Environmental biases in the study of ecological networks at the planetary scale

Introduction

Ecological networks are a useful way to think about ecological systems in which species or organisms interact (**???**; Poisot et al. 2016c; Delmas et al. 2018), and recently there has been an explosion of interest in their dynamics across large temporal scales (Tylianakis & Morris 2017; Baiser et al. 2019), especially alongside environmental gradients (Trøjelsgaard & Olesen 2016; Pellissier et al. 2017). As ecosystems and climates are changing rapidly, ecologists realized that networks are at risk or unravelling, being invaded by exotic species that can destabilize them (Strong & Leroux 2014; Magrach et al. 2017), or adopt entirely novel configurations (Guiden et al. 2019; Hui & Richardson 2019). Simulation studies seem to suggest that knowing the shape of the extant network is not sufficient (Thompson & Gonzalez 2017), and that it needs to be supplemented by additional data on species properties, climate, and climate projection.

This renewed interest in ecological networks has prompted several methodological efforts. First, an expansion of the analytical tools to study ecological networks and their variation in space (Poisot et al. 2012, 2015, 2017). Second, improvements in large-scale data-collection, through increased adoption of molecular biology tools (Evans et al. 2016; Eitzinger et al. 2019) and crowdsourcing of data collection (Pocock et al. 2015; Bahlai & Landis 2016; Roy et al. 2016). Finally, a surge in the development of tools that allow to *infer* species interaction (Morales-Castilla et al. 2015) based on limited but complementary data on existing network properties (Stock et al. 2017), species traits (Gravel et al. 2013; Desjardins-Proulx et al. 2017; Brousseau et al. 2017; Bartomeus et al. 2016), and environmental conditions (Gravel et al. 2018). These approaches tend to perform well in data-poor environments (Beauchesne et al. 2016), and can be combined through ensemble modeling or model averaging to generate possibly more robust predictions (Pomeranz et al. 2018). The later task of inferring interactions is particularly important, knowing that ecological networks are difficult to adequately sample in nature (Banašek-Richter et al. 2004; Gibson et al. 2011; Chacoff et al. 2012; Jordano 2016a).

The common goal to these efforts is to facilitate the prediction of network structure, particularly over space, both extant [Poisot et al. (2016b); Gravel et al. (2018); **MARINE FOODWEB**] and future (Albouy et al. 2014), and to appraise its possible variation in response to environmental changes. All of these developments share the need to be supported by state of the art data management: novel quantitative tools demand a higher volume of network data; novel collection techniques demand powerful data repositories; novel inference tools demand easier integration between different types of data, including but not limited to interactions, species traits, taxonomy, occurrences, and local bioclimatic conditions. In short, advancing the science of ecological networks requires not only to increase the volume of available data, but to pair it with ecologically relevant metadata, in a way that facilitates programmatic interaction so that they can be consumed by data analysis pipelines.

Borrett et al. (2014) identified network ecology as one of the fastest growing sub-field in the ecological sciences.

Synthesizing ecological data presents important challenges and also some exciting opportunities. Mangal is well suited to offer such opportunities in the study of ecological networks.

- A major challenge to ecological synthesis is generalizing from samples to the behaviour of ecological systems
- two obstacles to such generalizing in ecological systems: data coverage and data quality
  - data coverage: are data collected from every relevant system?
  - data quality: are data fit-for-purpose? Two particular aspects of quality
    - taxonomic resolution
    - sampling effort

Main question, is the data fit for purpose, what can we do and cannot do with it?

Poisot et al. (2016a) introduced `mangal.io` as a first step in this direction; in the years since the tool was originally published, we continued development of the data representation, amount and richness of metadata, and digitized and standardized as much ecological data as we could find. The second major release of this database contains over 1300 networks, 120000 interactions across close to 7000 taxa, and represents what is to our best knowledge the most complete collection of species interactions available. We seek to assess the fitness for purpose of ecological networks at the global scale to support synthesis research at large scales. Based on temporal and spatial biases in the description of some types of interactions, we conclude that while there is an increasing amount of available data, most of the planet's surface is poorly described. In particular, Africa, South America, and most of Asia have very sparse coverage. This suggests that the accuracy of synthesis efforts on the worldwide structure or properties of ecological networks will have very low predictive values within these areas, and reinforces the need to digitize available information, but also prioritize sampling towards these locations.

Global trends in ecological networks description

Network coverage is accelerating but spatially biased

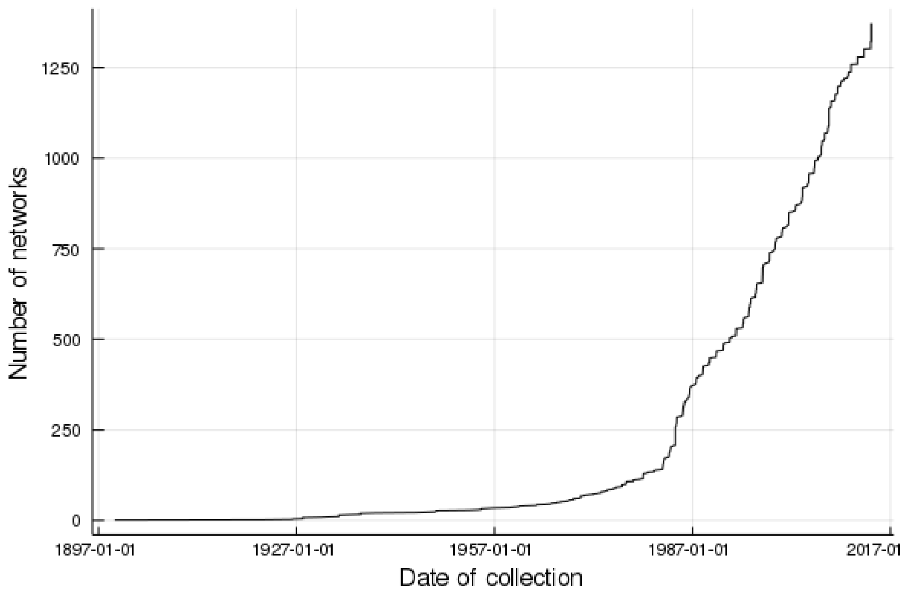

Figure 1: Cumulative number of ecological networks available in `mangal.io` as a function of the date of collection. About 1000 unique networks have been collected between 1987 and 2017, a rate of just over 30 networks a year.

The earliest recorded ecological networks date back to the late nineteenth century, with a strong increase in the rate of collection around the 1980s (fig. 1). Although the volume of available networks has increased over time, the sampling of these networks in space has been uneven. In fig. 2, we show that globally, network collection is biased towards the Northern hemisphere, and than different types of interactions have been sampled in different places. As such, it is very difficult to find a spatial area of sufficiently large size in which we have networks of predation, parasitism, and mutualism. The inter-tropical zone is particularly data-poor, either because data producers from the global South correctly perceive massive re-use of their data by Western world scientists as a form of scientific neo-colonialism (as advanced by Mauthner & Parry 2013), thereby providing a powerful incentive *against* their publication, or because ecological networks are subject to the same data deficit that is affecting all fields on ecology in the tropics (Collen et al. 2008). As Bruna (2010) identified almost ten years ago, improved data deposition requires an infrastructure to ensure they can be repurposed for future research, which we argue is provided by `mangal.io` for ecological interactions.

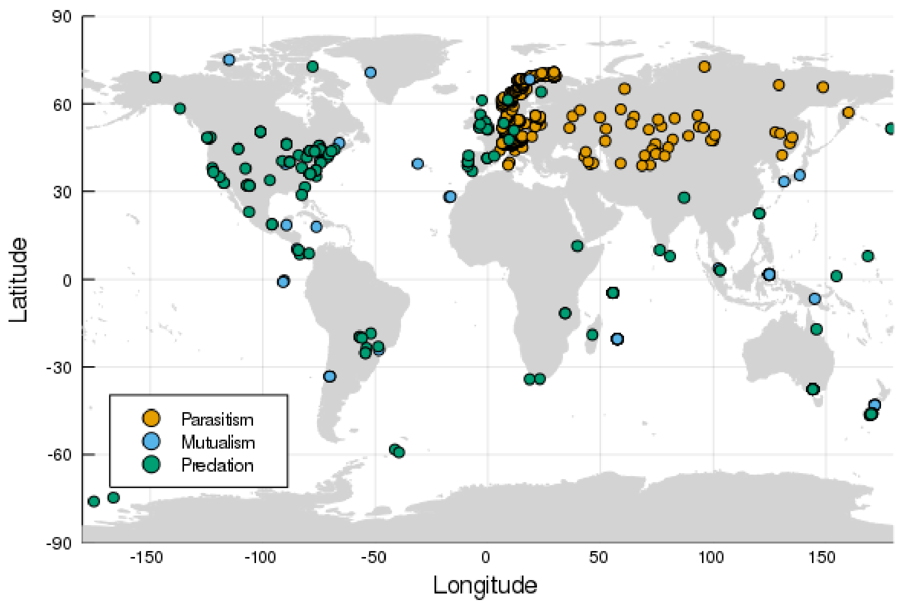

Figure 2: Each point on the map corresponds to a network with parasitic, mutualistic, and predatory interactions. It is noteworty that the spatial coverage of these types of interactions is uneven; the Americas have almost no parasitic network, for example. Some places have barely been studied at all, including Africa and Eastern Asia. This concentration of networks around rich countries speaks to an inadequate coverage of the diversity of landscapes on Earth.

Different interaction types have been studied in different biomes

Whittaker (1962) suggested that natural communities can be partitioned across biomes, largely defined as a function of their relative precipitation and temperature; in fig. 3, we show that even though networks, overall, capture the diversity of the precipitation/temperature climate well, types of networks have been studied in sub-spaces only. Specifically, parasitism networks have been studied in colder and drier climates; mutualism networks in wetter climates; predation networks display less of a bias.

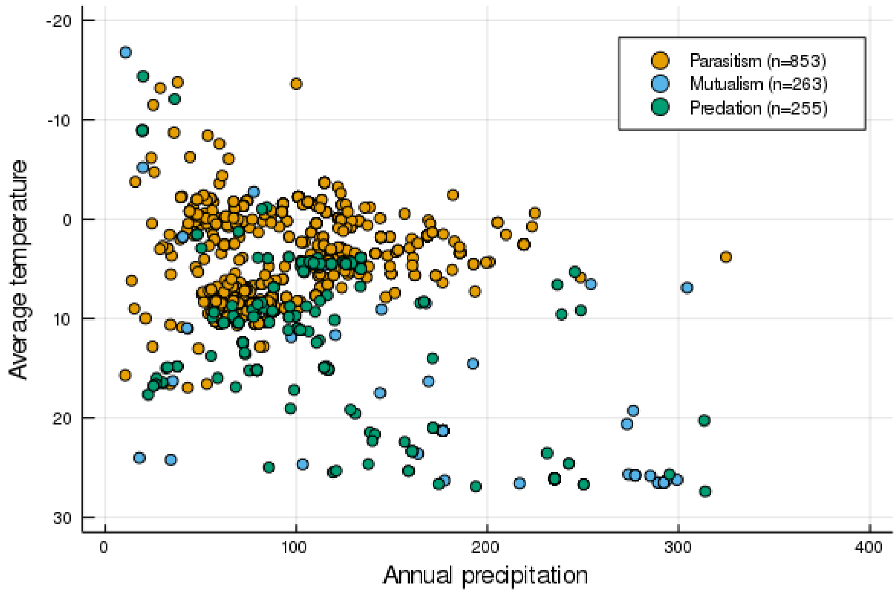

Figure 3: List of networks across biomes

Some locations on Earth have no climate analogue

Climate analogue

Mutualistic networks are biased towards more unique environments

Conclusions

Data quality: sampling effort and taxonomy

Jordano (2016b) – importance of taxonomic resolution

Sampling effort and taxonomic detail are two very challenging but important part of any ecological dataset. The datasets in Mangal represent some of the most detailed studies of ecological networks available. * measures of network structure may be particularly sensitive to the amount of sampling effort * repeat sampling may be necessary to capture a "saturation" of interactions. * we present some visualization of the sampling coverage of Mangal [tk] * High taxonomic resolution is difficult to achieve in ecology, especially depending on the sampling method used (e.g. gut contents vs observations). We present a

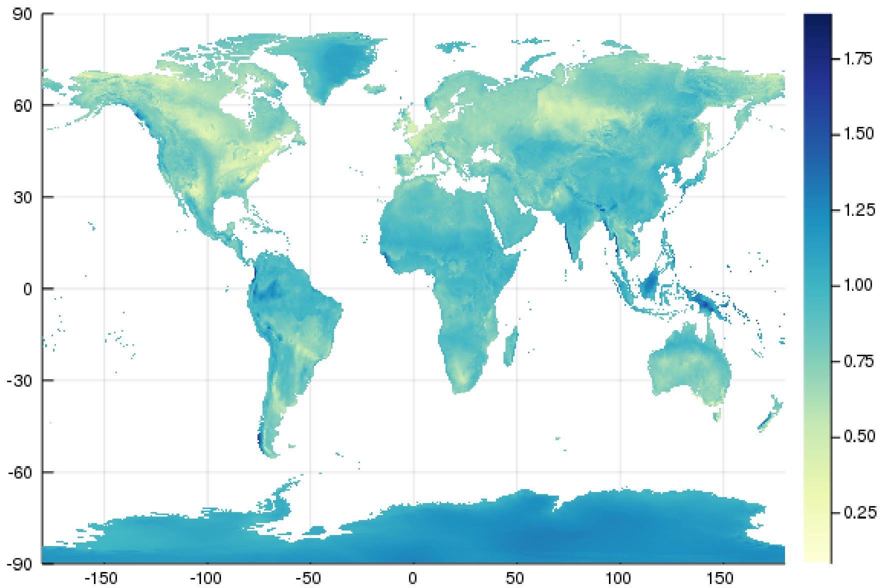

Figure 4: tk

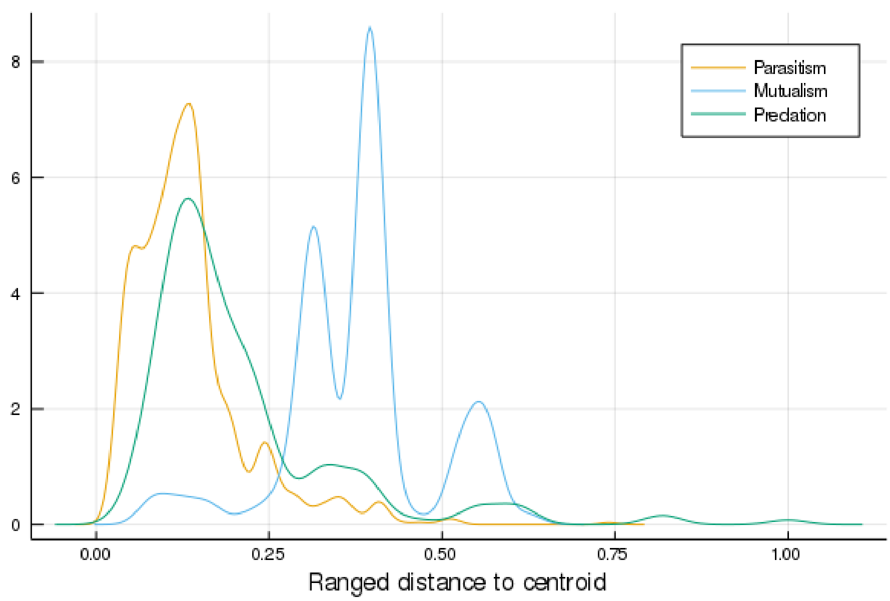

Figure 5: tk

breakdown of the taxonomic resolution of Mangal. * Ecological networks occur in various kinds, but they are not all equally well sampled. We present a breakdown of the number of parasitic, mutualistic and predator-prey networks sampled in Mangal

Can we predict the future of ecological networks under climate change?

Perhaps unsurprisingly, most of our knowledge on ecological networks is derived from data that were collected after the 1990s (fig. 1). This means that we have worryingly little information on ecological networks dating before the acceleration of the climate crisis, and therefore lack a robust baseline. Dalsgaard et al. (2013) provide strong evidence that the extant shape of ecological networks emerged in part in response to historical trends in climate change; in this perspective, the lack of reference data before the acceleration of the effects of climate change is of particular concern, as we may be deriving intuitions on ecological networks structure and assembly rule from networks that are in the midst of important ecological disturbances. Although there are some research on the response of co-occurrence and indirect interactions to climate change (Araújo et al. 2011; Losapio & Schöb 2017), these are a far cry from actual direct interactions; similarly, the data on "paleo-foodwebs", *i.e.* from deep evolutionary time (Nenzén et al. 2014; Yeakel et al. 2014; Muscente et al. 2018) represent the effect of more progressive change, and may not adequately inform us about the future of ecological networks under severe climate change.

Possibly more concerning is the fact that the spatial distribution of sampled networks shows a clear bias towards the Western world, specifically the Atlantic coast of the USA and Canada, and Western Europe (fig. 2). While this can to some degree be circumvented by working on the networks sampled in places that are close analogues to regions without direct information (almost all of Africa, most of South America, a large part of Asia), 4 suggests that this approach will rapidly be limited: the diversity of bioclimatic combinations on Earth leaves us with some areas lacking suitable analogues. These regions are expected to bear the worse of the socio-economical (*e.g.* Indonesia) or ecological (*e.g.* polar regions) consequences of climate change. All things considered, our current knowledge about the structure of ecological networks at the global scale leaves us under-prepared to predict their response to a warming world.

What purpose are global ecological network data fit for?

This begs the question of *what* can be achieved with our current knowledge of ecological networks. **TK**

Active development and data contribution

This is an open-source project: all data and all code supporting this are available on the Mangal project GitHub organization. Our hope is that the success of this project will encourage similar efforts within other parts of the ecological community. In addition, we hope that this project will encourage the recognition of the contribution that software creators make to ecological research.

References

Albouy et al. (2014). From projected species distribution to food-web structure under climate change. *Global Change Biology.* 20:730–41.

Araújo et al. (2011). Using species co-occurrence networks to assess the impacts of climate change. *Ecography.* 34:897–908.

Bahlai & Landis. (2016). Predicting plant attractiveness to pollinators with passive crowdsourcing. *Royal Society Open Science.* 3:150677.

Baiser et al. (2019). Ecogeographical rules and the macroecology of food webs. *Global Ecology and Biogeography.* 0.

Banašek-Richter et al. (2004). Sampling effects and the robustness of quantitative and qualitative food-web descriptors. *J Theor Biol.* 226:23–32.

Bartomeus et al. (2016). A common framework for identifying linkage rules across different types of interactions. *Funct Ecol.* 30:1894–903.

Beauchesne et al. (2016). Thinking Outside the Box–predicting Biotic Interactions in Data-poor Environments. *Vie et milieu-life and enVironment.* 66:333–42.

Borrett et al. (2014). The rise of Network Ecology: Maps of the topic diversity and scientific collaboration. *Ecological Modelling.* 293:111–27.

Brousseau et al. (2017). Trait-matching and phylogeny as predictors of predator-prey interactions involving ground beetles. *Functional Ecology.*

Bruna. (2010). Scientific Journals can Advance Tropical Biology and Conservation by Requiring Data Archiving. *Biotropica.* 42:399–401.

Chacoff et al. (2012). Evaluating sampling completeness in a desert plant-pollinator network. *J Anim Ecol.* 81:190–200.

Collen et al. (2008). The Tropical Biodiversity Data Gap: Addressing Disparity in Global Monitoring. *Tropical Conservation Science.* 1:75–88.

Dalsgaard et al. (2013). Historical climate-change influences modularity and nestedness of pollination networks. *Ecography.* 36:1331–40.

Delmas et al. (2018). Analysing ecological networks of species interactions. *Biological Reviews.*:112540.

Desjardins-Proulx et al. (2017). Ecological interactions and the Netflix problem. *PeerJ.* 5.

Eitzinger et al. (2019). Assessing changes in arthropod predator–prey interactions through DNA-based gut content analysis—variable environment, stable diet. *Molecular Ecology.* 28:266–80.

Evans et al. (2016). Merging DNA metabarcoding and ecological network analysis to understand and build resilient terrestrial ecosystems. *Functional Ecology.*

Gibson et al. (2011). Sampling method influences the structure of plant–pollinator networks. *Oikos.* 120:822–31.

Gravel et al. (2018). Bringing Elton and Grinnell together: a quantitative framework to represent the biogeography of ecological interaction networks. *Ecography.* 0.

Gravel et al. (2013). Inferring food web structure from predator-prey body size relationships. Freckleton, ed. *Methods in Ecology and Evolution.* 4:1083–90.

Guiden et al. (2019). Predator–Prey Interactions in the Anthropocene: Reconciling Multiple Aspects of Novelty. *Trends in Ecology & Evolution.* 0.

Hui & Richardson. (2019). How to Invade an Ecological Network. *Trends in Ecology & Evolution.* 34:121–31.

Jordano. (2016a). Chasing Ecological Interactions. *PLOS Biol.* 14:e1002559.

Jordano. (2016b). Sampling networks of ecological interactions. *Functional Ecology.*

Losapio & Schöb. (2017). Resistance of plant–plant networks to biodiversity loss and secondary extinctions following simulated environmental changes. *Functional Ecology.* 31:1145–52.

Magrach et al. (2017). Plant-pollinator networks in semi-natural grasslands are resistant to the loss of pollinators during blooming of mass-flowering crops. *Ecography.*:n/a–a.

Mauthner & Parry. (2013). Open Access Digital Data Sharing: Principles, Policies and Practices. *Social Epistemology.* 27:47–67.

Morales-Castilla et al. (2015). Inferring biotic interactions from proxies. *Trends in Ecology & Evolution.*

Muscente et al. (2018). Quantifying ecological impacts of mass extinctions with network analysis of fossil communities. *PNAS.*:201719976.

Nenzén et al. (2014). The Impact of 850,000 Years of Climate Changes on the Structure and Dynamics of Mammal Food Webs. *PLOS ONE.* 9:e106651.

Pellissier et al. (2017). Comparing species interaction networks along environmental gradients. *Biol Rev Camb Philos Soc.*

Pocock et al. (2015). The Biological Records Centre: a pioneer of citizen science. *Biol J Linn Soc.* 115:475–93.

Poisot et al. (2016a). mangal - making ecological network analysis simple. *Ecography.* 39:384–90.

Poisot et al. (2012). The dissimilarity of species interaction networks. *Ecol Lett.* 15:1353–61.

Poisot et al. (2016b). Synthetic datasets and community tools for the rapid testing of ecological hypotheses. *Ecography.* 39:402–8.

Poisot et al. (2017). Hosts, parasites and their interactions respond to different climatic variables. *Global Ecol Biogeogr.*:n/a–a.

Poisot et al. (2015). Beyond species: why ecological interaction networks vary through space and time. *Oikos.* 124:243–51.

Poisot et al. (2016c). Describe, understand and predict: why do we need networks in ecology? *Funct Ecol.* 30:1878–82.

Pomeranz et al. (2018). Inferring predator-prey interactions in food webs. *Methods in Ecology and Evolution.* 0.

Roy et al. (2016). Focal Plant Observations as a Standardised Method for Pollinator Monitoring: Opportunities and Limitations for Mass Participation Citizen Science. *PLOS ONE.* 11:e0150794.

Stock et al. (2017). Linear filtering reveals false negatives in species interaction data. *Scientific Reports.* 7:45908.

Strong & Leroux. (2014). Impact of Non-Native Terrestrial Mammals on the Structure of the Terrestrial Mammal Food Web of Newfoundland, Canada. *PLOS ONE.* 9:e106264.

Thompson & Gonzalez. (2017). Dispersal governs the reorganization of ecological networks under environmental change. *Nature Ecology & Evolution.* 1:0162.

Trøjelsgaard & Olesen. (2016). Ecological networks in motion: micro- and macroscopic variability across scales. *Funct Ecol.* 30:1926–35.

Tylianakis & Morris. (2017). Ecological Networks Across Environmental Gradients. *Annual Review of Ecology, Evolution, and Systematics.* 48:25–48.

Whittaker. (1962). Classification of Natural Communities. *Botanical Review.* 28:1–239.

Yeakel et al. (2014). Collapse of an ecological network in Ancient Egypt. *PNAS*. 111:14472–7.