

RESEARCH ARTICLE

Maximising the clustering coefficient of networks and the effects on habitat network robustness

Henriette Heer^{1*}, Lucas Streib¹, Ralf B. Schäfer¹, Stefan Ruzika²

1 Department of Quantitative Landscape Ecology, iES Landau, University of Koblenz-Landau, Landau i.d. Pfalz, Germany, **2** Department of Mathematics, University of Kaiserslautern, Kaiserslautern, Germany

* heer@uni-koblenz.de**OPEN ACCESS**

Citation: Heer H, Streib L, Schäfer RB, Ruzika S (2020) Maximising the clustering coefficient of networks and the effects on habitat network robustness. PLoS ONE 15(10): e0240940. <https://doi.org/10.1371/journal.pone.0240940>

Editor: Irene Sendiña-Nadal, Universidad Rey Juan Carlos, SPAIN

Received: June 24, 2020

Accepted: October 6, 2020

Published: October 20, 2020

Copyright: © 2020 Heer et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: All input data, code and results are available on <https://www.kaggle.com/henrietteheer/maximising-robustness>.

Funding: H.H. and L.S. have been supported by the B-M-U Graduate Academy of the University of Koblenz-Landau <https://www.uni-koblenz-landau.de/de/graduieretenakademie/graduieretenakademie>. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing interests: The authors have declared that no competing interests exist.

Abstract

The robustness of networks against node failure and the response of networks to node removal has been studied extensively for networks such as transportation networks, power grids, and food webs. In many cases, a network's clustering coefficient was identified as a good indicator for network robustness. In ecology, habitat networks constitute a powerful tool to represent metapopulations or -communities, where nodes represent habitat patches and links indicate how these are connected. Current climate and land-use changes result in decline of habitat area and its connectivity and are thus the main drivers for the ongoing biodiversity loss. Conservation efforts are therefore needed to improve the connectivity and mitigate effects of habitat loss. Habitat loss can easily be modelled with the help of habitat networks and the question arises how to modify networks to obtain higher robustness. Here, we develop tools to identify which links should be added to a network to increase the robustness. We introduce two different heuristics, Greedy and Lazy Greedy, to maximize the clustering coefficient if multiple links can be added. We test these approaches and compare the results to the optimal solution for different generic networks including a variety of standard networks as well as spatially explicit landscape based habitat networks. In a last step, we simulate the robustness of habitat networks before and after adding multiple links and investigate the increase in robustness depending on both the number of added links and the heuristic used. We found that using our heuristics to add links to sparse networks such as habitat networks has a greater impact on the clustering coefficient compared to randomly adding links. The Greedy algorithm delivered optimal results in almost all cases when adding two links to the network. Furthermore, the robustness of networks increased with the number of additional links added using the Greedy or Lazy Greedy algorithm.

Introduction

Habitat loss and fragmentation due to changes in climate and land use are one of the main drivers of the ongoing global biodiversity crisis [1–5]. The loss and fragmentation of habitat lead to a decrease in habitat connectivity, impeding the movement of individuals between

patches [6, 7]. This dispersal is crucial for species survival, as it facilitates interaction such as the exchange of genes between different populations and thus allows for the existence of metapopulations—a “population of populations” [8–10]. As a consequence of the constantly intensifying climate and land-use change, it is important for species conservation that we particularly try to preserve and improve habitat connectivity by creating dispersal corridors increasing a landscape’s permeability [11–14].

Graph theory provides powerful tools to represent and analyse habitat connectivity in highly fragmented landscapes [15–18]. Here, metapopulations are represented by habitat networks where nodes represent habitat patches and links indicate how these are connected [16, 19, 20]. With the help of habitat networks, the loss of habitat can easily be represented by removing nodes and reduced connectivity by removing links from the network [21–23]. Accordingly, many studies apply graph-theoretic tools to evaluate the effect of climate and land-use change and to find solutions for these effects in landscape planning [24–26].

The resilience of networks against node and link removal, also called network robustness, has been studied in a variety of networks, such as transportation networks, power grids, and food webs [27–31]. A network’s clustering coefficient was identified as a good proxy for robustness in a variety of networks such as habitat networks of herbivores and brown bears [32–34]. The clustering coefficient of a network was proposed by Watts and Strogatz [35] and is defined as the average of the local clustering coefficient of its nodes. A node’s clustering coefficient measures how close its neighbourhood is to a complete network in terms of the relative density of links in its neighbourhood. We exploit the relationship between the clustering coefficient and network robustness and improve a network’s robustness by maximising the network’s clustering coefficient.

The question we pose in this work is: Where should additional links best be created within a habitat network to maximise its clustering coefficient? We propose an algorithm to identify the missing link of a network that leads to the biggest increase in network robustness when added to the network, by using the clustering coefficient as an indicator. We introduce two different heuristics, a Greedy algorithm [36] and a deducted Lazy Greedy algorithm, to maximize the clustering coefficient if multiple links can be added. To speed up the two algorithms, we developed a method to update the clustering coefficient of a network after adding one link as opposed to calculating it without any prior knowledge. Both approaches can be applied to any network, regardless of whether or not it is based on a spatial component. We test these approaches and compare the results to the optimal solution for different generic networks including a variety of standard networks independent of space as well as spatially explicit landscape based habitat networks.

In a last step, we simulate the robustness of habitat networks against habitat loss as proposed by Heer et al. [51] before and after adding multiple links and investigate the increase in robustness depending on both the number of added links and the used heuristic. The robustness simulation combines the simulation of habitat loss by randomly removing habitat patches from the network with the simulations of metapopulation dynamics to evaluate the metapopulation’s robustness. Our proposed methods thus provide tools to facilitate landscape restoration by identifying which location leads to the largest improvement when additional links are added in these places.

Methods

Outline of analysis

We present the algorithm to update the clustering coefficient after one link is added and the Greedy and Lazy Greedy algorithms to add more than one link. We evaluated the effect of

adding links using the proposed algorithms on the clustering coefficient and therefore on the habitat network’s robustness. To evaluate the effect of the proposed algorithms on the clustering coefficient, we added two links to a variety of networks using (1) the Greedy algorithm, (2) the Lazy Greedy algorithm, and (3) a purely random approach. The clustering coefficients of the resulting networks were then compared to the clustering coefficient of the original network as well as the optimal solution, which was found by complete enumeration, i.e. iterating over all pairs of potential links. We tested our algorithms on different network types, including sparse standard networks (random, small-world, and regular) [37], dense standard networks, and habitat networks based on artificial landscapes and a generic insect species with both terrestrial and aquatic life stages created by Streib et al. [38].

Finally, we evaluated the effect of modified habitat networks on metapopulation robustness. To this end, we simulated and evaluated the metapopulation robustness as presented by Heer et al. [51] and studied the increase in robustness after adding links using the Greedy algorithm, the Lazy Greedy algorithm, and a random insertion approach. For these simulations, only the landscape-based habitat networks were taken into account as the standard networks are in general poor representatives of habitat networks.

Notation

We use the following notation throughout the manuscript. Let $G = (V, E)$ be a simple, undirected, loopless network with node set V and link set $E \subset V \times V$.

Let $(u, v) \in V \times V \setminus E$ be a pair of unconnected nodes in G . To be able to compare the network G with the extended network that arises from G by adding the link (u, v) to G , we use the following notation and set $G' := (V, E \cup \{(u, v)\})$. If we want to emphasize the link (u, v) , we will write $G + uv := G'$.

For a node $w \in V$, we set $N(w) := \{v \in V : (w, v) \in E\}$ as the set of neighbours of w , $d_w := |N(w)|$ as the degree, i.e. the number of neighbours, of w in G and d'_w the degree of w in G' . A triangle in a network G is a clique of three nodes $\{u, v, w\}$, i.e. all three nodes are connected with each other by links: $(u, v), (u, w), (v, w) \in E$. We set $T(w) := |\{(u, v) \in E : u, v \in N(w)\}|$ as the number of triangles in G that involve w and $T'(w)$ as the number of triangles in G' . Furthermore, with $N(u, v) := N(u) \cap N(v)$ we denote the set of common neighbours of u and v and $k := |N(u, v)|$ the number of common neighbours (Fig 1).

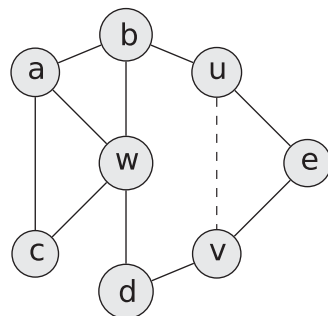


Fig 1. Example network to illustrate notation. $G = (V, E)$ with $n = 8$ nodes, 10 links, $V = \{a, b, c, d, e, u, v, w\}$, and $E = \{(a, b), (a, w), (a, c), (b, u), (b, w), (u, e), (w, c), (w, d), (e, v), (d, v)\}$. We choose $m = 1$ link from the set $\mathcal{E} = V \times V \setminus E$ of all links not included in G . G' is the network G after link (u, v) (represented as dashed line) was added: $G' := G + uv$. Then $N(w) = \{a, b, c, d\}$, $d_w = d'_w = 4$, $T(w) = T'(w) = 2$ (the triangles abw and acw) and $C(w) = \frac{2 \cdot 2}{4(4-1)} = \frac{1}{3}$. For u we obtain $d_u = 2$ and $d'_u = 3$ and similarly $T(u) = 0$ and $T'(u) = 1$ (the triangle uev). $N(u, v) = \{e\}$ and $k = |N(u, v)| = 1$. The clustering coefficient of G equals $C_G = \frac{1}{8} \cdot (\frac{1}{3} + 0 + 0 + \frac{1}{3} + \frac{2}{3} + 1 + 0 + 0) = \frac{1}{8} \cdot \frac{7}{3} = \frac{7}{24}$ and the clustering coefficient of the extended network is $C_{G'} = \frac{1}{8} \cdot (\frac{1}{3} + \frac{1}{3} + \frac{1}{3} + \frac{1}{3} + \frac{2}{3} + 1 + 0 + 1) = \frac{1}{8} \cdot 4 = \frac{1}{2}$.

<https://doi.org/10.1371/journal.pone.0240940.g001>

The clustering coefficient of a node $v \in V$ is defined as

$$C(v) = \begin{cases} \frac{2T(v)}{d_v(d_v-1)} & \text{if } d_v > 1 \\ 0 & \text{if } d_v \leq 1 \end{cases}.$$

It measures how close its neighbourhood is to a complete network in terms of the relative density of links in its neighbourhood. If all links between neighbours of v are present, then $T(v) = \frac{1}{2}d_v(d_v - 1)$ and the clustering coefficient takes its maximum value of 1. If no links between neighbours are present, then $T(v) = 0$ and thus $C(v) = 0$.

The clustering coefficient of a network G with $n := |V|$ nodes is defined as the average over the clustering coefficient of its nodes:

$$C_G = \frac{1}{n} \sum_{v \in V} C(v)$$

and can take any value between 0 and 1. Computing the clustering coefficient of a network with $n := |V|$ nodes has an $\mathcal{O}(n^\omega)$ complexity with $\omega \leq 2.376$ [39]. The most complex part of computing the clustering coefficient is finding triangles in a network, which can be done in $\mathcal{O}(n^\omega)$ using the adjacency matrix and fast matrix multiplication. Let $m \in \mathbb{N}$ be the number of links we want to add to the network and $\mathcal{E} \subseteq V \times V \setminus E$ a set of missing links to choose these m links from. Fig 1 gives an example for each variable introduced here.

Our aim is to improve a network’s robustness by adding links to the network. As the clustering coefficient is a good proxy for robustness [23, 32–34], we want to identify those links that should be added to the network to maximize the clustering coefficient. Mathematically, we want to solve the following problem:

Problem 1. Let $G = (V, E)$ be a network as above, $\mathcal{E} \subseteq V \times V \setminus E$, and $m \geq 1$ be given. Find a subset $\{e_1, \dots, e_m\} \subseteq \mathcal{E}$ such that $G' := (V, E \cup \{e_1, \dots, e_m\})$ has maximum clustering coefficient. In other words, find a solution to

$$\begin{aligned} \max \quad & C_{G'} \\ \text{s.t.} \quad & G' = (V, E \cup \{e_1, \dots, e_m\}) \\ & \{e_1, \dots, e_m\} \subseteq \mathcal{E}. \end{aligned}$$

Example 1. Consider the network $G = (V, E)$ from Fig 1. We set $\mathcal{E} = V \times V \setminus E$ and $m = 1$, i.e. we allow all unconnected pairs of nodes to be connected and the task is to identify $m = 1$ pair that maximizes the clustering coefficient when connecting the pair and adding the link to G . Problem 1 has two solutions, the pair (u, v) as well as the pair (d, e) , which both increase the clustering coefficient to 0.5. If we set $m = 2$ in the same problem, we obtain the unique solution (b, e) and (d, e) with a new clustering coefficient of 0.625.

In some cases, we want to add any link to the network in order to maximize the clustering coefficient and it makes sense to find those potential links $\{e_1, \dots, e_m\}$ within all pairs of unconnected nodes. In this case we set $\mathcal{E} := V \times V \setminus E$. However, especially when considering habitat networks, we may want to restrict this set to only some pairs of unconnected nodes. For habitat networks, for example, we may want to restrict the set to those pairs of unconnected nodes that are within a certain (Euclidean) distance from each other. This represents the assumption that the species in focus has a limited dispersal distance independent of the underlying land-use class [40].

Update clustering coefficient

We first aim to solve Problem 1 for $m = 1$, i.e., we want to find the pair of nodes $(u, v) \in \mathcal{E}$, such that the network $G' = G + uv$ has maximum clustering coefficient.

A naïve approach to find the relevant nodes u and v is to iterate over all unconnected pairs of nodes, connect those, and calculate the clustering coefficient of the extended network from scratch. This has a run time of $\mathcal{O}(|\mathcal{E}|n^2)$, as we iterate over $|\mathcal{E}|$ pairs and calculate the clustering coefficient each time from scratch. To speed up the process, however, we can exploit the fact that adding the link does not affect the clustering coefficient in most nodes. To see this, consider the degree of each node in G as well as the number of triangles it is part of. The degrees of the nodes in G' equal the degrees of the nodes in G , except for the two nodes u and v , as adding (u, v) to G increases the degrees of u and v by exactly one. The number of triangles in u and v each increases by the number of common neighbours of u and v , as each common neighbour $w \in N(u, v)$ introduces the triangle uvw and every triangle that does not use the link (u, v) also exists in G . Similarly, the number of triangles for each common neighbour of u and v increases by exactly one. The number of triangles does not change for every other node that is not u, v or a common neighbor of u and v . Accordingly, we can calculate the clustering coefficient of G' by adding the difference caused by u, v and every common neighbour w of u and v to the original clustering coefficient C_G :

$$C_{G'} = C_G + \frac{1}{n} \left(\Delta C(u) + \Delta C(v) + \sum_{w \in N(u,v)} \frac{2}{d_w(d_w - 1)} \right) \tag{1}$$

with

$$\Delta C(u) = \begin{cases} \frac{2k(d_u-1)-4T(u)}{d_u(d_u-1)} & \text{if } d_u > 1 \\ 1 & \text{if } d_u = 1 \end{cases}.$$

See S1 File for the proof of Eq 1.

It follows from Eq 1 and Fig 2, that adding a link to a network may also result in a smaller clustering coefficient compared to the original network. If u and v have no common neighbours, the sum over all common neighbours in Eq 1 is empty (and thus equals 0) and

$$\Delta C(u) = \frac{2 \cdot 0 \cdot (d_u - 1) - 4T(u)}{d_u(d_u^2 - 1)} = \frac{-4T(u)}{d_u(d_u^2 - 1)} \leq 0.$$

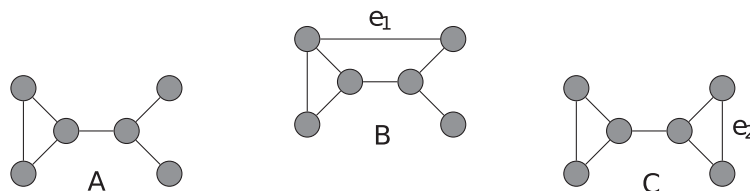


Fig 2. Varying effects of adding a link to a network on the clustering coefficient. (a) Original network with clustering coefficient $C_G = 0.3\bar{8}$. (b) Network after connecting two nodes with no common neighbours, $C_{G+e_1} = 0.2\bar{7}$. (c) Network after connecting two nodes with a common neighbour, $C_{G+e_2} = 0.7$.

<https://doi.org/10.1371/journal.pone.0240940.g002>

Similarly, $\Delta C(v) \leq 0$ and

$$C'_G = C_G + \frac{1}{n}(\Delta C(u) + \Delta C(v) + 0) \leq C_G.$$

Using Eq 1, we can update the clustering coefficient after adding a new link (u, v) to a network $G = (V, E)$ with known clustering coefficient C .

Algorithm 1 Update clustering coefficient

```

1: procedure UPDATECLUSTERING( $G = (V, E), C_G, (u, v)$ )
2:    $C_{\max} \leftarrow 0$ 
3:    $T \leftarrow \text{Triangles}(G)$ 
4:    $CN \leftarrow \text{CommonNeighbours}(u, v)$ 
5:    $k = |CN|$ 
6:   if  $d_u > 1$  then
7:      $C_{\max} \leftarrow C_{\max} + \frac{2k(d_u-1)-4T(u)}{d_u(d_u^2-1)}$ 
8:   else  $C_{\max} \leftarrow C_{\max} + 1$ 
9:   if  $d_v > 1$  then
10:     $C_{\max} \leftarrow C_{\max} + \frac{2k(d_v-1)-4T(v)}{d_v(d_v^2-1)}$ 
11:   else  $C_{\max} \leftarrow C_{\max} + 1$ 
12:   for  $w \in CN$  do
13:      $C_{\max} \leftarrow C_{\max} + \frac{2}{d_w(d_w-1)}$ 
14:    $C_{\max} \leftarrow \frac{C_{\max}}{|V|}$ 
15:    $C_{\max} \leftarrow C_{\max} + C_G$ 
16:   return  $C_{\max}$ 

```

Algorithm 1 takes a network $G = (V, E)$, its clustering coefficient C , and a pair of unconnected nodes u and v as input and returns the clustering coefficient of the extended network $G + uv$ using Eq 1. It finds the set of common neighbours of u and v , calculates $\Delta C(u)$ and $\Delta C(v)$, and then iterates over the set of common neighbours of u and v and increases the sum of $\Delta C(u)$ and $\Delta C(v)$ by $\frac{2}{d_w(d_w-1)}$ for each common neighbour w . The result is then averaged over the number of nodes in G and added to the original clustering coefficient. Eq 1 proves the correctness of this algorithm.

We use Algorithm 1 to develop a faster algorithm than the naïve one to find a solution of Problem 1 for $m = 1$. It iterates over the set \mathcal{E} of all possible pairs of nodes and calculates the new clustering coefficient by updating the clustering coefficient of the original network.

Algorithm 2 Maximize clustering coefficient

```

1: procedure MAXIMIZECLUSTERING( $G = (V, E), \mathcal{E}$ )
2:    $C_G \leftarrow \text{Clustering}(G)$ 
3:    $T \leftarrow \text{Triangles}(G)$ 
4:    $C_{\max} \leftarrow C_G$ 
5:   for  $(u, v) \in \mathcal{E}$  do
6:      $C = \text{UpdateClustering}(G, C_G, (u, v))$ 
7:     if  $C > C_{\max}$  then
8:        $C_{\max} \leftarrow C$ 
9:        $e_{\max} \leftarrow (u, v)$ 
10:  return  $C_{\max}, e_{\max}$ 

```

Algorithm 2 iterates over all potential links, uses Algorithm 1 to update the clustering coefficient and returns a link that yields the maximum clustering coefficient. As Algorithm 1 with input (u, v) returns the clustering coefficient of the extended network $G + uv$, and Algorithm 2 iterates over all potential links, it returns an optimal solution of Problem 1 for $m = 1$ in $\mathcal{O}(n^{\omega} + |\mathcal{E}|d_{\max})$, where d_{\max} is the maximum degree of the nodes in V .

The algorithm solves Problem 1 reasonably fast for $m = 1$. When adding multiple links, however, every combination of potential links needs to be checked, slowing the procedure substantially down even for only two links: There are $\binom{|\mathcal{E}|}{m}$ combinations of potential links and executing Algorithm 2 for each combination has a complexity of $\mathcal{O}(n^\omega + \binom{|\mathcal{E}|}{m}|\mathcal{E}|d_{\max})$. We thus introduce two heuristics, Greedy and Lazy Greedy, that identify the maximum clustering coefficient of a network when multiple links can be added.

Greedy

The Greedy algorithm successively adds one link that maximizes the clustering coefficient of the current network. Starting with a network G , the algorithm iterates over the set \mathcal{E} of all possible pairs of nodes and connects the pair u, v with the biggest increase in the clustering coefficient (see Algorithm 2). It then iterates again over all possible pairs of nodes in G' to find the second link and continues, until m links were found.

Algorithm 3 Greedy

```

1: procedure GREEDY ( $G = (V, E), \mathcal{E}, m$ )
2:   for  $i \in [1, m]$  do
3:      $C, e_i = \text{MaximizeClustering}(G = (V, E), \mathcal{E})$ 
4:      $G = G + e$ 
5:   return  $e_1, \dots, e_m$ 

```

We can calculate the clustering coefficient and the number of triangles once and then update these numbers. In that case, the Greedy algorithm calculates the solution in $\mathcal{O}(n^\omega + |\mathcal{E}|d_{\max}m)$, as it executes Algorithm 2 exactly m times. However, the solution found by the Greedy algorithm 3 is not necessarily optimal: Consider the network depicted in Fig 3a and assume we can add two links. The Greedy algorithm will add the links shown in Fig 3b, while the links shown in Fig 3c lead to a higher clustering coefficient.

Lazy Greedy. For even faster calculations—at the cost of optimality—we introduce a second heuristic, that iterates over all potential links once and then picks the m links that have the highest increase in the clustering coefficient if they were to be added individually.

The Lazy Greedy algorithm executes Algorithm 2 once and sorts the results afterwards. Using quick sort, sorting can be done in $\mathcal{O}(|\mathcal{E}| \log(|\mathcal{E}|))$ and we obtain a run time of $\mathcal{O}(n^\omega + |\mathcal{E}|d_{\max} + |\mathcal{E}| \log(|\mathcal{E}|))$ [41, 42]. Similar to the Greedy algorithm, the Lazy Greedy algorithm does not necessarily find the optimal solution to Problem 1. Fig 3 also serves as example of a

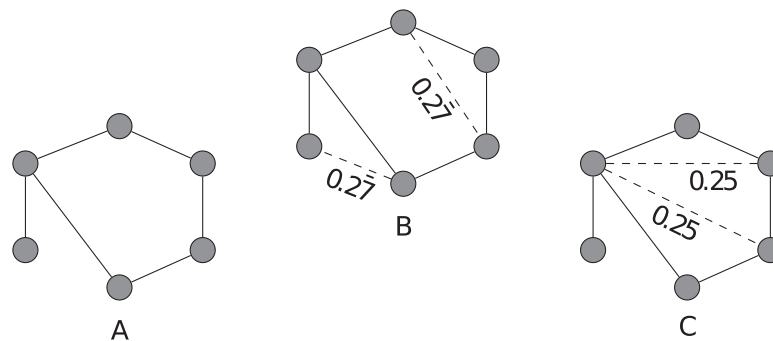


Fig 3. Example of non-optimal behaviour of the Greedy algorithm. (a) Original network with clustering coefficient $C = 0$. (b) Network with two links selected using the Greedy algorithm and clustering coefficient $C = 0.5$. (c) Optimal solution with clustering coefficient $C = 0.605$. The value corresponding to the dashed lines show the increase of clustering coefficient by adding the corresponding link to the network in (a). After adding one of the links depicted as dashed lines to the network in (c), the contribution of the other link increases to 0.35, as the two nodes incident to that link now have one common neighbour more (see Eq 1).

<https://doi.org/10.1371/journal.pone.0240940.g003>

non-optimal solution, as the Lazy Greedy algorithm will select the same links as the Greedy algorithm.

Algorithm 4 Lazy Greedy

```

1: procedure LAZYGREEDY( $G = (V, E), \mathcal{E}, m$ )
2:    $C_G \leftarrow \text{Clustering}(G)$ 
3:    $T \leftarrow \text{Triangles}(G)$ 
4:    $\text{results} \leftarrow \text{new Array}$ 
5:   for  $(u, v) \in \mathcal{E}$  do
6:      $C = \text{UpdateClustering}(G, C_G, (u, v))$ 
7:      $\text{append}(u, v, C)$  to  $\text{results}$ 
8:    $\text{results} \leftarrow \text{sort results by } C$ 
9:   return  $\text{results}[1], \text{results}[m]$ 

```

Random approach

We compared the described heuristics to the results of a random approach, where links were added uniformly at random to a network.

Networks

We tested the described heuristics on a variety of network types, namely landscape-based habitat networks created by Streib et al. [38] with random, clustered, and contiguous allocation of habitat patches / nodes as well as networks common in mathematics (random, regular, and small-world networks) [37]. The random, regular, and small-world networks represent a variety of network structures and are widely used in many disciplines, such as engineering, social sciences, finance, biology, and also ecology [35, 37, 43–46]. Fig 4 shows examples of the networks.

Landscape-based habitat networks. The landscape-based habitat networks were set up by Streib et al. [38] based on a generic insect species with aquatic and terrestrial life stages, landscapes consisting of different landscape types associated with varying dispersal cost, and a 50 km × 50 km section of a real stream network from southwest Germany. The stream network section was divided into 25 tiles of 10 km × 10 km areas and intersected with an artificial landscape consisting of open agricultural land, forestry land, and urban area with associated dispersal costs. A subset containing 10% of the pixels in the real stream network were chosen as habitat patches. We considered 3 types of habitat patch arrangements leading to 3 types of landscape-based habitat networks, namely (1) random (with all habitat patches randomly selected along streams with equal probability), (2) clustered (with only some habitat patches randomly selected along streams with equal probability and the others randomly selected along streams with equal probability within a given radius around any of the initially selected habitat patches), and (3) contiguous (with a smaller fraction of habitat patches randomly selected along streams with equal probability and a larger fraction of others randomly selected along streams with equal probability within a given radius around any of the initially selected habitat patches, leading to a more contiguous arrangement of the habitat patches compared to the clustered allocation). Reflecting the different stream structures in the different landscape tiles, this results in habitat networks with 54 to 111 habitat patches. Habitat patches were connected with the help of a least-cost path analysis based on the dispersal cost in the underlying landscape. If the cumulative dispersal cost between two habitat patches was less than the maximum dispersal cost, the two patches were considered to be connected and a corresponding link was added to the network. Differing from Streib et al. [38], we assumed shorter dispersal ranges of about 1300m through open agricultural land to simulate particular sensitive species. These dispersal ranges translated to maximum dispersal costs of 650 (as we assumed a

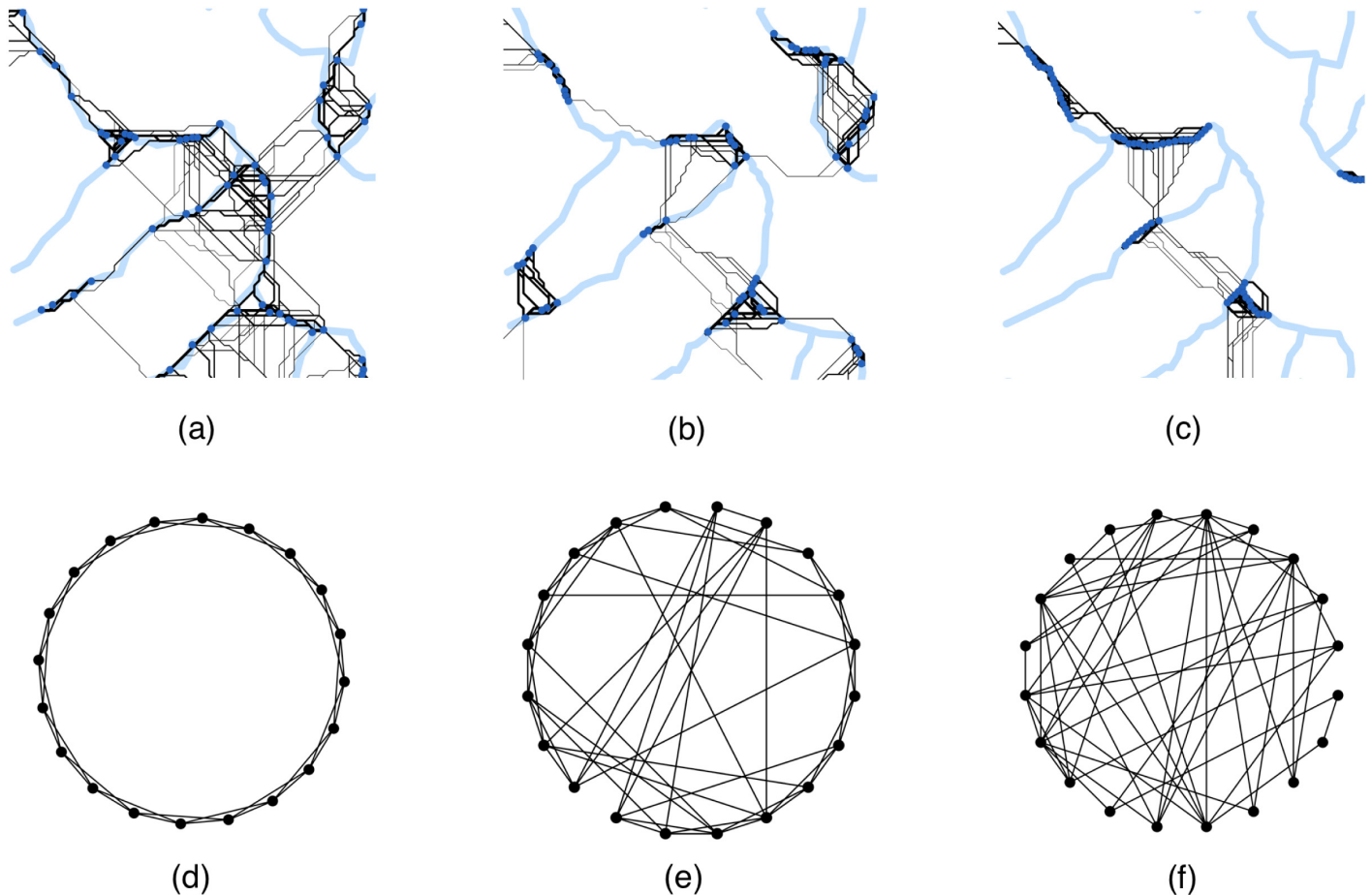


Fig 4. Networks examined. (a)–(c): Landscape-based networks. Dark-blue dots indicate nodes (habitat patches), black lines indicate links (dispersal pathways). The light-blue lines indicate the underlying stream network structure. (a) random allocation of habitat patches, (b) clustered allocation, (c) linear allocation. (d)–(f): Standard networks. (d) regular network, (e) small-world network, (f) random network.

<https://doi.org/10.1371/journal.pone.0240940.g004>

cost of 50 to traverse a $25\text{m} \times 25\text{m}$ area of open agricultural land, see [38] for further information). To ensure that all network types have similar distributions of the number of links, we finally adjusted the maximal dispersal costs to 900 for random, 650 for clustered, and 400 for contiguous habitat allocation. In total we analysed 250 networks per network type random, clustered, and contiguous. See Fig 4(a)–4(d) for examples of the networks.

Standard networks. We created standard networks (random, regular, and small-world) using algorithms from the Python package NetworkX [47]. In random networks, two nodes are connected purely at random with uniform distribution and nodes usually have very similar degrees. They were generated using the algorithm proposed by [48]. Regular networks are networks, where every node has the same degree [37]. Small-world networks are a mixture of regular and random networks and represent the small-world phenomenon from the social sciences [37, 49]. While most nodes are not connected to each other, neighbours of a node are connected with particularly high probability. In other terms, small-world networks are highly clustered and at the same time also exhibit particularly low average shortest path distances. We used the algorithm proposed by [50] to construct small-world networks.

We created two sets of these standard networks varying in their number of links per network. For sparse standard networks, all parameters were set to create networks with a number

of nodes and corresponding links similar to the landscape-based networks. This led to very sparse networks with only 4% of links present. Dense standard networks were also created with a number of nodes similar to the landscape-based networks, however the parameters were chosen such that about 75% of the potential links were present. S3.1 Table in [S1 File](#) shows the parameters and algorithms used to create the standard networks and [Fig 4\(e\)–4\(f\)](#) show examples of the networks.

In total, we analysed 250 networks per network type with the number of nodes between 50 and 111.

Effect of used algorithms on the clustering coefficient

We evaluated the effect of the two proposed algorithms on the clustering coefficient and compared the results to randomly adding links. To this end $m = 2$ links were added to each of the created networks using (1) the Greedy algorithm, (2) the Lazy Greedy algorithm, and (3) a purely random approach. We compared these results with the clustering coefficient of the original network and the optimal solution, which was found by iterating over all pairs of potential links.

In this analysis, we considered both standard and landscape-based networks, as the heuristics to maximize the clustering coefficient can be applied to any network. We defined the set of potential links to be the set of all unconnected pairs of nodes $\mathcal{E} = V \times V \setminus E$.

Effect of used algorithms on robustness of habitat networks

In a last step, we evaluated how much the added links improved the robustness of landscape-based habitat networks against habitat loss. We applied the simulations introduced by Heer et al. [51] to simulate habitat loss and evaluate the habitat network's robustness. For the simulations, a random habitat loss scenario was assumed where habitat patches (i.e., nodes) and corresponding links get lost permanently purely at random. On the remaining networks, random local extinctions were simulated, in a way that depends on the local-extinction risk of species and each patch's neighbourhood. Empty habitat patches could then be recolonised through dispersal from connected colonised habitat patches, in a way that depends on the dispersal range of species and each patch's neighbourhood. These extinction and recolonization processes were continued until a stationary distribution was reached. From this we obtain the fraction of colonised habitat patches. These simulations of habitat loss and subsequent extinction and recolonization processes were repeated for different degrees of habitat loss to obtain a robustness curve describing the fraction of colonised habitat patches in dependence on the fraction of lost habitat patches. Based on this robustness curve, we used the 'area under the curve' (AUC) as a measure to quantify metapopulation robustness. See [S1 File](#) and [51] for more details on the robustness simulation.

We compared the heuristics Greedy and Lazy Greedy with randomly adding links to the network and added 5 to 30 links in increments of 5. Baseline of these simulations was the robustness of the original habitat networks and we compared the increase of robustness originating from adding links using the different algorithms.

As the robustness simulations were specifically designed to evaluate the robustness of metapopulations on habitat networks, we considered the landscape-based habitat networks in this section only. We restrict the set of potential links \mathcal{E} to those unconnected pairs of patches that are at most 2500 m apart from each other:

$$\mathcal{E} := \{(u, v) \in V \times V \setminus E \mid \text{dist}_{\text{Eucl}}(u, v) < 2500 \text{ m}\}.$$

This represents the real world assumption, previously used in [38], that our generic species can

traverse at maximum 2500 m of open agricultural land with the dispersal distance reducing for areas with lower permeability such as urban area and forestry.

Results & discussion

Effect of used algorithms on the clustering coefficient

To compare the different algorithms, we added two links to the networks using each of the algorithms and calculated the difference in the clustering coefficient between the extended network and the original one. The optimal solution of adding two links to the landscape-based networks increases the clustering coefficient by 0.05 on average. For the sparse networks, the optimal solution resulted in a mean increase between 0.02 (regular networks) to 0.04 (small-world networks). All three dense network types showed no increase in the clustering coefficient after two links were added (Fig 5).

Our proposed algorithms Lazy Greedy and Greedy return results close to the optimal solution with Lazy Greedy being slightly worse. For both the Greedy and optimal solution the mean increase in the clustering coefficient was 0.030 over all network types and for the Lazy Greedy solution the mean increase was 0.029.

Adding two links randomly decreases the clustering coefficient for almost all landscape-based networks with a mean decrease of 0.15. The clustering coefficient for standard networks (both sparse and dense) remains unchanged by adding two links randomly.

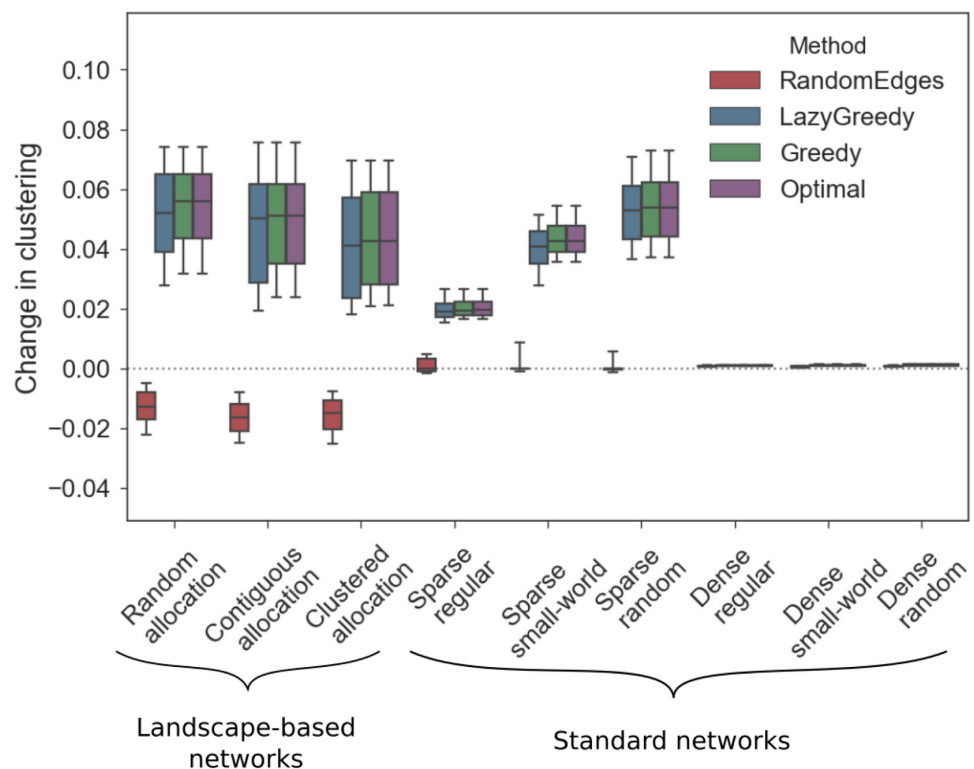


Fig 5. Greedy and Lazy Greedy algorithm applied to landscape-based and sparse networks lead to a higher increase in the clustering coefficient compared to randomly adding links. The horizontal axis shows the different network types, the vertical axis shows the change in clustering compared to the original network. The colour coding of the box-plots indicates the different algorithms.

<https://doi.org/10.1371/journal.pone.0240940.g005>

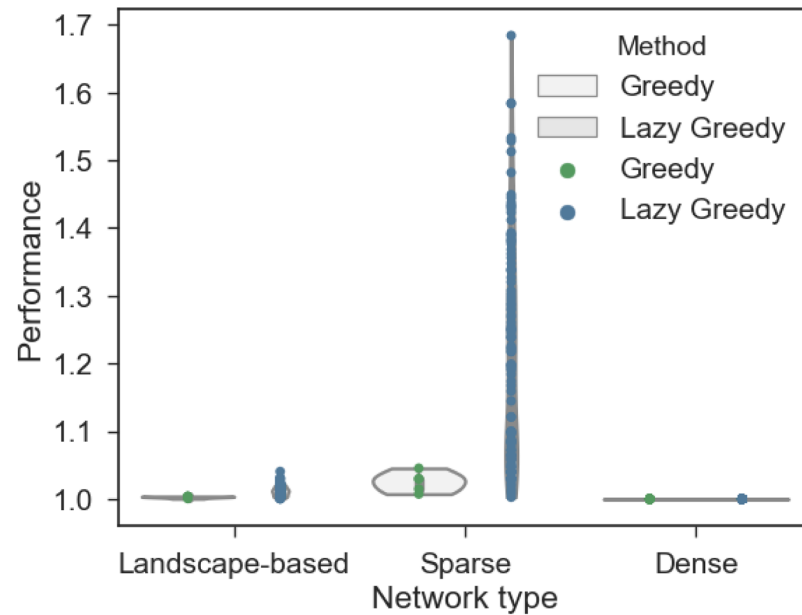


Fig 6. The Greedy algorithm returns an optimal solution in almost all cases. The vertical axis shows the quotient between optimal solution and solution of the heuristic. Only non optimal results are shown.

<https://doi.org/10.1371/journal.pone.0240940.g006>

For sparse networks, this implies that applying our heuristics to identify new links has a much larger impact on the clustering coefficient compared to the random approach. The same holds for habitat networks, which are usually sparse, leading to the conclusion that both the Greedy and Lazy Greedy heuristic are preferable to randomly adding links to a habitat network. For dense networks, however, adding two links has almost no impact on the clustering coefficient, independent from the considered method. As the majority of nodes in dense networks has a particularly high degree, the impact of an additional link decreases (see Eq 1), which explains the different results for dense networks. Furthermore, the clustering coefficient of dense networks is already rather high, leading to a smaller potential increase as well.

To quantify, how close the Greedy and Lazy Greedy algorithms approximate the optimal solution, we compared the clustering coefficient of the optimal solution with that produced by the Greedy and Lazy Greedy algorithm. The Greedy algorithm returned the optimal solution in 97.6% of the 2250 networks and the discrepancy between the clustering coefficient of the optimal solution and that produced by the Greedy algorithm was at most 3.8%. The Lazy Greedy algorithm, on the other hand, returned the optimal solution in only 76.0% of all networks and the discrepancy went up to 63.6%, increasing the clustering coefficient to 0.03 instead of 0.05 in that particular case (Fig 6).

Effect of used algorithms on robustness of habitat networks

The robustness of networks increased with the number of additional links, when the links were added with the help of the Greedy or Lazy Greedy algorithm. The correlation between the mean increase in robustness and number of additional links is $r = 0.8$ for the Greedy algorithm and $r = 0.76$ in case of the Lazy Greedy algorithm. If the links are added randomly, the increase in robustness is much smaller and the correlation between robustness and number of additional links drops to $r = 0.54$ (Fig 7).

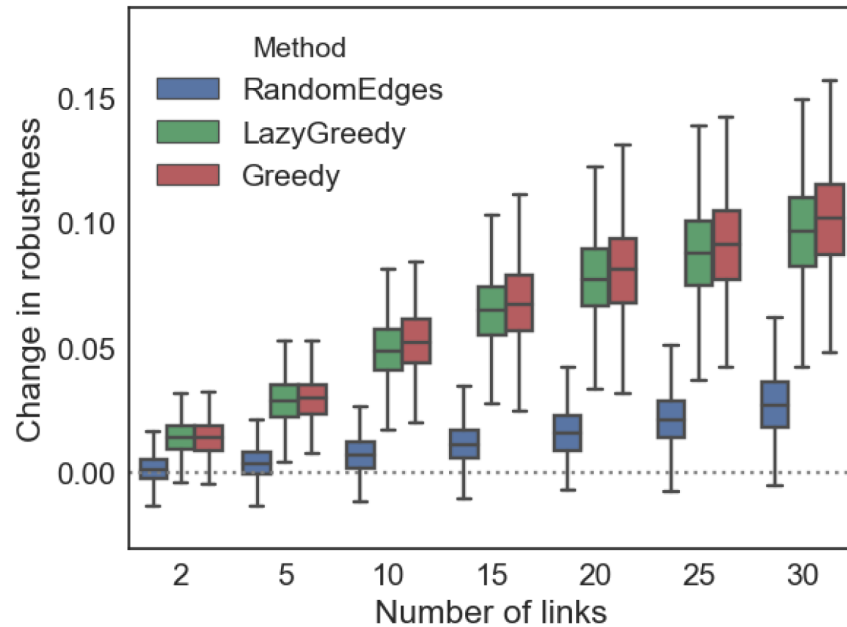


Fig 7. The robustness of networks increases with the number of additional links added using the Greedy or Lazy Greedy algorithm. The horizontal axis shows the number of links added to the network, the vertical axis the change in robustness. Colours indicate the algorithm used and each box shows the results over all landscape-based habitat networks.

<https://doi.org/10.1371/journal.pone.0240940.g007>

These results strongly suggest that using the presented algorithms to identify the links that should be added to a habitat network results in a much higher increase in robustness compared to randomly adding links.

Conclusion

We introduced two heuristics to maximise the clustering coefficient of a network by adding links. These methods work particularly well for sparse networks and yield a much higher increase in habitat network robustness compared to randomly adding links. Both the Greedy and Lazy Greedy heuristic return results close to the optimal solution for adding $m = 2$ links. While the Lazy Greedy algorithm is faster for large m , the Greedy algorithm returns results closer to the optimal solution and we suggest to apply the Greedy algorithm if possible.

Habitat connectivity is crucial for species survival, and habitat restoration efforts need to consider the robustness of habitat networks against habitat loss to increase connectivity and mitigate effects of future habitat loss. Our study shows that the location of links—and not only the number of links—has a large impact on metapopulation robustness, and presents a fast way to determine the best location for further links. It is the first study—to the best of our knowledge—that maximises the clustering coefficient of networks by adding links.

The heuristics presented here can be used to plan restoration efforts and increase habitat connectivity, as they provide locations in the habitat network that would lead to the largest increase of metapopulation robustness if they were connected. Simultaneously, our study shows that the location of links has a large impact on metapopulation robustness and thus emphasizes the importance of further mathematical models to improve habitat restoration strategies.

In summary, we presented two heuristics that identify which parts of a network need to be connected to obtain a higher network robustness. These heuristics work particularly well for

habitat networks and increase metapopulation robustness with increasing number of links added.

Supporting information

S1 File.
(PDF)

Author Contributions

Conceptualization: Henriette Heer, Stefan Ruzika.

Data curation: Henriette Heer.

Formal analysis: Henriette Heer.

Investigation: Henriette Heer.

Supervision: Ralf B. Schäfer, Stefan Ruzika.

Visualization: Henriette Heer.

Writing – original draft: Henriette Heer, Lucas Streib, Ralf B. Schäfer, Stefan Ruzika.

Writing – review & editing: Henriette Heer, Lucas Streib, Ralf B. Schäfer, Stefan Ruzika.

References

1. Sala OE, Chapin FS, Armesto JJ, Berlow E, Bloomfield J, Dirzo R, et al. Global biodiversity scenarios for the year 2100. *Science*. 2000; 287(5459):1770–1774. <https://doi.org/10.1126/science.287.5459.1770> PMID: 10710299
2. Fahrig L. Effects of habitat fragmentation on biodiversity. *Annual Review of Ecology, Evolution, and Systematics*. 2003; 34(1):487–515. <https://doi.org/10.1146/annurev.ecolsys.34.011802.132419>
3. Foley JA, DeFries R, Asner GP, Barford C, Bonan G, Carpenter SR, et al. Global consequences of land use. *Science*. 2005; 309(5734):570–574. <https://doi.org/10.1126/science.1111772> PMID: 16040698
4. Titeux N, Henle K, Mihoub JB, Regos A, Geijzendorffer IR, Cramer W, et al. Biodiversity scenarios neglect future land-use changes. *Global Change Biology*. 2016; 22(7):2505–2515. <https://doi.org/10.1111/gcb.13272> PMID: 26950650
5. Lechner AM, Sprod D, Carter O, Lefroy EC. Characterising landscape connectivity for conservation planning using a dispersal guild approach. *Landscape Ecology*. 2017; 32(1):99–113. <https://doi.org/10.1007/s10980-016-0431-5>
6. Turner MG, Ruscher CL. Changes in landscape patterns in Georgia, USA. *Landscape Ecology*. 1988; 1(4):241–251. <https://doi.org/10.1007/BF00157696>
7. Kindlmann P, Burel F. Connectivity measures: A review. *Landscape Ecology*. 2008; 23(8):879–890.
8. Levins R. Some demographic and genetic consequences of environmental heterogeneity for biological control. *American Entomologist*. 1969; 15(3):237–240.
9. Hanski I. Metapopulation dynamics. *Nature*. 1998; 396(6706):41–49. <https://doi.org/10.1038/23876>
10. Perry GL, Lee F. How does temporal variation in habitat connectivity influence metapopulation dynamics? *Oikos*. 2019; 128(9):1277–1286. <https://doi.org/10.1111/oik.06052>
11. Fahrig L, Merriam G. Conservation of fragmented populations. *Conservation Biology*. 1994; 8(1):50–59. <https://doi.org/10.1046/j.1523-1739.1994.08010050.x>
12. Hanski I. *Metapopulation Ecology*. Oxford University Press; 1999.
13. Urban MC. Accelerating extinction risk from climate change. *Science*. 2015; 348(6234):571–573. <https://doi.org/10.1126/science.aaa4984> PMID: 25931559
14. Petsas P, Tsavdaridou AI, Mazaris AD. A multispecies approach for assessing landscape connectivity in data-poor regions. *Landscape Ecology*. 2020; p. 1–16.
15. Urban DL, Keitt T. Landscape connectivity: A graph-theoretic perspective. *Ecology*. 2001; 82(5):1205–1218.

16. Dale M, Fortin MJ. From graphs to spatial graphs. *Annual Review of Ecology, Evolution, and Systematics*. 2010; 41(1):21–38. <https://doi.org/10.1146/annurev-ecolsys-102209-144718>
17. Erős T, Olden JD, Schick RS, Schmera D, Fortin MJ. Characterizing connectivity relationships in freshwaters using patch-based graphs. *Landscape Ecology*. 2012; 27(2):303–317. <https://doi.org/10.1007/s10980-011-9659-2>
18. DeAngelis DL, Yurek S. Spatially explicit modeling in ecology: A review. *Ecosystems*. 2017; 20(2):284–300. <https://doi.org/10.1007/s10021-016-0066-z>
19. Urban DL, Minor ES, Treml EA, Schick RS. Graph models of habitat mosaics. *Ecology Letters*. 2009; 12(3):260–273. <https://doi.org/10.1111/j.1461-0248.2008.01271.x> PMID: 19161432
20. Galpern P, Manseau M, Fall A. Patch-based graphs of landscape connectivity: a guide to construction, analysis and application for conservation. *Biological Conservation*. 2011; 144(1):44–55. <https://doi.org/10.1016/j.biocon.2010.09.002>
21. Martensen AC, Saura S, Fortin MJ. Spatio-temporal connectivity: Assessing the amount of reachable habitat in dynamic landscapes. *Methods in Ecology and Evolution*. 2017; 8(10):1253–1264. <https://doi.org/10.1111/2041-210X.12799>
22. Kun Á, Oborny B, Dieckmann U. Five main phases of landscape degradation revealed by a dynamic mesoscale model analysing the splitting, shrinking, and disappearing of habitat patches. *Scientific Reports*. 2019; 9(1):1–11. <https://doi.org/10.1038/s41598-019-47497-7>
23. Prima MC, Duchesne T, Fortin A, Rivest LP, Drapeau P, St-Laurent MH, et al. A landscape experiment of spatial network robustness and space-use reorganisation following habitat fragmentation. *Functional Ecology*. 2019; 33(9):1663–1673. <https://doi.org/10.1111/1365-2435.13380>
24. Mazaris AD, Papanikolaou AD, Barbet-Massin M, Kallimanis AS, Jiguet F, Schmeller DS, et al. Evaluating the connectivity of a protected areas' network under the prism of global change: the efficiency of the European Natura 2000 network for four birds of prey. *PLoS One*. 2013; 8(3):e59640. <https://doi.org/10.1371/journal.pone.0059640> PMID: 23527237
25. Dilts TE, Weisberg PJ, Leitner P, Matocq MD, Inman RD, Nussear KE, et al. Multiscale connectivity and graph theory highlight critical areas for conservation under climate change. *Ecological Applications*. 2016; 26(4):1223–1237. <https://doi.org/10.1890/15-0925>
26. Pietsch M. Contribution of connectivity metrics to the assessment of biodiversity—Some methodological considerations to improve landscape planning. *Ecological Indicators*. 2018; 94:116–127. <https://doi.org/10.1016/j.ecolind.2017.05.052>
27. Albert R, Jeong H, Barabasi AL. Error and attack tolerance of complex networks. *Nature*. 2001; 409(6819):542–542. <https://doi.org/10.1038/35054111>
28. Solé RV, Montoya JM. Complexity and fragility in ecological networks. *Proceedings of the Royal Society of London Series B: Biological Sciences*. 2001; 268(1480):2039–2045. <https://doi.org/10.1098/rspb.2001.1767> PMID: 11571051
29. Rosas-Casals M, Valverde S, Solé RV. Topological vulnerability of the European power grid under errors and attacks. *International Journal of Bifurcation and Chaos*. 2007; 17(07):2465–2475. <https://doi.org/10.1142/S0218127407018531>
30. Berche B, Von Ferber C, Holovatch T, Holovatch Y. Resilience of public transport networks against attacks. *The European Physical Journal B*. 2009; 71(1):125–137. <https://doi.org/10.1140/epjb/e2009-00291-3>
31. Cuadra L, Salcedo-Sanz S, Del Ser J, Jiménez-Fernández S, Geem Z. A critical review of robustness in power grids using complex networks concepts. *Energies*. 2015; 8(9):9211–9265. <https://doi.org/10.3390/en8099211>
32. Ash J, Newth D. Optimizing complex networks for resilience against cascading failure. *Physica A: Statistical Mechanics and its Applications*. 2007; 380:673–683. <https://doi.org/10.1016/j.physa.2006.12.058>
33. Almpnidou V, Mazaris AD, Mertzanis Y, Avraam I, Antoniou I, Pantis JD, et al. Providing insights on habitat connectivity for male brown bears: A combination of habitat suitability and landscape graph-based models. *Ecological Modelling*. 2014; 286:37–44. <https://doi.org/10.1016/j.ecolmodel.2014.04.024>
34. Fox RJ, Bellwood DR. Herbivores in a small world: network theory highlights vulnerability in the function of herbivory on coral reefs. *Functional Ecology*. 2014; 28(3):642–651. <https://doi.org/10.1111/1365-2435.12190>
35. Watts DJ, Strogatz SH. Collective dynamics of 'small-world' networks. *Nature*. 1998; 393(6684):440–442. <https://doi.org/10.1038/30918> PMID: 9623998
36. Krumke SO, Noltemeier H. *Graphentheoretische Konzepte und Algorithmen*. Springer-Verlag; 2009.
37. Newman MEJ. *Networks: An Introduction*. Oxford University Press; 2010.

38. Streib L, Kattwinkel M, Heer H, Ruzika S, Schäfer RB. How does habitat connectivity influence the colonization success of a hemimetabolous aquatic insect?—A modeling approach. *Ecological Modelling*. 2020; 416(108909).
39. Green O, Bader DA. Faster clustering coefficient using vertex covers. In: 2013 International Conference on Social Computing. IEEE; 2013. p. 321–330.
40. Keller D, Van Strien MJ, Holderegger R. Do landscape barriers affect functional connectivity of populations of an endangered damselfly? *Freshwater Biology*. 2012; 57(7):1373–1384. <https://doi.org/10.1111/j.1365-2427.2012.02797.x>
41. Ottmann T, Widmayer P. *Algorithmen und Datenstrukturen*. Springer; 2017.
42. Hoare CA. Quicksort. *The Computer Journal*. 1962; 5(1):10–16. <https://doi.org/10.1093/comjnl/5.1.10>
43. Wasserman S, Faust K. *Social network analysis: Methods and applications*. vol. 8. Cambridge University Press; 1994.
44. Barabasi AL, Oltvai ZN. Network biology: Understanding the cell's functional organization. *Nature reviews genetics*. 2004; 5(2):101–113. <https://doi.org/10.1038/nrg1272> PMID: 14735121
45. Holland MD, Hastings A. Strong effect of dispersal network structure on ecological dynamics. *Nature*. 2008; 456(7223):792–794. <https://doi.org/10.1038/nature07395> PMID: 18931656
46. Thompson PL, Rayfield B, Gonzalez A. Robustness of the spatial insurance effects of biodiversity to habitat loss. *Evolutionary Ecology Research*. 2015; 16(6):445–460.
47. Hagberg A, Swart P, Chult DS. Exploring network structure, dynamics, and function using NetworkX. Technical Report LA-UR-08-5495, Los Alamos National Laboratory (LANL), Los Alamos, NM, USA. 2008;.
48. Erdős P, Rényi A. On the evolution of random graphs. *Publications of the Mathematical Institute of the Hungarian Academy of Sciences*. 1960; 5(1):17–60.
49. Boccaletti S, Latora V, Moreno Y, Chavez M, Hwang DU. Complex networks: Structure and dynamics. *Physics Reports*. 2006; 424(4-5):175–308. <https://doi.org/10.1016/j.physrep.2005.10.009>
50. Newman MEJ, Watts DJ. Renormalization group analysis of the small-world network model. *Physics Letters A*. 1999; 263(4-6):341–346. [https://doi.org/10.1016/S0375-9601\(99\)00757-4](https://doi.org/10.1016/S0375-9601(99)00757-4)
51. Heer H, Streib L, Schäfer RB, Dieckmann U. Indicators for Assessing the Robustness of Metapopulations against Habitat Loss. *Ecological Indicators*, in press.