

# TRENDS IN EDUCATIONAL RESEARCH:

## SEMANTIC ANALYSIS OF PAPER ABSTRACTS

Stefan Trausan-Matu\*, Philippe Dessus\*\*,  
Cristian Ionut Paraschiv\*, & Mihai Dascalu\*

\*RACAI, “Politehnica” Univ. of Bucharest, Romania

\*\*LSE, Univ. Grenoble Alpes, France

# Outline

1. Considerations
2. Experiment
3. Further work

Refs available at

<http://www.citeulike.org/user/pdessus/tag/educmap>



# 1. CONSIDERATIONS

Pros & Cons of Co-citation/biblio. Coupling-based  
Analyses

vs.

Text-based Analysis

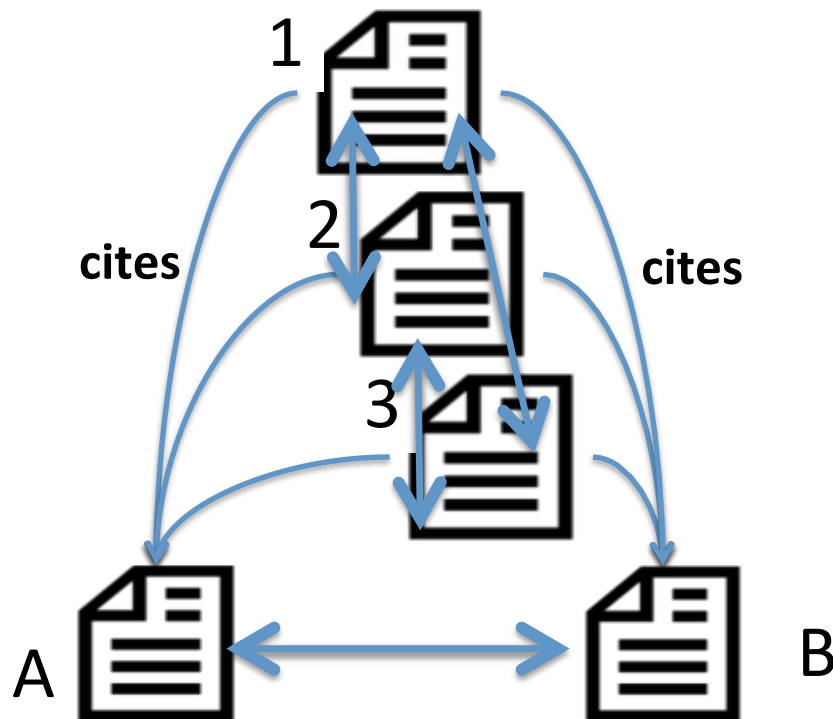
# <sup>4</sup>Co-citation Analysis, Biblio. Coupling and Semantic Comparison

- How to measure the relatedness between papers?

Bet. A & B: co-citation analysis

Bet. 1-3: biblio. coupling analysis

Semantic comparison of  
papers' abstracts



# 5 Co-citation/Biblio. Coupling Analyses

## Pros

## Cons

Co-occurrence-based and bibliography-centered, thus fast to process with very large sets of papers

Citation strategies are influenced by researchers' practices [Sword 12] and motivations [Ding et al. 14]

- Citation number differ across domains
- Influenced by authors' impact factor increasing endeavors
- "Homage" effect may entail overcitation (a.k.a., "preferential attachment")

Bib. coupling captures collaboration & communities  
=>network extraction easier

Assume all references are of similar value [Ding et al. 14], e.g., don't address the number of times a ref. is cited in a paper [Wan & Liu 14], or how it is cited (where in the paper)

Two citations of a given paper may refer to different aspects of the cited paper

# Text-based Approach

## Pros

Content-based, thus more accurate than Co-Citation Analysis to assess paper's similarity, at least to measure cohesion, topic similarity

Reduced variability of summaries content and size across domains

Co-cit.: Texts already there (citations come over time)

## Cons

Semantic analysis is more computationally demanding: fewer data is computable in one pass

Network extraction is more difficult; no obvious relationship with impact factor measures

Semantic similarity bet. papers doesn't necessarily mean authors' "community of thought" , rather replicating, copying



## 2. EXPERIMENT

Use the semantic analysis (LSA/LDA/WordNet) of paper summaries to compare them directly (instead of using references as a proxy).

# LSA Example [Landauer & Dumais 97]

- §1: Human machine interface for ABC computer applications  
 §2: A survey of user opinion of computer system response time  
 §3: The EPS user interface management system  
 §4: System and human system engineering testing of EPS  
 §5: Relation of user perceived response time to error measurement  
 §6: The generation of random, binary, ordered trees  
 §7: The intersection graph of paths in trees  
 §8: Graph minors IV: Widths of trees and well-quasi-ordering  
 §9: Graph minors: A survey

H-M  
Interface

Graphs

|           |   |   |   |   |   |   |   |   |   |      |      |      |      |     |      |      |      |      |
|-----------|---|---|---|---|---|---|---|---|---|------|------|------|------|-----|------|------|------|------|
| Human     | 1 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | .16  | .40  | .38  | .47  | .18 | -.05 | -.12 | -.16 | -.09 |
| interface | 1 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | .14  | .37  | .33  | .40  | .16 | -.03 | -.07 | -.10 | -.04 |
| computer  | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | .15  | .51  | .36  | .41  | .24 | .02  | .06  | .09  | .12  |
| user      | 0 | 1 | 1 | 0 | 1 | 0 | 0 | 0 | 0 | .26  | .84  | .61  | .70  | .39 | .03  | .08  | .12  | .19  |
| system    | 0 | 1 | 1 | 2 | 0 | 0 | 0 | 0 | 0 | .45  | 1.23 | 1.05 | 1.27 | .56 | -.07 | -.15 | -.21 | -.05 |
| response  | 0 | 1 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | .16  | .58  | .38  | .42  | .28 | .06  | .13  | .19  | .22  |
| time      | 0 | 1 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | .16  | .58  | .38  | .42  | .28 | .06  | .13  | .19  | .22  |
| EPS       | 0 | 0 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | .22  | .55  | .51  | .63  | .24 | -.07 | -.14 | -.20 | -.11 |
| survey    | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | .10  | .53  | .23  | .21  | .27 | .14  | .31  | .44  | .42  |
| trees     | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 1 | 0 | -.06 | .23  | -.14 | -.27 | .14 | .24  | .55  | .77  | .66  |
| graph     | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 1 | -.06 | .34  | -.15 | -.30 | .20 | .31  | .69  | .98  | .85  |
| minors    | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | -.04 | .25  | -.10 | -.21 | .15 | .22  | .50  | .71  | .62  |

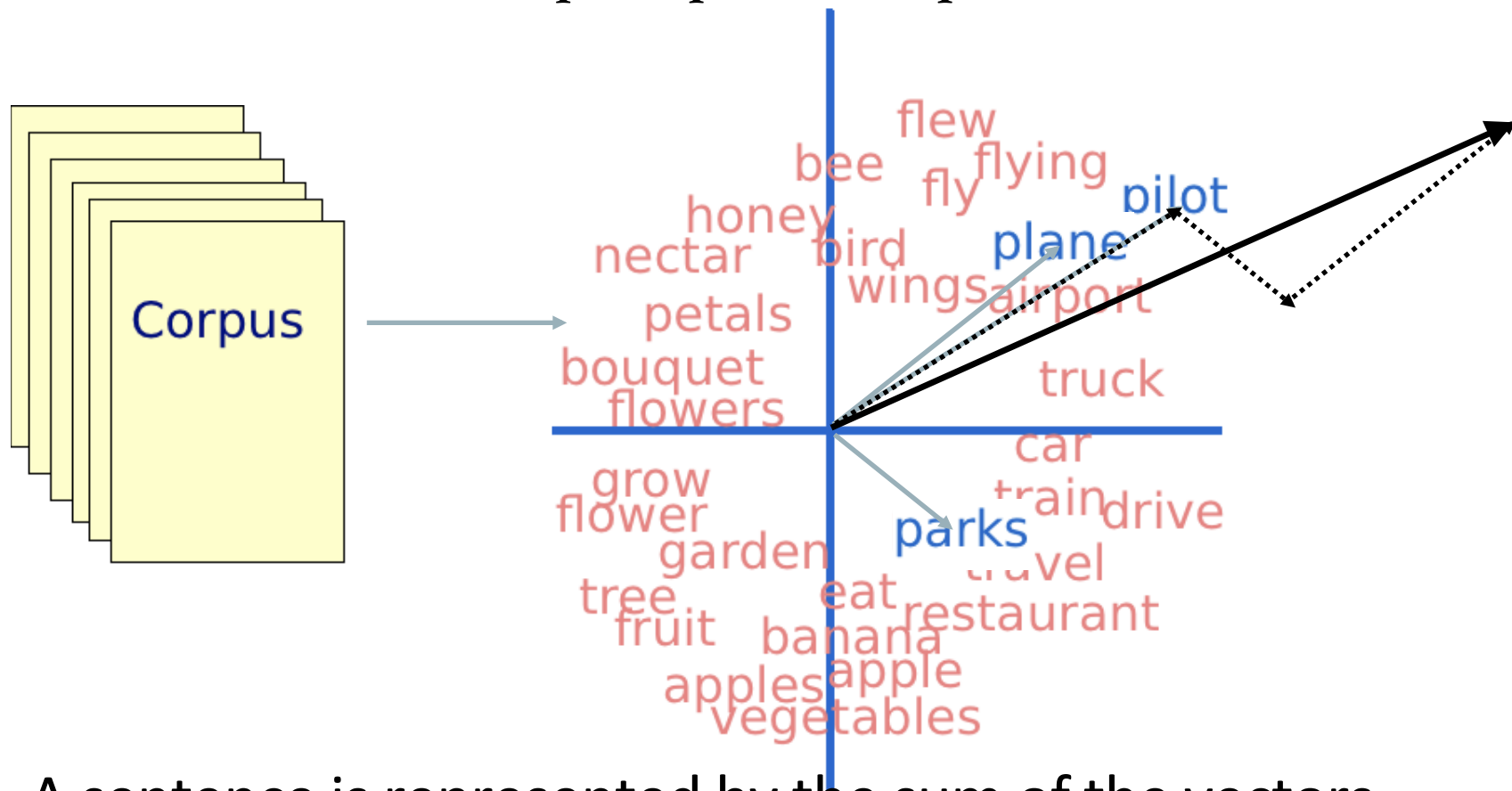
$r(\text{human}, \text{user}) = -.38$

$r(\text{human}, \text{user}) = .94$



# Latent Semantic Analysis: Sentence/Document Representation [Landauer & Dumais 90]

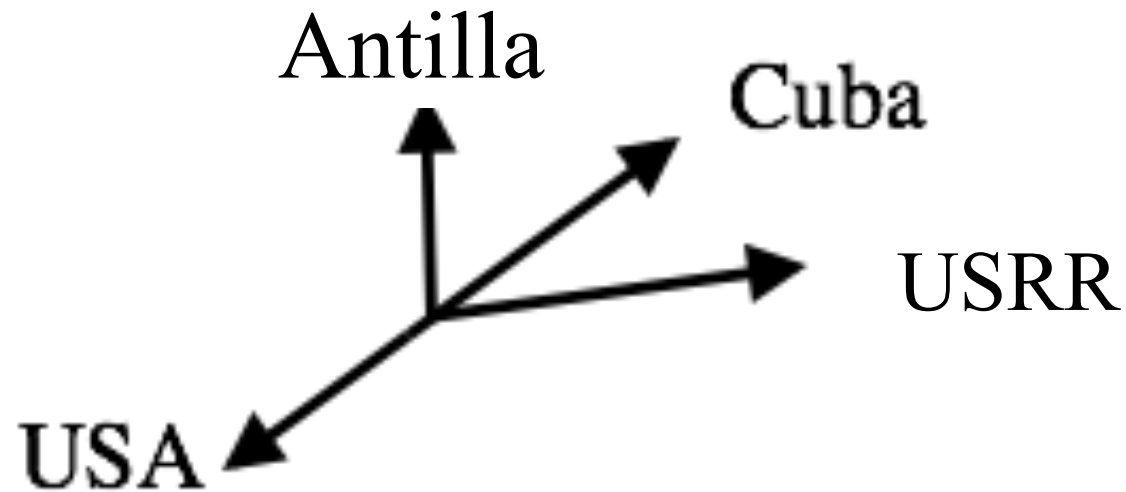
“The pilot parks the plane”



A sentence is represented by the sum of the vectors of its words [Lemaire & Denhière '05]

# <sup>10</sup> LSA Captures some (but not all!) Complex Word Semantic Relationships

Hypothetical Vectors [After Tversky 77]



# 11 Topic Modelling with Latent Dirichlet Allocation [Blei 2012]

- Probabilistic method, each analysed document is a “mix” of topics, of decreasing probabilities.
- Three example topics :

père(1750.0) famille(1267.0) mère(1221.0) fils(1139.0)  
enfant(1088.0) jeune(771.0) grand(644.0)  
parent(589.0) ...

guerre(1074.0) armée(381.0) soldat(294.0)  
résistance(227.0) combat(214.0) général(211.0)  
officier(187.0) ...

hôpital(975.0) médecin(774.0) médical(549.0)  
service(407.0) santé(405.0) malade(363.0)  
docteur(306.0) ...

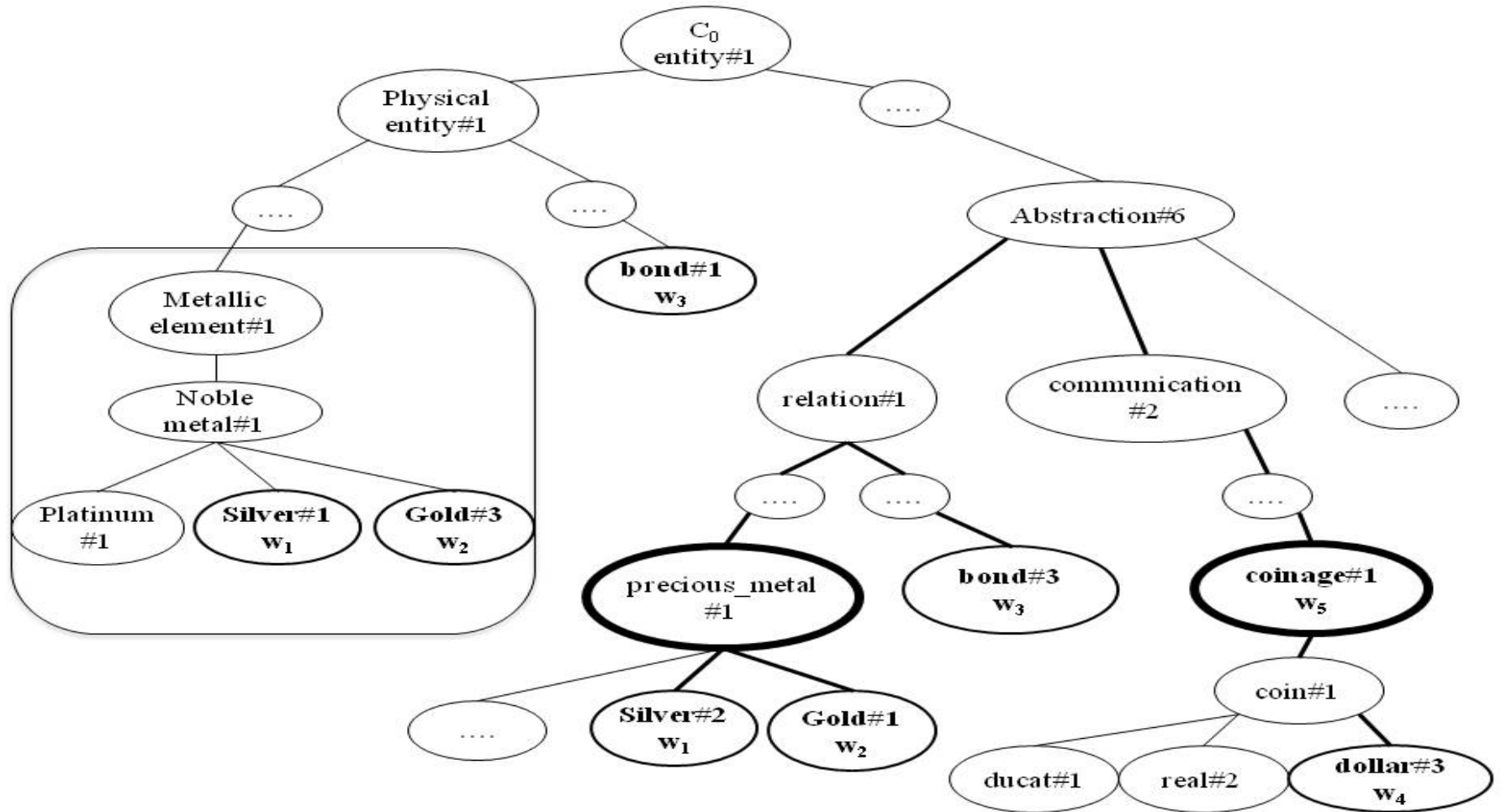


# WordNet

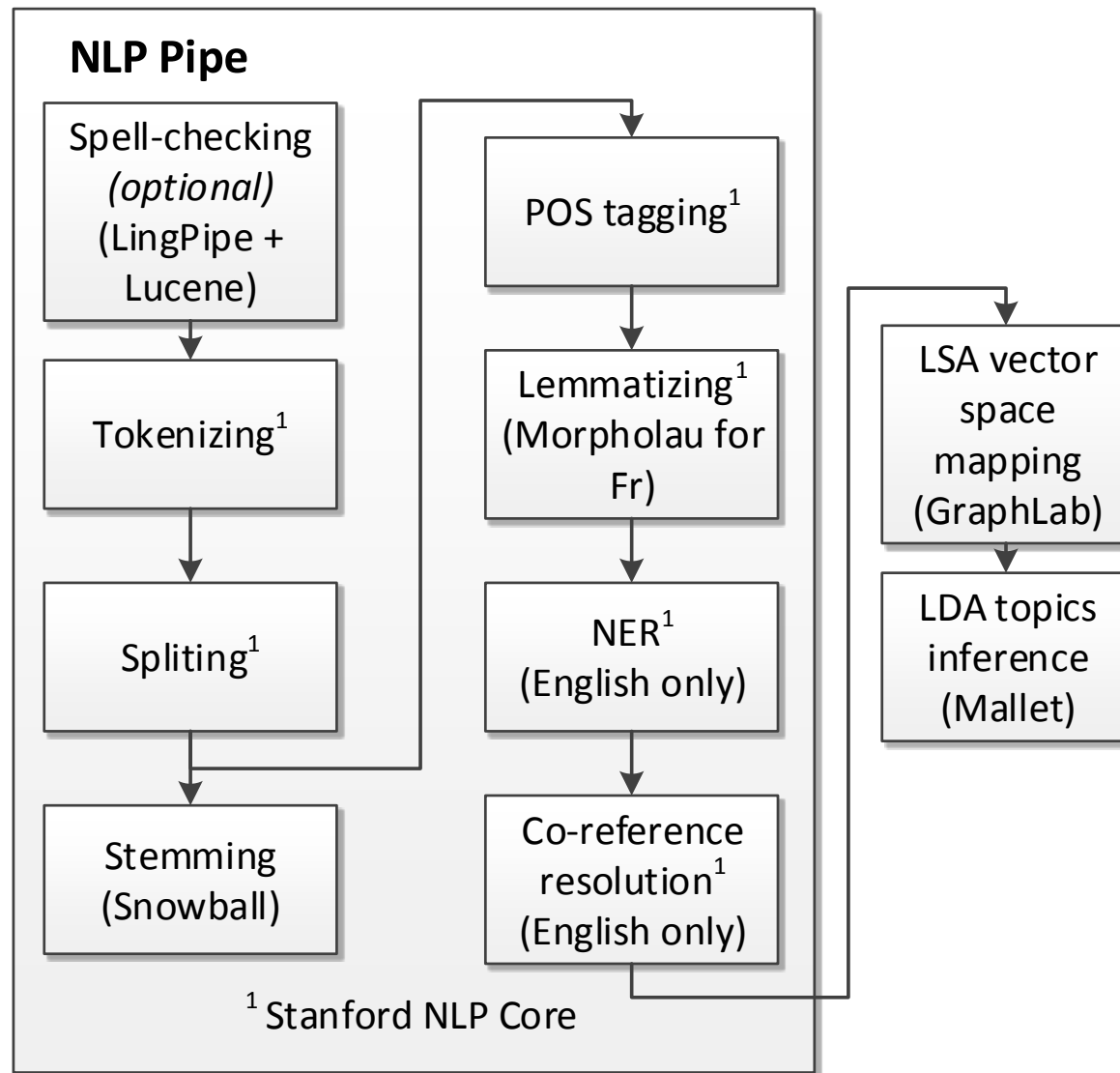
([wordnet.princeton.edu](http://wordnet.princeton.edu))

- Ontology > 150,000 concepts
- Nouns, verbs, adjectives, adverbs
- Built from psycholinguistics experiments → a semantic network of common concepts used in language

# WordNet: Topical Subtrees – An Example



# Text Pre-processing





# Corpora Description

- 1,000 abstracts from Jensen & Grauwin database
- Abstracts from educational science domain
- Problems of scalability & processing of entire corpora using full NLP pipe tools

# New Functionalities (1)


- Corpus similarity view
  - Conceptualize the corpora through semantic similarity between abstracts
  - Finds most similar pairs of articles
  - Determine most central articles
- Document centrality view
  - Find most similar articles to a selected article
  - Expand using a BFS strategy the graph by using the connected articles

## New Functionalities (2)

- Concept view
  - Identify most relevant concepts for the entire corpora
  - Relational meta models
- Keyword-abstract overlaps
  - Determine best overlap between keywords and abstract content
  - Compound score: occurrences of keywords (30%) and the semantic relatedness between keyword "phrase" and abstracts (70%)



# Article View


ReaderBench - Document Visualization

**Title:** Subject departments as professional communities?

Source: LSE    URI:    Sentiment polarity:

### Contents

a growing body of literature suggests that when schools become professional communities there are expected benefits in terms of teacher learning, school improvement and student achievement. in this article the concept of professional communities is examined for certain subject departments in dutch secondary schools. the authors report on research into the extent to which mathematics departments operate as professional communities. at the same time, it was investigated whether the level at which departments operate as professional communities is related to student achievement. the results indicate that departments are cohesive bodies regulating teacher behaviour in several respects. however, as professional communities they do not focus on improving the quality of their teachers and instruction. some characteristics of professional communities prove to be beneficial for student achievement, while others are not. the authors offer recommendations on how departments can develop into more professionally organised communities. [31.949]

### Topics

Filter only:

☒ Nouns    ☒ Verbs    0    25    50

| Topics         | Relevance |
|----------------|-----------|
| community      | 7.05      |
| student        | 3.9       |
| school         | 3.86      |
| teacher        | 3.84      |
| department     | 3.52      |
| achievement    | 2.74      |
| body           | 2.35      |
| learning       | 1.73      |
| author         | 1.69      |
| operate        | 1.43      |
| literature     | 1.4       |
| time           | 1.23      |
| recommendation | 1.11      |
| improvement    | 1.08      |
| instruction    | 1.05      |
| characteristic | 1.03      |
| extent         | 0.96      |
| respects       | 0.95      |

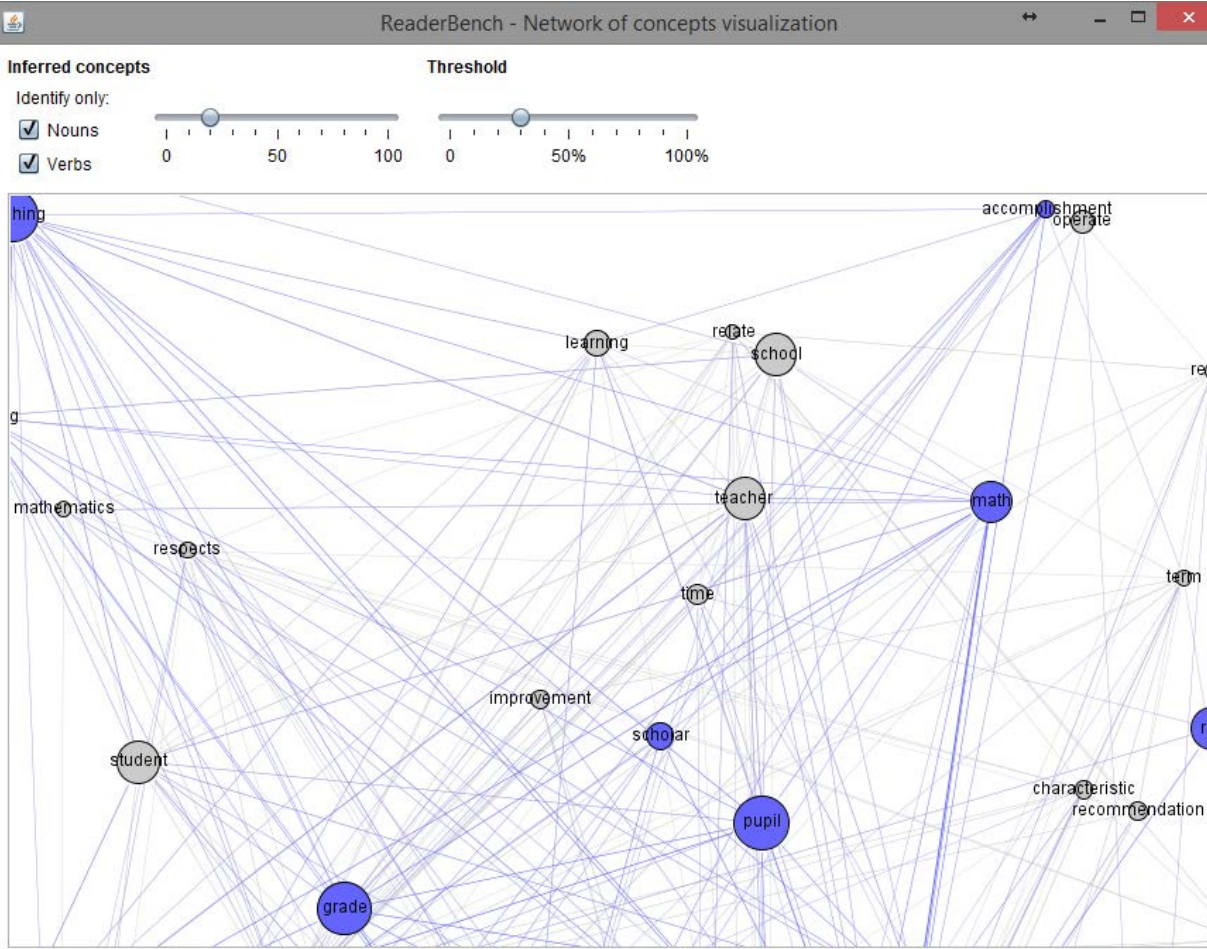
Advanced View

Visualize Multi-Layered Cohesion Graph

Select Voices

Display Voice Inter-animation

Generate network of concepts



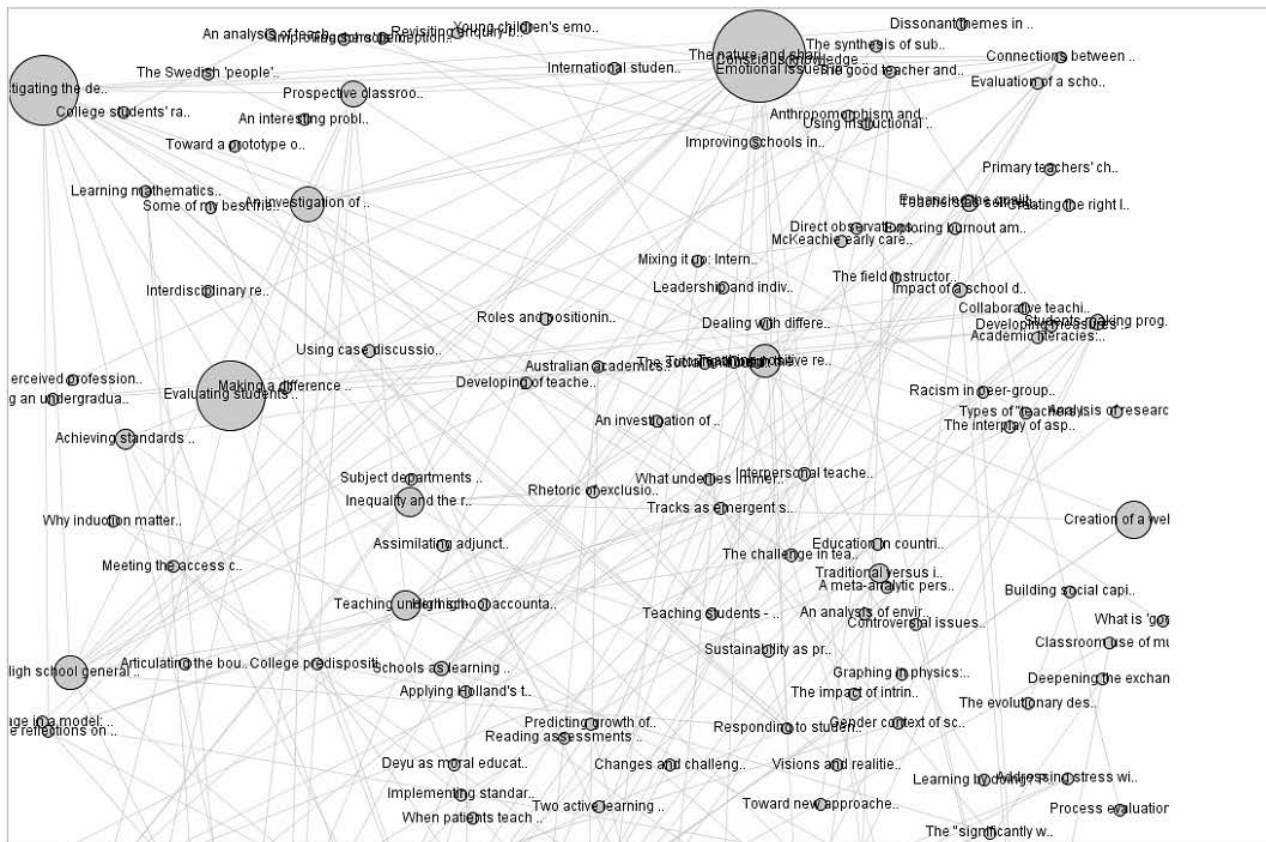


# Corpus Similarities View

Threshold



Distance to Central Document



| Document   | Distance |
|--|----------|
| Novice and expert teachers' conceptions of learners' prior knowledge               | 0.77     |
| Connections between teacher and student attitudes regarding script choice...       | 0.77     |
| An investigation of science teaching practices in Indonesian rural seconda...      | 0.76     |
| Message in a model: Teachers' responses to a court-ordered mandate for...          | 0.75     |
| Investigating the decision-making skills of cooperating teachers and stud...       | 0.74     |
| Using case discussion on the web to develop student teacher problem so...          | 0.73     |
| South Korean high school English teachers' code switching: Questions an...         | 0.73     |
| The good teacher and good teaching: Comparing beliefs of second-grade ...          | 0.72     |
| Teaching evolutionary biology: Pressures, stress, and coping                       | 0.72     |
| The challenge in teaching biotechnology  | 0.72     |
| High school general physical education teachers' behaviors and beliefs a...        | 0.71     |
| Responding to students' views about university teaching: the experience of...      | 0.71     |
| Relations between EFL teachers' formal knowledge of grammar and their ...          | 0.71     |
| The field instructor as group worker: Managing trust and competition in gro...     | 0.71     |
| The synthesis of subject and pedagogy for effective learning and teaching ...      | 0.70     |
| Evaluating students' science notebooks as an assessment tool                       | 0.70     |
| Prospective classroom teachers' metaphorical images of selves and com...           | 0.70     |
| Perceived professional needs of Korean science teachers majoring in che...         | 0.70     |
| Some reflections on learning from early school experience                          | 0.70     |
| Analysing cognitive or non-cognitive factors involved in the process of phys...    | 0.70     |
| Developing teachers' measures mathematics knowledge for teaching                   | 0.70     |
| Developing measures of content knowledge for teaching reading                      | 0.70     |
| Interpersonal teacher behaviour and student outcomes                               | 0.70     |
| Primary teachers' changing attitudes and cognition during a two-year scie...       | 0.69     |
| Impact of a school district's science reform effort on the achievement and ...     | 0.69     |
| Inequality and the right to learn: Access to qualified teachers in California's... | 0.69     |
| Enhancing the quality of argumentation in school science                           | 0.69     |
| Controversial issues - teachers' attitudes and practices in the context of cit...  | 0.69     |
| The interplay of aspirations, enjoyment, and work habits in academic ende...       | 0.68     |
| Teacher scoring of large-scale assessment: professional development or...          | 0.68     |
| Learning mathematics in a classroom community of inquiry                           | 0.67     |
| What is 'good teaching' in engineering education in India? A case study            | 0.67     |
| Situated learning in teacher training  | 0.67     |
| Visions and realities of Internet use in schools: Canadian perspectives            | 0.67     |
| What underlies immersion students' production: The case of Avior Besoin ...        | 0.67     |
| Dissonant themes in preservice portfolio development                               | 0.66     |
| Achieving standards in urban systemic reform: An example of a sixth grad...        | 0.66     |
| Tutoring through the internet: how students and teachers interact to constr...     | 0.66     |
| Evaluation of a school-based intervention for HIV/AIDS prevention among ...        | 0.65     |
| Between idealism and pragmatism: a case study of student teachers' ped...          | 0.65     |
| Teaching under high-stakes testing - Dilemmas and decisions of a teach...          | 0.65     |
| Toward new approaches for evaluating student field performance: Tappin...          | 0.65     |
| Traditional versus integrated preservice teacher education curriculum - A c...     | 0.65     |
| Teaching students - "Idea about science": dimensions of effective practice         | 0.65     |

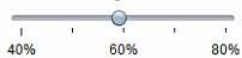
## Top Similar Articles

| Article 1   | Article 2   | Semantic Similarity Score |
|---|---|---------------------------|
| Developing of teachers' measures mathematics knowledge for teaching   | Developing measures of content knowledge for teaching reading   | 0.86                      |
| Investigating the decision-making skills of cooperating teachers and student teachers of English in a Turkish context                 | Interpersonal teacher behaviour and student outcomes  | 0.79                      |
| South Korean high school English teachers' code switching: Questions and challenges in the drive for maximal use of English in tea... | Connections between teacher and student attitudes regarding script choice in first-year Japanese language classrooms  | 0.79                      |
| Gender context of schooling and levels of stress among early adolescent pupils  | Gender organization of schooling and television viewing among early adolescents: a test of two alternative hypotheses | 0.79                      |
| Teaching under high-stakes testing - Dilemmas and decisions of a teacher educator   | Teachers as self-regulated learners   | 0.78                      |
| College students' ratings of student ability and teacher input as correlates of student performance on multiple-choice e...           | An analysis of teachers' rating scales as sources of evidence for a standardised Greek reading test                   | 0.77                      |
| The nature and sharing of teacher knowledge of technology in a student teacher/mentor teacher pair                                    | Novice and expert teachers' conceptions of learners' prior knowledge  | 0.77                      |



# Document Centrality View

Threshold among documents

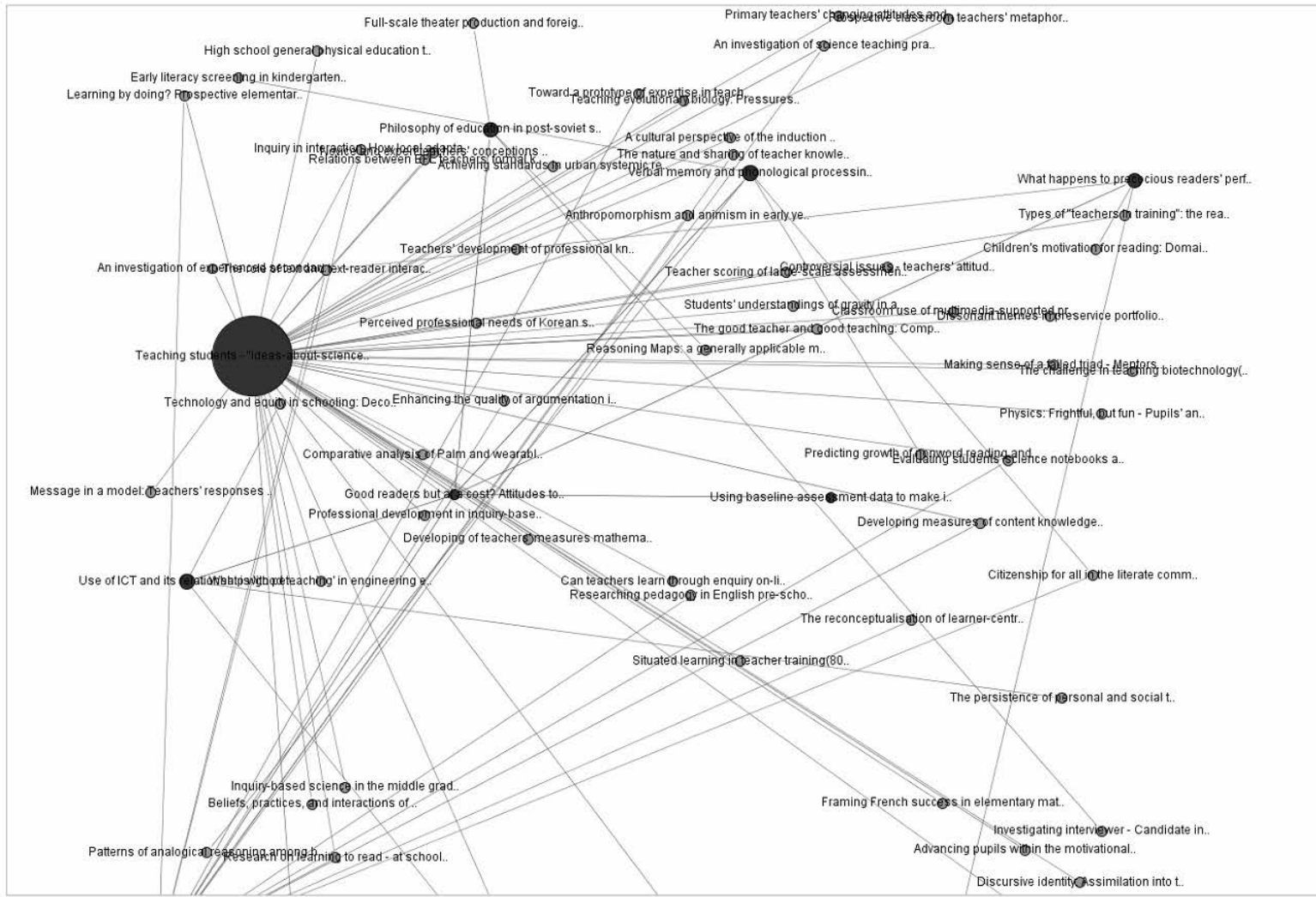


Depth Level

2

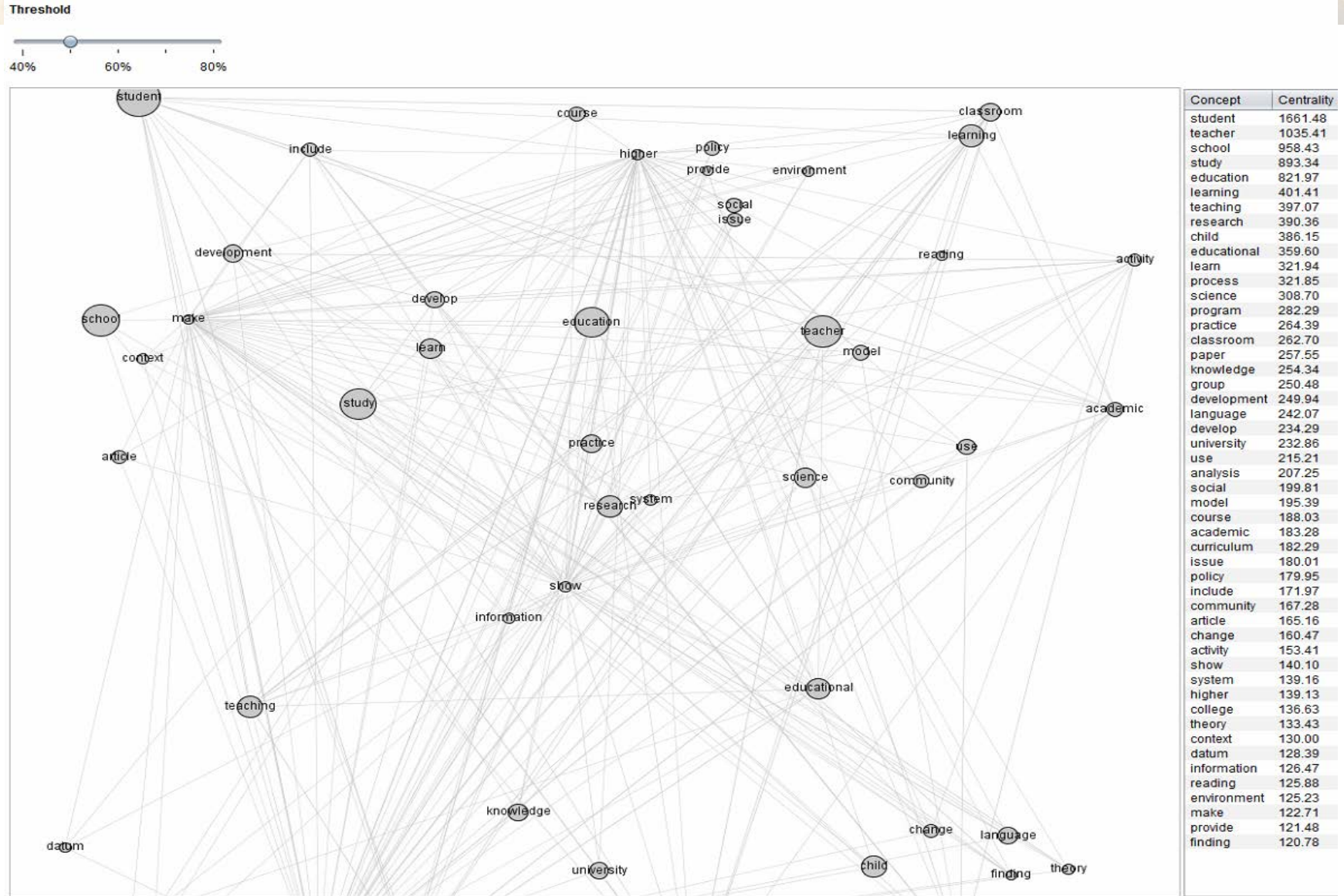
Top similar articles

| Article  | Similarity |
|--|------------|
| Verbal memory and phonological proce...        | 0.68       |
| Philosophy of education in post-soviet s...    | 0.65       |
| Use of ICT and its relationship with perf...   | 0.62       |
| What happens to precocious readers' p...       | 0.61       |
| Reading assessments in Kindergarten ...        | 0.61       |
| Using baseline assessment data to ma...        | 0.60       |
| Teaching students - "Ideas-about-scienc...     | 0.59       |
| Citizenship for all in the literate commun...  | 0.58       |
| The persistence of personal and social ...     | 0.58       |
| Patterns of analogical reasoning among...      | 0.56       |
| The role of text and text-reader interacti...  | 0.56       |
| Early literacy screening in kindergarten: ...  | 0.56       |
| Predicting growth of nonword reading a...      | 0.55       |
| Framing French success in elementary ...       | 0.55       |
| Developing measures of content knowl...        | 0.55       |
| Research on learning to read - at schoo...     | 0.55       |
| Researching pedagogy in English pre-s...       | 0.54       |
| Situated learning in teacher training          | 0.52       |
| Students' understandings of gravity in a...    | 0.51       |
| Beyond phonological skills: broader lan...     | 0.51       |
| Children's motivation for reading: Doma...     | 0.51       |
| Investigating interview - Candidate int...     | 0.50       |
| Anthropomorphism and animism in earl...        | 0.50       |
| The challenge in teaching biotechnology(...    | 0.50       |
| Can teachers learn through enquiry on-l...     | 0.50       |
| Impact of a school district's science refo...  | 0.50       |
| Inquiry in interaction: How local adaptati...  | 0.50       |
| Teachers' development of professional ...      | 0.49       |
| Enhancing the quality of argumentation i...    | 0.49       |
| Relations between EFL teachers' formal...      | 0.49       |
| A study of science teaching self-efficacy ...  | 0.49       |
| Professional development in inquiry-ba...      | 0.48       |
| Primary teachers' changing attitudes an...     | 0.48       |
| Evaluation of a school-based interventio...    | 0.48       |
| Prospective classroom teachers' metap...       | 0.48       |
| Message in a model: Teachers' respon...        | 0.47       |
| An investigation of science teaching pra...    | 0.47       |
| Developing of teachers' measures math...       | 0.47       |
| Beliefs, practices, and interactions of te...  | 0.47       |
| Evaluating students' science notebooks...      | 0.47       |
| Full-scale theater production and foreig...    | 0.47       |
| Inquiry-based science in the middle gra...     | 0.47       |
| Novice and expert teachers' conception...      | 0.47       |
| Classroom use of multimedia-supporte...        | 0.46       |
| Teacher scoring of large-scale assess...       | 0.46       |
| Toward a prototype of expertise in teach...    | 0.46       |
| Learning by doing? Prospective elemen...       | 0.46       |
| Achieving standards in urban systemic r...     | 0.45       |
| Types of "Teachers in training": the react...  | 0.45       |
| The difficult relationship between theory ...  | 0.45       |
| Physics: Frightful, but fun - Pupils' and t... | 0.44       |
| The nature and sharing of teacher know...      | 0.44       |
| High school general physical education ...     | 0.44       |
| Technology and equity in schooling: De...      | 0.44       |
| Comparative analysis of Palm and wear...       | 0.44       |
| The good teacher and good teaching: C...       | 0.44       |
| Teaching evolutionary biology: Pressure...     | 0.43       |





# Concept View





# Keywords-abstract Overlap

Best Articles - Keyword&Abstract Overlap

Keyword-Abstract overlap score

| Article   | Overlap Score | Semantic Score | Aggregated S... |
|---|---------------|----------------|-----------------|
| Exchange of information about physical education to support the transition of pupils from primary and secondary school                              | 1.00          | 0.64           | 0.75            |
| Perceived professional needs of Korean science teachers majoring in chemical education and their preferences for online and on-site training        | 0.81          | 0.70           | 0.73            |
| Connections between teacher and student attitudes regarding script choice in first-year Japanese language classrooms                                | 0.83          | 0.69           | 0.73            |
| Students at risk: Students' general study orientations and abandoning/prolonging the course of studies  | 0.73          | 0.71           | 0.71            |
| Views of elementary and middle school Turkish students toward environmental issues  | 0.88          | 0.61           | 0.69            |
| The good teacher and good teaching: Comparing beliefs of second-grade students, preservice teachers, and inservice teachers                         | 0.67          | 0.68           | 0.68            |
| Novice and expert teachers' conceptions of learners' prior knowledge  | 0.75          | 0.65           | 0.68            |
| Impact of a school district's science reform effort on the achievement and attitudes of third- and fourth-grade students                            | 0.62          | 0.71           | 0.68            |
| A longitudinal study of the educational and career trajectories of female participants of an urban informal science education program               | 0.62          | 0.71           | 0.68            |
| Types of "teachers in training": the reactions of primary school science teachers when confronted with the task of implementing an innovation       | 0.65          | 0.68           | 0.67            |
| College students' ratings of student effort, student ability and teacher input as correlates of student performance on multiple-choice exams        | 0.79          | 0.61           | 0.67            |
| An analysis of burnout and job satisfaction among Turkish special school headteachers and teachers, and the factors effecting their burnout and ... | 0.85          | 0.60           | 0.67            |
| Different worlds? A comparison of young people's home and school ICT use  | 0.75          | 0.62           | 0.66            |
| An analysis of teachers' rating scales as sources of evidence for a standardised Greek reading test   | 0.73          | 0.63           | 0.66            |
| Thinking differently about cultural diversity: Using postcolonial theory to (re)read science education  | 0.65          | 0.65           | 0.65            |
| Teachers' perceptions of in-service teacher training to support curriculum change in physical education: The Hong Kong Experience                   | 0.69          | 0.63           | 0.65            |
| HIV testing and awareness of care-related services among a group of HIV-positive Asian Americans and Pacific Islanders in te United States: Fin...  | 0.60          | 0.68           | 0.65            |
| Triple scheme of learning support design for scientific discovery learning based on computer simulation: experimental research                      | 0.62          | 0.65           | 0.64            |

South Korean high school English teacher  
Similarities and differences in role conflict  
Learning by doing? Prospective elementary school teachers' beliefs  
Are we creating separate and unequal tracks?  
Variation in the alcohol content of a 'drink'  
The development of study orientations and  
The development and validation of a framework  
Implementing mathematics reform in urban schools  
Evaluating students' science notebooks as  
Effects of neighborhood socioeconomic conditions  
Conceptions of learning science among high school students  
Anthropomorphism and animism in early childhood  
An evaluation of the relative efficiency for visualizing  
Analysis of research findings on environmental education  
The protective effect of parental expectations on children's academic achievement  
The effects of discovery learning and experiential learning on students' learning outcomes  
Research on literacy policy and professional development  
Developing understanding of image formation in physics  
Can middle-school science textbooks help students understand the scientific process?

Message

**Article:** Exchange of information about physical education to support the transition of pupils from primary and secondary school

**Keywords:** continuity, liaison activities, progression, transfer from primary to secondary, exchange, information, physical, education, support, transition, pupils, primary, secondary, school

**Abstract:** the purpose of this study was to identify how information about physical education is exchanged between secondary schools and their respective feeder primary schools, what information is exchanged and how this information is used. a secondary purpose was to look at whether there is any relationship between schools engaging in liaison activities and exchanging information about physical education, and between exchanging information and the number of associated secondary schools to which pupils are sent or feeder primary schools from which pupils are received. questionnaires were sent to secondary and feeder primary schools. responses from secondary schools and primary schools showed that the highest percentage of teachers exchanged information through written documentation, followed by discussion at cross phase liaison meetings. the type of information exchanged by the highest percentage of teachers was identified as generic information about key stage and of the national curriculum for physical education ncpce areas of activity and schemes of work, rather than information about the specific physical education content covered or information about individual pupils, such as levels of attainment or ability. further, results suggest that information may be used for pastoral purposes and that only a small percentage of teachers used the information exchanged to plan for continuity and progression in the physical education curriculum. there was a significant positive relationship between engagement in liaison activities and information received about the physical education curriculum followed by pupils, but a significant negative relationship for primary teachers between the number of different secondary schools to which pupils' progress and knowledge about the key stage schemes of work that year pupils will follow in their associated secondary schools. these results are discussed in relation to continuity and progression in physical education in the transfer of pupils from primary to secondary schools.

OK



### 3. FURTHER WORK

## Towards a *Content*-based Citation Analysis (CAA) [Ding et al. 14]

There is at least 15 reasons why to refer to a paper (practices, purpose, or motivation) [Garfield 64]. Full-text analyses allow to dig deeper into these motivations

- Syntactic CAA: Citation Proximity Analysis [Gipp & Beel 09]: accounts for co-occurrences within sentences, paragraphs or paper sections (contexts) [Elkiss et al. 08]
- Semantic CAA: Account for semantic similarity across full texts

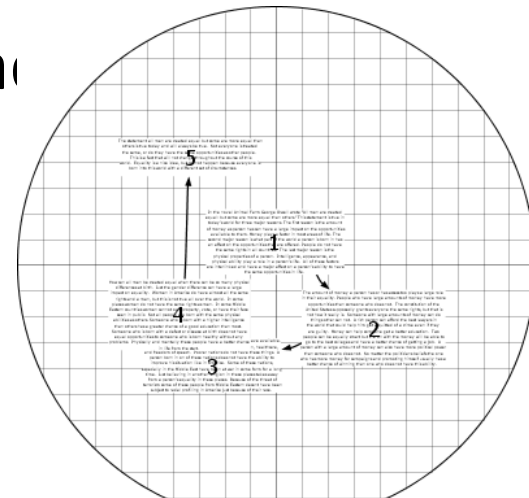


# Go Full Text!

- Semantic analyses of full text:
  - Use sentiment analysis to detect authors' citation motivation
  - Inter-paragraph cohesion to display overall paper's semantic flow [O'Rourke & Calvo 09]

• How to decrease the computational demand?

⚠ Consider using either LSA *or* LDA instead



# Further Work—Collaboration

- Domain representation readability (semiotic maps)
- Server setup to play with data
- Focus on specific sub-fields, like TEL -> ANR Orphee
- Merge our 2 approaches
  - 1<sup>st</sup> tier: Semantically compare the summaries-> Map
  - 2<sup>nd</sup> tier: Adjust the map with regard to papers' bib coupling

## Further Work—Likely Milestones

- Dynamic impact assessment: effect of the introduction of bulks of published papers on the paper's space (over years)
- LDA-based measurement [Song & Ding 14]
  - to tag clusters
  - to extract key-words and compare with the actuals
- Use researchers' networks and classification (Mendeley, CiteULike, etc.), a.k.a. alt-metrics, to uncover sub-disciplines [Kraker et al., in press]

## Even Further

- Recommender systems (which paper to read)
- Writing advices (how to write: coherence, which paper to cite)
- Citation genetic genealogy: a cited paper is the “son” of the papers citing it and shares their “ADN” [Sun & Zhu 12]
- Adding RDF papers metadata (e.g., CiTO tags <http://purl.org/spar/cito/> & SPAR, Semantic Publishing and Referencing initiative)



# Discussion

- Bib. coupling and semantic analysis are complementary approaches
- Relate computer-based analyses with human-based ones (results acceptability)

# THANK YOU FOR YOUR ATTENTION!

Refs available at

<http://www.citeulike.org/user/pdessus/tag/educmap>

Contact

trausan@gmail.com

philippe.dessus@upmf-grenoble.fr

paraschiv.ionut@gmail.com

mikedascalu@yahoo.com

# Aggregated Cohesion Function

- Normalized value in  $[0; 1]$  considering the following:
  - Inverse normalized distance
  - Lexical proximity – identical lemmas and semantic distances in WordNet / WOLF (Path, *Wu-Palmer*, Leacock-Chodorow)
  - Semantic similarity
    - Cosine similarity – LSA (GraphLab)
    - Jensen – Shannon Divergence – LDA (Mallet)

# Topics Extraction

- Individual normalized term frequency - *Statistical presence*
- Semantic similarities (LSA + LDA) - *Semantic relatedness*
  - Analysis element (sentence > block > document)
  - Entire document
- Weighted similarity with semantic chain - *Semantic coverage*



Topic Relevance