-Preprint-

On Networks, Trees and Traits

Evaluating Explanatory Power of Network Models in Cultural Evolutionary Research

Karim Baraghith[1]

Abstract[2]

In this paper, I assess the explanatory power of four graph- and network-based models recently used in the domain of Cultural Evolutionary Theory (CET): (a) Phylogenetic Trees, (b) Death–Birth Graphs, (c) Interaction Graphs, and (d) Trait Networks. First, I show that Interaction Graphs, Trait Networks and (empirically-calibrated) Death–Birth Graphs can function as genuine mechanistic explanations, while Phylogenies remain largely at the level of correlational evidence. I then broaden the analysis to two non-mechanistic accounts: Kitcher's unificatory perspective and Kostić's counterfactual theory of topological explanation. Interaction graphs excel mechanistically and topologically, death–birth graphs yield strong topological explanations, phylogenetic trees lead in unification, and trait networks offer mixed, moderate strengths. Because no single model or mode dominates, explanatory strength in CET is to a large extent question relative. I argue for a pluralistic approach in CET: rather than competing, these three explanatory axes ("modes of explanation") complement one another and suggest a roadmap for future hybrid network models that could combine these complementary virtues and invite closer collaboration between CET modellers and philosophers of science.

Keywords: cultural evolution, network, explanation, topological, causal, unification

1 Network Models in Cultural Evolution: a pluralist introduction

Cultural Evolutionary Theory (CET) attempts to understand the dynamics and diversity of culture from an evolutionary perspective. Small-scale transmission of cultural variants results in complex population structures, represented by trees and dynamic networks of cultural information. Cultural change can be regarded as a process resembling genetic change. This analogy

---

[1] Leibniz University Hannover (Germany), Institute of Philosophy, karim.baraghith@philos.uni-hannover.de

[2] This article will shortly appear in the *Synthese* Topical Collection "New Work on Network", edited by Daniel Kostić & James Woodward.

gave rise to a wide range of conceptual and methodological tools, often adapted from evolutionary biology. Pioneers in the field were authors like Boyd & Richerson (1988) and Cavalli-Sforza & Feldman (1981), but ever since they laid the foundations of CET, much work has been added to the framework – last but not least from philosophy of science, see e.g. Mesoudi (2011), Lewens (2015; 2020), Sterelny (2017), Boon et al. (2021), and others.

Recent years saw an increase in the use of *graph- and network models* within the interdisciplinary research programme of CET. Cultural traits rarely travel in single, well-defined lineages, instead they branch, fuse, hitch-hike, and recombine across social communities. However, the CET-network approaches available on the market make the impression of one big patchwork, rather than a unified framework. In the study of biological evolution, evolutionary graph theory rose in popularity since the works of Nowak (2006) or Doolittle (2009), but it is not clear which of its models can fully or partly be applied to CET. I think that four families of models now dominate the CET literature:

1. Phylogenetic Trees (e.g. Evans et al. [2021]), which reconstruct historical descent among languages, tool traditions, or ritual practices.
2. Death–Birth Graphs (e.g. Smolla & Akçay [2019]), which simulate coping and turnover in structured populations.
3. Interaction Graphs (e.g. Marjieh et al. [2025]), which record who interacts with whom in real time and with what frequency.
4. Trait Networks (e.g. Janson et al. [2021]), which map compatibility or incompatibility relations among the cultural variants themselves.

Each type of model carries an (often implicit) promise of explanation, yet the promises differ. A tree aims to explain a given distribution of traits by revealing common ancestry. A dynamical Death–Birth simulation explains by showing how local coping-relations between agents aggregate into long-run population level outcomes. An Interaction-Graph explains by pinpointing the detailed causal pathways along which information flows in network and where communities can be distinguished from each other in such a network. Finally, a trait network explains by exposing the internal logic of cultural systems that makes some variants co-occur. Across CET, recurring explanatory patterns can be grouped into several canonical[3] research questions (see Mesoudi 2011), for example:

Q1. Why do historical lineages branch as they do, and when is apparent similarity due to reticulation (horizontal transfer/borrowing across lineages) rather than descent (vertical transmission along lineages)?

---

[3] I call these questions "canonical" in a modest sense: they recur (often implicitly) across the whole CET network literature and are broad enough to cover the dominant explanatory uses of the four model families discussed here

Q2. Why does a trait fix (or fail to fix) with a particular probability and timescale under a given population structure?

Q3. Why do diffusion paths take the routes and speeds we observe, e.g. who influences whom, when, and how strongly?

Q4. Why do similar macro-patterns recur across domains, and could a common modelling schema account for different families of phenomena?

Q5. Why do specific structural invariants, i.e. network features (e.g., degree heterogeneity, modularity) robustly shape evolutionary outcomes?

I think that these questions *do not invite a single style of answer*. Some (Q2, Q3) target causal-mechanical understanding: identifying components, activities, and organizational relations that are difference-making for adoption, fixation, and flow of cultural traits. Others (Q4) call for unificatory payoffs: showing how a common modelling schema, with reusable variables and update rules, subsumes diverse cases. Still others (Q1, Q5) hinge on topological considerations, where structural properties—branching vs. reticulation, constraint networks among traits, or invariants like modularity—constrain what outcomes are reachable at all.

What aspects and features of cultural evolution do which particular network models explain, and how well do they do that? Philosophers of science have long debated what counts as a "scientific explanation". Relevant for us are:[4]

a. Causal-mechanical accounts (Craver 2016) demand an organised set of parts and activities that produce the phenomenon.
b. Unification accounts (Kitcher 1989) emphasize theoretical economy: one formal pattern explaining many cases.
c. Topological or structural accounts (Huneman 2018; Kostić 2020) contend that a phenomenon can be explained by the structural properties[5] of a network's shape, independent of mechanism.

The approach I suggest in this paper is deliberately pluralist: I map question types to model families and assess their explanatory contributions in three complementary explanatory modes,

---

[4] I center on causal-mechanical, unificatory, and topological explanations because each comes with clear, operational criteria that can be applied directly to CET network models: (i) organized components/activities/relations that are difference-making and testable via admissible interventions (causal-mechanical), (ii) schema reuse that confers explanatory economy across cases (unificatory), and (iii) structural invariants of networks that constrain reachable outcomes (topological). Other accounts of explanation are either folded into this triad or unsuitable for uniform scoring here. In particular, interventionism is treated as the diagnostic within causal-mechanical explanation rather than as a separate mode; dynamical and computational approaches specify how systems evolve or are computed, but presently lack domain-general, necessary-and-sufficient evaluative conditions comparable to the three modes above; and mathematical/optimality/universality explanations were developed for different targets and do not map cleanly onto the heterogeneous network cases I evaluate. The focus on these three therefore reflects operational tractability and fit to CET's network repertoire, not a denial of the value of the other approaches.

[5] I use "structural" in a neutral, practice-oriented sense; nothing in what follows presupposes a strong metaphysical thesis about the ontological status of mathematical entities.

without presuming a single “best” model across all questions – or a single “best” theory of (scientific) explanation. No single explanatory mode captures all the virtues of all network models, yet each mode has clear normative standards that a model either meets, partly meets, or fails to meet. *My core aim is to provide a transparent yard-stick for judging how well each of four CET network models explains what, having a certain Why-question in mind, and show that no single philosophical account captures all their strengths.*

To achieve this, **Section 2** anchors each network type in the causal-mechanical (“mechanistic”) sense of Craver (2016) and Bechtel (2020). I ask: Does the model provide a structural-connectivity map of real causal interactions? Does it contain causal motifs, i.e. recurrent subgraphs that act as “mechanism sketches” (sensu Piccini & Craver [2011])? Does it support Bechtelian hierarchical decomposition, linking lower-level coping or interaction events to higher-level cultural phenomena?

Note that a model with, for example, a modest mechanistic profile may still deliver strong unificatory or topological payoffs, since all profiles are question-relative. Thus, **section 3** introduces two non-mechanistic modes of explanation. I argue that “unificatory” explanations also matter because CET borrows tree and population-genetic formalisms wholesale from biology, promising cross-domain (as well as CET in-domain) economy. Likewise, “topological explanations” matter because the very promise of network science is that topology constrains behaviour, and Kostić’s (2020) counterfactual criterion makes that promise testable. Both, unificatory and topological explanations are not reducible to causal-mechanical explanations.

Section 4 presents the operational rubric. For each of the three modes and four network types I define and justify High/Moderate/Poor/None scores, flagging the intrinsic affordances and supplement load for each explanatory mode. A model scores “High” in a specific mode when its basic structure already fulfils that mode’s criteria; “Moderate” when some auxiliary assumptions suffice; “Poor” when extensive supplementation would be needed; and “None” when the mode is simply irrelevant. I connect my rubric to the rationale of Woodward (2025).

Section 5 synthesises the results into a pluralist thesis: explanatory power of the four CET network models is question-relative. If a researcher wants manipulable levers to change cultural outcomes, mechanistic strength is decisive. If she seeks theoretical economy across domains, unificatory strength dominates. Finally, if she aims to show why certain outcomes are inevitable given network structure, topological strength comes in. I outline a research agenda for possible hybrid network models that could integrate these virtues, inviting collaboration between CET modellers and philosophers of science.

2 Four Prominent Network Approaches in Cultural Evolutionary Research: a causal-mechanical analysis

In general, “culture” has a social structure (sensu Blau [1989]), i.e. traits bear structured relationships that dynamically influence transmission and selection, shaped by individual interactions and higher-level “agents” like institutions or firms. Graphs and networks (where nodes typically represent agents or traits and edges their interactions or relations) offer a natural way to capture this evolving structure. Different network types yield distinct long-run patterns of innovation diffusion, convention formation, and population change, raising the question: *Which of these network models truly explain cultural phenomena, and which merely describe them?* To answer this question, however, we must clarify what “truly explain” means, and the answer is pluralistic.

For now, we ground our analysis of the four network types (Phylogenetic Trees, Death–Birth Graphs, Interaction Graphs, and Trait Networks) in Carl Craver’s (2016) theory of how network models explain phenomena and in Bechtel’s (2020) notion of “higher and lower” mechanisms. Both authors can be situated in the “mechanist” (or causal-mechanical) tradition in philosophy of science, i.e. a mode of explanation, which treats explanations as accounts of how organized parts and activities produce a phenomenon.

We begin with the causal-mechanical account because it provides a widely used, network-friendly baseline—components, activities, and organization as difference-making structures—against which unificatory and topological virtues can be assessed without implying a single “best” model. Although I begin with causal-mechanical explanation, this reflects its role as a familiar entry point in discussions on scientific explanations—not any priority claim; the three explanatory modes are complementary and non-ordered, and the assessment is ques-tion-relative throughout. It functions as the baseline vocabulary and diagnostic, and is the most widely used explanatory framework in adjacent sciences (cognitive neuroscience, systems biology). Craver’s (2016) articulation specifically targets network models (components, activities, and organization as difference-making structures) making it a natural entry point for CET network analyses. The causal-mechanical provides a “lower bound” on explanatory standing (can the model identify components/activities/organization that are difference-making?). The later sections then evaluate complementary virtues, i.e. unification (schema reuse across domains) and topology (structural invariants constraining outcomes), that are not reducible to the mechanistic checklist.

Craver (2016) argues that whether, and how, network models explain depends not on features of graph theory or network analysis per se but on the explanandum under consideration, i.e. how the model is applied to a concrete system, and which kinds of relations (causal, mathematical, correlational) are treated as explanatory. He then distinguishes three distinct explanatory roles that any graph representation can play:

1. *Structural Connectivity*: When a network's nodes and edges accurately map the real causal pathways of a system, the model locates which components and interactions are difference-makers for the observed outcome.
2. *Causal-Motif Connectivity*: Certain subgraph patterns (or motifs) act as mechanism prototypes, illustrating how repeated configurations of parts and relations systematically generate the explanandum.
3. *Functional Connectivity*: Graphs that capture statistical or correlational patterns among elements serve as evidence for causal hypotheses but, without further causal interpretation, do not by themselves reveal how or why an effect occurs.

According to Craver, only those models fulfilling 1. and/or 2. can genuinely explain; models confined to 3. remain *at the level of description* or merely support mechanistic claims. It is not surprising that Craver (being a strong proponent of the causal-mechanical theory of scientific explanation) concludes that network analysis in general advances scientific investigation (by uncovering modules, hubs, correlational patterns), but:

> "[…] *does not seem to fundamentally alter the norms of explanation. The problem of directionality and the puzzle of correlational networks signal that, at least in many cases, the explanatory power of network models derives from their ability to represent how phenomena are situated,* […]*, in the causal and constitutive structures of our complex world."* cf. Craver 2016: 707.

For him, explanatory power of networks arises when network representations are embedded within causal-mechanical frameworks or constrained by ontic commitments that distinguish mere description from genuine explanation.[6]

Additionally, Bechtel (2020) articulates how mechanisms consist of parts (at a "lower" level) whose operations together yield a phenomenon (a "higher" level mechanism). Importantly, these parts can themselves be decomposed into sub-mechanisms, yielding a hierarchy of levels, each defined only relative to the mechanism in question.

2.1. Phylogenetic Trees

A phylogenetic tree represents each cultural lineage as a node and each directed edge as a hypothesised descent relation, and its specific target feature is the historical branching order and timing of divergence events among those lineages. Just like biological organisms or species, cultural artifacts, languages and techniques stand in genealogical relationships. To trace such

---

[6] In contrast to Craver (2016), who demands constitutive part–whole mapping for explanatory status, Woodward (2025) requires only interventionist dependence, tolerating macro-level or abstract variables so long as they are difference-makers. Both reject purely correlational "functional connectivity" graphs as explanatory *per se* and insist on specifying what relations count as difference-making for the explanandum.

historical pathways of cultural transmission, CET researchers frequently apply so called *phylogenetic methods*, which originate in the life sciences (palaeontology or paleogenetics). They aim to reconstruct the possible evolutionary history of recent forms, traits and cultural patterns, as Figure 1 depicts. For a recent overview, see Evans et al. (2021).

Phylogenetic trees adapt methods from biological systematics to reconstruct how cultural traits branch over time, depicting a uni-directional flow of information through vertical descent. In CET, researchers often employ Bayesian inference with Markov Chain Monte Carlo to sample from the vast space of possible trees, using metrics such as the Consistency Index (CI) and Retention Index (RI) to assess homoplasy and the fit between data and topology.[7] They further incorporate models like Pagel's DISCRETE to estimate trait-gain and loss as a Markov process, calibrating divergence dates by combining linguistic evidence with archaeological data (see e.g.

---

[7] *The consistency index* (*CI*) quantifies the level of homoplasies (analogies, i.e. convergent evolution or diffusion) in a tree. It is given by $CI = \frac{m}{o}$, where *m* denotes the minimum number of changes required to get from the pool of ancestral traits to the explanandum, which consists in the recent distribution and *o* is the observed number of changes. A *CI* close to 1 indicates minimal homoplasy. It is to be preferred against a tree with a lower *CI* as an explanation according to the principle of parsimony. *The retention index* (*RI*) on the other hand measures how well the tree retains the structure of the data while accounting for reversals and parallel changes, given by $RI = \frac{maxS-oS}{maxS-minS}$, where *maxS* stands for the maximum possible steps to get from the ancestral to the recent trait distribution, *oS* is the number of observed steps and *minS* the number of minimal steps required. Higher *RI* values suggest that the tree preserves much of the original data structure, making it a useful measure despite homoplasy.

Mace & Holden 2005). There are extensions to the classical “Tree framework”, deepening and introducing tools for statistical analysis, see e.g. Bortolussi et. al. (2006).

Despite concerns, see e.g. Maynard Smith (1986), that cultural borrowing (i.e. horizontal transmission) violates the strictly branching structure, many scholars have shown that phylogenetic methods remain robust in their application to sociocultural systems. Biological systems themselves exhibit reticulation (e.g. bacterial gene transfer), and advances like “partially reticulated” tree models accommodate both splits and occasional cross-links (Gray et al. 2007; Evans et al. 2021). In practice, CET phylogenies successfully model long-term vertical dynamics - such as the spread of pastoralism among Bantu languages (Mace & Holden 2005) - while quantifying the impact of horizontal exchange on tree accuracy and inferred adaptive trajectories.

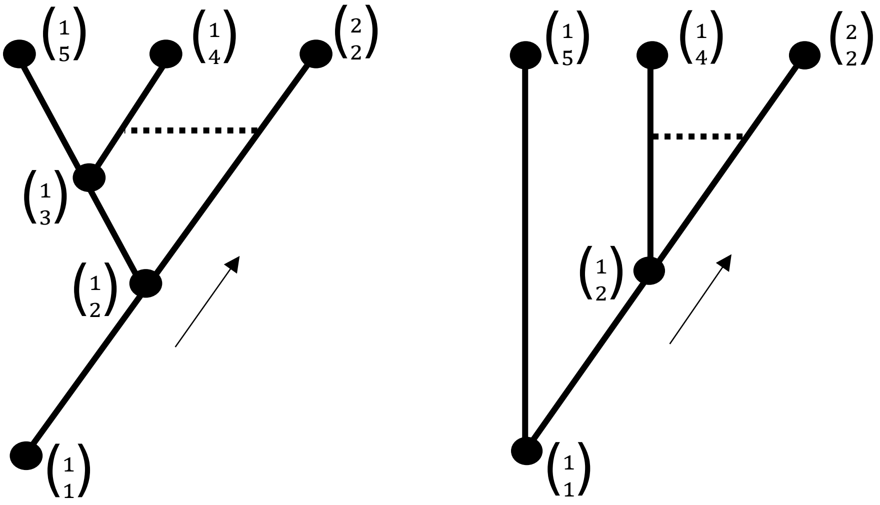

Figure 1: Two hypothetical phylogenetic trees illustrate two competing theories about the ancestral relationships among three recent individuals, each characterized by two traits $\binom{x}{y}$. Solid lines indicate vertical transmission, while dashed lines represent horizontal transmission. Mutation events occur at each branching point (solid dots), where one trait transforms into another. The tree on the right requires fewer mutation events compared to the tree on the left, suggesting a more parsimonious evolutionary path.

Within the mechanistic framework however, phylogenetic trees occupy no strong explanatory role. On Craver’s (2016) taxonomy, I think they are not structural-connectivity maps - because their branches depict hypothesised historical ancestry rather than contemporary causal interactions. They could be seen as causal-motif schemas, but in a very weak sense, as they contain no recurrent, difference-making subgraphs beyond generic bifurcations. Mechanistic accounts explain a phenomenon by showing how currently organised parts and activities produce it. They therefore function chiefly as functional connectivity devices: they organise correlational evidence for common descent that must be supplemented by an explicit transmission model before a full causal explanation emerges.[8]

Bechtel's (2020) notion of hierarchical mechanisms reinforces this verdict. The ancestor–descendant links in a tree are temporal successions, not part–whole compositions; a lineage is not made of its branches in the mechanistic sense but merely follows them in time. Consequently, phylogenies remain, from a mechanistic standpoint, poor causal representations.

2.2. Death-Birth Graphs

A death–birth graph represents each agent as a node and each directed edge as a potential coping route, and its specific target feature is the population-level fixation or loss of traits as shaped by local replacement dynamics (see Nowak 2006 or Smolla & Akçay 2019). It models (cultural) microevolution by treating individuals as nodes on a network and transmission as a "death–birth" update: at each timestep, one node is removed and replaced by an offspring of a randomly selected neighbour, with edge weights w_{ij} giving the probability that node *i* copies node *j,* see Figure 2. Smolla & Akçay (2019) extend earlier work (Cantor & Whitehead 2013; Allen & Nowak 2014) by allowing networks to rewire dynamically (offspring either inherit or randomly form social ties) and by comparing generalist ecosystems (high fitness = large repertoire size) to specialist ones (high fitness = peak proficiency). Their simulations show that dense networks foster high proficiency at the cost of diversity, whereas sparse networks produce the opposite, illustrating how individual learning dynamics aggregate into population-level cultural

---

[8] Interestingly, the so-called "Causal Bayes Nets" approach, see e.g. Pearl (2000), bears some structural resemblance to "Tree like" approaches, but a key difference between the two lies in explanatory focus. Both frameworks impose directed, acyclic graph structures, ensuring that information flows only in one direction: from parent to descendant in trees, and from cause to effect in Bayesian networks (where even the terms of causal "parent" or "descendant" are being used). Phylogenetic models estimate likelihoods of descent relationships, while Bayesian networks quantify conditional probabilities of causal effects. This distinction clarifies why phylogenetic trees are not causal models, even though they superficially resemble Bayesian networks. Causal Bayes Nets have explicitly suggested in the literature to represent mechanisms, see Gebharter (2014). Phylogenetic trees operate at a macro-level, explaining historical lineage relationships rather than causal mechanisms and describe which traits descend from which ancestors but do not specify how traits are transmitted or why they succeed. The focus is on historical reconstruction rather than actual process.

structures. The population dynamics are defined by the likelihood proportional to the edge weight, which determines whether an offspring node copies an adjacent node.[9]

When mapped onto Craver's (2016) taxonomy, I think that Death–Birth graphs serve as structural connectivity models, since their edges represent actual difference-making transmission routes (specific agents either coping or not coping other agents), and they can also play a causal-motif role by isolating simple replacement loops or fixation pathways as mechanism schemas that explain trait persistence or loss. They go beyond mere functional connectivity by pinpointing the "causal skeleton" of cultural change through coping.

Death–Birth graphs are dynamical networks. They start with individual agents and coping edges (lower-level parts). Aggregating update dynamics over time yields fixation or diversity patterns at the population level: a genuine higher-level phenomenon produced by organized lower-level operations. Because the same update rule applies recursively inside any sub-population, these

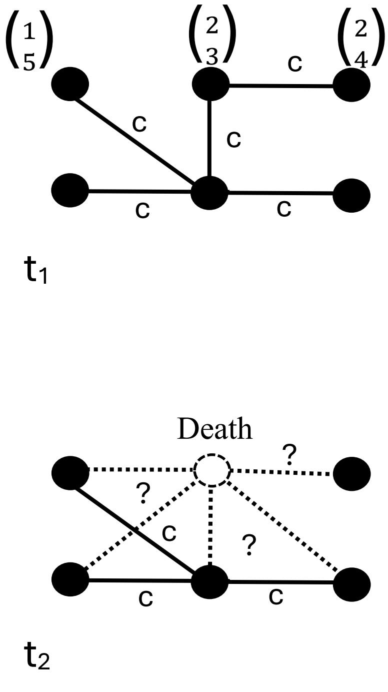

Figure 2: Two stages (t_1/t_2) of a dynamic Death-Birth graph, consisting of six individuals. At t_1 the nodes randomly copy each other (c). At t_2, one node (dashed-white) is chosen to "die" and is replaced by a new one ("birth"). The interesting question is then, how the newcomer connects to (copies) which other nodes, and this depends on its trait repertoire in comparison to that of its neighbours.

---

[9] Smolla & Akçay (2019) assume that cultural traits are acquired through: (1) asocial learning (innovation), where traits are learned independently, with success probability g , or through (2) social learning (coping), where traits are copied based on their frequency in the individual's neighbourhood. The probability is quadratic to emphasize "complex contagion", given as: $P_s(t) = s \times p_t^2$, where p_t is the proportion of neighbours exhibiting the trait, and s is the coping success rate.

graphs naturally support a “Bechtelian hierarchy”: agent-level replacement loops nest within neighbourhoods, which nest within the whole population.

However, since the edges only depict coping relations and each round only one node “dies” and is replaced by a new one (“birth”), any causal analysis (of what actually happens in the real world) is quite limited, especially when compared to the next network model.

2.3. Interaction Graphs & Community Discovery

In contrast to Death-Birth graphs, where the edges always depict coping, another type of network is much more flexible when it comes to the interpretation of the evolving edges. Interaction graphs represent agents as nodes and their social interactions - collaboration, imitation, competition, signalling, etc. - as edges, often with dynamic rewiring to reflect evolving ties. Its specific target feature is the real-time diffusion pathways and emergent community structure: how patterns of interaction generate, sustain, or reorganise cohesive cultural subpopulations, see Figure 3. Community-discovery algorithms like TILES (Rossetti et al. 2017) identify “nearly decomposable” clusters in those networks (high internal vs. low external interaction rates, see Simon 2002) that correspond to cohesive cultural subpopulations, which can then be grouped hierarchically via modularity optimization. For example, Youngblood et al. (2021)

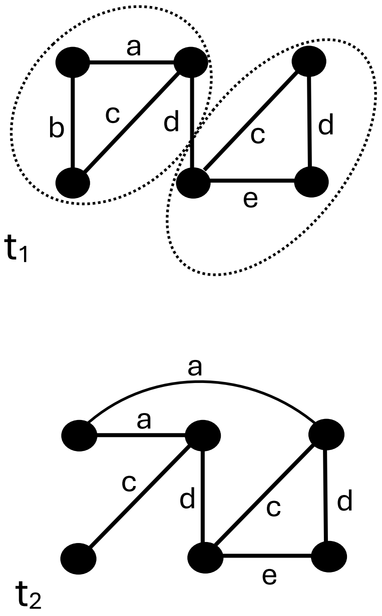

Figure 3: Two stages (t_1/t_2) of an Interaction Graph’s evolution. Nodes represent agents, edges (a,…,e) represent any kind of repeatable interaction. At t_1 the population can clearly be divided into two distinct groupings via “community discovery” (dashed circles), based on the (triangle) edge distribution; at t_2 however, this pattern changes, since one edge disappears, and a new one is added.

used hundreds of artist-collaborations from the DISCOGS database to trace electronic music styles. Extensions that integrate network-based game theory (Baraghith 2023) or experimental melody diffusion (Marjieh et al. 2025) further show how local reproduction and selection biases, mediated by topology, drive cultural complexity (and even aesthetic outcomes).

Seen through Craver's (2016) framework, interaction graphs unambiguously satisfy the structural-connectivity criterion: every edge records a real, non-idealized, difference-making social interaction. Because common subgraphs (triangles, dense cores, bridging ties) recur across the network and reliably shape processes like diffusion, cohesion, or innovation, the model also realises the causal-motif role, providing mechanism schemas rather than mere correlations. Thus, interaction graphs offer the granular causal structure required for a fully mechanistic explanation of a social network of specific agents, not just functional description.

Bechtel's (2020) hierarchy requirement deepens this assessment. Concrete interactions form the lowest level; community-detection algorithms compress them into cohesive groups on the "meso-level", and those groups can themselves organise into higher-order institutions or conventions. Each layer is defined only relative to the mechanism beneath it, "core" versus "peripheral" members within a community, for instance, precisely the context-dependent, multi-level decomposition Bechtel identifies as the hallmark of hierarchical mechanisms. Accordingly, Interaction Graphs deliver the richest mechanistic hierarchy among the four CET network types.

2.4. Trait Networks or "Cultural Systems"

A trait network maps individual cultural variants as nodes and their compatibility or incompatibility as weighted edges, and its specific target feature is the self-organisation of trait repertoires: how internal compatibility patterns drive the formation, stability, and diversity of coherent cultural packages. Trait networks abstract away from agents to focus on how traits co-evolve within a "cultural system" (Buskell et al. 2019; Janson et al. 2021). Often, these relationships represent compatibility, incompatibility or mutual synergy. Therefore, Trait-Networks are mostly *complete graphs*, where every node is connected to all other nodes, because every trait has a specific relationship to any other one, see Figure 4.[10] Pars pro toto, let us take a closer look at a recent Trait-Network model by Janson et al. (2021). In their model, cultural agents are more likely to adopt traits that are consistent with their existing beliefs and values (i.e., their *trait repertoire*). The authors evaluate network effects on cultural evolution based on three criteria: diversity, consistency and stability. After drawing a compatibility matrix (see footnote 3),

---

[10] In principle, such a complete graph could also be depicted as a compatibility matrix (w). It would look as follows in the case of Figure 4: $w = \begin{matrix} 0 & 1 & -1 & 1 \\ 1 & 0 & -1 & 1 \\ -1 & -1 & 0 & -1 \\ 1 & 1 & -1 & 0 \end{matrix}$, where 0 indicates, that the trait is neutral towards itself.

a logistic function is established to determine the coping probability.[11] A filtering mechanism makes sure that only traits are copied, which fit in an agent's existing repertoire. Population size (*N*) and number of traits (*T*) are fixed, which already provide two strong idealisations of the model. The outcome of the simulation measures every agent's repertoire size as well as internal consistency (i.e. the average compatibility within a repertoire)[12]. Trait filtering determines which traits are adopted based on their compatibility. Janson et al. (2021) transform a static compatibility matrix into a dynamical cultural-system mechanism: each edge weight w_{ij} (compatible = +1, incompatible = –1) modulates the probability that an agent will copy, display, or even invent trait *j* given its current repertoire. Those probabilities, fed into repeated social interactions and population turnover, generate higher-level outcomes such as repertoire size, internal consistency, cultural diversity, and system stability.

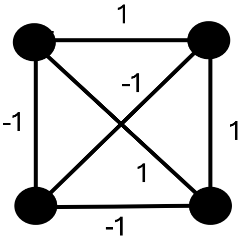

Figure 4: A Trait network, where the nodes stand for traits and the edges represent compatibility (1) or incompatibility (-1). Since two traits are always either compatible or incompatible, such a graph is often complete.

From Cravers (2016) perspective, this means the graph furnishes genuine structural connectivity: each edge is a difference-maker for the process of adoption. Self-reinforcing triads, cycles of mutual (in)compatibility serve as causal-motif schemas that explain why certain clusters of traits flourish while others are blocked. Bechtel's (2020) hierarchy requirement is likewise satisfied: trait-level compatibilities (lower level) feed into individual learning filters; repeated filtering operations aggregate into meso-level cultural niches; and those niches yield macro-level outcomes such as repertoire size, internal consistency, and long-term diversity. Thus, although trait networks lack agent-to-agent pathways, they still instantiate a multi-level mechanism in which organised part-relations (trait compatibilities) generate higher-level cultural structure, warranting a moderate mechanistic score in our evaluation.

---

[11] Their function takes the form: $p(s) = \frac{1}{1+e^{-k\times s}}$, where s is the average compatibility score of a new trait with an agent's repertoire and k is the strength of preference for compatibility. Population size (*N*) and number of traits (*T*) are fixed.

[12] Internal consistency (*I*) is given by: $I = \frac{1}{|R|}\sum_{ij\in R} w_{ij}$, where R is the set of traits already in an agent's repertoire and *i* and *j* are two traits.

Remember that Craver's taxonomy tells us that any network type will only *truly explain* if it identifies the relevant parts and activities (structural connectivity) and/or isolates the core difference-makers or motifs (causal-motif). Among the four network types, *Interaction Graphs* score highest in this category, because they explicitly model very specific agent-to-agent interactions, allowing for a detailed causal reconstruction of how cultural traits spread within structured populations. Unlike other models, they can incorporate complex dyadic interactions, making them well-suited for mechanistic explanations of cultural transmission on the level of cultural microevolution. By contrast, *Phylogenetic Trees* rank very bad in this category, since they are only to a very limited extend about mechanisms, but about a reconstruction of macro-level long term patterns of evolutionary descent. This is so, not because they are "macro-level descriptions", but because their branches represent historical succession rather than manipulable interactions. However, Phylogenetic trees have *other* explanatory strengths, as I will show in the next section.

3 Alternative Modes to measure Explanatory Power of CET Network Models

These four network approaches in CET do play *complementary* roles. Each model highlights different *aspects* of cultural evolution rather than competing for a single "true" explanation. I think that this requires engaging with different notions of "explanation" found in the philosophy of science literature, see e.g. Sprenger & Hartmann (2019). There is more than one mode of scientific explanation and that explanatory power very likely is not a singular concept but varies depending on the kind of explanation under consideration.

Section 2 assessed each CET network through a mechanistic lens, asking whether its structure maps real causal interactions (Craver 2016) and nests into multi-level mechanisms (Bechtel 2020). Yet mechanistic insight is only one way to "explain" a phenomenon. In this third section, I broaden the yard-stick by introducing three further, non-exclusive modes of scientific explanation: *Unificatory explanations* show how a single argument pattern compresses diverse cases and *topological explanations* derive the phenomenon as a mathematically necessary consequence of the network's shape. Re-examining the four network models under these lenses will reveal explanatory virtues that remain invisible when we focus solely on mechanism, and will set the stage for the pluralist synthesis in Section 5.

3.1. Unification as a Mode of Explanation

Another explanatory approach I want to look at is *unificatory explanation*, based on Kitcher (1989), where an explanation is considered strong if it integrates diverse phenomena under a small set of theoretical principles. Kitcher's main thesis is that explanation is the process of describing a wide variety of phenomena using as few argumentative patterns as possible repeatedly. Kitcher's unificatory account regards an explanation as strong to the extent that it subsumes diverse phenomena under a minimal set of argument patterns or “explanatory arguments”. The key measure is unificatory strength, which has two dimensions:

a. Breadth: How many distinct explananda are covered by the same schema?
b. Depth: How many layers of theory (surface patterns versus deeper principles) are linked by that schema?

Network models that perform well in this category should provide broad formal frameworks that systematize multiple aspects of cultural evolution under a single, coherent modelling approach. They would not only explain specific cases of cultural transmission but also provide a generalizable structure that applies across different domains, such as language evolution, technological diffusion, or the spread of social or political norms. The degree of unification achieved by a model depends on how well it reduces the number of independent assumptions needed to explain diverse cultural processes. In this sense, unificatory explanations favour models that reveal deep structural similarities across different cultural domains.

Following Humphreys’ (2001) distinction between analytic and synthetic understanding, note that I treat “unification” not as a strict alternative to causal-mechanical or topological explanation but as a synthetic virtue that can accompany either. Analytic understanding targets difference-making causal organization; synthetic understanding integrates results by showing how diverse cases fall under a reusable modelling schema. A network model may be causally informative and unificatory at once (and likewise topologically informative and unificatory), because unification concerns scope and schema-reuse, not a distinct “kind” of causal or structural relation. In other words, “unificatory power” tracks a cross-cutting explanatory virtue: the extent to which an explanatory schema is portable across cases and supports understanding by integrating diverse phenomena under a common pattern of reasoning. In this sense, both causal explanations and topological explanations can be unificatory. Topological explanations unify disparate systems insofar as the same structural property constrains outcomes across otherwise different causal realizations. Likewise, causal explanations can unify by identifying a common cause or common causal organization across heterogeneous realizers (as Humphreys [2001] stressed). The unificatory dimension therefore concerns scope and schema-reuse, not a distinct kind of dependence relation “over and above” causal or structural dependence.

We should further note here that the whole CET research enterprise *as such*, already provides a high level of unificatory potential, since it aims to reveal deep structural similarities between biological and cultural transmission/proliferation. For a more detailed investigation, see Baraghith & Feldbacher-Escamilla (2021). As a consequence of this, one should expect that all models used in the CET domain (whether they are network-like or not) bear a certain degree of unificatory explanatory power. A model has high unificatory potential if it successfully integrates diverse cultural phenomena under a coherent theoretical framework.

1. I think that *Phylogenetic Trees* perform best in this category, since they impose a structured evolutionary history onto cultural traits, providing a broad, systematic account of how different traditions/languages/technologies (or biological genes!) have evolved. Their unificatory strength makes them a popular framework in historical linguistics and

comparative anthropology, for a representative overview, see Evans et al. (2021). They yield a clear "one schema" account of macro-level cultural history – once stablished, it is even hard to see *how else* evolutionary history in any cultural domain should be formally depicted.

2. *Death-Birth Graphs* provide moderate unification, they partially unify cultural transmission with population (replacement) dynamics, offering a generalizable statistical framework that can be applied across different domains.
3. Trait Networks unify cultural evolution through internal trait compatibility structures, making them useful for explaining why some cultural elements persist while others are filtered out. However, because they focus on internal interactions rather than external transmission, their unificatory scope is more constrained than that of Trees.
4. Likewise, *Interaction Graphs* warrant a moderate unificatory potential. While they are to a large extend data-driven and context sensitive, and do not capture large-scale historical lineages like phylogenetic trees, they offer a super-flexible modelling framework applicable across diverse cultural domains. Whether tracking the diffusion of musical styles, opinions, or norms, they employ a shared structure - agents interacting through edges - and common transmission rules, such as conformist or payoff-biased learning. This reuse of formal patterns aligns with Kitcher's notion of explanatory unification. Marjieh et al. (2025) exemplify this by applying Interaction Graphs to musical evolution, demonstrating that selection and reproduction processes can be generalized across cultural forms. Further, Baraghith (2023) shows that even evolutionary game theory can be integrated into Interaction Graphs, extending their theoretical reach. Although they are limited in macrohistorical scope, their adaptability and structural coherence justify their classification as having moderate unificatory potential.

3.2. Topological Explanations

Last but not least, we must consider a mode of explanations, that appeared only quite recently in the literature: Topological Explanations. They directly address our general topic and have been shown to be a distinct mode of explanation that is not reducible to mechanistic explanations.[13]

Already Huneman (2018) offered an account of what he called "structural explanations": an explanation is structural when it leverages mathematical propositions about entire

---

[13] Kostić & Khalifa (2022) show that topological explanation can (in certain cases) be autonomous from mechanism. They reject the "Mechanistic Interpretation of Topological Explanation" (MITE), which says a network explains only if its nodes and edges denote the parts and interactions of an underlying mechanism that is counterfactually responsible for the effect at a higher level. Their counter-example (the correlation between motif frequency (MF3) in the macaque cortex and long-range functional connectivity) meets MITE's node-and-edge requirements but fails the responsibility and inter-level tests, yet still explains: varying MF3 would vary the functional links. Hence network shape alone can carry explanatory force, reinforcing the case for a topological mode distinct from, though sometimes overlapping with, mechanistic explanation.

model-classes to directly entail empirical phenomena, rather than merely representing or simulating underlying mechanisms. Not all structural explanations are "pure." Some (like e.g. Bergmann's rule) remain partly embedded in specific mechanisms, giving rise to a continuum from mechanism-dominant to purely structural explanations. Within the structural family, one can further distinguish sub-families (topological, algebraic, statistical), though these often overlap (e.g. graph matrices unify topology and algebra). This clarifies and systematizes a wide array of non-mechanistic explanations across the sciences. Furthermore, Hunemans idea of an "explanatory continuum" is in deep alignment with our idea of the scoring measure of explanatory strength, which we develop in the next section. This measure relies on Woodward (2025), who also worked on topological explanations. His two-folded approach will be considered in detail in section 4.

Kostić (2020) then formulated a "general theory of Topological Explanation" that states exactly how a property of a graph (the explanans A) genuinely explains a physical or dynamical phenomenon (the explanandum B). Three conditions must hold:

a. Facticity: both A and B are true of the target system.
b. Counterfactual dependence: had A not obtained, B would not have obtained; this can be vertical/global (e.g. small-worldness) or horizontal/local (e.g. high communicability between two nodes).
c. Explanatory perspectivism: the choice of global versus local counterfactual must match the scientist's "why-question."

Because A is structural and B behavioural, the dependency is asymmetric; reversing explanans and explanandum would fail these criteria, so no causal machinery is needed to block symmetry. Topological explanations thus differ from mechanistic ones: they prove that a phenomenon is mathematically necessary given network shape rather than tracing how parts and activities produce it. The same three-part template applies to any network domain, including (some) CET graphs. Kostić (2020) provides a crisp, criterion-based alternative to mechanistic, or unificatory accounts, made for network models. His facticity, counterfactual, and perspectival requirements tell us when "the shape of the graph alone" explains, and why that explanation remains one-directional.

Can we conclude that all of the network models have high explanatory strength in the topological mode, simply because they are all networks? In what follows, we will show that this is not the case, because the models differ in this respect, as well:

1. *Death-Birth Graphs* score high, since in a death–birth model the network is specified by an adjacency matrix W whose entries w_{ij} give the probability that individual i copies individual j. Global topological properties, such as isothermality (all nodes have equal "temperature", i.e. incoming weight), are therefore concrete, verifiable facts of W. Empirically, one can measure interaction frequencies in a real community and instantiate the matrix directly. Thus, the topological property has the potential to be factive in the sense Kostić requires. Concerning "counterfactual dependence", things get a little complicated: Remember that Kostic (2020) makes a distinction between "vertical-" and

"horizontal dependence". *The Isothermal Theorem* (Nowak 2006) states that on any isothermal graph, that realizes a Death-Birth Process, the fixation probability of a neutral mutant equals $1/N$. If the same population becomes non-isothermal (by adding a star hub or varying degrees), the fixation probability shifts to $\pi' \neq 1/N$ and may be amplified or suppressed depending on heterogeneity. So, the Counterfactual here would be: "*Had the network not been isothermal (e.g. had we introduced a high-degree hub) the global fixation probability would have changed*." Concerning the horizontal mode, we must understand that for a specific node i, fixation probability π_i is a function of its temperature or degree. Proven results (Allen & Nowak 2014) show that reducing i's degree lowers π_i. Counterfactual: "*Had node i possessed fewer outgoing links, its mutant would have had a lower chance of taking over.*" Because these dependencies are purely derived from graph-theoretic theorems - not empirical correlations - I think they satisfy Kostić's necessitating requirement to a high extent. Death–birth graphs ground their explanans in verifiable network facts and link those facts to fixation phenomena via proven counterfactual theorems.

2. *Phylogenetic trees* are not easy to evaluate in the topological mode. On the one hand, they provide a very concrete structure: branching topology, branch lengths, and standard summary indices such as consistency and retention indices (see footnote 5). It is also important, however, that a phylogeny is typically inferred rather than given: the evolutionary process generates a branching structure as lineages split, and CET studies normally report posterior support for clades. In that sense, "facticity" is satisfied only conditionally: the explanans is factual given acceptance of the best-supported tree (and the uncertainty attached to it). On Kostić's vertical/global reading, phylogenies can support topological explanations. For example, *coalescent theory* connects global tree-shape properties such as imbalance (A) to statistical properties of divergence-time patterns (B) under neutral evolution; in idealized settings, the counterfactual "*Had the tree been more balanced, divergence times would have been less clustered*" is mathematically tractable (cf. Pybus & Harvey 2000), though the strength of the result depends on substantive assumptions. On the horizontal/local reading, the relevant counterfactuals concern parts of a single tree (nodes, branches, subtrees). A change in local branch shape or length does not by itself entail a change in a particular trait event. It becomes difference-making only given additional modelling, i.e., an explicit character/transmission model that links trait histories to the tree (and, where appropriate, an explicit borrowing/reticulation model). Ordinary cultural "horizontal transmission" (borrowing across lineages) is therefore not an instance of Kostić's "horizontal" dependence within a fixed tree. Rather, it motivates moving beyond strict trees to reticulated structures ("phylogenetic networks"). For these reasons (conditional facticity, moderate vertical/global counterfactual support under idealizations, and only limited horizontal/local dependence without added character models), I think that Phylogenetic trees yield moderate explanatory strength in the topological mode.

3. In an *Interaction Graph,* the adjacency matrix records observed social contacts - who talks, trades, collaborates - during a fixed time-window or over repeated observations. Topological descriptors such as *global small-world coefficient σ* (ratio of clustering to path length), modularity (strength of community structure) or local communicability (weighted sum of all walks between two nodes) are therefore true properties of the actual network once the data are collected. Thus, interaction graphs totally satisfy Kostić's facticity requirement: the explanans A is an empirically grounded feature of the very system whose behaviour we seek to explain. Interaction Graphs also support necessitating counterfactuals. At the global level, diffusion theory (e.g. Newman 2018) proves that networks with high σ exhibit mean adoption times $T_{diff} \propto logN$ remove the long-range shortcuts and T_{diff} inflates to linear growth, so "*Had σ been low, the norm would have spread far more slowly."* At the local level, network-control results (see Gu et al. 2015) show the energy E_{ij} required to steer a signal from node *i* to *j* is inversely proportional to communicability C_{ij}; hence "*Had communicability between "sender" node A and "receiver" node B been lower (e.g. if several intermediary ties were absent) the energy (or time) needed to transmit the cultural variant would have been higher, making evolutionary adoption unlikely*." Because one and the same graph can answer a global "Why did trait X spread so fast in the population?" or a local "Why exactly those two nodes?" question simply by shifting focus, interaction graphs satisfy the perspectival requirement in a strong sense. No extensive recalibration is needed; the explanatory leverage flows from structural properties already contained in the data. For that reason, Interaction Graphs align exceptionally well with Kostić's facticity, dependence, and perspectivism requirements and justifiably receive a high rating on the topological axis.
4. Finally, *Trait Networks* consist of nodes that are themselves cultural variants and edges weighted by whether any two traits are judged to be compatible or incompatible. Although the resulting adjacency matrix is a perfectly legitimate network, it seldom meets Kostić's three conditions for a robust topological explanation. Facticity: many compatibility scores are researcher-assigned or inferred from sparse co-occurrence data, so the topological property A (e.g. a dense compatibility clique or a high average edge weight) is at best conjecturally true of the target system. Jansson et al.'s (2021) simulations reveal strong empirical dependencies regarding counter factuality. Global: Lowering mean compatibility *c* reliably shrinks culture size and raises internal consistency. Local: Flipping a single edge from +1 to –1 can block a trait's entry into an agent's repertoire. These links are *demonstrated by simulation, not proved to be mathematically necessary*, no general theorem shows that lowering a specific compatibility weight (or re-wiring trait links) necessarily changes the probability that those traits co-occur in repertoires. Finally, while Kostić's perspectival requirement could in principle be met - one might ask a local question ("Why do traits X and Y cluster?") or a global question ("Why is the cultural system fragmented?") - the answer would still hinge on unproven empirical

regularities rather than structural necessity. Because all three criteria are therefore satisfied only weakly, if at all, trait networks earn a poor score on the topological axis.

4 Can Explanatory Power be Quantified?

So far, I spoke of “high”, “moderate” or “poor” explanatory strength without really providing a rationale for such a measure. It is time to address this issue. In the philosophical literature, various attempts have been made to *quantify* explanatory power, most of which rely on probabilistic approaches. Some of the key contributions include McGrew (2003), Schupbach & Sprenger (2011), and Crupi & Tentori (2012), who propose measures of how much an explanation increases the probability of observed data. Other accounts, such as those developed by Myrvold (2003) or Lange (2004), link explanatory power to unification by analysing how well explanations “link up” or “screen off” unnecessary complexity.

However, due to significant challenges in developing a universal quantitative measure applicable to all three explanatory approaches, see e.g. Gebharter & Feldbacher-Escamilla (2023), I shall introduce a comparative ranking system here, which is based on “explanatory strength in a particular context” rather than absolute explanatory power. Instead of a numerical measure, I categorize the explanatory strength of each network model as (1) high, (2) moderate, (3) poor, or (4) none, depending on how well it aligns with the respective explanatory mode. This ranking system allows for a somewhat meaningful comparative analysis while avoiding the assumption that a single quantitative measure can be uniformly applied across all types of explanation, *for I do not see how to construct such a universal measure at this point*.

My four-level scale is not a popularity poll or a record of how researchers usually deploy a model. Instead, it rates how naturally the representation, as specified, satisfies each explanatory mode’s own normative standards from a philosophy of science perspective. I want to introduce two concepts, which govern these scoring:

(a) Intrinsic Affordance and
(b) Supplement Load.

What do I mean with that? *Intrinsic Affordance* refers to the explanatory resources a modelling framework provides by virtue of its built-in structure alone. It is what one can read straight off the representation - its variables, formal relations, and theorems that follow without further assumptions. If a model, as specified, already meets the normative criteria of a given explanatory mode, it does so by intrinsic affordance. *Supplement Load* is the additional empirical, conceptual, or mathematical work required to bring the same model up to those criteria when the intrinsic affordance is insufficient. This load may take the form of new data for calibration, auxiliary hypotheses, parameter fitting, intervention studies, or bespoke proofs that connect the model’s formalism to the explanandum. For example, a network model has high mechanistic affordance if its nodes and edges already pick out manipulable entities and interactions (e.g., an interaction graph whose edges are observed). It has high unificatory affordance if its formalism can be applied across a vast number of different (evolutionary) domains.

For example, consider interaction networks for Q3 (diffusion routes/speeds). Their *intrinsic affordance* is a direct representation of exposure structure (who interacts with whom). The main supplement load is (i) specifying/estimating a transmission rule (copying, payoff, prestige) and (ii) adding time-resolution or exposure controls to separate influence from homophily. By contrast, population-update graphs for Q2 (fixation under structure) intrinsically supply both a population structure and an explicit update rule. Their *supplement load* is typically parameter calibration (mutation and bias) rather than additional structural machinery. Finally, phylogenetic trees for Q1 (branching vs. reticulation) intrinsically provide lineage structure, but require a non-trivial supplement load to become explanatory: an explicit character/transmission model and (when relevant) tests for borrowing/reticulation.

Recently, also Woodward (2025) defined an interventionist counterfactual approach on network explanations: a network truly explains if outcome E depends on the pair $\langle N, D\rangle$, where N is structure and D is dynamics under ideal interventions (Schema W). He distinguishes IDE (Independent-Dynamics Explanation that needs both N & D) from rare DTE (Distinctively-Topological Explanations, where N alone suffices). Woodward's W highlights that any network explanation hinges on two ingredients: the connectivity structure *N* and a dynamics *D*. In my rubric, the structural features supplied by the model itself -its N- correspond to an intrinsic affordance, while the additional dynamical assumptions required to make the explanation run (D) constitute the typical supplement load. Thus, Woodward's framework can be seen as a further conceptual rationale for separating what a representation gives "for free" from the extra work needed to reach full explanatory force.

In judging explanatory strength, *the higher the intrinsic affordance and the lighter the supplement load, the stronger the score*. I define a scoring rule:

- *High*: ⟨N,D⟩ is intrinsically specified enough to test W with minimal supplement load (IDE satisfied; or DTE for the "topological mode" when N alone constrains E).
- *Moderate:* One of N or D is intrinsic; the other needs light–moderate supplements (data, parameters, identification) to test W.
- *Poor:* Only a proxy for N or D is intrinsic; testing W needs heavy supplementation or yields only indirect probes.
- *None*: The mode is inapplicable (no route to W even with substantial supplementation).

In short, Woodward's W turns my 'intrinsic affordances' into the ⟨N,D⟩ a model already gives you, and 'supplement load' into what must be added to complete ⟨N,D⟩ so that dependence of E on the right ingredients can be probed via ideal interventions.

The following Table 1 makes clear what I mean by the particular ranking categories for the respective modes of scientific explanation.

	Causal Mechanical	**Unificatory**	**Topological**

High	The model depicts step-by-step processes by which traits are transmitted or transformed (e.g., imitation biases, social interactions, feedback loops).	The model integrates multiple sub-phenomena (e.g., different cultural domains, multiple timescales) under a single theoretical scheme that reduces complexity.	The model pinpoints a factive network property A (global or local) and demonstrates that explanandum P counterfactually depends on A: “*Had A not obtained, P would not have obtained*.” Both vertical/global modes are matched to the explanatory question; asymmetry is secured.
Moderate	Some mechanistic details are present (e.g., rules for who copies whom), but larger causal structures remain black-boxed or simplified.	It captures some aspects of cultural phenomena (e.g. language) under a unified approach, but other phenomena (e.g. social norms) remain outside its scope.	It quantifies a relevant topological metric (degree distribution, betweenness, modularity, etc.) and links it to P, but the link is only empirical or heuristic; the counterfactual necessity is not rigorously shown, or the perspectival fit (global vs. local) is left implicit.
Poor	The model is mostly descriptive or correlational, lacking real insight into why or how the mechanism unfolds. The causal pathways described are very limited.	It is narrowly applicable, addressing just one phenomenon or scenario.	Topology is present only as descriptive ornament: metrics are reported or visualised, yet no attempt is made to argue that P follows necessarily from A, nor is any counterfactual tested.
None	The model is not at all about mechanism. It may track cultural frequencies or historical lineage without any attempt to specify the causal steps behind adoption or transmission.	No attempt is made to unify anything beyond the single scenario in question; the model is purely domain-specific or historically particular and cannot be generalized.	The model is non-topological: it either lacks a network representation or treats the graph purely as a timeline or intuitive taxonomy.

Table 1: A comparative ranking system (high; moderate; poor; none) for each of the three explanatory modes. Note that “unificatory power” is non-exclusive: it can be exhibited by causal-mechanical and/or topological explanations, and measures schema portability and integrative scope, not a separate kind of explanatory relation.

5 Explanatory Pluralism and a Roadmap for Hybrid CET Models

Section 2 and 3 showed that Phylogenetic trees, Death-Birth Graphs, Interaction Graphs and Trait Networks each emphasize different aspects of cultural evolution: lineage-based inheritance, network-driven diffusion, individual interactions/group formation or trait compatibility relations. By systematically mapping them onto different modes of explanation, we highlighted their respective contributions and limitations, providing a roadmap for researchers to select the most appropriate model for *their specific analytical goal.*

As already mentioned in section 1 of this paper, there are several canonical questions that CET researchers ask, e.g.:

- Q1. Why do historical lineages branch as they do, and when is apparent similarity due to reticulation (horizontal transfer/borrowing across lineages) rather than descent (vertical transmission along lineages)?
- Q2. Why does a trait fix (or fail to fix) with a particular probability and timescale under a given population structure?
- Q3. Why do diffusion paths take the routes and speeds we observe, e.g. who influences whom, when, and how strongly?
- Q4. Why do similar macro-patterns recur across domains, and could a common modelling schema account for different families of phenomena?
- Q5. Why do specific structural invariants, i.e. network features (e.g., degree heterogeneity, modularity) robustly shape evolutionary outcomes?

In this sense, *choosing a particular explanatory mode over another is choosing a type of answer, given the specific type of question*. The ranking (high, moderate, …) indicates the *quality of a possible answer* that the explanatory mode is able to deliver. A single explanatory mode does not cover the full problem space. Causal-mechanical analysis is the right tool when we need difference-making components, activities, and organization (notably Q2–Q3). Unificatory virtues matter when we aim for schematic economy across domains (Q4): the same model applies to many different domains. Topological virtues become decisive where structural invariants constrain what outcomes are reachable (Q5) or when we must distinguish convergence from real descent (Q1). The pluralist rubric makes explicit when a model's explanatory force depends on minimal vs. heavy add-ons (e.g., transmission models for trees; dynamic rules for interaction graphs).[14]

Having these questions and our ranking of (question-relative) explanatory strengths in mind, we can finally put all pieces together. What follows is a profiled characterisation of each model's explanatory contributions by question and mode; the framework is pluralist by design and eschews any aggregation into a single "best" model. Table 2 summarizes how the network models in CET align with the modes of scientific explanation, given the metric we introduced in section 4. Ratings are question-relative and mode-specific; they are not additive or comparable across modes, and no overall ranking ("total explanatory strength") is implied.

	Phylogenetic Trees	**Death-Birth Graphs**	**Interaction Graphs**	**Trait Networks**

---

[14] Our question-relative, multi-mode rubric treats CET models as inferential blueprints in Massimi's (2022) sense—tools for exploring possibility space and for fixing reliable, cross-perspectival claims about modally robust phenomena—so pluralism here is principled rather than eclectic.

Causal - Mechanical **Q2; Q3**	**Poor** Focuses on large-scale descent and historical branching / classification. Offers only minimal causal pathways. No recurrent difference-making motifs.	**Moderate** Includes local replacement rules (birth/death), providing an idealized mechanism. Coping edges become causal routes once calibrated, replacement loops serve as motifs, but the link to empirical mechanisms requires supplements.	**High** Model shows detailed social interactions (who imitates whom, at what rate), edges trace real causal interactions; recurring triads/bridges act as mechanism motifs about how cultural traits spread or stabilize.	**Moderate** Spells out how traits interact (compatibility, synergy), but omits who carries them; partial view of the underlying causal processes, when dynamical cultural systems are constructed based on their compatibility.
Unificatory **Q4**	**High** Unifies broad historical and linguistic data under a single genealogical framework, giving a clear “one schema” account of macro-level cultural history.	**Moderate** Unifies local birth–death processes across multiple contexts, but less capable of unifying *all* cultural phenomena under a single large-scale framework.	**Moderate** Can unify various short-term or small-scale social processes in a single interaction model, but doesn’t necessarily tie into deep historical or macro patterns.	**Moderate** Unifies diverse *intra-cultural* trait relationships within one conceptual map (synergy, conflict), clarifying large sets of co-evolving traits.
Topological **Q1; Q5**	**Moderate** Indices (Consistency/Retention) and tree diameter are topological properties of the inferred tree. More imbalanced trees necessarily raise variance in trait divergence times; however, proofs assume neutral models and ignore reticulation, so dependence is partial and global only.	**High** *Facticity*: the model’s adjacency matrix is explicit; graph properties are well-defined. Isothermal Theorem shows that altering degree heterogeneity *necessarily* raises or lowers fixation probability; both vertical and horizontal counterfactuals are mathematically proved. Both perspectives are possible.	**High** Highly supply global metrics (clustering, small-world coefficient) and local metrics (communicability). Researcher can pose either a global or a local question and obtain a counterfactual answer from the *same, unmodified* interaction graph.	**Poor** Compatibility matrices yield complete graphs with weighted edges, but many weights are hypothetical or researcher-assigned. No general theorem shows that altering a compatibility weight *necessitates* a change in co-occurrence; results are simulation-based.

Table 2: Summary of the comparison between the four types of network models and the three prominent modes of scientific explanations.

What can we learn from this overall comparison? Table 2 makes one point with stark clarity: no CET network model scores “High” across all three explanatory axes. Mechanistic power clusters in interaction graphs; unificatory power in phylogenetic trees; topological necessity in death-birth- and interaction graphs. These asymmetries are not accidents of current practice but reflections of intrinsic affordances: what a representation gives us for free and what must be supplemented.

The literature on network explanation often frames a contrast between mechanistic and topological explanation. My paper adds a further point: unificatory scope cuts across that contrast. Mechanistic models can unify when they identify common causal organization across heterogeneous realizers, and topological models can unify when the same structural constraint explains outcomes across systems with different causal details. Making this orthogonality explicit helps explain why "pluralism" in CET is not just a patchwork of models, but a structured division of explanatory labor. Philosophically, that pattern underwrites explanatory pluralism in two ways:

First, it shows that explanation is question relative. The researcher who asks "Which edges must we rewire to prevent misinformation?" needs the high mechanistic affordance of interaction graphs; the historical linguist, who asks "Can one bifurcation schema cover language, tool, and ritual evolution?" exploits the unificatory breadth of trees; the theorist who asks "Why is rapid diffusion inevitable in this topology?" turns to topological necessity. Different research questions (Q1-Q5) activate different columns of Table 2.

Second, the table reveals complementarity, not competition. A model that scores "Moderate" mechanistically but "High" unificatory can be *paired* with one that shows the opposite profile, yielding a composite explanation that is both manipulable and predictive.

This motivates a research agenda for *hybrid network models*: e.g. multiplex frameworks that, for example, embed empirically grounded interaction layers inside lineage trees, such as Youngblood et al.'s (2021) approach, which combines an Interaction-Graph model (micro), with a Phylogenetic Tree model (macro). Ideally, building and evaluating such hybrids will benefit from collaboration between CET modellers, who supply the empirical structure, and philosophers of science, who clarify which explanatory virtues are being added and at what supplement load.[15]

By showing that explanatory modes are complementary rather than competing, this paper clarifies why methodological diversity in CET is not a weakness but a resource, and how that diversity can be marshalled strategically. My framework is designed for two audiences: CET practitioners, who might need guidance matching scientific why-questions to model families and the minimal add-ons needed for defensible explanatory claims, and philosophers of science, who gain a do-main-grounded testbed for comparing explanatory virtues without collapsing them into a single metric. My tool functions like a diagnostic matrix: it does not rank models

---

[15] This roadmap is intentionally schematic: the aim is to mark promising integration points and the kinds of supplementary assumptions they would require, not to propose a fully worked modelling programme within the present word limit of this paper.

overall; it helps researchers (a) choose a model that fits their question, and (b) report limits and supplement load explicitly.

In short: I do not claim to locate a single, best theory of explanation for networks. Instead, I provide a principled way to measure which explanatory virtues each CET network model delivers and to decide, case by case, which virtues matter for a given scientific question.

6 References

Allen, B., Nowak, M. (2014): Games on Graphs. EMS Surv. Math. Sci. 1, p. 113–151.

Baraghith, K. (2023). From Games to Graphs: Evolving Networks in Cultural Evolution. In A. du Crest, M. Valković, A. Ariew, H. Desmond, P. Huneman, & T. A. C. Reydon (Eds.), Evolutionary Thinking Across Disciplines (Synthese Library, vol. 478, pp. 161–181). Springer, Cham.

Baraghith, K. & Feldbacher-Escamilla, C. (2021): The Many Faces of Generalizing the Theory of Evolution, in: American Philosophical Quarterly, Vol. 58, Issue 1, 35-49.

Bechtel, W. (2020): Hierarchy and Levels: Analysing Networks to Study Mechanisms in Molecular Biology. Philosophical Transactions of the Royal Society B: Biological Sciences 375 (1796): 20190320.

Blau, J. (1989): The Shape of Culture: A Study of Contemporary Cultural Patterns in the United States (American Sociological Association Rose Monographs). Cambridge: Cambridge University Press. doi:10.1017/CBO9780511570865

Boon E., van den Berg P., Molleman L., Weissing F.J. (2021): Foundations of cultural evolution. Philos Trans R Soc Lond B Biol Sci. 2021 Jul 5;376(1828): 20200041.

Bortolussi, N.; Durand, E.; Blum, M.; François, O.: (2006) apTreeshape: statistical analysis of phylogenetic tree shape. Bioinformatics. 1;22(3):363-4.

Boyd, R. & Richerson, P. J. (1988): Culture and the Evolutionary Process. Chicago: The University of Chicago Press.

Buskell, A., Enquist M., Jansson, F. (2019): A systems approach to cultural evolution, Palgrave Commun. 5, 131. (doi:10.1057/s41599-019-0343-5)

Cantor, M., Whitehead, H. (2013): The interplay between social networks and culture: Theoretically and among whales and dolphins. Philos. Trans. R. Soc. B Biol. Sci. 368, 20120340.

Cavalli-Sforza, L. & Feldman, M. (1981): Cultural Transmission and Evolution: a quantitative approach. Princeton: Princeton University Press.

Craver, Carl F. (2016): The Explanatory Power of Network Models. Philosophy of Science 83 (5): 698–709.

Crupi, V., & Tentori, K. (2012): A second look at the logic of explanatory power (with discussion). Philosophy of Science, 79(3), 365–385.

Doolittle, W. (2009): The practice of classification and the theory of evolution, and what the demise of Charles Darwin's tree of life hypothesis means for both of them. Philosophical Transactions of the Royal Society, B 2009, 364, pp. 2222-2228.

Evans, C.L., Greenhill, S.J., Watts, J., List, J-M., Botero, C.A., Gray, R.D., Kirby, K.R. (2021): The uses and abuses of tree thinking in cultural evolution. Phil. Trans. R. Soc. B 376: 20200056.

Gebharter, A. (2014): A Formal Framework for Representing Mechanisms? Philosophy of Science , Volume 81/Issue 1 , pp. 138 – 153.

Gebharter, A., & Feldbacher-Escamilla, C. J. (2023). Unification from a causal perspective. The British Journal for the Philosophy of Science, 99, 28-36.

Gu, S.; Pasqualetti, F.; Cieslak, M.; Grafton, S. T.; Betzel, R. F., & Bassett, D. S. (2015): Controllability of structural brain networks. *Nature Communications* 6 (2015): 8414.

Gray, R.; Greenhill, S. and Robert M. Ross (2007): The Pleasures and Perils of Darwinizing Culture (with Phylogenies), Biological Theory 2(4) 2007, 360–375.

Humphreys, Paul W. (2001): Analytic versus Synthetic Understanding, In Paul W. Humphreys, Philosophical Papers (2019), pp. 200–221. New York: Oxford University Press.

Huneman, P. (2018): Outlines of a theory of structural explanations. Philosophical Studies, 175(3), pp.665-702.

Janson, F., Aguilar, E., Acerbi, A., Enquist, M. (2021): Modelling cultural systems and selective filters. Phil. Trans. R. Soc. B 376, 20200045.

Kostić, D. (2020): General Theory of Topological Explanations and Explanatory Asymmetry. Philosophical Transactions of the Royal Society B: Biological Sciences, 375: 20190321.

Kostić, D., Khalifa, K. (2022): Decoupling Topological Explanations from Mechanisms. Philosophy of Science, 90(2), 245-268.

Kitcher, P. (1989): Scientific Explanation (Minnesota Studies in the Philosophy of Science, Volume 13), Minneapolis, MN: University of Minnesota Press.

Lange, M. (2004). Unification and scientific explanation. Philosophy of Science, 71(3), 286–307.

Lewens, T.: (2015): Cultural Evolution - Conceptual Challenges, Oxford: Oxford University Press.

Lewens, T. (2020): Cultural Evolution, The Stanford Encyclopedia of Philosophy (Summer 2020 Edition), Edward N. Zalta (ed.) <https://plato.stanford.edu/archives/sum2020/entries/evolution-cultural/>.

Mace, R., & Holden, C. J. (2005): A phylogenetic approach to cultural evolution. Trends in Ecology & Evolution, 20(3), 116–121.

Marjieh, R., Anglada-Tort, M., Griffiths, T. L., & Jacoby, N. (2025): Characterizing the Interaction of Cultural Evolution Mechanisms in Experimental Social Networks. Preprint available at: https://arxiv.org/abs/2502.12847.

Massimi, M. (2022): Perspectival Realism. 1st ed. Oxford University Press, New York.

McGrew, T. (2003): Confirmation, heuristics, and explanatory reasoning. British Journal for the Philosophy of Science, 54(4), 553–567.

Maynard Smith, J. (1986): Natural selection of culture? New York Review of Books, Nov. 6, p. 33.

Mesoudi, A. (2011): Cultural evolution: how Darwinian theory can explain human culture and synthesize the social sciences, Chicago, IL: University of Chicago Press.

Myrvold, W. (2003): A Bayesian account of the virtues of unification. Philosophy of Science, 70(3), 399–423.

Newman, M. (2018): Networks, Oxford University Press.

Nowak, M. (2006): Evolutionary Dynamics: exploring the equations of life. Cambridge (Mass.); London : Harvard University Press.

Pearl, J. (2000): Causality: Models, Reasoning, and Inference. Cambridge University Press.

Piccinini, G. & Craver, C. (2011): Integrating Psychology and Neuroscience: Functional Analyses as Mechanism Sketches. Synthese 183 (3): 283–311.

Pybus, O.G. & Harvey, P.H. (2000): Testing macro-evolutionary models using incomplete molecular phylogenies. Proceedings of the Royal Society B 267: 2267–2272.

Rossetti, G., Pappalardo, L., Pedreschi, D. & Giannotti, F. (2017): Tiles: an online algorithm for community discovery in dynamic social networks. Machine Learning 106, p. 1213–1241.

Schupbach, J. N., & Sprenger, J. (2011). The logic of explanatory power. Philosophy of Science, 78(1), 105–127.

Simon, H. A. (2002). Near decomposability and the speed of evolution. Industrial and Corporate Change, 11(3), 587–599.

Smolla, M., Akçay, E. (2019): Cultural selection shapes network structure. Sci. Adv. 5, eaaw0609. (doi:10. 1126/sciadv.aaw0609).

Sprenger, Jan and Stephan Hartmann (2019): Bayesian Philosophy of Science: Variations on a Theme by the Reverend Thomas Bayes, Oxford/New York: Oxford University Press.

Sterelny, K. (2017): Cultural evolution in California and Paris: a comparative review of Joseph Henrich's "the secret of our success" (Princeton University Press, 2016; pp xv, 445) and Olivier Morin's "How traditions live and die" (Oxford University Press, 2016; pp xvi, 300). Studies in history and philosophy of science part C: studies in history and philosophy of biological and biomedical sciences.

Woodward, J. F. (2025): Networks, Dynamics and Explanation. Synthese, 205(5), pp.1-28.

Youngblood, M., Baraghith, K. & K. P. Savage (2021): Phylogenetic Reconstruction of the Cultural Evolution of Electronic Music via Dynamic Community Detection (1975-1999), Evolution and Human Behavior 42/6.