

bipy user manual

Timothée Poisot

Version pre 1.0

Contents

1	Introduction	5
1.1	Installing bipy	5
1.2	Do I need to know Python ?	5
1.3	How do I cite bipy ?	6
2	Loading data	7
2.1	What should the data look like ?	7
2.2	Reading data from a local file	8
2.3	Reading data from the web	8
2.4	Adding species names	9
2.5	Adding bibliographical references	9
3	Species-level metrics	11
3.1	Degree	11
3.2	Specificity	11
3.2.1	Specificity	11
3.2.2	Species-specialization index	11
3.2.3	Resource-range	11
4	Network-level metrics	13
4.1	Nestedness	13
4.2	Modularity	13
4.3	Robustness to extinction	13
5	Null models	15
5.1	The wrapper for null models	15
5.2	Null models	15
5.2.1	Null model 1 – Bernoulli fill (null_1)	15
5.2.2	Null model 2 – interaction average (null_2)	15
5.2.3	Null model 3 – row averages (null_3row)	16
5.2.4	Null model 3 – column averages (null_3col)	16
5.3	Tests of the deviation of common measures	16

Contents

6	Graphics	17
6.1	Nested webs	17
6.2	Modular webs	17
7	Example scripts	19

1 Introduction

bipy is a set of functions written in Python, whose goal is to make it easy to analyze and visualize bipartite networks. This document describes the use of the bipy. This documentation is valid only for the versions before the 1.0 release (scheduled to occur sometime in the fall or winter of 2011). The aim of this documentation is to present only the “high-level” functions, i.e. these that the user will need to perform the analyses (although bipy was conceived to automate most of the analyses, so there are a very small number of functions that need to be known). In addition, some background about the measures used and the most important references will be given.

This guide should be viewed as a quick walkthrough to the features, and is intended primarily for the people that want to analyze their data without spending much effort in understanding the mechanisms behind the analyses. It also serves the purpose of detailing which measures are used, and why they were chosen.

1.1 Installing bipy

bipy can be obtained on *GitHub* (<http://github.com/tpoisot/bipy>). There are a number of required additional software. Obviously, bipy requires a Python installation, preferably 2.6 or 2.7. For those of you with a Mac, you need to download the version from python.org (<http://www.python.org/getit/>).

Once this is done,

To optimize performance, bipy makes an intensive use of the `numpy` and `scipy` packages.

1.2 Do I need to know Python ?

This is a good question! If you just want to have basic informations about your data, the short answer is **no**. bipy was designed from the beginning to automate most of the calculations, and use the most robust statistics available, so you don't have to play too much with the code. Most of what you will have to do is start python, load your data, and use the summary functions.

If you want to use bipy for more complicated analyses, then **yes**, learning python will definitely be a huge plus. But if you are willing to adopt a computational approach,

1 Introduction

chances are you already know a programming language, and you should be able to pick up python really quick.

1.3 How do I cite bipy ?

For the moment, just cite (by replacing the date with the day you downloaded it):

Poisot, T. (2011) bipy – bipartite networks analysis using python.
<http://github.com/tpoisot/bipy>. Downloaded on November 21, 2011.

A paper formally describing bipy (along with alternatives in other languages) is currently being written.

2 Loading data

This chapter describes how to load data in a format that can be understood by `bipy`. There are two different ways to load data in `bipy`: *via* a text file, or *via* the web. We will cover both in this chapter.

2.1 What should the data look like ?

`bipy` assumes that a bipartite network is nothing more than a matrix, in which each square represents the strength of the interaction between a species from the top trophic level, and a species from the bottom trophic level. Thus, a bipartite network is defined by a matrix \mathbf{M} , in which \mathbf{M}_{ij} is the strength of the link between upper trophic level species i and lower trophic level species j . It is assumed that the rows of \mathbf{M} are the upper trophic level species. This is a notational convention, but we will stick to it through all the documentation.

The advantage of this organization of the data is that it can easily be stored in a text file, with each interaction strength separated by a space, and the different organisms separated by a new line. Thus, the file

```
1 0 0 0 1
0 0 1 1 1
1 1 0 0 1
0 1 0 0 0
```

described the network illustrated in the margin. Of course, the values within the file can be anything (as long as they are numeric): zeroes and ones, number of visits, predation rate, percent of the diet, and so on.

However, an interaction network is much more than the interaction matrix. When an interaction matrix is loaded, `bipy` will automatically initiate the calculation of several important metrics. This has two advantages. First, these metrics are needed for multiple calculations, so instead of being re-calculated several times, they are calculated only once. This speeds-up calculation time. Second, the object is easier to manipulate, as it keeps all the relevant informations together. For example, the connectance of a bipartite object called `w` is `w.connectance`.

Loading data will read the interaction matrix, and generate this bipartite object, which will be used in nearly all the analyses after. This requires no input from the user, so getting the data ready to work is really straightforward.

2.2 Reading data from a local file

Assuming that your data are in this format, and the file is called `mydata.web`, you can load them

Listing 2.1: loading data from a file

```
1 web = loadweb('mydata.web')
```

At this point, the most difficult part is done. Your data are loaded, and you will be able to see what story they tell using the metrics described in the other chapters.

If your data are not in the “correct” format, i.e. the top level organisms are in columns, do not panic. You just need to pass a supplementary argument `t` to the `loadweb` function:

Listing 2.2: loading data from a file and transposing the data

```
1 web = loadweb('mydata.web', t=True)
```

This will tell `bipy` that your data need to be transposed before going on with the analysis. For a non-console version, just use `loadweb` with no parameters (you can still specify `t=True` if you want to transpose your data):

Listing 2.3: opening a file dialog

```
1 web = loadweb()
```

`loadweb` will open the file selection dialog of your operating system, so you can pick your dataset.

2.3 Reading data from the web

The alternative way is to load data stored on the internet. `bipy` is able to do so by using the `readRemoteWeb` function, which requires two arguments. The first one is the URL of a text file, which stores the interaction matrix. The second tells if the network should be transposed (false by default).

Listing 2.4: reading a dataset from the web

```
1 web = readRemoteWeb('http://server/mydata.web'), False)
```

This method can be used to load networks made publicly available on the `rwdb` (<http://www.nceas.ucsb.edu/interactionweb/>), for example.

2.4 Adding species names

bipy has the ability to store the names of the species in the network. Currently, the way to do it is to assign a value to the properties `lonames` and `upnames` of a `bipartite` object. An easier way to load the data is currently being implemented.

The code to assign species names is:

Listing 2.5: attributing species names

```
1 web = bipartite(readweb('mydata.web'))
2 web.lonames = ['plant1','plant2','plant3',...]
3 web.upnames = ['herb1','herb2','herb3',...]
```

bipy internally takes care of re-ordering the species names when the matrix is sorted (to reflect nestedness or modularity). In addition, the species names are used in some visualizations, and when generating species-level metrics summaries.

If you do not specify species names, then bipy will assign number to each of the species, starting with 0 for the first species in your file, and increasing.

2.5 Adding bibliographical references

When comparing several datasets, it may be interesting to know where the data come from. bipy is able to extract bibliographic information from several identifiers, and to generate a link to the original paper.

If you have a dataset loaded, you can add the original references and see them by doing:

Listing 2.6: attributing a reference

```
1 web.ref = ref({'doi':'10.1098/rsbl.2010.0774','pmid':'20961886'})
   })
2 output_citinfo(w)
```

This uses some NCBI functions to get the original data, so you need to be online to use it.

3 Species–level metrics

3.1 Degree

3.2 Specificity

`bipy` implements three measures of specificity, which are especially suited to the analysis of bipartite networks. SAY MORE.

3.2.1 Specificity

Specificity is measured with the “paired-differences index” (PDI, Poisot et al.).

3.2.2 Species-specialization index

The “species-specialization index” of Julliard et al. (2006) is implemented. When using the `bipartite` class, the values of `ssi` for all species are available in the `ssi` attribute.

3.2.3 Resource–range

4 Network-level metrics

4.1 Nestedness

The measure of nestedness used is `NODF`, as described by Almeida-Neto et al. (2008). This measure has the advantage of being robust to most of the variations in network structure (shape and size), and displays a consistent behaviour with regard to connectance (namely, a linear increase). In addition, it is possible to separate the nestedness of the whole network, and of each of the trophic levels.

`NODF` can be calculated using the `NNNNNNNN` function.

When using the `bipartite` class, the values of nestedness can be accessed with `.nodf`, `.nodf_up`, and `.nodf_low`. These attributes are also used by the `mini_bipartite` class.

4.2 Modularity

4.3 Robustness to extinction

The extinction robustness analysis used notably by Memmott et al. (2004) are implemented in `bipy`. Three scenarios of extinction are available: best case (increasing degree, species with a low degree go first), random, and worst case (decreasing degree, highly-connected species go first).

5 Null models

bipy proposes (as of now) three main classes of different null models. This chapter presents the way they work and the specificity of each. The purpose of a null model is to use some features of the original network to create a given number of random networks. All null models implemented in bipy share the property that the networks they return will have the exact same size than the original network. A consequence of this is that there are no species with no interactions – for this reason, networks with a low connectance cannot be meaningfully randomized, and bipy will issue a warning. For networks which can be randomized, but still have a low connectance, the null model will likely require a large number of iterations to find a solution with all rows and columns with at least one interaction, so the analysis might take a long time.

5.1 The wrapper for null models

5.2 Null models

The common feature of the null models that are currently implemented is that they will, for each interaction, generate a probability that it is realized, and then draw at random whether it happens or not using a binomial process. For the sake of notational simplicity, we call the probability that $M_{ij} = 1$ (i.e. species i and j are interacting) P_{ij} . It should be noted that both null models produce adjacency matrices, i.e. filled only with 0s and 1s.

5.2.1 Null model 1 – Bernoulli fill (null_1)

The first null model was proposed by Fortuna and Bascompte (2006). It accounts for the connectance Co of the network, such that for each interaction, $P_{ij} = Co$. As such, this null model only accounts for the connectance of the original network.

5.2.2 Null model 2 – interaction average (null_2)

The second null model was proposed by Bascompte et al. (2003). For each species n , we calculate its range as being the proportion of the organisms with which it interacts. Let R_n be this range.

For each interaction, the probability of it happening at random is

$$P_{ij} = \frac{1}{2} (R_i + R_j) \quad (5.1)$$

5.2.3 Null model 3 – row averages (null_3row)

5.2.4 Null model 3 – column averages (null_3col)

5.3 Tests of the deviation of common measures

The general approach for testing the deviation of the structure from a null expectation follows the same procedure. First, we measure N the metric of interest (e.g. modularity or nestedness), then we conduct a given number of replicates of one of the four proposed null models (see section 5.2), and measure N' (the same metric on the random network) for each. The deviation between the original structure and its random expectation is given by $N - N'$, whose mean should be 0 if the structure is random. To assess any deviation from the random expectation, we use a one-sample two-tailed t-test.

The

6 Graphics

bipy comes with several ways to visualize the data, which are meant to reflect the properties of the networks. There are two broad ways to visualize networks: as a matrix, and as “balls and strings”. The visualizations relies on the *Pyx* library, because it produces high quality (publication ready) graphics. This means that the *Pyx* library, along with *TEX*, need to be installed on your system.

In order to simplify the use of bipy, all of the plotting is handled by the function `plotWeb`. The number of arguments you can pass to is varies as a function of wether you want your visualization to reflect the nested or modular properties of the data.

6.1 Nested webs

6.2 Modular webs

In order to plot a network so as to reflect its modularity, you need to pass a second argument to `plotWeb`:

Listing 6.1: plotting a modular network

```
1 web = loadweb('mydata.web')
2 mod = findModules(mod,1000)
3 plotWeb(web,mod,filename='mydata-modular')
```

Note that when plotting a modular network, within each module, the species will be sorted by degree (so within each module, the visualization reflects the nestedness).

7 Example scripts

Bibliography

- Almeida-Neto, M., P. Guimaraes, P. Guimaraes Jr, R. Loyola, and W. Ulrich. 2008. A consistent metric for nestedness analysis in ecological systems: reconciling concept and measurement. *Oikos* 117:1227–1239.
- Bascompte, J., P. Jordano, C. J. Melián, and J. M. Olesen. 2003. The nested assembly of plant–animal mutualistic networks. *Proceedings of the National Academy of Sciences* 100:9383–9387.
- Fortuna, M. A., and J. Bascompte. 2006. Habitat loss and the structure of plant-animal mutualistic networks. *Ecology Letters* 9:281–286.
- Julliard, R., J. Clavel, V. Devictor, F. Jiguet, and D. Couvet. 2006. Spatial segregation of specialists and generalists in bird communities. *Ecology Letters* 9:1237–1244.
- Memmott, J., N. M. Waser, and M. V. Price. 2004. Tolerance of pollination networks to species extinctions. *Proceedings of the Royal Society B: Biological Sciences* 271:2605–2611.
- Poisot, T., E. Canard, N. Mouquet, and M. E. Hochberg. 2012. A comparative study of ecological specialization estimators. *Methods in Ecology and Evolution* .