

# Graph Layout by Path-Guided Stochastic Gradient Descent

Simon Heumos<sup>1\*</sup> , Andrea Guarracino<sup>2\*</sup> , Sven Nahnsen<sup>1,3</sup> , and Erik Garrison<sup>4</sup> 

<sup>1</sup>Quantitative Biology Center (QBiC) Tübingen, University of Tübingen, Tübingen, Germany

<sup>2</sup>Genomics Research Centre, Human Technopole, Milan, Italy

<sup>3</sup>Biomedical Data Science, Dept. of Computer Science, University of Tübingen, Tübingen, Germany

<sup>4</sup>University of Tennessee Health Science Center, Memphis, TN, USA

\* Contributed equally.

---

[Pangenome graphs](#) built from raw sets of alignments may have complex structures generated by common patterns of genome variation. These structures can introduce difficulty in downstream analyses, visualization, mapping, and interpretation. Graph sorting aims to find the best node order for graph layouts in 1 and 2 dimensions (1D and 2D) to simplify these complex regions. Our new path-guided (PG) [stochastic gradient descent](#) (SGD) algorithm exploits the biological information of the paths in a pangenome graph during the sorting process. In our [implementation](#), the algorithm moves a single pair of nodes at a time, optimizing the disparity between the layout distance of a node pair and the actual nucleotide distance of a path traversing these nodes. We apply the algorithm to gigabase-scale pangenome graphs producing graph layouts in 1D and 2D. The exploration of the path-guided SGD parameter space is in progress, in order to obtain the best layout as quickly as possible. We also evaluate path-guided metrics in order to measure a graph's stress level. The 1D PG-SGD algorithm is already a key step in general pangenome analysis such as the pangenome graph linearization pipeline [smoothxg](#).