High pathogenicity avian influenza virus transmission is shaped by inter-specific 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 Lycett[2] (0000-0003-3159-596X), Rowland Kao[2,6] (0000-0003-0919-6401), Ian A. 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
*A.I.Ward@leeds.ac.uk

Emerging zoonotic and epizootic pathogens that spill over from wildlife are a persistent, and often intractable, threat to human and animal health[1,2] with far-reaching consequences for global economic and food security[3]. This is particularly true for multi-host systems like wild birds, where ecological and behavioural diversity across many species interact, giving rise to complex transmission pathways[4,5] and, presenting a challenge for proactive surveillance[5]. Here we show that wild bird social network structure predicted viral genetic distance between high pathogenicity avian influenza viruses (HPAIV/s). Links in the network predicted close genetic distance between viral genomes, consistent with HPAIV transmission among species that co-occur in the wild. More so, species connection strength, over-and-above the effect of abundance, also predicted viral genetic distance. Our study provides an empirical demonstration of the drivers of disease transmission through complex multi-species communities of wild animals and, the use of social network analysis as framework for wildlife disease surveillance.

Diseases like Ebola, SARS-CoV-2 and avian influenza viruses all originate from multi-species groups of wild animals, often without detection in those populations prior to outbreaks in humans or livestock[6–8]. This is because, in non-human animal communities, patterns of disease transmission are highly variable, and cannot be solely explained by pathogen virulence. Instead, just as human mobility and interaction shapes the trajectory of an epidemic[9,10] animal social network structure, shaped by the social[4,11,12] and physical environment[13,14] influences epizootic disease propagation through multi-species networks of wild animals. It must therefore be drivers of these dynamic social systems, that, in turn, drive the spread of disease. This complexity is magnified further in multi-host systems, where many diverse species, with varied behavioural traits, and spatial preferences interact[4,5,15], giving rise to the pathways of transmission. For example, during the HPAI outbreak of 2021/22, where H5N1 clade 2.3.4.4b began to infect European seabirds, disease presence in otherwise isolated breeding colonies prompted abandonment of those sites, increasing between-colony

connectivity and driving further spread[16]. These expanded networks, drove host expansion with devastating consequences for the conservation of wild birds[17,18], driving incursion of HPAIV into new continents[19] and novel mammalian hosts[20–22].

Perhaps due to the challenges in understanding the drivers of disease through multi-species networks, research into animal disease systems has typically focused on specific traits (e.g. population dynamics[23,24] or host traits[25–28]), represented by network nodes, rather than between-species social structure, represented by the interconnecting edges between species nodes. However, the mechanism by which host traits, or environmental components influence 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[29].

Understanding the patterns of disease spread through multi-host networks, and the traits shaping those networks, represents an important knowledge gap for informing zoonotic and epizootic disease surveillance and prevention[5,12,15,30]. Both aspects have been challenging to test empirically, leaving evaluation of multi-host network disease dynamics reliant on theoretical models[28,31]. Furthermore, surveillance and predictive modeling remain limited by reactive sample collection, processing constraints, and taxonomic sampling biases[32,33]. Social network approaches offer promise[15], but have, so far, tended to only be applied to simple, single-species transmission systems[29,34–36]. The use of social network models to measure transmission is yet to be validated empirically, and their application to more complex, multi-species transmission systems remains a critical knowledge gap[5,15]. There is therefore a pressing need for predictive approaches to identify the species and pathways driving zoonotic and epizootic disease spread, in order to inform risk-based policy interventions.

Here, we used data from HPAIV outbreaks in Great Britain (GB), a serious public health threat with trans-boundary transmission between hundreds of interacting hosts and multiple transmission pathways[8,37], as a rich study system to test the relationship between multi-species social linkage and disease transmission. We combined social network analyses with viral phylogenetics and Bayesian multilevel modelling to test our prediction that social links predict viral genome distance (as a proxy for transmission probability) in multi-species networks of wild birds. Previous phylodynamic studies have matched H5N1 infection in poultry with identical genotypes in nearby wetland sediments[38], implying landscape transmission without identified pathways. We therefore used fine-scale species co-occurrence, derived from national monitoring data of wetland birds, to build social networks to explore relationships between network structure and host virus phylogeny. Our results demonstrate that bird species co-occurrence patterns drive viral genome distance at landscape scales, providing empirical evidence that network structure governs pathogen spread in complex communities of wild animals.

Aligning social networks with HPAI viral phylogenetics

We constructed a multi-species social network from five years of Wetland Bird Survey (WeBS) data from the British Trust for Ornithology (BTO), a database describing the spatial spread of wetland birds at the landscape scale[39]. From these data we used 3,722 individual surveys within a 50 km buffer of a farm cluster with recurrent HPAI outbreaks (nine outbreaks in the farm cluster and 69 identified within the buffer since 2018) in GB, between 2018 - 2022, which we combined into a single global network (for a summary see Table 1a-c). We used this network to describe patterns of inter-specific interaction by inferring 8,162 species dyads (edges, representing a species co-occurrence event) between 159 species nodes. Network structure, that is, species composition and the relative number of interconnecting edges between species pairs was consistent between survey years (Table 1b).

To measure transmission probability, we accessed HPAI viral genomic data from the Global Initiative on Sharing All Influenza Data (GISAID)[40], and calculated distance between virus sequences, which we then aligned with hosts. After filtering out domestic and wild host species, not also present in our WeBS data, we used 22,791 combinations of 205 viral genome sequences, representing 214 combinations of 20 host species (See supplementary table 1 for GISAID accession numbers and associated metadata). We calculated genetic distance between viral genomes, where identical pair-wise genomes (those with no genetic distance between them) were scored as 0. While we limited our viral phylogenetic data to the H5N1 subtype, we also identified reassortant events between H5N1 and other low and high pathogenicity sub-types, resulting in two mechanisms of viral evolution: small, incremental changes driven by gradual mutation 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[41].

For each species pair, we extracted two measures of edge centrality (a measure representing an individual species' contribution to global network structure), both of which we predicted were associated with transmission probability: 1) frequency of species co-occurrence in the landscape (dyadic edge-weight, global median = 63, range = 1 - 1,837); and 2) species-level connection strength (Simple Ratio Index, SRI, global median = 0.003, range = 9.63×10^{-5} - 2.72×10^{-2}). SRI represents the connectedness of a species pair relative to opportunity for association, and regardless of abundance. A greater SRI is therefore a measure of dyadic clustering probability in the social network. We calculated each measure from the global network, before filtering and scaling against our viral data. Each instance also comprised three measures from the viral phylogeny: conservative genetic distance (maximum genetic distance between viral sequences, median = 0.01; range = 0.001 - 0.02), distance between samples in meters (mean geometric distance, median = 251.9km, range = 88.6 - 682.9km) and distance between samples in days (temporal distance, mean = 280d, range = 8 - 537d).

To test the relationship between network centrality measures and viral genetic distance we ran two Bayesian Generalized Linear Mixed Models (GLMMs), controlling for spatial and temporal distances between sampling events. To sample from complex, high-

dimensional probability distributions we used Hamiltonian Monte Carlo sampling. We built separate models to evaluate associations between genetic distance as a proxy for transmission probability and two centrality measures, given as fixed effects, co-occurrence frequency and species connection strength. 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. We accounted for species specific variance on viral genetic distance independently of their dyad by linking a 'multi-membership' species variable within our random effects (see methods).

Network centrality predicts transmission pathways

We found that genetic distance was negatively associated with greater frequency of co-occurrence in the landscape (estimate = -0.06; 95% CI = -0.09 to -0.04), consistent with species that frequently co-occurred in the landscape having lower genetic distances (i.e., more closely related viral sequences) than those that co-occurred less frequently. The time elapsed between sampling (estimate = 0.19; 95% CI = 0.17 to 0.22) was strongly associated with greater genetic distance, while spatial distance showed no effect (estimate = -0.03; 95% CI = -0.06 to 0.01), possibly explained by differences in the spatial ecology of species in the network. Using a multi-membership random effect structure, we found high variation in individual species contributions to the model, suggesting that context and individual species ecology played an important role in transmission probability (estimate = 1.40; 95% CI = 1.02 to 1.97).

Similarly to co-occurrence, we found that maximum genetic distance was also negatively associated with greater species connection strength (estimate = -0.08; 95% CI = -0.11 to -0.05). This suggested that not only are species that co-occur more frequently in the landscape associated with lower genetic distances, but species pairs that cluster consistently, irrespective of abundance, are too. Again, time elapsed between samples was strongly associated with greater genetic distance (estimate = 0.20; 95% CI = 0.18 to 0.22), while spatial distance had only a small effect (estimate = -0.03; 95% CI = -0.07 to -0.00). Again, species identity explained significant variation in the model (estimate = 1.43; 95% CI = 1.04 to 2.04). Year, as a random effect, did not explain significant variance in either model, supporting the assumption that network structures remained broadly consistent across years.

To test if our associations could be explained by social network structure alone, we also rewired the network and repeated our models over a series of permutations while maintaining the degree distribution for each species node. In both the co-occurrence and species connection strength models, the observed effect estimates fell outside the 95% confidence intervals of the permuted (null) distributions (co-occurrence: 95% CI = -0.01 to 0.01; and connection strength: 95% CI = -0.01 to 0.01) indicating that the observed associations were stronger than expected by chance alone, given the network structure. The effect of space and time were consistent between observed and permuted models, consistent with our expectation that sequences sampled closer in space and/or time were more similar. We also repeated our models using full genetic

distances (irrespective of reassortant genomes), the results of which were consistent with our partial distance models (co-occurrence: 95% CI = -0.149 - -0.049; and connection strength: 95% CI = -0.19 - -0.09), implying limited influence of reassortant genomes in our sample.

Conclusion and Discussion:

We demonstrated that inter-specific social network structure in a multi-species community of wild animals significantly predicted genetic distance between HPAI samples, a proxy for disease transmission probability. Our study represents an empirical application of social network analysis to predict viral transmission pathways across multiple wild species in this way.

We used two, large, open-data repositories, underscoring their value to research into wildlife disease surveillance[39,40]. Although we used WeBS data from the BTO, recent studies have also used similar open-data to model multi-species networks[31,42]. For example, Vickers et al.[23] used the eBird global repository of bird survey data[43] to link abundance of wild birds in the landscape with HPAI incursion into poultry, but without specifying transmission pathways through a social network. In a social network context, previous studies have modeled pathogen transmission within simple or single species transmission systems[29,35,36,44], or adopted a trait-based approach within multi-species systems, requiring parameterization of network nodes[26,28]. Our results suggest that transmission of a disease of pandemic potential is also measurable through a multi-species network of wild animals using social-links (edges) as a proxy for transmission. This result is also predicted in the context of host or environment traits, which drive pathogen transmission, because they influence the shape of the social network, facilitating transmission by driving potential hosts into close proximity facilitating social linkage. Our study demonstrates an empirical baseline for further theoretical, predictive modelling of multi-species host-pathogen networks, by using social links as a proxy for transmission. This could help to fill gaps in phylodynamic modelling, such as those resulting from lack of virus samples from species with infected individuals. Network stability over time, which we have also demonstrated here, and speed of survey data availability will be crucial to apply the concepts we have demonstrated here to predicting future outbreak risk[33], using network edges as a proxy for transmission probability.

Although our study represents a complex, multi-species transmission system, we did not investigate the role of specific host (node-based) traits in shaping contagion, although much previous work has done so[23,24,26,27], without defining specific pathways of transmission. Host traits likely drive transmission through social links, by, for example, influencing immunocompetence[45], or behaviours triggered by the presence of disease in the community[16], or indirectly, by shaping the social network through which disease is transferred, irrespective of the presence of disease[12,30,46]. It is also likely that host traits drive disease transmission at different scales, such as long-distance migration, which has been associated with intercontinental HPAI transmission[28,47], whereas local outbreaks in poultry have been associated with short-distance dispersal in more sedentary birds[23,24,38]. This may also explain the minimal effect of distance between

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

While we used 205 viral genome sequences, representing 214 combinations of 20 host species, our study was limited by the availability of genome data for a large number of potential host species. The inferred social network, by comparison, comprised 8,162 edges between 159 species nodes. Although the sub-graph (blue, within the graph given in Fig.1a) offers proof of concept, we suggest proactive sampling of viruses from a sufficient number of a more comprehensive range of species would be required to map past transmission pathways of pathogen spread in order to enable greater confidence in prediction at multiple spatial and temporal scales. This may be particularly important for disease systems, such as HPAI, where species responsible for bridging the transmission gap between primary wild hosts and livestock are unidentified.

We identified two mechanisms of mutation within our viral phylogenetic data, gradual mutations that accumulated across transmission events, and genetic reassortant events where two viruses co-infect the same host cell, resulting in greater genetic distances. Harvey et al.[41] suggested that such events have driven the recent diversification of host species and the seasonality of emergence, and both of these components have contributed to the complexity of defining networks of wild hosts and surveillance of disease transmission. Future studies therefore should consider the potential for new reassortment opportunities and the confounding effect of reassortment of viral phenotypes. For the purpose of this study, where we have demonstrated an empirical link between measures of co-occurrence in the landscape and inter-specific transmission of HPAI, we used partial genetic distances between segments unaffected by reassortment of the viruses in our phylogeny. Repeating our models using full genetic distances found results consistent with those of our partial models, implying little influence of reassortants on the relatively limited genetic data used in our study.

We also note that a multi-species network of wild animals may include multiple mechanisms of transmission, and our study is appropriate only when applied to non-vector borne infectious disease systems. For example, our co-occurrence result may be explained by direct interaction or shared environmental exposure, but our HPAI viral genome database also included samples taken from predatory birds, raptors, that had close viral genetic distances to colonial seabirds. However, those links were likely tropic, rather than social, probably representing ingestion of infected prey, rather than indirect transmission. We therefore suggest that future studies should define network edges by the mechanism of transmission, for example social links, mobility, or predator-prey interactions, to evaluate patterns of transmission, as we have done here.

Combined, our results support the hypothesis that species co-occurrence in a landscape is linked with pathogen genetic distance, implying predictable transmission through a complex, multi-species social network of wild birds. Our social network approach, allied with increasing availability of open data on wild species occurrence and relative abundance, offers significant potential for evaluating and predicting pathogen

transmission pathways among wildlife communities, and hence opportunities to better inform surveillance and epizootic disease control policy.

Methods:

To test the hypothesis that host species' social linkages 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)[39], 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, which recorded 159 species. We built a single network for each year, where the nodes represented species recorded within the WeBS events, 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 159 species nodes and 8,162 interconnecting edges (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, which is 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[48]), which measures between 0 (no overlap between a species) and 1 (where one species is recorded so is the other, without exception). Although both measure interaction between species, pairwise weight and SRI represent different aspects of interaction in the landscape. Where a high pairwise weight denotes more frequent co-occurrence in the landscape, a higher SRI value represents stronger relative association, regardless of abundance. For example, in the WeBS data, the highest SRI values were between relatively scarce species whose habitat overlaps, for example Dipper Cinclus cinclus and Grey Wagtail Motacilla cinerea, which both favor fast flowing upland streams. Although they were both rare in the landscape, they were almost always recorded at the same sites. Finally, we restricted the network used in our models to only include the species also present in the viral phylogenetic data (see below), while weights and SRI measures were calculated and carried over from the global social network.

Network similarity: We tested the assumption that our networks are comparable between years, that is 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)[40]. We aligned 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 22,791 combinations of 205 viral genome sequences from 20 host species.

We aligned each sequence per segment using MAFFT v7.511[49], 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. The segments were concatenated into whole genomes which were used to quantify pairwise genetic distances between viruses. We compared each sequence with all others in the database to quantify a genetic distance between pairwise viral genomes using the Tamura and Nei 1993 model[50] using the ape package in R[51,52] that calculates genetic distance as a proportion that differ between each pair of sequence, accounting for difference in rates of transition and transversion mutations. To compensate for the potential of reassortant genomes in our data, which may confound the gradual accumulation of mutations through transmission in the landscape, we limited our models to only non-reassorting Haemagglutinin, Neuraminidase and Matrix protein genome fragments[41]. However, to assess the extent to which reassortant gene segments influenced our models, we also ran each using the full, rather than partial, genetic distances. We used the maximum genetic, least similar, distance between viral genomes for each species pair as a conservative proxy for transmission probability. 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 between sampling events[24,41]

Pairwise weight: We modeled pairwise weight, 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 z-transformed all fixed effects to better align scales between other variables in the model. 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. 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 speciesAB) by including species A and species B as a linked ‘multi-membership’ variable inside our random effects. This approach also avoided over-dispersion.

SRI: 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.

Model fit: To test the relationship between wild bird social network structure and maximum 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[51,53]. All models used the maximum genetic distance (least similar) between pairwise viral genomes, as a conservative proxy for transmission between hosts, as the response variable. Otherwise, our models only differed only by 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 (Ř) 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.

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 54. 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[54]. 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.

***

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 and model specification, and project supervision; AG contributed data, writing and editing; ABB contributed data, writing and editing; WH and SL contributed to study design primarily sequence selection, processing, and editing, RK 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

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).

References

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

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

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

4. 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).

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

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

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

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

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

10. 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).

11. 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).

12. 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).

13. 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).

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

15. 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).

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

17. 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).

18. 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).

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

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

21. 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).

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

23. 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).

24. 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).

25. 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).

26. 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).

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

28. 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).

29. 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).

30. 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).

31. 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).

32. 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).

33. 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).

34. 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).

35. 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).

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

37. 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).

38. 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).

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

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

41. 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 https://doi.org/10.1101/2025.08.20.670882 (2025).

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

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

44. Gomes, A. C. R., Sorenson, M. D. & Cardoso, G. C. Speciation is associated with changing ornamentation rather than stronger sexual selection. *Evolution* **70**, 2823–2838 (2016).

45. 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).

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

47. 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).

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

49. Katoh, K. & Standley, D. M. MAFFT: Iterative refinement and additional methods. in *Multiple sequence alignment methods* 131–146 (Springer, 2013).

50. 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).

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

52. Paradis, E. *et al.* Package ‘ape’. *Analyses of phylogenetics and evolution, version* **2**, 47 (2019).

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

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

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.

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	55.1; 51.8	0.00352; 0.00141	4
Anas platyrhynchos	3442	154	189722	516.2; 535.3	0.01074; 0.0089	3
Anser anser	2079	153	116590	344.3; 396.7	0.00933; 0.00734	28
Anser brachyrhynchus	246	136	28400	83.5; 85.6	0.00435; 0.00218	14
Branta canadensis	2001	149	86352	259.2; 299.9	0.00825; 0.00633	18
Branta leucopsis	188	134	16459	49.1; 51.7	0.00293; 0.00127	26
Buteo buteo	503	141	33185	109.9; 102.1	0.00555; 0.0025	20
Corvus corone	316	132	22654	71.2; 90.7	0.00361; 0.00214	1
Cygnus cygnus	154	119	9016	27.4; 25.3	0.00209; 0.00075	7
Cygnus olor	2448	150	128933	376.1; 431.9	0.00961; 0.00791	47
Falco peregrinus	106	140	8869	28.4; 29.5	0.00201; 0.00096	8
Falco tinnunculus	414	143	32792	106.7; 100.5	0.0054; 0.00247	2
Fulica atra	2386	150	113663	333.2; 390	0.00914; 0.00757	1
Fulmarus glacialis	1	24	24	1; 0	0.00547; 0.01054	1
Hydrocoloeus minutus	39	111	2332	7.1; 4.6	0.00087; 0.00075	1
Larus argentatus	915	152	43718	136.9; 148.1	0.00598; 0.00351	12
Larus canus	780	148	39136	121.6; 121.9	0.00571; 0.00323	3
Morus bassanus	2	76	275	3.6; 3.4	0.0015; 0.00466	4
Sterna hirundo	255	122	9023	31.1; 30.7	0.00218; 0.00104	12
Vanellus vanellus	1043	146	109460	317.1; 346.8	0.00896; 0.00658	1

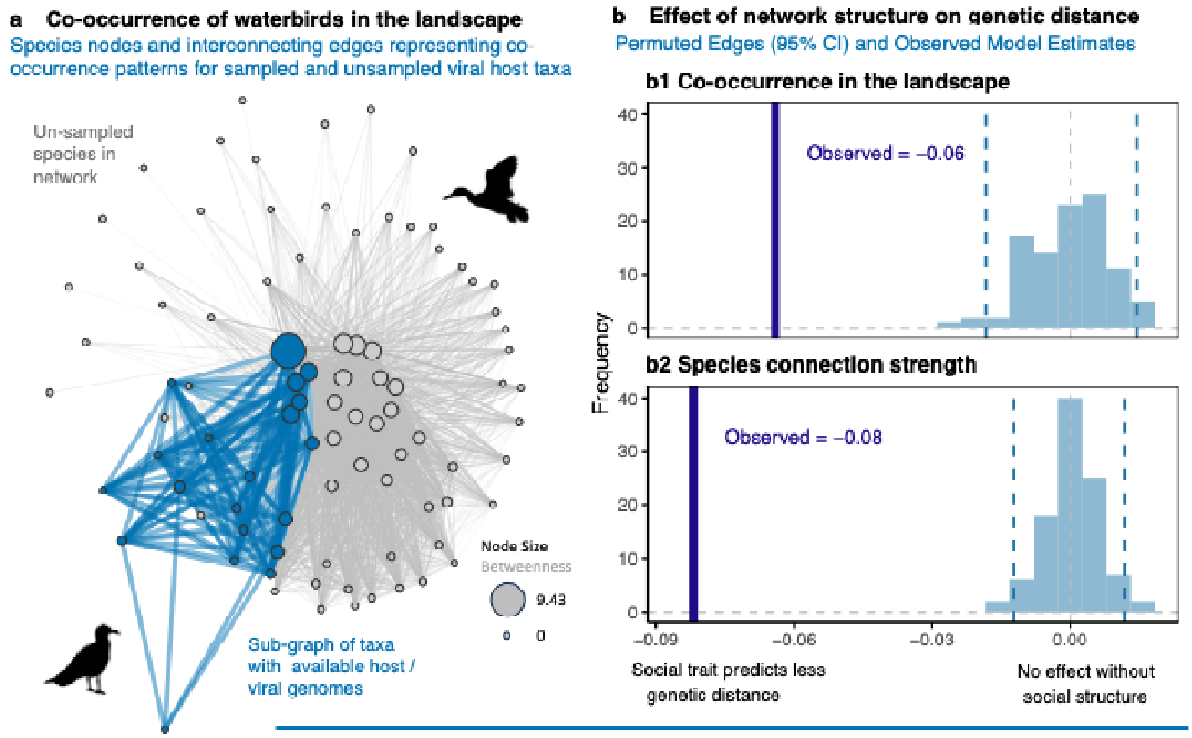

Figure 1. Network centrality predicts transmission pathways. a. Of the 159 species in the co-occurrence network, 20 were also associated with a genome for high pathogenicity avian influenza clade 2.3.4.4b, of which we had access to 205 unique sequecnes. The colour of the nodes and edges denote that a species pair had viral genome data available, blue, or not, grey. The node size denotes normalised betweenness value, where 9.43, sits on the greatest number of shortest paths between all of dyads in the network, in this case Mute Swan *Cygnus olor*, and 0, on no of the shortest paths (here including Northern Lapwing *Vanellus Vanellus*, Little egret *Egretta garzetta*, and Northern pintail *Anas acuta*). **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 connection strength). In both cases, the observed estimate falls outside the permuted distribution, suggesting that closer social proximity predicts less distant HPAI genomes - our results therefore differ from what is expected when social structure is removed from the network.