Response to the Reviewer Comments on Manuscript "Local Graph Clustering with Network Lasso"

corresponding author: Alexander Jung (first.last (at) aalto.fi)

June 24, 2023

We express our sincere gratitude for the insightful and constructive comments and suggestions provided by the reviewers. We have tried to address all these comments to the extent possible. In the revised manuscript, modifications and amendments are highlighted in red (the parts that have been removed are not visible in the revised manuscript, though).

Major modifications we implemented in the revised manuscript include the following:

- We have significantly revised the introduction in Section ?? to more transparently position our work relative to existing local graph clustering methods.
- We now dedicate ...
- We now define ...
- We have added a numerical experiment ...

In the following, we respond to the reviewer comments in a point-by-point manner. Section, page, equation, and reference numbers in the copied action editor and reviewer comments (typeset in italic print) refer to the original manuscript whereas those in our response (typeset in upright print) refer to the revised manuscript unless indicated otherwise.

Comments of Reviewer #1

1.1 *The author should emphasize the contributions of this work in the end of the section 1, more clearly. What the advantages and differences between this method and the existed ones? More fast or more efficient using the proposed methods? It is still unclear for readers.*

We have revised the introduction in Section I to better clarify the contributions made relative to prior work (including our own work).

1.2 *Authors should provide more clear and detailed explanations of Fig.1 i.e., such as the meaning of the rain drop, and the color of nodes, on the caption of this figure.*

We have expanded the caption of Figure 1 to clarify the meaning of the water drops and the node shading.

1.3 *In section V and VI, author should analyze how to set the parameters in the framework. Do they have*

the "optimal" choice?

blablabla...

1.4 *It will be valuable to provide some analysis or discussion on the computational complexity for the proposed algorithm, comparing some classic algorithms. Is it O(n^2) or O(n^3)?*

1.5 *Some references are old and the following new references related to network clustering may be useful for improving this manuscript, such as [1] IEEE Transactions on Industrial Informatics, 16, 8, 5327-5334, 2020. [2] Physical Review E 91 (1), 012801, 2015. [3] New Journal of Physics, 21, 015005, 2019. [4] IEEE Transactions on Emerging Topics in Computational Intelligence, 2(3), 214-223, 2018. [5] Physica A , 542, 123514, 2020.*

1.6 *Please improve language presentation carefully.*

We have carefully revised the manuscript and tried to improve the use of language and clarity of presentation.

Comments of Reviewer #2

2.1 *The paper does well to contrast with spectral methods. Please elaborate on the drawbacks of the Laplacian quadratic minimization and elaborate on why total variation is a better conceptualization for local clusters.*

We now compare spectral methods with our flow-based approach in Section ?? to demonstrate how our method is able to recover the ground-truth cluster while spectral methods fail.

2.2 *The example provided is not informative, since its solution is going to depend on the selected parameterization, and no intuitive solution exists that should be recovered. provide 2-3 examples that illustrate how this reconceptualization of a local cluster improves the understanding of meso-scale structure in the graph. The penultimate example would indicate how this conceptualization of local clusters are superior to global clustering method(s) and/or other local clusters. See, for example the original Lasso paper [5]. Good examples typically come from collaboration networks (arxiv dblp), social networks (YouTube friendster), brain networks (fmri, structural mri) or product networks (amazon).*

2.3 *Since the method is particularly well suited for very sparse structures, there are interesting classes of graphs with long chains. Perhaps these are also more appropriate examples.*

2.4 *Fig 1 is highly informative, yet its caption does little to help understand what's happening. Expand caption to highlight the meaning convey by the figure.*

We have expanded the caption of Figure 1 to improve its clarity.

Comments of Reviewer #3

3.1 *Why give Eq. (3) suddenly? Where is it original from? who did define it? It needs a reference?*

We have reformulated the end of Section ?? and the beginning of Section ?? to better motivate using nLasso (??) as a model for local graph clustering. In particular, we point out the relation to our recent

paper at p. ??:

3.2 *Eq. (4) is somehow similar to the objective of spectral clustering, only the metric is changed to street distance. So your means the seeds are designed for semi-supervised learning?*

We try to make the relation between our approach and spectral method more explicit in the introduction Section ?? of the revised manuscript. Moreover, we have added the following after we define the nLasso problem on page ??,

3.3 *Since Eq. (5,6, and 7) are the dual problem of Eq. (3)., you give a solution. Then, what is different the paper from your [7] and [8]?*

We now point out the difference between this approach and our previous work more explicitly in the revised manuscript after Eq (??).

3.4 *You give very good analysis at section 4 and 5, but the experiments is very simple. Can you conduct experiment on real datasets, e.g. in citation datasets?*

....

3.5 *I strongly recommend authors to release the source code along with the submission, since the optimization based projects are typically open-source oriented to facilitate a fair assessment of the performance of the proposed methods for the community.*

.....

3.6 *Do you consider each connected component is a cluster in your algorithm?*

The clusters delivered by our algorithm are typically strict subsets of the connected components. However, as we now discuss after Eq (??), for certain parameter value ranges, our algorithm will deliver the connected components as clusters: