



# Using KOS as a basis for text analytics and trend forecasting

**NKOSS Networked Knowledge Organization Systems and Services**  
**The 9th European NKOS Workshop at the 14th ECDL Conference,**  
**Glasgow, Scotland**  
**10th September 2010**

Marjorie M.K. Hlava  
President  
[mhlava@accessinn.com](mailto:mhlava@accessinn.com)  
[www.accessinn.com](http://www.accessinn.com)

# Agenda

## ❖ IEEE Challenge

- ❖ Where are our publication strengths?
- ❖ What are the emerging topics?
- ❖ Use our own data to address these question

## ❖ Access Innovation's Response

- ❖ Expanding and Mapping the IEEE Thesaurus
- ❖ Use term instead of text analytics to investigate

## ❖ Findings



The world's leading professional association  
for the advancement of technology

[Home](#) | [Shop](#) | [Join](#) | [myIEEE](#) | [Contact](#) | [Sitemap](#) | [IEEE Xplore](#)

Google™ Custom Search

Search »

[About Us](#) | [Membership](#) | [Publications](#) | [Conferences](#) | [Standards](#) | [Careers](#) | [Education](#) | [Volunteers](#) | [Societies](#) | [Geographic Activities](#)

Text size » A A A

## Advancing Technology for Humanity

Find an article in the IEEE Xplore digital library

Search »

Sign in to view your IEEE Member information  
through [myIEEE](#) »

[Forgot your web account username or password?](#)

### IEEE News

**IEEE Spectrum Wins Grand  
Neal Award for Excellence  
in Business Journalism**

[Read More»](#)

**AMA and IEEE Announce  
First Conference on Medical  
Technology in  
Individualized Healthcare**

[Read More»](#)

**IEEE Announces Major  
Upgrade of IEEE Xplore**

### Inside IEEE

**Conferences, Meetings, and  
Events**

Locate [upcoming conferences](#) or  
find [volunteer/corporate meetings](#)  
and [society, section & chapter  
activities](#) in your local area.

[Learn More»](#)

**Education**

Register for [online learning](#)  
programs like [IEEE Expert Now](#)  
and [Education Partners](#). Or find  
other [IEEE education activities](#)

### Product Highlights

**The New IEEE Xplore Digital  
Library**

New interface, new search  
tools, better research results

[Learn more»](#)

[Visit the new IEEE Xplore»](#)

**IEEE/IET Electronic Library**

The complete collection for  
corporations and institutions

[Learn more»](#)

[Organizational trial»](#)

### What do you want to do?

- [Renew membership](#)
- [Join IEEE or an IEEE Society](#)
- [Donate to Haiti fund](#)
- [Set up IEEE Web account](#)
- [Log in to My Account](#)
- [Find an IEEE standard](#)
- [Get member e-mail alias](#)
- [View IEEE.tv](#)
- [Obtain member insurance](#)
- [Locate a job](#)
- [Find student resources](#)



# About IEEE...

- ❖ Founded in 1884, IEEE is the world's largest professional association advancing technology for the benefit of humanity.
- ❖ We publish 148 technical journals, transactions and magazines, sponsor nearly 800 conferences annually, develop technology standards, and support the professional interests of more than 400,000 members in over 160 countries.
- ❖ Members participate in 38 societies and 7 councils
- ❖ The **IEEE Xplore® digital library** provides access to IEEE journals, transactions, letters, magazines and conference proceedings, IET and other 3<sup>rd</sup> Party journals and conference proceedings, IEEE Standards and IEEE educational courses.
  - Over 2.5 million documents

# The New IEEE Xplore

[IEEE.org](#) | [IEEE Xplore Digital Library](#) | [IEEE Standards Association](#) | [Spectrum Online](#) | [More IEEE Sites](#)

**IEEE Xplore®**  
DIGITAL LIBRARY

Delivering full text access to the world's highest quality technical literature in engineering and technology



**BROWSE**

[Journals & Magazines](#)  
[Conference Proceedings](#)  
[Standards](#)  
[Books](#)  
[Educational Courses](#)  
[Technology Surveys](#)

**MY SETTINGS** ▾

**CART**

**SIGN IN** ▾

[Feedback](#) [?](#) [Help](#)

**Search 2,590,153 documents**

 **SEARCH**

[Advanced Search](#) | [Preferences](#) | [Search Tips](#)

**SIGN IN** [?](#)

*Username*

*Password*

**Sign In**

[Forgot Username/Password?](#)

[Register](#)

**QUICK LINKS**

**Highlights**



**Welcome to the New IEEE Xplore!**

New Interface, New Search Tools, **Better Results.**

- Spend less time researching with a powerful, new search engine
- Find better information with faceted search results
- Personalize your account with custom saved searches, RSS feeds and e-mail alerts

# Specific Challenges

- ❖ Is there a way, using IEEE information, to forecast future direction?
- ❖ Where is the industry headed? What about by technology sector?
- ❖ Does our coverage match the IEEE mission and vision?
- ❖ Can IEEE become smarter about their data and potential markets using their collection in new ways?  
Are the societies publishing and talking about what their charter indicates they cover?
- ❖ What are the trends – are topics emerging/cooling?
- ❖ Can IEEE use technology and their own data to explore these questions while enhancing their data?

# Access Innovation's Response

- ❖ SciTech Strategies' Maps
- ❖ Access Innovation's Tools
- ❖ IEEE Xplore data
- ❖ Test with several thesauri

# Access Innovations / Data Harmony

- ❖ Founded in 1978
- ❖ Suite of Semantic Enrichment tools
- ❖ Updated the IEEE Thesaurus in 2005
- ❖ Built a rule base to auto index IEEE content
  - ❖ “90 % accuracy out of the box on journal data”\*
  - ❖ “80% out of the box on proceedings data”\*
  - Auto indexed 1.2 million Xplore records
    - ❖ With the IEEE thesaurus terms rule base
    - ❖ With the MeSH rule base
    - ❖ With DTIC rule base

\*Adam D. Philippidis, Manager, Indexing & Database Production, IEEE



# SciTech Strategies, Inc.

- ❖ Founded in 1982 (Center for Research Planning)
- ❖ Bibliometric Modeling of very large datasets
  - ❖ Thomson/ISI data (1982-2004)
  - ❖ Elsevier/Scopus data (2004-present)
- ❖ Focus on Accuracy
  - ❖ Disciplines > Research Communities > Researchers

# Disciplinary Map of Science

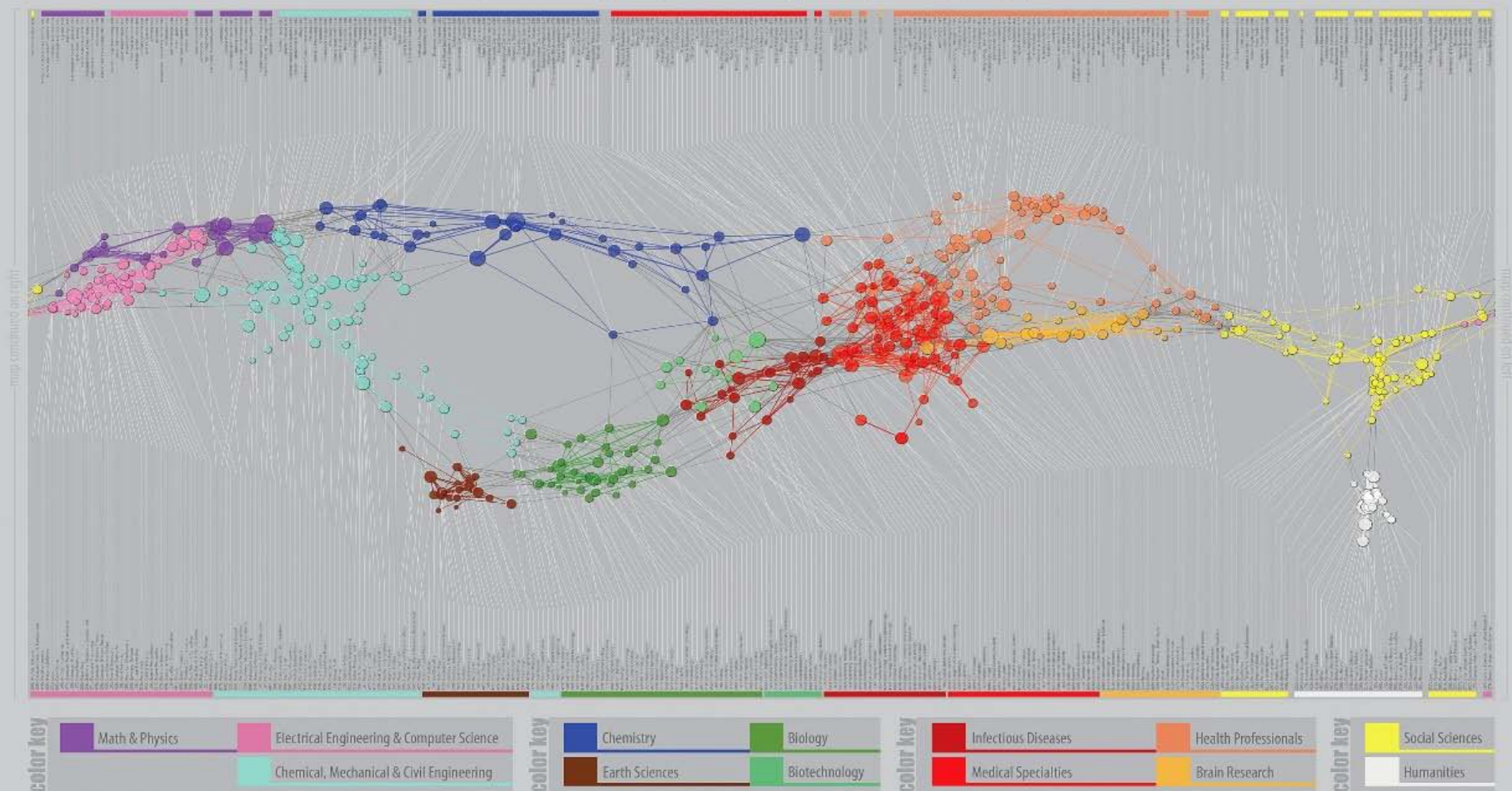
These maps were constructed by sorting more than 10,000 journals into disciplines, disciplines, specialties, or sub-specialties, and then plotting each journal on a sphere. The spheres are then projected onto a 2D plane, creating a map of science. The maps are based on a hierarchical clustering algorithm that groups journals into disciplines, specialties, or sub-specialties. The hierarchical clustering algorithm is based on the hierarchical clustering of journals into disciplines, specialties, or sub-specialties. The hierarchical clustering algorithm is based on the hierarchical clustering of journals into disciplines, specialties, or sub-specialties.

These maps are based on a hierarchical clustering algorithm that groups journals into disciplines, specialties, or sub-specialties. The hierarchical clustering algorithm is based on the hierarchical clustering of journals into disciplines, specialties, or sub-specialties. The hierarchical clustering algorithm is based on the hierarchical clustering of journals into disciplines, specialties, or sub-specialties.

## MAPS OF SCIENCE

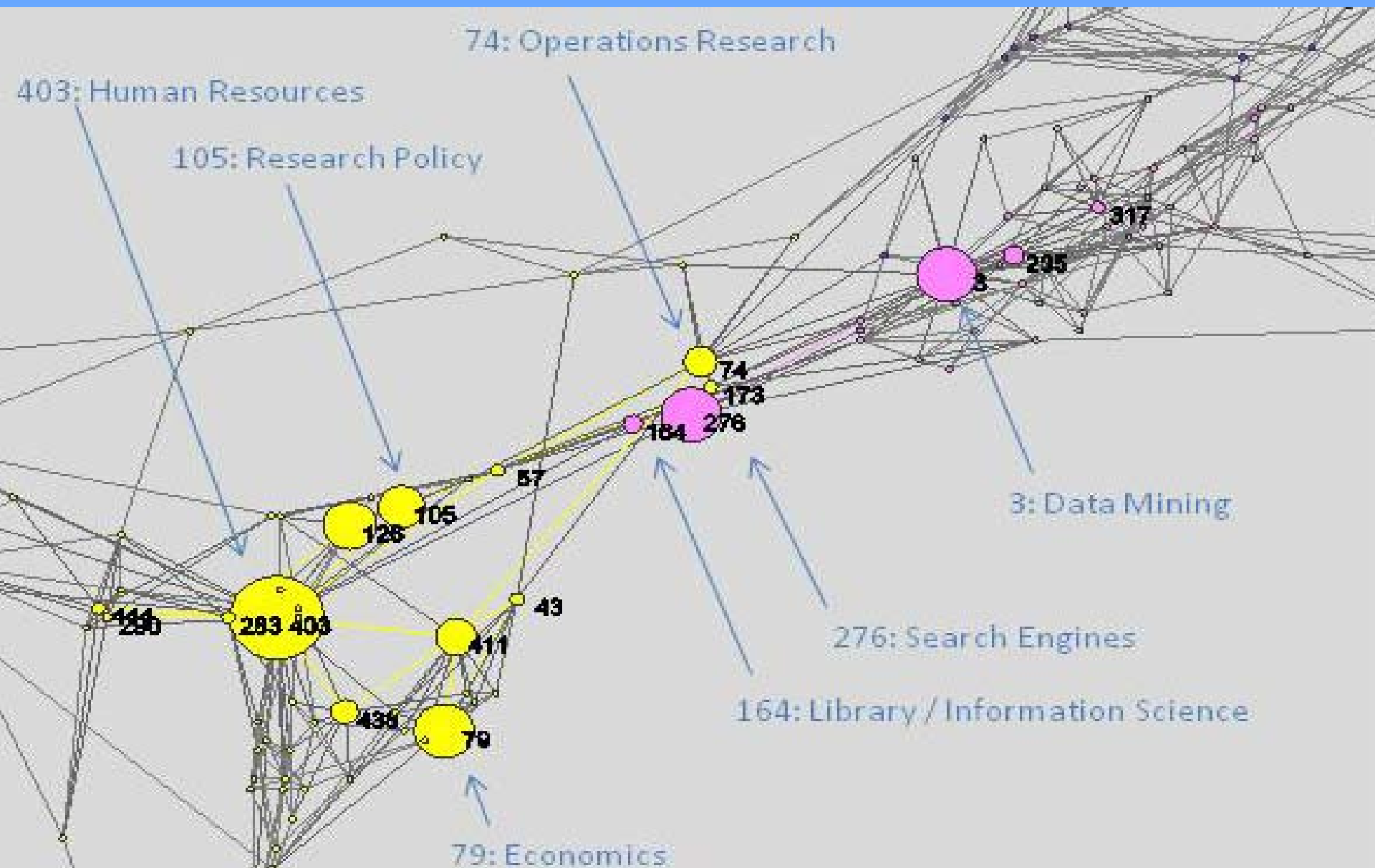


A visualization of 7.2 million scholarly documents appearing in over 16,000 journals, proceedings or symposia between Jan, 2001 and Dec, 2005.

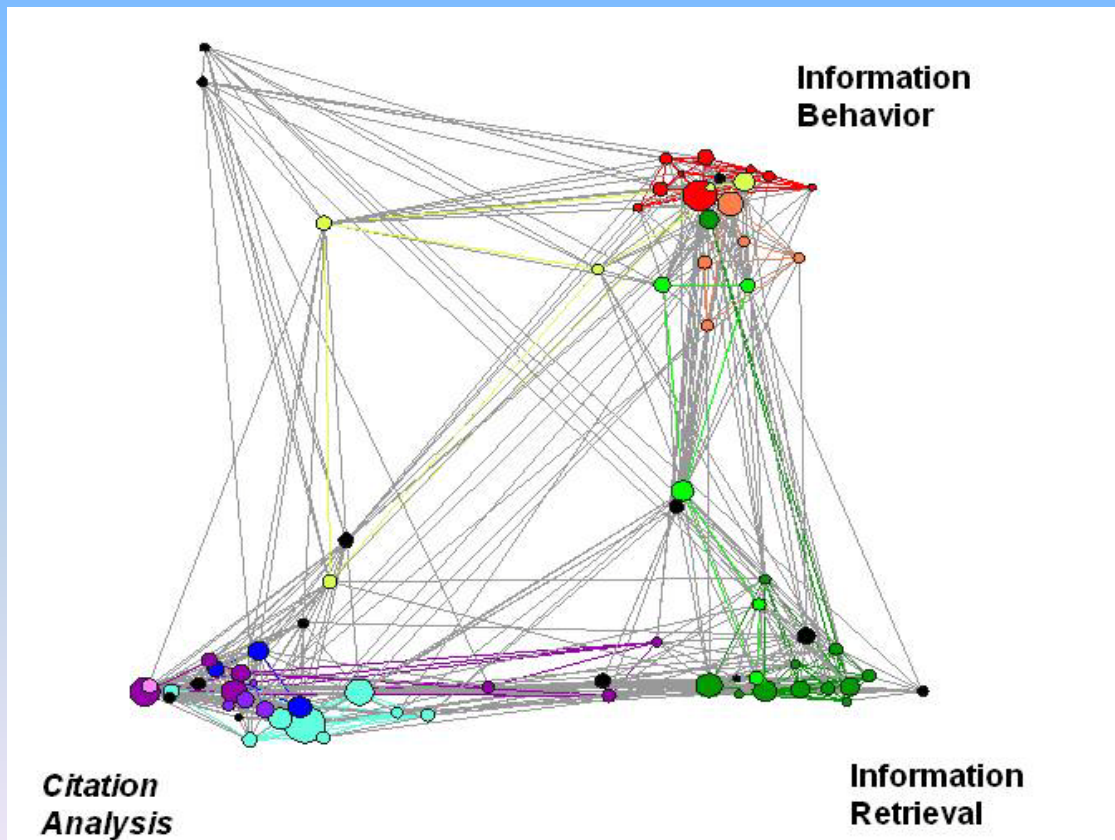
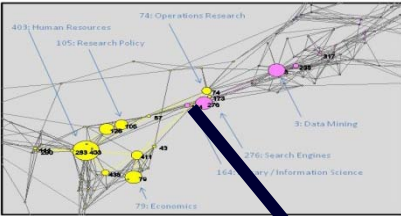


See <http://www.accessinnovations.com> for more information. © 2010 Access Innovations, Inc. All Rights Reserved. This visualization is based on data from the ISI/Clarivate Analytics database. The data is provided for informational purposes only and is not intended to be used for any other purpose. The data is provided for informational purposes only and is not intended to be used for any other purpose. The data is provided for informational purposes only and is not intended to be used for any other purpose.

# Relevant Disciplines for Science Mapping

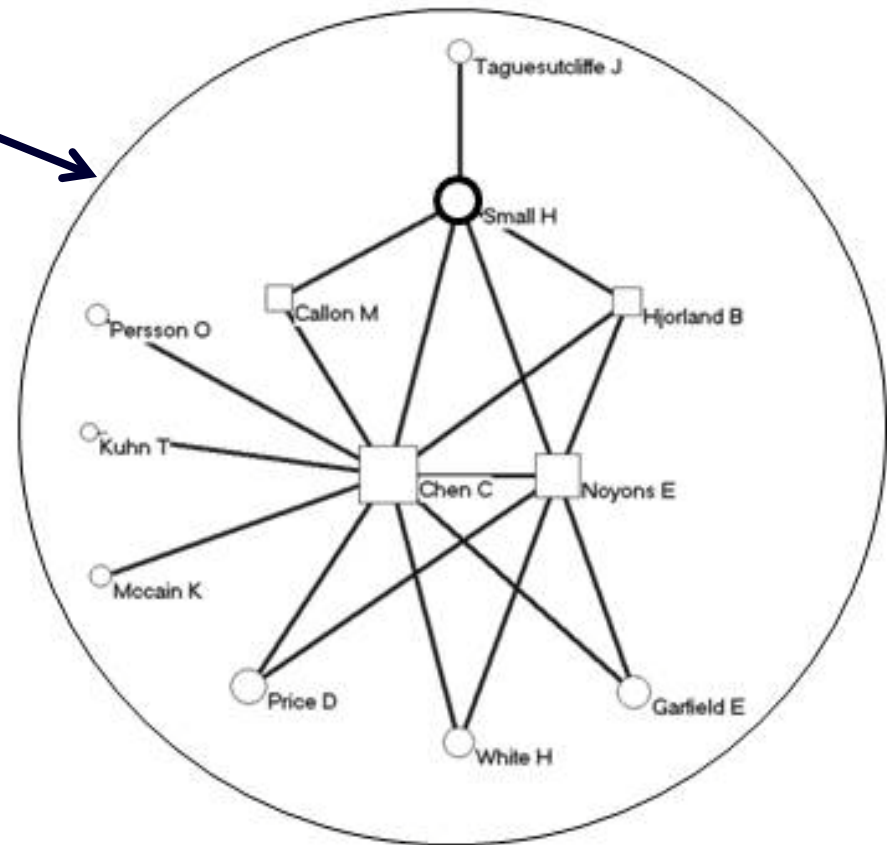
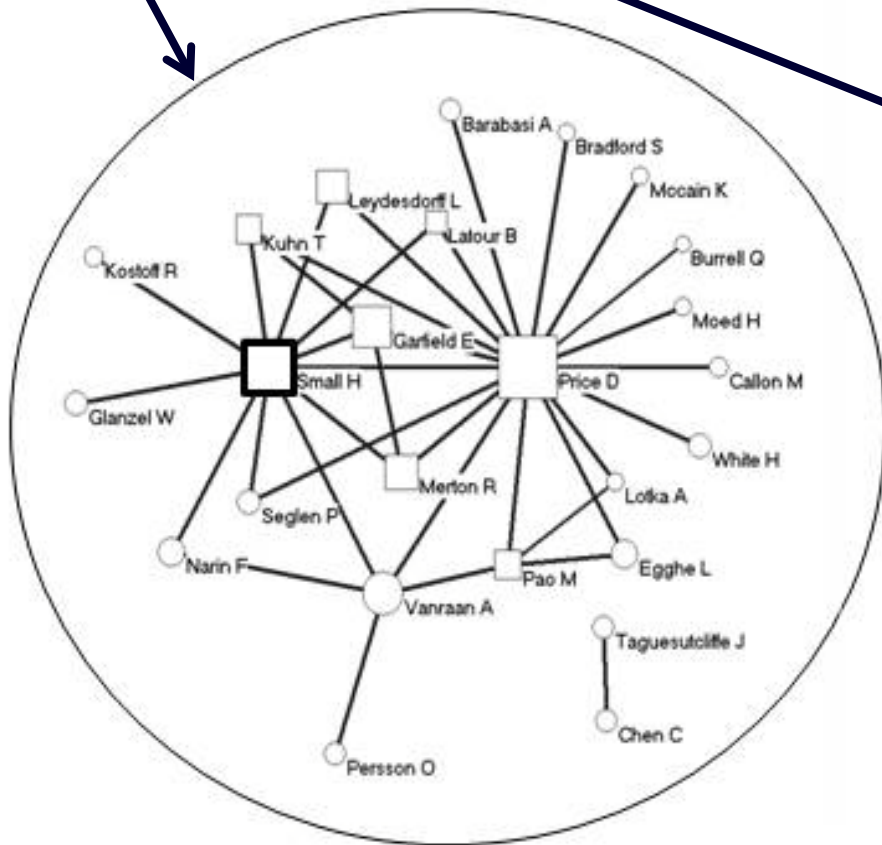
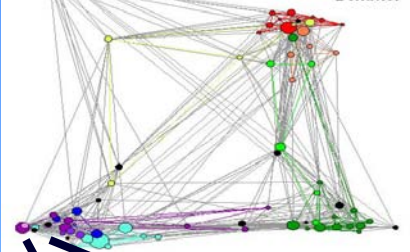


# Research Communities in the Information Science Discipline





# Intellectual Base of Two Research Communities



# From a society / publisher perspective visual answers

- Which topical areas form our core?  
Periphery?
- Where is the coverage dense? Thin?
- Which topical areas are most active? Least active?
- Which topical areas seem to be emerging?  
Declining?
- Which topical areas are interrelated? Isolated?
- What are the overlaps between journals / segments?
- Where are the potential expansion points?

# Questions with visual answers

- ❖ From a thesaurus perspective
- ❖ What terms are too broadly defined?
- ❖ How do actual topical relationships differ from the thesaurus structure?

# Preparing the data

- ❖ Index 1.2 Million Xplore records
  - ❖ Using the IEEE Thesaurus
  - ❖ Using the MeSH - Medical Subject Headings
  - ❖ Using the DTIC Thesaurus
- Normalize and enrich the XML as needed
- Create an XML / SQL Database



# Mapping IEEE thesaurus space

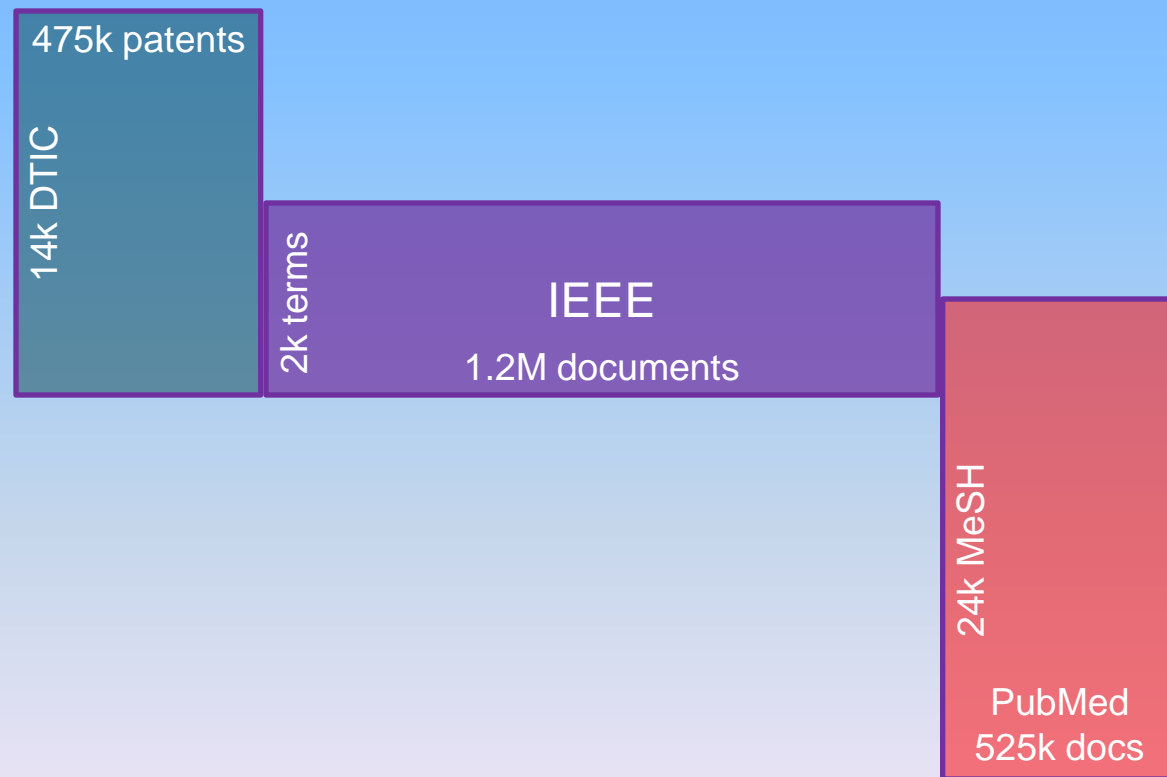
- ❖ We are more interested in an expanded map that includes adjacencies to the IEEE data
  - ❖ Expanded term set shows adjacent white space; opportunities for expansion
  - ❖ Similar process to that for simple map except ...
  - ❖ We need additional terms to add

# Mapping IEEE thesaurus space

- ❖ Criteria for additional terms
  - ❖ Low occurrence rate in IEEE documents
  - ❖ Linkage to terms in IEEE documents
  - ❖ Similar level of detail to current IEEE thesaurus terms
- ❖ Where do we find these terms? How can we add them?

# Defining expanded term space

## 1. Select related corpus'



# Defining expanded term space

## 2. Identify related terms



# Defining expanded term space

## 2. Identify related terms



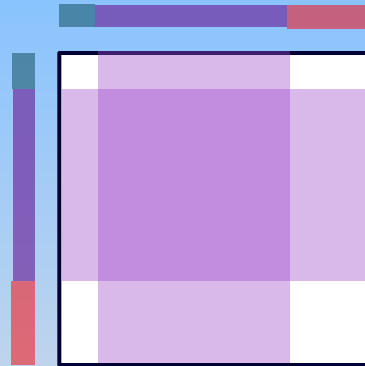
# Defining expanded term space

## 3. Resulting term set



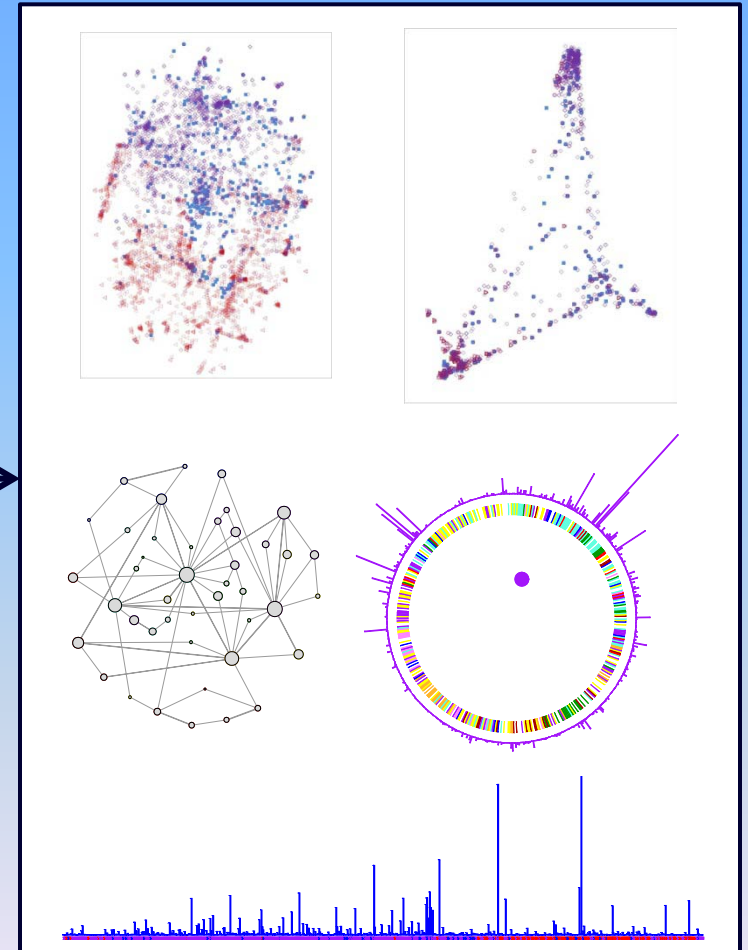
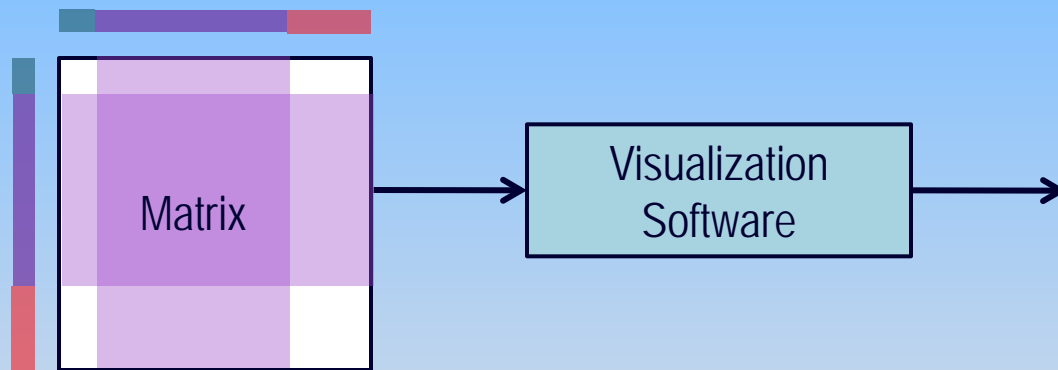
# Defining expanded term space

## 4. Term:Term Matrix



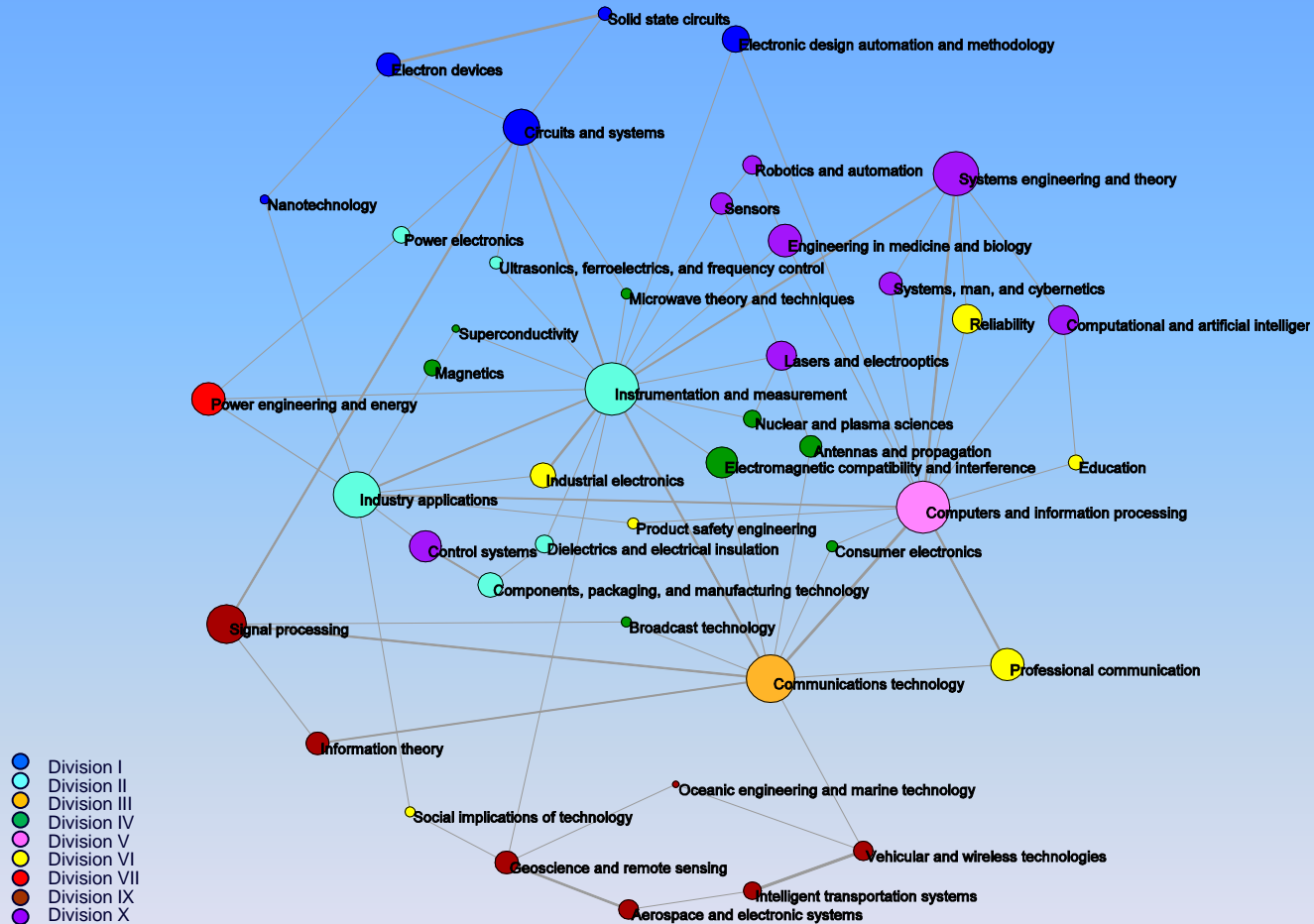
# Defining expanded term space

## 5. Visualization

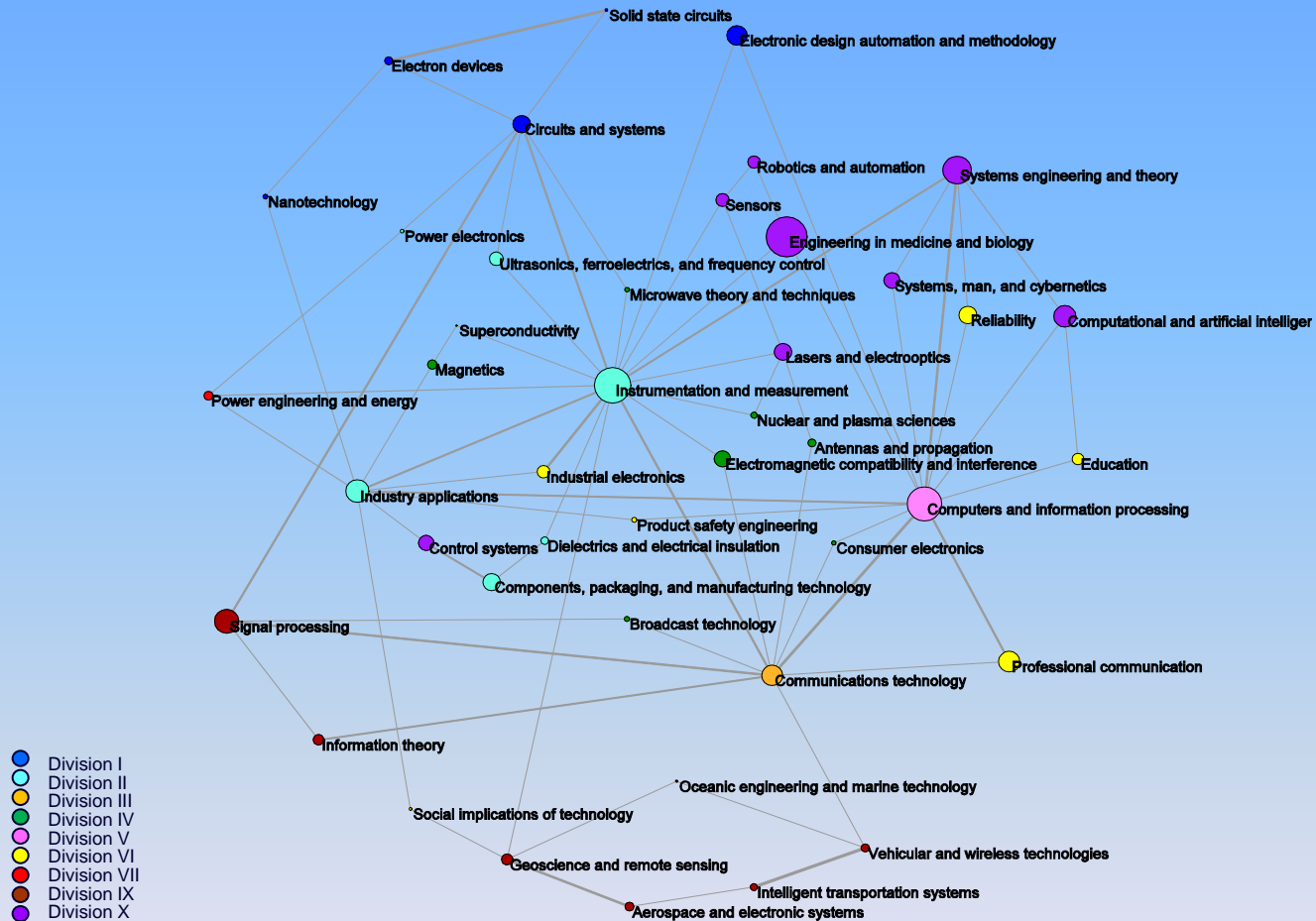




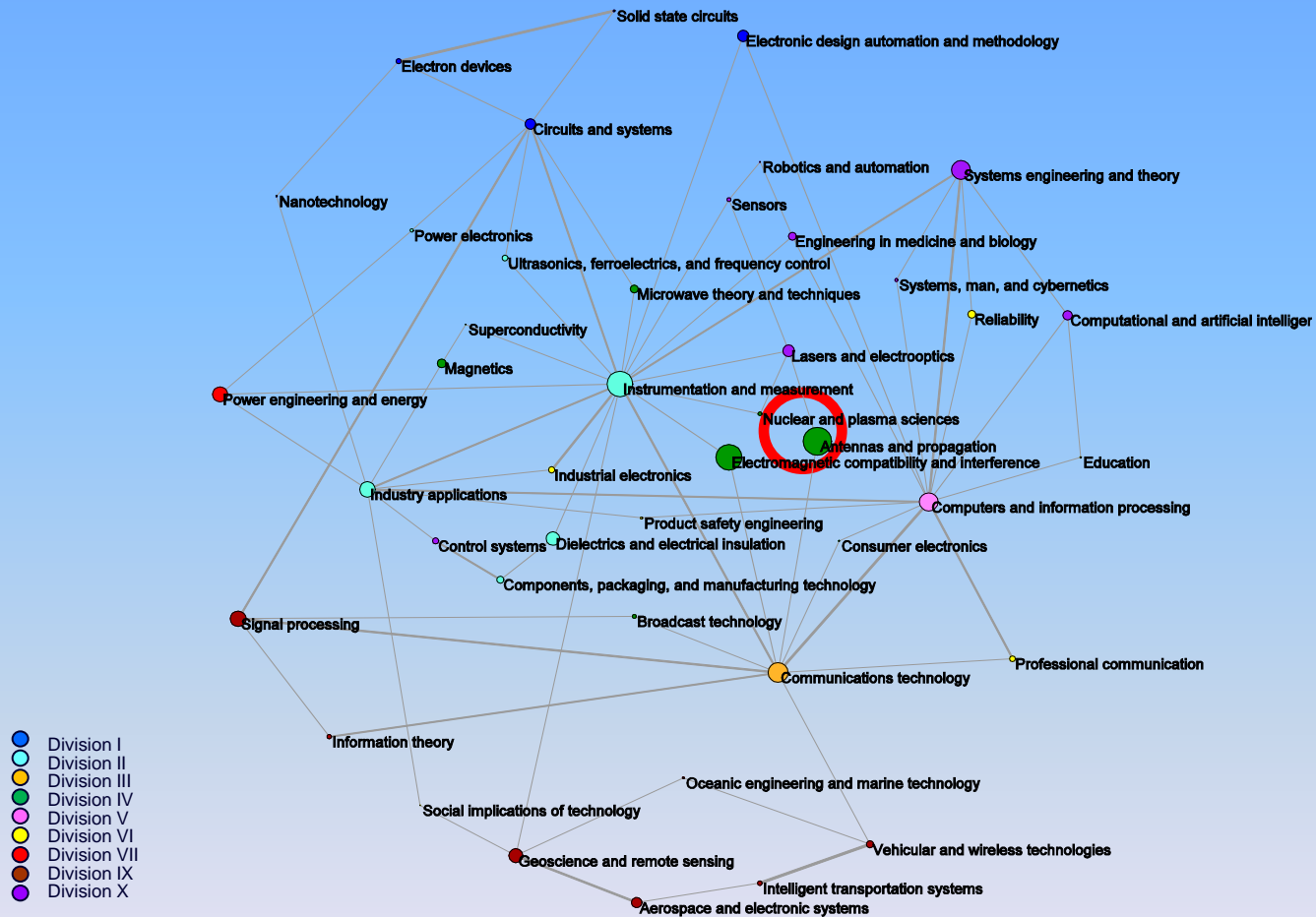
# Sample term and data maps



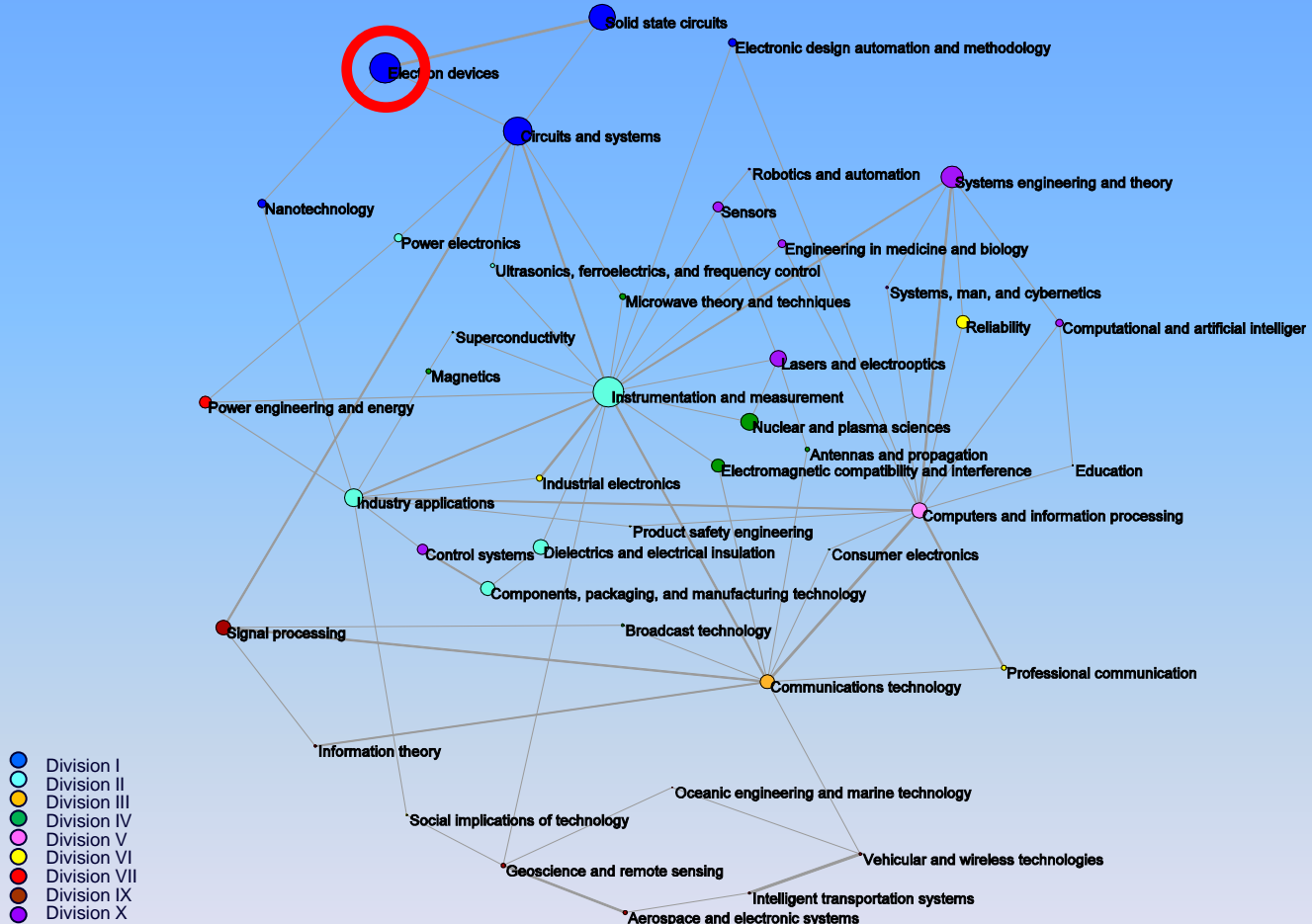
# PubMed



# IEEE Transactions on Antennas and Propagation



# IEEE Transactions on Electron Devices



# Findings

- ❖ Term space can be mapped effectively
- ❖ The mapped space can be used to show distributions and trends that give answers to questions
  - ❖ Database distribution comparisons
  - ❖ Journal / segment distribution comparisons (overlaps)
  - ❖ Journal / segment trending
  - ❖ Identify groups of terms that need trimming (rule base changes)

# Answering the IEEE questions

- ❖ Yes, we can use IEEE information, to forecast future directions
- ❖ Yes, look at each industry by technology sector over time to see where it is headed.
- ❖ IEEE coverage does not match the IEEE mission and vision by industry sector
- ❖ Provides new ways for IEEE to become smarter about their data and potential markets using their collection
- ❖ The societies are not all publishing and talking about what their charter indicates they cover.

# Looking to the future

- ❖ We can see specific trends and which topics are emerging/cooling
- ❖ Using the IEEE data and these term analytics technology we can explore the future and the boundaries of the IEEE future



# Thank You

Marjorie M.K. Hlava, President,  
Access Innovations / Data  
Harmony

[mhlava@accessinn.com](mailto:mhlava@accessinn.com)

Access Innovations  
4725 Indian School NE Suite 100  
Albuquerque, NM 87110

[www.accessinn.com](http://www.accessinn.com)

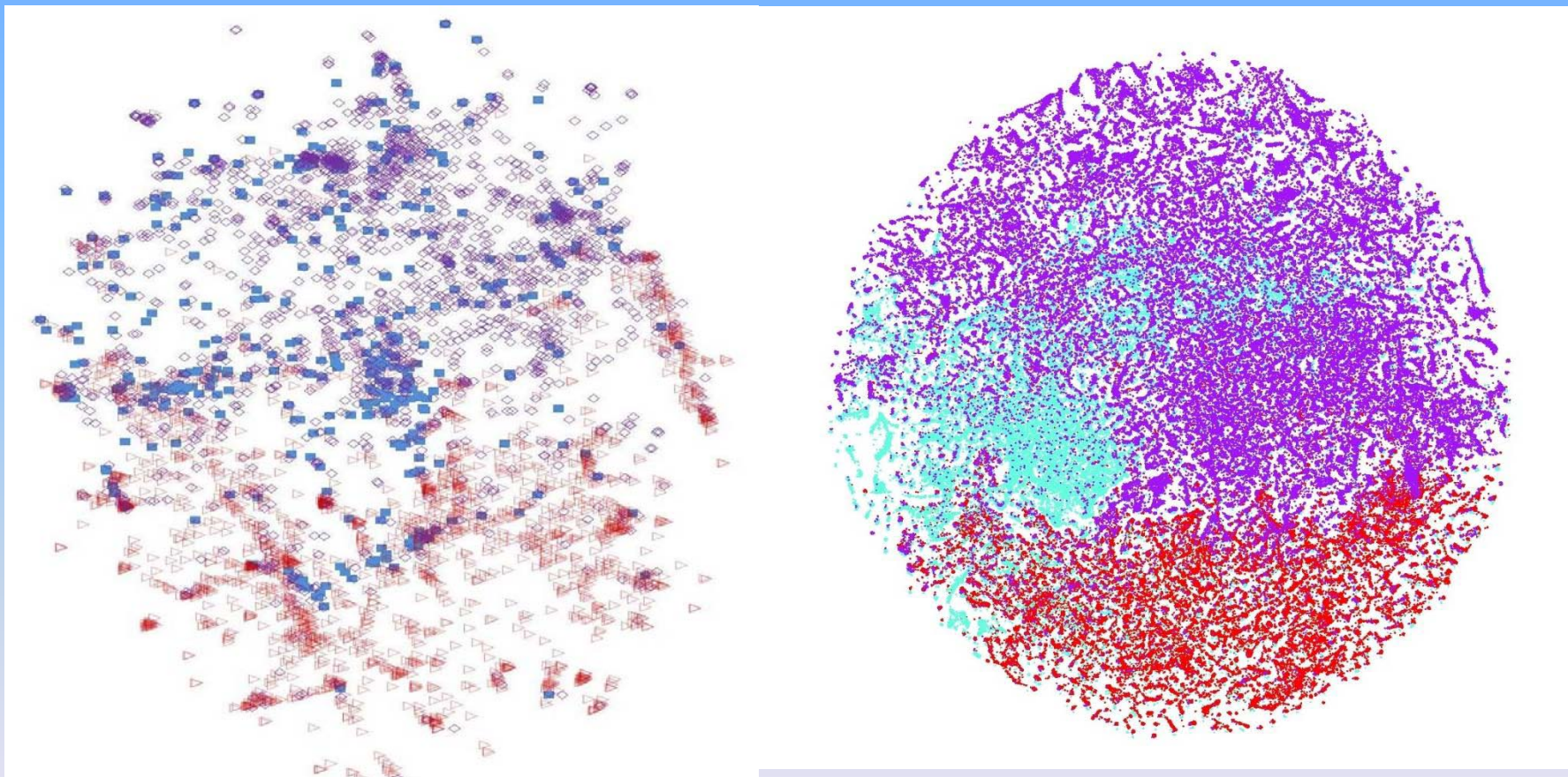
(505) 998-0800 office

(505) 256-1080 fax

[www.taxodiary.com](http://www.taxodiary.com) the taxonomy news blog



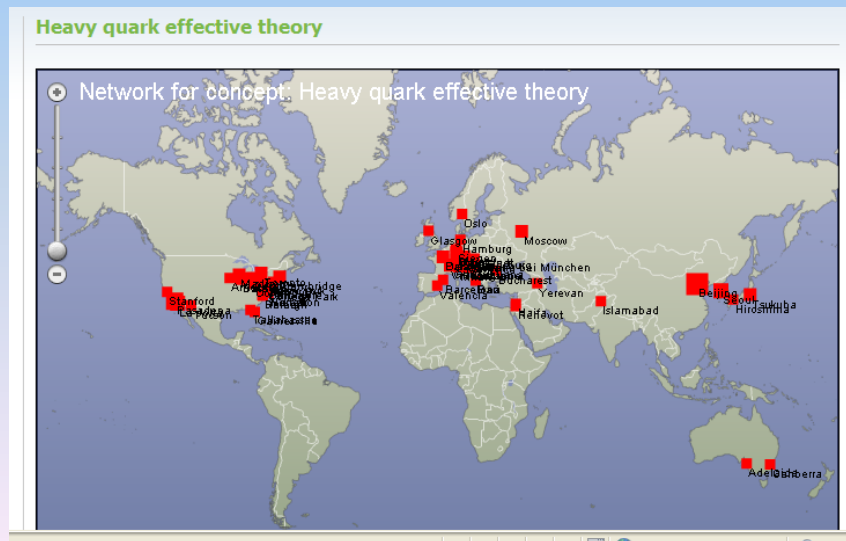
# Consensus: Thesaurus vs. Titles



# Scientific Social Networking Based on Metadata

- ❖ Idea has been here - Who is citing who like
  - ❖ ISI does it with references
  - ❖ API UniPHY does it using semantics
- ❖ Expand your options using
  - ❖ good metadata and descriptors

**Map who is working in the  
field and where**



**See the authors connections**

