

Network modularity reveals critical scales for connectivity in ecology and evolution

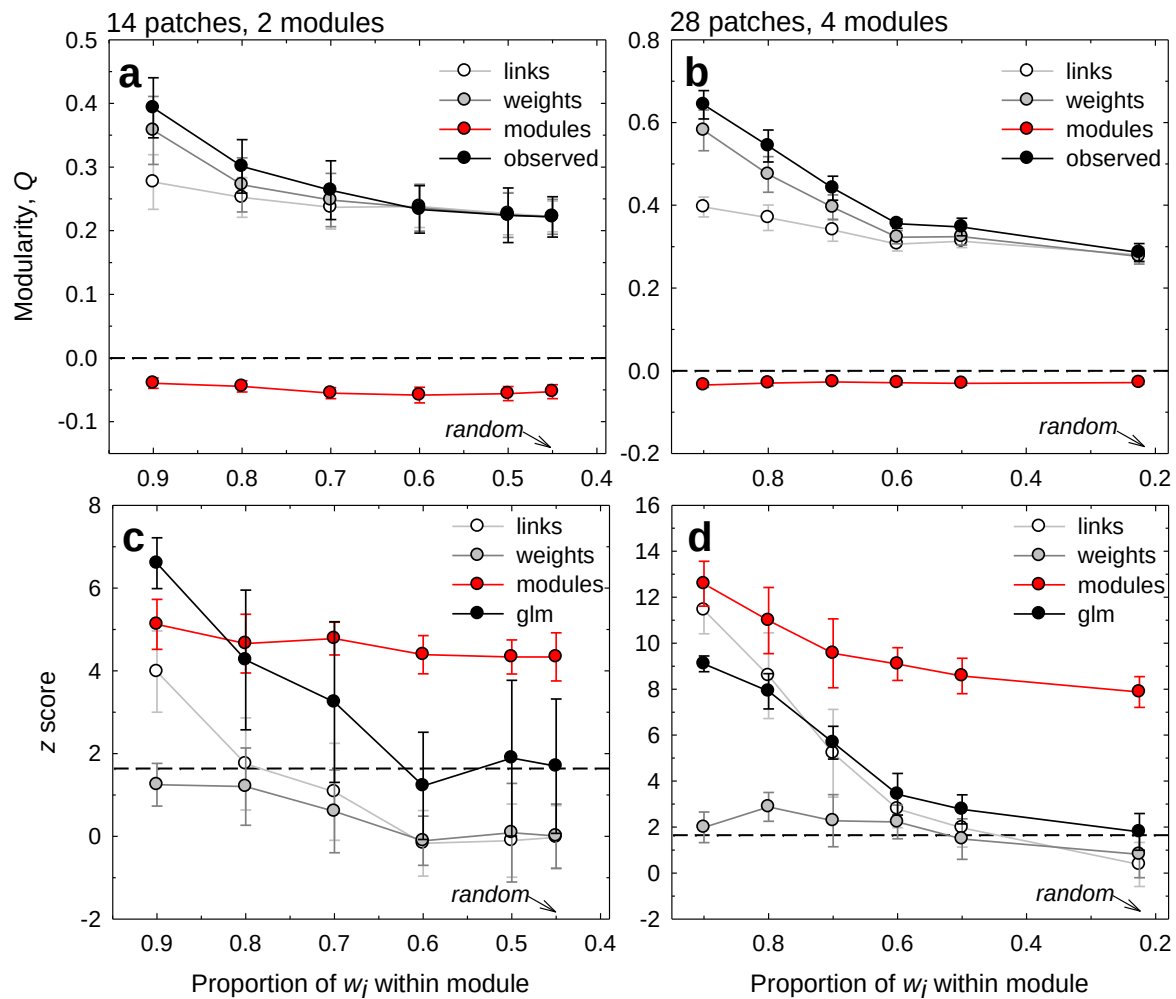
Robert J. Fletcher, Jr.¹, Andre Revell¹, Brian E. Reichert¹, Wiley M. Kitchens², Jeremy D. Dixon³, and James D. Austin¹

¹Department of Wildlife Ecology and Conservation, PO Box 110430, 110 Newins-Ziegler Hall, University of Florida, Gainesville, FL 32611-0430.

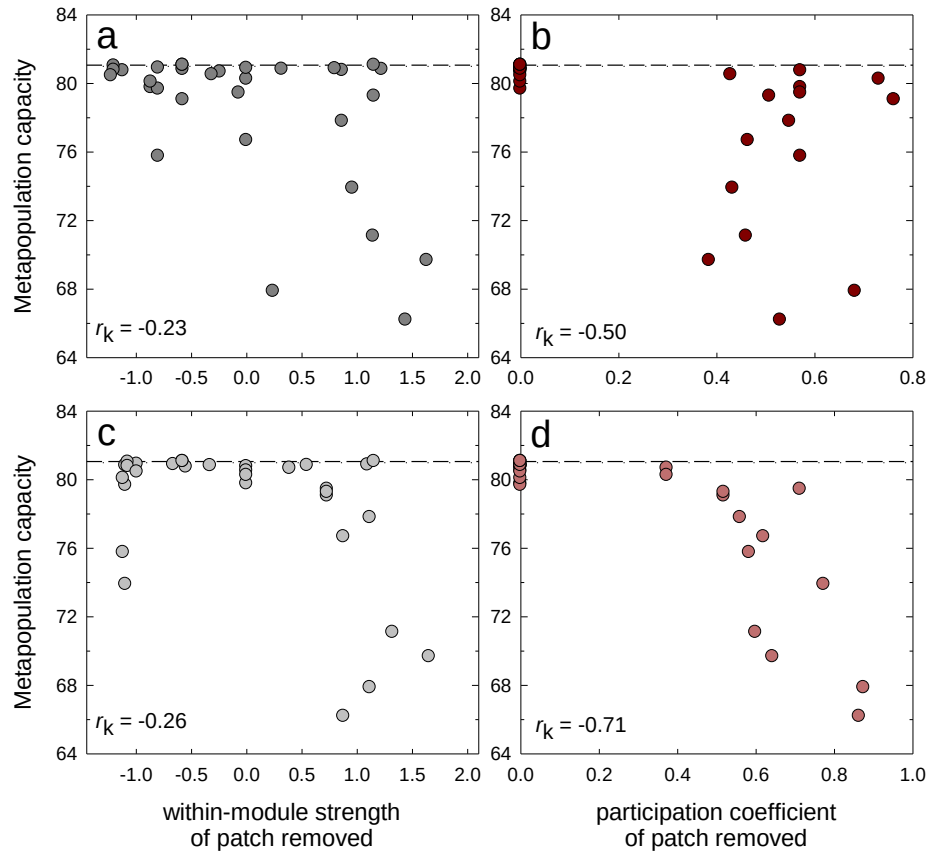
²U.S. Geological Survey, Florida Cooperative Fish and Wildlife Research Unit, University of Florida, Gainesville, FL 32611-0430

³Crocodile Lake National Wildlife Refuge, 10750 County Rd 905, Key Largo, FL 33037

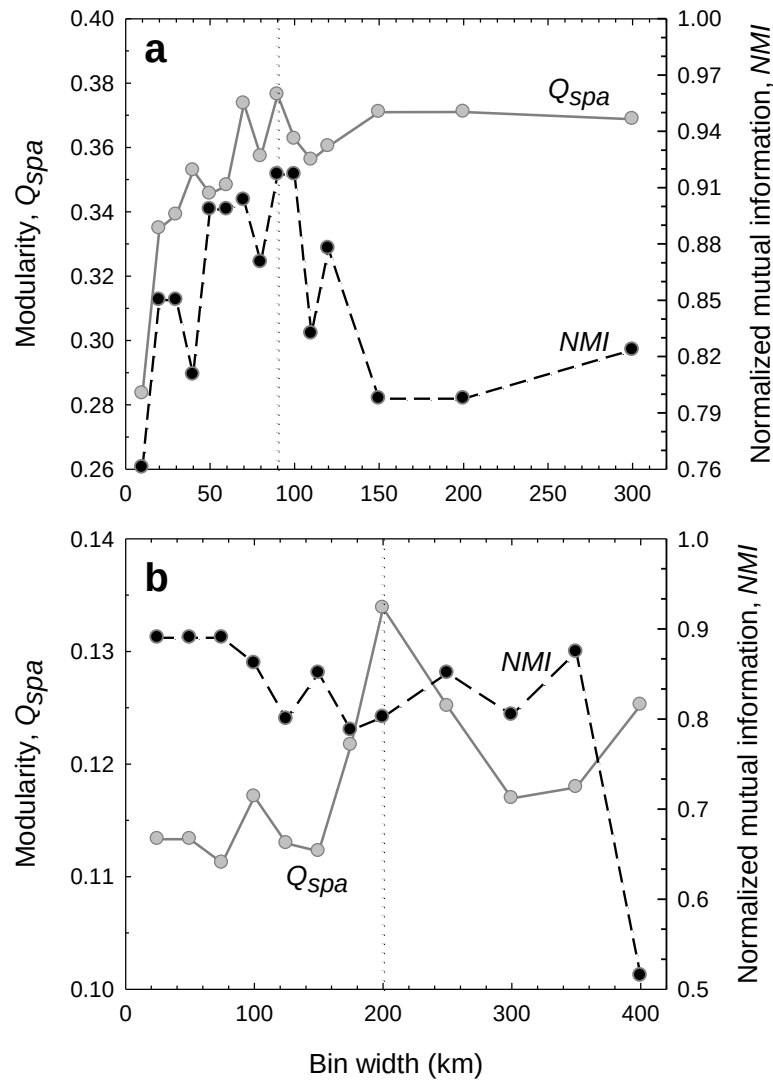
Supplementary Figures



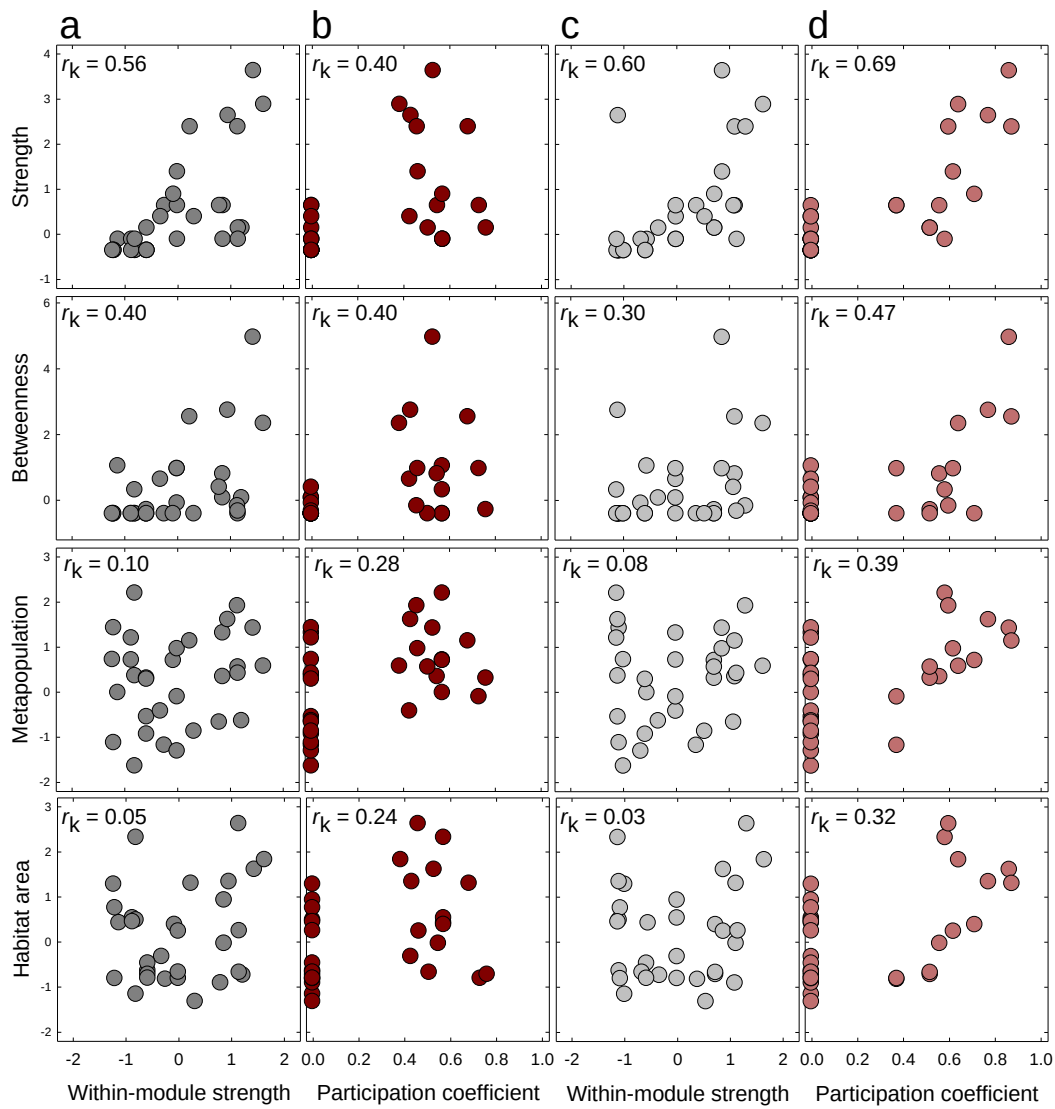
Supplementary Figure S1 | Ability of a variety of significance tests to identify known modules of different strengths. **(a, b)** Modularity ($\pm$ SD) of the observed simulated network and modularity of randomized networks based on randomizing links, weights, or module labels as a function of module strength described as the proportion of movements within modules. **(c,d)** Z scores ($\pm$ SD) for link randomization, weight randomization, module (membership) randomization, and a Poisson GLM. Dashed line indicates a z-score with $P = 0.05$. Arrow denotes random networks for each network size, where $E(A_{ij})$ is equal among all links.



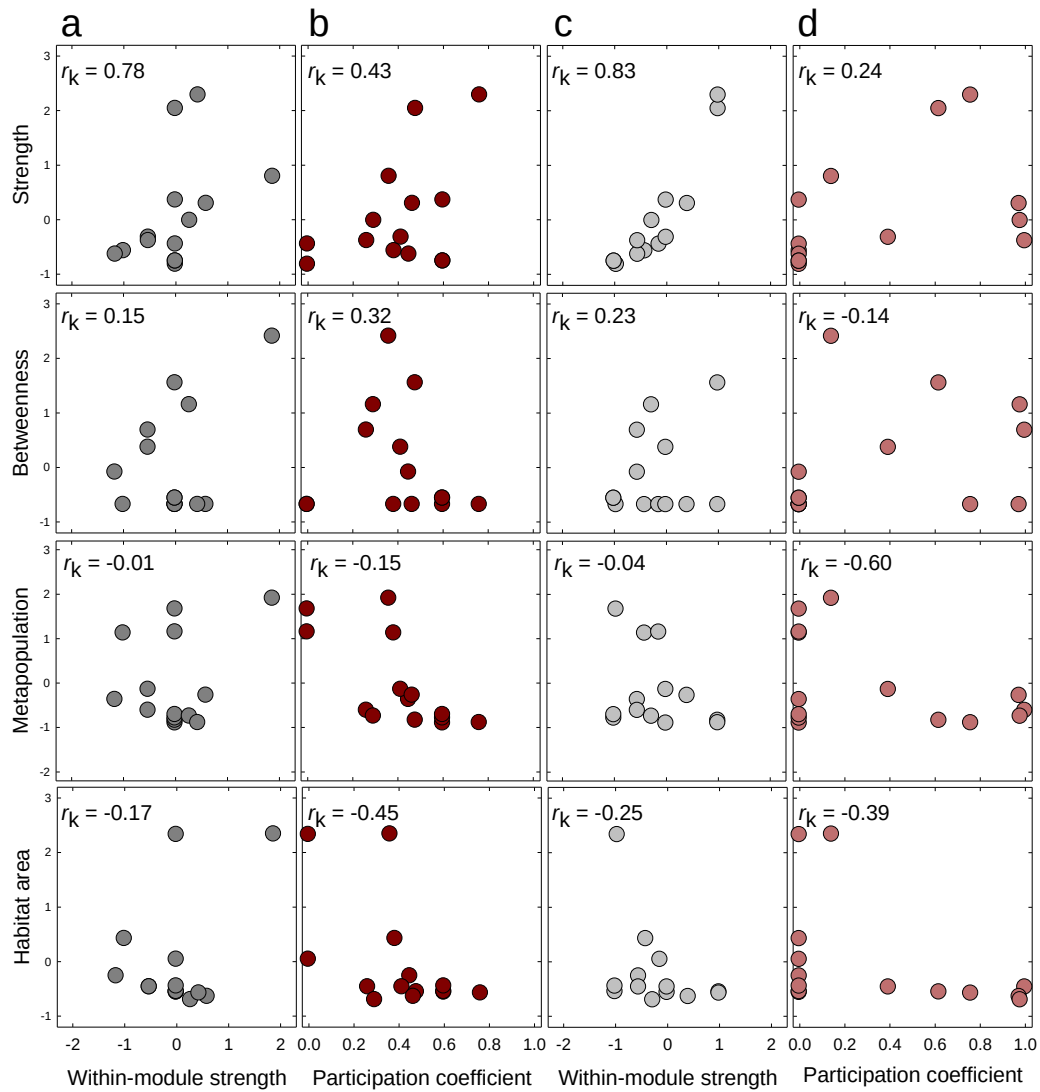
Supplementary Figure S2 | Metapopulation capacity of cactus bugs as a function of the within-module strength and participation coefficient of removed patches. **(a,b)** Module identification does not account for distance effects (Q_{ng} , see Fig. 1a). **(c, d)** Module identification accounts for distance effects (Q_{spa} , see Fig. 1c). Dashed line shows metapopulation capacity in the absence of patch removal; r_k = Kendall's tau rank correlation. A model based on modularity metrics from Q_{spa} fit the data better than for metrics from Q_{ng} (Akaike's information criterion, 151.1 versus 165.5). For both models, participation coefficient (Q_{ng} : $\beta = -6.5 \pm 2.1$ SE, $P = 0.004$; Q_{spa} : $\beta = -9.9 \pm 1.7$, $P < 0.001$) better predicted metapopulation capacity than did within-module strength (Q_{ng} : $\beta = -2.0.99 \pm 0.7$, $P = 0.006$; Q_{spa} : $\beta = -2.7 \pm 0.6$, $P = 0.653$).



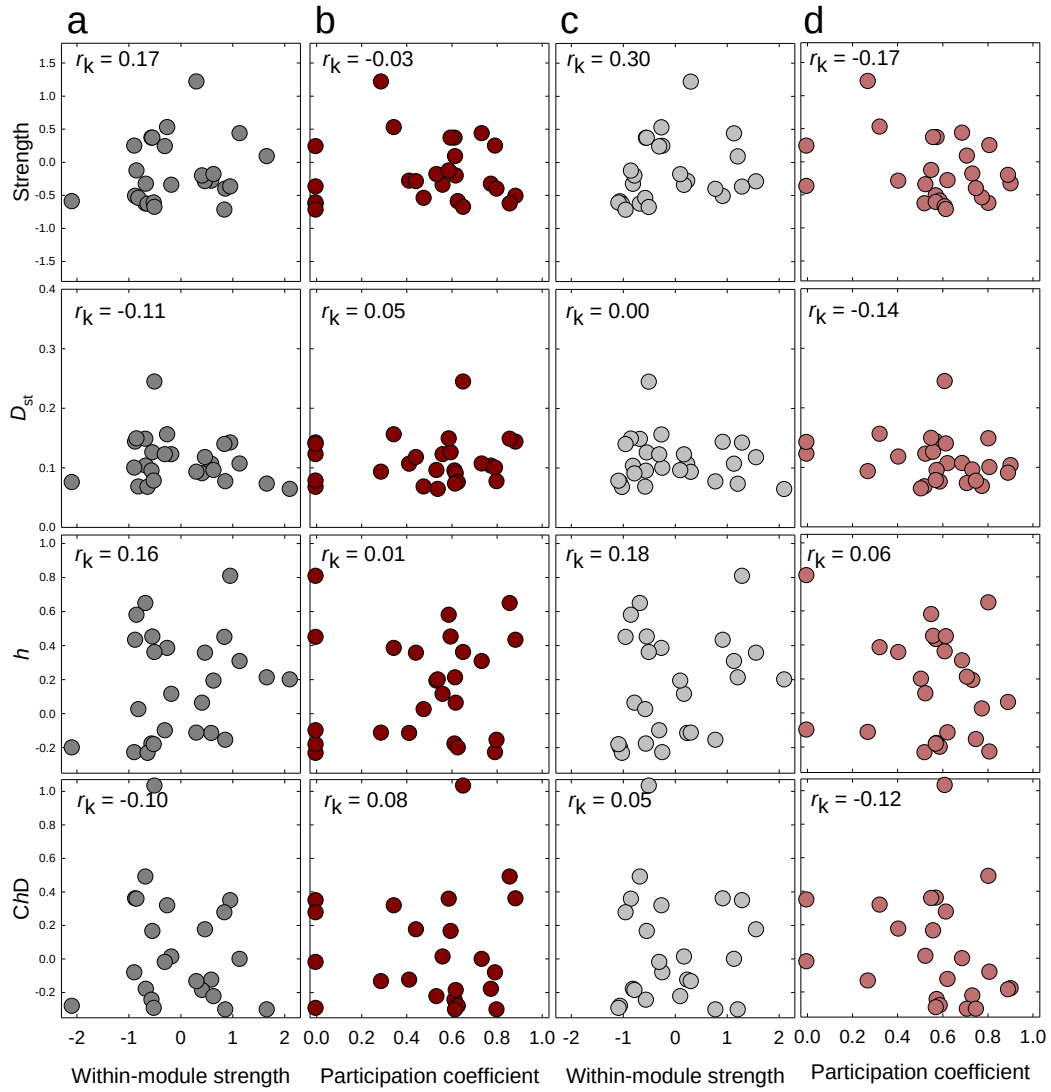
Supplementary Figure S3 | Sensitivity of modularity to bin size. Shown are changes in modularity values, Q_{spa} , with increasing bin size distances for **(a)** bullfrog and **(b)** black bear, as well as the normalized mutual information (NMI) between each bin size and all other bin sizes considered (the relative similarity of modules identified). Dotted line highlights bin size, with maximum Q_{spa} , which was used for assessments shown in the main text.



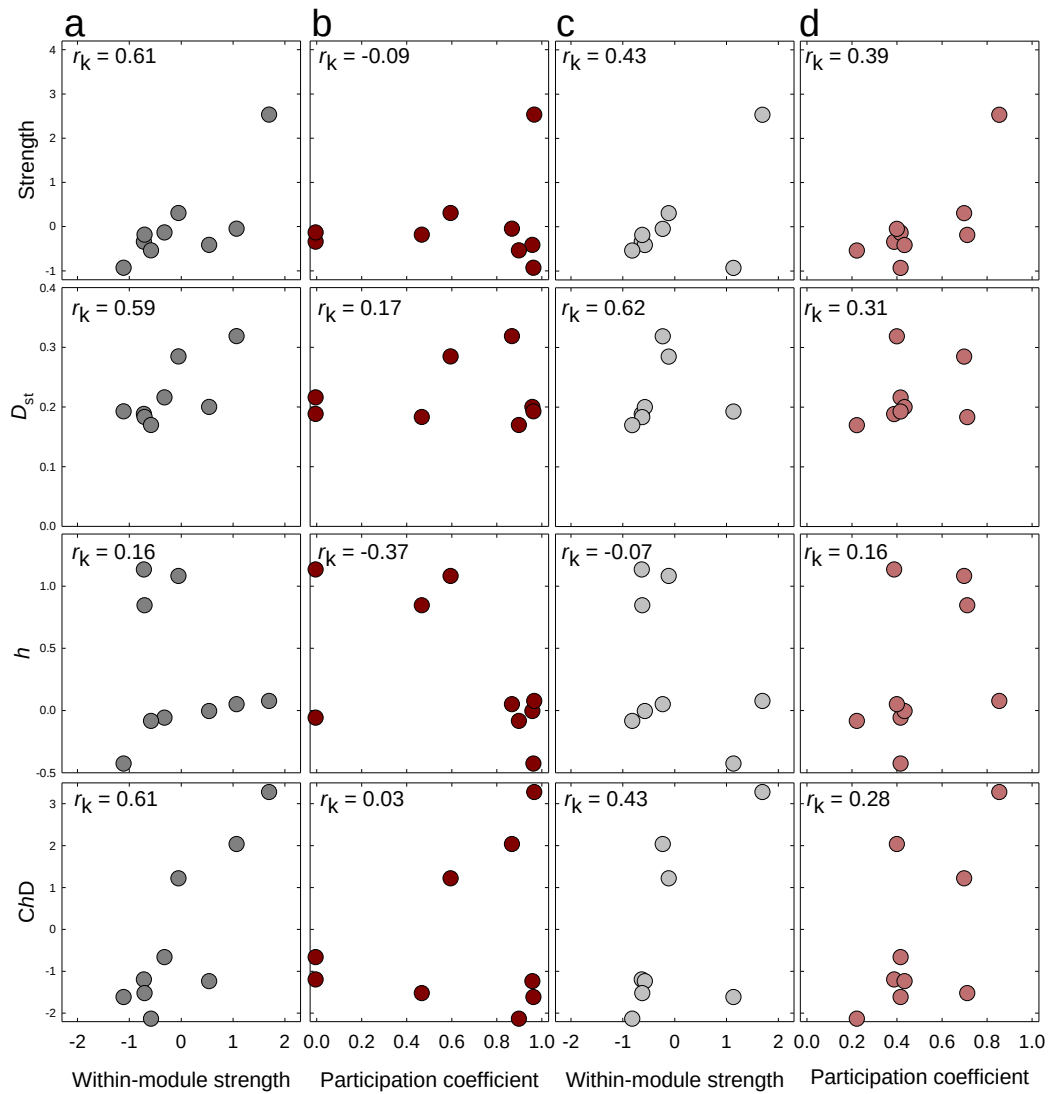
Supplementary Figure S4 | Comparison of within-module strength and participation coefficient with four common connectivity measures in cactus bugs. **(a,b)** Module identification does not account for distance effects (Q_{ng} , see Fig. 1a). **(c, d)** Module identification accounts for distance effects (Q_{spa} , see Fig. 1c). Patch strength, betweenness, metapopulation connectivity, and habitat area were centered and scaled (mean = 0, var = 1). r_k = Kendall's tau rank correlation.



Supplementary Figure S5 | Comparison of within-module strength and participation coefficient with four common connectivity measures in snail kites. **(a,b)** Module identification does not account for distance effects (Q_{ng} , see Fig. 1a). **(c, d)** Module identification accounts for distance effects (Q_{spa} , see Fig. 1c). Patch strength, betweenness, metapopulation connectivity, and habitat area were centered and scaled (mean = 0, var = 1). r_k = Kendall's tau rank correlation.



Supplementary Figure S6 | Comparison of within-module strength and participation coefficient with four common genetic connectivity measures in bullfrogs. **(a,b)** Module identification does not account for distance effects (Q_{ng} , see Fig. 1a). **(c, d)** Module identification accounts for distance effects (Q_{spa} , see Fig. 1c). Patch strength was centered and scaled (mean = 0, var = 1). r_k = Kendall's tau rank correlation.



Supplementary Figure S7 | Comparison of within-module strength and participation coefficient with four common genetic connectivity measures in black bears. **(a,b)** Module identification does not account for distance effects (Q_{ng} , see Fig. 1a). **(c, d)** Module identification accounts for distance effects (Q_{spa} , see Fig. 1c). Patch strength was centered and scaled (mean = 0, var = 1). r_k = Kendall's tau rank correlation.

Supplementary Tables

Supplementary Table S1. Summary of generalized linear models testing for differences in movement or genetic covariance within versus between identified modules.

Network	Z-value/t-value*	P-value	$w_{\text{within}}/w_{\text{between}}$
Q_{ng}			
Cactus bug	3.88	0.0001	2.89
Snail kite	8.07	<0.0001	3.91
Bullfrog	2.97	0.0045	2.33
Black bear	1.12	0.2814	1.94
Q_{spa}			
Cactus bug	2.54	0.019	1.50
Snail kite	-3.50	0.0004	0.61
Bullfrog	2.62	0.012	1.88
Black bear	0.66	0.517	1.19

*Z-value for Poisson GLMs, t-value for zero-adjusted gamma GLMs. Both types of models were fit with the `gamlss` package in R 2.15.

Supplementary Methods

Modularity optimization. Several modularity optimization techniques have been proposed. It is frequently argued that for small networks, optimization based on simulated annealing⁶⁵ is preferred (e.g., <200 patches), for moderate-sized networks (~200-1000 patches), spectral methods⁶⁶ are preferred, and for very large networks (>1000 patches), so-called ‘greedy’ techniques are preferred⁶⁷. Simulated annealing has been consistently shown to be a powerful approach for identifying modules under a variety of scenarios^{65,68}.

We used a simulated annealing algorithm for optimizing the modularity function based on the general approach of Guimera and Amaral⁶⁵. Note that our code is slightly different (and more general) than the NETCARTO program of Guimera and Amaral⁶⁵, because it allows for the ability to consider a variety of null models of relevance to ecology and evolution (see main text), which facilitates comparisons among networks and null models. Our simulated annealing algorithm was written in R 2.15 (Any use of trade, firm, or product names is for descriptive purposes only and does not imply endorsement by the U.S. Government).

Simulated annealing is useful for optimizing the modularity function because of the large number of possible module combinations. Simulated annealing uses an iterative process relying on a unique feature T (“temperature”) that decreases by a factor of c (“cooling factor”) after each iteration. This feature allows the algorithm to explore areas of high modularity without getting stuck in local maxima because at an initial high T , the algorithm can explore multiple local maxima in order to eventually find the absolute maximum. Each iteration of the algorithm is divided into two main sections, which themselves are iterated fN^2 times and fN times, respectively, where N is the number of nodes (patches) in the network and we set $f = 1$ ⁶⁵. In the

first section, a random patch is chosen and moved to another module. The algorithm either accepts or rejects this update based on the new modularity value and T^{65} . In the second section, the algorithm randomly chooses to merge two modules or to split one module into two modules. Within each iteration of the second section, only one update is performed on one randomly selected module (in the case of splitting) or two randomly selected modules (in the case of merging). For all analyses, we started the algorithm with $T = 1$, $c = 0.95$, and ran the algorithm until $T = 0.00003$. We initially tested the performance of the algorithm with random networks with known modules. These random networks varied in size similar to the range our paper used and the algorithm reliably found the known modules (see below).

We also note that because simulated annealing is a stochastic optimization function, we ran the optimization 100 times and used the maximum Q from these runs. The modules identified consistent among these runs. To compare among module partitions, we used a variant of normalized mutual information⁶⁸, NMI , calculated with the `clue` package in R. Normalized mutual information is an index based on entropy and ranges between 0-1. Comparing two identical module assignments (partitions) will result in $NMI = 1$, whereas two module assignments that show no overlap of mutual assignments will result in $NMI = 0$.

Significance tests. While the modularity, Q , is zero for situations of no identified modularity (i.e., one module), and is bound between -1 and 1, random networks can often generate non-zero Q values⁶⁹. However, there has been limited consideration of appropriate ways to assess the significance of observed modularity. Here, we describe four different ways to understand if Q and the modules identified are statistically meaningful. We then use simulations to determine the ability of such approaches to identify significant modularity when modularity is known.

Randomization tests are often used to assess significance of network patterns in biology⁷⁰, and are the most common approach for assessing the significance of observed modularity⁷¹. We consider three potential randomization tests: randomizing links of the observed network⁷¹, randomizing the weights of the observed network⁷², and randomizing the module assignments⁷⁰. The general question that these randomization tests address is: what are the range of Q values that would be expected from a random network of the same dimension and movement? Each test constrains the randomization process in different ways. Randomizing links is most commonly done by constraining the degree distribution of the randomized networks to be the same as the observed distribution⁷¹. This has the effect that randomized networks have patches with the same frequency of movements among patches. For randomizing links, we used a local rewiring algorithm⁷³. For weighted networks, like those considered here, one could alternatively randomize the observed weights, while leaving the overall topology intact⁷². Finally, another approach to randomization tests is to shuffle observed module assignments, leaving the network intact. This approach is similar in concept to the use of Analysis of Similarity Matrices in community ecology.

As an alternative to randomization tests, observed modules can be tested for significance by comparing the amount of movement within modules to the amount of movement between modules⁷⁴. This approach addresses the question: is there a significant variation in movement within versus among modules? While the simulated annealing algorithm finds the best partition, it does not guarantee that this partition has a strong classification of within versus between module movements. As such, this test allows for understanding if the observed modules are biologically meaningful, in terms of partitioning variation in movements across networks.

Simulations assessing significance tests. We assessed the ability of these approaches to identify significant modularity using simulations based on networks with similar properties to those considered here. We simulated networks with two scenarios. In the first scenario, we considered a network containing 14 patches and 2 modules (7 patches/module). In the second scenario, we considered a network with 28 patches and 4 modules (7 patches/module). Based on the observed movement in snail kites, where average patch strength w_i was 7.2 movements, we simulated a gradient of modularity strength where we varied the relative amount (based on a Poisson distribution) of within versus between module movements, ranging from 90% of movements being within modules to 50% of movements being within modules^{30,65}. For each scenario and modularity strength, we simulated 10 replicate networks. For each network, we assessed statistical significance of modularity using link randomization, weight randomization, membership randomization, and testing within versus between module movement using a generalized linear model (GLM) with a log link function and a Poisson error distribution. For randomization tests, we used 100 replicate randomizations. We report z scores for each test as a metric of statistical significance³².

Overall, these statistical methods varied in their ability to identify significant modularity (Supplementary Fig. S1). We found that weight randomization and membership randomization were poor tests for assessing modularity significance, with weight randomization consistently underestimating the significance of modularity whereas membership randomization tests concluded modularity was always significant, even for entirely random networks (where $E(A_{ij})$ was the same within and between modules). For small networks, link rewiring tests only identified significant modularity in the most extreme situations, where 90% of movement was contained within modules, whereas for larger networks this test concluded significant modularity

> 60% of movement was within modules. Note that results from randomizing links and weights also illustrate that random networks can have values of $Q > 0$. Using a simple GLM to test for within versus between module movement was the most powerful test for small networks, where it concluded no significant modularity on random networks but significant modularity for networks with > 60% of movement within modules and for large networks with > 50% of movement within modules. Based on these simulations, we focus on the use of GLMs to assess significance of modularity in our empirical datasets.

Modularity metrics relative to common connectivity metrics. Several connectivity metrics have been proposed in ecology and evolution⁷⁵⁻⁷⁷. In the main text, we compare connectivity metrics for patch importance based on within-module strength and the ‘participation coefficient’ (equations 5-6 of main text). To do so, we contrasted these module-derived metrics with that of patch strength ($w_i = \sum_j A_{ij}$). We used patch strength, w_i , because it is the metric most directly comparable to within-module importance and participation coefficient.

Here we also contrast within-module importance and participation coefficient to other patch-based metrics for connectivity in ecology and evolution. For mark-recapture data of individual movements, we focus on three metrics: betweenness centrality⁷⁷, habitat area in the surrounding landscape^{78,79}, and a common metapopulation metric for connectivity^{79,80}. We focus on betweenness centrality (i.e., the number of shortest paths going through a focal patch) because it is thought to be a highly relevant metric for identifying stepping stones or key bottlenecks in landscapes⁵¹, such that we expected betweenness centrality may be highly correlated with between-module connectivity (participation coefficients). Indeed, the initial approach to

optimizing the modularity algorithm used a heuristic based on betweenness centrality³⁰. Habitat area (where the buffer radius, r , is $= 1/\alpha$,⁷⁹) is a common metric for landscape investigations that does not require information on movement. Finally, because metapopulation theory has made major advances in our knowledge of connectivity and its effects on metapopulations⁸¹, we consider the commonly used metric:

$$mp_i = \sum_j (\exp(-\alpha d_{ij})) S_j I_j \quad (S1)$$

where I_j is an indicator of species presence in patch j .

For genetic data, we focused on comparisons with genetic metrics that assess patch diversity including the relative differentiation of individual patches (D_{ST}) and gene diversity (h)⁸². We decompose the contribution of each patch to the total h into components relating to the contribution due to divergence (ChD) following Petit et al.⁸³. These measures quantify the effect of a specific patch (k) to all others ($n-1$), excluding k .

For each of these metrics, we contrast patch prioritization for connectivity with within-module strength and participation coefficients calculated from Q_{ng} and Q_{spa} . In the main text, we illustrate patch connectivity roles for within-module and between module connectivity (Fig. 2 a-c) and changes in rank importance (Fig. 2d-f) in comparison to patch strength. Here, we also show rank correlations of within-module strength and participation coefficient with these other commonly used metrics for connectivity.

For mark-recapture data, within-module strength and participation coefficient were generally most correlated with patch strength, which was expected given that this metric shares the most similarity with the module-based connectivity metrics, and least with metapopulation connectivity, mp_i , and habitat area (Supplementary Figures S3-S4). Beyond patch strength correlations, all other correlations were ≤ 0.47 , with some correlations being strongly negative

for snail kites. For example, some of the most connected patches according to the metapopulation connectivity metric exhibited the lowest participation coefficients in the snail kite network. For genetics data, patch diversity correlations were generally weaker than for connectivity metrics used in the mark-recapture data. Correlations were weaker for the bullfrog than for the black bear (Supplementary Figures S6-S7), and were generally weaker in reference to the participation coefficient than within-module strength.

Supplementary References

- 65 Guimera, R. & Amaral, L. A. N. Cartography of complex networks: modules and
universal roles. *Journal of Statistical Mechanics-Theory and Experiment* (2005).
- 66 Newman, M. E. J. Modularity and community structure in networks. *Proc. Natl. Acad.
Sci. USA* **103**, 8577-8582 (2006).
- 67 Blondel, V. D., Guillaume, J.-L., Lambiotte, R., & Lefebvre, E. Fast unfolding of
communities in large networks. *Journal of Statistical Mechanics-Theory and Experiment* (2008).
doi:10.1088/1742-5468/2008/10/P10008
- 68 Didham, R. K. & Ewers, R. M. Predicting the impacts of edge effects in fragmented
habitats: Laurance and Yensen's core area model revisited. *Biol. Conserv.* **155**, 104-110
(2012).
- 69 Guimera, R., Sales-Pardo, M., & Amaral, L. A. N. Modularity from fluctuations in
random graphs and complex networks. *Physical Review E* **70(2)**, 1-4 (2004).
- 70 Croft, D. P., Madden, J. R., Franks, D. W., & James, R. Hypothesis testing in animal
social networks. *Trends Ecol. Evol.* **26**, 502-507 (2011).
- 71 Olesen, J. M., Bascompte, J., Dupont, Y. L., & Jordano, P. The modularity of pollination
networks. *Proc. Natl. Acad. Sci. USA* **104**, 19891-19896 (2007).
- 72 Barrat, A., Barthelemy, M., Pastor-Satorras, R., & Vespignani, A. The architecture of
complex weighted networks. *Proc. Natl. Acad. Sci. USA* **101**, 3747-3752 (2004).
- 73 Kimura, M. & Weiss, G. H. Stepping stone model of population structure and decrease of
genetic correlation with distance. *Genetics* **49**, 561-576 (1964).
- 74 Wang, Z. & Zhang, J. Z. In search of the biological significance of modular structures in
protein networks. *Plos Computational Biology* **3**, 1011-1021 (2007).

- 75 Calabrese, J. M. & Fagan, W. F. A comparison-shopper's guide to connectivity metrics.
Front Ecol Environ **2**, 529-536 (2004).
- 76 Minor, E. S. & Urban, D. L. A graph-theory framework for evaluating landscape
connectivity and conservation planning. *Conserv. Biol.* **22**, 297-307 (2008).
- 77 Estrada, E. & Bodin, O. Using network centrality measures to manage landscape
connectivity. *Ecol. Appl.* **18**, 1810-1825 (2008).
- 78 Fahrig, L. Effects of habitat fragmentation on biodiversity. *Ann Rev Ecol Evol Syst* **34**,
487-515 (2003).
- 79 Moilanen, A. & Nieminen, M. Simple connectivity measures in spatial ecology. *Ecology*
83, 1131-1145 (2002).
- 80 Hanski, I. A practical model of metapopulation dynamics. *J. Anim. Ecol.* **63**, 151-162
(1994).
- 81 Hanski, I. Metapopulation dynamics. *Nature* **396**, 41-49 (1998).
- 82 Gahl, M. K., Calhoun, A. J. K., & Graves, R. Facultative use of seasonal pools by
American bullfrogs (*Rana catesbeiana*). *Wetlands* **29**, 697-703 (2009).
- 83 Dixon, J. D. et al. Effectiveness of a regional corridor in connecting two Florida black
bear populations. *Conserv. Biol.* **20**, 155-162 (2006).