Diagnosing Emerging Science: The Cases "New Science of Networks" and Scientometrics

Haiko H. Lietz

lietz@forschungsinfo.de

Institute for Research Information and Quality Assurance, Bonn, Germany

Competence Centre for Bibliometrics, Bonn, Germany

University of Duisburg-Essen, Faculty of Social Sciences, Duisburg, Germany

Hochschule Mittweida, Faculty of Mathematics, Physics and Computer Science, Germany

January 31, 2010

Abstract - What is emerging science, and how can it be measured if a field, sub-field, or subject area is emerging? Often emerging science is diagnosed as a research front using citation analysis. Bettencourt et al. employ collaboration analysis and concentrate on structural properties of the process of emergence itself. According to the model the establishment of a paradigm in a field shows as a topological transition in its social structure. In this paper the model will be applied to the "New Science of Networks" and the Field of Scientometrics. Differences in their evolutionary processes show as expected. Model and methods of network and scaling analysis are discussed against the background of science studies, percolation theory, and relational sociology. A definition of emerging science, based on social structural concepts, is given. Special attention is paid to the self-similarity of the science system. Effects of using different counting methods on the results are also discussed.

Keywords - New Science of Networks, Field of Scientometrics, emerging science, co-authorship networks, scaling analysis

1 Introduction

What is emerging science, and how can it be measured if a field, sub-field, or subject area is emerging? Contributions to answers can be found in the classics of science studies, most visibly in de Solla Price [1963] who studied growth dynamics of science and coined the term research front. A research front as an identifiable part of emerging science is the small set of scientific publications – and its associated authors – of a field which cite each other more strongly than publications not belonging to the set. Another approach is by Kuhn [1962] who described the *modus operandi* of normal science and occasional paradigm shifts where a new order emerges. The prevailing order guides scientists in their daily work, for example in their citation behavior [Small, 1980], which can again hint at research fronts. Frequently emerging science has been diagnosed using various methods of citation analysis [Persson, 1994, White and McCain, 1998, Lazer et al., 2009, Shibata et al., 2009], but also semantic methods [Courtial, 1994], mapping [Börner and Scharnhorst, 2009], and combinations of methods [Chen, 2005, Besselaar and Heimeriks, 2006] have been employed.

Recently, Bettencourt et al. [2006, 2008, 2009] have made an innovative contribution to answering the introductory questions. They have shown for various fields that their exponential or logistic growth can well be described by a diffusion model [2006, 2008] and that they undergo a topological transition in their collaboration structure on their way to maturity. This transition can be measured in terms of a densification of a field's co-authorship network and, therein, the emergence of a giant component, in which a large part of the entire network's authors collaborate at least indirectly [Bettencourt et al., 2009]. Theoretically this transition in the social structure can be understood as the establishment of a paradigm and, as such, a dynamic on a level of meaning. Attention is therefore placed on the structural diagnosis of the process of emergence itself, rather than the detection of research fronts. This approach is innovative because science is diagnosed using collaboration analysis instead of citation analysis while conceptionally connecting social action and meaning. It is Bettencourt et al.'s merit to have brushed up de Solla Price's quantitative work using today's technical possibilities and to have tied it to Kuhn's theory.

In this paper the model is applied to two further fields for which different results are expected. The "New Science of Network" (NSoN) deals with the analysis of complex networks, is home in physics, mathematics, and computer science, and has received much attention [Freeman, 2004, Bonacich, 2004] since the publication of two influential papers [Watts and Strogatz, 1998, Barabási and Albert, 1999]. Citation analyses [Chen, 2005, Shibata et al., 2007, 2008, Lazer et al., 2009] give reasons to expect that the NSoN is based on a coherent research concept, which should be identifiable in terms of a giant component in its social structure. The second case is scientometrics. This field of information science studies the science system mainly using bibliometric methods [White and McCain, 1989, 1998, Besselaar and Heimeriks, 2006]. Bibliometric studies [Schubert and Maczelka, 1993, Courtial, 1994, Persson, 1994, Wouters and Leydesdorff, 1994] have painted a diffuse picture of the field. Based on these studies and expert judgements [Glänzel and Schoepflin, 1994, Larsen, 2008] it is not expected that a giant component can be identified in its social structure.

The paper is structured as follows: In the following, second, section the model [Bettencourt et al., 2006, 2008, 2009] is laid out against the background of science studies [Goffman, 1966, Kuhn, 1962, de Solla Price, 1963], percolation theory [Wilson, 1979, Newman, 2005], and relational sociology [White, 2008]. The third section introduces the two cases. In the fourth section, emerging science is defined, the delineation of the fields using keyword searches in the *Web of Science*® is introduced, and the methods of network definition [Newman, 2001b,a] and scaling analysis [Leskovec et al., 2005, Bettencourt et al., 2008, 2009] are discussed. Finally, the results are presented and discussed.

2 Structure and Dynamics of Science

de Solla Price [1963] was among the first to note that two fundamental laws govern the social structure and dynamics of science. First, the publication output of a field grows exponentially until it reaches a critical point from which on the growth rate starts to decrease, constituting logistic growth. Multiple logistic growth periods can accumulate to a prolonged exponential growth period. Second, the socio-cultural structure of the science system is *self-similar*, i.e. characteristics are not defined on a certain scale. Self-similarity is generally signified by power-laws [Mandelbrot, 1982, Bak, 1996, Newman, 2005] and specially by various scaling laws, such as those of Lotka, Zipf, and Bradford.

Kuhn [1962] described an aspect of the cultural evolution of scientific fields. According to this view, cumulative knowledge production periods (normal science) inevitably lead into a crisis as unexpected results (anomalies) pile up which cannot be explained by the prevailing theoretical concept (paradigm). The paradigm shift is then the relatively sudden and abrupt event when competing interpretations that try to explain the anomalies percolate into a coherent new paradigm. Kuhn noted that paradigm shifts are not necessarily revolutions of the Newtonian kind, but can also occur on small scales [Kuhn, 1962, pp.7]. Therefore the dynamics of change of the science system are also self-similar.

From a phenomenological perspective de Solla Price's and Kuhn's accounts resemble similar dynamics, albeit on different levels. In logistic growth, the period of time in which the growth rate of a field is close to its maximum is relatively short compared to its whole lifetime. Therefore, fields that have grown logistically have at one point in time evolved rather suddenly. In comparison, paradigm shifts are non-linear processes either. Both are phenomena of *emergence* in which, generally, a new quality comes into existence by interaction of a system's components [Sawyer, 2005]. To use a frequently used phrase, the whole is more than the sum of its parts. A classical example of emergent phenomena is ice forming when temperature decreases and reaches the critical temperature of water, zero degrees centigrade at normal pressure, where molecules stop moving freely. In percolation theory such processes resemble phase transitions and logistic functions are tell-tale signatures of such phenomena. Intriguingly, the structure of many systems, such as water, is only self-similar at the critical point [Wilson, 1979, Newman, 2005]. Following this school of thought, Solla Price's "invisible colleges" emerge out of the interaction of the scientists in a field and Kuhn's paradigms emerge out of the convergence of competing interpretations to a new paradigm. Both phenomena can be interpreted as a phase transition, in which time takes the role of temperature [cf. White, 2008, pp.70].

At about the same time of de Solla Price and Kuhn, Goffman [1966] studied the field of mast cell research and showed that, measured in terms of authors and publications, it had grown logistically. He showed that the population dynamics could be described using a diffusion model where the underlying assumption is that novel ideas spread from researcher to researcher just like a virus infects members of a susceptible population. Recently, Bettencourt et al. [2008] have shown an improved diffusion model [2006] to correctly describe the (logistic) population growth of six different fields. The model is based on the assumption that a scientific idea can recruit individuals, e.g. Ph.D. students or other scientists who regard this idea as relevant for their work, which causes the personnel of the field to grow. In a follow-up paper [2009] they diagnose the social structure and dynamics of these six plus two fields using complex network analysis. They show that seven fields' collaboration networks densify and grow a giant component. The model proposes that this topological transition in the social structure from a "liquid" to a more "solid" state mirrors the diffusion of a paradigm on the cultural level of the field. Supporting this conclusion, Condensed Matter Nuclear Science, which can not offer a coherent explanation of the Cold Fusion anomaly, is neither densifying nor growing a cohesive network core.

Bettencourt et al. have developed a sociologically promising model of a socio-cultural system. With *Identity and Control*, White [2008] has developed a relational sociology which offers the necessary concepts to bridge social action and cultural frames for action. Acting units are not persons in principle, but *identities* which can be positioned on all levels of socio-cultural space-time. Identities are linked in ordered networks and collectively try to *control* otherwise chaotic environments. As they interact they form new emergent identities which interact on a higher level of complexity. Feedback is a counter process to emergence: Emergent levels feed back onto lower levels by creating path dependencies and constraining identities socially and culturally.

When a paradigm diffuses in a population, it has become a "way of practicing" [Kuhn, 1962, pp.134] science which determines "community life" [pp.94]. In this sense Kuhn transcends his definition of a paradigm as "the entire constellation of beliefs, values, techniques, and so on shard by the members of a given community." [pp.175] When it comes to its implementation, a paradigm is also a *style*, a process-related sensiblity or simply a program which is executed in self-organization [cf. White, 2008, ch.4]. On an aggregated level a style leads to a field having an own identity. The concept bridges the gap between social structures of action and cultural structures of meaning. Insofar it is plausible that the diffusion of a common understanding of research should manifest on the level of social collaboration of scientists. Introducing the Matthew Effect, Merton [1988] has proposed a mechanism how the structural self-similarity of the science system could emerge by way of an institutionalized style of concentration of resources and attention [cf. Barabási and Albert, 1999, White 2008, pp.149].

3 The Cases "New Science of Networks" and Scientometrics

3.1 "New Science of Networks" (NSoN)

For half a century and longer social networks have been studied in the *Social Network Analysis* (SNA) program in sociology and mathematics [Freeman, 2004]. The end of the 90s brought breakthroughs [Watts and Strogatz, 1998, Barabási and Albert, 1999] in the physics, mathematics, and computer science based studies of complex networks. This *Complex Network Analysis* (CNA) is different from SNA inasmuch not only relatively small social networks but – due to the availability of huge data sources and processing capabilities – large complex networks with emergent properties are studied.

Indeed the cited SNA literature can be divided into three large clusters, the youngest of which holds the CNA paradigm [Chen, 2005]. Following the year 2000, works of physicists have increasingly been picked up in sociology, which also brought classics like *small world* research to new life [Lazer et al., 2009]. In sociology this was sometimes perceived as an "invasion of the physicists" [Bonacich, 2004]. Employing citation analysis, Shibata et al. could show that this perception formed because "the activated center of research shifted from social science to physics, and there was a wall among domains that prevented bridging among domains, like between articles by sociologists and articles by physicists in CN." [Shibata et al., 2007, p.881] Having originated in SNA, CNA widely made itself independent [Shibata et al., 2008]. This suggests that a "New Science of Networks" has formed around the CNA paradigm. According to the model by Bettencourt et al. a giant component should have emerged in the field's social structure.

3.2 Field of Scientometrics (FoS)

As a field of information science, scientometrics studies the science system mainly using bibliometric methods. Classically, information retrieval is the other large field of information science [White and McCain, 1989]. This split has been demonstrated bibliometrically, at least for the representative *Journal of the American Society for Information Science* of the second half of the 80s, with respect to both citing and cited authors [Persson, 1994]. In the widest sense both fields have in common that they deal with literature [White and McCain, 1998]. Again following the year 2000, the third field web studies emerged as a sign of the increasing importance of electronic information [Besselaar and Heimeriks, 2006], a dynamic which Courtial [1994] could not have foreseen in his early co-word analysis.

For scientometrics a crystallization in the course of the 80s around the journal *Scientometrics* has been diagnosed, even though it was too early to speak of the establishment of a paradigm [Schubert and Maczelka, 1993]. Between 1978 and 1992 the field had the character of a social science, with a largely fragmented collaboration structure, a weakly differentiated semantic structure, and relatively old references. A homogenous use of citations, however, was indicative of a field's own identity [Wouters and Leydesdorff, 1994]. But the bibliometric results from the three last cited studies are only restrictedly valid because only publications in the journal *Scientometrics* have been studied. Bibliometric results that assign the FoS an own identity are contradicted by expert judgements. Particularly according to Glänzel and Schoepflin [1994] scientometrics was in a crisis in the middle of the 90s even thought the field had grown rapidly and attracted much attention: "subfields are drifting apart, the field is lacking consensus in basic questions and of internal communication, the quality of scientometric research is questioned by other disciplines." [p.375] But also in 2008 Larsen concluded that "40 years of publication counting have not resulted in general agreement on definitions of methods and terminology nor in any kind of standardization." [p.235] In sum, the emergence of a giant component in the field's social structure is not expected.

4 Definitions, Data, and Methods

4.1 Definitions

Other than existing methods where research fronts are identified using citation analyses, emerging science will be diagnosed in terms of social structure. Leaning on percolation theory, the strong definition of emerging science is: *A field, sub-field, or subject area is emerging when it is at the critical point of growing a giant component in its collaboration structure.* A weaker definition, that does not require the social structural growth rate to be close to its maximum, is: *A field, sub-field, or subject area is emerging when its collaboration structure is densifying.* An indicator which can assist the diagnosis of emerging science is the productivity of a field, sub-field, or subject area. These concepts will be operationalized in the Methods sub-section.

Collaboration structures are operationalized as co-authorship networks where a co-authorship is present if two authors have published together [Newman, 2001b]. A *component* is a maximum subset of the network, in which all authors collaborate at least indirectly. A *giant component* is that component which is markedly bigger that all other components. When such a component is present, an emergent process has taken place [Newman, 2005].

Co-authorships are counted using whole and fractional counting methods. Because the real collaboration structure of authors co-authoring a publication is unknown, it is assumed that all authors per publication collaborate. Using whole counting, all co-authorships are assigned a value of 1. Obviously, however, the value 1, being independent of the number of co-authors, does not express a communication load. To assign network edges the probabilistic meaning of a communication load, Newman [2001a] has proposed a fractional counting method. In this case all co-authorships per publication are assigned the value $1/(n\text{-}1)$, where n is the number of authors per publication. To restore the measuring scale, edges are rounded to the next integer.

4.2 Data

Fields have to be captured as complete as possible because emergent effects on the field level are to be identified. A delineation using selected, particularly important journals of a field, so called core journals, would be to restrictive because peripheral publications can considerably contribute to network connectivity. Therefore, fields were delineated using keywords which were used to identify relevant publications.

Data was extracted from the *Web of Science*® as part of the *ISI Web of Knowledge*℠, from the databases SCI-EXPANDED and SSCI in the timespan 1900-2008.

Keywords for the NSoN are:

```
Topic=(“small-world network” OR “small world network”
OR “small-world networks” OR “small world networks”
OR “scale-free network” OR “scale free network” OR “scale-free networks”
OR “scale free networks” OR “complex network” OR “complex networks”)
```

This search strategy is supposed to avoid that purely social scientific publications about “social networks” are broadly captured. 6,321 publications of the document types

Article, Bibliography, Discussion, Editorial Material, Letter, Meeting Abstract, Note, Proceedings Paper, and Review were identified.

Keywords for the FoS are:

```
Topic=(bibliometric* OR informetric* OR scientometric*
OR “citation analys*s” OR “cocitation analys*s” OR “co-citation analys*s”
OR “self citation” OR “self citations” OR self-citation OR self-citations
OR “citation map*” OR “citation visuali*” OR “collaboration network”
OR “collaboration networks” OR “coauthorship network”
OR “coauthorship networks” OR “co-authorship network”
OR “co-authorship networks” OR “journal impact factor”
OR “journal impact factors” OR h-index OR “h index” OR “Hirsch index”
OR “S&T indicator” OR “S&T indicators”)
OR [Topic=(Lotka OR Zipf OR Bradford)
AND Subject Areas=(INFORMATION SCIENCE & LIBRARY SCIENCE
OR COMPUTER SCIENCE, INFORMATION SYSTEMS)]
```

These keywords are geared to the ones used by Bar-Ilan [2008]. This search strategy captures publications both in basic and applied research. To largely reduce the field to bibliometrics, keywords like `patent analys*s` and `science policy` were excluded. The confinement of the search for `Lotka OR Zipf OR Bradford` to the disciplines `INFORMATION SCIENCE & LIBRARY SCIENCE` and `COMPUTER SCIENCE, INFORMATION SYSTEMS` led to the exclusion of irrelevant publications not belonging to the FoS. In total, 4,544 publications were identified.

Author names were hand-cleaned: Two authors with the same family name and same initial of the first given name, but one having one or more further initials of given names, were unified if co-author, affiliation, or keyword contexts were similar. Authors carrying double family names at a later point in time, e.g. due to marriage, were unified if identified. Authors whose family names carry “van” or “de” could sometimes be unified when spaces were removed. Chinese names were generally not cleaned due to time constraints. Authors carrying the same name were not checked for different identities.

4.3 Methods

The main goal of the analyses is to check if co-authorship networks densify, which would satisfy the weak definition of emerging science, and if they grow a giant component, which would satisfy the strong definition. Scaling analysis is used to check for network densification [Leskovec et al., 2005, Bettencourt et al., 2009], i.e. the number of nodes and the number of edges of single years' networks are correlated non-linearly. If the exponent α of the power-law

$$edges = A * (nodes)^{\alpha}$$

is equal to 1, the number of co-authorships per author is constant. If $\alpha > 1$ ($\alpha < 1$) there are increasing (decreasing) returns to scale and the network is (not) densifying. α is estimated using the least squares method. A is a normalization constant. The size of the giant component is the number of nodes in the largest network component.

Field productivity is analyzed using another scaling law. If the exponent β of the power-law

$$\Delta\ publications = B * (\Delta\ authors)^{\beta}$$

is equal to, or smaller than, 1 the field is not productive. For $\beta > 1$ the field is productive because the *per capita* publication output is showing increasing returns to scale. B is a normalization constant.

Networks are visualized using the *ORA Network Visualizer* which is based on the *TouchGraph* technology.

5 Results

The first publication for the NSoN was recorded in 1965, for the FoS in 1948. Counting publications, figure 1 shows that the NSoN started to grow later but grew faster than the FoS. Both fields still grow exponentially with no overall tendency to saturate. The NSoN shows two marked growth periods, measured in terms of annual publication increase, which slowed down towards their end. The first started around 1990 and lasted ten years. The second started around 2000 and seems to slow down from 2007 on. The FoS shows a first growth period starting around 1975 and lasting about 15 years. The second started around 1991 and also lasted about 15 years. A third growth burst seems to be occurring right now [cf. tab.1].

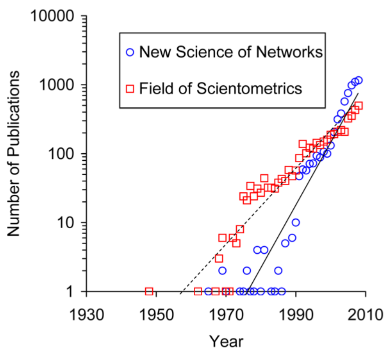

Figure 1: Growth of the "New Science of Networks" (NSoN) and the Field of Scientometrics (FoS) in publications per year.

Counting authors, 1991 was the first year that the NSoN counted more than 100 active authors. The FoS had passed this line already in 1990. In 2003, the NSoN permanently passed the 1,000 author line. The FoS counted that many authors once in 2007 but fell back to below 1,000 the next year [cf. tab.1].

Looking at productivity, both fields at first glance are not productive, with $\beta \approx 0.9$. The FoS even seems to be a little more productive. However, looking only at the years 1990-2008, when both fields counted more than approximately 100 active authors each

	"New Science of Networks"					Field of Scientometrics				
Year	Publications	Cumulative Publications	Active Authors	New Authors	Cumulative Authors	Publications	Cumulative Publications	Active Authors	New Authors	Cumulative Authors
1948						1	1	1	1	1
1962						1	2	1	1	2
1965	1	1	1	1	1					
1967						1	3	1	1	3
1968						3	6	4	4	7
1969	2	3	2	2	3	6	12	7	6	13
1970	1	4	2	2	5	1	13	1	1	14
1971						1	14	1	1	15
1972						6	20	5	5	20
1973						5	25	6	5	25
1974	1	5	1	1	6	8	33	9	8	33
1975	1	6	2	2	8	24	57	25	22	55
1976	2	8	4	4	12	21	78	22	19	74
1977	1	9	1	1	13	34	112	33	26	100
1978	1	10	2	2	15	24	136	27	23	123
1979	4	14	7	5	20	31	167	46	38	161
1980	1	15	1	1	21	27	194	39	27	188
1981	4	19	6	6	27	44	238	58	44	232
1982						32	270	32	25	257
1983	1	20	1	1	28	32	302	42	29	286
1984	1	21	2	2	30	31	333	39	28	314
1985	2	23	6	6	36	38	371	51	35	349
1986	1	24	1	1	37	43	414	56	38	387
1987	5	29	13	12	49	40	454	53	39	426
1988						58	512	78	49	475
1989	6	35	9	8	57	47	559	72	42	517
1990	10	45	19	19	76	57	616	100	76	593
1991	47	92	119	119	195	86	702	147	108	701
1992	59	151	176	176	371	138	840	210	161	862
1993	57	208	175	172	543	101	941	155	114	976
1994	71	279	223	212	755	122	1063	169	113	1089
1995	72	351	242	237	992	118	1181	202	133	1222
1996	93	444	286	271	1263	143	1324	235	168	1390
1997	88	532	256	235	1498	133	1457	225	138	1528
1998	106	638	358	345	1843	150	1607	246	165	1693
1999	100	738	356	335	2178	159	1766	277	180	1873
2000	131	869	415	394	2572	188	1954	323	237	2110
2001	204	1073	595	541	3113	192	2146	377	264	2374
2002	314	1387	860	730	3843	208	2354	393	266	2640
2003	381	1768	1034	885	4728	220	2574	402	278	2918
2004	570	2338	1531	1275	6003	207	2781	433	303	3221
2005	754	3092	1943	1518	7521	323	3104	658	479	3700
2006	976	4068	2388	1771	9292	354	3458	784	546	4246
2007	1098	5166	2646	1850	11142	428	3886	1024	760	5006
2008	1155	6321	2892	1928	13070	496	4382	996	680	5686

Table 1: Numbers of publications and authors for the "New Science of Networks" (NSoN) and the Field of Scientometrics (FoS).

year, trends reverse: The NSoN is productive ($\beta \approx 1.2$) while the FoS is not ($\beta \approx 0.8$) [cf. fig.2, tab.2].

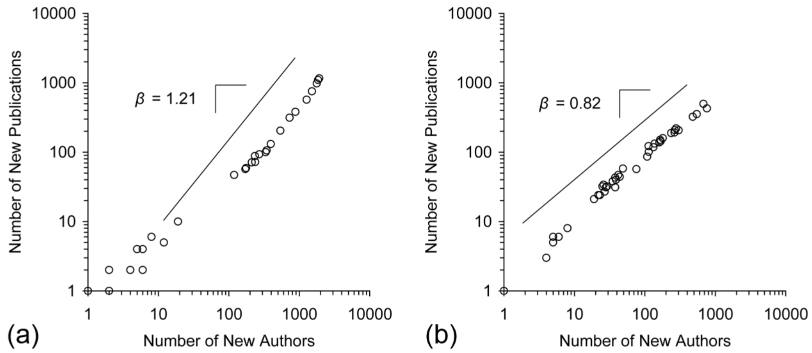

Figure 2: Productivity of the (a) "New Science of Networks" (NSoN) and the (b) Field of Scientometrics (FoS). Slopes of lines indicate scaling exponents for time periods 1990-2008.

Eye inspection of the evolution of co-authorship networks [fig.3] reveals that in the NSoN a giant component started to emerge in 2006 which held 595 of 2,892 and thus 21% of all authors active in 2008. The FoS has not grown a giant component. Only 23 of 996 active authors in 2008 form the largest component. This result is clearly confirmed by the sudden increase of the size of the largest component in figure 4a, but not in figure 5a.

"New Science of Networks"	**Field of Scientometrics**
$\alpha_{NSoN1965-2008whole} = 1.12$ ($R^2 = 0.99$)	$\alpha_{FoS1948-2008whole} = 1.10$ ($R^2 = 0.94$)
$\alpha_{NSoN1990-2008whole} = 1.08$ ($R^2 = 0.99$)	$\alpha_{FoS1990-2008whole} = 1.42$ ($R^2 = 0.96$)
$\alpha_{NSoN1965-2008fract} = 0.84$ ($R^2 = 0.98$)	$\alpha_{FoS1948-2008fract} = 0.93$ ($R^2 = 0.96$)
$\alpha_{NSoN1990-2008fract} = 1.19$ ($R^2 = 0.98$)	$\alpha_{FoS1990-2008fract} = 1.02$ ($R^2 = 0.97$)
$\beta_{NSoN1965-2008} = 0.90$ ($R^2 = 0.99$)	$\beta_{FoS1948-2008} = 0.94$ ($R^2 = 0.99$)
$\beta_{NSoN1990-2008} = 1.21$ ($R^2 = 0.99$)	$\beta_{FoS1990-2008} = 0.82$ ($R^2 = 0.97$)

Table 2: Densification and productivity indicators for different time periods (and counting methods).

Using whole counting and looking at all years, the social structure of both fields seem to be densifying, with $\alpha \approx 1.1$ [cf. tab.2]. Looking only at the growth following 1990, it turns out that the FoS densified quite heavily in that time ($\alpha \approx 1.4$), whereas the NSoN densified constantly over its whole lifetime ($\alpha \approx 1.1$). However, results are not robust. When fractional counting is used, both fields dip below 1 when all years are studied, but for the years 1990-2008 densification of the NSoN rises to its maximum ($\alpha \approx 1.2$) while the FoS hardly makes it above 1 [cf. figs.3b and 4b, tab.2].

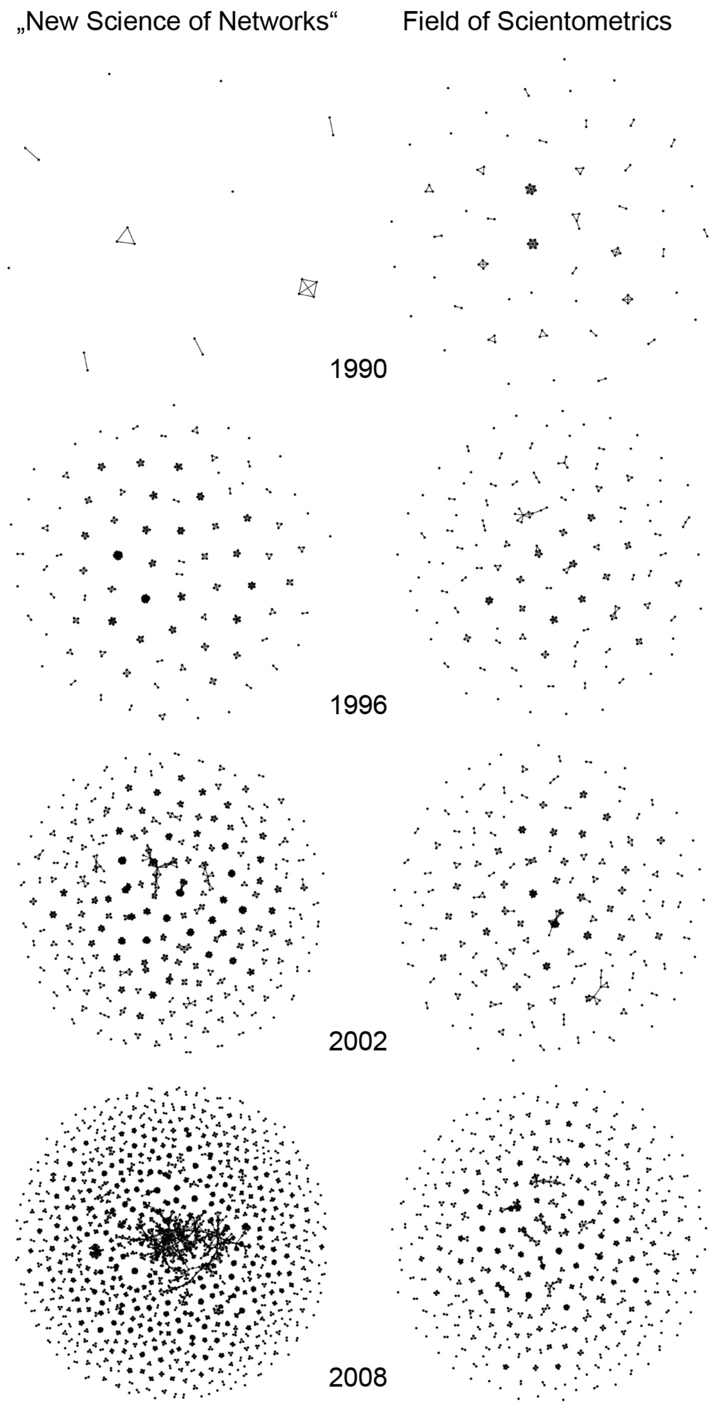

Figure 3: Temporal evolution of co-authorship networks of the "New Science of Networks" (NSoN) and the Field of Scientometrics (FoS): Only in the NSoN a giant component has emerged.

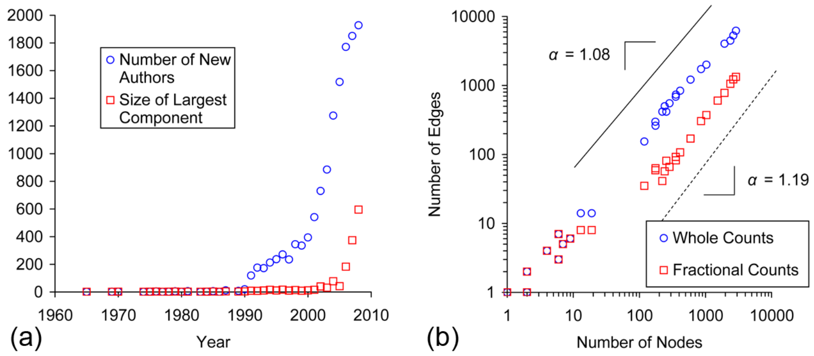

Figure 4: (a) Evolution, in authors newly entering the field and size of largest component in co-authorship network (whole counting), and (b) network densification (whole and fractional counting) of the "New Science of Networks" (NSoN). Slopes of lines indicate scaling exponents for time periods 1990-2008.

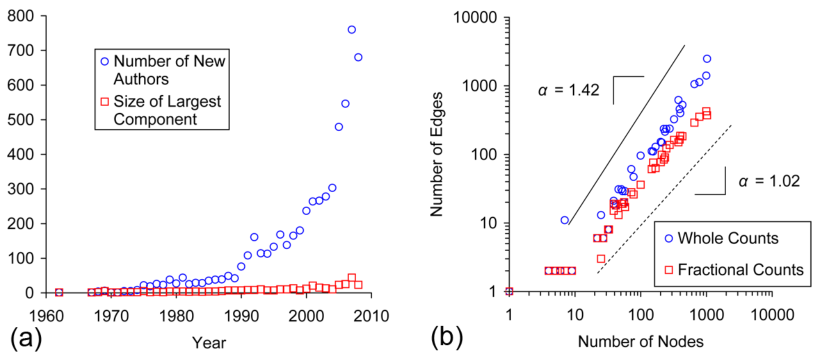

Figure 5: (a) Evolution, in authors newly entering the field and size of largest component in co-authorship network (whole counting), and (b) network densification (whole and fractional counting) of the Field of Scientometrics (FoS). Slopes of lines indicate scaling exponents for time periods 1990-2008. In (b) one outlier was removed in whole counting because a paper having 77 authors distorted the edge count.

6 Summary and Discussion

Departing from the model of Bettencourt et al. [2006, 2008, 2009], that the establishment of a paradigm in a field goes along with a transition of the field's social structure from a "liquid" to a more "solid" state, the "New Science of Networks" (NSoN) and the Field of Scientometrics (FoS) were studied. Different results were expected for the two fields. The expectations were based on previous analyses and expert judgements that the NSoN had formed relatively quickly around the *Complex Network Analysis* paradigm, whereas the FoS' establishment seems to be hampered by a lack of conceptual and methodological coherence.

The difference is measurable as expected. The NSoN started to emerge in the early 90s and growth accelerated following the publication of two key papers [Watts and Strogatz, 1998, Barabási and Albert, 1999] which were each cited more than 4,000 times in the *Web of Science*® at the end of 2009. The field's social structure densified continuously, independent of counting method. In 2006 a giant component, in which a large part of the entire network's authors collaborate at least indirectly, started to emerge. Since growth took off in 1990, the field is productive. These results and the results of the bibliometric studies described earlier [Chen, 2005, Shibata et al., 2007, 2008, Lazer et al., 2009] can easily be reconciled with the model. According to the definition, the NSoN is emerging science because it is at the critical point of growing a giant component.

It is more difficult with the FoS: This field is older and also experienced a growth burst beginning in the early 90s, even though it was already its second. A giant component could not be identified, so it is not emerging science according to the strong definition. The social structure seems to be densifying, at least starting in the 90s, but whole and fractional counting show considerably different trends. What is more, the field was never productive according to the scaling indicator used, which adds to the uncertainty that the field is emerging science according to the weak definition. This result agrees with the diffuse picture painted by earlier bibliometric analyses. On the one hand, it does not contradict the apparent crystallization in the course of the 80s around the journal *Scientometrics* [Schubert and Maczelka, 1993] or the homogenous use of citations which indicated an own identity of the field [Wouters and Leydesdorff, 1994]. The field, as represented by this core journal, may indeed have been more coherent at that time. But it has changed markedly due to the advance of technical possibilities and the increasing importance of electronic information [Besselaar and Heimeriks, 2006]. On the other hand, the result does support the expert judgements that state a lack of conceptual coherence [Glänzel and Schoepflin, 1994, Larsen, 2008].

Taken together, the conclusion is that the three metrics used here should be added to the bibliometric tool-box if emerging science is to be diagnosed or research fronts are to be identified. Using the strong definition, emerging science can only be identified when the process of emergence, which is relatively short compared to a field's lifetime, is already underway. The densification indicator α and the productivity indicator β can give early indications. Being scale-independent [Katz, 2000], they are in principle suitable for the diagnosis of self-similar growth dynamics. As has been shown here and elsewhere [Bettencourt et al., 2009], fields that sooner or later have grown a giant component have a value $\alpha > 1$. Densification likely is a condition for a giant component to emerge, but to what extent this indicator has prognostic power remains to be studied. It also remains to be seen how robust it is when counting methods change. This study has shown again that the counting method does have an effect on the results [Gauffriau and Larsen, 2005]. β seems to require more thoroughness in interpretation because not all fields that eventually grew a giant component were productive [Bettencourt et al., 2008].

Does the FoS not grow a giant component because the social sciences generally have largely fragmented collaboration structures [Wouters and Leydesdorff, 1994]? Moody [2004] did find a giant component consisting of 42% of 87,731 authors in sociology's social structure, but co-authorship networks had been aggregated for the years 1989 to 1999. So should co-authorships only be studied in 1 year windows, as done here, in cases of hard science, like the NSoN, but not in cases of soft science, like the FoS? This is future work.

Is it possible that the FoS is simply not large enough for its social structural phase transition? Does it have to count more than 2,000 active authors per year for a giant component to emerge? This would totally contradict the self-similarity of the science system, that processes of emergence are *not* scale-dependent. The critical points of the fields diagnosed by Bettencourt et al. [2008] are not always in the same order of magnitude.

Finally, the growth periods identified in both cases not only confirm de Solla Price's observation that multiple logistic growth periods can accumulate to an exponential growth period. They also demonstrate that the phenomena of emergence is also scale-independent. Fields can still be emerging even if subfields have already emerged. Ultimately, one arrives at a model where science is always dynamic, permanently emerging, and self-organizedly critical [Bak, 1996, Van Raan, 2000].

7 Acknowledgements

The author is grateful to Diego Rybski, Andre Müller, Sybille Hinze, and numerous commentators on three workshops and colloquia for their helpful assistance.

References

Per Bak. *How Nature Works: The Science of Self-Organized Criticality.* Springer, Berlin, 1996.

Judit Bar-Ilan. Informetrics at the beginning of the 21st century – a review. *Journal of Informetrics*, 2:1–52, 2008.

Albert-László Barabási and Réka Albert. Emergence of scaling in random networks. *Science*, 286:509–512, 1999.

P. van den Besselaar and G. Heimeriks. Mapping research topics using word-reference co-occurrences: A method and an exploratory study. *Scientometrics*, 68:377–393, 2006.

Luís M. A. Bettencourt, Ariel Cintrón-Arias, David I. Kaiser, and Carlos Castillo-Chávez. The power of a good idea: Quantitative modeling of the spread of ideas from epidemiological models. *Physica A*, 364:513–536, 2006.

Luís M. A. Bettencourt, David I. Kaiser, Jasleen Kaur, Carlos Castillo-Chávez, and David E. Wojick. Population modeling of the emergence and development of scientific fields. *Scientometrics*, 75:495–518, 2008.

Luís M. A. Bettencourt, David I. Kaiser, and Jasleen Kaur. Scientific discovery and topological transitions in collaboration networks. *Journal of Informetrics*, 3:210–221, 2009.

Phillip Bonacich. The invasion of the physicists. *Social Networks*, 26:285–288, 2004.

Katy Börner and Andrea Scharnhorst. Visual conceptualizations and models of science. *Journal of Informetrics*, 3:161–172, 2009.

Chaomei Chen. Measuring the movement of a research paradigm. *Proceedings of SPIE*, 5669:63–76, 2005.

J. Courtial. A coword analysis of scientometrics. *Scientometrics*, 31:251–260, 1994.

Derek J. de Solla Price. *Little Science, Big Science... and Beyond*. Columbia University Press, 1963.

Linton C. Freeman. *The Development of Social Network Analysis: A Study in the Sociology of Science*. Empirical Press, 2004.

Marianne Gauffriau and Peder O. Larsen. Counting methods are decisive for rankings based on publication and citation studies. *Scientometrics*, 64:85–93, 2005.

W. Glänzel and U. Schoepflin. Little scientometrics, big scientometrics... and beyond? *Scientometrics*, 30:375–384, 1994.

William Goffman. Mathematical approach to the spread of scientific ideas - the history of mast cell research. *Nature*, 212:449–452, 1966.

J. Sylvan Katz. Scale-independent indicators and research evaluation. *Science and Public Policy*, 27:23–36, 2000.

Thomas S. Kuhn. *The Structure of Scientific Revolutions*. University of Chicago Press, 3rd edition, 1962.

Peder O. Larsen. The state of the art in publication counting. *Scientometrics*, 77:235–251, 2008.

David Lazer, Ines Mergel, and Allan Friedman. Co-citation of prominent social network articles in sociology journals: The evolving canon. *Connections*, 29:43–64, 2009.

Jure Leskovec, Jon Kleinberg, and Christos Faloutsos. Graphs over time: Densification laws, shrinking diameters and possible explanations. *Proceedings of the eleventh ACM SIGKDD international conference on Knowledge discovery in data mining*, pages 177–187, 2005.

Benoit B. Mandelbrot. *The Fractal Geometry of Nature*. W. H. Freeman, 1982.

Robert K. Merton. The matthew effect in science. ii. cumulative advantage and the symbolism of intellectual property. *ISIS*, 79:606–623, 1988.

James Moody. The structure of a social science collaboration network: Disciplinary cohesion from 1963 to 1999. *American Journal of Sociology*, 69:213–238, 2004.

M. E. J. Newman. Scientific collaboration networks. ii. shortest paths, weighted networks, and centrality. *Physical Review E*, 64:016132, 2001a.

M. E. J. Newman. Scientific collaboration networks. i. network construction and fundamental results. *Physical Review E*, 64:016131, 2001b.

M. E. J. Newman. Power laws, pareto distributions and zipf's law. *Contemporary Physics*, 46:323–351, 2005.

O. Persson. The intellectual base and research fronts of jasis 1986-1990. *Journal of the American Society for Information Science*, 45:31–38, 1994.

R. Keith Sawyer. *Social Emergence: Societies as Complex Systems*. Cambridge University Press, 2005.

A. Schubert and H. Maczelka. Cognitive changes in scientometrics during the 1980s, as reflected by the reference patterns of its core journal. *Social Studies of Science*, 23: 571–581, 1993.

N. Shibata, Y. Kajikawa, and K. Matsushima. Topological analysis of citation networks to discover the future core articles. *Journal of the American Society for Information Science and Technology*, 58:872–882, 2007.

N. Shibata, Y. Kajikawa, Y. Takeda, and K. Matsushima. Detecting emerging research fronts based on topological measures in citation networks of scientific publications. *Technovation*, 28:758–775, 2008.

N. Shibata, Y. Kajikawa, Y. Takeda, and K. Matsushima. Comparative study on methods of detecting research fronts using different types of citation. *Journal of the American Society for Information Science and Technology*, 60:571–580, 2009.

Henry Small. Co-citation context analysis and the structure of paradigms. *Journal of Documentation*, 36:183–196, 1980.

A. F. J. Van Raan. On growth, ageing, and fractal differentiation of science. *Scientometrics*, 47:347–362, 2000.

Duncan J. Watts and Steven H. Strogatz. Collective dynamics of 'small-world' networks. *Nature*, 393:440–442, 1998.

H. D. White and C. W. McCain. Bibliometrics. *Annual Review of Information Science and Technology*, 24:119–186, 1989.

H. D. White and C. W. McCain. Visualizing a discipline: An author co-citation analysis of information science, 1972-1995. *Journal of the American Society for Information Science*, 49:327–355, 1998.

Harrison C. White. *Identity and Control: How Social Formations Emerge*. Princeton University Press, 2nd edition, 2008.

Kenneth G. Wilson. Problems in physics with many scales of length. *Scientific American*, 241(August 1979):140–157, 1979.

P. Wouters and L. Leydesdorff. Has price's dream come true: Is scientometrics a hard science? *Scientometrics*, 31:193–222, 1994.