Journal of Complex Networks (2018) **6**, 215–242
doi: 10.1093/comnet/cnx024
Advance Access Publication on 2 August 2017

A general Markov chain approach for disease and rumour spreading in complex networks

GUILHERME FERRAZ DE ARRUDA, FRANCISCO APARECIDO RODRIGUES AND PABLO MARTÍN RODRÍGUEZ
Departamento de Matemática Aplicada e Estatística, Instituto de Ciências Matemáticas e de Computação, Universidade de São Paulo—Campus de São Carlos, Caixa Postal 668, 13560-970 São Carlos, SP, Brazil

EMANUELE COZZO
Institute for Biocomputation and Physics of Complex Systems (BIFI), University of Zaragoza, 50018 Zaragoza, Spain and Department of Theoretical Physics, University of Zaragoza, 50018 Zaragoza, Spain

AND

YAMIR MORENO[†]
Institute for Biocomputation and Physics of Complex Systems (BIFI), University of Zaragoza, 50018 Zaragoza, Spain, Department of Theoretical Physics, University of Zaragoza, 50018 Zaragoza, Spain and Complex Networks and Systems Lagrange Lab, Institute for Scientific Interchange, Turin, Italy
[†]Corresponding author. Email: yamir.moreno@gmail.com

Edited by: Ernesto Estrada

[Received on 28 May 2017; editorial decision on 21 June 2017; accepted on 23 June 2017]

Spreading processes are ubiquitous in natural and artificial systems. They can be studied via a plethora of models, depending on the specific details of the phenomena under study. Disease contagion and rumour spreading are among the most important of these processes due to their practical relevance. However, despite the similarities between them, current models address both spreading dynamics separately. In this article, we propose a general spreading model that is based on discrete time Markov chains. The model includes all the transitions that are plausible for both a disease contagion process and rumour propagation. We show that our model not only covers the traditional spreading schemes but that it also contains some features relevant in social dynamics, such as apathy, forgetting, and lost/recovering of interest. The model is evaluated analytically to obtain the spreading thresholds and the early time dynamical behaviour for the contact and reactive processes in several scenarios. Comparison with Monte Carlo simulations shows that the Markov chain formalism is highly accurate while it excels in computational efficiency. We round off our work by showing how the proposed framework can be applied to the study of spreading processes occurring on social networks.

Keywords: complex networks; Markov chain models; epidemic spreading; rumour propagation.

1. Introduction

Disease spreading and information dissemination are two processes intimately linked that have been the subject of intense study since long time ago. These contagion phenomena are pervasive in nature, society

© The authors 2017. Published by Oxford University Press. All rights reserved.

Downloaded from https://academic.oup.com/comnet/article-abstract/6/2/215/4060524
by University of Manchester user
on 26 March 2018

and engineering [1]. As a result, we nowadays can describe and understand many of the mechanisms that are behind the propagation of pathogens among humans and other species, of digital viruses and malware through the Internet and diverse socio-technical systems, and of rumours among individuals, to mention a few examples [1]. Moreover, with the advent of modern communication technologies and transportation means and the increasing availability of data (often in real time or with a highly detailed time resolution), previous theoretical-only models are being fed with data [2], making data-driven simulations an effective tool for decision-making and for the design of efficient viral algorithms in the case of information dissemination.

The early models dealing with epidemic and rumour spreading considered only homogeneous mixing, in which the probability that a given node interacts with any other node in the network is the same for all nodes [3]. Although the homogeneous approximation facilitates the theoretical analysis of contagion processes, this approach turned out to be too simple as to capture the architecture of real-world complex systems, whose structure is largely heterogeneous [4]. Indeed, the availability of data about the topology of real systems spurred the development of network models [5], which in turn led to the inclusion of network's interaction patterns into disease spreading models [6].

One of the first and most used methods to study disease (and rumour) propagation is the so-called degree-based mean-field (DMF) approximation [6]. This approach groups vertices into classes and assumes that all nodes with the same degree are equivalent from a dynamical point of view. However, the DMF does not provide information about the probability of individual nodes. Recently, a formalism based on probabilistic discrete-time Markov chains was introduced to generalize existing MF approximations [7]. Differently from the MF approximation, discrete-time Markov chains enable the description of individual node dynamics as well as the determination of the macroscopic critical properties and the whole phase diagram [7].

Epidemic and rumour spreading processes are similar in many aspects. Indeed, the creation mechanism is the same in classical models: with a given (spreading) probability, a disease (rumour) is transmitted to any of the neighbours of an infected (spreader) individual. On the contrary, the annihilation mechanisms are completely different by the very nature of the processes being studied. In disease contagion, spreaders die out because they recover from the infection with a given probability. This is independent of their neighbours' dynamical states and of any interaction. Rumour-like processes are fundamentally different: in traditional models, rumours decay as a result of the interactions between spreaders and other individuals that already know the rumour—no matter whether they are actually propagating it (spreaders) or have already stopped (stiflers). Despite the differences between these two contagion processes, they have been studied using the same methodological approaches, albeit independently [1]. Surprisingly, the phenomenological similarities between them have not been fully exploited and there is not a general framework that allows studying both processes under the same formulation. Admittedly, rumour-like dynamics on complex topologies have been studied only at the mean-field level, and approaches like the discrete-time Markov chains are not available yet.

Here, we fill the existent gap and propose a general information-spreading model that captures the dynamical behaviour of epidemic and rumour spreading processes. Our model includes, as particular cases, not only the traditional models of rumour and epidemic spreading, but also other mechanisms such as apathy [8], forgetting [9], lost of interest [9, 10] and a new mechanism proposed here that characterizes cases in which the interest in the rumour can be recovered once lost. We thoroughly study, both analytically and numerically, the dynamics of the model in random synthetic graphs as well as in real social networks. Moreover, we analyse several plausible scenarios and obtain the corresponding critical spreading probabilities for the contact and reactive limiting cases, in addition to analyse the early time dynamics of the spreading process. We also perform extensive numerical simulations and show that the

Downloaded from https://academic.oup.com/comnet/article-abstract/6/2/215/4060524
by University of Manchester user
on 26 March 2018

discrete Markov chain approach developed here is highly accurate, both at the micro and the macroscopic levels. Finally, we discuss potential applications of this framework in the context of social contagion and the design of new viral algorithms for efficient information dissemination.

The rest of the article is organized as follows. In the next section, we present some related works summarizing the main previous results and approaches that have been adopted so far. Section 3 presents our model. Its early time behaviour and the steady state analysis are presented in Section 4. Section 5 analyses several special cases before turning our attention to the numerical analysis of the model, which is performed in Section 6. Finally, applications are discussed in Section 7, whereas our conclusions are presented in Section 8.

2. Previous spreading models

Spreading processes in networks, such as the propagation of diseases or rumours, are based on (i) spontaneous processes, in which each node might change its state without any external interference, or on (ii) contact based processes, in which the state of each individual might change following a contact with its neighbours. In the simplest case, each element performs only one contact per time unit. This process is called contact process (CP). On the other hand, when every neighbour of a node is contacted in one time step, the process is called reactive process (RP). Intermediate situations between CP and RP can be defined by considering a parameter that defines the activity of each node [7, 11]. Here we only consider the CP and RP schemes.

We assume that a contagion process could refer to a disease or a rumour that spreads on top of complex networked systems. Although one can study other variants of disease compartmental models, here we explore epidemic models in which there are at most three different dynamical states: (i) susceptible, which accounts for subjects that do not have the disease; (ii) infected, which denotes individuals who have the disease and are transmitting it, and (iii) recovered or removed, which stands for immunized or cured subjects. The simplest epidemic model is the susceptible-infected (SI), which models a disease with no cure. On the other hand, in the susceptible-infected-susceptible (SIS) model, individuals recover but do not acquire immunity in front of the disease, and, therefore, they can catch the disease again, that is, once recovered, they go back to the susceptible state. At variance with the SIS scenario, in the susceptible-infected-recovered (SIR) model, the recovery is permanent, that is, each individual acquires lifetime immunity. Above the critical point, the SIS model exhibits a steady state, where the number of susceptible and infective individuals are constant proportions of the population. More precisely, above the threshold, the SIS model presents a meta-state, where the probability that a macroscopic fraction of the population is infected is large than zero. It is worth mentioning that the only absorbing state of such a dynamics is the disease free configuration. In contrast, in the same regime, the SIR model presents infinitely many absorbing states when the fraction of infective is zero, being the number of recovered larger than zero.

The rumour models here considered also have three classes: (i) ignorant, which represents individuals that have not heard the rumour; (ii) spreader, which refers to those individuals that are aware of the rumour and are actively spreading it further, and (iii) stifler, who are those subjects that already know the rumour, but are not disseminating it any longer. The two main models of rumour spreading are due to Daley and Kendall (DK) [12–14] and Maki and Thompson (MT) [14, 15]. In the DK model, an edge of a network is sampled randomly, characterizing a contact. If this edge has a spreader and an ignorant, then the rumour is propagated from the spreader to the ignorant at a rate λ. On the other hand, if this edge is composed by two spreaders (or a spreader and a stifler), then the two spreaders turn into stiflers at a rate α. In the MT model, on the other hand, directed contacts are accounted for. Thus, if a spreader contacts a stifler

Downloaded from https://academic.oup.com/comnet/article-abstract/6/2/215/4060524
by University of Manchester user
on 26 March 2018

or another spreader, then only the initial spreader, who performed the contact, becomes a stifler at the same rate α. The traditional DK and MT models have been adapted to allow for heterogeneous contact patterns [16] and spontaneous lost of interest [8, 10].

For the sake of generality, given that we shall aim at developing a general model for spreading processes −no matter if a disease or a piece of information−, we consider throughout this paper the following states: (i) susceptible or ignorant, which we denote by state X, (ii) infected or spreader, which is represented by state Y; and (iii) recovered or stifler, corresponding to state Z. The reader is also referred to Table 1, where we have summarized the main rumour and epidemic models studied so far. The transitions between states in each model and the thresholds, when available, are also shown for completeness.

3. The general model

Here we introduce a general model that captures all the features of the models sketched in the previous Section 2. We consider a spreading dynamics on a population of N individuals whose contact pattern is given by a network. The structure of the interaction graph is encoded by a network with adjacency matrix $\mathbf{A}$, where $A_{ij} = 1$ if there is a connection between nodes i and j, and $A_{ij} = 0$ otherwise. We consider undirected networks, that is, $A_{ij} = A_{ji}$, for all i, j. The spreading process described by the model could refer to the transmission of a rumour, a disease or information that could be disseminated (also, a malware over a communication network, innovations, etc). In order to be more clear and precise, we will use henceforth a terminology that mostly refers to information transmission, but we stress that depending on the transitions that one allows to take place, the dynamics could represent other spreading processes. Therefore, let's assume that a 'piece of information' is being transmitted. An individual holding this piece of information and willing to spread it is called a spreader, whose state is represented by Y. On the other hand, a stifler, whose state is denoted by Z, is a subject who knows the information and does not spread it. An individual who is not aware of the information is called ignorant, and its state is denoted as X. In this way, the notation X, Y and Z, refers to subjects (or system's elements like devices) that have not participated in the spreading process, that are active, and that have taken part on the dynamics but are not active any longer, respectively. Hence, when we deal with an infectious disease, X, Y , and Z are, respectively, the susceptible, infective, and removed states.

The general model here described includes the transitions represented in Fig. 1. Specifically, at each time step, the spreading dynamics proceeds as follows for any given node i:

(i) An individual in state Y changes to state X with probability δ_1. This, for instance, refers to the case in which in a rumour process, a spreader forgets the rumour (meaning the piece of news becomes old and therefore Y is ignorant again). It also represents transitions of the type infected→susceptible in a SIS model.

(ii) An individual in state Y changes to state Z with probability δ_2. For rumour processes, it represents the case in which a spreader spontaneously (that is, not as a result of interactions with others Y or Z individuals) loses the interest in further propagating the rumour and becomes stifler. It also accounts for transitions of the type infected→recovered in a SIR model.

(iii) If transitions (i) and (ii) do not happen, which occurs with probability $(1 - \delta_1 - \delta_2)$, then the spreader i interacts with its neighbours. The outcome of such interactions are:

Downloaded from https://academic.oup.com/comnet/article-abstract/6/2/215/4060524
by University of Manchester user
on 26 March 2018

TABLE 1 *A brief literature review: a summary of previous epidemic and rumour spreading models. The states are (i) susceptible or ignorant (X), (ii) infected or spreader (Y) and (iii) recovered or stifler (Z)*

	Model	Interactions	Threshold	Networks	Comments	References
Epidemic	SI [1]	$Y+X \xrightarrow{\lambda} Y+Y$	–	Yes	Only two fixed points.	[14], [17]
	SIR	$Y+X \xrightarrow{\lambda} Y+Y$ $Y \xrightarrow{\delta} Z$	$\lambda > \frac{\delta}{\Lambda_{max}}$ [2]	Yes	Absorbing state, Z; presence of influential spreaders.	[14], [17]
	SIS	$Y+X \xrightarrow{\lambda} Y+Y$ $Y \xrightarrow{\delta} X$	$\lambda > \frac{\delta}{\Lambda_{max}}$ [3]	Yes	Presents an active steady state; discrete and continuous time.	[6], [18], [19], [20], [21], [14], [22], [23], [17], [7], [24]
	SIRS	$Y+X \xrightarrow{\lambda} Y+Y$ $Y \xrightarrow{\delta} Z$ $Z \xrightarrow{\gamma} X$	$\lambda > \delta \frac{\langle k \rangle}{\langle k^2 \rangle}$	Yes	Presents an active steady state; short-term immunity.	[17], [25], [26]
Rumour	Maki – Thompson	$Y+X \xrightarrow{\lambda} Y+Y$ $Y+Y \xrightarrow{\alpha} Z+Y$ $Y+Z \xrightarrow{\alpha} Z+Z$	$\frac{\lambda}{\alpha} > 0$	Yes	Absorbing state, Z; directed contact.	[15], [14], [27]
	Nekovee *et al.*	$Y+X \xrightarrow{\lambda} Y+Y$ $Y+Y \xrightarrow{\alpha} Z+Y$ $Y+Z \xrightarrow{\alpha} Z+Z$ $Y \xrightarrow{\delta} Z$	$\frac{\lambda}{\delta} \geq \frac{\langle k \rangle}{\langle k^2 \rangle}$	Yes	Absorbing state, Z; presents a 'lost of interest' mechanism.	[10]
	Borge *et al.* [4]	$Y+X \xrightarrow{\lambda\eta} Y+Y$ $Y+X \xrightarrow{\lambda(1-\eta)} Z+Y$ $Y+Y \xrightarrow{\alpha} Z+Z$ $Y+Z \xrightarrow{\alpha} Z+Z$	–	Yes	Absorbing state, Z; presents an apathy mechanism; models activity.	[8]
	Kawachi *et al.* [5]	$Y+X \xrightarrow{\alpha_{xy}\theta_{xy}} Y+Y$ $Y+X \xrightarrow{\alpha_{xy}(1-\theta_{xy})} Y+Z$ $Z+X \xrightarrow{\alpha_{xz}\theta_{xz}} Z+Y$ $Z+X \xrightarrow{\alpha_{xz}(1-\theta_{xz})} Z+Z$ $Y+Y \xrightarrow{\beta} Y+Z$ $Z+Y \xrightarrow{\gamma} Z+Z$ $Z+Z \xrightarrow{\lambda p} Z+X$ $Y \xrightarrow{\eta_y} X$ $Z \xrightarrow{\eta_z} X$	–	No	Presents an active steady state.	[9]

[1]This models is similar to the Eden model [28].

[2]This expression is obtained in details in [17], assuming uncorrelated networks. In [14], the authors show another compartmental based approach, which yields $\frac{\lambda}{\delta} > \frac{\langle k \rangle}{\langle k^2 \rangle - \langle k \rangle}$. The first follows a quenched mean field (QMF) approach, where the process takes place on a fixed network, while the second expression is obtained considering the degree-based mean field (DMF) approach, where we assume that every node with the same degree is statistically equivalent.

[3]This expression is obtained in details in [17], assuming uncorrelated networks. In [14] the authors show that $\frac{\lambda}{\delta} > \frac{\langle k \rangle}{\langle k^2 \rangle}$, that is completely analogous to the expression in the table, since $\Lambda_{max} = \frac{\langle k^2 \rangle}{\langle k \rangle}$ for a random network generated by the configuration model [29] considering scale-free networks in which $P(k) \sim k^{-\zeta}$ and $2 < \zeta < \frac{2}{5}$. Observe that those expressions are valid on finite networks and the threshold tends to zero in the thermodynamic limit, since $\langle k^2 \rangle \to \infty$.

[4]Considering the Model II in [8], which takes into account the apathy of the individuals.

[5]The authors considered even more interactions, but did not evaluate most of the possibilities numerically. In addition, the transitions follow the notation used in [9] and are rates, not necessarily probabilities, since the authors follow a continuous time approach.

Downloaded from https://academic.oup.com/comnet/article-abstract/6/2/215/4060524
by University of Manchester user
on 26 March 2018

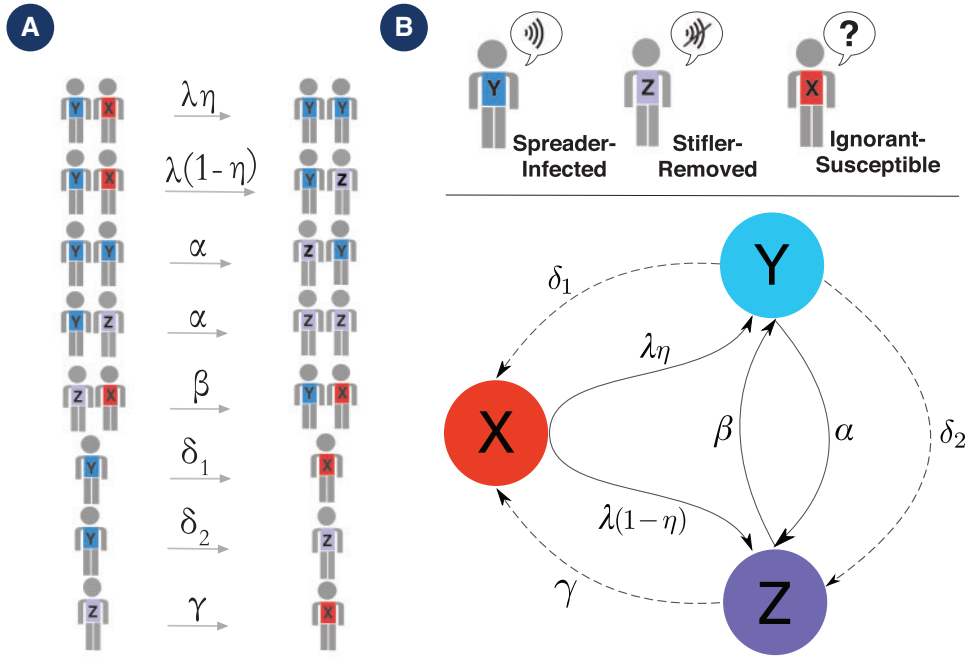

FIG. 1. A simplified diagram of all possible transitions between the three different dynamical states in our model. X, Y and Z stand for ignorant or susceptible, spreader and stifler or recovered, respectively. Most of the transitions involve interactions between two individuals, however, spontaneous ones are also allowed (represented by dashed lines).

(iii.1) If the individual contacted (the one at the other end of the edge) is in state X, then the latter turns into Y with probability $\lambda\eta$. This transition is the classical susceptible→infected one in disease models for $\eta = 1$. Traditional rumour models also include the very same transition (ignorant→spreader, also for $\eta = 1$), but here, in order to be as general as possible, we also consider that an ignorant X can directly go to the stifler class Z with probability $(1 - \eta)$, thus the probability that a transition $X \to Z$ occurs is $\lambda(1 - \eta)$. This mimics information dissemination in systems like Twitter, in which reading the tweet does not imply that the user spreads it —as a matter of fact, the most common situation is that in which the user does not retweet the piece of news [8].

(iii.2) On the contrary, if the neighbour of the spreader i is in state Y or Z, then i turns into Z with probability α. Note that this spreader→stifler transition is only found in rumour models and it is not present in disease contagion.

(iv) Finally, at variance with traditional rumour models [12, 15], we also ascribe an active role to stiflers Z. We assume that individuals in state Z can go back to state X with probability γ. This represents scenarios in which stiflers might spontaneously 'forget' the rumour, thus becoming ignorants again. We also note that this transition can be identified with a recovered→susceptible one in SIRS disease models, in which it is assumed that after some time, individuals might lose their acquired immunity and become susceptible to catch the same disease again.

(v) If (iv) does not happen, which occurs with probability $(1 - \gamma)$, then the node i in state Z interacts with its neighbours. If the contact is with a subject in state X (ignorant), the stifler (state Z) might

Downloaded from https://academic.oup.com/comnet/article-abstract/6/2/215/4060524
by University of Manchester user
on 26 March 2018

recover the interest on the rumour propagation and with probability β turns into Y. This transition mimics cases in which an individual who knows the rumour but is not transmitting it, learns that the rumour is still active and new, and therefore starts spreading it again.

In summary, our model presents the following transitions:

$$\begin{aligned}
Y + X &\xrightarrow{\lambda\eta} Y + Y,\\
Y + X &\xrightarrow{\lambda(1-\eta)} Y + Z,\\
Y + Y &\xrightarrow{\alpha} Z + Y,\\
Y + Z &\xrightarrow{\alpha} Z + Z,\\
Z + X &\xrightarrow{\beta} Y + X,\\
Y &\xrightarrow{\delta_1} X,\\
Y &\xrightarrow{\delta_2} Z,\\
Z &\xrightarrow{\gamma} X,
\end{aligned}$$

and our next goal will be to describe a process with such transitions through a suitable Markov chain approach.

3.1 *First order approximation*

In order to describe the evolution of this phenomenon in a given network made up by a set of nodes $[N] := \{1, 2, \ldots, N\}$, and adjacency matrix $\mathbf{A}$, we construct a discrete time Markov chain $(\xi_t)_{t\geq 0}$ with state space $\mathscr{S} = \{(1,0,0),(0,1,0),(0,0,1)\}^{[N]}$. More precisely, we define $\xi_t := \{(X_i(t), Y_i(t), Z_i(t)) : i \in [N]\}$, where $X_i(t)$, $Y_i(t)$ and $Z_i(t)$ are Bernoulli random variables indicating whether the node $i \in [N]$ is an ignorant (susceptible), a spreader (infected), or a stifler (recovered) at time t, respectively. So, $(1,0,0)$, $(0,1,0)$ and $(0,0,1)$ represent the states X, Y and Z, respectively. Each point $\xi \in \mathscr{S}$ is called configuration. To construct the Markov chain, we consider random objects defined on the same suitable probability space $(\Omega, \mathscr{F}, \mathbb{P})$, where Ω is the sample space, $\mathscr{F}$ is a σ-algebra of subsets of Ω, that is, the set of configurations, and $\mathbb{P}$ is a probability measure function. For every $i, j \in [N]$ we take the following independent collections, each of independent and identically distributed (i.i.d.) random variables

Collection	Associated to	Influence a choice on
$\{U_i(t)\}_{t\geq 0}$	Y	state Y
$\left\{I_i^{\gamma}(t)\right\}_{t\geq 0}$	Z	state Z
$\left\{I_{ij}^{\lambda}(t)\right\}_{t\geq 0}$	$Y + X$	state Y
$\left\{I_i^{\alpha}(t)\right\}_{t\geq 0}$	$Y + Y$ or $Y + Z$	state Y
$\left\{I_i^{\beta}(t)\right\}_{t\geq 0}$	$Z + X$	state Z
$\left\{I_i^{\eta}(t)\right\}_{t\geq 0}$	$Y + X$	state X

(3.1)

Downloaded from https://academic.oup.com/comnet/article-abstract/6/2/215/4060524
by University of Manchester user
on 26 March 2018

where $U_i(1)$ is a random variable uniformly distributed on $(0, 1)$, and $I_i^{\gamma}(1), I_{ij}^{\lambda}(1), I_i^{\alpha}(1), I_i^{\beta}(1)$ and $I_i^{\eta}(1)$ are Bernoulli random variables with parameter $\gamma, \lambda, \alpha, \beta$ and η, respectively. In addition, for each node i, we consider a sequence of i.i.d. random objects uniformly distributed on the neighbours of i, that is, a sequence $U_i^{nb}(1), U_i^{nb}(2), \ldots$ such that $P(U_i^{nb}(1) = j) = 1/k_i$, for all $j \in [N]$ satisfying $A_{ij} = 1$, where $k_i = \sum_{j\in[N]} A_{ij}$ is the number of neighbours of node i.

The main idea is to define a stochastic process that evolves according to the realization of the random variables defined above. For example, let's think of a rumour process. If at a fixed time t the node i is in state Y, then it forgets the rumour if $U_i(t) < \delta_1$ (transition (i) above), it loses the interest on the propagation whenever $\delta_1 \leq U_i(t) < \delta_1 + \delta_2$ (transition (ii) above), or contacts its neighbours when $U_i(t) \geq \delta_1 + \delta_2$ (i.e., transition (iii) above). In the last situation, if node i contacts node j, and j is in state X, then the rumour is propagated from i to j if $I_{ij}^{\lambda}(t) = 1$. On the other hand, if the contacted node j is a spreader, then i turns into a stifler, that is, $I_i^{\alpha}(t) = 1$. We can proceed in a similar way to represent the remaining transitions and interactions of the process. Given the above description, it is not difficult to see that the transitions of this stochastic process can be written, for each $i \in [N]$ and $t \geq 0$, as:

$$\begin{cases} X_i(t+1) &= X_i(t)A_i(t) + Y_i(t)1_{\{U_i^t<\delta_1\}} + Z_i(t)I_i^{\gamma}(t) \\ Y_i(t+1) &= X_i(t)(1-A_i(t))I_i^{\eta}(t)+ \\ & \quad Y_i(t)1_{\{U_i^t\geq\delta_1+\delta_2\}}B_i(t)+ \\ & \quad Z_i(t)(1-I_i^{\gamma}(t))(1-C_i(t)), \\ Z_i(t+1) &= X_i(t)(1-A_i(t))(1-I_i^{\eta}(t))+ \\ & \quad Y_i(t)\left(1_{\{U_i^t\geq\delta_1+\delta_2\}}(1-B_i(t)) + 1_{\{\delta_1\leq U_i^t<\delta_1+\delta_2\}}\right)+ \\ & \quad Z_i(t)(1-I_i^{\gamma}(t))C_i(t), \end{cases} \tag{3.2}$$

where $1_{\{A\}}$ is the indicator function, which is equal to 1 if the condition A is satisfied and 0 otherwise and $A_i(t)$, $B_i(t)$ and $C_i(t)$ are Bernoulli random variables indicating that a node i, given the influence of its neighbours, will not be informed, will not become a stifler, or will not recover the interest in rumour propagation, from time t to time $t + 1$, respectively. Observe that these variables depend on the contacts between node i and its neighbours.

Here we study two limiting cases, namely, the CP, in which each node performs only one contact per unit time; and the fully RP, in which each vertex contacts all its neighbours at each time step. The contact based variables for the CP are given by

$$\begin{cases} A_i(t) = \prod_{j=1}^{N}\left[1 - I_{ji}^{\lambda}(t)1_{\{U_j^{nb}(t)=i\}}1_{\{U_j^t\geq\delta_1+\delta_2\}}Y_j(t)\right], \\ B_i(t) = 1 - \sum_{j=1}^{N}\left[I_i^{\alpha}(t)1_{\{U_i^{nb}(t)=j\}}\left(Y_j(t)+Z_j(t)\right)\right], \\ C_i(t) = 1 - \sum_{j=1}^{N}\left[I_i^{\beta}(t)1_{\{U_i^{nb}(t)=j\}}X_j(t))\right]. \end{cases} \tag{3.3}$$

Downloaded from https://academic.oup.com/comnet/article-abstract/6/2/215/4060524
by University of Manchester user
on 26 March 2018

On the other hand, the contact-based variables for the RP are

$$\begin{cases} A_i(t) = \prod_{j=1}^{N} \left[1 - I_i^{\lambda}(t) A_{ji} 1_{\{U_j^t \geq \delta_1 + \delta_2\}} Y_j(t)\right], \\ B_i(t) = \prod_{j=1}^{N} \left[1 - I_i^{\alpha}(t) A_{ij} \left(Y_j(t) + Z_j(t)\right)\right], \\ C_i(t) = \prod_{j=1}^{N} \left[1 - I_i^{\beta}(t) A_{ij} X_j(t))\right]. \end{cases} \tag{3.4}$$

Therefore, the contact-based variables of the node i, that is, $A_i(t)$, $B_i(t)$ and $C_i(t)$ in equations (3.2), (3.3) and (3.4) are dependent on the parameters λ, α and β, the state of its neighbours and the process, that is, whether it is a CP or a RP. The variables $B_i(t)$ and $C_i(t)$ describe the feedback from the contacts to the node. Note that in rumour dynamics, interactions can also change the state of the spreading node, returning an immediate feedback, in contrast to epidemic spreading models [1] where the state of the spreader changes at a rate that is independent of the interaction network. These relations modelled by $B_i(t)$ and $C_i(t)$ are absent in the recent work by Stanoev *et al.* [30], since their formalism does not allow for instantaneous feedback over node i. We also highlight that the CP and RP mimic different situations in, for instance, social networks. The CP models one-to-one communication, when the rumours spread in friendship networks, email networks or networks in which each individual interacts with only one contact at each time step. On the other hand, the RP is best suited for one-to-many information dissemination, as in Twitter-like networks, since when a user posts a tweet, all its followers receive this information.

Finally, we also point out that Equation (3.2) ensures the existence of a function f such that

$$\xi_{t+1} = f\left(\xi_t, \mathscr{I}_t\right),$$

where

$$\mathscr{I}_t := \bigcup_{i=1}^{N} \left\{U_i(t), I_i^{\gamma}(t), I_{ij}^{\lambda}(t), I_i^{\alpha}(t), I_i^{\beta}(t), I_i^{\eta}(t), U_i^{nb}(t)\right\}$$

is the collection of all the (independent) random variables at time t. This in turns implies that $(\xi_t)_{t\geq 0}$ is a discrete-time Markov chain.

3.2 *The mean-field approximation*

Although the above Markovian description of the model is exact, its analysis is rather complex. This is because, while the Markov chain $(\xi_t)_{t\geq 0}$ is defined from the realizations of independent random variables through (3.2), the state of each node as well as of its neighbours have dependencies, whose complexity might change according to the network. We solve this difficulty by considering a mean-field approximation for the Markovian description. We denote the probabilities that a node i is in state X, Y and Z at time t by $p_i^X(t)$, $p_i^Y(t)$ and $p_i^Z(t)$, respectively, and note that

$$\begin{aligned} p_i^X(t) &:= \mathbb{P}\left(X_i(t) = 1\right) = \mathbb{E}(X_i(t)), \\ p_i^Y(t) &:= \mathbb{P}\left(Y_i(t) = 1\right) = \mathbb{E}(Y_i(t)), \\ p_i^Z(t) &:= \mathbb{P}\left(Z_i(t) = 1\right) = \mathbb{E}(Z_i(t)), \end{aligned} \tag{3.5}$$

Downloaded from https://academic.oup.com/comnet/article-abstract/6/2/215/4060524
by University of Manchester user
on 26 March 2018

where $\mathbb{E}$ denotes expected values. Therefore, the mean-field approximation is obtained by considering expected values in equations (3.2), (3.3) and (3.4). We also assume that there are no dynamical correlations at first order. In other words, we assume that the expected values of variable pairs factorize. The resulting system of equations obtained for a node $i \in [N]$, for all $t \geq 0$, is given by

$$\begin{cases} p_i^X(t+1) = & p_i^X(t)a_i(t) + p_i^Y(t)\delta_1 + p_i^Z(t)\gamma \\ p_i^Y(t+1) = & p_i^X(t)(1-a_i(t))\eta + \\ & p_i^Y(t)(1-\delta_1-\delta_2)b_i(t) + \\ & p_i^Z(t)(1-\gamma)(1-c_i(t)), \\ p_i^Z(t+1) = & p_i^X(t)(1-a_i(t))(1-\eta) + \\ & p_i^Y(t)(1-\delta_1-\delta_2)(1-b_i(t)) + p_i^Y(t)\delta_2 \\ & p_i^Z(t)(1-\gamma)c_i(t), \end{cases} \tag{3.6}$$

where $a_i(t)$, $b_i(t)$ and $c_i(t)$ are the probabilities that a node i, given a contact with its neighbours, remains at state X (e.g., it's not informed), does not change to state Z from state Y (that is, will not become a stifler), or remains at state Z (i.e., will not recover the interest in rumour propagation) in the time interval from t to $t+1$, respectively (see also Fig. 1). These probabilities depend on the number of contacts per unit time. Thus, the contact based probabilities for the CP are given by

$$\begin{cases} a_i(t) = \prod_{j=1}^N \left[1 - \lambda P_{ij}(1-\delta_1-\delta_2)p_j^Y(t)\right], \\ b_i(t) = 1 - \sum_{j=1}^N \left[\alpha P_{ij}\left(p_j^Y(t) + p_j^Z(t)\right)\right], \\ c_i(t) = 1 - \sum_{j=1}^N \left[\beta P_{ij} p_j^X(t))\right], \end{cases} \tag{3.7}$$

where $P_{ij} := A_{ij}/k_i$. On the other hand, the contact based probabilities for the RP are

$$\begin{cases} a_i(t) = \prod_{j=1}^N \left[1 - \lambda A_{ji}(1-\delta_1-\delta_2)p_j^Y(t)\right], \\ b_i(t) = \prod_{j=1}^N \left[1 - \alpha A_{ij}\left(p_j^Y(t) + p_j^Z(t)\right)\right], \\ c_i(t) = \prod_{j=1}^N \left[1 - \beta A_{ij} p_j^X(t))\right]. \end{cases} \tag{3.8}$$

The system of equations (3.6) describes the micro-state evolution of the system, that is, the evolution of the probabilities of each node to be in a given state. The macro-state variables can be defined as the average of the individual probabilities, namely,

$$\rho^\ell = \frac{1}{N}\sum_{i=1}^N p_i^\ell(\infty), \tag{3.9}$$

where $p_i^\ell(\infty)$ is the probability of node i to be in state ℓ, with $\ell \in \{X, Y, Z\}$, as $t \to \infty$. Note that the generalization for weighted and directed networks is straightforward. In weighted networks, it is

Downloaded from https://academic.oup.com/comnet/article-abstract/6/2/215/4060524
by University of Manchester user
on 26 March 2018

necessary to consider the weight matrix whose rows must sum up to one. In directed networks, the degree must be substituted by the out-degree, k_{out}. For the CP case, the directed network must be an ergodic Markov chain. Another possibility of generalization is the consideration of heterogeneous parameters, which is treated naturally in the set of equations (3.6), (3.7) and (3.8). In the following analysis we assume that the parameters have the same values for every nodes. In addition, we consider all networks as undirected and unweighted.

Furthermore, it is important to emphasize the differences between our model and the continuous formulation and consequently its embedded Markov chain, which is also a discrete time approach. In the continuous formulation, the probability of two events taking place at the same time vanishes as $\Delta t \to 0$. This implies that just one contact or one spontaneous process happens at a time. In the same spirit, the embedded Markov chain also captures this dynamics, however the time in this approach is discrete. Finally, we also note that in our approach every node performs its contacts in a time step -no matter whether a CP or a RP. Note that our formulation is similar to a cellular automaton, since every node changes its probabilities based on their neighbour probabilities.

3.3 *Homogeneously mixed population*

For the sake of comparison and completeness, we next study our model on a homogeneously mixed population, that is, we evaluate our model on a complete graph when the number of nodes goes to infinite. For a complete graph with N nodes, and for all $i \in [N]$, we have that $A_{ij} = 1$, for all $j \in [N] \setminus \{i\}$, which implies $P_{ij} = \frac{1}{N-1}$. We obtain for the CP, from the set of equations (3.7), and by taking the thermodynamic limit $N \to \infty$, that the probabilities that a node does not perform a transition to another state after a pairwise interaction at time t are given by

$$\begin{cases} a(t) = e^{\left[-\lambda(1-\delta_1-\delta_2)p^Y(t)\right]}, \\ b(t) = 1 - \alpha\left[p^Y(t) + p^Z(t)\right], \\ c(t) = 1 - \beta p^X(t). \end{cases} \tag{3.10}$$

Note that we omit the index i in our notation, since the probabilities are the same for every node. We observe that the probability $a(t)$ decays exponentially fast as a function of $p^Y(t)$, with a parameter $\tau^{-1} = \lambda(1 - \delta_1 - \delta_2)$. After including the probabilities (3.10) into equation (3.6), we obtain the steady state behaviour of this simplified model shown in Fig. 2. As it can be seen, the expected values for the densities of ignorant, spreaders and stiflers depend on α, which is the tuning parameter in this scenario.

Interesting enough, the behaviour of the RP scenario is radically different. For the RP, after taking the limit $N \to \infty$ in Equation (3.8), we get that $a(t) = 0$, $b(t) = 0$ and $c(t) = 0$, since those probabilities are based on the product of infinite terms with absolute values less than unity. As a consequence, the system of equations (3.6) for the RP becomes in the limit $N \to \infty$

$$\begin{cases} p^X(t+1) = p^Y(t)\delta_1 + p^Z(t)\gamma, \\ p^Y(t+1) = p^X(t)\eta + p^Z(t)(1-\gamma), \\ p^Z(t+1) = p^X(t)(1-\eta) + p^Y(t)(1-\delta_1). \end{cases} \tag{3.11}$$

The evolution of the fraction of spreaders, ignorants and stiflers are also shown in Fig. 2. Note that these quantities are constant in α in sharp contrast to the results for the CP.

Downloaded from https://academic.oup.com/comnet/article-abstract/6/2/215/4060524
by University of Manchester user
on 26 March 2018

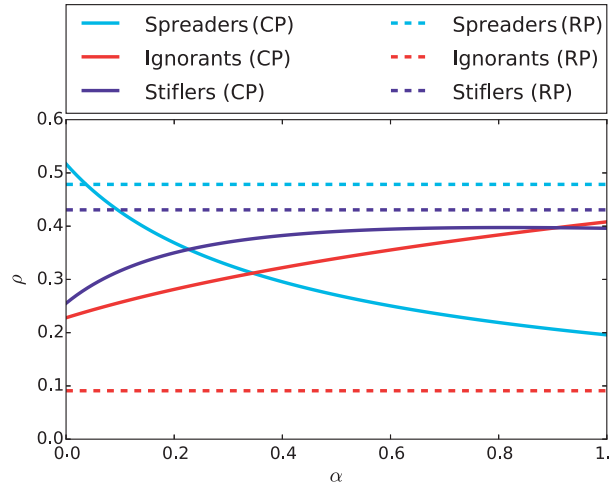

FIG. 2. Phase diagram of the steady state ($t \to \infty$) considering the complete infinity graph. The dynamical parameters have been set to $\delta_1 = \delta_2 = \gamma = 0.1$, $\lambda = 1$, $\beta = 0.5$ and $\eta = 1$.

4. Analytical analyses

In this section, we evaluate our model in terms of its early time behaviour and perform the steady state analysis. We derivate some closed expressions for the thresholds of the CP and RP cases in terms of the spectra of the probability transition and the adjacency matrices, respectively.

4.1 *Early time behaviour*

Consider the first steps of the process. For a small time $\tilde{t}$ we may assume that $p_i^Y(t) \approx \epsilon_i^Y$, and $p_i^Z(t) \approx \epsilon_i^Z$, for any $t \leq \tilde{t}$, where ϵ_i^Y and ϵ_i^Z are constants such that $0 \leq \epsilon_i^Y \ll 1$, and $0 \leq \epsilon_i^Z \ll 1$. This in turns implies that $p_i^X(t) = 1 - p_i^Y(t) - p_i^Z(t) \approx 1 - \epsilon_i^Y - \epsilon_i^Z \approx 1$. This approximation takes into account that the information spreading starts from at most a few spreaders. Neglecting second-order terms for the RP in equation (3.8) we obtain, for $t \leq \tilde{t}$

$$\begin{cases} a_i(t) \approx 1 - \sum_{j=1}^N \left[\lambda A_{ji}(1-\delta_1-\delta_2)p_j^Y(t)\right], \\ b_i(t) \approx 1 - \sum_{j=1}^N \left[\alpha A_{ij}\left(p_j^Y(t) + p_j^Z(t)\right)\right], \\ c_i(t) \approx 1 - \sum_{j=1}^N \left(\beta A_{ij} p_j^X(t)\right)). \end{cases} \tag{4.1}$$

Notice that

$$c_i(t) \approx 1 - \beta k_i, \tag{4.2}$$

Downloaded from https://academic.oup.com/comnet/article-abstract/6/2/215/4060524
by University of Manchester user
on 26 March 2018

since $p_i^X(t) \approx 1$, for $i = 1, 2, \ldots, N$. Substituting equations (4.1) and (4.2) in the system of equations (3.6), we obtain, for the equations of p_i^X and p_i^Y,

$$\begin{cases} -\lambda(1-\delta_1-\delta_2)\sum_{j=1}^N A_{ij}\epsilon_j^Y + \delta_1\epsilon_i^Y + \gamma\epsilon_i^Z \approx 0, \\ \eta\lambda(1-\delta_1-\delta_2)\sum_{j=1}^N A_{ij}\epsilon_j^Y + (1-\delta_1-\delta_2)\epsilon_i^Y + \beta k_i(1-\gamma)\epsilon_i^Z \approx \epsilon_i^Y, \end{cases} \tag{4.3}$$

where $\epsilon_j^Y + \epsilon_j^Z \approx 0, \forall j = 1, 2, \ldots, N$. Isolating ϵ_i^Z on the second equation, we get

$$\epsilon_i^Z = \frac{(\delta_1+\delta_2)\epsilon_i^Y - \eta\lambda(1-\delta_1-\delta_2)\sum_{j=1}^N A_{ij}\epsilon_j^Y}{\beta k_i(1-\gamma)}, \tag{4.4}$$

which describes ϵ_i^Z as a combination of ϵ_j^Y, $\forall j = 1, 2, \ldots N$. Substituting equation (4.4) on the first equation of (4.3) we have

$$\left(-1 - \frac{\eta\gamma}{\beta k_i(1-\gamma)}\right)\lambda(1-\delta_1-\delta_2)\sum_{j=1}^N A_{ij}\epsilon_j^Y + \delta_1\epsilon_i^Y + \gamma\frac{\epsilon_i^Y(\delta_1+\delta_2)}{\beta k_i(1-\gamma)} \approx 0, \tag{4.5}$$

whose factorization is

$$\sum_{j=1}^N \left[A_{ij} - \delta_{ij}\left(\frac{\gamma(\delta_1+\delta_2) + \delta_1(1-\gamma)\beta k_i}{\lambda(1-\delta_1-\delta_2)((1-\gamma)\beta k_i + \gamma\eta)}\right)\right]\epsilon_i^Y \approx 0. \tag{4.6}$$

Notice that equation (4.6) does not decouple the structure and the dynamics of the system, since it is not possible to isolate the terms depending on k_i. However, it is possible to evaluate the threshold for $\beta(1-\gamma) \approx 0$. In this case, the system has non-trivial solutions when $\left(\frac{\delta_1+\delta_2}{\eta\lambda(1-\delta_1-\delta_2)}\right)$ is an eigenvalue of A. Thus, the critical value, which depends on the largest eigenvalue of the adjacency matrix $\mathbf{A}$, is given by

$$\Lambda_{\max}^A \approx \left(\frac{\delta_1+\delta_2}{\eta\lambda(1-\delta_1-\delta_2)}\right)_c, \tag{4.7}$$

where $(\cdot)_c$ denotes the critical point. Considering the critical point as a function of λ, that is, assuming that the other dynamical parameters are fixed, the threshold vanishes in the thermodynamic limit for scale-free networks with a divergent second moment, similarly to epidemic spreading [6].

The numerical evaluation of equation (4.7) is shown in Fig. 3. As it can be seen, there is a good agreement between the theoretical predictions (equation (4.7)) and the numerical solution of the system (3.6). The calculation was performed for a scale-free network with $\zeta \approx 2.7$, $N = 10^4$ and $\langle k \rangle \approx 10$. As initial conditions, we considered that each node is set as a spreader with uniform probability $Y_i(0) = 0.01$, $\forall i \in \{1, 2, \ldots N\}$. The values of λ in the figure are close to zero, since λ depends on the inverse of the leading eigenvalue. We can also see that the limiting cases for α do not change the critical point. Furthermore, the density of nodes that holds the rumour, that is, of spreaders and stiflers, is higher for $\alpha = 0$ and $\eta = 1$, while the lowest density is obtained for $\alpha = 1$ and $\eta = 0.5$. Such observation reinforces the role of stifling rates, similarly to what happens in classical rumour models.

The approximation of the critical point in the RP is restricted by $\beta(1-\gamma) \approx 0$, since there is an explicit dependence on the degree k_i. However, for the CP the critical point can be evaluated without

Downloaded from https://academic.oup.com/comnet/article-abstract/6/2/215/4060524
by University of Manchester user
on 26 March 2018

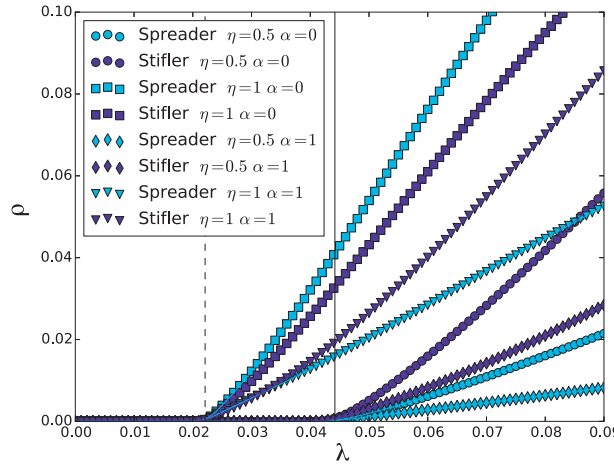

FIG. 3. Numerical evaluation of equation (3.6) considering the RP for $\beta = 0$, $\delta_1 = \gamma = 0.25$, $\delta_2 = 0.2$ and varying the spreading probability λ. The network considered is a scale-free network with $\zeta \approx 2.7$, $N = 10^4$ and $\langle k \rangle \approx 10$. The dashed line indicates the critical point (equation (4.7)) for $\eta = 1$, whereas the continuous line corresponds to the critical value for the case in which $\eta = 0.5$.

any constraint. The contact based probability in equation (3.7) can be obtained using the same set of equations as in the RP (equation (4.1)). Notice that only the first equation, that is, that for $a_i(t)$, is an approximation, whereas the equations for $b_i(t)$ and $c_i(t)$ are exact. Following the same approach as for the RP, the expression for the critical point for the CP is similar to that in equation (4.6) — the only two changes are (i) the use of the matrix $\mathbf{P}$ instead of $\mathbf{A}$ and (ii) $\sum_j P_{ij} = 1, \forall j = 1, 2, ...N$. Another important result is that the leading eigenvalue of the transition probability matrix is always equal to unity [31]. In this way, the critical point for the CP is given by

$$\Lambda^P_{max} = 1 \approx \left(\frac{\gamma\,(\delta_1 + \delta_2) + \delta_1(1-\gamma)\beta}{\lambda(1-\delta_1-\delta_2)\,((1-\gamma)\beta + \gamma\eta)} \right)_c . \tag{4.8}$$

Note that such expression does not depend on the network structure, since the leading eigenvalue Λ^P_{max} is the same for every connected network.

The numerical evaluation of the expression (4.8) is shown in Fig. 4 considering a scale-free network with $\zeta \approx 2.7$, $N = 10^4$ and $\langle k \rangle \approx 100$. In (a) we consider $\eta = 0.5$ and $\alpha = 1$, while in (b) we assume $\eta = 1$. Similarly to the RP case, here we also observe a very good agreement between the theoretical results and the numerical evaluation of the system (3.6) for all parameter values explored. The conclusions thus are similar as for the RP scenario, being the only difference the average degree of the network, a dependency that we shall analyse in more details later on.

The parameters of the model (with the exception of α that does not affect first order terms) can be used as control parameters of the system's dynamics. In particular, it is interesting to analyse the effect of β, which has been introduced here to account for the possibility of recovering the interest in the rumour. Isolating β on equation (4.8) we obtain

$$\beta_c \approx \frac{\gamma(\delta_1 + \delta_2 - \eta\lambda(1-\delta_1-\delta_2))}{\lambda(1-\delta_1-\delta_2) - \delta_1(1-\gamma)}, \tag{4.9}$$

Downloaded from https://academic.oup.com/comnet/article-abstract/6/2/215/4060524
by University of Manchester user
on 26 March 2018

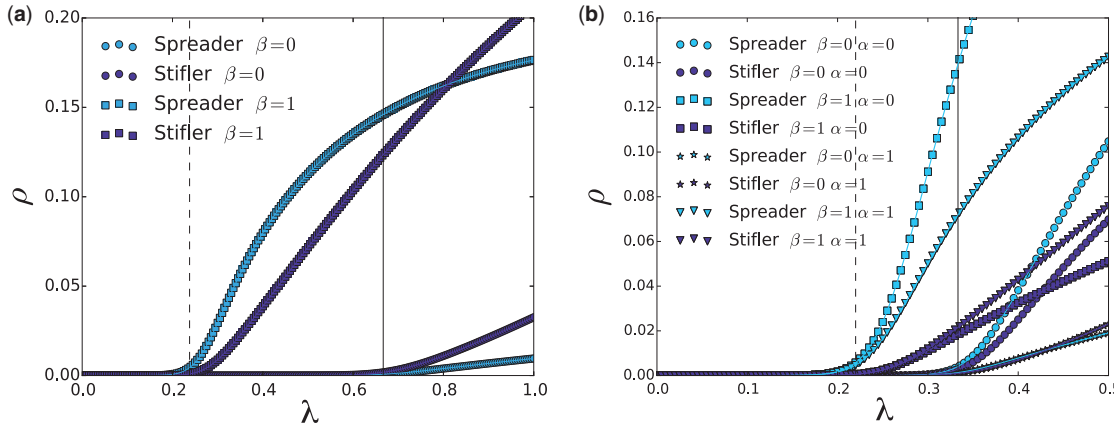

FIG. 4. Numerical evaluation of equation (3.6) considering the CP with parameters $\delta_1 = \gamma = 0.15$, $\delta_2 = 0.1$ and varying the spreading probability λ. The network considered is a scale-free network with $\zeta \approx 2.7$, $N = 10^4$ and $\langle k \rangle \approx 100$. In panel (a) $\eta = 0.5$ and $\alpha = 1$, whereas in (b) we have set $\eta = 1$ and also explored the limiting cases of $\alpha = 0$ and $\alpha = 1$. The vertical lines correspond to the critical point. Equation (4.7) evaluated for the parameters used here.

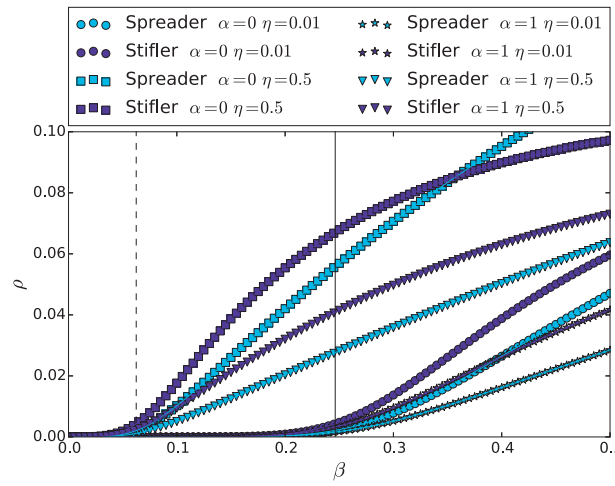

FIG. 5. Numerical evaluation of equation (3.6) considering the CP and, without loss of generality, fixing the parameters $\delta_1 = \gamma = \delta_2 = 0.2$, $\lambda = 1$ and varying the parameter β. The network considered is a scale-free network with $\zeta \approx 2.7$, $N = 10^4$ and $\langle k \rangle \approx 100$.

which defines the minimum value of β that allows for the spreading of the rumour to a macroscopic fraction of the population for a given spreading rate λ —as it is the case in most applications. Figure 5 shows the evaluation of equation (3.6) for the CP near the critical point considering a scale-free network with $\zeta \approx 2.7$, $N = 10^4$ and $\langle k \rangle \approx 100$, and, without loss of generality, fixing the parameters $\delta_1 = \gamma = \delta_2 = 0.2$, $\lambda = 1$ and varying the parameter β. Similarly to the results shown in Figs 3 and 4, our approximations agree with the simulated values.

As for the other important parameter, α, that defines at which rate spreaders decay into stiflers after $Y - Y$ or $Y - Z$ interactions, we stress that any approximation neglecting second-order terms does not

Downloaded from https://academic.oup.com/comnet/article-abstract/6/2/215/4060524
by University of Manchester user
on 26 March 2018

involve the parameter α, since it controls the stifling rate on the expression for $b_i(t)$. In other words, b_i is a conditional probability, and it is multiplied by the probability of an individual being an spreader. Thus, α does not affect the thresholds for the RP, since it always appears on second order terms (see Fig. 3). The same occurs for the CP, as shown in Fig. 4. Note that while the macroscopic behaviour of the system (the expected final densities of the different classes) is different for each parameter analysed, the threshold is always the same. An equivalent analysis was done in [10], where the authors proposed a model with a rate similar to δ_2 (in our model) and showed the independence of the threshold on the parameter α.

4.2 *Steady state analysis*

On the previous section, we analysed the early time evolution of the system aiming at finding its critical point. In this section, for the sake of completeness, we perform a similar analysis for the steady state solution. An analogous approach was used in [7], but for the specific case of a SIS model. At the steady state, we can assume, for t large enough, $p_i^I(t) \approx \pi_i^I$, $p_i^Y(t) \approx \pi_i^Y$ and $p_i^R(t) \approx \pi_i^R$, for $i = 1, 2, \ldots, N$. Neglecting second-order terms, inserting (4.1) on the system (3.6) and after some algebra, the steady-state solution is

$$\begin{cases} \pi_i^I = \frac{(\delta_1+\delta_2\gamma+\delta_1(-1+\gamma)c_i)}{(\delta_2(1+\gamma-a_i)+\delta_1(2+\eta(-1+a_i)-a_i+(-1+\gamma)c_i)-(-1+a_i)(1-c_i+\gamma(-1+\eta+c_i)))} \\ \pi_i^Y = \frac{(-1+a_i)(1-c_i+\gamma(-1+\eta+c_i))}{(\delta_2(-1-\gamma+a_i)+\delta_1(-2+\eta+a_i-\eta a_i+c_i-\gamma c_i)+(-1+a_i)(1-c_i+\gamma(-1+\eta+c_i)))} \\ \pi_i^R = \frac{-(((\delta_1+\delta_2-\delta_1\eta)(-1+a_i))}{(\delta_2(1+\gamma-a_i)+\delta_1(2+\eta(-1+a_i)-a_i+(-1+\gamma)c_i)-(-1+a_i)(1-c_i+\gamma(-1+\eta+c_i))))}, \end{cases} \tag{4.10}$$

where the time dependence does not appear any more.

Considering that $Y_i = \epsilon_i$, where $0 \leq \epsilon_i \ll 1$, and after substitution in the second equation of the system (4.10), we obtain

$$\epsilon_i(\delta_2\gamma + \delta_1(c_i - \gamma c_i - 1)) = (a_i - 1)(1 - c_i + \gamma(\eta + c_i - 1)). \tag{4.11}$$

Taking into account the approximation for $a_i(t)$ in equation (4.1), we have

$$\epsilon_i\,(\delta_2\gamma + \delta_1(k_i\beta(\gamma - 1) - \gamma)) = -\lambda(1 - \delta_1 - \delta_2)((1 - \gamma)k_i\beta + \gamma\eta)\sum_{j=1}^{N} A_{ij}\epsilon_j, \tag{4.12}$$

which is the same equation as (4.6). This result is for the RP process. For the CP case, we only need to change the adjacency matrix **A** by the transition probability matrix **P**. It is also worth highlighting that for an arbitrary choice of model parameters, in general, and at variance with classical rumour models, there is no absorbing state corresponding to the absence of spreaders. Regarding the parameter configurations where we have infinitely many absorbing states in the thermodynamic limit, we mention the classical SIR and Maki–Thompson models. Besides, observe that the SI model also has an absorbing state, which is the configuration in which every node is infected.

5. Special cases

The model proposed includes new transitions, formulates rumour dynamics in terms of discrete Markov chains and generalizes several previous spreading models (see Section 2). It is instructive to show how to obtain some of the main epidemic and rumour models from our approach, which we do next.

Downloaded from https://academic.oup.com/comnet/article-abstract/6/2/215/4060524
by University of Manchester user
on 26 March 2018

5.1 *Disease spreading*

We can obtain a particular case of the SIS model by setting $\eta = 1$, $\delta_2 = \gamma = 0$, $\beta = 0$ and $\alpha = 0$ in our model (see Fig. 1). This model is not exactly a SIS model, since we assume that each node cannot spread the disease and become susceptible at the same time step. Such modification implies that the parameters used on the traditional models change. However, in both cases the SIS dynamics can be studied in terms of the final fraction of spreaders. From equation (4.7), the epidemic threshold is given by

$$\Lambda_{max} \approx \left(\frac{\delta_1}{\lambda(1-\delta_1)} \right)_c . \tag{5.1}$$

Comparing with the results in Table 1, the recovery probability in this SIS model is thus given by $\delta = \frac{\delta_1}{(1-\delta_1)}$, while the spreading probability is λ.

In [7], the authors used a discrete time Markov chain approach to model the SIS dynamics considering a reinfection term and a parameter that allows to explore a family of contact-based scenarios, including as limiting cases the CP and the fully RP. Concerning the reinfection, our model does not include such a feature, however we note that reinfection within the same time step is rarely taken into account in SIS like models. On the other hand, the second ingredient can be easily incorporated into our framework by exchanging the matrix **A** by the matrix

$$R_{ij} = 1 - \left(1 - \frac{A_{ij}}{k_i}\right)^{\psi_i} \tag{5.2}$$

where ψ_i is the activity parameter. Observe that the limiting cases are obtained setting $\psi_i = 1$, $\forall i = 1, 2, ..., N$ for the CP, and $\psi_i \to \infty$, $\forall i = 1, 2, ..., N$ for the fully RP.

The SIR model is obtained from our model by setting $\eta = 1$, $\delta_1 = \gamma = 0$, $\beta = 0$ and $\alpha = 0$. In this case, the recovery rate is $\delta = \frac{\delta_2}{(1-\delta_2)}$. Additionally, our framework also includes the SIRS scenario as advanced before. By considering the model introduced in [26] (see also Table 1) the SIRS scheme is recovered by setting $\eta = 1$, $\delta_1 = 0$, $\beta = 0$ and $\alpha = 0$. The system can then be written as a function of λ, δ_2 and γ. From equation (4.7), such reduction implies that the threshold is given by

$$\Lambda_{max} \approx \left(\frac{\delta_2}{\lambda(1-\delta_2)} \right)_c \tag{5.3}$$

which depends only on δ_2 and λ.

5.2 *Rumour spreading*

All rumour models in Table 1 can be obtained from our general model. However, it is worth mentioning that the DK model is not covered by our model, since it considers edge sampling and undirected contacts (see Section 2). The difference between our approach and the previous models is that our approach does not allow an individual to perform two transitions (or two attempts) at the same time step.

The MK model can be obtained by considering $\eta = 1$, $\beta = 0$, $\delta_1 = \delta_2 = \gamma = 0$ in equation (3.6), letting the system be a function of α and λ. Such model does not have a critical threshold. On the other hand, it is possible to obtain the threshold for the variant introduced by Nekovee *et al*. [10], which considers an spontaneous process to model the lost of interest in the propagation of the rumour. The

Downloaded from https://academic.oup.com/comnet/article-abstract/6/2/215/4060524
by University of Manchester user
on 26 March 2018

probability of turning a spreader into a stifler is called $\bar{\delta}$ in the original paper. Such parameter is similar to δ_2 here, which leads us to a threshold equivalent to that in [10]

$$\Lambda_{max} \approx \left(\frac{\delta_2}{\lambda(1-\delta_2)} \right)_c . \tag{5.4}$$

Finally, the model by Borge-Holthoefer *et al*. [8] considers a feature that is very common in human communication through online social networks, namely, apathy, in which an ignorant might turn into a stifler after being informed—this is often the case observed in Twitter for instance, where the fact that a tweet appears in a user's timeline does not directly imply that the user spreads it further. This feature has been used to identify influential spreaders in rumour-like dynamics [8, 32–34]. The model by Borge-Holthoefer *et al*. [8] is recovered from our approach by setting $\beta = 0$, $\delta_1 = \delta_2 = \gamma = 0$ in equation (3.6). Note that the system's dynamics depends on η, α and λ in this scenario.

6. Numerical analysis and simulations

Once we have got some analytical insights, we next compare results from extensive Monte Carlo (MC) simulations with the numerical solution of the system of equations (3.6) that describe the dynamics of the model, showing that they agree at the micro (e.g., at the individual level) and macro (e.g., at the system level) scales. We obtain the respective phase diagrams for the limiting cases of CP and RP for several combinations of the model's parameters as well as for different topologies of the underlying network of contacts.

6.1 *Phase diagrams*

First, we present the (λ, β) phase diagrams for the CP and the RP cases. They are obtained by solving the dynamical set of equations (3.6) and varying these parameters from 0 to 1 in intervals of $\Delta\lambda = \Delta\beta = 5 \times 10^{-2}$. In addition, when $\beta < 0.05$ ($\lambda < 0.05$), that is, near the critical region, we set $\Delta\lambda = \Delta\beta = 5 \times 10^{-3}$.

Figures 6 and 7 show the (λ, β) phase diagrams obtained from the numerical evaluation of the system of equations (3.6) for the CP and the RP, respectively. We represent (colour-coded) in the two figures the fraction of ignorants, spreaders and stiflers in panels (a), (b) and (c), respectively. The underlying network is the same in both cases (scale-free graph with $\zeta \approx 2.7$, $N = 10^4$), except for the average degree, which is $\langle k \rangle \approx 100$ in Fig. 6 and $\langle k \rangle \approx 10$ in Fig. 7. The rest of parameters have been set to $\eta = 0.01$, $\delta_1 = \delta_2 = \gamma = 0.1$ and $\alpha = 1.0$ for the CP and to $\eta = 0.01$, $\delta_1 = \delta_2 = \gamma = 0.25$ and $\alpha = 1.0$ for the RP. As discussed in the previous section, it is possible to obtain an analytical expression for the critical point in the CP scenario (equation (4.8)), which is indicated by the continuous black line. This is not possible for the RP.

The comparison of the results shown in Figs 6 and 7 indicates that there are important differences in the system's behaviour for the CP and RP schemes, notably with respect to the final fraction of stiflers, which is in most practical scenarios the quantity that we would like to be as higher as possible—as that would mean that the piece of information reached a large fraction of the population and therefore that it was spread efficiently. The CP exhibits a critical point that does not depend on the network structure, and although it cannot be evaluated for the RP in a closed form, the numerical solutions shows that this scenario is more complex with some interesting dependencies on the network structure and the dynamical parameters of the model. As a matter of fact, as it can be seen in Fig. 7c, there is a non-linear effect in

Downloaded from https://academic.oup.com/comnet/article-abstract/6/2/215/4060524
by University of Manchester user
on 26 March 2018

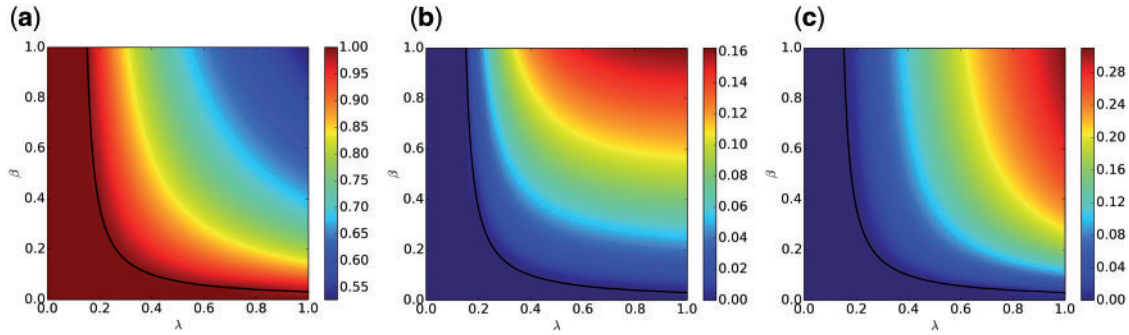

FIG. 6. (λ, β) phase diagrams at the stationary state for the CP case on a scale-free network with degree distribution $P(k) \sim k^{-\zeta}$ and $\zeta \approx 2.7$. The network is made up of $N = 10^4$ nodes and has an average degree $\langle k \rangle \approx 100$. The rest of dynamical parameters have been set to $\eta = 0.01$, $\delta_1 = \delta_2 = \gamma = 0.1$ and $\alpha = 1.0$. The intensity of the colour (as given by the legend) represents the final fraction of ignorants (panel (a)), spreaders (panel (b)) and stiflers (panel (c)). The continuous lines are the analytical values for the critical point.

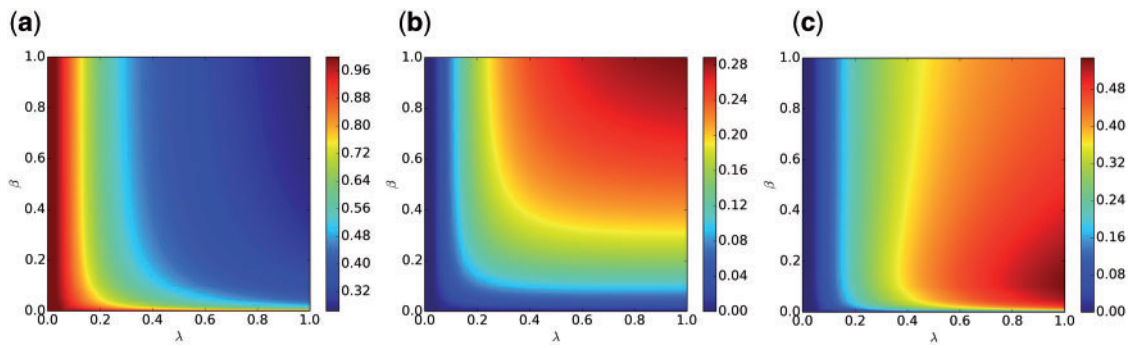

FIG. 7. (λ, β) phase diagrams as in Fig. 6 but for the RP. The underlying network of contacts has the same features except for the average degree that is $\langle k \rangle \approx 10$. The rest of parameters are $\eta = 0.01$, $\delta_1 = \delta_2 = \gamma = 0.25$ and $\alpha = 1.0$. The intensity of the colour (as given by the legend) represents the final fraction of ignorants (panel (a)), spreaders (panel (b)) and stiflers (panel (c)).

β for fixed values of λ. Admittedly, for large values of $\lambda \geq 0.7$, when β starts to increase from zero, the final density of stiflers also grows. However, at some point (roughly around $\beta \approx 0.1$), this density reaches its maximum value and starts decreasing beyond that value of β.

6.2 *Monte Carlo simulations*

In order to check whether our analytical and numerical solutions are accurate at describing both individual states and the macroscopic behaviour of the system, we have performed large-scale Monte Carlo simulations. All results reported henceforth are averages taken over at least 100 MC simulations with an initial fraction of spreaders and ignorants equal to $\rho^Y(0) = 0.01$ and $\rho^X(0) = 0.99$, respectively.

In Fig. 8, we present results for the dependency of the densities of X, Y and Z as a function of α with all the other model parameters fixed. The $\rho \times \alpha$ plots correspond to the CP (panel (a)) and the RP (panel (b)) scenarios. As it can be seen, the continuous lines, which are obtained by numerically evaluating equation (3.6) perfectly match results from MC simulations in both limiting cases. In order to check

Downloaded from https://academic.oup.com/comnet/article-abstract/6/2/215/4060524
by University of Manchester user
on 26 March 2018

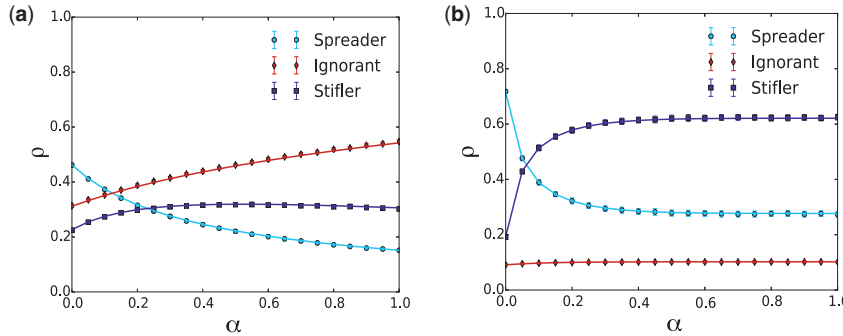

FIG. 8. Phase diagram at the steady state for (a) a CP considering a scale-free network with $\zeta \approx 2.7$, $N = 10^4$ and $\langle k \rangle \approx 100$, and (b) a RP simulated on top of a scale-free network with $\zeta \approx 2.7$, $N = 10^4$ and $\langle k \rangle \approx 10$. The simulations have been carried out using the following dynamical parameters: $\delta_1 = \delta_2 = \gamma = 0.1$, $\lambda = 1$, $\beta = 0.5$ and $\eta = 1$. The continuous lines are the theoretical predictions, whereas the symbols are the results of MC simulations. The standard deviation has approximately the size of the symbols.

whether this agreement is also verified at the individual scale, we have represented in a scatter plot, see Fig. 9, the probability of finding an individual in the ignorant, spreader or stifler class at the stationary state. Despite of finite size effects and stochastic fluctuations, the analytical results capture the micro dynamical states in the large t limit, since there is a strong correlation between the probabilities obtained from MC and the solutions of the system equation(3.6). These results thus convincingly show that one can explore the model in terms of either the individual probabilities or the macroscopic expected values by solving the equations describing the system's dynamics, without the need to rely on extensive and costly MC simulations. This is, from a practical point of view, an added value of the proposed framework, as we shall discuss later on.

Spreading processes have been shown to be greatly affected by the topological features of the networks on top of which they take place. It is therefore of further interest to investigate network effects on the dynamical evolution of the spreading phenomenon. To this end, we have explored the impact of three characteristics of the contact networks, namely, the exponent of the degree distribution $P(k)$ for scale-free graphs—for which $P(k) \sim k^{-\zeta}$—, the size of these networks and finally how dense they are by tuning the average degree of the network's nodes. Figure 10 shows the results of these analyses. First, we note that in all cases, the previous agreement between the numerical solution of the system's equations and MC simulations still holds. Other aspects worth highlighting include the fact that the value of ζ influences the outcomes of the spreading process mainly for the CP, unless ζ is larger than 3. This is because for larger exponents, the networks are effectively equivalent to homogeneous graphs, and therefore there are no hubs anymore, leading to fairly similar results independently of the specific value of ζ. Another interesting effect of the network structure, and in particular of ζ, is the one observed in Fig. 10(b). Admittedly, the final density of ignorants is independent of this network parameter. In other words, variations of the exponent mainly affect the number of spreaders and stiflers, with the number of ignorants remaining roughly constant.

As for finite size effects, our results, see Fig. 10(c), show that they do not appear to play a major role on the rumour spreading dynamics. This is not the case with respect to the last parameter analysed, the average degree of the network. As it can be seen in Fig. 10(d), the steady-state fractions of spreaders and stiflers strongly depend on the density of connections, while the fraction of ignorants is (as it happened with respect to ζ) almost the same for different average degrees. In summary, it seems that the most

Downloaded from https://academic.oup.com/comnet/article-abstract/6/2/215/4060524
by University of Manchester user
on 26 March 2018

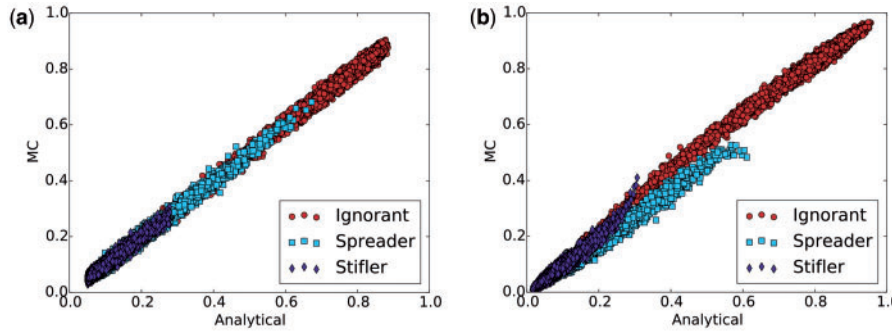

FIG. 9. Probability of finding a node i in any of the dynamical states of the model (X, Y or Z) obtained by solving analytically the system of equations (3.6) as compared with Monte Carlo simulations. The probabilities are calculated by averaging 10^3 simulations. Panel (a) corresponds to the CP, whereas the RP is represented in panel (b).

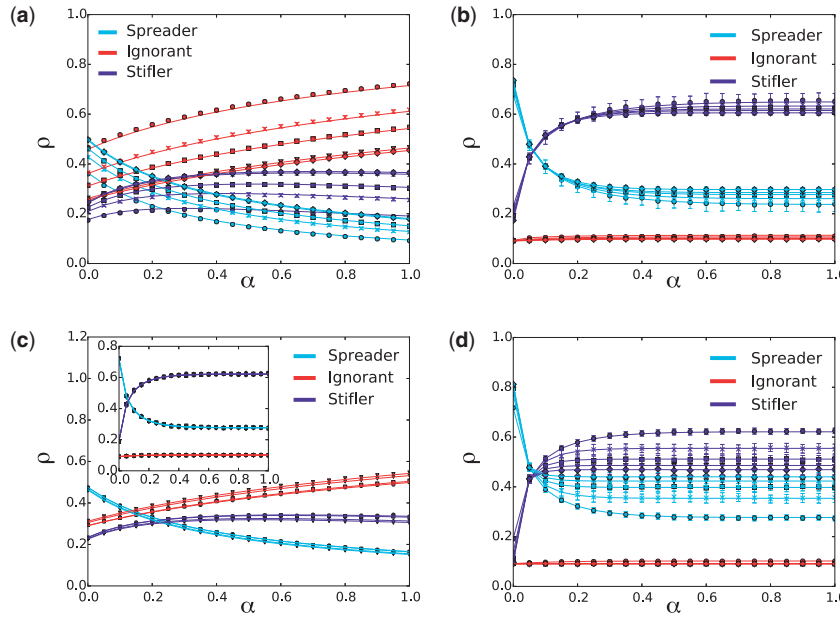

FIG. 10. Phase diagram of the system's dynamics at the steady state as a function of α. In all panels, the rest of parameters have been fixed to $\delta_1 = \delta_2 = \gamma = 0.1$, $\lambda = 1$, $\beta = 0.5$ and $\eta = 1$ and the continuous lines are the theoretical predictions, while the symbols correspond to the MC simulations. Panels (a) and (b) show results obtained varying the exponent ζ of the power law degree distribution, $P(k) \sim k^{-\zeta}$, of the underlying networks of contacts of size is $N = 10^4$ and $\langle k \rangle \approx 100$ (CP, panel (a)) and $\langle k \rangle \approx 10$ (RP, panel (b)). We considered the following exponents ζ: 2.3 (•), 2.5 (×), 2.7 (□), 3.3 (▽) and 3.5 (◇). The effects of the network size are shown in Panel (c) for scale-free networks ($P(k) \sim k^{-\zeta}$ with $\zeta \approx 2.7$) with $\langle k \rangle \approx 100$ (main plot) and $\langle k \rangle \approx 10$ (inset) and the following sizes: $N = 5 \times 10^2$ (•), $N = 10^3$ (□), $N = 5 \times 10^3$ (◇) and $N = 10^4$ (▽). Finally, in panel (d), we represent results obtained for the RP and different average degrees: $\langle k \rangle \approx 10$ (•), $\langle k \rangle \approx 20$ (×), $\langle k \rangle \approx 35$ (□), $\langle k \rangle \approx 45$ (▽) and $\langle k \rangle \approx 60$ (◇). The rest of network's parameters are the same as in panel (c).

Downloaded from https://academic.oup.com/comnet/article-abstract/6/2/215/4060524
by University of Manchester user
on 26 March 2018

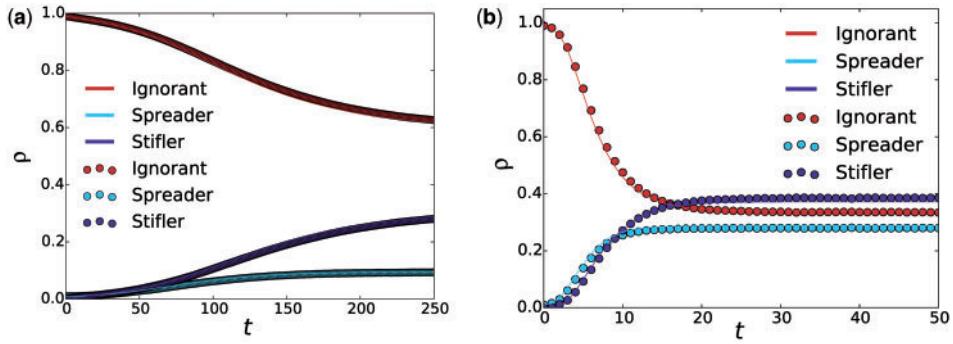

FIG. 11. Time evolution of the average probabilities for the (a) CP and (b) RP. The continuous lines are the theoretical predictions, while the symbols are obtained by Monte Carlo simulations.

important dependencies with respect to the networks' topological properties are given by its heterogeneity (characterized by ζ) and the density of connections (e.g., $\langle k \rangle$). Both features mainly affect the ratio between spreaders and stiflers at the stationary state, but do not significantly affect the final number of ignorants. Altogether, and from a practical point of view, one can then conclude that, with the exception of the heterogeneity of the degree distribution, the size and average degree of the networks have a somewhat minimal impact on the propagation of the rumour, at least in what concerns its main outcome: the number of individuals that learned the rumour—both spreaders and stiflers represent states in which individuals are aware of the news, that is, they have been reached by the contagion process (but also see below).

Finally, we also show Fig. 11 that the numerical solution of the system (3.6) and Monte Carlo simulations are in good agreement not only in the large t limit, but also when we explore the system's behaviour at intermediate times, that is, in transient states. The figure shows the evolution in time of the fraction of ignorants, spreaders and stiflers for both the CP and RP limiting cases when the substrate network is a scale-free graph with $\zeta \approx 2.7$, $N = 10^4$ and $\langle k \rangle \approx 100$. As we will see in the next section, the fact that the temporal dynamics can be captured accurately by numerically solving the set of equations (3.6) could be used to study real spreading dynamics for which highly resolved temporal data is available.

We round off this section by pointing out some limitations of the mean field approximation. Gleeson *et al*. [35] discussed that the accuracy of the mean field theory is higher when three main assumptions are satisfied, that is, (i) vanishing local clustering, (ii) non-modular network organization and (iii) absence of dynamical correlations. Conditions (i) and (ii) depend on the structure of the network. They are satisfied in our analysis, because we consider the configuration model, which satisfies the properties (i) and (ii). We will next analyse how our model performs in real online networks. However, the dynamical correlations might represent a major source of error, since we assume that the state of each node is independent of the state of the rest of vertices. Nevertheless, a node is more likely to be informed from one of its neighbours than from other vertices. In [35], the authors exemplified such effect by considering a SIS dynamics, suggesting that the error is reduced with the increase of the mean first neighbour degree. To quantify the error between our numerical experiments and MC simulations, we set the dynamical parameters to $\delta_1 = \delta_2 = \gamma = 0.1$, $\lambda = 1$ and vary α. The differences between the theoretical predictions and MC simulations are then quantified by the absolute error. We found (results not shown) that, for the RP, the error vanishes as soon as $\langle k \rangle \geq 10$, whereas for the case of a CP, the error is close to zero only for $\langle k \rangle \geq 100$, as shown in Fig. 12.

Downloaded from https://academic.oup.com/comnet/article-abstract/6/2/215/4060524
by University of Manchester user
on 26 March 2018

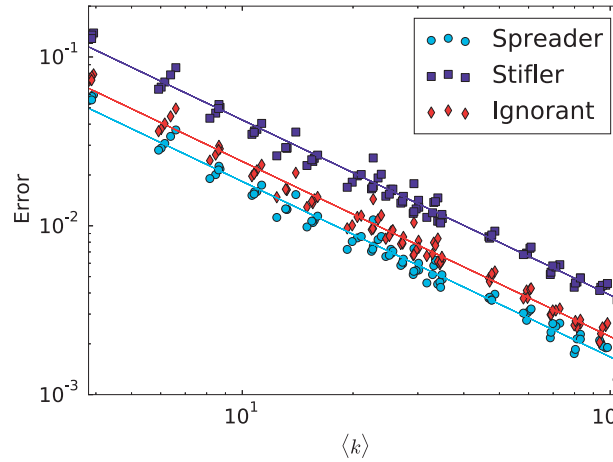

FIG. 12. Phase diagram at the steady state for the CP considering a scale-free network with $\zeta \approx 2.7$, $N = 10^4$ and the dynamical parameters $\delta_1 = \delta_2 = \gamma = 0.1$ and $\lambda = 1$. The continuous lines are obtained by the least squares method.

7. Applications to real social systems

Up to now, we have explored the model in computer-generated networks. While these networks share some of the topological features that have been found in real-world systems, they do not account for all of them, specially when it comes to clustering properties or several kinds of correlations. It is therefore important to run our model on top of real systems and check whether the results reported so far also hold for more realistic topologies. In doing so, we will also show how the model could be used to get a better understanding of the mechanisms driving real spreading phenomena. In what follows, we inspect whether the model gives accurate results at the microscopic scale for both the CP and the RP scenarios, as we have done before for the synthetic networks. Finally, we also reproduce the temporal dynamics of a real spreading process that took place over online social networks (Twitter in our analysis) when the confirmation of the existence of the Higgs boson was announced.

To verify that our model performs well also in real networks, we use two social networks. The first is an email network, in which messages are mostly spread as a contact processes (that is, one-to-one communication at a time), and the second network considered is the contact patterns of Twitter. The latter is an online social system where the information dynamics is mainly of the form one-to-all, that is, a user posts a message that reaches out all the user's followers at the same time, thus corresponding to the RP limit in our formulation. We also consider $\eta = 0.01$ to simulate the apathy as in [8].

The email contact network was created from emails exchanged between users within the Universitat Rovira i Virgili [36]. The network is composed by $N = 1133$ nodes and $\langle k \rangle = 9.6$; connections are directed and unweighted. However, here we consider an undirected version of this network. The Twitter network was extracted from the mobilizations in Spain during 2011 [33, 37, 38]. Here we consider a simplified version of this network, composed by $N = 85712$ nodes and $\langle k \rangle = 109.9$, with undirected and unweighted contacts. Both networks have an average clustering coefficient $\langle cc \rangle \approx 0.22$. In addition, while the email network is slightly assortative, $r = 0.08$, the Twitter network is disassortative $r = -0.14$.

We compare the results obtained via MC simulations of the model with the numerical solution of equation (3.6) in Fig. 13(a) and (b) using the Twitter and the email networks as underlying graphs, respectively. The results show that despite the new topological features of the real networks, the agreement

Downloaded from https://academic.oup.com/comnet/article-abstract/6/2/215/4060524
by University of Manchester user
on 26 March 2018

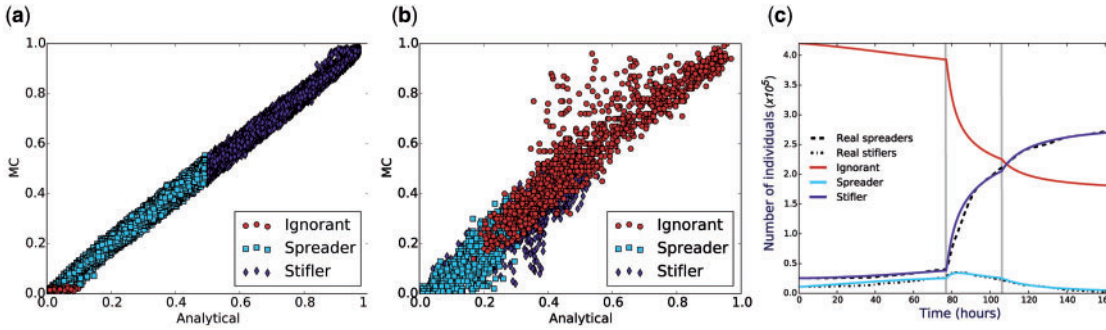

FIG. 13. Application of the proposed discrete time Markov chain formulation to the analysis of spreading processes that take place on top of social networks. Panels (a) and (b) show the probability of node i belonging to state X, Y or Z obtained by solving numerically the system of equations (3.6)) and by Monte Carlo simulations. Panel (a) corresponds to a RP that takes place over the Twitter network, where as panel (b) has been obtained simulating a CP on the email network. The model parameters: $\lambda = 1.0$, $\eta = 0.01$, $\delta_1 = \delta_2 = \gamma = 0.01$, $\beta = 1.0$ and $\alpha = 1.0$ for the RP; and $\lambda = 1.0$, $\eta = 0.01$, $\delta_1 = \delta_2 = \gamma = 0.1$, $\beta = 1.0$ and $\alpha = 1.0$. Results for MC simulations are averages over 10^3 different simulations. Panel (c) depicts the time evolution of the Higgs boson rumour spreading, taking into account tweets of the dataset shown in [32]. The dashed lines represent the real data, while the continuous lines are the numerical solutions of our model. The vertical lines mark the three time windows used, with the following parameters: (i) $0 \leq t \leq 77$, $\lambda = 0.00025$, $\alpha = 0.0002$, $\delta_2 = 0.0001$ and $\eta = 0.85$; (ii) $77 < t \leq 106$, $\lambda = 0.021$, $\alpha = 0.00075$, $\delta_2 = 0.0015$ and $\eta = 0.17$; (iii) $t > 106$, $\lambda = 0.065$, $\alpha = 0.002$, $\delta_2 = 0.002$ and $\eta = 0.01$.

is still very good. Panel (a) corresponds to a RP and the probabilities of finding a node in each of the three possible states at the large time limit match fairly well. This indicates that the new topological ingredients that were not present in the synthetic networks play a minor role when it comes to evaluate the accuracy of the discrete Markov chain formulation—or at the very least for this particular network. Panel (b) on the contrary shows the results obtained when simulating a CP on top of the email network. Again, the numerical solutions of the set of equations describing the dynamics of the system agree well with MC simulations, albeit having larger deviations and more dispersion along the diagonal line. This could be due to the fact that $\langle k \rangle = 9.6$ for the email network, and as discussed and shown earlier (see Fig. 12), for a CP errors are vanishingly small only beyond $\langle k \rangle \approx 100$. Other new topological characteristics or even some dynamical correlations [35] might have an impact as well. Additionally, note that another possible source could be the relative small size ($N = 1133$) of the email network. Admittedly, the Twitter network also has triangles, modular organization and degree–degree correlations [8, 37], but it is an order of magnitude larger.

We also apply the formalism to simulate the time evolution of a real information dissemination process. Specifically, we have modelled the temporal dynamics of the rumour spreading on Twitter during and after the announcement of the discovery of a new particle with the features of the elusive Higgs boson on 4 July 2012. Such dataset was formerly analysed in [32]. Here, we consider the giant component of an undirected network of the friend/follow network, which is composed by $N = 456626$ individuals, with $\langle k \rangle = 54.79$, an average clustering coefficient $\langle cc \rangle = 0.189$ and assortativity $r = -0.098$. This database describes the timestamps (in seconds) of mentions, replies and retweets. In our analysis, we consider these three timestamps as events. However, if two events occur at the same time and are generated by the same user, we consider them as just one event. For instance, if an user retweets a mention we would have two events at the same time, one for the mention and the other for the retweet. In such a case we consider it as just one tweet.

Downloaded from https://academic.oup.com/comnet/article-abstract/6/2/215/4060524
by University of Manchester user
on 26 March 2018

Moreover, how to assign the state a node belongs to (that is, spreader or stifler) is not trivial, because we cannot distinguish between: (i) a spreader turns into a stifler, then recovers the interest in the rumour and becomes a spreader again or (ii) a spreader that remains as a spreader during the observation time. Here, we assume that an individual is an ignorant if he/she did not tweet about the Higgs boson. An ignorant becomes a spreader the time he/she first tweets about this topic, and remains as such up to its last tweet, when it becomes a stifler. The only special case considered are the users that tweeted just once. In such a case, we consider that these individuals are spreaders during just one time step, that is, 1 unit time, and they become stiflers the next time-stamp. Additionally, the initial conditions are given by the activities before the observation time window, which implies to start with a certain fraction of spreaders and also stiflers, since some users that tweeted about the Higgs boson do not tweet during the observation time window. Note that the modelling assumes that a user is aware of the rumour only if he/she tweets about it. Finally, it is important to define Δt. For our analysis, we have assumed that each discrete time window represents 1 h of the real data. Such a choice implies that we are not able to distinguish events that occur at a faster rate. However, setting $\Delta t = 1h$ drastically reduces the computational cost. We remark that our goal here is to verify whether our model is capable of describing the behaviour observed in the real data, not to perform a detailed forensic analysis of the actual rumour spreading process.

In order to fit simulations of the model with the data, we assume that the forgetting mechanisms can be neglected, since the total observation time is reasonable small (1 week), implying that $\delta_1 = \gamma = 0$. Moreover, we assume $\beta = 0$, since we consider that there is no transition from stifler to spreader, that is, the possibility of recovering the interest in the rumour is neglected—this is in part also due to the constraint that we cannot distinguish such transitions in real data. Additionally, observe that δ_2 is the probability that a spreader loses the interest about the rumour spontaneously, while α represents the probability of turning into stiflers after contacting spreaders or stiflers. Note that η, γ and α might produce similar effects, depending on the defined time-steps.

Figure 13(c) compares the time evolution of the spreading dynamics as extracted from the real data with results obtained from our model using three different time windows. The latter is needed as it is known that in viral processes like the one we are analysing, there are different phases of the dynamics: an early stage in which the number of messages exchanged increases slowly (subcritical regime) followed by an explosive period (critical and supercritical regimes) that signals the moment beyond which the piece of news goes viral and finally a phase in which the active spreaders die out and stop propagating the rumour any farther [8]. For the first time window, we thus assign initial probabilities of being in state X, Y or Z as given by the real data, but considering them to be the same for every node. For the following time windows, these probabilities come from the output of the previous simulation window. On the first time window, $0 \leq t \leq 77$, we consider $\lambda = 0.00025$, $\alpha = 0.0002$, $\delta_2 = 0.0001$ and $\eta = 0.85$. On the second, $77 < t \leq 106$, $\lambda = 0.021$, $\alpha = 0.00075$, $\delta_2 = 0.0015$ and $\eta = 0.17$. Finally, for $t > 106$, $\lambda = 0.065$, $\alpha = 0.002$, $\delta_2 = 0.002$ and $\eta = 0.01$.

The results show that the model is indeed able to accurately reproduce the temporal evolution of the real spreading dynamics. We however stress that we chose those sets of parameters by simple inspection, that is, we did not apply any fitting algorithm. In order to obtain a better fit, one may use statistical inference tools or even a simulated annealing algorithm. As this is beyond the scope of this work, we leave this line of research as a potential future work. The simulation performed is nevertheless worth carrying out. The fact that a model like the one discussed here could be adapted without the use of sophisticated fitting algorithms to describe the temporal dynamics of a real rumour spreading process is an important step towards getting a better understanding of the mechanisms at work in this real contagion dynamics. Admittedly, without knowing the parameters, real-time projections of the temporal evolution of the rumour dynamics cannot be done. But once we know the parameters that give the best matching

Downloaded from https://academic.oup.com/comnet/article-abstract/6/2/215/4060524
by University of Manchester user
on 26 March 2018

with the real data, we are in the position to know which mechanisms (that is, what transitions) are more important and which do not, thus gaining valuable phenomenological insights. In turn, given the universal features behind spreading processes, this would allow to perform other analyses, such as detecting who are potential candidates to be influential spreaders in other online social systems or for the propagation of other rumours in the very same network.

8. Conclusions

Spreading processes play an important role in nature, society and engineering [1]. Due to their relevance, several models have been developed aiming at understanding, modelling and predicting how viruses, diseases, rumours and information propagate through complex networked systems. Despite the theoretical approaches that have been developed to model, for instance, epidemic and rumour dynamics [1], there are still important mechanisms that have not been taken into account, nor there exists a coherent and unifying theoretical and computational framework that deals with as many of these mechanisms as possible at the same time. In this paper, we provide such a unifying methodology using a discrete Markov chain approach, which includes and generalizes previous epidemic and rumour models by accounting for non-traditional behaviours, such as apathy, forgetting, and lost and recovering of interest.

We have focused our study on the theoretical and numerical analysis of the model, and have shown that results obtained by numerically solving the system of equations describing the system's dynamics are in good agreement with extensive MC simulations for three different scales: macro, micro and temporal. Regarding analytical results, we have obtained closed forms for the thresholds and the steady state densities of individuals in the different dynamical classes for the CP scenario and some special cases of the RP. Additionally, we have thoroughly analysed the influence of the model parameters as well as several network properties on the spreading dynamics. Our findings indicate that using synthetic networks could help getting a first insight into what are the effects of the different mechanisms at work with a high degree of accuracy. Finally, we have studied the propagation of rumours considering real networks on top of which both a contact process and a reactive process might take place. Importantly, we were able to reproduce the time evolution of a rumour propagation using Twitter's user activity during and after the announcement of the discovery of the Higgs boson.

Our formalism is general and covers many models on the literature. This opens new opportunities for the analysis of real data. It is also worth stressing that in information spreading on real systems like online social networks, rumour models play an important role. With this work, we have provided a framework that paves the way to developing new algorithms that could explore accurately and very fast different mechanisms and scenarios for viral information spreading. At variance with disease spreading, in which one is constrained to model a real outbreak, when it comes to design new ways to efficiently disseminate information, one is free to design the mechanisms that would optimize such a spreading process. In other words, one can design the viral process without being constrained to fit a given past or ongoing outbreak. To this end, performing MC simulations would be prohibitively costly given the size of the parameters' phase space. This practical hurdle might be surmounted by using the discrete Markov chain approach proposed here, as the computational cost of solving the set of equations is significantly low as compared to MC simulations. This is of special relevance when we are dealing with online social systems, whose sizes go from a few hundreds of individuals to millions of users. We plan to explore this line of research in the future.

Downloaded from https://academic.oup.com/comnet/article-abstract/6/2/215/4060524
by University of Manchester user
on 26 March 2018

Funding

CNPq (grant 305940/2010-4) and Fapesp (grant 2016/25682-5) to F.A.R.; Fapesp for the sponsorship provided (grants 2012/25219-2 and 2015/07463-1) to G.F.A.; CNPq (304676/2016-0) and Fapesp (15/03868-7,16/11648-0) to P.M.R.; EC FET-Proactive Project Multiplex (grant 317532) to E.C.; Government of Aragón, Spain through a grant to the group FENOL, MINECO and FEDER funds (grant FIS2014-55867-P) and by the European Commission FET-Proactive Project Multiplex (grant 317532) to Y. M.

References

1. Pastor-Satorras, R., Castellano, C., Van Mieghem, P. & Vespignani, A. (2015) Epidemic processes in complex networks. *Rev. Mod. Phys.*, **87**, 925–979.
2. Gleeson, J. P., O'Sullivan, K. P., Baños, R. A. & Moreno, Y. (2016) Effects of Network Structure, Competition and Memory Time on Social Spreading Phenomena. *Phys. Rev. X*, **6**, 021019.
3. Keeling, M. J. & Eames, K. T. (2005) Networks and epidemic models. *J. Roy. Soc. Interface*, **2**, 295–307.
4. Boccaletti, S., Latora, V., Moreno, Y., Chavez, M. & Hwang, D. (2006) Complex networks: structure and dynamics. *Phys. Rep.*, **424**, 175–308.
5. Barabási, A.-L. & Albert, R. (1999) Emergence of scaling in random networks. *Science*, **286**, 509–512.
6. Pastor-Satorras, R. & Vespignani, A. (2001) Epidemic Spreading in scale-free networks. *Phys. Rev. Lett.*, **86**, 3200–3203.
7. Gómez, S., Arenas, a., Borge-Holthoefer, J., Meloni, S. & Moreno, Y. (2010) Discrete-time Markov chain approach to contact-based disease spreading in complex networks. *EPL (Europhysics Letters)*, **89**, 38009.
8. Borge-Holthoefer, J., Meloni, S., Gonçalves, B. & Moreno, Y. (2012) Emergence of Influential Spreaders in Modified Rumor Models. *J. Stat. Phys.*, **151**, 383–393.
9. Kawachi, K., Seki, M., Yoshida, H., Otake, Y., Warashina, K. & Ueda, H. (2008) A rumor transmission model with various contact interactions. *J. Theoret. Biol.*, **253**, 55–60.
10. Nekovee, M., Moreno, Y., Bianconi, G. & Marsili, M. (2007) Theory of rumour spreading in complex social networks. *Physica A Stat. Mech. Appl.*, **374**, 457–470.
11. Gomez, S., Arenas, A., Borge-Holthoefer, J., Meloni, S. & Moreno, Y. (2011) Probabilistic framework for epidemic spreading in complex networks. *Int. J. Complex Syst. Sci.*, **1**, 47–54.
12. Daley, D. J. & Kendall, D. G. (1964) Epidemics and Rumours. *Nature*, **204**, 1118.
13. Daley, D. J. & Kendall, D. G. (1965) Stochastic rumours. *J. Inst. Math. Appl.*, **1**, 42–55.
14. Barrat, A., Barthlemy, M. & Vespignani, A. (2008) *Dynamical Processes on Complex Networks*. New York, NY: Cambridge University Press.
15. Maki, D. P. & Thompson, M. (1973) *Mathematical Models and Applications*. Englewood Cliffs, NJ: Prentice-Hall Inc..
16. Moreno, Y., Nekovee, M. & Pacheco, A. F. (2004) Dynamics of rumor spreading in complex networks. *Phys. Rev. E*, **69**, 066130.
17. Newman, M. (2010) *Networks: An Introduction*. Oxford University Press, Inc.
18. Boguñá, M. & Pastor-Satorras, R. (2002) Epidemic spreading in correlated complex networks. *Phys. Rev. E*, **66**, 047104.
19. Moreno, Y., Pastor-Satorras, R. & Vespignani, A. (2002) Epidemic outbreaks in complex heterogeneous networks. *Eur. Phys. J. B*, **26**, 521–529.
20. Boguñá, M., Pastor-Satorras, R. & Vespignani, A. (2003) Absence of epidemic threshold in scale-free networks with degree correlations. *Phys. Rev. Lett.*, **90**, 028701.
21. Moreno, Y., Gómez, J. B. & Pacheco, A. F. (2003) Epidemic incidence in correlated complex networks. *Phys. Rev. E*, **68**, 035103.
22. Gómez-Gardeñes, J., Latora, V., Moreno, Y. & Profumo, E. (2008) Spreading of sexually transmitted diseases in heterosexual populations. *Proc. Nat. Acad. Sci.*, **105**, 1399–1404.

Downloaded from https://academic.oup.com/comnet/article-abstract/6/2/215/4060524
by University of Manchester user
on 26 March 2018

23. Mieghem, P. V., Omic, J. & Kooij, R. (2009) Virus spread in networks. *IEEE/ACM Trans. Netw.*, **17**, 1–14.
24. Boguñá, M., Castellano, C. & Pastor-Satorras, R. (2013) Nature of the epidemic threshold for the susceptible-infected-susceptible dynamics in networks. *Phys. Rev. Lett.*, **111**, 068701.
25. Li, C.-H., Tsai, C.-C. & Yang, S.-Y. (2014) Analysis of epidemic spreading of an SIRS model in complex heterogeneous networks. *Commun. Nonlinear Sci. Numer. Simul.*, **19**, 1042–1054.
26. Chen, L. & Sun, J. (2014) Global stability and optimal control of an {SIRS} epidemic model on heterogeneous networks. *Phys A Stat. Mech. Appl.*, **410**, 196–204.
27. Lebensztayn, E., Machado, F. & Rodríguez, P. (2011) Limit Theorems for a General Stochastic Rumour Model. *SIAM J. Appl. Math.*, **71**, 1476–1486.
28. Eden, M. (1961) A Two-dimensional growth process. *Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability*, vol. 4. Contributions to Biology and Problems of Medicine. Berkeley, CA: University of California Press, pp. 223–239.
29. Chung, F. R. K., Lu, L. & Vu, V. H. (2003) The spectra of random graphs with given expected degrees. *Proc. Nat. Acad. Sci.*, **100**, 6313–6318.
30. Stanoev, A., Daniel, T. & Ljupco, K. (2014) Modeling the spread of multiple concurrent contagions on networks. *PLoS ONE*, **9**, e95669.
31. Mieghem, P. V. (2011) *Graph Spectra for Complex Networks*. New York, NY: Cambridge University Press.
32. De Domenico M., Lima A., Mougel P. & Musolesi M. (2013) The anatomy of a scientific rumor. *Sci. Rep.*, **3**.
33. González-Bailón S., Borge-Holthoefer J., Rivero A. & Moreno Y. (2011) The dynamics of protest recruitment through an online network. *Sci. Rep.*, **1**.
34. Arruda, G. F., Barbieri, A. L., Rodríguez, P. M., Rodrigues, F. A., Moreno, Y. & da Fontoura Costa, L. (2014) Role of centrality for the identification of influential spreaders in complex networks. *Phys. Rev. E*, **90**, 032812.
35. Gleeson, J. P., Melnik, S., Ward, J. A., Porter, M. A. & Mucha, P. J. (2012) Accuracy of mean-field theory for dynamics on real-world networks. *Phys. Rev. E*, **85**, 026106.
36. Guimera, R., Danon, L., Díaz-Guilera, A., Giralt, F. & Arenas, A. (2003) Self-similar community structure in a network of human interactions. *Phys. Rev. E*, **68**, 065103.
37. Borge-Holthoefer, J., Rivero, A., García, I., Cauhé, E., Ferrer, A., Ferrer, D., Francos, D., Iñiguez, D., Pérez, M., Ruiz, G., Sanz, F., Serrano, F., Viñas, C., Tarancón, A. & Moreno, Y. (2011) Structural and dynamical patterns on online social networks: the Spanish May 15th movement as a case study. *PloS one*, **6**, e23883.
38. Borge-Holthoefer, J., Rivero, A. & Moreno, Y. (2012) Locating privileged spreaders on an online social network. *Phys. Rev. E*, **85**, 066123.

Downloaded from https://academic.oup.com/comnet/article-abstract/6/2/215/4060524
by University of Manchester user
on 26 March 2018