Contagion dynamics in real-life and digital settings

Adam J. Kucharski[1*]

[1]Centre for the Mathematical Modelling of Infectious Diseases, London School of Hygiene & Tropical Medicine, UK.

*To whom correspondence should be addressed. Email: adam.kucharski@lshtm.ac.uk.

Abstract

Many fields face the challenge of understanding the diffusion of events and behaviours within a host population, from finance and politics to technology and health. In recent years, novel methodology and data have created new opportunities to study how inter-host contagion processes shape diffusion, and how these processes influence the success of interventions to seed or prevent such contagion. Given the complexity of identifying and analysing contagion, attention has typically focused on specific situations and settings. However, through a combination of large-scale data, mathematical modelling, and statistical inference, there is increasingly potential to characterise key aspects of contagion dynamics across fields and identify areas for linked research efforts. Such progress will be crucial for improving predictions of diffusion, design of seeding strategies, and evaluation of control measures.

Main text

The idea that events or behaviours can diffuse through a population via an inter-host transmission process has itself spread widely, from studies of voting (*1*) and suicide (*2*) to product purchases (*3*) and violence (*4*). Efforts to investigate such diffusion have gradually evolved from describing broad patterns of spread to examining the potential underlying processes driving diffusion. Within this wider field of diffusion research, a large number of studies have focused on contagion processes, whereby hosts in a given 'affected' state can transmit this state to other unaffected hosts (*5*). This work has included

quantification of the magnitude of contagion, as well as attempts to reconstruct its path. As online connectivity has grown in recent decades, new opportunities for the diffusion of events within population have also arisen, alongside novel data sources with which to study the potential role of contagion processes (*6, 7*).

Historically, practical and methodological constraints have meant that research into contagion has typically focused on quantifying contagion in specific settings, such as pairwise person-to-person transmission events in households or friendship groups (*8*). In recent years, however, new methodological developments (*9–11*), the rising availability of detailed network data (*12–14*), and ability to run large-scale online experiments (*15–18*) have led to substantial advances in the analysis of contagion dynamics. These developments are providing opportunities to quantify contagion in broader populations and obtain generalisable insights into inter-host transmission.

The ability to quantify the dynamics of contagion has long been a central component of infectious disease outbreak analysis (*19*). Such knowledge has made it possible to characterise and compare different outbreaks, stimulating methodological developments and research insights that have been successfully applied across pathogens and settings; the rapid growth in disease modelling in recent years has benefitted greatly from identification of these common underlying methods and principles (*20, 21*). In turn, this had led to an improved ability to forecast outbreak dynamics and evaluate the comparative effectiveness of potential control measures. As a result, policy decisions around infectious disease interventions are increasingly based on outbreak analysis and quantitative modelling (*19*). Similar approaches are now becoming feasible for other forms of contagion: recent progress in studies of contagion dynamics has spanned fields ranging from technology (*22*) and health (*17*) to violence (*23*) and politics (*24*). There is also growing awareness of how one form of contagion may shape another, as with the spread of online misinformation and outbreaks of vaccine-preventable diseases (*25*). As knowledge of contagion processes within and between disciplines improves, there will increasing potential to link insights and methods across fields, improving our ability to uncover the processes driving observed diffusion, forecast contagious events, and inform the design of effective interventions.

From connections to contagion

Despite diversity in the nature and effects of contagion in different settings, analysis efforts share a fundamental challenge: how to reconstruct the underlying unidirectional process of inter-host transmission from available data on the diffusion of a particular state within a population. Specifically, there is a need to identify how connections between hosts in some affected state (i.e. 'cases') are related to inter-host contagion events, which are typically unobserved (Figure 1A). Because individual may share a state for non-contagious reasons, such as homophily – whereby similar hosts are more likely to form links – or common exposure to an external source, the first step is to account for such confounding. One approach to separating contagion effects from alternative explanations has been to use prospective randomised experiments (*10*). These experiments have either randomised at the host-level – by introducing a specific state to a subset of hosts – or at the network level, by breaking and forming new network links. Such randomisation has historically been impractical to implement for many forms of contagion, but there have been two major sources of progress. First, host or network-level randomisation can be obtained via natural experiments rather than being imposed experimentally (*11*); for example, changes in weather can alter individual exercise behaviours (*26*), and reassignment of military families can result in formation of new network links, which has made it possible to study the spread of specific health attributes (*27*). Second, online social networks have made it feasible to run large-scale randomised experiments in which both the full contact network and timing of user-level actions can be directly observed following the introduction of a specific state (*15, 16, 18*).

In many situations, it is not feasible to conduct prospective studies or to employ an appropriate source of randomisation when analysing contagion; studies have therefore relied on retrospective analysis of observational data, which means it is often challenging to rule out homophily or common exposure as alternative explanations (*5, 10*). However, there has also been progress with this form of analysis. With online diffusion, it can be possible to infer the chain of transmission, either by direct observation – as is the case with nominated-based content (*12, 28*) – or by making simplifying assumptions about how potential exposures via network contacts relate to transmission activity (*7*). Alterna-

tively, mathematical models such as self-exciting point processes can be used to estimate transmission chains from contact networks and time series data, and evaluate the plausibility of contagion as an explanation for observed diffusion patterns (*23, 29*) (Figure 1B). However, it is important when using such models to examine the strength of evidence for a contagion processes against a null hypothesis of no contagion. Robust detection of contagion also requires appropriate statistical methods, such as a likelihood function that reflects the probability distribution underlying stochasticity in transmission. For example, it may be necessary to use an overdispersed rather than Poisson offspring distribution when analysing the distribution of secondary cases (*30*). Unbinned likelihood methods can also help make optimal use of available data (*31*), while model comparison methods, such as the likelihood ratio test for nested models or information criteria for non-nested frameworks, can be used to assess the relative performance of different hypotheses about contagion effects (*32*).

Another important consideration when analysing social contagion is that diffusion may not occur via a simple, unique transmission route through a network. Whereas biological infections typically involve a single transmission event from one host to another, certain actions may depend on multiple exposures. The spread of malware and certain forms of online sharing behave like 'simple' biological contagions, allowing for a direct analogy with infectious disease transmission (*12*). However, behaviours such as adoption of technology or sharing of political content may act as 'complex contagions', requiring multiple exposures for contagion to occur (*22, 33*). This highlights the value of measuring the underlying network structure alongside transmission (Figure 1A), as diffusion may be influenced by a combination of exposure types, and these exposure types may have a similar influence on transmission among analogous forms of contagion.

Quantifying transmission dynamics

As insights into transmission dynamics grow across fields, there is increasing scope to quantify and compare different forms of contagion. Such comparison can help identify likely patterns of contagion-driven diffusion for different events. In the case of exponentially growing diffusion, a well-established metric for such comparison is the doubling

time (*34*). There is evidence this can vary substantially depending on the type of diffusion: outbreaks that spread through digital networks, such as malware or reposting of online content, can double on a timescale of seconds or minutes; in contrast, outbreaks of infectious diseases may take days or weeks to double in size, and innovations can take years (Figure 2A).

A major limitation of the doubling time as metric is that it is only applicable to events that are growing exponentially, which is neither a necessary nor sufficient condition for diffusion to be driven by a contagion process. Recent efforts to characterise contagion in different settings have obtained deeper insights by deconstructing transmission into two other metrics: the reproduction number and the serial interval. Originally employed in infectious disease epidemiology, the reproduction number, R, measures the average number of secondary cases generated by a typical infectious case while the serial interval gives the average time between onset of symptoms in an infector and the infectee. The threshold $R = 1$ provides a natural separation between two transmission regimes: supercritical transmission ($R > 1$), with the expectation of exponential growth, and subcritical ($R < 1$), with an expected decline in infection over time.

One advantage of such information is it separates the magnitude of contagion from the timescale. Estimation of R and serial interval for different forms of contagion can therefore allow identification of key transmission characteristics (Figure 2B). Transmission of online content generally occurs on a fast timescale, albeit with a relatively low R; there is evidence that even the largest information cascades on Facebook during 2014–16 had R of around 2 (*12*), with the serial interval highly dependent on the sharing method. Studies of contagious violence suggest subcritical transmission is common, resulting in small 'stuttering' outbreaks rather than large epidemics; the serial interval for such transmission can be weeks or months (*23*). Contagious physical reactions such as yawns and laughter can have a serial interval in the order of seconds or minutes, much like transient forms of online contagion. At the other extreme, it has been estimated that many prominent scientific ideas had a high R, but – given the nature of the scientific citation process – a very long serial interval (*35*).

Forms of contagion with a large value of R and small serial interval will typically

lead to rapid diffusion, with events occurring at high rate per unit time. Such events are generally well characterised, because studies require few seed nodes and a relatively short period of observation to accumulate sufficient data (Figure 2B). In contrast, a low value of R and long serial interval will lead to rare – and likely often undetected – instances of contagion. A prospective study would therefore require multiple seed nodes as well as a long duration of observation. Combined with the limited feasibility of conducting randomised experiments for many form of potential contagion, this can lead to substantial variability in both the volume of available evidence and uncertainty associated with quantitative estimates of contagion dynamics (*35*).

Although specific types of outbreak – such as infectious diseases or online content sharing – tend to cluster together in the R and serial interval parameter space (Figure 2B), there are examples of overlap between categories. Based on these metrics, e-mail marketing chains are closer to subcritical contagion processes like monkeypox and gunshot violence than the spread of popular Facebook content. Identification of such similarities can allow for common methodology: if transmission is subcritical, outbreak size can be used to estimate R for emerging pathogens (*30*), gunshot violence (*23*), and e-mail marketing campaigns (*36*). Likewise, it can be possible to capture the simple contagion driving online nomination-based games, such as the ice bucket challenge, with adapted forms of the transmission dynamic models used to study acute outbreaks such as pandemic influenza (*37*).

By comparing the structure of the host contact network with the inferred transmission chains (*12, 17*), it is possible to further deconstruct the reproduction number into its four main components: mean duration of infectiousness (d); effective rate of contact between individuals per unit time (c); probability of transmission per contact (β); and mean proportion of the population who are susceptible (S). Because the reproduction number is the product of these four values, i.e. $R = d \times \beta \times c \times S$, such information can identify the main processes driving transmission. For example, among large Facebook cascades, there is evidence of an inverse relationship between β and c; during cascades of nomination-based games, only a few users were exposed on average, but many subsequently adopted, whereas popular transient posts on a Newsfeed were typically seen by a large number

of people, but the probability of adoption was very low (*12*). It has been observed than transient content on other platforms has a similarly low β, with cascades typically driven by a high effective contact rate, c (*7, 38*).

In real-life networks, a large value of c is often the result of heterogeneity in the degree distribution (Table 1). Such heterogeneity increases the effective contact rate because highly connected nodes are both more likely to be exposed and to expose others. This network property, originally popularised in the context of HIV transmission modelling (*39*), has since been used to explain the persistence of malware online (*6*) as well as the structure of risk in financial networks (*40*). This variance in degree distribution also gives rise to the 'friendship paradox', whereby, on average, an individuals' friends have more friends than they do (*41*). During an outbreak, heterogeneity in the degree distribution can lead to substantial variation in R at the individual level, with a small number of individuals generating superspreading or 'broadcast' events that produce a large number of secondary cases (*23, 30, 40, 42*). The structure of the underlying contact network can influence the broader transmission dynamics as well. Assortative networks – with highly connected nodes mostly connected to other highly connected ones – can generate rapid but constrained outbreaks, whereas disassortative networks can produce slower, larger outbreaks (*43*). With improved measurement of real-life networks, there is potential to gain a better understanding of what contagion dynamics should be expected; there is evidence that real-life human communities and social networks tend to be assortative (*44–46*), whereas the interbank loan network and the world wide web are dissassortative (*47, 48*).

Other network features can also influence outbreak dynamics. 'Small-world' links can create exposures between localised clusters, enabling transmission to spread faster to different parts of a network (*49, 50*). However, in the case of complex contagion, such links may inhibit transmission, because they do not allow sufficient social reinforcement form multiple contacts (*33*). Clustered networks can therefore enhance the diffusion of complex contagion by providing such reinforcement, while inhibiting the spread of simple contagion. Studies have also found that highly clustered networks can enable co-existence of competing pathogen strains (*51*); online information also has to compete for users'

attention (*52*), suggesting that clustered online networks may be more likely to sustain multiple competing forms of content (*53*). There is also evidence of an inverse correlation between susceptibility and influence on social media (*54*), which can reduce the relative role of highly connected nodes in the diffusion of contagious states (*26*).

As well as network structure and susceptibility, the reproduction number also depends on the inherent transmissibility of a contagious process, which is captured by β. For example, the characteristics of online content can influence the extent of its spread, with content that generates a strong emotional reaction more likely to be shared (*14, 55*). Similarly, the transmissibility of information in real-life can depend on the style and features of the message being communicated (*56*). There are also many documented instances of the appearance of online content evolving in a manner that increases transmissibility (*57*). However, there may still be strong underlying constraints on the extent of overall transmissibility that is possible with a given online sharing protocol (*12, 17*). With better characterisation of the relationship between network structure and other drivers of transmission, it will be possible to further elucidate the impact of these potential constraints in different settings.

Forecasting the dynamics of contagion

The tools required to characterise transmission can also be applied to questions of forecasting (*23, 58*). There are several ongoing challenges involved in forecasting the evolution of a contagion process, with three main factors influencing the success of efforts to date. First, there is the issue of data availability, with effective predictions reliant on the openness, timeliness and completeness of relevant data sources. If these data limitations are accounted for, it is likely to increase uncertainty in conclusions; if data limitations are not considered, it could lead to biased or otherwise flawed results. Even with access to high resolution public social media data, there are limitations to the specific features and interactions that can be extracted, particularly with regards to potential routes of exposure (*59*).

A second limiting factor is the extent to which a diffusion process is driven by predictable transmission mechanisms. Even when detailed data are available – as for online

platforms – if diffusion is strongly influenced by extrinsic randomness, there will be an inherent limit to *ex ante* prediction accuracy (*59*). In some cases, predictive performance can be improved by incorporating information on early patterns of diffusion with a 'peeking' approach (*60, 61*). However, given the often rapid timescales of digital diffusion (Figure 2B), such approaches may in general be more of theoretical rather than practical interest. Alternatively, if transmission follows a process that can be easily modelled mechanistically with a transmission dynamic model, as with nomination-based games online, it can enable *ex ante* predictions about diffusion (*37*). Better characterisation of similarities in transmission dynamics (Figure 2) could allow for further interdisciplinary applications of mechanistic forecasting models, especially for transmission involving simple contagion, which in many cases has a direct biological analogy. Deconstruction of the reproduction number could also benefit prediction efforts, because certain components of R will influence diffusion in predictable ways, such as the lack of susceptibility following a large cascade inhibiting recurrent contagion in the same population (*62*). Such approaches could allow for broad conclusions to made about transmission potential, even if some components of R cannot be independently estimated. However, some prediction tasks may still be more feasible than others; it has been suggested that it may be easier to predict the size of large online cascades than their temporal shape (*60*). These constraints may reduce the usefulness of predictions for outbreak planning in situations where the distribution of events over time influences their overall impact, such as with misinformation during elections (*63*).

A final limitation is the interpretability of the prediction task itself. Despite increasing interest in predicting social and digital contagion, approaches for evaluating the performance of these predictions have varied greatly, with studies using a range different metrics, data sources and methods (e.g. *ex ante* vs peeking) (*64*). Better standardisation would improve reproducibility and interpretation of results, as well as helping to identify the best promising approaches for future prediction efforts. (*24*)

Seeding contagion

As knowledge of a contagion process improves, it can help inform decisions about interventions (*65, 66*). When contagion in a network is desirable – as can be the case with adoption of specific resources that improve health or bring other benefits (*17, 67*) – a key question is how to choose seed nodes to initiate transmission (*68*). In recent years, several studies have expanded on what has historically been a limited empirical evidence base for optimal seeding. In particular, randomised experiments have found that selecting seed nodes based on simple measures of network connectivity like the degree of a node does not necessarily lead to increased diffusion, and may actually perform worse than seeding at random (*67, 69*). This is likely to be due to network features such as assortativity, which can lead to redundant clustering among seeds (*26, 67*). Clustering can also have an impact on the success of seeding complex contagions, which typically require social reinforcement to diffuse (*70*). However, such experiments can involve substantial effort and expense, particularly when conducted in face-to-face communities.

Selection of nodes can exhibit a 'submodular' property, whereby there can be diminishing returns to selecting larger numbers of nodes (*71*). Given the practical challenges of measuring and monitoring full contact networks, there has been increasing attention on approximate methods for identifying optimal seed nodes. A prominent example is the nomination-based approach, whereby a subset of participants each nominate a random contact (*72*); in networks with a heavy-tailed degree distribution, this approach will produce a set of individuals with much higher degree than average (*73*). This feature arises directly from the friendship paradox (*41*). As well as increasing the reach contagion in the limited randomised experiments that have assessed the method (*67, 74*), such approaches have the advantage of introducing an element of randomness, thereby reducing the chance of choosing seeds in a small clustered part of the network. Randomness can also have implications for the cost-effectiveness of seeding strategies. Because there can be a high level of variation in the extent of transmission – even for supposedly influential seeds – it may be more cost-effective to seed contagion via a large number of less prominent individuals to better capture average effects (*75*). This suggestion is consistent with the observation that high degree nodes are not necessarily optimal seeds in real-life networks,

and further illustrates the limitations of focusing only on egocentric network properties when designing interventions.

Recent efforts have extended experimental seeding strategies beyond analysis of network structure, with the aim of targeting the other components of the reproduction number, such as the duration of contagiousness or probability of transmission (*17*). By identifying similarities in the transmission dynamics of different forms of contagion, the conclusions of such studies could be further generalised beyond the specific intervention tested. Simulation studies could also be employed to provide relatively cheap, rapid insights into how optimal seeding depends on the different features observed in real-life networks (*26, 76, 77*), generating testable hypotheses about novel candidate interventions that could then be tested in field experiments.

Preventing contagion

As with strategies to seed contagion, interventions aiming to control contagious processes typically work by targeting one or more components of the reproduction number. In reality, some components may be easier to target than others. In the example of HIV, case identification and antiretroviral therapy can reduce the duration of infection; reduction in partners decreases the contact rate; condoms reduce transmission probabiliity; and pre-exposure prophylaxis can reduce susceptibility (*78*). Taken alone, the first two approaches have historically been insufficient to contain most outbreaks, leading to increased focus on latter two interventions as well. Other fields have similarly needed to identify which components of transmission are optimal to target with interventions. In finance, ring-fencing between investment and retail banking has been designed to remove potential routes of transmission (*79*), while capital requirements aim to reduce the susceptibility of individual banks (*40*). On online platforms, the diversity of interventions has spanned the full transmission process, albeit with varying success (*80, 81*). Some efforts to tackle contagion have aimed to reduce the number of people a user can share content with (*80*) or reduce the probability of interacting with such content (*82*); others have introduced pre-emptive messaging against misinformation, in an attempt to lower susceptibility (*83, 84*), or reduced the duration of time for which harmful content is visible (*81, 85*). In

some cases, intervention options depend on the user data that is available. On encrypted social media platforms, it is not possible to analyse the information being shared, which makes it difficult to identify and remove harmful content; such platforms have therefore adopted broader approaches such as reducing the overall contact rate (*80*). With a better understanding of the relative importance of the different components of transmission – and feasibility of targeting each one – there will be increasing scope to develop reliable *ex ante* predictions about which interventions will be most effective.

Knowledge of the underlying transmission dynamics can also help inform the timing and type of control measures. Interventions against contagion can be either reactive (i.e. rolled out during an outbreak) or pre-emptive (i.e. introduced ahead of time). For example, if an infectious disease has a long serial interval and clearly defined symptoms, contact tracing, quarantine and treatment can be an effective reactive strategy mid-outbreak, particularly if a pathogen has a low R (*21*). The existence of similar characteristics for gunshot violence (Figure 2B) provides theoretical support for violence interruption methods (*86*); these interventions implicitly rely on the fact that the serial interval between retaliatory attacks is long enough to intervene.

If an outbreak grows quickly – as with online sharing of information – successfully identifying and responding to contagious events in real-time can be far more challenging (*87*). In this case, proactive measures may therefore be more successful. This might involve efforts to reduce susceptibility, or restructuring the network to limit features that could amplify contagion (Table 1). Broader approaches like these may also be required for certain network structures. If contagion is driven by a high effective contact rate, rather than a high inherent transmissibility, a large number of nodes will need to be targeted to prevent each subsequent case. In such situations, it may be more efficient to employ a mass intervention than attempt to follow up every potential contact.

Evaluations of interventions for contagion should account for the dynamics of transmission to the fullest extent possible. If a behaviour is contagious, a control measure will have a direct effect – protecting individuals who receive the intervention – as well as an indirect effect, preventing onward transmission. In social science, the concept of a 'network multiplier' has been proposed to capture the indirect impact of an intervention on

first degree contacts of a case; the aim is for indirect effects to be measurable and comparable across interventions, from voting messages to health campaigns (*8*). As knowledge of transmission mechanisms develops further, generalisable methods like these could be extended to examine subsequent generations of transmission as well. For subcritical transmission, a single case with $R < 1$ will on average generate an outbreak with a total of $1/(1 - R)$ cases, assuming transmission events are independent (*36*). This provides a straightforward estimate of the indirect effect; preventing a single case will also prevent $1/(1 - R) - 1$ additional cases overall on average, which is more than the R first generation cases as measured by the network multiplier. In some situations, particularly when R is larger, network structure will have a non-negligible effect on transmission, which means more detailed studies will be required to rigorously assess effectiveness for many forms of contagion. This could include empirical methods such as randomised trials (*88*), as well as mathematical modelling studies, which are routinely used to inform vaccination policy (*89*), and which have been applied to study the effectiveness of different seeding strategies for social contagion (*26, 77*). As a more detailed quantitative understanding of social and digital contagion develops, it will be possible to assess the effectiveness of a range of historical interventions, as well as predict the likely impact of future ones. Such developments could substantially improve the ability to make evidence-based decisions about the management and control of contagious processes.

Future directions

There are several areas for future development in our understanding of contagion. First, there is a need to expand analysis of contagion beyond specific localised settings. Studies of contagion have historically been limited by the ability to observe outbreaks within wider populations, with data often subject to selection bias and missingness. This can also introduce selection bias in research, resulting disproportionate focus on certain settings. For example, research into the diffusion of information on social media has grown dramatically in recent years (Figure 3A), but this research has not been evenly distributed between platforms. Comparing research activity with active users, there are two extremes: Twitter, where posting is mostly public, is over-represented and WhatsApp, which is en-

crypted, is under-represented (Figure 3B). Given the potential for cross-platform transmission of content, if the selection bias in research activity translates into a bias in targeting of interventions against harmful content, it may undermine the effectiveness of these interventions (*90*). The ability to accurately measure factors driving contagion can also affect predictions about the impact of interventions; if data on network structure are required for risk assessment – as in central banking – measurement errors can result in biased estimates of risk (*91*), just as flawed assumptions or estimates in the analysis of STI transmission can result in incorrect conclusions about transmission potential (*39*).

A second area for development is linking measured contagion with its subsequent effects. For example, studies of online contagion have typically focused on events that are relatively easily to measure, such as sharing of content (*12, 24, 42, 60*), even though such events may be a poor proxy for underlying outcomes of interest, such as consumption of information (*92*) or voting decisions (*18*). Bridging this gap will require linked data on temporal measurements of contagion and resulting behaviours or actions, combined with causal inference, as well as identification of key summary indicators that can be deployed more broadly to track behaviours of interest. In recent years, there has been considerable progress in the design and implementation of experiments on online platforms (*10, 93*); by reducing barriers to implementation, field experiments could help further characterise the links between contagion and its consequences (*10*). In particular, such work could provide valuable insights into the real-life impact of information diffusion, from health protection (*25*) to election integrity (*63*).

A third area for development is the use of generalisable insights to evaluate prospective interventions against the spread of contagion. Such efforts could span violence (*23, 94*), health (*5, 9*), misinformation (*83*) and malware (*95*), and could include combinations of empirical and simulation studies (*26, 90, 96, 97*), with results feeding into cost-effectiveness analysis for different interventions, as is common in infectious disease epidemiology. Despite an extensive theoretical literature on malware transmission modelling, for example, few studies have applied these models to data to infer the properties of real-life malware outbreaks or their impact (*98*). As a result, there is still a relatively limited data-driven evidence base for interventions against different forms of malware

transmission, other than attempts to reduce susceptibility of individual machines. At the other end of the temporal scale, contagion with a long serial interval, such as intergenerational transmission of violence, has historically been challenging to characterise (*99*). With improved analysis methods for long-term transmission dynamics, it may be possible to quantify other forms of social contagion that are currently little studied – or perhaps not even identified yet – and propose novel interventions targeting key aspects of this contagion. Such methodology and evidence could also be linked across fields. If one form of contagion influences another, joint studies of such processes may prove more productive than unilateral efforts: the spread of vaccine misinformation online may influence disease transmission (*25*); the diffusion of news reports may shape the risk of contagion in mass shootings (*29*); and liquidity hoarding resulting from indirect confidence effects may exacerbate the direct impact of loan losses resulting from bank failures (*40*). These examples have involved collaborations between researchers working on social media, infectious disease epidemiology, crime, and finance. These joint approaches could help identify measurable forms of contagion that have a disproportionate impact on the overall burden and cost of an outbreak. Quantification of such feedback processes could therefore inform broader risk estimation efforts: national risk assessments typically have to compare diverse social, biological and digital threats, including contagious – and potentially interdependent – events such as public disorder, infectious disease pandemics, and cyber attacks (*100*).

Finally, there is need for an evidence base that operates on a similar timescale to key contagious threats, if not fully in real-time. Despite concerns about misinformation during the 2016 US Presidential Election, for example, many high profile studies examining this issue were not published until 2018 or 2019 (*14, 24, 92, 101*). Although some types of study would not be feasible to conduct on faster timescales, where possible the research community should find ways to maximise availability of evidence in the early stages of an outbreak, when it could have a greater impact than a later retrospective analysis, while also accounting for the inevitable uncertainties involved in real-time data streams (*102*). This could include real-time monitoring and quantifying of contagion on networks of interest (*103, 104*), as well as studies exploring potential pre-emptive interventions for

future outbreaks. As with infectious diseases, preventative approaches like these are likely to be far more cost-effective than a delayed, reactive response.

References

1. D. W. Nickerson, *American political Science review* **102**, 49 (2008).
2. M. Gould, P. Jamieson, D. Romer, *American Behavioral Scientist* **46**, 1269 (2003).
3. F. M. Bass, *Management science* **15**, 215 (1969).
4. M. I. Midlarsky, M. Crenshaw, F. Yoshida, *International Studies Quarterly* **24**, 262 (1980).
5. N. A. Christakis, J. H. Fowler, *Statistics in medicine* **32**, 556 (2013).
6. R. Pastor-Satorras, A. Vespignani, *Physical Review Letters* **86**, 3200 (2001).
7. S. Goel, A. Anderson, J. Hofman, D. J. Watts, *Management Science* **62**, 180 (2015).
8. T. J. VanderWeele, N. A. Christakis, *International Journal of Epidemiology* (2019).
9. S. Aral, C. Nicolaides, *Nature Communications* **8** (2017).
10. S. J. Taylor, D. Eckles, *Complex Spreading Phenomena in Social Systems* (Springer, 2018), pp. 289–322.
11. E. L. Ogburn, *Complex Spreading Phenomena in Social Systems* (Springer, 2018), pp. 47–64.
12. J. Cheng, *et al.*, *Twelfth International AAAI Conference on Web and Social Media* (2018).
13. S. B. Rosenthal, C. R. Twomey, A. T. Hartnett, H. S. Wu, I. D. Couzin, *Proceedings of the National Academy of Sciences* **112**, 4690 (2015).
14. S. Vosoughi, D. Roy, S. Aral, *Science* **359**, 1146 (2018).
15. D. Centola, *Science* **329**, 1194 (2010).

16. A. D. Kramer, J. E. Guillory, J. T. Hancock, *Proceedings of the National Academy of Sciences* **111**, 8788 (2014).

17. N. K. Cobb, M. A. Jacobs, P. Wileyto, T. Valente, A. L. Graham, *American journal of public health* **106**, 1130 (2016).

18. R. M. Bond, *et al.*, *Nature* **489**, 295 (2012).

19. H. Heesterbeek, *et al.*, *Science* **347**, aaa4339 (2015).

20. N. M. Ferguson, *et al.*, *Nature* **425**, 681 (2003).

21. C. Fraser, S. Riley, R. Anderson, N. Ferguson, *Proceedings of the National Academy of Sciences of the United States of America* **101**, 6146 (2004).

22. D. Guilbeault, J. Becker, D. Centola, *Complex Spreading Phenomena in Social Systems* (Springer, 2018), pp. 3–25.

23. B. Green, T. Horel, A. V. Papachristos, *JAMA Internal Medicine* **177**, 326 (2017).

24. N. Grinberg, K. Joseph, L. Friedland, B. Swire-Thompson, D. Lazer, *Science* **363**, 374 (2019).

25. P. Piot, *et al.*, *Nature* **575**, 119 (2019).

26. S. Aral, P. S. Dhillon, *Nature Human Behaviour* **2**, 375 (2018).

27. A. Datar, N. Nicosia, *JAMA Pediatrics* **172**, 239 (2018).

28. M. Y. Ni, B. H. Y. Chan, G. M. Leung, E. H. Y. Lau, H. Pang, *BMJ* **349**, g7185 (2014).

29. S. Towers, A. Gomez-Lievano, M. Khan, A. Mubayi, C. Castillo-Chavez, *PLOS ONE* **10**, e0117259 (2015).

30. J. O. Lloyd-Smith, S. J. Schreiber, P. E. Kopp, W. M. Getz, *Nature* **438**, 355 (2005).

31. S. Towers, A. Mubayi, C. Castillo-Chavez, *PloS one* **13** (2018).

32. K. Burnham, D. Anderson, *Model selection and multimodel inference: a practical information-theoretic approach.* (Springer-Verlag, New York, 2002), second edn.

33. D. Centola, M. Macy, *American Journal of Sociology* **113**, 702 (2007).

34. C. Starr, R. Rudman, *Science* **182**, 358 (1973).

35. Materials and methods are available as supplementary material on Science Online.

36. D. J. Watts, J. Peretti, M. Frumin, *Viral marketing for the real world* (Harvard Business School Pub., 2007).

37. D. A. Sprague, T. House, *PloS one* **12**, e0180802 (2017).

38. M. De Domenico, A. Lima, P. Mougel, M. Musolesi, *Scientific reports* **3**, 2980 (2013).

39. R. M. May, R. M. Anderson, M. Irwin, *Philosophical Transactions of the Royal Society of London. B, Biological Sciences* **321**, 565 (1988).

40. A. G. Haldane, R. M. May, *Nature* **469**, 351 (2011).

41. S. L. Feld, *American Journal of Sociology* **96**, 1464 (1991).

42. S. Goel, D. J. Watts, D. G. Goldstein, *Proceedings of the 13th ACM conference on electronic commerce* (ACM, 2012), pp. 623–638.

43. S. Gupta, R. M. Anderson, R. M. May, *AIDS (London, England)* **3**, 807 (1989).

44. J. Ugander, B. Karrer, L. Backstrom, C. Marlow, *arXiv* (2011).

45. C. L. Apicella, F. W. Marlowe, J. H. Fowler, N. A. Christakis, *Nature* **481**, 497 (2012).

46. N. Meghanathan, *Computer and Information Science* **9**, 7 (2016).

47. M. E. Newman, *Physical Review Letters* **89**, 208701 (2002).

48. K. Soramäki, M. L. Bech, J. Arnold, R. J. Glass, W. E. Beyeler, *Physica A: Statistical Mechanics and its Applications* **379**, 317 (2007).

49. M. E. Newman, D. J. Watts, *Physical review E* **60**, 7332 (1999).

50. A. Dhand, *et al.*, *Nature Communications* **10**, 1206 (2019).

51. C. Buckee, K. Koelle, M. Mustard, S. Gupta, *Proceedings of the National Academy of Sciences* **101**, 10839 (2004).

52. P. Lorenz-Spreen, B. M. Mønsted, P. Hövel, S. Lehmann, *Nature communications* **10**, 1759 (2019).

53. A. Kucharski, *Nature* **540**, 525 (2016).

54. S. Aral, D. Walker, *Science* **337**, 337 (2012).

55. J. Berger, K. L. Milkman, *American Marketing Association* (2011).

56. A. Mesoudi, A. Whiten, *Philosophical Transactions of the Royal Society B* **363**, 3489 (2008).

57. L. A. Adamic, T. M. Lento, E. Adar, P. C. Ng, *WSDM'16* (2016).

58. A. Tatar, M. D. De Amorim, S. Fdida, P. Antoniadis, *Journal of Internet Services and Applications* **5**, 8 (2014).

59. T. Martin, J. M. Hofman, A. Sharma, A. Anderson, D. J. Watts, *Proceedings of the 25th International Conference on World Wide Web* (International World Wide Web Conferences Steering Committee, 2016), pp. 683–694.

60. J. Cheng, L. Adamic, P. A. Dow, J. M. Kleinberg, J. Leskovec, *Proceedings of the 23rd international conference on World wide web* (ACM, 2014), pp. 925–936.

61. B. Shulman, A. Sharma, D. Cosley, *Tenth International AAAI Conference on Web and Social Media* (2016).

62. J. Cheng, L. A. Adamic, J. Kleinberg, J. Leskovec, *Proceedings of the 25th International Conference on World Wide Web (WWW'16)* (2016), pp. 671–681.

63. S. Aral, D. Eckles, *Science* **365**, 858 (2019).

64. J. M. Hofman, A. Sharma, D. J. Watts, *Science* **355**, 486 (2017).

65. T. W. Valente, *Science* **337**, 49 (2012).

66. E. Muller, R. Peres, *International Journal of Research in Marketing* **36**, 3 (2019).

67. D. A. Kim, *et al.*, *The Lancet* **386**, 145 (2015).

68. D. Kempe, J. Kleinberg, É. Tardos, *Proceedings of the ninth ACM SIGKDD international conference on Knowledge discovery and data mining* (2003), pp. 137–146.

69. A. Banerjee, A. G. Chandrasekhar, E. Duflo, M. O. Jackson, *Science* **341**, 1236498 (2013).

70. D. Centola, *How behavior spreads: The science of complex contagions*, vol. 3 (Princeton University Press, 2018).

71. J. Leskovec, *et al.*, *Proceedings of the 13th ACM SIGKDD international conference on Knowledge discovery and data mining* (2007), pp. 420–429.

72. N. A. Christakis, J. H. Fowler, *PloS one* **5** (2010).

73. R. Cohen, S. Havlin, D. Ben-Avraham, *Physical Review Letters* **91**, 247901 (2003).

74. E. L. Paluck, H. Shepherd, P. M. Aronow, *Proceedings of the National Academy of Sciences* **113**, 566 (2016).

75. E. Bakshy, J. M. Hofman, W. A. Mason, D. J. Watts, *Proceedings of the fourth ACM international conference on Web search and data mining* pp. 65–74 (2011).

76. D. B. Bahr, R. C. Browning, H. R. Wyatt, J. O. Hill, *Obesity* **17**, 723 (2009).

77. S. Iyer, L. A. Adamic, *International Conference on Complex Networks and their Applications* (Springer, 2018), pp. 273–286.

78. Centers for Disease Control and Prevention, `https://wwwn.cdc.gov/hivrisk/decreased_risk` (2009).

79. J. S. Vickers, *Independent Commission on Banking final report: recommendations* (The Stationery Office, 2011).

80. C. Arun, *Economic & Political Weekly* **54** (2019).

81. A. Friggeri, L. Adamic, D. Eckles, J. Cheng, *Eighth International AAAI Conference on Weblogs and Social Media* (2014).

82. H. Allcott, M. Gentzkow, C. Yu, *Research and Politics* (2019).

83. M. Takayasu, *et al.*, *PLOS ONE* **10**, e0121443 (2015).

84. J. Roozenbeek, S. van der Linden, *Palgrave Communications* **5**, 12 (2019).

85. M. Kelly, *The Verge* (2019).

86. J. M. Whitehill, D. W. Webster, S. Frattaroli, E. M. Parker, *Journal of urban health* **91**, 84 (2014).

87. J. Kahn, *Bloomberg News* (2019).

88. R. Campbell, *et al.*, *The Lancet* **371**, 1595 (2008).

89. M. Baguelin, *et al.*, *PLOS Medicine* **10** (2013).

90. N. Johnson, *et al.*, *Nature* pp. 1–5 (2019).

91. S. Battiston, G. Caldarelli, R. M. May, T. Roukny, J. E. Stiglitz, *Proceedings of the National Academy of Sciences* **113**, 10031 (2016).

92. C. Bail, *et al.*, *PNAS* (2019).

93. M. J. Salganik, *Bit by bit: social research in the digital age* (Princeton University Press, 2017).

94. G. Wood, A. Papachristos, *Nature Human Behaviour* (2019).

95. M. Antonakakis, *et al.*, *26th USENIX Security Symposium USENIX Security 17* (2017), pp. 1093–1110.

96. J. D. Farmer, D. Foley, *Nature* **460**, 685 (2009).

97. D. S. Burke, *Science* **354**, 529 (2016).

98. A. M. del Rey, *Security and Communication Networks* **8**, 2561 (2015).

99. D. M. Patel, M. A. Simon, R. M. Taylor, *et al.*, *Contagion of violence: Workshop summary* (National Academies Press, 2013).

100. Cabinet Office, *https://www.gov.uk/government/publications/national-risk-register-of-civil-emergencies-2017-edition* (2017).

101. A. Guess, J. Nagler, J. Tucker, *Science Advances* **5**, eaau4586 (2019).

102. J. A. Polonsky, *et al.*, *Philosophical Transactions of the Royal Society B* **374**, 20180276 (2019).

103. A. Alaphilippe, C. Ceccarelli, L. Charlet, M. Mycielski, *EU DisinfoLab Report* (2018).

104. A. G. Haldane, *Workshop Socio-Economic Complexity, Lorentz Center, Leiden* (2015).

105. G. E. Leventhal, A. L. Hill, M. A. Nowak, S. Bonhoeffer, *Nat Commun* **6**, 6101 (2015).

106. C. Kubelec, F. Sá, *Bank of England Working Paper* (2010).

107. S. Blumberg, J. O. Lloyd-Smith, *PLoS Comput Biol* **9**, e1002993 (2013).

108. L. A. Adamic, *International Conference on Theory and Practice of Digital Libraries* (Springer, 1999), pp. 443–452.

109. S. Clémençon, H. De Arazoza, F. Rossi, V. C. Tran, *Social Network Analysis and Mining* **5**, 58 (2015).

Acknowledgements A.J.K. is supported by a Sir Henry Dale Fellowship jointly funded by the Wellcome Trust and the Royal Society (grant Number 206250/Z/17/Z).

Author contributions A.J.K. conceived the review article and wrote the manuscript.

Author information All data used in the analysis are publicly available. Correspondence should be addressed to A.J.K. (adam.kucharski@lshtm.ac.uk).

Supplementary materials

Materials and Methods

Tables S1 to S2

Figures/Tables

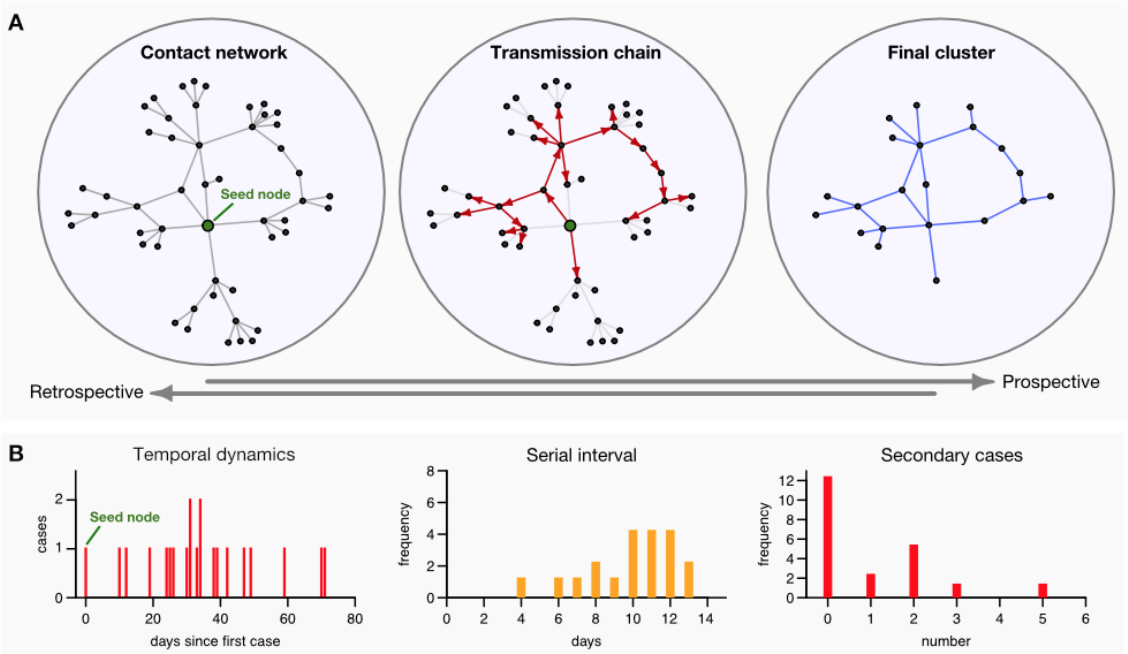

Figure 1: **Analysing the relationship between contact networks and transmission dynamics**. **(A)** Example scenario showing the three stages of a simple contagion process: an underlying contact network, a unidirectional transmission process through the network (red arrows), and the final cluster of cases linked by their original contacts (blue lines). In prospective randomised studies, one or more seed nodes will often be selected to initiate transmission, with subsequent transmission or cluster size used to measure contagion; in retrospective observational studies, the underlying transmission chain will typically need to be inferred from the final cluster data. **(B)** In retrospective studies, key summary statistics can be used to reconstruct the transmission process in **(A)**. If the temporal pattern of cases is known, for example, models can be used to infer the transmission chain and serial interval of infection (*23, 29*), and hence estimate the distribution of secondary cases arising from each infected node; the mean of this distribution is the reproduction number.

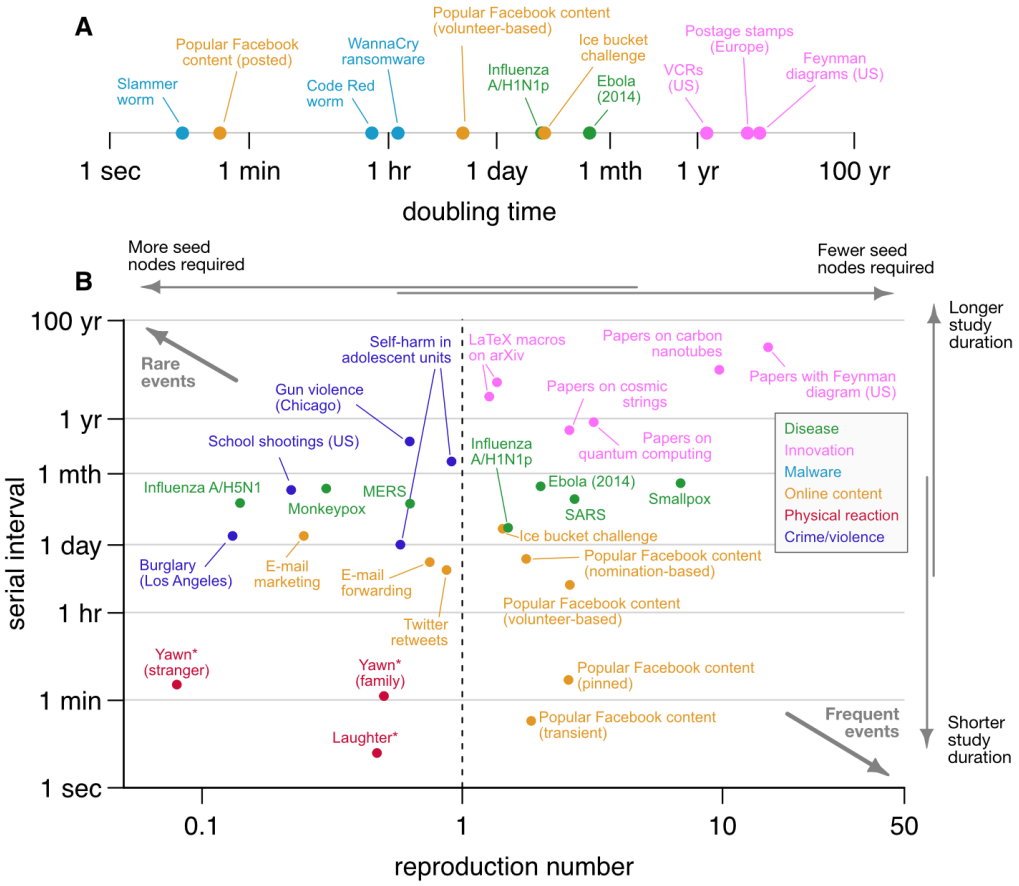

Figure 2: **Illustrative transmission characteristics of contagious events. (A)** Estimated doubling times for different diffusion events during their period of exponential growth. **(B)** Reproduction numbers and serial intervals for putative contagion processes in different settings, estimated from published datasets (*35*). Dashed line shows critical threshold separating exponential and subcritical outbreak growth. Diagonal arrows illustrate the expected frequency of contagious events for different parameter combinations, with vertical arrows on the axes showing the implications for prospective study designs. *Note these events were measured under specific experimental conditions, rather than in the wild; see (*35*) for more details.

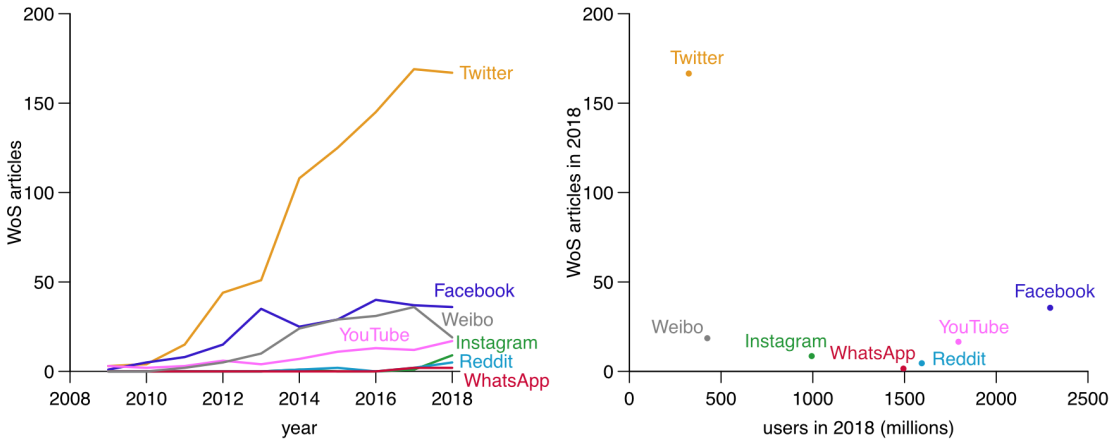

Figure 3: **Comparison of Web of Science papers on contagion and user activity on different online platforms, 2009–2018.** **(A)** Number of WoS articles studying contagion or diffusion on a given platform, by year of publication (*35*). **(B)** Comparison of the number of papers published in 2018 with number of monthly active users on each platform, or viewers in the case of YouTube.

Network property	Examples with property	Impact on contagion
Heavy-tailed degree distribution	Banks (*48*), online social networks (*44*), sexual contacts (*105*), financial networks (*106*).	Smaller transmission probability β required to sustain transmission (*6*); increased variance in outbreak size distribution (*107*).
Degree assortativity	Assortative: online social networks (*44*), human social contacts (*45, 46*). Disassortative: computer networks (*47*), interbank loan network (*48*).	Assortative: faster outbreak but smaller overall size. Disassortative: slower outbreak but larger size (*43*), world wide web.
Small-world links	International financial network (*106*), online social networks (*44*), world wide web (*108*).	Single small-world links can enable simple contagions to reach other parts of the network, while potentially inhibiting complex contagions (*22*).
Clustering	High clustering coefficient: human social contacts (*105*), online social networks (*44*), co-offender network (*23*), world wide web (*108*). Low clustering coefficient: sexual contacts (*105, 109*).	High clustering can be conducive for seeding complex contagions (*22*), and can enable co-existence of competing infections (*51*).
Joint distribution of susceptibility and influence	Online social networks (*54*).	An inverse association between susceptibility and influence can inhibit diffusion (*26*).

Table 1: **Common network properties and impact on contagion dynamics.** Many real-life networks will have a combination of these properties, which can influence the effectiveness of interventions such as seeding or control measures.