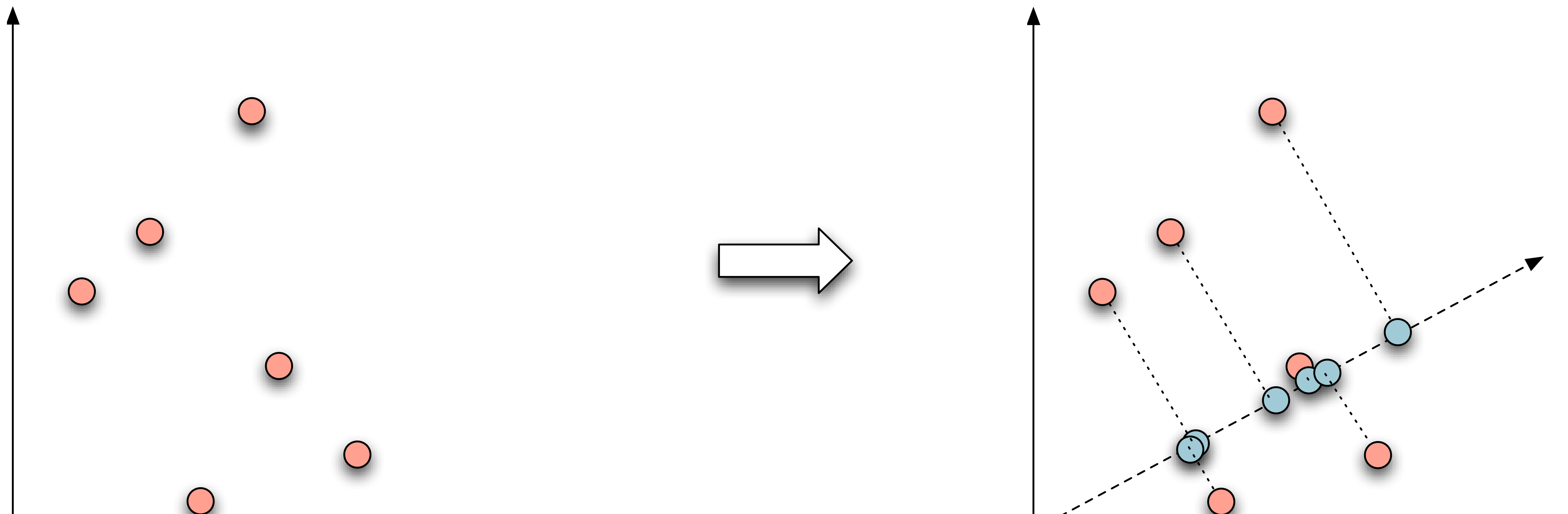
Matrix Decomposition and Latent Semantic Indexing (LSI)

Introduction to Information Retrieval
CS 221
Donald J. Patterson

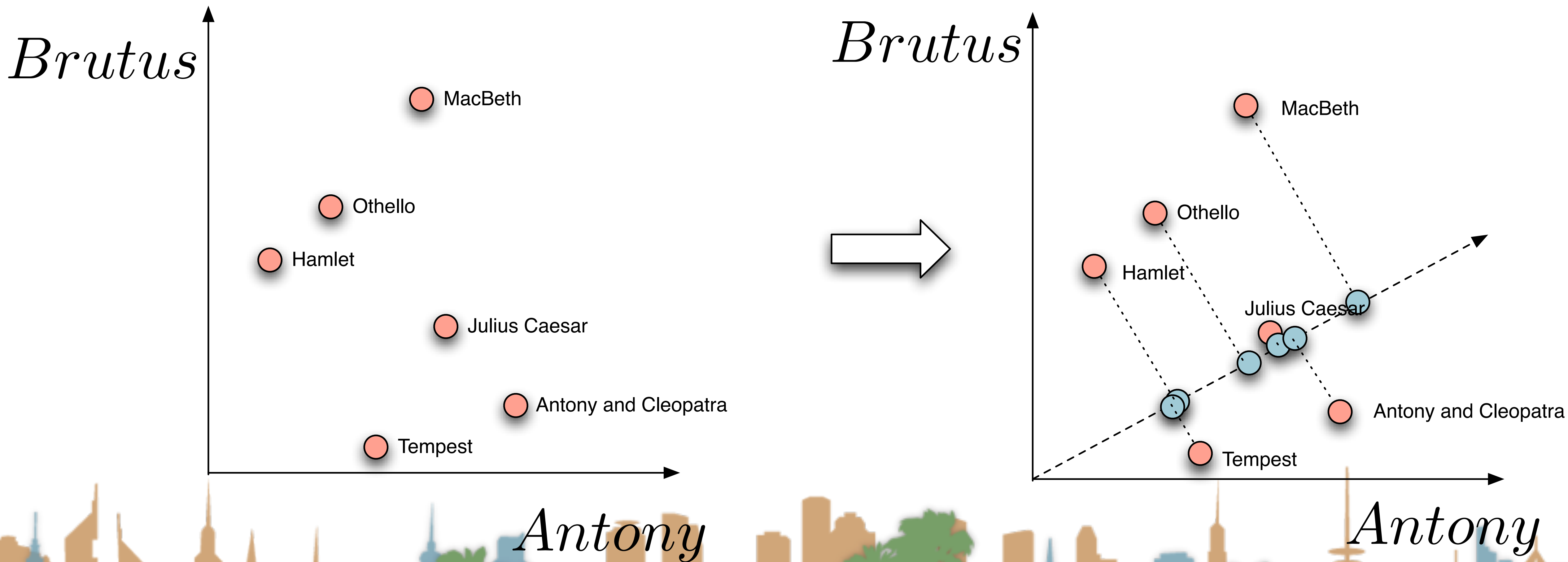
Mathematically speaking

- Latent Semantic Indexing can project on an arbitrary axis, not just a principal axis



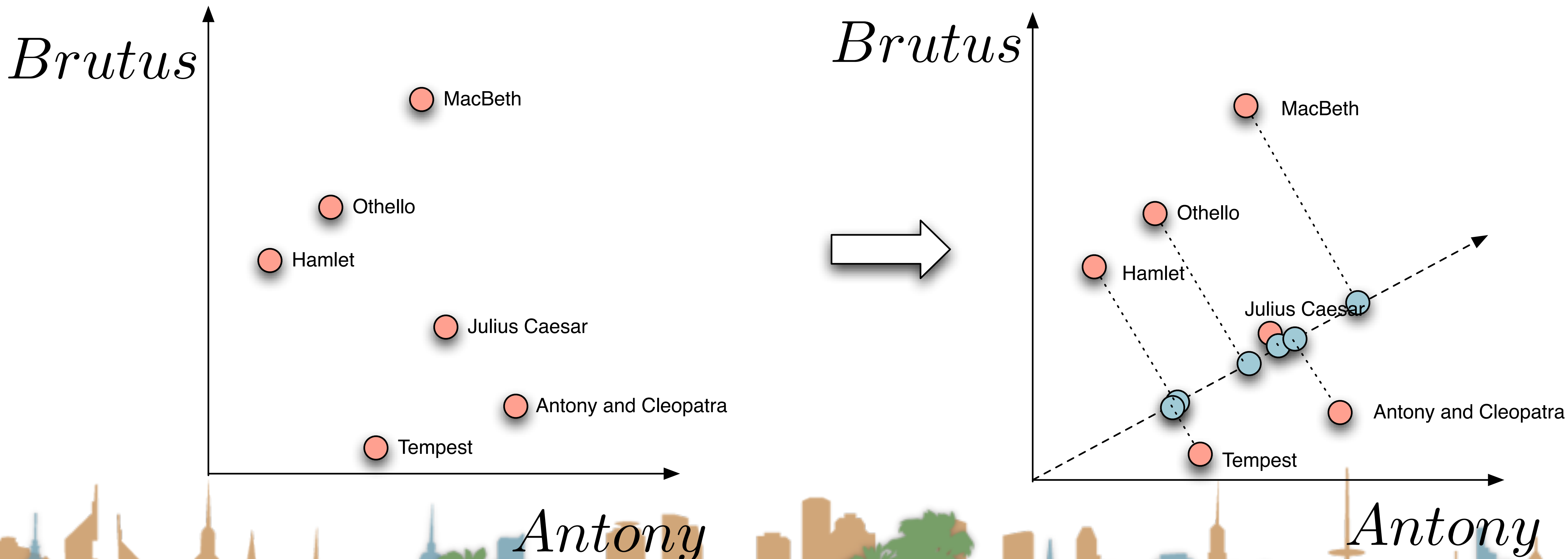
Mathematically speaking

- Our documents were just points in an N-dimensional term space
- We can project them also



Mathematically speaking

- Latent Semantic Indexing makes the claim that these new axes represent semantics - deeper meaning than just a term



Matrix Decomposition

- Singular Value Decomposition
  - Splits a matrix into three matrices
  - Such that
    - If
    - then
    - and
    - and
    - also Sigma is almost a diagonal matrix

Matrix Decomposition

- Singular Value Decomposition
  - Splits a matrix into three matrices $U \quad \Sigma \quad V^T$
  - Such that
    - If
    - then
    - and
    - and
    - also Sigma is almost a diagonal matrix

Matrix Decomposition

- Singular Value Decomposition
  - Splits a matrix into three matrices $U \quad \Sigma \quad V^T$
  - Such that $C = U\Sigma V^T$
    - If
    - then
    - and
    - and
    - also Sigma is almost a diagonal matrix

Matrix Decomposition

- Singular Value Decomposition
  - Splits a matrix into three matrices $U \quad \Sigma \quad V^T$
  - Such that $C = U\Sigma V^T$
    - If $C\ is\ (M\ by\ N)$
    - then
    - and
    - and
    - also Sigma is almost a diagonal matrix

Matrix Decomposition

- Singular Value Decomposition
  - Splits a matrix into three matrices $U \quad \Sigma \quad V^T$
  - Such that $C = U\Sigma V^T$
    - If $C\ is\ (M\ by\ N)$
    - then $U\ is\ (M\ by\ M)$
    - and
    - and
    - also Sigma is almost a diagonal matrix

Matrix Decomposition

- Singular Value Decomposition
  - Splits a matrix into three matrices $U \quad \Sigma \quad V^T$
  - Such that $C = U\Sigma V^T$
    - If $C\ is\ (M\ by\ N)$
    - then $U\ is\ (M\ by\ M)$
    - and $\Sigma\ is\ (M\ by\ N)$
    - and
    - also Sigma is almost a diagonal matrix

Matrix Decomposition

- Singular Value Decomposition
  - Splits a matrix into three matrices $U \quad \Sigma \quad V^T$
  - Such that $C = U\Sigma V^T$
    - If $C\ is\ (M\ by\ N)$
    - then $U\ is\ (M\ by\ M)$
    - and $\Sigma\ is\ (M\ by\ N)$
    - and $V^T\ is\ (N\ by\ N)$
    - also Sigma is almost a diagonal matrix

Matrix Decomposition

- Singular Value Decomposition
  - the columns of V are orthogonal eigenvectors of $C^T C$
  - the columns of U are orthogonal eigenvectors of CC^T
  - V represents our "term space"
  - U represents our "document space"
  - We want to transform our vectors from or original term space into our new term space

$$U^T C = U^T U \Sigma V^T$$

$$\Sigma^{-1} U^T C = \Sigma^{-1} \Sigma V^T$$

$$\Sigma^{-1} U^T C = V^T$$

Matrix Decomposition

- Singular Value Decomposition

$$C = U\Sigma V^T$$

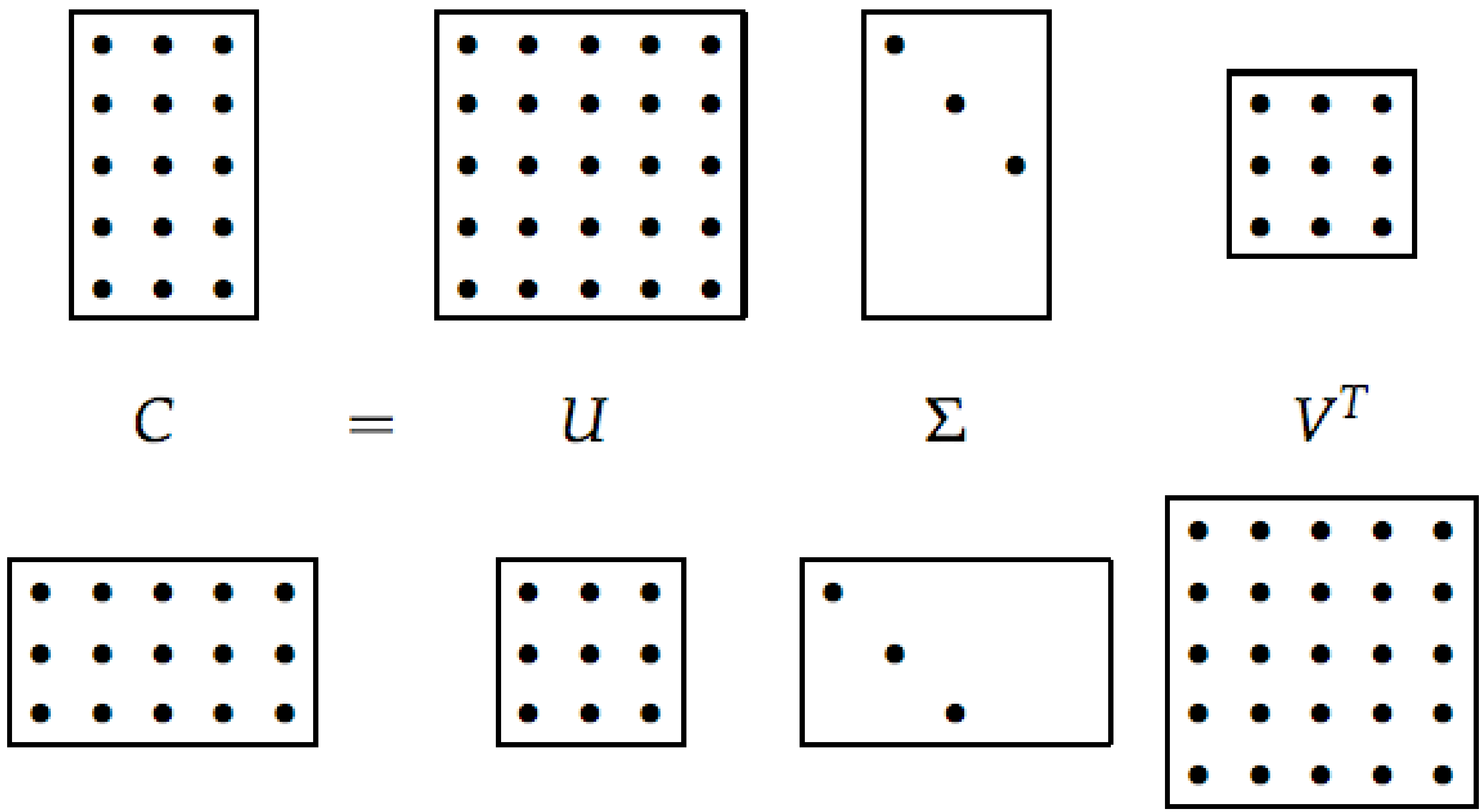
  - the columns of V are orthogonal eigenvectors of $C^T C$
  - the columns of U are orthogonal eigenvectors of CC^T
  - V represents our "term space"
  - U represents our "document space"
  - We want to transform our vectors from or original term space into our new term space

$$U^T C = U^T U \Sigma V^T$$

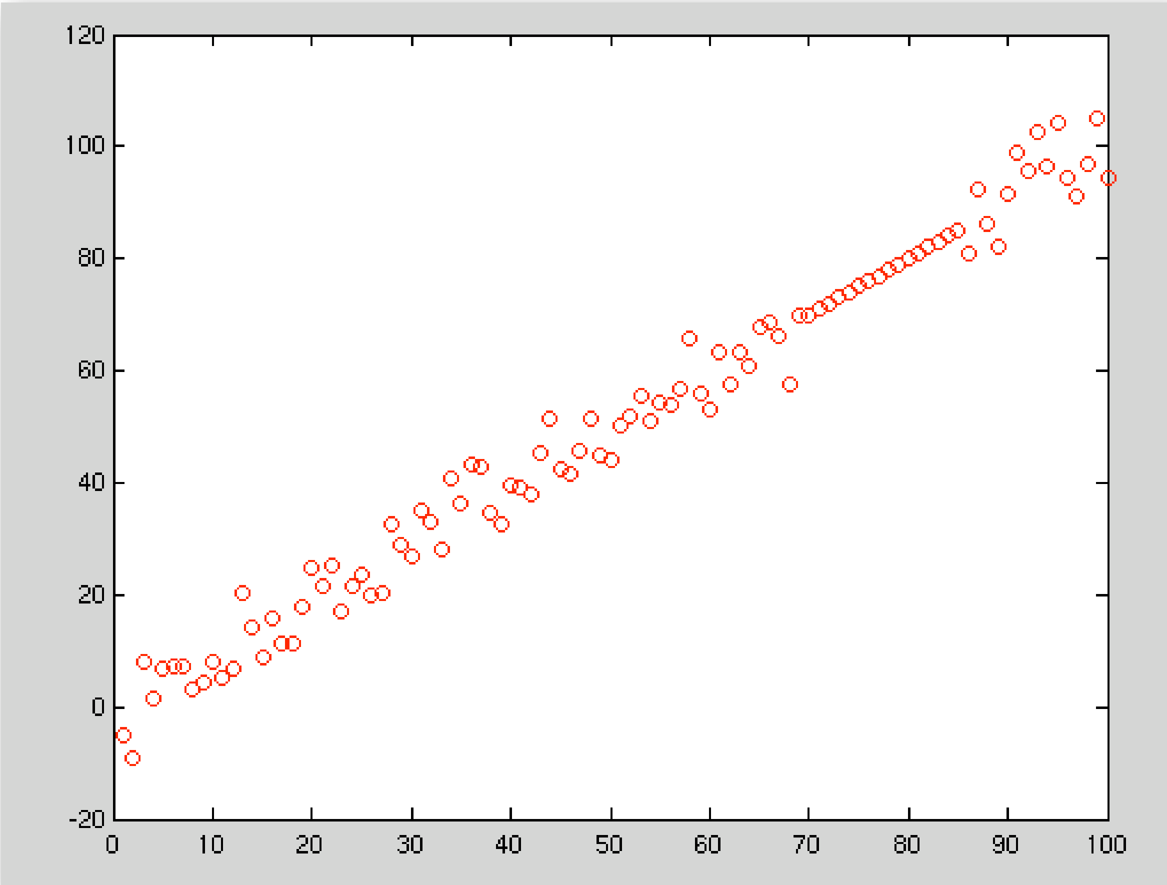
$$\Sigma^{-1} U^T C = \Sigma^{-1} \Sigma V^T$$

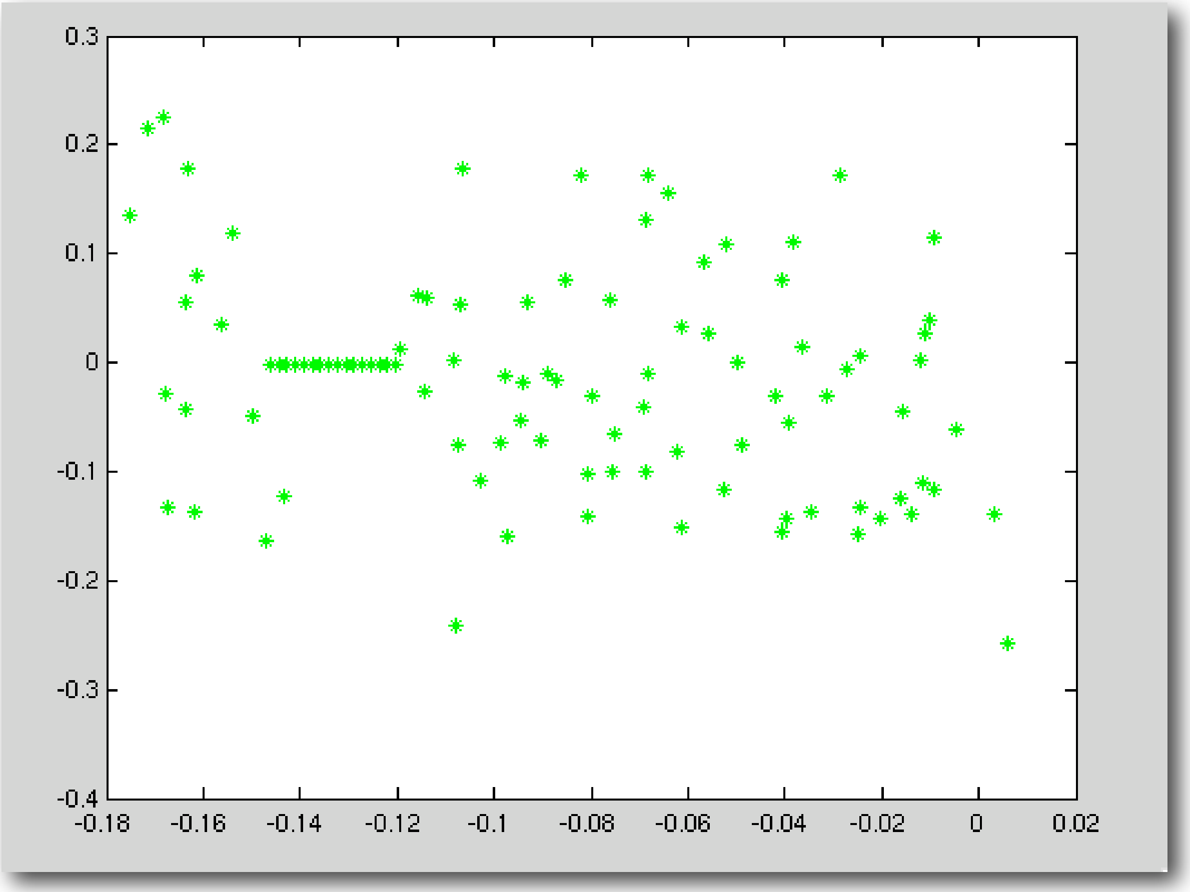
$$\Sigma^{-1} U^T C = V^T$$

Matrix Decomposition



Matrix Decomposition





Matrix Decomposition

- Singular Value Decomposition

$$\begin{vmatrix} SVD_{dim_1} \\ SVD_{dim_2} \\ SVD_{dim_3} \end{vmatrix} = \begin{vmatrix} a & b & c & d \\ a' & b' & c' & d' \\ a'' & b'' & c'' & d'' \end{vmatrix} * \begin{vmatrix} td_{dim_1} \\ td_{dim_2} \\ td_{dim_3} \\ td_{dim_4} \end{vmatrix}$$

$$SVD_{ConceptSpace} = M * query_{TermSpace}$$

$$M = \Sigma_k^{-1} U_k^T$$

$$query_{ConceptSpace} = \Sigma_k^{-1} U_k^T query_{TermSpace}$$

next...

LUCI