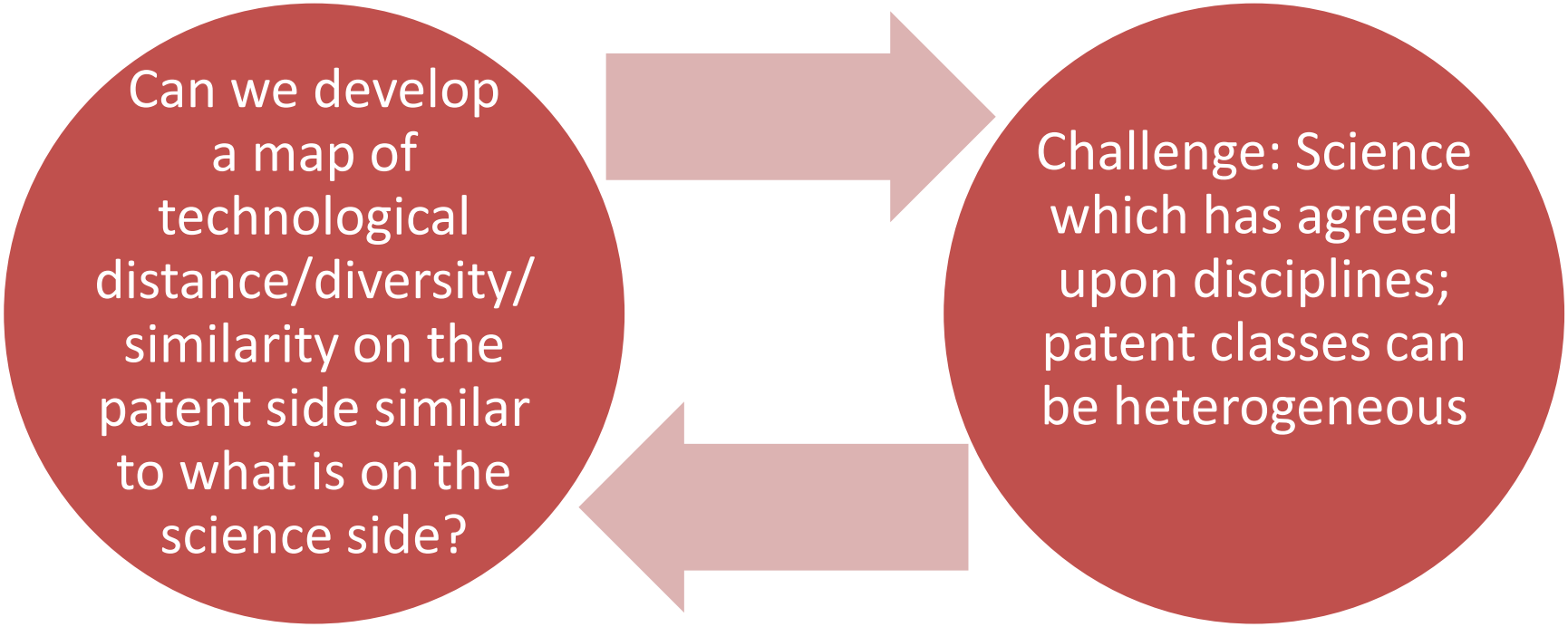
Hierarchy versus Similarity: Visualizing Technological Distance Using Patent Overlay Mapping

Jan Youtie, Georgia Tech, USA

from L. Kay, N. Newman, J. Youtie, A.L. Porter, I. Rafols, 2013. Patent Overlay Mapping: Visualizing Technological Distance, *Journal of the American Society for Information Science and Technology*

Can we make a map of patents?



What is Technological Distance/Diversity?

A particular patent class represents a technological position	• Common knowledge base, heuristic, scientific principle • Similar knowledge flows: patents in a class more likely to cite other patents in that same class
Distribution of patents across classifications in a patent hierarchy proxies technological diversity/distance	• Firms rely on distributed technological capabilities rather than just a core set
Technological diversity concept examples	• Indicator of more radical innovations • "Structural holes" • Specialization v. diversity in regional capabilities • Similarity of 2 patent portfolios – inverted "U"

But there are measurement issues → Our contribution focuses on 2 issues

Jaffe 1989; Breschi, Lissoni, Malerba 2003; Hinze et al. 1997; Granstand, Patel, Pavitt, 1997; Olsson 2004; Hollingsworth, 2000; Nooteboom et al. 2007

Issue 1. Is there similarity within hierarchy?

- Sections

A HUMAN NECESSITIES
B PERFORMING OPERATIONS; TRANSPORTING
C CHEMISTRY; METALLURGY
D TEXTILES; PAPER
E FIXED CONSTRUCTIONS
F MECHANICAL ENGINEERING; LIGHTING; HEATING; WEAPONS; BLASTING
G PHYSICS
H ELECTRICITY

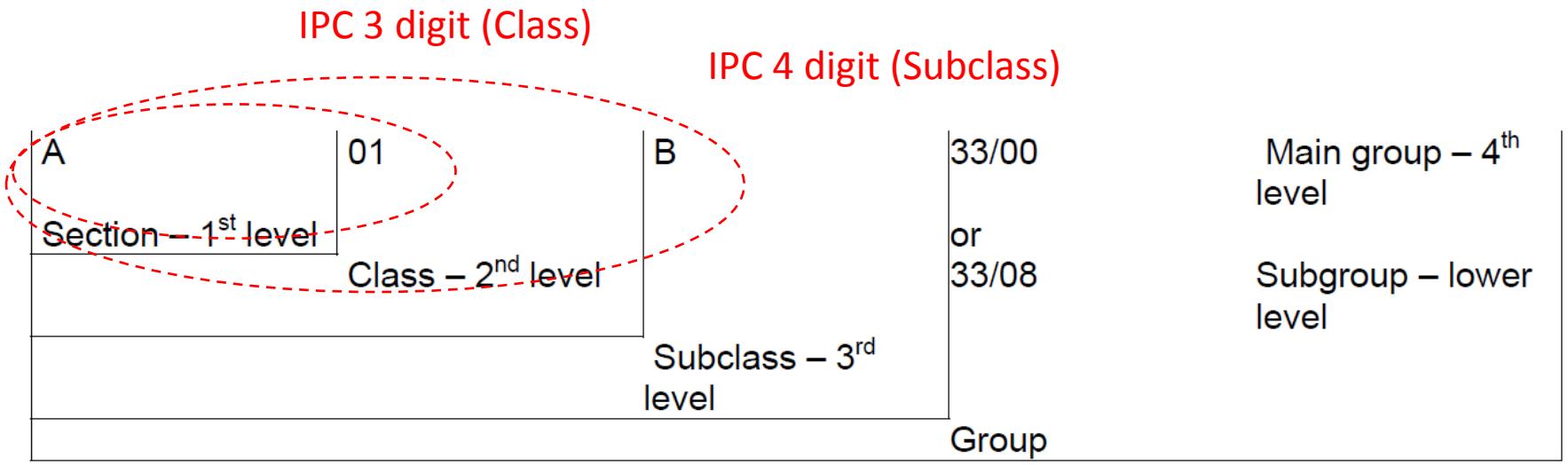
- Classifications in the same broad category are more similar than those in different broad categories
  - Counts
  - Shares
  - Herfindahl

Source: http://www.wipo.int/classifications/ipc/en/

Issue #1 Illustrated

- IPC classes in different sections maybe similar based on co-occurrence or citation similarity
  - A61K (incl. drugs) ≈ C07 (organic chemistry)
  - A61K ≉ A42B (hats)

Issue 2. There are patent maps*, but most are at a given hierarchical level



* e.g., Hinze et al. 1997, Leydesdorff et al. in press.

Source: Guide to the IPC (2013), p. 6.

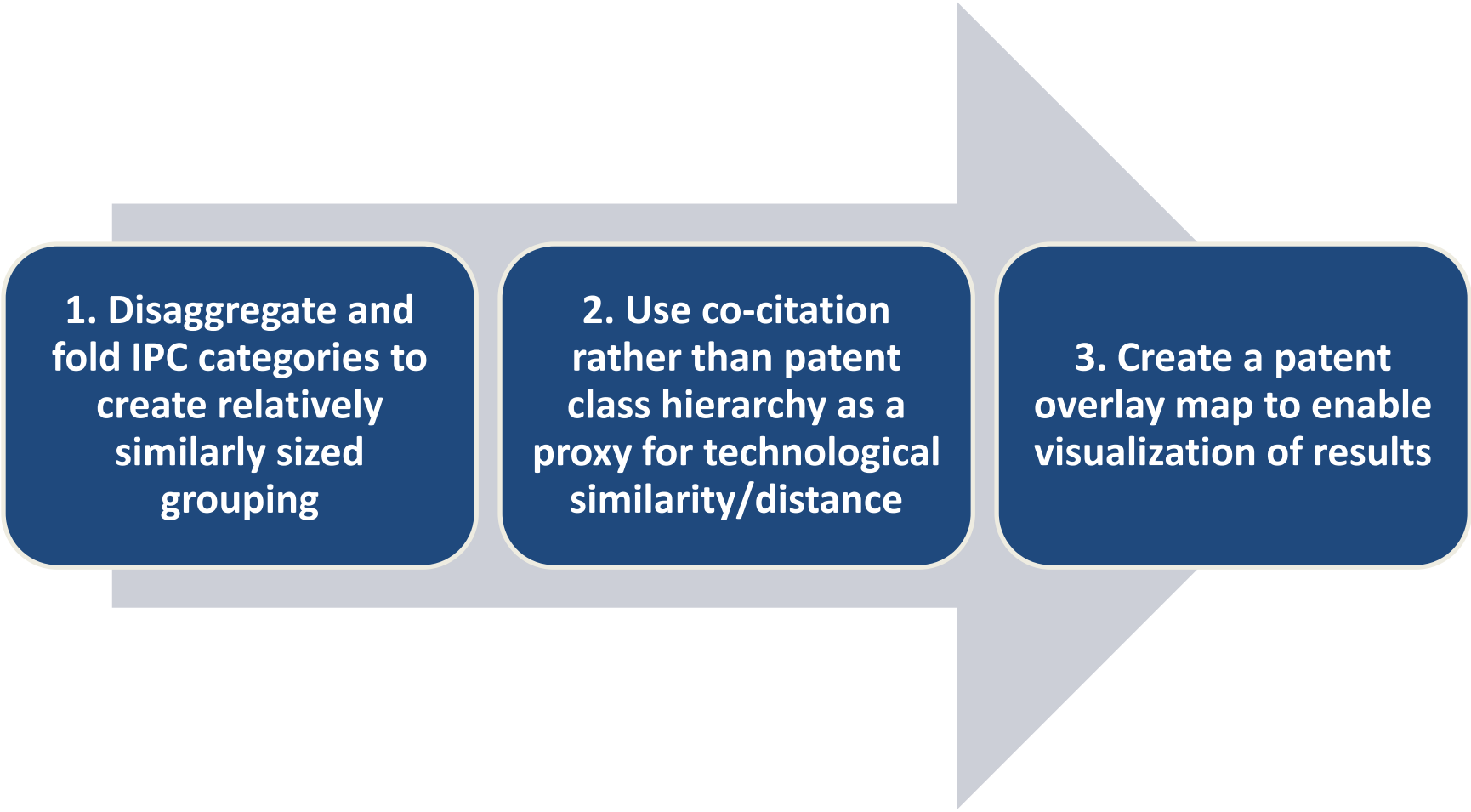
Issue #2 Illustrated

- Different IPC classes have different propensity to patent

A61K (Preparations for Medical, Dental, or Toilet Purposes, e.g., Drugs) =

A42B (Hats) =

Our Approach



Step 1: Fold the IPC categories

- Pulled the EP Authority population of 760,000+ records (grants, drops, etc.), PatStat issued 2000-2006 (IPC 7)
- 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
- If at the Class level, the population was less than 1,000, the IPC code was dropped for being too small to map
- This resulted in 466 categories
- Converted IPC codes to text labels using a manual process based on the IPC definitions/catchwords

Data pre-processing to group IPC categories, selected examples

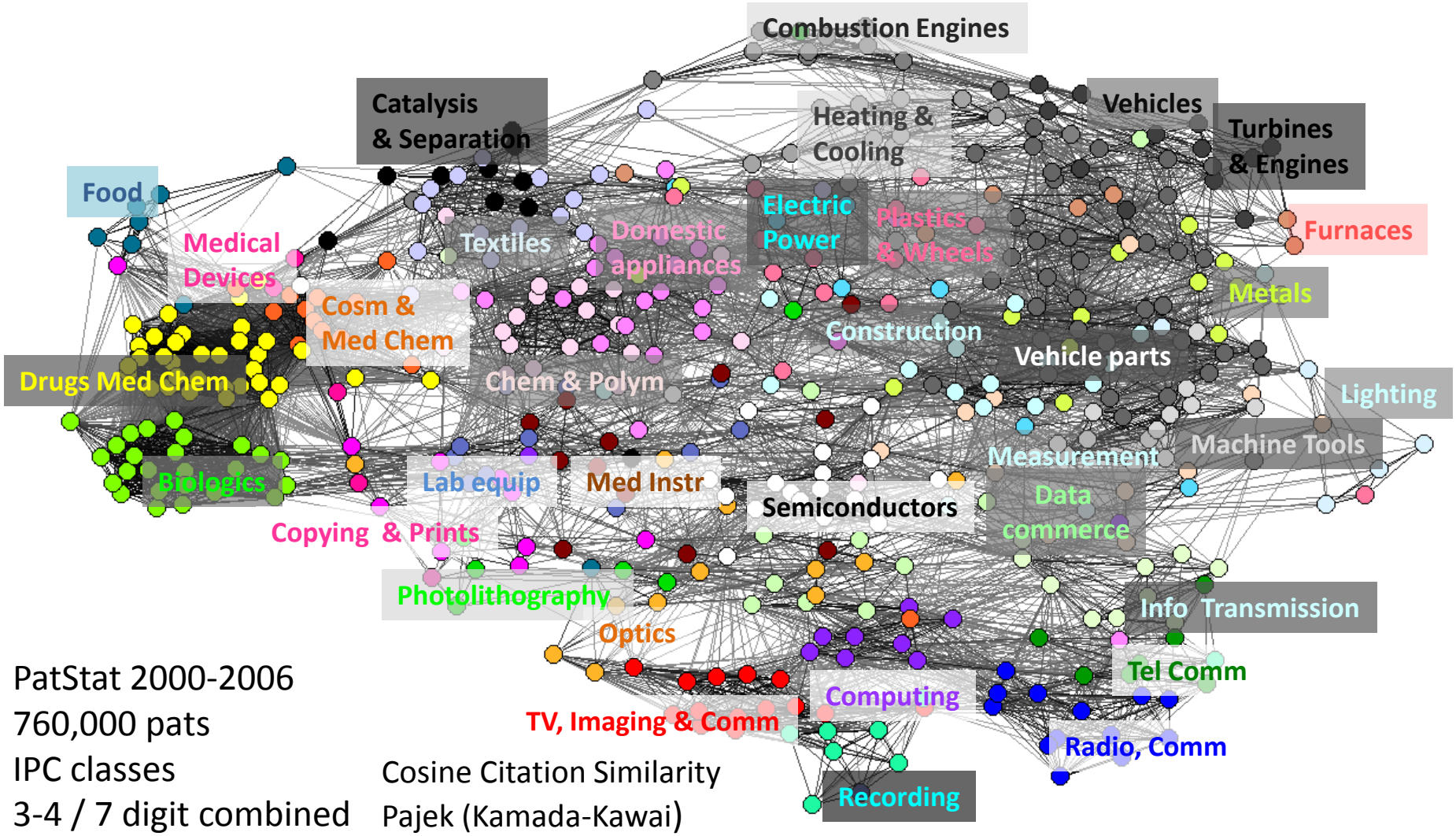
Original IPC in dataset	Catchwords	Original Record Count
A61B	Diagnosis; Surgery; Identification	25,808
		Authors' process splits this out into:
A61B 5/00	Measuring for diagnostic purposes	1,415
A61B 17/00	Surgical instruments, devices or methods, e.g. tourniquets	1,493
A61B 19/00	Instruments, implements or accessories for surgery or diagnosis not covered by any of the groups	1,444
		and a remainder:
A61B (e.g., A61B 7/00 stethoscopes; A61B 8/00 blood pressure)		21,456

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.

Step 2: Cosine similarity matrix

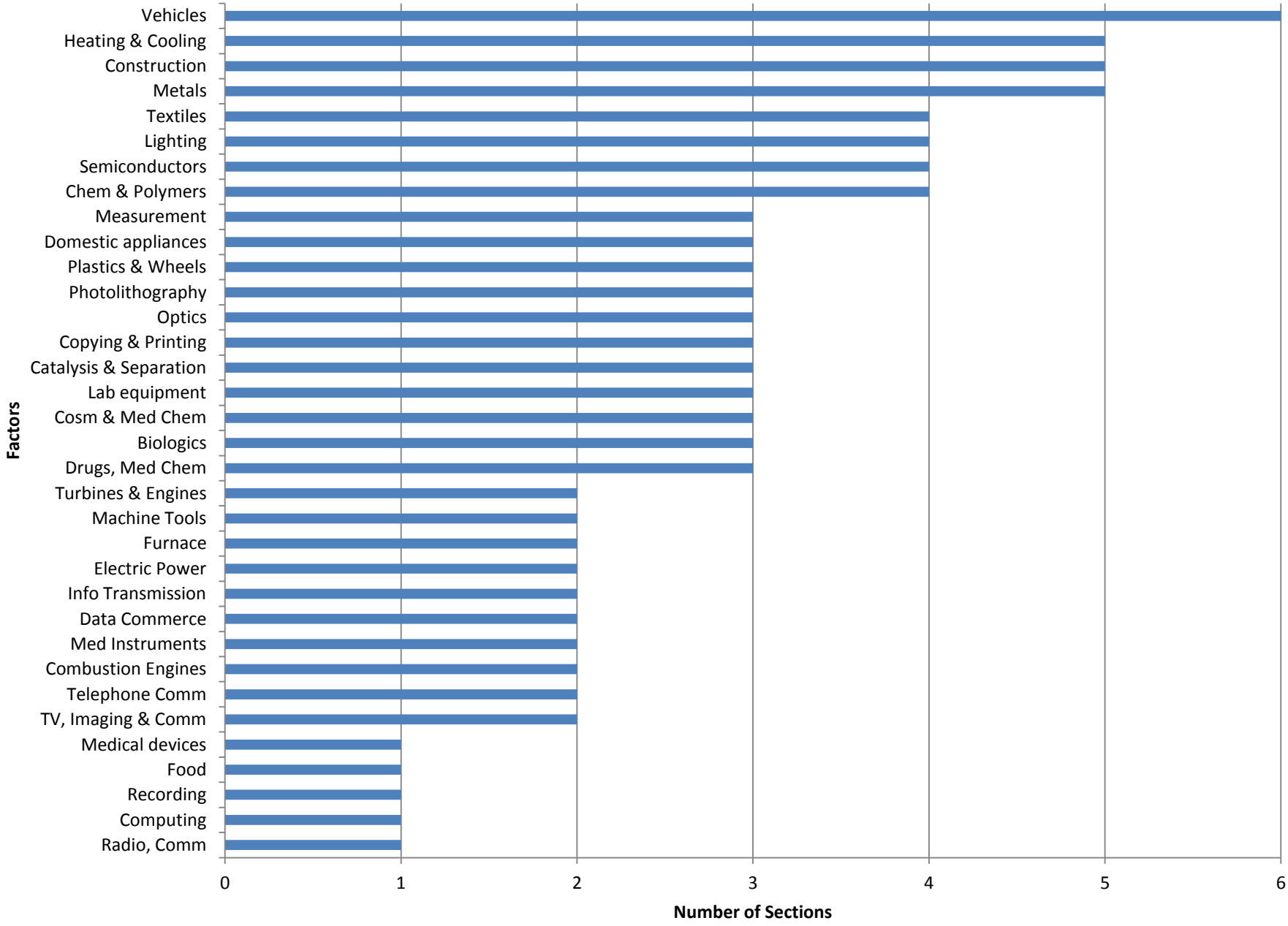
- Reduced the 466 categories into 35 “macro patent groupings” through factor analysis of the Cosine similarity matrix of cited to citing patents
  - Factors from 10 to 40 were tested
  - 35 factors appeared to have the greatest face validity
  - The 466 categories were color coded based on the factors
- Used the 466 category text labels to create a text label for each factor

Step 3: Patent Overlay Map
(35 factors color coded)

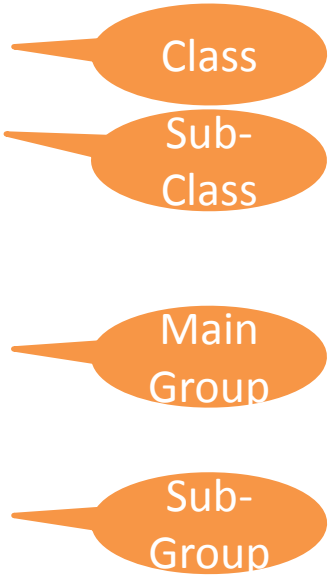


PatStat 2000-2006
760,000 pats
IPC classes
3-4 / 7 digit combined

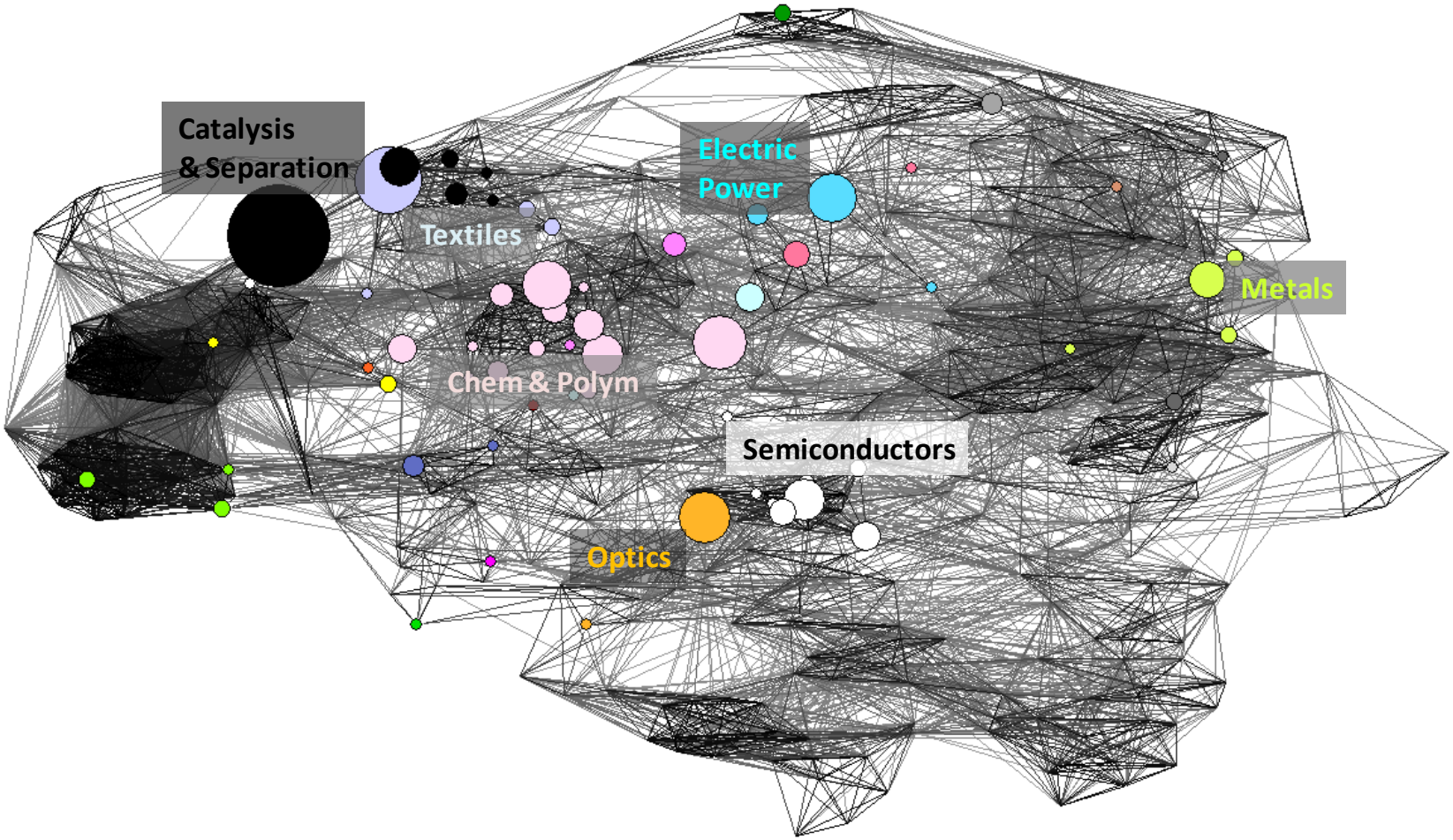
Cosine Citation Similarity
Pajek (Kamada-Kawai)


Vehicles
Heating & Cooling
Construction
Metals
Textiles
Lighting
Semiconductors
Chem & Polymers
Measurement
Domestic appliances
Plastics & Wheels
Photolithography
Optics
Copying & Printing
Catalysis & Separation
Lab equipment
Cosm & Med Chem
Biologics
Drugs, Med Chem
Turbines & Engines
Machine Tools
Furnace
Electric Power
Info Transmission
Data Commerce
Med Instruments
Combustion Engines
Telephone Comm
TV, Imaging & Comm
Medical devices
Food
Recording
Computing
Radio, Comm
Factors
0
1
2
3
4
5
6
Number of Sections

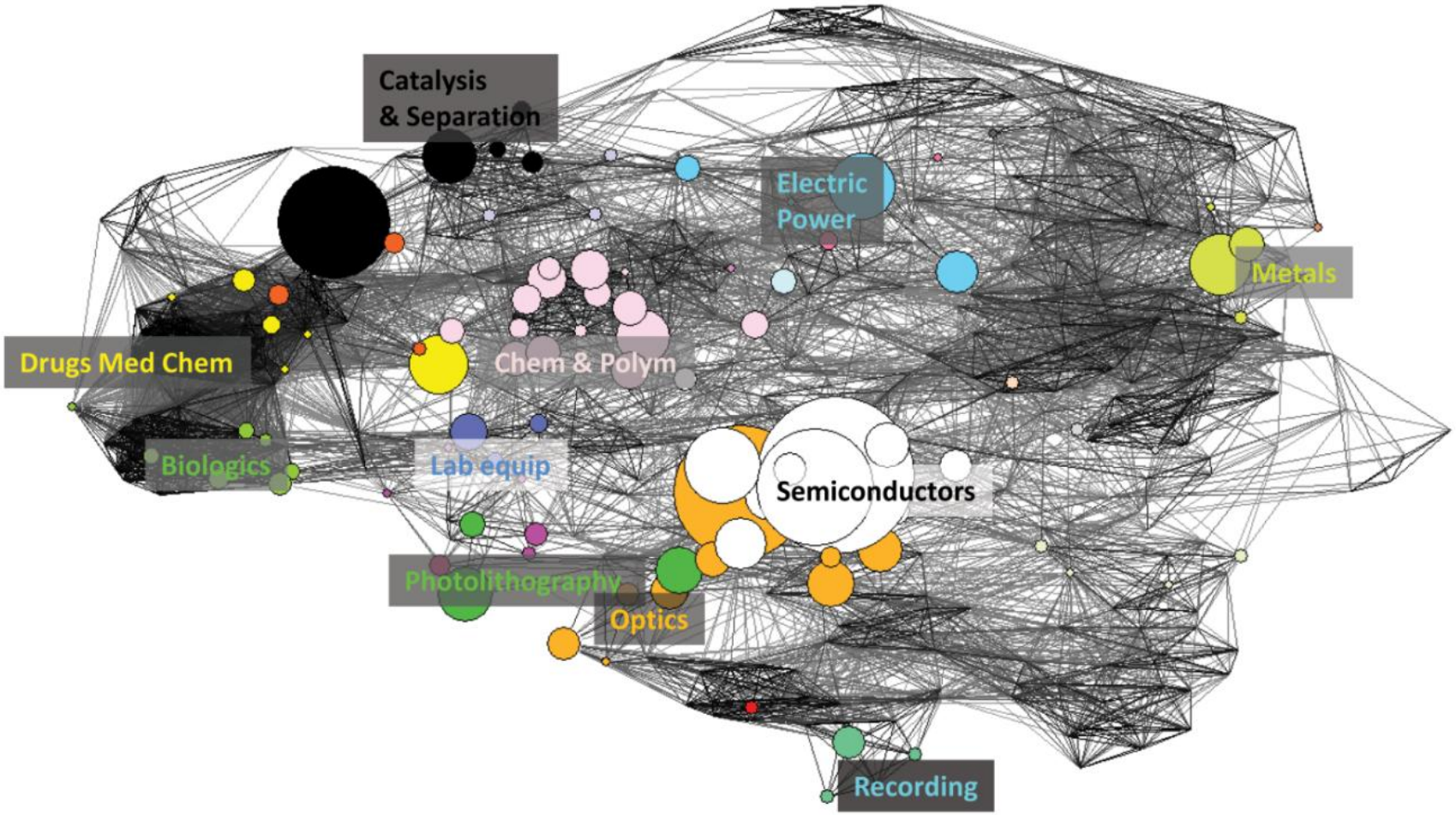
"Biologics": Category Label	IPC	Documents
Agriculture	A01	45,126
Animal husbandry	A01K	14,548
Peptides, medical	A61K 38/00	482,120
Antigens	A61K 39/00	20,010
Antibodies	A61K 39/395	47,662
Gene therapy	A61K 48/00	15,899
Saccharides	C07H 21/00	14,578
Peptides, compounds	C07K	58,219
Peptides from humans	C07K 14/435	43,462
Peptides from animals	C07K 14/47	14,602
Immunoglobulins	C07K 16/18	27,481
Extractions from organisms	C12N	26,627
Modified fungi	C12N 1/15	47,884
Modified yeasts	C12N 1/19	32,469
Cellulose processes	C12N 1/21	13,631
Virus transformed cells	C12N 5/10	10,402
Recombinant DNA	C12N 15/09	21,345
Genes encoding animal proteins	C12N 15/12	25,010
Fermentation for food	C12P	29,202
Testing, microorganisms	C12Q	18,442
Testing, nucleic acids	C12Q 1/68	22,731
Bacteriology	C12R	48,984
Measuring biological material	G01N 33/50	517,367
Immunoassay	G01N 33/53	43,835
Measuring using proteins, amino acids, lipids	G01N 33/68	119,957



Graphene Overlay Map



Samsung Overlay Map



Observations

- Similarity of structure of our patent map to others (e.g., Klavans & Boyack, 2009; Rafols & Leydesdorff, 2009)
- More fine-grained categorizations and differentiation ability

Medical and Veterinary Medicine (3 digit IPC)

(1) Drugs, Med. Chemistry,
(2) Biologics,
(3) Cosmetics and Med. Chemistry,
(4) Medical Instruments,
(5) Medical Devices

- IPC Hierarchy ≉ Co-citation Similarity
- Not as easy to use our structure (though thesaurus available at: http://www.sussex.ac.uk/Users/ir28/patmap/KaySupplementary1.xls)

Thank you

Acknowledgement: US National Science Foundation (NSF) through the Center for Nanotechnology in Society (Arizona State University; Award No. 0531194) and NSF Award No. 1064146 to Georgia Tech ("Revealing Innovation Pathways: Hybrid Science Maps for Technology Assessment and Foresight"). Revisions were partially supported in collaboration with the Center for Nanotechnology in Society, University of California Santa Barbara (NSF Awards No. 0938099 and No. 0531184). The findings and observations contained in this paper are those of the authors and do not necessarily reflect the views of the US National Science Foundation.