

PAVER: Pathway Analysis Visualization with Embedding Representations

William Ryan^{1*}, Ali Sajid Imami, MD¹, Hunter Eby¹, Robert McCullumsmith, MD, PhD^{2,3}, Rammohan Shukla, PhD²

¹College of Medicine and Life Sciences, University of Toledo, Toledo, OH 43614

²Department of Neurosciences, The University of Toledo, Toledo, OH 43614

³Department of Psychiatry, The University of Toledo, Toledo, OH 43614

*Corresponding author: wryan3@rockets.utoledo.edu

Keywords: Bioinformatics, Pathway Analysis, Gene Ontology, Omics

Published: 14 December 2023

Interpreting pathway analysis often poses a significant challenge due to the extensive lists of gene ontology (GO) terms that require meticulous manual curation to identify underlying themes. We developed PAVER, a novel R software package, to address this issue by automating theme generation and clustering of GO terms. By utilizing embedding representations and advanced machine learning techniques, PAVER discerns patterns within the GO terms, creating an intuitive visual landscape of clusters for ease of functional interpretation. This method significantly minimizes the time and effort traditionally required for manual curation. We applied PAVER to a previously published dataset, where it demonstrated robustness by generating themes that closely mirrored those produced by manual curation. With PAVER, we present a powerful tool that not only enhances the efficiency of pathway analysis but also broadens its accessibility across various fields, including disease pathway modeling, drug target identification, and comparative genomics. Our work with PAVER marks a significant step towards simplifying the pathway analysis interpretation process in bioinformatics research.