Practical Analysis of Ecological Networks

Timothée Poisot

May 20, 2020

Contents

Chapter 1

Introduction

When we first attempted to synthetize the literature on *how to best analyse ecological networks?* (Delmas et al. 2018), two things became clear. First, an increasing number of fields in ecology were using networks as a formalism to represent and analyse their data. Second, methods were multiplying at an alarming rate, and there was a great degree of methodological confusion about what to use, how to interpret it, and when not to use it. In short, the field of ecological network analysis was long due for a textbook.

1.1 Why networks?

1.2 What this book is about

1.2.1 The structure of ecological networks

1.2.2 Best practices in ecological networks research

1.2.3 Programming

1.3 What this book is not about

1.3.1 The ecology of ecological networks

1.3.2 The foundation of network theory

There are a number of fantastic textbooks on network science, where the mathematical foundations of this field are laid out in a clear and comprehensive way. One of the best ways to share the excitement about why the mathematics of

networks theory are worth looking into is to read a short, and very accessible, perspective by Strogatz (2001). Should this whet your appetite, there are multiple resources that are suitable to get a deeper education in graph theory. Two outstanding references are the now classic Newman (2010), and the more recent Barabási (2016). For readers seeking more introductory material, we recommed the short and very accessible volume by Chartrand (1985), as well as West (2001).

But fear not – this book is not about mathematics. The measures of ecological network structure we will discuss have are grounded in mathematics, but the point is *not* to go into much details into it. There will be a *fair* amount of mathematics, probably as much as during an advanced class in statistics (which is to say just enough to fill the average ecologist with apprehension, but not quite with dread), but it will not be the point of the material. This is merely something that comes with the analysis of ecological networks. You will realize through the different chapters that there is only a very small quantity of graph theory we need to know to do good ecology; this will have the highly desirable side effect to keep any mathematical posturing to a minimum.

1.3.3 Programming

1.4 How to read this book

1.4.1 The importance of primary literature

1.4.2 The importance of experimenting with code

This books involves *a lot* of computer code; in fact, it is mostly about writing computer code to analyse ecological networks. Although we rely extensively on a software package (**poisot_ecologicalnetworks.jl_2019**) to do this, there will be many applications where writing our own functions, or writing our own scripts, will be necessary. Copying and pasting this code, and running it exactly as presented here, is a terrible idea. Instead, we encourage readers to *adapt* the code to their own interests, uses, and questions.

How does one adapts code? **tk**

1.4.3 The importance of consistent notation

Throughout this book, we strive to use notation that remains as constant as possible. Networks are represented by capital letters, with the exact letter used representing some additional information about the type of network. Unipartite networks are U, bipartite networks are B, probabilistic networks are P, random networks are R, quantitative networks are G (only because Q is used for modularity already), and a network of an unspecified type is N. When there is a

collection of multiple networks, we use the same letter in bold face – for example, we can note that a function generates random networks from a bipartite network using $f(B) = \mathbf{R}$.

To refer to a specific interaction, we will use the notation N_{ij}, which represents an interaction *from* species i *to* species j. In some situations, we will also use $N(i, j)$ to represent the same information. The notation $N_{i\cdot}$ and $N_{\cdot j}$ indicate, respectively, the set of species with which i interacts, and the set of species that interact with j. For example, the number of predators of species 4 in a unipartite food web is $\|U_{\cdot 4}\|$.

Chapter 2

Comparing ecological networks

Although this might seem like a more advanced topic, comparing network is relatively easy in that it mostly involves enumerating elements – how many species, and how many interactions, are either unique to one, or shared by two networks? As such, a good understanding of the core concepts in **??** will be useful for this chapter.

To develop an intuition of why ecological networks differ, it is useful to think of them as sets. Not only does it allows defining a number of mathematical operations that will greatly simplify the notation, it also matches directly the formalism developped by Koleff, Gaston, and Lennon (2003) for the dissimilarity of presence-absence data, which we will follow throughout this chapter.

2.1 Pairwise network comparison

2.2 The metaweb

In the previous section, we have focused on the fact that different networks can share, or not, some of their species, and some of their interactions. This provided a powerful intuition to think about network dissimilarity in terms of α and β diversity - but what about γ? In this section, we will think about the "metaweb", defined by **Dunn06** as the aggregation of all regional interactions in a series of related networks. In other words, we can define a metaweb as the recursive union of a collection of networks.

2.2.1 Metaweb properties

2.2.2 Network distance to the metaweb

2.3 Multi-site network comparison

Chapter 3

Motifs in ecological networks

Chapter 4

Nestedness and interaction overlap

Chapter 5

The deeply flawed art of null hypothesis significance testing

```
using EcologicalNetworks
import Mangal
using Statistics
```

Ecologists, like sadly a great majority of scientists, decide on the worthiness of an observation by ignoring their best judgement, decades of training in biology, and expertise; instead, we rely on whether some arbitrary value (the p-value) is lower than some even more arbitrary threshold (0.05). It all seems very *ad hoc* (it is). And yet, a surprising amount of literature on ecological networks attempts to decide whether some observed value of a network measure is "significant", and so it is with great reluctance that we will dedicate a chapter to this practice.

5.1 Fundamentals of network NHST

Null Hypothesis Significance Testing (NHST) for ecological networks strives to answer the following question: "if I observed the value f_0 for the measure f on a network $\mathbf{N}$, is it in the range of values expected by chance?". Most often, we want to refine this question by asking if the value is larger or smaller than we expect, and this is why this process usually involves a one-tailed one-sample t-test.

5.1.1 Generating a sample

Given a function f and a network $\mathbf{N}$, NHST requires to measure the distribution of possible values of $f(\mathbf{N})$. For most measures (all of them, in fact), there is no direct expression of this distribution, and so we need to rely on an imperfect proxy: generating random networks. The exact models under which these models are generated is explained in section **TK**.

5.1.2 From the sample to the distribution

5.1.3 Estimating significance

5.2 Overview of the null models

5.2.1 Probabilistic models

5.2.2 Generalization of the probabilistic models

$$P_{ij} = \alpha_1 N_{ij} + \alpha_2 \frac{N_{.j}}{S_1} + \alpha_3 \frac{N_{i.}}{S_2} + \alpha_4 \frac{L}{S_1 S_2} \tag{5.1}$$

$$P = lf(N, \alpha = (0, 0, 0, 1)) \tag{5.2}$$

5.2.3 Permutational models

5.3 Application – nestedness of oil-flowers pollination networks

In this section, we will use the NHST process outline in figure **??** to determine whether bipartite networks are *more* nested than expected by chance, under two assumptions as to what drives the structure of the model

5.2

5.3.1 Preparing the data and measuring nestedness

We will first work on a dataset of pollination interactions between insects and oil-flowers, collected by Bezerra, Machado, and Mello (2009).

```
oilflower_network = Mangal.network(929)
N = convert(UnipartiteNetwork, oilflower_network)
B = convert(BipartiteNetwork, N)
```

```
13×13 bipartite  ecological network (Bool, Mangal.MangalNode) (L: 71)
```

We will use NODF (**tk**) to measure the nestedness of this network. This is an important step, as it will provide our f_0, the reference value, measured on the empirical network:

```
f0 = nodf(B)
```

```
0.849310643060643
```

Recall that NODF takes values between 0 and 1, where 1 indicates perfect nestedness – this network has therefore a rather high nestedness, and it makes sense to ask whether this value is larger than expected by chance. Our definition of *chance* here will be represented by two null models: either the nestedness is larger than expected given the value of connectance, or the nestedness is larger than expected given the degree distribution of the network.

5.3.2 Generating the null sample

```
P = null2(B)
R = rand(P, 5000)
```

TK

```
f = nodf.(R)
```

We can start looking at the average value of the nestedness of the random networks:

```
mean(f)
```

```
NaN
```

Well, that didn't work.

5.3.3 Fixing the null sample

This is probably because the NODF measure is failing on some networks, and returning a value of `NaN`. Let's see which networks are to blame:

```
X = R[findall(isnan.(f))]
```

We can look for all sorts of reasons to understand why the NODF measure is failing. The answer is as follows: as you will recall from section **??**, NODF involves dividing the number of overlapping interactions by the XXX number of interactions. So if one species has no interactions, we would end up dividing by 0, which would result in `NaN`. How could a species have no interaction? The null

model we use is in fact a probabilistic network, and we know from **probanet** that there is a chance (which can be very small), that a species will end up having no interactions if we perform independent random draws.

This provides an interesting hypothesis: that the networks for which we do not have a NODF score have "degenerate" matrices, after the **twosides** nomenclature.

```
all(isdegenerate.(X))
```

```
true
```

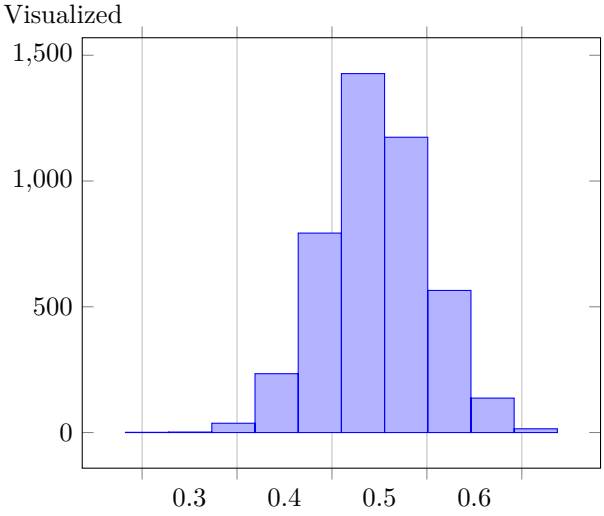
This can be solved in two ways. We can either decide to remove these networks, or we can decide to remove the species that are not connected. In section **??**, we will discuss the consequences of this choice, but for now, we will remove thes networks:

```
filter!(!isdegenerate, R)
```

We can now update the distribution of values:

```
f = nodf.(R)
```

Visualized



5.3.4 Measuring significance

```
sum(f .>= f0)/length(f)
```

```
0.0
```

5.4 Application – looking for generalities

5.5 Overview of the issues

5.5.1 Degenerate matrices

5.5.2 Strongly constrained matrices

5.5.3 Representativity of the sample

Bibliography

Barabási, Albert-LászÍo (2016). *Network Science*. Cambridge University Press.

Bezerra, Elisângela L. S., Isabel C. Machado, and Marco A. R. Mello (2009). "Pollination Networks of Oil-Flowers: A Tiny World within the Smallest of All Worlds". In: *Journal of Animal Ecology* 78.5, pp. 1096–1101. ISSN: 1365-2656. DOI: 10.1111/j.1365-2656.2009.01567.x.

Chartrand, Gary (1985). *Introductory Graph Theory*. Unabridged and corr. New York: Dover. 294 pp. ISBN: 978-0-486-24775-5.

Delmas, Eva et al. (June 20, 2018). "Analysing Ecological Networks of Species Interactions". In: *Biological Reviews*, p. 112540. ISSN: 14647931. DOI: 10.1111/brv.12433.

Koleff, Patricia, Kevin J. Gaston, and Jack J. Lennon (May 1, 2003). "Measuring Beta Diversity for Presence–Absence Data". In: *Journal of Animal Ecology* 72.3, pp. 367–382. ISSN: 1365-2656. DOI: 10.1046/j.1365-2656.2003.00710.x.

Newman, Mark E. J. (2010). *Networks. An Introduction*. New York, NY: Oxford University Press. URL: http://dl.acm.org/citation.cfm?id=1809753.

Strogatz, Steven H. (Mar. 8, 2001). "Exploring Complex Networks". In: *Nature* 410.6825, pp. 268–276. ISSN: 0028-0836. DOI: 10.1038/35065725.

West, Douglas Brent (2001). *Introduction to Graph Theory*. 2nd ed. Upper Saddle River, N.J: Prentice Hall. 588 pp. ISBN: 978-0-13-014400-3.