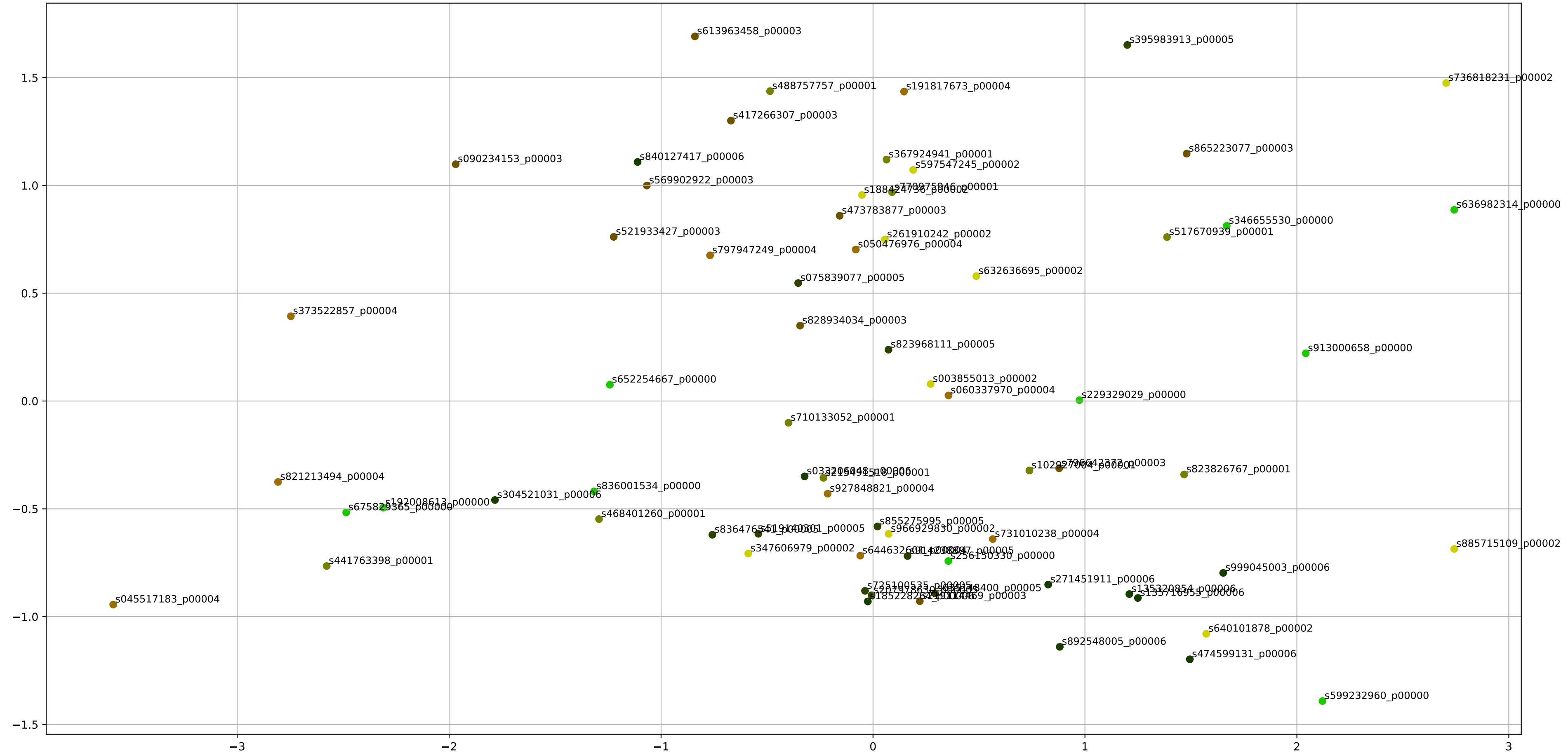
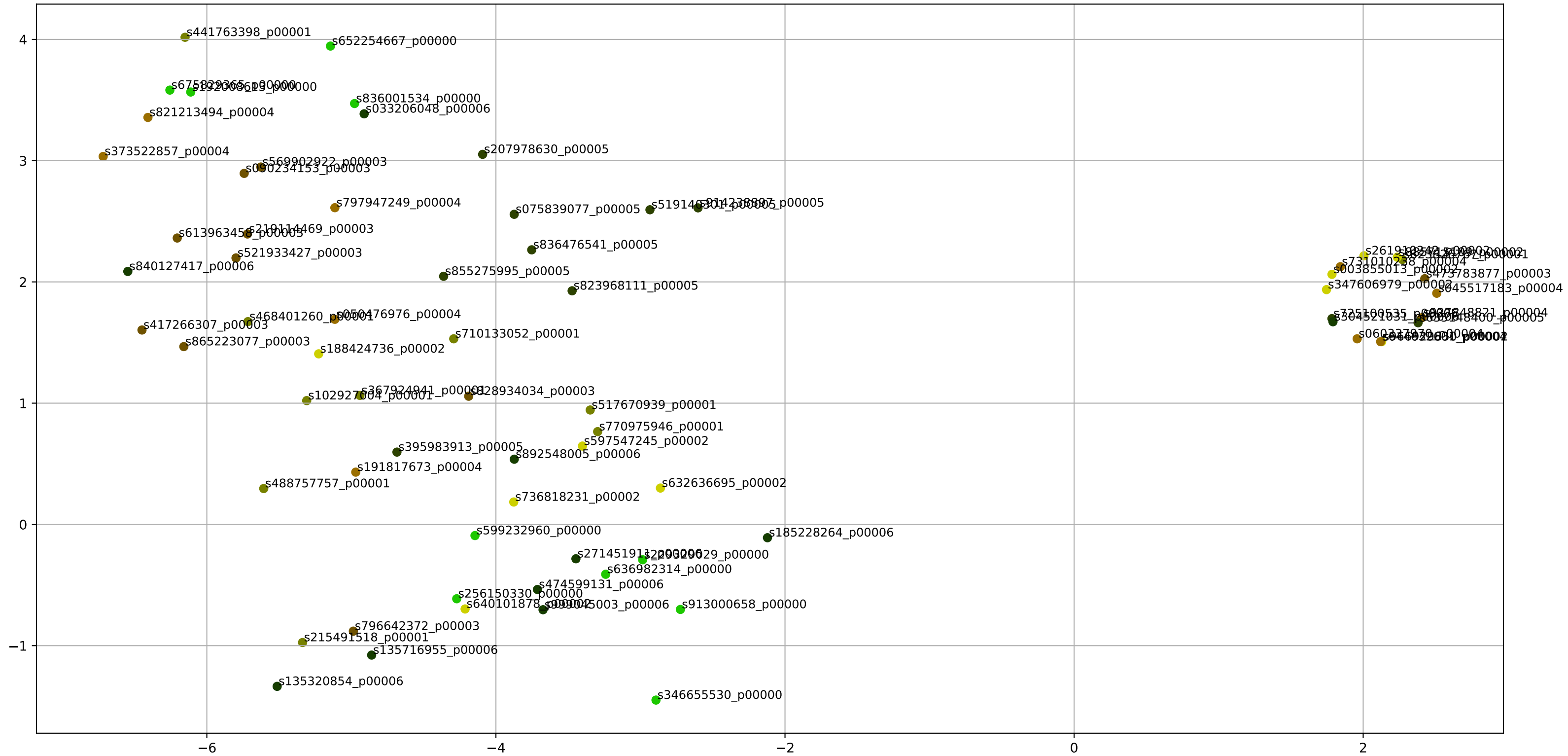


Code Embeddings (PCA)CODENET(CodeBERT)

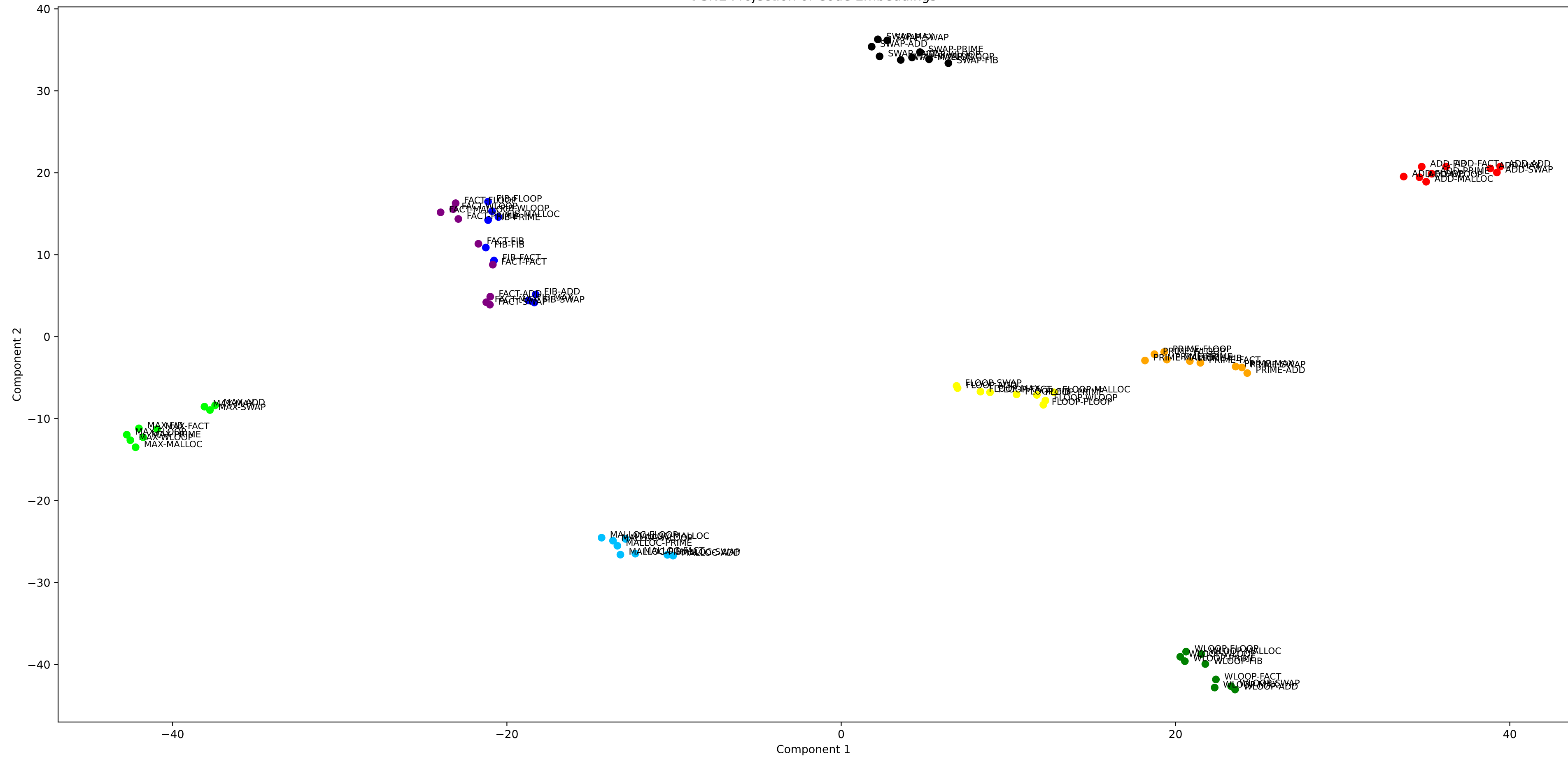


Code Embeddings (TSNE)Codenet

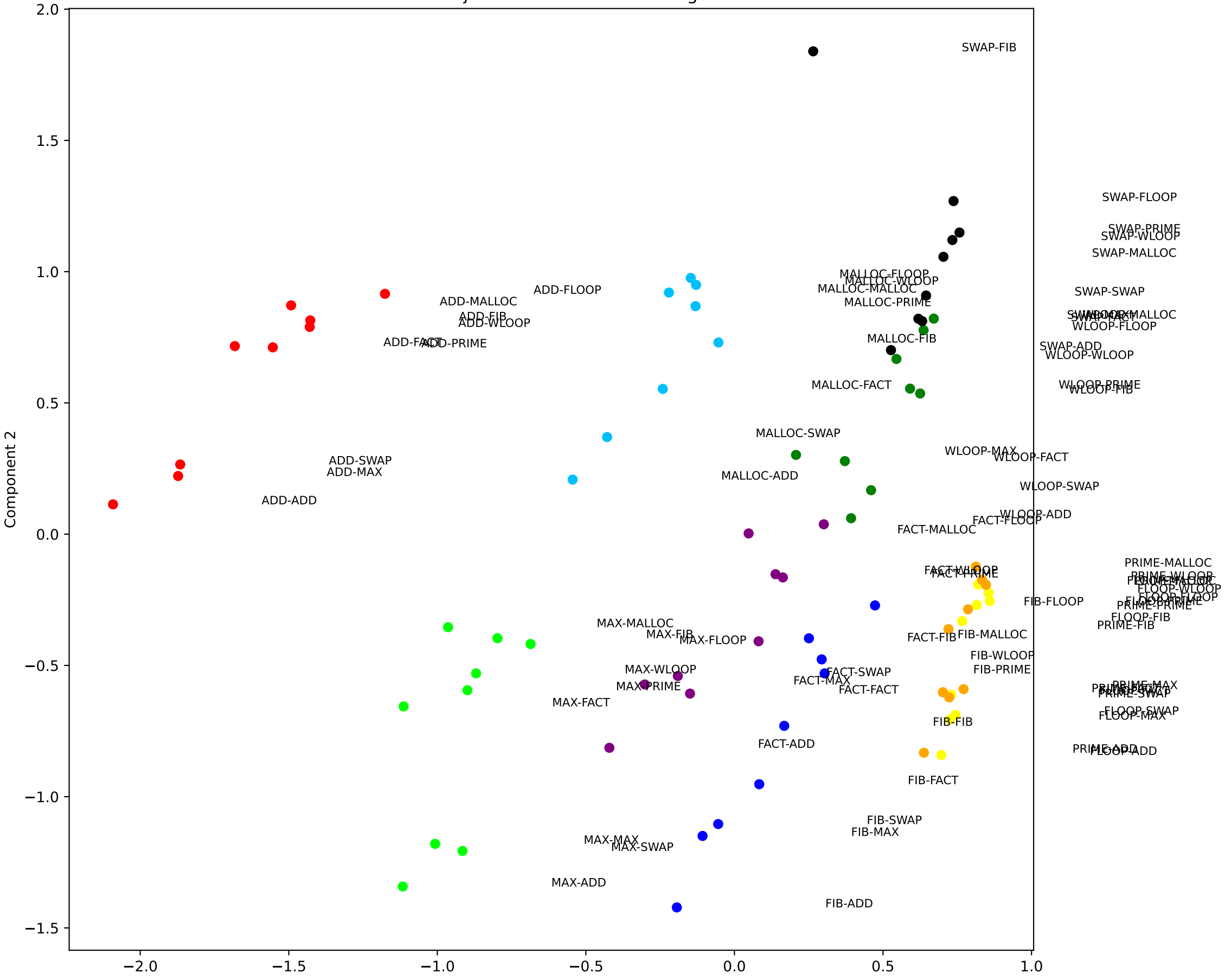


The scatter plot displays 1000 data points, each labeled with a unique identifier (e.g., s1000000_p000000, s1000001_p000000, etc.). The points are distributed across a coordinate system with the x-axis ranging from -1 to 4 and the y-axis ranging from -1 to 1. The points are colored in a gradient from dark blue to light blue. The distribution is characterized by several distinct clusters: a dense group on the left side (x < -0.5), a cluster on the right side (x > 3.5), and a large, more dispersed cluster in the center (x between -0.5 and 1.5, y between -0.5 and 0.5). The points are labeled with IDs such as s1000000_p000000, s1000001_p000000, s1000002_p000000, etc.

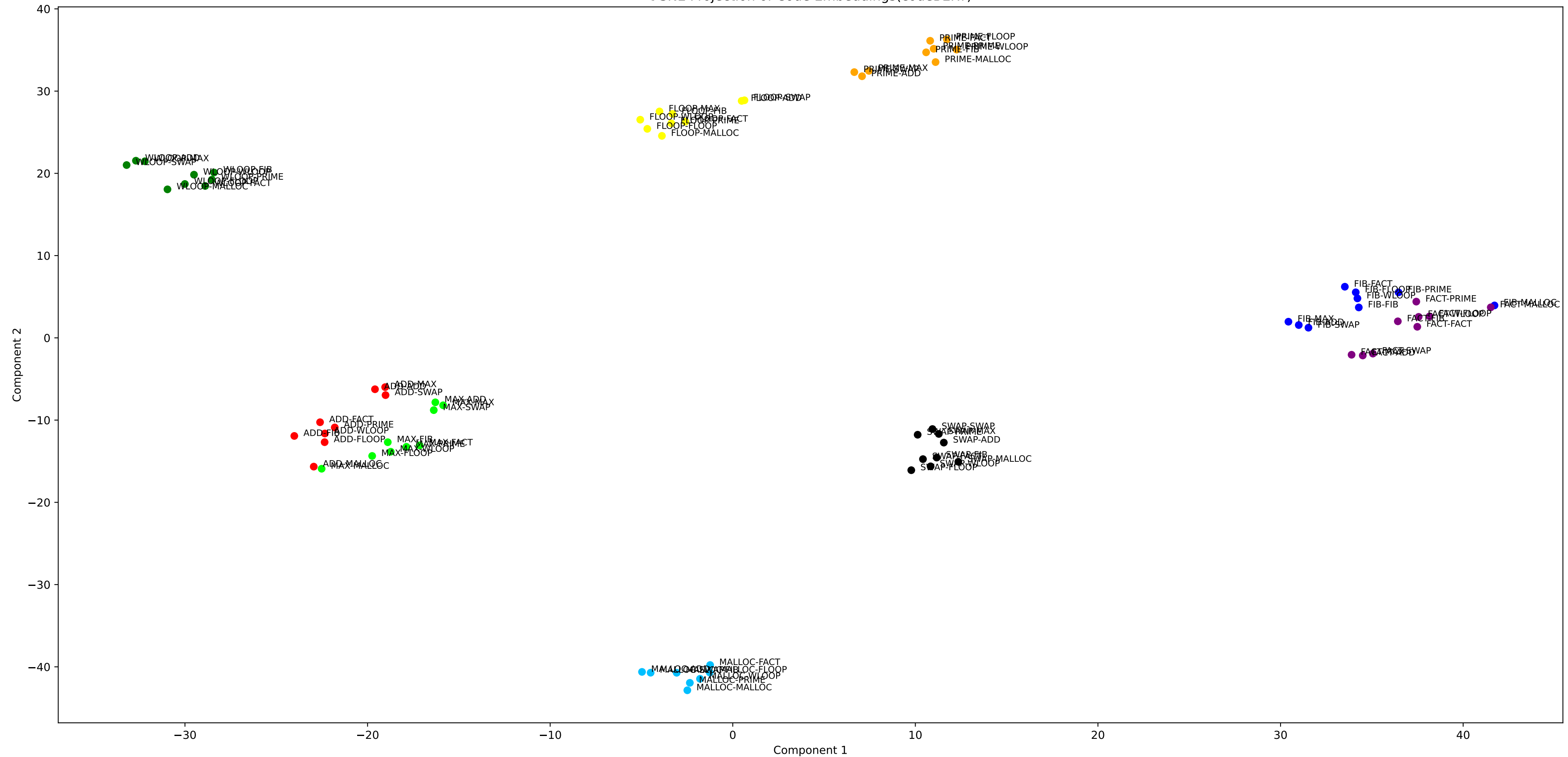
t-SNE Projection of Code Embeddings



PCA Projection of Code Embeddings



t-SNE Projection of Code Embeddings(codeBERT)



PCA Projection of Code Embeddings(CodeBERT)

