scientific reports

Check for updates

OPEN

Impact of proximity definitions and sampling rates on social networks in pigs based on tracking using computer vision

Clémence A. E. M. Orsini[1✉], Bernadett Hegedűs[1,2], Lisette E. van der Zande[3], Inonge Reimert[1], Piter Bijma[2] & J. Elizabeth Bolhuis[1]

In farm animals, social network analysis has become a popular approach to explore preferential associations. This study investigated how different spatial proximity definitions and sampling rates affect social networks based on proximity using computer vision. Video data collected over three days in 21 pens (6 pigs/pen), either enriched or barren, were analyzed using a tracking-by-detection method based on bounding boxes. Networks were constructed with five different definitions of proximity: (1) distance between centroids of bounding boxes < 50 cm, (2) occurrence of overlap of surfaces of bounding boxes, (3) surface overlap of bounding boxes > 20%, (4) a combination of (1) and (3), and (5) the harmonic mean of the distance between the two individuals. For each proximity definition, networks built with downsampled data were compared to a network built with 0.5 frames per second. The network metric degree centrality was less affected by proximity definitions compared to eigenvector centrality and clustering coefficient. To maintain high correlations with the complete network (r > 0.90), downsampling should not go beyond 1 frame every 6 min. This work shows how computer vision data can be used for social network analysis in livestock with limited space and choice of social environment, and how metrics depend on proximity definitions and sampling rates.

Social network analysis (SNA) is a powerful method to quantify inter-individual interactions and identify the social roles of each individual within a group[1]. The use of this method has become popular for studying the social dynamics within animal groups or populations and its application has increased considerably in recent years (e.g.,[1–3]). Previous studies on wild animals have, for example, shown the potential of SNA to detect temporal behaviour changes[4] and to identify key individuals whose removal or introduction strongly impacts the dynamics of the group[1,5,6]. In addition, in zoo-housed populations, SNA can help to develop effective conservation strategies[7] and improve welfare by reducing social stress[8]. In a social network, the individual animals are represented as nodes and are connected by edges. The edges can be direct when there is a clear initiator or receiver (e.g., in aggressive interactions), or indirect (e.g., during communal resting)[9]. Centrality metrics (e.g., degree, betweenness, eigenvector or clustering coefficient) describe the position of each individual within the network, representing different facets of behaviour such as connectivity or influence roles[1]. General network metrics (e.g., density, fragmentation, or radius) provide information about the overall network structure, such as the cohesion or the brokerage, and enable the comparison between different groups of similar sizes[10,11].

Besides its application on wild and zoo animals, SNA is a promising method to gain knowledge about social interactions in farm animals, such as pigs (*Sus scrofa domesticus*), and to positively impact their welfare. A better comprehension of social interactions in pigs may open doors to new management or breeding strategies to improve their group functioning[12]. Changes in measures of centrality and cohesion could be used to predict the escalation of abnormal behaviours within groups of pigs. These changes could also help in decision-making to remove key individuals involved in the spread of these behaviours[1,13]. In addition, when pigs are transferred between production stages, they are subjected to repeated changes in group composition, which often destabilize the social structure and results in aggression[14]. SNA can guide management practices to reduce social stress and minimize the destabilizing of the social structure. For example, removing pigs with a high centrality from poorly interconnected groups is likely to have a negative impact by destabilizing the group, compared with removing animals from more cohesive groups[1]. SNA can also help to identify preferential affiliations and their drivers (e.g.,

[1]Adaptation Physiology Group, Wageningen University & Research, Wageningen, The Netherlands. [2]Animal Breeding and Genomics Group, Wageningen University & Research, Wageningen, The Netherlands. [3]Topigs Norsvin Research Centre, 's-Hertogenbosch, The Netherlands. ✉email: clemence.orsini@wur.nl

sex, age, relatedness, dominance status)[15–17] and manage groups to keep preferred affiliative partners together, as they may provide social buffering against stressful events[18]. Previous studies revealed that domestic pigs are highly social animals and may develop social preferences, expressed especially by close spatial proximity during rest[19–21]. The choice of lying partners has, therefore, been suggested as a potential indicator of affiliative preferences.

When performing SNA, researchers have to make a multitude of choices for data collection and analysis. Behavioural data can be collected in a variety of ways, including differences in the sampling method, the period of study, and the definition of the connections between individuals (i.e. edge definitions)[4]. Depending on the research question being addressed, the edges can be defined based on specific behavioural interactions (e.g., grooming, agonistic behaviours) or on spatial proximity. However, in the literature, there is no consensus on how spatial proximity is defined. In the context of lying, for example, Durell et al. defined spatial proximity as two growing pigs spending time in physical contact or in very close parallel contact[19], whereas Li et al. specified a threshold for proximity of 50% of the pig's body in contact with the other pig[22]. Goumon et al. considered growing pigs as lying in spatial proximity when they were within 30 cm of each other[20], while Jowett and Amory used a threshold of 1 m in sows[23]. The definition of these thresholds may depend also on the size of the animals, on the environment in which they are housed, on the stocking density, and on the behaviour that is considered (for example lying in close proximity, versus exploring in close proximity)[20].

Different spatial proximity definitions might significantly affect the network metrics, and the choice of a definition has therefore important implications for the conclusions that can be drawn from the network[9]. For example, using a smaller distance threshold to capture spatial proximity may result in a lower frequency of individuals captured as in close proximity, and, thus, a lower individual average degree and lower network density[4]. A few studies have tested the impact of different protocols on social networks, including the effect of the sampling methods[4,24], or different social interactions and proximity techniques[15,25]. These studies demonstrated that using different methods significantly influences the results of the SNA. However, to our knowledge, the impact of different proximity definitions on proximity networks in pigs has not been investigated so far.

Apart from proximity definitions, sampling rates vary widely across SNA studies in pigs, with data collection periods ranging from one day[26,27] to a few hours across several days[22,28,29]. For SNA based on proximity in pigs, some studies used instantaneous scan sampling, where observations are made by scanning videos at fixed time points (e.g., every 10 min[22,30]), while others record proximity continuously over a defined observation window (e.g., 30 min[31]). Most of the SNA studies are based on manual observations which are highly time-consuming and therefore limit sampling duration. However, large quantities of data across all possible interacting individuals are necessary to build a robust and accurate representation of the social network[4]. Insufficient sampling might result in missed connections between individuals and can have significant implications on social network structure[9].

Recent advances in sensing technologies provide the opportunity to considerably increase the sampling rate of the data collection for SNA. Using such technologies, spatial positions of pigs may be investigated continuously at an individual level and over a long period. Location sensors, such as Global Positioning Systems (GPS), have been widely used in wild animals to assess contact patterns and study social networks based on spatial proximity[32]. These sensors are particularly well adapted to study behaviours of animals kept in large outdoor areas compared with cameras which have only a limited field of view. In farm animals, previous studies investigated the structure of the network of animals raised in relatively large areas (outdoor, free stall barn) with location sensors such as GPS[15,33]. However, the application of these sensors in dense groups of farm animals kept in relatively small areas is challenging as the quality of the data collected is limited by the accuracy of the sensor[15,33], which might be insufficient to provide the correct distance between animals when all distances are relatively short (e.g., within 30 cm, as the threshold used in Goumon et al.[20]). In addition, such sensors are associated with high costs and often require regular maintenance and therefore are not optimal for applications in large-scale commercial farms[32].

As an alternative to GPS, the use of computer vision algorithms is a non-invasive and relatively low-cost method that allows the collection of data at high sampling rates by providing the accurate location of individuals in each frame. Previous studies showed promising results of computer vision algorithms to track individual farm animals, such as cows[34], laying hens[35], and pigs[36,37] in various environments. In pigs, these algorithms have been applied to measure individual activity[38]. By detecting the pig's location in each frame, they also give the opportunity to calculate proximity to pen mates and location preferences. In previous studies, social interactions between animals have been extracted from computer vision tracking algorithms in different ways, such as identifying individuals that simultaneously occupy the same region of interest[39] or using a combination of thresholds on close distance, orientation of the head and proximity duration[40]. However, the use of computer vision tracking algorithms to study social proximity has been limited and predominantly focused on small animals, such as mice and *Drosophila*. In addition, multiple options are also available here to define proximity, and therefore decisions need to be made on how to build the social networks. Furthermore, by giving information at the frame level, these algorithms generate a very large amount of data and may require a long computing time. Therefore, downsampling of camera frames is necessary to study social networks based on tracking algorithms, in order to reduce computing time and to make outputs accessible efficiently to researchers, breeders, or farmers.

Thus, computer vision algorithms allow for many proximity definitions and sampling rates. However, to the best of our knowledge, no research has investigated the effect of different proximity definitions and sampling rates on the social networks of livestock, including pigs, based on computer vision tracking. Therefore, this study aims to explore how different proximity definitions and sampling rates affect proximity social networks in inactive pigs based on tracking data using computer vision in two different housing conditions. We expect that both proximity definition and sampling rate crucially impact the structure of the network resulting in notable changes in the network metrics.

Material and methods

All methods complied with the European Directive 2010/63/EU and Dutch law on the protection of animals used for scientific purposes. The Animal Care and Use Committee of Wageningen University approved the experiment (AVD1040020186245). All methods applied in the study were performed in accordance with the ARRIVE guidelines and regulations[41]. No anesthesia was performed during the experiment and the pigs were not euthanized after use.

Animals and housing

The pigs used in this study were the same as those described in van der Zande et al.[37] and Parois et al.[42]. The aim of these studies was to study resilience in pigs housed in different housing conditions. Hereto, pigs were subjected to several challenges between nine and 21 weeks of age. Pigs were 15 weeks of age when they were used for the current study. Three weeks before that, they had been exposed to a 2-h transport challenge with their pen mates[42]. Briefly, the experiment was carried out on 126 crossbred boars and gilts, equally divided into 3 different batches, and housed at CARUS research facility of Wageningen University and Research (Wageningen, the Netherlands) from 9 weeks of age. The pigs were grouped into 21 pens (n = 6 pigs per pen, balanced in sex) and kept in two different housing systems: 9 barren pens and 12 enriched pens. The unequal distribution of the pens across the two housing systems was due to technical issues with the cameras in the barren pens. In the barren system, the pigs were housed in pens with a partly slatted floor (1.2 × 4.7 m). The enriched pens were double in size (2.4 × 4.7 m) and the floor was covered with sawdust, straw, and peat, as bedding material, and several types of objects (toys, jute bags, rope, egg boxes) or substrate (hay, or alfalfa) were provided according to an alternating schedule. The temperature within the pens was recorded via temperature sensors during the three days of data collection. The temperature reached 23 °C during the first two days and 22 °C during the third day. Artificial light was provided between 07:00 and 19:00 h.

Data collection and computer vision algorithm

All pens were equipped with an RGB camera (Velleman: 1 lx/f2.0) mounted above the pen (top-down view) and plugged into electrics. Videos were recorded with a frame rate of 25 frames per second (FPS) and a resolution of 352 by 288 pixels. Videos were collected and analyzed on three consecutive days at 15 weeks of age using the computer vision algorithm for object detection and tracking described by van der Zande et al. (2021)[37]. This algorithm was trained on 4,000 annotated images spread across different days and pens. The algorithm You Only Look Once version 3[43], combined with Simple Online Realtime Tracking[44] was used to detect and track the location of each pig by a bounding box. The tracking data was cleaned manually by removing false positive detections (detections of more than six individuals per pen) and the sequential bounding box ID was connected to the real ID of the animal. More details about false positive detections can be found in van der Zande et al. (2021)[37].

Calculation and inactivity definition

To calculate proximity definitions during inactivity, we first needed to define when the pigs were inactive. Hereto, the distance moved by each pig in the three days was calculated by comparing the centroid of the bounding boxes of the same individual in two consecutive frames sampled at one frame per minute. Equation (1) from the Pythagorean theorem was used, where x_t and y_t are the coordinates of the centroid of the bounding boxes at frame t, and x_{t+1} and y_{t+1} are the corresponding coordinates at frame $t + 1$. An individual was considered inactive when the $distance\,moved$ between frame t and frame $t + 1$ was below 20 cm. This simple threshold was selected based on preliminary examinations of frames to capture instances when pigs were changing lying positions, such as transitioning between lateral and ventral lying, while filtering out pigs in locomotion.

$$distance\,moved = \sqrt{(x_{t+1} - x_t)^2 + (y_{t+1} - y_t)^2} \tag{1}$$

Calculations of proximity definitions

The proximity between two pigs was only calculated when they were simultaneously inactive.

Distance between individuals

For each pair of pigs, the distance between the two individuals was calculated from the centroids of their bounding boxes within the same frame. The centroid of the bounding box corresponds to the point of intersection of its two diagonals. Equation (2) was used, where x_{id_1} and y_{id_1} are the coordinates of the centroid of the bounding box of pig id_1, and x_{id_2} and y_{id_2}, the corresponding measure for pig id_2.

$$distance_{id_1 id_2} = \sqrt{(x_{id_1} - x_{id_2})^2 + (y_{id_1} - y_{id_2})^2} \tag{2}$$

Surface of overlap of the bounding boxes

The surface of overlap of two bounding boxes within the same frame (*S*) was calculated from the coordinates of the bounding boxes (Supplementary Fig. S1). The percentage of the overlap of the surface (surface overlap) was calculated as the surface of overlap divided by the sum of the areas of the two bounding boxes using Eq. (3).

$$surface\,overlap = \frac{S}{S_1 + S_2} \times 100\% \tag{3}$$

Proximity definitions

Five different definitions of proximity were used to build the social networks (Table 1); four binary definitions based on the distance between the centroids and/or surface overlap of the bounding boxes, and a weighted proximity, being the inverse of the harmonic mean of the distances between two individuals. This approach was chosen, as it assigns more significance to differences in distance when pigs are in close proximity compared to when they are further apart. For instance, if there are three distance records for two pigs (0.5 m, 1 m, and 1.5 m), the mean distance is 1 m. However, the mean proximity is calculated as $(1/0.5 + 1/1 + 1/1.5)/3$, resulting in a value of 0.82. This value is smaller than the reciprocal of the mean distance ($1/1 = 1$), indicating that shorter distances (such as 0.5 m) have a more pronounced impact than longer distances (like 1.5 m). P_{weig}, the weighted proximity (cm) was calculated using Eq. (4), where $distance_{id_1 id_2}(k)$ is the distance between the centroids of the bounding boxes of pig id_1 and pig id_2 (calculated similarly to Eq. (2)) in the frame *k*, and *n* is the total number of frames in which the two individuals were simultaneously inactive. The thresholds of 20% and 50 cm were chosen by combining insights from existing literature (50 cm corresponding to the 30 cm as used in Goumon et al.[20], with an additional 20 cm to account for the centroid of the bounding box, rather than the edge of the pig) and through preliminary examination of frames (Fig. 3).

$$P_{weig} = \frac{1}{n} \sum_{k=1}^{n} \frac{1}{distance_{id_1 id_2}(k)} \tag{4}$$

Agreement between proximity definitions

The percentage of agreement was calculated for each pair of the binary proximity definitions and each pair of bounding boxes across time using Eq. (5). *Agreement* refers to the instances where pairs of bounding boxes were classified similarly by the two proximity definitions (either both in close proximity or both not in close proximity). *Disagreement* refers to the instances where one proximity definition categorized the bounding boxes as in close proximity while the other did not.

$$\% \, of \, agreement = \frac{Agreement}{Agreement + Disagreement} \times 100 \tag{5}$$

Social network analysis

Construction of matrices for social network analysis

A matrix of proximity was generated for each pen and proximity definition. We filtered this matrix to keep only the associations when the individuals were simultaneously inactive and in close proximity (according to each definition). The edges of each network (for the four binary definitions) were built with a half weighted index (HWI)[45] using Eq. (6), where $T_{prox_{id_1 id_2}}$ is the total time during which pig id_1 and pig id_2 were simultaneously inactive and in close proximity according to one of the definitions, T_{id1} the total time pig id_1 spent inactive, and T_{id2} the total time pig id_2 spent inactive. This correction was necessary to account for variation in activity between pigs from the same pen. Otherwise, more active pigs might have appeared more central in a network simply due to their increased activity level, rather than their actual social preferences.

$$HWI = \frac{T_{prox_{id_1 id_2}}}{(T_{id1} + T_{id2})/2} \tag{6}$$

Similarly, we calculated the HWI for the quantitative proximity definition (WEIG) using Eq. (7), where $T_{id_1 id_2}$ is the total time during which pig id_1 and pig id_2 were simultaneously inactive.

$$HWI = \frac{P_{weig} \times T_{id_1 id_2}}{(T_{id1} + T_{id2})/2} \tag{7}$$

The HWI in Eq. (6) represents the proportion of time that two pigs were simultaneously inactive and in close proximity, adjusted for their individual inactivity durations. Similarly in Eq. (7), the HWI represents the proportion of time that two pigs were simultaneously inactive, adjusted for both individual inactivity duration and the proximity, defined as the average distance between the two pigs.

Network name	Type of definition	Proximity definition
DIST	Binary	Distance between the centroids of the bounding boxes of pig id_1 and pig id_2 is within 50 cm
OVER	Binary	Presence (yes or no) of overlap between the bounding boxes of pig id_1 and pig id_2
SURF	Binary	Percentage of overlap of the surface of the bounding boxes of pig id_1 and pig id_2 is above 20%
DIST_SURF	Binary	Distance between the centroids of the bounding boxes of pig id_1 and pig id_2 is within 50 cm and percentage of overlap of the surface of their bounding boxes is above 20%
WEIG	Quantitative	Harmonic mean of the distance between the centroids of the bounding boxes of pig id_1 and pig id_2

Table 1. Proximity definitions used to construct the five different social networks.

Metric	Level	Direct/Indirect	Local/Global	Measure	Description
Degree centrality	Individual	Direct	Local	Prominence	Weighted number (sum of the half-weighted degree) of direct connections the focal node has to other nodes in the network
Eigenvector centrality	Individual	Indirect	Global	Prominence	Influence of a node based on its neighbours' connections. A node is more central when connected to highly connected nodes
Clustering coefficient	Individual	Indirect	Local	Cohesion	Tendency of the neighbours of a node to be interconnected
Radius	Group	Direct	Global	Cohesion	Minimum distance between any pair of nodes in the network, considering the longest path lengths

Table 2. Description of the network metrics. Definitions are adapted from Makagon et al.[1] and McKee et al.[48].
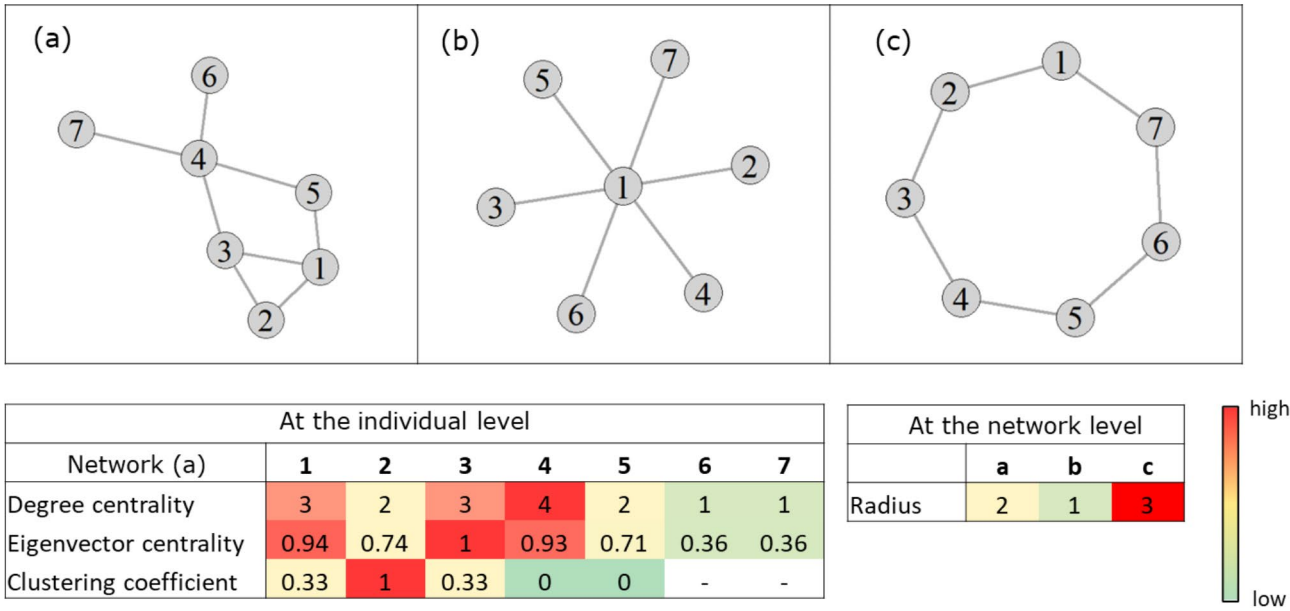
Fig. 1. Examples of networks (**a**–**c**) and their metrics. Individual level metrics are shown only for network a.

Construction of the complete and reduced networks

To estimate the effect of sampling rate on social networks, we compared a complete network to networks built with different sampling rates. For each proximity definition, the complete network was built using a frame rate of 0.5 FPS (corresponding to one frame every two seconds) to reduce potential noise due to micro-movements of the bounding boxes[38]. For each proximity definition, networks were built using different sampling rates, by using different sampling intervals (i.e. the interval between frames) ranging from one frame per 10 s to one frame per hour. From here on, we will refer to these networks as reduced networks, to distinguish them from the complete network (with one frame per 2 s).

Network visualization

Undirected social networks were visualized with R (version 4.1.2, R Foundation, Vienna, Austria), using the package "igraph"[46] where nodes represent individuals, and edges the HWI.

Network metrics

For each network, we calculated four different network metrics (Table 2 and Fig. 1) that measure different aspects of the network, both at individual and group level, based on direct and indirect connections. These metrics were selected based on their widespread use and established biological relevance[1]. They measure different aspects of the network and were therefore used as a representative case to illustrate the impact of proximity definitions and sampling rates on the network. Their characteristics are summarized in Table 2 and examples are given in Fig. 1. Direct/indirect relates to whether a metric measures properties using only immediate relationships (direct) or considers more complex network structures or relationships (indirect). Local/global, on the other hand, refers to whether a metric provides information about the network as a whole structure (global) or about the micro-structure of the network (local). Degree and eigenvector centrality reflect the prominence of a node, the degree centrality serving as an indicator of animals that are well connected with others and the eigenvector centrality as an indicator of influential animals[47]. In the context of the study, a pig with a high degree centrality is lying in proximity to many of its pen mates, whereas a pig with a high eigenvector centrality is lying in proximity to other pigs that are themselves lying in proximity to many of their pen mates. Clustering coefficient and radius evaluate the cohesiveness and the strength of bonds in a group[10,11]. The clustering coefficient evaluates the cohesiveness

between the neighbours of a specific node (pen mates it lies close to) and the radius between all group members. A pig with a high clustering coefficient is part of a closely connected group, where its neighbours frequently lie in proximity to each other. A low radius value indicates that most of the pigs are lying in proximity to most of their pen mates.

Comparison of social network metrics

A Mann–Whitney U test was used to assess whether the metric of the enriched pens (n = 12) differed significantly from the metric of the barren pens (n = 9) within each proximity definition. This non-parametric statistical test was used because the data were not normally distributed.

A bivariate linear model was fitted in ASReml 4.2[49] to estimate the correlation between the same metric for pairs of proximity definitions or pairs of sampling rates (Eq. 8),

$$\begin{bmatrix} \mathbf{y}_1 \\ \mathbf{y}_2 \end{bmatrix} = \begin{bmatrix} \mathbf{X}_1 & 0 \\ 0 & \mathbf{X}_2 \end{bmatrix} \begin{bmatrix} \boldsymbol{\beta}_1 \\ \boldsymbol{\beta}_2 \end{bmatrix} + \begin{bmatrix} \boldsymbol{\varepsilon}_1 \\ \boldsymbol{\varepsilon}_2 \end{bmatrix} \tag{8}$$

where $\mathbf{y}_1$ and $\mathbf{y}_2$ are the response vectors for the first and second trait, respectively, $\mathbf{X}_1$ and $\mathbf{X}_2$ are the design matrices for the vectors of fixed effects $\boldsymbol{\beta}_1$ and $\boldsymbol{\beta}_2$, and $\boldsymbol{\varepsilon}_1$ and ε_2 are the vectors of residuals. Traits 1 and 2 were either the metrics of a pair of proximity definitions, or the metrics of a pair of sampling rates. When analyzing sampling rates, trait 1 was always the metric of the complete network and trait 2 the metric of the reduced network. The fixed effects included housing and sex for the individual level metrics, and housing for the pen level metric (radius). The normality of the residuals was assessed for each univariate model by visual inspections (histograms and QQ-plots) and a Shapiro–Wilk test. A log transformation was applied to the radius, as the residuals of the models built with radius and DIST and DIST_SURF were not normally distributed. Residuals follow a bivariate normal distribution with (co)variance matrix:

$$var \begin{bmatrix} \boldsymbol{\varepsilon}_1 \\ \boldsymbol{\varepsilon}_2 \end{bmatrix} = \begin{bmatrix} \sigma_{\varepsilon_1}^2 & \rho_{12}\sigma_{\varepsilon_1}\sigma_{\varepsilon_2} \\ \rho_{12}\sigma_{\varepsilon_1}\sigma_{\varepsilon_2} & \sigma_{\varepsilon_2}^2 \end{bmatrix} \otimes \mathbf{I}$$

where $\otimes$ denotes the Kronecker product of two matrices and $\mathbf{I}$ is an identity matrix. The ρ_{12} is the parameter of interest i.e. the correlation between the metrics for trait 1 and trait 2. Hence, correlations were estimated after the correction of response variables for the fixed effects.

To assess the significance of the correlations, we calculated the Z-score by dividing the correlation coefficient by the standard error.

Results
Data description
Distribution of the distance and surface overlap of bounding boxes
Figure 2a shows the distribution of the distance between the centroids of bounding boxes among all possible pairs of individuals when both were inactive. In enriched pens, the average distance between bounding box centroids was 202.9 ± 96.6 cm (mean ± standard error), whereas, in barren pens, it was 129.9 ± 88.4 cm. Note that the enriched pens were twice as large as the barren pens. In enriched pens, the distribution was right-skewed without clear peaks (Fig. 2a). In barren pens, the distribution was more complex, with three local maxima.
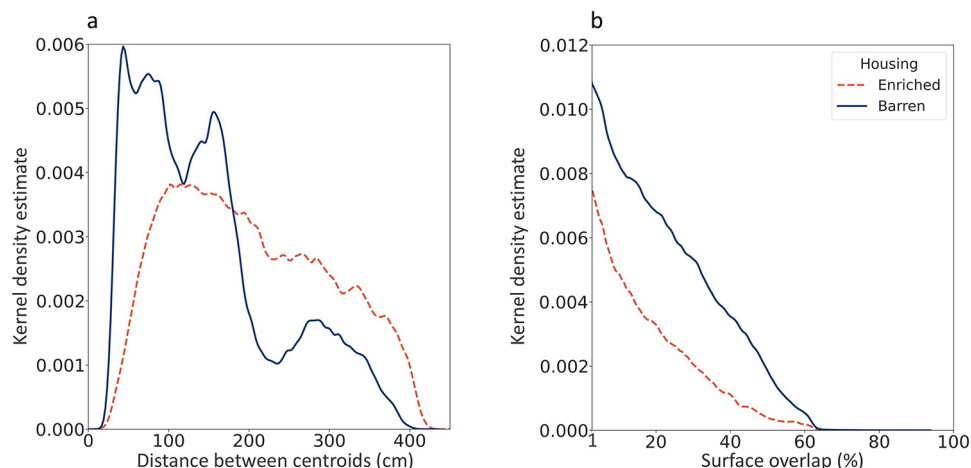

Fig. 2. Distributions of the distance between the centroids of pairs of bounding boxes (**a**) and the percentage of overlap of the surfaces of pairs of overlapping bounding boxes (**b**), separated for enriched and barren housing conditions. These distributions are based on n = 21 pens (12 enriched, 9 barrens), each pen containing 15 pairs of bounding boxes (for 6 individuals) at each time point (FPS = 0.5) and over 3 days. Note that panel b starts at values greater than 0% to focus on cases where bounding boxes overlap.

In addition, in the enriched pens, 2.3% of the pairs of bounding boxes were in close proximity based on their distance (< 50 cm), whereas in the barren pens, this was 11.3% (Fig. 2a).

Figure 2b shows the distribution of the percentage of overlap, between pairs of bounding boxes with more than 0% overlap. In total, 15.9% of the pairs of bounding boxes overlapped in enriched pens and 32.3% in barren pens (Fig. 2b). For both housing conditions, the distribution followed a right-skewed decreasing trend, with a higher probability for a small overlap than a large overlap. The probability of an overlap greater than ~60% was almost zero. Within the pairs that showed overlap, the percentage of overlap was smaller in enriched pens than in barren pens. In addition, in the enriched pens, 5.2% of the pairs of bounding boxes were in close proximity based on their overlap (> 20%), whereas in the barren pens this was 14.4% (Fig. 2b).

Agreement between proximity definitions
Table 3 shows the percentage of agreement of the four binary proximity definitions. The agreement varied from 81.9% between DIST_SURF and OVER to 98.9% between DIST and DIST_SURF (Table 3). The lowest agreements were found between OVER and the other definitions, and the highest between DIST_SURF and the two definitions DIST and SURF (Table 3).

Figure 3 shows four cropped frames with the bounding boxes of two pigs, illustrating agreement (Fig. 3a) and disagreement between the binary proximity definitions (Fig. 3b–d). As an example, we concentrate our result description on DIST and DIST_SURF. In Fig. 3b, the two pigs adopted a parallel ventral position, resulting in a short distance between the centroids of their bounding boxes (DIST criterion for proximity satisfied) but a limited surface overlap (SURF proximity criterion rejected). In Fig. 3c, d, only the SURF criterion was met and not DIST. This occurred for example, when the two pigs faced each other, with a portion of their body aligned in parallel (Fig. 3c), or when the two pigs lay at an angle (Fig. 3d).

Consequences of the proximity definitions for the network

The results reveal significant differences between enriched and barren pens for some of the network metrics (Table 4). Within the same proximity definition, barren pens exhibited a higher degree centrality, clustering coefficient and radius compared to the enriched pens (Table 4). This result indicates denser connectivity within the network of the barren pens in line with the distribution of the distance and surface overlap (Fig. 2).

Figure 4 shows the correlations, estimated by a bivariate model, of the complete networks built with the five different proximity definitions for each of the four metrics studied. All correlations were highly significant ($Z > 3.29$, $p < 0.001$), indicating strong associations for the same network metric across different proximity definitions. Overall, positive correlations were observed but the strength of these correlations varied largely. Strong correlation coefficients were observed for the direct metric degree centrality, ranging from 0.66 to 0.99 (Fig. 4). However, these correlations were lower for indirect metrics with values decreasing to 0.39 for eigenvector centrality and 0.35 for clustering coefficient. For the indirect metrics, the correlations were found to be weaker for the local metric clustering coefficient than the global metric eigenvector centrality. At the network level, the radius metric showed both weak and strong correlations, ranging from 0.52 to 0.96 (Fig. 4).

Among the binary proximity definitions, there was a trend to stronger correlations between the metrics when the proximity definition was built using the same geometric measure (Fig. 4). For instance, there were higher correlations between the metrics, when proximity was based on bounding box centroid distance (DIST and DIST_SURF) or between the ones based on surface overlap (SURF and OVER). On the other hand, correlations tended to be lower when comparing different geometric measures; particularly between DIST_SURF and OVER (correlation coefficients < 0.66).(Interestingly, the correlations between the proximity definition WEIG and the other four definitions differed largely depending on which metric was considered. The correlations between a metric of a pair of proximity definitions were particularly high with degree centrality and lower with the other metrics (Fig. 4).

A case study of an enriched pen

To illustrate the influence of the choice of a proximity definition on the network, a case study of an enriched pen is presented in Fig. 5. The DIST, DIST_SURF, and SURF networks show variations in edge strengths between different nodes within the network. These variations suggest that certain pairs of individuals spent more time in close proximity compared to other pairs (Fig. 5). In contrast, in the OVER and WEIG networks, all individuals appeared to be nearly equally connected within the network, suggesting a more uniform distribution of proximity across the group. This contrast between the networks is also reflected by the network metrics, with higher variations between the metrics values for different nodes in DIST, DIST_SURF, and SURF than in OVER

Proximity definition	DIST	SURF	DIST_SURF	OVER
DIST	100	–	–	–
SURF	94.8	100	–	–
DIST_SURF	98.9	95.9	100	–
OVER	82.9	86.1	81.9	100

Table 3. Percentage of agreement between the four binary proximity definitions. The percentages of agreement are calculated based on the tracking data with a sampling rate of 0.5 FPS. DIST: Distance between centroids of bounding boxes < 50 cm, SURF: Overlap of bounding boxes > 20%, DIST_SURF: Combination of DIST and SURF, OVER: Presence of overlap between bounding boxes.
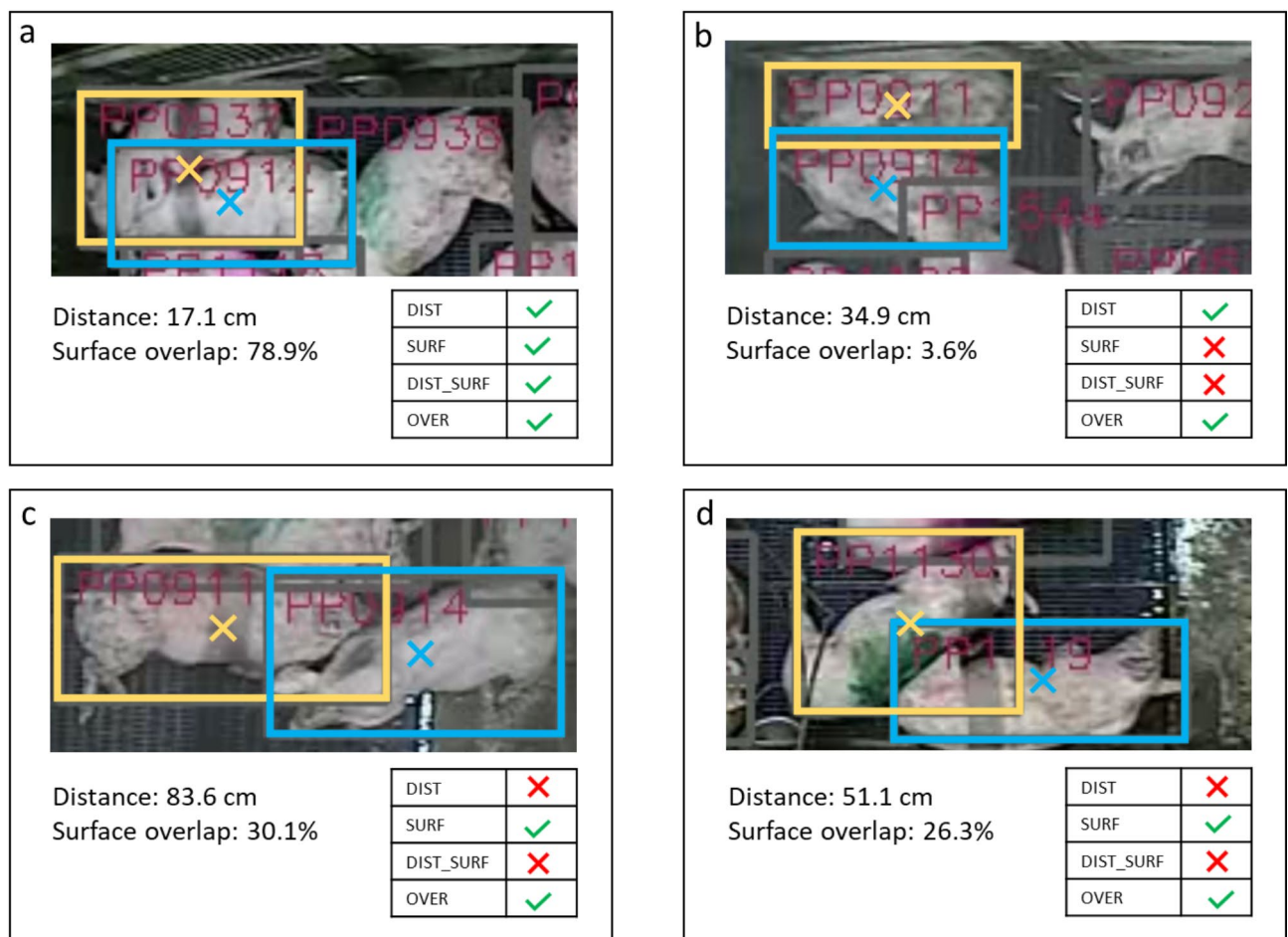
Fig. 3. Examples of cropped frames showing detected pairs of pigs with bounding boxes (**a**–**d**). In blue and yellow the bounding boxes and their respective centroids are highlighted. Below the frames, the distance between the centroids of the bounding boxes and surface overlap of the bounding boxes are given. The table indicates whether the pigs fulfil each proximity definition (marked by a green check) or not (indicated by a red cross). DIST: Distance between centroids of bounding boxes < 50 cm, SURF: Overlap of bounding boxes > 20%, DIST_SURF: Combination of DIST and SURF, OVER: Presence of overlap between bounding boxes.

and WEIG. The high connectivity between all pairs of individuals in OVER and WEIG is reflected by the higher clustering coefficient for both OVER and WEIG compared with DIST, DIST_SURF, and SURF (Fig. 5).

Furthermore, certain edges (e.g., 4–6 and 1–5) exhibited consistent strength levels across all networks, whereas others displayed varying strength between the five different networks (Fig. 5). For example, edge 3–4 was characterized by weak strength in OVER, SURF, and WEIG, but high strength in DIST and DIST-SURF (Fig. 5). As a result, we observed that pigs were ranked differently across networks, according to the network metrics. For example, in DIST and DIST_SURF, Pig 3 exhibited a noticeably high centrality (according to average degree and eigenvector) whereas Pig 1 had a low centrality (Fig. 5). This position was, however, not consistent across the five network definitions.

Consequences of downsampling for each network metric and proximity definition.
The impact of downsampling the tracking data for the two housing conditions combined is shown in Fig. 6 and presented separately for enriched and barred pens in Supplementary Figure S2. The impact of downsampling varied across network metrics (Fig. 6). Overall, the degree, eigenvector centrality, and radius were minimally impacted by downsampling. Especially for the degree centrality, the correlations with the complete network (1 frame every 2 s) remained high (> 0.90) even after reducing the sampling rate to 1 frame every hour (Fig. 6). Lower correlations between the reduced networks and the complete network were found for the clustering coefficient. To maintain high correlations for all the network metrics (> 0.90) without considering the housing conditions, the downsampling should not go beyond 1 frame every 6 min (Fig. 6). However, in barren pens, a larger sampling rate (1 frame every 25 min) would be sufficient to maintain high correlations for all network metrics, whereas in enriched pens, the downsampling should not go beyond 1 frame every 6 min (Supplementary Fig. S2).

The impact of downsampling on the network metrics varied across proximity definitions. For the networks built with WEIG, the metrics of the reduced networks maintained strong correlations with the complete network

Proximity definition	Housing	Degree centrality	Clustering coefficient	Eigenvector centrality	Radius
DIST	Barren	50.8 ± 6.4	0.42 ± 0.02	0.40 ± 0.04	10.1 ± 3.7
	Enriched	8.7 ± 1.4***	0.30 ± 0.02***	0.39 ± 0.04	1.2 ± 0.5***
DIST_SURF	Barren	40.4 ± 5.2	0.41 ± 0.02	0.39 ± 0.04	7.8 ± 3.5
	Enriched	7.6 ± 1.3***	0.30 ± 0.02***	0.39 ± 0.04	0.96 ± 0.5***
SURF	Barren	64.5 ± 5.5	0.45 ± 0.02	0.39 ± 0.03	13.2 ± 3.6
	Enriched	22.4 ± 2.5***	0.40 ± 0.02***	0.40 ± 0.03	4.1 ± 1.6***
OVER	Barren	**144.7** ± 5.8	0.59 ± 0.01	**0.41** ± 0.01	**31.7** ± 5.8
	Enriched	**68.4** ± 3.9***	0.55 ± 0.01***	**0.41** ± 0.01	**14.5** ± 2.4***
WEIG	Barren	4.0 ± 0.2	**0.73** ± 0.01	**0.41** ± 0.01	0.03 ± 0.01
	Enriched	2.5 ± 0.1***	**0.76** ± 0.01*	**0.41** ± 0.01	0.02 ± 0.01**

Table 4. Comparison of four network metrics (mean ± standard deviation) per housing condition (Barren vs. Enriched) built with different proximity definitions. The 10 networks were built with a sampling rate of 0.5 FPS (complete networks). Proximity between two pigs is defined with DIST: Distance between centroids of bounding boxes < 50 cm, SURF: Overlap of bounding boxes > 20%, DIST_SURF: Combination of DIST and SURF, OVER: Presence of overlap between bounding boxes, WEIG: harmonic mean of the distance between the centroids of the bounding boxes. Differences between housing conditions were tested with a Mann–Whitney U test for each metric and proximity definition and are indicated after the values for enriched pens, $*p < 0.05$, $**p < 0.01$, $***p < 0.001$. Values in bold indicate the highest value per metric. Note that the enriched pen is twice the size of the barren pen.

(ranging from 0.86 for clustering coefficient to 0.98 for degree centrality at a 1 frame per hour sampling rate, Fig. 6). For degree centrality and radius, the choice for a binary proximity definition seems to have less impact on the effect of downsampling on the correlations. For degree centrality, for example, the maximum difference between the lowest and highest correlations for different proximity definitions was 0.05 (observed at a sampling rate of 1 frame per 45 min). In contrast, the choice of a binary proximity definition has a clear impact on the effect of downsampling on the clustering coefficient. For the clustering coefficient, the maximum difference between the lowest and highest correlations for different proximity definitions was 0.55 (achieved with 1 frame per hour, Fig. 6). The clustering coefficient of the networks DIST and DIST_SURF were more sensitive to downsampling than the clustering coefficient of OVER and SURF.

Discussion

This study aimed to explore how different proximity definitions and sampling rates affect social network metrics of inactive pigs. Clear differences between networks, at both the individual and the network level, were revealed when using different proximity definitions. Network metrics differed in their sensitivity to downsampling, and this sensitivity was also affected by the chosen proximity definition.

In previous studies on spatial proximity in pigs with manual observation, lying in spatial proximity was defined in different ways[19,20,22,23]. The definition seems to depend on the interpretation of the observer whether two pigs are in proximity or not. Given the absence of one unique proximity definition, the goal of this study was not to identify the best definition for validating the detection of proximity. However, observers likely agree that pigs are in close proximity when there is full-body contact (as captured by DIST definition, see Fig. 3a, b). In contrast, part-body contact (captured by definition based on the overlap of bounding boxes, i.e. SURF and OVER, see Fig. 3c, d) is more likely to result in interobserver discrepancies, for example in the case when pigs are tail-to-tail.

The downside of DIST, SURF and DIST_SURF, the latter definition combining distance and overlap, is that the experimenter needs to decide on a minimal threshold that distinguishes whether pigs are in close proximity or not. The decision on this threshold might be arbitrary, as the 'ideal' distance is currently unknown[50]. This threshold needs to be decided carefully as a larger threshold, for example, may include more pairs of individuals considered as in close spatial proximity, resulting in a higher degree centrality[4]. The dependency of the results on the definition of the threshold also hampers comparisons of studies. In this study, we set thresholds for two definitions: DIST with a threshold of 50 cm distance between the centroids of two bounding boxes, and SURF with a threshold of 20% overlap of the surface of bounding boxes. DIST depends on the size of the animal, and the threshold should therefore be higher with larger pigs. SURF already takes into account the size of the animal, which makes it possible to use the same threshold for older pigs. An alternative to defining the threshold by an experimenter may utilize the size of the bounding box for the threshold.

Two of our definitions, OVER and WEIG, did not rely on a threshold set by the experimenter. OVER, i.e. the presence of overlap between two bounding boxes, might have included pigs that did not have any kind of contact. This definition can be relevant for a large population in a large area, but it may not be so suitable for pigs kept in relatively small pens. WEIG, the quantitative definition of proximity, was defined as the harmonic mean proximity between pigs and calculated as the average of the reciprocals of their distances. This definition assigns more significance to differences in distance when pigs are in close proximity compared to when they are further apart. As this approach includes all the distances, we expect it to be more powerful in distinguishing between
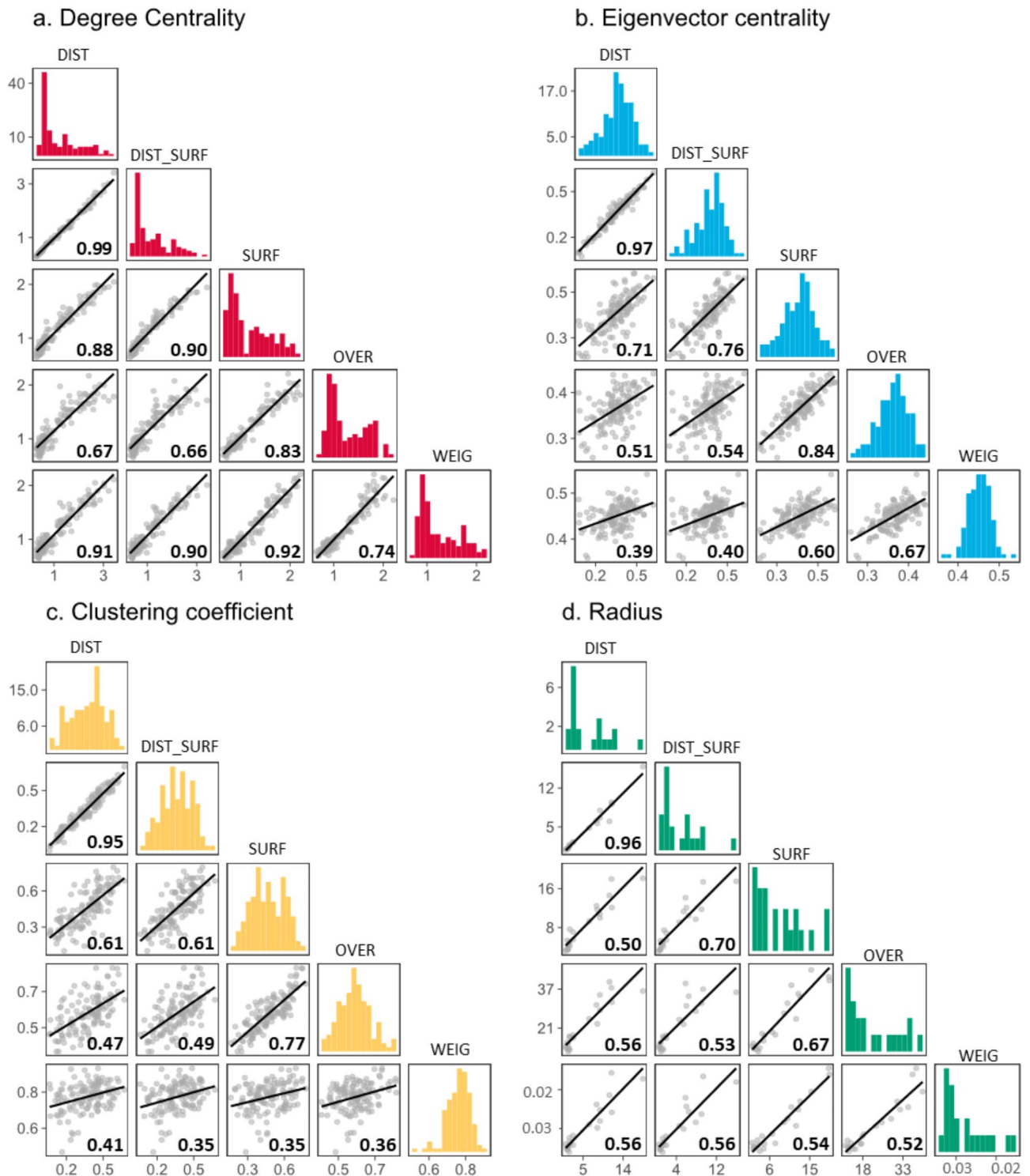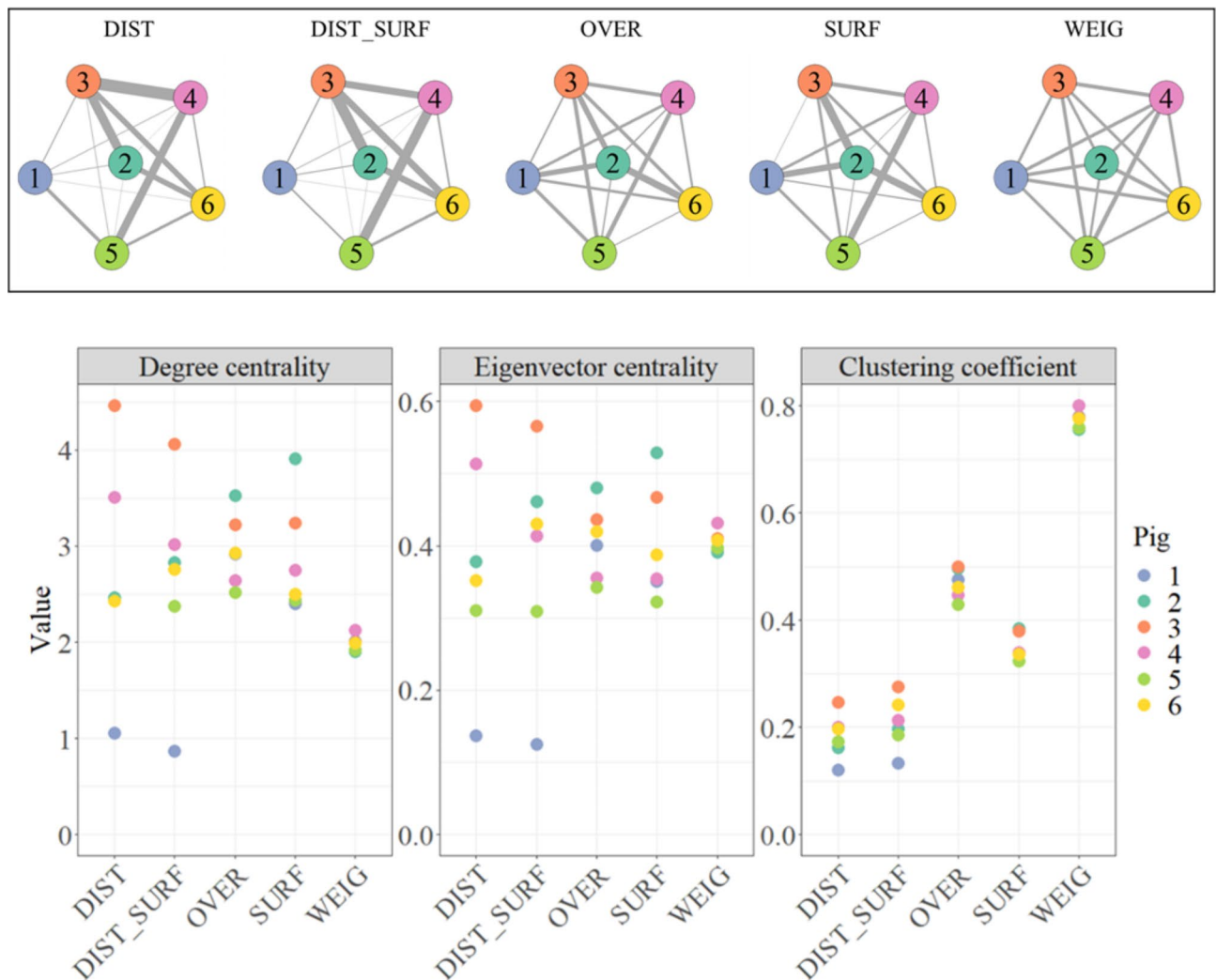
Fig. 4. Comparison of network metrics built with different proximity definitions. (a) Degree centrality, (b) Eigenvector centrality, (c) Clustering coefficient, (d) Radius. The network metrics are calculated based on the tracking data with a sampling rate of 0.5 FPS. The lower off-diagonal shows scatter plots with the linear fit in black and the correlation coefficients calculated from the bivariate model. All the correlations were significant ($p < 0.001$). The correlations coefficients for the radius were estimated with a log-transformation, due to the non-normal distribution of the residuals. The scatter plot and distribution of the radius values are plotted with the original data. Proximity between two pigs is defined with DIST: Distance between centroids of bounding boxes < 50 cm, SURF: Overlap of bounding boxes > 20%, DIST_SURF: Combination of DIST and SURF, OVER: Presence of overlap between bounding boxes, WEIG: harmonic mean of the distance between the centroids of the bounding boxes.

Fig. 5. An example of the structure of the networks for an enriched pen based on the different proximity definitions. The network metrics are calculated based on the tracking data with a sampling rate of 0.5 FPS. The width of the edge is proportional to the strength of the proximity. The degree centrality shown in this plot has been normalized by dividing the value by the average degree centrality of each proximity definition. The normalization was applied to the degree centrality to facilitate the comparison across different networks. In contrast, normalization was not necessary for eigenvector centrality and clustering coefficient, as these metrics are already bounded between 0 and 1. Proximity between two pigs is defined with DIST: Distance between centroids of bounding boxes < 50 cm, SURF: Overlap of bounding boxes > 20%, DIST_SURF: Combination of DIST and SURF, OVER: Presence of overlap between bounding boxes, WEIG: harmonic mean of the distance between the centroids of the bounding boxes.

certain pairs of individuals—for example, in distinguishing one pair that does not meet a binary definition of proximity but is often slightly above the threshold to another pair that consistently remains far apart.

At the individual level, changes in proximity definition had a minor effect on the direct metric degree centrality, whereas the indirect metrics (eigenvector centrality and clustering coefficients) were considerably affected. Degree centrality only considers a node's immediate connections, making it less sensitive to changes in the network[4] induced by different proximity definitions, compared with indirect metrics which consider the importance of a node based not only on its own connections but also on the connections of its neighbours[1]. A different proximity definition could result in changes in the weight of the edge and neighbouring nodes, potentially causing a cascading effect throughout the entire network structure and leading to significant shifts in the values of the indirect metric. In line with this result, Castles et al. (2014), who compared baboon networks based on three different proximity definitions (distance rule, chain rule, and nearest neighbour) also observed that the direct metrics degree and closeness were less affected by the chosen proximity definition than the indirect metric betweenness (another indirect metric that quantifies the extent to which a node serves as an intermediary between other nodes in a network)[25].

The local metric clustering coefficient was more affected by the choice of proximity definition than the global metric eigenvector centrality. The clustering coefficient is based on micro-level patterns within the network,
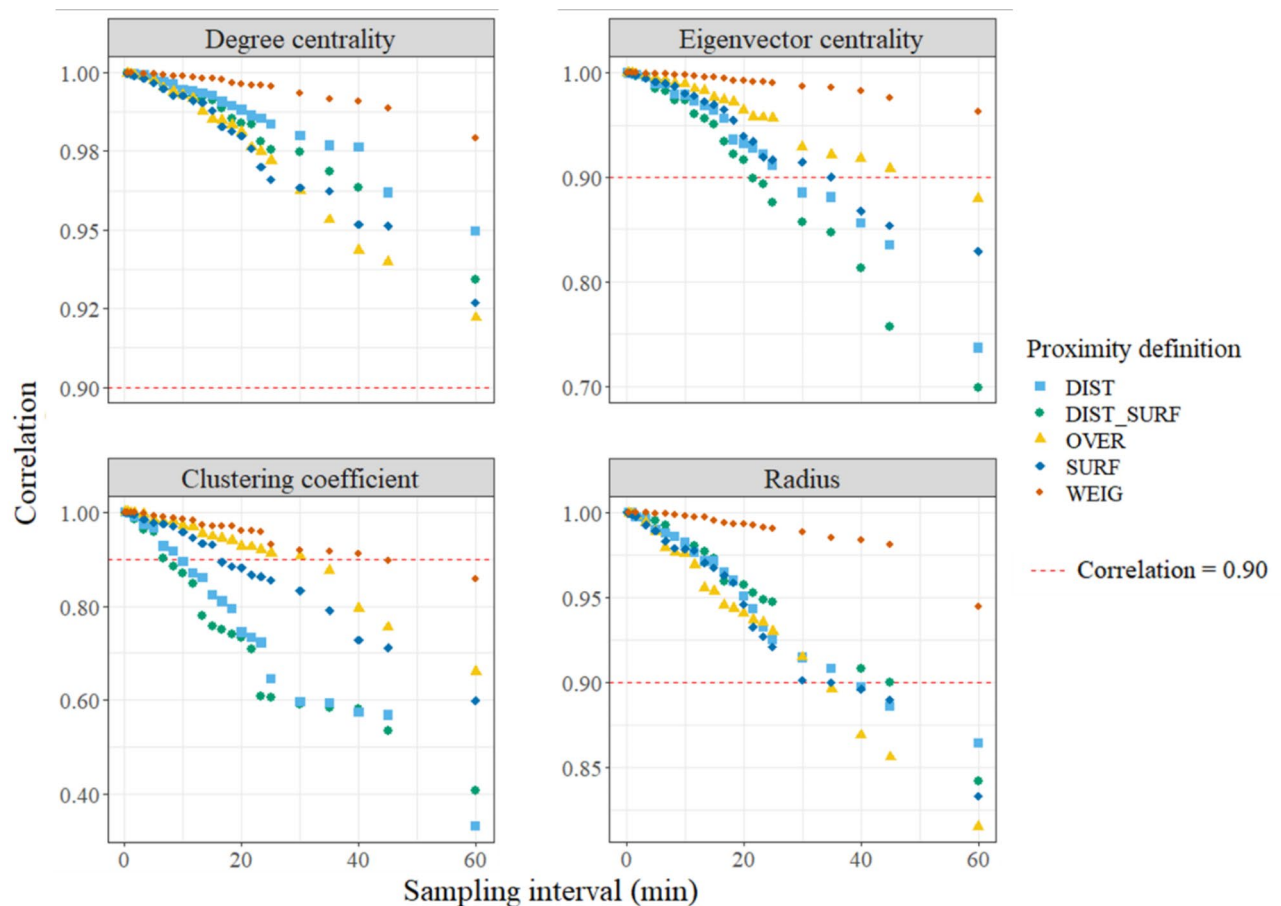
Fig. 6. Correlations between the metrics of the reduced network (built with different sampling intervals) and the metrics of the complete network (2 s sampling interval), for the degree centrality, eigenvector centrality, clustering coefficient, and radius. Note that the y-axis scale differs between panels. These plots show the results for both enriched and barren environments combined.

whereas eigenvector centrality takes into account a node's influence on the entire network and captures larger network patterns that may remain more consistent across proximity definitions. This could explain the lower correlation coefficients found between clustering coefficient metrics based on different proximity definitions and suggests that the networks based on the five different proximity definitions share more common information at a global network level, but may not align as closely at a local level.

Degree centrality, eigenvector centrality, and radius were also less impacted by downsampling than the clustering coefficient. For most of the proximity definitions (except DIST and DIST_SURF), downsampled tracking data still provided robust information (i.e. correlations between the yielded metrics and complete network remained > 0.90) on the positions of individuals within the network. These results agree with Davis et al. (2018), who found that low levels of sampling (10 samples/individual) can provide the same information as high levels of sampling (> 64 samples/individual in the same period) for social networks based on proximity in baboons[4]. Furthermore, downsampling had a negligible impact on the WEIG networks across various network metrics, compared with the binary definitions. This distinction lies in the nature of the proximity definition: WEIG integrates proximity between individuals in each frame, whereas the binary definitions consider individuals in proximity in a given frame only if they meet the specific condition. Downsampling increases the probability of missing a frame where this specific condition is met. Within the binary definitions, DIST and DIST_SURF were particularly affected by downsampling, given their restrictive condition occurring less frequently than OVER and SURF. Networks built with DIST and DIST_SURF require therefore a higher sampling rate to produce networks that reflect the social patterns on the complete period.

To produce networks that reflect similar social patterns as a high sampling rate, while reducing the data processing time, the downsampling should not go beyond 1 frame every 6 min in inactive pigs. A larger sampling interval may be sufficient when studying animals in limited space. In comparison, Larsen et al.[51] recommended a sampling interval of 1 frame every 30 s for assessing feeding duration in pigs. When investigating social networks based on proximity, longer sampling intervals remain reliable, given the higher frequency of proximity occurrences compared to feeding events and the focus on network metrics and not directly on duration.

The different values of the metrics across networks might also be a consequence of different types of proximity captured by the definitions. Although we found a high percentage of agreement between the binary proximity

definitions, especially between DIST, SURF, and DIST_SURF, when examining specific frames, these definitions appear to relate to distinct types of proximity. DIST seems to capture spatial proximity in cases where both individuals engaged in full-body contact, while SURF seems to capture situations where individuals were in part-body contact. In addition, OVER seems to capture both full-body and part-body contacts, including very small body contacts and even the absence of body contact.

It is unknown why a pig decides to be in full-body contact (as captured by DIST) or in part-body contact (as captured by SURF) with another pig, and which of these choices more accurately represents having a social preference for that other pig. The preference for full-body contact may be partially attributed to the pig's thermoregulatory needs[52,53]. However, in our study, the temperature was within the thermoneutral zone (22 °C and 23 °C), which reduced the likelihood of associations being driven by thermoregulation[54]. In addition, lying together for thermoregulation is more prevalent among newborns than among older pigs as used in this study[52,55]. In contrast, part-body contact, such as head-to-head, might mean being engaged in more specific social behaviours (e.g., nosing). In this study, no differentiation could be made between various body parts and information regarding the orientation of the pigs could not be considered. SURF might have captured part-body contact with all parts of the pig's body, including head-to-head or tail-to-tail contact, each potentially being driven by completely different motivations. Including the detection of specific key body points in the computer vision algorithm, for example, by detecting the head and the tail, could help to identify the direction of the pigs and to make the distinction between lying in close proximity head-to-head versus tail-to-tail. In a previous study, Wutke et al.[56] showed the potential of using key point detection to build networks based on specific contacts (head-to-head and head-to-tail).

In our study, we examined proximity during inactive periods, defined as instances where the distance moved between two frames one minute apart was less than 20 cm. We assumed that when inactive, pigs were predominantly in a lying posture, taking into account that fattening pigs spend the major part of their time lying[57]. However, our method could not distinguish between lying and standing pigs, therefore inactive defined pigs may have also included pigs that were standing still in the same location, for instance, when pigs in enriched environments were exploring straw. This method is therefore not suitable to assess welfare based on activity, as behaviours that are beneficial to pig's welfare, such as exploration of straw[58–61], are misclassified as inactivity. Additionally, our definition might have excluded pigs that were moving while lying, for example changing from lateral to sternal postures. We intentionally opted for a simple definition, as the main focus of our paper was on defining proximity and not on inactivity. Nevertheless, for research focused on networks based on spatial proximity during lying periods, more sophisticated criteria (such as analyzing consecutive active frames) or using alternative algorithms to differentiate between postures, could be investigated.

We observed a greater number of pairs of individuals with smaller distances and with larger overlapping values of their bounding box surfaces in barren pens compared to enriched pens. This suggests that when more space is given to the pigs, they do not choose to stay as close as in smaller environments. This questions the biological relevance of using spatial proximity to investigate social preferences in pigs kept under intensive housing conditions. Limited space allowance may inhibit the ability to express social preferences and maintain individual relationships reflected by spatial proximity[19,22]. In contrast, it has been shown that pigs with access to outdoor areas, generally characterized by more space, exhibited non-random social preferences in their lying behaviour[20]. Therefore, specific pro-social behaviours, such as play or social nosing contacts might be more relevant when investigating social preferences in restricted space[62,63]. Gan et al.[64] developed an algorithm based on convolutional neural networks to identify social behaviours between piglets (snout-snout, snout-body and play behaviour). However, data collection methods to study SNA based on an algorithm for such specific social behaviours and how the choice of one method affects the network metrics still have to be investigated.

In conclusion, computer vision algorithms based on tracking allow for the collection of data on location of individuals at a high sampling rate and are therefore useful to characterize social networks based on proximity. However, our results demonstrate that social network characteristics are influenced by both proximity definition and sampling rate and that their impact varies for different network metrics. Moreover, the effect of the sampling rate on network structure and the position of individuals within the network depends on the proximity definition chosen. This means that proximity definitions and sampling rates need to be selected carefully when using computer vision algorithms for social network analysis.

Data availability
Data can be made available upon request by contacting the corresponding author.

References
1. Makagon, M. M., McCowan, B. & Mench, J. A. How can social network analysis contribute to social behavior research in applied ethology? *Appl. Anim. Behav. Sci.* **138**, 152–161. https://doi.org/10.1016/j.applanim.2012.02.003 (2012).
2. Sosa, S., Sueur, C. & Puga-Gonzalez, I. Network measures in animal social network analysis: Their strengths, limits, interpretations and uses. *Methods Ecol. Evol.* **12**, 10–21. https://doi.org/10.1111/2041-210x.13366 (2021).
3. Buttner, K., Czycholl, I., Mees, K. & Krieter, J. Agonistic interactions in pigs-comparison of dominance indices with parameters derived from social network analysis in three age groups. *Animals* **9**, 929. https://doi.org/10.3390/ani9110929 (2019).
4. Davis, G. H., Crofoot, M. C. & Farine, D. R. Estimating the robustness and uncertainty of animal social networks using different observational methods. *Anim. Behav.* **141**, 29–44. https://doi.org/10.1016/j.anbehav.2018.04.012 (2018).
5. Kulahci, I. G., Ghazanfar, A. A. & Rubenstein, D. I. Consistent individual variation across interaction networks indicates social personalities in lemurs. *Anim. Behav.* **136**, 217–226. https://doi.org/10.1016/j.anbehav.2017.11.012 (2018).

6. Modlmeier, A. P., Keiser, C. N., Watters, J. V., Sih, A. & Pruitt, J. N. The keystone individual concept: an ecological and evolutionary overview. *Anim. Behav.* **89**, 53–62. https://doi.org/10.1016/j.anbehav.2013.12.020 (2014).
7. Lewton, J. & Rose, P. E. Social networks research in ex situ populations: Patterns, trends, and future directions for conservation-focused behavioral research. *Zoo. Biol.* **40**, 493–502. https://doi.org/10.1002/zoo.21638 (2021).
8. Rose, P. E. & Croft, D. P. The potential of Social Network Analysis as a tool for the management of zoo animals. *Anim. Welf.* **24**, 123–138. https://doi.org/10.7120/09627286.24.2.123 (2015).
9. Farine, D. R. & Whitehead, H. Constructing, conducting and interpreting animal social network analysis. *J. Anim. Ecol.* **84**, 1144–1163. https://doi.org/10.1111/1365-2656.12418 (2015).
10. Adelman, J. S., Moyers, S. C., Farine, D. R. & Hawley, D. M. Feeder use predicts both acquisition and transmission of a contagious pathogen in a North American songbird. *Proc. R. Soc. B Biol. Sci.* **282**, 20151429. https://doi.org/10.1098/rspb.2015.1429 (2015).
11. Aplin, L. M. et al. Consistent individual differences in the social phenotypes of wild great tits, Parus major. *Anim. Behav.* **108**, 117–127. https://doi.org/10.1016/j.anbehav.2015.07.016 (2015).
12. Finkemeier, M. A., Langbein, J. & Puppe, B. Personality research in mammalian farm animals: concepts, measures, and relationship to welfare. *Front. Vet. Sci.* **5**, 131. https://doi.org/10.3389/fvets.2018.00131 (2018).
13. McCowan, B., Anderson, K., Heagarty, A. & Cameron, A. Utility of social network analysis for primate behavioral management and well-being. *Appl. Anim. Behav. Sci.* **109**, 396–405. https://doi.org/10.1016/j.applanim.2007.02.009 (2008).
14. Coutellier, L. et al. Pig's responses to repeated social regrouping and relocation during the growing-finishing period. *Appl. Anim. Behav. Sci.* **105**, 102–114. https://doi.org/10.1016/j.applanim.2006.05.007 (2007).
15. Chopra, K. et al. Proximity interactions in a permanently housed dairy herd: Network structure, consistency, and individual differences. *Front. Vet. Sci.* **7**, 583715. https://doi.org/10.3389/fvets.2020.583715 (2020).
16. Clouard, C. et al. Evidence of stable preferential affiliative relationships in the domestic pig. *Anim. Behav.* **213**, 95–105. https://doi.org/10.1016/j.anbehav.2024.04.009 (2024).
17. Vázquez-Diosdado, J. A., Occhiuto, F., Carslake, C. & Kaler, J. Familiarity, age, weaning and health status impact social proximity networks in dairy calves. *Sci. Rep.* **13**, 2275. https://doi.org/10.1038/s41598-023-29309-1 (2023).
18. Reimert, I., Bolhuis, J. E., Kemp, B. & Rodenburg, T. B. Social support in pigs with different coping styles. *Physiol. Behav.* **129**, 221–229. https://doi.org/10.1016/j.physbeh.2014.02.059 (2014).
19. Durrell, J. L., Sneddon, I. A., O'Connell, N. E. & Whitehead, H. Do pigs form preferential associations? *Appl. Anim. Behav. Sci.* **89**, 41–52. https://doi.org/10.1016/j.applanim.2004.05.003 (2004).
20. Goumon, S., Illmann, G., Leszkowova, I., Dostalova, A. & Cantor, M. Dyadic affiliative preferences in a stable group of domestic pigs. *Appl. Anim. Behav. Sci.* **230**, 105045. https://doi.org/10.1016/j.applanim.2020.105045 (2020).
21. Stookey, J. M. & Gonyou, H. W. Recognition in swine: Recognition through familiarity or genetic relatedness? *Appl. Anim. Behav. Sci.* **55**, 291–305. https://doi.org/10.1016/S0168-1591(97)00046-4 (1998).
22. Li, Y. Z., Zhang, H. F., Johnston, L. J. & Martin, W. Understanding tail-biting in pigs through social network analysis. *Animals* **8**, 13. https://doi.org/10.3390/ani8010013 (2018).
23. Jowett, S. & Amory, J. The stability of social prominence and influence in a dynamic sow herd: A social network analysis approach. *Appl. Anim. Behav. Sci.* **238**, 105320. https://doi.org/10.1016/j.applanim.2021.105320 (2021).
24. Wilder, T., Krieter, J., Kemper, N. & Buttner, K. Network analysis of tail-biting in pigs—The impact of missed biting events on centrality parameters. *J. Agricul. Sci.* **160**, 107–116. https://doi.org/10.1017/s0021859622000090 (2022).
25. Castles, M. et al. Social networks created with different techniques are not comparable. *Anim. Behav.* **96**, 59–67. https://doi.org/10.1016/j.anbehav.2014.07.023 (2014).
26. Agha, S., Foister, S., Roehe, R., Turner, S. P. & Doeschl-Wilson, A. Genetic analysis of novel behaviour traits in pigs derived from social network analysis. *Genes* **13**, 561. https://doi.org/10.3390/genes13040561 (2022).
27. Foister, S. et al. Social network properties predict chronic aggression in commercial pig systems. *PLoS ONE* **13**, e0205122. https://doi.org/10.1371/journal.pone.0205122 (2018).
28. Agha, S., Fabrega, E., Quintanilla, R. & Sanchez, J. P. Social network analysis of agonistic behaviour and its association with economically important traits in pigs. *Animals* **10**, 2123. https://doi.org/10.3390/ani10112123 (2020).
29. Buttner, K., Scheffler, K., Czycholl, I. & Krieter, J. Network characteristics and development of social structure of agonistic behaviour in pigs across three repeated rehousing and mixing events. *Appl. Anim. Behav. Sci.* **168**, 24–30. https://doi.org/10.1016/j.applanim.2015.04.017 (2015).
30. Charles, K. M. S., VanderWaal, K. L., Anderson, J. E., Johnston, L. J. & Li, Y. Z. Evaluating social network metrics as indicators of tail injury caused by tail biting in growing-finishing pigs (Sus scrofa domesticus). *Front. Vet. Sci.* **11**, 1441813. https://doi.org/10.3389/fvets.2024.1441813 (2024).
31. Clouard, C. et al. Evidence of stable preferential affiliative relationships in the domestic pig. *Anim. Behav.* **213**, 95–105. https://doi.org/10.1016/j.anbehav.2024.04.009 (2024).
32. Nathan, R. et al. Big-data approaches lead to an increased understanding of the ecology of animal movement. *Science* **375**, 734. https://doi.org/10.1126/science.abg1780 (2022).
33. Ozella, L. et al. The effect of age, environment and management on social contact patterns in sheep. *Appl. Anim. Behav. Sci.* **225**, 104964. https://doi.org/10.1016/j.applanim.2020.104964 (2020).
34. Mar, C. C. et al. Cow detection and tracking system utilizing multi-feature tracking algorithm. *Sci. Rep.* **13**(1), 17423. https://doi.org/10.1038/s41598-023-44669-4 (2023).
35. van Putten, A., Giersberg, M. F. & Rodenburg, T. B. Tracking laying hens with ArUco marker backpacks. *Smart Agricult. Technol.* **10**, 100703. https://doi.org/10.1016/j.atech.2024.100703 (2025).
36. Guo, Q. H. et al. Enhanced camera-based individual pig detection and tracking for smart pig farms. *Comput. Electron. Agricult.* **211**, 108009. https://doi.org/10.1016/j.compag.2023.108009 (2023).
37. van der Zande, L. E., Guzhva, O. & Rodenburg, T. B. Individual detection and tracking of group housed pigs in their home pen using computer vision. *Front. Anim. Sci.* **2**, 669312. https://doi.org/10.3389/fanim.2021.669312 (2021).
38. van der Zande, L. E. et al. Estimation of resilience parameters following lps injection based on activity measured with computer vision. *Front. Anim. Sci.* **3**, 883940. https://doi.org/10.3389/fanim.2022.883940 (2022).
39. Shemesh, Y. et al. High-order social interactions in groups of mice. *Elife* **2**, e00759. https://doi.org/10.7554/eLife.00759 (2013).
40. Schneider, J., Dickinson, M. H. & Levine, J. D. Social structures depend on innate determinants and chemosensory processing in Drosophila. *Proc. Natl. Acad. Sci. USA* **109**, 17174–17179. https://doi.org/10.1073/pnas.1121252109 (2012).
41. du Sert, N. P. et al. The ARRIVE guidelines 2.0: Updated guidelines for reporting animal research. *Bmc Vet. Res.* https://doi.org/10.1186/s12917-020-02451-y (2020).
42. Parois, S. P. et al. A multi-suckling system combined with an enriched housing environment during the growing period promotes resilience to various challenges in pigs. *Sci. Rep.* **12**, 6804. https://doi.org/10.1038/s41598-022-10745-4 (2022).
43. Redmon, J. & Farhadi, A. YOLOv3: An Incremental Improvement. arXiv preprint arXiv:1804.02767 (2018).
44. Bewley, A., Gongyuan, Z., Ott, L., Ramos, F. & Upcroft, B. Simple online and realtime tracking. *IEEE International Conference on Image Processing (ICIP)*, 3464–3468. https://doi.org/10.1109/ICIP.2016.7533003 (2016).
45. Cairns, S. J. & Schwager, S. J. A comparison of association indexes. *Anim. Behav.* **35**, 1454–1469. https://doi.org/10.1016/s0003-3472(87)80018-0 (1987).
46. Csárdi, G. et al. igraph: Network Analysis and Visualization in R. R package version 2.1.4. https://CRAN.R-project.org/package=igraph (2025).

47. Cheney, D. L., Silk, J. B. & Seyfarth, R. M. Network connections, dyadic bonds and fitness in wild female baboons. *R. Soc. Open Sci.* **3**, 160255. https://doi.org/10.1098/rsos.160255 (2016).
48. McKee, J & Dallas, T. Structural network characteristics affect epidemic severity and prediction in social contact networks. *Infect. Dis. Model.* **9**, 204–213. https://doi.org/10.1016/j.idm.2023.12.008 (2024).
49. Butler, D. G., Cullis, B. R., Gilmour, A. R., Gogel, B. G. & Thompson, R. ASReml-R Reference Manual Version 4.2. (VSN International Ltd., Hemel Hempstead, UK, 2023).
50. Bartlett, E., Cameron, L. J. & Freeman, M. S. A preliminary comparison between proximity and interaction-based methods to construct equine (Equus caballus) social networks. *J. Vet. Behav. Clin. Appl. Res.* **50**, 36–45. https://doi.org/10.1016/j.jveb.2022.01.005 (2022).
51. Larsen, M. L. V., Liu, D., Kobek-Kjeldager Sigvartssøn, M., Norton, T. & Pedersen, L. J. Computer vision as a means to determine optimal sampling interval for feeding behaviour duration of finisher pigs. *International Society for Applied Ethology (ISAE).*, Vol. 18, 13. (Tallinn, Estonia, 2023).
52. Camerlink, I., Scheck, K., Cadman, T. & Rault, J. L. Lying in spatial proximity and active social behaviours capture different information when analysed at group level in indoor-housed pigs. *Appl. Anim. Behav. Sci.* **246**, 105540. https://doi.org/10.1016/j.applanim.2021.105540 (2022).
53. Spoolder, H. A. M., Aarnink, A. A. J., Vermeer, H. M., van Riel, J. & Edwards, S. A. Effect of increasing temperature on space requirements of group housed finishing pigs. *Appl. Anim. Behav. Sci.* **138**, 229–239. https://doi.org/10.1016/j.applanim.2012.02.010 (2012).
54. Huynh, T. T. T. et al. Thermal behaviour of growing pigs in response to high temperature and humidity. *Appl. Anim. Behav. Sci.* **91**, 1–16. https://doi.org/10.1016/j.applanim.2004.10.020 (2005).
55. Villanueva-García, D. et al. Hypothermia in newly born piglets: Mechanisms of thermoregulation and pathophysiology of death. *J. Anim. Behav. Biometeorol.* **9**(1), 2101. https://doi.org/10.31893/jabb.21001 (2021).
56. Wutke, M. et al. Detecting animal contacts-a deep learning-based pig detection and tracking approach for the quantification of social contacts. *Sensors* **21**, 7512. https://doi.org/10.3390/s21227512 (2021).
57. Ekkel, E. D., Spoolder, H. A. M., Hulsegge, I. & Hopster, H. Lying characteristics as determinants for space requirements in pigs. *Appl. Anim. Behav. Sci.* **80**, 19–30. https://doi.org/10.1016/s0168-1591(02)00154-5 (2003).
58. Beattie, V. E., O'Connell, N. E. & Moss, B. W. Influence of environmental enrichment on the behaviour, performance and meat quality of domestic pigs. *Livestock Product. Sci.* **65**, 71–79. https://doi.org/10.1016/s0301-6226(99)00179-7 (2000).
59. Bolhuis, J. E., Schouten, W. G. P., Schrama, J. W. & Wiegant, V. M. Behavioural development of pigs with different coping characteristics in barren and substrate-enriched housing conditions. *Appl. Anim. Behav. Sci.* **93**, 213–228. https://doi.org/10.1016/j.applanim.2005.01.006 (2005).
60. Luo, L., Reimert, I., Middelkoop, A., Kemp, B. & Bolhuis, J. E. Effects of early and current environmental enrichment on behavior and growth in pigs. *Front. Vet. Sci.* **7**, 268. https://doi.org/10.3389/fvets.2020.00268 (2020).
61. Van de Weerd, H. A., Docking, C. M., Day, J. E. L. & Edwards, S. A. The development of harmful social behaviour in pigs with intact tails and different enrichment backgrounds in two housing systems. *Anim. Sci.* **80**, 289–298. https://doi.org/10.1079/asc40450289 (2005).
62. Camerlink, I. & Turner, S. P. The pig's nose and its role in dominance relationships and harmful behaviour. *Appl. Anim. Behav. Sci.* **145**, 84–91. https://doi.org/10.1016/j.applanim.2013.02.008 (2013).
63. Portele, K. et al. Sow-piglet nose contacts in free-farrowing pens. *Animals* **9**(8), 53. https://doi.org/10.3390/ani9080513 (2019).
64. Gan, H. M. et al. Automated detection and analysis of social behaviors among preweaning piglets using key point-based spatial and temporal features. *Comput. Electron. Agricult.* **188**, 106357. https://doi.org/10.1016/j.compag.2021.106357 (2021).

Author contributions

Clémence A.E.M. Orsini: Conceptualization, methodology, analysis, writing original draft. Bernadett Hegedűs: Conceptualization, methodology, review. Lisette E. van der Zande: Development of the algorithm, review. Inonge Reimert: Supervision, conceptualization, methodology, review. Piter Bijma and J. Elizabeth Bolhuis: Supervision, conceptualization, methodology, review, project administration, funding acquisition.

Funding

Declarations

Competing interests

The authors declare no competing interests.

Ethical statement

The animal study was reviewed and approved by the Dutch Central Authority for Scientific Procedures on Animals.

Additional information

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1038/s41598-025-93830-8.

Correspondence and requests for materials should be addressed to C.A.E.M.O.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.