Complex Systems Science: Dreams of Universality, Reality of Interdisciplinarity
Sebastian Grauwin, Guillaume Beslon, Eric Fleury, Sara Franceschelli, Céline Robardet, Jean-Baptiste Rouquier, Pablo Jensen

To cite this version:
Sebastian Grauwin, Guillaume Beslon, Eric Fleury, Sara Franceschelli, Céline Robardet, et al.. Complex Systems Science: Dreams of Universality, Reality of Interdisciplinarity. Journal of the American Society for Information Science and Technology, Association for Information Science and Technology (ASIS&T), 2012, 63 (7), pp.10.1002/asi.22644. 10.1002/asi.22644 . hal-00706567

HAL Id: hal-00706567
https://hal.archives-ouvertes.fr/hal-00706567
Submitted on 11 Jun 2012

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L’archive ouverte pluridisciplinaire HAL, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d’enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.
Using a large database (~ 215 000 records) of relevant articles, we empirically study the “complex systems” field and its claims to find universal principles applying to systems in general. The study of references shared by the papers allows us to obtain a global point of view on the structure of this highly interdisciplinary field. We show that its overall coherence does not arise from a universal theory but instead from computational techniques and fruitful adaptations of the idea of self-organization to specific systems. We also find that communication between different disciplines goes through specific “trading zones”, ie sub-communities that create an interface around specific tools (a DNA microchip) or concepts (a network).

INTRODUCTION

Fundamental science has striven to reduce the diversity of the world to some stable building blocks such as atoms and genes. To be fruitful, this reductionist approach must be complemented by the reverse step of obtaining the properties of the whole (materials, organisms) by combining the microscopic entities, a notoriously difficult task [1–4]. The science of complex systems tackles this challenge, albeit from a different perspective. It adds the idea that “universal principles” could exist, which would allow for the prediction of the organization of the whole regardless of the nature of the microscopic entities. Ludwig Von Bertalanffy wrote already in 1968: “It seems legitimate to ask for a theory, not of systems of a more or less special kind, but of universal principles applying to systems in general” [5]. This dream of universality is still active: “[Complex networks science] suggests that nature has some universal organizational principles that might finally allow us to formulate a general theory of complex systems” [6]. Have such universal principles been discovered? Could they link disciplines such as sociology, biology, physics and computer science, which are very different in both methodology and objects of inquiry [7]?

RESULTS

In this paper, we empirically study the “complex systems” field using the quantitative tools developed to understand the organization of scientific fields [8] and their evolution [9–11]. Global science maps [8, 12–22] have become feasible recently, offering a tentative overall view of scientific fields and fostering dreams of a “science of science” [14]. Specifically, to collect a representative database of articles, we selected from the ISI Web of knowledge [23] all records containing topic keywords relevant for the field of complex systems (Table I). Table II contains the 20 most frequent cited references and journals within our dataset. To analyze the data, we build a network [14] in which the ~ 215 000 articles are the nodes. These nodes are linked according to the proportion of shared references (bibliographic coupling [24]). For this study, bibliographic coupling offers two advantages over the more usual co-citation link: it offers a faithful representation of the fields, giving equal weight to all published papers (whether cited or not) and it can be applied to recent papers (which have not yet been cited). For more details, the reader is referred to the section “Methods”.

Figure 1 shows the largest communities (thereafter also called “fields” or “disciplines”) obtained by modularity maximization of the network of papers published in the years 2000-2008. The layout of all the graphs is obtained thanks to a spring-based algorithm implemented in the Gephi visualization software [48, 49]. We first
TABLE I. Topic keywords used in our request in the ISI Web of Knowledge database and number of articles matching independently to each of these topic keywords. Each topic keywords except the first six where coupled with the topic keywords “complex*”. We moreover rejected the articles containing the topic keywords “complex scaling” or “linear search”, two terms referring respectively to (heavily used) specific methods of quantum chemistry and computer science.

<table>
<thead>
<tr>
<th>topic keywords</th>
<th>Results</th>
</tr>
</thead>
<tbody>
<tr>
<td>self organ*</td>
<td>32484</td>
</tr>
<tr>
<td>complex network*</td>
<td>6953</td>
</tr>
<tr>
<td>dynamical system</td>
<td>8205</td>
</tr>
<tr>
<td>econophysics</td>
<td>633</td>
</tr>
<tr>
<td>strange attractor</td>
<td>769</td>
</tr>
<tr>
<td>synergetics</td>
<td>379</td>
</tr>
<tr>
<td>adaptive system*</td>
<td>1141</td>
</tr>
<tr>
<td>artificial intelligence</td>
<td>1812</td>
</tr>
<tr>
<td>attractor</td>
<td>1034</td>
</tr>
<tr>
<td>bifurcation</td>
<td>3164</td>
</tr>
<tr>
<td>chaos</td>
<td>5370</td>
</tr>
<tr>
<td>control</td>
<td>116017</td>
</tr>
<tr>
<td>criticality</td>
<td>980</td>
</tr>
<tr>
<td>ecology</td>
<td>5869</td>
</tr>
<tr>
<td>economics</td>
<td>2243</td>
</tr>
<tr>
<td>epistemology</td>
<td>345</td>
</tr>
<tr>
<td>far from equilibrium</td>
<td>253</td>
</tr>
<tr>
<td>feedback</td>
<td>12881</td>
</tr>
<tr>
<td>fractal</td>
<td>3867</td>
</tr>
<tr>
<td>ising</td>
<td>975</td>
</tr>
<tr>
<td>multi agent</td>
<td>2032</td>
</tr>
<tr>
<td>multiagent</td>
<td>665</td>
</tr>
<tr>
<td>multi scale</td>
<td>779</td>
</tr>
<tr>
<td>multifractal</td>
<td>390</td>
</tr>
<tr>
<td>multiscale</td>
<td>1439</td>
</tr>
<tr>
<td>neural network*</td>
<td>12747</td>
</tr>
<tr>
<td>(non linear* OR nonlinear*) NOT equation*</td>
<td>10240</td>
</tr>
<tr>
<td>non linear dynamic*</td>
<td>560</td>
</tr>
<tr>
<td>non linear system*</td>
<td>391</td>
</tr>
<tr>
<td>nonlinear dynamic*</td>
<td>2245</td>
</tr>
<tr>
<td>nonlinear system*</td>
<td>1826</td>
</tr>
<tr>
<td>phase transition</td>
<td>5503</td>
</tr>
<tr>
<td>plasticity</td>
<td>6667</td>
</tr>
<tr>
<td>random walk</td>
<td>758</td>
</tr>
<tr>
<td>robustness</td>
<td>6498</td>
</tr>
<tr>
<td>scaling</td>
<td>7008</td>
</tr>
<tr>
<td>social system*</td>
<td>586</td>
</tr>
<tr>
<td>spin glass*</td>
<td>643</td>
</tr>
<tr>
<td>stability AND (lyapunov OR non linear* OR nonlinear*)</td>
<td>1399</td>
</tr>
<tr>
<td>stochastic</td>
<td>9184</td>
</tr>
<tr>
<td>synchronisation</td>
<td>4645</td>
</tr>
<tr>
<td>turbulence</td>
<td>4602</td>
</tr>
<tr>
<td>universality</td>
<td>861</td>
</tr>
<tr>
<td>cell* automaton*</td>
<td>1659</td>
</tr>
</tbody>
</table>

Note that all important complex systems subfields1 are present [25]. At the center, we find mostly theoretical domains: self-organized criticality, dynamical systems, complex networks, neural networks. These fields are connected to more experimental communities lying at the edges (materials science, biology or neurosciences). The links between theoretical and experimental fields suggest that complex systems science models have connections to the “real” world, as claimed by their practitioners.

To understand the inner structure of these large communities, we use recursive modularity optimization (see [26] and “Methods”). Most fields display a rich inner structure (Figure 2) with subcommunities (thereafter also called “subfields” or “subdisciplines”) organized around specific topics and references. The only exceptions are self-organized criticality and complex networks, where all articles cohere around a few references. For a short presentation of all the subfields, see Table III. For a more detailed presentation of the main subfields, including their authors, most used journals, references and keywords, see the Supplementary Information.

We analyze this complex structure at two levels. First, at the global scale, complex systems science appears to be a densely interconnected network. This is somewhat

1 In the following, we use italics to refer to the names of the communities.
TABLE II. The 20 references (including books and articles) and journals which are the more used by the articles of the whole database

<table>
<thead>
<tr>
<th>Reference</th>
<th>Times used</th>
<th>Journal (# distinct refs)</th>
<th>Times used</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bak P, 1987, PHYS REV LETT (59)</td>
<td>2131</td>
<td>NATURE (29166)</td>
<td>169309</td>
</tr>
<tr>
<td>Albert R, 2002, REV MOD PHYS (74)</td>
<td>2050</td>
<td>P NATL ACAD SCI USA (42504)</td>
<td>151140</td>
</tr>
<tr>
<td>Laemmli UK, 1970, NATURE (227)</td>
<td>1762</td>
<td>J BIOL CHEM (59436)</td>
<td>149042</td>
</tr>
<tr>
<td>Watts DJ, 1998, NATURE (393)</td>
<td>1732</td>
<td>SCIENCE (24880)</td>
<td>148002</td>
</tr>
<tr>
<td>Barabasi AL, 1999, SCIENCE (286)</td>
<td>1693</td>
<td>CELL (11044)</td>
<td>99168</td>
</tr>
<tr>
<td>Bak P, 1988, PHYS REV A (38)</td>
<td>1555</td>
<td>PHYS REV LETT (23269)</td>
<td>94861</td>
</tr>
<tr>
<td>Sambrook J, 1989, MOL CLONING LAB MANU</td>
<td>1439</td>
<td>J AM CHEM SOC (29807)</td>
<td>82569</td>
</tr>
<tr>
<td>Newman MEJ, 2001, SIAM REV (45)</td>
<td>1308</td>
<td>EMBO J (10926)</td>
<td>53049</td>
</tr>
<tr>
<td>Bradford MM, 1976, ANAL BIOCHEM (72)</td>
<td>1255</td>
<td>MOL CELL BIOL (12866)</td>
<td>52694</td>
</tr>
<tr>
<td>Lowry OH, 1951, J BIOL CHEM (193)</td>
<td>1106</td>
<td>J NEUROSCI (12313)</td>
<td>43152</td>
</tr>
<tr>
<td>Rundholt DE, 1986, PARALLEL DISTRIBUTED (1)</td>
<td>947</td>
<td>J IMMUNOL (18891)</td>
<td>41496</td>
</tr>
<tr>
<td>Strogatz SH, 2001, NATURE (410)</td>
<td>907</td>
<td>PHYS REV B (19367)</td>
<td>41450</td>
</tr>
<tr>
<td>Kohonen T, 1982, BIOL CYBERN (43)</td>
<td>901</td>
<td>J CELL BIOL (10239)</td>
<td>40560</td>
</tr>
<tr>
<td>Chomczynski P, 1987, ANAL BIOCHEM (162)</td>
<td>849</td>
<td>J CHEM PHYS (17136)</td>
<td>40074</td>
</tr>
<tr>
<td>Goldberg DE, 1989, GENETIC ALGORITHMS S</td>
<td>822</td>
<td>GENE DEV (4879)</td>
<td>38903</td>
</tr>
<tr>
<td>Lorenz EN, 1963, J ATMOS SCI (20)</td>
<td>726</td>
<td>BIOCHEMISTRY-US (16035)</td>
<td>32601</td>
</tr>
<tr>
<td>Mandelbrot BB, 1982, FRACTAL GEOMETRY NAT</td>
<td>721</td>
<td>BRAIN RES (15364)</td>
<td>30517</td>
</tr>
<tr>
<td>Kohonen T, 1990, P IEEE (78)</td>
<td>715</td>
<td>ANGEW CHEM INT EDIT (7572)</td>
<td>27718</td>
</tr>
<tr>
<td>Dorogovtsev SN, 2002, ADV PHYS (51)</td>
<td>688</td>
<td>NUCLEIC ACIDS RES (9738)</td>
<td>27242</td>
</tr>
<tr>
<td>Albert R, 2000, NATURE (406)</td>
<td>678</td>
<td>J EXP MED (8100)</td>
<td>27220</td>
</tr>
</tbody>
</table>

surprising since sharing references between subdisciplines means that they are able to read and understand these references, and moreover, that they find them useful. Would these shared references point to “universal” principles? Second, we focus on a more local scale, on the links that specifically connect two different disciplines (ie two different colors in Figure 1) to understand how they manage to exchange knowledge.

**Complex systems’ science overall coherence**

Let us start with the field’s overall coherence. We have looked for the references cited by many subfields. These form the “glue” that links many subdisciplines and connects the network. More precisely, we define the networking force of a reference $N(r)$ as the sum, over all pairs of subfields, of the proportion of their links explained by that reference (see Methods). Table IV shows that the references that glue the network are more methodological than theoretical: the most networking reference is “Numerical Recipes” [27], a series of books that gathers many routines for various numerical calculations and their implementation in computers. Most of the other linking references are mathematical handbooks or data analysis tools. If one looks for universality in the complex systems field, the computer – as a tool – seems to be a serious candidate. Among the leading contributors to the glue, we also find several references on self-organization (SO). Self-organization is not a predictive theory, but an approach that focuses on the spontaneous emergence of large-scale structures out of local interactions between the system’s subunits [28]. Several subdisciplines in Figure 2 can be related to this approach, as they use a keyword akin to “self-organization” (SO) in more than 10% of their articles (for a more complete list of the communities using this keyword, see the Supplementary Information). Among these we find *swarm SO*, *molecular SO* linking chemistry to biology, *growth SO* and *pattern formation SO* linking surface science to dynamical systems. This suggests that the field of complex systems focuses on the cases in which the link from microscopic to macroscopic can be analyzed through self-organization, which gave rise to several fruitful scientific programs, as we discuss below (section Discussion).

**Interdisciplinary trading zones**

At a more local scale, let us now look at the links that specifically connect two distinct disciplines. How are those connections established? It is widely accepted that scientific disciplines cannot easily communicate or be linked (in our case, share references) simply because it is difficult for a physicist to understand a biology paper and vice-versa. In addition, different disciplines have different definitions of what counts as a result or as an interesting research topic. For example, physical sciences look for universal laws, while social [29] and biological [30] sciences emphasize the variations in structure across different groups or contexts and use these differences to explain differences in outcomes. Physicians are interested
FIG. 2. Community structure obtained with a second run of the modularity maximization on the 2000-2008 network. This community structure is obtained by optimizing the internal modularity $Q$ of each community obtained by the first run of the modularity maximization algorithm on the 2000-2008 network, displayed on Figure 1 (See Methods for details on the procedure). The layout of the graph is obtained thanks to a spring-based algorithm implemented in the Gephi visualization software [48, 49]. The surface of each community is proportional to its number of articles and the width of the link between two communities $I$ and $J$ is proportional to the mean weight $\langle \omega \rangle_{IJ}$. For the sake of clarity, communities with less than 300 articles and links with a mean weight $\langle \omega \rangle_{IJ}$ less than $2 \times 10^{-5}$ are not displayed. The color of a community (see online) corresponds to the color of the field (Fig 1) it belongs to. Community labels generally correspond to the most frequent and/or significant keyword. For a detailed presentation of all the subfields, including their authors, most used journals, references and keywords, see the Supplementary Information.

in practical medical advances while physicists want to know whether physiological rhythms are chaotic or not [31].

Where do the links come from then? In an illuminating analogy, Peter Galison [32] compares the difficulty of connecting scientific disciplines to the difficulty of communicating between different languages. History of language has shown that when two cultures are strongly motivated to communicate - generally for commercial reasons - they develop simplified languages that allow for simple forms of interaction. At first, a “foreigner talk” develops, which becomes a “pidgin” when social uses consolidate this language. In rare cases, the “trading zone” stabilizes and the expanded pidgin becomes a creole, initiating the development of an original, autonomous culture. Analogously, biologists may create a simplified and partial version of their discipline for interested physicists, which may develop to a full-blown new discipline such as biophysics. Specifically, Galison has studied [32] how Monte Carlo simulations developed in the postwar period as a trading language between theorists, experimentalists, instrument makers, chemists and mechanical engineers. Our interest in the concept of a trading zone is to allow us to explore the dynamics of the interdisciplinary interaction instead of ending analysis by reference to a “symbiosis” or “collaboration”.
The acronyms and abbreviations in parenthesis correspond to the label of the subfields displayed on Fig 2. The inverse of the average of the weight of the inner links of a subfield, then two articles of this subfield chosen at random would share 1 reference over $2000 - 2008$ subfields' sizes $N < \omega^{-1}$ and modules $q$.

Table V gives a list of the main “trading zones” which connect theoretical and experimental fields in Figure 1 and capture a significant fraction of the links between these fields. The clearest example is transcriptomics data analysis, a subfield of neural networks which connects biologists interested in the interpretation of data retrieved from DNA chips and computer scientists interested in data analysis via methods from the neural networks field. The transcriptomics data analysis subfield represents $2.3\%$ of neural networks papers but accounts for $46.3\%$ of the connections between neural networks and biology and $16.5\%$ of the links between neural networks and complex networks. Other trading zones are computational systems biology, linking biology to many theoretical fields, among which dynamical systems, self-organized criticality and complex networks, neural synchronization linking dynamical systems and neurosciences, cytoskeleton self-organization linking biology to dynamical systems and self-organized criticality and calibration linking neural networks and material sciences. Note that a single
trading zone can be used by a fields to exchange with several other fields, as long as these other fields share the same “language”. For example, computational systems biology, links biology to dynamical systems, self-organized criticality and complex networks, three subfields which share the physicists’ toolkit. Since our map cannot cover all scientific fields, we may not recognize some subfields as trading zones, such as electrocardiogram which is likely to connect dynamical systems to medicine, or even miss a trading zone between geosciences and self-organized criticality.

By analyzing carefully the references used by trading zones and also the references that make the links between the trading zones and their neighbors, we can distinguish two types of trading zones, applicative and speculative. Let us start with transcriptomics data analysis, which is a clear example of “applicative” trading zone. The development of new measurement techniques in cellular biology (mainly DNA microarrays) produced huge amounts of data together with the need of new tools to analyze them. Since this new technique promised a better understanding of cell dynamics, a new scientific subdiscipline, able to understand data analysis and its biological interest was built around transcriptomic tools. The two references most used by this subfield stress the applicative side: the purpose of the first paper is “to describe a system of cluster analysis for genome-wide expression data from DNA microarray hybridization […] in a form intuitive for biologists” [33] while the second “describes the application of self-organizing maps for recognizing and classifying features in complex, multidimensional [transcriptomic] data” [34]. The transcriptomics data analysis papers are clustered together because they share references presenting this kind of applications. The applicative character of transcriptomics data analysis can also be seen in the origin of the references that link them to neighbor subfields (Figure 4). The common references between transcriptomics data

FIG. 3. Local “networking” force for four different references on the 2000-2008 network (Fig 2). Links established using the reference are shown in color. The number of citations corresponds to those included in papers of our database published between 2000 and 2008 a. Ref used: Albert & Barabasi [40] (2058 citations) b. Ref used: Press WH et al, Numerical Recipes - all editions [27] (1267 citations) c. Ref used: Nicolis [50] (342 citations) d. Ref used: Barabasi and Oltvai [42] (244 citations).
FIG. 4. Directed network. On this subset of the graph presented in Figure 2, the arrows are directed to the subfield that uses the other subfield’s references to establish the link. More precisely, the common references shared by two linked subfields are more similar to the internal references of the subfield from which the arrow originates than to the internal references of the subfield to which the arrow points (see “Methods” for more details). The figure shows that transcriptional data analysis (TDA) feeds from self-organizing maps (SOM) and neural networks (NN) methodological references, while biology subcommunities (mainly Transcriptomics) use transcriptional data analysis references. The orientation of the links is quite different for computational systems biology (CSB) and complex networks (CN), because these subfields tend to pump their neighbors’ references, while the other subfields do not find much use in computational systems biology and complex networks references.

The 20 most networking references in the 2000-2008 decade

<table>
<thead>
<tr>
<th>Reference</th>
<th>Topic</th>
<th>$N(r)$ (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Press et al. (1992)*</td>
<td>Numerical recipes (book)</td>
<td>1.250</td>
</tr>
<tr>
<td>Shannon (1948)*</td>
<td>Information theory</td>
<td>0.607</td>
</tr>
<tr>
<td>Metropolis et al. (1953)</td>
<td>Monte Carlo integration</td>
<td>0.509</td>
</tr>
<tr>
<td>Nicolis et al. (1977)*</td>
<td>Self organization (book)</td>
<td>0.420</td>
</tr>
<tr>
<td>Kauffman (1993)*</td>
<td>Self organization (book)</td>
<td>0.309</td>
</tr>
<tr>
<td>Hebb (1949)</td>
<td>Neuropsychology and behavior (book)</td>
<td>0.297</td>
</tr>
<tr>
<td>Alberts et al. (1994)</td>
<td>Molecular and cellular biology (book)</td>
<td>0.288</td>
</tr>
<tr>
<td>Abramowitz et al. (1968)*</td>
<td>Handbook of mathematical functions</td>
<td>0.269</td>
</tr>
<tr>
<td>Feller (1958)*</td>
<td>Introduction to probability theory (book)</td>
<td>0.268</td>
</tr>
<tr>
<td>Watson &amp; Crick (1953)</td>
<td>Structure of DNA</td>
<td>0.250</td>
</tr>
<tr>
<td>Lakowicz (1999)</td>
<td>Fluorescence spectroscopy</td>
<td>0.249</td>
</tr>
<tr>
<td>Turing (1952)</td>
<td>Morphogenesis</td>
<td>0.237</td>
</tr>
<tr>
<td>Witten et al. (1981)</td>
<td>Diffusion-limited aggregation</td>
<td>0.234</td>
</tr>
<tr>
<td>Cohen (1988)</td>
<td>Statistics and behavioral sciences (book)</td>
<td>0.223</td>
</tr>
<tr>
<td>Hopfield (1982)</td>
<td>Neural networks</td>
<td>0.217</td>
</tr>
<tr>
<td>Stanley (1971)</td>
<td>Phase transition (book)</td>
<td>0.202</td>
</tr>
<tr>
<td>Whitesides et al. (2002)</td>
<td>Self-assembly</td>
<td>0.188</td>
</tr>
<tr>
<td>Marquardt (1963)</td>
<td>Applied mathematics</td>
<td>0.174</td>
</tr>
<tr>
<td>Chomczynski (1987)</td>
<td>RNA isolation</td>
<td>0.167</td>
</tr>
<tr>
<td>Venter et al. (2001)</td>
<td>Human genome sequence</td>
<td>0.160</td>
</tr>
</tbody>
</table>

The references followed by a star correspond to books or papers which appeared in the database under several forms - essentially different publication years for the books - for which the networking power $N(r)$ have been summed. The complete references of these papers are given in Supplementary Information.

In computational systems biology and biology (mainly transcriptomics) are similar to references used by transcriptional data analysis papers themselves. This means that the link arises from biologists citing results obtained by transcriptional data analysis scientists or techniques they use. On the other hand, the common references between transcriptional data analysis and self-organizing maps (a subfield of neural networks) are similar to references used by self-organizing maps papers. Therefore, the link arises from transcriptional data analysis scientists citing classification techniques created by self-organizing maps scientists, while these scientists do not often use transcriptional data analysis references. Therefore, transcriptional data analysis allows self-organizing maps techniques to be understood and used to interpret biological data, with a relevance certified by biologists’ citations. The case of another trading zone, computational systems biology, is different. Its most used references point to computational methods - mainly Gillespie’s algorithm [35] or to experimental papers in which there is no explicit modeling but that show complex cellular dynamics, thus justifying indirectly the need for modeling. The link between experiments and modeling is still speculative, as summarized by one of the most used references in this subfield [36]: “we hope that this review will [...] promote closer collaboration between experimental and computational biologists.” Moreover, the common references between computational systems biology and biology are from biology, as if computational systems biology scientists were eager to quote potentially interesting biological applications for their modeling approach, while many biologists were still unaware of these models. In short, compared to transcriptional data analysis, computational systems biology seems a more speculative trading zone, at the frontier of biology and modeling, but presently lacking a specific object or concept to define an operational trading zone.
TABLE V. Strongest trading zones.

<table>
<thead>
<tr>
<th>Subfield</th>
<th>Fields</th>
<th>T (%)</th>
<th>T/T_{exp}</th>
</tr>
</thead>
<tbody>
<tr>
<td>TDA</td>
<td>Biology/Neural Networks</td>
<td>46.355</td>
<td>20.51</td>
</tr>
<tr>
<td>CSB</td>
<td>Biology/Dynamical Systems</td>
<td>49.704</td>
<td>8.87</td>
</tr>
<tr>
<td>CSB</td>
<td>Biology/SOC</td>
<td>49.255</td>
<td>8.79</td>
</tr>
<tr>
<td>ProteinStruct</td>
<td>Biology/Material Sciences</td>
<td>47.442</td>
<td>8.32</td>
</tr>
<tr>
<td>CSB</td>
<td>Biology/CN</td>
<td>42.931</td>
<td>7.66</td>
</tr>
<tr>
<td>TDA</td>
<td>CN/Neural Networks</td>
<td>16.543</td>
<td>7.32</td>
</tr>
<tr>
<td>Hemodyn</td>
<td>Neurosciences/FluidMech</td>
<td>54.552</td>
<td>7.22</td>
</tr>
<tr>
<td>NeuralSynchr</td>
<td>Neurosciences/Dynamical Systems</td>
<td>59.788</td>
<td>5.20</td>
</tr>
<tr>
<td>Hemodyn</td>
<td>Biology/FluidMech</td>
<td>39.113</td>
<td>5.18</td>
</tr>
<tr>
<td>CytoskSO</td>
<td>Biology/SOC</td>
<td>9.561</td>
<td>4.71</td>
</tr>
<tr>
<td>CytoskSO</td>
<td>Biology/Dynamical Systems</td>
<td>9.522</td>
<td>4.69</td>
</tr>
<tr>
<td>Calib</td>
<td>Material Sciences/Neural Networks</td>
<td>8.717</td>
<td>4.51</td>
</tr>
<tr>
<td>CellularNN</td>
<td>CN/Neural Networks</td>
<td>9.605</td>
<td>4.30</td>
</tr>
<tr>
<td>Transcript</td>
<td>Biology/CN</td>
<td>25.806</td>
<td>4.05</td>
</tr>
</tbody>
</table>

The trading force \( T \) of a subcommunity measures the fraction of the links between two fields (in Fig. 1) which goes through this subcommunity. More precisely, the trading force of \( I \), a subcommunity of \( \bar{I} \), towards any community \( J \) is the total weight of the article-article links between the subcommunity \( I \) and community \( J \), normalized by the total weight of the article-article links between \( I \) and \( J \):

\[
T_{IJ}(I) = \frac{\sum_{i \in I, j \in J} \omega_{ij}}{\sum_{i \in I, j \in J} \omega_{ij}}.
\]

The expected force \( T_{IJ}^{exp} \) is the value of the trading force one would expect if all the links between \( I \) and outside communities were equally shared among all sub-communities of \( I \), which is simply the fraction \( N_I/N_J \) of articles of \( I \) in \( J \). The acronyms of the subfields used here correspond to those explained in Table III.

DISCUSSION

Our empirical study of the “complex systems” field shows that its overall coherence does not arise from a universal theory but from computational techniques and successful adaptations of the idea of self-organization. The computer is important for advancing the understanding of complex systems because it allows scientists to play with simple but nonlinear models and to handle large sets of data obtained from complex systems. At a more local level — specifically the interdisciplinary level — trading zones allow for coordination between vastly different scientific cultures, who differ on their conception of an interesting topic, but who can work together around specific tools (a DNA microchip) or concepts (a network).

We now discuss how our study sheds light on the overall philosophy of the complex systems field. First, we examine the various claims to universality. A “general systems theory” would possess a collection of theoretical books or papers revealing the “universal” explanation; this would be evidenced in Figure 2 by a central group to which other groups would connect. Instead, our analysis shows a variety of modeling disciplines in a central position.

We argue that claims to universality are part of a rhetoric that legitimates the study of abstract and simple models [38]. Certainly, a few theoretical papers, such as Bak’s [39] (in SOC) or Albert and Barabasi’s [40] (in CN) point to “universal” mechanisms and are heavily cited. However, more than 90% of their citations arise from modelers themselves [41], suggesting that they may be universal... for theorists. Our data support the local character of these “universal” laws. First, Albert and Barabasi’s [40] paper is the most cited in the 2000-2008 decade but only links complex networks and self-organized criticality subfields (Figure 3a). The contrast with the global networking achieved by methodological references such as Metropolis’ algorithm or Numerical Recipes (Figure 3b) or self-organization references (Figure 3c) is clear. Second, the references that complex networks (Figure 4) and self-organized criticality communities share with experimental fields are similar to those of the experimental fields. This seizure of experimental references suggests that the links between modeling practices and their potential applications are mostly rhetorical: complex networks and self-organized criticality papers often quote experimental work as legitimizing their models, while experimentalists rarely refer to them. To try to become universal, theoretical approaches have to be “translated” into other disciplines. An example of this strategy is shown in Figure 3d which shows the links established between network science and biology thanks to Barabasi and Oltvai’s introduction of networks for biologists [42]. Regardless, many physicists are likely to continue looking for common patterns across systems to justify their neglect of the “details” of the system under study, the precise components and interactions [30, 52]. Universality is then another name for simplicity, a strong motivation for many physicists as expressed by the Santa Fe institute who aims at uncovering “the mechanisms that underlie the deep simplicity present in our complex world.” [53]. It is true that simple analytical models such as SOC or chaotic systems may lead to complicated behaviors and patterns. But this does not prove the reverse proposition, i.e. that all complex patterns can be explained by simple mechanisms. The “simplicity” approach turned out to be a successful strategy in the study of phase transitions, which can be studied through the very simple Ising model [51, 54, 55], but arguments for the usefulness of such an approach for biological or social systems are unconvincing [38].

It could be argued that links between these theories and experimental fields take time to establish and will be seen in the future. An interesting insight of the possible evolution of universality claims is given by the history of self-organization, which was considered by many as a universal key to Nature in the 1980’s [51]. This idea was fecund in that it gave birth to several active sub-disciplines (cytoskeleton SO, growth SO...) (Figure 2). However, it should be noticed that these heirs of self-organization are nowadays almost unrelated. The differ-
ent self-organization subfields are more linked to their own discipline (biology, materials science . . .) than between them. This is shown by the plain fact that community detection puts these SO subfields into different disciplines (different colors in Figure 2) instead of creating a single, unified, self-organization field. The reason is that these subfields use widely different references, as illustrated by the fact that there is no common reference among the 10 most used references for all the different self-organization subcommunities. Self-organization is therefore not a universal explanation but rather a kind of banner, which needs to be associated to references to specific elements (including techniques, microscopic entities and their interactions) to be fruitful.

To understand the essential role of the computer, it is important to distinguish between complex and complicated. A complex phenomenon has to be understood synthetically, as a whole, while a complicated phenomenon can be explained analytically [56]. In other words, a system is complex when the “parts” that are relevant to link the micro to the macro cannot be properly defined, as in social systems where humans cannot be defined without taking into account society (e.g., language, an essential part of the individual which is acquired through society). Similarly, synthetic biology aims at using functional components of living systems as building blocks to create artificial devices [57]. But many difficulties arise from the intertwining of the elements in a living being [58]. Biological parts are ill defined and their function cannot be isolated from the context that they themselves create. Despite its claims to complexity and holism, the “complex systems” field proposes a standard mechanistic vision of nature and society. As for most natural sciences, its aim is to transform complex systems into complicated systems that can be handled and eventually engineered by models and computer force.

This is confirmed by historical studies showing [51, 59] that the complex systems field is heir of the postwar sciences born around the computer: operational research, game theory and cybernetics. These fields started when physicists, mathematicians and engineers started collaborating to maximize the efficiency of WW II military operations [37, 59, 60]. These sciences extended the mechanistic, engineering vision of the physical world to the biological and social worlds. This view is still present in many today’s prominent CS scientists: “our knowledge of [social] mechanisms [...] is essential for self-optimization of the society as a whole” [61]; “We spend billions of dollars trying to understand the origins of the universe, while we still don’t understand the conditions for a stable society, a functioning economy, or peace” [62] or “[Systems biology] leads to a future where biology and medicine are transformed into precision engineering” [63].

In summary, we have obtained a global point of view on the structure of the “complex systems” field. This has allowed us to test empirically the idea of universality, showing that it remains a dream, albeit one which has lead to interesting but more modest realities. At the global scale, the whole domain is linked by the focus on self-organization and the use of computer-based methods for solving non-linear models. At a more local scale, the links between different disciplines are achieved through the development of “trading zones” [32]. These allow for coordination between vastly different scientific cultures, for example theoretical and experimental disciplines, which are only marginally connected. These disciplines may differ on the very conception of what is an interesting topic, but can work together around specific tools (a DNA microchip) or concepts (a network). Today, these interdisciplinary collaborations are a key to essential scientific challenges such as the analysis of the massive amount of data recently made available on biological and social systems [43, 44] and the understanding of the complex intertwining of different levels of organization that is characteristic of these systems.

**METHODS**

**Extraction of the data**

Our data have been extracted from the ISI Web of Knowledge database [23] in December 2008. The science of complex systems is particularly challenging as an epistemic object since there exists no consensual definition of the domain, nor any list of disciplines or journals that would gather all the relevant papers. Therefore, we selected all the articles of the database whose title, abstract (for articles published after 1990) or keywords contained at least one of a chosen list of topic keywords (Table I). These keywords were derived from discussions with experts of the field, mainly scientists working at the complex systems institute in Lyon (IXXI). We have retrieved 215 305 articles (141 098 between 2000 and 2008) containing 4 050 318 distinct references. Each record contains: authors, journal, year of publication, title, keywords (given by the authors and/or ISI Web of Science) and the list of references of the article. Any choice of keywords being potentially biased and partial, our strategy was to risk choosing too many of them – thus bypassing the lack of precise definition of the “complex systems” field and retrieving all its important subfields - and to trust the subsequent analysis to eliminate irrelevant articles.

In fact, as shown in table I, around 40% of the articles of the database comes solely from the combination of keywords “complex*” and “control”. While most of those articles were close to biology and not directly related to the field of complex systems, we chose to keep them in order to test the robustness of our analysis. As shown below, our strategy was successful, since most of
these “irrelevant” articles are grouped into a few communities (such as \textit{Apoptosis} or \textit{Immunology}) that lie at the network’s edges and do not bias the results.

\textbf{Links between articles}

Weight of links between articles are calculated through their common references (bibliographic coupling [24]). We define a similarity between two articles $i$ and $j$ as the cosine distance:

$$\omega_{ij} = \frac{|\mathcal{R}_i \cap \mathcal{R}_j|}{\sqrt{|\mathcal{R}_i||\mathcal{R}_j|}} \quad (1)$$

where $\mathcal{R}_i$ is the set of references of article $i$. By definition, $\omega_{ij} \in [0, 1]$, is equal to zero when $i$ and $j$ do not share any reference and is equal to 1 when their sets of references are identical. For this study, bibliographic coupling offers two advantages over the more usual co-citation link: it offers a faithful representation of the fields, giving equal weight to all published papers (whether cited or not) and it can be applied to recent papers (which have not yet been cited). Moreover, the links are established on the basis of the author’s own decisions (to include or not a given reference) rather than retrospectively from other scientists’ citations. Thus, bibliographic coupling can be used to analyze the community of research as it builds itself rather than as it is perceived by later scientists that cite its publications.

\textbf{Community detection and characterization}

In order to structure this network into groups of cohesive articles, we partition the set of papers by maximizing the modularity function. Given a partition of the nodes of the network, the modularity is the number of edges inside clusters (as opposed to crossing between clusters), minus the expected number of such edges if the network was randomly conditioned on the degree of each node. Community structures often maximize the modularity measure. We compute our partition using the algorithm presented in [47], which is designed to efficiently maximize the modularity function in large networks. More precisely, we used the weighted modularity $Q$ [45, 46], which is defined as $Q = \sum q_I$, where the module $q_I$ of a community $I$ is given by

$$q_I = \frac{\Omega_{II}}{\Omega} - \left(\frac{\sum_{J \neq I} \Omega_{IJ} + 2\Omega_{II}}{2\Omega}\right)^2 \quad (2)$$

where

$$\Omega_{II} = \frac{1}{2} \sum_{i \in I, j \in I} \omega_{ij}$$

is the total weight of the links inside community $I$,

$$\Omega_{IJ} = \sum_{i \in I, j \in J} \omega_{ij}$$

is the total weight of the links between communities $I$ and $J \neq I$

$$\Omega = \frac{1}{2} \sum_{i,j} \omega_{ij} = \sum_{(i,j)} \omega_{ij}$$

is the total weight of the links of the graph.

Each module $q_I$ compares the relative weight of edges $\Omega_{II}$ inside a community $I$ with the expected weight of edges $\left(\frac{\sum_{i \in I} \Omega_{IJ} + 2\Omega_{II}}{2\Omega}\right)^2$ that one would find in community $I$ if the network were a random network with the same number of nodes and where each node keeps its degree, but edges are otherwise randomly attached. See Ref [26] for a more explicit interpretation of the modularity, its properties and limits.

Applying the Louvain algorithm yields a first partition of the network into communities (also referred to as “fields” or “disciplines”, see Figure 1). To obtain the substructure of these communities, we apply the Louvain algorithm a second time on each of them. We find that most of these communities display a clear substructure with high values of internal modularity $Q^i$ (typically between 0.4 and 0.8). Only two of them (self-organized criticality and complex networks) are strongly bound around a few references and present much lower values of $Q^i$ (typically less than or around 0.2). Consequently they were not split into subfields which would not have much scientific relevance.

This recursive modularity optimization [26] leads us to a “subfield” graph (Figure 2). We have checked that all the obtained sub-communities satisfy the criterion $(q_I \geq 0)$ proposed by Fortunato and Barthélémy [26] to check their relevance (see Table III).

\textbf{Links between communities and their orientation}

The link between two communities $I$ and $J$ can be quantified by the average distance between an article $i \in I$ and an article $j \in J$:

$$<w>^{-1}_{IJ} = <\omega_{ij}>^{-1}_{i \in I, j \in J} = (\Omega_{IJ}/N_I N_J)^{-1} \quad (3)$$

A link between a community $I$ and a community $J$ exists if at least one reference is shared between an article of $I$ and an article of $J$. To analyze the scientific content conveyed by the link, it is important to know if the shared references are more similar to the references used by community $I$ or to the references used by community
Let \( n_{r,I} \) be the number of papers of community \( I \) using reference \( r \). Then,

- the number of article-article links inside community \( I \) which use reference \( r \) is \( L_{r,II} = n_{r,I} (n_{r,I} - 1)/2 \)
- the number of article-article links between communities \( I \) and \( J \) which use reference \( r \) is \( L_{r,IJ} = n_{r,I} n_{r,J} \)

We compare the set of references shared by the two communities \( I \) and \( J \) to the references used by \( I \) and \( J \) by computing the cosine similarity measures:

\[
\cos_{II,IJ} = \frac{\sum_r L_{r,II} L_{r,IJ}}{\sqrt{\sum_r L_{r,II}^2} \sqrt{\sum_r L_{r,IJ}^2}} \quad (4)
\]

\[
\cos_{JJ,IJ} = \frac{\sum_r L_{r,JJ} L_{r,IJ}}{\sqrt{\sum_r L_{r,JJ}^2} \sqrt{\sum_r L_{r,IJ}^2}} \quad (5)
\]

For example, if \( \cos_{II,IJ} < \cos_{JJ,IJ} \), the shared references are more similar to the references binding community \( J \) than to the references binding community \( I \). We then direct the link from community \( J \) to community \( I \), as community \( I \) “pumps” community \( J \) references to establish the link. See Figure 4 for examples of link orientation.

Visualizing linked communities

To obtain Figures 1 and 2, we use Gephi [48]. The layout of the graph is obtained thanks to a spring-based algorithm implemented in it [49]. ForceAtlas is a force directed layout: it simulates a physical system. Nodes repulse each other (like magnets) while edges attract the nodes they connect (like springs). These forces create a movement that converges to a balanced state, which helps in the interpretation of the data.

Networking power of references

To understand which references link the different sub-disciplines to form a connected network, we define the “glue” as the set of references shared between subfields. To give equal weight to all these links, we normalize each link to 1, leading to the normalized networking strength \( \mathcal{N}(r) \) of reference \( r \) as:

\[
\mathcal{N}(r) = \frac{1}{Z} \sum_{I \neq J} f_{IJ}(r) \quad (6)
\]

where \( f_{IJ}(r) \) is the fraction of links between an article of community \( I \) and an article of community \( J \) in which reference \( r \) is used and where \( Z \) is a normalization constant such that \( \sum_r \mathcal{N}(r) = 1 \). The normalization ensures that \( \mathcal{N}(r) \) represents the proportion of all the links of the complex systems field that can be assigned to reference \( r \).

ACKNOWLEDGMENTS

We acknowledge interesting discussions with Andrei Mogoutov, Léo Granger, Taras Kowaliw and Eric Bertin.


[56] For a telling example of this distinction in agriculture, see T Venturini (2008), Terminator contro terminator, in Bella G. e Diamantini D. (eds.) Studiare la Società dell Informazione. Milano: Guerini e Associati


