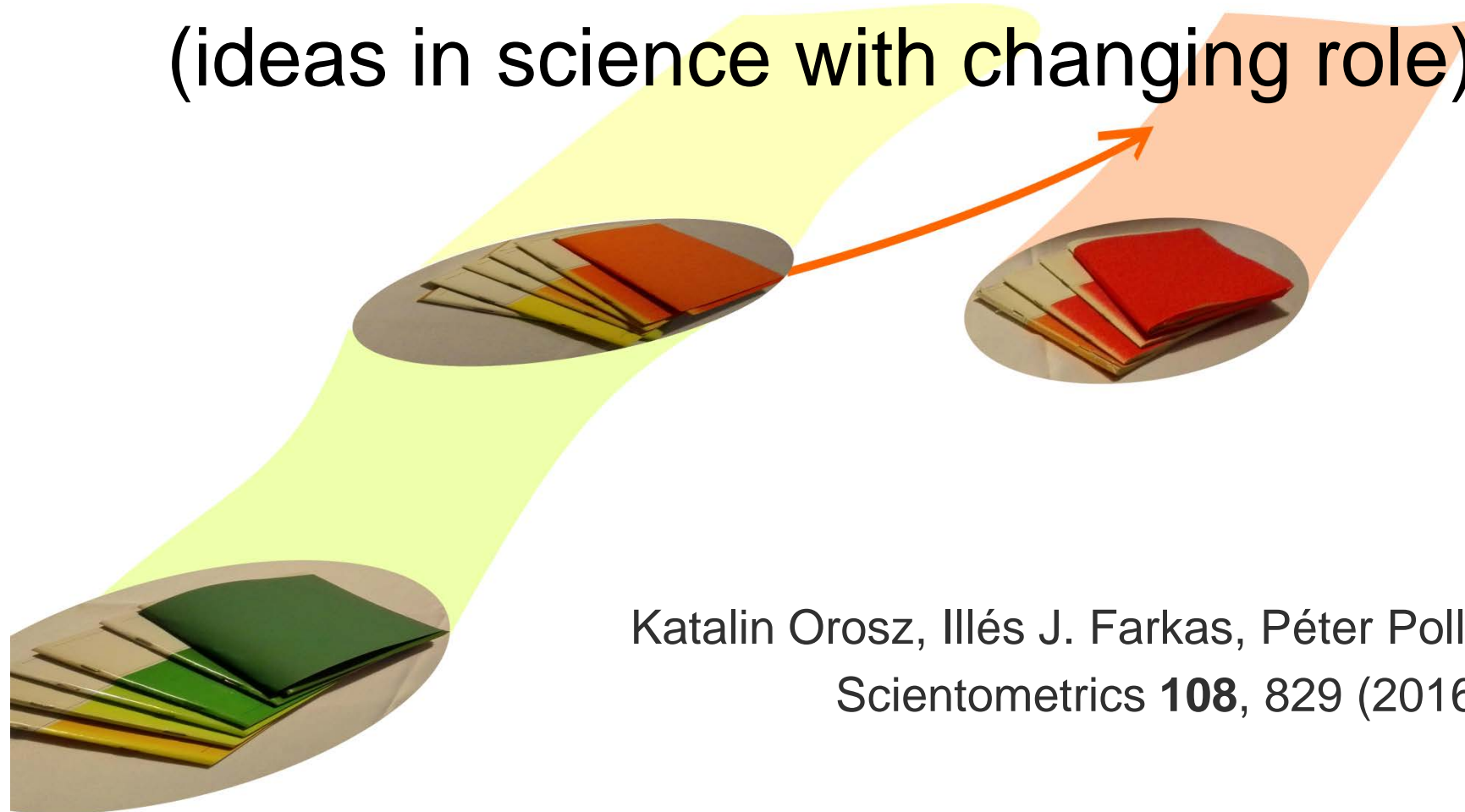




Quantifying the changing role of past publications

(ideas in science with changing role)



Katalin Orosz, Illés J. Farkas, Péter Pollner
Scientometrics **108**, 829 (2016)

Identification, location and temporal evolution of topics

Data and algorithm – comparison of approaches

KNOWeSCAPE, Budapest, 2016.08.30.

Motivation

“verba volant, scripta manent”
(words fly, writings remain)

Is it possible to change the topic of a published written material?

Motivation

“verba volant, scripta manent”
(words fly, writings remain)

Is it possible to change the topic of a published written material?

When we read it with a different point of view ...



(My wife and my mother-in-law, W.E.Hill, 1915)

Motivation

“verba volant, scripta manent”
(words fly, writings remain)

Is it possible to change the topic of a published written material?

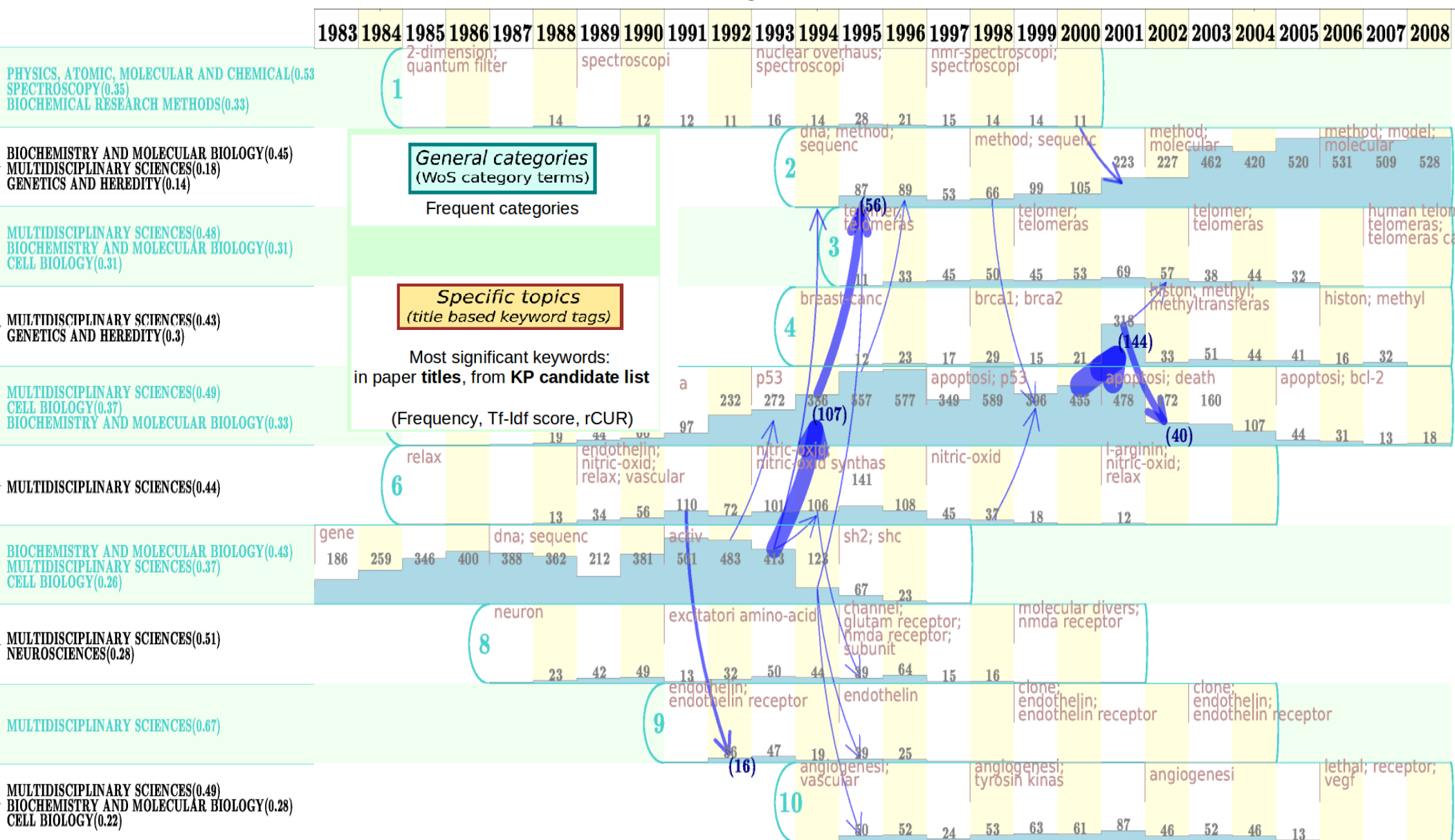
When we read it with a different point of view ...



(My wife and my mother-in-law, W.E.Hill, 1915)

How can we track and quantify these changes?

Map of groups



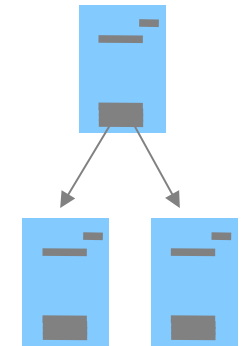
Changing and interacting groups of publications

Changing co-citation core – changing context of usage

Motivation

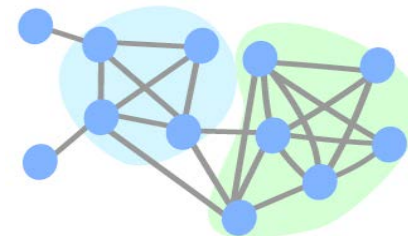
1. *Need a collection of large number of documents
(document = publish information for long lasting use)*

→ **PUBLICATION DATABASE WITH CITATION DATA**
(no wikipedia, no law collection)



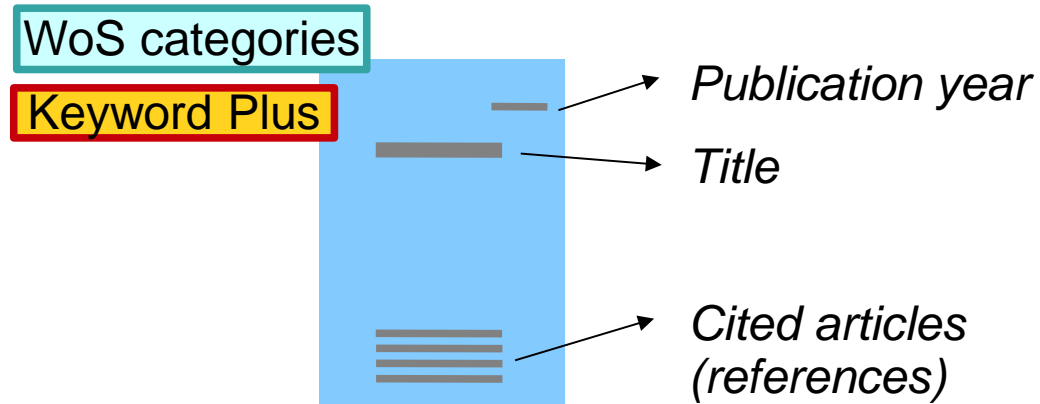
2. *Need to follow topic (actual information) evolution in time
(actual information = context of the document)*

→ **NETWORKS**



Web of Science data and co-citations

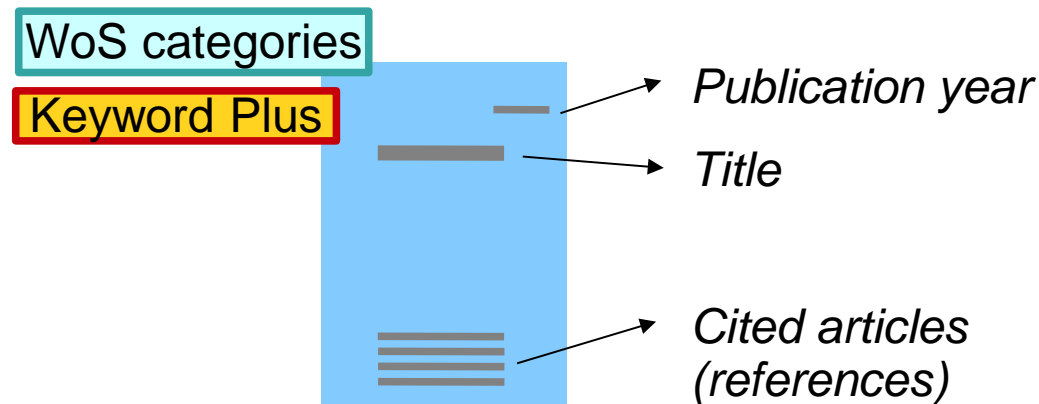
• **Data:**



• **Time dependent context?**

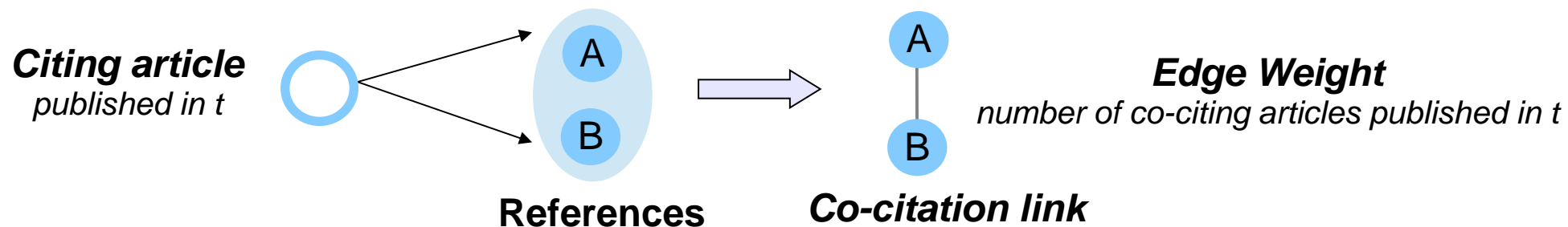
Web of Science data and co-citations

• Data:



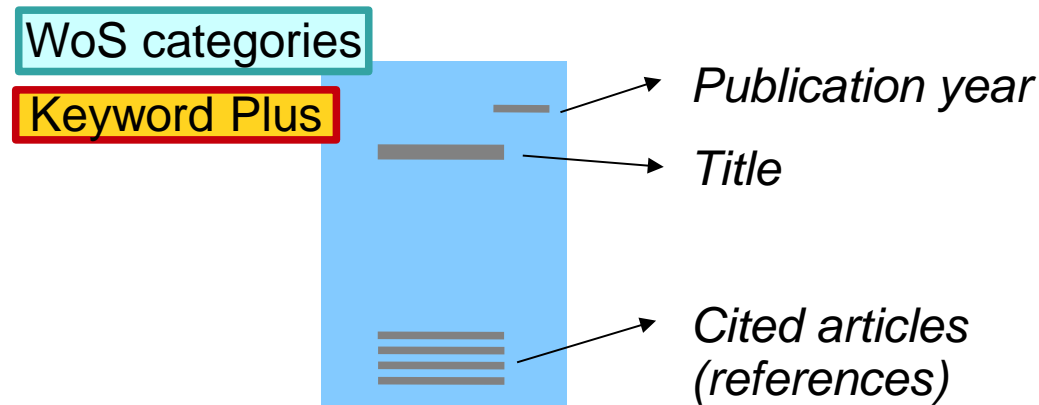
• **Co-citation link** between articles *A* and *B* in year *t* if a citing article published in *t* has both *A* and *B* in its reference list

Citing years: 1975 - 2008



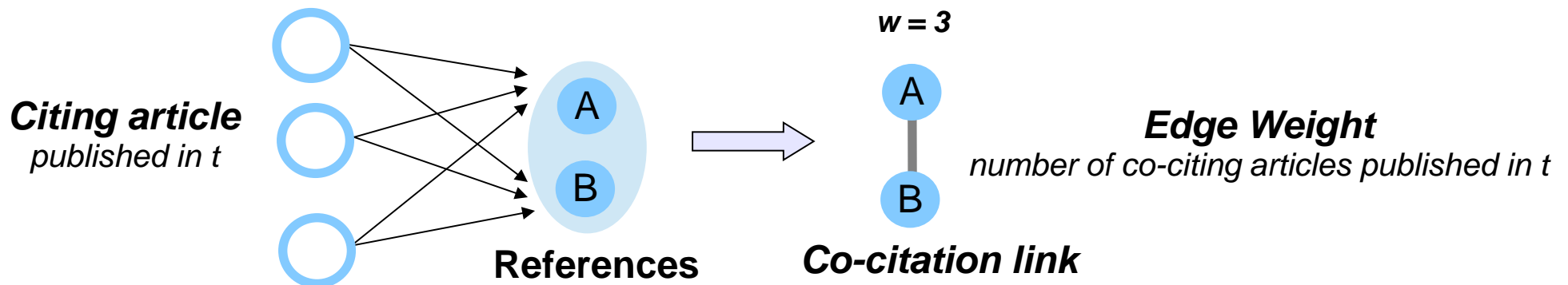
Web of Science data and co-citations

• Data:



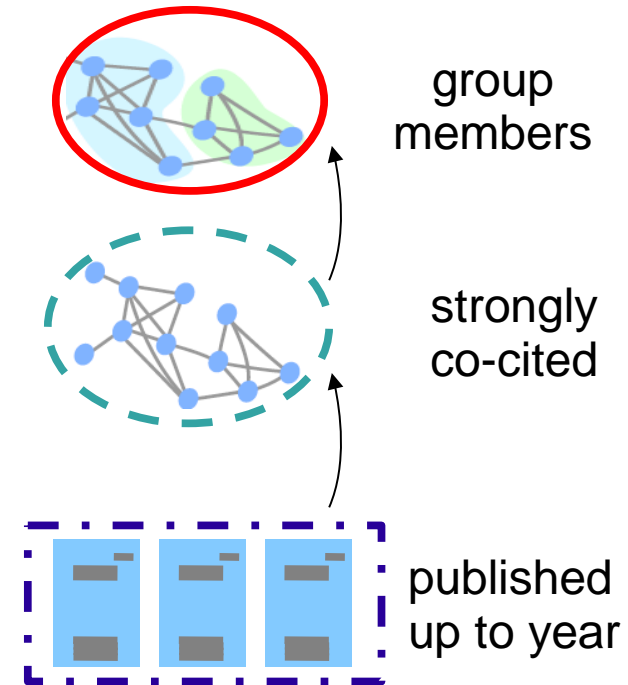
• **Co-citation link** between articles *A* and *B* in year *t* if a citing article published in *t* has both *A* and *B* in its reference list

Citing years: 1975 - 2008



Constructing time evolving groups

- Focusing on the “core” of science,*
- Topic (context) described by tightly cooperating articles,*
- Allow several topics for any article*
- Consequent coupling of groups between time steps*



Constructing time evolving groups

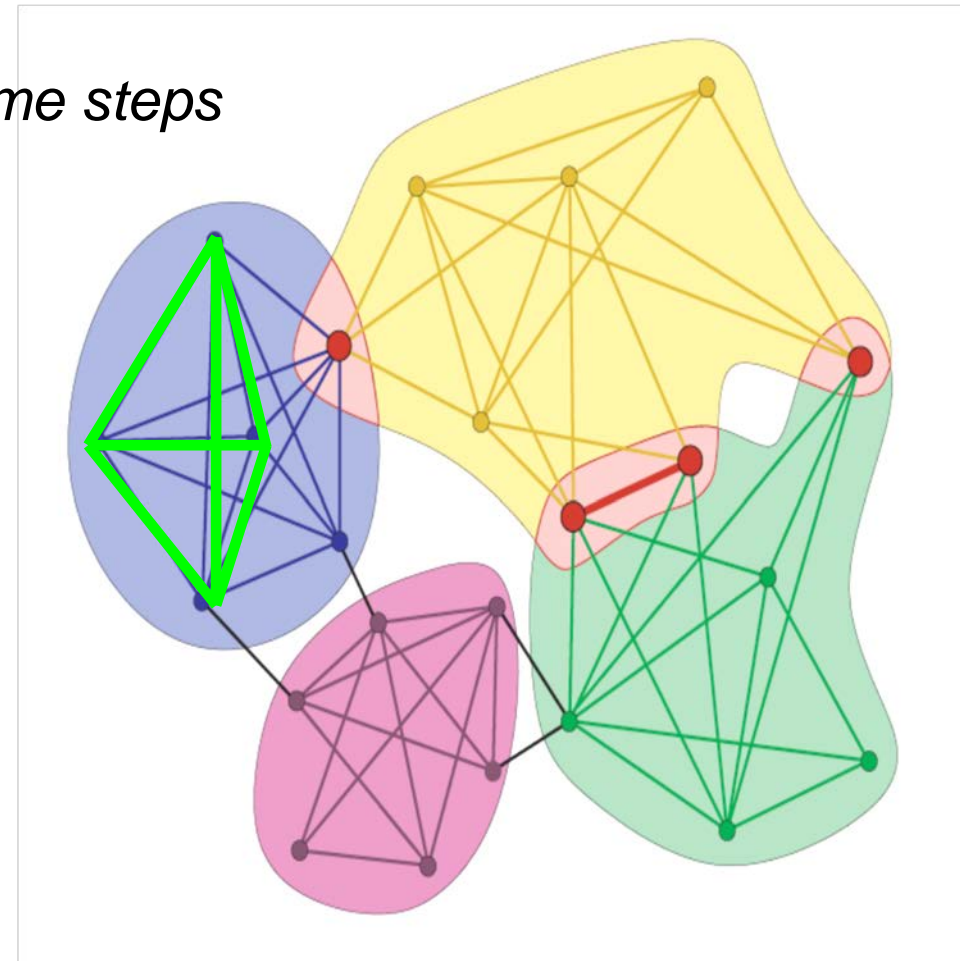
Focusing on the “core” of science,

Topic (context) described by tightly cooperating articles,

Allow several topics for any article

Consequent coupling of groups between time steps

→ *tCFinder algorithm*



Constructing time evolving groups

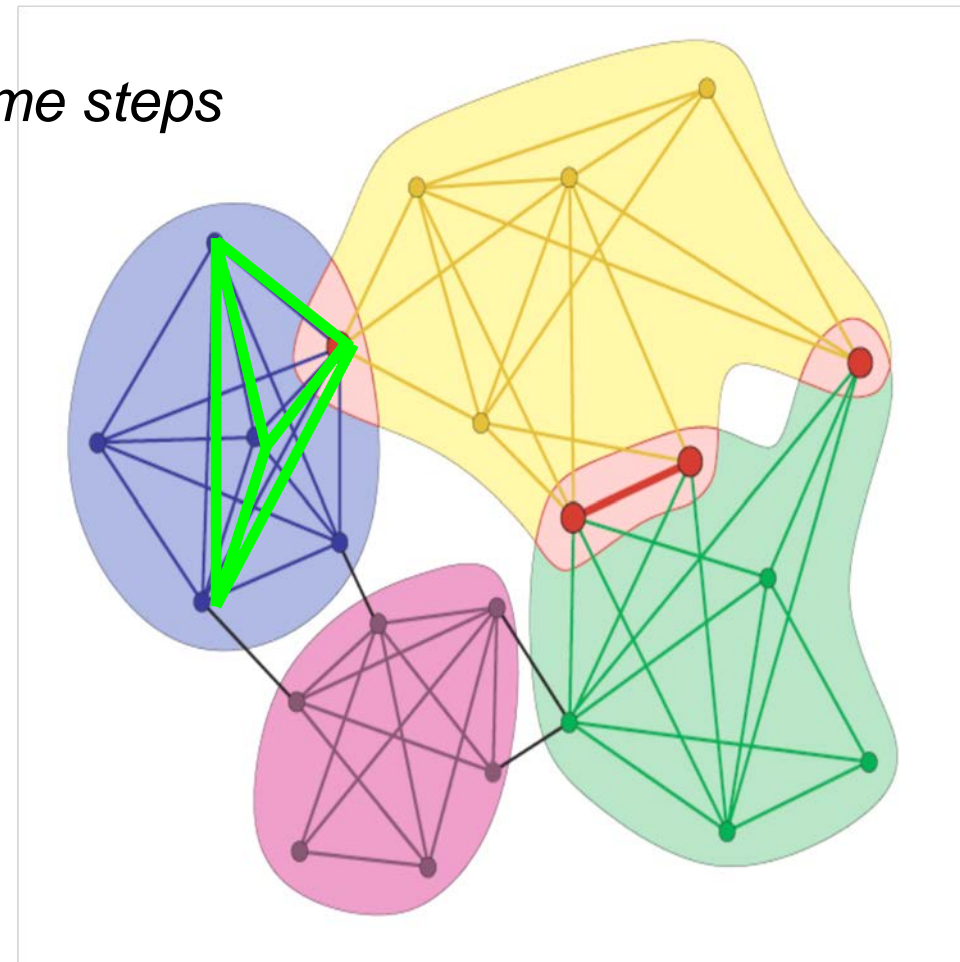
Focusing on the “core” of science,

Topic (context) described by tightly cooperating articles,

Allow several topics for any article

Consequent coupling of groups between time steps

→ *tCFinder algorithm*



Constructing time evolving groups

Focusing on the “core” of science,

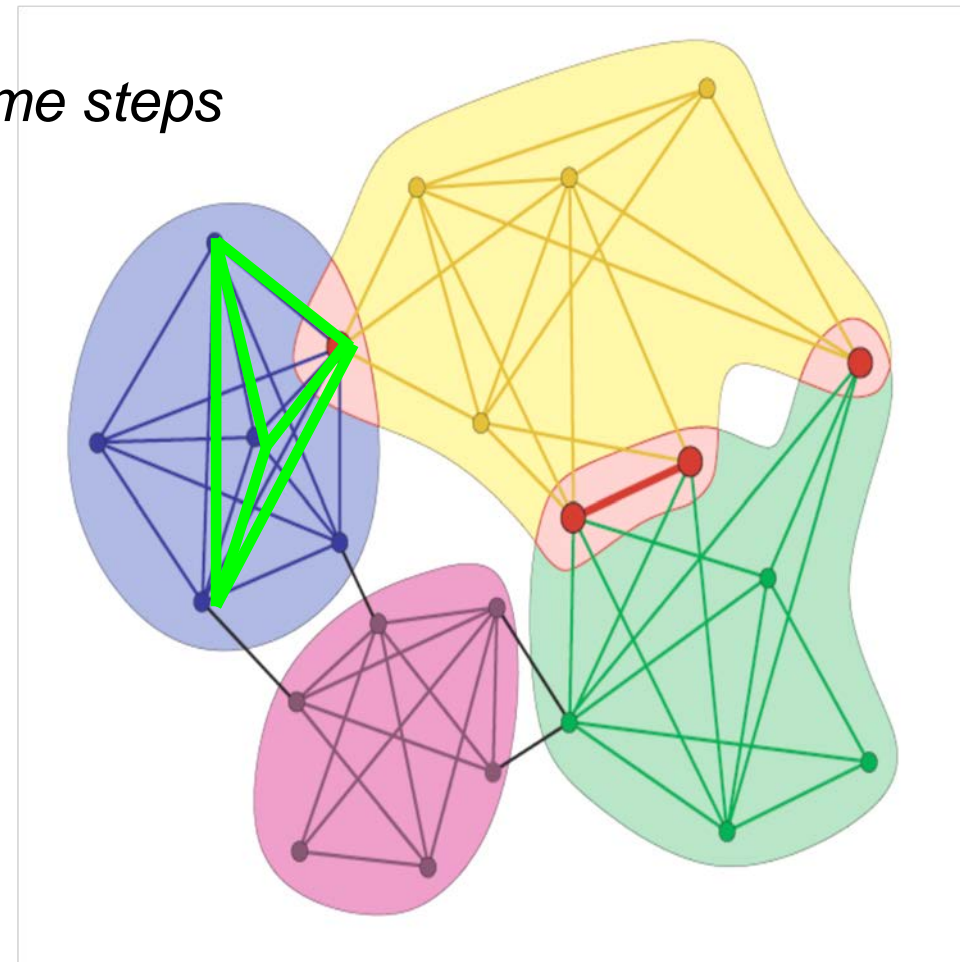
Topic (context) described by tightly cooperating articles,

Allow several topics for any article

Consequent coupling of groups between time steps

→ *tCFinder algorithm*

t and t+1 networks merged:
- Coupling of groups



Unsupervised topic selection

Selecting group labels (*topic description of articles in a group*):

Identify field of science:

WOS categories of journals *where articles appeared*
(*or a weighted combination if more categories*)

Describe time evolution of topics:

Set of **Keywords, Stemmed words of titles**

Selecting the most relevant labels (too many words in the above set):

- **TF-IDF** score (*top 10*)
- **CUR** selection (*top 3*)

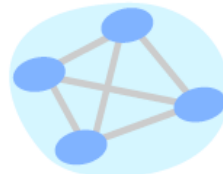
Features of groups and their states

State: the group at a certain time

For each state:

Size

Number of papers



General categories
(WoS category terms)

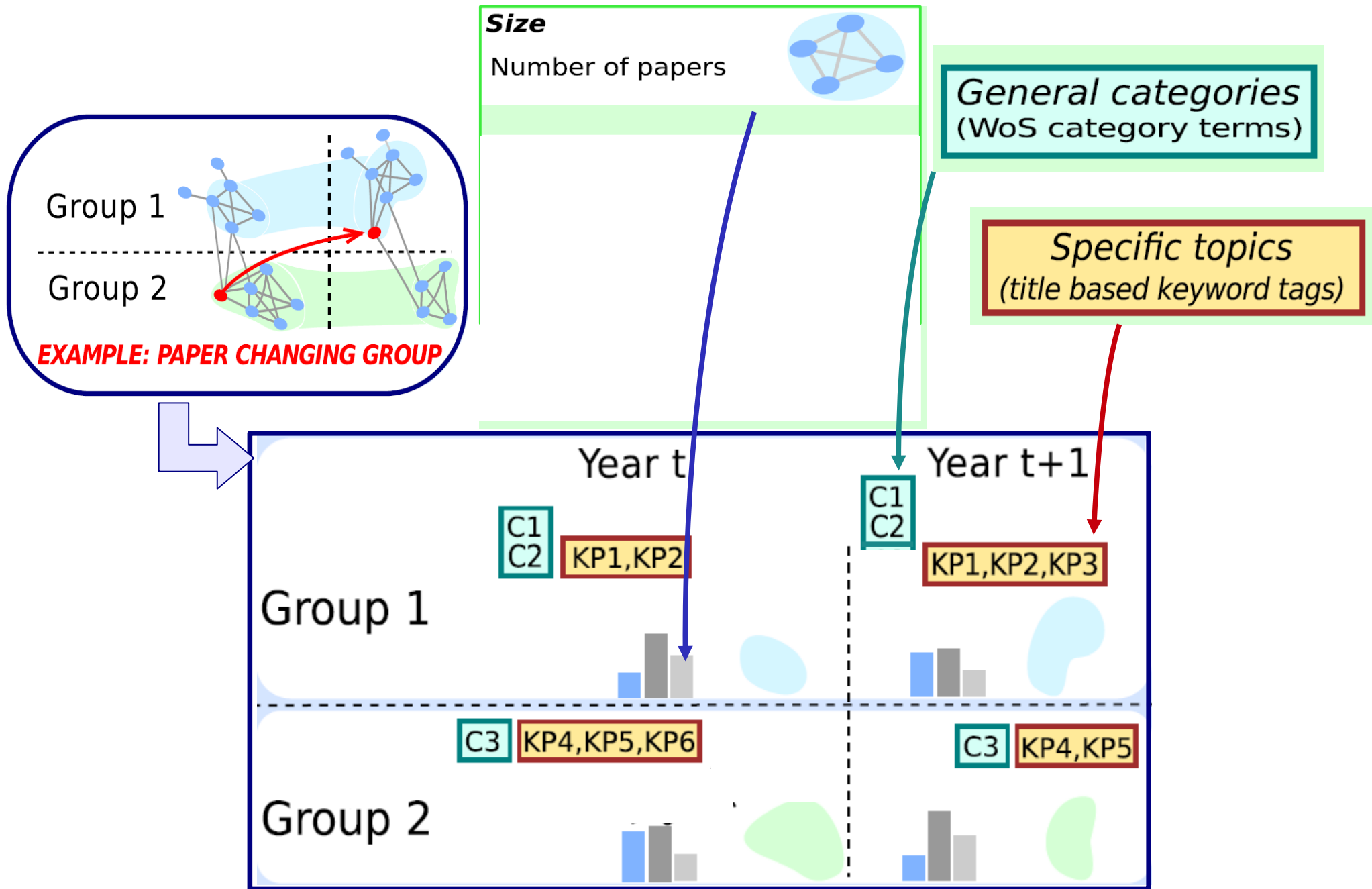
Frequent categories

Specific topics
(title based keyword tags)

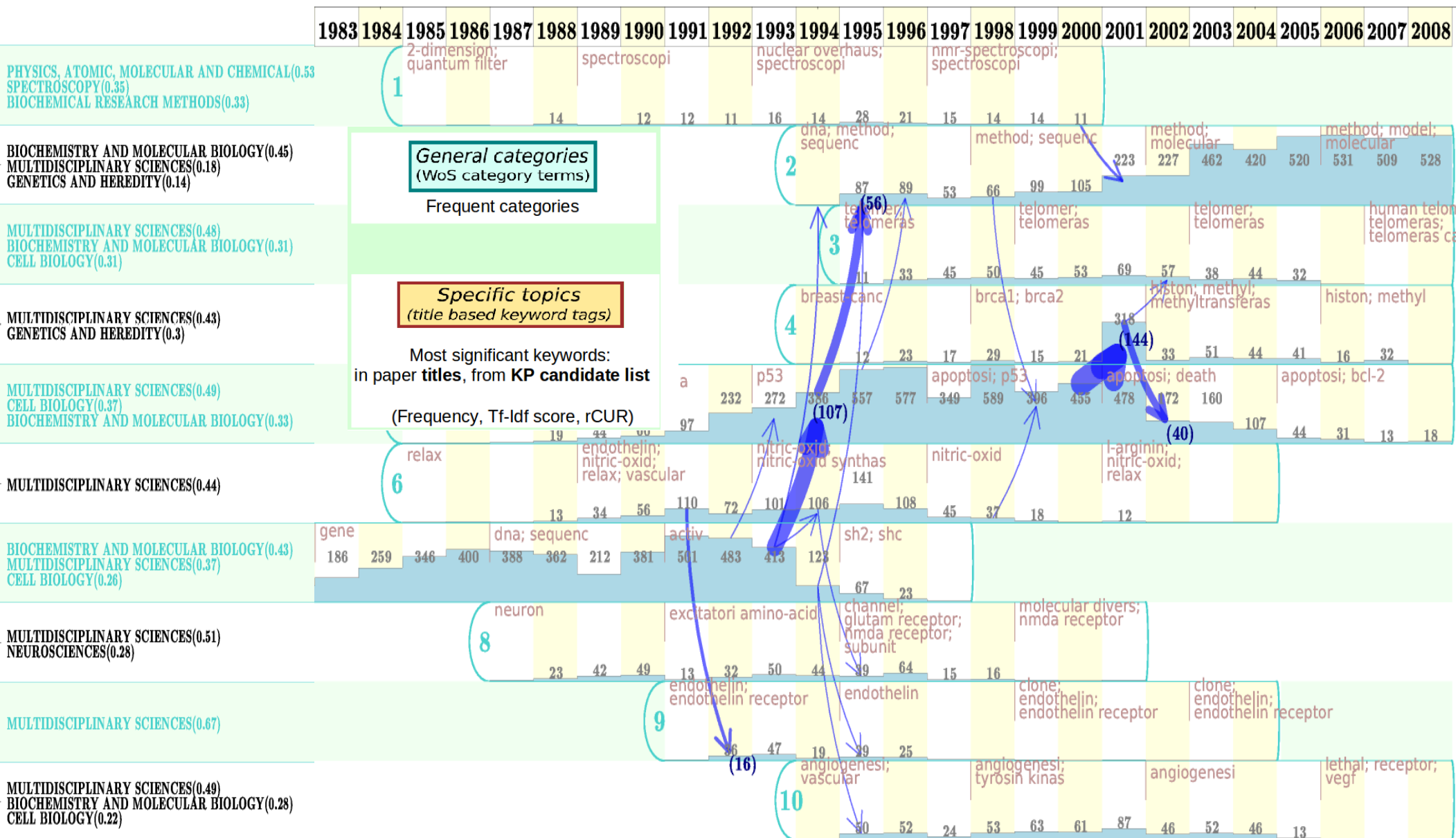
Most significant keywords:
in paper **titles**, from **KP candidate list**

(Frequency, Tf-Idf score, rCUR)

Timeline of groups



Map of groups



Changing and interacting groups of publications

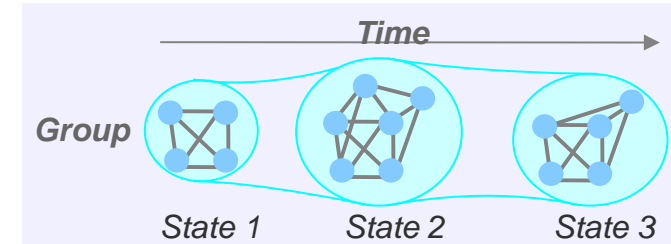
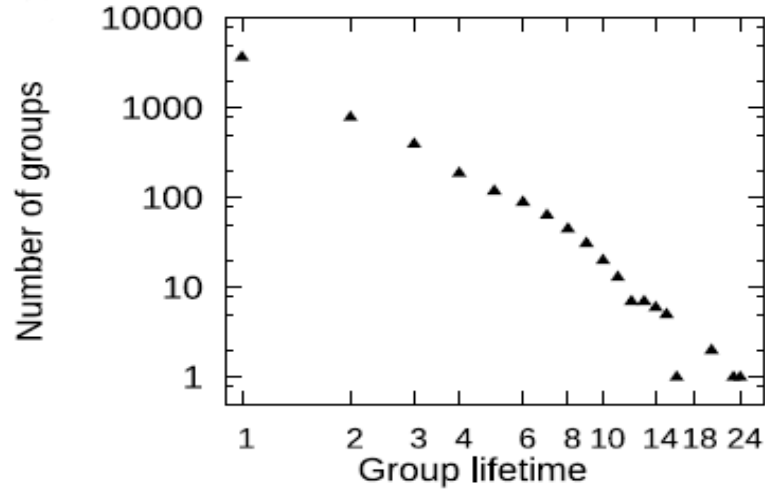
Changing co-citation core – changing context of usage

Results:

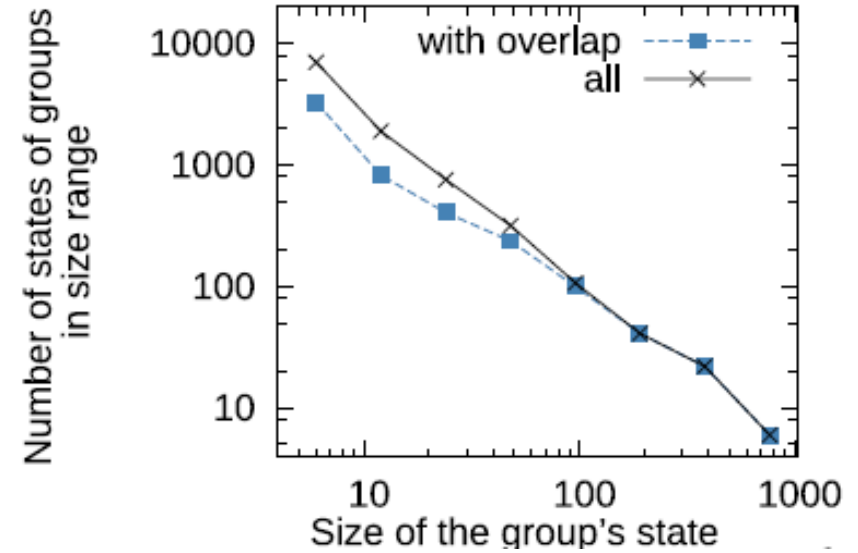
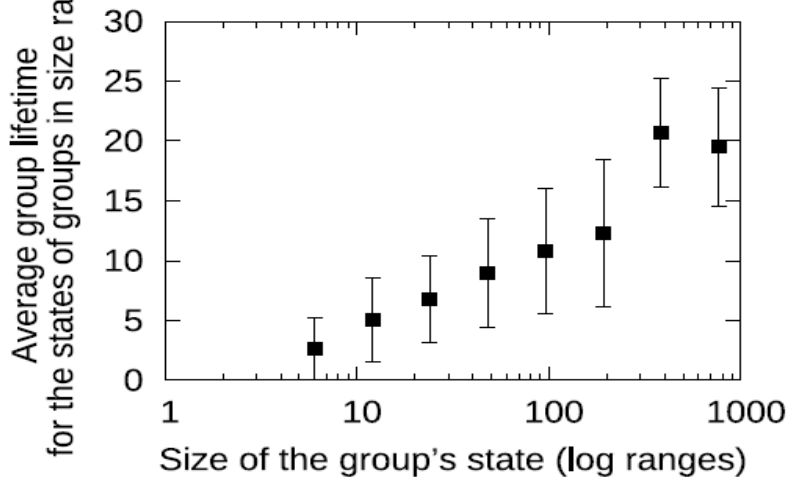
Group statistics: lifetime, size, overlap

5439 groups made up of **10160 states** (48% has overlapping article)

Mostly short group lifetimes, a few long ones



Larger groups live longer



many small states
Overlap size is small

MTA-ELTE Statistical and Biological Research Group
and ERC Advanced COLLMOT Group ELTE

Thank you for your attention!

Pollner
Péter

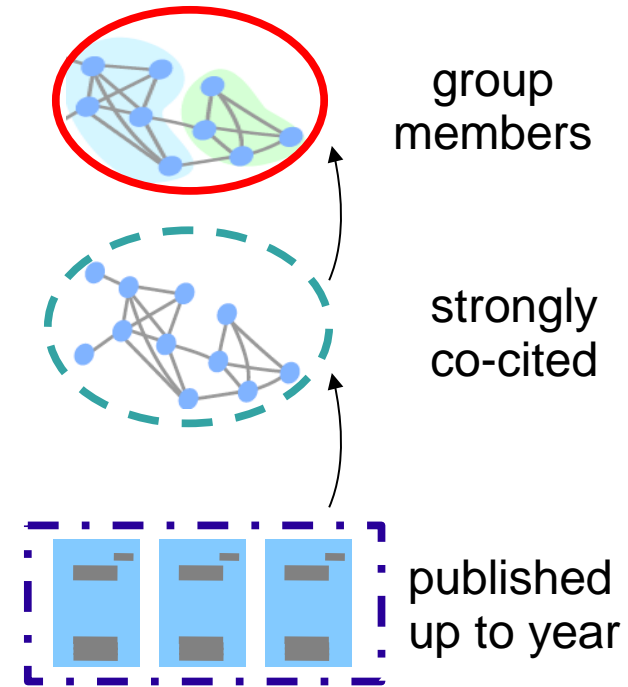
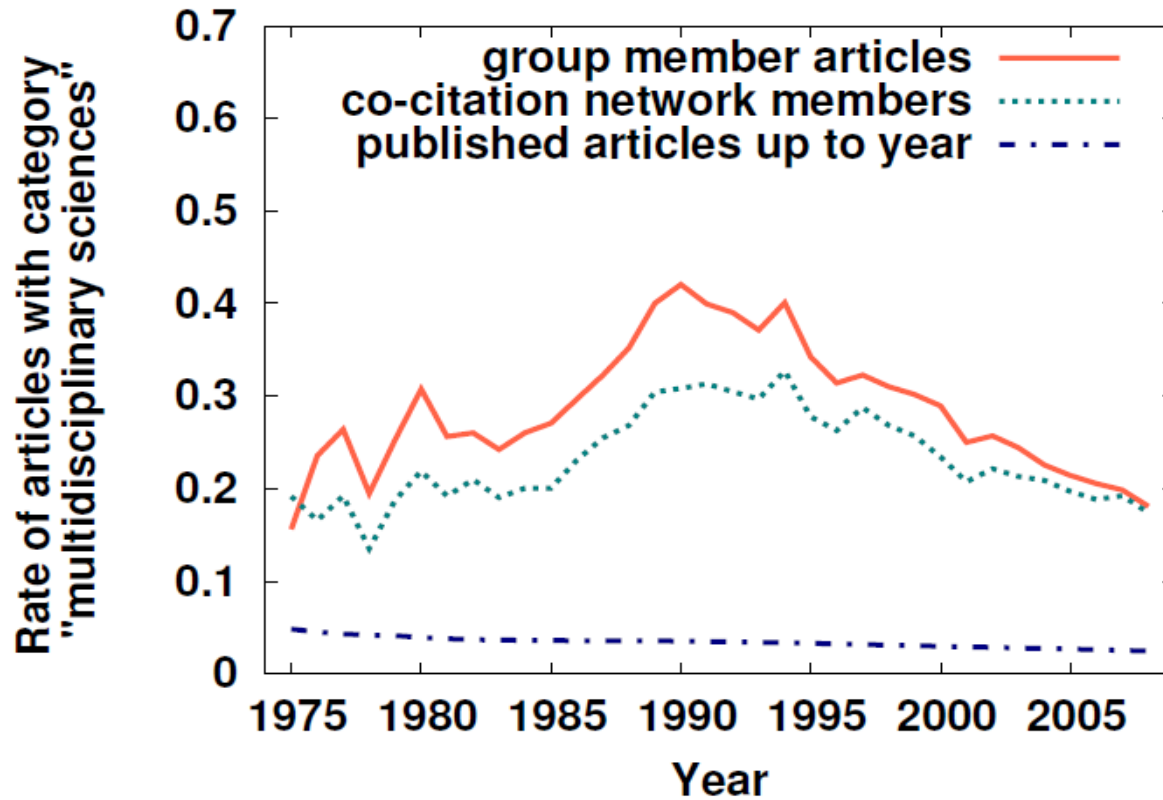
Farkas
Illés



Katalin
Orosz

Vicsek Tamás

Multidisciplinary



Enrichment of articles from multidisciplinary journal in the co-citation networks

Further enrichment in the groups