

# State-of-the-art bibliometric mapping techniques

Ludo Waltman

Centre for Science and Technology Studies, Leiden University

Graduate course, September 23, 2009



Universiteit Leiden

# The 7 steps of the bibliometric mapping process

1. Delineation of the domain of interest
2. Selection of the type of map
3. Selection of the items to be mapped
4. Calculation of co-occurrences
5. Mapping
6. Visualization
7. Evaluation

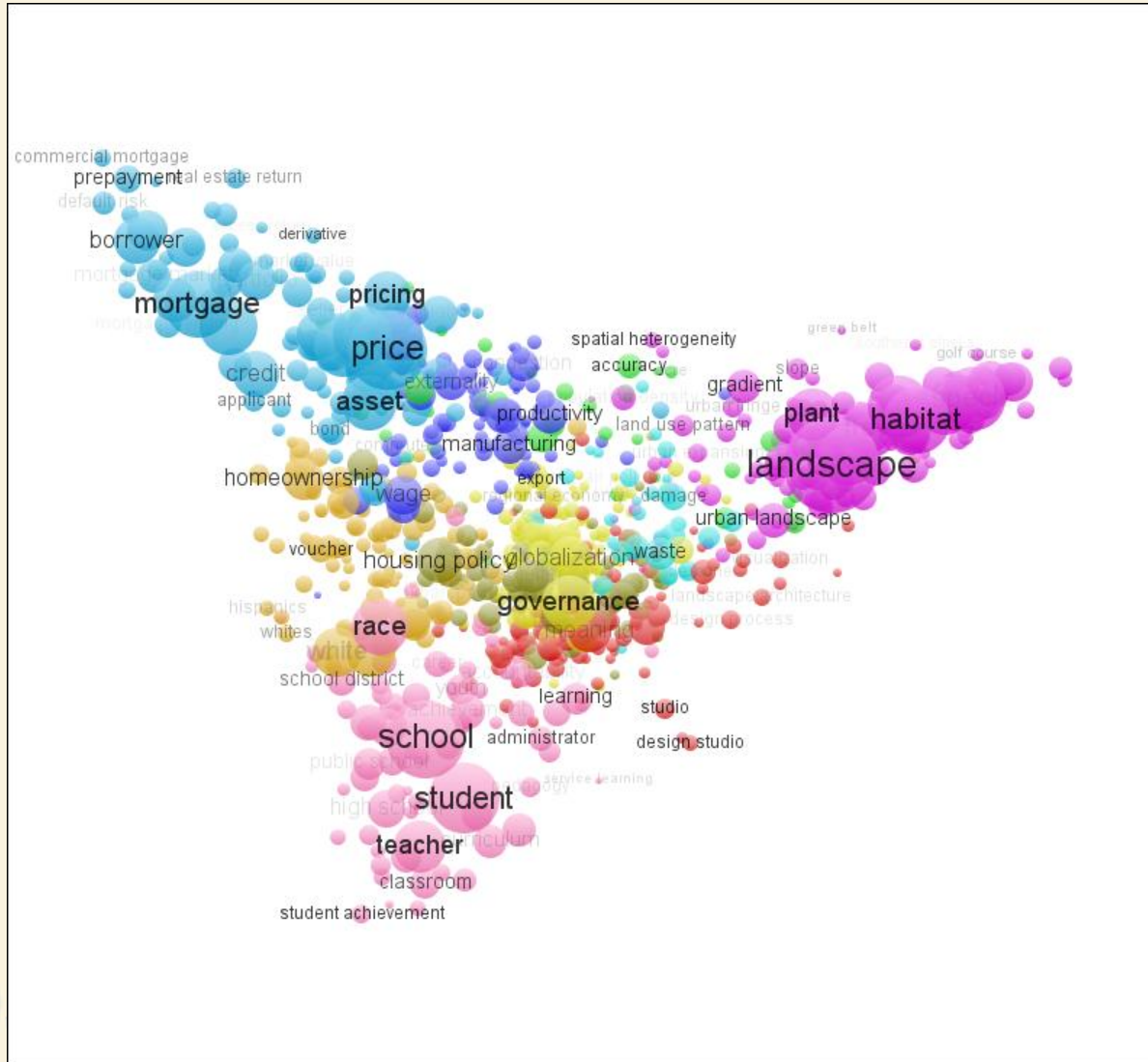


# Delineation of the domain of interest (1)

- Difficult and time-consuming issue
- Can sometimes be done based on Web of Science subject categories or based on a selected set of journals
- Especially difficult in the case of multidisciplinary domains; usually requires extensive expert input

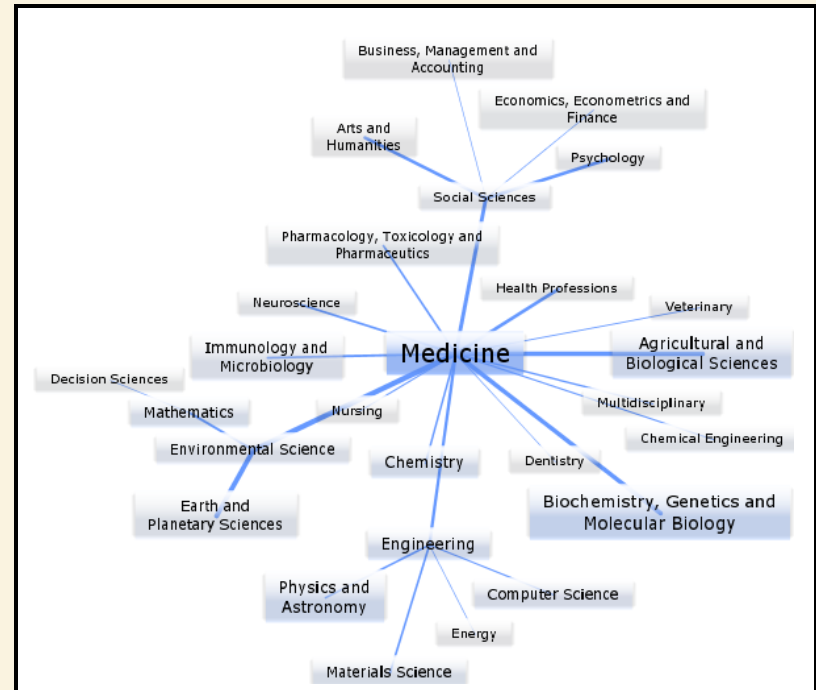
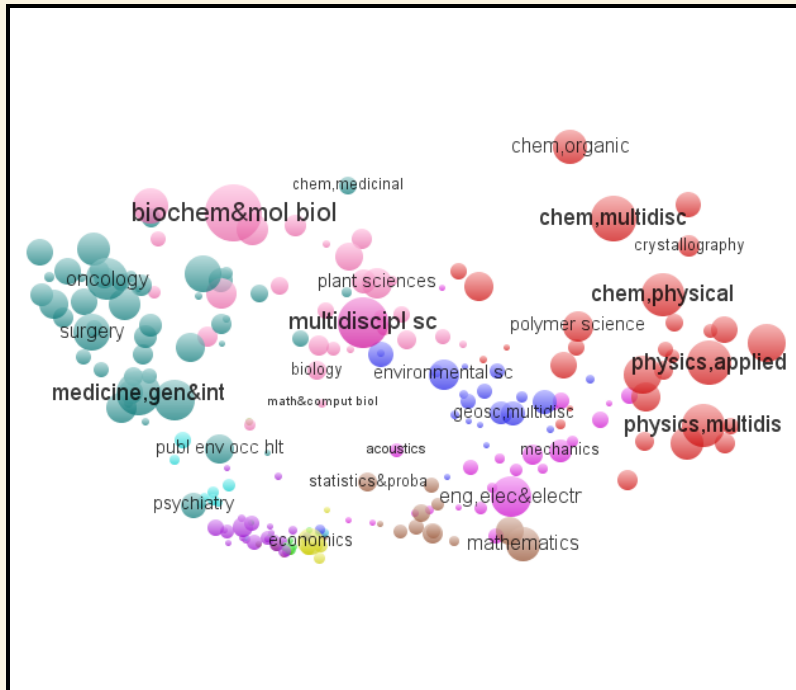


# Delineation of the domain of interest (2)



# Selection of the type of map (1)

- Distance-based or graph-based map?

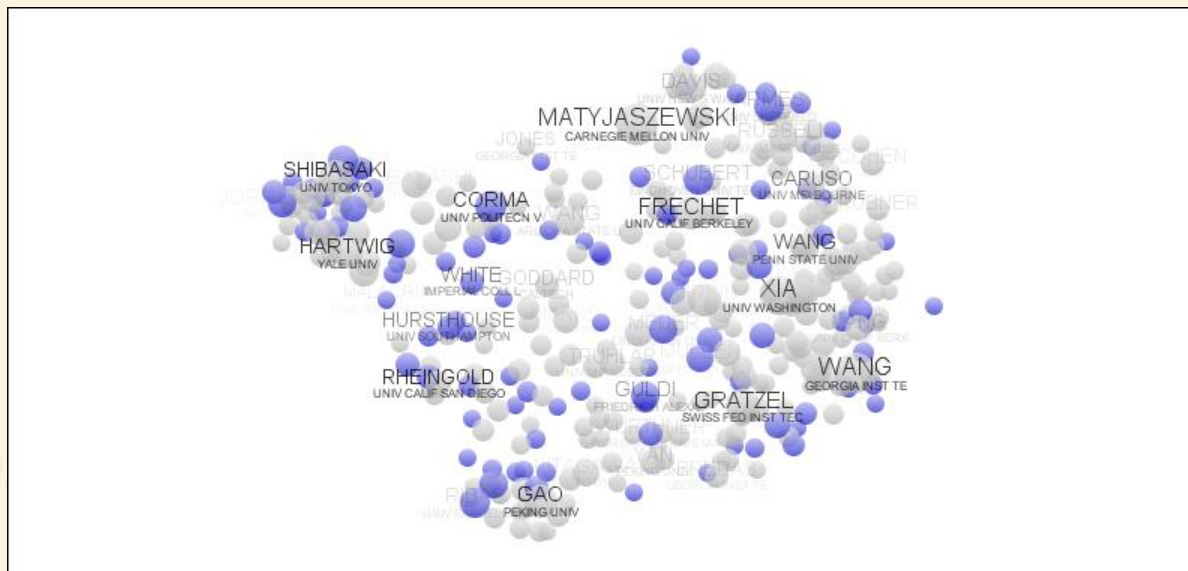
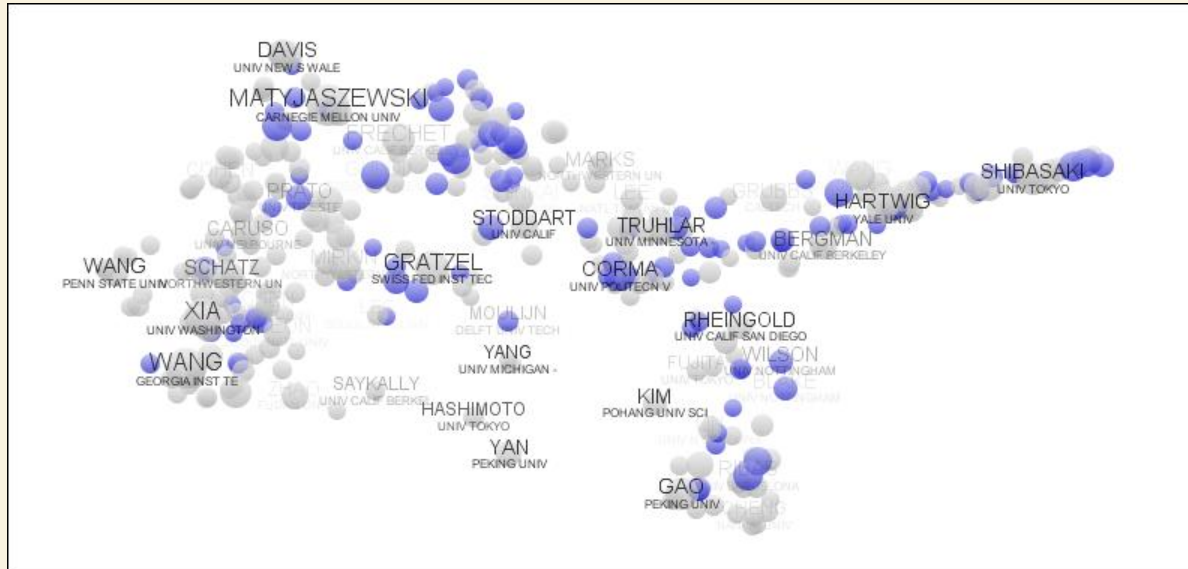


## Selection of the type of map (2)

- Type of item to be mapped?
  - Keywords
  - Documents
  - Authors
  - Journals
- How to measure the relatedness of items?
  - Co-occurrence of keywords in documents
  - Co-citation of documents, authors, or journals
  - Bibliographic coupling of documents, authors, or journals

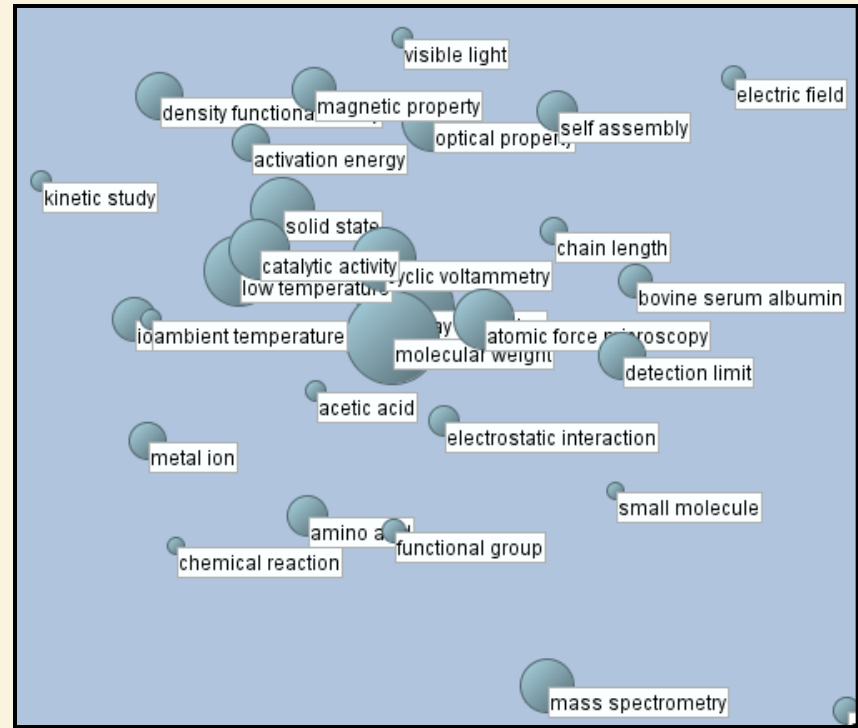
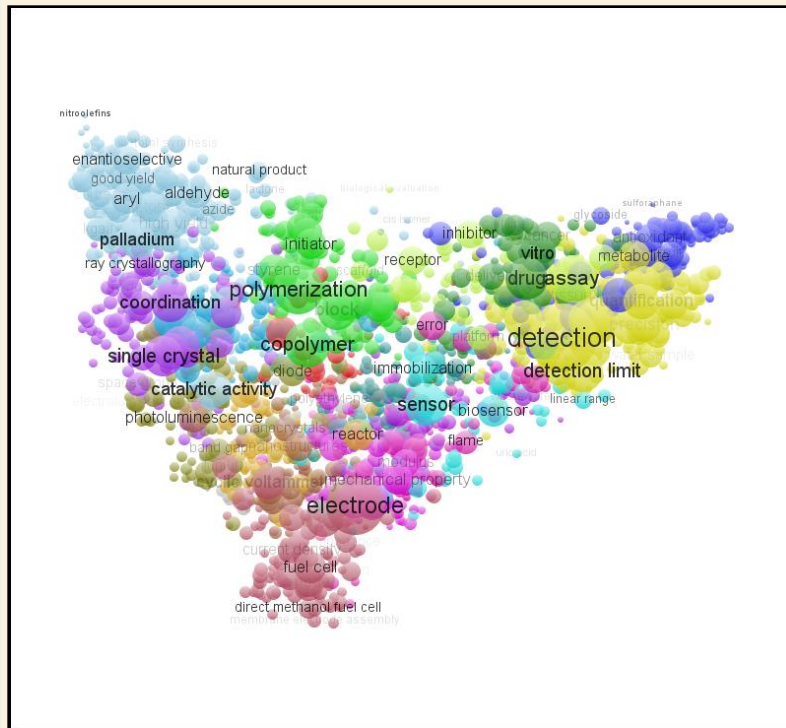


# Selection of the type of map (3)



# Selection of the type of map (4)

- Map of individual items or of clusters of items?



## Selection of the items to be mapped (1)

- Author selection can be difficult due to the problem of author identification
- Difficulties in keyword selection:
  - Keywords should not be too general
  - Keywords should not be ambiguous
  - Preferably, selection of synonyms should be avoided
  - List of relevant keywords is usually not available
  - Manual keyword selection is subjective and time-consuming



## Selection of the items to be mapped (2)

- Automatic keyword selection:
  1. Noun phrases are extracted from documents
  2. Documents are clustered into topics
  3. Noun phrases whose occurrences are biased towards one (or a few) topics are selected as relevant keywords



## Selection of the items to be mapped (2)

- Automatic keyword selection:
  1. Noun phrases are extracted from documents
  2. Documents are clustered into topics
  3. Noun phrases whose occurrences are biased towards one (or a few) topics are selected as relevant keywords

Topics

	Citation analysis	Bibliometric mapping	Patent analysis
Bibliometrics	54	32	23
Conclusion	108	132	76
Co-occurrence matrix	9	54	6
h-index	53	3	5
Impact factor	48	4	2
Innovation	11	13	52
Map	1	72	3
Patent	8	2	83
Result	76	99	52

Noun phrases



## Selection of the items to be mapped (2)

- Automatic keyword selection:
  1. Noun phrases are extracted from title/abstract data
  2. Documents are clustered into topics
  3. Noun phrases whose occurrences are biased towards one (or a few) topics are selected as relevant keywords

	Citation analysis	Bibliometric mapping	Patent analysis
Bibliometrics	54	32	23
Conclusion	108	132	76
Co-occurrence matrix	9	54	6
h-index	53	3	5
Impact factor	48	4	2
Innovation	11	13	52
Map	1	72	3
Patent	8	2	83
Result	76	99	52



# Calculation of co-occurrences (1)

Occurrence matrix

	Bibliometrics	h-index	Impact factor	Patent
Doc. 1		x	x	
Doc. 2	x	x		
Doc. 3	x			x
Doc. 4	x	x	x	
Doc. 5		x		x
Doc. 6		x		
Doc. 7		x	x	



# Calculation of co-occurrences (2)

Occurrence matrix

	Bibliometrics	h-index	Impact factor	Patent
Doc. 1		x	x	
Doc. 2	x	x		
Doc. 3	x			x
Doc. 4	x	x	x	
Doc. 5		x		x
Doc. 6		x		
Doc. 7		x	x	



Co-occurrence matrix

	Bibliometrics	h-index	Impact factor	Patent
Bibliometrics	3	2	1	1
h-index	2	6	3	1
Impact factor	1	3	3	0
Patent	1	1	0	2



# Calculation of co-occurrences (3)

Occurrence (citation) matrix

	Journal 1	Journal 2	Journal 3	Journal 4
Doc. 1		x	x	
Doc. 2	x	x		
Doc. 3	x			x
Doc. 4	x	x	x	
Doc. 5		x		x
Doc. 6		x		
Doc. 7		x	x	



# Calculation of co-occurrences (4)

Occurrence (citation) matrix

	Journal 1	Journal 2	Journal 3	Journal 4
Doc. 1		x	x	
Doc. 2	x	x		
Doc. 3	x			x
Doc. 4	x	x	x	
Doc. 5		x		x
Doc. 6		x		
Doc. 7		x	x	



Co-occurrence (co-citation) matrix

	Journal 1	Journal 2	Journal 3	Journal 4
Journal 1	3	2	1	1
Journal 2	2	6	3	1
Journal 3	1	3	3	0
Journal 4	1	1	0	2

# Mapping (1)

Mapping is done in two steps:

1. Normalization of co-occurrence frequencies
  - Association strength
  - Cosine
  - Jaccard index
2. Positioning of items in a 2D space
  - VOS (visualization of similarities)
  - MDS (multidimensional scaling)



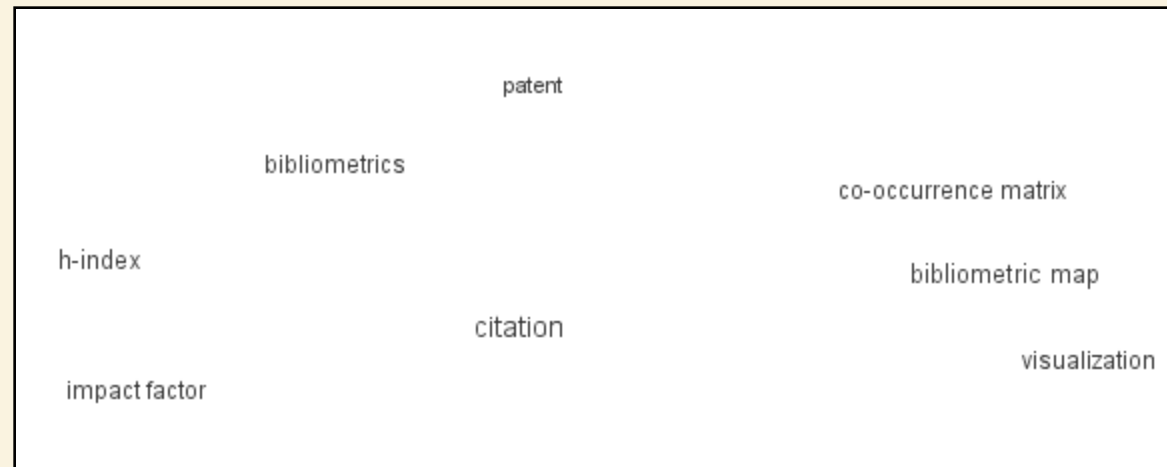
# Mapping (2)

	Bibliometrics	h-index	Impact factor	Patent	Citation	Bibliometric map	Co-occurrence matrix	Visualization
Bibliometrics	3							
h-index	2	5						
Impact factor	2	1	0					
Patent	3	5	4	1				
Citation	1	0	0	1	4			
Bibliometric map	1	0	0	1	3	4		
Co-occurrence matrix	0	0	1	0	2	5	3	
Visualization								



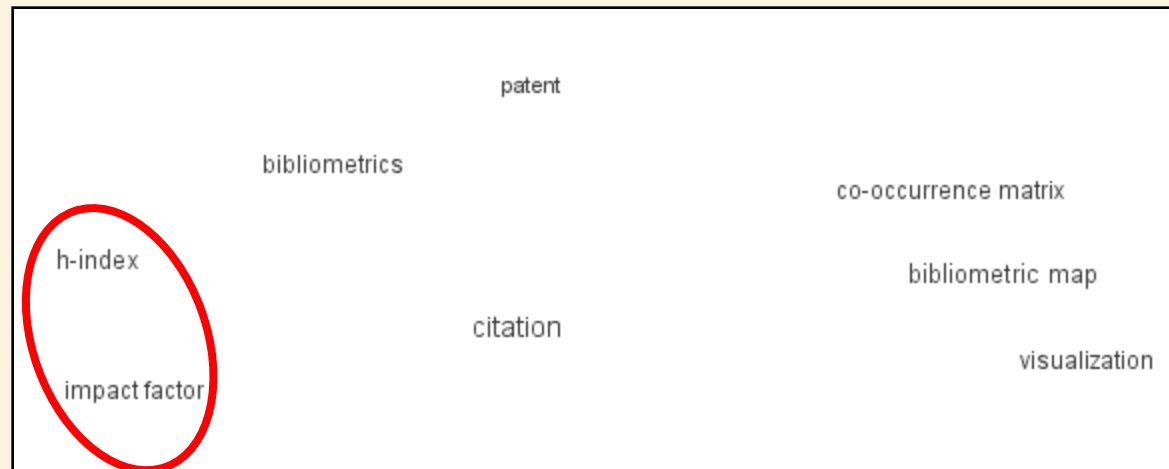
# Mapping (2)

	Bibliometrics	h-index	Impact factor	Patent	Citation	Bibliometric map	Co-occurrence matrix	Visualization
Bibliometrics	3							
h-index	2	5						
Impact factor	2	1	0					
Patent	3	5	4	1				
Citation	1	0	0	1	4			
Bibliometric map	1	0	0	1	3	4		
Co-occurrence matrix	0	0	1	0	2	5	3	
Visualization								



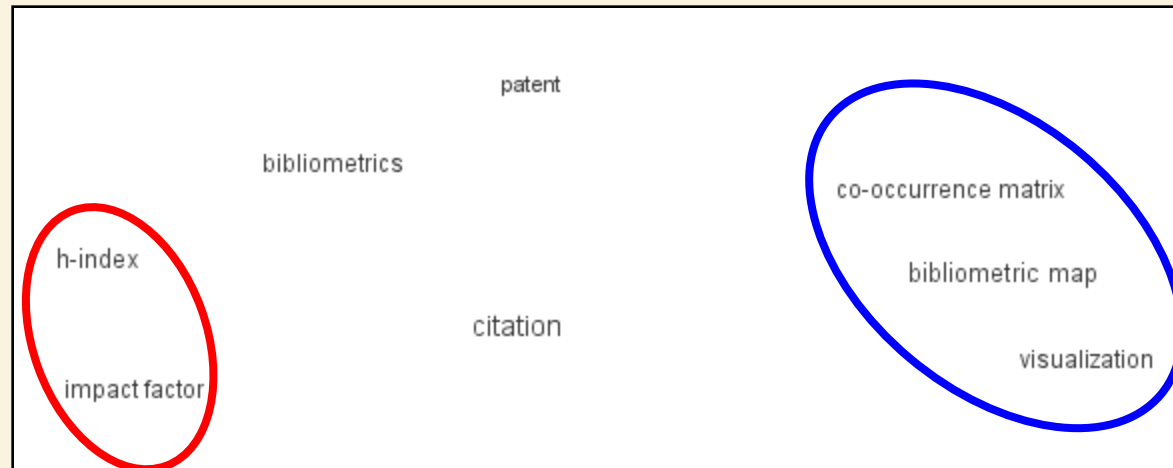
# Mapping (2)

	Bibliometrics	h-index	Impact factor	Patent	Citation	Bibliometric map	Co-occurrence matrix	Visualization
Bibliometrics	3	2	2	3	1	1	1	0
h-index	2	<b>5</b>	1	5	0	0	0	0
Impact factor	2	1	0	4	0	0	0	1
Patent	2	1	0	1	0	0	1	0
Citation	3	5	4	1	4	3	3	2
Bibliometric map	1	0	0	1	4	4	5	3
Co-occurrence matrix	1	0	0	1	3	4	5	3
Visualization	0	0	1	0	2	5	3	3



# Mapping (2)

	Bibliometrics	h-index	Impact factor	Patent	Citation	Bibliometric map	Co-occurrence matrix	Visualization
Bibliometrics	3	5						
h-index	2	5	0					
Impact factor	2	1	0					
Patent	2	1	0					
Citation	3	5	4	1				
Bibliometric map	1	0	0	1	4			
Co-occurrence matrix	1	0	0	1	3	4		
Visualization	0	0	1	0	2	5	3	







## Visualization (3)

- Label size proportional with an item's total number of co-occurrences
- Coloring of items based on a clustering technique
- Labeling algorithm that prevents labels from overlapping
- Functionality to zoom in and out on the map
- Density view



## Visualization (4)

- The visualization software that we use is called VOSviewer
- The software has been made freely available at [www.vosviewer.com](http://www.vosviewer.com)
- This afternoon the software will be used in the practice session



# Evaluation (1)

- Bibliometric mapping is an iterative process; several iterations may be required to obtain a satisfactory map
- Evaluation requires domain knowledge
- Important issues are:
  - Has the domain been delineated in a satisfactory way?
  - Have the correct keywords been selected?



# Evaluation (2)

