

Complex Systems Science: Dreams of Universality, Reality of Interdisciplinarity

Sébastien Grauwin^{1,2,3}, Guillaume Beslon^{1,2,4}, Eric Fleury^{1,2,5,6}, Sara Franceschelli^{1,2,6}, Céline Robardet^{1,2,4}, Jean-Baptiste Rouquier⁷, Pablo Jensen^{1,2,3,*}

1 Université de Lyon, France

2 IXXI, Rhône Alpes Institute of complex systems, 69364 Lyon, France

3 Laboratoire de Physique, UMR CNRS 5672, ENS de Lyon, 69364 Lyon, France

4 LIRIS, UMR CNRS 5205, INSA-Lyon, 69621 Villeurbanne, France

5 LIP, UMR CNRS 5668, INRIA and ENS de Lyon, 69364 Lyon, France

6 ENS de Lyon, 69364 Lyon, France

7 ISC-PIF, Paris, France

* E-mail: pablo.jensen@ens-lyon.fr

Abstract

Thanks to a large database ($\sim 215\,000$ 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 (Hayden, 2010; Anderson, 1972; Grauwin *et al.*, 2009a; Gannon, 2007). 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*” (Von Bertalanffy, 1976). 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” (Solé, 2000). 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 inquiry¹?

Results

In this paper, we empirically study the “complex systems” field using the quantitative tools developed to understand the organization of scientific fields (Small, 1999) and their evolution (Glänzel, 2003; Chavalar-ias & Cointet, 2009; Cambrosio *et al.*, 2006). Global science maps (Small, 1999; Klavans & Boyack,

¹The web site of the Santa Fe complex systems institute (<http://www.santafe.edu/about/>, accessed June 1st, 2010) defines its aim as promoting “multidisciplinary collaborations in the physical, biological, computational, and social sciences.”

2009; Small, 1973; Börner & Schernhost, 2009; Börner, 2010; Noyons, 2004; Leydesdorff & Rafols, 2009, 2010; Agarwal & Skupin, 2008; Cobo *et al.*, 2011; Leydesdorff & Persson, 2011; Grauwin & Jensen, 2011) have become feasible recently, offering a tentative overall view of scientific fields and fostering dreams of a “science of science” (Börner & Schernhost, 2009). Specifically, to collect a representative database of articles, we selected from the ISI Web of knowledge (<http://apps.isiknowledge.com/>) all records containing topic keywords relevant for the field of complex systems (Table 1). Table 2 contains the 20 most frequent cited references and journals within our dataset. To analyze the data, we build a network (Börner & Schernhost, 2009) in which the $\sim 215\,000$ articles are the nodes. These nodes are linked according to the proportion of shared references (bibliographic coupling (Kessler, 1963)). 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 Bastian *et al* (2009); Jacomy *et al* (2011). We first note that all important complex systems subfields² are present³. 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 (Fortunato & Barthélemy, 2007) 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 3. 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 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 $\mathcal{N}(r)$ as the sum, over all pairs of subfields, of the proportion of their links explained by that reference (see Methods). Table 4 shows that the references that glue the network are more methodological than theoretical: the most networking reference is “Numerical Recipes” (Press *et al.*, 2010), 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

²In the following, we use italics to refer to the names of the communities.

³As can be checked by consulting authoritative CS web sites such as Santa Fe’s and its “Exploring complexity” lectures: <http://www.santafe.edu/news/item/exploring-complexity-science-and-technology-santa-fe-institute-perspective/> (accessed June 1st, 2010).

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 (Mehdi *et al*, 2009). 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 (Borgatti *et al*, 2009) and biological (Fox-Keller, 2005) sciences emphasize the variations in structure across different groups or contexts and use these differences to explain differences in outcomes. Physicians are interested in practical medical advances while physicists want to know whether physiological rhythms are chaotic or not (Glass, 2001).

Where do the links come from then? In an illuminating analogy, Peter Galison (Galison, 1997) 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 (Galison, 1997) 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”.

Table 5 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*” (Eisen *et al*, 1998) while the second “*describes the application of self-organizing maps for recognizing and classifying features in complex, multidimensional [transcriptomic] data*” (Tamayo *et al*, 1999). 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 analysis* and *biology* (mainly *transcriptomics*) are similar to references used by *transcriptomics data analysis* papers themselves. This means that the link arises from biologists citing results obtained by *transcriptomics data analysis* scientists or techniques they use. On the other hand, the common references between *transcriptomics 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 *transcriptomics data analysis* scientists citing classification techniques created by *self-organizing maps* scientists, while these scientists do not often use *transcriptomics data analysis* references. Therefore, *transcriptomics 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 (Gillespie, 1977) 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 (Tyson *et al*, 2010): “*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 *transcriptomics 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.

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 (Edmonds, 2010). Certainly, a few theoretical papers, such as Bak’s (Bak *et al.*, 1987) (in SOC) or Albert and Barabasi’s (Albert & Barabasi, 2002) (in CN) point to “universal” mechanisms and are heavily cited. However, more than 90% of their citations arise from modelers themselves⁴, suggesting that they may be universal... for theorists. Our data support the local character of these “universal” laws. First, Albert and Barabasi’s (Albert & Barabasi, 2002) 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 legitimating 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 (Barabasi & Oltvai, 2004). 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 (Fox-Keller, 2005; Stanley *et al.*, 2000). 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”⁵. 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 (Fox-Keller, 2009; Castellano *et al.*, 2009; Solé *et al.*, 1999), but arguments for the usefulness of such an approach for biological or social systems are unconvincing (Edmonds, 2010).

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 (Fox-Keller, 2009). This idea was fecund in that it gave birth to several active subdisciplines (*cytoskeleton SO*, *growth SO*...) (Figure 2). However, it should be noticed that these heirs of self-organization are nowadays almost unrelated. The different *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

⁴Counted on Web of Science (January 28th 2011) by analyzing the citing papers by discipline. Specifically, “Subject Areas” (Web of Science name for subdisciplines) related to Physics (such as Physics Multidisciplinary or Physics Mathematical) account for 2272 out of 3158 citations (72%) for Bak’s paper (Bak *et al.*, 1987) while more “applied” subject areas (such as Geosciences and Applied mathematics) account for 676 citations (21%). Subject areas related to Physics or Mathematics account for 4201 out of 5281 citations (80%) of Albert and Barabasi’s paper (Albert & Barabasi, 2002), while subject areas related to biology or more applied fields account for 1060 citations (20%). Instead, areas related to biology account for most of citations to Barabasi and Oltvai’s introduction of network theory to biologists (Barabasi & Oltvai, 2004).

⁵<http://www.santafe.edu/about/> (accessed June 1st, 2010).

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 (For a telling example of this distinction, see Venturini (2008)). 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 (Benner & Sismour, 2005). But many difficulties arise from the intertwining of the elements in a living being (Serrano, 2007). 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 (Schweber & Wakchters, 2000; Fox-Keller, 2009) 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 (Pickering, 1995; Schweber & Wakchters, 2000; Bowker, 1993). 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” (Palla *et al.*, 2007); “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” (Helbing, 2009) or “[Systems biology] leads to a future where biology and medicine are transformed into precision engineering” (Kaneko, 2002).

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 led 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” (Galison, 1997). 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 (Lazer, 2009; Microsoft, 2006) 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 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 1). 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 1, 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 *Apoptosis* or *Immunology*) that lie at the network’s edges and do not bias the results.

Links between articles

Weight of links between articles are calculated through their common references (bibliographic coupling (Kessler, 1963)). 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.

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 (Blondel *et al*, 2008), which is designed to efficiently maximize the modularity function in large networks. More precisely, we used the weighted modularity Q (Newman, 2004; Fortunato, 2010), which is defined as $Q = \sum_I q_I$, 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

$$\begin{aligned} \Omega_{II} &= \frac{1}{2} \sum_{i \in I, j \in I} \omega_{ij} && \text{is the total weight of the links inside community } I, \\ \Omega_{IJ} &= \sum_{i \in I, j \in J} \omega_{ij} && \text{is the total weight of the links between communities } I \text{ and } J \neq I \\ \Omega &= \frac{1}{2} \sum_{i,j} \omega_{ij} = \sum_{(i,j)} \omega_{ij} && \text{is the total weight of the links of the graph.} \end{aligned}$$

Each module q_I compares the relative weight of edges $\frac{\Omega_{II}}{\Omega}$ inside a community I with the expected weight of edges $\left(\frac{\sum_{J \neq I} \Omega_{IJ} + 2\Omega_{II}}{2\Omega}\right)^2$ 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 (Fortunato & Barthélemy, 2007) 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 (Fortunato & Barthélemy, 2007) 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 Barthélemy (Fortunato & Barthélemy, 2007) to check their relevance (see Table 3).

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 $j \in J$:

$$\langle w \rangle_{IJ}^{-1} = \langle w_{ij} \rangle_{i \in I, j \in J}^{-1} = (\Omega_{IJ} / N_I N_J)^{-1} \quad (3)$$

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 J . To take into account this similarity, we define the *orientation* of a community-community link in the following way.

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:

$$\begin{aligned} \cos_{II,IJ} &= \frac{\sum_r L_{r,II} L_{r,IJ}}{\sqrt{\sum_r L_{r,II}^2 \sum_r L_{r,IJ}^2}} && \text{comparing the shared refs to those of } I \\ \cos_{JJ,IJ} &= \frac{\sum_r L_{r,JJ} L_{r,IJ}}{\sqrt{\sum_r L_{r,JJ}^2 \sum_r L_{r,IJ}^2}} && \text{comparing the shared refs to those of } J \end{aligned}$$

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 Bastian *et al* (2009). The layout of the graph is obtained thanks to a spring-based algorithm implemented in it Jacomy *et al* (2011). 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 subdisciplines 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:

$$\mathcal{N}(r) = \frac{1}{Z} \sum_{I \neq J} f_{IJ}(r) \quad (4)$$

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 Andreï Mogoutov, Léo Granger, Taras Kowaliw and Eric Bertin.

References

- Agarwal, P., Skupin, A. (2008). Self-organising maps: applications in geographic information science, Ed. John Wiley & Sons Inc.
- Albert, R. & Barabási, A.L. (2002). Statistical mechanics of complex networks. *Rev. Mod. Phys.* 74, 47-97.
- Anderson, P.W. (1972). More is Different. *Science*, 177, 393-396.
- Bak, P., Tang, C., Wiesenfeld, K. (1987). Self-organized criticality: an explanation of $1/f$ noise. *Phys Rev Lett* 59, 381-384.
- Barabási, A.L., Oltvai, Z. (2004). Network biology: understanding the cell’s functional organization. *Nature Rev. Genet.* 5:101-113
- Bastian, M., Heymann, S., Jacomy, M. (2009). Gephi: An Open Source Software for Exploring and Manipulating Networks. *Proc 3rd Intl ICWSM Conf.*
- Benner, S.A., Sismour, A.M. (2005). Synthetic biology. *Nat Rev Genet.* 6, 533-543.
- Von Bertalanffy, L. (1976) *General System Theory: Foundations, Development, Applications* (George Braziller).
- Blondel, V.D., Guillaume, J.-L., Lambiotte, R., Lefebvre, E. (2008). Fast unfolding of communities in large networks. *J Stat Mech* P10008.

- Borgatti, S.P., Mehra, A., Brass, D.J., Labianca, G. (2009). Network Analysis in the Social Sciences. *Science* 323, 892-895.
- Börner, K., Scharnhorst, A. (2009). Visual conceptualizations and models of science. *Journal of Informetrics* 3, 161-172.
- Börner, K. (2010). *Atlas of Science: Visualizing What We Know*. MIT Press.
- Bowker, G. (1993). How to Be Universal: Some Cybernetic Strategies, 1943-70. *Social Studies of Science*, 23, 107-127.
- Cambrosio, A., Keating, P., Mercier, S., Lewison, G. & Mogoutov, A. (2006). Mapping the emergence and development of translational cancer research. *European Journal of Cancer* 42, 3140-3148.
- Castellano, C., Fortunato, S., Loreto, V. (2009). Statistical physics of social dynamics. *Rev Mod Phys*, 81, 591-646.
- Chavalarias, D., Cointet, J.-P. (2009). The Reconstruction of Science Phylogeny, arXiv:0904.3154v3.
- Cobo, M. J., Lpez-Herrera, A. G., Herrera-Viedma, E., & Herrera, F. (2011). Science mapping software tools: review, analysis, and cooperative study among tools. *J Am Soc Inform Sci Tech*, 62, 1382-1402.
- Edmonds, B. (2010). Complexity and Context-Dependency, available at <http://bruce.edmonds.name/cacd/cacd.pdf>
- Eisen, M.B., Spellman, P.T., Brown, P.O., Botstein, D. (1998). Cluster analysis and display of genome-wide expression patterns. *Proc Natl Acad Sci USA* 95, 14863-14868.
- Fortunato, S. Barthélemy, M. (2007). Resolution limit in community detection. *Proc Natl Acad Sci USA* 104:36-41.
- Fortunato, S. (2010). Community detection in graphs. *Physics Reports*, 486:75-174.
- Fox-Keller, E. (2005). Revisiting “scale-free” networks. *BioEssays* 27, 1060-1068.
- Fox-Keller, E. (2009). Organisms, Machines, and Thunderstorms, a history of Self-Organization. *Historical studies in the natural sciences*, 38-39.
- Galison, P. (1997). *Image and Logic: A Material Culture of Microphysics* (University Of Chicago Press).
- Gannon, F. (2007). Too complex to comprehend? *EMBO Rep* 8:705.
- Gillespie, D.T. (1977). Exact Stochastic Simulation of Coupled Chemical Reactions. *J Phys Chem* 81(25), 2340-2361.
- Girvan, M., Newman, M.E.J. (2004). Finding and evaluating community structure in networks. *Phys Rev E* 69:026113.
- Glänzel, W. (2003). Bibliometrics as a research field: A course on theory and application of bibliometric indicators, available at http://nsdl.niscair.res.in/bitstream/123456789/968/1/Bib_Module_KUL.pdf
- Glass, L. (2001). Synchronization and rhythmic processes in physiology. *Nature* 410, 277-284.
- Grauwin, S., Bertin, E., Lemoy, R. & Jensen, P. (2009). Competition between collective and individual dynamics. *Proc Natl Acad Sci USA*, 106, 20622-20626.

- Grauwin, S. & Jensen P. (2011). Mapping scientific institutions. *Scientometrics*, 89(3), 943-954.
- Hayden, E.C. (2010). Life is complicated. *Nature*, 464, 664-667.
- Helbing, D. quoted in: Cho, A. (2009) Ourselves and Our Interactions: The Ultimate Physics Problem? *Science*, 325, 406-408.
- Jacomy, M., Heymann, S., Venturini, T. and Bastian, M., "ForceAtlas2, a graph layout algorithm for handy network visualization", preprint available at http://webatlas.fr/tempshare/ForceAtlas2_Paper.pdf
- Kaneko, (2002). Computational systems biology. *Nature*, 420, 206-210.
- Kessler, M.M. (1963). Bibliographic coupling between scientific papers. *American Documentation* 24:123-131.
- Klavans, R. & Boyack, K.W. (2009). Toward a consensus map of science. *J Am Soc Inform Sci Tech*, 60(3), 455-476.
- Lazer, D. *et al* (2009). Computational social science. *Science* 323:721-723.
- Leydesdorff, L., Rafols, I. (2009). A Global Map of Science Based on the ISI Subject Categories. *J Am Soc Inform Sci Tech*, 60(2).
- Leydesdorff, L. & Persson, O. (2010). Mapping the Geography of Science: Distribution Patterns and Networks of Relations among Cities and Institutes. *Journal of the American Society for Information Science & Technology*, 61(8), 1622-1634.
- Microsoft, (2006). Towards 2020 Science. Available at <http://research.microsoft.com/towards2020science>
- Moussaid, M., Garnier, S., Theraulaz, G. & Helbing, D. (2009). Collective Information Processing and Pattern Formation in Swarms, Flocks, and Crowds. *Topics in Cognitive Science* 1, 469-497.
- Nicolis, G., Prigogine, I. (1977). *Self-organization in nonequilibrium systems* (Wiley New York).
- Noyons, E.C.M. (2004). Science maps within a science policy context. In: Moed, H.F, Glanzel, W., Schmoch, U. (Eds.), *Handbook of Quantitative Science and Technology Research, The Use of Publication and Patent Statistics in Studies of S&T Systems*, pp. 237-256. Dordrecht: Kluwer Academic publishers.
- Palla, G., Barabási, A.L., Vicsek, T. (2007). Quantifying social group evolution. *Nature*, 446, 664-667.
- Pickering, A. (1995). *The mangle of practice* (University Of Chicago Press).
- Press, W.H. et al. (2007, 3rd edition). *Numerical Recipes. The Art of Scientific Computing* (Cambridge University Press, NY).
- Rafols, I., Leydesdorff, L. (2010). Science Overlay Maps: A New Tool for Research Policy and Library Management. *J Am Soc Inform Sci Tech*, 61(9), 1871-1887.
- Schweber, S., Wakhters, M. (2000). *Complex Systems, Modelling and Simulation. Studies in History and Philosophy of Science*, 31, 583-609.
- Serrano, L. (2007). Synthetic biology: promises and challenges. *Mol Syst Biol*. 3,158.
- Small, H., (1999). Visualizing Science by Citation Mapping. *J Am Soc Inform Sci*, 50(9), 799-813.

- Small, H. (1973). Co-citation in the scientific literature: A new measure of the relationship between two documents. *J Am Soc Inform Sci*, 24, 265-269.
- Solé, R.V., Manrubia, S.C., Benton, M., Kauffman, S., Bak, P. (1999). Criticality and scaling in evolutionary ecology. *Trends Ecol Evol* 14, 156-160.
- Solé, R.V. (2000). Quoted in the New York Times, December 26, First Cells, Then Species, Now the Web. (<http://www.nytimes.com/2000/12/26/science/26WEBS.html?printpage=yes?pagewanted=1>, last accessed Sept. 27th, 2010).
- Stanley, H.E. et al. (2000). Scale Invariance and Universality: Organizing Principles in Complex Systems. *Physica A*, 281, 60-68.
- Tamayo, P. et al (1999). Interpreting patterns of gene expression with self-organizing maps: Methods and application to hematopoietic differentiation. *Proc Natl Acad Sci USA* 96:2907-2912.
- Tyson, J.J., Chen, K.C., Novak, B. (2003). Sniffers, buzzers, toggles and blinkers: dynamics of regulatory and signaling pathways in the cell. *Curr Opin Cell Biol*, 15, 221-231.
- Venturini, T. (2008), Terminator contro terminator, in Bella G. e Diamantini D. (eds.) *Studiare la Società dell'Informazione*. Milano: Guerini e Associati.

Figure Legends

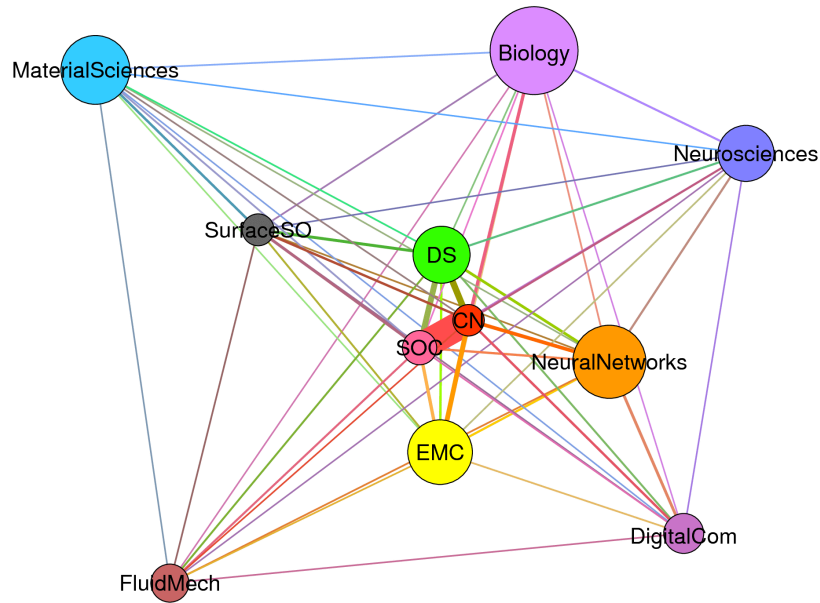


Figure 1: **Community structure obtained with a first run of the modularity maximization (Blondel *et al*, 2008) on the 2000-2008 network (141 098 articles).** The surface of a community I is proportional to its number of articles N_I and the width of the link between two communities I and J is proportional to the mean bibliographic coupling $\langle \omega \rangle_{IJ} = \sum_{i \in I, j \in J} \omega_{ij} / N_I N_J$. The layout of the graph is obtained thanks to a spring-based algorithm implemented in the Gephi visualization software (Bastian *et al*, 2009; Jacomy *et al*, 2011). For the sake of clarity, communities with less than 300 articles are not displayed. The label of a community represents the most frequent and/or significant keyword of its articles. *CN* stands for Complex Networks, *SOC* for Self Organized Criticality, *DS* for Dynamical Systems, *DigitCom* for Digital Communication and *SurfaceSO* for Self-organization on Surfaces. *EMC* is a more composite community where the three most representative keywords are *Ecology*, *Management* and *Computational Models*. See Figure 2 for details.

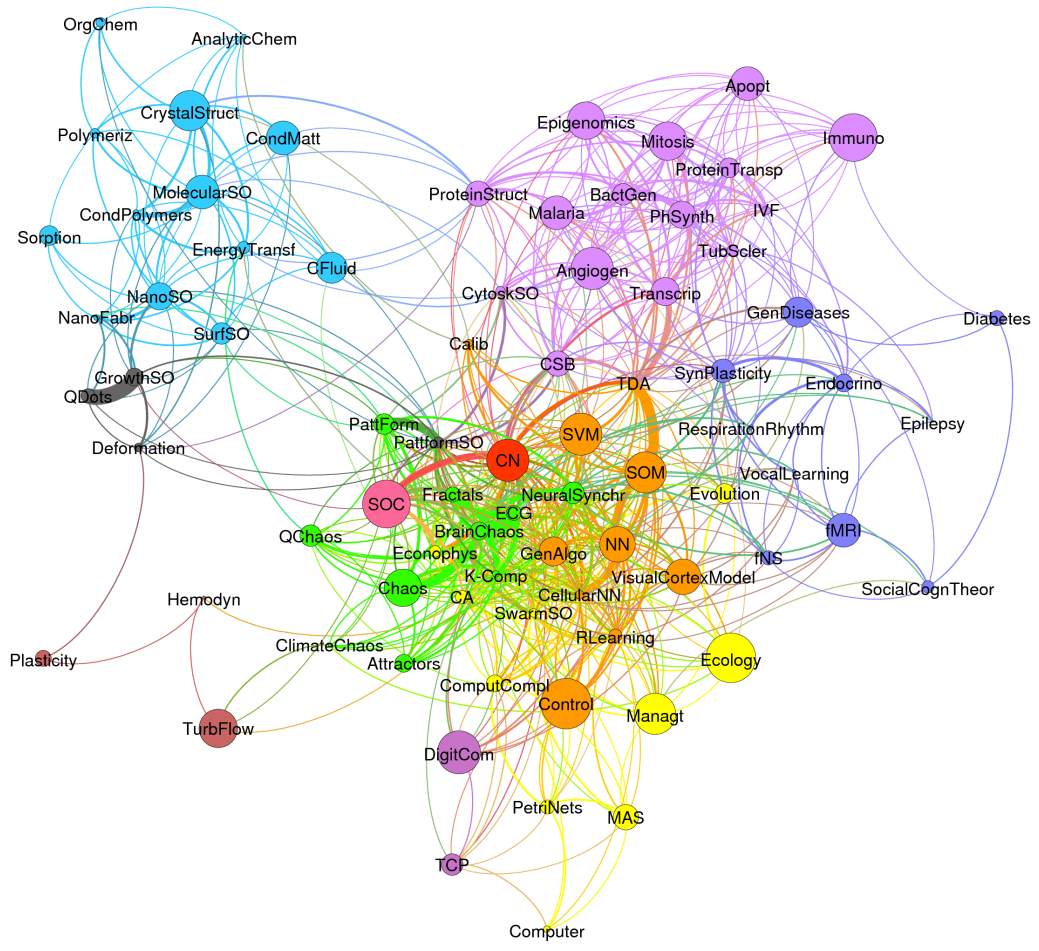


Figure 2: **Community structure obtained with a second run of the modularity maximization on the 2000-2008 network.** This community structure is obtained by optimizing the internal modularity Q^i of each community obtained by the first run of the modularity maximization algorithm on the 2000-2008 network, 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 (Bastian *et al* , 2009; Jacomy *et al* , 2011). 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 $2 \cdot 10^{-5}$ 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.

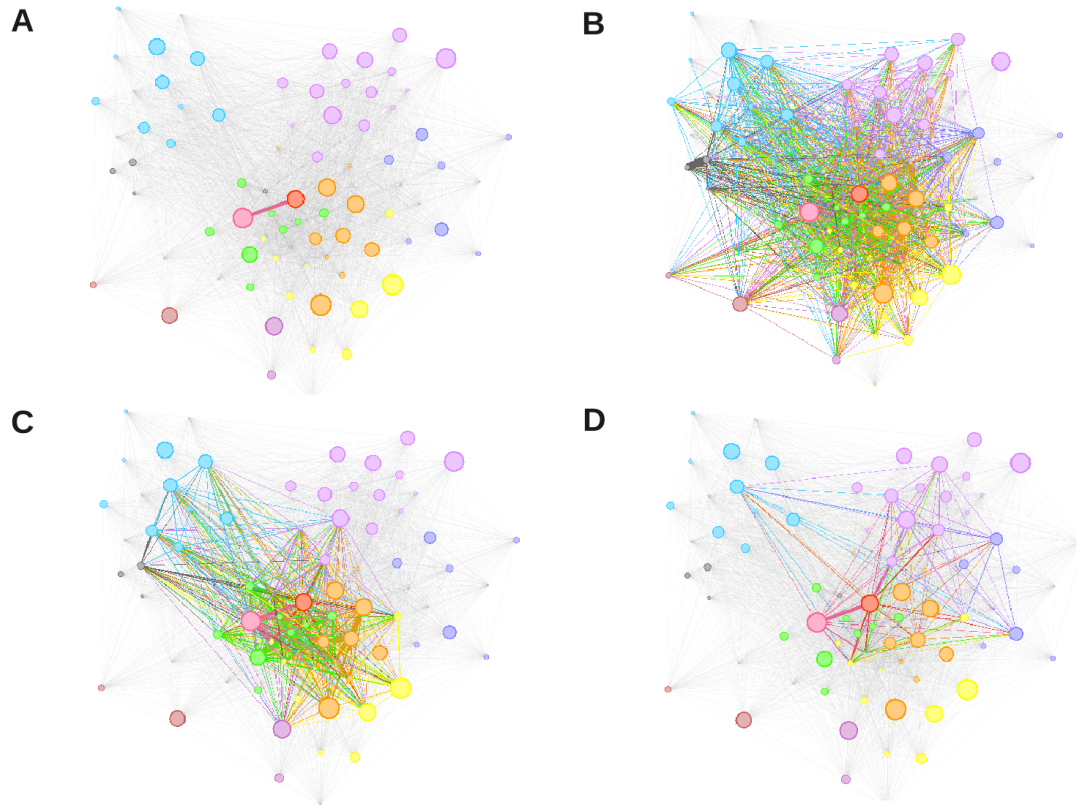


Figure 3: **Local “networking” force for four different references on the 2000-2008 network (Fig 2).** Links established using the reference are shown in color. The number of citations corresponds to those included in papers of our database published between 2000 and 2008 **a.** Ref used: Albert & Barabasi (Albert & Barabasi, 2002) (2058 citations) **b.** Ref used: Press WH *et al*, Numerical Recipes - all editions (Press *et al*, 2010) (1267 citations) **c.** Ref used: Nicolis (Nicolis & Prigogine, 1977) (342 citations) **d.** Ref used: Barabasi and Oltvai (Barábasi & Oltvai, 2004) (244 citations).

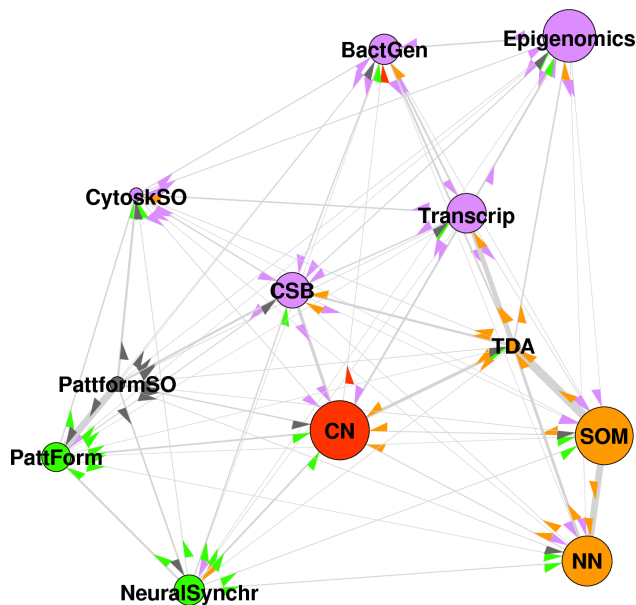


Figure 4: **Directed network.** On this subset of the graph presented in Figure 2, the arrows are directed to the subfield that uses the other subfield’s references to establish the link. More precisely, the common references shared by two linked subfields are more similar to the internal references 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.

Tables

Table 1: 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.

topic keywords	Results	topic keywords	Results
“self organ*”	32484	“multifractal”	390
“complex network*”	6953	“multiscale”	1439
“dynamical system”	8205	“neural network*”	12747
“econophysics”	633	(“non linear*” OR “nonlinear*”)	
“strange attractor”	769	NOT “equation*”	10240
“synergetics”	379	“non linear dynamic*”	560
“adaptive system*”	1141	“non linear system*”	391
“artificial intelligence”	1812	“nonlinear dynamic*”	2285
“attractor”	1034	“nonlinear system*”	1826
“bifurcation”	3164	“phase transition”	5503
“chaos“	5370	“plasticity”	6667
“control“	116017	“random walk”	758
“criticality”	980	“robustness”	6498
“ecology”	5869	“scaling”	7008
“economics”	2243	“social system*”	586
“epistemology”	345	“spin glass*”	643
“far from equilibrium”	253	“stability” AND (“lyapunov” OR	
“feedback”	12881	“non linear*” OR “nonlinear*”)	1399
“fractal”	3867	“stochastic”	9184
“ising”	975	“synchronization”	4645
“multi agent”	2032	“turbulence”	4602
“multiagent”	665	“universality”	861
“multi scale”	779	“cell* automat*”	1659

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 2: The 20 references (including books and articles) and journals which are the more used by the articles of the whole database

Reference	Times used	Journal (# distinct refs)	Times used
Bak P, 1987, PHYS REV LETT (59)	2131	NATURE (29166)	169309
Albert R, 2002, REV MOD PHYS (74)	2050	P NATL ACAD SCI USA (42504)	151140
Laemmler UK, 1970, NATURE (227)	1762	J BIOL CHEM (59436)	149042
Watts DJ, 1998, NATURE (393)	1732	SCIENCE (24880)	148002
Barabasi AL, 1999, SCIENCE (286)	1693	CELL (11044)	99168
Bak P, 1988, PHYS REV A (38)	1555	PHYS REV LETT (23269)	94861
Sambrook J, 1989, MOL CLONING LAB MANU	1439	J AM CHEM SOC (29807)	82569
Newman MEJ, 2003, SIAM REV (45)	1308	EMBO J (10926)	53049
Bradford MM, 1976, ANAL BIOCHEM (72)	1255	MOL CELL BIOL (12866)	52694
Lowry OH, 1951, J BIOL CHEM (193)	1106	J NEUROSCI (12313)	43152
Rumelhart DE, 1986, PARALLEL DISTRIBUTED (1)	947	J IMMUNOL (18891)	41496
Strogatz SH, 2001, NATURE (410)	907	PHYS REV B (19367)	41450
Kohonen T, 1982, BIOL CYBERN (43)	901	J CELL BIOL (10239)	40560
Chomczynski P, 1987, ANAL BIOCHEM (162)	849	J CHEM PHYS (17136)	40074
Goldberg DE, 1989, GENETIC ALGORITHMS S	822	GENE DEV (4879)	38903
Lorenz EN, 1963, J ATMOS SCI (20)	726	BIOCHEMISTRY-US (16035)	32061
Mandelbrot BB, 1982, FRACTAL GEOMETRY NAT	721	BRAIN RES (15364)	30517
Kohonen T, 1990, P IEEE (78)	715	ANGEW CHEM INT EDIT (7572)	27718
Dorogovtsev SN, 2002, ADV PHYS (51)	688	NUCLEIC ACIDS RES (9738)	27242
Albert R, 2000, NATURE (406)	678	J EXP MED (8100)	27220

Table 3: 2000-2008 subfields' sizes N , inner coherences $\langle \omega \rangle^{-1}$ and modules q .

Subfield	N	$\langle \omega \rangle^{-1}$	q
Analytic Chemistry (AnalyticChem)	419	336.14	0.0002
Angiogenesis (Angiogen)	3642	1408.78	0.0005
Apoptosis (Apopt)	2632	1424.23	0.0026
Attractors	1161	524.42	0.0013
Bacterial Genomics (BactGen)	1517	120.19	0.0103
Brain Chaos	1175	70.06	0.0105
Calibration (Calib)	538	371.17	0.0004
Cellular Automata(CA)	846	164.26	0.0023
Cellular Neural Networks (CellularNN)	620	86.17	0.0024
Chaos	3134	531.3	0.0099
ClimateChaos	352	205.31	0.0003
Complex Fluids (CFluid)	2310	994.14	0.0029
Complex Networks (CN)	3684	21.87	0.2235
Computational Complexity (ComputCompl)	1134	379.92	0.0018
Computational Systems Biology (CSB)	1799	323.99	0.0053
Computer	526	437.08	0.0003
Condensed Matter (CondMatt)	2629	631.99	0.0059
Condensed Matter - Polymers (CondPolymers)	471	131.85	0.0009
Control	4772	1086.35	0.0112
Crystal Structure (CrystalStruct)	3386	350.41	0.0175
Cytoskeleton Self-Organization(CytoskSO)	651	132.58	0.0017
Deformation	590	432.48	0.0004
Diabetes	1015	669.24	0.0008
Digital Communication (DigitCom)	3811	470.56	0.0165
Ecology	4751	1846.16	0.0066
Econophysics (Econophys)	738	96.12	0.0003
Electrocardiogram (ECG)	987	117.68	0.0044
Endocrinology (Endocrino)	1223	607.63	0.0013
Energy Transfer (EnergyTransf)	801	258.27	0.0013
Epigenomics	3055	677.49	0.0074
Epilepsy	316	273.67	0.0001
Evolution	1318	796.76	0.0011
Fractals	1015	192.02	0.0029
Functional MRI (fMRI)	2634	897.28	0.0041
Functional Neurosciences (fNS)	935	497.44	0.0009
Genetic Algorithm (GenAlgo)	2177	197.96	0.0128
Genetic Diseases (GenDiseases)	2273	387.34	0.0072
Growth Self-Organization (GrowthSO)	1192	113.84	0.0067
Hemodynamics (Hemodyn)	346	344.37	0.0001
Immunology (Immuno)	4403	2234.66	0.0047
In Vitro Fertilization (IVF)	591	281.01	0.0006
Kolmogorov Complexity (K-Comp)	501	121.94	0.0011
Malaria	2702	747.44	0.0052
Management (Managt)	3563	2159.31	0.0031
Mitosis	3171	564.98	0.0095
Molecular Self-Organization (MolecularSO)	2684	409.08	0.0094
Multi-agent System (MAS)	1787	1094.91	0.0015
Nanofabrication (NanoFabr)	457	45.28	0.0025
Nanosciences (Nano)	1995	418.01	0.0051
Neural Networks (NN)	2902	221.15	0.0201
Neural Synchronization (NeuralSynchr)	1451	453.59	0.0025
Organic Chemistry (OrgChem)	649	368.67	0.0006
Pattern Formation (PattForm)	1403	205.82	0.0051
Pattern Formation & Self-Organization (PattformSO)	691	142.82	0.0018
Petri Nets	957	275.76	0.0018
Photosynthesis (PhSynth)	2000	224.48	0.0096
Plasticity	1066	915.05	0.0006
Polymerization (Polymeriz)	645	98.54	0.0022
Protein Structure (ProteinStruct)	1830	237.07	0.0076
Protein Transport (ProteinTransp)	1305	308.77	0.0029
Quantum Chaos (QChaos)	1456	636.22	0.0018
Quantum Dots (QDots)	921	130.93	0.0035
Reinforcement Learning (RLearning)	891	287.57	0.0014
Respiration Rhythm	416	57.35	0.0016
Self-Organized Criticality (SOC)	4447	199.3	0.0509
Self-Organizing Maps(SOM)	3495	168.85	0.0376
Social Cognition Theory (SocialCognTheor)	800	680.59	0.0005
Sorption	1354	925.37	0.001
Support Vector Machines(SVM)	3660	867.91	0.0082
Surface Self-Organization (SurfSO)	1511	468.34	0.0026
Swarm Intelligence (SwarmIntel)	608	145.94	0.0013
Synaptic Plasticity (SynPlasticity)	1625	370.25	0.0038
Transcriptomics (Transcrip)	2043	439.37	0.0051
Transcriptomics Data Analysis (TDA)	628	43.32	0.0049
Transmission Control Protocol (TCP)	1473	718.93	0.0016
Tuberous Sclerosis (TubScler)	766	153.33	0.002
Turbulent Flow (TurbFlow)	3172	1212.32	0.0045
Visual Cortex Model	2851	845.22	0.0051
VocalLearning	389	162.56	0.0005

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 $\langle \omega \rangle^{-1}$ can be taken as an inner coherence measure. Indeed, would the weight of these links be homogeneously distributed between all pairs of articles of a given subfield, then two articles of this subfield chosen at random would share 1 reference over $\langle \omega \rangle^{-1}$.

Table 4: **The 20 most networking references in the 2000-2008 decade**

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

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 $\mathcal{N}(r)$ have been summed. The complete references of these papers are given in Supplementary Information.

Table 5: **Strongest trading zones.**

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

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 \bar{J} is the total weight of the article-article links between the subcommunity I and community \bar{J} , normalized by the total weight of the the article-article links between \bar{I} and \bar{J} : $T_{\bar{I}\bar{J}}(I) = \sum_{i \in I, j \in \bar{J}} \omega_{ij} / \sum_{i \in \bar{I}, j \in \bar{J}} \omega_{ij}$. The expected force T^{exp} is the value of the trading force one would expect if all the links between \bar{I} and outside communities were equally shared among all sub-communities of \bar{I} , which is simply the fraction $N_I/N_{\bar{I}}$ of articles of I in \bar{I} . The acronyms of the subfiels used here correspond to those explained in Table 3.