

Mathematical models of species dispersal and the resilience of metapopulations against habitat loss

by

HENRIETTE HEER
from Steinheim (Westf.)

Accepted Dissertation thesis for the partial fulfilment of the requirements for a
Doctor of Natural Sciences
Fachbereich 7: Natur- und Umweltwissenschaften
Universität Koblenz-Landau

Thesis examiners:

Prof. Dr. Ralf B. Schäfer, University of Koblenz-Landau, 1st examiner
Prof. Dr. Stefan Ruzika, Technical University of Kaiserslautern, 2nd examiner

Date of the oral examination: February 26, 2021

Acknowledgements

I would like to thank everyone who supported me during my PhD studies.

First and foremost, I want to thank my supervisor Prof. Dr. Ralf B. Schäfer. He took me on as a student after I was already one year into my PhD and has always been extremely supportive throughout this period. I am especially grateful for his guidance and appreciation of his openness and interest in a wide variety of research topics.

I would also like to thank Prof. Dr. Stefan Ruzika, who was the one to open the door for me and to accept me into the PhD programme in the first place. It was because of Stefan that I was introduced to the interesting and growing field of ecological modelling. He has always been very supportive and enthusiastic to discuss mathematical problems which is especially important in interdisciplinary research and his input helped shape my research throughout the project.

Many thanks to the Quantitative Landscape Ecology working group who have been very welcoming, approachable and always happy to help. In particular, thanks to my colleagues Stefan, Andreas and Anke for lightening up my life outside university with Volleyball games, trips to the climbing centre and nights out in Brennan's. It's been a pleasure guys, thanks for all the fun times! I especially want to thank Lucas Streib for his patience and helpfulness and for putting up with me as an office mate for the last two years.

Thanks also goes to the former working group of Stefan Ruzika at the Mathematical Institute in Koblenz, who supported me during the start of my PhD. In addition, I would like to thank the Optimisation Research Group at the University of Kaiserslautern for their hospitality and enthusiasm to discuss a wide variety of research problems during my visits.

Thanks also goes to Prof. Dr. Ulf Dieckmann for his support. His quick responses and useful comments were immensely helpful.

Finally, I want to thank my family and friends, for distracting me when I needed it and for their understanding during the more stressful times. Without the continuous support of my boyfriend, Luke Coburn, this thesis would not have been possible. Thank you for all your support and encouragement during this stressful time and for all of the interesting and helpful discussions about my research topic.

Summary

Habitat loss and fragmentation due to climate and land-use change are among the biggest threats to biodiversity, as the survival of species relies on suitable habitat area and the possibility to disperse between different patches of habitat. To predict and mitigate the effects of habitat loss, a better understanding of species dispersal is needed. Graph theory provides powerful tools to model metapopulations in changing landscapes with the help of habitat networks, where nodes represent habitat patches and links indicate the possible dispersal pathways between patches.

This thesis adapts tools from graph theory and optimisation to study species dispersal on habitat networks as well as the structure of habitat networks and the effects of habitat loss. In **chapter 1**, I will give an introduction to the thesis and the different topics presented in this thesis. **Chapter 2** will then give a brief summary of tools used in the thesis.

In **chapter 3**, I present our model on possible range shifts for a generic species. Based on a graph-based dispersal model for a generic aquatic invertebrate with a terrestrial life stage, we developed an optimisation model that models dispersal directed to predefined habitat patches and yields a minimum time until these patches are colonised with respect to the given landscape structure and species dispersal capabilities. We created a time-expanded network based on the original habitat network and solved a mixed integer program to obtain the minimum colonisation time. The results provide maximum possible range shifts, and can be used to estimate how fast newly formed habitat patches can be colonised. Although being specific for this simulation model, the general idea of deriving a surrogate can in principle be adapted to other simulation models.

Next, in **chapter 4**, I present our model to evaluate the robustness of metapopulations. Based on a variety of habitat networks and different generic species characterised by their dispersal traits and habitat demands, we modeled the permanent loss of habitat patches and subsequent metapopulation dynamics. The results show that species with short dispersal ranges and high local-extinction risks are particularly vulnerable to the loss of habitat across all types of networks. On this basis, we then investigated how well different graph-theoretic metrics of habitat networks can serve as indicators of metapopulation robustness against habitat loss. We identified the clustering coefficient of a network as the only good proxy for metapopulation robustness across all types of species, networks, and habitat loss scenarios.

Finally, in **chapter 5**, I utilise the results obtained in chapter 4 to identify the areas in a network that should be improved in terms of restoration to maximise the metapopulation robustness under limited resources. More specifically, we exploit our findings that a network's clustering coefficient is a good indicator for metapopulation robustness and develop two heuristics, a Greedy algorithm and a deduced Lazy Greedy algorithm, that aim at maximising the clustering coefficient of a network. Both algorithms can be applied to any network and are not specific to habitat networks only.

In **chapter 6**, I will summarize the main findings of this thesis, discuss their limitations and give an outlook of future research topics.

Overall this thesis develops frameworks to study the behaviour of habitat networks and introduces mathematical tools to ecology and thus narrows the gap between mathematics and ecology. While all models in this thesis were developed with a focus on aquatic invertebrates, they can easily be adapted to other metapopulations.

Contents

Acknowledgements	i
Summary	ii
1 Introduction	1
1.1 Ecology	1
1.2 Graph theory	2
1.3 Application to ecology	3
1.4 Objectives and outline of the thesis	4
2 Preliminaries	6
2.1 Notation and definitions	6
2.2 Landscape-based habitat networks	7
2.3 Standard networks	8
3 Optimisation model of dispersal simulations on a dendritic habitat network	10
3.1 Introduction	10
3.2 Methods	12
3.2.1 Simulation Model	12
3.2.2 Optimisation Model	14
Process Overview	14
Habitat Network	15
Mixed Integer Program	15
Time Horizon	18
3.2.3 Initialisation	19
3.2.4 Analysis of Models	19
3.3 Results & Discussion	20
3.3.1 Analysis of Model Results	20
3.3.2 Model Run Times	21
3.3.3 Example	22
3.3.4 Challenges & Outlook	23
4 Indicators for assessing the robustness of metapopulations against habitat loss	26
4.1 Introduction	27
4.2 Methods	29
4.2.1 Structure of analyses	29
4.2.2 Habitat networks	30
4.2.3 Habitat loss	31
4.2.4 Metapopulation dynamics	32
4.2.5 Statistical analyses	33
4.3 Results	35

4.3.1	Landscape-based habitat networks respond to habitat loss qualitatively differently than standard networks	35
4.3.2	Species with short dispersal ranges and high local-extinction risks are particularly vulnerable to habitat loss	36
4.3.3	The power of network metrics to predict robustness depends on the types of species, networks, and habitat loss	37
	Sensitive species	37
	Landscape-based habitat networks	37
	Contagious habitat loss	38
4.4	Discussion	39
4.5	Conclusions	42
5	Maximising the clustering coefficient of networks and the effects on habitat network robustness	48
5.1	Introduction	49
5.2	Methods	50
5.3	Results & Discussion	59
6	Discussion	67
6.1	Limitations & Outlook	68
	Bibliography	70
7	Publications	78
8	Author's contributions	79
	Curriculum vitae	80
	Declaration of Authorship	82

List of Figures

2.1	Examples of landscape-based habitat networks with (a) random, (b) clustered, and (c) contiguous habitat allocation. Blue dots represent habitat patches, light-blue lines depict the underlying stream landscape, and black lines indicate dispersal pathways. The width of these lines represents the dispersal costs with thick lines indicating lower costs and thin lines indicating higher costs.	8
2.2	Examples of standard networks with (a) random, (b) regular, (c) small-world, and (d) scale-free structure. Blue dots represent habitat patches, light-blue lines depict the underlying stream landscape, and black lines indicate dispersal pathways. The width of these lines represents the dispersal costs with thick lines indicating lower costs and thin lines indicating higher costs.	9
3.1	Flowchart of the simulation model. A detailed description of each subprocess (double rectangles) can be found in supplementary Information S1. (a) creation of habitat network, which is then used as input for the dispersal simulation depicted in (b)	14
3.2	(a) Habitat network (b) Corresponding time expanded network with time horizon 3 and source habitat A. Connections in (b) representing a specific connection in (a) use the same colour. Black connections in (b) are artificial connections indicating remaining in the same habitat patch from one time step to the next. Q represents the super source. . .	16
3.3	Relationship between the time steps found by the optimisation model and the cumulative dispersal cost from the nearest (i.e. most cost efficient) connected initial source habitat to the destination habitat of the corresponding model run. Each black mark in the plot corresponds to one of the 2,500 model runs. The grey line is the linear regression. . . .	20
3.4	Correlation between time steps needed for colonisation of destination habitat of simulation and optimisation. Each black mark corresponds to a single run of the optimisation model and the corresponding result from the simulation. The grey line is the linear regression.	21
3.5	Application of optimisation model on a habitat network. The red nodes represent initially colonised habitat patches, the yellow star is the corresponding (uncolonised) destination habitat. The thick, purple lines show the dispersal paths displayed by the optimisation model	23

- 4.1 Schematic overview of the structure of analyses in this study. We consider species with different traits inhabiting different habitat networks experiencing different habitat loss (left panels). For the resultant wide variety of ecological settings, we study metapopulation dynamics involving the local extinction and recolonization of habitat patches, for different degrees of habitat loss (middle panel). On this basis, we obtain a measure of metapopulation robustness, as the ‘area under the curve’ (AUC) of the dependence of the fraction of colonized habitat patches on the fraction of lost habitat patches (right panel, top-left part) and examine how well various graph-theoretic metrics of the investigated habitat networks (right panel, bottom-right part) can serve as reliable indicators of metapopulation robustness (right panel, central part). 30
- 4.2 Examples of the analysed seven types of habitat networks. (a-c) Landscape-based networks, with (a) random, (b) clustered, and (c) contiguous habitat allocation Streib et al. (2020). (d-g) Standard networks with (d) random, (e) regular, (f) small-world, and (g) scale-free structure. Blue dots represent habitat patches, light-blue lines depict the underlying stream landscape, and black lines indicate dispersal pathways. The width of these lines represents the dispersal costs with thick lines indicating lower costs and thin lines indicating higher costs. 32
- 4.3 Sensitive species, characterized by low metapopulation robustness to habitat loss, have short dispersal ranges and high local-extinction risks. (a) Dependence of metapopulation robustness on the dispersal ranges (vertical axes, parameter σ varied from 2 to 9 in steps of 1) and local-extinction risks (horizontal axis, parameter a varied from 2 to 9 in steps of 1) of species for seven different network types (columns) and three types of habitat loss (rows). Each cell shows the metapopulation robustness averaged over 10 replications (see Methods), with light-red colours indicating high robustness and dark-red colours indicating low robustness. (b) Corresponding distributions of metapopulation robustness according to network types (columns) and types of habitat loss (rows). The central line of each box indicates the median, the lower and upper edges of each box indicate the interquartile range, and the whiskers indicate the 10% and 90% quantiles. 36
- 4.4 When considering all types of species, networks, and habitat loss simultaneously, only the clustering coefficient can serve as a reliable indicator of metapopulation robustness. A total of 29 network metrics are shown in the rows, and the three types of habitat loss are shown in the panels. The colour of each cell indicates the strength of the relationship between metapopulation robustness and the considered network metric, measured as the standardized regression coefficient, averaged over all network types (this average strength of the relationship is set to 0 when the sign of the strength of the relationship is not consistent across network types). Within each row of each panel, the species-specific parameters governing local-extinction risks and dispersal ranges, respectively, are changing across the columns in steps of 1 between 2 and 9 first for a and then for σ . The last column shows averages over all cells of a row. 37

4.5	When considering sensitive species, the average clique size, the redundancy, the average degree, the connectance, the clustering coefficient, and the average closeness centrality are the best indicators of metapopulation robustness. All figure elements as in Fig. 4.	38
4.6	When considering landscape-based networks, the average clique size, the beta coefficient, the clustering coefficient, the redundancy, and the cyclomatic number are the best indicators of metapopulation robustness. All figure elements as in Fig. 4.4.	38
4.7	When considering contagious habitat loss, the clustering coefficient is the best indicator of metapopulation robustness. For sensitive species, the redundancy, the average clique size, the connectance, and the average degree are further appropriate indicators. All figure elements as in Fig. 4.4.	39
4.8	Variance in metapopulation robustness explained by the network type, dispersal range, and local-extinction risk. The total explained variance is high, reaching 85% or more for all types of habitat loss. The network type has a particularly strong impact on the robustness of metapopulations to contagious habitat loss.	40
4.9	Stream landscape in the German federal state of Rhineland-Palatinate used for deriving the landscape-based habitat networks analysed in the present study. Streams locations, indicated in light-blue, are obtained from the database 'Gewässernetz (gesamt)' (Rhineland-Palatinate Ministry of the Interior and for Sports, 2020a). The background shows a satellite image obtained from the web map service 'Luftbild RP Basisdienst' (Rhineland-Palatinate Ministry of the Interior and for Sports, 2020b). The grid lines indicate the division of the total 50 km × 50 km landscape into 25 landscape tiles each measuring 10 km × 10 km. The map uses the coordinate reference system 'ETRS89 / UTM zone 32N' with EPSG code 25832. Longitudinal positions are shown horizontally along the top edge and latitudinal positions are shown vertically along the left edge, both measured in metres. The overview map in the top right shows the location of the stream landscape (light-blue rectangle) within the state of Rhineland-Palatinate (satellite image) within Germany. The tile surrounded in red corresponds to the maps shown in Fig. 4.2a-c.	44
4.10	Full distribution of the number of nodes for the different network types. Only one distribution is shown for the three types of landscape-based habitat networks as the number of nodes in these networks does not depend on the types of habitat allocation, but only on the underlying stream landscapes: the distributions of the number of nodes are thus identical for all three types of landscape-based habitat networks. The standard networks are created using the same node distribution.	47

- 5.1 Example network to illustrate notation. $G = (V, E)$ with $n = 8$ nodes, 10 links, $V = \{a, b, c, d, e, u, v, w\}$, and $E = \{(a, b), (a, w), (a, c), (b, u), (b, w), (u, e), (w, c), (w, d), (e, v)\}$. 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 $C_{G'} = \frac{1}{8} \cdot (\frac{1}{3} + \frac{1}{3} + \frac{1}{3} + \frac{1}{3} + \frac{2}{3} + 1 + 0 + 1) = \frac{1}{8} \cdot 4 = \frac{1}{2}$ 52
- 5.2 Varying effects of adding a link to a network on the clustering coefficient. (a) Original network with clustering coefficient $C_G = 0.38$. (b) Network after connecting two nodes with no common neighbours, $C_{G+e_1} = 0.27$. (c) Network after connecting two nodes with a common neighbour, $C_{G+e_2} = 0.7$ 53
- 5.3 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 5.1). 56
- 5.4 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 58
- 5.5 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. 60
- 5.6 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. 61
- 5.7 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. 61

List of Tables

3.1	Dispersal costs and ratio per land cover class. The percentages refer to the underlying Neutral Landscape Model (NLM) that was used to create the landscape (Suppl. Inf. S1). A real stream network on a finer scale is added to the landscape and cells intersecting with the river network are declared as ‘aquatic’.	13
4.1	Graph-theoretic metrics investigated in this study. The listed quantities are network-level metrics, except for the node-level metrics of betweenness centrality, closeness centrality, and degree, for which our analysis considers the network-level metrics given by the average, minimum, maximum, and range of the node-level metrics.	35
4.2	Distribution of the number of nodes for each network type.	46
4.3	Distribution of the number of links for each network type.	46
4.4	Distribution of the dispersal costs for each network type.	46
5.1	Parameters to create standard networks. d is the degree of each node, p denotes the percentage of links present in the network and k is the degree of each node in the small-world network before rewiring.	66

Chapter 1

Introduction

In this thesis I present the research that I undertook during my PhD which was concerned with the degradation of ecosystems that results from climate change. I focused on metapopulations of aquatic invertebrates. In particular I looked at species dispersal and investigated this phenomenon using habitat networks. I examined how these habitat networks were structured, approximated these networks using graphs and then evaluated their robustness and developed a theoretical method to increase the robustness by adding new links to these virtual networks.

1.1 Ecology

As we progress into the 21st century, we are finding that some of the greatest challenges that humanity faces, and will face into the future, are degrading ecosystems, habitat loss and biodiversity decline. These problems have been caused or have been exasperated by climate change (Butchart et al., 2010; IPCC, 2019; Reid et al., 2005; Urban, 2015). Effects of climate change such as changes in species' abundance, distribution, and phenology, as well as an increase in extreme weather events are now visible in all ecosystems (Scheffers et al., 2016; Parmesan and Yohe, 2003; Parmesan, 2006) Furthermore, climate and land-use change have led to habitat degradation and fragmentation as well as range shifts as species abundance drifts northward or westward in the northern hemisphere (Alahuhta et al., 2019; Berg et al., 2010; Kuemmerlen et al., 2015; Scheffers et al., 2016; Parmesan and Yohe, 2003).

The decline of habitat area and connectivity due to the fragmentation of landscapes are among the biggest threats to biodiversity, as the survival of species relies on suitable habitat area as well as the individual's ability to disperse between different patches of habitat (Bruggeman et al., 2010; Fahrig, 2003; Foley et al., 2005; Urban, 2015). This dispersal is crucial for species survival, as it facilitates interactions such as the exchange of genes between different populations and thus allows for the existence of metapopulations – a “population of populations” (Levins, 1969; Hanski, 1998; Perry and Lee, 2019). To anticipate and mitigate the effects of climate and land-use change on biodiversity, a better understanding of species dispersal is thus necessary.

Field, lab, and mesocosm studies are important methods to investigate species traits and behaviours. However, extensive studies are time consuming and expensive. Simultaneously, methodical limitations complicate the reliable derivation of species specific data such as parameters governing dispersal processes or population growth (Didham et al., 2012; Tonkin et al., 2018). For these reasons, models are needed to gain a better understanding of species behaviour and to predict the response of metapopulations to a changing environment (Clark et al., 2001; Erős et al., 2012a; Evans et al., 2012; Reid et al., 2005).

Spatially explicit graph-based analysis has become popular in conservation biology and landscape ecology in the last decades (Erős and Lowe, 2019; Galpern et al., 2011; Urban et al., 2009; Zetterberg et al., 2010) and turned into one of the best approaches for modelling the dispersal of organisms (DeAngelis and Yurek, 2017; Erős and Lowe, 2019; Heino et al., 2017; Minor and Urban, 2007).

1.2 Graph theory

Graph theory provides powerful tools to model metapopulations in form of habitat networks. It combines ideas from a wide variety of fields such as mathematics, computer science, and social sciences to study these networks, which are often also called graphs (Newman, 2010; Cohen and Havlin, 2010; Barthélemy, 2014). In general, networks model the relationships within a given system by representing the system components as nodes (also called vertices) and their relationship as links (edges) that connect these nodes and allow us to build quantitative descriptions of the complex relationships that exist in the given system. Examples of networks are transportation networks such as the metro or airline networks, the world wide web, the electrical power grid, and social networks that model friendships or co-authorships. Ecological networks include predator-prey networks, food-webs and habitat networks of metapopulations or metacommunities (Newman et al., 2002; Cohen and Havlin, 2010; Barthélemy, 2014).

A special group of networks are the so-called *spatial networks* (Barthélemy, 2014). In contrast to other networks, the nodes of spatial networks are embedded in space and the probability of two nodes being connected decreases with distance (Gastner and Newman, 2006; Dale and Fortin, 2010). Naturally, habitat networks are spatial networks, as real-world habitat patches are located in a three-dimensional space that can often be approximated by a two-dimensional landscape, with patches at short distance having a higher probability of being connected than more distant patches. Food-webs on the other hand are independent of space.

The advantages of networks are numerous. Networks are particularly flexible as nodes can represent multiple properties, e.g. single individuals, whole populations or habitat patches (Galpern et al., 2011; Calabrese and Fagan, 2004). Networks allow for complex computations while only requiring relatively little data (Erős et al., 2012b; Rayfield et al., 2011). Simultaneously, spatially explicit data obtained from geographic information systems (GIS) can be added easily, allowing for a simple way to combine spatially explicit data with information on species specific dispersal characteristics. As networks are analysed in their abstract form, graph-theoretic tools from all disciplines can be applied to almost any system represented as a network. The resultant variety of existing tools can readily be applied in ecology (Urban et al., 2009; Newman, 2010) and can easily be used to evaluate different conservation scenarios or assess the impacts of a changing global environment.

Graph-theoretic tools can be used to analyse a network locally, by evaluating how central a node is in the network using a variety of centrality measures, or by identifying how clustered a node's neighbourhood is with the help of the local clustering coefficient. Other tools are targeted at the network structure and analyse the network as a whole (Boccaletti et al., 2006). Also, networks are often dynamic, and nodes and links can appear or disappear perhaps only temporarily. Depending on the affected region and the network itself, this can greatly affect the overall network structure, giving rise to an important question of how to determine the robustness of a network against node failure.

The robustness of a network quantifies the network's durability to the loss of nodes and links (Albert et al., 2001a) and has been subject to many studies in various areas such as technology, transportation, and trade (Callaway et al., 2000; Cohen et al., 2000; Albert et al., 2001a; Cuadra et al., 2015; Gephart et al., 2016). Network robustness has been analysed on various networks such as transportation networks, power grids, and food webs (Solé and Montoya, 2001; Rosas-Casals et al., 2007; Berche et al., 2009). As network robustness depends on a variety of factors and thus is complex and difficult to evaluate, it is important to find appropriate proxies for the robustness. Many related studies identified a network's clustering coefficient as a good proxy for robustness (Ash and Newth, 2007; Almpnidou et al., 2014; Fox and Bellwood, 2014; Prima et al., 2019). The clustering coefficient was proposed by Watts and Strogatz (1998). A node's clustering coefficient measures the relative density of links in its neighbourhood and is thus an indication of how close its neighbourhood is to being complete. A network's clustering coefficient is then defined as the average clustering coefficient of its nodes.

Another approach to account for the dynamic nature of networks is the use of *time expanded networks*. The time expansion of a network consists of one copy of the original network's nodes per time layer and directed connections between the copies of two nodes in consecutive layers, if the original two nodes are connected (Hamacher and Klamroth, 2006; Skutella, 2009). Time expanded networks can be used to model movement, such as traffic, the flow of goods or dispersing individuals, on networks over time (Baumann and Skutella, 2009; Ho et al., 2014; Köhler et al., 2009; Stasko et al., 2016).

Time expanded networks, and graph theory in general, are strongly connected to optimisation in various disciplines such as finance, logistics, engineering and transportation (Bondy and Murty, 1976; Hamacher and Klamroth, 2006). Optimisation problems arise in all quantitative disciplines from computer science and engineering to operations research and economics, and the development of solution methods has been of interest in mathematics for centuries. In the simplest case, an optimisation problem consists of maximising or minimising a real function by systematically choosing input values from within an allowed set and computing the value of the function.

1.3 Application to ecology

In ecology, graph theoretic techniques were mainly used for studying food webs and plant-animal mutualistic interactions (Landi et al., 2018; Urban and Keitt, 2001; Alba, 1973; Montoya and Solé, 2002; Erős et al., 2012b). However, they are also gaining more importance in landscape ecology and the use of habitat networks representing metapopulations and metacommunities is becoming more and more prevalent (Fortuna et al., 2006; Dale and Fortin, 2010; Ortiz-Rodríguez et al., 2019). In this context, landscapes are seen as habitat networks, with habitat patches represented as nodes and possible dispersal pathways represented as links.

Habitat networks provide a useful framework to study the dispersal of species. Furthermore, combined with techniques from the field of optimisation, we can obtain dispersal models that determine the fastest way for species to disperse and reach a specific habitable area. As opposed to common dispersal simulations, where the dispersal is assumed to be undirected, the optimisation approach models species dispersal directed to a specific area. The optimisation approach needs less information about species specific dispersal such as which habitats are preferably colonised.

This approach is exceptionally advantageous in ecology, as data is usually scarce and assumptions are often error-prone. On the other hand, applying optimisation to dispersal simulations can only provide general bounds to the questions usually answered by dispersal simulations, such as which patches are preferably colonised and how much time it takes for a specific patch to become populated. Although networks are now common in dispersal simulations and other ecological models, the potential of optimisation on graphs has not yet been explored in dispersal models (OBJECTIVE 1).

With the help of habitat networks, the loss of habitat can easily be represented by removing nodes and by removing links from the network we can simulate reduced connectivity. Habitat networks can thus be used to investigate the robustness of a metapopulation, represented by the habitat network, against climate and land-use change effects such as the loss of habitable area or decreased landscape permeability. Previous studies of the robustness of ecological networks have predominantly focused on food webs and mutualistic networks (Burgos et al., 2007; Evans et al., 2013). In the context of habitat networks, graph theory has been applied to study the effect of network connectivity on the dispersal of species (Calabrese and Fagan, 2004; Estrada and Bodin, 2008; Rayfield et al., 2011; Saunders et al., 2016; Upadhyay et al., 2017; Grech et al., 2018). These studies have usually modelled different species with varying traits such as dispersal ranges on a single underlying landscape. Furthermore, most studies have modelled static scenarios and ignored any metapopulation dynamics. Metapopulations, however, are highly dynamic and constantly adapt to changed conditions, for example after habitat loss. Therefore, analyses including the dynamic structure would appear to be more appropriate (Martensen et al., 2017; Kun et al., 2019; Prima et al., 2019; Shen et al., 2019). Recently, several studies of metapopulation robustness have utilized such an approach. Shen et al. (2019) simulated dispersal on regular networks, in which every habitat patch has the same number of neighbours. They observed that metapopulations become more resistant with an increasing number of connections.

There are different strategies to protect species from local extinctions due to habitat loss. Naturally, we would like to aim to protect habitable area and therefore slow down or even avoid the loss of habitat. However, as this is difficult to achieve, we should also focus on increasing habitat connectivity to soften the effects of habitat loss (Fahrig and Merriam, 1994; Hanski, 1999). Habitat connectivity can be increased by increasing the landscape permeability with the help of dispersal corridors. However, financial resources for designing and implementing conservation measures are usually limited and a quantitative assessment of habitat connectivity in consideration of future habitat loss is necessary to prioritise conservation efforts (Cowling et al., 1999) (OBJECTIVE 2)). Furthermore, methods to plan conservation measures are still scarce and the changing environment should be taken into account when allocating limited resources to gain the best possible outcome (OBJECTIVE 3).

1.4 Objectives and outline of the thesis

The overall goal of this thesis is to contribute to the emerging field of ecological modelling and to help building the bridge between ecology and mathematical optimisation. The thesis addresses the following objectives:

1. Deriving an optimisation model from a simulation model to answer similar research questions with less data requirements

2. Evaluating the metapopulation robustness of habitat networks and finding indicators for this robustness
3. Increasing the metapopulation robustness by adding links to already existing habitat networks

First, in **chapter 2**, I summarise definitions and results that build the base of this thesis and I present the (habitat) networks our research is based on. Then, in **chapter 3**, I present an optimisation model that applies tools from optimisation on graphs to a dispersal simulation model. Based on a dispersal simulation for a generic aquatic hemimetabolous insect, we develop an optimisation model as a surrogate for the simulation model. The optimisation model provides the minimum time for the species to colonise certain habitat patches under the given landscape structure. These results can be used to evaluate maximum possible range shifts in a given time frame and to estimate how fast restored or newly formed habitat patches can be (re-) colonised. Although our model was specifically developed for this simulation model, the same idea can easily be adapted to derive a surrogate for other simulation models.

Next, in **chapter 4**, I present our examination of the consequences of the permanent loss of habitat patches located on different types of habitat networks for metapopulations of different generic species characterised by traits describing their local-extinction risks and dispersal ranges. These results are compared among and between standard networks commonly studied in graph theory, such as random, regular, small-world, and scale-free networks (Boccaletti et al., 2006; Newman, 2010) on the one hand and spatial networks defined through random, clustered, and contiguous habitat allocation on riverine landscapes (Streib et al., 2020) on the other hand. On this basis, we investigate how well different graph-theoretic metrics of habitat networks can serve as indicators of metapopulation robustness against habitat loss.

Finally, in **chapter 5**, I address the question of where additional links should best be created within a habitat network to maximise its connectivity, for example in the context of a restoration measure. We propose an algorithm to identify the missing link of a network that leads to the biggest increase in network robustness. Here, we use the clustering coefficient as an indicator for robustness and thus identify the link that yields the biggest increase in the clustering coefficient. We introduce two heuristics, a Greedy algorithm (Krumke and Noltemeier, 2009) and a deducted Lazy Greedy algorithm to identify multiple missing links that increase the robustness the most when added to the network. Both approaches can be applied to any network, regardless of which system it represents. We test these heuristics 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.

Chapter 2

Preliminaries

During my PhD I investigated species dispersal on habitat networks, developed a simulation to evaluate the robustness of networks – or more precisely of the metapopulations represented by the networks – and developed and tested methods to increase this robustness. All of these calculations were performed on habitat networks developed to represent metapopulations of a generic hemimetabolous species with traits closely related to a dragon fly. The networks were created by Streib et al. (2020) and in this chapter I will briefly summarise their construction. Furthermore, I will summarise the most relevant tools and definitions in graph theory. Everything stated in this chapter can be found in Boccaletti et al. (2006), Krumke and Noltemeier (2009), and Streib et al. (2020).

2.1 Notation and definitions

An (undirected, simple) graph (also called network) is a pair $G = (V, E)$ of a non-empty set $V \neq \emptyset$ of nodes (also called vertices) and a set $E \subseteq V \times V$ of links (also called edges) that determines, which nodes in V are connected with each other. We call G loopless, if it does not allow for links that join a node to itself. We denote the link e connecting nodes u and v by $e = (u, v) = (v, u)$. If such link exists, u and v are said to be adjacent. Let v be a node in G . We call all nodes $u \in V$ that are adjacent to v the neighbours of v and set $N(v) := \{u \in V : (u, v) \in E\}$ the set of neighbours. The neighbourhood of v is the subgraph of G induced by $N(v)$, i.e. the subgraph (V_N, E_N) where $V_N = N(v)$ and $E_N = \{(u_1, u_2) \in E : u_1, u_2 \in N(v)\}$. The number of neighbours of v is called its degree: $\deg(v) := |N(v)|$.

A path is a sequence of distinct nodes that are adjacent to each other: $v_0, \dots, v_{n-1} \in V$ such that $(v_0, v_1), (v_1, v_2), \dots, (v_{n-2}, v_{n-1}) \in E$. It is a cycle, if additionally $(v_{n-1}, v_0) \in E$. We call a cycle of three nodes a triangle and set $T(v) := |\{(u, w) \in E : u, w \in N(v)\}|$ as the number of triangles in G that involve v . Furthermore, with $N(u, v) := N(u) \cap N(v)$ we denote the set of common neighbours of u and v .

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

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

It measures how close its neighbourhood is to a complete network in terms of the relative density of links. 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.

A directed graph $D = (V, E)$ is a graph, in which each link has a traversal direction. In particular is $(u, v) \neq (v, u)$ for nodes $u, v \in V$. The time expanded network $G_{\text{TEN}} = (V_{\text{TEN}}, E_{\text{TEN}})$ of a graph $G = (V, E)$ with time horizon $T \in \mathbb{N}$ is a directed graph with one copy v_t of each node $v \in V$ per time step $0 \leq t \leq T$. A link $(u, v) \in E$ of the original graph is represented by two direct links (u_t, v_{t+1}) and (v_t, u_{t+1}) . Furthermore, links between two subsequent copies of the same node are introduced: (v_t, v_{t+1}) .

A weighted graph is a graph $G = (V, E)$ with a function $w : E \rightarrow \mathbb{R}$ that assigns a weight to every link $e \in E$. In this thesis we assume weights to be positive, as they represent the energy an individual spends to travel from one habitat patch (node) to a neighbouring one, depending on the underlying landscape and the distance between patches. The shortest path between two nodes s and t – in ecology often called least-cost path – is a path $s = v_0, v_1, \dots, v_k = t$ from s to t with minimal cost, i.e. it minimises $\sum_{i=0}^{k-1} w((v_i, v_{i+1}))$. For an unweighted graph, we define a weight function $w : E \rightarrow \mathbb{R}, e \rightarrow 1$ and the shortest path between two nodes s and t counts the number of links that need to be traversed to reach t from s .

2.2 Landscape-based habitat networks

The landscape-based habitat networks are derived from a real-world stream landscape (Section 2.1.1. in Streib et al., 2020). For this purpose, we use a 50 km \times 50 km area of the stream landscape of Rhineland-Palatinate, Germany. The total stream landscape is divided into 25 landscape tiles each measuring 10 km \times 10 km to obtain a variety of dense or sparse stream landscapes with differing stream network structures. The landscapes surrounding the streams are created by randomly assigning one of the three land-use types ‘open agricultural land’, ‘forestry land’, and ‘urban area’ to each 25 m \times 25 m pixel of the 50 km \times 50 km area using the landscape-type configuration ‘random’ in Streib et al. (2020). We use land-use proportions of 25% for open agricultural land, 25% for forestry land, and 50% for urban area, to describe mixed landscapes with relatively large shares of urban area. The stream network is embedded into these landscapes by assigning a fourth land-use type ‘aquatic area’ to each pixel intersected by a stream. Each of the four land-use types are assigned dispersal costs to represent the permeability of the landscapes for a generic insect species inhabiting aquatic areas. We use dispersal costs of 25 for aquatic pixels, of 50 for agricultural pixels, of 75 for forestry pixels, and of 100 for urban pixels. Following these assignments, the landscapes are resampled to a new pixel size of 100 m \times 100 m.

On these landscapes, 10% of the aquatic pixels are selected as habitat patches using one of three different algorithms, giving rise to three types of landscape-based habitat networks with random, clustered, and contiguous habitat allocation. For random habitat allocation, 10% of all aquatic pixels are randomly selected with equal probability. For clustered habitat allocation, 5% of all aquatic pixels are randomly selected with equal probability, and 5% of all aquatic pixels are randomly selected

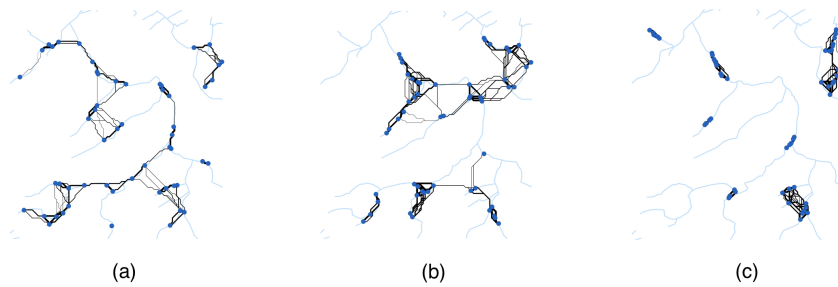


FIGURE 2.1: Examples of landscape-based habitat networks with (a) random, (b) clustered, and (c) contiguous habitat allocation. Blue dots represent habitat patches, light-blue lines depict the underlying stream landscape, and black lines indicate dispersal pathways. The width of these lines represents the dispersal costs with thick lines indicating lower costs and thin lines indicating higher costs.

with equal probability and without repetition within a radius of 500 m around the initially selected ones. For contiguous habitat allocation, we follow the same principle as for clustered habitat allocation, except that merely 2.5% of all aquatic pixels are randomly selected initially, resulting in a more contiguous allocation of the other habitat patches. Reflecting the different stream structures in the different landscape tiles, this results in habitat networks with 54 to 111 habitat patches.

To determine the dispersal costs between habitat patches, a least-cost path analysis is used and a link is introduced to the network, if the dispersal cost between two patches do not exceed a given maximum. Thus, the maximal dispersal costs directly affect the number of links in a network. See Fig. 2.1 for example of the habitat networks presented here.

2.3 Standard networks

We compared the landscape networks introduced above with networks common in mathematics, such as regular, random, small-world, and scale-free networks.

In a regular network $G = (V, E)$, each node is adjacent to the same number of nodes. Landscapes can easily be represented as regular networks by dividing the area in even tiles, such as rectangles or hexagons, and representing the tile as a node with links between neighbouring tiles.

The algorithm to create random networks was introduced by Erdős and Rényi (1960). In an Erdős–Rényi graph, two nodes are connected with a given probability. Given n nodes and a probability p , an Erdős–Rényi graph arises by instantiating n nodes and going through every possible connection between those nodes and including it into the network with probability p . Each pair of nodes thus has equal probability to be connected and the degree of nodes follow a binomial distribution.

Watts and Strogatz (1998) discovered that many real-world networks have rather different properties to the networks created using the Erdős–Rényi model and developed a different model to create so-called small-world networks. In small-world networks, most nodes are not connected with each other, but almost every node can be reached from any node in just a few steps. Small-world networks have a significantly higher clustering coefficient compared to random networks while at the same time exhibiting a similarly small shortest path length. These structures are

found in many real-world networks such as social networks, but also ecological networks (Fox and Bellwood, 2014; Prima et al., 2019). The Watts-Strogatz-model creates small-world networks from a regular network by randomly rewiring the links with a given probability p . For $p = 0$, no links are rewired and we obtain a regular network. For $p = 1$ every link is rewired and we obtain a random network. The small-world networks thus interpolate between regular and random networks.

Many real-world networks also exhibit a structure containing of a few, well connected nodes while most of the nodes have a rather small degree. This so-called scale-free structure cannot be generated using either of above mentioned algorithms. Barabási and Albert (1999) developed a model to create scale-free networks using the preferential attachment process. The algorithm successively adds nodes to a network and connects each node with a predefined number of nodes that already exist in the network with a probability proportional to the degree of the existing node. See Fig. 2.2 for example of the networks presented here.

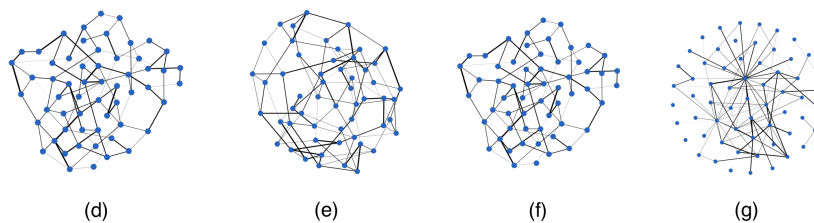


FIGURE 2.2: Examples of standard networks with (a) random, (b) regular, (c) small-world, and (d) scale-free structure. Blue dots represent habitat patches, light-blue lines depict the underlying stream landscape, and black lines indicate dispersal pathways. The width of these lines represents the dispersal costs with thick lines indicating lower costs and thin lines indicating higher costs.

Chapter 3

Optimisation model of dispersal simulations on a dendritic habitat network

The content of this chapter has already been published in an international reviewed journal and can be accessed via the following link:

Heer, H., Streib, L., Kattwinkel, M., Schäfer, R. B., & Ruzika, S. (2019). Optimisation Model of Dispersal Simulations on a Dendritic Habitat Network. *Sci Rep* 9, 8202 (2019). <https://www.nature.com/articles/s41598-019-44716-z>

Abstract

To predict and mitigate biodiversity loss, a better understanding of species distribution and reliable dispersal models are required. A promising approach in dispersal simulation is the method of spatially explicit graph-based analysis. While graph theory is strongly connected to the field of optimisation in a variety of disciplines, the potential of optimisation has not yet been exploited in dispersal models. We introduce an optimisation model built on a graph-based dispersal simulation of an aquatic invertebrate species with a terrestrial life stage. The model simulates a directed dispersal process and investigates the fastest route to colonise predefined vacant habitat patches. The optimisation model run-time is in general an order of magnitude faster than the underlying simulation and provides the minimum time until the considered habitat patches are colonised under the given landscape structure. These results can then be used to estimate how fast newly formed habitat patches can be reached and colonised. Our model can in principle be adapted to other simulation models and can thus be seen as a pioneer of a new set of models that may support landscape conservation and restoration.

3.1 Introduction

Climate change effects have now been measured throughout all ecosystems and include, but are not limited to, changes in species' phenology, abundance and distribution (Scheffers et al., 2016; Parmesan and Yohe, 2003; Parmesan, 2006). Widespread range shifts have been documented with range expansions in warm-adapted species and range contraction in cold-adapted species as well as a consistent trend of northward or westward range expansion of species in the northern hemisphere (Berg et al., 2010; Scheffers et al., 2016; Parmesan and Yohe, 2003) However, some species show little to no net range shifts and range shifts in general remain little understood

(Doak and Morris, 2010; Moritz et al., 2008). Range shifts are complex processes driven by population dynamics and dispersal, which themselves are determined by a variety of factors such as changes to the abiotic and biotic environment (Zurell et al., 2016; Sexton et al., 2009). To understand and mitigate the impacts of climate change on global biodiversity, reliable models of species dispersal are needed (Zurell et al., 2016; Bellard et al., 2012).

Spatially explicit dispersal models for freshwater insects are scarce (Heino et al., 2017). This scarcity is primarily due to the lack of data, as field studies are very costly and methodical limitations complicate the reliable derivation of dispersal distances for (freshwater) insects (Tonkin et al., 2018; Didham et al., 2012). A deeper knowledge in the field of species distribution and therefore species dispersal — as one of its key factors — is required to allow for prediction of the effects of climate change (Lowe and McPeck, 2014).

The method of spatially explicit graph-based analysis is one of the most promising approaches to model dispersal of aquatic individuals (Heino et al., 2017; DeAngelis and Yurek, 2017). This was adopted from the field of graph theory and gained popularity in landscape ecology and conservation biology in recent years (Galpern et al., 2011; Urban et al., 2009). While spatial graphs have become an important tool in terrestrial landscape ecology, they are still rarely used in aquatic ecosystem modelling (Erős et al., 2012b). The advantages of this graph-based structure are numerous. First, graphs are particularly flexible as vertices can represent multiple ecological properties, e.g. single individuals, whole populations or, as most appropriate in dispersal models, habitat patches. Second, vertices are connected by links which specify the connectivity relationship (Galpern et al., 2011; Urban et al., 2009; Calabrese and Fagan, 2004). Furthermore, spatially explicit data derived from geographic information systems (GIS) can be combined with information on dispersal characteristics of the considered species. At the same time only relatively few data are required (Erős et al., 2012b; Rayfield et al., 2011).

Although graph-based structures are now commonly used in ecological models such as dispersal simulations, the potential of optimisation on graphs has — to the best of our knowledge — not yet been exploited in dispersal models. Mathematical optimisation and graph theory are strongly connected in various other disciplines as finance, logistics, engineering and transportation and optimisation is used ubiquitously to solve a variety of problems in various disciplines (Hamacher and Klamroth, 2006; Bondy and Murty, 1976). In general, optimisation approaches are used to identify the ‘best’ solution for a given problem. In the context of dispersal simulations, optimisation can be used to find the fastest way for a species to disperse and related to this the minimum time required to colonise a habitat. Optimisation involves modelling a directed dispersal in contrast to the undirected dispersal that is usually simulated. This approach needs less information on dispersal strategies that define which habitats are preferably colonised and how to divide the dispersing biomass between all neighbouring habitat patches. This is an exceptional advantage, as collecting data is costly and making assumptions is error-prone. On the other hand, an optimal solution only determines bounds for a given problem and provides thus a more general and less specific solution to given research questions such as how long it takes a species to colonise a habitat or which habitats are colonised first.

This study applies tools from optimisation on graphs to a simulation model for dispersal. Given a graph-based model to simulate the spread of a generic aquatic invertebrate with a terrestrial life stage, an optimisation model is derived as a surrogate for the former model. It yields lower bounds on the colonisation time of specific habitats, which provide the minimum time until the considered habitat patches are

colonised under the given landscape structure. These results can be of great value, as they identify how far the considered species can disperse within a given time frame and thus give an indication of maximum possible range shifts. At the same time, the model can be used to estimate how fast newly formed habitat patches can be reached and colonised. This information can then be used to modify the underlying connectivity to make habitats more accessible or to study the impact of land use changes. Although being specific for this simulation model, the general idea of deriving a surrogate can in principle be adapted to other simulation models.

Our optimisation approach differs vastly from the least cost path method (Sawyer et al., 2011). The least cost path technique identifies a shortest connection between a pair of nodes, but does not consider the interaction of multiple source habitats. Our model also takes the possibility into account, that a habitat patch can be reached by more than just one neighbouring patch at a time and thus many patches can jointly colonise a habitat patch. Circuit theory (McRae et al., 2008) on the other hand incorporates the possibility of multiple pathways between habitat patches. In contrast to our model, however, it is largely applied to random walk theory. While it can be used to obtain an estimate of dispersal time, it is not designed to calculate lower bounds for these – the main feature of our optimisation model.

The interaction between optimisation and simulation is not a new field of study. However, note that our approach significantly differs from the topic of “simulation optimisation” (SO) (Amaran et al., 2016). SO is an umbrella term for techniques that search for specific settings of the input parameters to optimise stochastic simulations and often only depend on input - output data from these simulations. In contrast, our model is based on a deterministic simulation and can be classified as a traditional mathematical optimisation technique. Furthermore, we modify well-known mathematical optimisation techniques to develop a model as a surrogate of an ecological simulation model that answers different, but related questions to the simulation model.

We first present the graph-based simulation model for the distribution of a generic aquatic invertebrate with a terrestrial life stage. From this simulation model, we then derive a mathematical optimisation model in form of a mixed integer programming model (Dantzig, 2016; Schrijver, 1998; Wolsey, 2008). This model modifies and utilises the concept of dynamic network flows (Ford Jr and Fulkerson, 1958; Ford Jr and Fulkerson, 2015). Network flows are typically applied in transportation systems, air traffic control, production systems and financial flows (Skutella, 2009; Köhler et al., 2009; Kotnyek, 2003) but have not yet been used in ecology. Given some vacant habitats as targets, the optimisation model finds a route to colonise those habitats as quickly as possible.

3.2 Methods

3.2.1 Simulation Model

We developed a dynamic, spatially explicit dispersal model for a generic aquatic invertebrate species with a terrestrial life stage. The simulation model can also be adapted to vertebrates with both aquatic and terrestrial life stages (Grant et al., 2010; Searcy et al., 2013). The simulation is based on a habitat network embedded in an artificial landscape defined by four land cover classes (Tab. 3.1).

Since the overland dispersal of invertebrates during their terrestrial life stage is influenced by land cover (e.g. preference for specific land cover classes), we assigned dispersal-related costs to these land cover classes (Grönroos et al., 2013).

class name	percentage	dispersal costs
agriculture	66.6%	50.0
forest	11.1 %	75.0
urban	22.2%	100.0
aquatic	-	25.0

TABLE 3.1: Dispersal costs and ratio per land cover class. The percentages refer to the underlying Neutral Landscape Model (NLM) that was used to create the landscape (Suppl. Inf. S1). A real stream network on a finer scale is added to the landscape and cells intersecting with the river network are declared as ‘aquatic’.

These costs determine the spatial connectivity between habitats. They were chosen to represent landscape permeability with a relatively energy efficient dispersal through aquatic and open agricultural terrain, whereas forests and urban areas represent a rather costly dispersal path. Habitats are located along a stream network that is embedded in the landscape and are assigned with random habitat qualities which determine the maximum population that can be sustained in a habitat patch, called carrying capacity. Some of the habitat patches are randomly chosen as initial source habitats and considered colonised at the start of a simulation. The dispersal process from those patches is modeled as a dynamic process using a modified individual based model (Suppl. Inf. S1). The simulation is based on the demography-related processes population-growth (depending on habitat quality) and density-dependent emigration (Corbet, 1963; Córdoba-Aguilar, 2008; Bowler and Benton, 2005) (Fig. 3.1). Consequently, the amount of dispersing biomass primarily depends on population size and habitat quality as controlling factors of maximum population size (carrying capacity) (Amarasekare, 2004; Hodgson et al., 2011). In a colonised habitat patch, the population initially grows exclusively due to immigrating biomass from neighbouring source habitats. After a predefined threshold of biomass is reached, it turns into a source habitat and an additional population growth as well as emigration is simulated (Fig. 3.1). We assume that habitat patches that can be reached at low dispersal costs are preferably colonised (Van Nouhuys and Hanski, 2002) and thus receive a bigger share of biomass, where the dispersal costs depend on both the distance to a source habitat and the land cover classes traversed (Suppl. Inf. S1).

The maximum dispersal distance was set to 2500 m through open agricultural land (Keller et al., 2012; Hepenstrick et al., 2014). Consequently, on a cost raster with a cell-size of 100 m \times 100 m, our model species was assigned a maximum budget of 1250 cost units (‘agriculture’ 50 cost-units \cdot 25 raster cells) and two habitat patches were considered connected, if the dispersal costs between them was less than $C_{\max} = 1250$. This results in a graph-based habitat network $G = (V, E)$, where the set of vertices V subsumes all habitat patches and the edge set E contains all connections between them (Suppl. Inf. S1). The same habitat network is used as basis for the optimisation model.

One habitat network was created as basis for all following simulations. 50 sets of initial source habitats were randomly selected as simulation input (see *Initialisation*). As the simulation is deterministic, a single simulation per model input was sufficient. Although the design of a habitat network has a strong influence on species dispersal, we considered only one habitat network, as its influence was beyond the

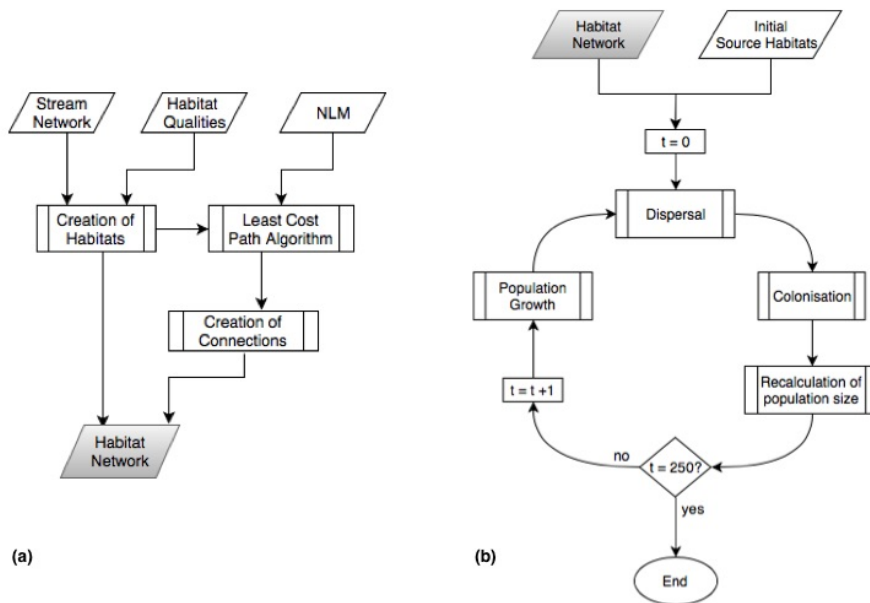


FIGURE 3.1: Flowchart of the simulation model. A detailed description of each subprocess (double rectangles) can be found in supplementary Information S1. (a) creation of habitat network, which is then used as input for the dispersal simulation depicted in (b)

focus of this study.

3.2.2 Optimisation Model

Process Overview

The simulation model assumes that close habitats are preferably colonised and applies a colonisation route accordingly (Suppl. Inf. S1). Here, a colonisation route is a detailed plan of the species's movement in the network over time, which leads to a colonisation success. In terms of the time expanded network (see next subsection), we define a colonisation route as a set of paths connecting the source to the corresponding copies of the destination habitats in the time expanded network combined with the information about the exact amount of biomass that is traveling along that path. Since this assumption has a strong impact on the colonisation time of vacant habitats, we design an optimisation model that identifies a route to colonise specific, predetermined habitat patches as quickly as possible. Analogously to the simulation model, a habitat network is created and a predefined share of habitat patches are randomly selected as initial source habitats. Additionally, a set of habitat patches is selected as destination habitats. The fully colonised initial source habitats initiate the dispersal process and dispersal is directed towards the selected destination habitats in contrast to the undirected dispersal in the simulation model. The model output is a provable lower bound on the colonisation time of the simulation model and guarantees that the predefined destination habitats will not be colonised earlier — independent of the dispersal route. Furthermore, the run-time of the optimisation model is substantially faster than the simulation model. The optimisation model is instantiated on the same landscape model as the simulation. The graph-based habitat network $G = (V, E)$ created by the simulation model (Suppl. Inf. S1) was used as a representation of the investigated area. Similar to the simulation model, population growth is not taking place before a specific threshold T_{SH} is reached. T_{SH} is

the minimum viable population — a simplified threshold that specifies the smallest amount of biomass needed for a species to persist in a habitat patch. However, to simplify the model, once the threshold is reached, the population will grow to the habitat specific carrying capacity $K(v)$ (Suppl. Inf. S1) within one time step. Thus, a habitat with a population size of at least T_{SH} units of biomass is considered to be fully occupied and a source habitat in the following time step. After a loss of biomass due to dispersal, the population of a habitat patch is set again to the carrying capacity in the same time step. To sum up, each source habitat v has a constant population size of $K(v)$ and can release an additional amount of up to $S_{DIS}K(v)$ biomass during the dispersal process.

The model utilises the method of time expanded networks (Ford Jr and Fulkerson, 1958; Ford Jr and Fulkerson, 2015) and solves a mixed integer program (MIP) (Wolsey, 2008; Schrijver, 1998; Dantzig, 2016) to compute the desired bounds.

Habitat Network

A time expanded network $G_{TEN} = (V_{TEN}, E_{TEN})$ (Fig. 3.2) is created to represent the graph-based habitat network $G = (V, E)$ and to store the population size of each habitat patch in every time step. A time expanded network is a directed network (i.e. connections between vertices have a direction and can only be traversed along this direction) (Hamacher and Klamroth, 2006) with one copy of each habitat patch of the underlying habitat network per time step (*time layer*) and connections between habitat patches in consecutive layers.

Let T be the time horizon, i.e. the maximum number of time steps considered in the model. For each habitat patch $v \in V$, $T + 1$ copies v_0, \dots, v_T are constructed which represent the habitat patch v at time steps $0, \dots, T$. For each connection $(u, v) \in E$ between two habitat patches u and v and each time step $t = 0, \dots, T - 1$, two directed connections (u_t, v_{t+1}) and (v_t, u_{t+1}) are introduced in the time expanded network. To model the possibility of remaining within a vertex between two consecutive time steps, connections (v_t, v_{t+1}) are introduced for all $t = 0, \dots, T - 1$ and for all habitat patches $v \in V$. As is typically done in time expanded networks (Ford Jr and Fulkerson, 1958; Ford Jr and Fulkerson, 2015), a super source Q is introduced together with connections (Q, q_0) to the first copy of each initial source habitat q .

All habitat patches and connections are equipped with the original data: each copy of a habitat patch $v \in V$ is assigned the same dispersal capacity value as the original, $u(v_t) := S_{DIS}K(v)$. Each copy of a connection is assigned the same dispersal costs as the original ones, while the artificial connections (connections between copies of the same habitat patch and all connections from Q) are assigned zero cost (Fig. 3.2).

Mixed Integer Program

In the following, an optimisation problem is formulated which yields the minimal colonisation time as described above. To this end, techniques from integer programming are applied and the result is a so-called mixed integer programming problem (MIP) (Dantzig, 2016; Schrijver, 1998; Wolsey, 2008) which will then be solved by some integer programming solver.

Binary decision variables $x(v_t)$ are introduced for each habitat $v \in V$ and each time step t . If $x(v_t) = 1$, then v_t is a source habitat and otherwise it is not. Furthermore, for each connection $e \in E_{TEN}$ in the time expanded network, a real variable

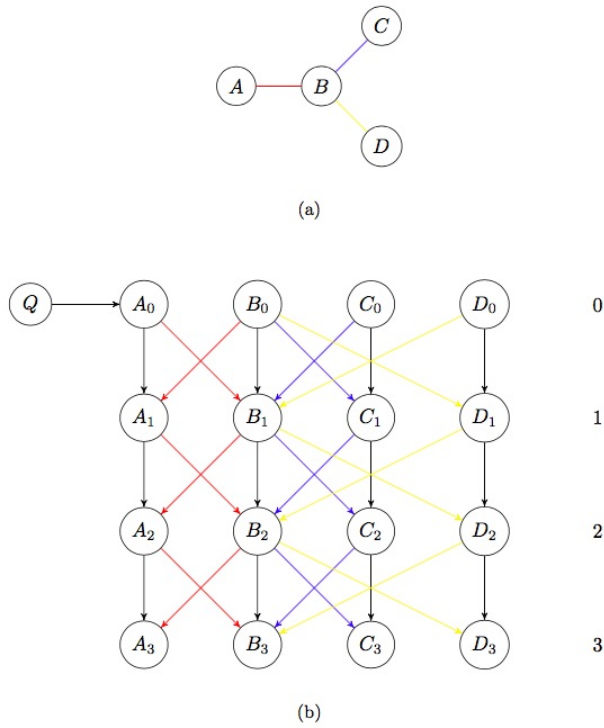


FIGURE 3.2: (a) Habitat network (b) Corresponding time expanded network with time horizon 3 and source habitat A . Connections in (b) representing a specific connection in (a) use the same colour. Black connections in (b) are artificial connections indicating remaining in the same habitat patch from one time step to the next. Q represents the super source.

$f(e) \geq 0$ specifies the amount of biomass traveling along this connection. Moreover, binary variables $x(t)$ are introduced for each time layer $t \in \{0, \dots, T\}$ indicating if all destination habitats in the corresponding time layer are source habitats. In the following, $\delta^+(v)$ denotes the set of all connections leaving v and, analogously, $\delta^-(v)$ specifies the set of incoming connections into v .

The objective function of the MIP minimises the sum of all time layer decision variables multiplied by t , over all t .

$$\min \sum_{t=0}^T tx(t)$$

Here, the cost coefficient t with which the decision variable is weighted, corresponds to the time and thus increases over time. Thus, in view of the minimisation objective, it is desirable, to send biomass to the destination habitats as quickly as possible. This objective function was adopted from models for the so-called quickest flow problem and the earliest arrival flow problem and guarantees that the fastest way to colonise the specific destination habitats will be found (Jarvis and Ratliff, 1982).

The first set of constraints

$$T_{SH}x(v) \leq \sum_{e \in \delta^-(v)} f(e) \quad \forall v \in V_{TEN} \setminus \{Q\} \quad (1)$$

ensures that a habitat v at time step t can only become a source habitat, if the incoming amount of biomass at time step t plus the biomass from the previous time step (represented as $f((v_{t-1}, v_t))$) are at least T_{SH} .

The second set of constraints

$$f(v_{t-1}, v_t) = \sum_{e \in \delta^-(v_{t-1})} f(e) \quad \forall v \in V, t \in \{0, \dots, T\} \quad (2)$$

sends all biomass of a habitat patch from the previous time step $t - 1$ to the current time step t .

The constraints

$$\sum_{e \in \delta^+(v)} \frac{f(e)}{1 - C(e) \frac{1}{C_{\max}}} \leq u(v)x(v) + \sum_{e \in \delta^-(v)} f(e) \quad \forall v \in V_{\text{TEN}} \quad (3)$$

are the crucial constraints of the model. They ensure that a source habitat does not emit more than an upper limit of biomass and simultaneously take into account that only a fraction of the biomass emitted reaches the connected habitats. On the left hand side of the inequality, the emitted biomass $f(e)$ is reduced by the mortality rate $C(e) \frac{1}{C_{\max}}$, where $C(e)$ represents the dispersal cost of a connection e and C_{\max} is the maximum dispersal cost (see simulation model). This reduction represents the mortality of dispersing biomass. The higher the dispersal costs $C(e)$ of a connection e , the smaller the share of biomass traversing connection e to reach the destination. The right hand side now ensures that a source habitat v does not emit more than $u(v)$ units of biomass. If v is no source habitat, then $x(v) = 0$ and no additional biomass can be emitted. The additional amount $\sum_{e \in \delta^-(v_t)} f(e)$ is the amount of biomass that stays in the habitat (constraint 2) and is sent into the next time step.

The constraints

$$f(Q, q_0) = T_{SH} \quad \forall q \in H_{\text{Start}} \quad (4)$$

ensure that all initial source habitats $q \in H_{\text{Start}}$ are fully colonised (according to T_{SH}) at time step 0, where H_{Start} is the set of all initial source habitats.

The fifth set of constraints

$$|H_{\text{dest}}|x(t) \leq \sum_{s \in H_{\text{dest}}} x(s_t) \quad \forall t \in \{0, \dots, T\} \quad (5)$$

ensures that the time layer variable $x(t)$ can only be set to one, if all destination habitats are colonised in that time layer, where H_{dest} is the set of destination habitats, and the constraint

$$\sum_{t=0}^T x(t) \geq 1 \quad (6)$$

requires that all destination habitats have to become source habitats eventually.

All in all, the following MIP is obtained, which can now be solved with the help of any MIP solver such as the one provided by Gurobi (Gurobi Optimization, Inc.,

2016).

$$\begin{aligned}
 & \min \sum_{t=0}^T tx(t) \\
 & \text{subject to (1) - (6)} \\
 & f(e) \in \mathbb{R}_+ \quad \forall e \in E_{\text{TEN}} \\
 & x(v) \in \{0, 1\} \quad \forall v \in V_{\text{TEN}} \\
 & x(t) \in \{0, 1\} \quad \forall t \in \{0, \dots, T\}
 \end{aligned}$$

Time Horizon

The choice of the time horizon is crucial to the model performance. Since the calculations are executed on a time expanded network, the model input is linear in T and thus a large time horizon will lead to an exorbitant model run-time, while a time horizon chosen too small will not return any information as the MIP will turn out to be infeasible. Thus, a good approximation of the maximum number needed will vastly improve the model performance. The following procedure was used to find the appropriate time horizon for a given habitat network and its specific initial source habitats and destination habitat.

With the help of the Python module 'Networkx' (Hagberg et al., 2008b) and taking the dispersal costs into account, a shortest path was calculated from each initial source habitat to the destination habitat. Based on these results, the nearest initial source habitat was identified and the destination habitat was colonised with successively colonising the habitat patches v_i from the nearest initial source habitat along the shortest path $P = (v_1, \dots, v_k)$ to the destination habitat, using the colonisation rules of the optimisation model. This can be calculated with the following formula:

$$TH_1 = \sum_{i=1}^n \left\lceil \frac{T_{SH}}{K(v_i)S_{DIS} \left(1 - \frac{C(v_i, v_{i+1})}{C_{\max}}\right)} \right\rceil,$$

Since this is only one of many feasible possibilities to colonise the specific destination habitat, the minimum of all possibilities is clearly smaller. To obtain an even closer bound, the same procedure was performed with the second nearest initial source habitat, if available, obtaining a second bound TH_2 for the time horizon. Although the cumulative dispersal costs from the second initial source habitat to the destination habitat is not smaller than from the first one, the second bound can be smaller than the first one due to rounding to integers in the formula, for instance. Thus, the minimum of both bounds is taken as the time horizon. In some cases both bounds TH_1 and TH_2 turned out to be too big and thus the minimum of both bounds and 30 was selected as time horizon for all model runs:

$$T = \min\{TH_1, TH_2, 30\}$$

If the MIP with this time horizon was infeasible, a new time horizon was set to be the minimum of TH_1 , TH_2 and 60 and the procedure was repeated with higher multiples of 30 if necessary. Although this led to a slower performance for model runs with an outcome bigger than 30 (due to solving a smaller, infeasible model and repeating the process), this method was used as it yielded a speedup for the majority of all model runs. Indeed only one percent of all model initialisations needed a time horizon bigger than 30.

3.2.3 Initialisation

The habitat network constructed by the simulation model was used for both models. To compare the optimisation model with the underlying simulation, one habitat network was chosen to represent the underlying landscape structure and both models were instantiated with the same model parameters (Suppl. Inf. Tab. S2). For each simulation model run a new set of initial source habitats was chosen. The same set was taken as optimisation model input for multiple model runs. Additionally a set of destination habitats was randomly chosen and each of the destination habitats was combined individually with each set of initial source habitats as input for an optimisation model run. As dispersal is undirected in the simulation model and the simulation model is deterministic, one model run per set of initial source habitats was sufficient to investigate the colonisation time of all possible habitat patches. For the optimisation model, dispersal is directed and different destination habitats have to be considered individually.

The considered extent of 50 km \times 50 km of the stream network accounts for a total of 19,490 pixels classified as 'aquatic'. As described in the simulation model (Suppl. Inf. S1), a random selection of 10% of these pixels were chosen as habitat patches. Together with the connections created by the least cost path algorithm (Suppl. Inf. S1), these habitat patches form the habitat network. One habitat network was created and then used for all model runs. For each simulation model run, 10% of those habitat patches were randomly selected as initial source habitats. In total, 50 distinct sets of initial source habitats were chosen and taken as model input of the simulation model. Additionally, 50 habitat patches were elected as destination habitats and each set of initial source habitats combined with each destination habitat individually were taken as model input for the optimisation model. Although the optimisation model was developed to determine the minimum colonisation time for a set of (multiple) destination habitats, we focus on a single destination habitat from here on. This makes it easier to compare outcome and run-time of the two models, as the combination of different destination habitats has a strong influence on them. Both models were implemented in Python 2.7. The MIP solver provided by Gurobi (Gurobi Optimization, Inc., 2016) was used to solve the optimisation problem. Both models were executed on a server with the Ubuntu release 16.04.3 LTS, Intel Xeon 16 core processor 2.50 GHz with memory of 31.4 GB and timed with the help of the Python module 'Timeit'.

3.2.4 Analysis of Models

First, we compare the optimisation model outcome with the model input by examining the dispersal distance of all initial source habitats from the specific destination habitat for all optimisation model runs. Second, to compare the outcome of both models, we investigate the number of time steps to colonise the considered destination habitats and compare the outcome of each of the 2,500 optimisation model runs to the time step in which the corresponding destination habitat changed its status into a source habitat in the simulation model for the first time (Suppl. Inf. S1). Third, we compare the run time of both models.

3.3 Results & Discussion

3.3.1 Analysis of Model Results

We found a positive correlation ($r = 0.85$) between the distance (in terms of dispersal costs) from the nearest initial source habitat to the destination habitat and the minimum colonisation time (in terms of time steps) calculated by the optimisation model (Fig. 3.3).

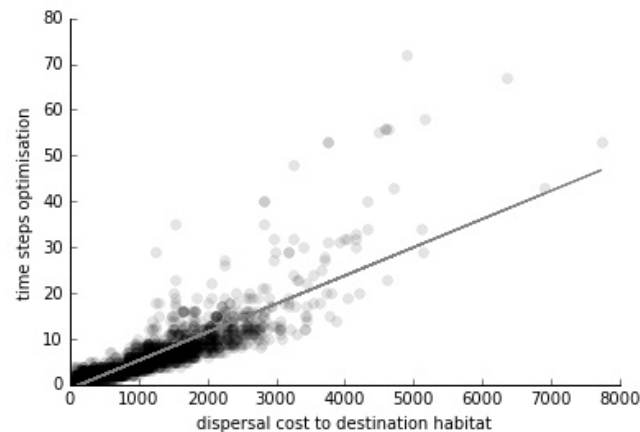


FIGURE 3.3: Relationship between the time steps found by the optimisation model and the cumulative dispersal cost from the nearest (i.e. most cost efficient) connected initial source habitat to the destination habitat of the corresponding model run. Each black mark in the plot corresponds to one of the 2,500 model runs. The grey line is the linear regression.

By contrast, the average dispersal costs from all initial source habitats (that are connected to the destination habitat) to the destination habitat is only weakly correlated with the optimisation model outcome ($r = 0.34$). This fact indicates that not all initial source habitats play an equally important role in the colonisation of the corresponding destination habitat. In fact, the dispersal costs of close initial source habitats have a much stronger influence on the optimisation model outcome. To conclude, the optimisation model colonises those habitat patches faster, that are more cost-efficient to reach. This was not implemented in the optimisation model and conforms to common literature and the assumption made for the simulation model that close habitat patches are preferably colonised (Van Nouhuys and Hanski, 2002; Kajzer et al., 2012). Thus, the optimisation model responds in a similar fashion to the simulation model and observations in common literature, which can be seen as a partial validation of the optimisation model.

In a second step, the outcome of both models was compared. In less than 2 % of the model inputs (41 out of 2,500), the destination habitat in the simulation model was not colonised after 250 time steps. These model instances were omitted in this analysis, as the simulation model outcome is unknown. However, they also had a considerable large optimisation model outcome with a mean of 32.4 time steps (range from 2 to 72, median 29).

The optimisation model colonises a destination habitat on average 6.8 times faster than the simulation model (Fig. 3.4). Thus, the optimisation model not only gives lower bounds on the colonisation time of the simulation model, but also gives an

estimate of the expected outcome of the simulation. However, this estimate is subject to considerable uncertainty and ranges from 1 to 98-fold for different model runs. Model runs with the highest deviation from this average have an optimisation

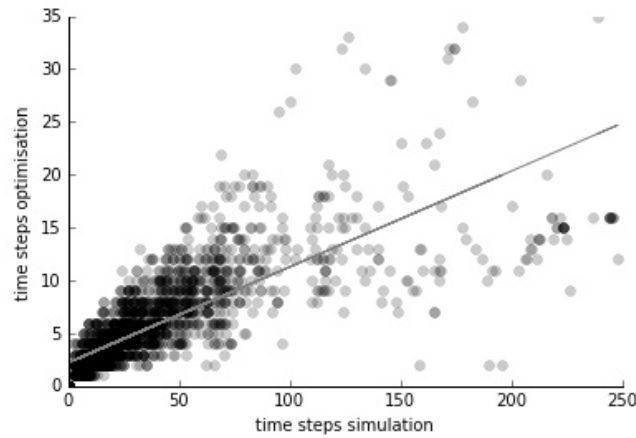


FIGURE 3.4: Correlation between time steps needed for colonisation of destination habitat of simulation and optimisation. Each black mark corresponds to a single run of the optimisation model and the corresponding result from the simulation. The grey line is the linear regression.

model outcome of 2 time steps (Fig. 3.4). These high deviations happen particularly in dense areas. In the simulation model, the initial source habitats will have many neighbouring habitat patches and dispersing biomass is distributed among many neighbours (Suppl. Inf. S1) — leaving only a small share for the designated destination habitat. The optimisation model on the other hand sends all available biomass directly towards the designated destination habitat. Accordingly, the destination habitat will be colonised much faster in the optimisation model compared to the simulation.

3.3.2 Model Run Times

The mean time of a simulation model run is 318 seconds with a standard deviation of only 5 seconds. The optimisation model was much faster on average, but also included some model runs with larger run-times. A run of the optimisation model takes 17 seconds on average and is thus almost 20 times faster than the corresponding simulation model. However, the performance varies vastly for different settings. The 90th percentile is 26 seconds and the 50th percentile is 3.5 seconds, while 16 of all 2,500 optimisation model runs (0.64%) were slower than the simulation model. The computationally most expensive model instances also have a rather large model outcome. This may be due to the way the time horizon was chosen and the fact that a higher time horizon was needed. Similar to a single destination habitat, the run time of the optimisation model with multiple destination habitats varies depending on the input. However, the run time takes on average 60 seconds and is thus considerably slower than the single destination case. On the other hand, this is still roughly 5 times faster than the simulation model run time. Thus, the run time advantage also holds for multiple destination habitats.

To conclude, the optimisation model is one order of magnitude faster than the simulation model. On the other hand, some model instances are hard to solve and

less than one percent of the instances needed more time than the corresponding simulation model.

It is important to point out that the two models pursue different goals and are thus difficult to compare. While the simulation model investigates, *inter alia*, the distribution of colonised habitat patches after a given number of time steps, the optimisation model examines the minimum number of time steps needed to reach a specific habitat patch. Thus our model is not a mere surrogate which answers the same questions with less accuracy, but provides results that cannot be found using the original model — in contrast to other surrogate models like (Gaussian process) emulators. Gaussian process emulators are statistical models that approximate unknown output of a complex and time-consuming simulation. Given some design data consisting of input - output pairs, the simulation output of further inputs are approximated by a Gaussian process (Bastos and O'Hagan, 2009). Emulators are orders of magnitude faster than their original model (Machac et al., 2016a; Machac et al., 2016b). Thus, considering the performance gains, the optimisation model can compete with emulators, but would be considered a slow speed-up.

On the other hand, the simulation model run-time strongly depends on the total number of simulated time steps. The total number of 250 time steps was chosen such that most habitat patches were reachable within that time frame and such that the number of time steps was not so high that run-time was needlessly increased. A better run-time comparison could be achieved by adjusting the fixed number of 250 time steps to an input-dependent number (for example by stopping the simulation when the destination habitat in focus is colonised). This adaptation, however, changes the focus of the simulation model and is not intended.

3.3.3 Example

In this section we demonstrate how our optimisation model can be applied in landscape management. Fig. 3.5 shows an artificial landscape (created as described in the Methods section) where a species is present in the southern area of the landscape (initially colonised habitat patches are represented by red circles). Due to climate change, more patches located in the northern part of the landscape become habitable. To evaluate how to facilitate the spread of our focal species to the newly formed habitat patches, a central patch is chosen as destination habitat (yellow star) as model input.

With the current underlying landscape scenario, the destination habitat is only colonised after 34 years, if the species disperses along the given paths. This result can then be used by landscape and freshwater managers to facilitate the colonisation by strengthening the connections as well as the habitat patches along them to allow for faster and easier traversal.

One should keep in mind, however, that the focus of this model is to determine a minimum colonisation time rather than identifying suitable areas in the landscape that yield the largest improvement (in terms of colonising the destination patch as fast as possible) if enhanced.

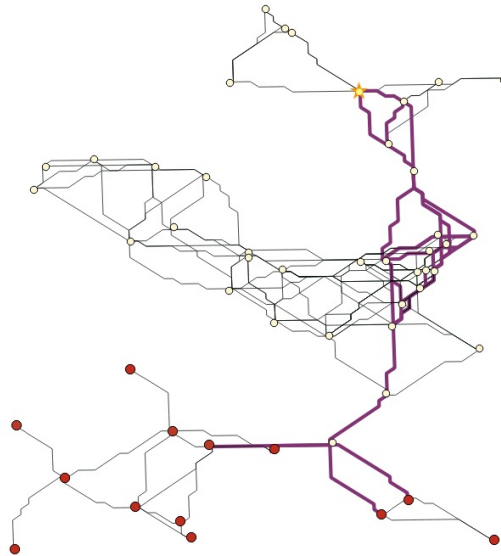


FIGURE 3.5: Application of optimisation model on a habitat network. The red nodes represent initially colonised habitat patches, the yellow star is the corresponding (uncolonised) destination habitat. The thick, purple lines show the dispersal paths displayed by the optimisation model

3.3.4 Challenges & Outlook

To construct the optimisation model from the simulation, the habitat network was transformed into a time expanded network. Additional decision variables were introduced to memorize the fully populated habitat patches. The time expansion results in an exponentially bigger input size, implying a loss of computational efficiency. At the same time, this is a common structure to monitor changes (here: of population size in habitat patches) over time (Skutella, 2009; Köhler et al., 2009; Kotnyek, 2003). Furthermore, this structure can later be exploited to integrate changes over time, for example in habitat quality, into the model.

Translating the dispersal process and population growth of the simulation model to linear constraints is the most challenging part in creating the optimisation model. In particular, integrating a more realistic population growth process into the optimisation model would increase the complexity considerably, as it demands additional decision variables and constraints. These were omitted in our optimisation model and the simulation of population growth was simplified to ensure a faster and simpler model. On the other hand, if the species in focus has very slow or complicated population dynamics, this simplification may lead to a huge underestimation of the colonisation time. Linear constraints are the core of linear programming and the main challenge in adapting the optimisation model to other simulation models will be to translate complex processes into linear equations.

Some studies also found inverse density dependent dispersal patterns for damselflies (Rouquette and Thompson, 2007). Inverse density dependent dispersal characterises the pattern that occurs when individuals from sparsely populated habitat patches gravitate towards more densely populated patches. In the current study, we focus on dispersal from colonised to empty patches, not between two colonised ones. Therefore, such inverse density-dependent dispersal is not relevant for our research question. Furthermore, the survival probability of a small population (i.e. a small amount of biomass in our study) reaching these patches and dispersing even further

to uncolonised habitat patches is negligible. Thus, both the simulation and the optimisation model focus on dispersal that occurs at the carrying capacity threshold. However, both models can be adapted to different dispersal patterns.

A detailed analysis of the model outcome can lead to a better understanding of range shifts. For example, the lower bounds found by the optimisation model can be used to identify important habitat patches for species dispersal and to evaluate the strength of the connection between certain habitat patches and their surroundings. This is especially interesting, as connectivity is a major concern for population survival and reduction of extinction risk (Saura and Pascual-Hortal, 2007; Fahrig and Merriam, 1985). The optimisation model allows to make decisions where and how to conserve habitat patches or landscape sections to secure a better habitat connectivity. At the same time, the model can be used to identify the optimal case to (re-)colonise habitat patches that arose or recovered due to climate change or other effects.

The model can readily be adapted to other dispersal simulation models and can thus be seen as a pioneer of a new set of models with a variety of applications such as dispersal prediction and habitat conservation and restoration.

Appendix

We prove that the MIP presented in this chapter provides the lower bounds on the dispersal time.

Theorem 3.3.1. *The MIP presented in this chapter finds a lower bound on the time steps needed by the simulation model to colonise the given destination habitats.*

Proof. Let a run of the simulation model that occupies the given destination habitats in $k \in \mathbb{N}$ time steps be given. We will construct a feasible solution for the MIP, which also occupies the destination habitats in k time steps.

In a first step set all variables equal to zero: $f(e) = 0$ for all $e \in E_{\text{TEN}}$, $f((s_t, S_t)) = 0$ for all destination habitats s and time steps t and $x(v) = 0$ for all $v \in V_{\text{TEN}}$.

Now, for each time step $t = 0, \dots, T$ set $f((u_t, v_{t+1}))$ as the amount of biomass that are sent from habitat u to v at time step t in the simulation.

Next set $f((Q, q_0)) = T_{SH}$ for all initial source habitats q . Beginning with time step $t = 0$, set $f((v_t, v_{t+1})) = \sum_{e \in \delta^-(v_t)} f(e)$ for all $v \in V$, $t = 0, \dots, T - 1$. For all nodes $v \in V_{\text{TEN}}$, if $\sum_{e \in \delta^-(v)} f(e) \geq T_{SH}$, set $x(v) = 1$. Next, set $f((s_t, S_t)) = x(s_t)$ for all destination habitats s and time steps t . For all $t = 0, \dots, T$, if $\sum_{s \in H_{\text{dest}}} f((s_t, S_t)) \geq |H_{\text{dest}}|$, set $x(S_t) = 1$ and $f((S_t, S)) = x(S_t)$.

We show that this is a feasible solution:

(1) $T_{SH}x(v) \leq \sum_{e \in \delta^-(v)} f(e) \forall v \in V_{\text{TEN}} \setminus \{Q, S_1, \dots, S_T, S\}$ holds, since

$$x(v) = \begin{cases} 1 & \text{if } T_{SH} \leq \sum_{e \in \delta^-(v)} f(e) \\ 0 & \text{else} \end{cases} \quad \forall v \in V_{\text{TEN}}.$$

(2) $f(v_t, v_{t+1}) = \sum_{e \in \delta^-(v_t)} f(e) \forall v \in V, t \in \{0, \dots, T\}$ holds by construction.

(3) Suppose that condition (3), $\sum_{e \in \delta^+(v)} \frac{f(e)}{1-c(e)p_m} \leq u(v)x(v) + \sum_{e \in \delta^-(v)} f(e) \forall v \in V_{\text{TEN}}$, is violated and there exists a habitat patch $v \in V$ and a time step t , such that $\sum_{e \in \delta^+(v_t)} \frac{f(e)}{1-c(e)p_m} > u(v_t)x(v_t) + \sum_{e \in \delta^-(v_t)} f(e)$. Then, since $c((v_t, v_{t+1})) =$

0, $\frac{f((v_t, v_{t+1}))}{1 - c((v_t, v_{t+1}))p_m} = f((v_t, v_{t+1})) = \sum_{e \in \delta^-(v_t)} f(e)$ and we obtain

$$\sum_{e \in \delta^+(v_t), e \neq (v_t, v_{t+1})} \frac{f(e)}{1 - c(e)p_m} > u(v_t)x(v_t)$$

Now two cases can occur.

- a) $x(v_t) = 0$. Then biomass is emitted, although $\sum_{e \in \delta^-(v_t)} f(e) < T_{SH}$. That is equivalent to v emitting biomass before becoming a source habitat in the simulation model:

Consider habitat v at time t in the simulation model. The amount of biomass in v at time step t is the amount of biomass in v at the previous time step $t - 1$ plus the amount of incoming biomass from connected source habitats plus a simulated population growth minus the outgoing amount of biomass, if v is a source habitat at time t . In the optimisation model this is by construction represented as

$$\sum_{e \in \delta^-(v_t)} f(e),$$

where the amount of biomass in v at time step $t - 1$ is given by $f((v_{t-1}, v_t))$ and the amount of incoming biomass from connected source habitats is given by

$$\sum_{e \in \delta^-(v_t), e \neq (v_{t+1}, v_t)} f(e).$$

Our assumption was that $x(v_t) = 0$. Thus, by construction $\sum_{e \in \delta^-(v_t)} f(e) < T_{SH}$ and again by construction this means that the amount of biomass in v at time $t - 1$ plus the incoming biomass in the simulation model also has to be less than T_{SH} . Now, according to the simulation model the population growth does not take part and the overall amount of biomass in v is less than T_{SH} in the current time step. Thus, in the simulation model biomass would have been emitted without v being a source habitat, which cannot happen.

- b) $x(v_t) = 1$. In this case more biomass is emitted than possible, which cannot happen in the simulation model either.

Thus, condition (3) is fulfilled for all nodes at all times.

(4)-(7) hold by definition.

- (8) All destination habitats are colonised at time step $k \leq T$ in the simulation. Thus, $\sum_{e \in \delta^-(s_k)} f(e) \geq T_{SH}$ and therefore $x(s_k) = 1$ for all destination habitats s . By construction $f((s_k, S_k)) = x(s_k) = 1$ for all $s \in \text{End}$ and thus $\sum_{s \in H_{\text{dest}}} f((s_k, S_k)) = |H_{\text{dest}}|$, which yields $x(S_k) = 1$. Again by construction $f((S_k, S)) = x(S_k) = 1$ and therefore $\sum_{t=0}^T f((S_t, S)) \geq 1$.

□

Chapter 4

Indicators for assessing the robustness of metapopulations against habitat loss

The content of this chapter has already been published in an international reviewed journal and can be accessed via the following link:

Heer, H., Streib, L., Schäfer, R. B., & Dieckmann, U. (2021). Indicators for assessing the robustness of metapopulations against habitat loss. *Ecological Indicators* 121 (2021): 106809. <https://doi.org/10.1016/j.ecolind.2020.106809>

Highlights

- Analysis of habitat loss coupled with metapopulation dynamics on habitat networks
- Different types of species, networks, and habitat loss studied
- Evaluation of 29 graph-theoretic indicators of metapopulation robustness against habitat loss
- Landscape-based networks respond to habitat loss qualitatively differently than standard networks
- Clustering coefficient is a good robustness indicator across all types of species, networks, and habitat loss

Abstract

Habitat loss and fragmentation resulting from changes in climate, land use, and pollution are main drivers of global biodiversity loss, as the survival of metapopulations relies on the ability of individuals to disperse among suitable habitat patches. To prioritize conservation efforts, methods are needed for evaluating the robustness of metapopulations against habitat loss. We therefore investigate this robustness for different degrees of habitat loss, different types of habitat loss (random, peripheral, and contagious), different types of habitat networks, and species differing in their local-extinction risks and dispersal ranges, with the latter two traits influencing metapopulation dynamics through local extinctions and the subsequent recolonization of patches, respectively. In particular, we analyse several standard network

types (with random, regular, small-world, or scale-free structure) and compare them with several alternative network types derived from real-world two-dimensional habitat landscapes (with random, clustered, or contiguous habitat allocation). To evaluate the robustness of metapopulations against habitat loss, we study how the fraction of colonized habitat patches changes with the fraction of lost habitat patches. Furthermore, we investigate how well 29 different graph-theoretic metrics of habitat networks can serve as indicators of metapopulation robustness against habitat loss – as this approach, where feasible, allows replacing complex simulation-based predictions with simple indicator-based predictions. We find that responses of species to habitat loss on the considered landscape-based habitat networks qualitatively differ from those on the considered standard habitat networks. This suggests that results obtained for the latter, albeit widely examined in the literature, can be unrepresentative and misleading. As expected, species with high risks of local extinction and short dispersal ranges are particularly vulnerable to habitat loss, across all considered types of habitat loss and habitat networks. The graph-theoretic network metric that best explains the robustness of metapopulations against habitat loss depends on the considered types of species, habitat networks, and habitat loss. None of the examined metrics give consistently reliable predictions under all circumstances. For sensitive species, characterized by high local-extinction risks and short dispersal ranges, a network's average clique size, redundancy, average degree, connectance, clustering coefficient, and average closeness centrality are the best indicators of metapopulation robustness. For landscape-based habitat networks, a network's average clique size, beta coefficient, clustering coefficient, redundancy, and cyclomatic number work best. For contagious habitat loss, the network type has a particularly strong impact on the species-specific robustness against habitat loss. In summary, our study introduces a method for evaluating the robustness of metapopulations against habitat loss and shows that a network's clustering coefficient, under a wide range of circumstances, is a particularly reliable indicator of this robustness.

Keywords

Habitat networks · Habitat loss · Network robustness · Metapopulation dynamics · Graph theory

4.1 Introduction

Habitat losses resulting from changes in climate and land use are among the main drivers of the ongoing global biodiversity crisis (Sala et al., 2000; Fahrig, 2003; Foley et al., 2005; Titeux et al., 2016; Lechner et al., 2017). The survival of metapopulations relies on suitable habitat and on the ability of individuals to disperse among different habitat patches. Small populations consisting of only few connected patches suffer from a risk of accidental local extinction through demographic stochasticity, are threatened by genetic impoverishment, and have lower chances of recovering from episodes of low abundance (Lande, 1988; Fagan and Holmes, 2006). Habitat loss tends to decrease habitat connectivity, impeding the movement of individuals among habitat patches (Turner and Ruscher, 1988; Kindlmann and Burel, 2008). A way to protect species from local extinctions due to habitat loss is to increase habitat connectivity by creating dispersal corridors increasing a landscape's permeability (Fahrig and Merriam, 1994; Hanski, 1999). As financial resources for designing and implementing conservation measures are usually limited, a quantitative assessment

of habitat connectivity in consideration of future habitat loss is necessary to prioritize conservation efforts (Cowling et al., 1999). Graph theory provides powerful tools to represent and analyse habitat connectivity in highly fragmented landscapes (Urban and Keitt, 2001; Dale and Fortin, 2010; Erős et al., 2012b; DeAngelis and Yurek, 2017). In this context, landscapes are seen as habitat networks, with habitat patches represented as nodes and possible dispersal pathways represented as links. Applying such tools from graph theory has numerous advantages: graph-based networks allow for complex computations using well-understood algorithms and can easily be used to evaluate different conservation scenarios or assess the impacts of global change. Furthermore, graph theory is being used ubiquitously over a wide range of disciplines ranging from social sciences, economics, and technology to biology. The resultant variety of existing tools can readily be applied in ecology (Urban et al., 2009; Newman, 2010). In ecology, techniques from graph theory and network analysis have mainly been used for investigating food webs and plant-animal mutualistic interactions (Landi et al., 2018), but are also common in landscape ecology (Fortuna et al., 2006). One of the main differences between habitat networks of metapopulations and food webs is the habitat networks' dependence on the underlying spatial landscape. While networks in general (such as food webs) can be independent of space, habitat networks are typically spatially embedded. Space is also relevant in many other disciplines and networks, giving rise to a focus across disciplines on so-called spatial networks. Spatial networks are defined as networks in which nodes are embedded in space and the probability of two nodes being connected decreases with distance (Gastner and Newman, 2006; Dale and Fortin, 2010; Barthélemy, 2014). Examples of such networks include the internet, power grids, transportation networks, and trade relations. Spatial networks tend to exhibit positive degree correlation, meaning that nodes have a higher probability of being connected with nodes of a similar degree (defined as the number of neighbours), as well as a high clustering coefficient, indicating that the neighbourhoods of nodes are particularly well connected (Watts and Strogatz, 1998; Newman, 2003; Boccaletti et al., 2006; Peyrard et al., 2008; Barthélemy, 2014). Naturally, habitat networks are spatial networks, as real-world habitat patches are located in a three-dimensional space that can often be approximated by a two-dimensional landscape, with patches at short distance having a higher probability of being connected than more distant patches. The robustness of networks against failure and attacks causing the loss of individual components (nodes or links) has been subject to many studies in various areas such as technology, transportation, and trade (Callaway et al., 2000; Cohen et al., 2000; Albert et al., 2001a; Cuadra et al., 2015; Gephart et al., 2016). A network's robustness is generally defined as "the ability of a network to avoid malfunctioning when a fraction of its constituents is damaged" (Boccaletti et al., 2006). Previous studies of the robustness of ecological networks have predominantly focused on food webs and mutualistic networks (Burgos et al., 2007; Evans et al., 2013). In the context of habitat networks, studies have applied graph theory to examine how the connectivity of networks affects the dispersal of various species (Calabrese and Fagan, 2004; Estrada and Bodin, 2008; Rayfield et al., 2011; Saunders et al., 2016; Upadhyay et al., 2017; Grech et al., 2018). While the majority of these studies have used variable species traits, such as dispersal ranges, to represent different species, each study has usually focused on only one underlying landscape. Furthermore, most preceding studies have modelled static scenarios, disregarding any metapopulation dynamics. Metapopulations, however, are highly dynamic and constantly adapt to changed conditions, e.g., after habitat loss. Therefore, analyses based on dynamic robustness would appear to be more appropriate (Martensen et al., 2017; Kun et al., 2019; Prima

et al., 2019; Shen et al., 2019). Recently, several studies of metapopulation robustness have taken such a perspective. Shen et al. (2019) simulated dispersal on regular networks, in which every habitat patch is connected to the same number of neighbouring patches, and observed that an increasing number of neighbours promotes metapopulation persistence. Prima et al. (2019) investigated habitat fragmentation and allowed for rewiring to account for dynamic robustness, finding that incorporating rewiring into network-robustness analyses is necessary and a naive application of graph theory to analyse habitat-network robustness may be inappropriate. Kun et al. (2019) examined the response of metapopulations to dynamic habitat loss and fragmentation, through which habitat sites transition from habitable to inhabitable and vice versa, and used their results to define five qualitatively different phases of landscape degradation. To examine dynamic robustness after the loss of habitat patches, our study therefore combines the simulation of metapopulation dynamics with the simulation of habitat-loss dynamics. We examine the consequences of the permanent loss of habitat patches located on different types of habitat networks for metapopulations of different generic species characterized by traits describing their local-extinction risks and dispersal ranges, by studying the resultant metapopulation dynamics including local extinction and recolonization. To evaluate the robustness of these metapopulations against habitat loss, we study how the fraction of colonized habitat patches changes with the fraction of lost habitat patches. These results are compared among and between standard networks commonly studied in graph theory, such as random, regular, small-world, and scale-free networks (Boccaletti et al., 2006; Newman, 2010) on the one hand and spatial networks defined through random, clustered, and contiguous habitat allocation on riverine landscapes (Streib et al., 2020) on the other hand. On this basis, we investigate how well different graph-theoretic metrics of habitat networks can serve as indicators of metapopulation robustness against habitat loss.

4.2 Methods

4.2.1 Structure of analyses

To evaluate metapopulation robustness against habitat loss, we study a model combining habitat loss with metapopulation dynamics. We consider three types of habitat loss – random, peripheral, and contagious – and remove a given fraction of habitat patches from a given habitat network accordingly. In our simulations, each remaining habitat patch, or network node, is either colonized or empty. All habitat patches removed by habitat loss are and stay empty, and all habitat patches remaining after habitat loss are colonized initially. For the latter patches, random local extinctions are considered, in a way that depends on the local-extinction risk of species and each patch’s neighbourhood. Empty habitat patches can be recolonized through dispersal from connected colonized habitat patches, in a way that depends on the dispersal range of species and each patch’s neighbourhood. These metapopulation dynamics are continued until a stationary distribution has been reached. From this we obtain the fraction of colonized habitat patches. We repeat such simulations of habitat loss and metapopulation dynamics for different degrees of habitat loss, to obtain a robustness curve describing the fraction of colonized habitat patches in dependence on the fraction of lost habitat patches. Based on this robustness curve, we use the ‘area under the curve’ (AUC) as a measure to quantify metapopulation robustness: the higher the fraction of colonized habitat patches across fractions of

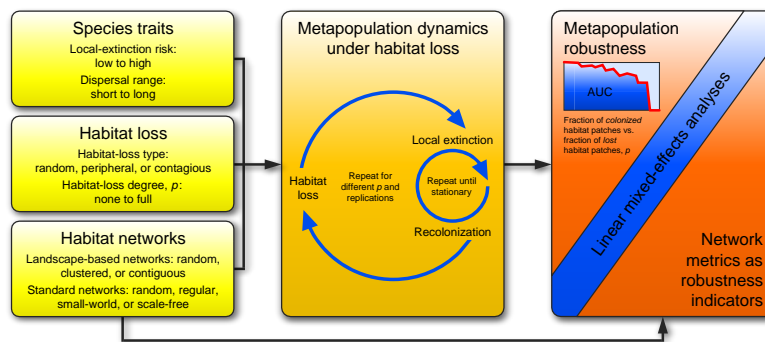


FIGURE 4.1: Schematic overview of the structure of analyses in this study. We consider species with different traits inhabiting different habitat networks experiencing different habitat loss (left panels). For the resultant wide variety of ecological settings, we study metapopulation dynamics involving the local extinction and recolonization of habitat patches, for different degrees of habitat loss (middle panel). On this basis, we obtain a measure of metapopulation robustness, as the ‘area under the curve’ (AUC) of the dependence of the fraction of colonized habitat patches on the fraction of lost habitat patches (right panel, top-left part) and examine how well various graph-theoretic metrics of the investigated habitat networks (right panel, bottom-right part) can serve as reliable indicators of metapopulation robustness (right panel, central part).

lost habitat patches, the higher the AUC, and thus the estimated metapopulation robustness. Metapopulation robustness is assessed in this way for different types of networks, different types of habitat loss, and different types of species. For each combination, simulations are replicated ten times to average over the sources of randomness affecting habitat networks, habitat loss, and metapopulation dynamics. The following subsections provide detailed specifications. Fig. 4.1 gives an overview of the structure of our analyses.

4.2.2 Habitat networks

We consider a wide range of habitat networks qualitatively differing in their structure. In particular, we compare three types of landscape-based networks with four types of standard networks commonly investigated in network theory. Landscape-based habitat networks are created as described by Streib et al. (2020) for stream landscapes from South-West Germany considering a generic insect species inhabiting habitat patches situated along the riverine parts of these landscapes. See Fig. 4.9 for a map of these stream landscapes and Appendix Section A.1 for detailed information on their specification. We examine three types of landscape-based networks varying in habitat-patch arrangements ranging from (1) random (with all habitat patches randomly selected along streams with equal probability), over (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) to (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 allocation of the habitat patches compared to the clustered allocation). Dispersal pathways linking these patches are constructed by a least-cost path analysis of land-use scenarios. See Fig. 4.2a-c for examples of the resultant landscape-based habitat networks and Appendix Section A.2 and Streib et al. (2020) for detailed information on their specification. Standard networks with (1) random, (2) regular, (3) small-world, and (4) scale-free structures are created using algorithms implemented in the Python package NetworkX version 1.10 (Hagberg et al., 2008a). Parameters are set to create networks that are similar to the landscape-based networks in terms of the number of habitat patches (nodes) and corresponding dispersal pathways (links). Every link is assigned random dispersal costs with a distribution similar to the landscape-based networks. In random networks, two nodes are connected with a given probability. The degrees of nodes, defined as the number of other nodes to which they are connected, thus follow a binomial distribution, which approaches a normal distribution with narrow variance as the number of nodes becomes large. These random networks are generated using the algorithm by (Erdős and Rényi, 1960). Regular networks are networks in which every node has the same degree (Newman, 2010). For example, triangular, square, and hexagonal grids define regular networks, with nodes in each grid cell and links existing wherever grid cells are adjacent. Small-world networks are a mixture of random and regular networks, capturing the small-world phenomenon well-known from the social sciences (Watts and Strogatz, 1998; Boccaletti et al., 2006; Newman, 2010). While most nodes are not connected to each other, the neighbours of a node are connected with a probability that is higher than in random networks. In other words, small-world networks are highly clustered while at the same time exhibiting low average shortest-path distances between nodes. We use the algorithm by Newman and Watts (1999) to construct small-world networks. Scale-free networks consist of a few nodes (so-called hubs) that are connected to many other nodes, while most nodes have only very few neighbours, with the degrees of nodes following a power-law distribution (Barabási and Albert, 1999; Barabási and Bonabeau, 2003). We use the Barabási–Albert preferential attachment model to construct scale-free networks (Barabási and Albert, 1999). See Fig. 4.2d-g for examples of the resultant standard habitat networks and Appendix Section A.3 for detailed information on their specification. In total, we analyse 1750 networks distributed over seven ensembles each containing 250 networks for each of the seven network types, with the number of nodes ranging between 54 and 111 and the number of links ranging between 39 and 324. See Fig. 4.2 for examples of all network types and Appendix Sections A.2 and A.3 for detailed information on the specification of the network ensembles.

4.2.3 Habitat loss

We investigate three types of habitat loss. Random habitat loss removes each habitat patch with equal probability p . Peripheral habitat loss removes peripheral habitat patches with higher probability than central habitat patches. For this purpose, we define the centrality of a patch as its betweenness centrality (Freeman, 1977; Estrada and Bodin, 2008), which measures how many of the shortest paths between all pairs of patches pass through a focal patch. When peripheral habitat loss with probability p is applied to a network with n habitat patches, a total of pn patches are removed, which are chosen with relative probabilities $1 - b(v)$, where $b(v)$ is the betweenness centrality of patch v . Contagious habitat loss removes adjacent habitat patches with higher probability than non-adjacent habitat patches. First, a single habitat patch

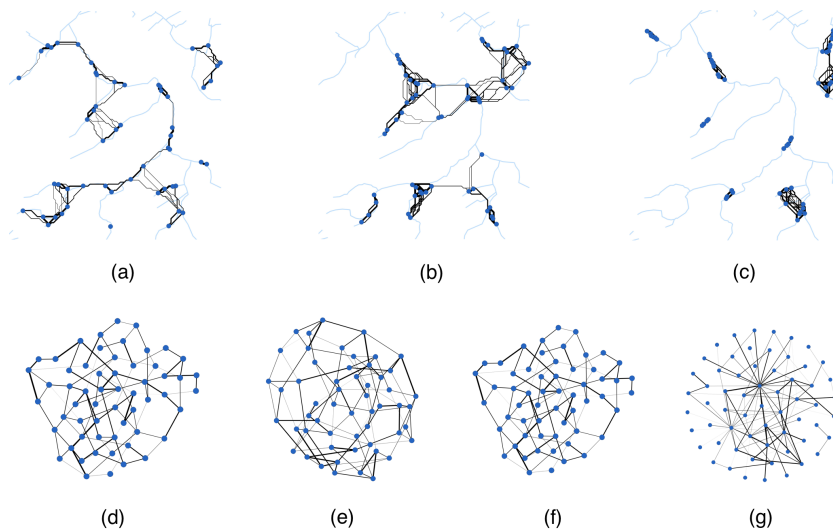


FIGURE 4.2: Examples of the analysed seven types of habitat networks. (a-c) Landscape-based networks, with (a) random, (b) clustered, and (c) contiguous habitat allocation Streib et al. (2020). (d-g) Standard networks with (d) random, (e) regular, (f) small-world, and (g) scale-free structure. Blue dots represent habitat patches, light-blue lines depict the underlying stream landscape, and black lines indicate dispersal pathways. The width of these lines represents the dispersal costs with thick lines indicating lower costs and thin lines indicating higher costs.

chosen with equal probability from among all habitat patches is removed. Next, habitat patches adjacent to a lost patch are removed recursively, from lowest to highest dispersal costs using the breadth-first search algorithm, stopping once pn patches are removed. These two steps are repeated until pn patches are removed.

4.2.4 Metapopulation dynamics

For a given level of habitat loss, local extinctions are simulated on the remaining habitat network, by considering local extinctions in habitat patches and the recolonization of habitat patches.

As the long-term survival of a population in a habitat patch highly depends on its potential to exchange individuals with neighbouring patches (Lande, 1988; Fagan and Holmes, 2006), we assume that the risk of local extinction in a focal patch is positively correlated with the number of patches connected to it within its neighbourhood. We use the size of cliques as a measure of these connections: a clique in a network is a subset of nodes such that every two nodes in this subset are connected with each other (Alba, 1973; Newman, 2010). Denoting by $c(v)$ the size of the largest clique that contains the node v we assume that the population in v goes extinct with probability

$$p_{\text{ext}}(v) = a^{1-c(v)},$$

where $a > 1$ is a species-specific parameter governing the local-extinction risk of a species. We can thus interpret a as the factor by which the local-extinction risk in v geometrically decreases with every additional node in $c(v)$. For example, when

v has only one clique-level neighbour, $c(v) = 2$ and $p_{\text{ext}}(v) = a^{-1}$, while when $c(v)$ increases by the increment 1 from 2 to 3, $p_{\text{ext}}(v)$ drops by the factor a from a^{-1} to a^{-2} . We investigate values of a ranging from 2 to 9 to account for the different local-extinction risks of different species. We can think of these risks decreasing with increasing $c(v)$ more slowly for habitat specialists (small values of a) and more rapidly for habitat generalists (large values of a).

Empty habitat patches can be recolonised from connected colonised patches. Recolonisation is modelled with the help of a Gaussian dispersal kernel (Nathan et al., 2012; Chapman et al., 2007) and we assume that an empty habitat patch v becomes recolonised from a colonised patch w with probability

$$p_{\text{col}}(v, w) = \frac{m_{vw}}{\sum_{u \in V} m_{uw}},$$

where $m_{vw} = \exp(-\frac{1}{2}d_{vw}^2/\sigma^2)$ is the dispersal kernel, V the set of all network nodes, d_{vw} the distance between habitat patches v and w in terms of dispersal costs, and $\sigma > 0$ a species-specific dispersal parameter governing the dispersal range of a species. We can thus interpret σ as the dispersal cost at which the dispersal probability drops to $1/\sqrt{e} = 60.7\%$ of its maximum. Similar to a , we investigate values of σ ranging from 2 to 9 to account for the different dispersal capacities of different species. We can think of these capacities as being low for poor dispersers (small values of σ) and high for good dispersers (large values for σ).

These local extinctions in and recolonizations of habitat patches are simulated alternately until a stationary frequency of colonized patches is reached.

4.2.5 Statistical analyses

We describe the structure of the habitat networks using a variety of metrics common in graph theory, such as the number of nodes and links, the average centrality of nodes, a network's clustering coefficient, and the average clique size. All of these metrics have been implemented in the Python package NetworkX version 1.10 (Hagberg et al., 2008a). Table 4.1 provides a list of the examined metrics and indicates the corresponding NetworkX algorithms. We use the statistical software R version 3.5.1 (R Core Team, 2018) in combination with the two R packages lme4 version 1.1.19 (Bates et al., 2014) and sjstats version 0.17.2 (Lüdtke, 2018) to perform a linear mixed-effects analysis of the relationship between a metapopulation's robustness, measured by the AUC as described above, and each considered network metric. For this, we fit linear mixed-effects models (LMMs) with metapopulation robustness as the response variable and one of the network metrics as the predictor variable (i.e., fixed effect). We account for the statistical dependence associated with the grouped structure in our simulation data that arises from using 10 replications per setup (each defined by a fixed combination of species, networks, and habitat loss) by treating these replication setups as random effects affecting the LMM intercepts. For details on LMMs, see Faraway (2016). Based on the estimated LMMs, we standardize the LMM slopes using standardized regression coefficients (denoted as b_{std} throughout the present study) to allow for a comparison of the strength of the relationship across different network metrics (Landis, 2005). These standardized regression coefficients describe the predicted change in metapopulation robustness resulting from a unit change in the considered network metric when both quantities are expressed in standard-deviation units. To identify network metrics that exhibit an overall strongly positive or negative relationship with metapopulation robustness irrespective of the considered network types, we average the regression coefficients

over all network types in the considered set of network types. In case of contradictory relationships (i.e., if one network type exhibits a positive relationship with a network metric whereas another network type exhibits a negative relationship with the same metric), we set the mean regression coefficient to 0 to indicate that the relationship is unreliable. To examine the relative importance of network type, dispersal range, and local-extinction risk for determining metapopulation robustness, we fit LMMs with metapopulation robustness as the response variable and network type, σ , and a as the predictor variables (i.e., fixed effects). As described above, we treat the replication setups as random effects affecting the LMM intercepts. Based on the estimated LMMs, we measure effect size as the proportion η^2 of the total variance in metapopulation robustness that can be attributed separately to each of the three predictor variables.

Metric	Specification	NetworkX functions involved
Betweenness centrality	Fraction of shortest paths between the two nodes of all pairs of different nodes that pass through a focal node	<code>betweenness_centrality</code>
Closeness centrality	Reciprocal of the sum of the shortest-path lengths from a focal node to every other node, averaged over all other nodes	<code>closeness_centrality</code>
Clique number	Size of largest clique	<code>graph_clique_number</code>
Degree	Number of neighbours of a focal node	<code>degree_centrality</code>
Average clique size	Size of largest clique a node is part of, averaged over all nodes	<code>node_clique_number</code>
Average shortest path length	Shortest-path length between the two nodes of all pairs of different nodes, averaged over all such pairs	<code>average_shortest_path_length</code>
Average connectivity	Minimum number of nodes that must be removed to disconnect two non-adjacent nodes, averaged over all such pairs	<code>average_node_connectivity</code>
Beta coefficient	Ratio of the number of links to the number of nodes	<code>number_of_nodes</code> , <code>number_of_edges</code>
Size of centre	Number of nodes with minimum eccentricity, with the eccentricity of each node defined as the maximum of its shortest-path lengths to all other nodes	<code>center</code>
Average number of cliques	Number of cliques a node is part of, averaged over all nodes	<code>cliques_containing_node</code>
Clustering coefficient	Ratio of the number of existing triangles in a node's neighbourhood to the number of all possible triangles in this neighbourhood, averaged over all nodes	<code>average_clustering</code>
Number of components	Number of connected components	<code>number_connected_components</code>

Cyclomatic number	Maximum number of links that can be removed without increasing the number of components	number_of_nodes, number_of_edges, number_connected_components
Number of links	Number of links (edges)	number_of_edges
Number of nodes	Number of nodes	number_of_nodes
Size of periphery	Number of nodes with maximum eccentricity, with the eccentricity of each node defined as the maximum of its shortest-path lengths to all other nodes	periphery
Periphery-to-centre ratio	Ratio of the size of the periphery to the size of the centre	periphery, center
Connectance	Ratio of the number of existing links to the number of all possible links	number_of_nodes, number_of_edges
Redundancy	Ratio of the number of existing circuits to the number of all possible circuits, with circuits defined as closed paths through distinct nodes	number_of_nodes, number_of_edges
Skewness	Skewness of degree distribution	degree.values, scipy.stats.skew
Degree correlation	Pearson correlation coefficient of the degree pairs of all adjacent nodes	degree_assortativity_coefficient

TABLE 4.1: Graph-theoretic metrics investigated in this study. The listed quantities are network-level metrics, except for the node-level metrics of betweenness centrality, closeness centrality, and degree, for which our analysis considers the network-level metrics given by the average, minimum, maximum, and range of the node-level metrics.

4.3 Results

4.3.1 Landscape-based habitat networks respond to habitat loss qualitatively differently than standard networks

We find that the responses of landscape-based habitat networks to habitat loss qualitatively differ from those of standard habitat networks, with smaller differences in responses occurring within these two groups of network types. First, metapopulations on landscape-based networks are more robust compared to those on standard networks. Across all types of habitat loss, metapopulation robustness on landscape-based networks has a mean of 0.61 with a standard deviation of 0.19, whereas on standard networks it has a mean of only 0.42 with a standard deviation of 0.17 (Fig. 4.3b). Second, several of the tested network metrics can serve as reliable indicators of metapopulation robustness across the three types of landscape-based networks (Fig. 4.5), but are not suitable to predict metapopulation robustness across all network types including the four types of standard networks. These findings may be explained by the structural differences we find between landscape-based and standard networks. In line with what is expected for spatial networks in general, the landscape-based networks exhibit particularly high clustering coefficients

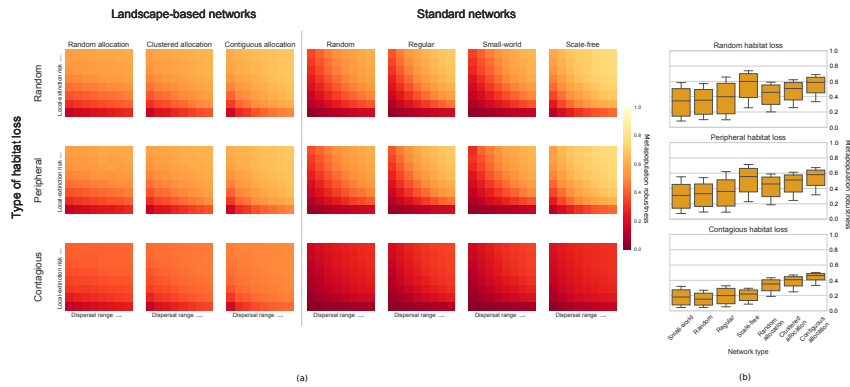


FIGURE 4.3: Sensitive species, characterized by low metapopulation robustness to habitat loss, have short dispersal ranges and high local-extinction risks. (a) Dependence of metapopulation robustness on the dispersal ranges (vertical axes, parameter σ varied from 2 to 9 in steps of 1) and local-extinction risks (horizontal axis, parameter a varied from 2 to 9 in steps of 1) of species for seven different network types (columns) and three types of habitat loss (rows). Each cell shows the metapopulation robustness averaged over 10 replications (see Methods), with light-red colours indicating high robustness and dark-red colours indicating low robustness. (b) Corresponding distributions of metapopulation robustness according to network types (columns) and types of habitat loss (rows). The central line of each box indicates the median, the lower and upper edges of each box indicate the interquartile range, and the whiskers indicate the 10% and 90% quantiles.

(mean of 0.65) and high positive degree correlation (mean of 0.7), whereas the standard networks show almost no clustering and a low negative to low positive degree correlation with mean clustering coefficients between 0.15 (scale-free networks) and 0.05 (all others) and mean degree correlations between -0.2 (scale-free networks) and +0.1 (small-world networks). Our results suggest that none of the standard network types – be they of random, regular, small-world, or scale-free structure – are suitable representatives of real-world habitat networks, despite their widespread use in a broad range of studies and disciplines. Overlooking this fundamental limitation when evaluating conservation measures in response to habitat loss is risky at best and misleading at worst.

4.3.2 Species with short dispersal ranges and high local-extinction risks are particularly vulnerable to habitat loss

Across all types of networks and habitat loss, species with short dispersal ranges or high local-extinction risks are especially vulnerable to habitat loss (Fig. 4.3a). Such species, which may be called sensitive, occupy an L-shaped parameter region in which either dispersal ranges are short or local extinctions are likely (Fig. 4.3a). As expected, increasing the dispersal range or decreasing the local extinction risk of a species results in increasing metapopulation robustness, again across all types of networks and habitat loss (Fig. 4.3a). Contagious habitat loss, which might be considered to describe best what is happening to real-world habitat networks, brings out particularly clearly the qualitative differences described in Section 3.1 between landscape-based and standard networks (Fig. 4.3b).

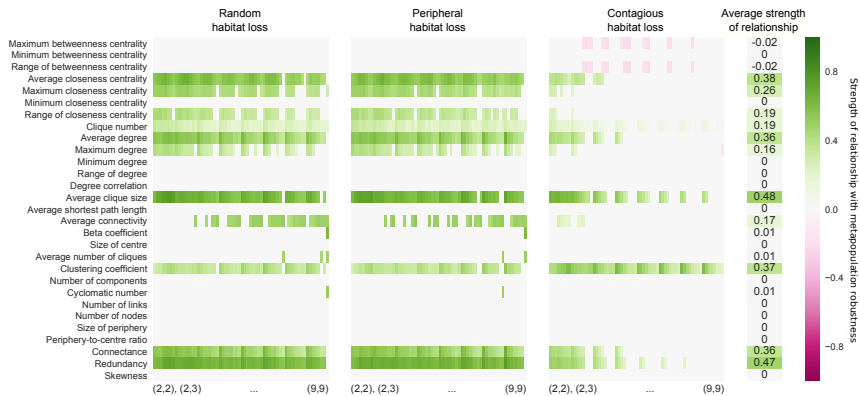


FIGURE 4.4: When considering all types of species, networks, and habitat loss simultaneously, only the clustering coefficient can serve as a reliable indicator of metapopulation robustness. A total of 29 network metrics are shown in the rows, and the three types of habitat loss are shown in the panels. The colour of each cell indicates the strength of the relationship between metapopulation robustness and the considered network metric, measured as the standardized regression coefficient, averaged over all network types (this average strength of the relationship is set to 0 when the sign of the strength of the relationship is not consistent across network types). Within each row of each panel, the species-specific parameters governing local-extinction risks and dispersal ranges, respectively, are changing across the columns in steps of 1 between 2 and 9 first for a and then for σ . The last column shows averages over all cells of a row.

4.3.3 The power of network metrics to predict robustness depends on the types of species, networks, and habitat loss

We find that the clustering coefficient is the only reliable indicator of metapopulation robustness across all considered setups in terms of species, networks, and habitat loss (Fig. 4.4), with a mean standardised regression coefficient of $b_{\text{std}} = 0.37$. At the same time, if we focus on certain subsets of setups, additional reliable indicators of metapopulation robustness can be identified (Figs. 4.5- 4.7).

Sensitive species

When focusing on species with short dispersal ranges (i.e., $\sigma=2$) or high local-extinction risks (i.e., $a=2$), several specific network metrics emerge as reliable indicators of metapopulation robustness (Fig. 4.5). The best indicators are the average clique size ($b_{\text{std}} = 0.65$), the redundancy ($b_{\text{std}} = 0.56$), the average degree ($b_{\text{std}} = 0.48$), the connectance ($b_{\text{std}} = 0.48$), the clustering coefficient ($b_{\text{std}} = 0.45$), and the average closeness centrality ($b_{\text{std}} = 0.44$). For interpreting the strong result for the average clique size, we ought to keep in mind that in our model of metapopulation dynamics the size of the largest clique a node is part of plays a direct role in determining the local-extinction probability in a habitat patch.

Landscape-based habitat networks

When focusing on landscape-based networks, we find a variety of network metrics that are reliable indicators of metapopulation robustness for all types of species and habitat loss (Fig. 4.6). The best indicators are the average clique size ($b_{\text{std}} = 0.61$),

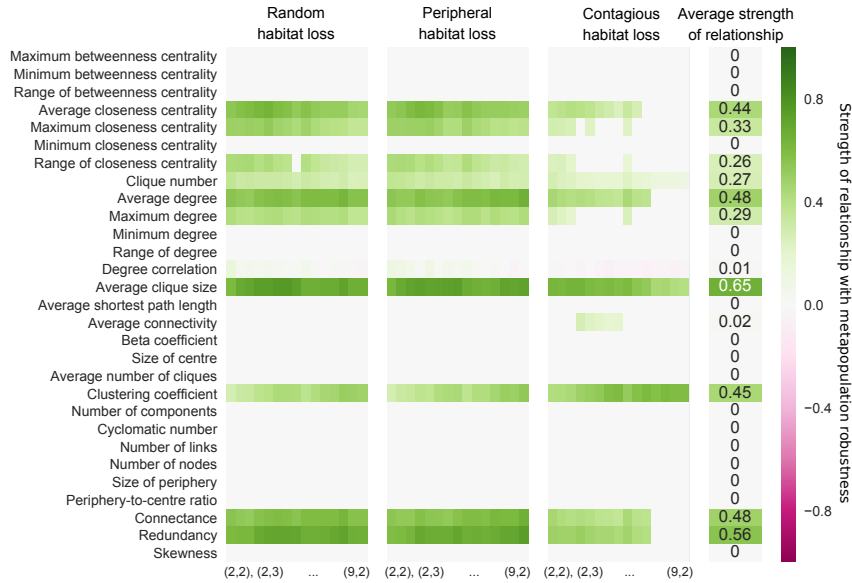


FIGURE 4.5: When considering sensitive species, the average clique size, the redundancy, the average degree, the connectance, the clustering coefficient, and the average closeness centrality are the best indicators of metapopulation robustness. All figure elements as in Fig. 4.

the ratio of links to nodes, also called the beta coefficient ($b_{std} = 0.58$), the clustering coefficient ($b_{std} = 0.54$), the redundancy ($b_{std} = 0.52$), and the cyclomatic number ($b_{std} = 0.48$). Similar to the results above, the relationships between metapopulation robustness and the different network metrics are often stronger for sensitive species, i.e., for species that are weak dispersers or habitat specialists.

Contagious habitat loss

When focusing on contagious habitat loss (Fig. 4.7), the best indicator of metapopulation robustness is the clustering coefficient ($b_{std} = 0.42$). For sensitive species, the average clique size ($b_{std} = 0.27$), the redundancy ($b_{std} = 0.15$), the connectance ($b_{std} = 0.11$), and the average degree ($b_{std} = 0.11$) are good indicators as well.

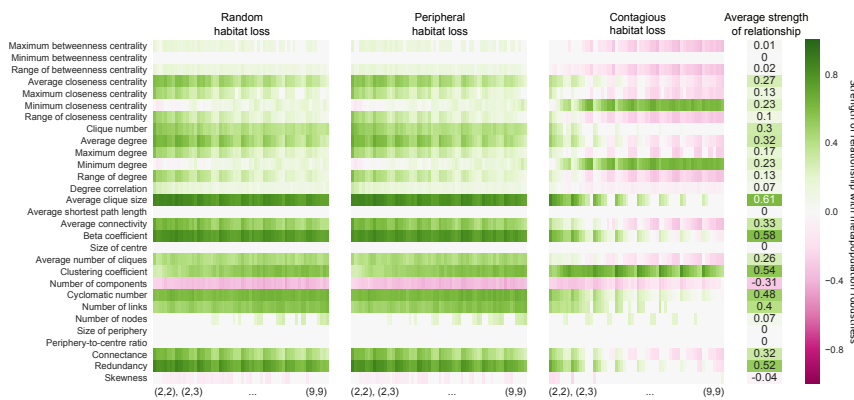


FIGURE 4.6: When considering landscape-based networks, the average clique size, the beta coefficient, the clustering coefficient, the redundancy, and the cyclomatic number are the best indicators of metapopulation robustness. All figure elements as in Fig. 4.4.

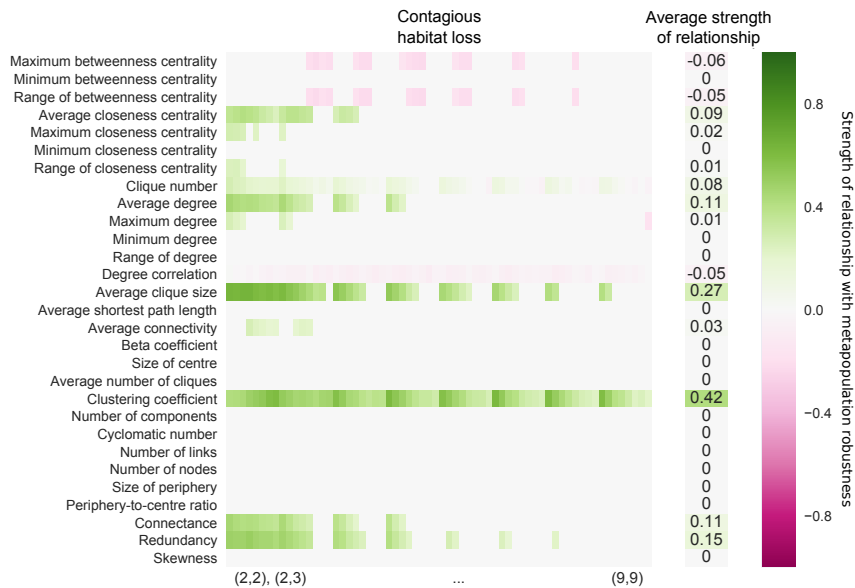


FIGURE 4.7: When considering contagious habitat loss, the clustering coefficient is the best indicator of metapopulation robustness. For sensitive species, the redundancy, the average clique size, the connectance, and the average degree are further appropriate indicators. All figure elements as in Fig. 4.4.

For contagious habitat loss, the network type, dispersal range, and local-extinction risk jointly account for 90% of the variance in metapopulation robustness (Fig. 4.8). Within these 90%, the network type has a particularly large impact on metapopulation robustness, accounting for 58% of the variance alone. For random and peripheral habitat loss, about 85% of the variance in metapopulation robustness are explained by network type, dispersal range, and local-extinction risk together. In contrast to contagious habitat loss, the network type here accounts for less than 20% of the variance, while the local-extinction risk accounts for about 50% of the variance.

4.4 Discussion

Our results have shown high metapopulation robustness against habitat loss for species with strong dispersal capacities and low local-extinction risks across all types of networks and habitat loss. We have found that none of the analysed network metrics are reliable indicators of metapopulation robustness across all considered types of species, networks, and habitat loss. At the same time, when focusing our analysis on sensitive species (characterized by short dispersal ranges and high local-extinction risks) or landscape-based networks (characterized by high clustering and degree correlations), we could identify a variety of reliable indicators of metapopulation robustness. This indicates that the structure of habitat networks has a higher impact on the metapopulation robustness of particularly vulnerable species inhabiting real-world habitat landscapes. Thus, the increase in the predictability of metapopulation robustness by network metrics occurs precisely where the practical needs for such predictability are highest. We have identified a network's clustering coefficient,

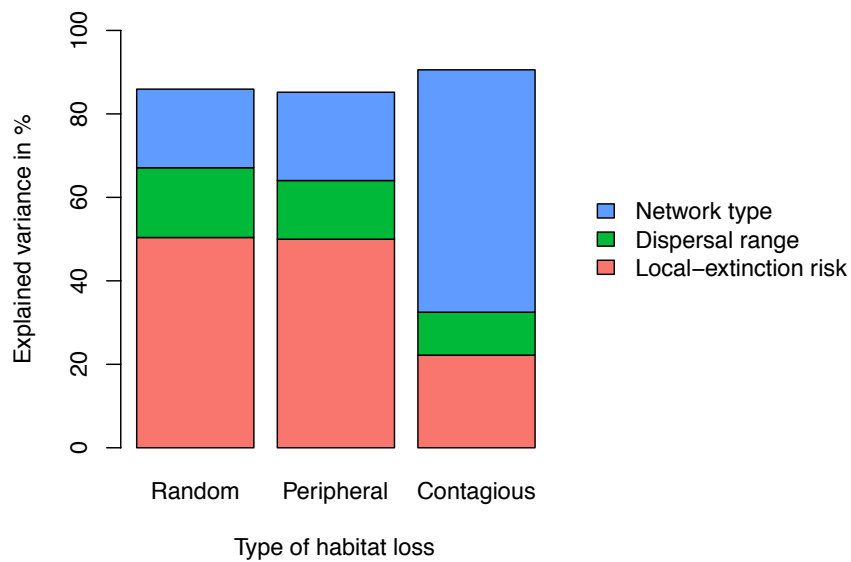


FIGURE 4.8: Variance in metapopulation robustness explained by the network type, dispersal range, and local-extinction risk. The total explained variance is high, reaching 85% or more for all types of habitat loss. The network type has a particularly strong impact on the robustness of metapopulations to contagious habitat loss.

its average clique size, and its beta coefficient as reliable indicators of metapopulation robustness on landscape-based habitat networks. These findings are in agreement with other studies suggesting the small-world characteristics of networks as a sign of robustness, because small-world networks are characterized by clustering coefficients that are higher than expected for random networks in conjunction with average shortest path lengths that are similar to or shorter than expected for random networks (Almpanidou et al., 2014; Fox and Bellwood, 2014; Prima et al., 2019). However, the small-world characteristic is not a quantitative metric and the clustering coefficient is, thus, a more informative and precise measure. According to our results, the average shortest path length, in contrast to the clustering coefficient, does not exhibit a reliable relationship with metapopulation robustness across the different network types examined in our study, even though it is sometimes also considered as being characteristic of small-world networks. In agreement with our findings, Ash and Newth (2007) developed networks that are particularly robust against cascading failures and noticed their high clustering coefficient. In general, scale-free networks are seen as being particularly robust against random node failure, since the probability of a random network node having a high degree – and its failure thus causing the loss of many network links – is especially low for networks of this type (Albert et al., 2001a; Barabási and Bonabeau, 2003). Our analysis, however, does not support the notion that metapopulations on scale-free networks or small-world networks are particularly robust to habitat loss. Instead, we find that metapopulations on landscape-based networks are most robust to habitat loss. This indicates that the embedding of a network's structure in two-dimensional landscapes (which also leads to higher clustering) may play a crucial role in determining its robustness. The seeming discrepancy between the surmised robustness of small-world networks and scale-free networks on the one hand and our results on the other could partially originate also from subtly different notions of robustness: while many studies and verbal accounts of network stability consider what prevents networks from fragmenting into disconnected subnetworks, such large-scale considerations are not necessarily very relevant for a metapopulation's response to habitat loss, provided the resultant fragments each remain large enough to support viable metapopulations. Studies of networks describing food webs have identified connectance (i.e., the ratio of existing links in a network to all possible links) as a good indicator of robustness against species invasion (Romanuk et al., 2017). Although connectance shows a strong relationship with the metapopulation robustness in our analysis of landscape-based networks, we have found it to be a rather poor indicator when considering all network types. Yet, when focusing on peripheral habitat loss alone, connectance showed, together with the average patch degree, the best results. We hope that our results will prove useful for evaluating and prioritizing species, metapopulations, and geographic areas that are particularly vulnerable to habitat loss. Once habitat networks are mapped out, practitioners can use the specific network metrics highlighted by our study to identify those networks that are least robust against habitat loss and thus need most protection or remedial action. Furthermore, conservation efforts can be improved by assessing which areas of a larger landscape need to be modified, and how, so as to raise a focal network metric positively related with metapopulation robustness, since this can be expected, based on our results, to improve the robustness of metapopulations inhabiting the whole landscape. As we have found the clustering coefficient to display a reliably strong relationship with metapopulation robustness throughout the majority of settings we have investigated and since it has been identified as a good proxy of network robustness also in other studies (Ash and Newth, 2007; Almpanidou et al., 2014; Fox and

Bellwood, 2014; Prima et al., 2019), we particularly recommend using the clustering coefficient in practical applications. If and when more information about a specific metapopulation is known – such as the applicable species traits, network type, or habitat-loss type –, our results can guide the selection of another network metric that is even better suited than the clustering coefficient for the scenario at hand. One of the main findings of our study is that standard networks, which lack a spatial reference to an underlying landscape, respond qualitatively differently to habitat loss compared to landscape-based habitat networks. We suggest that this fundamental limitation renders standard networks, despite their widespread use in many areas of network science, unsuitable for studying the impacts of habitat loss on metapopulations. When this fundamental limitation is not recognized, resultant assessments of conservation measures are likely to be seriously flawed, potentially leading to the avoidable sacrifice of biodiversity and/or loss of resources. Future studies could strive to investigate a broader range of spatial networks, as well as of habitat networks derived from real-world landscapes, to evaluate the wider generalizability of our results.

4.5 Conclusions

The following list provides a summary of our main conclusions:

- Responses of species to habitat loss on the considered landscape-based habitat networks are qualitatively different from those on the considered standard habitat networks.
- Species with high risks of local extinction and short dispersal ranges are particularly vulnerable to habitat loss, across all considered types of habitat loss and habitat networks.
- The graph-theoretic network metric that best explains the robustness of metapopulations against habitat loss depends on the considered types of species, habitat networks, and habitat loss.
- For sensitive species, characterized by high local-extinction risks and short dispersal ranges, a network's average clique size, redundancy, average degree, connectance, clustering coefficient, and average closeness centrality are the best indicators of metapopulation robustness.
- For landscape-based habitat networks, a network's average clique size, beta coefficient, clustering coefficient, redundancy, and cyclomatic number work best.
- For contagious habitat loss, the network type has a particularly strong impact on the species-specific robustness against habitat loss.
- Conservation biologists and landscape managers can benefit from this study as it helps evaluate conservation measures for protecting habitat networks and identifying species, metapopulations, and geographic areas that are in need of the most protection.

Acknowledgement

The authors thankfully acknowledge support by the B-M-U Graduate Academy of the University of Koblenz-Landau.

Appendix: Specification of Habitat Networks

In this appendix, we provide detailed information on the specification of the analysed seven types of habitat networks. For the landscape-based networks investigated in the present study, we use the habitat networks developed by Streib et al. (2020) with two modifications: First, we consider only one set of proportions of land-use types (as detailed below) and only one proportion of habitat patches (as detailed below). Second, we reduce the maximal dispersal costs, to obtain networks with fewer links (as detailed below). We briefly summarize how we have followed Streib et al. (2020), to make it easier for readers to understand and reproduce the analysed habitat landscapes (Section A.1) and landscape-based networks (Section A.2). We then specify the analysed standard networks (Section A.3). All network types are specified based on (1) habitat patches, (2) links between patches, and (3) dispersal costs along links. The landscape-based networks are defined by first specifying the habitat patches and dispersal costs between patches, with the links then jointly determined by these. The standard networks are defined by first specifying the habitat patches and links between patches, with the dispersal costs then assigned to the links. Tables 4.2- 4.4 show the distribution of the number of nodes, links, and dispersal costs for each of the seven network types. All analysed networks, as well as the Python code for all analyses in the present study, are available online at https://github.com/hheer/HabitatRobustness_Indicators.

A.1. Habitat landscapes

The landscape-based habitat networks are derived from a real-world stream landscape (Section 2.1.1 in Streib et al., 2020). For this purpose, we use a 50 km × 50 km area of the stream landscape of Rhineland-Palatinate, Germany, as this landscape is of local interest and associated data is readily available (Fig. 4.9). The total stream landscape is divided into 25 landscape tiles each measuring 10 km × 10 km to obtain a variety of dense or sparse stream landscapes with differing stream network structures (Fig. 4.9). These two-dimensional stream landscapes provide necessary, although not sufficient, information for specifying the analysed landscape-based habitat networks, which means that the real-world stream landscapes endow the analysed landscape-based habitat networks with important real-world properties. The landscapes surrounding the streams are created by randomly assigning one of the three land-use types ‘open agricultural land’, ‘forestry land’, and ‘urban area’ to each 25 m × 25 m pixel of the 50 km × 50 km area without spatial auto-correlation; this corresponds to the landscape-type configuration ‘random’ in Streib et al. (2020). We use land-use proportions of 25% for open agricultural land, 25% for forestry land, and 50% for urban area, to describe mixed landscapes with relatively large shares of urban area. The stream network is embedded into these landscapes by assigning a fourth land-use type ‘aquatic area’ to each pixel intersected by a stream. Each of the four land-use types are assigned dispersal costs to represent the permeability of the landscapes for a generic insect species inhabiting aquatic areas. We use dispersal costs of 25 for aquatic pixels, of 50 for agricultural pixels, of 75 for forestry pixels,

and of 100 for urban pixels (Section 2.1.1 in Streib et al., 2020). Following these assignments, the landscapes are resampled to a new pixel size of $100\text{ m} \times 100\text{ m}$ by taking the proportions of land-use types and the averages of dispersal costs over all $4 \times 4 = 16$ involved smaller pixels to obtain a broad variety of mixed land-use types and associated dispersal costs (Section 2.1.1 in Streib et al., 2020).

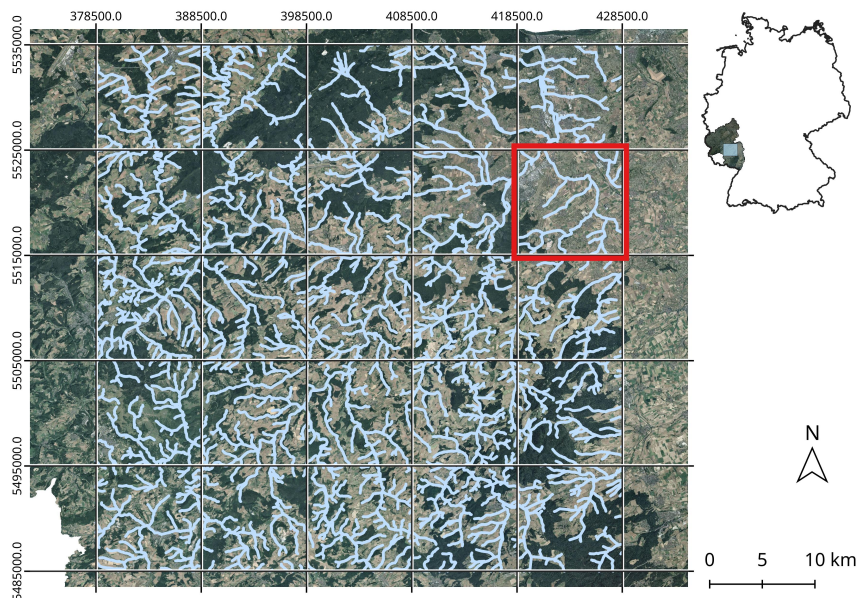


FIGURE 4.9: Stream landscape in the German federal state of Rhineland-Palatinate used for deriving the landscape-based habitat networks analysed in the present study. Streams locations, indicated in light-blue, are obtained from the database ‘Gewässernetz (gesamt)’ (Rhineland-Palatinate Ministry of the Interior and for Sports, 2020a). The background shows a satellite image obtained from the web map service ‘Luftbild RP Basisdienst’ (Rhineland-Palatinate Ministry of the Interior and for Sports, 2020b). The grid lines indicate the division of the total $50\text{ km} \times 50\text{ km}$ landscape into 25 landscape tiles each measuring $10\text{ km} \times 10\text{ km}$. The map uses the coordinate reference system ‘ETRS89 / UTM zone 32N’ with EPSG code 25832. Longitudinal positions are shown horizontally along the top edge and latitudinal positions are shown vertically along the left edge, both measured in metres. The overview map in the top right shows the location of the stream landscape (light-blue rectangle) within the state of Rhineland-Palatinate (satellite image) within Germany. The tile surrounded in red corresponds to the maps shown in Fig. 4.2a-c.

A.2. Landscape-based habitat networks

Habitat patches

For each of the 25 $10\text{ km} \times 10\text{ km}$ landscape tiles with embedded stream structures and pixels of size $100\text{ m} \times 100\text{ m}$, 10% of the aquatic pixels are selected as habitat patches (Section 2.1.2 in Streib et al., 2020). Reflecting the different stream structures in the different landscape tiles, this results in habitat networks with 54 to 111 habitat patches. The particular aquatic pixels selected as habitat patches are identified using one of three different algorithms, giving rise to three types of landscape-based

habitat networks with random, clustered, and contiguous habitat allocation. For random habitat allocation, 10% of all aquatic pixels are randomly selected with equal probability. For clustered habitat allocation, 5% of all aquatic pixels are randomly selected with equal probability, and 5% of all aquatic pixels are randomly selected with equal probability and without repetitions within a radius of 500 m around the initially selected ones. For contiguous habitat allocation, we follow the same principle as for clustered habitat allocation, except that merely 2.5% of all aquatic pixels are randomly selected initially, resulting in a more contiguous allocation of the other habitat patches.

Dispersal costs

For each of the 25 landscape tiles, a least-cost path analysis is used to determine the dispersal costs within all pairs of habitat patches, resulting in the minimal dispersal cost between two habitat patches being assigned to the link connecting them (Section 2.1.3 in Streib et al., 2020).

Habitat links

For each of the 25 landscape tiles, links are removed from the habitat network if the assigned dispersal costs exceed a given maximum. Thus, the maximal dispersal costs directly affect the number of links in a network. Differing from (Streib et al., 2020), we assume maximal dispersal costs of 900 for random, 650 for clustered, and 400 for contiguous habitat allocation to ensure that all network types have similar distributions of the number of links.

A.3. Standard habitat networks

Habitat patches

The number of habitat patches for each standard habitat network is chosen randomly between 54 and 111, with uniform probabilities so as to ensure that all network types have similar distributions of the number of nodes (Table 4.2 and Fig. 4.10).

Habitat links

Parameters to create the standard habitat networks are set so as to ensure that the distributions of the number of links for these networks are similar to those of the landscape-based networks (Table 4.3). Habitat patches in random networks are connected with probability $p = 0.04$ (networks are created using the NetworkX function `nx.erdos_renyi_graph`). Regular networks are set to have a degree of $d=4$ (networks are created using the NetworkX function `nx.random_regular_graph`). Habitat patches in small-world networks are set to have $k = 2$ neighbours to start with, and the probability of adding a new link for each link is $p = 0.6$ (networks are created using the NetworkX function `nx.newman_watts_strogatz_graph`). The scale-free networks are defined based on degree sequences: these sequences follow a power law, and the fraction $P(k)$ of habitat patches connected to k neighbours is proportional to $k^{-(2)}$ (networks are created using the NetworkX functions `nx.configuration_model`, `nx.utils.create_degree_sequence`, and `powerlaw_sequence`).

Dispersal costs

Each link in a standard habitat network is randomly assigned a dispersal cost so as to ensure that the distributions of dispersal costs for these networks are similar to the averaged distribution for landscape-based networks (Table 4.4). Therefore, the links are assigned dispersal costs randomly chosen between 46 and 400 for a first third of links, between 46 and 650 for a second third, and between 46 and 900 for the last third, in each case using a uniform probability density function.

Network type	Mean	Standard deviation	Minimum	Maximum
Random allocation	78	12.2	54	111
Clustered allocation	78	12.2	54	111
Contiguous allocation	78	12.2	54	111
Random	78	12.2	54	111
Regular	78	12.2	54	111
Small-world	78	12.2	54	111
Scale-free	78	12.2	54	111
Mean	78	12.2	54	111

TABLE 4.2: Distribution of the number of nodes for each network type.

Network type	Mean	Standard deviation	Minimum	Maximum
Random allocation	139.2	36.6	53	271
Clustered allocation	129.5	30.9	63	233
Contiguous allocation	139.0	27.9	82	261
Random	123.2	39.9	39	254
Regular	141.0	32.3	81	222
Small-world	116.8	18.7	76	176
Scale-free	225.1	36.7	153	324
Mean	144.8	31.9	78.1	248.7

TABLE 4.3: Distribution of the number of links for each network type.

Network type	Mean	Standard deviation	Minimum	Maximum
Random allocation	543.1	236.6	50	900
Clustered allocation	345.7	166.7	50	650
Contiguous allocation	228.9	98.5	50	400
Random	348.8	210.8	50	900
Regular	347.8	210.7	50	900
Small-world	348.0	209.5	50	900
Scale-free	346.8	209.3	50	900
Mean	358.4	191.7	50	792.8

TABLE 4.4: Distribution of the dispersal costs for each network type.

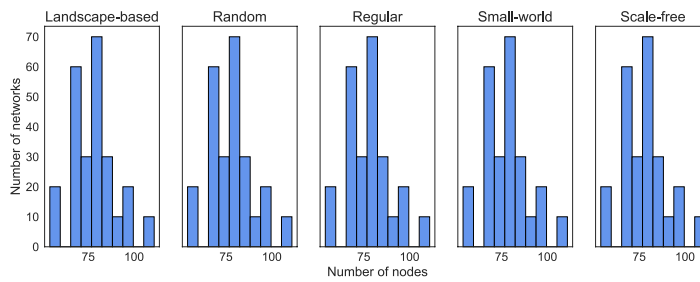


FIGURE 4.10: Full distribution of the number of nodes for the different network types. Only one distribution is shown for the three types of landscape-based habitat networks as the number of nodes in these networks does not depend on the types of habitat allocation, but only on the underlying stream landscapes: the distributions of the number of nodes are thus identical for all three types of landscape-based habitat networks. The standard networks are created using the same node distribution.

Chapter 5

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

The content of this chapter has already been published in an international reviewed journal and can be accessed via the following link:

Heer, H., Streib, L. Schäfer, R. B., & 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>

Abstract

The robustness of networks against node failure has been of interest in various disciplines 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 a higher robustness against habitat loss. Here, we develop tools to identify which links should be added to a network to increase the robustness. We introduce two different heuristics, a Greedy and Lazy Greedy algorithm, 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 independent of space 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 used heuristic. We found that applying our heuristics to add links to particularly sparse networks such as habitat networks has a much larger 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 link when links were added using the Greedy or Lazy Greedy algorithm.

Keywords

Clustering coefficient · Network robustness · Habitat networks · Habitat loss · Graph theory

5.1 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 (Sala et al., 2000; Fahrig, 2003; Foley et al., 2005; Titeux et al., 2016; Lechner et al., 2017). The loss and fragmentation of habitat lead to a decrease in habitat connectivity, impeding the movement of individuals between patches (Turner and Ruscher, 1988; Kindlmann and Burel, 2008). 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” (Levins, 1969; Hanski, 1998; Perry and Lee, 2019). 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 (Fahrig and Merriam, 1994; Hanski, 1999; Urban, 2015; Petsas et al., 2020).

Graph theory provides powerful tools to represent and analyse habitat connectivity in highly fragmented landscapes (Urban and Keitt, 2001; Dale and Fortin, 2010; Erős et al., 2012b; DeAngelis and Yurek, 2017). Here, metapopulations are represented by habitat networks where nodes represent habitat patches and links indicate how these are connected (Urban et al., 2009; Dale and Fortin, 2010; Galpern et al., 2011). 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 (Martensen et al., 2017; Kun et al., 2019; Prima et al., 2019). 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 (Mazaris et al., 2013; Dilts et al., 2016; Pietsch, 2018).

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 (Albert et al., 2001b; Solé and Montoya, 2001; Rosas-Casals et al., 2007; Berche et al., 2009; Cuadra et al., 2015). 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 (Ash and Newth, 2007; Almpanidou et al., 2014; Fox and Bellwood, 2014). The clustering coefficient of a network was proposed by Watts and Strogatz (Watts and Strogatz, 1998) 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 (Krumke and Noltemeier, 2009) 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. (2020) 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.

5.2 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) (Newman, 2010), 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. (Streib et al., 2020).

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. (2020) 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 5.1).

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 $\omega \leq 2.376$ (Green and Bader, 2013). 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 5.1 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 (Ash and Newth, 2007; Almpantidou et al., 2014; Fox and Bellwood, 2014; Prima et al., 2019), 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 5.2.1. *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*

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

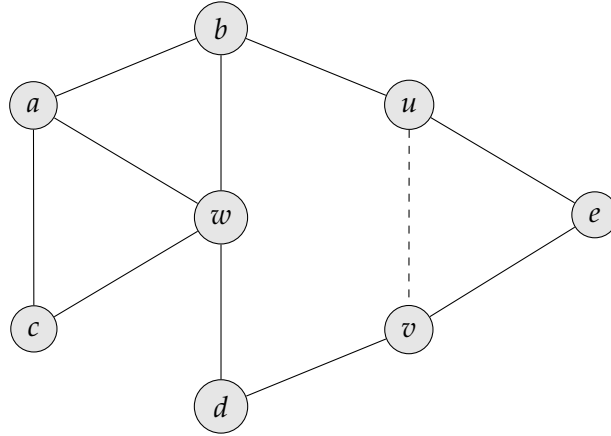


FIGURE 5.1: Example network to illustrate notation. $G = (V, E)$ with $n = 8$ nodes, 10 links, $V = \{a, b, c, d, e, u, v, w\}$, and $E = \{(a, b), (a, w), (a, c), (b, u), (b, w), (u, e), (w, c), (w, d), (e, v), (d, v)\}$. We choose $m = 1$ link from the set $\mathcal{E} = V \times V \setminus E$ of 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 $C_{G'} = \frac{1}{8} \cdot (\frac{1}{3} + \frac{1}{3} + \frac{1}{3} + \frac{1}{3} + \frac{2}{3} + 1 + 0 + 1) = \frac{1}{8} \cdot 4 = \frac{1}{2}$.

Example 5.2.2. Consider the network $G = (V, E)$ from Fig 5.1. 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 5.2.1 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 on the underlying land-use class (Keller et al., 2012).

Update clustering coefficient

We first aim to solve Problem 5.2.1 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 $\mathcal{O}(|\mathcal{E}|n^w)$, 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 :

$$C_{G'} = C_G + \frac{1}{n} \left(\Delta C(u) + \Delta C(v) + \sum_{w \in N(u, v)} \frac{2}{d_w(d_w - 1)} \right) \quad (5.1)$$

with

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

See Appendix for the proof of Eq 5.1.

It follows from Eq 5.1 and Fig 5.2, 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 5.1 is empty (and thus equals 0) and

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

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

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

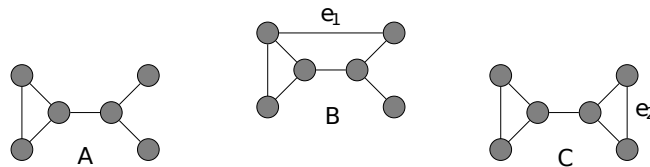


FIGURE 5.2: Varying effects of adding a link to a network on the clustering coefficient. (a) Original network with clustering coefficient $C_G = 0.38$. (b) Network after connecting two nodes with no common neighbours, $C_{G+e_1} = 0.27$. (c) Network after connecting two nodes with a common neighbour, $C_{G+e_2} = 0.7$.

Using Eq 5.1, 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 5.1. 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 5.1 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 5.2.1 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 5.2.1 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 5.2.1 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 5.3a and assume we can add two links. The Greedy algorithm will add the links shown in Fig. 5.3b, while the links shown in Fig. 5.3c 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 $\mathcal{O}(n^\omega + |\mathcal{E}|d_{\max} + |\mathcal{E}| \log(|\mathcal{E}|))$ (Ottmann and Widmayer, 2017; Hoare, 1962). Similar to the Greedy algorithm, the Lazy Greedy algorithm does not necessarily find the optimal solution to Problem 5.2.1. Fig 5.3 also serves as example of a non-optimal solution, as the Lazy Greedy algorithm will select the same links as the Greedy algorithm.

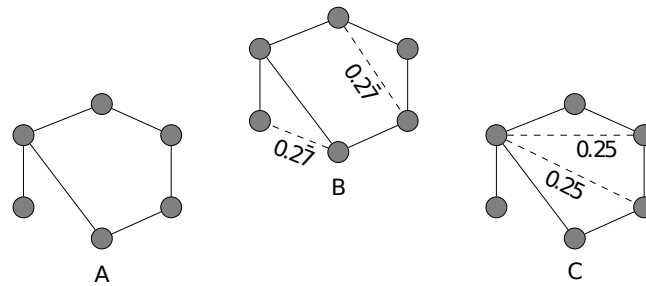


FIGURE 5.3: Example of non-optimal behaviour of the Greedy Algorithm. (a) Original network with clustering coefficient $C = 0$. (b) Network with two links selected using the Greedy algorithm and clustering coefficient $C = 0.5$. (c) Optimal solution with clustering coefficient $C = 0.605$. The value 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 5.1).

Algorithm 4 Lazy Greedy

```

1: procedure LAZYGREEDY( $G = (V, E), \mathcal{E}, m$ )
2:    $C_G \leftarrow$  Clustering( $G$ )
3:    $T \leftarrow$  Triangles( $G$ )
4:   results  $\leftarrow$  new Array
5:   for  $(u, v) \in \mathcal{E}$  do
6:      $C =$  UpdateClustering( $G, C_G, (u, v)$ )
7:     append  $(u, v, C)$  to results
8:   results  $\leftarrow$  sort results by  $C$ 
9:   return results[1], 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. (Streib et al., 2020) with random, clustered, and contiguous allocation of habitat patches / nodes as well as networks common in mathematics (random, regular, and small-world networks) (Newman, 2010). 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 (Wasserman and Faust, 1994; Watts and Strogatz, 1998; Barabasi and Oltvai, 2004; Holland and Hastings, 2008; Newman, 2010; Thompson et al., 2015). Fig 5.4 shows examples of the networks.

Landscape-based habitat networks

The landscape-based habitat networks were set up by Streib et al. (Streib et al., 2020) 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. (Streib et al., 2020), 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 cost of 50 to traverse a 25m × 25m area of open agricultural land, see Streib et al., 2020 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 5.4(a)-(d) for examples of the networks.

Standard networks

We created standard networks (random, regular, and small-world) using algorithms from the Python package NetworkX (Hagberg et al., 2008a). 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 Erdős and Rényi, 1960. Regular networks are networks, where every node has the same degree (Newman, 2010). Small-world networks are a mixture of regular and random networks and represent the small-world phenomenon from the social sciences (Boccaletti et al., 2006; Newman, 2010). 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 Newman and Watts, 1999 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. Table S3.1 shows the parameters and algorithms used to create the standard networks and Fig 5.4(e)-(f) show examples of the networks.

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

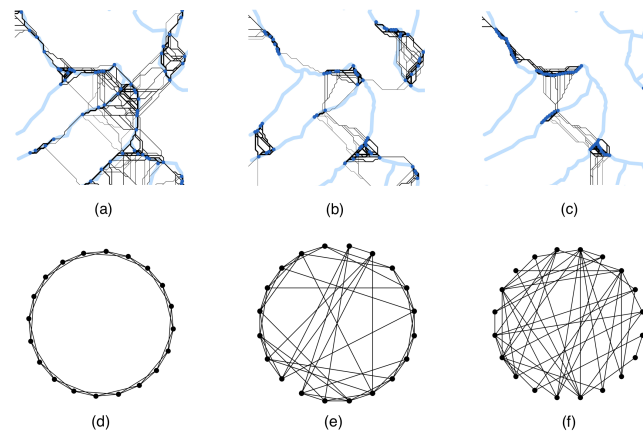


FIGURE 5.4: Networks examined. (a) - (c): Landscape-based networks. Dark-blue dots indicate nodes (habitat patches), black lines indicate links (dispersal pathways). The light-blue lines indicate the underlying stream network structure. (a) random allocation of habitat patches, (b) clustered allocation, (c) linear allocation. (d)-(f): Standard networks. (d) regular network, (e) small-world network, (f) random network

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. (2020) 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 Appendix and (Heer et al., 2020) 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 Streib et al., 2020, 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.

5.3 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 5.5).

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.

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

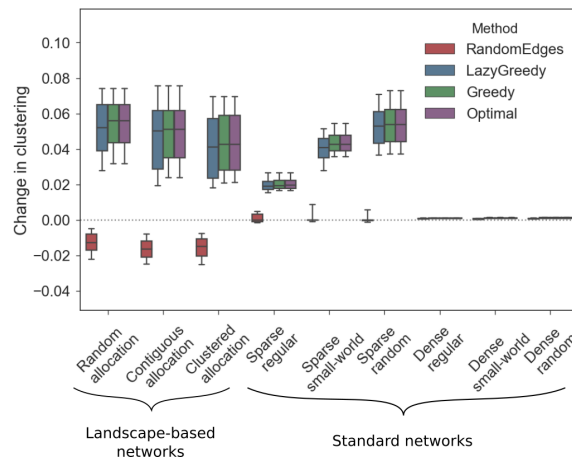


FIGURE 5.5: Greedy and Lazy Greedy algorithm applied to landscape-based and sparse networks lead to a higher increase in the clustering coefficient compared to randomly adding links. The horizontal axis shows the different network types, the vertical axis shows the change in clustering compared to the original network. The colour coding of the box-plots indicates the different algorithms.

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 5.1), 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 5.6).

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 5.7).

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.

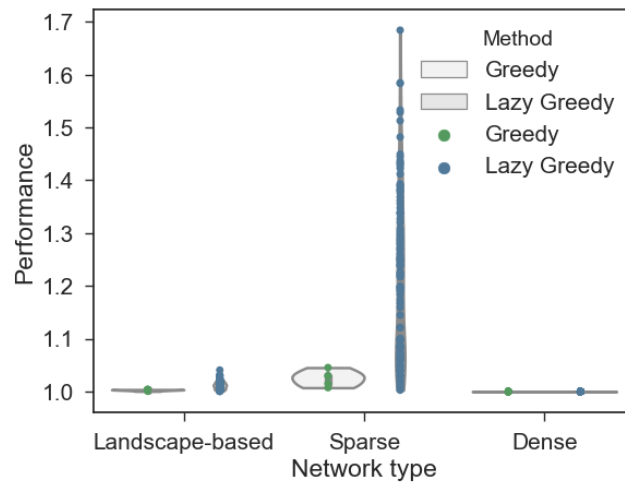


FIGURE 5.6: The Greedy algorithm returns an optimal solution in almost all cases. The vertical axis shows the quotient between optimal solution and solution of the heuristic. Only non optimal results are shown.

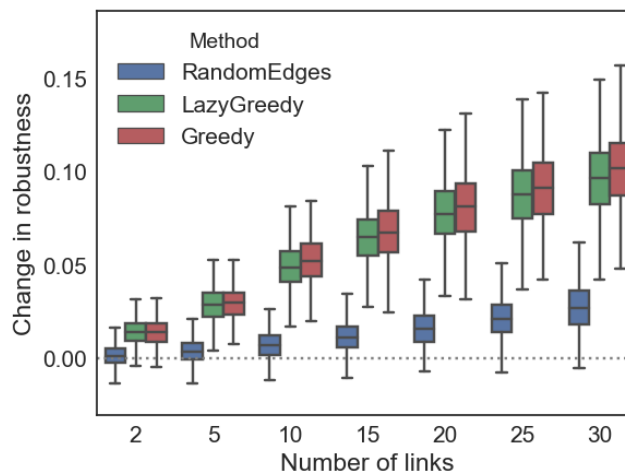


FIGURE 5.7: The robustness of networks increases with the number of additional links added using the Greedy or Lazy Greedy algorithm. The horizontal axis shows the number of links added to the network, the vertical axis the change in robustness. Colours indicate the algorithm used and each box shows the results over all landscape-based habitat networks.

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 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 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.

Appendix

A.1. Update clustering coefficient

We prove equation (5.1):

Lemma 5.3.1. *Let $G = (V, E)$ be a network as above, $(u, v) \in (V \times V) \setminus E$ and $G' = (V, E \cup \{(u, v)\})$ be the network resulting from G by inserting (u, v) . Then the following holds:*

1. $d'_u = d_u + 1$ and $d'_v = d_v + 1$
2. $T'(w) = T(w) + 1$ for all common neighbours $w \in N(u, v)$ of u and v
3. $T'(u) = T(u) + |N(u, v)|$ and $T'(v) = T(v) + |N(u, v)|$
4. $T'(w) = T(w) \forall w \in V \setminus N(u, v), w \neq u, v$
5. $d'_w = d_w$ for all $w \neq u, v$

Proof. Let G, e, G' as above.

1. Consider node u in G' . Then, v becomes a neighbour of u after (u, v) was inserted, i.e. v is a neighbour of u in G' , but not in G . All other neighbours do not change. Thus $d'_u = d_u + 1$.
2. Consider $w \in N(u, v)$. Then uvw is a triangle in G' , but not in G . It is the only triangle involving both w and (u, v) . All other triangles involving w do not involve (u, v) and are therefore also in G and thus $T'(w) = T(w) + 1$.
3. Consider node $u \in V$. Triangles involving u in G' that are not in G also have to involve v . The missing node in a triangle thus has to be connected to both u and v . Thus exactly the neighbours of both u and v are involved in triangles in G' that do not exist in G .
4. Let $u, v \neq w \in V \setminus N(u, v)$. Triangles in G' that are not in G have to involve (u, v) . Since w is not a neighbour of both u and v , no triangle uvw exists and thus $T'(w) = T(w)$.

□

We use Lemma 5.3.1 to calculate the clustering coefficient $C_{G'}$ of G' with help of C_G .

Lemma 5.3.2. *Let $G = (V, E)$ be a network as above with n nodes, $(u, v) \in \mathcal{E}$ and $G' = (V, E \cup \{(u, v)\})$. Let $k := |N(u, v)| \geq 1$ be the number of common neighbours of u and v . Then the difference in clustering is as follows:*

$$\Delta C = C_{G'} - C_G = \frac{1}{n} \left(\Delta C(u) + \Delta C(v) + \sum_{w \in N(u, v)} \frac{2}{d_w(d_w - 1)} \right) \quad (5.2)$$

with

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

Proof. Let $u, v \in V$ with $d_u, d_v > 1$ and set $k := |N(u, v)|$ as the number of common neighbours. Then, it holds

$$\begin{aligned} C_{G'} &= \frac{1}{n} \left(C_{G'}(u) + C_{G'}(v) + \sum_{w \in N(u, v)} C_{G'}(w) + \sum_{w \notin N(u, v), w \neq u, v} C_{G'}(w) \right) \\ &= \frac{1}{n} \left(\frac{2T'(u)}{d'_u(d'_u - 1)} + \frac{2T'(v)}{d'_v(d'_v - 1)} + \sum_{w \in N(u, v)} \frac{2T'(w)}{d'_w(d'_w - 1)} + \sum_{w \notin N(u, v), w \neq u, v} C_{G'}(w) \right) \\ &= \frac{1}{n} \left(\frac{2(T(u) + k)}{(d_u + 1)d_u} + \frac{2(T(v) + k)}{(d_v + 1)d_v} + \sum_{w \in N(u, v)} \frac{2(T(w) + 1)}{d_w(d_w - 1)} + \sum_{w \notin N(u, v), w \neq u, v} C_G(w) \right). \end{aligned}$$

The difference in clustering by inserting $e = (u, v)$ can be calculated as

$$\begin{aligned} n\Delta C &= C_{G'} - C_G \\ &= \frac{2(T(u) + k)}{(d_u + 1)d_u} - \frac{2T(u)}{d_u(d_u - 1)} + \frac{2(T(v) + k)}{(d_v + 1)d_v} - \frac{2T(v)}{d_v(d_v - 1)} \\ &\quad + \sum_{w \in N(u, v)} \left(\frac{2(T(w) + 1)}{d_w(d_w - 1)} - \frac{2T(w)}{d_w(d_w - 1)} \right) \\ &\quad + \sum_{w \notin N(u, v), w \neq u, v} \left(\frac{2T(w)}{d_w(d_w - 1)} - \frac{2T(w)}{d_w(d_w - 1)} \right) \\ &= \frac{2k(d_u - 1) - 4T(u)}{d_u(d_u^2 - 1)} + \frac{2k(d_v - 1) - 4T(v)}{d_v(d_v^2 - 1)} + \sum_{w \in N(u, v)} \frac{2}{d_w(d_w - 1)}. \end{aligned}$$

Now, consider $u, v \in V$ with degree 1 and a common neighbour w . Then the difference in clustering is:

$$\begin{aligned}
 n\Delta C &= C_{G'} - C_G \\
 &= \frac{2(T(u) + k)}{2} - C_G(u) + \frac{2(T(v) + k)}{2} - C_G(v) + \sum_{w \in N(u,v)} \frac{2}{d_w(d_w - 1)} \\
 &= 1 - 0 + 1 - 0 + \frac{2}{d_w(d_w - 1)} \\
 &= 2 + \frac{2}{d_w(d_w - 1)}.
 \end{aligned}$$

This equation holds, because $d_u = d_v = 1$ and thus $T(u)$, $T(v)$, $C(u)$, and $C(v)$ all equal zero and $k = 1$, as they have a common neighbour. \square

A.2. Robustness simulation

We simulated habitat loss and subsequent metapopulation dynamics as proposed by Heer et al. (under review) on the landscape-based habitat networks to evaluate the increase of metapopulation robustness on those networks. Here, we briefly summarize how the simulation was modelled to make it easier for readers to follow our findings.

Simulation overview

To evaluate the robustness of a habitat network against habitat loss, we first simulated the habitat loss by randomly removing habitat patches from the network. Habitat patches on the remaining network were assumed to be fully colonised. Then, metapopulation dynamics consisting of local extinctions and subsequent recolonisation from neighbouring patches were simulated until a stationary distribution was reached. This process of simulated habitat loss and subsequent metapopulation dynamics was then 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: the higher the fraction of colonised habitat patches across fractions of lost habitat patches, the higher the AUC, and thus the estimated metapopulation robustness. For each network, simulations were replicated ten times to average over the sources of randomness affecting habitat loss and metapopulation dynamics.

Habitat loss

We assumed a random habitat loss scenario, which removed each habitat patch with equal probability p .

Metapopulation dynamics

For a given level of habitat loss, metapopulation dynamics were simulated on the remaining habitat network, by considering local extinctions in habitat patches and the recolonization of habitat patches. We used the size of cliques to measure, how well a patch is connected within its neighbourhood, as the survival of a population in

a habitat patch depends on its potential to exchange individuals with neighbouring patches. Denoting by $c(v)$ the size of the largest clique that contains the node v we assumed that the population in v goes extinct with probability

$$p_{\text{ext}}(v) = a^{1-c(v)},$$

where $a > 1$ is a species-specific parameter governing the local-extinction risk of a species. We can think of these risks decreasing with increasing $c(v)$ more slowly for habitat specialists (small values of a) and more rapidly for habitat generalists (large values of a). We investigated species with three different levels of local-extinction risks - low ($a = 2$), medium ($a = 5$), and high ($a = 9$).

Empty habitat patches can be recolonised from connected colonised patches. Recolonisation was modelled with the help of a Gaussian dispersal kernel and we assumed that an empty habitat patch v becomes recolonised from a colonised patch w with probability

$$p_{\text{col}}(v, w) = \frac{m_{vw}}{\sum_{u \in V} m_{uw}},$$

where $m_{vw} = \exp(-\frac{1}{2}d_{vw}^2/\sigma^2)$ is the dispersal kernel, V the set of all network nodes, d_{vw} the distance between habitat patches v and w in terms of dispersal costs and $\sigma > 0$ a species-specific dispersal parameter governing the dispersal range of a species. We can think of these dispersal ranges as being low for poor dispersers (small values of σ) and high for good dispersers (large values for σ). Similar to a , we investigate values of $\sigma \in [2, 5, 9]$ to account for the different dispersal capacities of different species.

These local extinctions in and recolonizations of habitat patches were simulated alternately until a stationary frequency of colonized patches was reached.

A.3. Networks

Network	NetworkX algorithm	Parameter sparse	Parameter dense
Regular	<code>nx.random_regular_graph</code>	$d = 4$	$d = 58$
Random	<code>nx.erdos_renyi_graph</code>	$p = 0.04$	$p = 0.75$
Small-world	<code>nx.newman_watts_strogatz_graph</code>	$k = 2, p = 0.6$	$k = 39, p = 0.5$

TABLE 5.1: Parameters to create standard networks. d is the degree of each node, p denotes the percentage of links present in the network and k is the degree of each node in the small-world network before rewiring.

Chapter 6

Discussion

In this section, the results of the publications are discussed, followed by the limitations and outlook.

In chapter 3, I presented an optimisation model we developed built on a dispersal simulation of a generic hemimetabolous species. It calculates the minimum time needed for the species in focus to disperse to and colonise predefined empty habitat patches. We found that the outcome of the optimisation model highly depends on the distances of those initial source habitats that are closest to the destination habitat. In other words, habitat patches that are particularly cost-efficiently connected to the initial source habitats, get colonised the fastest under the optimisation model. This is in line with common literature as close habitat patches are preferably colonised (Van Nouhuys and Hanski, 2002; Kajzer et al., 2012). Furthermore, the optimisation model outcome can be viewed as an estimate of the expected simulation model outcome, as the colonisation time calculated by the optimisation model is on average seven times faster compared to the colonisation time obtained by the simulation model. However, the optimisation model should not be viewed as a mere substitute for the simulation model, as it answers questions that are different, yet closely related, to the simulation model.

In chapter 4, I presented the results of our work analysing the robustness of habitat networks. We found that species with short dispersal ranges and high local-extinction risks were particularly vulnerable to the loss of habitat across all types of networks and habitat loss. We identified the clustering coefficient to be a reliable indicator for robustness across all kinds of networks, habitat loss scenarios and species. As small-world networks are characterized by particular high clustering coefficients with simultaneously short average shortest path lengths, this is in agreement with other studies suggesting the small-world characteristics of networks as sign of robustness (Almpanidou et al., 2014; Fox and Bellwood, 2014; Prima et al., 2019). Furthermore, species with weak dispersal capacities and high local-extinction risks generally showed a higher strength of the relationship between network metrics and metapopulation robustness, indicating that the structure of habitat networks has a higher impact on the metapopulation robustness of particularly vulnerable species. Thus, the predictability of metapopulation robustness by network metrics is particularly high for species that need most protection.

The landscape-based habitat networks responded differently to habitat loss than standard networks and generally displayed a slightly higher metapopulation robustness. This suggests that the standard networks, which lack a spatial reference to an underlying landscape, may not be suitable representatives of real-world habitat networks, despite their widespread use in a broad range of studies and disciplines.

In chapter 5, I presented the results of our work on maximising the robustness of habitat networks by adding new links using the clustering coefficient as a proxy. We introduced two heuristics, the Greedy and a deducted Lazy Greedy algorithm to

select links that should be added to a network to create a network with the maximum increase in the clustering coefficient and in turn the robustness. Both algorithms return networks with a much higher clustering coefficient compared to randomly adding links to the network. Furthermore, the number of added links chosen by either algorithm has a high correlation to the resultant change in metapopulation robustness. This is — to the best of our knowledge — the first study that investigates how the clustering coefficient of a network changes in response to additional links in the network.

6.1 Limitations & Outlook

The concept of a time expanded network to model a network over time as presented in chapter 3 is a common concept in graph theory (Skutella, 2009; Köhler et al., 2009; Kotnyek, 2003). With the help of time expanded networks, dynamic procedures on a network can be modeled as static problems, which often reduces the complexity of the problem drastically. As the dispersal of species is similar to a network flow, it seemed natural to exploit the structure of time expanded networks often used for network flows for our model. This structure can also be exploited to integrate changes to the network, such as changes in habitat quality or the (temporarily) appearing and disappearing of further habitat patches or links between them. One has to be very careful with the application of time expanded networks, however, as time expanded networks increase the size of the network considerably. A careful consideration between the gains achieved due to the reduced complexity of the problem at the cost of the increase of the input size is thus inevitable.

When constructing the optimisation model, the main challenge was to translate the simulated processes such as the dispersal and the population growth to linear constraints. While the optimisation model requires the constraints to be linear, reducing the simulation to linear processes may oversimplify the modelled system. On the other hand, integrating a more realistic population growth process, for example, will require additional decision variables as well as more complex and more numerous constraints which, in turn, lead to an increase in the model complexity. The most challenging part of adapting the optimisation model to other dispersal simulations will thus remain to be the translation of complex processes into linear equations and finding the balance between modelling complex processes and keeping the optimisation model simple. When considering species with particularly complicated population processes an oversimplification may lead to an underestimated colonisation time, rendering the optimisation model unsuitable for these species.

Future use of the optimisation model thus has to carefully find a balance between the usability and adaptability of the model due to the time expanded structure and the loss of computational efficiency.

Further models and a more thorough and detailed analysis of the optimisation model and its outcome can help us to understand the phenomenon of range shifts more thoroughly. With the lower bounds given by the optimisation model, we can identify habitat patches that are important stepping stones for species dispersal, for example. Furthermore, the strength of the connection between different habitat patches could be evaluated with the help of lower bounds obtained from the optimisation model.

One of the main findings of chapter 4 is that standard networks respond qualitatively differently to habitat loss compared to landscape-based networks. These differences are very likely due to the lack of a spatial reference to an underlying

landscape for standard networks and we suggest that this fundamental limitation renders standard networks unsuitable for studying the impact of habitat loss on metapopulations. Future studies should thus investigate a broader range of spatial networks and consider habitat networks derived from real-world landscapes to evaluate the wider generalizability of our results. As standard networks are commonly used to represent habitat networks due to their ubiquity in graph theory (Kun et al., 2009; Shen et al., 2019), we strongly suggest to consider the spatial aspect of habitat networks for future studies.

The heuristics presented in chapter 5 deliver reliable ways to select links that should be added to a network to increase its robustness. However, the performance of both heuristics remains unknown and no estimate could be found to qualify the quality of the approximation of the optimal solution for multiple edges. Furthermore, the algorithms are not yet applicable to real-world problems and further research is needed to make our algorithms more applicable for conservation purposes. For example, the goal of a real-world habitat conservation scenario is not the addition of a single link. Instead, further habitat patches may be created by increasing the habitat quality of landscapes. Additionally, corridors might be created to facilitate movement between patches. However, these corridors will most likely correspond to multiple new links in the same area. More mathematical models are needed to identify where new habitat patches should be created as well as to study the effect of adding bundles of links to a network.

The goal of this thesis was to gain a better understanding of the structure of habitat networks and predict their behaviour in future scenarios. We introduced an optimisation model to identify bounds on possible range shifts over time and developed models to evaluate and increase the robustness of metapopulations against habitat loss considering a variety of habitat loss scenarios as well as different species characteristics and network structures. Overall this thesis contributed in narrowing the gap between mathematics and ecology by introducing mathematical tools to ecology and developing frameworks to study the behaviour of habitat networks. While all models in this thesis were developed with aquatic invertebrates in mind, they can easily be adapted to other metapopulations.

Bibliography

- Alahuhta, J. et al. (2019). "Understanding environmental change through the lens of trait-based, functional, and phylogenetic biodiversity in freshwater ecosystems". In: *Environmental Reviews* 27.2, pp. 263–273.
- Alba, R. D. (1973). "A graph-theoretic definition of a sociometric clique". In: *Journal of Mathematical Sociology* 3.1, pp. 113–126.
- Albert, R., H. Jeong, and A.-L. Barabási (2001a). "Error and attack tolerance of complex networks". In: *Nature* 409.6819, pp. 542–542.
- Albert, R., H. Jeong, and A.-L. Barabasi (2001b). "Error and attack tolerance of complex networks". In: *Nature* 409.6819, pp. 542–542.
- Almpanidou, V. et al. (2014). "Providing insights on habitat connectivity for male brown bears: A combination of habitat suitability and landscape graph-based models". In: *Ecological Modelling* 286, pp. 37–44.
- Amaran, S. et al. (2016). "Simulation optimization: a review of algorithms and applications". In: *Annals of Operations Research* 240.1, pp. 351–380.
- Amarasekare, P. (2004). "The role of density-dependent dispersal in source–sink dynamics". In: *Journal of Theoretical Biology* 226.2, pp. 159–168.
- Ash, J. and D. Newth (2007). "Optimizing complex networks for resilience against cascading failure". In: *Physica A: Statistical Mechanics and its Applications* 380, pp. 673–683.
- Barabási, A.-L. and R. Albert (1999). "Emergence of scaling in random networks". In: *Science* 286.5439, pp. 509–512.
- Barabási, A.-L. and E. Bonabeau (2003). "Scale-free networks". In: *Scientific American* 288.5, pp. 60–69.
- Barabasi, A.-L. and Z. N. Oltvai (2004). "Network biology: Understanding the cell's functional organization". In: *Nature reviews genetics* 5.2, pp. 101–113.
- Barthélemy, M. (2014). *Spatial Networks*. Springer.
- Bastos, L. S. and A. O'Hagan (2009). "Diagnostics for Gaussian process emulators". In: *Technometrics* 51.4, pp. 425–438.
- Bates, D. et al. (2014). "Fitting linear mixed-effects models using lme4". In: *arXiv Preprint*, 1406.5823.
- Baumann, N. and M. Skutella (2009). "Earliest arrival flows with multiple sources". In: *Mathematics of Operations Research* 34.2, pp. 499–512.
- Bellard, C. et al. (2012). "Impacts of climate change on the future of biodiversity". In: *Ecology letters* 15.4, pp. 365–377.
- Berche, B. et al. (2009). "Resilience of public transport networks against attacks". In: *The European Physical Journal B* 71.1, pp. 125–137.
- Berg, M. P. et al. (2010). "Adapt or disperse: understanding species persistence in a changing world". In: *Global Change Biology* 16.2, pp. 587–598.
- Boccaletti, S. et al. (2006). "Complex networks: Structure and dynamics". In: *Physics Reports* 424.4-5, pp. 175–308.
- Bondy, J. A., U. S. R. Murty, et al. (1976). *Graph theory with applications*. Vol. 290. Citeseer.

- Bowler, D. E. and T. G. Benton (2005). "Causes and consequences of animal dispersal strategies: relating individual behaviour to spatial dynamics". In: *Biological Reviews* 80.2, pp. 205–225.
- Bruggeman, D. J., T. Wiegand, and N. Fernandez (2010). "The relative effects of habitat loss and fragmentation on population genetic variation in the red-cockaded woodpecker (*Picoides borealis*)". In: *Molecular Ecology* 19.17, pp. 3679–3691.
- Burgos, E. et al. (2007). "Why nestedness in mutualistic networks?" In: *Journal of Theoretical Biology* 249.2, pp. 307–313.
- Butchart, S. H. et al. (2010). "Global biodiversity: indicators of recent declines". In: *Science* 328.5982, pp. 1164–1168.
- Calabrese, J. M. and W. F. Fagan (2004). "A comparison-shopper's guide to connectivity metrics". In: *Frontiers in Ecology and the Environment* 2.10, pp. 529–536.
- Callaway, D. S. et al. (2000). "Network robustness and fragility: Percolation on random graphs". In: *Physical Review Letters* 85.25, pp. 5468–5471.
- Chapman, D. S., C. Dytham, and G. S. Oxford (2007). "Modelling population redistribution in a leaf beetle: an evaluation of alternative dispersal functions". In: *Journal of Animal Ecology* 76.1, pp. 36–44.
- Clark, J. S. et al. (2001). "Ecological forecasts: an emerging imperative". In: *science* 293.5530, pp. 657–660.
- Cohen, R. and S. Havlin (2010). *Complex networks: structure, robustness and function*. Cambridge university press.
- Cohen, R. et al. (2000). "Resilience of the internet to random breakdowns". In: *Physical review letters* 85.21, p. 4626.
- Corbet, P. S. (1963). *Biology of dragonflies*. Quadrangle Books.
- Córdoba-Aguilar, A. (2008). *Dragonflies and damselflies: model organisms for ecological and evolutionary research*. OUP Oxford.
- Cowling, R. et al. (1999). "From representation to persistence: Requirements for a sustainable system of conservation areas in the species-rich mediterranean-climate desert of southern Africa". In: *Diversity and Distributions* 5.1-2, pp. 51–71.
- Cuadra, L. et al. (2015). "A critical review of robustness in power grids using complex networks concepts". In: *Energies* 8.9, pp. 9211–9265.
- Dale, M. and M.-J. Fortin (2010). "From graphs to spatial graphs". In: *Annual Review of Ecology, Evolution, and Systematics* 41.1, pp. 21–38.
- Dantzig, G. (2016). *Linear programming and extensions*. Princeton university press.
- DeAngelis, D. L. and S. Yurek (2017). "Spatially explicit modeling in ecology: A review". In: *Ecosystems* 20.2, pp. 284–300.
- Didham, R. K. et al. (2012). "Horizontal and vertical structuring in the dispersal of adult aquatic insects in a fragmented landscape". In: *Fundamental and Applied Limnology/Archiv für Hydrobiologie* 180.1, pp. 27–40.
- Dilts, T. E. et al. (2016). "Multiscale connectivity and graph theory highlight critical areas for conservation under climate change". In: *Ecological Applications* 26.4, pp. 1223–1237.
- Doak, D. F. and W. F. Morris (2010). "Demographic compensation and tipping points in climate-induced range shifts". In: *Nature* 467.7318, p. 959.
- Erdős, P. and A. Rényi (1960). "On the evolution of random graphs". In: *Publications of the Mathematical Institute of the Hungarian Academy of Sciences* 5.1, pp. 17–60.
- Erős, T. and W. H. Lowe (2019). "The landscape ecology of rivers: from patch-based to spatial network analyses". In: *Current Landscape Ecology Reports* 4.4, pp. 103–112.
- Erős, T. et al. (2012a). "Characterizing connectivity relationships in freshwaters using patch-based graphs". In: *Landscape ecology* 27.2, pp. 303–317.

- Erős, T. et al. (2012b). "Characterizing connectivity relationships in freshwaters using patch-based graphs". In: *Landscape Ecology* 27.2, pp. 303–317.
- Estrada, E. and Ö. Bodin (2008). "Using network centrality measures to manage landscape connectivity". In: *Ecological Applications* 18.7, pp. 1810–1825.
- Evans, D. M., M. J. Pocock, and J. Memmott (2013). "The robustness of a network of ecological networks to habitat loss". In: *Ecology Letters* 16.7, pp. 844–852.
- Evans, M. R., K. J. Norris, and T. G. Benton (2012). *Predictive ecology: systems approaches*.
- Fagan, W. F. and E. Holmes (2006). "Quantifying the extinction vortex". In: *Ecology Letters* 9.1, pp. 51–60.
- Fahrig, L. (2003). "Effects of habitat fragmentation on biodiversity". In: *Annual Review of Ecology, Evolution, and Systematics* 34.1, pp. 487–515.
- Fahrig, L. and G. Merriam (1985). "Habitat Patch Connectivity and Population Survival: Ecological Archives E066-008". In: *Ecology* 66.6, pp. 1762–1768.
- (1994). "Conservation of fragmented populations". In: *Conservation Biology* 8.1, pp. 50–59.
- Faraway, J. J. (2016). *Extending the Linear Model with R: Generalized Linear, Mixed Effects and Nonparametric Regression Models*. Chapman and Hall/CRC.
- Foley, J. A. et al. (2005). "Global consequences of land use". In: *Science* 309.5734, pp. 570–574.
- Ford Jr, L. R. and D. R. Fulkerson (1958). "Constructing maximal dynamic flows from static flows". In: *Operations research* 6.3, pp. 419–433.
- Ford Jr, L. R. and D. R. Fulkerson (2015). *Flows in networks*. Princeton university press.
- Fortuna, M. A., C. Gómez-Rodríguez, and J. Bascompte (2006). "Spatial network structure and amphibian persistence in stochastic environments". In: *Proceedings of the Royal Society B: Biological Sciences* 273.1592, pp. 1429–1434.
- Fox, R. J. and D. R. Bellwood (2014). "Herbivores in a small world: Network theory highlights vulnerability in the function of herbivory on coral reefs". In: *Functional Ecology* 28.3, pp. 642–651.
- Freeman, L. C. (1977). "A set of measures of centrality based on betweenness". In: *Sociometry* 40.1, pp. 35–41.
- Galpern, P., M. Manseau, and A. Fall (2011). "Patch-based graphs of landscape connectivity: A guide to construction, analysis and application for conservation". In: *Biological conservation* 144.1, pp. 44–55.
- Gastner, M. T. and M. E. J. Newman (2006). "The spatial structure of networks". In: *The European Physical Journal B – Condensed Matter and Complex Systems* 49.2, pp. 247–252.
- Gephart, J. A. et al. (2016). "Vulnerability to shocks in the global seafood trade network". In: *Environmental Research Letters* 11.3, no. 035008.
- Grant, E. H. C. et al. (2010). "Use of multiple dispersal pathways facilitates amphibian persistence in stream networks". In: *Proceedings of the National Academy of Sciences* 107.15, pp. 6936–6940.
- Grech, A. et al. (2018). "Predicting the cumulative effect of multiple disturbances on seagrass connectivity". In: *Global Change Biology* 24.7, pp. 3093–3104.
- Green, O. and D. A. Bader (2013). "Faster clustering coefficient using vertex covers". In: *2013 International Conference on Social Computing*. IEEE, pp. 321–330.
- Grönroos, M. et al. (2013). "Metacommunity structuring in stream networks: roles of dispersal mode, distance type, and regional environmental context". In: *Ecology and Evolution* 3.13, pp. 4473–4487.
- Gurobi Optimization, Inc. (2016). *Gurobi Optimizer Reference Manual*. URL: <http://www.gurobi.com>.

- Hagberg, A., P. Swart, and D. Schult (2008a). "Exploring network structure, dynamics, and function using NetworkX". In: *Technical Report LA-UR-08-5495*, Los Alamos National Laboratory (LANL), Los Alamos, NM, USA.
- Hagberg, A. A., D. A. Schult, and P. J. Swart (2008b). "Exploring Network Structure, Dynamics, and Function using NetworkX". In: *Proceedings of the 7th Python in Science Conference*. Ed. by G. Varoquaux, T. Vaught, and J. Millman. Pasadena, CA USA, pp. 11–15.
- Hamacher, H. W. and K. Klamroth (2006). *Lineare Optimierung und Netzwerkoptimierung*. Springer.
- Hanski, I. (1999). *Metapopulation Ecology*. Oxford University Press.
- Hanski, I. (1998). "Metapopulation dynamics". In: *Nature* 396.6706, pp. 41–49.
- Heer, H. et al. (2020). "Indicators for Assessing the Robustness of Metapopulations against Habitat Loss". In:
- Heino, J. et al. (2017). "Integrating dispersal proxies in ecological and environmental research in the freshwater realm". In: *Environmental Reviews* 25.3, pp. 334–349.
- Hepenstrick, D., B. Koch, and C Monnerat (2014). *Merkblätter Arten – Libellen – Coenagrion mercuriale*.
- Ho, K. et al. (2014). "Dynamic modeling and optimization for space logistics using time-expanded networks". In: *Acta Astronautica* 105.2, pp. 428–443.
- Hoare, C. A. (1962). "Quicksort". In: *The Computer Journal* 5.1, pp. 10–16.
- Hodgson, J. A. et al. (2011). "Habitat area, quality and connectivity: striking the balance for efficient conservation". In: *Journal of Applied Ecology* 48.1, pp. 148–152.
- Holland, M. D. and A. Hastings (2008). "Strong effect of dispersal network structure on ecological dynamics". In: *Nature* 456.7223, pp. 792–794.
- IPCC (2019). *Climate Change and Land: an IPCC special report on climate change, desertification, land degradation, sustainable land management, food security, and greenhouse gas fluxes in terrestrial ecosystems*.
- Jarvis, J. J. and H. D. Ratliff (1982). "Note — some equivalent objectives for dynamic network flow problems". In: *Management Science* 28.1, pp. 106–109.
- Kajzer, J. et al. (2012). "Patch occupancy and abundance of local populations in landscapes differing in degree of habitat fragmentation: a case study of the colonial black-headed gull, *Chroicocephalus ridibundus*". In: *Journal of biogeography* 39.2, pp. 371–381.
- Keller, D., M. J. Van Strien, and R. Holderegger (2012). "Do landscape barriers affect functional connectivity of populations of an endangered damselfly?" In: *Freshwater Biology* 57.7, pp. 1373–1384.
- Kindlmann, P. and F. Burel (2008). "Connectivity measures: A review". In: *Landscape Ecology* 23.8, pp. 879–890.
- Köhler, E., R. H. Möhring, and M. Skutella (2009). "Traffic Networks and Flows over Time." In: *Algorithmics of Large and Complex Networks*. Springer, pp. 166–196.
- Kotnyek, B. (2003). "An annotated overview of dynamic network flows". PhD thesis. INRIA.
- Krumke, S. O. and H. Noltemeier (2009). *Graphentheoretische Konzepte und Algorithmen*. Springer-Verlag.
- Kuemmerlen, M. et al. (2015). "An attack on two fronts: predicting how changes in land use and climate affect the distribution of stream macroinvertebrates". In: *Freshwater biology* 60.7, pp. 1443–1458.
- Kun, A., B. Oborny, and U. Dieckmann (2009). "Intermediate landscape disturbance maximizes metapopulation density". In: *Landscape ecology* 24.10, pp. 1341–1350.
- Kun, Á., B. Oborny, and U. Dieckmann (2019). "Five main phases of landscape degradation revealed by a dynamic mesoscale model analysing the splitting,

- shrinking, and disappearing of habitat patches". In: *Scientific Reports* 9.1, pp. 1–11.
- Lande, R. (1988). "Genetics and demography in biological conservation". In: *Science* 241.4872, pp. 1455–1460.
- Landi, P. et al. (2018). "Complexity and stability of ecological networks: A review of the theory". In: *Population Ecology* 60.4, pp. 319–345.
- Landis, R. S. (2005). "Standardized regression coefficients". In: In: Everitt, B. S. and Howell, D. C. eds. *Encyclopedia of Statistics in Behavioral Science*. Wiley.
- Lechner, A. M. et al. (2017). "Characterising landscape connectivity for conservation planning using a dispersal guild approach". In: *Landscape Ecology* 32.1, pp. 99–113.
- Levins, R. (1969). "Some demographic and genetic consequences of environmental heterogeneity for biological control". In: *American Entomologist* 15.3, pp. 237–240.
- Lowe, W. H. and M. A. McPeck (2014). "Is dispersal neutral?" In: *Trends in ecology & evolution* 29.8, pp. 444–450.
- Lüdecke, D (2018). "*sjstats: Statistical functions for regression models*". In: R Package, Version 0.14.3.
- Machac, D., P. Reichert, and C. Albert (2016a). "Emulation of dynamic simulators with application to hydrology". In: *Journal of Computational Physics* 313, pp. 352–366.
- Machac, D. et al. (2016b). "Fast mechanism-based emulator of a slow urban hydrodynamic drainage simulator". In: *Environmental Modelling & Software* 78, pp. 54–67.
- Martensen, A. C., S. Saura, and M.-J. Fortin (2017). "Spatio-temporal connectivity: Assessing the amount of reachable habitat in dynamic landscapes". In: *Methods in Ecology and Evolution* 8.10, pp. 1253–1264.
- Mazaris, A. D. et al. (2013). "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". In: *PLoS One* 8.3, e59640.
- McRae, B. H. et al. (2008). "Using circuit theory to model connectivity in ecology, evolution, and conservation". In: *Ecology* 89.10, pp. 2712–2724.
- Minor, E. S. and D. L. Urban (2007). "Graph theory as a proxy for spatially explicit population models in conservation planning". In: *Ecological applications* 17.6, pp. 1771–1782.
- Montoya, J. M. and R. V. Solé (2002). "Small world patterns in food webs". In: *Journal of theoretical biology* 214.3, pp. 405–412.
- Moritz, C. et al. (2008). "Impact of a century of climate change on small-mammal communities in Yosemite National Park, USA". In: *Science* 322.5899, pp. 261–264.
- Nathan, R. et al. (2012). "Dispersal kernels". In: In: Clobert, J., Baguette, M., Benton, T. and Bullock, J. M. eds. *Dispersal Ecology and Evolution*. Oxford University Press, pp. 187–210.
- Newman, M. E. J. (2003). "Mixing patterns in networks". In: *Physical Review E* 67.026126, pp. 1–13.
- (2010). *Networks: An Introduction*. Oxford University Press.
- Newman, M. E. J. and D. J. Watts (1999). "Renormalization group analysis of the small-world network model". In: *Physics Letters A* 263.4-6, pp. 341–346.
- Newman, M. E. J., D. J. Watts, and S. H. Strogatz (2002). "Random graph models of social networks". In: *Proceedings of the national academy of sciences* 99.suppl 1, pp. 2566–2572.
- Ortiz-Rodríguez, D. O. et al. (2019). "Predicting species occurrences with habitat network models". In: *Ecology and Evolution* 9.18, pp. 10457–10471.

- Ottmann, T. and P. Widmayer (2017). *Algorithmen und Datenstrukturen*. Springer.
- Parmesan, C. (2006). "Ecological and evolutionary responses to recent climate change". In: *Annu. Rev. Ecol. Evol. Syst.* 37, pp. 637–669.
- Parmesan, C. and G. Yohe (2003). "A globally coherent fingerprint of climate change impacts across natural systems". In: *Nature* 421.6918, p. 37.
- Perry, G. L. and F. Lee (2019). "How does temporal variation in habitat connectivity influence metapopulation dynamics?" In: *Oikos* 128.9, pp. 1277–1286.
- Petsas, P., A. I. Tsavdaridou, and A. D. Mazaris (2020). "A multispecies approach for assessing landscape connectivity in data-poor regions". In: *Landscape Ecology*, pp. 1–16.
- Peyrard, N., U. Dieckmann, and A. Franc (2008). "Long-range correlations improve understanding of the influence of network structure on contact dynamics". In: *Theoretical Population Biology* 73.3, pp. 383–394.
- Pietsch, M. (2018). "Contribution of connectivity metrics to the assessment of biodiversity—Some methodological considerations to improve landscape planning". In: *Ecological Indicators* 94, pp. 116–127.
- Prima, M.-C. et al. (2019). "A landscape experiment of spatial network robustness and space-use reorganisation following habitat fragmentation". In: *Functional Ecology* 33.9, pp. 1663–1673.
- R Core Team (2018). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing.
- Rayfield, B., M.-J. Fortin, and A. Fall (2011). "Connectivity for conservation: A framework to classify network measures". In: *Ecology* 92.4, pp. 847–858.
- Reid, W. V. et al. (2005). *Ecosystems and human well-being-Synthesis: A report of the Millennium Ecosystem Assessment*. Island Press.
- Rhineland-Palatinate Ministry of the Interior and for Sports (2020a). "Gewässernetz (gesamt), Ressourcenidentifikator 54712". In: Wasserwirtschaftsverwaltung Rheinland-Pfalz, *GeoPortal.rlp*.
- (2020b). "Luftbild RP Basisdienst, Ressourcenidentifikator 30692". In: Wasserwirtschaftsverwaltung Rheinland-Pfalz, *GeoPortal.rlp*.
- Romanuk, T. et al. (2017). "Robustness trade-offs in model food webs: Invasion probability decreases while invasion consequences increase with connectance". In: *Advances in Ecological Research* 56, pp. 263–291.
- Rosas-Casals, M., S. Valverde, and R. V. Solé (2007). "Topological vulnerability of the European power grid under errors and attacks". In: *International Journal of Bifurcation and Chaos* 17.07, pp. 2465–2475.
- Rouquette, J. R. and D. J. Thompson (2007). "Patterns of movement and dispersal in an endangered damselfly and the consequences for its management". In: *Journal of Applied Ecology* 44.3, pp. 692–701.
- Sala, O. E. et al. (2000). "Global biodiversity scenarios for the year 2100". In: *Science* 287.5459, pp. 1770–1774.
- Saunders, M. I. et al. (2016). "Human impacts on connectivity in marine and freshwater ecosystems assessed using graph theory: A review". In: *Marine and Freshwater Research* 67.3, pp. 277–290.
- Saura, S. and L. Pascual-Hortal (2007). "A new habitat availability index to integrate connectivity in landscape conservation planning: comparison with existing indices and application to a case study". In: *Landscape and Urban Planning* 83.2-3, pp. 91–103.
- Sawyer, S. C., C. W. Epps, and J. S. Brashares (2011). "Placing linkages among fragmented habitats: do least-cost models reflect how animals use landscapes?" In: *Journal of Applied Ecology* 48.3, pp. 668–678.

- Scheffers, B. R. et al. (2016). "The broad footprint of climate change from genes to biomes to people". In: *Science* 354.6313, aaf7671.
- Schrijver, A. (1998). *Theory of linear and integer programming*. John Wiley & Sons.
- Searcy, C. A., E. Gabbai-Saldate, and H. B. Shaffer (2013). "Microhabitat use and migration distance of an endangered grassland amphibian". In: *Biological conservation* 158, pp. 80–87.
- Sexton, J. P. et al. (2009). "Evolution and ecology of species range limits". In: *Annual Review of Ecology, Evolution, and Systematics* 40.
- Shen, Y. et al. (2019). "Species persistence in spatially regular networks". In: *Ecological Modelling* 406, pp. 1–6.
- Skutella, M. (2009). "An introduction to network flows over time". In: *Research trends in combinatorial optimization*. Springer, pp. 451–482.
- Solé, R. V. and J. M. Montoya (2001). "Complexity and fragility in ecological networks". In: *Proceedings of the Royal Society of London. Series B: Biological Sciences* 268.1480, pp. 2039–2045.
- Stasko, T., B. Levine, and A. Reddy (2016). "Time-expanded network model of train-level subway ridership flows using actual train movement data". In: *Transportation Research Record* 2540.1, pp. 92–101.
- Streib, L. et al. (2020). "How does habitat connectivity influence the colonization success of a hemimetabolous aquatic insect? – A modeling approach". In: *Ecological Modelling* 416, no. 108909.
- Thompson, P. L., B. Rayfield, and A. Gonzalez (2015). "Robustness of the spatial insurance effects of biodiversity to habitat loss". In: *Evolutionary Ecology Research* 16.6, pp. 445–460.
- Titeux, N. et al. (2016). "Biodiversity scenarios neglect future land-use changes". In: *Global Change Biology* 22.7, pp. 2505–2515.
- Tonkin, J. D. et al. (2018). "The role of dispersal in river network metacommunities: Patterns, processes, and pathways". In: *Freshwater Biology* 63.1, pp. 141–163.
- Turner, M. G. and C. L. Ruscher (1988). "Changes in landscape patterns in Georgia, USA". In: *Landscape Ecology* 1.4, pp. 241–251.
- Upadhyay, S. et al. (2017). "A network theoretic study of ecological connectivity in Western Himalayas". In: *Ecological Modelling* 359, pp. 246–257.
- Urban, D. L. and T. Keitt (2001). "Landscape connectivity: A graph-theoretic perspective". In: *Ecology* 82.5, pp. 1205–1218.
- Urban, D. L. et al. (2009). "Graph models of habitat mosaics". In: *Ecology Letters* 12.3, pp. 260–273.
- Urban, M. C. (2015). "Accelerating extinction risk from climate change". In: *Science* 348.6234, pp. 571–573.
- Van Nouhuys, S. and I. Hanski (2002). "Colonization rates and distances of a host butterfly and two specific parasitoids in a fragmented landscape". In: *Journal of Animal Ecology* 71.4, pp. 639–650.
- Wasserman, S. and K. Faust (1994). *Social network analysis: Methods and applications*. Vol. 8. Cambridge University Press.
- Watts, D. J. and S. H. Strogatz (1998). "Collective dynamics of 'small-world' networks". In: *Nature* 393.6684, pp. 440–442.
- Wolsey, L. A. (2008). "Mixed integer programming". In: *Wiley Encyclopedia of Computer Science and Engineering*.
- Zetterberg, A., U. M. Mörtberg, and B. Balfors (2010). "Making graph theory operational for landscape ecological assessments, planning, and design". In: *Landscape and urban planning* 95.4, pp. 181–191.

Zurell, D. et al. (2016). "Benchmarking novel approaches for modelling species range dynamics". In: *Global change biology* 22.8, pp. 2651–2664.

Chapter 7

Publications

This cumulative dissertation is adapted from the following three scientific publications:

1. Heer, H., Streib, L., Kattwinkel, M., Schäfer, R. B., & Ruzika, S. (2019). Optimisation model of dispersal simulations on a dendritic habitat network. *Scientific reports*, 9(1), 1-11. <https://www.nature.com/articles/s41598-019-44716-z>
2. Heer, H., Streib, L., Schäfer, R. B., & Dieckmann, U. (2021). Indicators for assessing the robustness of metapopulations against habitat loss. *Ecological Indicators* 121 (2021): 106809. <https://doi.org/10.1016/j.ecolind.2020.106809>
3. Heer, H., Streib, L., Schäfer, R. B., & 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>

Chapter 8

Author's contributions

Article I

Title: Optimisation model of dispersal simulations on a dendritic habitat network

Authors: Henriette Heer, Lucas Streib, Mira Kattwinkel, Ralf B. Schäfer, and Stefan Ruzika

Status: Published in 2019 in Scientific Reports, 9(1), 1-11.

Contributions: **Heer** developed the optimisation model in consultation with Ruzika, Streib developed the simulation model in consultation with Kattwinkel and Schäfer, **Heer** performed the analysis of the modelling results. **Heer** and Streib took the lead in writing the manuscript; All authors provided critical feedback and helped shape the research, analysis and manuscript.

Article II

Title: Indicators for assessing the robustness of metapopulations against habitat loss

Authors: Henriette Heer, Lucas Streib, Ralf B. Schäfer, and Ulf Dieckmann

Status: Published in 2021 in Ecological Indicators, 121(2021), 106809

Contributions: **Heer** performed the research in collaboration with Dieckmann; Schäfer and Streib helped with the statistical analysis; All authors wrote the manuscript.

Article III

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

Authors: Henriette Heer, Lucas Streib, Ralf B. Schäfer, and Stefan Ruzika

Status: Published in 2020 in PLoS One 15(10), e0240940

Contributions: **Heer** developed the models in consultation with Ruzika; Streib and Schäfer provided critical feedback and helped shape the research and analysis; **Heer** led the writing with contributions from all other authors.

Henriette Heer

Curriculum Vitae

EDUCATION

- since Dec. 2017 **PhD Candidate**, *Institute for Environmental Sciences at University of Koblenz-Landau, Campus Landau*
interdisciplinary project STRESSOR
- Oct. 2016 – Dec. 2017 **PhD Candidate**, *Mathematical Institute at University of Koblenz-Landau, Campus Koblenz*
interdisciplinary project STRESSOR
- Feb. 2013 – Dec. 2015 **Masters of Mathematics**, *Technical University of Kaiserslautern*
Specialization in Algebra and Number Theory
Minor in Business and Economics
Title of Masters Thesis: Classification of Torsion Points on Certain Families of Elliptic Curves over Quartic Number Fields
- Oct. 2009 – Feb. 2013 **Bachelor of Mathematics**, *Technical University of Kaiserslautern*
Specialization in Algebra, Geometry and Symbolic Computation
Minor in Business and Economics
Title of Bachelor Thesis: Variationen der Hermite Normal Form

RESEARCH VISITS

- Oct. 2018 – Dec. 2018 **Visiting Researcher**, *Evolution and Ecology Program*
International Institute of Applied Systems Analysis (IIASA)
Laxenburg, Austria

SCIENTIFIC PUBLICATIONS

1. **Heer, H.**, Streib, L., Schäfer, R. B., & Dieckmann, U. (2021). Indicators for assessing the robustness of metapopulations against habitat loss. *Ecological Indicators* 121 (2021): 106809. <https://doi.org/10.1016/j.ecolind.2020.106809>
2. **Heer, H.**, Streib, L., Schäfer, R. B., & 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>

3. Streib, L., Kattwinkel, M., **Heer, H.**, Ruzika, S., & Schäfer, R. B. (2020). How does habitat connectivity influence the colonization success of a hemimetabolous aquatic insect?-A modeling approach. *Ecological Modelling*, 416, 108909. <https://doi.org/10.1016/j.ecolmodel.2019.108909>
4. **Heer, H.**, Streib, L., Kattwinkel, M., Schäfer, R. B., & Ruzika, S. (2019). Optimisation model of dispersal simulations on a dendritic habitat network. *Scientific reports*, 9(1), 1-11. <https://www.nature.com/articles/s41598-019-44716-z>
5. **Heer, H.**, McGuire, G., & Robinson O. (2016). JKL-ECM: An Implementation of ECM using Hessian Curves. *LMS J. Comput. Math.* 19 (Special issue A) 83–99

Declaration of Authorship

I, Henriette Heer, hereby declare that I autonomously conducted the work presented in this PhD thesis entitled “Mathematical models of species dispersal and the resilience of metapopulations against habitat loss”. All used assistances and involved contributors are clearly declared. This thesis has never been submitted elsewhere for an exam, as thesis or for evaluation in a similar context; to any department of this university or any scientific institute. I am aware that a violation of the aforementioned conditions can have legal consequences.

Place, date

Signature