

PERSPECTIVE

# Telling ecological networks apart by their structure: A computational challenge

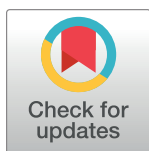
Matthew J. Michalska-Smith<sup>1‡\*</sup>, Stefano Allesina<sup>1,2</sup>

**1** Department of Ecology & Evolution, University of Chicago, Chicago, Illinois, United States of America,

**2** Northwestern Institute on Complex Systems, Northwestern University, Evanston, Illinois, United States of America

‡ Current address: Department of Veterinary Population Medicine, University of Minnesota, St. Paul, Minnesota, United States of America

\* [mjsmith@uchicago.edu](mailto:mjsmith@uchicago.edu)



## Abstract

Ecologists have been compiling ecological networks for over a century, detailing the interactions between species in a variety of ecosystems. To this end, they have built networks for mutualistic (e.g., pollination, seed dispersal) as well as antagonistic (e.g., herbivory, parasitism) interactions. The type of interaction being represented is believed to be reflected in the structure of the network, which would differ substantially between mutualistic and antagonistic networks. Here, we put this notion to the test by attempting to determine the type of interaction represented in a network based solely on its structure. We find that, although it is easy to separate different kinds of nonecological networks, ecological networks display much structural variation, making it difficult to distinguish between mutualistic and antagonistic interactions. We therefore frame the problem as a challenge for the community of scientists interested in computational biology and machine learning. We discuss the features a good solution to this problem should possess and the obstacles that need to be overcome to achieve this goal.

## OPEN ACCESS

**Citation:** Michalska-Smith MJ, Allesina S (2019) Telling ecological networks apart by their structure: A computational challenge. PLoS Comput Biol 15(6): e1007076. <https://doi.org/10.1371/journal.pcbi.1007076>

**Editor:** Tobias Bollenbach, Universität zu Köln, GERMANY

**Published:** June 27, 2019

**Copyright:** © 2019 Michalska-Smith, Allesina. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Funding:** MJM-S received funding from U.S. Department of Education (<https://www.ed.gov/>) grant P200A150101. SA and MJM-S received funding from the National Science Foundation (<https://www.nsf.gov/>) grant DEB-1148867 and the France Chicago Center's "France And Chicago Collaborating in The Sciences" program (<https://fcc.uchicago.edu/faccts>). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

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

## Author summary

In the late 1960s, Mark Kac asked, "Can one hear the shape of a drum?" challenging readers to reconstruct the geometry of the drum from a provided list of overtones. Physicists and mathematicians found that, although in general, one cannot hear the shape of a drum, many of its properties, such as its area and perimeter, can be "heard". Here, we ask whether the type of interaction being represented in a network can be "seen" when inspecting its shape—for example, whether we can distinguish networks reporting interactions between plants and their pollinators (mutually beneficial) from those representing interactions between plants and their herbivores (beneficial only to the herbivores). We show that many types of nonbiological networks can be easily separated based on structural properties, whereas determining the type of interaction represented by an ecological network is harder. We therefore turn this problem into a challenge for the scientific

community. We argue that solving this problem would greatly benefit the field, and we discuss what the consequences of a failure might be.

## Introduction

Since the early days of the field, ecologists found themselves detailing the multitude of interactions occurring between different populations, such as predation, parasitism, and herbivory (all "antagonistic" interactions) or pollination, seed dispersal, and symbiosis ("mutualistic" interactions) [1, 2]. To make sense of these data, they built networks in which nodes are species and edges stand for interactions between species. Depending on the interaction being depicted by the edges, we speak of food webs (the edge  $i \rightarrow j$  marks consumption of species  $i$  by  $j$ ), host–parasite networks ( $j$  parasitizes  $i$ ), pollination networks (animal  $j$  pollinates plant  $i$ ), herbivory networks (animal  $j$  feeds on plant  $i$ ), etc. Whereas the earliest published food web dates back more than a century [3], the past 20 years have seen a sharp increase in both the number of networks published and their quality (i.e., higher level of detail, larger number of species reported, and larger number of interactions, often including weighted edges).

Reducing a multidimensional object such as a network to a few numbers is a daunting task, but paralleling the progress of network analysis in other branches of science, ecologists set out to compute summary statistics on the empirical networks they collected; these metrics range from very simple measures, such as the size (number of nodes) and connectance (proportion of realized connections) of the network, to large-scale properties such as modularity (Are networks organized in blocks of dense connectivity loosely connected by few edges? [4]) and nestedness (Can interactions be organized as in a Russian doll—with specialist species choosing interactions among those of generalist species? [5–7]). As in other branches of science, ecologists also investigated degree distributions, motif profiles [8], k-cores [9], and many other network properties.

Here, we ask whether these metrics can be used to characterize the type of interaction being depicted by a network under the hypothesis that the type of interaction would affect network structure in a consistent and detectable way. Take a mutualistic network: both parties benefit from interacting, and they actively seek out this relationship from an evolutionary point of view—for example, plants reward pollinators with nectar, attract them with visual and olfactory cues, etc. Contrast this with herbivory, in which plants are desperate to avoid the interaction and display evolutionary strategies in the form of spines and thorns, toxic compounds, bark, etc. Given the contrasting forces at play, we might hypothesize that these processes give rise to networks with different structure [10, 11].

Previous analysis posited that mutualistic and antagonistic networks' properties would differ significantly (for example, mutualistic networks are believed to be "more nested" than antagonistic ones) [10–14]. However, this does not necessarily imply that network types can be separated based on these properties (even if the means of two distributions differ significantly, the overlap could be large enough to render a classification impossible). We frame the problem of detecting the type of interaction being represented in an ecological network by measuring its structure as a challenge to the community of scientists interested in networks, computational biology, and machine learning. We provide a database of networks of unprecedented size and completeness that can be used to test the quality of a solution, and we outline the principal hurdles that need to be overcome in order to solve this problem.

Before examining the details of the challenge, it is important to justify its timeliness. Classifying the type of an ecological network might seem like a pointless exercise: the ecologists that put it together knew perfectly well whether they were recording parasitism or pollination. However, solving this challenge is important for at least three reasons. First, showing that certain properties can be used to distinguish among networks is a post hoc justification for developing these metrics in the first place, thereby proving that they truly measure salient, relevant aspects of network structure. Second, a solution would provide a good justification for the representation of these data as networks—by compiling networks, we can infer properties of the system that would be difficult to assess otherwise. Third, with the advent of cheap, high-throughput molecular techniques, associations between species can be sampled at an unprecedented level of detail—for example, one can sequence the root of a tree to find all its fungal inhabitants, sequence insects to detect their endoparasites, etc. In this way, one could rapidly and reliably build networks of interactions between species—but what type of interactions would these species form? Some of the arbuscular mycorrhizal fungi sampled will be symbionts of the plant, whereas others will be parasites; in fact, the same species could act as either, depending on the ecological and environmental context. **If we could find universal indicators of mutualistic versus antagonistic interactions, then we could make an educated guess on the type of interaction represented in these networks by just measuring a few of their properties.**

Note that, to keep the task as simple as possible, here we ask the broad question, What type of interaction is represented in this network? However, the same question could be asked of each single edge, as ecological networks are typically composed of a variety of interactions [27, 30].

To illustrate the details of the computational challenge with a simple, solvable example, we draw upon a large data set of nonecological networks we have compiled for this purpose. We show that it is quite easy to separate these classes of networks, even using elementary methods. **We then turn to ecological networks, showing that the task is much more difficult: not only can we not separate cleanly mutualistic from antagonistic ecological networks, but we cannot even separate ecological networks from the remaining nonecological networks.**

## Materials and methods

**We** built undirected, unweighted bipartite networks recording the connections between actors and movies (“actor collaboration”), authors of scientific articles and the journals in which they have published (“authorship”), lawmakers votes on laws (“legislature”), microbial organisms and parts of the body where they are found (“microbiome”), and city neighborhoods and crime occurrences (“crimes”). These systems were chosen based on two characteristics: (1) for each category, multiple (100 or more) networks could be built using the same methodology and (2) the networks within each category would display a large range of sizes and levels of connectivity.

**We** also amassed a large database of more than 500 bipartite ecological networks, representing either antagonistic (host–parasite, host–parasitoid, bacteria–phage, plant–herbivore) or mutualistic (plant–pollinator, plant–seed disperser, ant–plant, anemone–fish) interactions.

The summary statistics on the size and connectivity for each class of networks are reported in Table 1; the details of how the networks were constructed and the references for the published networks are provided in the Supporting information.

## Choice of metrics

Given the great interest in network analysis, there are hundreds of network metrics one could use to attempt to characterize the different classes of networks. When it comes to ecological networks, however, two of their most accessible properties, the size (number of nodes) and

**Table 1. Summary statistics on the size and fill of collected networks.**

Type	Metric	Minimum	Maximum	Mean	Median
Actor Collaboration (155)	Connectance	0.0002	0.072	0.005	0.0026
	Links	258	78,145	12,658.652	6,753
	Rows	15	6,675	1,116.394	568
	Columns	239	43,872	7,808.723	4,832
Antagonism (197)	Connectance	0.0078	0.9301	0.27	0.2361
	Links	14	5,203	165.472	82
	Rows	5	749	38.67	18
	Columns	5	888	34.274	21
Authorship (109)	Connectance	0.0303	0.0588	0.04	0.0388
	Links	2,186	16,611	9,479.532	9,939
	Rows	86	712	389.349	396
	Columns	438	681	623.138	639
Crimes (1,816)	Connectance	0.0457	0.6053	0.25	0.2406
	Links	12	470	153.214	120
	Rows	7	77	39.834	33
	Columns	5	64	22.273	21
Legislature (245)	Connectance	0.1715	0.8455	0.593	0.6056
	Links	277	367,956	19,162.89	7,092
	Rows	16	972	167.898	138
	Columns	10	1,427	151.698	87
Microbiome (203)	Connectance	0.0621	0.1891	0.103	0.0982
	Links	7,585	46,936	19,396.808	18,500
	Rows	4,559	16,163	8,625.236	8,442
	Columns	9	44	22.616	23
Mutualism (360)	Connectance	0.0173	0.6875	0.202	0.1907
	Links	12	15,255	183.733	58
	Rows	5	1,044	47.244	27
	Columns	5	647	27.294	15

For references and a detailed description, see Supporting information.

<https://doi.org/10.1371/journal.pcbi.1007076.t001>

connectance (proportion of realized links), should not be used for this purpose. In fact, an ecological network's size and connectance mostly reflect the experimental design and sampling effort rather than structural differences from other networks. These networks typically integrate information through space (e.g., Where does my ecosystem "end"?) and time (Shall we build a network for each day of sampling? Week? Year?)—sampling more extensively in time and space will surely result in a larger number of species and connections.

Given these limitations, one should choose metrics that either are not influenced by size and connectance or can be rescaled to remove their effects. For the illustrative example we use to detail the computational challenge, we are going to follow the latter approach, though the former would be superior, if we were to find network properties that are insensitive to size and connectivity.

For each network, we measure the two largest eigenvalues of the adjacency matrix, which are associated with important network properties: it is well known that the first eigenvalue (spectral radius) of the adjacency matrix is maximized in perfectly nested networks [15], whereas the second eigenvalue will separate from the bulk of the spectrum in strongly modular networks [16]. Because both eigenvalues are expected to grow whenever we add nodes and

**Table 2. Metrics used in a PCA.**

Metric	Formula	PC1	PC2	PC3
Configuration model $\lambda_1$ relative error	$1 - \lambda_1^{cm} / \lambda_1$	0.578	-0.576	0.578
Erdős-Rényi $\lambda_1$ relative error	$1 - \lambda_1^{er} / \lambda_1$	0.646	-0.111	-0.755
Marchenko-Pastur $\lambda_2$ relative error	$1 - \lambda_2^{mp} / \lambda_2$	0.499	0.810	0.308
<b>Percent Explained Variance:</b>		<b>53.944</b>	<b>27.659</b>	<b>18.397</b>

PCA run on the correlation of 500 nonecological bipartite networks with each of the first three principal component vectors. Here,  $\lambda$  is an eigenvalue of the adjacency matrix of the network, and the subscript indicates its position when sorted in descending order; the superscript indicates that the value is an approximation of the eigenvalue under a given randomization protocol (Supporting information). Abbreviations: PC1, principal component 1; PC2, principal component 2; PC3, principal component 3; PCA, principal component analysis

<https://doi.org/10.1371/journal.pcbi.1007076.t002>

connections, we normalize them by computing their expectation under two null models: an Erdős-Rényi, random bipartite graph (in which the number of nodes and connections are preserved, but nodes are connected at random, [17]), and a configuration model (in which the nodes' degrees are preserved, but again, wiring is random [18, 19]). By computing the relative error we make when using these expectations rather than the observed values, we attempt to remove the trivial effect of size and connectance (Table 2).

## Mapping networks

Several methods of classifying objects have been developed in the literature on statistics and machine learning, and picking a single method from such an embarrassment of riches is hard. To keep the illustration of the challenge as simple as possible, here we measure the three spectral quantities introduced previously, thereby representing each network as a single point in a three-dimensional space. To better visualize the position of the networks in this space, we perform a principal component analysis (PCA), projecting the network positions on the plane defined by the first two components. Metaphorically, we are mapping out the "network space" such that, if we were to have chosen meaningful quantities, two networks belonging to the same class should be close, and the various network classes should form clusters that are well separated.

The use of a PCA facilitates the illustration of the three properties we believe any good solution to this challenge should possess:

**Generality.** Having trained on a specific data set, the method should be capable of classifying new networks without the need of repeating the calculation.

To illustrate this point, we defined the principal components using a set of 100 networks of each kind among the nonecological sets and projected them on the space defined by the first two components (Fig. A in S1 Supporting Information, left pane). The networks cluster by type, and the two axes explain much of the variation (Table 2). When we project onto this space all the remaining nonecological networks, we find that, despite being "out of fit", the new networks fall exactly where they should be: the method can be used to classify new networks based on their position in the plane.

**Specificity.** The method should not be able to classify the networks when the structure has been sufficiently perturbed.

In particular, if we were to randomize the networks, they should all be classified as "random networks" rather than retaining their identity: at its most basic level, specificity is needed to prove that the analysis is not overly influenced by size and connectivity.

The top panel shows the empirical, nonecological networks, a subset of which were used to construct the principal component space—which is maintained throughout the remaining

plots. The middle panel keeps the same axes, but instead of projecting the empirical networks, it uses the properties measured for their randomized version under the Erdős–Rényi model. Likewise, the bottom panel shows a point for each network once randomized using a configuration model; this randomization is more computationally taxing; hence, there are fewer data points in this panel, especially for the large and sparse actor collaboration networks (Supporting information).

To illustrate this point, we produced two randomized versions of each original network: an Erdős–Rényi random bipartite graph [17] with the same number of nodes in each class and the same number of connections as in the original graph and a network built with a configuration model [18, 19], preserving also the degree sequence of the original graph. When we compute the three metrics illustrated above for each randomization and project the resulting values on the PCA space, we find that almost all networks cluster together under the Erdős–Rényi model (Fig 1); note that ideally these should collapse to a point—the remaining scatter indicates that our approximation is not accurate for very sparse or very small networks, possibly because the size of the fluctuations of these quantities around their mean is influenced by size and connectance. When we analyze the data from the configuration model, we start seeing a clear separation of the clusters, meaning that degree distributions differ significantly between classes. In the same spirit, one could devise other randomizations preserving or foregoing certain properties, thereby probing whether they can be used to distinguish networks.

**Scalability.** There should be different levels of classification, displaying a hierarchical structure.

Previously, we stated that the method should be able to consistently distinguish networks of different classes—this language is intentionally imprecise. We use "class" loosely such that it could refer to separating networks representing mutualistic interactions from those depicting antagonisms, distinguishing hummingbird pollination networks from those in which the pollinators are bees, or even just recognizing ecological networks. This ambiguity is intentional in order to remain agnostic with respect to what the appropriate level of differentiation should be or even whether there should be one primary level. Analogous to the issue of classifying organisms in a taxonomy, there should be a hierarchy of differentiation that can be used to infer relevant aspects of these systems. For this reason, a successful method should be able to identify levels of segregation within the data according to some hierarchy of network similarity.

We demonstrate this property in Fig 2, in which we look more closely at the cluster of crime networks in the top panel of Fig 1. Although Fig 1 shows that crime networks cluster more closely to one another than to networks of actor collaboration, legislature voting records, or human microbiome networks, Fig 2 reveals substructure within this cloud of points corresponding to networks collected for the same city: although all crime networks cluster together, those in the same city cluster more tightly.

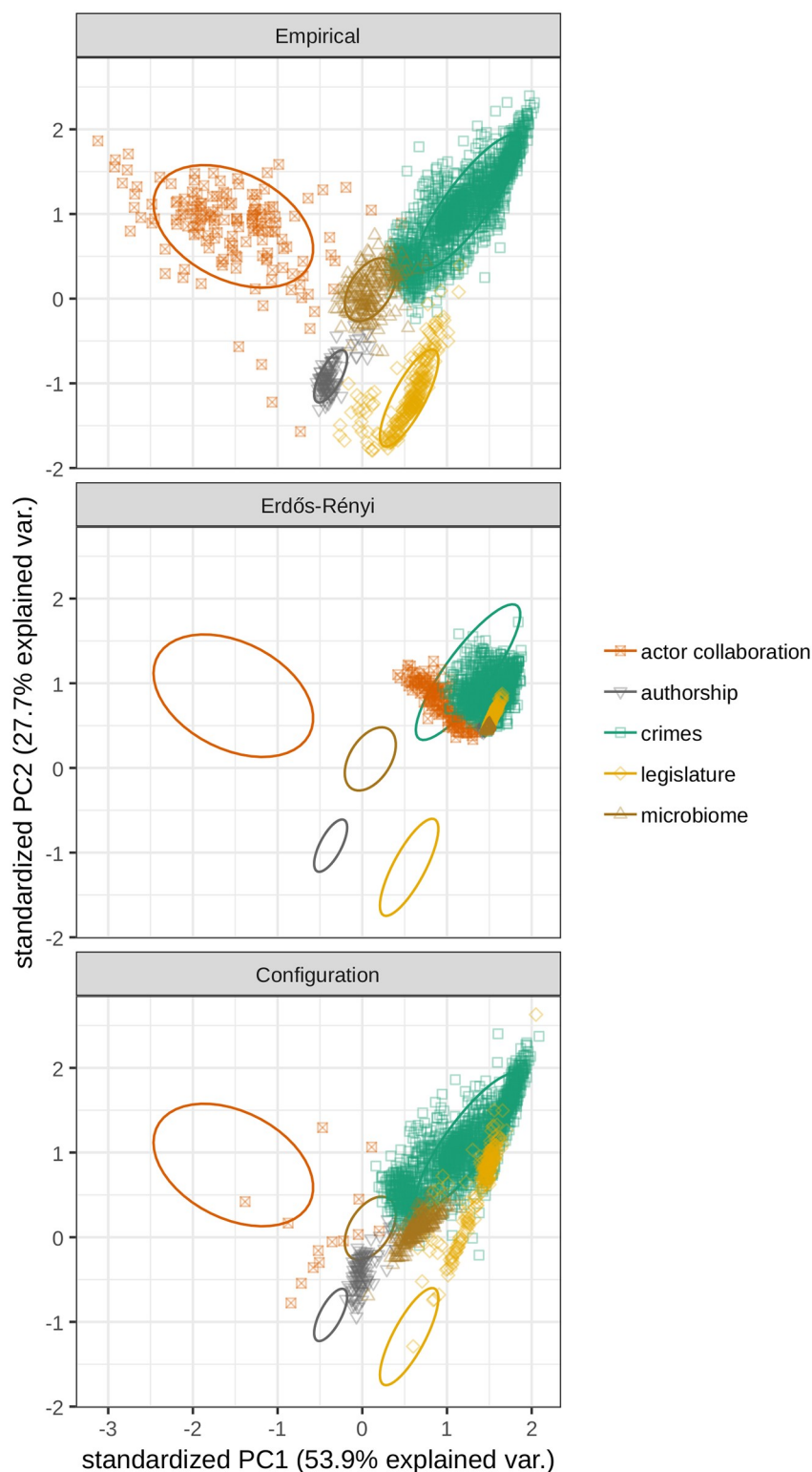
## Results and discussion

### Adding ecological networks to the map

Having demonstrated some level of success in achieving generality, specificity, and scalability, we now turn to the focus of our challenge: applying the method to ecological interaction networks.

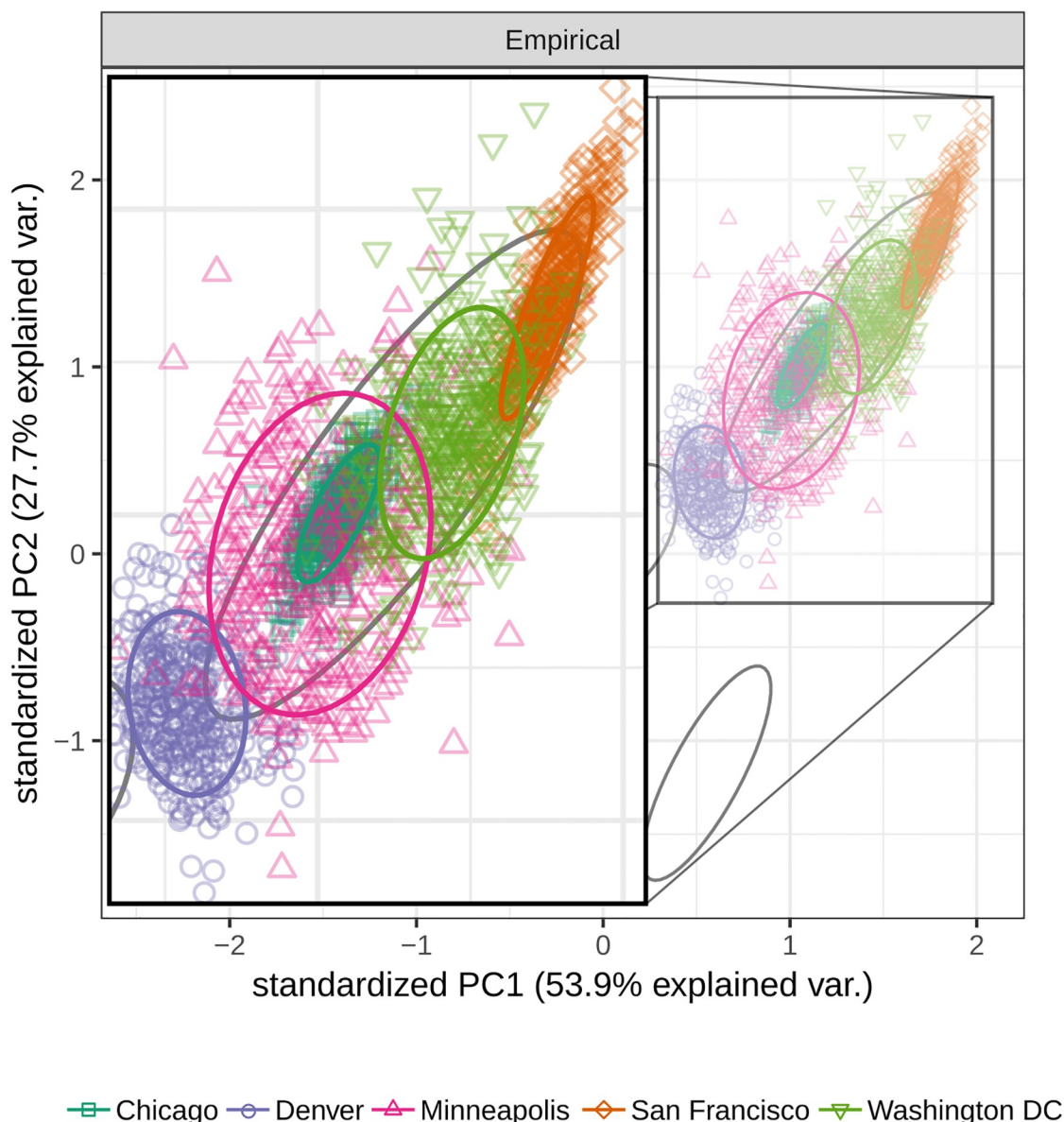
When we project our ecological networks onto the principal component space constructed above, we find that they span a large area of the map, overlapping many of the categories that were previously well separated. Moreover, antagonistic and mutualistic networks overlap substantially (Fig 3). These results persist when only highly replicated networks are used, when





**Fig 1. PCA biplots of the two most explanatory principal components.** Each point indicates a single network, and the ellipses are drawn to contain approximately 68% of the points in each network type, i.e., one standard deviation if the points were to follow a bivariate normal distribution. PC1, principal component 1; PC2, principal component 2; PCA, principal component analysis; var., variation.

<https://doi.org/10.1371/journal.pcbi.1007076.g001>



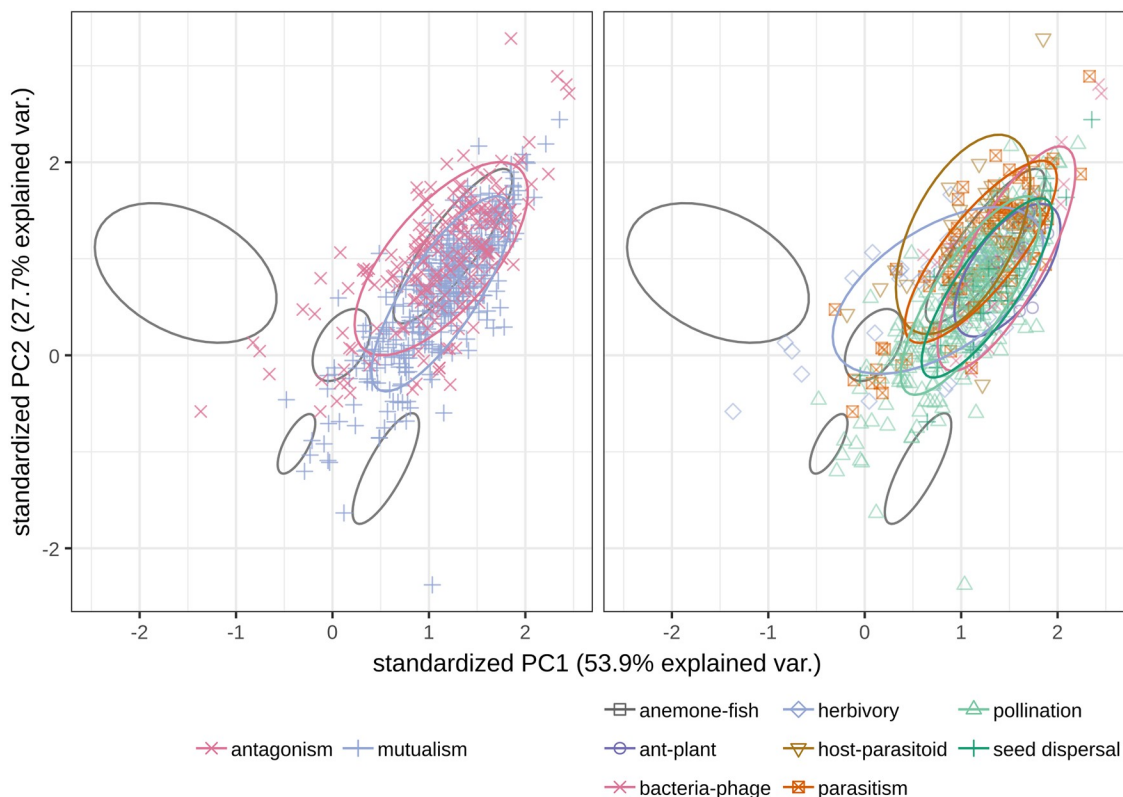
**Fig 2.** Subsetting the points from the top panel of Fig 1 to include only the networks of crime locations, we now color the points according to the city from which each crime network was collected. Note that, though all crime networks cluster together in Fig 1, those from the same city cluster more tightly, albeit with significant overlap between cities. This indicates that there is additional structure beyond that used to disambiguate crime networks from those of authorship, legislature, etc., and by looking at these finer differences, more subtle distinctions can be made. PC1, principal component 1; PC2, principal component 2; var., variation.

<https://doi.org/10.1371/journal.pcbi.1007076.g002>

ecological networks are utilized in the construction of the principal component space, and when alternative metrics are used to construct the PCA (Supporting information).

The results presented in Fig 3 are somewhat surprising—Why are ecological networks hard to classify when there have been previous publications claiming success? Using our new, larger data set, we can reassess the proposed differences between networks of antagonistic and mutualistic interactions. Using a smaller data set, Thébault and Fontaine [12] had found significant differences in nestedness [20] and modularity [4] between mutualistic and antagonistic networks. We repeat and extend this analysis by looking at additional measures of nestedness,





**Fig 3.** Left: overlaying more than 500 ecological networks representing either mutualistic or antagonistic interactions, we find that they do not cluster as nicely as the non-ecological ones did, suggesting that there is much more variation within ecological network structure than between the classes of non-ecological networks examined here. Right: this result holds when the networks are instead labeled according to the specific type of mutualism/antagonism they describe. PC1, principal component 1; PC2, principal component 2; var., variation.

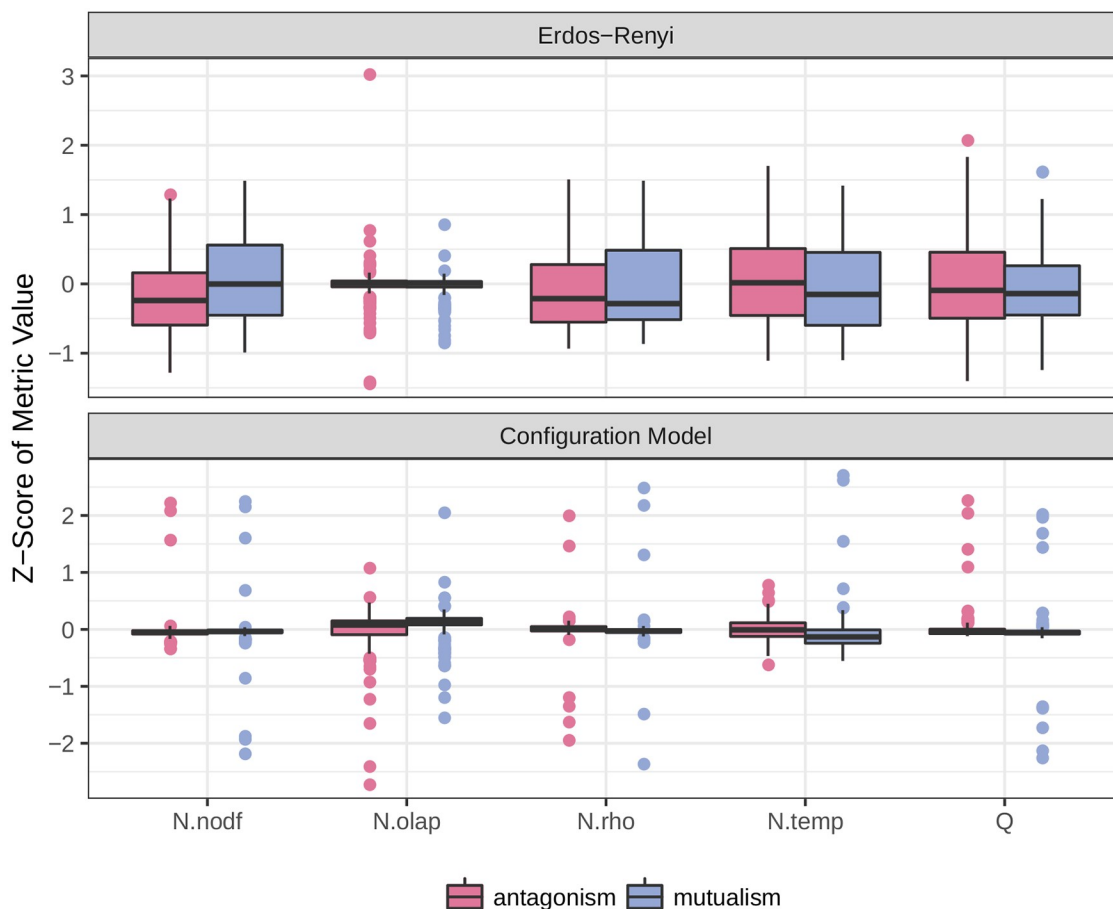
<https://doi.org/10.1371/journal.pcbi.1007076.g003>

constraining the null model to ensure connectedness, and adding a second null model (Fig 4 and Supporting information). We find conflicting support for the use of these measures in the classification of ecological networks. Indeed, both the size and direction of these differences can be influenced by the choice of metric. Moreover, even in the case of a significant difference, there is substantial overlap in the metric distributions, suggesting that these measures cannot alone reliably distinguish network types (i.e., though the means could be significantly different from a statistical standpoint, the high variance would render a classification based on these features difficult).

### Choice of method and metrics

For each potential method to solve the problem outlined here, one of the key decisions is which (and how many) metrics to include in the analysis, and a plethora of statistics are currently available. For illustrative purposes, we kept the number of metrics to a minimum, but in the Supporting information, we experiment with many more, obtaining the same qualitative results.

Note that the choice of metrics is not inconsequential: including too few will provide insufficient power to distinguish between similar network types, whereas too many can lead (in some modeling frameworks) to noise or multicollinearity that can obscure meaningful differences. There are also logistical concerns, as some metrics require much more computation



**Fig 4. We compared networks of ecological interactions based on their value of nestedness and modularity.** We randomized each empirical web 1,000 times according to an Erdős-Rényi (top) or configuration model (bottom), each modified slightly to produce only connected graphs (Supporting information). For each of these randomizations, we calculated nestedness [in a variety of ways, including NODF (N.nodf; [20]), overlap (N.olap; [21]), spectral radius (N.rho; [15]), and temperature (N.temp; [5])] and modularity (Q; [4]). We plot here the Z-score of the empirical value for each measure with respect to the 1,000 randomizations. For the Erdős-Rényi randomizations, the only significant (using Welch's [22] *t* test; Supporting information) difference is for NODF, showing mutualistic networks to be more nested than antagonistic ones. For the configuration null model, only the Overlap measure of nestedness shows significant differences (Supporting information), but in this case, networks of antagonistic interactions are deemed more nested than those depicting mutualisms. Modularity does not vary significantly between the two types of ecological interaction networks.

<https://doi.org/10.1371/journal.pcbi.1007076.g004>

time than others. Certain estimates of modularity, for instance, require first identifying the appropriate clustering of the species involved—a computationally intensive problem [23].

Similarly, there are a variety of methods that could be applied to this problem. Although we used simple spectral measures and PCA to render the network space in two dimensions here, a number of techniques, in particular those in the rapidly growing field of machine learning, are arguably better suited to this task. However, although popular machine learning techniques typically require the availability of a properly labeled "training set", here we achieved a good separation between different classes of nonecological networks without supplying any label—we simply mapped out the networks based on their spectral properties.

We used PCA to produce a two-dimensional "map" onto which we can place new networks. Yet PCA has a number of shortcomings. Foremost is that the space constructed varies with the data used. If we were to include more, or different, nonecological networks in the generation

phase, we would obtain a different map. Additionally, we have not run statistics on the clustering we observe in principal component space, introducing an element of subjectivity into the analysis as stated. Alternative or additional methods could be implemented that incorporate objective grouping to enhance the robustness of any results.

Finally, we have analyzed unweighted, undirected graphs, despite the fact that many modern ecological networks include measures of interaction strengths [24, 25]. Although weights are likely to help with separating different classes of networks, many of the most popular network properties (e.g., motif profiles) are difficult to extend to weighted networks.

## Conclusion

In 1966, Mark Kac asked, "Can One Hear the Shape of a Drum?" Is it possible to infer the shape of a drumhead from a list of the overtones it produces [26]? Although the short answer is no, many important properties of the drum, such as its area and perimeter, can be "heard" distinctly.

Here, we asked whether the type of interaction being represented in a network can be "seen" from its structure. We have shown that the classification is easy for **several classes of nonecological networks: even foregoing size and fill, the structure of these networks is distinctive enough to allow for a reliable classification.**

Yet the same task is more complicated when examining ecological networks. In this case, the naïve approach that successfully classified nonecological networks fails completely, and adding more metrics or using different approaches seems not to help with the classification (Supporting information). Because of this fact, we framed the problem as a computational challenge to the community of scientists interested in networks and machine learning. We have compiled a large data set that can be used to explore and validate potential solutions, and we have detailed three important properties that need to be fulfilled.

Succeeding in this challenge would have profound implications for the field: although networks are surely convenient objects for storing information about species' interactions, a positive solution would prove that, by representing ecological communities as networks, we gain insights that would be precluded otherwise. Moreover, as for other areas of machine learning (e.g., face recognition), with the advent of new techniques for the high-throughput production of networks, this question could rapidly move from the purely academic to the applied side. Although current approaches to constructing networks are often laborious, we know what these networks represent. **In the future, our ability to produce networks could surpass our ability to identify what interactions represent, leading to novel confusions and challenges.** Finally, network metrics that can capture essential aspects of the ecology of these systems would surely deserve a special place in the hearts and minds of ecologists.

But what if a solution to this challenge cannot be found? This failure could be due to a number of reasons. First, unweighted, bipartite graphs representing a single type of interaction might not carry sufficient information for this task. Many of the ecological networks currently being published quantify interactions (e.g., measuring the number of visits a pollinator pays to a plant), and a few started detailing the different types of interactions between the species using **multidimensional networks** [27–30]. As such, the solution could present itself once one were to extend the appropriate metrics to weighted, multidimensional networks. Second, it could be that ecological networks have greater within-class variation than the other classes examined here. Third, the difference between mutualistic and antagonistic relationships might not be the main driver of network structure, and other, more-meaningful classifications could emerge that would be easy to infer when examining ecological network structure. Ultimately, it is likely a combination of these explanations and others not listed here, the complete

description of which remains a topic for future investigation. In any case, there would be much to learn from a failure, and to celebrate in case of success.

## Supporting information

**S1 Supporting Information. Data collection sources and procedures, detailed and additional methods, and supplementary figures and tables.**  
(PDF)

## Acknowledgments

This work benefited from the feedback of Mercedes Pascual, Trevor Price, Sarah Cobey, Amos Maritan, Daniel Maynard, Zachary Miller, and Jacopo Grilli. Thanks to Phillip Staniczenko, Gyuri Barabás, Samraat Pawar, Elisa Thébault, and Liz Sander for discussion. This work utilized the computational resources of the Center for Research Informatics' Gardner HPC cluster at the University of Chicago ([cri.uchicago.edu](http://cri.uchicago.edu)). Data on authors networks were from Scopus ([scopus.com](http://scopus.com)); data on actor–movie networks were from the Movie Database ([themoviedb.org](http://themoviedb.org)); data on crime networks were from various city data portals ([data.cityofchicago.org](http://data.cityofchicago.org), [denvergov.org/opendata](http://denvergov.org/opendata), [minneapolisnmn.gov/opendata](http://minneapolisnmn.gov/opendata), [datasf.org/opendata](http://datasf.org/opendata), [opendata.dc.gov](http://opendata.dc.gov)); data on human microbiomes were from the NIH Human Microbiome Project ([hmpdacc.org](http://hmpdacc.org)); and data on legislature voting rolls were from GovTrack ([govtrack.us](http://govtrack.us)), the Harvard Dataverse ([dataverse.harvard.edu](http://dataverse.harvard.edu)), and VoteWorld ([voteworld.berkeley.edu](http://voteworld.berkeley.edu)).

## References

1. Bersier LF. A history of the study of ecological networks. In: Biological networks. Singapore: World Scientific; 2007. p. 365–421.
2. Ings TC, Hawes JE. The History of Ecological Networks. In: Ecological Networks in the Tropics. New York City: Springer; 2018. p. 15–28.
3. Camerano L. Dell'equilibrio dei viventi mercè la reciproca distruzione. Accademia delle Scienze di Torino. 1880; 15:393–414.
4. Newman ME. Modularity and community structure in networks. Proceedings of the National Academy of Sciences. 2006; 103(23):8577–8582.
5. Wright DH, Reeves JH. On the meaning and measurement of nestedness of species assemblages. Oecologia. 1992; 92(3):416–428. <https://doi.org/10.1007/BF00317469> PMID: 28312609
6. Bascompte J, Jordano P, Melián CJ, Olesen JM. The nested assembly of plant–animal mutualistic networks. Proceedings of the National Academy of Sciences. 2003; 100(16):9383–9387.
7. Ulrich W, Almeida-Neto M, Gotelli NJ. A consumer's guide to nestedness analysis. Oikos. 2009; 118(1):3–17.
8. Milo R, Shen-Orr S, Itzkovitz S, Kashtan N, Chklovskii D, Alon U. Network motifs: simple building blocks of complex networks. Science. 2002; 298(5594):824–827. <https://doi.org/10.1126/science.298.5594.824> PMID: 12399590
9. Seidman SB. Network structure and minimum degree. Social Networks. 1983; 5(3):269–287.
10. Thrall PH, Hochberg ME, Burdon JJ, Bever JD. Coevolution of symbiotic mutualists and parasites in a community context. Trends in Ecology & Evolution. 2007; 22(3):120–126.
11. Lewinsohn TM, Inácio Prado P, Jordano P, Bascompte J, Olesen JM. Structure in plant–animal interaction assemblages. Oikos. 2006; 113(1):174–184.
12. Thébault E, Fontaine C. Stability of ecological communities and the architecture of mutualistic and trophic networks. Science. 2010; 329(5993):853–856. <https://doi.org/10.1126/science.1188321> PMID: 20705861
13. Thompson JN. The coevolutionary process. Chicago: University of Chicago Press; 1994.
14. Thompson JN. The geographic mosaic of coevolution. Chicago: University of Chicago Press; 2005.
15. Staniczenko PP, Kopp JC, Allesina S. The ghost of nestedness in ecological networks. Nature Communications. 2013; 4:1391. <https://doi.org/10.1038/ncomms2422> PMID: 23340431

16. Newman ME. Spectral methods for community detection and graph partitioning. *Physical Review E*. 2013; 88(4):042822.
17. Erdős P, Rényi A. On random graphs. *Publicationes Mathematicae Debrecen*. 1959; 6:290–297.
18. Bender EA, Canfield ER. The asymptotic number of labeled graphs with given degree sequences. *Journal of Combinatorial Theory, Series A*. 1978; 24(3):296–307.
19. Strona G, Nappo D, Boccacci F, Fattorini S, San-Miguel-Ayaz J. A fast and unbiased procedure to randomize ecological binary matrices with fixed row and column totals. *Nature Communications*. 2014; 5:4114. <https://doi.org/10.1038/ncomms5114> PMID: 24916345
20. Almeida-Neto M, Guimarães P, Guimarães PR, Loyola RD, Ulrich W. A consistent metric for nestedness analysis in ecological systems: reconciling concept and measurement. *Oikos*. 2008; 117(8):1227–1239.
21. Strona G, Veech JA. A new measure of ecological network structure based on node overlap and segregation. *Methods in Ecology and Evolution*. 2015; 6(8):907–915.
22. Welch BL. The generalization of Student's problems when several different population variances are involved. *Biometrika*. 1947; 34(1/2):28–35.
23. Brandes U, Delling D, Gaertler M, Gorke R, Hoefer M, Nikoloski Z, et al. On modularity clustering. *IEEE Transactions on Knowledge and Data Engineering*. 2008; 20(2):172–188.
24. Vázquez DP, Aizen MA. Null model analyses of specialization in plant–pollinator interactions. *Ecology*. 2003; 84(9):2493–2501.
25. Bascompte J, Jordano P, Olesen JM. Asymmetric coevolutionary networks facilitate biodiversity maintenance. *Science*. 2006; 312(5772):431–433. <https://doi.org/10.1126/science.1123412> PMID: 16627742
26. Kac M. Can one hear the shape of a drum? *The American Mathematical Monthly*. 1966; 73(4P2):1–23.
27. Sander EL, Wootton JT, Allesina S. What can interaction webs tell us about species roles? *PLoS Comput Biol*. 2015; 11(7):e1004330. <https://doi.org/10.1371/journal.pcbi.1004330> PMID: 26197151
28. Mougi A, Kondoh M. Diversity of interaction types and ecological community stability. *Science*. 2012; 337(6092):349–351. <https://doi.org/10.1126/science.1220529> PMID: 22822151
29. Fontaine C, Guimarães PR, Kéfi S, Loeuille N, Memmott J, van Der Putten WH, et al. The ecological and evolutionary implications of merging different types of networks. *Ecology Letters*. 2011; 14(11):1170–1181. <https://doi.org/10.1111/j.1461-0248.2011.01688.x> PMID: 21951949
30. Kéfi S, Miele V, Wieters EA, Navarrete SA, Berlow EL. How structured is the entangled bank? The surprisingly simple organization of multiplex ecological networks leads to increased persistence and resilience. *PLoS Biol*. 2016; 14(8):e1002527. <https://doi.org/10.1371/journal.pbio.1002527> PMID: 27487303