The genetic basis for social network traits in the wild.

Jessica K. Hunter¹, Jamie Dunning¹, Alex Hoi Hang Chan²,³, Terry Burke⁴, Julia Schroeder¹

Affiliations:
¹Department of Life Sciences, Imperial College London, Silwood Park, Ascot, UK
²Centre for the Advanced Study of Collective Behaviour, University of Konstanz, Germany³
Department of Collective Behaviour, Max Planck Institute of Animal Behaviour, Konstanz, Germany
⁴Ecology and Evolution, School of Biosciences, The University of Sheffield, Sheffield, UK

Keywords: Heritability, Quantitative genetics, Social Network Analysis, Social behaviour

Correspondence: Jessica Hunter – jkhunter315@gmail.com

Abstract
Sociality is fundamental for many species, linked to an individual’s survival and reproductive success. The evolution of social behaviours has been studied, especially in complex social animals such as birds. However, discerning the genetic basis of these behaviours has been difficult due to the lack of ecological validity and life-history in controlled environments. In wild populations, however, there are the challenges of getting accurate pedigree and behavioural data. This paper used data from a long-term wild sparrow population in which we tracked the social behaviour and genetic pedigree, to assemble social networks. Here, we estimated the genetic architecture of three social network metrics—degree, strength, and betweenness—as a method of characterising social behaviour. The results show a clear genetic component, with heritability values of 8%, 6%, and 1.7% for strength, degree, and betweenness respectively. This study expands our understanding of how sociality is inherited from controlled settings to wild populations, showing a link between genetics and
social behaviour. The finding of a negative genetic correlation between strength and degree indicates a more complex genetic relationship underlying these social traits, calling for more research to understand the ecological and evolutionary impacts of these findings. Our results clearly show a genetic basis for social interactions in sparrows, inviting more research into sociality in other bird and non-bird species.

Introduction

The largely unexplored relationship between fitness, genetics, and social behaviour has garnered significant interest within the field of evolutionary biology (Bleakley et al., 2010, Lehmann & Rousset, 2014). Fitness, in evolutionary terms, refers to an individual's ability to survive and reproduce within a given environment (Orr, 2009). An individual's genetic makeup can influence trait variations that may be beneficial or detrimental to fitness in particular environments (Dunning et al., 2023). The transmission of these traits from one generation to the next is central to the conception of evolution and the basis of the study of genetics (Ellegren & Sheldon 2008).

Social behaviour, generally defined as interactions among individuals of the same species, significantly affects an individual's fitness, impacting their ability to find food, mates, and avoid predators (Riehl & Strong, 2018, Dunning et al., 2023). For instance, cooperative behaviours can enhance survival rates by aiding in predator avoidance or resource acquisition (Cornwallis et al., 2017). On the other hand, aggressive or competitive behaviours might hinder an individual's fitness if they lead to injury or social isolation (Arseneau-Robar et al., 2016, Szipl et al., 2019). This implies that social behaviour offer evolutionary advantages and thus should have gone through selection. There is evidence of intraspecific variation in social behaviour across many species, with consistent behavioural differences among individuals (Bell et al., 2009, Aplin et al., 2015, Plaza et al., 2019).
If there is selection for these behaviours, then, if they are also heritable they can evolve. Determining the genetic basis of social behaviour is pivotal, because it contributes to our understanding of the evolution of sociality, and it can elucidate how social behaviours contribute to an individual's fitness and, by extension, the evolutionary trajectory of species (Wice & Saltz, 2021, Hardie & Cooney, 2022). Moreover, knowing the genetic architecture of behaviour traits has practical applications for breeding and conservation efforts (Blumstein, 2010).

While some studies have explored the genetic basis of social behaviour in mice and invertebrates, these investigations have largely been conducted in controlled laboratory environments (Hammock & Young, 2005, Wice & Saltz, 2021). It is inherently difficult and expensive to obtain complete and accurate pedigree data of freely roaming, and potentially intermingling populations. While laboratory experiments have their merits, they however cannot replicate real-world situations and often do not represent the genetic diversity of wild populations (Calisi & Bentley, 2009). Furthermore, laboratory studies cannot incorporate population processes and life-history data (Pemberton, 2008). As a result, we lack data on the genetic architecture of sociality from wild populations.

To address this knowledge gap, this paper investigates the heritability of sociality in wild House sparrows (*Passer domesticus*) living on an isolated island off the coast of the UK. Here, migration to and from the island is minimal, allowing for following all individuals throughout their lives (Schroeder *et al*., 2011). The project's uniqueness lies in its opportunity to study the heritability of social behaviour in a wild population, which is only possible due to the availability of near-complete census and long-term nature of the study (Schroeder *et al*., 2011).

In this study, we use social network analysis (farine whitehead), to study and quantify sociality. Metrics calculated using social network analysis have been evidenced as a good proxy for personality traits in House sparrows, specifically in the population on Lundy Island.
(Plaza et al., 2019, Farine & Whitehead, 2015). This paper hypothesises that these traits are heritable in House sparrows.

Methods

Data collection

The data for the study was obtained from a long-term investigation conducted on Lundy Island (51°10′N, 4°40′W), located in the Bristol Channel. The study species was the House sparrow, a commonly used model organism known for its extensively studied biology and life history (Hansen et al., 2020). House sparrows are known to form social flocks during non-breeding periods and gather around food sources, making them an ideal species for studying social behaviour (Tuliozi et al., 2018).

Being located 19km off the North Devon coast limits migration, together with a near-complete, annual census and breeding monitoring, enabled us to build a comprehensive database of life-history traits of every individual in the population (Schroeder et al., 2011, Schroeder et al., 2015). DNA samples were collected from the sparrows after capturing birds in their natal nest box, or through mist netting. These DNA samples, along with regular nest monitoring, are used to build the genetic pedigree, which was constructed using <22 microsatellite markers (for details see Ockendon et al., 2009, Schroeder et al., 2012 and Dunning et al., 2023). At the time of writing, the pedigree spans 20 years, from 2000-2019.

Each sparrow is equipped with a distinctive combination of coloured rings and a Passive-Integrated Transponder (PIT tag, TROVANID100, for details see Schroeder et al., 2011). A Radio Frequency Identification (RFID) antenna measuring 19.8 x 19.8cm (DorsetID) was affixed to a seed reservoir positioned that is regularly visted by the sparrows. The antenna recorded instances of a tagged sparrow entering the feeder, as described in Dunning et al.
The reservoir design enables multiple birds to feed simultaneously, and PIT tags are recorded every ¼ of a second.

Social network analysis

Social network analysis is a process that quantifies the connections and interactions between individuals within a social group, represented as nodes (individuals) and edges (connections/relationships) in a network graph (Otte & Rousseau, 2002). We created social networks using presence data obtained from RFID readings during the winter non-breeding (NB) period of the specific year. We chose this approach to account for seasonal variations in social behaviour, as sparrows exhibit territorial behaviour during breeding seasons (Tuliozi et al., 2018). Data during the NB period was collected for 2013, 2015-16, 2016-17, and 2020, resulting in a total of 37 weeks of data in the final dataset.

We used the arrival time method to define non-random associations between individuals, where an overlap in arrival time at the bird feeder ($\Delta t$) is defined as an association (Dunning et al., 2023). This method is ideal for House sparrows, as other common methods tend to overestimate associations for gregarious species that aggregate around resources (Dunning et al., 2023, Ferreira et al., 2020). This method was also developed specifically on the Lundy Island sparrow population and their behaviour at feeders (Dunning et al., 2023). The arrival time method assumes that individuals who arrive at the feeder together are socially associated and were likely together before arriving (Dunning et al., 2023b).

We conducted the computation of social networks using the arrival method in R and RStudio (R Core Team, 2023, RStudio team, 2020) with a custom function, openly available in the Zenodo repository (Chan and Dunning, 2023). The function incorporates two parameters: $\Delta t$, representing a predefined time interval that starts when an individual’s arrival at the feeder, and $\Delta i$, the time after which the individual is considered to have left the feeder. By assigning $\Delta t$ to each arrival, the method considers only birds arriving within one another’s $\Delta t$ as associated, regardless of their duration at the feeder (Dunning et al., 2023). Individuals could only be recorded again after the $\Delta i$ period. $\Delta t$ and $\Delta i$ were assigned fixed values of 150s and
300s, respectively, as these intervals are most effective at capturing non-random associations of individuals who arrive in a group (Dunning et al., 2023).

We processed the data using the arrival time functions to construct weighted, non-directional social networks using the R package igraph (Csardi & Nepusz, 2006). To incorporate a greater amount of data into the wild animal model, we built a social network for each week. We used three node-based metrics (hereafter centrality measures), degree, strength and betweenness as measurements of sociality. Degree quantifies the number of associations an individual has with others and could be interpreted as how many friends an individual has. Strength further considers the intensity or frequency of those relationships, or how close an individual is to their friends. Betweenness is the number of shortest paths between other individuals in the group that pass through the focal individual, quantifying how central an individual is to the group (Sosa et al., 2020).

We computed degree, strength, betweenness and density with the igraph package (Csardi & Nepusz, 2006). Density, a global network metric referring to the ratio of associations an individual has over the number of potential associations it could have, was also calculated, as its use during modelling enables better comparison between different social networks (Hart, 2023).

Animal model

We utilized an animal model to quantify quantitative genetic parameters. Specifically, a generalised linear model using Markov Chain Monte Carlo (MCMC) approximation methods was used, with the MCMCglmm package in R (Hadfield et al., 2010). Animal models are a form of mixed-effect linear regression model, leveraging genetic information to estimate genetic correlations between traits as additive genetic variance (Wilson et al., 2010). Additive genetic variation is the measurable variance in a trait due to the combined effect of shared, multiple alleles (Kruuk et al., 2014). In the animal model, the additive genetic effect is treated as a random effect by incorporating the inverse relatedness matrix of the pedigree into the regression. This matrix mathematically portrays the extent of
genetic relatedness between individuals in a population, quantifying genetic relationships based on the pedigree.

The additive genetic effect is assumed to have a variance of $V_A$ (Kruuk et al., 2014). Any residual variance in phenotype is labelled as $V_R$, culminating in a total phenotypic variance of $V_A + V_R$ (Eq. 1, Dingemanse & Dochtermann, 2014), with $V_P$ being the total phenotypic variance.

\[
V_P = V_A + V_R
\]

The pedigree data obtained from the Lundy sparrow research team was processed to include only the columns of offspring, dam, and sire. A proportion of dams and sires were not present as offspring; these were added as offspring with parents labelled as NA. 10 individuals in the social data were not present in the pedigree, and excluded from the dataset, resulting in a total of 544 informative individuals and 816 observations.

A tri-variate MCMCglmm was used as the animal model. Degree, strength and betweenness were collectively modelled as the response variable. The inverse relatedness matrix of the pedigree, denoted as “animal” in MCMCglmm, was modelled as a random effect. Bird ID was also included as a random effect to account for permanent environment effects (differences between individuals not driven by additive genetic effects). Additionally, the network metric density was introduced as a fixed effect to improve model fit. Degree and betweenness were z-transformed to standardise the model output and improve model convergence. Strength was not scaled since it has a range of 0-1. All three centrality measures were modelled under a gaussian distribution. The modelling approach for pedigree and residuals was unrestricted to capture any potential covariance between traits arising from relatedness and to explore potential interactions among these traits within individuals. Priors were set as a matrix, with ones in the diagonal and zeroes elsewhere and given a minimal effect on posterior probability ($\nu = 0.002$). 1,000,000 iterations, a burn-in of 300,000, and a thinning interval of ten were used as the parameters for the tri-variate model.
Analysis

For each random effect outcome within the animal model, a variance-covariance (VCV) matrix was generated. The total phenotypic variance $V_P$ was calculated as the sum of these VCV matrices (Eq. 2, Wilson et al., 2010). The VCV matrix for permanent environment effects ($V_{PE}$) was represented as variance among individuals (modelled as random effect bird ID, a unique identifier of the birds).

$$V_P = V_A + V_{PE} + V_R$$

- $V_{PE} = \text{Permanent environment effects}$

These VCV matrices were used to calculate the narrow sense heritability ($h^2$) and repeatability of each network trait. $h^2$ quantifies the proportion of the variation in a trait within a population that is influenced by additive genetic factors (Wilson et al., 2010). $h^2$ is calculated as the fraction of phenotypic variance attributable to additive genetic variance, which can be represented as a percentage (Eq. 3).

$$h^2 = \frac{V_A}{V_P}$$

Repeatability refers to the fraction of phenotypic variance that remains consistent across multiple observations in an individual and indicates the stability of a trait (Wilson et al., 2010). It is often regarded as the upper limit of $h^2$, as it represents the total phenotypic variance observed among individuals, which includes additive genetic variance. Repeatability is calculated as the sum of overall individual variance $V_{ind}$ over total phenotypic variance of a trait (Eq. 4, Dingemanse & Dochtermann, 2014).
Moreover, to assess potential trait covariation, the phenotypic covariance between centrality measures resulting from additive genetic effects (COV\(_A\)) was used to compute genetic covariance (Wilson et al., 2010). We then computed the genetic correlation from the genetic covariance, denoted as \(r_A\) (Eq. 5, Åkesson et al., 2008), where COV\(_{A(x,y)}\) is the additive genetic covariance between trait x and y.

\[
\begin{align*}
\text{Eq. 4} & \quad \text{Repeatability} = \frac{V_{\text{ind}}}{V_p} = \frac{V_A + V_{PE}}{V_p} \\
\text{Eq. 5} & \quad r_A = \frac{\text{COV}_{A(x,y)}}{\sqrt{V_A(x)} \sqrt{V_A(y)}}
\end{align*}
\]

**Results**

Significant additive genetic variance was observed for all three centrality traits (Table 1). The covariance between traits due to pedigree was statistically significant for strength and degree (-0.0077, 95% confidence interval = -0.0149 – -0.0003).

Variance due to permanent environment was also statistically significant for all three traits (Table 1). Additionally, the residual variance and the covariance between traits was significant for all three traits. The MCMCglmm model and its variables met all assumptions, and diagnostic plots indicated convergence and independent sampling. The deviance information criterion was estimated to be 3,621. MCMC model outputs are given in Appendix 1.

| Table 1. Variance-covariances matrices of the animal model. 95% confidence intervals are indicated within the brackets. Genetic and phenotypic covariances are given above the diagonal, while correlations are provided below. The diagonal displays additive genetic variance first, followed by percentage variance. |
### Additive genetic variance

<table>
<thead>
<tr>
<th></th>
<th>Degree</th>
<th>Strength</th>
<th>Betweenness</th>
</tr>
</thead>
<tbody>
<tr>
<td>Degree</td>
<td>0.0628 (0.0054 - 0.1226), 2.42% (0.21% - 4.72%)</td>
<td>-0.0077 (-0.0149 - 0.0003)</td>
<td>0</td>
</tr>
<tr>
<td>Strength</td>
<td>-0.0254 (-0.0492 - -0.0009)</td>
<td>0.0028 (0.0007 – 0.0051), 0.11% (0.03% - 0.20%)</td>
<td>0</td>
</tr>
<tr>
<td>Betweenness</td>
<td>0</td>
<td>0</td>
<td>0.0174 (0.0017 – 0.0399), 0.67% (0.06 – 1.53)</td>
</tr>
</tbody>
</table>

### Permanent environment effects

<table>
<thead>
<tr>
<th></th>
<th>Degree</th>
<th>Strength</th>
<th>Betweenness</th>
</tr>
</thead>
<tbody>
<tr>
<td>Degree</td>
<td>0.0194 (0.0002 – 0.0691), 0.75% (0.01% - 2.66%)</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>Strength</td>
<td>NA</td>
<td>0.0021 (0.0003 – 0.0042), 0.08% (0.01% - 0.16%)</td>
<td>NA</td>
</tr>
<tr>
<td>Betweenness</td>
<td>NA</td>
<td>NA</td>
<td>0.0067 (0.0002 – 0.2043), 0.26% (0.01% - 0.79%)</td>
</tr>
</tbody>
</table>

### Residual variance

<table>
<thead>
<tr>
<th></th>
<th>Degree</th>
<th>Strength</th>
<th>Betweenness</th>
</tr>
</thead>
<tbody>
<tr>
<td>Degree</td>
<td>0.9621 (0.8597 – 1.0663), 37.05% (33.11% - 41.06%)</td>
<td>0.1145 (0.0964 – 0.1326)</td>
<td>0.3265 (0.2543 – 0.4025)</td>
</tr>
</tbody>
</table>
Heritability of degree, strength, and betweenness were determined to be 6.01%, 7.97% and 1.68% respectively (Table 2). Genetic correlation between degree and strength was -0.8169 (Table 2). As the covariance between degree and betweenness and betweenness and degree were not significant they were assumed to be 0.

Table 2. Heritability, Repeatability and Total Phenotypic Variance (Total) Values, along with Genetic Correlation between Centrality Measures.

<table>
<thead>
<tr>
<th>Centrality Measure</th>
<th>Heritability</th>
<th>Repeatability</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Degree</td>
<td>0.0601</td>
<td>0.0795</td>
<td>1.0442</td>
</tr>
<tr>
<td>Strength</td>
<td>0.0797</td>
<td>0.0818</td>
<td>0.0349</td>
</tr>
<tr>
<td>Betweenness</td>
<td>0.0168</td>
<td>0.0235</td>
<td>1.0338</td>
</tr>
</tbody>
</table>

Genetic correlation

- Degree – Strength  -0.8169
- Degree - Betweenness 0
- Strength - Betweenness 0

Discussion
The results of this study provide compelling evidence for a genetic basis underlying social centrality traits in sparrows, as all three centrality measures showed significant variance attributed to pedigree, showing that more closely related individuals had more similar trait measurements. This finding supports the idea that social behaviour in sparrows is heritable, since social centrality measures are reliable proxies for social behaviour (Plaza et al., 2019). The repeatability estimates provide strong support for the existence of consistent individual differences in social behaviour over time, as evidenced by significant and robust variation between birds.

Strength appeared to be the most significantly influenced by pedigree, with a heritability estimate of 8% (Table 2). On the other hand, betweenness was least affected by pedigree, with a heritability estimate of 1.7%. These findings could suggest that the quality and quantity of social interactions has a stronger genetic basis than the centrality of an individual in its population. Furthermore, heritability made up a significant proportion of repeatability.

The heritability of betweenness made up 70.83% of the repeatability of betweenness, whilst heritability of strength was 97.56% of the repeatability. This suggests that most of the variability in social network traits could be attributed to relatedness. Further research and exploration of the underlying genetic mechanisms could reveal the intriguing interplay between genetics and social behaviour.

An intriguing observation was the strong negative genetic correlation (-0.8169) between strength and degree (Table 2). This indicates that degree and strength co-vary in sparrow populations, and that individuals with more, but weaker connections, tend to be related to sparrows with fewer, but stronger associations. This might also suggest the presence of pleiotropy, where a single genetic locus influences both strength and degree, or a potential genetic linkage between the underlying genes for these two traits. This discovery merits further investigation to explore the underlying mechanisms and potential ecological and evolutionary implications.
While the variance attributed to pedigree was found to be statistically significant, the effect sizes for degree and betweenness were relatively small, measuring only 0.0628 and 0.0174 standard deviations from the mean, respectively (Table 1). The genetic variance for strength was also small (0.0028). The relatively wide confidence intervals, ranging from approximately 0.0017 to 0.1226 for all variance estimates, further underscore the modest effect sizes observed. Although the genetic influence on social centrality traits is significant, caution should be exercised in interpretations, and further validation is warranted. Moreover, the residuals exhibited large variances, surpassing the variation attributed to pedigree and individual differences. These significant residuals, which additionally had tighter confidence intervals, suggest the existence of additional unexplored factors that significantly influence centrality measures beyond those considered in the current model.

To address this, alternative models with different variables could be used to find other factors that influence social centrality measures, which could reduce residual variance and improve model fit. For example, including week as a random effect. When collecting a new dataset for social research, the age and sex of birds should be recorded and used since individuals of different ages and gender may interact differently. Additionally, social networks could be constructed on different time intervals, for example per month, or with intervals of a constant sample size, to determine if these factors change the findings found in this study.

Given more time, a null model employing randomized data sets could be run to validate the observed results (Farine, 2017). As the confidence intervals are relatively wide, assessing whether null results yield different confidence intervals would add further robustness to the significant findings.

Another aspect to consider is whether the variation in social traits we’ve observed is due to genetic factors or simply a result of closely related individuals sharing similar environments. This question requires more exploration, which could involve methods like cross-fostering or using data that contains extra-pair offspring. Furthermore, conducting models using data
collected during breeding seasons could offer valuable insights into the significance of social behaviour in different contexts.

In conclusion, despite the small impact and relatively large confidence intervals, there is compelling evidence that sociality is heritable in-house sparrows, adding valuable insights into the genetic basis of social behaviour. This finding warrants further investigation to understand the underlying mechanisms driving social behaviour.

**Statements and Declarations**

**Data Archiving Statement**

The data used in the study was provided by the Lundy Sparrow project, a long-term research initiative carried out by the Schroeder laboratory, run by Julia Schroeder, at Imperial College London, in collaboration with Sheffield University and Macquarie University. The data was largely cleaned before it was accessed for this study. Data sets will be available upon acceptance of the manuscript. The R script can be accessed at https://github.com/JKHunter315/GeneticBasisSocialitySparrowsICL. The arrival time method of constructing social networks was developed by Jamie Dunning and Alex Chan, from the Schroeder laboratory at Imperial College London, available publicly on Zenodo at https://zenodo.org/records/7527440 (Chan & Dunning, 2023).

**Competing interests**

The author declares no competing interests.

**References**


Dunning, J. et al. (2023) ‘Opposite-sex associations are linked with annual fitness, but sociality is stable over lifetime’, *Behavioral Ecology*, p. arac124. Available at: https://doi.org/10.1093/beheco/arac124.


R Core Team (2023). R: A Language and Environment for Statistical Computing. R


Wice, E.W. and Saltz, J.B. (2021) ‘Selection on heritable social network positions is context-dependent in Drosophila melanogaster’, *Nature Communications*, 12, p. 3357. Available at: [https://doi.org/10.1038/s41467-021-23672-1](https://doi.org/10.1038/s41467-021-23672-1).


**Appendix**

**Appendix 1. Summary statistics of each random effect. Covariances of interactions are displayed as trait x–trait y. The values in the table are scaled by a factor of 100 to accommodate the small effect sizes of the output. Significant values are denoted by *.*

<table>
<thead>
<tr>
<th>Summary Statistics</th>
<th>Pedigree</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean variance</td>
</tr>
<tr>
<td>Degree</td>
<td>6.28*</td>
</tr>
<tr>
<td>Strength</td>
<td>0.28*</td>
</tr>
<tr>
<td>Betweenness</td>
<td>1.74*</td>
</tr>
<tr>
<td>Degree - Betweenness</td>
<td>-2.26</td>
</tr>
<tr>
<td>Degree - Strength</td>
<td>-0.77*</td>
</tr>
<tr>
<td></td>
<td>Mean variance</td>
</tr>
<tr>
<td>----------------------</td>
<td>---------------</td>
</tr>
<tr>
<td><strong>Strength - Betweenness</strong></td>
<td>0.46</td>
</tr>
<tr>
<td><strong>Bird ID</strong></td>
<td></td>
</tr>
<tr>
<td>Degree</td>
<td>1.94*</td>
</tr>
<tr>
<td>Strength</td>
<td>0.21*</td>
</tr>
<tr>
<td>Betweenness</td>
<td>0.67*</td>
</tr>
<tr>
<td><strong>Residuals</strong></td>
<td></td>
</tr>
<tr>
<td>Degree</td>
<td>96.21*</td>
</tr>
<tr>
<td>Strength</td>
<td>3.00*</td>
</tr>
<tr>
<td>Betweenness</td>
<td>1.74*</td>
</tr>
<tr>
<td>Degree - Betweenness</td>
<td>32.65*</td>
</tr>
<tr>
<td>Degree - Strength</td>
<td>11.45*</td>
</tr>
<tr>
<td>Strength - Betweenness</td>
<td>5.07*</td>
</tr>
</tbody>
</table>