# Title Page

**Author**: Daniel Kostić, PHD

**Title**: Minimal structure explanations, scientific understanding and explanatory depth.

**Affiliation**: Institut d'histoire et de philosophie des sciences et des techniques (IHPST/CNRS/Université Paris 1 Panthéon-Sorbonne)

**Address**: IHPST/CNRS/Université Paris 1 Panthéon-Sorbonne, 13 rue du Four, 75006 Paris, France.

**E-mail**: daniel.kostic@gmail.com

**Phone**: +33 (0) 768890295

**Acknowledgements**: I thank Kareem Khalifa and Philippe Huneman for their extremely helpful comments on this paper. I owe a great deal of gratitude to Isabelle Drouet, Francesca Merlin, Umut Baysan and Cyrille Imbert for their comments on the very first draft of this paper. I also thank the participants of the workshops "Unifying the debates: mathematical and non-causal explanations" and "Understanding from models" as well as Duško Prelević for their very insightful discussions of various versions of this paper. I also thank two anonymous reviewers for their very helpful comments.

**Funding**: This research is funded by the European Commission, Horizon 2020 Framework Programme, Excellent Science, Marie Skłodowska-Curie Actions under REA grant agreement n° 703662; project: Philosophical Foundations of Topological Explanations (Proposal acronym: TOPEX).

# Minimal structure explanations, scientific understanding and explanatory depth

## Abstract

In this paper, I outline a heuristic for thinking about the relation between explanation and understanding that can be used to capture various levels of "intimacy", between them. I argue that the level of complexity in the structure of explanation is inversely proportional to the level of intimacy between explanation and understanding, i.e. the more complexity the less intimacy. I further argue that the level of complexity in the structure of explanation also affects the explanatory depth in a similar way to intimacy between explanation and understanding, i.e. the less complexity the greater explanatory depth and vice versa.

## 1. Introduction

Many philosophers maintain that explanation is intimately tied to understanding, specifically many hold that the goal of scientific explanation is to provide understanding of physical phenomena or in general of nature (de Regt 2009; Hempel 1948; Strevens 2008, 2013). The views about the relation between explanation and understanding range from largely dismissive (Hempel 1948; Trout 2007) which see the scientific understanding as a pragmatic or psychological by-product of explanation which is not a proper subject of philosophical inquiry, and which should rather belong to psychology; to proposals to treat the understanding independently from the explanation i.e. that there could be understanding without explanation (Lipton 2009; Schurz and Lambert 1994; Newman 2013, 2015), or on the other hand, that there could not be understanding without explanation (Strevens 2008, 2013; Khalifa 2012, 2017).

In this paper, I outline a heuristic for thinking about the relation between explanation and understanding that can be used to capture various levels of "intimacy", so to speak, between them, i.e. by using this heuristic we will be able to explain away some of the seemingly paradoxical cases in which it is claimed we could have the understanding without explanation, as well as cases where there can't be understanding without explanation. The idea is that the level of complexity in the structure of explanation is inversely proportional to the level of intimacy between explanation and understanding, i.e. the more complexity the less intimacy, and *vice versa*. The structure of explanation is understood in this paper as a description of the relation between the *explanans* and *explanandum*, and the complexity in this context should be understood as the number of components that are required to describe this relation. In this sense, the complexity of particular instances of explanation could possibly be measured, probably by using something like the minimum description length principle (Barron and Cover 1991; Barron et al 1998; Grünwald 2007), but developing such measure is out of the scope of this paper, because the primary goal of this paper is to point out the dependencies between the structure of explanation, scientific understanding and explanatory depth. The idea about the complexity in the structure of explanation could equally well be conceived in terms of a particular theory of explanation (mechanistic, functional, topological, semantic), in which case the theory of explanation will determine the complexity in the structure of explanation. This allows that the idea about the complexity in the structure of explanation be generalized from instance of explanations to theories of explanation as well, i.e. some theories of explanation describe more and some describe less complex structures of explanation. I further argue that the level of complexity in the structure of explanation also affects the explanatory depth in a similar way to intimacy between explanation and understanding, i.e. the less complexity the greater explanatory depth and *vice versa*. A more precise way to specify what is meant by the "structure" of explanation is to say that the structure of explanation is a description D of the relation R between the *explanans* p and *explanandum* q, for example in the general account of scientific explanation such as the D-N model the R represents logical entailment, and p has some subparts such as antecedent conditions and general laws, and q represents a description of the *explanandum* phenomenon. The D in this case has a form of a logical argument. Whereas in the minimal structure explanations such as the topological explanation, in describing the graph-theoretical dependency relations between topological variables we are simultaneously describing the relation between the *explanans* and *explanandum*.

All explanations, regardless of their kind or type, have some structure. I understand the structure in terms of a description of the relation between *explanans* and *explanandum*. Sometimes, as in Hempel's general theory of explanation, the structure will have the form of a deductive argument, that describes how the description of an *explanandum* is logically derived from a set of premises or why it should be expected given the premises (Hempel and Oppenheim 1948). Sometimes, as in a particular type of explanation such as the interventionist account, the explanation has a structure of "explanatory generalization" that describes invariant counterfactual dependency relations between the values of variables (Woodward and Hitchcock 2003). It seems at least intuitively plausible to think that different structures of explanations engage with counterfactual knowledge in different ways. For example, in the argument structure the connection between the counterfactual knowledge and explanation will depend on the truth of the premises and the counterfactual knowledge will be obtained in terms of derivation of the *explanadum* from a variety of premises. The amount of counterfactual knowledge in that case will be commensurate with the range of possible true premises. In the Woodwardian case, the connection between the structure of explanation and counterfactual knowledge is cast in terms of the amount or range of W-questions (what-if-thing-had-been-different) one could ask about the dependency of the value of the *explanandum* variable from the value of the *explanans* variable. In this sense, it would seem plausible to assume that different structures of explanation connect in different ways to counterfactual knowledge and can affect the scope of counterfactual knowledge that it connects to. I show that in the case of topological explanation this connection is more direct in virtue of which it covers much wider range of counterfactual knowledge than the explanations in which the connection between *explanans* and *explanandum* is less direct (such as for example interventionist type of explanation).

However, it is very important to distinguish what is minimal and what is complex in this context. I argue that it is the structure of explanation that can be very complex or minimally complex, not the explanation itself. Also, there is an important difference between simple and minimal here, in the sense that an explanation can be simple, but have a very complex structure, e.g. any explanation that has a deductive-nomological (D-N) structure. On the other hand, an explanation can be very complicated, but have a minimal structure, e.g. a topological explanation.

On this view, there are degrees of complexity in the structure of explanation, and so there could be very complex explanations which require a great deal of mediating knowledge to grasp the exact relation between the *explanans* and *explanandum*. Explanations with more complex structure would be Hempel's general account of explanation or the D-N model[1] (Hempel and Oppenheim 1948), Woodward's interventionist account (Woodward and Hitchcock 2003), mechanistic explanation (Craver 2007; Kaplan and Craver 2011; Machamer et al 2000), semantic explanation (Chalmers and Jackson 2001). But there could be also explanations that require very little or none of mediation to grasp the exact relation between the *explanans* and *explanandum*. The latter ones I will call the minimal structure explanations. The best example of minimal structure explanation is topological explanation (Darrason 2018; Huneman 2010, 2015; Kostić 2018a,b; Rathkopf 2018), but perhaps there could be other too, e.g. minimal model explanations (Batterman and Rice 2014), some accounts of mathematical explanations in science (Batterman 2009; Lange 2012), structural explanations (Huneman 2018).

Even though, there are many different ways to think of understanding and its relation to the explanation and knowledge, none of them have explicitly treated the relation between the structure of explanation and understanding

[1] I do think that the topological explanations and in general minimal structure explanations do not conform to the Hempel's general theory of explanation, just like mechanistic ones don't fit it either. I also think that topological explanations are different from mechanistic ones in a number of significant ways as it will become more evident later in the paper.

specifically. To avoid circularity when using the terms “grasping” and “understanding” in referring to the structure of explanation, following Strevens (2008, 2013) and Khalifa (2017) I distinguish between “*understanding-that*” and “*understanding-why*”. *Understanding-that* refers to some basic cognitive abilities such as being a competent speaker of a language, knowing what certain mathematical relations mean, grasping the mathematical axioms and knowing what it means to say that they are logically primitive, or knowing that something is a fact. The *understanding-why* is really what we are after here, and it refers to knowledge of why something is the case, which is based on the knowledge of counterfactuals. For example, an explanation that has an argument structure, the *explanandum* is the conclusion in the logical argument that is derived from the set of premises that constitute the *explanans*. Famously, Hempel and Oppenheim (1948) represented it thusly in figure 1.

In this case, we are talking about *understanding-that* of each of these premises. Furthermore, there is also the understanding-that of the rules of inference, order of derivation, validity and soundness. Of course, soundness or validity alone are not guarantees of a successful scientific explanation, i.e. one could have a correct understanding from the false explanation, if one was only following the explanatory relations in the D-N model for example. But in minimal structure explanations, topological being of them, such situation is not possible because explanatory relations in the topological explanation do not depend on the contingent causal facts that are particular to any of the physical systems in question. That is why the minimality and the abstractness and generality that they entail are so important to emphasize.

But the *understanding-why* comes from knowledge of all these relations and it is also supported by counterfactual thinking, i.e. had something in the argument been different in certain ways, the conclusion and thus the *explanandum* would have been affected in certain ways. The purpose of the above example is to illustrate the point about the relation between the structure of explanation and counterfactual information. In this sense, the understanding is facilitated by the knowledge requirements for grasping the exact relation between the *explanans* and *explanandum*. This further means that the explanation requires only the knowledge facilitated by the *understanding-that,* whereas proper understanding requires the knowledge facilitated by the *understanding-why*. Another way to put it is that the *understanding-why* comes from the structure of explanation, and it has a form of counterfactual information about the dependency relations between the *explanans* and *explanandum*.

What makes some structure of explanation more complex, is not the amount of background assumptions, but the number of components that are required to describe the relation between *explanans* and *explanandum*.[2] In this sense, it

[2] In terms of measuring the complexity in the structure of explanation, one can also distinguish between different levels and kinds of components. For example, in the D-n

means the more components the more complex the structure of explanation, and *vice versa*. For example, in the D-N model of explanation (Hempel and Oppenheim 1948) besides the statements about antecedent conditions and general laws, there are several other components that play an important role in the derivation of the *explanandum*, these are: the rules of inference (modus ponens, modus tollens), order of derivation (what is derived from what), soundness and validity of the argument. This kind of description of explanatory relations allows that there could be false explanation that provides correct understanding of explanatory relations. For example, if we substitute the Phlogiston theory as a general law in the D-N model, we will still be able to understand various counterfactual dependencies that the model postulates, and thus to have a correct *understanding-why* despite having a false explanation.

The complex structure of explanation can be represented schematically in the following way:

**(CSE):** $Understanding_{\text{that}}$ (X,Y,Z,W) → $Understanding_{why}$

Where X,Y,Z and W in D-N model may represent antecedent conditions, general laws, validity, soundness, order of derivation, and some additional explanatory component respectively. In the mechanistic

model, the components such as antecedent conditions and general laws seem to be different both in kind and in level from components such as rules of inference, soundness and validity.

explanation, these components would be: elements, activities, organizational principles, constitution relations, manipulability relations, variables. And in the semantic explanation these components could be: concepts, primary and secondary intensions, possible worlds, various possible world semantics that determine how intensions behave in various possible worlds. Based on all these explanatory components we are able to derive the *explanandum* from the *explanans* and to grasp various counterfactual dependency relations, i.e. to *understand-why*.

Whereas in minimal structure explanation just by *understanding-that* of the mathematical dependencies that describe a topology (in the case of topological explanation), we are able to understand various counterfactual dependencies in the very same noetic act of grasping the description of topology, and thus to have almost unmediated *understanding-why*.

The schematic representation of the minimal structure explanations would then look like this:

*Minimal structure explanation*

(**MSE**): $Understanding_{\text{that}}\ (T) \rightarrow Understanding_{why}$

Where T is a description of mathematical dependencies in a certain topology.

To better understand the point about the relation between the structure of explanation and understanding, consider an example of two different questions about the dependency of wiring costs and evolution of the brain dynamics on the network topology. In neuroscience, this issue often comes up in terms of questions how the wiring costs drive the evolution of brain networks, and about how the network topology in the brain constrains the wiring costs and dynamics. The former question would require a mechanistic explanation, that takes into account very specific details of the system and describes various causal dependencies across various time-scales, between the network modules, cognitive functions and how these dynamical and functional features constrain the evolution of brain networks. For example, the relevant question in this context will be about how the actual network connections that facilitate low wiring costs will be preferred in the evolution of brain's network structures. This explanation, by its very focus on the particular system and on its causal history will be less abstract and general. On the other hand, in the latter case, when explaining the topological constraints on the wiring costs, the explanation will have to take into account only the dependency relations between connectivity patterns in the network and the particular connections. By the very nature of this question, the explanation will be more abstract and general, because such dependency relations hold independently from any particular system, simply because they are mathematical dependencies that are actually describing the network model. Such explanation will have far fewer components and the relation between the *explanans* and *explanandum* in it will be much less mediated, because the very same dependency relations that are doing the explanatory work are the ones that are also used to describe the system in question.

These are two very different ways to answer two seemingly similar questions. However, it seems very difficult to resist the intuition that mechanistic and topological explanations, are of different levels of abstractness, generality and complexity (remember in topological explanation, the structure is rather minimal, because the relation between the *explanans* and *explanandum* is less mediated), and because of that they provide two different scopes of understanding. In this sense of level of complexity in the structure of explanation, the mechanistic explanation would conform to the CSE scheme of the structure of explanation, whereas the topological one would conform to the MSE scheme.

The minimal structure explanations also support an account of explanatory depth, that can be applied to both causal and non-causal explanations. The explanatory depth in this context is thought of in terms of richness of counterfactual explanatory relations that the explanation provides, so in this sense, the explanations which provide fewer counterfactual explanatory relations are less deep than the ones that provide more counterfactual relations.

Depending on the complexity of the structure of explanation, the relation between explanation and understanding can be more intimate or less intimate, the more complex the structure of explanation the less intimate the relation between the explanation and understating, and *vice versa*. Because of the minimal structure and more direct relation between explanation and understanding, these explanations will be deeper, and more universal, because they will provide more counterfactual dependency relations for our (armchair) grasping.

Having set all the important distinctions in this section, in the next section I discuss the topological explanation, which has a minimal structure in exactly this sense.

## 2. Minimal structure of topological explanation and scientific understanding

In order to make my case, I discuss an example of scientific explanation that is pervasively used in biology (Levy and Bechtel 2013; Green et al 2016; Huneman 2010, 2015), medicine (Darrason 2018), complexity theory (Rathkopf 2018) and neuroscience (Craver 2016[3]; Kostić 2018 a,b), i.e. the topological explanation. The knowledge requirements to grasp the relationship between the *explanans* and *explanandum* in this context are so minimal that it seems that there is a sense of immediacy between mentally grasping or apprehending the descriptions of mathematical properties and

[3] It should be noted that Craver (2016) doesn't accept the account of topological explanation as other authors that are cited here do. He argues that since topological explanations don't provide a norm for distinguishing good from bad explanations, they can't be considered explanations at all. At best, they constitute a new way to describe mechanisms.

relations that characterise the topology of a system in question (or the *understanding-that)* and the property or behaviour in the system that we want to explain.

Based on these considerations I argue that topological explanation has a minimal structure, which can be formulated in the following way:

*If a physical system or some of its aspects can be described as a network (by using graph theory, network analysis, network control theory and similar approaches), then just grasping the mathematical dependencies between topological properties and the network description suffices for the explanation of the behaviour or some properties of that physical system.*

For example, to explain the efficiency of the signal processing in the brain, one will only have to understand the mathematical dependency between the clustering coefficient and average path length and the network. Similarly, in explaining the dynamics of the epidemics such as the speed of the spread of infection and what portion of the population will be affected, one will also have to understand the same dependencies between the clustering coefficient and the path lengths. The explanation itself does not depend on the details of these two very different systems. That's why understanding the topology suffices for understanding the behaviour or properties of that physical system, without having to appeal to particular details of any of these systems.

This definition allows to make a further claim, that in topological explanation the relation between the explanation and understanding is more direct, i.e. in topological explanation describing the topological properties of networks at the same time, unmediated by the propositional structure, provides understanding of a physical fact we want to explain. More precisely, in topological explanation, an understanding of a mathematical characterization of topology allows us to grasp various counterfactual dependency relations between the *explanandum* and the description of topology (*the explanans*).

To lay out this account properly, I'll first explain what the topological explanation is, by using a simple example of Watts and Strogatz (1998) small-world model, and then show that the same explanatory relations hold in even more complex cases, such as for example use of topological hierarchical modularity in explaining various properties of the brain.

The topological explanation has a structure of a counterfactual that describes a mathematical dependency between a set of topological properties and a network representation of a real-world system (Kostić 2018c). Topology in this sense, refers to a specific global pattern of connectivity in a network or a graph. A network is a collection of nodes and edges, that are connected in certain ways, and a graph is a mathematical description of such a network (van den Heuvel and Sporns 2013, p. 683). The description of network topology and topological properties are obtained by quantifying networks.

There are many ways to quantify networks and analyse their topologies. The best known are the node and network degrees. A node degree is a measure of number of connections a node maintains, whereas a network degree is the average number of connections that nodes in a network maintain. Another example would include the measures of path lengths (average number of edges that have to be traversed to reach one node from the other) and clustering coefficient (measure of tendency of nodes that are connected among themselves to form connected triangles of nodes that are connected among themselves and therefore create very densely interconnected groups of nodes, that are called cliques). The path length is therefore a global property of the network and the clustering coefficient is the local network property. These measures are used to characterise network topology of various systems, regardless of what nodes and edges represent in those systems. For example, a network that has a low value for path lengths (i.e. short path length) and high clustering coefficient, is the way to characterise a small-world topology. Mathematically speaking, the small-world topology enables nodes that are in distant cliques, to be reachable from any other node in the network through significantly fewer steps than in any other kind of topology, and in that way, shorten the distance between the neighbourhood of nodes and neighbourhoods of neighbourhoods. This mathematical feature of small-world topology affects (mathematically) the network communication, because whatever process or activity or a mechanism we want to drive through such network the small-world topology will determine or in general constrain its dynamics (Kaiser and Hilgetag 2004, p. 312; van den Heuvel and Sporns 2013, p. 683).

Famously, the Watts and Strogatz (1998) small-world model was used to show the functional significance of small-world topology for dynamical systems (Watts and Strogatz 1998, p. 441). They used the example of the small-world model in the spread of infectious disease. They looked into a simple rewiring procedure of a family of graphs, so that starting from a ring lattice which has *n* nodes and *k* edges per node, they rewired each edge randomly with a probability *p*. The procedure allowed them to probe the graph properties between completely regular (*p=0*) to completely random (*p=1*), as is shown in figure 2.

As mentioned above, they quantified the structural properties of the graphs by using the measures of average path length (L*p*) and clustering coefficient (C*p*). In doing so they have found that the properties of a regular graph at (*p=0*) are that of a large-world where L grows linearly with *n* (the number of nodes). On the other hand, in random networks at (*p=1*) which are poorly clustered, the *L* grows only logarithmically with *n*. The topology of a graph in the region between the regular and random graphs (where the wiring probability distribution is 0<*p*<1) has surprisingly low L and high C. These properties obtain due to introduction of few long-range connections or edges which then shorten the distance not only between the pairs of nodes that they connect, but also between the neighbourhood of nodes that are connected to that pair of nodes, and thus further shortens the distance between the neighbourhoods and also neighbourhoods or neighbourhoods. A very important point to keep in mind is that at the local level of a clustered neighbourhood of nodes, the change from a regular to small-world topology is not detectable because replacing a short-range edge from such a highly-clustered neighbourhood with a long-range one, leaves the value of C (clustering coefficient) practically unchanged, but the L(*p*) drops dramatically.

To test the functional significance of small-world topology for dynamical systems they used a simplified model for the spread of infectious disease. They started with the same structure of family of graphs, where an infected individual is introduced into a healthy population and after a period of sickness which lasts a unit of dimensionless time, the infective individual is removed either by immunity or by death. During sickness, each of these individuals can infect their neighbours with some probability *r*. On each time step, the disease spreads through the graph (through the edges) until it either infects the whole population or it dies out and, in the process, infects only a portion of the population. The results these tests have shown are that:

1) Critical infectiousness *r-half* rapidly decreases for small *p;* and
2) The time T(*p*) that is required to infect the entire population, regardless of its structure, has the same functional form as a characteristic path length L(*p*).

To Watts and Strogatz this clearly shows that:

"All the models indicate that network structure influences the speed and extent of disease transmission, but our model illuminates the dynamics as an explicit function of structure, rather than for a few particular topologies, such as random graphs, stars and chains." (Watts and Strogatz 1998, p. 442).

In this example, the knowledge that is required to understand the description of small-world topology, or in Watts and Strogatz's vocabulary, to understand the "structure" is the very same knowledge that is required to understand the dynamics. Spreading of the infection along the edges of the graph is described by the very descriptions of its topological properties, i.e. the critical infectiousness decreases with topological randomness and the time T(*p*) to infect the entire population has the same functional form as the path length L(*p*). This feature of topological explanation allows us to understand the dynamics without any additional propositional apparatus or mediation. The structure of explanation in this case is minimal, in a sense that to grasp or to *understand-that* of the topological description of the dynamics is to grasp the explanatory relevant counterfactual dependencies or to *understand-why*.

The reason why understanding the topology of the network suffices for the explanation of the feature in question of that system is that the topological explanation describes the (counterfactual) dependency relations between the topological properties and the network representation of the system. Once the system is described as a network, the network is quantified to obtain the topological properties. The counterfactual relations between topological properties and the network representation is what provides the immediacy between *explanans* and *explanandum*. The "bridge" between mathematics and the physical system is the fact that topological properties and the dependency relations between them are the properties and the dependencies that mathematically define that network model in the first place.

In this case, one might object that in connecting the mathematical *explanans* to a physical fact requires some kind of an empirical premise, and adding the empirical premise gives some kind of propositional structure, which further shows that the understanding is mediated and that the topological explanation is not minimal in a sense I presented here. On this view, grasping the explanatory relations that are posited in the explanation is what constitutes the understanding. Strevens calls this view a *simple account of understanding* (Strevens 2008, 2013). According to this view, there cannot be understanding without explanation, because it is the very structure of explanation (the structure here should be understood as a structure of propositions that describe causal relations) that provides the correct explanatory relations between the propositions and mentally grasping those relations is what constitutes a scientific understanding. For example, one can know that the Newton's second law of motion is true, but without grasping its content, i.e. grasping the exact explanatory relations that the structure of explanation provides, they will not be able to understand a phenomenon that is explained by that law. Simply put, the structure of explanation merely supplies the explanatory relations for our (mental) grasping. The grasping in his sense is a tacit form of knowledge, more like a direct apprehension. For a lack of better definition Strevens claims that the understanding or direct apprehension is:

"the fundamental relation between mind and world, in virtue of which the mind has whatever familiarity it does with the way world is." (Strevens 2013, p. 511).

A possible objection at this point could be that one could always "translate" an explanation from minimal structure to more complex one and have the same result. For example, Craver (2016) puts the topological explanation in argument structure. He puts it this way:

"**Empirical Premise**. Königsberg's bridges form a connected network with four nodes. Three nodes have three edges; one has five.

**Mathematical Premise**. Among connected networks composed of four nodes, only networks containing zero or two nodes with odd degree contain Eulerian paths.

**Conclusion**. There is no Eulerian path around the bridges of Königsberg." (Craver 2016: 700).

Indeed, one can always do this kind of translation, but such translated explanation would not only be superfluous, because there is already a simpler one (the topological one that conforms to the MSE scheme), but it would also compromise explanatory depth. As we will see, greater complexity compromises depth, but deeper explanations provide more understanding. More precisely, in cases where both topological and mechanistic explanations can be given, the particular *explananda* will dictate which kind of explanation to use, and when the *explanandum* requires appealing to some general or abstract features, the topological explanation will be the one to go with. Topological explanations exemplify how minimal structure explanations provide greater depth, by having more intimate relation between explanation and understanding.

The same can be done with the Watts and Strogatz example, but as we have seen, in their case, the dynamics of epidemics is precisely described by the measure of critical infectiousness which increases or decreases with topological randomness, and the time to infect the entire population has the same functional form as the topological measure of path length, which makes the explanation conform to the MSE scheme. In other words, the dynamics of an epidemics is described topologically, and grasping the description of topology suffices for the understanding of the dynamics.

The minimal structure of topological explanation and its relation to understanding provide an interesting insight into the more general issue of explanatory depth, which I discuss in the next section.

## 3. **Minimal structure and explanatory depth**

As it was shown in the previous section, topological explanation has a minimal and noetic structure, i.e. the relation between the *explanans* and *explanandum* is less mediated or more direct. The more direct relation here means that just mentally grasping a mathematical dependency between topological properties and a mathematical representation of a system provides the *understanding-why* of properties or behaviours of that system that we want to explain.

Some topologies are very complex, so that in explanations based on them would seem to require that grasping the relation between the *explanans* and *explanandum* would be more mediated than Watts and Strogatz cases.

One way to answer this objection is that regardless of how complex the topology is, *understanding-that* of such topology would still belong to the *explanans*, and it's not an additional knowledge that is required to properly grasp the relation between the *explanans* and *explanandum*. This answer is based on the distinction between the simple and minimal from the beginning of the paper. According to this distinction, an explanation could be simple and still have a very complex structure, but explanation could be very complicated (in terms of grasping or *understanding-that* of highly abstract mathematical relations to describe certain topologies) but still have a minimal structure, i.e. to *understand-that* of the topology is to *understand-why* of a range of counterfactual dependencies between topological properties and a network representation of a system.

The other way could be that this account maybe doesn't apply to all cases of topological explanations, indicating a pluralist view about topological explanation, according to which there could be different kinds of topological explanations. This could very likely be the case, given the richness and complexity of connectivity patterns in real world systems.

However, minimal structure explanations provide greater explanatory depth in virtue of being more abstract and general, and in that way, they provide more counterfactual dependency relations for our grasping. This seems plausible, because an explanation that is less abstract and general would seem to provide fewer counterfactual dependency relations than the explanation that is more abstract and general, in virtue of being applicable to fewer classes of phenomena and in virtue of being able to cover fewer W-questions (what-if-things-have-been-different-questions). Something that is more abstract and more general by its very definition is encompassing something else that is more concrete, specific, localized and particular, and in virtue of that it can provide more counterfactual dependency relations than something that is more concrete, particular and specific.

This should be an uncontroversial claim because it is compatible with three influential accounts of explanatory depth in terms of generality and abstractness. Given that the minimal structure explanations hold independently from any actual system, the explanatory depth that they provide is compatible with the Deductive-Nomological account of explanatory depth (Hempel and Oppenheim 1948) in terms of applicability to a range of possible systems. This account is also compatible with the interventionist account of explanatory depth (Hitchcock and Woodward 2003) in terms of a range of counterfactual answers to what-if things-had-been-different questions. I consider Strevens' (2008, 2013) simple account of understanding as a variety of interventionist account, and that it offers a similar conception of explanatory depth. Finally, minimal structure explanation account of explanatory depth is also compatible with Weslake's (2010) abstractive account in terms of providing various levels of abstraction in the explanation itself (micro-macroscopic levels).

An interesting consequence of this view would be that the more complex the structure of explanation the more it is susceptible to cases where we get understanding from completely false explanations (Ptolemaic explanation of planetary orbits, Caloric theory explanation of heat, astrological explanation of personality traits or events, and many other). Because in those cases the *understanding-that* of various elements and explanatory relations in the *explanans* has a lot to do with understanding the rules of inference, validity, soundness, or in mechanistic explanation, it has to do with mapping a model onto a mechanism, that could all stand in correct explanatory relations, but some ontic detail might be false, and the false explanation would still produce correct understanding. Recall the example from the beginning in which just by replacing a law of nature in the D-N model with a Phlogiston theory, we can get a correct understanding from false explanation.

But in explanations with minimal structure, it is difficult to see how one would get a correct understanding from false explanation, because for example, in topological explanation, understanding the mathematical relations that describe the topological properties (*understanding-that*) is to understand their counterfactual mathematical dependency relations and thus to *understand-why*. In other words, in topological explanation the correct explanatory relations do not hinge on contingent causal or ontic facts, but instead hold in virtue of inherently mathematical dependencies, and so having a proper grasp of various counterfactual dependency relations between the topological properties and the network representation (*understanding-why*) is actually a part of having a proper grasp (*understanding-that*) of a mathematical description of a topology.

This explanatory pattern and the relation between the *explanans* and *explanandum* in topological explanation will be the case even in very complex systems such as the brain. The brain has a small world topology, and the small-worldliness perfectly describes its topology at the global level as well at the local level of connected triangles of nodes. The intermediate level of brain's network organisation is best described through the network measures of community structure or network modularity (Meunier et al 2010, p. 2). The modules in a network that are also called communities, are subsets of nodes that are densely interconnected among themselves in the same module, but sparsely connected to the nodes in other modules (ibid). Since the nodes in the same module are very densely connected, the existence of connections between the nodes in different modules plays a role in shortening the path lengths in the network architecture, thus providing another way to characterize small world topology, i.e. the high clustering within a module and existence of links between nodes in different modules is what significantly shortens the path lengths in the whole network, and thus constitute the small-world topology. It should be noted that even though the modular networks are small-world topologies, not all small-world networks are also modular, e.g. the Watts and Strogatz (1998) model is small-world but not modular (Meunier et al 2010, p. 2). In many natural systems, the brain being one of them, each module can be partitioned even further into sub-modules, so the brain has a hierarchical modularity which is approximately invariant over a number of levels in the hierarchy (ibid). Hierarchical modularity has a tremendous explanatory potential, especially when it comes to explaining dynamics, information processing at multiple scales, system's evolvability and stability. For example, small-worldliness that is rooted in hierarchically modular topology in the brain will be advantageous for the locally segregated processing in highly specialized functions (e.g. in visual motion detection) because the high clustering within the module will enable low wiring costs, and at the same time in such topology the short path lengths will more easily facilitate globally integrated processing of some of the more generic functions (e.g. working memory). Furthermore, hierarchical modular topology will be conducive for high dynamical complexity because it allows that both segregated and integrated activities co-exist in the system, or because such topology allows that both synchronisation and de-synchronisation coexist across the network. Topological modularity will also allow that marginally stable networks of submodules be combined or divided while at the same time preserve network stability at the global level.

In all of these cases, the explanatory pattern is the same. By grasping the mathematical dependency between a global pattern of hierarchical topological modularity and a network representation of the brain, we are immediately able to know (without any kind of mediation through propositional structure) that small-world topology that stems from network modularity at the same time enables low wiring cost (through high clustering within a module) and thus it is favourable for the locally segregated functions, while also supporting integrated processing through short path lengths when it comes to more generic functions such as working memory. In Strevens' terminology, a description of topology immediately provides explanatory relations for our grasping.

The hierarchical modular topology is a way to describe the small-world topology in greater detail than the original Watts and Strogatz model (1998), and the dependencies and constraints between topological properties and the dynamical features in the brain, provide incomparably richer patterns of counterfactual dependencies than the description of small-world topology based only on the path lengths and clustering coefficient. However, despite this, the relation between the topology and dynamics remains equally unmediated as in the Watts and Strogatz case, i.e. *understanding-that* of the hierarchical modular topology suffices for *understanding-why* of the dynamics. As we recall, the modules enable greater signal processing efficiency locally, and the connections between different modules in the whole network as well as within the hierarchy of submodules enables global integration of functions across the network, but also synchronisation and desynchronization activities across the whole network. An explanation of the various dynamical features of the brain in this context will only appeal to various mathematical dependencies between topological variables (such as node and network degree, path length, betweeness centrality) in order to provide a very rich set of counterfactual dependencies and thus to provide the greater understanding.

After having discussed the account of explanatory depth that the minimal structure explanations provide, in the last section I will summarise my argument.

## 4. Conclusion

As we have seen, in topological explanation the *understanding-that* of certain mathematical relations that describe topology in a system requires minimal mediation or no mediation through any kind of propositional structure at all, in order to obtain the *understanding-why*, e.g. mentally grasping various mathematical relations that describe network modularity allows to grasp the function integration across different scales in the very same noetic act. To that effect, topological explanation has a minimal structure. It is based on understanding mathematical descriptions of topologies.

This account of explanation covers various levels of intimacy between the explanation and understanding, from the ones in which explanation and understanding are the most distinct (meaning delivering understanding requires a great deal more of mediating knowledge), e.g. Strevens (2008, 2013) and Khalifa (2012, 2017); to cases where the explanation has a minimal structure and the delivery of understanding is less mediated.

This is a gradual view of explanation, according to which the less of the structure it has the more of the understanding it provides, and *vice versa*. This has to do with the explanatory depth, because the explanations with minimal structure provide greater explanatory depth in virtue of being more general, more abstract, and both of these features stem from the abstractness and generality of various topologies. The explanatory depth in this sense is based on the fact that minimal structure of topological explanation provides more possible explanatory relations, it casts the counterfactual net much wider so to speak, which also increases applicability of explanation across very diverse domains of cases. It provides answers to more counterfactual questions, and finally, it provides answers about counterfactual dependencies across multiple scales and levels of abstraction and organisation.

**References**:

- Barron, Andrew R., and Thomas M. Cover. 1991. "Minimum complexity density estimation." *IEEE transactions on information theory* 37, no. 4: 1034-1054.
- Barron, Andrew, Jorma Rissanen, and Bin Yu. 1998. "The minimum description length principle in coding and modeling." *IEEE Transactions on Information Theory* 44, no. 6: 2743-2760.
- Batterman, Robert W. 2009. "On the explanatory role of mathematics in empirical science." *The British Journal for the Philosophy of Science* 61, no. 1: 1-25.

- Batterman, Robert W., and Collin C. Rice. 2014. "Minimal model explanations." *Philosophy of Science* 81, no. 3: 349-376.
- Chalmers, David J., and Frank Jackson. 2001. "Conceptual analysis and reductive explanation." *The Philosophical Review* 110, no. 3: 315-360.
- Craver, Carl F. 2016. "The explanatory power of network models." *Philosophy of Science* 83, no. 5: 698-709.
- Craver, Carl F. 2007. *Explaining the brain: Mechanisms and the mosaic unity of neuroscience*. Oxford University Press.
- Darrason, Marie. 2018. "Mechanistic and topological explanations in medicine: the case of medical genetics and network medicine." *Synthese* 195, no. 1: 147-173.
- De Regt, Henk W. 2009. "The epistemic value of understanding." *Philosophy of Science* 76, no. 5: 585-597.
- Green, Sara, Maria Şerban, Raphael Scholl, Nicholaos Jones, Ingo Brigandt, and William Bechtel. 2018. "Network analyses in systems biology: new strategies for dealing with biological complexity." *Synthese* 195, no. 4: 1751-1777.
- Grünwald, Peter D. 2007. *The minimum description length principle*. MIT press.
- Hempel, Carl G., and Paul Oppenheim. 1948. "Studies in the Logic of Explanation." *Philosophy of science* 15, no. 2: 135-175.

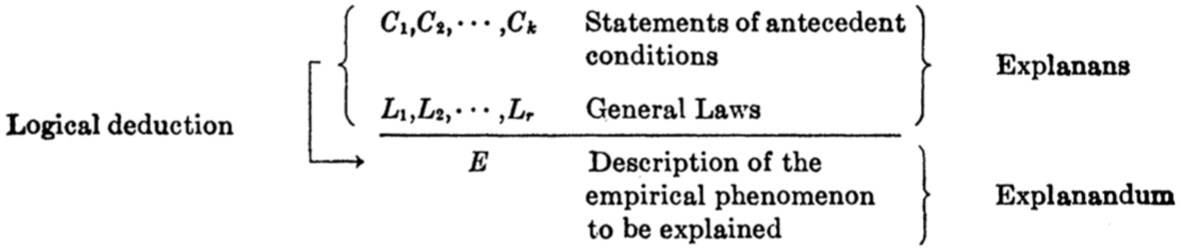
- Hitchcock, Christopher, and James Woodward. 2003. "Explanatory generalizations, part II: Plumbing explanatory depth." *Noûs* 37, no. 2: 181-199.
- Huneman, Philippe. 2018. "Outlines of a theory of structural explanations." *Philosophical Studies* 175, no. 3: 665-702.
- Huneman, Philippe. 2015. "Diversifying the picture of explanations in biological sciences: ways of combining topology with mechanisms." *Synthese*: 195: 115.
- Huneman, Philippe. 2010. "Topological explanations and robustness in biological sciences." *Synthese* 177, no. 2: 213-245.
- Kaiser, Marcus, and Claus C. Hilgetag. 2004. "Edge vulnerability in neural and metabolic networks." *Biological cybernetics* 90, no. 5: 311-317.
- Kaplan, David Michael, and Carl F. Craver. 2011. "The explanatory force of dynamical and mathematical models in neuroscience: A mechanistic perspective." *Philosophy of science* 78, no. 4: 601-627.
- Khalifa, Kareem. 2017. *Understanding, explanation, and scientific knowledge*. Cambridge University Press.
- Khalifa, Kareem. 2012. "The role of explanation in understanding." *The British Journal for the Philosophy of Science* 64, no. 1: 161-187.

- Kostić, Daniel. 2018a. "The topological realization." *Synthese* 195, no. 1: 79-98.
- Kostić, Daniel. 2018b."Mechanistic and topological explanations: an introduction." *Synthese* 195, no. 1: 1-10.
- Kostić, Daniel. 2018c. "“The “horizontal” and “vertical” ways of describing counterfactual dependency relations: a pluralist view about topological explanations”. *Manuscript.*
- Lange, Marc. 2012. "What makes a scientific explanation distinctively mathematical?." *The British Journal for the Philosophy of Science* 64, no. 3: 485-511.
- Levy, Arnon, and William Bechtel. 2013. "Abstraction and the organization of mechanisms." *Philosophy of science* 80, no. 2: 241-261.
- Lipton, Peter. 2009. "Understanding without explanation." *Scientific understanding: Philosophical perspectives*: 43-63.
- Machamer, Peter, Lindley Darden, and Carl F. Craver. 2000. "Thinking about mechanisms." *Philosophy of science* 67, no. 1: 1-25.
- Meunier, David, Renaud Lambiotte, and Edward T. Bullmore. 2010. "Modular and hierarchically modular organization of brain networks." *Frontiers in neuroscience* 4: 200.
- Newman, Mark P. 2015. "Theoretical Understanding in Science." *The British Journal for the Philosophy of Science* 68, no. 2: 571-595.

- Newman, Mark. 2013. "Refining the inferential model of scientific understanding." *International studies in the philosophy of science* 27, no. 2: 173-197.
- Rathkopf, Charles. 2018. "Network representation and complex systems." *Synthese* 195, no. 1: 55-78.
- Schurz, Gerhard, and Karel Lambert. 1994. "Outline of a theory of scientific understanding." *Synthese* 101, no. 1: 65-120.
- Strevens, Michael. 2013. "No understanding without explanation." *Studies in history and philosophy of science Part A* 44, no. 3: 510-515.
- Strevens, Michael. 2008. *Depth: An account of scientific explanation*. Harvard University Press.
- Trout, J.D., 2007. "The psychology of scientific explanation". *Philosophy Compass*, *2*(3), pp.564-591.
- van den Heuvel, Martijn P., and Olaf Sporns. 2013. "Network hubs in the human brain." *Trends in cognitive sciences* 17, no. 12: 683-696.
- Watts, Duncan J., and Steven H. Strogatz. 1998. "Collective dynamics of 'small-world'networks." *nature* 393, no. 6684: 440.
- Weslake, Brad. 2010. "Explanatory depth." *Philosophy of Science*77, no. 2: 273-294.
- Woodward, James, and Christopher Hitchcock. 2003. "Explanatory generalizations, part I: A counterfactual account." *Noûs* 37, no. 1: 1-24.

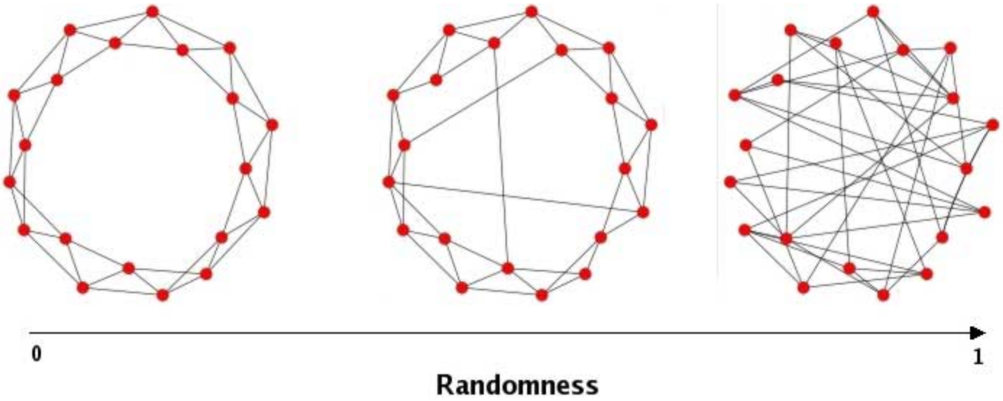
**Figures:**

- *Figure 1.*



(Hempel and Oppenheim 1948: 138).

- *Figure 2.*


0
1
Randomness