

METHODOLOGY

Open Access



# Temporal correlation coefficient for directed networks

Kathrin Büttner<sup>\*</sup>, Jennifer Salau and Joachim Krieter

<sup>\*</sup>Correspondence:  
kbuettner@tierzucht.uni-kiel.de  
Institute of Animal Breeding and Husbandry, Christian-Albrechts-University, Olshausenstr. 40, 24098 Kiel, Germany

## Abstract

Previous studies dealing with network theory focused mainly on the static aggregation of edges over specific time window lengths. Thus, most of the dynamic information gets lost. To assess the quality of such a static aggregation the temporal correlation coefficient can be calculated. It measures the overall possibility for an edge to persist between two consecutive snapshots. Up to now, this measure is only defined for undirected networks. Therefore, we introduce the adaption of the temporal correlation coefficient to directed networks. This new methodology enables the distinction between ingoing and outgoing edges. Besides a small example network presenting the single calculation steps, we also calculated the proposed measurements for a real pig trade network to emphasize the importance of considering the edge direction. The farm types at the beginning of the pork supply chain showed clearly higher values for the outgoing temporal correlation coefficient compared to the farm types at the end of the pork supply chain. These farm types showed higher values for the ingoing temporal correlation coefficient. The temporal correlation coefficient is a valuable tool to understand the structural dynamics of these systems, as it assesses the consistency of the edge configuration. The adaption of this measure for directed networks may help to preserve meaningful additional information about the investigated network that might get lost if the edge directions are ignored.

**Keywords:** Temporal network, Temporal correlation coefficient, Directed network, Topological overlap

## Background

Network theory has become a valuable framework in many different research areas, whenever the system under investigation can be described as a graph, thus a set of nodes and edges connecting these nodes. For instance, social contacts of individuals (Kasper and Voelkl 2009; Krause et al. 2007; Lewis et al. 2008; Makagon et al. 2012), disease transmission (Eames and Read 2008; Eubank 2005; Heckathorn et al. 1999), trade networks (Büttner et al. 2013, 2015; Guimerà et al. 2005; Kaluza et al. 2010; Nöremark et al. 2011), the World Wide Web and the internet (Albert et al. 1999; Barabási et al. 2000; Cohen et al. 2000) or citation networks (Newman 2001a, b), to name but a few.

These previous studies dealing with network analysis focused mainly on static network analysis, meaning an edge was drawn between a pair of nodes whenever there was a contact between these two nodes during the whole observation period.

A static network  $G = (V, E)$ , where  $V$  is the set of nodes and  $E$  is the set of edges, can be illustrated as so-called adjacency matrix  $(a_{ij})_{ij}$  with  $a_{ij} = 1$  if there is an edge between nodes  $i$  and  $j$ , and  $a_{ij} = 0$  otherwise. Thus, the temporal information is neglected. However, in order to avoid losing all temporal information of the system, one possible approach is to separate the observation window in smaller parts and aggregate the contacts only over the formed snapshots, which are then analysed separately (examples are Bajardi et al. 2011; Büttner et al. 2015; Dubé et al. 2011; Nöremark et al. 2011; Rautureau et al. 2012; Vernon and Keeling 2009). Therefore, not all temporal information gets lost.

Although the static network analysis neglects partly the temporal variation in the system under investigation, one of its big advantages lies in the huge toolbox of methods that has been developed in the last decades. They range from network parameters describing the whole network topology (e.g. diameter, degree distribution, density, fragmentation index) to centrality parameters which allow for node ranking (e.g. in- and out-degree, closeness centrality, betweenness centrality) (Newman 2010; Wasserman and Faust 1994).

Depending on the system under investigation, the static network analysis can capture its temporal dynamics sufficiently well, but if the temporal variation in the system is too high the static network analysis is unable to display this variation. This leads us to the so-called temporal networks (Holme and Saramäki 2012; Kempe et al. 2002). However, the analysis of temporal networks is an interdisciplinary field which is still in its infancy and therefore, the analytical and computational methods are still at an early stage of development (Masuda and Holme 2013).

Therefore, to assess the quality of the static aggregation compared to the temporal counterpart some measures have been developed. For example, the causal fidelity measures the fraction of the number of paths in a static network, which can also be taken in the temporal counterpart (Lentz et al. 2013). Another example is the temporal correlation coefficient. It is a measure of the overall average probability for an edge to persist across two consecutive snapshots and can be used for the quality assessment of the static aggregation (Nicosia et al. 2013; Tang et al. 2010). In the first case, the measure can be used for both undirected and directed networks, i.e. the edge direction is taken into account resulting in an asymmetric adjacency matrix. However, up to now, the temporal correlation coefficient is only defined for undirected networks. Taking the edge direction into account is of special importance for many research questions. For example, in behavioural sciences, it is important to know which individual is the active part and which is the passive part of an interaction or, in trade networks, which node is the supplier and which is the purchaser. Thus, working with directed networks provides more information and reveals further insights into the interaction between the individual nodes.

Therefore, we introduce in the present paper an adaption of the temporal correlation coefficient first presented by Tang et al. (2010) and Nicosia et al. (2013), that was modified by Pigott and Herrera (2014) and Büttner et al. (2016), to directed networks.

## Methods

The first part of the materials and methods section deals with the temporal correlation coefficient for undirected networks. Here, a detailed description of the individual calculation steps is entailed. In the second paragraph, the adaption of this measure for

directed networks is presented. In the last paragraph of the materials and methods section, both the undirected and the directed calculation of the temporal correlation coefficient are carried out on a real-word pig trade network of a producer community in Northern Germany. This comparison clarifies the importance of a distinction between the undirected and the directed temporal correlation coefficient.

**Temporal correlation coefficient**

The temporal correlation coefficient  $C$  measures the overall average probability for an edge to persist across two consecutive snapshots (Nicosia et al. 2013; Tang et al. 2010). The calculation of  $C$  is divided into three individual calculation steps.

Firstly, for all nodes  $i = 1, \dots, N$ , where  $N$  is the total number of nodes in the network, and all snapshots  $t_m$ , with  $m = 1, \dots, M - 1$ , where  $M$  is the total number of considered snapshots, the topological overlap  $C_i(t_m, t_{m+1})$  of node  $i$  between two consecutive snapshots  $t_m$  and  $t_{m+1}$  is calculated with the following formula:

$$C_i(t_m, t_{m+1}) = \frac{\sum_j a_{ij}(t_m)a_{ij}(t_{m+1})}{\sqrt{[\sum_j a_{ij}(t_m)][\sum_j a_{ij}(t_{m+1})]}}, \tag{1}$$

where  $a_{ij}$  denotes an entry in the unweighted adjacency matrix of the graph. Therefore, summing over  $a_{ij}$  illustrates the interaction between  $i$  and every other node for the two consecutive snapshots  $t_m$  and  $t_{m+1}$ .

Secondly, the average topological overlap of the graph  $C_m$  for two consecutive snapshots  $t_m$  and  $t_{m+1}$  is determined. According to the proposed adaption of the temporal correlation coefficient by Büttner et al. (2016),  $C_m$  is calculated as follows,

$$C_m = \frac{1}{\max[A(t_m), A(t_{m+1})]} \sum_{i=1}^N C_i(t_m, t_{m+1}), \tag{2}$$

where  $\max[A(t_m), A(t_{m+1})]$  denotes the maximum number of active nodes of the graph at  $t_m$  and  $t_{m+1}$ . A node  $i$  is called “active” at time  $t_m$ , if there exists a node  $j \neq i$  and an edge between  $i$  and  $j$  in the graph at  $t_m$ , i.e. node  $i$  has a degree greater than zero.

Despite the calculation of  $C_m$ , the average topological overlap of the nodes  $C_i$  for all snapshots can be calculated as follows,

$$C_i = \frac{1}{M - 1} \sum_{m=1}^{M-1} C_i(t_m, t_{m+1}). \tag{3}$$

In the third calculation step, summing up all results for the topological overlap gives the temporal correlation coefficient  $C$  of the network:

$$C = \frac{1}{M - 1} \sum_{m=1}^{M-1} C_m \tag{4}$$

The range of the values of all three calculation steps is between 0 and 1. One indicates that in the observed snapshots the edge configuration is identical, whereas zero means that none of the same edges are common in the observed snapshots.

Figure 1 depicts an example network consisting of 4 temporal snapshots with directional information given by the arrow tips of the edges. In the Additional file 1 the single calculation steps of the undirected temporal correlation coefficient  $C$  are explained in detail.

**Adaption of the temporal correlation coefficient to directed networks**

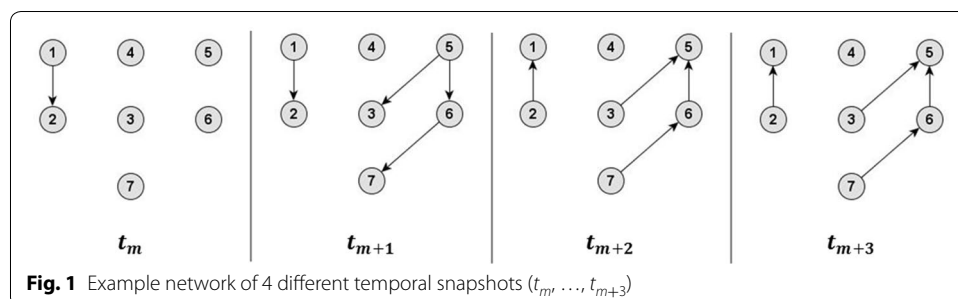
In directed networks a distinction between ingoing and outgoing edges is made. In undirected networks an edge from node  $i$  to node  $j$ —corresponding to  $a_{ij} = 1$ —is additionally considered as an edge from node  $j$  to node  $i$ , which implies  $a_{ji} = 1$ . Therefore, the adjacency matrices of undirected networks are symmetrical. This is no longer the case for directed networks, as there could be an edge from node  $i$  to node  $j$ , although no edge from node  $j$  to node  $i$  exists, which implies  $1 = a_{ij} \neq a_{ji} = 0$ . Figure 2 shows an example for the differences between the undirected and the directed representation of a network with regard to its adjacency matrices.

Therefore, the temporal correlation coefficient in directed networks should be calculated for the configuration of the ingoing edges (hereinafter named as ingoing temporal correlation coefficient  $C^{in}$ ) and for the configuration of the outgoing edges (hereinafter named as outgoing temporal correlation coefficient  $C^{out}$ ). Due to the fact that the maximum number of active nodes  $\max[A(t_m), A(t_{m+1})]$  is used to calculate the temporal correlation coefficient in undirected networks, where  $A(t_m)$  is the number of nodes with nonzero degree in the snapshot  $t_m$ , this has to be adapted while dealing with directed networks. In the calculation of  $C^{in}$  and  $C^{out}$ ,  $A(t_m)$  will be replaced by the number of nodes with nonzero in-degree  $A^{in}(t_m)$  and the number of nodes with nonzero out-degree  $A^{out}(t_m)$ , respectively.

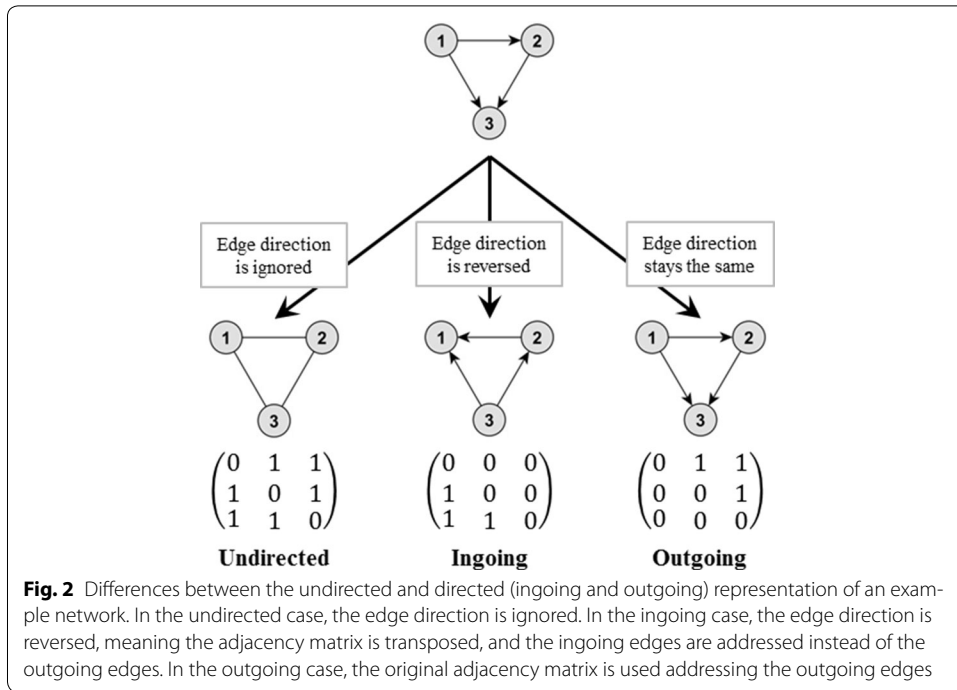
**Ingoing temporal correlation coefficient  $C^{in}$**

For the calculation of the ingoing temporal correlation coefficient  $C^{in}$ , Eq. (1) is used with the transposed adjacency matrix to focus on the ingoing edges (Fig. 2). The values  $C_i^{in}(t_m, t_{m+1})$  from this calculation step are then used in Eq. (5). In contrast to Eq. (2), instead of  $\max[A(t_m), A(t_{m+1})]$ ,  $\max[A^{in}(t_m), A^{in}(t_{m+1})]$  is used for the calculation of  $C_m^{in}$ :

$$C_m^{in} = \frac{1}{\max[A^{in}(t_m), A^{in}(t_{m+1})]} \sum_{i=1}^N C_i^{in}(t_m, t_{m+1}) \tag{5}$$



**Fig. 1** Example network of 4 different temporal snapshots ( $t_m, \dots, t_{m+3}$ )



In addition to  $C_m^{in}$  and similarly to Eq. (3), the average topological overlap of the nodes for all possible snapshots can also be calculated for the ingoing edges  $C_i^{in}$  and is calculated as follows,

$$C_i^{in} = \frac{1}{M-1} \sum_{m=1}^{M-1} C_i^{in}(t_m, t_{m+1}) \tag{6}$$

For the last calculation step no changes in Eq. (4) are carried out.

A detailed description of the single calculation steps for the ingoing temporal correlation coefficient  $C^{in}$  for the example network (Fig. 1) is illustrated in the Additional file 1.

**Outgoing temporal correlation coefficient  $C^{out}$**

For the calculation of the outgoing temporal correlation coefficient, the first calculation step (Eq. 1) stays the same, as the outgoing edges are represented by the untransposed adjacency matrix (Fig. 2). The values  $C_i^{out}(t_m, t_{m+1})$  are then used in Eq. (7). Here,  $\max[A^{out}(t_m), A^{out}(t_{m+1})]$  is used for the calculation of  $C_m^{out}$ :

$$C_m^{out} = \frac{1}{\max[A^{out}(t_m), A^{out}(t_{m+1})]} \sum_{i=1}^N C_i^{out}(t_m, t_{m+1}) \tag{7}$$

Additionally, the average topological overlap of the nodes for all possible snapshots for the outgoing case  $C_i^{out}$  can be calculated as follows,

$$C_i^{out} = \frac{1}{M-1} \sum_{m=1}^{M-1} C_i^{out}(t_m, t_{m+1}) \tag{8}$$

For the last calculation step no changes in Eq. (4) are carried out.

A detailed description of the single calculation steps for the outgoing temporal correlation coefficient  $C^{out}$  for the example network (Fig. 1) is illustrated in the Additional file 1.

#### **Convergence behaviour of the temporal correlation coefficient ( $C$ , $C^{in}$ and $C^{out}$ )**

As the topological overlap is a probability, it is expected that for all  $m = 1, \dots, M - 1$  the average topological overlap  $C_m$  in the undirected case, as well as  $C_m^{in}$  and  $C_m^{out}$  for the ingoing and outgoing case, respectively, are in the range of 0–1. The values for  $C_m$  between two consecutive snapshots equal 1, only if these snapshots show identical edge configuration. To reveal a convergence behaviour of the temporal correlation coefficients ( $C$ ,  $C^{in}$  and  $C^{out}$ ), an identical extension of the time series was generated by attaching the last snapshot, i.e. the graph at  $t_M$  of the example network of Fig. 1, repeatedly to the existing dynamic network until the length of this series of networks equalled 100. In Büttner et al. (2016) it has already been illustrated that the corresponding series of undirected temporal correlation coefficients  $C$ , calculated with the method proposed there, converges towards 1. The series of directed ingoing and outgoing temporal correlation coefficients corresponding to the series of identical extensions of the example network of Fig. 1 were calculated, and their convergence behaviour was plotted and analysed.

#### **Temporal correlation coefficient ( $C$ , $C^{in}$ and $C^{out}$ ) calculated for a pig trade network**

##### **Data basis**

From June 2006 to May 2009, pig movement data from a producer community in Northern Germany were recorded. This corresponds to an observation window length of 1096 days. The data contained the date of the movement, the supplier, the purchaser as well as the batch size and the type and age group of the delivered animals. In total, the data comprised 4635 animal movements between 483 farms which could be categorized in 29 multipliers, 34 farrowing farms, 153 finishing farms and 267 farrow-to-finishing farms. Due to the dead-end characteristic of the abattoirs they were excluded from the network analysis.

##### **Network construction**

In this pig trade network, the farms illustrate the nodes of the network and the trade contacts between the farms represent the edges. Networks with increasing time window lengths (1–548 days) were constructed in order to evaluate how the chosen time window length may affect the outcome of the temporal correlation coefficient. This means that for a time window length of 1 day 1096 single snapshots of the network were created. If the time window length is doubled to 2 days, the number of snapshots halves to 548. And finally for a time window length of 548 days, only 2 single snapshots were generated. Due to the predefined length of the observation window with 1096 days, the last snapshot may contain less days than the previous snapshots. This was the case if the length of the time window was not a proper divisor of 1096 which corresponds to the length of the whole observation period. These incomplete snapshots were excluded from the further analysis.

### Frequency distributions

In order to get more information about the frequency distribution of the average topological overlap of the nodes ( $C_i$ ,  $C_i^{in}$  and  $C_i^{out}$ ) the values were categorized for each time window length and each farm type into the following 6 categories: 0:  $C_i = 0$ , 1:  $0 < C_i \leq 0.2$ , 2:  $0.2 < C_i \leq 0.4$ , 3:  $0.4 < C_i \leq 0.6$ , 4:  $0.6 < C_i \leq 0.8$ , 5:  $0.8 < C_i \leq 1$ . Then the percentage of each category was calculated and plotted as a grouped bar chart.

### Statistical analysis

All statistical analyses were performed using SAS<sup>®</sup> statistical software package (SAS<sup>®</sup> Inst. Inc. 2008). The adaption of the temporal correlation coefficient was carried out with the programming language Python (van Rossum and Drake 2001).

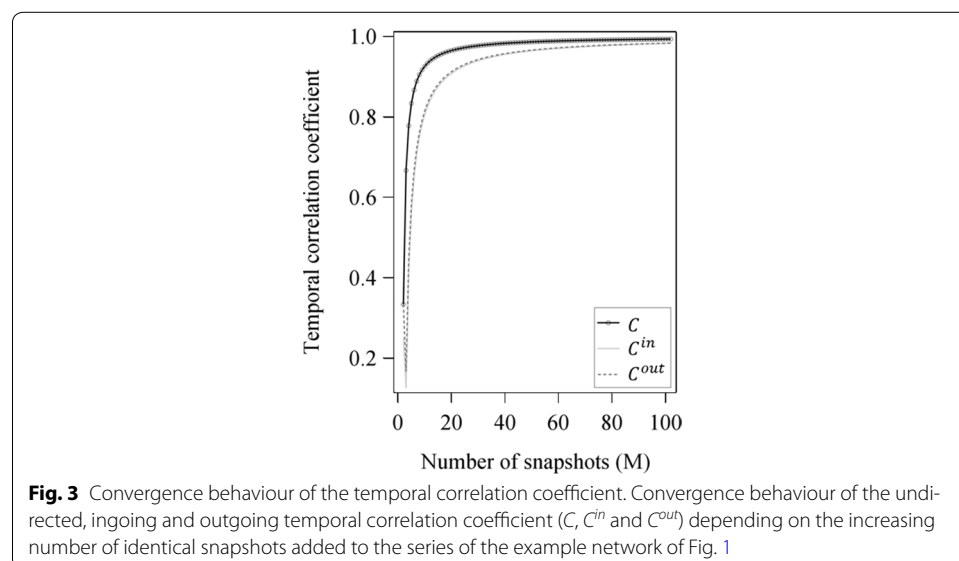
## Results

### Convergence behaviour of the undirected, ingoing and outgoing temporal correlation coefficient ( $C$ , $C^{in}$ and $C^{out}$ )

Figure 3 shows the convergence behaviour of the temporal correlation coefficient for the undirected and the directed (ingoing and outgoing) calculation for the example network of Fig. 1 depending on the increasing number of added identical snapshots. Both  $C^{in}$  and  $C^{out}$  showed lower values than  $C$ . However, with increasing number of snapshots added to the time series, all three possible calculations converged to 1 from below.

### Temporal correlation coefficient ( $C$ , $C^{in}$ and $C^{out}$ ) calculated for a pig trade network

For the pork supply chain under investigation, the highest mean value over all possible time window lengths ( $n = 548$ ) could be obtained for  $C^{in}$  with  $0.50 \pm 0.11$ , followed by  $C$  with  $0.47 \pm 0.11$ . The lowest mean value showed  $C^{out}$  with  $0.36 \pm 0.10$ . These results could be confirmed by Fig. 4, where  $C$  and  $C^{in}$  showed a quite similar course, whereas  $C^{out}$  had clearly lower values. Despite this, all three measures increased rapidly from close to 0 to their maximum values from the time window of length 1 to a window

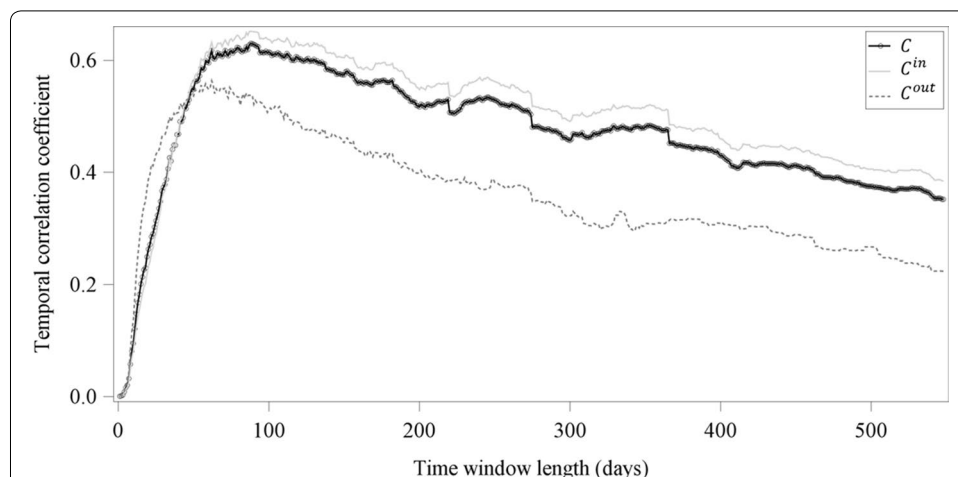


length of 62 days for  $C^{out}$  (max = 0.56) and 88 days for both  $C$  (max = 0.63) and  $C^{in}$  (max = 0.65). After reaching their maximum, the values decreased slowly with increasing length of the time window to 0.38 for  $C^{in}$ , 0.35 for  $C$  and 0.22 for  $C^{out}$ .

**Average topological overlap of the nodes ( $C_i$ ,  $C_i^{in}$  and  $C_i^{out}$ ) calculated for a pig trade network**

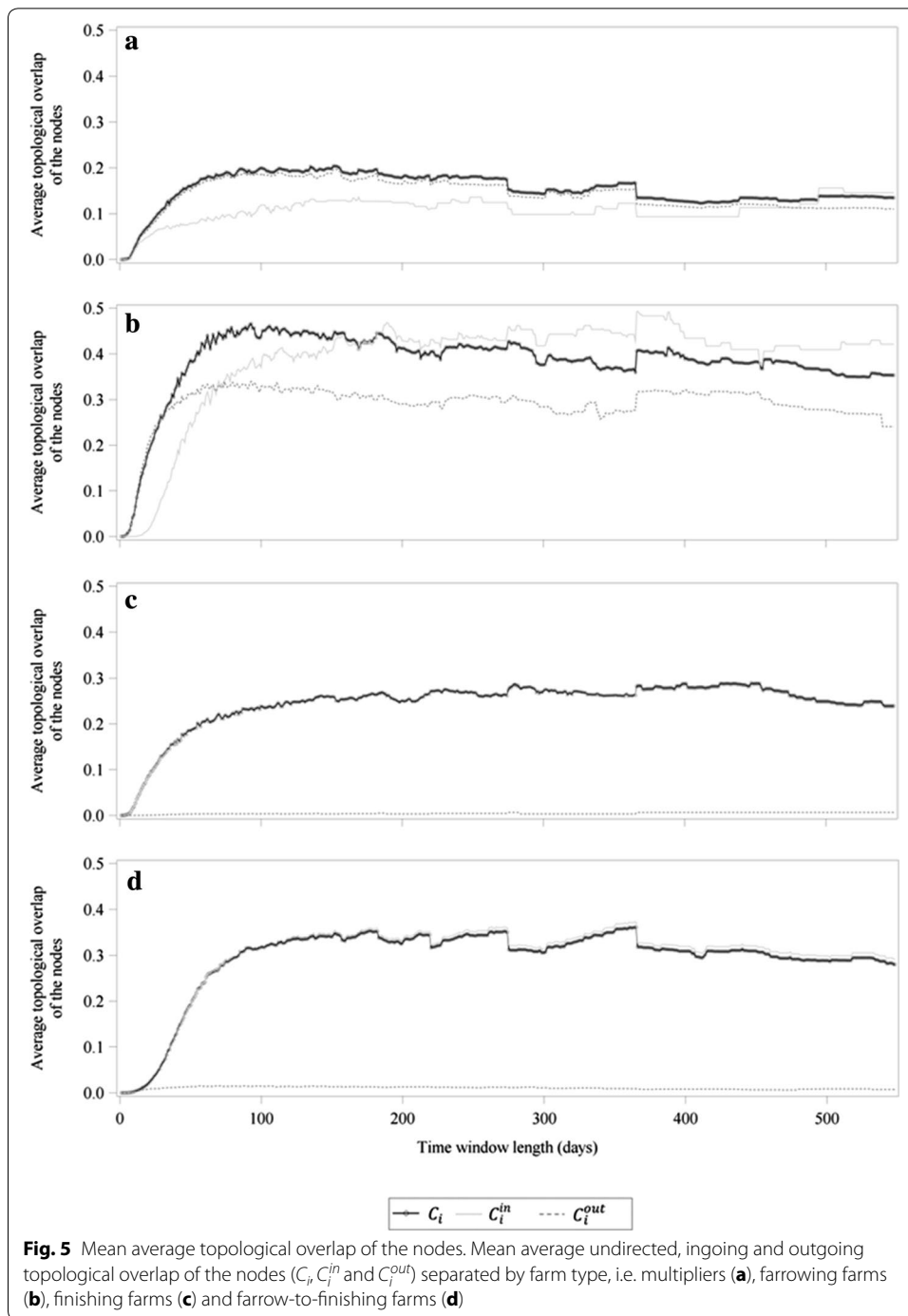
Figure 5 shows the mean average topological overlap of the nodes ( $C_i$ ,  $C_i^{in}$  and  $C_i^{out}$ ) separated by the farm types multipliers (Fig. 5a), farrowing farms (Fig. 5b), finishing farms (Fig. 5c) and farrow-to-finishing farms (Fig. 5d). The finishing farms and the farrow-to-finishing farms showed for the mean  $C_i$  and  $C_i^{in}$  relatively similar values, whereas the mean values of  $C_i^{out}$  equalled for almost all time window lengths 0. Contrarily, multipliers and farrowing farms had a more differentiated course of  $C_i$ ,  $C_i^{in}$  and  $C_i^{out}$ . For multipliers the mean values of  $C_i^{out}$  showed quite similar values compared to  $C_i$ , whereas the mean values of  $C_i^{in}$  were slightly lower than the values of  $C_i$  and  $C_i^{out}$  until a time window length of about 500 days, where  $C_i^{out}$  became larger than the other two measurements. For farrowing farms the course of  $C_i^{in}$  showed a smaller increase than  $C_i$  until a time window length of 180 days where  $C_i^{in}$  became larger than  $C_i$ .  $C_i^{out}$  had a quite similar course than  $C_i$  but moved down by about 0.10–0.15. Table 1 illustrates the mean average topological overlap of the nodes ( $C_i$ ,  $C_i^{in}$  and  $C_i^{out}$ ) separated by the farm types for the time window lengths of 62 and 88 days and gives therefore a detailed section of the values illustrated in Fig. 5 for the time window length with the maximum values for  $C$ ,  $C^{in}$  and  $C^{out}$ .

To get a better insight in the frequency distribution of  $C_i$ ,  $C_i^{in}$  and  $C_i^{out}$  depending on the chosen time window length and farm type, the values of  $C_i$ ,  $C_i^{in}$  and  $C_i^{out}$  were grouped into 6 different categories as illustrated in Figs. 6, 7 and 8. The first visual inspection reveals that, regarding  $C_i$  (Fig. 6) and  $C_i^{out}$  (Fig. 8), the distribution of the categories for multipliers and farrowing farms were very similar, whereas they differed strongly when these two measures are compared for finishing farms and farrow-to-finishing



**Fig. 4** Course of the temporal correlation coefficient for different time window lengths. Undirected, ingoing and outgoing temporal correlation coefficient ( $C$ ,  $C^{in}$  and  $C^{out}$ ) of the pork supply chain of a producer community in Northern Germany for different time windows lengths





farms. In contrast to this, the frequency distribution of  $C_i$  (Fig. 6) and  $C_i^{in}$  (Fig. 7) was nearly identical for finishing farms and farrow-to-finishing farms and at least similar for multipliers and farrowing farms. To specify some particular observations, Fig. 6b illustrating the undirected case of the farrowing farms held the smallest percentage of  $C_i$  values equal to 0 (category 0), not only in comparison between the farm types but also compared to the frequency distribution diagrams related to the directed case (Figs. 7b,

**Table 1 Mean (range) average topological overlap of the nodes**

Farm type	n	Mean (range) average topological overlap of the nodes		
		Undirected ( $C_i$ )	Ingoing ( $C_i^{in}$ )	Outgoing ( $C_i^{out}$ )
<i>Time window length of 62 days</i>				
Multiplier	29	0.18 (0–0.8)	0.09 (0–1)	0.17 (0–0.87)
Farrowing farm	34	0.44 (0–0.86)	0.32 (0–1)	0.33 (0–0.93)
Finishing farm	153	0.21 (0–1)	0.21 (0–1)	0 (0–0.56)
Farrow-to-finishing farm	267	0.26 (0–1)	0.26 (0–1)	0.01 (0–0.97)
<i>Time window length of 88 days</i>				
Multiplier	29	0.19 (0–0.82)	0.1 (0–1)	0.19 (0–0.82)
Farrowing farm	34	0.45 (0–0.89)	0.37 (0–1)	0.33 (0–0.91)
Finishing farm	153	0.23 (0–1)	0.23 (0–1)	0 (0–0.55)
Farrow-to-finishing farm	267	0.31 (0–1)	0.31 (0–1)	0.02 (0–0.96)

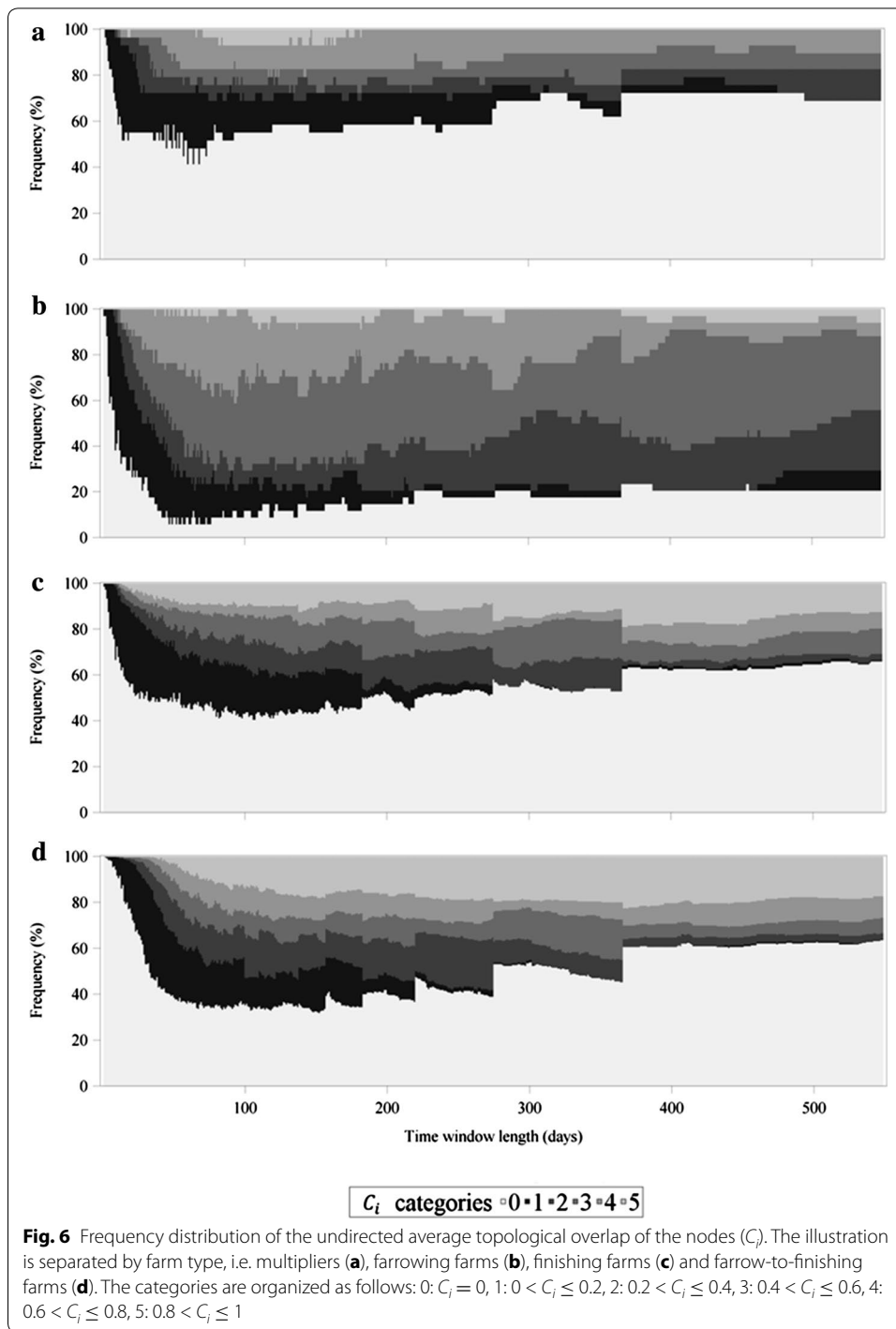
Mean (range) average undirected, ingoing and outgoing topological overlap of the nodes ( $C_i$ ,  $C_i^{in}$  and  $C_i^{out}$ ) separated by farm type for time window lengths of 62 and 88 days (corresponding to the maximum values of the temporal correlation coefficients  $C$ ,  $C^{in}$  and  $C^{out}$ )

8b). Furthermore, farrowing farms showed a higher percentage in the categories 4 and 5 in the undirected case compared to the other farm types. In contrast to that, finishing farms and farrow-to-finishing farms had, despite category 0, the majority of the farms in category 6 with the highest  $C_i$  values. Generally, the ratios between the higher categories differed between undirected, ingoing and outgoing cases. The frequency distribution of  $C_i$ ,  $C_i^{in}$  and  $C_i^{out}$  for the multipliers exemplarily illustrates that, in the undirected and the outgoing cases (Figs. 6a, 8a), category 5 occupied a larger area than category 6, but the frequency distribution regarding  $C_i^{in}$  (Fig. 7a) showed a larger area of category 6 than category 5. Looking at the union of nonzero values for temporal correlation coefficients of the nodes in the ingoing case, the percentage of accumulated categories 1–6, i.e.  $C_i^{in} > 0$ , was largest for time window lengths between 70 and 110 days independent of the chosen farm type, whereby the course was not that pronounced for multipliers (Fig. 7a) compared to other farm types (Fig. 7b–d). From Fig. 8 it becomes obvious that almost exclusively the farm types at the beginning of the pork supply chain obtained values of  $C_i^{out} > 0$ . In contrast to the distribution of  $C_i^{in}$ , the largest percentages of categories 1–6 for  $C_i^{out}$  could be observed for the time window lengths of 50–80 days.

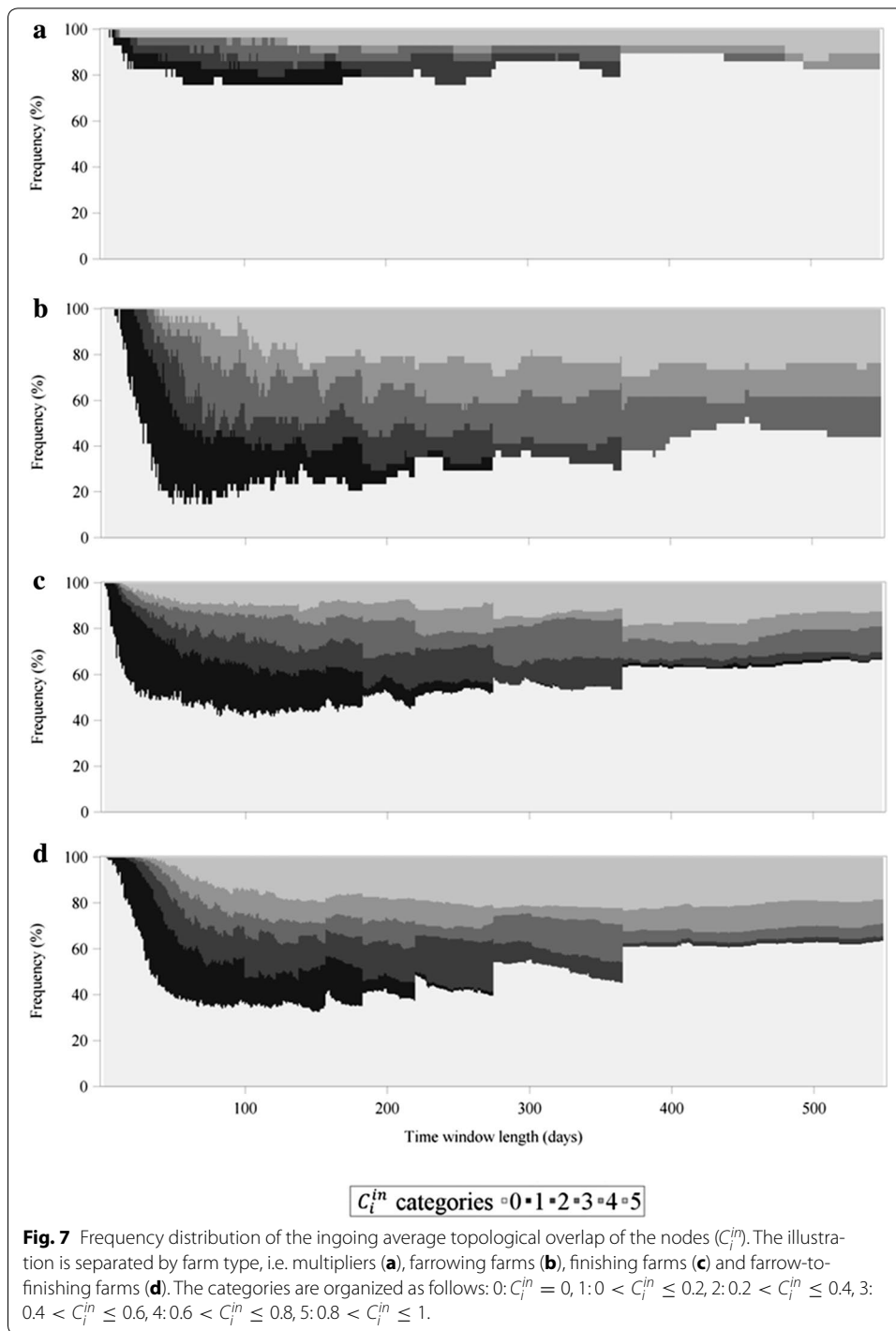
## Discussion

### Convergence behaviour of the undirected, ingoing and outgoing temporal correlation coefficient ( $C$ , $C^{in}$ and $C^{out}$ )

Due to the fact that the topological overlap represents the probability of an edge to persist between two consecutive snapshots and the temporal correlation coefficient is the average over all topological overlap values, both have to range between 0 and 1. In Fig. 3, the convergence behaviour of all three calculation methods (undirected, ingoing and outgoing) is illustrated for the example network of Fig. 1. All three calculation methods show the expected convergence against 1 with increasing number of added identical snapshots. The only difference for this example network can be seen between the

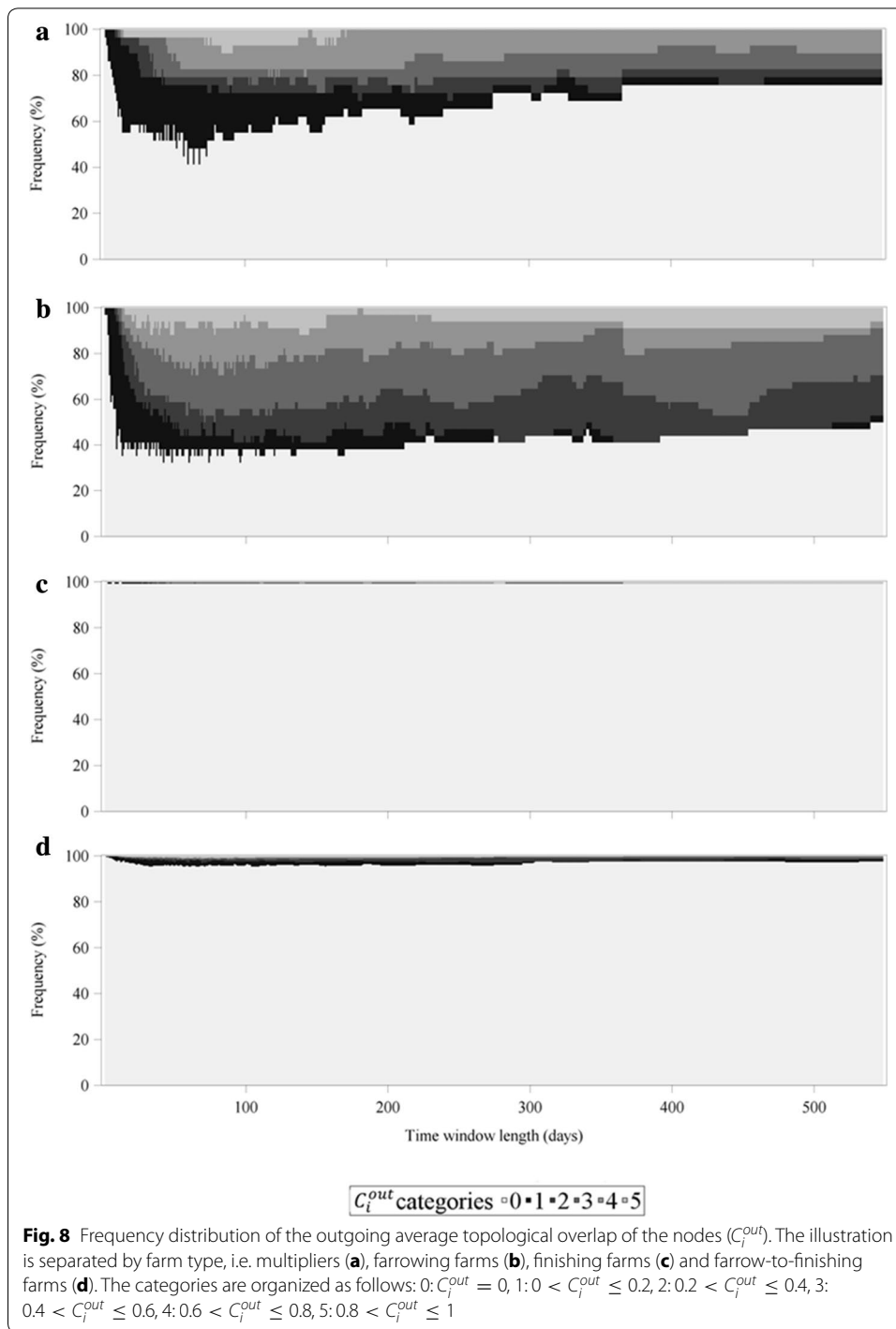


undirected and the directed case. Here, the undirected temporal correlation coefficient showed higher values compared to the directed cases. This can be explained by the fact that in the undirected case the edge direction is ignored and therefore the probability for an edge to persist between two consecutive snapshots doubles which lead to higher values.



**Temporal correlation coefficient ( $C$ ,  $C^{in}$  and  $C^{out}$ ) and average topological overlap of the nodes ( $C_i$ ,  $C_i^{in}$  and  $C_i^{out}$ ) calculated for a pig trade network**

The results of the temporal correlation coefficient showed for all three cases ( $C$ ,  $C^{in}$  and  $C^{out}$ ) that the choice of the time window length has an important impact on the edge persistence.  $C^{out}$  reached its maximum at a time window length of 62 days, indicating that



most of the outgoing trade contacts had periodic patterns on average every 2 months. In contrast, the maximum of both  $C$  and  $C^{in}$  was reached at a time window length of 88 days. This corresponds to periodic patterns of on average 3 months. However, Lentz et al. (2013) showed periodical patterns of 180 days, representing the characteristic time scale of a pork supply chain. This result is two to three times higher than the findings of

the present study. One explanation for these differences is that in the present study the abattoirs were excluded from the network analysis due to their dead-end characteristic for the transport of live animals. This also implies that a lot of farms, mainly the finishing and farrow-to-finishing farms, had no outgoing edges. This might be the main reason for the fact that  $C$  and  $C^{in}$  showed a similar course, whereas  $C^{out}$  could clearly be distinguished from them. Another point of consideration is that the fattening period including the trade contacts towards the abattoir, were also excluded from the network. This shortened the optimal time window length to 88 days for  $C$  and  $C^{in}$ , indicated by the maximum values of the temporal correlation coefficients. The optimal time window length of 62 days for  $C^{out}$  can be explained by the relative constant relationships between multipliers and farrowing or farrow-to-finishing farms, respectively. They mainly deliver or receive, respectively, gilts for replacement. Additionally, there are also constant trade relations between farrowing farms and finishing farms every 2 months if the finishing farms do not follow the all-in-all-out-procedure, but fatten their animals compartment specific. To protect their health status within the herds it is recommended to keep the number of suppliers as low as possible in order to reduce the possibility of a disease entry (Waldmann and Wendt 2004). Therefore, the outgoing trade contacts of the multipliers and the farrowing farms have relatively constant purchasers. Additionally, multipliers or farrowing farms got a relatively high mean  $C_i^{out}$ , whereas the mean  $C_i^{out}$  for the other farm types almost always equalled 0 (Fig. 5). However, these farm types only represent 13 % of the whole trade network. Due to the fact that for the calculation of the temporal correlation coefficients the edge configuration of all farms was taken into consideration, the values for  $C^{out}$  were lower than for  $C$  and  $C^{in}$ .

The results of the average topological overlap of the nodes (Fig. 5; Table 1) indicated that the different farm types showed clearly different courses of  $C_p$ ,  $C_i^{in}$  and  $C_i^{out}$ , depending on the position they occupy in the pork supply chain. Especially the farm types at the beginning of the pork supply chain, i.e. multipliers and farrowing farms had higher values of  $C_i^{out}$  compared to finishing farms and farrow-to-finishing farms. The frequency distributions are another way to illustrate the differences in farm types regarding their edge configurations via ingoing and outgoing temporal correlation coefficients of the nodes. They provide details on the orders of magnitude in addition to the means of  $C_p$ ,  $C_i^{in}$  and  $C_i^{out}$  (Figs. 6, 7, 8). Between the distributions of the categories of  $C_i$  and  $C_i^{out}$ , similarities for multipliers and farrowing farms but huge differences for finishing farms and farrow-to-finishing farms could be observed. The small amount of nonzero  $C_i^{out}$  values for finishing farms and farrow-to-finishing farms, i.e. the farm types at the end of the pork supply chain, can be explained by the exclusion of the abattoirs from the network analysis. Thus, these two farm types had almost no outgoing edges. To be more precise, only three finishing farms at all had outgoing edges. But these three farms could be classified as pure rearing farms which receive weaned piglets from a farrowing farm and delivered the piglets with about 30 kg body weight to another finishing farm. Therefore, they build a bridge between farrowing farms and the actual finishing farms with relatively stable trade contacts. However, all other finishing farms had no outgoing edges due to the exclusion of the abattoir. For a node without outgoing edges  $C_i^{out}$  cannot be computed and is therefore set to zero.

The distribution of the categories of  $C_i$  and  $C_i^{in}$  was nearly identical for finishing farms and farrow-to-finishing farms. In this case, no prominent differences for multipliers and farrowing farms could be observed. This indicates that the multipliers and the farrowing farms had ingoing edges with periodical patterns. Especially the farrowing farms received in constant time intervals gilts from the multipliers for their piglet production.

Looking at the frequency distributions of  $C_i$ ,  $C_i^{in}$  and  $C_i^{out}$  of the multipliers—as an example for the differing ratios in categories 1–6 between these measures—reveals a shift towards category 6 at the expense of mainly category 5 from  $C_i$ ,  $C_i^{out}$  compared to  $C_i^{in}$  (Figs. 6a, 7a, 8a). This can be explained by the fact that for the investigated pork supply chain most of the multipliers operated as hybrid breeding farms, whereas only a small amount could be assigned to nucleus breeding farms. The hybrid breeding farms received animals from the nucleus breeding farms in relatively constant intervals explaining the shift towards the higher amounts of  $C_i^{in}$  values to category 6.

Despite the differences between the undirected and the directed cases, all figures (Figs. 6, 7, 8) illustrate that category 0 was the most represented. The only exception is Fig. 6b representing the categorized values of  $C_i$  for the farrowing farms. In the presented pork supply chain, this farm type occupied a middle position between the multipliers on the one end and the finishing and farrow-to-finishing farms at the other end. This is the reason that for both  $C_i^{in}$  and  $C_i^{out}$ , and therefore also for  $C_i$  relatively high values for this farm type could be obtained.

#### Importance of the distinction between ingoing and outgoing movements

For the edge configuration of the pork supply chain under investigation a clear differentiation between the ingoing and outgoing movements could be observed. Especially the different farm types showed specific courses of  $C$ ,  $C^{in}$  and  $C^{out}$  as well as for  $C_i$ ,  $C_i^{in}$  and  $C_i^{out}$  depending on their position in the pork supply chain. These findings were confirmed by former studies of Büttner et al. (2013, 2015), where the centrality parameters also depend on the considered farm type. Especially the farm types at the beginning of the pork supply chain, i.e. multipliers and farrowing farms had significant higher values for the centrality parameters regarding the outgoing trade contacts compared to the farm types at the end of the pork supply chain, i.e. finishing and farrow-to-finishing farms. Furthermore, these studies showed that the rankings of the centrality parameters based on outgoing edges (e.g. out-degree, outgoing infection chain) had a more stable ranking than the centrality parameters based on the ingoing trade contacts (e.g. in-degree, ingoing infection chain). These findings gave evidence to the fact that there were some highly central farms delivering animals to the majority of this network. This results in a high value for the centrality parameters based on the outgoing trade contacts. However, it has to be considered that the values of the centrality parameters cannot directly be compared to the results of the temporal correlation coefficient. The reason for this is the fact that during the calculation of the centrality parameters only one node is focussed including the attached edges but it is unimportant which node stands on the other side of this edge. In contrast to this, during the calculation of the temporal correlation coefficient these pairs of nodes and especially the connection between two specific nodes are considered. Thus, an additional statement on stability of the connections between the nodes is made. This has to be kept in mind when talking about temporal correlation

coefficients. Despite the differences between the calculation of centrality parameters and the temporal correlation coefficient both can pass on more information if the direction of the edges is considered. The direction of the movements is of special importance for the analysis of disease spread within such directed networks. Here, a high number of ingoing movements increases the probability to get an infection, whereas a high number of outgoing movements means that the probability for spreading an infection increases. For the investigation of aggregated networks it is therefore very important to check the quality of this aggregation, in terms of high values for the temporal correlation coefficient. Furthermore, in order to keep the information of the directed network it is important to calculate not only the undirected temporal correlation coefficient  $C_i$  but also its directed counterparts  $C_i^{in}$  and  $C_i^{out}$  as more specific quality measures.

## Conclusions

The analysis of temporal networks as well as the investigation of their structural characteristics are still in their infancies. Thus, appropriate methods which help to examine how the structure of temporal networks may affect the dynamics of processes occurring on it, e.g. disease transmission, are still missing. Additionally, most of the newly described methods are initially based on undirected networks. However, especially if the direction of the edges is a main characteristic of the network, such as for the pig trade network, neglecting this property leads to an essential loss of information. Thus, the present paper provides methodologies to maintain this information by calculating the directed temporal correlation coefficients  $C_i^{in}$  and  $C_i^{out}$  representing an adaption of the temporal correlation coefficient to directed networks. Furthermore, it could be shown that depending on the position in the pork supply chain the edge configuration of specific farm types differentiated.

To conclude, considering the yet known dependencies and issues in dealing with the analysis of temporal network analysis, the temporal correlation coefficient is a valuable tool to understand the structural dynamics of these systems, as it assesses the consistency of the edge configuration. The adaption of this measure for directed networks may help to preserve meaningful additional information about the investigated network that might get lost, if the edge directions are ignored.

## Additional file

**Additional file 1.** Detailed calculation steps of the undirected, the ingoing and the outgoing temporal correlation coefficient ( $C$ ,  $C^{in}$ ,  $C^{out}$ ).

### Authors' contributions

KB, JS and JK designed the study. KB and JS participated in data analysis, developed the formula and drafted the manuscript. All authors read and approved the final manuscript.

### Acknowledgements

We gratefully acknowledge funding by the German Research Foundation (DFG).

### Competing interests

The authors declare that they have no competing interests.

Received: 22 February 2016 Accepted: 19 July 2016

Published online: 28 July 2016



## References

- Albert R, Jeong H, Barabási A (1999) Diameter of the World-Wide Web. *Nature* 401:130–131
- Bajardi P, Barrat A, Natale F, Savini L, Colizza V (2011) Dynamical patterns of cattle trade movements. *PLoS ONE* 6(5):e19869. doi:10.1371/journal.pone.0019869.t001
- Barabási A, Albert R, Jeong H (2000) Scale-free characteristics of random networks: the topology of the world-wide web. *Phys A* 281:69–77
- Büttner K, Krieter J, Traulsen A, Traulsen I (2013) Static network analysis of a pork supply chain in Northern Germany—characterisation of the potential spread of infectious diseases via animal movements. *Prev Vet Med* 110(3–4):418–428
- Büttner K, Krieter J, Traulsen I (2015) Characterization of contact structures for the spread of infectious diseases in a pork supply chain in northern Germany by dynamic network analysis of yearly and monthly networks. *Transbound Emerg Dis* 62(2):188–199
- Büttner K, Salau J, Krieter J (2016) Adaption of the temporal correlation coefficient calculation for temporal networks (applied to a real-world pig trade network). *SpringerPlus*. doi:10.1186/s40064-016-1811-7
- Cohen R, Erez K, Ben-Avraham D, Havlin S (2000) Resilience of the internet to random breakdowns. *Phys Rev Lett* 85(21):4626–4628
- Dubé C, Ribble C, Kelton D, McNab B (2011) Estimating potential epidemic size following introduction of a long-incubation disease in scale-free connected networks of milking-cow movements in Ontario, Canada. *Prev Vet Med* 99(2–4):102–111. doi:10.1016/j.prevetmed.2011.01.013
- Eames KTD, Read JM (2008) Networks in Epidemiology. In: Liò P, Yoneki E, Crowcroft J, Verma D (eds) *Bio-inspired computing and communication*, vol 5151. Springer, Berlin, pp 79–90
- Eubank S (2005) Network based models of infectious disease spread. *Jpn J Infect Dis* 58:59–S13
- Guimerà R, Mossa S, Turtschi A, Amaral LAN (2005) The worldwide air transportation network: anomalous centrality, community structure, and cities' global roles. *Proc Natl Acad Sci* 102:7794–7799
- Heckathorn DD, Broadhead RS, Anthony DL (1999) AIDS and social networks: HIV prevention through network mobilization. *Sociol Focus* 32(2):159–179
- Holme P, Saramäki J (2012) Temporal networks. *Phys Rep* 519(3):97–125
- Kaluza P, Kölzsch A, Gastner MT, Blasius B (2010) The complex network of global cargo ship movements. *J R Soc Interface* 7(48):1093–1103
- Kasper C, Voelkl B (2009) A social network analysis of primate groups. *Primates* 50(4):343–356
- Kempe D, Kleinberg J, Kumar A (2002) Connectivity and inference problems for temporal networks. *J Comput Syst Sci* 64(4):820–842
- Krause J, Croft DP, James R (2007) Social network theory in the behavioural sciences: potential applications. *Behav Ecol Sociobiol* 62(1):15–27
- Lentz HHK, Selhorst T, Sokolov IM (2013) Unfolding accessibility provides a macroscopic approach to temporal networks. *Phys Rev Lett* 110(11):118701
- Lewis K, Kaufman J, Gonzalez M, Wimmer A, Christakis N (2008) Tastes, ties, and time: a new social network dataset using Facebook.com. *Soc. Networks* 30(4):330–342
- Makagon MM, McCowan B, Mench JA (2012) How can social network analysis contribute to social behavior research in applied ethology? *Appl Anim Behav Sci* 138(3–4):152–161
- Masuda N, Holme P (2013) Predicting and controlling infectious disease epidemics using temporal networks. *F1000 Prime Rep* 5:6
- Newman ME (2001a) Scientific collaboration networks. I. Network construction and fundamental results. *Phys Rev E* 64(1 Pt 2):16131
- Newman ME (2001b) Scientific collaboration networks. II. Shortest paths, weighted networks, and centrality. *Phys Rev E* 64(1 Pt 2):16132
- Newman MEJ (2010) *Networks: an introduction*. Oxford University Press Inc., New York
- Nicosia V, Tang J, Mascolo C, Musolesi M, Russo G, Latora V (2013) Graph metrics for temporal networks. In: Holme P, Saramäki J (eds) *Temporal networks*. Springer, Berlin, pp 15–40
- 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. *Prev Vet Med* 99(2–4):78–90
- Pigott F, Herrera M (2014) Proposal for a correction to the temporal correlation coefficient calculation for temporal networks. arXiv preprint [arXiv:1403.1104](https://arxiv.org/abs/1403.1104)
- Rautureau S, Dufour B, Durand B (2012) Structural vulnerability of the French swine industry trade network to the spread of infectious diseases. *Animal* 6(7):1152–1162
- SAS® Inst. Inc. (2008) *User's guide (release 9.4)*, Cary, NC, USA
- Tang J, Scellato S, Musolesi M, Mascolo C, Latora V (2010) Small-world behavior in time-varying graphs. *Phys Rev E* 81:55101
- van Rossum G, Drake FL (2001) *Python reference manual (version 3.4.2)*. PythonLabs, Virginia
- Vernon MC, Keeling MJ (2009) Representing the UK's cattle herd as static and dynamic networks. *Proc R Soc B* 276(1656):469–476
- Waldmann K, Wendt M (eds) (2004) *Lehrbuch der Schweinekrankheiten*, 4th edn. Parey, Stuttgart
- Wasserman S, Faust K (1994) *Social network analysis: methods and applications*. Cambridge University Press, New York