

Explaining Species Endangerment Through Network Analysis

Nathan Hunsberger, Jeffrey Pan, Jovian Wang

Abstract

Objective

The goal of this paper is to examine how coral bleaching affects marine ecosystems off the coast of Jamaica and Cuba by applying network analysis tools on a food chain network in the Caribbean region. We will attempt to understand how network features correlate to endangerment levels of species in these trophic regions.

Materials and Methods

Using guild data obtained from REEF, GBIF, and Fishbase databases, we perform network analysis on top of this data through NetworkX to understand guild-level relationships and characteristics of the ecosystem. We then perform feature selection on top of these network characteristics to predict guild endangerment levels obtained from the IUCN RedList.

Results

Our feature selection tests returned results contrary to our hypothesis for guild out-degree, guild in-degree, and guild degree difference, as guild out-degree and guild degree difference positively correlated to guild endangerment. Additionally, the shortest path to a producer positively correlated to guild endangerment, demonstrating an increasing reliance on producers amongst the ecosystem.

Conclusion

Our findings suggest that the coral reef ecosystems in these trophic regions are at an imbalance, allowing coral bleaching to have a devastating impact on species in the area. Our results have some limitations, such as data timelines not aligning perfectly as well as some error caused by data structure or species grouping.

Background

Over the past decade, global warming has drastically affected coral reef ecosystems around the world. With rising temperatures, reefs have gone through the process of “coral bleaching”, where reefs expel the algae (zooxanthellae) inside their tissues. This phenomenon occurs because the zooxanthellae produce oxygen-based compounds, which are highly toxic to coral. Without the algae, the coral is more susceptible to diseases and other ailments, leading to a

major decline in the coral population. In fact, coral cover, the size of coral reefs, has decreased by 50.7% since 1985, with an average of 3.38% decrease per year from coral bleaching (De'ath, G. et al. 2012). This decrease in coral cover has led to instability in the population of algae, fish, and other organisms that reside in this marine ecosystem.

Introduction

In this paper, we will examine how coral bleaching affects marine ecosystems off the coast of Jamaica and Cuba. To accomplish this task, we first apply network analysis tools on a food chain network in the Caribbean region. We then analyze the network characteristics obtained from the tools, attempting to explain and understand the results. Lastly, we correlate these network characteristics with endangerment levels to better understand ecological interactions.

Through these methods, we hope to understand how network metrics can be applied to a given ecosystem or ecological network. Furthermore, we want to investigate how climate events, such as coral bleaching, affect ecosystems and interspecies interactions. We also aim to understand the resistance of the coral reef communities to both primary and secondary extinctions. Lastly, we want to examine ecological stability, an ecosystem's ability to recover from a catastrophic event.

Data Materials and Study Design

We use food chain data that focuses on the Greater Antillean Coral Reef Systems, which consist of the areas around the Cayman Islands, Cuba, and Jamaica. Specifically, we focus on coral reef communities off the coast of Cuba and Jamaica. This dataset, which was detailed and explained in Roopnarine, P. D., & Hertog, R. (2013), was constructed from REEF, GBIF, and Fishbase databases. This dataset describes relationships between guilds or groups of organisms that behave similarly, as shown in Table 1. Below are some measures used throughout the dataset:

- Guild Number - standard number among Jamaican and Cuban datasets which identifies a specific guild.

- Guild Description - Describes the common name of the species or the niche of the guild in the case of multispecies guilds. For example, most algae are grouped into one guild because they play a similar role as a producer.
- Foraging - The major foraging habitats for guild members: “r” for reef foragers, “g” for seagrass foragers, “rg” for reef-seagrass foragers, or “.” for none.
- Number of Prey - Number of other guilds that this guild feeds on.
- Prey - A list of guilds that this guild feeds on.

Guild Number	Guild Description	Foraging Habitat	Number of Prey	Prey
12	Endolithic sponges	.	30	1, 2
13	Ahermatypic benthic corals□	.	35	17, 18, 19, 20, 22, 23, 24
14	Hermatypic corals□	.	36	17, 18, 19, 20, 22, 23, 24
15	Micro-detritivores	.	1	0
16	Corallivorous polychaetes	.	76	14,31,32
17	Herbivorous micro-zooplankton	.	29	2

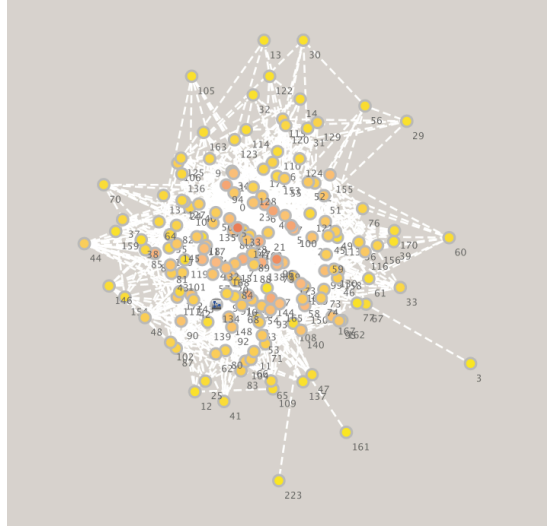
[Table 1: Sample of Guild Data](#)

Using this data, we can construct a weighted, directed food chain network, with nodes representing guilds and edges representing predator/prey relationships. Each edge weight represents the number of prey consumed in each interaction. Subnetworks may be highlighted to distinguish between producer and consumer guilds.

We also use the IUCN Red List API to determine the endangerment levels of the guilds involved in this study (The IUCN Red List). The IUCN Red List is largely recognized as the most authoritative guide to the status of biological diversity and endangerment. The database defines endangerment level per species by category. The classifications are: Not Evaluated (NE), Data Deficient (DD), Least Concern (LC), Near Threatened (NT), Vulnerable (VU), Endangered (EN), Critically Endangered (CR), Extinct in the Wild (EW), or Extinct (EX). The database also provides a similar categorical general population trend, which is either “Stable”, “Increasing”, “Decreasing”, or “Unknown”. Queries are pulled from the Red List API, and sent using RRedlist, an R client for the Red List.

Methods

Using NetworkX and Pandas, we constructed a weighted, directed graph. This graph is visualized using Cytoscape, and resembles a relatively connected network of guilds:



[Figure 1: Directed, Weighted Graph of Guild Relationships](#)

This network contains 247 nodes, with 4,042 total connections between nodes. We use a heat map to represent node colors, where nodes with higher out-degrees are represented by dark red while lower out-degrees nodes are represented by light yellow. Figure 1 shows the rough distribution of the network, with a few highly connected nodes and a relatively large number of nodes with smaller degrees.

Next, we create another dataset by linking our current network to endangerment information obtained from the Red List. We match all species in each guild to its corresponding information in the Red List. However, some species that are not on the Red List, such as singular cellular organisms in guilds 1 through 9, as well as DD species that have no endangerment assessment, such as *Agaricia fragilis*, are dropped from the dataset. The result is a dataset of 213 species belonging to 144 different guilds. Each species was then assigned a numerical endangerment score as follows. LC species have a score starting at 1, NT species have a score starting at 2, VU species have a score starting at 3, EN species have a score starting at 4, and CR species have a score starting at 5. Next, we account for current population trends. Species with a “Decreasing” population have an additional 0.5 added to their scores, while species with a “Stable” or “Increasing” population have their scores decrease by 0.5. Species with “Unknown” population trends retain their starting score. The resulting score effectively rates the risk of endangerment of each species for the foreseeable future. It should be noted that no species involved in the study are extinct in any capacity, which makes intuitive sense as they must have a current role in the food network.

Finally, scores are pooled by guild. Since few involved species are critically endangered, endangerment scores were maxed per guild to determine overall guild score. For example, if a guild had two species, one with endangerment level 3.0 and one with endangerment level 4.5, a score of 4.5 was assigned to that guild. The average guild endangerment level of our network is 1.45, with a minimum of 0.5 and a maximum of 5.5.

Lastly, we perform a feature selection of network characteristics to examine which network features best correlate to guild endangerment level. To perform this analysis, we bring together network characteristics on the 144 guilds which the IUCN Red List was able to score. We use Scikit Learn, a Python package, to perform an r correlation coefficient test on guild endangerment level using the following network features: guild degree, guild in degree, guild out degree, difference between guild out degree and guild in degree (guild degree difference), guild betweenness centrality, guild clustering coefficient, shortest path to producers per guild, and number of species preyed upon per guild. We are particularly interested in the R^2 coefficient per feature, as this determines the variation in guild endangerment level explained by that specific feature.

Results and Discussion

Scale Free Test

When working with real world data, it is important to test whether or not the network is scale-free (Patterns – the art, soul, and science of beholding nature). A scale-free network means the fraction of nodes with degree k follows a power-law distribution $k^{-\alpha}$, where $\alpha > 1$. In other words, a scale-free network means the vast majority of nodes have relatively few connections, while a select few important nodes, known as hubs, contribute heavily to total connectivity. Another characteristic of scale-free networks is known as dimension, where no matter the size or magnification of a network, the average path length from any node to another remains relatively unchanged. Most nodes should be able to reach each other by linking through hubs. For our guild network, being scale-free would imply some guilds (the hub guilds) are central to the food network and vital for its equilibrium, as a large number of guilds are connected through them. This can be viewed as dangerous to ecosystem resilience, as it brings about a dependence upon hub guilds and highly increases the likelihood of secondary extinctions. For example, if a hub central to the food chain is removed (i.e. a secondary

consumer), then it would affect any predators of this hub, leading to possible extinctions. It would also increase the population of any of its prey, leading to overpopulation and possibly population instability. Both of these consequences are a threat to the ecological network and to ecosystem resilience.

To determine if the network is scale-free, we ran a Kolmogorov-Smirnov test on the network's degree distribution with the null hypothesis being that the degree distribution resembles a power-law distribution (GeeksForGeeks, 2022). We determined a p-value of $7.561534611548206e-211$, which is significantly smaller than an alpha of 0.05, thus we are justified in rejecting the null hypothesis and reasoning that the node degree distribution does not follow a powerlaw distribution. Thus, the network is not scale-free.

Because the guild network is not scale-free, we reason that node degree is not centralized at a few hub guilds, thus the ecosystem does not have a select few guilds that are more important to the food chain than others. This demonstrates the idea of ecosystem resilience, the ability of an ecosystem to respond to disturbances relatively quickly in order to maintain its normal patterns of nutrient cycling and biomass production (Sasaki et al., 2015). In our case, the food network is not left divided if a few hub guilds go extinct/migrate out of the area.

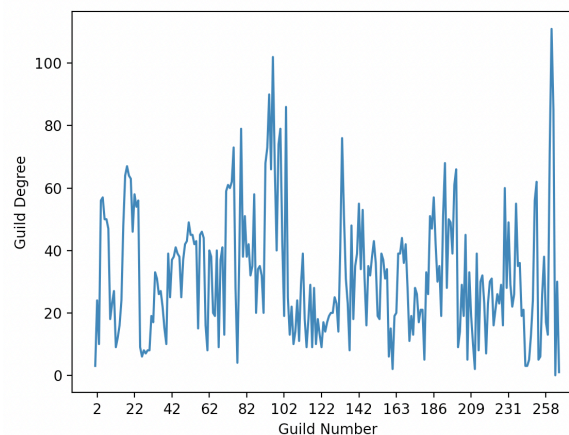
Network Metrics

To further understand any ecological relationships, we examine the network characteristics of betweenness, degree, homophily, and clustering coefficient, which were generated by NetworkX. We also examine the relationships between consumers and producers by finding the average shortest path to a producer node.

First, we will look at betweenness centrality, which is a measure of how interconnected a network is. In our network of a food chain, a higher betweenness index means that there would be more consequences or destruction if a certain guild were to disappear. Using our measure from NetworkX, we determine that there are no guilds with a large betweenness, with the largest betweenness score being guild 96 (omnivorous crustacea VI) with a score of 0.0157. This again demonstrates the idea of ecosystem resilience. Ecosystem resilience usually occurs in more complex or diverse ecosystems because there are more alternative resources, preventing a snowball effect of secondary extinctions. In the model created, because no guild controls the

interactions of a large proportion of other guilds, the extinction of one guild, such as guild 96, would not be expected to cause a large proportion of other extinctions.

Degrees are a measure of how many directed edges point to a node (in-degree) or how many edges originate from a node (out-degree). For our network structure, node in-degree represents how many guilds prey on this guild, and out-degree represents how many nodes this guild preys on. Separately, these measures can show us the top and bottom of community food chains, but together, the total degree can show us which guilds are most active in community food chains. We can use this to reveal trends between communities. The nodes with the largest total degree, and thus the most active in their communities, are guild 261 (the Caribbean Reef Shark with degree 111) and guild 96 (omnivorous crustacea VI with degree 102). The overall average node degree is 32.5. However, there is a standard deviation of 20.1, showing a larger number of outliers present as some guilds are more active than others. We hypothesized that this might be due to higher numbered guilds having a higher degree as they live in larger communities, but a graph of degree vs. guild number disproves this:

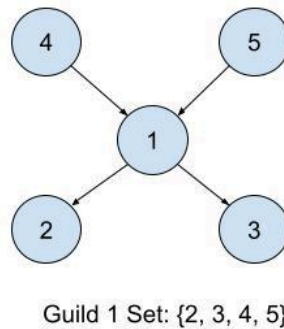


[Figure 2: Node Degree vs. Guild Number](#)

As shown in Figure 2, guild degree is not necessarily higher or lower depending on guild number.

We also use the homophily index to better understand network assortativity. The homophily index is a measure of a guild's predator/prey similarity. The implications are that a high homophily index will represent stronger ecosystem resilience, as the extinction of one guild won't have damaging effects on other guilds. For two connected guilds, we can find a homophily

score by using the Jaccard Similarity Index. The Jaccard Similarity Index relies on set comparisons. The similarity index between two sets A and B is defined as $|A \cap B| / |A \cup B|$. For the guild network, each guild has a set containing all the guilds this guild interacts with (all guilds this guild preys upon or is preyed upon by). This is illustrated in Figure 3.



[Figure 3: Example Jaccard Similarity Set](#)

For example, if guild 1 preys upon guilds 2 and 3, but is preyed upon by guilds 4 and 5, guild 1's set contains guilds 2, 3, 4, and 5. We chose Jaccard Similarity as a good measure of homophily because similar guilds should interact similarly within the ecosystem, thus the effectiveness of Jaccard Similarity at capturing Guild interactivity makes it useful in this situation.

For this network, we found an average Jaccard Similarity score of 0.065 per edge. Additionally, only 18% of edges between guilds had a Jaccard Similarity score above 0.10. These results imply low homophily across the network, an indication that the removal of some guilds from the network would have few negative effects on the ecosystem. A reason the average similarity scores could be so low is that guilds cover a range of species, not just one specific species, thus similar species will already be grouped together into guilds. This implies that interactions within the network across guilds are likely not similar, thus Jaccard Similarity scores should be low for this network in general.

Finally, we look at the clustering coefficient, a measure of the degree in which nodes cluster into tightly knit components with high densities of connections. We will use the global clustering coefficient, which is an overall indication of clustering in the entire network. This coefficient is calculated by taking the total amount of close triplets, or connected node triangles and dividing it among all possible triplets in the network. From our Jamaican guild network, we found a global clustering coefficient of 0.072. This score is surprisingly low considering the

nature of our ecological network, as we expect inter-species interactions. However, this is most likely due to our grouping into guilds, where similar species are put into one node. This would decrease the clustering coefficient because our network eliminates these interspecies interactions.

Additionally, we were interested in the relationship between consumers and producers, specifically how dependent consumers are on producers. To measure this, we examined the average shortest path length between each guild to a producer guild, as shorter path lengths between consumers and producers demonstrates a large role dependence on producers, and subsequently higher ecosystem instability to reduced coral reef abundance. The average shortest path between guilds and producers is a length of 1.80 with a standard deviation of 0.925. Thus, most guilds are at most 2-3 edges from producers in the ecosystem, and will quickly feel the effects of reduced coral reef abundance.

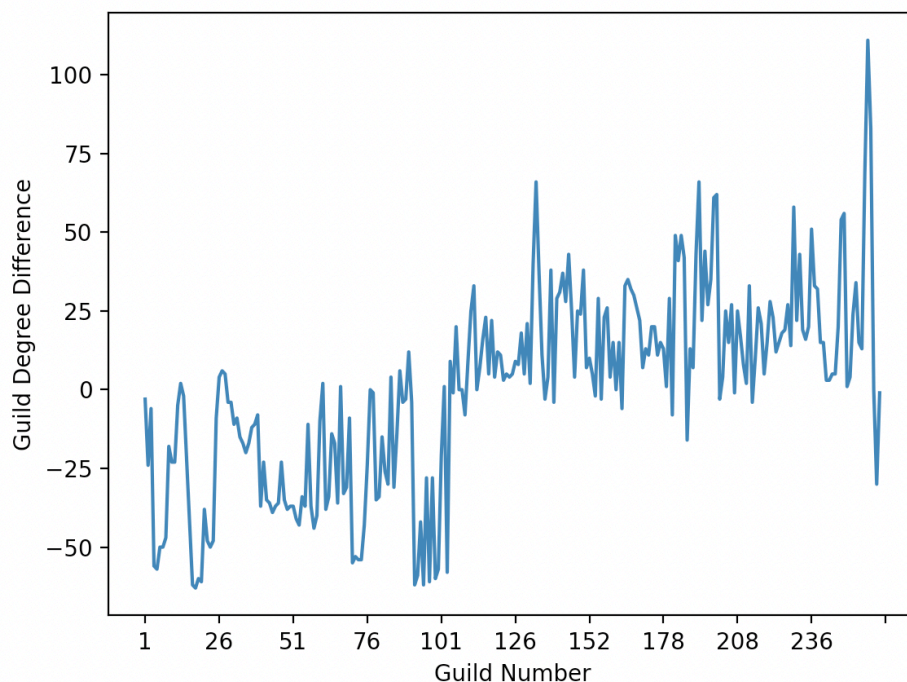
Correlation with Species Endangerment

The results of the r correlation coefficient test found three network features that best correlate with guild endangerment level: node out degree (with an r^2 coefficient of 0.25), guild degree difference (with an r^2 coefficient of 0.28), and shortest path to producers (with an r^2 coefficient of 0.24). This means that roughly 28% of the variation in guild endangerment level is explained by that guild's degree difference. While these correlation coefficients are relatively small, we attribute this to a few factors. First, there are only 144 guilds used to calculate r^2 , thus there is a lack of data available to get an accurate explanation of guild endangerment level. Second, guild endangerment level can be affected by numerous factors outside the guild's ecosystem interaction habits, such as water temperature or environmental effects (coral bleaching). Our r^2 test is specific to guild endangerment level brought about by ecosystem interaction. Thus, lower r^2 coefficients are expected but still useful in explaining the relationship between ecosystem interaction and guild endangerment.

Our r^2 test has also led to some results that contradict our hypothesis. Recall that for our network, guild out degree tells us how many guilds this guild preys upon, and guild in degree tells us how many guilds this guild is preyed upon by. Thus, it is natural to expect guild out degree to negatively correlate with guild endangerment level, as guilds that prey upon more guilds should be less exposed to danger in the ecosystem. Additionally, we expected guild in degree to positively correlate with guild endangerment level, as guilds that are preyed upon by

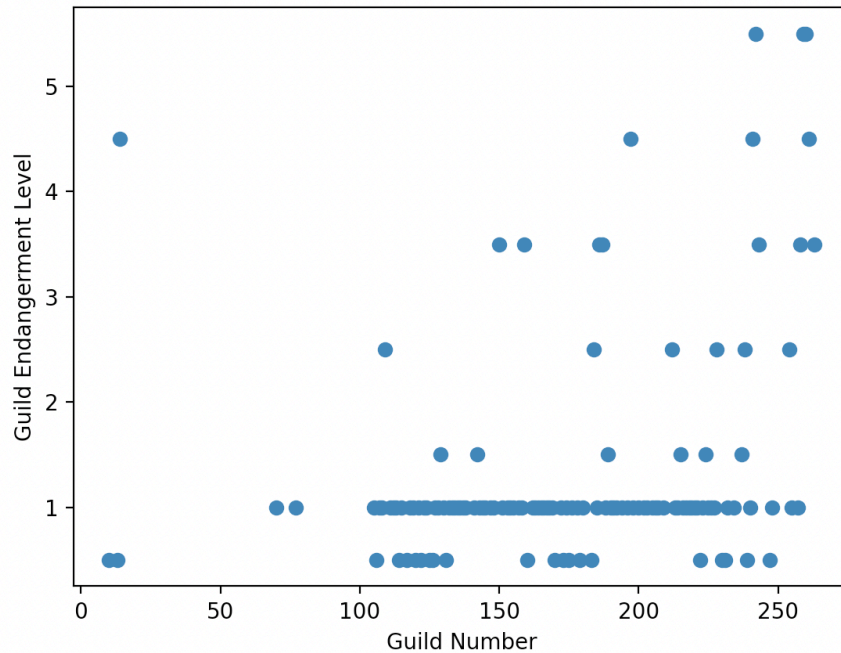
more guilds should be more exposed to danger in the ecosystem. Contrary to this hypothesis, we found an r^2 coefficient for guild out degree of .25, and r^2 coefficient for guild in degree of -.22 (we kept the negative after squaring to demonstrate negative correlation). One explanation we came up with was that guild degree values are naturally larger for some guilds, so simply comparing total out degree/in degree is not a sufficient measure, rather we should compare the difference in guild out degree and guild in degree (guild degree difference). This difference tells us how many more guilds this guild preys upon than guilds that prey upon this guild. For example, if a guild preys upon 10 guilds, and is only preyed upon by 5 guilds, the difference in guild out degree and guild in degree would be 5. We expected guild degree difference to negatively correlate with guild endangerment level, as guilds that prey upon more guilds than they are preyed upon by should not be in any danger in terms of ecosystem interaction. However, contrary to our belief, guild degree difference positively correlates with guild endangerment level, with an r^2 coefficient of 0.28.

We now try to understand why guild degree difference positively correlates with guild endangerment level. To understand this, let's examine how guild degree difference and guild endangerment level are affected by guild number, as guild number generally increases as average fish size per guild increases, which could be affecting our results.



[Figure 4: Guild Degree Difference vs. Guild Number](#)

As can be seen in Figure 4, guild degree difference increases with guild number, implying that higher numbered guilds are “on top” of the food chain, as they prey upon more guilds than they are preyed upon by. This is expected, and caused by the nature of the dataset being used. Now let’s examine the effect guild number has on guild endangerment level.



[Figure 5: Guild Endangerment Level vs. Guild Number](#)

First, note the lack of data points under 100 guilds. This is due to the fact that the IUCN Red List does not have data on any species in these guilds, which could have a small effect on our results. Additionally, note the spike on the far left side. This increased endangerment level is due to coral bleaching, something which affects all members of the ecosystem as the average shortest path between guilds and producers is 1.80 nodes (calculated earlier). We can then look at the results of the scatter plot in Figure 5, in which we notice a trend of increasing guild endangerment level as guild number increases. This implies that, per the IUCN Red List, guilds “on top” of the food chain in these trophic regions are also the ones in most danger to extinction. This would explain the positive correlation between guild degree distribution and guild endangerment level, as both metrics are affected by our dataset having guilds with bigger fish be numbered higher. However, it brings about the notion that guilds “on top” of the food chain are the ones in most danger, explaining why guild degree difference positively correlates with guild endangerment level. This means that trophic interactions in these regions may not be at

equilibrium. Just because one guild preys upon more guilds than another, it could still be facing extinction, something we would not expect in a healthy ecosystem.

The most significant result from our feature selection test is the positive correlation between the shortest path from one guild to a producer guild (guild shortest path) and guild endangerment level, with an r^2 correlation coefficient of 0.24. The longer a guild's shortest path is, the more likely the species in it will face extinction. This is as expected because these guilds are not preying upon producers like coral or algae. Rather, they are preying upon other fish species which is more difficult to do, leading to an increase in their guild endangerment level. This could explain the observation noted above that guilds "on top" of the food chain are the ones in most danger, but it also demonstrates a danger to the ecosystem as a whole as coral abundance decreases. These higher numbered guilds with larger guild shortest path metrics are already more likely to be in danger of extinction. As coral bleaching continues, the guilds that directly prey upon producers (guilds with guild shortest path less than 1) will have to move in order to find new food sources. This will have a disastrous impact on higher numbered guilds, as their food sources will have dispersed. Higher numbered guilds being already exposed to large guild endangerment levels means they will be affected the most by coral bleaching and the changing trophic interactions.

Conclusion

Given our findings, we conclude that there is definitive evidence of a risk posed to the species in the coral reef ecosystem due to the presence of a relationship between distance from producers and guild endangerment. Additionally, the relationship between guild degree difference and guild endangerment represents an imbalance in the ecosystem, one which makes rising coral bleaching an imminent threat to species survival.

This study has some limitations, however, as the data is organized in guilds which can have an effect on network structure and real-world ecosystem interactions. Additionally, guild data is concentrated on 2013 data while our Redlist endangerment level data ranges from 2004 to 2020. This could have a slight impact on our results as real-world endangerment trends could have been different at the time of guild data recording.

References

- De'ath, G., Fabricius, K. E., Sweatman, H., & Puotinen, M. (2012). The 27-year decline of coral cover on the Great Barrier Reef and its causes. *Proceedings of the National Academy of Sciences*, 109(44), 17995–17999. <https://doi.org/10.1073/pnas.1208909109>
- GeeksforGeeks. (2022, January 17). *ML | Kolmogorov-Smirnov Test*.
<https://www.geeksforgeeks.org/ml-kolmogorov-smirnov-test/>
- Graham, B. (n.d.). *Patterns – the art, soul, and science of beholding nature*.
- Hevey, D. (2018). Network analysis: a brief overview and tutorial. *Health Psychology and Behavioral Medicine*, 6(1), 301–328. <https://doi.org/10.1080/21642850.2018.1521283>
- Jaccard Similarity. (n.d.). <https://www.learndatasci.com/glossary/jaccard-similarity/>
- Liben-Nowell, D., & Kleinberg, J.M. (2007). The link-prediction problem for social networks. *J. Assoc. Inf. Sci. Technol.*, 58, 1019-1031.
- Roopnarine, P. D., & Hertog, R. (2013). Detailed Food Web Networks of Three Greater Antillean Coral Reef Systems: The Cayman Islands, Cuba, and Jamaica. *Dataset Papers in Ecology*, 2013, 1–9. <https://doi.org/10.7167/2013/857470>
- Sasaki, T., Furukawa, T., Iwasaki, Y., Seto, M., & Mori, A. S. (2015). Perspectives for ecosystem management based on ecosystem resilience and ecological thresholds against multiple and stochastic disturbances. *Ecological Indicators*, 57, 395–408.
<https://doi.org/10.1016/j.ecolind.2015.05.019>
- sklearn.feature_selection.r_regression*. (n.d.). Scikit-learn.
https://scikit-learn.org/stable/modules/generated/sklearn.feature_selection.r_regression.html
- The IUCN Red List*. (n.d.). IUCN Red List of Threatened Species. <https://www.iucnredlist.org/>

