Description of spreading dynamics by microscopic network models and macroscopic branching processes can differ due to coalescence

Johannes Zierenberg,[1,2] Jens Wilting,[1] Viola Priesemann,[1,2,*,†] and Anna Levina[3,4,*,‡]

[1]*Max Planck Institute for Dynamics and Self-Organization, Am Fassberg 17, 37077 Göttingen, Germany*

[2]*Bernstein Center for Computational Neuroscience, Am Fassberg 17, 37077 Göttingen, Germany*

[3]*University of Tübingen, Max Planck Ring 8, 72076 Tübingen, Germany*

[4]*Max Planck Institute for Biological Cybernetics, Max Planck Ring 8, 72076 Tübingen, Germany*

(Received 12 April 2019; accepted 17 December 2019; published 3 February 2020)

Spreading processes are conventionally monitored on a macroscopic level by counting the number of incidences over time. The spreading process can then be modeled either on the microscopic level, assuming an underlying interaction network, or directly on the macroscopic level, assuming that microscopic contributions are negligible. The macroscopic characteristics of both descriptions are commonly assumed to be identical. In this work we show that these characteristics of microscopic and macroscopic descriptions can be different due to coalescence, i.e., a node being activated at the same time by multiple sources. In particular, we consider a (microscopic) branching network (probabilistic cellular automaton) with annealed connectivity disorder, record the macroscopic activity, and then approximate this activity by a (macroscopic) branching process. In this framework we analytically calculate the effect of coalescence on the collective dynamics. We show that coalescence leads to a universal nonlinear scaling function for the conditional expectation value of successive network activity. This allows us to quantify the difference between the microscopic model parameter and established estimates of the macroscopic branching parameter. To overcome this difference, we propose a nonlinear estimator that correctly infers the microscopic model parameter for all system sizes.

DOI: 10.1103/PhysRevE.101.022301

I. INTRODUCTION

A multitude of spreading processes are influencing our lives. Examples include the spread of news, opinions, or rumors [1,2], the outbreak of diseases [3,4], the escalation of economic crises [5], and the propagation of spiking activity in neural systems [6,7]. Mathematically, the unifying feature of these processes is that some signal (infection, information, or spike) spreads through a system. Characterizing the collective dynamics of these processes, e.g., in terms of the average spreading rate (here described by the branching parameter $\hat{m}$), is of general interest to many fields.

Spreading can be modeled on a microscopic node-to-node level, which requires making assumptions about the interaction graph and the rules about how a signal may propagate from one node to the next. This typically involves stochastic processes, such as (probabilistic) cellular automata [8–10], contact processes [10,11], or interacting Hawkes processes [12]. In particular, infectious diseases have been modeled by so-called susceptible-infectious models or generalizations thereof [13], whereas spike propagation in neural networks has been modeled by so-called branching networks [14–22], Hawkes processes [23–25], or probabilistic integrate-and-fire networks [26,27]. The advantage of these microscopic models is that one can directly study the effect of network topology, such as country maps [28–30], daily transportation patterns [31], social links [32,33], or connectomes [34], and even time-varying networks, such as diffusive motion of nodes [35].

Microscopic spreading models can be classified into independent-spreading models and dependent-spreading models. For independent-spreading models, each link is associated with an independent spreading probability, resulting in an overall multiplicative probability to activate a node. The prime example for independent spreading is a probabilistic cellular automaton [9,10]. For dependent-spreading models, the activation of a node is determined by a function of the states of all neighboring nodes, e.g., threshold models [36] or integrate-and-fire models [26]. In this work we focus on the class of independent-spreading models represented by the branching network, a probabilistic cellular automaton [9,15] in the universality class of directed bond percolation [10,37].

Alternatively, spreading can be modeled on a macroscopic population level, assuming either that there is no explicit network or that network heterogeneity is averaged out. Macroscopic models describe the development of population or network activity without assuming a specific network topology, e.g., modeling the total number of infected people at each time step by simply assuming a general statistics of how each infected person spreads the disease. Classical examples include the branching process [6], Kesten process [38], and other processes of the autoregressive AR(n) family. These processes are frequently used to describe spreading dynamics

*These authors contributed equally to this work.

†viola.priesemann@ds.mpg.de

‡anna.levina@uni-tuebingen.de

in real-world systems, because parameter estimation for them has been well established. For example, branching processes have been used to explain data in neuroscience [21,39–42], epidemics [21,43], and economics [44].

Microscopic and macroscopic spreading models can however be quite different: Spreading processes in microscopic models (independent spreading) can interact with each other (e.g., when a node has already been activated by another node), while in macroscopic models this is typically not the case (e.g., in the branching process each element generates new descendants independently of the number of descendants of the other elements). As one of the consequences, the network activity of microscopic spreading models is upper bounded by the network size, whereas the population activity of macroscopic spreading models can in principle diverge. On the other hand, one can find equivalent behavior in the limit of low external and internal activation. As a result, macroscopic approximations have been used to describe typical behavior of microscopic models [14,21,22,43], in part because such approximations offer the advantage of simple analytical tractability. However, if one wants to model a real system, one has to carefully weight the assumptions one has to make in either microscopic or macroscopic models.

In this work we address the following question: Given that a real system (and its measured macroscopic data) can be approximated by a microscopic or a macroscopic model, what is the relation between the resulting macroscopic dynamics of both models? To approach this question (Sec. II), we simplify the problem by generating the data (network activity) directly with the microscopic model (branching network), thereby avoiding assumptions about the real system and how to model this with a microscopic model (Fig. 1). We can thus focus on the approximations in the macroscopic model (branching process). We reveal analytically and numerically that conventional estimators of the branching parameter can be biased, i.e., estimates do not agree with the microscopic model parameter (Secs. III A–III D). The reason for this bias is coalescence (the simultaneous activation of one node by multiple sources). We propose a nonlinear estimator that correctly infers the microscopic model parameter from the network activity (Sec. III E). Finally, we discuss our results with implications for general spreading processes (Sec. IV).

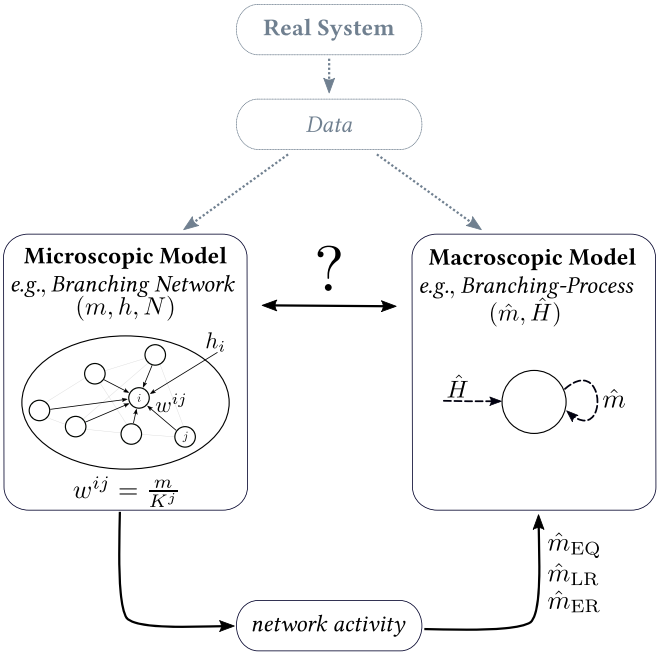

FIG. 1. Illustration of the question we address and how we approach it. The question is, given that a real system, for which we measure data (e.g., the number of incidences per time step), can be approximated by a microscopic model (assuming an underlying interaction network with update rules that produce incidences) or a macroscopic model (on the level of the number of incidences), what is the relation between the resulting macroscopic dynamics of both models? The approach is as follows. We use a branching network with annealed connectivity disorder to generate population activity. Then we approximate this activity by a branching process, creating a macroscopic description. Thereby, we avoid systematic errors from (i) an unknown real system and (ii) required approximations for the network model. We can thus focus on different branching-process approximations ($\hat{m}_{\mathrm{EQ}}$, $\hat{m}_{\mathrm{LR}}$, and $\hat{m}_{\mathrm{ER}}$; see Sec. II C) and can compare the resulting macroscopic dynamics of both models.

II. MODEL AND METHODS

We use branching networks as a microscopic model and generate macroscopic observables (network activity). We then approximate these observables by a branching process as a macroscopic model and compare the spreading rates between the microscopic model and macroscopic approximation (Fig. 1). In this framework, the spreading rate is called branching parameter m. To distinguish parameters on the microscopic and macroscopic levels, we denote macroscopic parameters by a circumflex, e.g., $\hat{m}$.

A. Branching network

Consider a network with N nodes. Time progresses in discrete time steps Δt (here $\Delta t = 1$). Each node i can be either silent ($s_t^i = 0$) or active ($s_t^i = 1$); thus the (macroscopic) network activity is given by $A_t = \sum_{i=1}^{N} s_t^i$. Activation of a node can be induced in two ways. First, internally, a node i can be activated with probability w^{ij} (connection weight) by another node j that was active in the previous time step. The connection weight determines the microscopic dynamics. Second, a node can be activated by an external Poisson input if one or more inputs arrive within Δt. For a Poisson process of rate h, the probability that no input arrives is $\exp(-h\Delta t)$ such that the activation probability through external input is $\lambda(h) = 1 - \exp(-h\Delta t)$. The networkwide external input rate is then $H = hN$. An activated node transitions back from $s_t^j = 1$ to $s_{t+1}^j = 0$ in the next time step unless activated again, which corresponds to a refractory period smaller than Δt in the modeling of neuronal activity.

The microscopic dynamics is controlled by the branching parameter m, which quantifies that a single activation of any node causes on average m active nodes in the next time step. The simplest way to achieve this is to set connection weights $w^{ij} = \frac{m}{K^j}$ if i is one of the K^j outgoing connections from node j and $w^{ij} = 0$ otherwise. We consider here the mean-field scenario of annealed connectivity disorder: Connections between nodes are redrawn in each step with probability K/N

and nodes are activated internally with probability $w = m/K$. For K sufficiently large, this is mathematically equivalent to an all-to-all connected network with $w^{ij} = w = \frac{m}{N}$, including potential self-coupling. This mean-field connectivity ignores spatial heterogeneity [45], but our results can be adapted to mean-field approximations for quenched disorder over static random (Erdős-Rényi-type) networks [15].

A naive implementation of our model dynamics on an all-to-all connected network is computationally very expensive. This is because for each active node we would have to draw N random numbers to check for internal activation. To reduce numerical complexity, we note that in the mean-field case the number of nodes k activated by a single active node is distributed binomially $P(k) = \binom{N}{k} w^k (1-w)^{N-k}$. Instead of going over all N connected nodes, we can thus first draw a number k from the binomial distribution and then draw k random nodes without repetitions to be activated. This procedure is significantly more efficient, especially for large system sizes. For our finite-size scaling analysis we simulated networks up to size $N = 2^{20} \approx 10^6$ for 10^7 times steps.

B. Branching process

The branching process [6] is a discrete-time stochastic Markov process: If at time t there are A_t elements, then at time $t+1$ each of these elements generates a random integer number of descendants x_t^i. If this internal generation of descendants is complemented by random external input y_t, one speaks about a branching process with immigration [46,47], or a driven branching process. The number of elements at time $t+1$ can be written as

$$A_{t+1} = \sum_{i=1}^{A_t} x_t^i + y_t. \tag{1}$$

Many results about branching processes only depend on the average number of internally generated elements $\hat{m} = \langle x_t^i \rangle$, called the branching parameter, and on the average number of externally generated elements $\hat{y}$ per time step. In order to compare with the branching network, we identify $\hat{y} = \hat{H}\Delta t$, where $\hat{H}$ is the total rate of the Poisson-distributed external input. Recall that we use $\hat{m}$ to distinguish parameters on the macroscopic level from those at a microscopic level. Using the Markovian nature of branching processes, we describe the time evolution of population activity A_t by the conditional expectation value

$$\langle A_{t+1} | A_t \rangle = \hat{m} A_t + \hat{H}\Delta t. \tag{2}$$

The branching process thus belongs to the class of processes with an autoregressive representation.

The population activity A of a driven branching process can be calculated in a mean-field approximation. Assuming a stationary population activity ($\hat{m} < 1$), we can neglect fluctuations and only consider the expectation value over Eq. (2). The law of total expectation then implies that $A = \langle\langle A_{t+1} | A_t \rangle\rangle = \hat{m}A + \hat{H}\Delta t$, where the expectation value of the population activity is the population rate $A = \langle A_t \rangle$. Solving for the population rate leads to

$$A = \frac{\hat{H}\Delta t}{1-\hat{m}}, \tag{3}$$

which is only well defined for $\hat{m} < 1$ and diverges for $\hat{m} \to 1$.

C. Branching-process approximation

The framework of a branching process can be used to infer the branching parameter as a proxy for the spreading dynamics from a time series of network activity $\{A_t\}$. This is clearly an approximation, because (i) the network sets an upper bound on A_t and (ii) nodes interact with each other. Within this approximation there are yet different possible approaches. We present three in the following.

The first and easiest approach we consider is to estimate the branching parameter via Eq. (3). Assuming that the networkwide external input rate is known from the microscopic dynamics ($\hat{H} = hN$), one can estimate the branching parameter from the expectation value of the network rate $\langle A \rangle$ (ER denotes the expected rate) as

$$\hat{m}_{\mathrm{ER}} = 1 - \frac{\hat{H}\Delta t}{\langle A \rangle}. \tag{4}$$

This clearly neglects fluctuations and will therefore be biased for $m \approx 1$ where fluctuations are large. In addition, it is obviously upper bounded, $\hat{m}_{\mathrm{ER}} \leqslant 1$.

The second approach we consider is more elaborate; it is based on the relationship for the conditional expectation value (2). Assuming a separation of timescales (STS), i.e., that no external activation is delivered while the network is active, one can neglect $\hat{H}$ and define the expectation value of the quotient of subsequent network activity (EQ denotes the expected quotient)

$$\hat{m}_{\mathrm{EQ}} = \left\langle \frac{A_{t+1}}{A_t} \right\rangle_{A_t > 0}. \tag{5}$$

This estimator has been widely applied to neural data [14,20,48]. However, as we will show in more detail later, it is strongly biased for networks subject to a non-negligible input rate ($H > 0$), as noted before [20].

The third approach we consider is again based on the conditional expectation value (2), but explicitly considers the presence of an external input rate. In fact, it is well known that the first moments of a driven branching process can be estimated with a linear regression [6,21,49] (LR denotes linear regression) as

$$\hat{m}_{\mathrm{LR}} = \frac{\mathrm{Cov}[A_{t+1}, A_t]}{\mathrm{Var}[A_t]}, \tag{6}$$

$$\hat{H}_{\mathrm{LR}} = \frac{1}{\Delta t}(\langle A_{t+1} \rangle - \hat{m}_{\mathrm{LR}} \langle A_t \rangle). \tag{7}$$

Besides the additional benefit of simultaneously estimating the input rate, this estimator can be extended to be invariant to subsampling, i.e., when activity is recorded from only (small) parts of the network [21]. For stationary activity $\langle A_t \rangle = \langle A_{t+1} \rangle = A$, Eq. (6) simplifies to

$$\hat{m}_{\mathrm{LR}} = \frac{\langle A_{t+1} A_t \rangle - A^2}{\langle A_t^2 \rangle - A^2}. \tag{8}$$

We define here the estimators in terms of expectation values $\langle \mathcal{O}_t \rangle$, formally defined for infinitely long time series of some observable $\mathcal{O}_t$. For a finite time series the expectation values themselves have to be estimated by the time average

$$\overline{\mathcal{O}_t} = \frac{1}{T}\sum_{t=1}^{T} \mathcal{O}_t \underset{T\to\infty}{\longrightarrow} \langle \mathcal{O}_t \rangle. \tag{9}$$

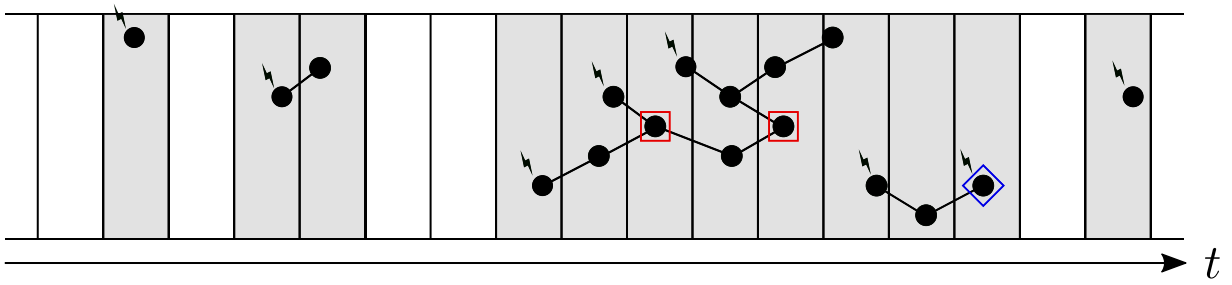

FIG. 2. Illustration of activity propagation (black circles) and coalescence in branching networks. Different activity propagation pathways converge onto each other, with the consequence that a node is activated by two or more sources. We distinguish internal coalescence (red squares), where two or more sources from within the network activate the same target, and external coalescence (blue diamond), where external input (denoted by black lightning bolts) contributes. The y axis denotes different nodes and the x axis time in discrete steps. Time steps without any activity are white; the others are gray.

III. RESULTS

A. Analytic results on effects of coalescence in driven branching networks

During spreading dynamics on a finite network, the activities of different nodes may interfere with each other. With increasing network activity, the probability increases that a node receives activation from two or more nodes in the same time step. One can call such multiple activations of a node coalescence [37] because different branches of the spreading process coalesce (Fig. 2). We further distinguish between internal coalescence, where a node gets activated by two or more nodes from within the network, and external coalescence, where a node gets additionally activated by external input.

As a consequence of coalescence, the effective number of internally activated nodes, or in other words the (microscopic) effective branching parameter $m_{\mathrm{eff}}(A_t)$, will be diminished. One may expect that in the limit $N \to \infty$ the effective branching parameter approaches the model branching parameter, but we will show that this is not always the case. For driven branching networks, one can imagine that the external input initiates independent spreading processes. The initiation can cause external coalescence with the present processes. In addition, the individual spreading processes interact in the sense of a neutral theory [50], which leads to internal coalescence.

In order to derive the effective branching parameter, we first derive the probability that a given node i is activated. Due to potential coalescence this is not straightforward, but we can compute the probability $p_{\mathrm{na}}^{\mathrm{int}}$ that node i is not activated by node j. For generality, we consider the annealed disorder with connection selection probability K/N and activation probability m/K. Then $p_{\mathrm{na}}^{\mathrm{int}} = (1 - K/N) + (K/N)(1 - m/K) = 1 - m/N$, which is independent of K. Also, node i is not activated by the external input with probability $p_{\mathrm{na}}^{\mathrm{ext}} = 1 - \lambda(h)$. Considering that there are A_t active nodes from which node i can be activated, the probability of activating node i is

$$P\left[s_t^i = 1 \middle| A_t, N, m, h\right] = 1 - \left(p_{\mathrm{na}}^{\mathrm{int}}\right)^{A_t} p_{\mathrm{na}}^{\mathrm{ext}} = 1 - \left(1 - \frac{m}{N}\right)^{A_t} [1 - \lambda(h)]. \quad (10)$$

Since this holds for any node in the network, we can generalize $P[s_t^i = 1|A_t, N, m, h] = p(A_t)$. Then the probability for network activity A_{t+1}, given network activity A_t in the previous time step, is expressed by the binomial distribution

$$P[A_{t+1}|A_t, N, m, h] = \binom{N}{A_{t+1}} p(A_t)^{A_{t+1}} [1 - p(A_t)]^{N - A_{t+1}}, \quad (11)$$

with the expectation value

$$\langle A_{t+1}|A_t\rangle = N p(A_t) = N - N\left(1 - \frac{m}{N}\right)^{A_t} [1 - \lambda(h)]. \quad (12)$$

We introduce the effective branching parameter $m_{\mathrm{eff}}(A_t)$ to satisfy Eq. (2), i.e.,

$$\langle A_{t+1}|A_t\rangle = m_{\mathrm{eff}}(A_t) A_t + \lambda(h) N, \quad (13)$$

such that

$$m_{\mathrm{eff}}(A_t, N, m, h) = \left(\frac{N}{A_t}\right)\left[1 - \left(1 - \frac{m}{N}\right)^{A_t}\right][1 - \lambda(h)]. \quad (14)$$

We can further compute the network activation rate $A = \langle A_t\rangle$. For this we recall $\langle A_{t+1}|A_t\rangle = N p(A_t)$ and assume stationary activity, where the law of total expectation yields $A = \langle A_t\rangle = \langle\langle A_{t+1}|A_t\rangle\rangle$. Combined with the approximation $\langle(1 - m/N)^{A_t}\rangle \approx (1 - m/N)^{\langle A_t\rangle}$, Eq. (12) results in the mean-field approximation

$$A = N - N\left(1 - \frac{m}{N}\right)^{A} [1 - \lambda(h)]. \quad (15)$$

To solve Eq. (15) for the network rate A, we first rewrite it as

$$(N - A)\ln\left(1 - \frac{m}{N}\right) e^{(N-A)\ln(1-m/N)} = \ln\left(1 - \frac{m}{N}\right) N\left(1 - \frac{m}{N}\right)^{N} [1 - \lambda(h)] \quad (16)$$

and make use of the Lambert W function, defined by $W(z)e^{W(z)} = z$ [51,52], to obtain

$$A(N, m, h) = N - \frac{W\left(\ln\left(1 - \frac{m}{N}\right) N\left(1 - \frac{m}{N}\right)^{N} [1 - \lambda(h)]\right)}{\ln\left(1 - \frac{m}{N}\right)}. \quad (17)$$

In the limit $N \to \infty$, we can replace $\ln(1 - \frac{m}{N})N \to -m$ and $(1 - \frac{m}{N})^N \to e^{-m}$ to obtain, for the activation rate per node

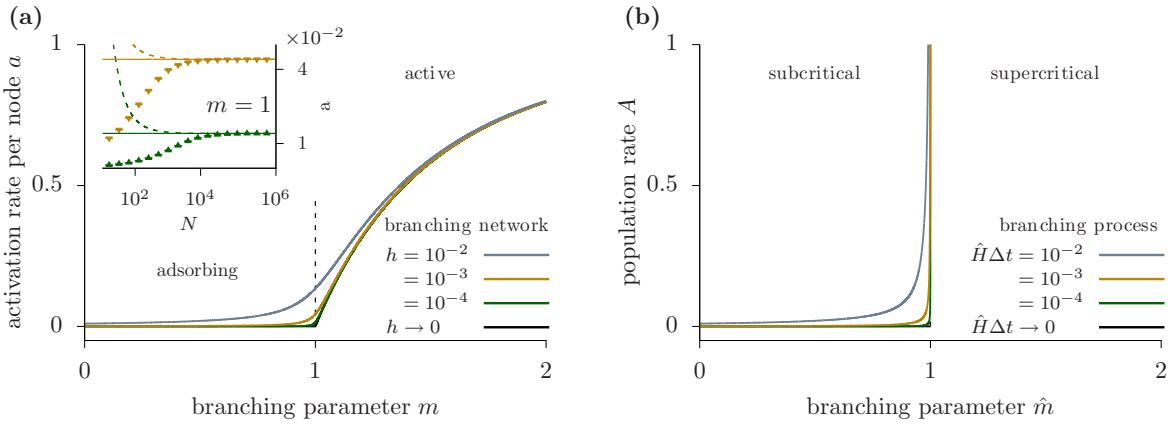

FIG. 3. (a) Analytic solution for the activation rate per node a as a function of the branching parameter m in a mean-field branching network ($N \to \infty$) for different external input rates h (encoded by color). For $h \to 0$, the branching network undergoes a critical nonequilibrium phase transition from an absorbing phase ($a = 0$) to an active phase ($a > 0$). The inset shows a as a function of network size N at the critical point ($m = 1$), comparing the analytic solution (17) (dashed lines) and asymptotic limit (18) (solid lines) with simulation results (data points) for exemplary external input rates (color as in the main plot). (b) Analytic solution for the population rate in a branching process (3). The activity diverges for $m \to 1$.

$a = A/N$,

$$a(m,h) \xrightarrow[N\to\infty]{} 1 + \frac{W(-me^{-m}[1-\lambda(h)])}{m}. \tag{18}$$

This asymptotic solution characterizes the phase diagram of the mean-field branching network [Fig. 3(a)]: For $h \to 0$ a critical point ($m = 1$) separates an absorbing phase ($m < 1$ and $a_t = 0$) from an active phase ($m > 1$ and $a_t \neq 0$). This transition belongs to the universality class of directed percolation. Formally, for $h \neq 0$ there is no critical absorbing phase transition [37], but quasicritical dynamics has been observed for specific connectivities [53]. Moreover, a recent study derived a critical line in the m-h plane that shows critical exponents from both directed and undirected percolation [54]. Still, we will here refer to $m = 1$ as critical-like dynamics, because for small h the dynamics on finite networks is very similar to true critical dynamics.

Our numerical results verify that in the limit $N \to \infty$ the rate per node converges to the analytic solution [Fig. 3(a), inset]. The difference for small system sizes can be explained by the presence of a temporally absorbing state ($A_t = 0$), where the system stays silent until reactivated by external input. This state is not captured by the mean-field approximation (assuming stationary activity) such that Eq. (15) overestimates a for small N.

The phase diagram of the mean-field branching network is qualitatively different from that of a branching process (Fig. 3). For the branching process, the critical point ($\hat{m} = 1$) separates a subcritical phase ($\hat{m} < 1$) from a supercritical phase ($\hat{m} > 1$). For a driven branching process this means that stationary activity is only possible in the subcritical phase [Eq. (3) and Fig. 3(b)], while activity always diverges in the supercritical phase. This hints at a potential bias when approximating network activity in a branching network by a branching process. In order to compare with the branching-process approximation, we nonetheless use the notion of subcritical-like ($m < 1$), critical-like ($m = 1$), and supercritical-like ($m > 1$) spreading dynamics for the branching network.

B. Analytic derivations for the result of branching-process approximations in driven branching networks

In the following we consider the three branching-process approximations introduced in Sec. II C and derive their systematic deviations from the model parameter (their bias) due to coalescence. We assume that the external input h is a system-size-independent constant.

The simplest case is the branching-process approximation through the expected network rate $\hat{m}_{\mathrm{ER}}$. Here we can simply insert the mean-field solution of the network rate (18) to obtain

$$\hat{m}_{\mathrm{ER}}(m,h) = 1 - \frac{mh\Delta t}{m + W(-me^{-m}[1-\lambda(h)])}. \tag{19}$$

If the external input rate h is known, which is usually not the case, this estimator is not as biased as one would naively expect (Fig. 4, gray dashed line). For $m < 1$, the estimator $\hat{m}_{\mathrm{ER}}$ works well. However, it starts to get biased around $m = 1$

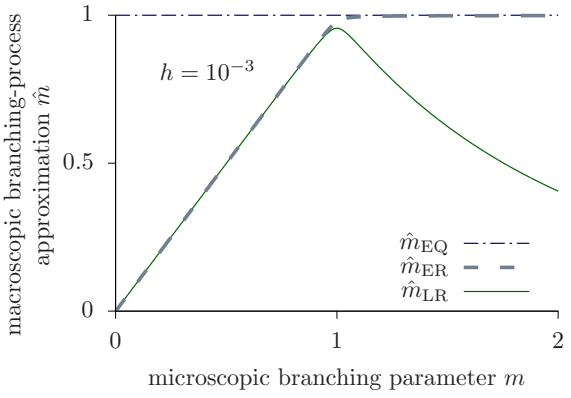

FIG. 4. Asymptotic ($N \to \infty$) macroscopic branching parameter estimates $\hat{m}$ as a function of the microscopic (model) branching parameter m of a driven branching network. The drive is chosen to be relatively large ($h = 10^{-3}$) to make the effects around $m = 1$ visible.

and for $m > 1$ saturates to its upper bound $\hat{m}_{\mathrm{ER}} \to 1$. This bias is due to the fact that in the branching process stationary activity can only be realized for subcritical dynamics ($m < 1$), while in the branching network stationary activity occurs also in the active phase ($m > 1$).

Next we calculate the asymptotic result of the branching-process approximation through the expected quotient of subsequent network activity $\hat{m}_{\mathrm{EQ}}$. This estimator has been frequently applied to process neural data. For our derivation, we assume a driven network with $Nh \gg 1$, so the network activity is practically always nonzero, i.e., $A_t > 0$. This is similar to the case of increasing bin size when processing neural data [20]. Assuming an effective branching process with $A_t > 0$ for all t, we can approximate Eq. (5) as $\hat{m}_{\mathrm{EQ}} \approx 1 + Nh(\langle \frac{1}{A_t} \rangle - \frac{1}{\langle A_t \rangle})$ (see the Appendix). By Jensen's inequality for the convex function $1/x$ we find $\langle \frac{N}{A_t} \rangle > \frac{N}{\langle A_t \rangle}$ such that $\hat{m}_{\mathrm{EQ}}$ should approach its limit from above. One can expect that fluctuations around A_t vanish in the limit of large system sizes, unless exactly at the critical point, such that $\langle \frac{N}{A_t} \rangle \approx \frac{N}{\langle A_t \rangle}$ and $\hat{m}_{\mathrm{EQ}} \approx 1$. Assuming a continuous behavior of the estimator, we expect the same result at the critical point such that

$$\lim_{N\to\infty} \hat{m}_{\mathrm{EQ}}(m, h) = 1 \,\forall\, m, \quad h > 0. \tag{20}$$

Therefore, in the limit $N \to \infty$ (fixed $h > 0$) any stationary activity in the driven regime is interpreted as a persistent internal spread, i.e., $\hat{m}_{\mathrm{EQ}}$ always infers $m = 1$ (Fig. 4, blue solid line). This is because the definition of $\hat{m}_{\mathrm{EQ}}$ implicitly assumes STS, or equivalently $h \to 0$, and under this assumption only $m = 1$ would produce stationary activity on expectation. Hence it is a correct estimator, as long as there is only internal activation, but fails as soon as a novel input h is applied while the network is still active ($A_t > 0$). The estimator thus does what it is supposed to do, but does not help us quantify the amount of internal activation in any driven regime like the living brain.

The third estimator $\hat{m}_{\mathrm{LR}}$ is based on a branching-process approximation through linear regression. It does not rely on knowledge of h and does not require any STS or specific regime for A_t. This estimator returns reliable results in the subcritical regime. In the vicinity of $m = 1$, we show that its asymptotic result can be fully attributed to nonvanishing coalescence effects. In detail, the asymptotic estimate can be calculated from the conditional expectation value $\langle A_{t+1}|A_t \rangle$. Normalizing Eq. (12), we find a system-size-independent scaling function $F(A_t/N)$ for the activity per node $a_t = A_t/N$:

$$\begin{aligned}
\langle A_{t+1}|A_t \rangle/N &= 1 - \left(1 - \frac{m}{N}\right)^{A_t} [1 - \lambda(h)] \\
&= 1 - \left(1 - \frac{m}{N}\right)^{NA_t/N} [1 - \lambda(h)] \\
&\simeq 1 - e^{-mA_t/N} [1 - \lambda(h)] \\
&= 1 - e^{-(h\Delta t + mA_t/N)} = F(A_t/N).
\end{aligned} \tag{21}$$

Indeed, numerical results of the normalized conditional expectation value covering system sizes from 2^5 up to 2^{20} all collapse onto this universal scaling function (Fig. 5, $m = 1$ and $h = 10^{-3}$). With increasing system size, the variance of the average activity decreases and the numerical results

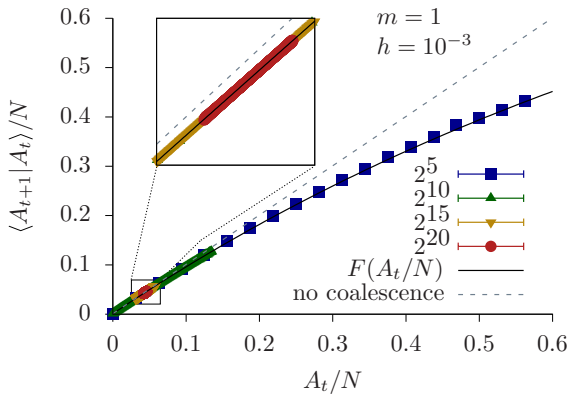

FIG. 5. Universal scaling function describes effective spreading of network activity for sufficiently large system sizes (here $N \geqslant 2^5$, $m = 1$, and $h = 10^{-3}$). Rescaling the conditional expectation value as $\langle A_{t+1}|A_t \rangle/N$ (data points) leads to a data collapse onto the universal scaling function $F(A_t/N)$ (solid line) defined in Eq. (21). Due to the nonlinear character of $F(A_t/N)$, a linear-regression estimate results in a slope that is slightly different from the case without coalescence (dashed line).

localize along the scaling function, justifying the mean-field assumptions above.

From the curvature of the scaling function we can derive the asymptotic result of the linear-regression estimator $\hat{m}_{\mathrm{LR}}$. The linear regression assumes a linear shape of the conditional expectation value (independence of A_{t+1} from A_t). As a consequence, $\hat{m}_{\mathrm{LR}}$ locally fits a straight line to the scaling function with diminished slope (Fig. 5). The slope of this line depends on the rate per node. For a nonzero rate, it will deviate from the ideal case without coalescence (Fig. 5, dashed line) and thereby the estimate will differ from the model parameter. Because the variance of the rate per node decreases with increasing system size, we can calculate the asymptotic estimate as the derivative of Eq. (21) at the average rate (18), i.e.,

$$\begin{aligned}
\hat{m}_{\mathrm{LR}}(m, h) &= \frac{d}{dA_t/N} F(A_t/N)\Big|_{A(m,h)/N} \\
&= m e^{-m\, a(m,h)} [1 - \lambda(h)] \\
&= m e^{-m - h\Delta t} e^{-W(-m e^{-m-h\Delta t})}.
\end{aligned} \tag{22}$$

To illustrate the bias, we can expand $\hat{m}_{\mathrm{LR}}$ for small $a \approx 0$, i.e., in the vicinity of the critical point, to obtain

$$\hat{m}_{\mathrm{LR}}(m, h) \approx [m - m^2 a(m, h)][1 - \lambda(h)]. \tag{23}$$

The asymptotic estimate $\hat{m}_{\mathrm{LR}}$ is thus biased ($\hat{m}_{\mathrm{LR}} < m$), which is most prominent in the vicinity of the critical point and the active phase (Fig. 4, solid green line, and Fig. 7, dashed lines). Importantly, the asymptotic bias vanishes for $h \to 0$ only within the absorbing phase ($m \leqslant 1$) where $a = 0$.

To summarize, in a driven regime all estimators of the microscopic dynamics are biased close to critical-like settings. Some are biased for the whole parameter range and reflect only the presence of drive (such as $\hat{m}_{\mathrm{EQ}}$). Others are deviating slightly from the true values for $m < 1$, but are strongly biased for $m \gtrsim 1$ (Fig. 4). The reason for the asymptotic bias is the coalescence in branching networks (Sec. III A). In the

following sections, we will numerically verify our analytical predictions on the estimation bias with finite-size scaling analyses of branching networks, for both the STS (Sec. III C) and the driven (Sec. III D) regime.

C. Finite-size scaling analysis in the STS regime reveals no asymptotic bias in the branching-process approximation

In the STS regime, activity (also called an avalanche) is initiated at a random node and evolves without any external input until the end. Formally, this corresponds to the limit of vanishing external input rate $h \to 0$. In numerical implementations one can skip periods of zero activity: Directly after activity has ceased, a new random node is initiated at $t+1$. In this regime the internal network dynamics depends only on the model parameter m. The finite-size scaling limit is well defined for $N \to \infty$ by keeping m fixed.

We focus on the two estimators $\hat{m}_{\mathrm{EQ}}$ and $\hat{m}_{\mathrm{LR}}$. The third one $\hat{m}_{\mathrm{ER}}$ requires stationary activity that is not present in the STS regime. In principle, both estimators are suitable for the formal STS regime, but the numerical implementation (skipping periods of zero activity) requires a little attention. The expected-quotient estimator $\hat{m}_{\mathrm{EQ}}$ is not affected by skipping periods with zero activity, because it only considers time steps with $A_t > 0$. The linear-regression estimate $\hat{m}_{\mathrm{LR}}$ is however strongly affected by the choice of skipping periods of zero activity, because periods of zero activity contribute to the estimated external input rate relevant for the linear regression. We thus need to modify the linear-regression estimator (8) for the STS regime by imposing a zero expectation of network activity $\langle A_t \rangle = 0$. This enters both the covariance and variance and we obtain

$$\hat{m}_{\mathrm{LR}}^{\mathrm{STS}} = \frac{\langle A_{t+1} A_t \rangle}{\langle A_t^2 \rangle}. \tag{24}$$

Keep in mind that this is only required because of the artificial implementation of the STS regime ($h \to 0$) as nonstationary activity. In addition, the implicit assumption of vanishing network rate $\langle A_t \rangle = 0$ confines our discussion to the absorbing phase $m \leqslant 1$.

In the STS regime, one may expect that the branching-process approximation from network activity is not biased in the limit $N \to \infty$. This is because small avalanches and equivalently small A_t occur statistically more often than large avalanches, even for critical spreading dynamics. More precisely, the same small-avalanche regime of the characteristic avalanche-size distribution remains dominant with increasing system size. At the same time, for annealed disorder the number of potential connections increases with system size, such that instances of internal coalescence become less probable. Together, we expect that in the STS regime macroscopic estimates reflect microscopic dynamics in the limit $N \to \infty$.

Indeed, the bias in the estimates of $\hat{m}_{\mathrm{EQ}}$ and $\hat{m}_{\mathrm{LR}}^{\mathrm{STS}}$ decreases rapidly with increasing system size N (Fig. 6). In the absorbing phase ($m = 0.9$), the bias diminishes rapidly and is below 0.001% for $N > 10^4$. For critical dynamics ($m = 1$), the bias decreases as a power law $1 - \hat{m}_{\mathrm{x}} \propto N^{\alpha_{\mathrm{x}}}$. Least-squares fits yield $\alpha_{\mathrm{EQ}} = 0.558(7)$, for $N > 32$ with $\chi^2 \approx 1.4$, and $\alpha_{\mathrm{LR}} = 0.490(2)$, for $N > 512$ with $\chi^2 \approx 0.3$. Upon changing the fit range, however, estimates vary outside of their statistical errors, compatible with an overall scaling of the form $1 - \hat{m} \propto N^{-1/2}$.

D. Finite-size scaling analysis in the (Poisson) driven regime reveals asymptotic bias in branching-process approximations

We now consider the driven regime of a network subject to homogeneous Poisson input. Considering additional stochastic external input, we need to specify how the infinite-size limit ($N \to \infty$) is approached: Here we choose to fix the microscopic model parameter m as well as the average external input rate per node h. This assumes that the external input rate scales with system size. We focus on the two estimators $\hat{m}_{\mathrm{EQ}}$ and $\hat{m}_{\mathrm{LR}}$, because they do not rely on knowledge of h.

We first show that in the driven regime the commonly employed expected-quotient estimator $\hat{m}_{\mathrm{EQ}}$ will always indicate critical-like dynamics ($m = 1$) for large system sizes as predicted by Eq. (20) [Fig. 7(a)]. If the dynamics is indeed critical like, $\hat{m}_{\mathrm{EQ}}$ first underestimates the microscopic dynamics for small N ($\hat{m}_{\mathrm{EQ}} < 1$), overestimates the dynamics for intermediate N ($\hat{m}_{\mathrm{EQ}} > 1$), and finally converges to the true microscopic

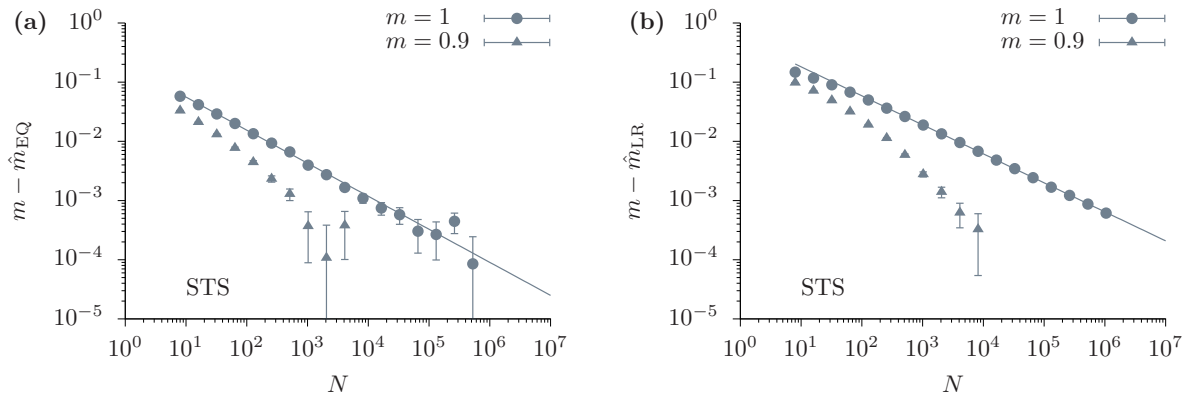

FIG. 6. Finite-size scaling in the separation-of-timescales regime for two standard estimators of the macroscopic branching parameter: (a) the expected-quotient estimator $\hat{m}_{\mathrm{EQ}}$ and (b) the linear-regression estimator $\hat{m}_{\mathrm{LR}}$. Both estimators converge to the microscopic branching parameter for $m \leqslant 1$. In the absorbing phase ($m = 0.9$), the macroscopic estimate quickly converges towards the microscopic model parameter. For critical dynamics ($m = 1$) the macroscopic estimates converge towards the microscopic parameter as a power law $1 - \hat{m}_x \sim N^{-\alpha}$ (see the text for details).

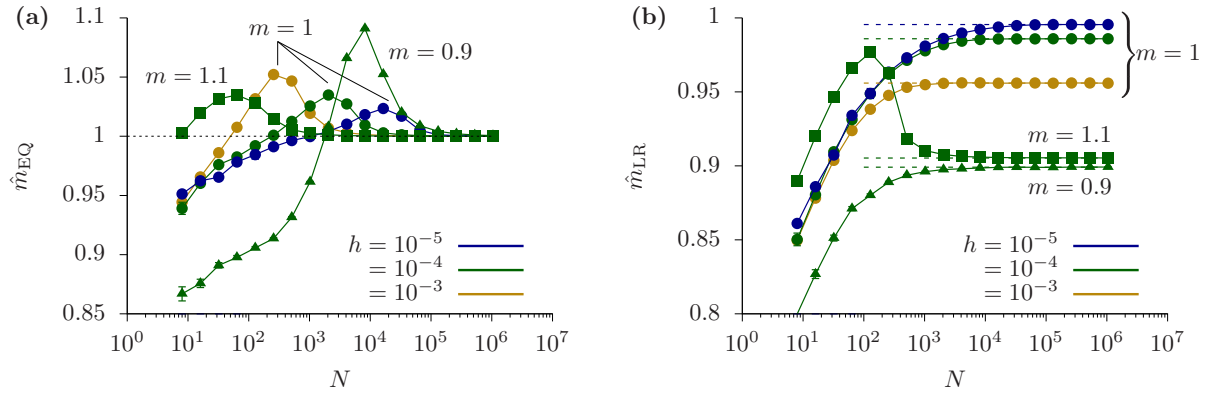

FIG. 7. Finite-size scaling in the driven regime. The two estimators of the macroscopic branching parameter, (a) the expected-quotient estimator $\hat{m}_{\mathrm{EQ}}$ and (b) the linear-regression estimator $\hat{m}_{\mathrm{LR}}$, both fail to measure the microscopic branching parameter due to coalescence effects (solid lines are a guide to the eye). Dashed lines are the calculated infinite-size limits of the respective estimates: (a) Eq. (20) and (b) Eq. (22).

dynamics for $N \to \infty$. The regime of overestimation shifts to larger systems sizes and decreases its amplitude as the input rate decreases. While this would seem reasonable for $m = 1$, the estimator $\hat{m}_{\mathrm{EQ}}$ fully fails for subcritical-like ($m = 0.9$) and supercritical-like ($m = 1.1$) dynamics. Independently of the true m, whether it is smaller than, larger than, or equal to unity, the estimator returns $\hat{m}_{\mathrm{EQ}} < 1$ for small N and a local maximum with $\hat{m}_{\mathrm{EQ}} > 1$ for intermediate N and eventually converges towards $\hat{m}_{\mathrm{EQ}} \to 1$ in the limit $N \to \infty$. This is clearly not the microscopic dynamics. Strikingly, $\hat{m}_{\mathrm{EQ}}$ would thereby predict critical-like dynamics for sufficiently large networks $N > 10^5$ even though the microscopic dynamics is clearly not critical-like.

The system-size dependence of $\hat{m}_{\mathrm{EQ}}$, showing a maximum ($\hat{m}_{\mathrm{EQ}} > 1$) and converging towards $\hat{m}_{\mathrm{EQ}} \to 1$ in the limit $N \to \infty$, is similar to results obtained when changing the bin size for neural spike recordings [20]. The initial increase and maximum can therefore be explained as extended periods of activity separated by a decreasing number of time steps with zero activity. The eventual convergence towards unity can then be explained by the resulting stationary activity and the absence of zero activity due to the increasing amount of network input hN (Sec. III B).

Next we show that in the driven regime the linear-regression estimator $\hat{m}_{\mathrm{LR}}$ underestimates microscopic dynamics (Fig. 7(b)). While this estimator is specifically constructed to infer m from driven systems, it does not consider coalescence. Coalescence leads to a bias even in the infinite-size limit as predicted by Eq. (22) and verified by our numerical results [Fig. 7(b)]. The system size above which $\hat{m}_{\mathrm{LR}}$ saturates corresponds to the system size above which also the rate per node saturates (Fig. 3, inset).

We want to point out two observations. First, the asymptotic bias of $\hat{m}_{\mathrm{LR}}$ depends on h, demonstrated here only for critical-like dynamics ($m = 1$), where the effect of h is strongest. Second, in the limit $h \to 0$ the asymptotic bias of $\hat{m}_{\mathrm{LR}}$ only vanishes for $m \leqslant 1$. This means that for supercritical-like dynamics ($m > 1$) the asymptotic estimator $\hat{m}_{\mathrm{LR}}$ remains biased even for $h \to 0$ and indicates subcriticality $\hat{m}_{\mathrm{LR}} < 1$ (Fig. 4). This is due to the stationary nonzero activity per node in the active phase ($m > 1$) of the branching network [cf. Fig. 3(a)].

E. Coalescence can be captured by a nonlinear estimator

We can use our analytic results to obtain the model parameter m without bias from the macroscopic network activity by directly fitting the nonlinear function (21) to the data. This defines our nonlinear regression estimator $\hat{m}_{\mathrm{NLR}}$ as a fit to the nonlinear scaling function

$$\langle A_{t+1}|A_t\rangle/N = 1 - e^{-(h\Delta t + \hat{m}_{\mathrm{NLR}} A_t/N)}. \tag{25}$$

We implemented this as a PYTHON curve fit. For our numerical data this nonlinear approach correctly infers the microscopic model parameter from the macroscopic network activity for almost all system sizes (Fig. 8). Of course, this relies on a universal scaling function for the conditional expectation value $\langle A_{t+1}|A_t\rangle = NF(A_t/N)$, here derived for an annealed disorder average. A similar scaling function has been derived within a mean-field approximation for a quenched disorder average over Erdős-Rényi networks with average degree K [15],

$$F(A_t/N) = 1 - \left(1 - \frac{mA_t}{NK}\right)^K [1 - \lambda(h)], \tag{26}$$

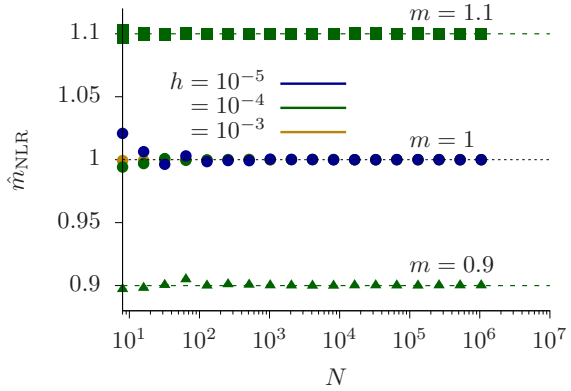

FIG. 8. The novel linear estimator $\hat{m}_{\mathrm{NLR}}$ correctly infers the microscopic model parameter m from the macroscopic dynamics in all-to-all connected branching networks for supercritical-like ($m = 1.1$), critical-like ($m = 1$), and subcritical-like ($m = 0.9$) dynamics.

where we neglected the refractory period to compare with our results. The expansion of Eq. (26) agrees to leading order with the expansion of Eq. (21).

IV. DISCUSSION

To summarize, we have shown that due to coalescence (the simultaneous activation of the same node from multiple sources) in a branching network, the approximation of network activity by a branching process can be biased. As the branching-process approximation of network activity is the basis for several linear estimators of spreading dynamics on networks, these estimators can be consequently biased as well. We verified this bias for an estimator based on the expected quotient of subsequent activities ($\hat{m}_{\mathrm{EQ}}$) and an estimator based on linear regression ($\hat{m}_{\mathrm{LR}}$). In the separation-of-timescale (STS) regime, which we argued is only well-defined in the absorbing phase ($m \leqslant 1$), the bias vanishes for $N \to \infty$. In the driven regime of nonvanishing input rate, there always remains an asymptotic bias for $N \to \infty$. We showed how to analytically compute the asymptotically remaining bias in the driven regime and verified it by a finite-size scaling analysis of simulation results.

If interested in an unbiased estimator of the microscopic branching parameter from the network activity, we provide a nonlinear estimator that explicitly takes coalescence into account and is thereby not biased. The nonlinear estimator is derived from the analytic solution of the conditional expectation value of subsequent network activities $\langle A_{t+1}|A_t\rangle$. This conditional expectation value has a universal scaling function for annealed disorder (derived here) and quenched disorder over random Erdős-Rényi networks (derived in Ref. [15]). If the scaling function is not known analytically, we propose that it can be obtained by inducing a data collapse (systematically rescaling the x and y axes) for the conditional expectation value measured in small system sizes with high numerical precision (cf. Fig. 5). While this approach is directly applicable to models (where it benefits from the large variance in average activity a for small system sizes), it cannot be applied trivially to experimental data. For one, the scaling function would need to be deduced from a representative model. In addition, our current results require fully sampled network activity. However, advances in recording techniques, e.g., optogenetic imaging of neural activity [55–57] or high report rates for measles in Germany [21], may enable to construct nonlinear estimators applicable to experimental data even in large systems.

When it comes to approximating real data with a macroscopic branching process, the potential bias (assuming that there is a static interaction network) can be evaluated on a case-to-case basis. For example, cortical neural network dynamics have been estimated (using a subsampling-invariant version of $\hat{m}_{\mathrm{LR}}$ [21]) to be slightly subcritical, $\hat{m} \approx 0.98$ [21,58], assuming spike propagation with $\Delta t = O(1\ \mathrm{ms})$. A branching parameter $m = 0.98$ means that in a stationary activity regime 98% of the firing rate is generated internally, while 2% is generated externally [see Eq. (7)]. For a firing rate per neuron of $O(10^{-3}\ \mathrm{ms}^{-1})$, the 2% external generation corresponds to an external input rate per neuron $h = O(10^{-5}\ \mathrm{ms}^{-1})$. These estimates are consistent with numerical predictions for cortical network dynamics suggesting $h\Delta t = O(10^{-4})$ or lower [22]. In this case, our results for the linear-regression estimator would predict a bias $m - \hat{m}_{\mathrm{LR}} = O(10^{-3})$. This bias is an order of magnitude smaller than the typical observed values for cortical brain networks *in vivo* ($\hat{m} \approx 0.98$) [21]. The largest observed values though reach $\hat{m} \approx 0.994$ [58], suggesting that certain neural networks approach the critical point almost as close as possible, given potential coalescence.

It is *a priori* unclear whether the microscopic model parameter m or the macroscopic parameter $\hat{m}$ is the correct description of the dynamics. We expect that the most reliable estimator of macroscopic dynamics is the linear-regression estimator $\hat{m}_{\mathrm{LR}}$, because it explicitly considers the external input rate present in many practical situations [21,49,59]. Whether inferring the microscopic or macroscopic branching parameter is "correct," however, depends on what is sought. On the one hand, if we want to infer the microscopic dynamics to understand microscopic processes, we seek m, for example, to characterize the impact of a single spike on the amount of subsequent spike initiations [60] or to predict probable routes of disease spreading in complex networks [29]. On the other hand, if we want to describe the time evolution of the collective network activity, we are interested in the macroscopic $\hat{m}$, for example, to characterize intrinsic timescales of cortical areas [61] or to estimate whether or not a disease has epidemic character [21,62]. The interpretation of the correct branching parameter thus depends on whether we want to know the microscopic or the macroscopic dynamics of a particular system.

The difference between microscopic and macroscopic dynamics of a system may be exploited by networks in reality. For example, a network might need to regulate directly the microscopic dynamics (for which it could adjust its local weights) or it might need to regulate the macroscopic dynamics. One option to achieve the latter is by implementing coalescence-compensating mechanisms [63,64], e.g., adaptive synaptic weights or a probabilistic (linear) integrate-and-fire mechanism. Using coalescence-compensating dynamics, the macroscopic dynamics is directly related to the local weights. It is conceivable that reality interpolates between both extremes. This is a regime where dependent-spreading models become relevant, which have not been covered in the present work.

Due to coalescence, the nonequilibrium phase transition in branching networks differs from that in branching processes. On the one hand, for branching networks without external input, the critical point separates an absorbing ($A = 0$) from an active ($A > 0$) phase, a critical phase transition in the universality class of directed percolation [10,37]. In fact, the branching network defined in this work is equivalent to mean-field directed bond percolation. Here the order parameter is the network activity. On the other hand, for branching processes the critical point separates a subcritical (zero probability for infinite avalanche or activity) from a supercritical (nonzero probability for infinite avalanche) phase [6]. Here the order parameter is the probability for infinite avalanches. However, the expected population activity for a subcritical branching process is indeed zero. Hence, the branching network and branching process share universal features in the absorbing

or subcritical phase, while their activities vary substantially in their active or supercritical phase.

We considered in our study homogeneous external input rates per node. This is clearly a leading-order approximation. In the context of neural networks, a homogeneous input rate per node can be motivated by networkwide input that cortical areas receive. In the context of infectious diseases, a homogeneous external input rate corresponds to a homogeneous external infection rate throughout the environment. It is natural to expect that input rates are in fact more heterogeneous. In the context of neural networks, the functional wiring of cortical layers induces heterogeneity (see, e.g., Refs. [65,66]). In the context of infectious diseases, it seems natural that external infection rates depend on local environmental variables (see, e.g., Ref. [67]). We expect that heterogeneous input rates may contribute additional sources of bias in the branching-process approximation of spreading dynamics. However, heterogeneous external input can also mean local input that does not scale proportionally to the system size, e.g., local stimulation of a neural culture or the initial onset of a disease outbreak. In this limit, our results are not directly applicable, e.g., we would expect that $\hat{m}_{\mathrm{EQ}}$ is no longer strongly biased.

We focused in our study on the case of annealed disorder, mathematically equivalent to an all-to-all connected network. This mean-field assumption turns out to be a good leading-order approximation for many complex networks [68]. Our annealed disorder also covers memoryless temporal networks [69]. However, we have also shown that several of our analytical results agree to leading order with those obtained for quenched disorder of random networks [15]. Indeed, in Ref. [63] we show that our analytic solution (18) describes to leading order the average rate a in a random Erdős-Rényi graph of size $N = 10^4$ with average degree $K = 10$. We thus expect that the majority of our conclusions are valid to leading order for general random networks, but at the same time we expect that details depend on the considered network topology. For example, the vanishing bias in the STS regime requires a large number of connections per node in the limit $N \to \infty$. This would be guaranteed if the average degree K would be coupled to the system size, i.e., K/N constant. However, reality may be quite different. In the scope of neuroscience, one expects sparsely connected networks with $K/N \to 0$ [70]. Also in the scope of infectious diseases in human contact networks, one expects a finite number of interaction links [71]. Moreover, there is an increasing number of studies that identify aspects of heterogeneous network topology that affect collective network dynamics in general [72–74], or specifically collective dynamics in neural networks [34,75–79] and for infectious diseases [80–82]. We expect that the branching-process bias remains a general leading-order effect in heterogeneous network topologies. For the large degree networks similar results were obtained in a recent study on the breakdown of treelike approximations [83]. A systematic future investigation of a branching-process bias in heterogeneous topologies could include the study of small-world networks [72] or spatially structured networks [75].

The simple branching networks we considered in this work have to be understood as a minimal example to study the effect of coalescence. Of course, models of real-world processes typically need refinement. For example, while many characteristic statistics for neuronal avalanches are captured by branching networks, the temporal correlations between avalanches and interavalanche intervals [84] cannot be captured in such a simple model. To account for such long-term dynamics a slow adaptation, e.g., a homeostatic regulation [22], is required. These refined models, however, are still subject to coalescence.

Our results are only applicable to fully sampled systems. However, typically, experimental measurements have access to only a small fraction of the system, resulting in subsampled data. Subsampling is a common problem in neuroscience [21,58,85–87], epidemiology [88–92], and networks in general [93]. In the case of subsampled network activity, there is an unbiased way to estimate the effective branching parameter m_{eff} by extending the linear-regression estimator to multiple regressions [21]. We leave it for future work to expand our results on coalescence effects to the subsampled regime.

ACKNOWLEDGMENTS

J.Z. would like to thank Peter Sollich, Peter Grassberger, and Malte Henkel for fruitful discussions. All authors received support from the Max Planck Society. J.Z. and V.P. received financial support from the German Ministry of Education and Research via the Bernstein Center for Computational Neuroscience Göttingen under Grant No. 01GQ1005B. J.W. was financially supported by Gertrud-Reemtsma-Stiftung. A.L. received funding from a Sofja Kovalevskaja Award from the Alexander von Humboldt Foundation, endowed by the Federal Ministry of Education and Research.

APPENDIX: EVALUATION OF THE BRANCHING-PROCESS APPROXIMATION THROUGH THE EXPECTED QUOTIENT $\hat{m}_{\mathrm{EQ}}$ IN THE DRIVEN SYSTEM

In the driven regime, we assume that all $A_t > 0$ for $Nh \gg 0$. We can then evaluate Eq. (5), using Bayes' rule $P[A_{t+1}|A_t] = P[A_{t+1}, A_t]/P[A_t]$,

$$\begin{aligned}\left\langle \frac{A_{t+1}}{A_t} \right\rangle &= \sum_{A_t} \sum_{A_{t+1}} P[A_t, A_{t+1}] \frac{A_{t+1}}{A_t} \\ &= \sum_{A_t} \frac{1}{A_t} P[A_t] \sum_{A_{t+1}} \frac{P[A_t, A_{t+1}]}{P[A_t]} A_{t+1} \\ &= \sum_{A_t} \frac{1}{A_t} P[A_t] \sum_{A_{t+1}} P[A_{t+1}|A_t] A_{t+1} \\ &= \sum_{A_t} \frac{\langle A_{t+1}|A_t \rangle}{A_t} P[A_t]. \end{aligned} \tag{A1}$$

Assuming an effective branching process, we insert $\langle A_{t+1}|A_t \rangle = \hat{m} A_t + hN$ with $\hat{m} = 1 - hN/\langle A_t \rangle$ and obtain

$$\begin{aligned}\left\langle \frac{A_{t+1}}{A_t} \right\rangle &\approx \sum_{A_t} \frac{\hat{m} A_t + hN}{A_t} P[A_t] \approx \hat{m} + hN \left\langle \frac{1}{A_t} \right\rangle \\ &\approx 1 + hN \left(\left\langle \frac{1}{A_t} \right\rangle - \frac{1}{\langle A_t \rangle} \right). \end{aligned} \tag{A2}$$

[1] D. J. Daley and D. G. Kendall, Epidemics and rumours, Nature (London) **204**, 1118 (1964).

[2] Y. Moreno, M. Nekovee, and A. F. Pacheco, Dynamics of rumor spreading in complex networks, Phys. Rev. E **69**, 066130 (2004).

[3] R. Pastor-Satorras, C. Castellano, P. Van Mieghem, and A. Vespignani, Epidemic processes in complex networks, Rev. Mod. Phys. **87**, 925 (2015).

[4] G. F. de Arruda, F. A. Rodrigues, and Y. Moreno, Fundamentals of spreading processes in single and multilayer complex networks, Phys. Rep. **756**, 1 (2018).

[5] A. Garas, P. Argyrakis, C. Rozenblat, M. Tomassini, and S. Havlin, Worldwide spreading of economic crisis, New J. Phys. **12**, 113043 (2010).

[6] T. E. Harris, *The Theory of Branching Processes* (Springer, Berlin, 1963).

[7] M. A. Muñoz, Colloquium: Criticality and dynamical scaling in living systems, Rev. Mod. Phys. **90**, 031001 (2018).

[8] S. Wolfram, Statistical mechanics of cellular automata, Rev. Mod. Phys. **55**, 601 (1983).

[9] E. Domany and W. Kinzel, Equivalence of Cellular Automata to Ising Models and Directed Percolation, Phys. Rev. Lett. **53**, 311 (1984).

[10] H. Hinrichsen, Non-equilibrium critical phenomena and phase transitions into absorbing states, Adv. Phys. **49**, 815 (2000).

[11] T. E. Harris, Contact interactions on a lattice, Ann. Probab. **2**, 969 (1974).

[12] A. G. Hawkes, Spectra of some self-exciting and mutually exciting point processes, Biometrika **58**, 83 (1971).

[13] H. Hethcote, The mathematics of infectious diseases, SIAM Rev. **42**, 599 (2000).

[14] C. Haldeman and J. M. Beggs, Critical Branching Captures Activity in Living Neural Networks and Maximizes the Number of Metastable States, Phys. Rev. Lett. **94**, 058101 (2005).

[15] O. Kinouchi and M. Copelli, Optimal dynamical range of excitable networks at criticality, Nat. Phys. **2**, 348 (2006).

[16] A. Levina, J. M. Herrmann, and M. Denker, Critical branching processes in neural networks, PAMM **7**, 1030701 (2008).

[17] D. B. Larremore, W. L. Shew, and J. G. Restrepo, Predicting Criticality and Dynamic Range in Complex Networks: Effects of Topology, Phys. Rev. Lett. **106**, 058101 (2011).

[18] D. B. Larremore, M. Y. Carpenter, E. Ott, and J. G. Restrepo, Statistical properties of avalanches in networks, Phys. Rev. E **85**, 066131 (2012).

[19] N. Friedman, S. Ito, B. A. W. Brinkman, M. Shimono, R. E. L. DeVille, K. A. Dahmen, J. M. Beggs, and T. C. Butler, Universal Critical Dynamics in High Resolution Neuronal Avalanche Data, Phys. Rev. Lett. **108**, 208102 (2012).

[20] V. Priesemann, M. Wibral, M. Valderrama, R. Pröpper, M. Le Van Quyen, T. Geisel, J. Triesch, D. Nikolić, and M. H. J. Munk, Spike avalanches *in vivo* suggest a driven, slightly subcritical brain state, Front. Syst. Neurosci. **8**, 108 (2014).

[21] J. Wilting and V. Priesemann, Inferring collective dynamical states from widely unobserved systems, Nat. Commun. **9**, 2325 (2018).

[22] J. Zierenberg, J. Wilting, and V. Priesemann, Homeostatic Plasticity and External Input Shape Neural Network Dynamics, Phys. Rev. X **8**, 031018 (2018).

[23] J. Chevallier, Mean-field limit of generalized Hawkes processes, Stoch. Process. Appl. **127**, 3870 (2017).

[24] J. Chevallier, Stimulus sensitivity of a spiking neural network model, J. Stat. Phys. **170**, 800 (2018).

[25] F. Y. K. Kossio, S. Goedeke, B. van den Akker, B. Ibarz, and R.-M. Memmesheimer, Growing Critical: Self-Organized Criticality in a Developing Neural System, Phys. Rev. Lett. **121**, 058301 (2018).

[26] D. B. Larremore, W. L. Shew, E. Ott, F. Sorrentino, and J. G. Restrepo, Inhibition Causes Ceaseless Dynamics in Networks of Excitable Nodes, Phys. Rev. Lett. **112**, 138103 (2014).

[27] J. Li and W. Shew, Tuning network dynamics from criticality to the chaotic balanced state, bioRxiv:551457.

[28] L. Hufnagel, D. Brockmann, and T. Geisel, Forecast and control of epidemics in a globalized world, Proc. Natl. Acad. Sci. USA **101**, 15124 (2004).

[29] D. Brockmann and D. Helbing, The hidden geometry of complex, network-driven contagion phenomena, Science **342**, 1337 (2013).

[30] J. T. Matamalas, A. Arenas, and S. Gómez, Effective approach to epidemic containment using link equations in complex networks, Sci. Adv. **4**, eaau4212 (2018).

[31] V. Belik, T. Geisel, and D. Brockmann, Natural Human Mobility Patterns and Spatial Spread of Infectious Diseases, Phys. Rev. X **1**, 011001 (2011).

[32] J. H. Fowler and N. A. Christakis, Dynamic spread of happiness in a large social network: Longitudinal analysis over 20 years in the Framingham Heart Study, BMJ **337**, a2338 (2008).

[33] B. F. Maier and D. Brockmann, Cover time for random walks on arbitrary complex networks, Phys. Rev. E **96**, 042307 (2017).

[34] P. Moretti and M. A. Muñoz, Griffiths phases and the stretching of criticality in brain networks, Nat. Commun. **4**, 2521 (2013).

[35] E. Bonabeau, Agent-based modeling: Methods and techniques for simulating human systems, Proc. Natl. Acad. Sci. USA **99**, 7280 (2002).

[36] S. Pei and H. A. Makse, Spreading dynamics in complex networks, J. Stat. Mech. (2013) P12002.

[37] M. Henkel, H. Hinrichsen, and S. Lübeck, *Non-Equilibrium Phase Transitions. Volume 1: Absorbing Phase Transitions* (Springer Science + Business Media, New York, 2008).

[38] H. Kesten, Random difference equations and renewal theory for products of random matrices, Acta Math. **131**, 207 (1973).

[39] J. M. Beggs and D. Plenz, Neuronal avalanches in neocortical circuits, J. Neurosci. **23**, 11167 (2003).

[40] V. Pasquale, P. Massobrio, L. L. Bologna, M. Chiappalone, and S. Martinoia, Self-organization and neuronal avalanches in networks of dissociated cortical neurons, Neuroscience **153**, 1354 (2008).

[41] T. Petermann, T. C. Thiagarajan, M. A. Lebedev, M. A. L. Nicolelis, D. R. Chialvo, and D. Plenz, Spontaneous cortical activity in awake monkeys composed of neuronal avalanches, Proc. Natl. Acad. Sci. USA **106**, 15921 (2009).

[42] A. Das and A. Levina, Critical Neuronal Models with Relaxed Timescale Separation, Phys. Rev. X **9**, 021062 (2019).

[43] F. Ball and P. Donnelly, Strong approximations for epidemic models, Stoch. Process. Appl. **55**, 1 (1995).

[44] R. van der Lans, G. van Bruggen, J. Eliashberg, and B. Wierenga, A viral branching model for predicting the spread of electronic word of mouth, Mark. Sci. **29**, 348 (2009).

[45] M. A. Lopes and A. V. Goltsev, Distinct dynamical behavior in Erdős-Rényi networks, regular random networks, ring lattices,

and all-to-all neuronal networks, Phys. Rev. E **99**, 022303 (2019).
[46] C. R. Heathcote, A branching process allowing immigration, J. R. Stat. Soc. Ser. B **27**, 138 (1965).
[47] A. G. Pakes, Branching processes with immigration, J. Appl. Prob. **8**, 32 (1971).
[48] S. Yu, A. Klaus, H. Yang, and D. Plenz, Scale-invariant neuronal avalanche dynamics and the cut-off in size distributions, PLoS One **9**, e99761 (2014).
[49] C. Z. Wei and J. Winnicki, Estimation of the means in the branching process with immigration, Ann. Stat. **18**, 1757 (1990).
[50] M. Martinello, J. Hidalgo, A. Maritan, S. di Santo, D. Plenz, and M. A. Muñoz, Neutral Theory and Scale-Free Neural Dynamics, Phys. Rev. X **7**, 041071 (2017).
[51] R. M. Corless, G. H. Gonnet, D. E. G. Hare, D. J. Jeffrey, and D. E. Knuth, On the Lambert *W* function, Adv. Comput. Math. **5**, 329 (1996).
[52] S. R. Valluri, D. J. Jeffrey, and R. M. Corless, Some applications of the Lambert *W* function to physics, Can. J. Phys. **78**, 823 (2000).
[53] R. V. Williams-García, M. Moore, J. M. Beggs, and G. Ortiz, Quasicritical brain dynamics on a nonequilibrium Widom line, Phys. Rev. E **90**, 062714 (2014).
[54] D. J. Korchinski, J. G. Orlandi, S.-W. Son, and J. Davidsen, Modelling zoonotic diseases, cooperation between spreading and spontaneous infection, arXiv:1908.08163.
[55] G. Nagel, D. Ollig, M. Fuhrmann, S. Kateriya, A. M. Musti, E. Bamberg, and P. Hegemann, Channelrhodopsin-1: A light-gated proton channel in green algae, Science **296**, 2395 (2002).
[56] E. S. Boyden, F. Zhang, E. Bamberg, G. Nagel, and K. Deisseroth, Millisecond-timescale, genetically targeted optical control of neural activity, Nat. Neurosci. **8**, 1263 (2005).
[57] C. K. Kim, A. Adhikari, and K. Deisseroth, Integration of optogenetics with complementary methodologies in systems neuroscience, Nat. Rev. Neurosci. **18**, 222 (2017).
[58] J. Wilting and V. Priesemann, Between perfectly critical and fully irregular: A reverberating model captures and predicts cortical spike propagation, Cereb. Cortex **29**, 2759 (2019).
[59] C. C. Heyde and E. Seneta, Estimation theory for growth and immigration rates in a multiplicative process, J. Appl. Probab. **9**, 235 (1972).
[60] M. London, A. Roth, L. Beeren, M. Häusser, and P. E. Latham, Sensitivity to perturbations *in vivo* implies high noise and suggests rate coding in cortex, Nature (London) **466**, 123 (2010).
[61] J. D. Murray, A. Bernacchia, D. J. Freedman, R. Romo, J. D. Wallis, X. Cai, C. Padoa-Schioppa, T. Pasternak, H. Seo, D. Lee, and X.-J. Wang, A hierarchy of intrinsic timescales across primate cortex, Nat. Neurosci. **17**, 1661 (2014).
[62] C. P. Farrington, M. N. Kanaan, and N. J. Gay, Branching process models for surveillance of infectious diseases controlled by mass vaccination, Biostatistics **4**, 279 (2003).
[63] J. Zierenberg, J. Wilting, V. Priesemann, and A. Levina, Tailored ensembles of neural networks optimize sensitivity to stimulus statistics, Phys. Rev. Research **2**, 013115 (2020).
[64] J. P. Neto, F. P. Spitzner, and V. Priesemann, A unified picture of neuronal avalanches arises from the understanding of sampling effects, arXiv:1910.09984.
[65] D. J. Felleman and D. E. Van, Distributed hierarchical processing in the primate cerebral cortex, Cereb. Cortex **1**, 1 (1991).
[66] K. D. Harris and G. M. G. Shepherd, The neocortical circuit: Themes and variations, Nat. Neurosci. **18**, 170 (2015).
[67] L. A. Real and R. Biek, Spatial dynamics and genetics of infectious diseases on heterogeneous landscapes, J. R. Soc. Interface **4**, 935 (2007).
[68] S. Gómez, A. Arenas, J. Borge-Holthoefer, S. Meloni, and Y. Moreno, Discrete-time Markov chain approach to contact-based disease spreading in complex networks, Europhys. Lett. **89**, 38009 (2010).
[69] A. Moinet, R. Pastor-Satorras, and A. Barrat, Effect of risk perception on epidemic spreading in temporal networks, Phys. Rev. E **97**, 012313 (2018).
[70] N. Brunel, Dynamics of sparsely connected networks of excitatory and inhibitory spiking neurons, J. Comput. Neurosci. **8**, 183 (2000).
[71] M. Salathe, M. Kazandjieva, J. W. Lee, P. Levis, M. W. Feldman, and J. H. Jones, A high-resolution human contact network for infectious disease transmission, Proc. Natl. Acad. Sci. USA **107**, 22020 (2010).
[72] D. J. Watts and S. H. Strogatz, Collective dynamics of 'small-world' networks, Nature (London) **393**, 440 (1998).
[73] M. Kitsak, L. K. Gallos, S. Havlin, F. Liljeros, L. Muchnik, H. E. Stanley, and H. A. Makse, Identification of influential spreaders in complex networks, Nat. Phys. **6**, 888 (2010).
[74] H. Ronellenfitsch, J. Dunkel, and M. Wilczek, Optimal Noise-Canceling Networks, Phys. Rev. Lett. **121**, 208301 (2018).
[75] J. G. Orlandi, J. Soriano, E. Alvarez-Lacalle, S. Teller, and J. Casademunt, Noise focusing and the emergence of coherent activity in neuronal cultures, Nat. Phys. **9**, 582 (2013).
[76] A. Levina, J. M. Herrmann, and T. Geisel, *Criticality in Neural Systems* (Wiley-VCH Verlag, Weinheim, Germany, 2014), pp. 417–436.
[77] F. Effenberger, J. Jost, and A. Levina, Self-organization in balanced state networks by STDP and homeostatic plasticity, PLoS Comput. Biol. **11**, e1004420 (2015).
[78] F. Mastrogiuseppe and S. Ostojic, Linking connectivity, dynamics and computations in low-rank recurrent neural networks, Neuron **99**, 609 (2018).
[79] G. Bondanelli and S. Ostojic, Coding with transient trajectories in recurrent neural networks, arXiv:1811.07592.
[80] M. E. J. Newman, Spread of epidemic disease on networks, Phys. Rev. E **66**, 016128 (2002).
[81] M. D. F. Shirley and S. P. Rushton, The impacts of network topology on disease spread, Ecol. Complex. **2**, 287 (2005).
[82] J. Hindes and I. B. Schwartz, Epidemic Extinction and Control in Heterogeneous Networks, Phys. Rev. Lett. **117**, 028302 (2016).
[83] S. Chandra, E. Ott, and M. Girvan, Critical network cascades with re-excitable nodes: Why tree-like approximations usually work, when they breakdown, and how to correct them, arXiv:1905.07433.
[84] F. Lombardi, H. J. Herrmann, D. Plenz, and L. de Arcangelis, Temporal correlations in neuronal avalanche occurrence, Sci. Rep. **6**, 24690 (2016).
[85] V. Priesemann, M. H. Munk, and M. Wibral, Subsampling effects in neuronal avalanche distributions recorded *in vivo*, BMC Neurosci. **10**, 40 (2009).

[86] T. L. Ribeiro, S. Ribeiro, H. Belchior, F. Caixeta, and M. Copelli, Undersampled critical branching processes on small-world and random networks fail to reproduce the statistics of spike avalanches, PLoS One **9**, e94992 (2014).

[87] A. Levina and V. Priesemann, Subsampling scaling, Nat. Commun. **8**, 15140 (2017).

[88] L. Papoz, B. Balkau, and J. Lellouch, Case counting in epidemiology: Limitations of methods based on multiple data sources, Int. J. Epidemiol. **25**, 474 (1996).

[89] R. M. Cormack, Problems with using capture-recapture in epidemiology: An example of a measles epidemic, J. Clin. Epidemiol. **52**, 909 (1999).

[90] D. Sethi, J. Wheeler, L. C. Rodrigues, S. Fox, and P. Roderick, Investigation of under-ascertainment in epidemiological studies based in general practice, Int. J. Epidemiol. **28**, 106 (1999).

[91] A. Chao, P. K. Tsay, S.-H. Lin, W.-Y. Shau, and D.-Y. Chao, The applications of capture-recapture models to epidemiological data, Stat. Med. **20**, 3123 (2001).

[92] K. Tilling, Capture-recapture methods—useful or misleading? Int. J. Epidemiol. **30**, 12 (2001).

[93] M. P. H. Stumpf, C. Wiuf, and R. M. May, Subnets of scale-free networks are not scale-free: Sampling properties of networks, Proc. Natl. Acad. Sci. USA **102**, 4221 (2005).