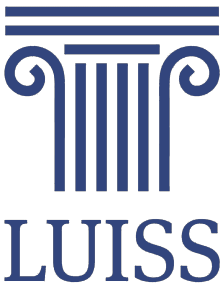
ID N. 37



UNIVERSITÀ CAMPUS BIO-MEDICO DI ROMA

DEPARTMENT OF ENGINEERING

LUISS GUIDO CARLI

DIPARTIMENTO DI IMPRESA E MANAGEMENT

Italian National Ph.D. in Artificial Intelligence
Health and Life Sciences
XXXVII Cycle

Exploring biological network structures: aquatic food webs and contact networks

Supervisors

Giuseppe F. Italiano

Blerina Sinaimeri

Candidate

Davide Torre

November, 2024

To Jun, with my unconditional love.

Acknowledgements

I am now sitting in my office in Turin at the ISI Foundation, writing what will be the last paragraph I compose for this thesis, yet the first that readers will encounter. I know that these three years have been transformative for both my future career and my personal development, and they would not have been the same without the people around me who chose to invest their time and energy to guide, support, and sustain me throughout this journey.

First of all, I want to thank my supervisors, Giuseppe F. Italiano and Blerina Sinaimeri. It is thanks to you that I have learned, in these three years, not to remain on the surface with respect to a research question, but to deepen into it, and to "think about it longer". I still have so much to learn, but the fact remains that every interaction we have had has been a reason for growth, even when it happened that I didn't accept it gracefully, and I apologize for this. The long time you have dedicated to me is now not only part of my personal baggage, but is also a supporting pillar on which my future research will be based, and I thank you from the bottom of my heart for the time, patience, opportunities, and care you have shown towards me.

A special thanks also goes to Stefano Guarino and Enrico Mastrostefano from the Institute for Applied Mathematics "Mauro Picone" (IAC) of the CNR. Not only for introducing me to research, but also for the time we spent together studying the generative model. I already carry in my heart the afternoons spent going crazy on the blackboard of the first floor of the IAC and the meetings on the third floor with the big screen. For this reason, I also thank Massimo Bernaschi for having accompanied me and made me passionate during my very first year of research, in 2020.

I thank the Campus Bio-Medico of Rome for having organized and administered this national doctorate in Artificial Intelligence, and LUISS for having made available not only the scholarship, but also the facilities where I have mainly carried out my research activity during these years.

I would like to thank Davide Chicco for his availability and collaboration, which has blossomed into a wonderful partnership.

I am grateful to Professor Robert Ulanowicz for answering my questions about food webs and for encouraging me, through his book, to view these networks from a more holistic perspective. I must say that I have so much more to learn!

I thank my friends. I thank Giovanni for his closeness throughout all these years, from our undergraduate degree in physics until today. You have always been by my side and our friendship is precious to me. You have always known how to encourage me when my morale was down, to give me a piece of your mind when it was necessary, and, always, you have been able to give me a different perspective on the things that happened to me. Giovanni, you are the demonstration of what is said in the book of Sirach: "A faithful friend is a precious support, and whoever finds one has found a treasure. A faithful friend is priceless and his value is incalculable." (Sir: 6,15-16)

I thank Francesco for his closeness during the most difficult moments of my life and for his lightheartedness and cheerfulness. Together we have changed so much since we met at WYD in Poland in 2016, we have gone through many different phases, from the super devout ones to finally realizing that God exists, you just shouldn't take him too seriously (just as you shouldn't take Christians and priests too seriously).

I thank my wonderful roommates from Via Calpurnio Fiamma. Erica, you are a wonderful person, with whom I shared so much silence and with whom I felt comfortable even without having to talk too much. Thank you for all the beautiful moments of these 7 years. Marilisa, you are an exceptional girl and woman and I really hope that more people around you realize this. I have seen you change so much in these years and you are living proof of the results that come from truly putting yourself out there. I am so grateful for all the laughter, the Simpsons episodes and references, and for freaking out together and face difficulties with a smile. Fabrizia, when you first arrived at the house you gave me the impression that you had your head in the clouds. Incredibly, this impression hasn't changed. The world needs so much lightness. The world needs people like you. Raffa, thank you for your determination and your ability to tell things as they are. Know that for me you are a great example. Thank you to Fabrizio Colucci, who is able to nourish our friendship with so much love and constancy. Thank you to Domenico Ferri from UJR, for his professionalism both as a mechanic and as a man, of the world!

I thank my family, especially my parents for having supported me and for always supporting me in this journey and for being close to me. I thank all my brothers and sisters (we Torres are a small army) for their example, which is very important for me in giving the right direction to my life. I thank my grandparents, the ones in the sky, and one down here.

Thanks to Debaish, Sakis, Manas, Anna, Fariba, Miro and Cosimo, because, in different ways and times, they have made this PhD better than it could have been. Thanks also to my

colleague from the 38th cycle Riccardo Sabbadini and to Gabriele Palma, first an excellent "student" during my time as Teaching Assistant, now an excellent colleague. Thanks to Fabio Angeletti, first a colleague in two Teaching Assistant adventures, then an excellent friend! I wish you only the best!

Thanks to Weng LaoShi for her support and professionalism throughout the Chinese course. Your dedication and expertise are truly precious and rare qualities, and I sincerely hope they will carry you far in all your endeavors.

Thanks to the fantastic PhD students of Via Lisbona, you are unique and the most resilient people I have ever met. Thanks to my colleagues from the 37th cycle. We have come a long way together, through many difficulties, but we made it. Thanks to Rosa, Mathematics professor at Einaudi in Rome, who never failed to provide me with her support and beautiful words of encouragement.

Thanks to the "Coro delle Stimmate" for being there and for living the passion of music together. In particular, thanks to Modestina, Alessandro, Andrea, Patrizia, Alba, Davide and Valeria, Federico and Mariateresa, Beatrice. Thanks to Enrico Petrillo, as we can still meet each other every year in Assisi, in June. Thanks to Ciro and Annarita and to the small group of the Sette Segni. Even though the course has ended and we see each other less, the relationships between us have not dimmed and still bind us so much! Thanks to Don Fabio.

Thanks to Don Carlo Russo, Salesian of the Don Bosco parish in Rome, who was my confessor during these years of PhD and welcomed my fragilities, weaknesses and moments of darkness, managing to guide me with calm and constancy. Thank you to the "young" choir of Don Bosco for these beautiful years we spent together during the PhD, I will carry so many moments in my heart, from the beautiful and light homilies of Cardinal Sarah, who didn't become pope and we're very sorry about that, to the masses for family days in the Vatican. Of these latter ones, for some unknown reason, we all mostly remember the moments after animating the masses, right? Thank you to Veronica Piscitelli, director of this choir who pushes us to bring out our voice and the best in us. Thank you to our legendary organist, Alessandro Capitani.

Thank you to MaiRong, ManQing and Li and NaiNai for welcoming me into their family, with so much joy and so much love.

And lastly, to Jun. Thank you for all the love you have given me, my beloved JuJu. I owe this thesis and this result mainly to you, to your unmatched standards in research, to your unfailing enthusiasm. In these years we have gotten to know each other, we have laughed, cried and, above all, we have loved each other, building, one step at a time, a path for a future together, a path so beautiful and so full of hope that it leaves me breathless every time I think about it. I love you with all my heart!

Abstract

Complex networks are graphs that represent systems of interconnected components, distinguished by non-trivial structural features that set them apart from regular or random graphs. They model systems by assigning nodes to each entity of the system, connected by their interactions (edges), and usually exhibiting emergent patterns and properties that are not easily inferred from individual components. Two important types of networks in biology are food webs and contact networks, with the former modelling "who eats whom" in ecosystems and the latter modelling "who meets whom" in populations.

Food webs are a specific type of complex networks that model energy and matter exchange in ecosystems. Understanding the patterns of these networks is important for identifying key species, assessing ecosystem health and stability, and predicting the effects of environmental disturbance. Contact networks represent one of the most important frameworks in fields such as epidemiology, where the particular structural properties of these networks have a significant impact on the dynamics of disease spread, making realistic models of these networks necessary to predict and control these dynamics.

The objective of this thesis is twofold: first, to measure and analyze the patterns of aquatic food webs, aiming to gain deeper insights into their structural and functional properties; second, to propose a network model to incorporate group mixing for constructing more realistic contact networks.

Contents

List of Figures

List of Tables

List of Algorithms

List of Symbols

G	graph representing a network of interactions
(V, E)	set of vertices (V) and edges (E) defining a graph
S	set of nodes in a food web
L	set of directed edges in a food web
C	connectance of a food web, defined as L/S^2 Equation 3.1
$\mathbb{S}^d$	d-sphere manifold
$\mathbb{H}^{d+1}$	$d+1$-hyperbolic manifold
$deg(v)$ or deg_v	the degree of a node v
$deg_{in}(v)$	the in-degree of a node v
$deg_{out}(v)$	the out-degree of a node v
TL_i	the trophic level of species i in a food web, as defined in Equation 3.2
a	binary interaction matrix of a food web
w	weighted interaction matrix of a food web
K	The edges between communities matrix. The element K_{IJ} represents the number of edges between communities I and J when $I \neq J$, and twice that number when $I = J$.
M	the expected total number of edges
Δ	the mixing matrix defined by $K_{IJ} = M\Delta_{IJ}$.

Chapter 1

Introduction

The purpose of this chapter is to introduce the reader to the problems addressed in this thesis, their significance and applications, the state of the art, the contributions of this work.

The main focus of this thesis is to analyze the structural organization of *complex networks* in ecological and epidemiological contexts, while also introducing a novel modeling framework to address challenges in network generation. *Complex networks* are graphs with non-trivial topological features and emergent properties, such as heterogeneous degree distributions, high clustering coefficients, community structures, and small-world effects. These characteristics distinguish them from regular or random graphs and are fundamental to understanding the dynamics and behaviors of many real-world systems [145, 152, 1].

In the study of *food webs*, this thesis aims to investigate the structural properties that govern aquatic ecosystems. Food webs are networks where nodes represent species and pools of organic matter and edges signify trophic interactions—essentially, "who eats whom" [190]. Although these networks are important for understanding ecosystem organization and resilience, existing datasets for aquatic food webs are often confined to a small number of ecosystem models. This scarcity limits systematic analyses, undermining understanding of ecosystem dynamics and resilience. Addressing these gaps is needed for multiple reasons: analyzing food web structures offers insight into global structural patterns common across ecosystems [34] and informs assessments of ecosystem resilience and responses to perturbations [66, 67], particularly in the face of global challenges such as climate change and biodiversity loss [135, 55].

In the domain of *contact networks*, this thesis addresses the challenge of defining realistic network models that capture known structural properties of human interactions. Contact networks, which represent physical proximity such as face-to-face interactions [14], provide a framework for understanding disease transmission dynamics. However, collecting compre-

hensive data on large-scale contact networks remains challenging: while surveys and wearable sensors provide valuable information, their utility is often constrained by scale, noise, and high costs [68]. To address these limitations, realistic models are important for capturing the characteristics of contact networks that influence epidemic dynamics—such as heterogeneous degree distributions, high clustering coefficients, and community structure. These models enable accurate simulations of outbreak scenarios and support effective public health interventions.

This thesis makes two contributions. First, it provides a detailed analysis of the structural features of aquatic food webs by analyzing 173 networks, including their organization, critical nodes, and three-species interaction patterns. Second, it proposes a synthetic modeling framework for contact networks that overcomes limitations in existing approaches, generating realistic graphs that align with observed social interaction patterns. The source code and datasets employed for both the food web analysis and synthetic network generation of this research are publicly accessible via two dedicated GitHub repositories.

1.1 Problem Definition and Motivation

Despite their important role in understanding ecosystem stability and biodiversity, the detailed structural characteristics of aquatic food webs are still underexplored. In this thesis, we focus on three interrelated challenges: analyzing core-periphery structures, identifying critical nodes, and examining the representation patterns of three-node motifs. The core-periphery structure is the partition of a network into two distinct sets: a densely connected core and a sparsely connected periphery, it reveals species' distinct roles, nutrient transfer pathways, and has been linked to ecosystem stability and persistence in the context of other ecological networks, such as pollinators [139, 119]. Critical nodes are the species whose removal severely disrupts network connectivity [154]. Their identification is essential for maintaining ecosystem resilience, especially in the face of threats such as habitat destruction, invasive species, climate change, and pollution [196]. Three-node motifs, which represent recurring patterns of interactions between three species [140], bridge the gap between pairwise connections and the entire network, offering a meso-scale perspective that complements controlled experimental studies [188]. Understanding these structural features is vital for assessing ecosystem resilience, informing conservation strategies, and addressing global challenges like climate change and biodiversity loss.

The second problem involves the generation of realistic synthetic contact networks. Contact networks, which model face-to-face interactions between individuals, are critical for understanding the dynamics of disease transmission [14, 185]. However, obtaining large-

scale contact network data remains challenging due to cost, privacy concerns, and noise in data collection methods such as surveys and wearable sensors [68]. This thesis addresses this limitation by proposing a random network model capable of incorporating three controllable topological properties: degree distribution, clustering, and community structure. Degree distribution reflects the variation in individual interactions, with high-degree nodes acting as potential super-spreaders in epidemic scenarios [153]. Clustering captures the localized nature of social interactions, influencing how diseases spread within tightly connected groups [28]. Community structure mirrors real-world stratifications, such as age or geographic groups observed in urban populations, and determines how outbreaks propagate within and across subpopulations [84]. These features are essential for simulating realistic outbreak scenarios, guiding public health interventions and supporting broader applications, such as in sociology, transportation, and modeling interaction networks like needle-sharing or sexual contacts [88]. By integrating these topological features into a single framework, the proposed model overcomes the limitations of existing approaches, providing a versatile tool for studying diverse network dynamics.

Both problems addressed in this thesis are united by their importance in tackling contemporary challenges. The analysis of aquatic food webs aids decision-makers in implementing science-based conservation actions, ensuring the sustainability of ecosystems that support human and ecological well-being. Similarly, realistic synthetic contact networks enable accurate simulations of epidemic outbreaks, informing interventions to mitigate their impacts and contributing to broader applications in network science.

1.2 Previous Research and Limitations

1.2.1 Aquatic Food Webs

Structural properties of food webs

Research on food web topology has revealed fundamental universal patterns, including a quantifiable scaling relationship between network connectance and species count, and a logarithmic correlation between network diameter and species abundance [65, 66, 34, 198, 122]. These empirical regularities suggest the existence of common organizational principles underlying ecosystem structure, providing a theoretical foundation for the development of robust models of trophic interactions.

Food webs exhibit a broad range of connectance values, defined as the fraction of all possible links that are realized (L/S^2). Typically it ranges between 0.03 and 0.3 across

different communities [18]. Higher connectance tends to correlate with greater stability, as densely interconnected webs are more robust to species loss as demonstrated by Jennifer Dunne and colleagues [66]. Many food webs exhibit small-world characteristics, meaning that most species are reachable from any other species through a small number of intermediate connections. One of these characteristics is short path lengths, which facilitate quick energy or nutrient transfer between species. However, clustering coefficients tend to be low, which may result from the limited size of these networks [206]. Unlike other networks, food webs generally do not display a power-law degree distribution and instead show an exponential or uniform distribution [214]. This pattern is attributed to the limited size and higher connectance of food webs, preventing the highly skewed distributions often seen in larger, sparser networks [206].

In addition to food webs, the analysis of other ecological networks has been of great interest to biologists in the past. For example it has been shown that the dynamics within ecological networks can exhibit core-periphery structures, as seen in plant-pollinator systems [139, 51, 179]. The *core-periphery structure* of an ecological network, which partitions the network into a densely connected core and a sparsely connected periphery, has been widely recognized as a framework for understanding species roles and ecosystem dynamics [51]. This structure highlights the functional differentiation within ecosystems, highlighting the roles species play in maintaining stability and facilitating nutrient transfer [216, 15, 170]. Miele et al. [139] explored this concept in a plant-pollinator network, showing that core species—those with many interactions—play a critical role in maintaining network cohesion, while peripheral species provide functional redundancy and adaptability. Such structures ensure that networks remain robust even when subjected to species loss or environmental changes.

Although core-periphery structures have been thoroughly investigated in mutualistic and pollinator networks, their application and analysis in food web contexts remain understudied. This represents a significant research gap, as these structural patterns have demonstrated considerable importance for understanding network architecture and the dynamics of stability within food web ecosystems.

Robustness of food webs due to extinction and critical nodes

The structure of a food web provides critical insights into its robustness against species loss [66, 65]. *Critical species* play a role in maintaining ecosystem stability and resilience through their presence. Their identification is particularly relevant given the increasing pressures on aquatic ecosystems, including habitat destruction, invasive species, climate change, and pollution [16, 12, 2]. For example, recent research highlights how ocean acidification and

warming disrupt plankton communities, with cascading effects on the entire food web [39]; while Baum et al. [17] emphasized how the decline of top predators can destabilize entire marine communities. These cases demonstrate the importance of using network analysis to understand and predict the consequences of environmental changes on aquatic food webs. This relationship extends to other types of ecological networks. Such knowledge is essential for guiding conservation efforts and mitigating the effects of biodiversity loss, particularly in the face of accelerating environmental changes driven by human activity and climate change [118, 167]. Sheykhali et al. in [179] investigated how mutualistic bipartite ecological networks—comprising resource (*e.g.*, plant) and consumer (*e.g.*, pollinator)—respond to species extinction through adaptation or rewiring of interactions. Using random and targeted extinction scenarios (that is, selecting the resource to remove proportionally to its degree), the authors examined the effects of rewiring based on resource affinity (that is, the extinct resource is replaced by another resource selected proportionally to the number of common consumers shared with it). They found that the robustness of the networks are strongly influenced by targeted species removal, especially when paired with affinity-driven rewiring. Notably, modularity and stability tend to increase as networks reorganize to buffer against co-extinctions, suggesting that resource-affinity rewiring helps preserve network structure by enhancing compartmentalization and limiting the spread of disruptions across the network. Pocock et al. [160] conducted a study on a 125-ha farm in Somerset, UK, where they sampled various types of ecological networks within the same ecosystem, including food webs, pollination networks, and host-parasitoid webs. Their findings revealed that networks involving pollinators were notably fragile. Additionally, they observed that the robustness of different network types did not strongly covary, indicating that ecological restoration efforts aimed at supporting one functional group may not necessarily benefit others.

When it comes to food webs, common strategies for testing robustness and identifying the sequence of the most critical species include: (1) targeting the most connected species; (2) randomly selecting species; (3) selecting the most connected species but excluding primary producers; and (4) targeting the least connected species [66, 67]. In these studies, robustness is typically defined as the fraction of species that must be removed to cause a total species loss (primary removals plus secondary extinctions) of $\geq 50\%$ of the original web. The main findings indicate that food webs are generally more resilient to random species loss than to the targeted removal of highly connected species. Furthermore, robustness tends to increase with higher connectance defined as the ratio between the nodes of a food web over the total amount of links, suggesting that food webs with denser connections among species are less susceptible to species loss.

Allesina et al. [5] proposed a method where nodes are removed gradually based on their

degree in a generalized dominator graph. In this graph, resource nodes serve as roots, and edges represent feeding interactions. This approach shifts the focus from individual species to their roles within the network.

Previous studies on food webs have investigated whether network robustness is influenced by factors beyond local properties. Their findings revealed that neither species richness nor omnivory (species feeding across multiple trophic levels) significantly correlate with food web robustness. However, when identifying the most critical node sequence, the procedures primarily rely on local properties, particularly node degree.

Three-nodes motifs

A motif is a small recurring pattern of interactions among a defined number of species. In food webs, for instance, there are 13 unique motifs involving three interacting species [140, 188]. The study of *three-node motifs* provides a meso-scale perspective on ecosystem dynamics, bridging the gap between pairwise interactions and the overall network structure [162]. These recurring patterns, such as exploitative and apparent competition [91], serve as fundamental building blocks for food webs, enabling researchers to investigate the factors contributing to dynamical stability [187]. Additionally, the analysis of three-node motifs complements controlled experimental approaches, such as mesocosm studies [188], by offering a scalable framework for extrapolating findings to more complex ecological systems.

Motif roles capture both direct and indirect ecological interactions, enabling a deeper understanding of the roles of the species within ecosystems beyond traditional metrics such as degree or trophic level [45]. Their effectiveness comes from their ability to analyze network-wide interaction patterns, including complex dynamics like trophic loops and intraguild predation [131]. Studies have revealed universal topological patterns in empirical food webs, independent of their size and connectivity. Milo et al. in [140], building on Williams and Martinez's approach in [207], developed a method to detect "network motifs" - recurring, significant interconnection patterns. Their analysis of three- and four-node subgraphs across various networks revealed that while food webs exhibit unique three-node motif patterns, they share four-node motif characteristics with neuronal and electronic networks. The autors suggest that these motifs likely emerge from evolutionary constraints on network development. Stouffer et al. [188] analyze a broader set of food webs by examining the *representation* of three-node motifs, which measures how frequently a motif appears in a network when randomized under specific constraints. They identified two main types of food webs: one that is rich in intraguild predation and has low levels of competition, and another with the opposite representation. The authors suggest that this observed pattern may stem from biological factors and conclude that it "would be interesting to explore further the ecological or environ-

mental reasons why these two food webs exhibit over-representation of isolated exploitative competition and isolated generalist predation". Additionally, motif roles have been applied to understand the stability o by linking the frequency of specific motifs to network stability [187]. Klaise et al. [104] developed a food web model that can replicate these distinct representation patterns. By adjusting a single parameter, the *trophic coherence*, their model can generate food webs that belong to one of the two motif families,.

The study of motif representation often complements the analysis of community modules, which are minimal yet realistic models describing groups of three interacting species [188]. Community modules aim to reveal how ecological patterns are shaped by both the structural organization of ecosystems and species interactions. For example, in freshwater ecosystems, omnivores—frequently involved in intraguild predation—adjust their foraging behavior to target the most abundant resources. This behavioral flexibility enables them to persist under varying environmental conditions, particularly as system productivity fluctuates [210]. Moreover, they offer valuable insights into changes in population dynamics and their effects on stability. For example, intraguild predation, a common motif, interacts with competition to shape population dynamics [161, 204]. A notable case occurs when threadfin shad are introduced to boost largemouth bass populations. While shad initially serve as prey for adult bass, they also compete with juvenile bass and bluegills for plankton. This competition reduces juvenile bass survival and growth, leading to fewer individuals reaching the piscivorous adult stage. Consequently, the smaller adult bass population preys on the shad, demonstrating how competition and predation within a module directly influence ecosystem stability [120].

Despite their utility, community modules alone cannot fully capture the dynamics of entire food webs. For instance, intraguild predation persists only under specific ecological conditions [187], even though it frequently appears in natural food webs. This highlights the importance of integrating external ecological factors into the analysis of community modules. Resource availability [92] and habitat complexity [99], for instance, significantly shape species interactions and converge in stabilizing or destabilizing these modules.

Beyond Network Analysis: Information Theory Meets Ecology

Traditional network analysis primarily focuses on structural characteristics, such as connectivity, degree distribution, and community detection. However, information theory offers a set of complementary tools that can enhance the analysis of food webs by focusing on the dynamics of interactions within these networks.

Information theory provides a powerful framework for analyzing food webs by quantifying the uncertainty and complexity of species interactions. For example, Robert Ulanowicz lever-

aged these concepts to examine ecosystem development and health, offering a systematic way to assess the balance between order and adaptability in energy flows within ecosystems [194]. These indices enable researchers to evaluate an ecosystem's resilience and adaptability in response to external pressures, such as climate change, provided that data from multiple time periods is available.

Applying information theory to time-series data on species abundances and interactions offers a powerful way to capture how the roles of species within food webs evolve over time and under changing environmental conditions. This dynamic perspective allows researchers to understand the transient states of ecosystems—such as during recovery from disturbances or adaptation to new environmental regimes [176].

These measures are particularly valuable for informing policy-makers and businesses on regulating activities related to fishing and resource exploitation. An illustrative example comes from [159], where methods to improve restoration policies are discussed as critical for addressing the fisheries crisis. The author suggests that whole-ecosystem indices, when combined with biodiversity metrics, can be used to assess an ecosystem's resilience. By employing simulation models, the historical state of an ecosystem can be estimated, allowing for predictions of future scenarios. These simulations help guide management decisions, aiming to balance economic objectives, such as net present value (NPV), with maintaining ecosystem health as indicated by biodiversity.

Beyond Food Webs: Ecological Networks

Research on the structure of food webs is just one example of a broad interdisciplinary research agenda on the structure of all types of networks, both biotic and abiotic. Food webs describe the biological organization in ecological communities, in particular they model how a diverse array of species interact through feeding relationships. However, understanding ecosystems requires a more comprehensive approach that goes beyond trophic interactions to include various types of ecological networks. These networks encompass a wide range of relationships such as mutualistic interactions, competition, and even the role of abiotic factors like nutrient flows.

For instance, ecological networks have been expanded to include non-trophic interactions, such as competition for resources or mutualistic relationships between species like pollinators and plants. This broader framework helps in understanding how these interactions collectively shape the dynamics and stability of ecosystems [31]. Mutualistic networks, such as those involving plant-pollinator relationships, have been shown to possess structural properties that confer robustness against disturbances. The architecture of these networks, characterized by nestedness and high connectivity, has been identified as a critical factor in

maintaining their stability, as observed by Thébault and Fontaine [189]. Their work highlights that while mutualistic networks benefit from nested structures to maintain stability, trophic networks often rely on compartmentalization to buffer against disturbances.

In addition, the study of spatially structured ecological networks, that couple localized resources (food) and habitats, highlights the importance of spatial separation and dispersal among species in influencing network structure and stability, highlighting how localized interactions and movements shape broader ecological patterns [168].

Applications of food webs analysis

Memmott [134] identified different areas where food webs and other ecological networks such as pollinators, host-parassitoids and mutualistic webs have emerged as applied tools, demonstrating their versatility and importance in addressing ecological challenges. Food webs serve as essential tools in *ecological restoration*, guiding the rebuilding of damaged ecosystems. Beyond just tracking species presence, restoration projects must consider functional relationships and interactions between species [69]. Since direct structural comparisons between sites are challenging due to varying species composition [209], ecological networks provide valuable insights into community structure, function, and resilience during ecosystem recovery.

Alien species represent the second greatest threat to global biodiversity after habitat loss [173]. While most research examines their impact at single trophic levels, food webs reveal complex indirect effects like apparent competition [93] that traditional survey methods miss. The interconnected nature of ecosystems means that losing native species to alien invasions can trigger cascading extinctions affecting parasitoids, parasites, and pathogens - groups that significantly influence community structure [35].

Habitat management often involves removing alien plants to protect native species. However, this practice's benefits are largely untested. Research by Carvalheiro et al. [35] on Trinia glauca revealed that its ant pollinators also fed on alien plants. Their network simulations, which allowed for host shifting, demonstrated that removing alien plants could inadvertently harm T. glauca by disrupting its pollinator populations, suggesting that alien plant removal strategies require careful consideration of potential indirect effects on native species.

Climate change threatens species not only through direct habitat loss [80] and disrupted ecological partnerships, but also through mass disruption of ecological interactions. Memmott et al. [135] demonstrated this by studying plant-pollinator networks, finding that climate-induced shifts in flowering and pollinator activity timing could reduce floral resources for 17% to 50% of pollinator species. Their analysis of a network involving 1419 pollinators and 429 plant species suggested that these phenological mismatches could trigger cascading

extinctions, affecting both pollinators and their dependant plants.

1.2.2 Contact Networks

Structural properties of contact networks

Studies on large-scale contact patterns, such as POLYMOD [143] and SOCRATES [205], have demonstrated that face-to-face interactions in a population are often stratified into *communities* featuring dense in-group connections with sparse intergroup ties. This effect, often called *group mixing*, reflects age-based, occupational, or geographical divisions within populations. Many real-world social networks often display a form of group mixing, driven by individuals' natural inclination to socialize with others who share similar interests or attributes, as highlighted by McPherson et al. [133]. Other studies of social networks, which model various types of interactions between individuals (*e.g.*, friendships or professional collaborations [145]), serve as proxies for understanding contact networks. These studies reveal two key feature: *heterogeneity*, which refers to the uneven distribution of connections across nodes in the network, and high *clustering coefficient*, which measures the proportion of fully connected triplets in a network, indicating the prevalence of tightly-knit groups.

These structural features—heterogeneous degree distribution, community structure, and high clustering—profoundly influence the dynamics of epidemic spread. Heterogeneous degree distributions imply that a small number of highly connected nodes (hubs) can disproportionately influence the spread of a disease, acting as super-spreaders and facilitating rapid transmission across the network [153]. Conversely, individuals with fewer connections are less likely to contribute significantly to disease propagation. Group stratification suggests that diseases initially propagate rapidly within tightly connected in-groups but require intergroup ties to spread to the broader population. This dynamic can slow the overall progression of an epidemic, creating natural barriers that may delay its spread [14]. High clustering coefficient indicates a higher prevalence of redundant connections, which can diminish the spread within tightly connected groups, as infections are less likely to exploit redundant ties when other direct pathways exist [24].

1.2.3 Contemporary methods for recreating contact patterns

Accurate modeling of contact patterns is crucial for understanding disease transmission and designing effective epidemiological interventions. These patterns form the backbone of network epidemiology, offering insights into the dynamics of infectious diseases within populations. Modern methodologies leverage advancements in computational techniques,

real-world data, and theoretical modeling to recreate contact networks. This section reviews three primary approaches—data-driven methods, agent-based models, and synthetic network generation—highlighting their methodologies, advantages, and limitations.

Data-driven methods rely on empirical datasets, such as school attendance records, workplace rosters, and mobility data, to infer contact patterns. For example, Mossong et al. [143] constructed age-specific contact matrices across Europe using survey data, while Prem et al. [165] applied Bayesian hierarchical modeling to integrate datasets with varying levels of granularity. Statistical techniques such as exponential random graph models (ERGMs) and latent space models allow researchers to reconstruct network structures from observed interactions [95]. Hens et al. [89] mined social mixing patterns from Belgian population surveys, providing critical insights for epidemiological models. Despite offering high realism and context-specific applicability, these methods are often limited by data availability and ethical concerns surrounding the use of sensitive information.

Agent-based models (ABMs) simulate populations computationally, where agents represent individuals with attributes such as age, occupation, and health status. These models draw input parameters from census data, epidemiological surveys, and demographic studies, incorporating features like transmission probabilities and mobility patterns [137, 9, 48]. Validated against historical outbreaks or synthetic benchmarks, ABMs align simulated dynamics with real-world observations. Ferguson et al. [72] utilized an ABM to evaluate the impact of interventions like school closures and vaccination during influenza outbreaks in the US and UK. Similarly, Fumanelli et al. [75] leveraged demographic data to infer social networks for disease simulations, while Iozzi et al. [97] combined ABMs with serological data to validate transmission estimates. Although ABMs excel in capturing behavioral and temporal heterogeneities, their computational demands and sensitivity to assumptions pose significant challenges.

Synthetic network generation focuses on creating artificial networks that replicate the structural properties of observed systems, emphasizing metrics like degree distributions, clustering coefficients, and assortativity. Classical models, including Erdős-Rényi and Barabási-Albert, produce networks with random and scale-free properties, respectively [10]. Configuration models preserve observed degree distributions while randomizing other connections, balancing realism and flexibility [146]. Hierarchical approaches, which incorporate community detection algorithms and clustering techniques, introduce realistic clustering patterns [84]. These methods enable "what-if" scenarios for testing interventions in hypothetical populations. For instance, Salathé and Jones [172] assessed targeted vaccination strategies using synthetic networks, while Mistry et al. [141] constructed high-resolution networks from mobility and demographic data to model disease dynamics. Meyers et al. [138] demonstrated the

predictive power of synthetic networks by applying percolation theory to epidemic spread in directed contact graphs. Despite their scalability and adaptability, synthetic network models often oversimplify complex social behaviors due to limited empirical grounding.

Each methodology addresses unique research needs. Data-driven methods excel in realism but are constrained by data availability and ethical considerations. Agent-based models provide unmatched flexibility for scenario testing but require careful parameterization and significant computational resources. Synthetic network generation offers scalability and adaptability, thus being an important tool for sensitivity analyses and hypothetical interventions, though its reliance on assumptions can limit applicability.

The approach adopted in this thesis falls under the category of synthetic network generation. By integrating structural properties such as group mixing and clustering, which can also be inferred from empirical data [87], this work bridges gaps between theoretical modeling and real-world applications. Hybrid methodologies, as demonstrated in prior studies [38], enhance the realism and robustness of synthetic models, paving the way for innovative applications in network epidemiology.

1.3 Thesis Contribution

This thesis contributes to the study of complex networks in ecological and epidemiological contexts by addressing critical limitations in current research and developing innovative methodologies to overcome them.

In the field of aquatic food webs, a longstanding challenge lies in the scarcity of large datasets and the absence of systematic analyses of structural properties. This thesis addresses these gaps by analyzing 173 aquatic food webs, the largest dataset examined in this domain to the best of our knowledge. The analysis explores three core structural features. First, it examines the core-periphery structure, extending previous research to food webs and providing insights into the roles of the species, energy and matter transfer pathways, and the general organization of these ecosystems. Second, it identifies critical nodes, *i.e.*, species whose removal disproportionately impacts network connectivity. This study introduces a network-level perspective to node removal analysis and develops a novel robustness metric to systematically compare the stability of different food webs. Finally, it conducts an in-depth examination of three-node motifs, exploring their distribution across aquatic ecosystem types and their ecological significance. Together, these analyses offer a comprehensive understanding of the structural properties governing food webs and their implications for ecosystem resilience.

In the domain of contact networks, this thesis introduces a novel framework for generating

synthetic networks with controllable topological properties. The proposed model incorporates degree heterogeneity, high clustering, and community structure into a unified approach, addressing key limitations of existing random geometric and entropy-based models. By providing precise control over group mixing and structural attributes, the framework enables the realistic simulation of interaction networks. Its adaptability is validated through applications to real-world scenarios, showcasing its ability to replicate observed network properties while remaining versatile across diverse contexts. Furthermore, this framework extends existing models by demonstrating that meso-scale properties can emerge independently of underlying geometric assumptions, thereby clarifying an unresolved aspect in latent geometry networks. This advancement not only broadens the applicability of synthetic network models in areas such as epidemiology and sociology, but also enhances the theoretical understanding of network structure. These contributions extend beyond epidemiological modeling, offering valuable tools for studying other interaction networks, such as those modeling needle-sharing or sexual contact patterns.

Through its dual focus on advancing the analysis of aquatic food webs and developing versatile synthetic models for contact networks, this work delivers actionable insights that support biodiversity conservation, public health strategies, and broader applications in the study of complex systems.

Chapter 2

Background

The purpose of this chapter is to provide context on the key areas of research explored in this thesis, focusing primarily on networks and their applications in ecology and contact networks. We will present an overview of network theory, discuss its use in ecological community research as well as in contact networks, and provide the classes of mathematical methods that we will employ in this thesis, highlighting their importance in overcoming the limitation in the literature.

To begin, we define what a network is and introduce key metrics used to study its structure and behavior.

At its most fundamental level, a network consists of a set of points connected in pairs by lines [145]. In network terminology, these points are called nodes (or vertices), while the lines are called edges. The versatility of network theory lies in its power to abstract and represent various forms of interactions, providing a unified framework for describing complex systems. In recent decades, research on networks has expanded rapidly, largely due to the effectiveness of graph theory and statistical mechanics in modeling and analyzing a diverse range of systems and their topologies.

2.1 Networks

2.1.1 Mathematical Definitions

A network is mathematically represented as a graph G, defined as an ordered pair $G = (V, E)$. Here, V is the set of nodes (or vertices), and E is the set of edges (or links), where each edge is a pair of nodes, such that $E \subseteq V \times V$. Nodes correspond to entities or objects within the system, while edges represent the interactions or relationships between them.

The interpretation of nodes and edges varies with context. For example, in contact networks, nodes can represent individuals engaged in various social activities (*e.g.*, at school [186], during a conference [185], or in urban areas [38]), and edges may indicate forms of contact, such as face-to-face interactions [81], strong ties like kinship or friendship [87], or even sexual relationships [88]. Conversely, in ecological networks, nodes often represent species (*e.g.*, in food webs [65] or plant-pollinator networks [139]), while edges describe relationships such as biomass transfer.

Edges in a network may be directed or undirected. Mathematically, an *undirected edge* is an unordered pair $(u, v) \in E$, where u and v are vertices in V. Undirected edges represent bidirectional relationships, indicating mutual connections between two nodes. For instance, in social networks, a friendship can be modeled as an undirected edge since both individuals equally acknowledge the relationship. *Directed edges* is represented as an ordered pair $(u, v) \neq (v, u)$. represent unidirectional relationships, such as information flow or hierarchical structures, where the relationship is not reciprocated.

Edges can also carry *weights*, numerical values that quantify specific aspects of the relationship, such as strength, capacity, or frequency. For instance, in contact networks, the weight might represent the cumulative duration of face-to-face interactions between two individuals over a defined period.

The inclusion of directionality and weight significantly enhances the modeling capabilities of networks. In ecological studies, for example, food webs represent feeding relationships within ecosystems. Here, nodes correspond to species, groups of organisms, or non-living organic pools consumed by detrivores [65]. Edges are directed, pointing from prey to predator to signify the flow of energy and nutrients. These edges can also be weighted to reflect the rate of predation or biomass transfer, providing detailed insights into ecosystem dynamics.

In a network, two nodes are said to be *adjacent* if they are connected by an edge. For undirected graphs, adjacency is mutual, meaning that if u is adjacent to v, then v is also adjacent to u. In directed graphs, adjacency is defined in terms of the direction of the edge: node u is *adjacent to* node v if there is a directed edge (u, v), *i.e.*, an edge leaving u and coming to v.

An *edge* is said to be *incident to* a node if it connects to that node. In directed graphs, we further distinguish edges as *entering* (incoming edges) or *leaving* (outgoing edges) a node. For a directed edge (u, v), node u is called the *tail*, and node v is called the *head* of the edge.

An *adjacency matrix* a is a square matrix used to represent a network, where the element a_{ij} indicates the presence or weight of an edge between nodes i and j. For unweighted graphs, $a_{ij} = 1$ if an edge exists and 0 otherwise. In undirected graphs, the matrix is symmetric ($a_{ij} = a_{ji}$) while in directed graph is, in general, asymmetric, capturing directionality. An

adjacency list, by contrast, organizes the graph by maintaining a list for each node, detailing all the nodes to which it connects.

The concept of *degree* in a network quantifies the number of connections associated with a node. In directed graphs, this can be further divided into *in-degree* (number of incoming edges) and *out-degree* (number of outgoing edges).

Degree The *degree* of a node v in an undirected network refers to the number of edges incident to the node. It is commonly denoted as $deg(v)$ or deg_v. For adjacency matrix representations, the degree can be computed by summing the values in the v-th row (or equivalently column) of the matrix. For example, in contact networks, the degree of a node represents the number of individuals a person interacts with. High-degree nodes, often referred to as "super-spreaders," play a significant role in the dynamics of disease transmission.

In-degree The *in-degree* of a node v in a directed network is the number of edges entering that node. It is denoted as $deg_{in}(v)$. For adjacency matrix representations, this measure corresponds to the sum of the values in the v-th column of the matrix. In food webs, the in-degree of a species represents the number of prey species consumed by a predator. Nodes with high in-degree are often generalist predators, preying on a wide variety of species.

Out-degree The *out-degree* of a node v in a directed network is the number of edges leaving that node. It is denoted as $deg_{out}(v)$. In adjacency matrix terms, it is calculated by summing the values in the v-th row of the matrix. In food webs, the out-degree of a species represents the number of predators that consume it. Species with high out-degree are typically more vulnerable as they are consumed by multiple predators.

2.1.2 Quantitative Network Analysis

While a complete map of the connections of a network [145] provided by the adjacency matrix or the adjacency list yields comprehensive structural information, interpreting this raw data can be challenging, just like trying to understand a city by looking at a detailed street map without knowing the significance of landmarks or traffic patterns. While the map shows every road and intersection, it provides limited information about how people move through the city, which streets experience the highest traffic, or where key hubs of activity are located. Similarly, the complete structure of a network does not immediately disclose the critical nodes, clusters, or patterns of interaction that influence its overall dynamics. Visual representations may offer qualitative insights for small networks; however, they quickly become impractical and less informative as networks grow in size and complexity.

To overcome this limitation, researchers have developed quantitative measures that distill complex structural data into manageable numerical values. Many of these measures originated in social network analysis (see, for example, [175]), a field developed to enhance our understanding of social structures. In this thesis, we will apply these metrics to study biological networks, focusing on food webs and contact networks, in order to uncover key structural patterns and interactions that influence ecosystem dynamics and population behavior.

Clustering Coefficient of a Node A *triangle* in an undirected network is a set of three nodes that are all pairwise connected by edges. In other words, if nodes u, v, and w form a triangle, then the edges (u, v), (v, w), and (w, u) all exist in the network.

The *local clustering coefficient* of a node v, typically denoted as C_v, quantifies the extent to which the neighbors of v are also connected to each other. Mathematically, it is defined as the ratio of the number of triangles that include v to the total number of possible triangles through v. The total number of possible triangles is given by $\binom{deg(v)}{2}$, where $deg(v)$ is the degree of node v.

For instance, in a social network, a high clustering coefficient suggests that many of your friends are also friends with each other.

Motifs Motifs are small, recurring patterns of interactions that highlight the meso-scale structure of a network [140]. In food webs, motifs are fundamental building blocks that capture processes extending beyond simple pairwise interactions. They include configurations such as apparent competition, intraguild predation, and exploitative competition [161, 90, 131].

Connected Components In an undirected network, a *connected subgraph* is a subset of nodes and edges belonging to G such that there exists at least one path between any pair of nodes within the subset. A connected component is a *maximal* connected subgraph, meaning that it is not possible to add any additional nodes or edges from the network without violating the connectivity property. In simpler terms, a connected component is a subgraph in which all nodes are connected, and no additional nodes from outside the subgraph can be included without breaking the connection.

Strongly Connected Components Similarly, a *strongly connected subgraph* in a directed network is a subset of nodes where every node is reachable from every other node within the subgraph. In other words, for any pair of nodes u and v in the subgraph, there exists a directed path from u to v and a directed path from v to u. A *strongly connected component*

(SCC) is defined as a *maximal* strongly connected subgraph, meaning that no additional edges or vertices from G can be included in the subgraph without breaking its property of being strongly connected.

2.1.3 Metrics Organized by Scale

Network metrics can be further categorized based on the scale of the information they provide. Returning to the map analogy, these scales help distinguish between local landmarks, neighborhood structures, and the city-wide layout.

Microscopic Measurements These metrics focus on individual nodes or edges as isolated entities within the network. For instance, the *degree* of a node in an undirected graph quantifies the number of its direct neighbors, providing insight into local connectivity.

Mesoscopic Measurements Mesoscopic metrics emphasize the organization of groups of nodes or edges. For example, *clustering* fully connected three nodes triples in undirected networks and *three-node motifs* in directed networks reveal local interaction patterns beyond pairwise relationships.

Macroscopic Measurements Macroscopic metrics describe the overall structure of the network. An example is the *giant component*, which represents the proportion of nodes in the largest connected subgraph.

In directed networks, the giant component corresponds to the largest strongly connected component, where every node is reachable from every other node within the subgraph.

In undirected networks, the giant component is the largest connected subgraph where any two nodes are linked by a path.

In figure Figure 2.1 we give an illustration of the metrics organised according to this scale.

2.1.4 Patterns and Properties in Real-World Networks

With various network investigation methods introduced, we now examine common patterns in real-world networks. This section reviews studies that investigate common patterns in real-world networks, emphasizing their influence on system behavior and the resilience of networked systems. In undirected networks, the *giant component* typically includes a significant proportion of the nodes, though the exact fraction varies depending on the structure and context of the network. In networks like the Internet, the giant component underpins

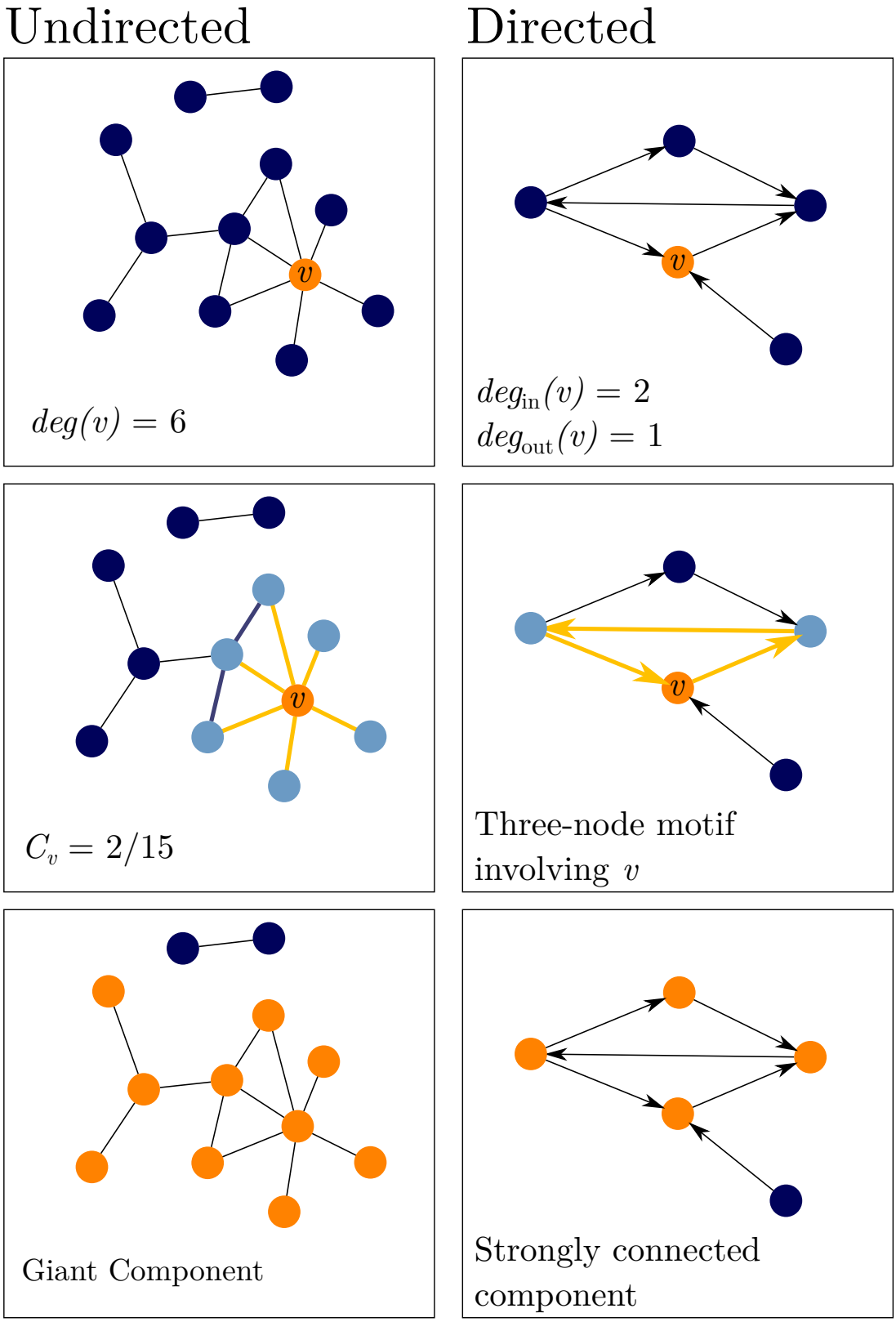


Figure 2.1: Illustration of network metrics categorized by scale: microscopic (first row), mesoscopic (second row), and macroscopic (third row). Each row highlights example metrics and their respective role in network analysis.

robustness by maintaining connectivity, even under random failures, but its fragmentation can lead to catastrophic system breakdowns. The giant component has been studied extensively to understand network fragmentation and the effects of node removal on overall connectivity [145].

In directed networks, the *bow-tie* structure, observed in systems like the World Wide Web, consists of a strongly connected core from which peripheral components radiate. This structure supports efficient information flow while isolating failures to the periphery, enhancing network robustness [29].

The *small-world effect*, where any two nodes are connected by short paths, promotes rapid information or resource dissemination across networks [203]. In real-world networks, this effect is often driven by high-degree nodes, or hubs, which create shortcuts across otherwise distant regions. These hubs also give rise to a scale-free degree distribution, a hallmark of resilience in networks. While scale-free networks can withstand random failures effectively, their dependence on hubs makes them particularly vulnerable to targeted attacks [62].

Another fundamental mesoscale pattern observed in many real-world networks is *community structure*. Communities (or modules) are sets of nodes that are more densely connected to each other than they are to nodes in other parts of the network [73]. Clustering algorithms and community detection methods are designed to identify these subgroups. The significance of these extracted modules lies in their composition: they consist of cohesive sets of nodes and the dense collection of edges linking these nodes, often representing functional units such as ecological guilds, social circles, or regulatory modules.

We conclude by highlighting a key distinction between social networks—networks where nodes can represent individuals, organizations, companies or other entities, and edges capture some form of interaction or relationship [145]—and non-social networks. Social networks typically exhibit a positive correlation in the degrees of connected nodes, a property known as *assortative mixing*. This means that highly connected nodes are more likely to be linked to other highly connected nodes. In contrast, non-social networks often display *disassortative mixing*, where highly connected nodes are preferentially connected to nodes with lower degrees. Furthermore, social networks are characterized by *high clustering* or transitivity, reflecting the tendency of an individual's neighbors to also be connected to each other. In non-social networks, however, clustering is often no greater than what would be expected by chance, given the observed degree distribution [148].

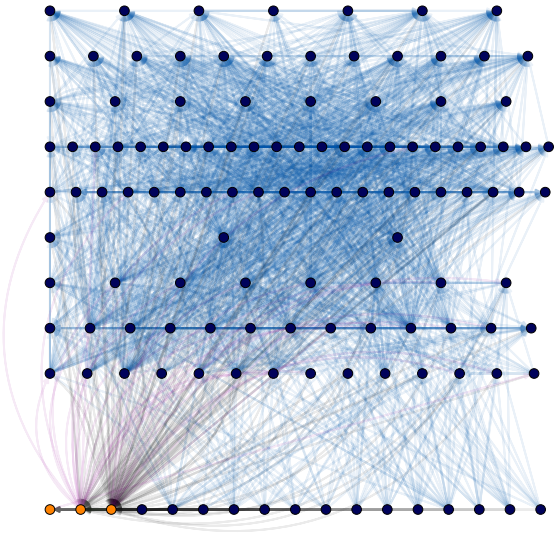


Figure 2.2: The Chesapeake Bay Mesohaline [7] food web, with nodes positioned by trophic level. Blue: living species; orange: non-living deposits of organic matter.

2.2 Food Webs

The purpose of this section is to define food webs and provide a review of their role in ecosystem research. Food webs are a subset of ecological networks, which offer valuable insights into the complex interactions among species in an ecosystem. These networks should be placed within the more general context of ecological networks. Ecological networks provide valuable insights into the complex interactions among species within an ecosystem (see, *e.g.*, [52, 58, 109]).

Over the past few decades, traditional network-based methods have gained popularity in ecological research offering tools to study ecosystem dynamics, such as biodiversity evolution and population changes, and their responses to external disturbances. Understanding the interplay between diverse organisms is important for clarifying the roles of individual species, particularly in the context of perturbations or species invasions [57, 85]. Given the ecological emergencies posed by climate change and human activities [12, 61, 196], it is essential to study trophic interactions at both species and ecosystem levels to predict and mitigate biodiversity loss [190]. Network analysis methods offer powerful tools to address these challenges [145]. Research in this context focusses on mainly three types of networks, based on the interactions they describe: food webs [64], host-parasitoid webs [179, 189], and mutualistic webs, such as plant-pollinator networks [139, 136, 54].

Food webs describe the networks of trophic interactions within ecological communi-

ties [65] modeling species and non-living organic matter compound as nodes, with interactions represented by directed edges that indicate *who eats whom* [65].

Food web analysis is a valuable framework for addressing environmental challenges, such as the potential collapse of global fisheries and the impacts of climate change. This approach has proven effective in detecting shifts in food web structures resulting from climate change, informing policy and management strategies aimed at mitigating negative effects on fisheries and other ecosystem services [193, 157, 132].

2.2.1 Food Webs Model Ecological Communities

In their simplest form, food webs model predation relationships in an ecological community. Consequently, a node represents a trophic species (*i.e.*, a set of taxa sharing same predators and preys) while an edge indicates a *direct* predation relationship from prey to predator. Therefore, in this basic form we might think of food webs as those networks that answer the question "who eats whom" in an ecological community [65].

However, a more complete description of an ecosystem involves the use of some nodes that describe pools of non-living organic material, from which detrivores and other creatures derive biomass that is put back into the ecosystem in a simpler form. In this context, a link from a living organism directed to a detrital component describe a flow of energy and matter that comes from the living tissues of an organisms into a pool of non-living matter [65, 142, 181]. By including these nodes, the edges take on a slightly different meaning: they not only indicate predation relationships but, more generally, how a certain medium (typically carbon) flows in an ecological community, similar to how a transport network works.

These trasfers of organic matter provide energy and nutrients necessary to sustain the metabolism of *heterotrophs*, organisms that cannot produce their own food. To include the energy that fuel *autotroph* - organisms that can convert abiotic sources of energy into energy stored in organic compounds - some models of food web include external inputs and their effective mass transfers. In this thesis, we will neglect these kinds of interactions, focusing only on those nodes that model actual pool of organic matter, either living or non-living.

2.2.2 Modeling and Validating Food Webs

The most typical approach to model a food web is to take a "census" of species in an ecosystem and then determine the network of diets [65] and on-field observation are enough to quantify the species in an ecological community and their feeding relationships. To quantify

the diets, traditionally the measures come from stomach content analysis as shown in Figure 2.3.

A more recent method relies on stable isotope analysis [158], the use of remotely operated vehicles [41] and the use of DNA in feces and gut contents [182].

Figure 2.3: Stomach content analysis. Figure from *Quantifying Food Webs* [112] is distributed under CC-BY licence.

A key question arises: how can these models be validated? This challenge introduces numerous interesting issues about effective ecosystem sampling that go beyond the scope of this thesis. Generally, model limitations fall into two categories: uncertainties originating from the data used during model construction, such as the quality and quantity of the data; and uncertainties related to model structure, which include decisions about the number of species, the level of taxonomic aggregation, and the specific marine domain being modeled (*e.g.*, benthos, pelagos, or the entire continuum) [110].

Furthermore, empirical observations often reflect biases introduced by the compilers of food webs [123]. Such biases may result in food webs with low resolution, exclusion of certain species, or incomplete information on species biomass and dietary quantities [65]. These challenges highlight the importance of using diverse datasets in comparative studies to mitigate individual authors' biases [3].

2.3 Contact Networks

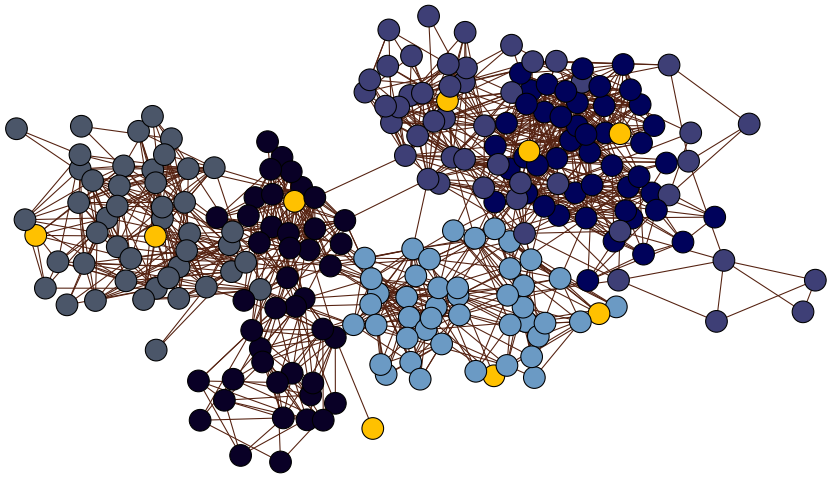


Figure 2.4: A network representing the interactions between students and teachers [186] Nodes form clusters that reflect the community structure of the network. Yellow nodes represent teachers, while blue and grey nodes represent students, divided by age groups.

Contact networks represent interactions among individuals within a population. These interactions often involve physical proximity or direct contact, such as face-to-face meetings or shared spaces [100, 116].

Although contact networks are not inherently biological, they serve as powerful tools for modeling biological processes. This is particularly true in epidemiology, where patterns of interaction between individuals critically influence disease transmission dynamics. For example, in a school environment, a contact network can represent interactions among students and teachers, providing valuable insights into the spread of infectious diseases within such a population [186].

To connect the conceptual framework of contact networks to their epidemiological applications, we examine the mathematical models that leverage these networks to analyze disease transmission patterns. Network-based models represent individuals as nodes in a graph, enabling the study of how social interaction structures influence disease propagation [88, 38, 146, 102]. While these theoretical models provide significant insights, their practical application relies on the quality and scale of data collected from real-world inter-

actions.

Despite advances in data collection techniques, such as wearable sensors and contact tracing, tracking interactions across large or dynamic populations presents challenges. These include incomplete data capture, recall biases in self-reported surveys, and the logistical and financial constraints of large-scale monitoring [68].

In this section, we define contact networks, explore methods for their collection and analysis, and highlight the key topological characteristics that influence disease transmission. To address the challenges of large-scale data collection, we also introduce a model designed to replicate prominent features of real-world contact networks, such as degree distribution, high clustering coefficients, and community structure.

2.3.1 Contact networks model human activity

Contact networks provide a framework for representing human interactions based on physical proximity. Nodes in these networks represent individuals, while edges indicate instances of close contact. These edges are typically *undirected* and may include weights to represent the cumulative duration or frequency of interactions over a specified time period.

Over the past two decades, several methods have been developed to collect data for constructing contact networks. A prominent approach involves using Radio-Frequency Identification (RFID) badges [70], which detect face-to-face proximity between individuals within a range of 1–1.5 meters. These devices exchange radio packets only when individuals are in close range and facing each other, as the human body acts as a shield to the carrier frequency. Encounter data collected by these devices is transmitted to an RFID reader and aggregated on a server for analysis. Figure 2.5 of the operation of this device is shown in the figure.

Other methods include tracking wireless devices [202], contact diaries and interviews [163], and video surveillance [116]. While these approaches vary in scope and precision, they collectively aim to capture the structure of human interactions and generate reliable data for modeling (see *e.g.*, [143, 205]). While these methods capture fine-grained intra-individual contacts, mixing pattern studies focus on understanding population-level contact structures, particularly age-stratified interactions. Seminal examples include the POLYMOD study [143], which surveyed social contact patterns across eight European countries, and the SOCRATES project [205], which extended this work with updated datasets and methodologies.

Another valuable source of network data comes from recorded human movements. Unlike interaction-based networks, these networks typically use locations as nodes, with edges representing the volume of movement between locations [156]. These edges are often weighted

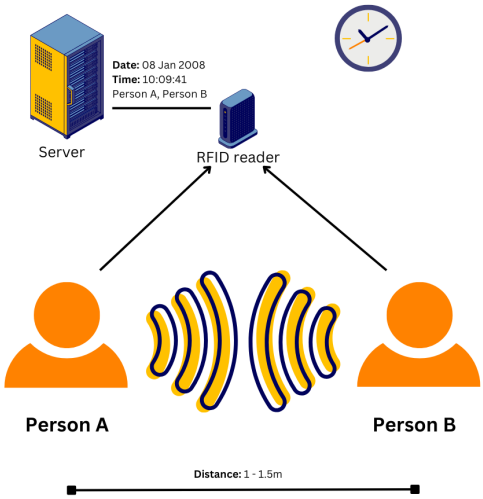


Figure 2.5: A schematic representation of RFID-based proximity tracking: Person A's RFID device detects the proximity of Person B's device when they are within a range of 1–1.5 meters. Both devices transmit data to an RFID reader, which records the interaction, including time and participants, on a central server.

and directional, reflecting the asymmetry of real-world travel patterns. Examples include Airline Transportation Networks [48] - which track the global movement of passengers and Commuter Networks [111], capturing daily travel to and from workplaces.

The resulting data can be organized in two primary ways: as *cumulative networks*, where all contacts within a specific time frame are aggregated [186], or as *temporal networks*, where the network evolves over time, capturing changes in nodes and edges at different moments [212]. This thesis focuses exclusively on static networks, where nodes and edges are fixed, representing the *cumulative* aggregation of contacts over a defined time span.

Challenges in Collecting Contact Network Data

Collecting empirical data on direct interactions between individuals presents significant challenges. Surveys and diaries offer rich contextual details, such as the duration and nature of contacts, but are often biased due to inaccurate recall and are expensive to administer. Wearable sensors provide an objective alternative, detecting close-range proximity [124] and enabling large-scale, temporally resolved studies. However, these methods are typically restricted to groups of a few hundred individuals [124, 74, 186], limiting their scalability and making it impractical to accurately map contact networks for populations numbering in the tens of thousands.

A common solution is to rely on synthetic population models [71, 87] designed to replicate key topological features of real-world networks. However, the degree of realism required depends on which network features significantly influence the dynamics of disease spread. In the next section, we introduce the mathematical framework for modeling outbreaks on networks and identify the topological features most critical to understanding and predicting epidemic dynamics.

2.3.2 Epidemic Dynamics Across Network Structures

This section provides a foundational understanding of traditional epidemic modeling approaches, focusing on compartmental models like SIR. It introduces basic epidemiological models and explores their dependence on network topological features such as degree distribution, clustering, and community structure. By doing so, the section aims to make the thesis accessible to readers with diverse backgrounds while highlighting the limitations of traditional models. This serves to motivate the transition to more advanced network-based approaches discussed later, emphasizing their reliance on topological properties of networks in shaping epidemic dynamics.

With contact networks and their collection methods established, we now explore their application in epidemiology. Compartmental models, such as SIR, form the foundation of infectious disease modeling by dividing populations into groups—susceptible (S), infected (I), and recovered (R)—and describing transitions between these compartments using differential equations and agent-based modeling.

Although these models offer valuable insights into how infections spread across a network, the primary purpose of network epidemiology is not to predict a single, definitive outcome. Predictions can paradoxically alter human behavior and public policy the moment they become widely known, thereby reshaping the very dynamics the models aim to forecast. Instead, the strength of network-based epidemic models lies in their ability to generate *scenarios* that account for the complexity of contagion dynamics and the heterogeneity of contact patterns within a population. By simulating possible trajectories under varying assumptions, researchers and policymakers can better anticipate potential challenges, evaluate intervention strategies, and devise adaptive responses that reflect real-world complexities. Achieving these scenarios requires two interconnected components: a realistic model of transmission and recovery dynamics, and an accurate representation of contact patterns within the population.

Compartmental Models

Compartmental models are an important framework in the context of mathematical epidemiology, offering a systematic approach to studying how infectious diseases spread through populations. In this framework, the population is divided into compartments based on disease states—such as susceptible (S), infected (I), and recovered (R). Transitions between compartments are driven by rates that encapsulate biological and social processes, such as transmission, recovery, and social interactions.

What makes compartmental models particularly powerful is their versatility. At their simplest, they assume a *well-mixed* population, where all individuals interact uniformly—a simplifying assumption that may not hold in heterogeneous real-world settings. However, their framework can be expanded to reflect real-world complexities, incorporating heterogeneities such as age groups, spatial distributions, or contact networks [14]. This adaptability allows these models to simulate disease dynamics across diverse scenarios, making them indispensable for understanding and predicting outbreaks.

A variety of compartmental models have been developed to address specific epidemiological questions:

- **SIR Model:** Individuals transition through the states of Susceptible (S), Infected (I), and Recovered (R), often assuming immunity after recovery.
- **SEIR Model:** Adds an *Exposed* (E) compartment to account for the latency period during which individuals are infected but not yet infectious.
- **SIS Model:** Models diseases that do not confer immunity (*e.g.*, common colds), with individuals cycling between Susceptible (S) and Infected (I) states.
- **MSEIR Model:** Extends the SEIR framework by including maternal immunity in newborns.

Among these frameworks, the SIR model stands out for its simplicity and its ability to capture essential aspects of epidemic dynamics. As such, it serves as the foundation for the discussion that follows.

SIR in Mean-Field Approximation In its simplest form, the SIR model divides the population into three compartments: *S*usceptible individuals who can become infected, *I*nfected individuals who can transmit the disease, and *R*ecovered individuals who have developed immunity. The dynamics of these compartments are driven by two primary processes: transmission, where susceptible individuals contract the disease through interactions

with infected individuals, and recovery, where infected individuals either recover or gain immunity. These interactions and transitions are governed by specific rates and are mathematically expressed as a system of ordinary differential equations. The dynamics of these compartments are governed by transmission and recovery rates, described mathematically by the following system of ordinary differential equations:

$$\begin{cases} \frac{dS}{dt} = -\beta \frac{SI}{N}, \\ \frac{dI}{dt} = \beta \frac{SI}{N} - \gamma I, \\ \frac{dR}{dt} = \gamma I, \end{cases} \tag{2.1}$$

where:

- $S(t), I(t), R(t)$ represent the number of individuals in each *compartment* at time t.
- $N = S(t) + I(t) + R(t)$: the total population, assumed constant.
- β is the *transmission rate*, indicating the likelihood of infection during contact between a susceptible and an infected individual.
- γ is the *recovery rate*, representing the fraction of infected individuals who recover per unit time.

This model assumes *homogeneous* (or well-mixed) interactions, whereby every individual is equally likely to contact every other individual. In Equation 2.1, the fraction of infected individuals $I(t)/N$ multiplies the fraction of susceptible individuals $S(t)/N$, generating new infections at a rate proportional to $\beta\left(\frac{S}{N}\right)\left(\frac{I}{N}\right) \times N$. Under this approximation, one obtains the threshold condition

$$R_0 \;=\; \frac{\beta}{\gamma},$$

where $R_0 < 1$ prevents the spread of infection and $R_0 > 1$ drives an epidemic.

A convenient way to see why $R_0 = \frac{\beta}{\gamma}$ acts as the threshold is through a stability analysis near the disease-free equilibrium (DFE). At the DFE, $I = 0$ and $R = 0$, implying $S = N$. Linearizing around $I \approx 0$ and $S \approx N$:

$$\frac{dI}{dt} \;\approx\; \beta \frac{N \cdot I}{N} - \gamma I \;=\; (\beta - \gamma)\, I.$$

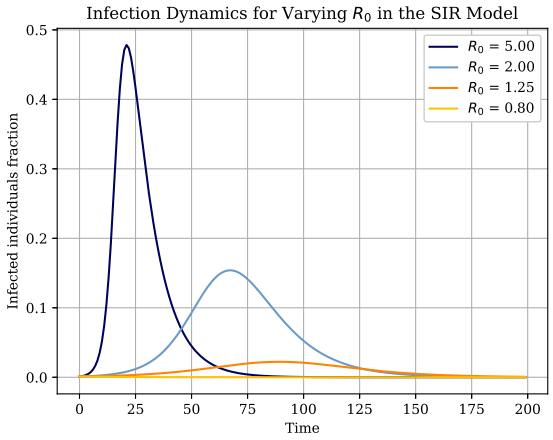
- If $\beta - \gamma > 0$ ($\frac{\beta}{\gamma} > 1$), dI/dt is positive for small I, so the infection grows and can invade the population.
- If $\beta - \gamma < 0$ ($\frac{\beta}{\gamma} < 1$), dI/dt is negative for small I, so the infection tends to die out.

Thus, we define the basic reproduction number as

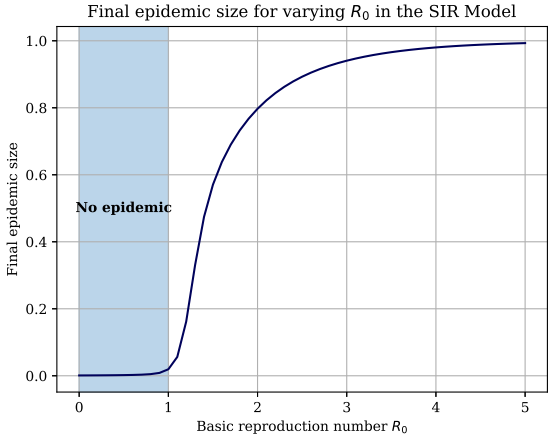
$$R_0 = \frac{\beta}{\gamma}.$$

When $R_0 < 1$, the epidemic cannot sustain itself and eventually dies out. Conversely, when $R_0 > 1$, a single infectious individual introduced into a fully susceptible population can, on average, infect more than one other person, leading to the spread of the epidemic. In Figure 2.6, we illustrated these distinct regimes through numerical integration for varying values of R_0.

This parameter can be estimated from real epidemic data [218]. For instance, during the COVID-19 outbreak in Italy, R_0 was estimated between 2.43 and 3.10 in March 2020 [53], with public health measures helping to reduce this value [83].



(a) Solution of the SIR model under the homogeneous mixing assumption for the infected population, $I(t)$, over time for varying basic reproduction numbers (R_0). The total population size is $N = 1000$, with a single initially infected individual.



(b) Asymptotic number of infected individuals as a function of the basic reproduction number R_0.

Figure 2.6: Epidemic dynamics and final epidemic size as functions of the basic reproduction number R_0 under the homogeneous mixing assumption. The size of the infected population as a function of R_0 highlights the threshold behavior at $R_0 = 1$, below which the infection dies out and above which the epidemic affects a significant portion of the population.

This type of differential equation system, often referred to as *mean-field* equations, is straightforward but ignores any underlying contact heterogeneity or network structure.

Heterogeneous Mixing Equations Many real-world systems exhibit variability in the number of contacts that each individual has. This variability can be captured by classifying

individuals according to their degree k (*i.e.*, the number of contacts). Although still an approximation of real epidemic dynamics, adding degree information is highly instructive because it illuminates the effect of the *degree distribution* on the spread of infection.

Let $p(k)$ denote the probability that a randomly chosen individual has degree k. In the *heterogeneous mean-field* (or *degree-based*) SIR model, we track $S_k(t)$, $I_k(t)$, and $R_k(t)$: the fractions of degree-k nodes in each compartment at time t, where $S_k + I_k + R_k = 1$ for each k. The total fraction of the population that has degree k is $p(k)$.

The infection and recovery dynamics are then

$$\begin{aligned}\frac{dS_k}{dt} &= -\beta\, S_k\, k\, \Theta,\\ \frac{dI_k}{dt} &= \beta\, S_k\, k\, \Theta - \gamma\, I_k,\\ \frac{dR_k}{dt} &= \gamma\, I_k,\end{aligned}$$

where

$$\Theta(t) \;=\; \sum_{k'} \frac{k'-1}{\langle k\rangle}\, p(k')\, I_{k'}(t)$$

represents the fraction of edges leading to infected nodes in the neighborhood of a susceptible node of degree k [11]. This term quantifies the average probability that a neighbor of a degree-k vertex is infected. Under a fully homogeneous assumption, this probability would simply be the overall density of infected individuals. However, in a heterogeneous network, $\Theta(t)$ is more complex since it must account for the variability in node degrees and the influence of those degrees on how infection spreads [14]. Here, we consider a simpler case in which *degree correlations* are absent. An infected node of degree k' can transmit infection through its remaining $k'-1$ links, and $\langle k\rangle$ is the average degree over all individuals.

Epidemic Threshold for Heterogeneous Mixing At the disease-free equilibrium (DFE), we have

$$I_k^{(\mathrm{DFE})} = 0, \quad S_k^{(\mathrm{DFE})} = 1, \quad R_k^{(\mathrm{DFE})} = 0 \quad \text{for all } k.$$

To determine whether a small perturbation away from this DFE grows or decays, one can linearize the equations around $I_k = 0$ and set $S_k(0) = 1 - I_k(0)$ [14]. In this linear regime, it is possible to integrate the equations for $S_k(t)$ and $R_k(t)$, obtaining

$$\begin{aligned}S_k(t) &= e^{-\beta\, k\, \phi(t)},\\ R_k(t) &= \gamma \int_0^t I_k(\tau)\, d\tau,\end{aligned}$$

where

$$\phi(t) \;=\; \int_0^t \Theta(\tau)\, d\tau \;=\; \frac{1}{\langle k\rangle}\,\frac{1}{\gamma}\sum_k (k-1)\,p(k)\,R_k(t).$$

Infinite-Time Limit of $R_k(t)$ A common approach to identifying the threshold is to analyze the limiting behavior of $R_k(t)$ as $t \to \infty$. If there is no outbreak, then $R_k(t)$ remains zero for all practical times; otherwise, $R_k(t)$ grows above zero. This is convenient because in the limit $t \to \infty$ we expect no active infections ($I_k(\infty) = 0$), and thus

$$\lim_{t\to\infty} R_k(t) \;=\; \lim_{t\to\infty}\big(1 - S_k(t)\big).$$

To simplify notation, let $\lim_{t\to\infty} f_k(t) = f_k(\infty)$ and $\lim_{t\to\infty}\phi(t) = \phi_\infty$. We define the total epidemic prevalence as $r_\infty = \sum_k p(k)R_k(\infty)$. We obtain:

$$r_\infty \;=\; \sum_k p(k)\,\big(1 - e^{-\beta\,k\,\phi_\infty}\big).$$

To find ϕ_∞, consider its derivative:

$$\frac{d\phi(t)}{dt} = \frac{1}{\langle k\rangle}\sum_k (k-1)\,p(k)\,I_k(t) \;=\; 1 \;-\; \frac{1}{\langle k\rangle} \;-\; \mu\,\phi(t) \;-\; \frac{1}{\langle k\rangle}\sum_k (k-1)\,p(k)\,e^{-\beta\,k\,\phi(t)},$$

where we used $I_k(t) = 1 - S_k(t) - R_k(t)$, and μ is a parameter reflecting how ϕ evolves alongside recovery and infection terms. In the limit $t \to \infty$, $I_k(\infty) = 0$ and hence $\frac{d\phi_\infty}{dt} = 0$. This yields

$$\mu\,\phi_\infty \;=\; 1 \;-\; \frac{1}{\langle k\rangle} \;-\; \frac{1}{\langle k\rangle}\sum_k (k-1)\,p(k)\,e^{-\beta\,k\,\phi_\infty}.$$

Asymptotic Solution via Graphical Analysis One solution is $\phi_\infty = 0$, which implies $r_\infty = 0$ (no outbreak). A nonzero solution $\phi_\infty > 0$ (hence $r_\infty > 0$) exists if

$$\frac{d}{d\phi_\infty}\Big[1 - \frac{1}{\langle k\rangle} \;-\; \frac{1}{\langle k\rangle}\sum_k (k-1)\,p(k)\,e^{-\beta\,k\,\phi_\infty}\Big]\Big|_{\phi_\infty=0} \;\geq\; \mu,$$

which can be verified using a standard graphical argument[1] and leads to

$$\frac{\beta}{\langle k\rangle}\sum_k k\,(k-1)\,p(k) \;\geq\; \mu,$$

[1]See, for example, Fig. 9.5 in [14].

defining the threshold condition:

$$\frac{\beta}{\mu} = \frac{\langle k \rangle}{\langle k^2 \rangle - \langle k \rangle}. \tag{2.2}$$

This result shows that, in degree-heterogeneous populations, the epidemic threshold depends crucially on both the first and second moments of the degree distribution. The factor $\langle k^2 \rangle - \langle k \rangle$ becomes larger for networks with higher degree variance, which lowers the threshold and makes outbreaks more likely. Consequently, even a small fraction of highly connected "hubs" can facilitate widespread contagion, underscoring the profound effect that degree heterogeneity and thus the topology of a network can have on epidemic dynamics.

Quenched Mean-Field Equations (Network-Based) When the explicit structure of a contact network is known, the epidemic dynamics can be modeled at a more granular level by assigning each node its own set of equations. In this approach, often referred to as the *quenched mean-field* (QMF) model, the adjacency matrix $\mathbf{A} = \{a_{ij}\}$ encodes the network structure, where $a_{ij} = 1$ if there is an edge (contact) between nodes i and j, and $a_{ij} = 0$ otherwise. The state of each node is described by the probabilities $S_i(t)$, $I_i(t)$, and $R_i(t)$, representing the likelihood that node i is susceptible, infected, or recovered, respectively. The governing equations for these probabilities are given by:

$$\begin{aligned}
\frac{dS_i}{dt} &= -\beta\, S_i \sum_j a_{ij} I_j, \\
\frac{dI_i}{dt} &= \beta\, S_i \sum_j a_{ij} I_j - \gamma\, I_i, \\
\frac{dR_i}{dt} &= \gamma\, I_i,
\end{aligned}$$

with the normalization condition $S_i + I_i + R_i = 1$ for all nodes i. In this formulation, the infection risk for a node depends explicitly on its neighbors, as encoded by the adjacency matrix. Once the network topology is known and the epidemic dynamics are defined, this system of differential equations can be solved numerically using standard integration techniques.

Epidemic Threshold for Quenched Mean-Field In the case of the SIS compartmental model, Wang et al. [201] demonstrated that the epidemic threshold depends on the inverse of the largest eigenvalue of the adjacency matrix, $\lambda_{\max}$. This result highlights the critical role of network structure in determining the onset of epidemics. Further insights were provided by Chung et al. [42], who derived an expression for $\lambda_{\max}$ in finite networks

with power-law degree distributions. Their findings show that for networks where node degrees follow a power-law distribution, the largest eigenvalue scales proportionally to the ratio $\frac{\langle k^2 \rangle}{\langle k \rangle}$, where $\langle k^2 \rangle$ and $\langle k \rangle$ are the second and first moments of the degree distribution, respectively. This scaling relationship aligns with the epidemic threshold predicted by the heterogeneous mean-field approximation [36], thereby bridging the gap between mean-field models and network-specific formulations.

Agent-Based SIR Model In parallel with the QMF approach, a fully agent-based model captures the epidemic process by simulating each individual (or node) in one of three possible states: susceptible (S), infected (I), or recovered (R). At any point in time, the transition of a state of an agent depends on its infected neighbors as well as the infection (β) and recovery (γ) rates. Specifically, a susceptible agent becomes infected with probability proportional to β whenever it encounters an infected neighbor, and infected agents transition to the recovered state at rate γ. Unlike the QMF system of differential equations, this method does not require solving an ODE system; instead, it relies on stochastic or rule-based updates of discrete agents, which can more closely reflect individual-level heterogeneity and real-world transmission patterns.

To facilitate flexible experimentation with different compartmental models and transition rules, we have developed a fully customizable agent-based platform named *EpiDict*, available on GitHub at https://github.com/davidetorre92/EpiDict. In *EpiDict*, users can specify the compartmental structure (*e.g.*, SIR, SEIR) and define custom transitions by editing a dictionary that governs infection, recovery, and other relevant processes. The diagram in Figure 2.7 illustrates the per-agent transition rates between the S, I, and R compartments. In an agent-based approach, each agent updates its state based on interactions with infected neighbors at the specified infection rate β, as well as a recovery rate γ.

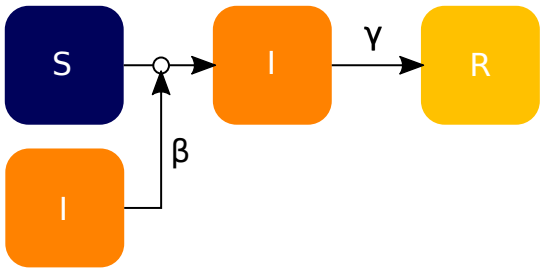
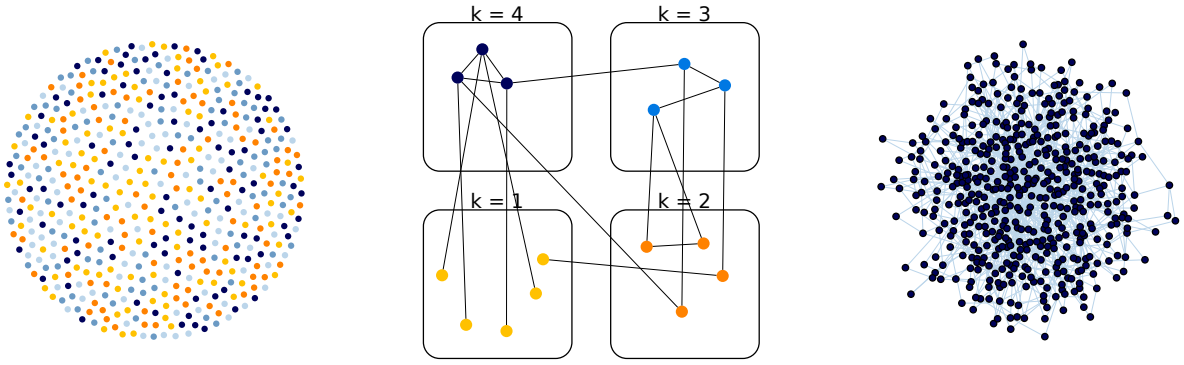


Figure 2.7: A schematic of the agent-based SIR model. Each agent is in one of three states: susceptible (S), infected (I), or recovered (R). Infection occurs with rate β, triggered by contact with infected neighbors, and recovery occurs with rate γ.

These three SIR formulations—(i) *mean-field* ODEs, (ii) *heterogeneous mixing* degree-based



(a) Mean-field approximation assumes a fully mixed population where every individual is equally likely to interact with any other.

(b) Heterogeneous mean-field approximation incorporates degree-based variability, categorizing individuals by their number of contacts.

(c) Quenched mean-field approximation utilizes the explicit structure of a contact network, modeling dynamics at the level of individual nodes.

Figure 2.8: Different approximation schemes for the SIR model in network epidemiology, illustrating the increasing granularity from mean-field to quenched mean-field approaches.

equations, and (iii) *quenched mean-field* node-level equations—demonstrate the role of population structure and network topology in shaping the epidemic threshold. The classical mean-field approach yields a straightforward threshold, $R_0 = \beta/\gamma$, assuming uniform mixing. In contrast, degree heterogeneity reveals a threshold dependent on the ratio $\frac{\langle k \rangle}{\langle k^2 \rangle - \langle k \rangle}$, reflecting the impact of variable connectivity. Finally, full adjacency knowledge introduces a threshold governed by the largest eigenvalue of the contact networks. The mathematical formulations of epidemic models provide foundational insights into how diseases spread across populations and show the importance of the underlying network topology to determine the dynamics of this process. However, understanding the practical implications requires evaluating key quantities, such as the epidemic threshold and the temporal evolution of infections, which are influenced by network topology. In the next section we illustrate these quantities, and show how simulations can inform epidemic preparedness and response strategies.

Epidemic Threshold, Evolution, and Reproduction Number

Simulating the spread of infectious diseases enables researchers and public health officials to analyze key aspects of epidemic dynamics and assess a population's susceptibility to outbreaks. Beyond the basic reproduction number [38], several metrics provide critical insights into the progression and control of epidemics, including:

(i) the *time evolution of infected and recovered individuals*, represented as $I(t)$ and $R(t)$. These trajectories provide essential information on the epidemic's growth rate, peak infection

levels, and duration, which are vital for resource allocation, such as hospital beds and medical supplies;

(ii) the *epidemic threshold*, a critical parameter that combines the characteristics of the pathogen and the network topology. This threshold determines whether an outbreak dies out or sustains itself. In computational epidemiology, it is mathematically defined based on conditions such as the largest eigenvalue of the contact network's adjacency matrix [40]. Real-world applications often interpret the threshold as the point at which interventions, such as vaccination campaigns or mobility restrictions, must be urgently implemented to prevent further spread;

(iii) the *effective reproduction number* [108], $R(t)$, which quantifies the average number of secondary infections caused by an infectious individual at time t. Unlike the basic reproduction number, $R(t)$ evolves dynamically as the epidemic progresses and interventions or behavioral changes alter transmission patterns. Tracking $R(t)$ provides real-time feedback on the effectiveness of public health measures;

(iv) the *attack rate* [108, 38], which measures the proportion of the population that becomes infected over the course of the epidemic. This metric is particularly relevant for assessing the overall impact of an outbreak and planning long-term recovery strategies;

(v) the *case fatality rate* (CFR) and *infection fatality rate* (IFR) [108], which quantify the lethality of the disease by measuring the proportion of deaths among confirmed cases and among all infections, respectively. These metrics are essential for understanding the severity of the disease and prioritizing vulnerable populations;

(vi) the *serial interval* and *generation time* [108], which characterize the timing between successive cases in the transmission chain. These temporal metrics help refine models of disease spread and improve predictions of future case trajectories.

The ability to calculate and interpret these metrics depends heavily on the structure of the population's contact network. Features such as degree distribution, clustering, and assortativity play a pivotal role in shaping the dynamics of disease transmission. By combining these measurements with real-world data, researchers can better inform public health policies, optimize intervention strategies, and predict the trajectory of an outbreak with greater precision.

2.3.3 The Impact of Network Topology on Epidemic Dynamics

The topology of a contact network profoundly influences the spread and dynamics of infectious diseases by shaping interaction patterns and determining how pathogens propagate. In the context of compartmental models, we have demonstrated that the variability and average

of the degree distribution a crucial role in the evolution of epidemics. This section further explores the topological features of contact networks, illustrating their relevance in modeling epidemic dynamics and predicting outbreak outcomes.

Degree Distribution The *degree distribution* quantifies the likelihood of nodes in a network having a specific number of connections. Contact networks exhibit diverse degree distributions, ranging from *scale-free networks*, dominated by a few highly connected nodes [113], to broader distributions [217], and occasionally resembling Erdős–Rényi (ER) random networks [116]. Empirical studies of human social interactions have identified a characteristic upper bound on degree, known as the Dunbar number—approximately 150 connections—reflecting cognitive constraints on maintaining stable social relationships [63]. The degree distribution significantly impacts epidemic thresholds, particularly through metrics like $\langle k(k-1)\rangle/\langle k\rangle$, which influence the basic reproduction number R_0.

Clustering and the Global Clustering Coefficient The *global clustering coefficient* measures the tendency of nodes to form tightly-knit groups or triangles, providing insights into the local density of connections. It is mathematically defined as:

$$c = 3 \times \frac{\text{number of triangles}}{\text{number of connected triples}}.$$

High clustering is prevalent in environments like workplaces and schools, where individuals form tightly connected subgroups [116, 186, 185]. While clustering reduces the spread of disease beyond the first few generations by limiting new contacts [101, 103], outbreaks can become more intense within tightly clustered communities once the pathogen penetrates these boundaries.

Degree Assortativity *Degree assortativity* reflects the tendency of nodes to connect with others of similar degree. Networks with positive assortativity often exhibit a *rich-club effect*, where highly connected nodes preferentially link to one another [127]. In contact networks, this structure can amplify R_0 by creating densely connected hubs that facilitate rapid pathogen transmission [88, 81]. Consequently, degree-assortative networks are more vulnerable to outbreaks, as these interconnected hubs act as efficient conduits for disease spread.

Community Structure *Community structure* and *group mixing* describe the partitioning of a network into cohesive subgroups or *communities*, characterized by denser intra-community connections relative to inter-community links. Newman and Park [148] demonstrated how community structure shapes clustering coefficients and degree correlations in

social networks. Many contact networks exhibit strong community structures, driven by individuals' natural propensity to form groups based on shared attributes or interests [133].

The concept of community structure is evident in friendship networks, where dense intra-community ties coexist with sparse inter-community connections. Granovetter's "The Strength of Weak Ties" [86] highlighted how weak ties—connections between individuals in different communities—serve as critical bridges for information flow and, by extension, disease transmission. Community organization can be quantitatively characterized using the *edge mixing matrix* K, where K_{IJ} enumerates edges between communities I and J, with intra-community edges ($I = J$) counted twice to reflect bidirectionality [21].

Age-stratified mixing patterns, extensively studied in POLYMOD [143] and SOCRATES [205], provide further insights into community structure. For example, school-aged children often form highly clustered networks characterized by intense intra-group interactions, while inter-generational contacts, though less frequent, act as bridges for transmitting diseases between age cohorts. This hierarchical organization profoundly influences epidemic dynamics, determining how pathogens spread across and within communities.

2.3.4 Generation of synthetic network ensambles

Having explored the impact of key topological properties of networks on epidemic dynamics, we now turn to frameworks that enable the generation of network ensembles incorporating these properties. Since reconstructing the complete structure of real-world networks is often infeasible, these frameworks provide a powerful alternative by producing probabilistic representations that capture essential topological characteristics. Two such approaches are the *maximum entropy* method and *latent geometry* networks. The maximum entropy framework allows for the generation of ensembles based on predefined macroscopic constraints, such as degree distribution or group mixing, ensuring that the generated networks remain unbiased within those constraints. In contrast, latent geometry networks model contact patterns by embedding nodes in a geometric space, where proximity governs connectivity. Together, these methods offer complementary tools for generating realistic network scenarios, providing a robust framework for constructing contact networks useful in developing outbreak scenarios within an epidemiological context.

Maximum entropy framework

In this section we introduce the methodology that will enable us to define a network model with the topological characteristics we have discussed so far. To introduce this method, it is good to start with an analogy.

Consider the scenario of preparing for a trip with incomplete information about the weather at your destination. You know it is autumn, the location is coastal, and there is a possibility of rain. In this context of uncertainty, the objective is to pack efficiently without overloading your luggage. To achieve this, you rely on the available information: selecting layered clothing suitable for cooler temperatures, including a raincoat to account for potential rainfall, and choosing versatile footwear appropriate for both wet and dry conditions. This approach exemplifies a balance between practicality and uncertainty, ensuring that your preparations align with the limited information while maintaining adaptability to a range of potential weather scenarios.

This simple decision-making process illustrates the essence of the Maximum Entropy Method. When dealing with systems where information is incomplete, the Maximum Entropy principle provides a systematic way to make the best possible predictions or inference is based on the constraints we know, while avoiding unwarranted assumptions. In this example, the constraints are the weather clues, and the principle ensures you don't overcommit to a specific scenario (*e.g.*, packing only for rain or only for sunshine). Instead, you maximize your adaptability within the boundaries of your knowledge.

The Maximum Entropy Method is rooted in information theory and is widely used in scientific and engineering disciplines for problems involving uncertainty [184]. Its fundamental idea is to select a probability distribution that maximizes the entropy—a measure of uncertainty or randomness—while satisfying known constraints. By doing so, the method ensures the least biased estimate possible, honoring the given information while refraining from incorporating assumptions not supported by the data.

The analogy of packing for a trip mirrors how the method is applied in practice. For instance, in statistical mechanics, the Maximum Entropy principle provides a framework to infer the distribution of microscopic states of a system given macroscopic constraints, such as energy or particle density [166]. Similarly, in network science, it can be used to generate ensemble networks that satisfy specific topological properties, like degree distributions, while avoiding unjustified assumptions about other structural characteristics [183]. The principle's versatility has also made it invaluable in fields as diverse as machine learning [94], natural language processing [20], and biology [56].

In this section, we introduce the general framework and outline how to derive the probability distributions for the $G(n, M)$ and $G(n, p)$ models from the maximum entropy principle, subject to appropriate constraints.

General Setup Consider a set $\mathcal{G}$ of all possible graphs on n vertices (labeled $1, 2, \ldots, n$). Each graph $G \in \mathcal{G}$ may be represented by its adjacency matrix $\{x_{ij}\}$, where $x_{ij} = 1$ if

there is an edge between nodes i and j, and $x_{ij} = 0$ otherwise. For simplicity, we consider undirected, simple graphs, so $x_{ij} = x_{ji}$ and $x_{ii} = 0$.

We seek a probability distribution $P(G)$ over $\mathcal{G}$ that maximizes the Shannon entropy:

$$S(P) = -\sum_{G \in \mathcal{G}} P(G) \ln P(G),$$

subject to constraints on one or more network statistics.

The general problem can be framed as:

$$\max_P \; S(P) = -\sum_{G \in \mathcal{G}} P(G) \ln P(G)$$

subject to:

$$\sum_{G \in \mathcal{G}} P(G) = 1 \quad \text{(normalization constraint)}$$

and possibly constraints of the form

$$\sum_{G \in \mathcal{G}} P(G) C(G) = C^*,$$

where $C(G)$ is some network statistic (*e.g.*, number of edges) and C^* is the desired (fixed or expected) value.

To solve this constrained optimization problem, we use the method of Lagrange multipliers. Define the Lagrangian:

$$\mathcal{L}(P, \alpha, \{\lambda_i\}) = -\sum_{G \in \mathcal{G}} P(G) \ln P(G) + \alpha \left(1 - \sum_{G \in \mathcal{G}} P(G)\right) + \sum_i \lambda_i \left(C_i^* - \sum_{G \in \mathcal{G}} P(G) C_i(G)\right),$$

where α and $\{\lambda_i\}$ are Lagrange multipliers enforcing the constraints.

Taking the functional derivative with respect to $P(G)$, we get:

$$\frac{\partial \mathcal{L}}{\partial P(G)} = -\ln P(G) - 1 - \sum_i \lambda_i C_i(G) - \alpha = 0.$$

Application: Derivation for the $G(n, M)$ Model (Hard Constraint) In the $G(n, M)$ model, also known as the Erdős–Rényi-Gilbert $G(n, M)$ model, we impose a *hard constraint* on the total number of edges. Let $E(G)$ denote the number of edges in G. Our constraints are:

$$\sum_{G \in \mathcal{G}} P(G) = 1 \quad \text{and} \quad P(G) > 0 \iff E(G) = M,$$

In other words, we restrict the support of $P(G)$ to only those graphs with exactly M edges. Solving the entropy maximization problem for $P(G)$ yields:

$$-\ln P(G) = 1 + \alpha \implies P(G) = e^{-1-\alpha}.$$

Notice that the right side of the equation doesn't depend on G. Using the constraint on the normalization we get:

$$\sum_{G \in \mathcal{G}} P(G) = 1 \implies \sum_{G \in \mathcal{G}} e^{-1-\alpha} = 1 \implies e^{-1-\alpha} = \frac{1}{|\{G : E(G) = M\}|}$$

By noticing that the number of graphs with n vertices and M edges is $\binom{\binom{n}{2}}{M}$, the only way to maximize entropy given this strict condition is to distribute the probability mass *uniformly* across all graphs with M edges:

$$P(G) = \begin{cases} \frac{1}{\binom{\binom{n}{2}}{M}}, & \text{if } E(G) = M, \\ 0, & \text{otherwise.} \end{cases}$$

To see this from the Lagrangian perspective, consider that the maximum entropy solution will not prefer any particular graph with M edges over another, since all satisfy the constraint identically and no other constraints differentiate them. Hence, the final solution is uniform. This scenario is analogous to the microcanonical ensemble in statistical mechanics, where the "energy" (here, the number of edges) is fixed exactly.

Application: Derivation for the $G(n, p)$ Model (Soft Constraint) For the $G(n, p)$ model, we enforce a *soft constraint* on the number of edges. Instead of fixing $E(G) = M$ exactly, we require only that the *expected* number of edges equals M:

$$\sum_{G \in \mathcal{G}} P(G)E(G) = M,$$

while still requiring

$$\sum_{G \in \mathcal{G}} P(G) = 1.$$

Solving for $P(G)$ yields:

$$-\ln P(G) = 1 + \alpha + \sum_i \lambda_i C_i(G) \implies P(G) = e^{-1-\alpha-\sum_i \lambda_i C_i(G)}.$$

Define the normalization constant:

$$Z = e^{1+\alpha} = \sum_{G \in \mathcal{G}} e^{-\sum_i \lambda_i C_i(G)},$$

so that

$$P(G) = \frac{e^{-\sum_i \lambda_i C_i(G)}}{Z}.$$

The values of the Lagrange multipliers $\{\lambda_i\}$ and the partition function Z are determined by the specified constraints C_i^*.

Here, all graphs are technically possible, but their probability decays exponentially with the number of edges, controlled by the Lagrange multiplier λ. To find Z, sum over all possible graphs. There are $\binom{n}{2}$ possible edges, each of which can be present or absent. Thus:

$$Z = \sum_{G \in \mathcal{G}} e^{-\lambda E(G)} = \sum_{E=0}^{\binom{n}{2}} \binom{\binom{n}{2}}{E} e^{-\lambda E} = (1 + e^{-\lambda})^{\binom{n}{2}}.$$

Hence:

$$P(G) = \frac{e^{-\lambda E(G)}}{(1 + e^{-\lambda})^{\binom{n}{2}}}.$$

This probability factorizes over edges because the presence/absence of each edge can be considered independently:

$$P(G) = \prod_{\{i<j\}} \frac{e^{-\lambda x_{ij}}}{1 + e^{-\lambda}}.$$

Set $p = \frac{1}{1+e^{\lambda}}$. Then:

$$P(G) = p^{E(G)}(1-p)^{\binom{n}{2} - E(G)}.$$

The expectation constraint $\sum_{G \in \mathcal{G}} P(G)E(G) = M$ determines λ and hence p:

$$\mathbb{E}[E(G)] = p\binom{n}{2} = M \implies p = \frac{M}{\binom{n}{2}}.$$

Thus, the final form is:

$$P(G) = \left(\frac{M}{\binom{n}{2}}\right)^{E(G)} \left(1 - \frac{M}{\binom{n}{2}}\right)^{\binom{n}{2} - E(G)},$$

which matches the $G(n, p)$ model with $p = M/\binom{n}{2}$. This solution corresponds to the canonical ensemble in statistical mechanics, where the "energy" (number of edges) is specified only in

expectation, allowing for fluctuations.

Analytical limitations Maximum entropy models are a powerful tool for generating probabilistic ensembles of networks that satisfy predefined macroscopic constraints. By maximizing the entropy of the network ensemble while adhering to these constraints, such models ensure unbiased representations of the specified properties. However, their applicability is limited by certain analytical and practical challenges.

One significant limitation arises when the constraints are difficult to handle analytically. For instance, consider the problem of generating a network ensemble where the average global clustering coefficient is fixed. The corresponding constraint can be expressed as:

$$\langle C \rangle = \frac{1}{N} \sum_i c_i,$$

where c_i, the clustering coefficient of node i, is defined as:

$$c_i = \frac{2T_i}{k_i(k_i - 1)},$$

with T_i representing the number of triangles node i participates in, and k_i being the degree of node i. The number of triangles incident to a node i is evaluated with $T_i = 1/2\,(A^3)_i$, where A^3 is the cube of the adjacency matrix. Substituting this definition into the constraint yields:

$$\langle C \rangle = \frac{1}{N} \sum_i \frac{(A^3)_i}{k_i(k_i - 1)}.$$

This non-linear expression introduces significant analytical complexity, as the clustering coefficient depends not only on local properties like degree but also on higher-order correlations between nodes (*e.g.*, shared neighbors). Incorporating such constraints into a maximum entropy framework is challenging, as it requires solving a system of equations that intertwines local and global network properties.

In practice, networks generated using maximum entropy methods, such as the fitness-corrected block model or the degree-corrected block model [21], tend to exhibit low clustering coefficients, which may not accurately reflect the structural properties of many real-world networks.

To address the challenge of generating networks with tunable clustering coefficients or other non-linear topological properties, it becomes necessary to move beyond the maximum entropy framework. Alternative approaches, such as latent geometry models, offer a promising solution by embedding nodes in a geometric space where clustering emerges naturally

as a function of proximity. These methods, discussed in the next section, provide a flexible framework for constructing networks with realistic clustering and other desirable features, bridging the gap left by maximum entropy models.

Network with Latent Geometry

In this section, we introduce an alternative class of network models in which vertices lie in an underlying latent space and edges are more likely to form between nodes that are close in that latent geometry. Much like the Maximum Entropy viewpoint, these models leverage incomplete information (unknown latent coordinates) and simple constraints (closeness in geometric space) to produce flexible, probabilistic rules of edge formation.

Imagine you are organizing a social event where every guest has a "personal space" preference. These preferences are invisible but dictate how comfortable guests feel talking to each other. Two people whose personal space requirements align well (*i.e.*, they have similar "latent coordinates") are more likely to interact, whereas those whose preferences differ significantly may stay apart.

This analogy mirrors the essence of *latent-geometry* network models. Instead of personal preferences, each node has a position in a latent geometric space (for instance, on a circle $\mathbb{S}^1$ or in a hyperbolic disk $\mathbb{H}^2$). The probability of an edge between any two nodes depends predominantly on their distance in that latent space. This approach balances specificity—capturing structured, often community-like connectivity—against generality, by avoiding assumptions about explicit communities or top-down grouping. Like the raincoat analogy in the Maximum Entropy framework, we do not over-commit to a single structure (such as tightly predefined clusters); instead, geometry guides the likelihood of connections while allowing network-wide heterogeneity to emerge naturally.

The $\mathbb{S}^1$ Model The $\mathbb{S}^1$ model [178] is one of the simplest frameworks for constructing networks in latent geometric spaces. Each node i is assigned two attributes: a hidden degree κ_i, representing its popularity or intrinsic attractiveness, and an angular coordinate θ_i, which defines its position on a one-dimensional sphere (*i.e.*, a circle). The radius of the circle is set to $R = \frac{N}{2\pi}$, where N is the number of nodes. This choice normalizes the density of nodes on the circle to 1 without loss of generality [177].

The probability of a connection between any two nodes i and j is given by:

$$p_{ij} = \frac{1}{1+\chi_{ij}^{\beta}} = \frac{1}{1+\left(\frac{d_{ij}}{\mu\kappa_i\kappa_j}\right)^{\beta}},$$

where $d_{ij} = R\Delta(\theta_i, \theta_j)$ is the angular distance between the nodes on the circle, and $\Delta(\theta_i, \theta_j) = \pi - |\pi - |\theta_i - \theta_j||$ ensures periodicity. The parameters μ and β control the average degree and clustering coefficient of the generated network, respectively.

It is useful to highlight a property of $\mathbb{S}^1$ and other geometric network models: these models inherently promote high clustering coefficients compared to other random network models. This is a direct consequence of the triangular inequality: nodes that are close to each other in the latent space are more likely to share common neighbors, resulting in densely connected triads.

The connection probability follows a form akin to a gravity law, where the likelihood of an edge increases with the product of the hidden degrees ($\kappa_i\kappa_j$) and decreases with the angular distance (d_{ij}). In this framework, nodes with higher κ-values (greater popularity) are more likely to connect, even if they are angularly distant, whereas nodes with lower popularity are more likely to connect only when angularly close.

A key property of the $\mathbb{S}^1$ model is its ability to generate networks with tunable clustering. The parameter β, referred to as the inverse temperature, dictates the clustering behavior. For $\beta < 1$, the networks are unclustered, while for $\beta > 1$, the networks exhibit finite clustering in the thermodynamic limit. This structural phase transition at $\beta = 1$ marks a critical point where the topology of the resulting networks shifts dramatically [178].

The $\mathbb{H}^2$ Model The $\mathbb{H}^2$ model [107] is a purely geometric extension of the $\mathbb{S}^1$ model, where the popularity and similarity dimensions are integrated into a single distance metric in the hyperbolic plane. This distance governs the likelihood of connections between nodes, with hyperbolically closer nodes being more likely to connect. The hyperbolic plane's unique properties, such as exponential growth of circumference and area with radius, make it an ideal framework for modeling small-world, scale-free, and clustered networks.

In this model, the hyperbolic distance between two points, i and j, with radial coordinates r_i, r_j and angular separation $\Delta\theta_{ij}$, is computed using the hyperbolic law of cosines:

$$\cosh x_{ij} = \cosh r_i \cosh r_j - \sinh r_i \sinh r_j \cos \Delta\theta_{ij}.$$

For large r_i and r_j, this simplifies to:

$$x_{ij} \approx r_i + r_j + 2 \ln \frac{\Delta\theta_{ij}}{2}.$$

The connection probability between nodes is determined by a Fermi-Dirac function:

$$p_{ij} = \frac{1}{1 + e^{\frac{\beta}{2}(x_{ij} - R_{\mathbb{H}^2})}},$$

where β controls clustering, and $R_{\mathbb{H}^2}$ is the disk's radius, which scales with the number of nodes N as $R_{\mathbb{H}^2} \propto \ln N$.

Nodes are distributed within the hyperbolic disk according to a density function:

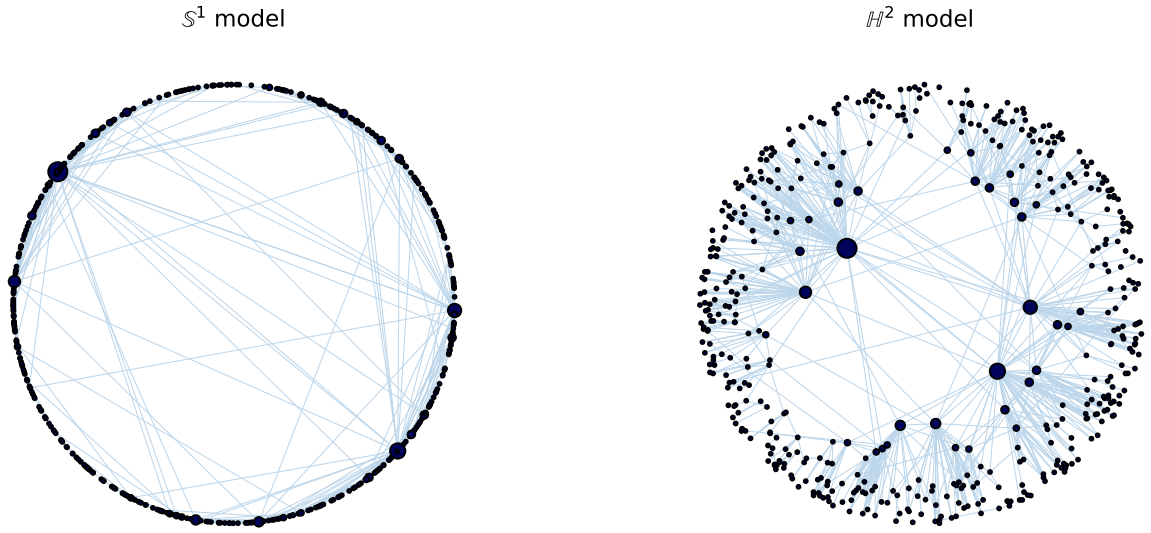
$$\rho(r) = \frac{\sinh r}{\cosh R_{\mathbb{H}^2} - 1},$$

ensuring homogeneity. High-degree nodes are placed near the disk's center, while low-degree nodes are closer to its periphery. This radial placement, combined with angular similarity, governs the network's topology.

Overall, the $\mathbb{H}^2$ model offers a powerful geometric framework for understanding the structural and functional properties of complex networks, including their navigability and clustering, while maintaining analytical tractability. Importantly, hyperbolic embeddings often capture hierarchical community structures: nodes near the origin ($r_i \approx 0$) are highly connected "hubs," and the angular dimension encodes more fine-grained "similarities." As a result, $\mathbb{H}^2$ networks can closely mimic real-world topologies, exhibiting strong clustering and broad degree distributions.

Quasi-Isometry of $\mathbb{S}^1$ and $\mathbb{H}^2$ Although $\mathbb{S}^1$ and $\mathbb{H}^2$ might seem fundamentally different (flat circle *vs.* curved hyperbolic disk), they are, in fact, *quasi-isometric* in many parameter regimes [177]. Concretely, the $\mathbb{S}^1$ model with certain radial scalings and a scale-free distribution of the fitnesses $\{\kappa_i\}$ can approximate much of the behavior of a hyperbolic disk of appropriate curvature. Conversely, a small radius hyperbolic model can appear very similar to a circle-based embedding. Formally, there exist mappings between $\mathbb{H}^2$ and $\mathbb{S}^1 \times \mathbb{R}^+$ that preserve distances up to bounded multiplicative factors. Hence, one can often translate results and inferences from one model to the other.

Soft Communities Latent geometry-based models, such as the $\mathbb{S}^1$ and $\mathbb{H}^2$ frameworks, provide a robust foundation for studying complex networks. These models embed nodes in a hidden geometric space, where the probability of connections depends on the geometric distance between nodes. A key feature of these frameworks is the spontaneous emergence of community-like structures, which arise naturally from the spatial arrangement of nodes without requiring explicit clustering mechanisms [105].



(a) Visualization of the $\mathbb{S}^1$ model: nodes are characterized by an angular position (representing similarity) and a fitness parameter (depicted by node size), influencing their likelihood of forming connections.

(b) Visualization of the $\mathbb{H}^2$ model: nodes are embedded in hyperbolic space, defined by an angular position and a radial coordinate, reflecting similarity and popularity, respectively.

Figure 2.9: Comparison of the $\mathbb{S}^1$ and $\mathbb{H}^2$ models. Both frameworks rely on latent geometric spaces to define the probability of edge formation, governed by node similarity and popularity metrics.

Traditionally, community structure in networks is identified through clustering and community detection algorithms that aim to partition or cover the graph into explicit subgroups (see, e.g., the comprehensive review by Fortunato [73]). These methods are generally based on the principle that communities are defined as sets of nodes with a higher density of internal connections compared to connections leading outside the group. Techniques such as hierarchical clustering, partitional clustering, and spectral clustering [219] work by defining similarity measures between nodes or groups, optimizing quality functions (like modularity [147]), or leveraging matrix properties to group nodes based on the network's structure. The outcome of these approaches is typically a discrete assignment of nodes to specific, often non-overlapping, modules, effectively enforcing a rigid partition of the network.

Unlike traditional approaches that enforce rigid partitions, these emergent communities reflect the intrinsic topology dictated by parameters such as degree distribution, clustering coefficients, and spatial constraints.

Nodes with similar angular positions or shorter hyperbolic distances tend to cluster into densely connected subgroups, mimicking real-world phenomena where entities with shared

characteristics naturally group together. These communities are often referred to as *soft communities* [220]. In the $\mathbb{S}^1$ and $\mathbb{H}^2$ models, such communities correspond to angular sectors characterized by high local clustering. These structures emerge as a byproduct of probabilistic connection rules rather than explicit design.

Empirical analyses highlight the robustness of these emergent communities. High modularity scores are consistently observed across various community detection algorithms, including Louvain, Infomap, and asynchronous label propagation. Furthermore, metrics like the angular separation index (ASI), specifically designed for hyperbolic networks, demonstrate that these communities occupy distinct angular regions with minimal overlap [105].

Garcia et al. [79] demonstrated that adopting a growing network paradigm, where link formation is governed by a competition between similarity and popularity, allows for the generation of networks with targeted topological features and soft communities. In this framework, nodes are sequentially added to the network, and connections are established based on a trade-off between proximity in a latent space (similarity) and the degree of existing nodes (popularity). In this dual mechanism hidden variables are now dependent and captures the hierarchical and modular structures often observed in real-world networks.

By sampling angular coordinates from non-uniform distributions, the sizes of communities can be effectively controlled, as shown by Muscoloni et al. [144]. For instance, denser regions in the angular space correspond to larger communities, whereas sparsely populated regions define smaller ones. This flexibility is particularly advantageous for modeling networks with heterogeneous community structures where group sizes vary significantly. However, in this method, communities are defined implicitly through their average angular coordinate, with the result that edges predominantly form within groups or between neighboring groups in the linear ordering of the angular space. This limitation stems from the nature of the angular distance metric, which inherently biases connections towards closer angular coordinates. Consequently, while the model effectively generates networks with localized clustering and modularity, it struggles to capture long-range inter-community edges often observed in empirical networks.

Despite their advantages, latent geometric frameworks lack general solutions for enforcing arbitrary mixing patterns. Recent studies have identified a dimensionality problem in these models [59]: while soft communities can be created by aggregating nodes within similarity space in D dimensions, the interplay between space dimensionality and community structure remains significant, and it is not yet clear whether increasing the size of the latent space can effectively control community structures. Addressing this limitation is important to advance the applicability of latent geometry-based models in scenarios where precise community control is needed.

Chapter 3

Exploring Aquatic Food Webs

Food webs, a specific type of ecological networks, model the flow of energy and matter in an ecosystem. In this section of the thesis, we analyze the structural properties of 173 aquatic food webs from freshwater, marine, and coastal ecosystems. This work was presented at the Computational Intelligence Methods for Bioinformatics and Biostatistics (CIBB 2023) [192], where further details are available. We investigate key network features such as core-periphery structure, which sheds light on the energy and matter flow from primary producers to apex predators. Additionally, we assess the robustness of these networks by examining their responses to species removal, identifying critical species that maintain the integrity of the ecosystem. We further explore the motif representation of these webs to reveal patterns of interaction at a mesoscopic scale, providing insights into the stability and resilience of ecological modules such as intraguild predation and competition. To the best of our knowledge, this represents the largest network analysis of aquatic food webs. The code and dataset used in this study are available at `https://github.com/davidetorre92/AquaNet`.

3.1 Foundational Concepts in Food Web Analysis

This section outlines concepts and terminologies foundational to the analysis of food webs. Nodes in the networks of the dataset of this study represent either living species or non-living pools of organic material. The edges are associated with trophic links and describe the flow of energy and matter between nodes.

These definitions are integral to the subsequent discussion of this chapter, where they are applied in the context of specific research questions and analyses. The following paragraphs provide a comprehensive examination of each term employed in this study.

Taxa richness Taxa richness refers to the number of nodes (S) in a food web [190]. These nodes extend beyond individual biological species: they include taxonomic groups at varying levels of aggregation, from single species to entire kingdoms (*e.g.*, mites or invertebrates); mixed groups or communities (such as zooplankton or bacterial communities); specific life stages of organisms (*e.g.*, juvenile or adult bass); parts of organisms (such as leaves or roots); and non-living organic components (like detritus; particulate organic carbon, POC; or dissolved organic carbon, DOC). While the term "taxa" traditionally refers to classifications of living organisms, its broader application here includes both living and non-living nodes for consistency with established definitions in food web studies (see *e.g.*, [190] Box 1). To minimize confusion, we distinguish between *living nodes*, representing living organisms, and *non-living nodes*, representing non-living organic components. This distinction becomes particularly important in specific analyses, such as the identification of critical node sequences or the study of three-node motifs, where the focus is typically restricted to living nodes. Retaining the term "taxa richness" while clearly defining its broader scope ensures consistency with the literature while accommodating the unique composition of the dataset.

Stock Density Stock density refers to the quantity of a specific medium, such as carbon or nitrogen, present within an ecological network, expressed per unit area or volume (*e.g.*, $\mathrm{mgC} \cdot \mathrm{m}^{-2}$, milligrams of carbon per square meter) [195]. In food web models, this measure includes the biomass of living organisms as well as the mass of non-living components, such as detritus or particulate organic matter.

Number of Trophic Links The number of trophic links (L) represents the exchange of matter or energy within an ecosystem [190]. It corresponds to the total number of directed edges between nodes in a food web, representing pathways through which matter and energy flow within an ecological community. In our study, these interactions include *grazing*, defined as the transfer of matter from primary producers (e.g., plants or algae) to herbivores through the grazer pathway; *predation*, describing the consumption of one organism by another at higher trophic levels; and *cannibalism*, where individuals of the same species prey on one another. The food web also accounts for *decomposition*, where matter flows from organic detritus to decomposers and detritivores, driving the recycling of nutrients; *autochthonous* flow of matter into detritus pools arising from the natural death of organisms and unassimilated prey [142]; and flow of matter and energy between non-living nodes due to *abiotic* factors such deposit of suspended matter [142]. In contemporary food web models, trophic links often quantify the flow of matter or nutrients between species. This flow is typically expressed as a rate of stock density transfer over time, measured in units such as $mg \cdot C \cdot m^{-3} \cdot \mathrm{yr}^{-1}$

(milligrams of carbon per square meter per year).

Matrix of interactions In our dataset, detailed information on the stock density of nodes and their flow is not always available. To address this, we adopt two matrices: the binary interaction matrix a, where $a_{ij} = 1$ denotes the presence of a link from i to j, and the weighted interaction matrix w, where w_{ij} quantifies the magnitude of the interaction (*e.g.*, biomass flow). In cases where weight data is unavailable, we assume $w_{ij} = a_{ij}$ for all pairs (i, j), treating the interaction as unweighted.

Connectance A measure of network complexity, representing the fraction of realized trophic links (actual interactions) out of all possible links between species [65]:

$$C = L/S^2 \tag{3.1}$$

Trophic Level The trophic level of a species defines its position within a food web [191]. *Primary producers*, such as plants and other basal resources [208], occupy the lowest trophic level, as they do not consume any other species. At the other end of the spectrum, *top predators*, which lack predators themselves, represent the highest trophic level in the web.

The concept of trophic level was first introduced by Lindeman in 1942 [115], who defined it in discrete terms: basal species occupy the first trophic level, herbivores the second, and successive levels are assigned to higher-order consumers in integer increments. This approach, based on the transfer of energy and matter through successive trophic levels, provided the foundation for much of modern ecosystem ecology and emphasized the role of energy flow as the driving mechanism behind food web dynamics.

While Lindeman's definition remains influential, more refined methods have been developed to accommodate the complexity of real-world food webs. In this thesis, we adopt the prey-averaged trophic level [208]:

$$TL_i = 1 + \frac{1}{\sum_j w_{ji}} \sum_j w_{ji} TL_j, \tag{3.2}$$

which iteratively calculates the trophic level TL_i of a species i as the *weighted average* of the trophic levels of its prey (j), plus 1. Basal species and *non-living nodes*, by definition, are assigned a trophic level of 1 [208].

This prey-averaged method provides a finer resolution compared to Lindeman's discrete approach, capturing complexities such as mixed diets and omnivory. It is particularly suitable

for our dataset, as many of the food webs lack link-strength information, and the prey-averaged method can be applied to food webs with binary interactions [208]. By using this definition, we capture the nuanced trophic dynamics inherent in real-world ecosystems while preserving compatibility with datasets of varying levels of detail.

Three-node motifs A three-node motif is a connected subgraph M of a graph G with three nodes and a specific edge arrangement. For directed graphs, there are 13 unique configurations of three-node motifs. Among these 13 possible interactions, four have been extensively studied in past decades.

Tri-trophic chain can be represented as $M = \{(u, v), (v, w)\}$, where $u \to v \to w$. It represents a simple linear food chain among three species, where the top predator (w) preys upon an intermediate species (v), which, in turn, consumes the basal species (u).

Intraguild predation can be represented as $M = \{(u, v), (v, w), (u, w)\}$, where $u \to v \to w$ and $u \to w$. It occurs when a top predator (w) consumes both an intermediate predator (v) and a basal prey species (u).

Apparent competition can be represented as $M = \{(u, w), (v, w), (u, w)\}$, where $u \to w \leftarrow v$. It occurs when two prey species (u and v) are hunted by a shared predator (w). In this scenario, an increase in one prey species can inadvertently increase predation pressure on the other, as the shared predator's population may rise due to the abundance of food.

Exploitative competition can be represented as $M = \{(u, v), (u, w), (u, w)\}$, where $v \leftarrow u \to w$. It is a motif where two predator species (v and w) compete for the same basal prey (u) without directly interacting with each other. The predators indirectly affect one another by consuming shared resources.

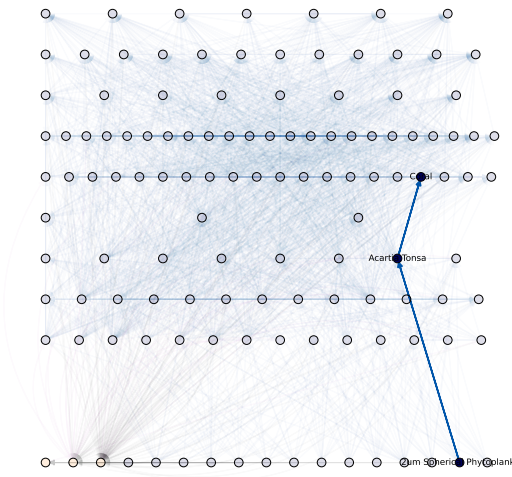
These three-node motifs, as depicted in Figure 3.1.

3.2 Data collection

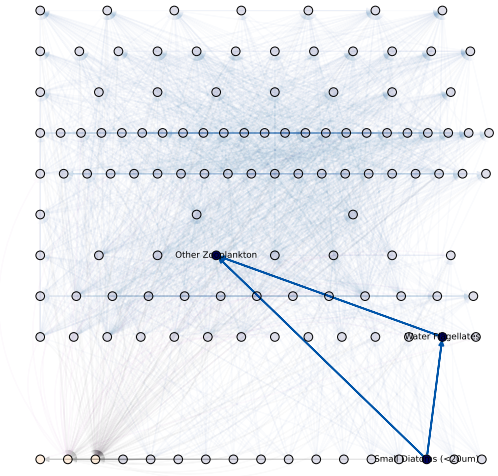
For this study, we collected a dataset of 173 aquatic food webs. Of these, 18 were sourced from the `igraphdata` library [50], 12 were manually extracted from [47], and 2 were directly obtained from published papers [123, 96]. The remaining 141 food webs were downloaded from the Ecopath With Ecosim repository [49] using the API available at `https://ecobase.ecopath.org/`.

All food webs are provided in XML `graphml` format and are accessible at `https://github.com/davidetorre92/AquaNet/`. A summary of the most representative metrics for these webs is available in Table 3.1

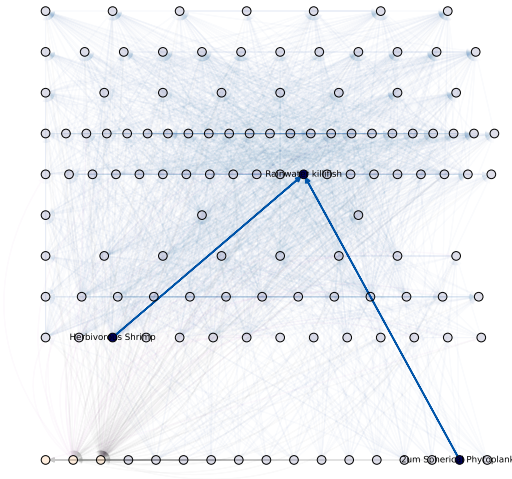
We categorized the food webs into six ecosystem types: lagoons/coasts, gulfs, lakes/ponds,



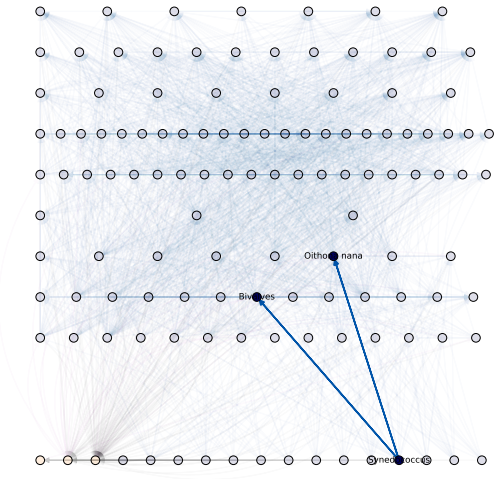
(a) Tri-trophic chain: Phytoplankton is eaten by Acartia Tonsa (a marine copepod) which is captured and consumed by Coral.



(b) Intraguild predation: Phytoplankton is eaten by Water Flagellates which is eaten by Zooplankton, that eats on Phytoplankton too.



(c) Apparent competition: Rainwater killifish, the top predator, eats both Herbivorous Shrimp and Phytoplankton which affect each other indirectly by being prey for the same predator.



(d) Exploitative competition: Oithona nana and Bivalves interact indirectly as they compete for common resources, Synedococcus.

Figure 3.1: Example of three-node motifs in the Chesapeake Bay Mesohaline [7] food web model.

open sea, rivers/streams, and swamps/marshes. The ecosystem type for each food web was inferred from the model's name or by checking in the original paper that introduced the

web, the specific type of ecosystem it represents. For example, the model named *Gulf of Mexico (1950)* was categorized under the *gulf* ecosystem, while *Charca de Maspalomas* was assigned to *lagoons*, and the two *Cypress* models (wet and dry seasons) were categorized as *swamps*. This classification resulted in 96 open sea ecosystems, 43 coast ecosystems, 11 gulfs, 11 rivers, 6 lakes, and 6 swamps in our dataset.

In the food webs manually extracted from [47, 123, 96], no information was available regarding the flow from living nodes directed to non-living nodes, that is the flow of energy and matter resulting from the natural death of organisms and unassimilated prey. To ensure consistency with food webs derived from these sources, we established links from living nodes to non-living nodes where appropriate. However, to align with models available in [50] and [49], we examined whether all living nodes were indeed connected to the non-living nodes. Upon investigation, we found that living nodes representing birds, as well as those for primary producers such as phytoplankton, bacteria, and plants, didn't have any links connected towards the non-living nodes. Based on this observation, we chose not to include these links, thereby maintaining consistency with the rest of the dataset while respecting the structural nuances of the original sources.

Metric	**Mean**	**Std Dev**	**Min**	**Max**
Taxa richness (S)	41.77	23.12	18	182
Number of trophic links (L)	424.94	448.04	55	2614
Connectance L/S^2	2.01×10^{-1}	7.55×10^{-2}	2.63×10^{-2}	3.85×10^{-1}
Basal species	4.2	7.86	1	90
Non-living nodes	1.50	0.9	1	7
Largest strongly connected component proportion	0.87	0.12	0.33	1.0
Max trophic level	4.41	0.48	3.14	5.67

Table 3.1: Food webs metrics in the dataset.

3.3 Core-periphery structure

The core-periphery structure is a basic pattern in a network, marked by two distinct components: a "core" with closely linked nodes, and a "periphery" where nodes are loosely connected, both to the core and to each other [76]. In ecological networks, the core-periphery structure is linked to the concept of nestedness (see, *e.g.*, [139]). This arrangement plays an important role in the network adaptability, enabling them to cope with environmental variations and internal disturbances. To identify this structure in aquatic food webs, we classified *core* nodes as those in the largest strongly connected component of the graph, with

Figure 3.2: Global map depicting the approximate locations of 173 food webs. Markers found in a continental zone indicate a food web that relates to one of the coasts of that region.

all other nodes designated as the *periphery*. Allesina et al. in [4] analyzed the flow of energy in food webs by studying the strongly connected components. Following a standard terminology (see, *e.g.*, [30]), the periphery can be further decomposed into five distinct subsets: the *IN-periphery*, consisting of all nodes that can reach the core but cannot be reached from the core; the *OUT-periphery*, consisting of nodes that are reachable from the core but cannot reach the core; the *Tubes* structure, that consists of nodes that do not interact with the core and that connect the IN-periphery to the OUT-periphery; the *Tendrils-IN* structure, consisting of nodes that are reachable from the IN-periphery, but do not interact with the core; the *Tendrils-OUT* structure, consisting of nodes that reach the OUT-periphery but do not interact with the core; and all the remaining nodes, which are not connected with the previous nodes and are thus referred to as *other* nodes. Figure 3.3 presents a schematic representation of this type of structure with the aforementioned definitions.

We computed the core-periphery structure for all the food webs in our dataset. In Figure 3.4, we present for each food web the proportion of nodes in the core and each peripheral

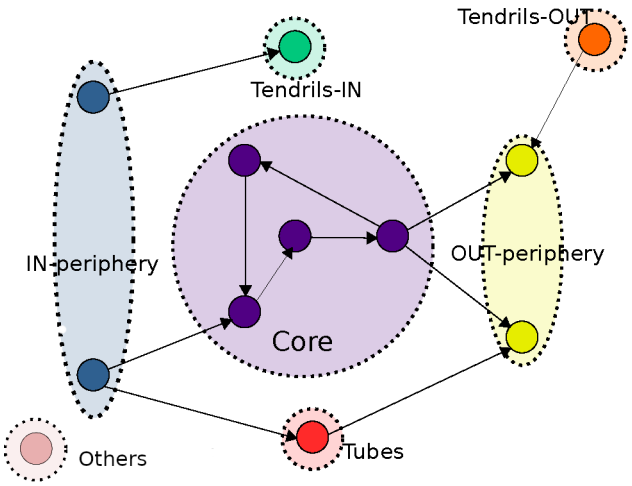


Figure 3.3: A sketch of the core-periphery structure. Core nodes belong to the giant strongly connected component, while other nodes belong to the periphery.

structure. Our analysis shows that the networks in our dataset share a very similar structure: a large core in which compartments are highly interconnected and a smaller IN-periphery that carries energy and matter to the core nodes. More than 75% of the analyzed food webs have a core that contains over 85% of their nodes. However, there is variability in the size of the core. A small subset of food webs, specifically 2.9%, present a core that includes only $40 - 50\%$ of the nodes.

The nodes in the IN-periphery correspond mainly to producers which supply with their energy and matter the strongly connected component. Producers derive their energy from sources like photosynthesis and feed energy and matter into the system through herbivores. In our food webs, producers are defined as nodes at the lowest trophic level with an in-degree equal to zero (i.e., they do not receive energy from other species) and an out-degree greater than zero. Besides producers, the IN-periphery includes plankton communities (in various aggregations), small fishes, and invertebrates, which export energy and nutrients to species with higher trophic levels. Occasionally, this structure also contains non-living nodes (*e.g.*, Sediment POC in the Upper Chesapeake Bay model, DOC in the St. Marks River model, and Fishery Offal in the Northern Californian Current model). These nodes are coupled with living organisms that recycle nutrients most of times into the core. On the other hand, nodes in the OUT-periphery represent consumers that feed directly on the core of the food web. These nodes are limited in number (24 in the dataset) and mainly consist of birds (such as gulls, kingfishers, and various raptors), turtles, jellyfish, and large fish such as whales, sharks, and killer whales.

As for the other peripherical structures, we recall that Tendrils-IN, Tendrils-OUT and Tubes contain species that neither reach the core nor are reached from the core and do not

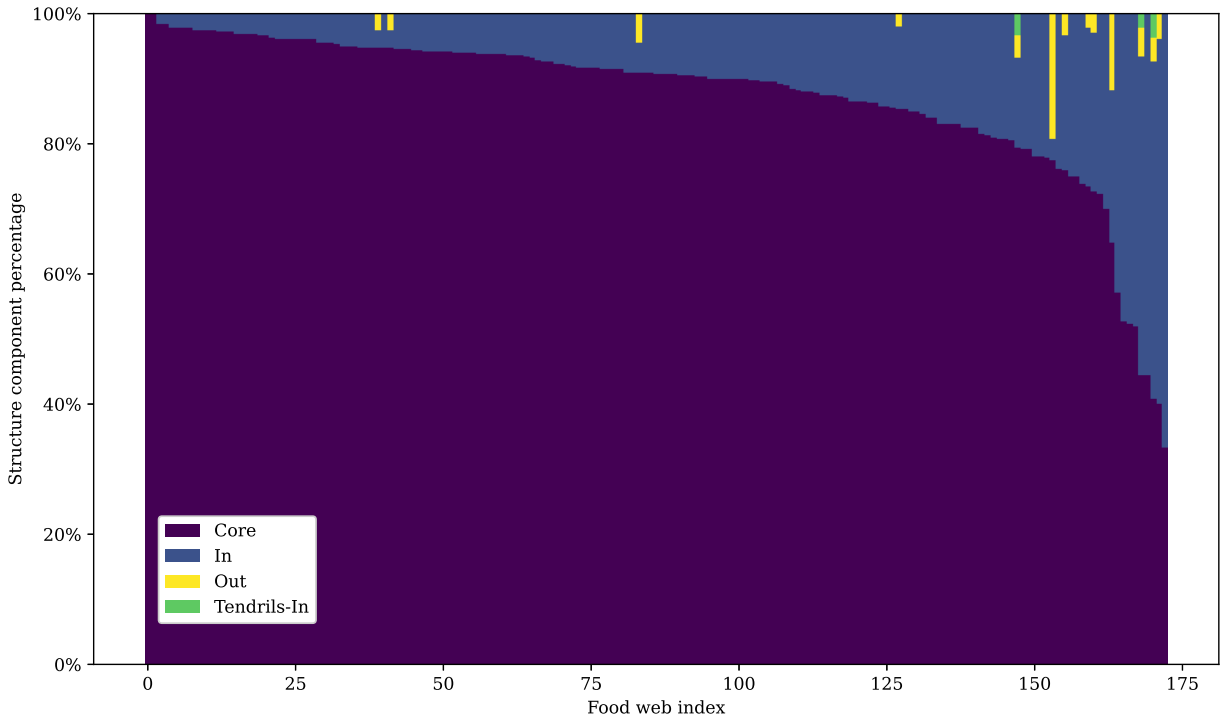


Figure 3.4: Proportion of core periphery relative to the food web size. The x-axis indicates the index of the food web, while the y-axis indicates the proportion of the nodes belonging to the the relative structure component. The food webs in the dataset are sorted according to their core size proportion.

seem to have a strong overall influence on the network. In our datasets only 3 food webs contain Tendrils-IN structure and none of the them contains Tubes and Tendrils-OUT. The three food webs that contain Tendrils-IN along with the species that form these structure, are as follows: in the Lower Chesapeake Bay, rotifers feeding on phytoplankton; in the Tasek Bera swamp of Malaysia, swallows feeding on various organisms of the IN-periphery; and in the Salt meadow, New Zealand, redpolls, feeding on seeds.

Our datasets include 22 networks with information spanning various contexts. For some networks, data has been collected over multiple years, while for others, we have data covering both the wet and dry seasons. We analyzed the differences in species composition, links, and the composition of core and periphery in these networks to test if this structure remains constant over time. In the six food webs modeling the same ecosystems across two different seasons—Cypress, Mangrove Estuary, and Florida Bay—we found that while the species remained consistent between the dry and wet seasons, there were minor differences in the links. However, the overall core and periphery structure remained unchanged. The same applies to the remaining 16 networks modelling ecosystems in different years.

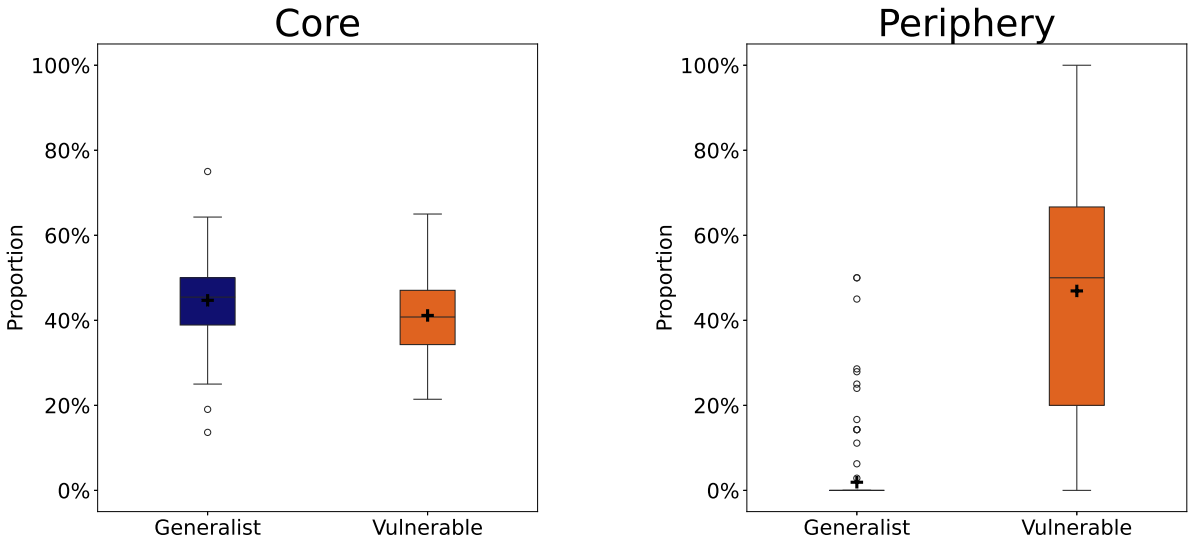
3.3.1 Horizontal and vertical diversity of core and periphery

Biodiversity is another crucial factor affecting the functionality and stability of food webs [126]. The biodiversity of food webs is composed of horizontal (*i.e.*, within trophic levels) and vertical diversity (*i.e.*, the number of trophic levels) [199] and they are believed to have opposing effects on stability. Vertical diversity tends to reduce stability in simple food chains by lengthening recovery times, suggesting why natural food webs have few trophic levels [171]. Whereas, horizontal generality, which usually considers the measure of generality and vulnerability of the species, is believed to influence positively the stability of the network [174]. Indeed, generalists, with their broad diets, can provide robustness to a food web by filling ecological gaps when certain species are removed or environmental conditions change [46, 149].

We define now formally these concepts. Consider a food web represented by its adjacency matrix a, that is $a_{ij} = 1$ if the is a directed edge from i to j, and $a_{ij} = 0$ otherwise. Given a species i, the amount $\sum_{j=1}^{S} a_{ji}$, is an indicator of the generality of i and measures the breadth of its diet, that is, the set of relationships that a consumer has with its prey. The vulnerability of i, is the amount $\sum_{j=1}^{S} a_{ij}$, and measures the number of consumers that prey upon this species. Note that in graph-theoretical terms, the generality of a node corresponds to its in-degree, while the vulnerability corresponds to its out-degree. To be able to compare these quantities across different networks, they are usually normalized by the average number of links per species L/S. Formally, to measure the generality and vulnerability of a species i in the food web G we use the following definitions provided in [207]: $Gen_i = \frac{S}{L} \sum_{j=1}^{S} a_{ji}$ and $Vul_i = \frac{S}{L} \sum_{j=1}^{S} a_{ij}$. A species i is a generalist if $Gen_i > 1$ and non-generalist otherwise, and vulnerable if $Vul_i > 1$ and non-vulnerable otherwise.

To better understand the food web robustness we explore the vertical and horizontal diversity of the core and periphery of food networks. In examining the horizontal diversity, we focus on how generalist and vulnerable species are distributed across the cores and peripheries of each food web. Figure 3.5a illustrates that in the core of food webs, the proportions of generalist and vulnerable species are quite similar, with generalists making up an average of 42.4% and vulnerable species 39.8%. Conversely, Figure 3.5b reveals that, with few exceptions, food web peripheries predominantly consist of vulnerable nodes. Specifically, on average, peripheries contain 1.8% generalist species and 48.1% vulnerable nodes. We reported here the proportions of species categorized as generalists and vulnerable, considering only the living nodes. An alternative approach would involve calculating the generality and vulnerability of species in a subgraph of the food web that includes only living nodes, excluding non-living nodes and their interactions. This adjustment shifts the proportions slightly: generalist compartments in the core rise to 48.6% (from the previously observed

42.4%), vulnerable species in the core increase to 40.5% (from 39.8%), while generalists in the periphery rise to 2.5% (from 1.8%) and vulnerable species reach 52.3% (up from 48.0%). These findings align with previous observations, although excluding non-living nodes may slightly affect the observed distribution of generalist and vulnerable species across core and peripheral network structures.



(a) Proportion of generalist (mean = 42.4%) and vulnerable (mean = 39.8%) living nodes in the core of food webs.

(b) Proportion of generalist (mean = 1.8%) and vulnerable (mean = 48.1%) living nodes in the peripheries of food webs.

Figure 3.5: Comparison of generalist and vulnerable species in the core (left) and periphery (right) of food webs. The '+' sign indicates the mean value across the dataset. The percentages represent the average proportion of each species type across all analyzed food webs.

Regarding vertical diversity in each food web, we calculated the maximum trophic level across the entire food web, as well as within its core and periphery. Figure 3.6 presents the distribution of these maximum trophic levels. The figure clearly shows that species with the highest trophic levels are predominantly found in the core, while those with lower trophic levels belong to the periphery. A compelling explanation for this observed pattern can be found in the structural composition of the peripheries - the subgraphs existing outside the core network. These peripheral regions are predominantly constituted by the IN-periphery, where organisms functioning as primary energy and matter suppliers to the core are situated. Due to their fundamental position in the food web, these species necessarily maintain lower trophic levels. The statistical outliers observed in our analysis, characterized by notably elevated trophic levels, are primarily associated with the OUT-periphery. This distinct region includes all those species that facilitate the export of energy and matter from the core

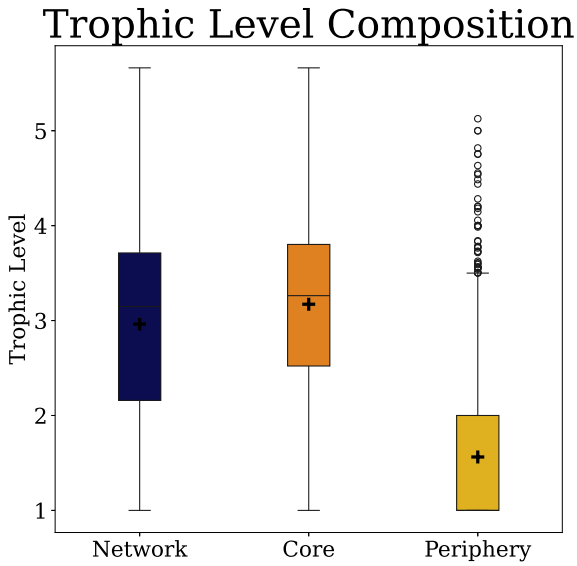


Figure 3.6: Distribution of trophic level across the entire network, core, and periphery subgroups of the food web. The symbol '+' denotes the mean trophic level for each group. Outliers are represented as individual points.

network, thus naturally occupying higher trophic level in the food web.

3.3.2 Flow of energy and matter

In this section, we classify the diet composition of each species in the various peripheries of the food web. We explore the core and periphery of aquatic food webs by systematically classifying the links between and within the structures. To achieve this, we classified all the links in the 173 food webs according to two criteria: the structures they connect (core, IN-periphery, OUT-periphery, or Tendrils-IN) and the type of node (living or non-living nodes) that serves as the source and target. To provide a sense of scale, our dataset contains a total of 73,514 links across all 173 food webs.

The data is organized in Table 3.2, where columns represent links between structures and rows represent links between types. Links between living nodes ($L \to L$) represent energy exchanges through strategies such as predation and herbivory. Links from living to non-living nodes ($L \to D$) describe a flow of energy and matter that comes from the living tissues of an organisms into a pool of non-living matter modeling the transfer of matter from living species to detrital pools after the death of organisms or from unassimilated prey [142, 65]. Links from detrital to living nodes ($D \to L$) represent the consumption of non-living organic material by living organisms including detritivores, which are the organisms that recycle

nutrients into the ecosystem. Finally, links between non-living nodes ($D \to D$) represent further decomposition within the detrital pool caused by abiotic factors such as deposition of suspended organic matter into sediment.

Links	$IN \to IN$	$IN \to C$	$IN \to TIN$	$C \to C$	$C \to OUT$	$OUT \to OUT$	$IN \to OUT$	**Total**
$L \to L$	1020	6483	6	54774	142	7	13	62445
$L \to D$	59	968	0	6970	0	0	0	7997
$D \to L$	29	216	0	2743	3	0	0	2991
$D \to D$	5	26	0	50	0	0	0	81
Total	1113	7693	6	64537	145	7	13	73514

Table 3.2: Link classification by food web structure, and matter and energy exchange interaction. Legend: C: core, IN: IN-periphery, OUT: OUT-periphery, TIN: Tendrils-IN. L living nodes, D non-living nodes, *e.g.*, detriti.

Regarding the links connecting species within the IN-periphery, we found that these links typically connect species at lower trophic levels to those at higher trophic levels, as defined in Equation 3.2. Links connecting the IN-periphery to the core total $7,693$, with $6,483$ (84.27% out of all the links from the IN-periphery to the core) being $L \to L$ links. This indicates that energy and nutrients predominantly flow between the IN-periphery and the core through herbivory and predation. Another significant contributor to the energy and matter flow between these two structures are the ($L \to D$) process from the IN-periphery to the core, with 968 interactions (12.58% out of all the links between IN-periphery and core), and $D \to L$ links, accounting for 216 interactions (2.81%). The remaining interactions involve exchanges between non-living nodes. The links within the core are $64,537$, accounting for 87.79% of the total dataset. Among these, $54,774$ (84.87% out of all the links within the core) are $L \to L$ interactions, $6,970$ (10.80% out of all the links within the core) are $D \to L$; $2,743$ (4.25% out of all the links within the core) are $L \to D$, and 50 (0.08% out of all the links within the core) are $D \to D$. Notably, of all $7,997$ $L \to D$ links, $6,970$ are within the core, indicating that most flow of matter due to death of organisms or from unassimilated prey occur in the core. From the core to the OUT-periphery there are mostly $L \to L$ relations, and between the OUT-periphery there are only $L \to L$ relations.

These measurements help to understand the organization of food webs: the core not only hosts the majority of species within a food web but also serves as the primary site for predation and the recycling of organic matter.

If we could trace a volume of matter from producers to OUT-periphery species, it would start in the IN-periphery, moving through this structure via herbivores and into the core through predation. Within the core, the volume of matter undergoes various transformations, with some of it being recycled by decomposers. Finally, it is exported by OUT-periphery predators, exiting the ecosystem cycle [181, 37]. The sketch in Figure 3.7 provides a depiction

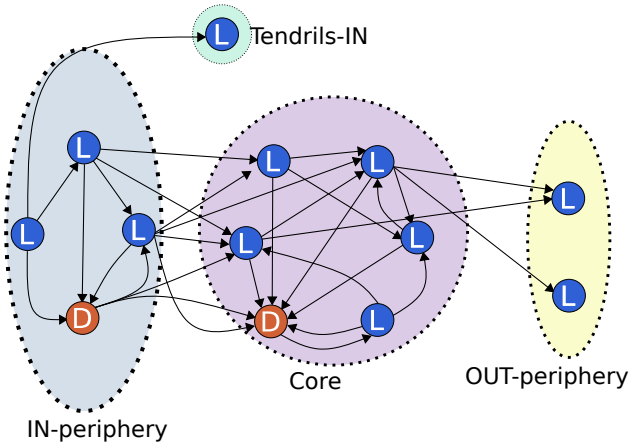


Figure 3.7: A sketch of the core periphery structure and the energy flow. The structures are coloured purple for the core, blue for the IN-periphery, yellow for the OUT-periphery and green for the Tendrils-In. Nodes with an orange fill, represent non-living nodes (*e.g.*, detriti, debris, DOC). living nodes are indicated with a capital L.

of the core-periphery structure of aquatic food webs and their peripheries.

3.4 Identification of critical nodes

Aquatic ecosystems experience loss of population and species, most of the times with unknown consequence in terms of resource use (that is, how organisms in an ecosystem use available resources such as light, nutrients, water, and food), primary production (that is, energy from the sun is converted into organic matter by autotrophic organisms), secondary production (that is, generation of biomass by heterotrophic organisms) [211]. The cause of these changes include direct human interventions such as fishing [211] or seabed mining [117], or from climate change [164], or other events such as the invasion of an alien species into an ecosystem [13]. Understanding and predicting the relationship between the topological structure of ecological networks and extinction patterns or cascading effects is important at this time because it helps guide conservation efforts and mitigate the impacts of biodiversity loss in the face of accelerating environmental changes [167]. In this section we evaluate the impact of living nodes on the stability of each food web.

Since in general this problem is hard (see, *e.g.*, [154]), this evaluation is done empirically, by assessing how much the removal of a node (corresponding to the disappearance of a species in the ecological network) may disrupt the structure of the network. The removal of a node from a food web corresponds to multiple scenarios: it may represent the extinction of a species, its migration, or a change in its function within the ecosystem [197, 180]. This can lead to cascading effects, such as the loss of a key predator or prey altering the structure

and function of the entire ecosystem. The impact may vary depending on the role of the removed node, potentially causing shifts in species abundance, changes in nutrient cycling, or even the collapse of the food web's [197]. This leads to the following natural notion of robustness of the network: if removing a few nodes has no noticeable impact, then the network structure is robust. On the other hand, if removing a few nodes quickly affects the underlying structure of the network, then the network is more fragile. Currently, the most used criteria are based on: (1) the most connected species; (2) randomly chosen species; (3) the most connected species excluding primary producers; and (4) the least connected species [66, 67, 149, 19]. In these studies the robustness is defined as the fraction of species that had to be removed in order to result in total species loss (i.e. primary species removals plus secondary extinctions) of $\geq 50\%$ of the species in the original web. Similarly, Allesina et al. in [5] proposed another method in which nodes are gradually removed according to their degree in the generalized multiple dominators graph in which the resource nodes represent are the root of the dominator graph and an edge between two nodes represents a domination relationship. Our approach differs from that seen in the past because we look for which species have a disruptive effect globally on the entire network, rather than picking a sequence based upon the *local* features of a node. To identify the presence of such species, we search for that sequence of nodes that results in the highest minimization of the *reachable pair fraction* upon their removal. This quantity is defined as [25]:

$$r(G) := R(G)/S^2 \tag{3.3}$$

where $R(G)$ are the reachable pairs of the graph G, *i.e.*, the number of pairs (u, v) such that there exists a directed path from u to v, and S is the number of nodes of the graph G. This quantity is normalized after dividing by the number of all possible pairs of nodes, S^2. A path from u to v indicates that a fraction of the energy originating at u can reach v (see, for example, [195], section 3 on Indirect Effects). The reachable pair fraction defined in Equation 3.3 is capable of capturing cascading effects across trophic levels in a food web: when a species is removed, the disruption can propagate *indirectly* through multiple trophic levels, affecting species that are not directly connected to it. This metric, therefore, quantifies the extent of these potential cascading effects by measuring how the network withstands species loss.

3.4.1 Notion of robustness

In the context of mutualistic ecological networks, Sheykhali et al. [179] provided a notion of robustness by measuring the area under the curve of the relative size of the largest connected

component vs. the fraction of nodes that have been removed in the web. Inspired by this idea, we established a notion of robustness by measuring the area under the $r(G)$ vs. the fraction of removed nodes. Consider a sequence of nodes in a network, denoted as $CSeq_G = \{s_1, \ldots, s_i, \ldots, s_{|S_G|}\}$. We define

$$\rho_G = \frac{1}{|S_G|} \sum_i r(G^{(i)}) \tag{3.4}$$

where $G^{(i)}$ is the graph obtained after removing from G the edges incident to the vertices $\{s_1, \ldots, s_i\}$. We set $G^0 = G$. Notice that, in this context, we are looking for a sequence of critical nodes of living nodes. Therefore, S_G in Equation 3.4 is the number of living nodes of the graph G and the sequence of nodes $CSeq_G$ is a set of nodes containing all living nodes of the graph G.

3.4.2 Description of the algorithm for computing the robustness

To determine the robustness ρ_G of a food web G, we need to find the sequence of most critical nodes. More specifically, given a food web $G = (S, L)$, the goal is to find a sequence of living nodes, $CSeq_G$, whose removal minimizes $\rho(G)$. We proceed in a greedy manner. Starting from an empty sequence of nodes $CSeq_G$, for each living node s of G, we measure $r(G^{(1)})$ where $G^{(1)}$ is obtained from G by removing all the links that are incident to s. We add to $CSeq_G$, the node that minimizes the quantity $r(G^{(1)})$. In case of ties, i.e., if there is more than one node minimizing $r(G^{(1)})$, we choose the node with the lowest trophic level defined in Equation 3.2. This choice reflects our focus on identifying nodes critical for maintaining bottom-up energy flows in the network, as species with lower trophic levels often act as primary pathways for energy transfer from basal resources to higher levels. We continue this process until there are no more links incident to the nodes representing living nodes. Once the sequence is set, we compute ρ_G accordingly using Equation 3.4.

We computed the robustness of all the food webs in our dataset. Table 3.3 presents the networks arranged in decreasing order of robustness. Furthermore, for each network, we provide the first three nodes found in the sequence, corresponding to the top three most critical nodes according to our algorithm.

We investigated whether the robustness ρ_G is correlated to other measures of the food web such as the size of the network, the fraction of the core calculated in the previous section, and also with the connectance, defined as the proportion of possible links (S^2) that are realized (L) [65]: $C = L/S^2$. The measure of robustness of a food web appears to be positively correlated (Pearson correlation = 0.60) with the proportion of nodes belonging to the core,

Rank	Food Web	$[s_1^G, s_2^G, s_3^G]$	ρ_G
1	Sand Beach, South Africa	Gastrosaccus, Callianassa, Donax; bacteria; Cumacea	0.369615
2	Sítios Novos Reservoir (2011)	Camarão; Cladocera; Copepoda	0.335623
3	Terminos Lagoon (1980)	Meiofauna; Polychaetes; Zooplankton	0.321184
...	...	...	...
171	Chesapeake Bay Mesohaline	bacteria in sediment poc; zooplankton; blue crab	0.063669
172	St. Marks River (Florida)	Benthic bact; Bacterio plankton; Micro protozoa	0.061475
173	Salt Meadow, New Zealand	amphipods; collembola; mites	0.043401

Table 3.3: The food webs are sorted in non-increasing order of robustness, with the highest values corresponding to the top rows. The first column corresponds to the rank in the ordering. The third column contains the sequence of the most critical nodes.

and the connectance (Pearson correlation = 0.59). This measure is not directly correlated with other simple food web metrics, such as the taxa richness (Pearson correlation = -0.22), the number of trophic links (Pearson correlation = 0.01), the maximum trophic level (Pearson correlation = -0.27). This finding is in agreement with Dunne et al. in [66]: our definition of robustness ρ_G increases with the connectance of the network. However, the authors' strategy for node removal involved starting with the species that had the highest number of trophic links to other species and then measuring the number of secondary extinctions. In contrast, our approach measures a global connectivity metric of the network, as defined in Equation 3.4. The resulting sequence of nodes differs from a simple sorting based on in-degree or out-degree alone. To quantify this difference, we calculated the Kendall rank correlation coefficient index between the sequence of nodes obtained by our method and those ordered by out-degree for each network (i.e., the species with highest number of trophic links to other species). Averaging these values yielded 3.38×10^{-2}, indicating a lack of correlation between the node sequences.

Moreover, the critical nodes identified by our algorithm for computing the robustness of food webs also appear to be biologically significant for their stability. In the Sand beach, South Africa food web, shrimp-like species and clams significantly contribute to the biota, as demonstrated by their rich concentration, serving as vital food sources for many fish species [32]. In the St. Mark River food web, benthic microbial communities are crucial for sustaining connectivity by driving organic matter decomposition, nutrient regeneration, and

influencing dissolved oxygen levels [150]. Similarly on the Salt meadow, New Zealand, food web, small arthropods like amphipods, collembola and mites which occupy both aquatic and terrestrial habitats, have a significant role in the ecosystem, as either detritivores or herbivores [155, 121].

Importantly, the method illustrated here has a tendency to detect communities of bacteria and plankton as the most critical nodes. This phenomenon can be attributed to the role these nodes play in the transformation of resources from non-living components, such as detrital matter, particulate organic matter, and dissolved nutrients, into bioavailable forms accessible to living organisms, thereby maintaining the integrity of the trophic structure [180, 213]. The disruption of edges connected to these nodes has cascading effects on the recycling mechanisms, effectively severing the pathways through which resources are reintroduced into the ecosystem.

The importance of microbial communities in aquatic ecosystems has been extensively documented, and it has been identified that factors adversely affecting microbial populations and bacterial function in general include ocean acidification, effects related to the penetration of ultraviolet radiation into the water column, climate warming, and increased CO_2 levels [180]. A similar argument applies to larger size plankton communities: plankton abundance varies strongly with human activities, ecosystem temperature, rainfall abundance, and water level in the ecosystem [213]. As a result, a change in the population of microbites or to their functionality in an ecosystem due to external factors is far from unlikely, and the results of this study can prove what the consequences might be for the entire ecosystem.

3.5 Representation of three-node motifs

We further investigated the structural pattern of the aquatic food webs by considering three nodes food web motifs, i.e., connected subgraphs containing interactions of three nodes [140, 188].

The relative frequencies of these motifs provide insights into the network's structure. In ecological networks, motifs are considered the fundamental building blocks of communities because they represent typical species interactions [140, 104, 26]. The primary strength of using motif roles lies in their ability to incorporate both direct and indirect interactions, offering a more nuanced understanding of species' ecological roles compared to simpler measures like degree or trophic level [45]. While degree focuses on immediate interactions, motif roles capture the broader context of how species interact within the network, considering effects like intraguild predation [131]. This makes motif roles particularly useful in complex ecological dynamics and for assessing a species' niche in a more holistic manner, exploring

effects beyond two-species interactions [43]. Analyzing their distribution (i.e., the quantity of each motif type within the network) allows us to connect two traditional approaches: the study of dynamics in simple modules and the analysis of aggregated metrics that describe the community as a whole: motifs are *meso-scale* structures that have been shown to affect the population size and dynamics of the food web as well as their stability, the role of omnivory and the predator-prey body-mass ratio [58, 44, 27]. More specifically, we quantified the over- and under-representation of 13 distinct motifs when neglecting cannibalism and non-living nodes. The complete set of three-node motifs is shown in Figure 3.8.

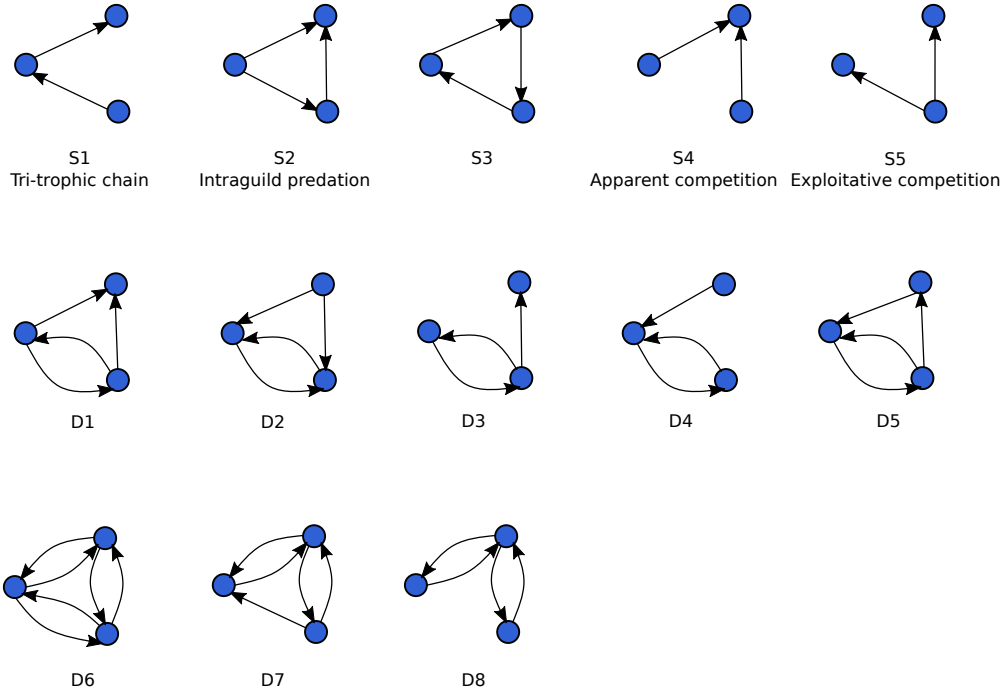


Figure 3.8: Connected three-node motifs representing different types of interactions in food webs. "S" motifs correspond to single-link interactions, while "D" motifs represent structures with double links. Only motifs named in existing literature (S_1, S_2, S_4, S_5) are assigned specific names. Arrows indicate the direction of energy or nutrient flow between nodes.

A short description of motifs follows. S_1 represent a tri-trophic chain [130] in which a top predator eats a single consumer, which in turn eats a single resource; S_2 intraguild predation [60], where the species at the top of the chain feeds on both the middle and bottom species; S_3 represent a *loop* [169] in which it is not possible to define a hierarchy between species; S_4 is the apparent competition where two resources are fed by the same predator; and S_5 is the exploitative competition where a resource is shared by two consumers [82]. Note that the term *intraguild predation* typically refers to predation between competitors of different species. However, since nodes in the food webs can represent the same species at different developmental stages (*e.g.*, juvenile, adult), we will extend the use of the S_2 motif

and the term intraguild predation to include this pattern of predation, whether it occurs within the same species or between different species.

The degree to which each motif is over- or under-represented in the food web, relative to what is expected by chance in a reference null model [8], has been linked to the dynamic properties of the community, and over-represented motifs constitute the basic building blocks of the food web [27] and are linked to the overall stability of food webs and their persistence [187, 27]. To quantify the over- and under-representation of such motifs, we employed a method defined in [188], which consists of evaluating the quantity:

$$z_i = \frac{N_{i,real} - \langle N_i \rangle}{\sigma_i} \tag{3.5}$$

where $N_{i,real}$ is the number of the i-th motif in a given network, $\langle N_i \rangle$ and σ_i are the average number of the i motif and its standard deviation measured on an ensemble of 50 random networks. These networks were generated from the original graph via the swap method described in [188], which preserves the in-degree and out-degree of each node as well as the number of single and double links. To generate each random network, we performed $100 \times L_G$ swaps where L_G indicates the links between living organisms of the graph G. Given a food web G, say that a motif i is under-represented in G if $z_i < 0$ and over-represented if $z_i > 0$.

To compare the representation of the motifs in the food webs of our dataset, we evaluated the profile of the z-score of each motif in each food web [140, 188]:

$$\mathcal{P}_i = z_i / \sqrt{(\sum_i z_i^2)}. \tag{3.6}$$

The distribution of the profile of the z-score is shown in Figure 3.9. These pattern of representation can be explained quite comprehensively through comparison to a static food-web model, the niche model [188]. We followed the analysis of Stouffer et al. [188] and investigated the relationship between the representation of the S_2 motifs (intraguild predation), and S_4 and S_5 motifs (apparent and exploitative competition, respectively). In agreement to the results of [188], aquatic food webs display a strong relationship between the representation of motif S_2, S_4 and S_5. More specifically, when in a food web the motif S_2 is over-represented then the motifs S_4 and S_5 are under-represented, whereas an under-representation of motif S_2 corresponds to an over-representation of motifs S_4 and S_5. Thus, we consider two groups of food webs: those in which S_2 is under-represented and those in which it is over-represented. We will refer to this division of networks based on intraguild predation as two distinct *families* of networks. In Figure 3.10 we report the correlation patterns between motifs representation, and in Figure 3.11 we plot the profile distribution

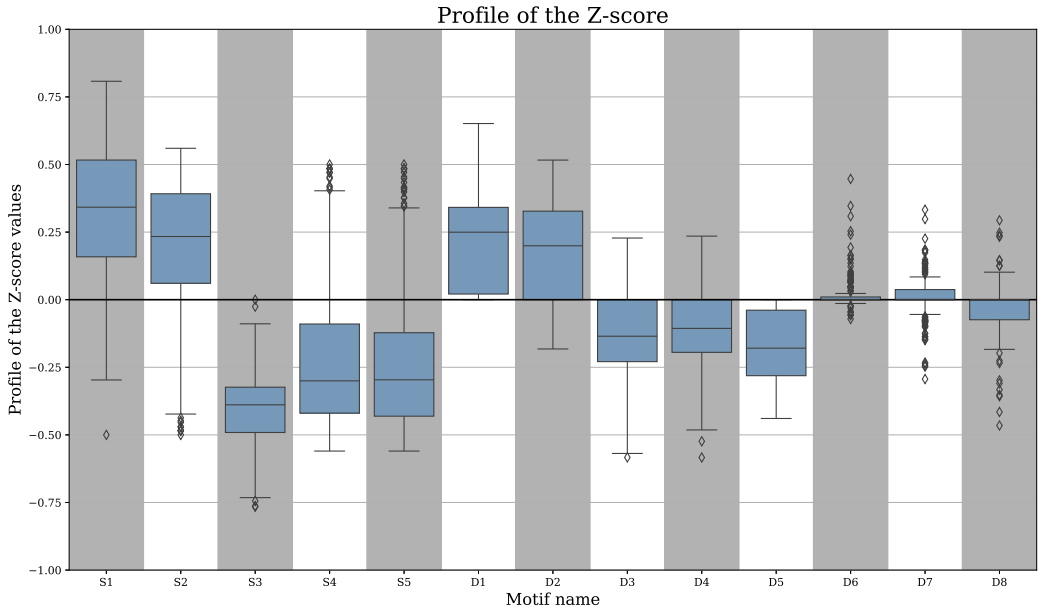


Figure 3.9: Z-score profiles for each motif across all food webs.

of the z-score based in this categorisation. In our dataset we measured that intraguild predation (S_2) is over-represented in most cases, (77% of the dataset). This is also is in agreement with [187] where it has been showed that intraguild competition motif (as well as tri-trophic food chain) is positively correlated with food-web persistence, while exploitative and apparent competition is negatively correlated with food web persistence. Polis in [161] suggests a mechanism in which a population interferes to a competitor species (that is, a species feeding on the same resource) via predation. This mechanism may be an explaination to the anti-correlation between intraguild predation and competition.

3.5.1 Intraguild predation representation accross different ecosystem types

In this section, we investigate why some food webs show an over-representation of S_2 (intraguild predation) and an under- representation of S_4 and S_5, (apparent and exploitative competition, respectively) while others display the opposite pattern.

In our dataset of 173 aquatic food webs, 134 (77.46%) exhibit an over-representation of intraguild predation, while 39 (22.54%) show an under-representation of the same motif. We evaluated the Pearson correlation coefficient between the the Z-score of the S_2 motif representation (intraguild predation), and the other metrics used in this study. These metrics

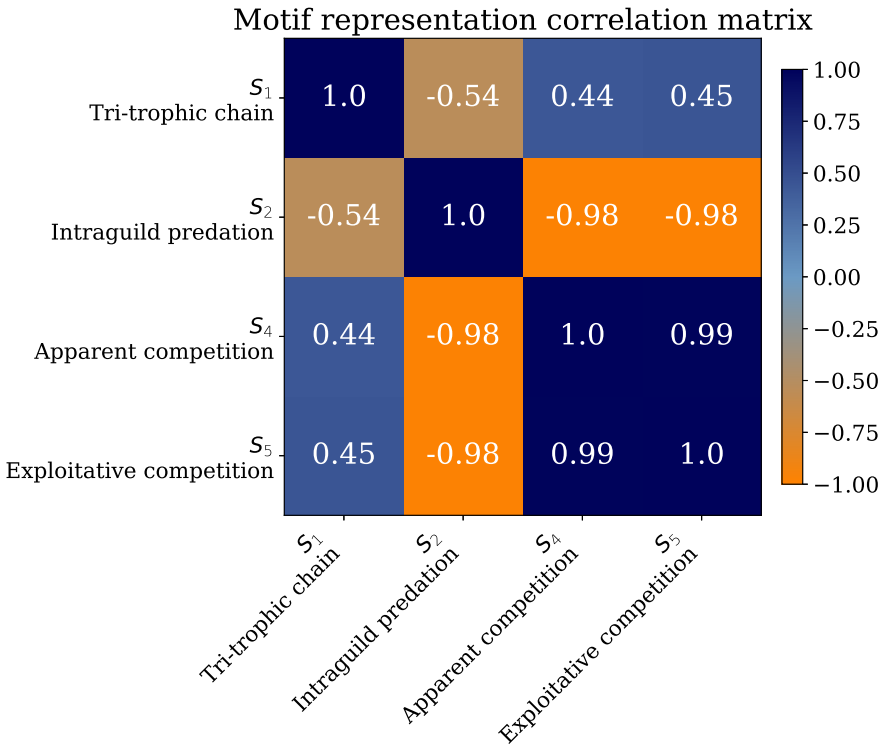


Figure 3.10: Pearson correlations of the z-scores of different motifs.

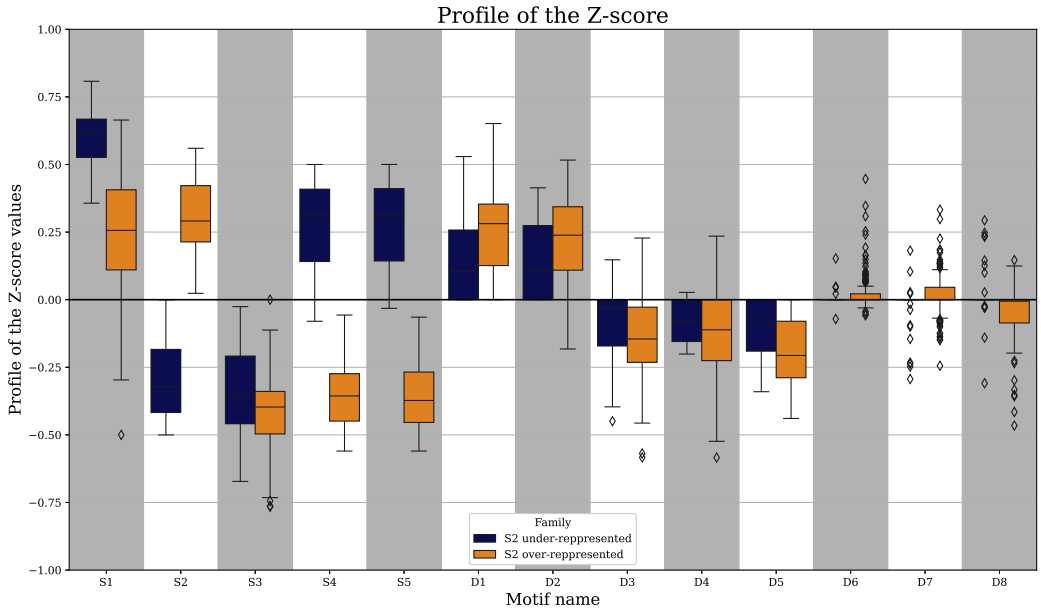


Figure 3.11: Z-score profiles for motifs in food webs with under-representation (blue boxes) and over-representation (red boxes) of intraguild predation (motif S_2).

include robustness (ρ_G), taxa richness (S), number of trophic links (L), connectance (C), the size of the giant strongly connected component, and maximum trophic level. The results are reported in Table 3.4. In all cases, the correlation coefficients (r) were below 0.3, indicating weak relationships, and no statistically significant correlations were observed ($p > 0.05$). To investigate whether this pattern varies across different aquatic ecosystems, we categorized the food webs into six ecosystem types with the method mentioned in section 3.2. Given this classification, we examined whether motif representation varied across ecosystem types. The results are reported in Table 3.5. Surprisingly, we observed that the pattern of over-representation of intraguild predation does not hold in lakes and rivers, two freshwater ecosystem types.

Given the lack of direct correlations between the food web metrics analyzed in this paper and the representation of intraguild predation, and acknowledging that ecosystems can shift from being intraguild-predation-dominated to competition-dominated depending on system productivity [60, 210], we propose that smaller ecosystems with limited resources may naturally favor smaller species that are highly specialized in their feeding habits. Klaise et al. [104] demonstrated that trophic coherence is higher in food webs with under-represented intraguild predation motifs and lower in those with over-represented intraguild predation. They suggest that a higher proportion of basal species (primary producers) leads to more

coherent food webs, characterized by a low incoherence parameter. A similar conclusion is drawn in [82], where the authors show that with increasing plant species richness, tri-trophic chains and both apparent and exploitative competition motifs became more highly represented. However, in our dataset, we observed no direct correlation between the ratio of basal species and the Z-score, or the Z-score profile, of the networks.

Another reason for which we observe this division into two families, might be partly due to how the network was compiled. The distinction between ecosystems rich in intraguild predation and those rich in competition could be an artifact of the network's low resolution [191]. For example, as shown in Figure 3 of [162], if we were to simplify the network by collapsing multiple species into single nodes—such as grouping bass into one node and combining phytoplankton, small zooplankton, and large zooplankton into a single "plankton" node—we would observe a decrease in the number of intraguild predation interactions. This suggests that some differences may be due to the level of detail captured in the network rather than actual ecological dynamics.

Metric	**Correlation Coefficient with z_{S_2}**
Robustness (ρ_G)	1.38×10^{-1}
Taxa Richness (S)	6.80×10^{-2}
Number of Trophic Links (L)	1.50×10^{-1}
Connectance (C)	1.77×10^{-1}
Giant Strongly Connected Component (GSCC)	1.02×10^{-1}
Maximum Trophic Level	3.00×10^{-1}

Table 3.4: Pearson correlation coefficients between the Z-score of the S_2 motif (intraguild predation) and various food web metrics. The GSCC represents the relative size of the giant strongly connected component—the Core.

Ecosystem Type	**S_2 Over-Represented**	**S_2 Under-Represented**	**Total**
Marine	81 (84%)	15 (16%)	96
Coast	34 (79%)	9 (21%)	43
Gulf	10 (91%)	1 (9%)	11
River	3 (27%)	8 (73%)	11
Swamp	4 (67%)	2 (33%)	6
Lake	1 (17%)	5 (83%)	6
Total	**134 (77%)**	**39 (23%)**	**173**

Table 3.5: Counts (percentages) of ecosystem types by S_2 representation.

Chapter 4

A Model for Contact Networks with Known Community Structure

Accurate contact network models are crucial for studying phenomena such as infectious disease spread [200], epidemic containment [125], and zoonotic spillover [114]. These networks capture complex interaction patterns and provide a robust framework for addressing biological challenges.

The influence of network structure on dynamic processes is well-known [103, 101, 102], yet designing models that incorporate all topological features of real-world networks remains challenging. Geometric models excel in controlling degree distribution and clustering, but lack parameters for controlling group mixing, limiting their ability to replicate community structures observed in real-world networks [79, 144].

To address this, we propose the **Hidden-degree Geometric Block Model** (**HGBM**), a generalization of the $\mathbb{S}^1$ model. HGBM introduces (i) hidden fitness normalized by community connectivity, and (ii) consistent latent features to preserve centrality with the specified community structure. Simulations demonstrate that HGBM uniquely achieves tunable community structures, degree distribution, and high clustering.

4.1 Models and methods

In subsubsection 2.3.4, we introduced the $\mathbb{S}^1/\mathbb{H}^2$ models in the framework of networks with latent geometry. One of their limitations is that they struggle to replicate the meso-scale organization characteristic of real-world networks and communities, where communities of densely connected nodes are a common feature. To date, no generalization of these models fully supports arbitrary mixing patterns within and between communities.

In that context, we introduced the concept of *soft* community structures that naturally emerge when nodes cluster together in the same region of the $\mathbb{S}^1$ and $\mathbb{H}^2$ space. Garcia et al. in [79] showed that by adopting a growing network paradigm, where link formation is driven by a competition between similarity and popularity, it is possible to generate networks with targeted topological features and soft communities. By sampling angular coordinates from non-uniform distributions, the size of each community can be effectively controlled [144]. However, this method defines communities based on their *average angular coordinate*, defined as the mean angular position of a group of nodes within a circular or spherical latent space region (see e.g. [23] Box 1).

The generalization of the $\mathbb{S}^1$ and $\mathbb{H}^2$ models to D-dimensional latent spaces extends the geometric framework by embedding nodes in higher-dimensional hyperspheres or hyperbolic spaces. In this D-dimensional setting, nodes are distributed according to specific density functions, but the probability of the existance of edge between two nodes depends on the distance in this D-dimensional space [33]. Evidence suggests that using higher-dimensional latent spaces could mitigate these limitations [59] : higher-dimensional models introduce a significant qualitative improvement over their lower-dimensional counterparts by reducing the restrictive influence of similarity on connection probabilities. In higher dimensions, the number of nearest neighbors for angular clusters representing communities increases, which allows for more flexible and realistic community structures. However, increasing dimensionality comes at the cost of reducing the maximum achievable clustering coefficients in geometric models [78].

Thus, developing a purely geometric generalization of the $\mathbb{S}^1$ and $\mathbb{H}^2$ models that supports flexible community structures while maintaining key features—such as high clustering coefficients and heterogeneous degree distributions—remains an open and significant challenge.

In this section, we define the geometric spaces and constraints necessary to generate the $\mathbb{S}^1$ model [22] and extend it to incorporate community structure. It will be useful to first outline the formulation of the $\mathbb{S}^1$ model using the maximum entropy principle, followed by a generalization of this framework to embed group mixing within the latent geometry.

4.1.1 The $\mathbb{S}^1$ model

In the $\mathbb{S}^1$ model, nodes are embedded in a one-dimensional sphere—a that is, each node is equiped with an angular position on the circle with radius R [107, 98] . Within this space, N nodes are uniformly distributed with a fixed density, normalized to one for simplicity, ensuring that ($N = 2\pi R$). Each node is assigned a hidden degree κ, a continuous real-valued parameter which is proportional to its expected degree, allowing the model to capture

heterogeneous degree distributions.

The angular position θ of a node on the circle can be correlated with κ, and both variables can follow an arbitrary joint distribution $\rho(\kappa, \theta)$. By appropriately selecting $\rho(\kappa, \theta)$, the $\mathbb{S}^1$ model is capable of reproducing a wide range of degree–degree correlation patterns and clustering spectra, making it a versatile framework for modeling complex networks [106]. Once all nodes are assigned a tuple (κ, θ), each pair of nodes is connected with probability:

$$p_{ij} = \frac{1}{1 + (\frac{x_{ij}}{\kappa_i \kappa_j \mu})^\beta} \tag{4.1}$$

Where x_{ij} is the angular distance defined as $x_{ij} = R\Delta(\theta_i, \theta_j)$. We derived this expression through the principle of maximum entropy in in the subsection below where we clarify all the details and definitions.

The probability of an edge existing between nodes i and j can be understood as a balance between two primary factors: the distance x_{ij} - where smaller distances increase the likelihood of an edge - and the hidden variables of the nodes κ_i and κ_j - where a higher product of these fitness values increases the edge probability. The parameters β and μ control the clustering and the average degree, respectively [22]. This model's flexibility in tuning the clustering coefficient via β makes it particularly valuable for generating graphs with a desired degree distribution $\rho(\kappa)$, zero-degree assortativity, and tunable clustering.

Derivation of the Probability Distribution from Max Entropy Principles

In this section, we provide the detailed derivation that leads from the maximum entropy principle under the given constraints to the probability distribution in Equation 4.1. We start from the constraints on expected energy and expected degrees and apply the method of Lagrange multipliers to obtain the maximum-entropy distribution with the same methodology in subsubsection 2.3.4. In that case we applied the principle of maximum entropy to obtain a definition of the probability of extraction of a graph G from the constraints on the number of edges. In this case we can use similar constraints, adding information about the distribution of nodes in the latent space $\mathbb{S}^1$. This will allow us to obtain an equation for the probability on the edge between node i and j rather than the probability on the graph G.

We consider a probability distribution $P(G)$ over all possible graphs G on n vertices. The graphs are determined by adjacency matrices (a_{ij}), where $a_{ij} = 1$ if there is an edge between i and j, and $a_{ij} = 0$ otherwise. We assume undirected graphs with no self-loops, so $a_{ij} = a_{ji}$ and $a_{ii} = 0$.

We define the following constraints. The first is on the normalization of the distribution

$P(G)$:

$$\sum_G P(G) = 1. \tag{4.2}$$

Let $E(G) = \sum_{i<j} \varepsilon_{ij} a_{ij}(G)$ denote the total energy of graph G, where $\varepsilon_{ij} = \varepsilon(x_{ij})$ is a given function of the distance x_{ij} on the ring. We impose:

$$\sum_G P(G)E(G) = E^*. \tag{4.3}$$

For each vertex i, let $\deg_i(G) = \sum_j a_{ij}(G)$ denote the degree of node i in G. We impose:

$$\sum_G P(G) \deg_i(G) = \kappa_i^*, \quad \forall i. \tag{4.4}$$

Our goal is to find the $P(G)$ that maximizes the Shannon entropy

$$S(P) = -\sum_G P(G) \ln P(G) \tag{4.5}$$

subject to the constraints in Eqs. (4.2)-(4.4).

Maximizing Entropy via Lagrange Multipliers We introduce a Lagrange multiplier α for the normalization constraint, a Lagrange multiplier β for the energy constraint, and a set of Lagrange multipliers $\{\lambda_i\}$ for the degree constraints. The constrained optimization problem is then formulated through the Lagrangian:

$$\begin{aligned}\mathcal{L}(P, \alpha, \{\lambda_i\}, \beta) = &-\sum_G P(G) \ln P(G) + \alpha\left(1 - \sum_G P(G)\right) \\ &+ \sum_i \lambda_i \left(\kappa_i^* - \sum_G P(G) \deg_i(G)\right) + \beta\left(E^* - \sum_G P(G)E(G)\right).\end{aligned} \tag{4.6}$$

To find the maximum, we take the functional derivative of $\mathcal{L}$ with respect to $P(G)$ and set it to zero:

$$\frac{\partial \mathcal{L}}{\partial P(G)} = -\ln P(G) - 1 - \sum_i \lambda_i \deg_i(G) - \beta E(G) - \alpha = 0. \tag{4.7}$$

Solve for $P(G)$:

$$-\ln P(G) = 1 + \alpha + \sum_i \lambda_i \deg_i(G) + \beta E(G). \tag{4.8}$$

This gives:

$$P(G) = e^{-1-\alpha} e^{-\sum_i \lambda_i \deg_i(G) - \beta E(G)}. \tag{4.9}$$

Define $Z = e^{1+\alpha}$ as a normalization constant (often called the partition function in analogy with statistical mechanics). Then:

$$P(G) = \frac{e^{-\sum_i \lambda_i \deg_i(G) - \beta E(G)}}{Z}. \tag{4.10}$$

Expressing the constraints in terms of edges Recall that:

$$\deg_i(G) = \sum_j a_{ij}(G) \quad \text{and} \quad E(G) = \sum_{i<j} \varepsilon_{ij} a_{ij}(G).$$

Substitute these into Eq. (4.10):

$$P(G) = \frac{e^{-\sum_i \lambda_i \sum_j a_{ij}(G) - \beta \sum_{i<j} \varepsilon_{ij} a_{ij}(G)}}{Z}. \tag{4.11}$$

We can rearrange the double sums. Notice that:

$$\sum_i \sum_j a_{ij}(G)\lambda_i = \sum_{i<j} a_{ij}(G)(\lambda_i + \lambda_j)$$

because each edge (i, j) contributes to both $\deg_i(G)$ and $\deg_j(G)$, and for undirected graphs we can symmetrize the sums.

Hence:

$$P(G) = \frac{e^{-\sum_{i<j} a_{ij}(G)(\lambda_i + \lambda_j + \beta\varepsilon_{ij})}}{Z}. \tag{4.12}$$

Factorization into independent edges Since the graph is defined by independent binary variables $a_{ij} \in \{0, 1\}$ for $i < j$, we can write $P(G)$ as a product over edges:

$$P(G) = \frac{1}{Z} \prod_{i<j} \exp\big[-a_{ij}(G)(\lambda_i + \lambda_j + \beta\varepsilon_{ij})\big]. \tag{4.13}$$

This factorization suggests that each edge is formed independently with some probability that depends on i, j, ε_{ij}, and the Lagrange multipliers.

For a single pair (i, j), the marginal probability that $a_{ij} = 1$ is given by:

$$p_{ij} = P(a_{ij} = 1) = \sum_{G: a_{ij}=1} P(G).$$

Since all edges factorize and each is a binary variable with only two possible states $a_{ij} = 0$ or 1, we can write:

$$p_{ij} = \frac{e^{-(\lambda_i+\lambda_j+\beta\varepsilon_{ij})}}{e^0 + e^{-(\lambda_i+\lambda_j+\beta\varepsilon_{ij})}}. \tag{4.14}$$

This simplifies to:

$$p_{ij} = \frac{1}{1 + e^{\lambda_i+\lambda_j+\beta\varepsilon_{ij}}}, \tag{4.15}$$

Thus, starting from the maximum entropy principle and imposing the constraints on expected energy and expected degrees, we have derived the logistic form of the link probability.

Suppressing degree correlations In this context, to suppress any nonstructural degree correlations, one must take $\varepsilon_{ij} = \ln(x_{ij})$ [22], so that p_{ij} can be rewritten as

$$p_{ij} = \frac{1}{1 + \left(x_{ij} e^{\frac{\lambda_i+\lambda_j}{\beta}}\right)^\beta} \tag{4.16}$$

Linking the degree of the node to the Lagragian multiplier We are now interested in finding the relationship between λ_i and the expected degree of the node deg_i in the graph obtained via the probability matrix p_{ij}. In this way, we can adjust the set $\{\lambda_i\}$ to obtain a desired degree distribution.

If the nodes are placed uniformly at random on the $\mathbb{S}^1$ ring, then $\rho(\theta) = \frac{1}{2\pi}$ and the density of nodes on $\mathbb{S}^1$ is $\delta = \frac{N}{2\pi R}$. Let the relative distance $x_{ij} = \Delta(\theta_i, \theta_j) = (\pi - |\pi - |\theta_i - \theta_j||)$ be the absolute angular distance between the latent coordinates θ_i and θ_j. Without loss of generality, one can take $\theta_i = 2\pi$ so that $\Delta(\theta_i, \theta_j) = \theta_j$. By integrating over θ_j and λ_j, one obtains

$$\langle p_{ij} \mid \lambda_i \rangle = \int \rho(\lambda) d\lambda \int_0^\pi \frac{1}{\pi} \frac{d\theta}{1 + \left(R\theta e^{\frac{\lambda_i+\lambda}{\beta}}\right)^\beta} = \frac{1}{\pi R} e^{-\frac{\lambda_i}{\beta}} \langle e^{-\frac{\lambda}{\beta}} \rangle \mathcal{I}_\beta \tag{4.17}$$

where $\mathcal{I}_\beta = \frac{\pi}{\beta \sin(\pi/\beta)}$.[1]

Since $\deg_i = \sum_j a_{ij}$, and therefore $\langle deg_i \rangle = \sum_j p_{ij}$, one finds

$$\kappa_i = \langle \deg_i \mid \lambda_i \rangle = N \langle p_{ij} \mid \lambda_i \rangle = 2\delta e^{-\frac{\lambda_i}{\beta}} \langle e^{-\frac{\lambda}{\beta}} \rangle \mathcal{I}_\beta \tag{4.18}$$

[1]The r.h.s. of Equation 4.17 follows from $\int_0^X \frac{dt}{1+t^\beta} \overset{X\to\infty}{\approx} \frac{\pi}{\beta \sin(\pi/\beta)} \cdot \frac{1}{X} = \frac{\mathcal{I}_\beta}{X}$ and the fact that $N \to \infty$ means $R \to \infty$ for fixed δ. In Equation 4.17, $\rho(\lambda)$ is the density over λ induced by $\rho(\kappa)$.

and, by further integrating over λ_i, one gets $\langle e^{-\frac{\lambda}{\beta}} \rangle = \sqrt{\langle \kappa \rangle (2\delta \mathcal{I}_\beta)^{-1}}$, so that

$$e^{-\frac{\lambda_i}{\beta}} = \frac{\kappa_i}{\sqrt{\langle \kappa \rangle}} \frac{1}{\sqrt{2\delta \mathcal{I}_\beta}} \tag{4.19}$$

Plugging (4.19) into (4.16), one finally finds

$$p_{ij} = \frac{1}{1 + \left(\frac{x_{ij}}{\frac{\kappa_i \kappa_j}{\langle \kappa \rangle} \frac{1}{2\delta \mathcal{I}_\beta}} \right)^\beta} = \frac{1}{1 + \left(\frac{x_{ij}}{\kappa_i \kappa_j \mu} \right)^\beta} \tag{4.20}$$

where $\mu = \frac{1}{2\delta \mathcal{I}_\beta \langle \kappa \rangle}$. Equation 4.20 matches the final expression in Equation 4.1.

In the following, it will be useful to rewrite Equation 4.1 in a slightly different way. Let $M = \frac{1}{2} \sum_i \kappa_i = \frac{N}{2} \langle \kappa \rangle$ be the expected number of edges in the graph, and let $f_i = \frac{\kappa_i}{2M}$ be the normalized hidden degree of v_i, a measure of the *relative popularity* of v_i with respect to the rest of the network. Note that f_i can be interpreted as the probability that a random edge is incident on v_i, and that $\sum_i f_i = 1$. Since $x_{ij} = R\Delta(\theta_i, \theta_j)$ and $\delta = \frac{N}{2\pi R}$, Equation 4.1 can be rewritten as

$$p_{ij} = \frac{1}{1 + \left(\frac{\Delta(\theta_i, \theta_j) \mathcal{I}_\beta}{2\pi M f_i f_j} \right)^\beta} \tag{4.21}$$

Equation 4.21 clarifies that, for fixed M, p_{ij} decreases with the angular distance between v_i and v_j, and increases with the popularity of the nodes.

It is worth noting that the $\mathbb{S}^1$ model can be generalized from one to D dimensions by assigning each node a position on the D-sphere, $\mathbf{v}_i$, such that $\sum_{d=1}^{D} v_{i,d}^2 = R^2$. In this case, the angular distance between two nodes i and j is defined as:

$$x_{ij} = R\,\Delta(\theta_i, \theta_j) = R \arccos\left(\frac{\langle \mathbf{v}_i, \mathbf{v}_j \rangle}{R^2} \right),$$

where $\langle \cdot, \cdot \rangle$ denotes the scalar product. This generalization extends the latent space embedding to higher dimensions while preserving the geometric principles of the original model.

4.1.2 The Hidden-degree Geometric Block Model (HGBM)

After deriving $\mathbb{S}^1$ and identifying its limitations in reconstructing community partitions from original data, we introduce a geometric model with an *imposed* community structure.

The central idea is to divide the nodes into q communities, denoted as $\{b_1, \ldots, b_q\}$ and enforce constraints not only on the degree of each node but also on the distribution of edges

both within and between these communities. To streamline the notation, we adopt the convention of using lowercase Latin letters to denote individual nodes and their corresponding uppercase letters to represent the communities to which they belong. For instance, node i is associated with community I, node j with community J and so on.

The constraints are established as follows. First, we define a $q \times q$ matrix K, where each element K_{IJ} represents the expected number of edges between communities I and J for $I \neq J$, and twice that value for intra-community edges ($I = J$) (see, for example, [21]) .

In the constraints definition we will impose that the matrix K is respected. Second, we redefine the constraints on the degree of each node, modifying the formulation originally defined for $\mathbb{S}^1$:

$$\langle deg_i \rangle = \kappa_i = f_i \sum_J K_{IJ} \tag{4.22}$$

In this definition, the hidden degree κ_i is expressed as the product of a *hidden fitness* f_i and the total number of edges leaving the community I to which node i belongs, $\sum_J K_{IJ}$. To ensure that the edge distribution prescribed by the matrix K_{IJ} is respected, it is necessary to normalize the hidden fitness values such that $\sum_{i \in I} f_i = 1$. This implies:

$$f_i = \frac{\kappa_i}{\sum_{i \in I} \kappa_i} \tag{4.23}$$

Defining constraints in this way achieves two key objectives. First, it replaces the concept of hidden degree with hidden fitness, which is normalized based on the inter-community edge distribution encoded in K. Second, the attributes of each node, such as hidden fitness and hidden angular position, can be effectively precomputed to preserve patterns of centrality while ensuring consistency with the specified community structure.

We present the result for the edge probability between nodes i and j, with the detailed calculation provided in subsubsection 4.1.2.

$$p_{ij} = \frac{1}{1 + \left(\frac{\Delta(\theta_i, \theta_j) \mathcal{I}_\beta}{\pi K_{IJ} f_i f_j} \right)^\beta} \tag{4.24}$$

Equation 4.24 has the same functional form of Equation 4.1. Here, β serves the same purpose as previously described, acting as a tuning parameter for adjusting the clustering coefficient.

Derivation of HGBM from Max Entropy Principles

In this section, we derive the probability distribution for the Hyperbolic Block Model (HGBM) from the maximum entropy principle, subject to the given constraints on the expected energy, the expected number of inter-block edges, and the expected degrees. We follow a procedure similar to that used for the standard $\mathbb{S}^1$ model in subsubsection 4.1.1, ensuring that the final probability factorizes into independent probabilities per link.

Setup and Constraints Consider a vertex set V of size N partitioned into q disjoint blocks (or communities) $B_1, \ldots, B_q$. We label blocks by uppercase indices $I, J, \ldots$ and vertices by lowercase indices $i, j, \ldots$, where $i \in I$ means vertex v_i belongs to block B_I. Let $N_I = |B_I|$ be the size of block B_I, and let $K = \{K_{IJ}\}$ be a symmetric matrix encoding the expected number of edges between blocks B_I and B_J.

Each vertex v_i is assigned a latent angular coordinate θ_i on the $\mathbb{S}^1$ ring and a fitness f_i such that $\sum_{i \in I} f_i = 1$. Similar to the standard $\mathbb{S}^1$ model, we impose three sets of constraints:

$$\langle E \rangle_P = E^*, \tag{4.25}$$

$$\langle L_{IJ} \rangle_P = K_{IJ} \quad \forall I, J, \tag{4.26}$$

$$\langle \deg_i \rangle_P = f_i \sum_J K_{I_i J} \quad \forall i. \tag{4.27}$$

Here, $E(G) = \sum_{i<j} \varepsilon_{ij} a_{ij}(G)$ is the total energy of graph G, with $\varepsilon_{ij} = \varepsilon(x_{ij})$ a function of the geometric distance $x_{ij} = R\Delta(\theta_i, \theta_j)$ on the ring of radius R. The quantities $L_{IJ} = \sum_{i \in I, j \in J} a_{ij}(G)$ count the number of edges between blocks I and J, and $\deg_i(G) = \sum_j a_{ij}(G)$ is the degree of vertex i.

Maximum Entropy Distribution and Lagrange Multipliers As before, we seek a probability distribution $P(G)$ that maximizes the Shannon entropy

$$S(P) = -\sum_G P(G) \ln P(G)$$

subject to the constraints in Eqs. (4.25)-(4.27) and the normalization $\sum_G P(G) = 1$.

Introducing a Lagrange multiplier α for normalization, a multiplier β for the energy constraint (4.25), a set of multipliers $\{\eta_{IJ}\}$ for the block-edge constraints (4.26), and a set $\{\lambda_i\}$ for the degree constraints (4.27), the Lagrangian is:

$$\begin{aligned}\mathcal{L}(P,\alpha,\{\lambda_i\},\{\eta_{IJ}\},\beta) = &-\sum_G P(G)\ln P(G) + \alpha\Big(1-\sum_G P(G)\Big)+\\ &+\sum_i \lambda_i\big(f_i\sum_J K_{I_iJ} - \sum_G P(G)\deg_i(G)\big) + \beta(E^* - \sum_G P(G)E(G)).\end{aligned} \tag{4.28}$$

Taking the derivative with respect to $P(G)$ and setting it to zero yields:

$$-\ln P(G) - 1 - \sum_{I,J}\eta_{IJ}L_{IJ}(G) - \sum_i \lambda_i \deg_i(G) - \beta E(G) - \alpha = 0.$$

Solving for $P(G)$, we obtain an exponential form. Since $L_{IJ}(G)$ and $\deg_i(G)$ both depend linearly on $a_{ij}(G)$, and $E(G)$ is also a linear function in $a_{ij}(G)$, the resulting maximum-entropy distribution factorizes over edges. This leads to the link probability:

$$p_{ij} = \frac{1}{1+\exp(\lambda_i+\lambda_j+\eta_{I_iJ_j}+\beta\varepsilon_{ij})}.$$

Following the same reasoning as in the standard $\mathbb{S}^1$ model, one can show that choosing $\varepsilon_{ij} = \ln(x_{ij})$ is necessary to suppress non-structural degree correlations, yielding:

$$p_{ij} = \frac{1}{1+\left(x_{ij}e^{\frac{\lambda_i+\lambda_j+\eta_{I_iJ_j}}{\beta}}\right)^\beta}.$$

Determining the Lagrange Multipliers We assume that θ_i are uniformly distributed on $[0,2\pi)$ so that $\rho(\theta) = \frac{1}{2\pi}$ and $\delta = \frac{N}{2\pi R}$ is the node density on the ring. For large N, each block B_I has N_I nodes with density $\delta_I = \frac{N_I}{2\pi R}$, and $\frac{N_I}{N}$ remains constant as $N\to\infty$.

Consider $v_i \in B_I$ and $v_j \in B_J$. For fixed λ_i and η_{IJ}, we integrate over θ_j (taking $\theta_i = 2\pi$ so that $\Delta(\theta_i,\theta_j) = \theta_j$) and over the distribution of λ_j to find:

$$\langle p_{ij} \mid \lambda_i, \eta_{IJ}, i\in I, j\in J\rangle = \frac{\mathcal{I}_\beta}{\pi R}e^{-\frac{\eta_{IJ}}{\beta}}e^{-\frac{\lambda_i}{\beta}}\langle e^{-\frac{\lambda}{\beta}}\rangle,$$

where $\mathcal{I}_\beta$ is a known integral factor arising from the asymptotic integration over angles.

Since $L_{IJ} = \sum_{i\in I, j\in J} a_{ij}$, taking expectations and using the independence of edges:

$$K_{IJ} = \langle L_{IJ} \mid \eta_{IJ}\rangle = N_I N_J \frac{\mathcal{I}_\beta}{\pi R}e^{-\frac{\eta_{IJ}}{\beta}}\langle e^{-\frac{\lambda}{\beta}}\rangle^2.$$

From this, we solve for $e^{-\frac{\eta_{IJ}}{\beta}}$:

$$e^{-\frac{\eta_{IJ}}{\beta}} = \frac{\pi R}{\mathcal{I}_\beta} \frac{K_{IJ}}{N_I N_J} \frac{1}{\langle e^{-\frac{\lambda}{\beta}} \rangle^2}.$$

Next, consider the expected degree of a node $v_i \in B_I$:

$$f_i \sum_J K_{I_i J} = \langle \deg_i \mid \lambda_i, K_{I_i J}, i \in I \rangle = \sum_J N_J \langle p_{ij} \mid \lambda_i, K_{IJ}, i \in I, j \in J \rangle.$$

Substituting the expression for $e^{-\frac{\eta_{IJ}}{\beta}}$ back into the integrated probability leads to:

$$f_i \sum_J K_{I_i J} = \frac{e^{-\frac{\lambda_i}{\beta}}}{\langle e^{-\frac{\lambda}{\beta}} \rangle} \frac{1}{N_I} \sum_J K_{IJ}.$$

Isolating $e^{-\frac{\lambda_i}{\beta}}$:

$$e^{-\frac{\lambda_i}{\beta}} = f_i N_I \langle e^{-\frac{\lambda}{\beta}} \rangle.$$

Final Probability Distribution Finally, substituting $e^{-\frac{\eta_{IJ}}{\beta}}$ and $e^{-\frac{\lambda_i}{\beta}}$ back into the expression for p_{ij}, and recalling $x_{ij} = R\Delta(\theta_i, \theta_j)$ and $\delta = \frac{N}{2\pi R}$, we obtain:

$$p_{ij} = \frac{1}{1 + \left(\frac{\Delta(\theta_i, \theta_j) \mathcal{I}_\beta}{\pi K_{IJ} f_i f_j} \right)^\beta}.$$

This final form mirrors the functional structure of the standard $\mathbb{S}^1$ model's probability distribution. The difference is that the fitness scores f_i are normalized per block, and the total number of edge-stubs $2M$ in the simpler model is replaced here by K_{IJ}, reflecting the block structure.

Adjusting the hidden degrees in a finite-size network

We demonstrate an important property of these models (both $\mathbb{S}^1$ and HGBM): in the thermodynamic limit, where $N \to \infty$, the expected degree of a node matches its hidden degree exactly, defined as $\kappa_i = f_i \sum_J K_{IJ}$. However, this relationship breaks down in finite-size networks, introducing discrepancies between the target degree distribution and the realized degrees in the generated network.

Accurately reconstructing a network with a desired degree distribution or degree sequence—averaged over an ensemble of graphs—requires compensating for this finite-size effect. To address this, the hidden degrees must be adjusted to account for the deviation be-

tween the expected and target degrees. The solution usually adopted (see, for example, [77] section A.2) is to adjust the hidden degree iteratively.

To begin, we calculate the expected degree of a node i with hidden degree κ_i, denoted as $\langle deg_i \rangle(\kappa_i)$. This requires evaluating the probability of the existence of an edge between two nodes with hidden degrees κ_i and κ_j in the communities I and J:

$$p_{ij} = \int_0^{\pi} \frac{1}{\pi} \frac{1}{1 + \left(\frac{\mathcal{I}_\beta \sum_I K_{IJ} \sum_J K_{IJ}}{\pi K_{IJ} \kappa_i k_j} \theta_j\right)^\beta} d\theta_j \tag{4.29}$$

In this integral, we set the angular position of the node with hidden degree κ_i to 2π without loss of generality and use the relationship $f_i = \kappa_i / \sum_J K_{IJ}$.

The exact solution of this integral is:

$$p_{ij} = \int_0^{\pi} \frac{1}{\pi} \frac{1}{1 + \left(\frac{\mathcal{I}_\beta \sum_I K_{IJ} \sum_J K_{IJ}}{\pi K_{IJ} \kappa_i \kappa_j} \theta_j\right)^\beta} d\theta_j = {}_2F_1\left(1, \frac{1}{\beta}, 1 + \frac{1}{\beta}, -\left(\frac{\mathcal{I}_\beta \sum_I K_{IJ} \sum_J K_{IJ}}{\pi K_{IJ} \kappa_i \kappa_j}\right)^\beta\right) \tag{4.30}$$

where ${}_2F_1$ is the hypergeometric function.

Given a set of hidden degrees, the expected degree for node i can be written as:

$$\langle deg_i \rangle(\kappa_i) = \sum_j {}_2F_1\left(1, \frac{1}{\beta}, 1 + \frac{1}{\beta}, -\left(\frac{\mathcal{I}_\beta \sum_I K_{IJ} \sum_J K_{IJ}}{\pi K_{IJ} \kappa_i \kappa_j}\right)^\beta\right) \tag{4.31}$$

To ensure the hidden degrees align with the desired degree distribution or sequence, we iteratively adjust κ_i as follows:

$$\kappa_i = \kappa_i + (deg_i - \langle deg_i \rangle(\kappa_i)) \cdot \mathcal{U}(0, 1) \tag{4.32}$$

Here, $\mathcal{U}(0, 1)$ is a random number uniformly distributed between 0 and 1. This scaling factor prevents the adjustments from becoming trapped in local minima while progressively aligning the hidden degrees with those values that allow the target degrees reconstruction. To further optimize this procedure, nodes can be grouped into classes according to degree. In this way, the evaluation in Equation 4.31 can be performed only once for each degree class. To make this procedure effective, the integer part of the hidden degree κ_i must be taken at each iteration.

The complete algorithm is summarized in Algorithm 1.

4.1.3 Implementing HGBM

Now that we have defined the model, we describe here a computational procedure to sample networks based on the probability between nodes (i, j) as defined in Equation 4.31.

First, we introduce the *mixing matrix*, a $q \times q$ matrix, Δ, such that $K_{IJ} = M\Delta_{IJ}$, where M is the expected total number of edges. The mixing matrix specifies the proportion of edges between communities I and J and is normalized such that $\sum_{IJ} \Delta_{IJ} = 2$. This concept allows the user to specify the proportions of edges between communities rather than their absolute values, which can be easily derived from a graph once the node partition is known.

Algorithm 2 generates a random network with a given community partition, mixing matrix Δ, average degree $\langle deg \rangle$, degree distribution exponent α, minimum degree κ_0, and inverted temperature β.

Regarding the degree of a node i, two approaches are available:

- **Global degree distribution:** The user specifies the average degree $\langle deg \rangle$, the power-law exponent α, and the minimum degree κ_0. This generates an ensemble of networks with a global degree distribution following the specified parameters. However, each node i will not have a fixed degree across realizations, as a new hidden fitness is extracted for each generation.

- **Node-specific degree sequence:** The user provides a specific degree sequence for the nodes. In this case, the generated ensemble ensures that the expected degree of each node matches the input sequence.

In both situations, it is important to adjust the degrees using the Algorithm 1.

The detailed steps for generating such an ensemble of networks are outlined in Algorithm 2.

Algorithm 1 Iterative algorithm to adjust the hidden degrees κ_i

Input: Target degrees $\{\deg_i\}$, tolerance ϵ, maximum iterations T
Output: Adjusted hidden degrees $\{\kappa_i\}$
Initialize hidden degrees: $\kappa_i \leftarrow \deg_i$
Set iteration counter $t \leftarrow 0$
repeat
    Collect the hidden degree classes $\{\kappa\}$
    for each class κ **do**
        Compute the expected degree for class κ, $\langle \deg \rangle(\kappa)$, using Equation 4.31
    end for
    for each node i **do**
        Assign each node i its corresponding expected degree:

$$\langle \deg_i \rangle = \langle \deg \rangle(\kappa_i), \quad \text{where } \kappa_i = \kappa$$

        Compute error: $\Delta_i \leftarrow \deg_i - \langle \deg_i \rangle(\kappa_i)$
        Update hidden degrees:

$$\kappa_i \leftarrow \kappa_i + \Delta_i \cdot \mathcal{U}(0, 1)$$

    end for
    Increment iteration counter: $t \leftarrow t + 1$
until Maximal error $\max(|\Delta_i|) < \epsilon$ or $t \geq T$
Return: Adjusted hidden degrees $\{k_i\}$

Algorithm 2 Generate a HGBM ensemble of random networks

Read input parameters
if κ_i are specified **then**
    Measure K_{IJ} based on the community partition and the degrees κ_i
    $\Delta_{IJ} \leftarrow K_{IJ}$
end if
Normalize Δ: $\Delta_{IJ} \leftarrow \frac{2\Delta_{IJ}}{\sum_{IJ}\Delta_{IJ}}$
for each node i **do**
    Assign community I according to community partition
    Assign angle $\theta_i \sim \mathcal{U}(0, 2\pi)$
    if κ_i are not specified **then**
        if κ_0 is not specified **then**
            $\kappa_0 \leftarrow \frac{\alpha-2}{\alpha-1}\langle \deg \rangle$
        end if
        Assign hidden degree $\kappa_i \sim \mathcal{PW}(\alpha, \kappa_0)$
    end if
    Set the target degrees $\deg_i \leftarrow \kappa_i$
    Adjust the hidden degree κ_i as showed in Algorithm 1
    Normalize fitness $f_i \leftarrow \kappa_i / \sum_{i \in I} \kappa_i$
end for
for each pair of nodes $i \neq j$ **do**
    Evaluate edge probability p_{ij} using Equation 4.24
end for
return θ, κ, and p_{ij}

4.2 Results

In this section we report our campaign of simulations to validate our model. Our validation strategy is divided into two main parts: first, we simulate an ensemble of synthetic networks with a predefined community structure; second, we use a real contact network to validate the results [186]. To evaluate these methods, we employ D-Mercator [98], a model-based embedding method that produces multidimensional maps of real networks into the $(D+1)$-hyperbolic space, where the similarity subspace is represented as a D-sphere. Using this approach, we demonstrate that while D-Mercator successfully replicate key topological features—such as degree distribution and clustering—they do not effectively capture the

community structure.

4.2.1 HGBM is capable to reconstruct the desired mixing matrix

This section demonstrates that our model reproduces a graph with a power-law degree distribution and a desired group mixing while maximizing the clustering coefficient. We generate an ensemble of graphs using the HGBM with a combination of parameters, including the community structure and inverse temperature.

The results are reported in Figure 4.1 and Table 4.1.

Options	Assortativity	Giant Component	Clustering Coefficient
$\beta = 5.0$ K ratio: 6 / 1	-0.0863 ± 0.0350	0.9891 ± 0.0026	0.6518 ± 0.0510
$\beta = 10.0$ K ratio: 6 / 1	-0.0791 ± 0.0260	0.9863 ± 0.0027	0.6760 ± 0.0381
$\beta = 5.0$ K ratio: 3 / 1	-0.0671 ± 0.0267	0.9877 ± 0.0033	0.6840 ± 0.0245
$\beta = 10.0$ K ratio: 3 / 1	-0.0847 ± 0.0279	0.9855 ± 0.0037	0.7257 ± 0.0282

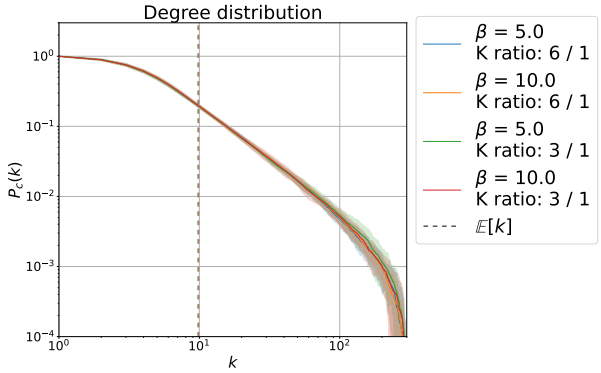
Table 4.1: Global metrics of the network generated with different parameter settings.

For these examples, we generated graphs with 10,000 nodes partitioned into four communities, each containing 2,500 nodes. The average degree was set to $\langle deg \rangle = 10$ across all tests, while the inverse temperature (β) and the mixing matrix were varied. For each parameter set, the measures were averaged over 10 realizations.

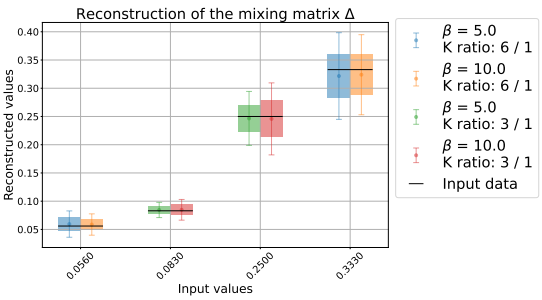
We explored a total of four scenarios. In the first two (K ratio = 3 / 1), the diagonal of the mixing matrix was three times the off-diagonal, with β set to 5 and 10, respectively. In the remaining two scenarios (K ratio = 6 / 1), the diagonal was six times the off-diagonal, again with β set to 5 and 10. This diversity reflects varying levels of community assortativity, which is defined as the tendency of nodes within the same community to preferentially form connections with one another rather than with nodes from different communities. This phenomenon is commonly observed in real-world networks [215].

In Figure 4.1a we report the measurements of the degree distributions for all the 4 scenarios. They respect the values imposed as input: the average is about 10.0 for each combination of parameters. In the 4 experiments, the exponent of the degree distribution was measured with the Python `powerlaw` library and has values 2.54, 2.57, 2.49, 2.58, against an expected value of 2.50. These values are in a range within 3.2% of the expected value.

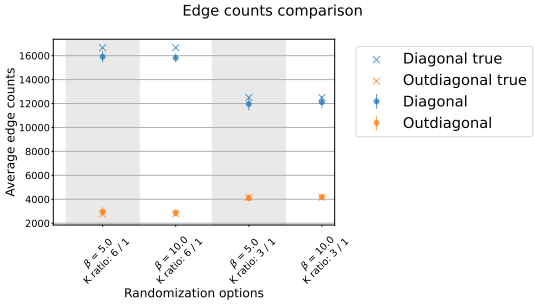
Communities are also faithfully reproduced. Figure 4.1b and Figure 4.1c show the performance of the mixing matrix reconstruction under different experimental conditions. In Figure 4.1b, reconstructed mixing matrix values are compared to real values. The points highlight the average over the ensamble, while the boxes and the caps highlight one and two



(a) Complementary cumulative distribution function (CCDF) of degree distribution. Solid lines show the average degree, with shaded regions representing the standard deviation. The dashed vertical line indicates the expected average degree, $\langle k \rangle$.



(b) Reconstructed vs. real mixing matrix values across various conditions. Boxes and caps represent one and two standard deviations, respectively.



(c) Average diagonal and off-diagonal elements of the mixing matrix K. 'x' markers indicate real values, while dots and error bars show the mean and standard deviation.

Figure 4.1: Topological validation of networks generated with different parameter settings. We tested two mixing matrices and two β values, resulting in four simulation scenarios. The K ratio represents the ratio of diagonal to off-diagonal elements in the edge matrix K. Ten graphs were generated for each parameter combination.

standard deviation respectively. In all the scenarios the values of the reconstructed mixing matrix fall in the range of a single standard deviation. In Figure 4.1c, we present the same results more compactly, showing the average K values for diagonal (blue) and off-diagonal (orange) elements, with 'x' markers for real values and dots with error bars for the average and standard deviation. This figure highlights that links are reconstructed both within communities (diagonal) and between communities (off-diagonal) with reasonable accuracy in all scenarios.

4.2.2 It is not possible to reconstruct a mixing matrix in $\mathbb{S}^D$

In this section we want to test whether or not it is possible to distribute the nodes in a way that respects the mixing matrix of a graph with four communities of equal size, each strongly connected internally and equally connected to the other three communities.

To this end, we embedded one of the graphs obtained in the previous section, specifically the 'K ratio = 6 / 1' case with $\beta = 10$, using D-Mercator [98], a model-based embedding method that maps networks into (D)-sphere, *i.e.* $\mathbb{S}^D$. We tested five different dimensions: $D = \{1, 2, 3, 5, 10\}$ to empirically examine how the embedding space dimension affects the reconstruction of the mixing matrix K. For each dimension, we generated 10 embeddings. In each case, the model inferred the inverted temperature β, the parameter μ controlling the average degree, the radius of the space, the reconstructed hidden degree κ_i for each node, and each node's position on the D-sphere. Once embedded, we measured angular distances between nodes and calculated connection probabilities. The result are shown in Figure 4.2 and Table 4.2.

In figure Figure 4.2a we observe that D-Mercator accurately reconstructs the degree distribution across all dimensions, with notable deviations only in the $D = 2$ case. Interestingly, the embedding in $D = 1$ yields a good reconstruction of the degree distribution. We speculate that this could be due to the fact that the original graph was generated in a one-dimensional latent space, which may align more naturally with the $D = 1$ embedding configuration. Similar observations can be made regarding the reconstruction of average neighbor degree (Figure 4.2d) and clustering (Figure 4.2c). In each case, D-Mercator achieves satisfactory reconstructions, with small variations depending on the embedding dimension. However, as shown in Figure 4.2b, D-Mercator fails to accurately reconstruct the mixing matrix in any dimension. The figure displays the diagonal values (in blue) and off-diagonal values (in orange) for each dimension. Dots with error bars represent the mean and standard deviation, while 'x' markers indicate the expected values from the HGBM output. Notably, although D-Mercator consistently captures a higher value for the diagonal elements

Options	Assortativity	Giant Component	Clustering Coefficient
Original	-0.0887	0.9886	0.6757
D = 1	-0.0880 ± 0.0001	0.9823 ± 0.0007	0.6837 ± 0.0173
D = 2	-0.0808 ± 0.0007	0.7798 ± 0.4110	0.6616 ± 0.0085
D = 3	-0.0879 ± 0.0020	0.7684 ± 0.4050	0.5920 ± 0.0086
D = 5	-0.0867 ± 0.0006	0.7763 ± 0.4091	0.4567 ± 0.0050
D = 10	-0.0909 ± 0.0002	0.6879 ± 0.4747	0.3257 ± 0.0022

Table 4.2: Global metrics of a synthetic network generated with our method in D-Mercator with different dimensions D.

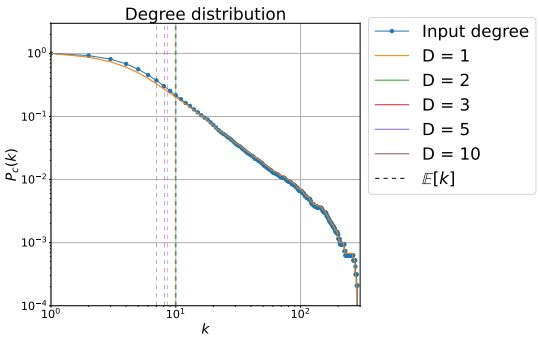
(indicating intra-community mixing), it fails to properly reconstruct the overall community structure imposed by the initial graph generated by HGBM.

Intuitively, we would expect that a mixing matrix of this type would be impossible to fully reconstruct in a one-dimensional space. With four equally sized communities, each with strong internal connectivity and uniform connectivity with nodes from other communities, no configuration in $\mathbb{S}^1$ can satisfy these conditions for all communities simultaneously. Simply put, if the communities are arranged in four sectors of the hyperbolic disk, intra-community links can be properly constructed. However, the inter-community links are distorted due to the fact that one community will inevitably be farther from the remaining two. This asymmetry prevents us from fulfilling all pairwise constraints at once. Consequently, higher-dimensional spaces (two or more dimensions) are likely required to balance the distances between community centers and accurately reconstruct the mixing matrix, ensuring equal connectivity across all communities.

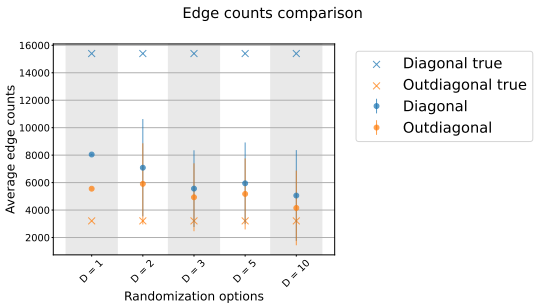
By analyzing the results in Figure 4.2b, we concluded that even increasing the dimensionality to $D = 2, 3, 5, 10$ does not resolve the reconstruction issue. These findings suggest that it is not trivial to decouple the concept of community from that of similarity, at least within this framework. This leads us to speculate that, in general, there may not be a way to distribute nodes on a D-sphere, regardless of its dimension, that satisfies all the constraints imposed by the mixing matrix.

4.2.3 HGBM preserves the mixing matrix in real network randomization

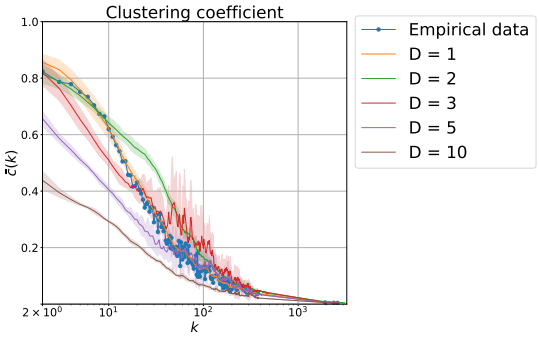
The goal of this section is showing that our model is capable of randomizing a real-world network while preserving its mixing matrix, degree distribution, and clustering. To accurately reconstruct the clustering, we used β as the only free parameter and tested six scenarios: $\beta = 2, 2.5, 5, 10, 50, 100$. Each value of β was chosen to balance the clustering while ensuring



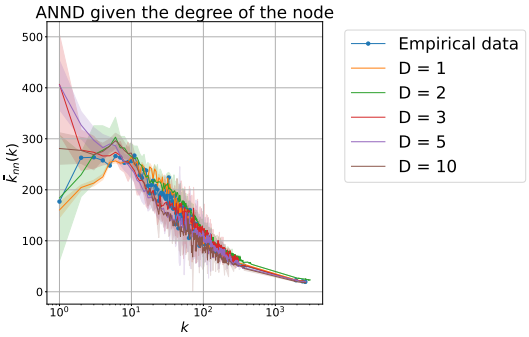
(a) Complementary cumulative distribution function (CCDF) of the degrees. The dashed vertical line indicates the expected average degree, $\langle k \rangle$.



(b) Average diagonal and off-diagonal elements of the mixing matrix K. 'x' markers indicate real values, while dots and error bars show the mean and standard deviation.



(c) Average clustering distribution given the degree class k.



(d) Average nearest neighbors degree given the degree class k.

Figure 4.2: Topological validation of HGBM networks embedded with D-Mercator with different embedding dimensions. For each dimension we generated 10 embeddings. In all panels, solid lines represent average values, and shaded regions indicate standard deviations

the key structural properties of the network were maintained. In future work, we will extend this approach by reconstructing β with a maximum likelihood approach, here we will reconstruct the clustering coefficient of the network empirically by testing several values of β.

We selected one dataset from SocioPatterns [186], which provides face-to-face interaction data of children and teaching staff in a school environment. The edges in this graph represent the cumulative durations (in seconds) of face-to-face interactions measured with RFID technology described in subsection 2.3.1

The range of contact durations spans from 20 seconds to 9,300 seconds (2 hours and 35 minutes) and is measured over the course of a single school day. The HGBM model introduced earlier does not support weighted graphs, necessitating a method for edge selection. For our analysis, we selected only the top 25% of contacts by duration. This choice was made to focus on the topological properties of the contact network most relevant to potential contagion scenarios. Including edges with cumulative contact durations as short as 20 seconds seemed unrealistic for modeling significant transmission events. By considering only the longest 25% of contacts, we emphasize interactions most likely to drive the spread of an outbreak, especially in scenarios where transmission requires sustained exposure.

The threshold of 25% was chosen to strike a balance between retaining a sufficient number of edges for robust analysis and focusing on meaningful, high-risk interactions. Moreover, given that the data covers only one school day, this threshold accounts for the potential significance of these longer-duration contacts in the context of outbreaks that might take several days to develop. Contacts of this duration are more likely to represent consistent interactions that could contribute significantly to sustained transmission over time.

Out of the initial 236 nodes and $5,899$ edges, the resulting graph contains 236 nodes and $1,596$ edges, with contact durations ranging from 160 seconds to 9,300 seconds. We report the results in Figure 4.3 and Table 4.3.

To compare our result, we made five embeddings with D-Mercator with the same dataset and proceed with a similar topological validation. For the embedding space we choose $D = \{1, 2, 3, 5, 10\}$. Results of the network randomized with D-Mercator are shown in Figure 4.4 and Table 4.4. By comparing the results, we observe that while both methods are capable of reproducing similar degree distributions and assortativity, D-Mercator struggles to accurately reconstruct the mixing matrix compared to HGBM. The average relative error for the mixing matrix is 47% for $D = 1$, 56% for $D = 2$, 70% for $D = 3$, 174% for $D = 5$, and 234% for $D = 10$. In contrast, the average relative error for HGBM remains consistently around 7% for each value of β.

Regarding the degree distribution, both methods perform well, with D-Mercator con-

taining the error within acceptable limits, and HGBM slightly understimating the average degree. A similar trend is observed for the clustering spectrum (*i.e.*, the average clustering coefficient given the degree class k) and the average nearest-neighbors degree, where both methods yield comparable results.

For the clustering coefficient, HGBM with $\beta = 5$ is capable of reconstructing a value similar to the original graph.

Regarding the giant component, D-Mercator and HGBM perform well, achieving a giant component that includes up to 99% of the nodes.

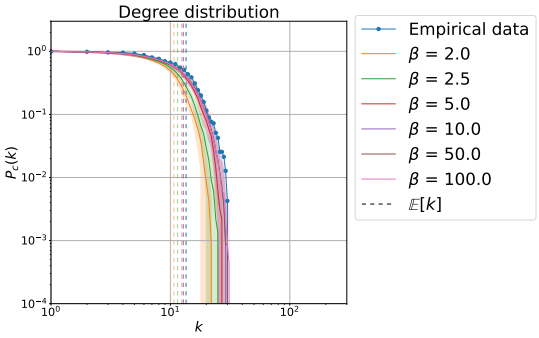
A final observation concerns degree assortativity, which measures the correlation between node degrees in a network. Our model, which uses a uniform distribution of latent angles, should guarantee zero degree assortativity, in line with theoretical predictions from [22]. While HGBM exhibits positive assortativity, D-Mercator maintains degree assortativity values closer to zero for $D = 5$, demonstrating greater consistency in this respect.

Options	Assortativity	Giant Component	Clustering Coefficient
Original	0.2394	1.0000	0.4736
$\beta = 2.0$	0.0928 ± 0.0254	0.9942 ± 0.0027	0.3468 ± 0.0077
$\beta = 2.5$	0.1237 ± 0.0330	0.9932 ± 0.0021	0.4050 ± 0.0085
$\beta = 5.0$	0.1274 ± 0.0369	0.9927 ± 0.0047	0.4970 ± 0.0057
$\beta = 10.0$	0.1423 ± 0.0482	0.9954 ± 0.0030	0.5173 ± 0.0105
$\beta = 50.0$	0.1220 ± 0.0245	0.9957 ± 0.0026	0.5263 ± 0.0141
$\beta = 100.0$	0.1210 ± 0.0351	0.9920 ± 0.0041	0.5262 ± 0.0104

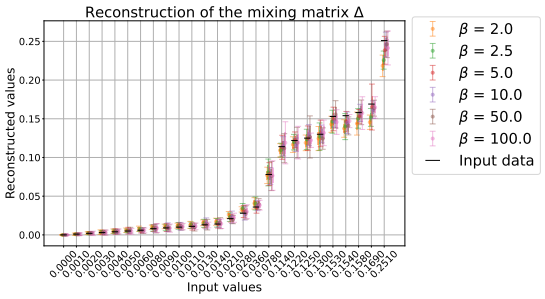
Table 4.3: Global metrics of SocioPattern generated with HGBM presented here with different β.

Options	Assortativity	Giant Component	Clustering Coefficient
Original	0.2394	1.0000	0.4736
D = 1	0.1273 ± 0.0058	0.9980 ± 0.0004	0.4674 ± 0.0031
D = 2	0.1577 ± 0.0119	0.9982 ± 0.0006	0.4650 ± 0.0080
D = 3	0.1496 ± 0.0094	0.9982 ± 0.0005	0.4401 ± 0.0056
D = 5	0.0059 ± 0.0540	0.9985 ± 0.0020	0.3437 ± 0.0096
D = 10	0.0251 ± 0.0221	0.9995 ± 0.0005	0.2266 ± 0.0058

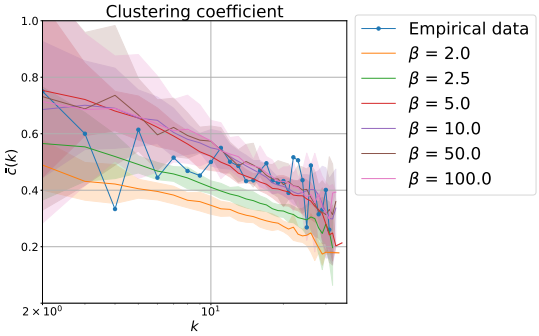
Table 4.4: Global metrics of SocioPattern generated with D-Mercator with different dimensions, D.



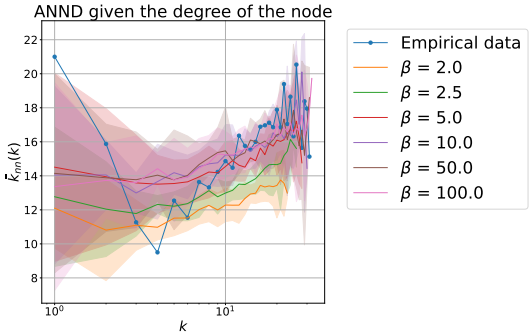
(a) Complementary cumulative distribution function (CCDF) of the degree distribution. The dashed vertical line shows the expected average degree, $\langle k \rangle$.



(b) Reconstructed mixing matrix values compared to real values under varying conditions. Boxplots represent entries of Δ, with one and two standard deviation errors shown by boxes and caps.

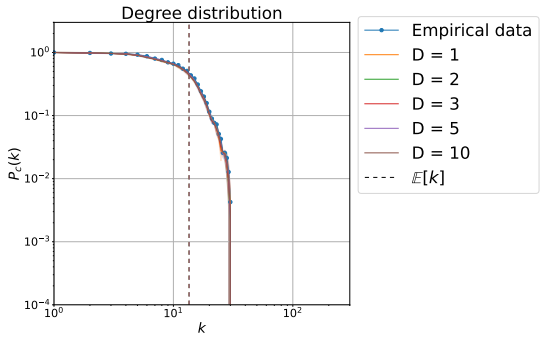


(c) Average clustering distribution given the degree class k.

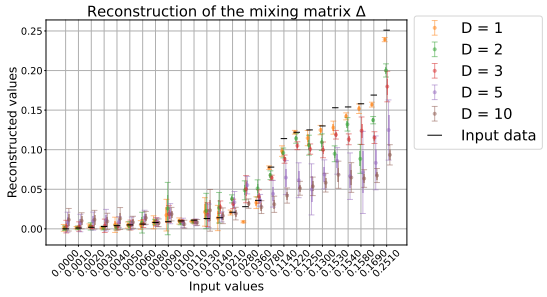


(d) Average nearest neighbors degree given the degree class k.

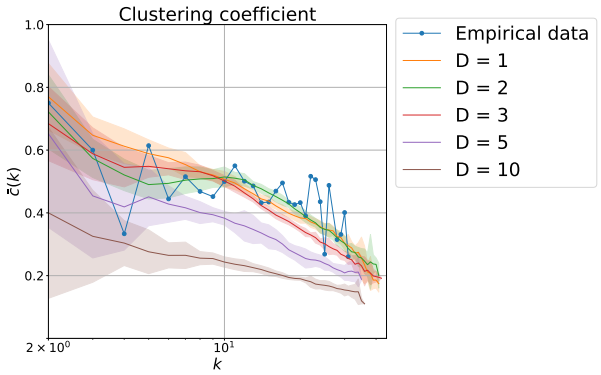
Figure 4.3: Topological validation of networks generated by randomizing the SocioPatterns dataset using the HGBM model. In all panels, solid lines represent average values, and shaded regions indicate standard deviations



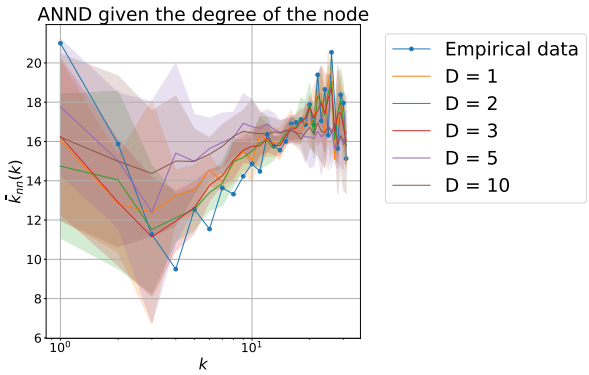
(a) Complementary cumulative distribution function (CCDF) of degree distribution. Solid lines show the average degree, with shaded regions representing the standard deviation. The dashed vertical line indicates the expected average degree, $\langle k \rangle$.



(b) Reconstructed mixing matrix values vs. real values under different conditions.



(c) Average clustering distribution given the degree class k.



(d) Average nearest neighbors degree given the degree class k.

Figure 4.4: Topological validation of networks randomizing the SocioPatterns dataset, generated using D-Mercator.

Chapter 5

Conclusions and Future Directions

5.1 Exploring Aquatic Food Webs

We analyzed the core-periphery structure of food webs, defined a method to obtain a critical sequence of nodes and investigated the representation of three-node motifs. Energy and matter flow predictably through the food web, beginning with the primary producers in the IN-periphery, passing through the core, and, when present, reaching the OUT-periphery, where they are exported from the ecosystem. This pattern is consistent with established findings in food web ecology (see, *e.g.*, [181, 37]). This flow not only sustains the species within the core but also ensures nutrient recycling, primarily through detrivores such as bacteria and plankton, which break down organic matter into simpler compounds recycled to organisms at higher trophic levels. In the core of food webs, the proportions of generalist and vulnerable species are quite similar while peripheries predominantly consist of vulnerable nodes. This is in accordance with what is observed in other ecological networks where generalist nodes of the network predominantly inhabit the core of ecological networks, emphasizing their crucial role in ecosystem stability [6, 151].

We found that the most critical nodes are often bacterial and planktonic communities, and that the robustness of a food web is positively correlated with the proportion of nodes belonging to the core and connectance (Equation 3.1), in agreement with [66]. It is important to note, however, that critical nodes in a food web are not limited to those at lower trophic levels. Many top predators play a roles in maintaining ecosystem stability and resilience through top-down control mechanisms, such as predation pressure that regulates prey populations and prevents trophic cascades [17]. While our approach emphasizes bottom-up energy flow as one important aspect of network robustness, this should not be conflated with a comprehensive measure of ecosystem stability or vulnerability, as energy flow represents

just one of many interacting mechanisms that govern food web dynamics.

We observed that the over-representation of the intraguild predation motif (S_2) implies the under-representation of apparent and exploitative competition motifs (S_4 and S_5, respectively). To explore this, we divided the food webs into two families based on the representation of intraguild predation. Intraguild predation is generally more prevalent in our dataset, supporting the suggestion that omnivory contributes to the stability of ecosystems [128, 129, 27]. Polis [161] proposed a hypothetical progression from pure competition to intraguild predation as ecosystems evolve, driven by aggression to monopolize local resources, which justifies why the over-representation of S_2 is associated with the under-representation of S_4 and S_5 in our data. We observed that in lakes and rivers food web, intraguild predation becomes under-represented, in contrast to what happens in other ecosystem types. Since we observed no other direct relationship between the food web measurements reported in this paper and the representation of intraguild predation, we proposed that ecosystems may shift from being dominated by intraguild predation to pure competition based on system productivity [60, 210] and that in lakes and rivers this motif is not sustained due to the presence of smaller specialized species. However, to confirm this hypothesis, we will need to include information on productivity levels in our food webs, and possibly repeat this analysis on a dataset that includes a larger number of freshwater ecosystems. This study contributes to the growing body of knowledge on food webs by providing an algorithmic framework to explore the interplay between biodiversity, species roles within ecosystems, and ecosystem stability.

Building upon the systematic structural characterization presented in this thesis, future research could expand its scope by exploring a broader range of topological features. This includes shifting focus beyond node-centric perspectives to emphasize the properties and significance of edges, their interactions, and their roles in shaping network dynamics. Such an approach would enhance the understanding of ecological implications derived from these metrics. Specifically, quantifying the direct and indirect effects between network components, such as between species in food webs, represents a valuable direction for further works. This could involve leveraging matrix-based approaches like the dietary proportions matrix, where element (i, j) represents the fraction that species i comprises of the total intake by species j. Raising this matrix to successive powers allows for the characterization of pathways of increasing length connecting entities, while summing these powers (leading to the Leontief structure matrix [195]) can quantify the total flow originating from one component that ultimately contributes to another, considering all direct and indirect routes. More in general, applying the measurements related with information theory explained in 1.2.1 can be a valuable direction for further investigation. Such analyses provide powerful insights into the

flow dynamics and distributed influence within the network. However, it is important to note that the full application and interpretation of these matrix-based flow measures are often most effective when detailed weighted interaction data, such as biomass flow between species, is available – a factor that poses a limitation for their comprehensive use with purely binary datasets and highlights the value of collecting rich, quantitative network data in future empirical studies. Future research could investigate network robustness through edge perturbation or removal. Analyzing the impact of removing individual or sets of trophic links would provide insights into the importance of specific dietary relationships and characterize links as functional or redundant based on their contribution to system stability [5].

5.2 A Model for Contact Networks with Known Community Structure

We introduced the Hidden-degree Geometric Block Model (HGBM), an extension of the $\mathbb{S}^1$ model that effectively incorporates group mixing. The HGBM builds on the intuition that any partition of the vertex set into communities induces a partition of their hidden degrees into hidden fitnesses to be normalized based on the connectivity of the pertaining community and that the hidden features (degrees and angular positions) can be extracted once for all to preserve the centrality of each node. HGBM introduce a community partition into the framework of maximum-entropy models with hidden geometries.

To verify that the HGBM serves the purpose it was designed for, we performed a comparative analysis between our model and the state-of-art method D-Mercator to embed the nodes of a graph into a D-sphere or a $D+1$-hyperboloid [98]. By conducting a series of simulations, we showed that the HGBM generates synthetic networks that concurrently exhibit three main features often found in real-world networks: a specific group mixing, a degree distribution, and a high clustering coefficient. Our work enables the integration of hyperbolic latent spaces into data-driven network models, generating more realistic contact networks while preserving empirical features like age-based or distance-based mixing [87]. Future work will focus on two directions: studying latent geometric models in higher dimensions, and investigating whether existing models reproduce additional real-world social network properties such as degree assortativity.

A Python implementation of the HGBM and all experimental software is available at our GitHub repository.

Bibliography

[1] Réka Albert and Albert-László Barabási. Statistical mechanics of complex networks. *Rev. Mod. Phys.*, 74:47–97, Jan 2002.

[2] Camille Albouy, Philippe Archambault, Ward Appeltans, Miguel B Araújo, David Beauchesne, Kevin Cazelles, Alyssa R Cirtwill, Marie-Josée Fortin, Nuria Galiana, Shawn J Leroux, et al. The marine fish food web is globally connected. *Nature Ecology & Evolution*, 3(8):1153–1161, 2019.

[3] Stefano Allesina, Antonio Bodini, and Cristina Bondavalli. Ecological subsystems via graph theory: the role of strongly connected components. *Oikos*, 110(1):164–176, 2005.

[4] Stefano Allesina, Antonio Bodini, and Cristina Bondavalli. Ecological subsystems via graph theory: the role of strongly connected components. *Oikos*, 110(1):164–176, 2005.

[5] Stefano Allesina, Antonio Bodini, and Mercedes Pascual. Functional links and robustness in food webs. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 364(1524):1701–1709, 2009.

[6] Tavis K. Anderson and Michael V. K. Sukhdeo. Host centrality in food web networks determines parasite diversity. *PLoS ONE*, 6(10):e26798, October 2011.

[7] Daniel Baird and Robert E Ulanowicz. The seasonal dynamics of the chesapeake bay ecosystem. *Ecological monographs*, 59(4):329–364, 1989.

[8] Benjamin Baiser, Rasha Elhesha, and Tamer Kahveci. Motifs in the assembly of food web networks. *Oikos*, 125(4):480–491, July 2015.

[9] Duygu Balcan, Hao Hu, Bruno Goncalves, Paolo Bajardi, Chiara Poletto, Jose J Ramasco, Daniela Paolotti, Nicola Perra, Michele Tizzoni, Wouter Van den Broeck, et al. Seasonal transmission potential and activity peaks of the new influenza a (h1n1): a monte carlo likelihood analysis based on human mobility. *BMC medicine*, 7:1–12, 2009.

[10] Albert-László Barabási and Réka Albert. Emergence of scaling in random networks. *Science*, 286(5439):509–512, 1999.

[11] Albert-László Barabási and Márton Pósfai. *Network science*. Cambridge University Press, Cambridge, 2016.

[12] Manuel Barange. *Marine ecosystems and global change*. Oxford University Press, Oxford, UK, 2010.

[13] Roberta Bardelli, Giorgio Mancinelli, Antonio Mazzola, and Salvatrice Vizzini. The atlantic blue crab callinectes sapidus spreading in the tyrrhenian sea: Evidence of an established population in the stagnone di marsala (sicily, southern italy). *NAŠE MORE: znanstveni časopis za more i pomorstvo*, 70(3 Special issue):177–183, 2023.

[14] Alain Barrat, Marc Barthelemy, and Alessandro Vespignani. *Dynamical processes on complex networks*. Cambridge university press, 2008.

[15] Jordi Bascompte, Pedro Jordano, Carlos J. Melián, and Jens M. Olesen. The nested assembly of plant–animal mutualistic networks. *Proceedings of the National Academy of Sciences*, 100(16):9383–9387, July 2003.

[16] Ishrat Bashir, Farooq A Lone, Rouf Ahmad Bhat, Shafat A Mir, Zubair A Dar, and Shakeel Ahmad Dar. Concerns and threats of contamination on aquatic ecosystems. *Bioremediation and biotechnology: sustainable approaches to pollution degradation*, pages 1–26, 2020.

[17] Julia K Baum and Boris Worm. Cascading top-down effects of changing oceanic predator abundances. *Journal of animal ecology*, 78(4):699–714, 2009.

[18] Andrew P Beckerman, Owen L Petchey, and Philip H Warren. Foraging biology predicts food web complexity. *Proceedings of the National Academy of Sciences*, 103(37):13745–13749, 2006.

[19] Michele Bellingeri and Antonio Bodini. Threshold extinction in food webs. *Theoretical Ecology*, 6(2):143–152, June 2012.

[20] Adam Berger, Stephen A Della Pietra, and Vincent J Della Pietra. A maximum entropy approach to natural language processing. *Computational linguistics*, 22(1):39–71, 1996.

[21] Massimo Bernaschi, Alessandro Celestini, Stefano Guarino, Enrico Mastrostefano, and Fabio Saracco. The fitness-corrected block model, or how to create maximum-entropy data-driven spatial social networks. *Scientific Reports*, 12(1):18206, 2022.

[22] Marián Boguñá, Dmitri Krioukov, Pedro Almagro, and M. Ángeles Serrano. Small worlds and clustering in spatial networks. *Phys. Rev. Res.*, 2:023040, Apr 2020.

[23] Marian Boguna, Ivan Bonamassa, Manlio De Domenico, Shlomo Havlin, Dmitri Krioukov, and M Ángeles Serrano. Network geometry. *Nature Reviews Physics*, 3(2):114–135, 2021.

[24] Marián Boguná and Romualdo Pastor-Satorras. Epidemic spreading in correlated complex networks. *Physical Review E*, 66(4):047104, 2002.

[25] Paolo Boldi, Marco Rosa, and Sebastiano Vigna. Robustness of social networks: Comparative results based on distance distributions. In Anwitaman Datta, Stuart Shulman, Baihua Zheng, Shou-De Lin, Aixin Sun, and Ee-Peng Lim, editors, *Social Informatics*, pages 8–21, Berlin, Heidelberg, 2011. Springer Berlin Heidelberg.

[26] Jonathan J Borrelli. Selection against instability: stable subgraphs are most frequent in empirical food webs. *Oikos*, 124(12):1583–1588, 2015.

[27] Jonathan J Borrelli, Stefano Allesina, Priyanga Amarasekare, Roger Arditi, Ivan Chase, John Damuth, Robert D Holt, Dmitrii O Logofet, Mark Novak, Rudolf P Rohr, et al. Selection on stability across ecological scales. *Trends in Ecology & Evolution*, 30(7):417–425, 2015.

[28] Tom Britton, Maria Deijfen, Andreas N Lagerås, and Mathias Lindholm. Epidemics on random graphs with tunable clustering. *Journal of Applied Probability*, 45(3):743–756, 2008.

[29] Andrei Broder, Ravi Kumar, Farzin Maghoul, Prabhakar Raghavan, Sridhar Rajagopalan, Raymie Stata, Andrew Tomkins, and Janet Wiener. Graph structure in the web. *Computer networks*, 33(1-6):309–320, 2000.

[30] Andrei Z. Broder, Ravi Kumar, Farzin Maghoul, Prabhakar Raghavan, Sridhar Rajagopalan, Raymie Stata, Andrew Tomkins, and Janet L. Wiener. Graph structure in the web. *Comput. Networks*, 33(1-6):309–320, 2000.

[31] Ulrich Brose, Eric L. Berlow, and Neo D. Martinez. 2.1 - from food webs to ecological networks: Linking non-linear trophic interactions with nutrient competition. In Peter de Ruiter, Volkmar Wolters, John C. Moore, and Kimberly Melville-Smith, editors, *Dynamic Food Webs*, volume 3 of *Theoretical Ecology Series*, pages 27–36. Academic Press, Burlington, 2005.

[32] A.C. Brown. Food relationships on the intertidal sandy beaches of the cape peninsula. *South African Journal of Science*, 60(2):35–41, 1964.

[33] Gabriel Budel, Maksim Kitsak, Rodrigo Aldecoa, Konstantin Zuev, and Dmitri Krioukov. Random hyperbolic graphs in d+ 1 dimensions. *Physical Review E*, 109(5):054131, 2024.

[34] J Camacho and A Arenas. Universal scaling in food-web structure? *Nature*, 435(7044):E3–E4, 2005.

[35] Luisa G Carvalheiro, Yvonne M Buckley, Rita Ventim, Simon V Fowler, and Jane Memmott. Apparent competition can compromise the safety of highly specific biocontrol agents. *Ecology Letters*, 11(7):690–700, 2008.

[36] Claudio Castellano and Romualdo Pastor-Satorras. Thresholds for epidemic spreading in networks. *Physical review letters*, 105(21):218701, 2010.

[37] Just Cebrian and Julien Lartigue. Patterns of herbivory and decomposition in aquatic and terrestrial ecosystems. *Ecological Monographs*, 74(2):237–259, 2004.

[38] Alessandro Celestini, Francesca Colaiori, Stefano Guarino, Enrico Mastrostefano, and Lena Rebecca Zastrow. Epidemic risk assessment from geographic population density. *Applied Network Science*, 7(1):39, 2022.

[39] Sonia Chaabane, Thibault de Garidel-Thoron, Julie Meilland, Olivier Sulpis, Thomas B Chalk, Geert-Jan A Brummer, P Graham Mortyn, Xavier Giraud, Hélène Howa, Nicolas Casajus, et al. Migrating is not enough for modern planktonic foraminifera in a changing ocean. *Nature*, pages 1–7, 2024.

[40] Deepayan Chakrabarti, Yang Wang, Chenxi Wang, Jurij Leskovec, and Christos Faloutsos. Epidemic thresholds in real networks. *ACM Transactions on Information and System Security (TISSEC)*, 10(4):1–26, 2008.

[41] C Anela Choy, Steven HD Haddock, and Bruce H Robison. Deep pelagic food web structure as revealed by in situ feeding observations. *Proceedings of the Royal Society B: Biological Sciences*, 284(1868):20172116, 2017.

[42] Fan Chung, Linyuan Lu, and Van Vu. Eigenvalues of random power law graphs. *Annals of Combinatorics*, 7(1):21–33, 2003.

[43] Alyssa R. Cirtwill, Giulio Valentino Dalla Riva, Marilia P. Gaiarsa, Malyon D. Bimler, E. Fernando Cagua, Camille Coux, and D. Matthias Dehling. A review of species role concepts in food webs. *Food Webs*, 16:e00093, September 2018.

[44] Alyssa R Cirtwill, Giulio Valentino Dalla Riva, Marilia P Gaiarsa, Malyon D Bimler, E Fernando Cagua, Camille Coux, and D Matthias Dehling. A review of species role concepts in food webs. *Food Webs*, 16:e00093, 2018.

[45] Alyssa R. Cirtwill and Anna Eklöf. Feeding environment and other traits shape species' roles in marine food webs. *Ecology Letters*, 21(6):875–884, 2018.

[46] G.P. Closs, S.R. Balcombe, and M.J. Shirley. Generalist predators, interaction strength and food-web stability. volume 28 of *Advances in Ecological Research*, pages 93–126. Academic Press, 1999.

[47] Joel E. Cohen, Frédéric Briand, and Charles M. Newman. *Data on 113 Community Food Webs*, pages 203–279. Springer Berlin Heidelberg, Berlin, Heidelberg, 1990.

[48] Vittoria Colizza, Alain Barrat, Marc Barthélemy, and Alessandro Vespignani. The role of the airline transportation network in the prediction and predictability of global epidemics. *Proceedings of the National Academy of Sciences*, 103(7):2015–2020, 2006.

[49] Mathieu Colléter, Audrey Valls, Jérôme Guitton, Morissette Lyne, Francisco Arreguín-Sánchez, Villy Christensen, Didier D Gascuel, and Daniel Pauly. *EcoBase: a repository solution to gather and communicate information from EwE models*. PhD thesis, Fisheries Centre, University of British Columbia, Canada, 2013.

[50] Gabor Csardi and Tamas Nepusz. The igraph software package for complex network research. *InterJournal*, Complex Systems:1695, 2006.

[51] P. Csermely, A. London, L.-Y. Wu, and B. Uzzi. Structure and dynamics of core/periphery networks. *Journal of Complex Networks*, 1(2):93–123, October 2013.

[52] M. R. T. Dale. *Applying Graph Theory in Ecological Research*. Cambridge University Press, 2017. ISBN: 9781316105450.

[53] Marco D'Arienzo and Angela Coniglio. Assessment of the sars-cov-2 basic reproduction number, r0, based on the early phase of covid-19 outbreak in italy. *Biosafety and health*, 2(2):57–59, 2020.

[54] Wesley Dáttilo, Nubia Lara-Rodríguez, Pedro Jordano, Paulo R Guimarães Jr, John N Thompson, Robert J Marquis, Lucas P Medeiros, Raul Ortiz-Pulido, Maria A Marcos-García, and Victor Rico-Gray. Unravelling darwin's entangled bank: architecture and robustness of mutualistic networks with multiple interaction types. *Proceedings of the Royal Society B: Biological Sciences*, 283(1843):20161564, 2016.

[55] Patrice David, Elisa Thebault, Orlane Anneville, P-F Duyck, Elodie Chapuis, and Nicolas Loeuille. Impacts of invasive species on food webs: a review of empirical data. *Advances in ecological research*, 56:1–60, 2017.

[56] Andrea De Martino and Daniele De Martino. An introduction to the maximum entropy approach and its application to inference problems in biology. *Heliyon*, 4(4), 2018.

[57] D Matthias Dehling and Daniel B Stouffer. Bringing the eltonian niche into functional diversity. *Oikos*, 127(12):1711–1723, 2018.

[58] Eva Delmas, Mathilde Besson, Marie-Hélène Brice, Laura A. Burkle, Giulio V. Dalla Riva, Marie-Josée Fortin, Dominique Gravel, Paulo R. Guimarães, David H. Hembry, Erica A. Newman, Jens M. Olesen, Mathias M. Pires, Justin D. Yeakel, and Timothée Poisot. Analysing ecological networks of species interactions. *Biological Reviews*, 94(1):16–36, June 2018.

[59] Béatrice Désy, Patrick Desrosiers, and Antoine Allard. Dimension matters when modeling network communities in hyperbolic spaces. *PNAS nexus*, 2(5):pgad136, 2023.

[60] Sebastian Diehl and Margit Feissel. Intraguild prey suffer from enrichment of their resources: a microcosm experiment with ciliates. *Ecology*, 82(11):2977–2983, 2001.

[61] Scott C Doney, Mary Ruckelshaus, J Emmett Duffy, James P Barry, Francis Chan, Chad A English, Heather M Galindo, Jacqueline M Grebmeier, Anne B Hollowed, Nancy Knowlton, et al. Climate change impacts on marine ecosystems. *Annual review of marine science*, 4(1):11–37, 2012.

[62] Gaogao Dong, Jianxi Gao, Ruijin Du, Lixin Tian, H Eugene Stanley, and Shlomo Havlin. Robustness of network of networks under targeted attack. *Physical Review E—Statistical, Nonlinear, and Soft Matter Physics*, 87(5):052804, 2013.

[63] Robin IM Dunbar. Constraints on the evolution of social institutions and their implications for information flow. *Journal of Institutional Economics*, 7(3):345–371, 2011.

[64] Jennifer A Dunne. The network structure of food webs. *Ecological networks: linking structure to dynamics in food webs*, pages 27–86, 2006.

[65] Jennifer A Dunne. Food webs. *Encyclopedia of complexity and systems science*, 1:3661–3682, 2009.

[66] Jennifer A Dunne, Richard J Williams, and Neo D Martinez. Network structure and biodiversity loss in food webs: robustness increases with connectance. *Ecology letters*, 5(4):558–567, 2002.

[67] Jennifer A Dunne, Richard J Williams, and Neo D Martinez. Network structure and robustness of marine food webs. *Marine Ecology Progress Series*, 273:291–302, 2004.

[68] Ken Eames, Shweta Bansal, Simon Frost, and Steven Riley. Six challenges in measuring contact networks for use in modeling. *Epidemics*, 10:72–77, 2015.

[69] Joan G Ehrenfeld and Louis A Toth. Restoration ecology and the ecosystem perspective. *Restoration ecology*, 5(4):307–317, 1997.

[70] Timon Elmer, Krishna Chaitanya, Prateek Purwar, and Christoph Stadtfeld. The validity of rfid badges measuring face-to-face interactions. *Behavior research methods*, 51:2120–2138, 2019.

[71] Stephen Eubank, Hasan Guclu, VS Anil Kumar, Madhav V Marathe, Aravind Srinivasan, Zoltan Toroczkai, and Nan Wang. modeling disease outbreaks in realistic urban social networks. *Nature*, 429(6988):180–184, 2004.

[72] Neil M Ferguson, Derek AT Cummings, Christophe Fraser, James C Cajka, Philip C Cooley, and Donald S Burke. Strategies for mitigating an influenza pandemic. *Nature*, 442(7101):448–452, 2006.

[73] Santo Fortunato. Community detection in graphs. *Physics reports*, 486(3-5):75–174, 2010.

[74] Julie Fournet and Alain Barrat. Contact patterns among high school students. *PloS one*, 9(9):e107878, 2014.

[75] Laura Fumanelli, Marco Ajelli, Piero Manfredi, Alessandro Vespignani, and Stefano Merler. Inferring the structure of social contacts from demographic data in the analysis of infectious diseases spread. *PLOS Computational Biology*, 8(9):1–10, 09 2012.

[76] Ryan J. Gallagher, Jean-Gabriel Young, and Brooke Foucault Welles. A clarified typology of core-periphery structure in networks. *Science Advances*, 7(12), 2021. Cited by: 23; All Open Access, Gold Open Access, Green Open Access.

[77] Guillermo García-Pérez, Antoine Allard, M Ángeles Serrano, and Marián Boguñá. Mercator: uncovering faithful hyperbolic embeddings of complex networks. *New Journal of Physics*, 21(12):123033, 2019.

[78] Guillermo García-Pérez, Marián Boguñá, and M Ángeles Serrano. Multiscale unfolding of real networks by geometric renormalization. *Nature Physics*, 14(6):583–589, 2018.

[79] Guillermo García-Pérez, M. Ángeles Serrano, and Marián Boguñá. Soft communities in similarity space. *Journal of Statistical Physics*, 173(3–4):775–782, June 2018.

[80] Colin J. Garroway, Evelien de Greef, Kyle J. Lefort, Matt J. Thorstensen, Andrew D. Foote, Cory J. D. Matthews, Jeff W. Higdon, Caila E. Kucheravy, Stephen D. Petersen, Aqqalu Rosing-Asvid, Fernando Ugarte, Rune Dietz, and Steven H. Ferguson. Climate change introduces threatened killer whale populations and conservation challenges to the arctic. *Global Change Biology*, 30(6):e17352, 2024. e17352 GCB-23-3010.R1.

[81] Mathieu Génois, Christian L Vestergaard, Julie Fournet, André Panisson, Isabelle Bonmarin, and Alain Barrat. Data on face-to-face contacts in an office building suggest a low-cost vaccination strategy based on community linkers. *Network Science*, 3(3):326–347, 2015.

[82] Darren P Giling, Anne Ebeling, Nico Eisenhauer, Sebastian T Meyer, Christiane Roscher, Michael Rzanny, Winfried Voigt, Wolfgang W Weisser, and Jes Hines. Plant diversity alters the representation of motifs in food webs. *Nature Communications*, 10(1):1226, 2019.

[83] Giulia Giordano, Franco Blanchini, Raffaele Bruno, Patrizio Colaneri, Alessandro Di Filippo, Angela Di Matteo, and Marta Colaneri. modeling the covid-19 epidemic and implementation of population-wide interventions in italy. *Nature medicine*, 26(6):855–860, 2020.

[84] Michelle Girvan and Mark EJ Newman. Community structure in social and biological networks. *Proceedings of the national academy of sciences*, 99(12):7821–7826, 2002.

[85] Oscar Godoy. Coexistence theory as a tool to understand biological invasions in species interaction networks: Implications for the study of novel ecosystems. *Functional Ecology*, 33(7):1190–1201, 2019.

[86] Mark S Granovetter. The strength of weak ties. *American journal of sociology*, 78(6):1360–1380, 1973.

[87] Stefano Guarino, Enrico Mastrostefano, Massimo Bernaschi, Alessandro Celestini, Marco Cianfriglia, Davide Torre, and Lena Rebecca Zastrow. Inferring urban social networks from publicly available data. *Future Internet*, 13(5):108, 2021.

[88] Sunetra Gupta, Roy M Anderson, and Robert M May. Networks of sexual contacts: implications for the pattern of spread of hiv. *Aids*, 3(12):807–818, 1989.

[89] Niel Hens, Nele Goeyvaerts, Marc Aerts, Ziv Shkedy, Pierre Van Damme, and Philippe Beutels. Mining social mixing patterns for infectious disease models based on a two-day population survey in belgium. *BMC infectious diseases*, 9:1–18, 2009.

[90] Robert D Holt. From metapopulation dynamics to community structure: some consequences of spatial heterogeneity. In *Metapopulation biology*, pages 149–164. Elsevier, 1997.

[91] Robert D Holt, James Grover, and David Tilman. Simple rules for interspecific dominance in systems with exploitative and apparent competition. *The American Naturalist*, 144(5):741–771, 1994.

[92] Robert D Holt and Gary R Huxel. Alternative prey and the dynamics of intraguild predation: theoretical perspectives. *Ecology*, 88(11):2706–2712, 2007.

[93] Robert D Holt and Gary A Polis. A theoretical framework for intraguild predation. *The American Naturalist*, 149(4):745–764, 1997.

[94] Yuheng Huang, Jiayang Song, Zhijie Wang, Shengming Zhao, Huaming Chen, Felix Juefei-Xu, and Lei Ma. Look before you leap: An exploratory study of uncertainty measurement for large language models. *arXiv preprint arXiv:2307.10236*, 2023.

[95] David R Hunter, Steven M Goodreau, and Mark S Handcock. Goodness of fit of social network models. *Journal of the american statistical association*, 103(481):248–258, 2008.

[96] M. Huxham, S. Beaney, and D. Raffaelli. Do parasites reduce the chances of triangulation in a real food web? *Oikos*, 76(2):284–300, 1996.

[97] Fabrizio Iozzi, Francesco Trusiano, Matteo Chinazzi, Francesco C. Billari, Emilio Zagheni, Stefano Merler, Marco Ajelli, Emanuele Del Fava, and Piero Manfredi. Little

italy: An agent-based approach to the estimation of contact patterns- fitting predicted matrices to serological data. *PLOS Computational Biology*, 6(12):1–10, 12 2010.

[98] Robert Jankowski, Antoine Allard, Marián Boguñá, and M. Ángeles Serrano. The d-mercator method for the multidimensional hyperbolic embedding of real networks. *Nature Communications*, 14(1):7585, Nov 2023.

[99] Arne Janssen, Maurice W Sabelis, Sara Magalhães, Marta Montserrat, and Tessa Van Der Hammen. Habitat structure affects intraguild predation. *Ecology*, 88(11):2713–2719, 2007.

[100] Felichism Kabo, Yongha Hwang, Margaret Levenstein, and Jason Owen-Smith. Shared paths to the lab: A sociospatial network analysis of collaboration. *Environment and Behavior*, 47(1):57–84, 2015.

[101] Matt Keeling. The implications of network structure for epidemic dynamics. *Theoretical population biology*, 67(1):1–8, 2005.

[102] Matt J Keeling and Ken TD Eames. Networks and epidemic models. *Journal of the royal society interface*, 2(4):295–307, 2005.

[103] Matthew J Keeling. The effects of local spatial structure on epidemiological invasions. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, 266(1421):859–867, 1999.

[104] Janis Klaise and Samuel Johnson. The origin of motif families in food webs. *Scientific reports*, 7(1):16197, 2017.

[105] Bianka Kovács and Gergely Palla. The inherent community structure of hyperbolic networks. *Scientific Reports*, 11(1), August 2021.

[106] Dmitri Krioukov. Clustering implies geometry in networks. *Physical review letters*, 116(20):208302, 2016.

[107] Dmitri Krioukov, Fragkiskos Papadopoulos, Maksim Kitsak, Amin Vahdat, and Marián Boguná. Hyperbolic geometry of complex networks. *Physical Review E—Statistical, Nonlinear, and Soft Matter Physics*, 82(3):036106, 2010.

[108] Ellen Kuhl. *Computational epidemiology*. Springer, 2021.

[109] Pietro Landi, Henintsoa O. Minoarivelo, Åke Brännström, Cang Hui, and Ulf Dieckmann. Complexity and stability of ecological networks: a review of the theory. *Population Ecology*, 60(4):319–345, July 2018.

[110] Géraldine Lassalle, Pierre Bourdaud, Blanche Saint-Béat, Sébastien Rochette, and Nathalie Niquil. A toolbox to evaluate data reliability for whole-ecosystem models: Application on the bay of biscay continental shelf food-web model. *Ecological modeling*, 285:13–21, 2014.

[111] Maxime Lenormand, Sylvie Huet, Floriana Gargiulo, and Guillaume Deffuant. A universal model of commuting networks. *PLoS ONE*, 7(10):e45985, October 2012.

[112] LibreTexts. Quantifying food webs, 2024. Accessed: 2024-10-09.

[113] Fredrik Liljeros, Christofer R Edling, Luis A Nunes Amaral, H Eugene Stanley, and Yvonne Åberg. The web of human sexual contacts. *Nature*, 411(6840):907–908, 2001.

[114] L. L. Lima and A. P. F. Atman. Complexity in the dengue spreading: A network analysis approach. *PLOS ONE*, 18(8):e0289690, August 2023.

[115] Raymond L. Lindeman. The trophic-dynamic aspect of ecology. *Ecology*, 23(4):399–417, 1942.

[116] Cheng Liu, Yina Yao, and Min Zhu. Structural statistics of a human contact network in a research institute. In *Proceedings of the 6th ACM SIGSPATIAL International Workshop on Emergency Management Using GIS*, EM-GIS '20, New York, NY, USA, 2020. Association for Computing Machinery.

[117] Michael Lodge, David Johnson, Gwenaëlle Le Gurun, Markus Wengler, Phil Weaver, and Vikki Gunn. Seabed mining: International seabed authority environmental management plan for the clarion–clipperton zone. a partnership approach. *Marine Policy*, 49:66–72, 2014.

[118] Michel Loreau, Shahid Naeem, Pablo Inchausti, Jan Bengtsson, J Philip Grime, Andrew Hector, David U Hooper, Michael A Huston, David Raffaelli, Bernhard Schmid, et al. Biodiversity and ecosystem functioning: current knowledge and future challenges. *science*, 294(5543):804–808, 2001.

[119] Xueke Lu, Clare Gray, Lee E. Brown, Mark E. Ledger, Alexander M. Milner, Raúl J. Mondragón, Guy Woodward, and Athen Ma. Drought rewires the cores of food webs. *Nature Climate Change*, 6(9):875–878, Sep 2016.

[120] Sean Christopher Lusk. Evaluating the effects of threadfin shad on largemouth bass and bluegill populations in small impoundments. Master's thesis, Auburn University, 2016.

[121] Malcolm Luxton. The marine littoral mites of the new zealand region. *Journal of the Royal Society of New Zealand*, 20(4):367–418, December 1990.

[122] Tomás Ignacio Marina, Leonardo A. Saravia, Georgina Cordone, Vanesa Salinas, Santiago R. Doyle, and Fernando R. Momo. Architecture of marine food webs: To be or not be a 'small-world'. *PLOS ONE*, 13(5):e0198217, May 2018.

[123] Neo D. Martinez. Artifacts or attributes? effects of resolution on the little rock lake food web. *Ecological Monographs*, 61(4):367–392, 1991.

[124] Rossana Mastrandrea, Julie Fournet, and Alain Barrat. Contact patterns in a high school: a comparison between data collected using wearable sensors, contact diaries and friendship surveys. *PloS one*, 10(9):e0136497, 2015.

[125] Joan T. Matamalas, Alex Arenas, and Sergio Gómez. Effective approach to epidemic containment using link equations in complex networks. *Science Advances*, 4(12):eaau4212, 2018.

[126] Robert M May. Qualitative stability in model ecosystems. *Ecology*, 54(3):638–641, 1973.

[127] Julian J McAuley, Luciano da Fontoura Costa, and Tibério S Caetano. Rich-club phenomenon across complex network hierarchies. *Applied physics letters*, 91(8), 2007.

[128] Kevin McCann and Alan Hastings. Re–evaluating the omnivory–stability relationship in food webs. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, 264(1385):1249–1254, 1997.

[129] Kevin McCann, Alan Hastings, and Gary R Huxel. Weak trophic interactions and the balance of nature. *Nature*, 395(6704):794–798, 1998.

[130] Kevin McCann and Peter Yodzis. Biological conditions for chaos in a three-species food chain. *Ecology*, pages 561–564, 1994.

[131] Kevin S McCann, JB Rasmussen, and James Umbanhowar. The dynamics of spatially coupled food webs. *Ecology letters*, 8(5):513–523, 2005.

[132] Eve McDonald-Madden, Regis Sabbadin, Edward T Game, Peter WJ Baxter, Iadine Chades, and Hugh P Possingham. Using food-web theory to conserve ecosystems. *Nature communications*, 7(1):10245, 2016.

[133] Miller McPherson, Lynn Smith-Lovin, and James M Cook. Birds of a feather: Homophily in social networks. *Annual review of sociology*, 27(1):415–444, 2001.

[134] Jane Memmott. Food webs: a ladder for picking strawberries or a practical tool for practical problems? *Philosophical Transactions of the Royal Society B: Biological Sciences*, 364(1524):1693–1699, 2009.

[135] Jane Memmott, Paul G Craze, Nickolas M Waser, and Mary V Price. Global warming and the disruption of plant–pollinator interactions. *Ecology letters*, 10(8):710–717, 2007.

[136] Jane Memmott, Nickolas M Waser, and Mary V Price. Tolerance of pollination networks to species extinctions. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, 271(1557):2605–2611, 2004.

[137] Stefano Merler and Marco Ajelli. The role of population heterogeneity and human mobility in the spread of pandemic influenza. *Proceedings of the Royal Society B: Biological Sciences*, 277(1681):557–565, 2010.

[138] L. Meyers. Contact network epidemiology: Bond percolation applied to infectious disease prediction and control. *Bulletin of the American Mathematical Society*, 44:63–86, 2006.

[139] Vincent Miele, Rodrigo Ramos-Jiliberto, and Diego P. Vázquez. Core–periphery dynamics in a plant–pollinator network. *Journal of Animal Ecology*, 89(7):1670–1677, April 2020.

[140] R. Milo, S. Shen-Orr, S. Itzkovitz, N. Kashtan, D. Chklovskii, and U. Alon. Network motifs: Simple building blocks of complex networks. *Science*, 298(5594):824–827, 2002.

[141] Dina Mistry, Maria Litvinova, Ana Pastore y Piontti, Matteo Chinazzi, Laura Fumanelli, Marcelo FC Gomes, Syed A Haque, Quan-Hui Liu, Kunpeng Mu, Xinyue Xiong, et al. Inferring high-resolution human mixing patterns for disease modeling. *Nature communications*, 12(1):323, 2021.

[142] John C. Moore, Eric L. Berlow, David C. Coleman, Peter C. de Ruiter, Quan Dong, Alan Hastings, Nancy Collins Johnson, Kevin S. McCann, Kim Melville, Peter J.

Morin, Knute Nadelhoffer, Amy D. Rosemond, David M. Post, John L. Sabo, Kate M. Scow, Michael J. Vanni, and Diana H. Wall. Detritus, trophic dynamics and biodiversity. *Ecology Letters*, 7(7):584–600, 2004.

[143] Joël Mossong, Niel Hens, Mark Jit, Philippe Beutels, Kari Auranen, Rafael Mikolajczyk, Marco Massari, Stefania Salmaso, Gianpaolo Scalia Tomba, Jacco Wallinga, et al. Social contacts and mixing patterns relevant to the spread of infectious diseases. *PLoS medicine*, 5(3):e74, 2008.

[144] Alessandro Muscoloni and Carlo Vittorio Cannistraci. A nonuniform popularity-similarity optimization (npso) model to efficiently generate realistic complex networks with communities. *New Journal of Physics*, 20(5):052002, may 2018.

[145] Mark Newman. *Networks: An Introduction*. Oxford University Press, 06 2018.

[146] Mark EJ Newman. Spread of epidemic disease on networks. *Physical review E*, 66(1):016128, 2002.

[147] Mark EJ Newman. Modularity and community structure in networks. *Proceedings of the national academy of sciences*, 103(23):8577–8582, 2006.

[148] Mark EJ Newman and Juyong Park. Why social networks are different from other types of networks. *Physical review E*, 68(3):036122, 2003.

[149] Savannah Nuwagaba, Feng Zhang, and Cang Hui. Robustness of rigid and adaptive networks to species loss. *PLOS ONE*, 12(12):e0189086, December 2017.

[150] Ömer K. Coskun, Volkan Özen, Scott D. Wankel, and William D. Orsi. Quantifying population-specific growth in benthic bacterial communities under low oxygen using h218o. *The ISME Journal*, 13(6):1546–1559, February 2019.

[151] Rubén Darío Palacio, Carlos Valderrama-Ardila, and Gustavo H. Kattan. Generalist species have a central role in a highly diverse plant–frugivore network. *Biotropica*, 48(3):349–355, 2016.

[152] Juyong Park and Mark EJ Newman. Statistical mechanics of networks. *Physical Review E—Statistical, Nonlinear, and Soft Matter Physics*, 70(6):066117, 2004.

[153] Romualdo Pastor-Satorras and Alessandro Vespignani. Epidemic spreading in scale-free networks. *Phys. Rev. Lett.*, 86:3200–3203, Apr 2001.

[154] Nilakantha Paudel, Loukas Georgiadis, and Giuseppe F. Italiano. Computing critical nodes in directed graphs. *ACM J. Exp. Algorithmics*, 23, 2018.

[155] K Paviour-Smith. The biotic community of a salt meadow in new zealand. *Transactions and Proceedings of the Royal Society of New Zealand*, 83:525–554, 1956.

[156] Emanuele Pepe, Paolo Bajardi, Laetitia Gauvin, Filippo Privitera, Brennan Lake, Ciro Cattuto, and Michele Tizzoni. Covid-19 outbreak response, a dataset to assess mobility changes in italy following national lockdown. *Scientific Data*, 7(1):230, Jul 2020.

[157] Daniel M Perkins, Julia Reiss, Gabriel Yvon-Durocher, and Guy Woodward. Global change and food webs in running waters. *Hydrobiologia*, 657:181–198, 2010.

[158] Bruce J. Peterson and Brian Fry. Stable isotopes in ecosystem studies. *Annual Review of Ecology and Systematics*, 18:293–320, 1987.

[159] Tony J Pitcher. Back-to-the-future: a fresh policy initiative for fisheries and a restoration ecology for ocean ecosystems. *Philos Trans R Soc Lond B Biol Sci*, 360(1453):107–121, January 2005.

[160] Michael JO Pocock, Darren M Evans, and Jane Memmott. The robustness and restoration of a network of ecological networks. *Science*, 335(6071):973–977, 2012.

[161] GA Polis. Exploitation competition and the evolution of interference, cannibalism, and intraguild predation in age/size-structured populations. In *Size-structured populations: ecology and evolution*, pages 185–202. Springer, 1988.

[162] Gary A Polis, Christopher A Myers, and Robert D Holt. The ecology and evolution of intraguild predation: potential competitors that eat each other. *Annual review of ecology and systematics*, pages 297–330, 1989.

[163] Gail E. Potter and Niel Hens. A Penalized Likelihood Approach to Estimate Within-Household Contact Networks from Egocentric Data. *Journal of the Royal Statistical Society Series C: Applied Statistics*, 62(4):629–648, 03 2013.

[164] Sadguru Prakash. Impact of climate change on aquatic ecosystem and its biodiversity: An overview. *International Journal of Biological Innovations*, 3(2), 2021.

[165] Kiesha Prem, Alex R Cook, and Mark Jit. Projecting social contact matrices in 152 countries using contact surveys and demographic data. *PLoS computational biology*, 13(9):e1005697, 2017.

[166] Steve Pressé, Kingshuk Ghosh, Julian Lee, and Ken A Dill. Principles of maximum entropy and maximum caliber in statistical physics. *Reviews of Modern Physics*, 85(3):1115–1141, 2013.

[167] David Raffaelli. How extinction patterns affect ecosystems. *Science*, 306(5699):1141–1142, 2004.

[168] Neil Rooney, Kevin S McCann, and John C Moore. A landscape theory for food web architecture. *Ecology letters*, 11(8):867–881, 2008.

[169] Axel G Rossberg. *Food webs and biodiversity: foundations, models, data*. John Wiley & Sons, 2013.

[170] Serguei Saavedra, Daniel B Stouffer, Brian Uzzi, and Jordi Bascompte. Strong contributors to network persistence are the most vulnerable to extinction. *Nature*, 478(7368):233–235, 2011.

[171] John L. Sabo, Jacques C. Finlay, Theodore Kennedy, and David M. Post. The role of discharge variation in scaling of drainage area and food chain length in rivers. *Science*, 330(6006):965–967, November 2010.

[172] Marcel Salathé and James H Jones. Dynamics and control of diseases in networks with community structure. *PLoS computational biology*, 6(4):e1000736, 2010.

[173] Don C Schmitz and Daniel Simberloff. Biological invasions: a growing threat. *Issues in Science and Technology*, 13(4):33–40, 1997.

[174] Thomas W. Schoener. Food webs from the small to the large: The robert h. macarthur award lecture. *Ecology*, 70(6):1559–1589, December 1989.

[175] John Scott. *Social Network Analysis: A Handbook*. Sage Publications, London, 2nd edition, 2000.

[176] Marco Scotti, Cristina Bondavalli, Giampaolo Rossetti, and Antonio Bodini. Flow network indices signal a directional change in ecosystems: Evidence from a small mountain lake (lake santo, northern italy). *Ecological Indicators*, 139:108896, 2022.

[177] M Ángeles Serrano and Marián Boguñá. *The Shortest Path to Network Geometry: A Practical Guide to Basic Models and Applications*. Cambridge University Press, 2021.

[178] M Ángeles Serrano, Dmitri Krioukov, and Marián Boguná. Self-similarity of complex networks and hidden metric spaces. *Physical review letters*, 100(7):078701, 2008.

[179] Somaye Sheykhali, Juan Fernández-Gracia, Anna Traveset, Maren Ziegler, Christian R. Voolstra, Carlos M. Duarte, and Víctor M. Eguíluz. Robustness to extinction and plasticity derived from mutualistic bipartite ecological networks. *Scientific Reports*, 10(1), June 2020.

[180] Man Shi, Jiangye Li, Qi Zhou, Guibin Wang, Weiguo Zhang, Zhenhua Zhang, Yan Gao, and Shaohua Yan. Interactions between elevated co2 levels and floating aquatic plants on the alteration of bacterial function in carbon assimilation and decomposition in eutrophic waters. *Water Research*, 171:115398, 2020.

[181] Jonathan B Shurin, Daniel S Gruner, and Helmut Hillebrand. All wet or dried up? real differences between aquatic and terrestrial food webs. *Proceedings of the Royal Society B: Biological Sciences*, 273(1582):1–9, 2006.

[182] Allison M Snider, Andrea Bonisoli-Alquati, Anna A Pérez-Umphrey, Philip C Stouffer, and Sabrina S Taylor. Metabarcoding of stomach contents and fecal samples provide similar insights about Seaside Sparrow diet. *Ornithological Applications*, 124(1):duab060, 12 2021.

[183] Tiziano Squartini and Diego Garlaschelli. Analytical maximum-likelihood method to detect patterns in real networks. *New Journal of Physics*, 13(8):083001, 2011.

[184] Tiziano Squartini and Diego Garlaschelli. *Maximum-entropy networks: Pattern detection, network reconstruction and graph combinatorics*. Springer, 2017.

[185] Juliette Stehlé, Nicolas Voirin, Alain Barrat, Ciro Cattuto, Vittoria Colizza, Lorenzo Isella, Corinne Régis, Jean-François Pinton, Nagham Khanafer, Wouter Van den Broeck, et al. Simulation of an seir infectious disease model on the dynamic contact network of conference attendees. *BMC medicine*, 9:1–15, 2011.

[186] Juliette Stehlé, Nicolas Voirin, Alain Barrat, Ciro Cattuto, Lorenzo Isella, Jean-François Pinton, Marco Quaggiotto, Wouter Van den Broeck, Corinne Régis, Bruno Lina, et al. High-resolution measurements of face-to-face contact patterns in a primary school. *PloS one*, 6(8):e23176, 2011.

[187] Daniel B Stouffer and Jordi Bascompte. Understanding food-web persistence from local to global scales. *Ecology letters*, 13(2):154–161, 2010.

[188] Daniel B Stouffer, Juan Camacho, Wenxin Jiang, and Luís A Nunes Amaral. Evidence for the existence of a robust pattern of prey selection in food webs. *Proc Biol Sci*, 274(1621):1931–1940, August 2007.

[189] Elisa Thébault and Colin Fontaine. Stability of ecological communities and the architecture of mutualistic and trophic networks. *Science*, 329(5993):853–856, 2010.

[190] Ross M Thompson, Ulrich Brose, Jennifer A Dunne, Robert O Hall, Sally Hladyz, Roger L Kitching, Neo D Martinez, Heidi Rantala, Tamara N Romanuk, Daniel B Stouffer, et al. Food webs: reconciling the structure and function of biodiversity. *Trends in ecology & evolution*, 27(12):689–697, 2012.

[191] Ross M. Thompson, Martin Hemberg, Brian M. Starzomski, and Jonathan B. Shurin. Trophic levels and trophic tangles: the prevalence of omnivory in real food webs. *Ecology*, 88(3):612–617, 2007.

[192] Davide Torre, Giuseppe F. Italiano, and Blerina Sinaimeri. A network approach to aquatic food web dynamics. In Martina Vettoretti, Erica Tavazzi, Enrico Longato, Giacomo Baruzzo, and Massimo Bellato, editors, *Computational Intelligence Methods for Bioinformatics and Biostatistics*, pages 1–15, Cham, 2025. Springer Nature Switzerland.

[193] Rowan Trebilco, Jess Melbourne-Thomas, and Andrew John Constable. The policy relevance of southern ocean food web structure: Implications of food web change for fisheries, conservation and carbon sequestration. *Marine Policy*, 115:103832, 2020.

[194] Robert E Ulanowicz. Information theory in ecology. *Computers & chemistry*, 25(4):393–399, 2001.

[195] Robert E Ulanowicz. Quantitative methods for ecological network analysis. *Computational biology and chemistry*, 28(5-6):321–339, 2004.

[196] Hadayet Ullah, Ivan Nagelkerken, Silvan U Goldenberg, and Damien A Fordham. Climate change could drive marine food web collapse through altered trophic flows and cyanobacterial proliferation. *PLoS biology*, 16(1):e2003446, 2018.

[197] Alfonso Valiente-Banuet, Marcelo A Aizen, Julio M Alcántara, Juan Arroyo, Andrea Cocucci, Mauro Galetti, María B García, Daniel García, José M Gómez, Pedro Jordano, et al. Beyond species loss: the extinction of ecological interactions in a changing world. *Functional Ecology*, 29(3):299–307, 2015.

[198] Jan E. Vermaat, Jennifer A. Dunne, and Alison J. Gilbert. Major dimensions in food-web structure properties. *Ecology*, 90(1):278–282, 2009.

[199] Shaopeng Wang and Ulrich Brose. Biodiversity and ecosystem functioning in food webs: the vertical diversity hypothesis. *Ecology Letters*, 21(1):9–20, October 2017.

[200] Wei Wang, Ming Tang, H Eugene Stanley, and Lidia A Braunstein. Unification of theoretical approaches for epidemic spreading on complex networks. *Reports on Progress in Physics*, 80(3):036603, feb 2017.

[201] Yang Wang, Deepayan Chakrabarti, Chenxi Wang, and Christos Faloutsos. Epidemic spreading in real networks: An eigenvalue viewpoint. In *22nd International Symposium on Reliable Distributed Systems, 2003. Proceedings.*, pages 25–34. IEEE, 2003.

[202] Yi Wang, Bhaskar Krishnamachari, and Thomas W. Valente. Findings from an empirical study of fine-grained human social contacts. In *2009 Sixth International Conference on Wireless On-Demand Network Systems and Services*, pages 153–160, 2009.

[203] Duncan J Watts and Steven H Strogatz. Collective dynamics of 'small-world'networks. *nature*, 393(6684):440–442, 1998.

[204] Earl E Werner and James F Gilliam. The ontogenetic niche and species interactions in size-structured populations. *Annual review of ecology and systematics*, 15:393–425, 1984.

[205] Lander Willem, Thang Van Hoang, Sebastian Funk, Pietro Coletti, Philippe Beutels, and Niel Hens. Socrates: an online tool leveraging a social contact data sharing initiative to assess mitigation strategies for covid-19. *BMC research notes*, 13(1):293, 2020.

[206] Richard J. Williams, Eric L. Berlow, Jennifer A. Dunne, Albert-László Barabási, and Neo D. Martinez. Two degrees of separation in complex food webs. *Proceedings of the National Academy of Sciences*, 99(20):12913–12916, September 2002.

[207] Richard J Williams and Neo D Martinez. Simple rules yield complex food webs. *Nature*, 404(6774):180–183, 2000.

[208] Richard J Williams and Neo D Martinez. Limits to trophic levels and omnivory in complex food webs: theory and data. *The American Naturalist*, 163(3):458–468, 2004.

[209] RJ Williams, GA Duff, DMJS Bowman, and GD Cook. Variation in the composition and structure of tropical savannas as a function of rainfall and soil texture along a large-scale climatic gradient in the northern territory, australia. *Journal of Biogeography*, 23(6):747–756, 1996.

[210] KL Wootton. Omnivory and stability in freshwater habitats: Does theory match reality? *Freshwater Biology*, 62(5):821–832, 2017.

[211] Boris Worm, Edward B Barbier, Nicola Beaumont, J Emmett Duffy, Carl Folke, Benjamin S Halpern, Jeremy BC Jackson, Heike K Lotze, Fiorenza Micheli, Stephen R Palumbi, et al. Impacts of biodiversity loss on ocean ecosystem services. *science*, 314(5800):787–790, 2006.

[212] Bo Yang, Hongbin Pei, Hechang Chen, Jiming Liu, and Shang Xia. Characterizing and discovering spatiotemporal social contact patterns for healthcare. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 39(8):1532–1546, 2017.

[213] Jun R Yang, Xiaoqing Yu, Huihuang Chen, Yi-Ming Kuo, and Jun Yang. Structural and functional variations of phytoplankton communities in the face of multiple disturbances. *Journal of Environmental Sciences*, 100:287–297, 2021.

[214] Jiang Zhang and Liangpeng Guo. Scaling behaviors of weighted food webs as energy transportation networks. *Journal of Theoretical Biology*, 264(3):760–770, 2010.

[215] Lizhi Zhang and Tiago P Peixoto. Statistical inference of assortative community structures. *Physical Review Research*, 2(4):043271, 2020.

[216] Xiao Zhang, Travis Martin, and Mark EJ Newman. Identification of core-periphery structure in networks. *Physical Review E*, 91(3):032803, 2015.

[217] Kun Zhao, Juliette Stehlé, Ginestra Bianconi, and Alain Barrat. Social network dynamics of face-to-face interactions. *Physical Review E—Statistical, Nonlinear, and Soft Matter Physics*, 83(5):056109, 2011.

[218] Xinhe Zhu, Bingbing Gao, Yongmin Zhong, Chengfan Gu, and Kup-Sze Choi. Extended kalman filter based on stochastic epidemiological model for covid-19 modeling. *Computers in Biology and Medicine*, 137:104810, 2021.

[219] Eric R Ziegel. *The elements of statistical learning*. Taylor & Francis, 2003.

[220] Konstantin Zuev, Marián Boguná, Ginestra Bianconi, and Dmitri Krioukov. Emergence of soft communities from geometric preferential attachment. *Scientific reports*, 5(1):9421, 2015.