Ljubljana, March 17th, 2021

Dear Editor,

We would like to submit our manuscript on a modular Python toolbox for t-SNE dimensionality reduction and embedding.

Science, according to Wikipedia, is "a systematic enterprise that builds and organizes knowledge". One of the earliest means of information organization is a map. Today, an important subfield of data science strives to map multidimensional data onto a two-dimensional plane to expose the structure, relations, provide predictions of functions, and uncover temporal trends. A prominent example of such a technique is t-SNE. For instance, t-SNE can embed expression-profiled single cells into a two-dimensional map, exposing regions of same-typed cells or charting cell development. Over 1,400 articles from Nature journals alone recently used t-SNE for data visualization. Nature Methods has published a number of papers that use t-SNE, including those with a t-SNE depiction of single-cell data maps. t-SNE visualizations have appeared on the magazines' covers and have been embraced by the entire data science community.

In the paper, we report on openTSNE, a Python-based open-source library for t-SNE. openTSNE runs orders of magnitudes faster than comparable Python-based implementations (e.g., scikit-learn) and can handle data sets containing millions of data points. t-SNE has been recently criticized for poor scalability and susceptibility to batch effects. Our proposed implementation addresses these concerns and includes the lately-proposed methodological advancements, which we review in the paper's online methods section.

Additionally, openTSNE is currently the only t-SNE library that can place new data points into a constructed embedding. The paper shows that such embedding of new data effectively mitigates batch effects and is an approach for visual prediction.

Our t-SNE implementation is open, available on GitHub, and fosters extensibility and experimentation. The library has already attracted significant interest within the Python community; it gained 758 GitHub stars, on par with prominent data science packages in Python, such as scanpy for single-cell analysis (854 stars).

We are submitting an original manuscript that has not been considered for publication before. **Lin Tang** would be an excellent choice for an editor for this manuscript because of her broad knowledge in the field and work in single-cell analytics. For reviewers, we kindly propose **Dmitry Kobak** (U Tuebingen), an author of recently proposed t-SNE tricks that we implemented in the library, **Barbara di Camillo** (U Padova), single-cell data scientists, **Marinka Žitnik** (Harvard), a data integration expert, or **Cagatay Turkay** (U Warwick), a specialist in explainable data visualizations.

Yours faithfully,
Pavlin G. Poličar, Martin Stražar, Blaž Zupan