High pathogenicity avian influenza virus transmission is shaped by multi-species social network structure

Jamie Dunning[1] (0000-0001-8234-8526), Anna Gamża[2] (0000-0003-3243-4543), Josh A. Firth[1,3] (0000-0001-7183-4115), Adham Ashton-Butt[1,4,5] (0000-0002-6926-6099), William T. Harvey[2] (0000-0001-9529-1127), Samantha J. Lycett[2] (0000-0003-3159-596X), Rowland R. Kao[2,6] (0000-0003-0919-6401), Ian H. Brown[7] (0000-0003-3909-1301) and *Alastair I. Ward[1] (0000-0002-3305-3323).

[1] School of Biology, University of Leeds, Leeds, United Kingdom
[2] The Roslin Institute, University of Edinburgh, Edinburgh, United Kingdom
[3] Department of Biology, University of Oxford, Oxford, United Kingdom
[4] British Trust for Ornithology, Thetford, United Kingdom
[5] School of Environmental Sciences, University of East Anglia, Norwich, United Kingdom
[6] School of Physics and Astronomy, University of Edinburgh, Edinburgh, United Kingdom
[7] The Pirbright Institute, Pirbright, United Kingdom
*jamiedunning8@googlemail.com

Abstract: The emergence of zoonotic and epizootic disease has had devastating consequences for human and animal health, including wildlife conservation. Yet, surveillance of multi-host disease systems is particularly challenging due to complex transmission pathways across many species. Social network analysis has been applied to simple transmission systems, but empirical applications to wild, multi-species systems are scarce. Here, we combined high pathogenicity avian influenza (HPAI) viral genomes, a zoonotic virus of pandemic potential, with a large citizen-science database of wild bird co-occurrence to test how multi-species social network structure predicts transmission dynamics. We calculated 1,687 pairwise genetic distances from 356 viral genomes across 33 host species and related these to network metrics. Species pairs with greater co-occurrence frequency showed significantly lower viral genetic distances, over and above relative species association. Our results demonstrate that social network models can predict zoonotic pathogen transmission in wild multi-species systems, enabling targeted disease surveillance.

Introduction: Emerging zoonotic and epizootic pathogens that spill over from wildlife are a persistent, and often intractable, threat to human and animal health (Kruse 2004; Jones et al. 2008) with far-reaching consequences for global economic and food security (Bernstein et al. 2022). Many such diseases, including Ebola and SARS-CoV-2, have originated from multi-species groups of wild animals, often without detection in those populations prior to disease emergence in humans or livestock (Marí et al. 2015; Holmes 2022; Xie et al. 2023). This is particularly true in multi-host-pathogen systems like high pathogenicity avian influenza viruses (HPAI) in wild birds, where ecological and behavioural diversity across many host species gives rise to complex transmission pathways (Bansal et al. 2007; Buhnerkempe et al. 2015; Dunning et al. 2025), presenting a challenge for proactive surveillance (Buhnerkempe et al. 2015). In communities of non-human animals, patterns of disease transmission are highly variable and cannot be solely explained by pathogen virulence. Instead, just as human mobility and interaction shape the trajectory of an epidemic (Firth et al. 2020; Kostandova et al. 2024), animal social network structure, patterns of association between individuals and species shaped by the social (Bansal et al. 2007; Sah et al. 2017; Albery et al. 2021) and physical environment (Carlson et al. 2022; King et al. 2023), influence epizootic disease propagation. For example, during the HPAI outbreak of 2021/22, H5N1 clade 2.3.4.4b began infecting European seabirds. The subsequent redrawing of host transmission networks through new flyways and behavioural ecologies had devastating consequences for the conservation of wild birds (Falchieri et al. 2022; Tremlett et al. 2025) and drove incursion of HPAI to new continents (Kandeil et al. 2023) and novel mammalian hosts (Sidik 2023; Plaza et al. 2024; Sah et al. 2024).
The mechanism by which host traits, or environmental components influence disease transmission is often by shaping social network structures - forcing potential hosts together or driving them apart. This spatial and temporal clustering either facilitates or constrains non-vector borne pathogen transmission through social links (Hamede et al. 2009).

Research into animal disease systems has typically focused on single-species or trait-based approaches, such as examining population dynamics (Vickers et al. 2024; Gamża et al. 2024) or host behaviour (Yin et al. 2023; Luis et al. 2015; Olival et al. 2017; McDuie et al. 2024) in isolation. Understanding how multi-species interactions shape disease spread therefore represents an important knowledge gap for informing zoonotic and epizootic disease surveillance (Buhnerkempe et al. 2015; Sah et al. 2018; Albery et al. 2021; Dunning et al. 2025). Evaluation of multi-host network disease dynamics has remained reliant on theoretical models (Albery et al. 2020; McDuie et al. 2024), while surveillance efforts are limited by reactive sample collection, taxonomic sampling biases, and processing constraints (Wade et al. 2022; Atkinson et al. 2025). Complex network approaches offer promise (Dunning et al. 2025) but have so far only been applied to simpler transmission systems in relatively few species (Hamede et al. 2009; Gómez et al. 2013; Wilson-Aggarwal et al. 2019; Proboste et al. 2024).

Here, we used data from HPAI outbreaks in Great Britain (GB), a serious public health threat with trans-boundary transmission between hundreds of interacting hosts and multiple transmission pathways (Xie et al. 2023; Reid et al. 2024) to test the relationship between multi-species social links and disease transmission. We defined social links broadly, encompassing both intentional social interaction and non-social

aggregation. We combined social network analyses with viral phylogenetics and Bayesian multilevel modelling to test our prediction that social links predict shorter distance between viral genomes (as a proxy for transmission probability) in multi-species networks of wild birds. To capture transmission pathways, we used fine-scale spatiotemporal co-occurrence between commonly occurring waterbird species, derived from national monitoring data. Previous phylodynamic studies have matched H5N1 infection in poultry with identical genotypes in nearby wetland sediments (Giacinti et al. 2024), implying landscape transmission without identified pathways. Our results demonstrate that patterns of co-occurrence drive transmission probability at landscape scales, providing scarce empirical evidence that network structure governs pathogen spread in multi-species groups of wild animals.

Materials and Methods: To test the hypothesis that host species' social links predict viral genetic distance across multi-species networks of wild birds, we combined social network analysis with phylodynamic and Bayesian mixed modelling approaches. We also tested if our results differed from those derived from an expected network, without social structure, by running a series of permutations that mirrored our observed data models.

Survey effort and Networks: To build species co-occurrence networks, we accessed five years of data from the British Trust for Ornithology Wetland Bird Survey (WeBS; Woodward et al. 2024), 2018 - 2022 for two overlapping regions of East Yorkshire, Great Britain, each defined by a 50km circular buffer around two recurrent HPAI outbreak sites. In total 3,722 individual surveys, including 159 species. WeBS follows a standardized survey methodology and aligns surveys with high tides to maximise detection probability of priority wetland birds.
We built a single network for each year, where the nodes represented species recorded within a WeBS survey (an event), and their interconnecting edges represented a shared event between a species pair. We combined these networks into a global network representing co-occurrence in the landscape across the study period, comprising 8,162 interconnecting edges between 159 species (for a summary see Table 1a-c).

From these networks we extracted two established dyadic (pairwise - between two species) measures of association. To measure the extent to which two species co-occur in the landscape, and therefore, may be expected to share viruses more easily, we used pairwise edge weight, or simple connection strength, which represented the absolute frequency of co-occurrence in the network. Then, to capture the extent to which a species pair is associated in space and time, accounting for absolute co-occurrence frequency relative to the opportunity for co-occurrence, we also used the Simple Ratio Index (SRI; Farine & Whitehead 2015).

SRI represents the connectedness of a species pair relative to opportunity for association, over-and-above the effect of abundance. A greater SRI is therefore a measure of species clustering probability in the social network.
For example, in the WeBS data, the highest SRI values were between relatively scarce species whose habitats overlap, in this case, Dipper *Cinclus cinclus* and Grey Wagtail *Motacilla cinerea*, which both use fast flowing upland streams. Although they were both rare in the landscape, they were almost always recorded within the same

recording events. We calculated each measure from the global network, before filtering and scaling against our viral data.

Network similarity: We tested the assumption that our networks were comparable across years - that the dyads that make up the general network structure were similar across our five years of data collection. We did this using two measures designed to compare non-uniform networks, in this case, with variation in the species recorded (the network actors, nodes) between years: 1) Jaccard distance describes the proportion of shared dyads between two networks, regardless of weight; and, 2) a Spearman rank correlation test to measure distance of dyadic edge weights between species pairs present in both networks. In both cases 1 denotes identical network structure, and 0 denotes complete difference (Table 1b).

Viral genetic distance: We accessed HPAI viral genomes collected in Great Britain from the Global Initiative on Sharing All Influenza Data (GISAID; Shu & McCauley 2017). We associated each sequence with host information and filtered out mammals, domestic birds and taxonomically vague entries (none of which are recorded as part of WeBS surveys), as well as non-H5N1 strains (H5N5 and H5N8), leaving 356 viral genome sequences across 33 host species. After restricting to species present in both the WeBS network and viral datasets, we obtained 1,687 pairwise genetic distance comparisons.
We aligned each sequence per segment using MAFFT v7.51149, trimmed them to coding regions, removed insertions present in fewer than 10% of sequences and rejected sequences with nucleotide coverage lower than 80% for all segments. We concatenated segment sequences and compared sequences with all others in the database to quantify a genetic distance between pairwise viral genomes using the Tamura and Nei 1993 model (Tamura & Nei 1993) in the ape package for R (Paradis et al. 2019; R Core Team 2025) that calculates genetic distance as a proportion that differ between each pair of sequence, accounting for difference in rates of transition and transversion mutations. While we limited our viral phylogenetic data to the H5N1 subtype, we identified genomes that had acquired genomic segments from influenza viruses of other low pathogenicity sub-types. Thus, there were two mechanisms of viral evolution influencing genetic distances among these H5N1 viruses: small, incremental changes driven by gradual accumulation of mutations during host-to-host transmission, and greater genetic shifts where reassortment occurs between co-infecting viruses. To test if social network centrality predicted transmission probability, we limited our study to these smaller incremental changes by using a measure of partial genetic distance (hereafter genetic distance), using only non-reassorting haemagglutinin, neuraminidase and matrix protein genome fragments (Harvey et al. 2025). We recorded the date associated with each sampling event and a location, provided by the Animal and Plant Health Agency, that is, measures of geometric (space between sampling events) and temporal (time between sampling events) distance (Gamża et al. 2024; Harvey et al. 2025)

Model fit: To test the relationship between wild bird social network structure and genetic distance between HPAI viruses (the response variable), and to accommodate multi-membership sampling in our species dyads, we ran a series of Generalized Linear Mixed Models (GLMMs) using Bayesian Markov Chain Monte Carlo (MCMC) simulations with the Hamiltonian Monte Carlo (HMC) and No-U-Turn Sampler (NUTS) algorithms in the brms package in R (R Core Team 2025; Bürkner 2017).

To capture transmission events, which are potentially rare in the wild, with divergent viral genomes we used the maximum genetic distance as a conservative estimate of genetic distance between a species pair, but, because samples were positively correlated with maximum genetic distance ($r = 0.52$), we controlled for the number of samples as a fixed effect in our models. We also confirmed that our network metrics and the number samples were only weakly correlated (Spearman's $\rho = 0.19$), indicating these variables captured largely independent sources of variation.

Otherwise, our models only differed the first predictor (fixed effect) variable, which was either pairwise weight or SRI. For all models, we used weakly informative priors, and ran each over 4 chains of 4,000 iterations, with a warm-up of 1,000 iterations and a thinning interval of 1. We used Effective Sample Sizes of >5000 in the posterior and tail distributions to maintain sampling efficiency and a scale reduction factor ($\check{R}$) values of 1 to ensure convergence. We also manually checked posterior trace plots to ensure model mixing. The fixed effects were considered statistically significant when the 95% credible interval (CI) excluded zero. Because our study focused on transmission pathways, rather than host susceptibility to infection we did not constrain the phylogenetic relationship between hosts in our models.

Weight model: We modeled the pairwise weight of birds present in both the WeBS and the viral phylogenetic data, maintaining weights and SRI measures from the global social network. We included the mean of geometric distance between viral pairwise sample locations and the mean of pairwise temporal distance (time elapsed between sampling events) as fixed effects. We used the mean (rather than the maximum) of geometric- and temporal distance to account for wide variation in sampling and to reduce marginal collinearity. We z-transformed all fixed effects to better align scales between other variables in the model. Because most species appeared across many dyads, we accounted for species specific effects on genetic distance independently of their dyad (i.e. species A + species B, rather than species AB) by including species A and species B as a linked 'multi-membership' variable inside our random effects. This approach also avoided over-dispersion.

SRI model: We repeated the first model, using the same model structure but with SRI instead of pairwise weight, as a fixed effect. We used the log-odds (using a logit link function) of SRI to better accommodate a linear relationship between SRI and other predictors in the model.

Permutations: In order to test that our results differed from the expected distribution without social structure in the network, we repeated our models with a series of network permutations. We did this by 're-wiring' the observed network, shuffling the edges between species pairs while maintaining the degree distribution for each species node (Farine 2017). We rewired networks 1000 times within years, before combining them into a global network within each permutation cycle. Edges were rewired between species nodes, breaking the structure of the observed network, but, because some species are much more frequent in our data -and therefore our networks have some strong ties- we sampled edges from the empirical distribution of edge weights (Farine 2017). This controlled for the chance that some effects are the result of some species having more interactions over all, rather than who those interactions were with. We maintained the structure of our original models and ran the permutation over 100 cycles. We inferred significance, that the effect of social

structure on genetic distance in our observed models was stronger than expected under a system of random mixing, where our observed credible intervals fell outside of the 95% confidence intervals of the permuted estimate.

Results:

We derived a global social network from 3,722 individual wetland bird surveys over five years (for a summary see Table 1a-c), comprised of 8,162 species dyads between 159 commonly occurring wetland bird species. Of these, 33 species were represented in both the social network and the viral genomic dataset (See supplementary Table 1 for GISAID accession numbers and associated metadata). Network structure was consistent across our survey years (Table 1b).

For each species pair, we extracted two network metrics, both of which we predicted were associated with transmission probability: 1) frequency of species co-occurrence in the landscape (simple connection strength measured as dyadic edge-weight, median = 36, range = 1 - 1,958), capturing social association or aggregation on wetlands; and 2) species-level association patterns (SRI, median = 0.0027, range = $2.4 \times 10^{-}$ - 0.028), capturing closer social association and/or shared habitat niches over and above abundance. Each pair also comprised four measures from the viral phylogeny: maximum genetic distance between viral sequences (median = 0.014; range = 0.000 - 0.028), number of samples (median = 15, range = 2 - 103), mean distance between samples in meters (geometric distance, median = 247.7 km, range = 0 - 1,105.1 km) and mean distance between samples in days (temporal distance, median = 243 d, range = 0 - 1,409 d).

Our models demonstrate that network metrics predicted transmission probability of an infectious disease through a multi-species community of wild birds. Species pairs that co-occurred more frequently had more similar viral genomes than those that co-occurred less often, strongly suggesting opportunity for transmission (estimate = -0.02, 95% CI = -0.05 to -0.00). The effect persisted after controlling for sampling effort, but repeating this model using the mean (rather than max) genetic distance resulted in no significant effect (estimate = 0.01, 95% CI = -0.03 to 0.05). Permutation testing demonstrated that co-occurrence predicted genetic distance beyond what would be expected from random network structure (permuted 95% CI = -0.01 to 0.01, $p < 0.001$). Patterns of relative species association (SRI) did not significantly predict more similar viral genomes (estimate = -0.01, 95% CI = -0.02 to 0.01), but the result did suggest a similar, albeit marginal, effect to the co-occurrence models. This marginal observed effect also predicted genetic distance beyond what would be expected from random network structure (permuted 95% CI = -0.008 to 0.007, $p = 0.01$), further suggesting a weak effect of species association, over and above abundance, on maximum genetic distance.

The effect of time between samples, distance between samples and the number of samples were consistent across both models. Time between samples strongly predicted more divergent viral genomes (estimate = 0.29, 95% CI = 0.27 to 0.30). Likewise, but to a lesser extent, distance between samples also predicted more divergent genomes (estimate = 0.03, 95% CI = 0.01 to 0.05), both consistent with gradual viral evolution over time and space. The number of samples contributed a

significantly positive effect to maximum genetic distance (estimate = 0.57, 95% CI = 0.28 to 0.85), justifying its addition as a covariate effect in our models. Using a multi-membership random effect structure, we found high variation in individual species contributions to both models, suggesting that context and ecology played important roles in transmission probability (weight model: sd = 1.10; 95% CI = 0.86 to 1.44; SRI model: sd = 1.11, 95% CI = 0.87 to 1.44). Time and space effects were consistent between observed and permuted models, both of which are independent of social network structure.

Discussion:

We demonstrated that social network structure in a multi-species community of wild animals predicted genetic distance between HPAI samples, a proxy for disease transmission probability. The stronger effect of raw co-occurrence frequency compared to relative association strength suggests that transmission is primarily driven by opportunity for contact rather than preferential association between species pairs. Species sharing space more frequently share more similar viruses, regardless of whether that association is socially structured. This provides scarce empirical evidence that network edges, rather than node-level host traits, can predict pathogen spread in a complex, multi-species wild animal system.

Our results suggest that HPAI transmission is measurable through a multi-species network of wild animals using social links as a proxy for transmission pathways. We suggest this approach provides an empirical baseline for further predictive process modelling of multi-species host-pathogen networks. Network stability over time, which we have also demonstrated here, and the speed of survey data availability will be crucial to applying these concepts to predicting future outbreak risk (Atkinson et al. 2025). We used two large, open-data repositories, underscoring their value to research into wildlife disease surveillance (Woodward et al. 2024; Shu & McCauley 2017). Recent studies have also used similar open data to model multi-species networks (Albery et al. 2020; Van Doren et al. 2025). For example, Vickers et al. (2024) used the eBird global repository of bird survey data (Sullivan et al. 2009) to link abundance of wild birds in the landscape with HPAI incursion into poultry.

Although our study represents a complex, multi-species transmission system, we did not investigate the role of specific host traits in shaping contagion, although much previous work has done so (Vickers et al. 2024; Gamża et al. 2024; Luis et al. 2015; Olival et al. 2017), without defining specific pathways of transmission. Host traits likely drive transmission through social links, by influencing immunocompetence (Hill et al. 2016), or behaviours triggered by the presence of disease in the community (Jeglinski et al. 2024), or indirectly by shaping the social network through which disease is transferred, irrespective of the presence of disease (Albery et al. 2021; Sah et al. 2018; Kulahci & Quinn 2019). It is also likely that host traits drive disease transmission at different scales. Long-distance migration has been associated with intercontinental HPAI transmission (McDuie et al. 2024; Tian et al. 2015), whereas local outbreaks in poultry have been associated with short-distance dispersal in more sedentary birds (Vickers et al. 2024; Gamża et al. 2024; Giacinti et al. 2024). This may also explain the relatively modest effect of geographic distance between samples in our models, which operated at the landscape rather than continental

scale. Our study therefore suggests how these two scales are linked, from intercontinental incursion to outbreaks in local farmyards. Evaluation of the role of host traits in HPAI transmission may further support identification of putative bridge taxa; those critical intermediary hosts between waterfowl and poultry.

We used maximum viral genetic distance as a proxy for transmission probability between species pairs, rather than describing actual transmission events in the landscape. The positive correlation between sample size and maximum genetic distance ($r = 0.52$) reflected the increased probability of sampling more divergent sequences when more data are available, but controlling for sampling effort as a covariate allowed us to separate biological signal from methodological artefact, evidenced by the stability of our results and the weak correlation between co-occurrence frequency and number of sequences (Spearman's $\rho = 0.19$). When repeating models using the mean, rather than the max, genetic distance, the direction of the co-occurrence effect was reversed in both models, with co-occurrence predicting slightly higher rather than lower genetic distance. Although beyond the scope of this study, this mean result, that may be constraining divergent genomes between species pairs, may suggest that frequent contact could also broaden viral diversity.

While we used 1,687 pairwise comparisons between 356 viral genome sequences across 33 host species, our study was limited by the availability of genomic data across the full range of potential host species. We do not expect this to have materially affected our results, which included only common wetland birds with nationally broad distributions. We suggest this aggregated measure nonetheless offers a strong proof-of-concept for network approaches in multi-species wild animal disease systems. Proactive sampling of viruses from a more comprehensive range of species would be required to map transmission pathways with greater confidence at multiple spatial and temporal scales, particularly for disease systems such as HPAI, where species responsible for bridging the transmission gap between primary wild hosts and livestock remain unidentified.

In addition to the gradual accumulation of mutations, genetic distances between influenza viruses may be influenced by reassortment, by which genomic segments are exchanged between viruses co-infecting the same host cell. Reassortment results in larger, instantaneous increases in genetic distance that can be accompanied by significant phenotypic change. There is evidence to suggest that reassortment of recently circulating HPAI viruses has driven diversification of host specificity and seasonality of transmission (Harvey et al. 2025), contributing to the complexity of defining networks of wild hosts. For this study, we used partial genetic distances between segments unaffected by reassortment, and repeating our models using full genetic distances found consistent results, implying little influence of reassortants on our dataset. Future studies should nonetheless consider whether genotypes generated by reassortment events occupy different positions within host social networks.

We also note that a multi-species network of wild animals may include multiple mechanisms of transmission, and our approach is appropriate only for non-vector-borne infectious disease systems. Our results may be explained by direct interaction or shared environmental exposure, but our viral genome database also included

samples from predatory birds with close genetic distances to colonial seabirds. Those links were likely trophic, representing ingestion of infected prey rather than indirect transmission. Future studies should therefore define network edges by the mechanism of transmission, for example social links, mobility, or predator-prey interactions, to more precisely evaluate patterns of spread.

Our social network approach, allied with the increasing availability of open data on wild species occurrence and relative abundance, offers significant potential for evaluating and predicting pathogen transmission pathways in wild animal systems, and hence opportunities to better inform surveillance and epizootic disease control policies.

***

Author Contributions: JD was involved in all elements of this study - design, data acquisition, formal analysis, writing and editing; JAF was involved in all aspects of this study, but primarily social network, model specification and project supervision; AG contributed data, writing and editing; ABB contributed data, writing and editing; WH and SJL contributed to study design primarily sequence selection, processing, and editing, RRK and IAB contributed to study design, data acquisition and editing; AIW was involved in all elements of this study.
Competing interests: The authors declare no competing interests
Data Availability statement: All of our processed data and relevant code files have been made available in the supplementary files - 10.5281/zenodo.17161964. Raw data files can be accessed from the Wetland Bird Survey via the British Trust for Ornithology, and viral phylogenetic data from GISAID. Sequences and accession numbers are given in supplementary table 1.

Funding and Acknowledgements: Data were provided by WeBS, a partnership jointly funded by the British Trust for Ornithology, Royal Society for the Protection of Birds and Joint Nature Conservation Committee with fieldwork conducted by volunteers and previous support from Wildfowl and Wetlands Trust. We acknowledge all data contributors, i.e., the Authors and their Originating laboratories responsible for obtaining the specimens, and their submitting laboratories for generating the genetic sequence and metadata and sharing via the GISAID Initiative. We thank Amanda Bretman, Bill Kunin, Talya Hackett, Elizabeth Duncan and Ashley C Banyard, Joe James and Holly Coombes from the Animal and Plant Health Agency, for comments on this manuscript or the data therein. Our project was supported by Flu-Trial Map Avian: JD, IAB and AIW under grant no. BB/Y007042/1; AG, RK, WH, SL were funded by the Strategic Program grant to Roslin Institute (grant no.BBS/E/RL/230002D), Flu-MAP project (grant no. BB/X006123/1), Flu-TrailMap (grant number BB/Y007271/1, BB/Y007298/1) and FluTrailMap- One Health (MR/Y03368X/1). JAF also acknowledges funding from NERC (NE/V013483/2) and WildAI (C-2023-00057). Icons used in figure 1 were sourced from PhyloPic and were donated to the public domain.

References:

Albery, G. F., Eskew, E. A., Ross, N. & Olival, K. J. Predicting the global mammalian viral sharing network using phylogeography. Nature communications 11, 2260 (2020).

Albery, G. F., Kirkpatrick, L., Firth, J. A. & Bansal, S. Unifying spatial and social network analysis in disease ecology. Journal of Animal Ecology 90, 45–61 (2021).

Atkinson, P. W. et al. Evaluating the use of carcass and testing data to assess the high pathogenicity avian influenza (HPAI) related mortality in wild birds in the united kingdom and crown dependencies between 2021–2023. Bird Study 0, 1–19 (2025).

Bansal, S., Grenfell, B. T. & Meyers, L. A. When individual behaviour matters: Homogeneous and network models in epidemiology. Journal of the Royal Society Interface 4, 879–891 (2007).

Bernstein, A. S. et al. The costs and benefits of primary prevention of zoonotic pandemics. Science Advances 8, eabl4183 (2022).

Buhnerkempe, M. G. et al. Eight challenges in modelling disease ecology in multi-host, multi-agent systems. Epidemics 10, 26–30 (2015).

Bürkner, P.-C. Brms: An r package for bayesian multilevel models using stan. Journal of statistical software 80, 1–28 (2017).

Carlson, C. J. et al. Climate change increases cross-species viral transmission risk. Nature 607, 555–562 (2022).

Dunning, J., Firth, J. A. & Ward, A. I. The spread of highly pathogenic avian influenza virus is a social network problem. PLoS Pathogens 21, e1013233 (2025).

Falchieri, M. et al. Shift in HPAI infection dynamics causes significant losses in seabird populations across great britain. The Veterinary record 191 7, 294–296 (2022).

Farine, D. R. & Whitehead, H. Constructing, conducting and interpreting animal social network analysis. Journal of animal ecology 84, 1144–1163 (2015).

Farine, D. R. A guide to null models for animal social network analysis. Methods in Ecology and Evolution 8, 1309–1320 (2017).

Firth, J. A. et al. Using a real-world network to model localized COVID-19 control strategies. Nature medicine 26, 1616–1622 (2020).

Gamża, A., Lycett, S., Sanchez, A. & Kao, R. Using sequence data to study spatial scales of interactions driving spread of highly pathogenic avian influenza in great britain. arXiv preprint arXiv:2411.10424 (2024).

Giacinti, J. A. et al. Transmission dynamics of highly pathogenic avian influenza virus at the wildlife-poultry-environmental interface: A case study. One Health 19, 100932 (2024).

Gómez, J. M., Nunn, C. L. & Verdú, M. Centrality in primate–parasite networks reveals the potential for the transmission of emerging infectious diseases to humans. Proceedings of the National Academy of Sciences 110, 7738–7741 (2013).

Hamede, R. K., Bashford, J., McCallum, H. & Jones, M. Contact networks in a wild tasmanian devil (sarcophilus harrisii) population: Using social network analysis to reveal seasonal variability in social behaviour and its implications for transmission of devil facial tumour disease. Ecology letters 12, 1147–1157 (2009).

Harvey, W. T. et al. Genetic reassortment and diversification of host specificity have driven evolutionary trajectories of lineages of panzootic H5N1 influenza. bioRxiv doi: https://doi.org/10.1101/2025.08.20.670882 (2025).

Hill, S. C. et al. Antibody responses to avian influenza viruses in wild birds broaden with age. Proceedings of the Royal Society B: Biological Sciences 283, 20162159 (2016).

Holmes, E. C. COVID-19—lessons for zoonotic disease. Science 375, 1114–1115 (2022).

Jeglinski, J. W. E. et al. HPAIV outbreak triggers short-term colony connectivity in a seabird metapopulation. Scientific Reports 14, (2024).

Jones, K. E. et al. Global trends in emerging infectious diseases. Nature 451, 990–993 (2008).

Kandeil, A. M. et al. Rapid evolution of a(H5N1) influenza viruses after intercontinental spread to north america. Nature Communications 14, (2023).

King, K. C., Hall, M. D. & Wolinska, J. Infectious disease ecology and evolution in a changing world. Philosophical Transactions of the Royal Society B vol. 378 20220002 (2023).

Kostandova, N. et al. A systematic review of using population-level human mobility data to understand SARS-CoV-2 transmission. Nature communications 15, 1–12 (2024).

Kruse, H., Kirkemo, A.-M. & Handeland, K. Wildlife as source of zoonotic infections. Emerging infectious diseases 10, 2067 (2004).

Kulahci, I. G. & Quinn, J. L. Dynamic relationships between information transmission and social connections. Trends in ecology & evolution 34, 545–554 (2019).

Luis, A. D. et al. Network analysis of host–virus communities in bats and rodents reveals determinants of cross-species transmission. Ecology letters 18, 1153–1162 (2015).

Marí Saéz, A. et al. Investigating the zoonotic origin of the west african ebola epidemic. EMBO molecular medicine 7, 17–23 (2015).

McDuie, F. et al. Mitigating risk: Predicting H5N1 avian influenza spread with an empirical model of bird movement. Transboundary and Emerging Diseases 2024, 5525298 (2024).

Olival, K. J. et al. Host and viral traits predict zoonotic spillover from mammals. Nature 546, 646–650 (2017).

Paradis, E. et al. Package 'ape'. Analyses of phylogenetics and evolution, version 2, 47 (2019).

Plaza, P. I., Gamarra-Toledo, V., Euguí, J. R. & Lambertucci, S. A. Recent changes in patterns of mammal infection with highly pathogenic avian influenza a (H5N1) virus worldwide. Emerging Infectious Diseases 30, 444 (2024).

Proboste, T. et al. Quantifying feral pig interactions to inform disease transmission networks. bioRxiv 2024–08 (2024).

Reid, S. M. et al. A multi-species, multi-pathogen avian viral disease outbreak event: Investigating potential for virus transmission at the wild bird–poultry interface. Emerging Microbes & Infections 13, 2348521 (2024).

Sah, P., Leu, S. T., Cross, P. C., Hudson, P. J. & Bansal, S. Unraveling the disease consequences and mechanisms of modular structure in animal social networks. Proceedings of the National Academy of Sciences 114, 4165–4170 (2017).

Sah, P., Mann, J. & Bansal, S. Disease implications of animal social network structure: A synthesis across social systems. Journal of Animal Ecology 87, 546–558 (2018).

Sah, R. et al. Concerns on H5N1 avian influenza given the outbreak in US dairy cattle. The Lancet Regional Health–Americas 35, (2024).

Shu, Y. & McCauley, J. GISAID: Global initiative on sharing all influenza data–from vision to reality. Eurosurveillance 22, 30494 (2017).

Sidik, S. M. Bird flu outbreak in mink sparks concern about spread in people. Nature 614, 17 (2023).

Sullivan, B. L. et al. eBird: A citizen-based bird observation network in the biological sciences. Biological conservation 142, 2282–2292 (2009).

Tamura, K. & Nei, M. Estimation of the number of nucleotide substitutions in the control region of mitochondrial DNA in humans and chimpanzees. Molecular biology and evolution 10, 512–526 (1993).

R Core Team. R: A Language and Environment for Statistical Computing. (R Foundation for Statistical Computing, Vienna, Austria, 2023).

Tian, H. et al. Avian influenza H5N1 viral and bird migration networks in asia. Proceedings of the National Academy of Sciences 112, 172–177 (2015).

Tremlett, C. J., Cleasby, I. R., Bolton, M. & Wilson, L. J. Declines in UK breeding populations of seabird species of conservation concern following the outbreak of high pathogenicity avian influenza (HPAI) in 2021–2022. Bird Study 1–18 (2025).

Van Doren, B. M. et al. Social associations across species during nocturnal bird migration. Current Biology (2025).

Vickers, S. H., Raghwani, J., Banyard, A. C. & Brown, I. H. Utilising citizen science data to rapidly assess potential wild bridging hosts and reservoirs of infection: Avian influenza outbreaks in great britain. bioRxiv (2024).

Wade, D. et al. High pathogenicity avian influenza: Targeted active surveillance of wild birds to enable early detection of emerging disease threats. Epidemiology and Infection 151, (2022).

Wilson-Aggarwal, J. K. et al. High-resolution contact networks of free-ranging domestic dogs canis familiaris and implications for transmission of infection. PLoS neglected tropical diseases 13, e0007565 (2019).

Woodward, I. D. et al. 2022/23: The Wetland Bird Survey and Goose & Swan Monitoring Programme. (2024).

Xie, R. et al. The episodic resurgence of highly pathogenic avian influenza H5 virus. Nature 622, 810–817 (2023).

Yin, S. et al. Functional traits explain waterbirds' host status, subtype richness, and community-level infection risk for avian influenza. Ecology Letters 26, 1780–1791 (2023).

Figures and Tables:

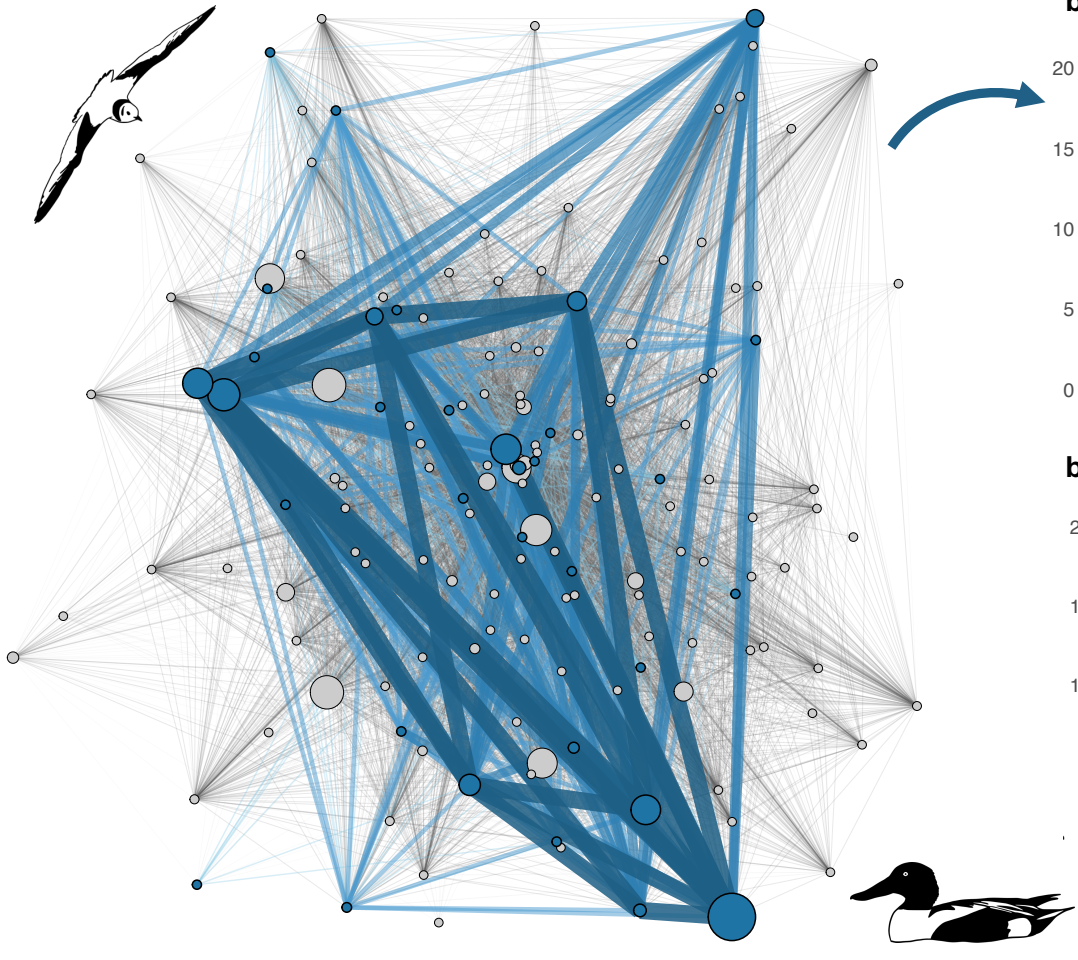
Table 1a-c. Summary statistics for social networks within (a) and between (b) years, and node (viral host species) attributions (c). a shows the number of species involved in each network, edges between species pairs representing co-occurrence within a recording event, and, the number of recording events. A recording event represents a single survey visit in time and space within the study area. b Shows between year network similarity, where the Upper right matrix shows Jaccard similarity (J), the number of shared dyads between years irrespective of edge weight; and the Lower left matrix, Spearman rank correlation coefficients (ρ), between dyadic edge weights. c shows the nodes included in the viral phylogeny that make up those used in our models. Here, host is the species from which virus was sequenced, n events shows the number of events that host was recorded in, degree is the number of distinct connections (over-laps) in the social network and Strength is the sum of edge weights across those connections, Dyadic weight and Dyadic SRI give the mean and sd of those centrality metrics, and n Seq denotes the number of sequences accessed for each host species.

Figure 1. Network centrality predicts transmission pathways. a. Of the 159 species in the co-occurrence network, 33 were also associated with a genome for high pathogenicity avian influenza clade 2.3.4.4b, of which we had access to 356 unique sequences. The colour of the nodes and edges denote that a species pair had a viral genome available, blue, or not, grey. The node size denotes normalised betweenness centrality, where the largest nodes are those most central to network structure, in this case Mute Swan *Cygnus olor*, and 0, on none of the shortest paths (here including Northern Lapwing *Vanellus vanellus*, Little egret *Egretta garzetta*, and Northern pintail *Anas acuta*). Edge width denotes extent of co-occurrence, where thicker edges link species pairs often recorded together. **b.** The expected distribution of permuted estimates in blue, with 95% confidence intervals; blue dotted lines, plotted against the observed estimate, dark vertical line, for two measures of network structure; edge weight (a measure of co-occurrence in the landscape) and Simple Ratio Index (a measure of species association). When organic social structure is removed from the network, species connection strength predicts less distant HPAI genomes, suggesting a biological link with patterns of species association and disease transmission.

Supplementary files:
S1 – GISAID sequences and accession numbers.
S2 – Combined data files used to run models.
S3 – Annotated code files.

a Co-occurrence of waterbirds in the landscape

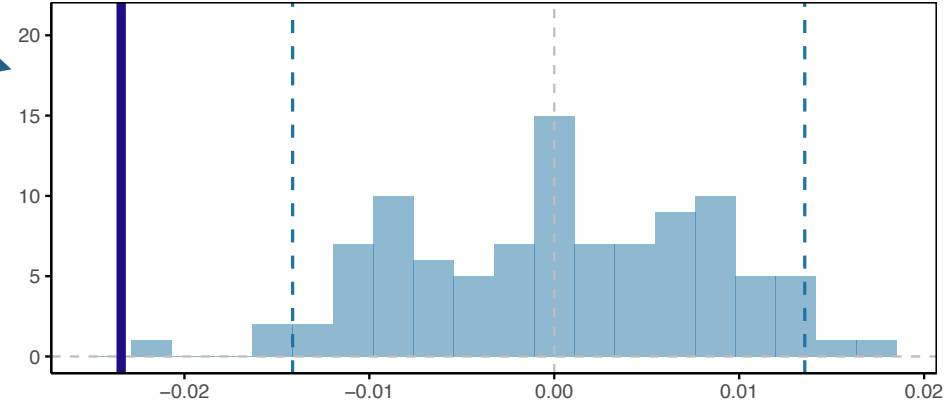
Landscape scale co-occurrence networks show HPAI transmission pathways between wild wetland birds using viral genetic distance as a proxy.



b Effect of network structure on genetic distance

Permuted Edges (95% CI) and Observed Model Estimates

b1 Co-occurrence in the landscape (Edge Weight)



b2 Relative association (Simple Ratio Index)

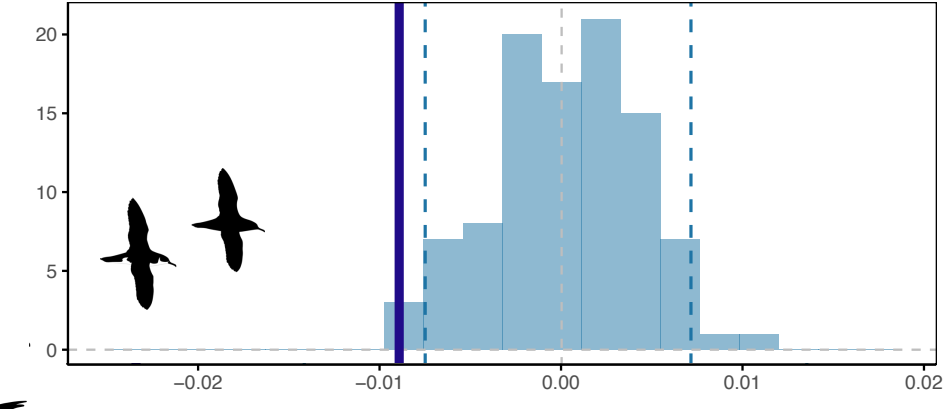


Table 1a-c. Summary statistics for social networks within (a) and between (b) years, and node (viral host species) attributions (c). 1a shows the number of species involved in each network, edges between species pairs representing co-occurrence within a recording event, and, the number of recording events. A recording event represents a single survey visit in time and space within the study area. **1b.** Shows between year network similarity, where the **Upper right matrix** shows Jaccard similarity (J), the number of shared dyads between years irrespective of edge weight; and the **Lower left matrix**, Spearman rank correlation coefficients (ρ), between dyadic edge weights. **1c** shows the nodes included in the viral phylogeny that make up those used in our models. Here, host is the species from which virus was sequenced, n events shows the number of events that host was recorded in, degree is the number of distinct connections (over-laps) in the social network and Strength is the sum of edge weights across those connections, Dyadic weight and Dyadic SRI give the mean and sd of those centrality metrics, and n Seq denotes the number of sequences accessed for each host species.

a. Network structure

Year	Species	Edges	Events
2018	115	4132	354
2019	131	5321	729
2020	127	4938	717
2021	134	6378	782
2022	137	6154	748

b. Network similarity between years

J → ρ ↓	**2018**	**2019**	**2020**	**2021**	**2022**
2018	-	0.67	0.65	0.56	0.56
2019	0.89	-	0.71	0.64	0.67
2020	0.86	0.91	-	0.66	0.63
2021	0.83	0.87	0.86	-	0.73
2022	0.79	0.85	0.85	0.91	-

c. Node attributes

Host	n Events	Degree	Strength	Dyadic Weight	Dyadic SRI	n Seq*
Accipiter nisus	279	136	17466	52.7; 52.2	0.00324; 0.00161	4
Acrocephalus scirpaceus	80	113	5184	18.7; 18.1	0.00144; 0.00095	1
Alca torda	6	34	88	1.5; 0.7	0.0041; 0.00773	1
Anas platyrhynchos	3442	154	189722	511.7; 570.1	0.01041; 0.00989	3
Anser albifrons	7	83	495	1.4; 0.5	0.00023; 0.00043	2
Anser anser	2079	153	116590	332.2; 387.4	0.00886; 0.0078	28
Anser brachyrhynchus	246	136	28400	89.8; 87	0.00441; 0.00247	14
Ardea cinerea	1795	151	99292	284.4; 326.2	0.00822; 0.00724	1
Branta canadensis	2001	149	86352	248.9; 288.7	0.00782; 0.0067	18
Branta leucopsis	188	134	16459	49.5; 50.6	0.00283; 0.00138	26
Buteo buteo	503	141	33185	104.3; 101.8	0.00506; 0.003	20
Chroicocephalus ridibundus	2065	153	88087	249.4; 289.7	0.00782; 0.00691	56
Corvus corone	316	132	22654	69.2; 89.1	0.00338; 0.00235	1
Cygnus cygnus	154	119	9016	29.8; 25.3	0.00206; 0.00077	7
Cygnus olor	2448	150	128933	363.9; 426.4	0.00917; 0.00851	47
Falco peregrinus	106	140	8869	28.3; 30.1	0.00192; 0.001	8
Falco subbuteo	52	110	3028	11.1; 13	0.00111; 0.00101	1
Falco tinnunculus	414	143	32792	101.5; 100.5	0.00493; 0.00296	2
Fulica atra	2386	150	113663	324.2; 380	0.0088; 0.00805	1
Fulmarus glacialis	1	24	24	1; 0	0.00678; 0.0093	1
Gallinula chloropus	2946	151	121677	345.9; 410	0.00906; 0.00845	1
Hydrocoloeus minutus	39	111	2332	7.1; 4.4	0.00102; 0.00122	1
Larus argentatus	915	152	43718	130.1; 146.2	0.0055; 0.00401	12
Larus canus	780	148	39136	117.6; 122.8	0.0053; 0.00375	3
Morus bassanus	2	76	275	3.9; 3.6	0.00229; 0.00555	4
Numenius arquata	474	149	110599	298.1; 343.6	0.00818; 0.00722	1
Rissa tridactyla	4	85	477	4.6; 3.6	0.00065; 0.00085	10
Somateria mollissima	5	30	101	1.6; 0.8	0.0043; 0.00857	1
Sterna hirundo	255	122	9023	29.9; 29.7	0.00215; 0.00109	12
Sterna paradisaea	12	97	1372	5.3; 5.2	0.00069; 0.00081	3
Strix aluco	47	108	2024	8; 5.5	0.00112; 0.00134	2
Thalasseus sandvicensis	7	98	950	5.3; 5.3	0.00065; 0.00074	5
Vanellus vanellus	1043	146	109460	314.3; 346.1	0.00872; 0.00729	1