

## Accepted Manuscript

New Tools for the Visualization of Biological Pathways

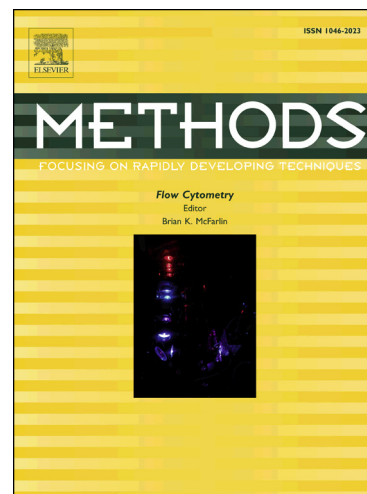
Tomojit Ghosh, Xiaofeng Ma, Michael Kirby

PII: S1046-2023(17)30043-9

DOI: <http://dx.doi.org/10.1016/j.ymeth.2017.09.006>

Reference: YMETH 4319

To appear in: *Methods*



Please cite this article as: T. Ghosh, X. Ma, M. Kirby, New Tools for the Visualization of Biological Pathways, *Methods* (2017), doi: <http://dx.doi.org/10.1016/j.ymeth.2017.09.006>

This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

# New Tools for the Visualization of Biological Pathways

Tomojit Ghosh, Xiaofeng Ma, Michael Kirby

Colorado State University, Fort Collins, Colorado, USA

---

## Abstract

This paper presents several geometrically motivated techniques for the visualization of high-dimensional biological data sets. The Grassmann manifold provides a robust framework for measuring data similarity in a subspace context. Sparse radial basis function classification as a visualization technique leverages recent advances in radial basis function learning via convex optimization. In the spirit of deep belief networks, supervised centroid-encoding is proposed as a way to exploit class label information. These methods are compared to linear and nonlinear principal component analysis (autoencoders) in the context of data visualization; these approaches may perform poorly for visualization when the variance of the data is spread across more than three dimensions. In contrast, the proposed methods are shown to capture significant data structure in two or three dimensions, even when the information in the data lives in higher dimensional subspaces. To illustrate these ideas, the visualization techniques are applied to gene expression data sets that capture the host immune system's response to infection by the Ebola virus in non-human primate and collaborative cross mouse.

*Keywords:* Data visualization, Grassmannian embedding, Autoencoder, Centroid-encoder, Sparse radial basis functions, Ebola virus

---

<sup>1</sup>This paper is based on research partially supported by the National Science Foundation under Grants No. DMS-1513633, and DMS-1322508 as well as DARPA awards N66001-17-2-4020 and D17AP00004. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation or DARPA.

Download English Version:

<https://daneshyari.com/en/article/8340151>

Download Persian Version:

<https://daneshyari.com/article/8340151>

[Daneshyari.com](https://daneshyari.com)