Burst the Filter Bubble

Using Semantic Web to Enable Serendipity

Valentina Maccatrozzo
The Network Insitute - Department of Computer Science

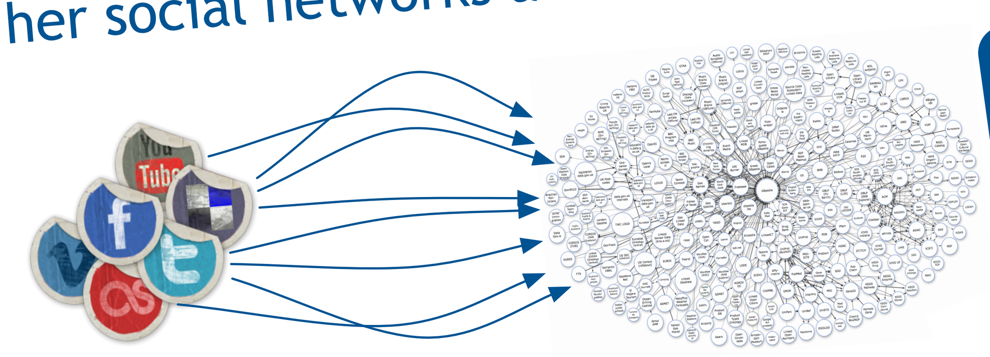
The problem
Personalization filters tend to show only information which agrees with the user's past viewpoint, effectively building a **filter bubble** around the user that tends to exclude contrary information.

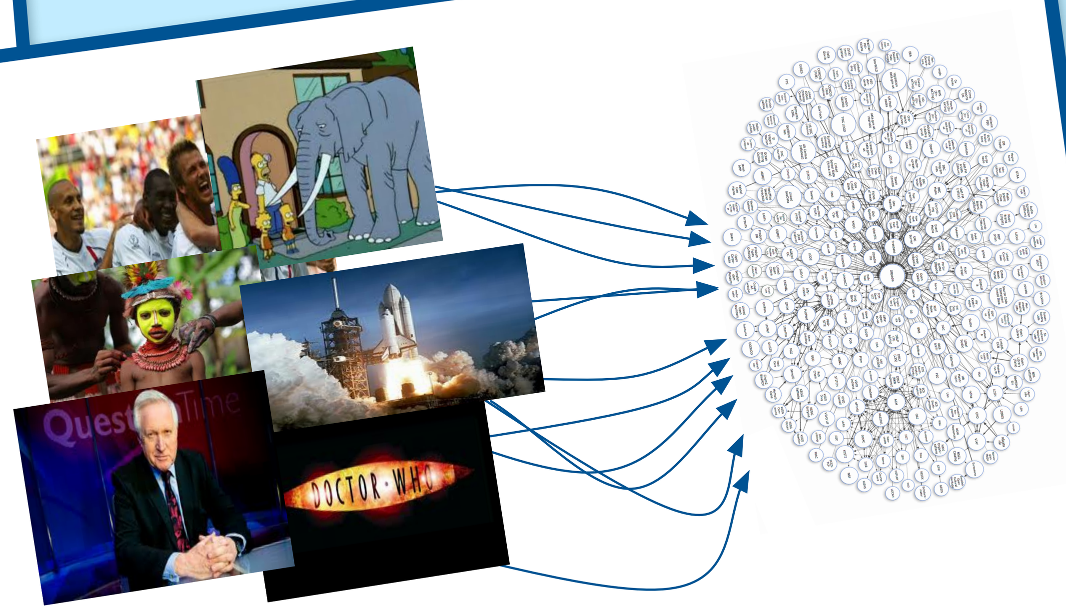
Users are missing the joy of **serendipity:** find something they did not know were looking for.

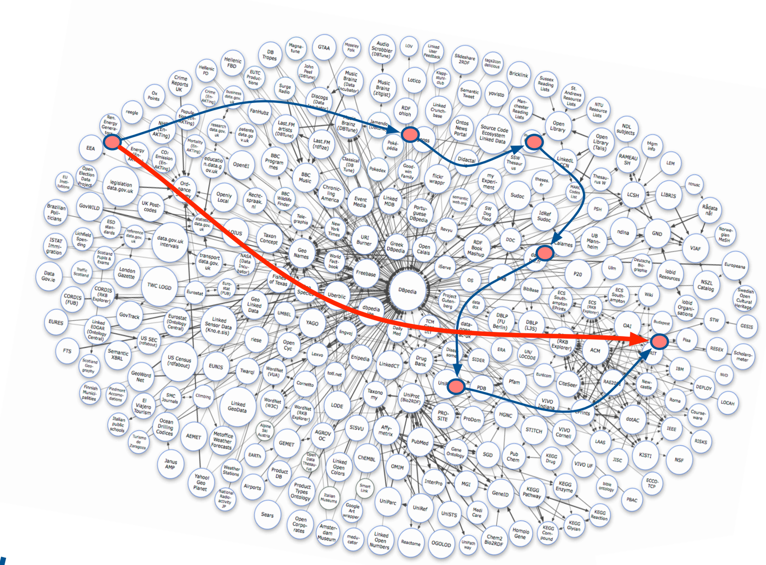
The solution
TV case

1

Jana watches TV and shares on social networks her thoughts

We build an up-to-date Jana's profile mapping her social networks activities to LOD

2

3

TV metadata is mapped to LOD too

4

We browse the LOD cloud to link new interests

5


We build an up-to-date and serendipitous EPG

The idea
We see the LOD cloud as a structured knowledge space covering a multitude of different domains. Creating new bindings between different knowledge elements can lead to the discovery of new knowledge paths.

We apply creative thinking models to the LOD cloud:
- ✓ variation & selection
- ✓ diverging & converging
- ✓ analogy & metaphor

The Network Institute
vrije Universiteit amsterdam

Collecting, Modeling and Sharing Green Knowledge

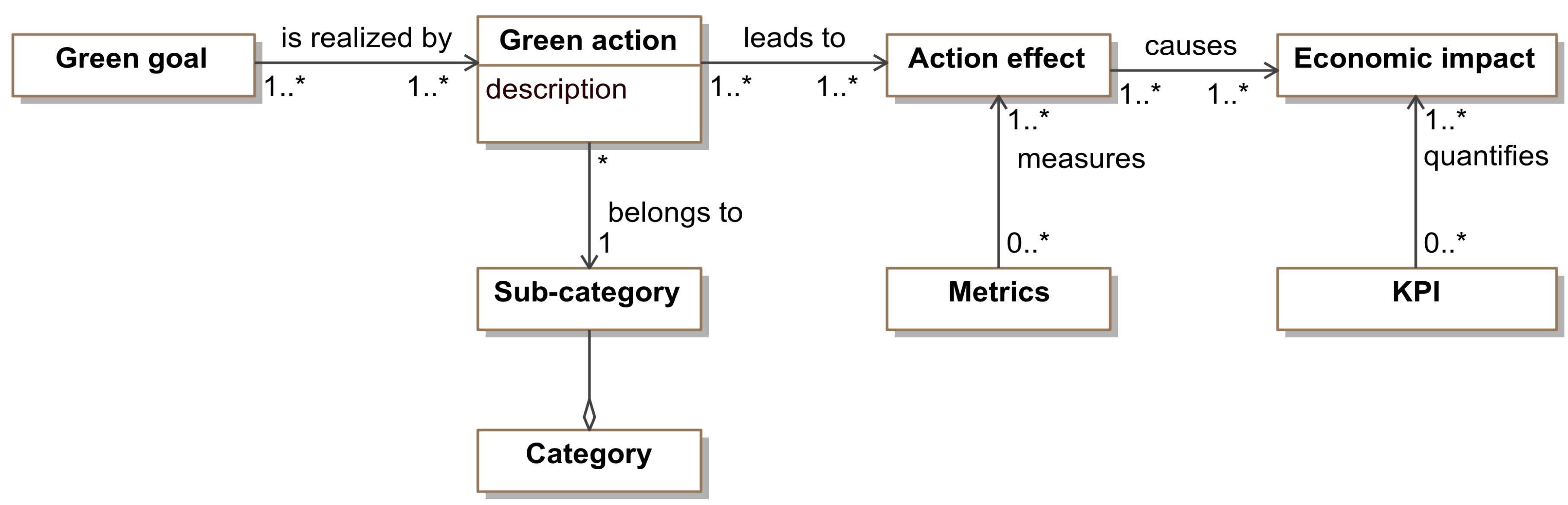
Qing Gu, Patricia Lago [q.gu, p.lago]@vu.nl
VU University Amsterdam

Sustainable green ICT requires alignment between economic impact and environmental effects.

Companies and educators lack reusable green practices including operational actions to re-green ICT, metrics and benchmarks, and experience testimonials.

The goal is to make available reusable practices for energy-efficient ICT systems and more sustainable ICT supported processes

An ontology for green practices



[Q. Gu, P. Lago & S. Potenza, Aligning Economic Impact with Environmental Benefits: A Green Strategy Model, Greens workshop, IEEE/ACM ICSE, 2012]

The ontology provides means to codify green practices and make explicit the link between green practices, their economic impact and environmental effects.

It enables researchers and practitioners to exchange green knowledge accumulated from academia and industry.

It may be used as an instrument for decision makers to justify and reason about the selection and adoption of green practices.

The Network Institute

Understanding the emerging networked world in its technological, economic and social aspects

De Boelelaan 1081a
1081 HV Amsterdam,
The Netherlands

Social & Sustainable Software Services (S4) Research Group

Knowlege Network Green Software (KNGS) aims to collect knowledge, raise awareness, present interesting studies.

The public-private cooperation to green IT, reduce CO2 emissions and build a Green Collar Economy in the Amsterdam region

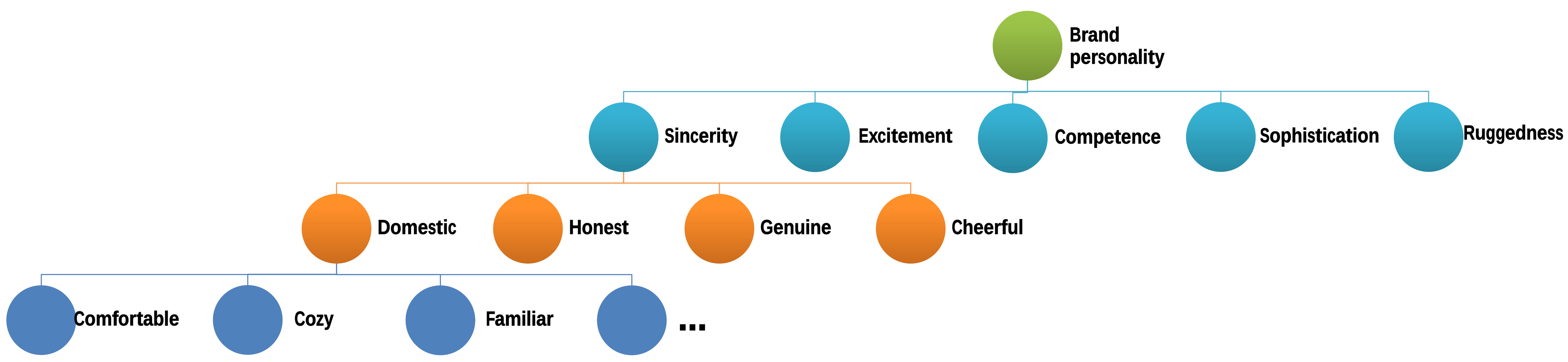
A subsidiary of a not-for-profit foundation SURF (Co-operative University Computing Facilities)

Decentralized Partnership Formation: Applications in Branding, Product Development and Product Distribution

Andreea Sandu, Bogdan Aurelian, Sophie Corlay, Anna Chmielowiec, Ivar Vermeulen, Maarten van Steen
{a.i.sandu, b.aurelian, s.j.s.corlay}@student.vu.nl, {a.a.chmielowiec, i.e.vermeulen, m.r.van.steen}@vu.nl
Vrije Universiteit, Amsterdam, The Netherlands

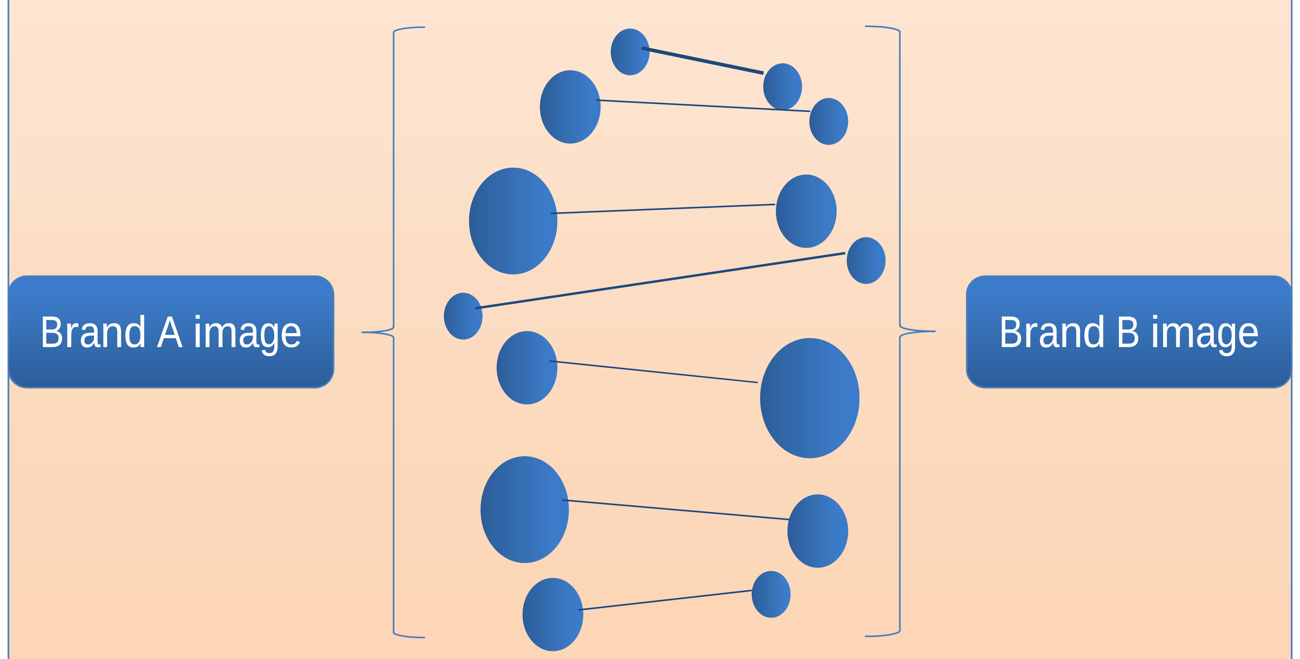
Web mining for co-branding potentials

1.Extracting brand images – by determining the strength of the associations between brand names and several attributes on the web. These attributes include the facets of the five core brand personality dimensions, expanded with selected synonyms from a thesaurus.

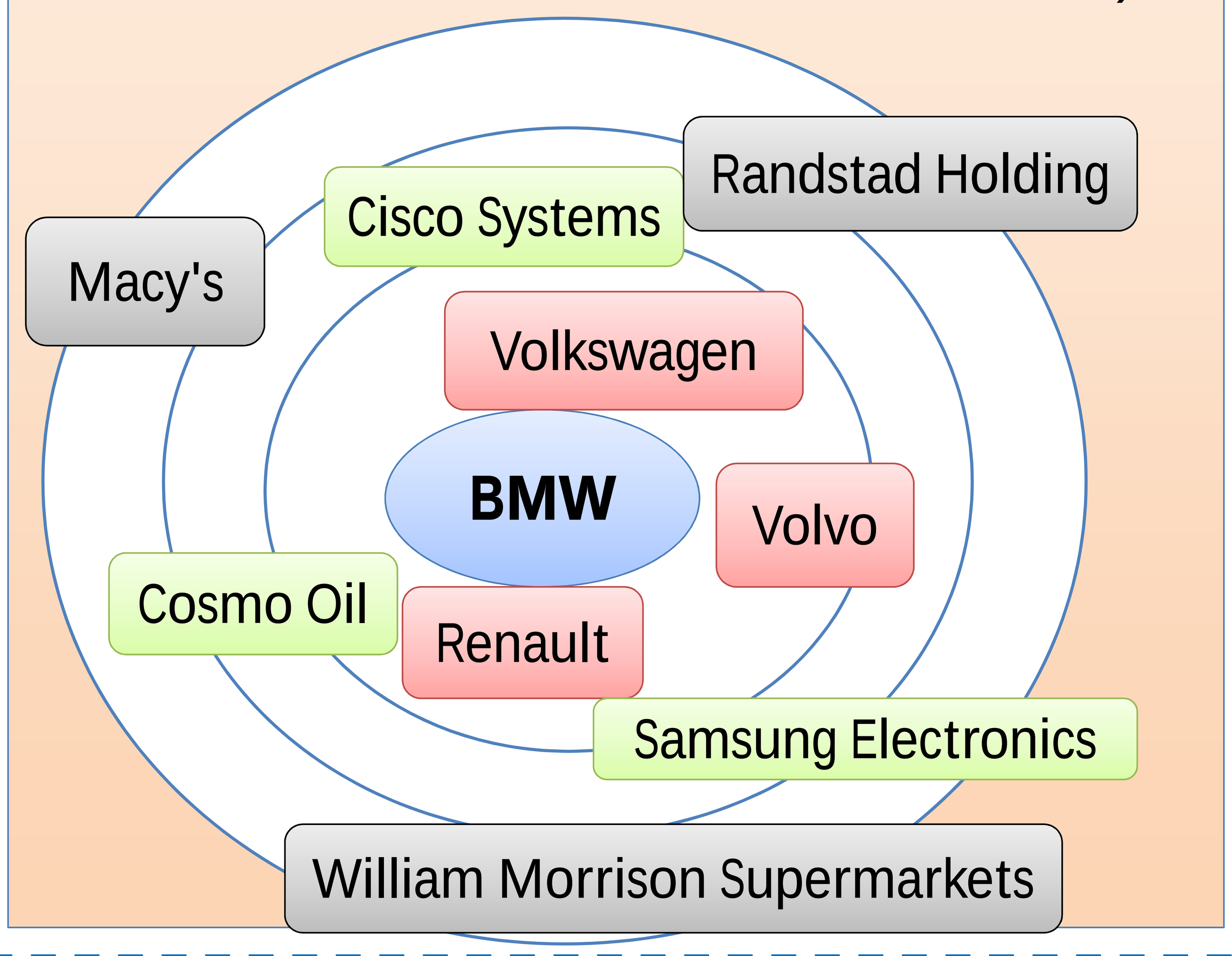


2. Assessing brand associations

Symbolic overlap - whether two brand images have ties of similar strength to their corresponding attributes



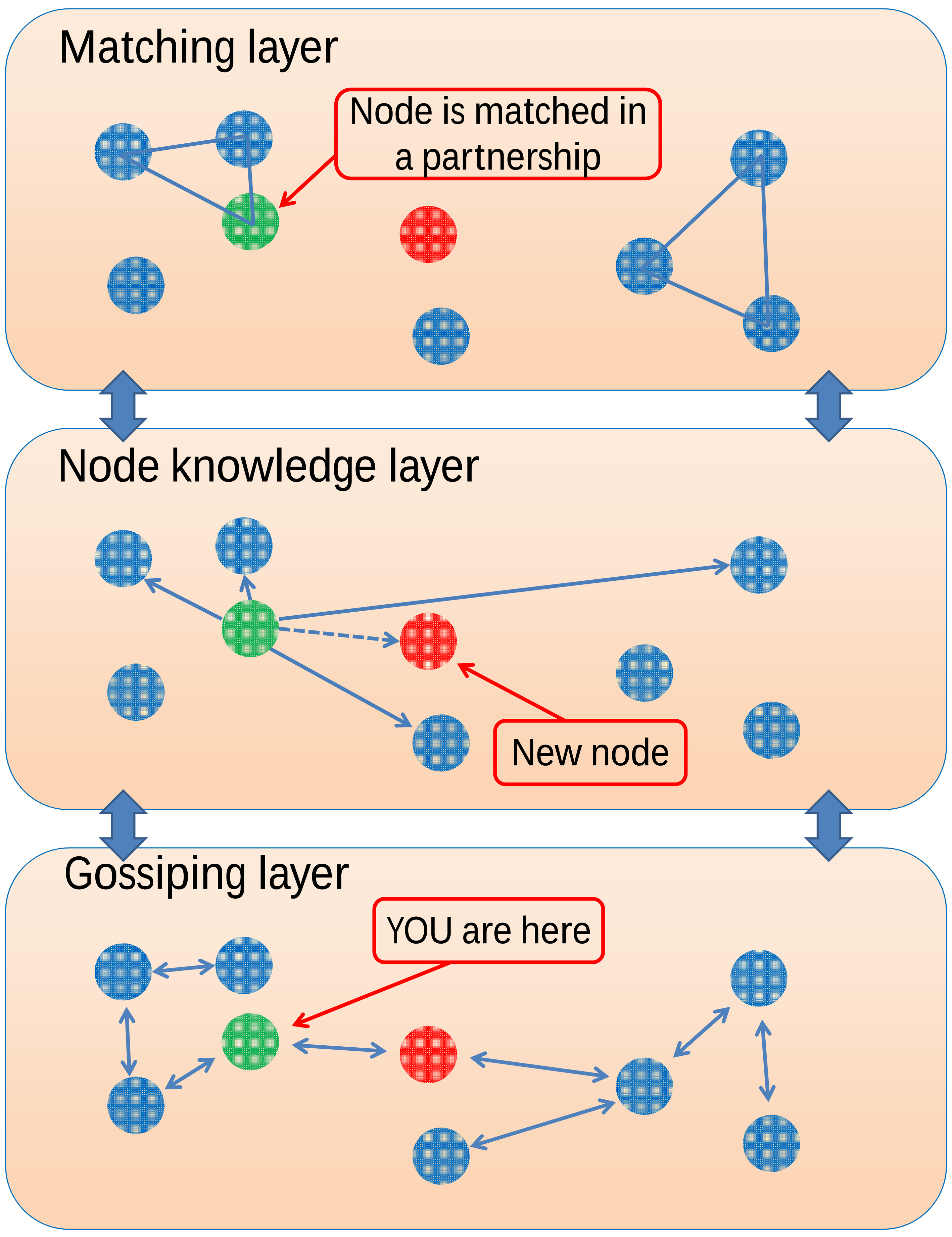
Functional overlap – the competitive pressure that two brands exert on each other (based on their number of co-occurrences on the web)



3. Computing the co-branding potential

- ❖Combines the two brand association metrics such that a high symbolic overlap and a moderate functional overlap lead to a high co-branding potential
- ❖A high symbolic fit ensures compatibility between the two brand images
- ❖ A moderate functional fit ensures complementarity of the contribution of the two brands to the marketed product

Decentralized matching



1. Matching protocol

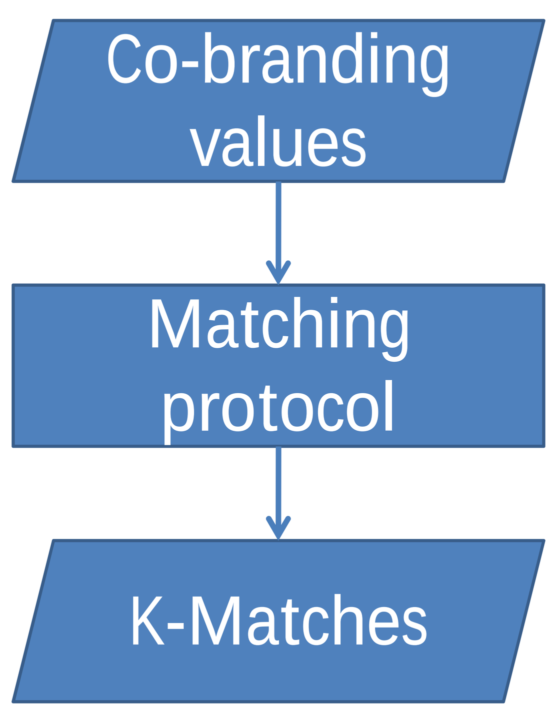
- ❖Finds groups of k partners
- ❖Each partner which forms a matched group cannot be part of another group
- ❖Uses co-branding values to assess the partnership value
- ❖Uses heuristic to avoid searching through all the solution states
- ❖Matching is run independently at the site of each company

2. Node knowledge layer

- ❖Provides stable node list to the matching protocol
- ❖Saves all known nodes and updates constantly with new nodes
- ❖Samples peers from the gossiping protocol

3. Gossiping layer

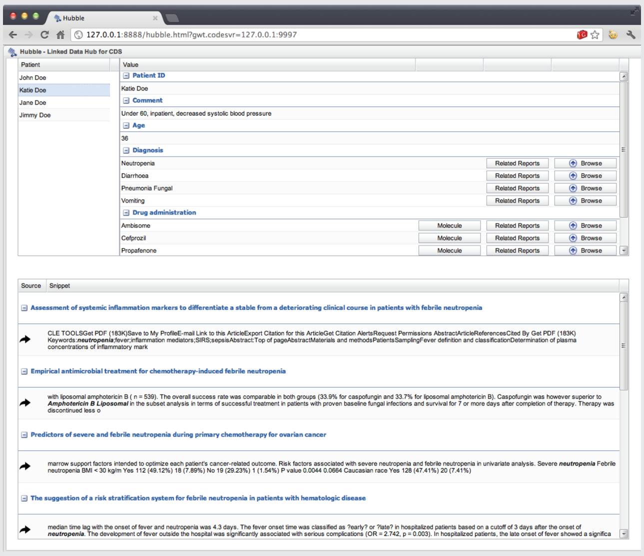
- ❖Every company is free to join or leave the matching process
- ❖The number of companies taking part in the process can become very large
- ❖Who keeps track of the companies and at what cost?
- ❖Gossiping solves this problem because it is decentralized and every company keeps track of its own set of potential partners



HUBBLE Linked Data Hub for Clinical Decision Support

Rinke Hoekstra[1,3], Sara Magliacane[1], Laurens Rietveld[1], Gerben de Vries[2], Adianto Wibisono[2] and Stefan Schlobach[1]

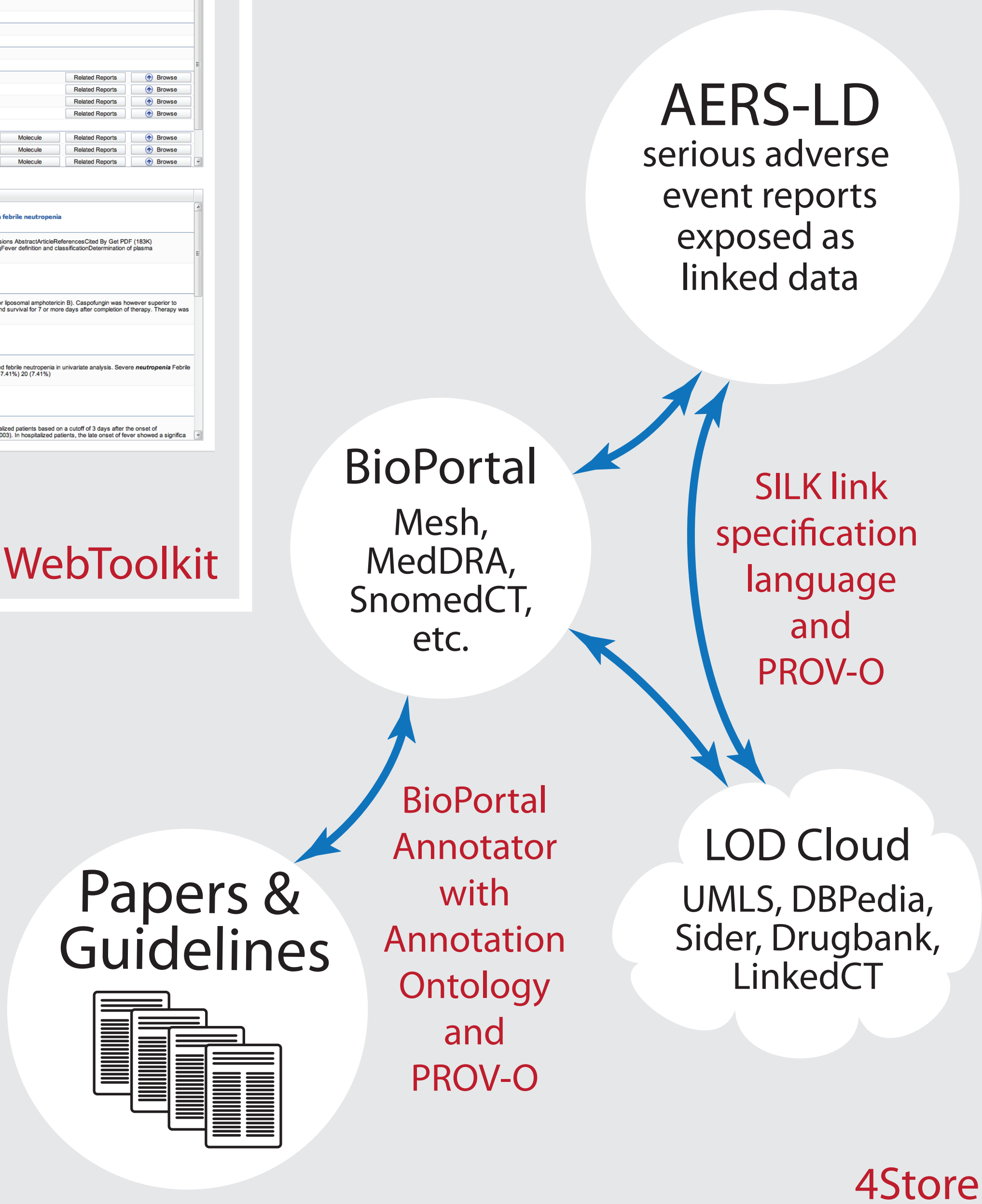
[1]Department of Computer Science, VU University Amsterdam, The Netherlands
[2]Department of Computer Science, University of Amsterdam, The Netherlands
[3]Leibniz Center for Law, University of Amsterdam, The Netherlands



Google WebToolkit

Hubble demonstrates three ‘sales pitches’ of linked data: inter-operability, interlinking and tool availability.

From patient to:
- Relevant publications
- Related adverse events
- Clinical trials
- Drug information
- Known side effects
- Statistical analysis



4Store

PHILIPS

COMMIT/

VU UNIVERSITY AMSTERDAM

UNIVERSITY OF AMSTERDAM

ELSEVIER

data2semantics

aers.data2semantics.org
www.data2semantics.org

Semantically Mapping Science: Results

Collaborators

- Computer Science: Paul Groth, Shenghui Wang, Ravindra Harige, Stefan Schlobach, Frank van Harmelen
- Organization Science: Peter van den Besselaar, Julie Birkholz
- Rathenau Institute: Thomas Gurney, Edwin Horlings

Web: http://www.sms-project.org

Project Motivation

Scientometrics is the field of Social Sciences that studies the evolution of scientific fields: how they grow, shrink, merge, appear or disappear, if they are inward or outward-looking, how they are clustered, if they have a high or low in- and outflux of people etc.

Typically, Scientometrics studies are done on the basis of bibliometric data: co-citation patterns, co-authoring pattersn, citation-impact studies, etc. The field has progressed rapidly since the widespread on-line availability of such bibliometric data (in the last 15 years or so). Such studies can now be done routinely.

However, publishing is only one of the many activities of scientists. They also do things like: review papers, have discussions, change jobs, interact with companies, organise and participate in events, are members of boards (conference, professional organisations), etc. With the advent of the Web, these other activities of scientists now also leave on-line traces that can be used for scientometrics purposes. The question is:

Can we use Semantic Web techniques to meaningfully detect, retrieve and manipulate such web-traces of activities of scientists in order to improve Scientometrics studies?

Results

The project ran from September 2009 – 2011. There were three core areas of results.

1) Using the Web for Science Studies

During the project, we icataloged web-based data sources that would provide new insights into science. Data sources included web-crawls (e.g. science blogs), web-available databases (e.g. DBLP), and APIs (e.g. Microsoft Academic Search, Yahoo Geolocation). These were then transformed into data usable for statistical and network analysis..

2) Developing New Semantic-Web based Methods

We developed new methods for being able to both acquire and analyze network data. Network data is a key data input for studying science dynamics. This work led to a best paper award at the International Semantic Web Conference.

3) Bootstrapping a new community

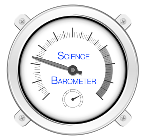
Working with collaborators in the US and Europe, we helped support the creation of a new community to study science impact measures based on the social web. Altmetrics has received media attention in Nature, Times Higher Education, Forbes and The Chronicle for Higher Education.

The Future

The project has established a strong collaboration within the VU. This is supported by the Network Institute through its student assistant program. Current work includes:

- application of developed methods to new science datasets;
- development of Semantic Web wrappers for Microsoft Academic Search and the CORDIS dataset of european grants;
- use of the created datasets for studies of interdisciplinary collaboration

The project team is currently pursuing funding to expand on the initial work of the project in order to create a *Science Barometer.* A robust tool for studying scientific activity as it happens through the use of web data sources.

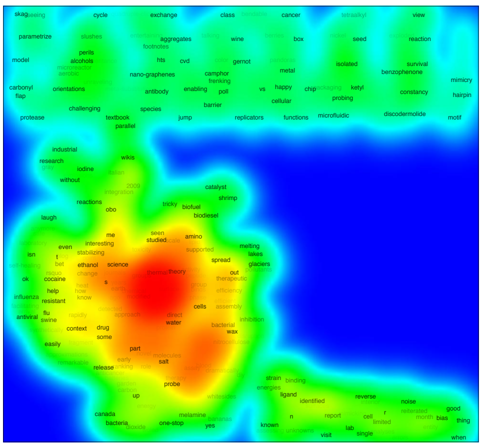


Examples

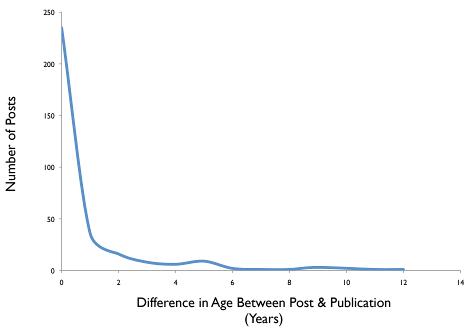
Science Studies: Chemistry Blogging

We studied 336 blog posts on chemistry from the blog aggregator site researchblogging.org. Each post at this site is required to have a citation to the published literature. Through this connection we were able to study the posts using biblometric techniques. Some results are below

Map of Blog Descriptions of Topically Similar Papers. Terms are grouped together according to the papers they discuss. Hotter colors denote more papers in a topic.



A plot showing the difference in age between when a blog post was made and when the paper the post cites was made. Blog posts are much more immediate.



Paul Groth, Thomas Gurney (2010) Studying Scientific Discourse on the Web using Bibliometrics: A Chemistry Blogging Case Study. In WebSci10: Extending the Frontiers of Society On-Line

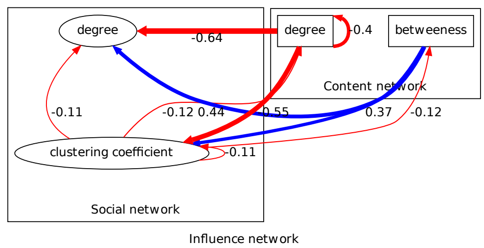
Method: Measuring influence between content and social networks

We developed a general framework for measuring the dynamic bi-directional influence between communication content and social networks. The framework leverages the idea that knowledge about both kinds of networks can be represented using the standard Semantic Web knowledge representation standards.

Application: World Wide Web Conference Proceedings

Question: How does the status of an authorship network influence a topic network over time and vice versa?

Content Network:	Topic network as represented by keyword co-occurrence
Social Network:	Co-authorship network
Binding:	Networks arebound via article
Data Source:	Publications of the WWW Conference
Time series:	four years, from 2007 to 2010



Influence network

Shenghui Wang and Paul Groth. 2010. Measuring the dynamic bi-directional influence between content and social networks. In Proceedings of the 9th international semantic web conference on The semantic web - Volume Part I (ISWC'10) – Best Paper Award

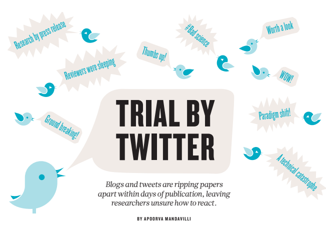
Community: altmetrics

altmetrics is the creation and study of new metrics based on the Social Web for analyzing, and informing scholarship.

Summarized in: J. Priem, D. Taraborelli, P. Groth, C. Neylon (2010), Alt-metrics: A manifesto, (v.1.0), 26 October 2010. http://altmetrics.org/manifesto

To support the development of the altmetrics community, we helped organize the following activities:

- How is the Web changing Scientific Impact? Science Online 2010
- Altmetrics11 workshop at Web Science 2011
- Altmetrics12 workshop at Web Science 2012
- PLOS One Collection on Altmetrics



Altmetrics appeared in this article in Nature Volume 469