# The complexity of knowledge production

Christopher Watts

COST TD1210 workshop

Warsaw, 9-10th November 2016

Disclaimer

- 3 days to deadline is rather short notice

- I'm an agent-based simulation modeller, not a statistician or network analyst

- Science modelling is a hobby
  - Funded by COST Actions!
- Not my day job (land-use models, at present)

- But I do have an interest in innovation

- What follows, therefore, is very preliminary…
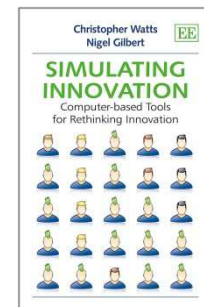
Thanks to my colleagues

The location of knowledge and know how production

- Where do knowledge production capabilities reside?
  - In institutions?
    - Facilities, Social networks, Organisational structure
  - In researchers?
    - Skills, Experience
- Policy implications
  - Metrics assessing institutions assume one thing
  - Institutions hiring staff assume another

- How easy is it for capabilities to be transferred?
  - Without also transferring the facilities, colleagues, culture, stimulations and motivations …

Cesar Hidalgo, information and economic complexity

- As seen in last year's KNOWeSCAPE meeting!

- Hidalgo, C. A. (2015). Why information grows : the evolution of order, from atoms to economies.

- Hidalgo, C. A., Klinger, B., Barabási, A.-L., & Hausmann, R. (2007). The Product Space Conditions the Development of Nations. Science, 317(5837), 482-487. doi:10.1126/science.1144581
- Hidalgo, C. A., & Hausmann, R. (2009). The building blocks of economic complexity. Proceedings of the National Academy of Sciences, 106(26), 10570-10575. doi:10.1073/pnas.0900943106
- Hausmann, R., & Hidalgo, C. (2011). The network structure of economic output. Journal of Economic Growth, 16(4), 309-342. doi:10.1007/s10887-011-9071-4
- Bustos, S., Gomez, C., Hausmann, R., & Hidalgo, C. A. (2012). The Dynamics of Nestedness Predicts the Evolution of Industrial Ecosystems. Plos One, 7(11), e49393. doi:10.1371/journal.pone.0049393

Hidalgo's concept of knowledge and know how

- Knowledge and know how are contained in human brains, and expressed in goods, texts, behaviour etc.
    - Crystallised imagination
    - Tacit knowledge (M. Polanyi)

- Although they are information, they cannot be copied easily
    - Training and experience take time
    - Communicating and interpretation are non-trivial
    - They are context dependent, especially social

Hidalgo's concept of knowledge complexity

- Some forms of knowledge are more complex, and hence harder to acquire, than others

- There is a limit to how much can be stored in one person (the "personbyte")
  - So complex knowledge may have to be spread over multiple people

- There are similarities between some forms of knowledge that make it easier for a knower of one form to learn another
  - So expect clustering, or co-location, in knowledge forms

From economies and exports to universities and research outputs

Economic complexity	?Research complexity?
Input-Output trade data	E.g. REF data, Publication data, Patents
Countries	Universities & other HE research institutions
Products for export	Papers, Patents, Graduates
Plot GDP per capita against economic complexity	Plot quality of knowledge production against research complexity
Use countries' relative economic complexity to explain economic growth	Explain dynamics in research output E.g. compare REF2014 with RAE2008

THE UK'S RESEARCH EXCELLENCE FRAMEWORK (REF) 2014

UK's 2014 Research Excellence Framework (REF)

- Successor to the Research Assessment Exercises (RAE) of 2008, 2001, ... 1986
- Higher Education institutions submit evidence of research outputs (publications), environment and impact
  - Up to 4 publications per researcher
  - Controlled by the institution (e.g. Head of Department)
- Grouped by unit of submission (subject area)
  - E.g. "Business & Management Studies"
- Evaluated by panels of peer experts
  - About 15 minutes per paper
- Used to determine allocation of £1.57 billion (2016/17) in funds
- Expensive!
  - £246 million, up from £66 million for RAE2008
  - Researchers' administration time
 - Disruption...

REF2014 data: more specifically

- Links from 154 institutions to 36 units of assessment (subject areas)

- For each institution-unit link we have:
  - % work that was rated 4*, 3*, 2*, 1* or U in Output, Environment, Impact and Overall
  - Multiple publications submitted as outputs
    - Papers, chapters, books
    - DOI and ISBN
  - (Information on research environment)
    - # staff as # Full-time equivalent (FTE)
  - (Case reports of impact)

Critiqueing the 2014 REF

- 2016: Nicholas Stern's review of the REF
  - https://www.gov.uk/government/publications/research-excellence-framework-review
  - Hobbs, F. D. R., & Roberts, L. M. (2016). The Stern review of the Research Excellence Framework. BMJ, 354.
  - http://www.nature.com/news/major-review-calls-time-on-gaming-in-uk-research-assessment-1.20343

- Mentions the possibility of gaming the system
- Recommends limiting publications to those published while employed at the institution

Nestedness, according to REF2014
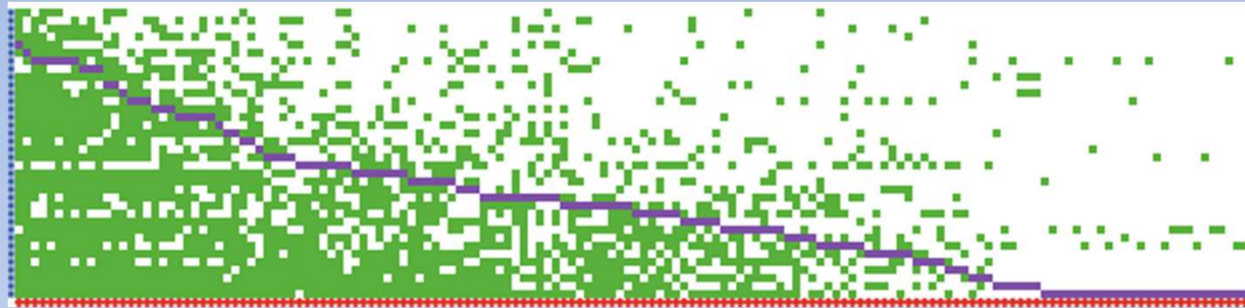
- The idea of nestedness
  - Least diversified institutions offer a subset of the units offered by the most diversified
  - Least common units to be offered almost exclusively by the most diverse institutions
- Exceptions:
  - Specialists: Music, Art, Agricultural colleges
  - Science/Social studies: Imperial College, LSE, SOAS
  - "Less academically challenging": Sport
 - Medicine: Offered by medical schools and top universities

Complexity and the method of reflections

- Economic complexity = "a measure that includes the average diversity of the countries that export the products exported by countries that export what an economy exports" (Hidalgo, 2015, p.157)

- Calculations based on
  - the *diversity* offered by institutions (# units submitted)
  - the *ubiquity* of units (# institutions offering them)
- Bipartite network metrics

M_{iu} : Institution i submitted research in unit u

$$k_{i,N} = \frac{1}{k_{i,0}} \sum_u M_{i,u} k_{N-1}$$

$$k_{u,N} = \frac{1}{k_{u,0}} \sum_i M_{i,u} k_{N-1}$$

$$k_{i,0} = \sum_u M_{i,u}$$

$$k_{u,0} = \sum_i M_{i,u}$$

Comparing complexity and quality

- Quality correlates positively with $k_{I,0}$
- Negatively with $k_{I,1}$, $k_{I,2}$ and $k_{I,3}$
- ...if various groups are excluded:
  - The low-diversity, specialist colleges: medical schools, music, art, agriculture
  - Imperial College, LSE
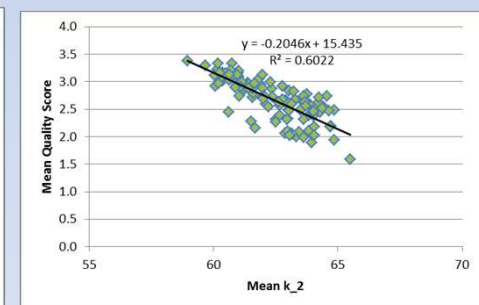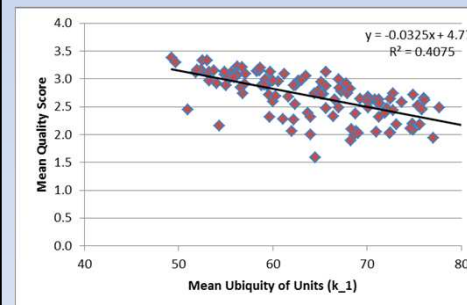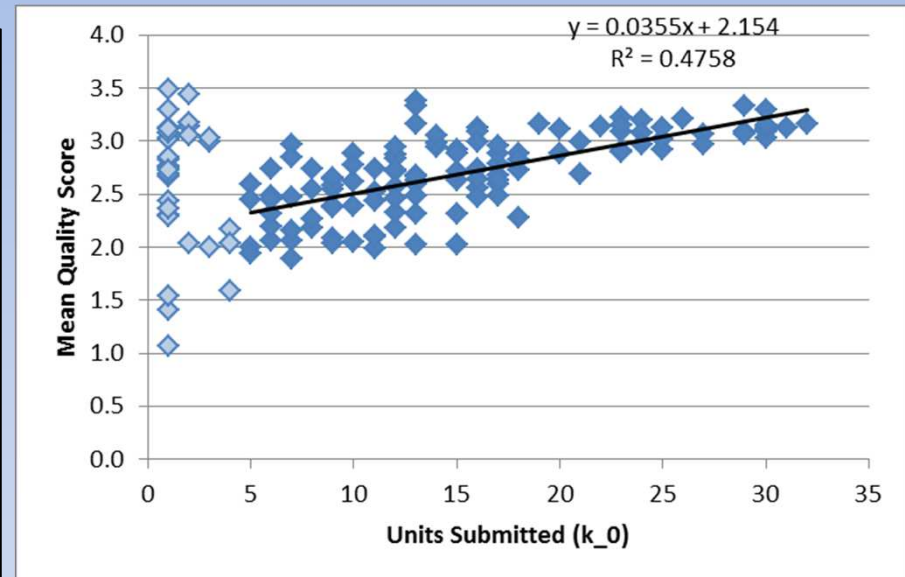
Comparing RAE2008 with REF2014

- RAE2008 data also show nestedness
- RAE2008 complexity metrics and quality rank correlate well with their REF2014 equivalents

- But neither set of complexity metrics correlates with *improvement* in quality rank
  - Contrast Hidalgo's explanation of economic growth

STAFF TURNOVER,
OR GAMING THE SYSTEM

Christopher Watts

Tracing the authors of REF submissions

- For each REF submission institution and unit (subject area):
  - Use DOI to trace submitted papers
  - Did any of their authors claim affiliation to the institution at time of publication?
  - Are any of the authors employed there according to the current (November 2016) staff list on the website?
  - Record numbers of authors Affiliated & Employed (A_E), Affiliated & Not employed (A_NE), Not affiliated & Employed (NA_E), Not affiliated & Not employed (NA_NE)

- Examined 20 randomly chosen papers
  - So no chapters or books

REF submissions as evidence of knowledge production capabilities

- Estimated three "sins" based on proportions of the papers submitted
  - No author-institution link found: All authors NA_NE
    - The authors did their work elsewhere, came for a short while, and have already gone
    - Institution gaming the system?
    - Submissions from junior staff on temporary contracts?
    - Institution has had high staff turnover?
  - No authors affiliated: A_E=0 & A_NE=0
    - The authors did their work elsewhere
    - The data tell us nothing about the institution's past production capabilities
  - No authors still employed: A_E=0 & NA_E=0
    - The authors did not stay
    - The data tell us nothing about current production capabilities

Preliminary results: statistics from 20 random papers

Institution	Unit	No author-institution link	No authors affiliated	No authors still employed
Warwick	Business & Mgt. Studies	25%	57%	57%
Loughborough	Business & Mgt. Studies	5%	35%	10%
Surrey	Business & Mgt. Studies	25%	60%	40%
Surrey	Sociology	5%	10%	15%

Why it matters for funding councils

- REF2014 is supposed to determine share of funds from Research Councils
- RCs not funding knowledge production, but the appearance of production

- REF encouraging churn and *disruption* of production
- Funding salary inflation
  - Zero-sum game, Winner-take-all competition
  - Funds earned from last RAE/REF used to attract staff and next REF-based funds, and so on…

Questions?

- Are REF submissions data suitable for Hidalgo's method of reflections?
  - Is (/Was) there a better suggestion for data?
- Does the "research complexity" concept make as much sense as Hidalgo's "economic complexity"?
  - If not, what's wrong with it?
- Why doesn't improvement in quality rank correlate with research complexity metrics?

- Is it worth persevering with the author-trace work?
  - Could we automate this?
  - E.g. bibliometric database with affiliations + Current staff lists
- Could we use it to test the transferability of researchers?

More radical questions (Shh! Don't ask these!)

- Is past performance that good a guide to future success?
  - Self-fulfilling prophesy, because we make it so with our funding
  - Perpetuating a status hierarchy, not maximising ROI

- Does giving the most resources in response to the best outputs represent the best return on investment?
  - Entry barriers, lack of economies of scale may be holding back less well funded institutions
  - Growing institutions hit traffic, communications and hiring bottlenecks
  - Does UK PLC benefit most from giving more resources to Oxford and Cambridge?
    - Contrast the more regional distribution in Germany
    - Better synergies and knowledge exchange with business to be had elsewhere and at lower levels?