

# Codebook: Unveiling the ecosystem of science

Nicolas Robinson-Garcia<sup>a,\*</sup>, Rodrigo Costas<sup>b</sup>, Thed N. van Leeuwen<sup>b</sup>, Tina Nane<sup>a</sup>

<sup>a</sup>*Applied Mathematics (DIAM), TU Delft, Delft, Netherlands*

<sup>b</sup>*CWTS, Leiden University, Leiden, Netherlands*

---

## Abstract

There is increasing evidence on the misuse and abuse of quantitative indicators in the current scientific reward system. Alternatively, more qualitative approaches, use of case studies or the design of indicators more sensitive to societal and scientific needs have been suggested. In this article we analyze the bases of such criticisms and motivations for changing the reward system by focusing on the assessment of individuals. We explore alternative models proposed or in use and identify common characteristics. Based on this we propose a valuation model by which we can systematically organize and prioritize performative aspects of scientists and consider other factors which may affect or might relevant for research policy. We finally test our model in a series of case studies based on six academic units.

---

Here I describe the selection of case studies and the data retrieval process and description of sources. The purpose of this is to show that 1) there are indeed a variety of different profiles, 2) these profiles co-exist and complement each other, 3) the diversity and typologies of profiles is field dependent, 4) personal features are key to understand this diversity. Here we emphasize two specific personal features: age and gender. Some of the research questions that could be answered descriptively are the following:

1. *Is there team science?*
- i. *Are these teams stable over time?* Some data that could be retrieved from WoS

- `Cluster_id`
- All classic bibliometric indicators
- Number of collaborators from same institution
- Number of external collaborators
- Number of collaborators from same institution by publication
- Number of external collaborators by publication

---

\*Corresponding author

Email address: N.Robinson@tudelft.nl (Nicolas Robinson-Garcia)

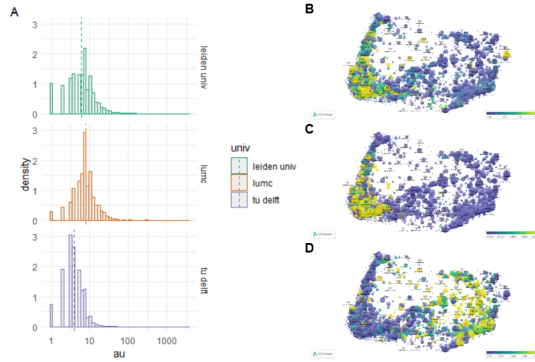
- ii. *Is there activity visible through co-authorship?* Check groups identified with manual checking in the web and maybe interviews
- 2. *Do research teams operate in a coordinated way?* Here it would be interesting to understand dependence relationship. Nederhof & van Raan (1993) refer to the star effects as what happens when a PI retires and the research group disappears. Here we should go beyond bibliometrics and see if there is someone in charge of the funding, someone of hiring and finding opportunities, someone who is more of a public face, etc. Also it would be interesting to use network analysis to determine authorities, hubs, etc. and contrast with their judgment. Some variables from WoS and interviews plus manually checking: authorship position (WoS); acknowledgments data (WoS); clusters from Ludo's subject classification to identify areas of specialization per subject (WoS); social media activity; Google Scholar data.
  - i. *Do they have a common research agenda?*
  - ii. *How is this agenda established?*
- 3. *How does team science affect individual trajectories?*
  - i. *How is credit shared?*
  - ii. *What is the relation between the role exerted and academic status?* A cohort analysis?
  - iii. *How is continuity of supporting scientists ensured?* Probably also from interviews. How do scholars change from institution or are maintained if they are not able to 'make the next step'.

#### *Selection of case studies*

The identification of research groups is done bibliometrically and based on Web of Science publications. For this I have selected all publications between 2008 and 2017 by LUMC, Leiden University or TU Delft. The following table includes some descriptives. Here I must note that researchers belonging to institutions are not based on the specific affiliation linkage of docs (which uses the Leiden Ranking affiliation already cleaned up), but based on `cluster_id` with either of the three institutions as their main or alternative address. This should be checked to see if there is a way to link to the `cluster_id` organizations to the cleaned affiliations from Leiden Ranking. I had to clean up this data myself. In any case this should not be a concern for the selection of case studies.

	TU Delft	Leiden Univ	LUMC
Publications	24,233	49,149	3,188
Researchers	9,975	14,116	279
Collaborators	34,409	117,307	14,153
Mean au/p	4.7	9.6	10.1
Median au/p	4	6	8
Sd au/p	5.4	31.5	18.7

The next figure shows the distribution of papers based on the number of authors by paper (A) and the thematic profile of each of the three institutions. Leiden University is the largest of the three institutions with a more comprehensive portfolio, although mostly focuses on Biomedical Sciences and Social Sciences. This focus on biomedical sciences is obviously more noticeable in the case of LUMC, although it still has some publications in fields of the social sciences, mostly related with Public Health. Finally, TU Delft shows a profile focused on Physics, Engineering and Mathematics. While there might be an overlap between LUMC and Leiden University, the high preponderance of biomedical literature might also be due to a close relation between these universities.



Six case studies will be selected. Three for each university and two by field. The purpose of this is not only to identify differences by discipline but also by institutional type. The fields are:

1. Physics and Engineering
2. Social Sciences and Humanities
3. Biomedical Sciences

Following I include the collaboration networks for each university and field. I have included a threshold of at least 10 publications by `cluster_id`, filtered by the giant component and calculated the betweenness centrality of each node. I have selected as seed researcher the one with the highest centrality.

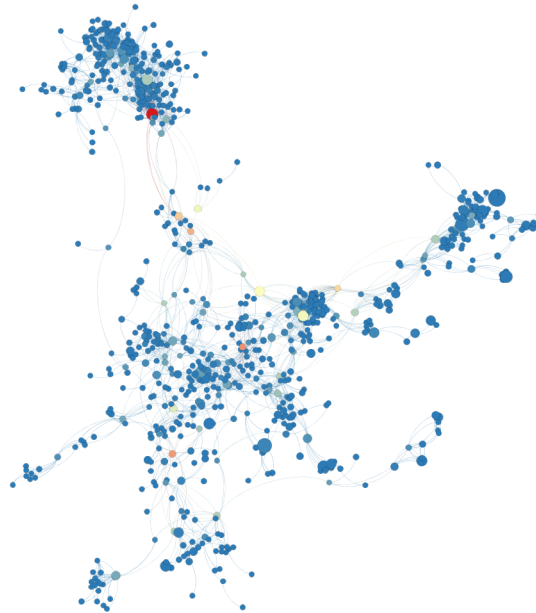
*Physics & Engineering*



1. *TU Delft.*

*Notes:*

- 1260 nodes (21.9% visible); 4702 edges (27.1% visible).
- `cluster_id` with highest betweenness = 33800547; Betweenness centrality: 0.11; Total publications = 159; Age = 32
- Name: Frans D. Tichelaar; First year: 1986; Last year: 2018
- PURE: [https://pure.tudelft.nl/portal/en/persons/fd-tichelaar\(56299c58-b6ec-478b-b188-b8744b69d954\).html](https://pure.tudelft.nl/portal/en/persons/fd-tichelaar(56299c58-b6ec-478b-b188-b8744b69d954).html)
- Institution: <http://nchrem.nl/people/dr-ir-f-d-tichelaar-frans/>

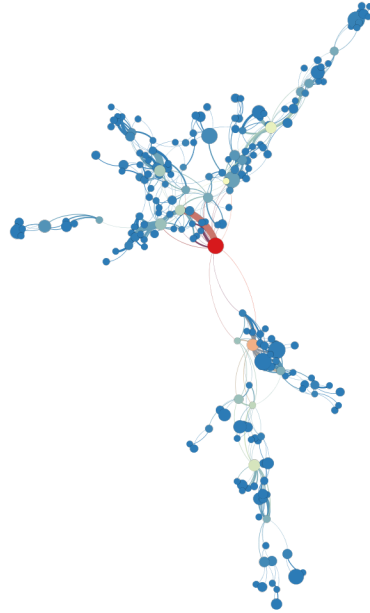


## 2. *Leiden Univ.*

### *Notes:*

- 765 nodes (35.1% visible); 3409 edges (40.6% visible).
- `cluster_id` with highest betweenness = 25501410; Betweenness centrality: 0.23; Total publications = 590; Age = 38
- Name: Ewine F. van Dishoeck; First year: 1980; Last year: 2018
- Institution: <https://local.strw.leidenuniv.nl/people/touchscreen2/persinline.php?id=16>
- Personal website: <https://home.strw.leidenuniv.nl/~ewine/>

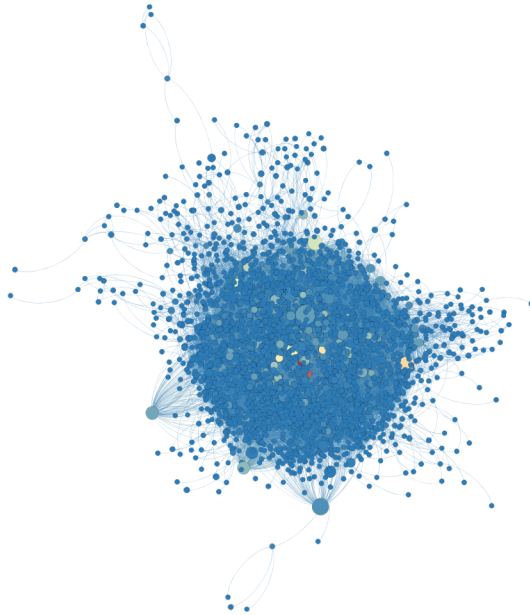
*Biomedical and Health Sciences*



1. *TU Delft.*

*Notes:*

- 214 nodes (18.9% visible); 615 edges (25.49% visible).
- **cluster\_id** with highest betweenness = 43204348; Betweenness centrality: 0.50; Total publications = 341; Age = 31
- Name: Harrie H. Weinans; First year: 1987; Last year: 2018
- Google Profile: <https://scholar.google.com/citations?user=di4NUp8AAAAJ&hl=en>
- PURE: [https://pure.tudelft.nl/portal/en/persons/hh-weinans\(f31bd75b-1863-4202-b64b-7356538284a7\)/publications.html](https://pure.tudelft.nl/portal/en/persons/hh-weinans(f31bd75b-1863-4202-b64b-7356538284a7)/publications.html)
- Co-affiliated to UMC Utrecht and TU Delft.

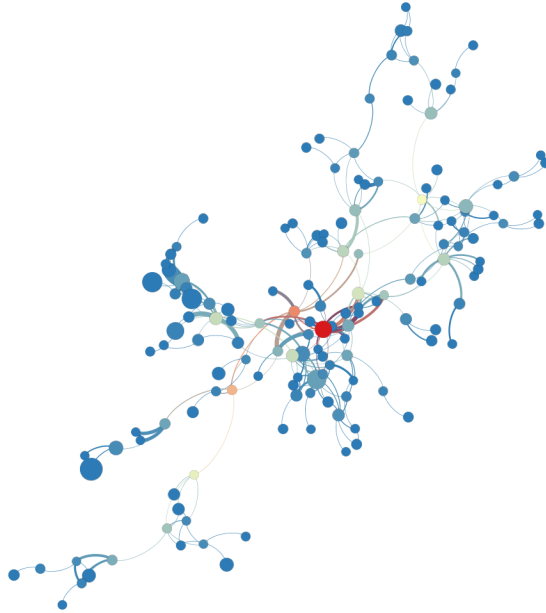


2. *Leiden Univ.*

*Notes:*

- Due to the density of the network the selected node can scarcely be seen.
- 3304 nodes (38.8% visible); 43647 edges (59.6% visible).
- `cluster_id` with highest betweenness = 19936939; Betweenness centrality: 0.47; Total publications = 185; Age = 27
- Name: Ron Wolterbeek; First year: 1991; Last year: 2018
- Institution: <https://www.lumc.nl/org/bds/medewerkers/rwolterbeek>
- Affiliated to LUMC.

*Social Sciences and Humanities*

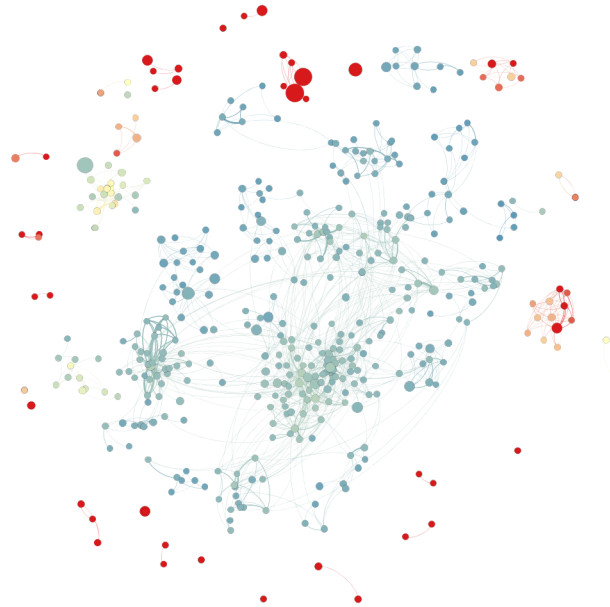


1. *TU Delft.*

*Notes:*

- 151 nodes (12.6% visible); 281 edges (15.8% visible).
- **cluster\_id** with highest betweenness = 12392841; Betweenness centrality: 0.41; Total publications = 134; Age = 18
- Name: Bert van Wee; First year: 1999; Last year: 2018
- Google Profile: <https://scholar.google.es/citations?user=dYUiqMYAAAAJ&hl=en>
- Institution: <https://www.tudelft.nl/en/tpm/about-the-faculty/departments/engineering-systems-and-services/people/full-professors/profdr-gp-bert-van-wee/>





## 2. *Leiden Univ.*

### *Notes:*

In this case, the selection of the seed researcher was not based on network indicator. The network above has the OpenOrd layout (instead of Yi ) and K-Core=1 group. The main issue here is that this field is largely populated by biomedical scientists and psychologists and psychiatrists, fields which are not good representations of Social Sciences and Humanities. What I have done is looked at those pairs of scholars with the highest shares of co-authored papers and go down the list until I found someone who was not from these fields nor from CWTS (Ludo and Nees are the third pair with more co-authored papers)

- 478 nodes (26.5% visible); 1294 edges (43.1% visible).
- `cluster_id` selected = 36871407; Betweenness centrality: 0.00; Total publications = 78; Age = 18
- Name: Judi Mesman; First year: 2000; Last year: 2018
- Institution: <https://www.universiteitleiden.nl/en/staffmembers/judi-mesman/publications#tab-1>
- Lab1: <http://www.diversityinparenting.nl/>
- Lab2: <https://www.societalchallengeslab.com/>

### *Expansion from seed to complete team*

Based on the six individuals selected in the first phase. I know search for their complete research team. The data retrieval process started on March, 2019. Here I include for each case how I have proceeded.

*Physics & Engineering - TU Delft*

Dr. Ir. F.D. Tichelaar belongs to the National Centre for High Resolution Electron Microscopy. According to its website, he is not the head of the institute which is formed by 9 researchers (5 staff and 4 researchers and post docs)

*Physics & Engineering - Leiden Univ*

Prof. Ewine van Dishoeck belongs to the Leiden Observatory. According to its website, there are 179 workers: 33 are staff members, 50 postdocs, 63 PhD students and 26 supporting staff.

*Biomedical Sciences - TU Delft*

Harrie H. Weinans is associate professor at UMC Utrecht in the department of Orthopaedics and Professor at TU Delft at the Biomechanical Engineering department. Here information is gathered from TU Delft. He is a parttime professor. Here I will not include the whole department, but only researchers working on the Biomaterials & Tissue Biomechanics group. Only supporting staff assigned to this group is included.

*Biomedical Sciences - Leiden University*

Ron Wolterbeek is first-line consultant in the area of Medical Statistics at LUMC. I could not find him actually in the staff page, so decided to look into the staff included there. There is not much about him in the Internet other than his publications.

*Social Sciences and Humanities - TU Delft*

Bert van Wee is professor in Transport Policy at the department of Engineering Systems and Services. In this case I do not find anything as research teams or groups and hence I am including the whole department.

*Social Sciences and Humanities - Leiden Univ*

Judi Mesman is dean of Leiden University College The Hague and professor of the interdisciplinary study of societal challenges. She directs the Societal Challenges Lab which is part of both the Faculty of Governance and global Affairs and the Faculty of Social Sciences at Leiden University. This case study is based on her lab.

*Data sources*

Data sources are selected based on dimensions from the *valuation model*. Following I include the main ones selected:

- **CWTS-Web of Science.** Publication and citation indicators are retrieved since 1980 or first publication to 2018 by matching individual to `cluster_id` following the author name disambiguation developed by Caron & van Eck (2014).
-

## References