Received: 15 April 2021 | Accepted: 3 January 2022

DOI: 10.1111/2041-210X.13805

Methods in Ecology and Evolution BRITISH ECOLOGICAL SOCIETY

APPLICATION

Exploring noise, degeneracy and determinism in biological networks with the `einet` package

Brennan Klein[1,2] | Anshuman Swain[3] | Travis Byrum[3] | Samuel V. Scarpino[1,4,5,6] | William F. Fagan[3]

[1]Network Science Institute, Northeastern University, Boston, MA, USA; [2]Laboratory for the Modeling of Biological and Socio-Technical Systems, Northeastern University, Boston, MA, USA; [3]Department of Biology, University of Maryland, College Park, MD, USA; [4]Santa Fe Institute, Santa Fe, NM, USA; [5]Vermont Complex Systems Center, University of Vermont, Burlington, VT, USA and [6]Pandemic Prevention Institute, Rockefeller Foundation, Washington, DC, USA

Correspondence
Brennan Klein
Email: b.klein@northeastern.edu

Funding information
John Templeton Foundation, Grant/Award Number: 61780; National Science Foundation, Grant/Award Number: DGE-1632976

Handling Editor: Timothée Poisot

Abstract

1. Understanding noise in networks and finding the right scale to represent a system are important problems in network biology. Most research focuses on the raw, micro-scale network from data/simulations and seldom explores the scale dependence of properties.
2. Here, we introduce the `einet` package, which looks at the most informative scale in a biological network using recent concepts from information theory and network science.
3. `einet` uses two metrics: Effective information, which measures the interplay between degeneracy and determinism in a network's edges, and causal emergence, which finds the scale of the network with the highest effective information.
4. `einet` is available in R and Python and provides tools to explore noise and scale dependency in networks as well as compare information flow and noise across networks.

KEYWORDS
biological networks, causal emergence, community interactions, information theory

1 | INTRODUCTION

The functioning of biological systems depends heavily on the interaction of its constituents at various levels of organization, be it at the subcellular level or at the ecosystem scale. To understand the complexity of such interactions in a more abstract yet workable way, people have turned to networks as a useful representation (Gosak et al., 2018; Gysi & Nowick, 2020). The use of network science and its tools in biology has exploded in the past two decades, owing to the burst of high-resolution data and improved computational capabilities. The application of network science has improved our fundamental understanding of many biological systems and structures, such as food webs (Dunne et al., 2002; Pascual & Dunne, 2006; Shaw et al., 2021), neuronal functioning and signalling (Bassett & Sporns, 2017; Sporns, 2014), gene interactions and regulation (Costanzo et al., 2019), mutualistic interactions (Vázquez

Brennan Klein and Anshuman Swain contributed equally to this work.

et al., 2009), epidemiological networks (Wang et al., 2015), microbial interactions (Faust & Raes, 2012), and animal social networks (Pinter-Wollman et al., 2014).

Most network studies focus on the structure and/or the dynamics of the constructed network itself, and probe how its various properties can explain or predict certain biological activity. However, sometimes due to noise—either inherent to the system or introduced via measurement and observation—the structure and dynamics of networks can be affected by the way the networks are constructed (Tsimring, 2014). The kind of noise introduced through observation will often be specific to particular systems under study. Past work has focused on the estimation, and in rare cases, the correction of such noise (Newman, 2018; Freilich et al., 2020; Swain, Devereux, et al., 2021). Beyond such estimation, it can often be difficult to deal with the noise that is inherently present in complex biological systems regardless of whether the noise is from measurement or is inherent to the system in question.

Because of the high levels of noise and stochasticity associated with biological systems, and because such systems are often subjected to random fluctuations and perturbations, biological systems tend to be redundant and degenerate in their functioning. This has led to a lot of research on the resilience and robustness of biological systems through degeneracy and other similar features (Edelman & Gally, 2001; Tononi et al., 1999). These studies have changed our view about the function of degeneracy in such systems, emphasizing increased resilience to sudden perturbations, and functional equivalence of degenerate entities (see Ahn et al., 2006; Edelman & Gally, 2001; Klein et al., 2021). Research on biological networks often seeks to connect certain deterministic or partially deterministic processes through the lens of redundancy, resilience and degeneracy (Ahn et al., 2006). Indeed, the interplay between complexity and degeneracy/determinism in biological systems has received a great deal of attention in network science and the complex systems literature (Hoel et al., 2013).

Generally, network methods focus on the literal constructed network to explore the effects of noise, perturbations, degeneracy and deterministic processes—but such an approach might not be able to tease apart certain informative aspects of the structure of the network (Klein & Hoel, 2020). Most biological networks have a certain form of hierarchy, and it is quite possible that despite the noise at the micro-scale (or the literal network), a network can possess more informative structure at a different scale (Klein & Hoel, 2020; Klein et al., 2021). The big question, then, is how to infer these higher, content-rich scales and to decide which scale is the most informative to use in analyses.

The concepts of *effective information* and *causal emergence* are network metrics devised to confront these problems (Klein & Hoel, 2020; Klein et al., 2021). These two indices, which are based on information theory (Hoel et al., 2013), measure the information associated with the structure of a network, which can be used to identify the presence of informative higher scales. In this work, we briefly describe the metrics, their interpretations and usage, through implementation both as an R package and a Python library.

2 | OVERVIEW OF THE METRICS

The deterministic nature of a network can be measured in terms of the uncertainty in the interactions among the nodes. That is, if activity on a given node influences or induces changes in another node's behaviour, we can study the inherent uncertainty in this relationship. This influence may come in the form of neuronal activity, where one neuron's firing could induce firing in a neighbouring neuron; in social networks, influence could come in the form of a connection strength, where my activity may depend more on a close friend than a mere acquaintance; or in biological systems, where the co-expression of two biological processes may indicate a dependency between the two processes.

If node *j* is connected to node *i*, and the behaviour of *i* suddenly changes, what are the chances that node *j*'s behaviour is also changed? To quantify this, we can calculate the average Shannon entropy of the out- and in-edges of nodes in a network. The entropy (or uncertainty) of out-edges corresponds to the degree of *determinism* in the network connections, while *degeneracy* captures how evenly the in-edges are distributed across the network (see following section). Effective information (*EI*) brings these two concepts (i.e. *determinism*, *degeneracy*) together, and is defined as the difference between the entropy of the average out-weight in the network minus the average entropy of the out-weights (see Klein & Hoel, 2020, for a detailed discussion):

$$EI = H\left(\left\langle W_i^{\text{out}}\right\rangle\right) - \left\langle H\left(W_i^{\text{out}}\right)\right\rangle \tag{1}$$

where H refers to the entropy, and $\langle\rangle$ is the average over all nodes, denoted by i. Effective information varies with network size, so researchers typically use a size-normalized measure known as *effectiveness* to characterize networks:

$$\text{Effectiveness} = \frac{EI}{log_2\left(n\right)} \tag{2}$$

where *EI* is normalized with respect to the number of nodes, n. Using these measures, we can compare different networks, highlighting structural properties that are associated with more or less uncertainty in the edges of a network. This gives researchers a lens through which to view network dynamics, robustness and growth, and can inform why we may expect different systems across nature and society to vary in the amount of noise contained in their connectivity structures. For example, researchers have studied the effectiveness of various biological or socio-technical systems (Klein & Hoel, 2020), protein interaction networks (Klein et al., 2021), animal social networks (Swain, Williams, et al., 2021; van der Marel et al., 2021) and more.

2.1 | Breaking down effective information: Determinism and degeneracy

As previously mentioned, the *EI* is defined by two key components: *determinism* and *degeneracy*. Determinism refers to the average

uncertainty over outgoing edges in a network. If a given node *i* has a single outgoing edge, there is no uncertainty about where a random walker on node *i* will traverse to next—in other words, node *i* increases the network's *determinism*. On the other hand, if *i* has outgoing edges to the other *n* nodes in a network, we are maximally uncertain about the subsequent location of the random walker (Figure 1a,b). Whereas determinism is a measure that quantifies uncertainty about the *future* states of random walkers in the network, *degeneracy* captures uncertainty about the source of the random walkers on a given node. Consider a node with only one incoming edge compared to a node with *n* incoming edges; when few nodes have disproportionately many connections, the network becomes more uncertain, on average, about past states of random walkers (see Figure 1c,d). Note: the notion of random walkers is a useful—if incomplete—way of formalizing flow or transitions in networks. Because networks are such a ubiquitous tool for representing complex systems, there are a number of ways to represent dynamical processes on networks from different domains; random walks have the benefit of being well-studied mathematically and are general enough such that they add few domain-specific assumptions. The prospect of re-defining a version of the effective information metric by replacing random walk dynamics with domain-specific dynamics is a rich and promising future direction of research.

Different networks display different behaviour when it comes to *determinism* and *degeneracy*. For example, networks with heavy-tailed degree distributions, which are common across nature and society, typically have higher *degeneracy* (and lower *EI* as a result). Similarly, networks containing dense subgraphs and/or community structure typically have lower *determinism* (and lower *EI* as a result). Networks that are sparse and have low variance in degree have higher *EI* and have less uncertainty about the micro-scale interactions that take place in the network.

In practice, determinism and degeneracy may correspond to different facets of biological systems. For example, a highly deterministic relationship in a gene expression network would correspond to almost one gene almost certainly causing the expression of another. A degenerate structure in, for example, an ecological network would be more *sensitive*, a connectivity structure where fluctuations in the prevalence of any of a number of species would influence the prevalence of a target species.

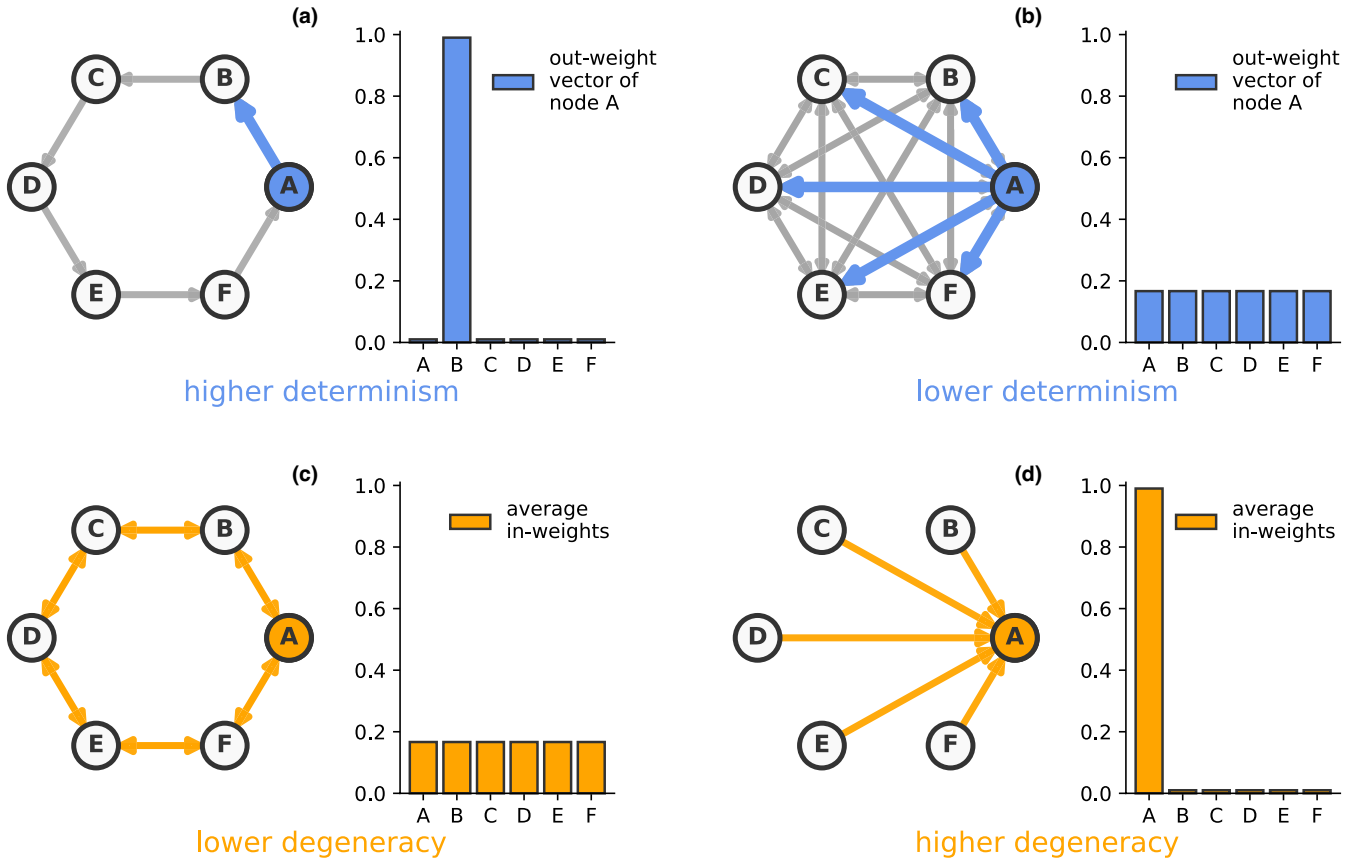

FIGURE 1 Determinism and degeneracy. Simple networks highlighting the differences between high/low determinism and degeneracy. (a) In this network, random walkers on node *A* will traverse to node *B* with probability of 1.0. (b) In this network, random walkers on node *A* will traverse to any other node in the network with probability *1/n*, where *n* is the size of the network. (c) In this network, no single node has a higher in-weight than any other node, which means that degeneracy has been minimized. (d) In this network, every node has a single outgoing edge, each of which connects to node *A*; this is an example of a maximally degenerate network

2.2 | Causal emergence: Informative higher scales

Networks that have higher *determinism* and lower *degeneracy* are less noisy and have more certainty in the interactions among their nodes; in this sense they are more effective (i.e. they have higher *effectiveness*). One recent development in network science and information theory has been the introduction of *causal emergence* to the study of complex networks (Klein & Hoel, 2020); this is a phenomenon where the original micro-scale network is recast as a coarse-grained, macro-scale network so as to increase the *effective information* of the original network. Algorithmically, this involves selecting a subgraph of nodes and their edges—a subgraph with low *determinism* and/or high *degeneracy*—from the original network and replacing the subgraph with a single macro node. A core functionality of the `einet` package is that it automates this coarse-graining process, using a variety of techniques from spectral graph theory to find and isolate regions of low *determinism* and/or high *degeneracy*. Users simply input a network into the `causal_emergence` function, and if there are highly noisy subgraphs, the function will use those subgraphs to create a higher scale macro representation of the network. We illustrate this process in Figure 2, using the protein–protein interaction network of *Mycoplasma putrefaciens* as an example.

Causal emergence (the quantity) is defined as the difference between the *EI* of the original, micro-scale network and the *EI* of a coarse-grained, higher scale representation of the network. In the context of ecological or biological systems, causal emergence of a network can be interpreted in many ways. For example, in insect social networks, we might expect to see more informative higher scales that capture more of the emergent coordination of the collective as a whole (e.g. in ant colonies; see Swain, Devereux, et al., 2021).

Higher order networks are of growing interest to network scientists, and they appear to be relatively ubiquitous, appearing in a variety of contexts from biology to economics to social structures and more. What causal emergence adds to this growing body of work is the ability to quantify why certain higher scale structures emerge, identifying common structural properties at the micro-scale that tend to facilitate higher order structure. How to interpret a given higher order structure (i.e. output from the `einet` package) can vary greatly across different disciplines, but put simply: *causal emergence* uses the pattern of noise/uncertainty at the micro-scale to define the most informative scale at which to represent a complex system. These higher scale structures may hold theoretical import (e.g. by comparing the amount of causal emergence across different species groups or ecosystems, as in Klein et al., 2021), or they help uncover measurement noise that must be addressed before studying networks from a given system (similar to recent advances in network permutation testing, as in Puga-Gonzalez et al., 2021). Future work may also explore the use of causal emergence in network comparison (Hartle et al., 2020) or parameter selection when modelling ecological networks (e.g. animal social networks, species interaction networks, etc.), to constrain the structure of networks produced by a generative model. Ultimately, a comprehensive, prescriptive description of how to interpret higher order structure in networks requires more theoretical development; we view `einet` as important to such future theoretical advances.

3 | IMPLEMENTATION AND USAGE

These tools are available through `einet`, both as an R package (GitHub) and a Python library (Github). For details about the implementation of the spectral algorithm for computing the micro-to-macro-scale mappings, see Griebenow et al. (2019).

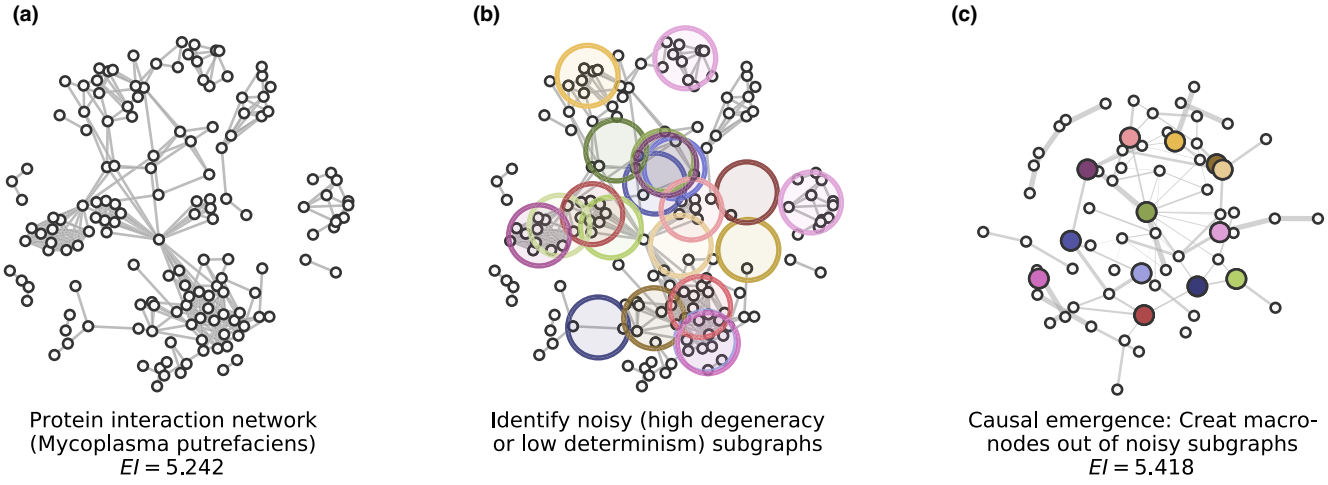

FIGURE 2 Causal emergence in protein networks. To illustrate the basic steps for calculating causal emergence in networks, we use the protein–protein interaction network of *Mycoplasma putrefaciens* as an example. (a) The original (micro-scale) network. (b) The different colour circles are meant to highlight subgraphs in the original network that are especially noisy—either because they consist of densely connected nodes or because they are sources of degeneracy. (c) To calculate causal emergence, we construct a macro-scale network by grouping the subgraphs from (b) into macro nodes, while preserving the total in/out edge weights into the selected subgraph. Note here that the *EI* of the network in (a) and the *EI* from (c) are different: the macro-scale network has more effective information than the original micro-scale network

3.1 | Example usage (R)

Effective information and effectiveness

```
>>> library(devtools)
>>> library(igraph)
>>> install_github("travisbyrum/einet")
>>> library(einet)
>>> G <- karate
>>> ei_micro <- effective_information(G)
>>> sprintf("EI micro: %.6f",ei_micro)
EI micro: 2.350095

>>> eff_micro <- ei_micro / log2(gorder(G))
>>> sprintf("Effectiveness: %.6f",eff_micro)
Effectiveness: 0.461939
```

Causal emergence

```
>>> library(einet)
>>> library(igraph)
>>> set.seed(2)
>>> G <- karate
>>> CE <- causal_emergence(G)
>>> ei_macro <- CE$ei_macro
>>> sprintf("EI macro: %.6f",ei_macro)
EI macro: 2.410400
```

The `CE` object returned by the `causal_emergence` function is a list object composed of `g_micro` (original network), `g_macro` (coarse-grained network), `mapping` (a list mapping from the micro nodes to their corresponding macro node, after coarse-graining happened), `ei_macro` (effective information of the coarse-grained network), `ei_micro` (original network's effective information), `effectiveness` (original network's effective information, divided by log(*n*) (normalization), where *n* is the number of nodes), and `ce` (the normalized difference in effective information between the micro- and macro-scale networks). Because the `causal_emergence()` function depends on the initial conditions of the search for the higher scale mapping, we recommend running the function several times and selecting the mapping with the largest effective information (note that this is the source of the small differences between the causal emergence in the R implementation and the Python below).

3.2 | Example usage (Python)

Effective information and effectiveness.

```
>>> from ei_net import effective_information
>>> import networkx as nx; import numpy as np
>>> G = nx.karate_club_graph()
>>> EI_micro = effective_information(G)
>>> print("EI micro: %.6f"%EI_micro)
EI micro: 2.350095
```

Causal emergence

```
>>> from ei_net import causal_emergence
>>> import networkx as nx; import numpy as np
>>> np.random.seed(2)
>>> CE = causal_emergence(G)
>>> EI_macro = CE["EI_macro"]
>>> print("EI macro: %.6f"%EI_macro)
EI macro: 2.415379

>>> N = G.number_of_nodes()
>>> eff_gain = (EI_macro-EI_micro)/np.log2(N)
>>> print("Effectiveness gain: %.6f"%eff_gain)
Effectiveness gain: 0.012832
```

The `CE` object returned by the `causal_emergence` function is a dictionary with the following keys: `G_micro` (original network), `G_macro` (coarse-grained network), `mapping` (a list mapping from the micro nodes to their corresponding macro node, after coarse-graining happened), `EI_macro` (effective information of the coarse-grained network), and `EI_micro` (original network's effective information). In the Python implementation of this software, there is an additional `causal_emergence_spectral()` function that leverages tools from spectral graph theory to compute causal emergence, and does so much faster than the greedy search used in `causal_emergence()` (see Griebenow et al., 2019 for more details).

ACKNOWLEDGEMENTS

B.K. and S.V.S. acknowledge the support of a grant from the John Templeton Foundation (61780). The opinions expressed in this publication are those of the author(s) and do not necessarily reflect the views of the John Templeton Foundation. A.S. thanks National Science Foundation award DGE-1632976 for training and support.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

AUTHORS' CONTRIBUTIONS

B.K., A.S., S.V.S. and W.F.F. conceptualized the project; B.K., T.B. and A.S. wrote the software in R and Python; A.S., B.K., S.V.S. and W.F.F. wrote and edited the manuscript. All authors provided feedback and approved the publication.

PEER REVIEW

The peer review history for this article is available at https://publons.com/publon/10.1111/2041-210X.13805.

DATA AVAILABILITY STATEMENT

The R package is available at github.com/travisbyrum/einet and archived on Zenodo (Byrum et al., 2021; https://doi.org/10.5281/zenodo.5784492). Python code is available at github.com/jkbren/einet and archived on Zenodo (Klein, 2021; https://doi.org/10.5281/zenodo.5236550); this repository also includes data as well as introductory Jupyter notebooks detailing various aspects of effective information, causal emergence and networks.

ORCID

Brennan Klein https://orcid.org/0000-0001-8326-5044
Anshuman Swain https://orcid.org/0000-0002-9180-2222
Samuel V. Scarpino https://orcid.org/0000-0001-5716-2770
William F. Fagan https://orcid.org/0000-0003-2433-9052

REFERENCES

Ahn, A. C., Tewari, M., Poon, C. S., & Phillips, R. S. (2006). The limits of reductionism in medicine: Could systems biology offer an alternative? *PLoS Medicine*, *3*(6), e208.

Bassett, D. S., & Sporns, O. (2017). Network neuroscience. *Nature Neuroscience*, *20*(3), 353–364.

Byrum, T., Swain, A., Klein, B., & Fagan, W. F. (2021). travisbyrum/einet: einet. Version v1.0.2. https://doi.org/10.5281/zenodo.5784492

Costanzo, M., Kuzmin, E., van Leeuwen, J., Mair, B., Moffat, J., Boone, C., & Andrews, B. (2019). Global genetic networks and the genotype-to-phenotype relationship. *Cell*, *177*(1), 85–100.

Dunne, J. A., Williams, R. J., & Martinez, N. D. (2002). Food-web structure and network theory: the role of connectance and size. *Proceedings of the National Academy of Sciences of the United States of America*, *99*(20), 12917–12922.

Edelman, G. M., & Gally, J. A. (2001). Degeneracy and complexity in biological systems. *Proceedings of the National Academy of Sciences of the United States of America*, *98*(24), 13763–13768.

Faust, K., & Raes, J. (2012). Microbial interactions: From networks to models. *Nature Reviews Microbiology*, *10*(8), 538–550.

Freilich, M. A., Rebolledo, R., Corcoran, D., & Marquet, P. A. (2020). Reconstructing ecological networks with noisy dynamics. *Proceedings of the Royal Society A*, *476*(2237), 20190739.

Gosak, M., Markovič, R., Dolenšek, J., Rupnik, M. S., Marhl, M., Stožer, A., & Perc, M. (2018). Network science of biological systems at different scales: A review. *Physics of Life Reviews*, *24*, 118–135.

Griebenow, R., Klein, B., & Hoel, E. (2019). Finding the right scale of a network: Efficient identification of causal emergence through spectral clustering. *arXiv* [Preprint], 1908.07565.

Gysi, D. M., & Nowick, K. (2020). Construction, comparison and evolution of networks in life sciences and other disciplines. *Journal of the Royal Society Interface*, *17*(166), 20190610.

Hartle, H., Klein, B., McCabe, S., Daniels, A., St-Onge, G., Murphy, C., & Hébert-Dufresne, L. (2020). Network comparison and the within-ensemble graph distance. *Proceedings of the Royal Society A*, *476*, 20190744.

Hoel, E., Albantakis, L., & Tononi, G. (2013). Quantifying causal emergence shows that macro can beat micro. *Proceedings of the National Academy of Sciences of the United States of America*, *110*(49), 19790–19795.

Klein, B. (2021). jkbren/einet: einet. Version v1.0. *Zenodo*, https://doi.org/10.5281/zenodo.5236550.

Klein, B., & Hoel, E. (2020). The emergence of informative higher scales in complex networks. *Complexity*, *2020*, 1–12.

Klein, B., Hoel, E., Swain, A., Griebenow, R., & Levin, M. (2021). Evolution and emergence: Higher-order information structure in protein interactomes across the tree of life. *Integrative Biology*, *1-12*, 283–294. https://doi.org/10.1093/intbio/zyab020

Newman, M. E. (2018). Network structure from rich but noisy data. *Nature Physics*, *14*(6), 542–545.

Pascual, M., & Dunne, J. A. (2006). *Ecological networks: Linking structure to dynamics in food webs*. Oxford University Press.

Pinter-Wollman, N., Hobson, E. A., Smith, J. E., Edelman, A. J., Shizuka, D., De Silva, S., Waters, J. S., Prager, S. D., Sasaki, T., Wittemyer, G., Fewell, J., & McDonald, D. B. (2014). The dynamics of animal social networks: Analytical, conceptual, and theoretical advances. *Behavioral Ecology*, *25*(2), 242–255.

Puga-Gonzalez, I., Sueur, C., & Sosa, S. (2021). Null models for animal social network analysis and data collected via focal sampling: Pre-network or node network permutation? *Methods in Ecology and Evolution*, *12*, 22–32.

Shaw, J. O., Coco, E., Wootton, K., Daems, D., Gillreath-Brown, A., Swain, A., & Dunne, J. A. (2021). Disentangling ecological and taphonomic signals in ancient food webs. *Paleobiology*, *47*(3), 385–401. https://doi.org/10.1017/pab.2020.59

Sporns, O. (2014). Contributions and challenges for network models in cognitive neuroscience. *Nature Neuroscience*, *17*(5), 652–660.

Swain, A., Devereux, M., & Fagan, W. F. (2021). Deciphering trophic interactions in a mid-Cambrian assemblage. *iScience*, *24*(4), 102271. https://doi.org/10.1016/j.isci.2021.102271

Swain, A., Williams, S. D., Di Felice, L. J., & Hobson, E. A. (2021). Interactions, information and emergence: Exploring task allocation in ant colonies using network analysis. *bioRxiv* [Preprint]. https://doi.org/10.1101/2021.03.29.437501

Tononi, G., Sporns, O., & Edelman, G. M. (1999). Measures of degeneracy and redundancy in biological networks. *Proceedings of the National Academy of Sciences of the United States of America*, *96*(6), 3257–3262.

Tsimring, L. S. (2014). Noise in biology. *Reports on Progress in Physics*, *77*(2), 026601.

van der Marel, A., Prasher, S., Carminito, C., O'Connell, C. L., Phillips, A., Kluever, B. M., & Hobson, E. A. (2021). A framework to evaluate whether to pool or separate behaviors in a multilayer network. *Current Zoology*, *67*(1), 101–111.

Vázquez, D. P., Blüthgen, N., Cagnolo, L., & Chacoff, N. P. (2009). Uniting pattern and process in plant-animal mutualistic networks: A review. *Annals of Botany*, *103*(9), 1445–1457.

Wang, Z., Andrews, M. A., Wu, Z. X., Wang, L., & Bauch, C. T. (2015). Coupled disease-behavior dynamics on complex networks: A review. *Physics of Life Reviews*, *15*, 1–29.

How to cite this article: Klein, B., Swain, A., Byrum, T., Scarpino, S. V., & Fagan, W. F. (2022). Exploring noise, degeneracy and determinism in biological networks with the `einet` package. *Methods in Ecology and Evolution*, 13, 799–804. https://doi.org/10.1111/2041-210X.13805