Co-occurrence and Knowledge Mapping to Identify Hot Topics and Key Players in the Fields of Mobility and Transport

8th ISSI, Sydney, 2001

International Society for Scientometrics and Informetrics

C. Widhalm[1], U. Gigler, A. Kopcsa and E. Schiebel

1) Neulerchenfelder Str. 19/1/6,

A-1160 Vienna, Austria

E-mail: clemens@widhalm.co.at

http://www.widhalm.co.at

Some Questions ...

If we study a research field we may be interested in ...

- What are the relevant topics and subtopics?
- How did they develop?
- Who and where are the relevant keyplayers?
  - institutions
  - authors

Where do We Find the Answers?

We might find experts with their specific points of view

or we go for wide-spread published literature such as

- scientific journals
- research project descriptions
- patents

We took documents related to mobility and transport...

Source DB for scientific publications (# of documents)

- INSPEC (563) physics, electronic & computer engineering
- NTIS (404) technology
- Compendex (348) engineering
- TRIS (1910) transport and traffic
- Enviroline (78) environment(al technologies)
- Env.Bib. (29) environment(al technologies)
- CORDIS (977) project descriptions of 4th Frameworkprogr. of EC

TOTAL (4309)

How can We Find the Answers?

- We look for the types of objects

  offered in the documents

- We calculate relations between theese objects

- We visualise these relations and evaluate them

Original document (example)

82/5/2831 (Item 1250 from file: 63)
DIALOG(R)File 63:Transport Res(TRIS)
(c) fmt only 2000 Dialog Corp. plc. All rts. reserv.
00742783 DA
TITLE: RECENT DEVELOPMENTS IN WEATHER RELATED TRAFFIC MANAGEMENT
AUTHOR(S): PAPAGEORGIOU, M(ED); POULIEZOS, A(ED); KULMALA, R
CORPORATE SOURCE: INTERNATIONAL FEDERATION OF AUTOMATIC CONTROL (IFAC), 2361 LAXENBURG , SCHLOSSPLATZ 12, AUSTRIA
JOURNAL: TRANSPORTATION SYSTEMS. PREPRINTS OF THE 8TH IFAC/IFIP/IFORS SYMPOSIUM, CHANIA, CRETE, GREECE, 16-18 JUNE 1997, VOL 2 Pag: 725-8
PUBLICATION DATE: 19970000 PUBLICATION YEAR: 1997
LANGUAGE: ENGLISH SUBFILE: IRRD (I)
IRRD DOCUMENT NUMBER: 891539
REFERENCES: 15
DATA SOURCE: Transport Research Laboratory (TRL)
ABSTRACT: Poor weather conditions such as fog, ice, snow etc are causing substantive safety problems as well as considerable efficiency and environment problems in European transport. Transport telematics systems can be a potential solution. Three different approaches are identified: improving weather-related road maintenance, improving drivers' knowledge of current weather conditions, and influencing driver behaviour in adverse conditions. Recent developments in thes e areas of application are discussed. (A) For the covering abstract see IRRD 891457.
DESCRIPTORS: CONFERENCE; WEATHER; TRAFFIC FLOW; SAFETY; INTELLIGENT TRANSPORT SYSTEM; TELECOMMUNICATION; DATA PROCESSING; MAINTENANCE; DRIVER INFORMATION; DRIVER BEHAVIOUR

We Reduce Contents of Documents to Keywords

- Given descriptors
- Automatic indexing of titles and abstracts
  - remove stopwords
  - stemming
  - phrase recognition
  - manual standardisation

Extracted keywords of example document:
application
behaviour
data processing
driver information
drivers
environmental aspects
european transport
ice
intelligent transport systems
maintenance
safety
telecommunication
traffic flow
weather

What are the Document Objects We Use?

- keywords (and clusters of keywords (see later))
- publication year
- institutions (plus respective country and region)
- authors

Relations: some Concepts of Co-occurrencies

- Which keywords are related to each other?

  co-occurrence of keywords in documents

  → co-word analysis

- Which institutions publish on same topics?

  co-occurrence of institutions in documents of similar content

  → co-institution analysis

- Which authors publish together?

  co-occurrence of authors in common documents

  → co-author analysis

Co-occurrency of keywords in documents

In how many documents a pair of keywords is used?

co-occurrency	application	drivers	highway traffic control	intelligent transport systems	line	maintenance	management	monitoring	network	real time systems	safety	traffic flow
application	453	8	19	48	19	78	37	16	42	19	42	20
drivers	8	116	5	16	4	18	12	5	12	11	32	12
highway traffic control	19	5	112	12	3							8
intelligent transport systems	48	16	12	292	6							20
line	19	4	3	6	140							5
maintenance	78	18	16	40	38							20
management	37	12	27	26	12							12
monitoring	16	5	5	12	11	38					13	6
network	42	12	11	9	15	28					10	15
real time systems	19	11	13	13	2	22					3	12
safety	42	32	7	30	23	84	24	13	10	3	351	12
traffic flow	20	12	8	20	5	20	12	6	15	12	12	168

$$J_{ij} = \frac{c_{ij}}{c_{ii} + c_{jj} - c_{ij}}$$

Jaccard

Cluster analysis of keywords

- Each keyword can be described as a vector, indicating the relation to all the other keywords in terms of the Jaccardindex
- This vector is just the respective column of the Jaccardmatrix
- If keywords can be described as vectors, a cluster analysis can be performed

Elements of two Clusters:

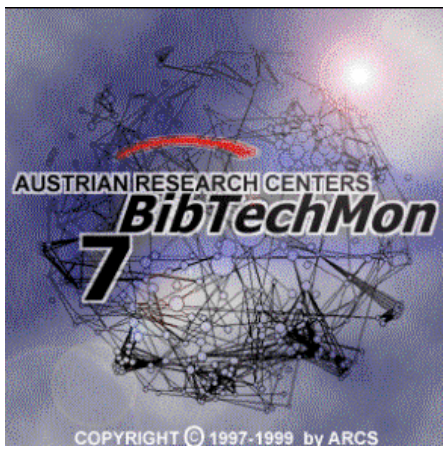
Fuels

Traffic Management and Control

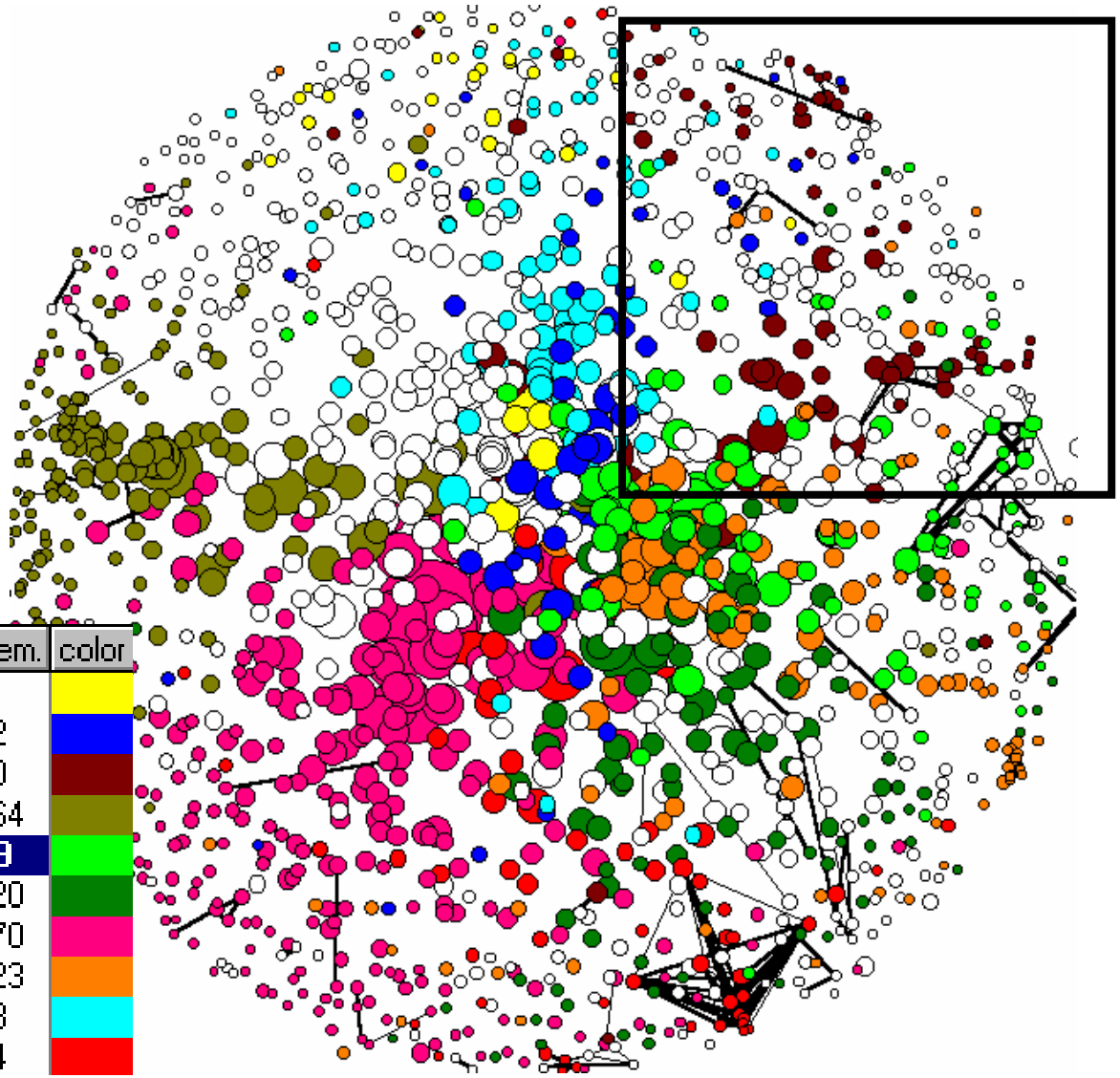
fuels	freq.	TMC	freq.
costs	325	traffic control	733
production	248	traffic management	323
benefits	221	intelligent transport systems	292
fuel	108	traffic	280
transport sector	100	control	263
energy	99	road	229
exploitation	97	road traffic	217
technical aspects	91	traffic engineering computing	183
flexible	86	traffic signals	178
prediction	82	traffic flow	168
alternative fuels	81	standards	155
electric vehicles	72	traffic control systems	153
expected achievements	59	location	148
cost benefit analysis	56	response	143
diesel	54	control systems	141
fuel cells	48	advanced traffic management systems	130
methanol	43	traffic congestion	121
electricity	40	travel	118
durability	36	highway traffic control	112
natural gas	36	speed	110

How can we Visualize Matrix and Clusters?

- Mapping based on co-occurency of keywords in documents and a mechanical spring model
- objects with high relation in terms of Jaccardindex are positioned closely together on the map.
- Objects of one cluster have same colour in the map
- Tool applied : *BibTechMon™*



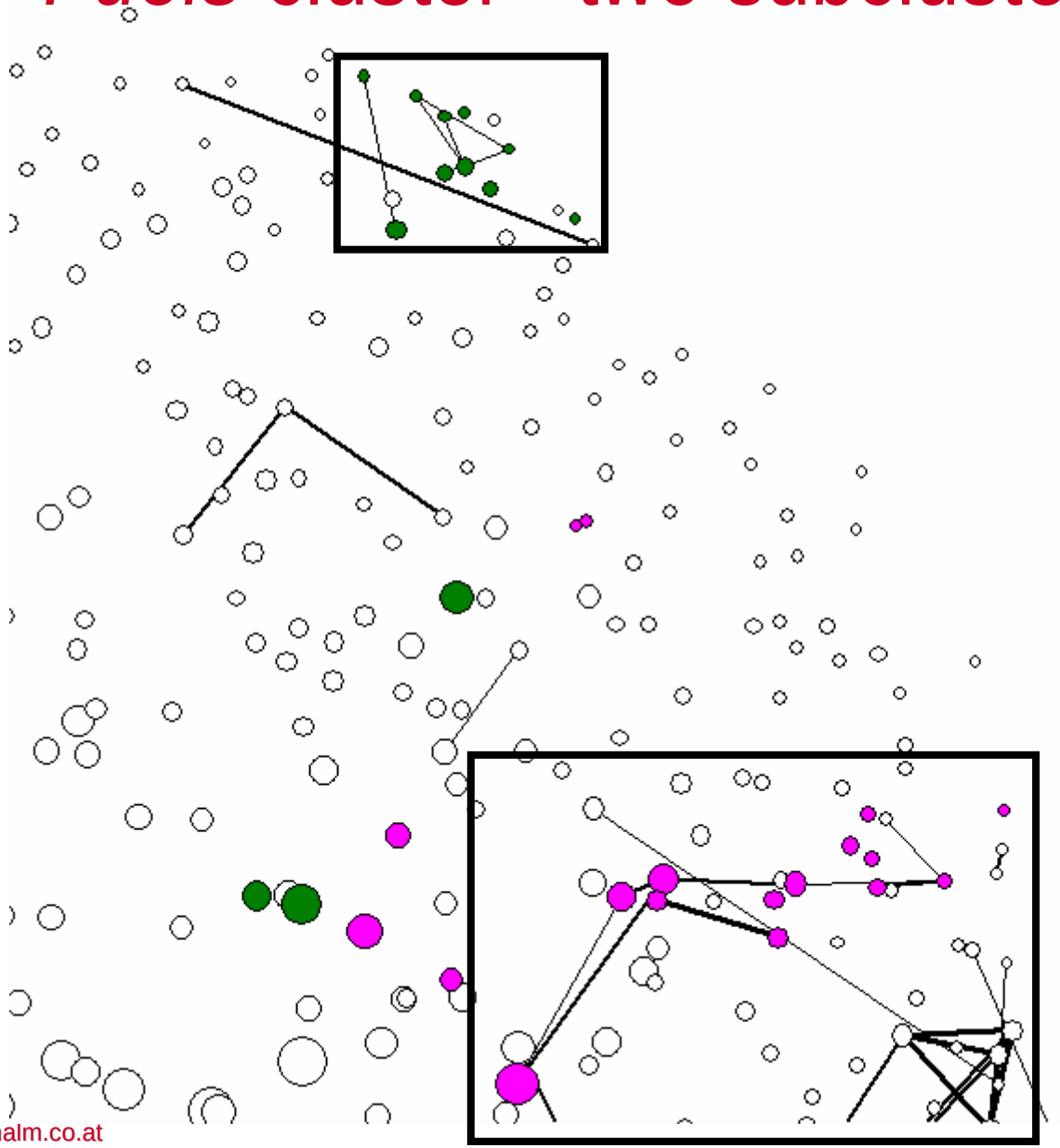
Network of keywords



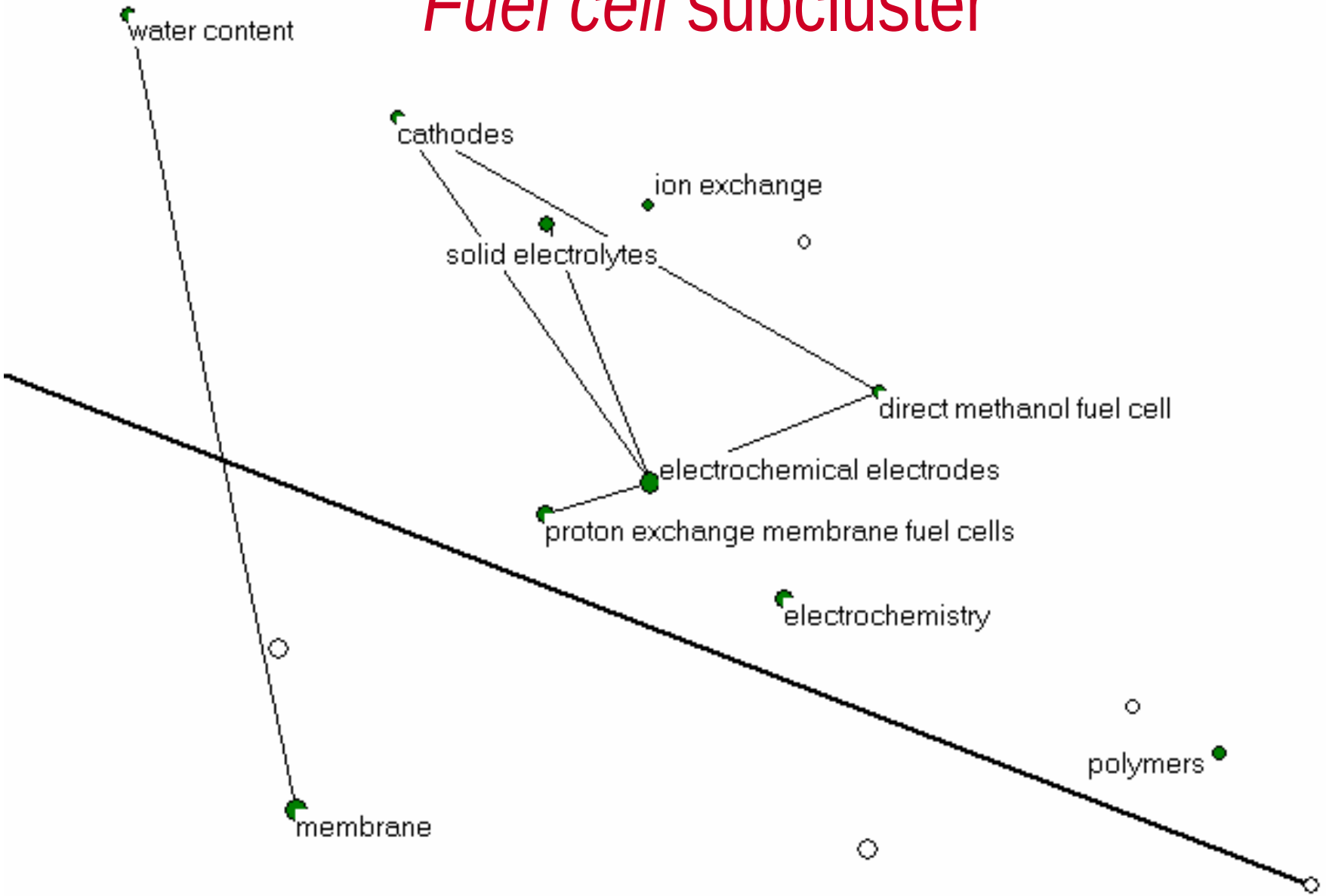
name Clusters	elem.	color
air traffic / navigation	31	
eu	52	
fuels	80	
networks/telecommunication	164	
sustainable development	**99**	
traffic / transport / urban planning	120	
traffic management/controll	270	
transport modes	123	
Industry	88	
safety	64	

clemens@widhalm.co.at

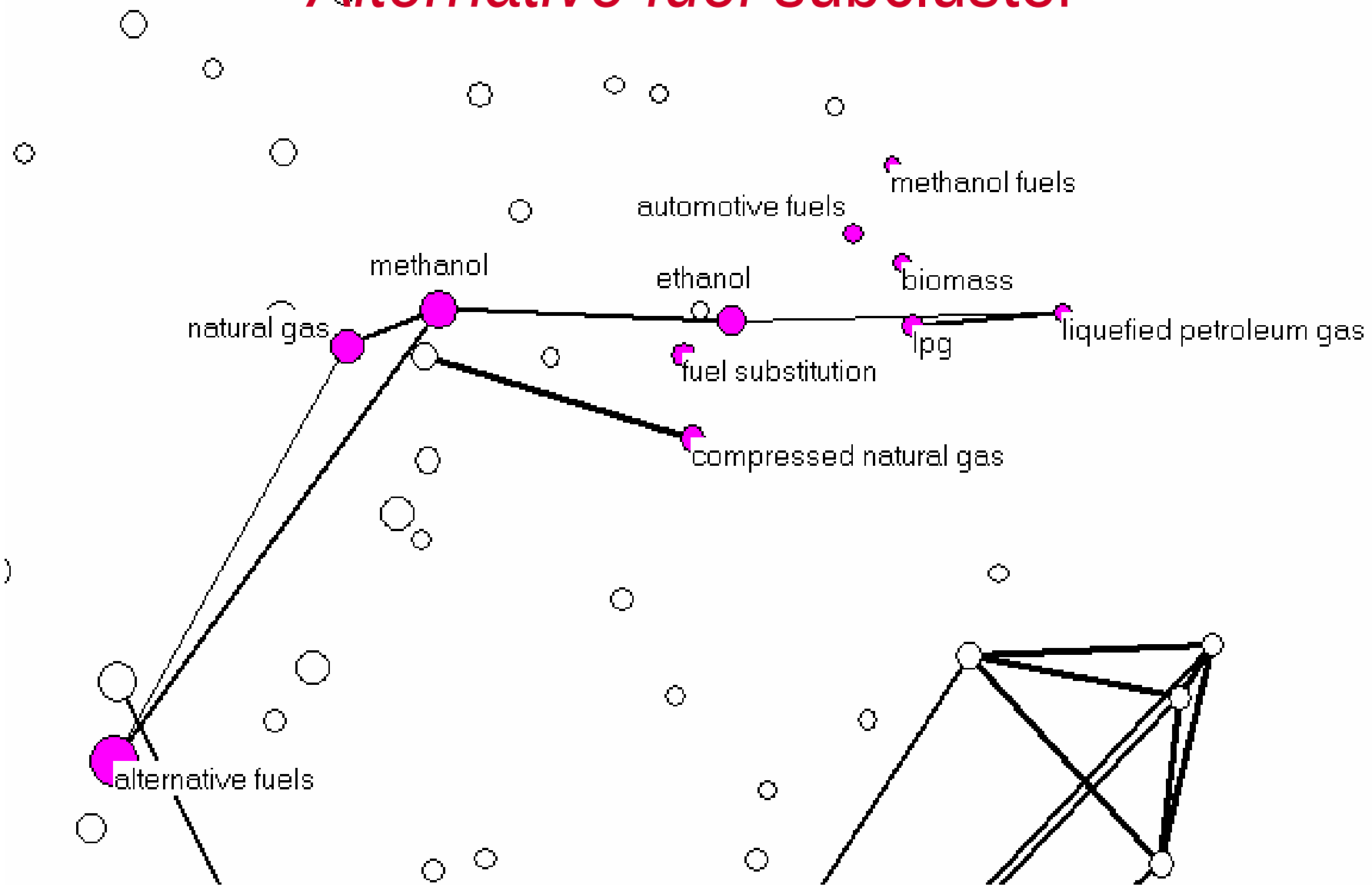
Fuels cluster - two subclusters



Fuel cell subcluster



Alternative fuel subcluster



clemens@widhalm.co.at

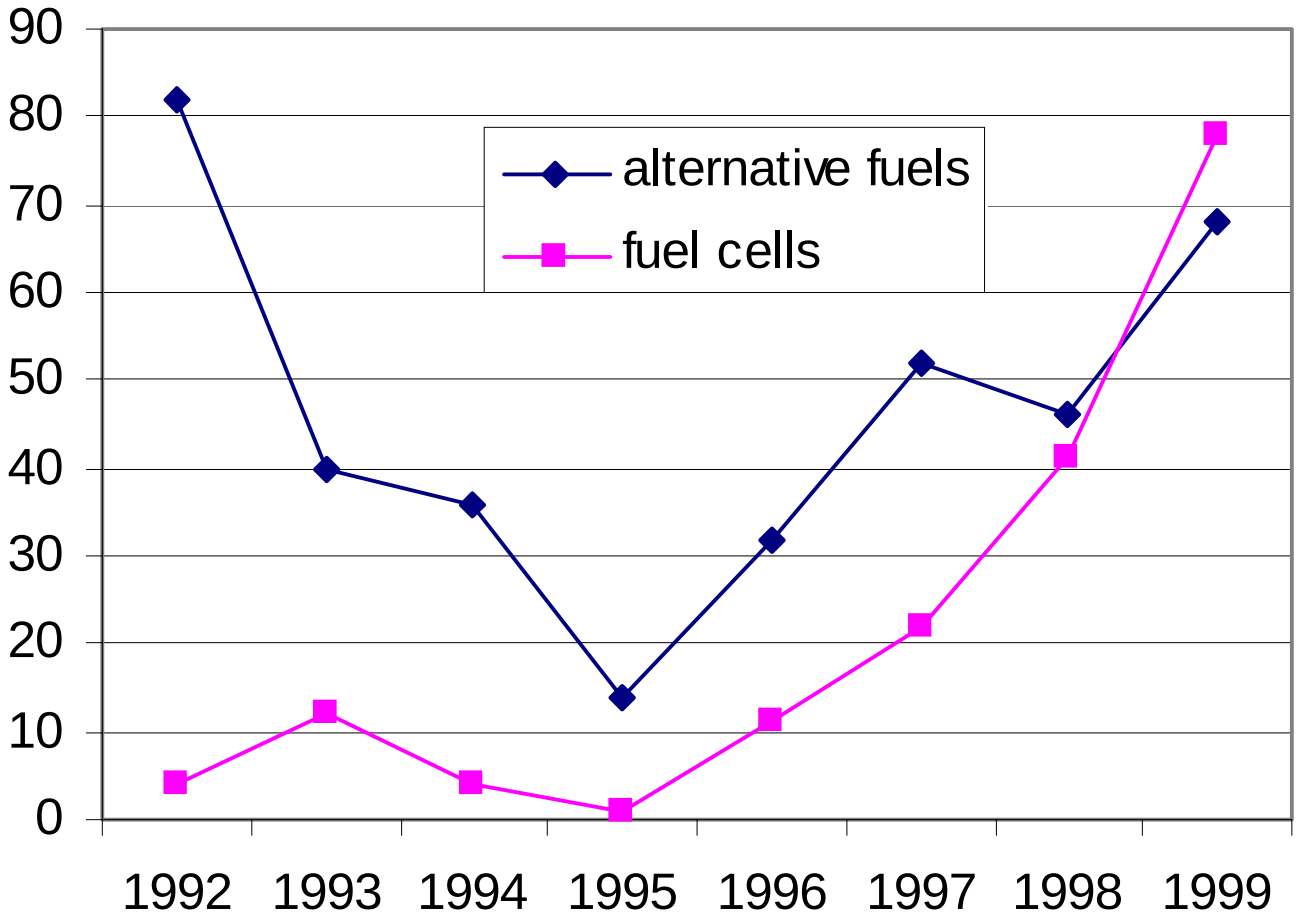
What did we find?

Clusters seem to have a substructure.

If we want to know the reason for that we might ask:

- Is there a different development of these subclusters?
- Are different keyplayers responsible for these subclusters?

Development of *alternative fuels* and *fuel cells*



Who are the keyplayers behind these topics?

High Ranking Keyplayers

(# of keywords of respective subcluster used by them)

alternative fuels:

192	USA	department of energy, washington, dc.
82	USA	iowa state university, ames, midwest transportation center
44	USA	california state dept. of transportation, sacramento.
36	USA	booz-allen and hamilton, inc., usa
36	USA	federal transit administration, washington, dc.
14	USA	transportation research board, nw , washington, dc, usa
14	EU	swedish transport and communications research board, stockholm (sweden)

fuel cells:

15	USA	department of energy, washington, dc.
11	EU	dept. of chem. & process eng., newcastle upon tyne univ., uk
6	USA	los alamos nat. lab., nm, usa
6	USA	arthur d. little inc., cambridge, ma, usa
5	Asia	[illegible]
5	EU	forschungszentrum julich gmbh, germany
5	USA	iowa state university, ames, midwest transportation center
4	EU	kommunikationsforskningsberedningen, stockholm, sweden

Which regions are responsible for that?

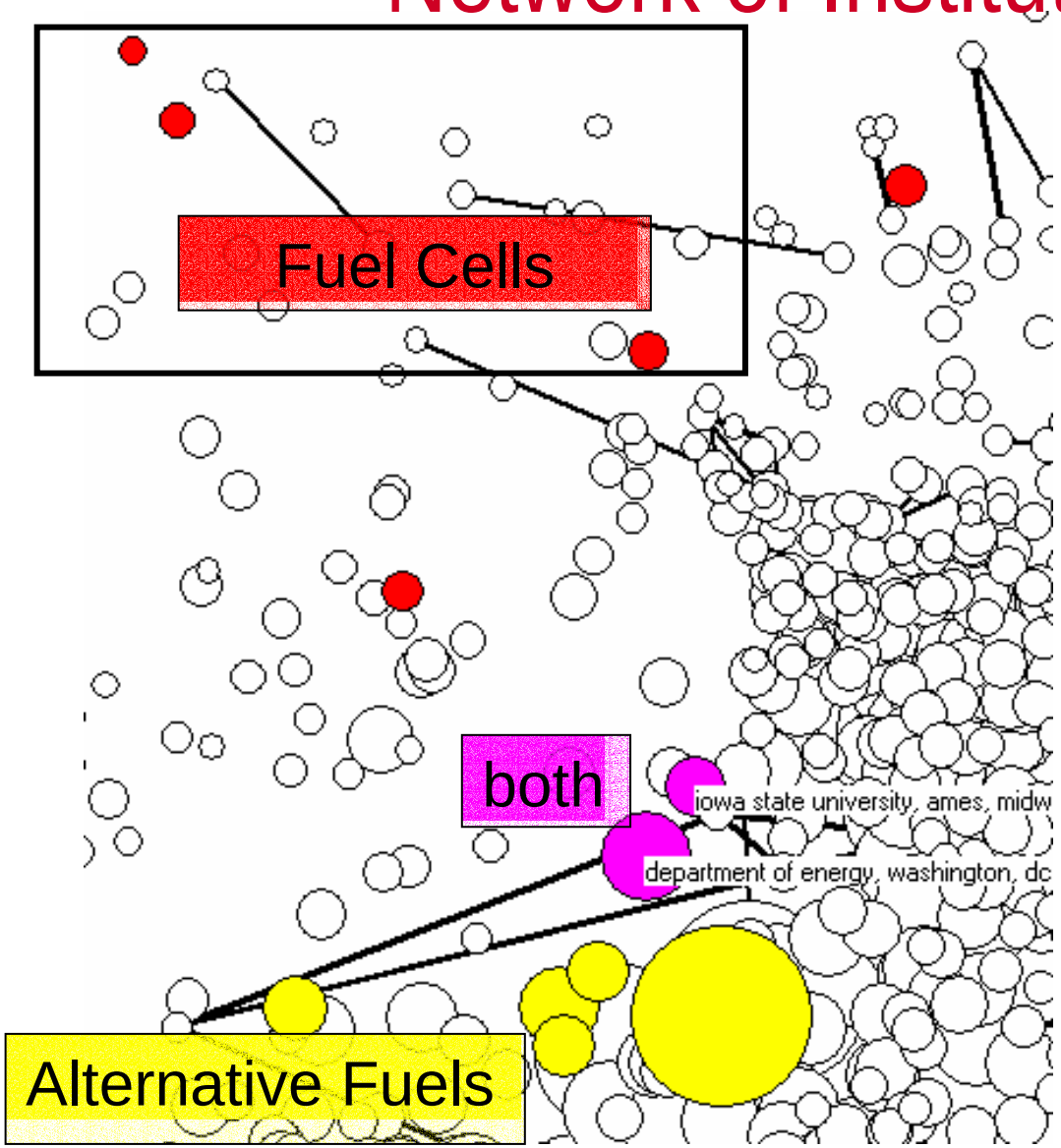
No. of Keywords of Subclusters used by Institutions of Regions

	alternative fuels	*fuel cells*
EU	114	45
USA	470	61

- USA clearly dominates *alternative fuel* topic
- EU comes along in *fuel cells*

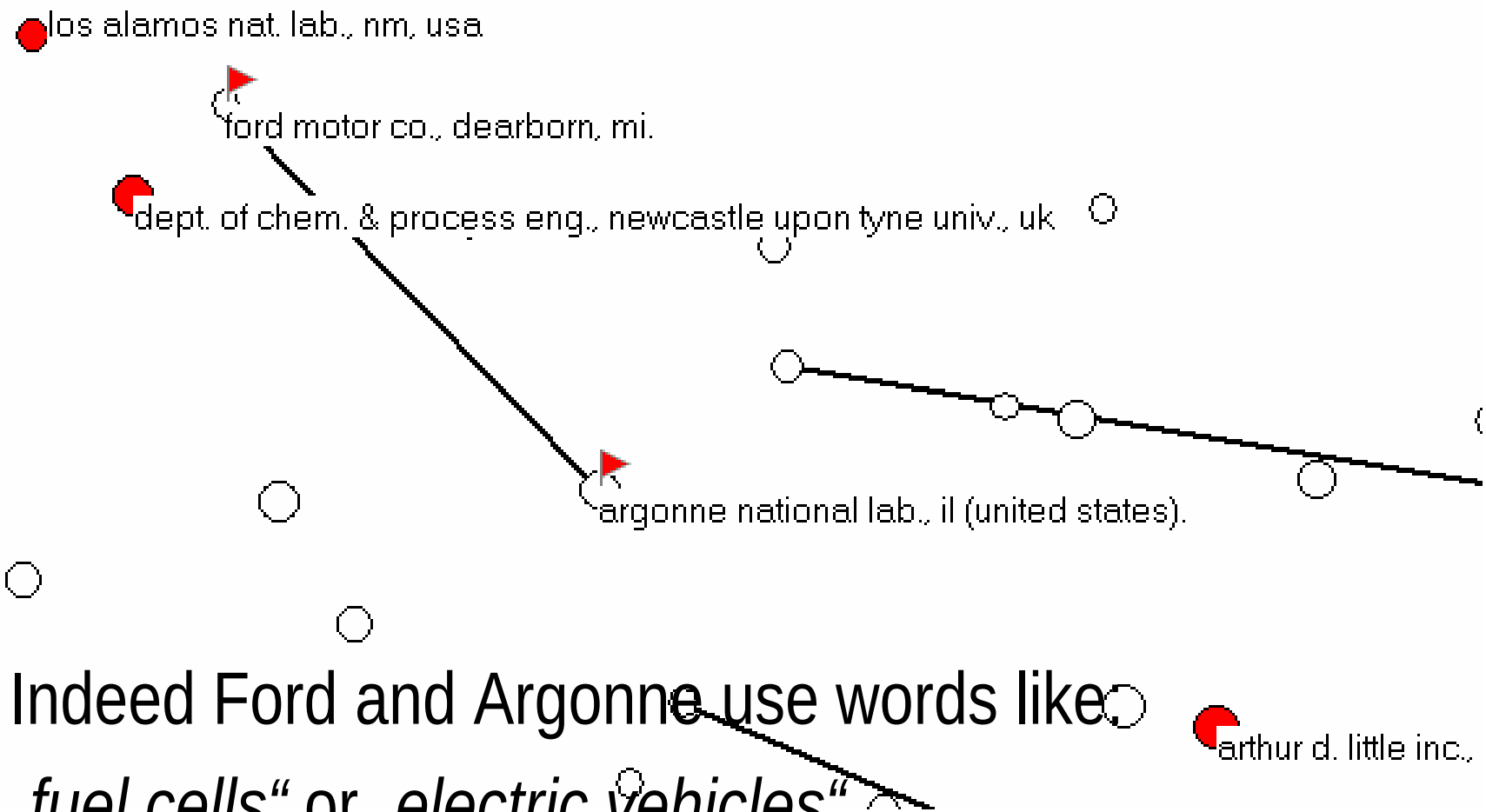
Where are the keyplayers located on a Knowledge Map?

Network of Institutions



Can we find more actors relevant for fuel cells?

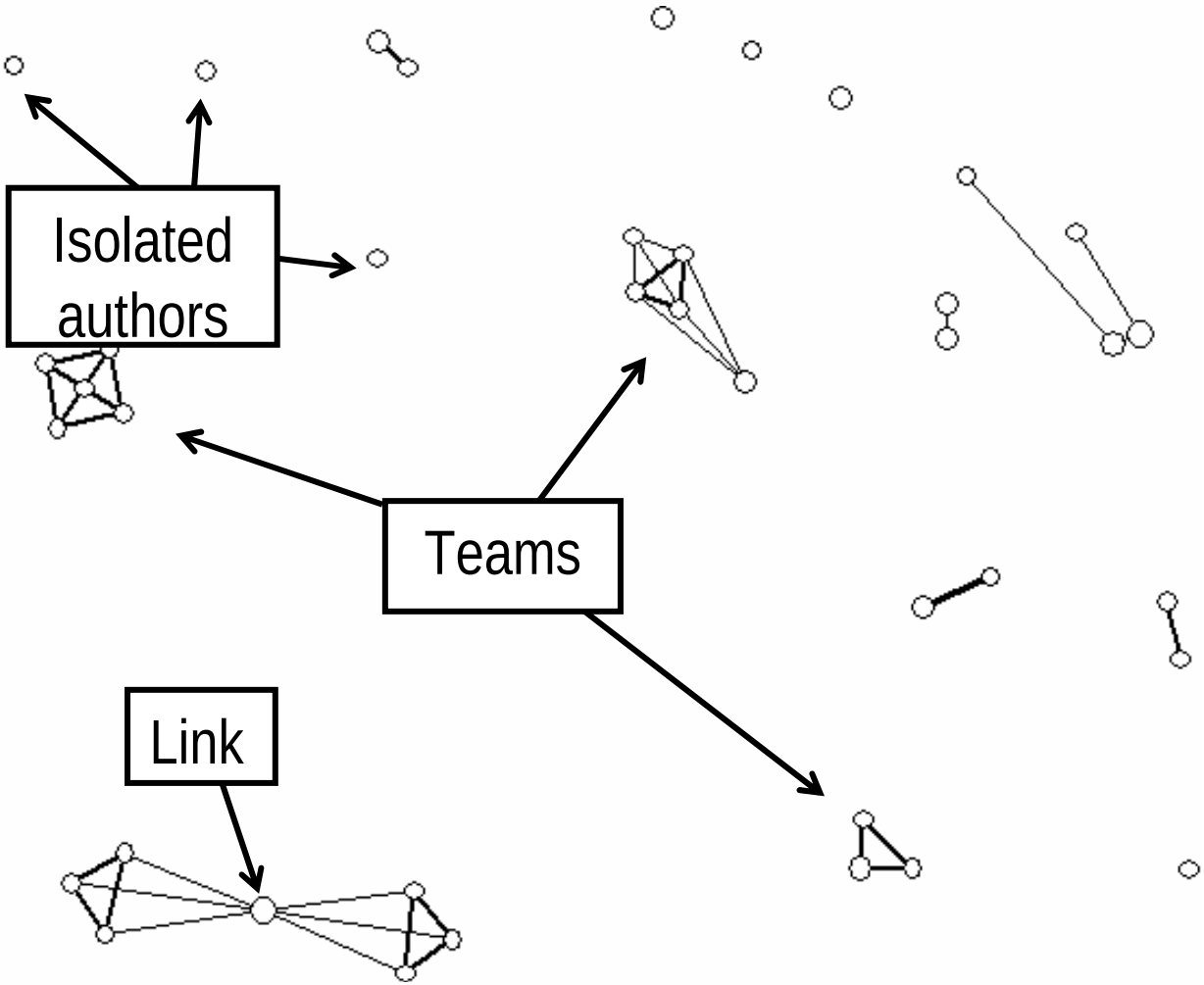
Fuel Cells Actors



Indeed Ford and Argonne use words like

„*fuel cells*“ or „*electric vehicles*“

Network of Authors



Conclusions

- For a comprehensive analysis of a research topic different aspects of documents should be considered: **What, When, Who** and **Where**
- **Relations** between respective objects can be calculated by **co-occurrence** analysis
- Easy interpretation even of huge respective matrices can be supported by **visualisation** in knowledge **maps**