ON THE EPISTEMOLOGY OF COMPLEX NETWORKS THEORY

ALBERTO PASCUAL-GARCÍA

Centro de Biología Molecular "Severo Ochoa" (CSIC-UAM), Madrid, (Spain)

apascual@cbm.uam.es

COMPLEX NETWORKS THEORY IS A DISCIPLINE that aims to understand those systems composed by a large number of elements interacting non linearly. Its starting point consists on modeling these systems as networks where the interacting entities are represented by nodes, and interactions among nodes as links if a relation exist. The mathematical framework further developed in order to explore both the static and dynamical properties of these networks is what is known as complex networks theory (Newman 2006). The conceptual simplicity of this approach facilitates its application to very different areas of knowledge such as physics, biology or social sciences, with an spectacular increase in the number and variety of publications during the last years.

The conceptual simplicity seems to partially explain its success for two reasons. On one hand we observe that some common (non trivial) patterns arise from the analysis of very different systems. In the other hand, it is possible to generate models in an *empty-of-content* fashion intended to reproduce a given pattern. Therefore, when the modeled process generates the desired pattern, it may be considered as a suitable starting point to explain its emergence in a variety of systems. This fact confers an extraordinary generality to the analysis and modeling of complex systems. A paradigmatic example comes from the observation that, in many systems, the probability distribution of the number of links per node (known as the degree distribution) follows an scale free behaviour (Barabási 2009). It has been proposed within the same formalism a simple model to explain this pattern, the so called *preferential attachment process* (Barabási 1999), that generates scale free networks and it is easily modified to account for a particular scaling exponent.

Therefore if, as it happens in this example, the modeled process generates the desired pattern, it may be considered as a suitable starting point to explain the emergence of the observation under analysis in a variety of systems. This fact confers to this theory an extraordinary versatility for the analysis and modeling of complex systems.

But, although the application of complex networks theory implies a commitment with strong assumptions arising from the very first modeling step -i.e. the system definition- the epistemological aspects of complex networks theory have received poor attention. Our aim here is to explore the epistemological basis of the complex networks theory.

The approach we will follow highlights the ability of the complex networks framework to deal with *dialectic* concepts (Georgescu-Roegen 1971). Dialectic concepts are better understood in contraposition to *arithmomorphic* concepts, which are those concepts that can be discretely differentiated. Following the words of Georgescu-Roegen: "(arithmomorphic concepts) conserve a differentiate individuality identical in all aspects to that of a natural number within the sequence of natural numbers". Arithmomorphic concepts are suitable for formal reasoning, and therefore compatible with a quantitative treatment. On the other hand, dialectic concepts intrinsically make reference to processes where qualitative changes are present. An accessible example may be the concept *phase transition*. For instance, in the transition from water to vapour both phases are conceptually well characterized. But it is the process of change itself what challenges our ability to quantitatively describe the system. How can we properly characterize both concepts, water and vapour, if we are in a regime where both phases coexist? Scientific modeling may be viewed as an activity where the construction of an arithmomorphic scheme is critical, and systems that must be described through dialectic concepts challenge scientific modeling. Concepts such as the adjective *democratic*, being defined with a wide variety of implicit qualitative variables changing in space and time, make difficult to propose an arithmomorphic scheme oriented to provide an objective measure (therefore discretely differentiated) of this concept, and must be understood as dialectic.

An illustrative example of what is understood as arithmomorphic scheme and its relation with dialectic concepts comes from the *classification* problem. The classification problem can be considered within the complex networks framework if we link objects when they share certain similarity, and this leads to disjoint clusters where elements within each cluster are transitively connected. In general, when the entities considered in any attempt of classification contain an increasing amount of dialectic concepts, there are also increasing chances of dealing with a frustrated problem (Binder 1986). In biology the problem of classification has been historically fundamental. Species definition is a classical example where discrepancies do exist nowadays, see for instance the discrepancies between Ecdysozoa and Coleomata hypothesis (Philip 2005; Philippe 2005), or the difficulties in the definition of bacterial species (Cohan 2002). And frustration in the classification problem is not specific of species definition. It also appears when we deal with other entities such as expression profiles or protein structures, and this is why it has been recognized as a central epistemological problem in computational biology (Dougherty 2006).

We argue that it is precisely in those scientific areas where dialectic concepts are found more frequently, as it is in biology, sociology or economy, where the formal development is being slower than in other sciences such as physics. It has been pointed out that a specific feature for the maturity of any science comes from a growing interest on “processes in which things have become what they are, starting out from what they once were, and in which they continue to change and to become something else in the future”, rather than an interest on “the basic qualities and properties that define the mode of being of the things treated in that science” (Bohm 1971), typical of earlier stages of its development. This may be the reason why Schrödinger predicted that biology would have increasing attention for experts coming from formal disciplines, as an also increasing amount of available data would facilitate any attempt of formalization (Schrödinger 1992).

With the following epistemological approach to complex networks theory we will try to show that the success of complex networks theory is not only due to its conceptual simplicity. We will show that, after the very hard process of conceptual reduction of the system to construct an arithmomorphic description, the analysis that follows obtains patterns which are compatible with a dialectic interpretation. We start with a brief introduction to the system description constructed in the standard modeling process within complex networks theory, in order to understand the parallelism we propose with an epistemological approach.

Modeling within the framework of complex networks theory necessarily begins with the definition of the system, which is an unavoidable exercise for any standard observer dealing with scientific knowledge (Maturana 1975). This exercise implies that an important number of variables are either neglected or simplified with operations such as averaging, what constitutes a dimensionality reduction process and defines system boundaries. In our case, we consider a set of nodes which are characterized by a finite set $x_i \in X$ of variables. These variables can be quantitative or qualitative (including the observation *the node A interacts with B*), being binary variables in the latter case. From the specific values α of the considered variables, x_i^α, links between nodes are defined after a suitable operation defined by the observer, what allows her to determine what is understood as relation, typically an interaction. The operations intended to define links constitute an additional dimensionality reduction process and highlight the relevant variables and values in the linkage process. Our proposal consists on relating these ingredients with an epistemological formal description.

We will follow the formal epistemological approach proposed in (Boniolo 2008), particularized to our problem. We will consider that the system *O* is composed by objects $o \in O$ that correspond to nodes in the standard modeling process. We will not ask ourselves whether the observer has direct access to the objects or not as in (Boniolo 2008), what has been proposed as a Kantian approximation to formal epistemology, although it is absolutely compatible with this interpretation. As objects are fully described by the different variables x_i together with its correspondent values x_i^α, any analysis can be accomplished just by knowing these variables and values, what will constitute the conceptual apparatus of the knowing subject. Therefore, we can define the potential conceptual apparatus *C* of the knowing subject as composed by a set of elementary concepts $c \in C$ (Frege 1892), once the system has been defined. But the actual conceptual apparatus Δ, consists just on particular values associated with the measurements provided to the subject that constructs the model. We propose a parallelism where the characteristics are associated with particular viable values of the variables ($x_i^\alpha \rightarrow c_i^\alpha$) and we further define concepts ν, which are finite subsets of characteristics: $\nu \equiv \{c_i^\alpha \dots c_j^\beta\}$. For a finite network we can consider a finite number of characteristics as its construction requires a static snapshot for the values of the variables, or alternatively thresholds can be imposed. For instance, a characteristic c_i^α can be defined as *a variable* x_i *taking a value larger than α*.

Following this parallelism, we define a binary constitution relation $\Vdash$ that, given any $o \in O$ and any $\nu \in \Delta$, the statement $o \Vdash \nu$ means that ν is one of the concepts cognitively constituting o. Of course the constitution relation for a given concept involves its set of characteristics, what leads to the following *constitution rules*:

$$\frac{o \Vdash \nu \quad c \in \nu}{o \Vdash \{c\}} \qquad \frac{(\forall c \in \nu) \;\; o \Vdash \{c\}}{o \Vdash \nu}$$

From the constitution relations we want now to discretely describe any object. Let $\nu \in \Delta$ be a concept. We define as the extension of a concept the subset of the objects in O constituted by ν :

$$Ext(\nu) \equiv \{o \in O \mid o \Vdash \nu\}$$

In order to guarantee that each object is completely grasped with a subset of concepts, which is the initial scenario we consider before linking objects, let us call U the subset of concepts constituting a given object o. This object is discretely differentiated of any other if the following condition holds:

$$\bigcap_{\nu \in U} Ext(\nu) = \{o\}$$

A complete parallelism between both formalisms would be reached if we finally relate the link definition in the standard modeling with an operation between sets of concepts in the epistemological counterpart. Any operation oriented to construct a link between two objects o_1 and o_2 will need to consider the subsets of concepts describing these objects, namely U_1 and U_2 respectively, in such a way that the extensions of the final concepts relate both objects according to the standard models. The operation:

$$Ext(U_1 \wedge U_2) = Ext(U_1) \cap Ext(U_2)$$

maps the binary relation defined by a link in the standard formalism with the epistemological formalism and implies a dimensionality reduction, as the concepts that allow us to discretely differentiate both objects lie out the intersection.

It is now possible to show (Valentini 2005), that the set of concepts Δ together with a binary relation $\Vdash$ from Δ to the set of objects *O* constitute a concrete topological space $\chi = (O, \Delta, \Vdash)$, and allows for a formal definition of open and closed sets. The topological objects constructed in this way have been used to propose a formal definition of vagueness (Boniolo 2008).

Equipped with the above definitions, we will investigate whether the formal definition of vagueness is compatible with the identification of dialectic concepts. We will analyze the sets obtained through the proposed parallelism for different patterns found in real networks.

In particular, we will analyze three types of *3-node* subgraphs that have been widely found in different systems and may be viewed as building blocks in complex networks (Milo 2002). We will further relate these patterns with real biological examples containing dialectic concepts. First, the objective definition of protein *fold*, which is directly related with the classification problem and that will allow us to recover the paradigmatic example of scale free networks (Sadreyev 2009; Pascual-García 2009). The second example will address the problem of the identification of remote protein homologs, that can be solved through transitive sequence similarity operations (Roessler 2008). The third example will consider a recent proposal for species coexistence through transitive competition (Allesina 2011).

We will finally discuss those critiques claiming that disciplines that make use of these kind of approaches, such as systems biology, are reductionist (Mazzocchi 2008; Van Regenmortel 2004). The fact that the description of the system is necessarily reduced does not imply that it is a reductionist approach. It is rather a necessary epistemological exercise to deal with complex systems that allows to the scientist to propose general questions and that otherwise would not be possible to handle. We claim that these approaches circumvent the difficulties arising from the study of systems with intrinsically dialectic concepts, opening a door to the establishment of general laws. And it would be nothing but time and experiments what will allow us to test both the predictive power of complex networks theory and the skepticism.

REFERENCES

(Allesina 2011) S. Allesina and J. M. Levine. A competitive network theory of species diversity. *Proceedings of the National Academy of Sciences*, 108(14):5638-5642 (2011).

(Barabási 2009) A.L. Barabási. Scale-Free networks: A decade and beyond. *Science*, 325(5939):412-413, (2009)

(Barabási 1999) A.L. Barabási and A. Réka. Emergence of scaling in random networks. *Science*, 286(5439):509-512, (1999).

(Binder 1986) K. Binder and A. P. Young. Spin glasses: Experimental facts, theoretical concepts, and open questions. *Reviews of Modern Physics*, 58(4):801-976, (1986).

(Bohm 1971) D. Bohm. Causality and Chance in Modern Physics. *University of Pennsylvania Press*, (1971).

(Boniolo 2008) G. Boniolo and S. Valentini. Vagueness, kant and topology: a study of formal epistemology. *Journal of Philosophical Logic*, 37(2):141-168, (2008).

(Cohan 2002) F. M. Cohan. What are bacterial species? *Annu Rev Microbiol*, 56:457-487, (2002).

(Dougherty 2006) E. R. Dougherty and U. Braga-Neto. Epistemology of computational biology: Mathematical models and experimental prediction as the basis of their validity. *Journal of Biological Systems*, 14(1):65-90, (2006).

(Frege 1892) G. Frege, Über Begriff und Gegenstand (1892), Engl. Transl. P. T. Geach, and M. Black (eds.) On concept and object. *Mind*, 60(238):168-180, (1951).

(Georgescu-Roegen 1971) N. Georgescu-Roegen. The Entropy Law and the Economic Process. *Harvard University Press*, 1st edition, (1971).

(Maturana 1975) H. R. Maturana, E. H. Lenneberg, and E. Lenneberg. Biology of Language, the Epistomology of Reality. In *Foundations of Language Development, a Multidisciplinary approach*, volume 2. The UNESCO Press, (1975).

(Mazzocchi 2008) F. Mazzocchi. Complexity in biology, exceeding the limits of reductionism and determinism using complexity theory. *EMBO Reports*, 9(1):10-14, (2008).

(Milo 2002) R. Milo, S. Shen-Orr, S. Itzkovitz, N. Kashtan, D. Chklovskii, and U. Alon. Network motifs: Simple building blocks of complex networks. *Science*, 298(5594):824-827, (2002).

(Newman 2006) M. Newman, A.-L. Barabási, and D. J. Watts. The Structure and Dynamics of Networks. *Princeton University Press*, 1st ed, (2006).

(Pascual-García 2009) A. Pascual-García, D. Abia, A. R. Ortiz, and U. Bastolla. Cross-Over between discrete and continuous protein structure space: Insights into automatic classification and networks of protein structures. *PLoS Comput Biol*, 5(3):e1000331, (2009).

(Philip 2005) G. K. Philip, C. J. Creevey, and J. O. McInerney. The opisthokonta and the ecdysozoa may not be clades: Stronger support for the grouping of plant and animal than for animal and fungi and stronger support for the coelomata than ecdysozoa. *Molecular Biology and Evolution*, 22(5):1175-1184, (2005).

(Philippe 2005) H. Philippe, N. Lartillot, and H. Brinkmann. Multigene analyses of bilaterian animals corroborate the monophyly of ecdysozoa, lophotrochozoa, and protostomia. *Molecular biology and evolution*, 22(5):1246-1253, (2005).

(Roessler 2008) C. G. Roessler, B. M. Hall, W. J. Anderson, W. M. Ingram, S. A. Roberts, W. R. Montfort, and M. H. J. Cordes. Transitive homology-guided structural studies lead to discovery of cro proteins with 40% sequence identity but different folds. *Proceedings of the National Academy of Sciences*, 105(7):2343-2348, (2008).

(Sadreyev 2009) R. I. Sadreyev, B.-H. Kim, and N. V. Grishin. Discrete-continuous duality of protein structure space. *Current Opinion in Structural Biology*, 19(3):321-328, (2009).

(Schrödinger 1992) Erwin Schrödinger. What is Life?: With Mind and Matter and Autobiographical Sketches. *Cambridge University Press*, (1992).

(Valentini 2005). The problem of the formalization of constructive topology. *Archive for Mathematical Logic*, 44(1):115-129, (2005).

(Van Regenmortel 2004) M.H.V. Van Regenmortel. Reductionism and complexity in molecular biology. *EMBO Reports*, 5(11):1016-1020, (2004).