

Exploring modularity in biological networks

Abstract.

Network theoretical approaches have shaped our understanding of many different kinds of biological modularity. This essay makes the case that to capture these contributions, it is useful to think about the role of network models in exploratory research. The overall point is that it is possible to provide a systematic analysis of the exploratory functions of network models in bioscientific research. Using two examples from molecular and developmental biology, I argue that often the same modelling approach can perform one or more exploratory functions, such as: introducing new directions of research, offering a complementary set of concepts, methods and algorithms for individuating important features of natural phenomena, generating proofs of principle demonstrations and potential explanations for phenomena of interest, and enlarging the scope of certain research agendas.

1. Introduction

One would be hard pressed to deny that mathematical models have contributed to our understanding of biological phenomena. The application of mathematical ideas in fields like evolutionary biology, developmental biology, molecular and cell biology, and neurobiology has enabled scientists to identify and characterize a wide variety of important features of biological systems. Some of these applications have changed the kinds of questions biologists ask about living organisms. The formulation of specific research hypotheses in different mathematical frameworks has also facilitated explorations of biological structure, and its links to characteristic biological functions. Moreover, within systems biology and systems medicine, a rich toolbox of mathematical techniques and ideas are required to get computationally tractable models of biological phenomena.

This essay makes a case for the importance of the exploratory roles of mathematical models in bioscientific research. It claims that often mathematical modeling is useful because it is an integral and active part of exploratory science. Recent philosophical analyses have argued that exploration should be considered alongside explanation, prediction, and unification as a core function of scientific models in general (Gelfert 2016, 2018, Massimi forthcoming). To articulate and illustrate some of these exploratory functions, I will focus on mathematical models from network science used in the investigation of modularity across different biological scales.

Mathematical models, methods and algorithms from network science have made their way into biomedical research with the rise of omics projects and their focus on the collection, analysis, and management of huge amounts of biological and medical data. Networks abstractly represent very large amounts of data as sets of nodes and links, thereby facilitating the analysis and comparison of diverse biological systems: metabolic systems, gene regulatory systems, protein interactions, signaling pathways, neurobiological systems, and ecological systems. In particular, network science methods have proved to be very fruitful in the exploratory investigation of modularity, understood as a key organizational property that spans many scales in biological systems.

I start in section 2 by sketching an analytic framework for thinking about the exploratory roles of models in general. Section 3 provides an informal and brief presentation of some of the research

questions concerning modularity that have been approached with the tools of network science. Section 4 makes the case that these modelling efforts are best understood as an important part of exploratory science. The conclusions reflect on the implications of this analysis for the scientific and philosophical debates concerning the role of network theoretical approaches in bioscientific research.

2. Exploratory mathematical models

A widespread view of modelling among philosophers of science is that models are instruments that perform a series of epistemic functions in scientific inquiry (prediction, explanation, unification, integration, etc.) in virtue of their capacity to represent accurately at least some of the relevant features of the systems under investigation. Philosophy of modeling comprises a wide range of views according to which models are epistemically successful by directly or indirectly representing their target systems (e.g., Giere 1988, Godfrey-Smith 2006, Weisberg 2013, Levy 2015, Frigg and Hartmann 2018). Given this preference for analyzing model-target relationships, modelling efforts in which scientific inquiry is still in search of a specific target, or in which the same model has the potential to answer multiple research questions or is being used to compare multiple targets, have received comparatively less philosophical attention (although see Knuuttila and Loettgers 2016; O'Malley and Parke 2018).

In this paper, I make the case that philosophers should cast their nets wider and scrutinize how models (and in particular mathematical models) contribute to exploratory science. This requires a refinement and extension of the terms in which one analyzes the epistemic contributions of scientific models. Much of what I have to say about the exploratory functions of network models in bioscientific research however is meant to complement rather than challenge previous philosophical analyses (e.g., Levy and Bechtel 2013, Jones 2014, Green et al. 2017, Kostić 2019, Rathkopf 2018). But even without doing a one-eighty in ongoing debates, I maintain that differentiating and articulating the exploratory functions of network models affords a better understanding of how researchers use these tools and methods to investigate in greater depth¹ and detail various aspects of biological systems.

The philosophical literature on scientific experimentation is a good starting point for sketching a more general notion of exploratory research. Scientists as well as philosophers and historians of science have acknowledged that exploratory experiments are not only common but also important for discovering new and interesting scientific phenomena, typically when theoretical frameworks are still under construction (Steinle 1997, Waters 2007, Feest 2012). Exploratory experimentation has been initially characterized in contrast with confirmatory or hypothesis-testing experiments which has also reinforced the idea that theories have no substantial role to play in this mode of experimentation. In more recent debates, and particularly in relation to contemporary examples, many agree that exploratory experiments can be informed by theories. Although they do not aim to evaluate theories, exploratory experiments are said to be “guided” or “directed” by theories. To make the analysis of

¹ The notion of ‘depth’ in this context reflects the fact that mathematical methods may be used in bioscientific research to investigate whether some formally defined property (e.g., functional robustness or modularity) occurs in different kinds of biological systems or how such a property is related to other systemic properties. The notion is akin to Kostić’s (2018) idea of ‘explanatory depth’ in so far as it can be understood in terms of the richness of counterfactual relations that an investigation supports.

exploratory experiments more precise, Franklin-Hall (2005) has introduced the distinction between *background* and *local* theories²: exploratory experimentation is not immune to background theories, but it does lack direction from local theories (Elliott 2007, O'Malley 2007).

While the distinction between background and local theories is an important one³, relying solely on it risks to provide too narrow a picture of how theories are used in exploratory experimentation or research more generally. Also, the distinction does not shed light on why scientists turn to exploratory experimentation or modelling when they do. Saying that exploratory experiments are not designed to test hypotheses is not the same thing as saying that they lack “direction” from background or local theories (see Colaço 2018). This lesson, I think, transfers to the practice of modelling. Theories, either general or local, need not be the target of confirmation to have their uses in exploratory experimentation or modelling.

Thus I claim that the general picture of exploratory research need not be reduced to: “that mode of inquiry which scientists revert to whenever there are no ‘finished’, full-fledged theories waiting to be tested.” Instead, a positive characterization should include at least the following features. Exploratory research marks the starting point of new research projects. It deals in providing proofs-of-principle demonstrations, sketching how-possible explanations of observable phenomena and assessing the suitability of some chosen target system which is thought to be the appropriate object of scientific investigation (Gelfert 2016). Exploratory models can also be used to produce modal knowledge, about what is *causally* or *objectively* possible (Massimi forthcoming).⁴ Both via experimental and modelling practices, exploratory research is often concerned with concept formation, hypothesis formulation, and theory construction (Feest 2012). In some instances, it aims to fill in gaps in current theorizing about specific entities or phenomena. That is, exploratory research is sometimes directly linked to the goal of providing more detailed or in-depth descriptions of the objects of scientific investigation. This list of epistemic functions of exploratory modes of investigation is meant to be representative rather than exhaustive. How background and local theories are used to perform each of these functions is a question that can be answered only against the backdrop of a specific research context.

² According to Franklin-Hall’s account, in investigating a target system S, a theory T_1 counts as a *background* theory if it represents a series of features of S, which the system has in virtue of being part of a larger class of systems. In contrast, a theory T_2 counts as a *local* theory if it represents something more specific about the components or organization of the target system S. For example, a theory about neural signaling counts as a background theory when experimenting on the hippocampus, whereas a mechanism schema of the hippocampus counts as a local theory in this context. So, local theories are typically more detailed accounts of the inner workings and organization of the particular object of a scientific investigation.

³ For one, the distinction draws attention to the fact that not all theoretical content which might be relevant in designing an exploratory experiment must be *uniquely* or *directly* about the particular system under investigation. Sometimes theories about the class of systems suffice to initiate and conduct exploratory research. Also, the distinction is useful in highlighting (if somewhat indirectly) that exploratory experiments are not pursued *only* in the initial stages of some research venture. Scientists engage in exploratory research even when they have theories relevant to the kind of system they investigate, but they want to construct or formulate more specific local theories.

⁴ Focusing on targetless fictional models and hypothetical perspectival models, Massimi (forthcoming) argues that sometimes scientific advances come from practices such as assigning properties to fictional targets (which stand as proxies for the phenomena of interest) or bootstrapping how-possible inferences that are progressively transformed and improved as more evidence becomes available. For instance, Maxwell’s honeycomb model of the ether counts as a targetless model which was able to generate knowledge about a *causal possibility* in the context of studying the phenomenon of electromagnetic induction, i.e., about how a magnetic field *might cause* a change in charge through a wire. Perspectival models used by physicists to search for supersymmetric particles have hypothetical targets and their primary function is to establish what *might exist*, or *objective possibilities*.

Raising this contextualist issue should help clarify what the previous characterization of the functions of exploratory research does and does *not* claim. On the one hand, despite the differences between exploratory and confirmatory or hypothesis-testing research, background and local theories figure in both contexts. In exploratory research, they implicitly or explicitly justify the use of the chosen (often theoretically complex) experimental techniques, or of the theoretical abstractions and idealization assumptions that make, for instance, mathematical modelling possible. Exploratory experiments and models can be used to develop hypotheses about possible causal relations and components. This supports the idea that exploratory research is often pursued with the aim of formulating more substantive hypotheses (theories, concepts) about the systems under investigation.

On the other hand, confirmatory and exploratory research are neither at odds with each other nor competing modes of investigation. Thinking of them as complementary is more fitting. Some exploratory experiments are followed up by confirmatory experiments. Ditto for models. Sometimes confirmatory experiments are not in the books for a long time and scientific inference is driven primarily by the results of exploratory experimentation and modelling. And, other times, the boundaries are more blurry. Again, the contextualist proviso is key here. Characterizing what is distinctive about the exploratory mode of research requires a detailed analysis of the research context, the role of background and local theories, of the methods and techniques used, as well as of the specific aims pursued by the scientists engaged in a specific type of investigation. The contrast with the hypothesis-testing, theory-evaluation mode of doing research is not intended to deny its importance, but to provide a more well-rounded picture of scientific practice.

With this general characterization of exploratory research in place, I turn next to showing that exploration is an important mode of investigation in network approaches to biological phenomena. In particular, I make the case that the contributions of some network models to the investigation of biological modularity are to be understood in terms of the exploratory functions that these models perform. To clarify the scope of my thesis, I should stress that I do not take the exploratory functions identified in this essay to exhaust the epistemic contributions of network theoretical methods to bioscientific research in general. Neither do I claim that network models are the only exploratory models used by biologists to tackle the problem of modularity. But they do constitute an important class of mathematical models which are well-suited to perform a series of exploratory functions.

In particular, the chosen case studies from contemporary research show how network modelling contributes to investigations of biological modularity by performing the following exploratory functions:

- I. generating new research questions about previously studied phenomena;
- II. constructing proofs-of-principle for certain empirical or theoretical possibilities;
- III. providing concepts for describing the distinctive organizational features of biological modularity (e.g., hierarchical organization and nestedness) and their relation to structure.

My analysis is compatible with the idea that the same network model often fulfills more than one function in exploratory research. While the list above does not aim to be a comprehensive taxonomy of exploratory functions that models can play in scientific research, it should give a sense of how central exploratory modelling is in the biomedical domain. Mathematical methods and models from network science have inspired and enabled significant advances in understanding the structure and

function of biological systems. This essay focuses on the contributions of network science to the investigation of the phenomenon of biological modularity. The case study introduced in section 4 suggests that scientists insistently use network models to explore the conditions under which modularity can emerge, support functional stability and influence the development of structure. Two sets of preliminary considerations will set the stage for this discussion: (i) a selective introduction of some of the research questions concerning the phenomenon of biological modularity pursued in contemporary research, and (ii) a general and informal overview of the methodology of network science, focusing on the concepts and methods that figure in the chosen set of case studies.

3. The search for biological modularity

Modularity has become a central concept in recent debates in many subfields of the life sciences, like evolutionary, molecular, and developmental biology, neuroscience and cognitive science. The concept has been applied to capture the fact that across many different scales, biological systems can be divided into smaller systems or components, with limited or controlled interactions between them. So while these biological components are themselves sets of internally strongly interacting parts, they are relatively autonomous with respect to each other. Different types of elements (nucleotides, proteins, morphological characters, etc.) can constitute a module, and different types of connections and links (physical, dynamical, or statistical) can be taken to determine the unity of a module. For instance, developmental modules are typically considered to be either different parts of the embryo that interact with each other (in induction or morphogenesis) or as sets of interacting molecules that act independently in the patterning of different types of tissues. In adult organisms, modules are taken to be made of cohesive parts which together perform some physiological or behaviourally relevant function. For evolutionary biology, variational modules are sets of traits that vary together and are somewhat independent from the evolution of other traits. Thus, a module, be it a variational, functional or developmental, is a part of an organism that is *integrated* with respect to a certain kind of process (natural variation, physiological function, developmental process, etc.) and *relatively autonomous* with respect to other parts of the organism (Melo et al. 2016).

This schematic definition of modularity is key to entire research programs which investigate the link between modularity and other systemic biological properties such as functional robustness, plasticity, disease, and evolvability. Experimental and theoretical work in fields like evolutionary biology, molecular biology, cell biology, developmental biology, neuroscience, and cognitive psychology has generated a wide array of ideas and hypotheses about these links, as well as more precise classifications of different types of modularity exhibited by biological systems.

While not all biomedical research on modularity falls neatly in the category of exploratory science, there are more than enough examples which show that experimental and modelling efforts in this area are not always pursued with the aim of confirming a specific hypothesis or testing a local theory. Some models are primarily concerned with comparing different ways of measuring biological modularity, others with refining existing measures, or developing new more precise ways to quantify the degrees of modularity in a system at a given scale⁵, and yet others with articulating causal

⁵ For relevant reviews of current methods used for module detection and modularity measures, see e.g., Fortunato 2010; Sporns and Betzel 2016; Gosak et al. 2018.

hypotheses about how systemic properties like modularity and functional stability, or modularity and evolvability are interconnected. All these latter contributions that current modelling strategies make the investigation of biological modularity fall under the umbrella of exploratory research.

For a more concrete articulation of these epistemic contributions of exploratory models, I will focus on network approaches to modularity. As pointed out above, a system is modular if it can be divided into internally strongly interacting components which in turn are relatively autonomous with respect to one another. Elements in a module are said to be more highly connected among each other than with the elements of other such groups. Thus network science notions such as those of patterns and strengths of connectedness will play a key role in formulating hypotheses about biological modularity.

The first motivation for expecting that network science will generate exploratory models is that it makes available concepts and measures that contribute to the articulation of relevant empirical hypotheses and research questions about biological modularity. Here is a sample of such questions. Are all types of modules (e.g., in metabolic, protein-protein interaction, gene regulatory networks) to be conceptualized as groups of tightly connected nodes? Must biological modules, across different scales, have similar connectivity profiles? Can nodes belong to more than one module? What does this imply about the organization of the system or the type of modularity it possesses?

The second motivation for looking at network science as a source of exploratory models has to do with the mathematical character of the framework itself. As a rich collection of mathematical resources, network science makes possible the construction of models in which the idealization, abstraction and approximation assumptions are clearly articulated. Some of these assumptions can be justified with respect to some research questions but not with respect to others, and methods for measuring modularity can shift, e.g., within a multiscale modelling study (Gosak et al. 2018), depending on the data one is dealing with. While the mathematical formalism of network science has its own internal dynamics (i.e., rules and conventions about how certain formulas can be modified, parameters defined, etc.), its application to the biological domain is constrained by a series of conceptual and empirical considerations. The choice of a specific mathematical method for detecting how modular a biological system is, depends on how modules are conceptualized in the first place. Are they sets of nodes with similar connectivity profiles? Can a node belong to more than one module? Also, not all methods will work for exploring different kinds of data or different empirical questions. For instance, in brain network science, functional connectivity data⁶ is represented in signed correlation matrices⁷, which in turn require mathematical approaches that can deal with signed

⁶ In network neuroscience functional networks are constructed from estimates of statistical dependencies between neuronal and regional time series data, and functional connectivity is typically measured using Pearson cross-correlations of hemodynamic or electrophysiological time courses (Lorenz, Jeng and Deem 2011).

⁷ Signed correlation matrices represent both positive and negative correlation coefficients between variables. While some network approaches disregard negative correlations altogether, some recommend introducing new indices for generalizing already existing formulas for computing network measures like the clustering coefficient (ibid.).

correlations. Or, if one is interested in investigating the time evolution of a network, one is justified to prefer multislice approaches⁸ to other analysis methods.

More generally, network science has the potential to evolve into a *mature mathematical formalism*⁹ that will be an integral part of how scientific understanding of biological modularity (among other features of complex biological systems) is formulated and further developed. In other words, network science might well turn out to be an essential element in concept formation and theory construction in the biological sciences (cf. Feest 2012). To bolster further the motivation for choosing network science to ground the argument for exploratory modelling, I preface the analysis of my chosen case study with an informal presentation of some key features of network science methodology.

4. Network science and modularity

4.1. Preliminaries

Applications of network science to biology rely on methods and concepts from many mathematical fields such as: graph theory, statistics, information theory, dynamic systems theory, machine learning, and measurement theory (Börner, Sanyal, and Vespignani 2007). Among these, graph theory is the primary analytical framework used for representing the complex systems studied by biologists. Within this framework, a graph is a set of nodes and edges. Nodes are used to represent the elements or parts of a system of interest, while edges stand for connections between pairs of nodes. Nodes can be connected directly by edges or indirectly via a set of nodes and edges. Edges can be directed or undirected, and they can be binary (present or absent) or they can take fractional values.

The methodology of network modelling typically comprises four distinct steps: (1) define network nodes, (2) define edges, (3) represent the complete set of nodes and edges in a connection matrix that specifies which pairs of nodes are directly connected by an edge, and (4) analyze the resulting network with appropriate network-theoretical tools (see Bullmore and Sporns 2009: 187). While some of these steps remain implicit in the presentation of scientific studies, they are important for evaluating the epistemic contributions of network models to the investigation of a specific biological phenomenon. If one is to show that some of these contributions fall under the umbrella of exploratory research, one must be able to identify the research question pursued in a particular investigation, the aim of the researchers, and the background and local theories informing the different modelling choices made in steps (1)-(4): about what is represented in the model and what is not, which algorithm

⁸ A multislice network represents a system as a series of slices in which each slice stands for a separate observation of the network. Multislice approaches can detect variations of modular organization across sets of data or time points (ibid.).

⁹ Gelfert (2011) conceives of a mathematical formalism as “a combination of a mathematical framework (*operator algebra*) with a fundamental physical theory (*quantum mechanics*), in conjunction with additional assumptions and heuristic rules that are neither entailed by fundamental theory nor tailored to any empirical problem in particular.” (25) In a ‘mature’ mathematical formalism, mathematical resources have evolved from mere shorthands to powerful heuristic devices and theoretical tools. Mathematical formalisms are also developed in the biosciences; they are sufficiently general to be applicable to a range of phenomena or systems, but not universal; they come with an empirical interpretation, so they amount to more than using the formal language of mathematics. They are an integral part of how new knowledge about the biological objects of investigation is produced and formulated.

is chosen to analyze the results of the model and why, how the results are cross-validated with data from existing studies, and so on.

Network-theoretical approaches have repeatedly and robustly identified modularity as an important organizational feature of biological systems. Within a network theoretic framework, modularity is topologically defined in terms of specific patterns of connectivity between nodes. Modules are “communities of nodes that share greater numbers of mutual connections within each community and fewer connections between them” (Sporns 2011, 113). Moreover, modularity is identified as an organizational property that spans multiple biological scales. Many complex systems display the *fractal* property of hierarchical modularity in which “roughly the same kind of community structure is expressed repeatedly at different hierarchical levels or topological scales of the network” (Meunier, Lambiotte, and Bullmore 2010, 2; Ravasz et. al. 2002).

4.2. Network models in exploratory research

4.2.1. Some general features

The potential of network theoretical approaches to play distinctive exploratory functions in the investigation of biological modularity transpires from some of the general characteristics outlined above. Bringing these features into focus will also help refine some characteristics of exploratory research that have been glossed over in section 2.

First, as a mathematical formalism, network science comes with its own toolbox of concepts, methods and algorithms that are used to identify and formulate different types of connectivity in biologically complex systems. Defining and investigating modularity in a network-theoretical framework is therefore *not exclusively* based on previous background or local theories or hypotheses about the internal organization of a biological system. That is not to say that network modelling results are not confronted with empirical evidence or with previously accepted hypotheses. But these “comparisons” are not to be interpreted (in every research context) as confirmations or disconfirmations of said hypotheses. Exploratory modelling needs background theories to establish the relevance or adequacy of specific methodological choices, while local theories function as auxiliary hypotheses in the interpretation of the results or findings of particular models. Experimental evidence and existent empirical hypotheses are also key in justifying the interpretation given to certain network theoretical measures. So, exploratory modelling does not operate in some sort of epistemic or theoretical vacuum. Rather, it is a mode of inquiry that makes different uses of local and background theories, and expects different types of results.

Second, the exploratory character of network theoretical approaches to modularity is also linked to its commitment to data-driven methodologies. But the two are not interchangeable. We could not make sense of the many functions of exploratory research (as emphasized in section 2) merely in terms of the features of big data methods (Ratti 2015). The latter offer a fertile ground for applications of exploratory modelling in at least two senses. They provide a larger search space for the identification of novel or “missed” empirical regularities and they allow for the application of certain mathematical techniques that would be methodologically inappropriate to smaller collections of data. Moreover, exploratory studies are possible even without appealing to huge collections of data (Gelfert 2016). So,

despite the affinity between network approaches and big data research we should resist equating the exploratory uses of network models with the affordances of the latter type of methodologies.

Third, due to their mathematical character network approaches make it easier to compare and integrate data pertaining to different levels of structural, functional and causal organization in biological systems. This characteristic has been already discussed by a number of philosophers of biology (Levy and Bechtel 2013; Green et al. 2017; Kostic 2018). But while previous philosophical analyses highlight how this integration facilitates explanation of important biological phenomena, I emphasize its exploratory potential. By connecting different levels of biological organization, network theoretical analyses of modularity can be used to explore whether certain organizational principles, e.g., topological efficiency, global efficiency of information transfer, and functional robustness are conserved “over different scales and types of measurement, across different species and for functional and anatomical networks” (Bullmore and Sporns 2009, 196).

Finally, modularity analyses provide a general framework for addressing anew some key theoretical questions concerning the link between a modular architecture and evolvability (Wagner, Pavlicev, and Cheverud 2007) or between the level of modularity of a system and its robustness. It is important to insist though that network models by themselves do not answer these questions. Their contributions are best understood in terms of expanding the set of concepts, methods and ideas that enable scientists to address these questions. They are not a substitute, but a complement to the experimental and theoretical approaches that have more traditionally dealt with them.

Now I turn to the analysis of an example of contemporary research on biological modularity in metabolic systems. I keep the discussion of the methodological details of this case study largely informal. My primary aim is to bring out the exploratory character of the modelling strategies involved and to show how this perspective enriches our understanding of the contributions of network theoretical approaches to biological knowledge.

4.2.2. Modularity in metabolic networks

One of the early influential network analyses of the modularity of metabolic systems is Ravasz et. al 2002. Their graph theoretical approach was based on a large collection of empirical data which revealed that metabolic networks have both the property of having a scale-free architecture¹⁰ as well as the property of being highly clustered¹¹. While the first network measure suggests that metabolic systems have only some hub nodes with many dispersed connections, the second measure is standardly associated with the presence of fairly isolated modules. This raises the research question explored in Ravasz et al. 2002: are metabolic systems organized as a collection of functional modules or are they highly integrated systems?

¹⁰ A network is called scale-free if the features of the network are independent of its size (the number of nodes). The structure of a scale-free network is characterized by a power-law distribution of the number of edges connecting its nodes. In a power-law distribution, the number of nodes with a really high number of edges is much bigger than in a normal law distribution.

¹¹ The clustering coefficient is a measure of the degree to which nodes in a graph tend to cluster together. When the connections in a graph are dense, the clustering coefficient is high.

To address this issue, Ravasz and colleagues constructed a model of a network that would exhibit both properties. They began with a module of four completely interconnected nodes arranged as in Figure 1(A). They then created three additional replicas, connecting the three peripheral nodes of each replica to the central node of the original module (Figure 1(B)). This procedure is then iterated (the next iteration is shown in Figure 1(C)). The network that results after several iterations exhibits high average clustering, due to it being built by replicating highly-clustered units, and is scale-free, as the peripheral units are connected in a replica only to a common unit in the original.

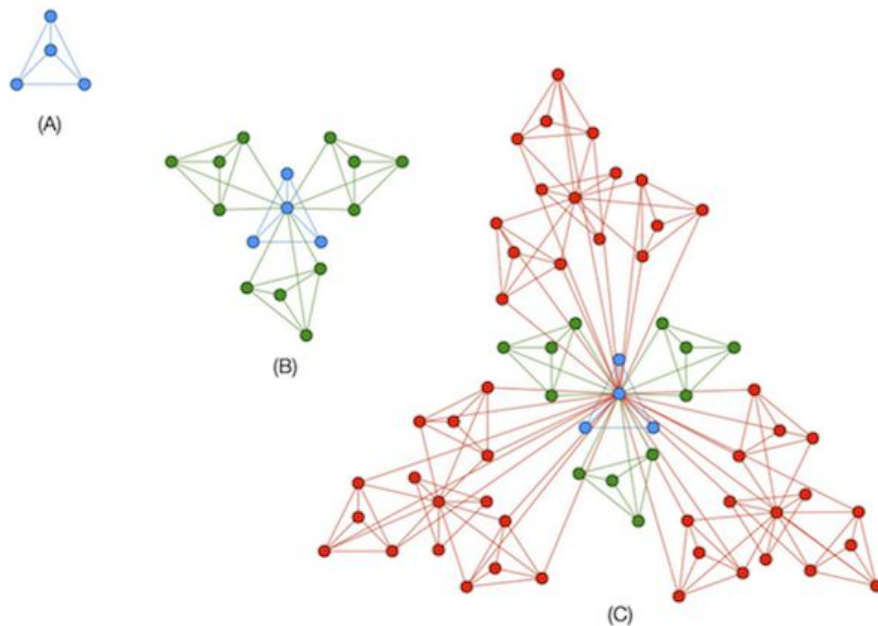


Figure 1. The method Ravasz et al. employed to build hierarchical modular networks that are both highly clustered and scale-free. The module shown in (A) is replicated three times, where in (B) the peripheral nodes of each replica are connected to the central node of the original module. This step is iterated one more time in (C), but can be repeated indefinitely.

The researchers observed that nodes that appear in the center of the network model have the highest degree¹² (i.e., are hubs) and the lowest clustering coefficient, while the nodes at the periphery of each of the super-modules (shown in different colors) have low degree and the largest clustering coefficient. In between are nodes with moderate degree and moderate clustering; these serve to integrate the smallest modules into larger modules. This creates a hierarchy in which the nodes generated at each replication step are connected not only to their own central node but also to the central node of the entire structure. Ravasz et al. refer to this network structure as a *hierarchical modular network* and propose that networks organized in this way might account for the combination of high clustering and scale-freeness in the metabolic networks of the 43 species they examined.

¹² The degree of a node in a network is the number of connections it has to other nodes; and the degree distribution is the probability distribution of these degrees over the whole network.

Having provided a proof-of-principle that modular networks can have both high clustering and the scale-freeness property, Ravasz et al. next represented the metabolic network of *E. coli*: each substrate corresponding to a node, and chemical interactions between substrates corresponding to edges. After several simplifying procedures that did not change the topology of the network, they constructed an overlap matrix to identify the modules in the metabolic system of *E. coli*. In informal terms, the overlap matrix shows, for each pair of substrates, the probability that any two substrates i and j are connected to the same other substrates. From the overlap matrix one can easily identify regions with high overlap and regions with less substrate overlap. The former were taken to correspond to the modules in the proposed type of hierarchical modular organization.

To explore whether these modules play a functional role in the metabolism of *E. coli*, the investigators identified the chemical classes to which the substrates belong: carbohydrates, lipids, amino acids, nucleotides and nucleic acids, and coenzymes and vitamins. In this representation, it becomes clear that most of the overlaps involve substrates from the same classes. In the next modelling step, Ravasz et al. used a clustering algorithm to group substrates based on their overlap; the results are shown along the top and right-hand edges. Because these clusters closely correspond to known chemical groups, they were viewed as picking out modules that correspond to biochemical functions. Using this same coloring scheme, the researchers represented the substrates in a three-dimensional network model, focusing in more detail on a specific module (involved in pyrimidine metabolism) and linked each sub-module to particular pathways in the module.

Overall, this exploratory model showed that the two features of metabolic networks are not incompatible. In fact, they point to a *hierarchical modular* organization of metabolic systems. The study also showed that modules closely corresponding to traditionally characterized biochemical pathways or mechanisms can be recovered from large-scale networks. Moreover, since the fit between the modules and previously established metabolic pathways is significantly close, Ravasz and colleagues emphasized that the boundary of modules is sometimes at variance with standard boundaries in biochemical pathway models, suggesting that “further experimental and theoretical analyses will be needed to understand the relation between the decomposition of *E. coli* metabolism offered by our topology-based approach and the biologically relevant subnetworks” (Ravasz et al. 2002: 1555).

In this research context, network models have the exploratory function of providing new concepts and methods to identify the type of modularity exhibited by metabolic systems. While the identification of the *hierarchical* character of modular organization of metabolic networks was driven by another network theoretic result rather than by a specific local theory about *E. coli*'s metabolism, the results of the network analysis were confronted with previous classifications of metabolic modules in terms of their biochemical functions. This is in line with the previous general claim that exploratory research done with network models is guided both by background and local theories.

A second exploratory function of this modelling strategy was to provide a proof-of-principle for the possibility of hierarchical modular organization in biological systems. A cursory look at current research shows also that this study served as a starting point in the investigation of new questions about the gradual and hierarchical character of modularity. In other words, this exploratory model generated a set of novel research questions: Are some metabolic networks more modular than others? Does the modularity of a system depend on its size? What role does environmental variability play in

determining the degree of modularity of a system? Are processes like horizontal gene transfer relevant for the nested character of the modularity observed in metabolic systems? (Lorenz et al. 2011)

Finally, while network models by themselves may not be able to establish how the modular structure of bacterial metabolic networks is linked to environmental variability and horizontal gene transfer, they have the potential of providing again important proofs of principle that such links exist. This can then trigger the search for *how-possible* explanations of how this link came to be established evolutionarily.

5. Conclusions

Network theoretical approaches have shaped our understanding of different kinds of modularity and their importance for a variety of biological phenomena. This essay made the case that to capture these contributions, it is useful to think about the role of network models in exploratory research. The overall point is that it is possible to provide a systematic analysis of the exploratory functions of network models in bioscientific research. Using examples from molecular biology and developmental biology, I argued that often the same modelling approach can perform one or more exploratory functions, such as: providing new directions of research, offering a complementary set of concepts, methods and algorithms for individuating important features of natural phenomena, and enlarging the scope of certain research agendas. The list of examples could have been easily extended to include network theoretic studies from evolutionary biology, cell biology, ecology, physiology and neuroscience, to count just a few of the biological fields where these approaches are currently highly influential.

Focusing on the exploratory functions of network models in bioscientific research provides a complement to other contemporary philosophical analyses which have been primarily concerned with vindicating the explanatory virtues of network theoretic approaches (e.g., Levy and Bechtel 2013; Colombo 2013; Craver 2016; Green et al. 2017). I suggest that thinking about network modelling in terms of their exploratory functions provides a more robust analytic framework for understanding both the effervescence of contemporary network theoretic research in the biomedical domain and the sobering methodological cautions raised by practicing scientists.

References

Börner, K., Sanyal, S. and Vespignani, A. (2007). “Network Science.” *Annual Review of Information Science and Technology* 41:537–607.

Bullmore, E., and Sporns, O. (2009). “Complex Brain Networks: Graph-Theoretical Analysis of Structural and Functional Systems.” *Nature Reviews Neuroscience* 10:186–98.

Burian, R. M. (1997). “Exploratory Experimentation and the Role of Histochemical Techniques in the Work of Jean Brachet, 1938-1952”. *History and Philosophy of the Life Sciences* 19 (1):27 - 45.

Burian, R. M. (2007). "On MicroRNA and the Need for Exploratory Experimentation in Post-Genomic Molecular Biology". *History and Philosophy of the Life Sciences* 29 (3):285 - 311.

Callebaut, W., and & Rasskin-Gutman, D. (eds.) (2005). *Modularity. Understanding the Development and Evolution of Natural Complex Systems*. MIT Press.

Colaço, D. (2018). "Rethinking the role of theory in exploratory experimentation". *Biology and Philosophy* 33 (5-6):38.

Colombo, M. (2013). "Moving forward (and beyond) the modularity debate: A network perspective". *Philosophy of Science* 80 (3):356-377.

Craver, C. F. (2016). "The Explanatory Power of Network Models". *Philosophy of Science* 83 (5):698-709.

Elliott, K. (2007). "Varieties of Exploratory Experimentation in Nanotoxicology". *History and Philosophy of the Life Sciences* 29 (3):313 - 336.

Feest, U. (2012). "Exploratory Experiments, Concept Formation, and Theory Construction in Psychology". In Uljana Feest & Friedrich Steinle (eds.), *Scientific Concepts and Investigative Practice*. De Gruyter. pp. 167-189.

Frigg, R. and Hartmann, S. (2018). "Models in Science", *The Stanford Encyclopedia of Philosophy* (Summer 2018 Edition), Edward N. Zalta (ed.), URL = <https://plato.stanford.edu/archives/sum2018/entries/models-science/>.

Gelfert, A. (2018). "Models in Search of Targets: Exploratory Modelling and the Case of Turing Patterns". In A. Christian, David Hommen, N. Retzlaff & Gerhard Schurz (eds.), *Philosophy of Science. European Studies in Philosophy of Science, vol 9*. Springer International Publishing. pp. 245-269.

Gelfert, A. (2016). *How to Do Science with Models: A Philosophical Primer*. Springer.

Gelfert, A. (2011). "Mathematical formalisms in scientific practice: From denotation to model-based representation". *Studies in History and Philosophy of Science Part A* 42 (2):272-286.

Giere, R. (1988). *Explaining Science. A Cognitive Approach*. Chicago University Press: Chicago.

Green, S., Şerban, M., Scholl, R., Jones, N., Brigandt, I. and Bechtel, W. (2018). "Network analyses in systems biology: new strategies for dealing with biological complexity". *Synthese* 195 (4):1751-1777.

Godfrey-Smith, P. (2006). "The strategy of model-based science". *Biology and Philosophy* 21 (5):725-740.

- Gosak, M., Markovic, R., Dolensek, J., Slak Rupnik, M., Mahrl, M., Stozer, A., and Perc, M. (2018). "Network science of biological systems at different scales: A review." *Physics of Life Review* 24: 118-135.
- Fortunato, S. (2010). "Community Detection in Graphs." *Physics Reports* 486:75–174.
- Franklin-Hall, L. R. (2005). "Exploratory experiments". *Philosophy of Science* 72 (5):888-899.
- Klingenberg CP. (2008). "Morphological integration and developmental modularity". *Annu Rev Ecol Evol S.* 39:115–132.
- Knuuttila, T. & Loettgers, A. (2016). "Model templates within and between disciplines: from magnets to gases – and socio-economic systems". *European Journal for Philosophy of Science* 6 (3):377-400.
- Kostić, D. (2019). "Minimal structure explanations, scientific understanding and explanatory depth". *Perspectives on Science* 27(1):48-67.
- Levy, A. (2015). "Modeling without models". *Philosophical Studies* 172 (3):781-798.
- Levy, A., and Bechtel, W.(2013). "Abstraction and the Organization of Mechanisms". *Philosophy of Science* 80 (2):241-261.
- Lorenz, D. M., Jeng, A., and Deem, M. W. (2011). "The Emergence of Modularity in Biological Systems". *Phys Life Rev.* 8(2): 129–160.
- Massimi, M. (forthcoming). "Two kinds of exploratory models". *Philosophy of Science*. <https://doi.org/10.1086/705494>
- Melo, D., Porto, A., Cheverud, J.M., and Marroig, G. (2016). "Modularity: Genes, Development and Evolution". *Annu. Rev. Ecol. Evol. Syst.* 47:463–86.
- Meunier, D., Lambiotte, R. and Bullmore, E. (2010). "Modular and Hierarchically Modular Organization of Brain Networks." *Frontiers in Neuroscience* 4:200.
- Newman, M. E. J. 2006. "Modularity and Community Structure in Networks." Proceedings of the National Academy of Sciences of the USA 103:8577–8696.
- O'Malley, M. (2007). "Exploratory Experimentation and Scientific Practice: Metagenomics and the Proteorhodopsin Case". *History and Philosophy of the Life Sciences* 29 (3):337 - 360.
- O'Malley, M. A. & Parke, E. C. (2018). "Microbes, mathematics, and models". *Studies in History and Philosophy of Science Part A* 72:1-10.
- Rathkopf, C. (2018). "Network representation and complex systems". *Synthese* (1).

Ratti, E. (2015). “Big Data Biology: Between Eliminative Inferences and Exploratory Experiments”. *Philosophy of Science* 82(2): 198-2018.

Ravasz E, Somera AL, Mongru DA, Oltvai ZN, Barabási AL. (2002). “Hierarchical organization of modularity in metabolic networks”. *Science*. 297:1551–1555.

Sporns, O., 2007. “Brain Connectivity.” *Scholarpedia* 2 (10): 4695.

Sporns, O. and Betzel, R. F. (2016). “Modular Brain Networks” *Annu Rev Psychol.* 67: 613-640.

Steinle, F. (1997). “Entering new fields: Exploratory uses of experimentation.” *Philosophy of Science* 64 (4):74.

Wagner, G. P., Pavlicev, M., and Cheverud, J.M. (2007). “The Road to Modularity.” *Nature Review Genetics* 8:921–31.

Waters, C. K. (2007). “The Nature and Context of Exploratory Experimentation: An Introduction to Three Case Studies of Exploratory Research.” *History and Philosophy of the Life Sciences* 29 (3):275 - 284.

Weisberg, M. (2013). *Simulation and Similarity: Using Models to Understand the World*. Oxford University Press.