, R. Norman, and D. Cartwright, "Structural Models," *An Introduction to the Theory of Directed Graphs*. New York: Wiley, 1965.
[99] M. H. DeGroot, "Reaching a consensus," *J. Amer. Statistical Assoc.*, vol. 69, no. 345, pp. 118–121, 1974. doi: 10.1080/01621459.1974.10480137.
[100] K. Lehrer, "When rational disagreement is impossible," *Noûs*, vol. 10, no. 3, pp. 327–332, 1976. doi: 10.2307/2214612.
[101] K. Lehrer and C. Wagner, *Rational Consensus in Science and Society*. London: D. Reidel, 1981.
[102] N. E. Friedkin, A. V. Proskurnikov, W. Mei, and F. Bullo, "Mathematical structures in group decision-making on resource allocation distributions," *Sci. Rep.*, vol. 9, no. 1, pp. 1377, 2019. doi: 10.1038/s41598-018-37847-2.
[103] L. Li, A. Scaglione, A. Swami, and Q. Zhao, "Consensus, polarization and clustering of opinions in social networks," *IEEE J. Select. Areas Commun.*, vol. 31, no. 6, pp. 1072–1083, 2013. doi: 10.1109/JSAC.2013.130609.
[104] N. Friedkin, "A formal theory of social power," *J. Math. Sociol.*, vol. 12, no. 2, pp. 103–126, 1986. doi: 10.1080/0022250X.1986.9990008.
[105] J. Wolfowitz, "Products of indecomposable, aperiodic, stochastic matrices," *Proc. Amer. Math. Soc.*, vol. 14, no. 5, pp. 733–737, 1963. doi: 10.1090/S0002-9939-1963-0154756-3.
[106] F. Gantmacher, *The Theory of Matrices*, vol. 2. New York, NY: AMS Chelsea, 2000.
[107] R. Horn and C. Johnson, *Topics in Matrix Analysis*. New York: Cambridge Univ. Press, 1991.
[108] W. Ren and R. Beard, *Distributed Consensus in Multi-Vehicle Cooperative Control: Theory and Applications*. London: Springer-Verlag, 2008.
[109] 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
[110] 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.
[111] R. Hegselmann and U. Krause, "Opinion dynamics and bounded confidence, models, analysis and simulation," *J. Artif. Soc. Soc. Simul.*, vol. 5, no. 3, p. 2, 2002.
[112] G. Deffuant, D. Neau, F. Amblard, and G. Weisbuch, "Mixing beliefs among interacting agents," *Adv. Complex Syst.*, vol. 3, nos. 1–4, pp. 87–98, 2000. doi: 10.1142/S0219525900000078.
[113] 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.
[114] D. Urbig, "Attitude dynamics with limited verbalisation capabilities," *J. Artif. Soc. Soc. Simul. (JASSS)*, vol. 6, no. 1, p. 2, 2003.
[115] A. Fläche and M. Macy, "Small worlds and cultural polarization," *J. Math. Sociol.*, vol. 35, nos. 1–3, pp. 146–176, 2011. doi: 10.1080/0022250X.2010.532261.
[116] 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. doi: 10.1371/journal.pcbi.1000959.
[117] 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. doi: 10.1371/journal.pone.0074516.
[118] S. Motsch and E. Tadmor, "Heterophilious dynamics enhances consensus," *SIAM Rev.*, vol. 56, no. 4, pp. 577–621, 2013. doi: 10.1137/120901866.
[119] 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. doi: 10.1137/16M1083402.
[120] J. Jin, Y. Yuan, W. Pan, C. Tomlin, A. A. Webb, and J. Gonçalves, "Identification of nonlinear sparse networks using sparse Bayesian learning," in *Proc. IEEE 56th Annu. Conf. Decision Control (CDC)*, 2017, pp. 6481–6486. doi: 10.1109/CDC.2017.8264636.
[121] N. M. Mangan, S. L. Brunton, J. L. Proctor, and J. N. Kutz, "Inferring biological networks by sparse identification of nonlinear dynamics," *IEEE Trans. Mol. Biol. Multi-Scale Commun.*, vol. 2, no. 1, pp. 52–63, 2016. doi: 10.1109/TMBMC.2016.2633265.
[122] E. Kaiser, J. N. Kutz, and S. L. Brunton, "Sparse identification of nonlinear dynamics for model predictive control in the low-data limit," *Proc. Roy. Soc. A: Math., Phys. Eng. Sci.*, vol. 474, no. 2219, p. 20180335, 2018. doi: 10.1098/rspa.2018.0335.
[123] C. Altafini, "Dynamics of opinion forming in structurally balanced social networks," *PLoS ONE*, vol. 7, no. 6, p. e38135, 2012. doi: 10.1371/journal.pone.0038135.
[124] C. Altafini, "Consensus problems on networks with antagonistic interactions," *IEEE Trans. Autom. Control*, vol. 58, no. 4, pp. 935–946, 2013.

[125] Z. Cao et al., "Rebels lead to the doctrine of the mean: A heterogeneous DeGroot model," *J. Syst. Sci. Complexity*, vol. 31, no. 6, pp. 1498–1509, 2018. doi: 10.1007/s11424-018-7136-6.
[126] 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. doi: 10.1109/TCNS.2015.2437528.
[127] 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. doi: 10.1109/TAC.2017.2700523.
[128] W. Xia and M. Cao, "Clustering in diffusively coupled networks," *Automatica*, vol. 47, no. 11, pp. 2395–2405, 2011. doi: 10.1016/j.automatica.2011.08.043.
[129] G. Shi, A. Proutiere, M. Johansson, J. Baras, and K. Johansson, "Emergent behaviors over signed random dynamical networks: State-flipping model," *IEEE Trans. Control Netw. Syst.*, vol. 2, no. 2, pp. 142–153, 2015. doi: 10.1109/TCNS.2014.2378915.
[130] G. Shi, A. Proutiere, M. Johansson, J. Baras, and K. Johansson, "The evolution of beliefs over signed social networks," *Oper. Res.*, vol. 64, no. 3, pp. 585–604, 2016.
[131] G. Shi, C. Altafini, and J. Baras, "Dynamics over signed networks," *SIAM Rev.*, vol. 61, no. 2, pp. 229–257, 2019. doi: 10.1137/17M1134172.
[132] C. Hovland, I. Janis, and H. Kelley, *Communication and Persuasion*. New Haven: Yale Univ. Press, 1953.
[133] 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. doi: 10.1371/journal.pone.0157948.
[134] A. Proskurnikov, A. Matveev, and M. Cao, "Consensus and polarization in Altafini's model with bidirectional time-varying network topologies," in *Proc. IEEE CDC 2014*, Los Angeles, CA, 2014, pp. 2112–2117. doi: 10.1109/CDC.2014.7039710.
[135] A. Proskurnikov and M. Cao, "Opinion dynamics using Altafini's model with a time-varying directed graph," in *Proc. IEEE ISIC 2014 (Part of IEEE MSC 2014)*, Antibes, 2014, pp. 849–854. doi: 10.1109/ISIC.2014.6967611.
[136] D. Meng, "Bipartite containment tracking of signed networks," *Automatica*, vol. 79, pp. 282–289, May 2017. doi: 10.1016/j.automatica.2017.01.044.
[137] D. Meng, M. Du, and Y. Wu, "Extended structural balance theory and method for cooperative-antagonistic networks," *IEEE Trans. Autom. Control*, vol. 65, no. 5, pp. 2147–2154, 2020. doi: 10.1109/TAC.2019.2937759.
[138] 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. doi: 10.1109/TAC.2015.2471655.
[139] 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. doi: 10.1287/moor.1120.0570.
[140] C. Childress and N. Friedkin, "Cultural reception and production," *Amer. Soc. Rev.*, vol. 77, no. 1, pp. 45–68, 2012. doi: 10.1177/0003122411428153.
[141] 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. doi: 10.1016/j.ifacol.2017.08.1424.
[142] 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.
[143] B. Gawronski and F. Strack, Eds., *Cognitive Consistency: A Fundamental Principle in Social Cognition*. New York: Guilford Press, 2012.
[144] H. Noorazar, M. J. Sottile, and K. R. Vixie, "An energy-based interaction model for population opinion dynamics with topic coupling," *Int. J. Modern Phys. C*, vol. 29, no. 11, p. 1850115, 2018. doi: 10.1142/S0129183118501152.
[145] F. Xiong, Y. Liu, L. Wang, and X. Wang, "Analysis and application of opinion model with multiple topic interactions," *Chaos: An Interdisciplinary J. Nonlinear Sci.*, vol. 27, no. 8, p. 083113, 2017. doi: 10.1063/1.4998736.
[146] H. Noorazar, "Recent advances in opinion propagation dynamics: A 2020 survey," *Eur. Phys. J. Plus*, vol. 135, no. 6, p. 521, 2020. doi: 10.1140/epjp/s13360-020-00541-2.
[147] 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. IEEE Conf. Decision Control (CDC)*, 2015, pp. 3475–3480. doi: 10.1109/CDC.2015.7402757.
[148] 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. doi: 10.1126/science.aag2624.
[149] S. Boyd, A. Ghosh, B. Prabhakar, and D. Shah, "Randomized gossip algorithms," *IEEE Trans. Inform. Theory*, vol. 52, no. 6, pp. 2508–2530, 2006. doi: 10.1109/TIT.2006.874516.
[150] C. Ravazzi, S. Hojjatinia, C. M. Lagoa, and F. Dabbene, "Ergodic opinion dynamics over networks: Learning influences from partial observations," *IEEE Trans. Autom. Control*, vol. 66, no. 6, pp. 2709–2723, June 2021. doi: 10.1109/TAC.2021.3056362.
[151] M. Rao, T. Javidi, Y. C. Eldar, and A. Goldsmith, "Estimation in autoregressive processes with partial observations," in *Proc. IEEE Int. Conf. Acoustics, Speech Signal Process. (ICASSP)*, 2017, pp. 4212–4216. doi: 10.1109/ICASSP.2017.7952950.
[152] X. Wu, H. Wai, and A. Scaglione, "Estimating social opinion dynamics models from voting records," *IEEE Trans. Signal Process.*, vol. 66, no. 16, pp. 4193–4206, Apr. 2018. doi: 10.1109/TSP.2018.2827321.
[153] H.-T. Wai, A. Scaglione, and A. Leshem, "Active sensing of social networks: Network identification from low-rank data," in *Cooperative and Graph Signal Processing*. New York: Academic Press, 2018, ch. 2, pp. 601–622. [Online]. Available: http://www.sciencedirect.com/science/article/pii/B9780128136775000225
[154] H.-T. Wai, S. Segarra, A. Ozdaglar, A. Scaglione, and A. Jadbabaie, "Blind community detection from low-rank excitations of a graph filter," *IEEE Trans. Signal Process.*, vol. 68, pp. 436–451, Jan. 2020. doi: 10.1109/TSP.2019.2961296.
[155] H. Wai, A. Scaglione, B. Barzel, and A. Leshem, "Joint network topology and dynamics recovery from perturbed stationary points," *IEEE Trans. Signal Process.*, vol. 67, no. 17, pp. 4582–4596, Sept. 2019. doi: 10.1109/TSP.2019.2929471.
[156] B. Yilmaz, K. Bekiroglu, C. Lagoa, and M. Sznaier, "A randomized algorithm for parsimonious model identification," *IEEE Trans. Autom. Control*, vol. 63, no. 2, pp. 532–539, Feb. 2018. doi: 10.1109/TAC.2017.2723959.
[157] M. Timme, "Revealing network connectivity from response dynamics," *Phys. Rev. Lett.*, vol. 98, no. 22, p. 224101, May 2007. doi: 10.1103/PhysRevLett.98.224101.
[158] X. Wang, X. Chen, and L. Wang, "Evolutionary dynamics of fairness on graphs with migration," *J. Theoretical Biol.*, vol. 380, pp. 103–114, 2015. doi: 10.1016/j.jtbi.2015.05.020.
[159] A. De, S. Bhattacharya, P. Bhattacharya, N. Ganguly, and S. Chakrabarti, "Learning a linear influence model from transient opinion dynamics," in *Proc. 23rd ACM Int. Conf. Inf. Knowl. Manag (CIKM '14)*. New York: Association for Computing Machinery, 2014, pp. 401–410. doi: 10.1145/2661829.2662064.
[160] 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, no. 2, p. 021021, Dec 2011. doi: 10.1103/PhysRevX.1.021021.
[161] G. Strang, *Introduction to Linear Algebra*. Wellesley, MA: Wellesley-Cambridge Press, 2009.
[162] M. Grant and S. Boyd, "CVX: Matlab software for disciplined convex programming, version 2.1," Mar. 2014. [Online]. Available: http://cvxr.com/cvx
[163] 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. doi: 10.1002/cpa.20124.
[164] A. Mirtabatabaei, P. Jia, and F. Bullo, "Eulerian opinion dynamics with bounded confidence and exogenous inputs," *Siam J. Appl. Dynamical Syst.*, vol. 13, no. 1, pp. 425–446, 2014. doi: 10.1137/130934040.
[165] J. Lorenz, "Universality in movie rating distributions," *Eur. Phys. J. B*, vol. 71, pp. 251–258, Sept. 2009. doi: 10.1140/epjb/e2009-00283-3.
[166] D. J. Watts and S. H. Strogatz, "Collective dynamics of'small-world'networks," *Nature*, vol. 393, no. 6684, pp. 409–410, 1998. doi: 10.1038/30918.
[167] R. Z. Albert, "Statistical mechanics of complex networks," Ph.D. dissertation, Univ. of Notre Dame, Notre Dame, IN, 2001, aAI3000268.
[168] M. Rao, A. Kipnis, T. Javidi, Y. C. Eldar, and A. Goldsmith, "System identification from partial samples: Non-asymptotic analysis," in *Proc. IEEE 55th Conf. Decision Control (CDC)*, 2016, pp. 2938–2944. doi: 10.1109/CDC.2016.7798707.
[169] M. S. Granovetter, "The strength of weak ties: A network theory revisited," *Sociol. Theory*, vol. 1, pp. 201–233, 1983. doi: 10.2307/202051.
[170] A. C. R. Martins, "Discrete opinion models as a limit case of the coda model," *Phys. A: Statist. Mech. Appl.*, vol. 395, pp. 352–357, Feb. 2014. doi: 10.1016/j.physa.2013.10.009.
[171] J. M. Hendrickx and S. Martin, "Open multi-agent systems: Gossiping with random arrivals and departures," in *Proc. IEEE 56th Annu. Conf. Decision Control (CDC)*, 2017, pp. 763–768. doi: 10.1109/CDC.2017.8263752.
[172] M. Franceschelli and P. Frasca, "Proportional dynamic consensus in open multi-agent systems," in *Proc. IEEE Conf. Decision Control (CDC)*, 2018, pp. 900–905. doi: 10.1109/CDC.2018.8619639.