



It's all about KNOWLEDGE

Gaston Heimeriks

"Measuring
and modeling
dynamics in
innovation
systems"

Twitter:
Email:
WWW:

@GastonHeimeriks
G.J.Heimeriks@uu.nl
Heimeriks.net



Outline

9.00-10.45 Gaston Heimeriks

- It's all about knowledge
- Strengths and weaknesses of using patents and publication data to measure knowledge dynamics
- Empirical examples

11.00-12.45 Edwin Horlings

- science and policy
- the use of Saint
- network analysis

13.45-15.30 Edwin Horlings and Gaston Heimeriks

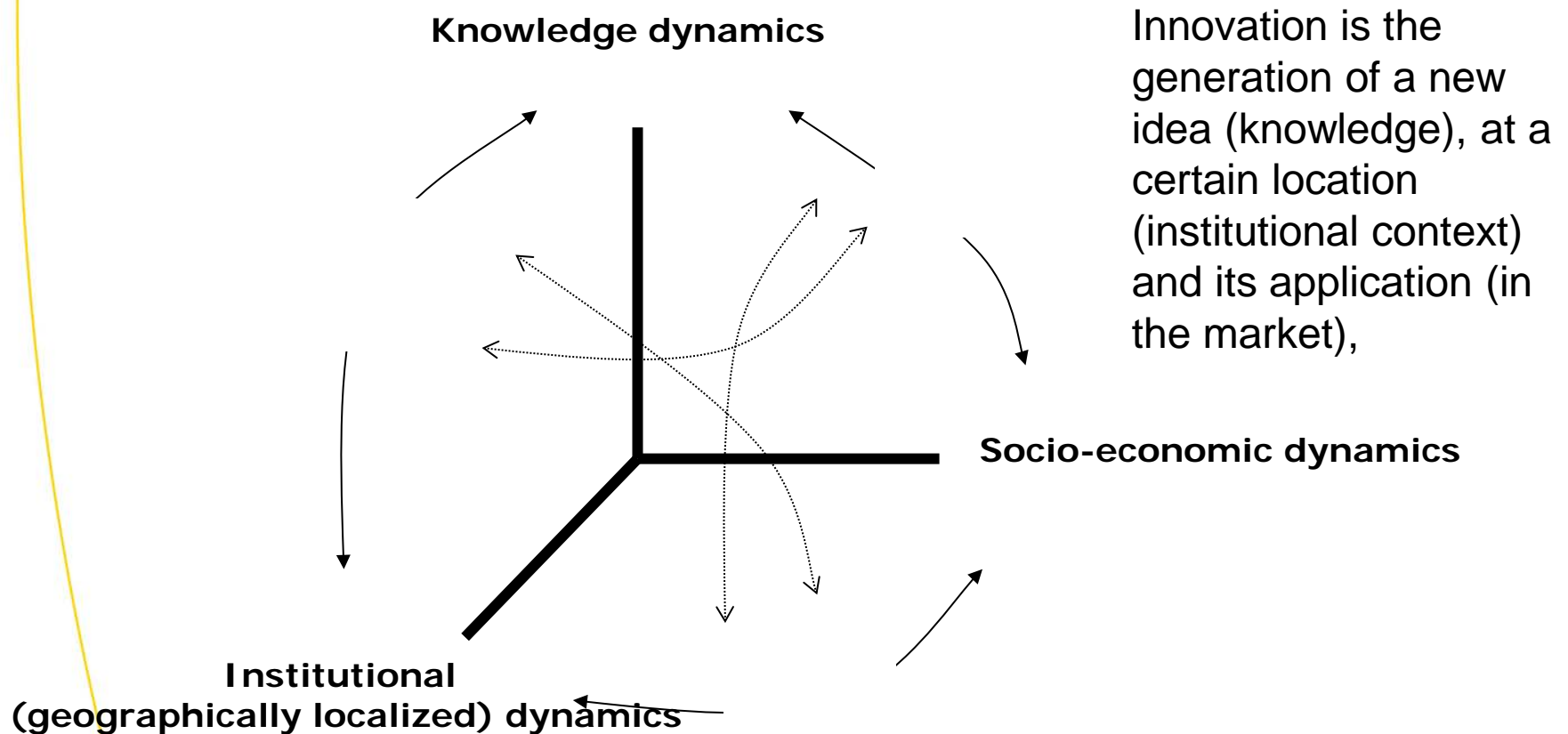
- Hands-on exercise scientometric data using SAINT



It's all about knowledge



What is innovation?





It's all about knowledge

- Knowledge is central to economic performance (Nelson and Winter, 1982; Romer, 1994; Schumpeter, 1943).
- Rich countries and regions are rich because they exploit progress in knowledge.
- The importance of knowledge production has further increased because of economic globalisation, and the ease of transmitting codified information across geographical space (David & Foray, 2002; Heimeriks & Vasileiadou, 2008).
- The term 'knowledge-based economy' reflects this fuller recognition of the place of organised knowledge in modern societies (OECD, 1996).



Knowledge is complicated

- Producing knowledge is different from the production of other goods and services.
- Knowledge can be shared freely, but it also is very individualized and difficult to transmit to others.
- Some knowledge is embodied in machinery, capital equipment, and complex systems.
- Other knowledge must be learned through experience, observation, research, or apprenticeship.
- From an evolutionary perspective, existing knowledge provides building blocks for further knowledge production.
- New knowledge evolves from the chaotic and constant recombining of already existing knowledge building blocks (Arthur, 2007).



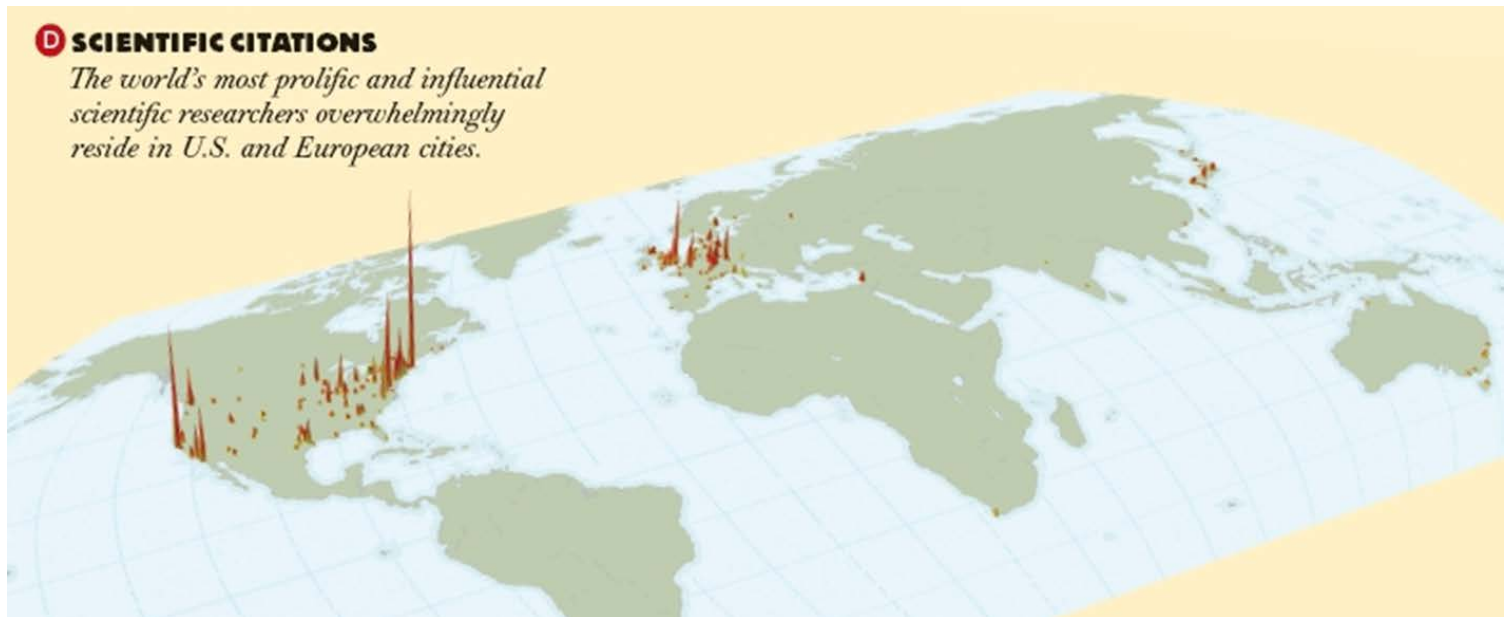
Knowledge is distributed

- knowledge is differentiated among locations, given that it is specific to the context in which it is created.
- proximity matters in lowering the barriers and costs of knowledge sharing and transmission
- Due to its tacit nature, knowledge has unique and characteristic features in each new learning environment.
- Furthermore, knowledge developments are partially irreversible: once new topics and the accompanying skills and routines have moved on, previous or simpler topics are 'forgotten'



It's all about knowledge

Knowledge – and innovative activity – are geographically clustered, and the “tendency toward spatial concentration has become more marked over time, not less” (Asheim and Gertler, 2005, p. 291)



‘Spiky World’ Richard Florida (2005)



Historical context

- Not long ago, the economics of knowledge had an ill-defined, “black box” character. Foray (2004) describes the “comfortable world” of standard models, in which only some agents and institutions (such as research and development (R&D) laboratories) and sectors (“knowledge industries”) were specialized in the production of knowledge.
- The output of this production, modeled by the knowledge production function (Griliches, 1979; Pakes and Griliches, 1984) is invention, typically represented by patents. See Audretsch (2003) and Feldman (1999) for summaries of the conventional approach.



Knowledge: Public good?

- Seminal work by Arrow (1962) presented a basic model of invention, R&D innovation and imitation.
- He argues that there is a tendency in industry to underinvest in R&D from society's point of view, due to problems for a firm to appropriate the economic benefits of its R&D. Patent protection would be one means for coping with this.



Collective action problem

- The state must play some role in the provision of public goods; otherwise they will be undersupplied.
- Collective action problem
 - why invest in something if the returns cannot be appropriated? (free rider problems)
 - how to invest in something that makes us all better off?

The benefits from the development of the transistor, the laser, or the mathematical algorithms that underlay the modern computer have been an enormous, extending well beyond benefits accruing to those who made or financed these innovations and discoveries.



Two strategies....

- Governments have pursued two different strategies in addressing these concerns.
 1. To increase the degree of appropriability of the return to knowledge, by issuing patents and copyright protection.
 2. Direct government support for basic research



.. with disadvantages

1. Tragedy of the anti-commons

- can raise the price of one of the most vital inputs into the innovative process and thus reduce the pace of follow-on innovations, even as it may provide returns to those making the original innovation. As a result, the overall pace of technical progress may be slowed

2. Information problem

- What to fund?

Necessary where the costs of the improved appropriability strategy are high. This is particularly true in the case of basic research because its benefits are widespread and diffuse and because attempts to appropriate its returns may significantly slow the overall pace of innovation. Indeed, many advances in basic knowledge--such as mathematical theorems--are not patentable, in spite of their importance and their potential practical applications.



	Rival consumption	Non-rival consumption
High appropriability	Private goods Milkshake	Club goods Sport stadiums
Low appropriability	Common pool resources Forests	Public goods National security

Two logics in **research**:

Disclosure: free circulation of information

Private property: retention of information



Science: a public good?

- Open science? Absorptive capacity (science as a club good?)!



Traditional neo-classical economy

- Economists label the all-purpose improvement factor responsible for growth “technology”—though it includes things like better laws and regulations as well as technical advance—and measure it using a technique called “growth accounting”.
- In this accounting, “technology” is the bit left over after calculating the effect on GDP of things like labour, capital and education. And at the moment, in the rich world, it looks like there is less of it about.



Innovation and economic growth



$$Y(t) = [K(t)]^\alpha A(t) L(t)^{1-\alpha}$$

where:

- $Y(t)$ represents the total production in an economy (the GDP) in some year, t .
- $K(t)$ is capital in the productive economy - which might be measured through the combined value of all companies in a capitalist economy.
- $L(t)$ is labour; this is simply the number of people in work, and since growth models are long run models they tend to ignore cyclical unemployment effects, assuming instead that the labour force is a constant fraction of an expanding population.
- $A(t)$ represents multifactor productivity (often generalized as "technology"). The change in this figure from $A(1960)$ to $A(1980)$ is the key to estimating the growth in labour 'efficiency' and the Solow residual between 1960 and 1980, for instance.

Almost 90%!!!



knowledge production function

ignores the fact that knowledge production is different from the production of goods in several ways.

1. uncertainty is inherent in the entire process of technological change.
2. knowledge is embodied not only in capital goods, as commonly modeled, but also in people
3. knowledge also is embodied in organizations, taking the form of organizational routines. Within firms, the resultant knowledge is greater than the sum of the individual knowledge



Innovation Systems

- The SI approach has its roots in the evolutionary theory (Nelson and Winter 1982).
- Knowledge is not only information, but also tacit knowledge; and can be both general and specific and is always costly. Knowledge can be specific to the location, firm or to the industry (Smith 2000).
- While in the neoclassical approach information asymmetries are considered to be a market failure, under the evolutionary theory and the SI approach asymmetric information is essential to provide novelty and variety.



Evolutionary economics; the localised institutional perspective

- Markets are evolutionary systems that each day carry out millions of simultaneous experiments on ways to make our lives better. The essential role of capitalism is not allocation—it is **creation**.
 - Life isn't drastically better for billions of people today than it was in 1800 because we are allocating the resources of the 19th-century economy more efficiently.
- It is better because we have life-saving antibiotics, refrigerators, access to vast amounts of information, and an enormous number of technical and social innovations that have become available to much (if not yet all) of the world's population.
- The genius of capitalism is that it both creates incentives for solving human problems and makes those solutions widely available.



From an evolutionary perspective

- There is no rational choice and equilibria.
- The efficient operation of markets is limited by the uncertainty of conditions and outcomes of innovation.
- The stuff out of which 'market failures' are made from the perspective of mainstream economics – such as asymmetric information, radical uncertainty, cumulative knowledge, path dependence, lack of equilibrium, and rigidities – are from an evolutionary perspective the stuff from which markets and innovation are made



- The Systems of Innovation (SI) approach shifts the focus away from actions at the level of individual and isolated units within the economy (firms) towards that of the collective underpinnings of innovation.
- It addresses the overall system that creates and distributes knowledge, rather than distributed individual components, and innovations are seen as outcomes of evolutionary processes within these systems.
- Government intervention to build up institutional context
 - E.g. Korea (Stiglitz)



Rationales and uncertainty?

- While the market failure argument provides a useful framework for existing products and services (e.g. bread, milk, cars), it is less suited for knowledge and innovation.
- Time dimension, the Knightian uncertainty
 1. Can the knowledge be developed (e.g. a new drug, a new energy technology)
 2. Can we develop the knowledge here in this institutional context, with our skills, infrastructures and resources?
 3. Can it be sold?
- These three uncertainties translate broadly into the three dominant rationales in science and innovation policy.

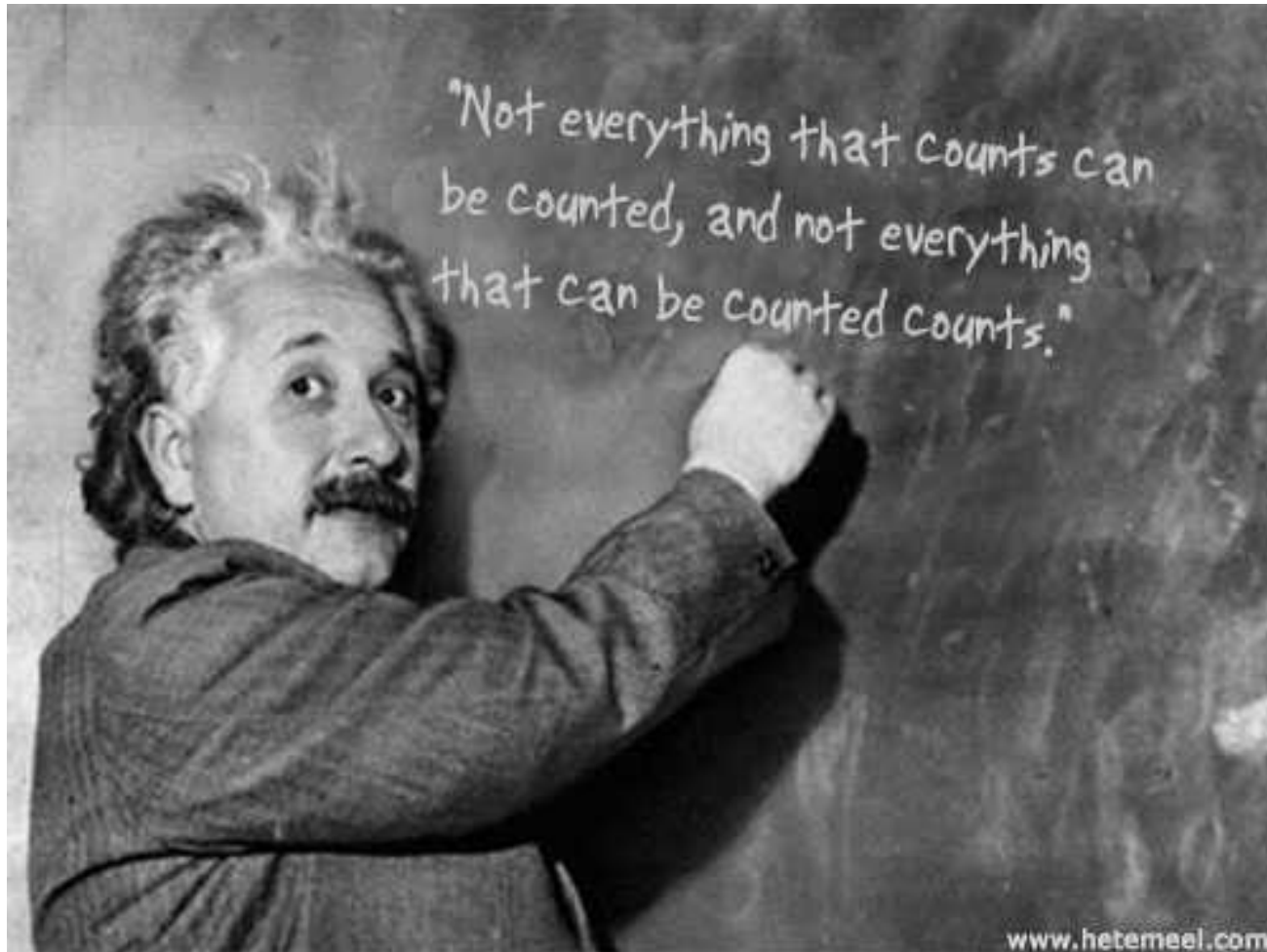


Using patent and publication data



Victor Mayer-Schönberger on Big Data

- As the scale of information increases, so does the number of inaccuracies. In a sample, it is important that the figures are as correct as possible.
- With big data, Mayer-Schönberger argues that the amount of data gives us a more valuable output, even though more errors may occur.
- A tendency to move from causality to correlations is the third characteristic described by Mayer-Schönberger. New data-mining techniques can give us information about what is happening, without explaining why





Scientometrics

- Scientometrics is the science of measuring and analysing science
- Scientific articles and citations between articles are the main objects in scientometrics
- Scientometrics was found in the sixties with the development of the Science Citation Index (SCI)
 - "The Citation Index is an alphabetic list of references given in bibliographies and footnotes of source articles arranged by first author. Each reference is followed by brief descriptions (citations) of the source articles which cite it."
- However, the online availability of articles including references (SCI) led to large opportunities for scientometrics



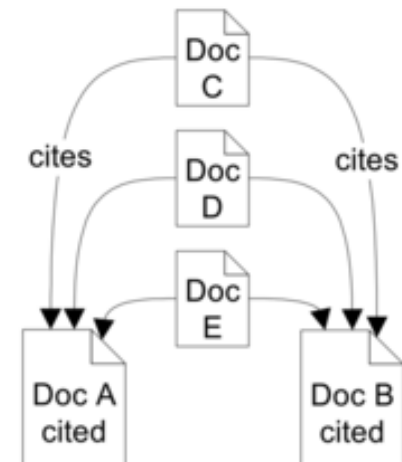
Relational analysis

- So we can access huge amounts of data, but how do we make sense of it?
- A useful way of analyzing the data is by investigating relations between the various attributes (author, year, citations etc.)
- Five types of relational analysis will be discussed in turn:
 - Co-citation analysis
 - Bibliographic coupling
 - Direct citation
 - Co-words
 - Word-reference co-occurrences
- Mixed modes of these analysis may also be used, because results out of different analyses can be complementary



Co-citation analysis

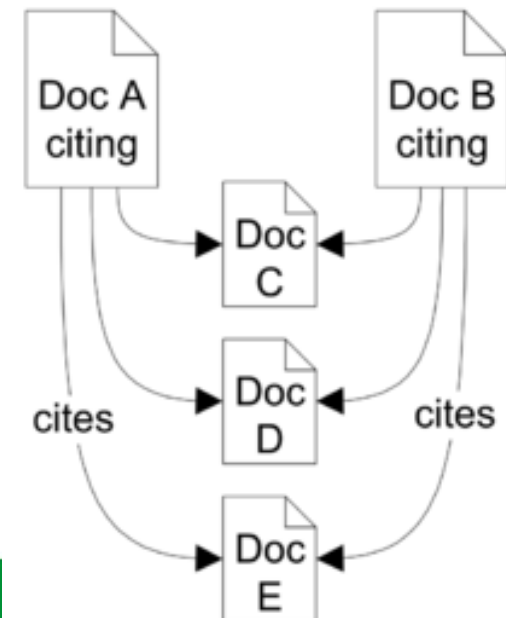
- Co-citation = frequency with which two documents are cited together by other documents.
- The more co-citations two documents receive, the higher their co-citation strength.
- The higher the co-citation strength the stronger the relation between articles.
- Two sorts of co-citation:
 - Co-citation clustering: which is simply the formation of clusters of co-cited documents (including A&B)
 - Co-citation analysis: which takes the result of co-citation clustering, and then assigns current papers (or papers from the research front) to the co-citation clusters (including A-E)
- Co-citation not suited for relatively new articles, because they have not yet received many citations
- Co-citation analysis benefits from the use of external references





Bibliographic coupling

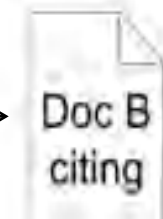
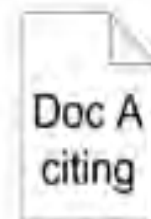
- Bibliographic coupling occurs when two works reference a common third work in their bibliographies. It is an indication that a probability exists that the two works treat a related subject matter
- The "coupling strength" of two given documents is higher the more citations to other documents they share
- Bibliographic coupling is able to cluster recent papers but clusters fewer of the old papers
- Opposite to co-citation analysis





Direct citation

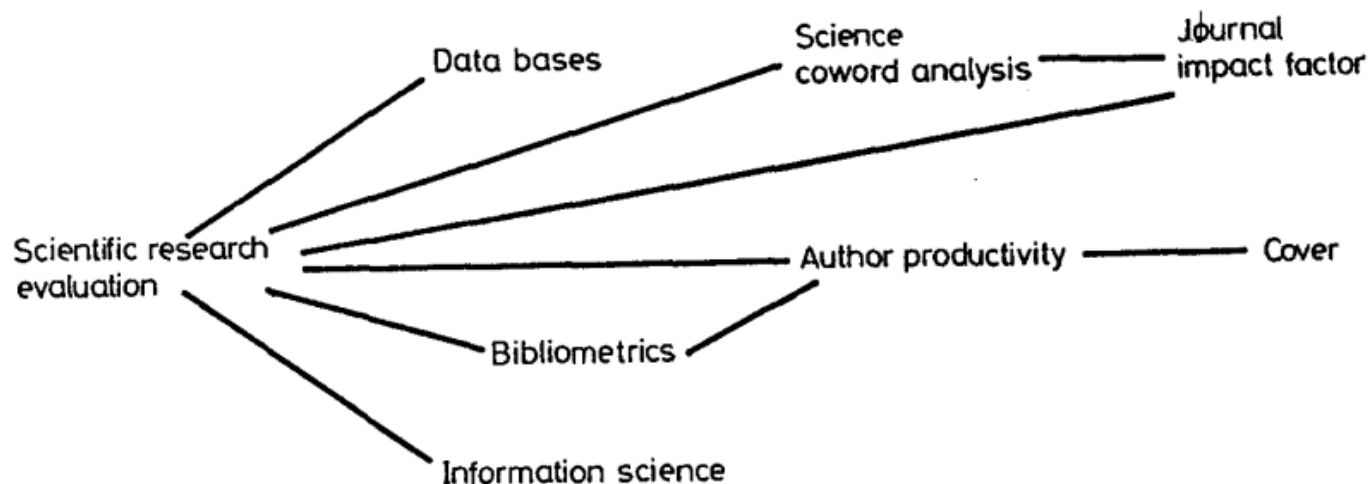
- Incorporates all direct citation for an article into the analysis
- Only considers links from within the set, e.g. no external references
- Need to use very long time windows to obtain a sufficient linking signal for clustering
- Direct citation clusters documents more evenly across the time window
- Tends to cluster a large number of documents than either coupling or co-citation





Co-words

- Not based on citation but on key-words attached to articles
- Subsequently, network of articles can be built by sorting based on strength and combining clusters
- Co-word analysis gives the same weight



fields
ey words,
s are often
)



Word-reference co-occurrences

- Combination of title words and cited references
- References are needed because words can have different meanings
- Lead to sets of papers that are similar in terms of these word-reference combinations
- Advantage is that it relies on both citing behavior and word choice of researchers
- Higher percentage of source documents are included in the analysis than co-citation mappings



Patents

- Cover a broad range of technologies
- Close link to invention
- Contains detailed information on invention process
- Spatial and temporal coverage of patent data is unique
- Patent data is public and readily available at low cost
- Direct outcome of inventive process (those expected to have commercial impact)
- Patented inventions expected to provide



Drawbacks

- Not all inventions are technologically patentable
- Not all inventions are patented
- Different propensity to patent (across industries; domestic vs foreign)
- Value distribution of patents is highly skewed (patent counts misleading)
- Differences in patent law and practices around the world limit comparability
- Changes in patent laws over time call for caution when analyzing trends



Universiteit Utrecht

2014 STI conference
3-5 September 2014
Leiden,
The Netherlands

The evolution of PV technology

Patents as Instruments for
Exploring Innovation Dynamics:
Different Perspectives on
“Photovoltaic Cells”

- Leydesdorff, Alkemade, Heimeriks and Hoekstra (2014) *Geographic and Technological Perspectives on "Photovoltaic Cells: " Patents as Instruments for Exploring Innovation Dynamics* [arXiv:1401.2778](https://arxiv.org/abs/1401.2778)
- Heimeriks, Alkemade and Leydesdorff (in preparation) *The path and place dependent evolution of PV technology*

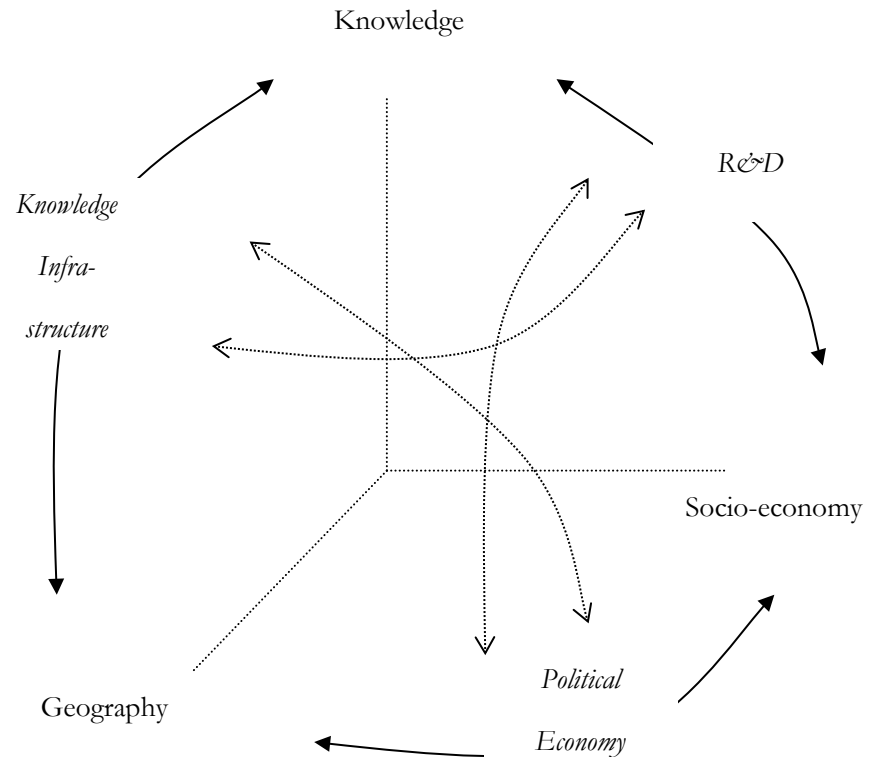
Introduction

- Technological invention is undoubtedly one of the keys to addressing societal challenges such as sustainability transitions.
- How does technological knowledge evolve over time, and to what extent do different locations contribute to this evolution in the context of globalisation?

Meta Model of Innovation

All models of innovation (implicitly) stress three analytical dimensions:

- Innovation is the creation of novelty (e.g. new knowledge)
- Innovation adds value in new ways (e.g. new market opportunities)
- Innovation is geographically localised (e.g. clusters in the spiky world)



Interacting dynamics

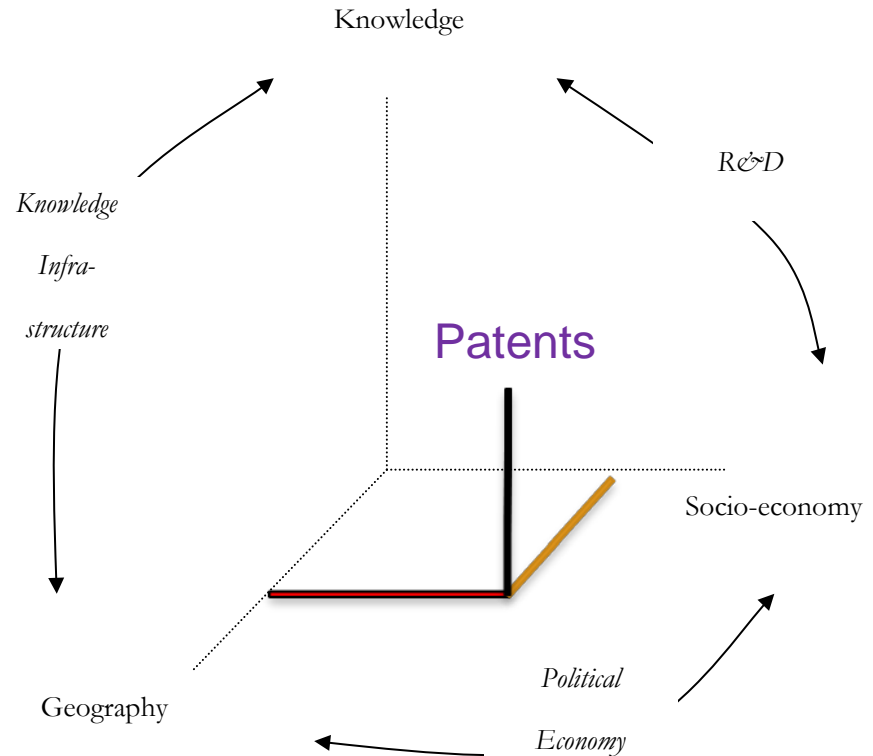
- Innovation has been considered 'lumpy', or discontinuous, in both a spatial, technological, economic, and temporal sense.
- We explore these interacting geographical, cognitive (technological) and economic developments over time

Hypotheses

- Technological change over time gives rise to a lumpy pattern of technological complexity: different generations can be identified;
- Geographical concentration occurs in combination with incremental technological knowledge developments and stable economic patterns;
- When novel technological building blocks become available, opportunities for new locations arise, associated with different local capabilities that may contribute to disruptive socio-economic opportunities, with new entrants and start-ups.

Data and Methods

- **Three environments are relevant to innovation:**
the context of technological knowledge production
the economic context
the geographical context
- **Patents reflect these different contexts;**
the addresses of inventors
patent classifications
assignee organisations



Patents as a lens

- The Cooperative Patent Classification system (CPC) was used to identify classes that match PV technologies
- **These new classifications have been backtracked into the existing databases (IPC) for indexing. The tag and its subclasses are now operational in both USPTO and PatStat data.**

Priority patents

PATSTATOCT13

IPC code	description	#PATSTAT
Y02E 10/541	CuInSe ₂ material PV cells	2212
Y02E 10/542	Dye sensitized solar cells	11627
Y02E 10/543	Solar cells from Group II-VI materials	1952
Y02E 10/544	Solar cells from Group III-V materials	3567
Y02E 10/545	Microcrystalline silicon PV cells	1029
Y02E 10/546	Polycrystalline silicon PV cells	2219
Y02E 10/547	Monocrystalline silicon PV cells	5065
Y02E 10/548	Amorphous silicon PV cells	4315
Y02E 10/549	organic PV cells	13247

		PATSTAT	USPTO
Y02E 10/541	CuInSe2 material PV cells	2212	422
Y02E 10/542	Dye sensitized solar cells	11627	532
Y02E 10/543	Solar cells from Group II-VI materials	1952	294
Y02E 10/544	Solar cells from Group III-V materials	3567	850
Y02E 10/545	Microcrystalline silicon PV cells	1029	146
Y02E 10/546	Polycrystalline silicon PV cells	2219	262
Y02E 10/547	Monocrystalline silicon PV cells	5065	1158
Y02E 10/548	Amorphous silicon PV cells	4315	742
Y02E 10/549	organic PV cells	13247	1340

Nine material technologies for photovoltaic cells distinguished in the Cooperative Patent Classifications (CPC).

The new Y02 classification scheme for climate change mitigation technologies makes it possible to explicitly identify PV patents and geographical locations with a strong PV knowledge base.

USPTO PATENT FULL-TEXT AND IMAGE DATABASE

[Home](#)
[Quick](#)
[Advanced](#)
[Pat Num](#)
[Help](#)
[View Cart](#)

CPC/Y02E10/5
41

Data current through September 17, 2013.

Query [\[Help\]](#)

Examples:

t1l/(tennis and (racquet or racket))

isd/1/8/2002 and motorcycle

in/newmar-julie

Select Years [\[Help\]](#)

Search

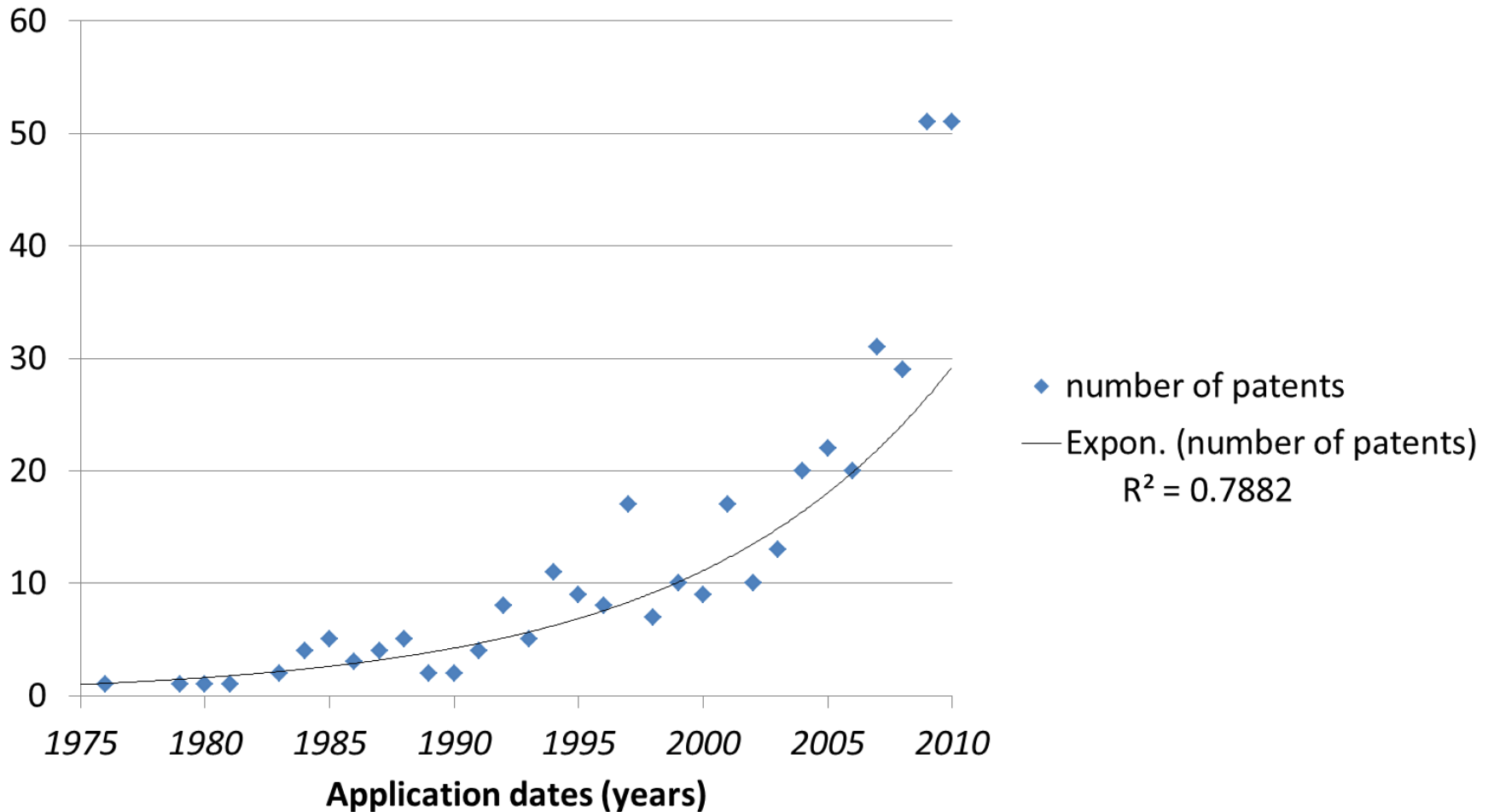
Reset

Patents from 1790 through 1975 are searchable only by Issue Date, Patent Number, and Current US Classification.

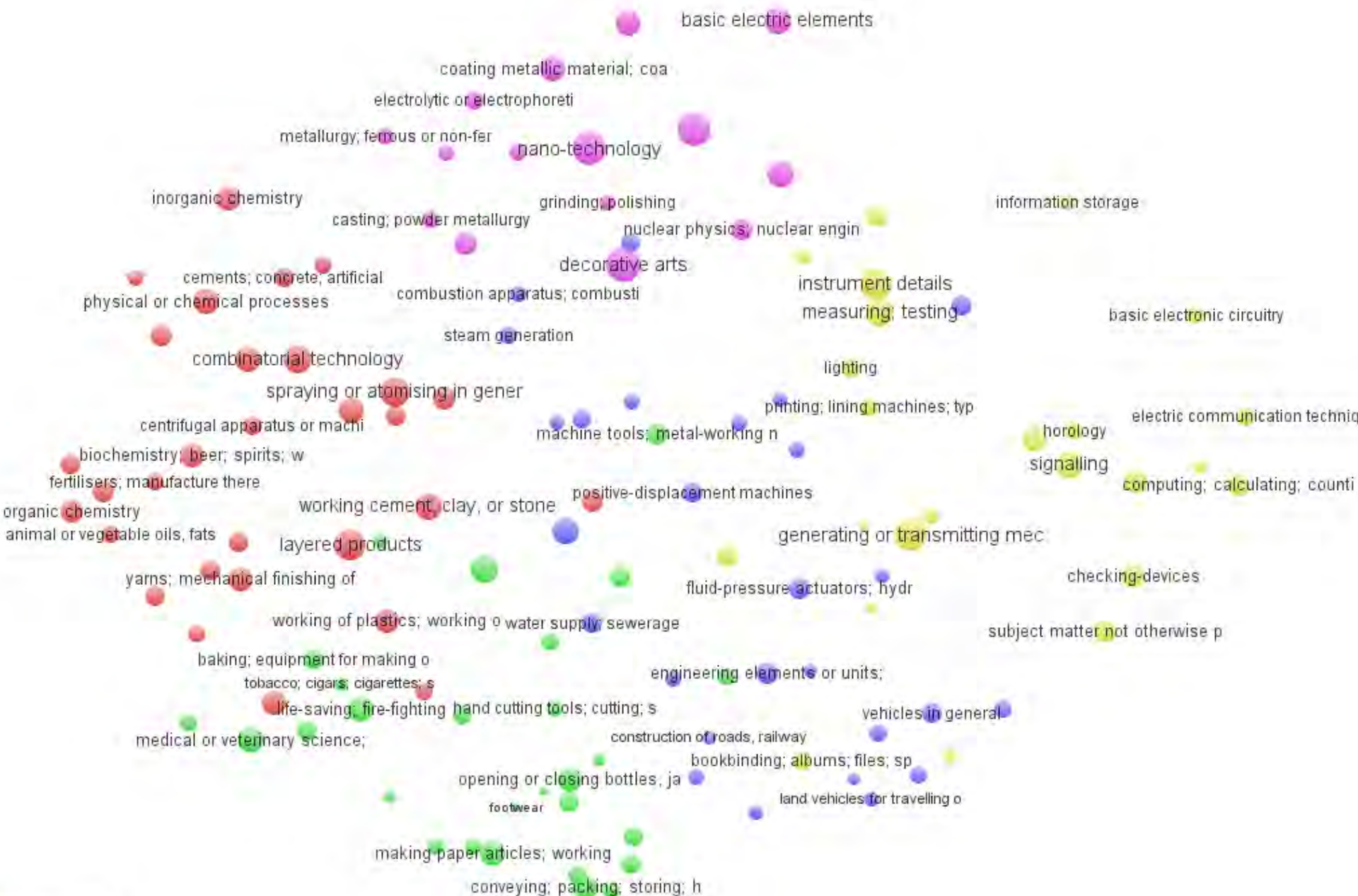
When searching for specific numbers in the Patent Number field, patent numbers must be seven characters in length, excluding commas, which are optional.

Field Code	Field Name	Field Code	Field Name
PN	Patent Number	IN	Inventor Name
ISD	Issue Date	IC	Inventor City
TTL	Title	IS	Inventor State
ABST	Abstract	ICN	Inventor Country
ACLM	Claim(s)	AANM	Applicant Name
SPEC	Description/Specification	AACI	Applicant City
CCL	Current US Classification	AAST	Applicant State
CPC	Current CPC Classification	AACO	Applicant Country
ICL	International Classification	AAAT	Applicant Type

419 patents in USPTO with "CPC/Y02E10/541"; Sep. 5, 2013
CuInSe₂ material PV cells



Cognitive dynamics



Base map for 124 IPC categories at the 3-digit level, using > 39 million US patents (1970)

basic electric elements

1974-
1978

physical or chemical processes

electric communication t

basic electric elements

1976-
1980

physical or chemical processes

basic electric elements

1988-
1992

coating metallic material; coa

electrolytic or electrophoreti

casting; powder metallurgy

spraying or atomising in gener

presses

basic electric elements

1989-
1993

coating metallic material; coa

electrolytic or electrophoreti

casting; powder metallurgy

spraying or atomising in gener

presses

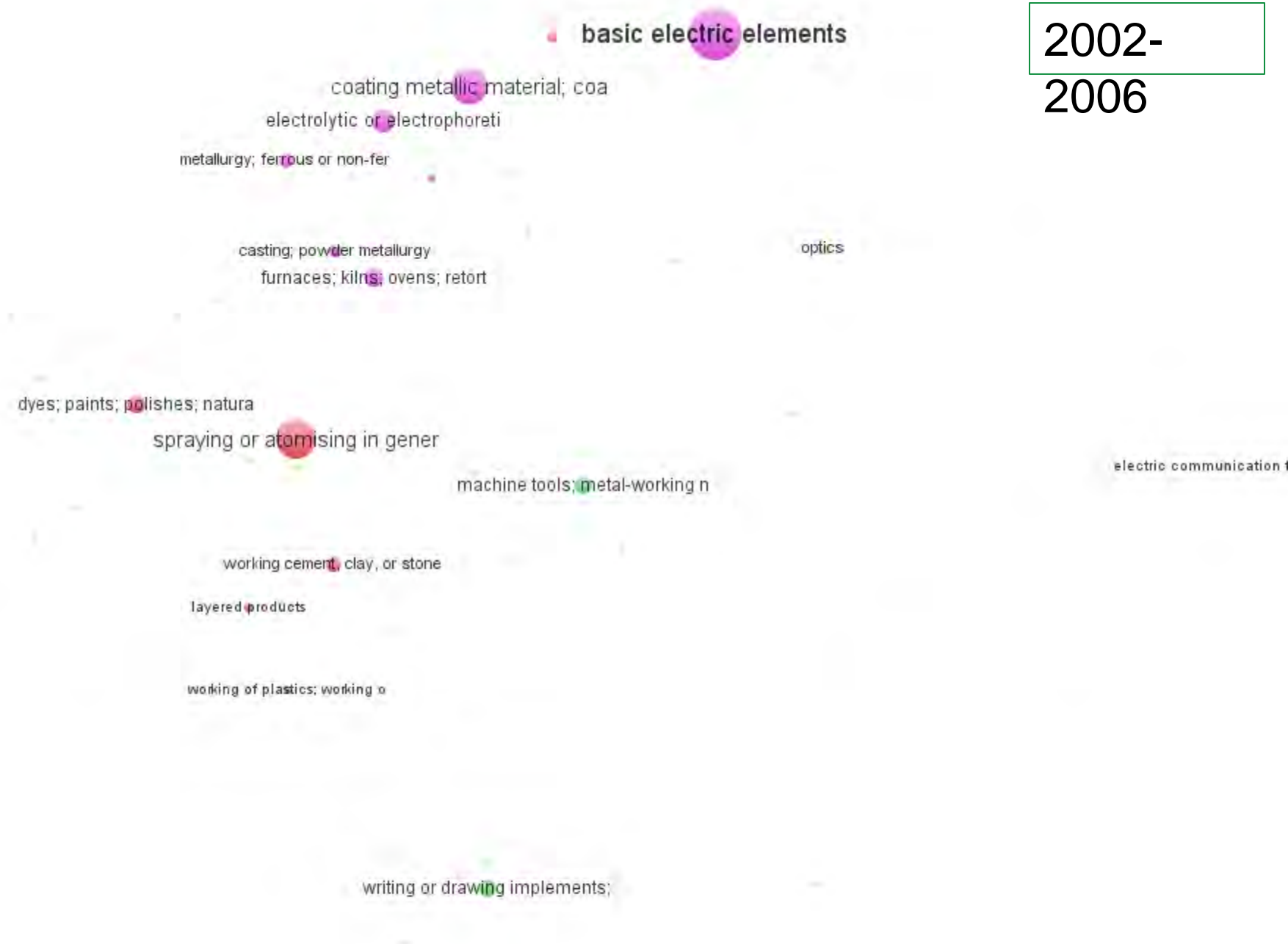
2000-
2004



2001-
2005



2002-
2006



2003-
2007

basic electric elements



2004-
2008

basic electric elements

coating metallic material; coa
electrolytic or electrophoreti

metallurgy; ferroous or non-fer

electric techniques not otherw

inorganic chemistry

casting; powder metallurgy

furnaces; kilns; ovens; retort

decorative arts

physical or chemical processes

measuring; testing

dyes; paints; polishes; natura

spraying or atomising in gener

printing; lining machines; typ

machine tools; metal-working n

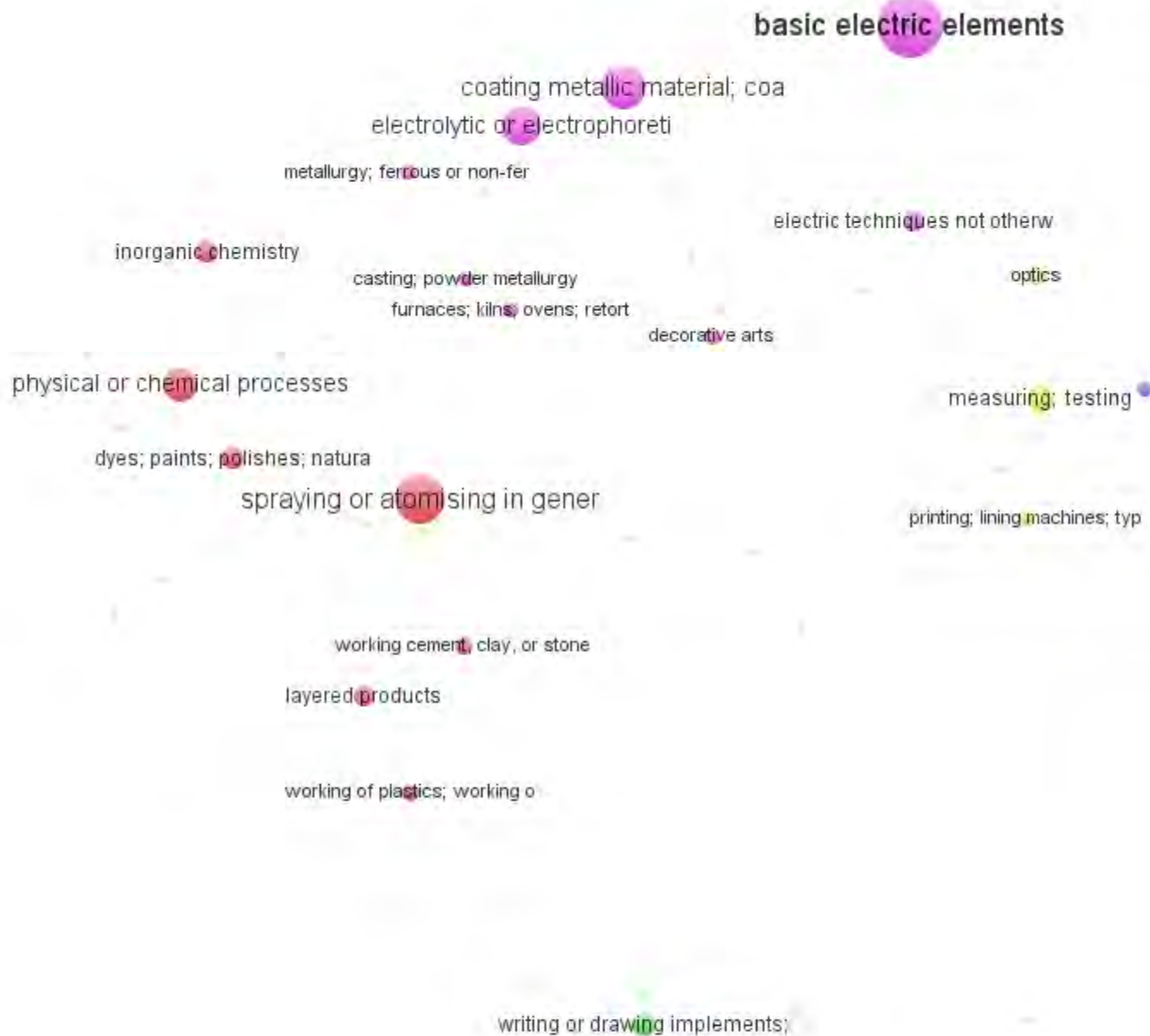
working cement, clay, or stone

layered products

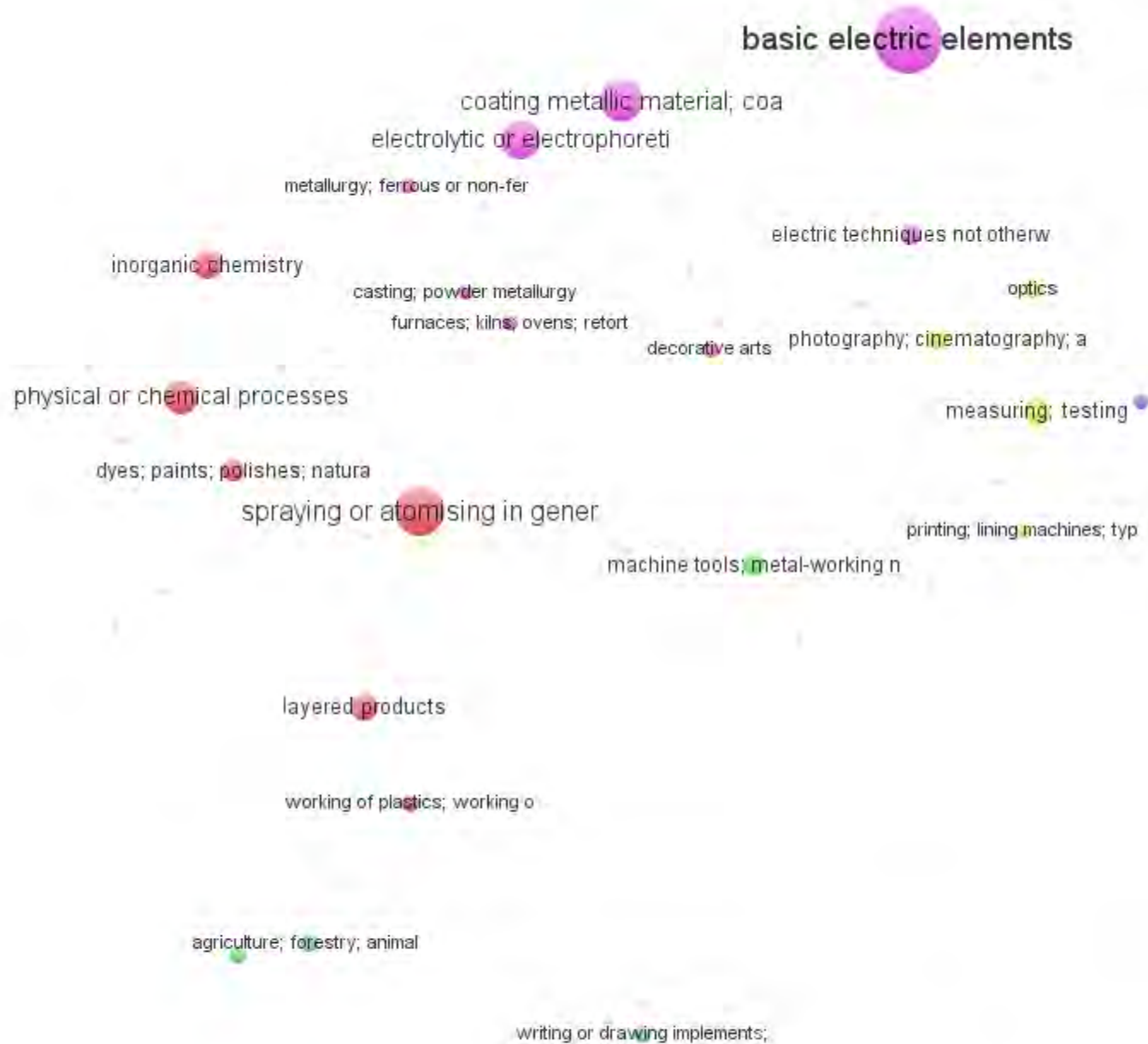
working of plastics; working o

writing or drawing implements;

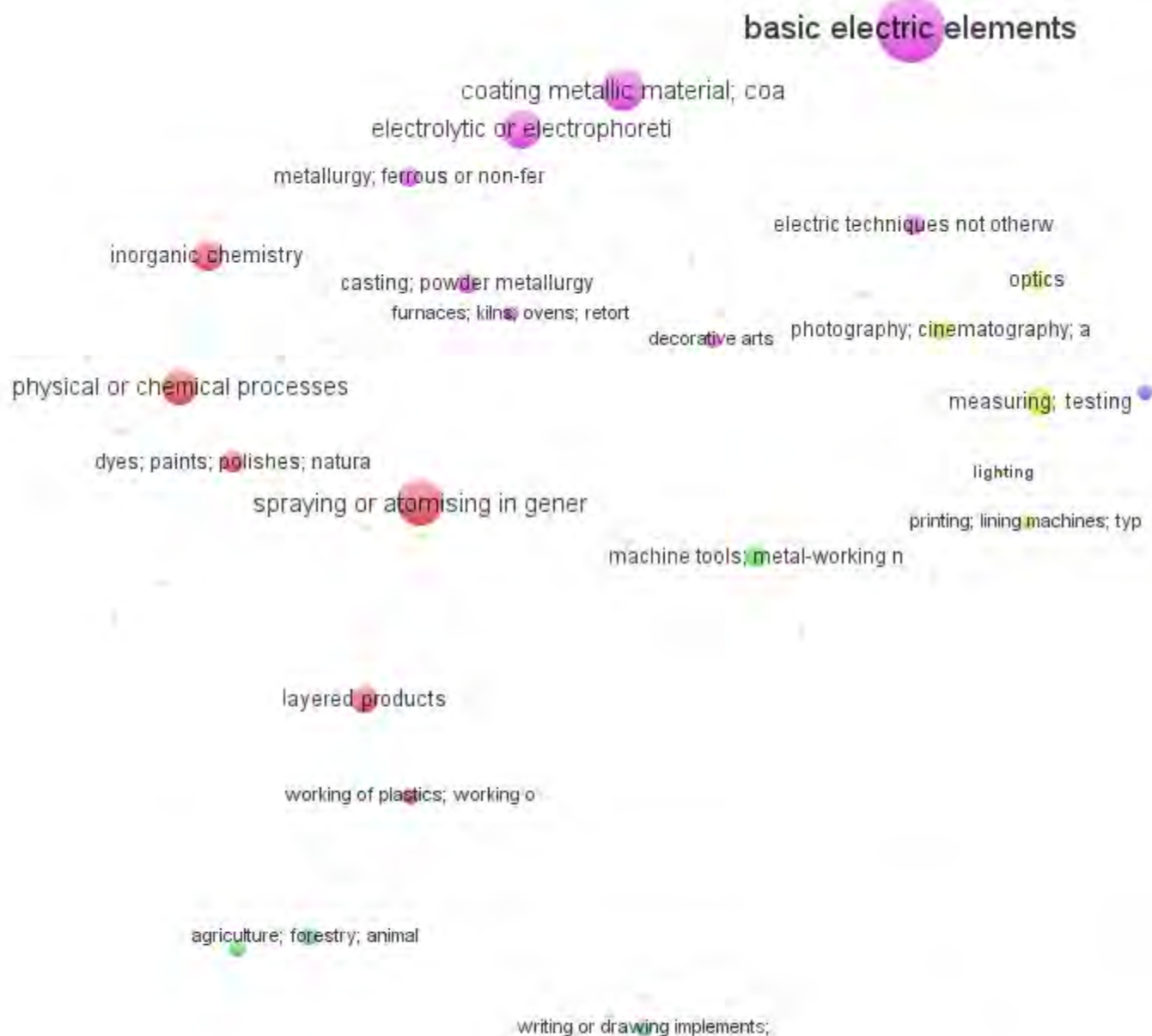
2005-
2009



2006-
2010



2007-
2011



2008-
2012

basic electric elements

coating metallic material; coa

electrolytic or electrophoreti

metallurgy; ferrous or non-fer

inorganic chemistry

casting; powder metallurgy

furnaces; kilns; ovens; retort

optics

decorative arts

photography; cinematography; a

physical or chemical processes

measuring; testing

dyes; paints; polishes; natura

lighting

spraying or atomising in gener

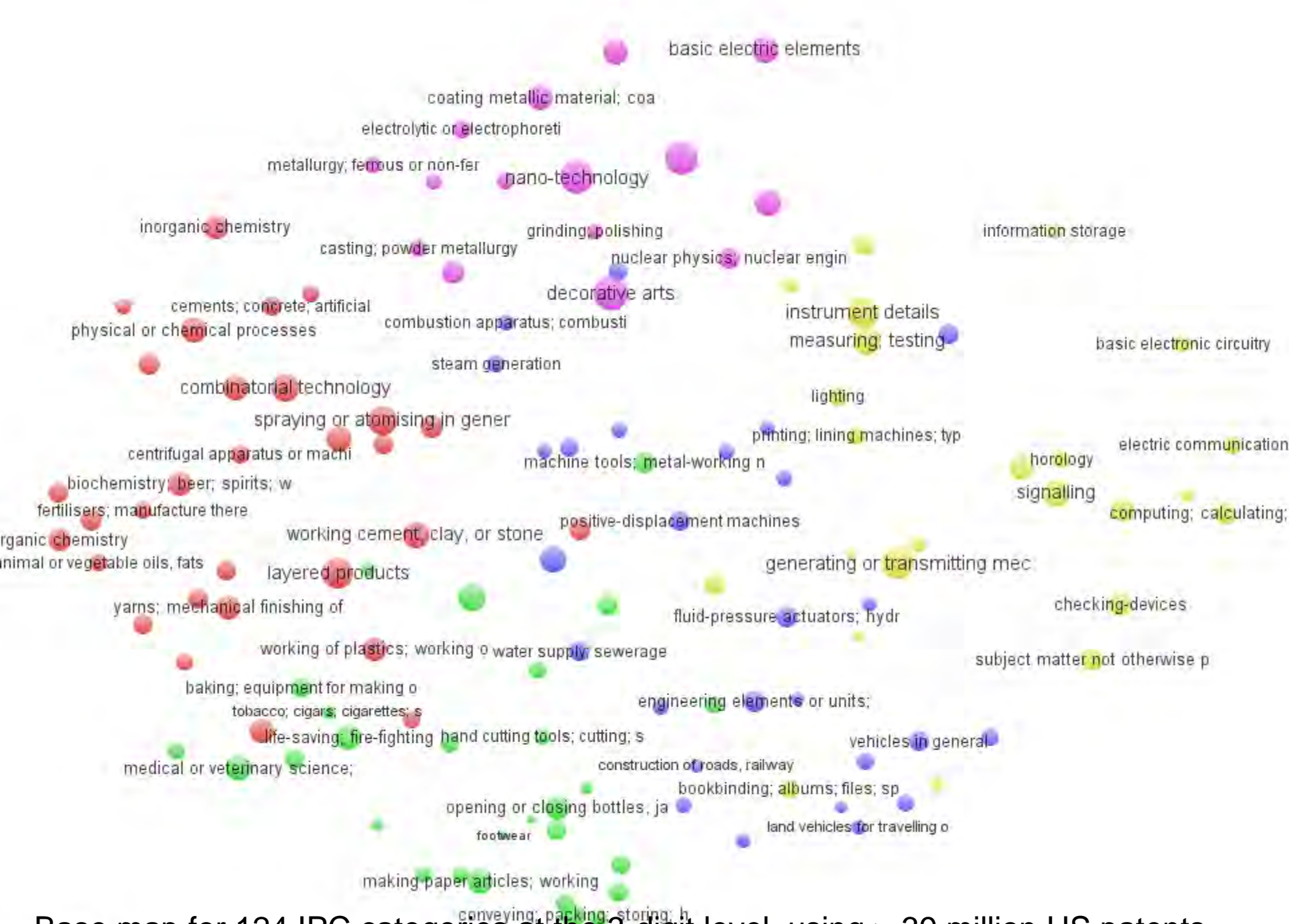
machine tools; metal-working n

layered products

working of plastics; working o

agriculture; forestry; animal

writing or drawing implements;



Base map for 124 IPC categories at the 3-digit level, using > 39 million US patents

Stirling diversity

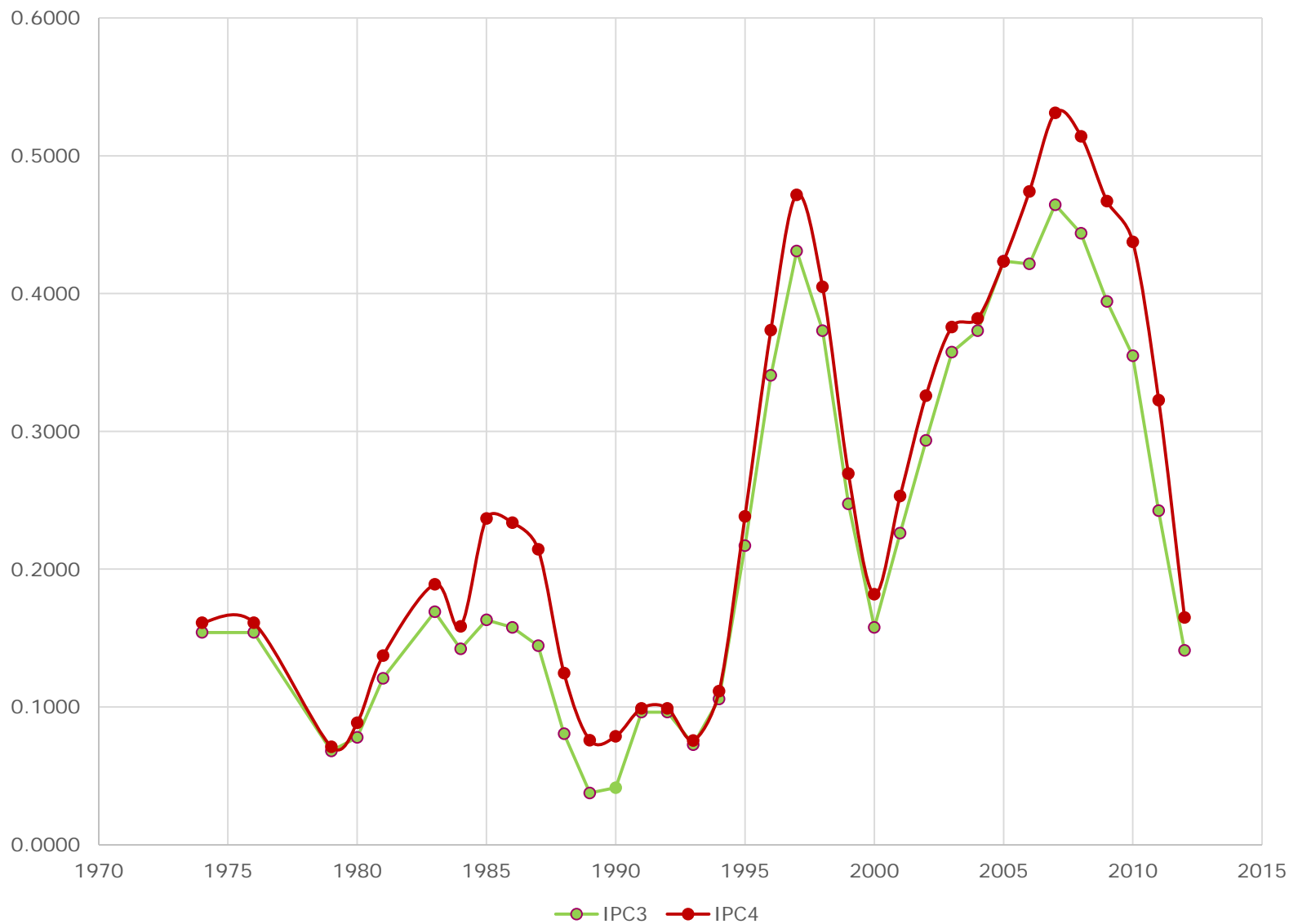
Technological diversity includes three entangled aspects; variety, balance, and disparity (Stirling, 2007), that are reflected in these maps:

1. the maps capture the variety (number) of technological classes;
2. they capture the technological balance by plotting the different sizes of the nodes;
3. maps convey the disparity (i.e. the cognitive distances) among technologies

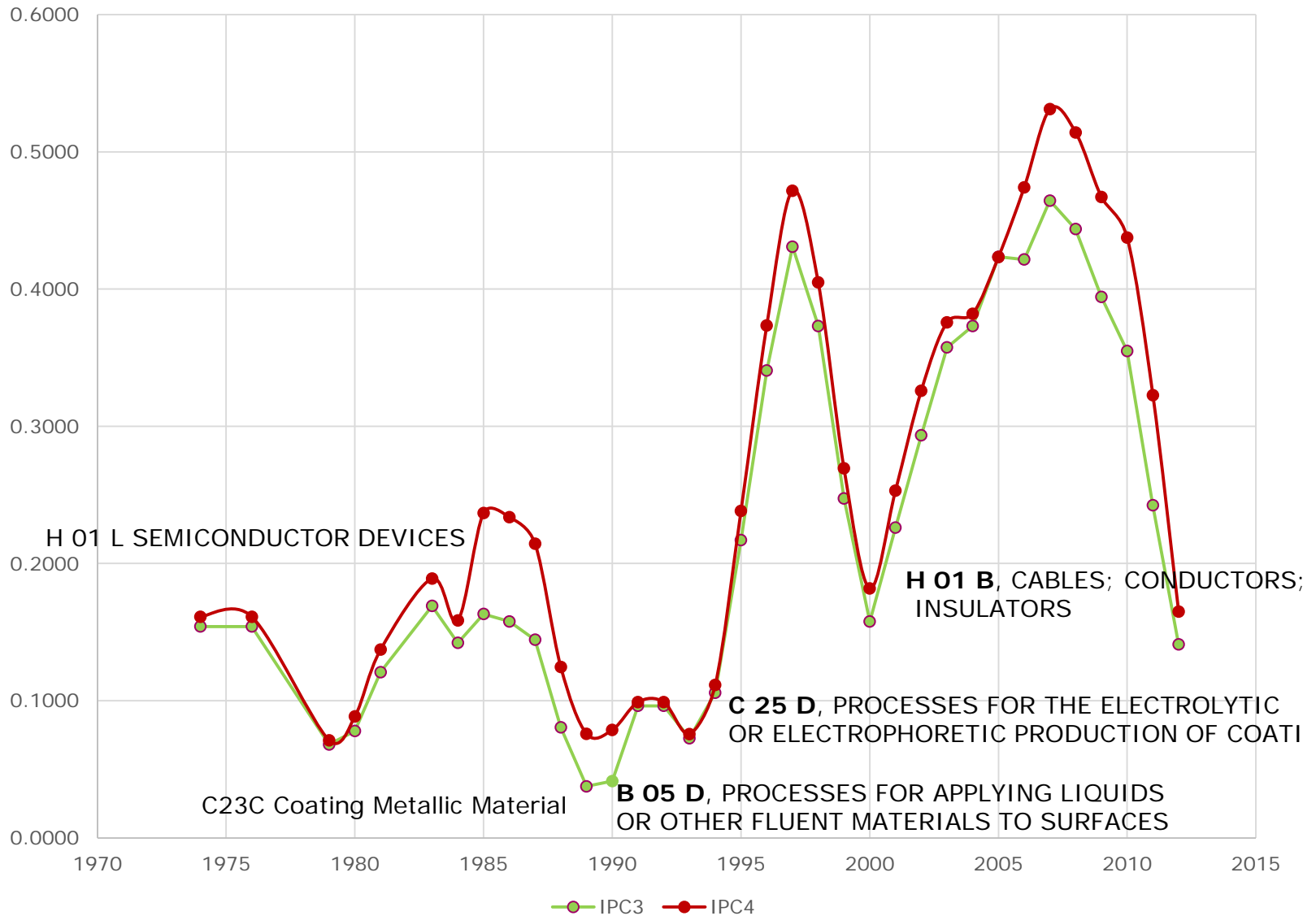
$$\Delta = \sum_{ij} p_i p_j d_{ij}$$

where d_{ij} is a disparity measure between two classes i and j —the categories are in this case IPC classes at the respective level of specificity—and p_i is the proportion of elements assigned to each class i

Rao-Stirling Diversity



Rao-Stirling diversity



Geographical dynamics

Development of inventor addresses in USPTO patents; Y02E10/541

- <http://www.leydesdorff.net/photovoltaic/cuinse2/animate.html>
- <http://data2semantics.github.io/PatViz>



1976-1980
next year

50%

Google map





1979-1983
next year

50%

Google map



Center: 28.00000,9.00000
Google 2000 km 2000 mi

Map created at [GPSVisualizer](#)
Map data ©2013 MapLink - Terms of



1980-1984
next year

50%



Google map



Center: 28.00000, 9.00000
2000 km
2000 mi

Map created at GPSVisualizer.com
Map data ©2013 MapLink - Terms of



1981-1985
next year

50%

Google map





1983-1987
next year

50%

Google map



Center: 28.00000,9.00000
Google 2000 km 2000 mi

Map created at [GPSVisualizer](#)
Map data ©2013 MapLink - Terms of





1985-1989
next year

50%

Google map



Center: 28.00000,9.00000
Google 2000 km 2000 mi

Map created at [GPSVisualizer](#)
Map data ©2013 MapLink - Terms of



1986-1990
next year

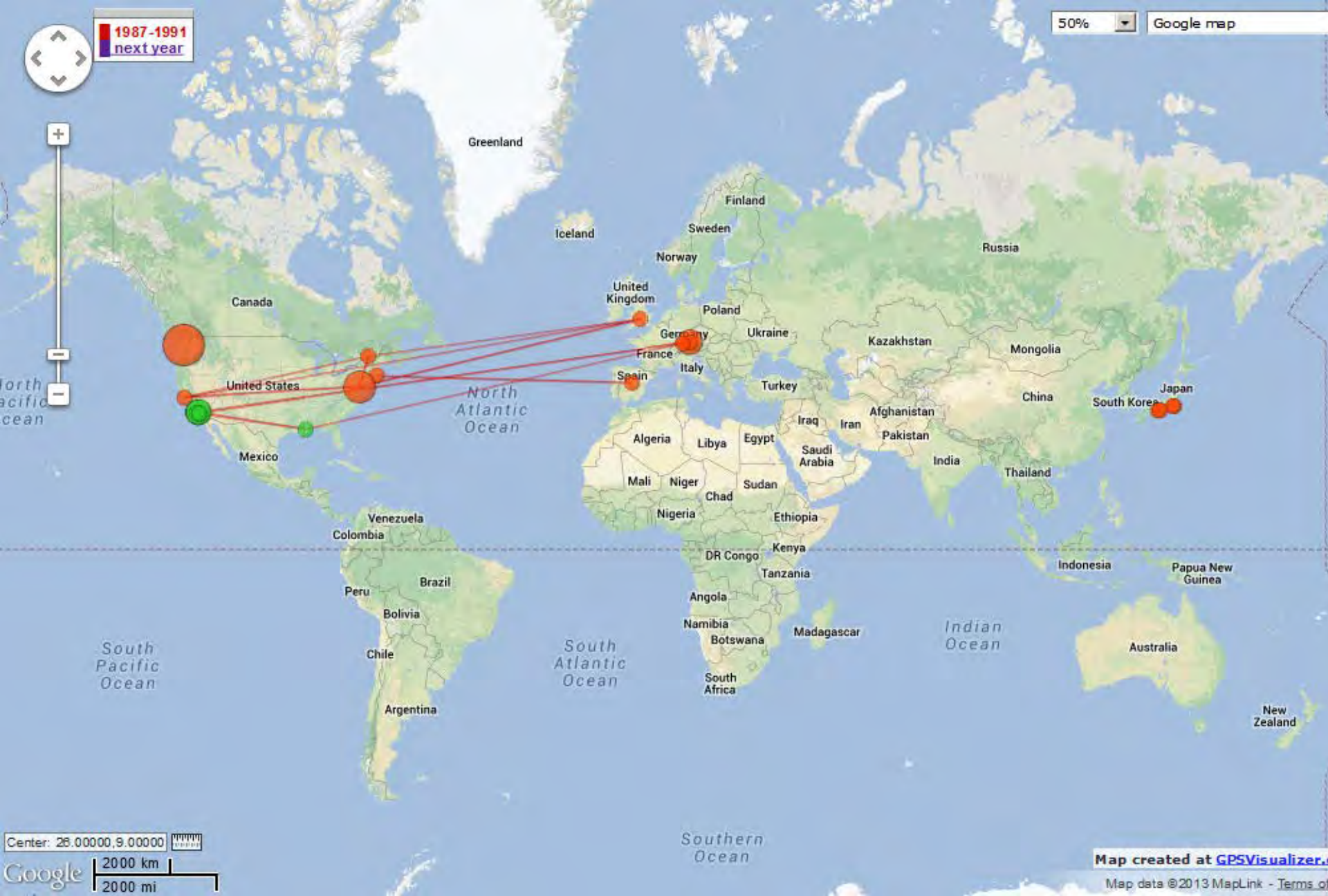




1987-1991
next year

50%

Google map



Center: 28.00000,9.00000
Google 2000 km 2000 mi

Map created at [GPSVisualizer](#)
Map data ©2013 MapLink - Terms of



1988-1992
next year

50%

Google map



Center: 28.00000,9.00000
2000 km
2000 mi

Map created at GPSVisualizer.com
Map data ©2013 MapLink - Terms of Use



1988-1992
next year

50%

Google map





1989-1993
next year

50%

Google map



Center: 28.00000,9.00000
2000 km
2000 mi

Map created at [GPSVisualizer](#)
Map data ©2013 MapLink - Terms of



1990-1994
next year

50%

Google map



Center: 28.00000,9.00000
2000 km
2000 mi

Map created at [GPSVisualizer.com](#)
Map data ©2013 MapLink - Terms of Service



1991-1995
next year

50%



Google map



Center: 28.00000,9.00000
Google 2000 km 2000 mi

Map created at [GPSVisualizer](#)
Map data ©2013 MapLink - Terms of



1992-1996
next year

50%

Google map





1993-1997
next year

50%

Google map



Center: 28.00000,9.00000
Google
2000 km
2000 mi

Map created at [GPSVisualizer](#)
Map data ©2013 MapLink - Terms of



1994-1998
next year

50%

Google map



Center: 28.00000,9.00000
Google
2000 km
2000 mi

Map created at [GPSVisualizer](#)
Map data ©2013 MapLink - Terms of



1995-1999
next year

50%

Google map



Center: 28.00000,9.00000
Google
2000 km
2000 mi

Map created at GPSVisualizer.com
Map data ©2013 MapLink - Terms of Service



1996-2000
next year

50%

Google map



Center: 28.00000,9.00000
Google
2000 km
2000 mi

Map created at [GPSVisualizer](#)
Map data ©2013 MapLink - Terms of





1999-2003
next year

50%

Google map



Center: 28.00000,9.00000
2000 km
2000 mi

Map created at [GPSVisualizer](#)
Map data ©2013 MapLink - Terms of



2000-2004
next year

50%

Google map



Center: 28.00000, 9.00000
Google 2000 km 2000 mi

Map created at [GPSVisualizer](#)
Map data ©2013 MapLink - Terms of



2001-2006
next year

50%



Google map



Center: 28.00000,9.00000
Google 2000 km 2000 mi

Map created at [GPSVisualizer](#)
Map data ©2013 MapLink - Terms of



2002-2006
next year

50%

Google map



Center: 28.00000,9.00000

Google 2000 km 2000 mi

Map created at [GPSVisualizer](#)

Map data ©2013 MapLink - Terms of



2003-2007
next year

50%

Google map



Center: 28.00000,9.00000
Google 2000 km 2000 mi

Map created at [GPSVisualizer](#)
Map data ©2013 MapLink - Terms of



2004-2008
next year

50%

Google map



Center: 28.00000, 9.00000
Google
2000 km
2000 mi

Map created at [GPSVisualizer](#)
Map data ©2013 MapLink - Terms of



2005-2009
next year

50%

Google map



Center: 28.00000,9.00000

Google 2000 km 2000 mi

Map created at GPSVisualizer.com

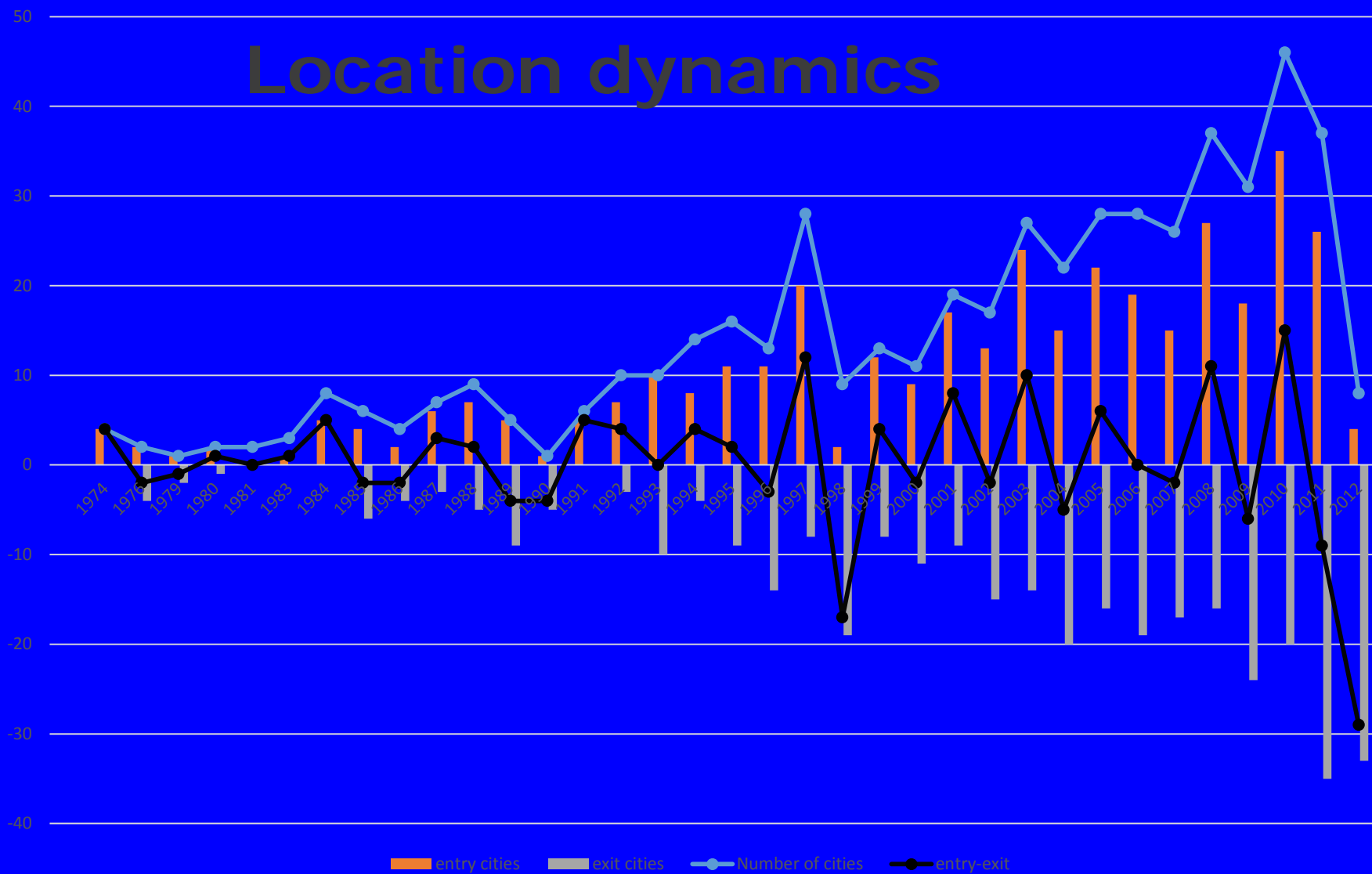
Map data ©2013 MapLink - Terms of



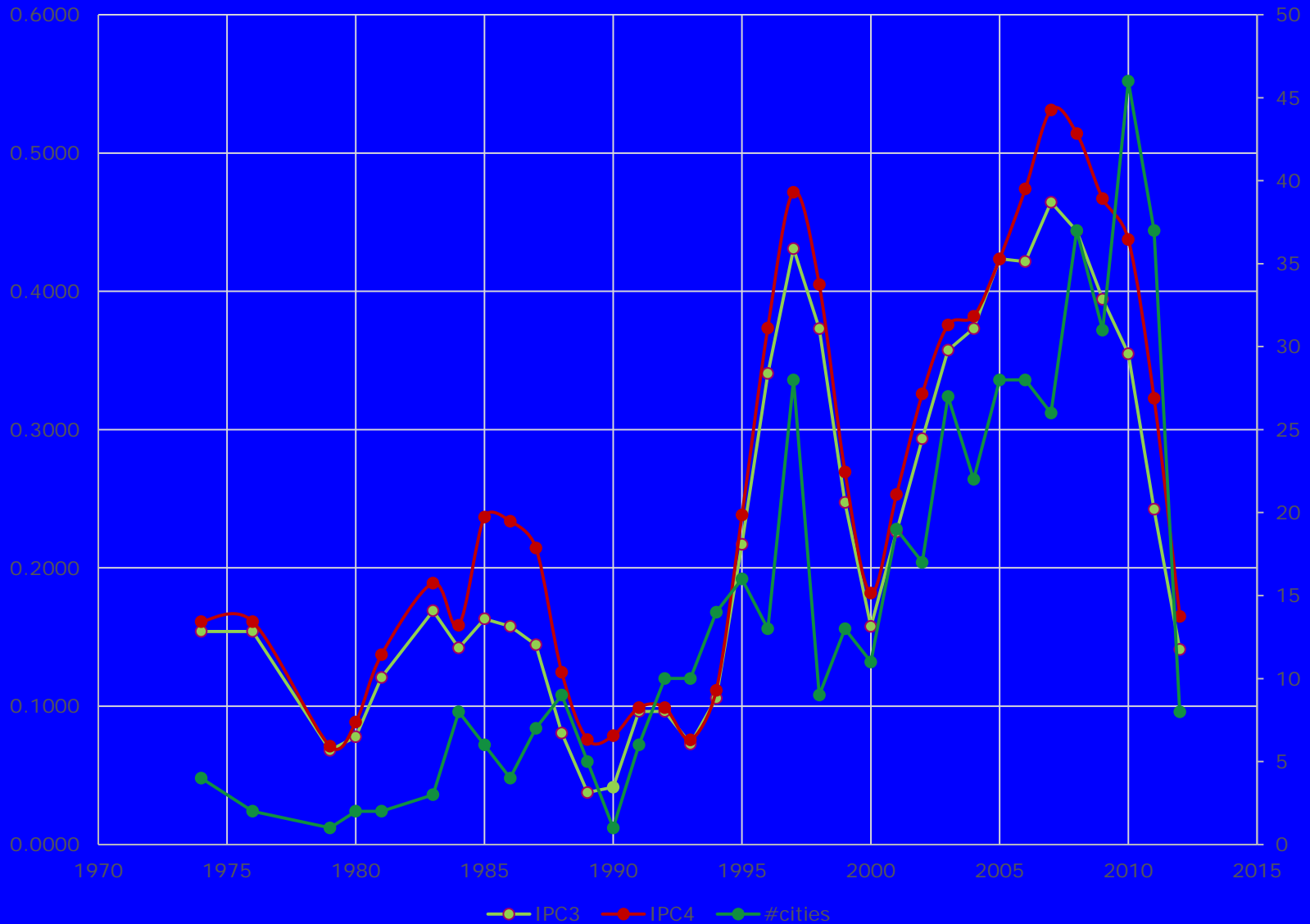




Location dynamics

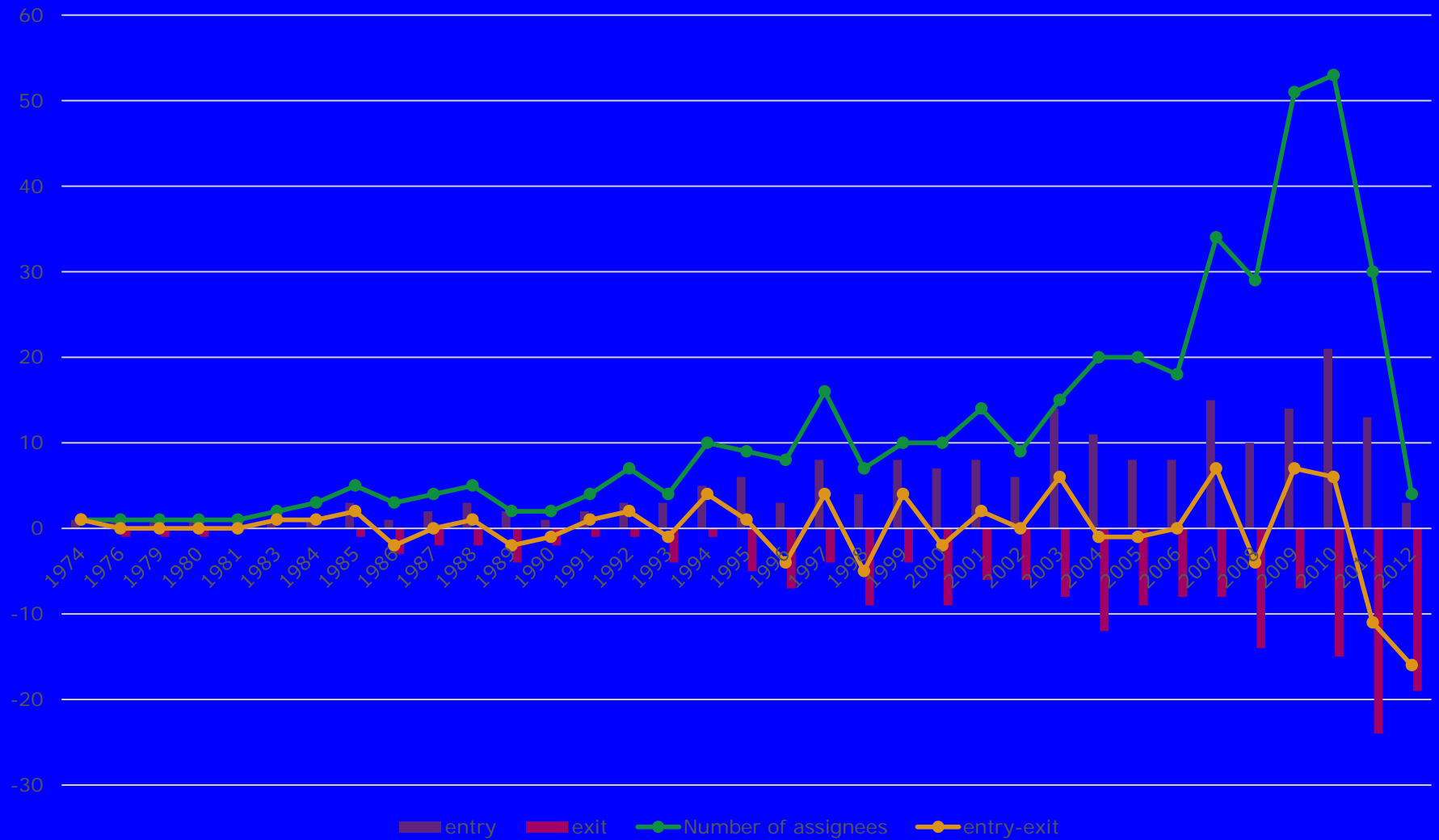


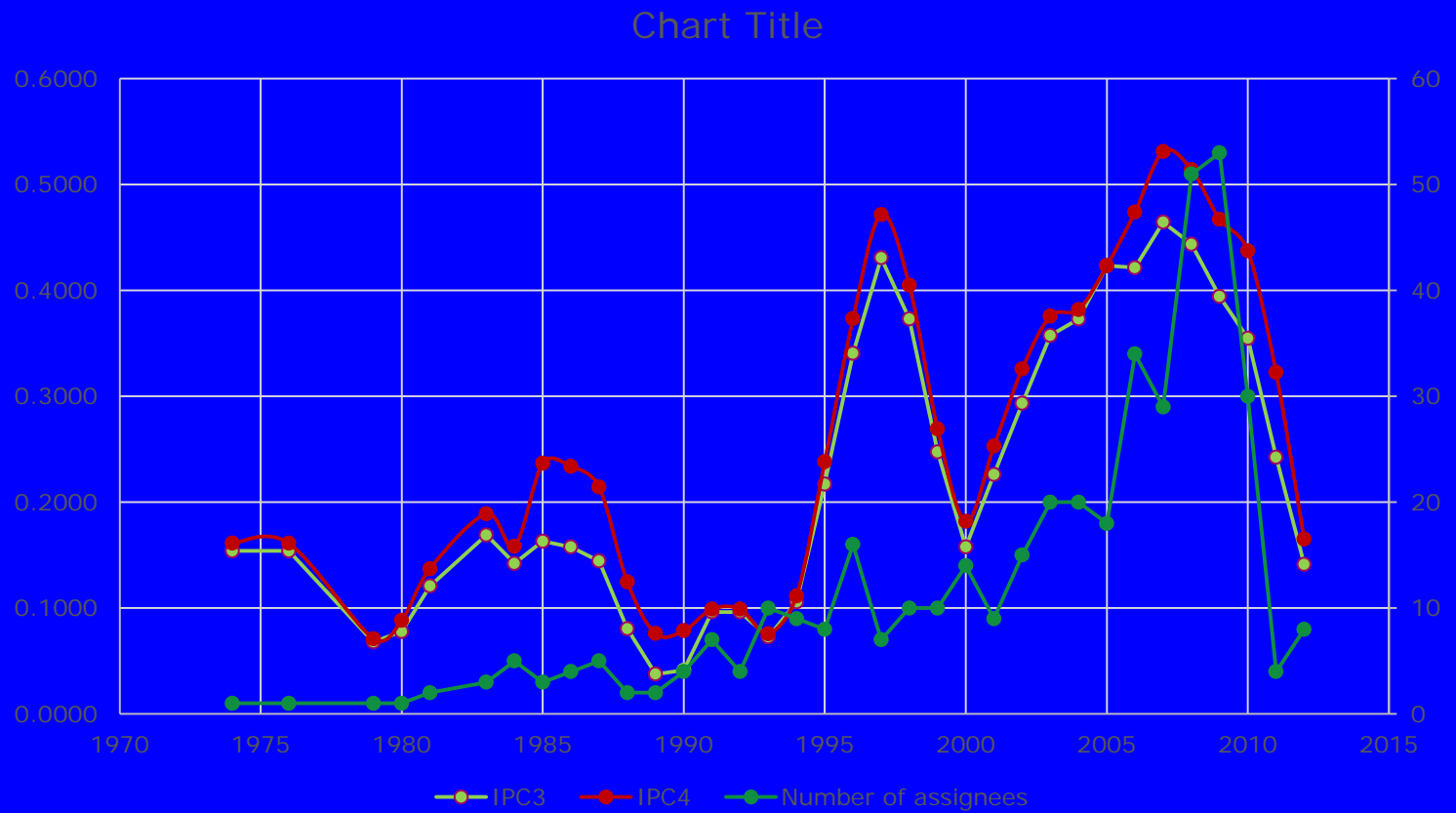
Cities vs Technological development



Economic dynamics

Chart Title





Discussion

- Technological innovations “tumble” through Triple Helix landscapes that change;
- The (three) selection environments are heavily institutionalized
 - **basemaps** (publications, patents, etc.)
- We need flexible tools for moving from one context to another → **interactive overlays**
 -
 -
 -

Conclusion

- We can identify different generations of PV technologies through Stirling diversity;
- The generations are associated with different locations of knowledge production;
- Tentative relationship can be established between maturation stage of technologies and entry/exit patterns of companies (and different locations).

Email:

WWW: Heimeriks.net

Twitter: