Social network models predict movement and connectivity in ecological landscapes

Robert J. Fletcher, Jr.[a,1], Miguel A. Acevedo[a,b], Brian E. Reichert[a], Kyle E. Pias[a], and Wiley M. Kitchens[c]

[a]Department of Wildlife Ecology and Conservation, [b]School of Natural Resources and Environment, and [c]US Geological Survey, Florida Cooperative Fish and Wildlife Research Unit, University of Florida, Gainesville, FL 32611

Network analysis is on the rise across scientific disciplines because of its ability to reveal complex, and often emergent, patterns and dynamics. Nonetheless, a growing concern in network analysis is the use of limited data for constructing networks. This concern is strikingly relevant to ecology and conservation biology, where network analysis is used to infer connectivity across landscapes. In this context, movement among patches is the crucial parameter for interpreting connectivity but because of the difficulty of collecting reliable movement data, most network analysis proceeds with only indirect information on movement across landscapes rather than using observed movement to construct networks. Statistical models developed for social networks provide promising alternatives for landscape network construction because they can leverage limited movement information to predict linkages. Using two mark-recapture datasets on individual movement and connectivity across landscapes, we test whether commonly used network constructions for interpreting connectivity can predict actual linkages and network structure, and we contrast these approaches to social network models. We find that currently applied network constructions for assessing connectivity consistently, and substantially, overpredict actual connectivity, resulting in considerable overestimation of metapopulation lifetime. Furthermore, social network models provide accurate predictions of network structure, and can do so with remarkably limited data on movement. Social network models offer a flexible and powerful way for not only understanding the factors influencing connectivity but also for providing more reliable estimates of connectivity and metapopulation persistence in the face of limited data.

dispersal | graph theory | habitat fragmentation | latent space models | landscape ecology

Network analysis has recently exploded across scientific disciplines, including the social sciences, physics, cellular biology, and ecology (1–4). Topics as divergent as the stability of the Internet and the structure of metabolic reactions can be depicted through network analysis (1, 3). Such analysis is beneficial because it can facilitate the identification of complex, and often emergent, patterns, and can provide hypotheses for relationships between structure and function in many systems (2, 3). Nonetheless, a growing, widespread concern in the topic of network analysis is the reliability of data used in constructing networks (4–8).

In ecology and conservation, network analysis is increasingly being used to assess population connectivity across landscapes (9–13). Because of the importance of connectivity in conservation and its relevance to population and community ecology (14–16), network analysis and the accompanying use of graph theory are often emphasized as powerful approaches that have modest data requirements for assessing connectivity (10, 11, 13). In this spatial context, resource patches are considered nodes (or vertices) and movements and/or flows between patches are described as links (or edges/arcs), such that the network is a landscape representation of patches that potentially interact (17).

Although movement is a critical component of the dynamics of populations and communities across landscapes (15, 16), movement is a notoriously difficult parameter to estimate at a landscape scale (13, 18). As a consequence, spatial network analysis is predominantly applied based on assumptions or indirect information regarding movement rather than on estimates of movements among resource patches. Links are typically assumed to occur if individuals can potentially move between patches based on maximum known dispersal distances of species or, less commonly, through assumptions regarding dispersal kernels (11, 13, 19). Given the difficulty to empirically quantify connectivity via experiments or observed movements, network constructions that use limited information on movement could provide valuable, cost-effective approaches for assessing connectivity (13). Despite this potential, direct tests of whether network analysis can provide meaningful estimates of actual connectivity have yet to emerge (17).

Statistical models developed for social networks (20, 21) provide a promising alternative approach for predicting links and assessing connectivity, although such models have not been applied in a landscape or conservation setting. These models were originally developed to test for factors influencing social relationships among individuals, but their generality offers applicability beyond the social sciences. Not only can these models predict different types of linkages, they can address formal hypotheses regarding network structure (20, 21). Social network models may be particularly relevant to connectivity assessments because they can be applied in situations where only limited data are collected on a network (20), which could be useful given the difficulties of measuring movements at landscape scales.

We consider two types of social network models, a sender–receiver model and a latent space model, to predict landscape connectivity (*Materials and Methods*) (20, 21). Both models come from stochastic social network models termed "latent position, random graphs," which allow for different types of link responses (e.g., binary, Poisson) and can account for spatial dependencies in connections across networks (20–22). For connectivity assessments, a sender–receiver model allows for unobserved movement heterogeneity and directional movement among patches (i.e., movement from patch i to j is not equal to movement from patch j to i). In contrast, a latent space model estimates an unknown or "latent connectivity space" by leveraging the observed similarities in movement across patches within networks (20). Given the potential for highly directional movement across landscapes in many circumstances (23, 24) and the difficulty to estimate the complex nature of factors influencing connectivity

(18, 25), these models may prove highly useful for connectivity assessments.

We test whether network constructions derived from social network models and recent approaches in spatial ecology can quantify observed connectivity and predict actual linkages in real-world, landscape networks. We contrast these network constructions using two empirical, mark-recapture datasets that span several orders of spatial magnitude: within-field movements of a cactus-feeding insect (*Chelinidea vittiger*) on patchy *Opuntia* cactus and breeding-season movements of the endangered Everglades snail kite (*Rostrhamus sociabilis plumbeus*) across wetlands in peninsular Florida, USA. These examples provide a rare opportunity to quantify the extent to which network constructions can predict linkages and whether constructions can recover observed landscape connectivity. We further contrast the ability of these approaches to predict linkages and recover landscape structure based on limited data by randomly removing an increasing proportion of the observed movement data and requantifying network metrics and link prediction accuracy (*Materials and Methods*) (cf. 6, 8).

Results

Both of these empirical examples are notable for assessing the utility of network analysis for ecological landscapes because in both situations, all patches (or nodes) were sampled in each time step, the networks captured spatial scales relevant to movements of each species (26, 27), and movement data were collected at fine temporal resolutions within each network (*Materials and Methods*). Based on these mark-recapture data, we observed 70 movements of *C. vittiger* and 108 movements of *R. sociabilis plumbeus*. The frequency of movements declined exponentially with distance for both species (Fig. S1). *C. vittiger* moved locally across cactus patches (median distance moved = 4.5 m) (cf. 26), whereas *R. sociabilis plumbeus* moved widely across its geographic range within breeding seasons (median distance moved = 43 km) (cf. 27). The resulting portrayal of these observed landscape networks showed that observed movements tended to be geographically localized (Fig. 1A). Movements were also highly directional: The fraction of links between patches that were reciprocal (i.e., movement in both directions) was extremely low (0.09 for *C. vittiger* and 0.07 for *R. sociabilis plumbeus*). These observed networks provided a powerful means to assess the ability of different network constructions to predict observed linkages.

We first constructed networks based on the maximum distance of movements observed in the literature (26, 28), which is the most common approach for constructing landscape networks when movement data are not available (9, 10, 19). This approach consistently had the lowest accuracy in predicting linkages compared with other methods, and predicted no better than chance for *C. vittiger* (Fig. 2 and Fig. S2). This low predictive accuracy arose because it consistently had high false-positive error rates, where it overpredicted linkages (Fig. 1B and Fig. S2).

An alternative to a maximum distance construction is to take information on a dispersal kernel (i.e., a movement probability density function based on geographic distance) to construct a landscape network (11, 12). We contrasted two kernel-based constructions for spatial networks (*Materials and Methods*). First, we constructed networks assuming a negative exponential kernel, which is frequently used in metapopulation ecology (16) and has been suggested as a useful model for spatial networks in other disciplines (29, 30). We contrasted this theoretical kernel construction (Fig. 1C) that does not require movement data to using the empirical movement kernel (Fig. 1D) that we estimated with mark-recapture data to construct landscape networks (Fig. S1). Both of these constructions were moderate in predictive accuracy, with each kernel-based method performing similarly (Fig. 2 and Fig. S2). This comparison suggests that, in the absence of
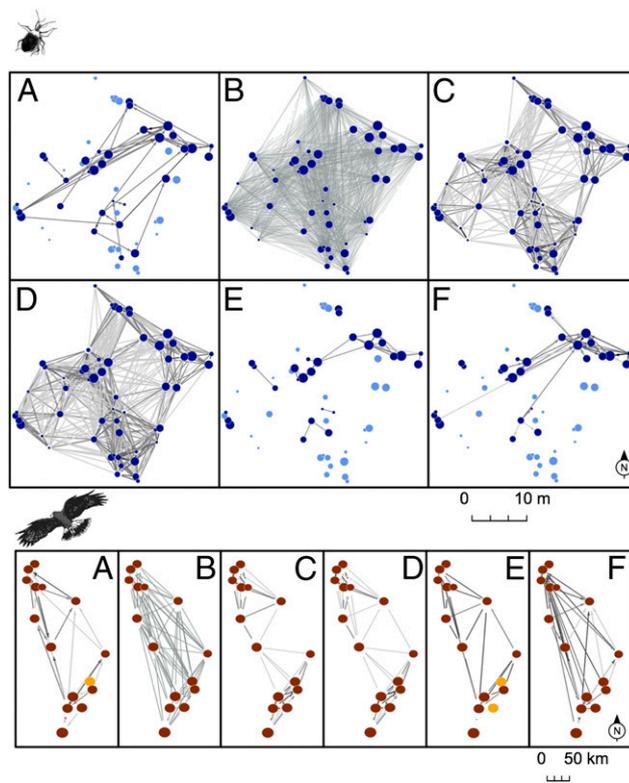
Fig. 1. The observed movement networks and network constructions for within-field movements of *C. vittiger* (*Upper*) on *Opuntia* cactus (n = 56 patches) and within-breeding season movements of *R. sociabilis plumbeus* (*Lower*) across wetlands within its geographic range in Florida (n = 15 wetlands). (A) The observed network, (B) maximum distance, (C) theoretical kernel, (D) empirical kernel, (E) latent space, and (F) sender–receiver constructions for each species. Node size proportional to the log(patch size) and link grayscale denotes increased relative intensity of observed and predicted movements. For kernel-based constructions, links greater than the average link prediction are shown; for latent space and sender–receiver models, link predictions greater than the threshold used for truncation are shown (based on maximum κ). Light-colored nodes highlight patches with no links to other patches.

movement data, using a theoretical kernel based on a negative exponential distribution can capture empirical kernel constructions and improve inference on connectivity over using the maximum known movement distance.

We then applied social network models to construct networks, which were generally best at predicting linkages (Fig. 1 *E* and *F*), with the sender–receiver model providing the highest predictive accuracy in both networks (Fig. 2 and Fig. S2). Interestingly, the sender–receiver models were generally more accurate at predicting linkages than other constructions, even when up to 50–80% of the data were missing when constructing the network (Fig. 2 and Fig. S2). The sender–receiver model was the only construction considered that could account for the observed directionality in movement, although it still predicted more reciprocity in movement than what was observed (the predicted fraction of reciprocal links between patches when no data were removed was 0.32 for *C. vittiger* and 0.18 for *R. sociabilis plumbeus*). In comparison with the sender–receiver model, the latent space model provided similar trends in predictive accuracy for *C. vittiger*, but for *R. sociabilis plumbeus* the latent space model suffered lower accuracy and it became the worst network construction when only very limited data were available (Fig. 2 and Fig. S2).
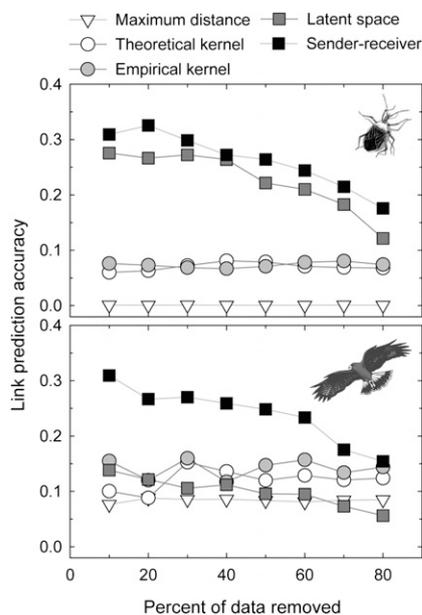
Fig. 2. The performance of network constructions, based on Cohen's κ, at predicting unknown linkages as a function of the amount of data used to construct the network. See Fig. S2 for other accuracy metrics.

We also determined whether these network constructions could recover the observed landscape structure by contrasting several network metrics (Table S1) based on the predicted network relative to the observed network (*Materials and Methods*). Sender–receiver models consistently recovered the observed landscape structure and could do so with remarkably little data on movement, whereas the maximum distance construction was consistently the worst approach for recovering observed landscape structure (Fig. 3). For some network metrics, such as connectance (i.e., link density), kernel-based methods recovered the observed structure. We further calculated the metapopulation lifetime (31) based on these network constructions and found that the maximum distance construction predicted much higher metapopulation lifetime for both species than the other network constructions (Fig. 3).

Although the utility of network constructions for assessing connectivity lies in their ability to predict linkages in landscapes (Fig. 2) and reconstruct known connectivity (Fig. 3), the network constructions we considered differ substantially in model complexity. Consequently, we also compared network constructions using model selection criteria (*SI Text*), which explicitly penalize for model complexity. Using this approach (Table S2), we found the same rank support for network constructions as we found when predicting links (Fig. 2), with the one exception being that the latent space model was the best fit for *C. vittiger*.

Discussion

Understanding and predicting connectivity are crucial for several questions in ecology and problems in conservation (14, 16). Network analysis and associated use of graph theory have been emphasized as offering a promising approach to connectivity analysis (10, 11, 13), yet we found that the way in which networks are constructed profoundly influences the resulting network and inferences on connectivity and metapopulation persistence. Although there have been several calls to revisit the foundations of network construction in other disciplines (5), this issue is extremely relevant to the use of network analysis for assessing connectivity in ecology.
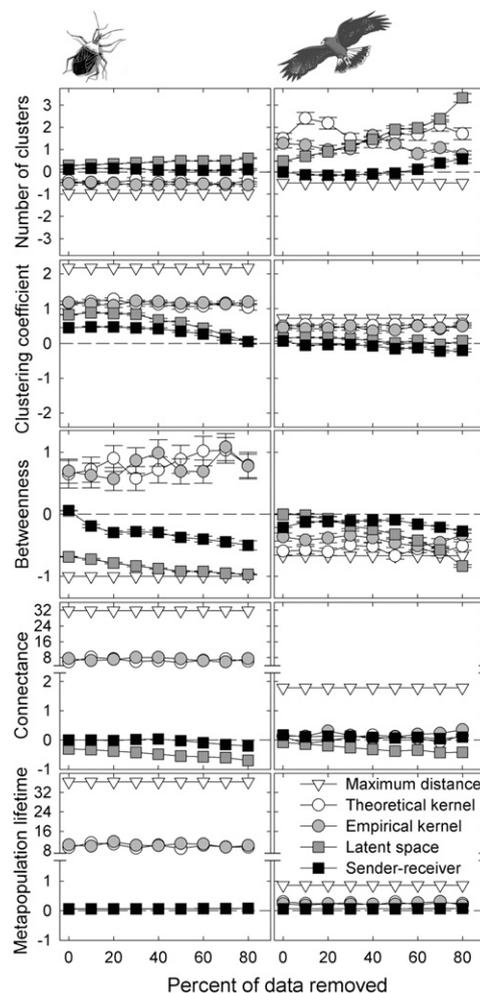
Fig. 3. Standardized network and metapopulation metrics based on different network constructions as a function of the amount of data used to construct the network. Shown for each metric is a measure of relative deviation of the predicted metric to the observed as $(x_{pred} - x_{obs})/x_{obs}$. The number of clusters provides a measure of the number of units within the network above the scale of patches. Clustering coefficients capture clustering of movement in networks. Betweenness describes the average number of shortest paths traveling through a patch, whereas connectance measures the observed number of links relative to the total possible number of links in the network. Metapopulation lifetime provides a measure of the predicted persistence of the network (Table S1).

The primary approach currently being used for problems in connectivity conservation—what we term the maximum distance construction—consistently overpredicted connectivity in our examples, suffered from the lowest accuracy in predicting linkages, and could not recover observed connectivity structure. This approach is frequently used because of the difficulty of collecting data on movement (e.g., 9, 10, 19). Although estimates from the maximum distance construction are often referred to as "potential connectivity" to acknowledge that actual connectivity is not measured (13), we found that this potential bears little resemblance to actual connectivity in real landscapes. This approach was worse in predicting linkages and recovering connectivity structure for *C. vittiger* than for *R. sociabilis plumbeus*, which was likely due to the differences in the prevalence of observed movements relative to the number of potential links in each landscape. This overprediction of connectivity is particularly relevant from a conservation perspective, because it translates to misleadingly high estimates of metapopulation lifetime in

landscape networks. Consequently, the use of maximum distance constructions for connectivity assessment may overestimate the viability of metapopulations.

Even in the absence of empirical data on movement, our results suggest that using a theoretical kernel-based construction can provide a more viable alternative to a maximum distance construction. Theoretical kernels, such as the negative exponential distribution used here (11, 16), provide a more realistic assessment of connectivity by providing less weight to infrequent, long-distance movements. Other kernel distributions could also be used to emphasize different types of movement (32).

We found that social network models more accurately predicted linkages and recovered network structure than other constructions. Although these models required empirical movement data, the models were still more parsimonious (in terms of model selection criteria) (Table S2) than other constructions, and the sender–receiver model was accurate even when remarkably limited data on movement were available to construct landscape networks. The utility of the sender–receiver model for both networks was likely due to its ability to capture the highly directional aspect of observed connectivity and unmeasured heterogeneity of movements to and from different patches, which is likely common in many spatial networks in ecology (23, 24). These models also provide a means to test hypotheses regarding connectivity, such as the role of patch size or matrix resistance in connectivity. For instance, by accounting for the potential effects of patch size on movement, the predictive accuracy of the sender–receiver models increased for both species (Table S3). Nonetheless, link prediction accuracy from social network models could still be improved considerably. Future applications of such models may improve predictions with the addition of other relevant factors (e.g., matrix resistance, path redundancy) that may influence movement among patches (e.g., 33).

Social network models may also be useful for other types of ecological network problems (4, 34), because they provide a general and flexible framework for modeling network data. For example, Chiu and Westveld (35) recently used such models to understand the role of phylogeny in food web structure. Such models may not only advance our understanding of the causes and consequences of ecological network structure but may also improve conservation and management strategies that rely on network inferences.

Materials and Methods

Focal Species and Movement Data. The cactus bug *C. vittiger* (Hemiptera: Coreidae) is dependent upon prickly pear cactus (*Opuntia* spp.), where it feeds, breeds, and aggregates throughout its life. We conducted a mark-recapture study in central Florida (29.4°N, 82.0°W), USA. In this area, *C. vittiger* uses *O. humifusa*, and has two or three generations per year. Adults are winged but rarely fly; instead, adult *C. vittiger* typically walk between cactus patches through an unsuitable matrix (26). Movements of adults are thus tractable and localized (Fig. S1) (26).

From September 2008 to December 2009, we censused all cactus patches (*n* = 56) in a 30 × 30-m network every 2–3 wk (except winter; 21 total visits), individually marking all adults using a nontoxic marker; these mark-recapture data allow estimation of movements among cactus patches. This network was not a closed system, but additional recapture rates surrounding this area were low. We mapped cactus patches (26) within the plot using a Trimble Global Positioning System (error <0.5 m).

The Everglades snail kite is dependent upon shallow freshwater ecosystems dominated by sparsely emergent vegetation, with its current range including suitable marsh and lake littoral habitats within a fragmented network of geographically distinct wetland units. Although often described as a panmictic population (36), short-term movement probabilities among wetlands are heterogeneous and within-breeding season movements appear more limited. We focused on within-breeding season movements of kites, because habitat suitability and prey availability are closely linked to local hydrology and can change dramatically within breeding seasons (37). Low water levels can cause kites to move in search of suitable habitat (37); these movements are associated with survival costs for younger, inexperienced individuals (27).

As a part of a long-term capture-mark-resight study, the entire breeding range of the snail kite in Florida was systematically searched via airboat during six intra-annual survey occasions, 2005–2009. Much of the population is banded with individually identifiable color-coded leg bands. During surveys, each detected individual was scanned to determine whether it was banded and the wetland location of each resighted individual was recorded, which allows for assessing movements among wetlands (for more details on field methods, see refs. 27 and 36). All techniques were approved by the Institutional Animal Care and Use Committee (approval F#149).

Maximum Distance Construction. Most applications of network analysis to connectivity conservation use a binary matrix (with elements p_{ij}) that describes pairwise connections between patches. Frequently, pairwise connections are assumed if Euclidean or other distance measures (e.g., least-cost distance) are within the known maximum dispersal distances of a species of interest (9, 10, 19), what we refer to as a maximum distance construction. For both datasets, we constructed binary networks using this formulation based on the maximum observed movement distance taken from the literature (26, 28), such that when patches *i* and *j* were within this maximum distance they were considered connected ($p_{ij} = 1$), whereas patches beyond this distance were considered unconnected ($p_{ij} = 0$). Using the maximum distances observed in these datasets (rather than those taken from the literature) resulted in similar conclusions for *C. vittiger* and less accurate predictions for *R. sociabilis plumbeus*.

Kernel-Based Constructions. We contrasted two kernel-based constructions for spatial networks. First, we constructed networks assuming a negative exponential kernel, which is frequently used in metapopulation ecology (16). We refer to this construction as a theoretical kernel (Eq. **1**),

$$p_{ij} = \exp(-\alpha d_{ij}),$$ **[1]**

where p_{ij} is the probability of a link between patches *i* and *j*, d_{ij} is the Euclidean distance between patches, and α is a scaling factor based on the average assumed movement distance (Fig. S1 shows movement kernels). Second, we contrasted theoretical kernels to the distribution of observed movement distances, what we term an empirical kernel construction. To allow for seamless comparisons of kernel constructions to the maximum distance construction, we drew 100 samples from kernels (with replacement) and for each sample we constructed a binary network based on that distance, where patches within the distance were considered connected and patches greater than that distance were considered unconnected.

Social Network Constructions. We modeled movement on spatial networks using social network models for three primary reasons. First, such models can predict linkages on relatively large, sparse networks compared with multistate models developed specifically for mark-recapture data (38). Second, the social network models we used allow for different types of movement data of the exponential family (e.g., binary, Poisson, etc.), which can be directed or undirected, and covariates regarding both nodes and links. Third, social network models have been developed to account for a wide variety of link dependencies that can occur within networks (22) and to leverage such dependencies to improve predictions. Link dependencies may be common on landscape networks, such as movements across stepping stones and movement variation in source-sink dynamics.

Hoff and colleagues (20, 22) have advanced a general latent position, random graph model for analyzing social networks that is highly relevant for network analysis in ecology and conservation. We considered two types of latent position, random graph models: *a latent space model* and a *sender–receiver model*. In the context of landscape connectivity, latent space models leverage observed similarities in movements within networks to predict links based on an unobservable or latent connectivity space ("social space" in social networks) (20). We modeled the probability of observed movement between patches as a Bernoulli distribution with a mean μ_{ij} and included distance, d_{ij}, between patches (pairwise distances) as the only fixed covariate in the model to provide direct comparisons with other constructions. We used a Euclidean distance measure based on similarities of between-patch movements to estimate latent connectivity space, such that our model formulation is $y_{ij} \sim \text{Bernoulli}(p_{ij})$ and (Eq. **2**)

$$\text{logit}\left(p_{ij}\right) = \alpha + \beta d_{ij} - |z_i - z_j|,$$ **[2]**

where y_{ij} is the observed presence or absence of a link (movement) between patches *i* and *j*, α is an intercept, β is the coefficient for distance, and (Eq. **3**)

$$|z_i - z_j| = \left(\sum_{k=1}^{K} (z_{ik} - z_{jk})^2 \right)^{1/2}, \qquad \text{[3]}$$

where K is the number of dimensions in the Euclidean latent space (in our examples, we set $K = 2$) (20, 21). As a consequence, when high rates of movement occur for patches i and j, Euclidean distances in the latent space will be small compared with patches with low movement rates. This formulation of latent space is inherently symmetrical (21). Hoff (21) shows that in a generalized linear mixed-model context, the latent space variable can be described as a random effect regarding the relational data, $z \sim N(0, I_k \times \sigma^2)$.

A sender–receiver model can be formulated in a similar way, where random effects are specified regarding within-patch dependencies and movement heterogeneity of source and destination patches (sender and receiver patches) (21) (Eq. 4):

$$\text{logit}\left(p_{ij} \right) = \alpha + \beta d_{ij} + \delta_i + \gamma_j, \qquad \text{[4]}$$

where δ and γ represent random effects of patches regarding emigration (δ_i) and immigration (γ_j) and $\delta \sim N(0, \sigma_\gamma^2)$ and $\gamma \sim N(0, \sigma_\gamma^2)$. The sender–receiver formulation allows for estimating directed movements in networks. The latent space and sender–receiver models could be integrated into a single model, but we treated these formulations separately for parsimony in model building. We also note that a logistic regression model that did not include these random effects (sender–receiver or latent space effects) was less accurate at predicting linkages than all other constructions for *R. sociabilis plumbeus* and was less accurate than both social network models for *C. vittiger*. These models were fit through Bayesian Markov chain Monte Carlo; see *SI Text* for a description of priors and assessment of model convergence.

Comparing Link Predictions. To assess the ability of different network constructions to predict linkages with limited data, we used cross-validation where we randomly removed 10–80% of the relational data from each dataset and constructed networks with the remaining data. For each removal amount, we used 100 randomly sampled replicates. We then asked whether constructions accurately predicted the withheld data. The network constructions we considered make different types of predictions for links, including binary predictions and proportion/probability predictions. To make accuracy assessment comparable among network constructions, we reduced constructions to binary predictions, which is the most common type of network used in spatial ecology (9, 10, 19). Maximum distance constructions are binary networks, whereas in kernel-based constructions, each of 100 random draws from the kernel produced a binary network; we report average estimates for predicting network linkages and structure based on these samples. Because the latent space and random-effects models provide predictions of link probability, we truncated these predictions to [0, 1] data, based on the threshold that maximized Cohen's κ (39) (*SI Text*).

We then used Cohen's κ (κ hereafter), the true skill statistic (TSS), and the area under the receiver-operating curve (AUC) to assess predictive accuracy of network constructions (40). κ And TSS each range from −1 to 1, with values of zero indicating performance no better than random (40). AUC is a threshold-independent metric that ranges from 0 to 1, with values of 0.5 indicating performance no better than random. We also calculated false-positive and false-negative error rates for each construction to better interpret the sources of errors when predicting network linkages (Fig. S2).

Comparing the Recovery of Landscape Structure and Metapopulation Measures. We assessed the potential for various network constructions to recover observed landscape structure and predict indices of metapopulation persistence. We calculated several metrics that describe connectivity and spatial structuring at different scales: number of clusters (components), clustering coefficient, average degree, average betweenness, average shortest path, and connectance (see Table S1 for definitions and examples of these metrics). We also calculated two measures that reflect metapopulation viability: metapopulation mean lifetime (31) and metapopulation capacity (41) (*SI Text*). We do not report average degree, average shortest path, or metapopulation capacity, because these metrics were redundant with other metrics. The network metrics capture spatial structuring (clustering coefficient, number of clusters) and connectivity at different scales (average betweenness, connectance). Clustering coefficients capture clustering of movement within networks (29), whereas the number of clusters provides a measure of the number of units within the network above the scale of patches (17). We chose betweenness as a measure of local-scale connectivity, and connectance as a network-scale measure of connectivity. For each metric, x, we calculate a measure of relative deviation of the predicted value to the observed value as $(x_{pred} - x_{obs})/x_{obs}$ (6). Consequently, network constructions with deviations close to 0 capture observed network structure.

1. Guimerà R, Nunes Amaral LA (2005) Functional cartography of complex metabolic networks. *Nature* 433:895–900.
2. Borgatti SP, Mehra A, Brass DJ, Labianca G (2009) Network analysis in the social sciences. *Science* 323:892–895.
3. Strogatz SH (2001) Exploring complex networks. *Nature* 410:268–276.
4. Olesen JM, et al. (2011) Missing and forbidden links in mutualistic networks. *Proc Biol Sci* 278:725–732.
5. Butts CT (2009) Revisiting the foundations of network analysis. *Science* 325:414–416.
6. Guimerà R, Sales-Pardo M (2009) Missing and spurious interactions and the reconstruction of complex networks. *Proc Natl Acad Sci USA* 106:22073–22078.
7. Handcock MS, Gile KJ (2010) Modeling social networks from sampled data. *Ann Appl Stat* 4(1):5–25.
8. Clauset A, Moore C, Newman MEJ (2008) Hierarchical structure and the prediction of missing links in networks. *Nature* 453(7191):98–101.
9. Lookingbill TR, Gardner RH, Ferrari JR, Keller CE (2010) Combining a dispersal model with network theory to assess habitat connectivity. *Ecol Appl* 20:427–441.
10. Minor ES, Urban DL (2008) A graph-theory framework for evaluating landscape connectivity and conservation planning. *Conserv Biol* 22:297–307.
11. Urban D, Keitt T (2001) Landscape connectivity: A graph-theoretic perspective. *Ecology* 82:1205–1218.
12. Bodin O, Tengö M, Norman A, Lundberg J, Elmqvist T (2006) The value of small size: Loss of forest patches and ecological thresholds in southern Madagascar. *Ecol Appl* 16:440–451.
13. Calabrese JM, Fagan WF (2004) A comparison-shopper's guide to connectivity metrics. *Front Ecol Environ* 2:529–536.
14. Crooks KR, Sanjayan M, eds (2006) *Connectivity Conservation* (Cambridge Univ Press, New York).
15. Leibold MA, et al. (2004) The metacommunity concept: A framework for multi-scale community ecology. *Ecol Lett* 7:601–613.
16. Hanski I (1998) Metapopulation dynamics. *Nature* 396:41–49.
17. Urban DL, Minor ES, Treml EA, Schick RS (2009) Graph models of habitat mosaics. *Ecol Lett* 12:260–273.
18. Bélisle M (2005) Measuring landscape connectivity: The challenge of behavioral landscape ecology. *Ecology* 86:1988–1995.
19. Bunn AG, Urban DL, Keitt TH (2000) Landscape connectivity: A conservation application of graph theory. *J Environ Manage* 59:265–278.
20. Hoff PD, Raftery AE, Handcock MS (2002) Latent space approaches to social network analysis. *J Am Stat Assoc* 97:1090–1098.
21. Hoff PD (2005) Bilinear mixed-effects models for dyadic data. *J Am Stat Assoc* 100:286–295.
22. Krivitsky PN, Handcock MS, Raftery AE, Hoff PD (2009) Representing degree distributions, clustering, and homophily in social networks with latent cluster random effects models. *Soc Networks* 31:204–213.
23. Campbell Grant EH, Nichols JD, Lowe WH, Fagan WF (2010) Use of multiple dispersal pathways facilitates amphibian persistence in stream networks. *Proc Natl Acad Sci USA* 107:6936–6940.
24. Treml EA, Halpin PN, Urban DL, Pratson LF (2008) Modeling population connectivity by ocean currents, a graph-theoretic approach for marine conservation. *Landscape Ecol* 23:19–36.
25. Winfree R, et al. (2005) Testing simple indices of habitat proximity. *Am Nat* 165:707–717.
26. Schooley RL, Wiens JA (2004) Movements of cactus bugs: Patch transfers, matrix resistance, and edge permeability. *Landscape Ecol* 19:801–810.
27. Martin J, Nichols JD, Kitchens WM, Hines JE (2006) Multiscale patterns of movement in fragmented landscapes and consequences on demography of the snail kite in Florida. *J Anim Ecol* 75:527–539.
28. Sykes PW, Jr., Rodgers JA, Jr., Bennetts RE (1995) *The Birds of North America Online*, ed Poole A (Cornell Lab Ornithol, Ithaca, NY).
29. Costa LD, Rodrigues FA, Travieso G, Boas PRV (2007) Characterization of complex networks: A survey of measurements. *Adv Phys* 56:167–242.
30. Wong LH, Pattison P, Robins G (2006) A spatial model for social networks. *Physica A* 360:99–120.
31. Kininmonth S, Drechsler M, Johst K, Possingham HP (2010) Metapopulation mean life time within complex networks. *Mar Ecol Prog Ser* 417:139–149.

32. Clark JS, Silman M, Kern R, Macklin E, HilleRisLambers J (1999) Seed dispersal near and far: Patterns across temperate and tropical forests. *Ecology* 80:1475–1494.
33. McRae BH, Beier P (2007) Circuit theory predicts gene flow in plant and animal populations. *Proc Natl Acad Sci USA* 104:19885–19890.
34. McDonald DB (2007) Predicting fate from early connectivity in a social network. *Proc Natl Acad Sci USA* 104:10910–10914.
35. Chiu GS, Westveld AH (2011) A unifying approach for food webs, phylogeny, social networks, and statistics. *Proc Natl Acad Sci USA* 108:15881–15886.
36. Dreitz VJ, et al. (2002) The use of resighting data to estimate the rate of population growth of the snail kite in Florida. *J Appl Stat* 29:609–623.
37. Takekawa JE, Beissinger SR (1989) Cyclic drought, dispersal, and the conservation of the snail kite in Florida: Lessons in critical habitat. *Conserv Biol* 3:302–311.
38. Hestbeck JB, Nichols JD, Malecki RA (1991) Estimates of movement and site fidelity using mark resight data of wintering Canada geese. *Ecology* 72:523–533.
39. Liu CR, Berry PM, Dawson TP, Pearson RG (2005) Selecting thresholds of occurrence in the prediction of species distributions. *Ecography* 28:385–393.
40. Allouche O, Tsoar A, Kadmon R (2006) Assessing the accuracy of species distribution models: Prevalence, kappa and the true skill statistic (TSS). *J Appl Ecol* 43:1223–1232.
41. Hanski I, Ovaskainen O (2000) The metapopulation capacity of a fragmented landscape. *Nature* 404:755–758.

ECOLOGY

Supporting Information

Fletcher et al. 10.1073/pnas.1107549108

SI Text

This supporting information includes: (*i*) a model selection approach for comparing among network constructions; (*ii*) descriptions of priors in hierarchical formulations of social network models and assessments of model convergence; (*iii*) accuracy assessments based on the true skill statistic, area under the receiver-operating curve (ROC), and false-positive and false-negative error rates; (*iv*) description of metapopulation metrics used for assessing population-level effects of predicted connectivity; (*v*) descriptions of network connectivity metrics; and (*vi*) empirical and theoretical movement kernels for the two networks.

1. Model Selection Comparison.

The network constructions we considered differ dramatically in model complexity. Whereas the cross-validation approach shown in Fig. 2 *implicitly* penalizes for model complexity (i.e., an overly complex model is unlikely to generalize to new situations), this approach did not explicitly penalize for model complexity. Consequently, we also contrasted constructions based on model selection criteria that directly penalize for model complexity rather than assess models based on cross-validation alone.

We compared models by using maximum likelihood methods to fit each network construction: maximum distance, theoretical kernel, latent space, and sender–receiver model. Note that we do not formally compare the empirical kernel shown in Figs. 2 and 3 to the other constructions, because it is a nonparametric formulation based on sampling the observed movement distances. We use likelihood-based estimation for model selection comparisons rather than Bayesian methods, because the use of model selection criteria is more debated with Bayesian estimation than with likelihood-based estimation (refs. 1–4 and accompanying discussion). Furthermore, we compared social network models by parameterizing random effects as fixed effects. We used this approach because model selection with random effects, particularly when interest is on conditional prediction (using random effects estimates to improve subject-specific predictions, as was the case here), is not resolved regarding estimation of the appropriate number of effective parameters for penalizing likelihoods (i.e., the number of effective parameters lies between 1 and R, where R is the number of random effects) (5, 6). Reformulating the likelihoods with fixed effects avoids the difficulty of model selection with random effects based on conditional inference, and it is conservative for model selection in the sense that social network models will be penalized more by using fixed effects than with selection criteria using random effects (5, 6). We used Akaike's information criterion (AIC) to compare among network constructions (7), but note that the Bayesian information criterion provided similar results.

Likelihood of the maximum distance construction. The likelihood functions for each of these network constructions can be derived from the likelihood for a Bernoulli distribution, with the differences among constructions being how we constrain the probability of a link, p. The maximum distance construction can be described as a Bernoulli model, where an observed link between patches i and j is a success ($x_{ij} = 1$) and there are $n \times (n - 1)$ possible links observed, where n is the number of patches. The maximum distance can be considered a fixed, categorical covariate. In this case, the likelihood can be described as (Eq. **S1**)

$$L\left(\theta | X, Y\right) = \prod_{ij=1}^{n \times (n-1)} p_{ij}^{x_{ij}} \left(1 - p_{ij}\right)^{1 - x_{ij}} \quad \text{[S1]}$$

and (Eq. **S2**)

$$\text{logit}\left(p_{ij}\right) = \beta_1 d_{\max}, \quad \text{[S2]}$$

where $d_{\max} = 1$ when the distance between patches i and j is less than the maximum known movement distance and $d_{\max} = 0$ when the distance between patches is greater than the maximum known movement distance, Y is the connectivity matrix of the observed data regarding the presence or absence of links, X is the covariate(s) considered ($d_{\max}$ in this case), and θ are the parameters to be estimated (β_1 for this case).

Based on the maximum movement distances reported in the literature (8, 9), we defined $d_{\max}$ as 52 m for *Chelinidea vittiger* and 225 km for *Rostrhamus sociabilis plumbeus*. Because the conventional approach in the literature is to assume $p_{ij} = 1$ when $d_{ij} < d_{\max}$ and $p_{ij} = 0$ with $d_{ij} > d_{\max}$, we fixed β_1 in Eq. **S2** to produce such predictions ($\beta_1 = 7$, such that $p_{ij} = 0.99$ when $d_{ij} < d_{\max}$). Note that using this literature-driven approach is not likely the maximum likelihood estimate (MLE) for β_1 in this model; we also contrast this model to a model based on the MLE of β_1 (Table S2).

Likelihood of the theoretical kernel construction. The theoretical kernel construction can also be described as a Bernoulli model. In this case, the likelihood function is the same as Eq. **S1**, but we alter Eq. **S2** to include the nonlinear function shown in Eq. **1** of the text. Because Eq. **1** bounds p_{ij} on the 0–1 interval, we did not use a logit link function. Again, we contrast the likelihood of the model when using α taken from the literature. As with the maximum distance construction, using the literature-driven approach for estimating α is not likely the MLE for this model; we also contrast a model using the estimate of the maximum likelihood of α.

Likelihood of the latent space model. In the latent space model, the likelihood function is the same as Eq. **S1**, but we constrain the probability of a link based on Eqs. **2** and **3** of the text. For direct, likelihood-based comparisons, we fit this model using the maximum likelihood procedure described in ref. 10, where the latent space is considered a fixed effect rather than a random effect. Initial values for the latent space parameters were taken using multidimensional scaling on the Euclidean similarity matrix (10).

Likelihood of the sender–receiver construction. In the sender–receiver model, the likelihood function is the same as Eq. **S1**, but we constrain the probability of a link based on Eq. **4** of the text. The sender–receiver random effects can be recast in the context of fixed, patch-specific effects of emigration and immigration to estimate likelihoods and model selection criteria comparable with other network constructions.

Model selection results. We found that for both species, variation in AIC among network constructions (Table S2) showed similar patterns to that of their predictive accuracy (shown in Fig. 2), and that there was strong support for social network models compared with other network constructions. For *C. vittiger*, the primary difference was that the latent space construction fit the data better than the sender–receiver model, whereas with assessments based on cross-validation, the sender–receiver construction was slightly better at predicting links. For *R. sociabilis plumbeus*, the rank order of network constructions based on AIC was the same as for cross-validation. We note that for both species, estimates taken from the literature for both the maximum distance and theoretical kernel constructions had much worse fit to the data

than the maximum likelihood estimates for those constructions (Table S2). See Table S3 for further comparison of social network models that included patch-size effects.

2. Priors Used in Hierarchical Formulation of Social Network Models and Interpreting Model Convergence. For network predictions shown in Figs. 2 and 3, we used a Bayesian formulation of social network models and Markov chain Monte Carlo (MCMC) methods (4, 10, 11). In a Bayesian context, the latent space model can be described as a hierarchical model, as shown in Eqs. **2** and **3** of the text. Using the methods in refs. 11 and 12, priors for the fixed effects were $N \sim (0, 9)$. The latent space distance, $Z = |z_i - z_j|$, is specified as a multivariate normal distribution, $Z \sim MVN_k(\mu, \sigma_z^2 I_k)$, where K is the number of dimensions in the latent space and I_k is a $K \times K$ identity matrix. Hyperpriors for this latent space were $\mu \sim MVN_k(0, \omega^2 I_k)$ and $\sigma_z^2 \sim \sigma_{0,z}^2 Inv\chi_{\alpha,z}^2$, where α is the degrees of freedom for the inverse χ^2 distribution and σ_0^2 is a scaling factor (11, 12). The scaled inverse χ^2 distribution can be shown as being a special case of the inverse gamma distribution (13). Following refs. 11 and 12, hyperparameters were initially set to $\sigma_{0,z}^2 = 1/8(n)^{K/2}$, $\alpha_z = n^{1/2}$, and $\omega^2 = 1/4(n)^{K/2}$, where n is the number of patches (nodes).

In a Bayesian context, the sender–receiver model can also be described as a hierarchical model, as shown in Eq. 4 of the text. Using the information in refs. 11 and 12, priors for the fixed effects were $N \sim (0, 9)$ and the hyperpriors were specified as $\sigma_\delta^2 \sim \alpha_\delta \sigma_{0,\delta}^2 Inv\chi_{\alpha,\delta}^2$ and $\sigma_\gamma^2 \sim \alpha_\gamma \sigma_{0,\gamma}^2 Inv\chi_{\alpha,\gamma}^2$. Using suggestions in refs. 11 and 12, hyperparameters were initially set to $\alpha_\delta = 3$ and $\sigma_{0,\gamma}^2 = 1$.

We used the Raftery–Lewis diagnostic (12, 13) to tune model runs, resulting in 10,000 samples for burn-in and 50,000 iterations, saving every 10th sample for estimating the posterior distributions. We assessed model convergence using the Gelman–Rubin diagnostic (14), calculated on the basis of two MCMC chains. This diagnostic compares posterior distributions between MCMC chains for each parameter in the model; at convergence, this diagnostic is ~1. A common rule is that convergence is likely when the diagnostic is <1.1 (15). For both network constructions, we assessed convergence as a function of the amount of data removed (0–80%).

For all parameters and amounts of data removed, Gelman–Rubin diagnostics were 1.0–1.012 for *C. vittiger*, suggesting model convergence. Similarly, Gelman–Rubin diagnostics ranged between 1.0 and 1.06 for *R. sociabilis plumbeus* across all parameters and amounts of data removed, suggesting model convergence.

3. Accuracy Assessments of Network Constructions. A challenge regarding the comparison of different network constructions for predicting links is how to seamlessly compare constructions that make predictions on different scales. In the network constructions we considered, the maximum distance construction made binary predictions, the kernels also made binary predictions for each sample from the distribution (which could be summed to provide proportions or could be assessed at the sample level), and social network constructions predicted probabilities ranging between 0 and 1. Our validation data were binary data regarding the presence or absence of an observed link. Whereas the maximum distance construction and kernel constructions could be assessed directly against the binary validation data, the social network constructions could not.

Three common solutions are often used to compare model probabilities to binary validation data: (*i*) use a threshold-independent measure for comparing observed data versus model predictions; (*ii*) use some sort of pooling, such as "calibration plots," where one compares the frequency of observed links to the probabilities that are predicted; and (*iii*) use a threshold to dichotomize predictions. The most common threshold-independent measure for assessing predictive accuracy is the area under the receiver-operating curve (AUC), and this measure has been used recently in comparing link predictions in other contexts (16, 17).

However, the AUC has less value for models making binary predictions (18), such that it is less appropriate to compare AUCs for the social network models with AUCs for the other network constructions. Assessments using calibration plots can be qualitatively useful, but require pooling and are still limited for comparing among network constructions and summary statistics regarding network structure (Fig. 3). Using a threshold to dichotomize predictions was the most natural solution to allow assessments at the individual link level, because the use of thresholds in the validation of probabilistic models is common across disciplines (19) and provides a means to seamlessly compare among all of the network constructions we considered by truncating probabilities to binary predictions, consistent with other network constructions. However, we also provide an AUC for comparisons among other recent attempts to predict links in networks (16, 17).

Decisions for selecting thresholds have received a considerable amount of interest, particularly in species distribution modeling, which has similar challenges regarding model validation as the network constructions we considered (20, 21). Freeman and Moisen (21) found that choosing a threshold that maximizes Cohen's κ worked well, particularly in situations where a goal is to preserve observed prevalence, or the observed proportion of links observed (i.e., connectance in binary networks), in predictions (22). κ measures the proportion of correctly classified links after accounting for the probability of chance agreement (19). Consequently, we selected thresholds that maximized κ for each species to dichotomize predictions of social network constructions.

With these thresholds, we then used four metrics for predictive accuracy assessment: Cohen's κ, the true skill statistic, false-positive error rates, and false-negative error rates. κ has a long history of use in assessing accuracy of maps (23). The true skill statistic (TSS) is frequently used in the medical literature (18). TSS has a similar interpretation as κ, but is thought to be less sensitive to prevalence. Both κ and TSS assess overall predictive accuracy; we also considered false-positive and false-negative error rates to better interpret the sources of prediction error.

4. Metapopulation Measures. We used two relevant metapopulation measures aimed at predicting metapopulation viability: the metapopulation mean lifetime (24) and the metapopulation capacity (25). Both of these measures focus on extinction–colonization dynamics in metapopulations and can be applied to the predicted network constructions to make inferences regarding the differences among network constructions in the predicted viability of metapopulations. *Metapopulation mean lifetime.* The metapopulation mean lifetime is derived from a stochastic metapopulation model that aims to approximate the viability of metapopulations. This model was initially developed by Frank and Wissel (24) and later extended to metapopulations on complex networks (26, 27).

This model starts by assuming that the extinction rate v of a local patch i can be described as (Eq. **S3**)

$$v_i = \varepsilon A_i^{-\eta}, \qquad \qquad \textbf{[S3]}$$

where A_i is the area of patch i, ε is a parameter relating to minimum patch size, and η is a scaling parameter that describes the relative amount of environmental variation on population growth, where a smaller value indicates more environmental variability and thus relatively larger local extinction rates. Aggregate local extinction rates of patches for the metapopulation can be described as the geometric mean of location extinction rates (Eq. **S4**):

$$v = \prod_{i=1}^{n} v_i^{1/n}. \qquad \qquad \textbf{[S4]}$$

Kininmonth et al. (27) described colonization ability of patch i from other patches using a "strength" network measure, or

a valued measure regarding the sum of link values for each patch (analogous to degree) (Table S1), w_{ij}, on a network as (Eq. **S5**)

$$u_i^{in} = \frac{1}{\mu}\left(\sum_{j=1}^{n} \frac{w_{ij}}{v_j}\right), \qquad \textbf{[S5]}$$

where μ is the number of immigrants required for successful colonization (in our application, we simply set $\mu = 1$). In the context of the network constructions assessed here, the strength is the predicted value, p_{ij}, of the link based on the network construction. Note that this weighs the flow/movement from j to i by the local extinction risk of j. Similarly, the colonization strength of patch i on other patches can be described as (Eq. **S6**)

$$u_i^{out} = \frac{1}{\mu v_i}\left(\sum_{j=1}^{n} w_{ij}\right). \qquad \textbf{[S6]}$$

The harmonic mean of u_i^{out} and u_i^{in} for each patch is then (Eq. **S7**)

$$U_i = \left(\frac{1}{2}\left(u_i^{in}\right)^{-2} + \frac{1}{2}\left(u_i^{out}\right)^{-2}\right)^{-1/2}. \qquad \textbf{[S7]}$$

The aggregated colonization:extinction ratio is calculated as the geometric mean of U_i for each patch (26, 27) (Eq. **S8**):

$$q = \prod_{i=1}^{n} U_i^{1/n}. \qquad \textbf{[S8]}$$

With these measures, the metapopulation mean lifetime (MLT) is approximated as (24, 26, 27) (Eq. **S9**)

$$MLT = \frac{1}{v}\sum_{i=1}^{n}\sum_{k=i}^{n}\frac{1}{k}\left(\frac{(N-i)!}{(N-k)!}\right)\frac{1}{(N-1)^{k-i}}q^{k-i}. \qquad \textbf{[S9]}$$

We calculated $\log_{10}(MLT)$ for each of the network constructions and compared it to the $\log_{10}(MLT)$ using the observed connectivity matrix. After setting $\mu = 1$, there were two additional parameters in the model: ε and η. We considered $1 < \varepsilon < 10$ and $0.05 < \eta < 1$ (cf. 27). In all comparisons, the relative magnitude of differences among network constructions was similar. In Fig. 3, we report results where $\varepsilon = 5$ and $\eta = 0.5$.

Metapopulation capacity. As an alternative to metapopulation mean lifetime, we also considered what has been termed the "metapopulation capacity" by Hanski and Ovaskainen (25). This metric has a similar goal as the metapopulation mean lifetime, but it is a relative metric that assumes deterministic extinction. Nonetheless, it has proven useful and does not have additional free parameters, like the metapopulation mean lifetime described above.

The metapopulation capacity was defined in ref. 25 as the leading real eigenvalue of an area-weighted connectivity matrix M consisting of the following elements (Eq. **S10**):

$$m_{ij} = \exp(-\alpha d_{ij})A_i A_j \qquad \textbf{[S10]}$$

and $m_{ii} = 0$. We calculated the metapopulation capacity using the predicted connectivity matrix (with elements p_{ij}) for each network construction instead of $\exp(-\alpha d_{ij})$, multiplied by $A_i A_j$. Metapopulation capacity was highly correlated with metapopulation mean lifetime ($r > 0.93$), so we only show results of metapopulation mean lifetime.

1. Raftery AE (1995) Bayesian model selection in social research. *Sociol Methodol* 25: 111–163.
2. Gelman A, Rubin DB (1995) Avoiding model selection in Bayesian social research. *Sociol Methodol* 25:165–173.
3. Spiegelhalter DJ, Best NG, Carlin BR, van der Linde A (2002) Bayesian measures of model complexity and fit. *J R Stat Soc Ser B Stat Methodol* 64:583–616.
4. Handcock MS, Raftery AE, Tantrum JM (2007) Model-based clustering for social networks. *J R Stat Soc Ser A Stat Soc* 170:301–354.
5. Vaida F, Blanchard S (2005) Conditional Akaike information for mixed-effects models. *Biometrika* 92:351–370.
6. Greven S, Kneib T (2010) On the behaviour of marginal and conditional AIC in linear mixed models. *Biometrika* 97:773–789.
7. Burnham KP, Anderson DR (1998) *Model Selection and Inference: A Practical Information-Theoretic Approach* (Springer, New York).
8. Sykes PW, Jr., Rodgers JA Jr., Bennetts RE (1995) *The Birds of North America Online*, ed Poole A (Cornell Lab Ornithol, Ithaca, NY).
9. Schooley RL, Wiens JA (2004) Movements of cactus bugs: Patch transfers, matrix resistance, and edge permeability. *Landscape Ecol* 19:801–810.
10. Hoff PD, Raftery AE, Handcock MS (2002) Latent space approaches to social network analysis. *J Am Stat Assoc* 97:1090–1098.
11. Krivitsky PN, Handcock MS, Raftery AE, Hoff PD (2009) Representing degree distributions, clustering, and homophily in social networks with latent cluster random effects models. *Soc Networks* 31:204–213.
12. Krivitsky PN, Handcock MS (2008) Fitting position latent cluster models for social networks with latentnet. *J Stat Softw* 24(5):1–23.
13. Gelman A, Carlin JB, Stern HS, Rubin DB (2004) *Bayesian Data Analysis* (CRC, London).
14. Gelman A, Rubin DB (1992) Inference from iterative simulation using multiple sequences. *Stat Sci* 7:457–511.
15. Gelman A, Hill J (2007) *Data Analysis Using Regression and Multilevel/Hierarchical Models* (Cambridge Univ Press, New York).
16. Clauset A, Moore C, Newman MEJ (2008) Hierarchical structure and the prediction of missing links in networks. *Nature* 453(7191):98–101.
17. Guimerà R, Sales-Pardo M (2009) Missing and spurious interactions and the reconstruction of complex networks. *Proc Natl Acad Sci USA* 106:22073–22078.
18. Allouche O, Tsoar A, Kadmon R (2006) Assessing the accuracy of species distribution models: Prevalence, kappa and the true skill statistic (TSS). *J Appl Ecol* 43:1223–1232.
19. Fielding AH, Bell JF (1997) A review of methods for the assessment of prediction errors in conservation presence/absence models. *Environ Conserv* 24:38–49.
20. Liu CR, Berry PM, Dawson TP, Pearson RG (2005) Selecting thresholds of occurrence in the prediction of species distributions. *Ecography* 28:385–393.
21. Freeman EA, Moisen GG (2008) A comparison of the performance of threshold criteria for binary classification in terms of predicted prevalence and kappa. *Ecol Modell* 217 (1-2):48–58.
22. Manel S, Williams HC, Ormerod SJ (2001) Evaluating presence-absence models in ecology: The need to account for prevalence. *J Appl Ecol* 38:921–931.
23. Congalton RG (1991) A review of assessing the accuracy of classification of remotely sensed data. *Remote Sens Environ* 37(1):35–46.
24. Frank K, Wissel C (2002) A formula for the mean lifetime of metapopulations in heterogeneous landscapes. *Am Nat* 159:530–552.
25. Hanski I, Ovaskainen O (2000) The metapopulation capacity of a fragmented landscape. *Nature* 404:755–758.
26. Drechsler M (2009) Predicting metapopulation lifetime from macroscopic network properties. *Math Biosci* 218(1):59–71.
27. Kininmonth S, Drechsler M, Johst K, Possingham HP (2010) Metapopulation mean life time within complex networks. *Mar Ecol Prog Ser* 417:139–149.
28. Bunn AG, Urban DL, Keitt TH (2000) Landscape connectivity: A conservation application of graph theory. *J Environ Manage* 59:265–278.
29. Urban D, Keitt T (2001) Landscape connectivity: A graph-theoretic perspective. *Ecology* 82:1205–1218.
30. Jordan F, et al. (2003) Characterizing the importance of habitat patches and corridors in maintaining the landscape connectivity of a *Pholidoptera transsylvanica* (Orthoptera) metapopulation. *Landscape Ecol* 18(1):83–92.
31. Fortuna MA, Gómez-Rodríguez C, Bascompte J (2006) Spatial network structure and amphibian persistence in stochastic environments. *Proc Biol Sci* 273:1429–1434.
32. Minor ES, Urban DL (2008) A graph-theory framework for evaluating landscape connectivity and conservation planning. *Conserv Biol* 22:297–307.
33. Cantwell MD, Forman RTT (1993) Landscape graphs: Ecological modeling with graph theory to detect configurations common to diverse landscapes. *Landscape Ecol* 8: 239–255.
34. Estrada E, Bodin O (2008) Using network centrality measures to manage landscape connectivity. *Ecol Appl* 18:1810–1825.
35. Bodin O, Norberg J (2007) A network approach for analyzing spatially structured populations in fragmented landscape. *Landscape Ecol* 22(1):31–44.
36. Ferrari JR, Lookingbill TR, Neel MC (2007) Two measures of landscape-graph connectivity: Assessment across gradients in area and configuration. *Landscape Ecol* 22:1315–1323.

Fig. S1. Observed movement distributions (gray bars) and theoretical kernels taken from the literature for *C. vittiger* (*Upper*) and *R. sociabilis* plumbeus (*Lower*). Theoretical kernels based on a negative exponential function, where the probability of a linkage, p_{ij}, between two patches is $\exp(-\alpha d_{ij})$. We calculated α as the inverse of the average movement distance observed in the literature (*Materials and Methods*).

Fig. S2. Alternative accuracy assessment metrics for predicting linkages in each network construction as a function of the amount of data used to construct the network for *C. vittiger* (*Left*) and *R. sociabilis* plumbeus (*Right*). Shown are the true skill statistic and the AUC as measures of overall prediction accuracy and false-positive and false-negative error rates.

Table S1. Descriptions of network connectivity metrics and metapopulation measures considered and examples of their use in spatial ecology

Metric	Description (examples in spatial ecology)
Number of clusters	Simply the number of clusters or components in a network. A cluster or component is a group of interconnected nodes for which a path exists between every pair of nodes. Paths between nodes can be either direct or indirect connections (1–3).
Clustering coefficient	Characterizes transitivity in the network by measuring the number of triangles, or the situation where if patch i is connected to j and j is connected to k, then i and k are connected (3–5).
Degree*	The number of direct (immediate) links to a node (4, 6, 7).
Betweenness	Number of geodesics going through a focal node. The geodesic (or shortest path between patches i and j) can be calculated as either the path that traverses the shortest number of links between the two nodes or the path that minimizes the sum of the weights of the links in the path (7, 8).
Shortest path distance*	The length of the geodesic, measured as either the number of links traversed or the total distance traversed (5, 9).
Connectance	A measure of link density, or the number of observed links relative to the total possible number of links in a network (3).
Metapopulation mean lifetime	An approximation formula of metapopulation viability in heterogeneous landscapes, assuming stochastic extinction–colonization dynamics (10–12).
Metapopulation capacity	The leading, real eigenvalue of an area-weighted connectivity matrix. Measures the relative safety of a metapopulation from deterministic extinction (13).

*Degree and shortest path distance are not reported, because they were redundant ($r > 0.97$) with connectance and the number of clusters, respectively. In addition, metapopulation capacity is not reported, because it was redundant with metapopulation mean lifetime ($r > 0.93$).

1. Bunn AG, Urban DL, Keitt TH (2000) Landscape connectivity: A conservation application of graph theory. *J Environ Manage* 59:265–278.
2. Urban D, Keitt T (2001) Landscape connectivity: A graph-theoretic perspective. *Ecology* 82:1205–1218.
3. Jordan F, et al. (2003) Characterizing the importance of habitat patches and corridors in maintaining the landscape connectivity of a *Pholidoptera transsylvanica* (Orthoptera) metapopulation. *Landscape Ecol* 18(1):83–92.
4. Fortuna MA, Gómez-Rodríguez C, Bascompte J (2006) Spatial network structure and amphibian persistence in stochastic environments. *Proc Biol Sci* 273:1429–1434.
5. Minor ES, Urban DL (2008) A graph-theory framework for evaluating landscape connectivity and conservation planning. *Conserv Biol* 22:297–307.
6. Cantwell MD, Forman RTT (1993) Landscape graphs: Ecological modeling with graph theory to detect configurations common to diverse landscapes. *Landscape Ecol* 8:239–255.
7. Estrada E, Bodin O (2008) Using network centrality measures to manage landscape connectivity. *Ecol Appl* 18:1810–1825.
8. Bodin O, Norberg J (2007) A network approach for analyzing spatially structured populations in fragmented landscape. *Landscape Ecol* 22(1):31–44.
9. Ferrari JR, Lookingbill TR, Neel MC (2007) Two measures of landscape-graph connectivity: Assessment across gradients in area and configuration. *Landscape Ecol* 22:1315–1323.
10. Frank K, Wissel C (2002) A formula for the mean lifetime of metapopulations in heterogeneous landscapes. *Am Nat* 159:530–552.
11. Drechsler M (2009) Predicting metapopulation lifetime from macroscopic network properties. *Math Biosci* 218(1):59–71.
12. Kininmonth S, Drechsler M, Johst K, Possingham HP (2010) Metapopulation mean life time within complex networks. *Mar Ecol Prog Ser* 417:139–149.
13. Hanski I, Ovaskainen O (2000) The metapopulation capacity of a fragmented landscape. *Nature* 404:755–758.

Table S2. Model selection, based on the AIC, for each network construction predicting the probability of a link

	P	−2LL	AIC	ΔAIC
C. vittiger				
Maximum distance from literature	1	42,411.6	42,413.6	42,001.0
Maximum distance (MLE)	1	519.4	521.4	108.8
Theoretical kernel from literature	1	1,527.2	1,529.2	1,116.6
Theoretical kernel (MLE)	1	629.0	631.0	218.3
Latent space	111	190.6	412.6	0.0
Sender–receiver	112	215.4	439.4	26.7
R. sociabilis plumbeus				
Maximum distance from literature	1	1,914.8	1,916.8	1,789.6
Maximum distance (MLE)	1	221.8	223.8	96.5
Theoretical kernel from literature	1	192.6	194.6	67.4
Theoretical kernel (MLE)	1	180.0	182.0	54.8
Latent space	29	102.2	160.2	33.0
Sender–receiver	30	67.3	127.3	0.0

Fixed effects analysis was used for all network constructions to facilitate accurate model selection comparisons. Using fixed effects in social network constructions is more conservative than using random effects (i.e., it will generally penalize latent space and sender–receiver models more than formulating models with random effects) but avoids the debate of how to contrast models based on conditional inference. See Hoff et al. (1) for a discussion on identifying the number of estimated parameters in the latent space model. −2LL, −2× the log likelihood (the deviance); AIC, −2LL + 2P; ΔAIC_i, AIC for construction *i* minus the minimum AIC in the candidate network constructions that we compared; P, number of estimated parameters. Smaller AIC values reflect more parsimonious constructions given the observed data. The total number of possible links being modeled for *C. vittiger* is 3,080, and the total number of possible links being modeled for *R. sociabilis plumbeus* is 210.

1. Hoff PD, Raftery AE, Handcock MS (2002) Latent space approaches to social network analysis. *J Am Stat Assoc* 97:1090−1098.

Table S3. Predictive accuracy and model selection, based on the AIC, for each network to test for patch-size effects influencing the probability of a link (movement) between patches in the network

Network construction	ΔAIC	AUC	κ	TSS
C. vittiger				
Sender–receiver	21.9	0.882	0.293	0.328
Sender–receiver + size$_{sender}$	16.5	0.887	0.270	0.369
Sender–receiver + size$_{receiver}$	5.3	0.876	0.263	0.307
Sender–receiver + size$_{sender}$ + size$_{receiver}$	0.0	0.899	0.314	0.383
R. sociabilis plumbeus				
Sender–receiver	0.0	0.778	0.272	0.310
Sender–receiver + size$_{sender}$	0.9	0.800	0.397	0.370
Sender–receiver + size$_{receiver}$	2.0	0.789	0.284	0.307
Sender–receiver + size$_{sender}$ + size$_{receiver}$	2.8	0.786	0.294	0.329

We compared the sender–receiver model shown in Figs. 2 and 3, because it had higher predictive accuracy than the latent space model. Predictive accuracy based on cross-validation using 50 replicate samples with replacement, where 80% of the data were used for model building and 20% were used for model testing (external validation). ΔAIC_i, AIC for construction *i* minus the minimum AIC of the models considered; κ, Cohen's κ.