CHAPTER

Information diffusion and rumor spreading

4

Argyris Kalogeratos[a,*], Kevin Scaman[‡*,], Luca Corinzia[‡*,†] and Nicolas Vayatis[*]**

[*]*CMLA, ENS Cachan, CNRS, Université Paris-Saclay, 94235 Cachan, France*

[**]*MSR, Inria Joint Center, 91120 Palaiseau, France*

[†]*ETH Zürich, 8092 Zürich, Switzerland*

[a]*Corresponding:* `kalogeratos@cmla.ens-cachan.fr`

ABSTRACT

This chapter studies information cascades on social networks with a special focus on types of diffusion processes such as rumors and false news. The complex temporal dynamics of information cascades and rapid changes in user interests require flexible mathematical modeling to properly describe the diffusion dynamics. After mentioning the modeling advancements of recent decades, we get to modern models, such as the Information Cascade Model (ICM), that are indeed capable of describing such time-dependent user interests and are thus particularly suited to the analysis of information diffusion. We provide a theoretical analysis of ICM, relating the dynamics of the cascade to structural characteristics of the social network, and then use that analysis to design control policies capable of efficiently reducing the undesired diffusion. The presented framework for activity shaping is generic while enjoying a simple convex relaxation. Finally, we present an algorithm for the control of Continuous-Time Independent Cascades which is evaluated and compared against baseline and state-of-the art approaches through diffusion simulations on real and synthetic social networks.

Keywords: Information diffusion networks, rumors, information cascades, propagation, diffusion control, social interaction

[‡] The main part of the work has been conducted while author was at CMLA[*].

4.1 INTRODUCTION

Modern societies understand the world, manifest different viewpoints, and test their objectiveness, by exchanging information through direct communication or, in more recent years, through online social networks. On a larger scale, this process may also create consensus and mitigate social friction through public debate, two essential aspects of a healthy democracy. Information diffusion is often represented by *pieces of information* (e.g. news, scientific or historical facts) that spread through a network. As for the network, that consist of interacting entities, such as individuals, institutions (e.g. governments, authorities, or other organizations), and private entities (e.g. media, marketing agencies).

The Internet era has offered new means to produce and share information through large-scale online social networks. The disposition of large amount of data coming from diffusion traces has helped scientific research to improve our understanding of diffusion processes arising in various disciplines, including sociology, epidemiology, marketing, and computer systems' security. However, the *democratization* of content creation and sharing has not been adequately coupled with effective (self-, collective, or automatic) moderation, correction and filtering mechanisms. Consequently, the explosive volume of the available content brings forward huge challenges regarding the human capacity to process that fast-paced and gigantic information stream, as well as regarding the technical aspects of data management.

Our daily information diet tends to promote the variety in the content we consume to the expense of its precision and detail. During moments of crisis, the scarcity of trustworthy information and lack of time to analyze it leads to the proliferation of false rumors. There are also various psychological factors that impact the way we participate in this exchange. For instance, people get influenced by others, but also tend to search and recall information and facts that align with their already formed belief system (*confirmation bias*).

Furthermore, users interact preferably with people of similar profiles and opinions (*homophily*), a tendency that greatly reduces the heterogeneity of the user's perceived public debate. In addition, members of any online group receive social pressure to conform to group's beliefs; that tends to radicalize opinions and allow questionable ideas to gain momentum (*echo chambers*). Then, the relative isolation of small online communities may lead them to believe in false rumors, even create a false consensus against what is considered as verifiable by the majority of society. The situation may get considerably aggravated in periods of political tension where polarization and partisanship grows in well-segregated groups that reduce significantly their exposure to counterarguments.

Rumor spreading and control. There are many types of misinformation: bad or 'yellow' journalism, fake news, rumors and unverified information, hoaxes, and others (for a discussion on the taxonomy see [1]). Despite the fact that many studies hardly distinguish these types, there are still notable differences with regards to the *actors* propagating unverified or untruth information (e.g. individuals, media, politicians or authorities), their *motives* (e.g. ignorance, desire to be part of a movement, gaining visibility and revenues, or as part of a speculative communication campaign), and the way people interact with a new piece of information in each of those cases, especially during its verification process. As it has been pointed out, terms like "fake news" *are just new names for very old problems*. The particular recent concern of public opinion on fake news is however due to the fact that the cascading effects of misinformation gain magnitude and speed in online social networks, and thus their short-term negative impact is boosted. These effects have been recorded in numerous major events, such as terrorist attacks, social demonstrations, elections, natural disasters, and war conflicts.

In this chapter we mainly refer to *untruth rumors*[1] that represent false information and may have malicious motives. Such rumors are usually proven false shortly after their appearance. However, the debunking may not propagate fast enough in the social network to prevent a rumor from pursuing its diffusion (this is also the case, for example, of long-lasting rumors such as conspiracy theories) and that is exactly the point where computational tools can be beneficial.

There have been many developments in recent decades concerning both information dissemination and viral epidemics on networks. Despite the particular properties of rumor spreading, it is still a type of information diffusion for which many generic models and results are therefore relevant. Early models originated from the Susceptible-Infected-Removed (SIR) epidemic model [2, 3] and a detailed related work is provided in the next section. Worth to mention though the modern family of Information Cascade Models (ICM) [4] which considers heterogeneous node-to-node transmission probabilities. ICM fits well to problems related to information diffusion on social networks and, among others, finds straightforward applications in digital marketing [5]. Indeed, ICMs were used to fit real information cascade data and observed node 'infection' times in the MemeTracker dataset [6]. In another work, the aim was to infer the edges of a diffusion network and estimate the transmission rates of each edge that best fits the observed data [7].

Theoretical studies have given valuable insights on diffusion processes by defining quantities tightly related with the systemic behavior (e.g. epidemic threshold, extinction time) and describing how a diffusion unfolds from an initial set of contagious nodes. Most notably, a number of studies highlighted the crucial role that the network structure plays in how the diffusion process unfolds, which is also the subject on which this chapter is largely devoted. The relation between the network structure and the behavior of SIR epidemics has been shown in [8]. Follow-up works did verified this relation and broadened the discussion to other types of diffusion models [9, 10]. Similar theoretical results have then been given for ICM as well [11, 12].

The quantification of systemic properties can help on the direction of risk assessment (e.g. economic, health, social risks) and, furthermore, enable *diffusion process*

[1] According to Oxford English Dictionary, a rumor is "*a currently circulating story or report of uncertain or doubtful truth*". Thus, a rumor is by definition uncertain and may eventually be true or false. However, what will always be problematic is the fact that rumors gain disproportional circulation speeds to their level of certainty.

engineering whose aim could be either to suppress or enhance a spreading. Under ICM, this engineering task is also called in literature as *influence optimization* or *activity shaping*, whereas the maximization has received a lot of attention for its direct marketing applications. In recent years, the suppression of information diffusion processes has also become a hot topic since it is related to various security hazards, e.g. due to cascades of misinformation like harmful rumors and fake news. Suppressive scenarios of the latter type are also possible in the ICM modeling context; the optimization problem would be the minimization of the spread of a piece of information in the network, e.g. by decreasing the probability for certain users to share the false content to their contacts. To the best of our knowledge there is no prior work on this direction and part of the contribution of this chapter is exactly on covering this gap by developing computational approaches that are able to reduce an undesired spread under the ICM.

Contribution and summary. The rest of the chapter keeps its focus on information diffusion and is structured as follows. We commence with the detailed related work (Sec. 4.2), the technical background regarding diffusion models (Sec. 4.3), and their dynamics as stochastic processes (Sec. 4.4). The reader may find helpful the Tab. 4.1 which lists the main notations we use in this chapter. Then, we discuss one of the interesting tasks arising in diffusion networks: the offline influence optimization through local intervention actions that affect the information spread (Sec. 4.5). The purpose can be either to minimize or maximize the influence by means of suppressive or enhancive actions, respectively. An efficient strategy should decide where on the network to perform a number of available actions (limited by a budget of resources) in order to better serve one of those two opposing aims.

To this end we extend the discussion with the novel approach first appeared in [13] which frames this task as a generalized optimization problem under the ICM and enjoys a convex continuous relaxation. In particular, we present a class of algorithms based on the optimization of the spectral radius of the Hazard matrix using a projected subgradient method (Sec. 4.6). For these algorithms, which can address both the maximization and the minimization problem, we provide theoretical analysis. The suppressive case is however more interesting in the context of this chapter as it is straightforwardly related to the control of undesired diffusion processes such as the spread of rumors. Hence, we investigate two standard case-studies of the minimization problem (Sec. 4.7): the *quarantine* (e.g. see [10, 14]) and the *node immunization* problem (see [15]).

Notably, among the major strengths of this framework is the fact that it can describe complex strategies that are able to use several immunization options by deploying simultaneously resources of different types (partial or total immunization of edges and nodes, etc). We also discuss how such strategies could find practical application to rumor control scenarios. In a section with experimental results (Sec. 4.8), the main presented control algorithm, called *NetShape*, is compared to standard baselines and state-of-the-art competitors in synthetic and benchmark network datasets. In the last section (Sec. 4.9), we give our conclusions and directions of future research.

4.2 RELATED WORK

Modeling information and rumor spreading. Phenomena like rumors are part of an old story which is adapted to the current technological context. Scientists started studying rumors and stories related to the two World Wars. Knap [2], and soon later Allport and Postman [3, 16], were among the first to analyze rumors and pose the question of their control. In the work of the latter two, it was pointed out that, loosely speaking, the spread of rumors is somewhat proportional to the general interest of the story and the ambiguity of the related evidence. The similarities between rumor and disease spreading were also noted in later literature, though Daley and Kendal were the first to connect epidemics and rumors in mathematical terms [17, 18]. However, they noted that their dynamics may be strikingly different due to the particularly complex rumor spreading mechanism. Specifically, they introduced a variant of the Susceptible-Infected-Removed (SIR) epidemic model, where stochastic recoveries are triggered either when a) an infected node interacts with an already recovered one, or b) two infected nodes interact and both may then recover. A slight modification was proposed in [19] concerning case (b) where only the infected node that initiates the interaction may recover.

These alterations to the basic epidemic model try to incorporate mechanisms where a person is probable to lose its motivation in continuing to spread a rumor when he realizes that it is no more novel and interesting, or has already been debunked. Interesting to note, though, there is no assumed self-recovery process and the recovery is rather brought about by crowdsourcing. This is in accordance to follow-up and recent data-driven studies on rumor spreading on `twitter` which from one side observed self-correction to be very weak and slow to take effect, while from the other side they observed almost 1:1 ratio of users promoting important false rumors and users trying to debunk them [20, 21].

In the course of the years more refined SIR-like epidemic models were proposed for information diffusion, including rumors, that still have a permanent recovered state (for a survey on compartmental models see [22, 23]). One example is the SEIR model that introduces the (E)xposed state in which the individual is infected but incubating before getting to (I) and become infectious to others. Another example is SEI[R]Z [24, 25] that introduces competition among adopters at state (I) and those at state (Z) who after infection have become skeptics. Both adopters and skeptics recruit from the susceptible population; nodes can ‘exit’ the system and change the population size over time. However, the state (S) also recruits from a general population which is out of the system, and one could assume that previously departed individuals may later become susceptible again.

Evidently, the most popular epidemic modeling choice for information cascades, including rumors, are the *monotonically increasing stochastic models* like SIR, that allow node transitions only towards more critical states and eventually lead to permanent recovery or removal (i.e. as if the node dies out). Indeed, such modeling fits to what is observed in high-frequency information circulation with short life, a setting that covers the majority of the content reaching users: from social networks, news broadcasts, entertainment industry, and advertising. Nevertheless, for an information spread that spans in longer time periods and may come and go to the current affairs (e.g. political issues, ideas, competing products, long-lasting rumors), models that allow reinfection, are definitely more relevant. In this sense, the Susceptible-Infected-Susceptible (SIS), or the more information-oriented SEI[R]Z [24, 25], could be fit better and enable also dynamic approaches for suppressing a diffusion, e.g. the *priority-planning* [26] or the greedy approach of [27].

More recently, Information Cascade Models (ICM) were introduced that have higher detail and can take advantage of the wealth of available social interactions data to fine-tune their parameters. First, Independent Cascades have emerged as a relevant model for viral diffusion of ideas and opinions [28, 29, 7, 30]. Similarly to SIR, Independent Cascades are also increasing stochastic processes. However, contrary to epidemic models, they capture the precise temporal dependencies between infection events of neighboring nodes, but require larger training datasets to infer them properly. Second, multivariate Hawkes processes are self-exciting point processes that are considered as the gold standard to deal with sequences of correlated events in many scientific fields, e.g. for earthquake prediction [31], in biological [32], financial [33, 34] and social interactions studies [35]. They were thus naturally adapted to information diffusion in social networks with the main advantage of allowing multiple events on a single node (e.g. posts, likes or shares in the case of a social network) [36, 37]. Finally, Linear Threshold models were developed to account for more complex diffusion dynamics in which users may require more than one concordant piece of information to accept it [28].

Influence optimization. The first attempts to put forward computational approaches for assessing the influence of users in social networks were those in [38, 39]. The *influence maximization* problem under the ICM was first formulated in [5]. It was proved that it is an NP-hard problem and remains NP-hard to approximate it within a factor $1 - 1/e$. It was also proven that the influence is a submodular function of the set of initially contagious nodes (referred to as *influencers*) and the authors proposed a greedy Monte Carlo-based algorithm as an approximation. A number of subsequent studies were focused on improving that technique [40, 41]. Notably, today's state-of-the-art techniques on influence control under the ICM are still based on Monte Carlo simulations and a greedy mechanism to select the actions sequentially.

Besides the popularity of influence maximization, various questions regarding how one could apply suppressive interventions have also become a hot topic in recent years. However, to the best of our knowledge, there is no existing work under the ICM and, as mentioned in the introduction, the methodological contribution of this chapter is on the development of computational approaches under the ICM that are able to efficiently reduce an undesired spread (see Sec. 4.5).

Network structure, information spread, and control approaches. Recent theoretical studies have highlighted how crucial the structure of the underlying network is for the behavior of a diffusion process. Specifically, they have studied the way structural characteristics of the network do appear in quantities that are tightly related with the process behavior, such as the epidemic threshold and the extinction time.

An early work that drew a line between epidemic spreading and the structural properties of the underlying network is that in [8]. Under a mean field approximation of an SIR epidemic model on a graph, they found that the epidemic threshold is proportional to the *spectral radius* of the adjacency matrix. Follow-up works verified this relation and broadened the discussion to more types of diffusion and related models. In [9] the $S^*I^2V^*$ model was presented as a generalization of numerous virus propagation models of the literature. It was also made possible to generalize the result of [8] to that generic virus models. Based on these works, several research studies have been presented on the epidemic control on networks, mainly focusing on developing *immunization* strategies (elimination of nodes) and *quarantine* strategies (elimination of edges). The eigenvalue perturbation theory was among the main analytical tools used, see for example [15, 14, 10].

Similar theoretical results to those discussed above have been given for ICM as well. Under discrete- or continuous-time ICM, it has been shown that the epidemic threshold depends on the *spectral radius* of a matrix built upon the edge transmission probabilities, termed as *Hazard matrix* [11, 12].

Related applications. Dealing with information diffusion and rumors gives rise to a series of computational and inference problems, namely among others: credibility assessment of posts and users [42]; sentimental analysis on how individuals receive a piece of information; stance/role identification of users towards it; detection of rumors and their spreaders in content streams [43, 44, 45]; identification of influential users that could maximize the reach of a campaign, by examining structural properties of the network alone or in combination to historical data (interaction traces) [5, 46, 47]; finally, the development of countermeasures to suppress a rumor or information cascade [48, 13] which is discussed in the technical part of the chapter.

4.3 MODELS OF INFORMATION CASCADES

Information cascades describe the dynamics of communication between individuals of a social network by capturing the way messages are shared and propagate among users. In all generality, an information cascade on a graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ is a multivariate stochastic process $\{X_i(t) : i \in \mathcal{V}, t \geq 0\}$ where $X_i(t) \in \mathcal{S}$ denotes the state of user i at time t, and $\mathcal{S}$ is a state space that may be finite, countable or uncountable. Depending on the specific model, the state of a user may refer to a binary quantity (e.g. $\mathcal{S} = \{Unaware, Informed\}$), to the number of messages received during $[0, t]$ (in which case $\mathcal{S} = \mathbb{N}$), or something more detailed regarding the message spread (e.g. $\mathcal{S} = \mathbb{R}^d$ a low-dimensional representation of the content of the message). In all the models, we consider that users that did not participate at all in the cascade are in a default state $0 \in \mathcal{S}$. As a rumor propagates through the network, the number of individuals participating in the cascade, called *influence*, will grow and eventually reach

Symbol	Description
$\mathbb{1}\{\text{<condition>}\}$	indicator function
$\mathbf{1}$	vector with all values equal to one
$\Vert X\Vert_\ell$	ℓ-norm for a given vector X: e.g. $\Vert X\Vert_1 = \sum_{ij} X_{ij}$, or generally $\Vert X\Vert_\ell = (\sum_{ij} X_{ij}^\ell)^{1/\ell}$
$M \odot M'$	the Hadamard product between matrices (i.e. coordinate-wise multiplication)
$\mu_{\pi(1)} \geq \mu_{\pi(2)} \ldots$	ordered values of vector μ using the order-to-index bijective mapping π
$\mathcal{G}, \mathcal{V}, n, \mathcal{E}, E$	network $\mathcal{G} = \{\mathcal{V}, \mathcal{E}\}$ of $n = \lvert\mathcal{V}\rvert$ nodes and $E = \lvert\mathcal{E}\rvert$ edges
(i, j)	edge $(i, j) \in \mathcal{E}$ of the graph between nodes i and j
A	network's adjacency matrix $A \in \{0, 1\}^{n\times n}$
$\mathcal{S}$	state space. Example states: (S)usceptible, (I)nfeted, (R)ecovered
S_0, n_0	subset $S_0 \subset \mathcal{V}$ of $n_0 = \lvert S_0\rvert$ influencer nodes from which the IC initiates
$\mathcal{F}$	$n \times n$ *Hazard matrix* $[\mathcal{F}_{ij}]_{ij}$ of non-negative integrable *Hazard functions* over time
$\mathbb{F}$	set of feasible Hazard matrices $\mathbb{F} \subset \mathbb{R}_+ \to \mathbb{R}_+^{n\times n}$, where $\mathcal{F}$ is one of its elements
Δ	matrix of the integrated difference of two Hazard matrices: $\Delta = \int_0^{+\infty} (\hat{\mathcal{F}}(t) - \mathcal{F}(t))dt$
τ_i	time $\tau_i \in \mathbb{R}_+ \cup \{+\infty\}$ at which the information reached node i during the process
$\sigma(S_0)$	*influence*: the final number of contagious nodes when diffusion starts from the set S_0
$\rho_{\mathcal{H}}(\mathcal{F})$	the largest eigenvalue of the symmetrized and integrated Hazard matrix $\mathcal{F}$
$\hat{p}(s)$	Laplace transform of the function $p(t)$
X	control actions matrix $X \in [0, 1]^{n\times n}$ with the amount of action taken on each edge
x	control actions vector $x \in [0, 1]^n$ with the amount of action taken on each node
k	budget of control actions $k \in (0, E)$ for actions on edges, or $k \in (0, n)$ for nodes

Table 4.1 Index of main notations.

a saturation point. We use this quantity as our main quality metric:

Definition 1. *Influence* $\sigma(S_0, t)$ – Let $S_0 = \{i \in \mathcal{V} \ : \ X_i(0) \neq 0\} \subset \mathcal{V}$ be the set of *influencers*, i.e. users that are initially contagious. The influence of the set S_0 at time t is defined as the total number of messages received by users of the social network before time t:

$$\sigma(S_0, t) = \mathbb{E}\left[\sum_{i\in V} \mathbb{1}\{X_i(t) \neq 0\}\right]. \tag{4.1}$$

In the following, we denote as $n = |\mathcal{V}|$ the size of the social network, $E = |\mathcal{E}|$ the number of connections, $n_0 = |S_0|$ the number of initial influencers and the adjacency matrix of $\mathcal{G}$ as $A \in \{0, 1\}^{n\times n}$ s.t. $A_{ij} = 1 \Leftrightarrow (i, j) \in \mathcal{E}$. Moreover, we denote as long-term influence the total number of received messages after the diffusion $\sigma(S_0) = \lim_{t\to+\infty} \sigma(S_0, t)$.

4.3.1 EARLY MODELS: VIRUSES SPREADING THROUGH SOCIAL NETWORKS

Epidemics are usually modeled using *Markov processes* [49], i.e. *memoryless* stochastic processes entirely defined by their transition matrix. This transition matrix defines the probability for each node to change state during an infinitesimal time window $[t, t + dt]$ (the simultaneous change of more than one node's state is considered improbable). In the following, we thus use the notation:

$$X_i(t) : \mathrm{Y} \rightarrow \mathrm{Z} \text{ at rate } C_i(t) \tag{4.2}$$

to denote the stochastic transition rate $C_i(t) \geq 0$ of node $i \in \{1, ..., n\}$ at time $t \geq 0$ from state Y to state Z, with Y, Z $\in \mathcal{S}$.

Due to similarities between spreading phenomena, virus models have been also used to describe information cascades on social networks. We here focus on two standard such models: the SI and SIR models and we refer the reader to the recent review in [50] for more information on the vast epidemiology literature.

4.3.1.1 Susceptible-Infected model

The Susceptible-Infected (SI) model is the simplest epidemic model, in which nodes can be either (S)usceptible or (I)nfected. An infected node transmits the disease to one of its susceptible neighbor at a rate β, and once infected a node remains infected and thus contagious.

Model 1. *SI model* – Let $\mathcal{G}$ be a (possibly weighted) graph of n nodes and adjacency matrix A. The Susceptible-Infected model is a continuous-time Markov process $X(t) \in \{\mathrm{S}, \mathrm{I}\}^n$ with the following transition rate:

$$X_i(t) \colon \mathrm{S} \rightarrow \mathrm{I} \text{ at rate } \beta \sum_j A_{ji} X_j(t), \tag{4.3}$$

where β is the transmission rate of the epidemic.

Since the nodes remain infected, a connected network will be totally infected at the end of the diffusion, and hence any set S_0 has influence $\sigma(S_0) = n$.

4.3.1.2 Susceptible-Infected-Removed model

The Susceptible-Infected-Removed (SIR) model [51] is a widely used epidemic model designed for scenarios in which patients present immunity to the disease after their infection and recovery. A recovered person will not transmit the disease further, nor will it be subject to reinfections. An additional state is thus added to the SI model and each node of the network is either (S)usceptible, (I)infected, or (R)emoved. At $t = 0$, a subset S_0 of n_0 nodes is infected. Then, each infected node will transmit the disease to its neighbors at rate β, and recover at rate δ.

Model 2. *SIR model* – Let $\mathcal{G}$ be a (possibly weighted) graph of n nodes and adjacency matrix A. The Susceptible-Infected-Removed model is a continuous-time Markov process $X(t) \in \{\mathrm{S}, \mathrm{I}, \mathrm{R}\}^n$ with the following transition rates:

$$\begin{array}{l} X_i(t) \colon \mathrm{S} \rightarrow \mathrm{I} \text{ at rate } \beta \sum_j A_{ji} X_j(t) \\ X_i(t) \colon \mathrm{I} \rightarrow \mathrm{R} \text{ at rate } \delta, \end{array} \tag{4.4}$$

where β is the transmission rate of the epidemic and δ is the recovery rate of nodes.

Usually, the graph is undirected and all edges have the same rate. More complex scenarios can be modeled using the *inhomogeneous SIR* model, in which each edge has its own transmission rate β_{ij} and each node its own recovery rate δ_i.

An alternative definition for this model is possible using *infection times*. One may see that each node gets infected at most once and recovers at most once as well. We can thus define, for each node i, the time τ_i^{I} at which it gets infected and the time τ_i^{R} at which it recovers, with $\tau_i^{\mathrm{I}}, \tau_i^{\mathrm{R}} \in \mathbb{R}_+ \cup \{+\infty\}$. Then, $\tau_i^{\mathrm{I}} = 0$ would indicate that user i is an influencer, while $\tau_i^{\mathrm{I}} = +\infty$ would indicate that node i never got infected throughout the whole epidemic.

Proposition 1. *For an SIR epidemic, the infection times τ_i^{I} of not initially infected nodes verify the following equality:*

$$\forall i \notin S_0,\ \tau_i^{\mathrm{I}} = \min_{\{j \in \{1,\dots,n\}\,:\,T_{ji} < D_j\}} (\tau_j^{\mathrm{I}} + T_{ji}), \tag{4.5}$$

where T_{ji} and D_j are independent exponential random variables of expected value $1/\beta$ and $1/\delta$, respectively, and $\tau_i^{\mathrm{I}} = +\infty$ if the set $\{j \in \{1, \dots, n\} : T_{ji} < D_j\}$ is empty. Furthermore, the recovery time of each node i is:

$$\tau_i^{\mathrm{R}} = \tau_i^{\mathrm{I}} + D_i. \tag{4.6}$$

Proof. This result relies on the fact that a node is infected as soon as at least one of its infected neighbors transmits the infection to him. Since these events are independent, the times T_{ij} required for infection along the edges of the network are also independent. For more precisions, see e.g. [11]. □

4.3.2 INDEPENDENT CASCADES

Independent Cascades were initially introduced as discrete-time diffusion processes [28], and later refined to more flexible continuous-time processes [**?**].

Model 3. *Discrete-Time Independent Cascades $\mathcal{DTIC}(\mathcal{P})$* – At time $t = 0$, only a set S_0 of influencers is infected. Given a matrix $\mathcal{P} = (p_{ij})_{ij} \in [0, 1]^{n \times n}$, each node i that receives the contagion at time t may transmit it at time $t + 1$ along its outgoing edge $(i, j) \in \mathcal{E}$ with probability p_{ij}. Node i cannot infect its neighbors in subsequent rounds $t' > t + 1$. The process terminates when no more infections are possible.

The continuous version of Independent Cascades requires the definition of *Hazard functions* to describe the varying transmission rates along each edge of the network.

Definition 2. *Hazard function $\mathcal{F}_{ij}(t)$* – For every edge $(i, j) \in \mathcal{E}$ of the graph, $\mathcal{F}_{ij}$ is a non-negative integrable function that describes the time-dependent stochastic transmission rate from node i to node j, after i's infection.

Model 4. *Continuous-Time Independent Cascades $\mathcal{CTIC}(\mathcal{F})$* – The $\mathcal{CTIC}(\mathcal{F})$

model is a stochastic diffusion process defined as follows: at time $s = 0$, only the influencer nodes in S_0 are infected. Then, each node i that receives the contagion at time τ_i may transmit it at time $s \geq \tau_i$ along an outgoing edge $(i, j) \in \mathcal{E}$ with stochastic rate of occurrence $\mathcal{F}_{ij}(s - \tau_i)$.

The rest of this chapter will mainly focus on the analysis and control of such information cascades. For notational purposes, we denote as $\mathcal{F} = [\mathcal{F}_{ij}]_{ij}$ the $n \times n$ *Hazard matrix* containing as elements the individual Hazard functions and, respectively, as $\mathcal{F}(t) = [\mathcal{F}_{ij}(t)]_{ij}$ the evaluation of all functions at a relative time-point t after each infection time τ_i. Essentially, network edges imply non-zero Hazard functions:

$$(i, j) \in \mathcal{E} \;\Leftrightarrow\; \exists t \geq 0 \text{ s.t. } \mathcal{F}_{ij}(t) \neq 0. \tag{4.7}$$

Note that each Hazard function $\mathcal{F}_{ij}$ is always evaluated at a *relative time-point* initialized at the infection time τ_i of the source node i.

Similarly to SIR, Independent Cascades are monotonically increasing stochastic processes, and each node can only be infected once. We can thus define, for each node i, the time τ_i of its first infection, which may be infinite if the node does never get infected during the contagion. Unlike SIR, no epidemic states are explicitly mentioned in the notations of $\mathcal{CTIC}$ (the reader may compare Eq. 4.5 and Eq. 4.8).

Proposition 2. *For a Continuous-Time Independent Cascade $\mathcal{CTIC}(\mathcal{F}, T)$, the infection times τ_i of non-influencer nodes verify the following equality:*

$$\forall i \notin S_0,\; \tau_i = \min_{j \in \{1,\dots,n\}} (\tau_j + T_{ji}), \tag{4.8}$$

where $T_{ij} \in \mathbb{R}_+ \cup \{+\infty\}$ are independent random variables of sub-probability density

$$p_{ij}(t) = \mathcal{F}_{ij}(t) \exp\left(- \int_0^t \mathcal{F}_{ij}(s) ds \right). \tag{4.9}$$

Proof. This result is similar to Proposition 1 and relies on the same observation: a node is active as soon as at least one of its active neighbors activated him. Since these events are independent (hence the name of the model), the times T_{ij} required for activation along the edges of the network are also independent. For more precisions, see for example [52]. □

In general, $p_{ij}(t)$ is *not* a probability density over $\mathbb{R}_+$ as it does not integrate to one, and $\mathbb{P}(T_{ij} = +\infty) = 1 - \int_0^{+\infty} p_{ij}(t)dt = \exp(- \int_0^{+\infty} \mathcal{F}_{ij}(t)dt)$. Proposition 2 provides a simple mechanism for simulating $\mathcal{CTIC}$, as one can first draw one independent random variable T_{ij} per edge, and then use a shortest path algorithm to compute the infection times τ_i for each node of the network.

In what follows, we focus on this model due to its expressiveness and broad use in modern social network studies. However, the large-scale dynamics of all diffusion models are relatively similar and exhibit the same threshold behavior.

4.4 LARGE-SCALE DYNAMICS OF INDEPENDENT CASCADES

At the scale of the network, the emergent behavior of information cascades display several typical characteristics which are common in most diffusion processes, including epidemics and computer viruses. For instance, Fig. 4.1 shows the number of identified cases of Ebola during a recent crisis, the number of queries for "pokemon go" when the game became viral, as well as the simulation of an Independent Cascade (see Model 4 in Sec. 3). All these diffusion processes exhibit similar behavior:

1. **Explosive start:** The cascade starts with an exponential increase and quickly reaches a non-negligible amount.
2. **Saturation point:** After a sharp increase during the early phase of the diffusion, the process reaches a saturation point and comes to a halt. Note that, for information cascades, a residual activity may produce a linear slope after the end of the diffusion. However, we ignore this aspect in our study.

As a consequence, we focus on four main characteristics of interest to describe the large-scale dynamics of information cascades:

1. **Existence:** Is the cascade powerful enough to enter the explosive phase?
2. **Saturation point:** What is the final reach of the cascade?
3. **Time for action:** When is the explosion taking place?
4. **Explosive rate:** How fast is the initial exponential increase of the cascade?

These four characteristics are summarized in a simulated toy example on Fig. 4.1(c). In the following sections, we provide estimates of these quantities depending on the diffusive properties of the process as well as the structure of the social network.

4.4.1 EXISTENCE OF A SUPERCRITICAL CASCADE

Intuitively, an information cascade may only sustain itself if, on average, people that receive the message share it to more than one of their neighbors. When the network connectivity is too low, the cascade cannot reach a large audience before dying out. This is highlighted by the following upper bound relating a measure of network connectivity introduced in [12], the *Hazard radius*, to the long-term influence.

Definition 3. *Hazard radius* $\rho_{\mathcal{H}}(\mathcal{F})$ – For a diffusion process $\mathcal{CTIC}(\mathcal{F})$, $\rho_{\mathcal{H}}(\mathcal{F})$ is the largest eigenvalue of the symmetrized and integrated Hazard matrix:

$$\rho_{\mathcal{H}}(\mathcal{F}) = \rho\left(\int_0^{+\infty} \frac{\mathcal{F}(t) + \mathcal{F}(t)^{\mathsf{T}}}{2} dt\right), \tag{4.10}$$

where $\rho(\cdot) = \max_i |\lambda_i|$ and λ_i are the eigenvalues of the input matrix.

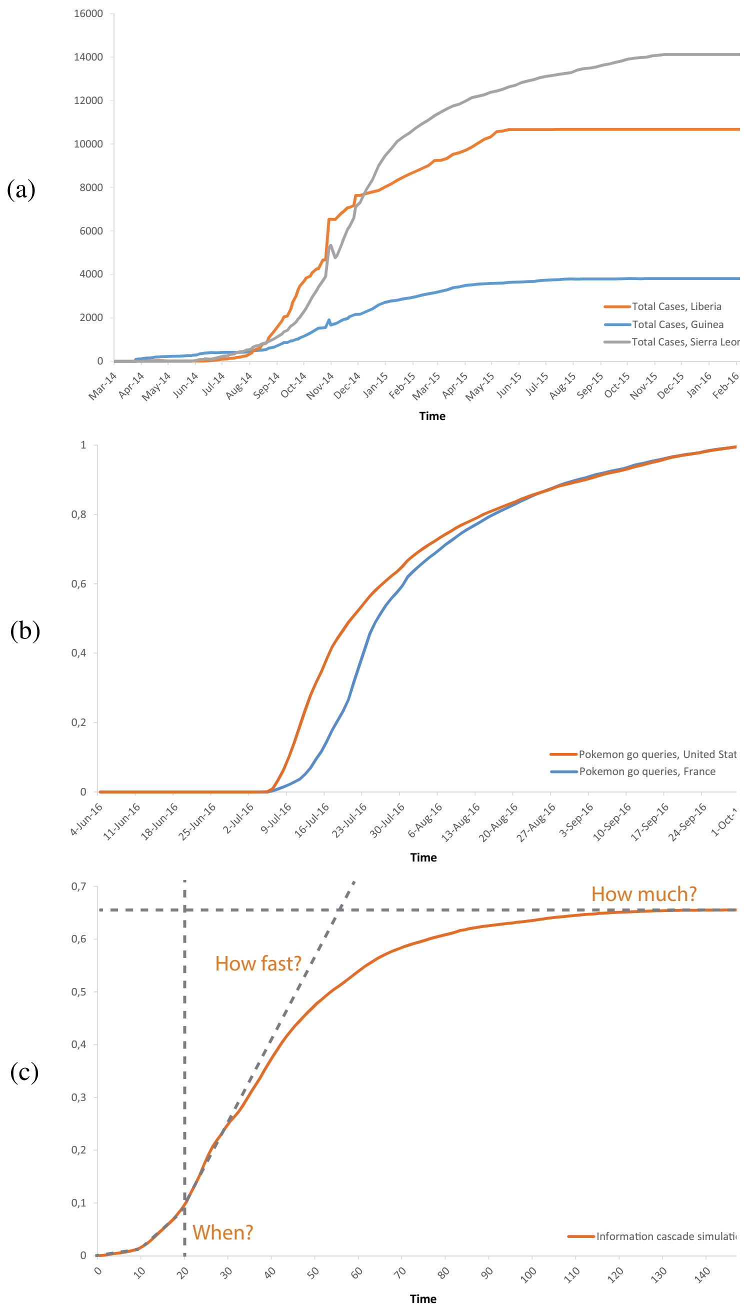


FIGURE 4.1 Main large-scale characteristics of diffusion processes appearing in real and simulated cascades

(a) Number of Ebola cases in Ginea, Liberia and Sierra Leone (source: World Health Organization); (b) number of searches for the query "pokemon go" on the Google search engine (source: Google Trend). (c) Simulation of a Continuous-Time Independent Cascade (see Model 4). The main large-scale characteristics highlighted in our analysis are also summarized: existence of outbreak, time before the explosion, explosive rate, and saturation point.

When all edges of the social network have identical Hazard function $\mathcal{F}_{ij}(t)$, the Hazard radius is proportional to the spectral radius of the adjacency matrix, which has been shown to drive the spread of epidemics [9]. The following proposition extends this result to Independent Cascades.

Proposition 3. *Let $S_0 \subset \mathcal{V}$ be a set of n_0 influencer nodes, and $\rho_{\mathcal{H}}(\mathcal{F})$ the Hazard radius of a $\mathcal{CTIC}(\mathcal{F})$. Then, if $\rho_{\mathcal{H}}(\mathcal{F}) < 1$, the influence of S_0 in $\mathcal{CTIC}(\mathcal{F})$ is upper bounded by:*

$$\sigma(S_0) \leq n_0 + \sqrt{\frac{\rho_{\mathcal{H}}(\mathcal{F})}{1 - \rho_{\mathcal{H}}(\mathcal{F})}} \sqrt{n_0(n - n_0)}. \tag{4.11}$$

Proof. This result relies on a non-trivial vector inequality between the activation probabilities Z_i at the end of the epidemic, defined as:

$$Z_i = \mathbb{P}(\tau_i < +\infty). \tag{4.12}$$

Note that

$$\|Z\|_1 = \sum_i \mathbb{E}[\mathbb{1}\{\tau_i < +\infty\}] = \sigma(S_0), \tag{4.13}$$

and any result on the vector Z will easily translate into a result on the influence. Proposition 2 leads to a relationship between the Z_i, as for any vector c, $\min_{j\in\{1,\dots,n\}} c_j < +\infty \Leftrightarrow \exists j \in \{1, ..., n\}$ s.t. $c_j < +\infty$, and thus

$$\begin{aligned} \mathbb{1}\{\tau_i < +\infty\} &= \mathbb{1}\{\min_{j\in\{1,\dots,n\}}(\tau_j + T_{ji}) < +\infty\} \\ &= 1 - \textstyle\prod_j \left(1 - \mathbb{1}\{\tau_j < +\infty\}\mathbb{1}\{T_{ji} < +\infty\}\right). \end{aligned} \tag{4.14}$$

Taking the expectation and using the Fortuin–Kasteleyn–Ginibre (FKG) inequality [53], a well-known result of mathematical physics, to prove the positive correlation between the variables $\mathbb{1}\{\tau_i < +\infty\}$, the following inequality arises after a short calculation:

$$\forall i \notin S_0, \;\; Z_i \leq 1 - \exp\left(-\sum_j \mathcal{H}_{ji} Z_j\right). \tag{4.15}$$

This inequality upper bounds the expected activation of a node with the expected activation of its neighbors, and can be turned into a bound on the norm of Z using the spectral radius of the matrix $\mathcal{H}$. The final step of the proof is rather calculatory and relies on Jensen's inequality and the definition of the spectral radius for symmetric matrices.The complete derivation is available in [12]. □

Hence, the Independent Cascade is subcritical when $\rho_{\mathcal{H}}(\mathcal{F}) < 1$, and the number of active users remains negligible compared to the size of the network: $\sigma(S_0) = O(\sqrt{n}) \ll n$. Note that we assume that the number of influencer nodes n_0 is bounded and does not depend on n.

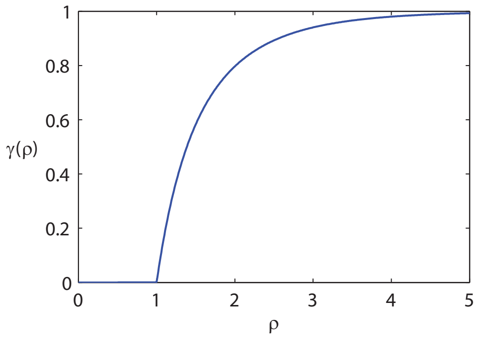


FIGURE 4.2 Upper bound on the saturation point

Function γ defined in Eq. 4.17. When $\rho_{\mathcal{H}}(\mathcal{F}) < 1$, the function is equal to 0, then increases and saturates to $\gamma = 1$ as $\rho_{\mathcal{H}}(\mathcal{F})$ tends to infinity.

4.4.2 LONG-TERM BEHAVIOR OF INDEPENDENT CASCADES

When the cascade is efficient enough to propagate to a large proportion of the network, it displays a sharp increase before saturating to a limit value. Although the precise value of this limit influence is hard to evaluate, several upper bounds have been provided and proven in the literature [12, 54]. We now provide such a result relating the long-term influence to the Hazard radius of the cascade.

Proposition 4. *Let $S_0 \subset \mathcal{V}$ be a set of n_0 influencer nodes, and $\rho_{\mathcal{H}}(\mathcal{F})$ the Hazard radius of a $\mathcal{CTIC}(\mathcal{F})$. Then, if $\rho_{\mathcal{H}}(\mathcal{F}) > 1$, the long-term influence of S_0 in $\mathcal{CTIC}(\mathcal{F})$ is upper bounded by:*

$$\sigma(S_0) \leq n_0 + \gamma(n - n_0) + c_n \sqrt{n_0(n - n_0)}\,, \tag{4.16}$$

where $c_n = \sqrt{\frac{\eta}{1-\eta}}$, $\eta = (1 - \gamma)\rho_{\mathcal{H}}(\mathcal{F})$ and $\gamma \in [0, 1]$ is the unique positive solution of the equation:

$$\gamma = 1 - \exp\left(-\rho_{\mathcal{H}}(\mathcal{F})\gamma\right). \tag{4.17}$$

Proof. This result is also a consequence of Eq. 4.15 relating the expected activations Z_i. See [12]. □

In essence, the proportion of active nodes after the cascade is negligible when $\rho_{\mathcal{H}}(\mathcal{F}) < 1$, and at most γ when $\rho_{\mathcal{H}}(\mathcal{F}) > 1$, where γ is defined by the implicit equation $\gamma = 1 - \exp\left(-\rho_{\mathcal{H}}(\mathcal{F})\gamma\right)$. Fig. 4.2 shows the proportion γ of Proposition 4 with respect to the Hazard radius $\rho_{\mathcal{H}}(\mathcal{F})$.

4.4.3 EXPLOSIVE DYNAMICS IN THE SUPERCRITICAL REGIME

Finally, the intermediate regime when the cascade grows exponentially can be analyzed using a modified version of the Hazard radius, known as *Laplace Hazard radius*.

Definition 4. *Laplace Hazard matrix* $\mathcal{L}(s)$ – Let p_{ij} be the edge transmission probabilities defined in Eq. 4.9. For $s \geq 0$, let $\mathcal{L}(s)$ be the $n \times n$ matrix, called as *Laplace Hazard matrix*, whose coefficients are:

$$\mathcal{L}_{ij}(s) = \begin{cases} -\hat{p}_{ij}(s)\left(\int_0^{+\infty} p_{ij}(t)dt\right)^{-1} \ln\left(1 - \int_0^{+\infty} p_{ij}(t)dt\right) & \text{if } (i,j) \in \mathcal{E} \\ 0 & \text{otherwise} \end{cases}, \tag{4.18}$$

where $\hat{p}_{ij}(s)$ denotes the Laplace transform of p_{ij} defined for every $s \geq 0$ by $\hat{p}_{ij}(s) = \int_0^{+\infty} p_{ij}(t)e^{-st}dt$.

Definition 5. *Laplace Hazard radius* $\rho_{\mathcal{L}}(s)$ – For a diffusion process $\mathcal{CTIC}(\mathcal{F})$ and $s \geq 0$, $\rho_{\mathcal{L}}(s)$ is the largest eigenvalue of the symmetrized Laplace Hazard matrix:

$$\rho_{\mathcal{L}}(s) = \rho\left(\frac{\mathcal{L}(s) + \mathcal{L}(s)^{\mathsf{T}}}{2}\right), \tag{4.19}$$

where $\rho(\cdot) = \max_i |\lambda_i|$ and λ_i are the eigenvalues of the input matrix.

This concept is slightly more complicated than the Hazard radius. When $s = 0$, the Laplace Hazard radius coincides with the Hazard radius: $\rho_{\mathcal{L}}(0) = \rho_{\mathcal{H}}(\mathcal{F})$. However, when s is large, the Laplace Hazard radius captures the short-term behavior of the hazard function by reducing the impact of long times through the Laplace transform. Quite surprisingly, the explosive rate of the cascade is upper bounded by the inverse value $\rho_{\mathcal{L}}^{-1}(1)$. This is discussed by the following proposition.

Proposition 5. *Let $t \geq 0$, $S_0 \subset \mathcal{V}$ be a set of n_0 influencer nodes, and $\rho_{\mathcal{L}}$ the Laplace Hazard radius. Then, the short-term influence of S_0 in $\mathcal{CTIC}(\mathcal{F})$ at time t is upper bounded by:*

$$\sigma(S_0, t) \leq n_0 + (2n_0)^{1/3}(n - n_0)^{2/3} \exp\left(\rho_{\mathcal{L}}^{-1}(1)t\right). \tag{4.20}$$

Proof. This result relies on a similar equation to Eq. 4.15 describing the dynamics of the cascade instead of its long-term stable regime. More specifically, Proposition 2 shows that, for any $t \geq 0$, the variables $\mathbb{1}\{\tau_i < t\}$ are related according to:

$$\mathbb{1}\{\tau_i < t\} = 1 - \prod_j \left(1 - \mathbb{1}\{\tau_j + T_{ji} < t\}\right). \tag{4.21}$$

Now, denoting as $Z_i(t) = \mathbb{P}(\tau_i < t)$ the probability that node i is active at time t, one may show the following vectorial inequality relating the variables $Z_i(t)$:

$$Z_i(t) \leq 1 - \exp\left(-\sum_j (\mathcal{F}_{ji} * Z_j)(t)\right), \tag{4.22}$$

where $(f * g)(t) = \int_{\mathbb{R}} f(s)g(t-s)ds$ is the convolution product. From this inequality, one may prove an upper bound on the Laplace transform of the influence $\widehat{\sigma}(s) = \int_0^{+\infty} \sigma(S_0, t)e^{-st}dt$, directly translating into an upper bound on the exponential increase of the influence. Again, the complete derivation is available in [11]. □

This result has two implications (for more precise results see [11]):

- First, the influence is at most increasing at an exponential rate of $\rho_{\mathcal{L}}^{-1}(1)$.
- Second, this also provides a characteristic time under which the cascade is still in its early phase. More precisely, before the critical time

  $$t \leq \frac{\ln n}{3\rho_{\mathcal{L}}^{-1}(1)}, \tag{4.23}$$

  the cascade is *subcritical* and the influence is negligible: $\sigma(S_0, t) = O(n^{2/3})$.

4.5 MONITORING INFORMATION CASCADES

Having presented the fundamental theoretical properties of diffusion processes related to information propagation over networks, we now discuss an efficient approach to the generic problem of optimizing influence (maximizing or minimizing) using actions that can shape, i.e. modify, the activity of single users. For instance, a marketing campaign may have a certain advertisement budget that can be used on targeted users of a social network. While these targeted resources are usually represented as new influencer nodes that will spread the piece of information, we rather consider the more refined and general case in which each resource will essentially alter the Hazard functions $\mathcal{F}_{ij}$ associated to a target node i, thus increasing, or decreasing, the probability for i to propagate by sharing the information with its neighbors.

Our generic framework assumes that a *set of feasible Hazard matrices* $\mathbb{F} \subset \mathbb{R}_+ \to \mathbb{R}_+^{n\times n}$ is available to the administrator. This set virtually contains all admissible policies that one could apply to the network. Then, the concern is to find the Hazard matrix $\mathcal{F} \in \mathbb{F}$ that minimizes, or maximizes depending on the task of interest, the influence. In Sec. 4.7 we show that two problems that have been a major focus of the literature so far, namely the edge-deletion problem [14] and the node-immunization problem [15] are particular instances of this framework. Note that this framework is generic enough to describe complex strategies that may use several immunization options by deploying simultaneously resources of different types (removal of edges, nodes, partial immunization, etc).

Problem 1. *Determining the optimal feasible policy* – Given a graph $\mathcal{G}$, a number of influencers n_0 and a set of admissible policies $\mathbb{F}$, find the optimal policy:

$$\mathcal{F}^* = \operatorname*{argmin}_{\mathcal{F}\in\mathbb{F}} \ \sigma^*_{n_0}(\mathcal{F}), \tag{4.24}$$

where $\sigma^*_{n_0}(\mathcal{F}) = \max\{\sigma(S_0) : S_0 \subset \mathcal{V} \text{ and } |S_0| = n_0\}$ is the optimal influence (according to Eq. 4.24 this is the minimum) over any possible set of n_0 influencer nodes.

Problem 1 cannot be solved exactly in polynomial time. The exact computation of the maximum influence $\sigma^*_{n_0}(\mathcal{F})$ is already a hard problem on its own, and minimizing this quantity adds an additional layer of complexity due to the non-convexity of the maximum influence w.r.t. the Hazard matrix (note: $\mathcal{F} \mapsto \sigma^*_{n_0}(\mathcal{F})$ is positive, upper bounded by n and not constant).

Proposition 6. *For any size of the set of influencers n_0, the computation of $\sigma^*_{n_0}(\mathcal{F})$ is #P-hard.*

Proof. We prove the theorem by reduction from a known #P-hard function: the computation of the influence $\sigma(S_0)$ given a set of influencers S_0 of size n_0 (see Theorem 1 of [55]). Indeed, let $\mathcal{CTIC}(\mathcal{F})$ be an Independent Cascade defined on $\mathcal{G} = (\mathcal{V}, \mathcal{E})$. We can construct a new graph $\mathcal{G}' = (\mathcal{V}', \mathcal{E}')$ as follows: for each influencer node $i \in S_0$, add a directed chain of n nodes $\{v_{i,1}, ..., v_{i,n}\} \subset \mathcal{V}'$ and connect $v_{i,n}$ to i by letting the transmission probabilities along the edges be all equal to one. Then, the maximum influence $\sigma^*_{n_0}$ is achieved with the nodes $S_0' = \{v_{i,1} \ : \ i \in S_0\}$ as influencer, and $\sigma^*_{n_0} = n\, n_0 + \sigma(S_0)$. The result follows from the #P-hardness of computing $\sigma(S_0)$ given S_0. □

The standard way to approximate the maximum influence is to employ incremental methods where the quality of each potential influencer is assessed using a Monte Carlo approach. In the following, we assume that the feasible set $\mathbb{F}$ is convex and included in a ball of radius R. Also, the requirement of Eq. 4.7, that network edges correspond to non-zero Hazard functions, holds for every feasible policy $\mathcal{F} \in \mathbb{F}$. Therefore, the number of edges E upper bounds the number of non-zero Hazard functions for any $\mathcal{F} \in \mathbb{F}$.

Remark 1. Although Problem 1 focuses on the minimization of the maximum influence, the algorithm presented in this paper is also applicable to the opposite task of influence maximization. Having a common ground for solving these opposite problems can be useful for applications where both opposing aims can interest different actors, e.g. in market competition. For the maximization, our algorithm would use a gradient ascent instead of a gradient descent optimization scheme. While the performance of the algorithm in that case may be competitive to state-of-the-art influence maximization algorithms, the non-convexity of this problem prevents us from providing any theoretical guarantees regarding the quality of the final solution.

4.6 AN ALGORITHM FOR REDUCING INFORMATION CASCADES

As it has been mentioned, solving exactly the influence optimization problem is computational intractable. Here, we propose to exploit the upper bound given in Proposition 4 as a heuristic for approximating the maximum influence. This approach can be seen as a *convex relaxation* of the original NP-Hard problem, and allows the use of convex optimization algorithms for this particular problem. The relaxed optimization problem thus becomes:

$$\mathcal{F}^* = \underset{\mathcal{F}\in\mathbb{F}}{\operatorname{argmin}}\ \rho_{\mathcal{H}}(\mathcal{F}). \tag{4.25}$$

When the feasible set $\mathbb{F}$ is convex, this optimization problem is also convex and our proposed method called *NetShape* uses a simple *projected subgradient descent* (see e.g. [56]) in order to find its minimum and make sure that the solution lays in $\mathbb{F}$. However, special care should be taken to perform the gradient step since, although the objective function $\rho_{\mathcal{H}}(\mathcal{F})$ admits a derivative w.r.t. the norm

$$\|\mathcal{F}\| = \sqrt{\sum_{i,j}\left(\int_0^{+\infty}\left|\mathcal{F}_{ij}(t)\right| dt\right)^2}, \tag{4.26}$$

the space of matrix functions equipped with this norm is only a Banach space in the sense that the norm $\|\mathcal{F}\|$ cannot be derived from a well chosen scalar product. Since gradients only exist in Hilbert spaces, gradient-based optimization methods are not directly applicable.

In the NetShape algorithm, the gradient and projection steps are performed on the *integral* of the Hazard functions $\int_0^{+\infty}\mathcal{F}_{ij}(t)dt$ by solving the optimization problem bellow:

$$\mathcal{F}^* = \underset{\hat{\mathcal{F}}\in\mathbb{F}}{\operatorname{argmin}}\left\|\int_0^{+\infty}\left(\hat{\mathcal{F}}(t) - \mathcal{F}(t)\right) dt + \eta\, u_{\mathcal{F}} u_{\mathcal{F}}^\top\right\|_2, \tag{4.27}$$

where $\eta > 0$ is a positive gradient step, $u_{\mathcal{F}}$ is the eigenvector associated to the largest eigenvalue of the matrix $\int_0^{+\infty}\frac{\mathcal{F}(t)+\mathcal{F}(t)^\top}{2}dt$, and $u_{\mathcal{F}} u_{\mathcal{F}}^\top$ is a subgradient of the objective function, as provided by the following proposition.

Proposition 7. *A subgradient of the objective function $f(M) = \rho(\frac{M+M^\top}{2})$ in the space of integrated Hazard functions, where M is a matrix, is given by the matrix:*

$$\nabla f(M) = u_M u_M^\top, \tag{4.28}$$

where u_M is the eigenvector associated to the largest eigenvalue of the matrix $\frac{M+M^\top}{2}$.

Proof. For any matrix M, let $f(M) = \rho(\frac{M+M^\top}{2}) = \max_{x:\|x\|_2=1} x^\top M x$, and u_M be such

Algorithm 1 – NetShape meta-algorithm

```
Input: feasible set $\mathbb{F} \subset \mathbb{R}_+ \to \mathbb{R}_+^{n\times n}$, radius $R > 0$ of $\mathbb{F}$, initial Hazard matrix $\mathcal{F} \in \mathbb{F}$, approx. parameter $\epsilon > 0$
Output: Hazard matrix $\mathcal{F}^* \in \mathbb{F}$
1: $\mathcal{F}^* \leftarrow \mathcal{F}$
2: $T \leftarrow \lceil \frac{R^2}{\epsilon^2} \rceil$
3: for $i = 1$ to $T - 1$ do
4:     $u_{\mathcal{F}} \leftarrow$ compute the eigenvector associated to the spectral radius $\rho_{\mathcal{H}}(\mathcal{F})$
5:     $\eta \leftarrow \frac{R}{\sqrt{i}}$
6:     $\mathcal{F} \leftarrow \mathrm{argmin}_{\hat{\mathcal{F}} \in \mathbb{F}} \left\| \int_0^{+\infty} \left( \hat{\mathcal{F}}(t) - \mathcal{F}(t) \right) dt + \eta\, u_{\mathcal{F}} u_{\mathcal{F}}^\top \right\|_2$
7:     $\mathcal{F}^* \leftarrow \mathcal{F}^* + \mathcal{F}$
8: end for
9: return $\frac{1}{T}\mathcal{F}^*$
```

an optimal vector. Then, we have $f(M+\varepsilon) = u_{M+\varepsilon}^\top (M+\varepsilon) u_{M+\varepsilon} \geq u_M^\top (M+\varepsilon) u_M = f(M) + u_M^\top \varepsilon u_M$, and, since $u_M^\top \varepsilon\, u_M = \left\langle u_M u_M^\top, \varepsilon \right\rangle$, $u_M u_M^\top$ is indeed a subgradient for $f(M)$. □

The projection step of line 6 in Alg. 1 is an optimization problem on its own, and NetShape algorithm is practical if and only if this optimization problem is simple enough to be solved. In the next sections we will see that, in many cases, this optimization problem can be solved in near linear time w.r.t. the number of edges of the network (i.e. $O(E \ln E)$), and is equivalent to a projection on a simplex.

4.6.1 CONVERGENCE AND SCALABILITY

Due to the convexity of the optimization problem in Eq. 4.25, NetShape finds the global minimum of the objective function and, as such, may be a good candidate to solve Problem 1. The complexity of the NetShape algorithm depends on the complexity of the projection step in Eq. 4.27. Each step of the gradient descent requires the computation of the first eigenvector of an $n \times n$ matrix, which can be computed in $O(E \ln E)$, where E is the number of edges of the underlying graph. In most real applications, the underlying graph on which the information is diffusing is *sparse*, in the sense that its number of edges E is small compared to n^2.

Proposition 8. *Assume that $\mathbb{F}$ is a convex set of Hazard matrices included in a ball of radius $R > 0$ w.r.t. the norm in Eq. 4.26, and that the projection step in Eq. 4.27 has complexity at most $O(E \ln E)$. Then, the NetShape algorithm described in Alg. 1 converges to the minimum of Eq. 4.25. Moreover, the complexity of the algorithm is $O(\frac{R^2}{\epsilon^2} E \ln E)$.*

Algorithm 2 – NetShape partial quarantine problem

```
Input: graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$, matrices of Hazard functions before and after treatment
    $\mathcal{F}, \hat{\mathcal{F}} \in \mathbb{F}$, approximation parameter $\epsilon > 0$, number of treatments $k$
Output: matrix of Hazard functions $\mathcal{F}^* \in \mathbb{F}$
$X \leftarrow 0, X^* \leftarrow 0$
$F \leftarrow \int_0^{+\infty} \mathcal{F}(t)dt$
$\Delta \leftarrow \int_0^{+\infty} (\hat{\mathcal{F}}(t)dt - \mathcal{F}(t))dt$
$R \leftarrow \sqrt{k} \max_{ij} \Delta_{ij}$
$T \leftarrow \lceil \frac{R^2}{\epsilon^2} \rceil$
for $i = 1$ to $T - 1$ do
    $M \leftarrow F + X \odot \Delta$
    $u \leftarrow$ the largest eigenvector of $\frac{1}{2}(M + M^\mathsf{T})$
    $Y \leftarrow X \odot \Delta - \frac{R}{\sqrt{i}} uu^\mathsf{T}$
    $X \leftarrow \mathrm{argmin}_{X' \in [0,1]^{n \times n}, \|X'\|_1 \leq k} \|X' \odot \Delta - Y\|_2$    // projection step (Alg. 3)
    $X^* \leftarrow X^* + X$
end for
return $\mathcal{F}^* = (1 - \frac{1}{T} X^*) \odot \mathcal{F} + \frac{1}{T} X^* \odot \hat{\mathcal{F}}$
```

Proof. This is a direct application of the projected subgradient descent to the problem:

$$\mathcal{H}^* = \underset{\mathcal{H} \in \mathbb{H}}{\mathrm{argmin}}\, \rho \left(\frac{\mathcal{H} + \mathcal{H}^\mathsf{T}}{2} \right), \tag{4.29}$$

where $\mathbb{H} = \left\{ \int_0^{+\infty} \mathcal{F}(t)dt \in \mathbb{R}^{n \times n} \,:\, \mathcal{F} \in \mathbb{F} \right\}$ is the set of feasible Hazard matrices. The convergence rate of such an algorithm can be found in [56]. □

Remark 2. The corresponding maximization problem is not convex anymore and only convergence to a local maximum can be expected. However, when the changes in the Hazard functions are relatively small (e.g. inefficient control actions, or only a limited number of treatments available to distribute), then NetShape achieves fairly good performance.

4.7 CASE STUDIES

In this section, we illustrate the generality of our framework by reframing well-known diffusion suppression problems that can find application in rumor control that has been discussed extensively in this chapter. Using Problem 1 we derive the corresponding variants of the NetShape algorithm.

Algorithm 3 – Projection step for the partial quarantine problem

```
Input: $\delta, y \in \mathbb{R}^E$, budget $k \in (0, E)$
Output: control actions vector $x'$
for $i = 1$ to $E$ do
    $\mu_i \leftarrow 2\delta_i y_i$
    $\mu_{E+i} \leftarrow 2\delta_i (y_i - \delta_i)$
end for
sort $\mu$ into $\mu_{\pi(1)} \geq \mu_{\pi(2)} \geq ... \geq \mu_{\pi(2E)}$
$d \leftarrow 0, s \leftarrow 0, i \leftarrow 1$
while $s < k$ and $\mu_{\pi(i)} \geq 0$ do
    $d \leftarrow d + \mathbb{1}\{\pi(i) \leq E\}\frac{1}{2\delta^2_{\pi(i)}} - \mathbb{1}\{\pi(i) > E\}\frac{1}{2\delta^2_{\sigma(i)-E}}$
    $s \leftarrow s + d(\mu_{\pi(i)} - \mu_{\pi(i+1)})$
    $i \leftarrow i + 1$
end while
$z \leftarrow \max\{0, \mu_{\sigma(i)} + \frac{s-k}{d}\}$
return $x'$ s.t. $x'_i = \max\{0, \min\{\frac{2\delta_i y_i - z}{2\delta^2_i}, 1\}\}$
```

For simplicity, we denote as $M \odot M'$ the Hadamard product between the two matrices (i.e. coordinate-wise multiplication), as $\Delta = \int_0^{+\infty} \left(\hat{\mathcal{F}}(t) - \mathcal{F}(t)\right) dt$ the matrix with the integrated coordinate-wise difference of two Hazard matrices in time, and as $\mathbf{1} \in \mathbb{R}^n$ the all-one vector (see notations in Tab. 4.1).

4.7.1 PARTIAL QUARANTINE

The *quarantine* approach aims to remove a small number of edges in order to minimize the spread of the contagion. This strategy is highly interventional in the sense that it totally removes edges, but in order to be practical it has to remain at low scale and affect a small amount of edges. This is the reason why it is mostly appropriate for dealing with the initial very few infections. The *partial quarantine* setting is a relaxation where one is interested to decrease the transmission probability along a set of targeted edges by using local and expensive actions.

Definition 6. *Partial quarantine* – Consider that a marketing campaign has k control actions to distribute in a network $\mathcal{G} = (\mathcal{V}, \mathcal{E})$. For each edge $(i, j) \in \mathcal{E}$, let $\mathcal{F}_{ij}$ and $\hat{\mathcal{F}}_{ij}$ be the Hazard matrices *before* and *after* applying control actions, respectively. If $X \in [0, 1]^{n \times n}$ is the control actions matrix and X_{ij} represents the amount of suppressive action taken on edge (i, j), then the set of feasible policies can be expressed as:

$$\mathbb{F} = \left\{(1 - X) \odot \mathcal{F} + X \odot \hat{\mathcal{F}} \ : \ X \in [0, 1]^{n \times n}, \ \|X\|_1 \leq k\right\}. \tag{4.30}$$

Example: For a non-negative scalar $\epsilon \geq 0$, we may consider $\hat{\mathcal{F}} = (1 - \epsilon)\mathcal{F}$ in order to model the suppression of selected transmission rates; formally:

$$\mathbb{F} = \{(1 - \epsilon X) \odot \mathcal{F} : X \in [0,1]^{n \times n},\ \|X\|_1 \leq k\}. \tag{4.31}$$

Importantly, for the special case where $\epsilon = 1$, this problem becomes equivalent to the setting discussed in [10] and [14].

A straightforward adaptation of Alg. 1 to this setting leads to the NetShape algorithm for partial quarantine described in Alg. 2. The projection step is performed by Alg. 3 on the flattened versions $x', \delta, y \in \mathbb{R}^E$ of the matrices X', Δ and Y, and the parameter R is chosen to upper bound $\max_{\mathcal{F}' \in \mathbb{F}} \|\mathcal{F}' - \mathcal{F}\|_2 = \max_{X \in [0,1]^{n \times n}, \|X\|_1 \leq k} \|X \odot \Delta\|_2$.

Lemma 1. *The projection step of Alg. 1 for the partial quarantine setting of Definition 6 is:*

$$X^* = \arg\min_{x' \in [0,1]^E,\ \|x'\|_1 \leq k} \left\| x' \odot \delta - y \right\|_2, \tag{4.32}$$

where δ and y are flattened version of, respectively, Δ and $Y = X \odot \Delta - \eta u_{\mathcal{F}} u_{\mathcal{F}}^{\mathsf{T}}$. Moreover, this problem can be solved in time $O(E \ln E)$ with Alg. 3, where E is the number of edges of the network.

Proof. Eq. 4.32 directly follows from Eq. 4.27 and the definition of $\mathbb{F}$. Alg. 3 is an extended version of the L_1-ball projection algorithm of [57]. Karush–Kuhn–Tucker (KKT) conditions for the optimization problem of Eq. 4.32 imply that $\exists z > 0$ s.t. $\forall i$, $x'_i = \max\{0, \min\{\frac{2\delta_i y_i - z}{2\delta_i^2}, 1\}\}$. The algorithm is a simple linear search for this value. Finally, the sorting step (Alg. 3, line 5) has the highest complexity $O(E \ln E)$, and the loops perform at most $2E$ iterations, hence an overall complexity $O(E \ln E)$. □

4.7.2 PARTIAL NODE IMMUNIZATION

More often, control actions can only be performed on the nodes rather than the network edges that was the case of the previous section. For example, imagine advertising campaigns that aim to enhance the diffusion of a product or, more relevant to the suppressive scenario, decision makers that debunk false information targeting specific influencer nodes. In that case, the effect of the control actions must be aggregated over nodes in the following way.

Definition 7. *Partial node immunization* – Consider that a control campaign has k control actions to distribute in a network $\mathcal{G} = (\mathcal{V}, \mathcal{E})$. For each edge $(i, j) \in \mathcal{E}$, let $\mathcal{F}_{ij}$ and $\hat{\mathcal{F}}_{ij}$ be the Hazard matrices *before* and *after* applying control actions, respectively. If $x \in [0,1]^n$ is the control actions vector and x_i represents the amount of suppressive action taken on node i, then we express the set of feasible policies as:

$$\mathbb{F} = \left\{ (1 - x\mathbf{1}^{\mathsf{T}}) \odot \mathcal{F} + x\mathbf{1}^{\mathsf{T}} \odot \hat{\mathcal{F}} : x \in [0,1]^n, \|x\|_1 \leq k \right\}. \tag{4.33}$$

This setting corresponds to partial quarantine in which all outgoing edges of a node are impacted by a single control action. When $\hat{\mathcal{F}} = 0$, this problem corresponds to the node removal problem (or vaccination), that consists in removing k nodes from the graph in advance in order to minimize a future contagion (see [15]).

Given a vector x, the projection problem to solve is:

$$
\begin{aligned}
x^* &= \operatorname*{argmin}_{x' \in [0,1]^n, \|x'\|_1 \le k} \left\| (x' \mathbf{1}^\mathsf{T}) \odot \Delta - Y \right\|_2 \\
&= \operatorname*{argmin}_{x' \in [0,1]^n, \|x'\|_1 \le k} \sum_i x_i^2 \Big(\sum_j \Delta_{ij}^2 \Big) - 2 x_i \Big(\sum_j \Delta_{ij} Y_{ij} \Big) \\
&= \operatorname*{argmin}_{x' \in [0,1]^n, \|x'\|_1 \le k} \left\| x' \odot \delta' - y' \right\|_2 ,
\end{aligned}
\tag{4.34}
$$

where $\delta'_i = \sqrt{\sum_j \Delta_{ij}^2}$ and $y'_i = \frac{\sum_j \Delta_{ij} Y_{ij}}{\sqrt{\sum_j \Delta_{ij}^2}}$. Hence we can apply the projection step of Alg. 3 for the partial node immunization problem using δ' and y', and its complexity is $O(n \ln n)$.

Remark 3. Since the upper bound of Proposition 4 holds as well for SIR epidemics [51] (see also [11]), this setting may also be used to reduce the spread of a disease using, for example, medical treatments or vaccines. More specifically, the Hazard matrix for an SIR epidemic is the following:

$$
\mathcal{H} = \ln\Big(1 + \frac{\beta}{\delta}\Big) A, \tag{4.35}
$$

where δ is the recovery (or removal) rate and β is the transmission rate along edges of the network, and A the adjacency matrix. Then, a medical treatment may increase the recovery rate δ for targeted nodes, thus decreasing all Hazard functions on its outgoing edges, and the partial node immunization setting is applicable.

Network	Nodes	Edges	Nodes in largest SCC
`SBD10ER`	500	2,701	497 :: 99.4%
`Facebook`	4,039	88,234	4,039 :: 100.0%
`Gnutella`	62,586	147,892	14,149 :: 22.6%
`Epinions`	75,879	508,837	32,223 :: 42.5%

TABLE 4.2 Datasets

Details of the benchmark real networks. The last column is the size of the strongly connected component.

4.8 EXPERIMENTS

4.8.1 EXPERIMENTAL SETUP AND EVALUATION

In this section, we provide empirical evidence for the discussion of this chapter on controlling Independent Cascades under the ICM. We set the focus of this empirical evaluation in the offline partial node immunization problem under the ICM, as described in Sec. 4.7.2, and we are interested to see in practice the performance gains of the *NetShape* algorithm when compared to other baseline and state-of-the-art alternative policies.

Compared policies. We provide comparative experimental results against several strategies, namely:

i) *Rand*: random selection of nodes;

ii) *Degree*: selection of k nodes with highest out-degree;

iii) *WeightedDegree*: selection of k nodes with highest sum of outgoing edge weight $w_{ij} = \int_0^{+\infty} \mathcal{F}_{ij}(t)dt$. This strategy can also be seen as the optimization of the first influence lower bound LB_1 of [54].

iv) *NetShield* algorithm [15]. Given the adjacency matrix of a graph, this outputs the best k-nodes to totally immunize so as to decrease the vulnerability of the graph. This is done by assigning to each node a *shield-value* that is high for nodes with high eigenscore and no edges connecting them. Note that, despite the fact that *NetShield* is tailored for immunization on unweighted graphs, it is not general enough to account for weighted edges and partial immunization as in our experimental setting.

Network datasets. The evaluation is performed on three benchmark real datasets (see Tab. 4.2) and the results are presented in subfigures of Fig. 4.4:

(a) a network of 'friends lists' from `Facebook` [58];

(b) the `Gnutella` peer-to-peer file sharing network [58],

(c) the who-trust-whom online review site `Epinions.com`;

(d) a synthetic random network of $n = 500$ nodes forming group structure (stochastic block-diagonal) that has been generated as follows. First, 10 equally-sized Erdös-Rényi clusters were independently formed with intra-cluster edge creation probability $p_{\text{inter}} = 0.1$. Then, their adjacency matrices were used to compose a block-diagonal structure with uniform inter-cluster rewiring probability $p_{\text{intra}} = 0.001$. Fig. 4.3a shows the structure of the final adjacency matrix (as having binary edge weights).

Note that the above networks only provide an unweighted adjacency matrix, thus only the existence, or not, of an edge between a pair of nodes is known. NetShape and the analysis of Sec. 4.5 is general covering time-variable propagation functions between nodes. However, without loss of generality and for the sake of simplifying the experimental setup, we decided to use a simple class of propagation functions. For the generation of the matrix of edge-transmission probability rates $\{p_{ij}\}$ we use a *trivalency model*, according to which, the p_{ij} values are drawn chosen uniformly at random from a small set of constants. In our case that is set to $\{p_{\text{low}}, p_{\text{med}}, p_{\text{high}}\}$ and the specific used values are mentioned explicitly for each dataset at the figures' captions.

Each treatment unit of the budget can be assigned to a single node and, here, we assume that it can cause a fixed decrease to the node's transmission probability rates along all of its outgoing edges (70% for the SBD10ER and 50% for the real networks).

In the experiments we evaluate the efficiency of the immunization policies on the basis of two measures for both of which lower values are better:

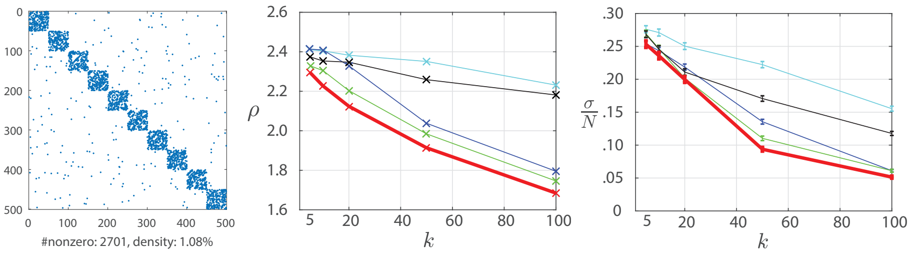
- *Spectral radius decrease*. We examine the extend of the decrease of the spectral radius of the Hazard matrix $\mathcal{F}$ and, hence, the decrease of the bound of the max-influence as described in Proposition 4.
- *Expected influence decrease*. We compare the performance of policies in terms of Problem 1. To this end, for each Hazard matrix $\mathcal{F}$, the influence is computed as the *average number of infected nodes* at the end of over 1,000 runs of the Independent Cascade $\mathcal{CTIC}$ while applying that specific Hazard matrix $\mathcal{F}$. Each time a single initial influencer is selected by the influence maximization algorithm Pruned Monte Carlo [40] by generating 1,000 vertex-weighted directed acyclic graphs (DAGs).

In the empirical study, we focus on the scenario where the spectral radius of the original network is approximately one, which is the setting in which decreasing the spectral radius has the most impact on the upper bounds in Proposition 4 and [12]. We believe that this intermediate regime is the most meaningful and interesting in order to test the different algorithms.

4.8.2 RESULTS

The results on the synthetic network are shown in Fig. 4.3 and those on the three real network datasets in subfigures of Fig. 4.4. The subfigures correspond to the two evaluation measures that we use, for a wide range of budget size k in proportion to the number of nodes of that network.

Firstly, we should note that the influence and the spectral radius measures correlate generally well across all reported experiments; they present similar decrease w.r.t. budget increase and hence 'agree' in the *order of effectiveness* of each policy when examined individually. As expected, all policies perform more comparably when very few or too many resources are available. In the former case, the very 'central' nodes are highly prioritized by all methods, while in the latter the significance of node selection diminishes. Even simple approaches perform well in all



(a) (b) (c)

FIGURE 4.3 Comparison of policies on a synthetic network

Comparison of NetShape’s performance against competitors on the synthetic network SBD10ER which is a composition of 10 Erdös-Rényi clusters (see details in Sec. 4.8.1). The values used for the trivalency model to generate edge weights are $p \in \{.1, .2, .5\}$. The tested budget values are: $k \in \{5, 10, 20, 50, 100\}$. (a) the structure of the generated non-symmetric, block-diagonal adjacency matrix (here plotted as a binary matrix); (b) spectral radius $\rho_H(\mathcal{F})$ vs. budget k, (c) influence: the expected proportion of infected nodes $\frac{\sigma}{n}$ vs. k. Lower values are better.

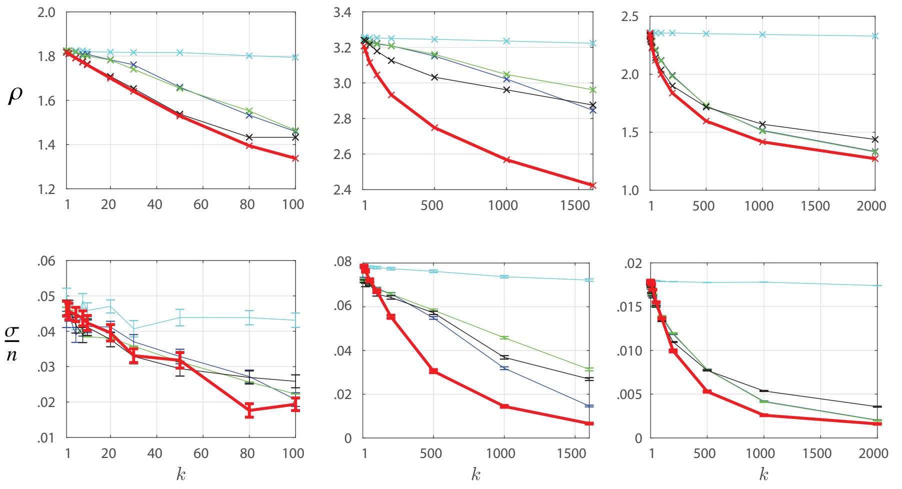
but Gnutella network where we get the most interesting results. NetShape achieves a sharp drop of the spectral radius early (i.e. for small budget k) in Gnutella and Epinions networks, which drives a large influence reduction. With regards to influence minimization, the difference to competitors is bigger though in Gnutella which is the most sparse and has the smallest strongly connected component (see Tab. 4.2). In Facebook, the reduction of the spectral radius is slower and seems less closely related with the influence, in the sense that the upper bound that we optimize is probably less tight to the behavior of the process.

Overall, the performance of the proposed NetShape algorithm is mostly as good or superior to that of the competitors, achieving up to a 50% decrease of the influence on the Gnutella network compared to its best competitor. Similar findings can be claimed for the experiments on the synthetic network SBD10ER.

4.9 CONCLUSION

The future of the diffusion networks field is full of interesting problems and potential applications. It will continue to enrich our understanding of diffusive phenomena and, at a second level, is expected to also change how information is circulated in online social networks.

The subject of this chapter was first to analyze the way information diffusion takes place in modern large-scale online social networks and the challenges regarding the control of certain types of undesired diffusion such as rumors, fake news,



(a) Facebook (b) Gnutella (c) Epinions

FIGURE 4.4 Comparison of policies on real networks

The evaluation is conducted on benchmark real networks in terms of two evaluation measures namely the spectral radius and the expected influence reduction. For each network, at the top row is plotted the $\rho_{\mathcal{H}}(\mathcal{F})$ vs. budget k, and at the bottom row the expected proportion of infected nodes $\frac{\sigma}{n}$ vs. k. (a) Facebook network, by generating infection rates $p \in \{.0001, .001, .01\}$; (b) Gnutella network with $p \in \{.1, .3, .6\}$; (c) Epinions network with $p \in \{.005, .005, .05\}$. Lower values are better.

and others. We have presented an overview of the complex context in which these information-related diffusive phenomena appear and how individuals participate in the process acting as users of online social platforms.

To present the background of related problems, we went through various approaches for modeling information cascades, including the early used virus models and the more recent Independent Cascades model. Specifically for the latter model, we spoke about its large-scale dynamics and how that relates to the network properties, the existence of a threshold value that defines the point of transition between subcritical and supercritical behavior, and the connection of that threshold value to the spectral radius of the Hazard matrix of the network.

Subsequently, we discussed a framework that we proposed recently for *spectral activity shaping* under the Continuous-Time Independent Cascades Model [13] that allows the administrator for local control actions by allocating targeted resources which can alter locally the spread of the process. The activity shaping is achieved via the optimization of the *spectral radius of the Hazard matrix* which enjoys a simple convex relaxation when used to minimize the influence of the cascade. In addition, by reframing a number of use-cases, we explained that the proposed framework is general and includes tasks such as partial quarantine that acts on edges and partial node immunization that acts on nodes. Notably, this generic framework can describe complex strategies that may use several immunization options by deploying simultaneously resources of different types (removal of edges, nodes, partial immunization, etc). Specifically for the influence minimization that is the one directly related to rumor spreading control, we presented the *NetShape* method which was compared favorably to baseline and a state-of-the-art method on real benchmark network datasets.

Among the interesting and challenging future work directions, on the same line to the presented framework, there can be the introduction of an ‘aging’ feature to each piece of information that would model its loss of relevance and attraction through time, and the theoretical study and experimental validation of the maximization counterpart of *Netshape* method.

REFERENCE

[1] A. Zubiaga, A. Aker, K. Bontcheva, M. Liakata, and R. Procter. Detection and resolution of rumours in social media: A survey. *ACM Computing Surveys*, 51(2):32:1–32:36.

[2] R.H. Knapp. A psychology of rumor. *The Public Opinion Quarterly*, 8(1):22–37, 1944.

[3] G.W. Allport and L. Postman. An analysis of rumor. *Public Opinion Quarterly*, 10(4):501–517, 1946.

[4] W. Chen, L.V.S. Lakshmanan, and C. Castillo. Information and influence propagation in social networks. *Synthesis Lectures on Data Management*, 5(4):1–177, 2013.

[5] D. Kempe, J. Kleinberg, and É. Tardos. Maximizing the spread of influence through a social network. In *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 137–146, 2003.

[6] J. Leskovec, L. Backstrom, and J. Kleinberg. Meme-tracking and the dynamics of the news cycle. In *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 497–506, 2009.

[7] M. Gomez-Rodriguez, D. Balduzzi, and B. Schölkopf. Uncovering the temporal dynamics of diffusion networks. In *Proceedings of the International Conference on Machine Learning*, pages 561–568, 2011.

[8] Y. Wang, D. Chakrabarti, C. Wang, and C. Faloutsos. Epidemic spreading in real networks: An eigenvalue viewpoint. In *Proceedings of the IEEE International Symposium on Reliable Distributed Systems*, pages 25–34, 2003.

[9] B.A. Prakash, D. Chakrabarti, N.C. Valler, M. Faloutsos, and C. Faloutsos. Threshold conditions for arbitrary cascade models on arbitrary networks. *Knowledge and Information Systems*, 33(3):549–575, 2012.

[10] H. Tong, B.A. Prakash, T. Eliassi-Rad, M. Faloutsos, and C. Faloutsos. Gelling, and melting, large graphs by edge manipulation. In *Proceedings of the ACM International Conference on Information and Knowledge Management*, pages 245–254, 2012.

[11] K. Scaman, R. Lemonnier, and N. Vayatis. Anytime influence bounds and the explosive behavior of continuous-time diffusion networks. In *Advances in Neural Information Processing Systems*, pages 2017–2025, 2015.

[12] R. Lemonnier, K. Scaman, and N. Vayatis. Tight bounds for influence in diffusion networks and application to bond percolation and epidemiology. In *Advances in Neural Information Processing Systems*, pages 846–854, 2014.

[13] K. Scaman, A. Kalogeratos, L. Corinzia, and N. Vayatis. A spectral method for activity shaping in continuous-time information cascades. *ArXiv e-prints*, abs/1709.05231, September 2017.

[14] P. Van Mieghem, D. Stevanović, F. Kuipers, C. Li, R. Van De Bovenkamp, D. Liu, and H. Wang. Decreasing the spectral radius of a graph by link removals. *Physical Review E*, 84(1):016101, 2011.

[15] H. Tong, B.A. Prakash, C. Tsourakakis, T. Eliassi-Rad, C. Faloutsos, and D.H. Chau. On the vulnerability of large graphs. In *Proceedings of the IEEE International Conference on Data Mining*, pages 1091–1096, 2010.

[16] G. W. Allport and L. Postman. The psychology of rumor. *Journal of Clinical Psychology*, 3(4), 1947.

[17] D.G. Kendall D.J. Daley. Epidemics and rumours. *Nature*, 204(8):1118, 1964.

[18] D.J. Daley and D.G. Kendall. Stochastic rumours. *IMA Journal of Applied Mathematics*, 1(1):42–55, 1965.

[19] D.P. Maki and M. Thompson. *Mathematical models and applications : with emphasis on the social, life, and management sciences*. Englewood Cliffs, N.J. : Prentice-Hall, 1973.

[20] C. Castillo, M. Mendoza, and B. Poblete. Predicting information credibility in time-sensitive

social media. *Internet Research*, 23(5):560–588, 2013.

[21] R. Procter, F. Vis, and A. Voss. Reading the riots on twitter: methodological innovation for the analysis of big data. *International Journal of Social Research Methodology*, 16(3):197–214, 2013.

[22] H.W. Hethcote. The mathematics of infectious diseases. *SIAM Review*, 42(4):599–653, December 2000.

[23] M. Newman. *Networks: An Introduction*. Oxford University Press, New York, NY, USA, 2010.

[24] L.M.A. Bettencourt, A. Cintrón-Arias, D.I. Kaiser, and C. Castillo-Chávez. The power of a good idea: Quantitative modeling of the spread of ideas from epidemiological models. *Physica A: Statistical Mechanics and its Applications*, 364(Supplement C):513 – 536, 2006.

[25] F. Jin, E. Dougherty, P. Saraf, Y. Cao, and N. Ramakrishnan. Epidemiological modeling of news and rumors on twitter. In *Proceedings of the Workshop on Social Network Mining and Analysis*, pages 1–9, 2013.

[26] K. Scaman, A. Kalogeratos, and N. Vayatis. Suppressing epidemics in networks using priority planning. *IEEE Transactions on Network Science and Engineering*, 3(4):271–285, 2016.

[27] K. Scaman, A. Kalogeratos, and Vayatis N. A greedy approach for dynamic control of diffusion processes in networks. In *IEEE International Conference on Tools with Artificial Intelligence*, pages 652–659. IEEE, 2015.

[28] D. Kempe, J. Kleinberg, and É. Tardos. Maximizing the spread of influence through a social network. In *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 137–146. ACM, 2003.

[29] W. Chen, Y. Wang, and S. Yang. Efficient influence maximization in social networks. In *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 199–208. ACM, 2009.

[30] Manuel Gomez-Rodriguez and Bernhard Schölkopf. Influence maximization in continuous time diffusion networks. In *Proceedings of the International Conference on Machine Learning*, pages 313–320, 2012.

[31] D. Vere-Jones. Earthquake prediction – statistician's view. *Journal of Physics of the Earth*, 26(2):129–146, 1978.

[32] P. Reynaud-Bouret, V. Rivoirard, F. Grammont, and C. Tuleau-Malot. Goodness-of-fit tests and nonparametric adaptive estimation for spike train analysis. *Journal of Mathematical Neurosciences*, 4(1):1–41, 2014.

[33] L. Bauwens and N. Hautsch. *Modelling financial high frequency data using point processes*. Springer, 2009.

[34] A. Alfonsi and P. Blanc. Dynamic optimal execution in a mixed-market-impact Hawkes price model. *Finance and Stochastics*, 20(1):183–218, 2015.

[35] R. Crane and D. Sornette. Robust dynamic classes revealed by measuring the response function of a social system. *Proceedings of the National Academy of Sciences*, 105(41):15649–15653, 2008.

[36] M. Farajtabar, N. Du, M. Gomez-Rodriguez, I. Valera, H. Zha, and L. Song. Shaping social activity by incentivizing users. In *Advances in neural information processing systems*, pages 2474–2482, 2014.

[37] R. Lemonnier, K. Scaman, and A. Kalogeratos. Multivariate Hawkes processes for large-scale inference. In *Proceedings of the Thirty-First AAAI Conference on Artificial Intelligence*, pages 2168–2174, 2017.

[38] P. Domingos and M. Richardson. Mining the network value of customers. In *Proceedings of*

the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pages 57–66, New York, NY, USA, 2001. ACM.

[39] M. Richardson and P. Domingos. Mining knowledge-sharing sites for viral marketing. In *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 61–70. ACM, 2002.

[40] N. Ohsaka, T. Akiba, Y. Yoshida, and K. Kawarabayashi. Fast and accurate influence maximization on large networks with pruned Monte-Carlo simulations. In *Proceedings of the AAAI Conference on Artificial Intelligence*, pages 138–144, 2014.

[41] J. Leskovec, A. Krause, C. Guestrin, C. Faloutsos, J. VanBriesen, and N. Glance. Cost-effective outbreak detection in networks. In *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 420–429, 2007.

[42] A. Gupta, P. Kumaraguru, C. Castillo, and P. Meier. Tweetcred: Real-time credibility assessment of content on twitter. In *Proceedings International Conference of Social Informatics*, pages 228–243, Cham, 2014. Springer.

[43] D. Shah and T. Zaman. Rumors in a network: Who's the culprit? *IEEE Transactions on Information Theory*, 57(8):5163–5181, 2011.

[44] E. Seo, P. Mohapatra, and T. Abdelzaher. Identifying rumors and their sources in social networks. In *Proceedings of the Ground/Air Multisensor Interoperability, Integration, and Networking for Persistent ISR III*, volume 8389, page 83891I, 2012.

[45] J. Ma, W. Gao, Z. Wei, Y. Lu, and K.-F. Wong. Detect rumors using time series of social context information on microblogging websites. In *Proceedings of the ACM International Conference on Information and Knowledge Management*, pages 1751–1754, 2015.

[46] M. Kitsak, L.K. Gallos, S. Havlin, F. Liljeros, L. Muchnik, H.E. Stanley, and H.A. Makse. Identification of influential spreaders in complex networks. *Nature*, 6(11):888–893, 2010.

[47] A. Goyal, F. Bonchi, and L.V.S. Lakshmanan. A data-based approach to social influence maximization. *Proceedings of VLDB Endowment*, 5(1):73–84, 2011.

[48] Zaobo H., Zhipeng C., and Xiaoming W. Modeling propagation dynamics and developing optimized countermeasures for rumor spreading in online social networks. In *Procceedings of the IEEE International Conference on Distributed Computing Systems*, pages 205–214, 2015.

[49] P. Van Mieghem, J. Omic, and R. Kooij. Virus spread in networks. *Networking, IEEE/ACM Transactions on*, 17(1):1–14, 2009.

[50] R. Pastor-Satorras, C. Castellano, P. Van Mieghem, and A. Vespignani. Epidemic processes in complex networks. *Reviews of Modern Physics*, 87:925–979, 2015.

[51] W.O. Kermack and A.G. McKendrick. Contributions to the mathematical theory of epidemics. II. the problem of endemicity. *Proceedings of the Royal society of London. Series A*, 138(834):55–83, 1932.

[52] N. Du, L. Song, M. Gomez-Rodriguez, and H. Zha. Scalable influence estimation in continuous-time diffusion networks. In *Advances in Neural Information Processing Systems*, pages 3147–3155, 2013.

[53] Cees M Fortuin, Pieter W Kasteleyn, and Jean Ginibre. Correlation inequalities on some partially ordered sets. *Communications in Mathematical Physics*, 22(2):89–103, 1971.

[54] J.T. Khim, V. Jog, and P.-L. Loh. Computing and maximizing influence in linear threshold and triggering models. In *Advances in Neural Information Processing Systems 29*, pages 4538–4546, 2016.

[55] C. Wang, W. Chen, and Y. Wang. Scalable influence maximization for independent cascade model in large-scale social networks. *Data Mining and Knowledge Discovery*, 25(3):545–576, 2012.

[56] S. Bubeck. Convex optimization: Algorithms and complexity. *Foundations and Trends in Machine Learning*, 8(3-4):231–357, 2015.

[57] J. Duchi, S. Shalev-Shwartz, Y. Singer, and T. Chandra. Efficient projections onto the L1-ball for learning in high dimensions. In *Proceedings of the International Conference on Machine Learning*, pages 272–279. ACM, 2008.

[58] J. Leskovec and A. Krevl. SNAP Datasets: Stanford large network dataset collection. `http://snap.stanford.edu/data`, June 2014.