

Dear Dr. Hocine Cherifi,

Thank you for editing our manuscript. Also, we sincerely thank the reviewers for carefully reading and commenting on our work. The manuscript is better because of it.

Below, we address the reviewers' comments point-by-point; we have updated the main text to reflect our responses.

We hope that the new manuscript satisfies the reviewers and can be published in Applied Network Science.

Yours sincerely,

Ulf Aslak, Søren Føns Vind Nielsen, Morten Mørup and Sune Lehmann

## **Point-by-point response to the reviewers' comments**

### **Response to Reviewer #1**

1. It is not clear what the authors mean under "need not be modeled at the node-level". Especially in the analyzed data set / application field. A short explanation or enlightening example would be helpful.

The context, where this wording occurs:

"However, an important limitation of this method - and to our knowledge all other methods for community detection in temporal networks - is the assumption that inter-temporal dependencies between states of the system need not be modeled at the node-level."

We thank the reviewer for noticing this problematic sentence. We revised the wording in the original sentence and added an extra line to clarify. It now reads:

"However, an important limitation of this method – and to our knowledge all other methods for community detection in temporal networks – is the assumption that dependencies between individual nodes across time can be modeled as the dependence between entire layers. Real systems with multiple asynchronous concurrent events must have varying dependencies between the same layers, and by ignoring these node-level dependencies important temporal dynamics are washed out."

The added sentence should also clarify what was previously meant by *node-level dependencies*.

2. Temporal networks:

Most of the parameters that define the temporal networks are discussed. Some of them are not. Providing some arguments, why the authors have chosen that particular values would help other researchers to apply the NFC method with optimal parameters.

- \* window size:  $w=25$
- \* activation-threshold=0.6,
- \* link-weight-threshold=top 1%
- \* subject-independent link density

Similar arguments as presented in the last page for the the relax rate and similarity threshold would be sufficient with a remark whether the results are sensitive or robust to small variations in the parameters.

We have added explanations in the *Materials and methods/Temporal networks* section that address our parameter choices and clarifies the consequential limitations when appropriate.

3. The following statement was not clear for me: "a node which at two times are embedded in neighborhoods"

In temporal networks, a node can (and usually do) exist at multiple timesteps. In the time-aggregated network nodes are often referred to as "physical nodes" and in the time-resolved temporal network each node at each time is a "state node". We did not emphasize this terminology in the paper because we felt it might distract from the main line of reasoning and also the method it relies on (NFC) and its accompanying paper is very clear on this. So if this was well established terminology in the network science literature, we would have written "two state nodes of the same physical node" or simply "two sibling state nodes". We still feel that establishing the full terminology is not a good solution, so for clarity we have revised the entire sentence and broken it into two:

"A low threshold allows two states of a node, residing in different temporal layers, to have an interlayer link that connects them even when they only have weakly similar neighborhoods. Reversely, a high threshold sets a high requirement for neighborhood similarity for interlayer links to be created between two states of a node."

We hope the reviewer agrees that this reads more clearly and we thank her/him for pointing out the issue with our initial formulation.

4.

- (i) Figure numbering is a bit confusing. They should follow the order as they appear in the text.
- (ii) Description of Fig 1,4,5 should refer to Fig 3 for explanation of color coding.
- (iii) The text in section "Task-synchrony reveals two prominent functional networks" refers to Fig 2 too late, and not in the first sentence which first refers to it.
- (iv) What is show with white violin plots in Fig 1A? One can guess it from the referenced literature, but it would help the reader if the caption or the referring text would explain it.
- (v) In Fig 1C the Community 0 and 1 seem to have some weak overlap and are not disjunctive as stated in the manuscript. Or the threshold value was chosen so high, that - based on expert opinion - they are not overlapping in thresholded components?

- (i) The confusion was likely caused by references to supporting figures. The referenced supporting figures do not belong in the main text as they are quite big, but we must reference them throughout the text. To clear up the confusion we now renumber supporting figures with the prefix "S". Furthermore, we have moved Figure 3 up so that it is now labeled Figure 1. Now, main text figures are references in order of appearance, and, in a separate enumeration, supporting figures are too.
- (ii) We have added references to the Fig. 3 (now Fig. 1) legend in the mentioned figures.
- (iii) The first sentence in "Task-synchrony reveals two prominent functional networks" did in fact not refer to Fig. 2 but to Fig. 1 (old labeling). It was meant to indicate that one can actually see that the modules seem to follow the different experiment tasks. This observation is important as it motivates the consequent quantitative search for the correlation strength between communities and tasks, and the reviewer's confusion is a strong signal that we should be more clear in our writing. We have now inserted a reference to Fig. 1 (now Fig. 2) in the sentence in question so it reads:

"From visual inspection of Figure 2A, it is clear that the activity patterns of communities 0 and 1 are modulated by the experimental tasks that subjects undertake."

- (iv) It appears we indeed never explained the white violins, and we thank the reviewer for noticing. We opt for calling them "strips". We have now committed an extra sentence to the caption of Fig. 1 (now Fig. 2):

"Each row pertains to an individual subject and visualizes the temporal dynamics of the community as a white strip of varying height."

- (v) Figure 1C (now 2C) visualizes in what fraction of community occurrences each component was active in the given community. Indeed, it shows that the two largest communities share some component activity. We understand that this can be confusing given our statement "Communities 0 and 1 are distinctly different...", and have therefore changed the sentence to "Communities 0 and 1 are, *while not entirely disjoint*, distinctly different", because this aligns better with what we intended to communicate. Different functional brain networks are expected to share some components, and Neighborhood Flow Coupling is in fact built to allow overlapping communities.

On the matter of thresholding: The overlap occurs mostly on components with high indices. In *Materials and methods/ICA* we state that we "rank the components by their temporal consistency over subjects". Components with low rank (and high index) are therefore temporally more spurious and likely to, when a community is active, randomly correlate with core community components. The reviewer keenly observes that (if we interpret his/her comment accurately) this may indicate the correlation threshold was too high. This is sensible reasoning and it is actually likely that using a lower threshold will remove some of these noisy correlations, however, brain fMRI data inherently has a relatively low signal-to-noise ratio (often close to one, see Lieberman and Cunningham 2009) and the 1% threshold we chose is at the low end of the spectrum compared with the literature. So while reducing the threshold would reduce the noise, it would also remove a lot of signal, and we would simultaneously challenge state-of-the-art approaches. Indeed, if we try lowering this threshold to values like 0.5%, 0.6% and 0.7%, we find that the consistency across subjects, that we otherwise observe, is lost.

5. The task performance and the synchrony was correlated in the manuscript for testing the hypothesis that subjects performance was in connection with their brain activity. Though, the performance of a subject depends heavily on the abilities of the subject. What was the relation between the performance gain (difference or ratio of performance 0-back and 2-back) and the synchrony?

We had a strong expectation that there would be a significant relation between task performance accuracy and task-community synchrony, so we were surprised to observe there was in fact none. In our discussion we reason that this may be due to the small variance in performance across subjects, and further hypothesize that even performing poorly may require significant focus on the task.

But the reviewer raises an important question we did not consider. We test his/her query—where gain is 2-back performance divided by 0-back performance—against the data and find no significant correlation (see Fig. 1). Computing gain as the difference between 0-back and 2-back, rather than the fraction, also yields no significant correlations.

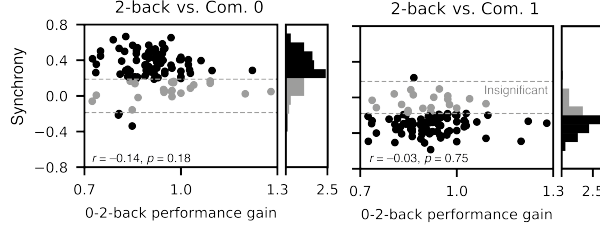


Figure 1: **Task performance gain versus community-task synchrony.** Gain is computed as 2-back performance divided by 0-back performance.

We choose not to include this figure in the paper, but since it is an important test we include the following sentence:

"Finally, as a way to correct for individual subject ability, we test the relation between performance gain (2-back performance divided by 0-back performance) and synchrony, yet find that it is not significant. We therefore reject the hypothesis that community-task synchrony is indicative of performance level and vice versa."

## Response to Reviewer #2

1. This is a well-written manuscript and the application of novel methods for temporal community detection in functional brain networks is potentially of great interest to the neuroimaging community.

We thank the reviewer for this kind remark.

2. My primary concern is that communities #2 and #3 contain only a very small number of components relative to communities #0 and #1. This must explain why these communities are virtually never activated compared to communities #0 and #1 in all subjects? What is then the relevance of these communities if they contain such few nodes and are never active as a result? The authors state that they do not search for parameters in any rigorous way, but rather choose a corresponding parameter set that gives meaningful results. This is a reasonable approach in the context of a proof of concept study, but I am not convinced that communities #2 and #3 carry any meaningful neurobiological information whatsoever, as indicated by the fact that they are neither activated in the resting state nor the task state. In my view the results would be much strengthened if communities #2 and #3 were of comparable size to the other two, and similar conclusions were still reached. I would strongly suggest that the authors explore this avenue.

Among the authors, we did discuss this problem, so are happy to have the reviewer's perspective. In our "non-rigorous" parameter exploration, we never recovered solutions with more than two communities exhibiting consistency across subjects. In spite of this, we still chose to keep communities #2 and #3 in the figures because we felt it was the most honest

way to present our findings. In the main text, we commented:

"... we also show plots for communities 2 and 3, but there are no trends that distinguish them from noise."

With fresher eyes, however, we do feel the need to revise our presentation. Therefore we have decided to entirely remove communities #2 and #3 from the main text figures, and instead elaborate further. The appendix figure displaying communities of all 100 subjects (Fig. S2) is left intact so the reader can get an idea of what the remaining communities look like. In our opinion this does away with some of the distraction, and makes for a clearer paper. We hope the reviewer is satisfied with this edit.

3. The chosen time window for the functional connectivity analysis is indeed much shorter than what is traditionally employed in neuroimaging experiments. The authors provide a reasonable justification for doing so, pertaining to compatibility with the timing of the experimental design. While it is reassuring that some of the same communities are found across subjects, that could simply be due to the large number of elements they contain (see point #1). I therefore remain concerned that the unusually short sliding window size may have influenced the findings. A 'rule of thumb' for selecting window sizes in fMRI data is to set the sliding time window to  $\text{length} = 1/f$ , where  $f$  is the high-pass filter. Thus, with a high-pass filter of 0.008 Hz as reported in the Methods, a 'standard' window size should be on the order of 125 secs. I acknowledge that window lengths shorter than  $1/f$  min can certainly be used to detect connectivity dynamics, but in the present case the chosen window size of 18 secs seems exceedingly far of the norm. By comparison, the study by Gonzalez-Castillo et al. (which the authors refer to in the introduction) used a sliding window of 22.5 sec and a highpass filter of 0.18Hz. I would therefore be satisfied if the authors could demonstrate, that community #1 (the one present during the fixation period) can be recovered, broadly, by keeping the sliding window length identical but using a high-pass cutoff in the 0.05-0.2 Hz range, and hence a  $1/f$  ratio closer to that employed by Gonzalez-Castillo et al.

As the reviewer notes, we used a short time-window to allow comparison with the experimental design. At the same time, it is true that our time-window and bandpass filtering frequency violates the rule of thumb proposed by Leonardi and Ville. To investigate the robustness of our conclusions to this, we have carried out another ICA analysis where we changed the high-pass filter cutoff to 0.05 Hz as suggested by the reviewer. After running Infomap with NFC again, we are able to recover a community with similar dynamics to the original community 1 (Fig. 2). We speculate that the component activation threshold, which we impose at 0.6 (recall components are activated between 0 and 1), is a measure to mitigate the spurious correlations in the case where the highpass filter was set to 0.008 Hz. We have consequently made this connection clear in the text.

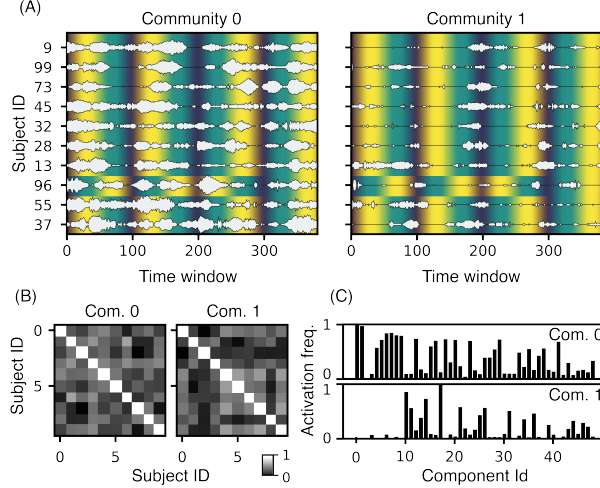


Figure 2: Revised Figure 1 (now 2), where ICA has been performed with a high-pass filter at 0.05 Hz. (B) Average inter-subject correlations for community 0 and 1 are 0.277 and 0.230, respectively. Note that since rerunning ICA with different parameters changes the order the components (which are in a post-processing step reordered by average inter-subject correlation), the distribution plots in (C) cannot be compared to those of the main text figure.

4. My last point concerns the neuroanatomical interpretation of these findings. I realize that 'Applied Network Science' does not specifically target a neuroscientific audience. Nevertheless it is difficult to make sense of these results without providing at least a 'bird's eye view' of what they could mean in the context of brain function. Without veering deep into neuroscientific territory, I would encourage the authors to address whether nodes from community community #0 (active during task performance) comprises brains regions traditionally associated with the n-back task (for which a wealth of prior fMRI studies have been conducted) and/or with working memory tasks in general.

We initially wanted this paper to be very concise and mostly focus on method. Our motivation to include Figure 6 (now S1) was that readers from neuroscience should have the necessary material to make interpretations themselves. Nevertheless, the reviewer does have a valid point. And in reviewing the core community components and matching them against brain domains that are known to activate during working memory and default mode function (using *Neurosynth*) we did actually find a good correspondence. We have now added a short paragraph at the end of *Task-synchrony reveals two prominent functional networks* which briefly summarizes our anatomical interpretation. Moreover, we have included a new figure, S4, which plots the components of each community that best match "working memory" and "default mode" maps synthesized using [neurosynth.org](http://neurosynth.org). We did not go further and estimate correlation scores between the communities (as distributions aggregated from weighted ICA components) and the activation maps for "working memory" and "default mode" keywords, because we believe the methodological steps and required documentation

(possible a new section) this would remove too much focus from the key contribution of our paper. Nonetheless, we are happy with this update which we feel strengthens the paper significantly.