Complexity and Self-organization

prepared for the *Encyclopedia of Library and Information Sciences*,

edited by Marcia J. Bates and Mary Niles Maack (Taylor & Francis, 2008)

Francis Heylighen

Free University of Brussels, Krijgskundestr. 33, 1160 Brussels, Belgium

fheyligh@vub.ac.be

Abstract: this article introduces some of the main concepts and methods of the science studying complex, self-organizing systems and networks, in a non-technical manner. Complexity cannot be strictly defined, only situated in between order and disorder. A complex system is typically modeled as a collection of interacting agents, representing components as diverse as people, cells or molecules. Because of the non-linearity of the interactions, the overall system evolution is to an important degree unpredictable and uncontrollable. However, the system tends to self-organize, in the sense that local interactions eventually produce global coordination and synergy. The resulting structure can in many cases be modeled as a network, with stabilized interactions functioning as links connecting the agents. Such complex, self-organized networks typically exhibit the properties of clustering, being scale-free, and forming a small world. These ideas have obvious applications in information science when studying networks of authors and their publications.

INTRODUCTION

In the last two decades, a new paradigm for scientific inquiry has been emerging: complexity. Classical science, as exemplified by Newtonian mechanics, is essentially reductionist: it reduces all complex phenomena to their simplest components, and then tries to describe these components in a complete, objective and deterministic manner (3,8). The philosophy of complexity is that this is in general impossible: complex systems, such as organisms, societies or the Internet, have properties—emergent properties—that cannot be reduced to the mere properties of their parts. Moreover, the behavior of these systems has aspects that are intrinsically unpredictable and uncontrollable, and cannot be described in any complete manner. At best, we can find certain statistical regularities in their quantitative features, or understand their qualitative behavior through metaphors, models, and computer simulations.

While these observations are mostly negative, emphasizing the traditional qualities that complex systems lack, these systems also have a number of surprisingly positive features, such as flexibility, autonomy and robustness, that traditional mechanistic systems lack. These qualities can all be seen as aspects of the process of self-organization that typifies complex systems: these systems spontaneously organize themselves so as to better cope with various internal and external perturbations and conflicts. This allows them to evolve and adapt to a constantly changing environment.

Processes of self-organization literally create order out of disorder (3). They are responsible for most of the patterns, structures and orderly arrangements that we find in the natural world, and many of those in the realms of mind, society and culture. The aim of information science can be seen as finding or creating such patterns in the immense amount of data that we are confronted with. Initially, patterns used to organize information were simple and orderly, such as "flat" databases in which items were ordered alphabetically by author's name or title, or hierarchically organized subject indices where each item was assigned to a fixed category. Present-day information systems, such as the world-wide web, are much less orderly, and may appear chaotic in comparison. Yet, being a result of self-organization, the web possesses a non-trivial structure that potentially makes information retrieval much more efficient. This structure and others have recently been investigated in the science of networks, which can be seen as part of the sciences of complexity and self-organization.

The concept of self-organization was first proposed by the cyberneticist W. Ross Ashby (1) in the 1940s and developed among others by his colleague Heinz von Foerster (2). During the 1960s and 1970s, the idea was picked up by physicists and chemists studying phase transitions and other phenomena of spontaneous ordering of molecules and particles. These include Ilya Prigogine (3), who received a Nobel Prize for his investigation of self-organizing "dissipative structures", and Hermann Haken (4), who dubbed his approach "synergetics". In the 1980s, this tradition cross-fertilized with the emerging mathematics of non-linear dynamics and chaos, producing an investigation of complex systems that is mostly quantitative, mathematical, and practiced by physicists. However, the same period saw the appearance of a parallel tradition of "complex

adaptive systems" (5), associated with the newly founded Santa Fe Institute for the sciences of complexity, that is closer in spirit to the cybernetic roots of the field. Building on the work of John Holland, Stuart Kauffman, Robert Axelrod, Brian Arthur and other SFI associates, this approach is more qualitative and rooted in computer simulation. It took its inspiration more from biology and the social sciences than from physics and chemistry, thus helping to create the new disciplines of *artificial life* and *social simulation*. The remainder of this paper will mostly focus on this second, simulation-based tradition, because it is most applicable to the intrinsically social and cognitive processes that produce the systems studied by information science. Although the other, mathematical tradition sometimes uses the term "complex systems" to characterize itself, the labels of "non-linear systems" or "chaos theory" seem more appropriate, given that this tradition is still rooted in the Newtonian assumption that apparently complex behavior can be reduced to simple, deterministic dynamics—an assumption which may be applicable to the weather, but not to the evolution of a real-world social system. Extending both traditions, the turn of the century witnessed a surging popularity of research into complex networks. This was inspired mostly by the growth of the world-wide web and the models proposed by Watts and Strogatz (6), and Barabasi and Albert (7).

At present, the "science of complexity" taken as whole is still little more than a collection of exemplars, methods and metaphors for modeling complex, self-organizing systems. However, while it still lacks integrated theoretical foundations, it has developed a number of widely applicable, fundamental concepts and paradigms that help us to better understand both the challenges and opportunities of complex systems. The present article will try to introduce the most important of these concepts in a simple and coherent manner, with an emphasis on the ones that may help us to understand the organization of networks of information sources.

COMPLEX SYSTEMS

There is no generally accepted definition of complexity (13): different authors have proposed dozens of measures or conceptions, none of which captures all the intuitive aspects of the concept, while they are applicable only to a very limited type of phenomena, such as binary strings or genomes. For example, the best-known measure, "Kolmogorov complexity", which is the basis of algorithmic information theory, defines the complexity of a string of characters as the length of the shortest program that can generate that string. However, this implies that random strings are maximally complex, since they allow no description shorter than the string itself. This contradicts our intuition that random systems are not truly complex. A number of more complex variations on this definition have been proposed to tackle this issue, but they still suffer from the fact that they are only applicable to strings, not to real-world systems. Moreover, it has been proven that the "shortest possible" description is in general uncomputable, implying that we can never be sure that we really have determined the true complexity of a string.

In spite of these fundamental problems in formalizing the notion of complexity, there are a number of more intuitive features of complex systems that appear again and again in

the different attempts to characterize the domain (8). One that is more or less universally accepted is that complexity must be situated *in between order and disorder*: complex systems are neither regular and predictable (like the rigid, "frozen" arrangement of molecules in a crystal), nor random and chaotic (like the ever changing configuration of molecules in a gas). They exhibit a mixture of both dimensions, being roughly predictable in some aspects, surprising and unpredictable in others. This intermediate position, balancing between rigidity and turbulence, is sometimes called the "edge of chaos". A number of theorists have proposed that this precarious balance is precisely what is necessary for adaptation, self-organization, and life to occur, and that complex systems tend to spontaneously evolve towards this "edge" (5).

Another fundamental feature is that complex systems consist of many (or at least several) parts that are connected via their interactions. Their components are both *distinct* and *connected*, both autonomous and to some degree mutually dependent. Complete dependence would imply order, like in a crystal where the state of one molecule determines the state of all the others. Complete independence would imply disorder, like in a gas where the state of one molecule gives you no information whatsoever about the state of the other molecules.

The components of a complex system are most commonly modeled as *agents*, i.e. individual systems that act upon their environment in response to the events they experience. Examples of agents are people, firms, animals, cells and molecules. The number of agents in the system is in general not fixed as agents can multiply or "die". Usually, agents are implicitly assumed to be goal-directed: their actions aim to maximize their individual "fitness", "utility" or "preference". When no specific goal can be distinguished, their activity still follows a simple cause-and-effect or condition-action logic: an agent will react to a specific condition perceived in the environment by producing an appropriate action. The causal relation or rule connecting condition and action, while initially fixed for a given type of agent, can in some cases change, by learning or evolutionary variation.

The environmental conditions to which an agent reacts are normally affected by other agents' activity. Therefore, an action by one agent will in general trigger further actions by one or more other agents, possibly setting in motion an extended chain of activity that propagates from agent to agent across the system. Such interactions are initially *local*: they start out affecting only the agents in the immediate neighborhood of the initial actor. However, their consequences are often *global*, affecting the system of agents as a whole, like a ripple produced by a pebble that locally disturbs the surface of the water, but then widens to encompass the whole pond.

NON-LINEARITY

The spreading of a wave is not a complex phenomenon, though, because its propagation is perfectly regular and predictable, and its strength diminishes as its reach widens. Processes in complex systems, on the other hand, are often *non-linear*: their effects are not proportional to their causes. When the effects are larger than the causes, we may say that there is an amplification or *positive feedback*: initially small perturbations reinforce

themselves so as to become ever more intense. An example is the spread of a disease, where a single infection may eventually turn into a global pandemic. Another example is the chain reaction that leads to a nuclear explosion. When the effects are smaller than the causes, there is a dampening or *negative feedback*: perturbations are gradually suppressed, until the system returns to its equilibrium state.

Interactions with positive feedback are very *sensitive to their initial conditions*: a change in that condition may be so small that it is intrinsically undetectable, yet result in a drastically altered outcome. This is called the *butterfly effect* after the observation that, because of the non-linearity of the system of equations governing the weather, the flapping of the wings of a butterfly in Tokyo may cause a hurricane in New York. The non-observability of the initial perturbations means that the outcome is *in principle* unpredictable, even if the dynamics of the system were perfectly deterministic: no weather monitoring system can be so accurate that it senses all the movements of butterfly wings… This explains why weather forecasts cannot be truly reliable, especially for the longer term. Positive feedback will amplify small, random fluctuations into wild, unpredictable swings, making the overall behavior of the system chaotic. An illustration can be found in the erratic up-and-down movements of quotations on the stock exchange.

In spite of the omnipresence of fluctuations, most systems around us appear relatively stable and predictable. This is due to the presence of negative feedback, which suppresses the effects of such fluctuations. However, while negative feedback makes a system more predictable, it also makes it less controllable: if we try to change the state of such a system, we may find that our changes are counteracted, and that whatever we do the system always returns to its own "preferred" equilibrium state. Examples can be found in social systems where attempts from leaders or governments to change the behavior often are actively resisted so that they eventually come to nothing.

The dynamics of complex systems typically exhibits a combination of positive and negative feedbacks, so that certain changes are amplified and others dampened. This makes the system's overall behavior both unpredictable and uncontrollable. Moreover, such systems are normally open, which means that they exchange matter, energy and/or information with their wider environment. For example, an economy or ecosystem is dependent on the climate, and the amount of sunlight, rain and heat that it produces. These in-going and out-going flows make the dynamics even more complicated, since we cannot know every external event that may affect the system. For example, a thriving economy or ecosystem may suddenly collapse because of the invasion by a foreign pest. Furthermore, the input of energy (such as sunlight) tends to feed amplification processes, so that they never reach the equilibrium that would otherwise follow the exhaustion of resources.

MODELLING COMPLEX SYSTEMS

For the above reasons, traditional deterministic models (such as systems of partial differential equations) of truly complex systems are in general impracticable (14), if not in principle uncomputable (15). In non-linear systems, simplifying the model by using approximations is dangerous as well. The common way to approximate the effect of

complex interactions by reducing it to the "mean field" (i.e. the average effect of many discrete actions performed by independent agents) can actually lead to fundamental errors. For example, a differential equation representing the "mean field" effect may predict that a certain perturbation will die out because it is too small, while a computer simulation of the individual agents finds that its effect is amplified by positive feedback until it dominates the system (16).

Because of these intrinsic difficulties with mathematical models, complexity researchers typically prefer computer simulations, which, while of course being approximations as well, are easier to manipulate, so that more different factors and variations of the model can be explored. Here, the system's evolution is traced step-by-step by iteratively applying the rules that govern the agents' interactions, thus generating the subsequent states of the system. Such simulations typically include a generator of random variations, to represent the effect of unpredictable perturbations. A typical setting is inspired by the Darwinian mechanism of natural selection, in which the rules that determine an agent's behavior are randomly "mutated" and sometimes recombined with the rules of another agent, after which the "fittest" or best performing agents or rules are selected to carry on, while the others are eliminated. To explore the possible behaviors of the system, many different "runs"—with different initial conditions or random variations during the process—of the simulation are performed. The main variable values for each run are collected. These results are then analyzed statistically to discover recurring trends.

This sometimes produces very robust results, in the sense that all runs, however different in their initial behavior, eventually appear to converge to the same type of stable pattern. In the majority of cases, the outcomes can be classified into a relatively small number of distinct categories. This provides the researchers with a qualitative picture of the most likely results—and hopefully an insight into the factors that promote one outcome rather than another one. It is only exceptionally that no clear pattern can be discerned in the outcomes of the different simulation runs. The reason that complex systems in spite of their intrinsic unpredictability tend to settle into a relatively small set of recognizable behaviors is their inherent tendency to *self-organize*.

SELF-ORGANIZATION

Self-organization can be defined as the spontaneous emergence of global structure out of local interactions. "Spontaneous" means that no internal or external agent is in control of the process: for a large enough system, any individual agent can be eliminated or replaced without damaging the resulting structure. The process is truly *collective*, i.e. *parallel* and *distributed* over all the agents. This makes the resulting organization intrinsically *robust* and resistant to damage and perturbations.

As noted, the components or agents of a complex system initially interact only locally, i.e. with their immediate neighbors. The actions of remote agents are initially independent of each other: there is no correlation between the activity in one region and the activity in another one. However, because all components are directly or indirectly connected, changes propagate so that far-away regions eventually are influenced by what happens here and now. Because of the complex interplay of positive and negative

feedbacks, this remote influence is very difficult to predict and may initially appear chaotic.

To explain the appearance of organization, we need to make one further assumption, namely that the outcome of interactions is not arbitrary, but exhibits a "preference" for certain situations over others. The principle is analogous to *natural selection*: certain configurations are intrinsically "fitter" than others, and therefore will be preferentially retained and/or multiplied during the system's evolution. When the agents are goal-directed, the origin of this preference is obvious: an agent will prefer an outcome that brings it closer to its goals. For example, in a market a firm will prefer the outcome that brings it more profit. In an ecosystem, an animal will prefer an outcome that brings it more food, or that reduces its risk of being attacked by a predator. But even inanimate, physical objects, such as molecules or stones, have an in-built "preference", namely for the state that minimizes their potential energy. Thus, a stone "prefers" the stable state at the foot of a hill to an unstable state on the top. Here, "preference" simply means that the unstable state will sooner or later be abandoned, while the stable one will be retained.

CO-EVOLUTION AND SYNERGY

Given such a preference, it is clear why an individual agent tends to "organize" itself so as to settle down in its preferred situation. The problem is that what is best for one agent is in general not best for the other agents. For example, more profit for a firm generally means less profit for its competitors, and an animal safe from attack by a predator means a predator that goes hungry. However, interaction is in general not a *zero-sum game*: a gain by one party does not necessarily imply an equivalent loss by the other party. In most cases, an outcome is possible in which both parties to some degree gain. For example, a firm may increase its profits by developing a more efficient technology, which it then licenses to its competitors, so that they too become more productive. In that case, we may say that the interaction exhibits *synergy*: the outcome is positive for all parties; all involved agents "prefer" the outcome to the situation without the interaction.

In general, such a collective solution is still a compromise, in the sense that not all agents can maximally realize their preferences. Not all the stones can end up in the same, lowest spot at the bottom of the hill, but they can all end up much lower than they were, by reducing the irregular hill to an even plain. Such a compromise reduces the tension or "conflict" between competing agents. (Such conflict would otherwise lead to instability as every action of the one triggers a counteraction by the other.) In that sense, we may say that the agents have *mutually adapted*; they have *coordinated* their actions so as to minimize friction and maximize synergy.

The achievement of this stable, synergetic state is in general a process of trial-and-error or variation-and-selection. Because agents are independent and interact locally, and because the dynamics of the system is unpredictable, they in general do not know what the effect of their actions on the other agents will be. They can only try out actions because they appear plausible, or even choose them at random, and note which ones bring them closer to their goals. Those actions can then be maintained or repeated, while the others are abandoned. This is the fundamental dynamics of natural selection. The main

difference with traditional Darwinian evolution is that trial-and-error happens simultaneously on different sides: the agents *co-evolve*, the one adapting to the other, until they mutually "fit".

FROM LOCAL TO GLOBAL ORGANIZATION

To shift from local coordination to global organization, we just need to note that all interactions between all agents in the complex system will tend towards such a coherent, stable state, until they are all mutually adapted. This process generally accelerates because of a positive feedback. The reason is that if two or more agents have reached a mutually fit state, this defines a stable assembly to which other agents can now adapt, by trying to "fit" into the assembly as well. The larger the assembly, the more "niches" it has in which other agents can fit. The more agents join the assembly, the larger it becomes, and the more niches it provides for even more agents to join. Thus, the assembly may grow exponentially until it encompasses the global system.

This growth is typically faster when the agents are identical (e.g. molecules of the same substance) or similar (e.g individuals from the same species), because the solution found by one agent will then suit the other agents as well, so that minimal further trial-and-error is needed once a good arrangement is locally found. This typically happens in processes of physical self-organization, such as crystallization, magnetization or the emergence of coherent light in a laser (4). When the agents are all different (e.g. species in an ecosystem), each in turn needs to explore in order to find its unique niche in an environment that continues to evolve, resulting in a much less explosive development.

In the case of identical agents, the global structure that emerges is typically uniform or regular, because the arrangement that is optimal for one agent is also the one optimal for the other agents. As a result, they all tend to settle into the same configuration. An example is a crystal, where all molecules are arranged at regular intervals and in the same orientation. In this case, self-organization produces a perfectly ordered pattern. In cases where the agents are diverse, like in an ecosystem or a market, the resulting structure is much more complex and unpredictable.

GLOBAL DYNAMICS

If we now consider the system as a whole—rather than the agents individually—we may note that the system too undergoes a process of variation. This can be seen as an exploration by the system of different regions of its state space, thus following an intricate trajectory. (The state space of the system is merely the Cartesian product of the state spaces of all its components). Self-organization then means that the system reaches an *attractor*, i.e. a part of the state space that it can enter but not leave. In that sense, an attractor is a region "preferred" by the global dynamics: states surrounding the attractor (the attractor *basin*) are unstable and will eventually be left and replaced by states inside the attractor.

A non-linear system has in general a multitude of attractors, each corresponding to a particular self-organized configuration. If the system starts out in a basin state, it will necessarily end up in the corresponding attractor, so that the long-term behavior can in principle be predicted (assuming we know what the attractor is, which is generally not the case). However, if it starts out in a state in between basins, it still has a "choice" about which basin and therefore which attractor it ends up in, and this will depend on unpredictable fluctuations. An attractor generally does not consist of a single state, but of a subspace of states in between which the system continues to move. The self-organized configuration, while more stable than the configuration before self-organization, is therefore in general not static but full of on-going activity.

Self-organization can be accelerated by augmenting the initial variation that makes the system explore its state space: the more different states it visits, the sooner it will reach a state that belongs to an attractor. The simplest way to increase such variation is to subject the system to random perturbations, i.e. "noise". For example, if you shake a pot filled with beans, the beans will explore a variety of configurations, while tending to settle into the one that is most stable, i.e. where the beans are packed most densely near the bottom of the pot. Thus, shaking will normally reduce the volume taken in by the beans. This principle was called "order from noise" by the cyberneticist von Foerster (2) and "order through fluctuations" by the thermodynamicist Prigogine (3).

EMERGENCE

The pattern formed by the stabilized interactions, mutual "fittings", or "bonds" between the agents determines a purposeful or functional structure. Its function is to minimize friction between the agents, and thus maximize their collective "fitness", "preference" or "utility". Therefore, we may call the resulting pattern "organization": the agents are organized or coordinated in their actions so as to maximize their synergy (4). However, such organization by definition imposes a *constraint* on the agents: they have lost the freedom to visit states outside the attractor, i.e. states with a lower fitness or higher friction. They have to obey new "rules", determining which actions are allowed, and which are not. They have lost some of their autonomy. The resulting mutual dependency has turned the collection of initially independent agents into an organization, i.e. a cohesive whole that is more than the sum of its parts. The goal of this "superagent" is to maximize overall synergy rather than individual utility. In a sense, the agents have turned from selfish individualists into conscientious cooperators. They have become subordinated (or "enslaved" in the terminology of Haken (4)) to the regulations of the collective.

This whole has *emergent* properties, i.e. properties that cannot be reduced to the properties of its parts. For example, a cell has the property of being alive, while the molecules that constitute it lack that property; gold has the properties of being shiny, malleable and yellow, but these properties do not exist for individual gold atoms (8). Rather than the parts individually, emergent properties characterize the pattern of interactions or relations between them. They typically include global or "holistic" aspects, such as robustness, synergy, coherence, symmetry and function.

Different attractor regimes imply different properties for the system obeying that regime. For example, a circulating convection current may rotate clockwise or counterclockwise. Since it cannot be a priori predicted which attractor the system will end up in, the emergent properties of the whole cannot be derived from the properties of its parts alone. Once the attractor regime has stabilized, the behavior of the parts is rather regulated or constrained by the properties of the higher-level whole. This is called *downward causation*. For example, the correspondence between DNA triplets and amino acids in the genetic code is not determined by the chemical properties of the molecules that constitute DNA, but by evolutionary history producing a particular mechanism for "reading" DNA triplets in living cells. A random variation of that history might well have produced a different mechanism and therefore a different code. The languages that different people speak are not determined by the neurophysiology of their brain, but by the self-organization of shared lexicons and grammatical rules within a community of communicating individuals.

While the self-organized whole is intrinsically stable, it is still flexible enough to cope with outside perturbations. These perturbations may push the system out of its attractor, but as long as the deviation is not too large, the system will automatically return to the same attractor. In the worst case, the system is pushed into a different basin but that will merely make it end up in a different attractor. In that sense, a self-organizing system is intrinsically *adaptive*: it maintains its basic organization in spite of continuing changes in its environment. As noted, perturbations may even make the system more robust, by helping it to discover a more stable organization.

COMPLEX NETWORKS

The structure emerging from self-organization can often be represented as a network. Initially, agents interact more or less randomly with whatever other agents happen to pass in their neighborhood. Because of natural selection, however, some of these interactions will be preferentially retained, because they are synergetic. Such a preferentially stabilized interaction may be called a bond, relationship, or *link*. A link couples or connects two agents, in the sense that linked agents preferentially interact with each other. The different links turn the assembly of agents into a *network*. Within the network, the agents can now be seen as *nodes* where different links come together. Perhaps the most intuitive example is a social network, which links people on the basis of friendship, trust or collaboration. Other well-known examples are the Internet, which connects computers via communication links, and the Web, which connects documents via hyperlinks. A more abstract example is the biochemical network that connects the molecules that react with each other within a cell in order to produce further molecules.

It is easy to define an abstract mathematical network. You just need a set N consisting of nodes n_i, and then select any subset L of links from the set of all possible connections between two nodes:

$$L = \{(n_i, n_j)\} \subset N \times N.$$

However, complexity researchers have observed that "natural" (i.e. self-organized rather than artificially designed) or "complex" networks tend to exhibit a number of specific features: they are *scale-free*, *small-world*, and *clustering*. These features are defined statistically: certain configurations of links appear with a much higher probability than chance. We will here try to explain these particular link distributions from the dynamics of self-organization of a network.

RANDOM NETWORKS

Let us assume that we start with a collection N of independent agents (future nodes of the network) that initially interact randomly, thus creating random links. This produces a *random network*, i.e. a network where the links have been selected by chance from the set N × N of all possible connections. Random networks have been extensively studied in mathematics. They exhibit the phenomenon of *percolation*: when links between randomly chosen nodes are added one by one, larger and larger subsets of N become connected into cohesive subnetworks. When more links are added, subnetworks will become connected to each other, defining a larger connected subset. When a certain threshold is passed, all subsets become connected so that there is now just a single connected network. It is said that the network *percolates*: imagine the links as tubes and a liquid being injected into one of the nodes; when the network percolates, the liquid will spread throughout the whole system, because any node is now directly or indirectly connected to any other node by an uninterrupted path or chain of links. Whatever happens in one node of the network can now in principle propagate to every other node in the network.

SMALL-WORLD NETWORKS

The maximum length of the shortest path connecting two nodes in a connected network is called the *diameter* of the network. If the diameter is small relative to the number of nodes, the network is said to be a *small-world network*. The notion derives from the "it's a small world" phenomenon in social networks: two people encountering each other will often find that they have one or more acquaintances in common. Studies of social networks have indicated that it is in general possible to find a short sequence of friend-of-a-friend links connecting two people. It has been estimated that on the scale of the world as a whole, two randomly chosen individuals are unlikely to be more than 6 such links removed from each other ("six degrees of separation").

Whereas random networks have the small-world property, the opposite applies to regular networks. An example of such a network is a two-dimensional lattice or grid, where each node is connected to its 4 direct neighbors (left-right-up-down), each of which is connected to its 4 neighbors, and so on. In a square grid of 100 000 × 100 000 = 10 billion nodes, the nodes at the opposite ends of a diagonal are 200 000 links apart. This is the diameter of the network. Compare this to the distance of a mere 6 links that apparently characterizes the world social network with its nearly 10 billion nodes! Regular networks, where nodes are linked according to strict, repetitive rules rather than

random connections, are typically large-world networks. This means that a change in one node will normally take a very long time to propagate to the rest of the network. As a result, the network will be slow to react to perturbations or innovations.

We may conclude that complex networks are not regular. But they are not random either: their linking patterns do obey certain regularities, albeit not strictly. In fact, it has been shown that a regular network can easily be turned into a small-world network by adding a small number of randomly chosen links to the otherwise strictly constrained links (6). These random links by definition do not care about the "distances" within the regular grid: e.g. they may directly connect nodes that are otherwise 100 000 links apart. Such random links create "wormholes" or "shortcuts" between otherwise remote regions, thus bringing them suddenly within easy reach. As a result, a small number of random links added to a regular network spectacularly decreases the shortest path length between nodes.

CLUSTERING

One of the non-random features that characterize complex networks is *clustering*. Clustering means that when A is linked to B, and B to C, then the probability is high (or at least much higher than could be expected in a random network) that A is also linked to C. In other words, two randomly chosen connections of B have a much higher than chance probability of being connected themselves.

The origin of this can best be explained by considering social networks. Here, the clustering property can be formulated as "the friends of my friends are (likely to be) my friends". In other words, friends tend to form a cluster or community in which everyone knows everyone. The reason is simple: when you regularly encounter your friends, you are likely to encounter their friends as well. More generally, if an agent A frequently interacts with an agent B, and B interacts with C, then the probability is high that A will sooner or later interact with C as well. If A and B have some similarity in aims that helps them to find synergy, and the same applies to B and C, then A and C are likely to discover a synergetic relationship as well.

SCALE-FREE NETWORKS

A less intuitive feature of complex networks is that their distribution of links tends to follow a *power law* (7): there are many nodes with few links, and few nodes with many links. More precisely, the number of nodes N with a given degree (i.e. number of links) K is proportional to a (negative) power of that degree:

$$N(K) \sim K^{-a}$$

(The values of the exponent *a* tend to vary between 1 and 3.) A network that obeys a power law is called *scale-free*. When $a = 1$, N is inversely proportional to K: in other words, as the number of links goes up, the number of nodes with that number of links goes down proportionally.

This property has been established empirically, by counting the number of links in various networks, such as the web or social networks. It turns out that a few nodes have an inordinate amount of links. They function as the *hubs* of the network, the central "cross-roads" where many different connections come together. The most common nodes, on the other hand, have just a few links. This means that nodes are strongly differentiated: something that happens to a hub will have a disproportionately large influence on the rest of the network, while something that happens to an ordinary node may have little or no consequences. This has great practical implications: an innovation or perturbation that appears in a hub (e.g. a central network server, a high-visibility web page, or a person who is known by many) may change the whole network in a short time, because it is immediately propagated far and wide. By identifying the hubs in a network, it becomes easier to manipulate its dynamics, for good or for bad. Obvious applications are the spread of computer viruses, contagious diseases, new ideas, or fashions.

Whereas clustering tends to increase distances in a network, by creating locally connected clusters that have few links outside the cluster, the presence of hubs has the opposite effect. Because hubs have a very large number of links they are likely to link into many different clusters, thus acting as shortcuts that reduce the distance between the clusters. But this also means that removing a hub may break the connections between otherwise remote regions of the network. This is a danger especially in communication networks such as the Internet, where the failure of a small number of hubs may split up the network into separate "islands" that no longer communicate with each other. Similar dangers exist in ecosystems where the disappearance of one or more key species—i.e. "hubs" on which many other species depend—may lead to a complete breakdown of the system.

Barabasi and Albert (7) have proposed a theoretical explanation for power-law distributions based on the mechanism of *preferential attachment*: new nodes joining the network preferentially establish links with nodes that already have a large number of links. They have shown that when the probability of linking to a node is exactly proportional to the number of links of that node the resulting network obeys a power law with $a = 3$.

For a more general scenario for the self-organization of a complex network, consider a collection of agents that initially only interact locally with those that happen to pass in their neighborhood. Some of these interactions will be stabilized into enduring links. Once they have some links, the locality principle entails that agents are more likely to forge links with the "friends of their friends" than with randomly chosen others, thus promoting clustering. But agents that already have a high number of links also have many "friends of friends" (i.e. nodes two links away) and therefore they will be more likely to develop additional links within this 2-step neighborhood or cluster. The more links an agent has, the larger its neighborhood, and therefore the larger the probability that it will receive even more links from within this neighborhood. Similarly, the larger the cluster, the more likely it is to receive random links from outside, thus extending the neighborhood outwards and linking it into other clusters. This determines a positive feedback, which leads to an explosive growth in the number of links. The agents that happen to be in the center of such a quickly growing cluster will become the hubs of the network.

APPLICATION TO KNOWLEDGE NETWORKS

Having reviewed key concepts typifying complex, self-organizing systems and networks, we will sketch some possible applications of these ideas in the area of information science. Information science focuses on the knowledge that is available in the documents that are available in libraries and databases across the world. These documents are typically produced by authors or researchers who investigate a domain, building further on the results of other authors, and publishing their results in new papers or books that refer to these used sources. This knowledge producing system can be viewed as a very complex network, formed by the researchers, the concepts they use and the publications they produce. All the "nodes" of the network are linked directly or indirectly, by relations such as citation, collaboration or information exchange. This complex system is intrinsically self-organizing: no individual or organization is in charge, or can decide in which direction knowledge should develop. Novel, globally available knowledge emerges out of the spontaneous, local interactions between the individual agents.

By applying the concepts and methods from the domain of complexity, we may hope to better understand the development and structure of this network. We can view it as a complex, adaptive system that generates new patterns (knowledge) through the complex, non-linear interactions between multitudes of autonomous agents (individual scientists and organizations). This system has the structure of a heterogeneous network (Fig. 1), consisting of three basic types of nodes (9): agents, i.e. the individuals or organizations who actively process and produce knowledge, containers, i.e. the documents, databases or journals in which the produced knowledge is stored and made available to other agents, and concepts, i.e. the abstract elements of knowledge itself, which are typically represented as keywords.
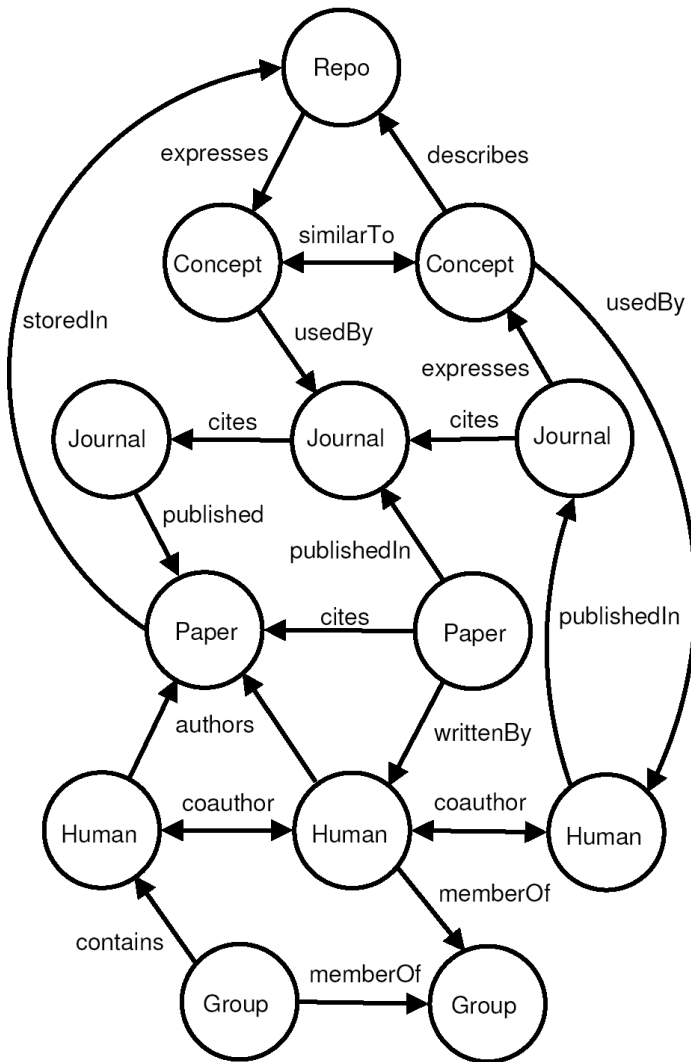
Fig. 1: a heterogeneous knowledge network, containing authors, concepts and documents

There already exists some preliminary work on subnetworks of this encompassing network, such as collaboration networks between authors (10) or citation networks between documents. This research has found that they possess typical features of complex networks, such as being scale-free and small-world. For example, citation networks typically contain a small number of hubs ("citation classics") with very many links, while most publications only gather a few citations. Some of the most successful recent methods for information retrieval, such as the PageRank algorithm underlying the Google search engine, or the HITS method developed by Kleinberg (11), implicitly use this network structure to identify the "hubs" of a hypertext network.

More interesting even than the static analysis of existing networks is the modeling of their evolution. We may assume that an information network will self-organize through the propagation of information between nodes across links, creating new links and nodes in the process. For example, assume that information is transferred from paper A to

researcher B. After reading the paper, B may decide to get some more information linked to paper A, e.g. by contacting A's author, or reading some of A's references. These in turn may refer B to other authors or papers relevant for B's interests, and so on. Some of these additional sources may turn out to be particularly important for B's research, inspiring B to develop a new concept, published in the form of one or more papers. This process will create links (e.g. B may start collaborating with another author, or refer in new papers to papers discovered in this way) and nodes (e.g. new papers, new concepts, new journals). Such links and nodes will tend to cluster around a small number of "hubs"—thus defining a new "community" of related authors, documents and ideas.

The emergence of a new scientific domain is a good example of the self-organization of such a community of knowledge (12), where people from initially diverse backgrounds find each other around a common interest, which gradually coalesces into a new paradigm. This process could be observed by mapping the network of authors, publications and keywords in a particular domain at regular intervals (e.g. every 2-5 years), and analyzing it in terms of clustering, hubs, average distances, etc. The change of these features over time may show processes of self-organization taking place. A good theory of the self-organization of knowledge communities would propose a number of processes and parameters that allow us to predict where, when and how such self-organization is most likely to take place. Such a theory would help us to find not only the presently most authoritative concepts, publications or authors (hubs), but those that are likely to become so in the future. This would provide a very powerful instrument to uncover emerging trends and to direct attention and investment towards the most promising people, ideas and information sources.

CONCLUSION

The science studying complex, self-organizing systems and networks is still in its infancy. Yet, it already provides us with a powerful new perspective and a number of promising conceptual and modeling tools for understanding the complex phenomena that surround us, including organisms, the Internet, ecosystems, markets and communities.

On the one hand, the complexity perspective reminds us to be modest in our aims: many phenomena in nature and society are simply too complex to be analyzed in the traditional scientific manner. Openness and non-linearity make a complex system in principle unpredictable and uncontrollable: the tiniest internal or external perturbations can be amplified into global changes. Therefore, we will never be able to capture it in a complete and deterministic model. Still, agent-based computer simulations can help us to get an insight into the qualitative dynamics of the system, and to classify and delimit the likely scenarios for its further evolution.

On the other hand, the complexity perspective gives us new reasons for optimism: while we cannot truly control a complex system, it tends to self-organize to a state where it regulates itself. This state tends to increase the utility or fitness of the system's active components or agents, by coordinating their interactions so as maximize synergy. The resulting organization is distributed over all the agents and their interactions, and thus much more robust and flexible than any centralized design. Moreover, it determines a

number of emergent, global properties that cannot be reduced to the properties of the individual components. By understanding the underlying mechanisms, we may be able to facilitate and stimulate such self-organization, or to drive it in one direction rather than another.

One of the most recent applications of the complexity perspective is the analysis of complex networks, such as the World-Wide Web, and the non-linear processes that generate them. This has led to the identification of common statistical features of such networks: small-world, clustering and scale-free link distributions. These notions promise a wealth of applications in the analysis of information networks, potentially helping us with the organization, management, retrieval and discovery of relevant knowledge within masses of ill-structured and continuously changing data.

FURTHER READING

Axelrod R.; M. D. Cohen. *Harnessing Complexity: organizational implications of a Scientific Frontier*, Simon and Schuster, 2000.

Bar-Yam, Y. *Dynamics of Complex Systems*. Westview Press, 2003. http://necsi.org/publications/dcs

Haken, H. *Information and Self-Organization: A Macroscopic Approach to Complex Systems*, Springer-Verlag: New York, 2000.

Heylighen, F. The Science of Self-organization and Adaptivity, in: L. D. Kiel, (ed.) Knowledge Management, Organizational Intelligence and Learning, and Complexity, in: *The Encyclopedia of (Life Support Systems* ((EOLSS), (Eolss Publishers, Oxford, 2002). (http://www.eolss.net)

Holland, J.H. *Hidden Order: How adaptation builds complexity*, Addison-Wesley, 1996.

Kelly, K. *Out of Control*, Addison-Wesley, New York, 1994.

Miller, J.H., Page, S.E. *Complex Adaptive Systems: an introduction to computational models of social life*, Princeton University Press, 2007.

Newman, M. E. J . The structure and function of complex networks. *SIAM Review* **2003**, 45, 167–256. http://arxiv.org/abs/cond-mat/0303516.

Prigogine, I.; Stengers, I. *Order out of Chaos*, Bantam Books, New York, 1984.

Waldrop, M. M. *Complexity: The Emerging Science at the Edge of Order and Chaos,* London: Viking, 1992.

REFERENCES

1. Ashby, W. R. Principles of the Self-organizing System. In von Foerster, H. and G. W. Zopf, Jr. (Eds.), *Principles of Self-organization*. Pergamon Press 1962, 255-278.

2. von Foerster, H.,. On self-organising systems and their environments, in: *Self-Organising Systems*, M.C. Yovits and S. Cameron (eds.), Pergamon Press, London, 1960 30-50.

3. Prigogine, I. and Stengers, I. *Order out of Chaos*, Bantam Books, New York, 1984.

4. Haken H. Information and Self-Organization: A Macroscopic Approach to Complex Systems, Springer-Verlag New York, 2000.

5. Waldrop, M. M. Complexity: The Emerging Science at the Edge of Order and Chaos, London: Viking, 1992.

6. Watts D. J.; S. H. Strogatz, Collective dynamics of small-world networks, Nature **1998**, 393, 440–442.

7. Albert R.; A.-L. Barabasi, Statistical mechanics of complex networks, Review of Modern Physics **2002**, 74, 47–97.

8. Gershenson, C.; F. Heylighen. How can we think the complex? in: Richardson, Kurt (ed.) *Managing the Complex Vol. 1: Philosophy, Theory and Application*.(Institute for the Study of Coherence and Emergence/Information Age Publishing, 2005), 47-62.

9. Rodriguez, M.A. A Multi-Relational Network to Support the Scholarly Communication Process, International Journal of Public Information Systems **2007**, 1, 13-29.

10. Newman M. The structure of scientific collaboration networks, Proceedings National Academy of Sciences USA **2001**, 98, 404–409.

11. Kleinberg J. Authoritative sources in a hyperlinked environment, Journal of the ACM **1999**, 46 (5), 604—632.

12. Wagner, C.S.; Leydesdorff, L. Network structure, self-organization, and the growth of international collaboration in science, Research Policy **2005**, 34 (10), 1608-1618.

13. Edmonds, B.. What is complexity?: the philosophy of complexity per se with application to some examples in evolution. In The Evolution of Complexity, F. Heylighen, J. Bollen, and A. Riegler, Eds. (Kluwer, Dordrecht **1999**), 1–18.

14. Holland, J.H. Studying Complex Adaptive Systems, Journal of Systems Science and Complexity **2006**, 19(1), 1—8.

15. Levin S.A. Complex Adaptive Systems: Exploring the Known, the Unknown and the Unknowable, Bulletin of the American Mathematical Society **2002**, 40 (1), 3–19.

16. Louzoun, Y., Solomon, S., Atlan, H., and Cohen, I.R. Modeling complexity in biology, Physica A: Statistical Mechanics and its Applications **2001**, 297 (1-2), 242--252