*Davide Cittaro, PhD  
Center for Omics Sciences, IRCCS San Raffaele Hospital*

*Via Olgettina 58*

*20132 Milan*

*Italy*

Prof. Fabian J. Theis

Associate Editor, Genome Biology

Dear Prof. Theis,

Please find enclosed our manuscript entitled, “Nested Stochastic Block Models Applied to the Analysis of Single Cell Data” by L. Morelli *et al*. submitted to your consideration for publication as a Method Article in Genome Biology.

Recent advances in the field of Single Cell technologies were mirrored by a rapid increase in the development of mathematical models and software packages to identify cell populations. It may be argued that one of the most popular approaches is based on the optimization of modularity of the neighborhood graph, representing cell-wise similarity. While extremely fast and intuitive, this approach has limitations, in particular the lack of a statistical measure of the outcome which leads to arbitrary choice of parameters (*e.g.* the resolution parameter).

In this manuscript we explore the application of Nested Stochastic Block Models (nSBM), a generative model for graphs organized into communities, to identify cell clusters in single cell data. Such approach has been extensively used to analyze graphs in other contexts and, to our knowledge, we are the first to extend it to single cell studies.

The main findings of our study are the following:

* nSBM is accurate in identifying cell populations in real data. Data are modeled in a hierarchical manner, removing the need to choose appropriate resolutions. Moreover, the entropy of a model is calculated, so that it is possible to perform model selection when parameters are changed.
* The hierarchy proposed by nSBM has a direct biological explanation. In a cell differentiation context, we show that the hierarchy is consistent with differentiation trajectories.
* We can calculate the difference in model entropy that occurs if cells are assigned to the wrong cluster. This allows us to compute a “cluster consistency” value which could be exploited to isolate spurious cell clusters.
* We developed a python library (*schist*) which is fully compatible with *scanpy*, to help with the adoption of nSBM to model single cell data.

For these reasons we believe that our work is of potential interest for a broad range of researchers working in this field, and therefore it may fit the scope of Genome Biology.

We acknowledge that a preprint version of this work has previously submitted to biorXiv platform (doi: 10.1101/2020.06.28.176180) and that the code of our library is publicly available (<https://github.com/dawe/schist>).

On behalf of all the authors, I would like to thank you since now for your time and consideration and look forward to hearing from you.

Please accept our best regards

Sincerely,

Davide Cittaro