Qualitative Citation Analysis Based on Formal Concept Analysis

Wiebke Petersen & Petja Heinrich
Institute of Language and Information
University of Düsseldorf

Overview

- **aim:** to present the FCA as an applicable method in the bibliometrics
- **method:** Formal concept analysis (FCA)
- **approach:** an example of bibliographic analisys using FCA
- tools and applications
- outlook

Some Basic Objectives and Aims of Bibliographic Studies

- citations and references (co-citations, co-references)
- co-authorship and scientific cooperation
- journal impact factor

Some Basic Objectives and Aims of Bibliographic Studies

- discovery of scientific influences and knowledge flows
- showing trends („hottest“ themes)
- determining the influential journals, authors and scientific ideas
- strategic publishing

Basic terms: Bibliographic Coupling

Publication P

Text text text text text text text text text text
text text text text text text text text text text
text text text text text text text text text text
text text text text text text text text text text
text text text text text text

References:

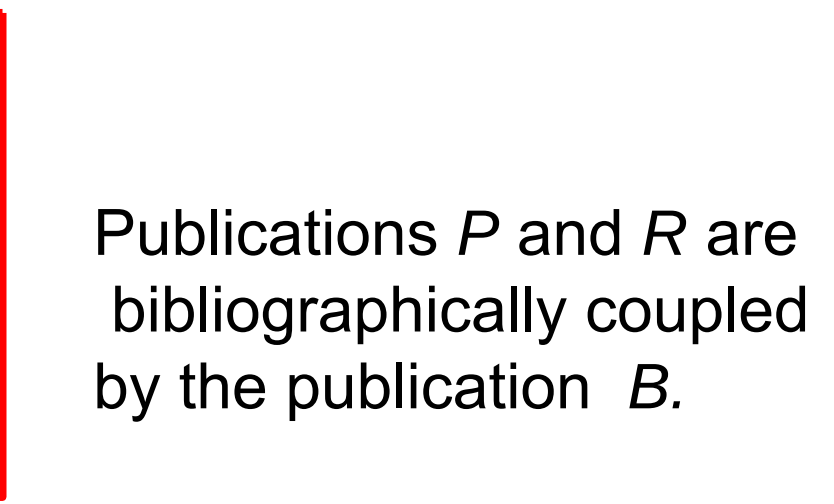
- Publication A
- Publication B

Publication R

Text text text text text text text text text text
text text text text text text text text text text
text text text text text text text text text text
text text text text text text text text text text
text text text text text text

References:

- Publication B
- Publication C



Publications *P* and *R* are bibliographically coupled by the publication *B.*

Basic Terms: Co-citation

Publication P

Text text text text text text text text text text
text text text text text text text text text text
text text text text text text text text text text
text text text text text text text text text text
text text text text text text

References:

- Publication A
- Publication B

Publications *A* and *B* are co-cited by the publication *P*.

Publication R

Text text text text text text text text text text
text text text text text text text text text text
text text text text text text text text text text
text text text text text text text text text text
text text text text text text

References:

- Publication B
- Publication C

Basic terms: Co-authorship

Publication P

Authors: **A1**
A2

Text text text text text text text text text text text text text text text text text text text

References:

- Publication A
- Publication B

Publication R

Authors: **A1**
A3, A4

Text text text text text text text text text text text text text text text text text text text

References:

- Publication B
- Publication C

The authors *A1* and *A2* are co-authors of the Publication P.

Quantitative and Qualitative Citation Analysis

Quantitative methods:

- applying of statistical methods

Qualitative methods:

- revealing of trends (citation or publication trends, patterns of co-authorship)
- visualisation – citation graphs, digrams

Formal Concept Analysis (FCA)

Formal Context

Def.: A **formal context** K is a tripel (G,M,I) where

- G is a set of objects
- M is a set of attributes
- I is a binary relation $I \subseteq G \times M$ where $(g,m) \in I$ is read as "object g has attribute m."

	direkt verwandt	älter	jünger	männlich	weiblich	eindeutig	andere Generation
Vater	×	×		×		×	×
Mutter	×	×			×	×	×
Bruder	×			×			
Schwester	×				×		
Kind	×		×				×
Sohn	×		×	×			×
Tochter	×		×		×		×
Onkel		×		×			×
Tante		×			×		×
Opa		×		×			×
Oma		×			×		×
Cousin				×			
Cousine					×		
Neffe			×	×			×
Nichte			×		×		×

Formal Concept Analysis (FCA)

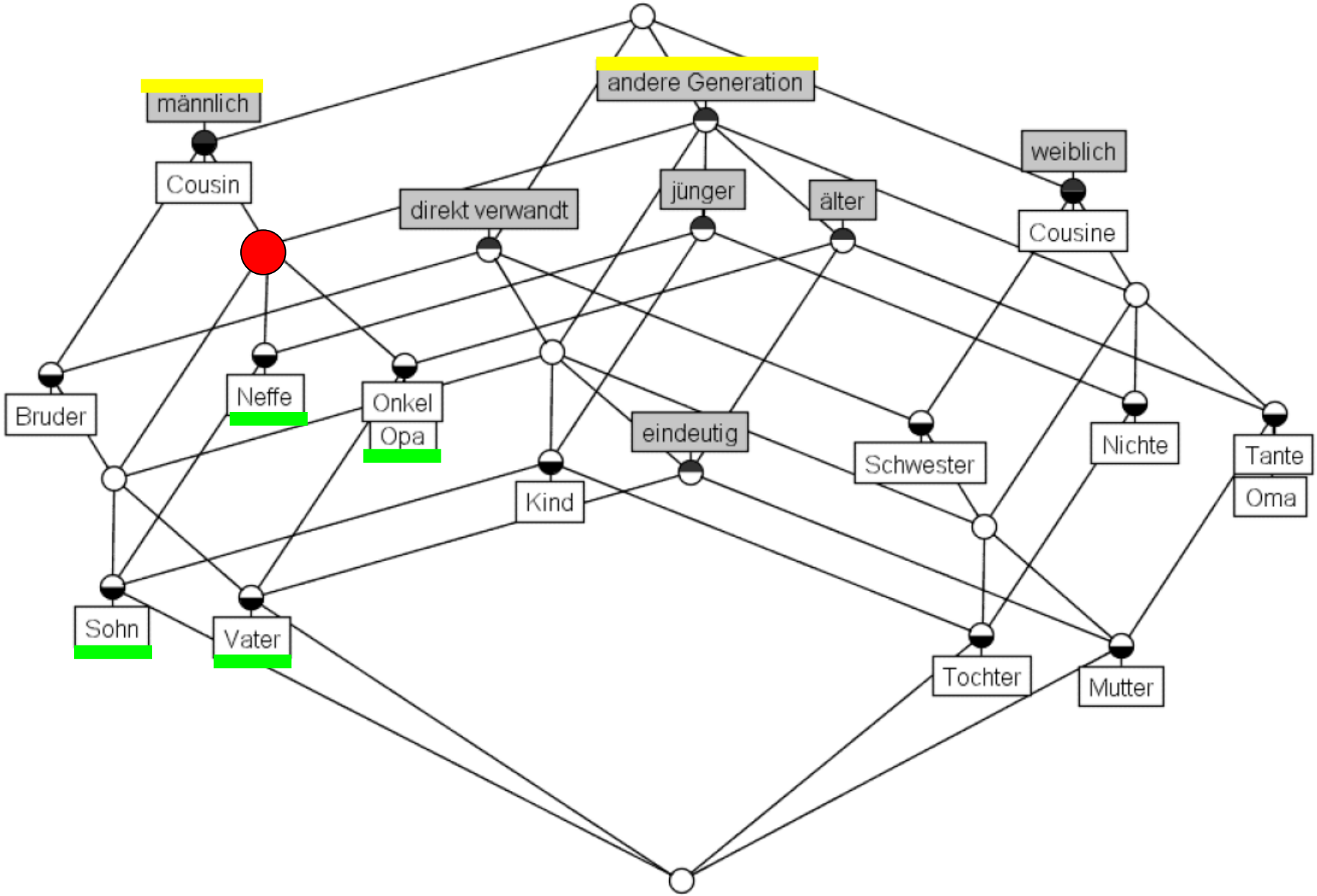
Formal Concept:

Def.: for A⊆G and B⊆M be **A'**={m∈M| ∀g∈A: (g,m)∈I} **B'**={g∈G| ∀m∈B: (g,m)∈I}

Def.: (A,B) is a **formal concept** of the formal context (G,M,I) if A⊆G, B⊆M, A'=B and B'=A.
A is called the **extent** and B the **intent** of the concept.

	direkt verwandt	älter	jünger	männlich	weiblich	eindeutig	andere Generation
Vater	×	×		×		×	×
Mutter	×	×			×	×	×
Bruder	×			×			
Schwester	×				×		
Kind	×		×				×
Sohn	×		×	×			×
Tochter	×		×		×		×
Onkel		×		×			×
Tante		×			×		×
Opa		×		×			×
Oma		×			×		×
Cousin				×			
Cousine					×		
Neffe			×	×			×
Nichte			×		×		×

FCA: Concept Lattice



Citation Analysis with FCA

	1	2	3	4	5	6
c1			x			
c2			x	x	x	x
c3				x		x
c4					x	
c5						
c6						

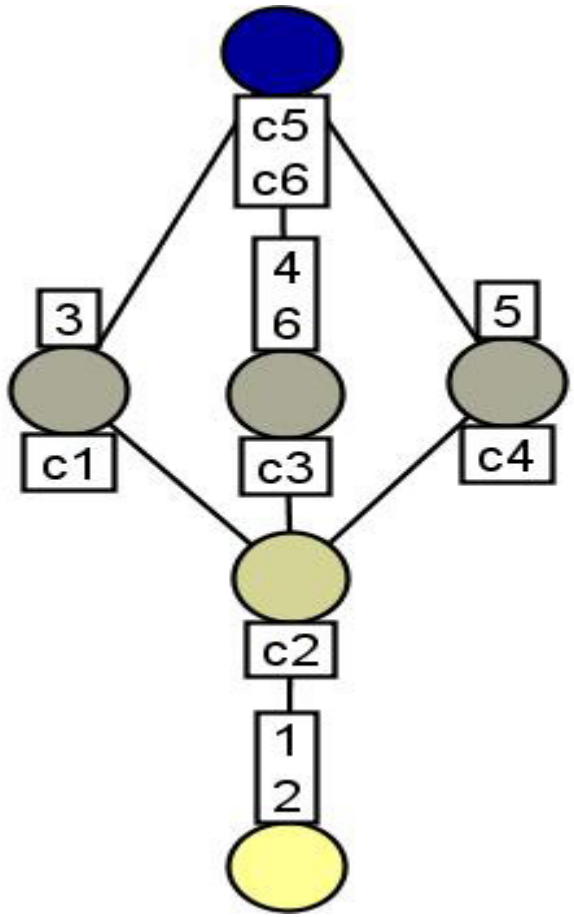
- The citation of a publication P is denoted by cite(P) or shorthand cP.
- The intent of the object concept of cite(X), i. e. the citation of publication X, is the set of publications citing X.
- Example: The intent of the object concept c2 is {3,4,5,6}.
- The extent of the atribute concept of a publication X is the set of citations in X
- Example: The extent of the atribute concept 3 is {c1,c2}.

Co-citation

•Two publications are co-cited if they both are cited by a third publication.

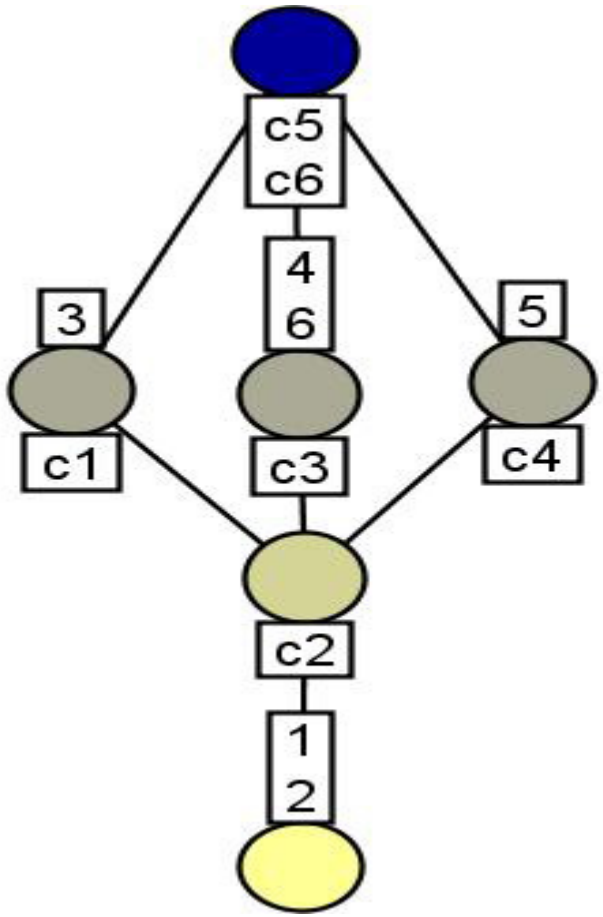
• P,R∈Pub are co-cited iff {cite(P), cite(R)}‘≠{}

• Example:
1 and 2 are co-cited (both are cited by 3).
cite(1)=c1, cite(2)=c2, {c1,c2}´= {3}

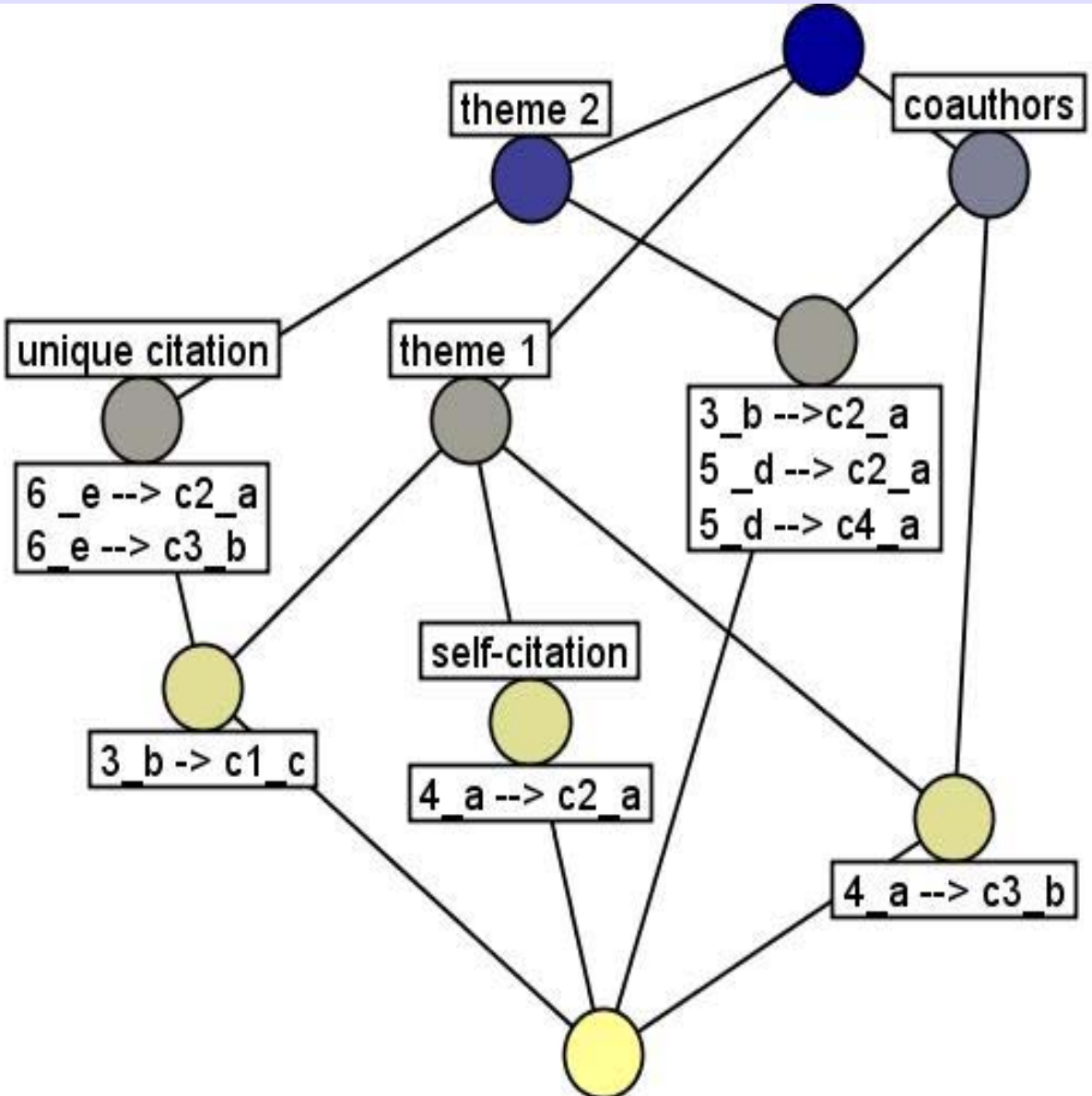


Bibliographic Coupling

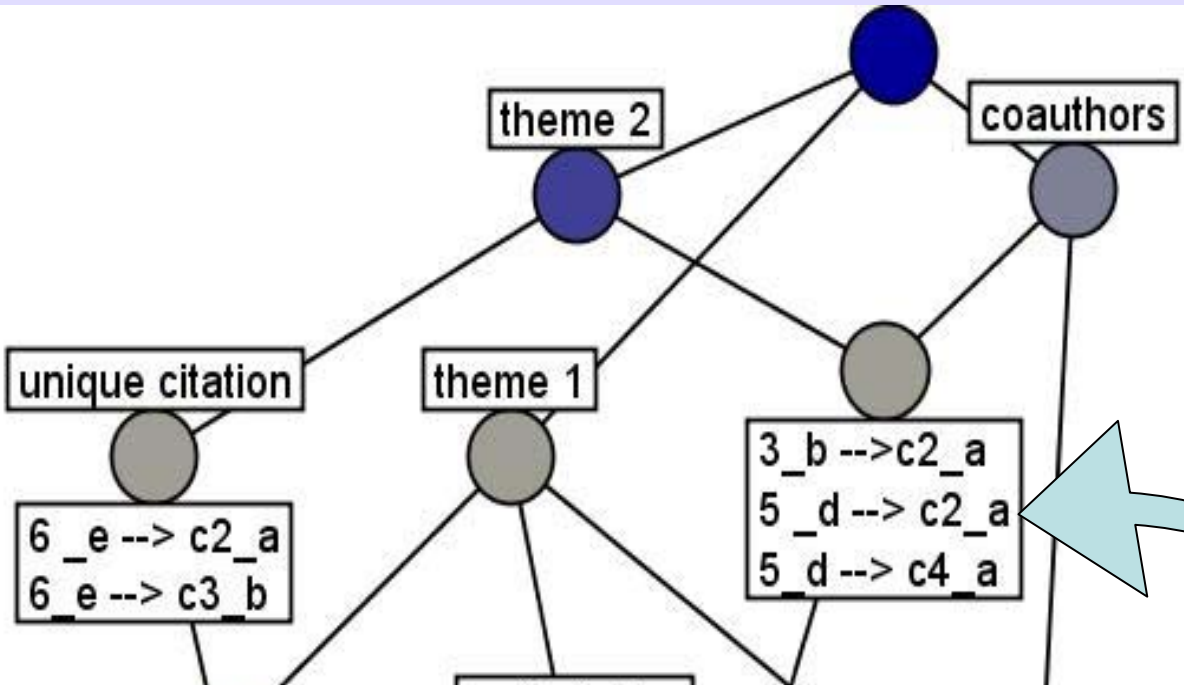
- Two publications are bibliographic coupled if they share a common reference
- P,R∈Pub are bibliographic coupled iff
  {P,R}´≠{}
- The strength of the bibliographic coupling is |{p,r}´| (the number of references which both the publications share)
- Example: 4 and 6 are bibliographic coupled by c3, c2 ({4,6}´= {c3,c2}).



A Model of Self-citations

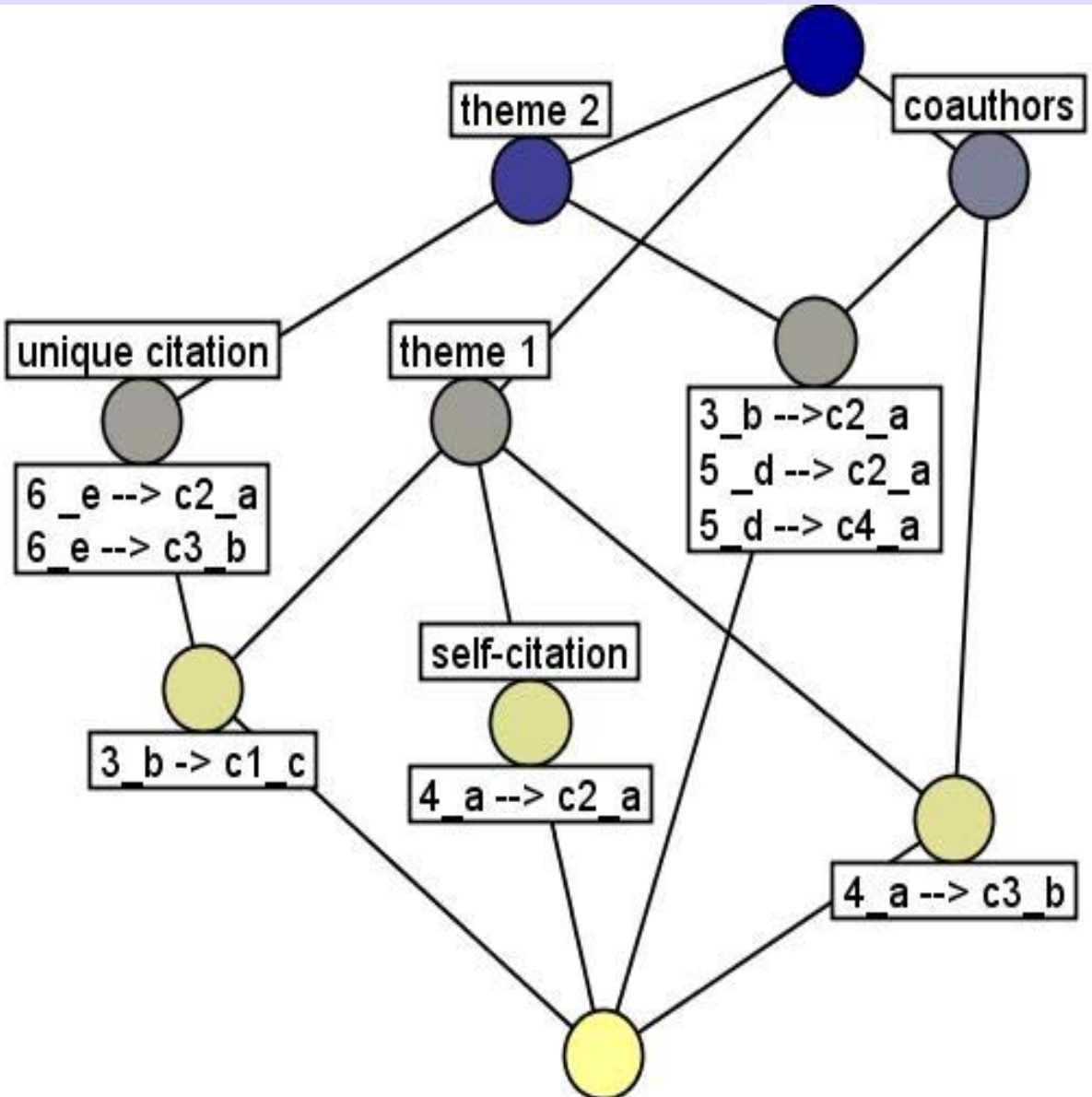


A Model of Self-citations



	1	2	3	4	5	6
c1			x			
c2			x	x	x	x
c3				x		x
c4					x	
c5						
c6						

Self-citations



Co-autorship

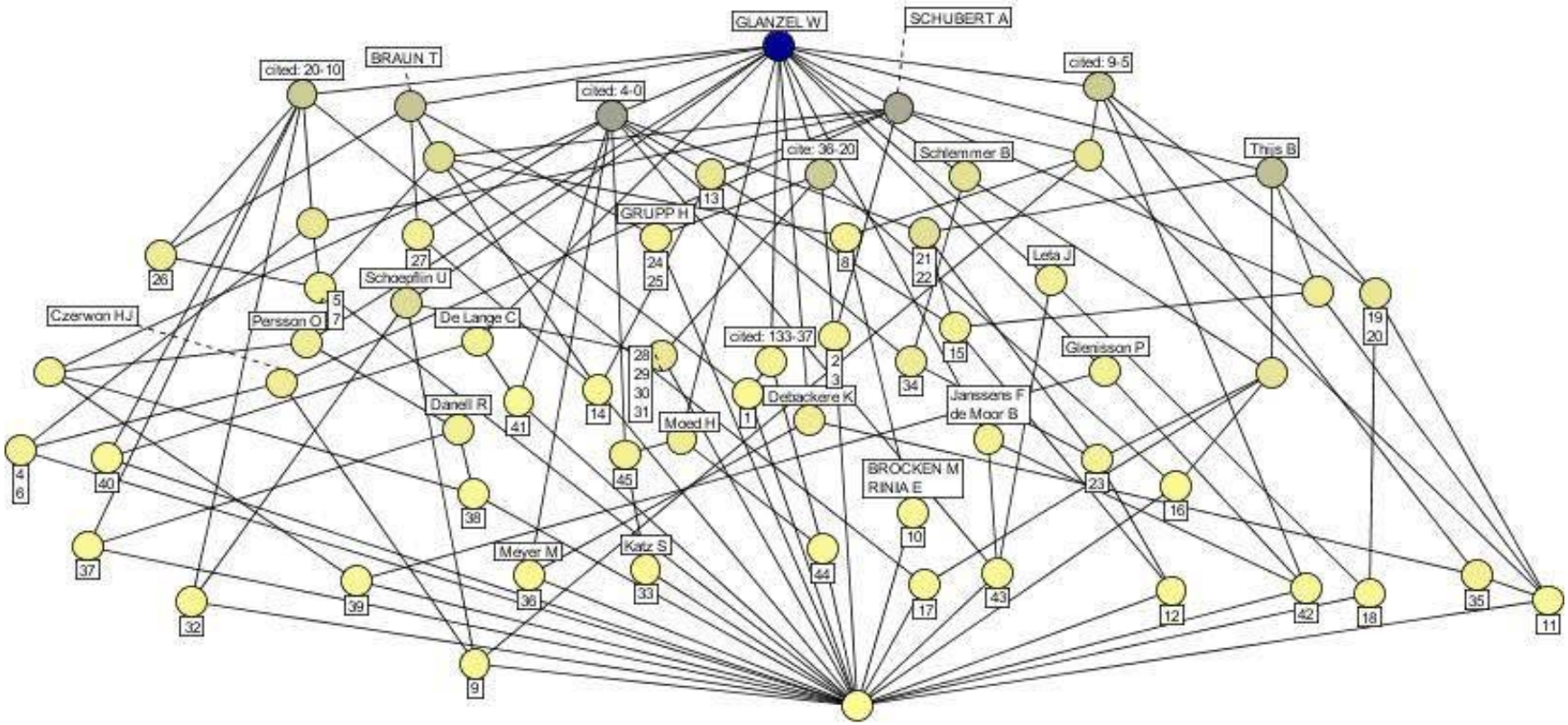
- The research: discovering and analising the scientific cooperations in the field of bibliometrics

- Query in *Web of Knowledge*:

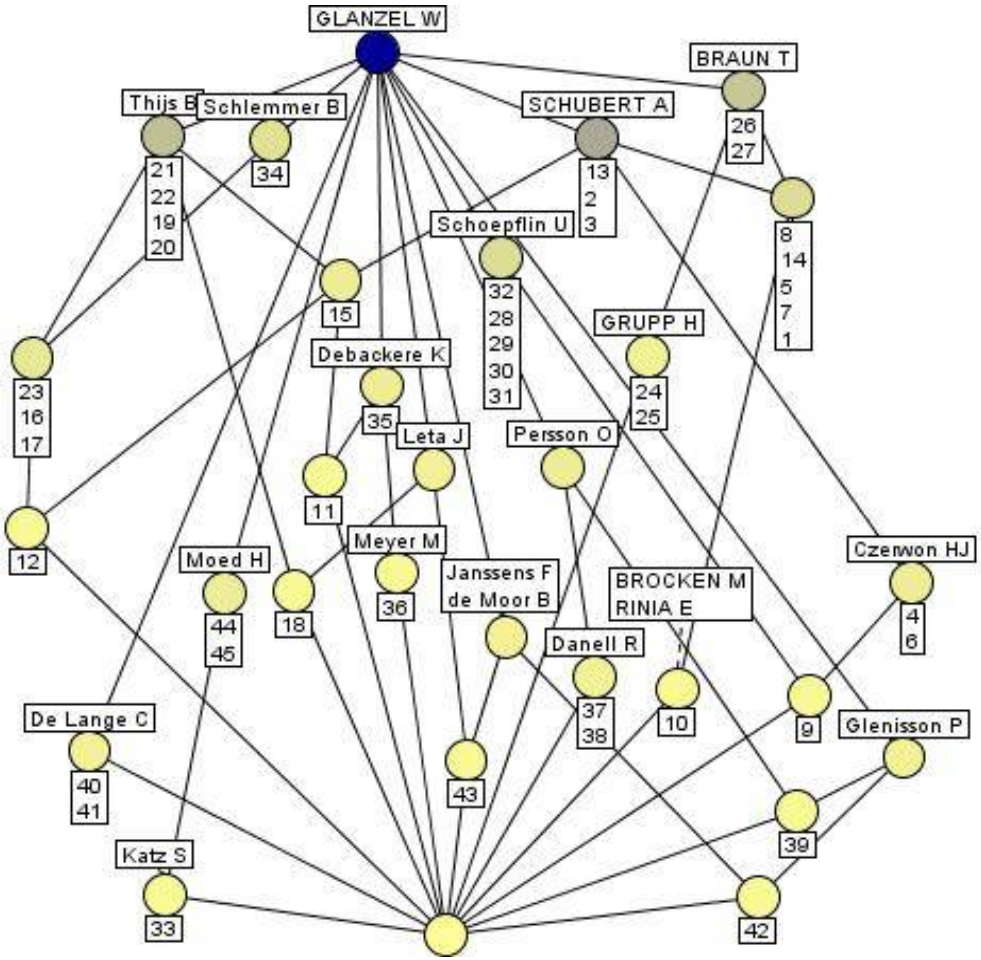
  informetr OR bibliometr* OR scientometr**
  Result: 2460 Documents

- Ranking of the authors and finding out who is the one with the biggest sets of publications: Glänzel, W.
- Exploring the cooperation patterns and showing a method for the creating an author‘s cooperation profile

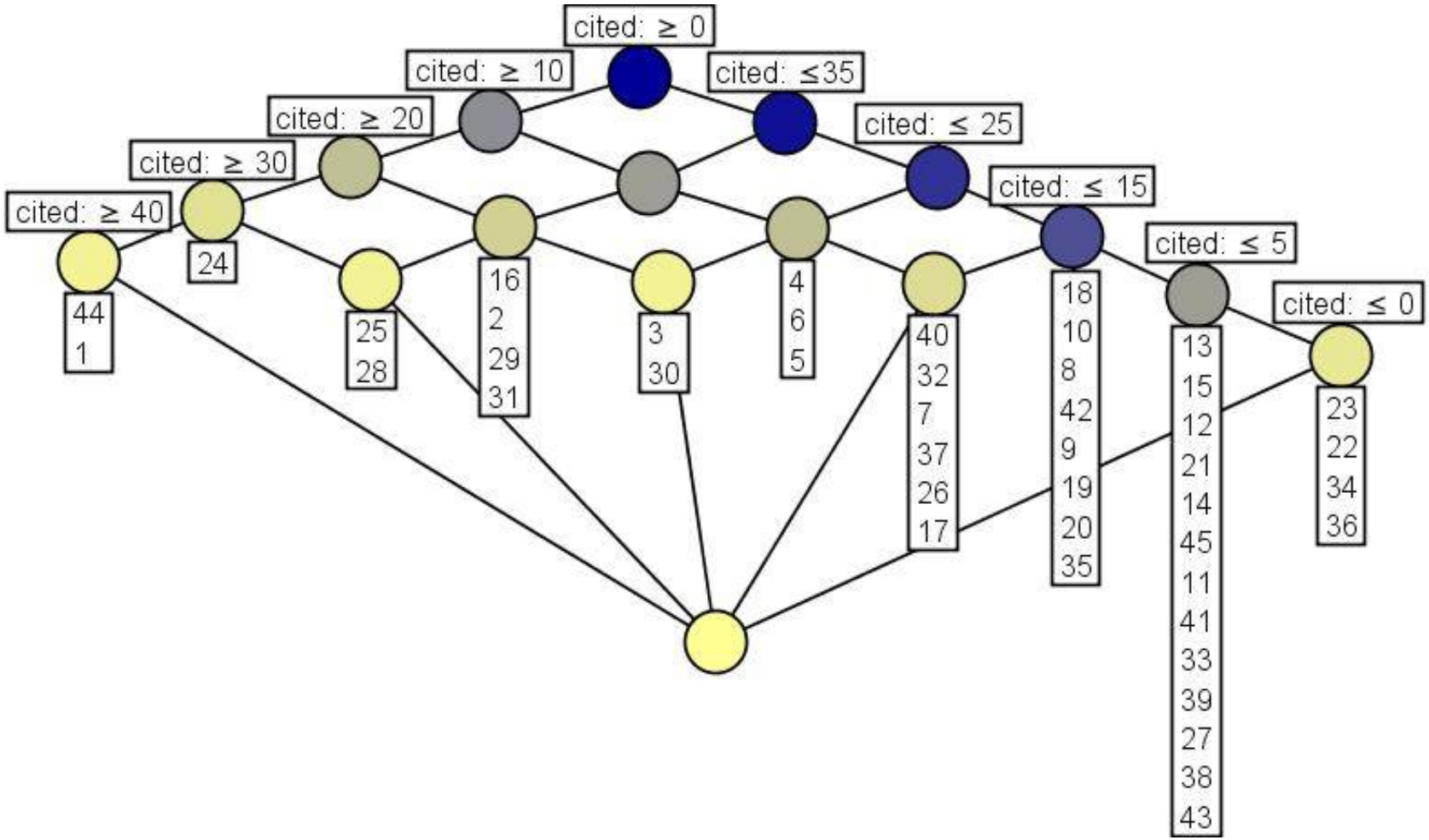
Cooperation profile of Glänzel, W.



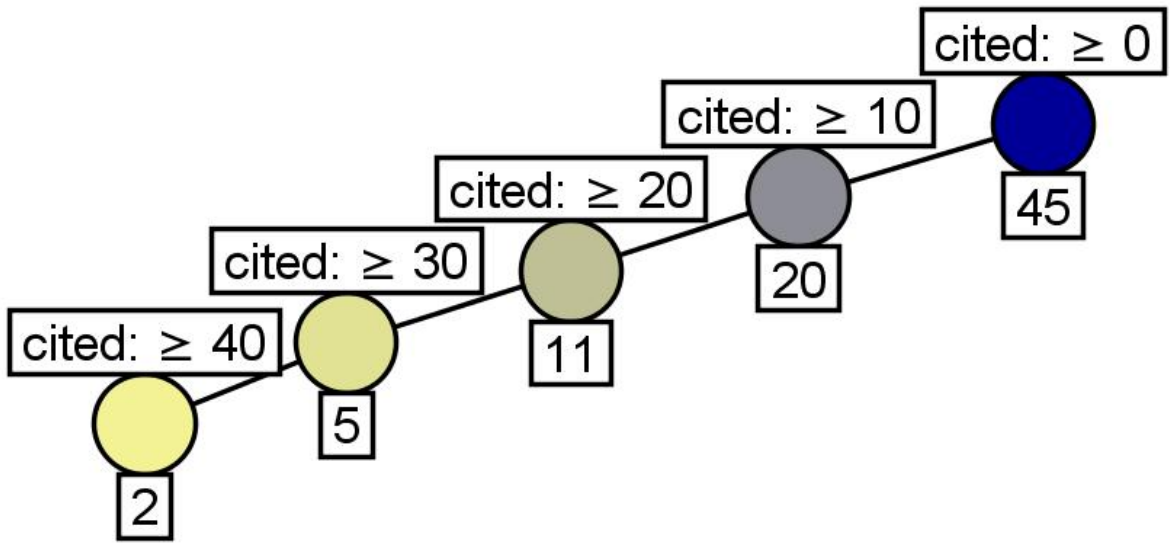
Cooperation Profile: Publications only



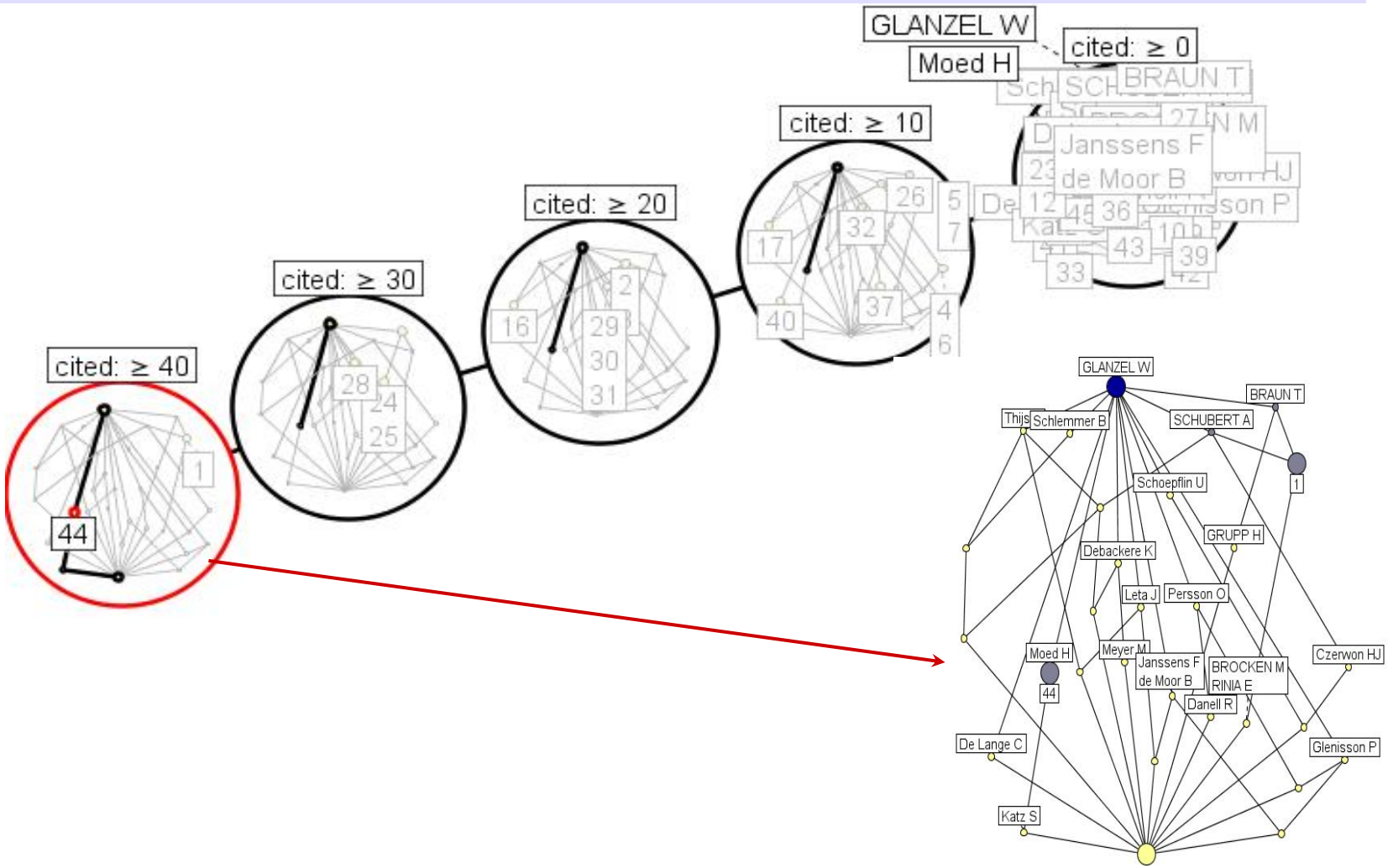
Interordinal Scale



Ordinal Scale



Ordinal Scale – Nested Diagram



FCA in the Bibliometics - Advantages

- an easy usage of different scales
- allows a dynamic browsing of the data
- shows logical implications
- „Zooming“ – adding attributes or objects is easy

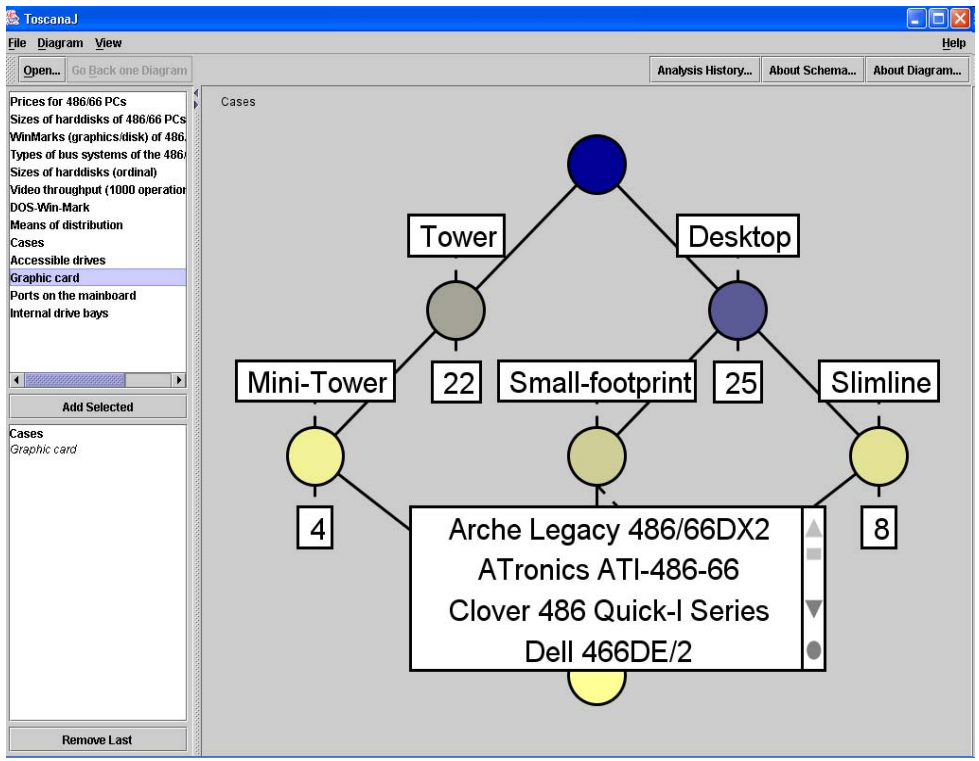
Applications

- Toscana J:

http://toscanaj.sourceforge.net/

.

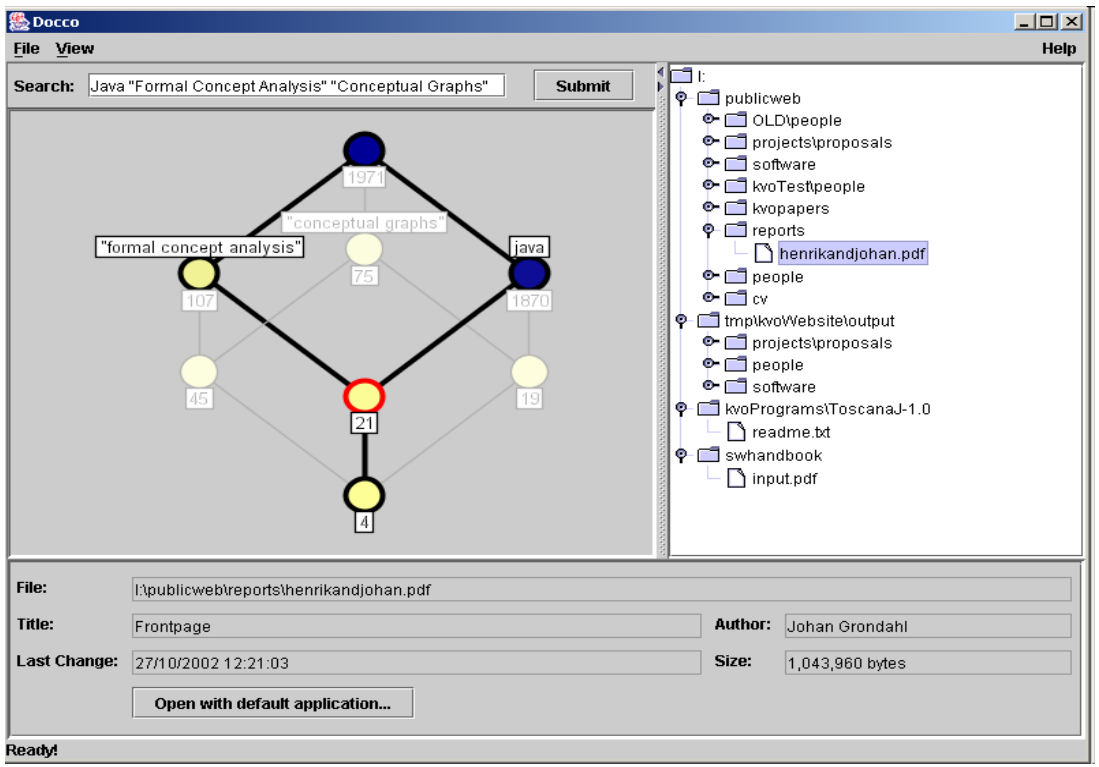
- First FCA literature retrieval system in a library (ZIT)



Applications

- Docco:

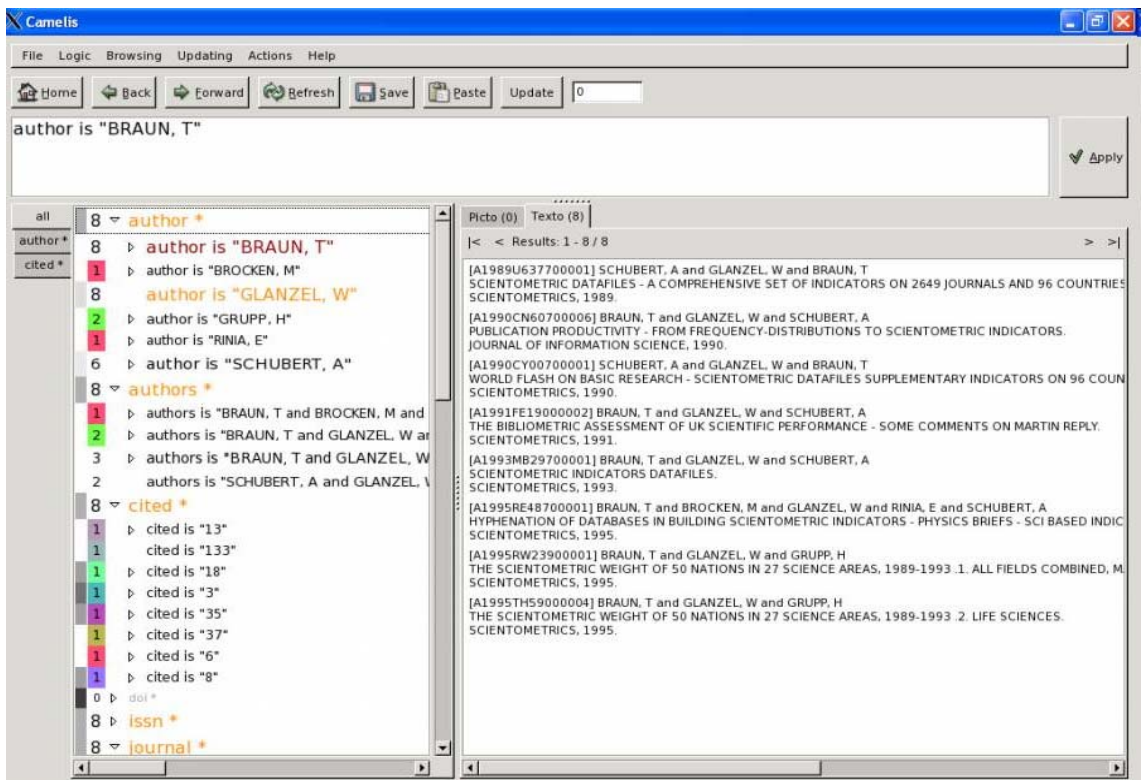
http://tockit.sourceforge.net/docco/



Applications

- Camelis:

http://www.irisa.fr/LIS/ferre/camelis/documentation.html



Outlook

- a model and a micro-bibliographic analysis
- high potential for visualisation of citation dependecies
- mining logical connection in sets of bibliographical data
- FCA as a supporting method
- tools for enrichment of bibliographical data and for personal archiving

Thank you for your attention!