**RESEARCH**                                                                                         **Open Access**

# Understanding the romanization spreading on historical interregional networks in Northern Tunisia

Margarita Kostré[1,2*], Vikram Sunkara[1], Christof Schütte[1,2] and Nataša Djurdjevac Conrad[1]

*Correspondence:
kostre@zib.de

[1] Modeling and Simulation of Complex Processes, Zuse Institute Berlin, 14195 Berlin, Germany
[2] Institut für Mathematik und Informatik, Freie Universität Berlin, 14195 Berlin, Germany

## Abstract

Spreading processes are important drivers of change in social systems. To understand the mechanisms of spreading it is fundamental to have information about the underlying contact network and the dynamical parameters of the process. However, in many real-wold examples, this information is not known and needs to be inferred from data. State-of-the-art spreading inference methods have mostly been applied to modern social systems, as they rely on availability of very detailed data. In this paper we study the inference challenges for historical spreading processes, for which only very fragmented information is available. To cope with this problem, we extend existing network models by formulating a model on a mesoscale with temporal spreading rate. Furthermore, we formulate the respective parameter inference problem for the extended model. We apply our approach to the romanization process of Northern Tunisia, a scarce dataset, and study properties of the inferred time-evolving interregional networks. As a result, we show that (1) optimal solutions consist of very different network structures and spreading rate functions; and that (2) these diverse solutions produce very similar spreading patterns. Finally, we discuss how inferred dominant interregional connections are related to available archaeological traces. Historical networks resulting from our approach can help understanding complex processes of cultural change in ancient times.

**Keywords:** Mesoscale spreading process, Network inference, Time-evolving network, Romanization spreading, Scarce data

## Introduction

Spreading dynamics has played a crucial role in shaping humankind since it emerged, as it led to the biggest human discoveries and societal transformations. Understanding the dynamics of spreading processes has thus been a topic of great interest in various scientific fields, such as diffusion of religion (Fousek et al. 2018) in humanities, rumor spreading (Daley and Kendall 1964) in social sciences and virus spreading in epidemiology (Hethcote 2000). Different types of models have been developed in order to study dynamics of spreading on various spatio-temporal scales based on available data-sets (Colizza and Vespignani 2008; Krause et al. 2018; Carrignon et al. 2020; Hethcote 2000).

Kostré *et al. Applied Network Science*     (2022) 7:53

Page 2 of 18

Typical data from real-world spreading processes includes observations of e.g. the number of infected individuals at different time points. More detailed information needed for understanding and predicting the spreading dynamics, such as the underlying contact structure between individuals and dynamical properties of the spreading, e.g. the spreading rate $\alpha$, are usually unknown. For large population numbers, it can be assumed that the considered population is well-mixed. Then, the spreading can be modelled by the so-called compartmental models with deterministic dynamics given by the system of ordinary differential equations (ODEs) (Kermack and McKendrick 1927; Hethcote 2000). These models are shown to be able to reproduce the data coming from real-world observations (Beira and Sebastião 2021; Maier and Brockmann 2020; Wulkow et al. 2020). Using appropriate parameter estimation methods, the spreading dynamics can be easily recovered by solving the ODE system. Due to their simple formulation and cheap computational cost, compartmental ODE models are often a model of choice for various spreading processes. However, these models don't offer detailed information about the dynamics on a resolution of each individual.

Recently, new models have been introduced with a focus on heterogeneous spatio-temporal interactions of individuals with realistic mobility and behavioral patterns. These models assume an underlying *network structure* that determines the possible interactions between individuals (Kiss et al. 2017; Karunakaran et al. 2017; Prasse et al. 2020; Schlosser et al. 2020; Volz 2008; Wu et al. 2020; Barthélemy et al. 2005; Newman 2002).

Additionally, by taking into account the mobility of individuals, e.g. using the airline travel data (Brockmann and Helbing  2013; Pei et al. 2018) or cell-phone data (Schlosser et al. 2020; Wesolowski et al. 2012), the underlying network naturally becomes time-evolving (Koher et al. 2019). A structure of such networks has been shown to have a strong impact on the dynamics of the spreading (Colizza et al. 2006; Brockmann and Helbing  2013).

Developing formal mathematical models for such systems has been a topic of many recent publications (Kiss et al. 2017). For example, ODE equations for compartmental models have been extended to account for the network structure, but the resulting ODE systems can usually not be analytically solved. When a network $G$ is known, then the ODE systems can be solved numerically, however, in many real-world examples these networks are not known and need to be inferred together with other dynamical parameters, e.g. $\alpha$, of the process. The inference of both $\alpha$ and $G$, based on partial observations poses a new parameter inference challenge (Prasse and Van Mieghem 2019; Brugere et al. 2018; Di Lauro et al. 2020; Ma et al. 2019). The difficulty arises due to the fact that different combinations of $\alpha$ and $G$ can produce very similar overall spreading patterns. Hence, the inference method has multiple valid parameter configurations to choose from. In this work, we shed light on the challenges of inferring the network and the dynamical parameters together, using a concrete real-world example.

In this paper, we study the historical process of Romanization spreading in Northern Tunisia during the period 146 BC–350 AD. Romanization is a complex meta-process with different interacting layers, e.g. on a social, political, and economic level. Processes of change on these different levels govern cultural change and are therefore considered to be important indicators of the Romanization. However, very little is known about

these processes and their complex *interaction mechanisms*, which makes modelling of Romanization spreading a challenging task. Here, we will focus on the changes in the civil administration system of Northern Tunisia and thus consider a city status to be the main indicator of the Romanization process. Cities with established roman administrative status will be considered to be romanized. Thus, we will model the Romanization process as an SI epidemic process, where susceptible will refer to non-romanized cities and infected to romanized cities. The spreading parameters and the underlying connections between these cities that govern the SI process are unknown. Observation data on this process was obtained from archaeological evidence about the cities in Northern Tunisia in the period 146 BC–350 AD. However, due to a limited amount of archaeological traces, available data is sparse, incomplete and uncertain. This makes the problem of inferring the unknown connections on a city-level infeasible.

To overcome this, we study the Romanization spreading on a mesoscale, i.e. between different regions in Tunisia. Grouping of cities into regions will have two main advantages compared to modelling on a city-level. Firstly, it will reduce the effect of data sparsity by considering several cities together instead of each city individually. Secondly, it will account for complex spreading mechanisms of cultural transmission, as an interregional model will essentially substitute two-city interactions by interactions between a number of cities from different regions. Aggregation of cities into regions will result in a model reduction, and the new dynamics on a mesoscale will be an SI epidemic process on an interregional network. Our model builds on standard epidemic spreading models on networks (Prasse et al. 2020) and extends these by considering time-evolving networks. The aim of this paper is to infer possible interregional networks underlying the Romanization spreading that could have produced the observed spreading patterns.

The remainder of the paper is organized in following way: In "Background" section, we discuss the historical background of the Romanization process in ancient Tunisia and present the main characteristics of the available dataset. In "Modeling romanization spreading between regions" section, we introduce our model, an SI epidemic model on an interregional network for the romanization process, and the respective inverse problem to infer its model parameters. In "Methods" section , we present the general methodology of the numerical methods used to perform the numerical experiments in the paper. In "Results" section, we collate the multitude of networks and dynamic parameters which exhibit very similar overall spreading behaviour. From these, we shortlist three contrasting networks, and show how these networks induce a similar spreading but over different edges in the networks. In "Validation" section, we cross-validate our inferred model with another historical data source. Lastly, we discuss the significance of our inferred model in understanding the romanization spreading in northern Tunisia, and discuss open problems with respect to the solving of such mixed–network parameters and dynamic parameter–based inverse problems.

## Background

Romanization process of northern Africa started in 146 BC when the region known nowadays as Tunisia was annexed by Rome in the aftermath of the Third Punic War. In the following centuries the African province expanded further with its greatest extend around 117 AD. During this period, cultural exchange, spreading of socio-economic and
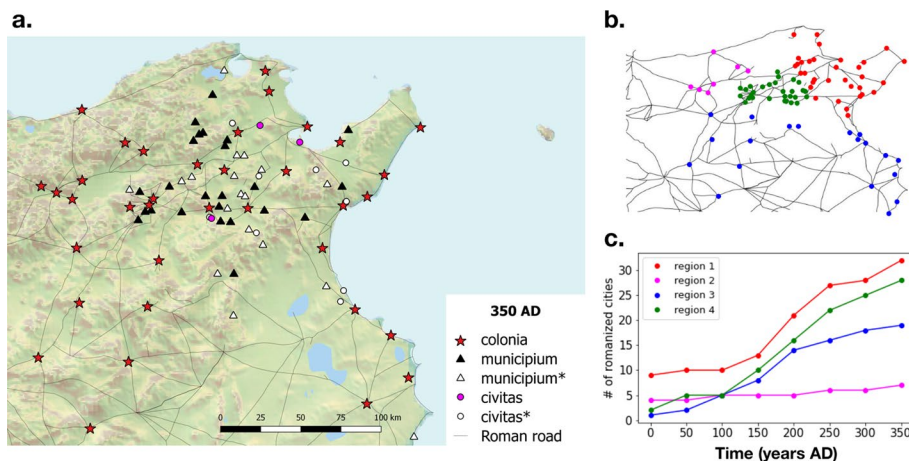
**Fig. 1** Visualization of available data and division of cities into regions. **a** Map of Roman roads and cities with Roman status in Northern Tunisia at 350 AD. City statuses marked with * denote cities with at least this status. (Map was made by Fleur Schweigart) **b** Geographical position of Roman cities and their division into four regions. Color of a city indicates a region it belongs to. **c** Observed number of romanized cities in the period 0–350 AD for each region corresponding to the division from Fig. (**b**). *Sources* Natural Earth; Ancient World Mapping Center (2012); iDAI. geoserver German Archaeological Institute (2022). Licensed under CC BY-NC 3.0. and Open Data Licence

technological innovations led gradually to adaptation to the Roman influence. The most import indicators of the cultural change, can be observed in the Romanization of towns, cities and entire settlement systems on an administrative level. Romans established a settlement hierarchy with a tripartite taxonomy of status: colonia, municipium and civitas. Colonias were cities with the highest ranks and initially these cities were strategically located. Municipia had a lower rank than colonias and their inhabitants did not have all Romans civil rights, but still had a certain amount of autonomy. In comparison to municipia, civitates had a lower rank and they were only semi-autonomous.

During the period 146 BC–350 AD cities in ancient Tunisia gradually got romanized, which was reflected in their administrative status, that often changed mostly to a higher rank. The data-base we are working with is publicly available on the server of the German Archaeological Institute (iDAI. geoserver German Archaeological Institute 2022). This data-base contains the administrative status of 88 cities from 146 BC until 350, AD collected through an extensive literature review and by querying archaeological online research databases (Babelon et al. 1893, iDAI. objects/Arachne; iDAI. gazetteer:Deutsches Archäologisches Institut). For a number of cities, the Roman administrative status is known through historical sources. In our database, preserved Latin documents were used as the only source of city status. In Fig. 1a we plot the romanized cities and mark them according to their status at 350 AD. Status of some cities (marked with *) is assumed, but could not be confirmed by archaeological records, so for these cities we plot only the last confirmed status. This reflects the uncertainty of already sparse data we have at hand. Note that in the following, we will focus only on the notion of a city being romanized or not and not on the particular status.

We will use the romanization data on a city level to uncover the interregional interactions and how these changed over time under Roman influence. Since we are interested in interpretation of the resulting interaction patterns based on archaeological

Kostré *et al. Applied Network Science*      (2022) 7:53

Page 5 of 18

information, we used a spatial division in 4 regions obtained from the expert knowledge, see Fig. 1b. These regions were shaped by features of the landscape, climate, economy, culture and politics. In particular, they cover areas that have different geographical properties, natural resources and socio-economic characteristics. The regions differ in size, as well as in the number of cities contained in the data-set, with a range between 7 and 33 cities per region. Additionally, we aggregate the temporal data, such at that at time point 0 we account for cities that obtained a roman status before 0AD; at time point 50 we aggregate information about cities that got romanized in the period [0, 50)AD and so on. For each region, we plot the number of romanized cities in these discrete time points, see Fig. 1c. We see diverse spreading curves for the four regions. For example, Region 1 is characterized by an almost constant number of romanized cities until 150 AD with a steep increase in the period 150–250 AD. On the contrary, the spreading curve of Region 2 is almost flat, since most of the cities got romanized in the first timeframe. Regions 3 and 4, both have a fast accelerating number of romanized cities. In the next sections, we will introduce an approach for inferring interregional interactions that could have produced spreading curves shown in Fig. 1c. **Remark:** The process of temporal aggregation explained above, was governed by the quality of available data. Namely, most of the considered data points cannot be dated directly with great precision. However, approximate time-spans during which a Roman municipality was inhabited can be extracted from the published literature. The choice of 50 year time intervals is made as the thinnest time slices that are still reasonably well supported by the data.

## Modeling romanization spreading between regions

### Model formulation

We model the Romanization spreading in ancient Tunisia by an SI epidemic model (Allen 1994) on a network of connected regions. In our context, this means that a city is considered to be infected if it has a Roman city status[1]. Likewise, a city is susceptible if it doesn't have a Roman status. Susceptible cities can become romanized through a cultural influence from already romanized cities. On a mesoscale, every region $m$ is characterized by the fixed number of cities inside this region $P_m$ and a regions state at time $t$ given by a pair $(s_m(t), i_m(t))$ denoting the number of susceptible and infected cities in $m$ at time $t \in [0, T]$. Formally, we model the Romanization spreading over a contact network of $N_R$ nodes.

The contact network is defined by

$$G_{m,n} := \begin{cases} \frac{W_{m,n}}{P_m} + \frac{W_{n,m}}{P_n} & \text{for } m \neq n, \\ \frac{W_{m,m}}{P_m} & \text{for } m = n, \end{cases} \tag{1}$$

where $W_{m,n}$ represents the connectivity between regions $m$ and $n$ and $W_{m,n}/P_m$ denotes the contact rate between cities in region $m$ and cities in region $n$. Our definition of the contact network $G$ builds on the idea that the Romanization process is driven by complex mechanisms of cultural influence and assumes symmetric contact patterns

---

[1] Here we do not distinguish which Roman status a city has.

$\frac{W_{m,n}}{P_m} + \frac{W_{n,m}}{P_n}$. Then for every region $m = 1, \ldots, N_R$ a change of its state is given by the following set of ordinary differential equations:

$$\frac{ds_m(t)}{dt} = -s_m(t)\,\alpha(t)\sum_{n=1}^{N_R} G_{m,n}\,i_n(t),$$
$$i_m(t) = P_m - s_m(t), \tag{2}$$

where $\alpha : \mathbb{R} \to \mathbb{R}$ denotes the spreading rate function. This model builds on existing work for SIR epidemic spreading on networks (Prasse et al. 2020; Wang et al. 2018), and extends these approaches by considering spreading rate function $\alpha(t)$ instead of a constant spreading rate. Using the definition of $G$, we can write a more detailed description of our model:

$$\frac{ds_m(t)}{dt} = -s_m(t)\,\alpha(t)\sum_{n=1,n\neq m}^{N_R}\left(\frac{W_{m,n}}{P_m} + \frac{W_{n,m}}{P_n}\right)i_n(t) - s_m(t)\,\alpha(t)\,\frac{W_{m,m}}{P_m}\,i_m(t), \tag{3}$$

$$i_m(t) = P_m - s_m(t). \tag{4}$$

This model was build considering the following types of interactions:

1  The first term in equation (3), namely $s_m(t)\,\alpha(t)\sum_{n=1,n\neq m}^{N_R}\left(\frac{W_{m,n}}{P_m} + \frac{W_{n,m}}{P_n}\right)i_n(t)$, accounts for the change in the number of romanized cities in $m$ caused by interactions with other regions.
2  The second term in equation (3), i.e. $s_m(t)\,\alpha(t)\,\frac{W_{m,m}}{P_m}\,i_m(t)$ models self-infections, i.e. influence that cities within the same region have on each other. Note that often the rate of self-infections are not considered (Wang et al. 2018). Here, like in Prasse et al. (2020), we assume that infections inside regions can have different dynamics and therefore we account for $W_{m,m}$ as well.
3  Finally, the second equation of our model (4) stands for the so-called conservation of population numbers in every region $m$, such that it holds $s_m(t) + i_m(t) = P_m$.

Since analytical solutions of the model given in (3) are not known, we solve the ODE numerically with the first order Runge-Kutta scheme (Alexander 1990). We parametrize the model (3) by the tuple containing the contact network and the spreading rate function, $\sigma := (G, \alpha)$. One can interpret the tuple $(G, \alpha)$ as a time-evolving network $\alpha(t)\,G$, that can be even time-continuous or time-discrete.

We fix our initial value to be $i(0) = (i_m(0))_{m=1}^{N_R}$. Using these, here on in, we denote the approximate solution to (3) for the given parameter set $\sigma$ and initial value $i(0)$ at time $t$ by,

$$\phi(t\,|\,\sigma, i(0)) = \left(P_m - \int_0^t \frac{ds_m(\tau)}{d\tau}\bigg|_{\sigma, P_m - i(0)} d\tau\right)_{m=1,\ldots,N_R}. \tag{5}$$

The function $\phi(\bullet\,|\,\sigma, i(0))$ describes a *spreading curve* through time determined by the parameter set $\sigma$. With this we can formulate the inverse problem, where given a set of

observations we need to infer the possible parameters $\sigma$ which most likely produced the data.

### Inferring model parameters from data

Our main goal is to infer $\sigma$, i.e. the weights of the underlying contact network $G$ and the spreading rate function $\alpha(t)$ given observations of the spreading in the discrete time points of the interval $[0, T]$. For a particular choice of $\sigma$, we can generate the corresponding spreading curves by solving the equation (2). Comparing these results to the given data via a distance measure, we obtain an error between the two results. Naturally, the parameter set which induces the smallest error to the data, is considered a viable candidate for a good fit. This idea is at the core of the so-called General Inverse Infection Model (Bóta and Gardner 2017), which in this way translates the inference problem to an optimization problem of finding the parameter set that minimizes the error between the corresponding spreading curves and the available data. We follow this methodology and adapt it to using an error function that takes the variance of data into account. This step is needed to due the nature of our data and model, as the number of cities in different regions varies strongly. Otherwise the method would tend to fit regions with larger populations more, than regions with smaller populations.

More formally, we formulate our parameter inference problem as follows: Let $N_T \in \mathbb{N}$ be the total number of time points at which observation were made. We denote $t_i$, for $i \in [0, \ldots, N_T]$, to be the *ith* time point and the vector $(\omega_{m,t_i})_{m=1}^{N_R} \in \mathbb{Z}_+^{N_R}$ to be the observed number of romanized cities in the *mth* region during the *ith* time interval. in the $N_R$ regions at this time, respectively. Considering all the data points, we define the loss function by Wulkow et al. (2020)

$$\ell(\sigma) := \sum_{i=1}^{N_T} \sum_{m=1}^{N_R} \frac{(\omega_{m,t_i} - \phi(t_i|\sigma, \omega_{\bullet,0}))^2}{C_{m,t_i}^2}, \tag{6}$$

where $C_{m,t_i} := \max\{\omega_{m,t_i}, STD(\omega_{m,\bullet})\}$ is the maximum between the data and the standard deviation for region $m$ observed in time. The loss function in (6) consists of two sums, which summing over the regions and the data-points. For each region $m$ and data-point $t_i$, we calculate the square distance between the data, $\omega_{m,t_i}$, and the predicted spreading time $t_i$ w.r.t. the parameter, $\phi(t_i|\sigma, \omega_{\bullet,0})$, weighted by $C_{m,t_i}$, as used in Wulkow et al. (2020). If the data $w_{m,t_i}$ is larger than the standard deviation, than the inner fraction in (6) is simply the relative distance. Otherwise the inner fraction is smaller for regions with a faster spreading.

Then, the constrained minimisation problem reads

$$\min_{\sigma=(G,\alpha)} \ell(\sigma),$$

subject to:

$$\forall\, m \neq n, \ 0 \leq G_{m,n} \leq 2, \text{ and } 0 \leq G_{m,m} \leq 1$$
$$\forall\, t \geq 0, \ 0 \leq \alpha(t) < \infty, \tag{7}$$

which we use to find candidate contact networks and spreading functions. In the particular example considered in this paper, we have four regions of interest, i.e. $N_R = 4$. We fit eight observations from 0 AD and 350 AD, with equidistant sampling of 50 years. Given we do not have a resolution below the 50 year period, we choose our spreading rate function to be a piece-wise constant function, with seven equidistant left clopen intervals between [0, 350]. We assume the spreading rate function has a range between [0, 0.1]. We only infer $G$, as only these terms are needed to solve for the spreading (2). Given $G$ is symmetric and the spreading rate function is piecewise constant function, in total, we have 17 unknown parameters which we minimize over in (7). In the context of this work we refer to problem (7) as a minimization problem or a parameter inference problem interchangeably.

## Methods

The parameter inference problem (7) is inherently ill-posed, that is, we have multiple parameter candidates which minimize the loss function of observing the data. Hence, we used the fast minimisation method prescaled Metropolis-adjusted Langevin algorithm (PMALA) with thousands random starts and collected the local minima. The inferred parameter tuple $(G, \alpha)$ characterises a time evolving network and the spreading on top of it, which we find eludes to give a simple visual interpretation. Therefore, to visualise what the landscape around the minima looks like, we projected the data using t-SNE method, with the distance function which discriminated parameters based on the spreading curves (sol. to (5)) they generated. This then gave an embedding where proximity between embedded parameter points produced similar trajectories in time. We give details of these steps below.

### Finding local minima with PMALA

The minimization problem (7) has multiple local minima. Common optimization methods, like, the stochastic gradient descent (Ketkar 2017), follow the gradient of the landscape. This makes the methods fast, but they tend to get stuck in a local minima. To avoid this issue, it is recommended to run the algorithm from different starting points or to change the learning rate during the iterations (Klein et al. 2009). Alternatively one can use Bayesian methods, which are appealing in their ability to capture uncertainty in parameters and avoid overfitting (Welling and Teh 2011).

In a Bayesian method we define the probability of parameters conditioned on the data, the so-called posterior distribution, and then sample from this distribution, to find the most likely parameters. This can be done efficiently with a controlled random walker exploring the parameter space, like the Metropolis-Hastings algorithm. The class of Metropolis-Hastings algorithms (Robert et al. 1999), is based on Markov chain properties. A Markov chain is constructed by defining a conditional density and having a property to converge to the desired distribution. The choice of the conditional density, that determines the rules of the walker, is crucial for the performance of the algorithm.

The Metropolis-adjusted Langevin algorithm (MALA) replaces the random walker with a walker following the Langevin diffusion, to make him more efficient (Kroese and Rubinstein 2012). Especially in high-dimensional spaces, the appropriate step-size can
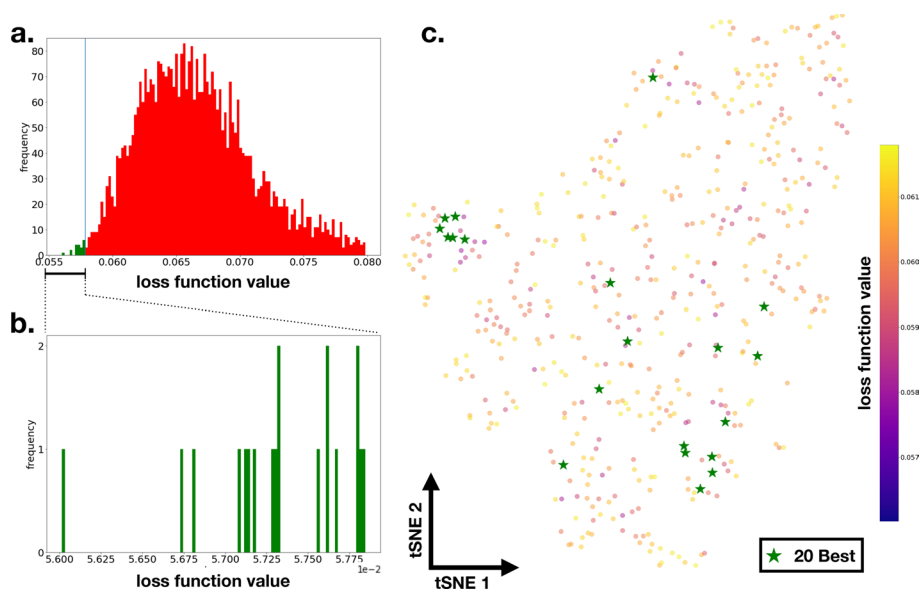
Kostré *et al. Applied Network Science*     (2022) 7:53

Page 9 of 18



**Fig. 2** Distribution of the inferred parameters. **a** Histogram of the inferred parameters with respect to the loss function. Shaded green shows the best 20 inferred parameters. **b** Zoomed axes of the histogram showing only the density of the best 20 inferred parameters. **c** t-SNE embedding of the top 500 inferred parameters, with the loss function colour overlay. The best 20 inferred parameters are marked with green stars

vary a lot in different dimensions. This makes an appropriate choice of the step-size difficult.

In light of this, the prescaled Metropolis-adjusted Langevin algorithm (PMALA) is a method that was developed for parameter estimation with parameters arising in different orders of magnitude. The scalar step-size of the walker in MALA is replaced by a diagonal matrix, which is prescribed beforehand, and is dependent on the average length of the gradient and the average length of a step in each of the parameter directions $\tau$. In this work, we only refer to how PMALA was used in the context of our minimization problem, for technical aspects we refer to Wulkow et al. (2021).

The advantage of PMALA is that it can find quickly many local minima even for problems with strong scale separation between parameters. Therefore, we treat the samples of each run of the PMALA algorithm as individual local minima of the loss function. We don't use any statistics of the samples, but simply utilize the larger variety of solutions found in comparison to other gradient descent approaches.

In the PMALA method has two free parameters which need to be tuned: the average length of a step in each parameter-direction $\tau$ and the number of iterations $n_{PMALA}$. We chose the parameter values to be $\tau = 0.03$ and $n_{PMALA} = 2000$. Given there are multiple minina, we chose 5000 random starts in the valid parameter space and ran the minimisation for 2000 steps.

### Projecting into low dimensional space

Each PMALA run gives a parameter, $\sigma$, which consist of a contact network $G$ and the spreading rate function $\alpha$, which together form a time evolving network. We picked the best 20 parameters w.r.t. the values of the loss function, see Fig. 2 b). However, it does

not necessarily imply that these 20 solutions exhibit similar contact network and spreading rate function. Given any two parameters, $\sigma_1$ and $\sigma_2$, it is difficult to measure the similarity between $\sigma_1$ and $\sigma_2$ using a simple distance function which captures similarity in both the contact networks and the spreading rate function. Hence, we opted to use the distance between the predicted spreading curves of the parameters as the distance function, that is,

$$d(\sigma_1, \sigma_2) := \int_0^T \left( \phi(\tau \mid \sigma_1, \bullet) - \phi(\tau \mid \sigma_2, \bullet) \right)^2 d\tau. \tag{8}$$

This distance function considers all points in between the observed time periods and should not to be confused by the loss function, which is a function of the data. Here the distance function is only a function of the difference in the respective spreading curves.

The remaining task is to project the solutions onto the two dimensional-space with respect to the distance in (5), such that solutions having a similar spreading curve, are also close in the two-dimensional space. For this purpose, we choose t-Distributed Stochastic Neighbor Embedding (t-SNE) (Zhou et al. 2018; Belkina et al. 2019). T-SNE preserves the local structure, by minimizing the Kullback-Leibler divergence between the joint probabilities of the low-dimensional embedding and the high-dimensional data. It is the projection algorithm with minimal structural information loss (Zhou et al. 2018). An important parameter for the projection is the so called perplexity, which describes how to balance the attention between local and global features of the data. Perplexity needs to be choosen depending on the structure of the data-points and their amount. If the perplexity is too small, then little variations of the data dominate the projection. For example, if we project points that are building a circle with a too low perplexity, then the circle looks twisted. On the other side if the perplexity is too large, then the points look randomly distributed. In this work, we were interested in the proximity of the best 20 inferred parameters, hence, we choose a perplexity of 30, to study them.

## Results

We recall that, based on expert knowledge, we divided the roman cities from our dataset into four regions ("Background" section), such that each represents a geographical region in Northern Tunisia, see Fig. 1. Available data spans the time period from 146 BC until 350 AD. We use the temporal information before 0 AD for initial conditions $i(0)$ and the rest we divide into seven time-frames with equal lengths of 50 years. We assume within each time-frame, a constant spreading rate, ergo the spreading rate function, $\alpha(t)$, to be piece-wise constant. Hence, we numerically solved the parameter inference problem (7), for the 17 parameters, which were needed to fit our adapted Romanization SI model (5), to our data (Fig. 1c). To avoid extensive numerical computations when solving the ODE, we scale the time period from [0, 350] to [0, 3.5]. This scaling only effects the range of the spreading function $\alpha(t)$.

To find the parameters $\sigma = (G, \alpha)$ which best fit the data in our parameter inference problem we used the PMALA numerical scheme (4). We ran the algorithm from 5000 different starting parameter configurations, to explore the parameter space and avoid local minima wells. We iterated the sampler over 2000 steps; we found this amount to be sufficient, as the average difference between the last two steps of the samplers were

of the order $10^{-3}$, and did not decrease with increased number of steps. We first investigated the distribution of the loss function of the 5000 parameter candidates given by PMALA, and found that the distribution had a long tail. We found that approximately 1000 starting points had a loss function strictly above 0.08. The remaining 4000 parameters, which yielded a loss function below 0.08, had the shape of a *Log-normal* distribution (see Fig. 2a). We did not observe multiple modalities in the distribution of the loss function, suggesting the landscape does not have multiple steep wells where the PMALA samplers would visit for a long period of time.

We wanted to investigate if the parameters found through PMALA came from a flat region or from a single well, hence, we studied if the parameters produced similar spreading curves. For this, we took the 500 best parameters with respect to the loss function and embedded them using the t-SNE projection with the spreading based distance measure (8) (see Fig. 2b–c). We observed very little visual correlation between a point's loss function and that of its neighbouring points. This suggests, that the parameters may not belong to a single minima well. Furthermore, when we overlayed the best 20 parameters, that is, the parameters yielding the smallest loss function, we observed them to be scattered over the t-SNE embedding (see Fig. 2c green star). Given that the t-SNE projection strongly preserves local distances and that we used a spreading based distance measure, implies that the 20 best parameters mostly induced distinctly different spreading curves.

By manual inspection of the 20 solutions, see Additional file 1: Fig. A, we observe that, according to their network structure and the shape of their spreading rate functions, the solutions can be divided into 3 main groups. From these groups we pick one representative solution each and analyze further properties of their contact networks *A*, *B* and *C* together with their corresponding spreading rate functions. We plot these in the top two plots of Fig. 3. Network edges are colored according to their weights in *G*, where light colors indicate small weights and dark red colors high weight. We observe that network patterns and shapes of spreading rate functions differ greatly between the three solutions. Solution A is characterized by high spreading rates and low, similarly distributed edge weights in Network A. This means that no specific interregional patterns govern the spreading and that the dynamics is driven by the high spreading rate. On the contrary, in solutions B and C, network structure has a bigger effect on the spreading and the changes in alpha become prominent only in the last time-frame accounting for the abrupt increase of romanized cities in region 1 and 4. Thus, we see that networks of different structure and distinct spreading rate functions lead to similar solutions wrt. spreading dynamics. We study this effect in more details by looking at the actual spreading curves. Namely, using the three networks *A*, *B*, *C* and the corresponding $\alpha(t)$ values, their spreading curves were obtained by solving equation (2), see the bottom plot of Fig. 3. We observe a good fit between the spreading curves and the available data, as the three solutions were chosen from the best 20 solutions and have the error bounded by 0.058. Furthermore, we see again that spreading curves with very different shapes, i.e. coming from distinct dynamics, can produce a good fit to available data.

Each solution $(G, \alpha)$ can be interpreted as a time-evolving network $(\alpha(0)G, \alpha(50)G, \ldots, \alpha(350)G)$. This network contains information on the interregional connectivity and the spreading rate, but not the information on the romanization flux
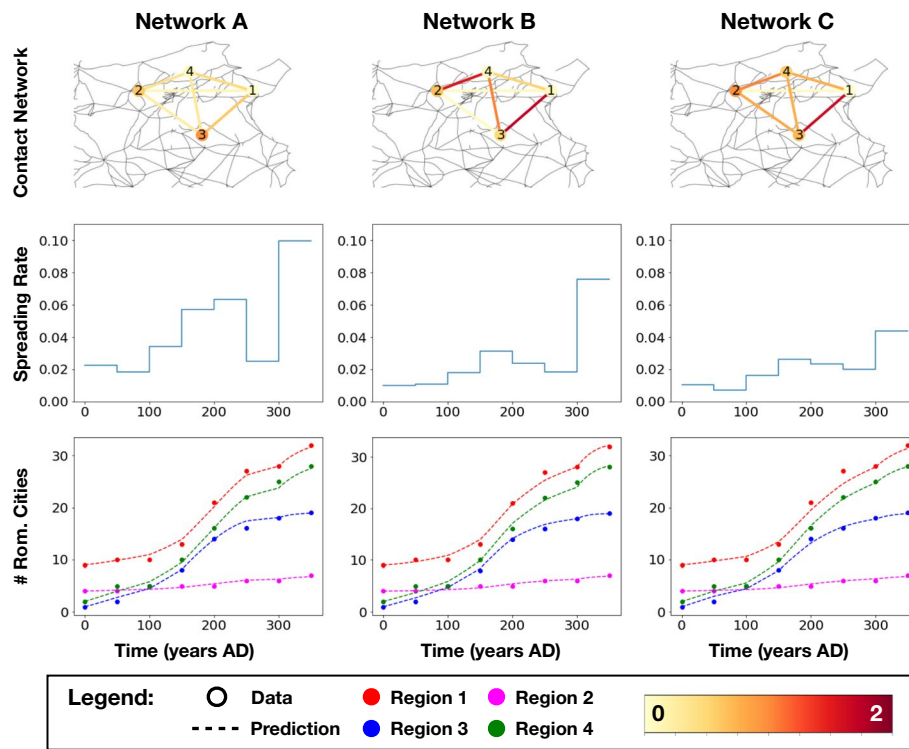
**Fig. 3** Representative solutions of the inference problem. Top plot: the contact networks A, B and C. Middle plot: corresponding spreading rate functions. Bottom plot: the resulting spreading curves w.r.t. the available data-points. Network edges are colored according to their weights, s.t. light colors indicate small weights and dark red colors stand for high edge weights. *Sources* Ancient World Mapping Center (2012); iDAI. geoserver German Archaeological Institute (2022). Licensed under CC BY-NC 3.0. and Open Data Licence

between regions that is governed by the number of romanized cities. To this end, we introduce a time-evolving, directed network

$$H_{m,n}(t) := \alpha(t)\, s_n(t)\, G_{m,n}\, i_m(t),  \tag{9}$$

that we will call the Romanization dynamics network.

Temporal evolution of the effective current between regions $m$ and $n$ stands for the amount of romanization influence and can be expressed by

$$H^*_{m,n}(t) := \max(H_{m,n}(t) - H_{n,m}(t), 0).$$

In Fig. 4 we plot the temporal evolution of $H^*(t)$ for the three selected networks A, B and C. Network edges are colored according to the value of the effective current $H^*(t)$ and the arrows point in the direction of positive current. Resulting effective currents differ for networks A, B and C, due to the varieties in their structural properties and $\alpha$. But, the main patterns are common in all three networks. Namely, we observe that in the beginning for all networks there is a high current from Region 1 to Region 3. Additionally, for networks A and C we see high $H^*$ from Region 1 to Region 4. Both observations agree with historical narratives about the political and economical importance of Region 1. Namely, a city Carthage that belonged to Region 1 was one of the largest cities of the Roman Empire, that certainly had a major influence on the cities close by. Also, Carthage
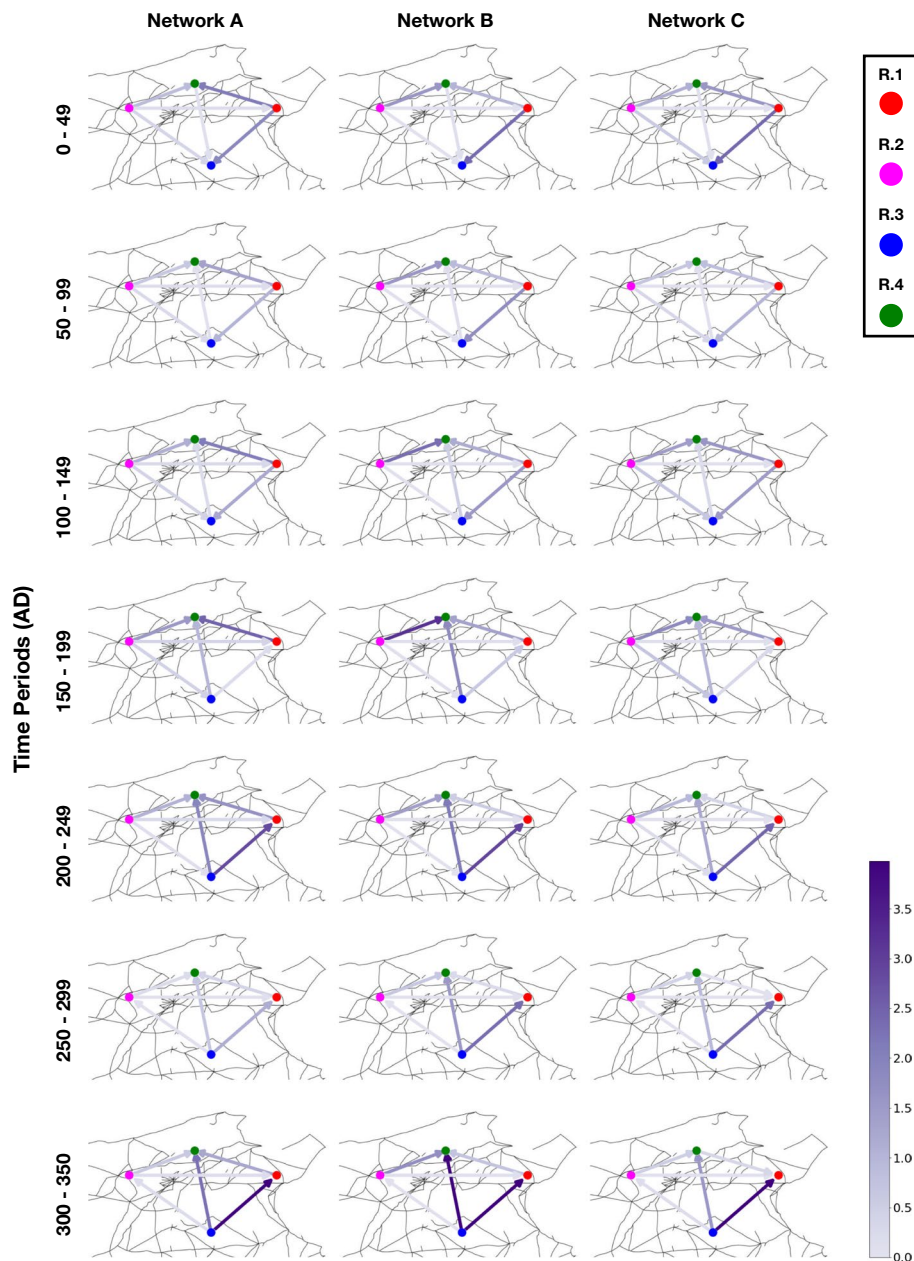
**Fig. 4** Temporal change of the effective current $H^*(t)$. Network edges are colored according to the value of the effective current $H^*(t)$ and the arrows point in the direction of positive current. Nodes are colored according to the original region color distribution. *Sources* Ancient World Mapping Center (2012); iDAI. geoserver German Archaeological Institute (2022). Licensed under CC BY-NC 3.0. and Open Data Licence

was the leading trade center and one of the major breadbaskets of the Empire. As such, it had a strong influence on other regions of Ancient Tunisia. From 150 AD, the effective current from Region 1 to Region 3 started strongly decaying. Furthermore, as most of the cities in Region 3 got romanized by 200 AD, the current from Region 3 to both Region 1 and Region 4 increases significantly, reaching its highest values before 350 AD. The early romanizaion of most cities in Region 2 determines the directionality of the current towards other regions. Additionally, it has an effect to other regions, e.g. a high

Kostré *et al. Applied Network Science* (2022) 7:53

Page 14 of 18
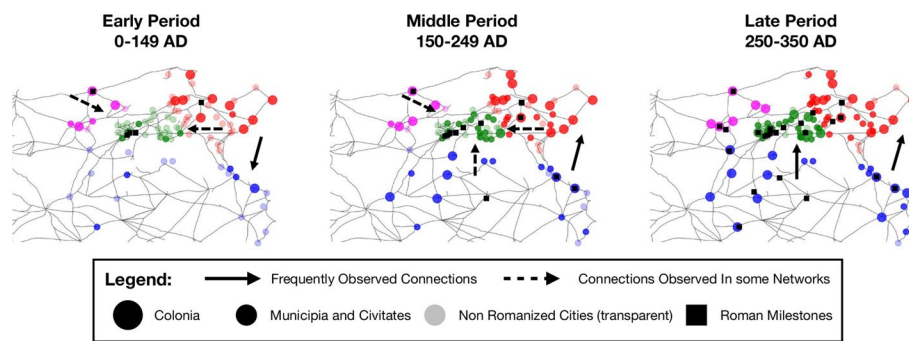


**Fig. 5** Validation data and dominant connections of *H**. Milestones are marked with black squares and cities with circles (large circles stand for colonias, small circles for municipia and civitates). Cities are colored according to the region they belongs to, were transparent colors indicate that a city has not yet been romanized at that time. Full arrows show the dominant connections of effective current *H** that all three solutions A, B, C have in common. Dashed arrows indicate dominant connections that only some of the solutions A, B, C have. *Sources* Ancient World Mapping Center (2012); iDAI. geoserver German Archaeological Institute (2022). Licensed under CC BY-NC 3.0. and Open Data Licence

current from Region 2 to Region 4 in network B. Commonalities from the three solutions discussed above, result in several dominant interregional connections that we will compare to archaeological traces as a method of validation, see "Validation" section.

## Validation

One of the major inventions of the Roman Empire were Roman roads. As a crucial part of the Empire infrastructure, Roman roads were excessively used for the movement of armies and civilians, but also for trading goods and communications between Roman cities. In particular, Roman roads were important connectors of Colonias, many of which were influential regional hubs of trade, politics and culture. Even now, Roman roads are being intensively studied as they are closely connected to modern-day infrastructures and economies (Flückiger et al. 2021). Thus, Roman roads can be considered to be carriers of cultural, political and economical influence.

We will use information about the Roman road network in ancient Tunisia to validate our results. The most reliable archaeological traces about the temporal evolution of the Roman road network that we found are collected in the data-set of milestones (iDAI. geoserver German Archaeological Institute 2022). Milestones are stone markers placed along Roman roads (Wilmanns 1881). Their inscriptions often indicated the distance to cities or other places, time reference to the road construction (or repair work) or commemoration to the ruling emperor. The data-set we use consists of 29 milestones for which we have information about their geographical position and time when they were (most likely) placed. We consider these milestones to be indicators for existence of important roads at a particular point in time. Specifically, as a mean of validation, we will compare the resemblance of milestones' position in space and time with the interregional connections with the highest values of the effective current *H**.

In Fig. 5, we plot the Roman road network, milestones and romanized cities aggregated in three time-frames 0–150 AD, 150–250 AD and 250–350 AD. Milestones are marked as black squares and cities are colored according to the region they belong to, as described in "Background" section, where large circles indicate Colonias and small circles stand for

cities of other roman status. Additionally, we plot dominant interregional connections as directed edges with the highest values of *H\**. We distinguish between interregional connections that all three networks have in common (marked with full arrows) and the ones that appear only some networks (marked with dashed arrows).

In the early period from 0–150 AD we observe a strong connection from region 1 in the east to a less romanized region 3 in the south. As discussed above, already at this time region 1 had many colonias many of which were important regional hubs. Similarly strong connections occur from region 1 to the central region 4, that we see in solution A. Another region with many colonias is the region 2 in the west for which our model infers a high effective current to the central region 4 for network B. Several milestones in the north-west part could explain the connection between the two regions. Whether the central region 4 was influenced in the beginning more from the east (region 1) or the west (region 2), should be further studied by archaeologist. In the next period from 150–250 AD, we observe strong romanization of the region 3, expressed by the appearance of many Colonias and the first milestones in this region. This could explain the high effective current we obtain from region 3 to region 1. In the central region 4 we find many new milestones. However, the main source of influence on region 4 in this period cannot be clearly identified, as it could be any of the other three regions. Finally, in the last step, our results show strong connections from region 3 to regions 1 and 4. Possible explanation is that by that time region 3 consisted of many important colonias, whereas in regions 1 and 4 from 250 AD we see many newly romanized cities and many milestones, indicating increased romanization of that area. More detailed evaluation of obtained results is needed to derive conclusions in archaeological context, which exceeds the focus of this work.

## Discussion

In this paper we inferred time-evolving networks from historical data. A problem when working with historical data is that, the data is scarce and has high uncertainty. Nevertheless, historical processes range over long time-scales and are governed by complex mechanisms, requiring a model that is able to capture such spatial and temporal attributes of the process.

To study the Romanization process in northern Tunisia, we introduced a mesoscale SI model with an underlying contact network and a temporal spreading rate , where the cities are divided into regions based on their features and location. Naturally, different choices of regions would lead to different contact networks and spreading rate functions. Then, using our expert curated Romanization data, we were able to infer several contact networks and the Romanization spreading rates during the period of 0–350 AD. In particular, we used the PMALA algorithm to solve the parameter inference problem, that is to find contact networks and spreading rate functions which reproduce the data. When we studied the best 20 solutions, we observed that the solutions could be divided into three main groups according to their network structure and the shape of their spreading rate functions. From these groups we picked one representative solution each and observed that optimal solutions consist of very different network structures and spreading rate functions, and furthermore, that these

diverse solutions produce very similar spreading patterns. Despite the limited amount of historical data, we were able to obtain partial information on the possible important interregional connections. We cross-validated these connections with the available archaeological traces. Our insights can now be further investigated and guide historical studies on the processes of cultural change during the ancient times.

Ideally, more data would lead to stronger inference and insights, however, typical information in historical research is hard to obtain and arduous. Hence, our approach is suitable and appropriate for the context of historical data, due to the fact, that despite the data being scarce, we still obtained plausible historical interregional connections. We believe that this way of approaching the study of historical questions can be easily generalized and applied to different historical examples.

One bottleneck of our approach is that the best found parameters had to be compared manually. Naturally, this is only feasible for small number of best parameters, like 20 in our case. To analyse larger solution sets, we would need a meaningful similarity measure which would take into account differences in both the contact network and the spreading parameters. With the right distance measure, we could cluster the parameter set to study only the distinct parameter configurations. This will be the focus of future research.

### Abbreviations
| | |
|---|---|
| BC | Before Christ |
| AD | Anno Domini |
| ODE | Ordinary differential equations |
| PMALA | Prescaled Metropolis-adjusted Langevin algorithm |
| MALA | Metropolis-adjusted Langevin algorithm |
| t-SNE | t-Distributed stochastic neighbor embedding |
| AWMC | Ancient World Mapping Center. |

## Supplementary Information
The online version contains supplementary material available at https://doi.org/10.1007/s41109-022-00492-w.

> **Additional file 1.** SupplementaryMaterial Figure a.

## Declarations

### Ethics approval and consent to participate
Not applicable.

## References

Alexander R (1990) Solving ordinary differential equations i: Nonstiff problems (e. hairer, sp norsett, and g. wanner). Siam Rev 32(3):485

Allen LJ (1994) Some discrete-time SI, SIR, and SIS epidemic models. Math Biosci 124(1):83–105

Ancient World Mapping Center Available at: http://awmc.unc.edu/wordpress/map-files/ (2012)

Babelon E, Cagnat R, Reinach S (1893) Atlas archéologique de la tunisie: édition speciale des cartes typographiques

Barthélemy M, Barrat A, Pastor-Satorras R, Vespignani A (2005) Dynamical patterns of epidemic outbreaks in complex heterogeneous networks. J Theor Biol 235(2):275–288

Beira MJ, Sebastião PJ (2021) A differential equations model-fitting analysis of covid-19 epidemiological data to explain multi-wave dynamics. Sci Rep 11(1):16312. https://doi.org/10.1038/s41598-021-95494-6

Belkina AC, Ciccolella CO, Anno R, Halpert R, Spidlen J, Snyder-Cappione JE (2019) Automated optimized parameters for T-distributed stochastic neighbor embedding improve visualization and analysis of large datasets. Nat Commun 10(1):1–12

Bóta A, Gardner L (2017) A generalized framework for the estimation of edge infection probabilities. arXiv preprint arXiv: 1706.07532

Brockmann D, Helbing D (2013) The hidden geometry of complex, network-driven contagion phenomena. Science 342(6164):1337–1342. https://doi.org/10.1126/science.1245200

Brugere I, Gallagher B, Berger-Wolf TY (2018) Network structure inference, a survey: motivations, methods, and applications. ACM Comput Surv (CSUR) 51(2):1–39

Carrignon S, Brughmans T, Romanowska I (2020) Tableware trade in the Roman East: exploring cultural and economic transmission with agent-based modelling and approximate bayesian computation. PloS One 15(11):0240414

Colizza V, Barrat A, Barthélemy M, Vespignani A (2006) The role of the airline transportation network in the prediction and predictability of global epidemics. Proc Natl Acad Sci 103(7):2015–2020

Colizza V, Vespignani A (2008) Epidemic modeling in metapopulation systems with heterogeneous coupling pattern: Theory and simulations. J Theor Biol 251(3):450–467

Daley DJ, Kendall DG (1964) Epidemics and rumours. Nature 204(4963):1118–1118

Di Lauro F, Croix J-C, Dashti M, Berthouze L, Kiss I (2020) Network inference from population-level observation of epidemics. Sci Rep 10(1):1–14

Flückiger M, Hornung E, Larch M, Ludwig M, Mees A (2021) Roman transport network connectivity and economic integration. Rev Econ Stud 89(2):774–810. https://doi.org/10.1093/restud/rdab036

Fousek J, Kaše V, Mertel A, Vỳtvarová E, Chalupa A (2018) Spatial constraints on the diffusion of religious innovations: the case of early Christianity in the Roman Empire. PloS One 13(12):0208744

Hethcote HW (2000) The mathematics of infectious diseases. SIAM Rev 42(4):599–653

iDAI. gazetteer:Deutsches Archäologisches Institut Available at: http://www.gazetteer.dainst.org. Last Accessd 2022

iDAI. geoserver German Archaeological Institute. In the process of publishing, will be available at: https://geoserver. dainst.org/maps/6002

iDAI. objects/Arachne Available at: http://www.arachne.dainst.org. Last Accessd 2022

Karunakaran RK, Manuel S, Satheesh EN (2017) Spreading information in complex networks: an overview and some modified methods. IntechOpen, London

Kermack WO, McKendrick AG (1927) A contribution to the mathematical theory of epidemics. Proc R Soc Lond Ser A Contain Papers Math Phys Character 115(772):700–721

Ketkar N (2017) Stochastic gradient descent. Deep learn Python, 113–132

Kiss IZ, Miller JC, Simon PL (2017) Mathematics of epidemics on networks: From exact to approximate models. Interdisciplinary Applied Mathematics

Klein S, Pluim JP, Staring M, Viergever MA (2009) Adaptive stochastic gradient descent optimisation for image registration. Int J Comput Vis 81(3):227–239

Koher A, Lentz HHK, Gleeson JP, Hövel P (2019) Contact-based model for epidemic spreading on temporal networks. Phys Rev X 9:031017. https://doi.org/10.1103/PhysRevX.9.031017

Krause AL, Kurowski L, Yawar K, Van Gorder RA (2018) Stochastic epidemic metapopulation models on networks: SIS dynamics and control strategies. J Theor Biol 449:35–52

Kroese DP, Rubinstein RY (2012) Monte carlo methods. Wiley Interdiscip Rev Comput Stat 4(1):48–58

Maier BF, Brockmann D (2020) Effective containment explains subexponential growth in recent confirmed covid-19 cases in china. Science 368(6492):742–746. https://doi.org/10.1126/science.abb4557

Ma L, Liu Q, Van Mieghem P (2019) Inferring network properties based on the epidemic prevalence. Appl Netw Sci 4(1):1–13

Natural Earth Available at: http://www.naturalearthdata.com. Last Accessd 2022

Newman ME (2002) Spread of epidemic disease on networks. Phys Rev E 66(1):016128

Pei S, Kandula S, Yang W, Shaman J (2018) Forecasting the spatial transmission of influenza in the United States. Proce Natl Acad Sci 115(11):2752–2757

Prasse B, Van Mieghem P (2019) Exact network reconstruction from complete SIS nodal state infection information seems infeasible. IEEE Trans Netw Sci Eng 6(4):748–759. https://doi.org/10.1109/TNSE.2018.2872511

Prasse B, Achterberg MA, Ma L, Van Mieghem P (2020) Network-inference-based prediction of the COVID-19 epidemic outbreak in the Chinese province Hubei. Appl Netw Sci 5(1):1–11

Robert CP, Casella G, Casella G (1999) Monte Carlo statistical methods, vol 2. Springer, New York

Schlosser F, Maier BF, Jack O, Hinrichs D, Zachariae A, Brockmann D (2020) Covid-19 lockdown induces disease-mitigating structural changes in mobility networks. Proc Natl Acad Sci 117(52):32883–32890. https://doi.org/10.1073/pnas.2012326117

Volz E (2008) Sir dynamics in random networks with heterogeneous connectivity. J Math Biol 56(3):293–310

Wang J, Wang X, Wu J (2018) Inferring metapopulation propagation network for intra-city epidemic control and prevention. In: Proceedings of the 24th ACM SIGKDD international conference on knowledge discovery & data mining, pp. 830–838

Welling M, Teh YW (2011) Bayesian learning via stochastic gradient Langevin dynamics. In: Proceedings of the 28th international conference on machine learning (ICML-11), pp 681–688 . Citeseer

Wesolowski A, Eagle N, Tatem AJ, Smith DL, Noor AM, Snow RW, Buckee CO (2012) Quantifying the impact of human mobility on malaria. Science 338(6104):267–270

Wilmanns G (1881) Inscriptiones Africae latinae. Corpus inscriptionum latinarum, vol 3. Georg Reimer, Berlin

Wulkow H, Conrad T, Conrad ND, Mueller SA, Nagel K, Schuette C (2020) Prediction of covid-19 spreading and optimal coordination of counter-measures: From microscopic to macroscopic models to pareto fronts. medRxiv . https://doi.org/10.1101/2020.12.01.20241885. https://www.medrxiv.org/content/early/2020/12/03/2020.12.01.20241885.full.pdf

Wulkow N, Telgmann R, Hungenberg KD, Schütte C, Wulkow M (2021) Deterministic and stochastic parameter estimation for polymer reaction kinetics i: theory and simple examples. Macromol Theory Simul 30:2100017

Wu B, Du J, Wang L (2020) Bridging the gap between opinion dynamics and evolutionary game theory: Some equivalence results. In: 2020 39th Chinese Control Conference (CCC), pp 6707–6714. IEEE

Zhou H, Wang F, Tao P (2018) t-distributed stochastic neighbor embedding method with the least information loss for macromolecular simulations. J Chem Theory Comput 14(11):5499–5510

## Publisher's Note