

**Distributed**

**Co-Clustering**

**Ralf Gehrer**

**Fabian Lindenberg**

**Distributed**

**Co-Clustering**

# Co-Clustering

Simultane Gruppierung  
der Zeilen und Spalten  
einer Matrix

# Warenkorb



# Szenario

Filmdaten	
Deutscher Titel:	<b>L.A. Crash</b>
Originaltitel:	<b>Crash</b>
Erscheinungsjahr:	2004
Länge:	ca. 110 Minuten
Originalsprache:	Englisch
Stab	
Regie:	Paul Haggis
Drehbuch:	Paul Haggis Robert Moresco
Produktion:	Don Cheadle Paul Haggis Mark R. Harris Robert Moresco Cathy Schulman Bob Yari

# Szenario

	Autor	Vorlage	Produktion	Verlag	Darsteller	Regie
Der Gefangene	1	0	0	1	0	0
L.A. Crash	0	0	1	0	1	1
Sakrileg	1	0	0	1	0	0
Minority Report	0	1	1	0	1	1

# Szenario

	Autor	Verlag	Regie	Produktion	Darsteller	Vorlage
Der Gefangene	1	1	0	0	0	0
Sakrileg	1	1	0	0	0	0
L.A. Crash	0	0	1	1	1	0
Minority Report	0	0	1	1	1	1

# Szenario

	Autor	Verlag	Regie	Produktion	Darsteller	Vorlage
Der Gefangene	1	1	0	0	0	0
Sakrileg	1	1	0	0	0	0
L.A. Crash	0	0	1	1	1	0
Minority Report	0	0	1	1	1	1



# Algorithmus

**repeat**

**for each** row  $i = 1..m$  **do**

**for each** row group label =  $1..k$  **do**

**if** error is smaller

$r(i) = p$

reorganize rows according to new labels

**for each** column  $j = 1..n$  **do**

...

reorganize columns

**until** error does not decrease

# Fehlerfunktion

Minimal loss in mutual information  
oder  
Minimum code length

# Algorithmus

**repeat**

**for each** row  $i = 1..m$  **do**

**M** **for each** row group label =  $1..k$  **do**  
**if** error is smaller  
     $r(i) = p$

**R** reorganize rows according to new labels

**M** **for each** column  $j = 1..n$  **do**  
    ...

**R** reorganize columns

**until** error does not decrease

Bestimmung des  
**bestmöglichen**  
Gruppenindexes

# Map

1	0	0	1	0	0
0	0	1	0	1	1
1	0	0	1	0	0
0	1	1	0	1	1

# Map

1	1	0	0	1	0	0
2	0	0	1	0	1	1
1	1	0	0	1	0	0
2	0	1	1	0	1	1

# Reduce

Umordnung der  
Zeilen bzw. Spalten

# Reduce

1	1	0	0	1	0	0
2	0	0	1	0	1	1
1	1	0	0	1	0	0
2	0	1	1	0	1	1



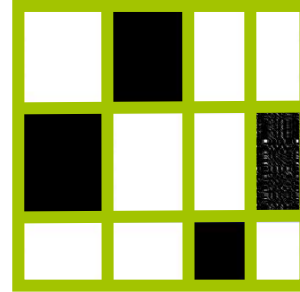
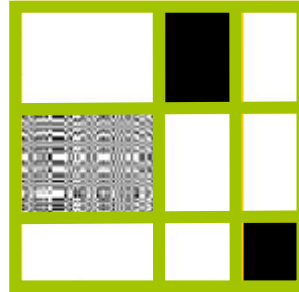
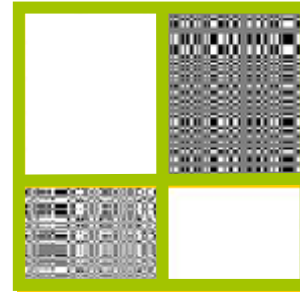
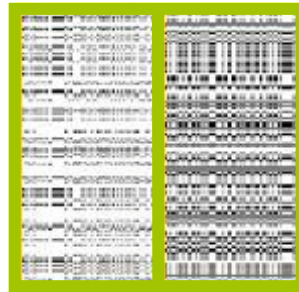
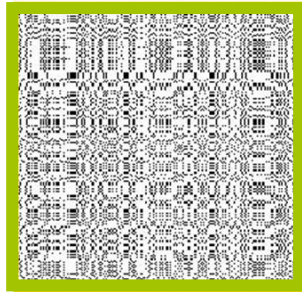
# Reduce

1	1	0	0	1	0	0
1	1	0	0	1	0	0
2	0	0	1	0	1	1
2	0	1	1	0	1	1

# Ergebnis

	1	1	2	2	2	3
1	1	1	0	0	0	0
1	1	1	0	0	0	0
2	0	0	1	1	1	0
2	0	0	1	1	1	1

# Gruppenanzahl



**Distributed**

**Co-Clustering**

Spiros Papadimitriou, Jimeng Sun

“DisCo: Distributed Co-clustering  
with Map-Reduce”