**Reply to the report of the Associate Editor**

Second AE review of "netgwas: An R Package for Network-Based Genome Wide
Association Studies"

- The introduction is much more readable and has improved a lot.
- Figure (1) really clarifies the workflow of the package.
- The methodology (technical details section) has much improved and is much more readable than it was. My compliments for that! I do have some notes on the description 'sparse latent graphical model' that I would like to see fixed. See below.
- I also have some notes on the section with examples, see below.

We thank the Associate Editor for his/her comments on the manuscript. All changes to the manuscript are marked in brown in the revised version.

Detailed notes:
p2. scan for pairs of loca and *the* $r^2$ measure.
We fixed it.

p2. the methods implemented in netgwas already account*s* for
We fixed it.

P3. the covariance matrix whose its diagonal...
--> the covariance matrix, which is obtained by rescaling the correlation
matrix C with it's own diagonal.
--> Note: In this section you use three different notations for the same information:
\Omega (covarance matrix), C (correlation matrix) and \Theta (precision matrix).
I think you can just use \Omega and use \Omega^{-1} everywhere you use the precision
matrix. That way the reader needs to parse less symbols.
In the revised manuscript, we have replaced C with \Omega to help the reader to parse less symbols.

p3. falls in the interval $t_j^{(K_j-1)}, t_^{(K)_j)}$
--> superscript should be (k-1) and (k), since K_j is fixed.
We fixed it.

p3. let th*e* parameter [...] hold_s_ the boundaries
We fixed it.

p3.Algorithm (1)
--> I do not understand the loop in line 5. Apparently there is a vector Z_* (or Z_\star, the notation is inconsistent), for each (i), where the length of the vector equals the number of iterations N.
Yes, that is correct.
Note. To stay consistent, we used the notation of Z_\star (rather than Z_*).

--> I do not see why there is a superscript (m) on \Theta in Alg 1. Eqs (1) and (2) suggest that they label the EM iterations, but they do not in the alogorithm.
We have adjusted the superscripts for \Theta in equations (1), (2), and the Algorithms 1and 2. Please note that the Algorithms 1 and 2 show the E-step of the EM algorithm.

p4. The last term in equation (1) impose*s* the
We fixed it.

p4. Equaition (1), what is P and what is \lambda in P_\lambda? It is mentioned in the last paragraph of the Technical details section, together with al \lambda^* but from the description I have no idea what it means.

The term P_\lambda(.) represents different possible penalty functions. We have highlighted this in the text below the equation (1) as follow, 'The last term in equation (1) represents different penalty functions. Here we impose the sparsity by means of L_1 penalty, on the jj′-th element of the precision matrix.'

Regarding the \lambda^*, we have rephrased the description. We hope this makes it more clear.

> 'Note that the sparsity of the estimated precision matrix in Equation (1) is controlled by a vector of penalty parameter \Lambda. We follow Foygel and Drton (2010) in using the extended Bayesian information criterion (eBIC) to select a suitable regularization parameter \Lambda^* to produce a sparse graph with a sparsity pattern corresponding to $\widehat{\Theta}_{\lambda^*}$.'

p4.Eq(1) should the superscript (m) on the left-hand-side be (m+1) to indicate progression over EM iterations? Should there be a \hat{} on the Q in the lhs of Eq. (1)?
In the revised manuscript, we have adjusted the equations (1) and (2) and the text below accordingly.

p4.Eq(1) Which norm are you using in |\Theta|? Is it the L2 norm?
We have adjusted this equation (in the revised manuscript, this is equation (2)).

p5. Algorithm 2. It says initialize r_{j,j'} using Line 1 above. It is a bit confusing/inconsistent that you use Y_{j}^(i) for elements of a data matrix and r_{j,j'} for elements of the matrix R. Why not use Y_{ij} and R_{jj'}?
The data matrix Z has been represented as Z_{j}^(i) in the entire Technical details section. To keep it consistent with Z, we have used Y_{j}^(i) as well to show the elements of the data matrix Y.

p5. Algorithm 2. From the Equations (1) and (2) I do not understand how you determine \Theta by Maximizing (1). I think you need to specify an initial value for \Theta to begin with.
In Algorithm 2, Lines 1-2 initialize the conditional expectation E(z^i_j | y^(i)) and the parameter estimate \Theta.

p6. ...in map construction for RIL populations
--> First use of RIL without explanation. Please write in full.
We have included the full name as '[...] in map construction for recombinant inbred line (RIL) populations.'

p6. Also it covers a wide range of possible population type*s*
We fixed it.

p6. \rho... which corresponds to the last term in *E*quation (1) --> There is no \rho in Equation (1), and as stated above, there is only one single unexplained value \lambda, so a reader will not understand at this point that there has to be a whole sequence of nonnegative decreasing parameters. The statement also implies that the last in the sequence can be 0, is this intended?
In theory, the last value in the sequence can be zero, which it corresponds to the non-penalized inference. In Equation (2) and the paragraphs below the equation, we have replaced the \lambda with \rho to match the notations between the theory and the R code.

p6. different algorithms is --> are
We fixed it.

p6. If it is known, user can specify --> the user can specify
We fixed it.

p7. The sentence "...allows to manually select an index among a list of indexes[*] of estimated graphs holds in the res object of class "netgwasmap" to build on a linkage map" Is completely incomprehensible, please rephrase. [*] the plural of index is indices.
We have rephrased the sentence as follow:

> ‘The function buildMap() allows users to interact with the map construction procedure and to build the linkage map on the manually selected penalty term. Whereas the function netmap() selects the penalty term \rho^⋆ using the eBIC method.

The function can be called via

buildMap(res, opt.index, min.m = NULL, use.comu = FALSE).

The argument opt.index can be chosen manually which is a number between one and the number of penalty parameter n.rho in netmap() . In the default setting, the n.rho is 6. So, the opt.index can get a value between 1 and 6.'

p7. This option helps to have a clear appearance of the linkage map
--> What is a 'clear appearance'? p7. where it removes a [...] of "linkage groups".
--> why the quotes? Are they not really linkage groups? Please be explicit, and patiently explain to the unknowing reader what is happening here.
We have adjusted the sentence as
'the argument min.m is an optional argument in buildMap() function, where it keeps the clusters of markers that at least have a size of min.m member of markers. The default value for this argument is 2.'

p7. ..that can be called via
--> and can be called via
We fixed it.

p7. This function returns a "netgwas" object with S3 class
--> This function returns an S3 object of class "netgwas"
We fixed it.

p7. Thus,to select an optimal path
--> The use of "Thus" indicates that what you state there follows logically from the previous sentence. This is not the case, so it is better to start with "To select an optimal path..."
We fixed it.

p7. Thus, to overcome the limitations
--> idem.
We fixed it.

p8. Fortunately, ge*n*otype-phenotype networks...
We fixed it.

p8. For the three main functions of the package there are plot, print, and summary methods available.
--> Methods apply to object types, not to functions. So what you want to say is that 'for objects of type netgwas' there are plot, print, and summary methods available. The fact that there are three functions that can produce such an object is of no issue here.
We replace the sentence with 'For objects of type 'netgwas ' there are plot, print and summary methods available'.

p8. The parallel package is not on CRAN. It has been part of core R since R 2.14. Also, if you do cite a CRAN package, include the proper citation, which you can get with citation('packagename')
We have adjusted accordingly as follow
'To speed up computations in all the three key functions of the netgwas package, we use the parallel package (R Core Team, 2022) on the Comprehensive R Archive Network (CRAN) at http://CRAN.R-project.org to support parallel computing [...]'

p8. Include the proper citation for the Matrix package.
The correct citation is included in the revised manuscript.

p9. Figure 3 Caption. All five chromosomes *are* detected
We fixed it.

p9. In line 2 of the first example, you can use sample(ncol(tetraPotato)) to shuffle the column numbers (e.g. try sample(10)).
We fixed it.

p9. I may be missing something, but I do not uderstand why the columns need to be permuted randomly?
The function netmap() orders and groups genetic markers along the genome. For the sake of illustrations, we used a well-known *A.thaliana* and tetraploid potato genotype data, where the markers are already ordered and grouped. Therefore, we first shuffle the order of markers in both datasets. Then, we apply the function netmap() to group and order markers. Later, we compare the original order with the estimated order from the function.

p9. As stated before: I do not understand what the opt.index argument means. In fact, I do not understand what buildMap does. The netmap funcion "construct[s] a linkage map" (p6). So what is happening when I call buildMap on the output of netmap?
In the revised manuscript, we have address this issue on page 4, and we have included the following

> 'the sparsity of the estimated precision matrix in Equation (2) is controlled by a vector of penalty parameter \rho. We follow Foygel and Drton (2010) in using the extended Bayesian information criterion (eBIC) to select a suitable regularization parameter \rho^* to produce a sparse graph with a sparsity pattern corresponding to \widehat{\Theta}_{\rho^*}. '

The argument opt.index corresponds to the index of \rho^* which is selected via one of the information criterion methods, like BIC, AIC, and eBIC. To give more flexibility to the package, the function buildMap () allows the users to choose the \rho^* manually, for example based on prior biological knowledge, and estimate the grouping and ordering of markers on the estimated network from the manually chosen \rho^*.

p10. Same note on the use of sample as on p9
We fixed it.

p10. The last line of the code provides the index of the selected graph using information criteria within the map construction procedure
--> Please rephrase, because it is very hard to understand what is said here.
We have rephrased this as below

> 'In the above code, the out$opt.index shows the index of the selected penalty term using the eBIC method. If one is interested in building linkage map, for instance, on the 4th estimated network then the buildMap() function can be used as follow'. (on page 9 the last paragraph)

p10. calculates the error LOD scores
--> first use of LOD. Please write in full
The first use of LOD is in line 18th of the second paragraph in the Introduction, "[...] recombination frequencies and LOD (logarithm of the odds ratio) scores [...]".

p10. Please use 'qtl' in stead of 'R/qtl'.
We fixed it.

p12. Please explain how the user would know how to color each node. Is this trial and error? Or can it be derived from information in the 'out' object? Similar with the order of the labels in the next line of the example.
The coloring of the nodes comes from a prior biological knowledge, like a genetic map. In the case of using functions netsnp() and netphenogeno() a known genetic map can be used to color each node, where nodes (i.e. genetic markers) whose belong to the same chromosome will be colored the same. To give an example, the following code on page 11

```
R>  cl <- c(rep("palegoldenrod", 24), rep("white",14), rep("tan1",17), rep("gray",16),
          rep("lightblue2",19))
```

is specified based on the A.thaliana's genetic map, where the first 24 markers in the genotype data belong to chromosome 1. So, all the markers from this chromosome are colored as 'pale goldenrod'.