ICIC Express Letters ICIC International ©2017 ISSN 1881-803X
Volume **11**, Number **1**, January **2017** pp. 183–188

A REVIEW ON EPIDEMIOLOGICAL MODELING OF INFORMATION

Daniel Moritz Marutschke

College of Information Science and Engineering
Ritsumeikan University
1-1-1 Nojihigashi, Kusatsu, Shiga 525-8577, Japan
moritz@fc.ritsumei.ac.jp

Received June 2016; accepted September 2016

Abstract. *Epidemiological modeling is a useful tool to investigate propagation dynamics of contagious agents. The field comes from epidemiology, the study of diseases. This research area has matured over decades and since being expanded to model the spread of other fields shown to have similar properties. The purpose of this review is to survey previous models available, what areas they have been applied to, investigate limitations, and offer directions for future research.*
Keywords: Knowledge propagation, Epidemiological modeling, SIR and SEIR models

1. **Introduction.** Cross-pollination of research areas is common practice – the adaptation of epidemiology, the study of diseases, poses no exception. The reasoning that diseases spread has been established some centuries ago, however limited to simple fever-spreading patterns and later to more significant discoveries of non-random patterns in cholera pandemic in Europe [1].

More delicate simulations – including competing diseases, vaccination campaigns, etc. – were made possible by computers in recent decades, growing in their computational power.

Using the existing models to explain dynamics in non-medical and non-biological fields was a natural evolution in research, as it often happens. In some areas, the initial enthusiasm was dampened by the realization that implementation proved to be difficult. Many business experts roll their eyes when hearing the phrase "viral marketing".

This paper provides an overview of the current state of research in deterministic models used in areas other than biological and medical epidemiology. Relevant general models in their basic form are provided as a system of ordinary differential equations. Discussing each adapted model, the author refers to the original and provides theoretical reasoning and discusses practical application and future directions where applicable.

Deterministic models in biological and medical epidemiology have been successfully used to track viral, bacterial, parasitic, and other contagious agents [2]. Hethcote's paper, being one of the most cited in regards to general state of epidemiology, shows the connection of not only different models and their extensions, but also in conditions such as vaccination campaigns. Models have been extended to fit special circumstances such as maternal immunity of new-born, vaccination campaigns and the delay thereof, pandemic spread with a combination of the aforementioned conditions, and many other scenarios. Numerous compartmental models were permutated to fit a given need – namely *SI*, *SIS*, *SIR*, *SEIR*, *SEIRS*, *MSEIRS*, etc., with the compartments represented as: S for susceptible, I for infected, R for recovered or removed, E for exposed (incubation time), and M for maternal immunity, respectively. It follows that these models end up in various states. Some manage to stay in an equilibrium, others fall into an oscillating state, and some others overrun by the infectious agent. To provide some rudimentary examples, these models are described in the following paragraphs.

SI models have two compartments and its dynamic develops, if every individual that comes into contact with the contagious agent will be infected – a community or system will be overrun.

SIR models have three compartments (an additional one for recovered/removed) to describe individuals who are removed from the equation after having been infected. Recovered or removed usually constitutes death by the disease or immunization.

SEIR models incorporate four compartments to account for an incubation period. This stage takes place before individuals become infected to compensate for a delay in becoming infectious and spreading the contagious agent further.

Without additional compartments, models such as the *SIS*, *SIRS*, and *SEIRS*. represent oscillating systems where once infected or recovered individuals enter the susceptible state again. This is being indicated by the trailing S at the end of the model to describe that the system is entering its original state after a given time period.

Transitions from one compartment to another can be described by Ordinary Differential Equations (ODE), leading to a system of differential equations, matching the number of compartments in a given model. These systems and their corresponding parameters are explained for the *SIR* and the *SEIR* models in the next section.

The premise that information propagates with similar mechanisms has been proposed over half a century ago by Rapoport and Bailey [3, 4, 5]. Goffman and Newill formulated a more specific notion of information diffusion in 1964 and 1967 [6, 7]. Several fields have adopted deterministic models to analyze marketing behavior (viral marketing) using clustered blog sites and the idea of a basic *SIS* model [8]; the spread of rumors in a finite, closed, and homogeneous population, following dynamics of the *SIR* model [9] and a detailed comparison of known models to that of rumors [10]; tracking of voters opinions by analyzing variations in the spread of hashtags on Twitter [11]; word-of-mouth propagation including herd behavior, propagation of word-of-mouth based e-mail forwarding, economics, etc. [12-14].

This paper is structured into the following sections: introduction, which lays down an overview of epidemiology with related models and its relatedness to other propagation mechanisms; basic models (Section 2), where the two most prevalent original models are described; research position (Section 3), which provides a brief history and the state of art in epidemiological modeling of non-biological contagious agents; future work (Section 4), with a focus on current limitations, possible approach to solve issues, and open problems; conclusions (Section 5), to summarize the given prospects.

2. **Basic Models.** Most of the approaches in modeling epidemiological dynamics fall back to the two models described in this section. Their idea, parameters, and graphical representation are given in their prevalent form. Some compartments that are renamed in other research can usually be reconstructed with these two models.

The basic epidemiological model shown in Figure 1 has four compartments: S for susceptible individuals, E for exposed individuals, i.e., individuals who have come into contact with a contagious agent, I for infected individuals, and R for removed individuals. Susceptible individuals are generally able to contract a contagious agent and move to the next compartment resembling the incubation time. Once they can themselves transmit

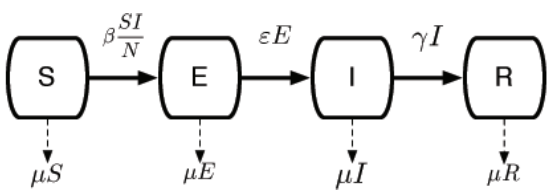

FIGURE 1. Basic *SEIR* model with transition parameters

the contagious agent to another individual, they move to the infectious compartment. Developing antibody and becoming immune or dying from the infection, individuals move to the removed compartment.

The equivalent *SIR* model can be described in the same way, without the E compartment (and without transition parameters εE and μE, respectively). The two systems of Ordinary Differential Equations detail the dynamic, including the parameters, in Equations (1a) to (1c) for the *SIR* model and Equations (2a) to (2c) for the *SEIR* model.

$$\frac{\mathrm{d}S}{\mathrm{d}t} = -\mu S - \beta\frac{SI}{N} \tag{1a}$$

$$\frac{\mathrm{d}I}{\mathrm{d}t} = \beta\frac{SI}{N} - (\gamma + \mu)I \tag{1b}$$

$$\frac{\mathrm{d}R}{\mathrm{d}t} = \gamma I - \mu R \tag{1c}$$

$$\frac{\mathrm{d}S}{\mathrm{d}t} = \mu N - \mu S - \beta\frac{SI}{N} \tag{2a}$$

$$\frac{\mathrm{d}E}{\mathrm{d}t} = \beta\frac{SI}{N} - (\varepsilon + \mu)E \tag{2b}$$

$$\frac{\mathrm{d}I}{\mathrm{d}t} = \varepsilon E - (\gamma + \mu)I \tag{2c}$$

$$\frac{\mathrm{d}R}{\mathrm{d}t} = \gamma I - \mu R \tag{2d}$$

Compartments and transition parameters can be identified easily by referring to the equations and the compartmental model in Figure 1. N is the number of individuals in the whole population that is investigated. S, E, I, and R are the number of individuals in the compartments for *susceptible*, *exposed*, *infected*, and *recovered*, respectively. In epidemiological terms, the parameters are defined as follows: the contact rate β; the incubation time $\frac{1}{\mu}$; the average infectious period $\frac{1}{\varepsilon}$; the life time of the infectious material γ.

3. **Research Position.** Goffman and Newill were one of the first authors to specifically propose simple deterministic models known from viral epidemiology to adapt to idea diffusion [6]. Goffman and Newill formulated the theoretical mathematical framework appropriate for information diffusion with characteristics and the corresponding reasoning. It is not until much later, however, that these mathematics could be put into practice. Some of these applications are described in the following paragraphs.

3.1. **Ideas and knowledge.** Bettencourt et al. [15] have presented a modified model they call *SEIZ* model, which incorporates the notion of skeptics or stiflers (compartment Z), which could actively weaken the propagation of a new idea or knowledge when worked against. The authors successfully improved the fitness of the *SEIR* model to the propagation of the Feynman diagram in theoretical physics. Interesting is in particular the comparison of the propagation in three different countries: Japan, USSR, and the U.S., having their own starting point in history and propagation characteristics.

Marutschke [16] did an extensive study on knowledge propagation in scientific publications of 88 initial keywords over a timespan of several decades using multiple extended models. Results included stable tracking of scientific keywords, keyword classification using principal component analysis, and demonstrating two less viable models (*SEIRK* and *SEIRKE* with a compartment K to deal with indirect influencers) to address knowledge propagation.

3.2. **Rumors.** Rumors and misinformation in several forms are analyzed by Shirai et al. [17] and Okada et al. [18]. These rumors and misinformation are namely regarding Cosmos Oil Company, power saving, iodine (chemical element), Turkey (country), Taiwan, Pokemon, and Fuji Television. The authors used their model to classify rumors into four categories, one-time simultaneous epidemic, one-time separate epidemic, recurring simultaneous epidemic, and recurring separate epidemic.

Jin et al. examine the established *SEIZ* model to favor a simpler *SIS* model [19]. In particular, news trends about the following topics were analyzed: Boston Marathon Bombings in 2013 on April 15 at 14:49:12 local time; Pope Resignation on the morning of February 11, 2013; Amuay Refinery Explosion on August 25, 2012 1:11 am local time; Michelle Obama at the 2013 Oscars.

3.3. **Social media.** Abdullah and Wu modified the basic *SIR* model to simulate the spread of news items on Twitter [20]. Their key to their approach is the rapid retweet (forwarding and spreading of the original information) of the news. The almost instantaneous retweet (forwarding) of news items makes the incubation time obsolete. Large textual data make for a fine-grained analysis. In addition, a two-stage Markov model is employed to separate epidemiological from non-epidemiological phases.

Xiong et al. proposed a diffusion model for the purpose of tracking information on microblogging sites using Twitter [21]. The *SCIR* model suggested in their paper resembles four states: susceptible, contacted, infected, and refractory. Their approach differs from Abdullah and Wu in the addition of a compartment similar to that symbolizing incubation time.

Meme tacking using an even more rudimentary *SI* model was done in [22].

Incorporating semantic information on social media to track actual diseases was investigated by Shao et al. [23]. Prakash uses the *SEIR* and an *SEIS* model to accomplish similar modeling [24]. Chen et al. use the *SIR* model for a public health indication model [25].

SIR model and an adapted *SIRI* model were used in Skaza and Blias to track hashtags on Twitter [26].

Study on the growth and decline of SNSs by using the infectious recovery SIR model by Tanaka et al. [27].

3.4. **Business and marketing.** Gurley and Johnson proposed a minor extension of the classic *SEIR* model using constants to better fit the propagation of subfields in economics [14]. In their paper, the fields of economics are tracked with high statistical significance, resulting in the ability to detect subfields that are likely to expand rapidly or are more likely to collapse. The comparison is done within one field of research without cultural adaptation.

Audestad proposed a model adapted by the classic susceptible-infectious-recovered (*SIR*) scheme to model dynamic market behavior with and without feedback [28]. The logical form of the model is presented as a buyer-player-quitter state description.

3.5. **Other areas.** Other fields formed from these mature mathematics, such as an $SEIQV$ model – susceptible, exposed, infected, quarantined, vaccinated – to track computer worms incorporating quarantine and vaccination parameters [29] or the spread in computer networks [30]. Noteworthy is the modeling of the forming of extremist opinion by Stauffer with a *GSEF* model – general, susceptible, excited, fanatic – and Sznajd with a closed community dogma of "united we stand divided we fall" [31, 32].

Political party growth is modeled with the *SIR* basis in [33].

The Internet of Things (IoT) becoming of growing importance, Makhoul et al. explore the *SIR* model to classify survivability of systems in the IoT era [34].

4. **Future Work.** The wide range of research provided shows viability to numerous fields. The propagation of scientific knowledge, rumors, opinions, economic fields, and others has been analyzed in several mediums, ranging from manual data selection, scientific publications, online database, social networks, etc. Although modeling can be done with high accuracy, the works done so far are mostly reactionary. Targeted fields can be represented well and show promising further insights, such as analysis of social media.

Business and Management on the other hand are naturally more difficult to focus. Repetition of products in a given market is less likely, due to the nature of competition. This also could explain the receding interest a few decades ago. With growing datasets, the ability to model systems of ODEs, and adaptations to agent-based modeling, the field indicates a direction for more exploration in this research discipline.

Most of presented researches that introduce new models to better fit the data would benefit from the perspective of model selection criteria. One way to test for future research would be the use of Akaike Information Criteria (AIC) or Bayesian Information Criteria (BIC). In some cases the increase in model complexity might not justify the increase in the model's fitness.

As the notion of big data is becoming an issue that needs addressing, information about feedback could be used to more accurately model epidemiological behavior. As foundational behavior is known in many fields and data is becoming more accessible, a more stable system, including the aforementioned statistics, could be a focus for upcoming research.

5. **Conclusions.** From medical and biological epidemiology, the use of deterministic models has steadily found the way into other areas. Complex idea diffusion and knowledge propagation has been difficult but recent approaches with representation using social media, business and management, scientific publications, and others have shown potential in the direction of more difficult contagious material. New fields such as the Internet of Things have now been involved as well.

Based on the current study and previous ones, extension on existing research is promising and the expansion to new fields with comparable dynamics is not explored enough, yet.

REFERENCES

[1] S. Boslaugh, *Encyclopedia of Epidemiology*, Sage Publications, 2008.

[2] H. W. Hethcote, The mathematics of infectious diseases, *SIAM Review*, vol.42, no.4, pp.599-653, 2000.

[3] A. Rapoport, Spread of information through a population with socio-structural bias: I. Assumption of transitivity, *The Bulletin of Mathematical Biophysics*, vol.15, no.4, pp.523-533, 1953.

[4] A. Rapoport, Spread of information through a population with socio-structural bias: II. Various models with partial transitivity, *The Bulletin of Mathematical Biophysics*, vol.15, no.4, pp.535-546, 1953.

[5] N. T. J. Bailey, A simple stochastic epidemic, *Biometrika*, vol.37, nos.3-4, pp.193-202, 1950.

[6] W. Goffman and V. A. Newill, Generalization of epidemic theory: An application to the transmission of ideas, *Nature*, vol.204, no.4955, pp.225-228, 1964.

[7] W. Goffman and V. A. Newill, Communication and epidemic processes, *Proc. of the R. Soc. Lond.*, vol.298, pp.316-334, 1967.

[8] J. Leskovec, L. A. Adamic and B. A. Huberman, The dynamics of viral marketing, *ACM Trans. the Web*, vol.1, no.1, p.5, 2005.

[9] D. J. Daley and D. G. Kendall, Epidemics and rumours, *Nature*, vol.204, no.4963, p.1118, 1964.

[10] K. Dietz, Epidemics and rumours: A survey, *Journal of the Royal Statistical Society. Series A (General)*, vol.130, no.4, pp.505-528, 1967.

[11] D. M. Romero, B. Meeder and J. Kleinberg, Differences in the mechanics of information diffusion across topics: Idioms, political hashtags, and complex contagion on Twitter, *Proc. of the 20th International Conference on World Wide Web*, New York, USA, pp.695-704, 2011.

[12] G. Ellison and D. Fudenberg, Word-of-mouth communication and social learning, *Quarterly Journal of Economics*, vol.110, no.1, pp.93-125, 1995.
[13] J. E. Phelps, R. Lewis, L. Mobilio, D. Perry and N. Raman, Viral marketing or electronic word-of-mouth advertising: Examining consumer responses and motivations to pass along email, *Journal of Advertising Research*, vol.44, no.4, pp.333-348, 2004.
[14] N. Gurley and D. K. N. Johnson, Viral economics: An epidemiological model of knowledge diffusion in economics, *Social Science Research Network*, 2011.
[15] L. M. A. Bettencourt, A. Cintronarias, D. Kaiser and C. Castillo-Chávez, The power of a good idea: Quantitative modeling of the spread of ideas from epidemiological models, *Physica A: Statistical Mechanics and Its Applications*, no.364, pp.513-536, 2006.
[16] D. M. Marutschke, *Epidemiological Modeling of Knowledge Propagation Represented by Scientific Publications*, Ph.D. Thesis, Kobe University, 2014.
[17] T. Shirai, S. Sakaki, F. Toriumi, K. Shinoda, K. Kazama, I. Noda, M. Numao and S. Kurihara, Estimation of false rumor diffusion model and prevention model of false rumor diffusion on Twitter, *The 26th Annual Conference of the Japanese Society of Artificial Intelligence*, pp.1-9, 2012.
[18] Y. Okada, T. Sakaki, T. Toriumi, K. Shinoda, K. Kazama, I. Noda, M. Numao and S. Kuriha, False rumor diffusion analysis based on the SIR-extended information diffusion model, *The 27th Annual Conference of the Japanese Society for Artificial Intelligence*, pp.1-4, 2013.
[19] F. Jin, E. Dougherty, P. Saraf, Y. Cao and N. Ramakrishnan, Epidemiological modeling of news and rumors on Twitter, *Proc. of the 7th Workshop on Social Network Mining and Analysis*, New York, USA, pp.1-9, 2013.
[20] S. Abdullah and X. Wu, An epidemic model for news spreading on Twitter, *IEEE International Conference on Tools with Artificial Intelligence*, pp.163-169, 2011.
[21] F. Xiong, Y. Liu, Z. Zhang, J. Zhu and Y. Zhang, An information diffusion model based on retweeting mechanism for online social media, *Physics Letters A*, vol.376, nos.30-31, pp.2103-2108, 2012.
[22] G. De Martino and S. Spina, Exploiting the time-dynamics of news diffusion on the Internet through a generalized susceptible-infected model, *Physica A: Statistical Mechanics and Its Applications*, no.438, pp.634-644, 2015.
[23] H. Shao, K. S. M. T. Hossain, H. Wu, M. Khan, A. Vullikanti, B. A. Prakash, M. Marathe and N. Ramakrishnan, Forecasting the flu: Designing social network sensors for epidemics, *arXiv.org*, 2016.
[24] B. A. Prakash, Prediction using propagation: From flu trends to cybersecurity, *IEEE Intelligent Systems*, pp.1-5, 2016.
[25] L. Chen, K. S. M. T. Hossain, P. Butler, N. Ramakrishnan and B. A. Prakash, Syndromic surveillance of flu on Twitter using weakly supervised temporal topic models, *Data Mining and Knowledge Discovery*, vol.30, no.3, pp.681-710, 2015.
[26] J. Skaza and B. Blais, Modeling the infectiousness of Twitter hashtags, *arXiv.org*, 2016.
[27] M. Tanaka, Y. Sakumoto, M. Aida and K. Kawashima, Study on the growth and decline of SNSs by using the infectious recovery SIR model, *The 10th Asia-Pacific Symposium on Information and Telecommunication Technologies*, pp.1-3, 2015.
[28] J. A Audestad, Some dynamic market models, *arXiv.org*, 2015.
[29] F. Wang, Y. Zhang, C. Wang, J. Ma and S. Moon, Stability analysis of a SEIQV epidemic model for rapid spreading worms, *Computers & Security*, vol.29, no.4, pp.410-418, 2010.
[30] J. R. C. Piqueira, B. F. Navarro and L. H. A. Monteiro, Epidemiological models applied to viruses in computer networks, *Journal of Computer Science*, vol.1, no.1, pp.31-34, 2005.
[31] D. Stauffer and M. Sahimi, Discrete simulation of the dynamics of spread of extreme opinions in a society, *Physica A: Statistical Mechanics and Its Applications*, no.364, pp.537-543, 2006.
[32] K. Sznajd-Weron and J. Sznajd, Opinion evolution in closed community, *International Journal of Modern Physics C*, vol.11, no.6, p.13, 2001.
[33] R. A. Jeffs, J. Hayward, P. A. Roach and J. Wyburn, Activist model of political party growth, *arXiv.org*, pp.359-372, 2015.
[34] A. Makhoul, C. Guyeux, M. Hakem and J. M. Bahi, Using an epidemiological approach to maximize data survival in the Internet of Things, *ACM Trans. Internet Technology*, vol.16, no.1, pp.1-15, 2016.