Spectral Analysis of the Vector Space Model (and Explicit Semantic Analysis)

Preliminary Report

Jérôme Kunegis & Thomas Gottron

Information Retrieval

Search results

From Wikipedia, the free encyclopedia

berlin Search

Content pages Multimedia Help and Project pages Everything Advanced

There is a page named "Berlin" on Wikipedia
For search help, please visit Help:Searching.

Berlin
Berlin. (bɜr'lɪn; bɛɐ̯'liːn | de-**Berlin**. ogg. is the capital city of Germany , and is one of the 16 states of Germany . **...**
105 KB (14,511 words) - 12:36, 23 April 2011

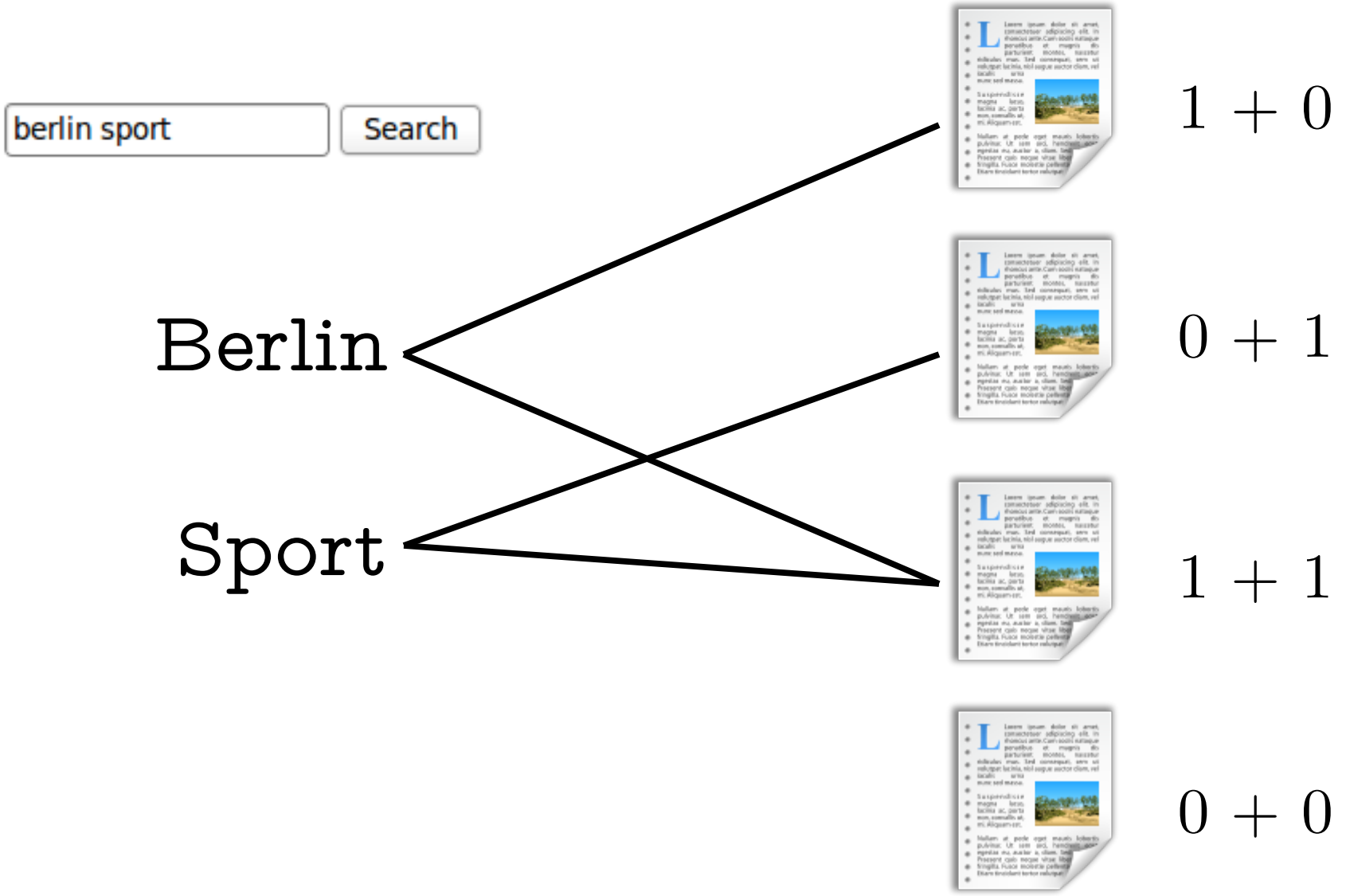
.berlin
.berlin (dotBERLIN) is a proposed new top level domain (TLD). It is a Sponsored top-level domain intended to be a top level domain for **...**
3 KB (465 words) - 17:15, 10 April 2011

Berlin Plus agreement (redirect from **Berlin+**)
The **Berlin** Plus agreement is the short title of a comprehensive package of agreements made between NATO and the EU on 16 December 2002 **...**
4 KB (562 words) - 04:57, 27 January 2011

Berlín
Berlín is a municipality in the Usulután department of El Salvador . Overview: The municipality of **Berlín** is made up of an urban center **...**
3 KB (431 words) - 06:22, 14 March 2011

Term Frequency

Term Frequency: Two Query Terms



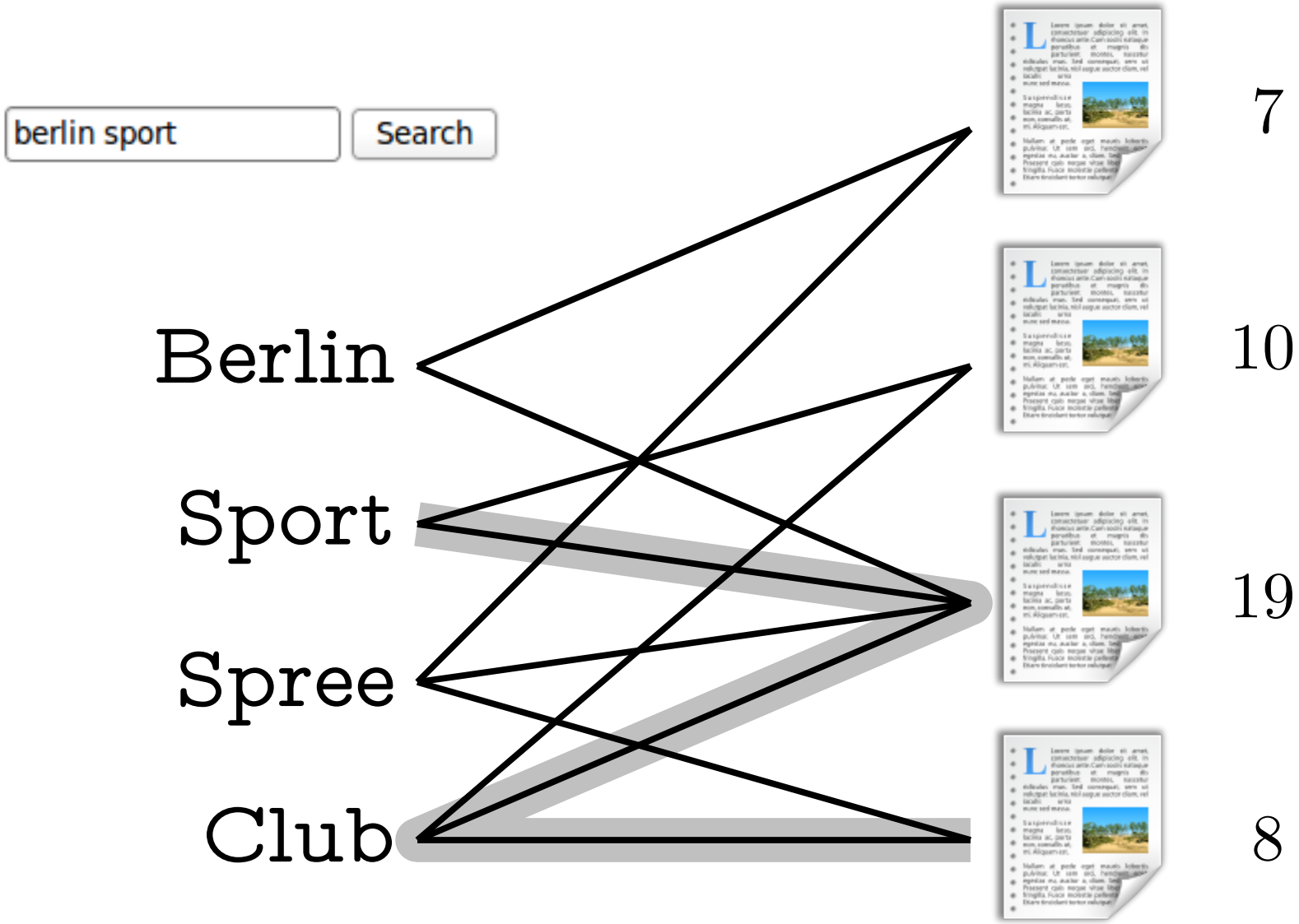
Vector Space Model

$$\mathbf{q} = \begin{bmatrix} \text{Asia} & \text{Koblenz} & \text{Weather} & \text{Money} & \text{Berlin} & \text{Christmas} & \text{Dresden} & \text{Spree} & \text{Club} & \text{Sport} & \text{Water} & \text{Hamster} & \text{Investment} \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \end{bmatrix}$$

$$\mathbf{M} = \left|\begin{array}{cccccccccccc} 0 & 0 & 0 & 1 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 1 & 0 & 0 & 1 & 1 & 1 & 1 & 1 & 0 \\ 0 & 0 & 1 & 1 & 0 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 1 & 1 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \end{array}\right|$$

$$\mathbf{M}\mathbf{q}^{\mathrm{T}} = \left|\begin{array}{c} 1 \\ 1 \\ 2 \\ 0 \\ 0 \\ 0 \\ 0 \end{array}\right|$$

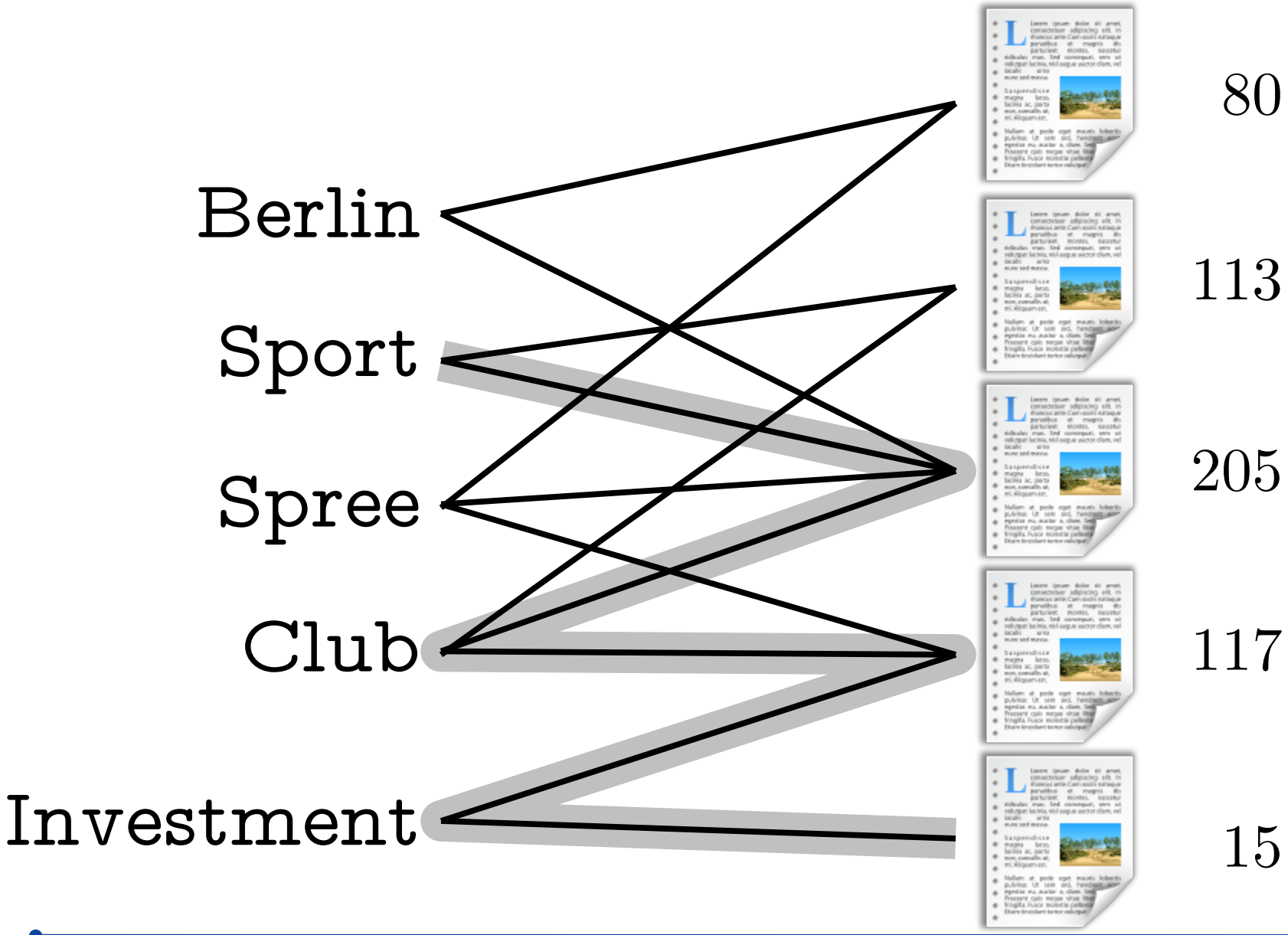
Extended Vector Space Model



Extended Vector Space Model

$$\mathbf{M}\mathbf{M}^{\mathrm{T}}\mathbf{M}\mathbf{q}^{\mathrm{T}} = \begin{vmatrix} 7 \\ 10 \\ 19 \\ 8 \\ 0 \\ 5 \\ 2 \end{vmatrix}$$

Extended Vector Space Model



Extended Vector Space Model

$$\mathbf{M}\mathbf{M}^{\mathrm{T}}\mathbf{M}\mathbf{M}^{\mathrm{T}}\mathbf{M}\mathbf{q}^{\mathrm{T}} = \begin{vmatrix} 80 \\ 113 \\ 205 \\ 117 \\ 15 \\ 73 \\ 23 \end{vmatrix}$$

Extended Vector Space Model: Computation

$$\mathbf{M} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^{\mathrm{T}}$$

Singular value decomposition

$$\mathbf{U}^{\mathrm{T}}\mathbf{U} = \mathbf{I}$$

$$\mathbf{V}^{\mathrm{T}}\mathbf{V} = \mathbf{I}$$

$\mathbf{\Sigma}$ diagonal

Extended Vector Space Model: Computation

$$\mathbf{M} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^{\mathrm{T}}$$

$$\begin{aligned}\mathbf{M}\mathbf{M}^{\mathrm{T}}\mathbf{M} &= \mathbf{U}\mathbf{\Sigma}\mathbf{V}^{\mathrm{T}}\,\mathbf{V}\mathbf{\Sigma}\mathbf{U}^{\mathrm{T}}\,\mathbf{U}\mathbf{\Sigma}\mathbf{V}^{\mathrm{T}} \\ &= \mathbf{U}\mathbf{\Sigma}^{3}\mathbf{V}^{\mathrm{T}}\end{aligned}$$

$$\begin{aligned}\mathbf{M}\mathbf{M}^{\mathrm{T}}\mathbf{M}\mathbf{M}^{\mathrm{T}}\mathbf{M} &= \mathbf{U}\mathbf{\Sigma}\mathbf{V}^{\mathrm{T}}\,\mathbf{V}\mathbf{\Sigma}\mathbf{U}^{\mathrm{T}}\,\mathbf{U}\mathbf{\Sigma}\mathbf{V}^{\mathrm{T}}\,\mathbf{V}\mathbf{\Sigma}\mathbf{U}^{\mathrm{T}}\,\mathbf{U}\mathbf{\Sigma}\mathbf{V}^{\mathrm{T}} \\ &= \mathbf{U}\mathbf{\Sigma}^{5}\mathbf{V}^{\mathrm{T}}\end{aligned}$$

Generalized Vector Space Model

$$a\mathbf{M} + b\mathbf{M}\mathbf{M}^{\mathrm{T}}\mathbf{M} + c\mathbf{M}\mathbf{M}^{\mathrm{T}}\mathbf{M}\mathbf{M}^{\mathrm{T}}\mathbf{M} + \ldots$$

$$= \mathbf{U}\,(\,a\boldsymbol{\Sigma} + b\boldsymbol{\Sigma}^3 + c\boldsymbol{\Sigma}^5 + \ldots\,)\,\mathbf{V}^{\mathrm{T}}$$

for $a > b > c > \ldots > 0$

Example:

$$\sinh(\boldsymbol{\Sigma}) = \boldsymbol{\Sigma} + (1/6)\,\boldsymbol{\Sigma}^3 + (1/120)\,\boldsymbol{\Sigma}^5 + \ldots$$

Does It Work?

TREC document collection:

$\mathbf{M}$:	$528{,}155 \times 829{,}883$	(document × term)
$\mathbf{Q}$:	$50 \times 829{,}883$	(query × term)
$\mathbf{S}$:	$528{,}155 \times 50$	(document × query)

Learning

$$\text{Let } \mathbf{M} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^{\mathrm{T}}$$

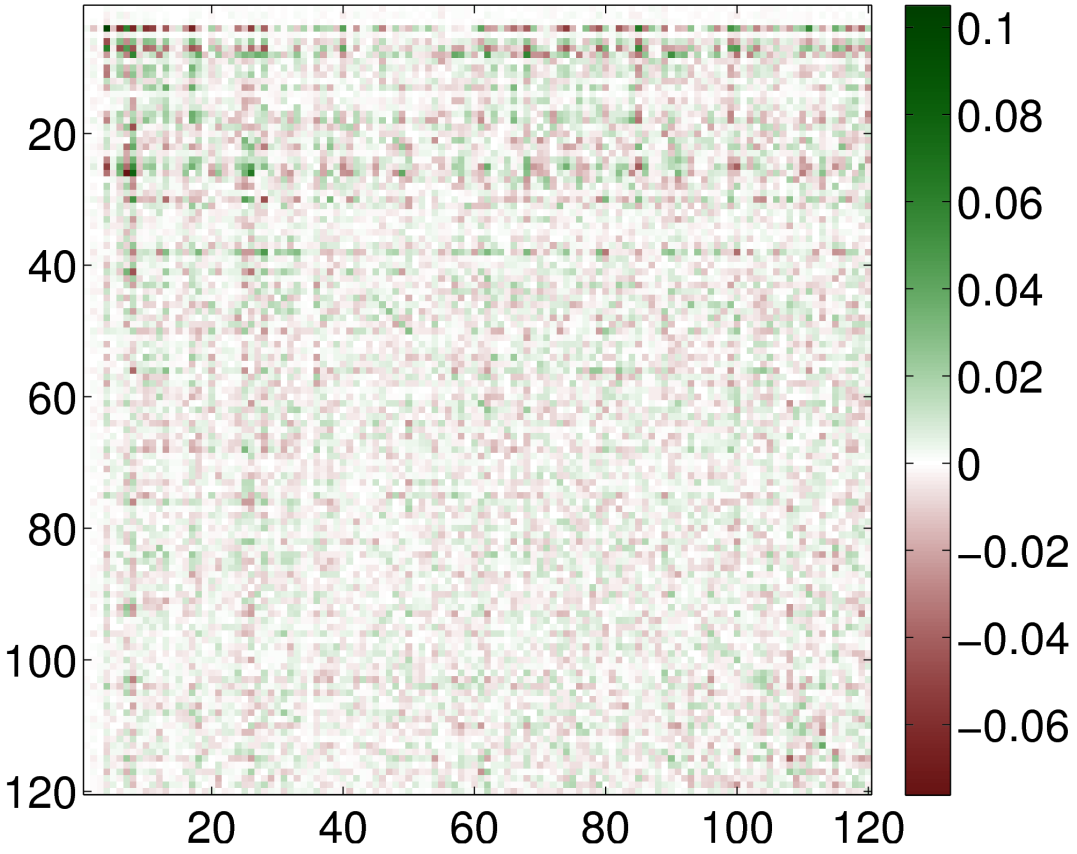
$$\mathbf{S} = \mathrm{f}(\mathbf{M})\mathbf{Q}^{\mathrm{T}}$$

$$\mathbf{S} = \mathbf{U}\mathrm{f}(\mathbf{\Sigma})\mathbf{V}^{\mathrm{T}}\mathbf{Q}^{\mathrm{T}}$$

$$\mathbf{U}^{\mathrm{T}}\mathbf{S}(\mathbf{Q}^{\mathrm{T}})^{-1}\mathbf{V} = \mathrm{f}(\mathbf{\Sigma})$$

Diagonality Test

The matrix $\mathbf{U}^{\mathrm{T}}\mathbf{S}(\mathbf{Q}^{\mathrm{T}})^{-1}\mathbf{V}$ should be diagonal:



Conclusions

- The vector space model will not work here
- The dataset is too small: only 50 queries
- To do: normalization (tf-idf, . . .)