

1 **An information-theory approach to geometry for animal groups**

2 Christoph D. Dahl<sup>1,2,3,\*</sup>, Elodie Ferrando<sup>3</sup>, and Klaus Zuberbühler<sup>3,4</sup>

3

4

5 <sup>1</sup> Graduate Institute of Mind, Brain and Consciousness, Taipei Medical University, Taipei,  
6 Taiwan

7 <sup>2</sup> Brain and Consciousness Research Center, Taipei Medical University Shuang-Ho Hospital,  
8 New Taipei City, Taiwan

9 <sup>3</sup> Institute of Biology, University of Neuchâtel, Neuchâtel, Switzerland

10 <sup>4</sup> School of Psychology and Neuroscience, University of St Andrews, United Kingdom

11

12 \* corresponding author:

13 Christoph D. Dahl < christoph.dahl@tmu.edu.tw, christoph.d.dahl@gmail.com >

## 14 **Abstract**

15 One of the hardest problems in studying animal behaviour is to quantify patterns of social  
16 interaction at the group level. Recent technological developments in global positioning system  
17 (GPS) devices have opened up new avenues for locating animals with unprecedented spatial  
18 and temporal resolution. Likewise, advances in computing power have enabled new levels of  
19 data analyses with complex mathematical models to address unresolved problems in animal  
20 behaviour, such as the nature of group geometry and the impact of group-level interactions  
21 on individuals. Here, we present an information theory-based tool for the analysis of group  
22 behaviour. We illustrate its affordances with GPS data collected from a freely interacting pack  
23 of 15 Siberian huskies (*Canis lupus familiaris*). We found that individual freedom in movement  
24 decisions was limited to about 4%, while a subject's location could be predicted with 96% median  
25 accuracy by the locations of other group members. Dominant individuals were less affected by  
26 other individuals' locations than subordinate ones, and same-sex individuals influenced each  
27 other more strongly than opposite-sex individuals. We also found that kinship relationships  
28 increased the mutual dependencies of individuals. Moreover, the network stability of the  
29 pack deteriorated with an upcoming feeding event. Together, we conclude that information  
30 theory-based approaches, coupled with state-of-the-art bio-logging technology, provide a powerful  
31 tool for future studies of animal social interactions beyond the dyadic level.

## 32 **Keywords**

33 selfish herd, group decision, dominance, kin relation, *Canis lupus familiaris*, Siberian husky,  
34 hierarchy, spatial cognition, social intelligence

## 35 **Introduction**

36 The combination of naturalistic observations in the field and systematic experimental testing in  
37 captive settings has long been the gold standard in studies of animal behaviour and cognition  
38 (Kummer, 1984). Behaviour, in turn, is usually collected by human observers, who follow a range  
39 of established observation techniques (Altmann, 1974) and experimental designs (Zuberbühler  
40 and Wittig, 2011). Although this combination has provided unprecedented progress, it has  
41 a number of inherent and highly problematic flaws. First, behaviour coding relies on human  
42 categorisation, which is inherently prone to bias. A common way to address this is to carry out  
43 inter-observer reliability tests, which remedies the problem to some extent. Second, behaviour  
44 experiments are often in interaction with humans, which can introduce 'Clever Hans' effects and

45 other forms of cuing (Pfungst, 1911). Third, human-led behaviour studies are usually restricted  
 46 to reporting from single individuals interacting with its socio-ecological Umwelt, which typically  
 47 excludes analyses at larger levels, including how individuals contribute to group-level behaviour  
 48 and decisions, or how they are influenced by them (Altmann, 1974; Snowdon, 1983). As a  
 49 consequence, current theories of animal social behaviour are largely based on data collected  
 50 during pair-wise interactions of animals (Branson et al., 2009; Dankert et al., 2009; De Chaumont  
 51 et al., 2012; Langford et al., 2006).

52 More recently, there have been efforts to studying patterns of social interactions at the group  
 53 level (Shemesh et al., 2013; Strandburg-Peshkin et al., 2015), but this is hardly possible without  
 54 moving away from subjectively coding animal behaviour by human observers. Bio-logging is  
 55 among the most promising tools for quantifying animal behaviour without subjective human  
 56 coding (Dahl et al., 2018; Gerencsér et al., 2013). For example, GPS-based inertial sensor  
 57 technology provides high-resolution data of movement patterns, which can be further processed  
 58 towards automated behavioural classification systems and descriptions of group-level interactions  
 59 (Ferdinandy et al., 2017; Ákos et al., 2014; Dahl et al., 2018; Nagy et al., 2010; Strandburg-Peshkin  
 60 et al., 2015), even in the absence of image-based tracking (Dell et al., 2014; Rasch et al.,  
 61 2016; Shemesh et al., 2013). Although these data are relatively easy to obtain, describing  
 62 social interactions in groups is computationally challenging, requiring mathematical models of  
 63 interaction that are complex (Bialek et al., 2012; Couzin et al., 2005, 2002; Lukeman et al.,  
 64 2010; Vicsek et al., 1995).

65 Here, we provide a means to study the social and environmental determinants of group formation  
 66 in free-ranging, group-living animals. Using GPS, we determined the individuals location  
 67 and assigned the location to one of 25 equally sized and spatially aligned patches. This  
 68 procedure guaranteed large enough sample conditions, while taking loss of information due  
 69 to data discretisation into account. We used information theory (Shannon, 1948) to determine  
 70 the degree to which individuals influence each other in choosing their location in the patch-grid  
 71 and how this is driven by social and environmental determinants. We test this tool on a pack of  
 72 15 dogs (*Canis lupus familiaris*; Siberian husky) living in a large outdoor enclosure. First, we  
 73 hypothesised social hierarchy to be an important factor in how animals choose their preferred  
 74 location, with lower ranking individuals' location being more strongly determined by others'  
 75 location than higher ranking ones. Second, we also hypothesised kinships to have an effect, albeit  
 76 in the opposite direction (Städele et al., 2016), resulting in greater predictability of related than  
 77 unrelated individuals. Third, we hypothesised sex to play a role, such that same-sex individuals  
 78 determined each other's position to a stronger degree than opposite-sex individuals, following

79 the idea of male-female role allocation in wolves (Mech, 1999). Lastly, we hypothesised a role of  
 80 important external events on the stability of the pack. We therefore analysed the predictability  
 81 of individuals' locations relative to an upcoming feeding event. Since each individual was  
 82 chained to a specific location during feeding, we predicted that, with increased proximity to  
 83 feeding time, group stability should decrease.

## 84 **Materials and Methods**

### 85 *Subjects*

86 Our subjects were 15 Siberian huskies (*Canis lupus familiaris*) living as a pack in an open  
 87 area of 750m<sup>2</sup> (25m x 30m) at the dog-sledding centre 'Les Attelages de la Roche Percée' at  
 88 'La Ferme de Nirveau' in 25510 Pierrefontaine-les-Varans, France (Fig. 1A). Subjects' ages  
 89 ranged from 5 to 10 years (M = 6.73, SD = 1.83, Table 1). None of the individuals suffered  
 90 from any known orthopaedic and/or neurological disorders. According to the owners, the dogs  
 91 have been living as one pack for at least five years - the youngest individuals were five at the  
 92 time of recording. Human interactions with the dogs were largely in terms of training for dog  
 93 sledding and feeding once per day. During the time of recording (October to November 2017)  
 94 there was no training. Social rank was provided by owners, who were also the care-takers,  
 95 based on qualitative assessments (Fig. 1C). In detail, we asked the owners to evaluate which  
 96 individual of each given pair is more dominant than the other. This resulted in a matrix of  
 97 scores (1 = dominant, .5 = equal, 0 = submissive, Table 2), from which we then calculated  
 98 the normalised David's scores (Gammell et al., 2003). Despite the fact that owners generally  
 99 rate the dominance of their dogs relatively reliably (Kubinyi and Wallis, 2019), in the current  
 100 study, social rank was only used as a measure to correlate and interpret the outcome from the  
 101 movement pattern analyses and had no influence on the analysis of the movement patterns per  
 102 se. Kinship was coded as the presence or absence of siblings in the group (Table 1).

103

### 104 *Data logging*

105 All 15 individuals were fitted with data loggers that provided location data via GPS, as latitudes  
 106 and longitudes in degrees, minutes and seconds. As loggers we used 'Adafruit Ultimate GPS  
 107 Featherwing' devices with an update rate of 5Hz, allowing positional accuracy of 1.5 m, a  
 108 velocity accuracy of .1 m/s and 'Adafruit Feather Adalogger' with an ATmega32u4 clocked at  
 109 8MHz (Adafruit Industries, NY 10013, USA). Data were locally stored on SanDisk Ultra 16GB  
 110 MicroSD cards (Western Digital Technologies, Inc., Milpitas, CA 95035, USA). We further used  
 111 3.7V 900mAh LiPo batteries to provide power of up to a full day of continuous recording. The

112 hardware was programmed using ‘Arduino Adafruit GPS Library’ to read out the unparsed  
 113 NMEA sentences. Time stamps were recorded using an in-built real-time clock (RTC).

114

115 *Data collection and analysis*

116 We recorded data during three sessions on separate days, lasting nine hours each. A session  
 117 started after the dogs were equipped with loggers attached to harnesses and it ended before the  
 118 dogs were tethered for feeding around 18:00. In each session we recorded positioning data from  
 119 all fifteen dogs. We first parsed the NMEA sentences into latitude, longitude and time. We  
 120 then applied a rotational transformation in the image plane to align the shape of the habitat  
 121 to an upright rectangle. The values were then converted into a normalized representation of  
 122 space ranging from 0 to 1 on both axes, while maintaining the spatial proportions. We then  
 123 equally spaced the habitat into 5 by 5 patches (smaller areas) and converted each normalized  
 124 GPS location into the corresponding patch within which the GPS location fell (Fig. 1B). The  
 125 number of patches in the patch-grid was chosen based on theoretical, practical and statistical  
 126 considerations: (a) In theory, the patch size must have a minimum area sufficient to allow all  
 127 individuals to be simultaneously present. In the case of 15 husky dogs at least  $5m^2$ -patches  
 128 are required. (b) Data discretisation, the process of converting continuous spatial information  
 129 into a finite set of spatial patches, increases the number of samples per condition at the cost  
 130 of a discretisation error. A patch size in the mid-range of the available patch size spectrum  
 131 is a reasonable choice that aims towards an optimal trade-off between statistical power and  
 132 discretisation error. (c) A finer patch-grid (i.e. smaller patches and higher number of patches)  
 133 generates a greater number of location patterns and conveys more information about the social  
 134 interactions from which to derive the social network. However, a finer patch-grid also generates  
 135 a greater amount of theoretically possible location patterns, e.g., 15 dogs and a 7 by 7 patch-grid  
 136 result in  $15^{49}$  location patterns. Refining the patch-grid is accompanied by a drastic increase of  
 137 computational power required to analyse the location patterns.

138

139 *Location patterns*

140 For each time point  $t$  we determined a location pattern consisting of the location patches of all  
 141 individuals (e.g., location pattern:  $(x_1, x_2, x_3, \dots, x_n) = [2, 1, 9, \dots, 4]$ ), where  $x_i$ , is the location  
 142 of dog  $i$  (ranging from 1 to 15) and  $x$  is the location (ranging from 1 to 25) (Shemesh et al.,  
 143 2013). We then determined the frequencies of finding the fifteen individuals in given location  
 144 patterns ( $p_{\text{empirical}} = (x_1, x_2, x_3, \dots, x_n)$ ) and compared these frequencies with an independent  
 145 model, assuming that each individual dog moves freely and solely according to its individual

146 preferences ( $p_{\text{independent}} = p(x_1) p(x_2) p(x_3) \dots p(x_n)$ ).

147

148 ‘Mutual information’

149 We based our analyses on the concept of ‘mutual information’ (Shannon, 1948). To this end, we  
 150 first calculated the uncertainty of the location of dog  $i$ , defined as the entropy of the distribution  
 151 of locations of dog  $i$ , which is

$$152 \quad H(x_i) = - \sum_{x_i} p(x_i) \log_2 p(x_i) \quad (1)$$

153 We then calculated the conditional entropy of locations of a pair of dogs ( $i, j$ )

$$154 \quad H(x_i|x_j) = - \sum_{x_i x_j} p(x_i, x_j) \log_2 p(x_i|x_j) \quad (2)$$

155 and subtracted that from one of the dogs ( $i$ ); that is, the mutual information between the  
 156 locations of dog  $i$  and dog  $j$ , described as follows:

157

$$158 \quad I(x_i; x_j) = H(x_i) - H(x_i|x_j) \quad (3)$$

159 The mutual information for dog  $i$  and dog  $j$  was then divided by the entropy of dog  $i$  to get a  
 160 normalized measurement of the uncertainty about the location of dog  $i$ , when considering the  
 161 joint positions of dogs  $i$  and  $j$ . Therefore,

$$162 \quad \frac{I(x_i; x_j)}{H(x_i)} \quad (4)$$

163 describes the fraction of the uncertainty regarding the locations of dog  $i$  that can be explained by  
 164 the locations of dog  $j$ . Accordingly, we calculated the mutual information of one dog considering  
 165 the joint locations of multiple dogs, given by the following equation:

166

$$167 \quad \frac{I(x_i; \{x_j, x_k, x_l, x_m, x_n, \dots x_w\})}{H(x_i)} \quad (5)$$

168 To determine the distribution of mutual information between fifteen dogs and 25 possible  
 169 locations, we used a resampling approach, by selecting 4'000 randomly determined configurations  
 170 of dog  $i$  (to-be-explained individual) and dogs  $j$  to  $w$  (number of explaining individuals was  
 171 varied from 1 to 14).

172

173 *Predicting individual location: rank and kinship*

174 Mutual information values were binned according to the individual subjects and number of  
 175 group members taken into consideration to explain a subject’s location. This allowed us to

176 systematically describe dependencies of a subject's social rank and the predictability of its  
 177 location by the locations of others. Using Spearman's rank correlation we hypothesized that  
 178 low-ranking individuals, more than high-ranking individuals, were subjected to the positions of  
 179 others and, hence, more predictable regarding their locations. Mutual information values were  
 180 also used to describe mutual information between siblings and non-relatives. Here, we predicted  
 181 that mutual information was larger between sibling than non-relatives, given that siblings build  
 182 a social bond and stay in close proximity to each other. To test this, we performed a two-way  
 183 analysis of variance with mutual information values as the dependent variable, 'Kinship' (yes,  
 184 no) and '# individuals', describing how many individuals' locations were used to explain the  
 185 locations of the individual to be explained, being the grouping (independent) variables.

186

187 *Social network*

188 We determined the social network (Fig 2AB) by calculating linkages (edges) between pairs of  
 189 individuals (nodes). Linkages were defined as mutual information larger than average of all  
 190 mutual information values. In other words, edges indicated that two individuals (nodes) were  
 191 influencing each other in terms of their locations above average. The underlying algorithm was  
 192 a so-called 'force-directed placement' plotting routine (Fruchterman and Reingold, 1991). These  
 193 algorithms follow an aesthetical principle, i.e., they aim at placing all nodes in a way that the  
 194 edges are close to equal lengths while minimising the number of edge-crossings. Based on the  
 195 relative positions of sets of nodes and edges, the algorithm assigns forces to them and repositions  
 196 the sets of nodes and edges to minimize force energies (Kobourov, 2012). In detail, forces push  
 197 or pull the nodes closer together or further apart, eventually reaching a state of equilibrium,  
 198 where the spatial repositioning of nodes becomes ineffective, i.e. nodes do not move anymore.  
 199 Based on these positions, the 2D-graphs were drawn and the sum of edges on the shortest path  
 200 between individuals with respect to intrinsic attributes, such as sex and kinship, were calculated.  
 201 We also performed a randomization test by computing the sum of edges of randomly assigned  
 202 individuals to the current positions, shuffling kinship labels while preserving the existing node  
 203 locations. We ran this iterative procedure 200 times and compared the actual values with it. We  
 204 computed a similar procedure for sex, reassigning male and female labels to current positions,  
 205 in order to calculate same- and opposite-sex distances under randomly assigned conditions. We  
 206 use two-way anova to analyse the shortest path between individuals (15) and the number of  
 207 individuals used to explain the remaining individual (1 to 14).

208

209 *External event*

210 The most relevant non-social event in subjects' daily activities was most certainly the advent  
 211 of feeding, which always took place around 18:00 local time. For this purpose, the dogs were  
 212 individually tethered to a chain spot on the ground in order to avoid food-related aggression.  
 213 Therefore, recording sessions regularly ended prior to the tethering of the dogs, when the loggers  
 214 were turned off and the harnesses were taken off. The actual feeding event and its influence  
 215 on group structure, for this reason, was not part of the study, although we could study the  
 216 effects of anticipation by time-logging the entire recording session towards the feeding event.  
 217 We subdivided the entire session into time frames of 54 min and calculated group information  
 218 within these time periods. We applied Gaussian Kernel linear regression on time frames (elapsed  
 219 time) and mutual information values.

220

221 *Estimate of misplacement*

222 The location of the experimental site was a fenced open field providing strong GPS reception  
 223 from multiple satellites with little to no signal obstruction. Multipath error, i.e. the satellite  
 224 signal bouncing off buildings or large natural formations and, thus, generating large positional  
 225 errors, did not occur. Additionally, the GPS receivers were initiated at least 10 minutes prior to  
 226 the beginning of the recordings to first acquire signals from satellites. To estimate inaccuracies  
 227 as a result of receiver errors, we calculated differences between consecutive measurement. Also,  
 228 we re-ran the main elements of the mutual information analysis to evaluate the effect due to the  
 229 unlikely scenario of having 25% misplacements in the patch-grid on the main results (Figure  
 230 4A-C). To this end, we randomly misplaced every fourth location point to one of the adjacent  
 231 patches prior to determining the mutual information values.

232

233 *Null distribution of mutual information*

234 We determined the null distributions of mutual information by individually permutating the  
 235 empirical data obtained from each animal. We then re-ran the main elements of the mutual  
 236 information analysis to evaluate the extent to which chance explains the empirical findings  
 237 (Figure 4DE).

238 **Results**

239 Preferences for locations were measured for each individual as fractions of time (Fig. 1C).  
 240 Probabilities of location patterns of the empirical data and of the independent model were  
 241 highly different (Fig 1D). Of possible  $25^{15}$  states (i.e., 25 locations by 15 individuals), only  
 242  $2 \cdot 10^4$  states (i.e. location patterns) were observed. However, the observed states occurred

243 at high frequencies, up to 382 times (Fig 1D, black line). In contrast, the individual model  
 244 predicted 31'096 location patterns, however, at lower probabilities (i.e. each location pattern  
 245 occurred one time only) (Fig 1D, red line). Hence, our data indicate that the group restricts  
 246 the number of socially accepted location patterns to about 7% for the average individual (2'104  
 247 out of 31'096 location patterns). The remaining 28'992 location patterns, predicted by a model  
 248 based on individual preferences, were socially avoided.

249 We then quantified the strength of dependencies between pairs of individuals. In detail, we  
 250 addressed the extent to which knowing the location of one dog helped inferring the location of  
 251 another dog. In a scenario where two individuals were fully independent on each other, knowing  
 252 one individual's location would not make any predictions about another individual's location.  
 253 We therefore determined the entropy (i.e. uncertainty) of locations for each individual to  
 254 predict the current location based on previous location information, and the mutual information  
 255 between two individuals, i.e. how much of the uncertainty about the location of one individual  
 256 could be reduced by knowing the location of another individual (see methods). We also  
 257 calculated the joint locations of multiple individuals to explain the locations of single individuals.  
 258 The results showed that knowing the locations of one individual allowed predicting a second  
 259 individual's location in a range from 11 to 67%, with a median of 30.2% (Figure 1E). Knowing  
 260 the location of two individuals allowed predicting the location of a third individual for up to  
 261 56.5% (median). With increasing number of individuals' joint locations, the explanatory power  
 262 of a further individual increased, reaching a median of 96.3% with 14 individuals (Figure 1F).  
 263 Additionally, we evaluated the extent to which these results can be explained by chance. As  
 264 shown in Figure 4D,E, there is a drastic reduction of mutual information when time sequences  
 265 were permuted for each individual prior to the calculation routine of mutual information.

266

267 *Social rank*

268 We then sorted the mutual information values according to the individual explained (e.g. dog  $i$ ),  
 269 independent of the number of individuals used to explain that individual (e.g. joint information  
 270 of dog  $j$ , dog  $k$  and dog  $l$ ). We found that two individuals (Dubble and Elska), determined as  
 271 submissive by the owners, turned out to be strongly influenced by other dogs' locations in terms  
 272 of choosing their own locations (Figure 1G). Two further individuals (Dusty and Gribouille),  
 273 classified as the most dominant dogs in the pack, were found to be least influenced by other  
 274 dogs' locations in terms of choosing their own locations. We correlated the social rank and  
 275 the mutual information values and found a positive correlation for most group sizes (Figure  
 276 1G, inset). The black line in Figure 1G indicates that group size of four individuals (i.e. one

277 individual to be explained and three individuals to explain) showed the strongest correlation  
 278 between social rank and mutual information values ( $rs(13) = .64, p < .05$ ) (Figure 1H).

279

280 *Kinship*

281 We also calculated the mutual information for siblings. Here, we determined the degree to which  
 282 we could explain an individual's location by the location information of all its siblings. As a  
 283 control, we calculated the predictability of the same individual but with non-kin individuals  
 284 contributing to the joint location information. We found that individual location could be  
 285 predicted to a larger degree by the location of siblings than unrelated group members (Kinship:  
 286  $F(1,29) = 4.78, p < .05, \text{Mean sq} = .11$ ; # individuals:  $F(2,29) = 3.41, p < .05, \text{Mean sq} = .08$ )  
 287 (Figure 1I).

288

289 *Social network*

290 We calculated the linkages between individuals based on the definition that a linkage exists  
 291 if it is above average. Social networks were determined using the 'Force' plotting routine  
 292 (Fruchterman and Reingold, 1991), in our case based on mutual information values derived from  
 293 four individuals (i.e. one individual to be explained and three individuals to explain, Figure  
 294 2AB). Results showed that nodes and linkages grouped according to the sex of the individuals  
 295 (Fig 2A), with sums of edges on the shortest path between individuals of same and opposite sexes  
 296 differing significantly ( $F(1,392) = 1.93, p < .05, \text{Mean sq} = 0.007$ ). The number of individuals  
 297 used to explain a single individual turned out to be a modulating factor ( $F(13,392) = 1.83,$   
 298  $p < .05, \text{Mean sq} = 0.006$ ) (Fig 2C). In an iterative t-test comparison (corrected for multiple  
 299 comparisons), we found that same-sex individuals were closer to each other than predicted by a  
 300 randomization control procedure when the social network was constructed on the basis of mutual  
 301 information values computed by the locations of 3 to 5 individuals to explain the locations of  
 302 another single individual (all  $p < .01$ ). Similarly, the same social network revealed a grouping  
 303 according to kinship, where siblings were spatially closer than non-siblings ( $F(1,392) = 1.77,$   
 304  $p < .05, \text{Mean sq} = 0.008$ ) (Fig 2BD). The number of individuals required to explain a single  
 305 individual's location was a modulating factor ( $F(13,392) = 2.31, p < .01, \text{Mean sq} = 0.007$ ) (Fig  
 306 2D). Critically, siblings were closer than predicted by an iterative procedure randomizing kinship  
 307 relationships when the social network was constructed on the basis of mutual information values  
 308 computed by the locations of 3 or 6 and more individuals to explain the locations of another  
 309 single individual (all  $p < .05$ ). The individuals Gribouille and, to a smaller extent, Dusty were  
 310 positioned slightly off from the general cluster. These individuals are dominant to all others and

311 have fewer dependencies on other individuals. Further Gribouille is primarily linked to male  
 312 individuals (Figure 2A) and appears to be separated from her own kin (Figure 2B).

313

314 *Anticipating non-social events*

315 Lastly, we determined the extent to which the group dependency was affected by external  
 316 events, introducing an interference factor on the group formation. We therefore subdivided  
 317 the recording session according to the relative time prior to feeding, which regularly occurred  
 318 around 18:00 local time, and determined the mutual information between each individual with  
 319 each other individual and each other pair or group of individuals, as described above, with  
 320 regard to these particular time windows. We found that the closer the time progressed toward  
 321 feeding, the smaller the probabilities of predicting one individual’s position by looking at other  
 322 individuals’ positions became (Figure 3A). We interpret this as a loss of group stability due  
 323 to the forthcoming feeding event that led to a group structure with fewer social constraints.  
 324 The greatest mutual information values, hence the strongest dependencies, were found during  
 325 a mid-day period. In contrast, early morning and evening periods – the latter being close  
 326 to feeding – showed less strong dependencies, hence reduced stability of the pack (Fig 3B).  
 327 Interestingly, this effect was strongest for mutual information values determined from pairs  
 328 of animals (i.e. addressing the question “how much of location of dog  $i$  can be explained by  
 329 location of dog  $j$ ?”). In contrast, mutual information values are more constant throughout the  
 330 day when determined based on a large group of animals (e.g. how much of the locations of dog  
 331  $i$  can we explain by looking at the remaining 14 dogs?).

332

333 *Error estimation*

334 We confirmed that GPS recordings were accurate and according to the specifications provided  
 335 by the manufacturer. We found that the 99% confidence intervals of differences of two consecutive  
 336 values were:  $CI = [.0451, .0456]$ , hence the number of two consecutive values deviating more  
 337 than 1.5m is neglectable small. Further, we created the unlikely scenario that a quarter of all  
 338 values is subjected to unspecified errors and results in a misplacement in the patch-grid. We  
 339 showed that this scenario, indeed, caused a drop in mutual information (Figure 4BC); however,  
 340 it did not affect the overall consistency of the results.

341 **Discussion**

342 We described movement patterns in a pack of dogs by collecting GPS (NMEA) data from  
 343 multiple animals in parallel and, using a real time clock (RTC) on each logger, synchronising

344 the individual movement patterns. We relied on a classic mathematical theory of communication  
 345 (Shannon, 1948) to quantify the extent to which one individual's movements can be explained  
 346 by other individuals' movements. We found highly correlated group behaviour, first, insofar as  
 347 the number of observed location patterns was very restricted and by a multitude smaller than  
 348 what would be predicted if individuals freely moved according to their individual preferences.  
 349 Hence, 'socially allowed' configurations are governed by the group, limiting individual freedom  
 350 in movement to about of 4%, suggesting that individuals' movement decisions were determined  
 351 by others up to 96%.

352

353 It must be noted that there is a trade-off between reducing the mathematical complexity  
 354 behind the analysis of group interactions and data discretisation by converting continuous spatial  
 355 information into a finite set of spatial patches. In this process, loss of information about the  
 356 exact location of individuals occurs. Interestingly, as the results of this study suggest, the social  
 357 characteristics constituting group behaviour remain observable.

358

359 We found that dominant individuals were influenced by other individuals to a lesser degree  
 360 than subordinate ones. We also found that, while same-sex individuals influenced each other  
 361 more strongly than opposite-sex individuals, the pack was spatially controlled by two dominant  
 362 individuals, an alpha male and an alpha female. Furthermore, we found that kin relationships  
 363 increased the mutual dependencies of individuals, with siblings moving together, a social structure  
 364 similar to family composition in free-ranging wolves (Gadbois, 2002; Mech, 1999; Packard, 2003).  
 365 Interestingly, the level at which grouping effects emerged varies depending on the attributes:  
 366 Social rank seems to play a major role in very small configurations of animals, emerging already  
 367 in the mutual information between one individual and two other individuals. Sex classes, too,  
 368 emerged at an early level of co-dependencies, i.e. with groups of 4 to 6 animals. In contrast,  
 369 kinship classes are more strongly subjected to the organization of the whole pack, emerging  
 370 most prominently at 7 and more individuals.

371

372 We also showed that group structure responded to an external factor, that is, upcoming feeding  
 373 events. Interestingly, as time progressed the mutual information between animals decreased,  
 374 suggesting reduced network stability due to the upcoming feeding event. This is to some extent  
 375 surprising in that it does not reflect food competition by means of dominating others, but  
 376 disintegration of social structure. Alternatively, the findings might be interpreted by considering  
 377 the given feeding ecology, where food is available on a daily basis and guaranteed to each

378 individual, irrespective of social rank and sex. It must be noted that we cannot rule out  
 379 changes in social structure due to the circadian rhythm, a biological process controlling and  
 380 regulating the sleep-wake cycle (Azzi et al., 2017).

381

382 *Evolution of social structure*

383 Studies in wild and captive wolves disagree on what should be considered the natural social and  
 384 hierarchical structure. While captive wolves and domestic dogs showed heightened agonistic  
 385 behaviour and a social structure best described by a linear hierarchical model (Cafazzo et al.,  
 386 2016, 2010), wild wolves have been described as living in family units, consisting of a breeding  
 387 pair with their offspring (Gadbois, 2002; Mech, 1999; Packard, 2003). Family compositions  
 388 suggest a non-linear and more complex but, at the same time, more flexible hierarchical structure  
 389 (Packard, 2003) with the parents playing the roles of leaders, making group decisions and  
 390 initiating group movements (Peterson et al., 2002). Little is known about social behaviour,  
 391 dominance and rank in natural packs of dogs. Free-ranging dogs normally build packs ranging  
 392 from 2 to 8 individuals, in rare cases over 20 individuals (Bonanni and Cafazzo, 2014), and  
 393 consist of males and females. Females tend to have multi-male mating preferences (Cafazzo  
 394 et al., 2014) and subsequently raise their offspring. Also noteworthy is that mating strategies,  
 395 raising preferences and rank hierarchies vary in dogs and wolves depending on the feeding  
 396 ecology (Marshall-Pescini et al., 2017).

397

398 Importantly, in our study, we used human-socialised dogs, as opposed to stray dogs with  
 399 unknown socialisation histories. Subjects had been raised in human social environments, but  
 400 kept in a relatively free-ranging and natural pack. As a consequence, their social structure might  
 401 differ from both wolves and free-ranging stray dogs, given the absence of influencing factors such  
 402 as foraging/hunting, mating/breeding and predators/danger sources, and the presence of factors  
 403 such as human-dependency, controlled feeding and care taking.

404

405 *Conclusions*

406 With recent developments in hard- and software technologies, traditional methods in animal  
 407 behaviour research are likely to experience fundamental changes towards automated continuous  
 408 recording procedures, allowing the quantification of not only single-animal activities, but interactions  
 409 of groups of animals with an unprecedented temporal accuracy. While many algorithms to  
 410 quantifying group interactions are highly sophisticated and mathematically complex (Bialek  
 411 et al., 2012; Lukeman et al., 2010; Vicsek et al., 1995), we here present a relatively simple

412 mathematical approach, rooted in information theory (Shannon, 1948). We were able to describe  
413 the social dynamics in a semi-natural pack of freely interacting dogs to a surprisingly complex  
414 degree, which revealed the influence of evolutionarily inherited social structures that included  
415 kin relations and dominance relations, as well as predictions made by individuals about relevant  
416 forthcoming external events.

## 417 **Acknowledgements**

418 We thank ‘Les Attelages de la Roche Percée’ at ‘La Ferme de Nirveau’ in Pierrefontaine-les-Varans,  
419 25510, France, for their support. We are grateful to the Swiss National Science Foundation  
420 for supporting this project via the Ambizione Fellowship (PZ00P3\_154741) awarded to CDD.  
421 KZ has also been supported by the Swiss National Science Foundation (31003A\_166458). We  
422 are also grateful to the Taipei Medical University for awarding a Startup- funding to CDD  
423 (108-6402-004-112). We thank Guillaume Dezecache and Malte J. Rasch for comments on the  
424 manuscript.

## 425 **Author contributions**

426 CDD: study design, data collection, analysis and interpretation, writing article, provision of  
427 necessary tools; EF: data collection, provision of resources; KZ: provision of necessary tools and  
428 resources, writing article.

## 429 **Funding**

430 This study was funded via the Ambizione Fellowship of the Swiss National Science Foundation  
431 (SNSF) (PZ00P3\_154741) and the Startup-funding of Taipei Medical University (108-6402-004-112)  
432 awarded to CDD as well as by project funding of the Swiss National Science Foundation  
433 (31003A\_166458) awarded to KZ.

## 434 **Ethical approval**

435 According to the local authorities (Comité d’Ethique de l’Expérimentation Animale Grand  
436 Campus Dijon, Université de Bourgogne, Maison de l’Université, Esplanade Erasme, 21078  
437 Dijon, France), non-invasive studies on dogs are allowed to be conducted without any special  
438 permission in France. ‘Les Attelages de la Roche Percée’ at ‘La Ferme de Nirveau’ in 25510

439 Pierrefontaine-les-Varans, France, responded to our enquiry and volunteered to participate in  
 440 this study.

## 441 **References**

442 Ákos Z, Beck R, Nagy M, Vicsek T, Kubinyi E (2014) Leadership and path characteristics during  
 443 walks are linked to dominance order and individual traits in dogs. *PLoS computational biology*  
 444 10(1):e1003446

445 Altmann J (1974) Observational study of behavior: sampling methods. *Behaviour*  
 446 49(3-4):227–266

447 Azzi A, Evans JA, Leise T, Myung J, Takumi T, Davidson AJ, Brown SA (2017) Network  
 448 dynamics mediate circadian clock plasticity. *Neuron* 93(2):441–450

449 Bialek W, Cavagna A, Giardina I, Mora T, Silvestri E, Viale M, Walczak AM (2012) Statistical  
 450 mechanics for natural flocks of birds. *Proceedings of the National Academy of Sciences*  
 451 109(13):4786–4791

452 Bonanni R, Cafazzo S (2014) The social organisation of a population of free-ranging dogs in a  
 453 suburban area of rome: a reassessment of the effects of domestication on dogs' behaviour. In:  
 454 *The social dog*, Elsevier, pp 65–104

455 Branson K, Robie AA, Bender J, Perona P, Dickinson MH (2009) High-throughput ethomics in  
 456 large groups of drosophila. *Nature methods* 6(6):451

457 Cafazzo S, Valsecchi P, Bonanni R, Natoli E (2010) Dominance in relation to age, sex,  
 458 and competitive contexts in a group of free-ranging domestic dogs. *Behavioral Ecology*  
 459 21(3):443–455

460 Cafazzo S, Bonanni R, Valsecchi P, Natoli E (2014) Social variables affecting mate preferences,  
 461 copulation and reproductive outcome in a pack of free-ranging dogs. *PLoS One* 9(6):e98594

462 Cafazzo S, Lazzaroni M, Marshall-Pescini S (2016) Dominance relationships in a family pack  
 463 of captive arctic wolves (*Canis lupus arctos*): the influence of competition for food, age and  
 464 sex. *PeerJ* 4:e2707

465 Couzin ID, Krause J, James R, Ruxton GD, Franks NR (2002) Collective memory and spatial  
 466 sorting in animal groups. *Journal of theoretical biology* 218(1):1–11

- 467 Couzin ID, Krause J, Franks NR, Levin SA (2005) Effective leadership and decision-making in  
 468 animal groups on the move. *Nature* 433(7025):513
- 469 Dahl CD, Wyss C, Zuberbühler K, Bachmann I (2018) Social information in equine movement  
 470 gestalten. *Animal cognition* 21(4):583–594
- 471 Dankert H, Wang L, Hoopfer ED, Anderson DJ, Perona P (2009) Automated monitoring and  
 472 analysis of social behavior in drosophila. *Nature methods* 6(4):297
- 473 De Chaumont F, Coura RDS, Serreau P, Cressant A, Chabout J, Granon S, Olivo-Marin JC  
 474 (2012) Computerized video analysis of social interactions in mice. *Nature methods* 9(4):410
- 475 Dell AI, Bender JA, Branson K, Couzin ID, de Polavieja GG, Noldus LP, Pérez-Escudero A,  
 476 Perona P, Straw AD, Wikelski M, et al. (2014) Automated image-based tracking and its  
 477 application in ecology. *Trends in ecology & evolution* 29(7):417–428
- 478 Ferdinandy B, Ozogány K, Vicsek T (2017) Collective motion of groups of self-propelled  
 479 particles following interacting leaders. *Physica A: Statistical Mechanics and its Applications*  
 480 479:467–477
- 481 Fruchterman TM, Reingold EM (1991) Graph drawing by force-directed placement. *Software:*  
 482 *Practice and experience* 21(11):1129–1164
- 483 Gadbois S (2002) The socioendocrinology of aggression-mediated stress in timber wolves (*Canis*  
 484 *lupus*).
- 485 Gammell MP, Vries Hd, Jennings DJ, Carlin CM, Hayden TJ (2003) David’s score: a more  
 486 appropriate dominance ranking method than clutton-brock et al.’s index. *Animal behaviour*  
 487 66(3):601–605
- 488 Gerencsér L, Vásárhelyi G, Nagy M, Vicsek T, Miklósi A (2013) Identification of behaviour in  
 489 freely moving dogs (*Canis familiaris*) using inertial sensors. *PloS one* 8(10):e77814
- 490 Kobourov SG (2012) Spring embedders and force directed graph drawing algorithms. arXiv  
 491 preprint arXiv:12013011
- 492 Kubinyi E, Wallis LJ (2019) Dominance in dogs as rated by owners corresponds to ethologically  
 493 valid markers of dominance. *PeerJ* 7:e6838
- 494 Kummer H (1984) From laboratory to desert and back: a social system of hamadryas baboons.  
 495 *Animal Behaviour* 32(4):965–971

- 496 Langford DJ, Cramer SE, Shehzad Z, Smith SB, Sotocinal SG, Levenstadt JS, Chanda ML,  
 497 Levitin DJ, Mogil JS (2006) Social modulation of pain as evidence for empathy in mice.  
 498 Science 312(5782):1967–1970
- 499 Lukeman R, Li YX, Edelstein-Keshet L (2010) Inferring individual rules from collective  
 500 behavior. Proceedings of the National Academy of Sciences 107(28):12576–12580
- 501 Marshall-Pescini S, Cafazzo S, Viranyi Z, Range F (2017) Integrating social ecology in  
 502 explanations of wolf–dog behavioral differences. Current Opinion in Behavioral Sciences  
 503 16:80–86
- 504 Mech LD (1999) Alpha status, dominance, and division of labor in wolf packs. Canadian Journal  
 505 of Zoology 77(8):1196–1203
- 506 Nagy M, Akos Z, Biro D, Vicsek T (2010) Hierarchical group dynamics in pigeon flocks. Nature  
 507 464(7290):890
- 508 Packard JM (2003) Wolf behaviour: reproductive, social and intelligent. Wolves: behavior,  
 509 ecology, and conservation pp 35–65
- 510 Peterson RO, Jacobs AK, Drummer TD, Mech LD, Smith DW (2002) Leadership behavior in  
 511 relation to dominance and reproductive status in gray wolves, *Canis lupus*. Canadian Journal  
 512 of Zoology 80(8):1405–1412
- 513 Pfungst O (1911) Clever Hans:(the horse of Mr. Von Osten.) a contribution to experimental  
 514 animal and human psychology. Holt, Rinehart and Winston
- 515 Rasch MJ, Shi A, Ji Z (2016) Closing the loop: tracking and perturbing behaviour of individuals  
 516 in a group in real-time. bioRxiv p 071308
- 517 Shannon CE (1948) A mathematical theory of communication. Bell system technical journal  
 518 27(3):379–423
- 519 Shemesh Y, Sztainberg Y, Forkosh O, Shlapobersky T, Chen A, Schneidman E (2013) High-order  
 520 social interactions in groups of mice. Elife 2:e00759
- 521 Snowdon CT (1983) Ethology, comparative psychology, and animal behavior. Annual Review  
 522 of Psychology 34(1):63–94
- 523 Städele V, Pines M, Swedell L, Vigilant L (2016) The ties that bind: maternal kin bias  
 524 in a multilevel primate society despite natal dispersal by both sexes. American journal of  
 525 primatology 78(7):731–744

- 526 Strandburg-Peshkin A, Farine DR, Couzin ID, Crofoot MC (2015) Shared decision-making  
527 drives collective movement in wild baboons. *Science* 348(6241):1358–1361
- 528 Vicsek T, Czirók A, Ben-Jacob E, Cohen I, Shochet O (1995) Novel type of phase transition in  
529 a system of self-driven particles. *Physical review letters* 75(6):1226
- 530 Zuberbühler K, Wittig RM (2011) Field experiments with non-human primates: a tutorial. In:  
531 Field and laboratory methods in primatology: a practical guide, Cambridge Univ. Pr., pp  
532 207–224

533 **Tables**

534 Table 1

<b>Name</b>	<b>Age</b>	<b>Rank</b>	<b>Sex</b>	<b>Siblings</b>
Dusty	10	1	m	Démon, Dubble
Démon	10	2	m	Dusty, Dubble
Jack	5	3	m	Torok, Kinai, Quebec
Torok	5	4	m	Jack, Kinai, Quebec
Kinai	5	4	m	Jack, Torok, Quebec
Quebec	5	5	m	Jack, Torok, Kinai
Dubble	10	6	m	Dusty, Démon
Gribouille	6	1	f	Siska, Babou
Babou	6	2	f	Gribouille, Siska
Siska	6	2	f	Gribouille, Babou
Laika	8	3	f	Elska, Friskies
Friskies	8	3	f	Laika, Elska
Tara	6	4	f	Banquise
Banquise	6	5	f	Tara
Elska	8	6	f	Laika, Friskies

535

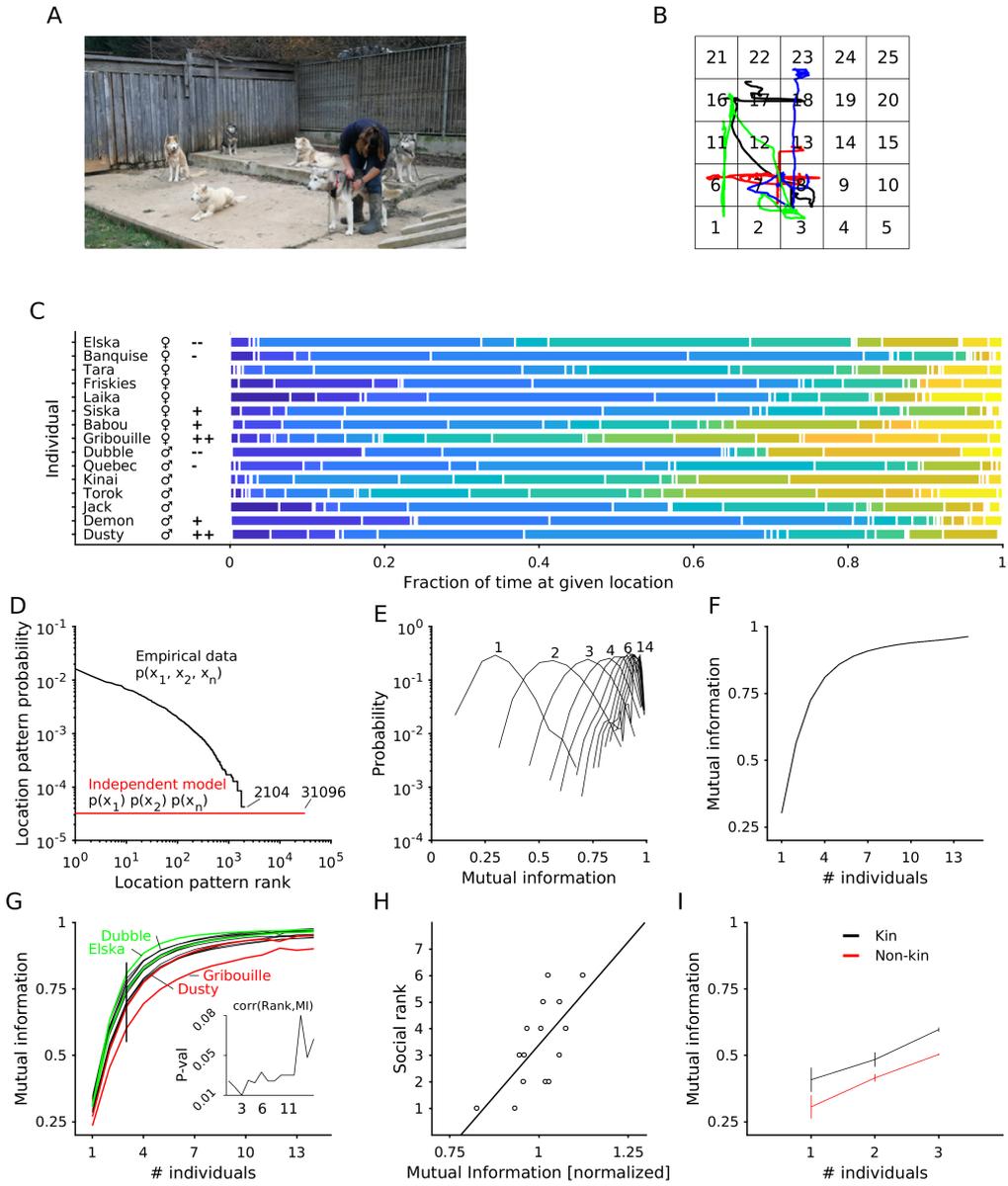
536 Table 2

	Dusty	Démon	Jack	Torok	Kinai	Quebec	Dubble	Gribouille	Babou	Siska	Laika	Friskies	Tara	Banquise	Elska
Dusty	NaN	0	0	0	0	0	0	.5	0	0	0	0	0	0	0
Démon	1	NaN	0	0	0	0	0	1	.5	.5	1	1	1	1	1
Jack	1	1	NaN	.5	.5	0	0	1	1	1	.5	.5	.5	0	0
Torok	1	1	.5	NaN	.5	0	0	1	1	1	.5	.5	.5	0	0
Kinai	1	1	.5	.5	NaN	0	0	1	1	1	.5	.5	.5	0	0
Quebec	1	1	1	1	1	NaN	0	1	1	1	1	1	1	.5	0
Dubble	1	1	1	1	1	1	NaN	1	1	1	1	1	1	1	.5
Gribouille	.5	0	0	0	0	0	0	NaN	0	0	0	0	0	0	0
Babou	1	.5	0	0	0	0	0	1	NaN	.5	0	0	0	0	0
Siska	1	.5	0	0	0	0	0	1	.5	NaN	0	0	0	0	0
Laika	1	1	.5	.5	.5	0	0	1	1	1	NaN	.5	.5	0	0
Friskies	1	1	.5	.5	.5	0	0	1	1	1	.5	NaN	.5	0	0
Tara	1	1	.5	.5	.5	0	0	1	1	1	.5	.5	NaN	0	0
Banquise	1	1	1	1	1	.5	0	1	1	1	1	1	1	NaN	0
Elska	1	1	1	1	1	1	.5	1	1	1	1	1	1	1	NaN

537

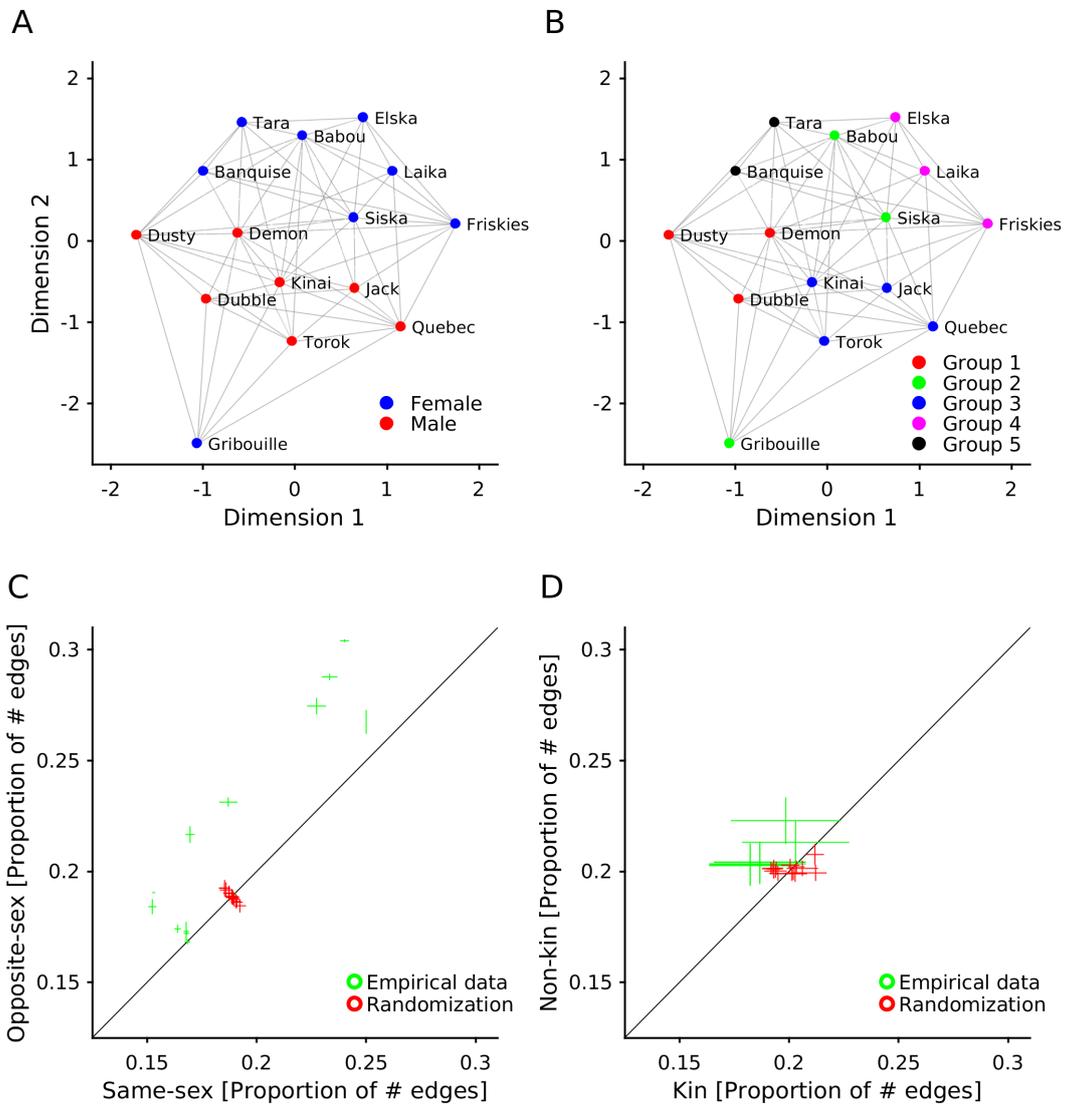
538 **Figures**

539 Figure 1



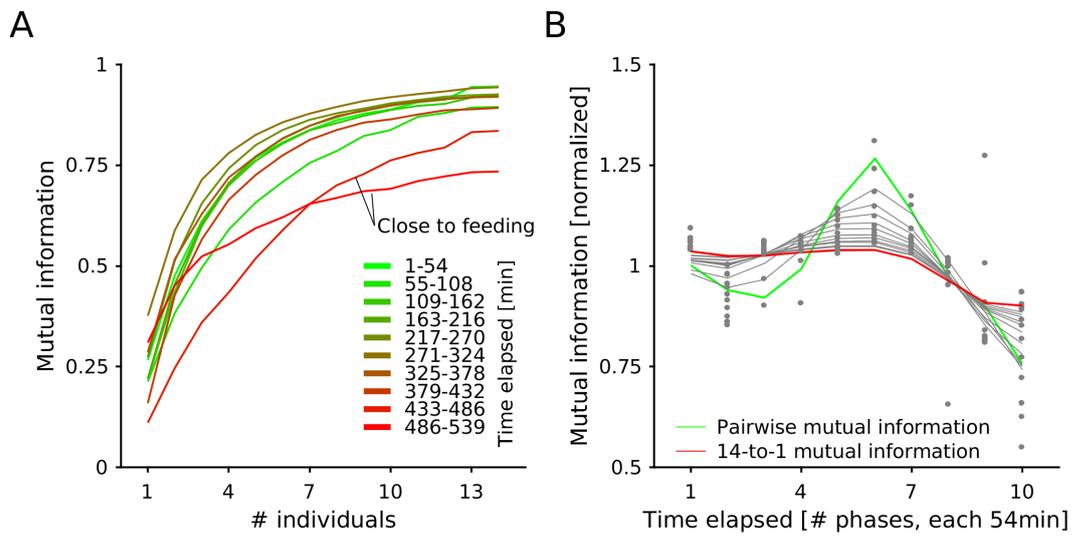
540

541 Figure 2



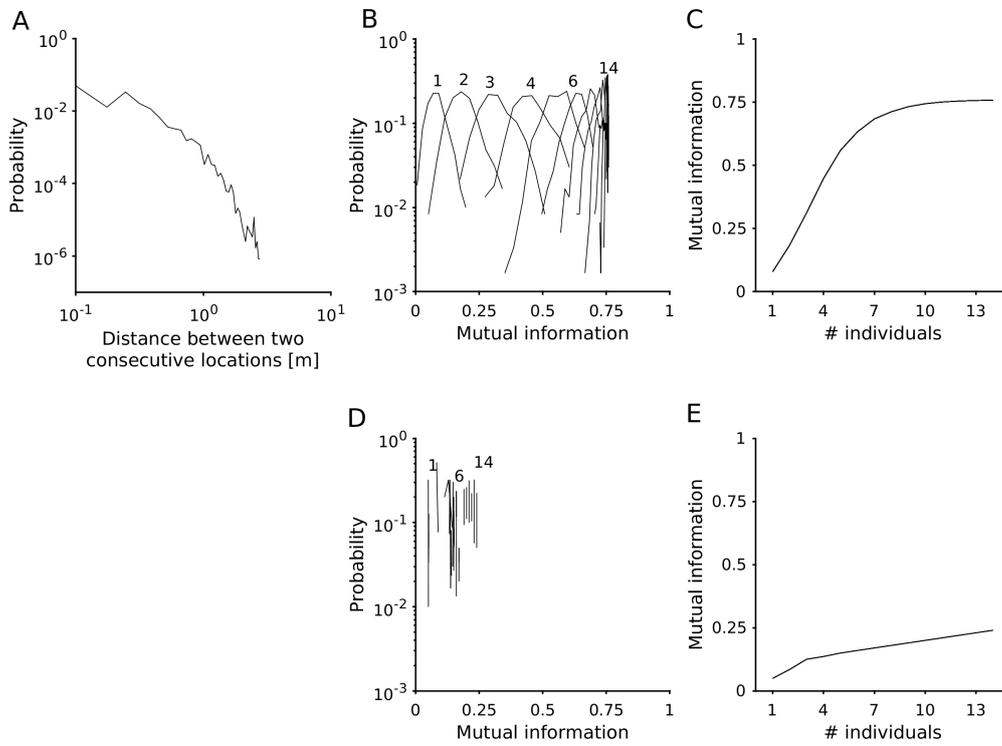
542

543 Figure 3



544

545 Figure 4



546

547 **Table captions**

548 Table 1

549 Identity and social information about individuals.

550

551 Table 2

552 Pairwise assessment of dominance (1 = dominant, .5 = equal, 0 = submissive).

553

554 **Figure captions**

555 Figure 1

556 Preparation phase and results of dog tracking. A. Preparation phase: dogs were equipped with  
 557 GPS loggers mounted to a collar. B. Example movement trajectories of three dogs overlaid onto  
 558 the 25 patch-grid. C. Fractions of time at given location were plotted for each dog. Colour-coded  
 559 the locations. Symbols indicate the sex and rank of the individuals. '+' symbols indicate  
 560 dominance; '-' symbols indicate submission. D. Location patterns were plotted according to  
 561 the number of occurrences with the most probable location pattern first and the least probable  
 562 location pattern last (x-axis) and the probability of each location pattern (y-axis). E. The  
 563 mutual information between the dog  $i$  (to be explained) and a varying number of other dogs  
 564 (ranging from 1 to 14) is shown in histograms. F. The median of each histogram (mutual  
 565 information) in E is plotted against the number of individuals used to explain dog  $i$ . G. Mutual  
 566 information is shown as a function of increasing number of individuals used to explain dog  $i$ .  
 567 Dominant (red) and submissive (green) individuals, which were subjected to be explained by  
 568 others, are highlighted. H. The social rank is plotted against and correlated with the normalized  
 569 mutual information. Normalization was done by dividing the values by the mean. I. Mutual  
 570 information is plotted against a varying number of kin or non-kin (control). Maximal number  
 571 of kin is limited to four; hence number of individuals to explain dog  $i$  is limited to three.

572

573 Figure 2

574 Social networks. A. shows a 2D map illustrating the social network of mutual influence between  
 575 members. An increasing number of links indicate an increased influence regarding dog  $i$ 's  
 576 position through others. Colour-coded is the sex of the individuals. B. Same as in A. Colour-coded  
 577 is the kinship, indicating the siblings. C. The sum of edges on the shortest path between  
 578 individuals of same and opposite sexes, as extracted from A, are shown. Means and standard

579 deviations are indicated by green lines. Red lines denote the results from the randomization  
 580 procedure. Numbers mark the number of individuals that were used for the mutual information  
 581 computation. D. Same as in C, but according to kinship.

582

583 Figure 3

584 Influence of external event. A. Mutual information values extracted from various time periods  
 585 are plotted against the number of individuals used to explain dog  $i$ . The colour-code indicates  
 586 the elapsed time from the start of the recording. ‘green’ time segments are closer to the onset of  
 587 recording, while ‘red’ time segments are close to an upcoming feeding event and therefore closer  
 588 to the end of the recording. B. Normalized mutual information is plotted against the elapsed  
 589 time from onset of recording. Normalization was done by dividing the mutual information  
 590 values by their mean. We applied a Gaussian Kernel linear regression function with a Kernel  
 591 bandwidth of 1 to describe the data.

592

593 Figure 4

594 Error estimation and null distribution. A. shows the probability distribution of distances of two  
 595 consecutive GPS measurements. B. The mutual information between the dog  $i$  (to be explained)  
 596 and a varying number of other dogs (ranging from 1 to 14) is shown in histograms. Due to GPS  
 597 errors, 25% of misplacements of location samples in the patch-grid are simulated. In reality,  
 598 this scenario is unlikely. C. The median of each histogram (mutual information) in B is plotted  
 599 against the number of individuals used to explain dog  $i$ . D. The mutual information, as in  
 600 B, is shown under the condition of permutating patch-grid assignments. E. The median of  
 601 each histogram (mutual information in D is plotted against the number of individuals used to  
 602 explain dog  $i$ .