

RESEARCH

Open Access



# Dynamic centrality measures for cattle trade networks

Patrick Hoscheit<sup>\*</sup> , Éric Anthony and Elisabeta Vergu

<sup>\*</sup>Correspondence:  
patrick.hoscheit@inrae.fr  
INRAE, MaIAGE,  
Université Paris-Saclay,  
78350 Jouy-en-Josas, France

## Abstract

We study network centrality measures that take into account the specific structure of networks with time-stamped edges. In particular, we explore how such measures can be used to identify nodes most relevant for the spread of epidemics on directed, temporal contact networks. We present a percolation study on the French cattle trade network, proving that time-aware centrality measures such as the TempoRank significantly outperform measures defined on the static network. In order to make TempoRank amenable to large-scale networks, we show how it can be efficiently computed through direct simulation of time-respecting random walks.

**Keywords:** Temporal networks, Centrality, Cattle trade, Epidemic mitigation

## Introduction

When dealing with epidemic processes spreading along the edges of networks, centrality measures can be used to identify the nodes most important for disease transmission. Those nodes can then be targeted with specific measures (vaccination, diagnostic focus or reduction in contact) for the purposes of outbreak prevention or mitigation, with an expected favourable impact on epidemic outcomes. The study of centrality measures on large networks therefore occupies a central place in network theory, with hundreds of measures defined for general or specific purposes (Jalili et al. 2015; Lü et al. 2016). Such measures have been extended to networks with additional structure, such as modularity (Ghalmane et al. 2019) or multiple interconnected layers (so-called *mutiplex* or *multi-layer* networks) (Taylor et al. 2019; Lv et al. 2019; Pedroche et al. 2016; Agryzkov et al. 2019).

Networks with time-stamped edges (so-called *temporal* or *dynamic* networks) have an additional level of heterogeneity which makes the study of their role as underlying structure for epidemic spread more challenging. Indeed, infections can only spread by following time-respecting paths, that is composed of edges with chronologically increasing time-stamps. This renders the use of statically computed centrality measures unable to provide useful insights for the control of outbreaks on such networks. Several centrality measures taking this specific structure into account have been defined in past years, mostly by generalizing well-studied static quantities (Kim and Anderson 2012; Hanke and Foraita 2017; Taylor et al. 2017; Bajardi et al. 2012). However, the additional

complexity of temporal networks is reflected by the high computational cost of some of these methods, many of which are prohibitively expensive for networks exceeding thousands of nodes, or more than a hundred time-slices.

In this work, our main contribution is to study SIR-type outbreaks on the French cattle trade network, in which nodes are cattle holdings, and time-stamped edges correspond to the commercial movement of animals between these holdings (Dutta et al. 2018; Natale et al. 2011). We focus on movements in 2014 and 2015, which gives rise to networks with over 170,000 vertices and over 2,250,000 time-stamped edges, with a daily granularity. Besides classical static measures, we will focus on the TempoRank measure, as defined in Rocha and Masuda (2014). This is an extension of PageRank centrality, defined using time-respecting random walks on the network. We will recall some of its properties and present a deterministic (exact) implementation. We will also present an extension of the TempoRank which might be more efficient in the context of epidemic prevention, by identifying vertices that have high downstream transmission potential. However, the matrix-based methods on which this implementation is based fail in the context of large networks, because of the absence of sparsity. We show how direct simulation can instead be used to approximate TempoRank computation, which makes this measure useful even for very large temporal networks. We show that our stochastic method is able to compute TempoRank centrality measures on the large BDNI network, and, using a numerical percolation study, that they outperform static centrality measures in mitigating outbreak sizes.

The paper is structured as follows: in the “[Centrality measures for temporal networks](#)” section, we will recall some of the definitions of centrality measures used and show how some of them are related to the properties of random walks on the network. Then, in the “[Random walk centralities: definition and properties](#)” section, we will present two versions of the TempoRank measure in the same framework, as well as a stochastic approximation algorithm that enables their computation even for large network sizes or large numbers of time-slices. Finally, in the “[Centrality measures in cattle trade networks](#)” section we will describe how these measures are correlated with the spread of infectious disease, using the French cattle trade network as an example.

## Centrality measures for temporal networks

In this section, we will describe the framework in which we will define and study centrality measures. Since most temporal network centralities are generalizations of static centralities, we will first review some classical notions of centrality, which will also be used later in the numerical percolation study on the French cattle trade network.

### Static PageRank centrality

*Static Networks.* Let  $\mathcal{G} = (V, E)$  be a directed graph with  $|V| = N$  vertices and let  $A = (a_{i,j})$  be its adjacency matrix. The aim of centrality measures is to associate to each vertex a numerical quantity representing its importance in the graph according to some criterion. In this paper, we will mainly consider centrality measures based on random walks, such that the most central vertices are the ones visited most often by random walkers on the network.

*PageRank.* The best-known such measure is *PageRank* centrality, defined as following. Let  $P$  be the transition matrix of a simple random walk on  $V$ , defined by  $p_{i,j} = a_{i,j} / \sum_{k \neq i} a_{i,k}$ . To avoid the possibility of random walkers getting stuck in vertices with no outgoing edges, we add a *teleportation* probability  $d$ , which corresponds to a transition from a vertex  $i$  to any other vertex  $j$ , uniformly chosen in  $V$ . The transition matrix of the random walk with teleportation is then:

$$\tilde{P} = \left( \frac{a_{i,j}}{\sum_{k \neq i} a_{i,k}} (1 - d) + \frac{d}{N}, 1 \leq i, j \leq N \right). \quad (1)$$

For  $d \in (0, 1]$ ,  $\tilde{P}$  is then an irreducible (row-)stochastic matrix, regardless of the connectivity structure of the original network  $\mathcal{G}$ . Hence, according to the Perron-Frobenius theorem, the left eigenspace associated to the dominant eigenvalue 1 is of dimension 1, and contains a positive eigenvector  $v$ . The PageRank vector  $v^{\text{PR}}$  is then the normalized eigenvector

$$v^{\text{PR}} = \frac{v}{\sum v_i}, \quad \text{such that} \quad v^{\text{PR}} \tilde{P} = v^{\text{PR}}. \quad (2)$$

The latter equality can be interpreted in probabilistic terms. Indeed,  $v^{\text{PR}}$  defines a probability distribution on  $V$ , and (2) is simply the stationarity of this distribution with respect to the simple random walk with teleportation on  $\mathcal{G}$ .

### Temporal network centralities

*Temporal networks.* We will consider temporal networks  $\mathcal{G} = (\mathcal{G}_t, t \in [1, T])$  defined over a discrete timespan  $t \in [1, T]$ . For each time  $t$ , we require  $\mathcal{G}_t$  to be a network  $\mathcal{G}_t = (V, E_t)$  on a fixed vertex-set  $V$  with cardinal  $N$ . For convenience, we will label the vertices of  $\mathcal{G}$  and assume that  $V = [1, N]$ . By default, the networks we consider are directed, since directionality of trade is of great importance in the main application we have in mind, namely cattle trade networks, in which nodes are holdings and edges represent cattle being moved from one holding to another.

Representing temporal networks by a sequence of discrete snapshots naturally leads to a matrix representation  $(A_t, t \in [1, T])$ , where  $A_t = (a_{t,i,j}, i, j \in V)$  is the adjacency matrix of the network  $\mathcal{G}_t$ . However, it is also possible to embed the snapshots  $\mathcal{G}_t$  inside a single *concatenated* network  $\mathcal{G}^{\text{conc}}$ , which has vertex-set  $[1, T + 1] \times V$ . In this representation, each edge  $(u, v) \in E_t$  is seen as a directed edge from vertex  $(t, u)$  to  $(t + 1, v)$ . This representation is inspired by the application to cattle trade movements, in which an edge  $(u, v)$  at time  $t$  represents the movement of an animal between two holdings, such that that animal is present in holding  $u$  at time  $t$ , then in holding  $v$  at time  $t + 1$ . Of course, it would also be possible to use edges in  $E_t$  to represent movements from time-slice  $t - 1$  to time-slice  $t$ . Using our convention however, in matrix form, the concatenated network  $\mathcal{G}^{\text{conc}}$  has an adjacency matrix defined by blocks:

$$A^{\text{conc}} = \begin{pmatrix} 0 & A_1 & 0 & \cdots & 0 \\ & 0 & A_2 & \cdots & 0 \\ & & & \ddots & \\ 0 & 0 & 0 & \cdots & A_T \end{pmatrix}. \quad (3)$$

Note that this matrix is of dimension  $TN \times (T+1)N$ . We will also sometimes consider the *wrapped* network  $\mathcal{G}^{\text{wrap}}$ . This network has vertex-set  $[1, T] \times V$  and edges from  $E_T$  go from layer  $T$  to layer 1. Its matrix representation is then

$$A^{\text{wrap}} = \begin{pmatrix} 0 & A_1 & 0 & \cdots & 0 \\ & 0 & A_2 & \cdots & 0 \\ & & & \ddots & 0 \\ 0 & 0 & 0 & \cdots & A_{T-1} \\ A_T & 0 & 0 & \cdots & 0 \end{pmatrix}. \quad (4)$$

Optionally, each of these networks can be augmented to include self-loops linking each vertex  $(t, i)$  to its representation  $(t+1, i)$  in the next time-slice (using the convention that  $T+1=1$  in the case of the wrapped network). This will be useful when considering lazy random walks, which can stay in a given vertex for some amount of time before moving using an outgoing edge.

**TempoRank centrality.** There are at least two extensions of PageRank centrality to the temporal networks. Here, we will focus on TempoRank centrality (Rocha and Masuda 2014), but we note that a different approach was considered in Rozenshtein and Gionis (2016), also taking advantage of the close relationship between PageRank scores and the stationary distributions of random walks on the network. The temporal PageRank approach of Rozenshtein and Gionis (2016) defines a time-dependent score  $\mathbf{r}(u, t)$  for each node  $u$  and each time-stamp  $t$  recorded in the temporal network by considering the number of time-respecting random walks reaching  $u$  before time  $t$ . Nodes that can be reached by many time-respecting walks will then have a high temporal PageRank score at time  $t$ . This approach is particularly useful in the case of streaming graphs, and the authors describe an efficient algorithm to update the temporal PageRank scores as more interactions are recorded, accurately reflecting rising and falling influence of a given node as time passes. However, such an approach is not well-suited to our context of outbreak mitigation on cattle trade networks, due to the periodic nature of such networks. Indeed, there is a strong annual and seasonal periodicity in the activity of certain holdings. The corresponding nodes will then have low temporal PageRank scores after long periods of inactivity, while being of critical importance in epidemic transmission during their activity period.

By contrast, the TempoRank score introduced in Rocha and Masuda (2014) summarizes all temporal information contained in a set of time-stamped edges into a single score by averaging over all observation times. TempoRank centrality considers the wrapped network  $\mathcal{G}^{\text{wrap}}$  associated with a given temporal network  $\mathcal{G}$ . With the application to cattle trade networks in mind, we will assume that  $\mathcal{G}$ , hence also  $\mathcal{G}^{\text{wrap}}$  are directed. Let  $q \geq 0$  be a *laziness* parameter and  $d \geq 0$  be a *teleportation* parameter. The transition matrix for the lazy random walk with teleportation is then defined by:

$$(B_t)_{ij} = \begin{cases} q + (1-q)d/N + (1-q)(1-d), & \text{if } \text{odeg}_t(i) = 0, i = j \\ (1-q)d/N, & \text{if } \text{odeg}_t(i) = 0, i \neq j \\ q + (1-q)d/N, & \text{if } \text{odeg}_t(i) > 0, i = j, \\ (1-q)[d/N + a_{t,ij}(1-d)/\text{odeg}_t(i)], & \text{if } \text{odeg}_t(i) > 0, i \neq j, \end{cases} \quad (5)$$

with the out-degree  $\text{odeg}_t(i)$  defined by:

$$\text{odeg}_t(i) = \sum_{j \neq i} a_{t,ij}. \quad (6)$$

In other words, random walkers stay in place with probability  $q$  (note that the authors of Rocha and Masuda (2014) used a degree-dependent probability  $q^{\text{deg}_t(i)}$  of staying in place). If they decide to move, they teleport to a uniformly chosen vertex with probability  $d$ , or move along a uniformly chosen outgoing edge with probability  $1-d$ . Note that if the current vertex is isolated at time  $t$  ( $\text{odeg}_t(i) = 0$ ), there is no such outgoing edge, and they stay in place (this occurs with probability  $(1-q)(1-d)$ ). Using the wrapped network representation of (4), we then get the following transition matrix on  $\mathcal{G}^{\text{wrap}}$ :

$$B^{\text{wrap}} = \begin{pmatrix} 0 & B_1 & 0 & \cdots & 0 \\ & 0 & B_2 & \cdots & 0 \\ & & 0 & \ddots & 0 \\ 0 & 0 & 0 & \cdots & B_{T-1} \\ B_T & 0 & 0 & \cdots & 0 \end{pmatrix}. \quad (7)$$

By definition,  $B^{\text{wrap}}$  is a stochastic matrix, and, by the Perron-Frobenius theorem, admits a (unique up to a multiplicative constant) left 1-eigenvector  $\Pi$  with nonnegative elements. Writing left 1-eigenvectors under the form  $\Pi = (\pi_1 \dots \pi_T) \in \mathbb{R}^{TN}$  with  $\pi_i \in \mathbb{R}^N$ , we see that they need to satisfy the equations:

$$\pi_T B_T = \pi_1, \pi_1 B_1 = \pi_2, \dots, \pi_{T-1} B_{T-1} = \pi_T, \quad (8)$$

which is equivalent to

$$\begin{aligned} \pi_T B_T B_1 \cdots B_{T-1} &= \pi_T, \\ \pi_1 B_1 B_2 \cdots B_T &= \pi_1, \\ &\dots \\ \pi_{T-1} B_{T-1} B_T \cdots B_{T-2} &= \pi_{T-1}. \end{aligned} \quad (9)$$

In other words, the  $\pi_t$  are the stationary distributions of the random walks  $(X_{mT+t})$  as  $m \rightarrow \infty$ , where  $X$  has one-step transition matrix given by (5). Averaging over time, we define:

$$\text{TR}(x) = \frac{1}{T} \sum_{t=1}^T \pi_t(x) \quad (10)$$

We will also consider the following version of TempoRank that takes into account the existence of outgoing edges in a given time-slice. We will call it *out-TempoRank* centrality. It gives no weight to vertices with no outgoing edges, by contrast with regular TempoRank centrality.

$$\text{oTR}(x) = \frac{1}{T} \sum_{t=1}^T \pi_t(x) \mathbf{1}_{\text{odeg}_t(x) > 0} \quad (11)$$

### Random walk centralities: definition and properties

#### Stochastic approximation

We will now see how direct stochastic simulations can be used to approximate random-walk based centrality measures. This makes their computation effectively possible on very large networks, for which matrix-based methods are not feasible. Indeed, eigenvalue methods for large matrices are relying on sparsity assumptions that are not satisfied here: for  $d > 0$ , the transition matrices  $B_t$  are always dense. Even if  $d = 0$ , while the individual transition matrices might be sparse, depending on the connectivity structure of the underlying network, the cyclical products used in the definition of the TempoRank will lose sparsity as the number of time-slices increases.

For  $q \in [0, 1]$  and  $d \in [0, 1]$ , let then  $(X_n, n \geq 0)$  be a Markov chain on  $\mathcal{G}^{\text{wrap}}$  with transition matrix  $B^{\text{wrap}}$  (7). There is an obvious projection map  $\pi : \mathcal{G}^{\text{wrap}} \rightarrow \mathcal{G}$  such that  $\pi(t, x) = x$ ,  $1 \leq t \leq T$ ,  $x \in V$ . We are interested in the asymptotic behavior of the *visit times*  $V_n$ , as well as the *exit times*  $O_n$ , defined respectively by:

$$V_n(x) = |\{k \leq n, \pi(X_k) = x\}|, \quad (12)$$

$$O_n(x) = |\{k \leq n, \pi(X_k) = x \text{ and } \pi(X_{k+1}) \neq x\}|. \quad (13)$$

We can use these quantities to approximate both TempoRank and out-TempoRank centrality. Indeed, if  $d > 0$ , then  $X$  is an irreducible Markov chain on  $V$ , and

$$\lim_{n \rightarrow \infty} n^{-1} V_n(x) = \text{TR}(x), \quad (14)$$

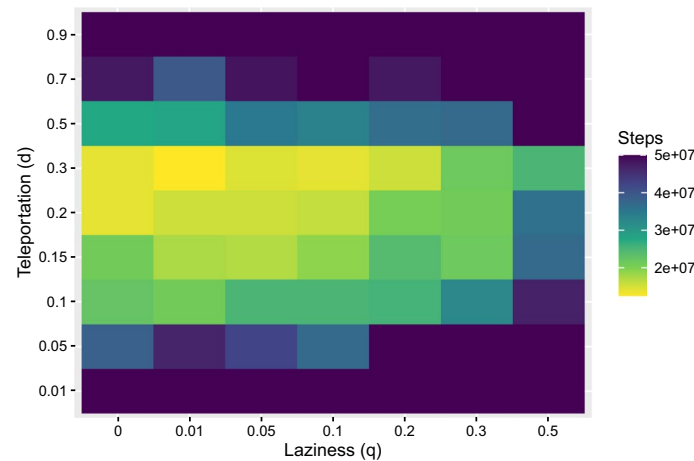
$$\lim_{n \rightarrow \infty} n^{-1} O_n(x) = (1 - q) \left[ d \frac{N-1}{N} \text{TR}(x) + (1 - d) \text{oTR}(x) \right]. \quad (15)$$

Hence, we can compute TempoRank and out-TempoRank centrality by simulating time-respecting random walks on  $\mathcal{G}^{\text{wrap}}$  and by counting the time spent in each vertex, resp. the number of movements exiting a given vertex through an outgoing edge (not through teleportation). Both the deterministic (matrix-based) and the stochastic algorithm are freely available<sup>1</sup>. The use of such methods to compute centralities when the size of the graph makes matrix-based methods unusable is classical (Avrachenkov et al. 2007; Broder et al. 2006; Bahmani et al. 2010) and has been extensively studied in the context of internet search engines.

#### Convergence of random walks

In order to quantify the effect of the laziness parameter  $q$  and the teleportation parameter  $d$  on the speed of convergence of the quantities described above to their deterministic limit, we used the well-known Primary School temporal network

<sup>1</sup> <https://github.com/phoscheit/DynPageRank>



**Fig. 1** Number of steps required to have total absolute percentage error between estimated and actual out-TempoRank score below 1%, as a function of laziness ( $q$ ) and teleportation ( $d$ ) parameters

(Gemmetto et al. 2014). This network is constituted by face-to-face close interactions between high school students and teachers, as measured in 20-second intervals. We aggregated all contacts occurring in a given hour into a single time-slice, which resulted in a temporal network of 242 nodes and 26,603 edges over 19 time-slices.

We computed the vector of exit times for a random walk with  $5 \times 10^7$  steps, logging interim values at increments of  $5 \times 10^5$  steps, for varying values of  $d$  and  $q$ . We also computed the exact value of the out-TempoRank vector using (11) and compared it to the stochastic approximation at each iteration using Total Absolute Percentage Error:

$$TAPE = \sum_{x \in V} \frac{|O_n(x) - \text{oTR}(x)|}{\text{oTR}(x)}.$$

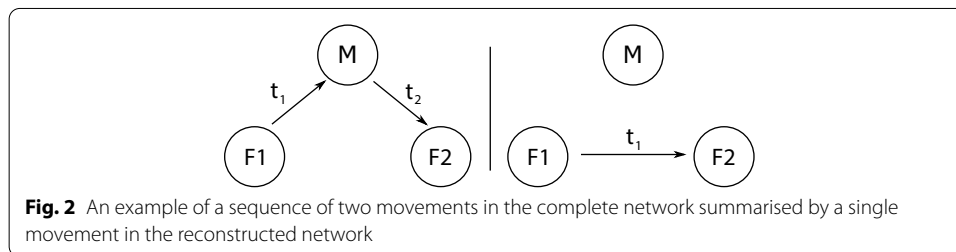
Results (shown in Fig. 1) indicate that, for this network at least, the fastest convergence is obtained for small values of  $q$  (including 0), which of course reflects the fact that laziness slows down the random walks. Most importantly, there also exists an optimal value of  $d$  for which convergence is fastest. For the schools network, this value lies around  $d = 0.3$ . Understanding how this optimal parameter choice is related to the geometry of the wrapped temporal network could be of great importance in tuning the approximation algorithm in settings for which deterministic computations are unfeasible.

### Centrality measures in cattle trade networks

In this section, we will use the previously defined centrality measures to identify the nodes most responsible for the spread of disease in a large network of cattle movements in France. Other than the static and temporal PageRank centralities defined above, we also consider Katz and betweenness centrality (definitions can be found in Additional file 1: section 1).

**Table 1** Summary statistics of the French cattle trade networks, aggregated over a year

	Complete		Reconstructed	
	2014	2015	2014	2015
Number of vertices	179,518	174,034	169,248	164,171
Number of edges	989,684	980,376	1,401,577	1,371,844
Number of movements	2,318,824	2,268,494	1,552,554	1,522,314
Greatest strongly c.c.	74,044	68,746	52,782	49,445
Average out-degree	5.51	5.63	8.28	8.35
Maximum in-degree	12,426	11,673	2,087	2,108
Maximum out-degree	2,767	2,252	351	389
Exponent of degree distribution	2.12	2.24	5.11	5.29
Average clustering	0.316	0.315	0.033	0.036
Degree assortativity	−0.120	−0.123	−0.30	−0.316



### The French BDNI network

The French Base de Données Nationale d'Identification (BDNI) is a national database logging all movements of cattle between two holdings in France. For the purpose of this work, we extracted all movements occurring in a two-year period between 2014-01-01 and 2015-12-31. We used these movements to construct a temporal network, with holdings as vertices and cattle movements as edges, with a time-stamp corresponding to the day of the movement. We summarized some statistics of the 2014 and 2015 networks in Table 1 below. A more detailed statistical analysis of the BDNI networks between 2005 and 2009 can be found in Dutta et al. (2014).

An important feature of this network and more generally of cattle trade networks is the presence of markets and assembly centres, through which a significant part of all trades occur. They are the obvious most central vertices in any cattle trade network (Vidondo and Voelkl 2018), but they may not be hotspots of disease transmission, in particular for pathogens that require prolonged close contact for infection, such as paratuberculosis or BVDV (Gates et al. 2014). Indeed, animals remain in these holdings for a single day (or up to a week in the case of assembly centres), which might not be long enough for significant transmission to occur. To account for such dynamics, we consider two versions of the BDNI network: the first (dubbed the *complete* network), includes all movements to and from markets and assembly centres, whereas the second (the *reconstructed* network) bypasses these vertices by reconstructing movements between farms that involve a market or assembly centre as intermediary. More precisely, if an animal moves from a farm  $F_1$  to a market or assembly centre at a date  $t_1$ , then subsequently moves to another



farm  $F_2$  at a later date  $t_2$ , we will log a single movement from  $F_1$  to  $F_2$  at time  $t_1$  in the reconstructed network (Fig. 2).

As can be seen in Table 1, the most striking difference between the complete and reconstructed networks lies in their degree distribution. Assuming a power-law distribution, we computed its tail exponent, which is such that

$$\mathbb{P}(D = n) \sim_{n \rightarrow \infty} n^{-\alpha}$$

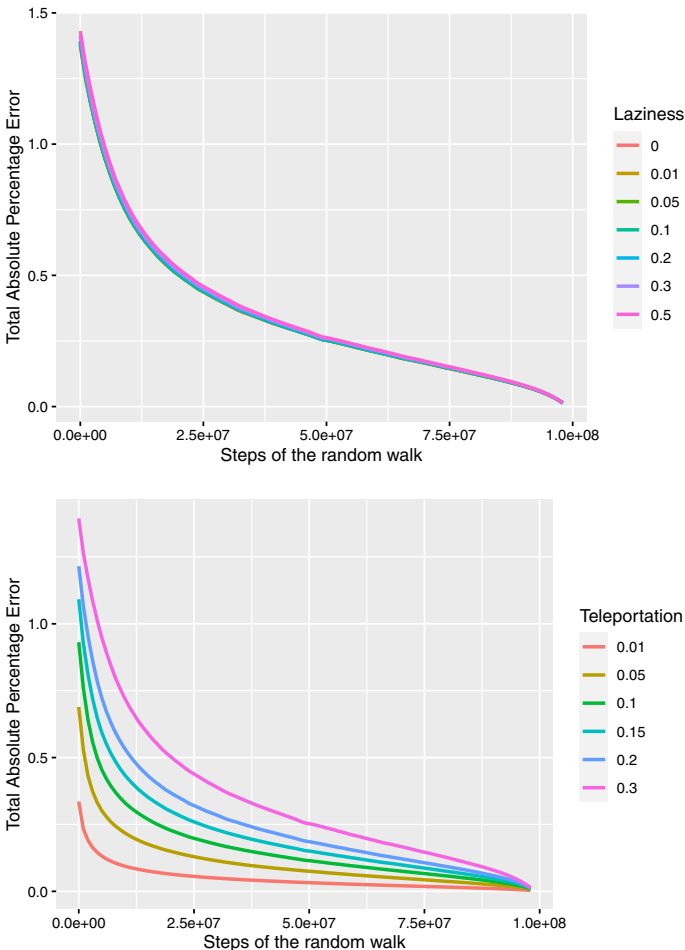
for the degree  $D$  of a uniformly chosen vertex in the network. While the complete networks exhibit so-called *scale-free* behavior ( $1 < \alpha < 3$ ), the reconstructed networks have  $\alpha > 3$ , indicating a network without large hubs. The same behavior can be seen with the higher average out-degree in the reconstructed networks, caused by the splitting of single farm-to-market or farm-to-assembly-centre movements into several farm-to-farm movements. As we will see below, the different connectivity structures lead to markedly different epidemic behavior and thus, to different hierarchies between centrality measures used to control the epidemics.

### Centrality in cattle trade networks

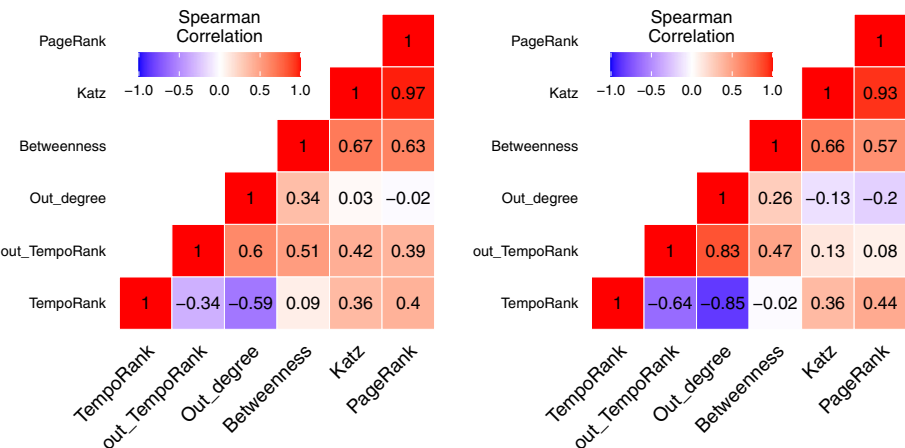
There have been several studies of centrality in cattle trade networks. Most approaches have focused on (temporal) path-counting methods, such as the Disease-Flow centralities in Natale et al. (2009), closely related to the outgoing and ingoing contact chains described in Nöremark et al. (2011). Such methods have proved useful in identifying epidemiologically important nodes (Büttner et al. 2013; Vidondo and Voelkl 2018) but fail for large-scale networks due to combinatorial explosion in the number of nodes or the number of time-steps. Other approaches are simulation-based, aiming at identifying suitable control strategies by directly simulating outbreaks (using, for instance, generic SI-type transmission dynamics) on the temporal networks (Payen et al. 2019; Bajardi et al. 2012).

We computed static and temporal centrality measures for the 2014 and 2015 BDNI networks, both in their complete and reconstructed version. Due to the size of the network, computation of the TempoRank measures using deterministic methods was impossible. We used the stochastic approximations of (14)–(15), with  $10^8$  steps for the random walks. We compared the centralities at increments of  $10^6$  steps to the final state to check convergence (Fig. 3). We observe that the laziness parameter has no influence on the speed of convergence, whereas for fixed  $q$ , the fastest convergence was achieved for low values of  $d$ .

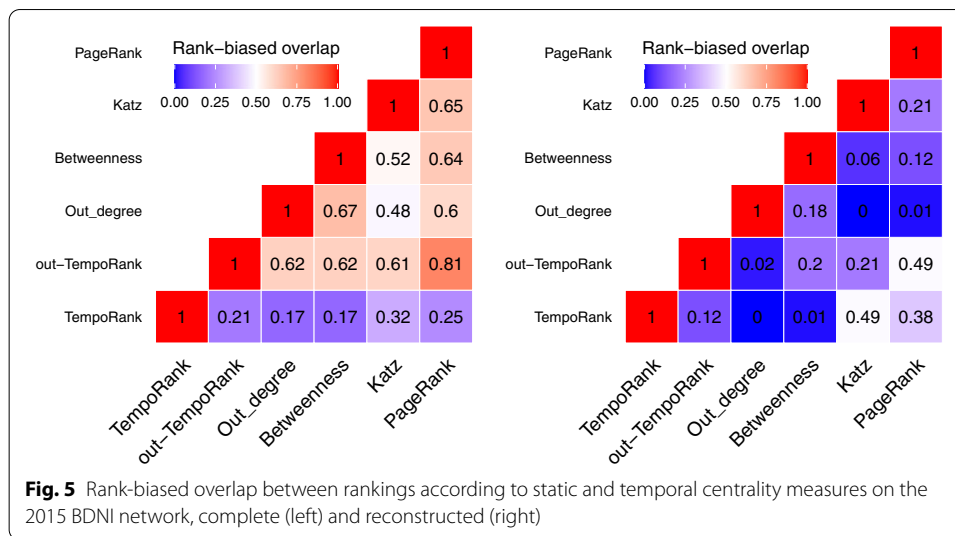
In order to assess correlations between measures, we computed Spearman (rank) coefficients (Fig. 4). This analysis shows the high level of similarity between Katz and PageRank centrality, as well as moderate levels of correlation between Katz centrality, PageRank centrality and betweenness, as well as out-degree centrality and out-TempoRank. Broadly speaking, this suggests a rough classification of the six measures considered here into three categories: those measuring the outgoing centrality of a node (outgoing degree and out-TempoRank) and those measuring its ingoing centrality (TempoRank, Katz and PageRank centrality). Betweenness centrality is moderately correlated to measures in both groups, since it takes both incoming and outgoing paths into account.



**Fig. 3** Total absolute percentage error between the state of the approximation algorithm after  $10^8$  steps and its state after increments of  $10^6$  steps, for the 2014 complete BDNI network, for  $d = 0.3$  and varying levels of  $q$  (top), and  $q = 0.1$  and varying levels of  $d$  (bottom)



**Fig. 4** Spearman correlation coefficients between static and temporal centrality measures on the 2015 BDNI network, complete (left) and reconstructed (right)



In addition to this analysis, we also compared the rankings of vertices according to the different centralities, using rank-biased overlap (RBO) (Webber et al. 2010). This measure of similarity between two rankings (permutations)  $\sigma$  and  $\tau$  of  $[1, N]$  is defined by

$$RBO(\sigma, \tau) = (1 - p) \sum_{k=1}^N p^{k-1} \frac{|\sigma_{1:k} \cap \tau_{1:k}|}{k},$$

where  $\sigma_{1:k}$  is the set of the  $k$  highest-ranked elements in  $\sigma$  and where  $p \in (0, 1)$  is a parameter quantifying how much weight should be placed on comparisons between the highest-ranking elements. In essence, the smaller  $p$ , the less importance given to the overlap  $\sigma_{1:k} \cap \tau_{1:k}$  with high  $k$ . Figure 5 shows the pairwise RBO scores for  $p = 0.997$ , chosen so that 95% of weight is given to the top 1000 vertices in the network. There is a striking dissimilarity between the complete and the reconstructed network. On the complete network, most centrality measures have moderately high overlap, with the exception of the TempoRank measure, which has low overlap with all other centralities. The latter is probably due to it highly ranking vertices with no or few outgoing edges, on which random walkers spend a disproportionate amount of time, thereby increasing their TempoRank score. On the reconstructed network however, almost all pairs of centralities have very low overlap, the most similar being TempoRank and Katz centrality as well as out-TempoRank and PageRank centrality. This shows that the problem of identifying high-centrality nodes is much more difficult on a diffuse network such as the reconstructed cattle trade network, whereas on scale-free networks, most measures have similar rankings, perhaps with slight differences in the order.

### Percolation analysis

Ultimately, we wish to understand how centrality of a node is correlated with the role this node plays in epidemic outbreaks and if it can be a good predictor of the spread of infection on the temporal network under study. This is a classical problem in network analysis (Lü et al. 2016) and is usually studied either by direct simulation (Vidondo and Voelkl 2018; Payen et al. 2019) or by computation of a proxy of epidemic spread, such

as the size of the largest connected component (Mweu et al. 2013; Büttner et al. 2013) or the so-called epidemic threshold (Valdano et al. 2015, 2018). In this section, we will use as a metric the mean final outbreak size of simulated SIR processes. We will use the `tsir` package for fast outbreak simulation (Holme 2020). This package simulates epidemic outbreaks on temporal networks in the following manner:

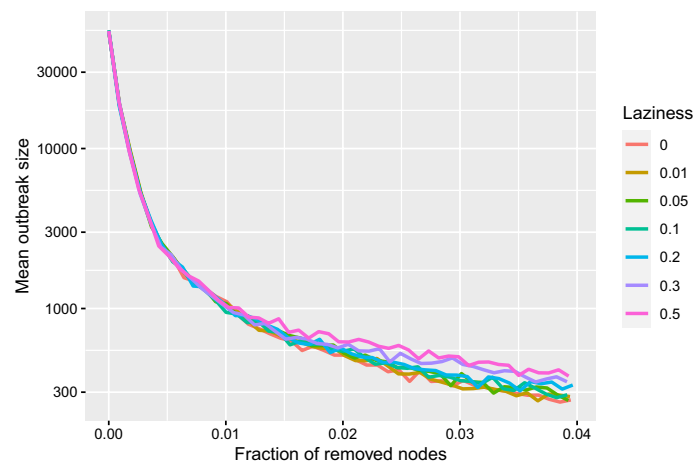
1. Choose a node uniformly at random in the network and set its status to “infected” at the start of the observation period (2015-01-01 in our case). All other nodes are set to “susceptible”.
2. Each infected node remains in that state for an exponentially distributed time with parameter  $\gamma > 0$ , after which it becomes “removed”.
3. For each time-stamped edge  $(t, u, v)$  such that  $u$  is infected and  $v$  is susceptible at time  $t$ ,  $v$  becomes infected with probability  $\beta \in (0, 1]$ .

For each centrality measure, we ranked vertices in decreasing order of centrality, the measures being computed on the corresponding 2014 movement network. We then removed an increasing proportion of vertices in the 2015 network, ranging from 0.1% to 5% of all vertices. When removing a vertex, we removed all edges from and to that vertex. We then simulated 10,000 iterations of a temporal SIR process on the percolated 2015 network, with per-contact infection probability  $\beta = 0.5$  and recovery rate  $\gamma = 1$  per day. On the static network aggregating all edges with timestamps in 2015, this corresponds to an epidemic with basic reproduction number (Durrett 2007) given by:

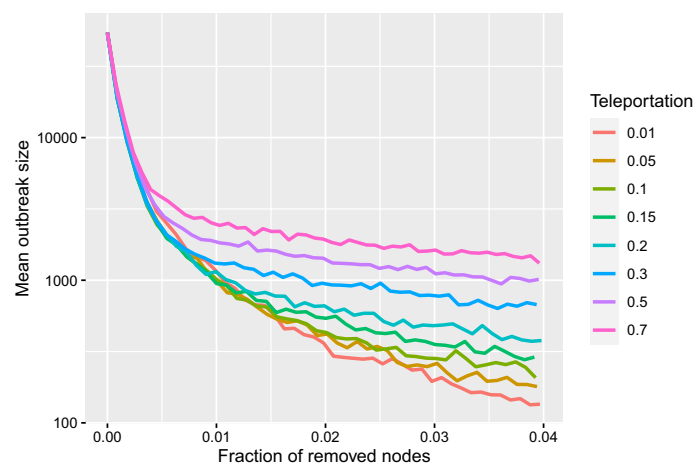
$$R_0 = \frac{\beta}{\gamma} \left( \frac{\langle k^2 \rangle}{\langle k \rangle} - 1 \right),$$

where  $\langle k \rangle$  and  $\langle k^2 \rangle$  are the first and second moment of the degree distribution. This would give  $R_0 = 28.75$  and  $R_0 = 10$  on the complete and reconstructed networks, respectively. However, when taking temporality into account, contagion dynamics are much less explosive since the number of contacts of a given node need only be considered over a short time period (Enright and Kao 2018). Exact computation of the epidemic threshold for general temporal networks is possible (Valdano et al. 2018) but can be computationally unfeasible for large networks. Interestingly, we found that final outbreak sizes had low variance (see Additional file 1: section 2 for more data), perhaps reflecting the fact that once a pathogen reaches a node with high connectivity and high activity, outbreaks tend to be quite similar.

We calibrated the TempoRank algorithm by computing centrality using different values of laziness  $q$  and teleportation  $d$  and comparing the mean final outcome curves (Figs. 6 and 7). Complete results for all examined combinations of  $q$  and  $d$  can be found in Additional file 1: section 6. Interestingly, the laziness parameter seemed to have only a very limited impact on final outbreak sizes, with higher laziness leading to higher outbreak sizes. This effect is only appearing for high fractions of removed nodes; it seems that, whatever value of  $q$  was chosen for out-TempoRank computation, the highest-ranked nodes were identical. A similar effect can be seen when comparing disintegration curves for varying levels of teleportation, although the differences are much more important. The best outcomes were obtained when



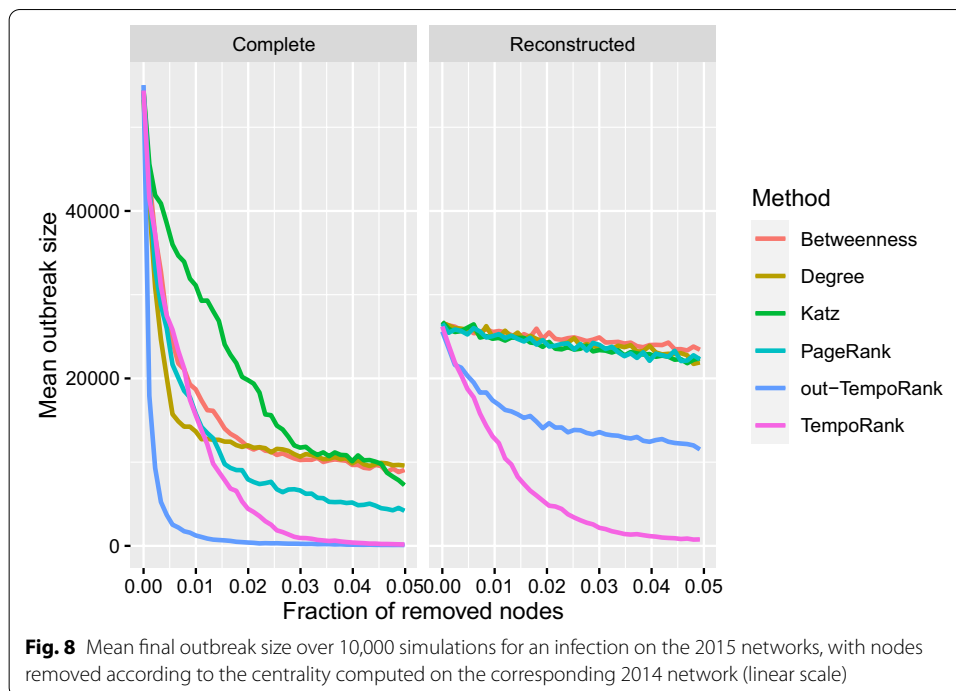
**Fig. 6** Mean final outbreak size over 10,000 simulations for an infection on the 2015 complete BDNI network, with nodes removed according to out-TempoRank centrality computed on the 2014 complete BDNI network, with fixed teleportation probability  $d = 0.15$  and varying levels of laziness  $q$  (logarithmic scale)



**Fig. 7** Mean final outbreak size over 10,000 simulations for an infection on the 2015 complete BDNI network, with nodes removed according to out-TempoRank centrality computed on the 2014 complete BDNI network, with fixed laziness  $q = 0.1$  and varying teleportation parameters  $d$  (logarithmic scale)

teleportation was set to the lowest levels ( $d = 0.01$ ). We then used TempoRank and out-TempoRank scores computed with the optimal values  $q = 0$  and  $d = 0.01$  for comparison with other centrality measures.

Results of the percolation experiment with all centrality measures taken into account are shown in Fig. 8. In the complete graph, the rapid decline in final outbreak sizes, even for small numbers of removed nodes, is a well-known feature of scale-free networks when nodes are targeted appropriately (not uniformly at random). For most centrality measures, the highest-ranked nodes tend to be nodes with high degree, and their removal leads to rapid disintegration of the connectivity structure of the network. Indeed, in cattle trade networks, markets and assembly centres act as trade hubs (Hoscheit et al. 2017; Natale et al. 2009; Bajardi et al. 2011; Salines et al. 2017)



and they are highly ranked by most centralities (see Additional file 1: section 3), which leads to similar disintegration profiles. Nevertheless, out-TempoRank centrality is the best-performing measure by this metric. By contrast, TempoRank centrality performs similarly at first than the static measures we included in our analysis. However, the disintegration curves associated to most static measures exhibit a plateau after a rapid initial decline, which is not the case for TempoRank centrality. Interestingly, comparison of the fraction of markets and assembly centres in the highest-ranked nodes (Additional file 1) shows that out-TempoRank manages to identify those nodes as influential spreaders. By contrast, there is a smaller fraction of markets and assembly centres in the highly ranked nodes according to TempoRank, yet it still manages to effectively prevent epidemic spread at around 3% of removed nodes. The results we obtained, in particular the hierarchy of centrality measures, were consistent across the parameter space for the epidemic process (Additional file 1).

In the reconstructed graph, most disintegration profiles are linear, with different slopes depending on the centrality measure used to select the removed nodes. This is again as expected, because large hubs have been removed and replaced with many farm-to-farm edges (as can be seen in the exponent of the degree distribution in Table 1). This makes the node removal problem more difficult, since there are potentially many redundant paths that can be taken by infections. We see that the temporal centrality measures significantly outperform the static measures on the reconstructed network, which shows that temporal information plays a major part in the spread of infection on cattle trade networks, once its strong connectivity properties have been muted. On this network, the regular TempoRank measure outperforms out-TempoRank, suggesting that the re-weighting of nodes by the fraction of time that they had active outgoing edges is a poor

predictor of epidemic spread on this network. The difference with the complete network is striking in that regard.

We observed no qualitative difference in the disintegration curves for outbreaks simulated on the 2015 networks, whether the centralities were computed on the 2014 or the 2015 data (see Additional file 1: section 4). This is indicative of high year-to-year similarity of the global structure of the networks, even though at the local level, there can be significant variation in the choice of trading partners (Valdano et al. 2015). Also, since contemporary data is not always available at the time of implementation of control measures, this shows that relying on past data can still produce high-quality outbreak prevention and mitigation.

## Conclusion

In this work, we showed how random-walk-based centrality measures are well-suited for large temporal networks, since they can be efficiently computed using stochastic simulations. We demonstrate their suitability for the purpose of epidemic mitigation by targeted removal of nodes, and their improved performance when compared to measures computed on the static (time-aggregated) network.

However, there are some limitations inherent to the definition of TempoRank and out-TempoRank centralities. Indeed, they are defined as time-averaged quantities as can be seen from their computation through the stationary distributions of random walks on the time-wrapped network  $\mathcal{G}^{\text{wrap}}$ . This makes them robust to small local changes in the network, but fails to take into account heterogeneous behaviour, such as seasonality (Vidondo and Voelkl 2018) or changes in network dynamics which would make the recent past a better predictor of future outbreaks than older temporal patterns. It should be possible to account for such ruptures, provided they can be accurately detected (Ranshous et al. 2015; Donnat and Holmes 2018; Monnig and Meyer 2018) by appropriately reweighting time-slices in the definition of TempoRank centrality (10). Other, so-called *online* centrality measures update the scores as new interactions are added to the dataset (Béres et al. 2018; Rozenshtein and Gionis 2016) but fail to identify older, periodically active nodes which might have an important role to play in epidemic spread in specific timeframes.

An important next step will be the use of centrality measures for targeted control measures more subtle than the outright removal procedure studied here. For instance, targeted vaccination campaigns or diagnostic strategies are some of the tools available for the control of cattle disease, but they need to be weighted against their economic impact. Better focused strategies could yield similar epidemiological outcome at lower economic cost.

In a similar fashion, the effect of parameters  $q$  and  $d$  on final epidemic size should be explored further. For instance, on temporal networks with different geometries (such as sexual contact networks), it would be interesting to understand whether low final outbreak sizes are always associated with low teleportation parameters in the TempoRank scores or if this is a specificity of networks that are similar to cattle trade networks.

It will also be important to obtain quantitative results on the speed of convergence of the stochastic algorithm to the TempoRank and out-TempoRank centrality vectors. It is well-known (Levin and Peres 2017) that mixing times of Markov chains are strongly

related, among other things, to the spectral gap of their transition matrices. It is not clear how these results carry over to the temporal setting, and we hope that further work in that direction will enable fine tuning of the number of steps required to compute good approximations of temporal centrality measures when deterministic algorithms are not practicable.

#### Abbreviations

BDNI: Base de Données Nationale d'Identification (National Cattle Identification Database); SIR: Susceptible-infective-recovered; TAPE: Total absolute percentage error; BVDV: Bovine viral Diarrhea virus; RBO: Rank-biased overlap.

#### Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1007/s41109-021-00368-5>.

**Additional file 1.** Supplementary information regarding precise definitions of static centralities (Section S1), variance in the simulated outbreak sizes (Section S2), the relative weight of markets and assembly centres in highly ranked nodes (Section S3), the impact of using different yearly datasets to compute centralities (Section S4), the impact of varying epidemic parameters (Section S5), and the impact of varying random walk parameters to approximate TempoRank measures (Section S6).

#### Acknowledgements

The authors would like to acknowledge the help of Gaël Beaunée and Simon Labarthe. We are grateful to the INRAE MIGALE bioinformatics facility (MIGALE, INRAE, 2020. Migale bioinformatics Facility, doi: 10.15454/1.5572390655343293E12) for providing computing and storage resources."

#### Authors' contributions

PH and EV conceived and designed the analysis. PH and EA contributed analysis tools and performed the analysis. PH and EV wrote the paper. All authors read and approved the final manuscript.

#### Funding

This work was carried out with the financial support of the French National Research Agency (ANR), project ANR-16-CE32-0007-01 (CADENCE).

#### Availability of data and materials

The "schools" dataset is available in the SocioPatterns repository <http://www.sociopatterns.org>. The BDNI dataset is available from the French Ministry of Agriculture but restrictions apply to the availability of these data, which were used under license for the current study, and so are not publicly available. Data are however available from the authors upon reasonable request, subject to permission of the Ministry of Agriculture.

#### Declaration

##### Competing interests

The authors declare that they have no competing interests.

Received: 29 October 2020 Accepted: 5 March 2021

Published online: 24 March 2021

#### References

- Agryzkov T, Curado M, Pedroche F, Tortosa L, Vicent JF (2019) Extending the adapted PageRank algorithm centrality to multiplex networks with data using the PageRank two-layer approach. *Symmetry* 11(2):1–17. <https://doi.org/10.3390/SYM11020284>
- Avrachenkov K, Litvak N, Nemirovsky D, Osipova N (2007) Monte Carlo methods in pagerank computation: when one iteration is sufficient. *SIAM J Numer Anal* 45(2):890–904. <https://doi.org/10.1137/050643799>
- Bahmani B, Chowdhury A, Goel A (2010) Fast incremental and personalized PageRank. *Proc VLDB Endowment* 4(3):173–184. <https://doi.org/10.14778/1929861.1929864>
- Bajardi P, Barrat A, Natale F, Savini L, Colizza V (2011) Dynamical patterns of cattle trade movements. *PLoS ONE* 6(5):19869. <https://doi.org/10.1371/journal.pone.0019869>
- Bajardi P, Barrat A, Savini L, Colizza V (2012) Optimizing surveillance for livestock disease spreading through animal movements. *J R Soc Interface* 9(76):2814–25. <https://doi.org/10.1098/rsif.2012.0289>
- Béres F, Pálovics R, Oláh A, Benczúr AA (2018) Temporal walk based centrality metric for graph streams. *Appl Netw Sci* 3(1):1–26. <https://doi.org/10.1007/s41109-018-0080-5>
- Broder AZ, Lempel R, Maghoul F, Pedersen J (2006) Efficient PageRank approximation via graph aggregation. *Inf Retrieval* 9(2):123–138. <https://doi.org/10.1007/s10791-006-7146-1>
- Büttner K, Krieter J, Traulsen A, Traulsen I (2013) Efficient interruption of infection chains by targeted removal of central holdings in an animal trade network. *PLoS ONE* 8(9):74292. <https://doi.org/10.1371/journal.pone.0074292>
- Donnat C, Holmes S (2018) Tracking network dynamics: a survey using graph distances. *Ann Appl Stat* 12(2):971–1012. <https://doi.org/10.1214/18-AOAS1176>



- Durrett R (2007) Random graph dynamics. Cambridge University Press, Cambridge
- Dutta BL, Ezanno P, Vergu E (2014) Characteristics of the spatio-temporal network of cattle movements in France over a 5-year period. *Prevent Veterin Med* 117(1):79–94. <https://doi.org/10.1016/j.prevetmed.2014.09.005>
- Dutta R, Mira A, Onnela J-P (2018) Bayesian inference of spreading processes on networks. *Proc R Soc Math Phys Eng Sci* 474(2215):20180129. <https://doi.org/10.1098/rspa.2018.0129>
- Enright J, Kao R (2018) Epidemics on dynamic networks. *Epidemics*. <https://doi.org/10.1016/j.epidem.2018.04.003>
- Gates MC, Humphry RW, Gunn GJ, Woolhouse MEJ (2014) Not all cows are epidemiologically equal: quantifying the risks of bovine viral diarrhoea virus (BVDV) transmission through cattle movements. *Vet Res* 45(1):110. <https://doi.org/10.1186/s13567-014-0110-y>
- Gemmetto V, Barrat A, Cattuto C (2014) Mitigation of infectious disease at school: targeted class closure vs school closure. *BMC Infect Dis* 14(1):1–10. <https://doi.org/10.1186/s12879-014-0695-9.1408.7038>
- Ghahmane Z, Cherifi C, Cherifi H, Hassouni ME (2019) Centrality in complex networks with overlapping community structure. *Sci Rep* 9(1):1–29. <https://doi.org/10.1038/s41598-019-46507-y>
- Hanke M, Foraita R (2017) Clone temporal centrality measures for incomplete sequences of graph snapshots. *BMC Bioinform* 18(1):261. <https://doi.org/10.1186/s12859-017-1677-x>
- Holme P (2020) Fast and principled simulations of the SIR model on temporal networks. *arXiv preprint* 1–15. 2007.14386
- Hoscheit P, Geeraert S, Beaunée G, Monod H, Gilligan CA, Filipe JAN, Vergu E, Moslonka-lefebvre M (2017) Dynamical network models for cattle trade: towards economy-based epidemic risk assessment. *J Complex Netw* 5(4):604–624. <https://doi.org/10.1093/comnet/cnw026>
- Jalili M, Salehzadeh-Yazdi A, Asgari Y, Arab SS, Yaghmaie M, Ghavamzadeh A, Alimoghaddam K (2015) CentiServer: a comprehensive resource, web-based application and R package for centrality analysis. *PLoS ONE* 10(11):1–8. <https://doi.org/10.1371/journal.pone.0143111>
- Kim H, Anderson R (2012) Temporal node centrality in complex networks. *Phys Rev E Stat Nonlinear Soft Matter Phys* 85(2):1–8. <https://doi.org/10.1103/PhysRevE.85.026107>
- Levin DA, Peres Y (2017) Markov chains and mixing times. American Mathematical Society, Providence, p 464
- Lü L, Chen D, Ren XL, Zhang QM, Zhang YC, Zhou T (2016) Vital nodes identification in complex networks. *Phys Rep* 650:1–63. <https://doi.org/10.1016/j.physrep.2016.06.007>
- Lv L, Zhang K, Zhang T, Li X, Zhang J, Xue W (2019) Eigenvector centrality measure based on node similarity for multilayer and temporal networks. *IEEE Access* 7:115725–115733. <https://doi.org/10.1109/ACCESS.2019.2936217>
- Monnig ND, Meyer FG (2018) The resistance perturbation distance: a metric for the analysis of dynamic networks. *Discrete Appl Math* 236:347–386. <https://doi.org/10.1016/j.dam.2017.10.007>
- Mweu MM, Fournié G, Halasa T, Toft N, Nielsen SS (2013) Temporal characterisation of the network of Danish cattle movements and its implication for disease control: 2000–2009. *Prevent Vet Med* 110(3–4):379–87. <https://doi.org/10.1016/j.prevetmed.2013.02.015>
- Natale F, Giovannini A, Savini L, Palma D, Possenti L, Fiore G, Calistri P (2009) Network analysis of Italian cattle trade patterns and evaluation of risks for potential disease spread. *Prevent Vet Med* 92(4):341–50. <https://doi.org/10.1016/j.prevetmed.2009.08.026>
- Natale F, Savini L, Giovannini A, Calistri P, Candeloro L, Fiore G (2011) Evaluation of risk and vulnerability using a disease flow centrality measure in dynamic cattle trade networks. *Prevent Veterin Med* 98(2–3):111–8. <https://doi.org/10.1016/j.prevetmed.2010.11.013>
- Nöremark M, Håkansson N, Lewerin SS, Lindberg A, Jonsson A (2011) Network analysis of cattle and pig movements in Sweden: measures relevant for disease control and risk based surveillance. *Prevent Vet Med* 99(2–4):78–90. <https://doi.org/10.1016/j.prevetmed.2010.12.009>
- Payen A, Tabourier L, Latapy M (2019) Spreading dynamics in a cattle trade network: size, speed, typical profile and consequences on epidemic control strategies. *PLoS ONE* 14(6):0217972. <https://doi.org/10.1371/journal.pone.0217972>
- Pedroche F, Romance M, Criado R (2016) A biplex approach to PageRank centrality: from classic to multiplex networks. *Chaos Interdiscip J Nonlinear Sci*. <https://doi.org/10.1063/1.4952955>
- Ranshous S, Shen S, Koutra D, Harenberg S, Faloutsos C, Samatova NF (2015) Anomaly detection in dynamic networks: a survey. *Wiley Interdiscip Rev Comput Stat* 7(3):223–247. <https://doi.org/10.1002/wics.1347>
- Rocha LEC, Masuda N (2014) Random walk centrality for temporal networks. *N J Phys*. <https://doi.org/10.1088/1367-2630/16/6/063023>
- Rozenshtein P, Gionis A (2016) Temporal PageRank. *Mach Learn Knowl Discov Databases*. <https://doi.org/10.1007/978-3-319-46227-1>
- Salines M, Andraud M, Rose N (2017) Pig movements in France: designing network models fitting the transmission route of pathogens. *PLoS ONE* 12(10):1–24. <https://doi.org/10.1371/journal.pone.0185858>
- Taylor D, Myers SA, Clauset A, Porter MA, Mucha PJ (2017) Eigenvector-based centrality measures for temporal networks. *Multiscale Model Simul* 15(1):537–574. <https://doi.org/10.1137/16M1066142>
- Taylor D, Porter MA, Mucha PJ (2019) Tunable eigenvector-based centralities for multiplex and temporal. *Networks* 1904:02059
- Valdano E, Ferreri L, Poletto C, Colizza V (2015) Analytical computation of the epidemic threshold on temporal networks. *Phys Rev* 5(2):021005. <https://doi.org/10.1103/PhysRevX.5.021005.1406.4815>
- Valdano E, Poletto C, Giovannini A, Palma D, Savini L, Colizza V (2015) Predicting epidemic risk from past temporal contact data. *PLoS Comput Biol* 11(3):1004152. <https://doi.org/10.1371/journal.pcbi.1004152.g001>
- Valdano E, Fiorentin MR, Poletto C, Colizza V (2018) Epidemic threshold in continuous-time evolving networks. *Phys Rev Lett* 120(6):068302. <https://doi.org/10.1103/PhysRevLett.120.068302>
- Vidondo B, Voelkl B (2018) Dynamic network measures reveal the impact of cattle markets and alpine summering on the risk of epidemic outbreaks in the Swiss cattle population. *BMC Veterin Res* 14(1):1–11. <https://doi.org/10.1186/s12917-018-1406-3>
- Webber W, Moffat A, Zobel J (2010) A similarity measure for indefinite rankings. *ACM Trans Inf Syst* 28(4):1–38. <https://doi.org/10.1145/1852102.1852106>