



---

Matthias Meyer (2011)

## Bibliometrics, Stylized Facts and the Way Ahead: How to Build Good Social Simulation Models of Science?

*Journal of Artificial Societies and Social Simulation* 14 (4) 4  
<<http://jasss.soc.surrey.ac.uk/14/4/4.html>>

Received: 20-Jun-2011 Accepted: 26-Jul-2011 Published: 31-Oct-2011

---

### Abstract

This paper discusses how stylized facts derived from bibliometric studies can be used to build social simulation models of science. Based on a list of six stylized facts of science it illustrates how they can be brought into play to consolidate and direct research. Moreover, it discusses challenges such a stylized facts based approach of modeling science has to solve.

Keywords:

Bibliometrics, Stylized Facts, Methodology, Model Comparison, Validation

---

### Introduction

- 1.1 For some time now, sociologists - and more recently also economists - have realized that science provides an excellent object of study (Hands 2001; Gilbert 1997). Not only do they have an intuitive connection to the object as they are scientists themselves. Science also leaves many traces such as publications, citations and biographies that make it an excellent test-bed for both probing theories and modeling (Gilbert 1997). Simulation models seem to be particularly attractive because of their capacity to examine highly relevant aspects, for instance processes, networks, heterogeneous actors and cognition.
- 1.2 By now, a considerable amount of simulation models of the science system exists. Therefore, consolidating and focusing these efforts seems to be worthwhile. Among the many questions such a process has to address are: What has been achieved by the different models published in the different streams of literature? What are the remaining issues with promise for further investigation? How to assess whether and in which respect one model is better than others?
- 1.3 This paper discusses how stylized facts derived from bibliometric studies could help answering these questions. Stylized facts are broad but not necessarily universal generalizations of empirical observations and describe essential characteristics of a phenomenon that require an explanation (Kaldor 1968; Meyer 2008). I will argue that the benefit of using stylized facts to construct and assess models of science applies not only to a single research project (see Gilbert 1997 for an example). In addition, stylized facts can also be drawn on for combining and directing the modeling efforts of a community of researchers building models of science.
- 1.4 The remainder of this paper is structured as follows. In the next section, I will present some candidates for stylized facts in science that can be derived from bibliometric studies. Then I will explain how stylized facts of science can be used for merging and guiding research especially within a community of researchers. On this basis, it will be discussed how a future path might look and which challenges might have to be overcome in the future.

### Bibliometrics and stylized facts in science

- 2.1 With respect to empirical data about science, researchers are in quite a fortunate position. Several data sources exist to gather information about scientific publications and citations, e.g. Science Citation Index, Scopus and Google Scholar. While the Science Citation Index has been considered as the gold standard for a long time, the latter have become interesting alternatives in terms of quality and scope. Bibliometric methods use the information from such data sources as a starting point and analyze them. Among the most prominent approaches are publication analyses, citation analyses and co-citation analyses. They yield

information about the most prolific or influential authors in a field or about structural properties such as the level of differentiation or the existence of invisible colleges (for an application to the field of social simulation see Meyer, Lorscheid and Troitzsch 2009 and Meyer, Zaggl and Carley 2011).

- 2.2 Studying publication and citation data has revealed several recurring patterns. They are sometimes called "bibliometric laws" or "stylized facts of science" due to the repetition. Without claiming to be exhaustive, among the most important are:
1. Lotka's law describing the number of authors contributing a particular number of papers to a field or journal (Lotka 1926; Simon 1957).
  2. Matthew effect: Prominent scientists receive disproportionately more credit than less eminent researchers (Merton 1968). This phenomenon can also be observed for citations in papers or journals.
  3. Exponential growth of the number of scientists and journals (de Solla Price 1963; Gilbert 1997).
  4. Invisible colleges of specialties for every 100 scientists (de Solla Price 1963).
  5. Short half-life of literature with recent literature being cited much more frequently. For the natural sciences, the time frame is about 5 years (Burton and Kebler 1960; Umstätter, Rehm and Dorogi 1982).
  6. Bradford's law of scattering describing the distribution of articles on a given subject across journals, again showing a high concentration on a few core journals (Cole 1993).

These stylized facts are "natural" targets for social simulation models of science in general and agent-based modeling in particular (Sun and Naveh 2009). They can be understood as global patterns in science and the question is how to explain them bottom-up through micro-behavior.



## Using stylized facts of science to consolidate and direct research

- 3.1 Several scientists have recognized the opportunity for deriving simulation model targets from these patterns (e.g. Gilbert 1997, Saam and Reiter 1999 and Sun and Naveh 2009). It should be noted, that no paper has claimed to predict or mimic actual data. They simulate the processes at work that lead to some of the main aspects of the phenomena observed. Hence, the output of their simulation models is validated by comparing it with the supposed stylized facts of science.
- 3.2 While my approach is similar, it differs in one aspect: It uses stylized facts of science to consolidate and direct the research efforts of an entire community of researchers (see Heine, Meyer and Strangfeld 2005; Meyer 2008). They provide an excellent reference point when building good social simulation models of science. This can be achieved at two main levels.
- 3.3 First, one can use a list of stylized facts to assess the state of the debate by examining which stylized facts the different models have reproduced. From this perspective, the contribution of a respective model lies in its ability to explain the stylized facts of a phenomenon. A model that can at least partially explain a stylized fact is therefore regarded as more valuable than one that is geared toward side issues. If there are several stylized facts, a model's merit increases with the number of stylized facts it is able to explain alongside the absence of contradictions relating to the unexplained stylized facts. Figure 1 illustrates the basic approach for the bibliometric laws mentioned above. In addition, it shows how this approach easily identifies issues left to be addressed in the future. In our example several stylized facts still need to be addressed by future research. Ideally, there are in the future models reproducing all stylized facts. Given that the list of stylized facts does not claim to be exhaustive, identifying further stylized facts would provide additional objectives for future research.



merely shifts from the model level to the level of stylized facts. Systematically deriving stylized facts from existing empirical evidence could also lead to refining our knowledge about these recurring patterns. This is how early statements about Lotka's law have been advanced by subsequent empirical investigations (Pao 1985). Similarly, one might find differences between fields. The half-time of literature might be longer for the social sciences than for the natural sciences. Such differences in themselves might be interesting aspects to be investigated in future simulation studies.

- 4.3 When a list of accepted and empirically well grounded stylized facts exists, several issues can be tackled. One element to be addressed is the question about the appropriate unit of analysis in social simulation models of science: scientists or papers. In Gilbert's model, papers represent the basic unit of analysis, whereas Saam and Reiter's model features individual scientists. Sun and Naveh are positioned in the middle between the other two. I believe that looking more closely "under the hood" (Hausman 1995) of the different models by using the stylized facts will help answer this question, because it focuses model analysis. Similarly, the related question about the elements to be included in such models can be addressed from a stylized facts perspective when assessing the contributions of more "realistic" models incorporating cognition and networks in terms of their added explanatory value. Stylized facts might even become a useful reference point for reducing the complexity of models by establishing which model elements can be simplified without losing the ability to reproduce a given set of stylized facts (Sun and Naveh 2009).
- 4.4 Besides the many possible benefits using a set of stylized facts as a reference point for a community of researchers, some challenges still exist. A first potential limitation for social simulation studies based on bibliometric information is the data quality of the different sources poses. It has been reported to be quite poor for SCI (Meyer, Lorscheid and Troitzsch 2009). Similarly, the data provided by Google scholar often requires further refinement. Another well-known problem of SCI data is that it only provides the name of the first author of references. This can make it difficult examining citation behavior and addressing questions such as whether scientists act based on direct and indirect reciprocity when making citations.
- 4.5 A second issue stems from the fact that stylized facts typically refer to macro-data. Therefore, currently stylized facts of science cannot be used for direct validation at the micro-level (only in the indirect way described above). If stylized facts were derived for the micro-level of the models as well, they would not be used for output validation but for input validation. Introducing an appropriate terminology would reduce confusion and help to distinguish between the two types of stylized facts with respect to their uses (e.g. stylized facts used for input validation versus stylized facts used for output validation).
- 4.6 A final caveat refers to an overly dogmatic and static application of stylized facts. Used in such a way for guiding the research efforts of a community stylized facts can be perceived as a methodological straightjacket and they might even hamper progress. Not all interesting aspects of a phenomenon are necessarily captured by the "accepted list of stylized facts". Similarly, models might produce some results which pinpoint new aspects of a phenomenon that have yet to be considered. This should remind researchers to subject stylized facts to scientific scrutiny and revision if necessary in the same way that scientific theories are.



## Notes

<sup>1</sup>Please note that the mapping of stylized facts to the respective models is mainly for illustration purposes and is solely based on the claims of the authors in the quoted papers. Moreover, other valid criteria to assess the quality of models do exist beyond the ability to reproduce stylized facts (e.g. Burton and Obel 1995).

<sup>2</sup>They call this model validation strategy "pattern-oriented modeling", but it follows a very similar basic logic.



## References

- BURTON, R and OBEL, B (1995). The Validity of Computational Models in Organization Science: From Model Realism to Purpose of the Model. *Computational and Mathematical Organization Theory*, Vol. 1, No. 1, pp. 57-71. [doi:10.1007/BF01307828]
- BURTON, R E and KEBLER, R W (1960). The "half life" of some scientific and technical literatures. *American documentation*, Vol. 11, No. 1, pp. 18-22. [doi:10.1002/asi.5090110105]
- COLE, P F (1993). A new look at reference scattering. *Journal of Documentation*, Vol. 18, No. 2, pp. 58-64. [doi:10.1108/eb026315]
- DE SOLLA PRICE, D J (1963). *Little science, big science... and beyond*. New York: Columbia University Press.
- GILBERT, N (1997). A Simulation of the Structure of Academic Science. *Sociological Research Online*, Vol. 2, No. 2. <http://www.socresonline.org.uk/2/2/3.html> [doi:10.5153/sro.85]
- GRIMM, V, REVILLA, E, BERGER, U, JELTSCH, F, MOOIJ, W M, RAILSBACK, S F, THULKE, H-H, WEINER, J, WIEGAND, T and DEANGELIS, D L (2005). Pattern-Oriented Modeling of Agent-Based Complex Systems: Lessons from Ecology. *Science*, Vol. 310, No., pp. 987-991. [doi:10.1126/science.1116681]

- HANDS, D W (2001). *Reflection without Rules: Economic Methodology and Contemporary Science Theory*. Cambridge: Cambridge Univ. Press. [doi:10.1017/CBO9780511612602]
- HAUSMAN, D M (1995). Why Look Under the Hood? In Hausman, D M (Ed.): *The Philosophy of Economics: an Anthology*. 2. ed., reprint. Cambridge: Cambridge Univ. Press, pp. 217-222.
- HEINE, B O, MEYER, M and STRANGFELD, O (2005). Stylised facts and the contribution of simulation to the economic analysis of budgeting. *Journal of Artificial Societies and Social Simulation*, 8 (4) 4 <http://jasss.soc.surrey.ac.uk/8/4/4.html>
- KALDOR, N (1968). Capital Accumulation and Economic Growth. In Lutz, F A and Hague, D C (Ed.): *The Theory of Capital*. Reprint. London: Macmillan, pp. 177-222.
- LOTKA, A J (1926). The frequency distribution of scientific productivity. *Journal of Washington Academy Sciences*, Vol. 16, No. 12, pp. 317-324.
- MERTON, R K (1968). The Matthew effect in science. *Science*, Vol. 159, No. 3810, pp. 56-63. [doi:10.1126/science.159.3810.56]
- MEYER, M (2008). Stylized Facts as a Solid Basis for Validating Simulation Models? - Some Epistemological Reflections. EPOS 2008. Lisbon.
- MEYER, M, LORSCHIED, I and TROITZSCH, K G (2009). The Development of Social Simulation as Reflected in the First Ten Years of JASSS: A Citation and Co-Citation Analysis. *Journal of Artificial Societies and Social Simulation*, 4 (12) 12. <http://jasss.soc.surrey.ac.uk/12/4/12.html>
- MEYER, M, ZAGGL, M A and CARLEY, K M (2011). Measuring CMOT's intellectual structure and its development. *Computational and Mathematical Organization Theory*, Vol. 17, No. 1, pp. 1-34. [doi:10.1007/s10588-010-9076-0]
- PAO, M L (1985). Lotka's law: A testing procedure. Vol. 21, No. 4, pp. 305-320. [doi:10.1016/0306-4573(85)90055-x]
- SAAM, N S and REITER, S (1999). Lotka's law reconsidered: The evolution of publication and citation distributions in scientific fields. *Scientometrics*, Vol. 44, No. 2, pp. 135-155. [doi:10.1007/BF02457376]
- SIMON, H A (1957). *Models of man, social and rational*. New York: Wiley.
- SUN, R and NAVEH, I (2009). Cognitive simulation of academic science. International Joint Conference on Neural Networks: 3011-3017. [doi:10.1109/ijcnn.2009.5178638]
- UMSTÄTTER, W, REHM, M and DOROGI, Z (1982). Die Halbwertszeit in der naturwissenschaftlichen Literatur. *Nachrichten für Dokumentation*, Vol. 33, No. 2, pp. 50-52.