OIKOS

Research

Host traits associated with species roles in parasite sharing networks

Tad A. Dallas, Barbara A. Han, Charles L. Nunn, Andrew W. Park, Patrick R. Stephens and John M. Drake

T. A. Dallas (http://orcid.org/0000-0003-3328-9958) (tad.a.dallas@gmail.com), Centre for Ecological Change, Univ. of Helsinki, Helsinki, Finland. TAD, A. W. Park, P. R. Stephens and J. M. Drake, Odum School of Ecology, Univ. of Georgia, Athens, GA 30602, USA. AWP and JMD also at: Center for the Ecology of Infectious Diseases, Univ. of Georgia, Athens, GA, USA. – B. A. Han, Cary Inst. of Ecosystem Studies, Millbrook, NY, USA. – C. L. Nunn, Dept of Evolutionary Anthropology, Duke Univ., Durham, NC, USA.

The community of host species that a parasite infects is often explained by functional traits and phylogeny, predicting that closely related hosts or those with particular traits share more parasites with other hosts. Previous research has examined parasite community similarity by regressing pairwise parasite community dissimilarity between two host species against host phylogenetic distance. However, pairwise approaches cannot target specific host species responsible for disproportionate levels of parasite sharing. To better identify why some host species contribute differentially to parasite diversity patterns, we represent parasite sharing using ecological networks consisting of host species connected by instances of shared parasitism. These networks can help identify host species and traits associated with high levels of parasite sharing that may subsequently identify important hosts for parasite maintenance and transmission within communities. We used global-scale parasite sharing networks of ungulates, carnivores, and primates to determine if host importance – encapsulated by the network measures degree, closeness, betweenness, and eigenvector centrality – was predictable based on host traits. Our findings suggest that host centrality in parasite sharing networks is a function of host population density and range size, with range size reflecting both species geographic range and the home range of those species. In the full network, host taxonomic family became an important predictor of centrality, suggesting a role for evolutionary relationships between host and parasite species. More broadly, these findings show that trait data predict key properties of ecological networks, thus highlighting a role for species traits in understanding network assembly, stability, and structure.

Keywords: ecological network, host–parasite interactions, node centrality

---

Synthesis

The interactions between host and parasite species can be represented as a network where host species are connected by instances of shared parasitism. This representation allows the identification of specific host species, and associated trait profiles, that contribute strongly to parasite sharing. We found that host range size, host population density, and taxonomic family (e.g. Cervidae) were important for predicting host species contributions to parasite sharing. These findings have relevance to a breadth of ecological questions, ranging from understanding pathogen spillover to predicting the effects of environmental change on species interactions.

NSO
NORDIC SOCIETY OIKOS

www.oikosjournal.org

Introduction

Although the importance of multi-host parasite species has been established (Rigaud et al. 2010), disease ecologists are still far from understanding the sharing of the same parasites by different host species (hereafter, parasite sharing) and the structure of the complex interactions between host communities and their resident parasite communities (Hatcher et al. 2012). The integration of network theory into studies of host–parasite interactions has lead to several advances in disease ecology (Poulin 2010, Pilosof et al. 2015). For example, treating host–parasite interactions as a bipartite network, in which host and parasite species represent nodes of two different classes that interact through parasitism, has provided connections between graph theoretic measures (e.g. nestedness, modularity) and pathogen transmission and diversity (Guegan and Hugueny 1994, Fortuna et al. 2010, Poulin 2010, Dallas and Cornelius 2015). Topological measures of host-parasite networks allow for the comparison of networks across space and time (Poisot et al. 2012, Morris et al. 2014), and thus enable testing of biogeographical and macroecological hypotheses at the level of entire networks.

Unipartite projections of bipartite host–parasite networks – commonly referred to as co-occurrence networks – offer a way to examine parasite sharing among host species (Griffiths et al. 2014, Dallas and Poisot 2018). Nodes in these unipartite networks correspond to host species, with links between nodes representing instances of shared parasitism. Links are often weighted by either the number of shared parasite species or the similarity in parasite community composition between pairs of hosts, where similarity is measured using common community similarity indices (Legendre and Caceres 2013). This network representation has previously been used to examine species co-occurrence as a function of environmental change (Araujo et al. 2011), the determinants of microbial community distributions (Williams et al. 2014), parasite sharing among host individuals or species (Luis et al. 2015, Pilosof et al. 2015), and estimating pathogen spillover potential to humans (Gomez et al. 2013). By using this conceptual and analytical approach, it may be possible to quantify the relative importance of host species to parasite diversity and transmission and host species' contributions to overall network structure (Pilosof et al. 2015).

The relative importance of a node to the overall network is commonly referred to as centrality. Measures of centrality have been used to estimate node importance to network structure (Newman 2010), examine aspects of network stability and collapse (Holme et al. 2002, Radicchi 2015), identify target genes potentially associated with disease (Navlakha and Kingsford 2010, Ma et al. 2016) and develop targeted vaccination strategies for social contact networks (Rushmore et al. 2014). This suggests that centrality is a useful measure of node importance across many different network types. Further, centrality is conceptually related to measures of beta diversity from community ecology, as central species should correspond to high site-level beta diversity (Box 1). With regards to parasite sharing networks, centrality may be used to identify host species that are important for parasite sharing (Walker et al. 2017). While previous studies have largely focused on identifying correlates of parasite species richness (a quantity related to some measures of centrality), one study of a primate parasite sharing network examined primate parasite sharing networks using the same centrality measures used here, finding that population density and geographic range size were associated with large host centrality (Gomez et al. 2013). We build on this previous effort, applying machine

---

Box 1: Centrality as a measure of beta diversity

Centrality in co-occurrence networks may have an under-appreciated role in measuring parasitism in a given host, as it captures aspects of both alpha and beta diversity. Host–parasite interaction data are typically recorded as a list of parasite species that infect a host species, which means that scientists only have information on the existence of an interaction between host and parasite species, but not on the frequency, duration, or intensity of the interaction. This restricts the options for measuring parasite diversity, as many site-level diversity indices require information on the occurrence frequency (Gotelli and Colwell 2011), which corresponds to the number of times the parasite was found on a host species. Pairwise indices of beta diversity designed for presence–absence data are useful for examining dissimilarity between sites, but do not provide a site-level (i.e. host species level) estimate of diversity. Koleff et al. (2003) decomposed a set of 24 beta diversity indices into three components (a, b and c). They defined a as the number of shared interactions between two sites, and b and c as those interactions unique to each site. Many of these dissimilarity metrics weight unique and shared portions of the communities equally. For many questions associated with the community ecology of infectious disease, however, it is more important to know the instances of shared parasitism between two host species (i.e. component a) instead of the parasites unique to either host species (i.e. components b and c). This information is important, for example, when considering generalist parasites and their distribution in a community, or when considering risks of infection to novel hosts in a community (Fenton and Pedersen 2005). Whereas the shared parasite diversity component is pairwise between host species, and therefore cannot be directly related to aspects of host ecology, centrality is a composite measure of shared parasite diversity for a given host species and can be linked to host ecology, phylogeny, behavior, and other characteristics. Further, the interpretation of centrality measures and site-level measures of beta diversity are conceptually similar, as beta diversity at the site level attempts to identify compositional differences among a set of sites, while centrality captures the instances of shared species among sites. Further, many of the centrality measures directly incorporate information on all sites, providing information not only on the shared community between sites, but on shared species across the network of sites i.e. measures borrow information on the shared components of completely different sites for many centrality measures.

learning techniques to an expanded set of host groups to develop cross-validated models capable of identifying host traits associated with centrality.

More generally, we aim to determine whether centrality is a predictable trait of a host species. The loss of these important host species could drastically alter host–parasite network structure (Dallas and Cornelius 2015), or alter infection dynamics in other host species (Lacroix et al. 2014), which could enhance pathogen spillover potential to humans. A highly central host species could be a source of many parasites to other species, but this need not be the case; for example, a host with higher centrality may simply be more susceptible to infection by a wide range of parasites, including parasites originating from hosts not included in the network. Another possibility is that a highly central host species could be a 'sink', i.e. able to become infected by many parasite species, but less able to transmit them to new host species. In this case, targeting central host species for removal or anti-parasite treatment would have little or no effect on disease dynamics in other hosts of that community.

Here, we examine four measures of centrality that capture different aspects of node importance: degree, betweenness, closeness, and eigenvector centrality (Fig. 1). Degree centrality is simply the summed number of links per node, or the total

number of shared parasite species with other host species. It is important to note that a single parasite species can be shared with multiple host species, and can be counted multiple times, such that degree centrality is not equivalent to parasite species richness (Supplementary material Appendix 1 Fig. A15). Betweenness centrality is the number of shortest paths between all other nodes that pass through a focal node (Newman 2010). Host species with high betweenness are those that share parasites with groups of hosts that do not share many parasites, representing a potential pathway for parasite sharing between host groups. Closeness centrality measures the relative distance of a focal node to all other nodes in the network; as such, this measure identifies nodes that are close to many other nodes (Kiss and Bichler 2008, Newman 2010, Poulin et al. 2013). Host species which share many parasites with many host species would have high closeness centrality. In this sense, closeness centrality contains more information than degree centrality, as closeness utilizes information on all nodes in the network, while degree is concerned only with the number of links to a focal node (and thus is similar to the distinction between local and global centrality). Lastly, we examined eigenvector centrality, which measures the tendency for a host species to share parasites with host species that share many parasites with many other host species. The underlying logic is that host species connected to well-connected hosts are important for understanding disease ecology, a nuance that many centrality measures fail to capture. Preferential removal of species with larger eigenvector centrality measures in food webs results in fast network collapse (Allesina and Pascual 2009), suggesting that hosts with large eigenvector centrality likely play a key role in food web dynamics and that hosts differ in their contributions to food web stability.

We used data from the Global Mammal Parasite Database (Nunn and Altizer 2005, Stephens et al. 2017) to create a large parasite sharing network of 385 host species derived from over 16 000 interactions between host and parasite species. Using these data, we estimated host centrality (i.e. degree, closeness, betweenness and eigenvector) for all host species, and independently for taxonomic subsets (carnivores, ungulates and primates). Under the hypothesis that centrality reflects the importance of a host to parasite diversity and transmission, we predicted that centrality measures would covary with host traits – i.e. attributes of particular nodes in our network – and phylogeny, focusing especially on traits that enhance parasite encounter probabilities and range overlap with other host species (e.g. host geographic range; Gomez et al. 2013). Under neutral dynamics – in which host-parasite associations occur through encounter processes since hosts are presumably equivalent – host population density and geographic range size would determine parasite sharing. Given that centrality measures capture different aspects of network structure, we expected to find variation in which host traits were important to parasite sharing. However, we found host trait importance was relatively consistent across both host taxonomic level and centrality measure used. Together, our findings provide a link between properties of host ecology and importance in parasite sharing networks, suggesting that host traits underlie patterns of parasite sharing and diversity.
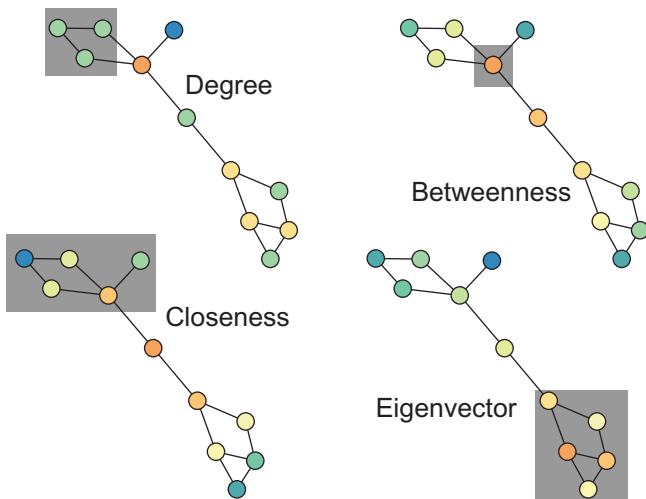

Figure 1. A simple representation of parasite sharing networks highlighting the centrality metrics examined. Circles represent host species, with links between circles representing instances of parasite sharing. Here, parasite sharing links are essentially unweighted for ease of display, though variation in link weight is important to all centrality measures considered. Warmer colors (yellow and orange) correspond to higher ranked centrality values for a given type of centrality. Key differences among centrality measures are highlighted using grey boxes. For instance, the grey box on the degree graph identifies three nodes with equivalent degree, but different values for other centrality measures. Further, the upper cluster of nodes has higher closeness centrality, whereas the lower cluster has higher eigenvector centrality (grey boxes in closeness and eigenvector centrality graphs), and betweenness identifies a single node of the upper cluster (grey box in betweenness graph) as the most important node in the graph.

Methods

Host–parasite interactions of the Global Mammal Parasite Database

We used the Global Mammal Parasite Database (GMPD) for our analyses (Nunn and Altizer 2005, Stephens et al. 2017), which is divided into host taxonomic subsets of carnivores, ungulates and primates. Following previous work with GMPD (Han et al. 2016), we removed several host species prior to analyses, including aquatic carnivores i.e. species in families Otariidae, Phocidae and Odobenidae, and twelve domesticated animal species (e.g. *Sus scrofa*; see Supplementary material Appendix 1 for a full list). Parasite species classified at the genus-level were included in the current analyses, though the results are similar when these parasite species are removed. Parasite species identified to the genus level make up 18% of the parasite occurrence records in the GMPD, and their removal did not strongly influence co-occurrence network structure and the resulting measures of centrality (Supplementary material Appendix 1). More information on removed host and parasite species is available in the Supplementary material Appendix 1, and the provided code (see data and code at <https://doi.org/10.6084/m9.figshare.5129980>). With these considerations, we were left with information on the interactions between 385 unique host species and over 1800 unique parasite species.

Species trait data (Supplementary material Appendix 1 Table A6) were obtained primarily from PanTHERIA (Jones et al. 2009), though some missing data for ungulates were included from <www.ultimateungulate.com/>. Host taxonomy was incorporated by using Order and Family as covariates, while host trait covariates included morphological (body length, body mass, brain weight), life history (maximum age, age at sexual maturity, diet breadth, activity cycle, terrestriality), geographic (home range size, geographic range, habitat breadth), and reproductive (gestation length, interbirth interval, litters per year, litter size, neonate body mass, and weaning age) characteristics. Terrestiality refers to species habitat preferences (either aboveground or belowground). Geographic and life history traits can influence transmission patterns, because features such as diet breadth and geographic range size are directly related to the likelihood of contacting parasite species, which is relevant to assessing whether host–parasite networks are the result of neutral processes. Morphological and reproductive covariates capture host species' pace of life, potentially also capturing variation relating to immune investment (Lee 2006). Further, increasing body length and body mass may provide more colonization opportunities for parasite species, which could influence centrality estimates. Traits with data available for at least 80% of host species were included in analyses.

In addition to species trait and taxonomic data, evolutionary distinctiveness was calculated for each species based on a published phylogenetic supertree of all mammals (Bininda-Emonds et al. 2007, Fritz et al. 2009), which captures evolutionary history among host species by identifying those from clades with slower diversification rates, or that have fewer close relatives. Previous work demonstrated that parasite species richness tended to decrease with host evolutionary distinctiveness (Park et al. 2018), which is likely to influence patterns of parasite sharing among host species. Collectively, taxonomic family and evolutionary distinctiveness capture two different aspects of evolutionary history, as taxonomic family typically captures more recent evolutionary history, while evolutionary distinctiveness may reflect longer term evolutionary trajectories. Evolutionary distinctiveness was calculated using both equal splits (Redding and Mooers 2006) and fair proportions (Isaac et al. 2007) methods implemented in the R (<www.r-project.org>) package picante (Kembel et al. 2010). Both measures were strongly correlated, here we report results using the fair proportion measure of distinctiveness.

We also investigated phylogenetic signal in model residuals using the K statistic. This measure captures variation in model performance as a result of phylogenetic relationships (Revell et al. 2008) by comparing variance of phylogenetically independent contrasts of empirical data and a null model. This is different from evolutionary distinctiveness (discussed above), which captures evolutionary isolation of species, and has previously been related to parasite diversity (Huang et al. 2015).

Addressing potential sampling biases

The degree to which host species were studied could influence the number of parasite species found, and could potentially influence the link weight in parasite sharing networks because link weight is influenced by the sampling effort for a given host species. Following Nunn et al. (2003), Altizer et al. (2007), we attempted to account for sampling biases by including citation counts for hosts as a predictor variable in our analyses. To do this, we searched Web of Science for the Latin binomial names of each host species plus the word 'parasite', and determined the number of associated citations. This method has been previously used to account for sampling effort in parasitological studies (Ezenwa et al. 2006, Stephens et al. 2017).

However, the association between the number of published scientific studies and the intensity of sampling for parasites may be quite weak, as parasitologists may preferentially sample host–parasite interactions in a single species context (e.g. plague in prairie dogs), which would serve to increase publication count while not influencing sampling intensity for the overall size of the parasite community of that host species. We examine this in the Supplementary material Appendix 1 Fig. A2. Additionally, citation counts obtained from Web of Science vary among Universities or research centers, as bibliographic database access is tailored to the institution (Dallas et al. 2018). While citation count did generally improve model accuracy (measured as correlation between model predicted and observed centrality), the effect was

marginal, and citation counts were rarely among the strongest contributors to model accuracy, suggesting that interactions between citation count and host traits were responsible for the slightly improved accuracy. For instance, citation count is related to host body mass ($\rho = 0.16$, $t = 2.79$, $p = 0.006$), geographic range size ($\rho = 0.12$, $t = 2.10$, $p = 0.04$), and home range size ($\rho = 0.17$, $t = 2.39$, $p = 0.018$). Given these concerns, we report models without citation counts in the main text, and place results that control for citation counts in the Supplementary material Appendix 1.

Parasite sharing network formation and host species centrality

We created three parasite sharing subnetworks, composed of taxonomic subsets of hosts, including ungulates (n = 104), carnivores (n = 108), and primates (n = 173), and one large parasite sharing network containing all host species (n = 385). Centrality measures (i.e. degree, betweenness, closeness and eigenvector centrality) for each host species were calculated using the igraph R package (Csardi and Nepusz 2006). We investigated the relationship between parasite species richness per host species and centrality measures in the Supplementary material Appendix 1.

Boosted regression tree models of host centrality

Boosted regression tree (BRT) models were used to assess how host and phylogenetic characteristics influence centrality. BRT models were trained on all groups together and each host group (e.g. carnivores), and for each centrality measure using the gbm R package (Ridgeway 2015). This analysis was chosen for its flexibility and numerous advantages over more traditional regression approaches, including allowing for non-linear responses, collinear predictors, and variable interactions (see Elith et al. 2008 for full discussion). The process of boosting allows for the combination of many 'weak learners' – trees with a small number of splits – where subsequent trees attempt to explain the remaining variation. Models were trained with learning rate of 0.001, a commonly used value (Elith et al. 2008), and an interaction depth of 4, which allows for interactions among covariates. Given the limited data available, we did not perform out-of-sample prediction. However, to reduce the possibility of overfitting, models were five-fold cross validated, and the optimal number of trees was chosen using a cross-validation procedure (max trees = 50 000).

The relative importance of each predictor variable in BRT models was determined by calculating the relative improvement in model fit as a result of including a given predictor, weighted by the number of times the predictor appeared in the collection of trees (Breiman et al. 2001). These relative contribution values were then scaled such that each predictor variable has a value between 0 and 100, and all values sum to 100, with larger numbers corresponding to a larger relative influence on model fit. The effect of each predictor variable on host species centrality was examined using partial

dependence plots, which visualize the relationship between predictor and response after accounting for other predictors (Elith et al. 2008). Model accuracy was examined using two correlation measures. First, we calculated Spearman's rank correlation values between predicted and observed centrality values to capture the ability of our models to correctly rank nodes in their respective centrality order based on host covariates. In many applications, it is more important to be able to rank nodes in order of importance to the network than to predict actual centrality values. However, raw centrality values may provide a continuous measure of parasite diversity at the host level that is sensitive to the existing host and parasite community. That is, changes in host centrality with the addition of a novel parasite or host species could provide a measure of relative effect on each host species. We report rank correlations in the main text, as the successful ranking of host species is our primary goal, but report Pearson's correlations in the Supplementary material Appendix 1.

Results

Important predictors of centrality

Across all host groups and centrality measures, traits describing host spatial habitat usage (i.e. home range size, geographic range size, and population density within geographic ranges) and host taxonomy (i.e. host family) were important predictors of host centrality in parasite sharing networks, suggesting that densely populated or more widespread host species tend to share a greater number of parasites with a greater number of other host species (Fig. 2–5). Covariates of host range use were largely positively correlated with centrality measures, though some non-linear relationships can be observed in the partial dependence plots (Supplementary material Appendix 1 Fig. A7–A10). For instance, the effect of ungulate species population density on betweenness centrality peaks at an intermediate value, creating a hump-shaped partial dependence plot (top panel of Supplementary material Appendix 1 Fig. A8).

Further, in some instances the effect was opposite, with larger home range size or population density actually decreasing the marginal effect on centrality (e.g. effect of home range size on closeness centrality in carnivores; Supplementary material Appendix 1 Fig. A9). However, geographic range size had a universal positive relationship with centrality measures in all models where it was in the top 3 predictor variables (7 of 16 models; Supplementary material Appendix 1 Fig. A7–A10), suggesting that widespread host species are also the most central in parasite sharing networks, even considering that parasite species richness may not be strongly related to species centrality (Supplementary material Appendix 1 Fig. A15).

In models containing all three host groups, Antilocapridae was the most important family for models of all four centrality measures. Within each group, host family was an important covariate, and many of the most important host families
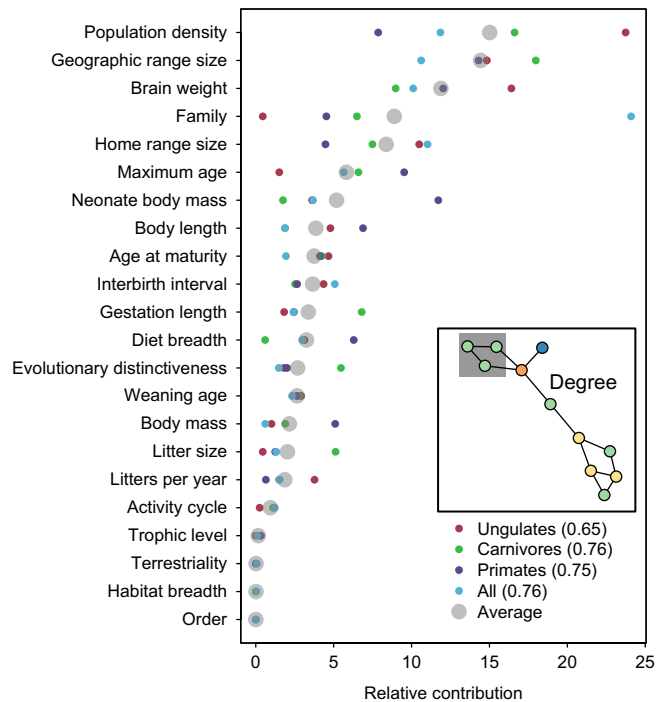
Figure 2. The relative importance of host trait covariates to the boosted regression tree model of degree centrality for ungulates (red), carnivores (green), primates (purple), and all hosts combined (blue). Host traits are sorted based on average variable importance, and numeric values in legend correspond to Spearman's rank correlation coefficients between predicted and observed centrality values.
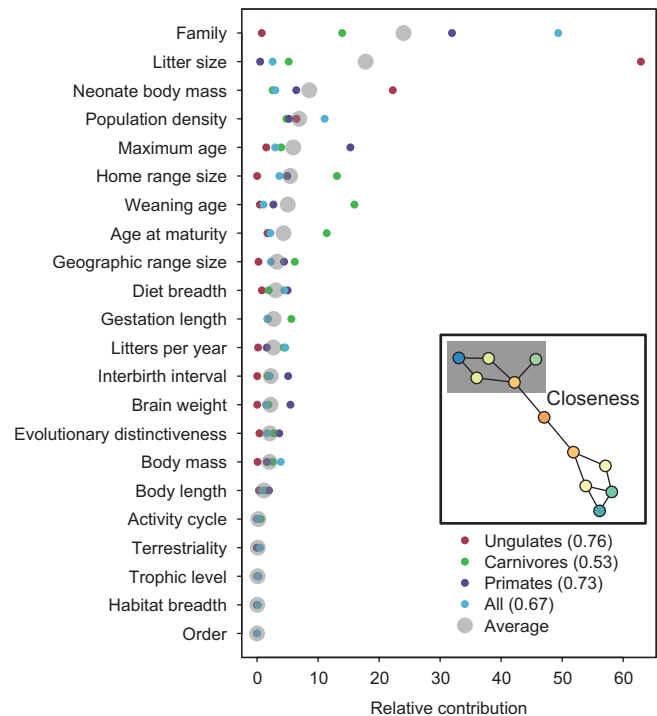

Figure 3. The relative importance of host trait covariates to the boosted regression tree model of closeness centrality for ungulates (red), carnivores (green), primates (purple), and all hosts combined (blue). Host traits are sorted based on average variable importance, and numeric values in legend correspond to Spearman's rank correlation coefficients between predicted and observed centrality values.

were conserved across centrality measures for primates (e.g. Atelidae and Cheirogaleidae), ungulates (e.g. Antilocapridae and Cervidae), and carnivores (e.g. Felidae and Ursidae). These families do not necessarily have the highest parasite species richness, or the most number of species, suggesting that species within these families may be infected by generalist parasites. Further research is warranted, as species within these families could be especially important to the maintenance of parasite diversity. While family-level taxonomic information was important, evolutionary distinctiveness was not consistently ranked as important, apart from betweenness centrality in ungulates (Fig. 4). These two measures are different in that evolutionary distinctiveness captures long term evolutionary history, while taxonomic family may capture more recent evolutionary processes as well as capturing unmeasured trait variation. This result, combined with our failure to detect a phylogenetic signal in the model residuals (Supplementary material Appendix 1 Table A8), suggest a limited effect of host evolutionary history on host importance to parasite sharing.

Model performance

Boosted regression models predicted the centrality rank of each node: The node centrality rankings predicted by the model were significantly correlated to the observed node

centrality rank across each of the 16 models, which included 4 measures of centrality for each of 4 host groups ($\bar{\rho} = 0.67$). Further, boosted regression models captured the raw centrality scores (Supplementary material Appendix 1 Table A7) for degree (Fig. 2), closeness (Fig. 3), betweenness (Fig. 4), and eigenvector centrality (Fig. 5).

Important covariates related to node importance tended to be consistent for host groups regardless of centrality measure. To estimate this consistency, we calculated rank correlation coefficients between the orders of predictor variables of different models. All rank correlations were significant and positive, suggesting that covariates maintained approximately the same rank in order of importance across centrality measures and host groups. Mean rank correlation coefficients, averaged across different measures of centrality, between the order of host trait covariates were high on average for ungulates ($\bar{\rho} = 0.59$), carnivores ($\bar{\rho} = 0.72$), primates ($\bar{\rho} = 0.52$), and all host groups together ($\bar{\rho} = 0.69$). Further, mean rank correlation coefficients, averaged across host groups, between the order of host covariates were high on average across host groups for degree ($\bar{\rho} = 0.80$), closeness ($\bar{\rho} = 0.43$), betweenness ($\bar{\rho} = 0.65$), and eigenvector ($\bar{\rho} = 0.65$) centrality. This suggests that important predictor variables were conserved within host groups across different measures of centrality, and among centrality measures across
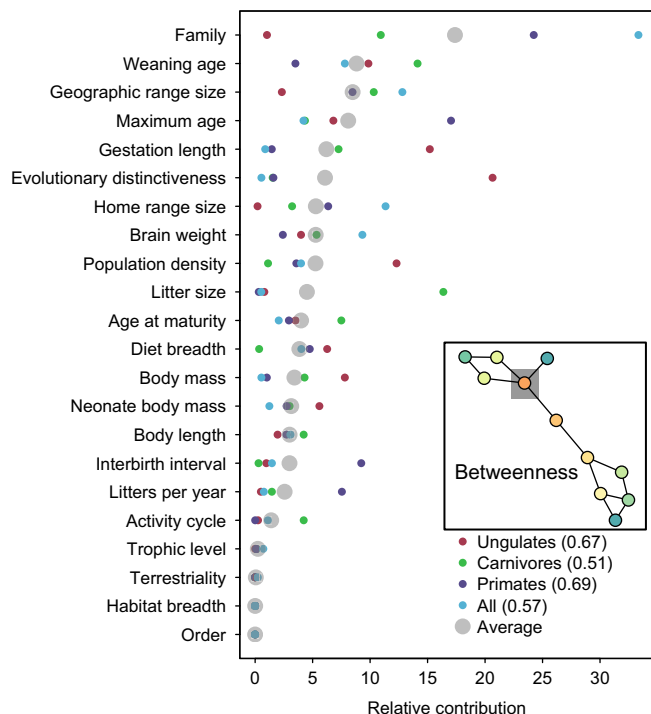
Figure 4. The relative importance of host trait covariates to the boosted regression tree model of betweenness centrality for ungulates (red), carnivores (green), primates (purple), and all hosts combined (blue). Host traits are sorted based on average variable importance, and numeric values in legend correspond to Spearman's rank correlation coefficients between predicted and observed centrality values.
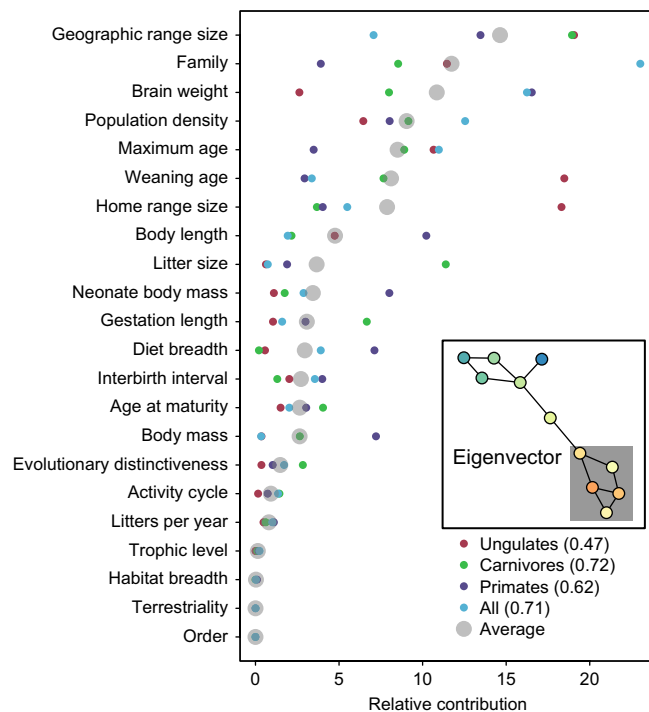

Figure 5. The relative importance of host trait covariates to the boosted regression tree model of eigenvector centrality for ungulates (red), carnivores (green), primates (purple), and all hosts combined (blue). Host traits are sorted based on average variable importance, and numeric values in legend correspond to Spearman's rank correlation coefficients between predicted and observed centrality values.

different host groups (Supplementary material Appendix 1 Fig. A16–A19). This finding is interesting, as centrality measures themselves were only weakly related to one another (Supplementary material Appendix 1 Fig. A11–A14). On the other hand, the small differences in covariate ranking between centrality measures may suggest that researchers select the most appropriate centrality index for the research question, as centrality measures vary in how much the focal site (host species) matters relative to the structure of the entire network of shared interactions. Lastly, the inclusion of Web of Science citation counts, a proxy for sampling bias, did not strongly increase model accuracy (Supplementary material Appendix 1 Fig. A3–A6).

Discussion

Central host species in parasite sharing networks are largely those that have a disproportionate number of parasite interactions (or those that connect otherwise unconnected groups of host species). High host centrality may result from several different pathways: a host species could be infected by many generalist parasite species (i.e. high degree and closeness centrality), a host could share parasites with two

host groups that do not share many parasites (high betweenness centrality), or a host species may share parasites with another host species that shares parasites with many other host species (high eigenvector centrality). Interestingly, we found that the host traits most important for host species centrality either do not strongly depend on such pathways, or that the traits are conserved such that the same set of host traits are important predictors of multiple centrality measures. Specifically, aspects of host ecology related to species habitat utilization – including home range size, geographic range size, population density, and activity patterns – were among the top predictors of host centrality in networks of parasite sharing. While this partly follows neutral expectations, the importance of taxonomy and other host traits suggests that host–parasite interactions are certainly not entirely neutral, as has been suggested previously (Canard et al. 2014). This deviation from neutrality is potentially the result of host specificity exhibited by parasite species, coevolutionary relationships between host and parasite species, and host–parasite trait mismatches which prevent infection (e.g. short parasite mouthparts and a thick host skin). Together, this suggests that host importance in parasite sharing networks is a quantity that has a basis in host traits, and that interaction patterns may be predictable based on host traits (Dallas et al. 2017).

Taxonomic family membership was an important predictor of host centrality in the full model (which combined all host groups) but not in the group-specific models (carnivores, primates, ungulates). This could indicate that phylogeny is an important predictor of host centrality among diverse and more phylogenetically distinct groups, but not within them. Our analyses reveal certain families that are the most influential to parasite sharing across species (e.g. Antilocapridae). It is possible that in the absence of directly incorporating phylogenetic information, some of the correlations that we observed may reflect the influence of shared descent rather than true evolutionary correlations. However, we found little evidence for the presence of phylogenetic signal in model residuals. Moreover, given that the biological variables with the highest importance scores were identical across three distantly related mammal groups, our results involving functional associations were unlikely to be driven purely by phylogenetic effects, and suggest that phylogenetic relationships are less important than host traits in determining host species centrality in parasite sharing networks (Supplementary material Appendix 1 Table A8).

In addition to looking across numerous taxa and numerous traits per taxa, an advantage of the statistical learning approach applied here is that our analyses account for multi-collinearity among traits, and provide a picture of how important particular traits are relative to others (both within and across multiple taxa). Some of the results we report here are corroborated by previous, taxa-specific studies which found host evolutionary history, host traits, and geographic range size to be important in predicting parasite sharing (Cooper et al. 2012, Gomez et al. 2013). For example, Gomez et al. (2013) used a subset of the GMPD data on primate species and found that geographic range size and phylogeny were important predictors of centrality in parasite sharing among primates. In addition, we found that traits related to species range usage were strongly associated with different centrality measures, and centrality for different host taxa. For example, traits related to species life history (e.g. maximum age), reproduction (e.g. neonate body mass), and morphology (e.g. brain weight) were also important predictors of primate host centrality, suggesting that parasite sharing is a reflection of intrinsic features dictating host life history. Meanwhile, some traits previously found to be important to parasite sharing (i.e. trophic level) did not contribute strongly to model fit (Chen et al. 2008), though this is potentially the result of how trophic level was defined (see Supplementary material Appendix 1 for further detail). Many of the traits identified as most important to parasite sharing were similar across host taxa, suggesting that some important drivers of parasite sharing may be consistent among all mammals. Lastly, our results collectively suggest that host centrality in parasite sharing networks is a quantity that can be captured and well predicted by host species traits, even when centrality measures are not strongly related to one another (Valente et al. 2008) (Supplementary material Appendix 1 Fig. A11–A14).

The topological approach to understanding parasite sharing that we adopt here does not account for changing parasite associations (i.e. link re-wiring), which could be important in the context of introduced species and spillover of parasites (Young et al. 2016). For instance, the infection of a novel host species would add a link, or change the weight of an existing link, in the parasite sharing network. It is not known how the role of parasite host switching, invasion of novel host or parasite species, or temporal dynamics of host–parasite interactions influence variation in host centrality in parasite sharing networks, yet these processes are likely to impact which host individuals or species are most likely to share parasites with other individuals or species. Given that host centrality is predictable based on host traits, the relative importance of introduced host species to parasite sharing should also be estimable on the basis of traits, even when no data on host–parasite interactions for the introduced host species are known (Schmidt et al. 2012). Contrary to the host-centric view employed here, a similar approach that takes a parasite-centric view may be successful in identifying parasite species able to form links between host species, which could be important in pre-empting parasite spillover events (Luis et al. 2015).

In conclusion, we applied a machine learning approach to investigate the predictors of parasite sharing in mammals, with parasite sharing defined as different host species being regularly infected with the same parasite species. We found that four measures of centrality were predictable from host characteristics involving population density, range size, and taxonomic family, with effects varying in strength among the different mammalian orders that we examined. These findings have relevance to a broad range of ecological topics, ranging from understanding pathogen spillover to predicting the effects of environmental change. We also showed how centrality can be treated as a measure of beta diversity. Future research could apply these approaches to a projection of host sharing by parasites (the parasite's perspective), while also applying these methods to current challenges in global change.

References

Allesina, S. and Pascual, M. 2009. Googling food webs: can an eigenvector measure species' importance for coextinctions. – PLoS Comput. Biol 5: e1000494.

Altizer, S. et al. 2007. Do threatened hosts have fewer parasites? A comparative study in primates. – J. Anim. Ecol. 76: 304–314.

Araujo, M. B. et al. 2011. Using species co-occurrence networks to assess the impacts of climate change. – Ecography 34: 897–908.

Bininda-Emonds, O. R. et al. 2007. The delayed rise of present-day mammals. – Nature 446: 507–512.

Breiman, L. et al. 2001. Statistical modeling: the two cultures. – Stat. Sci. 16: 199–231.

Canard, E. et al. 2014. Empirical evaluation of neutral interactions in host-parasite networks. – Am. Nat. 183: 468–479.

Chen, H.-W. et al. 2008. Network position of hosts in food webs and their parasite diversity. – Oikos 117: 1847–1855.

Cooper, N. et al. 2012. Phylogenetic host specificity and understanding parasite sharing in primates. – Ecol. Lett. 15: 1370–1377.

Csardi, G. and Nepusz, T. 2006. The igraph software package for complex network research. – InterJournal: Complex Systems, p. 1695.

Dallas, T. and Cornelius, E. 2015. Co-extinction in a host–parasite network: identifying key hosts for network stability. – Sci. Rep. 5: 13185.

Dallas, T. A. and Poisot, T. 2018. Compositional turnover in host and parasite communities does not change network structure. – Ecography 41: 1534–1542.

Dallas, T. et al. 2017. Predicting cryptic links in host–parasite networks. – PLOS Comput. Biol. 13: e1005557.

Dallas, T. et al. 2018. Variable bibliographic database access could limit reproducibility. – BioScience 68: 552–553.

Elith, J. et al. 2008. A working guide to boosted regression trees. – J. Anim. Ecol. 77: 802–813.

Ezenwa, V. O. et al. 2006. Host traits and parasite species richness in even and odd-toed hoofed mammals, Artiodactyla and Perissodactyla. – Oikos 115: 526–536.

Fenton, A. and Pedersen, A. B. 2005. Community epidemiology framework for classifying disease threats. – Emerg. Infect. Dis. 11: 1815–1821.

Fortuna, M. A. et al. 2010. Nestedness versus modularity in ecological networks: two sides of the same coin? – J. Anim. Ecol. 79: 811–817.

Fritz, S. A. et al. 2009. Geographical variation in predictors of mammalian extinction risk: big is bad, but only in the tropics. – Ecol. Lett. 12: 538–549.

Gomez, J. M. et al. 2013. Centrality in primate–parasite networks reveals the potential for the transmission of emerging infectious diseases to humans. – Proc. Natl Acad. Sci. USA 110: 7738–7741.

Gotelli, N. J. and Colwell, R. K. 2011. Estimating species richness. – In: Biological diversity: frontiers in measurement and assessment. Oxford Univ. Press, pp. 39–54.

Griffiths, E. C. et al. 2014. Analysis of a summary network of co-infection in humans reveals that parasites interact most via shared resources. – Proc. R. Soc. B 281: 20132286.

Guegan, J.-F. and Hugueny, B. 1994. A nested parasite species subset pattern in tropical fish: host as major determinant of parasite infracommunity structure. – Oecologia 100: 184–189.

Han, B. A. et al. 2016. Global patterns of zoonotic disease in mammals. – Trends Parasitol. 32: 565–577.

Hatcher, M. J. et al. 2012. Diverse effects of parasites in ecosystems: linking interdependent processes. – Front. Ecol. Environ. 10: 186–194.

Holme, P. et al. 2002. Attack vulnerability of complex networks. – Phys. Rev. E 65: 056109.

Huang, S. et al. 2015. Parasite diversity declines with host evolutionary distinctiveness: a global analysis of carnivores. – Evolution 69: 621–630.

Isaac, N. J. et al. 2007. Mammals on the EDGE: conservation priorities based on threat and phylogeny. – PLoS One 2: e296.

Jones, K. E. et al. 2009. PanTHERIA: a species-level database of life history, ecology, and geography of extant and recently extinct mammals: Ecological Archives E090-184. – Ecology 90: 2648–2648.

Kembel, S. W. et al. 2010. Picante: R tools for integrating phylogenies and ecology. – Bioinformatics 26: 1463–1464.

Kiss, C. and Bichler, M. 2008. Identification of influencers measuring influence in customer networks. – Decis. Support Syst. 46: 233–253.

Koleff, P. et al. 2003. Measuring beta diversity for presence–absence data. – J. Anim. Ecol. 72: 367–382.

Lacroix, C. et al. 2014. Non-random biodiversity loss underlies predictable increases in viral disease prevalence. – J. R. Soc. Interface 11: 20130947.

Lee, K. A. 2006. Linking immune defenses and life history at the levels of the individual and the species. – Integr. Comp. Biol. 46: 1000–1015.

Legendre, P. and Caceres, M. 2013. Beta diversity as the variance of community data: dissimilarity coefficients and partitioning. – Ecol. Lett. 16: 951–963.

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

Ma, B. et al. 2016. Geographic patterns of co-occurrence network topological features for soil microbiota at continental scale in eastern China. – ISME J. 10: 1891–1901.

Morris, R. J. et al. 2014. Antagonistic interaction networks are structured independently of latitude and host guild. – Ecol. Lett. 17: 340–349.

Navlakha, S. and Kingsford, C. 2010. The power of protein interaction networks for associating genes with diseases. – Bioinformatics 26: 1057–1063.

Newman, M. 2010. Measures and metrics. – In: Networks: an introduction. Oxford Univ. Press, pp. 168–220.

Nunn, C. L. and Altizer, S. M. 2005. The global mammal parasite database: an online resource for infectious disease records in wild primates. – Evol. Anthropol. Issues News Rev. 14: 1–2.

Nunn, C. L. et al. 2003. Comparative tests of parasite species richness in primates. – Am. Nat. 162: 597–614.

Park A. et al. 2018. Characterizing the phylogenetic specialism-generalism spectrum of mammal parasites. – Proc. R. Soc. B 285: 20172613.

Pilosof, S. et al. 2015. Potential parasite transmission in multi-host networks based on parasite sharing. – PLoS One 10: e0117909.

Poisot, T. et al. 2012. The dissimilarity of species interaction networks. – Ecol. Lett. 15: 1353–1361.

Poulin, R. 2010. Network analysis shining light on parasite ecology and diversity. – Trends Parasitol. 26: 492–498.

Poulin, R. et al. 2013. Phylogeny determines the role of helminth parasites in intertidal food webs. – J. Anim. Ecol. 82: 1265–1275.

Radicchi, F. 2015. Predicting percolation thresholds in networks. – Phys. Rev. E 91: 010801.

Redding, D. W. and Mooers, A. O. 2006. Incorporating evolutionary measures into conservation prioritization. – Conserv. Biol. 20: 1670–1678.

Revell, L. J. et al. 2008. Phylogenetic signal, evolutionary process and rate. – Syst. Biol. 57: 591–601.

Ridgeway, G. 2015. gbm: generalized boosted regression models. – R package ver. 2.1.1.

Rigaud, T. et al. 2010. Parasite and host assemblages: embracing the reality will improve our knowledge of parasite transmission and virulence. – Proc. R. Soc. B 277: 3693–3702.

Rushmore, J. et al. 2014. Network-based vaccination improves prospects for disease control in wild chimpanzees. – J. R. Soc. Interface 11: 20140349.

Schmidt, J. P. et al. 2012. Bioeconomic forecasting of invasive species by ecological syndrome. – Ecosphere 3: 1–19.

Stephens, P. R. et al. 2017. Global mammal parasite database ver. 2.0. – Ecology 98: 1476.

Valente, T. W. et al. 2008. How correlated are network centrality measures? – Connections 28: 16.

Walker, J. et al. 2017. Uncertain links in host-parasite networks: lessons for parasite transmission in a multi-host system. – Phil. Trans. R. Soc. B 372: 20160095.

Williams, R. J. et al. 2014. Demonstrating microbial co-occurrence pattern analyses within and between ecosystems. – Front. Microbiol. 5: 358.

Young, H. S. et al. 2016. Introduced species, disease ecology, and biodiversity–disease relationships. – Trends Ecol. Evol. 32: 41–54.

Supplementary material (available online as Appendix oik-05602 at <www.oikosjournal.org/appendix/oik-05602>). Appendix 1.