Resubmission letter

Dear Editor,

we would like to resubmit our paper "Spectrally gapped" random walks on networks: a Mean First Passage Time formula for publication in SciPost Physics. We wish to thank the two anonymous reviewers who read our first submission very thoroughly and provided valuable recommendations for further improvements, which we have taken fully on board, as detailed below.

Yours sincerely,
Silvia Bartolucci (on behalf of all coauthors)

Referee 1

The work by S. Bartolucci and coworkers provides an approximate but explicit formula for the Mean First Passage Time of a random walk between a source and a target node of a directed weighted network. The main idea behind this work is that of mapping the definition of MFPT onto the calculation of average quantities in a random matrix ensemble, in which a rank-1 deterministic $N \times N$ matrix (corresponding to the average weighted adjacency matrix) is perturbed by non-symmetric (and possibly correlated) gaussian random variables. In the case of i.i.d. gaussian variables, such an average can be carried out analytically using standard replica methods and a saddle-point calculation. The final formula for MFPT appears to be a sort of "noise-dressed" generalisation of known Sherman-Morrison formula for matrix inversion. It is worth stressing that, even in the large N limit, the final formula is approximate because it requires a sufficiently fast cumulant decay condition for large N. Such condition is usually verified in dense weighted networks and it can be phenomenologically related to the existence of a sufficiently large spectral gap between the isolated Perron-Frobenius eigenvalue and the bulk of the spectrum. As expected, the accuracy of the results obtained by the proposed explicit formula deteriorates with increasing network sparseness. The accuracy of the approximation is tested numerically with different types of random weighted directed graphs: fully-connected networks, Erdős-Rényi random graphs and random regular graphs. In random graphs with $N = 500$ nodes, the error of the approximate estimate compared to the exact one is just of the order of few percents for pretty high sparseness, e.g. average degree of $\mathcal{O}(10)$. The main advantage of this formula compared to existing approaches is that (1) it does not rely on matrix inversion, which can be cumbersome and numerically inaccurate, and (2) it is locally defined on the network. The fact that the approximate formula works well also on moderately sparse graphs (and that its limitations are pretty clearly established by a sound theoretical analysis) makes the present work an unquestionable contribution to the field of stochastic processes on random structures and complex networks.

The main result is sound and relevant, it provides a novel link between different research areas such as random matrices, disordered systems and complex networks. The paper is well organised, clearly written, with a very pedagogical discussion of the physical meaning of the formula, which turns out to be as useful for the general public as the more technical calculations. The introduction to the problem is non-technical and accessible to a broad interdisciplinary audience. Relevant literature is cited and the state of the art is explicitly discussed in the manuscript. All theoretical results are reproducible from the calculations and numerical results could be with a straightforward algorithmic implementation

of the main formula. Conclusions are clear, and in addition to summarise results, they provide some perspectives for future work. For all these reasons I think the manuscript meets all journal's acceptance criteria and strongly recommend its publication almost in the current form (some minor amendments are suggested below).

We thank the referee for their very positive assessment of our work.

Minor points:

— In Section 1.2, the sentence "The formula (5) is strikingly simple, not relying on a large-N approximation, ..." is a bit misleading. The derivation of the formula uses a saddle-point calculation with a decay condition on some cumulant; in this sense it seems to rely on N being large (on a specific random matrix ensemble). However, I understand that the formula can be applied on any graph, for finite N, and in practice its validity only requires that the spectral gap is large enough. Is this what you mean? Please, just explain a bit better this sentence in order to avoid confusion in the reader.

We fully agree with the referee, and have reformulated the sentence to remove any ambiguity, by stating that the main formula – although obtained in a large-N setting – in practice works well even for single instances at moderately low N.

— In the case of random regular graph the accuracy of the formula deteriorates considerably for low degree. In Table 2, the case for $c = 4$ is surprisingly bad if compared with the errors reported for larger degrees and to the case of Erdős-Rényi random graphs. Do you have any idea about the reason? Could it be related to the fact that in random regular graphs entries of the adjacency matrix are much more correlated than in Erdős-Rényi graphs?

We thank the referee for raising this point. We have checked again our codes very carefully, and can confirm that the poor performance of the approximate formula for $c = 4$ (with a decrease in accuracy as high as $\sim 50\%$) does occur in specific instances. However, the accuracy is highly sensitive to the specific instance of the graph at hand, as well as to the pair of nodes chosen. For RRG – all other things being equal – we could easily find instances and pairs of nodes for which the accuracy fell within just $\sim 10\%$ of the exact result, even in a "sparse" $c = 4$ case. The bottomline is that Tables 1 and 2 only provide *qualitative* evidence about how the accuracy of the approximate formula works as a function of the connectivity c, but the *quantitative* (numerical) values – e.g. the RRG $c = 4$ case highlighted by the referee – likely followed from a particularly unlucky instance. We have added a sentence at the end of the RRG Section to urge some caution and to somehow downplay the "quantitative" side of the evidence that can be gathered from the tables. It would of course be interesting to study in a more systematic (quantitative) way the statistics of the accuracies across instances and pairs of nodes, as a function of the connectivity, and we have included this task in the outlook section of the Conclusions.

— Again on the same point, a random graph of $N = 500$ nodes with an average degree of $\mathcal{O}(10)$ is not very sparse. In order to be more conclusive on the applicability of the formula on sparse graph it would be useful to show results (e.g. the scatterplots or some other metrics) at fixed average degree for different values of N (e.g. $N = 500, 1000, 5000$).

We include here in Fig. 1 a scatter plot of exact *vs* approximate MFPT on E-R graphs with fixed av-

erage degree ($c = 10$) and increasing $N = 500, 1000, 3000$, which corroborates – at least visually – that the accuracy of the formula indeed deteriorates considerably as the graph becomes sparser.

— The Authors say that "We do not report here on (dense) scale-free topologies, whose phenomenology is very similar to the other cases, albeit with significantly larger fluctuations: a detailed study of this (and other) heterogeneous cases is deferred to a separate publication." And also "While our approach continues to work for (sufficiently dense) heterogeneous networks, the intra-row fluctuations around the random matrix assumption $\langle a_{ij} \rangle = z_i/N$ may be very significant there and will thus require a more careful treatment". I do not criticise the decision of deferring the analysis of heterogenous structures to a different publication, however the present paper would really benefit of a deeper discussion on the limitations to the applicability of the formula in the case of heterogenous networks. In fact, from the fact that the formula only cares about incoming weights, one would expect that it could work pretty well when computing MFPT on hubs and central nodes and possibly much worse on poorly connected regions of the graphs. Is this correct? Is there any known result on the spectral property (spectral gap) of heterogenous graphs which prevents/limits the applicability of the formula to such graphs?

We thank the referee for alerting us on the possibility that the formula might work at different levels of accuracy on the *same* instance of a network, depending on whether a node is tightly or poorly connected with the rest of the web. To test this possibility, we have added a study of perhaps the simplest analytically solvable structure (a *star* graph) displaying great heterogeneity between the central node of degree $N-1$ (very well connected to the leaves) and the leaves themselves (having degree 1). The results – reported in the new Appendix B of the manuscript – indeed corroborates the referee's intuition that the accuracy of the formula can vary considerably within the *same* instance, depending on how well connected to the rest of the network the target node is. A more systematic study of this dependence would be very interesting and is included in the outlook section of the Conclusions.

Referee 2

Dear Editor

The manuscript deals with the interesting and important problem of the mean first passage time on random walks on networks. The main result, summarized in Eq. (5), is based on a previous work or the same authors (refs 36-37) and provides an elegant formula for the mean first passage time in terms of the transition matrix without any need to invert matrices, iterate them etc. In that sense, this work does provide a step forward in the field and a practical result. Furthermore, this formula is applicable to weighted and directed networks, which is an obvious strength. The main limitation of this work is the fact that it requires a significant spectral gap of the reduced transition matrix in order to perform well. In practical terms, this limits the applicability to very sparse networks. My main criticism about this manuscript in its current form is that it is not 100% well written, and could benefit from further clarifications and examples.

We thank the Referee for their very positive assessment of our work.

Overall, I think that this work represents an important contribution and should be published once the authors consider the following comments:

1. Right at the beginning (around eqs (1)-(2)) the authors define the adjacency matrix A and the

transition matrix T. traditionally the adjacency matrix is a binary matrix, with 0/1 entries. I believe that the authors intend to use here a weighted version of A, and this should be stated clearly. Furthermore, it would be useful to stress that A does not have to be symmetric so that it encodes directed graphs too.

We have clarified that A is real-valued and not necessarily symmetric in our setting.

2. In chapter 2 the authors use a random matrix A which, if I understand correctly, is not the adjacency matrix, but a different object - namely the reduced transition matrix. I find this notation confusing, and it would be best to replace the notation. It would also be useful to explain in section 2 that A is not the adjacency matrix.

The referee is perfectly right, and we have changed the notation $A \to H$ every time we deal with a sub-stochastic, noise-dressed matrix.

3. In Eq. (7) the authors define $\langle A \rangle$ in terms of the z_i. It is unclear at this point how the z_i are actually defined, nor what these number actually mean. The precise definition of the z_i's is given at the last paragraph of section 2, but should appear next to Eq. (7). Also, it would be a good idea to try and motivate or provide an intuition for that choice.

We have added a sentence after eq. (7) to clarify that the z_i are meant to be the row sums of the "unperturbed" matrix $\langle H \rangle$ – or equivalently, the average values of the row sums of H. The underlying idea is as follows: consider an *empirical* sub-stochastic matrix $T^{(j)}$, needed in Eq. (3). Compute its row sums $\{z_1, \ldots, z_N\}$, where $z_i \leq 1$. Now, form a "perfectly balanced" (rank-1) version of $T^{(j)}$ – called $\langle H \rangle$ – where such z_i's are distributed evenly across columns. The random matrix ensemble H, therefore, is nothing but a randomized version of the empirical matrix $T^{(j)}$ at hand, with the constraint that it retains the information about its row sums. Our results show that – under certain "spectrally gapped" conditions – the full knowledge of the matrix $T^{(j)}$ is not necessary nor required, and what really matters (as far as the MFPT is concerned) are just the row sums z_i of $T^{(j)}$.

4. In the context of Eq. (13) the authors say that "it is rather straightforward to show that...— It would be useful to add a reference here, or provide an explanation.

We have added that eq. (13) follows from the formula for the inverse of a block matrix, and pointed to our earlier ref. [37] where the calculation is spelt out in more detail.

5. In Eq. (21) the authors formulate the cumulant decay condition. While this is a sound technical definition, it would be useful to try and provide some insight into scenarios where this is obeyed vs. scenarios where this is not obeyed.

This is an interesting and fair request, but very hard to accommodate in an exhaustive manner. We have however expanded our footnote 6 into the new Appendix A to provide a full calculation of the cumulant decay condition in the case of i.i.d. Gaussian entries δh_{ij} with variance σ_N, which is one of the few calculations we can handle in a simple and straightforward way. The conclusion is that – depending on the way σ_N scales with N – the cumulant decay condition can be met or not (the variance should not be "too large"). We hope that this will be sufficient to give the reader at least a partial idea of how the cumu-

lant condition is expected to work, by forcing the probability density of the δh_{ij} to be "as narrow as needed".

6. In Sec. 3.2 the authors consider Erdős-Rényi networks. It is not completely clear if they use here a directed or undirected version of the network, and this should be explained. Also, I guess that the weights are incorporated into the adjacency matrix directly, and this should be stated more clearly.

The referee is perfectly right. We have amended the text to explain that we are considering a directed version of the ER network (restricted to its strongly connected component), obtained by imposing random, non-symmetric weights on the links of an otherwise symmetric "backbone".

7. A second issue with ER networks is regarding the reference to the strongly connected component of the ER network. The authors correctly mention at the end of the section that $c = \ln N = 6.21$ is the connectedness threshold. I believe it would be useful to mention this background fact at the beginning of the section to remind the readers of this fact. Furthermore, in the example presented in Fig. 3 the authors use $c = 7$ and $c = 190$. These values are already above $c = 6.21$ and hence the issues with the connectedness threshold are ignored. I believe that it would be useful to present a sparser ER network example as well, say for $c = 4$. I understands that in this limit the spectral gap is much smaller and the validity of the formula is weaker, but I believe it will be useful to probe more this limit in addition to the $c = 7$ and 190. Furthermore, the giant component of the ER network exhibits degree-degree correlations (it is disassortative) below the connectedness threshold, and it would be interesting to see the effect of this.

We have included a footnote at the beginning of section 3.2 to remind the reader about the connectedness threshold of E-R graphs. We provide here in Fig. 2 the requested scatter plot for $c = 4$, which (not surprisingly) is dramatically off. As the agreement is so poor, and the plot itself does not seem to be significantly more instructive than those we had already included, we would actually still prefer to leave it out of the paper. The point raised by the referee concerning the impact of assortativity on the accuracy of our formula is an interesting one. We have added a remark in the outlook section of the Conclusions as material for further investigations.

8. In the context of ER as well as Random Regular Graphs, in the limit of large c's - would it be possible to provide a large c approximation for Eq. (5)? As is well known, the spectrum of such dense ER/RRG networks approaches a simple form and it might yield a simple formula for the Mean First Passage Time in this limit.

We thank the referee for this suggestion, but unfortunately we have not found a way to make it work. It is worth remarking that the spectrum of the reduced transition matrix is only useful to assess the *accuracy* of the main formula (5), but does not enter it in any simple or straightforward way as far as the actual *values* of m_{ij} are concerned. So there does not seem to be an obvious path to convert the simplicity of the formulae expressing the spectral density in special cases into simplified (but nontrivial) large-connectivity expressions for our main formula.

9. In Fig. 4 the authors present results for $c = 7$ and $c = 190$ for Random Regular Graphs. It would be useful to include $c = 4$ too to probe the less dense limit, where the assumptions of Eq. (5) break down. I think this could add insight into the result. Furthermore, already for $c = 7$ we see deviations. Can the author explain whether these deviations are systematic? It seems that the approx-

imation tends to underestimate the exact result, hence most dots lie below the line. Is it a consistent result over many realizations or an accidental one?

We provide here in Fig. 3 the requested scatter plot for $c = 4$, which – as in the E-R case – is dramatically off. As the agreement is so poor, and the plot itself does not seem to be significantly more instructive than those we had already included, we would actually still prefer to leave it out of the paper as well. Concerning the second remark, we are confident that the deviations are indeed systematic, and our approximation tends to underestimate the exact result. This intuition stems from the following application of a version of Jensen's inequality for random matrices with independent entries[1], which in our notation would read $\langle (\mathbb{1} - H)^{-1} \rangle \geq (\mathbb{1} - \langle H \rangle)^{-1}$. The interpretation is as follows: if the reduced transition matrix were accurately modelled by a low-variance, "noise dressed" version of the "balanced" (rank-1) matrix $\langle H \rangle$, then the exact MFPTs would be consistently over-estimated w.r.t. our approximate formula (given by the r.h.s. of the above inequality). Clearly, we do not have a neat theorem to appeal to in the case of single, empirical instances (as in Fig. 3 and 4 of the manuscript), but the argument from Jensen's inequality – combined with a self-averaging assumption – convincingly supports the systematic nature of the under-estimation of the exact MFPTs by our formula.

10. In the captions of Tables 1 and 2, the authors report a comparison of "numerical", exact and approximate results. First, by "numerical" they really mean results of a simulations (as written in the main text). Therefore, it would be better to replace "numerical" by simulations. Also, it would be useful to add a specific reference to the equation used for the approximate and exact results - which equations were precisely used here? Another issue – the table reports a specific pair (i, j) – which pair? What were the degrees of the nodes i and j? why not average over all pairs?

We have corrected the table entries and included pointers to the relevant equations as suggested. We picked the pair $(4, 20)$ for no particular reason, but the qualitative conclusions (in terms of how relative accuracies generally deteriorate with higher sparsity) do not change if other pairs are picked. We did not store and thus do not have the information about the degrees of the nodes that were used for those simulations. Averaging over all pairs would indeed reduce fluctuations and lead to an even better agreement with our approximate formula (in the cases when it works), but we felt it was also important to stress that our formula *does not require averaging, or any smoothening procedure* to work well – in the "spectrally gapped" cases – for any specific pair of nodes.

11. On the broader picture - could the authors say anything about the distribution of first passage times? What quantity would be needed, from the random matrix side, in order to provide a similar result for the full distribution? Any comment here would be highly appreciated.

This is a very interesting question, but seemingly well beyond reach of the methods presented in our manuscript. Not surprisingly, there have been very few studies of the full distribution of the MFPT (see e.g. Ref. [50]), which did not lead in any case to very explicit and user-friendly expressions. The main technical hurdle stems from the very object to be computed, the joint distribution of MFPTs $P(\boldsymbol{m}) = \langle \delta \left(\boldsymbol{m} - (\mathbb{1} - H)^{-1} \mathbf{1} \right) \rangle$. After exponentiating the Dirac delta using its Fourier representation, and following again all the steps detailed after eq. (12), we would end up with an expression for $P(\boldsymbol{m})$ with replicated integrals *in the exponent*, which we do not know how to handle. So it seems that a new set of

[1] See A. Simonovits (1975). A Note on the Underestimation and Overestimation of the Leontief Inverse. Econometrica **43**, pp. 493-498.

ideas is needed to tackle this interesting problem.

To summarize, this work provides a useful result for the mean first passage time in weighted and directed networks. I would be happy to recommend publication of the paper once the authors consider the comments mentioned above.

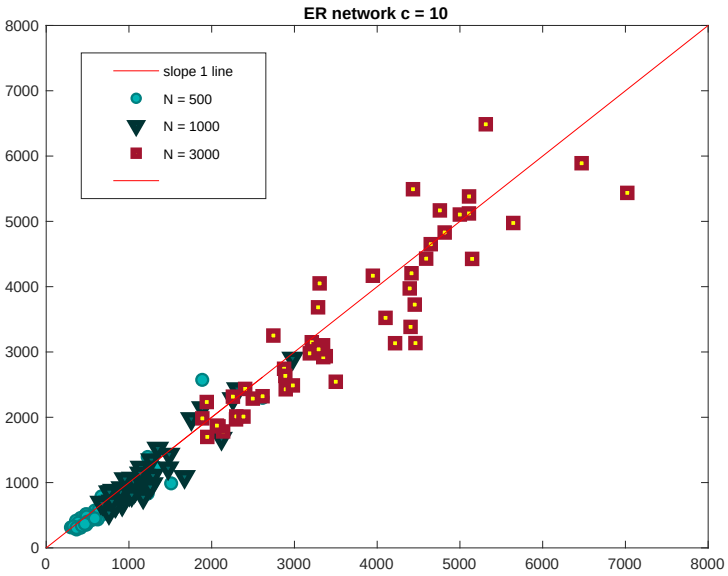

Figure 1: Scatter plot of approximate vs. exact MFPT on ER graphs with $c = 10$ for increasing values of N.

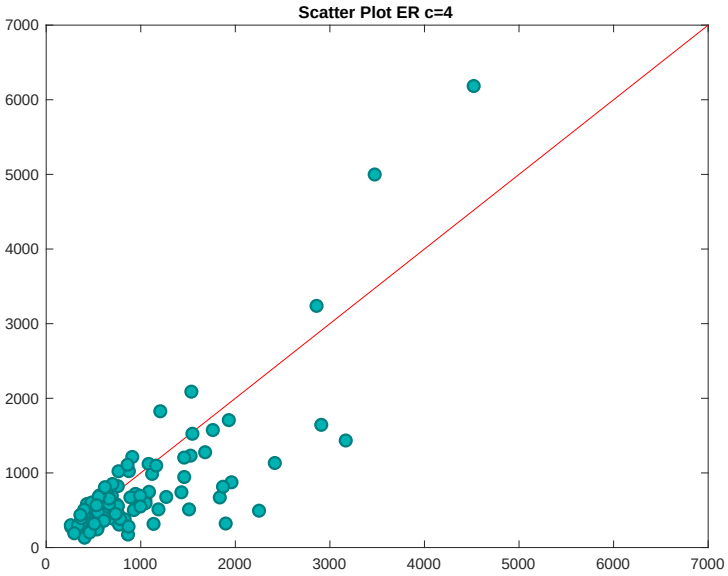

Figure 2: Scatter plot of approximate vs. exact MFPT on ER graphs with $c = 4$.

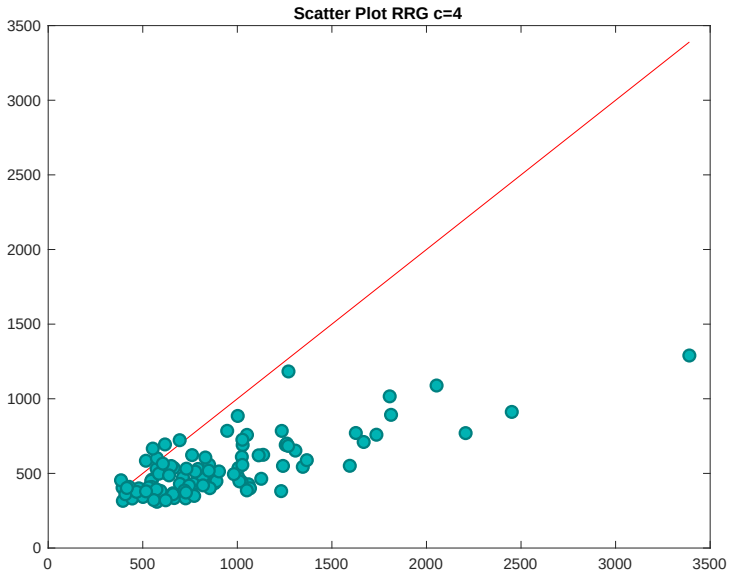

Figure 3: Scatter plot of approximate vs. exact MFPT on RRG graphs with $c = 4$.