A natural history of networks: Higher-order network modeling for paleobiology research

Alexis Rojas[1, 2*], Anton Eriksson[2], Magnus Neuman[2], Daniel Edler[2, 3, 4], Christopher Blöcker[2] and Martin Rosvall[2]

[1]Department of Computer Science, University of Helsinki, 32611, Finland

[2]Integrated Science Lab, Department of Physics, Umeå University, SE-901 87 Umeå, Sweden

[3]Gothenburg Global Biodiversity Centre, Box 461, SE-405 30 Gothenburg, Sweden

[4]Department of Biological and Environmental Sciences, University of Gothenburg, Box 461, SE-405 30 Gothenburg, Sweden

*** Correspondence:**
Corresponding Author
alexis.rojasbriceno@helsinki.fi

Keywords: paleobiology, geohistorical data, networks, higher-order, Infomap, Map Equation.

Abstract

Paleobiologists often employ network-based methods to analyze the inherently complex data retrieved from geohistorical records. However, they lack a common framework for designing, performing, evaluating, and communicating network-based studies, hampering reproducibility and interdisciplinary research. The high-dimensional and spatiotemporally resolved data also raise questions about the limitations of standard network models to represent the local-, regional-, and global-scale systems considered in paleobiology. By assuming independent pairwise links, standard network models wash out higher-order node interactions and can obscure paleontological patterns. These challenges provide an opportunity to move paleobiology research beyond standard network representations toward higher-order models better suited for the complex relational structure of the geohistorical data. Higher-order representations can represent the spatiotemporal constraints on the information paths underlying geohistorical data, capturing the high-dimensional patterns more accurately. Here we describe how to use the Map Equation framework for designing higher-order models of geohistorical data, address some practical decisions involved in modeling complex dependencies, and discuss critical methodological and conceptual issues that make it difficult to compare results across studies in the growing body of network paleobiology research. We illustrate multilayer networks, hypergraphs, and varying Markov time models for higher-order networks in case studies based on the fossil record and delineate future research directions for current challenges in the emerging field of network paleobiology.

1 Current challenges and opportunities for network-based paleobiology research

Network science is transforming scientific research and thinking in the twenty-first century. Many natural and social phenomena can be described as networks, where nodes represent individual components, and links indicate their interactions. Network science studies high-dimensional, heterogeneously structured, complex systems and their underlying processes (Barabási and Pósfai 2016). In recent years, standard network models based on pairwise or direct interactions between individual components have been applied to almost every area of paleontological research, including

biostratigraphy (Muscente et al. 2019), biogeography (Dunhill et al. 2016; Kiel 2017; Rojas et al. 2017; Kocsis et al. 2018; Jeon et al. 2021), macroecology (Roopnarine 2010), and macroevolution (Kocsis et al. 2018; Muscente et al. 2018). However, methodological inconsistencies and conceptual issues in the emergent interdisciplinary field of network paleobiology make it challenging to reproduce experiments and compare outcomes across studies. In addition, the complexity of the high-dimensional and spatiotemporally resolved data retrieved from geohistorical records raises questions about the limitations of the standard network models to accurately represent the variety of local (outcrops, stratigraphic sections), regional (geological basins), and global scale systems (The Earth-Life system, Rojas et al. 2021) examined in paleobiology.

Although network science provides methods for statistical analysis and machine learning of relational objects (Brandes et al. 2013; Lambiotte et al. 2019), paleobiologists often describe network analysis as a tool for visualization and qualitative assessment (Huang et al. 2016; Penn-Clarke and Harper 2020; Ye et al. 2021). While network visualization techniques are powerful tools for exploratory data analysis (Perri and Scholtes 2020), this misrepresentation reflects a lack of understanding of the broader field of network science that influences research practices in paleobiology. Recent studies (Eriksson et al. 2021; Rojas et al. 2021) suggest that the most critical conceptual issue in the growing body of network paleobiology is ignoring the extent to which the choice of a network impacts the results. There are also methodological inconsistencies, including inadequate description of the input network, incomplete explanation of the clustering approach, and uncritical acceptance of the network partition without validation. In addition, the lack of a common framework obstructs interdisciplinarity, reproducibility, and communicability. There is an urgent need to standardize research practices in the emergent field of network paleobiology, including how to design, perform, communicate, and evaluate network studies.

In this overview, we describe how to use the Map Equation framework (Rosvall and Bergstrom, 2008; Edler et al., 2017) for designing higher-order network models of geohistorical data. Specifically, we describe the concept of higher-order interactions in networks, address some practical decisions involved in network-based modeling of geohistorical data, and illustrate alternative network models, including multilayer networks, hypergraphs, and varying Markov time models for higher-order networks, with cases based on real geohistorical records. In addition, we delineate future research directions for current challenges in network paleobiology. We focus on the Map Equation framework and its applications in paleobiology research because it is an increasingly popular alternative to the standard statistical approaches currently used in paleobiology research (Rojas et al., 2021). The associated software called Infomap for finding community structure in standard and higher-order networks is also widely used in natural and social sciences (Rosvall and Bergstrom 2008; Lancichinetti and Fortunato 2009). The Map Equation framework for higher-order networks combines spatial, temporal, sedimentological, ecological, morphological, taxonomic, and any other geohistorical data, enabling standardized integrated investigation of the intricate interactions between plate tectonics, natural and anthropogenic climate change, and the evolution of life.

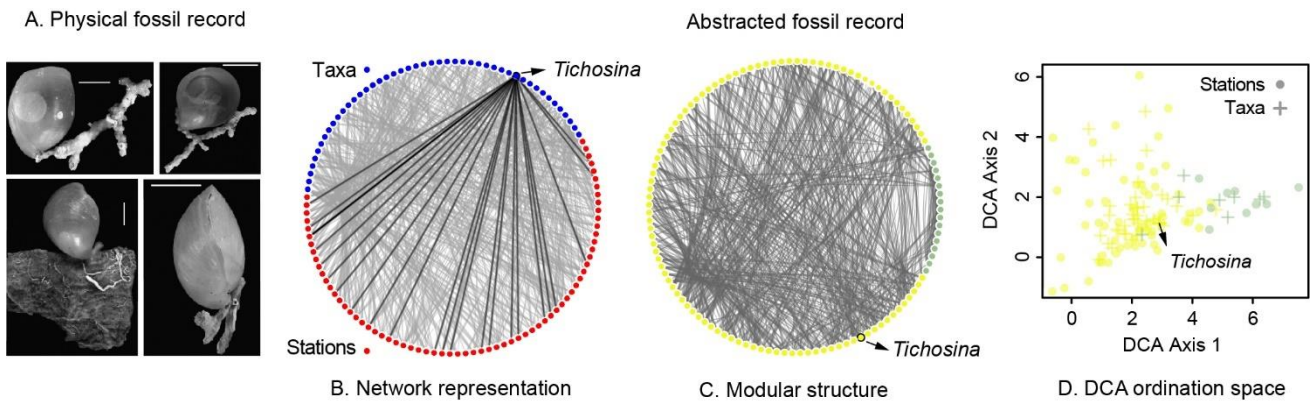


Figure 1. Network representations of geohistorical data are abstracted fossil records that can be improved. **A**. Physical Fossil record. These brachiopod shells aim to represent the benthic macrofauna from the R/V Pillsbury program in the Caribbean. Modified from Rojas et al. (2022, fig. 2). **B**. Abstracted fossil record. The abstracted record is a bipartite network representation of the underlying data (sampling stations × taxa matrix) (Rojas et al. 2022). The brachiopod *Tichosina*, the larger component of the Cenozoic brachiopod faunas in the Caribbean, is indicated. **C**. Modular structure delineated via community detection with the Map Equation framework and using a Markov Time = 2 (Kheirkhahzadeh et al. 2016). Modules include sampling stations and taxa. Nodes are rearranged in the circular layout by their module affiliation. Only the two larger modules are displayed; they represent 98% of the network flow. **D**. Modules mapped on a Detrended Correspondence Analysis (DCA) ordination space.

2 The Map Equation framework for higher-order networks

Research in network paleobiology has focused on delineating community structure that reveals, for example, biozones, bioregions, or evolutionary faunas, based on statistical regularities (Figure 1a-d). This unsupervised learning task known as community detection has been central to the field of network science, and researchers have developed different methods for different purposes (Schaub et al. 2017). The Map Equation framework models dynamics on networks using random walks to identify flow-based communities Rosvall and Bergstrom 2008). Random walk-based methods for community detection are attractive for geologist and paleobiologists because of their ability to explore the biosedimentary record at multiple scales and the intuitive interpretation of how network flows represent movement patterns among network's nodes (Lambiotte and Rosvall 2012) (e.g., species moving over their biogeographic range). Previous research has found that the Map Equation outperforms other approaches when operating on various benchmark networks (Lancichinetti and Fortunato 2009; Aldecoa and Marín 2013; Kheirkhahzadeh et al. 2016). The Map Equation framework consists of an objective function with the same name accompanied with its search algorithm Infomap.

2.1 The map equation

The Map Equation measures the modular description length of a random walk on a network (Rosvall and Bergstrom 2008). It exploits the equivalence between compressing data and finding regularities in the data (Rissanen 1978). The shorter the modular description of the random walk, the more detected modular regularities. The modules that minimize the Map Equation reveal the most modular regularities.

A random walk, consisting of a sequence of steps from node to node along randomly chosen links, possibly proportional to the link weights in weighted representation, is the simplest model of network dynamics. For example, on a bipartite occurrence network consisting of taxa and sampling stations where they occur, a random walk starts on a randomly selected station and moves to a random linked taxon. Then it moves to a station within the taxon's geographic range (Figure.1b), selected randomly, and repeats indefinitely. If this occurrence network has groups of highly interconnected stations and taxa, the random walk persists within those modules for a relatively long time. By describing a random walk on a network and not the network itself, the Map Equation highlights essential modular dynamics induced by the network structure (Figure 1c-d). These network-flow-based modules are essential for the network's function because they capture influence between nodes beyond nearest neighbors (Rosvall and Bergstrom 2008).

The Map Equation captures the statistics of random walks using their average visit rates of nodes and links. The per-step average codelength is the cost in bits of specifying steps within modules and between modules weighted by their usage rates (Rosvall & Bergstrom 2008). Small modules enable short descriptions within modules because they require little information to specify the visited nodes but may lead to long descriptions between modules from frequent and expensive module exits and entries. In contrast, large modules enable short descriptions between modules because between-module steps are rare and cheap but have long descriptions within modules from the many nodes to differentiate. Small modules where random walks persist long compress the network flows maximally and reveal the most modular regularities for the modeled network flows (Rosvall & Bergstrom 2008; Rosvall et al. 2009).

The Map Equation generalizes straightforwardly to higher-order network models, including memory networks (Rosvall et al. 2014), multilayer networks (De Domenico et al. 2015), hypergraphs (Eriksson et al. 2021), and varying Markov time models (Kheirkhahzadeh et al. 2016). Recent Bayesian generalizations of the Map Equation deals with incomplete data in a theoretically founded way (Smiljaníc 2020, Smiljaníc 2021).

2.2 The search algorithm Infomap

Because it is impossible to check all modular solutions in all but the smallest networks, Infomap uses a fast stochastic search algorithm to minimize the Map Equation over possible node assignments (Rosvall and Bergstrom 2008). The algorithm consists of multiple search procedures (Edler et al. 2017). The core algorithm starts by assigning each node to its own module. Then it repeatedly loops through each node in random order and moves it to the module that reduces the codelength the most. Infomap repeats this procedure until no move decreases the codelength, rebuilds the network with the modules forming nodes at a coarser level, moves these nodes into even coarser modules, and so on until no move reduces the codelength further.

To improve this two-level solution, Infomap alternates between a fine-tuning and a coarse-tuning procedure by moving individual nodes or sub-modules between modules. To find a hierarchical solution, Infomap starts from the two-level solution and iteratively builds super-module levels that compress the description of movements between modules. Then it clears the structure under each of the coarsest modules and recursively and in parallel builds sub-modules within each module until no finer structure can be found that decreases the hierarchical codelength. In this way, the resulting hierarchical structure of the network may have branches of different depth.

When modeling network dynamics on standard networks through random walks, a single node type usually represents a physical component of the system and describes the flows with the nodes' links

at the same time (Figure 2a-b). In contrast, the Map Equation framework for higher-order networks distinguishes physical nodes, which represent the system's components, from state nodes, which describe the system's internal flows (Edler et al. 2017) (Figure 2c). State nodes in the Map Equation framework can represent, for instance, memory of previous steps, layers in multilayer network representations (Rojas et al. 2021), lumped states, or any other complex relationship in the underlying geohistorical data. To capture the higher-order nature of flows on networks with state nodes, Infomap applies the same procedures to the state nodes with aggregated visit rates for all state nodes of the same physical node assigned to the same community.

3 Higher-order network models in paleobiology

3.1 The complex relational structure of the geohistorical data

Geohistorical records, either stratigraphic sections, boreholes, ice cores, or archaeological sites, are inherently complex. They provide spatiotemporal, environmental, and biological information about a particular system (Council 2005). Despite their limitations (Kidwell and Holland 2002; Kidwell and Tomasovych 2013), the high-dimensional and spatiotemporally resolved data retrieved from individual geohistorical records allow for evaluation of past biotic responses to natural and human-induced environmental changes at local to regional scales (Council 2005; Scarponi and Kowalewski 2007; Dietl and Flessa 2011; Dietl et al. 2015; Durham and Dietl 2015). Although high-precision chronological studies have improved our understanding of past biotic crises (e.g., Smith et al. 2018), compilations of individual geohistorical records are central for studies at larger spatiotemporal scales. Fossil occurrences of the benthic marine invertebrates in the Paleobiology Database (PaleoDB) (Peters and McClennen 2016) have become the benchmark data for network-based research on macroevolution, macroecology, and biogeography (Kocsis et al. 2018; Muscente et al. 2018; Rojas et al. 2021). In most cases, PaleoDB collections have geographic information and are assigned to a geological stage, making the modelling of the temporal constraints straightforward. PaleoDB collections also have lithostratigraphic and sedimentological information and sometimes include taphonomy and body-size data. Each fossil occurrence in PaleoDB belong to one collection, has a name with a certain taxonomic resolution, and is linked to an independent taxonomic classification with associated ecological information. These complex relational data describe the structure of the Phanerozoic life at multiple taxonomic levels and spatiotemporal scales.

There are also numerous databases covering specific taxonomic groups, time intervals, and/or geographic regions, for instance, the New and Old Worlds fossil mammal database, and the Strategic Environmental Archaeology Database. These sources compile high-resolution archaeological and paleontological data on individual geohistorical records. In most cases, the samples age is a value taken from the original literature sources and obtained from a range of dating methods varying in precision and accuracy (Buckland 2014). Because samples may have an age range larger than the length of the preferred bin interval, modeling temporal constraints of high-resolution data using multilayer networks is challenging. Pilotto et al. (2022) approached this question using the Map Equation framework to investigate the recent fossil record of European beetles. Their multilayer network analysis relaxes the temporal constraints, allowing a random walker to move toward neighboring layers without exceeding the age limits of the samples in the data. In practice, with ordered layers representing 500-year time intervals, and an accepted age range of 2000 years in the filtered samples, authors allowed a random walker to relax toward the first two layers in each direction. This approach accounts for the age uncertainty inherent to the samples, making it possible to explore high resolution geohistorical records using multilayer representations.

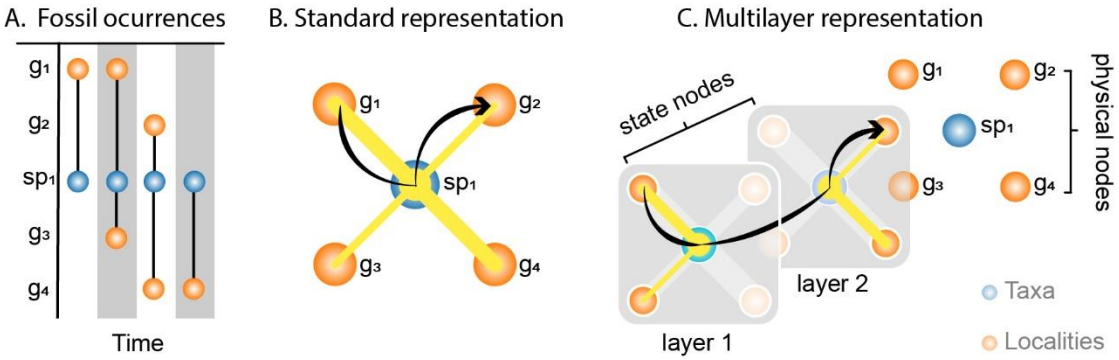


Figure 2. Network representations of geohistorical data. **A.** Temporal occurrence data. **B.** Standard bipartite network representation created by aggregating fossil occurrences (sp. 1) into arbitrary units (g1 to g4). This network uses the same nodes to represent the physical components of the system as well as to describe system' flows. A random walk as the simplest information flow model forms paths across independent pairwise links on this network washing out higher-order node interactions. **C.** Multilayer network representation. The Map Equation framework for higher-order networks distinguishes physical nodes that represent the system's components from state nodes describing the system's internal flows. This higher-order model better capture the constraints on the information paths and thus flows tend to stay among units from each layer.

3.2 Higher-order networks capture the complexity of the geohistorical data

One of the major conceptual changes in modern paleobiology research has been the distinction between the physical fossil record, consisting of in-situ or ex-situ specimens, and abstracted representations based on data retrieved from this physical record (Sepkoski 2013; Allmon et al. 2018). We capitalize on this idea by explicitly considering network representations of geohistorical data as abstracted records that can be described but also improved. Intuitively, a good model of the physical fossil record should be maximally parsimonious, but it must also be sufficiently complex to capture the complex interactions in the spatiotemporally resolved and high-dimensional geohistorical data. Unfortunately, researchers designing network representations of geohistorical data usually ignore that each network has a set of assumptions about the spatiotemporal structure and dynamics of the biosedimentary record (e.g., whether or not to describe temporal constraints) and their choice impacts the outcome of their studies (e.g., whether or not the solutions capture larger-scale patterns). In paleobiology, researchers tacitly assume that using benchmark data guarantees reproducibility. Here we argue that setting benchmark is required to improve reproducibility and communicability in network paleobiology. This challenge is also an opportunity to move beyond standard network representations, toward higher-order models (Benson et al. 2016) that better capture the complexity of the geohistorical data.

Network representations of geohistorical data are usually standard models based on pairwise connections between taxa (e.g., species, genus) and sampling units (e.g., collections, localities, grid cells) (Figure 2a). The standard approach consists in modeling dynamical processes on these networks with first-order flows. This modelling approach oversimplifies the underlying dynamics of the system because forms paths across independent pairwise connections, washing out higher-order node interactions (Figure 2b). Whereas standard models use a single node type to represents the physical components of the system (e.g., taxa, sampling units) and model flows through one-step dynamics on their links, the Map Equation framework for higher-order networks introduces abstract

nodes to describe the different states in which the physical nodes can be in the system (De Domenico et al. 2015) (Figure 2c). For instance, the multilayer network representation of the benchmark data on the Phanerozoic benthic marine faunas from Rojas et al. (2021), is a higher-order model in which a physical node representing a given taxon contains a number of state nodes carrying information on the geological stage where this taxon occurs. Therefore, this higher-order representation created through the Map Equation framework is a form of memory network (Edler et al., 2017). First-order models ignore the temporal constraints inherent to the underlying system, obscuring the large-scale modular patterns (Figure 3). When comparing unipartite, bipartite, and multilayer representations of the Phanerozoic benthic marine faunas, the multilayer network achieves the shortest codelength and the best compression (Table 1).

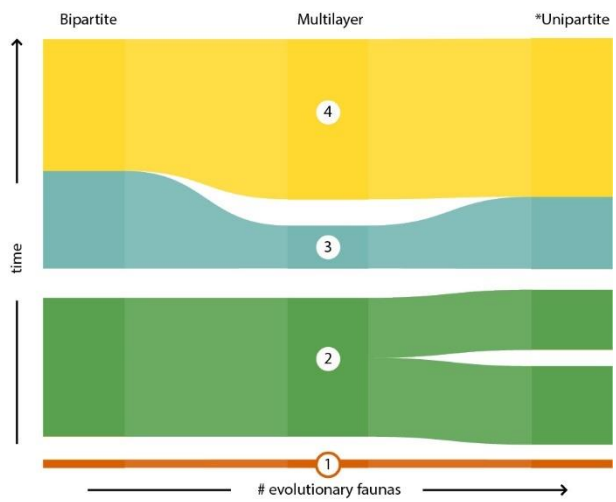


Figure 3. Different network representations capture different aspects of the benchmark data on the Phanerozoic marine faunas (Paleobiology Database). Bipartite and unipartite representations ignore the temporal constraints inherent to the biosedimentary record. *Unipartite projection obtained from the bipartite network by rescaling the Markov time (Kheirkhahzadeh et al. 2016). Alluvial diagram representing 98% of the network flow in each case.

Model	Physical nodes	State nodes	Levels	Top modules (99%)	Codelength (bits)	Compression (%)
Unipartite*	23203	—	2	4	11.0646	8.6009
Bipartite	23203	—	3	3	10.5169	13.1253
Multilayer	23203	64508	7	4	**6.2826**	**55.2804**

Table 1. Comparison of the modular structure of different network representations of the benchmark data on the Phanerozoic marine faunas. The multilayer representation achieves the shortest codelength and the best compression. For each partition, we measure the compression by using the corresponding one-level partition as a baseline. *Unipartite projection obtained from the bipartite network by rescaling the Markov time (Kheirkhahzadeh et al. 2016).

3.3 Alternative higher-order network models for geohistorical data

3.3.1 Multilayer networks

Multilayer networks are models used to represent complex systems with multitype interactions. In the Map Equation framework, multilayer networks can be used to model temporal and non-temporal data. Temporal constraints in real geohistorical systems includes ordered geological stages (Rojas et al. 2021) (Figure 3), arbitrary temporal bins (Pilotto et al 2022), and biostratigraphic frameworks

(Viglietti et al. 2022). Interlayer dynamics in multilayer networks is modeled based on the intralayer information (Eriksson et al. 2022). For instance, when layers represent geological stages, intralayer link structure describes the constraints on the network dynamics at a given stage, whereas interlayer link structure is created through neighborhood flow coupling between state nodes of the same physical node (Aslak et al. 2018; Rojas et al. 2021). In practice, the Map Equation framework can generate interlayer links using a relax rate (r), with a random walker moving between nodes within a given layer guided by intralayer links with probability $1-r$ and relaxing to adjacent layers guided by links between state nodes of the same physical node with a probability r (Edler et al. 2017). Previous studies show that a relax rate equal to 0.25 is large enough to capture temporal structures but small enough to preserve intralayer information (Aslak et al. 2018; Rojas et al. 2021). The bidirectional geo-historical time arrow in multilayers is obtained by limiting to 1 the number of adjacent layers in each direction to which a random walk can relax.

Multilayer networks can be also used to model geohistorical data lacking temporal constraints. In general, a geohistorical record with its multiple biotic and abiotic components can be conceptualized as a complex system with a multilayered structure, where each layer describes a particular interaction between sampling units, for instance, those resulting from the fossil composition and sedimentological features, independently. Physical nodes in this two-layered network can represent sampling units of any scale (e.g., samples, beds, members, formations), taxa of any resolution (e.g., species, genus), and sedimentologically defined groups (e.g., textural classes). Whereas each sampling unit in this multilayer network would be represented by a number of state nodes equal to the number of multitype interactions considered in the study, other physical components would be represented by a single state node because they occur only in one layer. In this network, movements between state nodes within each layer represent pairwise interactions, whereas movements between nodes across layers represent higher-order interactions (De Domenico et al. 2015). Our first example illustrates this approach on a case study based on the marine invertebrate fossil record (Holland and Patzkowsky 2004).

3.3.2 Varying Markov time models for higher-order networks

The fossil record does not have a unique and optimal level of description, but rather multiple levels representing different scales in the organization of life. Flow-based community detection approaches usually assumes one-step dynamics on the links. Instead, modeling network dynamics at different Markov times, an intrinsic time scale of the network (Delvenne et al. 2010), captures the larger-scale modular structure at different resolutions. Varying Markov time models can be used to explore structure and dynamics in both first- and higher-order networks (Kheirkhahzadeh et al. 2016). These models are especially useful when the modular structure of an empirical network does not show a hierarchical organization (e.g., Penn-Clarke and Harper 2020) or when a two-level solution is chosen in the Map Equation framework. Exploring varying time scales to reveal fine to coarse partitions allows to connect time scales of the dynamics to the structural scales present in the network (Lambiotte et al. 2014). In networks representing the deep-time fossil record, the specific relationship between the Markov time and time scales of the evolution of the Earth-Life system has not been explored.

In the Map Equation framework, the discrete (Kheirkhahzadeh et al. 2016) and continuous (Schaub et al. 2012) time evolution of a Markov process on the network is defined by the parameter Markov time. Intuitively, when using shorter Markov times (<1) the average transition rate of a random walk is lower than the encoding rate (see Map Equation section), the same node is encoded multiple times in its trajectory, and smaller modules are delineated. In contrast, when using longer Markov times (>1), the average transition rate is higher than the encoding rate, not every node on its trajectory is

encoded, and larger modules are delineated (Kheirkhahzadeh et al. 2016). In practice, using a parameter Markov time = 2, we explicitly explore the two-step dynamics on the network. In the second example, we illustrate varying Markov time models using a case study on the mid-Devonian biogeography of the brachiopods where one-step dynamics on the links (Markov time = 1) in a relatively small bipartite network does not capture a hierarchical nested organization.

3.3.3 Hypergraphs

Hypergraphs are network models used to study complex systems in which an arbitrary number of its components can interact. These so-called multigroup interactions, represented through hyperedges connecting all the nodes involved in the interaction, differ from binary contacts represented through links in conventional network models (Carletti and Fanelli 2022). Recently, Eriksson et al. (2021) provided the first hypergraph representation of data derived from the fossil record. They used stage-level global occurrences of the benthic marine animals from Cambrian (541 MY) to Cretaceous (66 MY), sourced from the PaleoDB, to create a hypergraph where physical nodes are fossil taxa linked through weighted hyperedges that connect all taxa occurring at each stage. In this representation, taxa occurring in multiple geological stages form connections between hyperedges, which allows to capture temporal structure and dynamics. This network explicitly models temporal constraints in the underlying paleontological data as hyperedges and can be used to evaluate large-scale temporal patterns in the fossil record when spatial relationships are not available (Figure 4).

Recently, Eriksson et al. (2022) derived unipartite, bipartite, and multilayer network representations of hypergraph flows and evaluated how the choice of network representation and specific random-walk model impact the number, size, depth, and overlap of multilevel communities. To create each different network models, they represented hyperedges as nodes in the bipartite network, projected the bipartite flow to create a unipartite flow representation (i.e., two-step dynamics obtained with a Markov time = 2) (Kheirkhahzadeh et al. 2016), and created state nodes for each hyperedge to which a node belongs to construct the multilayer network, showing that the Map Equation framework can be used to model multigroup interactions. Overall, their results showed the advantages of using multilayer network representations of data derived from the fossil record over bipartite and unipartite representations to quantify macroevolutionary patterns, with different random walk models (e.g., including and excluding self-links) providing similar solutions.

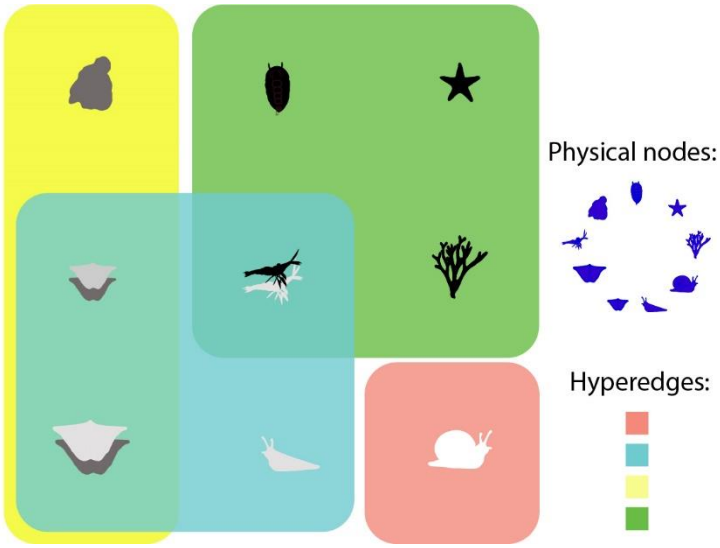


Figure 4. Schematic diagram illustrating a hypergraph representing temporal occurrence data. In this diagram, physical nodes are fossil taxa connected through hyperlinks. Hyperlinks are depicted with different colors and connect taxa from the same time interval. Taxa recorded in different time intervals are represented by state nodes, depicted with different shades of grey.

4 Visualizing higher-order multiscale structure in networks

Network visualizations are essential tools for understanding the modular structure and dynamics of higher-order networks describing geohistorical data. However, standard graphic tools developed to represent first-order interactions among network components fail to capture higher-order and multiscale community structure and depend on the arbitrary scale of analysis (Peixoto and Rosvall 2017; Perri and Scholtes 2020). To overcome these limitations, the Map Equation framework provides graphic tools for mapping higher-order and multiscale community structure in network partitions. They are freely available as a client-side web application at https://www.mapequation.org.

4.1 Network Navigator

Higher-order networks representing geohistorical systems are usually complex with hierarchical and multilevel modular structures. The Network Navigator tool was developed to explore such complex structures in real networks. This visualization tool creates interactive maps of hierarchical network partitions with aggregated inter-module links. The navigator draws bidirectional links with length inversely proportional to flow and width and color saturation proportional to flow. As a result, weakly connected modules are further apart and are connected through narrower links. The Network Navigator draws modules as circles with an area proportional to flow volume and border thickness proportional to exiting flow (Figure 5). It shows the leaf nodes assigned to the module at the lowest level. The Network Navigator was recently used to explore the multiscale organization of the Phanerozoic benthic marine faunas, highlighting how the large-scale evolutionary faunas are built up from lower-scale biogeographic entities (Rojas et al., 2021).

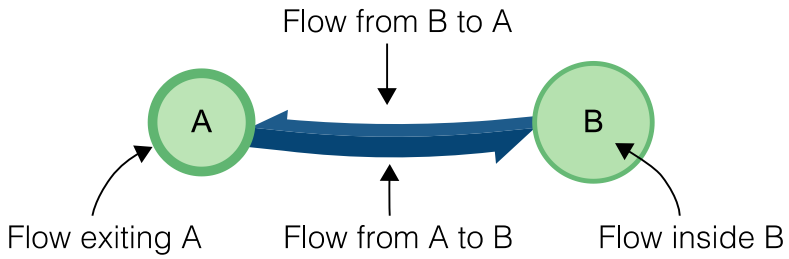


Figure 5. Aggregated inter-module links used by the Network Navigator. Modules and nodes as circles with areas proportional to flow and border widths proportional to exiting flow. Bidirectional links with width proportional to inter-module flow

4.2 Alluvial diagrams

When comparing network partitions, researchers are interested in comparing changes in the overall modular structure rather than link structure. Alluvial diagrams are visualization tools that highlight changes in modular structure across different network partitions, including, for instance, optimized, bootstrapped, suboptimal, and planted solutions. In the paleobiology literature, alluvial diagrams have been used primarily to compare alternative partitions obtained from bootstrapping (Rojas et al., 2021) and sensitivity analysis (Pilotto et al., 2022). Recently, alluvial diagrams have been used to compare solutions representing alternative biostratigraphy models for the late Permian-mid Triassic

Beaufort Group in South Africa (Viglietti et al., 2022). Because it is impractical and typically unnecessary to represent all nodes, an alluvial diagram highlights changes in the module assignment of the most important nodes across partitions. In the Map Equation framework, the flow volume determines node importance and is calculated when Infomap searches for the optimal partition (Rosvall and Bergstrom, 2008).

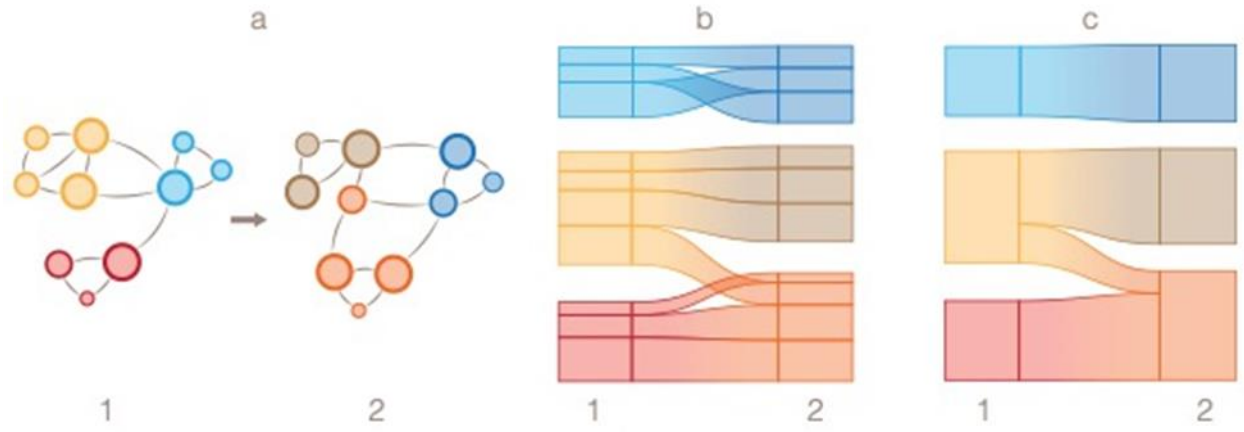


Figure 6. Visualizing change in network partitions using alluvial diagrams. To compare networks with the same sets of nodes, we assemble at least two network partitions (a). Then, we group nodes in the same module in stacked bars with height proportional to the node's flow volume and connect corresponding nodes with streamlines (b). Finally, to highlight how the partitions change, we aggregate the nodes into modules (c). To compare additional network partitions, we add more stacks of bars to the right and repeat the procedure b-c.

The alluvial diagram is constructed by grouping nodes with the same module assignment to simplify and highlight the change in network partitions (Figure 6). These groups of nodes are sized proportional to their flow volume, which might change between partitions. When drawing the diagram, we create streamlines between each node that occurs in both networks, with height proportional to the node's flow. For large networks, drawing streamlines between nodes can obscure the large-scale pattern. Instead, to highlight change between modules, we aggregate nodes and streamlines between nodes that stay together. The alluvial diagram showing the macroevolutionary patterns obtained from different network representations of a large paleontological dataset (Figure 3) illustrates how this graphic tool helps to better understand changes in the modular structure between partitions.

5 Robustness evaluation

5.1 Identifying significant assignments in network partitions

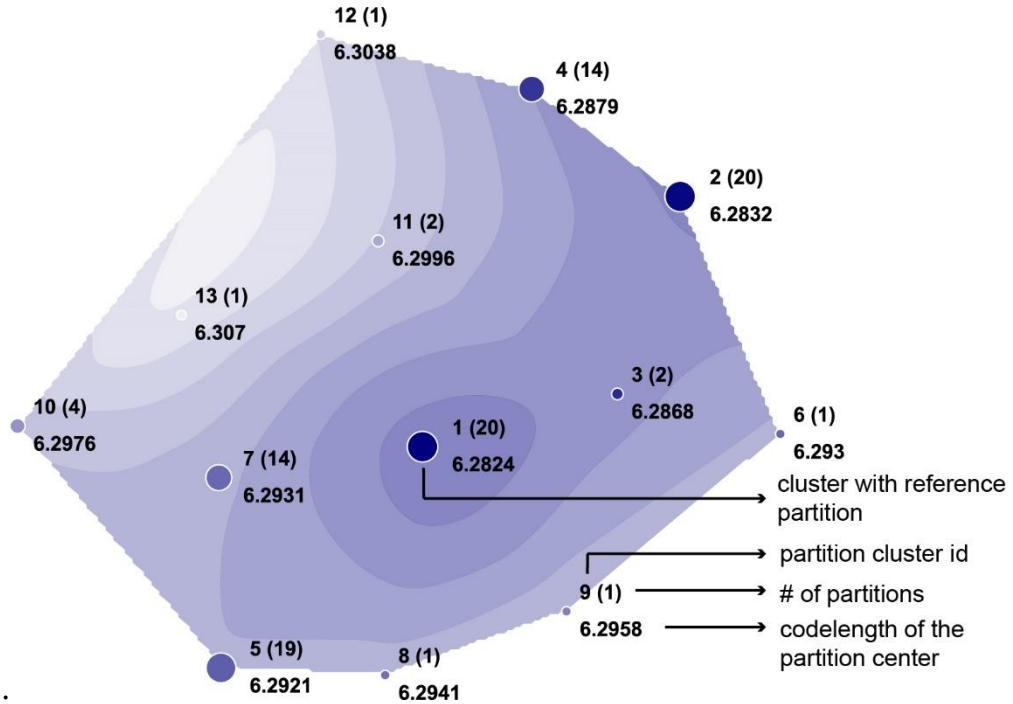
There are different approaches to measure the significance of network partitions. The general idea behind this task is to examine bootstrap replicates to identify which aspects of the modular description of the assembled network are best supported by the underlying data. Overall, those features that occur across a large proportion of the bootstrap replicates are robust. The Map Equation framework includes a significance clustering method that identifies nodes that are significantly assigned to a module in a reference partition. For a given module in the reference partition, the largest subset of nodes that are clustered together in at least 95% of all bootstrap partitions represent its significant core. This significance clustering approach allows to distinguish the modules that stand alone. In practice, it considers a module to be significant if its significant core is clustered with no

other significant core in at least 95% of all bootstrap partitions (Rosvall and Bergstrom, 2010). This approach has been used to distinguish gradual from abrupt biotic transitions in the fossil record (Pilotto et al., 2022).

Alternatively, having a set of bootstrap modular descriptions, the significance of the modules in the reference partition as well as any other aspect of the modular description can be evaluated using for instance distance-based similarity measures such as the Jaccard index (e.g., Calatayud et al. 2019) and measures built upon concepts from information theory (Vinh et al., 2009). Bootstrap replicates can be generated from the underlying data by considering a discrete distribution if the geohistorical data represent fossil counts, and also a truncated distribution to avoid false negatives (Rojas et al. 2021). This approach to create bootstrap replicates of the original network aims to imitate the process of collecting fossil through which paleontologists produce the physical fossil record. Overall, irrespective of the procedure employed, standard research practices should include the identification of significant assignments in network partitions.

5.2 Exploring alternative solutions

Finding the best partition of a network is in general a non-convex optimization problem and the practitioner therefore needs to consider the possibility of multiple solutions. To this end, methods exploring the solutions and their quality have been developed. One such example is the approach to explore the solution landscape, which determine the minimum number of searches required by search algorithm Infomap to obtain a good network partition given a user-specified resolution length that sets the accepted accuracy (Calatayud et al. 2019). We illustrate this approach with the multilayer network representing the Phanerozoic fossil record of the benthic marine faunas, where alternative partitions are embedded using a dimension reduction technique (McInnes et al. 2018) (Figure 7), and the distance between partitions is calculated using a weighted version of the Jaccard distance. In this example, the codelength varies between the partition clusters, with the best partition in the middle cluster, but in practice these variations have a very limited impact on the large-scale patterns with only a few nodes alternating between modules.



.

Figure 7. Quality of alternative partitions of the multilayer network representing the Phanerozoic fossil record of the benthic marine faunas (Rojas et al. 2021) mapped in a two-dimensional space. Circles represent clusters of network partitions, located at the cluster center, with size proportional to the number of partitions grouped into the cluster. Map isolines are constructed using the Jaccard distance between partitions. Despite differences in their quality (codelength), all partition clusters identified at the selected scale show a modular structure with four modules representing the Phanerozoic evolutionary faunas. At the selected scale, there is not a cluster of solutions representing a three-tier model (Sepkoski 1981).

Although the approach to explore the solution landscape was developed for practitioners to be confident that they find an optimal modular description when using the Map Equation framework, it highlights the importance of looking at nearly optimal or sub-optimal partitions when dealing with real geohistorical systems. Exploring such solutions rather than focusing on finding the optimum solution can provide a better understanding of complex patterns. For instance, the extinction at the end of the Cretaceous, despite being a significant and abrupt event, is not captured at the resolution of the landscape illustrated in figure 7. To find a solution showing a Cretaceous-Neogene global transition at the higher hierarchical level, we should look at less optimal solutions (see Rojas et al. 2021). Overall, by looking at a range of solutions ordered from highest to lowest quality, we can get some insights about the relative importance of the different events shaping the Phanerozoic history of life.

6 Case studies on the fossil record

6.1 Delineating litho-biofacies by using multilayer networks

Understanding how the distribution of organisms along environmental gradients change through time is a primary research area in paleobiology (Patzkowsky and Holland 2012). Indirect ordination techniques applied to species occurrence data have been successfully employed to recover environmental gradients in the sedimentary record. In general, environmental gradients are interpreted from mapping taxa and/or samples, coded by external factors (e.g., life habit for taxa and depositional environment for samples), into the reduced ordination space (Holland and Patzkowsky 2004). Here we provide a multilayer network analysis that combines taxon abundance and samples attributes into the modelling. Specifically, we conceptualize the biosedimentary record as a complex system with a multilayered structure, by creating a network representation with two layers, one describing taxonomic composition and the other sedimentologically-defined relationships between sampling units (Figure 8a). The underlying data were obtained from a basin-scale study carried out on Late Ordovician outcrops in central Kentucky (Holland and Patzkowsky 2004)

Physical nodes in this two-layered network represent sampling units, taxa, and sedimentological groups. Layers in this representation capture fossil assemblages (biofacies) and sedimentological groups (lithofacies) in the geohistorical record, independently. However, the multilayer network reveals the so-called litho-bio facies (Figure 8b). This higher-order approach provides a direct interpretation of the gradients obtained via ordination analysis in two ways, delineating modules that comprise strongly connected taxa, samples, and sedimentological descriptors, and partitioning the gradient into discrete regions (Figure 8c). Although the sedimentological information underlying this case study is relatively simple, our network model can be easily extended to represent geochemical, taphonomic, and any other complex relationship between sampling units, as well as relationship between taxa (i.e., ecology, body size). Although beds are the fundamental units of both stratigraphy and paleontology (Patzkowsky and Holland 2012), sampling units in this multilayer framework can

represent stratigraphic units of any scale. In practice, depending on the stratigraphic resolution of the underlying geohistorical data, this multilayer network model can be used to capture multitype relationships between samples, beds, members, or broader units.

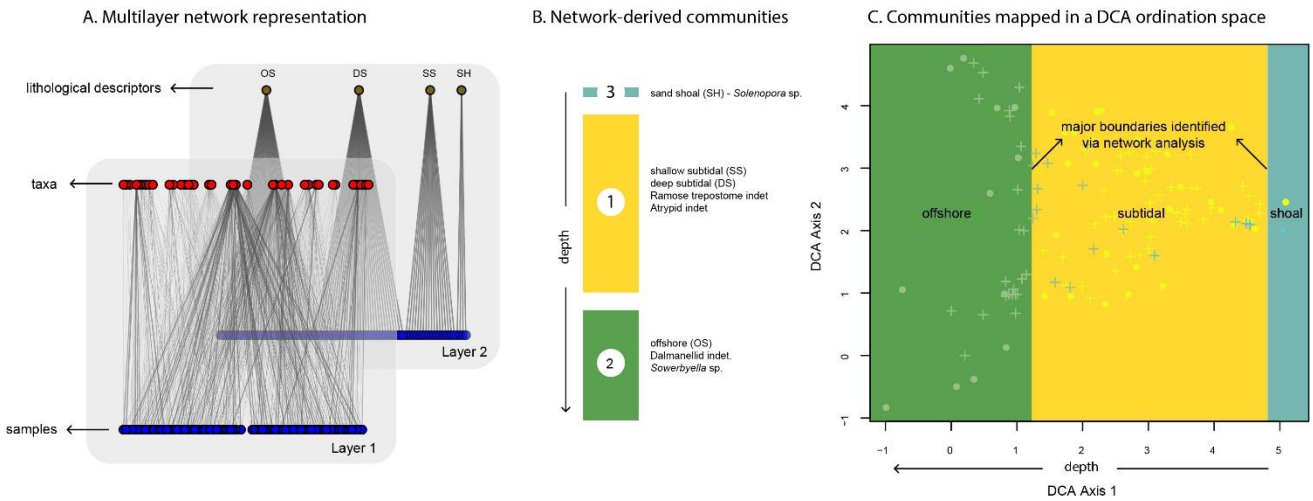


Figure 8. A. A multilayer network representing litho-biofacies from Late Ordovician outcrops in central Kentucky. One layer describes the biotic component (samples × taxa matrix) and the other describes the abiotic component (samples × environments matrix) of the geohistorical record. **B**. Litho-biofacies delineated via community detection using the Map Equation framework. Modules in the multilayer solution includes taxa, samples, and environments, and can be directly interpreted as litho-biofacies. **C**. Detrended Component Analysis (DCA) on the samples × taxa matrix. Network modules mapped in a DCA ordination space indicating their distribution along the depth water gradient. Background colored based on the module affiliation. Data from Holland and Patzkowsky (2004).

6.2 Validating mid-Devonian marine bioregionalization using varying Markov time models

Research employing network-based approaches to describe biogeography in the fossil record is overwhelmingly focused on describing one-step dynamics (Markov time = 1) in relatively small bipartite networks derived from the relatively limited fossil data (e.g., Penn-Clarke and Harper, 2020; Ye et al., 2021), to reveal continental to global scale marine bioregions. Although first-order network representations of fossil occurrences have been shown to capture a biogeographic signal at some geological stages in the Phanerozoic (Rojas et al. 2017; Kocsis et al. 2018), these studies are unable to identify transition zones, provide a single-scale description of the bioregions and obscure larger-scale patterns. They ignore that bioregions do not have a unique level of description but multiple levels reflecting the complex spatial structuring of biodiversity.

Here we describe how to use varying Markov time models through the Map Equation framework to overcome some of the limitations of the standard models that are currently used in paleobiology research. This case study is based on a standard bipartite network that is relevant for Devonian biogeography, but the approach can be applied directly to higher-order networks. We use varying Markov time models to examine the bioregionalization of the Middle Devonian Brachiopods from the Old-World Realm (Penn-Clarke and Harper 2020). This approach allows to reveal the larger-scale biogeographic structure at different resolutions (Kheirkhahzadeh et al., 2016). Results provide new insights into the biogeographic affinities of the brachiopod faunas from Southern Peru, which

remains as an open question (Figure 9). However, this case study shows that we can reveal the complex spatial organization of the biodiversity through varying Markov time models.

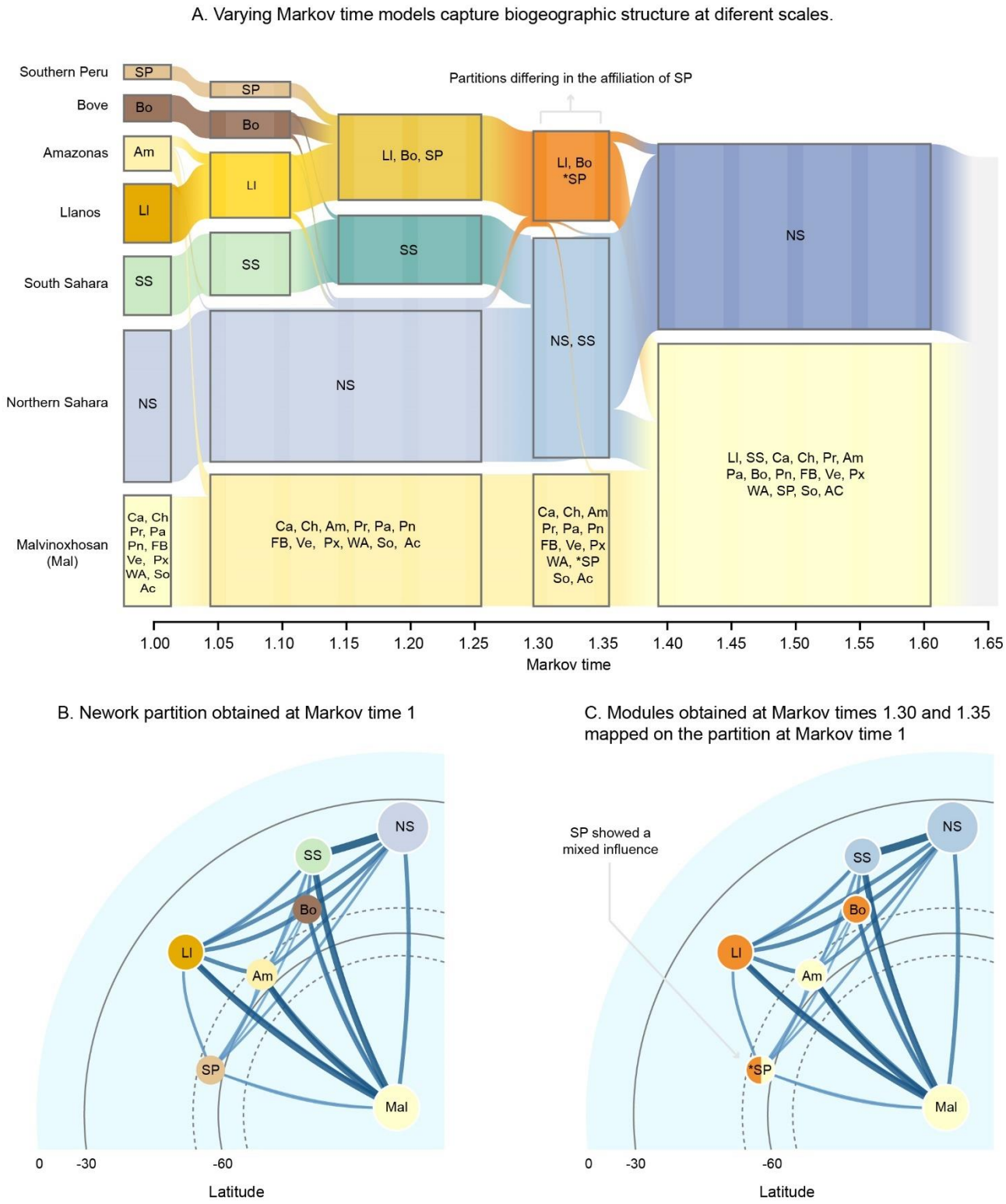


Figure 9. Bioregionalization of the Middle Devonian Old World Realm. **A**. Varying Markov time models on the biogeographic network constructed from the brachiopod data in Penn-Clarke and Harper (2020). Network partitions at different Markov times reveal the larger-scale biogeographic structures obtained at different resolutions. **B**. Network partition obtained at the Markov time 1 (7 modules). Circles represent the seven modules delineated when exploring the one-step dynamics on the assembled network. **C**. Modules obtained from partitions at Markov times 1.30 and 1.35 (3 modules) mapped on those obtained at the Markov time 1. These two partitions differ in the affiliation of the Southern Peru locality, placed alternatively into the modules representing higher (white) and lower (orange) latitudes. Overall, coarser partitions contain fewer clusters as the Markov time increases.

7 Future research: Benchmark networks for research in paleobiology

Different network representations capture different aspects of the geohistorical records and may exhibit different modular patterns. Higher-order representations better capture the spatiotemporal constraints on the information paths underlying geohistorical data, providing more accurate descriptions of paleontological patterns: Employing the Map Equation framework for higher-order networks also improve reproducibility and communicability. Establishing higher-order benchmark networks from well-established benchmark data enables researchers to compare results from different studies and methods. We believe these methodological efforts lay the ground for a fertile research direction in the emergent field of network paleobiology.

8 Conflict of Interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

9 Author Contributions

A.B.: conceptualization, data curation, formal analysis, methodology, visualization, writing—original draft, writing, editing; A.E.: visualization, writing; M.N.: visualization, witing, editing; D.E.: writing; M.R.: writing, editing.

10 Funding

A.R. was supported by the Kone Fundation funded project *Comparing evolutionary processes in nature and society* (project number 202007064). A.E. and M.R. were supported by the Swedish Foundation for Strategic Research, Grant No.\ SB16-0089. D.E. and M.R. were supported by the Swedish Research Council, Grant No.\ 2016-00796.

Acknowledgments

We thank the contributors to the Paleobiology Database who collected data. AR thanks Steve Holland for providing data to delineate Middle Upper Ordovician biofacies.

11 References

Aldecoa, R., and Marín, I. (2013). Exploring the limits of community detection strategies in complex networks. Sci Rep 3, 2216. doi: 10.1038/srep02216.

Allmon, W. D., Dietl, G. P., Hendricks, J. R., and Ross, R. M. (2018). “Bridging the two fossil records: Paleontology’s ‘big data’ future resides in museum collections,” in Museums at the Forefront of the History and Philosophy of Geology: History Made, History in the Making (Geological Society of America). doi: 10.1130/2018.2535(03).

Angel, S., Parent, J., Civco, D. L., Blei, A., and Potere, D. (2011). The dimensions of global urban expansion: Estimates and projections for all countries, 2000–2050. Progress in Planning 75, 53–107. doi: 10.1016/j.progress.2011.04.001.

Aslak, U., Rosvall, M., and Lehmann, S. (2018). Constrained information flows in temporal networks reveal intermittent communities. Phys. Rev. E 97, 062312. doi: 10.1103/PhysRevE.97.062312.

Barabási, A.-L., and Pósfai, M. (2016). Network science. Cambridge, United Kingdom: Cambridge University Press.

Barber, J. N., Nicholson, L. C., Woods, L. C., Judd, L. M., Sezmis, A. L., Hawkey, J., et al. (2022). Species interactions constrain adaptation and preserve ecological stability in an experimental microbial community. ISME J 16, 1442–1452. doi: 10.1038/s41396-022-01191-1.

Batty, M., and Longley, P. A. (1994). Fractal cities: a geometry of form and function. London: Acad. Press.

Batty, M., and Marshall, S. (2009). Centenary paper: The evolution of cities: Geddes, Abercrombie and the new physicalism. Town Planning Review 80, 551–574. doi: 10.3828/tpr.2009.12.

Brandes, U., Robins, G., McCRANIE, A., and Wasserman, S. (2013). What is network science? Netw. sci. 1, 1–15. doi: 10.1017/nws.2013.2.

Buckland, P. I. (2014). The Bugs Coleopteran Ecology Package (BugsCEP) database: 1000 sites and half a million fossils later. Quaternary International 341, 272–282. doi: 10.1016/j.quaint.2014.01.030.

Carletti, T., and Fanelli, D. (2022). “Pattern Formation on Hypergraphs,” in Higher-Order Systems Understanding Complex Systems., eds. F. Battiston and G. Petri (Cham: Springer International Publishing), 163–180. doi: 10.1007/978-3-030-91374-8_5.

Clark, P. J., and Evans, F. C. (1954). Distance to Nearest Neighbor as a Measure of Spatial Relationships in Populations. Ecology 35, 445–453. doi: 10.2307/1931034.

Council, N. R. (2005). The Geological Record of Ecological Dynamics: Understanding the Biotic Effects of Future Environmental Change. Washington, DC: The National Academies Press doi: 10.17226/11209.

Csárdi, G., and Nepusz, T. (2006). The igraph software package for complex network research. in.

De Domenico, M., Lancichinetti, A., Arenas, A., and Rosvall, M. (2015a). Identifying Modular Flows on Multilayer Networks Reveals Highly Overlapping Organization in Interconnected Systems. Phys. Rev. X 5, 011027. doi: 10.1103/PhysRevX.5.011027.

De Domenico, M., Lancichinetti, A., Arenas, A., and Rosvall, M. (2015b). Identifying Modular Flows on Multilayer Networks Reveals Highly Overlapping Organization in Interconnected Systems. Phys. Rev. X 5, 011027. doi: 10.1103/PhysRevX.5.011027.

De Domenico, M., Lancichinetti, A., Arenas, A., and Rosvall, M. (2015c). Identifying Modular Flows on Multilayer Networks Reveals Highly Overlapping Organization in Interconnected Systems. Phys. Rev. X 5, 011027. doi: 10.1103/PhysRevX.5.011027.

Delvenne, J.-C., Yaliraki, S. N., and Barahona, M. (2010). Stability of graph communities across time scales. Proc. Natl. Acad. Sci. U.S.A. 107, 12755–12760. doi: 10.1073/pnas.0903215107.

Dunhill, A. M., Bestwick, J., Narey, H., and Sciberras, J. (2016). Dinosaur biogeographical structure and Mesozoic continental fragmentation: a network-based approach. Journal of Biogeography. doi: 10.1111/jbi.12766.

Edler, D., Bohlin, L., and Rosvall, and (2017). Mapping Higher-Order Network Flows in Memory and Multilayer Networks with Infomap. Algorithms 10, 112. doi: 10.3390/a10040112.

Eriksson, A., Carletti, T., Lambiotte, R., Rojas, A., and Rosvall, M. (2022). “Flow-Based Community Detection in Hypergraphs,” in Higher-Order Systems Understanding Complex Systems., eds. F. Battiston and G. Petri (Cham: Springer International Publishing), 141–161. doi: 10.1007/978-3-030-91374-8_4.

Eriksson, A., Edler, D., Rojas, A., de Domenico, M., and Rosvall, M. (2021a). How choosing random-walk model and network representation matters for flow-based community detection in hypergraphs. Commun Phys 4, 133. doi: 10.1038/s42005-021-00634-z.

Eriksson, A., Edler, D., Rojas, A., and Rosvall, M. (2021b). Mapping flows on hypergraphs. Communications Physics.

Foote, M., Crampton, J. S., Beu, A. G., Marshall, B. A., Cooper, R. A., Maxwell, P. A., et al. (2007). Rise and Fall of Species Occupancy in Cenozoic Fossil Mollusks. Science 318, 1131–1134. doi: 10.1126/science.1146303.

Ghasemian, A., Hosseinmardi, H., and Clauset, A. (2019). Evaluating Overfit and Underfit in Models of Network Community Structure. IEEE Trans. Knowl. Data Eng., 1–1. doi: 10.1109/TKDE.2019.2911585.

Hansen, M. H., and Yu, B. (2001). Model Selection and the Principle of Minimum Description Length. Journal of the American Statistical Association 96, 746–774. doi: 10.1198/016214501753168398.

HOFFMEISTER, A. P., and KOWALEWSKI, M. (2001). Spatial and Environmental Variation in the Fossil Record of Drilling Predation: A Case Study from the Miocene of Central Europe. PALAIOS 16, 566–579. doi: 10.1669/0883-1351(2001)016<0566:SAEVIT>2.0.CO;2.

Holland, S. M. (2016). The non-uniformity of fossil preservation. Phil. Trans. R. Soc. B 371, 20150130. doi: 10.1098/rstb.2015.0130.

Holland, S. M., and Patzkowsky, M. E. (2004). Ecosystem Structure and Stability: Middle Upper Ordovician of Central Kentucky, USA. PALAIOS 19, 316–331. doi: 10.1669/0883-1351(2004)019<0316:ESASMU>2.0.CO;2.

Huang, B., Zhan, R.-B., and Wang, G.-X. (2016). Recovery brachiopod associations from the lower Silurian of South China and their paleoecological implications. Can. J. Earth Sci. 53, 674–679. doi: 10.1139/cjes-2015-0193.

Hunt, G. (2010). Evolution in Fossil Lineages: Paleontology and The Origin of Species. The American Naturalist 176, S61–S76. doi: 10.1086/657057.

Illian, J. ed. (2008). Statistical analysis and modelling of spatial point patterns. Chichester, England ; Hoboken, NJ: John Wiley.

Jablonski, D. (2017). Approaches to Macroevolution: 1. General Concepts and Origin of Variation. Evol Biol 44, 427–450. doi: 10.1007/s11692-017-9420-0.

Johnson, I. (1984). "Cell Frequency Recording and Analysis of Artifact Distributions," in Intrasite Spatial Analysis in Archaeology, ed. H. J. Hietala (Cambridge: Cambridge University Press), 75–96.

Kheirkhahzadeh, M., Lancichinetti, A., and Rosvall, M. (2016). Efficient community detection of network flows for varying Markov times and bipartite networks. Phys. Rev. E 93, 032309. doi: 10.1103/PhysRevE.93.032309.

Kidwell, S. M., and Holland, S. M. (2002). The Quality of the Fossil Record: Implications for Evolutionary Analyses. Annual Review of Ecology and Systematics 33, 561–588. doi: 10.1146/annurev.ecolsys.33.030602.152151.

Kiel, S. (2017a). Reply to Smith et al .: Network analysis reveals connectivity patterns in the continuum of reducing ecosystems. Proc. R. Soc. B. 284, 20171644. doi: 10.1098/rspb.2017.1644.

Kiel, S. (2017b). Using network analysis to trace the evolution of biogeography through geologic time: A case study. Geology, G38877.1. doi: 10.1130/G38877.1.

Kocsis, Á. T., Reddin, C. J., and Kiessling, W. (2018). The biogeographical imprint of mass extinctions. Proc. R. Soc. B. 285, 20180232. doi: 10.1098/rspb.2018.0232.

Kocsis, Á. T., Reddin, C. J., Scotese, C. R., Valdes, P. J., and Kiessling, W. (2021). Increase in marine provinciality over the last 250 million years governed more by climate change than plate tectonics. Proc. R. Soc. B. 288, 20211342. doi: 10.1098/rspb.2021.1342.

Kowalewski, M., and Novack-Gottshall, P. (2010). Resampling Methods in Paleontology. Paleontol. Soc. pap. 16, 19–54. doi: 10.1017/S1089332600001807.

Lambiotte, R., Delvenne, J.-C., and Barahona, M. (2014). Random Walks, Markov Processes and the Multiscale Modular Organization of Complex Networks. IEEE Trans. Netw. Sci. Eng. 1, 76–90. doi: 10.1109/TNSE.2015.2391998.

Lambiotte, R., and Rosvall, M. (2012a). Ranking and clustering of nodes in networks with smart teleportation. Phys. Rev. E 85, 056107. doi: 10.1103/PhysRevE.85.056107.

Lambiotte, R., and Rosvall, M. (2012b). Ranking and clustering of nodes in networks with smart teleportation. Phys. Rev. E 85, 056107. doi: 10.1103/PhysRevE.85.056107.

Lambiotte, R., and Scholtes, I. (2018). Beyond Graph Mining: Higher-Order Data Analytics for Temporal Network Data. in.

Lancichinetti, A., and Fortunato, S. (2009). Community detection algorithms: A comparative analysis. Phys. Rev. E 80, 056117. doi: 10.1103/PhysRevE.80.056117.

MacLeod, M. (2018). What makes interdisciplinarity difficult? Some consequences of domain specificity in interdisciplinary practice. Synthese 195, 697–720. doi: 10.1007/s11229-016-1236-4.

Mucha, P. J., Richardson, T., Macon, K., Porter, M. A., and Onnela, J.-P. (2010). Community Structure in Time-Dependent, Multiscale, and Multiplex Networks. Science 328, 876–878. doi: 10.1126/science.1184819.

Muscente, A. D., Bykova, N., Boag, T. H., Buatois, L. A., Mángano, M. G., Eleish, A., et al. (2019). Ediacaran biozones identified with network analysis provide evidence for pulsed extinctions of early complex life. Nat Commun 10, 911. doi: 10.1038/s41467-019-08837-3.

Newman, M. E. J. (2003). The Structure and Function of Complex Networks. SIAM Rev. 45, 167–256. doi: 10.1137/S003614450342480.

Newman, M. E. J. (2011). Resource Letter CS–1: Complex Systems. American Journal of Physics 79, 800–810. doi: 10.1119/1.3590372.

Patzkowsky, M. E., and Holland, S. M. (2012a). Stratigraphic paleobiology: understanding the distribution of fossil taxa in time and space. Chicago ; London: The University of Chicago Press.

Patzkowsky, M. E., and Holland, S. M. (2012b). Stratigraphic paleobiology: understanding the distribution of fossil taxa in time and space. Chicago ; London: The University of Chicago Press.

Peixoto, T. P., and Rosvall, M. (2017). Modelling sequences and temporal networks with dynamic community structures. Nat Commun 8, 582. doi: 10.1038/s41467-017-00148-9.

Penn-Clarke, C. R., and Harper, D. A. T. (2020). Early−Middle Devonian brachiopod provincialism and bioregionalization at high latitudes: A case study from southwestern Gondwana. GSA Bulletin. doi: 10.1130/B35670.1.

Perri, V., and Scholtes, I. (2020). “HOTVis: Higher-Order Time-Aware Visualisation of Dynamic Graphs,” in Graph Drawing and Network Visualization Lecture Notes in Computer Science., eds. D. Auber and P. Valtr (Cham: Springer International Publishing), 99–114. doi: 10.1007/978-3-030-68766-3_8.

Petrovic, L. V., and Scholtes, I. (2022). Learning the Markov Order of Paths in Graphs. in Proceedings of the ACM Web Conference 2022 (Virtual Event, Lyon France: ACM), 1559–1569. doi: 10.1145/3485447.3512091.

Pilotto, F., Rojas, A., and Buckland, P. I. (2022). Late Holocene anthropogenic landscape change in northwestern Europe impacted insect biodiversity as much as climate change did after the last Ice Age. Proc. R. Soc. B. 289, 20212734. doi: 10.1098/rspb.2021.2734.

Rissanen, J. (1978). Modeling by shortest data description. Automatica 14, 465–471. doi: 10.1016/0005-1098(78)90005-5.

Rojas, A., Calatayud, J., Kowalewski, M., Neuman, M., and Rosvall, M. (2021). A multiscale view of the Phanerozoic fossil record reveals the three major biotic transitions. Commun Biol 4, 309. doi: 10.1038/s42003-021-01805-y.

Rojas, A., Gracia, A., Hernández-Ávila, I., Patarroyo, P., and Kowalewski, M. (2022). Occurrence of the brachiopod Tichosina in deep-sea coral bottoms of the Caribbean sea and its paleoenvironmental implications. Bulletin of the Florida Museum of Natural History 59, 1–15. doi: 10.1101/2020.06.24.168658.

Rojas, A., Patarroyo, P., Mao, L., Bengtson, P., and Kowalewski, M. (2017). Global biogeography of Albian ammonoids: A network-based approach. Geology 45, 659–662. doi: 10.1130/G38944.1.

Roopnarine, P. (2010a). Networks, Extinction and Paleocommunity Food Webs. Nat Prec. doi: 10.1038/npre.2010.4433.2.

Roopnarine, P. D. (2010b). Networks, Extinction and Paleocommunity Food Webs. Paleontol. Soc. pap. 16, 143–161. doi: 10.1017/S1089332600001856.

Rosvall, M., Esquivel, A. V., Lancichinetti, A., West, J. D., and Lambiotte, R. (2014). Memory in network flows and its effects on spreading dynamics and community detection. Nat Commun 5, 4630. doi: 10.1038/ncomms5630.

Rzymski, C., Tresoldi, T., Greenhill, S. J., Wu, M.-S., Schweikhard, N. E., Koptjevskaja-Tamm, M., et al. (2020). The Database of Cross-Linguistic Colexifications, reproducible analysis of cross-linguistic polysemies. Sci Data 7, 13. doi: 10.1038/s41597-019-0341-x.

Scarponi, D., and Kowalewski, M. (2007). Sequence stratigraphic anatomy of diversity patterns: Late Quaternary benthic mollusks of The Po Plain, Italy. Palaios 22, 296–305. doi: 10.2110/palo.2005.p05-020r.

Scarponi, D., Rojas, A., Nawrot, R., Cheli, A., and Kowalewski, M. (2022). “Assessing biotic response to anthropogenic forcing using mollusk assemblages from the Po-Adriatic System (Italy),” in Conservation Palaeobiology of Marine Ecosystems: Concepts and Applications (The Geological Society Special Publications). In press

Schaub, M. T., Delvenne, J.-C., Rosvall, M., and Lambiotte, R. (2017). The many facets of community detection in complex networks. Appl Netw Sci 2, 4. doi: 10.1007/s41109-017-0023-6.

Schaub, M. T., Lambiotte, R., and Barahona, M. (2012). Encoding dynamics for multiscale community detection: Markov time sweeping for the map equation. Phys. Rev. E 86, 026112. doi: 10.1103/PhysRevE.86.026112.

Sepkoski, D. (2013). Towards “A Natural History of Data”: Evolving Practices and Epistemologies of Data in Paleontology, 1800–2000. J Hist Biol 46, 401–444. doi: 10.1007/s10739-012-9336-6.

Sepkoski, J. J. (1981). A factor analytic description of the Phanerozoic marine fossil record. Paleobiology 7, 36–53. doi: 10.1017/S0094837300003778.

Smiljanić, J., Blöcker, C., Edler, D., and Rosvall, M. (2021). Mapping flows on weighted and directed networks with incomplete observations. Journal of Complex Networks 9, cnab044. doi: 10.1093/comnet/cnab044.

Smiljanić, J., Edler, D., and Rosvall, M. (2020). Mapping flows on sparse networks with missing links. Phys. Rev. E 102, 012302. doi: 10.1103/PhysRevE.102.012302.

Smith, S. M., Sprain, C. J., Clemens, W. A., Lofgren, D. L., Renne, P. R., and Wilson, G. P. (2018). Early mammalian recovery after the end-Cretaceous mass extinction: A high-resolution view from McGuire Creek area, Montana, USA. GSA Bulletin. doi: 10.1130/B31926.1.

Stier, A. J., Schertz, K. E., Rim, N. W., Cardenas-Iniguez, C., Lahey, B. B., Bettencourt, L. M. A., et al. (2022). Reply to Huth et al.: Cities are defined by their spatially aggregated socioeconomic networks. Proc. Natl. Acad. Sci. U.S.A. 119, e2119313118. doi: 10.1073/pnas.2119313118.

Tenkanen, H., and Toivonen, T. (2020). Longitudinal spatial dataset on travel times and distances by different travel modes in Helsinki Region. Sci Data 7, 77. doi: 10.1038/s41597-020-0413-y.

Vermeij, G. J. (1977). The Mesozoic marine revolution: evidence from snails, predators and grazers. Paleobiology 3, 245–258. doi: 10.1017/S0094837300005352.

Viglietti, P., Rojas, A., Rosvall, M., Davis, B., and Angielczyk, K. (2022). Network-based biostratigraphy for the late Permian-Triassic Beaufort Group (Karoo Supergroup) in South Africa enhances biozone applicability and stratigraphic correlation. Palaeontology. doi: 10.1111/pala.12622

Vilhena, D. A., and Antonelli, A. (2015). A network approach for identifying and delimiting biogeographical regions. Nat Commun 6, 6848. doi: 10.1038/ncomms7848.

Vinh, N. X., Epps, J., and Bailey, J. (2009). Information theoretic measures for clusterings comparison: is a correction for chance necessary? in Proceedings of the 26th Annual International Conference on Machine Learning - ICML ’09 (Montreal, Quebec, Canada: ACM Press), 1–8. doi: 10.1145/1553374.1553511.

Xu, J., Wickramarathne, T. L., and Chawla, N. V. (2016). Representing higher-order dependencies in networks. Sci. Adv. 2, e1600028. doi: 10.1126/sciadv.1600028.

Yang, Z., Algesheimer, R., and Tessone, C. J. (2016). A Comparative Analysis of Community Detection Algorithms on Artificial Networks. Sci Rep 6, 30750. doi: 10.1038/srep30750.

Ye, F., Shi, G. R., and Bitner, M. A. (2021). Global biogeography of living brachiopods: Bioregionalization patterns and possible controls. PLoS ONE 16, e0259004. doi: 10.1371/journal.pone.0259004.