1-1-2013

Uncovering the Structures In Ecological Networks: Multiple Techniques For Multiple Purposes

Peng Gao
University of South Carolina - Columbia

Recommended Citation

Gao, P.(2013). *Uncovering the Structures In Ecological Networks: Multiple Techniques For Multiple Purposes.* (Doctoral dissertation). Retrieved from https://scholarcommons.sc.edu/etd/2567

UNCOVERING THE STRUCTURES IN ECOLOGICAL NETWORKS: MULTIPLE TECHNIQUES FOR MULTIPLE PURPOSES

by

Peng Gao

Bachelor of Agriculture
Yangzhou University, 2004

Master of Agriculture
Chinese Academy of Sciences, 2007

Submitted in Partial Fulfillment of the Requirements

For the Degree of Doctor of Philosophy in

Geography

College of Arts and Sciences

University of South Carolina

2013

Accepted by:

John Kupfer, Major Professor

Don Edwards, Committee Member

Diansheng Guo, Committee Member

Michael Hodgson, Committee Member

Lacy Ford, Vice Provost and Dean of Graduate Studies

DEDICATION

This work is dedicated to my grandmother who brought me up and supported me all through my life.

ACKNOWLEDGEMENTS

I have enjoyed a great time both personally and professionally, since I joined the Ph. D program at the Department of Geography, University of South Carolina six years ago. The accomplishment of the Ph. D degree cannot be achieved without the help from many people. First and foremost, I would like to thank my advisor, Dr. John A. Kupfer, for him insight, mentorship, patience and encouragement in these years and in the future. He is the greatest advisor that I ever have and has brought me a lot of knowledge and fun. I am grateful to my committee members, Dr. Don Edwards, Dr. Diansheng Guo, and Dr. Michael Hodgson for their guidance and suggestions on the dissertation. I wish to thank the team of CISA (the Carolinas Integrated Sciences and Assessments), department staff members, and graduate fellows for their encouragement and support.

I am fortunate to have Dr. Charles Adams' help with my speech that has made my teaching possible. I consider the teaching experience as important as the accomplishment of the dissertation. I am very grateful to continuous attentions and encouragement on me through many years from professors in Yangzhou University, especially Dr. Manrong Chen and Dr. Shixue Yin. Finally, I would like to give my deepest thanks to my grandparents, parents, and other family members for their passionate and

unconditional support. This work was supported by the Dean's Dissertation Fellowship from the College of Arts and Sciences at the University of South Carolina.

ABSTRACT

Ecosystem structure and function are the product of biological and ecological elements and their connections and interactions. Understanding structure and process in ecosystems is critical to ecological studies. Ecological networks, based on simple concepts in which biological and ecological elements are depicted as nodes with relationships between them described as links, have been recognized as a valuable means of clarifying the relationship between structures and process in ecosystems. Ecological network analysis has benefited from the advancement of techniques in social science, computer science, and mathematics, but attention must be paid to whether the designs of these techniques follow ecological principles and produce results that are ecologically meaningful and interpretable. The objective of this dissertation is to examine the suitability of these methods for various applications addressing different ecological concerns. Specifically, the studies that comprise this dissertation test methods that reveal the structure of various ecological networks by decomposing networks of interest into groups of nodes or aggregating nodes into groups. The key findings in each specific application are summarized below.

In the first paper, REgionalization with Clustering And Partitioning (GraphRECAP) (Guo 2009) and Girvan and Newman's method (Girvan and Newman 2002) were compared in the study of finding compartments in the habitat network of ring-tailed

lemurs (*Lemur catta*). The compartments are groups of nodes in which lemur movements are more prevalent among the groups than across the groups. GraphRECAP found compartments with a larger minimum number of habitat patches in compartments. These compartments are considered to be more robust to local extinctions because they had stronger within-compartment dispersal, greater traversability, and more alternative routes for escape from disturbance. The potential defect of the Girvan and Newman's method, an unbalanced partitioning of graphs under certain circumstances, was believed to account for its lower performance.

In the second study, Modularity based Hierarchical Region Discovery (MHRD) and Edge ratio-based Hierarchical Region Discovery (EHRD) were used to detect movement patterns in trajectories of 34 cattle (*Bos taurus*), 30 mule deer (*Odocoileus hemionus*), and 38 elk (*Cervus elaphus*) tracked by an Automated Telemetry at Starkey National Forest, in northeastern Oregon, USA. Both methods treated animal trajectories as a spatial and ecological graph, regionalized the graph such that animals have more movement within the regions than across the regions, and then investigated the movement patterns on the basis of regions. EHRD identified regions that more effectively captured the characteristics of different species movement than MHRD. Clusters of trajectories identified by EHRD had higher cohesion within clusters and better separation between clusters on the basis of attributes of trajectories extracted from the regions. The regions detected by EHRD also served as more effective predictors for classifying trajectories of different species, achieving a higher classification accuracy with more simplicity. EHRD had better performance, because it did not rely on the null model that MHRD compared to, but invalid in this application.

In the third study, a proposed Extended Additive Jaccard Similarity index (EAJS) overcame the weakness of the Additive Jaccard Similarity index (AJS) (Yodzis and Winemiller 1999) in the aggregation of species for the mammalian food web in the Serengeti ecosystem. As compared to AJS, the use of the EAJS captured the similarity between species that have equivalent trophic roles. Clusters grouped using EAJS showed higher trophic similarities between species within clusters and stronger separation between species across clusters as compared to AJS. The EAJS clusters also exhibited patterns related to habitat structure of plants and network topology associated with animal weights. The consideration of species feeding relations at a broader scale (i.e., not limited in adjacent trophic levels) accounted for the advantages of EAJS over AJS.

The concluding chapter summarizes how the methods examined in the previous chapters perform in different ecological applications and examines the designs of these algorithms and whether the designs make ecological sense. It then provides valuable suggestions on the selections of methods to answer different ecological questions in practice and on the development and improvement of more ecological-oriented techniques.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Chapter 1 Introduction

1.1 Introduction

The study of networks and application of network theory has become common in fields as diverse as physics, sociology, computer science, transportation, and economics over the last two to three decades. Network theory provides a means of understanding how complex groups of interrelated phenomena interact, function and produce unexpected kinds of behavior that may not be predictable from knowledge of the individual parts. In ecology, ecological networks have been recognized as powerful models to elucidate the relationship between structures and processes (Dale and Fortin 2010), and network theory has been applied both to examine fundamental ecological questions and as a tool for managing and protecting biodiversity (Cumming et al. 2010).

Network theory fundamentally concerns itself with the study of graphs, a rapidly growing area of interest in fields such as biogeography, landscape ecology and conservation biology, to name just a few (Kupfer 2012). In a network- (or graph-) based approach, biological and ecological entities are treated as nodes and their interactions are depicted as links among nodes. Ecological networks take different forms and have been used to address different ecological concerns, depending on the biological and ecological entities and their relationships that nodes and links represent. Food webs in which species (nodes) interact through trophic relationship (links) are perhaps the most typical ecological networks that have long been investigated in ecology (Elton 1927). In recent years, networks that consist of habitat patches represented by nodes and dispersal routes among habitat patches as links, have also been introduced into ecology (often known as species habitat networks) to examine habitat connectivity and fragmentation at a board scale (Urban and Keitt 2001).

Despite the many forms that ecological networks may take, interactions among the elements in the system are viewed as being key to producing structural complexity and determining and maintaining functionality of ecosystems. Feeding relations in food webs influence the dynamics and persistence of populations, shape food web structure, and govern ecological processes in a system (De Ruiter et al. 2005). The dispersal of species among habitat patches in a species habitat network, which influences recolonization of unoccupied habitat patches and rescue following local extinctions, is crucial to the robustness of ecosystems to disturbance and the persistence of metapopulations. Therefore, the patterns of relations in ecological networks and how these patterns relate to the characteristics of the networks are a central focus of network analysis (Webb and Bodin 2008; Cumming et al. 2010).

The rapid growth of network analysis principles and techniques from computer science, mathematics and social science has brought thoughtful theories and powerful tools to address problems in ecological network analysis. However, ecological networks have their own characteristics that distinguish them from other networks. For example, a node in a social network can connect to other nodes (e.g., one person can reach others) via a relatively low number of links (known as the ‘small world effect’: (Schnettler 2009)). In food webs, the paths between nodes are often even shorter than in social networks (Dunne et al. 2002; Williams et al. 2002). Social networks are commonly scale-free networks in which the distribution of node degree follows a power-law distribution. Ecological networks do not generally follow scale-free distributions (Cumming et al. 2010), so the design of an algorithm for ecological network analysis should follow or be compatible with fundamental principles in ecology. The results should not only be examined in a computational way, but also be able to reveal patterns that are ecologically meaningful and interpretable.

Network analysis has a long history in science, and many methods have been developed to reveal the structure, understand the complexity, and capture the dynamics of networks from various perspectives and for different purposes. One common application of network analysis is to decompose a network or graph into groups of nodes according to criteria or definitions. These groups usually have certain properties in common or play similar roles in the networks. This approach is meant to effectively and efficiently reduce the complexity of an otherwise complex system to reveal the structure of networks. For example, species in food webs can be categorized into producers, primary consumers, secondary consumers, tertiary consumers etc., which reveals energy pyramids and their trophic roles in food webs.

In this dissertation, I focus on one of the primary tasks of network analysis, the reduction of system complexity as a means for revealing the structure of ecological networks. I do so by employing methods that decompose the networks of interest into groups of nodes or aggregate nodes into groups. I compare multiple techniques and examine their suitability when these techniques are adapted to analyze various ecological networks for different applications. Some of the techniques explored in this dissertation were originally developed in computer science or mathematics but not specifically for ecological concerns. Their application here is, in part, an attempt to demonstrate their potential value in the realm of ecological applications. However, this dissertation does not only attempt to answer the question "which one is better for ecological applications", but also attempts to explore "why it is better for ecological applications" by examining the designs of the algorithms and whether the designs make ecological sense.

The dissertation consists of three separate, but related studies. Each study focuses on one type of ecological network and addresses ecological concerns for that particular type of ecological network. The general research questions "which one is better for ecological applications" and "why it is better for ecological applications" are specified in three manuscripts presented as Chapters 2-4 in this dissertation (Table 1.1). These are outlined below. Terminologies that are frequently used in this dissertation are list in Table 1.2.

Objective 1: *Compare the algorithm of Girvan and Newman with Graph-based REgionalization with Clustering And Partitioning (GraphRECAP) in the application of detecting compartments in a species habitat network*

Chapter 2 titled "Identifying Functionally-Connected Habitat Compartments with a Novel Regionalization Technique" addresses this question. Species habitat networks or

graphs are a type of ecological networks in which a set of nodes (habitat patches) are connected by links representing inter-patch-dispersal. Compartments are groups of habitat patches in which interactions (e.g., dispersal linkages) are more prevalent among the groups than between nodes across groups. Two methods, the algorithm of Girvan and Newman (Girvan and Newman 2002) and GraphRECAP (Guo 2009) were applied to habitat network of ring-tailed lemurs (*Lemur catta*) in southern Madagascar which consisted of habitats and dispersal data of ring-tailed lemurs. The evaluation was focused on the ecological traits of the compartments found by two methods. Compartment characteristics such as the number of habitat patches in the compartments were examined for the benefit that a larger number of habitat patches in a compartment facilitates patch recolonization of local losses from within-compartment sources. Three measures of network connectivity and traversability were also used for evaluation: the connection strength of habitat patches in the compartments (modularity), the ease of individual organism movements (Harary Index), and the degree of alternative route presence (Alpha Index). Compartments identified by GraphRECAP had stronger within-compartment dispersal, greater traversability, more alternative routes for escape from disturbance, and a larger minimum number of habitat patches within compartments, all of which are more desirable traits for ecological networks. GraphRECAP offers an improved means for characterizing the spatial structure of populations in terms of improving habitat connectivity and increasing the persistence of populations.

Objective 2: *Contrast modularity-based and edge ratio-based Hierarchical Region Discovery in terms of their abilities to detect movement patterns in animal trajectories*

Chapter 3, titled "Detection of Regions in Spatial Graphs: a New Approach to Animal Trajectory Analysis" answers this question. It examines patterns of animal movement by treating animal trajectories as a spatial and ecological graph and then regionalizing the trajectories such that animals have more movement within the regions than across the regions. Nodes are spatial clusters of telemetry locations in animal trajectories and links are the movement of animals among these telemetry locations. Such a spatial graph is unique, because nodes in the graph bear spatial information and are connected by the movement of animals. Hierarchical Region Discovery finds groups of nodes in the spatial graph built from animal trajectories that meet two requirements. First, the groups of nodes must be spatially contiguous. Therefore, each group of nodes forms a region. Second, animals have more movement within than across the regions. Modularity and edge ratio are two measures quantifying the movement within regions and governing the process of finding regions. The movement of cattle (*Bos taurus*), mule deer (*Odocoileus hemionus*), and elk (*Cervus elaphus*) tracked by an Automated Telemetry at Starkey National Forest, in northeastern Oregon, USA in June 1995 was analyzed by extracting attributes of the trajectories based on the regions found by modularity-based and edge ratio-based Hierarchical Region Discovery. The attributes were further used to cluster and classify these trajectories (assuming we do not know what species the trajectories represented). The quality of clusters and accuracy and simplicity of decision tree classification were used to evaluate the ability of detected regions to capture the characteristics of different species movement. While modularity has been widely used, edge ratio more effectively captured the characteristics of the animal movement. The reasons why regions defined by the edge ratio were more suitable for this particular ecological application were also explored by examining the designs of the modularity and edge ratio. Edge ratio-based Hierarchical Region Discovery provides an alternative approach to interpreting animal movement on the basis of regions and to discover unknown patterns.

Objective 3: *Contrast clusters of species in food webs aggregated on the basis of the Additive Jaccard Similarity (AJS) and the Extended Additive Jaccard Similarity (EAJS)*

Chapter 4 titled “Uncovering Food Web Structure Using a Novel Trophic Similarity Measure” answers this question. This study focused on food webs, the most typical ecological networks studied for a long time. Two trophic similarity measures (AJS and EAJS) were used to aggregate plants and mammalian species in the food web of the Serengeti ecosystem in northern Tanzania and southern Kenya. AJS only considers shared predators and prey at adjacent trophic levels to measure the trophic similarity between two species, while EAJS incorporates not only the similarity of shared predators and prey at adjacent trophic levels but all the trophic levels associated with each species. Compared to AJS, the clusters of species on the basis of EAJS had higher quality which means that species in the same clusters have higher similarity and species in different clusters have higher dissimilarity in terms of their trophic relationships in the food web. Clusters found on the basis of EAJS also reflected factors known to structure food webs. Plants of the same habitat tended to be grouped in same clusters. The grouping of animals was related to their weights. The advantage of EAJS lies in the fact that it is designed to consider species feeding relations in food webs in a broad scale (i.e., not limited to adjacent trophic levels). EAJS provides an approach to revealing the patterns of trophic relations among species in food webs and exploring known and unknown factors shaping food web structure.

1.2 Literature Review

1.2.1 Unique Properties of Ecological Networks

Network analysis has a long history. The investigation of Königsberg Bridge problem by the great mathematician Leonard Euler in 1736 is regarded as the earliest study of networks. Network analysis in recent decades has focused on the properties of real-world networks and the dynamics of networks (Newman et al. 2006).

As a branch of network analysis in general, the study of ecological network has followed the same trend. The analysis of ecological networks has found special properties that are regulated by fundamental ecological principles. In food webs, typically only 10% energy can be transferred from one trophic level to the next trophic level. Energy pyramid of a typical food web may contain producers, primary, secondary, and tertiary consumers. Species at the level beyond the tertiary consumer are rare. This rule limits the number of links that connect two species (known as the shortest path between two nodes in graph theory). In social networks, the number of links that connect two nodes (e.g. one person reach other persons via social relations) is small (known as the 'small world effect': Schnettler 2009). Compared to these social networks, the paths (the number of links) between species are often even shorter in food webs

(Dunne et al. 2002; Williams et al. 2002). Moreover, networks such as World Wide Web and social networks have the features of small world effect and/or scale free (the distribution of node degree follows a power-law distribution) (Strogatz 2001; Albert and Barabasi 2002). Food webs do not share these general features with other types of networks (Camacho et al. 2002). However, food webs have their own rules that shape their complexity. Cohen et al. (1990) summarized five laws (e.g., cycles are rare; chains are short) that shaped food web structures while Williams and Martinez (2000) succeeded in predicting twelve properties (e.g. the fraction of top, intermediate, and basal species in a food web) of food webs using only two parameters: species number and connectance.

Species habitat networks also have their own properties, because the formation of species habitat networks is different from other types of social networks. Social networks such as Facebook are typically built through adding people and their personal connections to social networks. Species habitat networks often arise from the fragmentation of formerly contiguous habitats into habitat patches (Fortuna and Bascompte 2006). Therefore, it is unlikely that species habitat networks have the same property as scale free networks, though some habitat patches may have many links connected to them (Norberg and Cumming 2008).

1.2.2 Roles and Functions of Individual Nodes and Groups of Nodes in Ecological Networks

The properties of ecological networks discussed above mainly focus on all nodes and links in entire networks. The role of individual nodes has also been evaluated. Hubs are nodes in the networks that have many connections with them. The number of links associated with a node is called the degree of the node. So hubs are nodes with high degrees in networks. They have important roles in preventing the entire network from being decomposed into pieces or subgraphs. These nodes are known as keystone species in food webs, the extinction of which produces great impacts on the abundance of other species in ecosystems (Jordán 2009). In a species habitat network, these nodes are habitat patches which are critical to maintain the habitat connectivity for the entire graph (Minor and Urban 2008).

Besides the degree of nodes, many indices have been developed to assess the importance of the nodes in ecological networks and to identify these critical nodes for the stability of networks. *Closeness Centrality* (measuring the average distance of the focal node from all others in the graph) and *Betweenness Centrality* (the proportion of the shortest paths between all pairs of nodes that contains the focal node) are two indicators used both in the studies of food webs and species habitat networks. The merits of the two indicators are their considering the position of a node at a “meso-scale’’. In food web studies, Estrada (2007) compared indices that identify keystone species at local, global and “meso” scale. The author found that the “meso-scale’’ indicators are more important than others in determining the relative importance of species in epidemic spread and parasitism rates. In the study of habitat network of ring-tailed lemurs (*Lemur catta*), Bodin and Norberg (2007) argued that habitat patches with high Betweenness Centrality are crucial to the landscape traversability and serve as backbones of the habitat network, because they decrease the overall network distance between pairs of habitat patches.

On one hand, all nodes and links in an entire network produce the properties for the ecological network. One the other hand, some individual nodes play critical roles in maintaining the structure and functions of the ecological network. However, other important functions of ecological networks are the products of groups of nodes and their interactions. A typical example is the producers in food webs that take the energy and nutrition from the environment to the ecosystem and serve as the fundaments of food webs. Actually, grouping nodes in ecological networks is one efficient way of reducing the complexity of ecological network to better understand their structures and how the structures provide different functions.

Nodes in ecological networks can be grouped according to various criteria or definitions. The structures of ecological networks are therefore exposed in different ways. In the studies of food webs, one classical way of aggregating species is clustering them according to their trophic similarity conducted by Yodzis and Winemiller (1999). They compared the performance of multiple criteria (e.g., additive and multiplicative Jaccard similarity) in aggregating 116 species in a food web from a tropical flood plain into trophic groups. They concluded that additive Jaccard similarity is better than multiplicative similarity in terms of producing more consistent and ecologically-interpretable patterns of aggregation. While additive Jaccard similarity is popular, it has documented weaknesses as well: the lower ability to identify species with equivalent trophic roles, especially when they do not share the same predators and prey. For example, two herbivores that feed on totally different plants or are eaten by different carnivores are separated into different groups on the basis of additive similarity, even though they may play equivalent trophic roles in a food web. The advance of social network analysis introduced the concept of “regular equivalence” to the studies of aggregating species in food webs according their trophic roles. Luczkovich et al. (2002) adapted this concept to aggregate species into isotrophic groups. Species in the same isotrophic group have the same or similar trophic roles in a food web, feeding on and being preyed upon by equivalent species (e.g., herbivores feed on plants and are eaten by carnivores).

Another criterion used to group species measures the strength of interactions among the species. The methods based on this criterion are called compartment detection in network analysis which finds groups of nodes such that nodes have more connections within groups than across groups. Raffaelli and Hall (1992) determined the compartments in food webs by examining the frequency distributions of trophic similarity coefficient of the species and mapping the species in ordination plots based on the assumption that species which are more similar in their trophic interactions will be closer together in ordination plots. Krause et al. (2003; 2009) adapted an odds ratio method (Frank 1995) which iteratively reassigns taxa to compartments to maximize the odds that links occur within compartments versus links between compartments. Studies over decades have suggested an intermediate level of compartmentalization enhance robustness of food webs (May 1972; Pimm 1979; Teng and McCann 2004).

Groups of habitat patches (nodes) in species habitat networks that are isolated from habitat patches in other groups (i.e., no links connecting habitat patches between groups) are called components. Components in species habitat networks are caused by habitat fragmentation. A habitat network where formerly every habitat patches can be connected to others may be broken down into isolated components, because habitat fragmentation removes the habitat patches or prohibits the movement among patches that are critical to the entire habitat connectivity. Number of components and the largest components are two indices used to assess the vulnerability of habitat connectivity to different levels of fragmentation. For example, Lookingbill et al. (2010) used the two indices to evaluate the habitat connectivity of Delmarva fox squirrel inhabiting forested areas on the Delmarva Peninsula, USA. They examined the change of the two indices under different scenarios of abilities that Delmarva fox squirrels disperse between habitat patches.

However, Bodin and Norberg (2007: p34) argued that "the binary perspective of components, wherein a set of nodes is completely isolated from the rest, is deficient in detecting a more continual degree of compartmentalization in the landscape". They suggested using the compartments to capture the continual degree of habitat connectivity. They adapted the widely-used Girvan and Newman method (Girvan and Newman 2002) to identify the compartments in habitat network of ring-tailed lemurs (*Lemur catta*) in southern Madagascar.

1.2.3 Compartment Detection Methods

Readers are referred to Fortunato (2010) for a comprehensive review on compartment detection methods. Methods that are highly related to this dissertation are introduced below with a particular emphasis on the compartment detection methods for spatial graphs which are examined in Chapter 3 of the dissertation.

The Girvan and Newman method (Girvan and Newman 2002) has been recognized as a milestone in the field of compartment detection (Fortunato 2010). They focused on the concept of edge betweenness, defined as the number of shortest paths between pairs of vertices that run along it. Compartments are defined by systematically removing edges that have high edge betweenness. Although the method has been applied to the study of a wide range of networks (e.g., marine food webs: Rezende et al. 2009, metabolic networks: Ono et al. 2005, protein interaction networks: Dunn et al. 2005), it has also been criticied that it may yiled unbanlanced partitioning under certain circumentances (Chen and Yuan 2006).

The other remarkable contribution by Newman and Girvan is that they introduced *modularity* to evaluate compartment methods by quantifing how good the detected compartments are (Newman and Girvan 2004). By definition,it measures "the fraction of the edges in the network that connect nodes of the same type (i.e., within-compartment edges) minus the expected value of the same quantity in a network with the same compartment divisions but random connections between the nodes" (Newman and Girvan 2004: p7). Their contribution lies in two facts. First, it offers the most popular null model where nodes are connected in a random manner, subjected to the constraint that the expected degree of each node matches the degree of the node in the original graph. Second, it quanitiatively expresses the strength of compartments (Fortunato 2010). Huge number of modularity based optimization methods have been created since then including GraphRECAP (Graph-based REgionalization with Clustering and Partitioning) method (Guo 2009) that is employed in the dissertation and Clauset et al. (2004) ‘s method that improves the computational efficiency.

When nodes in a network are located in a space equipped with a metric, the network is recognized as a spatial graph (Barthelemy 2011). The important role of spatial information in network analyses such as the evolution of transportation (Chorley and Haggett 1971), human migration (Guo 2011) has been recognized. Guo (2011) developed GraphRECAP (Graph-based REgionalization with Clustering and Partitioning) to discover spatially contiguous compartment patterns in the migration data of U.S. which contains over 700,000 county-to-county migration flows. The method efficiently reduced the complexity in the migration data and uncovered patterns that strongly related to space such as "core-suburban relationship" from a network perspective. Later on, Guo et al. (2010; 2012) applied the method on trajectory analysis by treating the movement of vehicles across space as spatial networks and regionalizing the spatial networks (i.e., finding spatially contiguous compartments in the spatial networks). Enforcing spatial constraint is also valuable in facilitating the visualization and interpreting ecological or biogeographic data. In the study of forest patterns for 2,109 watersheds in the continental U.S., Kupfer et al. (2012) identified hierarchical regions based on measures of forest extent, connectivity, and change by enforcing spatial constraint into a traditional hierarchical clustering method. The detected forest pattern regions had more desirable properties than those from non-spatial clustering methods and reflected the influence of natural and anthropogenic factors structuring forest extent and fragmentation.

Table 1.1 A summary of the types of ecological networks examined, study goals, and methods for Chapters 2-4.

Chapter	Type of Ecological Network	Goal(s)	Methods
2	Species habitat network	To identify groups of habitats (compartments) that are closely linked by dispersal of the species, *Lemur catta* (ring-tailed lemur)	Algorithm of Girvan and Newman and REgionalization with Clustering And Partitioning (GraphRECAP)
3	Spatial network of animal movement	To detect groups of nodes that are spatially contiguous and have more animal movement within groups than across groups (the groups of nodes are called regions in that they are spatially contiguous) To reveal movement pattern of different species based on detected regions	Modularity-based Hierarchical Region Discovery (MHRD) and Edge ratio-based Hierarchical Region Discovery (EHRD)
4	Food web	To aggregate species based on their trophic similarity	Additive Jaccard Similarity (AJS) and Extended Additive Jaccard Similarity (EAJS)

Table 1.2 Terminologies used in the dissertation

Terminology	Definition
Network	A mathematical model consists of nodes and links that represent the connections between pairs of nodes.
Ecological Network	A network in which nodes and links represent biological and ecological entities and their interactions.
Spatial Network	A network for which the nodes are located in a space equipped with a metric.
Compartment(s)	Groups of nodes in a network with many edges joining nodes of the same group but comparatively few edges joining nodes of different groups.
Region(s)	Compartments in a spatial network in which nodes are spatially contiguous.
Regionalization	the process that detects regions in a spatial graph.
Spatial Cluster(s)	A spatial cluster is an aggregation of telemetry locations in animal trajectories by Shared Nearest Neighbors (SNN) method. A spatial cluster has spatial information can be spatially adjacent to other spatial clusters. Its boundary is the merged thiessen polygons surrounding the telemetry locations belonging to it. Spatial clusters serve as nodes in a spatial graph. (See page 50 -51 and Figure 3.1)

Chapter 2 Identifying Functionally-Connected Habitat Compartments with a Novel Regionalization Technique[1]

Abstract

Landscape ecologists have increasingly turned to the use of landscape graphs in which a landscape is represented as a set of nodes (habitat patches) connected by links representing inter-patch-dispersal. This study explores the use of a novel regionalization method, GraphRECAP (Graph-based REgionalization with Clustering And Partitioning), to detect structural groups of habitat patches (compartments) in a landscape graph such that the connections (i.e. the movement of individual organisms) within the groups are greater than those across groups. Specifically, we mapped compartments using habitat and dispersal data for ring-tailed lemurs (*Lemur catta*) in an agricultural landscape in southern Madagascar using both GraphRECAP and the widely-used Girvan and Newman method. Model performance was evaluated by comparing compartment characteristics and three measures of network connectivity and traversability: the connection strength of habitat patches in the compartments (modularity), the potential ease of individual organism movements (Harary Index), and the degree of alternative route presence (Alpha Index). Compartments identified by GraphRECAP had stronger within-

---

[1] The manuscript is in view from *Landscape Ecology*

compartment connections, greater traversability, more alternative routes, and a larger minimum number of habitat patches within compartments, all of which are more desirable traits for ecological networks. Our method could thus facilitate the study of ecosystem resilience and the design of nature reserves and landscape networks to promote the landscape-scale dispersal of species in the fragmented habitats.

2.1 Introduction

Recent decades have been marked by efforts to understand the characteristics and dynamics of spatially-structured populations, local populations that occupy discrete habitat patches connected by individual dispersal (Hanski and Gilpin 1991; Gilarranz and Bascompte 2012). These efforts have been driven to a large degree by concerns related to the conservation and management of species in spatially-heterogeneous landscapes, often as they relate to the ease with which organisms or other phenomena can move across the landscape (i.e., landscape connectivity). Habitat loss and fragmentation caused by human activities remain the greatest ongoing threat to the survival of many species (Benton et al. 2003; Kerr and Deguise 2004), but maintaining functional linkages among habitat patches facilitates the acquisition of spatially- and temporally-variable resources (Clobert et al. 2009), helps to offset the inherent risks to smaller populations, and is thus crucial for the viability of vulnerable populations in landscapes transformed by human actions (Brooks 2003). The persistence of spatially-structured populations is also affected by the spread of diseases and invasive species, which in turn are influenced by the arrangement of habitat patches and landscape connectivity (Urban et al. 2009; Bellisario et al. 2010).

Network analysis has become a useful tool in the study of spatially-structured populations. In landscape ecological applications, a network or graph corresponds to a landscape in which the nodes represent habitat patches and the links indicate connections between them via dispersal. In much the same way that metrics based on island biogeography theory were used in the past (e.g., patch area, nearest patch distance: Kupfer 1995), new measures based on network theory are being developed and implemented to describe aspects of landscape pattern (e.g., Rayfield et al. 2011; Foltête et al. 2012) and gauge the potential impacts of habitat loss on biodiversity (Kupfer 2012). Interest has especially centered on the characteristics of individual elements, for example, nodes and links and their role in network pattern (e.g., measures of node centrality), or how overall network properties change with node or link removal (Saura and Rubio 2010; Reunanen et al. 2012; Ziolkowska et al. 2012). Less research has been conducted at the level of components, groups of interconnected nodes. In this paper, we focus on compartments, a variation of components'.

The fundamental difference between compartments and components involves the level of interaction among nodes (habitat patches, in this case) and node groups. Components, by definition, consist of linked nodes that are isolated from nodes in other components, that is, no paths exist between nodes of different components. Functionally, this means that an organism in a given patch could move to other patches in the same component, but would be unable to reach patches in other components (Bodin and Norberg 2007). While based on a similar premise as components, compartments represent groups of nodes in which interactions (e.g., dispersal linkages) are more prevalent among the member nodes than between nodes across groups, but the latter still occur. Bodin and Norberg (2007) argued that the use of analyses focused on compartments is preferable and more realistic to those using components because it better captures the varying degrees of connectivity that exist in most landscapes rather than imposing the strict limitation that node groups are isolated from one another. Compartments were first introduced to network analyses in the study of food web structure (Pimm 1979) but have been used in other fields, including landscape ecology (e.g., Bellisario et al. 2010). The contribution of compartments to metapopulation persistence, for example, has been recognized (e.g., Minor and Urban 2008; Urban et al. 2009).

One of the challenges to a more widespread focus on habitat compartments involves the detection of functional compartments in complex graphs, though several approaches and methodologies have been developed by scholars from multiple disciplines (Fortunato 2010). One of the most influential methods for delineating compartments is that developed by Girvan and Newman (2002), which has been applied to the study of a wide range of networks (e.g., marine food webs: Rezende et al. 2009, metabolic networks: Ono et al. 2005, protein interaction networks: Dunn et al. 2005). Their approach has also been used to decompose fragmented landscapes into compartments (Bodin and Norberg 2007) and is considered a viable means of characterizing habitat network structure and connectivity (Economo and Keitt 2010; Galpern et al. 2011).

In this study, we explore the use of a novel regionalization method, GraphRECAP (Graph-based REgionalization with Clustering and Partitioning: Guo 2009), to detect functionally-connected compartments in a landscape graph. This method decomposes the landscape graph by optimizing a measure of connection strength among compartments (modularity), but does so using a different, and potentially more straightforward, approach than that employed in other studies. We mapped compartments using habitat and dispersal data for ring-tailed lemurs (*Lemur catta*) in an agricultural landscape in southern Madagascar using both GraphRECAP and the Girvan and Newman method, and then compared the output of the methods using measures of network connectivity and traversability.

2.2 Methods

2.2.1 Compartmentalization Methods

Compartmentalization methods uncover groups of nodes in a network or graph such that the within-group connections are greater than between-group connections. The Girvan and Newman method (Girvan and Newman 2002) (hereafter GN) is intuitively a "bridge" cutting process (Figure 2.1). To find the bridges in a graph, Girvan and Newman (2002) extended the concept of vertex betweenness (Freeman 1977) to edge betweenness, defined as the number of shortest paths between pairs of vertices that run along it. In a graph that consists of compartments connected by a few inter-compartment edges (i.e., bridges), all shortest paths between different compartments must go along one of these inter-compartment edges. Thus, these bridges are characterized by high edge betweenness. Compartments are defined by systematically removing edges that have high edge betweenness. The pseudo-code for this process is simple and follows four basic steps, as outlined in Girvan and Newman (2002: p7823):

> "1. Calculate the betweenness for all edges in the network;
>
> 2. Remove the edge with the highest betweenness;
>
> 3. Recalculate betweennesses for all edges affected by the removal;
>
> 4. Repeat from step 2 until no edges remain."

The algorithm is available in the UCINET software (version 6.453) (Borgatti et al. 2002).

In this study, we use the GraphRECAP (Graph-based REgionalization with Clustering and Partitioning) method (Guo 2009) as a means for delineating habitat compartments. In line with the definition of compartments as having more within-compartment connections and fewer across-compartment connections, the objective of GraphRECAP is to maximize *modularity*, a measure of the strength of connections within compartments proposed by Newman and Girvan (2004), while decomposing the graph to compartments. The modularity of a compartment (i.e., within-compartment modularity) is calculated as the sum of modularity between all pairs of the nodes within the compartment. Specifically, let *a* and *b* be two nodes in a graph; the modularity between the two nodes of *a* and *b* is defined as (Equation 2.1):

$$\text{Modularity}\ (a, b) = \text{Actual Connections}\ (a, b) - \text{Expected Connections}\ (a, b) \quad (2.1)$$

The expected connection between *a* and *b* is calculated using the total connections associated with the two nodes (edges associated with *a* and *b* in the network) and the total connections in the graph (all the edges in the network). Let C_a and C_b represent the total connections incident on *a* and *b*, respectively, and let *C* be the total connections in the graph. The expected connection between *a* and *b* is calculated with Equation 2.2:

$$\text{Expected Connections } (a, b) = (Ca * C_b) / C \quad (2.2)$$

For example, the within-compartment modularity of a compartment *A* is the sum of modularity between each pair of nodes $a_1, a_2, \ldots, a_n$ within *A* (Equation 2.3):

$$\text{Within-compartment Modularity } (A) = \sum_{\mathrm{i}}^{\mathrm{n}} \sum_{\mathrm{j}}^{\mathrm{n}} \text{Modularity } (a_i , a_j) \quad (\mathrm{i} \neq \mathrm{j}) \ (2.3)$$

Similarly, we define the total modularity of a graph that is partitioned into a set of *k* compartments $A_1, A_2, \ldots, A_k$ as the sum of within-compartment modularity of each compartment (Equation 2.4).

$$\text{Modularity } (A_1, A_2, \ldots, A_k) = \sum_{\mathrm{i}}^{\mathrm{k}} (\text{within–compartment Modularity } (Ai)) \quad (2.4)$$

GraphRECAP first partitions a graph into two sub-graphs (compartments) such that the total modularity is maximized. Among the produced subgraphs, GraphRECAP chooses the best subgraph (which increases the total modularity the most if cut) and partitions it into two new subgraphs. This process is repeated to generate a specified number of compartments.

GraphRECAP begins by using a standard hierarchical clustering method (e.g., average linkage, complete linkage, or the Ward clustering method) to iteratively merge nodes that have highest modularity from the bottom up (Figure 2.2a). This step yields a dendrogram representing the nested grouping of nodes. It then iteratively examines all the edges of the tree or dendrogram built in the first step and cuts the one that maximizes the total within-compartment modularity when the tree is cut at that edge (Figure 2.2b). After each partition, a Tabu-based optimization method (Guo and Jin 2011) is used to further improve the partition. Tabu is a classic heuristic procedure for solving optimization problems. In this study, it checks the nodes that are immediate between the two partitioned compartments and switches them from one compartment to another in the hope of increasing the total within-compartment modularity. To improve the efficiency of searching, it uses memory structures that store the visited solutions and prevent them from being re-visited in a short time period. In other words, “Tabu” is a list of the recently visited solutions that are forbidden to be re-used in a short time period (Glover, 1990). By employing Tabu optimization, GraphRECAP attempts to avoid the trap of local optima and overcome a potential disadvantage of agglomerative algorithms.

2.2.2 Study Data

The two compartmentalization methods were applied to data on habitat and dispersal for ring-tailed lemurs (*Lemur catta*) in southern Madagascar that were utilized in a previous study (Bodin and Norberg 2007). The study area, target species, and initial data processing are briefly introduced here, but readers are referred to Bodin et al. (2006) and Bodin and Norberg (2007) for further details.

The study area is an agricultural landscape mosaic containing hundreds of small and dense dry-forest patches. Though the forest patches only occupy 3.5% of the study area, they provide habitat for several species of conservation interest, including *L. catta*, which feeds on fruits of more than 30 species of plants. Due to the low diversity of frugivores in the study area, especially compared to other tropical areas, *L. catta* is a key seed disperser for plant species in forest ecosystems threatened by habitat loss, for example, tamarind (*Tamarindus indica*: Mertl-Millhollen et al. 2011). The investigation of the movement of *L. catta* among forest patches thus provides valuable information for use in conservation efforts of not only *L.catta*, but also plant species dependent on it for seed dispersal services.

A supervised maximum likelihood classification was adopted to identify forest patches from Landsat 7 ETM+ satellite imagery obtained on 28 May 2000. Forest patches with areas > 1 ha were recognized as potential habitat for *L. catta*, and the capability for interpatch movement was assessed using a negative exponential dispersal kernel. The vagility of *L. catta* was estimated as the distance corresponding to a movement rate of only 5% of a normalized maximum, which was set at 1000 m, because previous studies have indicated that individuals can move this distance to forage per day. We assume that seeds can be dispersed between two patches by *L. catta* only when the estimated dispersal flux rate was higher than that estimated by the vagility.

Based on these assumptions, a graph was constructed with 259 nodes representing the forest patches and 1236 links connecting them between which *L. catta* can move and potentially disperse seeds. Because the full landscape graph contained a number of smaller disconnected components, GN and GraphRECAP were applied to just the largest component (Figure 2.3), which contained 183 nodes and 1058 connections. Landscape visualization was performed using Pajek (De Nooy et al. 2012).

2.2.3 Evaluation

Following Bodin and Norberg (2007), we partitioned the landscape graph using the two methods to generate ten compartments that achieved the highest sum of within-compartment modularity. The performance of the methods was evaluated by three indices. The first index was sum of within-compartment modularity of the ten compartments as proposed by Newman and Girvan (2004), which has been described in the Methods section. Modularity has been used as an index to evaluate the performance of compartmentalization methods, with higher values indicating denser connections between the nodes within compartments and sparser connections between nodes in different compartments (Newman 2006).

The second index that we adopt is the standardized Harary index, which measures how easy individual organisms can move within compartments and is intuitively and mathematically linked to landscape connectivity (Ricotta et al. 2000). In this case, a graph (G) that is composed of a set of m nodes N(G) and n links L(G) is represented by G(m, n). D (G) is defined as the distance matrix of G, where d_{ij} is the minimum number of links connecting nodes *i* and *j* (i.e., the shortest path between node *i* and *j*). R(G) is the reciprocal distance matrix whose elements r_{ij} are substituted by the reciprocal of d_{ij} in D(G). The Harary index (H) is the sum of the off-diagonal values in the upper triangular submatrix of R(G) (Ricotta et al. 2000). A higher r_{ij} indicates a lower number of links connecting nodes *i* and *j*. Ecologically, this suggests that organisms can move more easily from habitat patch *i* to *j*, because the two patches are more functionally connected. A higher H thus suggests a greater ease of traversability in a landscape graph.

The Standardized Harary index ($\overline{\mathrm{H}}$) in Equation 2.5 is used to make H comparable among landscape graphs that have differing numbers of nodes:

$$\overline{\mathrm{H}} = (\mathrm{H} - \mathrm{H}_{\mathrm{chain}})/(\mathrm{H}_{\mathrm{complete}} - \mathrm{H}_{\mathrm{chain}}) \quad (2.5)$$

where $\mathrm{H}_{\mathrm{chain}}$ and $\mathrm{H}_{\mathrm{complete}}$ are the Harary Index values of the chain and complete graphs. A chain graph is the least connected graph where no links could be removed without disconnecting the graph, while a complete graph is the most connected graph where no more links could be inserted into the graph. $\mathrm{H}_{\mathrm{chain}}$ and $\mathrm{H}_{\mathrm{complete}}$ are calculated as:

$$\mathrm{H}_{\mathrm{chain}} = (m-1)/1+(m-2)/2+(m-3)/3+\ldots+1/(m-1) \quad (2.6)$$

$$\mathrm{H}_{\mathrm{complete}} = m*(m-1)/2 \quad (2.7)$$

where m is the number of nodes. The possible value of H ranges from $\mathrm{H}_{\mathrm{chain}}$ to $\mathrm{H}_{\mathrm{complete}}$. $\overline{\mathrm{H}}$ is the difference between the actual H and the minimum possible H (i.e., $\mathrm{H}_{\mathrm{chain}}$) standardized by the possible range of H (i.e., the difference between $\mathrm{H}_{\mathrm{complete}}$ and $\mathrm{H}_{\mathrm{chain}}$). Therefore, $\overline{\mathrm{H}}$ is bounded within [0, 1]. The exception is that $\overline{\mathrm{H}}$ is not applicable when *m* is 2 because there is only one configuration for a graph having two nodes. Here, we calculated $\overline{\mathrm{H}}$for each compartment and the average $\overline{\mathrm{H}}$ of all the compartments, which was used to evaluate overall traversability in the partitioned graph. Higher values of $\overline{\mathrm{H}}$ suggest that a given definition of compartments better facilitates organism movement.

Finally, the Alpha Index (also known as Meshedness or Network Circuitry) is the ratio of the actual number of loops and the number of loops in the corresponding maximal planar network. Loops provide alternative routes for organisms to avoid disturbance and predation (Forman 1995; Rayfield et al. 2011). A higher Alpha Index suggests a greater opportunity for organisms to take various pathways to minimize the impact of disturbance and predation. Following Forman (1995), the Alpha Index is calculated as:

$$\alpha = \frac{\text{number of circuits}}{\text{maximum possible number of circuits}} = \frac{n - m + 1}{2m - 5} \qquad (2.8)$$

where m is the number of nodes and n is the number of links. The average Alpha Index of all the compartments was used to assess the overall robustness to disturbance and predation of compartments.

2.3 Results

Results of the two compartmentalization methods differed for 8 of the 10 compartments, with only Compartments 6 and 10 sharing the same set of patches in both classifications (Figure 2.4). The major differences between compartments identified by the two methods were in the central and western parts of the landscape graph. For example, Compartment 1 as defined by GN (Figure 2.4a) was partitioned into two compartments by GraphRECAP (Compartments 1 and 9: Figure 2.4b). Conversely, the nodes assigned to Compartments 2 and 7 by GN were all assigned to one compartment (Compartment 2) by GraphRECAP. Other discrepancies between the methods were marked by subtle differences in the locations of compartment boundaries.

In addition to differences in compartment membership, the sizes of compartments (i.e., the number of nodes in each compartment) partitioned by GraphRECAP were more consistent than those of compartments partitioned by GN. The smallest and largest compartments partitioned by GN contained 5 and 38 forest patches, respectively, while the size of compartments partitioned by GraphRECAP ranged from

12-34 patches (Table 2.1). Such differences are important not only from the standpoint of the amount of available habitat linked in each compartment, but also in compartment connectivity. For example, the assignment of nodes to Compartments 4 and 5 by GN resulted in two disproportionately-sized compartments, including a chain graph (Compartment 4) with the lowest possible values for both the Standardized Harary and Alpha Indices (Table 2.1), while the same nodes as partitioned by GraphRECAP resulted in more comparably-sized compartments (Compartments 3 and 8) with much greater average traversability and connectivity.

In terms of the overall structural measures of network and compartment characteristics, the modularity of the graph partitioned by GraphRECAP (0.813) was marginally higher than that partitioned by GN (0.804), indicating that nodes within compartments formed by GraphRECAP were more functionally-connected than those within compartments detected by GN. The mean values of $\overline{\mathrm{H}}$ and the Alpha Index (Table 2.1) for compartments partitioned by GraphRECAP were likewise higher, suggesting a greater degree of traversability and network circuitry within compartments than for compartments partitioned by GN.

2.4 Discussion

Graph- and network-based analyses of landscape connectivity have been advocated as valuable approaches for assessing and managing biodiversity in the face of habitat loss and fragmentation (Laita et al. 2010; Rubio and Saura 2012; Theobald et al. 2012). Previous research on habitat networks has tended to focus on either: 1) node-level patterns and processes, for example, identifying individual patches with specific characteristics (e.g., the greatest area or lowest isolation), or 2) network-level properties, for example, monitoring the number, arrangement, or connectivity of habitat patches in a landscape or identifying how network properties change with the loss of individual nodes or links. A focus on approaches at intermediate levels (components, compartments) that blend aspects of node- and network-level analyses can provide a useful perspective for habitat conservation and management by stressing the 'local' connections among nodes while still considering broader-scale, network-level linkages.

A landscape graph can be decomposed into compartments in numerous ways, but enumerating all the ways and choosing the best approach can be computationally expensive or infeasible (Fortunato 2010). The goal of this study was to partition an agricultural landscape with scattered dry-forest patches in southern Madagascar into habitat compartments using a graph regionalization technique, GraphRECAP, and contrast the resulting compartment properties with those identified by the more commonly used Girvan and Newman method. Ultimately, we believe that the identification of compartments and the key linkages that tie individual compartments to one another could contribute to management efforts that facilitate the persistence of structured populations; it is thus important that any noted differences in the results of different methods have potential ecological meaning.

Our results suggest that compartments generated by GraphRECAP have a number of characteristics that would be desirable from the standpoint of maximizing biodiversity and landscape connectivity. First, modularity values were higher for the

GraphRECAP partitioning, indicating that the landscape graph had a higher degree of compartmentalization than that partitioned by GN. GraphRECAP grouped the habitat patches in such a way that the patches had more interactions with other patches in the same compartments (i.e., denser connections within compartments), which would enhance the potential movements of and seed dispersal by ring-tailed lemurs while fostering greater resistance of compartments to disturbance. The strong connection among the neighboring habitat patches in a compartment is essential to the persistence of spatially-structured populations at the local scale because movement is more likely to take place in proximal habitat patches than those distributed farther away in other parts of the entire landscape graph. Although a high degree of compartmentalization may impede movement from the perspective of the entire landscape graph (Minor and Urban, 2007), local populations may benefit from higher within-compartment connectivity, especially when compartments contain enough habitat patches to support survival and reproduction. Our method stresses this definition of compartments. Meanwhile, a high degree of compartmentalization may reduce the potential effects of disturbance and disease (Minor and Urban, 2007).

Compartments created by GraphRECAP were also more consistent in size. With the GN partition, Compartment 4 contained only 5 patches, and two other compartments contained as few patches as the smallest compartment defined by GraphRECAP (Table 2.1; Figure 2.4). Though habitat patch quality and local population size were not considered, expectations from metapopulation theory would suggest that populations in compartments composed of a smaller number of patches could be more vulnerable to local extinctions due to lower effective population sizes and have a lower chance of being rescued by outside immigration (Hanski 1997). In contrast, the smallest compartment detected by GraphRECAP contained 12 habitat patches, which could both enhance the resistance of the compartment to habitat loss and facilitate patch recolonization of local losses from within-compartment sources.

Our results also suggest that compartments partitioned by GraphRECAP were better linked and more traversable. While there was compartment-to-compartment variability, the mean overall values for the standardized Harary Index ($\overline{\mathrm{H}}$), which is especially sensitive to changes in connectivity as metapopulations approach the minimal viable population size (MVP) (Jordán et al. 2003), and the Alpha Index, which is a measure of the degree of circuit presence and thus the number of options for organisms to traverse among habitat patches, were both higher for the GraphRECAP partitions (Table 2.1). The results were most extreme for Compartment 4 detected by GN, which was a chain graph with $\overline{\mathrm{H}}$=0 and no alternative movement routes for organisms in the event of disturbance.

The more favorable results for the GraphRECAP compartments stem directly from its method of deriving partitions. Although GN has been widely used, it has been criticized for producing unbalanced partitions under certain circumstances (Chen and Yuan, 2006). As it turned out in this study, Compartment 4 found by Girvan and Newman's method (Figure 2.4a) contained a small number of habitat patches. As discussed above, such a compartment is more vulnerable to local extinctions. GraphRECAP is a modularity optimization approach to compartment detection that tends to find compartments that have relatively even size in terms of number of links in the compartments (Fortunato 2010 and literature therein). It is thus somewhat less prone to unbalanced partitioning, which contributes to the greater connectivity and traversability within compartments found by GraphRECAP. Conversely, an advantage of GN in other ecological applications is its ability to identify links with high edge centrality, that is, those edges that are most central and thus most “between” compartments. Some of these edges (e.g. the edge that connects Compartment 5 and Compartment 7 in Figure 2.4a) are critical to the connectivity of the entire network. On the other hand, controlling the transmission of disease through these edges is an efficient way to prevent the spread of disease in the entire network.

Examples of studies targeted at intermediate network levels have become increasingly common and play an important role in studies of habitat networks (O'Brien et al. 2006; Vergara et al. 2013). Most often, researchers have focused on components, groups of linked habitat patches (nodes) that are isolated from patches in other components (e.g., Devi et al. 2013). Because habitat patches in one component are not linked to those in others, each component: 1) functions as a single sub-population, with individuals linked by dispersal within the component, but 2) exhibits population dynamics that are relatively independent from those in other components (Bodin and Norberg 2007; Vergara et al. 2013). By varying the threshold used to define interpatch linkages, it is possible to better understand component linkages and scale-dependent network properties (e.g., O’Brien et al. 2006). For example, McIntyre and Strauss (2013) calculated seven standard graph-theoretical metrics at multiple scales by varying window sizes and dispersal distances in their study of the habitat patch network of playa wetlands in the panhandle of Texas. Their approach revealed spatial patterns at the component level that could not be found either at the node- or network level and provided a useful means for examining habitat connectivity.

We chose to focus on compartments, rather than components, because we believe compartments better capture the varying degrees of connectivity that exist in most landscapes. Specifically, compartments recognize that all nodes in a landscape may be linked by dispersal, but that interactions (e.g., dispersal linkages) are more prevalent among certain patches than others. The process of partitioning a landscape graph into compartments is thus meant to identify clusters of habitat patches that are most closely linked, rather than identify which patches are or are not linked at a given threshold. In this respect, a compartment-based approach is compatible with principles implicit in island biogeography theory and metapopulation theory, which stress a certain degree of interactions among the patches and sub-populations in a given network or system. It is also consistent with recommendations for not only protecting ‘anchor areas’ of key remnant forests but also restoring smaller fragments in their neighborhood that could serve as stepping stones promoting connectivity (e.g., Holvorcem et al. 2011). In short, examining habitat connectivity from the component perspective answers the question “Are groups of habitat patches are connected?” while investigating habitat connectivity from the compartment perspective addresses the question “Which areas of the landscape are most highly connected?” (Galpern et al. 2011).

In this study, we recognized forest patches with areas > 1 ha as potential habitat patches and chose a dispersal threshold of 1000m to construct the habitat patch network based on previous studies of lemur vagility and the approach used by Bodin and Norberg (2007). However, Bodin et al. (2006) varied the thresholds determining habitat patches and dispersal linkages and analyzed changes in the values of several component metrics (e.g. the largest component, the area of habitat patches covered in the component) to investigate the habitat connectivity of species other than lemurs. Compartment detection methods can, in fact, be applied to habitat patch networks configured using any values of the minimum areas of habitat patches or dispersal thresholds to explore the continuous varying degrees of connectivity and capture the spatial patterns in between the node- and network-levels. A compartment-based approach thus provides a means of identifying groups of patches within which dispersal is most prevalent given specified assumptions about habitat arrangement and organism dispersal.

2.5 Conclusion

In this study, we used a novel regionalization method, Graph-based REgionalization with Clustering And Partitioning (GraphRECAP), to decompose a landscape graph to compartments. Compared to the compartments which were also detected by the more widely used Girvan and Newman method from the same landscape graph, the compartments found by our method had stronger within-compartment connections, greater traversability, more alternative routes, and a larger minimum size of habitat patches within compartments, all of which are more desirable traits for ecological networks. Our method thus offers an improved means for characterizing the spatial structure of populations in terms of improving habitat connectivity and increasing the persistence of populations.

Table 2.1 Compartment properties of the Madagascar landscape graph as partitioned by the Girvan and Newman method and GraphRECAP.

Compartment	Habitat Patches		Standardized Harary Index		Alpha Index	
	GN	GraphRECAP	GN	GraphRECAP	GN	GraphRECAP
1	38	13	0.356	0.431	0.359	0.306
2	10	25	0.213	0.472	0.217	0.380
3	12	12	0.734	0.703	0.400	0.398
4	5	12	0.000	0.203	0.000	0.226
5	27	20	0.422	0.543	0.372	0.386
6	15	15	0.601	0.601	0.390	0.390
7	24	34	0.276	0.220	0.280	0.256
8	26	23	0.263	0.324	0.326	0.341
9	13	16	0.151	0.115	0.120	0.097
10	13	13	0.537	0.537	0.362	0.362
Mean	-	-	0.355	0.415	0.283	0.314

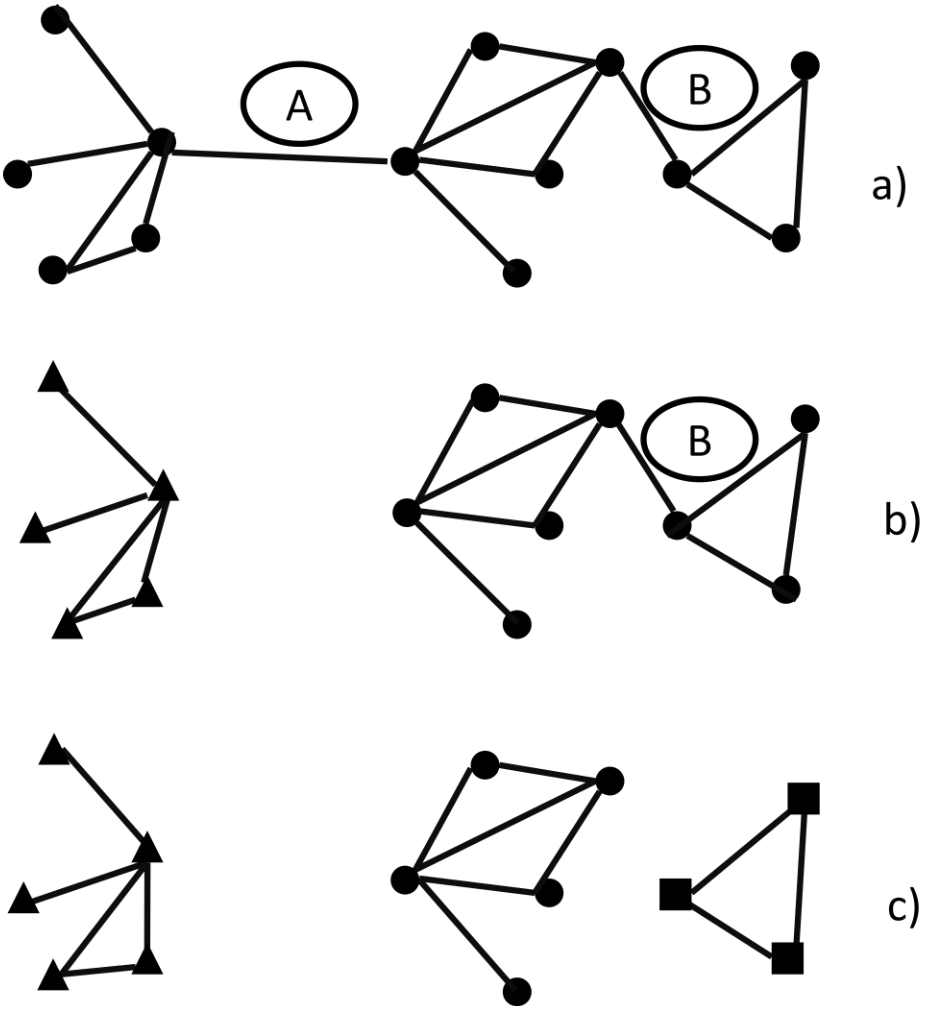
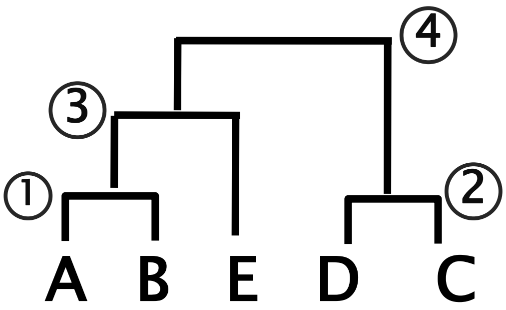
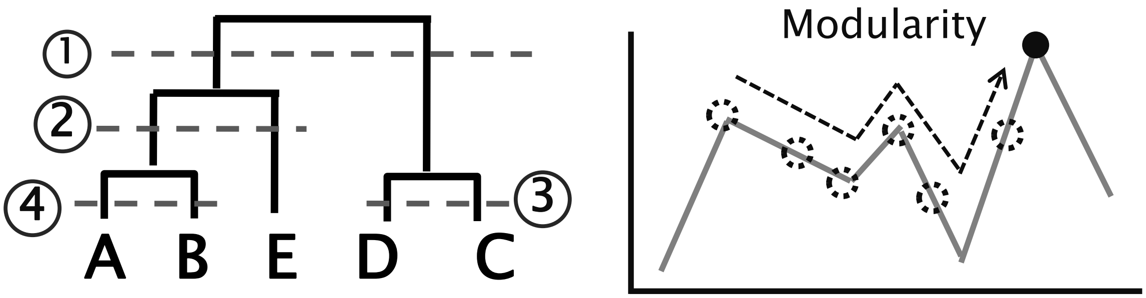


Figure 2.1 An illustration of the Girvan and Newman method. This method defines compartments by iteratively removing edges with high edge betweenness. In this case, edges A and B in the full landscape graph (a) are removed to obtain the compartments marked by different shapes in (b) and (c)



Node Pair	Modularity
AB	8
AC	-9
AD	-5
AE	5
BC	-4
BD	-1
BE	4
CD	6
CE	-7
DE	3

a)



b)

c)

Figure 2.2 An illustration of the GraphRECAP (Graph-based REgionalization with Clustering And Partitioning) method, which includes three steps: a) clustering nodes based on the modularity between pairs of nodes, b) partitioning the dendrogram, and c) optimizing within-compartment modularity after each partitioning. Numbers in circles indicate the sequence of clustering and partitioning.

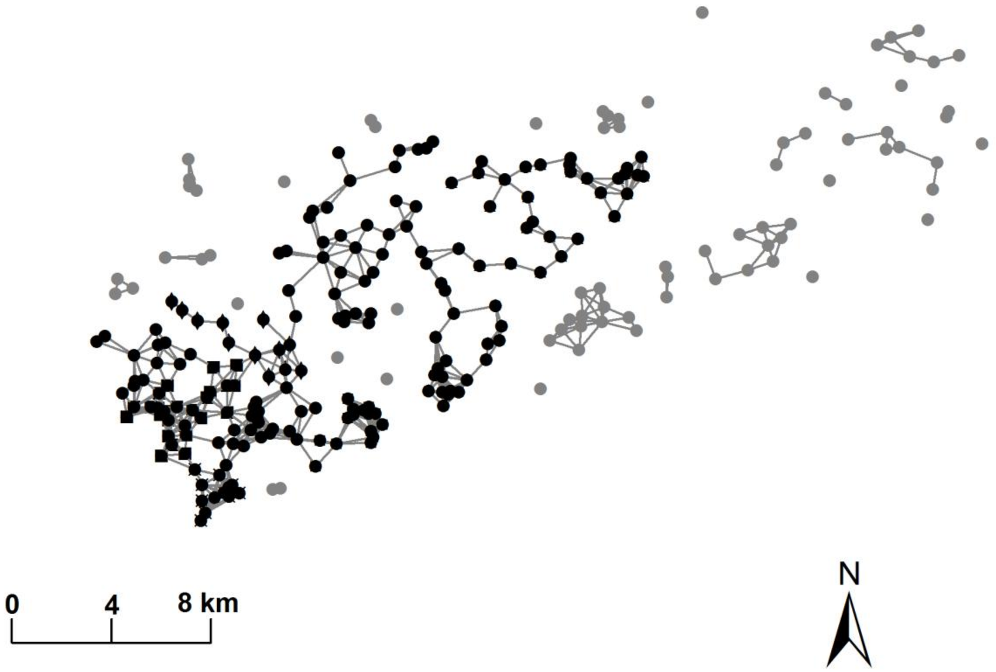
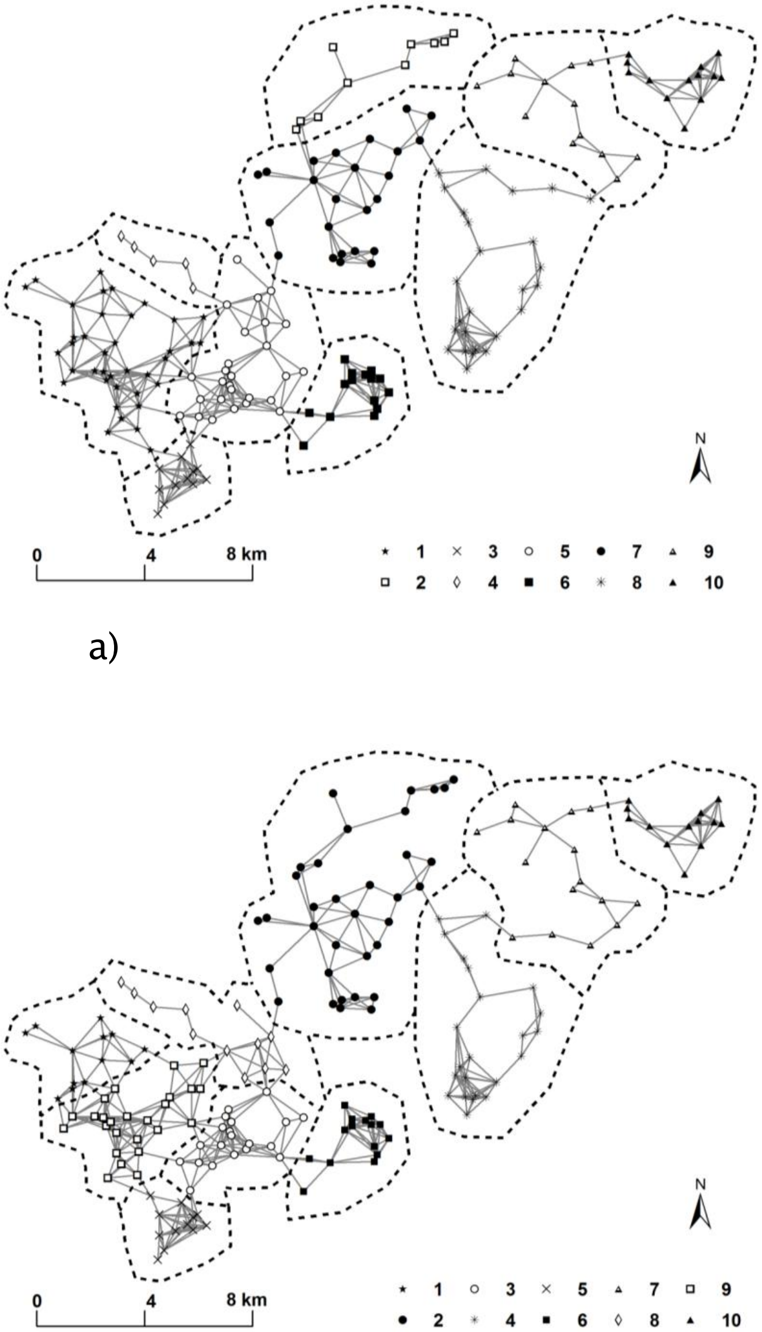


Figure 2.3 Network representation of ring-tailed lemur habitat patches in the study landscape. Only patches in the largest component, represented by black dots, were analyzed in this study



a)

b)

Figure 2.4 Decomposition of the largest component into ten compartments using a) the Girvan and Newman method and b) GraphRECAP

Chapter 3 Detection of Regions in Spatial Graphs: a New Approach to Animal Trajectory Analysis[2]

Abstract

The increasing availability of telemetry data with high spatial and temporal resolution promises to greatly advance scientific understandings of how spatial and temporal factors impact the movements of individual organisms and thereby affect species persistence in heterogeneous landscapes. The amount of data provided by such methods, however, can be challenging to analyze and interpret. In this study, we used a trajectory analysis approach based on Hierarchical Region Discovery (HRD) to investigate the movement of cattle (*Bos taurus*), mule deer (*Odocoileus hemionus*), and elk (*Cervus elaphus*) tracked by an Automated Telemetry at Starkey National Forest, in northeastern Oregon, USA in June 1995. Trajectories of the animals were partitioned into regions such that organisms had more movement within than across regions. Attributes of the trajectories were extracted based on the regions and were further used to cluster and classify these trajectories. Specifically, we evaluated two criteria that govern the process of finding regions (modularity and edge ratio) by comparing the quality of clusters and the accuracy and simplicity of classification using the attributes derived from different regions found by the two methods. While modularity has been

---

[2] This manuscript will be submit to *International Journal of Geographical Information Science*

widely used, we found that regions found by edge ratio more effectively captured the characteristics of the animal movement. We also discuss the differences in the designs of modularity and edge ratio and explore the reasons why regions defined by the edge ratio were more suitable for this particular ecological application. While it should not be viewed as a replacement for other methods of animal trajectory analysis, Edge ratio-based Hierarchical Region Discovery provides an alternative approach to capturing the characteristics of different species movement and to exploring hidden patterns.

3.1 Introduction

Animal movement is a fundamental process that determines the fate of individual organisms, the structure and dynamics of populations, and the nature of species interactions and community assembly (Nathan et al. 2008; Miller 2012). Detailed observation of the movement of individual animals coupled with the development and application of movement models serve as the basis for understanding spatial population processes and provide insights into spatial dynamics at higher levels of ecological and spatial organization such as patches, communities, and metapopulations (Bennett and Tang 2006; Schick et al. 2008; Eros et al. 2012; Rathore et al. 2012; Holdo and Roach 2013). Such knowledge is crucial to addressing some of the most pressing questions in conservation biology and biogeography today, for example, the potential impacts of habitat loss and climate change on species survival and persistence (Schick et al.).

Among the various mathematical models developed to analyze and understand animal movement paths, Correlated Random Walk (CRW) models have provided perhaps the strongest basis for the development of advanced movement models

(Codling et al. 2008; Miller 2012). CRW models predict individual movements by randomly selecting movement lengths and turn angles from empirical distributions that consider factors influencing the movement, such as the animal response to different habitats (Bailey and Thompson 2006). Better approximations of movement can be achieved by incorporating animal behavior into the movement models (Morales and Ellner 2002; Morales et al. 2004; Jonsen et al. 2005; Schick et al. 2008). For example, more complex movement can be modeled by letting behavioral modes govern the parameters in CRW models (Jonsen et al. 2005). By linking models that focus on fine-scale individual movement processes to broader-scale population processes, it becomes possible to integrate behavior, biogeography and population dynamics into mechanistic models that connect decision-making at the individual level with movement, and, ultimately with distribution and population structure (Patterson et al. 2008: 93).

Complimenting model-based approaches to quantifying and understanding organism movements, advances in wildlife telemetry over the last two decades have greatly increased the amount and quality of available data on animals' use of space (Aarts et al. 2008). A range of toolkits and algorithms have been developed specifically to quantify the spatial patterns of animal movement (e.g. Calenge et al. 2009; Tang et al. 2011), and methods developed to analyze the movement of a wider range of objects (e.g., vehicles) in fields such as computer science, GIS and geovisualization also provide valuable insights into animal movement patterns and processes. For example, Lee et al. (2008) used a region- and trajectory-based method to capture and differentiate the movement characteristics of different types of objects, including hurricanes, ships and large mammals. The means for incorporating temporal dynamics more effectively into static spatial data, one of the main tasks of temporal GIS and a fertile ground for further research in animal telemetry research, remains at the forefront of GIS research (Long and Nelson 2013).

While technological advances have made it relatively easy to quantify the basic geometric or quantitative properties of individual animal movements from radio-telemetry data in the form of movement metrics (e.g., speed, heading, turning angles between subsequent locations and rates of movement between regions: Patterson et al. 2008), the amount of information generated by tracking numerous individuals poses greater challenges to the quantification and interpretation of collective movement data. The most common way of integrating spatial and temporal information is to project individual movements in a three dimensional space-time cube where two axes represent geographic space and the third axis stands for time (Andrienko and Andrienko 2006). Interpretability can be further enhanced by color-coding trajectories according to their properties (e.g. the object types) or by using filter, query, and animation functions to identify, retrieve or map trajectories with specified attributes (e.g. Kwan 2000). However, the efficiency of such approaches declines with increasing spans of movement time or an increasing number of trajectories due to clutter and occlusion (Andrienko and Andrienko 2013).

Aggregation methods have been proposed as a means for reducing the complexity inherent in large data sets, improving the efficiency of trajectory visualization, and facilitating pattern recognition. Clustering is a widely used technique for aggregation that can be adapted for use in analyzing and comparing animal movement trajectories. In much the same way that non-spatial objects can be clustered on the basis of similarity measures, indices describing the path similarity of trajectories can be represented by measures such as the Hausdoff distance (Huttenlocher et al. 1993), the longest common subsequence (LCSS) (e.g. Cheriyadat and Radke 2008) and dynamic time wrapping (DTW) (e.g. Usabiaga et al. 2007). While popular, these measurements have documented weaknesses as well: the Hausdoff distance cannot consider chronological order of the points (Zhang 2006), the LCSS is sensitive to the threshold selected to determine whether two elements match, and the DTW is not robust to noise (Chen 2005).

Alternatively, telemetry locations can first be grouped into subsets such as spatial clusters (e.g. Andrienko and Andrienko (2011)) or regions (e.g. Guo et al. (2012)). By aggregating movements (i.e., flows) between locations, users are able to obtain an overall view of the spatial and temporal distribution of multiple movements and to uncover potential patterns (Andrienko and Andrienko 2013). For instance, Fosca Giannotti (2007) decomposed vehicle trajectories into the regions of interest visited during movements. Trajectories were then described as regions, and the time used to travel from region to region was analyzed from the view of the spatial - temporal sequence. The regions of interest in their study were detected based on *a priori* knowledge or, when no such knowledge was available, point density. This kind of approach, which is less influenced by the geometry of movement paths, can detect hidden patterns in the movement data and might be particularly suitable for animal movements because organisms, unlike vehicles, usually exhibit free movement. Verhein and Chawla (2008) established multiple spatiotemporal association rules to detect *stationary* and *high traffic* regions and described how mobile objects move between regions over time. When they applied their method to the movement of caribou in northern Canada, group and individual movements were distinguished by different regions where the movement occurred.

In this study, we first use a trajectory analysis approach based on Hierarchical Region Discovery (HRD), which detects regions of interest, and then investigate the movement patterns at the level of regions. The delineation of regions, however, is dictated by the criterion used to cluster movement trajectories. Here we introduce the use of ***edge ratio*** as a means for detecting regions, rather than ***modularity*** (Newman and Girvan 2004), as used in Guo et al. (in prep). To make a comparison, we analyze the same data set examined by Guo et al. (in prep) which includes the movements of cattle (*Bos taurus*), mule deer (*Odocoileus hemionus*), and elk (*Cervus elaphus*) in Starkey National Forest, in northeastern Oregon, USA. We explore the suitability of ***edge ratio*** in this particular ecological application for the designs of the two criteria.

3.2 Study Area

The study area is the Starkey Experimental Forest and Range, which is located in northeastern Oregon. The data set, which is described in more detail by (Rowland et al. 1998), contains 14,990 x-y coordinates for 34 cattle, 30 mule deer, and 38 elk tracked by an Automated Telemetry System based on LORAN-C navigation technology in June 1995. Temporal resolution of the data is 45 - 90 minutes, and the spatial error is ca. 200

m. Cattle movement was restricted by barbed-wire fences (Coe et al. 2001). This research builds on that from two previous studies (Lee et al. 2008; Guo et al. in prep), which analyzed the same dataset but using different types of trajectory analysis.

3.3 Methods

The method that we apply in this study is Hierarchical Region Discovery-based trajectory analysis (hereafter HRD), which is designed to detect regions from a huge number of telemetry locations based on the criterion that organisms exhibit greater movement within regions and less movement across regions. HRD has three steps (Guo et al. in prep). First, it groups points from all trajectories into small spatial clusters. Second, it constructs a weighted graph where nodes represent the spatial clusters and edges are the connections among spatial clusters based on trajectories that pass through the clusters. Third, it uses a contiguity-constrained graph partitioning method to discover regions. Additional information on HRD can be found in (Guo et al. 2012) and (Guo et al. in prep). The rest of the Methods section is divided into three parts which describe HRD, the difference between modularity and edge ratio, and how to evaluate the performance of modularity-based vs. edge ratio-based analyses.

3.3.1 Hierarchical Region Discovery: Building the Network of Spatial Clusters

Radio telemetry data take the form of time-indexed spatial locations of individual animals. Animal movement can be represented by a set of trajectories $T = \{T_i\}$ ($1<i<n$) of n individuals, wherein each T_i comprises m points $P_{ij} = \{<s_{ij}, t_{ij}$ ($1<j<m$) and s_{ij} and t_{ij} represent the spatial coordinates and time of an individual telemetry location. Determining the pairwise similarity between spatial points is prerequisite for any

clustering method that groups points into spatial clusters. The similarity of any pair of points (A and B) in this study is based on their Shared Nearest Neighbors (SNN). Specifically, let $NN(A, k)$ and $NN(B, k)$ be the k Nearest Neighbors (NN) of A and B in space, respectively, while SNN is the number of points shared by both NN (A, k) and NN (B, k). The similarity of A and B in k nearest neighbors is the ratio between the intersection of $NN(A, k)$ and $NN(B, k)$ (i.e., $SNN(A, B, k)$) and the union of $NN(A, k)$ and $NN(B, k)$:

$$\text{Similarity}(A, B) = (NN(A, k) \cap NN(B, k)) / (NN(A, k) \cup NN(B, k)) \quad (3.1)$$

Before clustering, a Delaunay triangulation (DT) is built for all points (Guo et al. 2003). A DT for a set P of points in the Euclidean plane is a triangulation DT (P) such that no point in P is inside the circumcircle of any triangle in DT (P) (Tsai 1993). The construction of a DT efficiently reduces the time complexity of finding the k nearest neighbor points of each point and grouping points into spatial clusters. To find the k nearest neighbors of each point, the algorithm first searches through the points that are directly connected to the focusing point by the edges in the DT (i.e., points connected to the focusing point in the first order). If the number of points that are found is less than k, then the algorithm examines second or higher order connections (points that are connected via their connections to the focusing point) until the nearest k points are found.

The clustering process is a bottom-up procedure in which single linkage clustering is applied only to pairs of points connected by edges in the triangulation. This limitation significantly quickens the clustering process because the total number of edges is linearly proportional to the number of points. Beginning with each point as a single cluster, clusters are merged according to their descending order of similarity. The clustering process continues until the number of points contained in each cluster reaches a user-defined minimum number of points. The clustering procedure is thus governed by two parameters, the number of nearest neighbors (*k*) and the minimum size of clusters (*q*). When *k* is held constant and *q* varies, the structure of the dendrogram does not change. The increase or decrease of *q* yields clusters in higher or lower hierarchy. When *q* is held constant, changing *k* impacts the smoothing effect. A larger value of k produces a stronger smoothing effect because the similarity of two points is considered in a larger neighborhood. Sensitivity tests have indicated that patterns change slightly with different settings of *k* and *q* (Guo et al. 2012). The KNN based clustering has two merits (Guo et al. 2012): 1) unlike other methods, KNN is generally not biased towards producing clusters of a particular shape (e.g., k-means clusters are likely to be circles), and 2) it is adaptive to the uneven distribution of points over space and able to find more clusters in areas of high point density and fewer clusters where point density is low.

The next step involves building connections among the clusters. After clustering (Figure 3.1 a)), each Point (*P*) corresponds to or is represented by a Spatial Cluster (*SC*) to which it belongs. Two spatial clusters are connected when a trajectory passes through points assigned to them. The number of the connections between two *SC*s is determined by the frequency that trajectories pass through points assigned to them. The resulting weighted graph is thus a spatial graph (i.e., one that contains spatial information) in which nodes represent the spatial clusters and edges indicate the number of connections among them.

Figure 3.1 illustrates how the algorithm examines all pairs of points traversed by trajectories and counts the connections between spatial clusters that represent the pairs of points. The trajectory in brown passes through SC_1, SC_2, SC_3, and SC_4 (Figure 3.1a). Therefore, connections are counted between pairs of SC_1 and SC_2, SC_1 and SC_3, SC_1 and SC_4, SC_2 and SC_3, SC_2 and SC_4, and SC_3 and SC_4. In the same way, connections built from the trajectory in purple passing through SC_1, SC_2, SC_4, SC_5 and SC_6 are pairs of SC_1 and SC_2, SC_1 and SC_4, SC_1 and SC_5, SC_1 and SC_6, SC_2 and SC_4, SC_2 and SC_5, SC_2 and SC_6 , SC_4 and SC_5, SC_4 and SC_6, and SC_5 and SC_6. A weighted graph is built by enumerating connections through the two trajectories, with the thickness of edges indicating the number of connections (Figure 3.1b). In this example, SC_1 and SC_2 have two connections between them because both trajectories pass through SC_1 and $SC_{2.}$

3.3.2 Hierarchical Region Discovery: Regionalization of Spatial Clusters

Once the weighted graph is built from analyses of the trajectories, spatial clusters are aggregated to level of regions or 'compartments'. Compartments are groups of nodes in networks or graphs with many edges joining nodes of the same group but comparatively few edges joining nodes of different groups (Fortunato 2010). Compartment detection serves as a means of reducing the complexity of networks and facilitating the search for patterns in an otherwise complex set of relationships (Bodin et al. 2007).

While there are many approaches and methodologies for compartment detection, we adopt a technique developed specifically for graphs with spatial information, GraphRECAP (Graph-based REgionalization with Clustering and Partitioning: Guo 2009). In contrast to other methods, GraphRECAP ensures that the nodes (i.e., spatial clusters) in detected compartments are spatially contiguous. Therefore, we use the term *regions* for the detected groups of spatial clusters. GraphRECAP involves contiguity constrained hierarchical clustering and spatially contiguous tree partitioning. Different criteria can be used to govern the processes of clustering and partitioning, which may results in different regions. Modularity and edge ratio, the two criteria used in this study, are introduced below.

Modularity and edge ratio quantify the connections among nodes within compartments (i.e., regions in this study) in different ways. Modularity measures "the fraction of the edges in the network that connect nodes of the same type (i.e., within-compartment edges) minus the expected value of the same quantity in a network with the same compartment divisions but random connections between the nodes” (Newman and Girvan 2004: 7). Given a Graph (*G*) that is partitioned to two compartments (*A* and *B*) which contain *n* and *m* nodes respectively, the modularity of *G* is defined as:

Modularity (*G*) = $\sum_{\mathrm{i}}^{\mathrm{n}} \sum_{\mathrm{j}}^{\mathrm{m}} \mathrm{Modularity}\ (node_i\ , node_j)\quad (\mathrm{i} \neq \mathrm{j},\ node_i \in A, node_j \in B)$ (3.2)

The modularity between nodes *i* and *j* is defined as the difference between the actual connections and expected connections between *i* and *j*:

Modularity $(i\ , j)$ = Actual Connections $(i\ , j)$ – Expected Connections $(i\ , j)$ (3.3)

The expected number of connections between node i and j is calculated using the total connections associated with the two nodes C_i and C_j (i.e., edges associated with i and j in the graph) and the total connections (C) in the graph (i.e., all the edges in the graph):

$$\text{Expected Connections } (i\ , j) = (C_i * C_j) / C \quad (3.4)$$

High modularity indicates that the connections among nodes within compartments are greater than those expected by random chance.

Edge ratio determines the strength of within-compartment connections by comparing them to connections between compartments. Specifically, it is the ratio of within-compartment connections to the between-compartment connections. For compartments A and B, edge ratio chooses the fewer connections within the two compartments as the within-compartment connections and compares them with the connections between A and B (Connections (A, B)):

$$\text{Edge ratio } (A, B) = \min\,(\text{Connections } (A), \text{Connections } (B)) / \text{Connections } (A, B) \quad (3.5)$$

The graph is partitioned by an iterative process that optimizes the selected criteria (i.e., modularity or edge ratio), which may result in different ways of partitioning a graph. The graph in figure 3.2a is decomposed to two compartments differently in figure 3.2b by maximizing modularity (Modularity = 10.92; Edge ratio= 5.5) and in figure 3.2c by maximizing edge ratio (Modularity = 8.75; Edge ratio = 6).

The clustering process in HRD is comparable to that with a standard hierarchical clustering method in that it involves iteratively merging nodes from the bottom up that have the highest similarity. The clustering process in HRD merges nodes that have the highest modularity, but it requires clusters to be merged at each hierarchical level to be spatially contiguous. A dendrogram is constructed after clustering. The difference between Guo et al. (in prep) and our method lies in the criterion that governs the partitioning process that cuts the dendrogram into regions. The former uses modularity (hereafter, MHRD: Modularity-based Hierarchical Region Discovery), while the later adopts edge ratio (hereafter, EHRD: Edge ratio-based Hierarchical Region Discovery). In the partitioning process, MHRD examines all the edges of the dendrogram and cuts the one that maximizes the modularity when the tree is cut at that edge. In contrast, EHRD examines all the edges of the dendrogram and cuts the one that maximizes the edge ratio when the tree is cut at that edge. The partitioning process continues until the desired number of subtrees (i.e., regions) is reached. After each partition, a Tabu-based optimization method (Guo and Jin 2011) is used to further improve the partition by fine-tuning the assignment of the nodes to the subtrees and avoid the trap of local optima.

3.3.3 Evaluation of modularity-based vs. edge ratio-based Hierarchical Region Discovery

Through the three steps of HRD described above, the study area was hierarchically partitioned to regions by EHRD and MHRD on the basis of animal movements. These regions (i.e., compartments) provided a means for exploring factors that may structure individual and collective animal movements based on the region(s) where animals stay and traverse.

To compare the performance of EHRD and MHRD, we extracted the attributes of trajectories based on the regions detected by MHRD and EHRD. We then used these attributes to cluster the 102 trajectories and classify them as cattle, deer, or elk, assuming that we don’t know what species the trajectories represent. The hypothesis is that a good partitioning of regions should be able to capture the characteristics of species movement and thus serve as an effective predictor of animal movement. More specifically, a good partitioning of regions may be able to find regions that are dominated by one species, or regions where a species seldom traversed. When attributes extracted from these regions are used to cluster and classify the trajectories, we should be able to achieve a high quality of clusters and achieve a high accuracy of classification. The procedures used to conduct these test are the same as those used in Guo et al. (in prep), to allow for exact comparisons.

Because both MHRD and EHRD produced hierarchical regions, we examined results for regionalizations of varying detail, ranging from two to ten regions. Thus, attributes associated with trajectories were extracted at each level of regions ranging from two to ten. However, MHRD and EHRD identified different regions due to different criteria of partitioning as described above, which resulted in different attributes of trajectories based on the detected regions. Therefore, the comparison focused on which method provides more informative attributes that improve the performance of clustering and classification.

Following Guo et al. (in prep), we simply extracted the percentage of telemetry locations of each individual within each detected regions as the attributes associated with each trajectory. In this way, each trajectory had the same number of variables as the level of hierarchy. For example, at the two region level, each trajectory had two variables (i.e., the percentage of telemetry locations of the trajectory in each of the two regions), three variables at the three region level, and so on. The attributes of the trajectories were then used to calculate the distance between pairwise trajectories for cluster analysis and to serve as input for a decision tree analysis in See5 (Rulequest Research, 2011) to classify trajectories of individuals for different species.

The distance between two trajectories is a modified Euclidian distance. For a study area partitioned into *n* regions, let *P(i,k)* and *P(j, k)* be the percent of telemetry locations from trajectories *i* and *j* in *k*th region (*1=<k=<n*). The distance (*d*) between trajectories *i* and *j* is defined as:

$$d = \sqrt{\sum_{\mathrm{k}}^{\mathrm{n}}\left(P(i,k) - P(j,k)\right)^2 \left(\left(P(i,k) + P(j,k)\right)\right)} \quad (3.6)$$

where the left term under the radical sign is the Euclidian distance, and the right term is a weight that emphasizes the presence of two trajectories in one region while deemphasizing the regions where neither of them appear. For example, at the five region level, if a trajectory is only within the first region and another trajectory stays only in the second region, they are very different in terms of the presence in different regions. The modified Euclidian distance can capture this difference, while the traditional Euclidian distance cannot because they are so similar in the other three regions (i.e., both are absent).

We applied Average Linkage clustering (ALK) to cluster trajectories. EHRD and MHRD produced different regions and thus resulted in different distance matrices and clusters. We used the Silhouette index (Rousseeuw 1987) to evaluate the quality of clusters produced by EHRD and MHRD. In a comparative study of thirty cluster validity indices, Arbelaitz et al. (2013) examined performance of these indices using synthetic and real datasets under different conditions of tests such as different clustering methods, various cluster densities and multiple levels of noise. This index has been recognized as the most suitable cluster validity index overall. The Silhouette index (S) is calculated as:

$$S = \frac{1}{n}\sum_{i}^{n}\frac{(b_i - a_i)}{\max(a_i, b_i)} \quad (3.7)$$

where *n* is the number of trajectories, a_i is the average distance between trajectory *i* and all other trajectories in its own cluster and b_i is the minimum of the average dissimilarities between *i* and trajectories in other clusters. a_i measures the overall dissimilarity of trajectory *i* to other trajectories in the same cluster (i.e., the cohesion of the clusters, with smaller values meaning higher cohesion), and b_i measures the average distance of trajectory *i* to trajectories in the cluster that is most similar or closest to it (i.e., the separation of the clusters, with larger values meaning better separation). When a_i is greater than b_i, the Silhouette index for trajectory *i* is negative, suggesting that trajectory *i* is more similar to trajectories in other clusters. When b_i is greater than a_i, it means that the average distance of trajectory *i* to trajectories in the 'nearest' cluster is larger than that of trajectory *i* to other trajectories in the same cluster. In this case, the Silhouette index measures the difference between the two distances, scaled to the former. Therefore, larger values suggest a higher quality.

Since the regions created by MHRD and EHRD are hierarchical, we extracted attributes at levels of three to ten regions. Clustering was conducted at each level of regions using attributes extracted from the regions at that particular level. For example, trajectories were clustered using three variables (i.e., the percent of telemetry locations of the trajectory in Regions 1-3) at the level of three regions. The Silhouette index also depends on the number of clusters. To avoid any bias in the comparison of the Silhouette index at a particular number of clusters or a particular level of regions (e.g. MHRD may achieve a higher Silhouette index for 5 clusters at the level of 3 regions, but EHRD may outperform for 4 clusters at the level of 6 regions), we calculated the Silhouette indices ranging from 2 to 20 clusters at levels of three to ten regions for a comprehensive comparison.

The classification of 102 trajectories into cattle, mule deer, and elk was conducted using the decision tree method as implemented in See5 (Rulequest Research, 2011). Following Guo et al. (in prep), we used the percentage of telemetry locations of trajectories in each region from the level of two to ten regions as variables for a decision tree analysis.

If EHRD and MHRD produced different regions resulting in different sets of variables, we posited that one useful measure of their quality would be which provides more informative variables. A decision tree that uses more informative variables should be able to achieve a better performance (i.e., accuracy and simplicity as *described below*). The performance of decision trees depends on training samples (Safavian and Landgrebe 1991). To avoid bias that may be caused by the selection of the training samples, we used *K- fold cross-validation* to test the performance of the decision tree. The entire dataset is divided into *K* roughly equal parts, $L^1, L^2, \ldots L^k$. The decision tree is then conducted *K* times, and each time the decision tree used the dataset with L^i ($1<i<K$) excluded as the training dataset and L^i as the validation dataset. K was set to ten in this study. The performance of classification was evaluated by the accuracy and simplicity.

The accuracy is measured by the average accuracy of ten-fold test. The simplicity is determined by the average number of leaves of ten-fold test (Osei-Bryson 2004).The method that achieves a higher accuracy with less number of leaves (i.e., higher simplicity) is considered as a better one.

In addition to comparing the performance of clustering and classification, we also explored the ecological implications or meanings associated with the regions delineated by MHRD and EHRD. Previous studies have suggested that cattle, mule deer, and elk compete for food and space. For example, in a study of the spatial distribution and use of different habitats by elk and mule deer, Coe et al. (2001) found that the number of locations of elk in pasture and ponderosa pine / Douglas fir forests decreased with the presence of cattle. They also found that as the increasing use of ponderosa pine / Douglas fir by elk, the use of ponderosa pine / Douglas fir by mule deer decreased. To explore whether such patterns were reflected in the detected regions by MHRD and EMRD, we compared the point density of the three species in each region, hypothesizing that point density can suggest or confirm these inter-species relationship as observed in the previous studies. For example, a higher point density of elk in one region should result in lower point density of deer due to competition.

3.4 Results

Using the settings employed by Guo et al. (in prep) (k = 50 and q = 30), we aggregated the 14,990 telemetry points into 203 spatial clusters (Figure 3.3). Partitions of these clusters into a small number of regions (2-3) by GraphRECAP were very similar whether the regionalization was based on modularity (Figure 3.4a-b) or edge ratio

(Figure 3.4e-f). At the four region level, MHRD resulted in a division of the region that contained most locations of cattle (Region 1 in Figure 3.4b) into two regions (Regions 1 and 2 in Figure 3.4c; see also Figure 3.5a-b). In contrast, EHRD retained this cattle-dominated area in the northern portion of the study area, instead splitting the eastern region dominated by deer and elk (Region 3 in Figure 3.4f) into two regions (Regions 3 and 4 in Figure 3.4g; see also Figure 3.5d-e). Additional regions captured discernible gaps in trajectories. Increasing the number of regions from four to five, EHRD detected the tight boundary that enclosed most locations of cattle (region 1 in Figure 3.4h, see also Figure 3.6d). MHRD divided the large region in southwest into two regions (Figure 3.4c-d), one of which (i.e., Region 4 in Figure 3.4b) had comparably less amount of mule deer movement.

The different regional boundaries (Figures 3.4-6) generated by MHRD and EHRD resulted in different attributes (i.e., the percent of telemetry locations of the trajectories in each region) associated with the trajectories, and thereby different clusters of trajectories. In general, Silhouette index values decreased as the number of regions increased for both MHRD and EHRD (Figure 3.7). The Silhouette index for clusters produced by MHRD had the greatest decrease from the level of three regions to four regions, the level at which the cattle-dominated region was subdivided. The largest decrease of Silhouette index values for EHRD classifications occurred when the number of regions increased from five to six, which divided the region that contained most of locations of cattle (Region 1 in Figure 3.6d) into two regions (not shown). Therefore, detecting the region that was dominated by cattle and keeping it from being further divided into subregions is important to maintaining the quality of clusters.

MHRD achieved the highest Silhouette index value for three clusters at the level of three regions (Figure 3.7). Three clusters created by MHRD at the level of three regions were similar to those generated by EHRD (Figure 3.8), probably because that the three regions detected by the two methods were similar (Figure 3.4b and 3.4f). However, a greater difference was observed for five clusters at the level of five regions (Figure 3.9). EHRD produced two almost pure clusters of mule deer trajectories (Figure 3.9h-i). The regions (Regions 2 and 4 in Figure 3.4h, see also Figure 3.6e) where most of the two clusters of trajectories occupied had the highest point density of mule deer (Table 3.1). Trajectories of cattle were in one cluster produced by EHRD (Figure 3.9j). However, they were separated into two clusters generated by MHRD (Figure 3.9d-e).

This different partition of regions by MHRD and EHRD also resulted in different point densities of cattle, mule deer and elk in each region. EHRD was better able to identify the most preferred regions of cattle, elk and mule deer suggested by the point density at the level of five regions (Table 3.1). Region 1 (the most preferred region of cattle with the highest point density of 235.4 among five regions) had a higher point density of cattle than Region 1 or 2 detected by MHRD (Figure 3.6a and 3.6d). Region 5 (the most preferred region of elk with the highest point density of 135.6 among five regions) had a higher point density than Region 3 detected by MHRD (point density: 134.8) (Figure 3.6c and 3.6f). Region 2 (the most preferred region of mule deer with the highest point density of 131.9 among five regions) had a higher point density than

Region 3 detected by MHRD (point density: 88.18) (Figure 3.6b and 3.6e). Region 2 is also regarded as the least preferred region of elk with lowest point density of 19.85 which is lower than that of Region 2 (point density: 43.0) detected by MHRD (Figure 3.6c and 3.6f).

Point density of elk and mule deer were negatively correlated at the level of four regions detected by both MHRD and EHRD, but the EHRD had a higher R square (Figure 3.10a-b). The relationship became weak at the level of five regions found by MHRD, but remained strong in the five regions generated by EHRD (Figure 10c-d and Table 3.1).

Attributes of trajectories extracted from regions detected by MHRD and EHRD also affected the performance of decision tree. In ten tests of the decision tree, the highest, average (mean), and lowest accuracy achieved by EHRD were all higher than those by MHRD (100.0%, 87.3%, and 70.0% respectively vs. 90.9%, 77.3%, and 40.0% respectively). Meanwhile, EHRD also had higher simplicity (i.e., using less number of leaves to achieve the accuracy). In ten tests of decision trees, the largest, average, and lowest number of leaves by EHRD were 9, 6, and 4 respectively, compared to 11, 8.5, and 6 by MHRD (Table 3.2).

3.5 Discussion

We investigated the movement patterns of cattle, mule deer, and elk at the Starkey Experimental Forest and Range by finding regions such that organisms exhibit greater movement within regions and less movement across regions. We compared the modularity and edge ratio in term of their abilities to find more informative and ecological meaningful regions to effective capture the characteristics of animal movement. We first address the differences between regions found by MHRD and EHRD and the influence of the differences on capturing the movement patterns. We then explore the reasons that caused the differences which underlie the designs of the modularity and edge ratio.

3.5.1 Regions found by Modularity-based and Edge ratio-based Hierarchical Region Discovery

The most distinct difference between the MHRD and EHRD is that EHRD detected the tight boundary of cattle movement (Figure 3.6d), while MHRD divided the region that contained most of locations of cattle (Region 1 in Figure 3.4b) into two regions (Regions 1 and 2 in Figure 3.5a) at the level of four regions. Note that even though region 1 identified by MHRD in Figure 3.4b contained most of locations of cattle, it was not a tight boundary of cattle movement, that is, it covered area where cattle did not traverse (Figure 3.5a). . EHRD was also more effective than MHRD in detecting the effects of barbed-wire fences on cattle (see, Coe et al., 2001).

Differences in regions yielded different attributes associated with trajectories, and therefore accounted for the different clusters of trajectories. At the level of five regions, all trajectories of cattle were assigned to one cluster by EHRD (Figure 3.9j), while they were separated into two clusters by MHRD (Figure 3.9d-e). The assignment of all trajectories of cattle to one cluster increased the cluster cohesion (Figure 3.9j), since the trajectories of cattle in this cluster are very similar to each other due to movement restrictions imposed by barbed-wire fences. In contrast, assigning trajectories of cattle into two clusters reduced the separation between the two clusters

(Figure 3.9d-e). EHRD also produced two almost pure clusters of trajectories of mule deer (Figure 3.9h-i). These explained the higher the Silhouette index of clusters produced by EHRD (Figure 3.7). The different attributes associated with trajectories also affected the performance of decision tree. The decision tree that used the attributes from EHRD achieved higher accuracy and simplicity, because the attributes effectively captured the characteristics of movement of different species.

The different partitioning of regions also resulted in different point densities of cattle, mule deer and elk in each region. The point density of mule deer and elk were negatively correlated at the level of the four and five regions detected by EHRD (Figure 3.10). This pattern may reflect the competition of elk and mule deer on the usage of forage resources (Coe et al., 2001). The point that is below the regression line in Figure 10d represented the point density of elk and mule deer in Region 1 of Figure 3.4h, where cattle dominated. With the presence of cattle in this region, the point density of both elk and mule deer was lower off the trend line. This might be explained by competition for forage between cattle and elk and cattle and mule deer (Coe et al., 2001).

3.5.2 Differences in the design of Modularity and Edge ratio

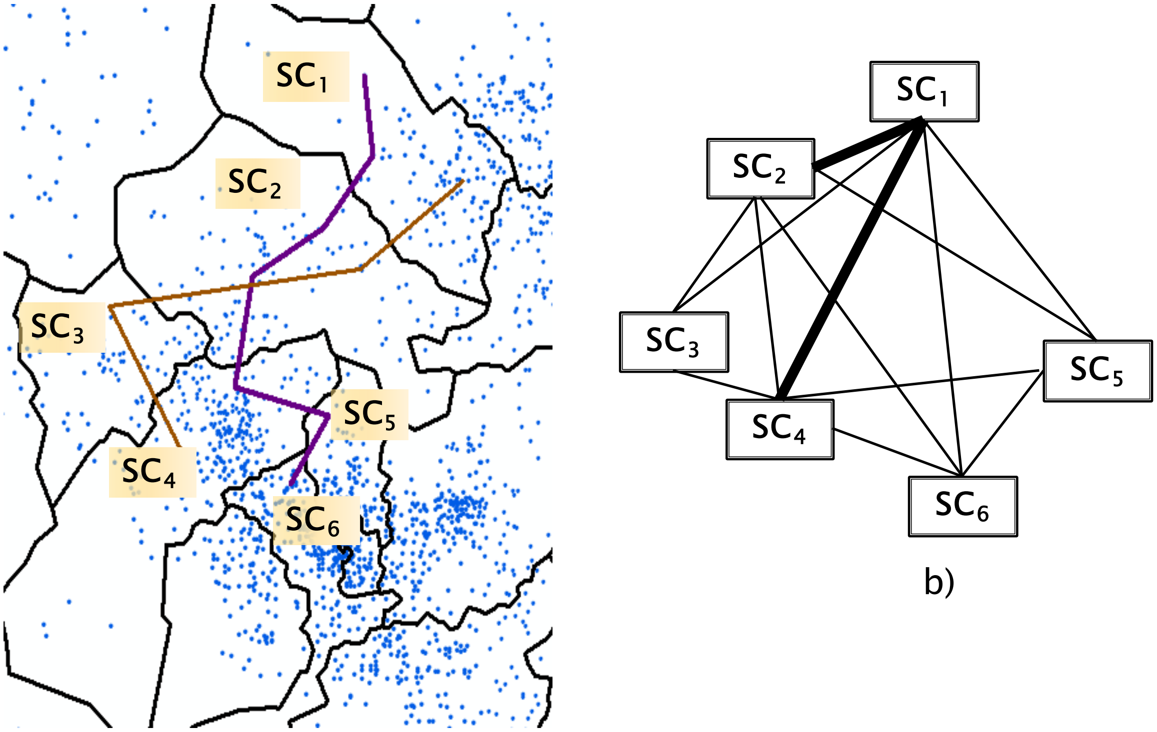
Compartments (i.e., regions in this study) are described as groups of nodes in networks or graphs with many edges joining nodes of the same group but comparatively few edges joining nodes of different groups (Fortunato 2010). Modularity and Edge ratio are two criteria have been developed to quantify edge connections within and between groups of nodes for the compartment delineation. Although modularity has been widely used as a criterion to optimize in compartment detection, the limits of modularity have also been recognized.

First, modularity compares the actual connections among nodes in graphs or networks with those in graphs (i.e., null models) where nodes are connected in random. It assumes that each node can be connected to other nodes in the null models. This assumption is tenuous at best for large systems (Fortunato 2010) and does not appear to be correct in this study. Due to the effects of barbed-wire fences, it is unreasonable to assume that cattle can move from spatial clusters within fenced off area to those outside of the fencing. In other words, spatial clusters within and outside of the barbed-wire fences should not be connected by the movement of cattle. Therefore, it is invalid to compare the actual connections to connections in null models where nodes (i.e., spatial clusters in this study) are assumed to be connected in a random manner. Edge ratio determines the strength of within-compartment connections by comparing the actual connections within the compartments and the actual connections between compartments. Therefore, edge ratio does not rely on any null model.

Second, modularity optimization has a low ability to detect compartments that are comparatively small with the respect to the graph as a whole (Fortunato 2010). More specifically, modularity optimization tends to find compartments that have a relatively even number of connections (as illustrated in Figure 3.2). The within-region connections were uneven in Regions 1 and 2 found by EHRD (701,698 vs. 88,140: Figure 3.6d). Region 1 contained most telemetry locations of cattle. In contrast, the within-region connections in Regions 1 and 2 found by MHRD in Figure 3.6a were relatively even (360,300 vs. 276,162). This accounts for the reason that Modularity-based Hierarchical Region Discovery did not delineate the tight boundary of cattle movement.

3.6 Conclusion

In this study, we investigated the movement patterns of cattle, mule deer, and elk at the Starkey Experimental Forest and Range by finding regions such that organisms exhibit greater movement within regions and less movement across regions. We compared our approach, which used the criterion of edge ratio to detect the regions, with more widely used modularity. Attributes of trajectories derived from regions detected according to the criterion of edge ratio produced higher quality of clusters of trajectories as evaluated by the Silhouette index and achieved higher accuracy and simplicity in classifying trajectories of different species compared to the previous work. We also found that the regions detected according to the criterion of edge ratio might suggest competition among cattle, mule deer, and elk**.** Edge ratio-based Hierarchical Region Discovery is more suitable to analyze animal trajectory and is a potentially useful tool for exploring hidden or unknown patterns in animal movement.



b)

a)

Figure 3.1 Creation of a weighted graph based on trajectories linking spatial clusters. a) the aggregation of points traversed by trajectories into Spatial Clusters (*SC*s). b) the resulting weighted graph created by counting connections between *SC*s that represent the pairs of points in the trajectories.

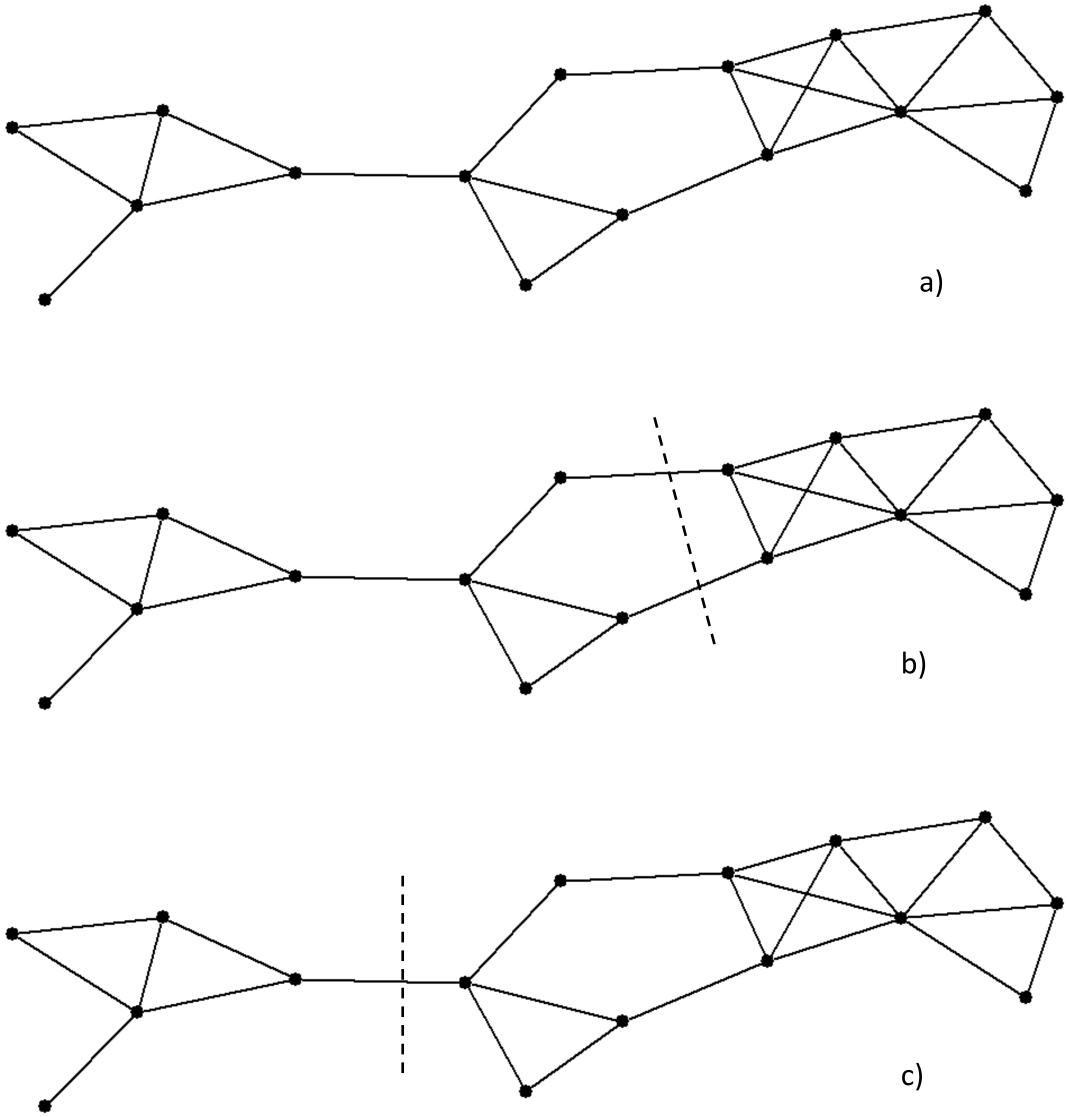


Figure 3.2 Partitioning of a graph (a) into two compartments by optimizing: b) modularity or c) edge ratio.

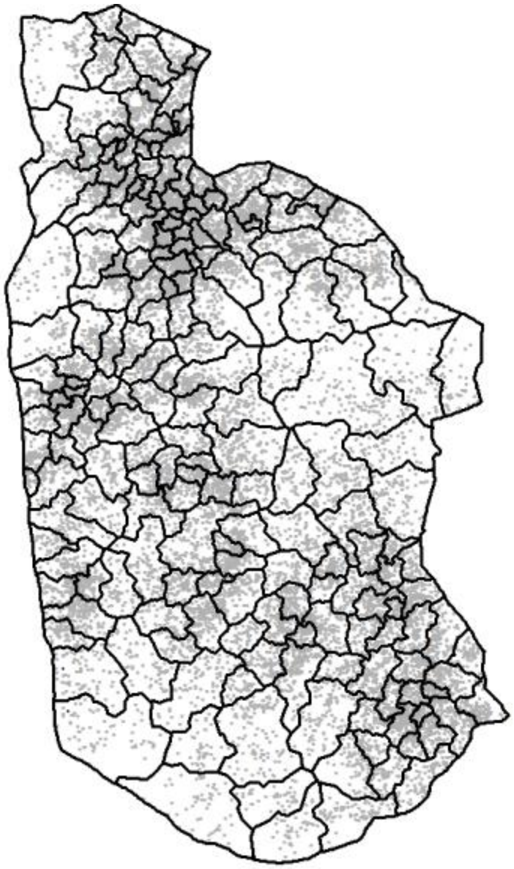


Figure 3.3 Aggregation of animal trajectory points (n=14,990) at Starkey Experimental Forest and Range into 203 spatial clusters.

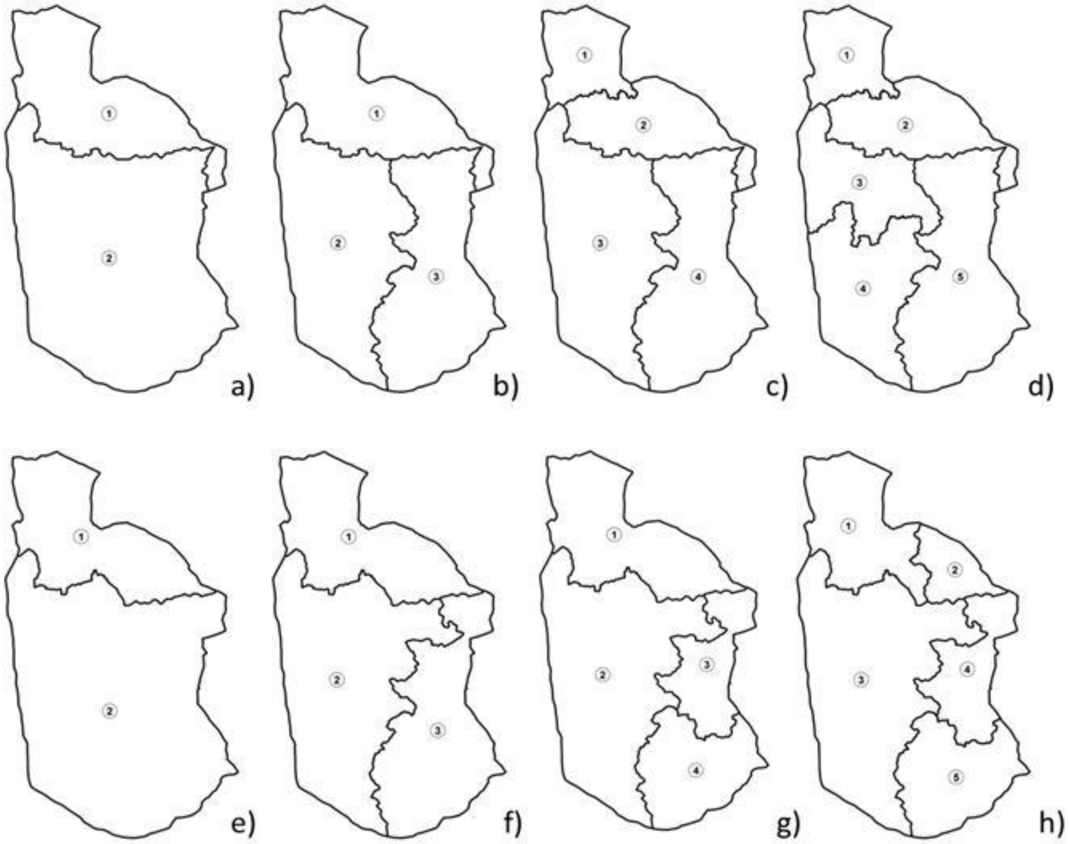


Figure 3.4 Hierarchical regions of animal trajectories at Starkey Experimental Forest and Range generated by a-d) Modularity-based Hierarchical Region Discovery (MHRD), and e) - h) Edge ratio-based Hierarchical Region Discovery (EHRD).

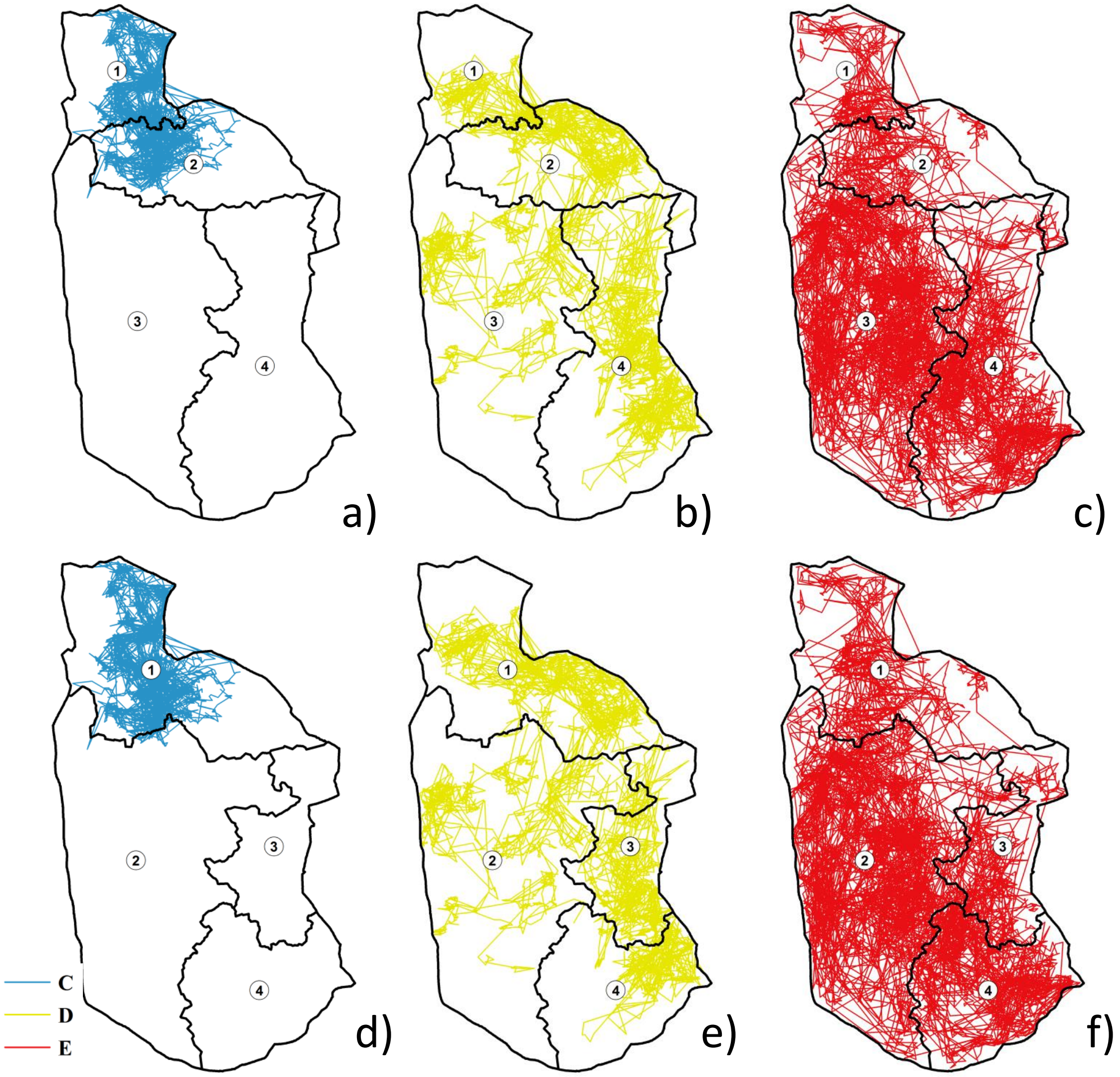


Figure 3.5 Trajectories of 34 cattle (C), 30 mule deer (D), and 38 elk (E) at Starkey Experimental Forest and Range overlaid on a four region classification generated by a) to c) Modularity-based Hierarchical Region Discovery (MHRD), and e) to f) Edge ratio-based Hierarchical Region Discovery (EHRD).

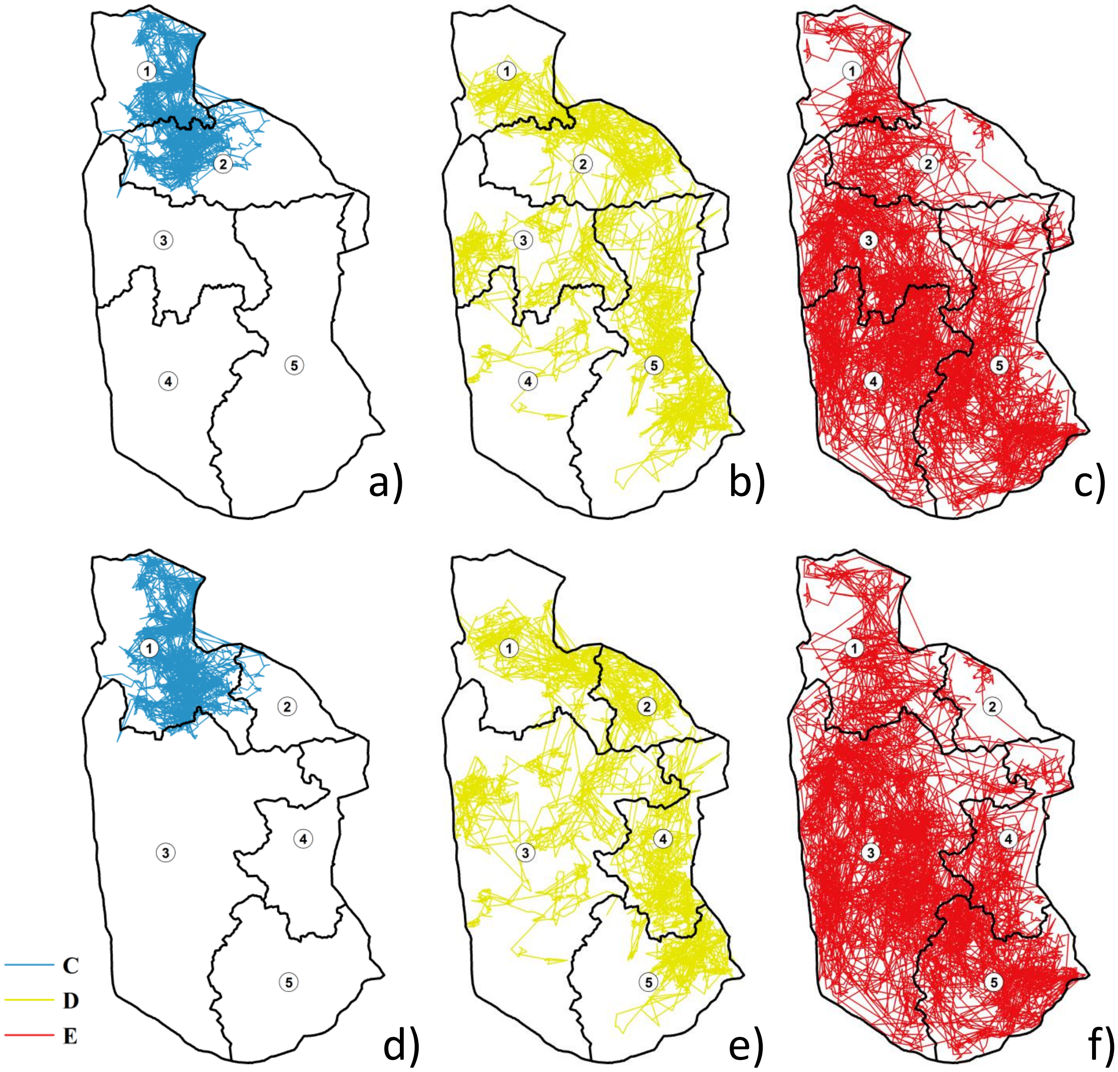


Figure 3.6 Trajectories of 34 cattle (C), 30 mule deer (D), and 38 elk (E) at Starkey Experimental Forest and Range overlaid on a five region classification generated by a) to c) Modularity-based Hierarchical Region Discovery (MHRD), and e) to f) Edge ratio-based Hierarchical Region Discovery (EHRD).

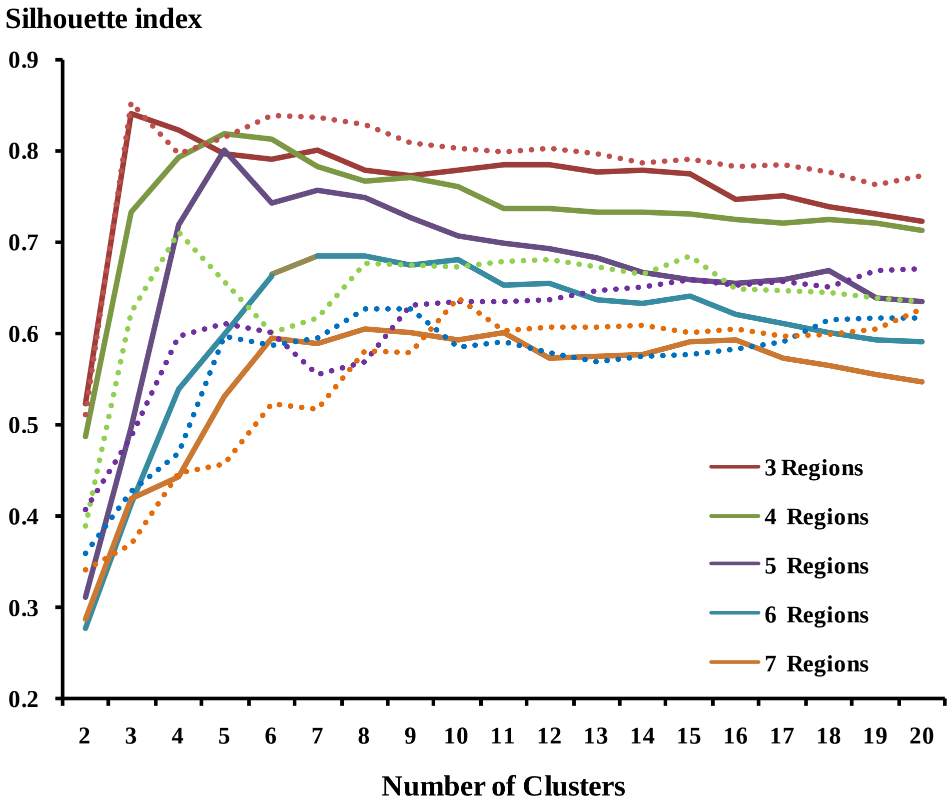


Figure 3.7 Silhouette index of 2 to 20 clusters produced by Modularity-based Hierarchical Region Discovery (MHRD, dashed lines) and Edge ratio-based Hierarchical Region Discovery (EHRD, solid lines) at the level of 3 to 7 regions

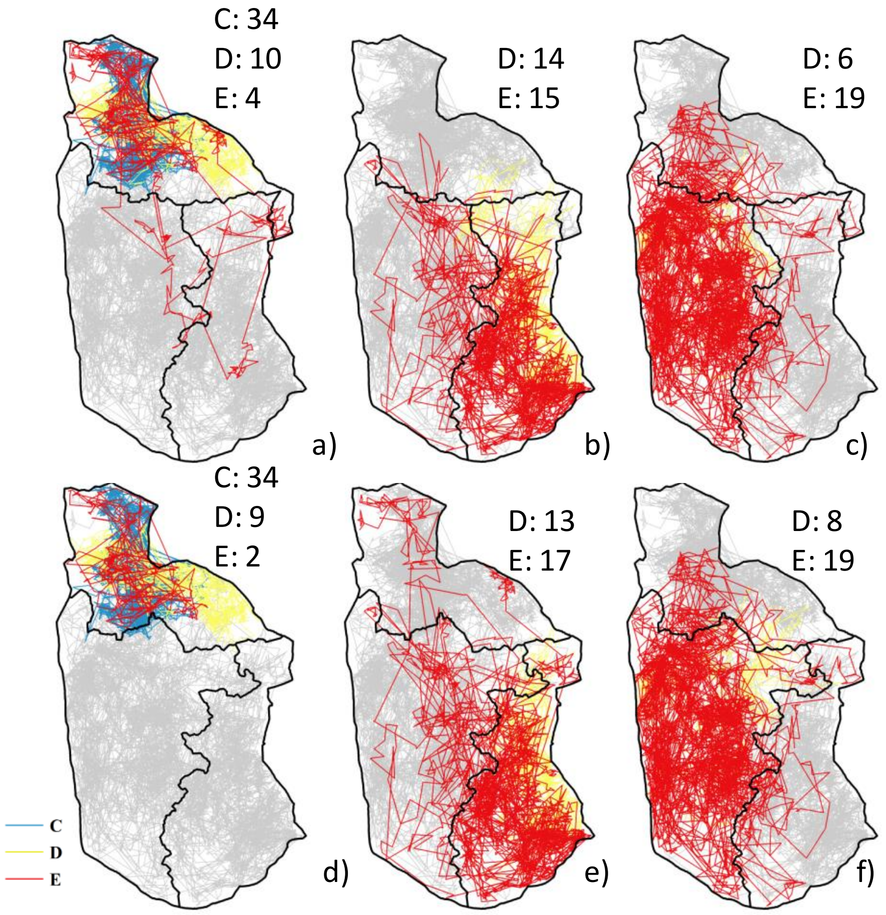


Figure 3.8 Three region classification of 102 trajectories including 34 cattle (C), 30 mule deer (D), and 38 elk (E) at Starkey Experimental Forest and Range at the level of three regions produced by a) to c) Modularity-based Hierarchical Region Discovery (MHRD), and d) to f) Edge ratio-based Hierarchical Region Discovery (EHRD). Labels indicate the number of cattle (C), mule deer (D), and elk (E) in each cluster.

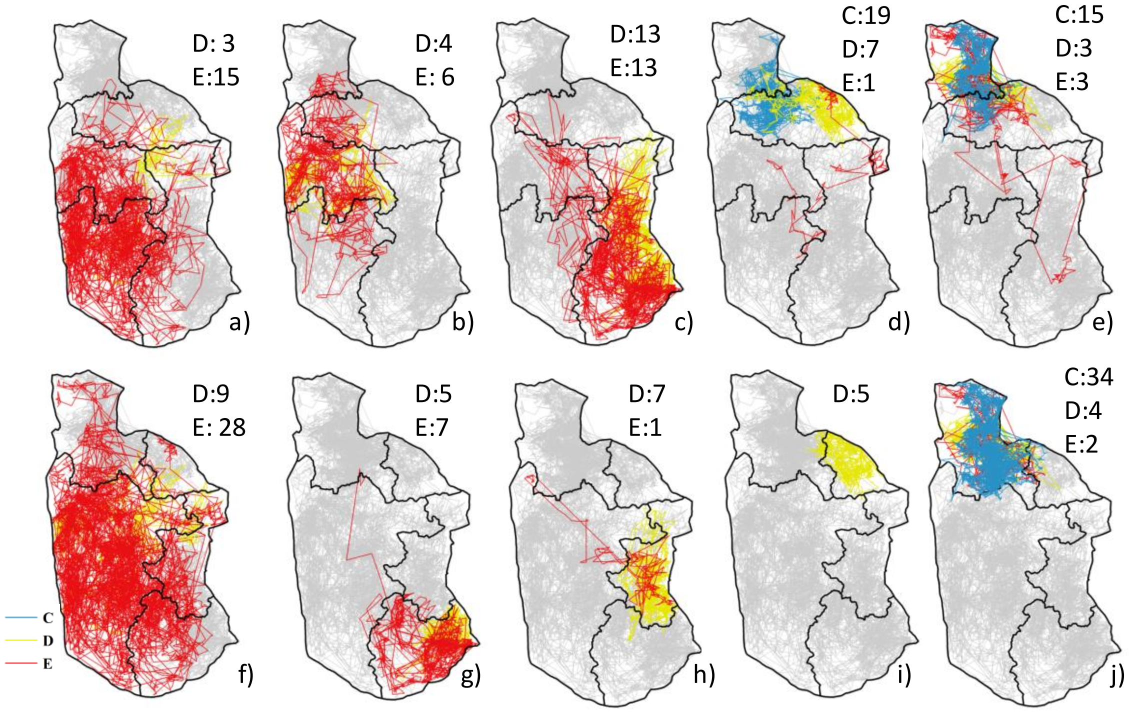


Figure 3.9 Five region classification of 102 trajectories including 34 cattle (C), 30 mule deer (D), and 38 elk (E) at Starkey Experimental Forest and Range at the level of five regions produced by a) to c) Modularity-based Hierarchical Region Discovery (MHRD), and d) to f) Edge ratio-based Hierarchical Region Discovery (EHRD). Labels indicate the number of cattle (C), mule deer (D), and elk (E) in each cluster.

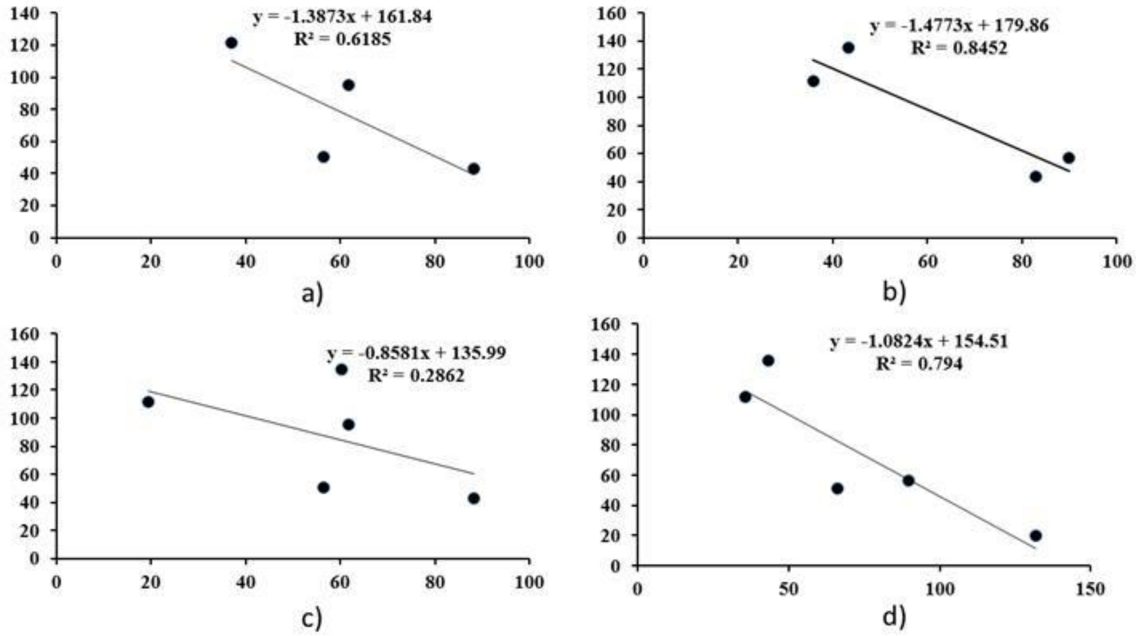


Figure 3.10 The relationship between point density (number of points per square kilometer) of mule deer (x axis) and elk (y axis) in: four regions generated by a) Modularity-based Hierarchical Region Discovery (MHRD), b) Edge ratio-based Hierarchical Region Discovery (EHRD), and five regions generated by c) Modularity-based Hierarchical Region Discovery (MHRD), and d) Edge ratio-based Hierarchical Region Discovery (EHRD).

Table 3.1 Point density (number of points per square kilometer) of cattle, mule deer, and elk at the level of five regions detected by Modularity-based Hierarchical Region Discovery (MHRD) and Edge ratio-based Hierarchical Region Discovery (EHRD). Region boundaries and animal trajectories are shown in Figure 3.6.

	MHRD			EHRD		
Region ID	Cattle	Mule Deer	Elk	Cattle	Mule Deer	Elk
1	199.8	56.4	50.3	235.4	66.2	51.5
2	137.3	88.2	43.0	9.2	131.9	19.9
3	0.1	60.2	134.8	1.6	35.8	111.8
4	0.0	19.4	111.3	0.0	89.9	56.5
5	0.0	61.7	95.1	0.0	43.3	135.6

Table 3.2 Accuracy and number of leaves (simplicity) from 10-folder decision tree test using variables derived from Modularity-based Hierarchical Region Discovery (MHRD) and Edge ratio-based Hierarchical Region Discovery (EHRD) to classify 102 trajectories into cattle, mule deer, and elk.

	Accuracy				Number of leaves			
	Average	Stand Deviation	Max	Min	Average	Stand Deviation	Max	Min
MHRD	77.3%	15.1%	90.9%	40.0%	8.5	1.6	11	6
EHRD	87.3%	11.5%	100.0%	70.0%	6.0	1.7	9	4

Chapter 4 Uncovering Food Web Structure Using a Novel Trophic Similarity Measure[3]

Abstract

Aggregation of species on the basis of their trophic roles in food webs is a fundamental step for uncovering the structure of food webs in ecosystems. Although the Additive Jaccard Similarity (AJS) has been widely used to measure trophic similarity between species, it has also been criticized for its low ability to find species with equivalent trophic roles, especially when they do not share the same predators and prey. In this study, we proposed a new trophic similarity measure, the Extended Additive Jaccard Similarity (EAJS), that incorporates not only the similarity of shared predators and prey at adjacent trophic levels but at all trophic levels. The two trophic similarity measures (AJS and EAJS) were used to aggregate species in the mammalian food web for the Serengeti ecosystem in northern Tanzania and southern Kenya. Compared to AJS, the clusters of species based on EAJS had higher quality, which means that species in the same clusters have higher similarity and species in different clusters have higher dissimilarity in terms of their trophic relationships in the food web. Clusters found on the basis of EAJS also better reflected ecological factors known to structure food webs. Plants of the same habitat tended to grouped in same clusters, and the grouping of

---

[3] This manuscript will be submit to *Ecological Research*

animals was more clearly related to their weights. The advantage of EAJS lies in the fact that it is designed to consider species feeding relations in food webs in a broad scale (i.e., not limited to adjacent trophic levels). Our approach provides a means for revealing the patterns of trophic relations among species in food webs and exploring known and unknown factors shaping food web structure.

4.1 Introduction

Food webs have been and continue to be a central research focus in many areas of ecology because of the importance of energy flows in structuring population dynamics, maintaining biodiversity and ecosystem integrity, and shaping network topology (Ruiter et al. 2003). The search for order and simplicity within food webs has attracted the attention of researchers for over a century (Elton 1927), including efforts to uncover their structural properties (e.g. Polis 1991; Havens 1992), reveal the rules shaping their intricacy (e.g. Willams and Martinez 2000), and capture species roles and interactions within them (e.g. Luczkovich 2003; Jordán 2009). While the interactions among species that form the basis of food webs may be complex (Polis 1991), food webs are non-random and highly patterned in nature (Pimm 1982) and are regulated by a limited number of biological processes. For example, Cohen (1990) summarized five laws that shaped food web structures while Williams and Martinez (2000) succeeded in predicting twelve properties of food webs using only two parameters: species number and connectance.

As with the taxonomic classification system established by biologists to hierarchically order organisms, one efficient way of reducing the complexity of food webs to better understand them is to decompose them into groups of species according to certain criteria or definitions. Approaches to grouping species in food webs have been examined from various angles, and controversy remains in terms of linking the detected groups to ecologically-meaningful structuring agents. Ideally, such approaches would be rooted in studies of the functional roles occupied by groups of species within food webs as determined by experimental manipulations (e.g., Paine 1980), but manipulative approaches are often impractical, or at least difficult, for many systems. Therefore, many studies of food web structure have focused on connectedness, that is, groups of nodes (e.g., individual species or an aggregation of species) and their observed trophic connections. There is a long tradition of the study of ‘connectedness webs’, in part because food web membership and trophic interrelations can be analyzed by observation alone.

One of the most influential and fundamental works on aggregating species into trophic groups based on their observed trophic connections is Yodzis and Winemiller (1999), who compared the performance of multiple criteria (e.g., additive and multiplicative Jaccard similarity) and clustering algorithms (e.g., average linkage, complete linkage) in aggregating species into trophic groups. They concluded that additive similarity is better than multiplicative similarity and that average linkage is one of the two methods that produce more consistent and ecologically-interpretable patterns of aggregation. Luczkovich et al. (2003), however, pointed out a fundamental limitation of additive similarity in the treatment of species with similar trophic roles; specifically, it lacks the ability to find species with equivalent trophic roles, if they do not share the same predators and prey. For example, two herbivores that feed on totally different plants or are eaten by different carnivores are separated into different groups on the basis of additive similarity, even though they may play equivalent trophic roles in a food web. Luczkovich et al. (2003) adapted the concept of regular equivalence from social networks to partition food webs into isotrophic groups. Species in the same isotrophic group have the same or similar trophic roles in a food web, feeding on and being preyed upon by equivalent species (e.g., herbivores feed on plants and are eaten by carnivores).

An alternative approach to partitioning food webs focuses on detecting compartments, groups of species that have many feeding relations within the groups but fewer feeding relations across groups (Krause et al. 2003). From an ecological standpoint, compartmentalization is thought to contribute to stability of the food web (Melian and Bascompte 2002). For example, Krause et al. (2003) investigated the response of a food web to two disturbance scenarios and found that compartmentalization could reduce the impact of disturbance on the other compartments by constraining its impact to a single compartment.

To merge research approaches that on the one hand focus on strong within group feeding relations (compartments) vs. those that emphasize little or no within group feeding relations (isotrophic groups), Allesina and Pascual (2009) developed a unique probabilistic model that simultaneously considers the two types of patterns. By optimizing an Akaike Information Criterion function of species interaction and their assignment to groups, the model is able to detect the dominant pattern in food webs, either compartments or isotrophic groups. However, as indicated by the authors, the biological interpretation of such patterns needs to be further examined.

In this study, we aggregate species in a Serengeti ecosystem food web into clusters based on pairwise species similarity values calculated using the Additive Jaccard Similarity (AJS) coefficient and a novel trophic similarity measure (Extended Additive Jaccard Similarity, EAJS), which extends the Additive Jaccard Similarity to consider higher- and lower-order trophic level relationships. We then evaluate the aggregations of species according to EAJS and AJS using a cluster validity index and explore the biological and ecological factors which may account for the clustering of species.

4.2 Study area and dataset

This study investigated the food web of the Serengeti ecosystem, which covers an area of plains and open woodlands in northern Tanzania and southern Kenya. Famous for its biodiversity, including the largest herds of grazing mammals in the world (Sinclair and Norton-Griffiths 1984), the Serengeti has been the site of several seminal studies in grassland and savanna ecology. For example, McNaughton (1978) investigated the composite environmental factors contributing to the community organization in the Serengeti National Park while Sinclair et al. (2003) examined patterns of predation within the diverse mammal community. The food web used in this study is the same as that analyzed by Baskerville et al. (2011), which readers can refer to for details. The food web was compiled by collecting the feeding links in the literature (Casebeer and Koss 1970; McNaughton 1978; Cooper et al. 1999; Sinclair et al. 2003) together with some links known from personal observation by Baskerville et al. (2011). The resulting food web is composed of 592 feeding links among 161 species, which include 129 plants, 23 herbivores, and 9 carnivores and omnivores.

4.3 Method

4.3.1 Calculating Pairwise Species Similarity Values

To aggregate species into clusters and reveal the structure of food webs, we first define the similarity between each pair of species based on predator-prey relationships. We do so using two measures of species similarity: the Additive Jaccard Similarity (AJS) coefficient used by (Yodzis and Winemiller 1999) and others, and a novel trophic similarity measure, Extended Additive Jaccard Similarity (EAJS, *described below*). Species were then aggregated into groups by applying Average Linkage clustering (ALK) to the species similarity matrix.

For two species *i* and *j*, AJS is defined as:

$$\mathrm{AJS}(i, j) = \frac{a}{a+b+c} \quad (4.1)$$

where a is the total number of prey or predator of both species *i* and species *j*; b is the number of prey or predator of species *i* but not of species *j*, and c is the number of prey or predator of species *j* but not of species *i*. Values equal 1.0 when two species share the same predators and prey, and decrease when species have few predator or prey species in common.

EAJS differs in that it incorporates not only the similarity of shared predators and prey at adjacent trophic levels (i.e., the direct predators or prey of two species) but all the trophic levels associated with each species (Figure 4.1). Thus, rather than just searching for predators and prey species, EAJS iteratively searches for all higher and lower level predators and prey (i.e., species preyed upon by a prey species or predators of a predator species) until no additional higher- or lower-level linkages are found. If a species appears on two levels or more (e.g., the species is the predator of species *i* and is also the predator's predator of species *i*), only the feeding relationship on the closer level is considered. In doing so, the predators and prey of species *i* and *j* in all trophic levels are identified. AJS of species *i* and *j* is first determined at each equivalent level (e.g., the prey of species *i* and *j*, the predator's predator of species *i* and *j*). EAJS between species *i* and *j* through all levels is then calculated as the sum of AJS at each level (Equation 1) divided by the maximum number of levels for species *i* and *j*.

Calculations of AJS and EAJS are illustrated in Figure 4.1. AJS is based on the shared prey and predators at adjacent trophic levels. In this case, the number of shared prey is 1, the number of shared predators is 2, and the number of prey or predators that are not shared by species *A* and *B* is 4. Therefore, AJS equals 0.43 (i.e., (1+2)/(1+2+4)). EAJS (0.46) is the sum of AJS at each equivalent level (i.e., 1/3, 1/2, 1, and 0 at the prey level and predator level 1-3 respectively) divided by the maximum number of levels (i.e., species *A* has 4 trophic levels of predators or prey).

ALK uses the pair-wise similarity matrices of species produced by AJS and EAJS to aggregate the species into clusters. ALK defines the distance between two clusters as the average dissimilarity between all cross-cluster pairs of species. It builds a dendrogram by iteratively merging the species or clusters which have the shortest distance. This approach is hierarchical, and any number of clusters can be obtained by cutting the dendrogram until a desired number is reached.

4.3.2 Evaluation of Food Web Clusters

Defining the similarity between each pair of species on the basis of AJS vs. EAJS will, in most cases, yield two different similarity matrices and thus two clustering results (i.e., two ways of aggregating species in the food web). While there is no 'best' cluster result, we evaluate the results of the food web partitions derived using AJS and EAJS based on two criteria: 1) the quality of clusters as a function of cohesion of species within clusters and separation of species in different clusters, and 2) the ability of the methods to identify clusters distinguished by ecological factors known to structure food webs.

We first evaluated the quality of clustering based on the two similarity matrices using a cluster validity index, the Silhouette index (Rousseeuw 1987). In an extensive comparative study of thirty cluster validity indices, Arbelaitz et al. (2013) examined index performance using synthetic and real datasets under different test conditions, such as different clustering methods, various cluster densities, and multiple levels of noise. The Silhouette index was recognized as the most suitable cluster validity index in terms of successful rates of recognizing the number of clusters in the different datasets under various test conditions.

The Silhouette index (S) is calculated as:

$$S = \frac{1}{n}\sum_{i}^{n} \frac{(b_i - a_i)}{\max(a_i, b_i)} \qquad (4.2)$$

where n is the number of species in the food web, a_i is the average distance between species i and all other species in its own cluster and b_i is the minimum of the average dissimilarities between i and species in other clusters. a_i measures the overall dissimilarity of species *i* to other species in the same cluster (i.e., the cohesion of the clusters, with smaller values meaning higher cohesion), and b_i measures the average distance of species *i* to species in the cluster that is most similar or closest to it (i.e., the separation of the clusters, with larger values meaning better separation). When a_i is greater than b_i, the Silhouette index for species *i* is negative, suggesting that species *i* is more similar to species in other clusters. When b_i is greater than a_i, it means that the average distance of species *i* to species in the 'nearest' cluster is larger than that of species *i* to other species in the same cluster. In this case, the Silhouette index measures the difference between the two distances, scaled to the former. Therefore, larger values suggest a higher quality.

While the Silhouette Index provides a means for assessing the mathematical quality of groups, it is important that clusters identified using the two measures capture ecologically-meaningful relationships, as well. The ecological and biological factors that structure food webs have been widely investigated. We focus on two factors: 1) habitat, and 2) body size. Pimm (1980) found that compartments exist in food webs that span major habitat divisions (e.g. forest and prairie, adjacent freshwater and terrestrial habitats). The role of habitat boundaries in shaping compartments was further investigated by Girvan and Newman (2002) and Krause et al. (2003). Such patterns may be expressed particularly at the producer level.

To test the degree of the overall clustering of habitat in the groups of producers, we compared Shannon entropy values for habitats in groups of plants identified by the clustering process against those in randomized groups. Primary habitats of the 129 plants species were assigned to one of eight classes (grassland, woodland, riparian, kopje, shrubland, thicket, disturbed, undetermined) on the basis of studies of plant community composition and personal knowledge of the system (Baskerville et al. 2011). Shannon entropy was used to measure the habitat signature in each group of plants, that is, the clustering of plants in different habitats among the groups. For group *i*, the Shannon entropy (H_i) is defined as:

$$H_i = -\sum_{\mathrm{j}} \frac{n_{ij}}{n_i} \log(\frac{n_{ij}}{n_i}) \quad (4.3)$$

where *j* is the habitat, n_i is the size of group *i*, and n_{ij} is the number of plants that are assigned to habitat *j* in group *i*. A low Shannon entropy value indicates clustering of habitats in the group. The overall clustering of habitats for all groups of plants is measured by the sum of Shannon entropy of each group, weighted by the size of each group:

$$H = \sum_{\mathrm{i}} \frac{n_i}{n} H_i \quad (4.4)$$

where *n* is the total number of plants in all groups.

To test the significance of the clustering of habitats in the groups, we calculate the *p*-value as the probability of a value lower than or equal to *H* drawn from the randomized partitions with groups of identical size. To determine whether a group of plants is overrepresented by plants of a certain habitat (i.e., significant clustering of a habitat in a group), we calculate the *p*-value as the probability that a randomized group of the same size would have as many or more plants of the habitat (Baskerville et al. 2011).

In addition to habitat, it is well documented that trophic patterns are shaped by body size, which influences predator-prey relationships. Predators typically consume prey that are smaller than themselves, although larger predators eat prey with a wider range of body sizes than smaller predators (Cohen et al. 1993; Brose et al. 2006a; Brose et al. 2006b; Riede et al. 2011). Therefore, it is not surprising that body size has been incorporated as a primary factor in understanding food web structure (Paine 1963; Williams and Martinez 2000; Emmerson and Raffaelli 2004; Stouffer et al. 2005; Petchey et al. 2008). We collected the weights of animals in this food web from Roberts (1951) and Sinclair et al. (2003). In the study of predation patterns of Serengeti ecosystem, Sinclair et al. (2003) found that the ungulates can be broken down into groups according to their vulnerability to predators which are related to their weights. The ungulates whose weights are less than 150 kg are more vulnerable to predators. They also found that groups of carnivores are distinguishable based on the weight range of herbivores they prey on. We compared the aggregation of animals with the grouping of animals in Sinclair et al. (2003) and examined whether the patterns exist between the clustering of species and their weights.

4.4 Results

4.4.1 Cluster Results

The clustering process is hierarchical, meaning that users can select any number of food web compartments depending on the level of detail desired. Here, we discuss results for 18 compartments, the level of detail presented by Baskerville et al. (2011) for the same dataset. This level of detail also had a high Silhouette Index and resulted in ecologically-meaningful compartments (*both discussed below*).

Partitioning of the Serengeti food web into eighteen groups based on EAJS values clearly distinguished three trophic levels: carnivores and omnivores (Groups 1-3), herbivores (Groups 4-11), and producers (Groups 12-18) (Table 4.1, Figure 4.3). At the highest trophic level (carnivores and omnivores), Group 1 contained just one species, *Caracal caracal*, which has no predator and relies on only four herbivores. This is in contrast to carnivores in Group 2, which included large carnivores that utilize a broader range of prey, and Group 3, which contained all carnivores eaten by *Panthera pardus*.

Species in the eight groups at the herbivore level (Groups 4-11) showed patterns that can be tied to differences in their predator and prey species. Groups 10 and 11 each contained only one species, *Loxodonta africana* and *Hippopotamus amphibious*, respectively, neither of which has a predator. *L. africana* is terrestrial and eats terrestrial plants, while *H. amphibius* is an aquatic animal and feeds on aquatic plants (Bigalke et al. 1954). Although the former eats 46 plants and the later feeds on 9 plants, they only share 5 plants in common. Of the remaining six groups, five (Groups 4, 5, 6, 8, and 9) were distinguished on the basis of specific predator or prey species. Species in Group 5 (*Procavia capensis*), Group 6 (*Heterohyrax brucei* and *Papio anubis*) and Group 9 (*Giraffa camelopardalis* and *Syncerus caffer*) are only eaten by one species (*Caracal caracal*, *Panthera pardus* and *Panthera leo* respectively), while species in Groups 4 and 8 have only have two predators: *Taurotragus oryx*in (Group 8) is eaten by *P. leo* and *P. pardus*, and *Pedetes capensis* (Group 4) is eaten by *C. caracal* and *P. pardus*. Group 7 contained the largest number of herbivores, each of which has at least three predators.

At the primary producer level, Groups 17 and 18 both comprised only one species that is eaten by a particular predator (i.e., *Olea spp.* in Group 17 is eaten only by *G. camelopardalis* and *Panicum repens* in Group 18 is eaten only by *H. amphibius*). In contrast, the consumers of plants in Group 13 included all of the herbivores and secondary consumers that eat these herbivores. Species in the remaining groups (Groups 12, 14, 15 and 16) were aggregated on the basis of different primary and secondary consumers. Predators of species in Group 14 included only three species (*Heterohyrax brucei, Papio anubis,* and *Loxodonta africana*) while the only secondary consumer of this group is *Panthera pardus*, which feeds on both *Heterohyrax brucei* and *Papio anubis*. Species in this group are similar mainly because *P. pardus* is the only secondary consumer of them and they have the highest Additive Jaccard Similarity (the value is 1) at the level of secondary consumer. In contrast, the predators of species in Group 12 are five totally different species (i.e., *Alcelaphus buselaphus, Damaliscus korrigum, Kobus ellipsiprymnus, Pedetes capensis,* and *Procavia capensis*). Five species (i.e., *C. caracal, Crocuta crocuta, Lycaon pictus, Panthera leo,* and *Pantherap ardus*) were the secondary consumers of species in Group 12, of which *P. pardus* is the only one that overlaps with the secondary consumers of Group 14.

In contrast to the food web identified using EAJS values, AJS identified two groups at the level of carnivores and omnivores (Figure 4.4). *P. pardus* and its two prey (*Acinonyx jubatus* and *Canis aureus*) were assigned to the same group (Group 2) while the other two prey of *P. Parus* were assigned to Group 1. The four prey species of *P.pPardus* were all assigned to Group 3 based on EAJS (Figure 4.3). AJS detected six groups at the level of herbivores, while EAJS found eight. AJS did not distinguish *L. africana* and *H. amphibious,* that have no predators, from other species as EAJS did. AJS recognized five groups of plants (Groups 14-18 in Figure 4.4) that only included one or two plants. Plants in these five groups have only one or two particular predators of their own, and they do not share any predators. They were separated from other groups of plants, because they do not share any predators other than their particular predators in common with other plants, and similarity of these plants to other plants is very low by the definition of AJS.

4.4.2 Cluster Validity

Clusters derived from EAJS consistently had higher Silhouette Index values than those from AJS. This suggests a higher quality of clusters, which means species in the same clusters exhibit greater cohesion, that is, they are more similar to each other while species in different clusters are more different from each other in terms of their trophic relationships in the food web.

Compared to AJS, the clusters of species found on the basis of EAJS not only had higher quality, but also revealed more detailed patterns related to the habitat segregation of plants and network topology associated with weights of the animals. Compared to the randomized mean value of 1.37, mean weighted Shannon entropy across 18 groups identified by EAJS was 1.21 (p = 0.00003), suggesting that plants of the same habitat type are significantly clustered in groups. Moreover, Groups 12, 13, and 16 were significantly overrepresented by different habitat types. In Group 12, riparian plants were significantly overrepresented, comprising 50% of the group, compared with a random expectation of 6.25% ($p < 0.005$). Group 13 was overrepresented by woodland plants, which occupied 36.21% of plants in this group, compared to a random expectation of 25.00% ($p < 0.005$). Group 16 contained 41.38% kopje plants compared to a random expectation of 19.64% ($p < 0.002$). Mean weighted Shannon of 18 groups that identified by AJS suggested an overall clustering of habitats within the groups of plants (1.12 vs. randomized mean value of 1.31, $p = 0.00001$), but no types of habitat were overrepresented in any group ($p > 0.05$).

The grouping of species (Groups 1-3) at the level of carnivores and omnivores reflected the expectation that predators typically consume prey that are smaller than themselves and that larger predators eat prey with a wider range of body sizes than smaller predators. Carnivores in Group 2 are large predators with weights > 50 kg that feed on a wide range of mammals. In contrast, species in Groups 1 and 3 are smaller predators who have a narrower range of smaller mammals that are less than 11 kg, except for *Acinonyx jubatus* in Group 3, which eats mammals up to 100 kg. The groups developed using AJS did not show the pattern that carnivores or omnivores were assembled on the basis of prey sizes, as those derived using EAJS did.

At the level of herbivores, the number of their predators decreases from Group 7 to Group 11, as determined using EAJS. Each species in Group 7 has at least three predators while those in Group 8 had two predators species, and those in Group 9 had just one. The species in Groups 10 and 11 had no predators. Meanwhile the average weights of the species increase from Group 7 to Group 11. In the study of predation patterns of Serengeti ecosystem, (Sinclair et al. 2003) found that the ungulates with mean weights less than 150 kg are more vulnerable to predators. It is interesting to find that all the species in Group 7 are less than 150 kg including all the five species reported by (Sinclair et al. 2003). The wider range of predators of species in Group 7 probably reflects the vulnerability of these species to predators. (Sinclair et al. 2003) also found that the chance of species' mortality caused by predators decreases, as weights of the species increase. *L. africana* (Group 10) and *H. amphibius* (Group 11) are two large mammals with no predator. The chance of their mortality caused by predators almost drops to zero due to the large body size (Sinclair et al. 2003). Once again, no such pattern was observed in groups found by AJS.

4.5 Discussion

Methods and approaches for reducing the complexity of food webs have grown in recent years because they provide a means for better understanding food web structure and stability and for projecting the potential effects of anthropogenic and natural disturbances on biodiversity and ecosystem integrity. In this study, we aggregated mammalian species in the Serengeti ecosystem using average linkage clustering based on two trophic similarity measures, the Additive Jaccard Similarity index and an Extended Additive Jaccard Similarity. The difference between these two measures is that the latter considers not only the similarity of shared predators and prey at adjacent trophic levels but all the trophic levels associated with the species. This broader interpretation of food web connectance provided by the way that EAJS determines similarity considers the interactions of one species with other species in the network as a whole, without limiting such interactions to direct feeding relations.

From an ecological standpoint, EAJS makes more ecological senses when we compare the clusters derived by AJS and EAJS at the producer level. For example, *Sporobolus festivus* was assigned to a single cluster by AJS (Group 18 in Figure 4.4 and Table 4.2), mainly because it is only eaten by *Aepyceros melampus* and does not share any predator other than *A. melampus* in common with other plants. The similarity between *Sporobolus festivus* and *Sporobolus fimbriatus*, which is eaten by *Nanger granti* and *Eudorcas thomsonii*, is zero according to AJS, because they share no predator in common. AJS failed to capture their similar trophic roles as producers in the food web. In contrast, their similarity is 0.417 according to EAJS, because they share five secondary consumers, including all of the four species in Group 2 and *Acinonyx jubatus* in Group 3 in Figure 4.3. They were thus assigned to the same group (Group 13 in Figure 4.3 and Table 4.1). Overall, AJS underestimated the similarity among the plants in terms of their similar trophic roles as producers in the food web, especially when two plants share few or no predators in common. The underestimation reduced the cohesion in the clusters of plants and accounted for the lower Silhouette Index compared to that derived from EAJS.

The food web examined in this study did not include any biological information aside from a set of nodes representing species and links representing their interactions. However, the aggregation of species on the basis of EAJS made more biological sense and revealed patterns associated with habitat types of plants and weight of animals.

Similar to a study on the same food web using a sophisticated Bayesian group modeling, we also found that the plants from the same habitats tend to be assembled. As suggested by Baskerville et al. (2011), different habitat types have distinct spatial distributions in the Serengeti. Therefore, the structure at the producer level may partially reflect the flow of energy and nutrition supplying the food web from different spatial location, with herbivores integrating spatially separated groups of plants, and carnivores integrating spatially widespread herbivores. What distinguishes our approach and results from those of Baskerville et al. (2011) is the aggregation of herbivores. Baskerville et al. (2011) claimed a group named "small herbivores" which included *Hippopotamus amphibius, a large mammal* and a miscellaneous group which included herbivores ranging from small (e.g., *Madoqua kirkii*) to large size (e.g. *Loxodonta africana*). The aggregation of herbivores on the basis of EAJS suggested a strong and clear pattern associated with animal weights.

In addition to patterns related to habitat structure of plants and network topology associated with animal weights, the clustering of species by EAJS is better able to identify ecological linkages across the entire trophic system than AJS. For example, the clustering based on EAJS identifies the groups of producers that support the carnivores and omnivores that comprise Groups 1-3 (Figure 4.3). Species in Group 3 (with only one exception) rely solely on producers in Group 13. Groups 1 and 2 are ultimately supported by plants in various producer groups, but they have different primary producer groups providing food sources. Group 2 primarily relies on plants in Group 13. In contrast, the four species (especially *Procavia capensis* in Group 5) that serve as prey of *C. caracal* in Group 1 eat not only plants in Group 13, but also many of the plants in Group 16. This indicates that the loss of plants in Group 13 will have greatest impact on the species in Group 3 because of their heavy reliance on producers in Group 13. Conversely, there would be less impact on *C. caracal* in Group 1 because it relies on other producers.

4.6 Conclusion

In this study, we developed a novel trophic similarity measure, Extended Additive Jaccard Similarity (EAJS), that considers not only the similarity of shared predators and prey at adjacent trophic levels but at all the trophic levels. Aggregation of species in the Serengeti ecosystem based on EAJS was compared to the clusters of species derived on the basis of the more widely used Additive Jaccard Similarity (AJS). We found that the clusters of species based on EAJS had higher quality compared to these based on AJS which means that species in the same clusters have higher similarity and species in different clusters have higher dissimilarity in terms of their trophic relationships in the food web. Clusters derived from EAJS values also better reflected ecological factors known to structure food webs. Plants of the same habitat tended to cluster in groups. The grouping of animals was related to their weights. The advantage of EAJS lies in the fact that it considers species feeding relations in food webs in a broad scale (i.e., not limited to adjacent trophic levels). Our approach provides a means for revealing the patterns of trophic relations among species in food webs and exploring known and unknown factors shaping food web structure.

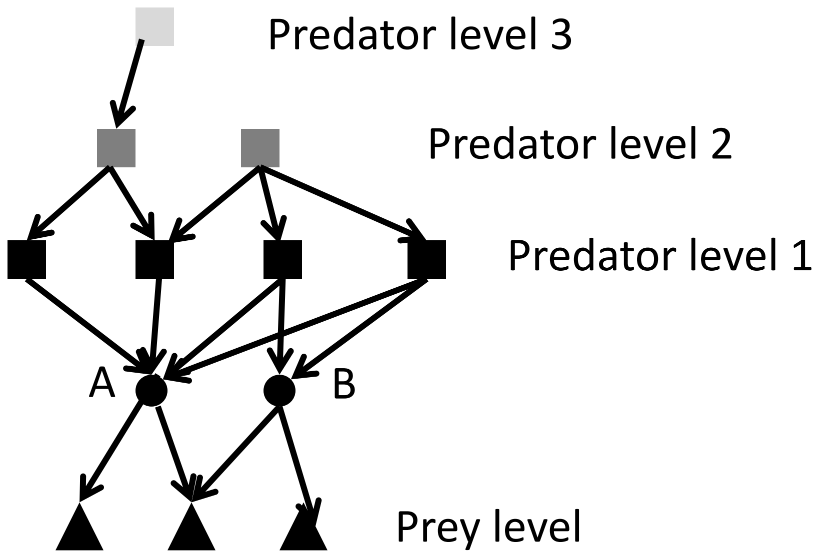


Figure 4.1 Comparison of Additive Jaccard Similarity (AJS) and Extended Additive Jaccard Similarity (EAJS). AJS is calculated based on prey and predators only at adjacent trophic levels, while EAJS is based on prey and predators in at all the trophic levels.

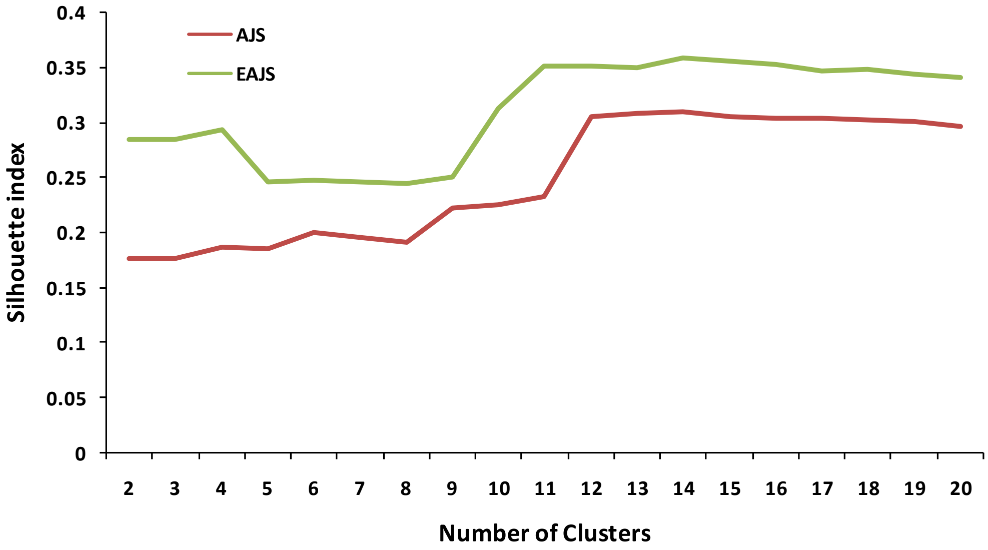


Figure 4.2 Silhouette index of 2 to 20clusters of species in a Serengeti food web based on Additive Jaccard Similarity (AJS) and Extended Additive Jaccard Similarity (EAJS)

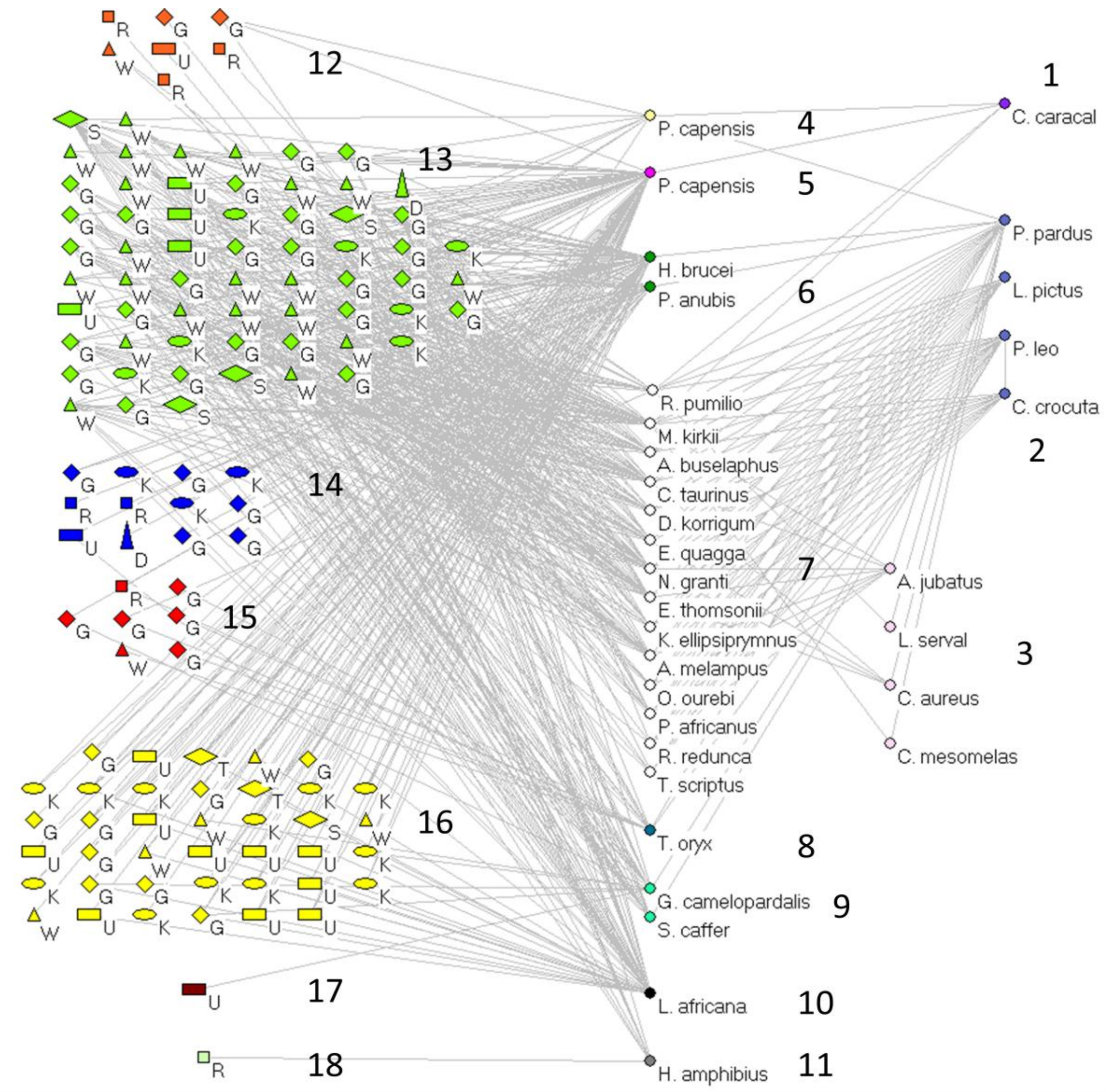


Figure 4.3 The Serengeti food web, with groups identified on the basis of Extended Additive Jaccard Similarity (EAJS) represented by different colors. Species are arranged by trophic level from plants (left) to herbivores (middle) to carnivores and omnivores (right). Different shapes indicate types of habitats of the plants including: (G)rassland, (W)oodland, (R)iparian, (K)opje, (S)hrubland, (T)hicket, (D)isturbed, (U)ndetermined.

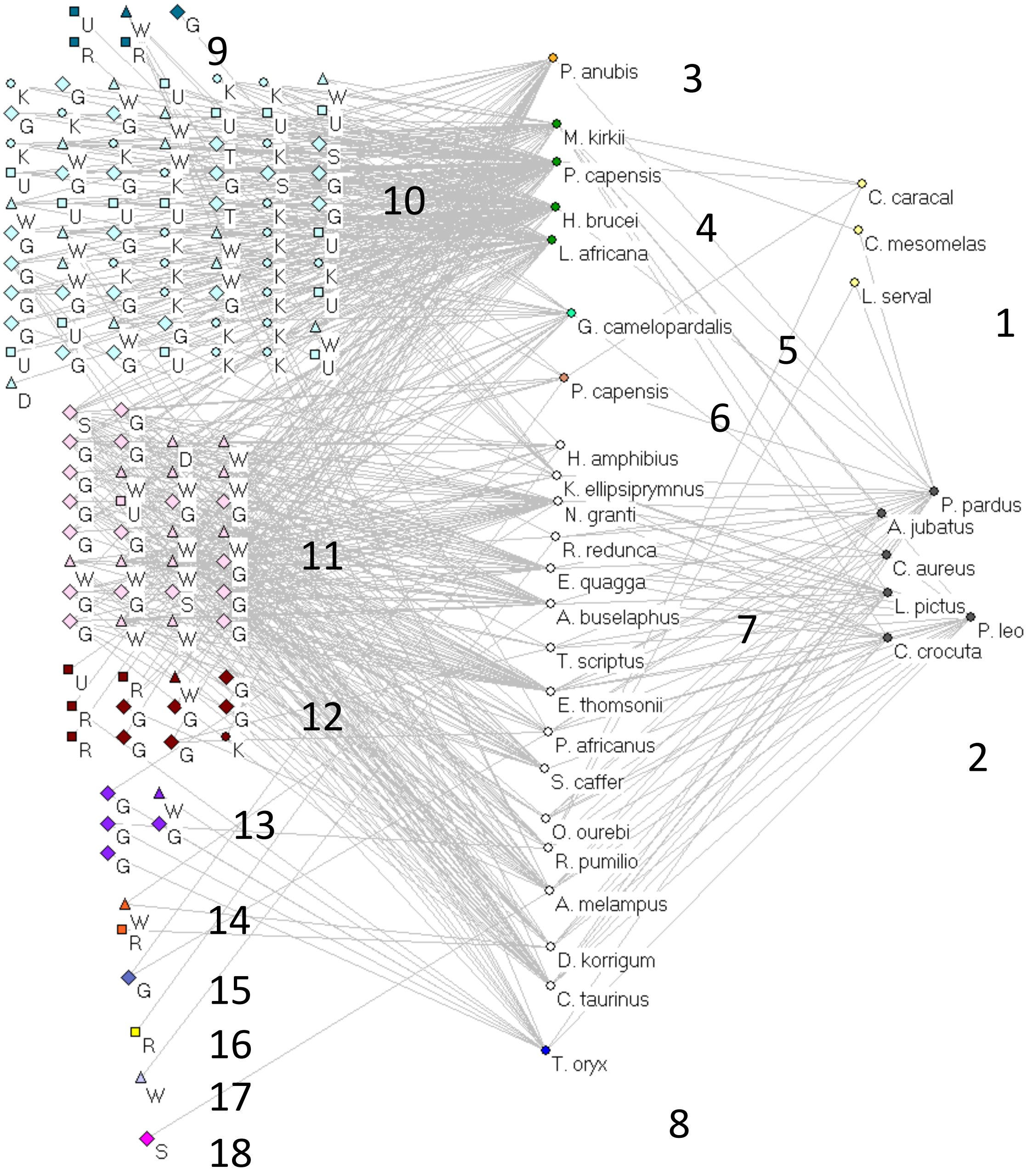


Figure 4.4 The Serengeti food web, with groups identified on the basis of Additive Jaccard Similarity (AJS) represented by different colors. Species are arranged by trophic level from plants (left) to herbivores (middle) to carnivores and omnivores (right). Different shapes indicate types of habitats of the plants including: (G)rassland, (W)oodland, (R)iparian, (K)opje, (S)hrubland, (T)hicket, (D)isturbed, (U)ndetermined.

Table 4.1 Species membership in eighteen groups identified on the basis of Extended Additive Jaccard Similarity (EAJS) for the Serengeti food web.

	Carnivores or Omnivores
group 1	*Caracal caracal*
group 2	*Crocuta crocuta, Lycaon pictus, Panthera leo, Panthera pardus*
group 3	*Acinonyx jubatus, Canis aureus, Canis mesomelas, Leptailurus serval*
	Herbivores
group 4	*Pedetes capensis*
group 5	*Procavia capensis*
group 6	*Heterohyrax brucei, Papio anubis*
group 7	*Aepyceros melampus, Alcelaphus buselaphus, Connochaetes taurinus, Damaliscus korrigum, Equus quagga, Nanger granti, Eudorcas thomsonii, Kobus ellipsiprymnus, Madoqua kirkii, Ourebia ourebi, Phacochoerus africanus, Redunca redunca, Tragelaphus scriptus, Rhabdomys pumilio*
group 8	*Taurotragus oryx*
group 9	*Giraffa camelopardalis, Syncerus caffer*
group 10	*Loxodonta Africana*
group 11	*Hippopotamus amphibious*
	Producers
group 12	*Andropogon schirensis, Cymbopogon excavatus, Digitaria ternata, Phragmites mauritianus, Psilolemma jaegeri, Sporobolus spicatus, Typha capensis*
group 13	*Acalypha fruticosa, Acacia senegal, Acacia tortilis, Achyranthes aspera, Allophylus rubifolius, Aloe macrosiphon, Andropogon greenwayi, Aristida spp., Balanites aegyptiaca, Boscia augustifolia, Bothriochloa insculpta, Brachiaria semiundulata, Capparis tomentosa, Pennisetum ciliare, Chloris gayana, Commelina africana, Commiphora trothae, Combretum molle, Cordia ovalis, Croton macrostachyus, Cynodon dactylon, Digitaria diagonalis, Digitaria macroblephara, Digitaria scalarum, Dinebra retroflexa, Duosperma kilimandscharica, Echinochloa haploclada, Eragrostis cilianensis, Eragrostis exasperata, Eragrostis tenuifolia, Eustachys paspaloides, Ficus glumosa, Grewia bicolor, Grewia trichocarpa, Harpachne schimperi, Heteropogon contortus, Hibiscus spp., Hibiscus lunariifolius, Hoslundia opposita, Hyperthelia dissoluta, Hyparrhenia filipendula, Hyparrhenia rufa, Indigofera basiflora, Indigofera hochstetteri, Kalanchoe spp., Maerua cafra, Microchloa kunthii, Ocimum spp., Panicum coloratum, Panicum maximum, Pennisetum mezianum, Pennisetum stramineum, Sansevieria ehrenbergii, Sida spp., Solanum dennekense, Solanum incanum, Solanum nigrum, Sporobolus festivus, Sporobolus fimbriatus, Sporobolus ioclados, Sporobolus pyramidalis, Themeda triandra*
group 14	*Acacia xanthophloea, Commiphora merkeri, Crotalaria spinosa, Digitaria velutina, Euphorbia candelabrum, Ficus thonningii, Heliotropium steudneri,*

	Kigelia africana, Lippia ukambensis, Sarga versicolor, Tricholaena teneriffae, Ziziphus spp.
group 15	*Acacia seyal, Chloris roxburghiana, Digitaria milanjiana, Lonchocarpus eriocalyx, Panicum deustum, Setaria pallide fusca, Setaria sphacelata*
group 16	*Abutilon spp., Acacia robusta, Albizia harveyi, Albuca spp., Aloe secundiflora, Blepharis acanthodioides, Chloris pycnothrix, Cissus quadrangularis, Cissus rotundifolia, Commiphora schimperi, Croton dichogamus, Cyperus spp., Cyphostemma spp., Diheteropogon amplectens, Emilia coccinea, Eragrostis aspera, Eriochloa nubica, Ficus ingens, Grewia fallax, Hypoestes forskaolii, Iboza spp., Ipomoea obscura, Jasminum spp., Kedrostis foetidissima, Kyllinga nervosa, Pappea capensis, Pavetta assimilis, Pavonia patens, Pellaea calomelanos, Phyllanthus sepialis, Pupalia lappacea, Rhoicissus revoilii, Sclerocarya birrea, Senna didymobotrya, Sansevieria suffruticosa, Sporobolus pellucidus, Sporobolus stapfianus, Turraea fischeri, Ximenia caffra*
group 17	*Olea spp.*
group 18	*Panicum repens*

Table 4.2 Species membership in eighteen groups identified on the basis of Additive Jaccard Similarity (AJS) for the Serengeti food web.

	Carnivores or Omnivores
Group 1	*Canis mesomelas, Caracal caracal, Leptailurus serval*
Group 2	*Acinonyx jubatus, Canis aureus, Crocuta crocuta, Lycaon pictus, Panthera leo, Panthera pardus*
	Herbivores
Group 3	*Papio anubis*
Group 4	*Heterohyrax brucei, Loxodonta africana, Madoqua kirkii, Procavia capensis*
Group 5	*Giraffa camelopardalis*
Group 6	*Pedetes capensis*
Group 7	*Aepyceros melampus, Alcelaphus buselaphus, Connochaetes taurinus, Damaliscus korrigum, Equus quagga, Nanger granti, Eudorcas thomsonii, Hippopotamus amphibius, Kobus ellipsiprymnus, Ourebia ourebi, Phacochoerus africanus, Redunca redunca, Rhabdomys pumilio, Syncerus caffer, Tragelaphus scriptus*
Group 8	*Taurotragus oryx*
	Producers
Group 9	*Andropogon schirensis, Chloris gayana, Cymbopogon excavatus, Phragmites mauritianus, Typha capensis*
Group 10	*Abutilon spp., Acalypha fruticosa, Acacia robusta, Acacia tortilis, Achyranthes aspera, Albizia harveyi, Albuca spp., Allophylus rubifolius, Aloe macrosiphon, Aloe secundiflora, Blepharis acanthodioides, Boscia augustifolia, Capparis tomentosa, Pennisetum ciliare, Chloris pycnothrix, Cissus quadrangularis, Cissus rotundifolia, Commelina africana, Commiphora merkeri, Combretum molle, Commiphora schimperi, Cordia ovalis, Croton dichogamus, Cyperus spp., Cyphostemma spp., Digitaria ternata, Digitaria velutina, Diheteropogon amplectens, Emilia coccinea, Eragrostis aspera, Eriochloa nubica, Ficus glumosa, Ficus ingens, Ficus thonningii, Grewia fallax, Grewia trichocarpa, Heliotropium steudneri, Hibiscus lunariifolius, Hoslundia opposita, Hypoestes forskaolii, Iboza spp., Indigofera basiflora, Ipomoea obscura, Jasminum spp., Kalanchoe spp., Kedrostis foetidissima, Kyllinga nervosa, Lippia ukambensis, Maerua cafra, Ocimum spp., Panicum maximum, Pappea capensis, Pavetta assimilis, Pavonia patens, Pellaea calomelanos, Pennisetum stramineum, Phyllanthus sepialis, Pupalia lappacea, Rhoicissus revoilii, Sclerocarya birrea, Senna didymobotrya, Sansevieria ehrenbergii, Sansevieria suffruticosa, Solanum dennekense, Solanum nigrum, Sporobolus pellucidus, Sporobolus stapfianus, Tricholaena teneriffae, Turraea fischeri, Ximenia caffra, Ziziphus spp.*
Group 11	*Andropogon greenwayi, Aristida spp., Balanites aegyptiaca,*

	Bothriochloa insculpta, Brachiaria semiundulata, Croton macrostachyus, Cynodon dactylon, Digitaria diagonalis, Digitaria macroblephara, Digitaria scalarum, Dinebra retroflexa, Eragrostis cilianensis, Eragrostis tenuifolia, Eustachys paspaloides, Grewia bicolor, Harpachne schimperi, Heteropogon contortus, Hibiscus spp., Hyparrhenia filipendula, Hyparrhenia rufa, Indigofera hochstetteri, Microchloa kunthii, Panicum coloratum, Pennisetum mezianum, Sida spp., Solanum incanum, Sporobolus fimbriatus, Sporobolus ioclados, Sporobolus pyramidalis, Themeda triandra,
Group 12	*Acacia senegal, Acacia seyal, Acacia xanthophloea, Commiphora trothae, Crotalaria spinosa, Digitaria milanjiana, Echinochloa haploclada, Euphorbia candelabrum, Kigelia africana, Olea spp., Panicum deustum, Sarga versicolor,*
Group 13	*Chloris roxburghiana, Duosperma kilimandscharica, Lonchocarpus eriocalyx, Setaria pallide fusca, Setaria sphacelata,*
Group 14	*Psilolemma jaegeri, Sporobolus spicatus*
Group 15	*Eragrostis exasperata*
Group 16	*Panicum repens*
Group 17	*Hyperthelia dissoluta*
Group 18	*Sporobolus festivus*

Chapter 5 Conclusion

Ecological networks exist in different forms and have been adapted to address different challenges and concerns in ecology. Ecological networks share same properties (e.g. complexity) as other networks, but they also have their own traits distinguishing themselves from other networks. In this dissertation, I focused on analyses that reduce the complexity and reveal the structure of ecological networks by decomposing them into groups of nodes or aggregating nodes into groups. I examined the performance and suitability of multiple techniques and algorithms when they were applied to different types of ecological networks focusing on particular goals in the studies presented in Chapters 2-4 of the dissertation (Table 1.1). One of the concerns of this dissertation is the fitness or suitability of these methods to ecological applications, especially for these methods that were not originally designed for the ecological studies.

5.1 Synthesis of Research Findings

In the study of decomposing the habitat network of *Lemur catta* (ring-tailed lemur) into compartments (Chapter 2), Graph-based REgionalization with Clustering And Partitioning (GraphRECAP) found compartments that had a higher modularity Harary Index, and Alpha Index than those detected by Girvan and Newman's method. Ecologically it suggested that these compartments had stronger within-compartment connections, greater traversability, and more alternative routes. Another desirable trait of these compartments is their relatively even size. In other words, compartments identified by GraphRECAP had a larger minimum number of habitat patches in compartments.

GraphRECAP is a modularity optimization approach for compartment detection, while Girvan and Newman's method works on removing edges that have high edge betweenness (intuitively it is a "bridge" cutting process) and does not directly optimize modularity. So it is expected that GraphRECAP achieved higher modularity. Fortunato (2010) and literature therein pointed out modularity optimization has low abilities to detect compartments that are comparatively small with the respect to the graph as a whole. In other words, modularity optimization tends to find compartments that have relatively even size. This weakness, from the view of computer science or mathematics, could be an “advantage” in this particular ecological application. Compartments generated by modularity optimization tend to be even sized. They could both enhance the resistance of the compartment to habitat loss and facilitate patch recolonization of local losses from within-compartment sources. In contrast, Girvan and Newman's method has been criticized because it may produce unbalanced partitions under certain circumstances (Chen and Yuan, 2006). Just as it turned out in this study, some of compartments found by Girvan and Newman's method contained small number of habitat patches. They are more vulnerable to local extinctions, due to lower effective population sizes, and have a lower chance of being rescued by outside immigration.

The trajectory analysis presented in Chapter 3 revealed patterns of animal movement by regionalizing animal trajectories such that animals had more movement within regions than across regions. It treated the animal movement tracked by radio telemetry as a spatial and ecological graph which was decomposed into groups of nodes for movement pattern recognition. These groups of nodes were spatially contiguous, forming regions. They were also ecologically-based in that animals had denser movement within regions compared to that across regions. Edge ratio-based Hierarchical Region Discovery (EHRD) was regarded as a better method than modularity-based Hierarchical Region Discovery (MHRD) in terms of finding more ecological-based regions that effectively captured the characteristics or traits of different species' movement. The clustering of animal trajectories based on regions detected by EHRD had higher similarity within clusters (e.g., all trajectories of cattle were grouped into one cluster). Classification of the trajectories (assuming we do not know what species that the trajectories represented) using attributes of trajectories derived from regions detected by EHRD achieved higher accuracy and simplicity.

The different ways that EHRD and MHRD partitioned the spatial graph of animal movement into regions lie in the designs of the two methods. To determine the strength of the connections via animal movement within regions, modularity compares the actual connections among nodes in graphs or networks with those in graphs (i.e., null models) where nodes are connected in random. It assumes that each node can be connected to other nodes in the null models. This assumption was invalid in this study. The movement of cattle was restricted by barbed-wire fences. Therefore, it was unreasonable to assume that cattle can move from the area within fenced off area to those outside of the fencing. It was not invalid either to compare the actual connections to connections in null models where nodes are assumed to be connected in a random manner. Edge ratio determines the strength of within-compartment connections by comparing the actual connections within the compartments and the actual connections between compartments. Therefore, edge ratio does not rely on any null model.

The tendency of modularity optimization in finding relatively even sized compartments as discussed above became a weakness in the trajectory analysis as illustrated in Figure 5.1. The most distinguishable and reasonable partition should be at the larger impedance in Figure 5.1a. However, modularity optimization finds the compartments with relatively even size and less ecologically meanings as shown in Figure 5.1c.

In the study of aggregating species in food webs according their trophic similarity (Chapter 4), the Extended Additive Jaccard Similarity (EAJS) was proposed to overcome the weakness of the Additive Jaccard Similarity (AJS) (i.e., the low ability to find species with equivalent trophic roles, if they do not share the same predators and prey). EAJS considers not only the similarity of shared predators and prey at adjacent trophic levels as AJS does, but at all the trophic levels. It turned out that EAJS succeeded in capturing species that have similar trophic roles (e.g. plants as producers) in the food web. The clusters of species on the basis of EAJS also exhibited patterns related to habitat structure of plants and network topology associated with animal weights. The advantage of EAJS lies in the fact that it considers species feeding relations in food webs at a broader scale.

5.2 Lessons learned

This dissertation examined two types of methods to reduce complexity of ecological networks. The first type of method decomposes the networks into compartments. The second type of method determines trophic similarity between pairs of species in food webs based on which clustering methods can be applied to aggregate species into groups. Among the three methods examined for compartment detection, the most widely used is the Girvan and Newman’s method which finds compartments by cutting the “bridges” connecting the compartments. The other two optimize modularity in the process of building the dendrogram of the nodes (clustering the nodes), but they differ in the criteria governing the process of partitioning the dendrogram (decomposing the graph). The performance of the three methods depends on their designs of the algorithms as well as the goals in different ecological applications.

A good example is the performance of modularity-based optimization approaches in the studies presented in Chapters 2 and 3. It is desirable to find relatively even-sized compartments in species habitat network for conservation purposes (Chapter 2). The modularity-based optimization approaches tend to find compartments that have relatively even size. Therefore, GraphRECAP is a better method for certain applications when compared to the Girvan and Newman’s method, which produced uneven partitions. However, there is no assumption that compartments should have relatively comparable numbers of nodes in the investigation of movement patterns of animals (Chapter 3). Therefore, partitioning the graphs based on edge ratio is a better option, because it does not have this assumption. Moreover, modularity compares the actual connections to those in a null model where nodes in the graph can connect to each other in a random manner. Although the comparison with the null model makes modularity more rigorous, it is not valid to assume that nodes in the graph of animal movement can connect to each other randomly. Thus, edge ratio has the advantage that it does not rely on such a null model.

Modularity-based optimization approaches may not be a better choice in other applications. The Girvan and Newman's method is better in terms of highlighting the linkages among habitats that are critical to the connectivity of entire habitat networks, because the Girvan and Newman's method identifies links with high edge centrality, that is, those edges that are most central and maintain the connectivity of entire habitat networks.

The Extended Additive Jaccard Similarity (EAJS) and the Additive Jaccard Similarity (AJS) are two methods to determine trophic similarity between species in food webs. Although EAJS succeeded in capturing species that have similar trophic roles (e.g. plants as producers) in the food web, it requires the feeding relations not only at adjacent trophic levels (i.e., the direct predators or prey of two species) but all the trophic levels. This method might be subjected to issues associated with data constraints. If only the feeding relations of the direct predators or prey are available, AJS is an efficient approach and probably the only choice to determine trophic similarity between species.

So to sum up, the answer to the question “Which method(s) is (are) better?” depends on the objectives of different studies and the constraints such as budget and data availability. One objective of this dissertation was thus to assess which method was better and why it is better for different goals in various ecological applications.

5.3 Future work

Future work will go in two directions. From an ecological perspective, more case studies need to be conducted to confirm the consistency of the performance of a method for a particular or specific goal. For example, it is interesting to apply the compartment detection methods to habitat networks of aquatic species and examine their performance and suitability. The increasing availability of data may also greatly enhance the algorithms and methods examined in this dissertation to better understand the processes in ecosystems. For example, the abundance of plants and animals in the Serengeti ecosystem will enable the estimation of the amount of energy flowing through the groups of species across the food web and the assessment of importance of each group to food web stability.

From a methodological perspective, the properties (e.g., the distribution of node degree) of these ecological networks, especially the spatial networks of animal movement proposed in this dissertation, need to be examined. The deep understanding of the properties or traits of the ecological networks and the capture of their uniqueness is a foundation for developing new methods customized to analyze these ecological networks and creating new criteria or indices for optimization or evaluation. For example, the Harary Index and Alpha Index were used to evaluate traversability and overall robustness to disturbance of the compartments. It might be possible to design a method to optimize these indices at the same time in compartment detection. New criteria that are more ecological network-oriented may also be created.

Applying methods such as those used in this dissertation in different ecological applications and exploring the linkages between the detected patterns and ecological processes yields insights into their performance, their ability to capture ecologically-meaningful patterns, and the their suitability and robustness to different situations. The feedbacks from these applications will provide valuable information and guidance on improving existing methods and developing ecologically-oriented indices and methods. This trial and error process will deepen insights on the research questions “which one is better for ecological applications” and “why it is better for ecological applications” addressed at the beginning of the dissertation.

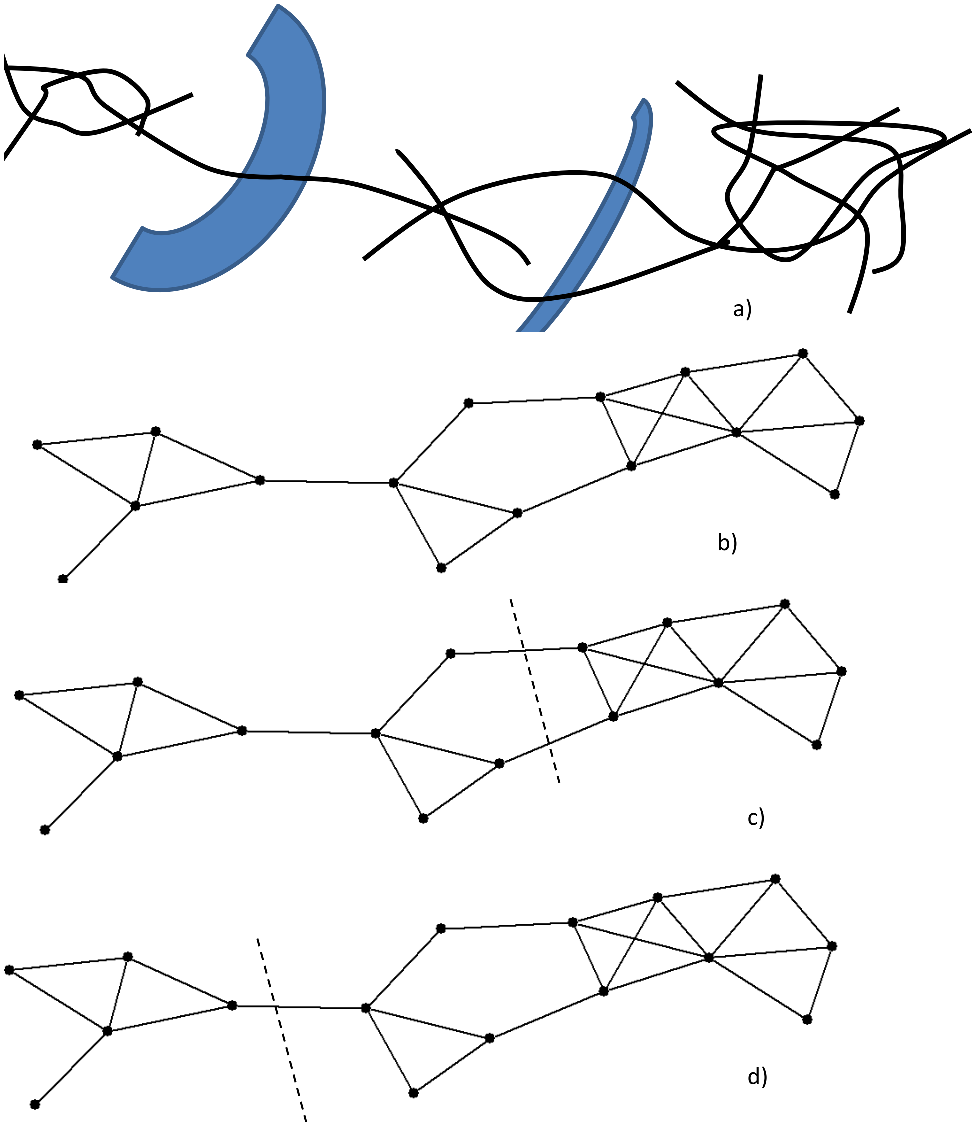


Figure 5.1 a) compartments in trajectories of animal movements caused by impedances (e.g., a large creek vs. a small creek); b) the conceptualized graph from a); c) and d) Partitioning of the graph (b) into two compartments by optimizing: c) modularity or d) edge ratio.

References

Aarts G, MacKenzie M, McConnell B, Fedak M, Matthiopoulos J (2008) Estimating space-use and habitat preference from wildlife telemetry data. Ecography 31(1):140-160

Albert R, Barabasi AL (2002) Statistical mechanics of complex networks. Rev Mod Phys 74(1):47-97

Allesina S, Pascual M (2009) Food web models: a plea for groups. Ecol Lett 12(7):652-662

Andrienko N, Andrienko G (2006) A Framework for Using Coordinated Displays for the Analysis of Multidimensional, Spatial, and Temporal Data. Geographic Hypermedia. Springer, pp. 293-308

Andrienko N, Andrienko G (2011) Spatial Generalization and Aggregation of Massive Movement Data. Ieee T Vis Comput Gr 17(2):205-219

Andrienko N, Andrienko G (2013) Visual analytics of movement: An overview of methods, tools and procedures. Info Vis 12(1):3-24

Arbelaitz O, Gurrutxaga I, Muguerza J, Perez JM, Perona I (2013) An extensive comparative study of cluster validity indices. Pattern Recogn 46(1):243-256

Bailey H, Thompson P (2006) Quantitative analysis of bottlenose dolphin movement patterns and their relationship with foraging. J Anim Ecol 75(2):456-465

Barthelemy M (2011) Spatial networks. Phys Rep 499(1-3):1-101

Baskerville EB, Dobson AP, Bedford T, Allesina S, Anderson TM, Pascual M (2011) Spatial Guilds in the Serengeti Food Web Revealed by a Bayesian Group Model. PLoS Comp Biol 7(12)

Bellisario B, Cerfolli F, Nascetti G (2010) Spatial network structure and robustness of detritus-based communities in a patchy environment. Ecol Res 25(4):813-821

Bennett DA, Tang W (2006) Modelling adaptive, spatially aware, and mobile agents: Elk migration in Yellowstone. Int J Geogr Inf Sci 20(9):1039-1066

Benton TG, Vickery JA, Wilson JD (2003) Farmland biodiversity: is habitat heterogeneity the key? Trends in Ecol Evol 18(4):182-188

Bodin Ö, Norberg J (2007) A network approach for analyzing spatially structured populations in fragmented landscape. Landsc Ecol 22:31-44

Bodin Ö, Tengö M, Norman A, Lundberg J, Elmqvist T (2006) The value of small size: Loss of forest patches and ecological thresholds in southern Madagascar. Ecol Appl 16(2):440-451

Borgatti SP, Everett MG, Freeman LC (2002) Ucinet forwindows: software for social network analysis.Analytic Technologies, Harvard

Brooks CP (2003) A scalar analysis of landscape connectivity. Oikos 102(2): 256-278

Brose U, Jonsson T, Berlow EL et al (2006a) Consumer-resource body-size relationships in natural food webs. Ecology 87(10):2411-2417

Brose U, Williams RJ, Martinez ND (2006b) Allometric scaling enhances stability in complex food webs. Ecol Lett 9(11):1228-1236

Calenge C, Dray S, Royer-Carenzi M (2009) The concept of animals' trajectories from a data analysis perspective. Ecol Inform 4(1):34-41

Camacho J, Guimera R, Amaral LAN (2002) Robust patterns in food web structure. Phys Rev Lett 88(22)

Casebeer R, Koss G (1970) Food habits of wildebeest, zebra, hartebeest and cattle in Kenya Masailand. Afr J Ecol 8(1):25-36

Chen J, Yuan B (2006) Detecting functional modules in the yeast protein-protein interaction network. Bioinformatics 22(18):2283-2290

Chen L, Özsu MT, Oria V Robust and fast similarity search for moving object trajectories. In: Proceedings of the 2005 ACM SIGMOD international conference on Management of data, 2005. ACM, p. 491-502

Cheriyadat AM, Radke RJ (2008) Detecting Dominant Motions in Dense Crowds. Ieee J Sel Top Signa 2(4):568-581

Chorley RJ, Haggett P (1967) Models in geography. Methuen, London

Clauset A, Newman MEJ, Moore C (2004) Finding community structure in very large networks. Phys Rev E 70(6)

Clobert J, Le Galliard JF, Cote J, Meylan S, Massot M (2009) Informed dispersal, heterogeneity in animal dispersal syndromes and the dynamics of spatially structured populations. Ecol Lett 12(3):197-209

Codling EA, Plank MJ, Benhamou S (2008) Random walk models in biology. J R Soc Interface 5(25):813-834

Coe PK, Johnson BK, Kern JW, Findholt SL, Kie JG, Wisdom MJ (2001) Responses of elk and mule deer to cattle in summer. J Range Manage 54(2):205-205

Cohen JE, Briand F, Newman CM (1990) Community food webs. Data and theory. (Biomathematics, vol 20) Springer, Berlin Heidelberg New York

Cohen JE, Pimm SL, Yodzis P, Saldana J (1993) Body sizes of animal predators and animal prey in food webs. J Anim Ecol 62(1):67-78

Cooper SM, Holekamp KE, Smale L (1999) A seasonal feast: long-term analysis of feeding behaviour in the spotted hyaena (Crocuta crocuta). Afr J Ecol 37(2):149-160

Cumming GS, Bodin O, Ernstson H, Elmqvist T (2010) Network analysis in conservation biogeography: challenges and opportunities. Divers Distrib 16(3):414-425

Dale MRT, Fortin MJ (2010) From Graphs to Spatial Graphs. In: Futuyma D. J., Shafer H. B., Simberloff D. (eds), Annual Review of Ecology, Evolution, and Systematics, Vol 41, Annual Review of Ecology Evolution and Systematics. pp. 21-38

De Nooy W, Mrvar A, Batagelj V (2012) Exploratory Social Network Analysis with Pajek, 2nd edn. Cambridge University Press, New York

De Ruiter PC, Wolters V, Moore JC (2005) Dynamic food webs: multispecies assemblages, ecosystem development and environmental change. Elsevier, Burlington

Devi BSS, Murthy MSR, Debnath B, Jha CS (2013) Forest patch connectivity diagnostics and prioritization using graph theory. Ecol Model 251:279-287

Dunn R, Dudbridge F, Sanderson CM (2005) The use of edge-betweenness clustering to investigate biological function in protein interaction networks. BMC Bioinformatics 6:1-14

Dunne JA, Williams RJ, Martinez ND (2002) Network structure and biodiversity loss in food webs: robustness increases with connectance. Ecol Lett 5(4):558-567

Economo EP, Keitt TH (2010) Network isolation and local diversity in neutral metacommunities. Oikos 119(8):1355-1363

Elton CS (1927) Animal ecology. Sidgwick and Jackson, London

Emmerson MC, Raffaelli D (2004) Predator-prey body size, interaction strength and the stability of a real food web. J Anim Ecol 73(3):399-409

Eros T, Olden JD, Schick RS, Schmera D, Fortin MJ (2012) Characterizing connectivity relationships in freshwaters using patch-based graphs. Landsc Ecol 27(2):303-317

Estrada E (2007) Characterization of topological keystone species local, global and "meso-scale" centralities in food webs. Ecol Complex 4(1-2):48-57

Foltête J-C, Clauzel C, Vuidel G (2012) A software tool dedicated to the modelling of landscape networks. Environ Modell Softw 38:316-327

Forman RTT 1995. Land mosaics: The ecology of landscapes and regions. Cambridge University Press, Cambridge

Fortuna MA, Bascompte J (2006) Habitat loss and the structure of plant-animal mutualistic networks. Ecol Lett 9(3):281-6

Fortunato S (2010) Community detection in graphs. Phys Rep 486(3-5):75-174

Fosca Giannotti MN, Fabio Pinelli, Dino Pedresch (2007) Trajectory pattern mining. International Conference on Knowledge Discovery and Data Mining, Proceedings of the 13th ACM SIGKDD international conference on Knowledge discovery and data mining:330 - 339

Frank KA (1995) Identifying cohesive subgroups. Soc Networks 17(1):27-56

Freeman LC (1977) Set of measures of centrality based on betweenness. Sociometry 40(1):35-41

Galpern P, Manseau M, Fall A (2011) Patch-based graphs of landscape connectivity: A guide to construction, analysis and application for conservation. Biol Conserv 144(1):44-55

Gilarranz LJ, Bascompte J (2012) Spatial network structure and metapopulation persistence. J Theor Biol 297: 11-16

Girvan M, Newman MEJ (2002) Community structure in social and biological networks. Proc Natl Acad Sci U S A 99(12):7821-7826

Girvan M, Newman MEJ (2002) Community structure in social and biological networks. Proc Natl Acad Sci USA 99(12):7821-7826

Glover F (1990) Tabu search - a tutorial. Interfaces 20(4):74-94

Guo D (2009) Flow Mapping and Multivariate Visualization of Large Spatial Interaction Data. Ieee T Vis Comput Gr 15(6):1041-1048

Guo, D., Gao P, and Zhu X. Trajectory Clustering and Classification with Hierarchical Region Detection, Int J Geogr Inf Sci, in development

Guo D, Jin H (2011) iRedistrict: Geovisual analytics for redistricting optimization. J Visual Lang Comput 22(4):279-289

Guo D, Liu S, Jin H (2010) A graph-based approach to vehicle trajectory analysis. Journal of Location Based Services 4(3-4):183-199

Guo D, Peuquet D, Gahegan M (2003) ICEAGE: Interactive Clustering and Exploration of Large and High-dimensional Geodata. GeoInformatica 7(3):229-253

Guo D, Zhu X, Jin H, Gao P, Andris C (2012) Discovering Spatial Patterns in Origin-Destination Mobility Data. Transactions in GIS 16(3):411-429

Hanski I (1997) Metapopulation dynamics from concepts and observations to predictive models. In: Hanski I, Gilpin M (eds) Metapopulation biology: ecology, genetics, and evolution. Academic Press, San Diego, pp 69-91

Hanski I, Gilpin M (1991) Metapopulation dynamics - brief-history and conceptual domain. Biol J Linnean Soc 42(1-2):3-16

Havens K (1992) Scale and structure in natural food webs. Science 257(5073):1107-1109

Holdo RM, Roach RR (2013) Inferring animal population distributions from individual tracking data: theoretical insights and potential pitfalls. J Anim Ecol 82(1):175-181

Holvorcem CGD, Tambosi LR, Ribeiro MC, Costa S, Bernardo Mesquita CA (2011) Anchor areas to improve conservation and increase connectivity within the Brazilian "Mesopotamia of Biodiversity". Natureza & Conservacao 9(2):225-231

Huttenlocher DP, Klanderman GA, Rucklidge WJ (1993) Comparing images using the hausdorff distance. Ieee T Pattern Anal 15(9):850-863

Jonsen ID, Flenming JM, Myers RA (2005) Robust state-space modeling of animal movement data. Ecology 86(11):2874-2880

Jordán F (2009) Keystone species and food webs. Philos T R Soc B 364(1524):1733-1741

Jordán F, Baldi A, Orci KM, Racz I, Varga Z (2003) Characterizing the importance of habitat patches and corridors in maintaining the landscape connectivity of a *Pholidoptera transsylvanica* (Orthoptera) metapopulation. Landsc Ecol 18(1):83-92

Kerr JT, Deguise I (2004) Habitat loss and the limits to endangered species recovery. Ecol Lett 7(12):1163-1169

Krause AE, Frank KA, Jones ML et al (2009) Adaptations in a hierarchical food web of southeastern Lake Michigan. Ecol Model 220(22): 3147-3162

Krause AE, Frank KA, Mason DM, Ulanowicz RE, Taylor WW (2003) Compartments revealed in food-web structure. Nature 426(6964): 282-285

Kupfer JA (1995) Landscape ecology and biogeography. Prog Phys Geogr 19(1):18-34

Kupfer JA (2012) Landscape ecology and biogeography: Rethinking landscape metrics in a post-FRAGSTATS landscape. Prog Phys Geogr 36(3):400-420

Kupfer JA, Gao P, Guo D (2012) Regionalization of forest pattern metrics for the continental United States using contiguity constrained clustering and partitioning. Ecol Inform 9:11-18

Kwan MP (2000) Interactive geovisualization of activity-travel patterns using three-dimensional geographical information systems: a methodological exploration with a large data set. Transport Res C-Emer 8(1-6):185-203

Laita A, Mönkkönen M, Kotiaho JS (2010) Woodland key habitats evaluated as part of a functional reserve network. Biol Conserv 143(5):1212-1227

Lee J-G, Han J, Li X, Gonzalez H (2008) TraClass: trajectory classification using hierarchical region-based and trajectory-based clustering. Proceedings of the VLDB Endowment 1(1):1081-1094

Long JA, Nelson TA (2013) A review of quantitative methods for movement data. Int J Geogr Inf Sci 27(2):292-318

Lookingbill TR, Gardner RH, Ferrari JR, Keller CE (2010) Combining a dispersal model with network theory to assess habitat connectivity. Ecol Appl 20(2):427-441

Luczkovich JJ, Borgatti SP, Johnson JC, Everett MG (2003) Defining and measuring trophic role similarity in food webs using regular equivalence. J Theor Biol 220(3):303-321

May RM (1972) Will a large complex system be stable. Nature 238(5364):413-414

McIntyre NE, Strauss RE (2013) A new, multi-scaled graph visualization approach: an example within the playa wetland network of the Great Plains. Landsc Ecol 28(4):769-782

McNaughton SJ (1978) Stability and diversity of ecological communities. Nature 274(5668):251-253

Melian CJ, Bascompte J (2002) Complex networks: two ways to be robust? Ecol Lett 5(6):705-708

Mertl-Millhollen AS, Blumenfeld-Jones K, Raharison SM, Tsaramanana DR, Rasamimanana H (2011) Tamarind tree seed dispersal by ring-tailed lemurs. Primates 52(4):391-396

Miller JA (2012) Using Spatially Explicit Simulated Data to Analyze Animal Interactions: A Case Study with Brown Hyenas in Northern Botswana. Transactions in GIS 16(3):271-291

Minor ES, Urban DL (2008) A graph-theory frarmework for evaluating landscape connectivity and conservation planning. Conserv Biol 22(2):297-307

Minor ES, Urban DL (2008) A graph-theory frarmework for evaluating landscape connectivity and conservation planning. Conserv Biol 22(2):297-307

Morales JM, Ellner SP (2002) Scaling up animal movements in heterogeneous landscapes: The importance of behavior. Ecology 83(8):2240-2247

Morales JM, Haydon DT, Frair J, Holsiner KE, Fryxell JM (2004) Extracting more out of relocation data: Building movement models as mixtures of random walks. Ecology 85(9):2436-2445

Nathan R, Getz WM, Revilla E et al (2008) A movement ecology paradigm for unifying organismal movement research. Proc Natl Acad Sci U S A 105(49):19052-19059

Newman M, Barabási A-L, Watts DJ (2006) The structure and dynamics of networks. Princeton University Press, Princeton

Newman MEJ (2006) Modularity and community structure in networks. Proc Natl Acad SciUSA 103(23):8577–8582

Newman MEJ, Girvan M (2004) Finding and evaluating community structure in networks. Phys Rev E 69(2): 1-15

Norberg J, Cumming GS (2008) Complexity theory for a sustainable future. Columbia University Press, New York

O'Brien D, Manseau M, Fall A, Fortin MJ (2006) Testing the importance of spatial configuration of winter habitat for woodland caribou: An application of graph theory. Biol Conserv 130(1):70-83

Ono N, Fujiwara Y, Yuta K (2005) Artificial metabolic system: An evolutionary model for community organization in metabolic networks. In: Capcarrère MS, Freitas AA, Bentley PJ, Johnson CG, Timmis J (eds) Advances in Artificial Life, 8th European Conference, ECAL 2005, Canterbury, U.K., September 5-9, 2005, Proceedings, volume 3630 of Lecture Notes of Computer Science, Springer, Berlin, p 716-724

Osei-Bryson KM (2004) Evaluation of decision trees: a multi-criteria approach. Comput Oper Res 31(11):1933-1945

Paine RT (1963) Trophic relationships of 8 sympatric predatory gastropods. Ecology 44(1):63-73

Paine RT (1980) Food webs - linkage, interaction strength and community infrastructure - the 3rd tansley lecture. J Anim Ecol 49(3):667-685

Patterson TA, Thomas L, Wilcox C, Ovaskainen O, Matthiopoulos J (2008) State-space models of individual animal movement. Trends Ecol Evol 23(2):87-94

Petchey OL, Beckerman AP, Riede JO, Warren PH (2008) Size, foraging, and food web structure. Proc Natl Acad Sci U S A 105(11):4191-4196

Pimm SL (1979) Structure of food webs. Theor Popul Biol 16(2):144-158

Pimm SL (1982) Food webs. Chapman and Hall, London

Polis GA (1991) Complex trophic interactions in deserts - an empirical critique of food-web theory. Am Nat 138(1):123-155

Raffaelli D, Hall SJ (1992) Compartments and predation in an estuarine food web. J. Anim. Ecol. 61(3):551-560

Rathore CS, Dubey Y, Shrivastava A, Pathak P, Patil V (2012) Opportunities of Habitat Connectivity for Tiger (Panthera tigris) between Kanha and Pench National Parks in Madhya Pradesh, India. PLoS ONE 7(7):13

Rayfield B, Fortin MJ, Fall A (2011) Connectivity for conservation: a framework to classify network measures. Ecology 92(4):847-858

Reunanen P, Fall A, Nikula A (2012) Spatial graphs as templates for habitat networks in boreal landscapes. Biodivers Conserv 21(14):3569-3584

Rezende EL, Albert EM, Fortuna MA, Bascompte J (2009) Compartments in a marine food web associated with phylogeny, body mass, and habitat structure. Ecol Lett 12(8):779-788

Ricotta CA, Stanisci A, Avena GC, Blasi C (2000) Quantifying the network connectivity of landscape mosaics a graph theoretical approach. Community Ecol 1(1): 89–94.

Riede JO, Brose U, Ebenman B et al (2011) Stepping in Elton's footprints: a general scaling model for body masses and trophic levels across ecosystems. Ecol Lett 14(2):169-178

Roberts A (1951) The mammals of South Africa. Trustees ‘The mammals of South Africa book fund’, Johannesburg

Rousseeuw PJ (1987) Silhouettes - a graphical aid to the interpretation and validation of cluster-analysis. J Comput Appl Math 20:53-65

Rowland MM, Coe PK, Stussy RJ, Ager AA, Cimon NJ, Johnson BK, and Wisdom MJ (1998) The Starkey habitat database for ungulate research: construction, documentation, and use. US Department of Agriculture, Forest Service, Pacific Northwest Research Station Portland, OR, USA

Rubio L, Saura S (2012) Assessing the importance of individual habitat patches as irreplaceable connecting elements: An analysis of simulated and real landscape data. Ecol Complex 11:28-37

Rulequest Research. (2011) See5/c5.0. [Online]. Available: http://www.rulequest.com/see5-info.htm

Safavian SR, Landgrebe D (1991) A survey of decision tree classifier methodology. Ieee T Syst Man Cyb 21(3):660-674

Saura S, Rubio L (2010) A common currency for the different ways in which patches and links can contribute to habitat availability and connectivity in the landscape. Ecography 33(3):523-537

Schick RS, Loarie SR, Colchero F et al (2008) Understanding movement data and movement processes: current and emerging directions. Ecol Lett 11(12):1338-1350

Schnettler S (2009) A structured overview of 50 years of small-world research. Soc Networks 31(3):165-178

Sinclair ARE, Mduma S, Brashares JS (2003) Patterns of predation in a diverse predator-prey system. Nature 425(6955): 288-290

Sinclair ARE, Norton-Griffiths M (1984) Serengeti: dynamics of an ecosystem. University of Chicago Press, Chicago

Stanisc A, Avena GC, Blasi C (2000). Quantifying the network connectivity of landscape mosaics a graph theoretical approach. Community Ecol 1: 89–94

Stouffer DB, Camacho J, Guimera R, Ng CA, Amaral LAN (2005) Quantitative patterns in the structure of model and empirical food webs. Ecology 86(5):1301-1311

Strogatz SH (2001) Exploring complex networks. Nature 410(6825):268-276

Tang W, Wang S, Bennett DA, Liu Y (2011) Agent-based modeling within a cyberinfrastructure environment: a service-oriented computing approach. Int J Geogr Inf Sci 25(9):1323-1346

Teng J, McCann KS (2004) Dynamics of compartmented and reticulate food webs in relation to energetic flows. Am Nat 164(1):85-100

Theobald DM, Reed SE, Fields K, Soulê M (2012) Connecting natural landscapes using a landscape permeability model to prioritize conservation activities in the United States. Conserv Lett 5(2):123-133

Tsai VJD (1993) Delaunay triangulations in TIN creation - an overview and a linear-time algorithm. Int J Geogr Inf Sci 7(6):501-524

Urban D, Keitt T (2001) Landscape connectivity: A graph-theoretic perspective. Ecology 82(5):1205-1218

Urban DL, Minor ES, Treml EA, Schick RS (2009) Graph models of habitat mosaics. Ecol Lett 12(3):260-273

Usabiaga J, Bebis G, Erol A, Nicolescu M (2007) Recognizing simple human actions using 3D head movement. Comput Intell 23(4):484-496

Vergara PM, Perez-Hernandez CG, Hahn IJ, Soto GE (2013) Deforestation in central Chile causes a rapid decline in landscape connectivity for a forest specialist bird species. Ecol Res 28(3):481-492

Verhein F, Chawla S (2008) Mining spatio-temporal patterns in object mobility databases. Data Min Knowl Disc 16(1):5-38

Webb C, Bodin Ö (2008) A network perspective on modularity and control of flow in robust systems. In: Norberg and Cumming (ed) Complexity theory for a sustainable future. Columbia University Press, West Sussex

Williams RJ, Berlow EL, Dunne JA, Barabasi AL, Martinez ND (2002) Two degrees of separation in complex food webs. Proc Natl Acad Sci U S A 99(20):12913-12916

Williams RJ, Martinez ND (2000) Simple rules yield complex food webs. Nature 404(6774):180-183

Yodzis P, Winemiller KO (1999) In search of operational trophospecies in a tropical aquatic food web. Oikos 87(2):327-340

Zhang Z, Huang K, Tan T Comparison of similarity measures for trajectory clustering in outdoor surveillance scenes. In: Pattern Recognition, 2006. ICPR 2006. 18th International Conference on, 2006. vol 3. Ieee, p. 1135-1138

Ziolkowska E, Ostapowicz K, Kuemmerle T, Perzanowski K, Radeloff VC, Kozak J (2012) Potential habitat connectivity of European bison (Bison bonasus) in the Carpathians. Biol Conserv 146(1):188-196