



Dynamics of information diffusion and its applications on complex networks



Zi-Ke Zhang^{a,b}, Chuang Liu^a, Xiu-Xiu Zhan^{a,c}, Xin Lu^{d,e}, Chu-Xu Zhang^f,
Yi-Cheng Zhang^{g,a,*}

^a Alibaba Research Center for Complexity Sciences, Hangzhou Normal University, Hangzhou 311121, PR China

^b Institute of Automation, Shanghai Jiaotong University, Shanghai 200030, PR China

^c Faculty of Electrical Engineering, Mathematics and Computer Science, Delft University of Technology, 2628 CD Delft, The Netherlands

^d College of Information System and Management, National University of Defense Technology, Changsha 410073, Hunan, PR China

^e Department of Public Health Sciences, Karolinska Institutet, Stockholm 17177, Sweden

^f Department of Computer Science, Rutgers University, NJ, 08854-8019, USA

^g Department of Physics, University of Fribourg, Fribourg, CH-1700, Switzerland

ARTICLE INFO

Article history:

Available online 19 July 2016

editor: I. Procaccia

Keywords:

Complex networks

Information diffusion

Network dynamics

ABSTRACT

The ongoing rapid expansion of the Word Wide Web (WWW) greatly increases the information of effective transmission from heterogeneous individuals to various systems. Extensive research for information diffusion is introduced by a broad range of communities including social and computer scientists, physicists, and interdisciplinary researchers. Despite substantial theoretical and empirical studies, unification and comparison of different theories and approaches are lacking, which impedes further advances. In this article, we review recent developments in information diffusion and discuss the major challenges. We compare and evaluate available models and algorithms to respectively investigate their physical roles and optimization designs. Potential impacts and future directions are discussed. We emphasize that information diffusion has great scientific depth and combines diverse research fields which makes it interesting for physicists as well as interdisciplinary researchers.

© 2016 Elsevier B.V. All rights reserved.

Contents

1.	Introduction.....	2
2.	Information diffusion in real applications.....	3
2.1.	Networking: Platforms and applications.....	3
2.1.1.	Instant messengers.....	3
2.1.2.	Blogging social networking applications.....	3
2.1.3.	Micro-blogging social networking applications.....	4
2.1.4.	Email networks.....	4
2.1.5.	Mobile communication.....	4
2.1.6.	Others applications.....	5
2.2.	Taxonomy of information spreading phenomenon.....	5

* Corresponding author at: Department of Physics, University of Fribourg, Fribourg, CH-1700, Switzerland.

E-mail addresses: zkz@hznu.edu.cn (Z.-K. Zhang), yi-cheng.zhang@unifr.ch (Y.-C. Zhang).

2.2.1.	Broadcasting.....	5
2.2.2.	Information sharing.....	5
2.2.3.	Crowdsourcing and collaboration	6
2.2.4.	Viral marketing.....	6
2.2.5.	Opinion formation	6
2.2.6.	Innovation diffusion	7
2.2.7.	Malicious spreading.....	7
3.	Empirical studies.....	7
3.1.	Spreading patterns.....	7
3.1.1.	Temporal dynamical patterns.....	7
3.1.2.	Diffusion structure pattern	9
3.1.3.	Human dynamical pattern	11
3.2.	Cascading effects.....	12
3.2.1.	Information cascades.....	12
3.2.2.	Cascade prediction.....	13
4.	Dynamical models.....	14
4.1.	Threshold models	14
4.1.1.	Linear threshold model	14
4.1.2.	Generalized threshold models.....	14
4.2.	Cascading models	15
4.2.1.	Independent cascade model	15
4.2.2.	Generalized cascade models.....	16
4.3.	Epidemic models.....	17
4.3.1.	Classical epidemic models	17
4.3.2.	Mathematical analysis for epidemic models	18
4.4.	Other models.....	20
5.	Applications studies.....	22
5.1.	Network reconstruction	22
5.1.1.	Compressed sensing.....	23
5.1.2.	CST-based reconstruction	23
5.2.	Rumor spreading.....	24
5.3.	Leader identification.....	25
5.4.	Dissemination management	25
5.4.1.	Epidemic threshold.....	26
5.4.2.	Node-level strategy	26
5.4.3.	Edge-level strategy	27
6.	Outlook	27
	Acknowledgments	28
	References.....	28

1. Introduction

Recently, tremendous achievements of the empirical analysis on information diffusion (also known as information spreading or transmission) have been obtained, for the fast development of the database and computational technology. Temporal motif analysis on the mobile phone communication system reveals the homophily feature in the social interaction, where communications tend to be more frequent between individuals with similar traits, such as sex and ages [1]. And human activity patterns, such as the large heterogeneity in response time [2], weight-topology correlations and the burst activity patterns on individuals [3], would slow down information spreading significantly. Apart from the mobile phone communication systems, the online social networks are the most important information spreading platform. On the online social system, users can access information through peer-to-peer interaction as well as the external (out-of-network) sources [4], and different kinds of users, such as media, bloggers, common individuals and so on, play significantly different roles in information spreading [5]. Contrary to the traditional realizations, most information items are just approved by very tiny fraction of population, while the number of the observed large cascades is very small [6]. For the case of large cascades, the number of the new approved individuals always follows power-law relaxation after the burst [7,8]. All these empirical observations can be used to suggest new directions for modeling and application of the information spreading process.

Modeling the information (fads, innovations, collective actions, viral memes, opinions and rumors) diffusion process is of outstanding interest for analyzing the information spreading patterns on social networks, stopping the transmission of viruses, as well as controlling the spread of rumors, etc. A variety of models have been proposed to characterize this process, in which the most classical models are the independent cascade model (ICM) where the information flows over the network through cascade [9,10] and the threshold model (including the linear threshold model and general threshold models) established based on the assumption that the neighbors play significant roles for the diffusion process [11,12]. Besides these models, perhaps the most commonly used models are epidemic spreading models, such as the SIS model and SIR model, as a piece of information can be transmitted from one individual to another which is the similar pattern as the

epidemic spreading [13]. Obviously, the models considered above all focus on static networks, which is called dynamics on networks. However, in many applications, the links of the networks can also evolve with time, which is treated as dynamics of the networks [14]. Consequently, the adaptive network appeared in order to combine these two fields. Most models concerning the adaptive process in information diffusion are used to characterizing the dynamic of opinion formation, such as the voter-like approach models and models concerning social segregation [15,16].

Thanks to the ever-increasing data collection of information spreading, applications of information spreading studies, ranging from network construction to opinion leader identification, benefit a lot for our human beings. Recent carefully designed work [17] based on compress sensing theory developed a general framework for reconstructing networks from time series propagation data. The uncovering complex structure provides the fundament to understand and manage collective dynamics, like minimizing the rumor propagation in the online virtual communities or controlling the outbreaks of global scale virus, which is the same target of the effort in locating the opinion sources or leaders [18,19]. Despite the data-driven approach without the network prior knowledge in understanding and controlling collective dynamics, many studies provide effective and efficient strategies in managing propagations via changing the complex structure based on node or edge level. The surprising finding is [20] that the largest (in module) eigenvalue of the network adjacency matrix is the only network parameter that determines the tipping point of the propagation process, inspires the various algorithms including node removal [21], node immunization [22], edge rewiring [23], etc., for minimizing/maximizing the information propagation. Many extensions [24,25] were designed and studied, which provide plenty desired choices to manage information spreading in real applications.

2. Information diffusion in real applications

2.1. Networking: Platforms and applications

As a fundamental need of human daily activities, information spreading is the goal, yet an emerging phenomenon from the complex of human communications [26]. In fact, facilitating information spreading is the core motivation of developing modern technique systems, such as instant messengers, Twitter, blogs, Facebook, telecommunication systems, etc. The vast development of online social networking platforms has, to the most extent, expanded and facilitated information spreading. Such expansion and facilitation have not only speeded up information spreading, but also changed the fundamental mechanisms underlying the emerging complex spreading patterns. In this section, we aim to summarize main streams of information spreading systems, which we categorize (noninclusive) as instant messengers, blogging and micro-blogging social networking applications, email networks, mobile communication systems, etc.

2.1.1. Instant messengers

Instant messaging (IM) is a type of online chat which offers real-time transmission over the internet. Typically, messages are sent, in the form of text, hyperlinks, picture, voice over IP, and video etc., between users who have added their counterparts on their contact list ("friends list"). Most IM apps also allow messages to be sent between strangers who are not on the list but have agreed on an authentication to allow receiving messages from specific users.

The development of IM comes along with the expanding of the internet. Early tools include peer-to-peer IMs such as talk, ntalk and ytalk, and client-server IMs such as talker and IRC. While many of these IMs have updated to support conversations between users on different machines later on, a list of other internet-wide, GUI-based (graphical user interface) IM clients were populated till 2000s, these include ICQ, AOL, PowWow, as well as other software type clients developed by different companies, such as Excite, MSN, Ubuque and Yahoo!. In addition to supporting inter-person communication with various services (e.g., voice over IP, video calling, web conferencing, file transferring, desktop sharing etc.), recent development of IM services also incorporates the function of social networking, such as allowing users to post web-based updates and suggesting new contacts through recommender algorithms. These include Facebook Chat, Google talks, QQ, WeChat, etc.

Even the typical communication with IMs is person to person, many IMs nowadays offer multicast transmission, e.g., "chat rooms" or "chat groups", where participants might be anonymous or might be previously known to each other (for example collaborators on a project that is using chat to facilitate communication). Obviously, it is more efficient to broadcast information with a group of people rather than with a one-to-one messaging.

2.1.2. Blogging social networking applications

Early versions of IMs are mostly stand-alone applications; the ability of social networking is limited that, information, in forms of various contents, can only be shared within limited group, such as one-to-one chat window, group-chat rooms, etc. Along with the development of online social media, web-based social networking services (SNS) have been proposed to extend the diversity and broadness of information sharing. In such platforms, e.g., Facebook, users can post text, images, video, hyperlinks, and the alike to their profile page such that the content can be visible to either a group of defined people (private), or to everyone (public). Besides posting and sharing interesting contents, most of such systems nowadays also provide many additional services, such as notes, chat, virtual gifts, marketplace, messaging, voice calls, video calling, encrypted services, etc.

Compared with traditional IM tools, which are relatively private, SNS allow users to easily keep in touch with friends and family by sending messages and posting status update, the later will be seen by lots of other people. The enhanced features of communication and networking have stimulated a vast increase in the use of SNS in the past decade, for example, Facebook was public accessible in 2006, two years later, it had overtaken Myspace and become the largest online social network in the world at the time with 500 million users. By the end of 2015, the number of monthly active users reached 1.59 billion. Similar SNS platforms have also been developed around the world, for example, the SNS version of the Chinese IM QQ, Qzone, enables users to upload content to their profile page and keep interaction with other users online. Due to the popularity of QQ in China, Qzone has 653 million monthly active users by the late of 2015.

2.1.3. Micro-blogging social networking applications

The aforementioned Facebook-like SNS platforms allow users to interact closely with friends and families who are on their friends list. The Twitter-like systems, however, make no assumptions about mutual friendship and extend the scope of interaction by allowing each user to “follow” people who are far-beyond their personal social networks, such that the relationships in twitter can be irreciprocal, as celebrities or accounts created by key agencies may attract much higher attention than ordinary users. For example, in 2016, the singer and actress Katy Perry on Twitter has more than 84 million followers, and president Barack Obama has more than 70 million followers.

Another distinct feature of Twitter-like systems is that the post is limited within a short number of characters. For example, the length limit of Twitter is 140 characters. In other words, Twitter-like systems encourage brevity, though pictures and urls can also be uploaded. In contrast, Facebook posts can run from a single line to about 63,000 characters, and the upper limit of Google+ is 100,000 characters. We term such platforms as micro-blogging SNS, i.e., it refers to frequent but very short posts, unlike regular blogging-writing long posts with photos, micro-blogging is meant to be quick, succinct, and pointed.

The increasing use of smart phones have shifted most web-based SNS to fast and easy applications on mobile devices [27]. Due to the simplicity and brevity, micro-blogging SNS systems grow fast on mobile devices, these including Twitter, foursquare, tumblr, friendfeed, plurk, Sina Weibo, etc. It is worth noting that, while initially designed to limit the size of the text that can be posted, many micro-blogging services are now offering posting with higher or no limit. On the other side, blogging SNS sites, such as Facebook, is now offering the option to “follow” users who is not on one's friends list, making the boundaries of blogging and micro-blogging more indistinguishable.

2.1.4. Email networks

Unlike instant messaging or blogging applications which aim to the speediness of communication and status updates, email, provides a technique for transferring of digital messages that with no push of immediacy. Though early email systems required the author and recipient to both be online at the same time, today's email systems are based on a “store-and-forward” model: email servers accept, forward, deliver, and store messages. The users only need to be connected to the mail server whenever they need to receive or send messages.

The content of each email message is formed of two major sections: the message header which is structured into fields, including email address of the author (*from*), local time when the email is written (*Date*), email address of recipients (*to*, *cc*), etc., and the message body which is written by the author and can consist either plain text or content of various format and type. Therefore, through the email header fields, it is possible to forward a piece of information through the chain of emailing links. Most companies and organizations have their own email server, while there is also a large selection of open email service providers, such as Gmail, Hotmail, Yahoo! Mail, etc. To today, there are an estimated 2.7 billion email users worldwide and 215 billion emails that are sent daily.

2.1.5. Mobile communication

Mobile communication, or mobile telephony, refers to the communication of mobile phone users through the cellular network of base transceiver stations (BTS or towers). Each mobile phone may dial or receive a call or text message (short message service, SMS) if there is network coverage, for which the typical range is from a few hundred meters in densely populated cities to a few dozens of kilometers in remote area. With smartphones, many internet-based services have also shifted to mobile devices, including IMs, SNS, email, and so on. As the most widely used technology and most common electronic device in the world, it was estimated that the unique number of mobile phone subscribers would increase from 4.7 billion in 2015 to 5.6 billion by 2020, accounting to a penetration rate of 63% and 72% of the global population, respectively [27–29].

Through mobile communication, each individual with a phone is readily connectable. In terms of spreading, mobile phones form a fundamental network for information flow of various types: dissemination of warning messages [30], broadcasting of advertisements [31], exchange of opinions [32], etc. Particularly, either cellular towers or GPS allows locating of the device, which provide rich contextual information on the environment and facilitate many location based services [33], such as advertising [34], recommendation [35], location-based social networking [34], navigation [36], etc.

Table 1

List of popular social networking applications in the world.

Name	Registered users	Active users	Date launched	Country of origin	Date of user stat.
Facebook	2+ billion	1.591 billion	Feb-04	UnitedStates	Jan-16
WhatsApp	1+ billion	1 billion	Jun-11	UnitedStates	Feb-16
TencentQQ	1+ billion	853 million	Feb-99	China	Dec-15
FacebookMessenger	2+ billion	800 million	Aug-11	UnitedStates	Jun-15
TencentQzone	1+ billion	653 million	May-05	China	Sep-15
WeChat	1+ billion	650 million	Jan-11	China	Dec-15
Google+	2.2 billion	540 million	Jun-11	UnitedStates	Jan-15
Instagram	400+ million	400 million	Oct-10	UnitedStates	Sep-15
Twitter	1+ billion	305 million	Mar-06	UnitedStates	Dec-15
Skype	663+ million	300 million	Aug-03	Estonia	Mar-14
BaiduTieba	1 billion	300 million	Dec-03	China	Jul-14
Viber	606 million	249 million	Dec-10	Israel	Jun-15
SinaWeibo	800+ million	222 million	Aug-09	China	Dec-15
LINE	600 million	215 million	Jun-11	Japan	Dec-15
YY	773 million	122 million	Dec-10	China	Jun-15
Snapchat	100+ million	100+ million	Sep-11	UnitedStates	May-15
BBM	190 million	100 million	Feb-07	Canada	Feb-15
Pinterest	100+ million	100 million	Mar-10	UnitedStates	Sep-15
LinkedIn	400+ million	100 million	May-03	UnitedStates	Apr-16
Telegram	100+ million	100 million	Aug-13	Germany	Feb-16
VK.COM	260 million	60 million	Sep-06	Russia	Jul-14
Myspace	1 billion	50+ million	Jan-04	America	Jan-2015
Kakao Talk	160 million	48 million	Mar-10	Korea	Sep-15
Renren	137 million	37million	Dec-05	China	Nov-13
Yixin	100+ million	30+ million	Aug-13	China	Jul-14
Odnoklassniki	65 million	25+ million	Mar-06	Russia	Jul-13
Momo	180 million	7.8 million	Aug-11	China	Oct-15
Kik Messenger	275 million	2.5 million	2009	Canada	Apr-16

2.1.6. Others applications

Beyond the platforms and applications mentioned above, there are a large number of other systems that embed rich information spreading phenomenon. For example, it is quite popular nowadays to use content-specific applications on smartphones for sharing of specific type of information, such as Instagram, Flickr for photos, YouTube, Vine for videos, Foursquare, Gowalla for local search and recommendations, etc.

BBS (bulletin board system) is an old fashion way for information broadcasting and sharing. The user can read news and bulletins, and exchange messages with other users through email, public message boards, and sometimes via direct chatting. It was one of the most popular internet service in the 1990s, and is still popular in many countries to today, for example, the Chinese Tianya Club, which offers BBS, blog, microblog and photo album services, has more than 100 million registered users and a daily unique visitors of 12 million.

It is worth noting that many of the above mentioned applications have integrated with smartphones and there is also a tendency that each communication tool is trying to provide additional features that others have. For example, being initially a pure photo-sharing tool, Instagram is now offering the functions of video sharing as well. We have collected a list of popular social networking applications in the world, see [Table 1](#).

2.2. Taxonomy of information spreading phenomenon

2.2.1. Broadcasting

Broadcasting is the operation of sending information from the source to as many targeted audience as possible. Online social networking services have been revolutionized the way for information broadcasting. Other than traditional one-way radio or television, broadcasters are using social media to innovatively enhance user engagement, traditional marketing campaigns and promotions, warning message distribution, etc. Through online social networks of the audience, the target information can be relayed, reinforced, and dissimilated rapidly. For example, during the 2012 Indonesia Earthquake, for Twitter followers of the central disaster warning agency (BMKG), even with less than 0.1% of all followers re-tweeting the initial reports, tsunami early warnings released by BMKG could potentially reach 4 million users in 15 min [[37,38](#)].

2.2.2. Information sharing

Information sharing is a core phenomenon on complex networks: the distinctive feature of modern online social media is its de-centralized structure such that each participant can generate, post and re-post others' content which they think of certain value (entertaining, learning, business, etc. [[39,40](#)]), such as the re-tweet behavior in Twitter. Unlike broadcasting, for which the purpose is to send the information to as many targeted audience as possible, in information sharing, the initial

motivation of composing the content may quite differ, however, as many users can forward and re-post one single piece of content. It can diffuse into the social network far beyond the author's personal network.

In a survey from Facebook, the authors found that the strongest motivator for providing more information for sharing is “having fun” and “revealing enough information so that necessary/useful to me and other people to benefit from Facebook” [40]. In [41], from an analysis of 41.7 million users, 1.47 billion social relations, and 106 million tweets, the authors found that the most shared trending topics (over 85%) are headline news or persistent news in nature, and they found that any retweeted tweets reached an average of 1000 users, regardless of the number of followers from the original tweet. Once retweeted, a tweet gets retweeted almost instantly on next hops, signifying fast diffusion of information after the first retweet.

2.2.3. Crowdsourcing and collaboration

Crowdsourcing is the process of obtaining needed services, ideas or content by soliciting contributions from a large group of people, and especially from an online community. As an innovative form of online collaboration, crowdsourcing has been applied in a wide range of fields: knowledge discovery and management, distributed human intelligence tasking, broadcast search, and peer-vetted creative production, etc [42,43].

One of the well-known applications of crowdsourcing is the 2009 DARPA Network Challenge, in which the competition teams had to locate ten red balloons, lofted 30.5 m into the air at locations scattered throughout the United States, and then report their findings to DARPA. The winning team from the Massachusetts Institute of Technology (MIT) was able to within nine hours identify the correct latitude and longitude of all ten balloons, with the help of Facebook, Twitter and a homemade Web site for rapidly gathering and disseminating very precise information [44]. In astronomy, NASA's photo organizing project asks internet users to browse photos taken from space and try to identify the location the picture is documenting. Other examples of crowdsourcing projects include crisis mapping [45], crowdfunding [46], etc.

2.2.4. Viral marketing

While consumers are more resistant to traditional forms of advertising, such as TV and newspaper ads, viral marketing becomes an effective alternative marketing strategy by exploiting existing social networks and encouraging customers to share product information with their friends [47–49]. It comes in many forms, from videos to games, images, blogs, text messages, and emails, as long as the message can be sent to another person. For example, you can post a video that includes a link to your website and a page where a message can be sent to other people.

Hotmail was one of the many cases that became extremely successful utilizing viral marketing. When it first began, they decided that the best way to entice new customers would be to reach the friends, family and colleagues of each user they had. So, in the footer of each and every email that was sent from a Hotmail account, the Hotmail team placed a link that read something like this: “Want a free email account? Sign-up for Hotmail today!” The strategy went extremely successful and hundreds of thousands of users signed up for their free email accounts. In fact, it recruited 12 million users in 18 months, a historically fast growth at that time [50]. Due to the advantages of easy to implement, social networking, low-cost, potential of exponential growth, and long-term effect, viral marketing has become a common business campaign nowadays. From Facebook, to Twitter, to YouTube and LinkedIn, it has shown the power of expanding business and boosting revenue. Other well-known cases include the Burger King's Subservient Chicken campaign [51], ALS Ice Bucket Challenge [52], Old Spice's “The Man Your Man Could Smell Like” [53] and Coca-cola's “Friendly Twist” campaign [54], etc.

2.2.5. Opinion formation

Opinion formation is an outcome of information spreading on social networks. Through exposure and exchange of individual views, attitudes, and beliefs about a particular topic, either a single individual would be influenced and form a personal opinion, or the population as a whole would exhibit a synthesis view of the topic, i.e., public opinion [55]. With billions of users, the social network has become an important and essential information channel which can spread very rapidly and have an extremely broad coverage. Social media and social networking sites are increasingly used by people to express their opinions, on breaking news, political issues, sports events, and new products, etc.

There has been an increasing interest on leveraging social media and social networking sites to sense and predict opinions as well as understand opinion dynamics. For example, Twitter has been used for broadcasting political messages, interacting with voters, and for predicting election [56,57]; quantitative investment firms measure and trade investor sentiment using social media [58–60]; and, large corporations increasingly leverage brand sentiment, estimated from users' posts, likes and shares in social media and social networking sites, to design their marketing campaigns [61]. In a study of two years' twitter corpus on 1 billion messages, it is found that the sentiment word frequencies in Twitter can correlate with surveys on consumer confidence and political to as high as 80% [62].

It is worth noting that the opinion formation is closely related to the human behavior contagion process on social networks, which has been observed and studied for a variety of aspects in the past years [63–65]. For example, in an experiment designed by Damon Centola, it is found that individual adoption of health behavior increases when participants received social reinforcement from multiple neighbors in the network, and the behavior spread farther and faster across clustered-lattice networks than across corresponding random networks [66].

2.2.6. Innovation diffusion

Similar to opinion formation, new ideas and ways of doing things, or adoption of new products can spread gradually through social networks, as people tend to adopt an innovation with increasing likelihood, depending on the proportion of their friends and neighbors who have adopted it [63,67]. In a classic study, Coleman et al. [68] showed how doctors' willingness to prescribe the new antibiotic tetracycline diffused through professional contacts. Other early discoveries of innovation diffusion on social networks include family planning methods, new agricultural practices, and a variety of other innovations [69].

The rapid growth and popularity of online social networking services provide new means for communication and interaction, and consequently make the Internet a unique environment for the emergence and spread of innovations. As physical proximity is no longer a constraint for interactions, the underlying mechanisms, speed and extent of spreading have changed substantially [67,70,71]. For example, for new technologies to reach 50 million users, it took 50 years for electricity, while MySpace used only one year. Online social media also enable researchers and the industry to quantify factors that may be crucial to innovation diffusion besides characters of the innovation itself, such as opinion leaders, network local clustering [64,67,72,73].

2.2.7. Malicious spreading

Malicious spreading refers to spreading of information that is harmful to the community, including the spread of malicious programs (computer viruses, worms, trojan horses, spyware, etc.), malvertising, rumors and gossips [13,74–76]. Through web services of the connected internet, such information can spread to millions of agents within a very short time. For example, the very early and notorious computer worm "ILOVEYOU", spreading as an email message with the subject line "ILOVEYOU" and the attachment "LOVE-LETTER-FOR-YOU.txt.vbs", infected fifty million computers within ten days. The other notorious one, *Storm Worm*, which is a backdoor Trojan horse also using e-mail messages to spread, had infected 1 to 10 million computers in 2007 and accounted for 8% of all malware infections globally three days after the infection began [77].

When it comes to rumor propagation, the mass connected online social networks today offer the potential of spreading with unimaginable speed. According to Benjamin et al., a rumor began at a random node of the Twitter network reaches on average 45.6 million of the total of 51.2 million members within only eight rounds of communication [78]. As rumors or misinformation may cause serious consequences in certain circumstance, e.g., in a disastrous situation [79,80], it is of great importance to detect, prevent and understand the mechanisms of rumor spreading.

3. Empirical studies

3.1. Spreading patterns

Information spreading process presents abundant and various patterns in real social systems because of the complicated communication mechanism coupling with the diversiform social interaction dynamics. Recently, for the fast development of the database and computational technology, multifarious spreading patterns based on the empirical analysis on information spreading have been obtained. The observation of the spreading patterns would help us to understand the dynamics of information diffusion more clearly. In this section, we review the information spreading patterns introduced in the literatures, which are essentially obtained from the online social systems and the mobile communication systems.

3.1.1. Temporal dynamical patterns

Based on the collective human effects [7], some information about the popular topic can quickly spread out across the social network through the peer-to-peer interaction as well as mainstream media. Understanding the temporal dynamic pattern for these popular information has been proved difficult, since there seems to be no significant shifts to describe the continuous appearance, growth and decay process [81] (as Fig. 1), which is also referred to as *the rise and fall pattern* in Ref. [82]. However, it is very important to clearly describe the temporal dynamical patterns, that people can reliably predict the overall dynamics of spreading process based on observing only a small number of spreaders within the initial steps [83, 84], with the direct application including rumor spreading [78], public opinion monitor, the popularity of newly emerging events and even prediction of the product information based on the collective attention [85]. The temporal dynamics were first addressed by Crane and Sornette [7], who studied the dynamical process of the data extracted from the daily views on YouTube. In their study, the temporal pattern can be described accurately as a self-excited Hawkes conditional Poisson process, and the functional forms of growth and decay process can be approximately considered as power-law rise and fall pattern, and similar patterns are also observed on the blog systems where the exponent values lie between -0.1 and -2.5 [86]. Another method to study the temporal pattern on information spreading is based on clustering methods. Yang and Leskovec [83] obtained six classes of temporal patterns using the K-Spectral Centroid clustering algorithm on two massive datasets, Twitter and blog posts and news media articles. In addition, Lehmann et al. [87] presented four groups of temporal patterns, which are activity concentrated before and during peak, during and after peak, symmetrically around the peak, and on the single day of the peak respectively, with coarse-graining the temporal dynamics of the hashtags on Twitter. Furthermore, differential stochastic equations were also a widely used method to describe the temporal patterns. Matsubara

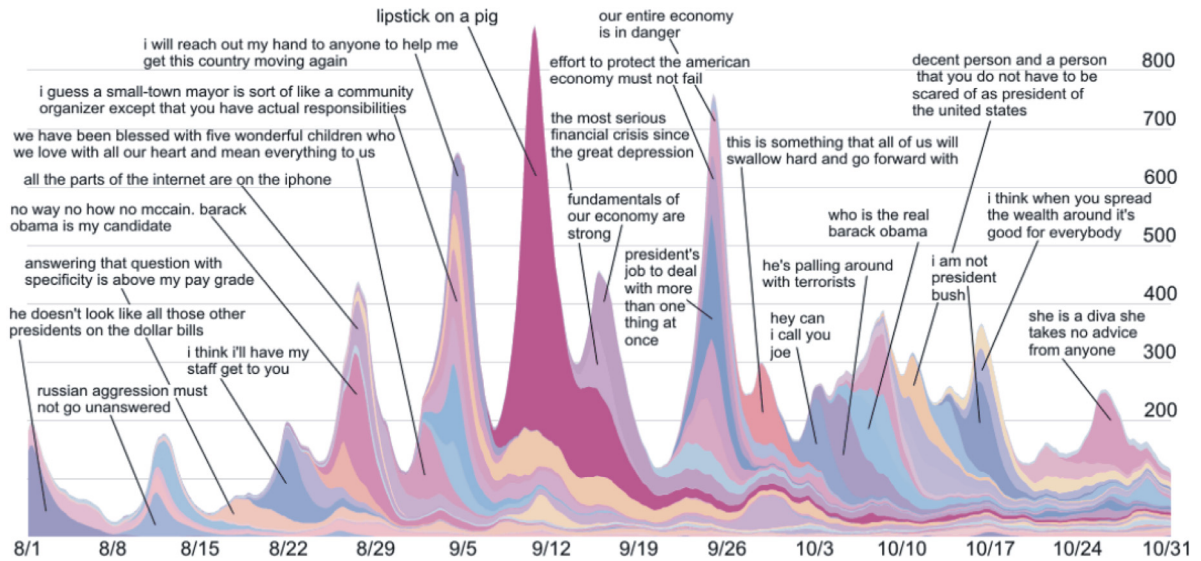


Fig. 1. Temporal dynamical patterns of 50 largest threads from mainstream online medias for the period Aug. 1–Oct. 31, 2008. The thickness of each strand corresponds to the volume of the corresponding thread over time. And the various dynamic patterns could be observed.
Source: From Ref. [81].

et al. [82] presented a SPIKEM model, which can describe all the patterns of the rise-and-fall burst in the data collected from the public domain. The dynamic mechanism of the temporal pattern seemed to be explicitly described with the differential stochastic equations, however too much assumption and simplification are considered in this kind of methods [88–90].

The temporal dynamical patterns of the information spreading are still unclear, leading to the critical resistance for the information cascading prediction [91,92]. And the diversiform temporal patterns would be caused by the complicated social systems, which could be generally classified into two domains [88]. The first factor is the *information content*, and information with different contents may have very different audiences and spreading paths, leading to different temporal patterns [93]. And some information would be very popular within very short time, such as the news of the gravitational waves detection very recently, while most messages are just adopted by very few people, such as some *grassroot* voices. The intrinsic infectiousness of different information contents (a piece of message, tweets or some others) are very different, and infectiousness itself is also time dependent [94]. Significant variation of spreading patterns was obtained in the widely-used hashtags on different topics spread on the Twitter systems [95]. Vicario et al. analyzed the diffusion process of information related to scientific and conspiracy news among the Facebook users, and the information cascade dynamics were very different, which indicated that news with different categories have different assimilation [96]. Coscia [97] indicated that a piece of information that located in the periphery of the information similarity space has the significant advantage in the potential popularity. Furthermore, the different expressions for the same thing also generate various spreading dynamics. And some information spreading patterns, especially the disease information, are still coupled with the diffusion of the corresponding disease among the population [98–100]. Although we can obtain many empirical analyses between some simple content features and spreading patterns, such as whether the tweets contain URL or hashtags [101,102], the length of the messages [103,104] and so on, it is still unclear that what features of the content are the critical factors to affect the spreading pattern. In addition, the different temporal patterns are also identified between various types of contents. It is generally known that rumor detection is a critical problem in information diffusion on social networks, and Kwon and Cha found that the temporal patterns were very different between rumor and no-rumor information [105], which is also an important method to detect the rumors where rumor-like messages have stronger fluctuations over time [106,107]. Online experiments of the behavior spreading reveal the social influence effect [66,108], leading to the different realizations from the simple spreading process assumption, such as news spreading.

The second factor is the diffusion environment of information spreading, which is also referred to as the *context* [88]. Many types of information spreading processes can be considered as the propagation through peer-to-peer interactions, such as reposting the information from someone you followed on Twitter, sending messages to your friends on Facebook, communicating through physical contacts and so on. And the underlying network topologies are the most important environment for such diffusion processes which is proved to be related to the temporal pattern [109]. Generally, users with more friends in the network would have greater impact on spread speed as well as the information popularity [110], and it is also a very important yet simple metric to evaluate the user's influence in social science. Other features of the network topology, including the degree distribution [111], small-world structure [93], ego structure [112,113], tie strength [114,115] and the ideological homophily in affiliation [116], all played very important roles in information spreading. For example, Bakshy et al. found that weak ties would play a more dominant role in the dissemination of information online than currently

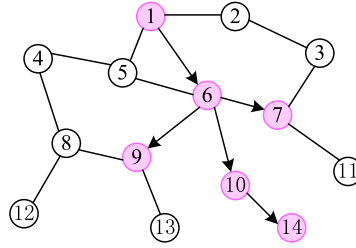


Fig. 2. A toy example of *information diffusion network* on a social network. The pink nodes represent the individuals who assess the information, and this information diffusion starts at node 1 and terminates on nodes 7, 9 and 14. And the arrows indicate the information transmission direction.

believing that more weak ties were responsible for the propagation of novel information [109]. However, the topologies of the social networks in real systems are quite diversity and complicated, leading to the abundant temporal patterns of information dissemination. Even more complicated, the edges are not continuously active in social system, which is also referred to as *temporal network* [117], and the temporal structure of edges can significantly affect the interaction dynamics through the network. In addition, the information diffusion can also affect the evolution of the social network [118] and the information become popular (or burst) when disjoint clusters of users begin to merge to form one giant component with discussing the corresponding topics [110]. Recently, although many theoretical model and numerical simulation methods [119] are proposed to interpret the spreading dynamics on the underlying network structure, the simulation results are always far away from the empirical observations [6].

3.1.2. Diffusion structure pattern

In this review, we just focus on the information spreading mainly via the social networks, and the spreading processes merely via broadcast or television are out of our discussion. As the information transmitted from the source to others through the peer-to-peer interaction (the links on the social network), all the propagation paths of the information could be extracted to form a connected subnetwork, which is also referred to as the *diffusion network* (or *information cascading*). In the diffusion network, nodes are the individuals who accept the corresponding information, and edges are information paths that individuals transmit information to others. As illustrated in Fig. 2, the diffusion network of each information can be considered as a tree-like structure [6,120], where the individual who posts the message first would be the root node (or the *information seed*, as node 1 in Fig. 2), while the individuals who do not transmit the information out would be regarded as the leaf nodes (as nodes 7, 9 and 14 in Fig. 2). And in some cases, the multiple *parents* are allowed [121], where an individual can receive the information from more than one individual. In general, we can dig a distinct diffusion network for each information on the social networks.

The empirical analyses on the diffusion network focus on two key quantities: the frequency of distinct diffusion structures and the diffusion network size distribution among all distinct diffusion events [6]. As the common sense of the theoretical diffusion models, such as the threshold model [9], disease-like diffusion model [122], “viral” marketing process [123] and so on (detailed review of the diffusion model could be found in Section 4 *Dynamical Models*), the model results showed that popular events or information can transmit multiple steps from the *information seed* and it is very likely to generate the viral spreading with large number of informed individuals.

However, the empirical observation of the real system is very different from the model generations. Goel et al. [6] studied the diffusion patterns with extracting the diffusion networks of the data collected from seven online domains, ranging from communications platforms (*Yahoo! Kindness*, *Zync*, *Friend Sense* and *Yahoo! Voice*) to networked games (*The Secretary Game*) to microblogging services (*Twitter News Stories* and *Twitter Videos*). With reconstructing all diffusion networks of the datasets, they found that the vast majority of diffusion trees (ranging from 73% to 95% across the seven domains) just have one node, which indicated no diffusion at all. And the total fraction of the several simple tree structures (as shown in Fig. 3(a)), accounted for more than 97% of all diffusion networks in each domain. The tree size is the number of individuals in the diffusion network, and the tree depth is defined as the number of layers the information spreads out from the *information seed*. And both distributions of tree size (Fig. 3(b)) and tree depth (Fig. 3(c)) were positively skewed, indicating the high concentration of small-size and short-layers diffusion trees, and the large diffusion networks seem to be the extremely rare phenomenon in the real diffusion processes. In addition, similar right-skewed and heavy-tailed size distribution of the diffusion networks are also observed on many other systems, such as Facebook [92], Digg [111], Sina Weibo Platform [124,125], Slashdot [126], Email Communications [127–130] and so on.

Recently, Goel et al. [120] proposed the *structural virality* to characterize the diffusion network structure. The structural virality (v) is defined as the average distance between all pairs of nodes (i, j) in a diffusion network (also referred to as the Wiener index),

$$v = \frac{1}{n(n-1)} \sum_{i=1}^n \sum_{j=1}^n d_{ij}, \quad (1)$$

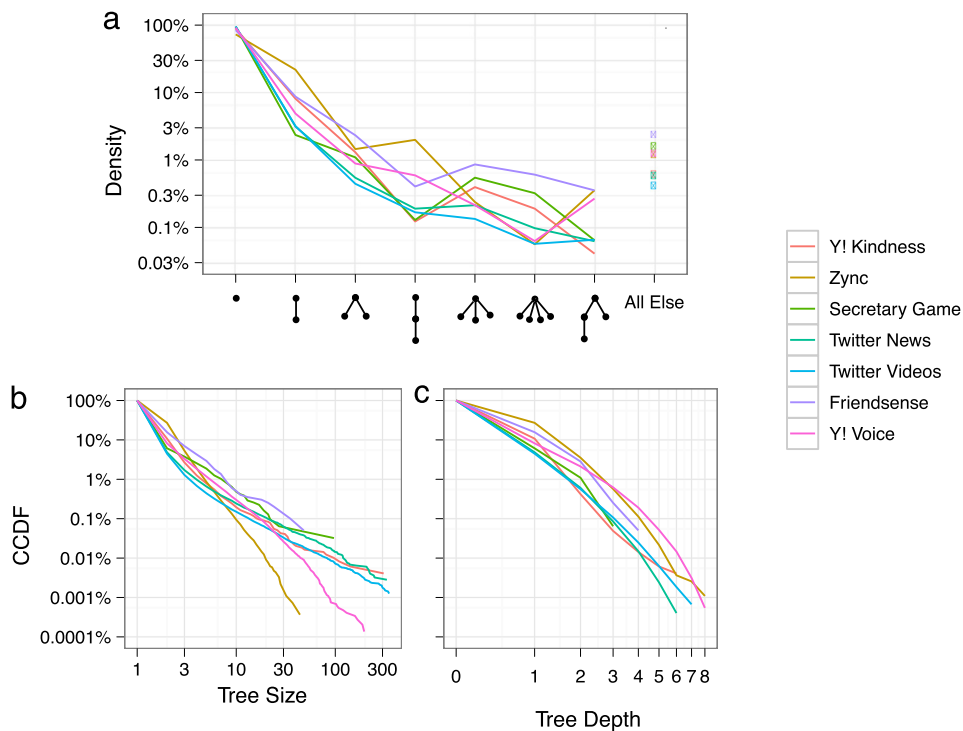


Fig. 3. The distribution of the diffusion networks on the seven domains. a, b and c represent the distribution of tree structure, tree size and depth respectively.
Source: From Ref. [6].

where $n > 1$ is the number of nodes in the diffusion network and d_{ij} is the length of the shortest path between node i and j . Intuitively, low structural virality shows that the diffusion network grows through a small number of hub nodes (broadcast effect), while high virality indicates many long information paths in the diffusion network (viral spreading). And the empirical analyses on many systems show the small value of structural virality as well as the low correlation between the structural virality and the size of the diffusion networks [6,92,120]. In contrast to the above-mentioned empirical results, Anderson et al. [131] extracted the very different diffusion structural pattern on the LinkedIn dataset, where the tree-size and tree-depth distribution decay significantly slower than the previously studied diffusion systems. The tree-size and structural virality are strongly correlated, indicating the multi-step diffusion on the large LinkedIn signup cascades.

Motivated by the observation of the heterogeneity of the diffusion size distribution, it would be an interesting problem that how the information seed would affect the final size of the information diffusion network, which would be very important in the markets such as the new product diffusion [132]. Initially, the widely accepted view is that the hubs [133] would be the most influential spreaders because the large number of their followers could receive information from them directly. However, the detailed structure of the social network is not well considered in this case, and Kitsak et al. [134] showed that the information initiated by nodes with large coreness would spread to more individuals using the disease-like model assumption, which was also proved with the empirical analysis on many online domains, including Twitter and Facebook [135]. And more recently, the extension of the H-index concept is also considered as an important parameter to quantify the influential spreaders [136]. Unlike considering the influence of individuals, Morone and Makse [137] mapped the influence maximization problem onto optimal percolation to identify the set of optimal influencers. In addition, rather than the mainly consideration on the network structure, Kim et al. found that information could be efficiently propagated using neighbors with a high propagation rate rather than those with a large number of neighbors, according to the analysis of Twitter dataset related to the 2010 UK general election between 5th and 12th of May [138].

Actually, the information diffusion among the social system would be generated by the combination of *broadcast* and *viral spreading* [120]. However, the most information diffusion literatures focus on the spreading process of peer-to-peer interactions, also referred to as the viral spreading, ignoring the influence of the broadcast mechanism, where a large number of individuals can receive the information directly from the same source, such as the mass media, news websites, video, TV and so on. Using the data of URL mentions in the Twitter network, Myers et al. [4] presented that about 71% of information by volume could be attributed to peer-to-peer interactions, and the rest is due to the external influence. More surprisingly, tracking the propagation of the hashtags in the Twitter network, Lehmann et al. [87] found that the peer-to-peer spreading process plays a minor role in hashtag popularity, while the exogenous factors such as communication in mass media would be more important to drive the hashtag propagation. And Liu et al. [139] obtained the similar results on the analysis of

the event diffusion on Sina Weibo Platform, and they also proposed a spreading model coupling with broadcast and viral spreading to show the various diffusion patterns determined by the events. Empirically, the structural virality is typically low and independent of diffusion network size in many systems, implying that the high information popularity would be largely caused by the broadcasting influence [120].

3.1.3. Human dynamical pattern

The heterogeneity and diversity of users' behavior on the social network would also generate different patterns in information spreading. The influence of the users' behavior can be classified into two respects.

(i) The diversified users' preference

It is widely accepted that users with similar interest or preference are likely to interact with each other, which in turn affects the information spreading significantly. Temporal motif analysis on the mobile phone communication system reveals the homophily feature in the social interaction, where communications tend to be more frequent between individuals with similar traits, such as sex and ages [1]. Bakshy et al. [116] observed the ideological homophily in friend networks based on the Facebook Data where the conservatives/liberals are more likely to make friends with similar political affiliation.

Besides the influence on the social network structure, the users' interest or preference would also affect information diffusion directly. Generally, users would like to assess and transmit the information which is consistent with their interest and preference, and the different preferences are usually interpreted as the spreading probability (such as the retweet ratio or mention ratio [140] in Tweeter or Tweeter-like systems). However, it would be very difficult to extract the spreading probability in the real systems, and some parameters could be roughly considered as the individuals' spreading probability, for example, the fraction of neighbors who accept the information [125]. More precisely, Gomez-Rodriguez et al. [132] studied the transmission rates of each pairwise with maximizing the likelihood of an observed set of information cascades with considering three well-known parametric models: exponential, power-law and Rayleigh. And the spreading environment, such as the information content, social influence [141] or even the emotion [142] would influence the users' interest and preference significantly. And the changeable spreading probability would also be critical factor that most spreading models cannot match the real information spreading process very well.

(ii) The bursty event dynamical

Generally, individuals' behaviors exhibit temporal bursty pattern [143], where frequently occurring events are typically observed over very short periods, followed by long periods of inactivity. And on the macro-scale, the users' inter-event time could be described as the power-law distribution or fat-tail, which is deviated largely from uniform or Poissonian statistics. This bursty temporal pattern is also a universal phenomenon in the information diffusion system, such as the phone communication [144,145], web browsing [146,129], online interactions [124] and so on.

Most empirical results show that the temporal bursty pattern would slow the information spreading. The seminal observation for the influence of the bursty on the spreading process was conducted by Vazquez et al. [147] with analyzing the spread of Email worms among Email users. The time interval (τ) between two consecutive emails sent by the same user in the Email data was proved to be the non-Poissonian process ($P_E(\tau) = A\tau^{-\alpha} \exp(-\frac{\tau}{\tau_E})$, Eq. (1) in Ref. [147]) while users interacted uniformly based on the Poisson process assumption. And the virus prevalence decay times of simulation using empirical email sequences were significantly larger than that predicted by the standard Poisson process based model, which indicated that the non-Poissonian activity pattern would slow the spreading process. In addition, they obtained that the prevalence decay time was given by the characteristic decay of the time interval distribution, according to the analysis of the correlation between the long time decay of the prevalence and the long time interval (τ) behavior. Similar result was also presented in Ref. [148], which concluded that the correlation was depending on the interaction of agents to some extent, but insensitive to the network topology. Iribarren and Moro [2,149] also found that the large heterogeneity in the response time should be the reason of the slow dynamics according to the branching model analysis, especially for the case of the much slower pace of the information spreading in the email experiment than the traditional models.

Besides the temporal bursty pattern, some other heterogeneities caused by the user behavior would also influence the information spreading pattern. Karsai et al. [3] proposed the *SI* model on the human communication network using the event sequences, where an infected individual would infect a susceptible neighbor if there was an event (a phone call, or an email) between them at the corresponding time. To gain the further understanding of the network topology and burstiness, they provided several reshuffled methods that preserve partly of the following correlations: community structure, weight-topology correlations, bursty event dynamics on single links, event–event correlation between links and the daily pattern. Simulation results showed that besides the bursty activity patterns, the weight-topology correlations were also an impeding effect to slow down the spreading. Backlund et al. [150] proposed a topological-temporal threshold model where a user would become adopter with a high rate of adopted neighbors within a chosen time, and performed the model on four different empirical temporal networks, ranging from call, SMS, Email to face-to-face interactions. Results showed that the bursty pattern would suppress the cascade size while timing correlations between contacts promote the adoption cascade. In addition, the long-lasting interactions [151], the correlated burst (correlation between the time intervals) [151] would also be the origin of slowing down of information spreading.

Very recently, some results indicated that the bursty pattern did not always slow the spreading process. In fact, the phenomenon that the heterogeneous contacts patterns that speed up the spreading process have been reported in the simulated infection on the sexual contacts network [152] and some other generated temporal networks [153]. The different

influence of the bursty pattern on the spreading process should call attention to the importance of small inter-event times or the role of lower bounds of inter-event times, whose effect has been largely ignored [154]. Motivated by this concerns, Jo et al. [154] performed the *SI* spreading model in infinite systems for arbitrary time interval distributions. For the non-Poissonian bursty processes, in order to test the lower bound effect, the shifted power-law distribution with an exponential cutoff was considered ($P_0(l) = \frac{l_c^{\alpha-1}}{\Gamma(1-\alpha, \frac{l_0}{l_c})} l^{-\alpha} e^{-l/l_c} \theta(l-l_0)$, Eq. (10) in Ref. [154]). The result indicated that the burstiness accelerated the spreading as compared to a corresponding Poissonian process for the early and intermediate time stages, while the dynamics in the late time stage was very different that the burstiness resulted in a slower convergence to a fully infected state ($\propto t^{-\beta}$) in contrast to the exponential decay of the Poissonian process ($\propto e^{-t}$). And this would be the origin of the conflicting spreading phenomenon with the bursty patterns observed in various systems. Furthermore, Horváth and Kertész [155] applied the *SI* model on temporal networks with considering burstiness, topology and non-stationary. For the stationary bursty process which was governed by power-law time interval distribution (the Pareto distribution $P_{pow}(t) = t_{min}^{\alpha} \alpha \frac{1}{t^{\alpha+1}}$, if $t \geq t_{min}$, Eq. (1) in Ref. [155]), there was a crossover (α_c) from slowing down to acceleration comparing with the Poissonian case, where the burstiness slowed down the spreading process for $\alpha \in (1, \alpha_c)$, while accelerated spreading process for $\alpha \in (\alpha_c, \infty)$. The non-stationarity had a significant effect on the spreading speed for strongly fat-tailed time interval distribution (with small α), and the process of young age can cause rapid spreading even for $\alpha \in (1, \alpha_c)$. The age of the process also can lead to an accelerated or decelerated spreading in the non-stationary process governed by the same power-law time interval distribution. In addition, the correlation between events, which cannot be attributed to the time interval distribution, can both slow-down or speed-up diffusion in systems with the same static topology [156].

At last, we also review the effect of the circadian and intra-day or weekly patterns on the information diffusion. Many empirical analyses show that the number of message on the systems varies over time with regular intra-day patterns, where users are active during day time (usually peaks round 10:00am or 8:00pm), and inactive in midnight [124,157]. And Gao et al. [157] proposed a time mapping process (referred to as *weibo time*) to eliminate the effect of users' circadian in prediction of information spreading dynamics. Using the public Twitter messages, Golder and Macy [158] identified the diurnal and seasonal mood rhythms, where the temporal affective mood pattern of positive affect was very similar to the information spreading behavior pattern on the social systems. In addition, the circadian is also considered as the origin of the power-law scaling of the time interval distribution [159]. However, using the de-seasoning methods, Jo et al. [160] found that the circadian and weekly patterns of users' activity was not a significant factor to cause the temporal communication patterns (inhomogeneous and bursty).

3.2. Cascading effects

3.2.1. Information cascades

Cascading dynamics are very important and yet ubiquitous in the complex systems, where one individual or a tiny population can affect a large number of nodes with the avalanching growth for the intricate individual interaction patterns [161]. The most studied cascading behavior is the *cascading failures*, which is extensively reported in many fields such as the infrastructure systems [162], traffic jams [163], financial systems [164] and so on.

In the information spreading domain, human natural inclination to share information with others drives individuals to mention or copy the information from one of their neighbors, and then transmit it to the other neighbors. When large number of people participate this information propagation, large information cascade emerges. In many literatures, the information cascades are also defined as the diffusion network, and the statistic properties such as the heterogeneous size distribution and low structure virality are reviewed in detail in Section 3.2. And the cascade size distribution could emerge using the threshold assumption, where only a tiny proportion of the cascade will break out for they pass the critical point and the remained large proportion would diminish before the critical point of the outbreak. In general, large cascades occur in a short period of time where the corresponding information content usually outbreak quickly after it firstly reported without lasting for a long time.

Recently, many theoretical models are proposed to describe the emergence of the large cascades, such as the sandpile model [165], the threshold model [9], simple spreading process and so on. However, the heterogeneous human behaviors are not fully considered in these models, where the users' differences are all presented in the underlying social structures. In fact, the emergence of the information cascade is associated with users' interest, emotion and profiles, including ages, genders, jobs and even races. It is challenging to describe the information cascade due to the heterogeneous human behavior pattern and users' diversified decisions to the information item. To this end, two layers of studies on the information cascading appear in recent literatures.

On the micro-level, more detailed users' behaviors are considered in the agent-based model of information diffusion. For example, Weng et al. [166] considered the competition among information in the use's screen, where each user has a memory list and information screen with limited size, which is also referred to as *limited attention*. The information would be repost if it catches the user's interest or it would be forgotten for the limited attention. According to this assumption, the model generations, including the information popularity (cascade size), users' activity and lifetime, were consistent with the statistics results from the empirical data collected from Twitter. In addition, using the critical branching process, the

cascade size generated from competition mechanism follows the power-law distribution with exponent smaller than 2, which is similar to the observation in empirical process [167–169].

On the macro-level, the data-driven methods are applied to detect the influence of users' behavior on information diffusion. Fortunately, using the machine learning algorithms, the users' hidden information could be mined according to users' online behavior which can be collected easily. For example, many highly sensitive personal features, such as the sexual orientation, personality traits, interest and so on, could be accurately predicted using only the *Facebook Likes* datasets [170]. And a user's interest feature [171] and emotion categories [172] could also be extracted from her/his information content, or even the posted information of the users that s/he followed in the social networks. With clustering the user's information content, the user's topical diversity could be denoted as the entropy of her/his posted information content among various topics [173], and the participation of the users with high topical diversity in the early stage tends to form large size information cascades in the future. Apart from the user's interest, user's behavior also influenced by the local social influence (the structure among his/her friends), and Zhang et al. [174] identified that the likelihood of user's retweeting behavior was significantly larger if his/her friends who also retweeted the message did not know each other by training a logistic regression classifier in the microblogging system. Although there are abundant empirical phenomena related to the cascade, the findings seem a bit fragmentary and non-systematic, where the data of the spreading process on different systems, or even different stages for the same systems are very different. In order to present the information cascading process more clearly, the combined understanding of these two layers would be a new direction.

3.2.2. Cascade prediction

In this part, we just focus on the prediction of the information cascades with large size, where the outbreaks of cascade would be the paramount important problem in information spreading. For example, the outbreak of the bad news such as the rumor would bring bad or even devastating effects. In this case, the prediction of the rare outbreak cascades in the early stages is a critical issue in the information cascading research. And we would ask the questions that why, when and how does a large information cascade form, for each individual focuses on the information related to his/her interests. Inspired by these questions, the following three research points have aroused considerable interests: the information cascade predictability problems, burst time prediction in cascades and the information popularity (size of the cascade in the final state).

Measuring the predictability of large cascade is difficult for the multiple factors that affect the information diffusion process [175]. The view of the unpredictable of the large cascade is proposed because of the rare occurrence of the large cascades in many systems [6] and the inherent inequality and unpredictability [176] in social systems. Salganik et al. [176] concluded that no measure of quality can precisely predict the outbreak cascades with creating an artificial “music market” experiment, which showed that the same songs with identical initial conditions could achieve very different levels of popularity. And Cheng et al. [92] observed the similar phenomenon that independent resharings of the same photo can generate very different cascade sizes using the photo-resharing data from Facebook. However, they showed that “which cascade ends up with largest size” could be predicted with high accuracy with the observation of temporal and structural features of the initial spreading process [92]. In addition, Holme et al. used the diversity of the out-break size (the standard deviation) as a key quantity for unpredictability, and presented the decay of unpredictability as a function of time using the *SIR* process on the static network [177] as well as the temporal network [178]. In general, more accurate prediction could be given if we have more information about the cascade, such as the more diffusion path in the cascade. But “to what extent of the known information is the cascade dynamic predictable” and “which features are the most significant for the cascade prediction task” are still unclear [92]. Though it would be very hard to identify the intrinsic properties of the predictability of information cascade, some fragmentary regularities are obtained based on the empirical analysis, for example, breadth rather than depth of the cascade structure [92], and the importance of the individuals that participate in the cascade [179] at the initial steps are well correlated with the final size of the cascades.

The general cascade prediction problem can be illustrated as how to estimate the cascade growth process with the initial portion of the cascade [92]. The burst time and the final size of the cascade are the most attractive in the information cascade analysis. As illustrated in Section 4.2, the information spreading dynamics exhibit the “rise and fall” temporal pattern, and the burst of the cascade should be the origin of the quick rise pattern. Predicting when the burst of cascade time will come is also critical important in social science. The traditional regression models seem to be powerless to address this task for the “rise and fall” pattern and the various lifetime of the cascades [180]. In this case, Wang et al. [91] proposed a classification based method to predict the burst time. Dividing the time span of the cascade into many equal time windows, they transformed the time prediction task to a classification problem with predicting which time window the burst would appear.

From the literatures up to date, we can roughly classify the methods of predicting the final size of the cascade (or information popularity) into three categories. The first type is based on the peer-to-peer spreading assumption on the underlying networks [181]. The key problem is how to mine the propagation probability of the links, and even sometimes the network structure is incomplete which need to infer the coupled of network structure and propagation probability [182]. The second type is feature based method, which needs to extract a list of features that might affect the information diffusion, including the content, users, network structure and temporal features [183,184], and then apply different machine learning methods to make the prediction of the information popularity in the final state [92,179]. Although the feature based method could generate good performance of the prediction, the challenge is that the result is sensitive to the quality

of the extracted features. The third type is based on the probabilistic model, for example, using the theory of exciting point processes (Hawkes process) [94,185]. Very recently, Zhao et al. [94] proposed the self-exciting model for information cascades (SEISMIC) following this framework, which required no feature engineering and minimal information about the information cascade as well as the underlying network structure. Apart from the high prediction accuracy of the final cascade size, it also could identify whether the cascade was in the supercritical state or not at each time point with checking that the transmission rate was above or below the critical threshold [94].

4. Dynamical models

4.1. Threshold models

4.1.1. Linear threshold model

The linear threshold model was first proposed to describe collective behavior [11] and has been applied to describe a series of binary decision phenomena in economics and sociology. The main idea of this model is that individuals in the network display a herd-like behavior, i.e., they make decisions based on the actions of their neighbors. This kind of decision-making strategy implies that the threshold model contains memory of the exposure history. Considering a network with N agents, each of the agents can be in one of the two states: *active* state and *inactive* state. Furthermore, a threshold ϕ is assigned to each individual in the network, which is randomly chosen from a distribution $f(\phi)$ ($f(\phi)$ can be selected as desired). The detailed model is described as follows:

- At the beginning, a small fraction of individuals are randomly selected and assigned to *active* state as the information seeds. All the others are in *inactive* state.
- During each time step, for an individual i in *inactive* state, denote ϕ_i as the threshold value of i . The state of i will change to *active* state if the fraction of its neighbors in *active* state is equal to or is larger than ϕ_i . All the individuals in *active* state stay unchanged.
- The diffusion process comes to an end when the number of individuals in *active* state becomes stable.

In terms of information diffusion, linear threshold model to investigate information cascade was firstly studied by Watts [9]. The paper demonstrated that when the model is performed on networks, information cascade will emerge. The main focus of this study is when a global cascade would be triggered by a small fraction of *initiators* and the expected size of the global cascades. Therefore, a generating function approach is applied to investigate the cascade conditions [122]. If there is only one initiator at the initial step, the diffusion will continue only if at least one of its neighbors has a threshold such that $\phi_i \leq 1/k_i$, where k_i is the degree of node i . In terms of this model, it is conjectured that a global cascade will be possible when the subnetwork of *vulnerable* vertices percolates throughout the network as a whole. When the connectivity of the network is low, the cascade sizes with power law distribution are obtained. However, when the network is highly connected, the cascade size distribution is bimodal, as the propagation of cascade is limited by the stability of the nodes in the network.

For a spreading model, one of the most important dynamical behaviors is the threshold value. Centola et al. [186] gave numerical solution of the threshold points in various complex networks, including empirical social networks. They concluded that different from the results of the independent interaction models, random links between otherwise distant nodes in fact reduce a networks ability to propagate collective behavior. Other factors that influence cascade spreading have also been studied in numerous literatures, such as the size of *initiators* [187,188] and clustering coefficient [189].

Recently, threshold model has been applied to various directions of complex systems. For example, in order to maximize the expected diffusion of information, the issue of choosing influential sets of individuals is a problem in discrete optimization. Therefore, Kempe et al. [190] obtain the first provable performance guarantees for approximation algorithms in a number of general cases by using threshold model and other simple but widely used models. Another application of threshold model is to study the influence of community structure on information diffusion [12,191]. By using the linear threshold model, Nematzadeh et al. [12] have revealed the roles of modular structure in information diffusion, which indicates that strong communities can enhance local spreading and weak communities can make global spreading easier. Moreover, they found an optimal modular structure which can both facilitate local and global spreading.

4.1.2. Generalized threshold models

In the process of spreading, generally, two kinds of spreading models characterizing the contagion between individuals are studied: (i) independent interaction models, such as *SIS* and *SIR* models, in which successive contacts result in contagion with independent probability; (ii) threshold models, in which adoption of behaviors depends on the states of current active neighbors. Dodds et al. [192,193] generalized the threshold model by introducing memory of past exposures to a contagious influence. The model also incorporated independent interaction models as a particular case. Considering a population with N individuals, the individuals can be in one of the three states: *S* (susceptible), *I* (infected) and *R* (recovered). Each susceptible node i is assigned a threshold d_{i^*} (randomly choosing from a distribution $g(d^*)$). The detailed model can be illustrated as follows:

- **Initialization:** At the initial step, a small fraction of nodes are randomly chosen as the seeds of the diffusion, which are in *I-state*. All other nodes are in *S-state*.

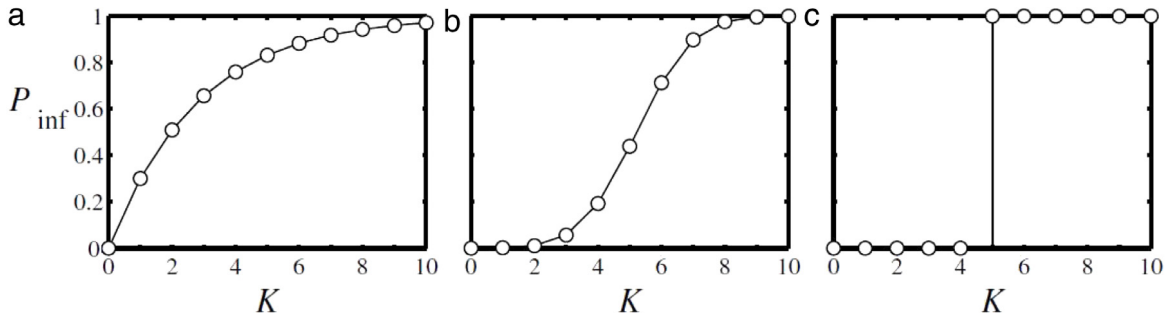


Fig. 4. The infected probability of an S node P_{inf} after K contacts with infected neighbors in T steps. (a) Independent interaction model (SIR). (b) Stochastic threshold model. (c) Deterministic threshold model.

Source: From Ref. [192].

- **Infected:** During each time step, for an S node i , it will receive a positive dose $d_i(t)$ (from some distribution $f(d)$) with probability p from each of its infected (I) neighbors. Each susceptible node will keep a memory of doses over the previous T steps and will turn to the infected state if the accumulated dose $D_i(t) = \sum_{t'=t-T+1}^t d_i(t')$ is larger than d_{i*} at some time.
- **Recovery:** For an infected node j , if the accumulated dose $D_j(t)$ falls below d_{j*} , it will recover to R state.

According to the description of this model, we can calculate the probability of an S node that has encountered $K \leq T$ infected individuals in T steps becomes infected, which is $P_{inf}(K) = \sum_{k=1}^K \binom{K}{k} p^k (1-p)^{K-k} P_k$, where $P_k = \int_0^\infty dd^* g(d^*) P(\sum_{i=1}^k d_i \geq d^*)$ is the average fraction of individuals infected after receiving k positive doses in T steps. By choosing different values of $p, f(d)$ and $g(d_*)$, we can get different kinds of models, three examples are given in Fig. 4. Besides the general solution for the model, the authors also obtained the conditions for when the three classes of collective dynamics exist and how the memory length affects the classification of the systems in this work.

Threshold model has been generalized to multiplex networks. Brummitt et al. [194] assumed that a node becomes active if the fraction of active neighbors in any layer exceeds the threshold. Moreover, Yağan et al. [195] proposed a threshold model in which nodes switch state if their *perceived* proportion of active neighbors exceeds the threshold on multiplex networks. In the field of temporal networks, a generalized deterministic threshold model with history-dependent contagion is simulated on empirical temporal networks, which turns out to be that burst activity patterns can facilitate spreading process [196]. In another literature, Karimi et al. [197] modified the linear threshold model by considering both the chronological order of events and the period a contact that can influence an actor, with the result that burst activity of individuals may hinder the spreading process. More recently, some researchers become to consider using non-Markovian process to describe social contagion in the assumption of threshold model [198,199].

4.2. Cascading models

4.2.1. Independent cascade model

Inspired by the theory of interacting particle systems [200], an independent cascade model was first proposed by Goldenberg et al. [123,201] to explore marketing model. Similar to the threshold model, two states are considered in this model, i.e., *active* or *inactive*. Generally, there are two basic hypotheses in this kind of models. The first one is that the probability of a node i accepts the infection from an active node is independent of the influence of other active nodes on i . This implies that the independent cascade model is sender-centered, whereas the threshold model is receiver-centered. The second hypothesis is that any active node i has only one chance to infect its inactive neighbor j , whether success or not, node i will not influence j in the subsequent steps. Based on these hypotheses, the detailed propagation algorithm can be concluded as follows:

- Initially, a small fraction of nodes are randomly chosen as the seeds of the diffusion process (in the *active* state).
- At step t , a newly active node i (has been activated at step $t - 1$) will activate its inactive neighbor j with probability p_{ij} . The neighbors of j that were activated in step $t - 1$ will tend to influence j in an arbitrary order.
- The process will repeat until there is not any active nodes that will influence the other inactive ones.

As described above, in the independent cascade model, the diffusion probability p_{ij} through each link should be specified in advance. However, in real networks, the true values of these probabilities are usually unavailable in practice. This poses a problem of how to predict each diffusion probability through each link from the real data. Saito et al. [202] addressed the problem of how to predict the probabilities for the independent cascade model from a set of past propagations. They defined this as a likelihood maximization problem and used the EM algorithm to predict the probabilities. The experiment on real blogroll network showed a good performance. Another observation [203] focused on using the independent cascade model to propose learning models of information flow in networks and predict activation probabilities from attributed and unattributed data. Since the exact flow calculations based on the independent model are exponential, a method for sampling

an approximation efficiently using the Metropolis–Hastings algorithm is proposed in this work to learn joint and conditional flow probabilities.

When considering the information diffusion process, an important aspect is how to control the information cascade size in the final state. For example, for a positive information, we may want to spread it as widely as possible. However, when it turns to a negative news, the problem is how to minimize the influence of this news. The first one is called influence maximization problem [190,204,205], and the second one is named contamination minimization problem [206].

The purpose of influence maximization problem is to find a limited number of nodes that are influential for the diffusion of information. Suppose A is a set of seed nodes, the expected number of active nodes in the final state is denoted as $\sigma(A)$. The influence maximization problem requires to find a set S , $|S| = k$, such that $\sigma(A)$ is maximum. Kempe et al. [190] studied the influence maximization problem as a discrete optimization problem based on two models, e.g., independent cascade model and linear threshold model. Computational experiments on large collaboration networks showed that their greedy hill-climbing algorithms significantly outperform the high-degree and centrality heuristics. However, a large amount of computation is required for the greedy algorithm, as the marginal gains (denoted as $\nabla\sigma(A)$) for $\sigma(A)$ should be calculated for many times under different set of initial nodes. Thus, an efficient method to estimate $\nabla\sigma(A)$ based on bond percolation and graph theory is proposed [204] and is applied to approximately solving the optimization problem under the greedy algorithm. Experiments performed on large-scale real-world networks showed that this method can reduce computational cost.

In terms of the contamination minimization problem, previous works investigated this problem by removing nodes from a network [207,208]. A new method of blocking a limited number of links in a network for reducing the cascade size was given by Kimura et al. [206]. In this work, a contamination degree $c(G)$ is defined on graph $G = (V, E)$, which is the average of influence degrees of all the nodes in G , i.e.,

$$c(G) = \frac{1}{|V|} \sum_{v \in V} \sigma(v; G) \quad (2)$$

where $\sigma(v; G)$ is the influence degree of node v in G , denoted as the expected number of active nodes at the end of the propagation process for initial active node v . As it is difficult to solve the contamination minimization problem directly on large networks, a greedy algorithm which outperforms traditional link-removal methods is proposed as an approximate solution to this problem. Besides physicists solutions, we also hold one dissemination management section in this review for discussing minimize/maximize information spreading from computer scientists perspective (see **Sec. Dissemination Management**).

4.2.2. Generalized cascade models

The parameter estimation and influence maximization problem in terms of the independent cascade model are usually with high computational complexity, so some generalized cascade models are also proposed to solve this problem. We will briefly introduce these models in this section.

As we know, when information spreads in the population, a time-delay phenomenon may emerge. However, the independent cascade model cannot describe the time-delays for propagation. A prototype to incorporate time-delay in the independent cascade model is given by Gruhl et al. [209], in which they used parameters to represent time-delay through links. An EM-like algorithm is used to estimate the parameters from the observed data, which shows good performance on *ER* network. The drawback of the model is that it considers time as a discrete variable, which means that each node can be activated at a specific time, and this is usually not true for information diffusion. Accordingly, a continuous-time independent cascade model (*CTIC* model) is proposed to solve this problem [210,211]. In the *CTIC* model, a parameter $r_{ij} > 0$ is given as a time-delay parameter for link ij . The model is similar to the independent cascade model, except for the contagion process. That is, suppose i is active at time t and has a single chance to activate its neighbor node j (in the inactive state). At the same time, a time-delay parameter δ is chosen from an exponential distribution with parameter r_{ij} . Assume that node j is in the inactive state before time $t + \delta$, then node i will activate node j and succeed with probability p_{ij} (diffusion probability through link (i, j)), thus node j will become active at time $t + \delta$. Considering time-delay in this model makes the time-sequence observed data structural. Thus a rigorous likelihood to obtain the observed data is formulated and the parameters (including time-delay parameters and diffusion probabilities) are estimated by maximizing this likelihood. The method is applied to solve the problem of ranking influential nodes and evaluating topics propagation, both of which work better than conventional methods.

Actually, a series of diffusion models in social network are based on concrete graph structure and ignores the temporal dynamics of the diffusion, which makes the estimation unreliable. Therefore, a model based on *CTIC* model involving semantics, social and time is proposed in [212], which is called Time-Based Asynchronous Independent cascade (T-BaSIC) model. In this model, a Bayesian logistic regression is used to infer time-dependent diffusion probabilities between nodes on Twitter. Experiments on real datasets show the model's effectiveness on predicting the dynamic of the diffusion.

In the preceding section, the independent cascade model is used to solve the influence maximization problem in several literatures. Simultaneously, generalized independent cascade models are also used to solve this problem. A representative one is the decreasing cascade model [213]. In this model, an incremental function $p_j(i, S) \in [0, 1]$ is defined as the success probability of node i activating node j , where S is a set of nodes that are j 's active neighbors who have already attempted

but failed to make j active. According to the definition of this model, the independent cascade model is a special case of this model when $p_j(i, S)$ is a constant. By adopting this model, a set of initiators are chosen as the targeted nodes to yield the largest expected cascade, in which they prove that the expected number of influence nodes is a monotone and submodular function of the targeted set.

Obviously, different information has different cascade sizes in social networks. This is because users' authoritativeness, expertise, trust and influence change with topics. Hence, a topic-aware propagation model based on the independent cascade model is proposed (i.e., *TIC* model) [214], which involves items' characteristics, users' interests and social influence in the model. Given a directed graph $G = (V, E)$, each link $(i, j) \in E$ and each topic $z \in [1, K]$, a probability $p_{i,j}^z$ is introduced as a weight of link (i, j) concerning topic z . For each item m , a distribution over the topics is given, i.e., $\gamma_m^z = P(Z = z|m)$ for a given topic z , where $\sum_{z=1}^K \gamma_m^z = 1$. Thus, the diffusion probability through link (i, j) is defined as

$$p_{i,j}^m = \sum_{z=1}^K \gamma_m^z p_{i,j}^z. \quad (3)$$

The other procedure of *TIC* model is similar to the independent cascade model. Moreover, an EM algorithm is derived to estimate the parameters from a dataset of past propagation, which shows high accuracy compared to the conventional methods.

4.3. Epidemic models

4.3.1. Classical epidemic models

The most widely used mathematical models for information diffusion are the epidemic models. The first attempt to use the mathematical method to analyze epidemic is by Daniel Bernoulli in 1760, when he studied the spreading process of smallpox. However, it was Kermack and McKendrick that laid a basic foundation for the mathematical models, as the widely used models (i.e., *SIR* and *SIS* compartment model) were introduced by them [215,216]. Later on, numerous epidemic models are proposed to explain different kinds of diseases. These epidemic models are applied to various other contagion processes, including information diffusion, innovation diffusion, promotion of commercial products as well as the spread of political movements.

Generally, people are divided into different compartments in the epidemic models, individuals in the same compartment are assigned with the same state. The most common states used in the information diffusion process are: (i) *S*: the susceptible state, represents people in this state is unaware of the information and would be infected in the future; (ii) *I*: the infected state, represents people in this state have already aware of this information and will transmit it to others; (iii) *R*: the recovered state, represents people in this state have already aware of this information, but are not care about it, thus they will not transmit the information to others any more. Different combinations of these states can result in different models, such as *SI*, *SIS* and *SIR* model.

Herein we first consider these epidemic models in a well-mixed population. The simplest case is the *SI* model, in which two states are considered, *S* and *I*. This model is similar to the threshold model and the cascade model, as when susceptible individuals know about the information (turns to *I* state), they will be in *I* state forever. Denote $s(t)$ and $i(t)$ as the proportion of susceptible and infected individuals at time t , thus we have $s(t) + i(t) = 1$. Suppose the probability of infected individuals infecting the susceptible individuals is β . Therefore, the *SI* model can be illustrated by ordinary differential equations as follows:

$$\begin{cases} \frac{ds(t)}{dt} = -\beta s(t)i(t) \\ \frac{di(t)}{dt} = \beta s(t)i(t). \end{cases} \quad (4)$$

The *SIS* model is used to characterize epidemics that have transient immunity, such as influenza. When it turns to the information process, it means that the individuals know about the information (in *I* state) will ignore the information and become susceptible again (in *S* state). We denote $s(t)$ and $i(t)$ as the fraction of susceptible and infected individuals in the population respectively. The transmission probability between *I* state individuals and *S* state individuals is β , and the recovered probability of the *I* state individuals is γ . Accordingly, the *SIS* model can also be expressed by the ordinary differential equations:

$$\begin{cases} \frac{ds(t)}{dt} = -\beta s(t)i(t) + \gamma i(t) \\ \frac{di(t)}{dt} = \beta s(t)i(t) - \gamma i(t). \end{cases} \quad (5)$$

Given an initial value $i(0)$, we can derive the expression of $i(t)$:

$$i(t) = \frac{i(0)(\beta - \gamma)e^{(\beta - \gamma)t}}{\beta - \gamma + \beta i(0)e^{(\beta - \gamma)t}}. \quad (6)$$

We define $\lambda \triangleq \beta/\gamma$ as the effective infection rate of this model. From Eq. (6), we can find that when $\lambda > 1$, the steady-state value of I is $i = (\beta - \gamma)/\beta = 1 - \frac{1}{\lambda}$, an endemic disease state is obtained in this case (indicates that there will be a number of infected individuals in the population at the final state). However, when $\lambda < 1$, we have $i \rightarrow 0 (t \rightarrow \infty)$, which means that there will be no I state individuals in the population finally and is called a healthy state. Therefore, $\lambda = 1$ is a threshold value of the SIS model, which is also known as the basic reproduction number [217].

The SIR model is introduced to explain the epidemics with permanent immunity in the population. Different from the SI and SIS model, a recovered state (R) is given in this model. Individuals in S state would be infected by the I state individuals with probability β , whereas the I state individuals would recover to R state with a recovery probability γ . Thus, the ordinary differential equations of SIR model are

$$\begin{cases} \frac{ds(t)}{dt} = -\beta s(t)i(t) \\ \frac{di(t)}{dt} = \beta s(t)i(t) - \gamma i(t) \\ \frac{dr(t)}{dt} = \gamma i(t). \end{cases} \quad (7)$$

$\lambda \triangleq \beta/\gamma$ can also be defined as the basic reproduction number of the SIR model. We can get similar conclusion as the SIS model, i.e., when $\lambda < 1$, $r = 0$, indicating that information cannot spread in the population. When $\lambda > 1$, we have $r > 0$, which means the information can spread out as the increase of λ . As a consequence, $\lambda = 1$ is a threshold value of SIR model.

4.3.2. Mathematical analysis for epidemic models

In the previous section, epidemic models are studied in a well-mixed population. These models can also be analyzed by other mathematical approaches in a network-structured demographic. In this section, we illustrate these approaches in detail in the case of SIR model. It should be emphasized that these approaches can also be generalized to other epidemic models. We focus our study on a network $G = (V, E)$ with the average degree of $\langle k \rangle$. The transmission probability between I state individuals and S state individuals is β , and the recovery rate of the I state individuals is γ .

Homogeneous mean field approach. Denoting $s(t)$, $i(t)$ and $r(t)$ as the fraction of susceptible, infected and recovered individuals in SIR model, thus we have $s(t) + i(t) + r(t) = 1$. In the homogeneous mean field approach, we assume that each individual is connected to $\langle k \rangle$ neighbors in average [218]. Thus the changes of $s(t)$, $i(t)$ and $r(t)$ are expressed by

$$\begin{cases} \frac{ds(t)}{dt} = -\langle k \rangle \beta s(t)i(t) \\ \frac{di(t)}{dt} = \langle k \rangle \beta s(t)i(t) - \gamma i(t) \\ \frac{dr(t)}{dt} = \gamma i(t). \end{cases} \quad (8)$$

From the equations of homogeneous mean field approach, we know that this approach considers the degree of the network to some extent compared to the classical approach in a well-mixed population. The threshold value of this approach is also different, which is $R_0 = \langle k \rangle \beta / \gamma$ (R_0 is the basic reproductive number) [219].

Heterogeneous mean field approach. The homogeneous mean field approach considers each individual can infect a constant number of neighbors, which is not suitable for the heterogeneous networks. A heterogeneous mean field approach is proposed to overcome this drawback [220,221]. In this approach, individuals are divided into different categories based on node degrees, i.e., $s_k(t)$, $i_k(t)$ and $r_k(t)$ represent the fraction of susceptible, infected and recovered individuals with degree k respectively. Consequently, the heterogeneous mean field approach can be governed by

$$\begin{cases} \frac{ds_k(t)}{dt} = -k\beta s_k(t)\theta_k(t) \\ \frac{di_k(t)}{dt} = k\beta s_k(t)\theta_k(t) - \gamma i_k(t) \\ \frac{dr_k(t)}{dt} = \gamma i_k(t) \end{cases} \quad (9)$$

where $\theta_k(t)$ is the probability that a random selected link is pointing to an infected individual. The specific form of $\theta_k(t)$ in terms of uncorrelated and correlated network is [222],

$$\theta_k(t) = \begin{cases} \sum_k (k-1)p(k)i_k(t)/\langle k \rangle & \text{if } G \text{ is uncorrelated network} \\ \sum_{k'} i_{k'}(t)((k'-1)/k')p(k'|k) & \text{if } G \text{ is correlated network.} \end{cases} \quad (10)$$

The threshold value of this model can also be calculated. When G is an uncorrelated network, the threshold is $\lambda^{ucr} = \frac{\langle k^2 \rangle}{\langle k^2 \rangle - \langle k \rangle}$, where $\langle k^2 \rangle$ is the second moment of the node degree distribution. When G is a correlated network, the threshold is $\lambda^{cr} = \frac{1}{\bar{A}_m}$, where \bar{A}_m is the maximum eigenvalue of the connectivity matrix $\bar{C}_{kk'} = \beta(k(k' - 1)/k')p(k'/k)$.

Pair-based approach. In order to study the epidemic models on networks, approaches consider the evolution of the edges are proposed, known as the pair-based approaches [223–225]. In this model, $[X]$ represents the expected number of individuals in different types. For example, $[S]$ expresses the expected number of susceptible individuals, $[SI]$ denotes the expected number of links connecting a susceptible individual to an infected individual, $[SIS]$ represents a triple with form of $S - I - S$. The changes of the variables can be described by the following differential equations:

$$\left\{ \begin{array}{l} \frac{d[S]}{dt} = -\beta[SI] \\ \frac{d[I]}{dt} = \beta[SI] - \gamma[I] \\ \frac{d[R]}{dt} = \gamma[I] \\ \frac{d[SS]}{dt} = -2\beta[SSI] \\ \frac{d[SR]}{dt} = -\beta[RSI] + \gamma[SI] \\ \frac{d[IR]}{dt} = \beta[RSI] + \gamma([II] - [IR]) \\ \frac{d[II]}{dt} = 2\beta([ISI] + [SI]) - 2\gamma[II] \\ \frac{d[SI]}{dt} = \beta([SSI] - [ISI] - [SI]) - \gamma[SI]. \end{array} \right. \quad (11)$$

System (11) can be closed at the level of pairs by assuming different distributions of neighbors, for instance, Poisson, binomial or multinomial distribution [223,224]. As triples are considered in this system, clustering effect of the network can also be involved by different pair approximation approaches. On the other hand, system (11) is a homogeneous pair-based approach, which can be generalized to a heterogeneous case. Like the heterogeneous mean field approach, the node degree is considered in the heterogeneous pair-based approach [225]. In addition, based on different closure method, the threshold values can be derived respectively.

Individual-based approach. The approaches introduced above have considered the network structure to some extent. However, these approaches are aggregate representations of the network, as they cannot distinguish among individuals with the same node degrees and ignores the central properties of the individuals. Thus an individual-based approach is proposed in terms of continuous time Markov chain SIR model [226].

We first give the definition of continuous time Markov chain SIR model. Suppose there are N individuals in the population, as each individual can be in one of the three states (i.e., S, I, R), thus there are 3^N possible network states. Denote $X = \{x_1, x_2, \dots, x_N\}$ as the network state, where $x \in \{S, I, R\}$. $W_X = p(X = \{x_1, x_2, \dots, x_N\})$ is the probability that the network is in state X . The state transition matrix is given by $Q = (q_{X,Y})_{3^N \times 3^N}$, in which $q_{X,Y}$ represents the transition rate from network state X to Y . At any time t , $W_X(t)$ represents the probability of the network is in state X ($\sum_{X \in \{X_1, \dots, X_{3^N}\}} W_X(t) = 1$), thus we can obtain the change of this probability with time as follows:

$$\frac{dW^T(t)}{dt} = W^T(t)Q \quad (12)$$

where $W^T(t)$ is the transpose of $W(t)$. Given an initial value $W^T(0)$, the solution of Eq. (12) is obtained

$$W^T(t) = W^T(0)e^{Qt}. \quad (13)$$

According to Eq. (13), the solution complexity of continuous time Markov chain SIR model is exponential $O(3^N)$. Therefore, an individual-based approach is given to decrease the complexity to $O(N)$. The $Q_{3^N \times 3^N}$ matrix is decomposed to N infinitesimal matrices, in which each matrix is within three states [226,227]. By using the effective average infection rate instead of random infection rate, the state change of each individual v is

$$\left\{ \begin{array}{l} \frac{dS_v(t)}{dt} = -S_v(t)\beta \sum_{z \in V} a_{v,z} I_z(t) \\ \frac{dI_v(t)}{dt} = S_v(t)\beta \sum_{z \in V} a_{v,z} I_z(t) - \gamma I_v(t) \\ \frac{dR_v(t)}{dt} = \gamma I_v(t) \end{array} \right. \quad (14)$$

where $A = (a_{v,z})_{N \times N}$ is the adjacent matrix of the network. The threshold value for this model is $R_0 = 1/\lambda_{\max,A}$, where $\lambda_{\max,A}$ is the maximum eigenvalue of A .

Generating function approach. As the *SIR* model demonstrates a transition between a phase of a finite epidemic size and a phase of only a limit individuals are infected, resulting in the *SIR* model can map into a bond percolation problem [228]. An occupation probability T of the links in the network is defined in the bond percolation theory, the ultimate size of the epidemic outbreak is exactly the size of the cluster of vertices that can be reached from the initial vertex by occupying edges with probability T . The average behavior of the networks under bond percolation can be solved by generating function. Suppose the degree distribution of the network is p_k , the generating function for the degree distribution and the excess degree generating function are

$$G_0(x) = \sum_{k=0}^{\infty} p_k x^k \quad (15)$$

$$G_1(x) = \sum_{k=0}^{\infty} (k+1) p_{k+1} x^k / \langle k \rangle. \quad (16)$$

Denote $\langle s \rangle$ and $S(T)$ as the average outbreak size and the giant component size of the bond percolation process. Thus

$$\langle s \rangle = 1 + \frac{T G'_0(1)}{1 - T G'_1(1)} \quad (17)$$

when $T G'_1(1) = 1$, Eq. (17) diverges, the threshold value can be obtained as

$$T_c = \frac{1}{G'_1(1)} = \frac{\sum_k k p_k}{\sum_k k(k-1) p_k} \quad (18)$$

when $T > T_c$, a giant component $S(T)$ emerges, which can be acquired by solving the equations below.

$$\begin{cases} S(T) = 1 - G_0(u; T), \\ u = G_1(u; T). \end{cases} \quad (19)$$

Experiments on networks show that the solution of Eq. (17) and Eq. (19) is well consistent with the simulation results. The percolation theory can also be applied to other *SIR*-like models, the results are shown in [229,230].

4.4. Other models

With the development of science and technology, we are now exposed to a large amount of data. Thus the requirement of new approaches to characterize the information process in online social network is more intense than ever before. Besides the traditional models above, there are also some new models to model the information process in terms of datasets on social networks. In this section, we will briefly introduce them.

Linear Influence Model. Even though various models have been proposed to investigate information diffusion, modeling the diffusion of social networks has always been a challenge task. As each node in the network plays different roles in the information process, a model concerned about studying the influence of a particular node and how the influence changes over time is proposed in [231]. The model which is designated as *Linear Influence Model (LIM)* is based on a number of empirical analyses on *Twitter*. The assumption of *LIM* is that the number of newly infected nodes depends on which other nodes got infected in the past.

Define $I_u(l)$ as the influential function of node u , which means the number of followup infected nodes l time units after node u is infected. $V(t)$ is the total number of infected nodes at time t . According to the assumption of *LIM*, $V(t)$ can be expressed as

$$V(t+1) = \sum_{u \in A(t)} I_u(t - t_u) \quad (20)$$

where $A(t)$ is the set of nodes that have been infected, and node u is infected at time t_u ($t \leq t_u$). The influential function is modeled in a non-parametric way, which leads the model to a simple least squares problem and can be solved on large real networks. Experiments on large scale datasets showed that the model can accurately assess the node influence and predict the temporal dynamics of information diffusion.

Diffusive Logistic Model. Most of the diffusion models focus on modeling information diffusion in temporal dimensions, while a model concerned about both temporal and spatial dimensions is given by Wang et al. [232]. They proposed a PDE-based diffusive logistic equation to model information diffusion in online social networks, which aims at modeling how information travels over time and space and the density of infected individuals at a fixed distance (Fig. 5). In this model, *friendship hops* are used as a distance metric, and the diffusion process is divided into two processes, that is, growth process

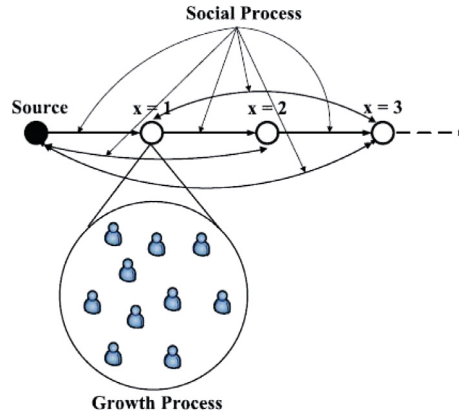


Fig. 5. Diffusion process of diffusive logistic model.
Source: From Ref. [232].

and social process. Growth process aims at modeling the diffusion process between the individuals at the same distance from the diffusion source. Social process is designed to model the diffusion process between individuals at different distances from the diffusion source. In addition, this model also incorporated the propagation between individuals who are not direct friends. The model was expressed by Partial Differential Equation and was validated on the Digg dataset. Experiments showed that the diffusive logistic model can predict the density of influenced users for a given distance effectively by choosing suitable initial condition and parameters. Moreover, this model is a first attempt to use PDE-based equation to model diffusion process both in temporal and spatial dimensions in online social networks.

Information diffusion with External Influence. Diffusion channel plays an important role in information diffusion. As we know, there are two ways for information reaching a person, that is, through the connections in the underlying network and an external influence from outside of the network. However, most of the present models focus on studying the peer-to-peer communications via the underlying networks, ignoring the external influence on the diffusion process. Recently, some researchers have incorporated the external influence to study how information spreads on social networks [4,6,139,233]. The authors [4] first empirically show that only around 71% of URL mentions on Twitter are attributed to the underlying network effects, the remaining 29% of the mentions are due to the external effects. Thus a probabilistic generative model both considering the internal and external effects is presented to illustrate information diffusion. In this model, they denote $\lambda_{int}(t)$ as an internal hazard function which governs the random amount of time it takes for an infected node to expose its neighbors and $\lambda_{ext}(t)$ as the event profile which is a stream of varying intensity of external exposures that a node receives, thus these two variables can be expressed as follows

$$\lambda_{int}(t)dt = P(i \text{ exposes } j \in [t, t + dt] | i \text{ hasn't exposed } j \text{ yet}) \quad (21)$$

for any node i and j , where t is the amount of time that has passed since i is infected.

$$\lambda_{ext}(t)dt = P(i \text{ receives exposure} \in [t, t + dt]) \quad (22)$$

for any node i , where t represents the amount of time since the diffusion first appears in the network. $\Lambda_{int}(t)$ and $\Lambda_{ext}(t)$ are the expected value of $\lambda_{int}(t)$ and $\lambda_{ext}(t)$, thus the exposure distribution which represents the probability that node i has received n exposure by time t is as follows

$$P_{exp}^i(n; t) \approx \binom{t/dt}{n} \left(\frac{\Lambda_{int}^i(t) + \lambda_{ext}(t)}{t} \cdot dt \right)^n \times \left(1 - \frac{\Lambda_{int}^i(t) + \lambda_{ext}(t)}{t} \cdot dt \right)^{t/dt - n} \quad (23)$$

From Eq. (23), we know that $P_{exp}^i(n; t)$ considers both the contacts from the underlying network and the influence of external sources. Experiments on Twitter network show that the model can be used to infer the shape of the influence functions and to show the influence of external sources on information diffusion.

In addition, the authors [139] begin with showing that the external influence plays an important role in information diffusion process by analyzing eight typical events in *Sina Weibo*. Therefore, a theoretical model is proposed to incorporate both internal and external influence, which contains three properties, i.e., memory effects, role of spreaders and non-redundancy of contacts (Fig. 6). They find that both the mathematical model and simulation results are consistent with the empirical analysis, which reveals that the spreading pattern is determined by the event's characteristic.

The interplay between information diffusion and epidemic spreading. When a disease is spreading in the population, the information about this disease will be diffused simultaneously. Thus the individuals who know about the disease will change their behavior to avoid being infected or prevent infecting the others. Seminal studies considering information diffusion in the epidemic dynamical process are given by Funk et al. [98,235,236]. They aimed at building theoretical models incorporating these two spreading processes in a well-mixed population. Besides, they found that the information diffusion

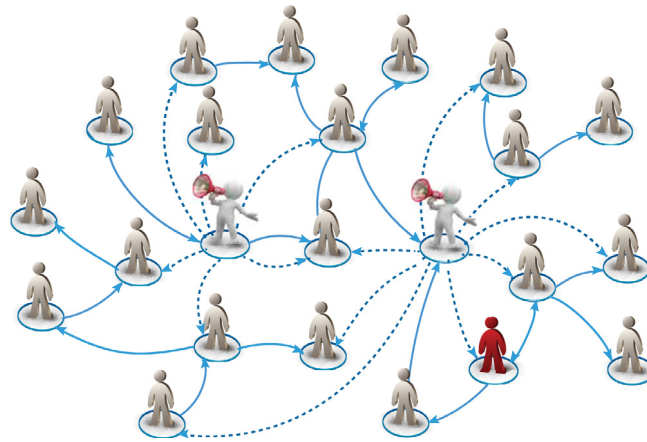


Fig. 6. Information diffusion model with internal and external influence. The individuals with loudspeakers are the media agents (external influence), the remaining are the common individuals in the population.
Source: From Ref. [139].

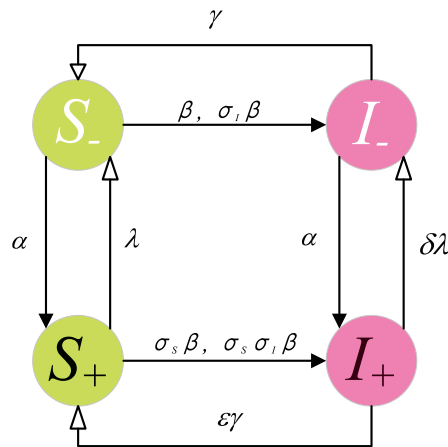


Fig. 7. Interplay between information diffusion and epidemic spreading. Four individual states are considered: (i) S_- : susceptible individuals who do not know about the disease; (ii) S_+ : susceptible individuals who know about the disease; I_- : infected individuals who do not know about the disease; I_+ : infected individuals who know about the disease.
Source: From Ref. [234].

process may affect the spreading of epidemics in three ways: (i) the individuals' disease state; (ii) the parameters used in the spreading process; (iii) the network structure. Whereas the mutual feedback phenomenon between information diffusion process and epidemic spreading process is empirically verified by analyzing the data of disease and disease information about *H7N9* and *Dengue* in [234]. They proposed a mathematical model using the pairwise approach to illustrate the interaction between the two spreading processes on *ER* network (Fig. 7), which shows a good agreement with the simulation results. In addition, they also proposed a model concerning these two processes on adaptive networks (Fig. 8), which shows that the adaptive behavior triggered by the information diffusion can significantly prevent the spread of disease spreading [237].

Besides the aforementioned models, there are also some other models describing information diffusion on social networks, such as strategic game models [238], voter models [239] and opinion formation models on adaptive networks [240,241]. These models allow us to have a better understanding and to further study the underlying dynamical process.

5. Applications studies

5.1. Network reconstruction

Reconstructing network structure from data, has become an important issue in recent network science research [242–248] and it is fundamental for us to understand and control collective dynamics in networks. Solutions to the problem lead to broad applications because the complex interacting dynamic exists in the systems of various disciplines [249–251]. There are some previous works [252–254] uncovered the nodal interaction patterns for coupled oscillator networks or

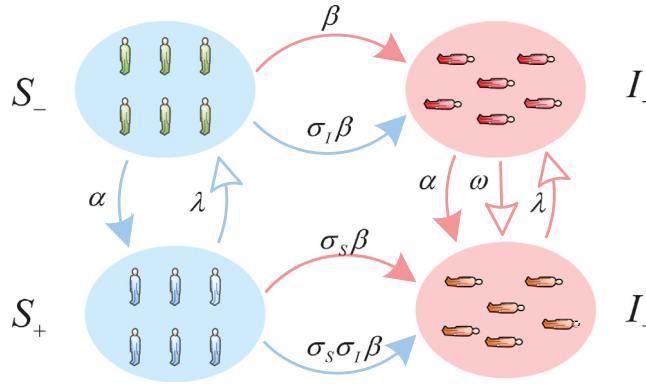


Fig. 8. Interaction between information and disease spreading process. The individuals are the same as Fig. 8, however, this model is performed on adaptive networks.
Source: From Ref. [139].

evolutionary games from time series data based on compressed sensing theory (CST) [255–258]. Different from those works, reconstructing the underlying contact network from the nodal states under the given entity propagation model, e.g., susceptible–infected–susceptible (SIS) [259], remains to be a great challenge since the dynamic of entity dissemination is highly stochastic [17]. It is difficult to cast the highly non-trivial transformation associated with the spreading dynamics into CST framework. To solve this challenge, Shen et al. [17] developed the CST framework from the time series data of nodal states of different entity spreading model, according to the law of large numbers. Then they solved the problem, i.e., reconstructing the underlying network, via using the standard solutions proposed by previous works [255–258]. We give a brief introduction about this work in this review.

5.1.1. Compressed sensing

The CST tries to reconstruct a vector $X \in R^n$ from linear responses Y of X with the following equation:

$$Y = \Phi \cdot X \quad (24)$$

where $Y \in R^m$ and Φ is an $m \times n$ matrix. X can be effectively reconstructed by solving the following convex-optimization problem:

$$\min \|X\|_1 \quad \text{s.t.} \quad Y = \Phi \cdot X \quad (25)$$

where $\|X\|_1 = \sum_{i=1}^n |X_i|$, is the L_1 form of X and Φ is restricted isometry. The key feature of CST is that the number of responses Y can be much less than the number of components of the unknown vector X , i.e., $m \ll n$, with the condition that X is sparse and the number of non-zero components in it is less than M . We need to form the problem of reconstructing propagation network as Eq. (24).

5.1.2. CST-based reconstruction

We consider the simple case that no hidden source exists and the propagation starts from a fraction of infected nodes, and give the CST-based reconstruction framework for SIS dynamics in this review (the case for CP model [260] can be found in [17]). For the SIS model, the state s_i of an arbitrary node i has only two possible cases:

$$s_i = \begin{cases} 0, & \text{susceptible} \\ 1, & \text{infected} \end{cases} \quad (26)$$

and i is infected by its neighbors at time t with probability $p_i^{01}(t)$:

$$p_i^{01}(t) = 1 - (1 - \alpha_i)^{\sum_{j=1, j \neq i}^n a_{ij} s_j(t)} \quad (27)$$

where α_i is the infected rate of node i and $s_j(t)$ is the state of node j at time t . The adjacent relationship between node i and j is denoted by a_{ij} ($a_{ij} = 1$ if i connects to j otherwise $a_{ij} = 0$). The superscript 01 represents the change from susceptible state to infected state. Similarly, the recovery probability of i from state 1 to 0 is denoted by $p_i^{10}(t)$: $p_i^{10}(t) = \beta_i$, where β_i is the recovery rate of i . We rewrite the log-form of Eq. (27) as:

$$\ln(1 - p_i^{01}(t)) = \ln(1 - \alpha_i) \sum_{j=1, j \neq i}^n a_{ij} s_j(t). \quad (28)$$

It can be written as standard CST form: $Y_{m \times 1} = \Phi_{m \times (n-1)} \cdot X_{(n-1) \times 1}$, where $Y_{m \times 1}$ equals $\ln(1 - p_i^{01}(t))$ at different times $t = t_1, t_2, \dots, t_m$, $\Phi_{m \times (n-1)}$ is determined by the state $s_j(t)$ of nodes except i at time t and $X_{(n-1) \times 1}$ contains the adjacent relationship between i and other nodes. With available nodal states information $\Phi_{m \times (n-1)}$, we can reconstruct adjacent relationship of nodes $X_{(n-1) \times 1}$ if we can infer the infection probabilities $p_i^{01}(t)$ at different time.

To estimate $p_i^{01}(t)$, firstly set a threshold Δ related to the normalized Hamming distance between strings composed of $s_j(t) (j \neq i)$ at different t to identify a base string at \hat{t}_λ and a set of strings subject to the base. Further obtain a set of base strings by setting another threshold θ associated with the normalized Hamming distance. Then the probability $\ln(1 - p_i^{01}(\hat{t}_\lambda))$ can be estimated by the average value over the state $s_i(t+1)$ at all suitable time. The process finally gives a set of reconstruction formulation in matrix form:

$$\begin{bmatrix} \ln(1 - \langle s_i(\hat{t}_1 + 1) \rangle) \\ \ln(1 - \langle s_i(\hat{t}_2 + 1) \rangle) \\ \vdots \\ \ln(1 - \langle s_i(\hat{t}_m + 1) \rangle) \end{bmatrix} = \begin{bmatrix} \langle s_1(\hat{t}_1) \rangle & \cdots & \langle s_{i-1}(\hat{t}_1) \rangle & \langle s_i(\hat{t}_1) \rangle & \cdots & \langle s_n(\hat{t}_1) \rangle \\ \langle s_1(\hat{t}_2) \rangle & \cdots & \langle s_{i-1}(\hat{t}_2) \rangle & \langle s_i(\hat{t}_2) \rangle & \cdots & \langle s_n(\hat{t}_2) \rangle \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \langle s_1(\hat{t}_m) \rangle & \cdots & \langle s_{i-1}(\hat{t}_m) \rangle & \langle s_i(\hat{t}_m) \rangle & \cdots & \langle s_n(\hat{t}_m) \rangle \end{bmatrix} \times \begin{bmatrix} \ln(1 - \alpha_i) a_{i,1} \\ \vdots \\ \ln(1 - \alpha_i) a_{i,i-1} \\ \ln(1 - \alpha_i) a_{i,i+1} \\ \vdots \\ \ln(1 - \alpha_i) a_{i,n} \end{bmatrix} \quad (29)$$

where $\hat{t}_1, \hat{t}_2, \dots, \hat{t}_m$ denote the time associated with m base strings and $\langle \cdot \rangle$ represents the average value of all satisfied t (refer to paper [17] for details of mathematical deduction). With the $Y_{m \times 1}$ and $\Phi_{m \times (n-1)}$ inferred from time series of nodal states, we can reconstruct all adjacent relationships between node i and other nodes by using standard solutions of CST to solve Eq. (29). Similarly, neighbors of all nodes can be uncovered, that is the full reconstruction of network can be achieved.

Besides CST-based methods, there are some other works [18,261,262] addressing the inverse problem of special topics of complex propagation networks. Specifically, the paths of diffusion from a single source constitute a tree-like structure if the propagation dynamics started from a single source. With the available information about the early stage of the dissemination process, it is feasible to infer all branches that reveal the connections from the source to neighbors as well as farther neighbors. In addition, the time delays in the diffusion process help infer the source of the propagation network through enumerating all possible hierarchical trees, which is another applicable topic different from constructing underlying contact network and has been studied by some works [18,19,263].

5.2. Rumor spreading

Rumor is a kind of harmful information that spreads in human societies in all times [264]. Recently, the development of online social networks (Twitter, Facebook, Digg etc.) further facilitates the spread of rumors in the population and makes the influence of rumors much wider than ever before.

The first rumor spreading model is proposed by Daley et al. [265] (known as the $D-K$ model), which is a variant of the SIR model. In $D-K$ model, individuals may be in one of the three states: ignorant (S , susceptible state in SIR), spreader (I , infected state), and stifter (R , recovered state). The ignorant individuals may be infected by the spreader individuals with some probability. The spreader individuals may turn to the stifter state with some probability when they encounter the stifter individuals or another spreader individual. Daley et al. [266] use stochastic process to analyze this model, in which they assume that the transition probability between different states obeys some mathematical distributions.

Suppose initially the number of individuals in state S, I, R are $S(0) = N, I(0) = 1$ and $R(0) = 0$ respectively. Thus, when $t \geq 0$, we have $S(t) + I(t) + R(t) = N + 1$. Transition probability ρ_{si} can be expressed as

$$\rho_{si}(t) = PS(t) = s, I(t) = i | S(0) = N, I(0) = 1. \quad (30)$$

The change of ρ_{si} with time is

$$\frac{d\rho_{si}}{dt} = \begin{cases} (s+1)(i-1)\rho_{s+1,i-1} + (N-s-i)(i+1)\rho_{s,i+1} \\ \quad + \frac{(s+1)(i+1)}{2}\rho_{s,i+1} - i\left(N - \frac{i-1}{2}\right)\rho_{si} & \text{if } 0 \leq s \leq N, 0 \leq i \leq N \\ 0 & \text{otherwise,} \end{cases} \quad (31)$$

$D-K$ model is reasonable under certain conditions despite it is not fully comply with the rumor spreading process in reality. According to $D-K$ model, some other modified models have also been proposed later on. For example, Maki et al. [267] consider the third process of the $D-K$ model as: when two spreader individuals encounter with each other, only one of them turns into stifter state.

As the structure of the population is heterogeneous, it is necessary to consider rumor spreading on networks. The first attempt to use complex networks to study the rumor dynamics is Zanette [268,269], when he investigated rumor dynamics on small-world networks. They used a simplified $D-K$ model, in which the ignorant individuals would become infected

when they encounter the spreader individuals, and the spreader individuals would turn to stifter state when they encounter stifter individuals or another spreader individual. A mean-field equation was given as follows

$$\begin{cases} \frac{ds(t)}{dt} = -s(t)i(t) \\ \frac{di(t)}{dt} = s(t)i(t) - i(t)(i(t) + r(t)) \\ \frac{dr(t)}{dt} = i(t)(i(t) + r(t)) \end{cases} \quad (32)$$

where $s(t)$, $i(t)$ and $r(t)$ are the fraction of ignorant, spreader and stifter individuals in the population. He found that a transition occurs at a critical value of the rewiring probability p between a regime where the rumor dies in a small neighborhood of its origin (small values of p), and a regime where it spreads over a finite fraction of the whole population (large values of p).

Numerous studies also considered rumor dynamics on scale-free networks [270,271]. In [270], they further developed $D - K$ model and used degree-based mean-field equation to describe rumor spreading on scale-free networks

$$\begin{cases} \frac{ds_k(t)}{dt} = -\beta k s_k(t) \sum_{k'} \frac{k' P(k') i_{k'}(t)}{\langle k \rangle} \\ \frac{di_k(t)}{dt} = \beta k s_k(t) \sum_{k'} \frac{k' P(k') i_{k'}(t)}{\langle k \rangle} - \gamma k i_k(t) \sum_{k'} \frac{k' P(k') (i_{k'}(t) + r_{k'}(t))}{\langle k \rangle} \\ \frac{dr_k(t)}{dt} = \gamma k i_k(t) \sum_{k'} \frac{k' P(k') (i_{k'}(t) + r_{k'}(t))}{\langle k \rangle} \end{cases} \quad (33)$$

where β and γ are the transmission and recovered probability, $s_k(t)$, $i_k(t)$ and $r_k(t)$ are the fraction of ignorant, spreader and stifter individuals with degree k . The final fraction of stifter individuals $r(\infty)$ and the cost of time are used as metrics to measure efficiency of the propagation. They found that $r(\infty)$ is related to transmission probability β . In addition, they also found that the value of $r(\infty)$ is larger in small-world networks compared to scale-free networks.

In addition to the models used above, rumor models are presented to control rumor spreading on social networks. For example, researchers proposed different models to model the spread of rumors and anti-rumors on the networks [272,264,273], in which the anti-rumors are used to combat rumor spreading. Furthermore, the role of influential spreaders in rumor spreading also attracted much attention recently [274,275], which is similar to the influential spreaders identification of other information and epidemic spreading process.

5.3. Leader identification

Considering the impact of information diffusion, messages from influential individuals [276], namely *leaders*, would be highly accepted and forwarded by their followers, thus can quickly spread to the whole society [277] according to small-world [278] and word-of-mouth [61] effects. Therefore, how to effectively identify those opinion leaders in large-scale networks brings a big challenge. In general, the problem of leader identification can be concluded as designing nodes ranking algorithms by denoting their respective importance. The present leader identification algorithms can be classified as (i) network-based; and (ii) machine-learning based approaches. The former focuses on using network topology [279] and dynamic models [280] to discover leaders, and the latter tends to extract useful features to find out target nodes [190]. Network topology, such as centrality [279], structural hole [281], k-shell [134], can statistically identify influential leaders with respect to their local roles in networks. Comparatively, dynamical models can make use of the information of global network structure. Pagerank based algorithms [282,280] hypothesizes that information flow can equally distribute along network edges, and the final equilibrium corresponds to ranking scores of every nodes. Computer scientists regard both static network topologies and dynamic flows as independent features, and optimize the objective function via iterative refinement [283]. Generally, methods based on network topologies are easy to implement and understand, models based on dynamical mechanisms can achieve robust results as more information is adopted. Feature extraction based approaches, as it further unifies both network and meta information (e.g., text, multi-layer network), can obtain more reliable results and be widely applied in engineering community. For more detailed work on leader identification, readers can find in the accompanying review article: **Vital nodes identification in complex networks**.

5.4. Dissemination management

The dynamic process on large network like retweets on Twitter [41,284] and H7N9 spreading [285] in a population are similar to entity dissemination. Given a network and a budget k , the key problem of dissemination management is how to quickly find the best k nodes to delete/immunize or best k edges to delete, for making the remaining network to be most

robust to the virus attack. Take the rumor control on Twitter as an application, we want to cut off the best k connections among user accounts to contain the rumor dissemination as much as possible. Similarly, given a population attacked by H7N9, which best k people should we immunize for minimizing the number of victims? In the following of this section, we divide the discussion of dissemination management problem into three parts: (1) Firstly quantify the robustness of the given network and find the epidemic threshold. In other words, how likely the given network will be attacked by a virus and what is the tipping point for virus to spread out? (2) Secondly derive the corresponding attacking value of selected k nodes of the given network and design node-based strategy for containing entity dissemination. (3) Thirdly derive the corresponding attacking value of a set of k edges of the given network and design edge-based strategy to minimize entity spreading.

5.4.1. Epidemic threshold

Given an undirected unweighted network with a virus propagation model, at which condition the virus will die out? The answer to the question is the epidemic threshold and it is the minimum level for preventing a viral propagation from dying out quickly [219,286,287]. For ease of presentation, the epidemic threshold is normalized as an effective strength s and the tipping point of propagation is reached when $s = 1$. The threshold theorem [20,288] proved that under threshold condition, i.e., $s < 1$, virus will be wiped out, and the effective strength is formulated as:

$$s = \lambda_1 \cdot c_{vp} \quad (34)$$

where λ_1 is the largest eigenvalue of adjacency matrix A of the network and c_{vp} is an explicit constant related with the given virus propagation model. Accordingly, the effective strength is a combination result of the particular underlying contact-network and the specific propagation model. For example, the largest eigenvalue of a clique like contact network is: $\lambda_1 = n - 1$, where n is the number of nodes. Thus the effective strength of popular susceptible/infected/susceptible (SIS) model [259] on such network is: $s = \lambda_1 \cdot \frac{\alpha}{\beta}$, where α and β denote the attack and healing probability over a contact-link, respectively. Likewise, the threshold theorem is applicable for all other standard virus propagation models [20,288].

According to the threshold theorem, the propagation is affected by the topology of contact-network and the property of the entity like virality or duration. In general, we assume that we cannot modify the entity's strength. Thus in order to manage the entity dissemination, i.e., control the value of s , we focus on manipulating the network connectivity, that is modifying λ_1 of the network.

5.4.2. Node-level strategy

Given an undirected unweighted network and budget k , which best k nodes should we delete/immunize to make the remaining network as robust as possible against virus propagation? The problem is closely related to the epidemic threshold theorem. Firstly, the "Vulnerability" [21] of the network G , i.e., $V(G)$, is measured by the largest eigenvalue of the adjacency matrix A ,

$$V(G) = \lambda_1. \quad (35)$$

The larger λ_1 is, the easier for virus propagation. Then in order to contain the virus propagation as much as possible, we should delete/immunize k nodes to decrease the $V(G)$ of network as much as possible since the drop of λ_1 leads to the decrement of effective strength. Specifically, the "Shield-Value" (SV) [21] quantifies the importance of a given k nodes set S in dissemination management:

$$SV(S) = \sum_{i \in S} 2\lambda_1 \mathbf{u}(i) - \sum_{i,j \in S} A(i,j) \mathbf{u}(i) \mathbf{u}(j) \quad (36)$$

where \mathbf{u} is the eigenvector corresponding to λ_1 . Obviously, nodes set S has larger $SV(S)$ if each of them has a large eigen-score ($\mathbf{u}(i)$) and they have few connections (most of $A(i,j)$ equal 0). In addition, $SV(S)$ is a good approximation of the eigen-drop of the largest eigenvalue, $\Delta\lambda_1(S)$, after deleting nodes set S :

$$\Delta\lambda_1(S) = SV(S) + \mathcal{O}\left(\sum_{j \in S} \|A(:,j)\|^2\right) \quad (37)$$

where $\Delta\lambda_1(S) = \lambda_1 - \hat{\lambda}_1$, $\hat{\lambda}_1$ is the largest eigenvalue for perturbed network after deleting nodes set S .

According to Eq. (37), it is natural for us to choose nodes set S that has the largest $SV(S)$ for reaching the largest $\Delta\lambda_1(S)$. An intuitive way is to compute the $SV(S)$ for each selection of k nodes and pick the set with the largest $SV(S)$. However, the high computational complexity $\mathcal{O}\left(\binom{k}{n} \cdot m\right)$ (n and m are the number of nodes and edges, respectively) makes the solution be not applicable for large network. To solve such problem, Tong et al. [21] proposed *NetShield* algorithm which reaches both efficacy and efficiency. We briefly illustrate the *NetShield* as:

Firstly initialize S as empty set and compute the largest eigenvalue λ_1 as well as the corresponding eigenvector \mathbf{u} of the given network, and compute the "Shield-Value" of each node i by: $SV(i) = (2\lambda_1 - A(i,i)) \cdot \mathbf{u}(i)^2$. Then greedily select one more node j ($j \notin S$) and add it into set S by: $j = \text{argmax}_{score}(i)$, where $score(j) = \mathbf{v}(i) \cdot 2 \cdot \mathbf{b}(j) \cdot \mathbf{u}(j)$ and $\mathbf{b} = A(:,S) \cdot \mathbf{u}(S)$. Repeat previous step and finally find the best k nodes for deletion/immunization. The good approximation of $SV(S)$ for

$\Delta\lambda_1(S)$ guarantees the effectiveness of *NetShield* and the greedy selection makes the computing complexity of *NetShield* become $\mathcal{O}(nk^2 + m)$, which is much smaller than $\mathcal{O}\left(\binom{k}{n} \cdot m\right)$.

Besides the *NetShield*, there are some other works of node-level strategy for dissemination management. Briesemeister et al. [289] focused on immunization of power law networks, Hayashi et al. [290] studied the growing network and derived the extinction conditions under random and targeted immunization for the *SHIR* (Susceptible, Hidden, Infectious, Recovered) model. Prakash et al. [291,292] proposed effective algorithms for node immunization on time-varying networks. Different from those works, the *NetShield* algorithm provides a general solution of node-level strategy for arbitrary static or time-varying networks.

5.4.3. Edge-level strategy

Given an undirected unweighted network and budget k , which best k edges should we delete to make the remaining network as robust as possible against virus propagation? Different from node-level strategy, we shift the problem to the level of edges here. Edge level strategy is more acceptable than node-level strategy in real application. Take rumor spreading in online social network as an example, deleting nodes means we need to shut down some legitimate user accounts while we can avoid this by deleting few connections among users. As we have seen in Sections 5.4.1 and 5.4.2, the propagation is affected by the “Vulnerability” of the network and the properties of the entity, and the “Vulnerability” can be represented by the largest eigenvalue λ_1 of the network. The problem we address is how we can contain the dissemination by deleting some edges of the network since we assume that we cannot change the properties of the propagating entity.

Similar to node-level strategy, the key issue to solve the problem is how to quantify the impact of deleting a set of edges S in terms of λ_1 . The naive way is to select k edges randomly and recompute λ_1 after deletion then find the best k edges for the largest $\Delta\lambda_1(S)$. However, it is computationally infeasible for large network since it takes $\mathcal{O}(m)$ time for each of $\binom{k}{m}$ possible selections. To solve this problem, Tong et al. [23] introduced eigen-score to quantify the impact of each edge in dissemination management and proposed *K-EdgeDeletion* algorithm for *NetMelt* problem. It reaches both computing efficiency and dissemination control efficacy. Specifically, let \mathbf{u} and \mathbf{v} be the corresponding left and right eigenvector of λ_1 , respectively. The $\Delta\lambda_1(S)$ after deleting a set of k edges S can be approximated by the following theorem:

$$\Delta\lambda_1(S) = c \cdot \sum_{e_{ij} \in S} \mathbf{u}(i)\mathbf{v}(j) + \mathcal{O}(k) \quad (38)$$

where $c = 1/\mathbf{u}^T\mathbf{v}$, and i and j denote the two ending nodes of deleted edge e_{ij} . Obviously, the eigen-score of edge e_{ij} is quantified by $\mathbf{u}(i)\mathbf{v}(j)$. Like *NetShield* algorithm, the *K-EdgeDeletion* algorithm greedily chooses the best k edges for deletion and we briefly illustrate it as:

Firstly initialize S as empty set and compute the largest eigenvalue λ_1 as well as corresponding left and right eigenvector \mathbf{u} and \mathbf{v} of the given network. If the minimum value of \mathbf{u} or \mathbf{v} is negative, we should revise \mathbf{u} or \mathbf{v} as $-\mathbf{u}$ or $-\mathbf{v}$. Then we compute the eigen-score of each edge e_{ij} by equation: $\text{score}(e_{ij}) = \mathbf{u}(i)\mathbf{v}(j)$. Finally find the best k edges with the top- k scores and add them to the deleted set S . The time cost of *K-EdgeDeletion* is $\mathcal{O}(mk + n)$ and it is much faster than random set selection. In addition, the good approximation of eigen-score value of deleted edges set makes the algorithm perform well in containing dissemination. On the contrary, the same approximation theorem can be applied to the *NetGel* problem which adds best k edges into the network so that its λ_1 will increase as much as possible, i.e., strengthening the entity propagation as much as possible. We recommend reader Ref. [23] which includes details of *K-EdgeAddition* algorithm for solving *NetGel* problem.

There are some other edge-level strategies for slowing down the entity spreading [293], reducing the average infection probability [294], evaluating the attack vulnerability [295] or studying contagion blocking in networks [296], etc. Similar to *K-EdgeDeletion* algorithm, Bishop et al. [297] proposed a convex optimization based solution to approximately minimize the largest eigenvalue of the network and Chan et al. [298] tracked multiple eigenvalues in order to measure network robustness. Long et al. [24] showed that *K-EdgeDeletion* performs poorly on networks with small eigen-gaps (like many social networks) and introduced *MET* (short for Multiple Eigenvalues Tracking) to overcome that problem. Zhang et al. [25] found the weakness of *K-EdgeDeletion* algorithm in maintaining community structure and proposed *CRlink* (short for Community Edge Relink) algorithm which performs well on both containing dissemination and maintaining community structure.

6. Outlook

In this section, we introduce a few challenges that the field of information diffusion to tackle in the future.

We live in the era of information, receive and send messages every moment. Besides traditional channels, e.g., TV and radio broadcasting, new ICT tools allow information to spread faster and wider than ever before, and dominate the way we obtain knowledge and form our opinions. We have to adapt to the ever-changing and challenging situations and fragment our time by hundreds of mobile applications (APPs). News-oriented applications can learn our personalized preferences and then push messages according to our historical information, which attract users via precise information diffusion. Social medias, tend to spread messages via the versatile social networking services, which employ the power of both the crowd and stars. Wearable devices collect our dynamical body information to infer the health status and share with sympathy friends. While there has been an undeniable progress in the information availability, a fundamental question remains elusive: how can

we get more useful information than before? Comparing with the great volume of input data, how to filter out information of less importance is a huge challenge. Recommender systems, a recently emergent technology, has been raised to solve the dilemma by collaborative filtering. However, the information diffusion is quite a dynamical and timely process (see the previous article [299]), classical recommender systems would fail to provide such real-time service in a detailed micro-level. Conversely, an effective yet efficient filtering algorithm would also promote the fast spreading of and suppress the diffusion of rumors or misinformation [96].

Although there is a vast class of researches studying the dynamics of information diffusion, most of them are based on classical epidemic spreading, which hypothesizes that contagion propagation can also lead to the patterns of information spreading. However, as information transmission requires much less cost and variate much faster than that of physical contagion with the help of modern internet tools, the modeling of information diffusion should also take those aspects into account. Generally, one will or will not forward a certain message are decided by many reasons, e.g. news content and timeliness, her/his preferences, or just random choice. If those factors are considered in modeling the corresponding dynamics, physics can thus prove more powerful in studying the driving universality of information spreading patterns.

Another interesting facet of information diffusion is related to its application in predicting various social phenomena. Firstly, the approach of information diffusion has been used to predict the financial turbulence and regional economical trends. The financial market can be partially predicted since people tends to accept messages which are confirmed by social environment and then invest according to what they have learned. Secondly, in traffic systems, we sometimes intend to imitate how others choose routes to walk or drive if some unknown accidents happen, as such explicit behaviors can easily spread to the waiting queue. Finally, the online information in social networks, saying awareness, can somehow reflect the situation of epidemic spreading, hence can be applied to model and predict the coupling dynamics between information and epidemics from both micro- and macro-perspectives. In a word, the incorporation of various kinds of diffusion data can largely enhance our ability of understanding the underlying mechanisms of corresponding occurrences.

The science of information diffusion is just starting—despite impressive progresses, much remains to be understood. For further advances intuition alone is no longer enough and a multidisciplinary approach will surely bring powerful tools that innovative matchmakers may turn the immense potential into real life applications.

Acknowledgments

We benefit a lot from the fruitful works collaborated with the following people, Luciano Pietronero, Matúš Medo, Pak Ming Hui, Wei Zeng, Chi Ho Yeung, etc. We also acknowledge the team lead by Linyuan Lü (some members are the authors of the accompanying review) for useful exchange of research. This work was partially supported by Natural Science Foundation of China (Grant Nos. 715222014, 11305043, 61503110, 71301165 and 61673151), Zhejiang Provincial Natural Science Foundation of China (Grant Nos. LY14A050001 and LQ16F030006), Swiss National Science Foundation (Grant No. 200020–143272), and the EU FP7 Grant 611272 (project GROWTHCOM).

References

- [1] L. Kovanen, K. Kaski, J. Kertész, J. Saramäki, Temporal motifs reveal homophily, gender specific patterns, and group talk in call sequences, *Proc. Natl. Acad. Sci. USA* 110 (45) (2013) 18070–18075.
- [2] J.L. Iribarren, E. Moro, Impact of human activity patterns on the dynamics of information diffusion, *Phys. Rev. Lett.* 103 (2009) 038702.
- [3] M. Karsai, M. Kivela, R.K. Pan, K. Kaski, J. Kertész, A.-L. Barabási, J. Saramäki, Small but slow world: how network topology and burstiness slow down spreading, *Phys. Rev. E* 83 (2011) 025102(R).
- [4] S.A. Myers, C.-G. Zhu, J. Leskovec, Information diffusion and external influence in networks, in: *Proceedings of the 18th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, ACM Press, New York, 2012, pp. 33–41.
- [5] S. Wu, J.M. Hofman, W.A. Mason, D.J. Watts, Who says what to whom on twitter, in: *Proceedings of the 20th ACM International Conference on World Wide Web*, ACM Press, New York, 2011, pp. 705–714.
- [6] S. Goel, D.J. Watts, D.G. Goldstein, The structure of online diffusion networks, in: *Proceedings of the 13rd ACM International Conference on Electronic Commerce*, ACM Press, New York, 2012, pp. 623–638.
- [7] R. Crane, D. Sornette, Robust dynamic classes revealed by measuring the response function of a social system, *Proc. Natl. Acad. Sci. USA* 105 (41) (2008) 15649–15653.
- [8] N. Hodas, K. Lerman, The simple rules of social contagion, *Sci. Rep.* 4 (2015) 4343.
- [9] D.J. Watts, A simple model of global cascades on random networks, *Proc. Natl. Acad. Sci. USA* 99 (9) (2002) 5766–5771.
- [10] P.L. Krapivsky, S. Redner, D. Volovik, Reinforcement-driven spread of innovations and fads, *J. Stat. Mech.* (2011) P12003.
- [11] M. Granovetter, Threshold models of collective behavior, *Am. J. Sociol.* 83 (6) (1978) 1420–1443.
- [12] A. Nematzadeh, E. Ferrara, A. Flammini, Y.-Y. Ahn, Optimal network modularity for information diffusion, *Phys. Rev. Lett.* 113 (8) (2014) 088701.
- [13] D. Trpevski, W.K. Tang, L. Kocarev, Model for rumor spreading over networks, *Phys. Rev. E* 81 (2010) 056102.
- [14] T. Gross, H. Sayama, *Adaptive Networks*, Springer, Berlin, Heidelberg, 2009.
- [15] T. Gross, B. Blasius, Adaptive coevolutionary networks: a review, *J. R. Soc. Interface* 5 (20) (2008) 259–271.
- [16] C. Nardini, B. Kozma, A. Barrat, Who's talking first? Consensus or lack thereof in coevolving opinion formation models, *Phys. Rev. Lett.* 100 (15) (2008) 158701.
- [17] Z. Shen, W.-X. Wang, Y. Fan, Z. Di, Y.-C. Lai, Reconstructing propagation networks with natural diversity and identifying hidden sources, *Nature Commun.* 5 (2014) 4323.
- [18] P.C. Pinto, P. Thiran, M. Vetterli, Locating the source of diffusion in large-scale networks, *Phys. Rev. Lett.* 109 (6) (2012) 068702.
- [19] Z. Shen, S. Cao, W.-X. Wang, Z. Di, H.E. Stanley, Locating the source of diffusion in complex networks by time-reversal backward spreading, *Phys. Rev. E* 93 (3) (2016) 032301.
- [20] D. Chakrabarti, Y. Wang, C. Wang, J. Leskovec, C. Faloutsos, Epidemic thresholds in real networks, *ACM Trans. Inf. Syst. Secur.* 10 (4) (2008) 1.
- [21] H. Tong, B.A. Prakash, C. Tsourakakis, T. Eliassi-Rad, C. Faloutsos, D.H. Chau, On the vulnerability of large graphs, in: *Proceedings of the 10th International Conference on Data Mining*, IEEE Press, 2010, pp. 1091–1096.

- [22] B.A. Prakash, L. Adamic, T.J. Iwashyna, H.H. Tong, F. Christos, Fractional immunization in networks, in: Proceedings of the 2013 SIAM International Conference on Data Mining, SIAM Press, 2013, pp. 659–667.
- [23] H. Tong, B.A. Prakash, T. Eliassi-Rad, M. Faloutsos, C. Faloutsos, Gelling, and melting, large graphs by edge manipulation, in: Proceedings of the 21st ACM International Conference on Information and Knowledge Management, ACM Press, New York, 2012, pp. 245–254.
- [24] L.T. Le, T. Eliassi-Rad, H. Tong, Met: A fast algorithm for minimizing propagation in large graphs with small eigen-gaps, in: Proceedings of the 2015 SIAM International Conference on Data Mining, SIAM Press, 2015, pp. 694–702.
- [25] C. Zhang, T. Eliassi-Rad, Minimizing dissemination in a population while maintaining its community structure, in: The SIGKDD Workshop on Population Informatics for Big Data, ACM Press, Sydney, 2015.
- [26] A.P. McAfee, Enterprise 2.0: The dawn of emergent collaboration, MIT Sloan Manage. Rev. 47 (3) (2006) 21–28.
- [27] M. Page, M. Molina, G. Jones, D. Makarov, The Mobile Economy, AT Kearny, 2016.
- [28] H. Zhao, Z. Dong, T. Li, X. Wang, C. Pang, Segmenting time series with connected lines under maximum error bound, Inform. Sci. 345 (2016) 1–8.
- [29] Q. Zhu, Q. Lin, Z. Du, Z. Liang, W. Wang, Z. Zhu, J. Chen, P. Huang, Z. Ming, A novel adaptive hybrid crossover operator for multiobjective evolutionary algorithm, Inform. Sci. 345 (2016) 177–198.
- [30] J.H. Sorensen, Hazard warning systems: Review of 20 years of progress, Nat. Hazards. Rev. 1 (2) (2000) 119–125.
- [31] S. Bellman, R.F. Potter, S. Treleven-Hassard, J.A. Robinson, D. Varan, The effectiveness of branded mobile phone apps, J. Interact. Mark. 25 (4) (2011) 191–200.
- [32] J.E. Katz, M. Aakhus, Perpetual Contact: Mobile Communication, Private Talk, Public Performance, Cambridge University Press, 2002.
- [33] J. Schiller, A. Voisard, Location-based Services, Morgan Kaufmann Publishers, 2004.
- [34] S. Dhar, U. Varshney, Challenges and business models for mobile location-based services and advertising, Commun. ACM 54 (5) (2011) 121–128.
- [35] M.-H. Park, J.-H. Hong, S.-B. Cho, Location-based recommendation system using bayesian user's preference model in mobile devices, in: Proceedings of the 4th International Conference on Ubiquitous Intelligence and Computing, Springer, 2007, pp. 1130–1139.
- [36] A. Küpper, Location-based Services: Fundamentals and Operation, John Wiley & Sons, 2005.
- [37] A. Chatfield, U. Brajawidagda, Twitter tsunami early warning network: a social network analysis of twitter information flows, in: Proceedings of the 23rd Australasian Conference on Information Systems, 2012, pp. 1–10.
- [38] Q. Lin, Z. Liu, Q. Yan, Z. Du, C.A.C. Coello, Z. Liang, W. Wang, J. Chen, Adaptive composite operator selection and parameter control for multiobjective evolutionary algorithm, Inform. Sci. 339 (2016) 332–352.
- [39] J. DiMicco, D.R. Millen, W. Geyer, C. Dugan, B. Brownholtz, M. Muller, Motivations for social networking at work, in: Proceedings of the 2008 ACM Conference on Computer Supported Cooperative Work, ACM Press, New York, 2008, pp. 711–720.
- [40] A. Acquisti, R. Gross, Imagined communities: Awareness, information sharing, and privacy on the facebook, in: Proceedings of the 6th International Conference on Privacy Enhancing Technologies, Springer, 2006, pp. 36–58.
- [41] H. Kwak, C. Lee, H. Park, S. Moon, What is twitter, a social network or a news media? in: Proceedings of the 19th International Conference on World Wide Web, ACM Press, New York, 2010, pp. 591–600.
- [42] D.C. Brabham, Crowdsourcing, MIT Press, 2013.
- [43] J. Howe, The rise of crowdsourcing, Wired Magazine 14 (6) (2006) 1–4.
- [44] J.C. Tang, M. Cebrian, N.A. Giacobe, H.-W. Kim, T. Kim, D.B. Wickert, Reflecting on the darpa red balloon challenge, Commun. ACM 54 (4) (2011) 78–85.
- [45] P. Meier, Crisis mapping in action: How open source software and global volunteer networks are changing the world, one map at a time, J. Map. Geogr. Libr. 8 (2) (2012) 89–100.
- [46] P. Belleflamme, T. Lambert, A. Schwenbacher, Crowdfunding: Tapping the right crowd, J. Bus. Venturing 29 (5) (2014) 585–609.
- [47] J. Leskovec, L.A. Adamic, B.A. Huberman, The dynamics of viral marketing, ACM. T. Web 1 (1) (2007) 5.
- [48] M. Richardson, P. Domingos, Mining knowledge-sharing sites for viral marketing, in: Proceedings of the 8th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, ACM Press, New York, 2002, pp. 61–70.
- [49] Z. Lai, W.K. Wong, Y. Xu, J. Yang, D. Zhang, Approximate orthogonal sparse embedding for dimensionality reduction, IEEE Trans. Neural Netw. Learn. Syst. 27 (4) (2016) 723–735.
- [50] S. Jurvetson, What exactly is viral marketing, Red Herring 78 (2000) 110–112.
- [51] D.J. Watts, J. Peretti, M. Frumin, Viral Marketing for the Real World, Harvard Business School Pub., 2007.
- [52] M. Vaidya, Ice bucket challenge cash may help derisk als drug research, Nat. Methods 20 (10) (2014) 1080.
- [53] A.M. Kaplan, M. Haenlein, Two hearts in three-quarter time: How to waltz the social media/viral marketing dance, Bus. Horiz. 54 (3) (2011) 253–263.
- [54] S. Fournier, J. Avery, The uninvited brand, Bus. Horiz. 54 (3) (2011) 193–207.
- [55] D.J. Watts, P.S. Dodds, Influentials, networks, and public opinion formation, J. Consum. Res. 34 (4) (2007) 441–458.
- [56] A. Tumasjan, T.O. Sprenger, P.G. Sandner, I.M. Welpe, Predicting elections with twitter: What 140 characters reveal about political sentiment, in: Proceedings of the Fourth International AAAI Conference on Web and Social Media, 2010, pp. 178–185.
- [57] T. Graham, M. Broersma, K. Hazelhoff, G. van't Haar, Between broadcasting political messages and interacting with voters: The use of twitter during the 2010 UK general election campaign, Inform. Commun. Soc. 16 (5) (2013) 692–716.
- [58] P.C. Tetlock, Giving content to investor sentiment: The role of media in the stock market, J. Financ. 62 (3) (2007) 1139–1168.
- [59] R. Mao, P. Zhang, X. Li, X. Liu, M. Lu, Pivot selection for metric-space indexing, Int. J. Mach. Learn. Cybern. 7 (2) (2016) 311–323.
- [60] J. Duda, P. Korus, N.J. Gadgil, K. Tahboub, E.J. Delp, Image-like 2d barcodes using generalizations of the Kuznetsov–Tsybakov problem, IEEE Trans. Inf. Forensics Secur. 11 (4) (2016) 691–703.
- [61] B.J. Jansen, M. Zhang, K. Sobel, A. Chowdury, Twitter power: Tweets as electronic word of mouth, J. Am. Soc. Inf. Sci. Tec. 60 (11) (2009) 2169–2188.
- [62] B. O'Connor, R. Balasubramanyam, B.R. Routledge, N.A. Smith, From tweets to polls: Linking text sentiment to public opinion time series, in: Proceedings of the 4th International AAAI Conference on Weblogs and Social Media, 2010, pp. 122–129.
- [63] R.S. Burt, Social contagion and innovation: Cohesion versus structural equivalence, Am. J. Sociol. 92 (6) (1987) 1287–1335.
- [64] J. Ugander, L. Backstrom, C. Marlow, J. Kleinberg, Structural diversity in social contagion, Proc. Natl. Acad. Sci. USA 109 (16) (2012) 5962–5966.
- [65] M. McPherson, L. Smith-Lovin, J.M. Cook, Birds of a feather: Homophily in social networks, Annu. Rev. Sociol. 27 (2001) 415–444.
- [66] D. Centola, The spread of behavior in an online social network experiment, Science 329 (2010) 1194–1197.
- [67] G.E. Kreindler, H.P. Young, Rapid innovation diffusion in social networks, Proc. Natl. Acad. Sci. USA 111 (Suppl. 3) (2014) 10881–10888.
- [68] J.S. Coleman, E. Katz, H. Menzel, Medical innovation: A diffusion study, Adm. Sci. Quart. 12 (2) (1967) 355–361.
- [69] E.M. Rogers, Diffusion of Innovations, The Free Press, New York, 2003.
- [70] S. Ciliberti, O.C. Martin, A. Wagner, Innovation and robustness in complex regulatory gene networks, Proc. Natl. Acad. Sci. USA 104 (34) (2007) 13591–13596.
- [71] W. Tang, H. Li, W. Luo, J. Huang, Adaptive steganalysis based on embedding probabilities of pixels, IEEE Trans. Inf. Forensics Secur. 11 (4) (2016) 734–745.
- [72] G. Wang, S. Zhang, K. Wu, Q. Zhang, L.M. Ni, Tim: Fine-grained rate adaptation in wlns, IEEE Trans. Mob. Comput. 15 (3) (2016) 748–761.
- [73] G. Wang, Y. Guan, Y. Wang, Z. Shao, Energy-aware assignment and scheduling for hybrid main memory in embedded systems, Computing 98 (3) (2016) 279–301.
- [74] N. Berger, C. Borgs, J.T. Chayes, A. Saberi, On the spread of viruses on the internet, in: Proceedings of the 16th Annual ACM-SIAM Symposium on Discrete Algorithms, Society for Industrial and Applied Mathematics, 2005, pp. 301–310.
- [75] P. Wang, M.C. González, C.A. Hidalgo, A.-L. Barabási, Understanding the spreading patterns of mobile phone viruses, Science 324 (5930) (2009) 1071–1076.
- [76] E. Serrano, C.A. Iglesias, M. Garijo, A survey of twitter rumor spreading simulations, in: Computational Collective Intelligence, Springer, 2015, pp. 113–122.

- [77] T. Holz, M. Steiner, F. Dahl, E. Biersack, F.C. Freiling, Measurements and mitigation of peer-to-peer-based botnets: A case study on storm worm, in: Proceedings of the 1st Usenix Workshop on Large-Scale Exploits and Emergent Threats, no. 9, 2008.
- [78] B. Doerr, M. Fouz, T. Friedrich, Why rumors spread so quickly in social networks, *Commun. ACM* 55 (6) (2012) 70–75.
- [79] O. Oh, K.H. Kwon, H.R. Rao, An exploration of social media in extreme events: Rumor theory and twitter during the haiti earthquake 2010, in: Proceedings of the 31st International Conference on Information Systems, 2010, pp. 1–13.
- [80] Y. Tanaka, Y. Sakamoto, T. Matsuka, Transmission of rumor and criticism in twitter after the great japan earthquake, in: Annual Meeting of the Cognitive Science Society, 2012, p. 2387.
- [81] J. Leskovec, L. Backstrom, J. Kleinberg, Meme-tracking and the dynamics of the news cycle, in: Proceedings of the 15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, ACM Press, New York, 2009, pp. 497–506.
- [82] Y. Matsubara, Y. Sakurai, B.A. Prakash, L. Li, C. Faloutsos, Rise and fall patterns of information diffusion: Model and implications, in: Proceedings of the 18th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, ACM Press, New York, 2012, pp. 6–14.
- [83] J. Yang, J. Leskovec, Patterns of temporal variation in online media, in: Proceedings of the 4th ACM International Conference on Web Search and Data Mining, ACM Press, New York, 2011, pp. 177–186.
- [84] X. Zhang, C. Xu, X. Sun, G. Baci, Schatten-q regularizer constrained low rank subspace clustering model, *Neurocomputing* 182 (2016) 36–47.
- [85] G.L. Ciampaglia, A. Flammini, F. Menczer, The production of information in the attention economy, *Sci. Rep.* 5 (2015) 9452.
- [86] Y. Sano, K. Yamada, H. Watanabe, H. Takayasu, M. Takayasu, Empirical analysis of collective human behavior for extraordinary events in the blogosphere, *Phys. Rev. E* 87 (2013) 012805.
- [87] J. Lehmann, B. Goncalves, J.J. Ramasco, C. Cattuto, Dynamical classes of collective attention in twitter, in: Proceedings of the 21st ACM International Conference on World Wide Web, ACM Press, New York, 2012, pp. 251–260.
- [88] P. Zhang, X.-F. Wang, B.-X. Li, On predicting Twitter trend: factors and models, in: Proceedings of the 2013 ACM International Conference on Advances in Social Networks Analysis and Mining, ACM Press, New York, 2013, pp. 1427–1429.
- [89] J.C.-W. Lin, W. Gan, P. Fournier-Viger, T.-P. Hong, V.S. Tseng, Efficient algorithms for mining high-utility itemsets in uncertain databases, *Knowl.-Based Syst.* 96 (2016) 171–187.
- [90] Q.F. Zhou, W.H. Mow, S. Zhang, D. Toupakaris, Two-way decode-and-forward for low-complexity wireless relaying: Selective forwarding versus one-bit soft forwarding, *IEEE Trans. Wireless Commun.* 15 (3) (2016) 1866–1880.
- [91] S.-Z. Wang, Z. Yan, X. Hu, P.S. Yu, Z.-J. Li, Burst time prediction in cascades, in: Proceedings of the 29th AAAI Conference on Artificial Intelligence, AAAI Press, 2015, pp. 325–331.
- [92] J.-T. Cheng, L. Adamic, P.A. Dow, J.M. Kleinberg, J. Leskovec, Can cascades be predicted? in: Proceedings of the 23rd ACM International Conference on World Wide Web, ACM Press, New York, 2014, pp. 925–936.
- [93] L.-Y. Lü, D.-B. Chen, T. Zhou, The small world yields the most effective information spreading, *New J. Phys.* 13 (2011) 123005.
- [94] Q.-Y. Zhao, M.A. Erdogdu, H.Y. He, A. Rajaraman, J. Leskovec, SEISMIC: a self-exciting point process model for predicting tweet popularity, in: Proceedings of the 21st ACM International Conference on Knowledge Discovery and Data Mining, ACM Press, New York, 2015, pp. 1513–1522.
- [95] D.M. Romero, B. Beeder, J. Kleinberg, Differences in the mechanics of information diffusion across topics: idioms, political hashtags, and complex contagion on twitter, in: Proceedings of the 20th ACM International Conference on World Wide Web, ACM Press, New York, 2011, pp. 695–704.
- [96] M.D. Vicario, A. Bessi, F. Zollo, F. Petroni, A. Scala, G. Caldarelli, H.E. Stanley, W. Quattrociocchi, The spreading of misinformation online, *Proc. Natl. Acad. Sci. USA* 113 (3) (2016) 554–559.
- [97] M. Coscia, Average is boring: how similarity kills a meme's success, *Sci. Rep.* 4 (2014) 6477.
- [98] S. Funk, E. Gilad, C. Watkins, V.A.A. Jansen, The spread of awareness and its impact on epidemic outbreaks, *Proc. Natl. Acad. Sci. USA* 106 (16) (2009) 6872–6877.
- [99] L. Cui, G. Li, Q. Lin, J. Chen, N. Lu, Adaptive differential evolution algorithm with novel mutation strategies in multiple sub-populations, *Comput. Oper. Res.* 67 (2016) 155–173.
- [100] W. Pan, S. Xia, Z. Liu, X. Peng, Z. Ming, Mixed factorization for collaborative recommendation with heterogeneous explicit feedbacks, *Inform. Sci.* 332 (2016) 84–93.
- [101] Z. Ma, A. Sun, G. Cong, Will this # hashtag be popular tomorrow? in: Proceedings of the 35th ACM International Conference on Research and Development in Information Retrieval, ACM Press, New York, 2012, pp. 1173–1174.
- [102] G. Zhou, L. Yu, C.-X. Zhang, C. Liu, Z.-K. Zhang, J.-L. Zhang, A novel approach for generating personalized mention list on micro-blogging system, in: Proceedings of the 15th IEEE International Conference on Data Mining Workshops, IEEE Press, 2015, pp. 1368–1374.
- [103] K.-L. Chen, T.-Q. Chen, G.-Q. Zheng, O. Jin, E.-P. Yao, Y. Yu, Collaborative personalized tweet recommendation, in: Proceedings of the 35th ACM International Conference on Research and Development in Information Retrieval, ACM Press, New York, 2012, pp. 661–670.
- [104] C.M. Alis, M.T. Lim, H.S. Moat, D. Barchiesi, T. Preis, S.R. Bishop, Quantifying regional differences in the length of twitter messages, *PLoS ONE* 10 (4) (2015) e0122278.
- [105] S. Kwon, M. Cha, Modeling bursty temporal pattern of rumors, in: Proceedings of the 8th International AAAI Conference on Weblogs and Social Media, AAAI Press, 2014, pp. 650–651.
- [106] S. Kwon, M. Cha, K. Jung, W. Chen, Y. Wang, Prominent features of rumor propagation in online social media, in: Proceedings of the 13th International Conference on Data Mining, IEEE Press, 2013, pp. 1103–1108.
- [107] J. Ma, W. Gao, Z.-Y. Wei, Y.-M. Lu, K.-F. Wong, Detect rumors using time series of social context information on microblogging websites, in: Proceedings of the 24th ACM International Conference on Information and Knowledge Management, ACM Press, New York, 2015, pp. 1751–1754.
- [108] D. Centola, An experimental study of homophily in the adoption of health behavior, *Science* 334 (2011) 1269–1272.
- [109] E. Bakshy, I. Rosenn, C. Marlow, L. Adamic, The role of social networks in information diffusion, in: Proceedings of the 21st ACM International Conference on World Wide Web, ACM Press, New York, 2012, pp. 519–528.
- [110] S. Ardon, A. Bagchi, A. Mahanti, A. Ruhela, A. Seth, R.M. Tripathy, S. Triukose, Spatio-temporal and events based analysis of topic popularity in twitter, in: Proceedings of the 22nd ACM International Conference on Information and Knowledge Management, ACM Press, New York, 2013, pp. 219–228.
- [111] K. Lerman, R. Ghosh, Information contagion: an empirical study of the spread of news on digg and twitter social networks, in: Proceedings of the 4th AAAI International Conference on Weblogs and Social Media, AAAI Press, 2010, pp. 90–97.
- [112] V. Arnaboldi, M. Conti, M. Gala, A. Passarella, F. Pezzoni, Ego network structure in online social networks and its impact on information diffusion, *Comput. Commun.* 76 (2016) 26–41.
- [113] J. Zhang, B. Liu, J. Tang, T. Chen, J.-Z. Li, Social influence locality for modeling retweeting behaviors, in: Proceedings of the 23rd International Joint Conference on Artificial Intelligence, AAAI Press, 2013, pp. 2761–2767.
- [114] M.S. Granovetter, The strength of weak ties, *Am. J. Sociol.* 78 (1973) 1360–1380.
- [115] J.-P. Onnela, J. Saramäki, J. Hyvönen, G. Szabó, D. Lazer, K. Kaski, J. Kertész, A.-L. Barabási, Structure and tie strengths in mobile communication networks, *Proc. Natl. Acad. Sci. USA* 104 (18) (2007) 7332–7336.
- [116] E. Bakshy, S. Messing, L.A. Adamic, Exposure to ideologically diverse news and opinion on facebook, *Science* 348 (2015) 1130–1132.
- [117] P. Holme, J. Saramäki, Temporal networks, *Phys. Rep.* 519 (2012) 97–125.
- [118] S. Myers, J. Leskovec, The bursty dynamics of the twitter information network, in: Proceedings of the 23rd ACM International Conference on World Wide Web, ACM Press, New York, 2014, pp. 913–924.
- [119] R. Pastor-Satorras, C. Castellano, P. Van Mieghem, A. Vespignani, Epidemic processes in complex networks, *Rev. Modern Phys.* 87 (3) (2015) 925–979.
- [120] S. Goel, A. Anderson, J. Hofman, D.J. Watts, The structural virality of online diffusion, *Manag. Sci.* 62 (1) (2016) 180–196.
- [121] J. Leskovec, L.A. Adamic, B.A. Huberman, The dynamics of viral marketing, *ACM T. WEB* 1 (1) (2007) 5.
- [122] M.E.J. Newman, Spread of epidemic disease on networks, *Phys. Rev. E* 66 (2002) 016128.

- [123] J. Goldenberg, B. Libai, E. Muller, Talk of the network: A complex systems look at the underlying process of word-of-mouth, *Mark. Lett.* 12 (3) (2001) 211–223.
- [124] P. Bao, H.-W. Shen, W. Chen, X.-Q. Cheng, Cumulative effect in information diffusion: empirical study on a microblogging network, *PLoS ONE* 8 (10) (2013) e76027.
- [125] S. Pei, L. Muchnik, S.-T. Tang, Z.-M. Zheng, H.A. Makse, Exploring the complex pattern of information spreading in online blog communities, *PLoS ONE* 10 (2015) e0126894.
- [126] V. Gómez, H. Kappen, A. Kaltenbrunner, Modeling the structure and evolution of discussion cascades, in: *Proceedings of the 22nd ACM International Conference on Hypertext and Hypermedia*, ACM Press, New York, 2011, pp. 181–190.
- [127] D.-S. Wang, Z. Wen, H.-H. Tong, C.-Y. Lin, C.-M. Song, A.-L. Barabási, Information spreading in context, in: *Proceedings of the 20th ACM International Conference on World Wide Web*, ACM Press, New York, 2011, pp. 735–744.
- [128] S. Bi, R. Zhang, Placement optimization of energy and information access points in wireless powered communication networks, *IEEE Trans. Wirel. Commun.* 15 (3) (2016) 2351–2364.
- [129] C. Guo, B. Liao, L. Huang, X. Lin, J. Zhang, On convexity of fairness-aware energy-efficient power allocation in spectrum-sharing networks, *IEEE Commun. Lett.* 20 (3) (2016) 534–537.
- [130] C. Guo, B. Liao, L. Huang, Q. Li, X. Lin, Convexity of fairness-aware resource allocation in wireless powered communication networks, *IEEE Commun. Lett.* 20 (3) (2016) 474–477.
- [131] A. Anderson, D. Huttenlocher, J. Kleinberg, J. Leskovec, M. Tiwari, Global diffusion via cascading invitations: structure, growth, and homophily, in: *Proceedings of the 24th ACM International Conference on World Wide Web*, ACM Press, New York, 2015, pp. 66–76.
- [132] M. Gomez-Rodriguez, D. Balduzzi, B. Schölkopf, Uncovering the temporal dynamics of diffusion networks, in: *Proceedings of the 28th ACM International Conference on Machine Learning*, ACM Press, New York, 2011, pp. 561–568.
- [133] R. Cohen, K. Erez, D. ben Avraham, S. Havlin, Breakdown of the internet under intentional attack, *Phys. Rev. Lett.* 86 (2001) 3682–3685.
- [134] M. Kitsak, L.K. Gallos, S. Havlin, F. Liljeros, L. Muchnik, H.E. Stanley, H.A. Makse, Identification of influential spreaders in complex networks, *Nat. Phys.* 6 (2010) 888–893.
- [135] S. Pei, L. Muchnik, J.S.J. Andrade, Z.-M. Zheng, H.A. Makse, Searching for superspreaders of information in real-world social media, *Sci. Rep.* 4 (2014) 5547.
- [136] L. Lü, T. Zhou, Q.-M. Zhang, H.E. Stanley, The H-index of a network node and its relation to degree and coreness, *Nature Commun.* 6 (2016) 10168.
- [137] F. Morone, H.A. Makse, Influence maximization in complex networks through optimal percolation, *Nature* 524 (2015) 65–68.
- [138] H. Kim, K. Beznosov, E. Yoneki, Finding influential neighbors to maximize information diffusion in twitter, in: *Proceedings of the 23rd ACM International Conference on World Wide Web*, ACM Press, New York, 2014, pp. 701–706.
- [139] C. Liu, X.-X. Zhan, Z.-K. Zhang, G.-Q. Sun, P.M. Hui, How events determine spreading patterns: information transmission via internal and external influences on social networks, *New J. Phys.* 7 (2015) 113045.
- [140] J. Yang, S. Counts, Predicting the speed, scale and range of information diffusion in twitter, in: *Proceedings of the 4th AAAI International Conference on Weblogs and Social Media*, AAAI Press, 2010, pp. 355–358.
- [141] M. Karsai, G. Iñiguez, K. Kaski, J. Kertész, Complex contagion process in spreading of online innovation, *J. R. Soc. Interface* 11 (2014) 20140694.
- [142] A.D.I. Kramer, J.E. Guillory, J.T. Hancock, Experimental evidence of massive-scale emotional contagion through social networks, *Proc. Natl. Acad. Sci. USA* 111 (24) (2014) 8788–8790.
- [143] A.-L. Barabási, The origin of bursts and heavy tails in human dynamic, *Nature* 435 (2005) 207–211.
- [144] Y. Wu, C.-S. Zhou, J.-H. Xiao, J. Kurths, H.J. Schellnhuber, Evidence for a bimodal distribution in human communication, *Proc. Natl. Acad. Sci. USA* 107 (44) (2010) 18803–18808.
- [145] Z.-Q. Jiang, W.-J. Xie, M.-X. Li, B. Podobnik, W.-X. Zhou, H.E. Stanley, Calling patterns in human communication dynamics, *Proc. Natl. Acad. Sci. USA* 110 (5) (2013) 1600–1605.
- [146] J. Ratkiewicz, S. Fortunato, A. Flammini, F. Menczer, A. Vespignani, Characterizing and modeling the dynamics of online popularity, *Phys. Rev. Lett.* 105 (2010) 158701.
- [147] A. Vazquez, B. Rácz, A. Lukács, A.-L. Barabási, Impact of non-Poissonian activity patterns on spreading processes, *Phys. Rev. Lett.* 98 (2007) 158702.
- [148] B. Min, K.-I. Goh, A. Vazquez, Spreading dynamics following bursty human activity patterns, *Phys. Rev. E* 83 (2011) 036102.
- [149] J.L. Iribarren, E. Moro, Branching dynamics of viral information spreading, *Phys. Rev. E* 84 (2011) 046116.
- [150] V.-P. Backlund, J. Saramäki, R.K. Pan, Effects of temporal correlations on cascades: threshold models on temporal networks, *Phys. Rev. E* 89 (2014) 062815.
- [151] N. Masuda, K. Klemm, V.M. Eguíluz, Temporal networks: slowing down diffusion by long lasting interactions, *Phys. Rev. Lett.* 111 (2013) 188701.
- [152] L.E.C. Rocha, F. Liljeros, P. Holme, Simulated epidemics in an empirical spatiotemporal network of 50185 sexual contacts, *PLoS Comput. Biol.* 7 (3) (2011) e1001109.
- [153] L.E.C. Rocha, V.D. Blondel, Bursts of vertex activation and epidemics in evolving networks, *PLoS Comput. Biol.* 9 (3) (2013) e1002974.
- [154] H.-H. Jo, J.I. Perotti, K. Kaski, J. Kertész, Analytically solvable model of spreading dynamics with non-Poissonian processes, *Phys. Rev. X* 4 (2014) 011041.
- [155] D.X. Horváth, J. Kertész, Spreading dynamics on networks: the role of burstiness, topology and non-stationarity, *New J. Phys.* 16 (7) (2014) 073037.
- [156] I. Scholtes, N. Wider, R. Pfizner, A. Garas, C.J. Tessone, F. Schweitzer, Causality-driven slow-down and speed-up of diffusion in non-Markovian temporal networks, *Nature Commun.* 5 (2014) 5024.
- [157] S. Gao, J. Ma, Z.-M. Chen, Modeling and predicting retweeting dynamics on microblogging platforms, in: *Proceedings of the 8th ACM International Conference on Web Search and Data Mining*, ACM, New York, 2015, pp. 107–116.
- [158] S.A. Golder, M.W. Macy, Diurnal and seasonal mood vary with work, sleep, and daylight across diverse cultures, *Science* 333 (2011) 1878–1881.
- [159] R.D. Malmgren, D.B. Stouffer, A.S.L.O. Campanharo, L.A. Amaral, On universality in human correspondence activity, *Science* 325 (2009) 1696–1700.
- [160] H.-H. Jo, M. Karsai, J. Kertész, K. Kaski, Circadian pattern and burstiness in human communication activity, *New J. Phys.* 14 (2012) 013055.
- [161] S.N. Dorogovtsev, A.V. Goltsev, J.F. Mendes, Critical phenomena in complex networks, *Rev. Modern Phys.* 80 (2008) 1275–1335.
- [162] S.V. Buldyrev, R. Parshani, G. Paul, H.E. Stanley, S. Havlin, Catastrophic cascade of failures in interdependent networks, *Nature* 464 (2010) 1025–1028.
- [163] D.-Q. Li, B.-W. Fu, Y.-P. Wang, G.-Q. Lu, Y. Berezin, H.E. Stanley, S. Havlin, Percolation transition in dynamical traffic network with evolving critical bottlenecks, *Proc. Natl. Acad. Sci. USA* 112 (3) (2015) 669–672.
- [164] R. Parshani, S.V. Buldyrev, S. Havlin, Critical effect of dependency groups on the function of networks, *Proc. Natl. Acad. Sci. USA* 108 (3) (2011) 1007–1010.
- [165] C.D. Brummitt, R.M. D'Souza, E.A. Leicht, Suppressing cascades of load in interdependent networks, *Proc. Natl. Acad. Sci. USA* 109 (12) (2012) 680–689.
- [166] L.-L. Weng, A. Flammini, A. Vespignani, F. Menczer, Competition among memes in a world with limited attention, *Sci. Rep.* 2 (2012) 335.
- [167] J.P. Gleeson, J.A. Ward, K.P. O'Sullivan, W.T. Lee, Competition-induced criticality in a model of meme popularity, *Phys. Rev. Lett.* 112 (2014) 048701.
- [168] Y. Yuan, B. Li, M.Q.-H. Meng, Bleeding frame and region detection in the wireless capsule endoscopy video, *IEEE J. Biomed. Health Inform.* 20 (2) (2016) 624–630.
- [169] H. Yuan, S. Kwong, X. Wang, Y. Zhang, F. Li, A virtual view psnr estimation method for 3-d videos, *IEEE Trans. Broadcast.* 62 (1) (2016) 134–140.
- [170] M. Kosinski, D. Stillwell, T. Graepel, Private traits and attributes are predictable from digital records of human behavior, *Proc. Natl. Acad. Sci. USA* 110 (15) (2013) 5802–5805.
- [171] P. Bhattacharya, M.B. Zafar, N. Ganguly, S. Ghosh, K.P. Gummadi, Inferring user interests in the twitter social network, in: *Proceedings of the 8th ACM Conference on Recommender Systems*, ACM Press, New York, 2014, pp. 357–360.
- [172] S.M. Mohammad, S. Kiritchenko, Using hashtags to capture fine emotion categories from tweets, *Comput. Intell.* 31 (2) (2015) 301–326.
- [173] L.-L. Weng, F. Menczer, Topicality and impact in social media: diverse message, focused messages, *PLoS ONE* 10 (2) (2015) e0118410.

- [174] J. Zhang, J. Tang, J.-Z. Li, Y. Liu, C.-X. Xing, Who influenced you? Predicting retweet via social influence locality, *ACM Trans. Knowl. Data Eng.* D 9 (3) (2015) 25.
- [175] J.M. Miotto, E.G. Altmann, Predictability of extreme events in social media, *PLoS ONE* 9 (2014) e111506.
- [176] M.J. Salganik, P.S. Dodds, D.J. Watts, Experimental study of inequality and unpredictability in an artificial cultural market, *Science* 311 (2006) 854–856.
- [177] P. Holme, T. Takaguchi, Time evolution of predictability of epidemics on networks, *Phys. Rev. E* 91 (4) (2015) 042811.
- [178] P. Holme, Information content of contact-pattern representations and predictability of epidemic outbreaks, *Sci. Rep.* 5 (2015) 14462.
- [179] P. Cui, S.-F. Jin, L.-Y. Yu, F. Wang, W.-W. Zhu, S.-Q. Yang, Cascading outbreak prediction in networks: A data-driven approach, in: *Proceedings of the 19th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, ACM Press, New York, 2013, pp. 901–909.
- [180] S.-Z. Wang, Z. Yan, X. Hu, P.S. Yu, Z.-J. Li, B. Wang, CPB: a classification-based approach for burst time prediction in cascades, *Knowl. Inf. Syst.* (2015) 1–29.
- [181] D.-B. Chen, R. Xiao, A. Zeng, Predicting the evolution of spreading on complex networks, *Sci. Rep.* 4 (2014) 6108.
- [182] M. Gomez-Rodriguez, J. Leskovec, D. Balduzzi, B. Schölkopf, Uncovering the structure and temporal dynamics of information propagation, *Netw. Sci.* 2 (2014) 26–65.
- [183] D.B. Kurka, A. Godoy, F.J.V. Zuben, Online social network analysis: a survey of reaserch applications in computer science, 2015, arXiv:1504.05655.
- [184] Y. Li, W. Liu, Q. Huang, Traffic anomaly detection based on image descriptor in videos, *Multimedia Tools Appl.* 75 (5) (2016) 2487–2505.
- [185] S.-H. Yang, H. Zha, Mixture of mutually exciting processes for viral diffusion, in: *Proceedings of the 30th ACM International Conference on Machine Learning*, ACM Press, New York, 2013, pp. 1–9.
- [186] D. Centola, V.M. Eguíluz, M.W. Macy, Cascade dynamics of complex propagation, *Physica A* 374 (1) (2007) 449–456.
- [187] J.P. Gleeson, D.J. Cahalane, Seed size strongly affects cascades on random networks, *Phys. Rev. E* 75 (5) (2007) 056103.
- [188] P. Singh, S. Sreenivasan, B.K. Szymanski, G. Korniss, Threshold-limited spreading in social networks with multiple initiators, *Sci. Rep.* 3 (2013) 2330.
- [189] D.E. Whitney, Dynamic theory of cascades on finite clustered random networks with a threshold rule, *Phys. Rev. E* 82 (6) (2010) 066110.
- [190] D. Kempe, J. Kleinberg, É Tardos, Maximizing the spread of influence through a social network, in: *Proceedings of the 9th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, ACM Press, New York, 2003, pp. 137–146.
- [191] L. Weng, F. Menczer, Y.-Y. Ahn, Virality prediction and community structure in social networks, *Sci. Rep.* 3 (2013) 2522.
- [192] P.S. Dodds, D.J. Watts, Universal behavior in a generalized model of contagion, *Phys. Rev. Lett.* 92 (21) (2004) 218701.
- [193] P.S. Dodds, D.J. Watts, A generalized model of social and biological contagion, *J. Theoret. Biol.* 232 (4) (2005) 587–604.
- [194] C.D. Brummitt, K.-M. Lee, K.-I. Goh, Multiplexity-facilitated cascades in networks, *Phys. Rev. E* 85 (4) (2012) 045102.
- [195] O. Yağan, F. Gligor, Analysis of complex contagions in random multiplex networks, *Phys. Rev. E* 86 (3) (2012) 036103.
- [196] T. Takaguchi, N. Masuda, P. Holme, Bursty communication patterns facilitate spreading in a threshold-based epidemic dynamics, *PLoS ONE* 8 (7) (2013) e68629.
- [197] F. Karimi, P. Holme, Threshold model of cascades in empirical temporal networks, *Physica A* 392 (16) (2013) 3476–3483.
- [198] W. Wang, M. Tang, H.-F. Zhang, Y.-C. Lai, Dynamics of social contagions with memory of nonredundant information, *Phys. Rev. E* 92 (1) (2015) 012820.
- [199] W. Wang, M. Tang, P. Shu, Z. Wang, Dynamics of social contagions with heterogeneous adoption thresholds: crossover phenomena in phase transition, *New J. Phys.* 18 (1) (2016) 013029.
- [200] T. Liggett, *Interacting Particle Systems*, Springer Science & Business Media, 2012.
- [201] J. Goldenberg, B. Libai, E. Muller, Using complex systems analysis to advance marketing theory development: Modeling heterogeneity effects on new product growth through stochastic cellular automata, *Acad. Market. Sci. Rev.* 2001 (2001) 9.
- [202] K. Saito, R. Nakano, M. Kimura, Prediction of information diffusion probabilities for independent cascade model, in: *Proceedings of the 12th International Conference on Knowledge-Based Intelligent Information and Engineering Systems*, Springer, 2008, pp. 67–75.
- [203] L. Dickens, I. Molloy, J. Lobo, P.-C. Cheng, A. Russo, Learning stochastic models of information flow, in: *Proceedings of 28th IEEE International Conference on Data Engineering*, IEEE Press, 2012, pp. 570–581.
- [204] M. Kimura, K. Saito, R. Nakano, Extracting influential nodes for information diffusion on a social network, in: *Proceedings of the 22nd AAAI International Conference on Artificial Intelligence*, AAAI Press, 2007, pp. 1371–1376.
- [205] M. Kimura, K. Saito, R. Nakano, H. Motoda, Finding influential nodes in a social network from information diffusion data, in: *Social Computing and Behavioral Modeling*, Springer, 2009, pp. 1–8.
- [206] M. Kimura, K. Saito, H. Motoda, Minimizing the spread of contamination by blocking links in a network, in: *Proceedings of 23rd AAAI Conference on Artificial Intelligence*, AAAI Press, 2008, pp. 1175–1180.
- [207] R. Albert, H. Jeong, A.-L. Barabási, Error and attack tolerance of complex networks, *Nature* 406 (6794) (2000) 378–382.
- [208] A. Broder, R. Kumar, F. Maghoul, P. Raghavan, S. Rajagopalan, R. Stata, A. Tomkins, J. Wiener, Graph structure in the web, *Comput. Netw.* 33 (1) (2000) 309–320.
- [209] D. Gruhl, R. Guha, D. Liben-Nowell, A. Tomkins, Information diffusion through blogspace, in: *Proceedings of the 13th ACM International Conference on World Wide Web*, ACM Press, New York, 2004, pp. 491–501.
- [210] K. Saito, M. Kimura, K. Ohara, H. Motoda, Learning continuous-time information diffusion model for social behavioral data analysis, in: *Advances in Machine Learning*, Springer, 2009, pp. 322–337.
- [211] K. Saito, M. Kimura, K. Ohara, H. Motoda, Selecting information diffusion models over social networks for behavioral analysis, in: *Machine Learning and Knowledge Discovery in Databases*, Springer, 2010, pp. 180–195.
- [212] A. Guille, H. Hacid, A predictive model for the temporal dynamics of information diffusion in online social networks, in: *Proceedings of the 21st ACM International Conference on World Wide Web*, ACM Press, New York, 2012, pp. 1145–1152.
- [213] D. Kempe, J. Kleinberg, É Tardos, Influential nodes in a diffusion model for social networks, in: *Automata, Languages and Programming*, Springer, 2005, pp. 1127–1138.
- [214] N. Barbieri, F. Bonchi, G. Manco, Topic-aware social influence propagation models, *Knowl. Inf. Syst.* 37 (3) (2013) 555–584.
- [215] W.O. Kermack, A.G. McKendrick, A contribution to the mathematical theory of epidemics, *Proc. R. Soc. Lond. Ser. A Math. Phys. Eng. Sci.* 115 (772) (1927) 700–721.
- [216] W. Kermack, A. McKendrick, Contributions to the mathematical theory of epidemics ii. The problem of endemicity, *Bull. Math. Biol.* 53 (1–2) (1991) 57–87.
- [217] J. Heffernan, R. Smith, L. Wahl, Perspectives on the basic reproductive ratio, *J. R. Soc. Interface* 2 (4) (2005) 281–293.
- [218] R.M. Anderson, R.M. May, *Infectious Diseases of Humans: Dynamics and Control*, Oxford University Press, 1991.
- [219] A. Barrat, M. Barthélemy, A. Vespignani, *Dynamical Processes on Complex Networks*, Cambridge University Press, 2008.
- [220] Y. Moreno, R. Pastor-Satorras, A. Vespignani, Epidemic outbreaks in complex heterogeneous networks, *Eur. Phys. J. B* 26 (4) (2002) 521–529.
- [221] R. Yang, B.-H. Wang, J. Ren, W.-J. Bai, Z.-W. Shi, W.-X. Wang, T. Zhou, Epidemic spreading on heterogeneous networks with identical infectivity, *Phys. Lett. A* 364 (3) (2007) 189–193.
- [222] M. Boguná, R. Pastor-Satorras, A. Vespignani, et al., Epidemic spreading in complex networks with degree correlations, in: *Proceedings of the 18th Sitges Conference on Statistical Mechanics*, in: *Lecture Notes in Physics*, Springer-Verlag, Berlin, 2003.
- [223] A.J. Morris, Representing spatial interactions in simple ecological models (Ph.D. thesis), University of Warwick, 1997.
- [224] M.J. Keeling, The effects of local spatial structure on epidemiological invasions, *Proc. R. Soc. Lond. B: Biol. Sci.* 266 (1421) (1999) 859–867.
- [225] T. House, M.J. Keeling, Insights from unifying modern approximations to infections on networks, *J. R. Soc. Interface* 8 (54) (2011) 67–73.
- [226] M. Youssef, C. Scoglio, An individual-based approach to sir epidemics in contact networks, *J. Theoret. Biol.* 283 (1) (2011) 136–144.
- [227] P. Van Mieghem, J. Omic, R. Kooij, Virus spread in networks, *IEEE ACM Trans. Netw.* 17 (1) (2009) 1–14.
- [228] M.E. Newman, S.H. Strogatz, D.J. Watts, Random graphs with arbitrary degree distributions and their applications, *Phys. Rev. E* 64 (2) (2001) 026118.
- [229] K.E. Hamilton, L.P. Pryadko, Tight lower bound for percolation threshold on an infinite graph, *Phys. Rev. Lett.* 113 (20) (2014) 208701.

- [230] B. Karrer, M.E. Newman, L. Zdeborová, Percolation on sparse networks, *Phys. Rev. Lett.* 113 (20) (2014) 208702.
- [231] J. Yang, J. Leskovec, Modeling information diffusion in implicit networks, in: *Proceedings of the 2010 IEEE International Conference on Data Mining*, IEEE Press, 2010, pp. 599–608.
- [232] F. Wang, H. Wang, K. Xu, Diffusive logistic model towards predicting information diffusion in online social networks, in: *Proceedings of the 32nd International Conference on Distributed Computing Systems Workshops*, IEEE Press, 2012, pp. 133–139.
- [233] K.-K. Kleineberg, M. Boguñá, Evolution of the digital society reveals balance between viral and mass media influence, *Phys. Rev. X* 4 (3) (2014) 031046.
- [234] X.-X. Zhan, C. Liu, G. Zhou, Z.-K. Zhang, G.-Q. Sun, J.J. Zhu, Mutual feedback between epidemic spreading and information diffusion, 2015, arXiv:1506.03932.
- [235] S. Funk, E. Gilad, V. Jansen, Endemic disease, awareness, and local behavioural response, *J. Theoret. Biol.* 264 (2) (2010) 501–509.
- [236] S. Funk, M. Salathé, V.A. Jansen, Modelling the influence of human behaviour on the spread of infectious diseases: a review, *J. R. Soc. Interface* 7 (50) (2010) 1247–1256.
- [237] X.-X. Zhan, C. Liu, G.-Q. Sun, Z.-K. Zhang, Epidemic dynamics on information-driven adaptive networks, 2015, arXiv:1505.04856.
- [238] S. Morris, Contagion, *Rev. Econ. Stud.* 67 (1) (2000) 57–78.
- [239] E. Even-Dar, A. Shapira, A note on maximizing the spread of influence in social networks, in: *Internet and Network Economics*, Springer, 2007, pp. 281–286.
- [240] G.C. Ehrhardt, M. Marsili, F. Vega-Redondo, Phenomenological models of socioeconomic network dynamics, *Phys. Rev. E* 74 (3) (2006) 036106.
- [241] I. Benczik, S. Benczik, B. Schmittmann, R. Zia, Lack of consensus in social systems, *Europhys. Lett.* 82 (4) (2008) 48006.
- [242] G. Caldarelli, A. Chessa, F. Pammolli, A. Gabrielli, M. Puliga, Reconstructing a credit network, *Nat. Phys.* 9 (3) (2013) 125–126.
- [243] T.S. Gardner, D. Di Bernardo, D. Lorenz, J.J. Collins, Inferring genetic networks and identifying compound mode of action via expression profiling, *Science* 301 (5629) (2003) 102–105.
- [244] M. Timme, Revealing network connectivity from response dynamics, *Phys. Rev. Lett.* 98 (22) (2007) 224101.
- [245] S. Hempel, A. Koseska, J. Kurths, Z. Nikoloski, Inner composition alignment for inferring directed networks from short time series, *Phys. Rev. Lett.* 107 (5) (2011) 054101.
- [246] Z. Levnajić, A. Pikovsky, Network reconstruction from random phase resetting, *Phys. Rev. Lett.* 107 (3) (2011) 034101.
- [247] A. Clauset, C. Moore, M.E. Newman, Hierarchical structure and the prediction of missing links in networks, *Nature* 453 (7191) (2008) 98–101.
- [248] J. Bongard, H. Lipson, Automated reverse engineering of nonlinear dynamical systems, *Proc. Natl. Acad. Sci. USA* 104 (24) (2007) 9943–9948.
- [249] R. Albert, A.-L. Barabási, Statistical mechanics of complex networks, *Rev. Modern Phys.* 74 (1) (2002) 47.
- [250] M.E. Newman, The structure and function of complex networks, *SIAM Rev.* 45 (2) (2003) 167–256.
- [251] S. Boccaletti, V. Latora, Y. Moreno, M. Chavez, D.-U. Hwang, Complex networks: Structure and dynamics, *Phys. Rep.* 424 (4) (2006) 175–308.
- [252] W.-X. Wang, R. Yang, Y.-C. Lai, V. Kovanis, C. Grebogi, Predicting catastrophes in nonlinear dynamical systems by compressive sensing, *Phys. Rev. Lett.* 106 (15) (2011) 154101.
- [253] W.-X. Wang, R. Yang, Y.-C. Lai, V. Kovanis, M.A.F. Harrison, Time-series-based prediction of complex oscillator networks via compressive sensing, *Europhys. Lett.* 94 (4) (2011) 48006.
- [254] W.-X. Wang, Y.-C. Lai, C. Grebogi, J. Ye, Network reconstruction based on evolutionary-game data via compressive sensing, *Phys. Rev. X* 1 (2) (2011) 021021.
- [255] E.J. Candès, J. Romberg, T. Tao, Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information, *IEEE Trans. Inform. Theory* 52 (2) (2006) 489–509.
- [256] E.J. Candès, J.K. Romberg, T. Tao, Stable signal recovery from incomplete and inaccurate measurements, *Comm. Pure Appl. Math.* 59 (8) (2006) 1207–1223.
- [257] D.L. Donoho, Compressed sensing, *IEEE Trans. Inform. Theory* 52 (4) (2006) 1289–1306.
- [258] E.J. Candès, M.B. Wakin, An introduction to compressive sampling, *IEEE Trans. Signal Process.* 25 (2) (2008) 21–30.
- [259] R. Pastor-Satorras, A. Vespignani, Epidemic spreading in scale-free networks, *Phys. Rev. Lett.* 86 (14) (2001) 3200.
- [260] C. Castellano, R. Pastor-Satorras, Non-mean-field behavior of the contact process on scale-free networks, *Phys. Rev. Lett.* 96 (3) (2006) 038701.
- [261] M. Gomez Rodriguez, J. Leskovec, A. Krause, Inferring networks of diffusion and influence, in: *Proceedings of the 16th ACM International Conference on Knowledge Discovery and Data Mining*, ACM Press, New York, 2010, pp. 1019–1028.
- [262] M. Gomez Rodriguez, J. Leskovec, B. Schölkopf, Structure and dynamics of information pathways in online media, in: *Proceedings of the 6th ACM International Conference on Web Search and Data Mining*, ACM Press, New York, 2013, pp. 23–32.
- [263] A.Y. Lokhov, M. Mézard, H. Ohta, L. Zdeborová, Inferring the origin of an epidemic with a dynamic message-passing algorithm, *Phys. Rev. E* 90 (1) (2014) 012801.
- [264] R.M. Tripathy, A. Bagchi, S. Mehta, Towards combating rumors in social networks: Models and metrics, *Intell. Data Anal.* 17 (1) (2013) 149–175.
- [265] D.J. Daley, D.G. Kendall, Epidemics and rumours, *Nature* 204 (1964) 1118.
- [266] D. Daley, D.G. Kendall, Stochastic rumours, *IMA J. Appl. Math.* 1 (1) (1965) 42–55.
- [267] D.P. Maki, M. Thompson, *Mathematical Models and Applications, with Emphasis on the Social, Life and Management Sciences*, Prentice Hall, 1973.
- [268] D.H. Zanette, Criticality of rumor propagation on small-world networks, 2001, arXiv:cond-mat/0109049.
- [269] D.H. Zanette, Dynamics of rumor propagation on small-world networks, *Phys. Rev. E* 65 (4) (2002) 041908.
- [270] Y. Moreno, M. Nekovee, A.F. Pacheco, Dynamics of rumor spreading in complex networks, *Phys. Rev. E* 69 (6) (2004) 066130.
- [271] Y. Moreno, M. Nekovee, A. Vespignani, Efficiency and reliability of epidemic data dissemination in complex networks, *Phys. Rev. E* 69 (5) (2004) 055101.
- [272] R.M. Tripathy, A. Bagchi, S. Mehta, A study of rumor control strategies on social networks, in: *Proceedings of the 19th ACM International Conference on Information and Management*, ACM Press, New York, 2010, pp. 1817–1820.
- [273] C. Budak, D. Agrawal, A. El Abbadi, Limiting the spread of misinformation in social networks, in: *Proceedings of the 20th International Conference on World Wide Web*, ACM Press, New York, 2011, pp. 665–674.
- [274] J. Borge-Holthoefer, Y. Moreno, Absence of influential spreaders in rumor dynamics, *Phys. Rev. E* 85 (2) (2012) 026116.
- [275] J. Borge-Holthoefer, A. Rivero, Y. Moreno, Locating privileged spreaders on an online social network, *Phys. Rev. E* 85 (6) (2012) 066123.
- [276] S. Aral, D. Walker, Identifying influential and susceptible members of social networks, *Science* 337 (6092) (2012) 337–341.
- [277] T.W. Valente, R.L. Davis, Accelerating the diffusion of innovations using opinion leaders, *Ann. Am. Acad. Pol. Sci.* 566 (1) (1999) 55–67.
- [278] D.J. Watts, S.H. Strogatz, Collective dynamics of small-world networks, *Nature* 393 (6684) (1998) 440–442.
- [279] D. Chen, L. Lü, M.-S. Shang, Y.-C. Zhang, T. Zhou, Identifying influential nodes in complex networks, *Physica A* 391 (4) (2012) 1777–1787.
- [280] G. Ghoshal, A.-L. Barabási, Ranking stability and super-stable nodes in complex networks, *Nature Commun.* 2 (2011) 394.
- [281] T. Lou, J. Tang, Mining structural hole spanners through information diffusion in social networks, in: *Proceedings of the 22nd ACM International Conference on World Wide Web*, ACM Press, New York, 2013, pp. 825–836.
- [282] L. Lü, Y.-C. Zhang, C.H. Yeung, T. Zhou, Leaders in social networks, the delicious case, *PloS ONE* 6 (6) (2011) e21202.
- [283] J. Tang, J. Sun, C. Wang, Z. Yang, Social influence analysis in large-scale networks, in: *Proceedings of the 15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, ACM Press, New York, 2009, pp. 807–816.
- [284] J. Yang, S. Counts, Predicting the speed, scale, and range of information diffusion in twitter, in: *Proceedings of the 4th International Conference on Weblogs and Social Media*, AAAI Press, 2010, pp. 355–358.
- [285] M. Gilbert, N. Golding, H. Zhou, G.W. Wint, T.P. Robinson, A.J. Tatem, S. Lai, S. Zhou, H. Jiang, D. Guo, et al., Predicting the risk of avian influenza a h7n9 infection in live-poultry markets across asia, *Nature Commun.* 5 (2014) 4116.
- [286] H.W. Hethcote, The mathematics of infectious diseases, *SIAM Rev.* 42 (4) (2000) 599–653.
- [287] J. Kleinberg, Computing: The wireless epidemic, *Nature* 449 (7160) (2007) 287–288.

- [288] B.A. Prakash, D. Chakrabarti, M. Faloutsos, N. Valler, C. Faloutsos, Threshold conditions for arbitrary cascade models on arbitrary networks, in: *Proceedings of the 11th IEEE International Conference on Data Mining*, IEEE process, 2011, pp. 537–546.
- [289] L. Briesemeister, P. Lincoln, P. Porras, Epidemic profiles and defense of scale-free networks, in: *Proceedings of the 2003 ACM Workshop on Rapid Malcode*, ACM Press, New York, 2003, pp. 67–75.
- [290] Y. Hayashi, M. Minoura, J. Matsukubo, Recoverable prevalence in growing scale-free networks and the effective immunization, 2003, [arXiv:cond-mat/0305549](#).
- [291] B.A. Prakash, H. Tong, N. Valler, M. Faloutsos, C. Faloutsos, Virus propagation on time-varying networks: Theory and immunization algorithms, in: *Proceedings of the 2010 European Conference on Machine Learning and Knowledge Discovery in Databases*, Springer Press, 2010, pp. 99–114.
- [292] N.C. Valler, B.A. Prakash, H. Tong, M. Faloutsos, C. Faloutsos, Epidemic spread in mobile ad hoc networks: Determining the tipping point, in: *Proceedings of the 10th International IFIP TC 6 Conference on Networking*, Springer Press, 2011, pp. 266–280.
- [293] J. Marcelino, M. Kaiser, Reducing influenza spreading over the airline network, *PLoS Curr. Influenza* 1 (2009) RRN1005.
- [294] C.M. Schneider, T. Mihaljev, S. Havlin, H.J. Herrmann, Restraining epidemics by improving immunization strategies, 2011, [arXiv:1102.1929](#).
- [295] P. Holme, B.J. Kim, C.N. Yoon, S.K. Han, Attack vulnerability of complex networks, *Phys. Rev. E* 65 (5) (2002) 056109.
- [296] C.J. Kuhlman, G. Tuli, S. Swarup, M.V. Marathe, S. Ravi, Blocking simple and complex contagion by edge removal, in: *Proceedings of the 13th IEEE International Conference on Data Mining*, IEEE Press, 2013, pp. 399–408.
- [297] A.N. Bishop, I. Shames, Link operations for slowing the spread of disease in complex networks, *Europhys. Lett.* 95 (1) (2011) 18005.
- [298] H. Chan, L. Akoglu, H. Tong, Make it or break it: Manipulating robustness in large networks, in: *Proceedings of the 2014 SIAM International Conference on Data Mining*, SIAM, 2014, pp. 325–333.
- [299] L. Lü, M. Medo, C.H. Yeung, Y.-C. Zhang, Z.-K. Zhang, T. Zhou, Recommender systems, *Phys. Rep.* 519 (1) (2012) 1–49.