

Figure 1: A typical pipeline of data visualization by first constructing a K-nearest neighbor graph and then projecting the graph into a low-dimensional space.

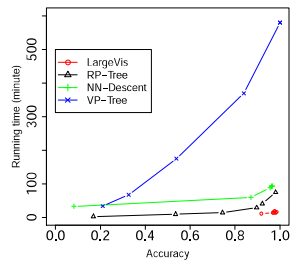
eration technique [26] for the t-SNE by first constructing a K-nearest neighbor (KNN) graph of the data points and then projecting the graph into low-dimensional spaces with tree-based algorithms. T-SNE and its variants, which represent a family of methods that first construct a similarity structure from the data and then project the structure into a 2D/3D space (see Figure 1), have been widely adopted recently due to the ability to handle real-world data and the good quality of visualizations.

Despite their successes, when applied to data with millions of points and hundreds of dimensions, the t-SNE techniques are still far from satisfaction. The reasons are three-fold:

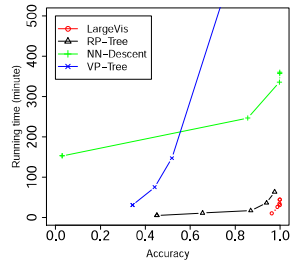




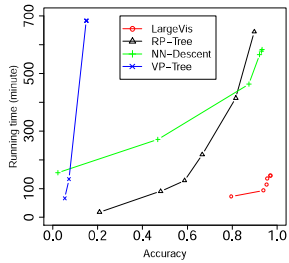




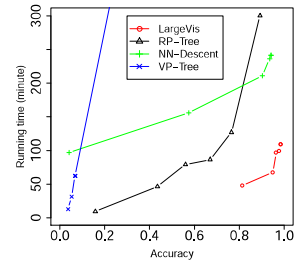
(a) WikiWord



(b) WikiDoc



(c) LiveJournal



(d) CSAuthor







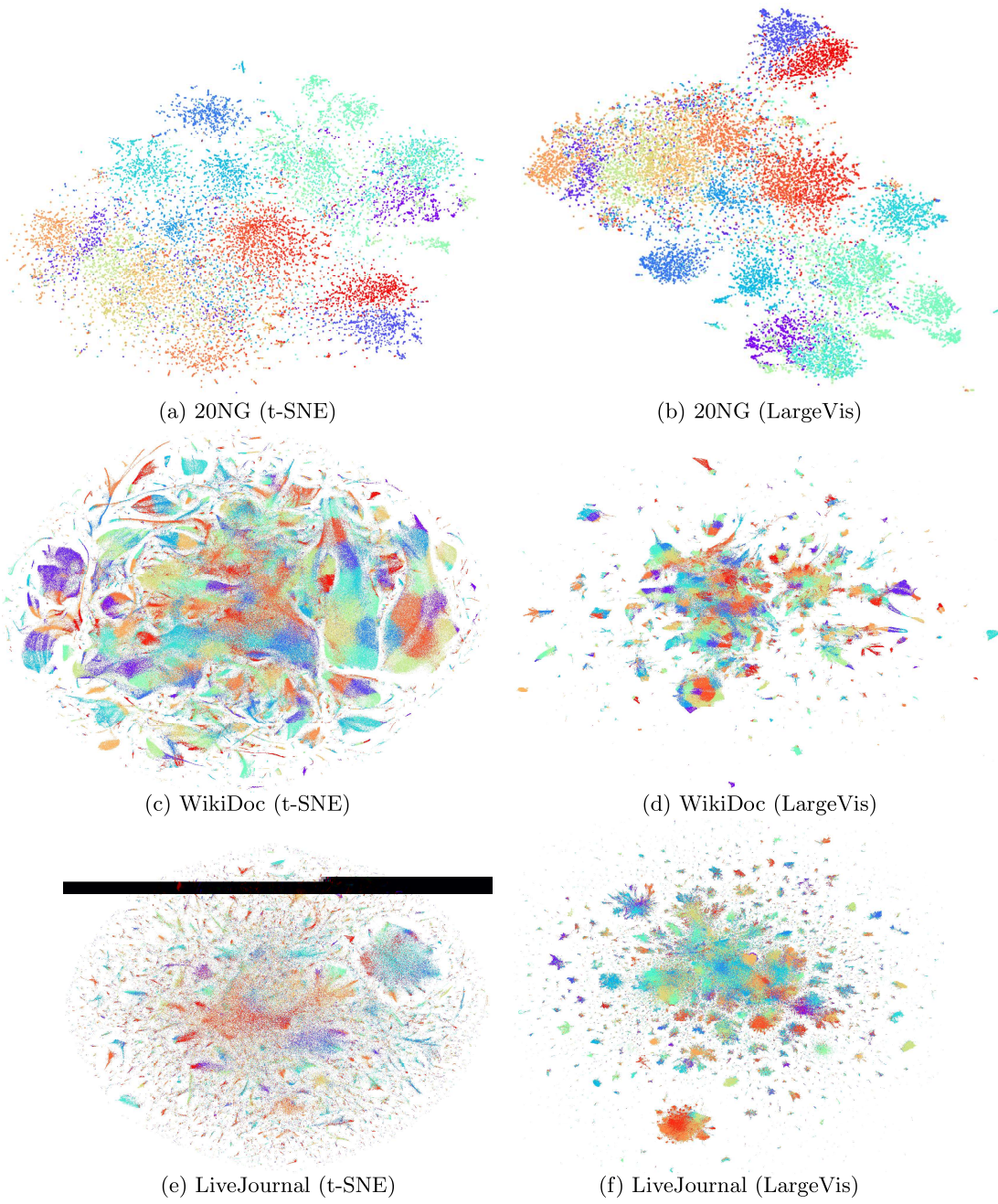


Figure 8: Visualizations of 20NG, WikiDoc, and LiveJournal by t-S



