Response to the Comments of the Associate Editor and the Reviewers on Manuscript SPL-28200-2020, "Local Graph Clustering with Network Lasso"

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

August 13, 2020

We express our sincere gratitude for the insightful and constructive comments and suggestions provided by the reviewers. Major modifications we implemented in the revised manuscript include the following:

- We have significantly revised the introduction in Section I to more transparently position our work relative to existing local graph clustering methods.

- We now sharpen the discussion of our contributions in Section I and Section VII. In particular, as stated in the beginning of Section VII,

  > We have studied the application of nLasso to local graph clustering. Our main technical result is a characterization of the nLasso solutions in terms of network flows between cluster boundaries and seed nodes. Conceptually, we provide an interesting link between flow-based clustering and non-smooth convex optimization.

- We have added a numerical experiment in Section VI of the revised manuscript. In this additional experiment, we apply our local graph clustering method to an image segmentation task.

In the following, we respond to each comment of the action editor and the reviewers 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 the Associate Editor

AE.1 *Figure 1 was not clear and needs more explanation.*

We have expanded the caption of Figure 1 in order to clarify the meaning of Figure 1.

AE.2 *Algorithm complexity needs some discussion.*

We have added the following discussion on the computational complexity of the proposed method in the revised manuscript after (19):

> The updates (14)-(19) can be implemented as a distributed message-passing method for jointly solving nLasso (5) and the dual network flow problem (7). The computational complexity of one full iteration is proportional to the number of edges in the empirical graph. The overall complexity also depends on the number of iterations required to ensure the iterate $\hat{x}_i^{(r)}$ being sufficiently close to the nLasso (5) solution.
>
> Basic analysis of proximal methods shows that the number of required iterations required scales inversely with the required sub-optimality of $\hat{x}_i^{(r)}$ (see [3], [6]). This convergence rate cannot be improved for chain graphs [12]. For a fixed number of iterations and empirical graphs with bounded maximum node degree, the computational complexity of our method scales linearly with the number of nodes (data points).

AE.3 *References need updating.*

We have updated the references. See also our response to reviewer comment 1.5

AE.4 *Please apply this method to some large dataset. Compare results to other known methods.*

We have added a numerical experiment using a larger graph representing images. This experiment compares our nLasso based local clustering method with local graph clustering methods available from https://github.com/kfoynt/LocalGraphClustering.

AE.5 *How does this method compare to known clustering methods such as global clustering methods?*

We have expanded the discussion of how our approach relates to existing methods in Section I. Moreover, we have added comparisons with existing methods such as spectral graph clustering in Section VI.

AE.6 *Clearly address difference in your paper and [7],[8].*
[7] A. Jung. On the duality between network flows and network lasso. ArXiv e-prints arXiv:1910.01805, Mar. 2020.
[8] A. Jung, A O. Hero, A. Mara, S. Jahromi, A. Heimowitz, and Y.C. Eldar. Semi-supervised learning in network-structured data via total variation minimization. IEEE Trans. Signal Processing, 67(24), Dec. 2019.

We try to make the relation of this paper to our previous work more explicit in Section 1 and also right after the definition of the nLasso problem (5). See also our response to reviewer comment 3.3.

AE.7 *Review the paper carefully for grammar and English.*

We have reviewed carefully the use of language to avoid any grammatical errors in the revised manuscript.

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 significantly revised 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?*

We have added a discussion about the choice for the nLasso parameters at the end of Section IV:

> The above interpretation helps to guide the choice for the parameters α and λ in (5). The edge capacities λW_e limit the rate by which the values $\hat{x}_i^{(r)}$ can be "build up". Choosing λ too small would, therefore, slow down the convergence of $\hat{x}_i^{(r)}$. On the other hand, using nLasso (5) with too large λ does not allow to detect small local clusters $\mathcal{C}_k$ (see Section V).

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)?*

See our response to the associate editor comment AE.2.

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.*

We thank the reviewer for pointing out these references which we have added in the revised manuscript.

1.6 *Authors are encouraged to applied their framework on some large scale datasets, to verify the efficiency.*

We have added a discussion on the scalability of our method in the revised manuscript (see our response to associate editor comment AE.2). Moreover, we have expanded the numerical experiments which now include a larger empirical graph representing an image (see our response to reviewer comment 2.2).

1.7 *Please improve language presentation carefully.*

We have carefully revised the manuscript 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 VI 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).*

We have added a numerical experiment revolving around an image segmentation task. Moreover, we have prepared a Python notebook. This notebook contains the source code allowing to reproduce all numerical experiments and moreover to apply our method to user-defined datasets.

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.*

We have slightly expanded the numerical experiment that considers a chain graph (see Section VI of the revised manuscript). In particular, we now compare our method with spectral graph clustering.

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 the 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 II and the beginning of Section III to better motivate using nLasso (5) as a model for local graph clustering. In particular, as we point out before (5),

> We have recently started to explore the relation between network flow problems and TV minimization [13], [14]. Loosely speaking, the solution of TV minimization is piece-wise constant over clusters whose boundaries have a small total weight. This motivates us to learn the indicator function for the cluster $\mathcal{C}_k$ around the seed nodes $\mathcal{S}_k$ by solving ...

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 Section I of the revised manuscript. Moreover, we have added the following after defining the nLasso problem (5),

> Spectral methods use optimization problems similar to (5) but with the Laplacian quadratic form $\sum_{\{i,j\}\in\mathcal{E}} W_{i,j}(x_i - x_j)^2$ instead of TV (6).

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 point out the difference between this approach and our previous work somewhat more explicitly in the revised manuscript after (5),

> The special case of (5) when $\alpha = 0$ is studied in [13]. Closely related to (5) is the constrained TV minimization studied in [14]. This constrained TV minimization enforces the learned graph signal to be equal to one for each seed node $i \in \mathcal{S}_k$. In contrast, (5) uses soft constraints allowing for $\hat{x}_i \neq 1$ even at seed nodes $i \in \mathcal{S}_k$. Spectral methods use optimization problems similar to (5) but with the Laplacian quadratic form $\sum_{\{i,j\} \in \mathcal{E}} W_{i,j}(x_i - x_j)^2$ instead of TV (6).

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?*

We have added a numerical experiment that illustrates how our method can be used for image segmentation. See also our response to reviewer comment 2.2.

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.*

We have prepared a Python notebook that contains the source code underlying our experiments. This notebook is available at https://github.com/alexjungaalto/ResearchPublic/blob/master/LocalClusteringNLasso/LocalGraphClusteringNLasso12Aug2020.ipynb. We will also try to submit the notebook as supplementary material in this journal.

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 Proposition 1, for certain parameter value ranges, our algorithm will deliver the connected components as clusters:

> Thus, choosing values for λ larger than some critical value makes nLasso (5) deliver a cluster (20) containing all connected components of the empirical graph $\mathcal{G}$ with seed nodes $\mathcal{S}_k$.

Comments of Reviewer #4

4.1 *My only concern is the computational aspect. The author only tested the proposed method on a very small synthetic dataset. Experiments on a larger (maybe real-world) graph with a more complex structure could make the statement more convincing. So my suggestion is to extend the empirical study with more comprehensive experiments.*

We have added some discussion on the computational aspects of our method (see our response to associate editor comment AE.2). Moreover, we have expanded the numerical experiments (see our response to reviewer comment 2.2.)

4.2 *Other minor comments: Page 1 Line 30 left: "depent" Page 1 Line 34 right: "This methods" Page 2 Line 19 left "to chose" Page 3 Line 52 left "result are"*

We have carefully revised the manuscript in order to avoid any such typos.