Scaling law for the impact of mutant contagion

Jonas S. Juul

Niels Bohr Institute, University of Copenhagen, Blegdamsvej 17, Copenhagen 2100-DK, Denmark

and

Center for Applied Mathematics, Cornell University, Ithaca, New York 14853, USA

Steven H. Strogatz

Center for Applied Mathematics, Cornell University, Ithaca, New York 14853, USA

(Dated: October 3, 2019)
Contagion, broadly construed, refers to anything that can spread infectiously from peer to peer. Examples include communicable diseases, rumors, misinformation, ideas, innovations, bank failures, and electrical blackouts. Sometimes, as in the 1918 Spanish flu epidemic, a contagion mutates as it propagates. Here, using a simple mathematical model, we quantify the downstream impact of a contagion that mutates exactly once as it travels. Assuming that this mutation occurs at a random node in the contact network, we calculate the distribution of the number of “descendants,” \( d \), downstream from the initial “Patient Zero” mutant. We find that the tail of the distribution decays as \( d^{-2} \) for complete graphs, random graphs, small-world networks and other infinite-dimensional networks. This prediction agrees with the observed statistics of memes propagating and mutating on Facebook, and is expected to hold universally for other effectively infinite-dimensional networks, such as the global human contact network. In a wider context, our approach suggests a possible starting point for a mesoscopic theory of contagion. Such a theory would focus on the paths traced by a spreading contagion, thereby furnishing an intermediate level of description between that of individual nodes and the total infected population. For every discipline concerned with contagion and its prevention, we anticipate that contagion pathways hold valuable lessons, given their role as the conduits through which single mutations, innovations, or failures can sweep through a network as a whole.

When a contagion spreads, it propagates from one or more “parent” nodes to a number of “descendant” nodes. Enumerating the descendants in all the paths stemming from a parent can reveal important and useful information. In particular, suppose the contagion mutates into a more pernicious form at some point along its travels. Then counting its descendants would tell us how many nodes will be confronted by this nastier strain. A mutation event of this sort occurred in 1918, and gave rise to the Spanish flu epidemic that killed millions of people worldwide. Similar (but less consequential) mutations happen online when users modify memes to make them funnier or stickier before sharing them with their peers.

To quantify the impact of such mutations, consider a simplified model of contagion in
which each node is either susceptible or permanently infected (Fig. 1). As the contagion spreads (Fig. 1(a)), we record which nodes caught it from which, and plot the resulting paths of infection as an epidemic tree (Fig. 1(b)). Then we count how many nodes would be affected by a mutation occurring at a random “Patient Zero” node. In the example shown in Fig. 1(c), the mutant infection occurs at node $B$ and is passed along to the two nodes below it. Of course, if the mutation had occurred elsewhere, it could have produced either more descendants (e.g., three descendants, had the mutation occurred at $A$) or fewer (zero descendants, had it occurred at $C$). Thus, the natural statistical quantity to study is the distribution of the number of descendants, aggregated over all possible Patient Zero nodes.

In one sense, the dynamics assumed here are trivial: one node after another gets infected until no susceptibles remain. But what is far from trivial are the descendant distributions implied by this model, as they depend on the network’s structure. To learn what to expect, we first compute descendant distributions numerically from Monte Carlo simulations. For a given random realization of the contagion process on a given network, like the one shown in Fig. 1(b), we count the number of descendants of each node and compile a histogram. This histogram, however, merely gives the descendant distribution for one realization of the dynamics. To extract a more robust statistical measurement, we average over the random location of the initially infected seed node, as well as the random decisions of whom to infect at each step, to obtain an average descendant distribution.

Figure 2 shows the average descendant distribution for the simplest possible network structure: a complete graph, in which each node is connected to all the others. The downward slope of the plot indicates that many nodes have few descendants, and a few nodes have many descendants. Of course, the seed $O$ has every other node as its descendant, as an artifact of the assumed initial conditions. Its corresponding data point in Fig. 2 lies off the curve for this reason.

The most striking feature of the descendant distribution in Fig. 2 is its apparent power-law decay for $d \gg 1$. To explain this scaling law intuitively, recall that one way of getting power-law distributions is through rich-get-richer effects, and observe:

(i) If node $i$ infected node $j$, the ancestors of $j$ will be $i$ and all the ancestors of $i$.

(ii) A node $i$ can acquire a new descendant $j$ if it passes the infection on to $j$, or if one of its descendants passes the infection on to $j$. 


FIG. 1. Simple model of contagion spreading on a network and its corresponding epidemic tree. Black filled circles denote susceptible nodes; red filled circles, infected nodes; red open circles, nodes infected by a mutant strain of the infection. a) Starting with a single infected seed $O$ at time $t = 0$, another node gets infected at random at the next time step. Any edge between an infected node and a susceptible node has an equal chance of being the next edge over which the contagion spreads. We keep track of which nodes transmitted and received the infection at every time step, until ultimately every node is infected. b) The epidemic tree shows who infected whom in the contagion process depicted in a). We draw this tree with the seed on top. The nodes that the seed infected are drawn in the second layer, and so on. A descendant of node $i$ is defined as any node that directly or indirectly received the infection from node $i$. Such a descendant node $j$ can be reached by starting at node $i$ and following a sequence of directed edges downward through the epidemic tree until the path ends at $j$. c) If a mutant infection occurs at some node ($B$, in the example shown here), that node passes the mutated strain on to all its descendants (two descendants, in this example).
The first point means that our model contagion process is equivalent to a network that grows by node copying. The second point suggests that the probability of a node acquiring more descendants should grow, loosely speaking, in proportion to the number of descendants it already has, thereby making the rich richer.

To sharpen this intuition, we calculate the descendant distribution $P_d$ analytically for some exactly solvable networks (for derivations, see Supplementary Sections 2, 3, 4, and 5). First, for a complete graph in the limit $N \to \infty$, we find

$$P_d = \frac{1}{(d+2)(d+1)}. \quad (1)$$

Figure 2 shows that this result agrees well with our simulation data. Likewise, for several
FIG. 3. Descendant distributions for the simple contagion process on random networks.

We simulated the simple contagion process on \( z \)-regular configuration models and Erdős–Rényi (ER) networks of \( N = 10^4 \) nodes. The descendant distributions have been rescaled to collapse on the analytical solution \( (2) \). This rescaling involved adding \( \tilde{x}(z) = (z-1)/(z-2) \) to \( d \), and multiplying \( P_d \) by \( \tilde{x}(z)^{-1} \), the inverse of the scaling factor of \( P_d \) (see Supplementary Section 3).

In the limit of infinite network size, the descendant distributions can be derived. We obtain the infinite-N solution

\[
P_d = \frac{z-1}{z-2} B \left( \frac{z-1}{z-2}, 2 \right),
\]

where \( B(a, b) \) denotes the beta function and \( z \) is the average degree. Figure 3 shows the simulation results for \( z \)-regular configuration models and Erdős–Rényi random graphs of size \( N = 10^4 \). When plotted in a manner suggested by equation \( (2) \), the simulation data for the different random networks collapse onto a single curve (Fig. 3), consistent with the analytical approximation. Finally, for a small-world network created by inserting random shortcuts in a ring lattice, with probability \( p \) of connecting a node with a node chosen uniformly at random, the analytical solution (Supplementary Section 4 and Extended Data Fig. 1) is

\[
P_d = \frac{2p + 1}{2p} B \left( \frac{2p + 1}{2p}, 2 \right).
\]
FIG. 4. **Descendant distributions of a contagion process simulated on real networks.** We ran simulations of the contagion process on two empirical undirected networks (see Supplementary Section 6): one with $N = 81,306$ nodes consisting of the combined edges of 973 Twitter ego-networks, and another with $N = 13,866$ nodes consisting of Facebook pages of athletes, in which edges indicate mutual likes among them. In both cases, we started the contagion at a random seed node, and let exactly 2,000 nodes get infected. Then we stopped the spreading, obtained the descendant distribution for the realization, and started a new simulation with a seed chosen uniformly at random. The descendant distributions shown here are averaged over $10^3$ such simulations. The tail of the distribution declines with an exponent close to $-2$.

Remarkably, all the descendant distributions we have calculated so far turn out to decay asymptotically according to the same power law:

$$P_d \propto d^{-2}$$

(4)

for $d \gg 1$. Further analysis (Supplementary Section 5) indicates that this inverse-square scaling follows from a property that the complete graph shares with the random networks: they all become infinite dimensional as $N \rightarrow \infty$. On this basis, we expect that the same $d^{-2}$ scaling should hold for other infinite-dimensional networks, but not for one-dimensional
chains, two-dimensional grids, three-dimensional lattices, or other networks whose dimensionality remains finite as the number of nodes tends to infinity. Simulations of the model contagion on two-dimensional square grids support this prediction: descendant distributions deviate significantly from the $d^{-2}$ scaling (Extended Data Fig. 2).

Conveniently, many real-world networks are effectively infinite dimensional. Consider the social network Facebook, which as of June 2019 had more than 2.4 billion active users. In a fascinating study, Adamic et al. examined memes spreading from friend to friend on the Facebook social graph. Typically, memes would propagate from one user to another without being altered, but occasionally a user would change the content of the meme before resharing it. This would make a new variant of the meme, which would then spread on the network along with previously existing copies. Adamic et al. examined the frequency of different variants of rarely-changing memes, and found that the frequency distribution of the most widely shared variants followed an inverse-square law. Specifically, they found the exponent to be $-2.01 \pm 0.15$. This exponent matched the prediction of a mean-field model (the Yule process), but it remained unclear why a model without any underlying network structure could account for the exponent obtained from the actual Facebook network.

Our work suggests that the observed exponent of $-2$ is a consequence of the approximate infinite-dimensionality of the Facebook network. Indeed, when we simulate our simple contagion process on sub-networks of Facebook or Twitter, an approximate power-law tail with a slope close to $-2$ emerges (Fig. 4 and Supplementary Section 6).

Our analysis can be viewed as a step toward a mesoscopic theory of contagion, in which infection pathways and epidemic trees would play the leading role, operating at a scale in between the local level of individual nodes and the global level of the entire network. To clarify these distinctions, consider the transition to a giant component in a susceptible-infected-removed model of contagion on a network. Above the transition, there exists a giant infected component of size proportional to $N$. Such macroscopic phenomena have been extensively and fruitfully studied in the literature on network contagion. But sizes of giant components and other macroscopic quantities lump all infected nodes together, and thus discard information about which nodes infected which. Such causal information is retained in epidemic trees, which show the pathways of contagion.

In this letter we have shown one way that epidemic trees can be used: they allow us to calculate descendant distributions, which quantify the impact of a mutant contagion.
occurring at a random place in the network. Our finding that the distribution has a universal $d^{-2}$ tail (for infinite-dimensional networks) means that the expected size of a mutant infected component is of size comparable to $\log N$ for $N \gg 1$. This size is intermediate in a precise sense; it is large compared to the $O(1)$ scale of individual nodes, but small compared to the $O(N)$ scale of giant infected components and the network itself. Note, however, that the variance of the mutant infected component size also diverges as $N \to \infty$. Hence the mean and variance do not adequately summarize the overall distribution, underscoring that one should rely only on the descendant distribution itself, as calculated here.

We expect that notions like contagion pathways, epidemic trees, and descendant distributions are just the beginning of a mesoscopic theory of contagion. Much remains to be discovered about the geometry and statistics of these and other quantities, both empirically for real contagions, and theoretically for a wide range of infection dynamics and network structures. Understanding this middle ground might also have practical benefits for the control of contagion processes, in contexts ranging from vaccination strategies for communicable diseases to methods for combating the spread of misinformation on social media.

**Code availability**

All scripts necessary to reproduce the simulated results are available at https://sid.erda.dk/wsgi-bin/ls.py?share_id=F8JmKmQryb.

**Acknowledgements**

J.S.J. acknowledges funding through the University of Copenhagen UCPH 2016 Excellence Programme for Interdisciplinary Research, the Danish Council for Independent Research, and thanks the Center for Applied Mathematics at Cornell University for hospitality while this work was carried out. S.H.S. was supported by NSF Grant CCF-1522054.

**Author Contributions**

J.S.J. and S.H.S. designed the study. J.S.J. performed the simulations and calculations. J.S.J. and S.H.S. wrote the manuscript.
Competing Interests

The authors declare that they have no competing financial interests.

Materials & Correspondence

Should be addressed to S.H.S.
Extended Data Fig. 1. **Descendant distributions of a contagion process on Newman-Watts small-world networks.** The networks are created by starting out with a ring in which every node is connected to its two nearest neighbors, and then connecting each node with probability $p$ to another node chosen uniformly at random. The resulting descendant distributions, plotted here for networks of size $N = 10^4$ nodes, show the same universal behavior discussed in the main text: the distribution $P_d$ decays in proportion to $d^{-2}$ for large $d$, followed by a finite-size cutoff. We simulate the system for two values of $p$ and plot the resulting descendant distributions $P_d$ along with the analytical approximation derived in Supplementary Section 4.
Extended Data Fig. 2. Descendant distributions of a contagion process on two-dimensional square grids with periodic boundary conditions. The networks consist of $N = 99^2$ nodes, and $10^3$ random realizations of the spreading process were simulated. Because the underlying network is two-dimensional rather than infinite-dimensional, the resulting descendant distributions do not show the scaling law discussed in the main text: the distribution $P_d$ does not decay in proportion to $d^{-2}$ for large $d$. 
Supplementary Information:
Scaling law for the impact of mutant contagion

Jonas S. Juul
Niels Bohr Institute, University of Copenhagen, Blegdamsvej 17, Copenhagen 2100-DK, Denmark
and
Center for Applied Mathematics, Cornell University, Ithaca, New York 14853, USA

Steven H. Strogatz
Center for Applied Mathematics, Cornell University, Ithaca, New York 14853, USA
(Dated: October 3, 2019)

* jonas.juul@nbi.ku.dk
† strogatz@cornell.edu (corresponding author)
I. PRIOR WORK ON CONTAGION PATHS AND EPIDEMIC TREES

The contagion paths and epidemic trees constructed in Fig. 1 in the main text have been studied previously in diverse disciplines. They have been called adoption paths\(^\text{15}\), dissemination trees\(^\text{29,30}\), spreading patterns\(^\text{31}\), causal trees of disease transmission\(^\text{32}\), diffusion structure patterns\(^\text{33}\), the structure of diffusion events\(^\text{34}\), and epidemic trees\(^\text{35}\). We adopt the latter term, which comes from epidemiology, with a single caveat: Generally the graph of the propagation paths for a contagion need not be a directed tree; in the case of a complex contagion\(^\text{36}\), where each child node has two or more parents, the graph could be a directed graph with no cycles. But for the simple contagions studied in this paper, where each child is assumed to have only one parent, the graph of the propagation paths is always a tree.

In the following sections, we show the details of our calculations of descendant distributions on complete graphs, configuration models, Erdős–Rényi random networks, small-world networks, and infinite-dimensional networks in general. In each case, we assume a simplified model of contagion dynamics in which each node is in one of two states: susceptible, or permanently infected and infectious, as described in the main text. The final section of the Supplementary Information contains information about the maximum-likelihood fit we performed on the descendant distributions for the spreading process on empirical networks.

II. COMPLETE GRAPH

In this section, we calculate the descendant distribution for a complete graph of \(N\) nodes, with \(N \gg 1\).

Suppose that nodes are infected one at a time, and that the descendant distribution after \(t\) nodes have been infected is given by \(P_{d,t}\). We wish to calculate the equilibrium distribution of descendants, \(P_d := \lim_{t \to \infty} P_{d,t}\). To do this, it is helpful to use language from network growth\(^\text{37}\). When a new node is infected, a number of already-infected nodes will gain this node as a descendant in the epidemic tree. If, say, 14 nodes acquire this node as a descendant, let us refer to this as introducing 14 descendants in the epidemic tree and then distributing these 14 descendants among the infected nodes. With this terminology in place, we proceed with the calculation.

First, because any edge that connects a susceptible and infected node is equally likely to
be the next edge over which the infection is transmitted, and because the graph is complete, the expected fraction of newly introduced descendants that nodes with $d$ descendants get is

$$\frac{(d + 1)P_{d,t}}{\sum_d (d + 1)P_{d,t}} = \frac{(d + 1)P_{d,t}}{(m_t + 1)},$$

(S1)

where

$$m_t := \sum_d dP_{d,t}$$

(S2)

is the mean number of descendants in the epidemic tree at time $t$. The numerator in Eq. (S1) expresses point (ii) in the main text, and the denominator is a normalisation factor. Next, to go from the expected fraction in Eq. (S1) to the expected number of new descendants that a node with $d$ descendants gets in the following time step, we must multiply the expected fraction (S1) by the total expected number of new descendants, aggregated over nodes with any number of descendants, that are added during the time step.

To find this total, we observe that every infected node has equal probability of being the next to pass on the infection, and there are $t$ infected nodes at time $t$. Thus the probability that nodes with $d$ descendants will get a new descendant is $(d + 1)P_{d,t}/t$. Summing over all $d$ then gives us the expected fraction of the infected nodes in total that will get a new descendant in the following time step; multiplying by $t$ gives us the corresponding expected number. This argument tells us, then, that

$$t \sum_d \frac{(d + 1)P_{d,t}}{t} = m_t + 1$$

(S3)

is the expected number of infected nodes, in total, that will get a new descendant in the following time step. Note that the underlying network did not influence this last part of the calculation.

By combining Eqs. (S1) and (S3) we find that, for the complete graph, the expected number of new descendants that a node with $d$ descendants gets in time step $t$ is

$$\frac{(d + 1)P_{d,t}}{(m_t + 1)}(m_t + 1) = (d + 1)P_{d,t}. \tag{S4}$$

This result leads us to the following master equation, which expresses the expected gain and loss of nodes with $d$ descendants between time steps $t$ and $t + 1$:

$$(t + 1)P_{d,t+1} - tP_{d,t} = \begin{cases} 1 - P_{0,t} & \text{for } d = 0, \\ dP_{d-1,t} - (d + 1)P_{d,t} & \text{for } d \geq 1. \end{cases} \tag{S5}$$
The case $d = 0$ is different from other values of $d$ since the newly infected node will have no descendants when it is added to the epidemic tree, thereby making the gain term in the master equation equal to 1. An equilibrium distribution must satisfy $P_{d,t} = P_{d,t+1} =: P_d$. Applying this condition and solving for $P_d$, we get:

$$P_0 = \frac{1}{2}, \quad P_d = \frac{d}{d+2}P_{d-1}. \quad (S6)$$

From this we conclude that the distribution of the expected number of descendants on the complete graph is

$$P_d = \frac{d!}{(d+2)!} = \frac{1}{(d+2)(d+1)}. \quad (S7)$$

As mentioned in the main text, keeping track of descendants can be mapped to growing a network by node copying\textsuperscript{25}. For the complete graph, this mapping means that equation (S7) is identical to the formula for the in-degree distribution calculated by Krapivsky & Redner\textsuperscript{25}. In their paper on network growth with node copying, Krapivsky & Redner derive geometrical properties of the grown networks. We refer the interested reader to the paper, and continue with calculating descendant distributions below.

### III. CONFIGURATION MODEL AND ERDŐS–RÉNYI RANDOM NETWORKS

In this section, we turn to two classes of random networks: configuration-model networks, and Erdős–Rényi random graphs.

In the configuration model that we consider, each of $N$ nodes has a certain number of “half edges” (or “stubs”) sticking out of it, with the number of stubs being chosen at random from a prescribed degree distribution. The network is then generated by connecting pairs of stubs, chosen uniformly at random from the list of all stubs, to make the full edges of the resulting network.

The Erdős–Rényi networks are constructed by considering each pair of nodes independently and, with probability $p$, connecting that pair with an undirected edge.

To understand Fig. 3 shown in the main text, we now calculate the descendant distribution for these random networks, using the same method as in the previous section. At time steps $t \geq 1$, an infected node with degree $k$ has at least one infected neighbor (its “parent”). If the infected node (denoted $I$), or one of its descendants, infects a neighbor, then $I$ loses one edge over which it could infect another node. By doing this, however, it gets a new
descendant, which might have a number of edges connecting it to susceptible nodes. If we assume that every one of the $k - 1$ edges that could connect an infectious degree-$k$ node with a susceptible node has equal probability of doing so (equal to 1 in the infinite-network limit), and if we assume that this probability is the same for every infected node, then an infected node has on average $(z - 2)d + (z - 1)$ edges which could connect it to susceptible nodes. Here $z$ is the mean degree of the network.

So the mean number of new descendants that a node with $d$ descendants gets when a new node is infected is

$$\frac{(z - 2)d + z - 1}{(z - 2)m_t + z - 1} P_{d,t}(m_t + 1).$$

Using this result, we can write down a master equation as we did when calculating the descendant distribution for the spreading process on the complete graph, and solve for a steady-state descendant distribution $P_d$, in the limit of infinite network size. After some algebra (see Supplementary Section V for details), we find that

$$P_d = \begin{cases} \frac{z-2}{2z-3} & \text{for } d = 0, \\ \frac{z-2}{2z-3} \left[ B \left( \frac{z-1}{z-2}, 2 \right) \right]^{-1} B \left( \frac{z-1}{z-2} + d, 2 \right) & \text{for } d \geq 1. \end{cases}$$

Here $B(a,b)$ is the beta function, which declines as $a^{-b}$ as $a \to \infty$ for fixed $b$. In our case, this means

$$P_d \propto B \left( \frac{z-1}{z-2} + d, 2 \right),$$

$$\propto d^{-2},$$

for $d \gg 1$. By invoking identities for the beta function, we can rewrite the expression for the descendant distribution as

$$P_d = \frac{z-1}{z-2} B \left( \frac{z-1}{z-2} + d, 2 \right),$$

which is the expression we list in the main text. Figure 3 in the main text collapses the simulated data on the curve $B(\tilde{d}, 2)$, where

$$\tilde{d} := \tilde{x}(z) + d = \frac{z-1}{z-2} + d.$$

Given a simulated data point $(d, P_d)$, this collapse is made by plotting the data point at $(d + \tilde{x}(z), [\tilde{x}(z)]^{-1} P_d)$ instead.
IV. A RING AND A SMALL-WORLD NETWORK

The networks of the previous section are related to each other, in that edges between nodes are created according to a random procedure. They all show the inverse-square scaling mentioned in the main text: $P_d \propto d^{-2}$ for large $d$.

On the other hand, some non-random graphs do not have this limiting behavior. If we consider the spreading process taking place on a ring, in which every node has only two neighbours, one to its left and one to its right, we can write down the expected distribution of number of descendants immediately. Starting from only a single seed, in each time step there will be exactly one possibility for the process to spread on the right hand side of the seed, and one possibility to spread on the left hand side. The resulting distribution of descendants in a ring consisting of $N$ nodes is

$$P_d = N \sum_{L=0}^{N} \left[ \Theta (d \leq L - 1) P_L + \Theta (d \leq N - L + 1) P_L \right].$$

(S13)

where $N$ is a normalization constant, $\Theta(x)$ is the Heaviside function equal to 1 if $x \geq 0$ and 0 otherwise, and $P_L$ is the probability of the contagion process spreading exactly $L$ times to the left along the periphery of the ring, given by

$$P_L = \binom{N}{L} \left( \frac{1}{2} \right)^N.$$

(S14)

One natural question to ask is then: How random does a network have to be to show the limiting behavior $P_d \propto d^{-2}$ we observed in the previous sections? In the rest of this section we analytically estimate the descendant distribution for the contagion process on small-world networks. Specifically, the small-world networks are Watts-Newman small-world networks in which all nodes are connected to their two immediate neighbours on a ring lattice, and each node gets a shortcut to a neighbour chosen uniformly at random with probability $p$.

First, we must estimate the expected number of new descendants that a node with $d$ descendants gets when a node gets infected. If the underlying network was simply a ring and no shortcuts had been added, every node would have equal chance of getting new descendants. This changes when the shortcuts are inserted: For each descendant a node has, the chance that one of its descendants has a shortcut increases. If the infection traverses such a shortcut link successfully, it can spread both to the right and to the left in this newly discovered part of the network. Hence, two more boundaries between infectious nodes and
susceptible nodes have been created, and every node that has descendants on this boundary now has a higher chance of getting more descendants. This effect alters the expected number of descendants received by a node with \( d \) descendants when a new node gets infected. The expected number now becomes

\[
\frac{P_{d,t}(1 + 2p(d + 1))}{1 + 2p(m_t + 1)} (m_t + 1).
\]

(S15)

Here the first term represents the shortcut-independent probability that every node has to get a new descendant, and the terms that are proportional to \( p \) correspond to the increased probability of getting new descendants that nodes get via shortcuts. With this, we can write down the master equation as in the two sections above. After some algebra (see Supplementary Section \( \text{V} \) below for details), we find

\[
P_d = \begin{cases} 
\frac{2p}{1 + 4p} & \text{for } d = 0, \\
\frac{2p}{1 + 4p} \left[ B \left( \frac{2p+1}{2p}, 2 \right) \right]^{-1} B \left( \frac{2p+1}{2p} + d, 2 \right) & \text{for } d \geq 0.
\end{cases}
\]

(S16)

For large \( d \), this analytical solution declines as

\[
P_d \propto B \left( \frac{2p + 1}{2p} + d, 2 \right),
\]

\[
\propto d^{-2}.
\]

(S17)

(S18)

In Extended Data Fig. 1, we see that the analytical solution indeed is in qualitative agreement with the simulations.

V. UNIVERSAL BEHAVIOR AND WHEN IT BREAKS DOWN

We have studied the descendant distributions for simple contagion on different networks: complete graphs, configuration-model networks, Erdős–Rényi networks, and small-world networks. On all of these networks, we have discovered a striking universality: The distributions decline as a power law with exponent \(-2\) for large \( d \). A natural question to ask is, then, what unifies these graphs: When does the universality exist, and when does it break down?

One thing that is true for all the graphs we have studied is that the probability of getting more descendants is linearly proportional to the number of descendants the node already has.
In other words, the expected number of descendants received by a node with \( d \) descendants, when a new node gets infected, is of the form

\[
P_{d,t} \frac{(c + fd)}{\sum_d P_{d,t} (c + fd)} (m_t + 1) = \frac{P_{d,t} (c + fd)}{c + f m_t} (m_t + 1),
\]

(S19)

for \( c, f > 0 \). We will now show that this, and \( m_t \to \infty \) as \( t \to \infty \), is sufficient to make the resulting distribution of the number of descendants decline as the power law with exponent \(-2\) for large \( d \). The condition \( m_t \to \infty \) is true for all the classes of random graphs we have examined, since the fewer edges compared to the complete graph decreases the interface between susceptible and infectious nodes. This makes the probability of nodes with many descendants getting additional descendants increase compared to the spreading process on the complete graph. Because \( m_t \) diverges for the complete graph, \( m_t \) also diverges for the random graph in question by the comparison test. As \( m_t \to \infty \), the right hand side of equation (S19) approaches \( P_{d,t}(d/f + d) \). With this, we get the master equation,

\[
(t + 1)P_{d,t+1} - tP_{d,t} = \begin{cases} 
1 - P_d & \text{for } d = 0, \\
\left[P_{d-1,t} \left(\frac{c}{f} + d - 1\right)
- P_{d,t} \left(\frac{c}{f} + d\right)\right] & \text{for } d \geq 1.
\end{cases}
\]

(S20)

Looking for steady-state solutions \( P_{d,t+1} = P_{d,t} =: P_d \), we obtain

\[
P_0 = \frac{f}{f + c}, \quad P_d = \frac{c/f - 1 + d}{1 + c/f + d} P_{d-1},
\]

(S21)

where the expression for \( P_d \) is valid for \( d \geq 1 \). Denoting \( c/f - 1 =: \alpha \), we can use the recursive nature of the expression to rewrite \( P_d \) as follows:

\[
P_d = P_0 \prod_{\lambda=1}^d \frac{\alpha + \lambda}{\alpha + 2 + \lambda} = P_0 \frac{\Gamma(\alpha + 3)\Gamma(\alpha + d + 1)}{\Gamma(\alpha + 1)\Gamma(\alpha + 3 + d)}.
\]

(S22)

If we increase the terms of the fraction by a factor of \( \Gamma(2) \), and use the relation between gamma functions and beta functions, \( \Gamma(x)\Gamma(y)/\Gamma(x + y) = B(x, y) \), we get

\[
P_d = \frac{f}{f + c} [B(\alpha + 1, 2)]^{-1} B(\alpha + d + 1, 2),
\]

(S23)

\[
= \frac{c}{f} B\left(\frac{c}{f} + d, 2\right).
\]

(S24)
The final step was made by inserting the value of \( \alpha \) and evaluating \( B(c/f, 2) = f^2/[c(c+f)] \).

The asymptotic behavior for large \( d \) is

\[
P_d \propto (c/f + d)^{-2} \\
\propto d^{-2}.
\] (S25)

Therefore, if the probability of getting more descendants increases linearly with the number of descendants a node already has, the descendant distribution will decline as \( d^{-2} \) for large \( d \). If we interpret the number of descendants an infected node has as a volume, and the interface separating infectious and susceptible nodes as a surface area, the descendant distribution will show the observed universality if the surface area and the volume increase equally fast (proportional to \( d \)); in other words, if the graph is infinite dimensional.

VI. EMPIRICAL NETWORKS AND MAXIMUM-LIKELIHOOD FIT

The two empirical networks discussed in the main text were obtained from the SNAP database. For the Twitter network, we converted all directed edges into undirected ones, not allowing parallel edges. To estimate the power-law exponent of the tail of the distributions, we used a maximum-likelihood method to fit a power-law to the data for \( d \in [100, 1900] \) (not including the extreme data points caused by the choice of having a single seed). We used the approximate expression for the maximum-likelihood power-law exponent \( \hat{\alpha} \) for binned data,

\[
\hat{\alpha} \approx 1 + n \left[ \sum_{i=1}^{n} \ln \frac{d_i}{d_{\text{min}} - 1/2} \right]^{-1},
\] (S26)

and checked that direct numerical evaluation gave a similar result. In this formula, \( n \) is the total number of data points, \( d_i \) is the \( i^{\text{th}} \) data point, and \( d_{\text{min}} \) is the smallest value for the data (in this case, 100). We estimated the standard error on \( \hat{\alpha} \) by using the corresponding formula,

\[
\sigma_{\hat{\alpha}} = \frac{1}{\sqrt{n} \left[ \frac{\zeta''(\hat{\alpha}, d_{\text{min}})}{\zeta'(\hat{\alpha}, d_{\text{min}})} - \left( \frac{\zeta'(\hat{\alpha}, d_{\text{min}})}{\zeta(\hat{\alpha}, d_{\text{min}})} \right)^2 \right]},
\] (S27)
where prime indicates differentiation with respect to the first variable. For both networks we obtained $\hat{\alpha} \approx 2.04 \pm 0.01$.

[1] Watts, D. J. A simple model of global cascades on random networks. *Proceedings of the National Academy of Sciences* **99**, 5766–5771 (2002).

[2] Dodds, P. S. & Watts, D. J. A generalized model of social and biological contagion. *Journal of theoretical biology* **232**, 587–604 (2005).

[3] Castellano, C., Fortunato, S. & Loreto, V. Statistical physics of social dynamics. *Reviews of Modern Physics* **81**, 591 (2009).

[4] Christakis, N. A. & Fowler, J. H. The spread of obesity in a large social network over 32 years. *New England Journal of Medicine* **357**, 370–379 (2007).

[5] Lehmann, S. & Ahn, Y.-Y. *Complex spreading phenomena in social systems* (Springer, 2018).

[6] Anderson, R. M. & May, R. M. *Infectious diseases of humans: dynamics and control* (Oxford university press, 1991).

[7] Keeling, M. J. & Rohani, P. *Modeling infectious diseases in humans and animals* (Princeton University Press, 2011).

[8] De Domenico, M., Granell, C., Porter, M. A. & Arenas, A. The physics of spreading processes in multilayer networks. *Nature Physics* **12**, 901 (2016).

[9] Fennell, P. G. & Gleeson, J. P. Multistate dynamical processes on networks: analysis through degree-based approximation frameworks. *SIAM Review* **61**, 92–118 (2019).

[10] Pastor-Satorras, R., Castellano, C., Van Mieghem, P. & Vespignani, A. Epidemic processes in complex networks. *Reviews of Modern Physics* **87**, 925 (2015).

[11] Daley, D. J. & Kendall, D. G. Stochastic rumours. *IMA Journal of Applied Mathematics* **1**, 42–55 (1965).

[12] Vosoughi, S., Roy, D. & Aral, S. The spread of true and false news online. *Science* **359**, 1146–1151 (2018).

[13] Bettencourt, L. M., Cintrón-Arias, A., Kaiser, D. I. & Castillo-Chávez, C. The power of a good idea: Quantitative modeling of the spread of ideas from epidemiological models. *Physica A: Statistical Mechanics and its Applications* **364**, 513–536 (2006).
[14] Mellor, A., Mobilia, M., Redner, S., Rucklidge, A. M. & Ward, J. A. Influence of luddism on innovation diffusion. *Physical Review E* **92**, 012806 (2015).

[15] Juul, J. S. & Porter, M. A. Hipsters on networks: How a minority group of individuals can lead to an antiestablishment majority. *Physical Review E* **99**, 022313 (2019).

[16] Krapivsky, P. L., Redner, S. & Volovik, D. Reinforcement-driven spread of innovations and fads. *Journal of Statistical Mechanics: Theory and Experiment* **2011**, P12003 (2011).

[17] Haldane, A. G. & May, R. M. Systemic risk in banking ecosystems. *Nature* **469**, 351 (2011).

[18] Brummitt, C. D., D’Souza, R. M. & Leicht, E. A. Suppressing cascades of load in interdependent networks. *Proceedings of the National Academy of Sciences* **109**, E680–E689 (2012).

[19] Andreasen, V., Viboud, C. & Simonsen, L. Epidemiologic characterization of the 1918 influenza pandemic summer wave in Copenhagen: implications for pandemic control strategies. *The Journal of Infectious Diseases* **197**, 270–278 (2008).

[20] Adamic, L. A., Lento, T. M., Adar, E. & Ng, P. C. Information evolution in social networks. In *Proceedings of the Ninth ACM International Conference on Web Search and Data Mining*, 473–482 (ACM, 2016).

[21] Yule, G. U. A mathematical theory of evolution, based on the conclusions of dr. j. c. willis, f.r.s. *Philosophical Transactions of the Royal Society of London. Series B, containing papers of a biological character* **213**, 21–87 (1925).

[22] Simon, H. A. On a class of skew distribution functions. *Biometrika* **42**, 425–440 (1955).

[23] De Solla Price, D. A general theory of bibliometric and other cumulative advantage processes. *Journal of the American Society for Information Science* **27**, 292–306 (1976).

[24] Barabási, A.-L. & Albert, R. Emergence of scaling in random networks. *Science* **286**, 509–512 (1999).

[25] Krapivsky, P. L. & Redner, S. Network growth by copying. *Physical Review E* **71**, 036118 (2005).

[26] Newman, M. E. J. & Watts, D. J. Renormalization group analysis of the small-world network model. *Physics Letters A* **263**, 341–346 (1999).

[27] Leskovec, J. & McAuley, J. J. Learning to discover social circles in ego networks. In *Advances in neural information processing systems*, 539–547 (2012).

[28] Rozemberczki, B., Davies, R., Sarkar, R. & Sutton, C. Gemsec: Graph embedding with self clustering. *arXiv preprint arXiv:1802.03997* (2018).
[29] Oh, S.-W. & Porter, M. A. Complex contagions with timers. *Chaos: An Interdisciplinary Journal of Nonlinear Science* **28**, 033101 (2018).

[30] Liben-Nowell, D. & Kleinberg, J. Tracing information flow on a global scale using internet chain-letter data. *Proceedings of the National Academy of Sciences* **105**, 4633–4638 (2008).

[31] Jang, S. M. *et al.* A computational approach for examining the roots and spreading patterns of fake news: Evolution tree analysis. *Computers in Human Behavior* **84**, 103–113 (2018).

[32] Vázquez, A. Causal tree of disease transmission and the spreading of infectious diseases. In *Discrete Methods in Epidemiology*, 163–180 (2004).

[33] Zhang, Z.-K. *et al.* Dynamics of information diffusion and its applications on complex networks. *Physics Reports* **651**, 1–34 (2016).

[34] Goel, S., Anderson, A., Hofman, J. & Watts, D. J. The structural virality of online diffusion. *Management Science* **62**, 180–196 (2015).

[35] Haydon, D. T. *et al.* The construction and analysis of epidemic trees with reference to the 2001 uk foot–and–mouth outbreak. *Proceedings of the Royal Society of London. Series B: Biological Sciences* **270**, 121–127 (2003).

[36] Mønsted, B., Sapieżyński, P., Ferrara, E. & Lehmann, S. Evidence of complex contagion of information in social media: An experiment using twitter bots. *PloS one* **12**, e0184148 (2017).

[37] Newman, M. E. J. The structure and function of complex networks. *SIAM Review* **45**, 167–256 (2003).

[38] Leskovec, J. & Krevl, A. SNAP Datasets: Stanford large network dataset collection. [http://snap.stanford.edu/data](http://snap.stanford.edu/data) (2014).

[39] Clauset, A., Shalizi, C. R. & Newman, M. E. J. Power-law distributions in empirical data. *SIAM review* **51**, 661–703 (2009).