Large Graph Mining: Power Tools and a Practitioner's guide

Task 1: Node importance

Faloutsos, Miller, Tsourakakis

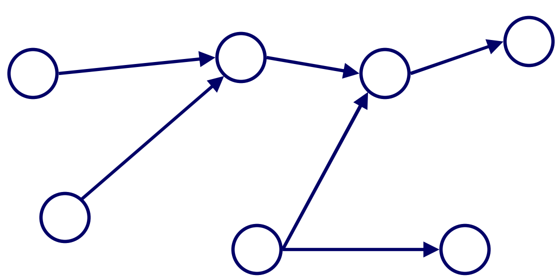
CMU

Outline

- Introduction – Motivation
- **Task 1: Node importance**
- Task 2: Community detection
- Task 3: Recommendations
- Task 4: Connection sub-graphs
- Task 5: Mining graphs over time
- Task 6: Virus/influence propagation
- Task 7: Spectral graph theory
- Task 8: Tera/peta graph mining: hadoop
- Observations – patterns of real graphs
- Conclusions

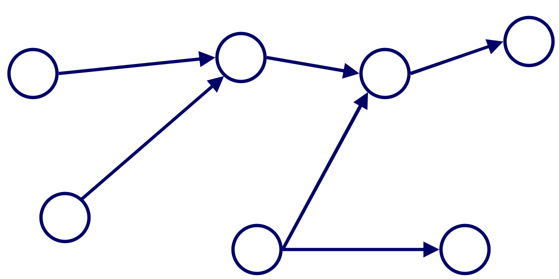
Node importance - Motivation:

- Given a graph (eg., web pages containing the desirable query word)
- Q: Which node is the most important?



Node importance - Motivation:

- Given a graph (eg., web pages containing the desirable query word)
- Q: Which node is the most important?
- A1: HITS (SVD = Singular Value Decomposition)
- A2: eigenvector (PageRank)



Node importance - motivation

- SVD and eigenvector analysis: very closely related
- See 'theory Task', later

SVD - Detailed outline

- Motivation
- Definition - properties
- Interpretation
- Complexity
- Case studies

SVD - Motivation

- problem #1: text - LSI: find ‘concepts’
- problem #2: compression / dim. reduction

SVD - Motivation

- problem #1: text - LSI: find ‘concepts’

term / document	data	information	retrieval	brain	lung
CS-TR1	1	1	1	0	0
CS-TR2	2	2	2	0	0
CS-TR3	1	1	1	0	0
CS-TR4	5	5	5	0	0
MED-TR1	0	0	0	2	2
MED-TR2	0	0	0	3	3
MED-TR3	0	0	0	1	1

SVD - Motivation

- Customer-product, for recommendation system:

$$\begin{array}{cc} & \begin{array}{ccccc} \text{bread} & \text{lettuce} & \text{tomatos} & \text{beef} & \text{chicken} \end{array} \\ \begin{array}{c} \\ \\ \text{vegetarians} \\ \\ \\ \text{meat eaters} \\ \\ \end{array} & \begin{bmatrix} 1 & 1 & 1 & 0 & 0 \\ 2 & 2 & 2 & 0 & 0 \\ 1 & 1 & 1 & 0 & 0 \\ 5 & 5 & 5 & 0 & 0 \\ 0 & 0 & 0 & 2 & 2 \\ 0 & 0 & 0 & 3 & 3 \\ 0 & 0 & 0 & 1 & 1 \end{bmatrix} \end{array}$$

details

SVD - Motivation

- problem #2: compress / reduce dimensionality

details

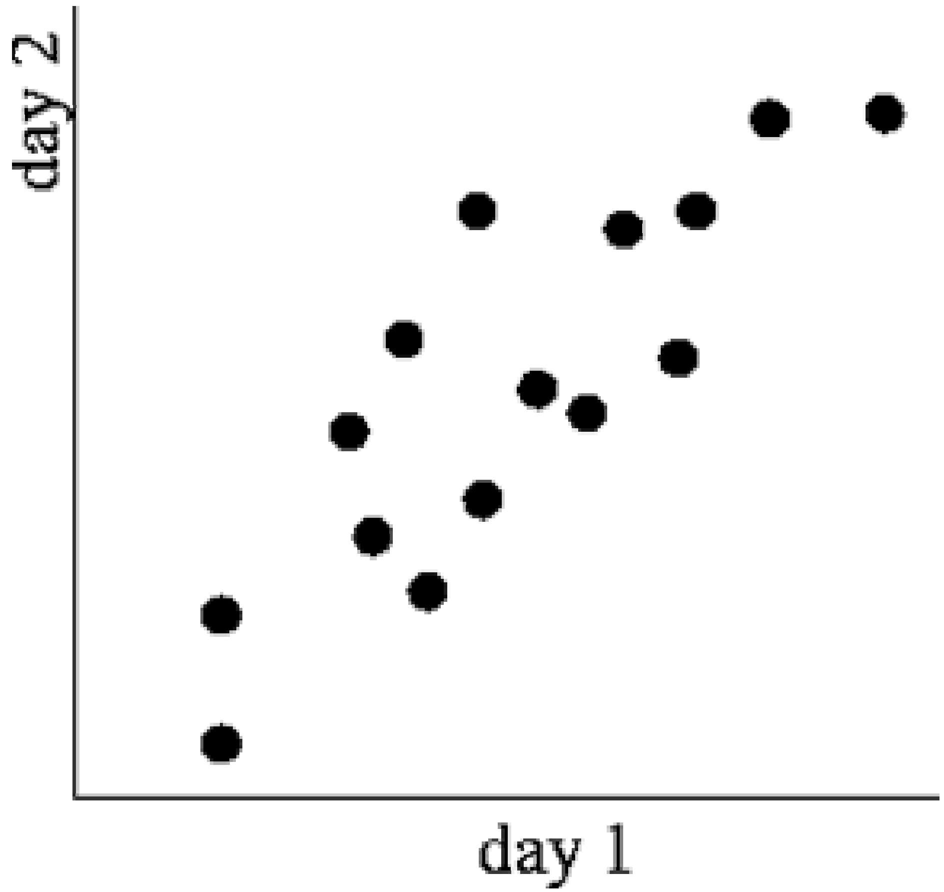
Problem - specs

- ~10**6 rows; ~10**3 columns; no updates;
- random access to any cell(s) ; small error: OK

day / customer	We 7/10/96	Th 7/11/96	Fr 7/12/96	Sa 7/13/96	Su 7/14/96
ABC Inc.	1	1	1	0	0
DEF Ltd.	2	2	2	0	0
GHI Inc.	1	1	1	0	0
KLM Co.	5	5	5	0	0
Smith	0	0	0	2	2
Johnson	0	0	0	3	3
Thompson	0	0	0	1	1

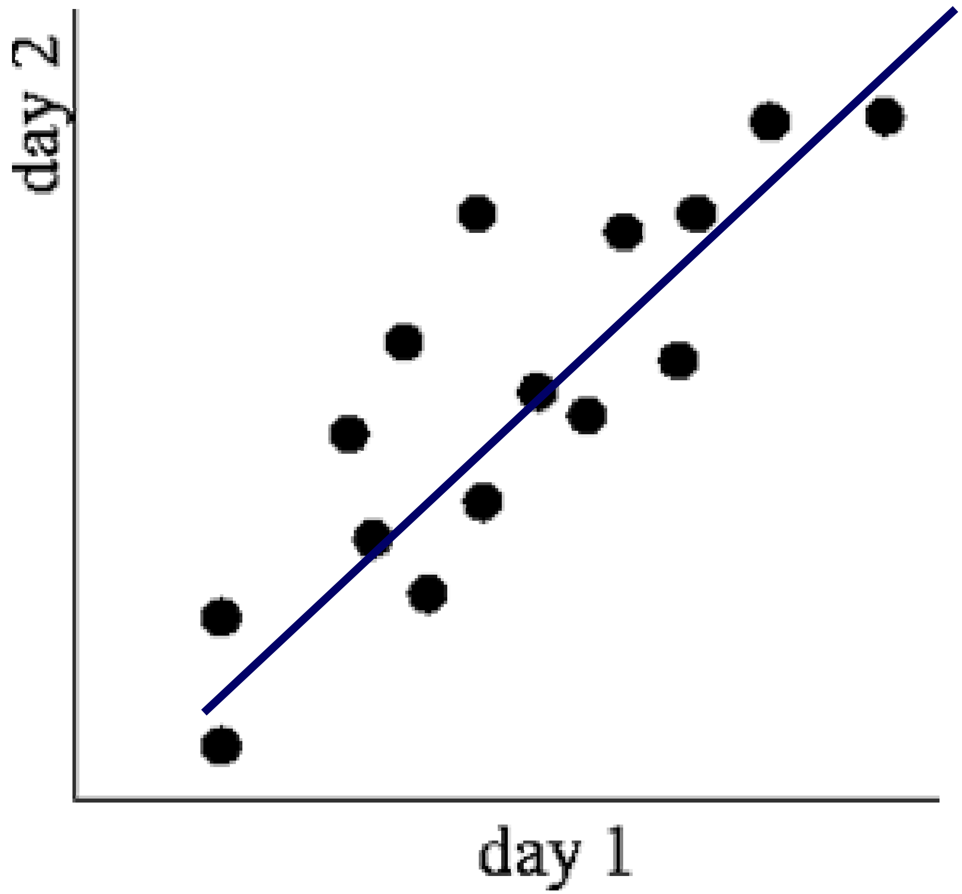
details

SVD - Motivation



details

SVD - Motivation



SVD - Detailed outline

- Motivation
- Definition - properties
- Interpretation
- Complexity
- Case studies
- Additional properties

SVD - Definition

(reminder: matrix multiplication

$$\begin{bmatrix} 1 & 2 \\ 3 & 4 \\ 5 & 6 \end{bmatrix} \text{x} \begin{bmatrix} 1 \\ -1 \end{bmatrix} = \begin{bmatrix} \\ \\ \\ \end{bmatrix}$$

3 x *2* *2* x *1*

SVD - Definition

(reminder: matrix multiplication

$$\begin{bmatrix} 1 & 2 \\ 3 & 4 \\ 5 & 6 \end{bmatrix} \times \begin{bmatrix} 1 \\ -1 \end{bmatrix} = \begin{bmatrix} \\ \\ \\ \end{bmatrix}$$

3 x *2*   *2* x *1*   *3* x *1*

SVD - Definition

(reminder: matrix multiplication

$$\begin{bmatrix} 1 & 2 \\ 3 & 4 \\ 5 & 6 \end{bmatrix} \times \begin{bmatrix} 1 \\ -1 \end{bmatrix} = \begin{bmatrix} -1 \\ \\ \\ \end{bmatrix}$$

3 x *2*   *2* x *1*   *3* x *1*

SVD - Definition

(reminder: matrix multiplication

$$\begin{bmatrix} 1 & 2 \\ 3 & 4 \\ 5 & 6 \end{bmatrix} \times \begin{bmatrix} 1 \\ -1 \end{bmatrix} = \begin{bmatrix} -1 \\ -1 \\ \ \end{bmatrix}$$

3 x *2*　　*2* x *1*　　*3* x *1*

SVD - Definition

(reminder: matrix multiplication

$$\begin{bmatrix} 1 & 2 \\ 3 & 4 \\ 5 & 6 \end{bmatrix} \text{ x } \begin{bmatrix} 1 \\ -1 \end{bmatrix} = \begin{bmatrix} -1 \\ -1 \\ -1 \end{bmatrix}$$

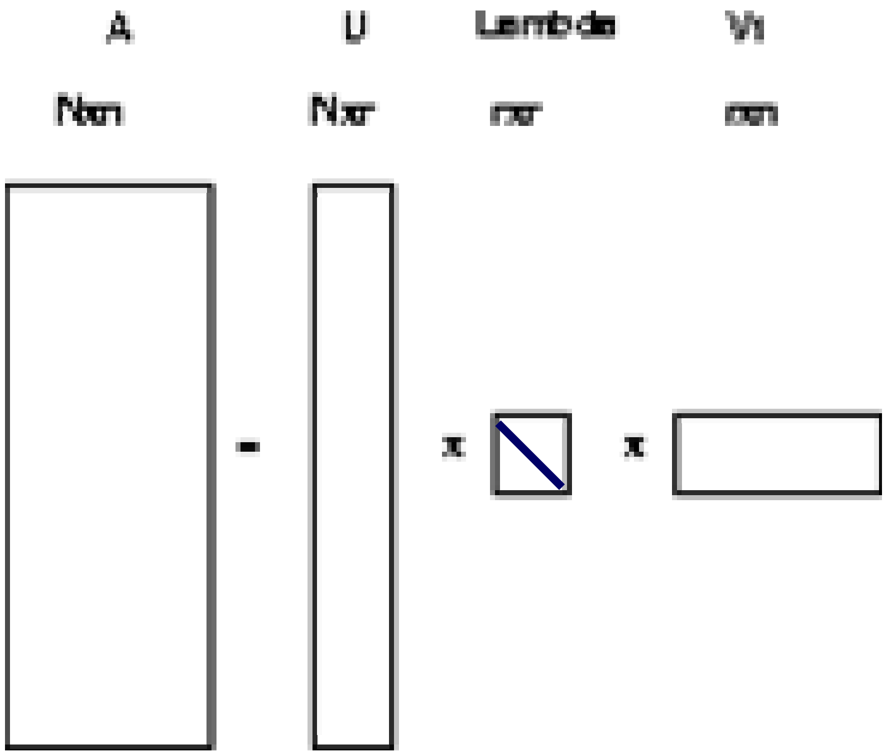
SVD - Definition

$$\boxed{\mathbf{A}_{[\mathbf{n} \times \mathbf{m}]} = \mathbf{U}_{[\mathbf{n} \times \mathbf{r}]} \, \mathbf{\Lambda}_{[\mathbf{r} \times \mathbf{r}]} \, (\mathbf{V}_{[\mathbf{m} \times \mathbf{r}]})^T}$$

- **A**: n x m matrix (eg., n documents, m terms)
- **U**: n x r matrix (n documents, r concepts)
- **Λ**: r x r diagonal matrix (strength of each 'concept') (r : rank of the matrix)
- **V**: m x r matrix (m terms, r concepts)

SVD - Definition

- $\mathbf{A} = \mathbf{U}\ \mathbf{\Lambda}\ \mathbf{V}^T$ - example:



SVD - Properties

THEOREM [Press+92]: always possible to decompose matrix **A** into $\mathbf{A} = \mathbf{U}\ \mathbf{\Lambda}\ \mathbf{V}^T$, where

- **U, Λ, V**: unique (*)
- **U**, **V**: column orthonormal (ie., columns are unit vectors, orthogonal to each other)
  - $\mathbf{U}^T\ \mathbf{U} = \mathbf{I}$; $\mathbf{V}^T\ \mathbf{V} = \mathbf{I}$ (**I:** identity matrix)
- **Λ:** singular are positive, and sorted in decreasing order

SVD - Example

- $\mathbf{A} = \mathbf{U}\,\mathbf{\Lambda}\,\mathbf{V}^T$ - example:

$$
\begin{array}{c}
\\ \text{CS} \\ \\ \\ \text{MD} \\ \\
\end{array}
\begin{array}{c}
\begin{array}{ccccc} \text{data} & \text{inf.} & \text{retrieval} & \text{brain} & \text{lung} \end{array} \\
\begin{bmatrix}
1 & 1 & 1 & 0 & 0 \\
2 & 2 & 2 & 0 & 0 \\
1 & 1 & 1 & 0 & 0 \\
5 & 5 & 5 & 0 & 0 \\
0 & 0 & 0 & 2 & 2 \\
0 & 0 & 0 & 3 & 3 \\
0 & 0 & 0 & 1 & 1
\end{bmatrix}
\end{array}
=
\begin{bmatrix}
0.18 & 0 \\
0.36 & 0 \\
0.18 & 0 \\
0.90 & 0 \\
0 & 0.53 \\
0 & 0.80 \\
0 & 0.27
\end{bmatrix}
\mathrm{x}
\begin{bmatrix}
9.64 & 0 \\
0 & 5.29
\end{bmatrix}
\mathrm{x}
\begin{bmatrix}
0.58 & 0.58 & 0.58 & 0 & 0 \\
0 & 0 & 0 & 0.71 & 0.71
\end{bmatrix}
$$

SVD - Example

- $\mathbf{A} = \mathbf{U}\, \boldsymbol{\Lambda}\, \mathbf{V}^T$ - example:

CS-concept

MD-concept

$$
\begin{array}{c|c}
 & \text{data} \quad \text{inf.} \quad \text{retrieval} \quad \text{brain} \quad \text{lung} \\
\begin{array}{c} \text{CS} \\ \\ \\ \\ \text{MD} \\ \\ \\ \end{array} &
\begin{bmatrix} 1 & 1 & 1 & 0 & 0 \\ 2 & 2 & 2 & 0 & 0 \\ 1 & 1 & 1 & 0 & 0 \\ 5 & 5 & 5 & 0 & 0 \\ 0 & 0 & 0 & 2 & 2 \\ 0 & 0 & 0 & 3 & 3 \\ 0 & 0 & 0 & 1 & 1 \end{bmatrix}
\end{array}
=
\begin{bmatrix} 0.18 & 0 \\ 0.36 & 0 \\ 0.18 & 0 \\ 0.90 & 0 \\ 0 & 0.53 \\ 0 & 0.80 \\ 0 & 0.27 \end{bmatrix}
\times
\begin{bmatrix} 9.64 & 0 \\ 0 & 5.29 \end{bmatrix}
\times
\begin{bmatrix} 0.58 & 0.58 & 0.58 & 0 & 0 \\ 0 & 0 & 0 & 0.71 & 0.71 \end{bmatrix}
$$

SVD - Example

- $\mathbf{A} = \mathbf{U}\, \Lambda\, \mathbf{V}^T$ - example:

doc-to-concept similarity matrix

CS-concept

MD-concept

$$
\begin{array}{c|c}
 & \begin{array}{ccccc} \text{data} & \text{inf.} & \text{retrieval} & \text{brain} & \text{lung} \end{array} \\
\begin{array}{c} \text{CS} \\ \\ \\ \\ \text{MD} \\ \\ \\ \end{array} &
\begin{bmatrix} 1 & 1 & 1 & 0 & 0 \\ 2 & 2 & 2 & 0 & 0 \\ 1 & 1 & 1 & 0 & 0 \\ 5 & 5 & 5 & 0 & 0 \\ 0 & 0 & 0 & 2 & 2 \\ 0 & 0 & 0 & 3 & 3 \\ 0 & 0 & 0 & 1 & 1 \end{bmatrix}
\end{array}
=
\begin{bmatrix} 0.18 & 0 \\ 0.36 & 0 \\ 0.18 & 0 \\ 0.90 & 0 \\ 0 & 0.53 \\ 0 & 0.80 \\ 0 & 0.27 \end{bmatrix}
\times
\begin{bmatrix} 9.64 & 0 \\ 0 & 5.29 \end{bmatrix}
\times
\begin{bmatrix} 0.58 & 0.58 & 0.58 & 0 & 0 \\ 0 & 0 & 0 & 0.71 & 0.71 \end{bmatrix}
$$

SVD - Example

- $\mathbf{A} = \mathbf{U}\,\mathbf{\Lambda}\,\mathbf{V}^T$ - example:

'strength' of CS-concept

$$
\begin{array}{c|c}
 & \text{data} \quad \text{inf.} \quad \text{retrieval} \quad \text{brain} \quad \text{lung} \\
\begin{array}{c} \text{CS} \\ \\ \\ \\ \text{MD} \\ \\ \\ \end{array} &
\begin{bmatrix} 1 & 1 & 1 & 0 & 0 \\ 2 & 2 & 2 & 0 & 0 \\ 1 & 1 & 1 & 0 & 0 \\ 5 & 5 & 5 & 0 & 0 \\ 0 & 0 & 0 & 2 & 2 \\ 0 & 0 & 0 & 3 & 3 \\ 0 & 0 & 0 & 1 & 1 \end{bmatrix}
\end{array}
=
\begin{bmatrix} 0.18 & 0 \\ 0.36 & 0 \\ 0.18 & 0 \\ 0.90 & 0 \\ 0 & 0.53 \\ 0 & 0.80 \\ 0 & 0.27 \end{bmatrix}
\times
\begin{bmatrix} 9.64 & 0 \\ 0 & 5.29 \end{bmatrix}
\times
\begin{bmatrix} 0.58 & 0.58 & 0.58 & 0 & 0 \\ 0 & 0 & 0 & 0.71 & 0.71 \end{bmatrix}
$$

SVD - Example

- $\mathbf{A} = \mathbf{U}\,\mathbf{\Lambda}\,\mathbf{V}^T$ - example:

term-to-concept similarity matrix

CS-concept

$$
\begin{array}{c|ccccc}
 & \text{data} & \text{inf.} & \text{retrieval} & \text{brain} & \text{lung} \\
\hline
\text{CS} & \\
\text{MD} &
\end{array}
\quad
\left[\begin{array}{ccccc}
1 & 1 & 1 & 0 & 0 \\
2 & 2 & 2 & 0 & 0 \\
1 & 1 & 1 & 0 & 0 \\
5 & 5 & 5 & 0 & 0 \\
0 & 0 & 0 & 2 & 2 \\
0 & 0 & 0 & 3 & 3 \\
0 & 0 & 0 & 1 & 1
\end{array}\right]
=
\left[\begin{array}{cc}
0.18 & 0 \\
0.36 & 0 \\
0.18 & 0 \\
0.90 & 0 \\
0 & 0.53 \\
0 & 0.80 \\
0 & 0.27
\end{array}\right]
\text{x}
\left[\begin{array}{cc}
9.64 & 0 \\
0 & 5.29
\end{array}\right]
\text{x}
\left[\begin{array}{ccccc}
0.58 & 0.58 & 0.58 & 0 & 0 \\
0 & 0 & 0 & 0.71 & 0.71
\end{array}\right]
$$

SVD - Example

- $\mathbf{A} = \mathbf{U}\, \mathbf{\Lambda}\, \mathbf{V}^T$ - example:

term-to-concept similarity matrix

$$
\begin{array}{c|c}
 & \text{data} \quad \text{inf.} \quad \text{retrieval} \quad \text{brain} \quad \text{lung} \\
\begin{array}{c} \text{CS} \\ \\ \\ \\ \text{MD} \\ \\ \\ \end{array} &
\begin{bmatrix} 1 & 1 & 1 & 0 & 0 \\ 2 & 2 & 2 & 0 & 0 \\ 1 & 1 & 1 & 0 & 0 \\ 5 & 5 & 5 & 0 & 0 \\ 0 & 0 & 0 & 2 & 2 \\ 0 & 0 & 0 & 3 & 3 \\ 0 & 0 & 0 & 1 & 1 \end{bmatrix}
\end{array}
=
\begin{bmatrix} 0.18 & 0 \\ 0.36 & 0 \\ 0.18 & 0 \\ 0.90 & 0 \\ 0 & 0.53 \\ 0 & 0.80 \\ 0 & 0.27 \end{bmatrix}
\times
\begin{bmatrix} 9.64 & 0 \\ 0 & 5.29 \end{bmatrix}
\times
\begin{bmatrix} 0.58 & 0.58 & 0.58 & 0 & 0 \\ 0 & 0 & 0 & 0.71 & 0.71 \end{bmatrix}
$$

CS-concept

SVD - Detailed outline

- Motivation
- Definition - properties
- Interpretation
- Complexity
- Case studies
- Additional properties

SVD - Interpretation #1

'documents', 'terms' and 'concepts':

- **U**: document-to-concept similarity matrix
- **V**: term-to-concept sim. matrix
- Λ: its diagonal elements: 'strength' of each concept

SVD – Interpretation #1

'documents', 'terms' and 'concepts':

Q**:** if **A** is the document-to-term matrix, what is $\mathbf{A}^T \mathbf{A}$?

A:

Q: $\mathbf{A} \mathbf{A}^T$?

A:

SVD – Interpretation #1

'documents', 'terms' and 'concepts':

Q**:** if **A** is the document-to-term matrix, what is $\mathbf{A}^T \mathbf{A}$?

A: term-to-term ([m x m]) similarity matrix

Q: $\mathbf{A}\,\mathbf{A}^T$?

A: document-to-document ([n x n]) similarity matrix

SVD properties

- **V** are the eigenvectors of the *covariance matrix* $\mathbf{A}^T\mathbf{A}$

- **U** are the eigenvectors of the *Gram (inner-product) matrix* $\mathbf{A}\mathbf{A}^T$

Further reading:
1. Ian T. Jolliffe, *Principal Component Analysis* (2nd ed), Springer, 2002.
2. Gilbert Strang, *Linear Algebra and Its Applications* (4th ed), Brooks Cole, 2005.

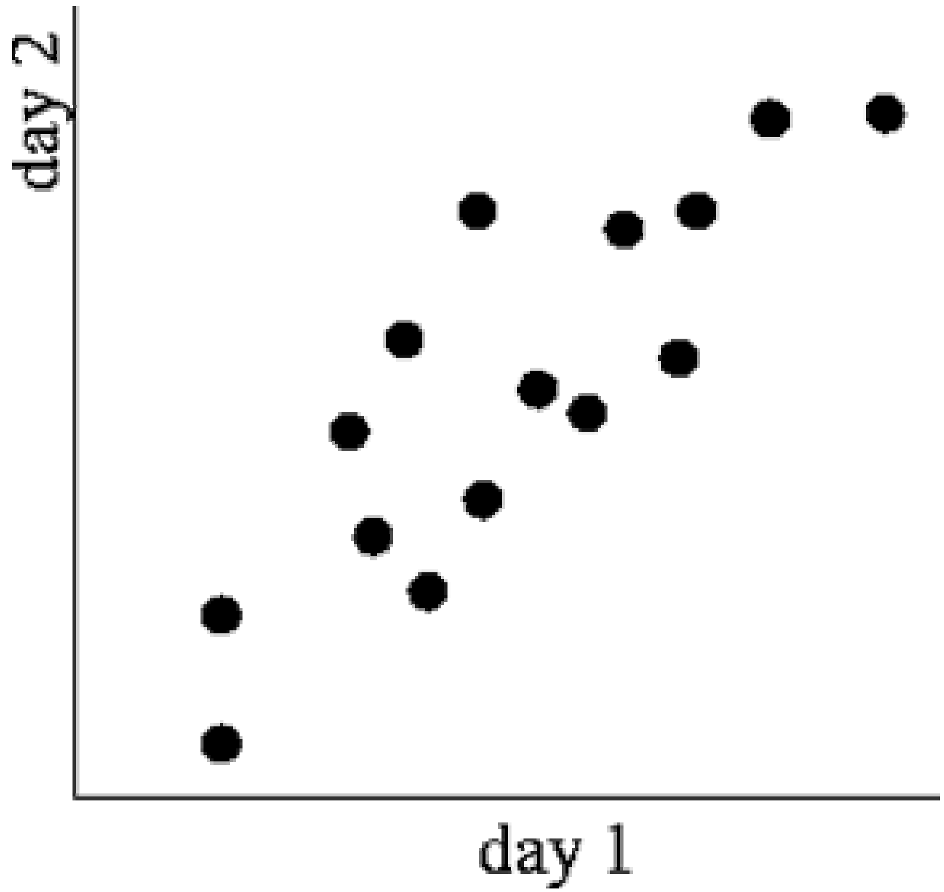
details

SVD - Interpretation #2

- best axis to project on: ('best' = min sum of squares of projection errors)

details

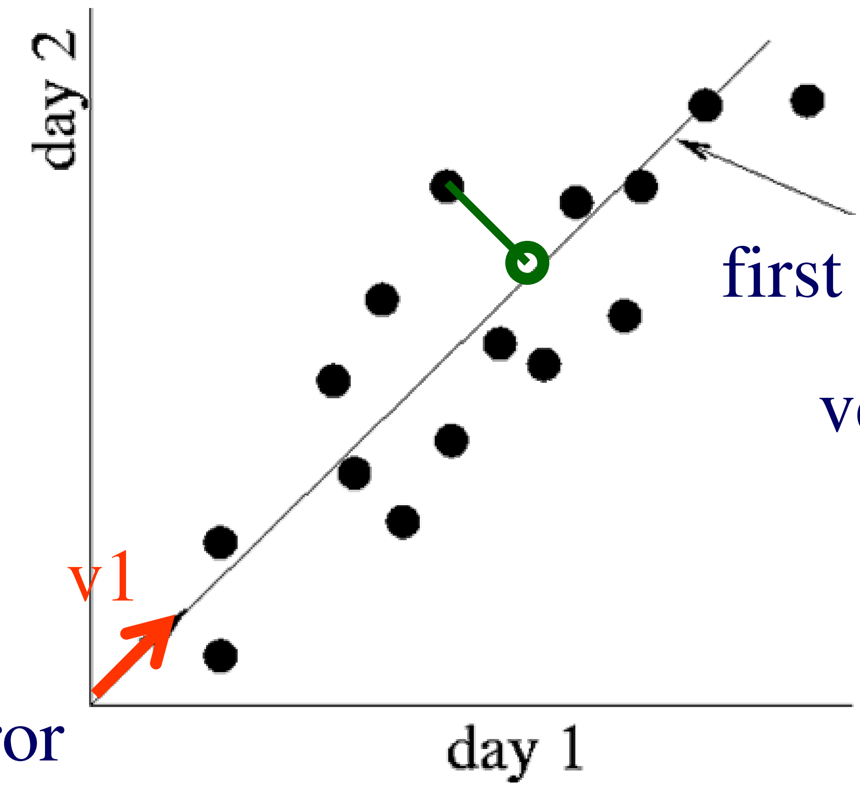
SVD - Motivation



details

SVD - interpretation #2

SVD: gives
best axis to project



- minimum RMS error

details

SVD - Interpretation #2

day / customer	We 7/10/96	Th 7/11/96	Fr 7/12/96	Sa 7/13/96	Su 7/14/96
ABC Inc.	1	1	1	0	0
DEF Ltd.	2	2	2	0	0
GHI Inc.	1	1	1	0	0
KLM Co.	5	5	5	0	0
Smith	0	0	0	2	2
Johnson	0	0	0	3	3
Thompson	0	0	0	1	1

details

SVD - Interpretation #2

- $\mathbf{A} = \mathbf{U}\ \mathbf{\Lambda}\ \mathbf{V}^T$ - example:

$$\begin{bmatrix} 1 & 1 & 1 & 0 & 0 \\ 2 & 2 & 2 & 0 & 0 \\ 1 & 1 & 1 & 0 & 0 \\ 5 & 5 & 5 & 0 & 0 \\ 0 & 0 & 0 & 2 & 2 \\ 0 & 0 & 0 & 3 & 3 \\ 0 & 0 & 0 & 1 & 1 \end{bmatrix} = \begin{bmatrix} 0.18 & 0 \\ 0.36 & 0 \\ 0.18 & 0 \\ 0.90 & 0 \\ 0 & 0.53 \\ 0 & 0.80 \\ 0 & 0.27 \end{bmatrix} \times \begin{bmatrix} 9.64 & 0 \\ 0 & 5.29 \end{bmatrix} \times$$

$$\begin{array}{l} \text{v1} \\ \begin{bmatrix} \boxed{0.58 \quad 0.58 \quad 0.58 \quad 0 \quad 0} \\ 0 \quad 0 \quad 0 \quad 0.71 \quad 0.71 \end{bmatrix} \end{array}$$

details

SVD - Interpretation #2

- $\mathbf{A} = \mathbf{U}\, \mathbf{\Lambda}\, \mathbf{V}^T$ - example:

variance ('spread') on the v1 axis

$$\begin{bmatrix} 1 & 1 & 1 & 0 & 0 \\ 2 & 2 & 2 & 0 & 0 \\ 1 & 1 & 1 & 0 & 0 \\ 5 & 5 & 5 & 0 & 0 \\ 0 & 0 & 0 & 2 & 2 \\ 0 & 0 & 0 & 3 & 3 \\ 0 & 0 & 0 & 1 & 1 \end{bmatrix} = \begin{bmatrix} 0.18 & 0 \\ 0.36 & 0 \\ 0.18 & 0 \\ 0.90 & 0 \\ 0 & 0.53 \\ 0 & 0.80 \\ 0 & 0.27 \end{bmatrix} \times \begin{bmatrix} 9.64 & 0 \\ 0 & 5.29 \end{bmatrix} \times \begin{bmatrix} 0.58 & 0.58 & 0.58 & 0 & 0 \\ 0 & 0 & 0 & 0.71 & 0.71 \end{bmatrix}$$

details

SVD - Interpretation #2

- $\mathbf{A} = \mathbf{U}\ \mathbf{\Lambda}\ \mathbf{V}^T$ - example:
  - $\mathbf{U}\ \mathbf{\Lambda}$ gives the coordinates of the points in the projection axis

$$
\begin{bmatrix}
1 & 1 & 1 & 0 & 0 \\
2 & 2 & 2 & 0 & 0 \\
1 & 1 & 1 & 0 & 0 \\
5 & 5 & 5 & 0 & 0 \\
0 & 0 & 0 & 2 & 2 \\
0 & 0 & 0 & 3 & 3 \\
0 & 0 & 0 & 1 & 1
\end{bmatrix}
=
\begin{bmatrix}
0.18 & 0 \\
0.36 & 0 \\
0.18 & 0 \\
0.90 & 0 \\
0 & 0.53 \\
0 & 0.80 \\
0 & 0.27
\end{bmatrix}
\times
\begin{bmatrix}
9.64 & 0 \\
0 & 5.29
\end{bmatrix}
\times
\begin{bmatrix}
0.58 & 0.58 & 0.58 & 0 & 0 \\
0 & 0 & 0 & 0.71 & 0.71
\end{bmatrix}
$$

details

SVD - Interpretation #2

- More details
- Q: how exactly is dim. reduction done?

$$\begin{bmatrix} 1 & 1 & 1 & 0 & 0 \\ 2 & 2 & 2 & 0 & 0 \\ 1 & 1 & 1 & 0 & 0 \\ 5 & 5 & 5 & 0 & 0 \\ 0 & 0 & 0 & 2 & 2 \\ 0 & 0 & 0 & 3 & 3 \\ 0 & 0 & 0 & 1 & 1 \end{bmatrix} = \begin{bmatrix} 0.18 & 0 \\ 0.36 & 0 \\ 0.18 & 0 \\ 0.90 & 0 \\ 0 & 0.53 \\ 0 & 0.80 \\ 0 & 0.27 \end{bmatrix} \times \begin{bmatrix} 9.64 & 0 \\ 0 & 5.29 \end{bmatrix} \times \begin{bmatrix} 0.58 & 0.58 & 0.58 & 0 & 0 \\ 0 & 0 & 0 & 0.71 & 0.71 \end{bmatrix}$$

details

SVD - Interpretation #2

- More details
- Q: how exactly is dim. reduction done?
- A: set the smallest singular values to zero:

$$\begin{bmatrix} 1 & 1 & 1 & 0 & 0 \\ 2 & 2 & 2 & 0 & 0 \\ 1 & 1 & 1 & 0 & 0 \\ 5 & 5 & 5 & 0 & 0 \\ 0 & 0 & 0 & 2 & 2 \\ 0 & 0 & 0 & 3 & 3 \\ 0 & 0 & 0 & 1 & 1 \end{bmatrix} = \begin{bmatrix} 0.18 & 0 \\ 0.36 & 0 \\ 0.18 & 0 \\ 0.90 & 0 \\ 0 & 0.53 \\ 0 & 0.80 \\ 0 & 0.27 \end{bmatrix} \times \begin{bmatrix} 9.64 & 0 \\ 0 & \cancel{5.29} \end{bmatrix} \times \begin{bmatrix} 0.58 & 0.58 & 0.58 & 0 & 0 \\ 0 & 0 & 0 & 0.71 & 0.71 \end{bmatrix}$$

details

SVD - Interpretation #2

$$
\begin{bmatrix} 1 & 1 & 1 & 0 & 0 \\ 2 & 2 & 2 & 0 & 0 \\ 1 & 1 & 1 & 0 & 0 \\ 5 & 5 & 5 & 0 & 0 \\ 0 & 0 & 0 & 2 & 2 \\ 0 & 0 & 0 & 3 & 3 \\ 0 & 0 & 0 & 1 & 1 \end{bmatrix}
\sim
\begin{bmatrix} 0.18 & 0 \\ 0.36 & 0 \\ 0.18 & 0 \\ 0.90 & 0 \\ 0 & 0.53 \\ 0 & 0.80 \\ 0 & 0.27 \end{bmatrix}
\text{x}
\begin{bmatrix} 9.64 & 0 \\ 0 & \cancel{0} \end{bmatrix}
\text{x}
\begin{bmatrix} 0.58 & 0.58 & 0.58 & 0 & 0 \\ 0 & 0 & 0 & 0.71 & 0.71 \end{bmatrix}
$$

details

SVD - Interpretation #2

$$\begin{bmatrix} 1 & 1 & 1 & 0 & 0 \\ 2 & 2 & 2 & 0 & 0 \\ 1 & 1 & 1 & 0 & 0 \\ 5 & 5 & 5 & 0 & 0 \\ 0 & 0 & 0 & 2 & 2 \\ 0 & 0 & 0 & 3 & 3 \\ 0 & 0 & 0 & 1 & 1 \end{bmatrix} \sim \begin{bmatrix} 0.18 & \cancel{0} \\ 0.36 & \cancel{0} \\ 0.18 & \cancel{0} \\ 0.90 & \cancel{0} \\ 0 & \cancel{0.53} \\ 0 & \cancel{0.80} \\ 0 & \cancel{0.27} \end{bmatrix} \text{ x } \begin{bmatrix} 9.64 & 0 \\ 0 & \cancel{0} \end{bmatrix} \text{ x } \begin{bmatrix} 0.58 & 0.58 & 0.58 & 0 & 0 \\ \cancel{0} & \cancel{0} & \cancel{0} & \cancel{0.71} & \cancel{0.71} \end{bmatrix}$$

details

SVD - Interpretation #2

$$
\begin{bmatrix} 1 & 1 & 1 & 0 & 0 \\ 2 & 2 & 2 & 0 & 0 \\ 1 & 1 & 1 & 0 & 0 \\ 5 & 5 & 5 & 0 & 0 \\ 0 & 0 & 0 & 2 & 2 \\ 0 & 0 & 0 & 3 & 3 \\ 0 & 0 & 0 & 1 & 1 \end{bmatrix} \sim \begin{bmatrix} 0.18 \\ 0.36 \\ 0.18 \\ 0.90 \\ 0 \\ 0 \\ 0 \end{bmatrix} \times \begin{bmatrix} 9.64 \end{bmatrix} \times \begin{bmatrix} 0.58 & 0.58 & 0.58 & 0 & 0 \end{bmatrix}
$$

details

SVD - Interpretation #2

$$\begin{bmatrix} 1 & 1 & 1 & 0 & 0 \\ 2 & 2 & 2 & 0 & 0 \\ 1 & 1 & 1 & 0 & 0 \\ 5 & 5 & 5 & 0 & 0 \\ 0 & 0 & 0 & 2 & 2 \\ 0 & 0 & 0 & 3 & 3 \\ 0 & 0 & 0 & 1 & 1 \end{bmatrix} \sim \begin{bmatrix} 1 & 1 & 1 & 0 & 0 \\ 2 & 2 & 2 & 0 & 0 \\ 1 & 1 & 1 & 0 & 0 \\ 5 & 5 & 5 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

details

SVD - Interpretation #2

Exactly equivalent:

'spectral decomposition' of the matrix:

$$\begin{bmatrix} 1 & 1 & 1 & 0 & 0 \\ 2 & 2 & 2 & 0 & 0 \\ 1 & 1 & 1 & 0 & 0 \\ 5 & 5 & 5 & 0 & 0 \\ 0 & 0 & 0 & 2 & 2 \\ 0 & 0 & 0 & 3 & 3 \\ 0 & 0 & 0 & 1 & 1 \end{bmatrix} = \begin{bmatrix} 0.18 & 0 \\ 0.36 & 0 \\ 0.18 & 0 \\ 0.90 & 0 \\ 0 & 0.53 \\ 0 & 0.80 \\ 0 & 0.27 \end{bmatrix} \text{x} \begin{bmatrix} 9.64 & 0 \\ 0 & 5.29 \end{bmatrix} \text{x} \begin{bmatrix} 0.58 & 0.58 & 0.58 & 0 & 0 \\ 0 & 0 & 0 & 0.71 & 0.71 \end{bmatrix}$$

details

SVD - Interpretation #2

Exactly equivalent:

'spectral decomposition' of the matrix:

$$\begin{bmatrix} 1 & 1 & 1 & 0 & 0 \\ 2 & 2 & 2 & 0 & 0 \\ 1 & 1 & 1 & 0 & 0 \\ 5 & 5 & 5 & 0 & 0 \\ 0 & 0 & 0 & 2 & 2 \\ 0 & 0 & 0 & 3 & 3 \\ 0 & 0 & 0 & 1 & 1 \end{bmatrix} = \begin{bmatrix} | & | \\ u_1 & u_2 \\ | & | \end{bmatrix} \text{ x } \begin{bmatrix} \lambda_1 & \emptyset \\ \emptyset & \lambda_2 \end{bmatrix} \text{ x } \begin{bmatrix} \text{---} & v_1 & \text{---} \\ \text{---} & v_2 & \text{---} \end{bmatrix}$$

details

SVD - Interpretation #2

Exactly equivalent:

'spectral decomposition' of the matrix:

$$\begin{array}{cc} & \longleftarrow\ m\ \longrightarrow \\ n & \begin{bmatrix} 1 & 1 & 1 & 0 & 0 \\ 2 & 2 & 2 & 0 & 0 \\ 1 & 1 & 1 & 0 & 0 \\ 5 & 5 & 5 & 0 & 0 \\ 0 & 0 & 0 & 2 & 2 \\ 0 & 0 & 0 & 3 & 3 \\ 0 & 0 & 0 & 1 & 1 \end{bmatrix} \end{array} = \lambda_1 \quad u_1 \quad v^T{}_1 + \quad \lambda_2 \quad u_2 \quad v^T{}_2 + \ldots$$

details

SVD - Interpretation #2

Exactly equivalent:

'spectral decomposition' of the matrix:

$$\begin{array}{cc} & m \\ n & \begin{bmatrix} 1 & 1 & 1 & 0 & 0 \\ 2 & 2 & 2 & 0 & 0 \\ 1 & 1 & 1 & 0 & 0 \\ 5 & 5 & 5 & 0 & 0 \\ 0 & 0 & 0 & 2 & 2 \\ 0 & 0 & 0 & 3 & 3 \\ 0 & 0 & 0 & 1 & 1 \end{bmatrix} \end{array} = \overbrace{\lambda_1 \; \underbrace{u_1}_{n \times 1} \; \underbrace{v^T_1}_{1 \times m} + \; \lambda_2 \; u_2 \; v^T_2 + \ldots}^{r \text{ terms}}$$

details

SVD - Interpretation #2

approximation / dim. reduction:
by keeping the first few terms (Q: how many?)

$$\begin{array}{cc} & \longleftarrow \; m \; \longrightarrow \\ n & \begin{bmatrix} 1 & 1 & 1 & 0 & 0 \\ 2 & 2 & 2 & 0 & 0 \\ 1 & 1 & 1 & 0 & 0 \\ 5 & 5 & 5 & 0 & 0 \\ 0 & 0 & 0 & 2 & 2 \\ 0 & 0 & 0 & 3 & 3 \\ 0 & 0 & 0 & 1 & 1 \end{bmatrix} \end{array} = \lambda_1 \; u_1 \; v^T_1 + \lambda_2 \; u_2 \; v^T_2 + \ldots$$

assume: $\lambda_1 >= \lambda_2 >= \ldots$

details

SVD - Interpretation #2

A (heuristic - [Fukunaga]): keep 80-90% of 'energy' (= sum of squares of λ_i 's)

$$\begin{array}{cc} & \longleftarrow m \longrightarrow \\ n & \begin{bmatrix} 1 & 1 & 1 & 0 & 0 \\ 2 & 2 & 2 & 0 & 0 \\ 1 & 1 & 1 & 0 & 0 \\ 5 & 5 & 5 & 0 & 0 \\ 0 & 0 & 0 & 2 & 2 \\ 0 & 0 & 0 & 3 & 3 \\ 0 & 0 & 0 & 1 & 1 \end{bmatrix} \end{array} = \lambda_1 \; u_1 \; v^T{}_1 + \lambda_2 \; u_2 \; v^T{}_2 + \ldots$$

assume: $\lambda_1 >= \lambda_2 >= \ldots$

SVD - Detailed outline

- Motivation
- Definition - properties
- Interpretation
  - #1: documents/terms/concepts
  - #2: dim. reduction
  - #3: picking non-zero, rectangular 'blobs'
- Complexity
- Case studies
- Additional properties

SVD - Interpretation #3

- finds non-zero 'blobs' in a data matrix

$$\begin{bmatrix} 1 & 1 & 1 & 0 & 0 \\ 2 & 2 & 2 & 0 & 0 \\ 1 & 1 & 1 & 0 & 0 \\ 5 & 5 & 5 & 0 & 0 \\ 0 & 0 & 0 & 2 & 2 \\ 0 & 0 & 0 & 3 & 3 \\ 0 & 0 & 0 & 1 & 1 \end{bmatrix} = \begin{bmatrix} 0.18 & 0 \\ 0.36 & 0 \\ 0.18 & 0 \\ 0.90 & 0 \\ 0 & 0.53 \\ 0 & 0.80 \\ 0 & 0.27 \end{bmatrix} \times \begin{bmatrix} 9.64 & 0 \\ 0 & 5.29 \end{bmatrix} \times \begin{bmatrix} 0.58 & 0.58 & 0.58 & 0 & 0 \\ 0 & 0 & 0 & 0.71 & 0.71 \end{bmatrix}$$

SVD - Interpretation #3

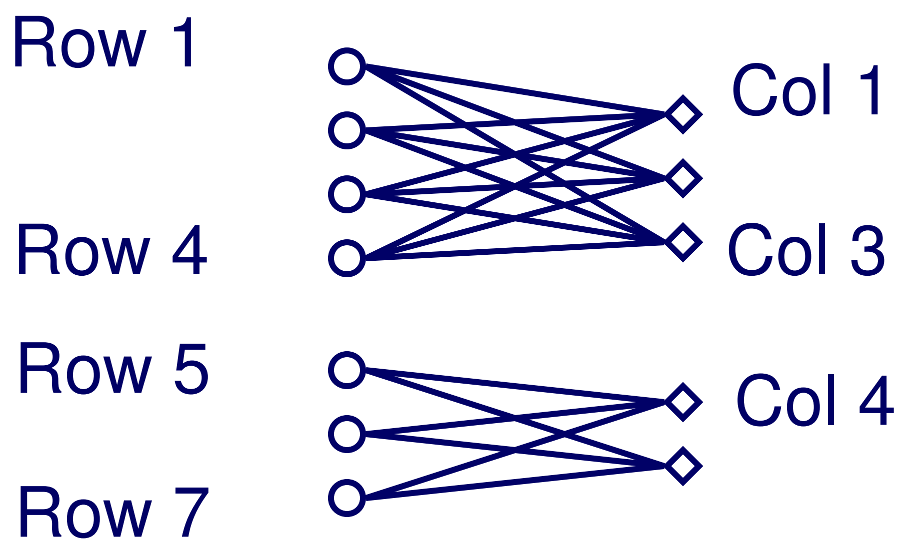
- finds non-zero 'blobs' in a data matrix

$$\left[\begin{array}{ccc|cc} 1 & 1 & 1 & 0 & 0 \\ 2 & 2 & 2 & 0 & 0 \\ 1 & 1 & 1 & 0 & 0 \\ 5 & 5 & 5 & 0 & 0 \\ \hline 0 & 0 & 0 & 2 & 2 \\ 0 & 0 & 0 & 3 & 3 \\ 0 & 0 & 0 & 1 & 1 \end{array}\right] = \left[\begin{array}{cc} 0.18 & 0 \\ 0.36 & 0 \\ 0.18 & 0 \\ 0.90 & 0 \\ 0 & 0.53 \\ 0 & 0.80 \\ 0 & 0.27 \end{array}\right] \times \left[\begin{array}{cc} 9.64 & 0 \\ 0 & 5.29 \end{array}\right] \times \left[\begin{array}{ccccc} 0.58 & 0.58 & 0.58 & 0 & 0 \\ 0 & 0 & 0 & 0.71 & 0.71 \end{array}\right]$$

SVD - Interpretation #3

- finds non-zero ‘blobs’ in a data matrix =
- ‘communities’ (bi-partite cores, here)

$$
\left[\begin{array}{ccc|cc}
1 & 1 & 1 & 0 & 0 \\
2 & 2 & 2 & 0 & 0 \\
1 & 1 & 1 & 0 & 0 \\
5 & 5 & 5 & 0 & 0 \\
\hline
0 & 0 & 0 & 2 & 2 \\
0 & 0 & 0 & 3 & 3 \\
0 & 0 & 0 & 1 & 1
\end{array}\right]
$$



SVD - Detailed outline

- Motivation
- Definition - properties
- Interpretation
- Complexity
- Case studies
- Additional properties

SVD - Complexity

- O(n * m * m) or O(n * n * m) (whichever is less)
- less work, if we just want singular values
- or if we want first *k* singular vectors
- or if the matrix is sparse [Berry]
- Implemented: in any linear algebra package (LINPACK, matlab, Splus, mathematica ...)

SVD - conclusions so far

- SVD: $\mathbf{A} = \mathbf{U}\, \boldsymbol{\Lambda}\, \mathbf{V}^{\mathbf{T}}$: unique (*)
- **U**: document-to-concept similarities
- **V**: term-to-concept similarities
- **Λ**: strength of each concept
- dim. reduction: keep the first few strongest singular values (80-90% of 'energy')
  - SVD: picks up linear correlations
- SVD: picks up non-zero 'blobs'

details

SVD - Detailed outline

- Motivation
- Definition - properties
- Interpretation
- Complexity
- SVD properties
- Case studies
- Conclusions

details

SVD - Other properties - summary

- can produce orthogonal basis (obvious) (who cares?)
- can solve over- and under-determined linear problems (see C(1) property)
- can compute 'fixed points' (= 'steady state prob. in Markov chains') (see C(4) property)

details

SVD -outline of properties

- (A): obvious
- (B): less obvious
- (C): least obvious (and most powerful!)

details

Properties - by defn.:

A(0): $\mathbf{A}_{[n \times m]} = \mathbf{U}_{[n \times r]} \, \mathbf{\Lambda}_{[r \times r]} \, \mathbf{V}^{\mathbf{T}}_{[r \times m]}$

A(1): $\mathbf{U}^{\mathrm{T}}_{[r \times n]} \, \mathbf{U}_{[n \times r]} = \mathbf{I}_{[r \times r]}$ (identity matrix)

A(2): $\mathbf{V}^{\mathrm{T}}_{[r \times n]} \, \mathbf{V}_{[n \times r]} = \mathbf{I}_{[r \times r]}$

A(3): $\Lambda^k = \mathrm{diag}(\lambda_1^k, \lambda_2^k, \ldots \lambda_r^k)$ (k: ANY real number)

A(4): $\mathbf{A}^{\mathbf{T}} = \mathbf{V} \, \mathbf{\Lambda} \, \mathbf{U}^{\mathbf{T}}$

details

Less obvious properties

A(0): $\mathbf{A}_{[n \times m]} = \mathbf{U}_{[n \times r]} \, \boldsymbol{\Lambda}_{[r \times r]} \, \mathbf{V}^{\mathbf{T}}_{[r \times m]}$

B(1): $\mathbf{A}_{[n \times m]} \, (\mathbf{A}^{T})_{[m \times n]} = ??$

details

Less obvious properties

A(0): $\mathbf{A}_{[n \times m]} = \mathbf{U}_{[n \times r]} \mathbf{\Lambda}_{[r \times r]} \mathbf{V}^{\mathbf{T}}_{[r \times m]}$

B(1): $\mathbf{A}_{[n \times m]} (\mathbf{A}^{T})_{[m \times n]} = \mathbf{U} \mathbf{\Lambda}^{2} \mathbf{U}^{T}$

symmetric; Intuition?

details

Less obvious properties

A(0): $\mathbf{A}_{[n \times m]} = \mathbf{U}_{[n \times r]} \mathbf{\Lambda}_{[r \times r]} \mathbf{V}^{\mathbf{T}}_{[r \times m]}$

B(1): $\mathbf{A}_{[n \times m]} (\mathbf{A}^{T})_{[m \times n]} = \mathbf{U} \mathbf{\Lambda}^{2} \mathbf{U}^{T}$

symmetric; Intuition?

'document-to-document' similarity matrix

B(2): symmetrically, for 'V'

(AT) [m x n] A [n x m] = V L2 VT

Intuition?

details

Less obvious properties

A: term-to-term similarity matrix

B(3): $((\mathbf{A}^T)_{[m \times n]} \mathbf{A}_{[n \times m]})^k = \mathbf{V} \, \mathbf{\Lambda}^{2k} \, \mathbf{V}^T$

and

B(4): $(\mathbf{A}^T \mathbf{A})^k \sim \mathbf{v}_1 \, \lambda_1^{2k} \, \mathbf{v}_1^T$ for k>>1

where

$\mathbf{v}_1$: [m x 1] first column (singular-vector) of **V**

λ_1: strongest singular value

details

Less obvious properties

B(4): $(\mathbf{A}^T \mathbf{A})^k \sim \mathbf{v}_1 \lambda_1^{2k} \mathbf{v}_1^T$ for k>>1

$\boxed{\text{B(5): } (\mathbf{A}^T \mathbf{A})^k \mathbf{v'} \sim (\text{constant})\, \mathbf{v}_1}$

ie., for (almost) any **v'**, it converges to a vector parallel to $\mathbf{v}_1$

Thus, useful to compute first singular vector/value (as well as the next ones, too...)

Less obvious properties - repeated:

details

A(0): $\mathbf{A}_{[n \times m]} = \mathbf{U}_{[n \times r]} \, \mathbf{\Lambda}_{[r \times r]} \, \mathbf{V}^{\mathbf{T}}_{[r \times m]}$

B(1): $\mathbf{A}_{[n \times m]} (\mathbf{A}^{T})_{[m \times n]} = \mathbf{U} \, \mathbf{\Lambda}^{2} \, \mathbf{U}^{T}$

B(2): $(\mathbf{A}^{T})_{[m \times n]} \mathbf{A}_{[n \times m]} = \mathbf{V} \, \mathbf{\Lambda}^{2} \, \mathbf{V}^{T}$

B(3): $((\mathbf{A}^{T})_{[m \times n]} \mathbf{A}_{[n \times m]})^{k} = \mathbf{V} \, \mathbf{\Lambda}^{2k} \, \mathbf{V}^{T}$

B(4): $(\mathbf{A}^{T} \mathbf{A})^{k} \sim v_1 \, \lambda_1^{2k} \, v_1^{T}$

B(5): $(\mathbf{A}^{T} \mathbf{A})^{k} \, \mathbf{v'} \sim$ (constant) $\mathbf{v}_1$

details

Least obvious properties - cont'd

A(0): $\mathbf{A}_{[n \times m]} = \mathbf{U}_{[\, n \times r \,]} \, \mathbf{\Lambda}_{[\, r \times r \,]} \, \mathbf{V}^{\mathbf{T}}_{[\, r \times m]}$

C(2): $\mathbf{A}_{[n \times m]} \, \mathbf{v_1}_{[m \times 1]} = \lambda_1 \, \mathbf{u_1}_{[n \times 1]}$

where $\mathbf{v_1}$, $\mathbf{u_1}$ the first (column) vectors of **V**, **U.** ($\mathbf{v_1}$ == right-singular-vector)

C(3): symmetrically: $\mathbf{u_1}^{\mathbf{T}} \mathbf{A} = \lambda_1 \mathbf{v_1}^{\mathbf{T}}$

$\mathbf{u_1}$ == left-singular-vector

Therefore:

details

Least obvious properties - cont'd

A(0): $\mathbf{A}_{[n \times m]} = \mathbf{U}_{[n \times r]} \mathbf{\Lambda}_{[r \times r]} \mathbf{V}^{\mathbf{T}}_{[r \times m]}$

C(4): $\mathbf{A}^{\mathbf{T}} \mathbf{A} \mathbf{v}_1 = \lambda_1^2 \mathbf{v}_1$

(**fixed point** - the dfn of eigenvector for a symmetric matrix)

Least obvious properties - altogether

details

A(0): $\mathbf{A}_{[n \times m]} = \mathbf{U}_{[n \times r]}\, \mathbf{\Lambda}_{[r \times r]}\, \mathbf{V}^{\mathbf{T}}_{[r \times m]}$

C(1): $\mathbf{A}_{[n \times m]}\, \mathbf{x}_{[m \times 1]} = \mathbf{b}_{[n \times 1]}$

then, $\mathbf{x}_0 = \mathbf{V}\, \Lambda^{(-1)}\, \mathbf{U}^T\, \mathbf{b}$**:** shortest, actual or least-squares solution

C(2): $\mathbf{A}_{[n \times m]}\, \mathbf{v}_{\mathbf{1}\,[m \times 1]} = \lambda_{\mathbf{1}}\, \mathbf{u}_{\mathbf{1}\,[n \times 1]}$

C(3): $\mathbf{u}_{\mathbf{1}}^{\mathbf{T}}\, \mathbf{A} = \lambda_{\mathbf{1}}\, \mathbf{v}_{\mathbf{1}}^{\mathbf{T}}$

C(4): $\mathbf{A}^{\mathbf{T}}\, \mathbf{A}\, \mathbf{v}_{\mathbf{1}} = \lambda_{\mathbf{1}}^{\mathbf{2}}\, \mathbf{v}_{\mathbf{1}}$

details

Properties - conclusions

A(0): $\mathbf{A}_{[n \times m]} = \mathbf{U}_{[\, n \times r \,]}\, \mathbf{\Lambda}_{[\, r \times r \,]}\, \mathbf{V}^{\mathbf{T}}_{[\, r \times m]}$

B(5): $(\mathbf{A}^{T} \mathbf{A})^{k}\, \mathbf{v'} \sim$ (constant) $\mathbf{v}_1$

C(1): $\mathbf{A}_{[n \times m]}\, \mathbf{x}_{[m \times 1]} = \mathbf{b}_{[n \times 1]}$

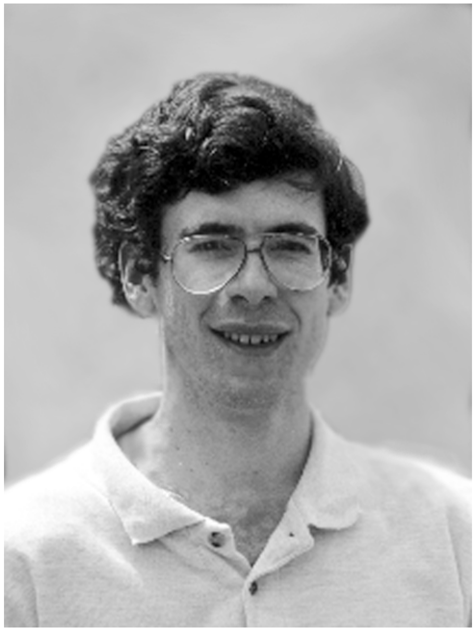
then, $\mathbf{x}_0 = \mathbf{V}\, \mathbf{\Lambda}^{(-1)}\, \mathbf{U}^{T}\, \mathbf{b}$**:** shortest, actual or least-squares solution

C(4): $\mathbf{A}^{\mathbf{T}} \mathbf{A}\, \mathbf{v}_{\mathbf{1}} = \lambda_{\mathbf{1}}^{\mathbf{2}}\, \mathbf{v}_{\mathbf{1}}$

SVD - detailed outline

- ...
- SVD properties
- case studies
  - Kleinberg's algorithm
  - Google's algorithm
- Conclusions

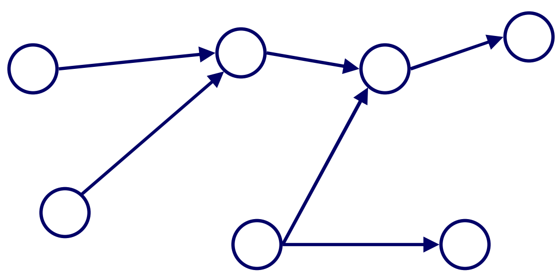
Kleinberg's algo (HITS)



Kleinberg, Jon (1998). *Authoritative sources in a hyperlinked environment*. Proc. 9th ACM-SIAM Symposium on Discrete Algorithms.

Recall: problem dfn

- Given a graph (eg., web pages containing the desirable query word)
- Q: Which node is the most important?



Kleinberg's algorithm

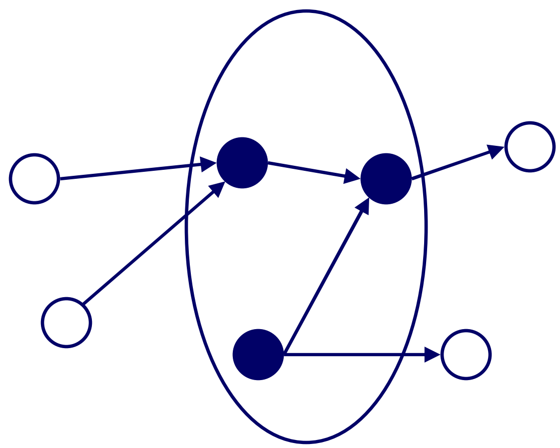
- Problem dfn: given the web and a query
- find the most 'authoritative' web pages for this query

Step 0: find all pages containing the query terms

Step 1: expand by one move forward and backward

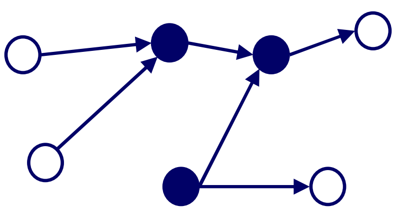
Kleinberg's algorithm

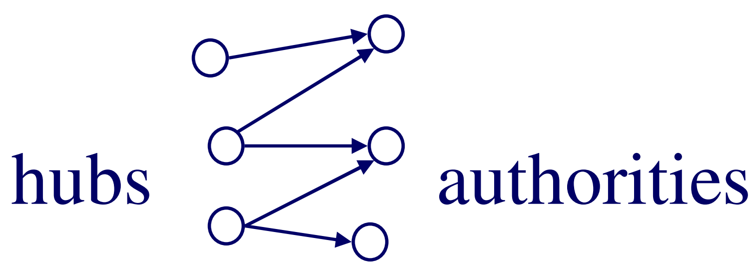
- Step 1: expand by one move forward and backward



Kleinberg's algorithm

- on the resulting graph, give high score (= 'authorities') to nodes that many important nodes point to
- give high importance score ('hubs') to nodes that point to good 'authorities')





Kleinberg's algorithm

observations

- recursive definition!
- each node (say, '*i*'-th node) has both an authoritativeness score a_i and a hubness score h_i

Kleinberg's algorithm

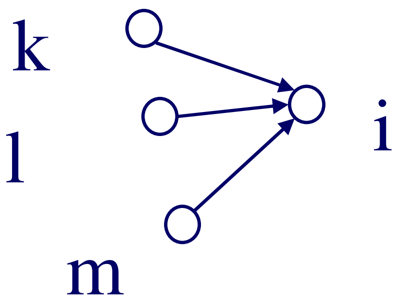
Let E be the set of edges and **A** be the adjacency matrix:

the (i,j) is 1 if the edge from i to j exists

Let h and a be [n x 1] vectors with the 'hubness' and 'authoritativiness' scores.

Then:

Kleinberg's algorithm



Then:

$$a_i = h_k + h_l + h_m$$

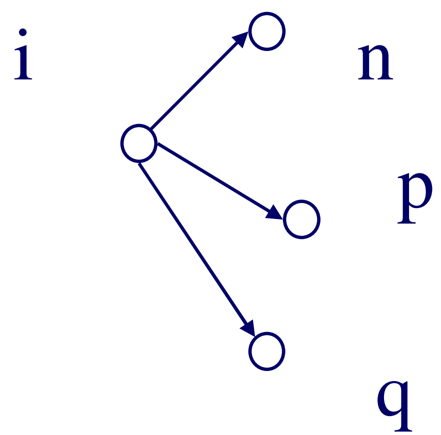
that is

a_i = Sum (h_j)   over all j that (j,i) edge exists

or

$\mathbf{a} = \mathbf{A}^T \mathbf{h}$

Kleinberg's algorithm



symmetrically, for the 'hubness':

$$h_i = a_n + a_p + a_q$$

that is

$h_i = \text{Sum}\ (q_j)$ over all j that (i,j) edge exists

or

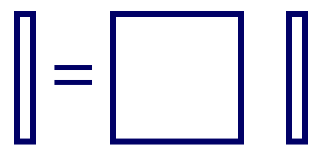
$$\mathbf{h} = \mathbf{A}\ \mathbf{a}$$

Kleinberg's algorithm

In conclusion, we want vectors **h** and **a** such that:

$$\mathbf{h} = \mathbf{A}\,\mathbf{a}$$

$$\mathbf{a} = \mathbf{A}^{T}\,\mathbf{h}$$



Recall properties:

C(2): $\mathbf{A}_{[n \times m]}\ \mathbf{v_1}_{[m \times 1]} = \lambda_1\ \mathbf{u_1}_{[n \times 1]}$

C(3): $\mathbf{u_1}^{\mathbf{T}}\,\mathbf{A} = \lambda_1\ \mathbf{v_1}^{\mathbf{T}}$

Kleinberg's algorithm

In short, the solutions to

$$\mathbf{h} = \mathbf{A}\ \mathbf{a}$$

$$\mathbf{a} = \mathbf{A}^T\ \mathbf{h}$$

are the <u>left- and right- singular-vectors</u> of the adjacency matrix **A.**

Starting from random **a'** and iterating, we'll eventually converge

(Q: to which of all the singular-vectors? why?)

Kleinberg's algorithm

(Q: to which of all the singular-vectors? why?)

A: to the ones of the strongest singular-value, because of property B(5):

$$\text{B(5): } (\mathbf{A}^T \mathbf{A})^k \mathbf{v'} \sim (\text{constant})\ \mathbf{v}_1$$

Kleinberg's algorithm - results

Eg., for the query 'java':

0.328 www.gamelan.com

0.251 java.sun.com

0.190 www.digitalfocus.com ("the java developer")

Kleinberg's algorithm - discussion

- 'authority' score can be used to find 'similar pages' (how?)

SVD - detailed outline

- ...
- Complexity
- SVD properties
- Case studies
  - Kleinberg's algorithm (HITS)
  - Google's algorithm
- Conclusions

PageRank (google)

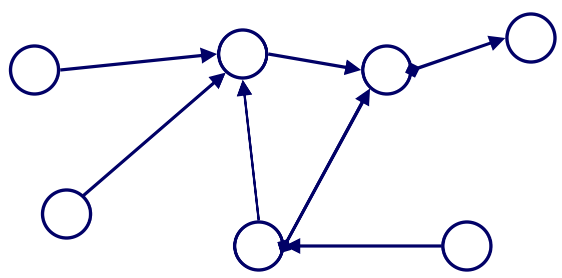
Larry Page

Sergey Brin

•Brin, Sergey and Lawrence Page (1998). *Anatomy of a Large-Scale Hypertextual Web Search Engine*. 7th Intl World Wide Web Conf.

Problem: PageRank

Given a directed graph, find its most interesting/central node

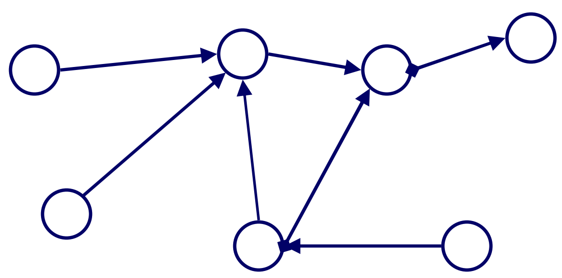


A node is important, if it is connected with important nodes (recursive, but OK!)

Problem: PageRank - solution

Given a directed graph, find its most interesting/central node

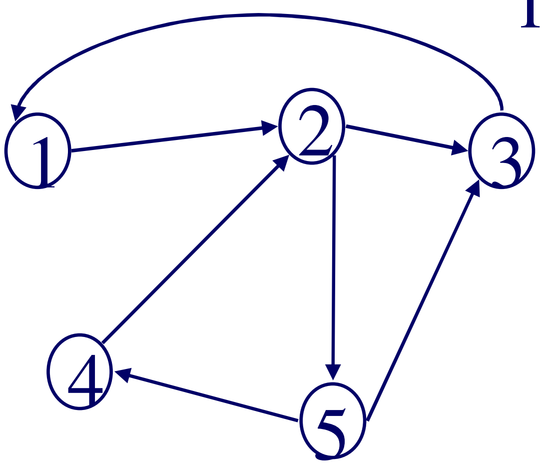
Proposed solution: Random walk; spot most 'popular' node (-> steady state prob. (ssp))



A node has high ssp, if it is connected with high ssp nodes (recursive, but OK!)

(Simplified) PageRank algorithm

- Let **A** be the adjacency matrix;
- let **B** be the transition matrix: transpose, column-normalized - then



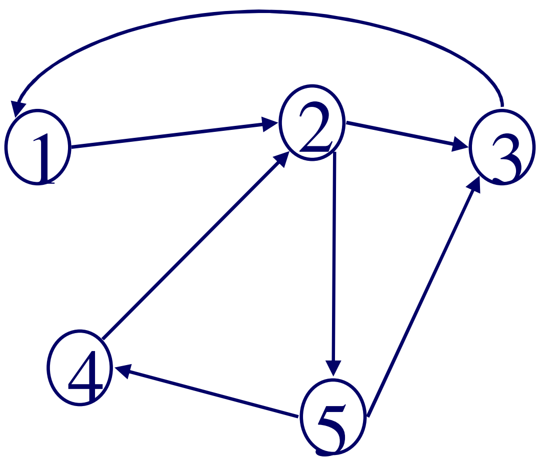
From

To

$$\overset{\mathbf{B}}{\begin{bmatrix} & & 1 & & \\ 1 & & & 1 & \\ & 1/2 & & & 1/2 \\ & & & & 1/2 \\ & 1/2 & & & \end{bmatrix}} \begin{bmatrix} p1 \\ p2 \\ p3 \\ p4 \\ p5 \end{bmatrix} = \begin{bmatrix} p1 \\ p2 \\ p3 \\ p4 \\ p5 \end{bmatrix}$$

(Simplified) PageRank algorithm

- **B p = p**



$$
\begin{array}{ccc}
\mathbf{B} & \mathbf{p} & = \quad \mathbf{p} \\
\begin{bmatrix} & & 1 & & \\ 1 & & & 1 & \\ & 1/2 & & & 1/2 \\ & & & & 1/2 \\ & 1/2 & & & \end{bmatrix} & \begin{bmatrix} p1 \\ p2 \\ p3 \\ p4 \\ p5 \end{bmatrix} & = \begin{bmatrix} p1 \\ p2 \\ p3 \\ p4 \\ p5 \end{bmatrix}
\end{array}
$$

Definitions

A Adjacency matrix (from-to)

D Degree matrix = (diag (d1, d2, …, dn))

B Transition matrix: to-from, column normalized

$$\mathbf{B} = \mathbf{A}^{T} \mathbf{D}^{-1}$$

(Simplified) PageRank algorithm

- **B p =** 1 * **p**
- thus, **p** is the **eigenvector** that corresponds to the highest eigenvalue (=1, since the matrix is column-normalized)
- Why does such a **p** exist?
  - **p** exists if **B** is nxn, nonnegative, irreducible [Perron–Frobenius theorem]

(Simplified) PageRank algorithm

- In short: imagine a particle randomly moving along the edges
- compute its steady-state probabilities (ssp)

Full version of algo: with occasional random jumps

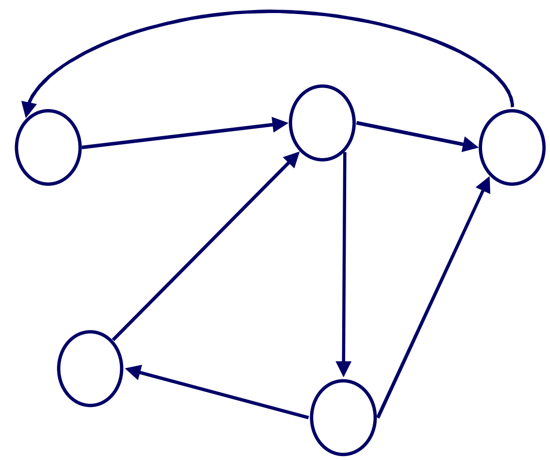
Why? To make the matrix irreducible

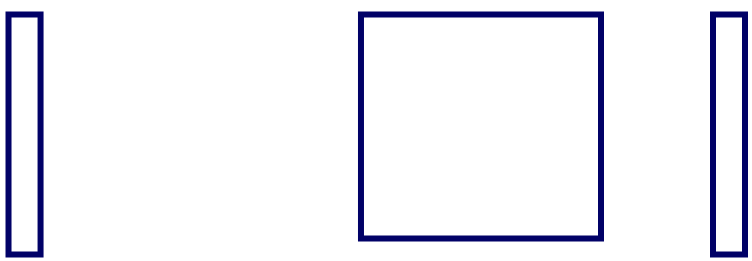
Full Algorithm

- With probability *1-c*, fly-out to a random node
- Then, we have

$$\mathbf{p} = c\,\mathbf{B}\,\mathbf{p} + (1-c)/n\ \mathbf{1} \Rightarrow$$

$$\mathbf{p} = (1-c)/n\ [\mathbf{I} - c\,\mathbf{B}]^{-1}\ \mathbf{1}$$





Alternative notation

M Modified transition matrix

$\mathbf{M} = c\,\mathbf{B} + (1\text{-}c)/n\ \mathbf{1}\ \mathbf{1}^T$

Then

$\mathbf{p} = \mathbf{M}\,\mathbf{p}$

That is: the steady state probabilities =

PageRank scores form the *first eigenvector* of the 'modified transition matrix'

Parenthesis: intuition behind eigenvectors

Formal definition

If **A** is a (n x n) square matrix
(λ , **x**) is an **eigenvalue/eigenvector** pair
of **A** if

$$\boxed{\mathbf{A}\,\mathbf{x} = \lambda\,\mathbf{x}}$$

CLOSELY related to singular values:

Property #1: Eigen- vs singular-values

if

$$\mathbf{B}_{[n \times m]} = \mathbf{U}_{[n \times r]} \, \Lambda_{[r \times r]} \, (\mathbf{V}_{[m \times r]})^T$$

then $\mathbf{A} = (\mathbf{B}^T\mathbf{B})$ is symmetric and

$$\text{C(4): } \mathbf{B}^T \mathbf{B} \, \mathbf{v_i} = \lambda_i^2 \, \mathbf{v_i}$$

ie, $\mathbf{v_1}$, $\mathbf{v_2}$, ...: eigenvectors of $\mathbf{A} = (\mathbf{B}^T\mathbf{B})$

Property #2

- If $\mathbf{A}_{[nxn]}$ is a real, symmetric matrix
- Then it has *n* real eigenvalues

(if **A** is not symmetric, some eigenvalues may be complex)

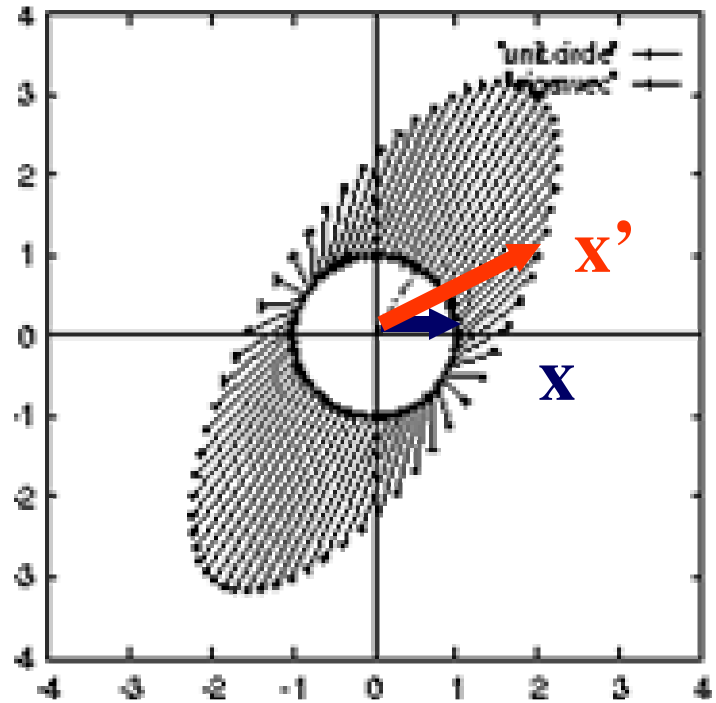
Property #3

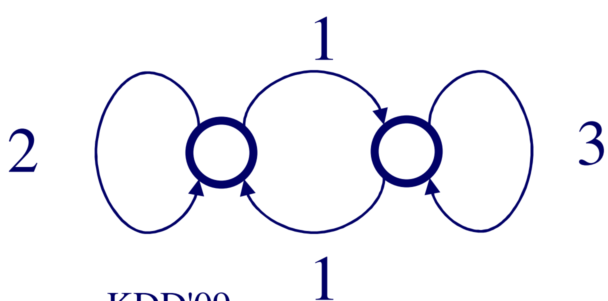
- If $\mathbf{A}_{[\mathbf{nxn}]}$ is a real, symmetric matrix
- Then it has *n* real eigenvalues
- And they agree with its *n* singular values, except possibly for the sign

Intuition

- **A** as vector transformation

$$
\begin{array}{ccccc}
\mathbf{x'} & & \mathbf{A} & & \mathbf{x} \\
\begin{bmatrix} 2 \\ 1 \end{bmatrix} & = & \begin{bmatrix} 2 & 1 \\ 1 & 3 \end{bmatrix} & & \begin{bmatrix} 1 \\ 0 \end{bmatrix}
\end{array}
$$

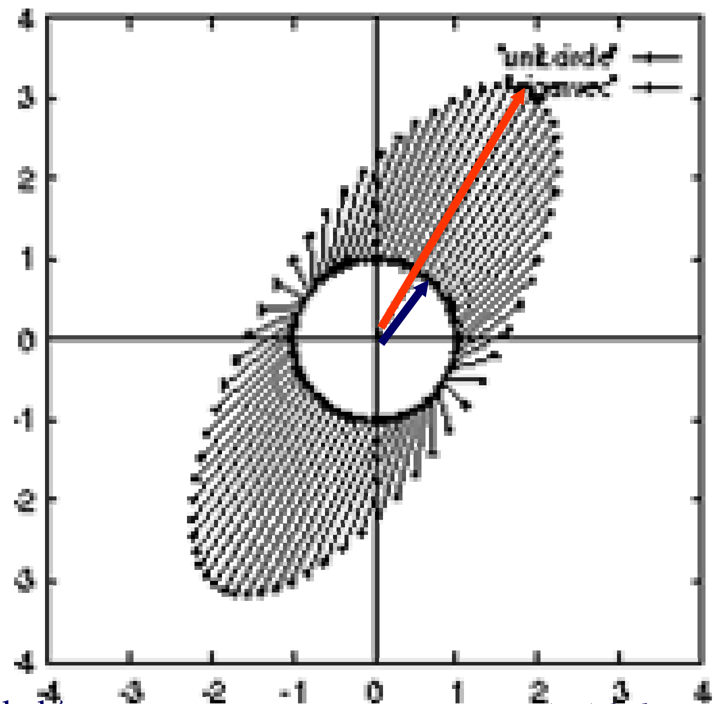




Intuition

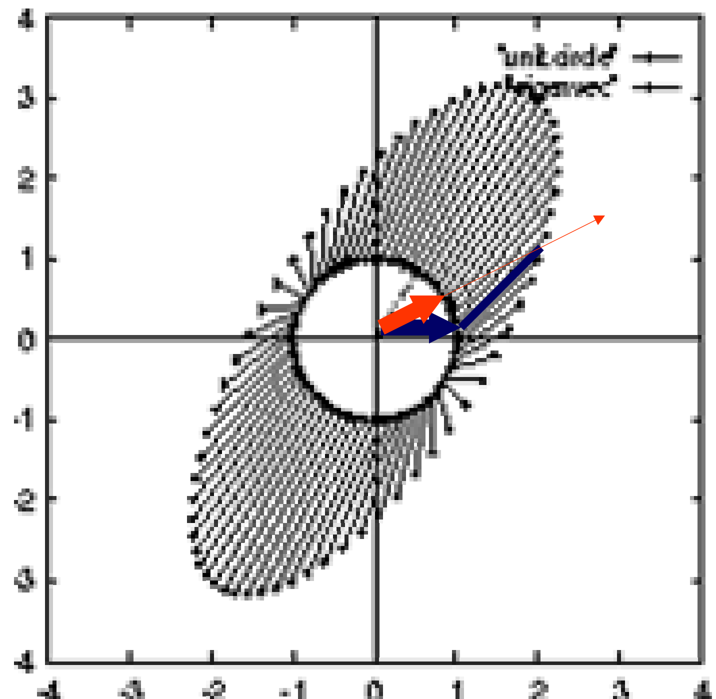
- By defn., eigenvectors remain parallel to themselves (‘**fixed points**’)

$$
\begin{array}{cccc}
\lambda_1 & \mathbf{v_1} & \mathbf{A} & \mathbf{v_1} \\
3.62 \; * & \begin{bmatrix} 0.52 \\ 0.85 \end{bmatrix} = & \begin{bmatrix} 2 & 1 \\ 1 & 3 \end{bmatrix} & \begin{bmatrix} 0.52 \\ 0.85 \end{bmatrix}
\end{array}
$$



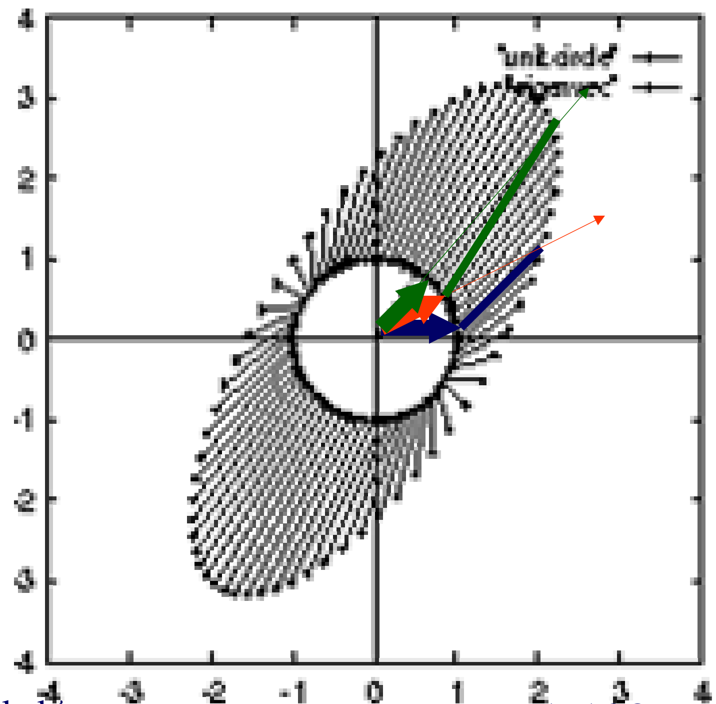
Convergence

- Usually, fast:



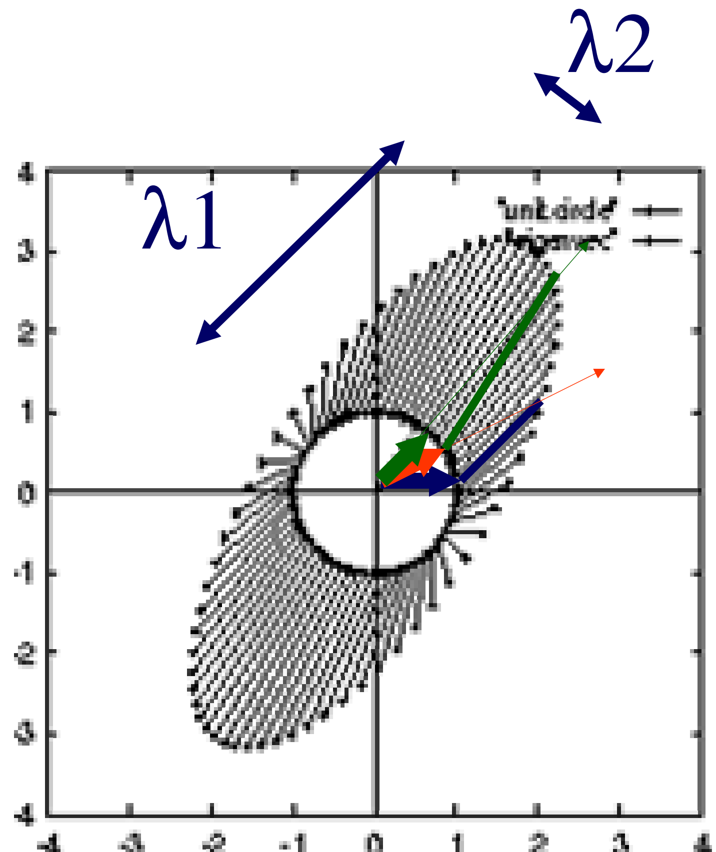
Convergence

- Usually, fast:



Convergence

- Usually, fast:
- depends on ratio
  $\lambda 1 : \lambda 2$



Kleinberg/google - conclusions

SVD helps in graph analysis:

hub/authority scores: strongest left- and right-singular-vectors of the adjacency matrix

random walk on a graph: steady state probabilities are given by the strongest eigenvector of the (modified) transition matrix

Conclusions

- SVD: a **valuable** tool
- given a document-term matrix, it finds ‘concepts’ (LSI)
- ... and can find fixed-points or steady-state probabilities (google/ Kleinberg/ Markov Chains)

Conclusions cont'd

(We didn't discuss/elaborate, but, SVD

- ... can reduce dimensionality (KL)
- ... and can find rules (PCA; RatioRules)
- ... and can solve optimally over- and under-constraint linear systems (least squares / query feedbacks)

References

- Berry, Michael: http://www.cs.utk.edu/~lsi/
- Brin, S. and L. Page (1998). *Anatomy of a Large-Scale Hypertextual Web Search Engine*. 7th Intl World Wide Web Conf.

References

- Christos Faloutsos, *Searching Multimedia Databases by Content*, Springer, 1996. (App. D)
- Fukunaga, K. (1990). *Introduction to Statistical Pattern Recognition*, Academic Press.
- I.T. Jolliffe *Principal Component Analysis* Springer, 2002 (2nd ed.)

References cont'd

- Kleinberg, J. (1998). *Authoritative sources in a hyperlinked environment*. Proc. 9th ACM-SIAM Symposium on Discrete Algorithms.
- Press, W. H., S. A. Teukolsky, et al. (1992). *Numerical Recipes in C*, Cambridge University Press. `www.nr.com`