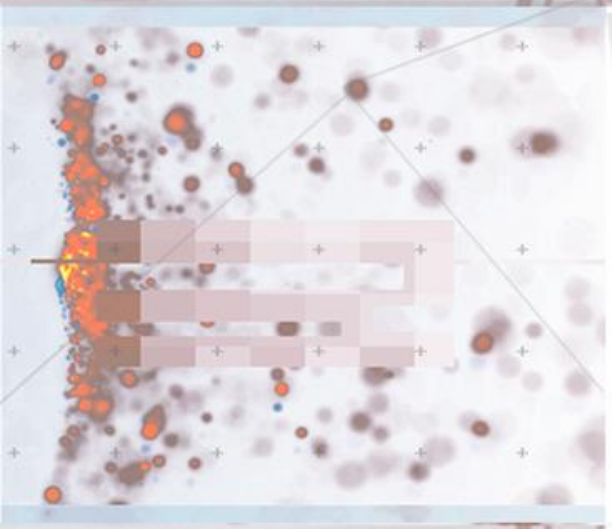

Session 9: Dimensionality Reduction Methods

In-video Quiz

- Suppose we have word-document co-occurrence matrix A with number of words n = 4M and number of documents d = 50M. We want to approximate matrix A via Nonnegative Matrix Factorization. Which of the following choices of k would be more suitable in this application?
  - k = 10
  - k = 512
  - k = 4M
  - k = 50M.
- Answer: k = 512
- Explanation: The goal is to approximately factorize the co-occurrence matrix A with two lower-rank matrices U and V. If k = 4M or 50M, the ranks of U and V could be large. In particular, for the extreme cases, U and V are full rank. If k = 10, the ranks of U and V are too small, leading to a bad approximation of matrix A. k = 512 could be a suitable choice for the rank, which could give good approximation of A while enjoying the low ranker property.