

## NOS: A software suite to compute node overlap and segregation ( ) in ecological networks

Strona, Giovanni; Matthews, Thomas; Kortsch, Susanne; Veetch, Joseph A

DOI:

[10.1111/ecog.03447](https://doi.org/10.1111/ecog.03447)

License:

Creative Commons: Attribution (CC BY)

*Document Version*

Publisher's PDF, also known as Version of record

*Citation for published version (Harvard):*

Strona, G, Matthews, T, Kortsch, S & Veetch, JA 2018, 'NOS: A software suite to compute node overlap and segregation ( ) in ecological networks', *Ecography*, vol. 41, no. 3, pp. 558–566.  
<https://doi.org/10.1111/ecog.03447>

[Link to publication on Research at Birmingham portal](#)

### General rights

Unless a licence is specified above, all rights (including copyright and moral rights) in this document are retained by the authors and/or the copyright holders. The express permission of the copyright holder must be obtained for any use of this material other than for purposes permitted by law.

- Users may freely distribute the URL that is used to identify this publication.
- Users may download and/or print one copy of the publication from the University of Birmingham research portal for the purpose of private study or non-commercial research.
- User may use extracts from the document in line with the concept of 'fair dealing' under the Copyright, Designs and Patents Act 1988 (?)
- Users may not further distribute the material nor use it for the purposes of commercial gain.

Where a licence is displayed above, please note the terms and conditions of the licence govern your use of this document.

When citing, please reference the published version.

### Take down policy

While the University of Birmingham exercises care and attention in making items available there are rare occasions when an item has been uploaded in error or has been deemed to be commercially or otherwise sensitive.

If you believe that this is the case for this document, please contact [UBIRA@lists.bham.ac.uk](mailto:UBIRA@lists.bham.ac.uk) providing details and we will remove access to the work immediately and investigate.

# ECOGRAPHY

## Software notes

### NOS: a software suite to compute node overlap and segregation in ecological networks

Giovanni Strona, Thomas J. Matthews, Susanne Kortsch and Joseph A. Veech

G. Strona (<http://orcid.org/0000-0003-2294-4013>) ([goblinshrimp@gmail.com](mailto:goblinshrimp@gmail.com)), European Commission, Joint Research Centre, Directorate D – Sustainable Resources – Bio-Economy Unit, Ispra (VA), Italy. – T. J. Matthews, GEES (School of Geography, Earth and Environmental Sciences), The Univ. of Birmingham, Birmingham, UK, and CE3C – Centre for Ecology, Evolution and Environmental Changes/Azorean Biodiversity Group and Univ. dos Açores – Depto de Ciências Agrárias, Angra do Heroísmo, Açores, Portugal. – S. Kortsch, (<http://orcid.org/0000-0002-6495-4781>), Norwegian College of Fishery Science, UIT the Arctic Univ. of Norway, Tromsø, Norway. – J. A. Veech, Dept of Biology, Texas State Univ., San Marcos, TX, USA.

#### Ecography

41: 558–566, 2018

doi: 10.1111/ecog.03447

Subject Editor: Brody Sandel

Editor-in-Chief: Miguel Araújo

Accepted 21 September 2017



Investigating the structure of ecological networks can help unravel the mechanisms promoting and maintaining biodiversity. Recently, Strona and Veech 2015 (A new measure of ecological network structure based on node overlap and segregation. – *Methods Ecol. Evol.* 6: 907–915) introduced a new metric ( $\bar{N}$ , pronounced ‘nos’), that allows assessment of structural patterns in networks ranging from complete node segregation to perfect nestedness, and that also provides a visual and quantitative assessment of the degree of network modularity. The  $\bar{N}$  metric permits testing of a wide range of hypotheses regarding the tendency for species to share interacting partners by taking into account ecologically plausible species interactions based on constraints such as trophic levels and habitat preference. Here we introduce NOS, a software suite (including a web interface freely accessible at <<http://nos.alwaysdata.net>>, an executable program, and Python and R packages) that makes it possible to exploit the full potential of this method. Besides computing node overlap and segregation ( $\bar{N}$ ), the software provides different functions to automatically identify a set of possible resource–consumer interactions in food webs based on trophic levels. As an example of application, we analyzed two well-resolved high-latitude marine food webs, showing that an explicit a priori consideration of trophic levels is fundamental for a proper assessment of food web structure.

---

#### Background

Natural systems are bound together by invisible wires connecting ecologically interacting species. These can be arranged into bipartite networks, i.e. networks where interactions can take place only between (but not within) the members of two distinguishable groups (e.g. plant–pollinator, seed–disperser and host–parasite



[www.ecography.org](http://www.ecography.org)

networks), or into unimode networks, where such distinction is not straightforward (e.g. food-webs, species co-occurrence networks). It has been widely hypothesized that the structure of these networks could play a leading role in promoting the diversity and preserving the stability of ecosystems, providing new momentum to the diversity-stability debate (May 1973, McCann 2000, Allesina and Tang 2012). In particular, various studies have identified the seemingly widespread tendency of species to share interacting partners (i.e. nestedness) as being key to species coexistence (Bastolla et al. 2009, Rohr et al. 2014). This has generated much interest in developing analytical tools to measure nestedness, most of which are implemented in R packages such as ‘bipartite’ (Dormann et al. 2008) and ‘vegan’ (Oksanen et al. 2017).

Recently, Strona and Veech (2015) introduced a new measure ( $\bar{N}$ , pronounced ‘nos’) capable of quantifying the tendency of species to share (or not to share) interacting partners, a more detailed description of the unique features of  $\bar{N}$  are presented in Box 1. A pattern where species tend to share more interacting partners than expected is commonly identified as nestedness (Ulrich et al. 2009), and can be measured by different available metrics. As previously demonstrated (Strona and Veech 2015),  $\bar{N}$  is well correlated with popular nestedness metrics such as the NODEF (Almeida-Neto et al. 2008), and the more recent spectral radius measure (Staniczenko et al. 2013). The concept of nestedness became popular in the study of presence-absence species/locality matrices before being used in ecological network analysis (Atmar and Patterson 1993, Brualdi and Sanderson 1999). Similar to these other nestedness metrics,  $\bar{N}$  can be applied to species/locality matrices, because these correspond to bipartite networks where localities are linked to the species inhabiting them.

However, and distinct from other available procedures,  $\bar{N}$  can capture in a continuous way not only nestedness, but also

the opposite tendency, i.e. a situation whereby species tend to share less partners than expected, as may happen, for example, in a food web where consumers are highly specialized on the resources they use, so to minimize competition. This pattern, that we term node segregation, has received much attention in the context of presence-absence matrices, and particularly in the study of species co-occurrence (Ulrich and Gotelli 2007), where it is usually measured using an index called the C-score (Stone and Roberts 1990). However, node segregation has received less attention in the study of ecological networks, possibly due to the absence of dedicated metrics.

$\bar{N}$  represents a synthetic measure of ecological network structure which is consistent with, and can therefore replace, common nestedness and segregation measures (Supplementary material Appendix 1 Fig. A1). Furthermore, and distinct from other available procedures,  $\bar{N}$  can be applied not only to bipartite networks (e.g. plant-pollinator, fruit-disperser and host-parasite networks, and species/locality matrices) but also to unimode networks, such as food webs (Strona and Veech 2017).

Similar to NODEF,  $\bar{N}$  is computed as the average of the standardized overlap observed in each pair of nodes (e.g. species) in the network (cf. Almeida-Neto et al. 2008, Strona and Veech 2015). The variability in the values of pairwise node overlap permits the discrimination between situations where all pairs of nodes show no tendency towards either more or less than expected overlap, from situations where a proportion of the node pairs show a tendency for high overlap, while the rest of the nodes show a tendency for segregation. The latter situation corresponds to a scenario of modularity, where nodes can be attributed to different clusters, and tend to share interactions with nodes in their cluster but not with nodes from different clusters (Newman 2006). Therefore, the computational procedure used to obtain  $\bar{N}$  can also provide a derivative measure of modularity.

### Box 1. Unique features of the $\bar{N}$ metric

The basic step in computing the  $\bar{N}$  metric relies on a probabilistic approach to analyze species co-occurrence (i.e. species sharing sites) previously implemented by Veech (2013). Rather than analyzing species co-occurrence,  $\bar{N}$  measures the extent to which species share other species, i.e. interacting partners.  $\bar{N}$  is conceptually similar to various metrics commonly used to assess the degree of species sharing between different sites or similarity of species composition, such as the many available  $\beta$ -diversity indices for presence-absence data (see Koleff et al. 2003 for a review). Nevertheless,  $\bar{N}$  differs considerably from other measures in the way it is computed. The most unique feature of  $\bar{N}$  is the way it is used to quantify pairwise node overlap/segregation as a standardized deviation from a probabilistic expectation. This allows for powerful significance testing, and for a direct comparison of  $\bar{N}$  values from different networks (Strona and Veech 2015). A second fundamental difference and advantage of  $\bar{N}$  compared to common overlap/segregation metrics is that  $\bar{N}$  can be applied to any kind of network (Strona and Veech 2015, 2017), while most of the latter are designed for bipartite networks (but see, for example, Johnson et al. 2013, Cantor et al. 2017). Third, the computational design of  $\bar{N}$  permits users to distinguish, beforehand and explicitly, between ‘permitted’ and ‘forbidden’ interactions, providing an innovative framework to test various ecological/evolutionary hypotheses (Strona and Veech 2017) – see also the section ‘Automatic rules for determining potential neighbours’. The fourth unique feature is the stability of  $\bar{N}$  towards ‘partial’ pairwise comparisons between nodes. That is, although the ‘formal’ computation of  $\bar{N}$  requires computing  $N_{ij}$  between all pairs of nodes  $i$  and  $j$ , the average of  $N_{ij}$  tends very rapidly toward the true value of  $\bar{N}$  (see Fig. 3 and the section ‘Estimating  $\bar{N}$  from a small fraction of pairwise node comparisons’). Lastly,  $\bar{N}$  assesses in a single measure, both overlap, i.e. nestedness, and segregation on a continuous, symmetric scale. This, together with the additional information on modularity offered by the standard deviation of  $N_{ij}$  values, provides a synthetic view of network structure (nicely illustrated by the histogram representation of  $N_{ij}$  values, Fig. 2).

Here we aim at demonstrating and promoting the use of  $\bar{N}$  by introducing NOS, a suite of software tools, including a user-friendly online interface accessible at: <http://nos.alwaysdata.net>, a Windows executable program, and a Python (van Rossum and de Boer 1991) and an R package ('nos', <https://cran.r-project.org/web/packages/nos/index.html>). An important feature of  $\bar{N}$  is its ability to assess the degree of overlap in a network relative to different scenarios of ecologically plausible species interactions (see also Strona and Veech 2017 for a thorough discussion on the theoretical implications of this feature). One of the main aims of NOS is to promote the application of this approach to ecological problems. In the following paragraphs, we will illustrate the main features of NOS, demonstrating its usage on two empirical food webs.

### Computation of $\bar{N}$

The general formula to compute the overlap in neighbours for a focal pair of nodes (e.g. species)  $V_i$  and  $V_j$  is:

$$N_{ij} = \frac{(S_{ij} - P_{ij})}{\min(d_i, d_j)} \times \frac{1}{\Omega_{ij}} \quad (1)$$

where  $S_{ij}$  is the actual number of neighbours (e.g. interacting partner species) shared by  $V_i$  and  $V_j$ ,  $P_{ij}$  is the expected

number of shared neighbours, and  $d_i$  and  $d_j$  are the respective node degrees (i.e. the number of neighbours) (Fig. 1).  $P_{ij}$  is computed, using a probabilistic approach derived from Veech (2013), as:

$$P_{ij} = \sum_{k=1}^{\min(d_i, d_j)} \frac{\binom{n}{k} \times \binom{n-k}{d_j-k} \times \binom{n-d_j}{d_i-k}}{\binom{n}{d_j} \times \binom{n}{d_i}} \times k \quad (2)$$

$\Omega_{ij}$  is a standardization parameter corresponding to the maximum possible value of  $N_{ij}$  (Fig. 1). The standardization parameter is needed to ensure that the measure is symmetric around 0 (and scales between -1 and 1). For any  $V_i$  and  $V_j$  sharing more nodes than expected by chance (i.e. having  $S_{ij} > P_{ij}$ ),  $\Omega_{ij}$  is computed as:

$$\Omega_{ij} = \frac{\min(d_i, d_j) - P_{ij}}{\min(d_i, d_j)} \quad (3)$$

For any  $V_i$  and  $V_j$  sharing fewer nodes than expected by chance (i.e. having  $S_{ij} < P_{ij}$ ), then if:

$$(d_i + d_j - n) < 0 \rightarrow \Omega_{ij} = \frac{P_{ij}}{\min(d_i, d_j)} \quad (4)$$

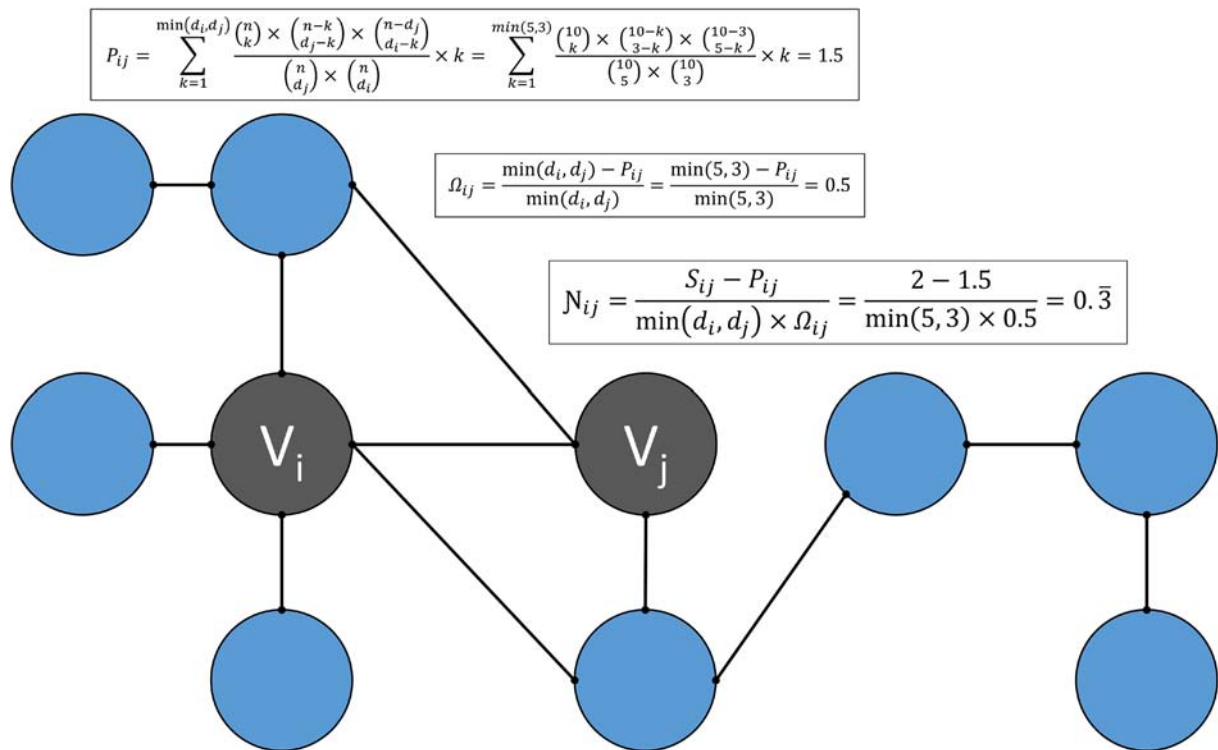


Figure 1. Example of computation of  $N_{ij}$  in a network with 10 nodes.  $S_{ij}$  is the actual number of neighbours (e.g. interacting partner species) shared by  $V_i$  and  $V_j$ .  $P_{ij}$  is the expected number of shared neighbours, computed using combinatorics and considering all nodes (including the two focal ones) as potential partners for  $V_i$  and  $V_j$ .  $d_i$  and  $d_j$  are the respective node degrees (i.e. the number of neighbours per node).

or else:

$$(d_i + d_j - n) \geq 0 \rightarrow \Omega_{ij} = \frac{P_{ij} - (d_i + d_j - n)}{\min(d_i, d_j)} \quad (5)$$

If the number of observed shared nodes is equal to that expected ( $S_{ij} = P_{ij}$ ), then Eq. 1 gives 0 regardless of the  $\Omega_{ij}$  parameter (which is set to 1).

$\bar{N}$  is then computed as the average of all  $N_{ij}$  pairs (with  $i > j$ ). For networks where links have directionality, two separate sets of  $N_{ij}$  are computed. For example, in a plant pollinator network, a set of  $N_{ij}$  is computed for all pairs of plants based on the overlap of their pollinators, and another set of  $N_{ij}$  values is computed for all pairs of pollinators based on the overlap of their associated plants. The two series of  $N_{ij}$  are then averaged to a  $\bar{N}_{in}$  and a  $\bar{N}_{out}$  values. These values, which are potentially informative of ecological patterns, can be averaged to provide an overall  $\bar{N}$  score.

$\bar{N}$  varies between  $-1$  and  $1$ , where  $-1$  represents a network with less than expected node overlap (i.e. a completely segregated network) and  $1$  represents a network with higher than expected node overlap (i.e. a completely nested network). An  $\bar{N}$  equal to zero could indicate either a situation where positive and negative values cancel each other out, or a situation where all node pairs have the randomly expected number of neighbours, or any intermediate scenario between these two.

A situation where a group(s) of nodes can be identified in a network where node overlap is substantially higher than expected between nodes belonging to the same group, and lower than expected between nodes belonging to different groups, indicates modularity (Newman 2006). Analyzing the distribution of  $N_{ij}$  values provides an immediate visual way to distinguish between random and modular networks, with the latter showing a bimodal distribution of  $N_{ij}$  values, with a negative peak arising from node segregation between different modules, and a positive one arising from node overlap within modules. As discussed in Strona and Veech (2015), a straightforward way to quantify this pattern is evaluating the statistical dispersion of  $N_{ij}$  values around the mean. Thus, the standard deviation of the  $N_{ij}$  values can serve as a simple measure of modularity (Mod). It should be highlighted that identifying modules (i.e. communities) in networks is a non-trivial problem (Radicchi et al. 2004). However, Mod is not meant to replace more refined, specific, measures of modularity, rather it can be used as a key to interpret and compare  $\bar{N}$  values (Fig. 2). As in the case of  $\bar{N}$ , if the target network is directed, two distinct measures of modularity can be computed for in- and out-nodes and then averaged to provide an overall measure of modularity.

## Adjusting the $n$ parameter

The key feature of the aforementioned method is the application of the probabilistic approach of Veech (2013) to

compute the expected number of nodes shared between two other nodes,  $P_{ij}$  (Eq. 2).

The value  $n$  in Eq. (2) indicates the number of nodes that can be potentially shared by  $V_i$  and  $V_j$ , and is key to understanding the potential of  $\bar{N}$  for robust and flexible analysis of user-specified hypotheses. In fact, adjusting  $n$  according to different criteria allows for the application of the method to different kinds of networks (as described in Strona and Veech 2015), as well as for testing different hypotheses.

As highlighted above, in directed networks, such as food webs or host–parasite networks, where links go from resources to consumers,  $\bar{N}$  is calculated as the average of two separate values ( $\bar{N}_{in}$  and  $\bar{N}_{out}$ ) computed for, respectively, all the nodes having in-coming links (e.g. consumers) and all the nodes having out-going links (e.g. resources). In bipartite networks,  $V_i$  and  $V_j$  always belong to the same category (e.g. pollinators), while the difference between observed and expected node overlap ( $N_{ij}$ ) is evaluated on the nodes belonging to the other category (e.g. plants). Thus,  $\bar{N}_{in}$  assesses the overall overlap between plants used by any pair of pollinators, while  $\bar{N}_{out}$  assesses the overlap between the pollinators using any pair of plants. In the standard setup of the NOS software, as well as in classical nestedness analysis, all plants and all pollinators in the network are used in the computation of  $\bar{N}_{in}$  and  $\bar{N}_{out}$ . Consequently, the parameter  $n$  in Eq. 2 is equal to the number of plants when computing  $\bar{N}_{in}$ , and to the number of pollinators when computing  $\bar{N}_{out}$ . However, NOS provides users with the option of reducing the set of potential shared partners on the basis of additional criteria (such as functional constraints) to test specific hypotheses.

The same reasoning outlined above also applies to uni-mode networks, that is, directed networks where nodes can have in-coming and out-going links simultaneously, such as food-webs where the same species can be both a consumer of some species and a resource consumed by other species. In this case, the broadest possible set of potential partners for a node consists of all nodes in the network (with  $n$  therefore being equal to the total number of nodes). This could make sense in some particular situations, but is unrealistic in many others. However,  $\bar{N}$  and NOS allow for added realism by limiting the set of potential partners for each node according to network-wide or node specific criteria. For example, one may assume that self-interactions (such as cannibalism) are not possible in a given network, hence excluding  $i$  and  $j$  from the computation of each  $N_{ij}$  value. Another possibility is that one may identify the set of potential (i.e. ecologically plausible) partners for each node in a network (based for example on the species' trophic level) and compute each  $N_{ij}$  value based only on the intersection between the potential partners of  $i$  and the potential partners of  $j$ , each time adjusting  $n$  accordingly. Besides permitting users to specify plausible interactions, the NOS software also includes functionality enabling automatic identification of possible trophic interactions in a food-web based on network topology (Williams and Martinez 2004).

We demonstrate the use and potential of our software on two well resolved marine high-latitude food webs (Kortsch et al. 2015), highlighting how accounting for



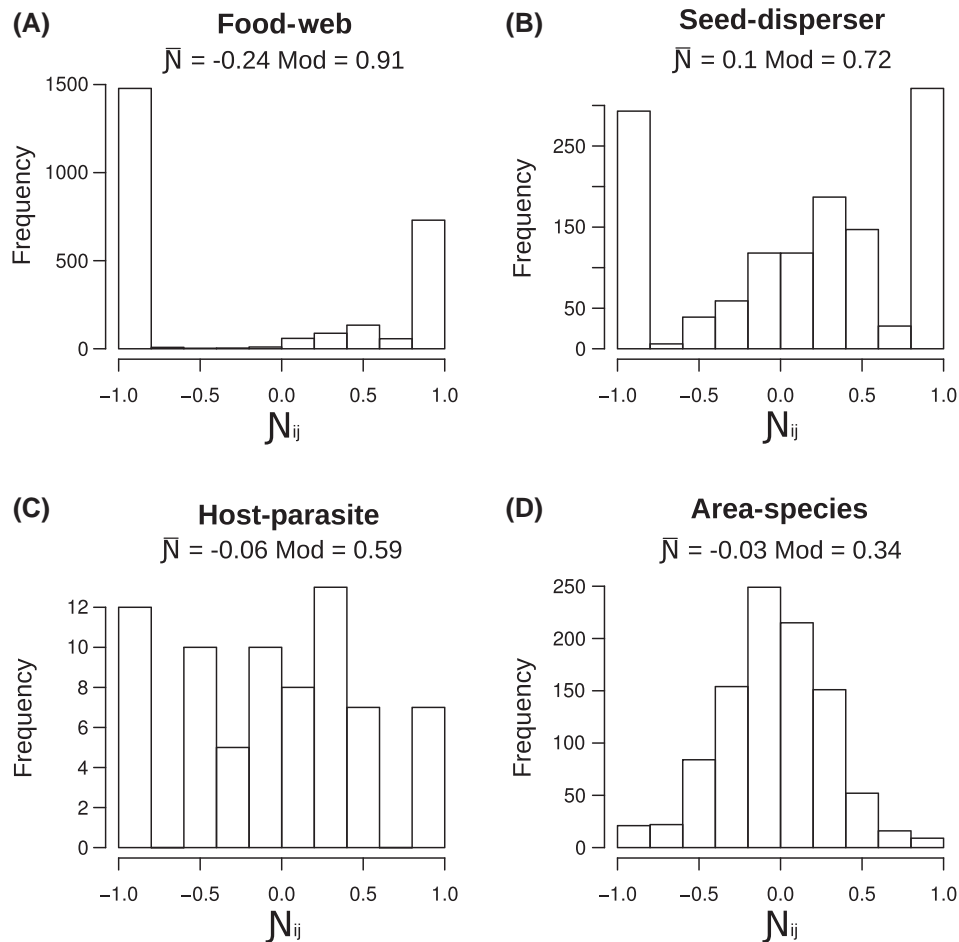


Figure 2. Distribution of  $N_{ij}$  values in networks having node overlap close to 0 (thus indicating a random pattern of overlap), and different degree of modularity (computed as the standard deviation of  $N_{ij}$  values). (A): food-web of an oak gall from Wytham Wood, UK (Askew 1961); (B): seed-disperser networks of tanagers and honeycreepers from Trinidad (Snow and Snow 1971); (C): mammal flea host-parasite network from middle Ural, Russia (Krasnov et al. 2012); (D): meta-community of marine fouling organisms on tile plates from Beaufort, North Carolina (Sutherland and Karlson 1977).

trophy structure can have a very strong impact while evaluating node overlap and segregation. In addition, we explore the possibility of estimating  $\bar{N}$  without performing all possible pairwise comparisons – a potential time-saving shortcut for users of the software. We show how a small random fraction of  $N_{ij}$  values (10%) permits one to obtain a very accurate estimate of  $\bar{N}$ . This property of  $\bar{N}$  reduces computing time while minimizing the loss of information, and thus makes the approach suitable for assessing the structure of very large networks, including non-ecological networks such as those built using microarray data (Broom et al. 2010).

## The web-interface

The web interface to NOS can be accessed at <http://nos.alwaysdata.net> free of charge, and without the need for registration. It has been designed to make usage straightforward whilst maintaining flexibility; providing the same

user-friendly interface as our previous software for nestedness analysis (NeD, <http://purl.oclc.org/ned>; Strona et al. 2014). As a minimum input, NOS requires a network in the form of an edge list, i.e. a list with a single pair of source (e.g. prey) and target (e.g. predator) nodes separated by a comma; source and target identifiers can be any combination of numbers, symbols or characters. If only this list is provided,  $\bar{N}$  will be computed without any particular assumption regarding the set of possible interactions; the only constraints deriving directly from the network structure (see next paragraph). Alternatively, if the user provides another edge list including all the possible (i.e. ecologically plausible) interactions between nodes (including the actual ones), the software will use this list to reduce the set of possible shared partners for each focal pair of nodes before computing  $\bar{N}$ . Edge lists can be either uploaded in the form of plain text files ('csv' or 'txt') or pasted into text-boxes.

Users need to indicate the type of network, e.g. undirected, directed unimode, or directed bipartite. In the case of directed unimode networks, the user is asked to specify whether the

network should be treated as a food web. If this is the case and the user provides no network of potential interactions, then NOS applies a routine to compute trophic levels of nodes based on the network topology (see next section), and then uses this information to build a network of potential/permitted interactions prior to computing  $\bar{N}$ . For exploratory purposes, the web-interface also permits the generation of random networks with a pre-selected degree of overlap/segregation.

## Automatic rules for determining potential neighbours

When the user does not provide a list of potential interactions, NOS automatically determines the set of potential neighbours (and hence  $n$ ) for each pair of nodes. In the case of bipartite networks, all source nodes (e.g. pollinators) are considered as potential partners for each target node (e.g. plants) and vice versa.

In the case of directed unimode networks, NOS provides two options. If the network is a food web, then the user can ask the software to estimate trophic levels for each species in the web, and then use these trophic levels to compile a network of potential interactions. Trophic levels (TL) for each species can be estimated as 1 plus the shortest, average, or maximum path distance between the species and a basal resource (Williams and Martinez 2004). Two species,  $a$  and  $b$ , are included in the list of potential interactions ( $a \rightarrow b$ ) according to two possible criteria. The first criterion is called 'step' and selects potential interactions if

$$TL(a) + l\_step \leq TL(b) < TL(a) + u\_step$$

with  $l\_step$  and  $u\_step$  being user selected parameters.

The second criterion is called 'threshold' and selects potential interactions if

$$TL(a) < TL(b).$$

Actually, the 'threshold' criterion can be considered as a special case of the 'step' criterion where  $l\_step = 1$  and  $u\_step$  is equal or higher than the maximum  $TL$  in the network. Nevertheless, since this may not be intuitive, we have opted for keeping the two options separate.

The user is also free to select whether or not cannibalistic interactions are allowed. If the user does not choose the option of specifying plausible trophic links, then all items in the web are considered as potential interacting partners, i.e.  $n$  is set equal to the number of all nodes in the network. This last criterion also applies to undirected networks.

## Output

The output page of the NOS web interface reports a summary of the analysis, including the user's setup choices,  $\bar{N}$  (and the separated values of  $\bar{N}_{in}$  and  $\bar{N}_{out}$  in the case of directed networks),  $Z$  and  $p$  values, and modularity values

(see Strona and Veech 2015 for details). In addition, the output provides histograms showing the frequency of  $N_{ij}$  values for all node pairs, and separated for in and out nodes in the case of directed networks. Networks with up to 1000 edges are also visualized. Furthermore, the software provides a visual depiction in the form of a heat-map showing  $N_{ij}$  values for all node pairs (arranged in a square adjacency matrix).

## Static software

The same functionalities of the NOS web-interface (except graph generation) are available in the form of an executable program for the Windows environment, a Python script (van Rossum and de Boer 1991), and an R package (available from CRAN) that can all be downloaded from <http://nos.alwaysdata.net/downloads>. In addition to permitting the analysis of networks without restrictions on the numbers of nodes and edges, these four procedures also allow batch (i.e. multiple) analyses. Detailed instructions on the use of the Windows program, and of the Python and R packages are provided alongside the applications.

## Estimating $\bar{N}$ from a small fraction of pairwise node comparisons

In general, ecological networks are relatively small ( $< 500$  nodes) and thus they do not pose computational challenges in using all possible node pairs in order to compute  $\bar{N}$ . However, examining node overlap is often conducted for other kinds of networks, such as metabolic, neural, social and technological networks (Jonhson et al. 2013). These networks are often very large, consisting of thousands to millions of nodes and up to a hundred million links, which makes computational efficiency a non-trivial issue. We examined whether  $\bar{N}$  could be estimated accurately using a random subset of  $N_{ij}$  values, rather than the full set representing all the focal node pairs. We compared 1000 nested and 1000 segregated random networks (obtained by using the method in Strona and Veech 2015) with their respective  $\bar{N}$  values obtained by only including a random 10% of the possible node pair comparisons in each network. We found that the exhaustive inclusion of all node pairs is not necessary, and that using a small fraction of node pairs is enough to obtain an estimate of  $\bar{N}$  very close to the actual value (Fig. 3). This property of the metric is useful as it allows it to be applied to very large networks, and is most likely a result of complex transitivity properties of pairwise comparisons. That is, it is likely that if  $N_{ij}$  and  $N_{kj}$  are similar, then they will also be similar to  $N_{ik}$ , and so on.

NOS software permits users to take advantage of this property and speed up computational time by reducing the number of performed pair comparisons (calculation of  $N_{ij}$  values) to a desired percentage. This could be a useful option for exploratory analysis, but for computationally tractable networks, we recommend the full analysis. For large networks, the online software automatically uses a maximum of

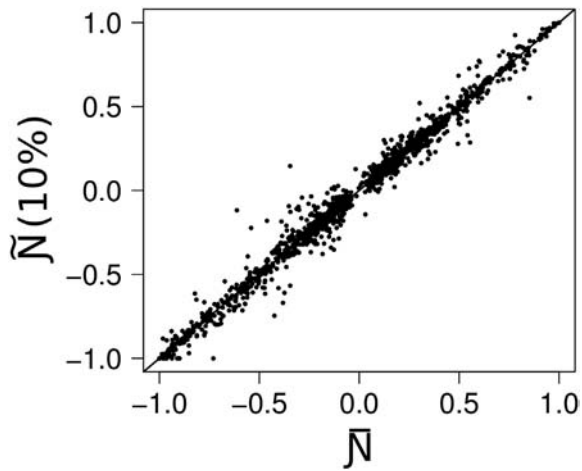


Figure 3. Relationship between  $\bar{N}$  values of 2000 simulated networks with different degree of overlap and segregation and their corresponding  $\bar{N}$  values obtained by performing a random sample of 10% of all possible node pair comparisons.

5000  $N_{ij}$  values. No limitation is imposed on the Windows, R and Python versions of the software, where users are free to select any percentage of pair comparisons to be performed.

### Examples of application

To ensure the correct functioning of the code, we tested it on the same set of networks investigated by Strona and Veech (2015), retrieving identical results. Additionally, we

performed a test on two detailed food webs (boreal and arctic) from a high-latitude marine ecosystem, the Barents Sea, available from Kortsch et al. (2015). The two networks can be downloaded in a format compatible with NOS from <http://nos.alwaysdata.net/downloads>, and hence used to replicate our analyses.

In a first experiment, we examined the networks by uploading them to the online software and running it with the default setting. This identifies the network as ‘directed unimode’ (which is correct for a food web), does not create a set of potential interactions, includes cannibalistic interactions in the computation, and performs all possible pairwise node comparisons to compute  $\bar{N}$ . Histograms showing the distribution of  $N_{ij}$  values for both networks are shown in Fig. 4A–B, while results of the analysis are summarized in Table 1. In both cases, the analysis revealed a tendency towards segregation, and a relatively high modularity. These results would suggest that the two networks are characterized by a high compartmentation, and by a reduced overlap in resource usage. However, the two networks are taxonomically broad, including detritus and bacteria, basal taxa, zooplankton, fish, sea birds and marine mammals.

We then verified if the observed lack of overlap in resource usage obtained with the standard setting (which assumes all pairwise interactions between nodes in the network are possible, such as, for example, zooplankton eating birds) persisted in a more restrictive and realistic scenario of possible interactions. For this, we replicated the analyses by using the NOS integrated feature of creating a network of potential interactions based on trophic rules. Thus, we selected ‘yes’

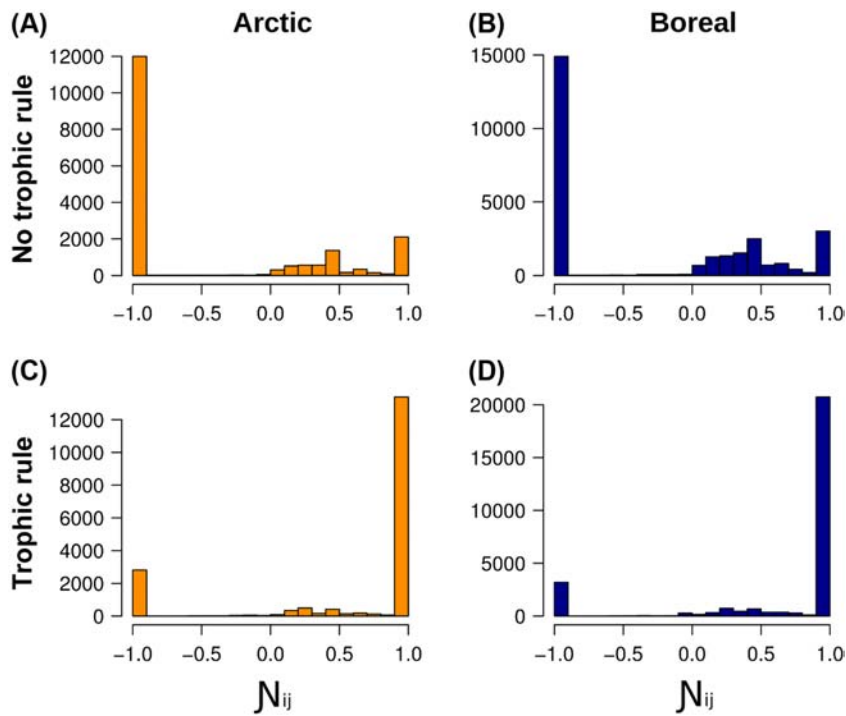


Figure 4. Histograms showing how the distribution of  $N_{ij}$  values in the arctic and boreal food webs changes when all nodes are considered as potential interacting partners (A, B), and when potential interactions are instead filtered according to trophic rules (C, D).



Table 1. Results of the tests on the two Barents Sea food webs (Arctic and Boreal) from Kortsch et al. (2015). The two sets of values indicate experiments where  $\bar{N}$  was computed assuming all nodes in the networks as potential partners for the focal nodes ('No trophic rule'), and experiments where the sets of potential partners of each node in the networks was identified according to trophic rules based on food-web structure ('Trophic rule').

|                    | No trophic rule |        | Trophic rule |        |
|--------------------|-----------------|--------|--------------|--------|
|                    | Arctic          | Boreal | Arctic       | Boreal |
| $\bar{N}$          | -0.474          | -0.291 | 0.637        | 0.694  |
| $\bar{N}_{in}$     | -0.398          | -0.412 | 0.554        | 0.552  |
| $\bar{N}_{out}$    | -0.549          | -0.17  | 0.721        | 0.836  |
| Z                  | -1.014          | -0.663 | 1.477        | 1.184  |
| p-value            | 0.311           | 0.508  | 0.14         | 0.07   |
| Mod                | 0.761           | 0.779  | 0.7          | 0.607  |
| Mod <sub>in</sub>  | 0.807           | 0.779  | 0.764        | 0.752  |
| Mod <sub>out</sub> | 0.715           | 0.779  | 0.635        | 0.461  |

in the option 'Compute potential network based on estimation of trophic level', we left the default option of 'Shortest path' to basal resource as the trophic measure, and the option of 'Threshold' trophic rule. Again, we chose to perform all possible pairwise node comparisons to compute  $\bar{N}$ . Taking into account potential interactions based on trophic levels led to very different results, revealing that, within trophic levels, resources tend to be shared by consumers and vice versa, producing a high overlap (Fig. 4B–C, Table 1). However, the networks again showed a high level of modularity, which is consistent with the notion that a modular pattern would ensure the stability of food-webs, a mechanism that was proposed, amongst others, by May (1972), and that is still a primary focus of ecological investigation (Grilli et al. 2016).

## Data deposition

Data available from the Dryad Digital Repository: <<http://dx.doi.org/10.5061/dryad.h58qc>> (Strona et al. 2017).

## Conclusions

Understanding whether species tends to co-occur at a given site more or less than expected by chance has been at the center of ecological debate for a very long time (Stone and Roberts 1990). With the emergence of network analysis in ecology the discussion has been extended to species interactions, with the question being whether species tend to share interacting partners more or less than random expectation (Strona and Veech 2015). Greater-than-expected sharing (i.e. nestedness) has been suggested to be a fundamental mechanism in promoting and maintaining biodiversity (Bastolla et al. 2009, Rohr et al. 2014).

Despite the obvious conceptual differences between site-sharing and species-sharing, this issue has been investigated with similar approaches (e.g. nestedness analysis) and it has posed parallel challenges to ecologists, generating conceptually overlapping debate. Much of this has revolved around

technical details about the metrics used to measure overlap, and the procedures to assess their statistical significance (Ulrich et al. 2009). Most available procedures are based on the underlying assumption that all species interactions (in a network) are ecologically possible, even if this may not always be a well justified approach given that constraints that limit the set of possible interactions may be important for our understanding of ecological network structure. These constraints may emerge from evolutionary processes such as the morphological co-adaptation of species, e.g. in the case of many plant–pollinator interactions in bipartite networks, or from the hierarchical structure of species trophic interactions in food webs.

Our metric ( $\bar{N}$ ) and software (NOS) tackle this issue explicitly by providing users with complete flexibility in defining whether any particular interaction is possible in relation to specific hypotheses. Additionally, the NOS software provides a set of automated features that permit the testing of a broad range of hypotheses, even when explicit information about possible interactions is not available, with minimum effort from the user's side. We hope that the NOS software (<<http://nos.alwaysdata.net>>) will provide new perspectives on patterns and processes shaping ecological networks.

To cite NOS or acknowledge its use, cite this Software note as follows, substituting the version of the application that you used for 'version 0':

Strona, G., Matthews, T. J., Kortsch, S. and Veech, J. A. 2017. NOS: a software suite to compute node overlap and segregation in ecological networks. – *Ecography* 40: 000–000 (ver. 0).

## References

- Allesina, S. and Tang, S. 2012. Stability criteria for complex ecosystems. – *Nature* 483: 205–208.
- Almeida-Neto, M. et al. 2008. A consistent metric for nestedness analysis in ecological systems: reconciling concept and measurement. – *Oikos* 117: 1227–1239.
- Askew, R. R. 1961. On the biology of the inhabitants of oak galls of Cynipidae (Hymenoptera) in Britain. – *Trans. Soc. Br. Entomol.* 14: 237–268
- Atmar, W. and Patterson, B. D. 1993. The measure of order and disorder in the distribution of species in fragmented habitat. – *Oecologia* 96: 373–382.
- Bastolla, U. et al. 2009. The architecture of mutualistic networks minimizes competition and increases biodiversity. – *Nature* 458: 1018–1020.
- Broom, B. M. et al. 2010. Building networks with microarray data. – In: Bang, H. et al. (eds), *Statistical methods in molecular biology*. Springer, pp. 315–343.
- Brualdi, R. A. and Sanderson, J. G. 1999. Nested species subsets, gaps, and discrepancy. – *Oecologia* 119: 256–264.
- Cantor, M. et al. 2017. Nestedness across biological scales. – *PLoS One* 12: e0171691.
- Dormann, C. F. et al. 2008. Introducing the bipartite package: analysing ecological networks. – *R News* 8: 8–11.

- Grilli, J. et al. 2016. Modularity and stability in ecological communities. – *Nat. Commun.* 7: 12031.
- Jonhson, S. et al. 2013. Factors determining nestedness in complex networks. – *PLoS One* 8: e74025.
- Koleff, P. et al. 2003. Measuring beta diversity for presence–absence data. – *J. Anim. Ecol.* 72: 367–382.
- Kortsch, S. et al. 2015. Climate change alters the structure of arctic marine food webs due to poleward shifts of boreal generalists. – *Proc. R. Soc. B* 282: 20151546.
- Krasnov, B. R. et al. 2012. Phylogenetic signal in module composition and species connectivity in compartmentalized host–parasite networks. – *Am. Nat.* 179: 501–511.
- May, R. M. 1972. Will a large complex system be stable? – *Nature* 238: 413–414.
- May, R. M. 1973. *Stability and complexity in model ecosystems* (vol. 6). – Princeton Univ. Press.
- McCann, K. S. 2000. The diversity–stability debate. – *Nature* 405: 228–233.
- Newman, M. E. 2006. Modularity and community structure in networks. – *Proc. Natl Acad. Sci. USA* 103: 8577–8582.
- Oksanen, J. et al. 2017. *vegan: community ecology package*. – R package ver. 2.4-3, <<https://CRAN.R-project.org/package=vegan>>.
- Radicchi, F. et al. 2004. Defining and identifying communities in networks. – *Proc. Natl Acad. Sci. USA* 101: 2658–2663.
- Rohr, R. P. et al. 2014. On the structural stability of mutualistic systems. – *Science* 345: 1253–1257.
- Snow, B. K. and Snow, D. W. 1971. The feeding ecology of tanagers and honeycreepers in Trinidad. – *Auk* 88: 291–322.
- Staniczenko, P. P. et al. 2013. The ghost of nestedness in ecological networks. – *Nat. Commun.* 4: 1391.
- Stone, L. and Roberts, A. 1990. The checkerboard score and species distributions. – *Oecologia* 85: 74–79.
- Strona, G. and Veech, J. A. 2015. A new measure of ecological network structure based on node overlap and segregation. – *Methods Ecol. Evol.* 6: 907–915.
- Strona, G. and Veech, J. A. 2017. Forbidden vs. permitted interactions: disentangling processes from patterns in ecological network analysis. – *Ecol. Evol.* doi: 10.1002/ece3.3102
- Strona, G. et al. 2014. Nestedness for dummies (NeD): a user-friendly web interface for exploratory nestedness analysis. – *J. Stat. Softw.* 59: 1–9.
- Strona, G. et al. 2017. Data from: NOS: a software suite to compute node overlap and segregation in ecological networks. – Dryad Digital Repository, <<http://dx.doi.org/10.5061/dryad.h58qc>>.
- Sutherland, J. P. and Karlson, R. H. 1977. Development and stability of the fouling community at Beaufort, North Carolina. – *Ecol. Monogr.* 47: 425–446.
- Ulrich, W. and Gotelli, N. J. 2007. Disentangling community patterns of nestedness and species co-occurrence. – *Oikos* 116: 2053–2061.
- Ulrich, W. et al. 2009. A consumer’s guide to nestedness analysis. – *Oikos* 118: 3–17.
- van Rossum, G. and de Boer, J. 1991. Interactively testing remote servers using the Python programming language. – *CWi Quart.* 4: 283–303.
- Veech, J. A. 2013. A probabilistic model for analysing species co-occurrence. – *Global Ecol. Biogeogr.* 22: 252–260.
- Williams, R. J. and Martinez, N. D. 2004. Limits to trophic levels and omnivory in complex food webs: theory and data. – *Am. Nat.* 163: 458–468.

Supplementary material (Appendix ECOG-03447 at <[www.ecography.org/appendix/ecog-03447](http://www.ecography.org/appendix/ecog-03447)>). Appendix 1.