BOOK REVIEW

The Rules of Contagion: Why Things Spread—And Why They Stop

Reviewed by David Aldous

The Rules of Contagion: Why Things Spread—And Why They Stop
Adam Kucharski

This book contains almost no mathematics, instead it provides extensive real-world background on a topic of intrinsic interest which, in the context of pandemics, is of immense real-world importance. While an epidemic is the spread of a disease, it has long been realized that the spread of opinions, preferences, or practices, and nowadays fake news or cat videos, is analogous. *Notices* readers will know, or not be surprised to learn, that there has been substantial mathematical modeling of such phenomena over the last century, enhanced in the 21st century by the increased availability of data (e.g., from social networks).

This book succeeds marvelously in illustrating the breadth of this analogy, via roughly 50 real-world stories told in greater or lesser detail. It has an exemplary style of serious popular science, avoiding the fluffy style of "I was sitting in Professor X's office one spring morning ..." in favor of just telling the story accurately. It is worth noting that it was written just before the COVID-19 pandemic, and that the author is an epidemiologist rather than a mathematician.

David Aldous is a professor emeritus of statistics and mathematics at UC Berkeley. His email address is `aldousdj@berkeley.edu`.

Communicated by Notices *Book Review Editor Katelynn Kochalski.*

DOI: https://doi.org/10.1090/noti2486

Despite the minimal mathematics in the book, I hope it will be widely read and appreciated by mathematicians. It provides both an overview of one of the major real-world contexts where elementary and advanced mathematics are manifestly useful, and a fertile source for engaging undergraduates with real-world modeling activities.

Outline of contents: Epidemics. The book opens with a detailed account of the work on malaria transmission by Ronald Ross, who argued in 1910 via mathematical modeling that mosquito control—reducing the prevalence of mosquitos below some critical level—would suffice to minimize the prevalence of malaria. Ross's subsequent interaction with Anderson McKendrick led to the first epidemiological paper [KM27] that we might regard as serious mathematical research. This 1927 paper—rather scandalously not included in *MathSciNet*—introduced the fundamental SIR (susceptible-infected-recovered) model. In the most basic form of this model, individuals are in one of the three SIR states, and each susceptible becomes infected at rate proportional to the number of infectives.

The concept of *herd immunity* was introduced also around this time. Such timeline assertions can quickly be checked for consistency with Google Ngram (see Figure 1 below). Mathematical analysis of variant models developed slowly over the next decades, and a milestone was the 1957 monograph by Bailey [Bai57]. To digress, though I am no expert, I have a personal fondness for the topic because this was the first monograph, rather than textbook, that I read as an undergraduate. I suspect that most of us remember our first monograph.

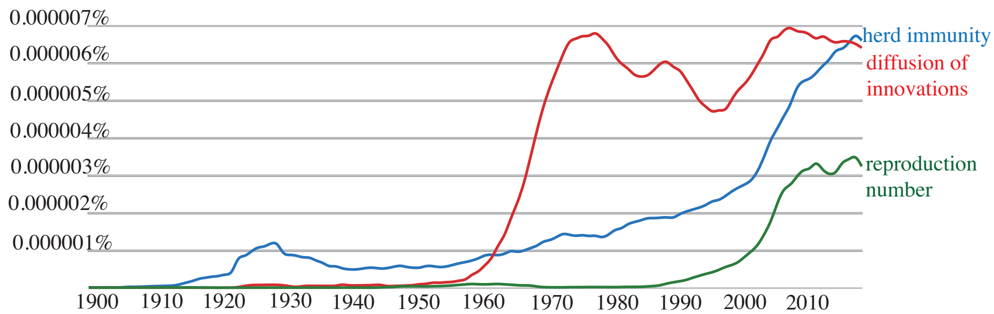

Figure 1. Trend in frequency of 3 phrases: "herd immunity," "reproduction number," and "diffusion of innovations." From Google Books Ngram Viewer, `https://books.google.com/ngrams`.

In the *MathSciNet* review, the monograph is praised as

> ...the first comprehensive account of the work that has been done in this field, and it is difficult to think how his task could have been better done. ...to the mathematical reader the most interesting parts of the book are those in which a model is defined and an attempt is then made to give a qualitative description of its behaviour.

Its subsequent influence on the mathematical literature is evidenced by 4000+ citations to the 1975 second edition. In contrast, consider the words of the book under review.

> Then [1957] progress stuttered. The obstacle was [the monograph [Bai57]]. ...It was almost entirely theoretical, with hardly any real-life data. [It] was an impressive survey of epidemic theory ...But here was a problem: Bailey had left out a crucial idea, which would turn out to be one of the most important concepts ...

That crucial idea was to focus attention on the *reproduction number R*, the average number of new individuals who get infected from one individual. This focus on R, which became prominent in applied epidemiology over the 1980s (a timeline assertion again consistent with Ngram—Figure 1) is here attributed to Robert May and Roy Anderson around 1980. While the latter attribution seems literally true for the phrase *reproduction number*, the concept was in fact well understood much earlier (under the term *threshold*) in the applied probability and theoretical epidemic modeling community—see [HD96] for history and an introduction to the mathematics.

The relevance of R to epidemics is clear: in the simplest possible model, from a few initial infectives there will be around R^n infectives after the nth stage of transmission, so a pandemic will occur if $R > 1$ but will not occur if $R < 1$. Of course, for any pretense of realism one needs to take account of the heterogeneity of populations. Heterogeneity arises in many ways. As a "spatial" issue, in the sense that disease contacts must be physically close, or online contacts are generally people "like you" in a relevant context. And as a "number of contacts" issue, which is affected by your occupation and sociability. Ideally, epidemics and analogous processes should be modeled via a network whose vertices are individuals and whose edges, representing possible transmission, are marked with the probability of transmission if one end-individual is infected. Needless to say, though inventing such models is easy, finding enough real-world data to make them realistic is very difficult.

For a mathematician, we have not given a precise definition of R. In the real world this is estimated from data. Indeed, many of us tracked graphics for such estimates during the COVID-19 pandemic. Within a mathematical model, R will be a function of the model parameters. In both cases what one obtains is better described as an *effective* reproduction number, matching the "R^n infectives after the nth stage" interpretation. The book's useful mnemonic is

> R depends on the 4 DOTS: *duration* of infection, *opportunities* for transmission, *transmission probability* during each opportunity, and average *susceptibility*.

Regarding "number N of contacts," COVID-19 has made us all familiar with the *superspreader* concept, roughly corresponding mathematically to a power law tail for the distribution of N. The book gives various examples from previous epidemics. For instance, regarding Ebola:

> The cases most likely to be involved with superspreading were the ones that couldn't be linked to existing chains of transmission. Put simply, the people driving the epidemic were generally the ones the health authorities didn't know about. These people went undetected until they sparked a new set of infections, making it near impossible to predict superspreading events.

Like other recent writers, the author debunks the persistent myth of a "patient zero" initiating the AIDS epidemic.

Outline of contents: Analogies. The most obvious analogy is surely malware—*computer virus*—though this is discussed only briefly.

Another widely known analogy is the notion of *diffusion of innovations* [Wik21] which was introduced by

Everett Rogers in 1962 and quickly gained traction (Figure 1 again). This models the proportion of users who have adopted a new technology product, from color televisions in the 1960s to smartphones in the 2000s. Mathematically, this is merely the even simpler SI model with I = number of adopters and S = number of non-adopters. It predicts the *S*-shaped logistic function $ce^{t-t_0}/(e^{t-t_0}+e^{t_0-t})$ as an approximation to the proportion adopting, as a function of time.

Another well known analogy involves finance, and the familiar story of CDOs (collateralized debt obligations) and systemic risk leading to the global financial crisis of 2007–2008 is recounted. In a conceptual sense, the financial system exists to move risk from those who are risk-averse to those who are willing to be paid to take on the risks, via a network of institutions which are borrowing and lending to each other, and the failure of one institution could cause a dramatic "house of cards" collapse.

Turning to the spread of ideas and habits, examples in Chapter 3 include the notorious and widely publicized Christakis-Fowler studies starting from [CF07] asserting a "social contagion" effect for conditions such as obesity, happiness, or loneliness. These studies concluded that one individual acquiring the condition tended to increase the chance of a friend acquiring the condition, and that this effect extended to friends-of-friends. It is noted that whether this is correct or even meaningful is debated – mathematician Russell Lyons [Lyo11] and others have published critiques. Another example concerns the *backfire effect*, the claim that being told additional information, logically contradicting existing beliefs, might actually reinforce them. This raises the interesting question of how one can get data to test this?

COVID-19 has made us familiar with ways to attempt to control a disease epidemic once started, but what about control of analogous processes? Chapter 4 includes stories about metaphorical "epidemics of violence" or opioid abuse, and describes attempts at control techniques such as predictive policing.

The longest chapter, *Going Viral*, gives fascinating details about matters for which (I guess) most of us have read only superficial accounts. Many millions of organizations are seeking to spread their message widely online—from advertising to political campaigning to fake news—but we don't have the time to watch all of these, so most attempts must fail to gain widespread attention. What makes the few succeed? The phrase "going viral" suggests some amateur humorous YouTube video, but in fact this is the exception. The author describes his own experience after giving a YouTubed talk at London's Royal Institution. From followers of that institution, it acquired a hundred or so views a day for a year, then suddenly it picked up more views in a few days than in all previous time. This was not because of a viral epidemic, but because it was featured on the YouTube homepage. Such stories provide background to an important finding.

> A common argument for featuring extreme views is that they would spread anyway, even without media amplification. But studies of online contagion have found the opposite: content rarely goes far without broadcast events to amplify it. If an idea becomes popular, it's generally because well-known personalities and media outlets have helped it spread . . . .

Another line of analysis argues that *influencers* have rather less effect than generally supposed. For instance, given a new tweet, can one predict whether it will create a large Twitter cascade? Analyses show that the content of the tweet has little relevance, while the tweeter's past record has more relevance, but no prediction algorithm was found to be very accurate.

There is a brief but admirably nuanced discussion of data privacy versus usefulness of mass data. Cell phone GPS data on individuals has been used for contact tracing during COVID-19, reasonably enough, but the general availablity of such data has obvious risks:

> In a 2014 survey, 85% of US domestic violence shelters were protecting people from abusers who had stalked them via GPS.

As another example, we might agree that "how does the content we see on social media affect our emotional state?" is an important question, but does this justify the Facebook experiment in which researchers secretly altered News Feeds to show happier or sadder posts?

Discussion. The book is organized appropriately around the contexts where epidemics and analogs occur, rather than around the underlying mathematical methodologies, which are hardly mentioned. In true scholarly fashion, there are 53 pages of references, mostly to sources for the stories; some contain relevant data, but few address the underlying mathematics. The curious mathematician might be interested to know that a huge relevant mathematical academic literature has appeared over the last 20 years, predominantly on what is best termed *spread of information on networks*. As a few representative books or surveys:

- [KMS17] replicates the spirit of [Bai57] by providing a rigorous comprehensive treatment of a range of basic SIR-style models on networks, immediately becoming the standard reference for such material.
- [DM10] has a broader introductory "applied math" style.
- [CFL09] takes a wide-ranging statistical physics approach.

In my opinion, the book under review is also valuable as extensive thought-provoking raw material for undergraduate research projects. As background, I am not enthusiastic about organized REU programs aimed at theorem-proving. Undoubtedly they serve to identify and encourage the very best future researchers. But proving new theorems is hard: I worry that typical attendees might get discouraged by lack of success, or conversely succeed and get the impression that theory research is easier than it really is. In contrast, devising and studying toy models of real-world phenomena promotes a wide range of skills—identifying a conceptual question, devising models, analysis of simple models and simulations of complex ones, comparison with data. Mathematical readers of this book surely have enough creative imagination to identify many such questions. For instance: given a limited budget for contact-tracing in an epidemic, is it better to use resources to trace forward (from a case testing positive today, who might they have already infected?) or backward (who might they have been infected by, and then who else might that person have infected?). Such questions have of course already been examined, but there is no "one right answer" given the multitude of models one might choose.

Regarding COVID-19, the significance of R and the necessity for measures to reduce R have perhaps been understood by those members of the public who take a rational view of medicine. The significance of exponential growth is often less apparent to non-mathematicians. For instance, the one or two week delay (compared to an alternative more aggressive strategy) in starting the U.K. lockdown in March 2020 clearly had a huge potential downside in permitting a substantially larger infected population via exponential growth, compared to a limited upside. The current Prime Minister ironically ignored the saying attributed to an earlier PM—"a week is a long time in politics."

References

[Bai57] Norman T. J. Bailey, *The mathematical theory of epidemics*, Hafner Publishing Co., New York, 1957. MR0095085

[CF07] Nicholas A. Christakis and James H. Fowler, *The spread of obesity in a large social network over 32 years*, New England Journal of Medicine **357** (2007), no. 4, 370–379, available at `https://doi.org/10.1056/NEJMsa066082`. PMID: 17652652.

[CFL09] Claudio Castellano, Santo Fortunato, and Vittorio Loreto, *Statistical physics of social dynamics*, Rev. Mod. Phys. **81** (2009), 591–646.

[DM10] Moez Draief and Laurent Massoulié, *Epidemics and rumours in complex networks*, London Mathematical Society Lecture Note Series, vol. 369, Cambridge University Press, Cambridge, 2010. MR2582458

[HD96] J. A. P. Heesterbeek and K. Dietz, *The concept of R_0 in epidemic theory*, Statist. Neerlandica **50** (1996), no. 1, 89–110, DOI 10.1111/j.1467-9574.1996.tb01482.x. MR1381210

[KM27] W. O. Kermack and A. G. McKendrick, *A contribution to the mathematical theory of epidemics*, Proc. Roy. Soc. London **115** (1927), 700–721.

[KMS17] István Z. Kiss, Joel C. Miller, and Péter L. Simon, *Mathematics of epidemics on networks: From exact to approximate models*, Interdisciplinary Applied Mathematics, vol. 46, Springer, Cham, 2017, DOI 10.1007/978-3-319-50806-1. MR3644065

[Lyo11] Russell Lyons, *The spread of evidence-poor medicine via flawed social-network analysis*, Statistics, Politics, and Policy **2** (2011), no. 1.

[Wik21] Wikipedia contributors, *Diffusion of innovations — Wikipedia, the free encyclopedia*, 2021. [Online; accessed 2-June-2021].

David Aldous

Credits

Figure 1 is courtesy of Google Books Ngram Viewer, `https://books.google.com/ngrams`.

Photo of David Aldous is courtesy of David Aldous.