Turn Information into Knowledge

www.TheVantagePoint.com

Search Technology
770.441.1457
VantagePoint@searchtech.com

Patent Overlay Mapping: Visualizing Technological Distance

Based on collaborations between
Nils C. Newman[1]
Ismael Rafols[2]
Alan L. Porter[3]
Jan Youtie[4]
Luciano Kay[5]

[1] Search Technology
[2] SPRU, University of Sussex
[3] Search Technology
[4] EII, Georgia Institute of Technology
[5] STIP, Georgia Institute of Technology

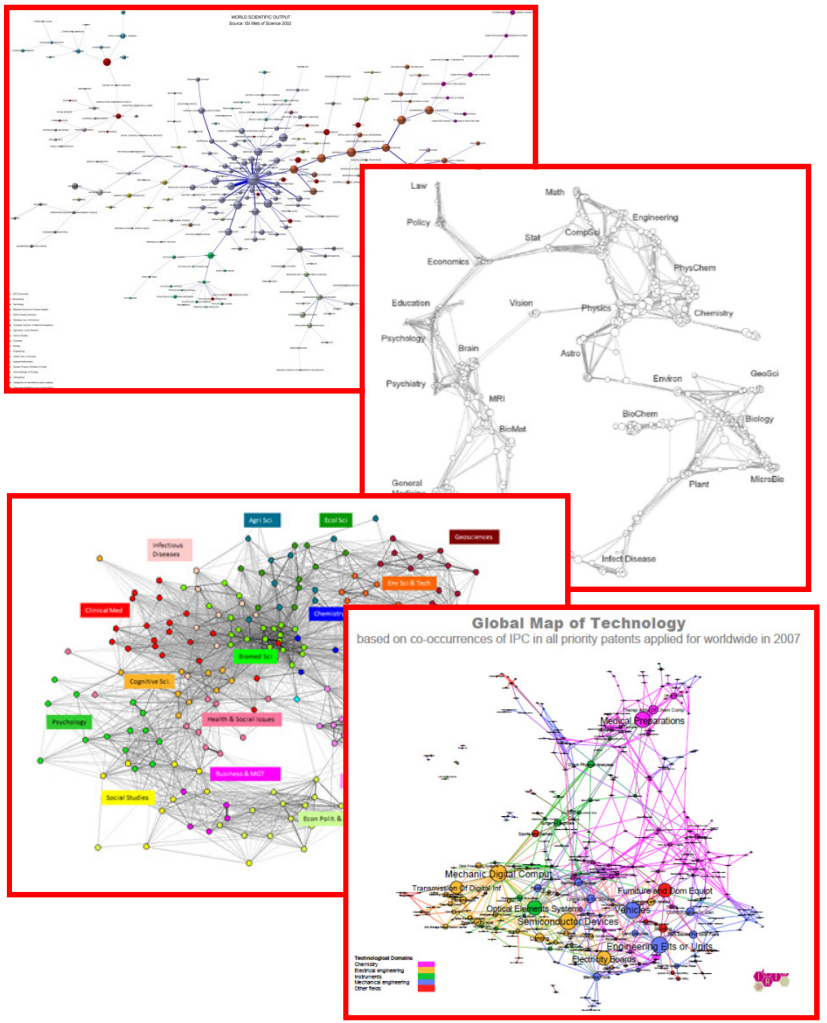
Context: Proliferation of global maps of science

Until 2000s, very few:

- Small et al. (1985), Small (1999), Bassecoulard and Zitt (1999)

In recent years many:

- Moya-Anegon et al. (2004, 2007), Klavans and Boyack (2005, 2009), Antoine Schoen et al (2011), Leydesdorff and Rafols (2009),Balaban and Klein (2006), Rosvall and Bergstrom (2009), Janssens et al. (2009), Bollen et al. (2009), etc.



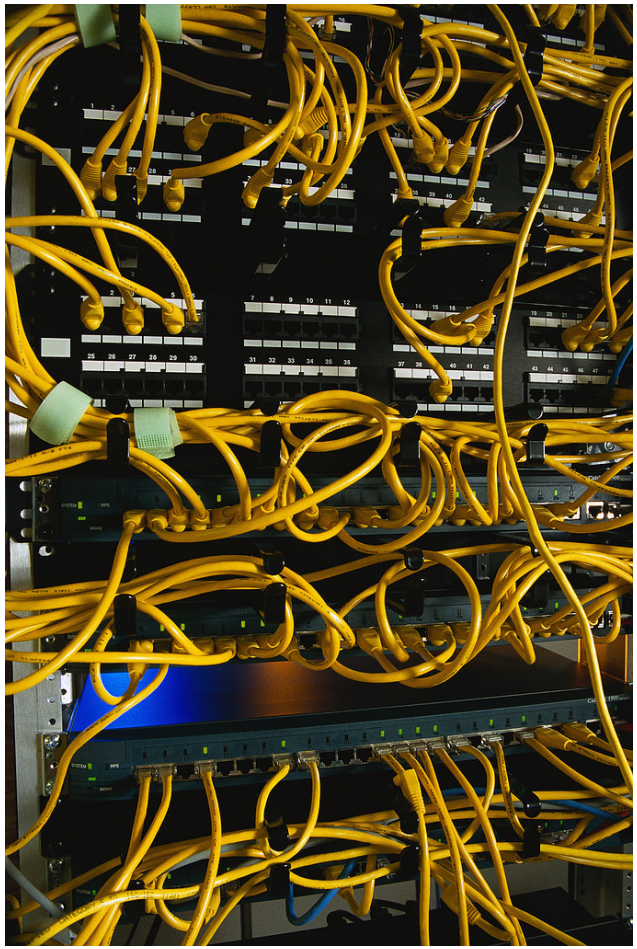
Motivation: The shifting structure of science

From ‘The **tree** of knowledge’

- Hierarchical
- Branching into subdisciplines and specialties
- Matching the social structure of university departments

... to a **web** of ‘ways of knowing’

- Changing social contract (Gibbons et al. 1994)
- Increasing institutional hybridity (Etkowitz and Leydesdorff, 2000)
- Dissonance between epistemic and social structures



Approaches to mapping the sciences

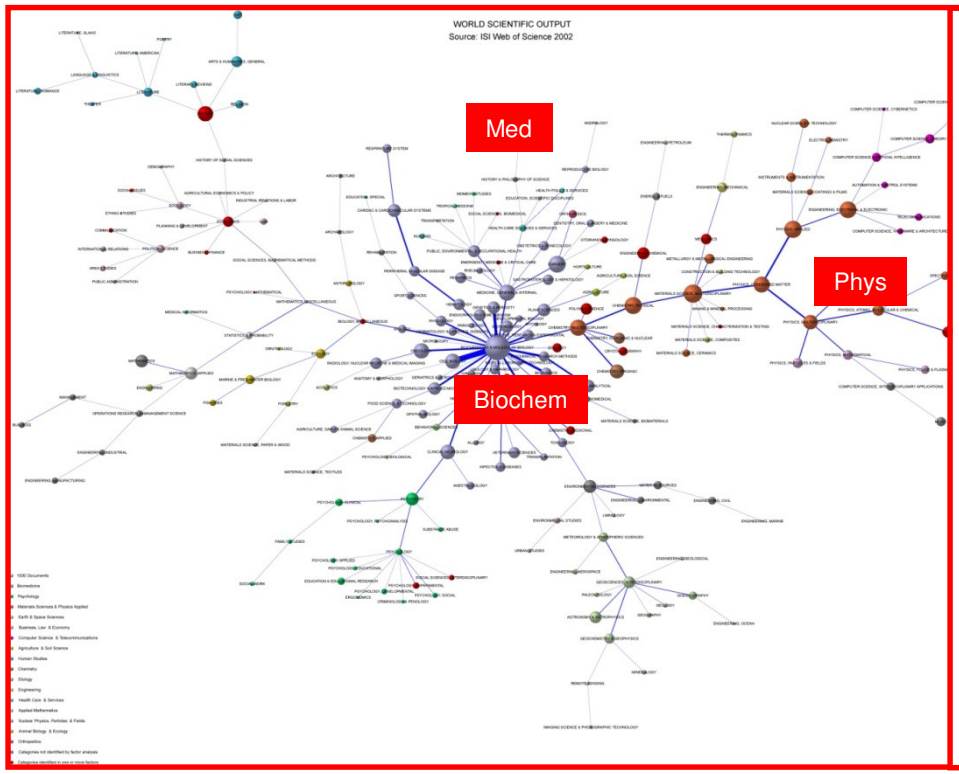
Local science maps portraying internal dynamics within a field

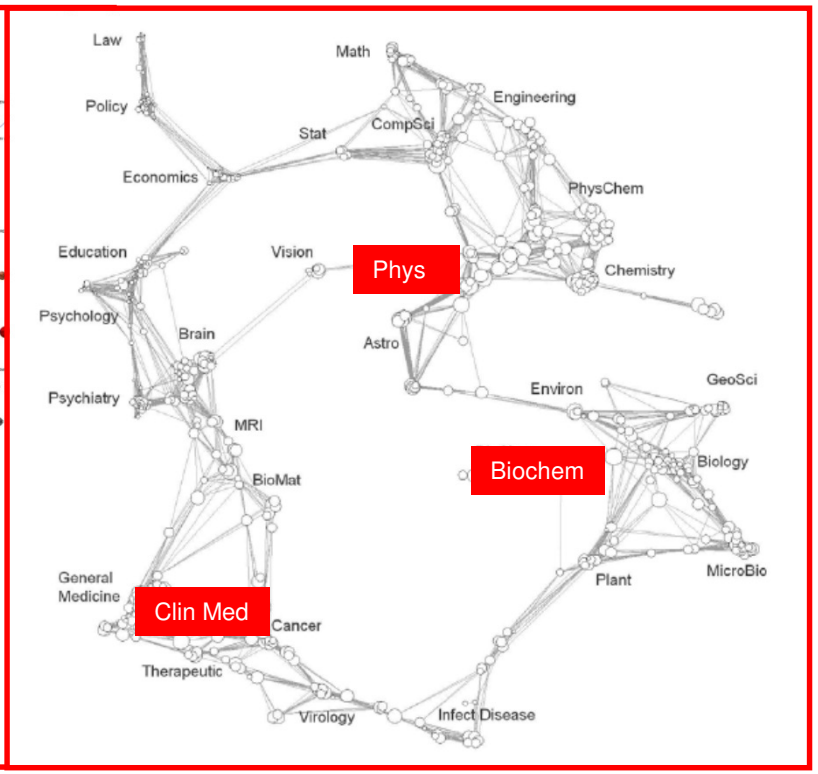
- Developed since the 70s (Small 1973)
- Widely used since late 80s
  - Co-citation of articles (Small, 1997)
  - Co-word analysis (Callon et al. 1986)
  - Co-classification of articles (Noyons and van Ran, 1998)
  - Co-citation of journals (van den Besselaar and Leyesdorff, 1996)
- Improvements in visualization tools in the ’00s

Global Science Maps:

- Small et al. (1985), Garfield (1987)
- But methods seen as unstable and problematic (Leydesdorff, 1987)
- Many choices:
  - Unit of analysis
  - Similarity measures
  - Dimension reduction
  - Visualisation technique and layout
- Expectations: Highly dependent on the specific choices

Surprise: Robustness in core structure





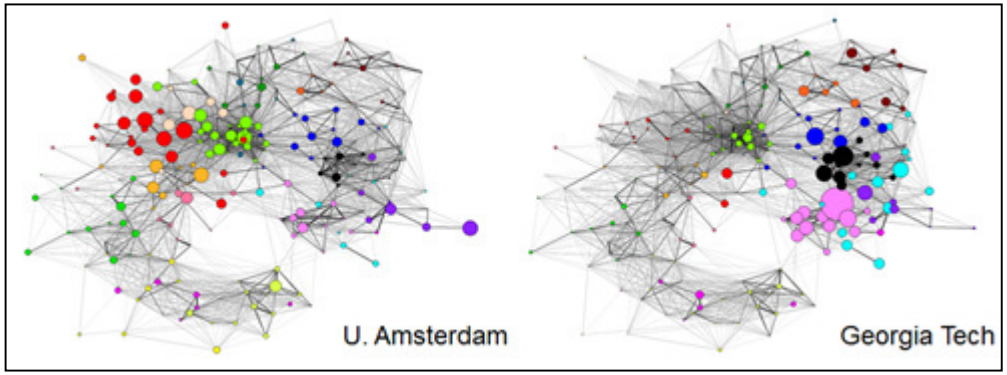
Moya-Anegon et al. 2007

Boyack, Borner and Klavans, 2009

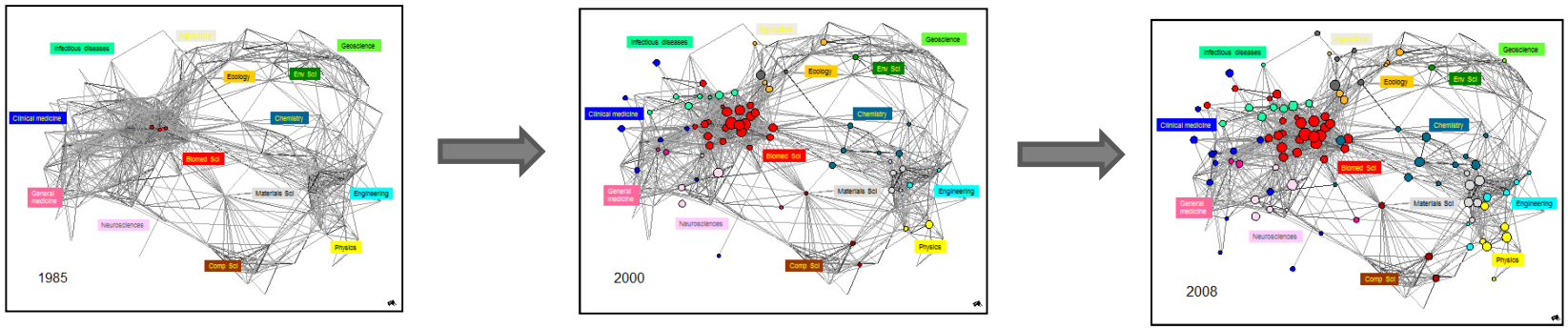
Layout can be appear very different, but strongest linkages still similar
Klavans and Boyack (2009)

And the science maps are useful

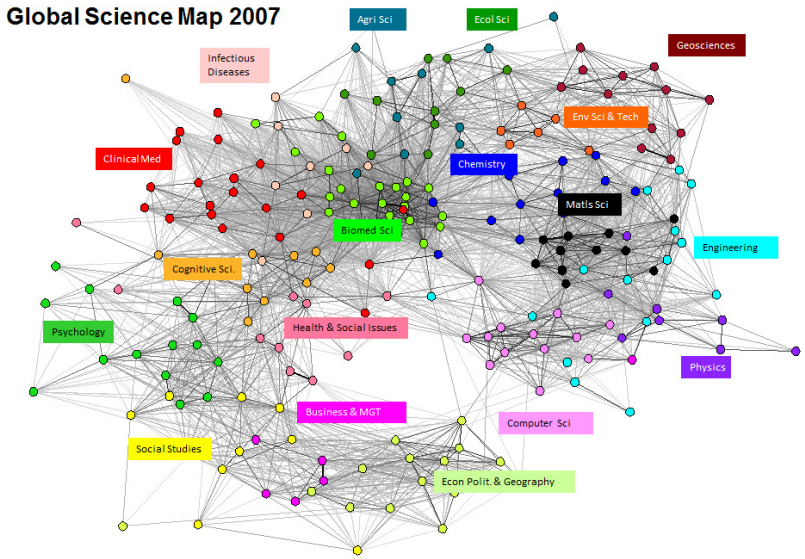
1. Benchmarking - How similar/different are organisations A & B?

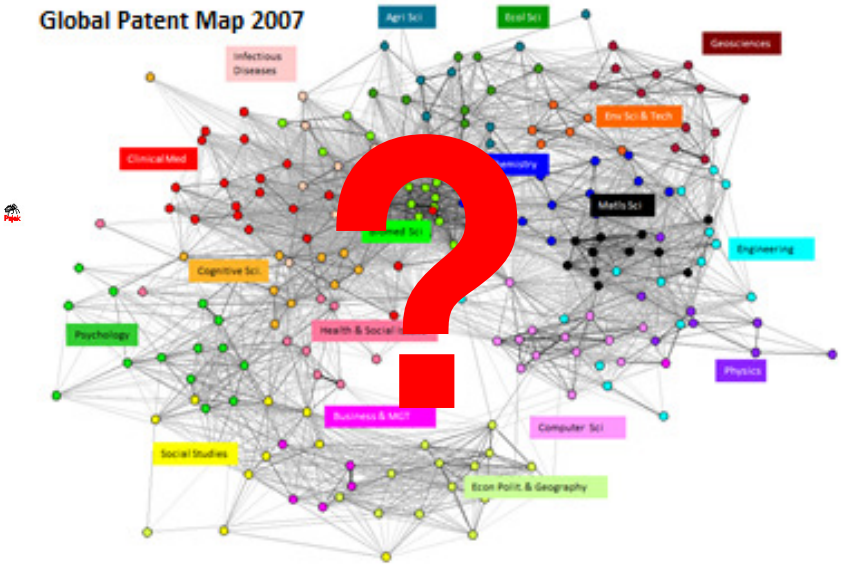


2. Capturing temporal change - How do the patterns evolve over time?



If we can make a map of Science, can we make a map of Patents?





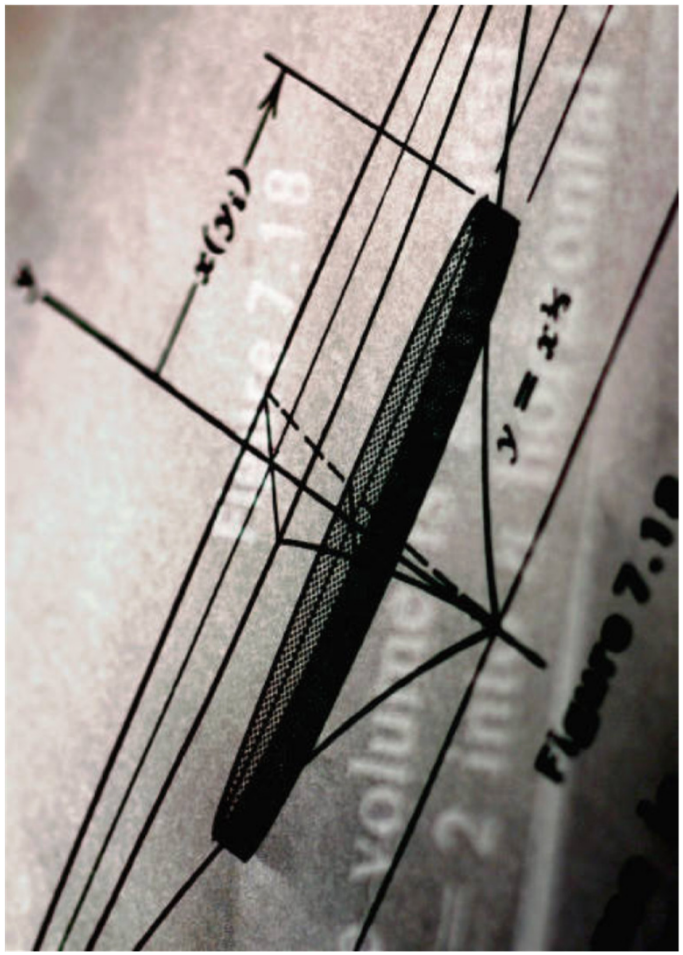
The problems…..

What to use as an equivalent to citation patterns?

- *Are patent citations functionally equivalent to journal citations?*

What to use as an equivalent to ISI subject categories?

- *Would IPC's work as suitable replacement?*



The IPC problem…..

If we use IPC's, what level should be used?

- Section, Class, Sub Class or Group?

The Sub Class level makes sense but…..

- The Sub Class level has the population problem

A61K =

A42B =

A potential solution

Fold the IPC category up to create relatively similar sized grouping

- For large population IPC's, use the smallest Sub Group level
- For small population IPC's, aggregate up to General Group level, Sub Class or Class
- Establish a floor cut-off and drop very small aggregated populations

85709	A61K
1982	A61K 8/00
724	A61K 8/02
1082	A61K 8/04

85709	A61K
2706	A61K 8/00
1082	A61K 8/04

The Experiment

To test this approach:

- Pulled the EP Authority population from PatStat for records issued from 2000-2006 (~760,000 records)
- Collected the instances of IPC for all the records
- Each IPC with an instance count greater than 1,000 was kept in its original state
- Each IPC with an instance count less than 1,000 was folded up to the next highest level until the count exceeded 1,000 or the Class level was reached
  - During the folding, any other IPC's with counts exceeding 1,000 in the same branch were left out of the folding count
- If at the Class level, the population was less than 1,000, the IPC code was dropped for being too small to map

The Experiment continues…

- The cut-off at 1,000 produced 466 IPC entities
  - Several other cut-offs were tested
- The 466 IPC entities were then mapped to thesaurus
- Returning to PatStat the patents cited by the initial set were pulled from PatStat and their IPC's (if available) were mapped the 466 IPC entities
  - No cut-off on Cited IPC's
- This data collection allowed the creation of table containing
  - Patent Number
  - IPC Number
  - Cited Patent Number
  - Cited IPC Number

The Experiment continues…

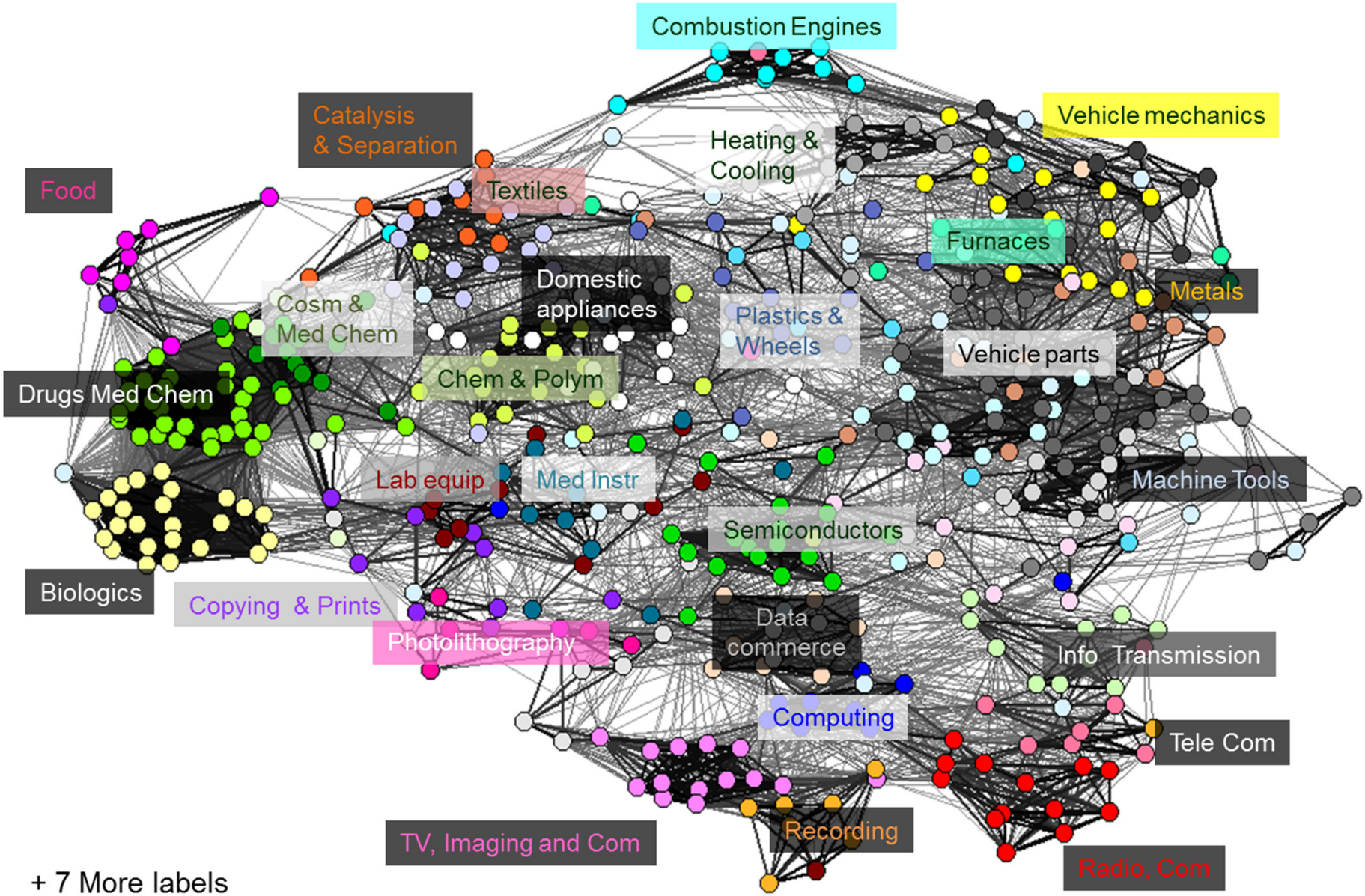
- The data was converted into Pajek format and R Stats was used to create a CosSim file in Cited and Citing directions

- A factor analysis of the Cosine similarity matrix of cited to citing patents was used to reduce the 466 categories into 35 “macro patent groupings”
  - Factors from 10 to 40 were tested
  - Using 35 factors appeared to have the greatest face validity
  - The 466 categories were color coded based on the factors

- The IPC codes were converted to text labels using a manual process based on the IPC definitions

- The text labels for the categories were used to create a text label for each factor

The result looks like…

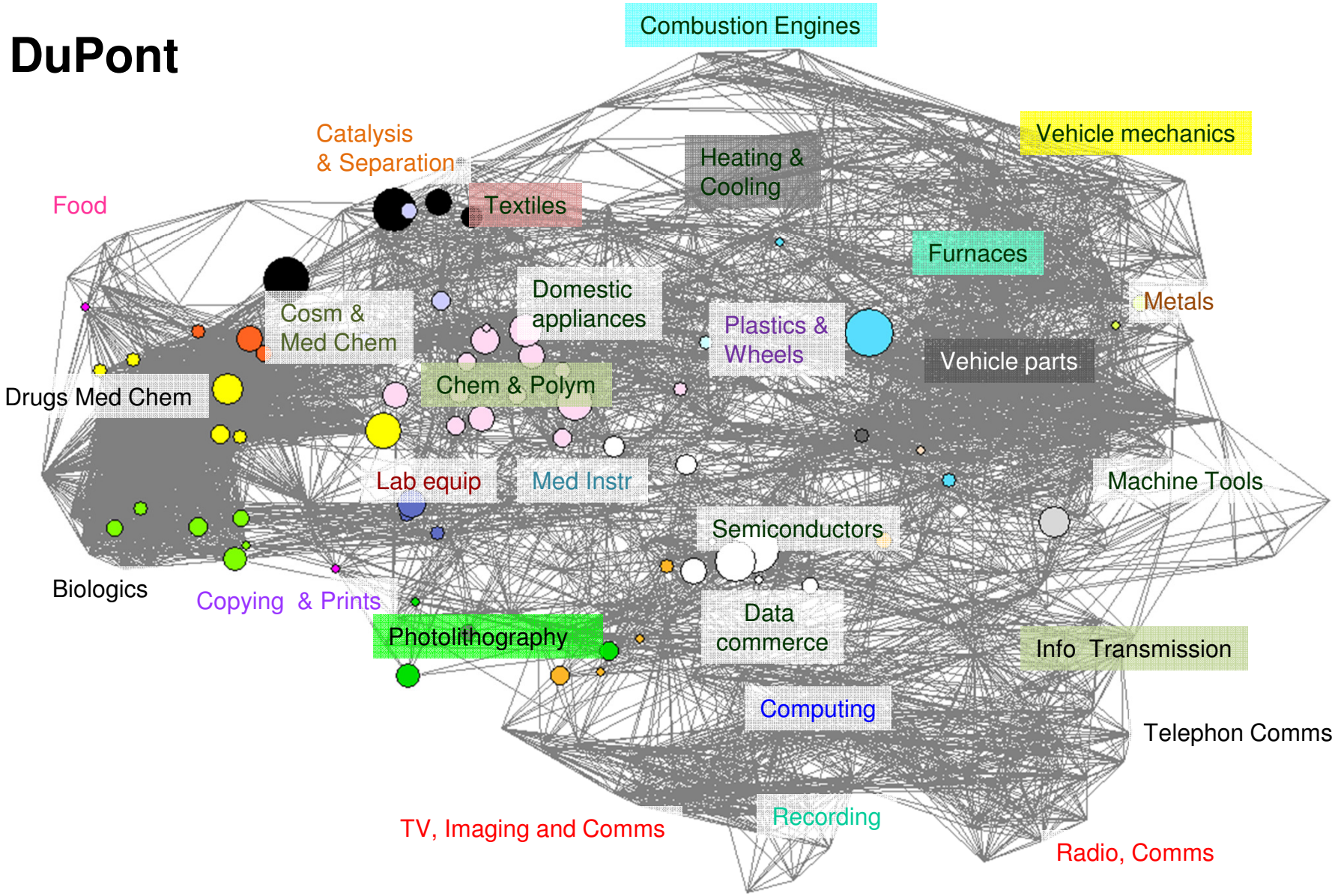


But is it useful?

- The following slides contain patent overlays for specific companies and specific technologies based on 466 categories
- Data: Two sets of overlays are based on existing datasets of nanotechnology patent applications (Coverage period: 2000-2006)
  - Companies
    - **Samsung**
    - **IBM**
    - **Philips**
    - **DuPont**
  - Technologies
    - **Nano-Biosensors**
    - **Dye-sensitized solar cells**
    - **Graphene**
- Graphs:
  - Same scale was sought for all overlays to be able to compare
  - Color of labels not always coincide, a few were modified to match node colors

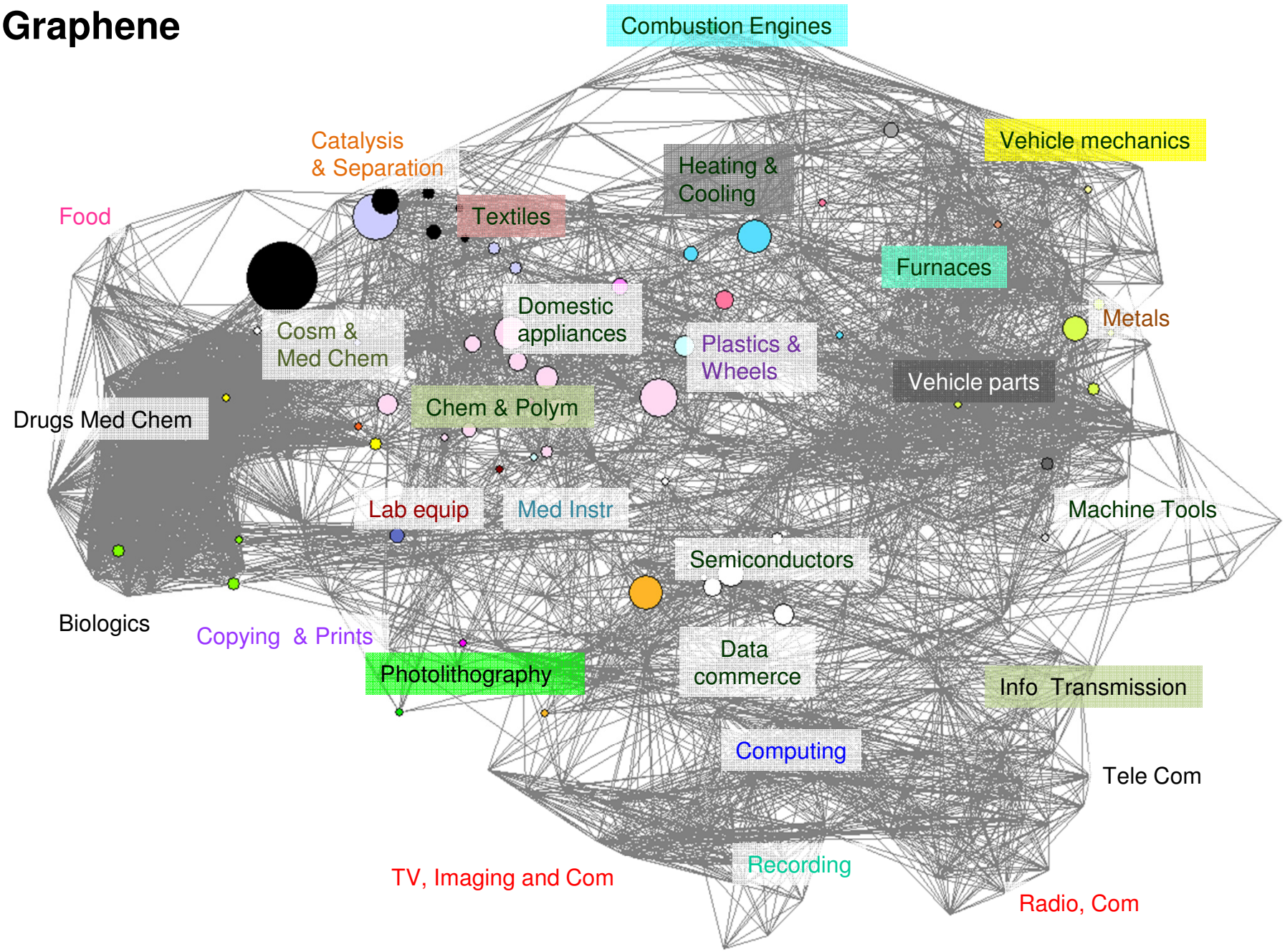
Company Overlay

DuPont


Combustion Engines
Catalysis
& Separation
Vehicle mechanics
Heating &
Cooling
Food
Textiles
Furnaces
Domestic
appliances
Metals
Cosm &
Med Chem
Plastics &
Wheels
Vehicle parts
Drugs Med Chem
Chem & Polym
Lab equip
Med Instr
Machine Tools
Semiconductors
Biologics
Copying & Prints
Data
commerce
Photolithography
Info Transmission
Computing
Telephon Comms
Recording
TV, Imaging and Comms
Radio, Comms

Technology Overlay

Graphene


Combustion Engines
Catalysis
& Separation
Vehicle mechanics
Heating &
Cooling
Food
Textiles
Furnaces
Domestic
appliances
Metals
Cosm &
Med Chem
Plastics &
Wheels
Vehicle parts
Drugs Med Chem
Chem & Polym
Lab equip
Med Instr
Machine Tools
Semiconductors
Biologics
Copying & Prints
Data
commerce
Photolithography
Info Transmission
Computing
Tele Com
Recording
TV, Imaging and Com
Radio, Com

Conclusions

- Potential new tool for research management
  - A somewhat intuitive and straightforward method
  - Based on conventional categories and the core structure of patents (*hopefully*)
- Proposed conditions of application:
  - Large set (>2000 pubs; low resolution >100)
  - Traceability, transparency, parsimony
  - Best contrasted with local maps

➢ ***Possible applications:***

- ❖ ***Benchmarking***
- ❖ ***Temporal Analysis***
- ❖ ***Competitive Technical Intelligence***

Next Steps

- Switch to IPC8
- Work with Grants only
- Expand beyond EP?
- Work on the labels
- Segment maps by Year or Year Ranges to look for stability or instability
- Compare results with other global patent mapping efforts
- Experiment with potential connections between Patent Map and Science Maps in technologies with strong science links
- Move beyond IPC's?
- Your suggestions….

Search Technology

Nils Newman
newman@searchtech.com

www.theVantagePoint.com