

mangal – making complex ecological network analysis simpler

T. Poisot, B. Baiser, J. Dunne, S. Kéfi, F. Massol, S. Wood, D. Gravel

Jan. 2014

1 The study of ecological networks is severely limited by (i) the difficulty to access
2 data, (ii) the lack of a standardized way to link meta-data with interactions, and (iii)
3 the disparity of formats in which ecological networks themselves are represented.
4 To overcome these limitations, we conceived a data specification for ecological net-
5 works. We implemented a database respecting this standard, and released a R
6 package (`rmangal`) allowing users to programmatically access, curate, and deposit
7 data on ecological interactions. In this article, we show how these tools, in conjunc-
8 tions with other frameworks for the programmatic manipulation of open ecological
9 data, streamlines the analysis process, and improves replicability and reproducibil-
10 ity of ecological networks studies.

11 `## Error: impossible de trouver la fonction "as"`

12 Introduction

13 Ecological networks enable ecologists to accommodate the complexity of natural communities,
14 and to discover mechanisms contributing to their persistence, stability, resilience, and function-
15 ing. Most of the “early” studies of ecological networks were focused on understanding how
16 the structure of interactions within one location affected the ecological properties of this local
17 community. This led to classic results, such as the buffering impact of modularity on species

1 loss (Stouffer & Bascompte 2011), the increase in robustness to extinctions along with increases
2 in connectance (Dunne *et al.* 2002), and the fact that organisation of interactions maximize bio-
3 diversity (Bastolla *et al.* 2009). More recently, new studies introduced the idea that different
4 networks can be meaningfully compared, either to understand the importance of environ-
5 mental gradients on the realisation of ecological interactions [tylianakis_habitat_2007], or to
6 understand the mechanisms behind variation in the structure of ecological networks (Poisot
7 *et al.* 2012). Yet, meta-analyses of a large number of ecological networks are still extremely
8 rare, and most of the studies comparing several networks do so within the limit of particular
9 systems (Schleuning *et al.* 2011; Dalsgaard *et al.* 2013). In part, this can be attributed to the
10 limited methods allowing to compare networks in which no species are in common. However,
11 the severe shortage of data in the field also restricts the power of large-scale analyses.

12 An increasing number of approaches are being put forth to *predict* the structure of ecological
13 networks, either relying on latent variables (Rohr *et al.* 2010) or actual traits (Gravel *et al.* 2013).
14 These approaches, so as to be adequately calibrated, require easily accessible data. Compar-
15 ing the efficiency of different methods is also facilitated if there is an homogeneous way of
16 representing ecological interactions, and the associated metadata. In this paper, we (i) estab-
17 lish the need of a data specification serving as a *lingua franca* among network ecologists, (ii)
18 describe this data specification. Finally, we (iii) describe `rmangal`, a R package and `compagnon`
19 database, relying on this data specification. We provide some use cases showing how this
20 new approach makes complex analyzes simpler, and allows for the integration of new tools to
21 manipulate biodiversity resources.

22 **Networks need a data specification**

23 Ecological networks are (often) stored as their *adjacency matrix* (or as the quantitative link
24 matrix), that is a series of 0 and 1 indicating, respectively, the absence and presence of an
25 interaction. This format is extremely convenient for *use* (as most network analysis packages,
26 *e.g.* `bipartite`, `betalink`, `foodweb`, require data to be presented this way), but is extremely

1 inefficient at *storing* meta-data. In most cases, an adjacency matrix informs on the identity of
2 species (in cases where rows and columns headers are present), and the presence or absence
3 of interactions. If other data about the environment (*e.g.* where the network was sampled) or
4 the species (*e.g.* the population size, trait distribution, or other observations) are available,
5 they are most either given in other files, or as accompanying text. In both cases, making
6 a programmatic link between interaction data and relevant meta-data is difficult and error-
7 prone.

8 By contrast, a data specification provides a common language for network ecologists to inter-
9 act, and ensure that, regardless of their source, data can be used in a shared workflow. Most
10 importantly, a data specification describes how data are *exchanged*. Each group retains the
11 ability to store the data in the format that is most convenient for in-house use, and only needs
12 to provide export options (*e.g.* through an API, that is a programmatic interface returning ob-
13 jects in response to pre-determined requests) respecting the data specification. This approach
14 ensures that *all* data can be used in meta-analyses, and increases the impact of data (Piwowar
15 *et al.* 2007; Piwowar & Vision 2013).

16 **Elements of the data specification**

17 The data specification (Fig. XX) is built around the idea that (ecological) networks are col-
18 lections of relationships between ecological objects, each element having particular meta-data
19 associated. In this section, we detail highlight the way networks are represented in the mangal
20 specification. An interactive webpage with the elements of the data specification can be found
21 online at <http://mangal.uqar.ca/doc/spec/>. The data specification is available either at the
22 API root (*e.g.* <http://mangal.uqar.ca/api/v1/?format=json>), or can be viewed using the
23 `whatIs` function from the R package (see *Supp. Mat. 1*). Rather than giving an exhaustive list
24 of the data specification (which is available online at the aforementioned URL), this section
25 serves as an overview of each element, and how they interact.

26 We propose JSON, a format equivalent to XML, as an efficient way to uniformize data repre-

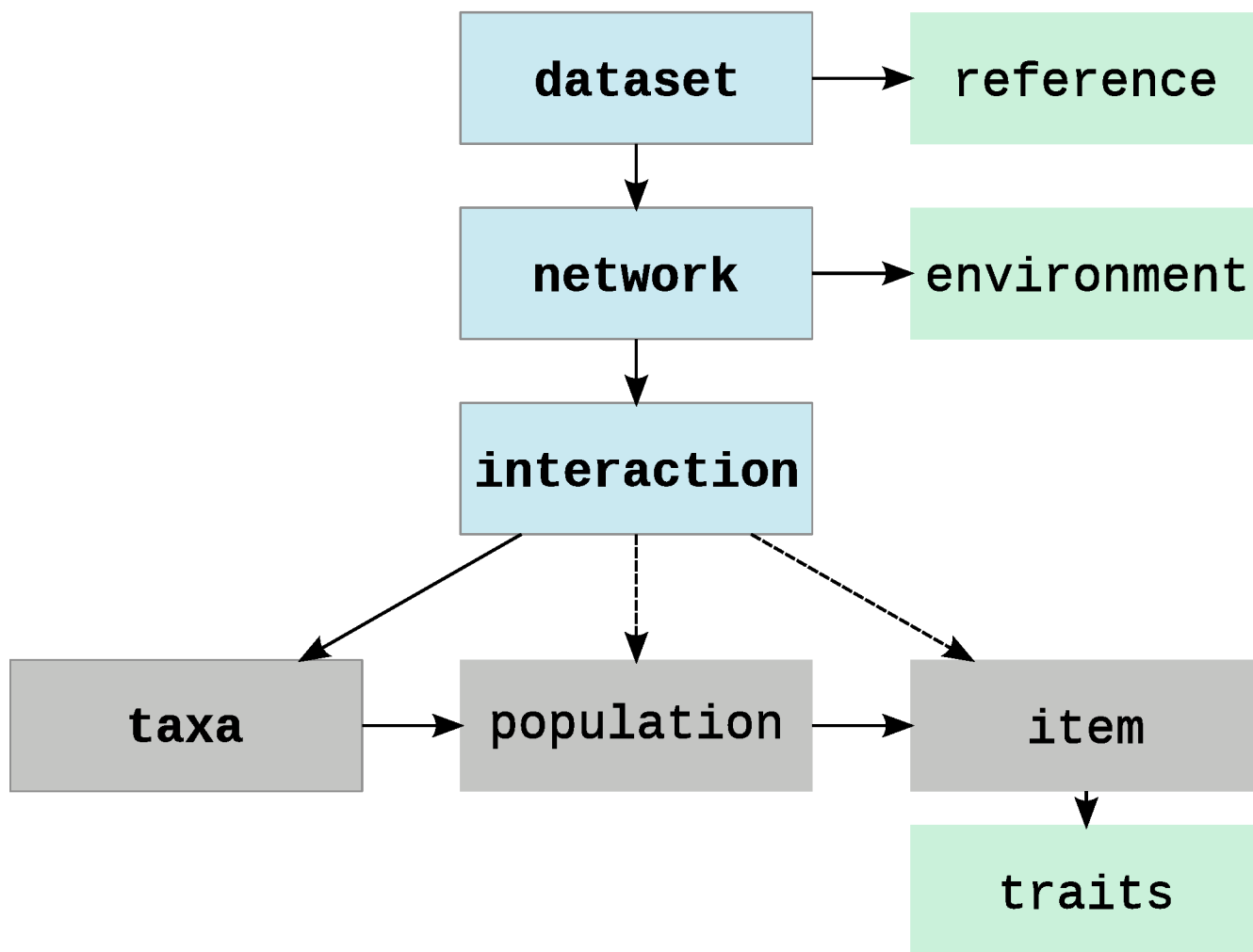


Figure 1: An overview of the data specification, and the hierarchy between objects. Each box correspond to a level of the data specification. Grey boxes are nodes, blue boxes are interactions and networks, and green boxes are metadata. The **bold** boxes (dataset, network, interaction, taxa) are the minimal elements needed to represent a network.

1 sentation for two main reasons. First, it has emerged as a *de facto* standard for web platform
2 serving data, and accepting data from users. Second, it allows *validation* of the data: a JSON
3 file can be matched against a scheme, and one can verify that it is correctly formatted (this
4 includes the possibility that not all fields are filled, as will depend on available data). Finally,
5 JSON objects are easily and cheaply (memory-wise) parsed in the most common programming
6 languages, notably R (equivalent to `list`) and python (equivalent to `dict`). For most users, the
7 format in which data are transmitted is unimportant, as the interaction happens within R – as
8 such, knowing how JSON objects are organised is only useful for those who want to interact
9 with the API directly.

10 **Node informations**

11 **Taxa**

12 Taxa are a taxonomic entity of any level, identified by their name, vernacular name, and
13 their identifiers in a variety of taxonomic services. Associating the identifiers of each taxa is
14 important to leverage the power of the new generation of open data tools, such as `taxize`
15 [`@chamberlain_taxize:_2013`]. The data specification currently has fields for `ncbi`, `gbif`, `itis`,
16 `eol` and `bold` identifiers.

17 **Population**

18 A population is one observed instance of a taxa object. If your experimental design is repli-
19 cated through space, then each taxa have a population object corresponding to each locality.
20 Populations do not have associated meta-data, but serve as “containers” for `item` objects.

21 **Item**

22 An `item` is an instance of a population. Items have a `level` argument, which can be either
23 `individual` or `population`; this allows to represent both individual-level networks (*i.e.* there

1 are as many items attached to a population than there were individuals of this population
2 sampled), and population-level networks. When item represents a population, it is possible
3 to give a measure of the size of this population. The notion of item is particularly useful
4 for time-replicated designs: each observation of a population at a time-point is an item with
5 associated trait values, and possibly population size.

6 **Network informations**

7 **Interaction**

8 An interaction links, *a minima*, two taxa objects (but can also link pairs of populations or
9 items). The most important attributes of interactions are the type of interaction (of which
10 we provide a list of possible values, see *Supp. Mat. 1*), and its nature, *i.e.* how it was observed.
11 This field help differentiate direct observations, text mining, and inference. Note that the
12 nature field can also take absence as a value; this is useful for, *e.g.*, “cafeteria” experiments in
13 which there is high confidence that the interaction did not happen.

14 **Network**

15 A network is a series of interaction object, along with (i) informations on its spatial posi-
16 tion (provided at the latitude and longitude), (ii) the date of sampling, and (iii) references to
17 measures of environmental conditions.

18 **Dataset**

19 A dataset is a collection of one or several network(s). Datasets also have a field for data
20 and papers, both of which are references to bibliographic or web resources describing, respec-
21 tively, the source of the data, and the papers in which these data have been significantly used.
22 Datasets are the preferred entry point in the resources.

1 **Meta-data**

2 **Trait value**

3 Objects of type `item` can have associated `trait` values. These consist in the description of the
4 trait being measured, the value, and the units in which the measure was taken.

5 **Environmental condition**

6 Environmental conditions are associated to `network`. These are defined by the environmental
7 property measured, its value, and the units.

8 **References**

9 References are associated to datasets. They accomodate the DOI, JSON or PubMed identifiers,
10 or a URL. When possible, the DOI should be preferred as it offers more potential to interact
11 with other on-line tools, such as the *CrossRef* API.

12 **Use cases**

13 In this section, we present use cases using the `rmangal` package for R, to interact with a
14 database implementing this data specification, and serving data through an API (<http://mangal.uqar.ca>).
15 It is possible for users to deposit data into this database, through the R package. Data are made
16 available under a *CC-0 Waiver* (Poisot *et al.* 2013). Detailed informations about how to upload
17 data are given in the vignettes and manual of the `rmangal` package. So as to save room in the
18 manuscript, we source each example; the complete `r` files to reproduce the examples of this
19 section are attached as *Suppl. Mat.*.

20 The data we use for this example come from Ricciardi *et al.* (2010). These were previously
21 available on the *InteractionWeb DataBase* as a single `xls` file. We uploaded them in the `mangal`
22 database at <http://mangal.uqar.ca/api/v1/dataset/{todo}>.

1 Link-species relationships

2 In the first example, we visualize the relationship between the number of species and the
3 number of interactions, which Martinez (1992) propose to be linear (in food webs).

```
source("usecases/1_ls.r")
```

4 Producing this figure requires less than 10 lines of code. The only information needed is
5 the identifier of the network or dataset, which we suggest should be reported in publications
6 as: "These data were deposited in the mangal format at <URL>/api/v1/dataset/<ID>". This
7 encourages data re-use.

8 Network beta-diversity

9 In the second example, we use the framework of network β -diversity (Poisot *et al.* 2012) to
10 measure the extent to which networks that are far apart in space have different interactions.
11 Each network in the dataset has a latitude and longitude, meaning that it is possible to measure
12 the geographic distance between two networks.

13 For each pair of network, we measure the geographic distance (in km.), the species dissimilar-
14 ity (β_S), the network dissimilarity when all species are present (β_{WN}), and finally, the network
15 dissimilarity when only shared species are considered (β_{OS}).

```
source("usecases/2_beta.r")
```

16 As shown in Fig. XX, while species dissimilarity and overall network dissimilarity increase
17 when two networks are far apart, this is not the case for the way common species interact.
18 This suggests that in this system, network dissimilarity over space is primarily driven by
19 species turnover. The ease to gather both raw interaction data and associated meta-data make
20 producing this analysis extremely straightforward.

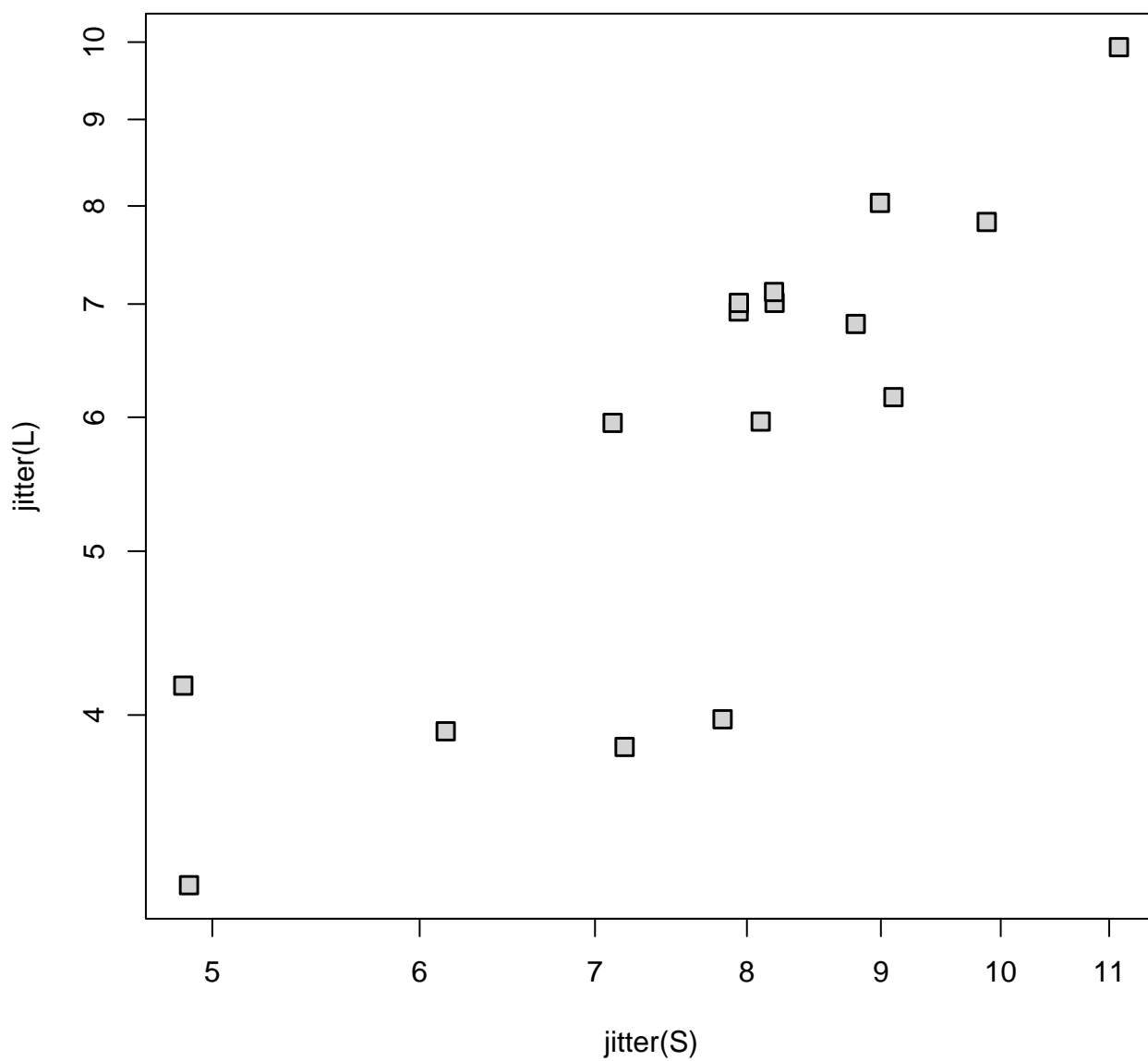


Figure 2: Relationship between the number of species and number of interactions in the anemonefish-fish dataset.

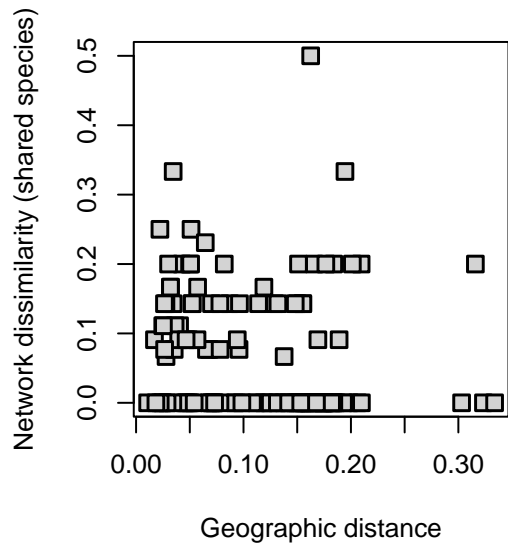
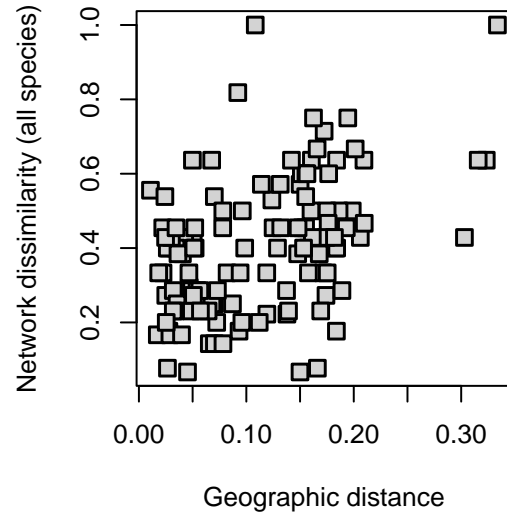
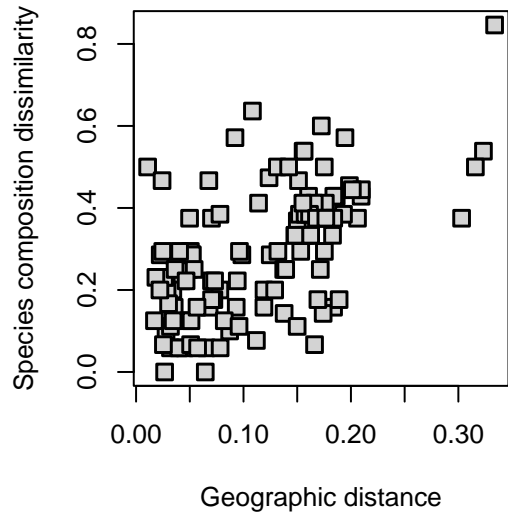


Figure 3: Relationships between the geographic distance between two sites, and the species dissimilarity, network dissimilarity with all, and only shared, species.

1 Spatial visualisation of networks

2 Bascompte (2009) proposes an interesting visualisation for spatialized networks, in which each
3 species is laid out on a map at the center of mass of its area of occurrence; interactions are then
4 drawn between species, to show how species distribution determines biotic interactions. In
5 this final use case, we propose to reproduce a similar figure, using the RgoogleMaps package.

```
source("usecases/3_spatial.r")
```

6 Conclusions

7 In this contribution, we presented mangal, a data format for the exchange of ecological net-
8 works and associated meta-data. We deployed an online database with an associated API,
9 relying on this data specification. Finally, we introduced rmangal, a R package designed to
10 interact with APIs using the mangal format. We expect that the data specification will evolve
11 based on the needs of the community. At the moment, users are welcome to propose such
12 changes on the project issue page: <https://github.com/mangal-wg/mangal/issues>.

13 References

- 14 Bascompte, J. (2009). Disentangling the web of life. *Science (New York, N.Y.)*, **325**, 416–9.
- 15 Bastolla, U., Fortuna, M.A., Pascual-García, A., Ferrera, A., Luque, B. & Bascompte, J. (2009).
16 The architecture of mutualistic networks minimizes competition and increases biodiversity.
17 *Nature*, **458**, 1018–1020.
- 18 Dalsgaard, B., Trøjelsgaard, K., González, A.M.M., Nogués-Bravo, D., Ollerton, J., Petanidou,
19 T., Sandel, B., Schleuning, M., Wang, Z., Rahbek, C., Sutherland, W.J., Svenning, J.-C. & Olesen,
20 J.M. (2013). Historical climate-change influences modularity and nestedness of pollination
21 networks. *Ecography*, no–no. Retrieved May 14, 2013,

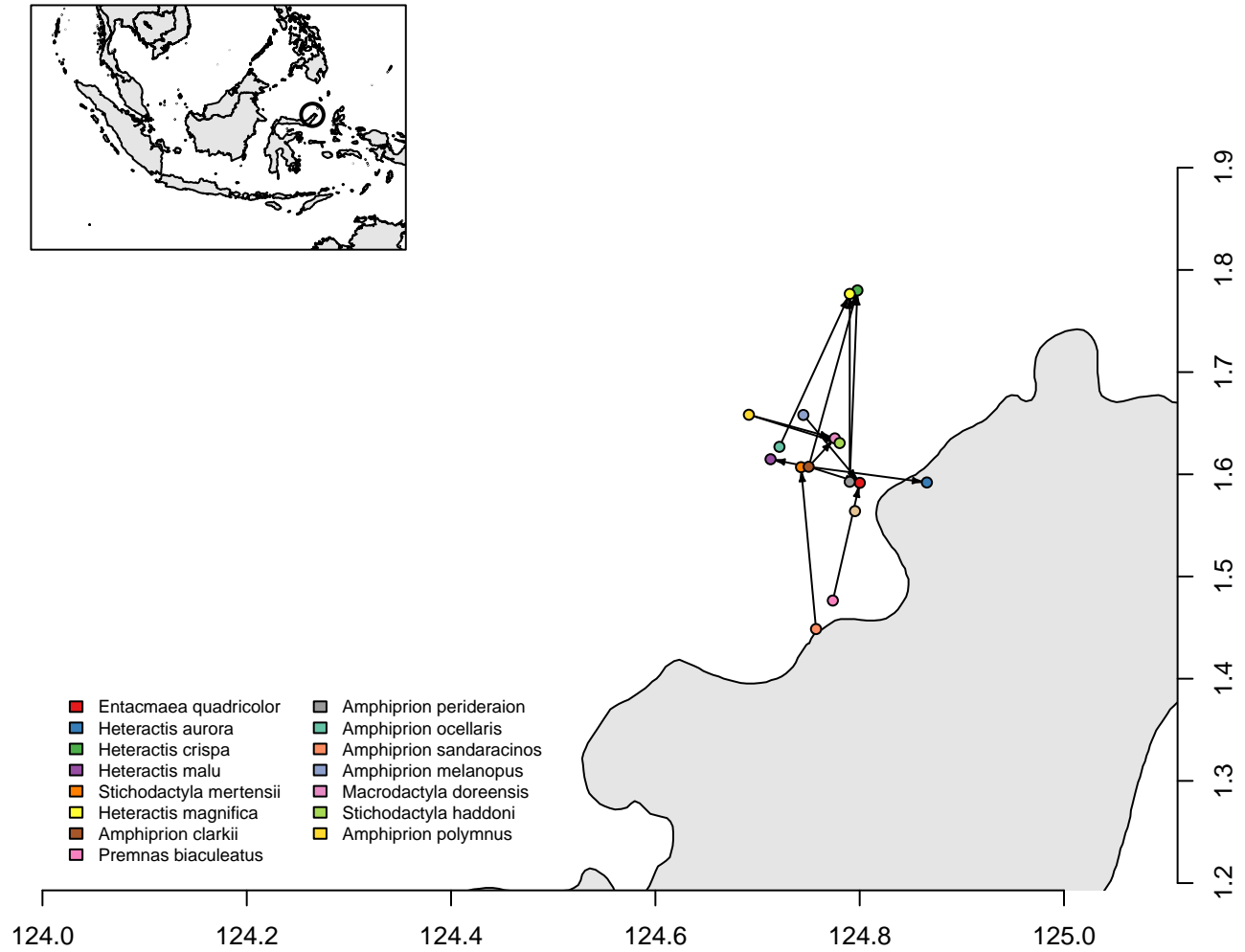


Figure 4: Spatial plot of a network, using the maps and rmangal packages. The circle in the inset map show the location of the sites. Each dot in the main map represents a species, with interactions drawn between them.

- 1 Dunne, J.A., Williams, R.J. & Martinez, N.D. (2002). Network structure and biodiversity loss
2 in food webs: robustness increases with connectance. *Ecology Letters*, **5**, 558–567.
- 3 Gravel, D., Poisot, T., Albouy, C., Velez, L. & Mouillot, D. (2013). Inferring food web structure
4 from predator-prey body size relationships. *Methods in Ecology and Evolution*.
- 5 Martinez, N.D. (1992). Constant connectance in community food webs. *The American Natural-*
6 *ist*, **139**, 1208–1218.
- 7 Piwowar, H.A. & Vision, T.J. (2013). Data reuse and the open data citation advantage. *PeerJ*, **1**.
8 Retrieved October 05, 2013,
- 9 Piwowar, H.A., Day, R.S. & Fridsma, D.B. (2007). Sharing detailed research data is associated
10 with increased citation rate. (J. Ioannidis, Ed.). *PloS one*, **2**, e308.
- 11 Poisot, T., Canard, E., Mouillot, D., Mouquet, N. & Gravel, D. (2012). The dissimilarity of
12 species interaction networks. *Ecology Letters*, **15**, 1353–1361.
- 13 Poisot, T., Mounce, R. & Gravel, D. (2013). Moving toward a sustainable ecological science:
14 don't let data go to waste!
- 15 Ricciardi, F., Boyer, M. & Ollerton, J. (2010). Assemblage and interaction structure of the
16 anemonefish-anemone mutualism across the Manado region of Sulawesi, Indonesia. *Environ-*
17 *mental Biology of Fishes*, **87**, 333–347. Retrieved January 10, 2014,
- 18 Rohr, R.P., Scherer, H., Kehrli, P., Mazza, C. & Bersier, L.-F. (2010). Modeling food webs:
19 exploring unexplained structure using latent traits. *The American naturalist*, **176**, 170–7.
- 20 Schleuning, M., Blüthgen, N., Flörchinger, M., Braun, J., Schaefer, H.M. & Böhning-Gaese,
21 K. (2011). Specialization and interaction strength in a tropical plant-frugivore network differ
22 among forest strata. *Ecology*, **92**, 26–36.
- 23 Stouffer, D.B. & Bascompte, J. (2011). Compartmentalization increases food-web persistence.
24 *Proceedings of the National Academy of Sciences of the United States of America*, **108**, 3648–3652.