KNOWLEDGE MEDIA KMi INSTITUTE

Rexplore

Making Sense of Research

Enrico Motta
Knowledge Media Institute
The Open University
United Kingdom

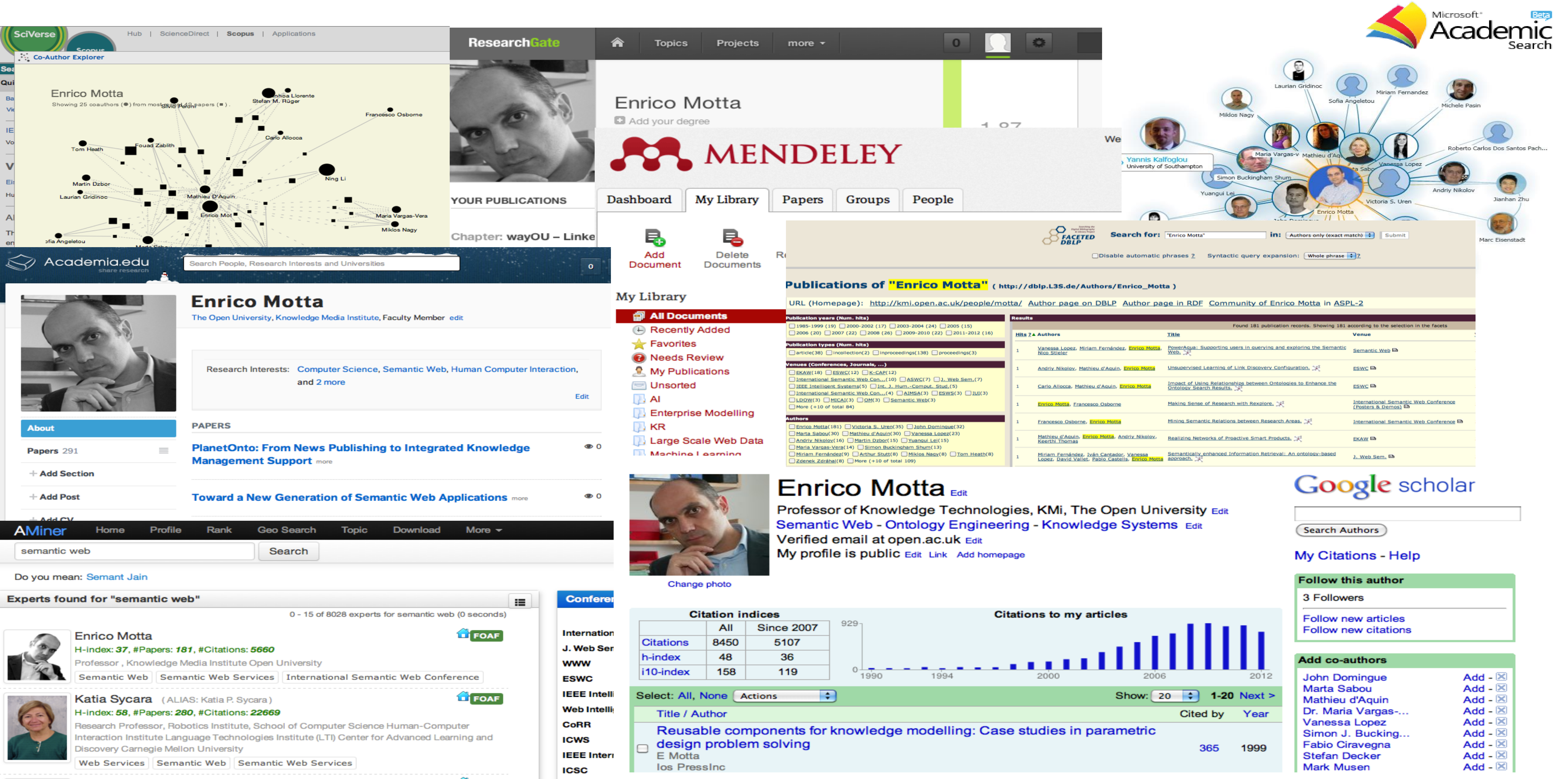
Hats I wear….

- Researcher
- Research Manager
- Supervisor/Mentor
- Editor-in-chief of a journal
- Advisor to strategic research programmes
- etc

Tasks

- **Academic Expert Search**.
  - E.g., “find me researchers with expertise in both Social Networks and Semantic Web, with at least some publications in CHI and ISWC, with more than 15 years research experience, a h-index greater than 15, etc”
- **Understanding Research Dynamics**
  - E.g., as EiC, I often need to make a decision about proposals for a special issue in a particular topic. This requires to understand whether the area is ‘hot’ right now or is decreasing in importance, who are the key people and groups, etc..

Exploring scholarly data: a variety of options….



Lack of comprehensive and integrated support

"There is still a need for an *integrated solution*, where the different functionalities and visualizations are provided in a coherent manner, through an environment able to support a seamless navigation between the different views and functionalities"

Dunne et al., 2012

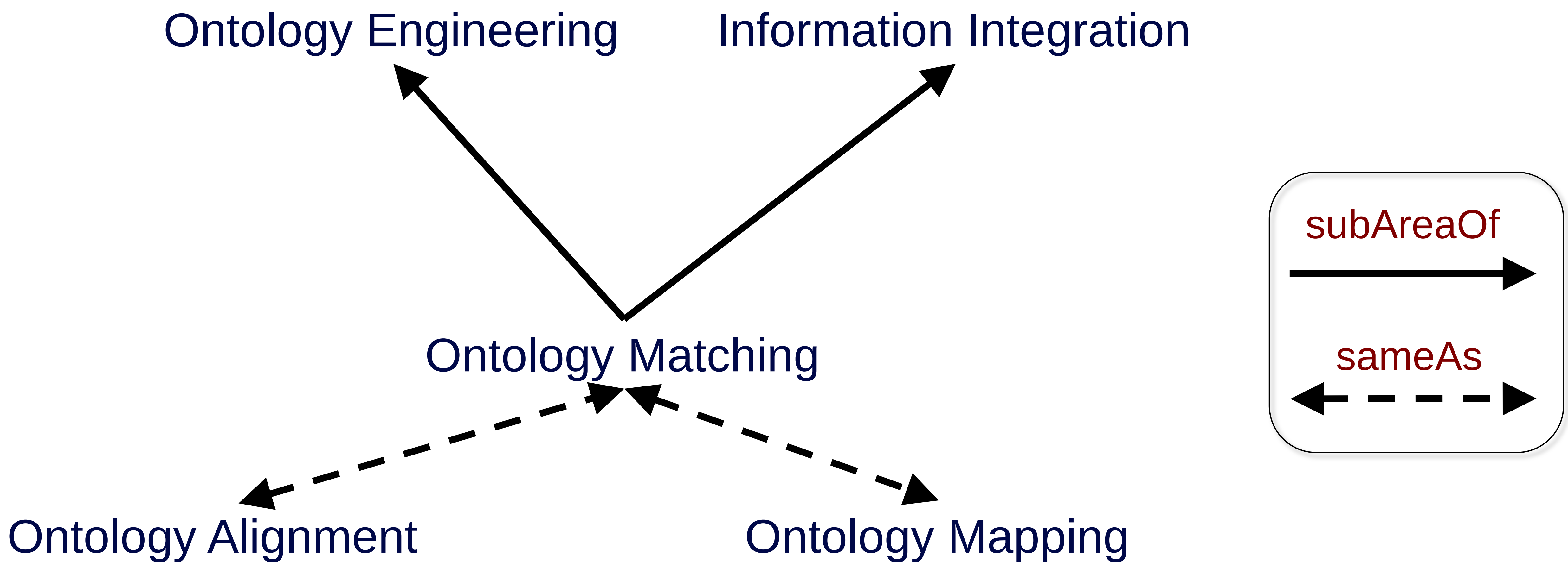
Digital library perspective

- Tools tend to focus primarily on authors' publications and citations rather than sensemaking or expert search (in particular highly-faceted expert search)

Lack of a semantic treatment of research topics

- Current tools do not treat research topics as ‘first class citizens’.
  - E.g., a tool may support a keyword search for papers on Ontology Matching, but by and large tools does not ‘understand’ that Ontology Matching is actually a research area
- Crucially, understanding what is a research area also means understanding what is <u>not</u> a research area
  - E.g., “case study” is often used as a tag for papers, but it is not actually a research area

Relations between research areas



ACM and other similar classifications

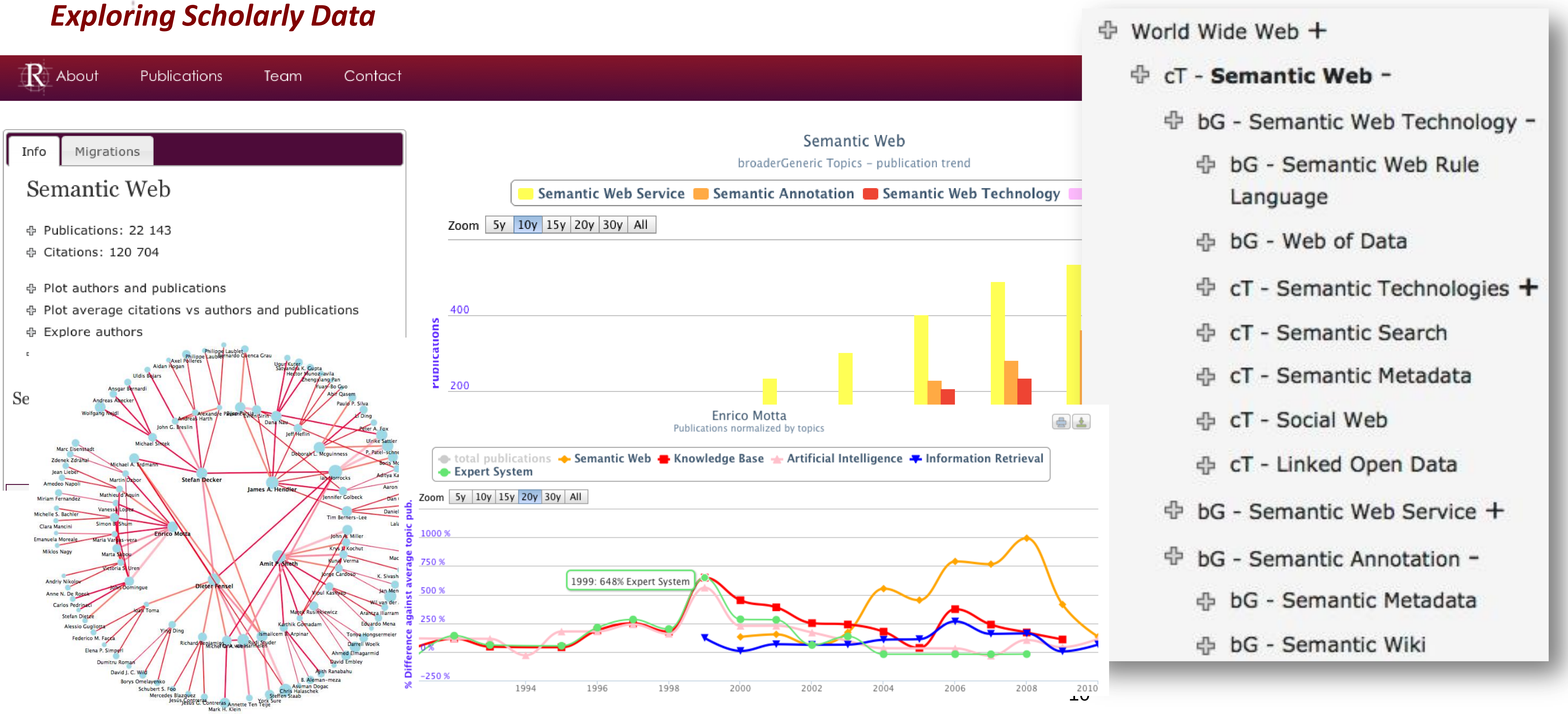
XII. Intelligent Web Services and Semantic Web

I. Intelligent Web service languages
II. Internet reasoning services
III. Ontology design
IV. Ontology languages

- The relations between entries are unclear
  - They are meant to be sub-areas, but for many of them it can be argued that they are not really sub-areas
- The different types of relationships are not distinguished
- Rather shallow
  - Most areas we know about are not listed – e.g., only 4 topics are classified under Semantic Web
- Static, manually defined, hence they get obsolete very quickly

Rexplore

Exploring Scholarly Data



Mining scholarly relations with Klink

- Klink takes as input a corpus of publications, annotated with keywords
  - Keywords can be user generated or can be automatically extracted from the abstract or the full text of the publication
  - In our experiments we used a corpus of 15M computer science publications obtained from Microsoft Academic Search
- Tidies up the set of keywords by removing keywords that do not denote a research area – e.g., “case study” or “NeOn Project”.
- Automatically computes three types of semantic relationships between the identified research areas.
- Returns a KB of semantic relationships between research areas

Relations mined by Klink

- *Skos:broaderGeneric (A, B)* – A is a sub-area of B.
  - E.g., “Semantic Web Services” is a sub-area of “Web Services”
- *relatedEquivalent (A, B)* – A and B are normally used to denote the same research area.
  - E.g., “Ontology Matching” and “Ontology Mapping” denote the same area
- *contributesTo (A, B)* – The outputs from area A are relevant to research in area B.
  - E.g., Research in “Ontology Engineering” contributes to research in “Semantic Web”

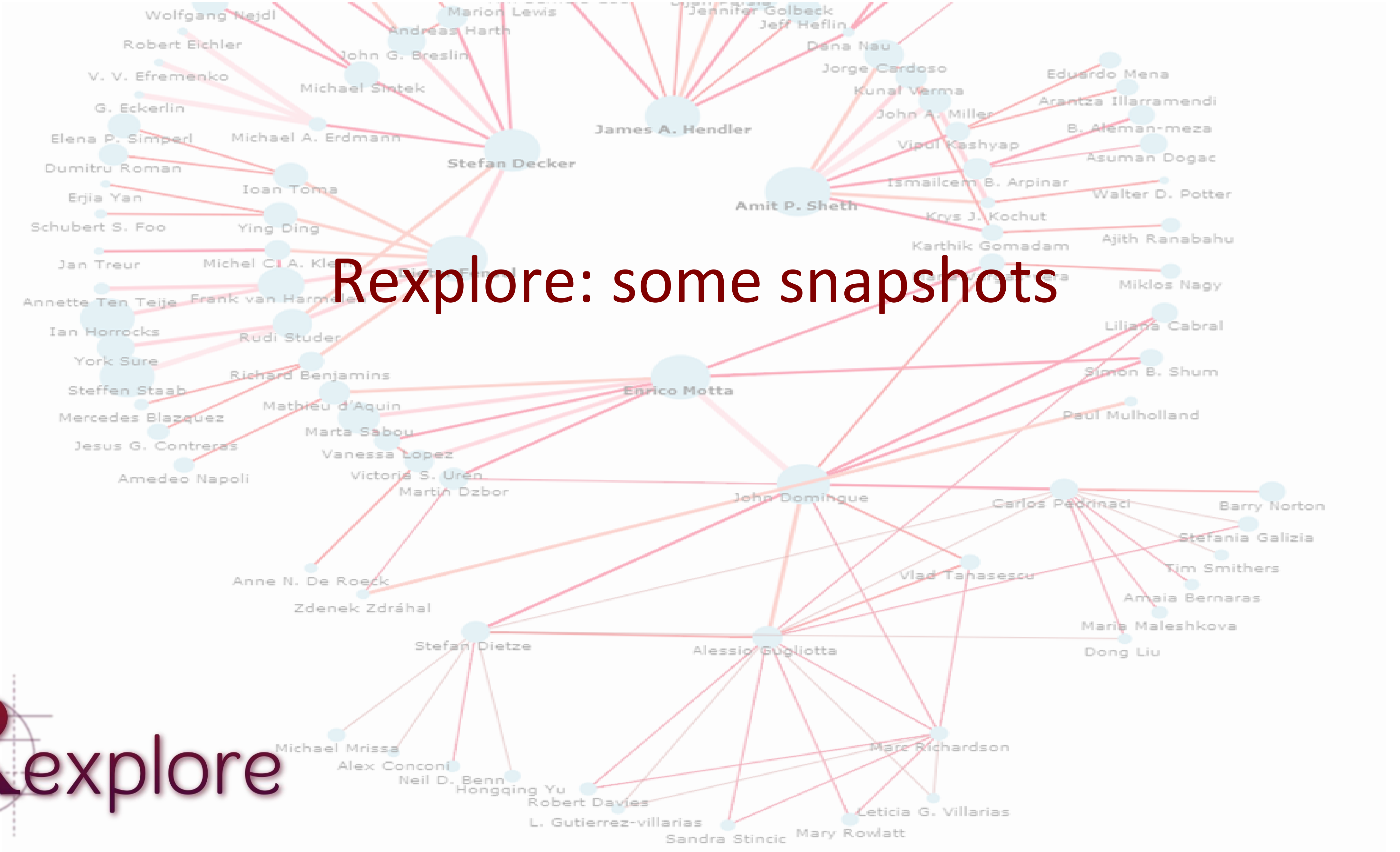
Semantic Relationships

- Reload Ontology Integration relationships

- Problem Solving +
  - bG - Domain Knowledge +
    - cT - Ontology +
      - bG - Ontology Mapping -
        - bG - **Ontology Integration**
      - bG - {Ontology Matching, Ontology Alignment} -
        - cT - **Ontology Integration**
  - cT - Knowledge Base +
    - bG - Knowledge Representation +
      - cT - Knowledge Acquisition +
        - bG - *Domain Knowledge* +
    - cT - Expert System +
      - cT - *Knowledge Acquisition* +
  - cT - Artificial Intelligence +
    - bG - *Knowledge Representation* +
    - cT - *Domain Knowledge* +
    - cT - *Expert System* +
- World Wide Web +
  - cT - Semantic Web +
    - bG - Semantic Interoperability +
      - cT - *Ontology Mapping* +
      - cT - *{Ontology Matching, Ontology Alignment}* +
    - cT - Ontology +
      - bG - *Ontology Mapping* +
      - bG - *{Ontology Matching, Ontology Alignment}* +
- Natural Language +
  - cT - *Knowledge Representation* +

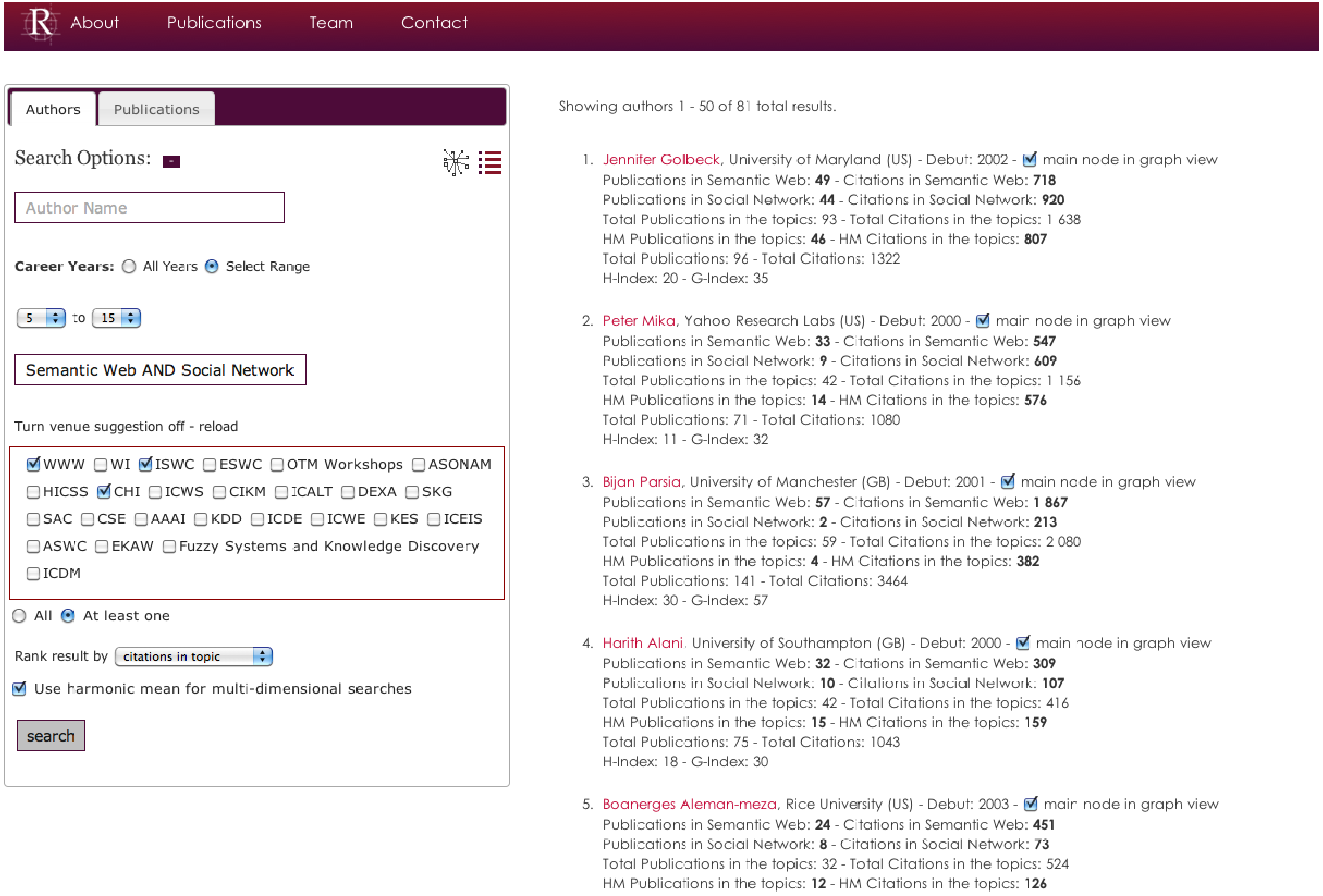
From a corpus of 15M papers accessed through the MAS API Klink identified about 1500 research topics and structured them by means of almost 3000 semantic relationships

Rexplore: some snapshots



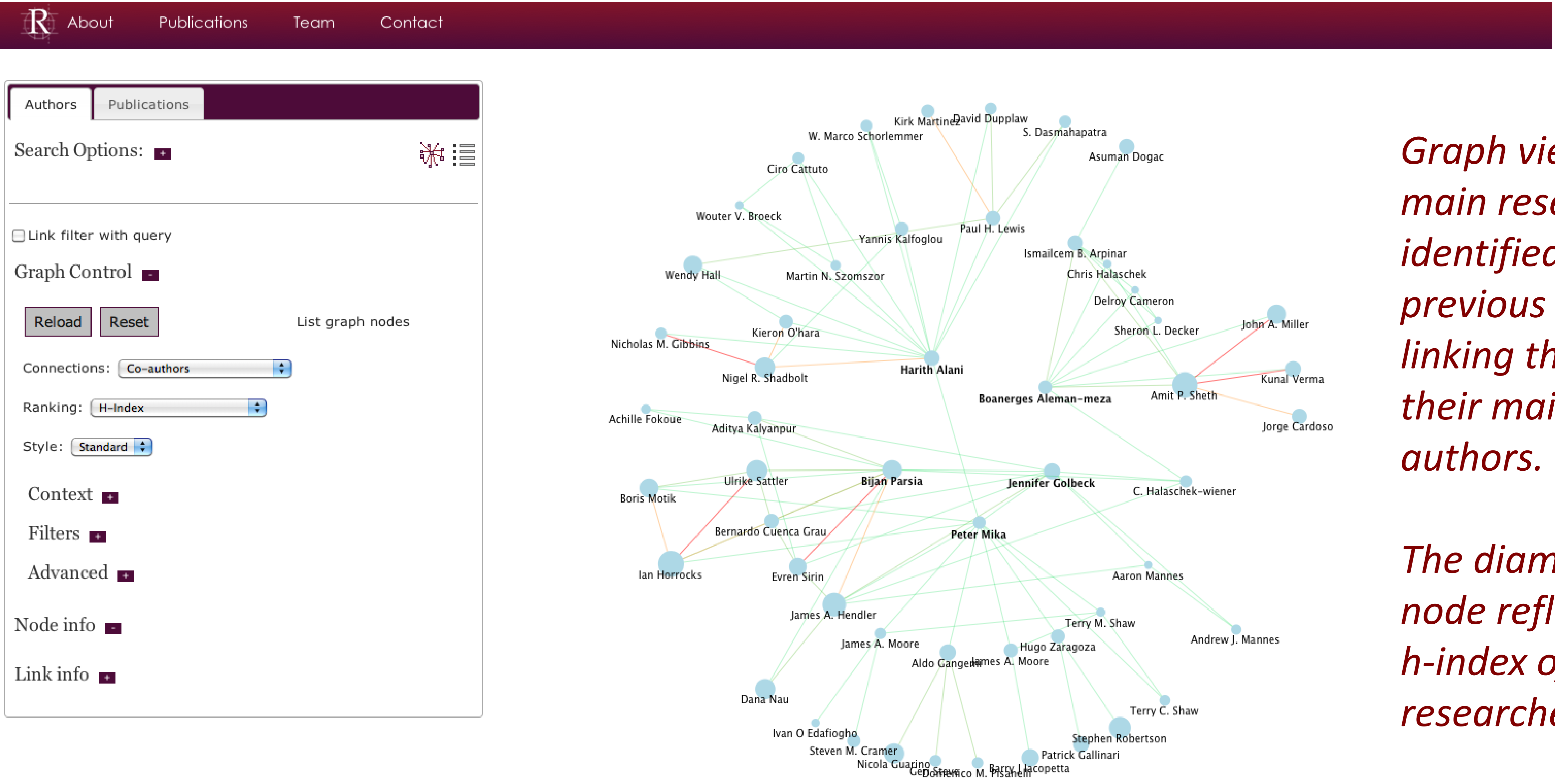
Rexplore

Expert Search (1a)



Researchers in the 5-15 career range with expertise in both semantic web and social networks, with publications in at least one of {CHI, ISWC, WWW), ranked with respect to the impact of their work in these two areas (using harmonic mean)

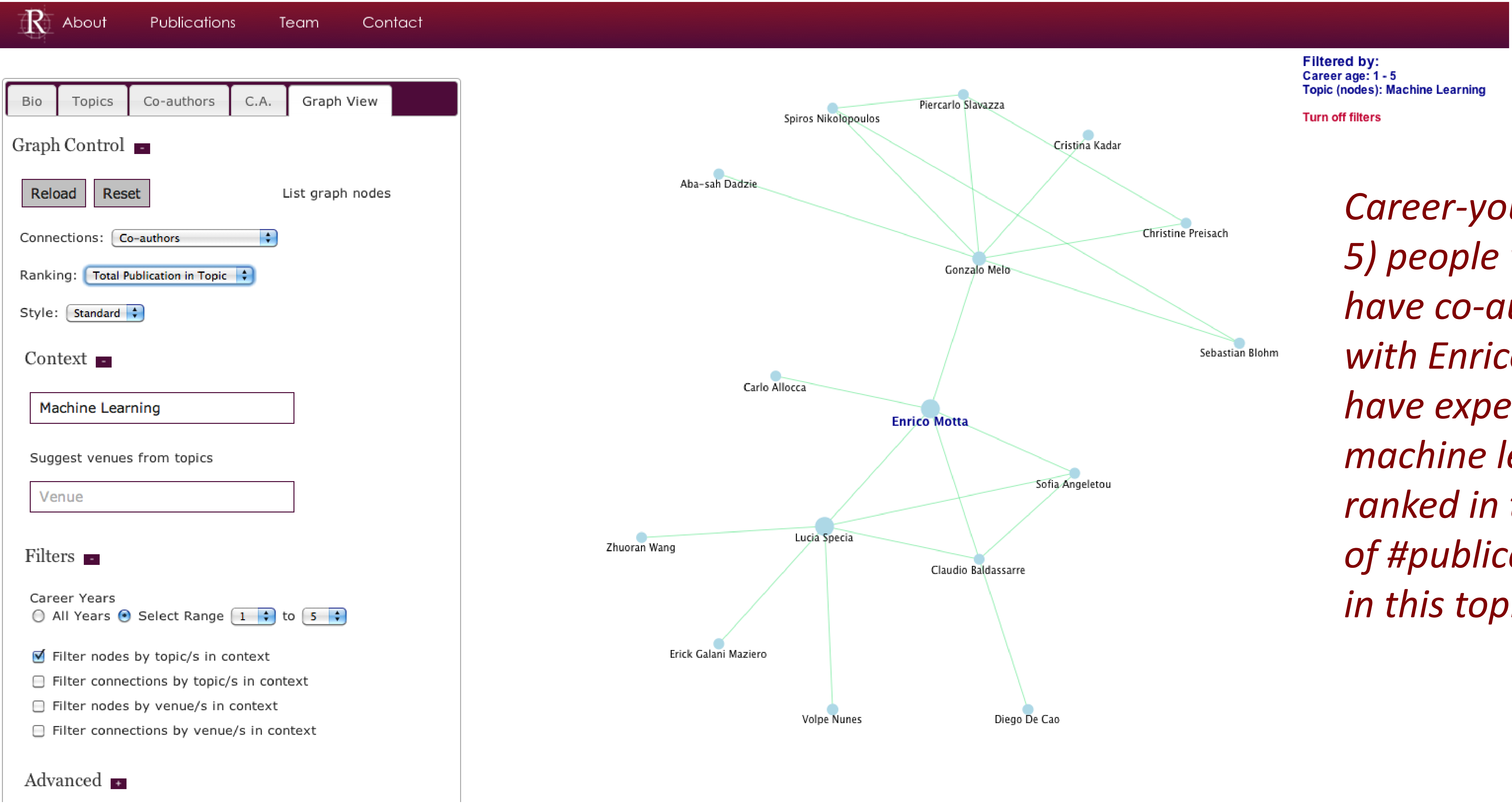
Expert Search (1b)



Graph view of main researchers identified in previous slide, linking them to their main co-authors.

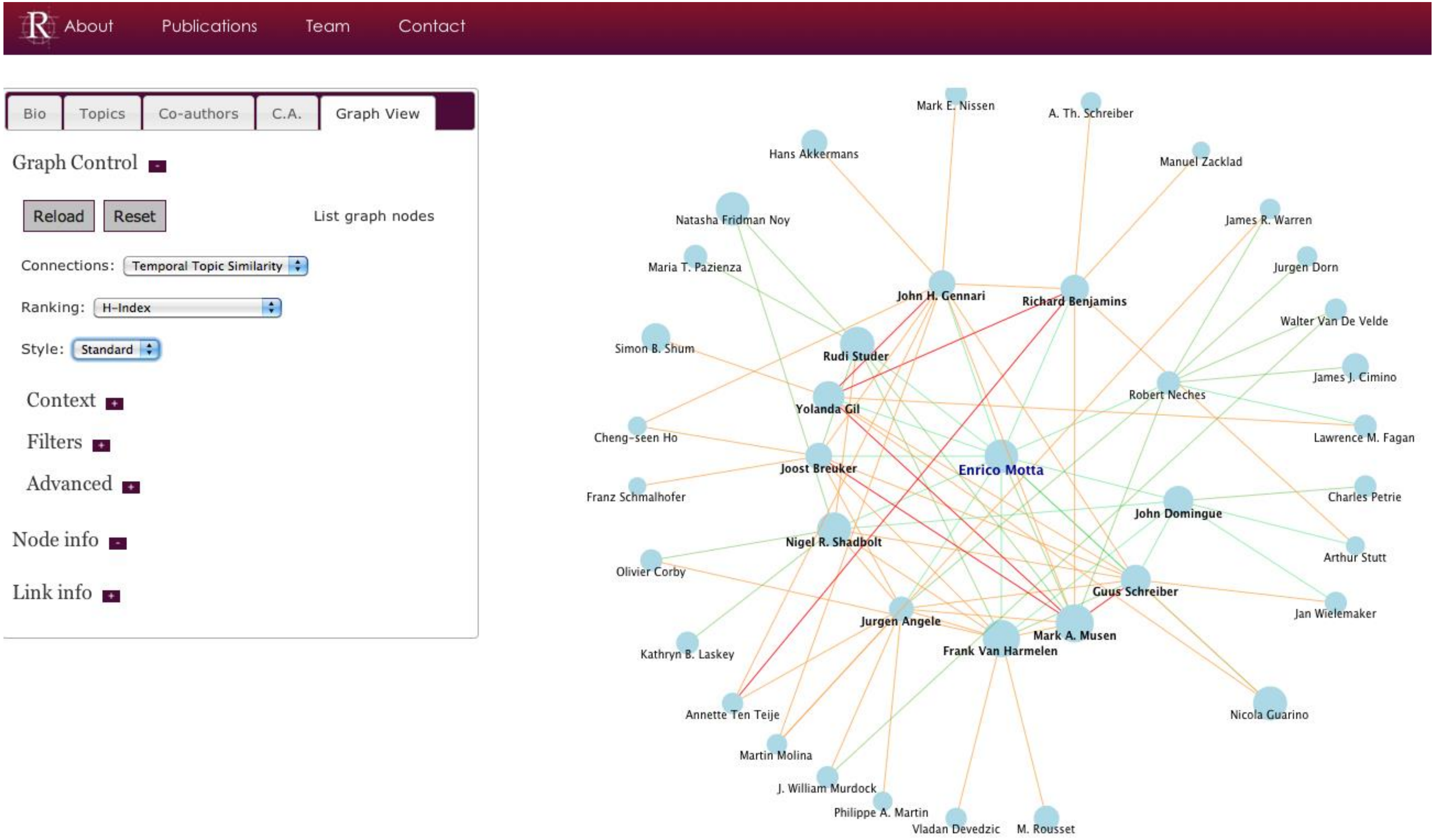
The diameter of a node reflects the h-index of the researcher

Expert Search (2)



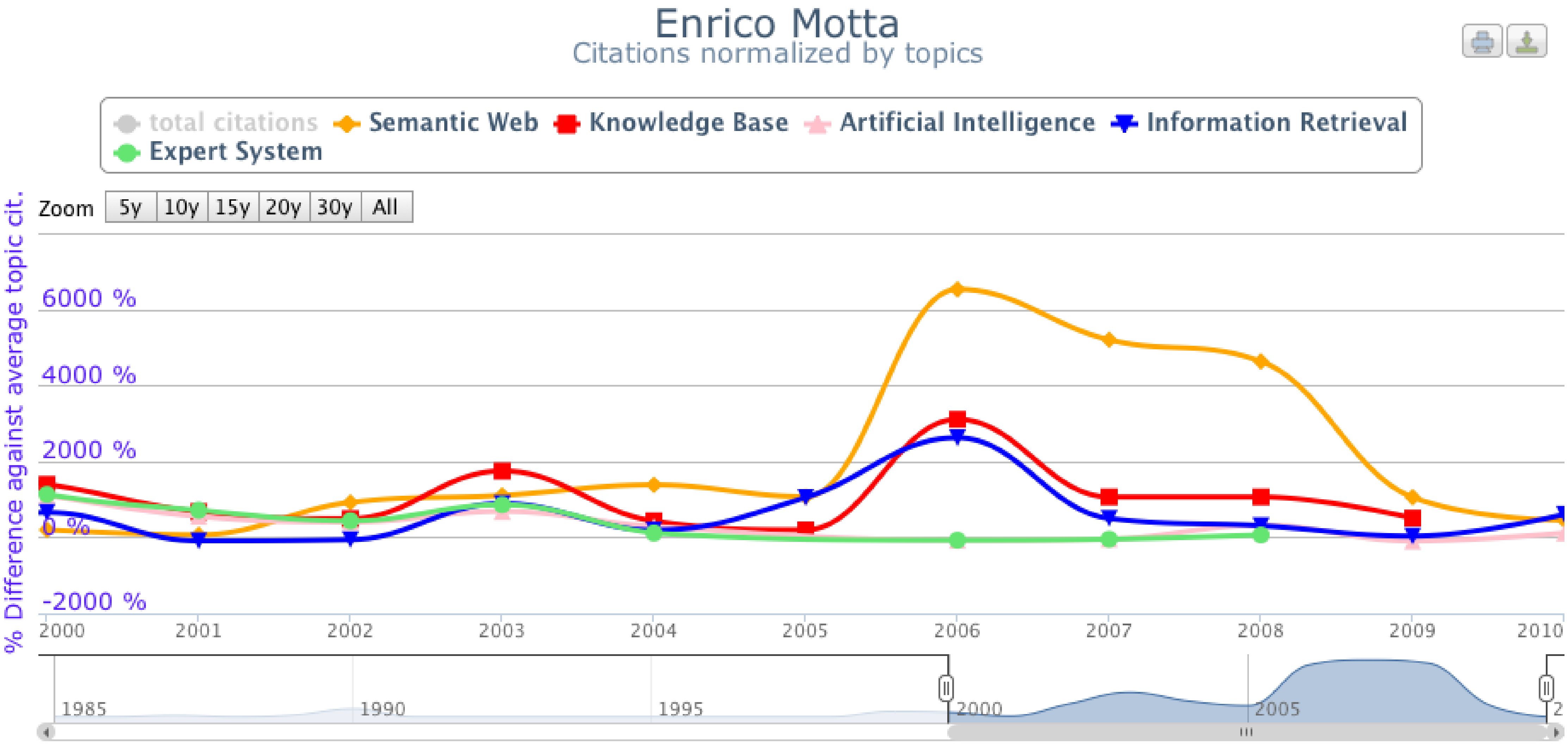
Career-young (1-5) people who have co-authored with Enrico and have expertise in machine learning, ranked in terms of #publications in this topic

Shared Research Trajectories

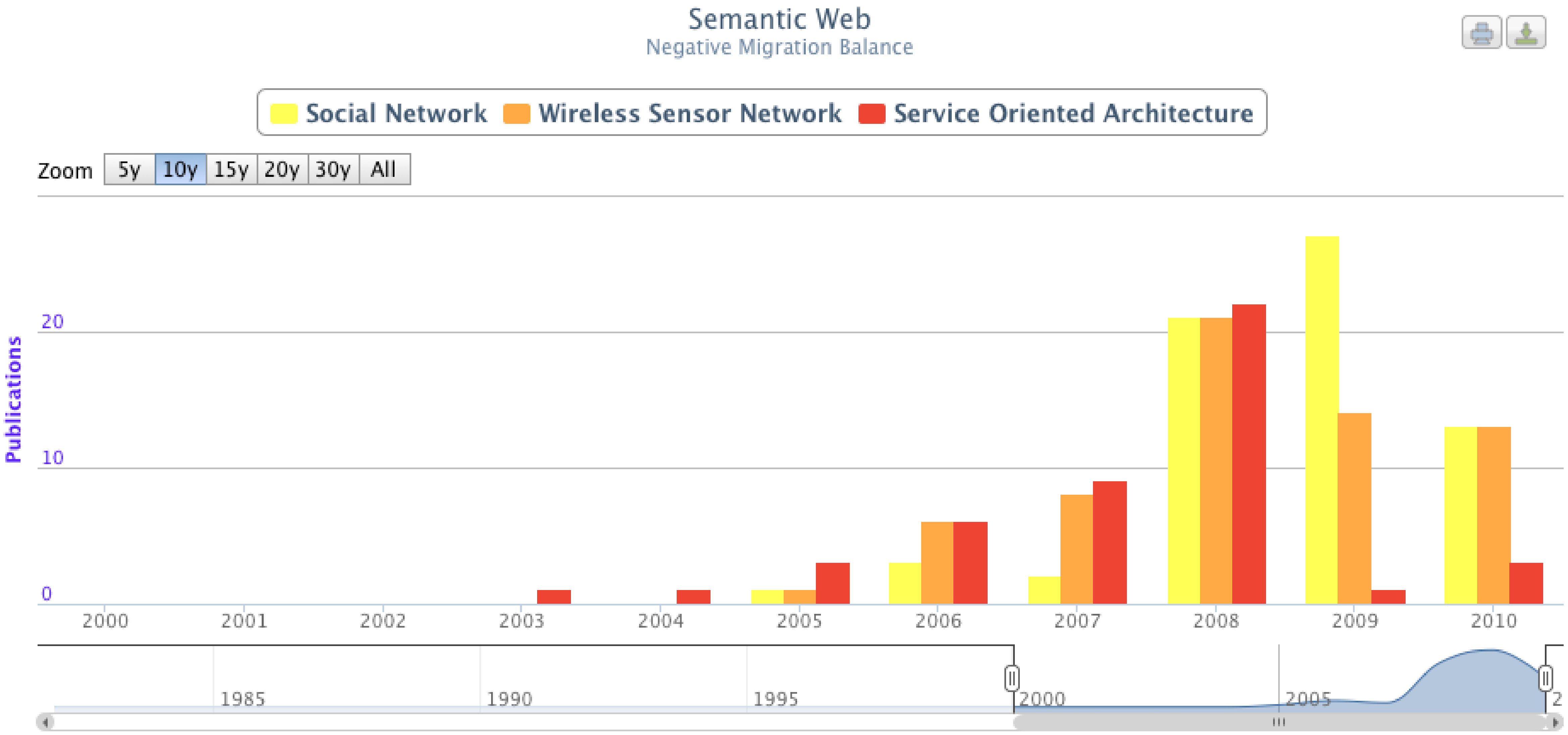


The authors who are most similar to Enrico with respect to the evolution of their research interests over time.

Normalised impact per topic over time



Where are SW authors going?...



Conclusions (1)

- Rexplore aims to provide an integrated solution to support tasks that require the exploration and analysis of scholarly data
- It does so by integrating a semantic foundation with statistical and visual analytics solutions

Conclusions (2)

- The fine-grained structure of research topics generated by Klink supports
  - Expert search, trend analysis, and exploration at a very fine grained level of granularity
  - The definition of fine-grained impact metrics, such as "citations in topics" or "normalised impact with respect to topic", which allow to measure very specific elements of academic impact

Conclusions (3)

- A rigorous empirical evaluation confirmed:
  - The effectiveness of the functionalities provided by the tool. 94% of the testers described Rexplore as “very effective”
  - The robustness of the tool with respect to tasks proposed by the users themselves. Rexplore was able to support satisfactorily 88% of the testers with respect to tasks proposed by them

KNOWLEDGE MEDIA

KMi

INSTITUTE