Ann Ist Super Sanità 2021 | Vol. 57, No. 4: 330-342
DOI: 10.4415/ANN_21_04_08

Network as a language for precision medicine

Lorenzo Farina

Dipartimento di Ingegneria Informatica, Automatica e Gestionale "A. Ruberti", Sapienza Università di Roma, Rome, Italy

Abstract

The article takes as a starting point the observation of a deep and long-standing gap between the views of biologists/physicians and that of physicists/data scientists when dealing with life sciences. This gap has been exacerbated by the advent of large-scale -omics technologies. Here, we focus on the impact of this gap in the field of precision medicine that impedes dialogue between omics data analysts and precision medicine physicians. To try to overcome this cultural divide, here we suggest a new possibility through the use of network science as a shared language composed of a vocabulary of words that have different meanings in each discipline but refer to the same biological entity. By doing so, one can move from biological concepts to network patterns and algorithms and backwards, thus generating a dialogue between "life scientists" and "number scientists". The article presents several simple network concepts with a straightforward biological interpretation as a starting point for such interdisciplinary dialogue.

Key words

- network science
- precision medicine
- computational biology
- bioinformatics
- omics data
- network medicine

INTRODUCTION

There is a growing awareness of the large and increasing gap between two visions of life: the vision of research scientists like biologists, physicians, molecular biologists, and so on, and that of research scientists like physicists, engineers, mathematicians, and so on [1]. For the sake of simplicity, we can call the first "life scientists" and the second "number scientists". The two groups are just an idealized and simplified classification, and a wide variety of different approaches are included in the same group, as well as a continuum between the two extremes is conceivable. Nevertheless, life scientists share a common vision that is well defined in the words of Ernst Mayr: "*Owing to their complexity, biological systems are richly endowed with capacities such as reproduction, metabolism, replication, regulation, adaptedness, growth, and hierarchical organization. Nothing of the sort exists in the inanimate world*" [2]. The fundamental separation between the two visions is just in front of us every day. It suffices to quote the philosopher of science Evelyn Fox Keller who wrote: "*I have had ample opportunity to observe failures of communication virtually whenever experimental and mathematical biologists happened to be in the same room*" [3]. Life scientists vision is centered on "concepts" like for example evolution, adaptation, development, speciation, purpose, which are not amenable of a rigorous and immediate mathematical formalization, while number scientists' vision is centered on fitting "patterns" to data [4]. Furthermore, it is a common belief that biology is subject to "universal laws" (yet to be discovered), and that from these laws one can derive mathematical models, and consequently computations on data, i.e., algorithms. In other words, the underlying idea is that, to make sense of biological data we need to find universal patterns and "general mathematical theories" of biology, just like in physics. Number scientists believe that biological information can be obtained from data only by establishing an all-encompassing mathematical framework from which derive equations that naturally lead to some calculation on data. Simply put, from this perspective, the path from biological properties to algorithms on data is necessarily mediated by a "universal mathematical theory" provided by the coming of the "Newton of biology" and guided by the "law of parsimony" (Ockham's razor) [5]. By contrast, life scientists have an opposing view of the problem of making sense of biological data. The only "law" of biology, although not written in the language of mathematics, is evolution, and their vision is centered on the uniqueness of life as a scientific discipline, with its own language and concepts [2], often (if not always) not amenable of precise mathematical formulation. Apart from the biochemical elements that make life possible on earth (e.g., DNA structure or protein structure), the life scientists usually strongly oppose universality, as clearly explained by Steven Jay Gould "*If nature teaches us any lesson, it loudly proclaims life's diversity. […] In any case, bursting diversity is nature watchword; it should never be submerged by careless abstraction*" [6]. Our aim is not to reconcile such visions or to take the side of one or the other, but to focus

Address for correspondence: Lorenzo Farina, Dipartimento di Ingegneria Informatica, Automatica e Gestionale "A. Ruberti", Sapienza Università di Roma, Via Ariosto 25, 00185 Rome, Italy. E-mail: lorenzo.farina@uniroma1.it.

on the almost complete absence of dialogue between life scientists and number scientists caused by an objective gap between the two cultures. Certainly, we do not support both the idea that biology's dignity as a science depends on its degree of mathematization and the opposite vision that considers mathematics as a simple tool-generator, so that the life scientist must choose the "best" one for his/her purposes, without any interactive dialogue with the number scientist.

Here, we focus on the impact of this gap in the field of precision medicine by presenting an alternative view and propose a contribution to reduce the gap (that we call the "complex data divide") and allow a dialogue, by looking at network science as a vocabulary generator (not tools). It is *not* enough to fit data to patterns if the patterns do not match a biological concept, and vice-versa. Mathematical patterns may be elegant but lacking biological plausibility (like, for example, the Turing mathematical model of morphogenesis [3] just to cite the most popular), and biological concepts may not result in a significant mathematical pattern on available data. Precisely, here we propose networks as a way to represent omics (large scale) data by focusing on relationships among elements and discuss some network pattern (the words of the language) as suggestive of biological features so that, the life scientist and the number scientist, can speak the same language even though they give a different meaning to the same words. It is worth noting that the recently introduced concept of "networks of networks" [7] can be fruitfully used to generate new "words" of the network vocabulary at different scales.

Here, we review some of the most promising examples of "words" of the network language and their precision medicine counterpart and show how they can be used to link biological concepts to algorithms on networks and vice versa, without the need for any "general theory" and, most importantly, to stimulate a truly interdisciplinary dialogue between life scientists and number scientists. As a final comment, we note that the network vocabulary is still in its infancy, but there is growing evidence that such network-based dialogue can effectively reduce the complex data divide.

"OMEXITY": OMICS DATA EXPLOSION AND DISEASE COMPLEXITY

Complexity as emergence. A complex disease is the result of many intertwined factors that include polygenic risk variants, physiological and environmental stresses, lifestyle habits, and many others. Moreover, even Mendelian disorders do not adhere to the one gene – one phenotype model, so that the awareness that virtually all diseases share these properties is increasing [8]. For these reasons, complex diseases have increasingly become the focus of modern medicine, especially in western countries, where age-related pathologies (cardiovascular and respiratory diseases, cancer, or type 2 diabetes) are the leading causes of death globally. According to the World Health Organization (WHO), 7 of the 10 leading causes of death are the so-called "non-communicable diseases". These seven causes accounted for 44% of all deaths or 80% of the top 10 and, together, accounted for 74% of deaths globally in 2019 [9]. Non-communicable diseases (NCDs) are chronic pathologies, that is they tend to develop and persist over a long period. Most importantly, they are all complex diseases, in that their onset and development are inherently multi-factorial.

Complexity in disease onset and development can be fully considered the "rule" rather than "the "exception". Indeed, the definition itself of a "complex" disease as "caused by the interaction of multiple genes and environmental factors" (taken from NIH glossary of genetics term), calls to mind the concept of *interconnection* among factors. As such, complexity arises from the crosstalk among a variety of molecular factors and pathways that prevents the understanding of pathogenesis as a linear causal route from genotype to phenotype. Complex diseases emerge from the interplay of many actors working together [10]. Such multi-factorial character implies that most causal players have just "weak" effects on the disease, whereas mainstream research studies assume the presence of risk elements with a "strong" effect [11].

The emergence of disease as a complex phenomenon makes it intractable from a reductionist perspective, which is the idea that systems can be understood by looking at every single component and disease as a linear chain of molecular interactions [12]. For example, using the terminology borrowed from complexity theory, cancer onset and progression is characterized by many "emergent" properties that reveal its behavior as an adaptive, self-organized system. The hallmarks of cancers include triggering proliferative signaling, evading growth suppressors, resisting cell death, enabling replicative immortality, inducing angiogenesis, and activating invasion and metastasis [13]. These biological properties cannot be traced back to a single cell or component, but to the web of interconnections of many factors that lead to the "emergence" of both normal and pathological cell behavior.

Omics networks. As already mentioned, a complex disease is also a complex system, in that its properties "emerge" from the interactions between components and the environment. From this perspective, it is quite natural to represent complex systems as *networks* where nodes represent parts (components), and links represent interactions or, more generally, associations among nodes [14], as illustrated in *Figure 1*.

The explosion of omics data has revolutionized the study of complex diseases. The big data produced by omics technologies are pervasive, and their variety and availability increase every day. They include genomics, transcriptomics, proteomics, epigenomics, microbiomics and many others. For example, Next Generation Sequencing (NGS) devices allow the study of inherited genetic factors using exome or targeted panels, or the effects of specific lifestyles by detecting changes that impact the global expression pattern, assessing epigenetic mechanisms (methylation, non-coding RNAs). It also allows to study the impact of environmental factors, such as the microbiome composition and its interactions with the immune system, to identify biomarkers and personalized drug response. The current situation is often referred to as era of "big" data,

and emphasis is given on the quantitative aspect, rather it should be clear that the real challenge is that such large amounts of data are linked to one another in very complex ways, so they should be called "complex data" to better highlight the key question at stake which is *not* merely quantitative, as discussed in what follows.

Types of omics networks. It would be unreasonable to give an exhaustive list of all possible networks that can be constructed using omics data. Indeed, besides *interaction networks*, that is networks in which a link is present due to physical interactions between the two biological entities represented by nodes, *e.g.*, map of protein-protein interaction (PPI), it is very common to build *association networks* where links represent any kind of association rule, thus making the universe of all possible association networks virtually infinite. Associations between two elements (the nodes of the network) can be derived in many ways, by looking at any commonality between them. For example, interesting association networks are the so-called "human disease network" or "diseasome" [15], which is a disease-disease network where diseases are connected if they share a common genetic component (a mutation) or the "patient similarity networks" [16] in which patients are linked based on their similarities in various clinical features, including genomic profiles. A comprehensive list of molecular networks of both types can be found in a recent review by Silverman *et al.* [17].

More is different. Omics data are certainly important because they provide an unprecedented view of a cell's life at the molecular level, but equally important is the role of the availability of large quantities of information, i.e., the large-scale feature of omics data. Indeed, as anticipated by Philip Anderson in 1972, "*more is different*" [18], which means that when we deal with a large number of highly interconnected entities, the properties of the single part fade into the background and new emergent properties arise, i.e. quantitative difference may become qualitative. This point is illustrated by *Figure 2*.

Omics technology explosion developed independently of awareness of the complexity of diseases. Indeed, complexity theory originated from the studies of the early fifties of the last century initiated by von Bertalanffy [19] and Boulding [20] in the field of holistic general dynamic systems theory [21], while the omics revolution started in the seventies of the last century fostered by discoveries in the field of biotechnology like recombinant DNA technology and Sanger sequencing [21, 22]. Interestingly, those two independent lines of research have now met in a single pathway leading to precision medicine. In other words, the omics/complexity (that we might call "omexity") era has just begun.

Omexity and precision medicine. The rise of "precision medicine" is crucially related both to widespread awareness of the complexity of virtually all pathological conditions and to the availability of increasingly large amounts of molecular omics data which call for improved information processing capabilities to extract relevant information [23]. Medical research in the era of precision medicine cannot but include data analysis and integration of large and heterogeneous sources like DNA/RNA sequencing, proteomics, imaging, digital pathology, laboratory medicine, vital signs, medical records, and so on. Simply put, complexity and omics together, naturally lead to the development of an integrative medical mindset that merges clinical observation and data pattern recognition. The real challenge of omexity and the associated precision medicine approach, is the buildup of abilities from somewhat separate worlds, i.e., the computational sciences and the life sciences. This amounts to saying that a new kind of interdisciplinary mindset is needed to tackle omexity, that is the ability to establish a dialogue between the

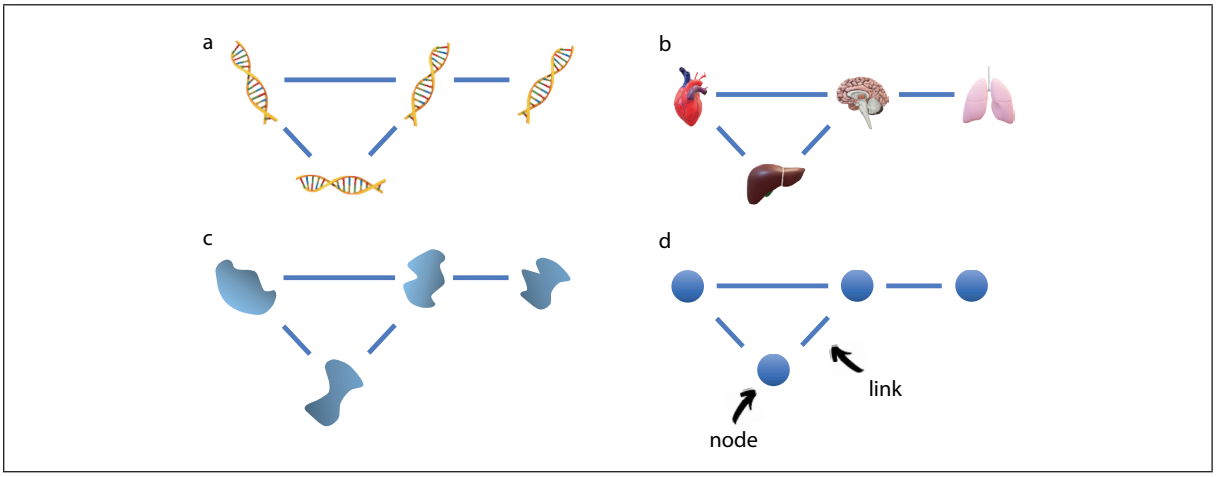

Figure 1
Networks. A network is a collection of nodes and links that connect them. The network representation simplifies reality in that it focuses on relationships, rather than on element properties. In fact, a) DNA-DNA associations may be obtained, for example, if mutations on a pair of genes are related to the same disease or if they are both targets of some transcription factors or miRNA or epigenetic modifier. b) A link between two organs may be established, for example, by considering comorbidities or for sharing some molecular driver. c) Two proteins could be linked if they interact, i.e., if they form a complex. d) the resulting abstract network is the same for all cases considered since the link pattern is the same, even though the elements are completely different.

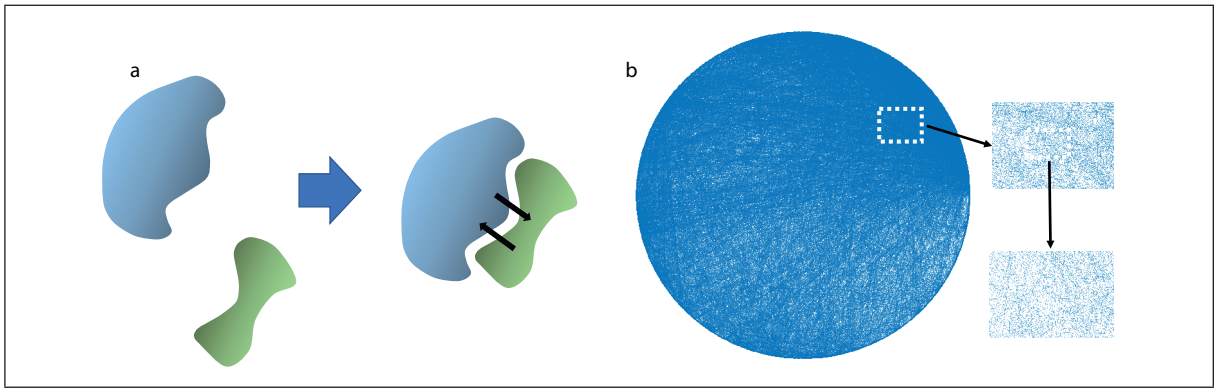

Figure 2
More is different. a) The binding of two proteins can be a very complicated process involving many spatial and biochemical properties of the subunits and the environment. b) When millions of binding events occur, new properties of the ensemble arise, and the properties of the whole cannot be attributable to single parts. Simply put, "more is different".

"calculative" and the "meditative" mind, using a terminology due to Martin Heidegger [24], that in our case corresponds to, on the one hand, the computational aspect of pattern recognition in omics data interpretation, and, on the other hand, to a more holistic view of disease and specificity of the single patient. It is perhaps the greatest challenge of the healthcare sector over the coming decade to integrate all these resources and translate them into clinical practice [23]. Such a deep gap between the medical and the data analyst mindsets can be termed as the "complex data divide".

In the next section, the cultural origins of such a divide will be discussed and possible roads will be suggested to reduce this gap that stalls knowledge advancement and interdisciplinary dialogue. As discussed in what follows, the network language could have a very special role in this dialogue.

THE "COMPLEX DATA DIVIDE" AND INTERDISCIPLINARY DIALOGUE

A document published in 1999 entitled *Medical and societal consequences of the Human Genome Project* [25] predicted that within 10 or 15 years, the impact of the human DNA sequencing would be a radical transformation of medicine. There is a general agreement that the genome project radically changed the rules of medical research, the way of practicing biological discovery, and the ubiquitous digitation of biological sciences. However, there is still a debate on whether there is a real impact on population's life expectancy or any other public health measures [26]. The omics data explosion has increased public expectations on the utility of molecular big data and on artificial intelligence (AI) methodologies for the discovery of new therapies for diseases which still lack effective treatments, like the vast majority of complex diseases, e.g., cancer or diabetes. Moreover, the recent pandemic made it clear to the general public, the urgent need for a personalized treatment based on multi-level profiling able to take into account the large variety of responses observed in the CoViD-19 manifestations, ranging from the absence of symptoms to pneumonia and death. This is a great challenge and an opportunity for omexity research. However, to meet such high expectations a radical perspective change is needed.

A key issue is the huge gap between the historical and cultural milieu of researchers trained in computational sciences (computer science, statistics, engineering, etc.) and those trained in the life sciences (biologists, physicians, biochemists, etc.). To tackle omexity, which lies at the interface between big heterogeneous omics data analysis (computation) and the extraction of relevant information (biological interpretation), a new interdisciplinary or intercultural dialogue must be established. Although the strict separation of the two domains is well known and often referred to as the difference from theoretical and applied sciences, the situation evidenced by omexity has peculiarities that make it something very different from the past. The omexity challenge cannot proceed directly from theory to applications, simply because we do not have theories like in physics or general quantitative laws that can provide an unambiguous framework for the development of the "best" algorithm for a given biological or medical question. A good illustrative example is the evidence that molecular systems of complex cells are inherently different from simple electronic circuits [27]. In other words, the role of mathematics and computation (algorithms) in the life sciences is extremely different from that of physics or engineering. As regards omexity, the point at stake is not to find universal laws but to establish a dialogue between omics data analysis and its biological interpretations. In other words, what is needed is a sort of "mapping" or "bidirectional flow" from emergent properties of data structures and emergent properties of biological systems. This key point is illustrated in *Figure 3*.

This mapping can be obtained, for example, by using "metaphorical projections", which is a metaphorical correspondence that can be established between two separate worlds, so that the finding in one domain can be translated into the other domain and *vice versa*. It is worth noting that such mapping cannot be devoid of any "theory" since a mechanistic explanation of the mapping is required to provide a solid background of such mapping. In other words, this is *not* the end of theory [28], but (hopefully) the beginning of a peer-to-peer

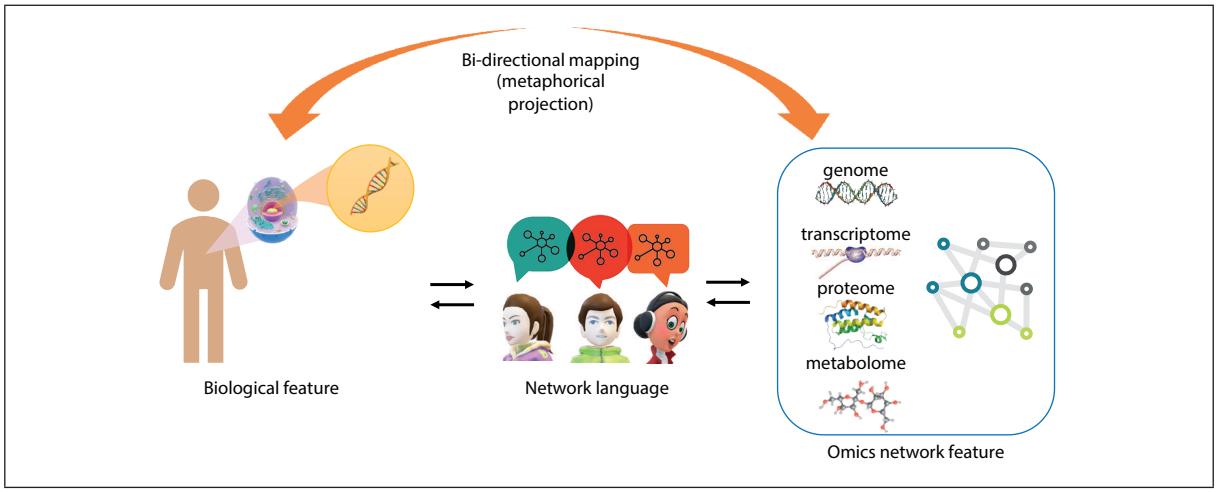

Figure 3
The complex data divide: networks as mediators from biology to omics and back. The bi-directional mapping (or metaphorical projection) from biological to network features is mediated by the network language. The availability of a set of words (language) that have a precise meaning, although different, both for the life scientist and the number scientist, makes the interdisciplinary dialogue possible.

dialogue between separate domains, each with its own rules and procedures that may be theories (like in statistics or physics) or interpretations in the light of biological concepts, like evolution, natural selection, reproduction, metabolism, replication, regulation, adaptedness, growth, hierarchical organization and so on, none of which have counterparts in the inanimate world. As pointed out by Ernst Mayr, the most fundamental difference between biology and the hard sciences is that biology theories are based on "concepts", while in the physical sciences they are based on "natural laws" [2]. Indeed, data do not speak for themselves since science is not about finding patterns but (biological) explanations for those patterns [29]. Most importantly, the goal is not the discovery of a unifying general theory but the finding of a link (a mapping of some sort) between the two domains, each with peculiar theories no matter how mathematical or conceptual they are.

Here we support the idea that network language, as defined and explained in the following sections, can effectively function as a bridge between the computational and the living domains, i.e., able to define single entities that can have a valuable, although different, meaning in both fields of research. Precisely, the large amounts of complex omics data and the awareness of disease complexity, lead straightforwardly to a true inter-disciplinarity dialogue which is possible only in the presence of a common language, like that of networks, and that results in a calculation or algorithm on data to implement precision medicine. This process is pictorially described in *Figure 4*.

THE NETWORK LANGUAGE

A metaphorical projection

A "network" is not necessarily a "real" thing, rather it must be considered a "cognitive schema" [30], which is an abstract collection of concepts used to make sense of the unknown world of life. Precisely, a general characterization of a network can be defined by a bunch of "nuclei" (where matter, as well as other activities are far more concentrated) linked to each other by edges (streets, power cables, etc.) passing in a much less dense environment [31], just like a lumped-element model of a spatially distributed physical system. From this perspective, in a more abstract sense, nodes are non-dimensional points and edges are one-dimensional lines. But, protein-protein interactions are not "lines" between "points" but very complex phenomena where many spatial and energetic factors are at work. For example, by substituting proteins with nodes and bindings with links, one is "projecting" the network schema onto the protein-protein interaction network and thus performing a "metaphorical projection" [30, 32]. The network is a functional abstraction, a way of rendering complex systems comprehensible using an oversimplified representation of data. And yet, at the same time, only through this distortion of reality operated by a metaphorical projection, the protein-protein interaction network becomes amenable to computations on data. The key point here is that "distortion" is by no means a re-creation of reality but, rather the inevitable re-organization of available knowledge into meaningful forms able to indicate solutions or useful directions of research for a specific purpose or problem of interest. An example of a successful metaphorical projection is the London underground map as designed by Harry Beck in 1933. "The map is not the territory" or "the menu is not the meal" are popular expressions to remember that we cannot confuse models of reality with reality itself. The construction of a map is not an easy task, since the goal is to represent only "relevant" information for the end-user. From the more general perspective of the scientific enterprise, it is worth quoting Gaston Bachelard's statement: "*Contemporary science maintains that quantities which are negligible must be neglected. It is not enough to say they can be neglected.*" [33, p. 220]. In the early 1930s, the map of the London underground was purely geographic and metro stations were represented

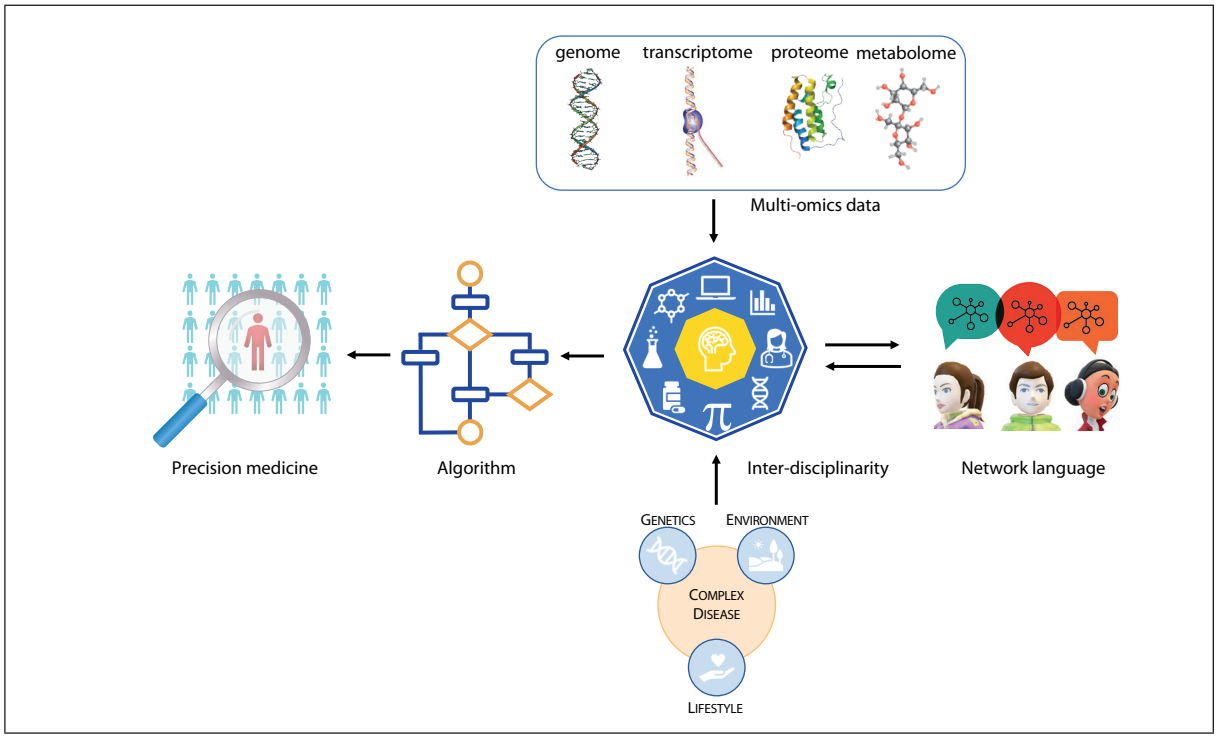

Figure 4
Network language for precision medicine. The network language enables a true interdisciplinary dialogue among the many skills required by the availability of complex data, like omics data. Questions on complex diseases can be answered using omics data through interdisciplinarity that, using the network language, may directly provide algorithms on data to provide answers without resorting to abstract theories of any kind.

on a true scale together with the city's street network, attractions, public institutions, parks, and waterways. The information content was overwhelming, and passengers found it difficult to use it. When the first abstract map developed by Beck was presented to Londoners in 1933, they found it useful and easy to understand [34]. The basic idea was to place stations without any direct correspondence with "true" geographic positions and to use only straight lines and orientations of 0°, 90°, or 45° degrees. An example of the resulting map is reported in *Figure 5*.

Beck's map is considered by many the most celebrated graphic design of the 20th century [34] and it is a perfect example of a metaphorical projection of "real data" (i.e., the geographic map) on an abstract and oversimplified representation which must serve the primary function of helping users to extract the information they need, without too many details that might lead the end-user astray. And this is not a specific problem with maps, but a general paradigm of how scientific enterprise proceeds by neglecting what *must* be neglected. Indeed, network representation of relationships among omics data is certainly an oversimplified vision of the tremendous complexity of life, but it is a necessary step to *make sense* of huge amounts of complex data. Once abstraction of concrete is performed to form representational elements (nodes and links), integration of these elements can be used to identify structures or patterns which, in turn, represent the words that constitute the "network language" able to express a global configuration, i.e. the act of arranging all the informational elements and their structures to create a whole: this "pattern language" approach is known as "theory of centers" and it has been developed for the design of visual artifacts [34].

In sum, a network is a visual representation of data focused on relationships (links) among elements (nodes) and link configuration is what really matters, whereas peculiarities of the nodes are neglected. The goal is therefore to identify network "patterns" to be metaphorically projected onto biological concepts, thus providing a language that may be able to overcome the complex data divide and allow computational and the biomedical researchers to talk to each other about the same reality from different perspectives. The next section is devoted to the illustration of three popular examples of network patterns that have been used to make sense of omics data in the recent literature.

The three C's of network language

When studying a network, many properties can be of interest, depending on the particular problem at hand. For example, in 1736, when Leonhard Euler studied the bridges of Konigsberg, he was interested in finding a path through the city that would cross each of its seven bridges over the river Pregel only once. However, special focus has been given by researchers in social and biological networks on three properties that are key if we consider the metaphorical projection on a situation in which "messages" or "information" flows from node to node through the links. Not surprisingly, a cell

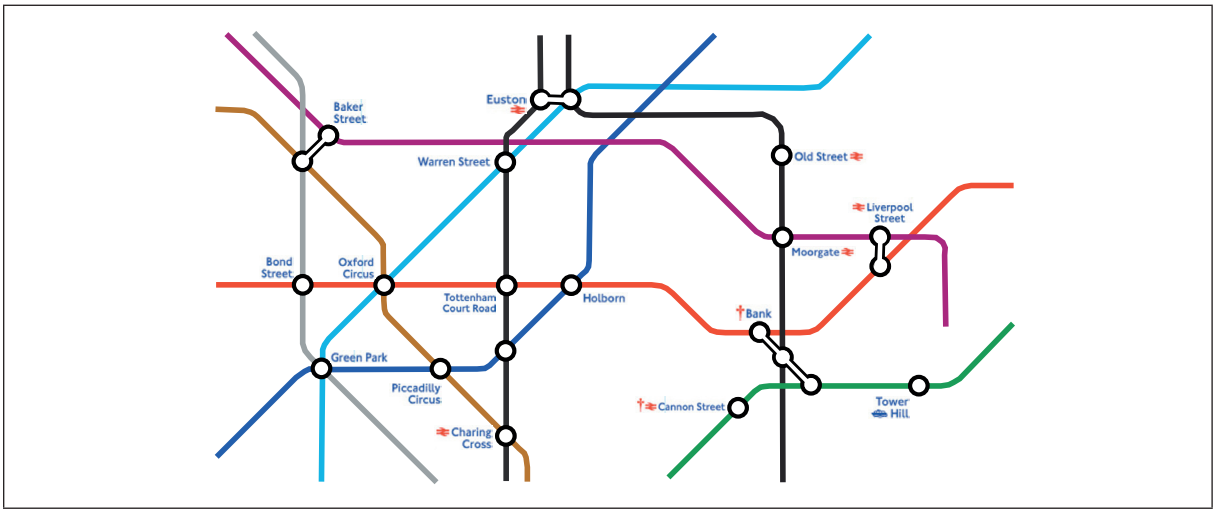

Figure 5
Metro map as metaphorical projection. The London Metro map is a paradigmatic example of a "metaphorical projection". Real geographic coordinates have been eliminated in favor of greater readability. The purpose of the map is not to faithfully reproduce reality but to represent relevant information suitable for finding the best route between stops.

is considered an "information processing unit" by many influential molecular biologists [35, 36]. From this perspective, it is customary to consider three features of a network: the modular organization in communities, the presence of "influential" (or central) nodes in a network or a community, and the presence of nodes connecting different communities. Therefore, the three C's of network analysis are communities, centralities, and connectors (see *Figure 6*).

Besides centrality measures, there are many other important topological properties tightly linked to biological concepts of special interest for precision medicine. The most promising certainly is the "interactome disease module" perturbation model of disease onset and development proposed by Barabasi, Gulbahce and Loscalzo [37] which has been validated on several real cases [38]. Here we focus on the 3C's for the sake of brevity, but the same arguments apply to any pattern that can enrich the vocabulary of the network language for precision medicine.

Communities. Using network science terminology, modularity is often referred to as having a "community structure", *i.e.*, their vertices are organized into groups, called *communities, clusters*, or *modules* [39] as shown in *Figure 7a*.

Modularity is a key feature of living systems. Every cellular event, such as signaling or DNA replication, is the result of the presence of "modules" composed of several molecular machineries or regulatory structures, coordinately interacting directly or indirectly [40]. Indeed, at the molecular scale, the presence of modules is often described as an ensemble of gene products highly coordinated at the functional level, interacting physically and subject to co-regulation [41, 42]. Moreover, modularity may support evolutionary forces and sustain change. The organization of functions in discrete modules (possibly partially overlapped) provide robustness to change but permit changes by modifications of the interconnections among modules. This is key to allow evolvability in uncertain and noisy environments and, at the same time, maintain adaptability [40, 43]. Modularity is an omnipresent property of genomic data of all living systems which can be found in many kinds of experimental datasets, such as protein-protein or protein-DNA interactions, gene expression measurements, and many others [44]. The modularity structure of a network and identification of communities can be formally characterized in many ways. The most widely used one is the "modularity measure" defined by Newman as the fraction of edges that belong to the given communities minus the expected fraction whether links were randomly distributed [45]. Community finding algorithms using the modularity measure are based, for example, on maximum likelihood [46] or local greedy approach [47].

The identification of modules in a network may provide useful information on how it is organized by emphasizing regions with a sort of "degree of autonomy"

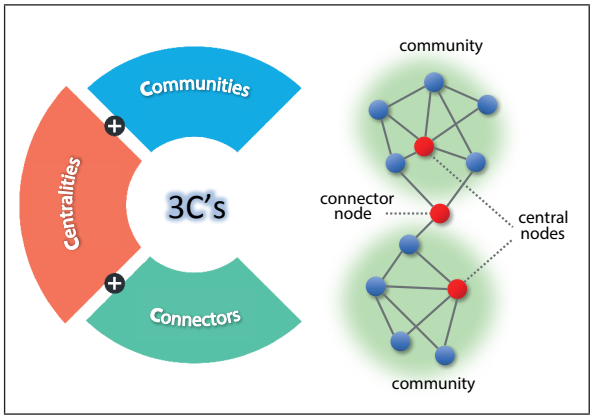

Figure 6
The three C's of network analysis. The most used "words" in the network language correspond both to mathematical patterns (community, centrality, connector) and biological concepts (co-operation, influence, mediator).

or "self-organization" within the network. It allows the classification of nodes, based on their "importance" with respect to their module. For example, we can highlight nodes that are totally embedded within their module from those which lies on the frontier between modules, which may act as "connectors" between them and, as such, play a key role both in holding them together and in the dynamics of spreading "information" throughout the network. Indeed, many tools to identify and study the biological/medical significance of modules are available in the literature, the most widely used are the weighted gene co-expression network analysis (WGCNA) [48] and the switch miner algorithm (SWIM) [49]. There is a long list of biological and clinical applications of the "module" identification, for example, WGCNA has been applied to hepatocellular carcinoma [50], calcific aortic valve disease [51], cervical cancer [52], and pulmonary artery hypertension [53], just to cite a few. SWIM algorithm has shown key modules (the so-called "switch genes") in drug response [54], miRNA cancer networks [55], glioblastoma stem cells [56], chronic obstructive pulmonary disease [57], breast cancer [58], cancer-miRNA networks [55] and disease/genes associations [59].

Centralities. A fundamental question when studying networks (both at the whole network-level or community-level) is to find candidate nodes for being the "most influential" of the whole network or of the community it belongs to. In the network science language, they are referred to as "central nodes", whilst the "mapped" biological concept can be that of a "key driver", "critical", "switch" gene or mutation, a drug "target" or a "lethal" protein, depending on the context. Therefore, measures of "centrality" summarize a node's involvement in or contribution to the cohesiveness of the network [60]. Most importantly, centrality values depend solely on the network topology, i.e., on its structure defined by how links and nodes are set up to relate to each other. A commonly used description of a node's centrality is based on three main properties: its connectedness, its role as a mediator, and its closeness to other nodes [61]. The first property may be metaphorically projected onto an interconnected social group exchanging messages (information), thus corresponding to its degree of potential communication activity, the second may be viewed as the potential to control such activity, and the third its efficiency in passing messages to all other nodes [62].

Degree centrality. One of the most popular ways to characterize node importance in terms of its connectedness is to compute its "degree centrality", i.e., the number of connections it has to other nodes (see *Figure 7c*). The underlying idea is clear and simple: degree centrality is a measure of importance based on the number of connections, the more the better. A typical metaphorical interpretation is that of an individual with many friends in a social network or that of an airport with many flights. A mapping between network properties and biological concepts is the well-known "lethality-centrality" correspondence in protein networks [63-65]. The underlying idea is that a single protein, although working as a catalyst or signaling molecule, or building block in a cell, also have a role defined by the network of interactions with other proteins (or DNA/RNA) in which it has a cellular function within functional modules [40]. By studying the *Saccharomyces cerevisiae* PPI, Jeong *et al.* [63] found that the phenotypic consequences of single gene deletion are affected by the number of interactions of its protein product, i.e., by its degree centrality in the PPI

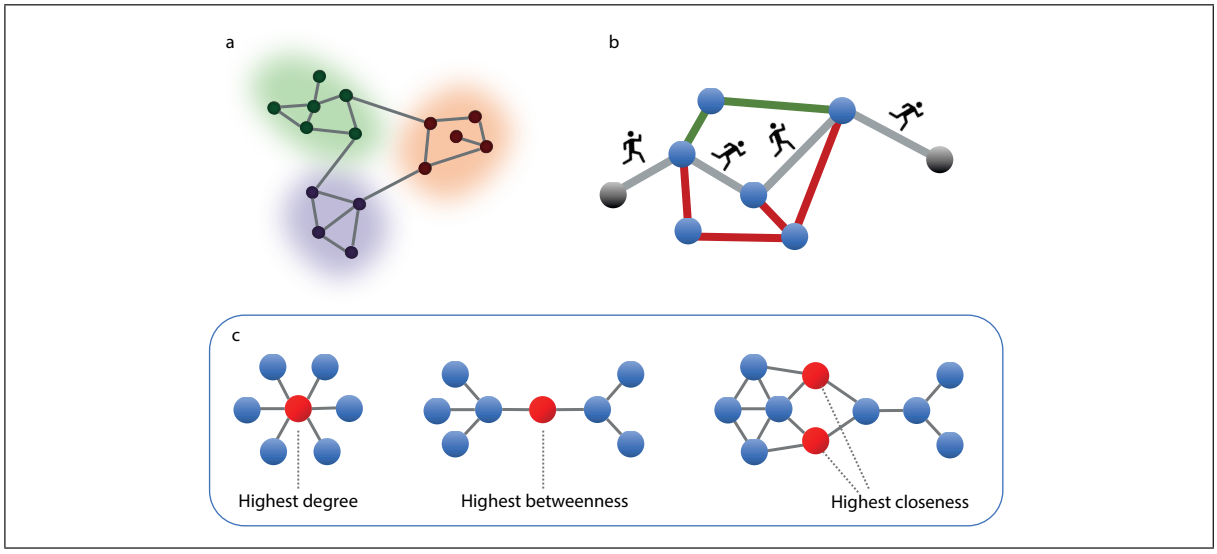

Figure 7
Network patterns. a) Network communities. Apart from computational aspects, modular networks reflect the inherent modularity of living systems and the network language provides a straightforward mapping from a biological concept (cooperating sub-units for a specific purpose or function) and topological properties and, the latter can be translated in an algorithm on data. b) Distance and shortest path(s) between two nodes in a network. The distance between source and target nodes is 4 (minimum number of hops) but there are two shortest paths with the same number of "hops" (green detour). c) Network centrality measures. Three ways to characterize "importance" from a network perspective (degree, betweenness, closeness) and from a biological perspective (e.g., lethality, integrity of functional cordination, effective information flow).

network. They showed that nodes degree is very inhomogeneously distributed with most proteins having few connections and a few of them being highly connected. The latter, it is argued, play a central role in mediating interactions within the global network. Highly connected proteins (called "hubs") are three times more likely to be essential than proteins with only a small number of links to other proteins [63]. The biological interpretation of hubs in the PPI network is still debated [66], but to our purposes, it is important to realize how the network language can function as a mapmaker between "computation on data" (degree of the PPI network) and "biological concept" (essentiality of a protein). The biological significance of degree centrality is witnessed by several publications identifying hubs as a critical feature of the interactome for disease as in diabetes mellitus [67], nephropathies [68], allergic response [69] or cancers [70-73]. Also, other types of networks have shown the relevance of degree centrality as in the gene-interaction network [74-76], miRNA-target gene networks [77], RNA-RNA networks [78], developmental regulatory networks [79], co-expression networks [49, 80, 81] or in the human brain network [82, 83].

Distance and shortest path. Other popular measures of importance are the so-called "betweenness" and "closeness" centralities. Before we discuss the mapping provided by these network characterizations, we preliminary need to define the concept of "shortest path" and "distance" between two nodes in a network (see *Figure 7b*). Given a set of nodes and links connecting them (i.e., a network) the "distance" between two nodes is usually defined as the minimum number of "hops" (links) needed to move from the source node to the target node and the corresponding path(s) are termed "shortest paths". Even if multiple shortest paths can be identified on the network from source to target nodes, the minimum number of links is a single value. An example of the mapping of the concept of "shortest path" on a biological problem, is that of finding repurposable drugs. The best candidate drug for a given set of proteins associated with a given disease, is that "close", i.e., having the smallest shortest paths, to its targets [84]. Now we can define the concept of "betweenness and closeness centrality" of a node in a network.

Betweenness centrality. A popular measure of "importance" of a node in terms of its ability as a "mediator", is called "betweenness centrality". It depends on its ability to allow nodes to reach other nodes, i.e., the extent to which a node lies between other nodes which depend on it [62]. Its formal definition coincides with the sum of the fractions of shortest paths passing through it (they may be more than one as previously shown) for all pairs of nodes [62]. A metaphorical projection on the concept of "information flow" is that a node with a high betweenness is potentially able to control such flow, that is it can facilitate, impede, or bias the "transmission" of "messages" [85] or, more generally, "information". Interestingly, using PPI data, Samokhin *et al.* [86] identified NEDD9 as a critical node in the phenotype transition from adaptive to pathogenic fibrosis using betweenness centrality, Joy *et al.* [87] found that proteins with high betweenness are more likely to be essential and that evolutionary age of proteins is positively correlated with betweenness and Duron *et al.* [88] showed results indicating robustness of betweenness centrality in the identification of target genes for drug development. Using mammalian transcription networks, Potapov *et al.* [89] showed that the top list of genes displaying high degree and high betweenness, such as P53, C-FOS, C-JUN, and C-MYC, is enriched with genes that are known as having tumor-suppressor or proto-oncogene properties.

Closeness centrality. The last example to show the use of the most popular measures of importance of a node, we consider now the so-called "closeness centrality". The terminology makes it clear that it provides information about the property of a node to be "close" to all other nodes, i.e. to be at the center of a network. The formal definition consists of two steps: first, the sum of its distances to other nodes is computed and, second, its value is defined by the inverse of such a value. In this way, high closeness values correspond to nodes that are close to all others, and the smaller the total distance of a node to other nodes, the higher its closeness is. Using the already mentioned metaphorical projection of a network as a web of "information flow", a node with a high closeness may be considered important since information can rapidly spread to all other nodes very quickly. In biological terms, one may think, for example, of a protein-protein interaction network where a misfolded protein may produce a perturbation that can produce some effect (e.g., by decreasing or increasing the binding strength) to its interaction partners and so on, thus resembling the situation in which a message rapidly spreads over a web of people starting from its "center". For example, Ozgur *et al.* [75], using closeness centrality in a gene interaction prostate cancer network, inferred the presence of 18 new potential disease genes and Amitai *et al.* [90], using a residue interaction network where amino acid residues are the nodes and their interactions with each other are the links, found that active site, ligand-binding and evolutionarily conserved residues, typically have high closeness values. Ma and Zeng [91] showed that nodes with a high closeness in a metabolic network belong to the central metabolism, namely the glycolysis and citric acid cycle pathway.

Clearly, many other centrality measures can be defined to characterize some topological property of a network. For example, we can mention barycenter, cluster rank, decay, diffusion degree, geodesic k-path, leverage, lobby, radiality, eccentricity, Kleinberg's authority scores, and Harary graph, just to cite a few [92]. The proliferation of centrality measures is not a problem, on the contrary, this is the normal and positive development of a language, where new words are "invented" every day thus increasing the vocabulary of the network language. We envisage the birth of a large dictionary of thousands of words that can be used as the building blocks for expressing biomedical properties in this new language. The situation resembles that of sign language for hearing and speech impaired people, where gestures are used to express concepts, feelings, and ideas. The key advantage of the network language is that, once the biologist/physician has found the way of expressing his/

her ideas on the medical problem of interest in terms of networks, then an algorithm on data can be readily obtained, thus avoiding the intermediate step of a unifying general theory of disease and network science.

The key issue here discussed is that the "network language" may be able to provide biological interpretations of such properties therefore drawing the attention of the experimentalist to specific nodes for further analysis.

Connectors. Connectors are nodes in the network that connect modules. This broad definition that needs to be precisely quantified – as discussed in the following – to set up an algorithm, is very interesting from a biological perspective. Indeed, communities of nodes (genes, proteins, etc.) are there because they cooperate for some purpose or function in a cell (a functional module). However, groups of cooperating entities cannot work in isolation, but they must be coordinated for proper global function. In other words, self-organization is required for an appropriate response to internal or external stimuli. The simplest way to map this biological feature on a network is to consider nodes through which different communities can "communicate", thus using the usual "information flow" metaphor. Such nodes are usually referred as "connectors", as shown in *Figure 6*. A very interesting application of this property is reported by Niss *et al.* [93] where the protein-protein interactome of dendritic cells has been studied. They found an intriguing group of 294 proteins each forming a "bow-tie" structure, that is a single protein connecting the majority of protein complexes. The latter are "communities" on the network, and such "knot" proteins at the center of the bow tie, act as connectors. Such proteins resulted to have fundamental biological properties, like multifunctional capabilities, enrichment in essential proteins, and wide expression in other cells and tissues [93].

Since connector nodes, at the community level, may not be connectors on a global scale, they do not have necessarily a high betweenness centrality value. Therefore, new formal definitions are needed and the most popular measures to characterize "connector nodes" have been provided by Guimerà and Amaral [94] and by Paci *et al.* [49]. To characterize connectors nodes in a modular network, Guimerà and Amaral [94] suggest considering two parameters: a measure of internal connectivity called "within-module degree" defined as the degree of a node by counting its links to members of the same community, and a measure of external connectivity called the "participation coefficient", which is defined in such a way that its values are close to one if its links are uniformly distributed among all the modules (or communities) and zero if all its links are within its own module. It is therefore clear, that connector nodes can be computationally identified as those having a low within-module degree and a high participation coefficient, as shown in *Figure 8*. Moreover, the figure makes it clear that four regions can be identified and the other three node's roles identified: local hubs, global hubs, and peripheral nodes.

Using the "clusterphobic coefficient" to measure external connectivity as in [49], instead of the participation coefficient, we can also identify on the co-expression network, a specific class of connectors called "switches" which has been shown to be associated with transitions of cell's state [54-57, 95].

CONCLUSIONS

There is a growing awareness of the large and increasing gap between two visions of life: that of the "life scientists", like biologists, physicians, molecular biologists, and so on, and that of the "number scientist" like physicists, engineers, mathematicians, and so on. Biologists' vision of life is centered on "concepts" like for example evolution, adaptation, development, speciation, purpose, which are not amenable of a rigorous and immediate mathematical formalization, whilst physicists/

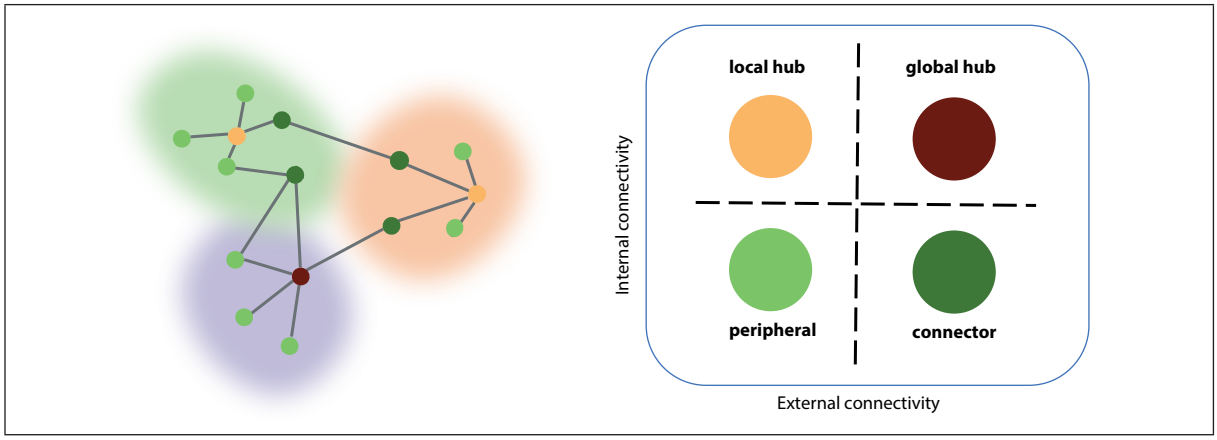

Figure 8
Connectors and other roles in modular networks. A "cartography" of a modular network can be constructed using measures of internal and external connectivity (degree). Given the modules, each node of the network corresponds to a point on the map and each quadrant corresponds to different roles. High external and internal connectivity (first quadrant) correspond to global hubs, low external and high internal connectivity (second quadrant) correspond to local hubs, low external and internal connectivity (third quadrant) correspond to peripheral nodes and high external and low internal connectivity (fourth quadrant) correspond to connectors.

data scientists' vision is centered on fitting "patterns" to data using "universal laws" or "universal organizing principles". The aim of the article is not to reconcile such visions or to take the side of one or the other, but to focus on the almost complete absence of dialogue and on how to stimulate effective interaction between the two in the field of precision medicine. An alternative view is suggested and a contribution to its solution is presented aiming to reduce the gap (that we called the "complex data divide") and allow a dialogue, by looking at network science as a vocabulary-generator filled with "words" that have different meanings in each discipline but refer to the same "thing" (cell behavior, health, disease, etc.). In this way, each researcher can continue to study his/her own discipline independently and, at the same time, engage in a true inter-disciplinary dialogue to implement precision medicine in a clinical setting.

Acknowledgements

The Author would like to thank Professor Joseph Loscalzo (Harvard Medical School) for critical insightful suggestions and Professor Elisabetta Ferretti (Sapienza, University of Rome) for her careful reading of the manuscript and comments.

This work has been partially supported by Sapienza University of Rome, "Progetto di Ateneo", grant number: RM11916B88C3E2DE.

Conflict of interest statement

The Author declares that there is no conflict of interest.

Received on 4 June 2021.
Accepted on 18 October 2021.

REFERENCES

1. Keller EF. A clash of two cultures. Nature. 2007;445(7128):603. doi: 10.1038/445603a
2. Mayr E. What makes biology unique? Considerations on the autonomy of a scientific discipline. Cambridge: Cambridge University Press; 2004. doi: 10.1017/CBO9780511617188
3. Keller EF. Making sense of life: explaining biological development with models, metaphors, and machines. Cambridge, Massachusetts; London, England: Harvard University Press; 2002. doi: 10.2307/j.ctvjz82ht
4. Steen LA. The science of patterns. Science. 1988;240(4852):611-6. doi: 10.1126/science.240.4852.611
5. Shmulevich I. Model selection in genomics. Environ Health Perspect. 2003;111(6):A328-9. doi: 10.1289/ehp.111-a328b
6. Gould SJ. Hen's teeth and horse's toes: Further reflections in natural history. New York: WW Norton & Company; 1984.
7. Uversky VN, Giuliani A. Networks of networks: An essay on multi-level biological organization. Front Genet. 2021;12:706260. doi: 10.3389/fgene.2021.706260
8. Hunter, D. J. Gene–environment interactions in human diseases. Nature Rev Genet. 6;287-98;2005. doi: 10.1038/nrg1578
9. The top 10 causes of death. https://www.who.int/news-room/fact-sheets/detail/the-top-10-causes-of-death.
10. Craig J. Complex diseases: Research and applications. Nature Educ. 1(1):184;2008.
11. Buchanan AV, Weiss KM, Fullerton SM. Dissecting complex disease: the quest for the Philosopher's Stone? Int J Epidemiol. 2006;35(3):562-71. doi: 10.1093/ije/dyl001
12. Greene JA, Loscalzo J. Putting the patient back together – social medicine, network medicine, and the limits of reductionism. N Engl J Med. 2017;377(25):2493-9. doi: 10.1056/NEJMms1706744
13. Hanahan D, Weinberg RA. Hallmarks of cancer: the next generation. Cell. 2011;144(5):646-74. doi: 10.1016/j.cell.2011.02.013
14. Tieri P, Farina L, Petti M, Astolfi L, Paci P, Castiglione F. Network inference and reconstruction in bioinformatics. Encyclopedia of bioinformatics and computational biology. Academic Press; 2019. p. 805-13, doi: 10.1016/B978-0-12-809633-8.20290-2
15. Goh KI, Cusick ME, Valle D, Childs B, Vidal M, Barabási AL. The human disease network. Proc Natl Acad Sci USA. 2007;104(21):8685-90. doi: 10.1073/pnas.0701361104
16. Pai S, Bader GD. Patient Similarity Networks for Precision Medicine. J Mol Biol. 2018;430(18 Pt A):2924-38. doi: 10.1016/j.jmb.2018.05.037
17. Silverman EK, Schmidt HHHW, Anastasiadou E, Altucci L, Angelini M, Badimon L, Balligand JL, Benincasa G, Capasso G, Conte F, Di Costanzo A, Farina L, Fiscon G, Gatto L, Gentili M, Loscalzo J, Marchese C, Napoli C, Paci P, Petti M, Quackenbush J, Tieri P, Viggiano D, Vilahur G, Glass K, Baumbach J. Molecular networks in network medicine: Development and applications. Wiley Interdiscip Rev Syst Biol Med. 2020;12(6):e1489. doi: 10.1002/wsbm.1489
18. Anderson PW. More is different. Science 1972;177(4047):393-6.
19. von Bertalanffy L. The theory of open systems in physics and biology. Science. 1950;111(2872):23-9. doi: 10.1126/science.111.2872.23
20. Boulding KE. General Systems Theory – The Skeleton of Science. Manag Sci. 1956;2(3):197-208. doi: 10.1287/mnsc.2.3.197
21. Manzoni C, Kia DA, Vandrovcova J, Hardy J, Wood NW, Lewis PA, Ferrari R. Genome, transcriptome and proteome: the rise of omics data and their integration in biomedical sciences. Brief Bioinform. 2018;19(2):286-302. doi: 10.1093/bib/bbw114
22. Kandpal R, Saviola B, Felton J. The era of omics unlimited. Biotechniques. 2009;46(5):351-2, 354-5. doi: 10.2144/000113137
23. D'Adamo GL, Widdop JT, Giles EM. The future is now? Clinical and translational aspects of "Omics" technologies. Immunol Cell Biol. 2021 Feb;99(2):168-76. doi: 10.1111/imcb.12404
24. Malloy DC, Martin R, Hadjistavropoulos T, Liu P, McCarthy EF, Park I, Shalani N, Murakami M, Paholpak S. Discourse on medicine: meditative and calculative approaches to ethics from an international perspective. Philos Ethics Humanit Med. 2014;9:18. doi: 10.1186/1747-5341-9-18
25. Collins FS. Shattuck lecture – medical and societal consequences of the Human Genome Project. N Engl J Med. 1999;341(1):28-37.

26. Joyner MJ, Paneth N. Promises, promises, and precision medicine. J Clin Invest. 2019;129(3):946-8. doi: 10.1172/JCI126119
27. Antebi YE, Linton JM, Klumpe H, Bintu B, Gong M, Su C, McCardell R, Elowitz MB. Combinatorial signal perception in the BMP Pathway. Cell. 2017;170(6):1184-1196.e24. doi: 10.1016/j.cell.2017.08.015
28. Anderson C. The end of theory: The data deluge makes the scientific method obsolete. Wired. Jun 23; 2008.
29. Pigliucci M. The end of theory in science? EMBO Rep. 2009;10(6):534. doi: 10.1038/embor.2009.111
30. Kilgore C. Rhetoric of the network: toward a new metaphor. Mosaic: a journal for the interdisciplinary study of literature. 2013;46:37-58. doi: 10.1353/mos.2013.0044
31. Palumbo MC, et al. Networks everywhere? Some general implications of an emergent metaphor. Curr Bioinformatics 1 (2006): 219-234.
32. Fauconnier G, Turner M. The way we think: Conceptual blending and the mind's hidden complexities. Basic Books; 2002.
33. Bachelard G, The formation of the scientific mind,: a contribution to a psychoanalysis of objective knowledge. Clinamen Press; 2002.
34. Degani A. A tale of two maps: Analysis of the London underground. Diagram. Ergonomics in Design. 2013;21(3):7-16. doi: 10.1177/1064804613489125
35. Nurse P. Systems biology: understanding cells. Nature. 2003;424(6951):883. doi: 10.1038/424883a
36. Brenner S. Sequences and consequences. Philos Trans R Soc Lond B Biol Sci. 2010;365(1537):207-12. doi: 10.1098/rstb.2009.0221
37. Barabási AL, Gulbahce N, Loscalzo J. Network medicine: a network-based approach to human disease. Nat Rev Genet. 2011;12(1):56-68. doi:10.1038/nrg2918
38. Lee LY, Loscalzo J. Network medicine in pathobiology. Am J Pathol. 2019;189(7):1311-26. doi: 10.1016/j.ajpath.2019.03.009
39. Santo Fortunato, Darko Hric, Community detection in networks: A user guide. Phys Rep. 2016;659:1-44.
40. Hartwell LH, Hopfield JJ, Leibler S, Murray AW. From molecular to modular cell biology. Nature. 1999;402(6761 Suppl):C47-52. doi: 10.1038/35011540
41. Mitra K, Carvunis AR, Ramesh SK, Ideker T. Integrative approaches for finding modular structure in biological networks. Nat Rev Genet. 2013;14(10):719-32. doi: 10.1038/nrg3552
42. Wagner GP, Pavlicev M, Cheverud JM. The road to modularity. Nat Rev Genet. 2007;8(12):921-31. doi: 10.1038/nrg2267
43. Hintze A, Adami C. Evolution of complex modular biological networks. PLoS Comput Biol. 2008;4(2):e23. doi: 10.1371/journal.pcbi.0040023
44. Fraser HB. Coevolution, modularity and human disease. Curr Opin Genet Dev. 2006;16(6):637-44. doi: 10.1016/j.gde.2006.09.001
45. Newman ME. Modularity and community structure in networks. Proc Natl Acad Sci USA. 2006;103(23):8577-82. doi: 10.1073/pnas.0601602103
46. Newman, M. (2011). Communities, modules and large-scale structure in networks. Nature Physics. 2011; 8:25-31. doi: 10.1038/nphys2162
47. Blondel VD, Guillaume J-L, Lambiotte R, Lefebvre, E. Fast unfolding of communities in large networks. Journal of Statistical Mechanics: Theory and Experiment. 2008, P10008.
48. Langfelder P, Horvath S. WGCNA: an R package for weighted correlation network analysis. BMC Bioinformatics. 2008;9:559. doi: 10.1186/1471-2105-9-559
49. Paci P, Colombo T, Fiscon G, Gurtner A, Pavesi G, Farina L. SWIM: a computational tool to unveiling crucial nodes in complex biological networks. Sci Rep. 2017;7:44797. doi: 10.1038/srep44797. Erratum in: Sci Rep. 2017;7:46843.
50. Li L, Xia S, Shi X, Chen X, Shang D. The novel immune-related genes predict the prognosis of patients with hepatocellular carcinoma. Sci Rep. 2021;11(1):10728. doi: 10.1038/s41598-021-89747-7
51. Sun JY, Hua Y, Shen H, Qu Q, Kan JY, Kong XQ, Sun W, Shen YY. Identification of key genes in calcific aortic valve disease via weighted gene co-expression network analysis. BMC Med Genom. 2021;14(1):135. doi: 10.1186/s12920-021-00989-w
52. Deng ZM, Dai FF, Zhou Q, Cheng YX. Hsa_circ_0000301 facilitates the progression of cervical cancer by targeting miR-1228-3p/IRF4 Axis. BMC Cancer. 2021;21(1):583. doi: 10.1186/s12885-021-08331-4.
53. Bai Z, Xu L, Dai Y, Yuan Q, Zhou Z. ECM2 and GLT8D2 in human pulmonary artery hypertension: fruits from weighted gene co-expression network analysis. J Thorac Dis. 2021;13(4):2242-54. doi: 10.21037/jtd-20-3069
54. Falcone R, Conte F, Fiscon G, Pecce V, Sponziello M, Durante C, Farina L, Filetti S, Paci P, Verrienti A. BRAFV600E-mutant cancers display a variety of networks by SWIM analysis: prediction of vemurafenib clinical response. Endocrine. 2019;64(2):406-413. doi: 10.1007/s12020-019-01890-4
55. Fiscon G, Conte F, Farina L, Pellegrini M, Russo F, Paci P. Identification of disease-miRNA networks across different cancer types using SWIM. Methods Mol Biol. 2019;1970:169-81. doi: 10.1007/978-1-4939-9207-2_10
56. Fiscon G, Conte F, Paci P. SWIM tool application to expression data of glioblastoma stem-like cell lines, corresponding primary tumors and conventional glioma cell lines. BMC Bioinformatics. 2018;19(Suppl. 15):436. doi: 10.1186/s12859-018-2421-x
57. Paci P, Fiscon G, Conte F, Licursi V, Morrow J, Hersh C, Cho M, Castaldi P, Glass K, Silverman EK, Farina L. Integrated transcriptomic correlation network analysis identifies COPD molecular determinants. Sci Rep. 2020;10(1):3361. doi: 10.1038/s41598-020-60228-7
58. Fiscon G, Pegoraro S, Conte F, Manfioletti G, Paci P. Gene network analysis using SWIM reveals interplay between the transcription factor-encoding genes HMGA1, FOXM1, and MYBL2 in triple-negative breast cancer. FEBS Lett. 2021; Apr 9. doi: 10.1002/1873-3468.14085. Epub ahead of print.
59. Paci P, Fiscon G, Conte F, Wang RS, Farina L, Loscalzo J. Gene co-expression in the interactome: moving from correlation toward causation via an integrated approach to disease module discovery. NPJ Syst Biol Appl. 2021;7(1):3. doi: 10.1038/s41540-020-00168-0
60. Borgatti SP, Everett MG. A graph-theoretic perspective on centrality. Soc Netw. 2006;8:466-84.
61. Freeman LC. Centrality in networks: I. Conceptual clarifications. Soc Netw. 1979;1:215-39.
62. Zweig KA. Network analysis literacy. Lecture Notes in Social Networks, springer, Vienna, 2016
63. Jeong H, Mason SP, Barabási AL, Oltvai ZN. Lethality and centrality in protein networks. Nature. 2001;411(6833):41-2. doi: 10.1038/35075138
64. Hahn MW, Kern AD. Comparative genomics of centrality and essentiality in three eukaryotic protein-interaction networks. Mol Biol Evol. 2005;22(4):803-6. doi: 10.1093/molbev/msi072
65. Joyce AR, Palsson BØ. Predicting gene essentiality using genome-scale in silico models. Meth Mol Biol.

2008;416:433-57. doi: 10.1007/978-1-59745-321-9_30
66. Goymer P. Network biology: why do we need hubs? Nat Rev Genet. 2008;9(9):650. doi: 10.1038/nrg2450.
67. Ambedkar C, Reddi KK, Muppalaneni NB, Kalyani D. Application of centrality measures in the identification of critical genes in diabetes mellitus. Bioinformation. 2015;11(2):90-5. doi: 10.6026/97320630011090
68. Abedi M, Gheisari Y. Nodes with high centrality in protein interaction networks are responsible for driving signaling pathways in diabetic nephropathy. Peer J. 2015;3:e1284. doi: 10.7717/peerj.1284
69. Lu X, Jain VV, Finn PW, Perkins DL. Hubs in biological interaction networks exhibit low changes in expression in experimental asthma. Mol Syst Biol. 2007;3:98. doi: 10.1038/msb4100138
70. Jonsson PF, Bates PA. Global topological features of cancer proteins in the human interactome. Bioinformatics. 2006;22(18):2291-7. doi: 10.1093/bioinformatics/btl390
71. Li Z, Ivanov AA, Su R, Gonzalez-Pecchi V, Qi Q, Liu S, Webber P, McMillan E, Rusnak L, Pham C, Chen X, Mo X, Revennaugh B, Zhou W, Marcus A, Harati S, Chen X, Johns MA, White MA, Moreno C, Cooper LA, Du Y, Khuri FR, Fu H. The OncoPPi network of cancer-focused protein-protein interactions to inform biological insights and therapeutic strategies. Nat Commun. 2017;8:14356. doi: 10.1038/ncomms14356. Erratum in: Nat Commun. 2017;8:15350.
72. Sun J, Zhao Z. A comparative study of cancer proteins in the human protein-protein interaction network. BMC Genomics. 2010;11(Suppl 3):S5. doi: 10.1186/1471-2164-11-S3-S5
73. [82. van den Heuvel MP, Sporns O. Network hubs in the human brain. Trends Cogn Sci. 2013;17(12):683-96. doi: 10.1016/j.tics.2013.09.012
74. Lehner B, Crombie C, Tischler J, Fortunato A, Fraser AG. Systematic mapping of genetic interactions in *Caenorhabditis elegans* identifies common modifiers of diverse signaling pathways. Nat Genet. 2006;38(8):896-903. doi: 10.1038/ng1844
75. Ozgür A, Vu T, Erkan G, Radev DR. Identifying gene-disease associations using centrality on a literature mined gene-interaction network. Bioinformatics. 2008;24(13):i277-85. doi: 10.1093/bioinformatics/btn182
76. Yu D, Lim J, Wang X, Liang F, Xiao G. Enhanced construction of gene regulatory networks using hub gene information. BMC Bioinformatics. 2017;18(1):186. doi: 10.1186/s12859-017-1576-1
77. Mukherjee S, Akbar I, Bhagat R, Hazra B, Bhattacharyya A, Seth P, Roy D, Basu A. Identification and classification of hubs in microRNA target gene networks in human neural stem/progenitor cells following Japanese Encephalitis Virus Infection. mSphere. 2019;4(5):e00588-19. doi: 10.1128/mSphere.00588-19
78. Banerjee S, Kalyani Yabalooru SR, Karunagaran D. Identification of mRNA and non-coding RNA hubs using network analysis in organ tropism regulated triple negative breast cancer metastasis. Comput Biol Med. 2020;127:104076. doi: 10.1016/j.compbiomed.2020.104076
79. Seo CH, Kim JR, Kim MS, Cho KH. Hub genes with positive feedbacks function as master switches in developmental gene regulatory networks. Bioinformatics. 2009;25(15):1898-904. doi: 10.1093/bioinformatics/btp316
80. Liu Y, Gu HY, Zhu J, Niu YM, Zhang C, Guo GL. Identification of hub genes and key pathways associated with bipolar disorder based on weighted gene co-expression network analysis. Front Physiol. 2019;10:1081. doi: 10.3389/fphys.2019.01081
81. Zhu Z, Jin Z, Deng Y, Wei L, Yuan X, Zhang M, Sun D. Co-expression network analysis identifies four hub genes associated with prognosis in soft tissue sarcoma. Front Genet. 2019;10:37. doi: 10.3389/fgene.2019.00037
82. van den Heuvel MP, Sporns O. Network hubs in the human brain. Trends Cogn Sci. 2013;17(12):683-96. doi: 10.1016/j.tics.2013.09.012
83. Palva S. The importance of hubs in large-scale networks. Nat Hum Behav. 2018;2(10):724-5. doi: 10.1038/s41562-018-0438-9
84. Yildirim MA, Goh KI, Cusick ME, Barabási AL, Vidal M. Drug-target network. Nat Biotechnol. 2007;25(10):1119-26. doi: 10.1038/nbt1338
85. Freeman LC. A set of measures of centrality based upon betweenness. Sociometry. 1977;40:35-41
86. Samokhin AO, Stephens T, Wertheim BM, Wang RS, Vargas SO, Yung LM, Cao M, Brown M, Arons E, Dieffenbach PB, Fewell JG, Matar M, Bowman FP, Haley KJ, Alba GA, Marino SM, Kumar R, Rosas IO, Waxman AB, Oldham WM, Khanna D, Graham BB, Seo S, Gladyshev VN, Yu PB, Fredenburgh LE, Loscalzo J, Leopold JA, Maron BA. NEDD9 targets COL3A1 to promote endothelial fibrosis and pulmonary arterial hypertension. Sci Transl Med. 2018;10(445):eaap7294. doi: 10.1126/scitranslmed.aap7294
87. Joy MP, Brock A, Ingber DE, Huang S. High-betweenness proteins in the yeast protein interaction network. J Biomed Biotechnol. 2005;2005(2):96-103. doi: 10.1155/JBB.2005.96
88. Durón C, Pan Y, Gutmann DH, Hardin J, Radunskaya A. Variability of Betweenness Centrality and Its Effect on Identifying Essential Genes. Bull Math Biol. 2019;81(9):3655-73. doi: 10.1007/s11538-018-0526-z
89. Potapov AP, Voss N, Sasse N, Wingender E. Topology of mammalian transcription networks. Genome Inform. 2005;16(2):270-8.
90. Amitai G, Shemesh A, Sitbon E, Shklar M, Netanely D, Venger I, Pietrokovski S. Network analysis of protein structures identifies functional residues. J Mol Biol. 2004;344(4):1135-46. doi: 10.1016/j.jmb.2004.10.055
91. Ma HW, Zeng AP. The connectivity structure, giant strong component and centrality of metabolic networks. Bioinformatics. 2003;19(11):1423-30. doi: 10.1093/bioinformatics/btg177
92. Ashtiani M, Salehzadeh-Yazdi A, Razaghi-Moghadam Z, Hennig H, Wolkenhauer O, Mirzaie M, Jafari M. A systematic survey of centrality measures for protein-protein interaction networks. BMC Syst Biol. 2018;12(1):80. doi: 10.1186/s12918-018-0598-2
93. Niss K, Gomez-Casado C, Hjaltelin JX, Joeris T, Agace WW, Belling KG, Brunak S. Complete topological mapping of a cellular protein interactome reveals bow-tie motifs as ubiquitous connectors of protein complexes. Cell Rep. 2020;31(11):107763. doi: 10.1016/j.celrep.2020.107763
94. Guimerà R, Amaral LA. Cartography of complex networks: modules and universal roles. J Stat Mech. 2005;2005(P02001):nihpa35573. doi: 10.1088/1742-5468/2005/02/P02001
95. Fiscon G, Conte F, Farina L, Paci P. SAveRUNNER: A network-based algorithm for drug repurposing and its application to COVID-19. PLoS Comput Biol. 2021;17(2):e1008686. doi: 10.1371/journal.pcbi.1008686