Agent Based Model of research as a social proces: Dynamics of disciplines' popularity as function of funding and fashion

Pawel Sobkowicz

Funds, fashions … and discoveries

background to modeling of science as a social process

- What motivates scientists?
  - Curiosity
  - Fame and recognition
  - Altruism
  - Money (survival?)
  - Influence of mentors
  - Influence of peers (fashions?)
  - Fear of other employment prospects?
- What motivates societies to fund science?
  - Direct utility of research (otherwise described as impact)
  - Specific needs (e.g. defense programs, healthcare)
  - Tradition (continuation of long established institutions)
  - National or local pride and reputation (showing that a community can spend money on useless but prestigious activities)
  - Decisions of authorities (democratic or otherwise), especially those **not** to fund certain research.

Other important factors

- (Hyper)Exponential growth of: numbers of scientists, research topics, publications, journals …
- High levels of competition for diminishing *per capita* funding
- Scientific community focus on success and neweness (aka innovation)
  - And the associated problems of lack or replication, fraud, statistical massaging of data…
- **Short term, grant based funding processes**
  - Short term vision
- Mixed evaluation processes:
  - **Political and economic pressures (focus on „impact")**
  - **Evaluating promises, not results**
  - Cliques, old-boy networks, hierarchical structures, parochialism (my discipline is better/more important than yours…)

Why Agent Based Models?

Modelling offers a chance to go beyond simple descriptions. J Epstein lists 16 reasons in his famous „Why model" paper:

1. **Explain (very distinct from predict)**
2. **Guide data collection**
3. **Illuminate core dynamics**
4. Suggest dynamical analogies
5. Discover new questions
6. Promote a scientific habit of mind
7. Bound (bracket) outcomes to plausible ranges
8. Illuminate core uncertainties.
9. Offer crisis options in near-real time
10. **Demonstrate tradeoffs / suggest efficiencies**
11. **Challenge the robustness of prevailing theory through perturbations**
12. **Expose prevailing wisdom as incompatible with available data**
13. Train practitioners
14. Discipline the policy dialogue
15. Educate the general public
16. Reveal the apparently simple (complex) to be complex (simple)

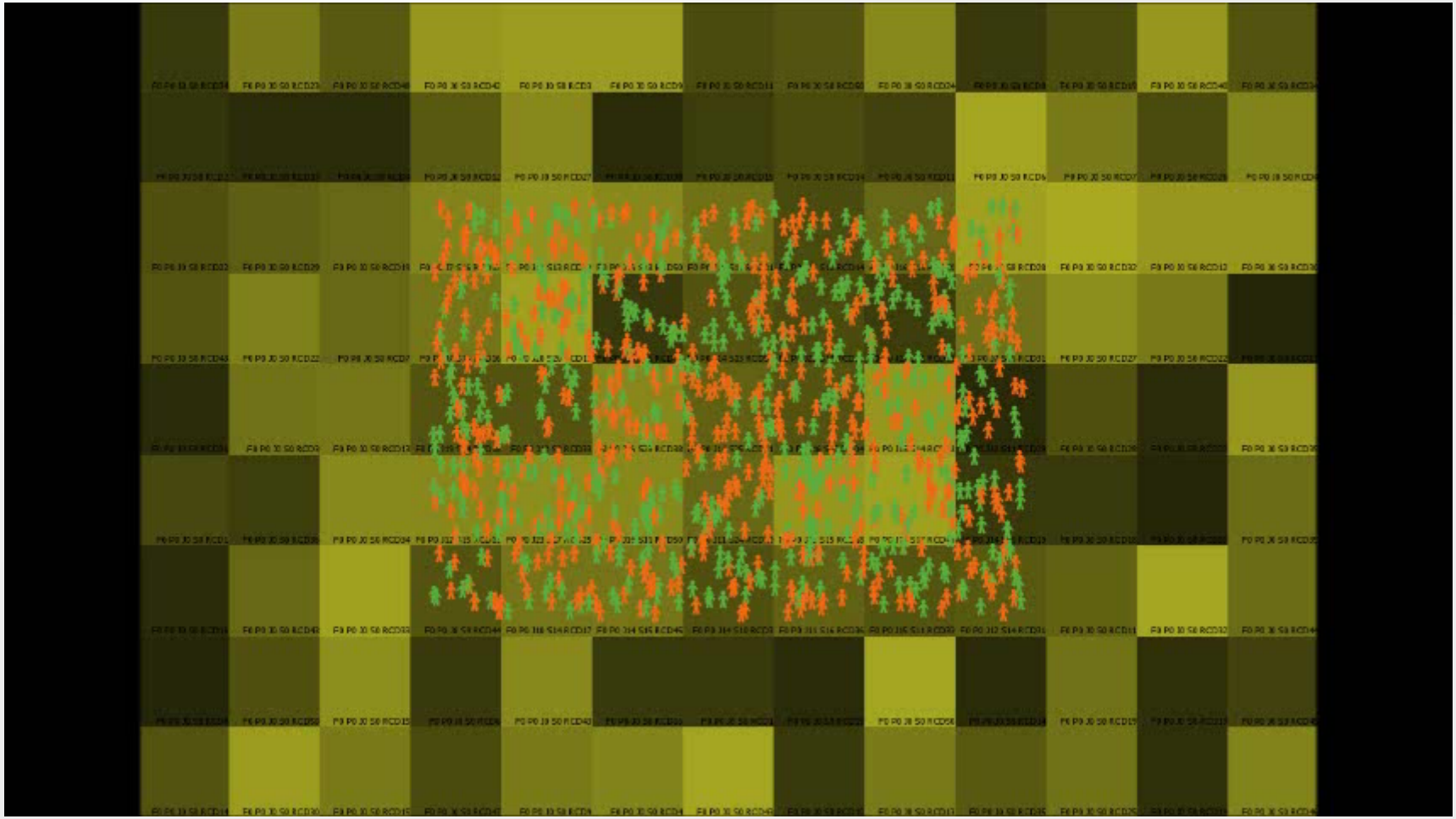
Why Agent Based Models?

- This view is reinforced by A Nowak, A Rachwalska and W Borkowski, in their paper „Why Simulate? To Develop a Mental Model".
- The computer simulations are especially important in situations where simple mental models are not able to capture the complexity of the phenomena involved.
- Especially, when the interactions and dynamics rules are nonlinear, i.e. when small causes can lead to large effects.
- Science as a proces, with its complexity, unpredicatability and multiple types of interactions almost begs for a simluation-based description.

Our starting point: a toy model of science (hopefully with some potential…)

- Each research topic (subdiscipline) = square patch
- Senior and junior scientist agent types
- Research progress due to „normal” and „crucial” discoveries, the latter driving the topic popularity **(curiosity and fame)**
- Funding for each topic depends on its societal impact **(utility)**, effects of crucial discoveries and lobbying **(money)**
- Agents apply for grants, conduct research or decide to „move” to a neighboring topic patch (where their chances could be better).
- Successful agents thrive, unsuccessful leave the system
- Net Logo model available from the author
  - Quite a few parameters to play with

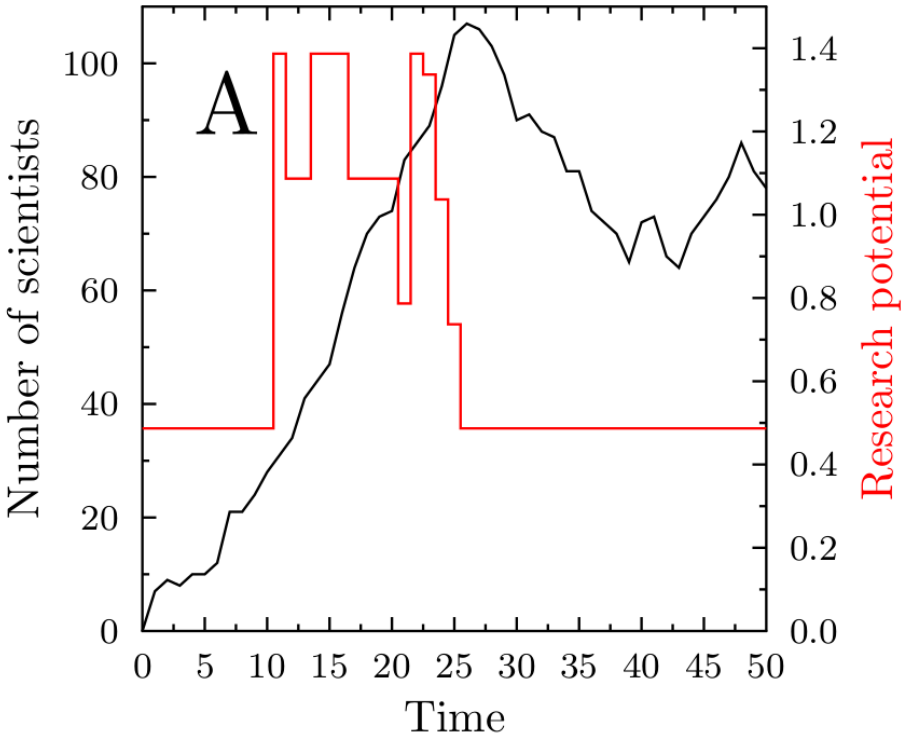
Toy model: scientists following the money, discoveries and fashions

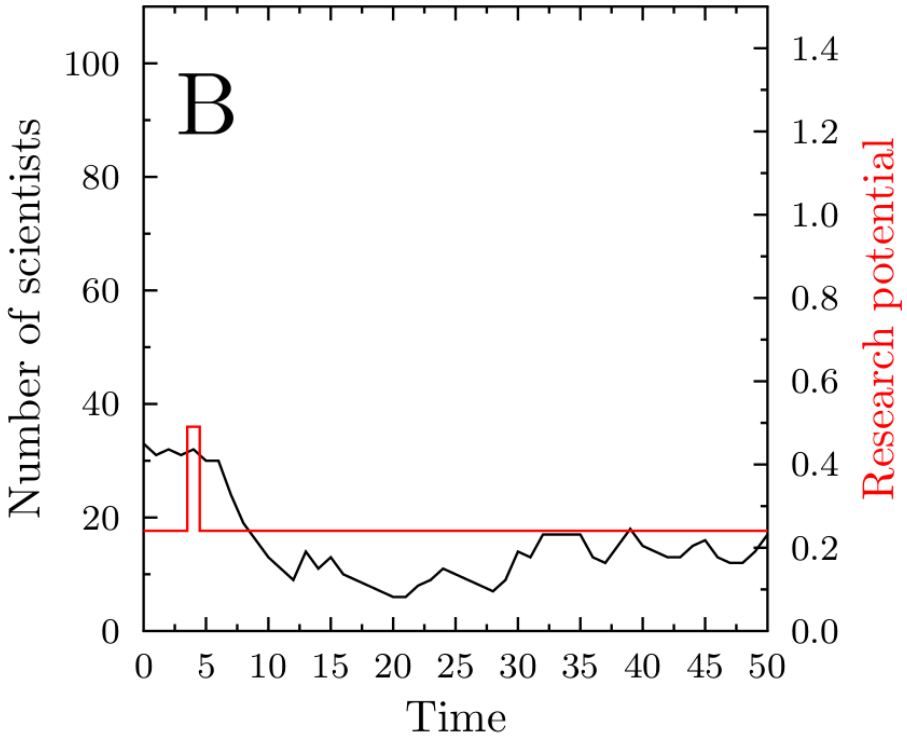


Popularity of topics: examples of time evolution

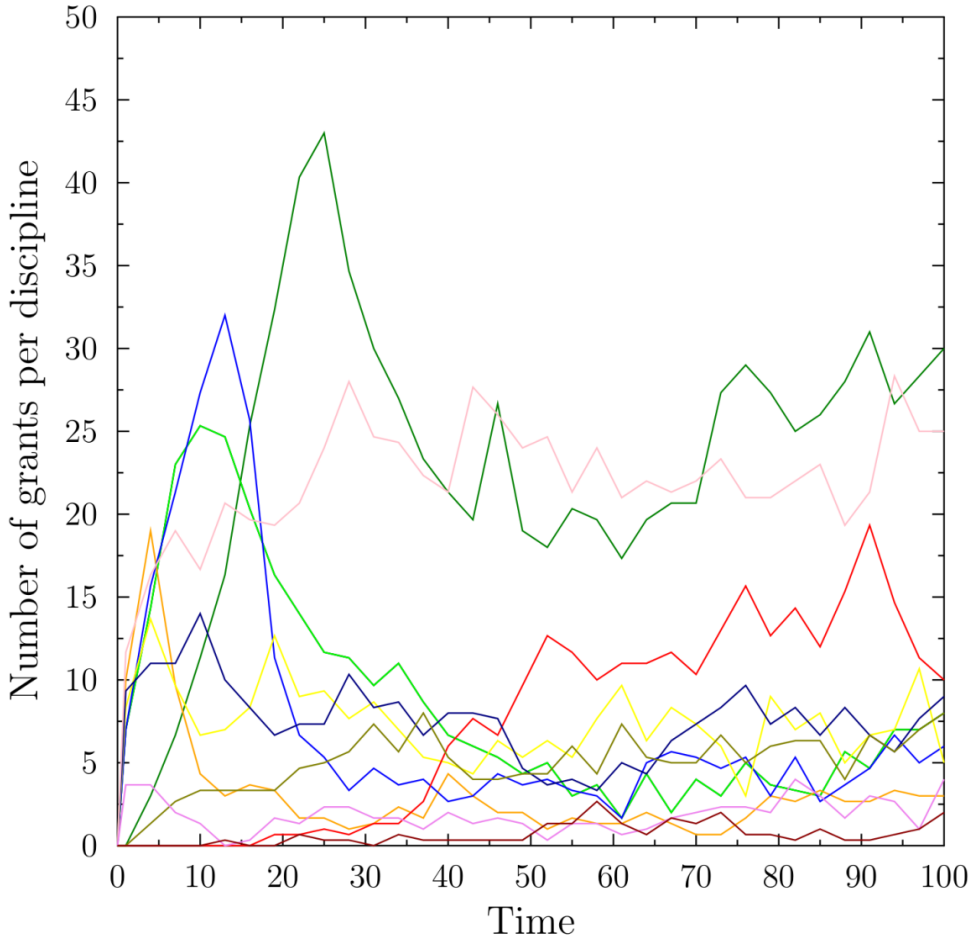
Important and interesting topic
(many crucial discoveries)

Much less important topic



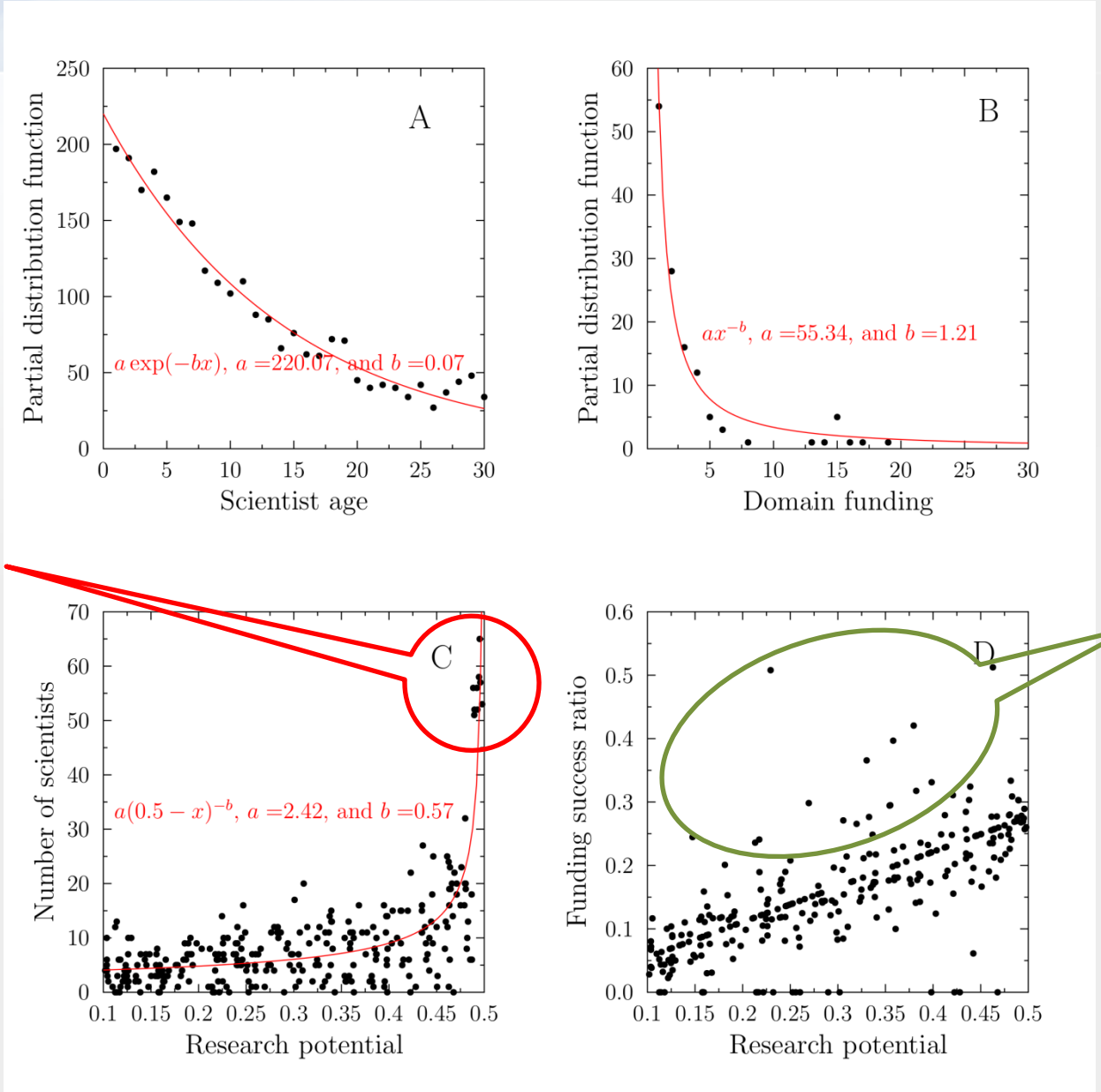


Funded grants for selected topics (patches)



Other selected simulation statistics

Following the money



Fashionable or well lobbied domains

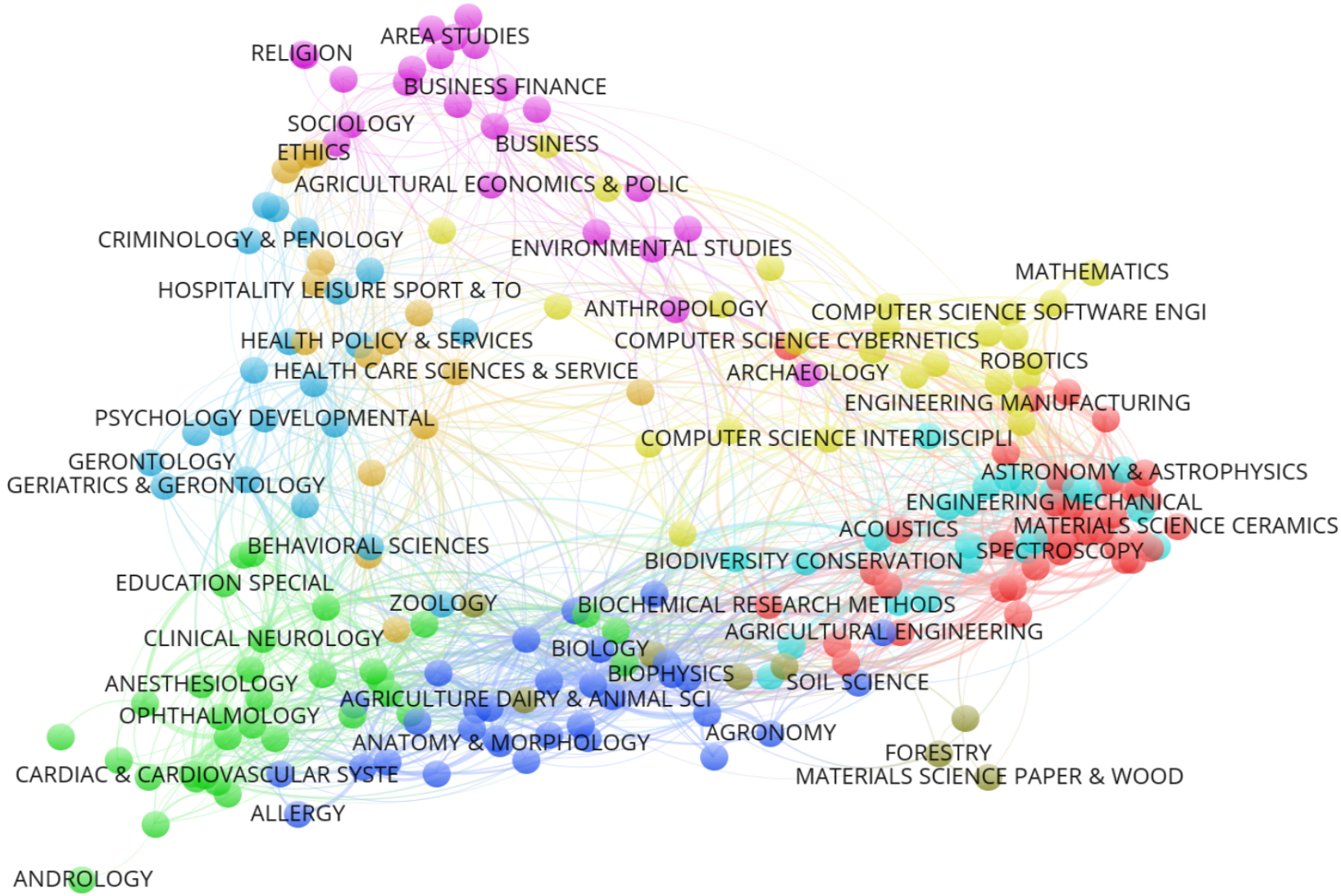
How to make the model realistic?

From abstract square lattice to topics network map

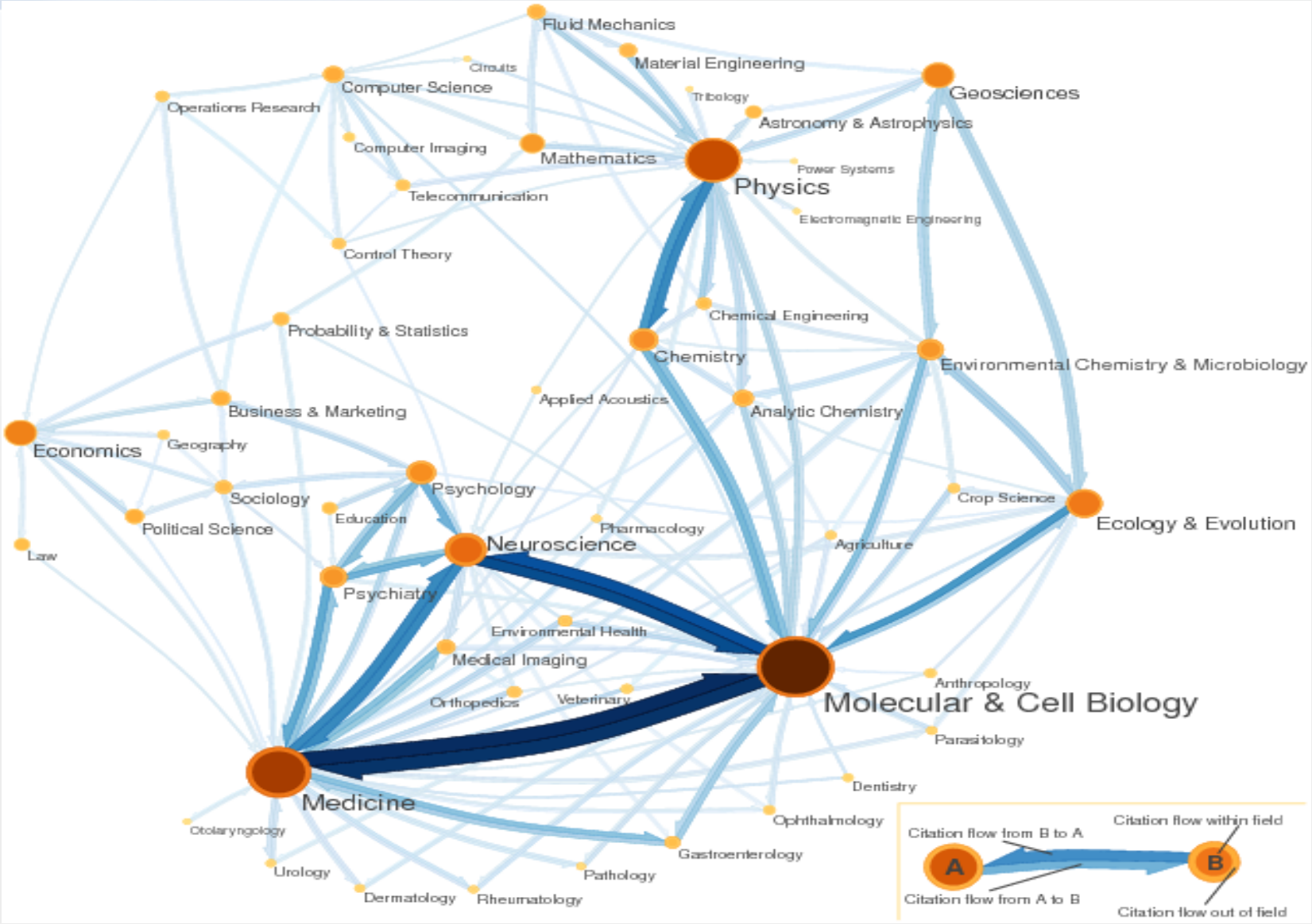
- Almost every country and funding agency uses its own list of disciplines and research topics. Some are rather general, some quite detailed.
- But how to describe the „distances" between them, the interactions and the possibility, for a scientist, to move from one topic to another?
- There are works creating such discipline networks based on various types of connections:
  - Analyses of common topics in research journals (e.g.Van Eck and Waltman)
  - Citation flows
  - Keyword based analyses
  - Topics declared in proposal submissions

Network of disciplines

based on research journals map by Van Eck & Waltman

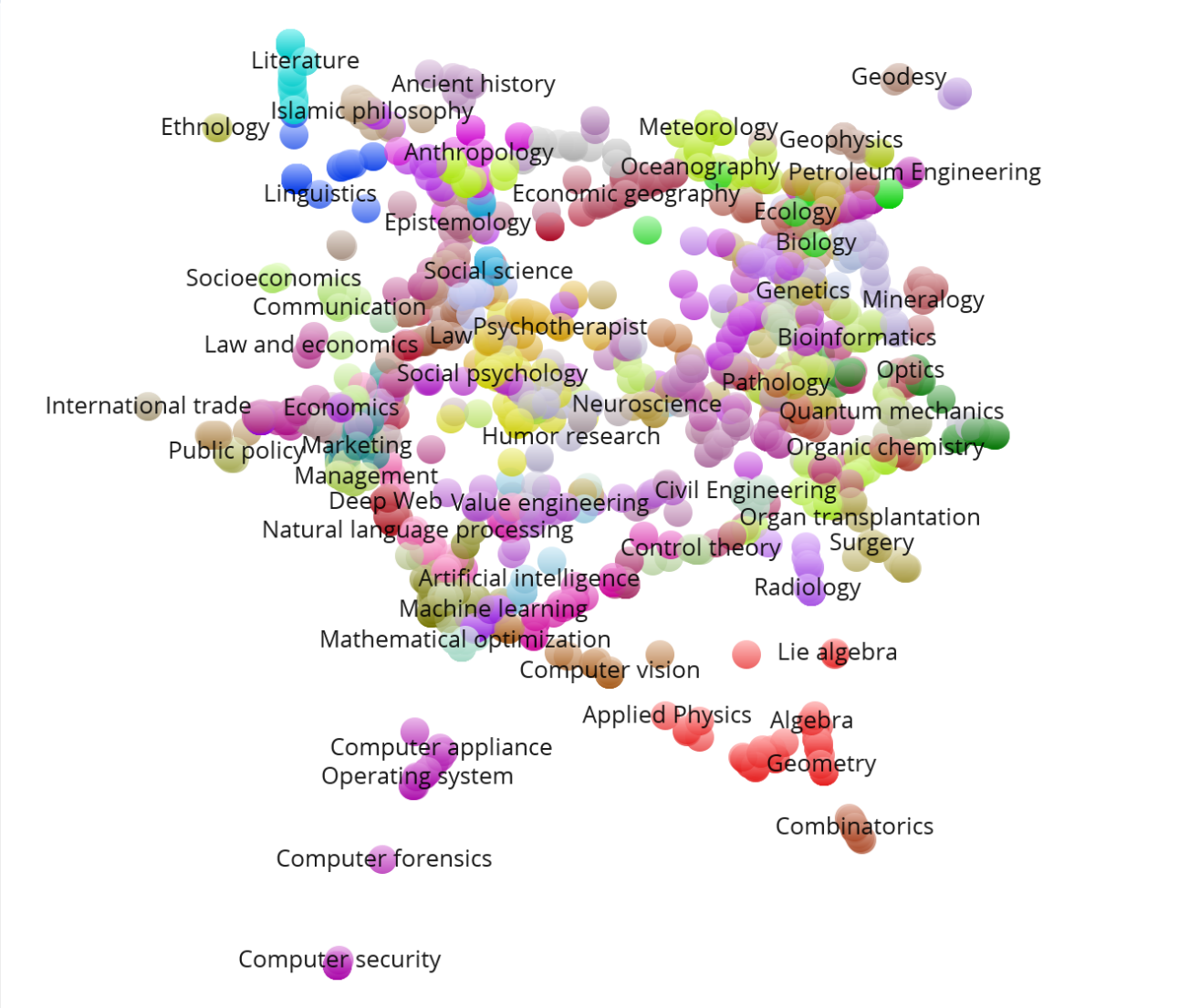


The EIGENFACTOR project maps
based on citation flows (West, Bergstrom, Rosvall)



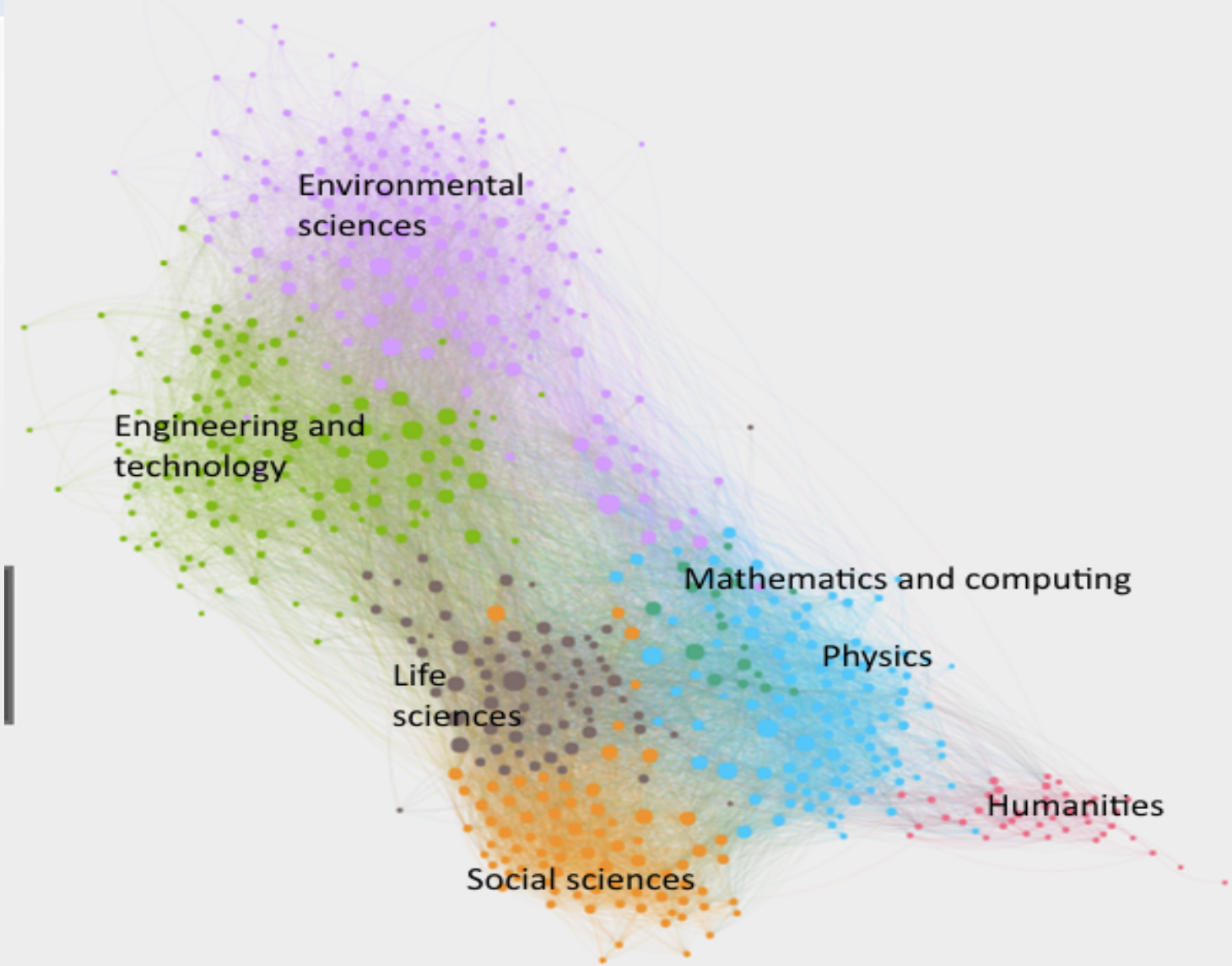
Microsoft Academic Search discipline map

based on work by Sinha, Shen, Song, Hao Ma, Eide, Hsu, and Wang



UK Gateway to Research (GtR) topics map

Juan Mateor Garcia, James Gardiner

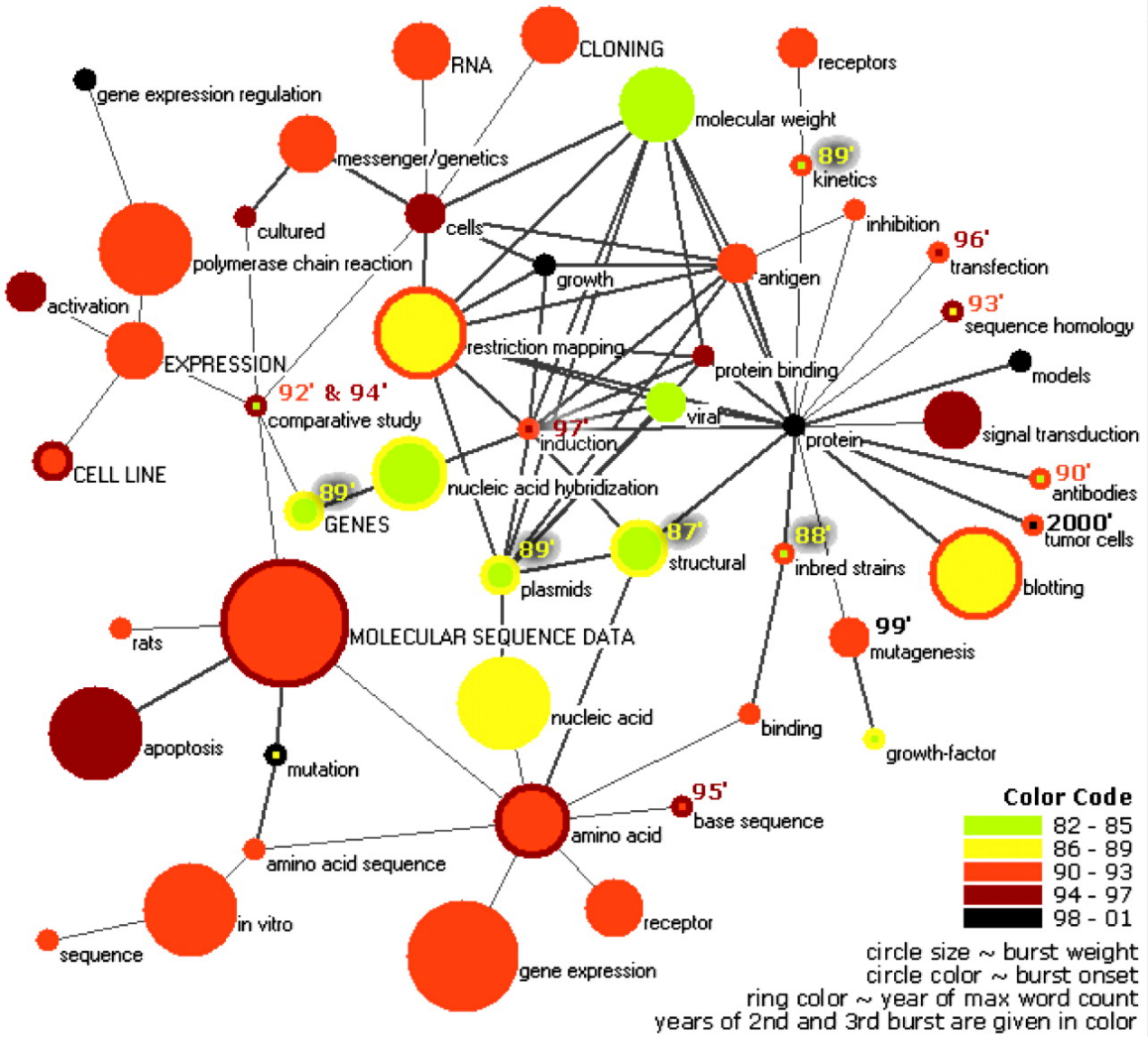


From static maps to dynamic picture (much more difficult)

- The time evolution data on the research effort (number of scientists) and funding in each small domain are even much harder to obtain
- Even assuming that disciplines and topics pre-exist (equivalent to the empty squares in the toy model), are there any data on the actual movements of scientists between the research topics?
- There are very few works on this. A good example is Mapping topics and topic bursts in PNAS by Ketan K. Mane and Katy Börner
  - they used co-word association maps of major topics based on highly frequent words and words with a sudden increase in usage, a phenomenon called "burst"
  - But the time resolution is low, and the keywords are a poor proxy for scientists…

Temporal dynamics of interest in topics

Ketan K. Mane and Katy Börner



From network maps to funding

- In any case, mapping the research topics is only the first step
- **We need data on the:**
  - **Funding devoted to specific subjects**
  - **Numbers of grant proposals in each of them**
  - **Success ratios / numbers of funded proposals**
  - **Numbers of researchers working on specific topics**
  - **Movement of these researchers between topics and disciplines**
- These data are rather difficult to gather because:
  - Funding comes from diverse sources (national and international) so the use of single database may fail to capture the process
  - Scientists are mobile (institution to institution, country to country)
  - Topic divisions are not the same everywhere and they change in time.

- Still, just to show what may be found in publicly available data …

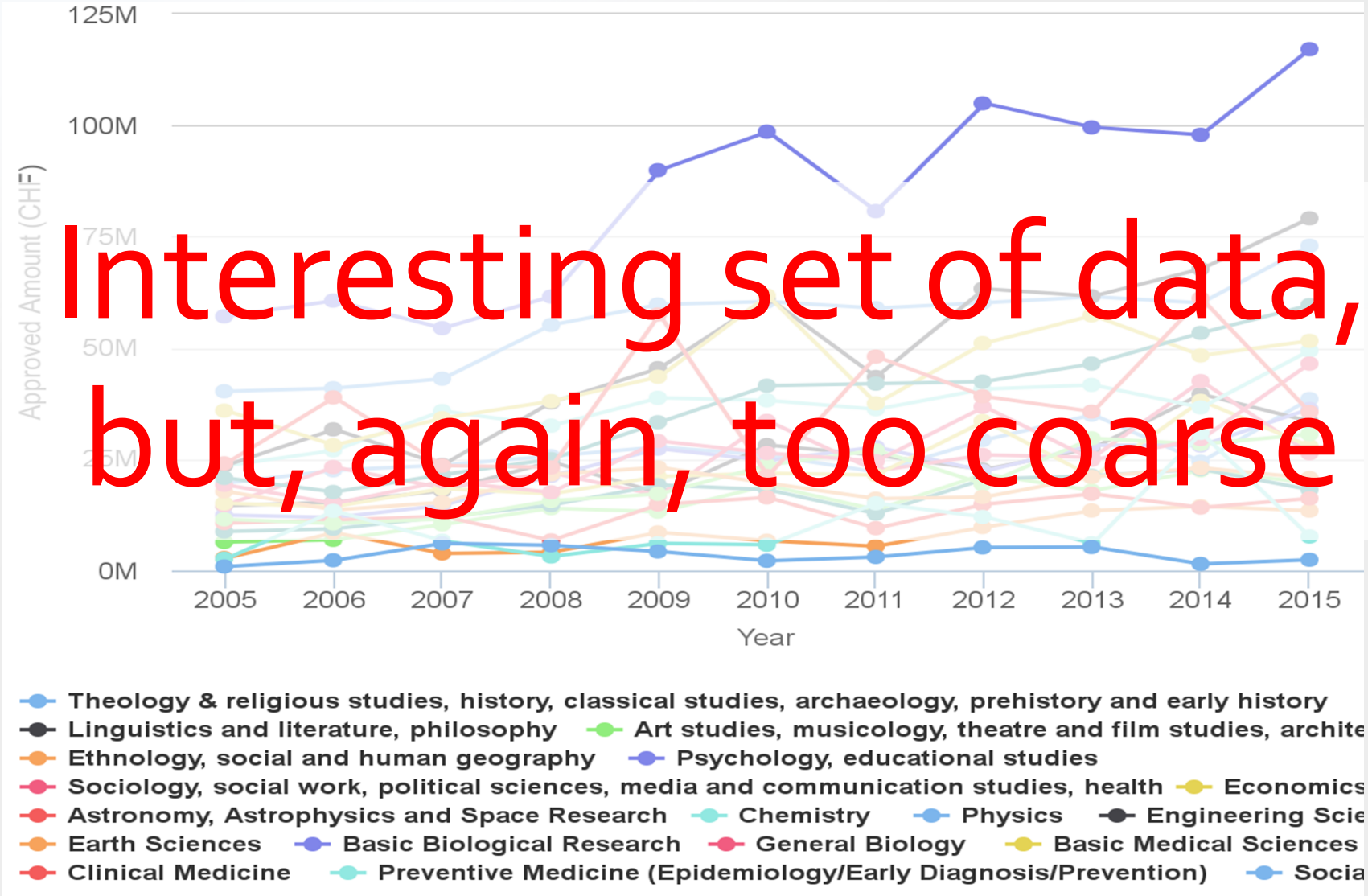
US funding by discipline (NSF)

Agarwala, R. & Teitelbaum, E., 2010. Trends in funding for dissertation field research: why do political science and sociology students win so few awards?

Interesting set of data, but too coarse

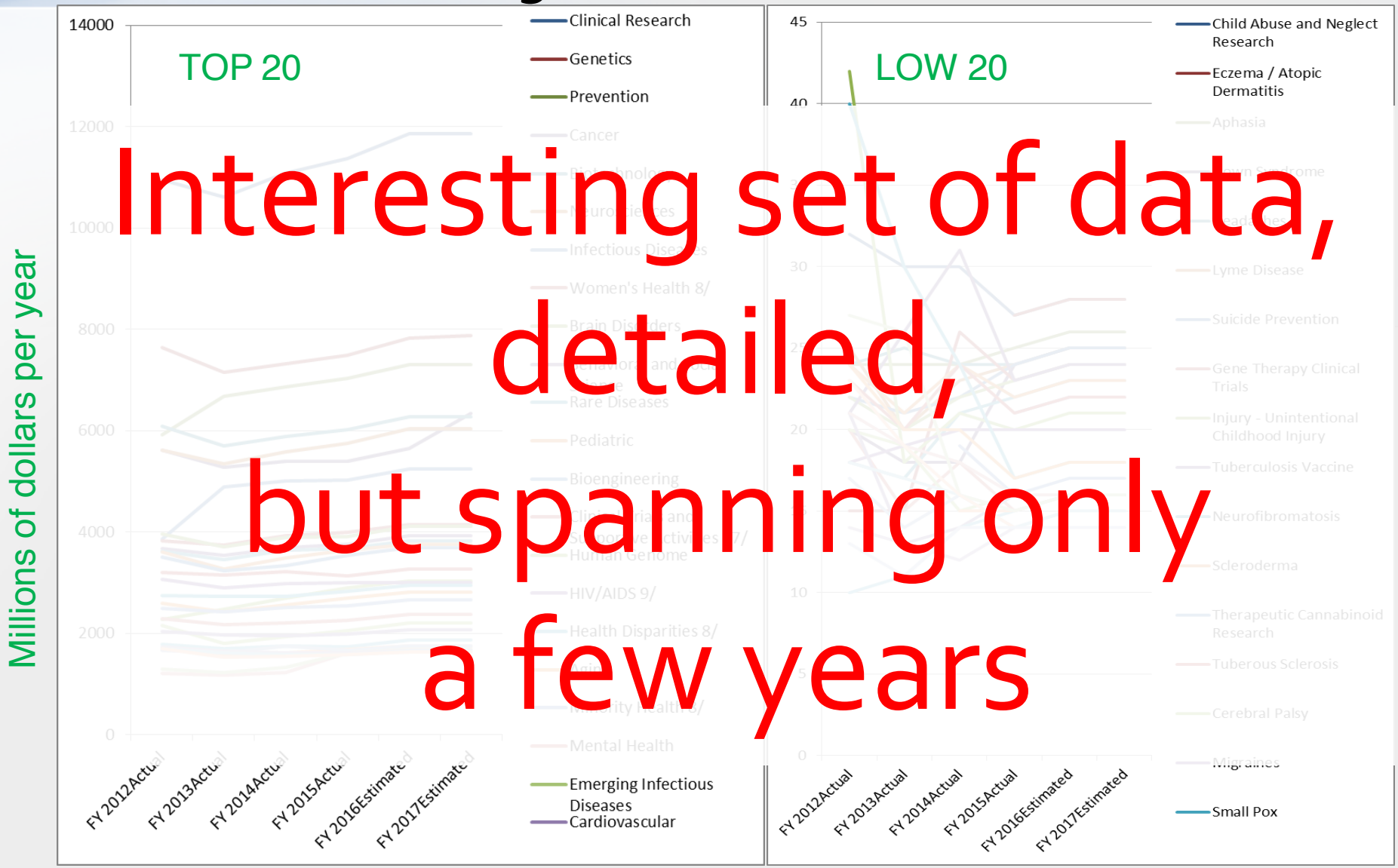
Swiss funding by discipline (# of grants)

http://p3.snf.ch/Default.aspx?id=grantsbydiscipline



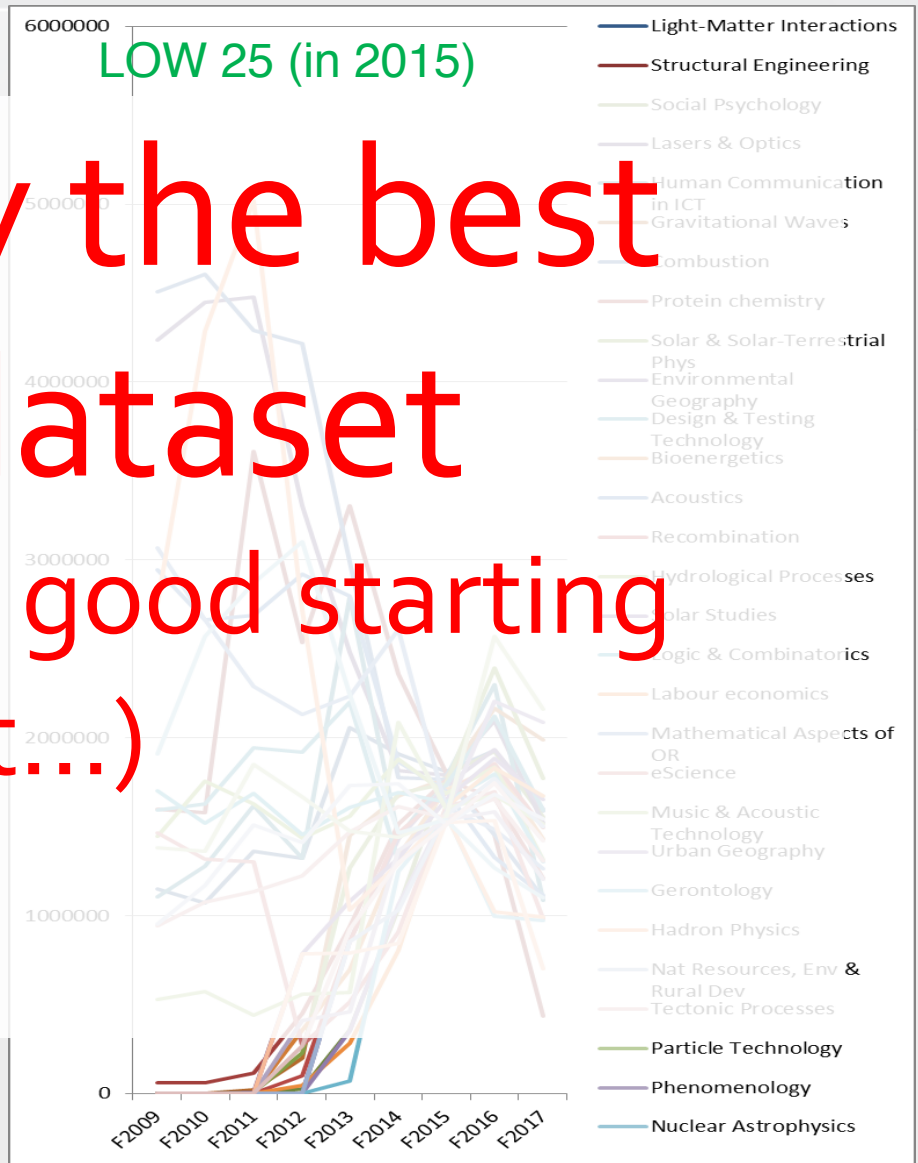
Interesting set of data, but, again, too coarse

NIH Funding for Various Research, Condition, and Disease Categories



Interesting set of data, detailed, but spanning only a few years

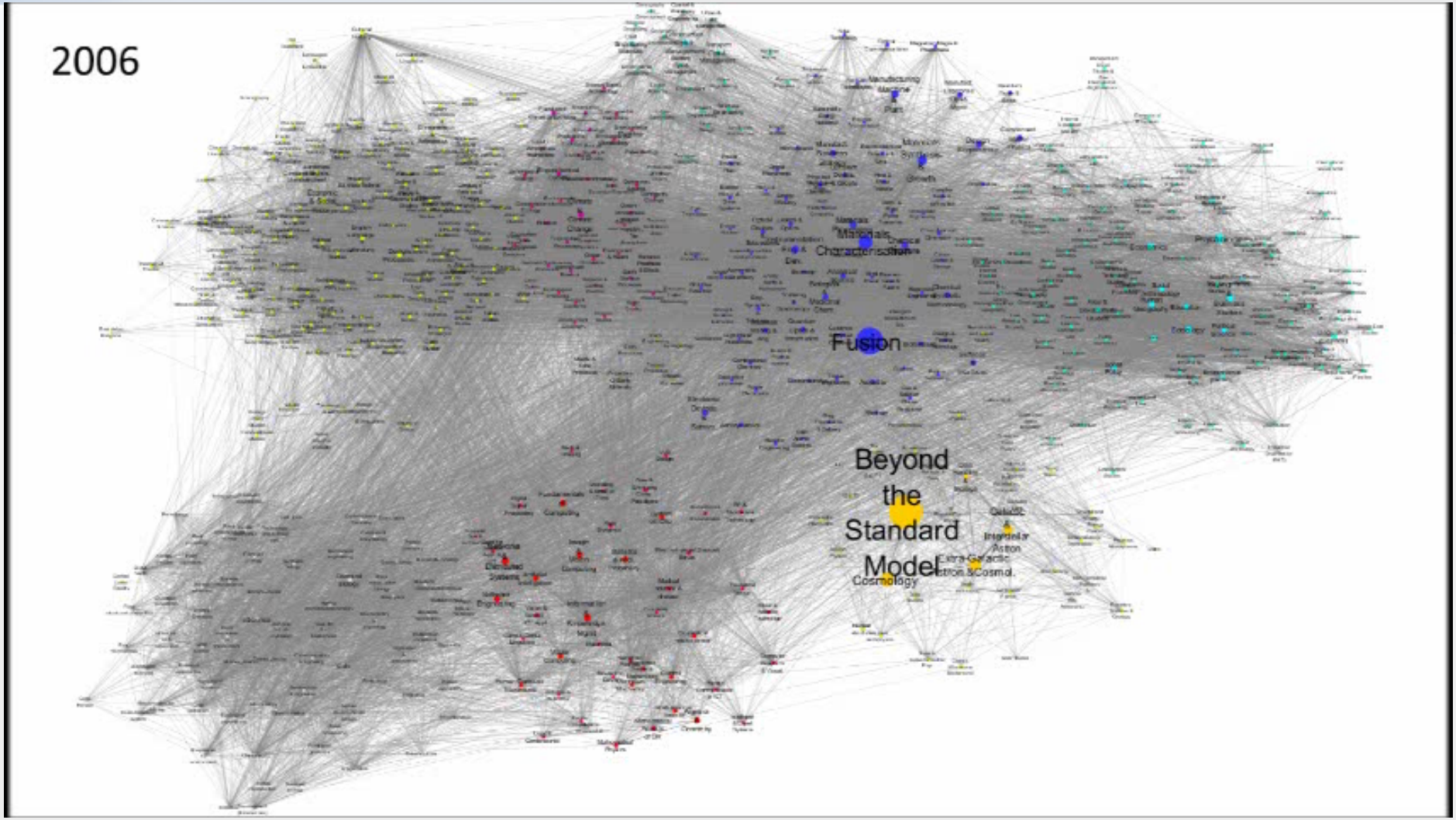
UK GtR funding by topic and year



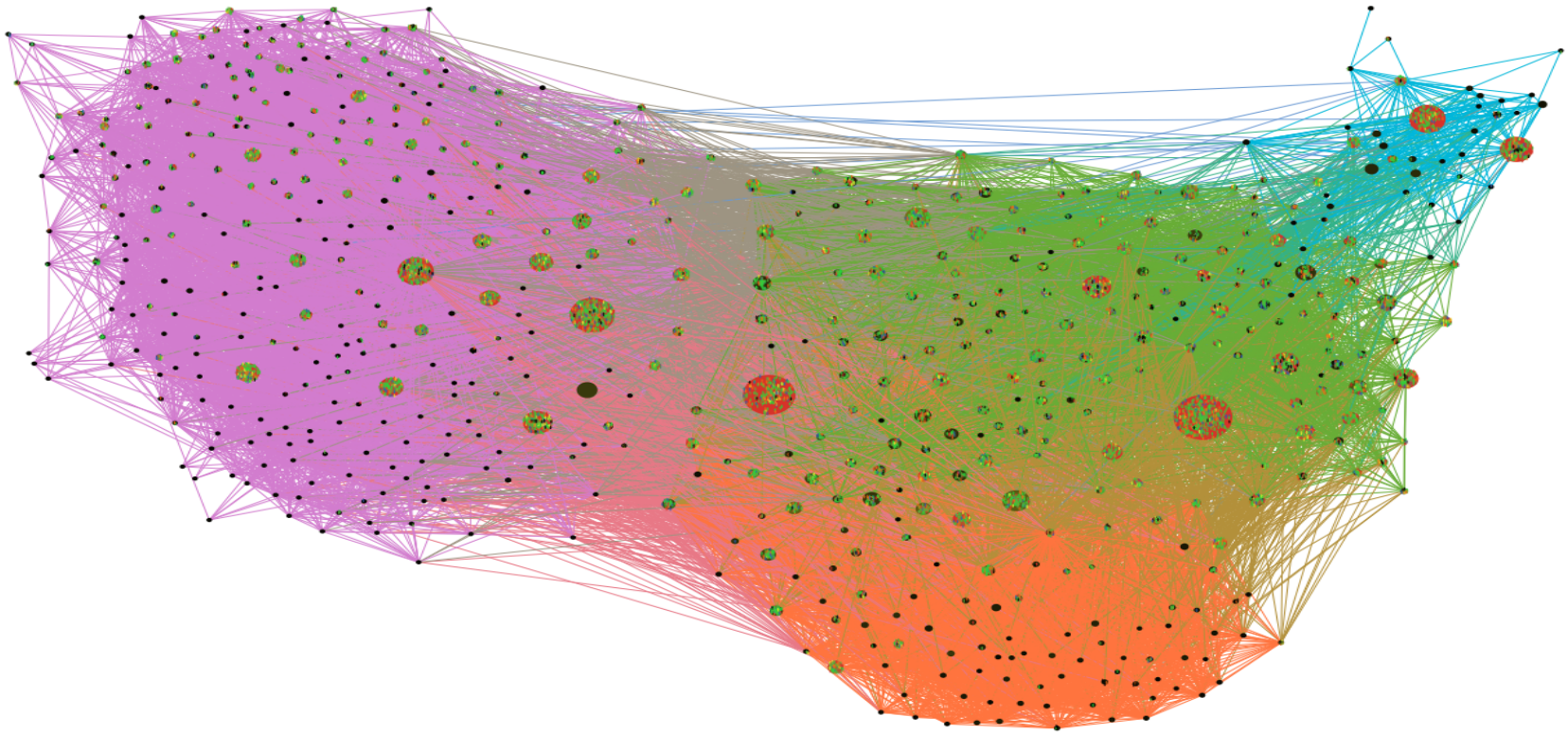
Potentially the best public dataset

(not ideal, but a good starting point…)

Combining network and temporal funding data UK GtR dataset



Next steps
(from the toy model to a realistic one)



- **The data is somewhere (funding agencies in particular)… but for some reason it seems to be treated as secret. Why? Secret from whom?**