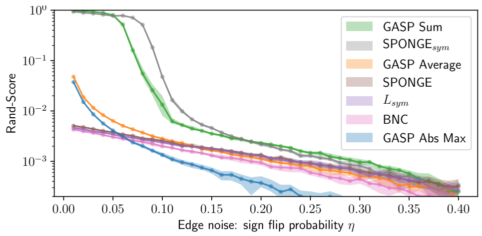

Figure 1: Experiments on SSBM synthetic graphs

Method	Rand-Score
GASP Average	0.8966
GASP Sum	0.8965
GASP Abs Max	0.8932
SPONGE$_{sym}$ [5]	0.4389
L_{sym} [8]	0.1931
SPONGE [5]	0.0789
BNC [3]	0.0074

Table 1: Experiments on CREMI

Method	AP	AP 50%
PANet [11]	**31.8**	**57.1**
GMIS + GASP Avg	28.3	47.0
GMIS	27.3	45.6
Mask R-CNN	26.2	49.9
SGN [10]	25.0	44.9
DIN [1]	20.0	38.8
DWS [2]	19.4	35.3

Table 2: Scores on CityScapes *test* set

We appreciate the very helpful comments by all reviewers and will address their concerns in the final version.

Comparison with spectral clustering (R3): As kindly suggested by Reviewer 3, we include additional experiments comparing GASP with the spectral clustering methods proposed in [5, 3, 8]. Similarly to most of spectral clustering approaches, these methods need to previously specify the number k of final clusters. On the other hand, in our article we focused on algorithms and tasks related to *correlation clustering*, which has the goal of partitioning a signed graph when the true number of clusters is unknown. As suggested by R3, we then generated synthetic graphs from a signed stochastic block model (SSBM) where the true number of clusters is previously known. In the experiments, we used an Erdős-Rényi random graph model $\mathcal{G}(N, p)$ with $N = 10^5$ vertices and edge probability $p = 0.1$. Following the approach in [5], we partitioned the graph into $k = 100$ equally-sized clusters, such that edges connecting vertices belonging to the same cluster (different clusters) had Gaussian distributed edge weights centered at $\mu = 1$ ($\mu =$-1) and with standard deviation 0.1. To model noise, we flipped the sign of each edge independently with probability η. Results are shown in Fig. 1. At the same time, we would also like to argue that computer vision is currently one of the most important machine learning application domains, so a success on real data from that domain should count no less than success on synthetic / toy data. Thus, we extended the comparison with spectral methods to the task of neuron segmentation. Since the spectral methods cannot scale to the full CREMI dataset, we considered a random tiny crop of 10x100x100 voxels resulting in a graph with 10^5 vertices and $\sim 10^6$ edges. Scores are summarized in Table 1.
We would like to stress that in these experiments we compared GASP with the strongest possible baseline, by giving the true number of GT clusters as an input to the spectral methods. Despite this, GASP significantly outperformed other methods on neuro-data and achieved comparable scores on the SSBM synthetic data. The specific design choice of a *sign flipping* noise used in the SSBM experiments turned out to favor GASP with *Sum* linkage, which is the one with the lowest tendency to over-cluster and tends to grow one cluster at the time. TODO: More?

Results on Cityscapes test (R1 and R2): To highlight the competitive and strong performances of GASP, in Table 2 we list the scores achieved on the CityScapes test set, as kindly suggested by R1. GASP with Average linkage outperforms all previously proposed proposal-free methods. The best performing method PANet [11] is a proposal-based method strongly related to Mask R-CNN.

GASP and objective function (R2): We agree with Reviewer 2 and we also believe that a generalized objective function would give more significance to the framework. Currently, only the *Sum* and the *Abs Max* linkage have been shown to optimize specific objective functions [13, 7]. At the same time, to our knowledge, the proposed framework is the first one that highlights a clear and simple connection between several partitioning algorithms for signed graph and hierarchical agglomerative clustering (HAC) of unsigned graphs. HAC algorithms have been introduced more than 50 years ago [9], but only recent work started to better understand its analytical foundation [12, 4, 6]. We then think that our framework can in principle suggest further connections between the recent findings about HAC and signed graph partitioning.

References

[1] A. Arnab and P. H. Torr. Pixelwise instance segmentation with a dynamically instantiated network. In *CVPR*, pages 441–450, 2017.
[2] M. Bai and R. Urtasun. Deep watershed transform for instance segmentation. In *CVPR*, pages 5221–5229, 2017.
[3] K.-Y. Chiang, J. J. Whang, and I. S. Dhillon. Scalable clustering of signed networks using balance normalized cut. pages 615–624. ACM, 2012.
[4] V. Cohen-Addad, V. Kanade, F. Mallmann-Trenn, and C. Mathieu. Hierarchical clustering: Objective functions and algorithms. 2019.
[5] M. Cucuringu, P. Davies, A. Glielmo, and H. Tyagi. SPONGE: A generalized eigenproblem for clustering signed networks. In *AISTATS*, 2019.
[6] S. Dasgupta. A cost function for similarity-based hierarchical clustering. *arXiv preprint arXiv:1510.05043*, 2015.
[7] M. Keuper, E. Levinkov, N. Bonneel, G. Lavoué, T. Brox, and B. Andres. Efficient decomposition of image and mesh graphs by lifted multicuts. In *ICCV*, 2015.
[8] J. Kunegis, S. Schmidt, A. Lommatzsch, J. Lerner, E. W. De Luca, and S. Albayrak. Spectral analysis of signed graphs for clustering, prediction and visualization. SIAM, 2010.
[9] G. N. Lance and W. T. Williams. A general theory of classificatory sorting strategies: 1. hierarchical systems. *The computer journal*, 9(4):373–380, 1967.
[10] S. Liu, J. Jia, S. Fidler, and R. Urtasun. Sgn: Sequential grouping networks for instance segmentation. In *ICCV*, pages 3496–3504, 2017.
[11] S. Liu, L. Qi, H. Qin, J. Shi, and J. Jia. Path aggregation network for instance segmentation. pages 8759–8768, 2018.
[12] B. Moseley and J. Wang. Approximation bounds for hierarchical clustering: Average linkage, bisecting k-means, and local search. In *NIPS*, 2017.
[13] S. Wolf, A. Bailoni, C. Pape, N. Rahaman, A. Kreshuk, U. Köthe, and F. A. Hamprecht. The mutex watershed and its objective: Efficient, parameter-free image partitioning. *arXiv preprint arXiv:1904.12654*, 2019.