

Associate Editor Comments to Author

Three reviewers have assessed this manuscript and all are, in general, enthusiastic regarding the ideas presented. However, all reviewers brought up issues regarding the accessibility, organization, and current level of transparency in the manuscript. In particular, there is a request for better explanations of the motivation for devising your new metrics and explicitly tackling the issue of what they add to our understanding of ecology. These concerns notwithstanding, your paper has the potential to be a very nice contribution to the literature and I encourage you to resubmit the paper once these revisions are incorporated.

We would like to thank the editor for allowing us to resubmit this manuscript. As you will note by reading the replies, and the revised manuscript, we have made numerous, substantial, and extensive changes. These strive to (i) make our intent clearer, especially in the introduction; (ii) clarify the formulation of some of the measures; (iii) discuss more in depth the consequences, especially for sampling, of our work. We hope that the manuscript, in revised form, meets or exceeds the criteria for publication in *Methods in Ecology & Evolution*. Please note that all page/line numbers refer to the track-changed version of the manuscript.

Reviewer(s)' Comments to Author

Reviewer 1

The ms by Poisot et al. addresses an important issue, that is, how to handle the variable nature of interaction networks. The authors propose to adapt a probability framework to describe interactions and thus propose a set of adaptations of known network properties to such a framework. In general I find the ms well-written, well-structured and well-argued. However, I do have some issues that I think needs to be addressed.

My first and overall concern is in respect to the accessibility of the proposed method and its use in practice. I think it would be important to explain the use of the software mentioned. Personally I am not familiar with the julia language and I expect that this presentation of the software would appear rather inaccessible to many ecologists. More practical information is needed in this regard, if not in the main text, then perhaps as a supplement (I did not find any supplemental material for the submission).

We appreciate this concern, but since the reviewer decided to give it that much emphasis, we have to raise two points. First, this is not an application paper, and we do not think that the code should

impact the evaluation of the merit of the methodological advance we accomplished. Second, the reviewer is not criticizing code quality, but language choice, which we do not think is a relevant measure of merit. Nevertheless, we have added a short section at the end of the paper, with a link to the documentation and user guide. The code we provide is thoroughly tested, automatically validated, and documented. We are confident that all of the elements are here to allow a motivated researcher to apply our methods to their data.

Second, I think the authors need to guide the reader a bit more. What precisely is the general difference between the proposed formulas and existing ones to calculate network properties? For instance, why can the original metrics not be readily used in the probabilistic framework. Try to explain this in a sentence or two to the (perhaps skeptical) first time reader.

We have added a few sentences at **X** and **Y** to emphasize on the differences. It was our impression that the first two sentences of the original second-to-last paragraph of the introduction established the need for probabilistic measures clearly: “The current way of dealing with probabilistic interactions are either to ignore variability entirely or to generate random networks. Probabilistic metrics are a mathematically rigorous alternative to both.”. As it was obviously not sufficient, we have added a sentence before, to explain that a modification in the methodology is needed.

Third, the overall important question of how to estimate the probability of interactions remains unanswered. I am not saying that the paper should find the solution to this problem but perhaps some more discussion of the practical implications of the framework is needed. At page 17 this is briefly discussed but leaves the reader with a lot of uncertainty and the message of the paper becomes very theoretical. Will we be able to reliably estimate the probability of a given interaction, e.g. between a pair of species with no existing interaction data? Would for example information on traits and phenology of species be enough to provide a good estimation of the probability of interaction between a species pair? Abundance of species would doubtless be important as well. This brings me on to my next point, that the probabilities of interactions are considered to be independent in the proposed framework. Why is this necessary? In real ecological communities interactions are not independent, so I think this point needs to be better explained/addressed in the text.

We agree with the referee that this paper should not solve the question of how to measure interaction probability. This is a question for people carrying out empirical measurements to decide. Nevertheless, we have added an entire section to the discussion, at **X**, to give some insights about when interactions are independent or not, and how they can be estimated.

How does the method relate to weighted networks? One argument for the probabilistic approach is to move beyond the description of interactions as either/or (0/1). But this is already being done with the use of weighted metrics. It would be relevant with some discussion on weighted vs probabilistic networks.

We agree with this point entirely. The issue is that the existence of an interaction (which we work on here) may not be related in trivial ways to its intensity (the weighting that the reviewer mentions). We give, at **X**, arguments for why the question of interaction weight should be modeled separately from the question of interaction existence.

Please explain better the use of Bernoulli trials.

We have added a clarification at **X**.

Improve/clarify terminology: On P13L17 it is stated that random networks are generated using Bernoulli trials. It is stated that these are binary networks (0/1) but are these not supposed to hold probabilities (A_{ij}) instead of 0/1? Then in the table (which has no number?) the authors distinguish between Binary, Bernoulli trials, and Probabilistic. It is not completely clear to me what is what.

We have added a clarification at **X**. To some extent, 0/1 networks are probabilistic networks, where all non-zero interactions happen with probability unity. This is a strength of our approach, that all types of data can be analyzed. We have also emphasized this point in the introduction.

P5L11-12: So these metrics are only applicable on unipartite networks? Please make this clear to the reader.

This is, in the original submission, made clear to the reader, in the second paragraph of “Suite of probabilistic network metrics”, where we explain how to make bipartite networks unipartite in case the measure requires it. We have rephrased this sentence to make it clearer.

P6L1-4: Not clear to me what is meant. What is “the exact value of the non-probabilistic measure”?

We have rephrased.

P6L19: Capital E.

Changed.

P8L20: Please choose a more appropriate/original reference on modularity.

Especially in the absence of suggestions, we think this reference is appropriate enough.

P9L8: measure without s

Changed.

P9L17: What does this sentence mean starting with %DC?

This was a typo and has been corrected.

Equation 10: I don't understand this. How can C_i be equal to C_i / C ?

We have clarified the text. This is also the original formulation in Katz's paper.

P10L3: $In = is$

We respectfully disagree.

P10L10-11: I would think that a species with no outgoing links would be a top predator (if edge direction shows the flow of energy)?

The convention is usually to show links as interactions, so species with no outgoing links establish no interactions (but it is true that some authors represent links as flow of energy, in which case the direction of the "arrow" is reversed). In any case, as we now explicitly mention, one should be cautious not to interpret the absence of outgoing links as the sign that a species is a top predator in a probabilistic context. We thank the referee for raising this point.

P14L20: a = of

Changed.

P15L2: acquired?

Changed.

P17L12: Cafeteria experiments?

We have added an explanation.

Figure 1: An explanation of the axes would be nice to have in the legend.

We have moved the description of axes from main text to the legend, and apologize for having made this choice in the first place.

Reviewer 2

I have read the manuscript “The structure of probabilistic networks” by Poisot et al. in which the authors derive probabilistic versions of current metrics of interaction network structure and provide examples for their implementation. The major innovation being made by the authors, both in developing these probabilistic network metrics and the analytic framework which inspired them (e.g. Poisot et al. 2015), is to shift the focus of network analyses away from characterizing static interaction data, and towards characterizing models of interaction probabilities. Although the authors are not the first to use probabilistic modelling approaches, they have developed a promising and mathematically unified framework elsewhere, and the metrics described in this ms are a cleverly constructed toolbox for describing some aspects of network structure in future studies using this framework. The article is generally well written, but lacks some background that would help clarify why the metrics being proposed are an important contribution. It also seems like some results are either missing, or are only described qualitatively (and briefly) in the main text. Below, I outline several questions I have regarding the background given, the presentation of the Applications in the ms, and potential limitations of the proposed metrics.

MAJOR COMMENTS

1. More background information needs to be given in the introduction regarding the conceptual framework motivating the development of these probabilistic metrics (this has been developed elsewhere: e.g. Poisot et al. 2015, Oikos). Very little of this background is provided in the introduction, and is only briefly discussed in the ‘Implications for data collection’ subsection at the end of the ms. This made it difficult to follow how one would generate a probability model that these metrics could be appropriate for, and therefore why we would need probabilistic metrics in the first place. An in-depth exploration of this background isn’t necessary or appropriate, but a clear and concise explanation should be given to help clarify the motivation for developing the metrics described in the ms. To improve clarity, I would also recommend that the examples provided in the ‘Applications’ section start by describing how the probability adjacency matrix

A is being estimated from data, and stating specifically what aspects of these metrics' performance each example is demonstrating.

We have reworked both the introduction and the "data collection" parts, and we are confident that the revised manuscript introduces both the concepts and the need for a new methodology in a more convincing way. We have added the required information on the estimation of initial adjacency matrices at \mathbf{x} .

2. Can the interaction probability models, and the associated metrics, describe the probability of interaction frequencies/rates? That is, can they describe variation in quantitative network structure? Throughout the ms, the authors refer to binary networks, but never state explicitly whether their methods are limited to the binary case, or if they can be generalized to deal with adjacency matrices describing variation of interaction frequencies/rates. If so, this needs to be clarified and explained in the ms (e.g. what constitutes a Bernoulli trial when estimating interaction probabilities from frequency data?). If not, this raises a potentially major limitation for the methods/metrics being proposed that should be addressed. There is an extensive literature on dynamical consequences of variation in interaction strength for ecological networks, including the development of many metrics of quantitative network structure. If extending these models/metrics to the quantitative case is a natural and intended next step, it might be worth either including them in this ms, or mentioning them as a natural direction for future development.

As we explain in the reply to referee 1, we do not think the frequency/rate of interaction is amenable to our framework (and we provide a rationale why at \mathbf{x}). This being said, we agree that this is a clear next step for research, and identify one potential approach from species distribution models to pursue it.

3. In the introduction [p.3, L.2-12] the authors cite studies which have demonstrated that many of the same network metrics they re-derive can be predicted by probabilistic and likelihood based models incorporating minimal information about species relative abundances and trait distributions. The authors interpret this as evidence that interactions should be represented as probabilities rather than binary or count data. However, another (but by no means mutually exclusive) interpretation is that current metrics are not particularly informative about the biology underlying species interactions. Why then should we expect the same metrics to be informative when derived for calculation directly from probabilistic models? What new insights about species biology do we gain from these metrics when calculated for Bernoulli probability matrices? In a field that is burdened with many redundant metrics, an article describing a suite of re-derived metrics should provide clear reasoning for why we need more.

We have edited the introduction to show why these measures are needed – the probabilistic approach to networks is gaining traction, as is clear from reading the literature, and the current metrics are not able to use this information. As we mention in **Y**, the measures we propose here are not new, but rather they are a re-expression of well-established ones in a probabilistic context. Taking a step back, we disagree with the idea that measures are not informative about ecology. The papers that make this claim usually overlook broad swathes of the literature, that demonstrate clear links between network measures and community properties. This is an idea we have no intent of publicizing further.

MINOR COMMENTS

1. assumption of invariant interactions and monte-carlo techniques as currently unsatisfactory methods for dealing with inherently variable species interactions. The authors provide examples of problems that can arise when using invariant networks, but they don't make a clear case for why current null-model/monte-carlo approaches, which are firmly rooted in probabilistic modelling, are unsatisfactory. Although they later provide some evidence in the 'Applications' section, it would be helpful if they could give some examples or at least foreshadow why these approaches may be problematic in the introduction.

This is correct, and we have added a short sentence at **X**, citing two recent studies that focused on these issues in ecological networks.

- 2.

Fixed.

3. data?

We have added an explanation at **Y**.

4. being reported? Sample variance? Monte-Carlo C.I.? Analytic Variance? This information should be provided, either in the column titles, or as a note at the foot of the table.

We think that this table is small enough that it can be kept inline, but will defer to the editor for this decision. We have added the required informations.

5. network ‘structure’ (2 of the nestedness metrics are non-independent, and the 3rd is a function of the other two). It isn’t necessary to report comparisons of all the proposed metrics that can be calculated for this network, but it would certainly give a better example of their usefulness and performance if more were made available (if not in the main text, then maybe as an appendix?). The authors highlight the fact that many of their probabilistic metrics have analytically derived variances, but they choose to report nestedness metrics that do not appear to have an analytically derived variance. They also choose a case study that, for the most part, results in variances of 0. It would be nice to see a more effective demonstration of this very novel capability of probabilistic networks and metrics.

We have added an example using motif counting, and discuss why, even in the absence of strong differences, the probabilistic approach should be preferred over Bernoulli trials.

6. interactions and network metrics seems like a very interesting and novel application of these methods. But I wonder if 0 or low variances, as reported in this case study, are likely to be a common result when interaction probabilities are being estimated from poorly-replicated binary interaction networks (which is likely to be the case given the practical challenges of replicated interaction sampling for ecological networks like food-webs and pollination networks).

We agree with the referee, and would rather not speculate on this in the absence of data. We added a section to the discussion, that explains how the confidence interval around each probability can be estimated from the sample size. These are, indeed, likely to be high, which reinforces the need for predictive models in addition to observational data.

7. of probabilistic network metrics in the context of null model hypothesis testing need to be clarified. As written, it is unclear what comparisons are being made, and less clear what the comparisons of interest are or should be. A couple specific points that need clarifying:
 - 1) What is ‘N’? This is only ever defined as ‘nestedness’, but does not appear anywhere in the earlier subsection describing the probabilistic measure of nestedness (identified by the Greek letter eta). Is this the nestedness (eta) of the empirical network?

The legend has been fixed.

- 2) Why are random networks being generated for this analysis (L.17)? I thought it was possible to calculate eta directly from the probability matrix corresponding to each null model. Is nestedness being calculated for these random networks, and the distribution compared to eta? If so, it is unclear where this fit into the analysis.

As we say in this part, and in the introduction of the applications section, we are interested in comparing probabilistic measures to the output of null models.

As it reads, and from Figure 1, it sounds like the authors are presenting 2 comparisons of the difference between eta for an empirical network, and a null model: one each for 2 pairs of null models (I vs. II, and IIIin vs. IIIout) (although they briefly discuss results of all pairwise comparisons in the text [p.15, L.9-12]). It is not clear why these particular comparisons are of interest, other than demonstrating that probabilistic metrics can indeed be calculated for both empirical networks and probability matrices generated by null-models.

As we now state in **y**, the key result is that the application of probabilistic measures reveal a potential bias in the null models output – we have also referred to previous literature in support of this claim.

8. There is no information given regarding the nestedness values calculated for the empirical networks. However, since all delta-N values are positive, the empirical networks that return large delta-N's must necessarily also have high nestedness values (although the reverse is not necessarily true for networks that return low delta-N values). This statement needs more explanation, or should be omitted. Given the similarity in the constraints between null models I & II, it seems unsurprising that this linear relation exists.

We have deleted this statement.

9. [p.15, L.1-3; Fig. 1]: This seems like an example of the strong (and well known) covariance between connectance and nestedness (and other metrics; Kay & Schemske 2004, Vermaat et al. 2009, James et al. 2011) rather than an illustration of why we need to re-think how null models are being interpreted. If a null-model is being constructed using constraints based on the connectance of the empirical network, it follows that it will return nestedness values that reflect this non-independence.

The existence of a covariance is reason enough to re-think our interpretation of null models, in our opinion. We have added references at other places (**x**) that support this point of view. No changes.

10. simulations in the ‘Null-model based hypothesis testing’ sub-section. Are the authors referring to their results from the ‘Comparison of probabilistic networks’ subsection? If so, they report strong agreement between the probabilistic metrics calculated directly from the probability matrices and Bernoulli trials (Table 1), and the bias-to-variance ratio between the Bernoulli Trials & Probabilistic columns appears to be miniscule for the ‘Links’ metric, even with 10^4 Bernoulli trials. As far as I can see, there is no derivation for the variance of the probabilistic nestedness metrics given earlier in the ms, and no variance reported for the Bernoulli Trials analysis of nestedness, so no comparison can be made here.

We have fixed this sentence.

11. It’s unclear what the relevance of these comparisons is, other than to demonstrate that probabilistic metrics can be calculated for Null models. It seems that a more interesting and useful demonstration of the scope of these metrics’ capabilities would be to show 1) how a probability model incorporating a combination of species relative abundances and trait distributions/experimental data can be constructed and characterized using probabilistic metrics, and then 2) compared with alternative and/or null models.

The suggestion made by the referee is far beyond the scope of this paper. We have instead added a broader explanation of what this example demonstrate at **x**.

- 12.

“respectively” – now spelled out when first used.

- 13.

We have added an explanation at **X**.

14. [p.16, L.20]: According to their methods, Olito & Fox (2014) performed temporally, but not spatially replicated interaction sampling.

We agree, but we believe this is a reasonable time-for-space substitution.

15. [Figure 1]: Models are identified using H-A numerals in axes titles (1, 2, 3i, 3o), but roman numerals (I, II, IIIin, IIIout) in the text. A minor issue, but these should be consistent.

This has been changed.

16. published and volume/page numbers (e.g. Poisot et al. 2014 -> 2015)

These have been changed.

TYPOGRAPHICAL/GRAMMATICAL

[p.3, L.24] ‘In the recent years...’ -> ‘In recent years’ instead of

[p.9, L.3] Comma does not need separate line.

[p.13, L.11] Sentence is missing a subject. Consider restructuring.

[p.15, L.24] ‘...does not depends...’ -> ‘...does not depend...’

[p.16, L.9-10] Not sure what is meant by ‘... the error is overall sensitive...’.

[p.17, L.28] ‘...emergent...’ -> ‘...emerging...’?

[p.20, L.14] Citation is missing volume & page number.

Reviewer 3

Below, I clarify some questions and concerns raised during the revision and point out some modifications or clarification which could be useful to improve the manuscript.

In this study, authors propose a suite of networks metrics (aggregated statistics) describing interactions as probabilistic events (Bernoulli distribution). They also provide a trial software in “julia” language where these metrics can be calculated. They also propose and applied two methods by means of which is possible estimate metrics from: 1) probability of interactions directly measured (fieldwork observations or lab experiments) or 2) random networks created using null models. The main merit and novelty of this study is sited on the effort to improve classical metrics by incorporating the natural variation in interactions, which is an important step for ecological networks. The mathematics behind these metrics seems consistent to me, and most equations were sufficient to understand the authors rationale (If they were not clear enough, I tried point out in detail below). Despite of be an interesting contribution, several improvements should be taken into account to have a manuscript suitable for publication in MEE.

I would highlight: 1) notable lack of support to several statements; 2) important concepts were overlooked; 3) some references were ill cited; 4) in “applications”, examples covers just a small fraction of the proposed metrics and potential sources of bias were not investigated; 5) the manuscript is dense both in the amount of information and equations, so simpler synthesis are necessary, 6)

applicability of each metric and the null models to each situation is not being explored clearly enough and 7) language “julia” is not often used by ecologists which will limit the applicability of these metrics.

We have provided answers, and made changes where necessary, to these points, as we detail below.

INTRODUCTION

Overview - Several statements are poorly supported by evidence. Several sentences are unclear and lack of references make some arguments weak. Self-citation is abundant and sometimes they are in unsuitable places. Some important literature in the subject (using approach related to probabilistic matrices) are missing and then some important concepts were overlooked. I specify below some of these concerns, which I believe to be useful to improve the manuscript:

Page 3. Line 8. I strongly suggest removal of the citation of Poisot 2012 here and throughout the manuscript. Why to cite a literature in French and published in a local journal, if there are several others more suitable papers? For instance, Schleuning et al 2014 *Ecography* 38 and Bascompte 2009 *Science* 325.

There has been a mistake when compiling the paper, for which we apologize. The correct paper (Poisot et al. 2012, *Ecol Lett*) is now cited in the text. This is a valid reference to the claim that interactions, and hence networks, vary, since it is in this paper that we develop the first systematic attempt to quantify this variation.

Page 3. Lines 8-10. authors mentioned “rich methodological literature” but are citing no one. Please, provide evidence to support your statements.

We are providing references to the methodological literature throughout the paper, whenever measures are introduced. Duplicating the references here would be awkward and unnecessary. We have added a general reference for ecology.

Page 3. Lines 11-23. There is no citation in this paragraph. Mathematical concepts are used but references are not provided to help reader to understand, specially the two last sentences. Here, argumentation also is quite confused.

We are not sure what to make of this comment since there are no concrete suggestions to go with it. No changes.

Page 3. Lines 66-28. “This happens because, contrary to a long-standing assumption of network studies, species from the same pool do not interact in a consistent way (Poisot et al. 2012)”. Is this a hypothesis or are there evidences? References which could be considered as empirical evidences are Maruyama et al 2014_Biotropica and Trojelsgaard et al 2015_Proceedings B.

Evidences are presented in Poisot et al. 2012 – as well as in the remainder of this paragraph. We have nonetheless added these references.

Page 3. Whole third paragraph. I suggest that authors explore the “forbidden link” concepts here. Turnover in interaction are determined by spatio-temporal overlaps but also non-distributional factors, such as morphological mismatching; see Vázquez et al. 2009_Ecology, Maruyama et al. 2014, Vizentin-Bugoni et al. 2014_Proceedings B, or other factors, e.g. Renoult et al 2015_Oikos. Also in relation to the relative contribution of neutral-based versus niche-based processes, two studies (Maruyama et al. 2014, Vizentin-Bugoni et al. 2014) have shown that niche based-processes (represented by yes-no events!) can be better than abundances in predicting interactions.

We have added some of the references suggested. We do not think that re-reviewing the literature here would be useful, since we have essentially done that in Poisot et al. 2015. In addition, “forbidden link” is nothing more than a catchy term for the consequence of mismatches in phenology, which we already mention.

Page 4. lines 13-24. “The current way of dealing with probabilistic interactions are either to ignore variability entirely or to generate random networks. Probabilistic metrics are a mathematically rigorous alternative to both.” Here, again there is no citation. I particularly agree with most arguments but references should be provided.

It is our hope that this manuscript will, in time, become the reference for this statement.

Page 4. lines 13-24. “The current way of dealing with probabilistic interactions are either to ignore variability entirely or to generate random networks. Probabilistic metrics are a mathematically rigorous alternative to both. When ignoring the probabilistic nature of interactions (henceforth binary networks), every non-zero element of the network is assumed to be 1. This leads to over-representation of some rare events, and increases the number of interactions. An alternative is to consider only the interactions above a given threshold...”. Why describe networks as quantitative matrices (i.e. frequency of interactions) is not a better solution to over represented rare events? By now, I disagreed with authors.

As we now explain in **X**, the issue of intensity and existence of the interactions are different ones.

Page 5. line 1-2. Great that it is a free and open source-source, but why “julia”. Most “network ecologists” are familiarized using R or Matlab. In this sense, I guess the receptivity and usefulness of this metrics will be very limited until they be available at some more popular languages. Would be excellent if someone could to create a package, perhaps linked to bipartite package in R.

The code is open source, and comes with a full documentation. Anyone with sufficient interest can port it to R if needed.

SUITE OF PROBABILISTIC NETWORK METRICS

Overview - The mathematics behind these metrics seems consistent to me, and most equations were sufficient to understand the authors rationale. Please, review carefully this session to make clearer when calculations are suitable for bipartite or unipartite networks and also should highlight mathematic differences in each situation (e.g. connectance) or the strong differences in biological distinct meaning (e.g. centrality measures, vulnerability).

We have edited the measures description throughout for clarity.

Should be interesting discuss if these metrics are useful for quantitative networks instead just binary ones?

We now give an explanation of why we think this is not the case, at **y**.

This session is very dense an focused in mathematics behind metrics. To reduce confusion in this strongly suggest authors to synthesize information in a table which should have: Metric’s names, their verbal definition, biological meaning (non mandatory), what class of network are applicable (bipartite or unipartite and binary or quantitative if possible).

We do think this would not improve the paper. The description of formula is the core of the paper, and this would result in a massive table, without any room to explain what the parameters means.

Page 5. line 16. Authors assumes that all interactions are independent. Seems important authors return to this assumption in the Discussion, and explain implication of this on the realism of the estimations. They just criticize previous studies assuming the same thing.

We have added a section to the discussion, at **X**.

Page 5. line 17. On “generality” and “vulnerability”. Seems necessary authors describe shortly the biological meanings of both metrics (or illustrate with some ecological example), since they do not come back to them in the discussions. The same for several other subsequent metrics like “Path length” and “Centrality”

We do not think that there is enough room to describe the ecological interpretation of each of there (very classical) measures of network structure.

Page 7. line 3. In order to avoid confusion, I suggest give other symbol to Species instead “k”, which was already used do “node degree”.

It is commonly held convention to use i , j , and k to denote elements that are iterated over, and we do not think this creates any confusion here.

Page 7. line 14-15. I suggest author illustrate with some ecological situation as example.

Again, we are confident that network ecologists will be aware of what these measures entail.

Page 8. Authors should identify the value of range both for nestedness and modularity.

The range of nestedness was already identified. We have added the range of values for modularity.

Page 10. line 17. What means “%DG: to the degree or generality?” ??

Fixed.

Metrics described from page 6 to page 10 (line 2) are widely used in non-probabilistic version, more intuitive and relatively ease to measure from observed networks. However, it is not clear to me, what is which data these metrics (“Species with no outgoing links”, “Species with no incoming links” and “Number of species with no interaction”) are based on. Furthermore, biologically what is the realism of a predator with no preys in the community, for example? Authors tried discuss a bit these issues in page 11 (lines 7-10). Perhaps authors should provide examples and also describe better these metrics in regard of their applicability to bipartite networks.

We have added a number of clarifications, see **x** and **y**.

Page 12. line 4. “Element-wise product” operation is also known by Hadamard product (for instance, see Vizentin-Bugoni et al 2014). Please use both terms to alert readers.

We have made this modification.

APPLICATIONS

Page 13. Lines 15-23. Please, define (and/or exemplify) what are naive and non-naive approaches. Furthermore, it is not clear the implications of the assumptions of independence between interactions on authors’ approach. Whether elements (nodes/species) are out of the estimated network of course this will affect metrics... Is it what authors want to say?

We have clarified this part.

Page 13 line 24. On “Comparison of probabilistic networks”. So, does researchers need some a priori measure of probability of interaction (as Poullain et al, 2008) to calculate the metrics proposed in the manuscript? In page 6 (line 1-2) authors states that it is applicable on binary data. Please, clarify in the beginning of you paper in which kind of data this metrics can be applied precisely. Some drawings or tables could be provided (in the Supplementary Materials?) to clarify this confusions. This confusion is making hard to follow the paper.

We have modified the introduction of the paper (**X**) to make this point clearer. As it said in the original submission, these measures *can* be applied on the binary networks, but they will work on probabilistic networks too. This is essentially a universal toolkit.

Page 14. About the table. Is there an observed network, right? So, the metric calculated values should be also provided. It is important to compare with the estimated values and could show some potential incongruence in the estimations (for instance, should be impossible the estimated number of links to be smaller than in the observed network).

The calculated values, as now explained in the paragraph immediately before the table, are given in the *Binary* column. In the contrary, it is expected that the estimated number of interactions using probabilistic data is lower than using binary ones; the only situation in which the two quantities should be equal are when all interactions have a probability of unity to happen.

Page 14-16. In regard of “Null model based hypothesis testing.” In the Fig.1, basically author calculate the delta (difference between observed metric and the expected by the null model). What is the difference between their approach and those already used by other authors (For instance, Schleuning et al. 2014_ELE)? In which extend, this study is overcoming the previous ones? This session do not seems novel. It is hard to understand the link between these analysis and the methods they are proposing. Authors highlighted: “In short, this analysis reveals that (i) the null expectation of a network property under randomization scenarios can be obtained through the analysis of the probabilistic matrix, instead of the analysis of simulated Bernoulli networks;(ii) Different models have different systematic biases, with models of the type III performing overall better for nestedness than any other models. This can be explained by the fact that nestedness of a network, as expressed by Bastolla et al. (2009), is the average of a row-wise and column-wise nestedness. These depend on the species degree, and as such should be well predicted by models III.” It is not precisely novel, most network papers have used null model analysis and made estimation of “deltas” analysis. The point to included this session should be improved to make clearer.

We have added a justification of the novelty at **Y**. Although it is true that most, if not all, recent network papers used some sort of null model, none that we know of explicitly treated the template matrix as a probability matrix from which properties can be measured, but instead resorted to simulations. As we now explain, this approach has some biases that cast a serious doubt on their relevance.

Page 17. line 1-3. Why authors used nestedness based in Bastolla et al (2009) instead of NODF of Almeida Neto et al. (2008)? NODF and wNODF are much widely used.

The eta of Almeida-Neto et al have also some moderate biases, which have been corrected by Bastolla et al. The two measures are often equivalent, but we have a higher confidence in eta. We have added this information at **X**. Wide use is no argument for quality, sadly.

IMPLICATIONS FOR DATA COLLECTION

Page 17. Lines 9. Replace all Olito & Fox (2014) citation by the original reference: Vázquez et al. 2009_Ecology.

We think that Olito and Fox 2014 is a more appropriate reference since they use a model that is far more elegant and predictive than the one used by Vazquez et al.

Page 17. Lines 5-18. Related to use of frequency of interaction as measures of abundances (which is a often used procedure in network paper) please consider

the issues and potential biases pointed out by Vizentin-Bugoni et al. (2014). Vizentin-Bugoni et al. (2014) is a more suitable reference here than Olito & Fox (2014) because they used abundances measured independently from the observed network and also simulated potential bias arising from use of frequency of interaction as a proxy for species abundance. The original reference, Vázquez et al. 2009, do not discuss these issues.

We do not think that this would improve the paper – we have also, in addition, cited the Vinzentin-Bugoni 2014 paper, which the referee apparently greatly enjoyed, several times. No changes were made.

Page 17. Line 14-17. “On the other hand, “cafeteria experiments” give truly independent probabilities of interactions; even a simple criteria, such as the frequency of interactions when the two species are put together, is a way of estimating probability.” Please, explain better what do you mean by “truly independent”. In a real community species are influencing reciprocally interactions of other species in several ways (competition, facilitation, etc.). This cannot be neglected.

We agree with this statement, but have provided arguments for why our assumption is both reasonable and the only way to move forward at the moment, at **X**.

Page 17. Line 14-17 “Using the approach outline by Poisot et al. (2014), both sources of information (species abundance and the outcome of experiments) can be combined to estimate the probability that interactions will happen in empirical communities.” Why not other mechanisms, as those proposed by forbidden links concept (Olesen et al 2011_Proceedings B), can be equally or more important than species abundances (for instance, Vizentin-Bugoni et al. 2014 and Maruyama et al. 2014, Olito & Fox 2014)?

We have added a reference to trait distribution (which is the mechanism underlying forbidden links).

Page 17. Lines 27-28. Please, provide support (references) for these statements. That is a recurrent problem through in the manuscript.

These are discussion statements that summarize the manuscript. We feel this paragraph, and the manuscript as a whole, are sufficiently referenced.