Continuous-Time Discrete-Distribution Theory for Activity-Driven Networks
Lorenzo Zino, Alessandro Rizzo, and Maurizio Porfiri
DOI: 10.1103/PhysRevLett.117.228302
A continuous-time discrete-distribution theory for activity-driven networks

Lorenzo Zino 1,∗, Alessandro Rizzo 2,†,1 and Maurizio Porfiri 3

1 Dipartimento di Scienze Matematiche “G. L. Lagrange”, Politecnico di Torino, Torino, Italy
2 Dipartimento di Automatica e Informatica, Politecnico di Torino, Torino, Italy
3 Department of Mechanical and Aerospace Engineering, New York University Tandon School of Engineering, Brooklyn NY, USA

(Dated: October 26, 2016)

Activity-driven networks are a powerful paradigm to study epidemic spreading over time-varying networks. Despite significant advances, most of the current understanding relies on discrete-time computer simulations, in which each node is assigned an activity potential from a continuous distribution. Here, we establish a continuous-time discrete-distribution framework toward an analytical treatment of the epidemic spreading, from its onset to the endemic equilibrium. In the thermodynamic limit, we derive a nonlinear dynamical system to accurately model the epidemic spreading and leverage techniques from the fields of differential inclusions and adaptive estimation to inform shortand long-term predictions. We demonstrate our framework through the analysis of two real-world case studies, exemplifying different physical phenomena and time scales.

PACS numbers: 89.75.Hc, 64.60.aq, 87.23.Ge

The study of time-varying networks has greatly contributed to our understanding of epidemic spreading, pushing the state of the art beyond the limitations imposed by time-invariant networks of contacts [1–9]. Activity-driven networks (ADNs) have emerged as a powerful paradigm to model the co-evolution of the network of contacts and the individual dynamics [10–21]. Most of the studies on ADNs are based on extensive Monte Carlo simulations, and analytical results are only limited to linearized mean-field approximations.

Here, we seek to establish an analytical framework to study the entire dynamics of the epidemic spreading at the population level (from the zero-infected condition to the endemic equilibrium). Differently from the original ADN formulation, where a discrete-time epidemic model is implemented with a continuous probability distribution for the nodes’ activities, we formulate a continuous-time model with a discrete distribution. This change of perspective leads to a rigorous analytical treatment, without the need of extensive Monte Carlo simulations that have constituted the primary tool for the study of ADNs. Our approach is not prone to the confounds associated with the selection of the time step, which has been proven to influence the dynamics of the discrete-time dynamical process [22]. Our theory relies on a reduced number of parameters with respect to traditional ADNs [10–20]. This is critical for robust parameter identification from real-world data [15, 23–26].

We consider a (large) population of $n$ individuals, each associated with a node of a time-varying undirected graph $\mathcal{G}(t) = (\mathcal{V}, \mathcal{E}(t))$, with $t \in \mathbb{R}^+$. $\mathcal{V} = \{1, \ldots, n\}$ is the node set and $\mathcal{E}(t)$ is the time-varying edge set, which is related to the network of contacts. We focus on a susceptible-infected-susceptible process [27]. Each node $v \in \mathcal{V}$ is assigned a time-invariant activity rate $a_v$, which represents the expected number of contacts that node $v$ generates in a unit time interval. Starting from $t = 0$, node $v$ becomes active after a time that is sampled from an exponentially distributed random variable with parameter $a_v$ [28]. When a node activates, it contacts exactly one node uniformly at random in $\mathcal{V}$, generating a single edge. If this edge connects an infected node with a susceptible one, then the epidemics propagates with a fixed probability $\lambda$, otherwise nothing happens. We suppose that the duration of the contact is instantaneous, so that $\lambda$ is considered a per-contact infection probability. The edge is instantaneously removed, and the node may activate again according to the same rule. Each infected node recovers after a time that is drawn from an exponentially distributed random variable with parameter $\mu$, becoming susceptible again. Thus, $\frac{1}{\mu}$ is the expected time needed by an individual to recover [29].

The proposed discrete activity distribution follows a power-law with $k$ equidistant activation classes, characterized by an activity rate $a_i$ ($a_1 < \cdots < a_k$). For the generic $i$-th class, we denote with $n_i$ its number of nodes and we let $n_i \propto a_i^{-\gamma}$. The parameter $\gamma$ controls the heterogeneity among individuals, similar to the classical ADN paradigm with a continuous distribution of activity potentials [30].

We indicate with $Y_v(t) \in \{S, I\}$ the state of node $v$ at time $t$, which can be either susceptible ($S$) or infected ($I$), and we assemble all the states in a vector $Y(t) \in \{S, I\}^\mathcal{V}$. Towards analyzing the epidemic spreading at the population level, we map $Y(t)$ to a $k$-dimensional stochastic process $Z(t) := Z[Y(t)]$, encapsulating the fraction of infected nodes in each activation class. The $i$-th compo-
number of activation classes

0

0.332

0

minute

j

flu

activation rate of nodes in the

0.430

2.10

2.09

macroscopic variables,

0

59

fraction of infected nodes in the

per-contact infection probability

nomena, modeled through ADNs: flu spreading in a uni-
we consider two different dynamics on real-world phe-
between continuous- and discrete-time epidemic models,
To verify this claim and demonstrate the correspondence
for the transient and endemic equilibrium of the system.

For (1), we appreciate that the drift in the fraction of
infected nodes in each class is determined by three effects:
the recovery of infected nodes (−µζi); the spreading
associated with active nodes in the i-th class generating
contacts toward infected nodes (λ(1 − ζi)x1); and the
spreading related to active infected nodes generating con-
tacts with the nodes of the i-th class (λ(1 − ζi)x2).

The elegant form of the system dynamics (1) in terms of
the variables ζ1, . . . , ζk lends itself into rigorous and
revealing schemes to gain insight into the physics of the
epidemic spreading. Here, we focus on two complementa-
ry strategies that could be systematically utilized for
short- and long- term predictions. First, we propose the
use of differential inclusions to establish rigorous bounds
for the transient and endemic equilibrium of the system.
Second, we explore the integration of estimation tech-
niques to accurately predict the population of infected
individuals from sporadic data which could be collected
in real-world scenarios.

Integrating (1) allows to closely simulate the epidemic
spreading without the need of Monte Carlo simulations.
To verify this claim and demonstrate the correspondence
between continuous- and discrete-time epidemic models,
we consider two different dynamics on real-world phe-
omena, modeled through ADNs: flu spreading in a uni-
versity campus and trend diffusion on Twitter. System
parameters are obtained from case studies [10, 32–35],
as detailed in the supplementary material, and are sum-
marized in Table II. We compare the outcome of Monte
Carlo simulations averaged over 200 trials for both the
continuous- and the discrete-time processes, along with
the integration of the deterministic system (1). In both
examples, the activity distribution is discretized over
k = 59 equidistant activation classes. Fig. 1 demon-
strates the equivalence of our approach with respect to
traditional ADNs in Monte Carlo simulations, along with
the validity of equation (1) to exactly predict the epi-
demic spreading.

To facilitate the mathematical treatment of the k-di-

dimensional system (1), we rewrite the system dynamics
in terms of the first k macroscopic variables, x1, . . . , xk.
Through this change of variables, the epidemic spreading
is governed by the following ODEs:

\[
\begin{align*}
\dot{x}_1 &= (\lambda\alpha_1 - \mu)x_1 + \lambda x_2 - 2\lambda x_1 x_2, \\
\dot{x}_2 &= \lambda\alpha_2 x_1 + (\lambda\alpha_1 - \mu)x_2 - \lambda x_1 x_3 - \lambda x_2^2, \\
\dot{x}_3 &= \lambda\alpha_3 x_1 + \lambda\alpha_2 x_2 - \lambda x_3 - \lambda x_1 x_4 - \lambda x_2 x_3, \\
&\ldots \\
\dot{x}_k &= \lambda\alpha_k x_1 + \lambda\alpha_{k-1} x_2 - \mu x_k - \lambda x_1 \sum \eta_h a^k h \zeta_h - \lambda x_2 x_k,
\end{align*}
\]

where αj = \sum \eta_h a^j h are the moments of the activity rates
distribution, whose first two values are also reported in

TABLE I: Notation.

| k | number of activation classes |
| a_i | activation rate of nodes in the i-th class |
| η_h | fraction of nodes in the i-th class |
| λ | per-contact infection probability |
| µ | recovery rate |
| ζ_i | fraction of infected nodes in the i-th class |
| x_j | macroscopic variables, \( x_j = \sum \eta_h a^j h \zeta_h \) |
| \alpha_j | j-th moment of the distribution \( \alpha, \alpha_j = \sum \eta_h a^j h \) |

TABLE II: Parameters of real-world case studies based on ADNs.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>flu</th>
<th>Twitter</th>
</tr>
</thead>
<tbody>
<tr>
<td>n</td>
<td>30896</td>
<td>531788</td>
</tr>
<tr>
<td>k</td>
<td>59</td>
<td>59</td>
</tr>
<tr>
<td>γ</td>
<td>2.09</td>
<td>2.10</td>
</tr>
<tr>
<td>λ</td>
<td>0.430</td>
<td>0.332</td>
</tr>
<tr>
<td>µ</td>
<td>0.138</td>
<td>0.0997</td>
</tr>
<tr>
<td>α1</td>
<td>0.317</td>
<td>0.536</td>
</tr>
<tr>
<td>α2</td>
<td>0.381</td>
<td>0.781</td>
</tr>
<tr>
<td>time unit</td>
<td>day</td>
<td>minute</td>
</tr>
</tbody>
</table>
The study of (2) offers important insight on the epidemic spreading, beyond the mere computation of the epidemic threshold \((\alpha_1 + \sqrt{\alpha_2})^{-1}\) from linear stability analysis \([10–12]\); details are presented in the supplementary material. However, numerical instabilities may emerge when considering power-laws with \(\gamma \in [2,3]\), where all statistical moments from the second onwards may blow up. Moreover, prescribing initial conditions for higher order macroscopic variables beyond \(x_1\) may be not feasible when dealing with experimental data.

A possible approach to address these issues is to project the \(k\)-dimensional dynamics to a lower dimensional space consisting of only \(k^* < k\) equations. We approximate the term \(x_{k^*+1}\) using two elementary bounds: \(a_1x_k \leq x_{k^*+1} \leq a_kx_{k^*}\). Using these bounds, we can reduce system of \(k\) ODEs in (2) to a system of \(k^*\) ordinary differential inclusions (ODIs) \([36]\), consisting of one inclusion and \(k^*-1\) equations.

If \(k^* = 1\), we bound \(a_1x_1 \leq x_2 \leq \min\{a_1, a_kx_1\}\), reducing (2) to a single ODI. This one-dimensional system should not be contemplated to accurately predict the evolution of the process during the transient, between the zero-infected condition and the endemic equilibrium, due to the conservativeness of the bounds during such a transient phase. However, it can be effectively used to analytically determine an interval \(\bar{x}\) for the endemic equilibrium \(\bar{x}_1\), which is

\[
\left[\max\left\{\frac{\lambda a_1}{\lambda a_1 + \mu}, \frac{\lambda (a_k + \alpha_1) - \mu}{2\lambda a_k}, \frac{\lambda (a_1 + \alpha_1) - \mu}{2\lambda a_1}\right\}, \frac{\lambda (a_1 + \alpha_1) - \mu}{2\lambda a_1}\right],
\]

(3a)

if \(\lambda a_1 > \mu\), and

\[
\left[\frac{\lambda (a_1 + \alpha_1) - \mu}{2\lambda a_1}, \min\left\{\frac{\lambda a_1}{\lambda a_1 + \mu}, \frac{\lambda (a_k + \alpha_1) - \mu}{2\lambda a_k}, \frac{\lambda (a_1 + \alpha_1) - \mu}{2\lambda a_1}\right\}\right],
\]

(3b)

if \(\lambda a_1 < \mu\). Notice that, if \(\lambda a_1 = \mu\), we analytically compute \(\bar{x}_1 = 1/2\).

To demonstrate the use of these bounds we refer, here and henceforth, to the two real-world case studies on flu spreading and trend diffusion on Twitter. From simulations in Fig. 2, we evince that the accuracy of the bounds depends on the system parameters. Specifically, our results suggest that the closer is the endemic state to \(\bar{x}_1 = 1/2\) (that is, \(\alpha_1 \lambda = \mu\)), the more precise the bounds are.

An improved prediction of the transient phase is obtained with \(k^* = 2\), which leads to an ODI for the evolution of \(x_2\), coupled to the first ODE in (2). As detailed in the supplementary material, we establish the two fol-
tion accuracy could be improved combining the bounds in Figs. 2 and 3.

An alternative strategy for the analysis of system (2) entails the use of epidemic data, sampled at the population level at a time period $T$, to drive the reduction of the dynamics on a lower dimensional space. Given the accuracy of (4) in estimating the transient response of the system, we focus on a two-dimensional dynamics in terms of $x_1$ and $x_2$. With reference to (2), we consider only the first two ODEs and we hypothesize that $x_3$ is linear in $x_1$ with a proportionality constant $C$ that is estimated from epidemic data. Specifically, we propose the following two-dimensional dynamics:

$$
\begin{align*}
\dot{x}_1 &= -\mu x_1 + \lambda \alpha_1 x_1 + \lambda x_2 - 2\lambda x_1 x_2, \\
\dot{x}_2 &= \lambda \alpha_2 x_1 + (\lambda \alpha_1 - \mu) x_2 - \lambda C x_1^2 - \lambda x_2^2.
\end{align*}
$$

(6)

As a first approximation, we hypothesize that $C$ is constant throughout the entire epidemic spreading and set to $C = \alpha_2$, which corresponds to a homogeneous distribution of infected individuals over all the activation classes. Our prediction of the infected population is informed by the knowledge of the over-time setting times $hT$. In particular, we denote the piece-wise predictions with $x_1^{(h)}(t)$ and $x_2^{(h)}(t)$, in the interval $t \in [hT, hT + T)$ [37, 38], where $h \in \mathbb{Z}^+$. These predictions are informed by the knowledge of the overall infected fraction of population $X_{hT}$ at the sampling times $hT$. We initialize the algorithm at time $t = 0$ by setting $x_1^{(0)}(0) = X_0$ and $x_2^{(0)}(0) = X_{0}\alpha_1$. The algorithm runs through the following steps: i) system (6) is integrated from $hT$ to $(h + 1)T$, producing the solutions $x_1^{(h)}(t)$ and $x_2^{(h)}(t)$; ii) at $t = (h + 1)T$, the initial conditions for marching in time are set as $x_1^{(h+1)}(t) = X_{(h+1)T}$ and $x_2^{(h+1)}(t) = x_2^{(h)}(t)$; and iii) $h$ is incremented by 1 and the process resumes to step i).

This algorithm is only based on the fraction of infected nodes at the inception of each sampling window of duration $T$, which is central for real-world applications. For example, it may be possible to periodically estimate the number of individuals affected by flu or the number of mentions and re-tweets of a specific trend. The knowledge of the detailed state of all the network nodes is not required by the algorithm, which dispenses with information about higher order microscopic variables. In Fig. 4, we demonstrate the use of the prediction algorithm against a simulation for the flu case study, by using a time window of a day or a week. Short-term forecasts (daily) are very close to the real dynamics (the average error is less than 1%), while forecasts on longer horizons (weekly) tend to be less accurate (with an average error around 10%).

To improve on the finite horizon forecast algorithm, we may treat $C$ as piece-wise constant in time and adaptively update it during each prediction window. We initiate the algorithm by setting $C_0 = \alpha_2$. Then, fixing a real constant $\beta > 0$, at the end of each iteration, $C_h$ is updated as [37, 38]

$$
C_{h+1} = C_h \left( 1 + \beta \frac{X_{(h+1)T} - x_1^{(h)}((h+1)T)}{1 - 2X_{(h+1)T}} \right).
$$

(7)

Here, $C$ is incremented by a term that is proportional to the prediction error at the inception of a new prediction window. The effect of the denominator is to change the sign of the increment when $X_{(h+1)T} > 1/2$, following a line of reasoning similar to the one used to define $\phi_{x_1 x_2}(x_1)$ in (5). In Fig. 4, we demonstrate the improvement of the approach, which is successful in closely predicting population level dynamics even with only data available on a weekly basis.

In summary, we have proposed a new framework to study epidemic spreading over ADNs. While intuition may conceive epidemic models to be executed in discrete-time on nodes whose activity is drawn from a continuous distribution, our approach posits a different view. By discretizing the activity distribution and considering a continuous-time evolution, we put forward a mathematically tractable approach to study epidemic spreading from its onset to the endemic equilibrium. In the thermodynamic limit, we have shown that epidemic spreading can be described through a set of coupled ODEs. Techniques from the field of differential inclusions were leveraged to gather insight on the transient response and endemic equilibrium. Toward connecting the theoretical framework with real data, we introduced an adaptive es-

\begin{figure}[h]
\centering
\begin{subfigure}{0.45\textwidth}
\centering
\includegraphics[width=\textwidth]{daily_forecasts}
\caption{Daily forecasts}
\end{subfigure}
\begin{subfigure}{0.45\textwidth}
\centering
\includegraphics[width=\textwidth]{weekly_forecasts}
\caption{Weekly forecasts}
\end{subfigure}
\begin{subfigure}{0.45\textwidth}
\centering
\includegraphics[width=\textwidth]{daily_forecasts}
\caption{Daily forecasts}
\end{subfigure}
\begin{subfigure}{0.45\textwidth}
\centering
\includegraphics[width=\textwidth]{weekly_forecasts}
\caption{Weekly forecasts}
\end{subfigure}
\caption{Simulation of a discrete-time continuous-distribution ADN (blue), for the flu case study, and our predictions over a finite time horizon of one day and one week (red). Predictions in (a) and (b) are obtained with a constant estimate for $C$, while those in (c) and (d) are based on the on-line adaptive update in (7) with $\beta = 1$. A similar result for the Twitter case study is presented in the supplementary material.}
\end{figure}
Our continuous-time discrete-distribution framework offers a rigorous mathematical basis for overcoming some of the limitations of the ADN paradigm. For example, the treatment of non-exponential inter-event times may be tackled by modifying the ODE system (1) to incorporate specific statistical properties of the inter-event times distribution [39]. Non-mean-field dynamics may be studied by partitioning the nodes in several classes of activation, differentiating the probability of contagion within each class and between different classes.

This work was supported by National Science Foundation under grant No. CMMI-1561134, the Army Research Office under grant No. W911NF-15-1-0267, with Drs. A. Garcia and S.C. Stanton as program managers, and by Compagnia di San Paolo.