Scientometrics and Social Network Analysis for Policy

Edwin Horlings

August 18, 2015

Rathenau Instituut

RATHENAU INSTITUTE

Mission and scope

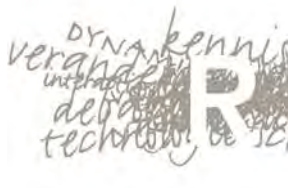
- Formal mission

The Rathenau Instituut promotes the formation of political and public opinion on science and technology. To this end, the institute studies the organization and development of science systems, publishes about social impact of new technologies, and organizes debates on issues and dilemmas in science and technology.

- Independent status

- Two departments
  - Technology Assessment (since 1986)
  - Science System Assessment (since 2005)

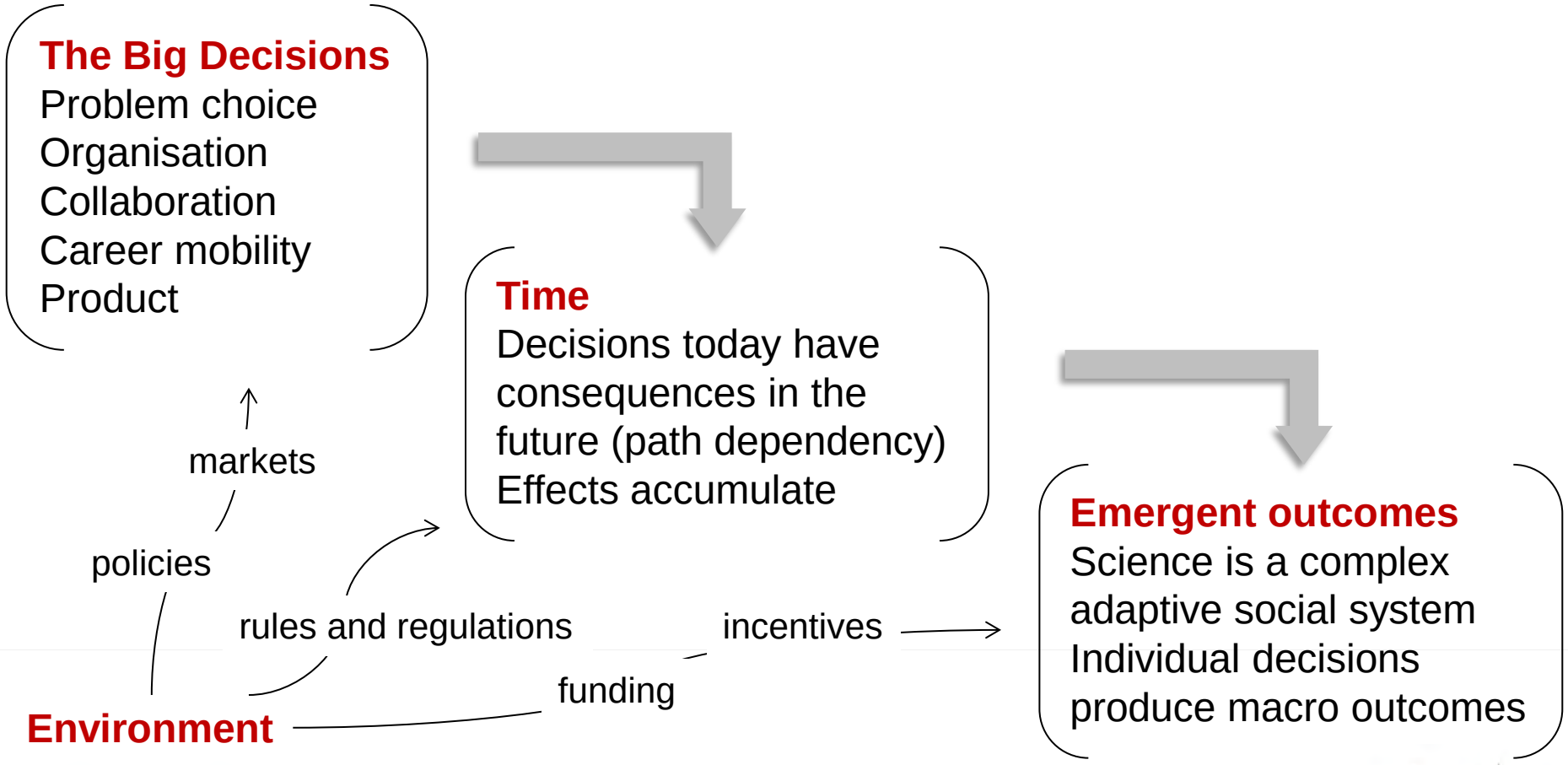
- Three functions
  - information
  - analysis
  - debate



What we do

- Information
  - Facts & Figures (e.g. annual overview of public R&D funding)
  - www.denederlandsewetenschap.nl or www.dutchscience.nl
- Analysis
  - Globalisation of R&D: is R&D leaving the Netherlands?
  - Transdisciplinary knowledge production in climate research
- Debate
  - Blogs (https://rathenaunl.wordpress.com/) and op-eds
  - Workshops and meetings (e.g. about valorisation)

From big decisions to emergent outcomes



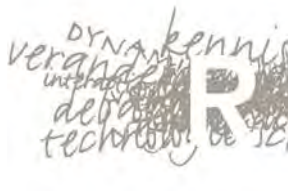
APPLIED SCIENTOMETRICS

Scientometrics

- Essentially, the science of measuring and analysing science.

- The unit of analysis
  - can range from an individual to entire countries
  - can be a topic within a scientific field or the whole field
  - can be an object (e.g. a piece of lab equipment) or it can be abstract (e.g. a potential concept to emerge in 10 years from a specific field).

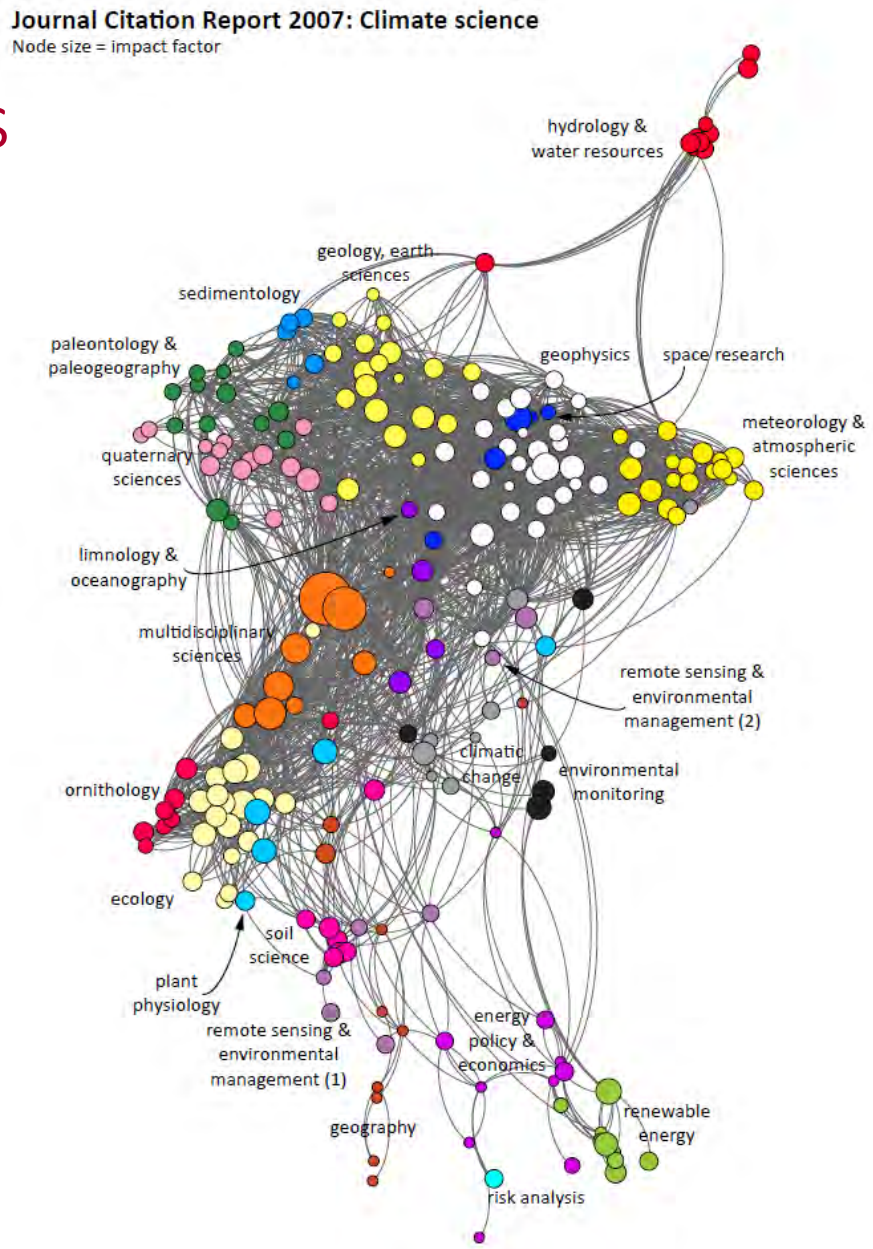
- Core journals
  - Scientometrics
  - JASIST
  - Journal of Informetrics



The magic of scientometrics

- Assumptions:
  - From raw data to answers in the blink of an eye
  - Cool, easy-to-make visuals
  - Quantitative beats qualitative anyday

Assumptions are the mother of all f***-ups



The magic of scientometrics

- Assumptions:
  - From raw data to answers in the blink of an eye
  - Cool, easy-to-make visuals
  - Qu
    an
- Reality:
  - Raw data must be cleaned before use
  - It takes skill and creativity to
    able for

Radboud Univ Nijmegen, Inst Mol & Mat, Nijmegen, Netherlands.
Radboud Univ Nijmegen, Inst Mol & Mat, NL-6500 GL Nijmegen, Netherlands.
Radboud Univ Nijmegen, Inst Mol & Mat, NL-6525 AD Nijmegen, Netherlands.
Radboud Univ Nijmegen, Fac Sci, Inst Mol & Mat, NL-6525 AJ Nijmegen, Netherlands.
Radboud Univ Nijmegen, Inst Mol & Mat, NL-6525 AJ Nijmegen, Netherlands.
Radboud Univ Nijmegen, High Feild Magnet Lab, NL-6525 ED Nijmegen, Netherlands.
Radboud Univ Nijmegen, High Field Magnet Lab, Inst Mol & Mat, NL-6525 ED Nijmegen, Netherlands.
Radboud Univ Nijmegen, High Field Magnet Lab, NL-6525 ED Nijmegen, Netherlands.
Radboud Univ Nijmegen, IMM, NL-6525 ED Nijmegen, Netherlands.
Radboud Univ Nijmegen, Inst Mol & Mat, NL-6525 ED Nijmegen, Netherlands.
Radboud Univ Nijmegen, Inst Mol & Mat, High Field Magnet Lab, NL-6525 ED Nijmegen, Netherlands.
Radboud Univ Nijmegen, Inst Mol Mat, High Field Magnet Lab, NL-6525 ED Nijmegen, Netherlands.
Radboud Univ Nijmegen, Inst Mol Mat, Condensed Matter Theory, NL-6525 ED Nijmegen, Netherlands.

Author disambiguation

- Thousands of researchers with identical names (e.g. Y. Zhang): how to tell the difference?
- Important for evaluation and for research
- Developed an algorithm with 95-100% accuracy
- Now developing software tool with University of Paris Est (ESIEE)

Rathenau Instituut

Scientometrics (2012) 91:435–449
DOI 10.1007/s11192-011-0589-1

Author disambiguation using multi-aspect similarity indicators

Thomas Gurney · Edwin Horlings · Peter van den Besselaar

Received: 5 December 2011 / Published online: 30 December 2011

Abstract Key to accurate bibliometric analyses is the ability to correctly link individuals to their corpus of work, with an optimal balance between precision and recall. We have developed an algorithm that does this disambiguation task with a very high recall and precision. The method addresses the issues of discarded records due to null data fields and their resultant effect on recall, precision and F-measure results. We have implemented a dynamic approach to similarity calculations based on all available data fields. We have also included differences in author contribution and age difference between publications, both of which have meaningful effects on overall similarity measurements, resulting in significantly higher recall and precision of returned records. The results are presented from a test dataset of heterogeneous catalysis publications. Results demonstrate significantly high average F-measure scores and substantial improvements on previous and stand-alone techniques.

Keywords Author disambiguation · Precision and recall · Homonyms · Community detection · Data discarding

Introduction

The use of scientometrics has become increasingly prevalent in many forms of scientific analysis and policy making. Key to good bibliometric analysis is the ability to correctly

An extended version of a paper presented at the 13th International Conference on Scientometrics and Informetrics, Durban (South Africa), 4–7 July 2011 (T. Gurney, E. Horlings, P. van den Besselaar, 2011).

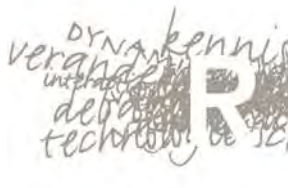
T. Gurney (✉) · E. Horlings
Rathenau Institute, Anna van Saksenlaan 51, 2593 HW The Hague, The Netherlands
e-mail: t.gurney@rathenau.nl

Applied Scientometrics

- Scientometrics as a tool
- Not about developing methods but about using methods to answer substantive questions
- It always starts with The Interesting Question
  - provides focus
  - prevents drowning in data
  - helps select the most efficient approach

Examples of interesting questions

- What is the profile of a successful researcher in terms of publications, author position, citations, grants, size of coauthor networks, international collaborations?
- What do the co-author networks of group leaders look like?
- How has scientific research on breast cancer developed and what is the role of one specific university in it?
- How strong is the Netherlands in a specific type of research (e.g. solar energy)?
- Has policy X affected behaviour Y?
- etcetera



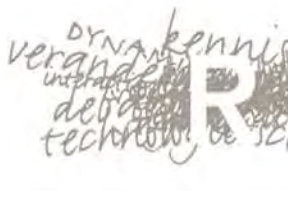
SAINT TOOLKIT

Three main tools

- Science Assessment Integrated Network Toolkit

- Main components
  - ISI Parser: convert raw Web of Science data into a relational database
  - Word Splitter: cuts full text into words, eliminating stop words, and shortening words to their stem using different algorithms
  - Network Tools: identify clusters in network using one of the best clustering algorithms (Blondel and Infomap)

EMERGENT OUTCOMES



Effects of priority setting

- Ambition of many policy makers is to guide the direction of research in specific directions
- Examined effects of **Focus and Mass**, a major policy principle in the Netherlands aimed at priority setting
- Expect effects on growth and specialisation of scientific output

Focus and Mass

- Eight priority areas for science and innovation
  - ICT
  - Chemistry and Chemical Technology
  - High-tech Systems
  - Advanced Materials
  - Nanotechnology
  - Water
  - Genomics
  - Food & Flowers

Data and methods

- Web of Science
  - Five citation databases (SCI, SSCI, A&HCI, CPCI, CPCI-SS&H)
  - Search for all publications containing an address from country X and published in year Y (e.g. cu=Netherlands and py=2008)
  - All types of publications (articles, conference proceedings, letters, etc.)
  - Results analysed for distribution among subject areas
  - Social sciences and humanities excluded from longitudinal data

- Double counting
  - Publications involving international collaboration
  - Subject areas overlap
  - Removed where possible

Four hypotheses

If the policy of Focus and Mass has had an effect:

1. The overall degree of concentration in the scientific output of the Netherlands will have increased.
2. Scientific areas targeted by the policy will have grown relative to other scientific areas.
3. In comparison with other countries, the specialisation pattern of Dutch science will have shifted towards F&M-areas.
4. The share of Dutch output in the world output of F&M-areas will have increased.

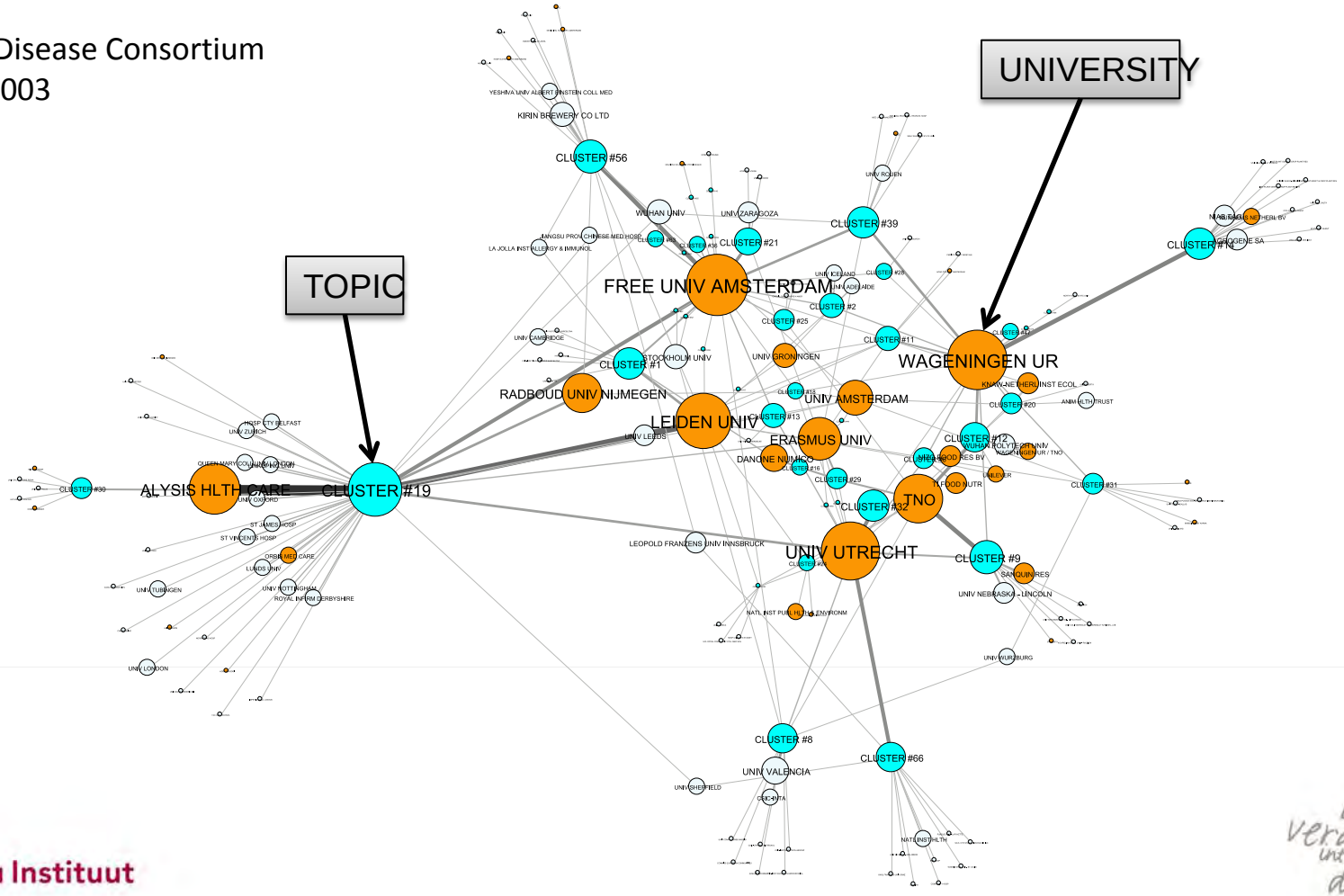
REJECTED

Hypothesis 2: Relative growth

	RELATIVELY GROWING	RELATIVELY STABLE	RELATIVELY DECLINING
LARGE (>5%)			
MEDIUM	ICT	Genomics High-tech systems	Chemistry & chemical technology Advanced materials Food & Flowers
SMALL <1.3%)	Nanotechnology	Water	

How Dutch universities work on scientific topics

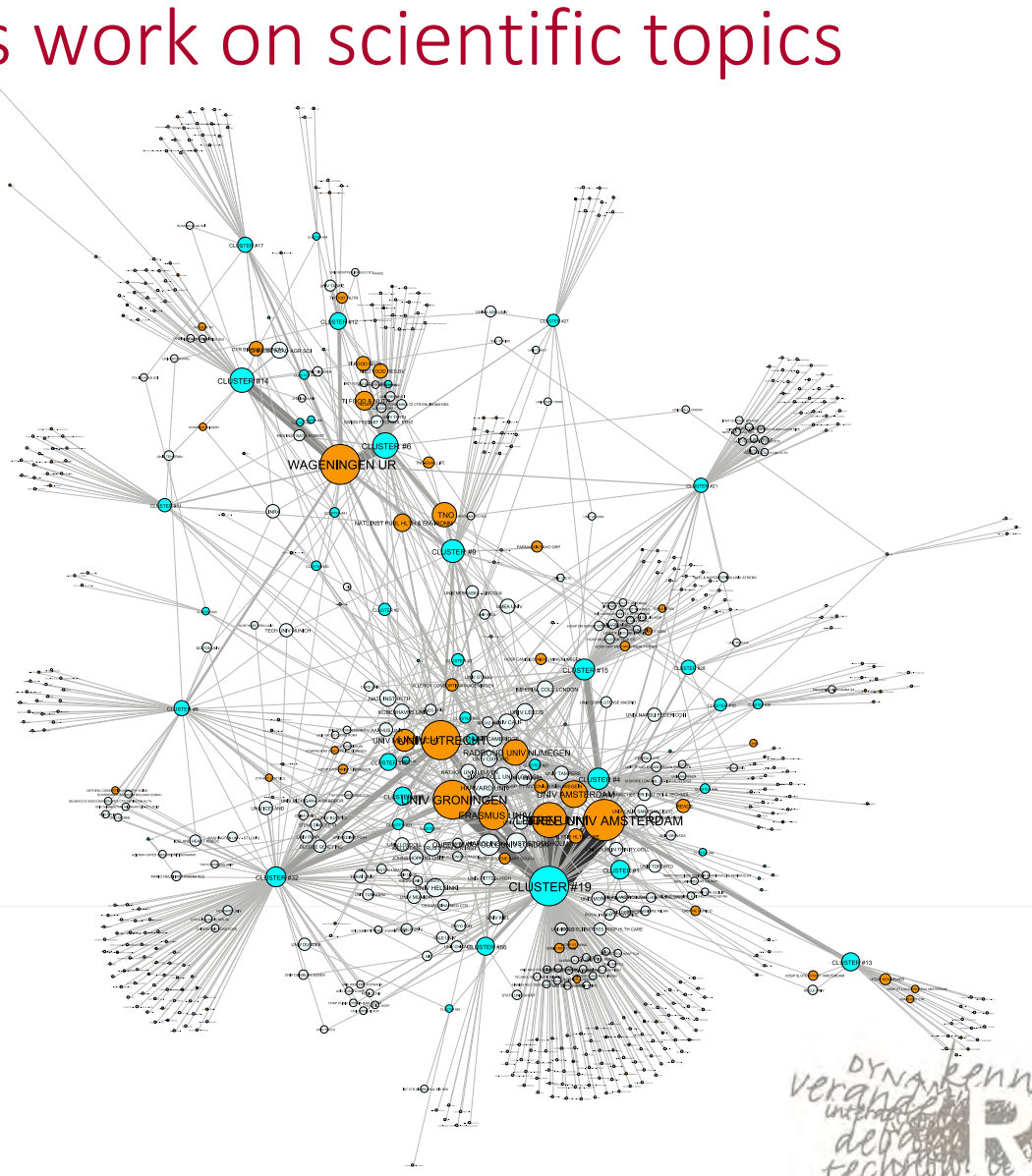
Celiac Disease Consortium
2000-2003



How Dutch universities work on scientific topics

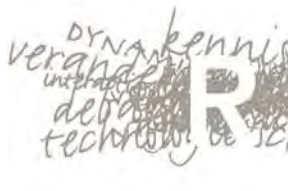
Celiac Disease Consortium
2007-2010

- Denser network
- More institutions involved
- More coherent: more universities work on the same small set of topics



No conclusion without statistical analysis

- A visualisation can be extremely informative...
- ....but be careful of the Rorschach effect!
- You must confirm what you think you see:
  - statistical analysis
  - interviews
  - other methods



INDIVIDUAL SEARCH STRATEGIES

Mapping individual portfolios

- Understanding how scientists specialise
- Measuring the effects of policies at individual level: adaptation
- Identifying effects related to life cycles, generations, career changes
- Interpreting performance in evaluation

Scientometrics
DOI 10.1007/s11192-012-0789-3

Search strategies along the academic lifecycle

Edwin Horlings · Thomas Gurney

Received: 29 March 2012

Abstract Understanding how individual scientists build a personal portfolio of research is key to understanding outcomes on the level of scientific fields, institutions, and systems. We lack the scientometric and statistical instruments to examine the development over time of the involvement of researchers in different problem areas. In this paper we present a scientometric method to map, measure, and compare the entire corpus of individual scientists. We use this method to analyse the search strategies of 43 condensed matter physicists along their academic lifecycle. We formulate six propositions that summarise our theoretical expectations and are empirically testable: (1) a scientist's work consists of multiple finite research trails; (2) a scientist will work in several parallel research trails; (3) a scientist's role in research trail selection changes along the lifecycle; (4) a scientist's portfolio will converge before it diverges; (5) the rise and fall of research trails is associated with career changes; and (6) the rise and fall of research trails is associated with the potential for reputational gain. Four propositions are confirmed, the fifth is rejected, and the sixth could not be confirmed or rejected. In combination, the results of the four confirmed propositions reveal specific search strategies along the academic lifecycle. In the PhD phase scientists work in one problem area that is often unconnected to the later portfolio. The postdoctoral phase is where scientists diversify their portfolio and their social network, entering various problem areas and abandoning low-yielding ones. A professor has a much more stable portfolio, leading the work of PhDs and postdoctoral researchers. We present an agenda for future research and discuss theoretical and policy implications.

Keywords Mapping science · Academic careers · Lifecycle · Agenda setting · Problem choice · Complex adaptive system

Reward system of science

- Priority and recognition
  - Merton (1957): search for priority and recognition
  - Elaborated in sociology of science: Cole 1970; Cole& Cole 1967; Hagstrom 1965, 1974; Reskin 1977
  - New economics of science: Dasgupta & David 1994; Stewart 1995; Stephan 1996
  - Empirically tested by Hagstrom 1974, Zuckerman & Cole 1994

- Problem choice is the instrument of competition between a scientist and his peers

Community

- Problem choice is driven by the possibility of gaining reputation
  - How complex is the problem? potential reputational gains vs. risk
  - How many other scientists are working on the problem? crowdedness

- Community of peers is required
  - To be able to compete a scientist has to have competitors with whom he has to reach a level of consensus on the basic premises of the problem area (Whitley, 1974, 2000; Lave & Wenger 1991)
  - Community structures the search for new problems
  - Access to resources (e.g. through collaboration)

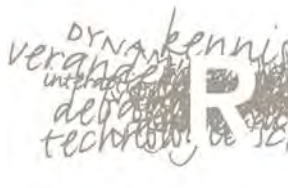
Search strategies

- **Search**: the process by which an individual scientist identifies, enters, develops, and exits a problem area and its associated community of peers.

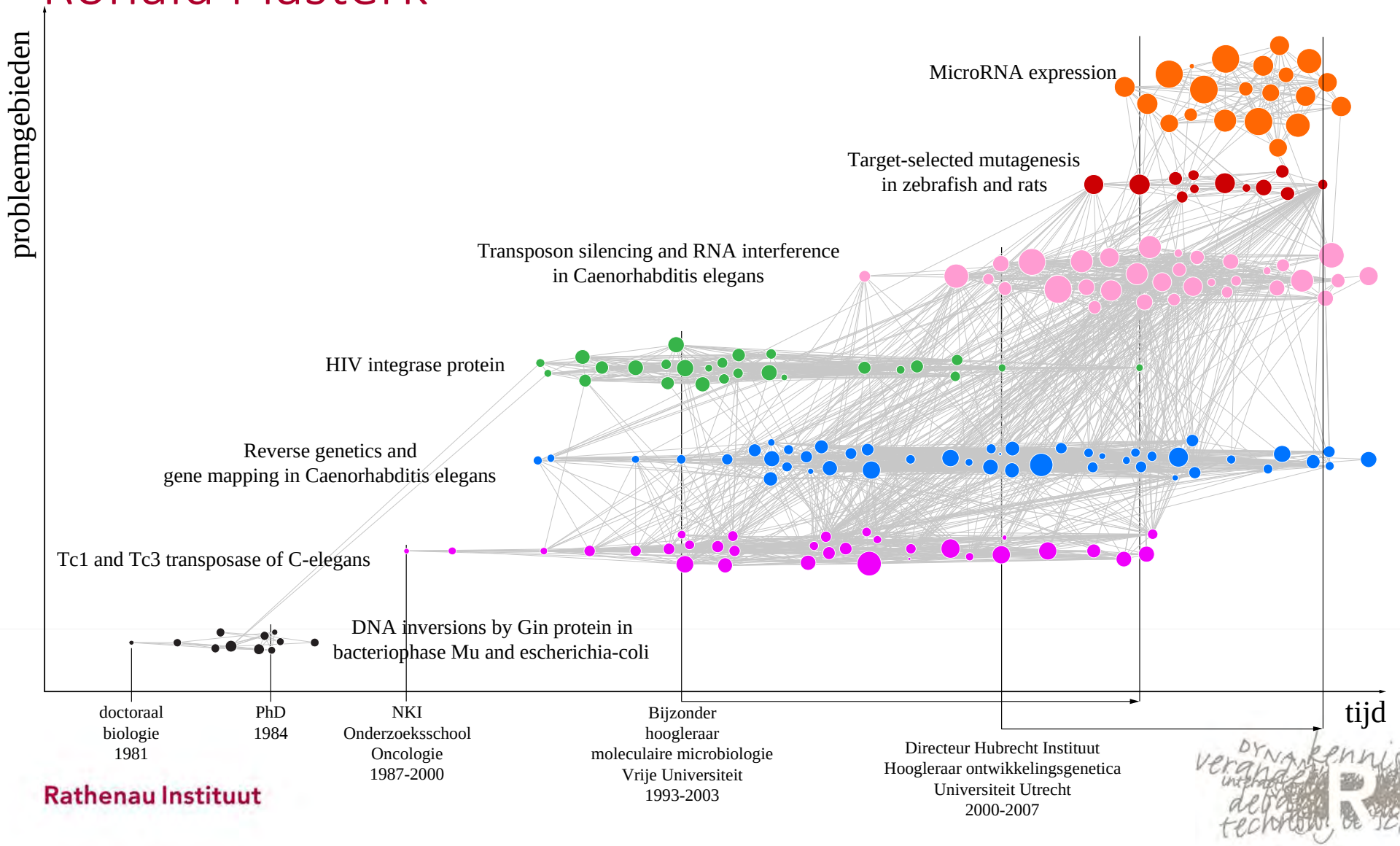
- **Strategy**: the scientist's strategic positioning in a competitive environment presumes a degree of planning, coherence and consistency to problem choice over time.

Six propositions derived from theory

1. A scientist's work consists of multiple finite research trails
2. A scientist will work in several parallel research trails
3. A scientist's role in research trail selection changes along the lifecycle
4. The start and end of research trails is associated with career changes
5. The start and end of research trails is associated with the potential for reputational gain
6. A scientist's portfolio will converge before it diverges



Ronald Plasterk

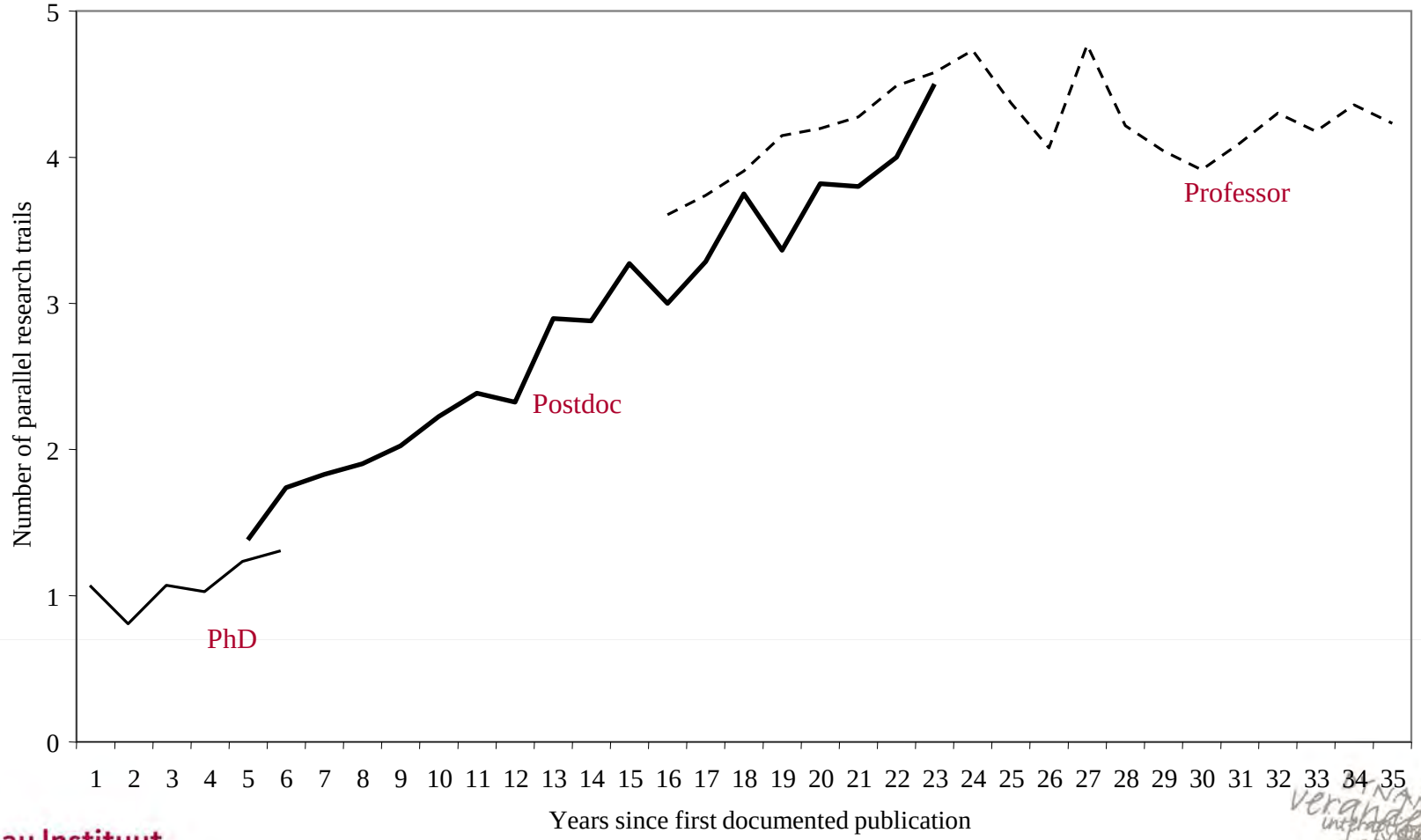


Rathenau Instituut

Six propositions

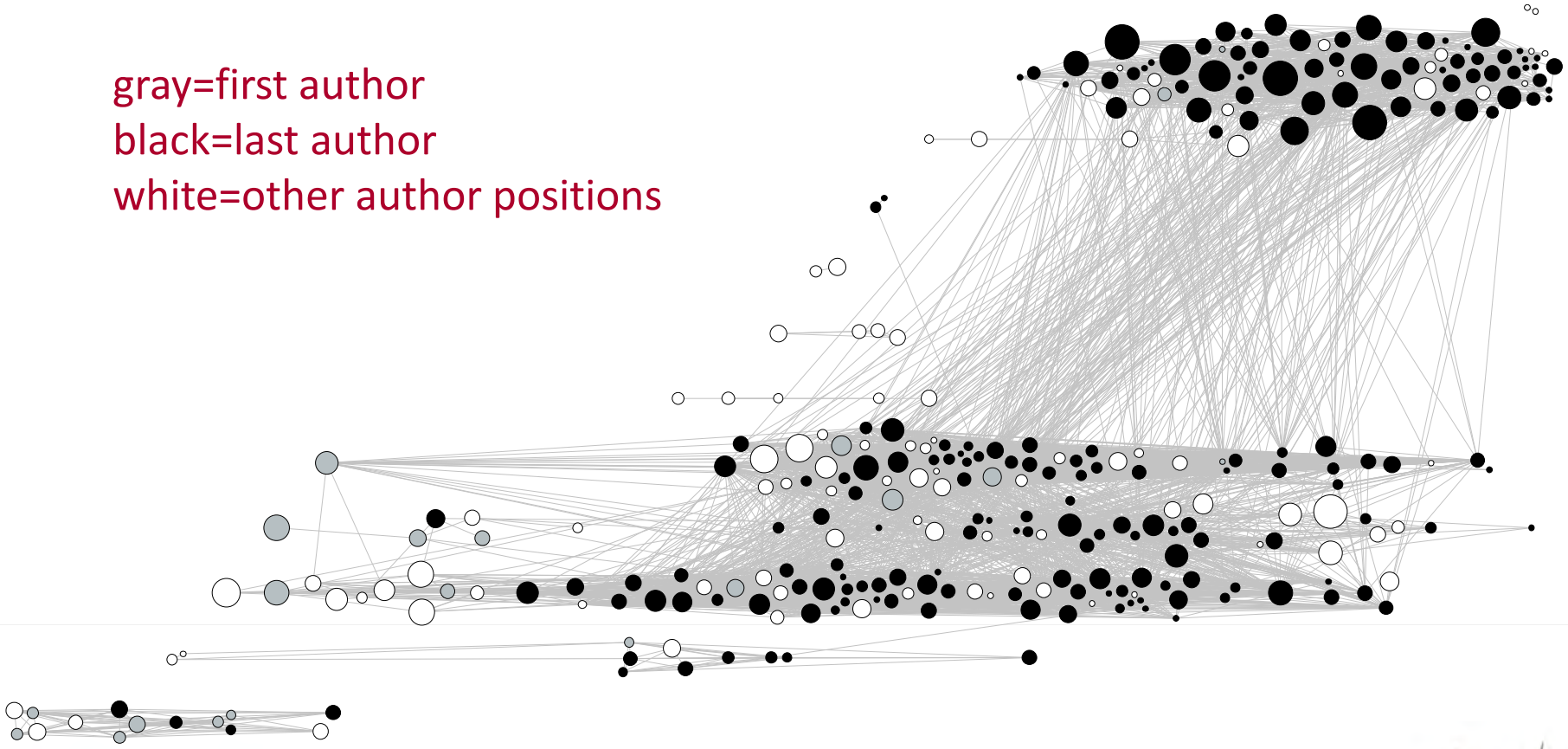
1. A scientist's work consists of multiple finite research trails ✓
2. A scientist will work in several parallel research trails ✓
3. A scientist's role in research trail selection changes along the lifecycle ✓
4. The start and end of research trails is associated with career changes ✗
5. The start and end of research trails is associated with the potential for reputational gain ✓
6. A scientist's portfolio will converge before it diverges ✗

Proposition 2: Parallel research trails



Proposition 3: Role in problem selection changes along the lifecycle

gray=first author
black=last author
white=other author positions

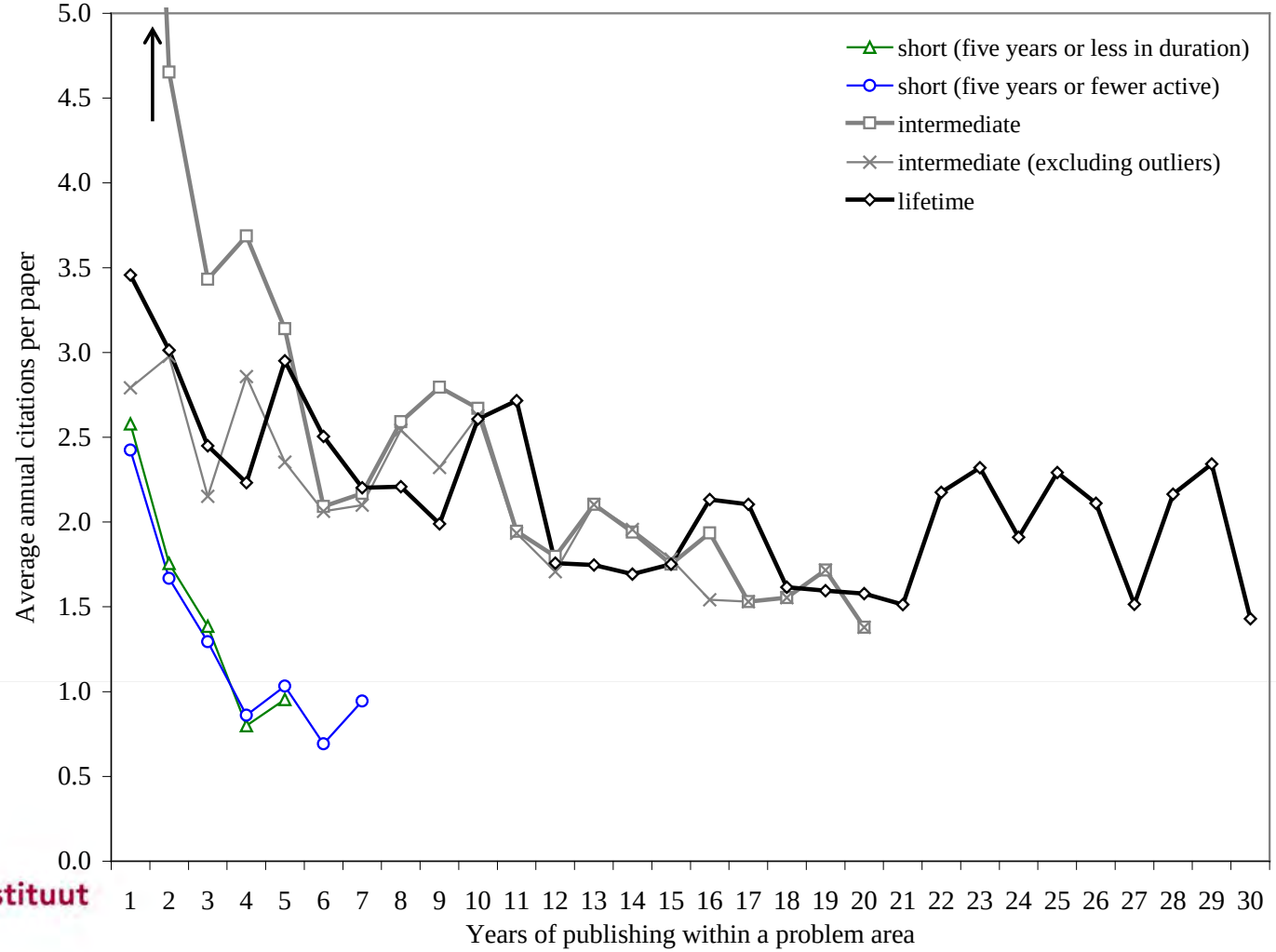


Proposition 3: Role in problem selection changes along the lifecycle

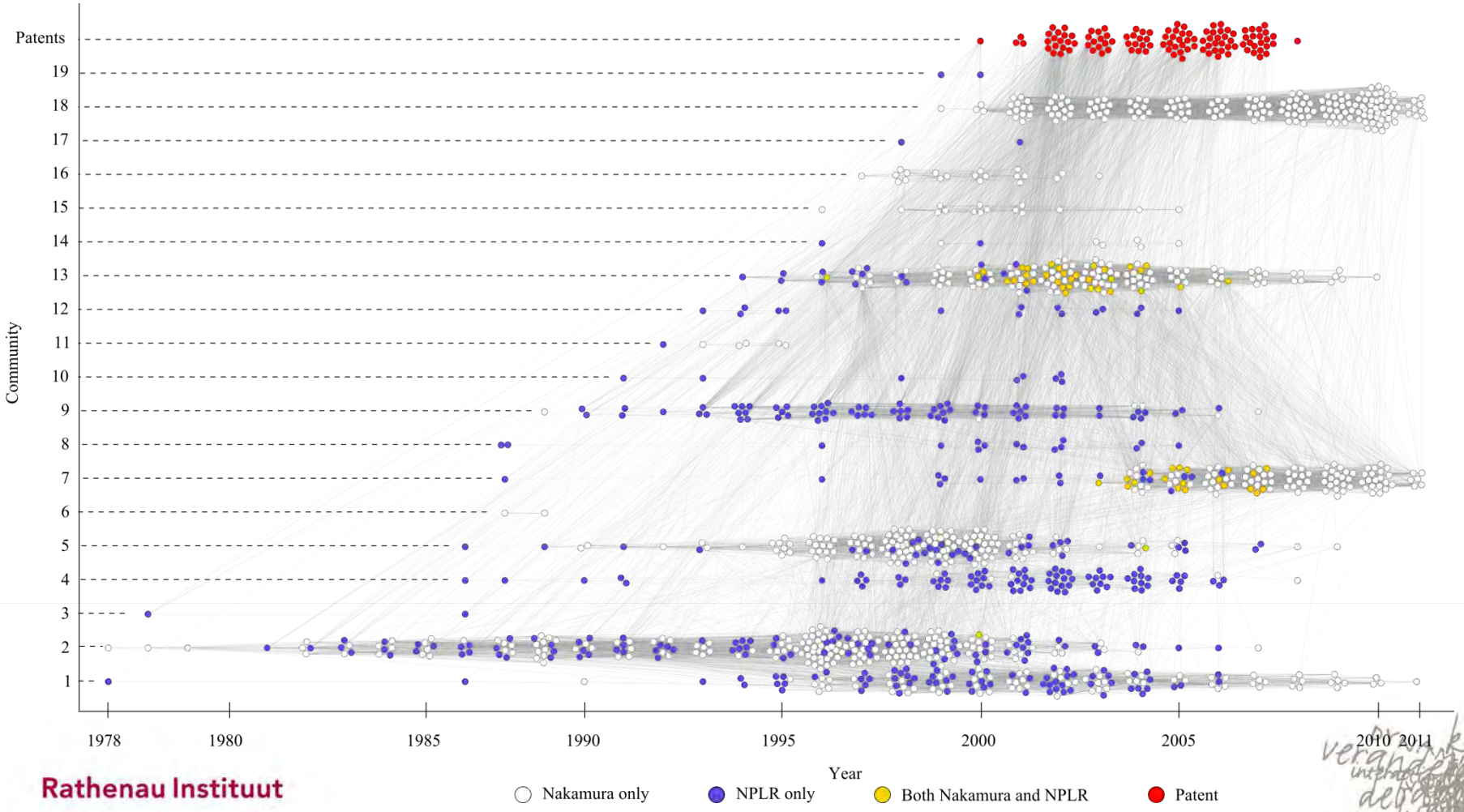
Table 1. Percentage of publications written in first, other or last author positions in three phases of the academic lifecycle

author position	PhD	Postdoc	Professor
first	56.4	22.8	19.7
other	28.9	39.6	29.4
last	14.8	37.6	50.8
total	100	100	100

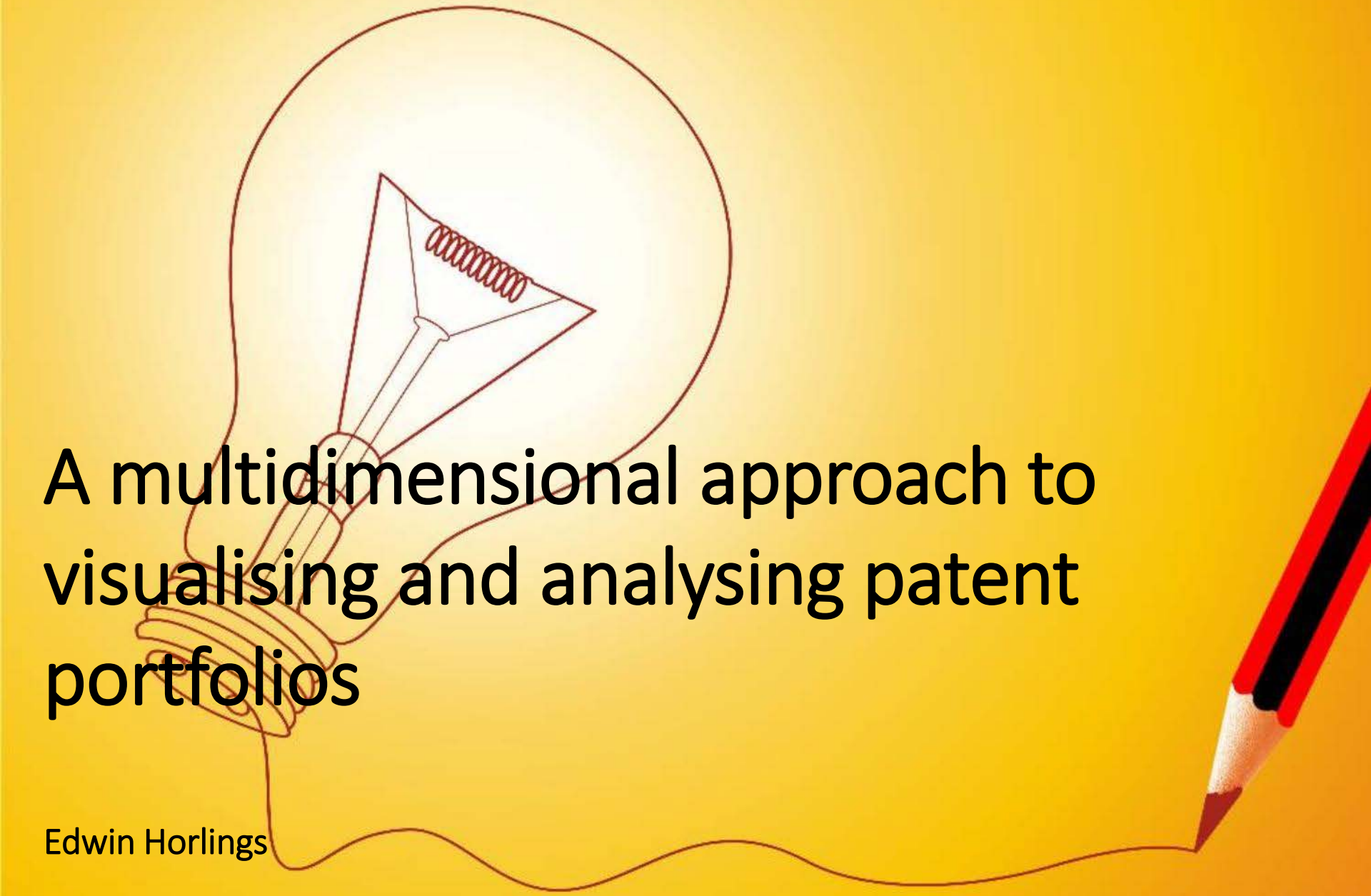
Proposition 5: Start and end of trails associated with reputational gain



Linking patents to publications



Rathenau Instituut



A multidimensional approach to visualising and analysing patent portfolios

Edwin Horlings

Global TechMining Conference, Leiden, 2 September 2014

A data infrastructure for PATSTAT

- **QUERY SET 1**: Pre-construct aggregated information for all of PATSTAT
  - per application, INPADOC family, and single priority family
  - basic information on application and publication
  - aggregate tables for citations between applications and families
- Why aggregate?
  - time saving
  - normalise globally rather than locally
- **QUERY SET 2**: for a specific set of patents collect information from aggregate tables, calculate local information (e.g. topics within the dataset), and produce output for statistical analysis and visualisation

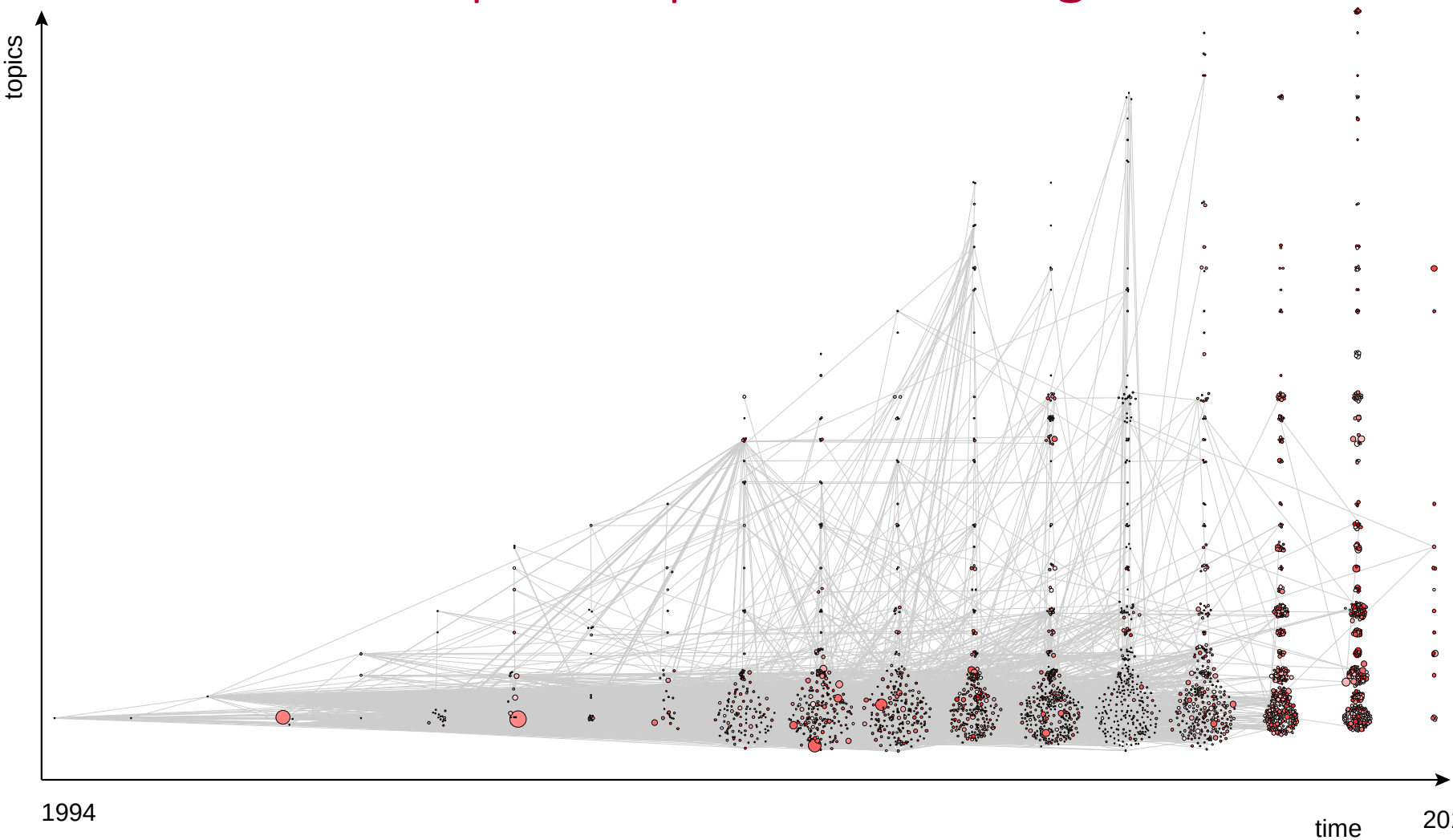
Five dimensions for analysis

Dimension	Examples
Time	date of application date of publications distance in time between application, publication, citation, and granting
Citation	forward and backward to patents and non-patent literature references
Topics	clusters of highly similar patents as measured, for example, by cooccurrence of IPC codes or words
Diversity	variety of topics distribution of applications among topics
Quality	economic value technical impact nature of the invention

Indicators for patent quality

Indicator	Interpretation	Reference
size	larger families are more valuable	Lerner (1994)
scope	broad patents are more valuable	Lanjouw et al. (1998)
backward citations	patents with more backward citations have higher value and are more incremental	Trajtenberg M. (1990) Lanjouw and Schankerman, (2001)
forward citations (within 5 years)	technological importance and economic value of inventions	Trajtenberg M. (1990)
number and share of NPLRs	distance to science, technical quality	Callaert et al. (2006) Branstetter (2005)
claims and adjusted claims	number of claims reflects expected patent value and technological breadth	Tong and Davidson (1994) Squicciarini et al. (2013)
grant		
grant lag	shorter lag indicates higher value	Czarnitzki, Hussinger & Schneider (2009)
generality	range of later generations of inventions that have benefitted from a patent	Trajtenberg, Henderson & Jaffe (1997)
originality	indicates diversity of knowledge sources	Trajtenberg, Henderson & Jaffe (1997)
radicalness	radical versus incremental	Shane (2001)
technology cycle time	pace of technological progress	Kayal & Waters (1999)

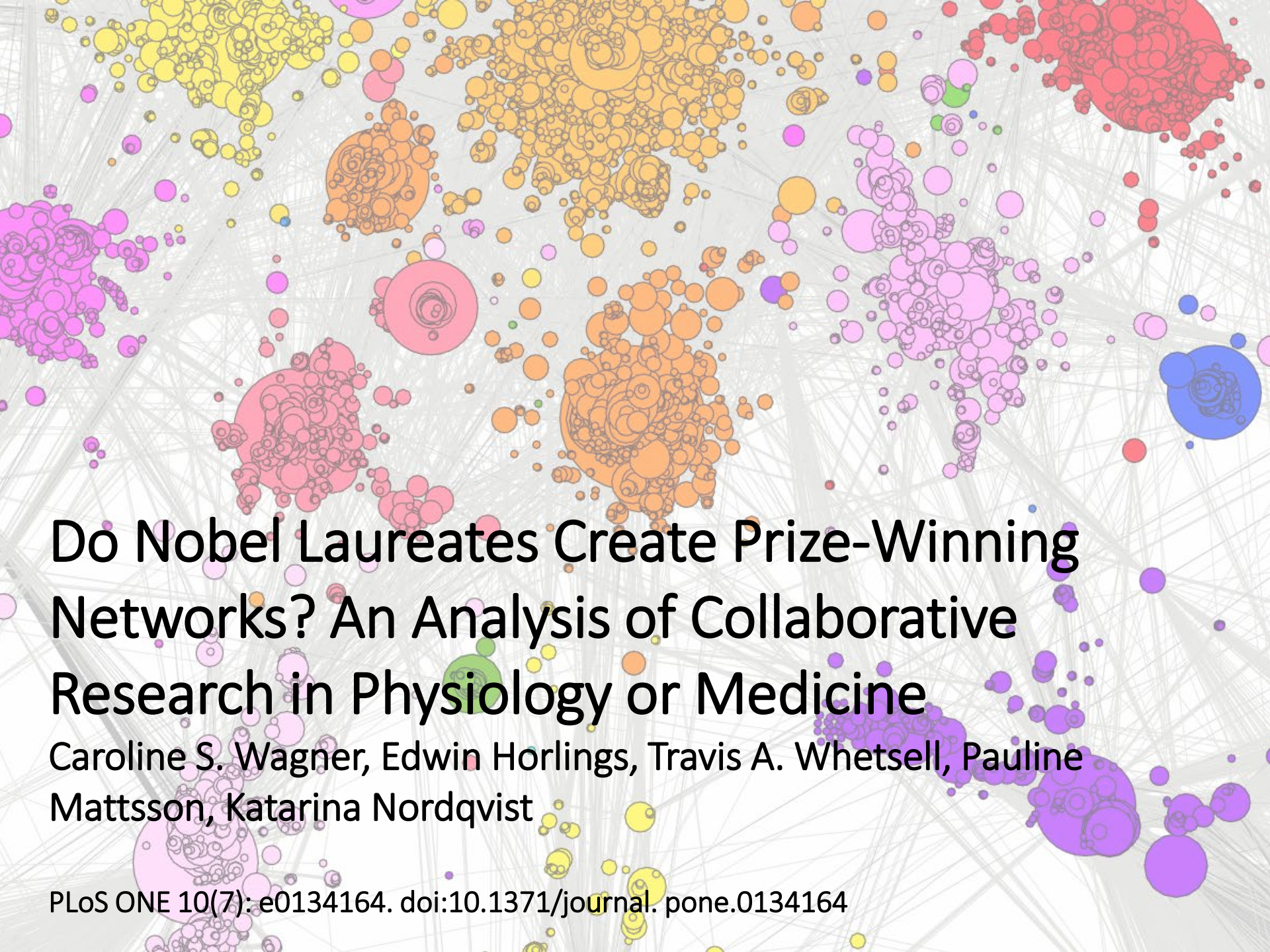
A firm's lifetime patent portfolio: Google



The quality of academic patents compared

	N (single priority families)	share of NPLRs = 0	mean share of NPLRs	standard deviation	median share of NPLRs
general universities	853	39.6%	.410	.392	.375
technical universities	606	60.2%	.193	.292	.000
non-university PROs	2,343	55.9%	.215	.305	.000
top-100 firms	50,367	86.1%	.042	.138	.000
other firms	29,891	81.9%	.070	.198	.000

Note: Estimates for 1990-2010. University-owned patents only.



Do Nobel Laureates Create Prize-Winning Networks? An Analysis of Collaborative Research in Physiology or Medicine

Caroline S. Wagner, Edwin Horlings, Travis A. Whetsell, Pauline Mattsson, Katarina Nordqvist

PLoS ONE 10(7): e0134164. doi:10.1371/journal. pone.0134164

Question

- Is the way in which Nobel laureates collaborate substantially different or are the embedded in different networks than equally excellent non-laureates?

- 68 Laureates compared to 68 Non-Laureates “of Nobel-class”

- We applied scientometrics methods and techniques, including SAINT: author disambiguation, visualisation with Gephi, similarity measures, SNA metrics

Sample

- 68 of 101 Nobel laureates in Physiology or Medicine
  - 33 difficult or impossible to distinguish because of their name
  - period 1969-2011

- 68 contemporaries with an H-index of similar magnitude

- Compared as **two groups** rather than as **matched pairs**

Data and methods

- Downloaded all their publications from the Web of Science
  - Laureates c. 15.000
  - Non-Laureates c. 21.000
- Formal check if they worked in the same domain
- Harmonised all coauthors using scientometric methods

  Gurney, T., Horlings, E., & Van Den Besselaar, P. (2011). Author disambiguation using multi-aspect similarity indicators. *Scientometrics*, *91*(2), 435-449.

- Visualisation in Gephi
- Social Network Analysis

Results: publications and coauthorship

	Pubs	Unique authors	Sole author	Authors per pub	Citations per pub
Laureates	222	344	10%	5.0	45
Non-Laureates	306	455	5%	5.1	32
Laureates --before the Prize			11%	4.4	
--in the 2 years after the Prize			17%	6.2	
--after the Prize			9%	6.4	

Results: network

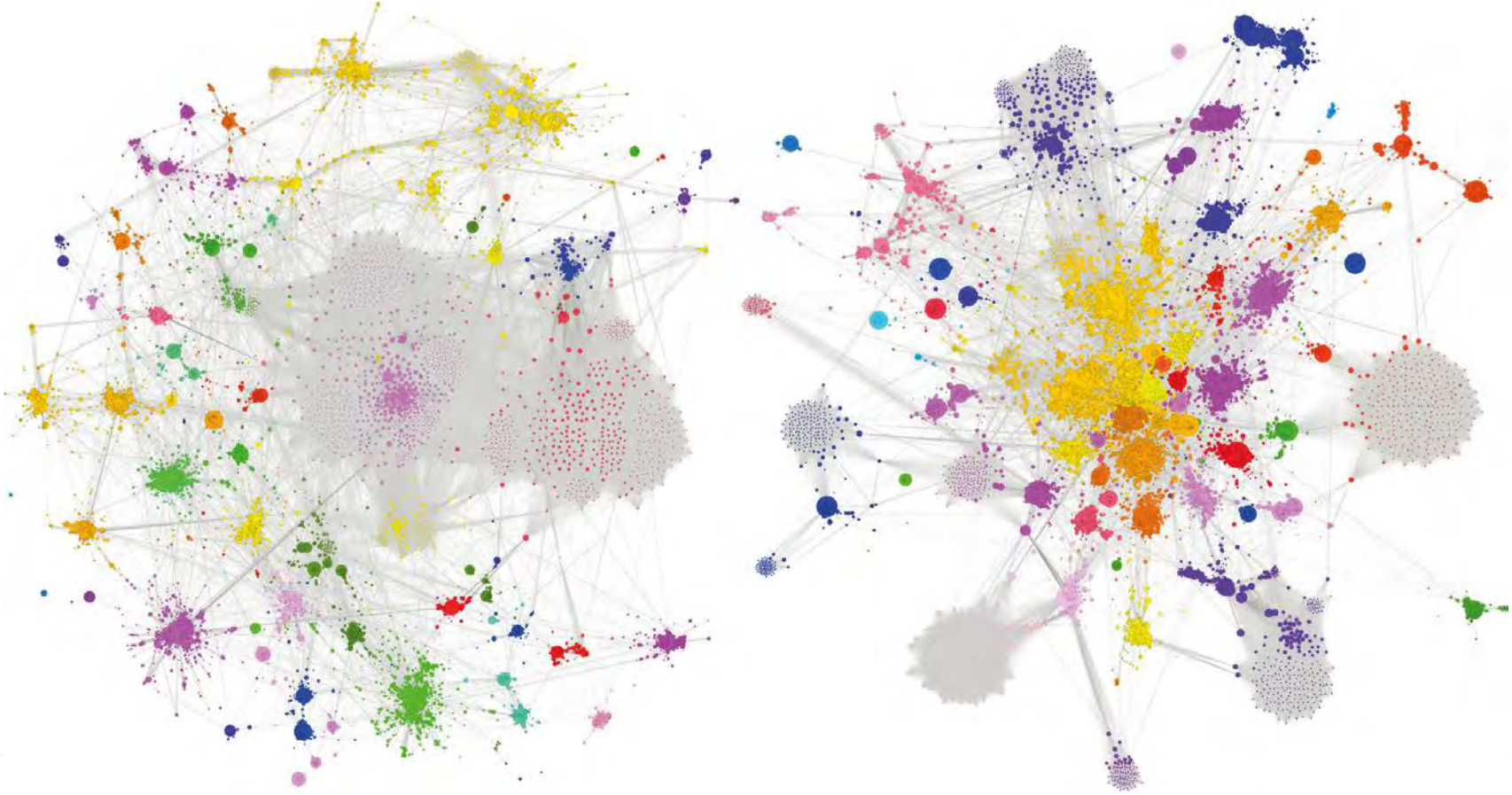


Fig 1. Nobel Laureate Coauthor Network (left) and Non-Laureate Coauthor Network (right). Network graphs were produced in Gephi, using the Force Atlas 2 layout, a variant of the Fruchterman-Reingold algorithm with stronger clustering. Linlog mode was used with scaling set to 0.2 and gravity to 1 0. After applying the Force Atlas 2 layout, Noverlap was used to prevent nodes from overlapping [27]. Node coloring is based on modularity class, identified using the Blondel et al. [18] algorithm.

Results: network

	Average degree	Modularity	Density	Average clustering coefficient	Average path length
Entire network					
Laureates	32.940	.795	.002	.870	3.996
Non-Laureates	23.115	.914	.001	.863	4.057
Only L and NL					
Laureates	3.912	.656	.058	.459	2.962
Non-Laureates	1.118	.828	.017	.441	3.374

Results: network

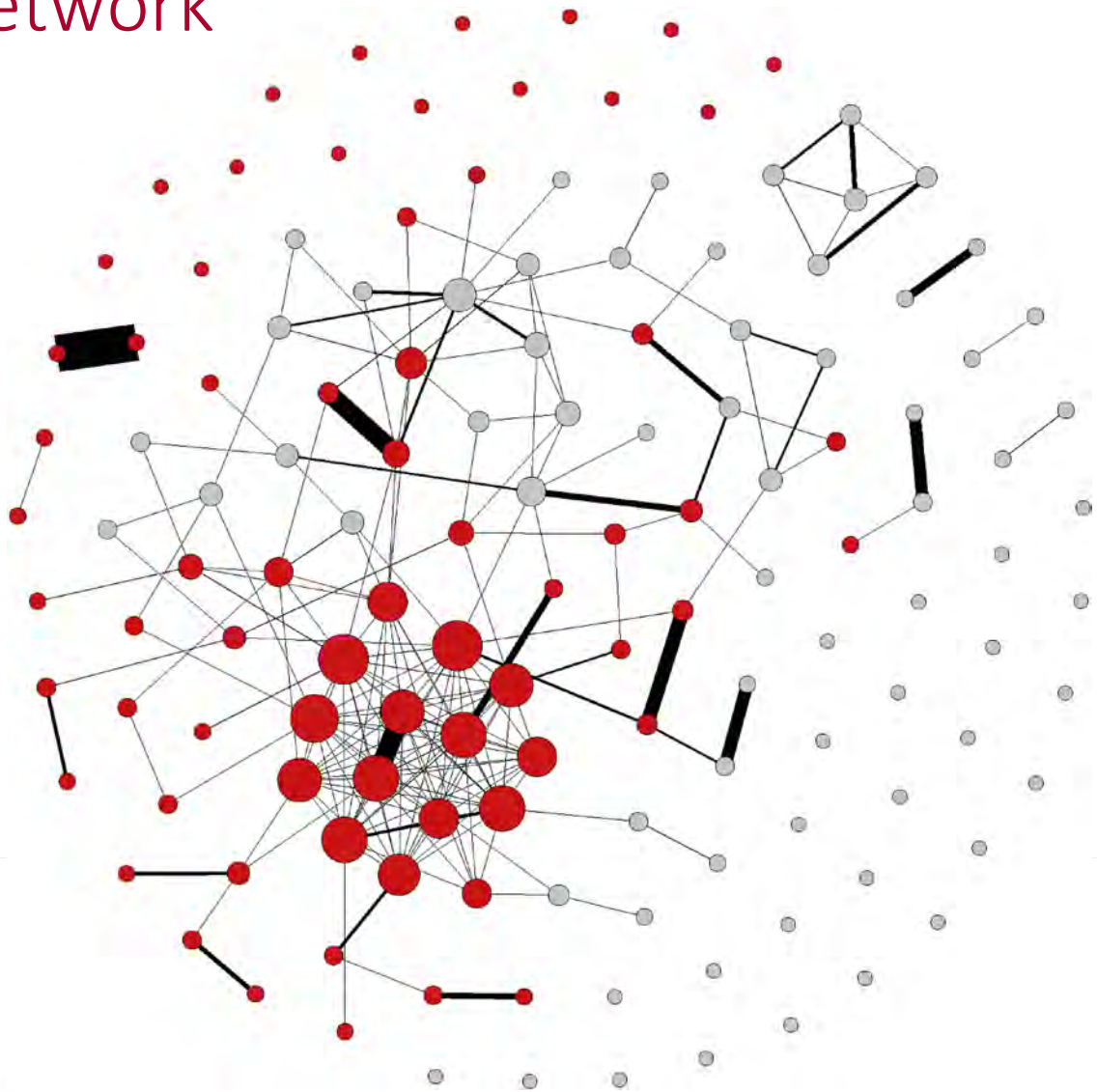


Fig 3. Coauthor Relations among Nobel Laureates (scarlet) and non-Laureates (grey). Fruchterman-Reingold layout was used. Node size is based on degree. Laureates are scarlet, and non-Laureates are grey. Edge size represents weight.

doi:10.1371/journal.pone.0134164.g003

Interpretation (1)

- Nobel laureates are **more selective**:
  - Fewer publications, but of higher quality
  - Publishing alone gives more visibility (reputation)
  - Select the best students and coauthors
    - preferential attachment as effect: Matthew Effect
    - preferential attachment as proces: preferences of actors looking to collaborate

Interpretation (2)

- Nobel laureates extract **more benefits from their network**
  - Better access to information
  - Use this information to distinguish themselves in competition with peers: their network is *"better structured for originality"*
  - Network of Nobel laureates shows their strategic behaviour:

    "highly creative researchers may seek one another out, perhaps not so much for cooperation (although that clearly occurs), but to stay abreast of what others are doing to ensure an advantageous position for originality relative to other high achievers. Investing social capital in network relationships (while 'costing' more in terms of time and attention than operating in a tight community) provides the pay-off of knowing what others are researching."

From 11.700 questions by citizens to a national research agenda for the Netherlands: a scientometric assessment

2025 - Vision for Science

Three main goals

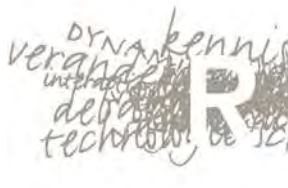
1. Science of worldwide significance
2. Maximum societal impact
3. Breeding ground for talent

Creation of a National Science Agenda

From 11700 questions to an agenda

- Too many to organise by hand
  - Wide variation in quality but bad questions are not less relevant
  - Some questions and topics appear more often which does not mean that they are more important
  - He who selects by hand, dominates the outcome

- A technical solutions may help
  - Look for structure (clustering) in the 11700 questions
  - Heuristic aid, not the solution for selecting and prioritising topics



Comparing questions

1. How similar is each pair of questions?
2. Which questions are more similar to each other than to all other questions?

- Question 1: find a measure for the degree of similarity

- Question 2: find a way to identify clusters in the similarity network

Similarity

- Cooccurrence of words in questions and explanations
  - Cleaned the words to harmonise quality and reduce word variation:
    - Removed stopwords (e.g. adjectives, prepositions)
    - Removed generic words with low distinctiveness (“relative”, “assume”, “people”, etc.)
    - Corrected spelling errors
    - Standardised spelling variants (in Dutch: kwantum – quantum) and conjugations (did – do – done; firm – firms)
    - From 55,935 unique words (without correction, without stopwords) to 36,857 unique words after standardisation
  - Every word was counted once (bit vectors)
  - Similarity in word cooccurrence: Jaccard coefficient

Practical example

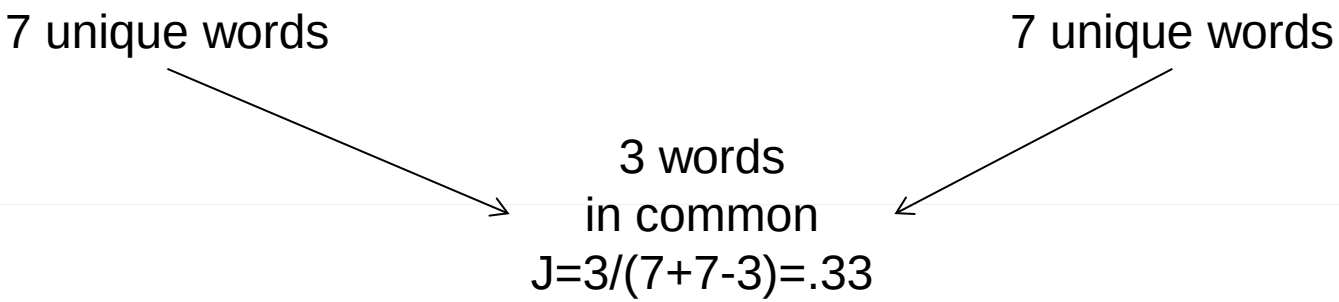
Is de werkwijze die kunstenaars en ontwerpers hanteren van waarde voor innovatie in andere sectoren?

Hoe meten we de bijdrage van creativiteit aan innovatie? Is creativiteit onder kunstenaars en ontwerpers iets anders dan in andere beroepen, oftewel: wat is creativiteit?

Practical example

Is de **werkwijze** die **kunstenaars** en **ontwerpers hanteren** van **waarde** voor **innovatie** in andere **sectoren**?

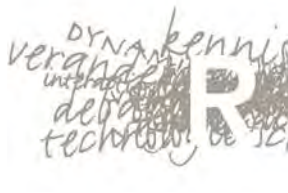
Hoe **meten** we de **bijdrage** van **creativiteit** aan **innovatie**? Is **creativiteit** onder **kunstenaars** en **ontwerpers** iets anders dan in andere **beroepen**, oftewel: wat is **creativiteit**?



Clusters

- 281 clusters
  - Largest cluster: 280 questions (astronomy)
  - Smallest clusters: isolated questions not similar to other questions
  - Three questions with all words removed (“What is a question?”, “Why is there something and not nothing?”, “find out what is totally useless”)
  - Identical questions (same question and explanation submitted several times)

- Interpretation of cluster content
  - spurious results but often homogenous
  - homogenous subject but different questions
  - Examine each cluster individually



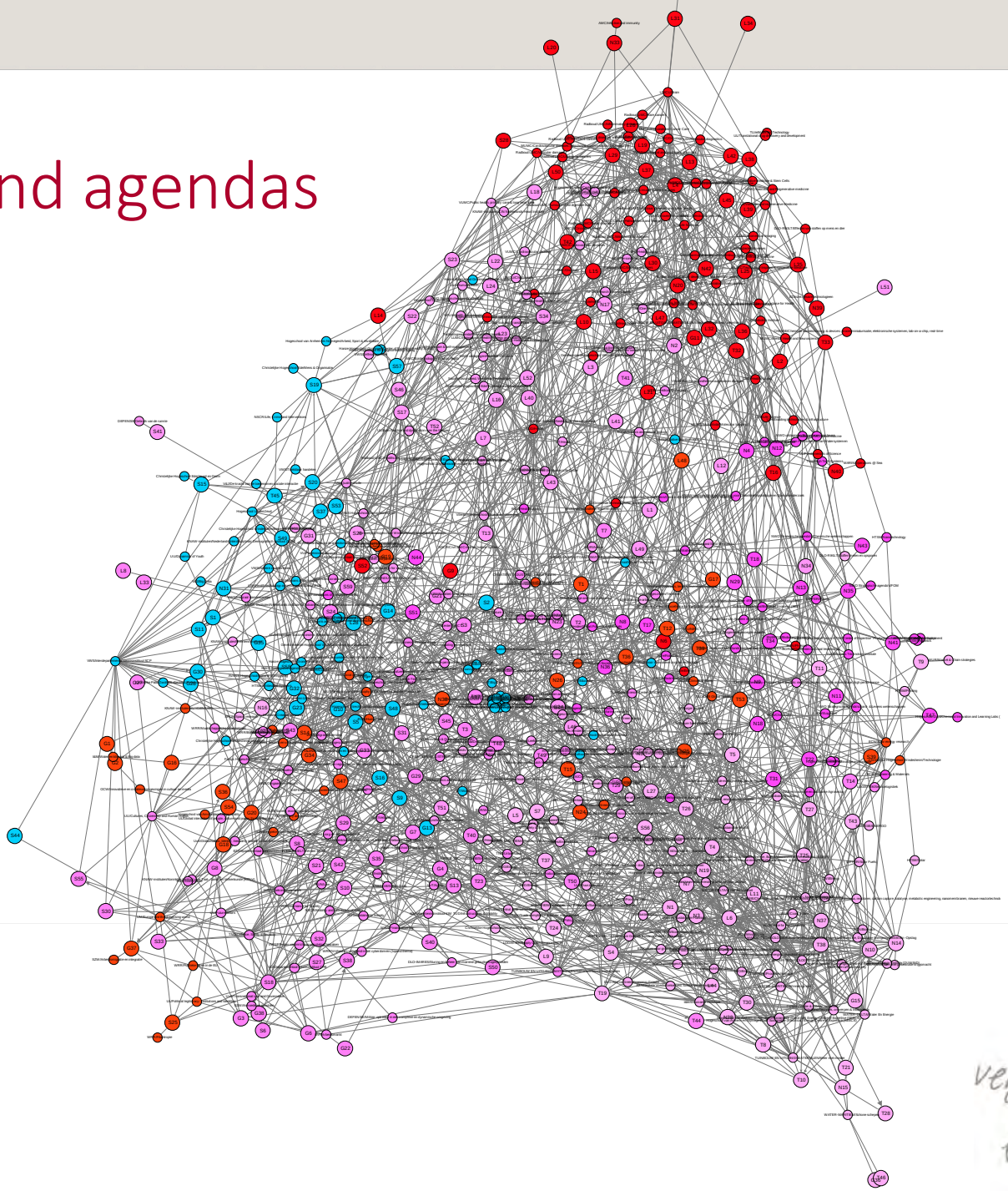
Matching the agenda to other agendas

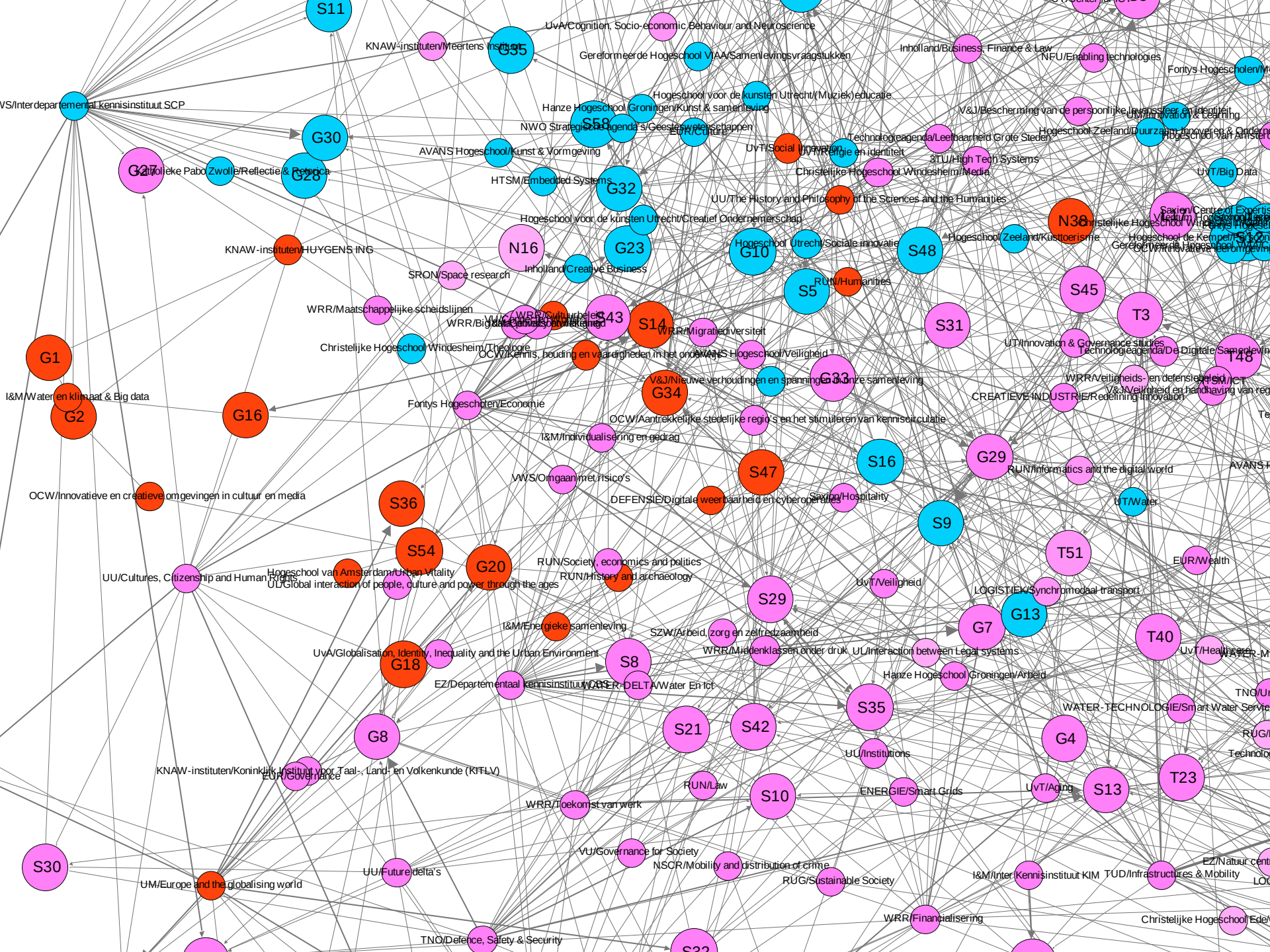
- Final questions produced by scientific juries using submitted questions

- Existing knowledge agendas
  - Scientific knowledge producers (universities, universities of applied science, PROs, university medical centres, etc.)
  - Societal knowledge users (government departments, regions, advisory bodies, UN, Horizon 2020, etc.)

- Currently 248 jury clusters and about 600 knowledge agendas

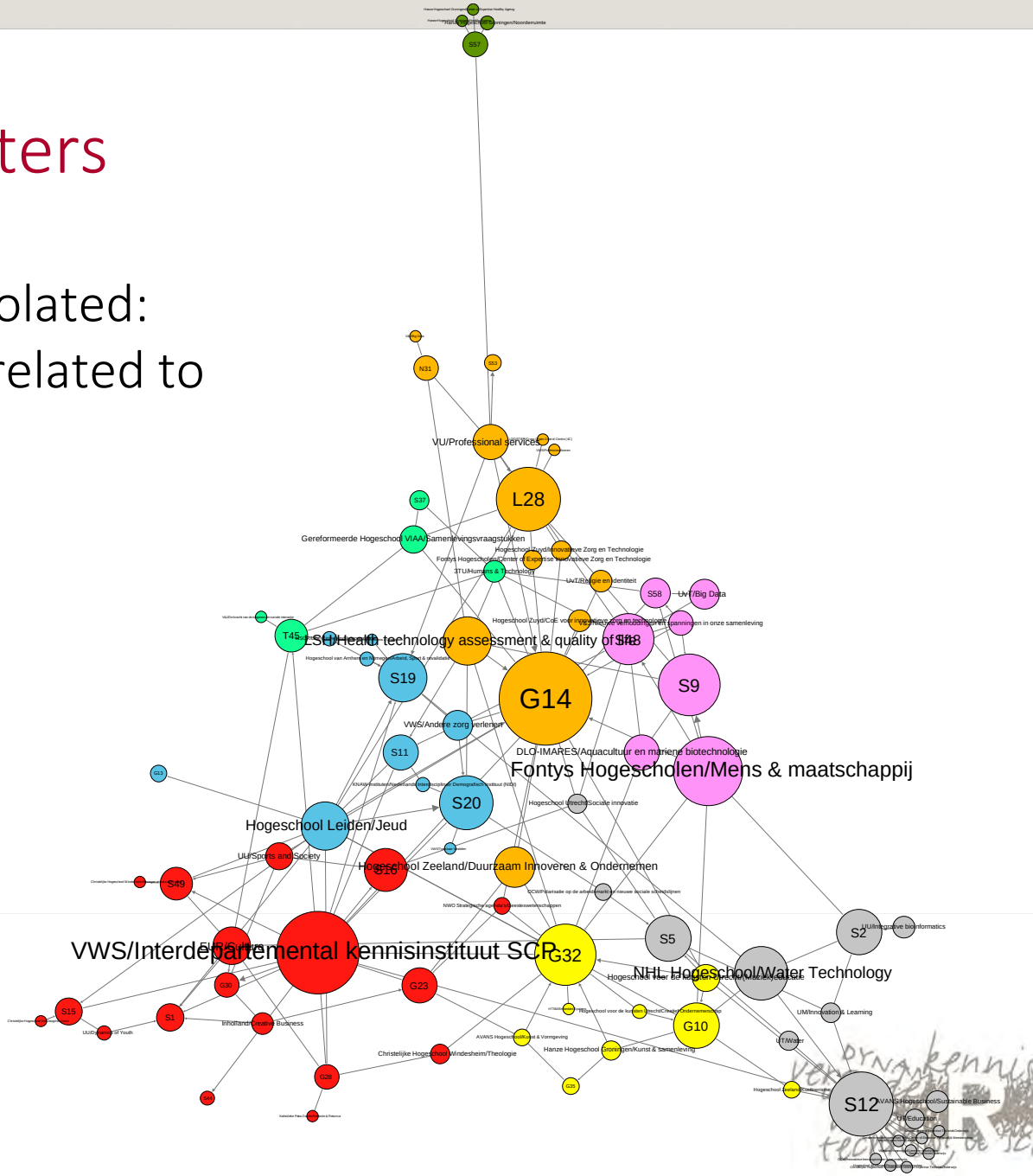
Questions and agendas




S11
KNAW-instituten/Meertens Instituut
G35
UvA/Cognition, Socio-economic Behaviour and Neuroscience
Gereformeerde Hogeschool VIAA/Samenlevingsvraagstukken
Inholland/Business, Finance & Law
NFU/Enabling technologies
Fontys Hogescholen/M
WS/Interdepartementaal Kennisinstituut SCP
G30
Hanze Hogeschool Groningen/Kunst & samenleving
Hogeschool voor de kunsten Utrecht/(Muziek)educatie
S58
NWO Strategische Agenda's/Geesteswetenschappen
EUR/Culture
V&J/Bescherming van de persoonlijke levenssfeer en identiteit
UM/Innovation & Learning
Hogeschool Zeeland/Duurzame Innovaties & Onderne
Hogeschool van Amsterd
Technologieagenda/Leefbaarheid Grote Steden
AVANS Hogeschool/Kunst & Vormgeving
UvT/Social Innovation
Identiteit en identiteit
3TU/High Tech Systems
G27
Katholieke Pabo Zwolle/Reflectie & Retorica
G28
HTSM/Embedded Systems
Christelijke Hogeschool Windesheim/Media
UvT/Big Data
G32
UU/The History and Philosophy of the Sciences and the Humanities
Hogeschool voor de kunsten Utrecht/Creatief Ondernemerschap
N38
Christelijke Hogeschool Windesheim/Media
Saxion/Centre of Expertis
Hogeschool Zeeland/Kusttoerisme
Hogeschool de Kempel/PS
Gereformeerde Hogeschool VIAA/
OCW/Innovatieve gebiedsgerichte
KNAW-instituten/HUYGENS ING
N16
G23
G10
Hogeschool Utrecht/Sociale innovatie
S48
SRON/Space research
Inholland/Creative Business
RUN/Humanities
S5
S45
WRR/Maatschappelijke scheidslijnen
S43
S14
T3
WRR/Migratiediversiteit
S31
G1
Christelijke Hogeschool Windesheim/Theologie
OCW/Kennis, houding en vaardigheden in het onderwijs
AVANS Hogeschool/Veiligheid
UT/Innovation & Governance studies
Technologieagenda/De Digitale Samenlevin
T48
G33
V&J/Nieuwe verhoudingen en spanningen binnen onze samenleving
WRR/Veiligheids- en defensiebeleid
HTSM/ICT
I&M/Water en klimaat & Big data
G34
V&J/Veiligheid en handhaving van reg
CREATIEVE INDUSTRIE/Redefining Innovation
G2
G16
Fontys Hogescholen/Economie
OCW/Aantrekkelijke stedelijke regio's en het stimuleren van kenniscirculatie
I&M/Individualisering en gedrag
S16
G29
S47
RUN/Informatics and the digital world
AVANS H
VWS/Omgaan met risico's
OCW/Innovatieve en creatieve omgevingen in cultuur en media
DEFENSIE/Digitale weerbaarheid en cyberoperaties
Saxion/Hospitality
UT/Water
S36
S9
S54
T51
EUR/Wealth
G20
RUN/Society, economics and politics
RUN/History and archaeology
Hogeschool van Amsterdam/Urban Vitality
UU/Cultures, Citizenship and Human Rights
UU/Global interaction of people, culture and power through the ages
UvT/Veiligheid
LOGISTIEK/Synchromodaal transport
S29
G13
G7
I&M/Energieke samenleving
SZW/Arbeid, zorg en zelfredzaamheid
T40
UvT/Healthcare
WRR/Middenklassen onder druk
UU/Interaction between Legal systems
UvA/Globalisation, Identity, Inequality and the Urban Environment
G18
S8
Hanze Hogeschool Groningen/Arbeid
EZ/Departementaal kennisinstituut OCW
WATER-DELTA/Water En Ict
S35
TNO/U
WATER-TECHNOLOGIE/Smart Water Servic
S21
S42
G8
G4
RUG/
UU/Institutions
KNAW-instituten/Koninklijk Instituut voor Taal-, Land- en Volkenkunde (KITLV)
EUR/Governance
T23
RUN/Law
S10
ENERGIE/Smart Grids
UvT/Aging
S13
WRR/Toekomst van werk
S30
VU/Governance for Society
NSCR/Mobility and distribution of crime
UU/Future delta's
UM/Europe and the globalising world
RUG/Sustainable Society
I&M/Inter Kennisinstituut KIM
TUD/Infrastructures & Mobility
EZ/Natuur cent
WRR/Financialisering
Christelijke Hogeschool Ede/
TNO/Defence, Safety & Security

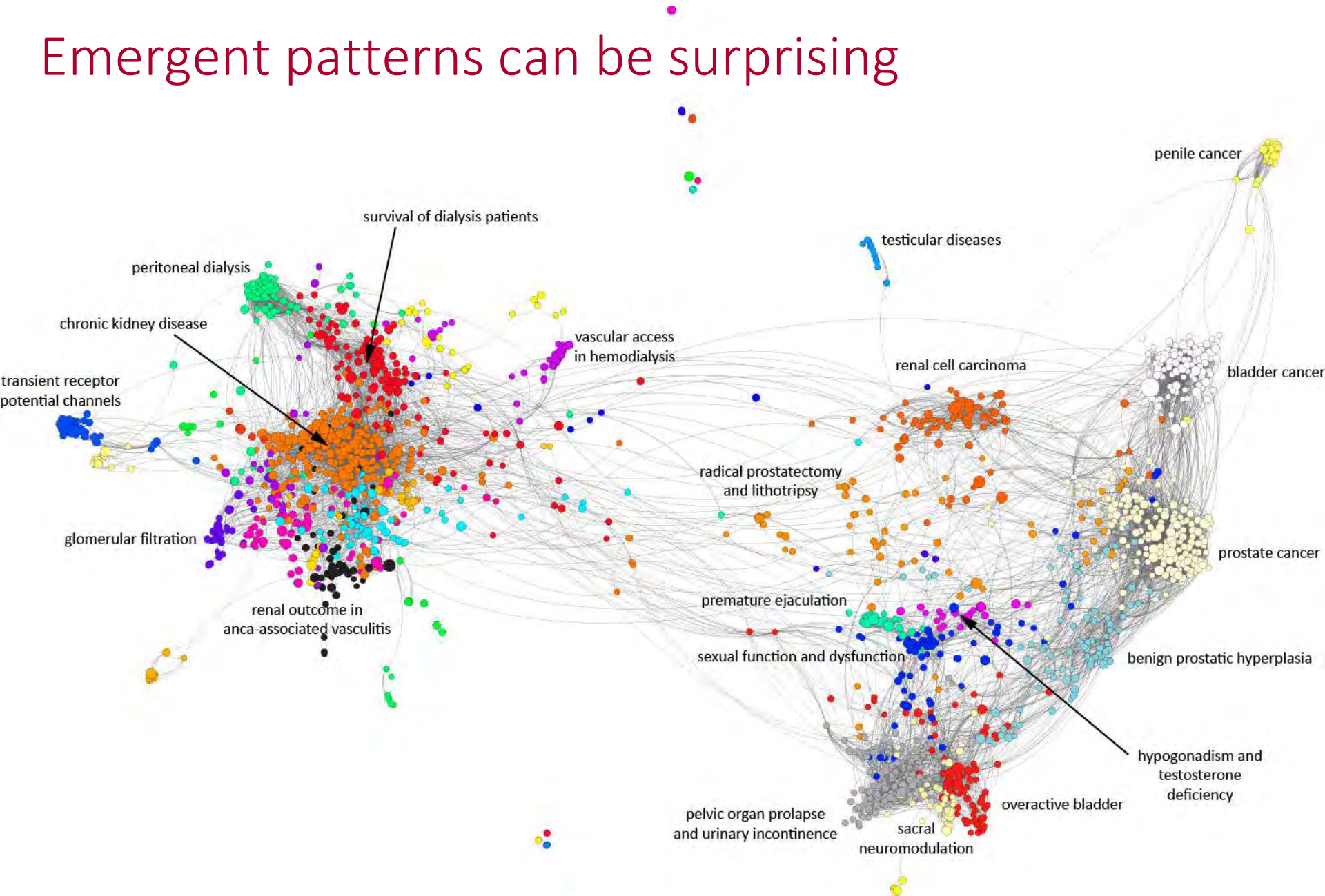
Clusters within clusters

- One particular cluster isolated: questions and agendas related to child development
- Mix of disciplines and organisations
- Reduce to something non-scientists can understand

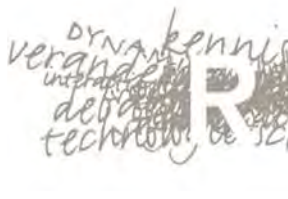


Rathenau Instituut

Emergent patterns can be surprising

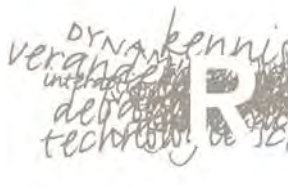


SO YOU WANNA APPLY SCIENTOMETRICS



Main lessons

- Focus on one interesting question and design the right approach
- Be precise and disciplined: keep a logfile to document what you are doing
- Do not be afraid of manual labour: tools cannot solve all problems
- Data are never clean: always clean and harmonise your data
- Make your assumptions explicit



Thanks for your attention

Edwin Horlings | e.horlings@rathenau.nl