

RESEARCH ARTICLE

Comparing Structure and Dynamics of Transition Graphs by the Symmetric Difference Metric Over an Edge-Filtration

BELÉN GARCÍA PASCUAL¹, LARS M. SALBU¹, JESSICA RENZ¹,
KONSTANTINOS GIANNAKIS¹, AND IAIN G. JOHNSTON^{1,2}

¹Department of Mathematics, University of Bergen, 5007 Bergen, Norway

²Computational Biology Unit, University of Bergen, 5006 Bergen, Norway

Corresponding author: Iain G. Johnston (Iain.Johnston@uib.no)

This work was supported in part by the Trond Mohn Foundation through the Project “HyperEvol” under Grant TMS2021TMT09, in part by the Centre for Antimicrobial Resistance in Western Norway (CAMRIA) under Grant TMS2020TMT11, and in part by the European Research Council (ERC) through the European Union’s Horizon 2020 Research and Innovation Programme under Grant 805046 (EvoConBio).

ABSTRACT Transition graphs or transition diagrams, describing the rates and probabilities with which a system changes between discrete states, are common throughout the sciences. In many cases, parameterisations of transition graphs are inferred from different datasets, for example in the context of Markov or hidden Markov models. An important task for followup analysis is to find efficient and effective ways to compare transition graphs with different parameterisations. Here, we introduce the Weight-Filtration Comparison Curve (WFCC), an approach by which the differences between two or more parameterisations of a transition graph can be quantified and compared. Borrowing from topological data analysis, the WFCC allows graphs learned from different datasets and/or null models to be systematically compared, and differences in both the fine- and coarse-grained structure and dynamics of transition graphs to be quantitatively assessed. We demonstrate WFCC with simple illustrative cases and real-world cases of transition graphs inferred from global data on the evolution of antimicrobial resistance in different countries, showing how different inferred dynamics, and different levels of uncertainty, are reported by structural aspects of these comparison curves.

INDEX TERMS Transition graphs, Markov models, filtration curves, topological data analysis, antimicrobial resistance.

I. INTRODUCTION

Systems across the natural sciences are often modelled with a discrete ‘state space’, describing every state the system can exist in, and a ‘transition graph’ (also ‘diagram’ or ‘matrix’) describing the system’s propensity to make transitions between these states. This picture is at the root of Markov and hidden Markov models [1], [2], [3], where the future behaviour of a system depends only on its current state and not its history. These models are used across the sciences: some examples from biology [4], for

instance, include modelling dynamics of disease spreading [5], animal movement [6], a wide range of uses in evolutionary biology from DNA sequence evolution [7] to the dynamics of particular evolving systems like mitochondrial DNA [8] or antimicrobial resistance [9], and tracking the dynamics of cancer progression in ‘accumulation modelling’ which explicitly or implicitly captures Markovian transitions between states [10], [11]. A common task in these fields is, given some observations of a system, to learn the parameters of the transition graph that best describe its behaviour. This parameterisation can then be interpreted scientifically, to describe and/or make predictions about the system. Specifically, if the state space is $S = \{S_1, S_2, \dots, S_n\}$, we are

The associate editor coordinating the review of this manuscript and approving it for publication was Orazio Gambino¹.

interested in the matrix θ , where θ_{ij} gives either a transition rate (in continuous time) or transition probability (in discrete time) from S_i to S_j .

Well-established approaches exist for estimating these transition parameters from observations, including estimation directly from observations for Markov chains [12], [13] and the Baum-Welch algorithm for hidden Markov models [14]. A subsequent task is to compare such parameter estimates across different cases. For example, if we estimate transition rates for a cancer progression model from data, do the parameters differ when we look at male versus female patients? Or if we estimate evolutionary transition rates for bacteria acquiring drug resistance, do the parameters differ for bacteria in different geographical regions? Addressing these questions requires a method to quantify the differences between differently-weighted transition graphs.

The general topic of graph comparison has several well-developed branches of literature. Graphs of different topology (such as degrees, connectivity, centrality, etc) can be compared using graph isomorphism approaches [15], set-theory [16], [17], by defining a distance that informs about structural similarities [18], by looking at matching subgraphs [19], [20] or by comparing neighbourhoods of vertices [21], [22]. All these methods can be used to compare unweighted graphs with different underlying structures. We highlight the work in [23] where a metric is defined also to compare directed weighted graphs with the same number of vertices [23, Eq 6], and a pseudometric is presented for directed weighted graphs of different dimension [23, Thm 1]. Recently, methods based on graph filtrations have been proposed to compare graphs [24], [25]. These methods put the focus on the weights of the edges, and they make it possible to consider different types of features (like connected components or edge repetitions) and to localise when differences between graphs appear or disappear.

In comparing transition graphs, however, we are usually more interested in comparing graphs with the same topology but different edge weightings. This task connects to the related field of comparing adjacency matrices (matrices describing the weights graph edges). Approaches here include spectral methods [26], [27], the general graph distance induced by a distance on the adjacent matrices [27, Eqn 20] and graph kernels [26], [27]. Our work is aligned to the literature employing the l_1 -metric [28, Def 3.6], also called the Manhattan distance [29], between two adjacent matrices, as we do use this metric in our method as a final step to give an overall readout of the dissimilarity of a pair of graphs. However, these adjacency matrix methods typically do not give information about when differences between graphs appear or disappear, nor about the magnitude of these local differences.

To facilitate this finer-grained analysis of differences between transition graphs, we here present a method, named Weight-Filtration Comparison Curve (WFCC), to visually and quantitatively compare graphs with the same vertices and

edges, but different weights on their edges. The framework of this comparison method is general and versatile, with the potential to be applied to a wide range of problems. Our method follows a current trend in graph comparisons connected to topological data analysis (TDA): transforming the weighted graph into an edge-filtration (a nested sequence of edge sets) formed by gradually adding edges at different thresholds or filtration values [24], [25]. We then quantify differences between edge sets with the symmetric difference metric, which has been applied to unfiltered graphs in the classical work [30, Sec 1.8.5], and recently, in [31] to unfiltered simplicial complexes that represent mathematical models.

II. METHODS

A. EDGE-FILTRATIONS ON GRAPHS

We begin with a graph $G = (V, E, w)$, with vertices V and edges E , and with each edge $e \in E$ having a corresponding weight $w(e)$.

Definition 1: Let $G = (V, E, w)$ be a weighted graph. The **edge-filtration** on G is the nested sequence of sets $\dots \subseteq K_{t_1} \subseteq K_{t_2} \subseteq K_{t_3} \subseteq \dots$ where $t_i \in \mathbb{R}$ (called the **filtration value** or **threshold**) and

$$K_{t_i} = \{e \in E \mid w(e) \leq t_i\}.$$

The edge-filtration $\{K_t\}_{t \in \mathbb{R}}$ is indeed nested, since if $t \leq s$ and $w(e) \leq t$, then $w(e) \leq s$, so we have inclusions of sets $K_t \subseteq K_s$. This filtration starts capturing the edges with the smallest weights and progressively adds stronger edges with higher weights. Note that the set $\{(V, K_t, w|_{K_t})\}_{t \in \mathbb{R}}$ defines a nested sequence of weighted graphs, also called graph-filtration.

B. SYMMETRIC DIFFERENCE METRIC AS A FUNCTION OF THE FILTRATION VALUE

We next compare edge-filtrations on two graphs by considering the symmetric difference of the filtration sets in each filtration value. The symmetric difference gives the cardinality of the non-common elements in two sets, that is, the number of elements that are in one of the sets but not in the other:

Definition 2: Let S be a set of sets. The **symmetric difference metric** on S is the function $d_\Delta : S \times S \rightarrow \mathbb{R}$ where

$$d_\Delta(K, L) = |K \Delta L| = |(K \setminus L) \cup (L \setminus K)| = |(K \cup L) \setminus (L \cap K)|.$$

Note that the symmetric difference metric is indeed a metric as it is positive definite, symmetric, reflexive and satisfies the triangle inequality, as shown in [30, page 16].

Now if $\{K_t\}$ and $\{L_t\}$ are edge filtrations of the same underlying graph $G = (V, E)$, we can for each filtration value $t \in \mathbb{R}$ consider the symmetric difference $d_\Delta(K_t, L_t)$. In this case we have S equal to the set of all subsets of E . By looking at the symmetric difference at each filtration step we localise

at which stages the weighted graphs differ and for how long those differences persist.

In our method, we plot the symmetric difference metric as a function of t , visualising it as a curve. Then, to give a readout of this metric across all filtration values, we compute the area under the curve. We show that this area corresponds to a metric (Proposition 1), and it turns out that it is exactly the l_1 -metric.

On the sets of the symmetric difference $\{K_t \Delta L_t\}$ we can measure other features besides the number of elements, like the number of connected components as in [24] and [25], the number of nodes, or the Euler characteristic (which might be negative). This shows that our framework is general, flexible and applicable for a wide variety of problems and contexts. Note that the sets of the symmetric difference $\{K_t \Delta L_t\}$ do not form a filtration nor a zigzag filtration [32, Sec 2.2], [33].

C. THE WFCC ALGORITHM FOR COMPARING WEIGHTED GRAPHS

In this section we present the algorithm of our Weight-Filtration Comparison Curve (WFCC) method (Algorithm 1) to compute the symmetric difference metric as a function of the filtration value.

Algorithm 1 Weight-Filtration Comparison Curve (WFCC): Algorithm for the Symmetric Difference as a Function of the Filtration Value

Input: A graph $G = (V, E)$ with two weightings w_1 and w_2 .

Output: A list of unique edge weights, together with a list of the cardinality of the symmetric difference of the weighted graphs for each of these edge weights.

- 1: Let fVals be the sets of unique edge weights in w_1 and w_2 . If computationally required, choose a smaller set of values (see text).
- 2: Let Card be a zero-array with each entry corresponding to a filtration value v in fVals.
- 3: **for** edge e in E **do**
- 4: Find the pair $(\min\{w_1(e), w_2(e)\}, \max\{w_1(e), w_2(e)\})$ giving the interval of filtration values where the edge e is in one graph but not in the other.
- 5: **for** v in fVals **do**
- 6: **if** $\min\{w_1(e), w_2(e)\} \leq v < \max\{w_1(e), w_2(e)\}$ **then:**
- 7: Card[v] += 1
- 8: **end if**
- 9: **end for**
- 10: **end for**
- 11: **return** (fVals, Card).

In Algorithm 1, every unique edge weight in the two graphs is considered. For large graphs this may be computationally intractable. In this case, the set fVals can be taken to be a coarser-grained alternative, accepting that the details of the WFCC curve between individual values may be lost. For the cases explored here, this coarse-graining is not necessary; if

it is used, we suggest exploring the influence of the choice of scale on the results to ensure robustness. Note that the complexity of WFCC 1 is $\mathcal{O}(|E| \cdot |\text{fVals}|)$, where $|\text{fVals}| \leq 2|E|$. If G is a hypercube (Def 3) of dimension n , then the number of edges is $|E| = 2^{n-1}n$. Plotting the output curve of WFCC 1 we get a read-out of how much and when the weightings on a graph differ, the idea being that the lower the curve the more similar the weightings are (see Fig. 1 for a concrete example). For a more quantitative approach, and to provide a read-out of the level of dissimilarity across all filtration values, we can consider the area under the curve of the symmetric difference metric as a function of the filtration value.

D. INTEGRATION TO SCALAR METRIC FOR WEIGHTINGS OF GRAPHS

As a final step in our approach, we consider the area under the WFCC 1 curve. This area turns out to be exactly the l_1 -metric [28, Def 3.6] between the weightings. We calculate the area in the exact case when $\text{fVals} = \mathbb{R}$. Let $G = (V, E)$ be a graph with two weightings w_1 and w_2 . Note that a single edge $e \in E$ influences the cardinality of the symmetric difference only for filtration values t between $w_1(e)$ and $w_2(e)$, and in this range it changes the cardinality by 1. Integrating over t gives a total contribution of $1 * |w_1(e) - w_2(e)|$, where $|w_1(e) - w_2(e)|$ is the length of the range where they differ. To get the total area, we simply sum the contribution of each of the edges.

Proposition 1: Consider a graph $G = (V, E)$. Let w_1 and w_2 be weightings of G with edge-filtrations $\{K_t\}$ and $\{K'_t\}$ respectively. The area under the curve $d_\Delta(K_t, K'_t)$ for all t is precisely the l_1 -metric between the weightings

$$d(w_1, w_2) = \sum_{e \in E} |w_1(e) - w_2(e)|.$$

The l_1 -metric is indeed a metric since the triangle inequality is inherited from the absolute value, and the rest of the properties are straightforward to check. The distance between two weightings can be calculated in linear time $\mathcal{O}(|E|)$ as it is just the sum over all edges. However, for the case of a hypercube of dimension n (Def 4), the number of edges is $2^{n-1}n$ which is exponential with respect to n . These fast computations are an advantage, as classical methods from computational topology often involve much higher complexity.

Note that to compute the l_1 -metric for two weightings, one does not need to construct a filtration first. However, if we ignore the filtration approach, we cannot detect local dissimilarities and their persistence in the structures of the graphs, or get simplified representations of the graphs at different thresholds.

E. DATA COLLECTION

We obtained data on sequenced, drug-resistant *Mycobacterium tuberculosis* isolates from the BV-BRC database [34]. Specifically, we downloaded the ‘Genome’ and ‘AMR

Phenotypes' datasets from the online interface at https://www.bv-brc.org/view/Taxonomy/1773#view_tab=genomes. Cross-referencing by Genome ID, we recorded countries of origin for each record, along with susceptible/resistant patterns for every recorded drug (in the database, some susceptibility data are experimentally measured, and some are inferred from genome information). We treated 'intermediate' resistance records as 'resistant' for this case study. We retained records for the nine drugs for which $> 100k$ records were available (ofloxacin, ethambutol, isoniazid, streptomycin, capreomycin, rifampin, kanamycin, amikacin, pyrazinamide), which overlap with the drugs analysed in previous work [9]. We discarded records with any missing data for these drugs, and subsetted by country of origin. In this illustrative case, we retain only unique drug resistance patterns, to guard against pseudoreplication (see Discussion in Sec IV).

F. NUMERICAL IMPLEMENTATION

Our method finding the filtrations on edges and plotting the curves of their symmetric difference metric, as described above in WFCC 1, is implemented in Python using the NumPy library [35], and for visualisations we use Pandas [36], Matplotlib [37] and Seaborn [38]. For additional visualisations in R we used the libraries ggplot2 [39] and ggraph [40]. The code can be found in <https://github.com/lar-sal/WFCC.git>.

HyperHMM was implemented according to [41]. The convergence criterion for the Hypercubic Baum-Welch Algorithm was set to 0.001. To estimate the uncertainty of the algorithm we used 10 bootstrap resamples of 10000 random walkers each of observed transitions and reported the standard deviation and mean of the resamples.

The weights of the hypercube are obtained by running HyperHMM on the different resistance datasets described above. As output we obtained a list with the transition probabilities from one possible resistance pattern to another, so exactly one number for every edge in the transition graph. This output is unique if we run the algorithm again.

III. RESULTS

A. TOY ILLUSTRATION OF WFCC

In Fig. 1 we see an example of how WFCC compares two different 2-dimensional hypercubes. The area under the curve can be computed as $d(w_1, w_2) = \sum_{i=1}^4 |w_1(e_i) - w_2(e_i)| = 1$ by Prop. 1.

This analysis applies WFCC to the 'raw' probabilities associated with each edge in the transition graph. However, the dynamics of a given system are not determined exclusively by these edge weights, but also by an initial condition. The probability 'flux' (Def 5) through a given edge in the graph depends both on the edge's weight and the probability that its source node is occupied. In many scientific applications, it is these patterns of flux – the probabilities that a system actually makes given transitions – that are of

most interest to compare and analyse. WFCC can naturally be applied to both the 'raw' edge weights and the edge fluxes, as we will illustrate further below.

B. INTERPRETING WFCC CURVES

Generally, a WFCC value W at a given t makes the statement that there are W edges with weights $\leq t$ in one graph that have weights $> t$ in the other. A WFCC of zero means that the same set of edges has weights $\leq t$ in both graphs; a high WFCC value means that the sets of edges with weight $\leq t$ in the two graphs are highly disjoint. The WFCC for $t = 1$, for transition graphs, is always zero, as all edges in both graphs will have weights ≤ 1 .

A strength of the WFCC method is that, while the difference curve integrates to the simple l_1 -metric, the details of the curve provide information on the similarities and differences of the transition graphs, like how big these differences are and where they appear. To help interpret these details, we will introduce some definitions. For a fixed filtration value, we say that an edge is *common* for the two weightings/graphs if it lies in the same (higher/lower) partition-set, and *non-common* if it lies in different partition-sets for the two graphs. The l_1 -metric gives a readout of the symmetric difference metric curve across all filtration values, as it is exactly the area under this curve (Prop. 1).

We now consider how a single edge affects the curve obtained by the WFCC method, and we discuss how multiple edges act together. Let (V, E, w_1) and (V, E, w_2) be two weighted graphs with the same graph structure. We first note that edges $e \in E$ whose weights in both graphs are the same, i.e. $w_1(e) = w_2(e)$, do not contribute to the curve. If the weights differ, say $w_1(e) < w_2(e)$, then the edge e will contribute to the curve by increasing its value by 1 in the interval $[w_1(e), w_2(e))$. In particular, a small difference in weight will give the curve a small 'bump', whereas bigger differences contribute to the curve over most filtration values, resulting in longer 'plateaus'. The symmetry of the symmetric difference means that interchanging the weights $w_1(e)$ and $w_2(e)$ does not change the final curve.

We make some important observations when it comes to how two or more edges can contribute together. First, if the intervals corresponding to two edges intersect, then swapping the start- or endpoints of the intervals does not change the curve. For the example in Fig. 1, we have two intervals $[0.2, 0.7)$ and $[0.3, 0.8)$ contributing to the curve. By setting $w_2(e_1) = 0.8$ and $w_1(e_2) = 0.7$, we get the intervals $[0.3, 0.7)$ and $[0.2, 0.8)$ which will give exactly the same curve. Bumps can also correspond to the intersection of such intervals. Second, it can happen that an interval starts exactly where another one ends, which would be indistinguishable to a single edge with a higher weight difference, in both cases producing a plateau. For example, if one edge has weights 0 and 0.5 and another has weights 1 and 0.5, they raise the curve on $[0, 0.5) \cup [0.5, 1) = [0, 1)$, the same as a single edge with weights 0 and 1.

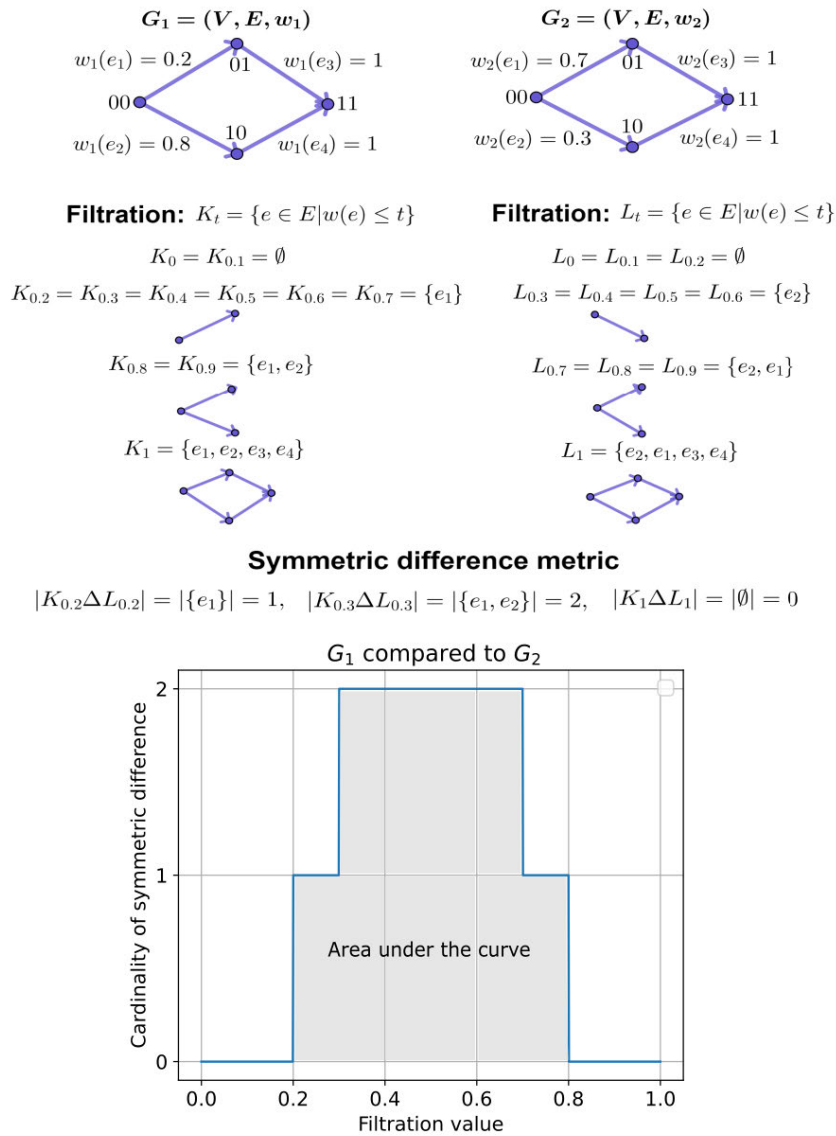


FIGURE 1. Toy example of Weight-Filtration Comparison Curve (WFCC). (top) Two transition graphs (here, 2D hypercubic transition graphs) with the same vertex and edge sets V and E but different edge weights w_1 and w_2 . (centre) The sets of edges retained for each graph through filtration with different thresholds t . (bottom) The symmetric difference metric comparing retained edge sets as a function of threshold t . The curve starts at $t = 0$ with value 0 (both graphs have identical, empty, sets of edges with $t \leq 0$), rises to 1 (G_1 has $w_1(e_1) \leq 0.2$, G_2 does not), then 2 (each graph has one edge with $w \leq 0.3$ that the other does not), before falling as the mismatched edges survive the higher threshold values. The area under the curve is the I_1 -norm.

More broadly, if two graphs share some similar features but differ in others, the position on the t -axis at which peaks and troughs occur can inform us about these differences. A high WFCC at low t that decreases at higher t arises from two graphs that differ in their complements of low-weight edges (low-probability transitions) but share similarities in higher-weight edges (high-probability transitions), and may therefore support similar dynamics with different ‘variations on a theme’. We will see that WFCC peaks at low t , arising from different presence/absence patterns of low-weight edges, are often matched by peaks at higher t values,

presenting a ‘trough’ at medium t values. This is because the low-weight edges necessarily (since probabilities sum to one) remove some probability from higher-weight edges from the same source, leading to differences at higher t values.

C. WFCC METHOD APPLIED TO SIMPLE SYNTHETIC TRANSITION GRAPHS

To illustrate the general ideas presented in Sec III-B, we examine the WFCC curves corresponding to differences between some simple transition graphs. Motivated by the field of evolutionary accumulation modelling, where a system

irreversibly acquires binary traits one-by-one [10], [11], we consider hypercubic transition graphs (Def 4). We first consider a variety of simple 3-dimensional hypercubic transition graphs with different edge weights (Fig 2).

We begin by considering the raw transition probabilities, and later we consider the flux. We start by comparing a graph supporting ‘branching’ dynamics (first state 1, then either state 3 or state 5, then state 7) to two alternatives: a ‘noisier’ graph supporting the same core dynamics but with alternative pathways supported at low probability, and a graph supporting only one deterministic pathway (first state 1, then state 5, then state 7) (Fig. 2 A). In the noisier case, the low-probability alternative edges (with weights close to zero) contribute to the symmetric difference at low t . The difference then vanishes for intermediate t , before reappearing at high t as a consequence of the alternate pathways removing probability from the original pathway (with high weights). This case illustrates how small differences can contribute to the WFCC curve at low and high t , with a trough in between. In the single-path case, a constant symmetric difference is observed across t , first consisting of the lower-weight edge for the branch that is absent, then the higher-weight edge for the branch that is present. This case illustrates how adjacent edge differences can contribute to a uniform symmetric difference profile. These two curves have equal l_1 -metric of value 1, but their unequal shapes inform about particular differences in the structures of the graphs.

In Fig. 2 B, we compare the graph supporting a single, deterministic pathway to a family of increasingly different graphs. The first two support the same core dynamics but with lower and higher probabilities of alternative pathways respectively. The third is a ‘null’ case, where all transitions from a given state are equally likely. The last is an ‘opposite’ case, where a single deterministic pathway is supported that is completely disjoint from the original pathway. In the WFCC curves, we see a similar influence of noise as in Fig. 2 A – lower and higher probabilities of alternative pathways produce symmetric differences over a smaller and larger range respectively of extreme t values, with a trough of zero difference at intermediate t . In both cases, the symmetric difference contribution is from the same set of edges – only the range of t values for which they differ changes. The comparison with the null, uniform case loses this trough: now all t values give a non-zero symmetric difference. Finally, the ‘opposite’ disjoint pathway gives a constant, high difference across all t values, corresponding to the absence of shared edge weights across the comparison.

In Fig. 2 C-D, we apply the same analysis to probability flux patterns (Def 5), reflecting the probability of visiting different edges given a start point at the state of all zeros. In both cases, we see similar general behaviours as for the raw probabilities in Fig. 2 A-B. Some different behaviours to the raw probabilities for the ‘branch’ graph and its ‘noisier’ version, are a peak at medium t values corresponding to the two edges at the end of the branching (3-7,5-7), now non-common, in addition to the edges at the start of the

branch (1)-3,1-5) that contributed to the curve from earlier. For the ‘branch’ and single-path case, the last two edges of the branch are non-common, and contribute in complementary intervals, raising the ‘plateau’.

Finally, in Fig. 3, we compute the l_1 -norm (the area under the WFCC curve) for all pairs of graphs in the family we consider, and embed these pairwise distances in a 2D space using multidimensional scaling [42], [43, Sec 3.2]. Intuitive trends are preserved throughout these comparisons. The disjoint deterministic pathways differ the most; increasing the noise around a given dynamic drives the flux towards the uniform null case; branching variations on single pathway themes are more similar to their single pathways than to each other; two-pathway graphs are intermediate between the one-pathway cases. Taken together, this picture suggests both that the WFCC curves provide detailed insight into the specific differences between graphs, and integrate to produce the interpretable and efficiently-calculated scalar metric to compare them. In Fig. 6 we show more examples of WFCC curve comparisons for these synthetic hypercubes.

D. CASE STUDY: GLOBAL VARIABILITY IN INFERRED EVOLUTIONARY DYNAMICS OF TUBERCULOSIS ANTIMICROBIAL RESISTANCE

One emerging application of accumulation modelling is in studying the evolution of antimicrobial resistance (AMR) [9], [41], [44], [45]. In one approach here, the state space describes drug resistance profiles of bacteria as binary strings, where a 1 at position i means that the bacterium has resistance to drug i . To illustrate how WFCC may be applied to real-world data, we consider observed patterns of antimicrobial resistance for *Mycobacterium tuberculosis* from the BV-BRC database [34] (see Methods). We use HyperHMM, an algorithm for reconstructing transition matrices from observed data [41], to estimate parameters of the (hypercubic) transition graphs in these different cases (see Methods), and compare the different generated hypercubes with the WFCC method (Algorithm 1). The set of inferred transition graphs by country is shown in Fig. 8.

We are interested in progression pathways and how the resistance evolves, which can better be captured by considering the probability flux through state space (Def 5). This is particularly important for inferred hypercubic transition graphs, where many edges (corresponding to transitions between unobserved states) will in general be unidentifiable given data. In these cases, we want to avoid the (arbitrary) values assigned to these unidentifiable edges contributing to our distance measure, and therefore consider probability fluxes, which are constrained by observations.

An illustrative subset of results is shown in Fig. 4. In the illustrated transition graphs, an instance of the system (a tuberculosis lineage) is assumed to start at state 000... , corresponding to no acquired drug resistances. Evolution proceeds by the stepwise acquisition of features, moving

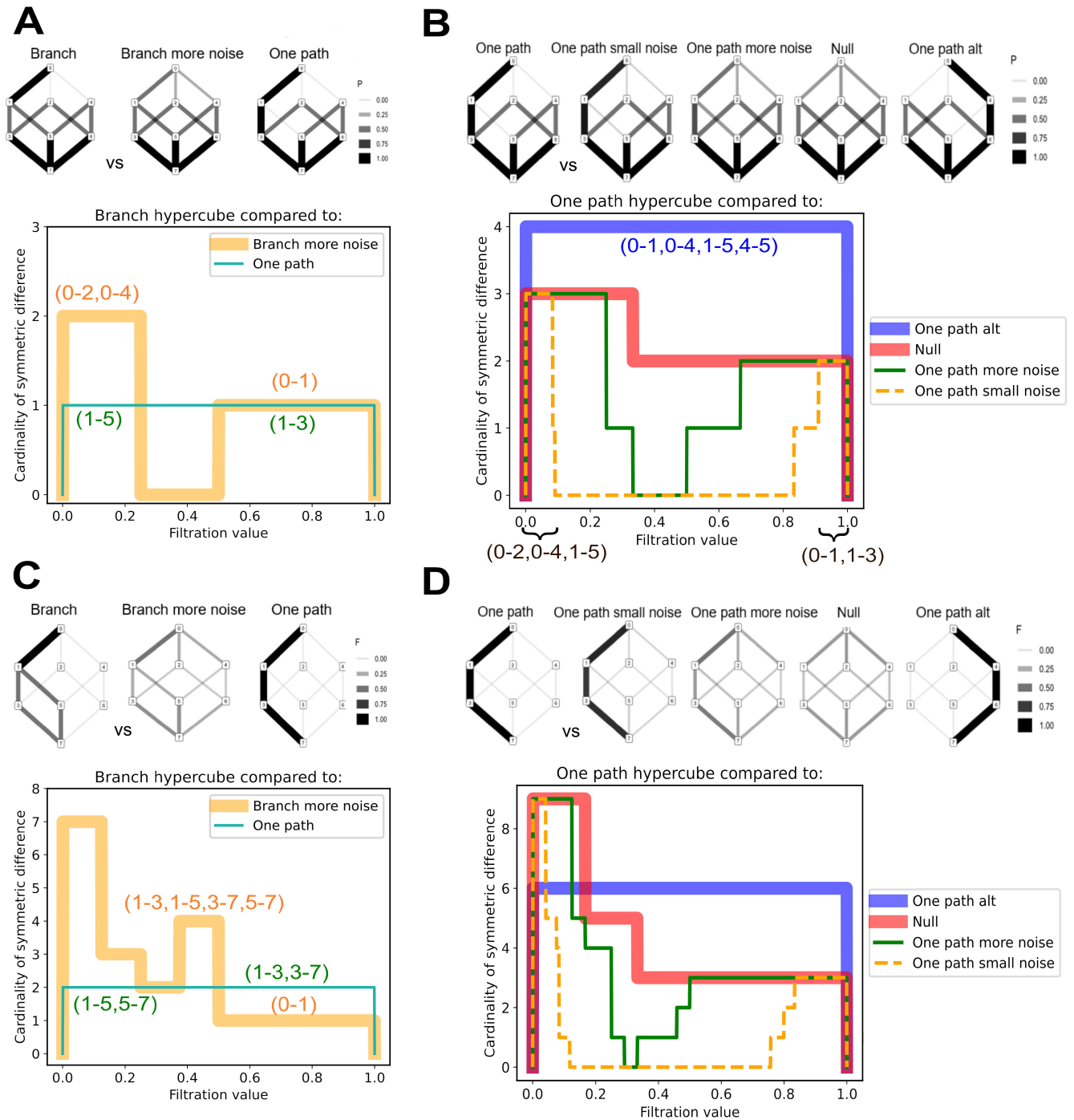


FIGURE 2. Illustration of WFCC features with simple hypercubic transition networks. We perform WFCC analysis on a family of constructed transition networks, in each case plotting the WFCC curve annotated with the edges that give rise to the given symmetric difference values. (A-B) WFCC analysis of transition probabilities. (A) A graph supporting distinct ‘branching’ dynamics compared to a noisier version supporting more competing pathways and one supporting a single, deterministic pathway. These two curves have different shapes, but equal I_1 -metric. (B) A graph supporting a single, deterministic pathway compared to a spectrum of partners: lower- and higher-noise variants supporting more competing pathways, a ‘null’ case supporting all pathways equally, and an opposite case where a different deterministic pathway is supported. (C-D) WFCC analysis of probability fluxes. (C) As (A); (D) as (B).

towards the state 111..., and the graphs in Fig. 4 show the probability fluxes, inferred from observations, through different transitions between states. Kenya illustrates the case where no isolates were found in the dataset with resistance

to any of the 9 drugs of interest. The inferred transition matrix describing evolutionary dynamics is therefore the ‘null’, uniform case: in the absence of observations, all pathways are equally likely. Sweden illustrates a case with

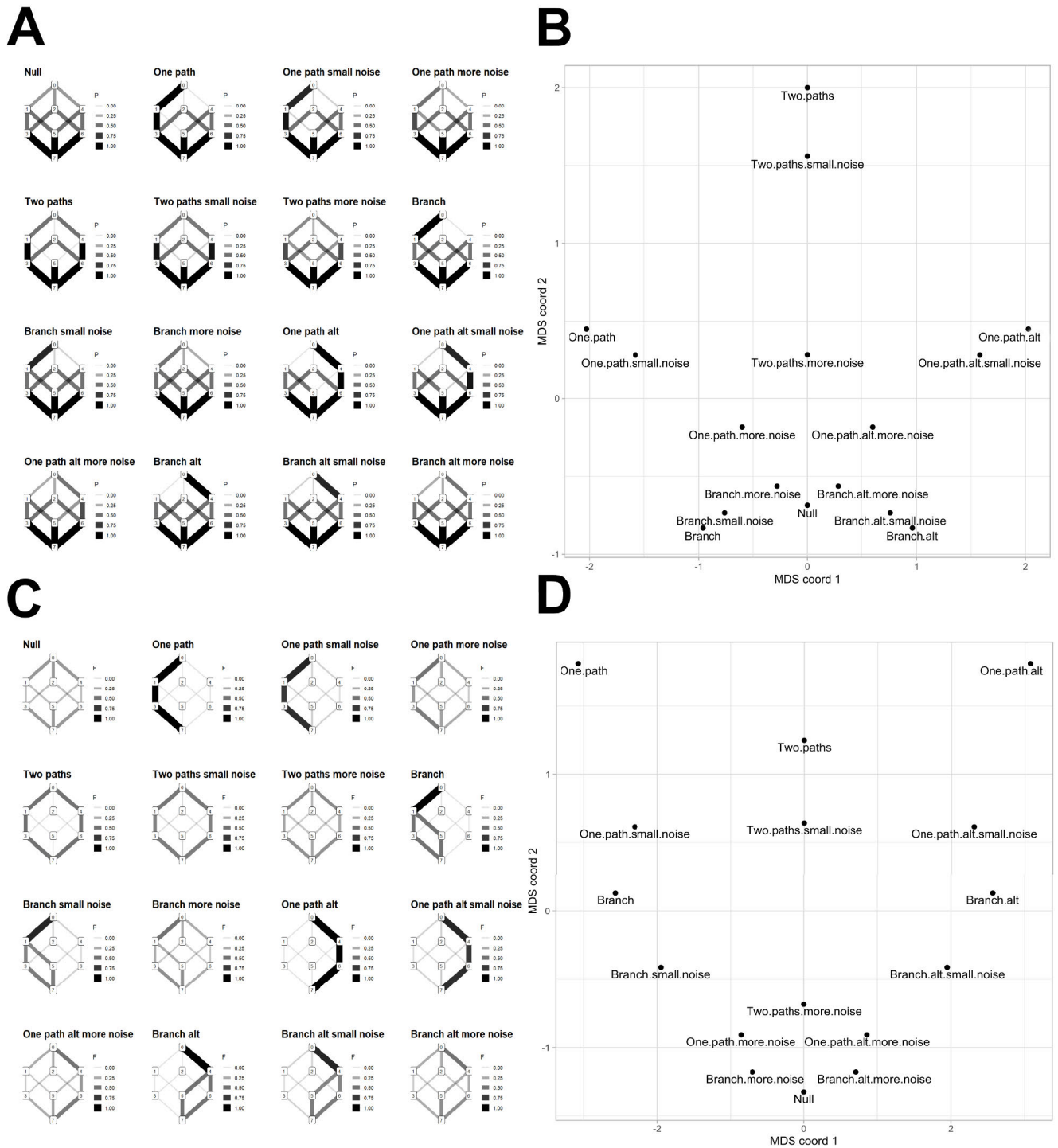


FIGURE 3. Embeddings of integrated WFCC measure (l_1 -norm) for simple transition graphs. (A) The family of example transition graphs and the raw probability weights through their edges: a ‘null’ case supporting all pathways equally; a ‘one path’ case supporting a single, deterministic pathway, and variations of it with lower or higher level of noise for alternative pathways supported at low probability; a ‘two paths’ case supporting two opposite deterministic pathways, and variations of it with lower or higher level of noise; a ‘branch’ case supporting an equal branching in the intermediate stage of the single pathway in ‘one path’, and variations of it with lower or higher level of noise; and alternative ‘alt’ structures mirrored to the ones above. (B) A multidimensional scaling (MDS) embedding of the scalar l_1 -norm distances between these probability sets from integrating the WFCC curve. (C) The same family of example transition graphs with flux edge weights; (D) an analogous MDS embedding using these probability flux sets. The values of the l_1 -norms for each pair of graphs are in Fig. 7.

observations of resistance to a single drug (here, and commonly, streptomycin). The first evolutionary step – streptomycin resistance – is clearly inferred, but no information is

present for subsequent dynamics. In Norway and Bulgaria, some instances of multi-drug resistance constrain early evolutionary steps, leaving later ones unconstrained; in Italy,

Spain, and Belgium, several highly multi-drug resistant isolates constrain longer evolutionary pathways.

These differences are clearly reflected in the WFCC curves comparing these instances to the uniform case (Fig. 4). First consider the behaviour at lower filtration values, shown with a linear scale. Kenya is identical to the uniform null case, with a WFCC of zero throughout. Sweden has a higher WFCC at lower filtration values, corresponding to the disappearance of many edges that do not follow from the constrained first step. The WFCC drops at higher filtration values as these differences (low-weight edges present in the uniform graph but not in the Sweden case) progressively disappear. The other cases have even higher early WFCC curves (as even more of their low-weight edges are constrained away). The longer constrained pathways in the cases of Italy, Spain, and Belgium keep their WFCC curves higher at higher threshold values than the other cases. The curves for Sweden and Norway are characterised by comparatively large deviations from the trend of the other countries at lower values (Sweden) and more intermediate values (Norway), corresponding to the focus of probability flux in one and two canalised early transitions respectively.

The behaviour of the WFCC curves for higher filtration values, shown with a logarithmic scale in Fig. 4, demonstrates the coarser-grained differences between countries. Sweden has a persistent symmetric difference of one over most of the unit interval. This ‘plateau’ reports Sweden’s key feature: the canalised first step, replacing 9 edges each with uniform weight of $1/9$ with a single edge of weight 1. Correspondingly, the WFCC height-1 ‘plateau’ begins at $1/9$ (neither the null nor Swedish case have weights above this for 8 of the initial steps) and extends to 1. Similar behaviour corresponding to early canalisation is observed in the Spanish case, where a height-1 plateau extends from $1/2$ to 1 – here, the same canalised first step is responsible for this plateau, but concentrated probability flux along other early pathways (not present in Sweden) mean that the plateau begins later (as the Spanish graph has several later edges up to weight $1/2$ that are not present in the null). The small number of distinct inferred pathways in the Spanish case are reported as the higher intermediate value prior to the height-1 plateau.

Only Spain and Sweden have a single inferred first step with weight 1. Other countries’ inferred graphs share probability flux across a set of competing initial pathways, and hence their WFCC curves drop to zero at lower filtration values. Norway, for example, drops to zero at $1/3$, reflecting the fact that its flux is shared between 3 initial steps each of weight $1/3$. These initial steps, paired with their following consequential transitions give Norway’s WFCC curve a value of 6 for a long range of WFCC values preceding this drop, with only limited intermediate steps between these larger-scale changes and the finer-grained changes at lower WFCC values.

Italy’s inferred graph supports a wider range of initial steps but shows substantial concentration of probability flux at later stages. Here, the WFCC curve’s drop to zero occurs at a

high filtration value around 0.4, corresponding to the highest weight among the edges in these concentrated later steps. The drop is from a cardinality of 2, reflecting the length of the most strongly constrained tract of transitions. Prior to this, the curve takes a higher cardinality value reflecting both the range of possible initial steps and the later steps surrounding this most constrained tract.

The WFCC traces for Italy (at lower filtration values), Bulgaria, and to a greater extent for Belgium, display more graded behaviour. The limited number of initial steps means that these curves drop to zero at higher filtration values, but their behaviour prior to these drops consists of a large collection of small changes (corresponding to differences in flux at later steps through the transition graph between the canalised country-specific case and the uniform null case). For example, between 4 and 10 pathways are clearly inferred at intermediate ‘levels’ of the Belgium graph (between around 3 and 7 steps from the original source node). Distribution of probability flux across this limited number of pathways, as opposed to a uniform distribution over many more pathways in the uniform case, leads to a collection of differences over small filtration ranges when compared to the null.

This collection of examples shows how specific features of the WFCC curve can quantitatively report differences in patterns of probabilistic flux (here, including the degree of canalisation and number of different pathways supported through a transition graph). The full set of inferred graphs by country could readily be compared in this way and in this detail to address and explore particular hypotheses; the integrated l_1 -norm picture can also be used to compare at scale. In Fig. 5 we plot the MDS visualisation of this integrated measure (distances in Fig. 9) and a principal components analysis (PCA) of the inferred parameterisations for comparison. It will immediately be seen that the countries yielding the uniform hypercube (Kenya, Madagascar, Hong Kong, etc Fig. 8) fall together at an extreme side of both plots; those with streptomycin as the single inferred step (Sweden, Finland, Ireland, etc) cluster together nearby. Ghana also has only a single inferred step, but this step is pyrazinamide resistance, leading to a separation in both embeddings. Those countries with highly constrained, canalised pathways (Spain, Belgium, etc) fall far from the uniform cases. However, the MDS plot based on the integrated WFCC metric (the l_1 -norm) separates these more canalised cases substantially more than the PCA plot, where they are all clustered together. By contrast, the MDS plot separates out those countries where evolutionary dynamics are inferred to be canalised down different pathways, potentially corresponding to the more scientifically interesting reporting of structure.

IV. DISCUSSION

Our WFCC method of filtering the graph and generating a curve as the symmetric difference metric at each filtration value, is aligned with the work in [24] and [25], where the authors form an analogous graph filtration based on the

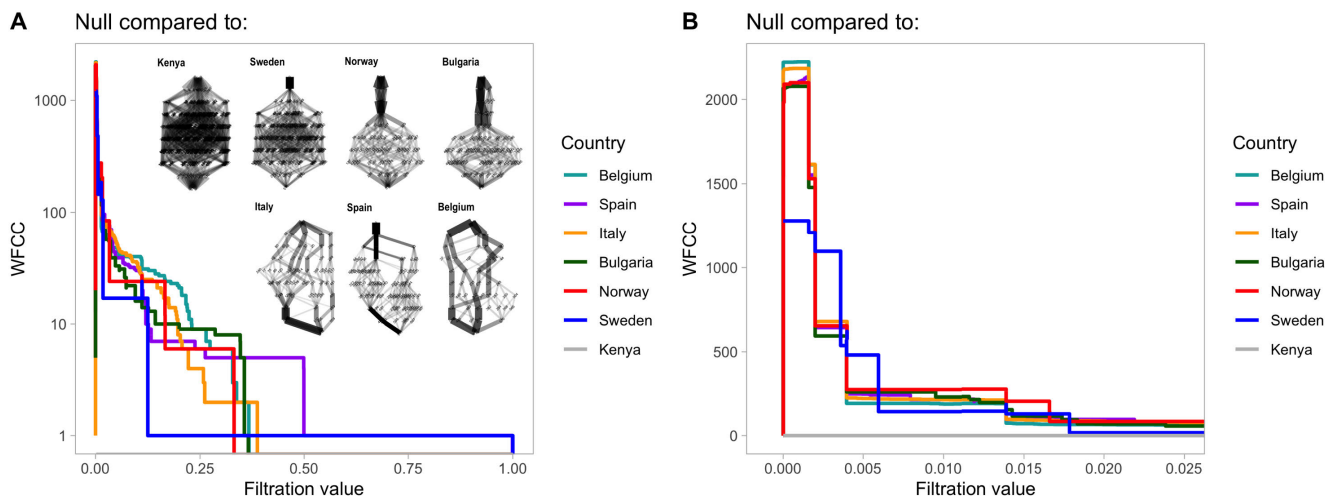


FIGURE 4. Inferred tuberculosis evolutionary dynamics and WFCC comparison. The WFCC curves comparing the null, uniform hypercubic transition graph with the transition graphs inferred for tuberculosis drug resistance evolution in 7 countries. The countries’ corresponding transition graphs are illustrated in the inset. In these graphs, the source node, corresponding to no drug resistances acquired, is at the top, and transitions run from one vertical layer to the one below, with each transition corresponding to the acquisition of resistance to another drug. The width of an edge corresponds to its probability flux; node labels give the decimal value for a binary string describing the presence/absence of resistance to different drugs (see text). The two traces show the same curves; (A) on a log scale, (B) on a linear vertical scale and truncated at threshold values above 0.025, where fewer changes occur. The WFCC curve for Kenya takes value zero throughout (not visible on the logarithmic scale) as that dataset contained no resistance patterns and hence its inferred dynamics are identical to the null case.

weighting. Then they construct a curve tracking features like the number of connected components or node label repetitions across the filtration. The main difference between their approach and ours is that they generate a curve per graph, while we generate a single curve per two graphs. This single curve already encodes dissimilarity information between the two graphs, with no need of doing further comparisons [25] or clustering [24]. In our approach, we count the number of elements in the symmetric difference of edge-sets across the filtration, but we can also look at different properties of the symmetric difference, as long as that property for the empty set is zero. Some examples of other properties we can consider are the number of connected components (as proposed in [24]), the number of nodes and the Euler characteristic [46], which are readily extracted from the analysis we consider and may be used as diagnostic statistics in more general comparison questions. Other TDA tools besides filtrations have recently been employed to specifically study the topology of unweighted hypercubes [47], [48], [49]. Our WFCC method has the advantage of being computationally fast, with runtime proportional to the number of edges times the number of filtration values considered.

In the context of transition graphs inferred from observations, we have considered only a single point estimate (the maximum likelihood parameterisation). In real situations any parameter estimates will have associated uncertainty, and the question of how to incorporate this uncertainty into our distance measure then arises. An obvious option is to report a range of distances and curves corresponding to the range of parameter estimates. Curve features that remain consistent

across this range would then likely be true regardless of the particular ‘true’ (population) parameterisation; features that differed across the range of estimates could be less robustly claimed. Hypothesis testing about the range of ‘true’ differences could then be accomplished by resampling methods (as are used for uncertainty quantification in HyperHMM [41]).

Of course, other methods for inferring transition graphs from data can also be used for a given problem [12], [13], [14]; we have used HyperHMM as an illustrative example. The inference of transition networks from data is the focus, for example, of the powerful *corHMM* package in evolutionary biology [3]. A wide range of approaches in accumulation modelling, developed particularly in the cancer modelling literature [10], [11], works towards the same goal: inferring transitions between states. These approaches often address a parameter space of reduced dimensionality, where relationships between features defining a system state (like the different drugs in our AMR example) are mapped to transition rates between states (although HyperHMM, as an instance of the Baum-Welch algorithm [14], directly reports a full set of estimated edge weights for a transition graph). However, a full parameterised transition graph can readily be constructed from the outcomes of these approaches too, and compared using the WFCC approach.

In the anti-microbial resistance case study, we have only considered unique resistance patterns for each region. This discards many repeated observations of the same resistance pattern across different strains. The rationale for this, in this illustrative case, is that isolates with identical resistance patterns may not have evolved those patterns independently,

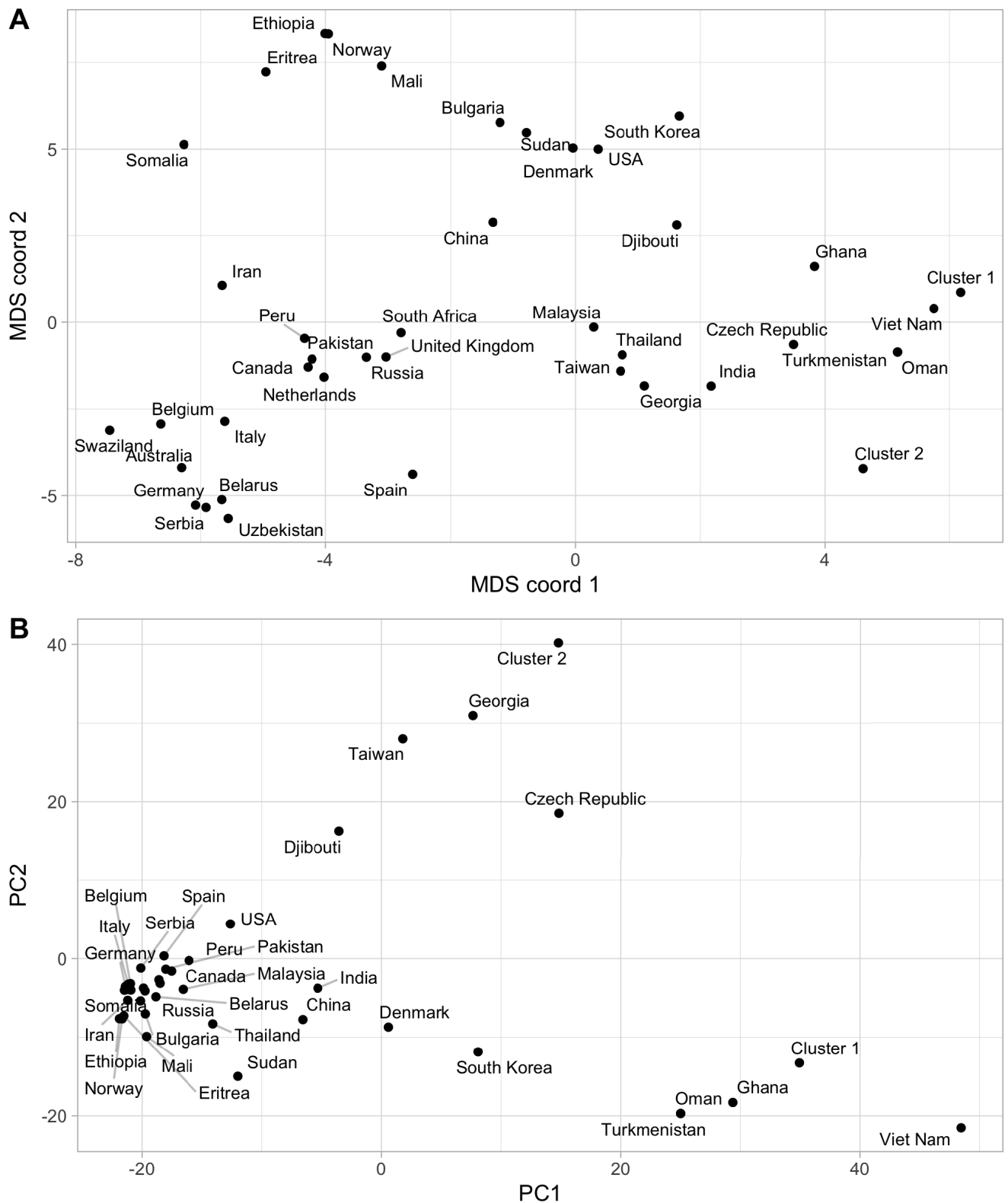


FIGURE 5. Embeddings of inferred tuberculosis dynamics. (A) MDS embedding of transition graphs describing tuberculosis evolution in different countries, using the l_1 -norm distances from the integrated WFCC curve (Fig. 9). (B) PCA embedding of the sets of edge weights for these transition graphs. The PCA plot shows largely homogeneous clusters of countries with some similar transition structures, whereas the MDS plot separates those countries substantially more, showing in more detail the relative differences of those countries, whose evolutionary dynamics are canalised down different pathways. ‘Cluster 1’ is Guatemala, Hong Kong, Kenya, Lebanon, Madagascar, Nepal, Philippines, Uganda; ‘Cluster 2’ is Brazil, Colombia, Ecuador, Finland, Ireland, Sweden; countries within a cluster have identical inferred dynamics (Fig. 9).

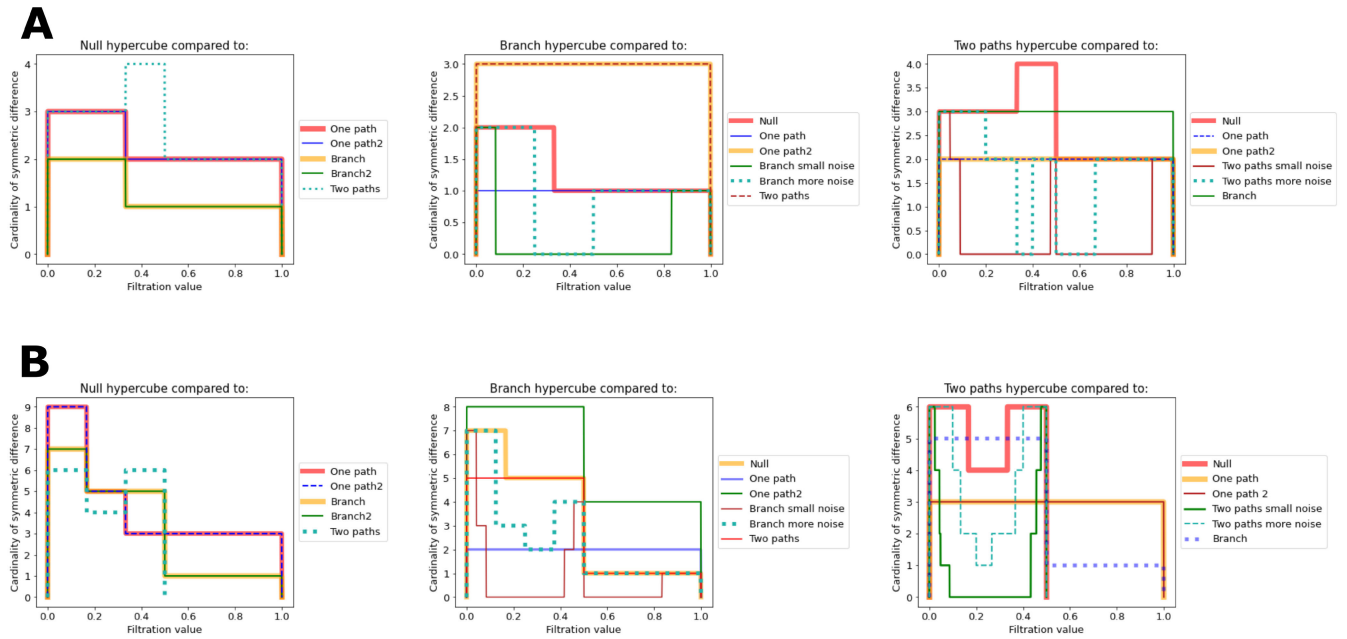


FIGURE 6. WFCC comparison for more cases of synthetic, deterministic hypercubes for A) raw probability weightings in Fig. 3 A and B) flux weightings in Fig. 3 C. Note that '2' here corresponds to 'alt' in the main text.

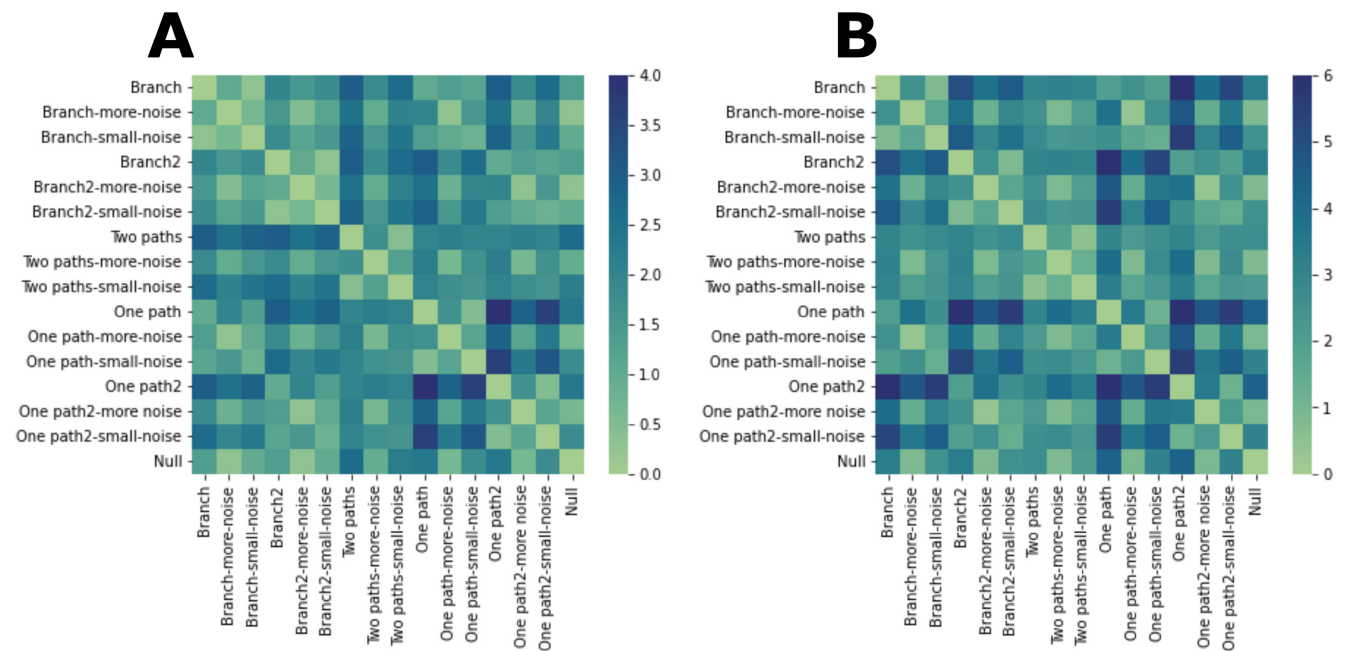


FIGURE 7. I_1 -metric for A) raw probability weightings and B) flux weightings of the synthetic hypercubes considered. The flux captures more strongly the main dissimilarities. Note that '2' here corresponds to 'alt' in the main text.

but may have inherited them from a common ancestor. In that case, they cannot be regarded as independent instances where that pattern has evolved. Of course, reducing the observation set to unique patterns only is likely an overcorrection: the correct approach would be to estimate a phylogeny linking the observations and consider the independent instances of each transition (as in [9], [41] for a tuberculosis case study

with an independently-constructed phylogeny). Additionally, as we are only considering tuberculosis samples after our own curation from the BV-BRC database [34], we cannot claim that our inferred transition graph for evolution in a given country gives a complete representation of current knowledge in that country. However, as this case study is designed to illustrate the comparison method and not report evolutionary

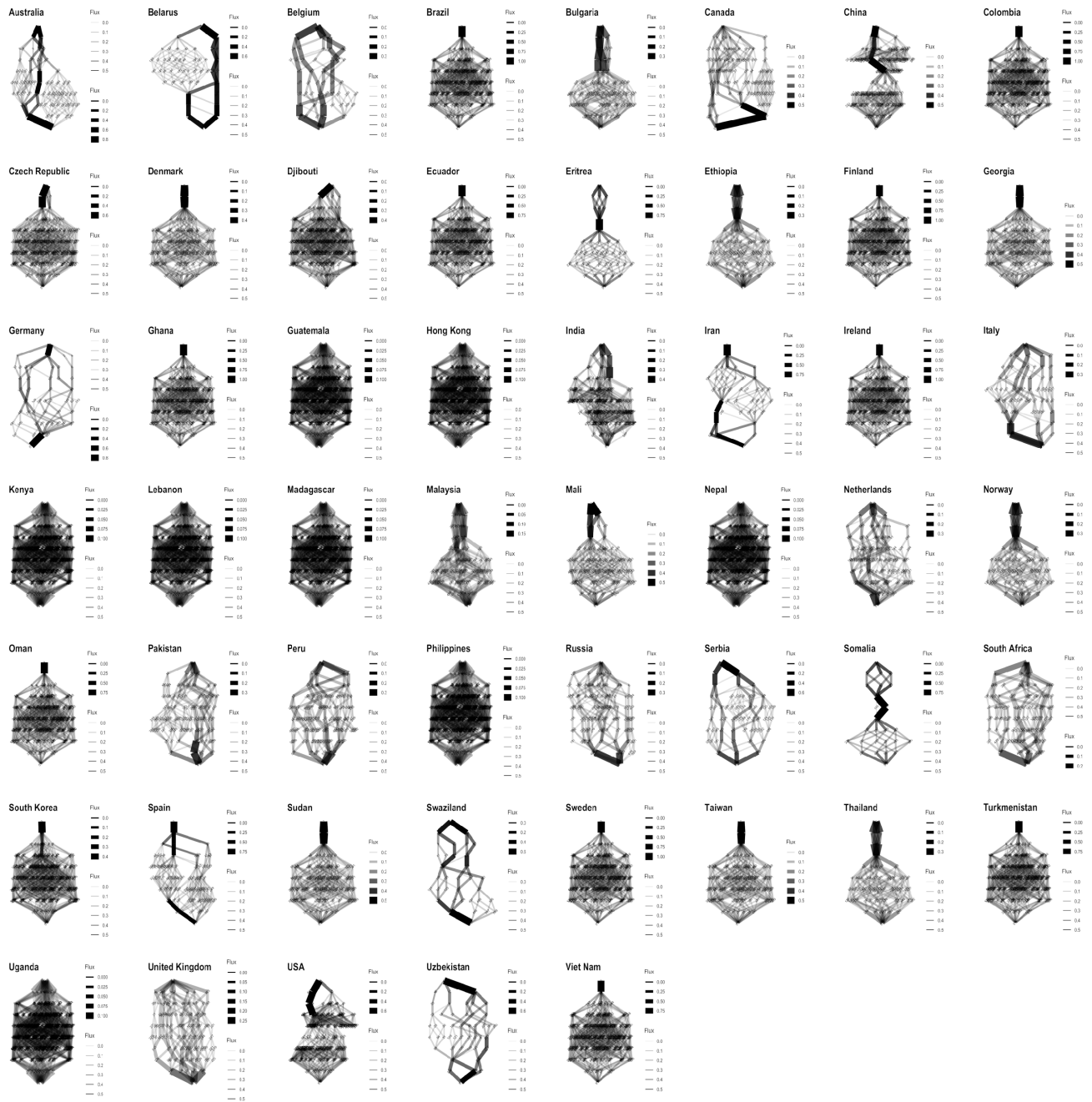


FIGURE 8. Transition graphs inferred for tuberculosis drug resistance evolution across countries. In each graph, the top state corresponds to the state where no drug resistance has been acquired. Lower levels correspond to the progressive acquisition of resistance to different drugs.

dynamics in depth, we take this simpler approach to yield a range of inferred transition graphs for comparison and bear in mind that their applied evolutionary interpretation is limited by these simplifications.

As future work, we are interested in finding full progression pathways. This can be approached by filling in small holes in the graph and keeping track of bigger holes. To do that, the clique complex can be a natural tool [50, Section 3], [51, Section 3.1]. By definition, clique complexes are

Vietoris-Rips complexes [52], and Vietoris-Rips complexes are proven to be an efficient way to compute persistent homology [53]. One can also consider other simplicial complexes from graphs, for example the ones in [51]. Another venue of future work is to explore stability conditions for the WFCC method. This consists of finding metrics on the data so that small perturbations in the data will produce small differences in the Markov transition graphs inferred from an inference algorithm like HyperHMM [41]. For now, we hope

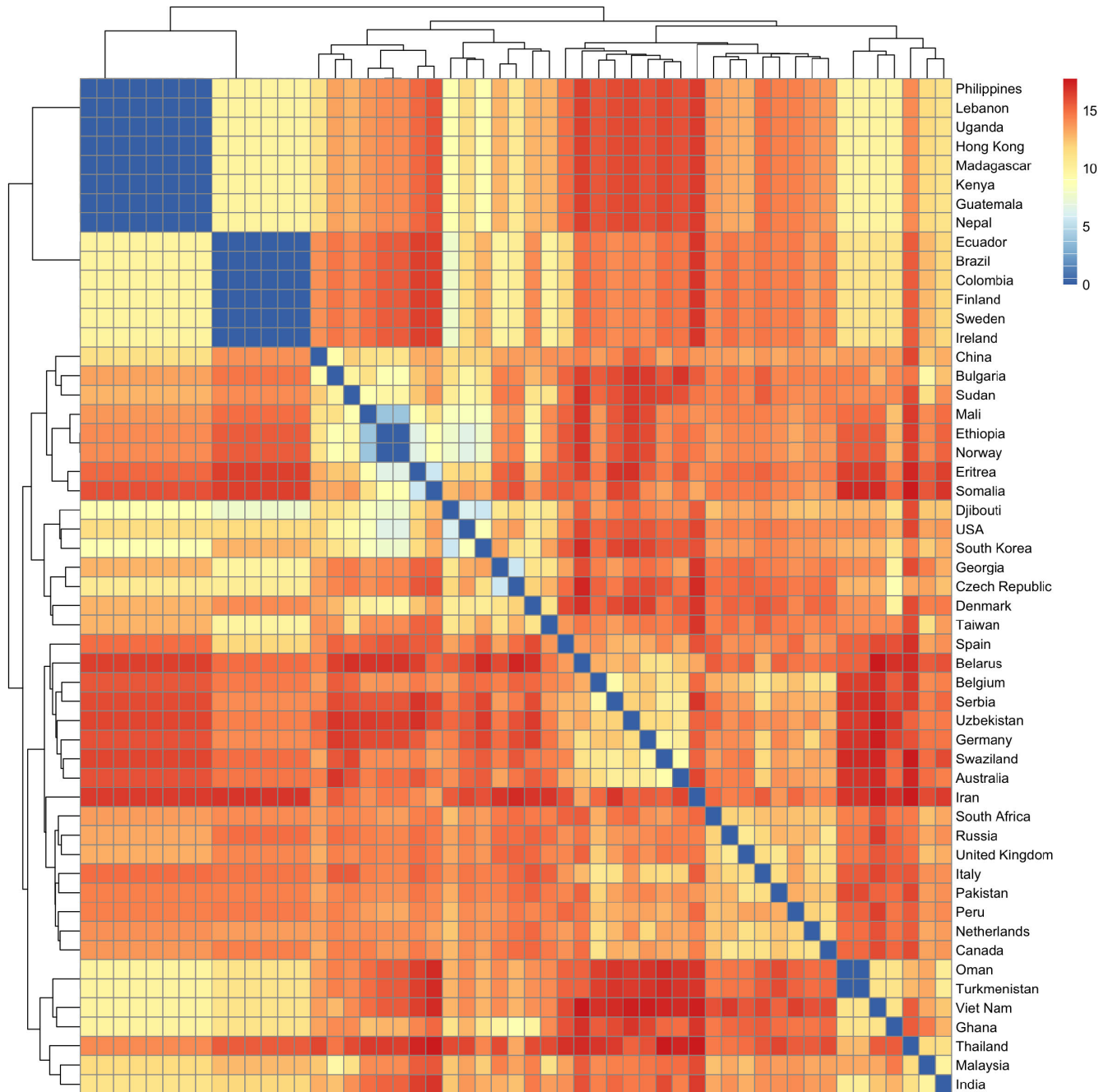


FIGURE 9. I_1 distances between transitions graphs inferred for tuberculosis evolution across countries.

that the insights provided from the WFCC approach will help to inform about scientifically interesting distinctions between inferred transition dynamics across a range of scientific disciplines.

**APPENDIX A
METHODS: HYPERCUBES AND FLUX WEIGHTINGS**

A **(directed) graph** $G = (V, E)$ consists of a **vertex set** V and an **edge set** $E \subseteq V \times V$ consisting of ordered pairs of

vertices. We write xy for the edge $e = (x, y) \in E$, and say that xy is an edge from x to y .

Definition 3: (Directed version of [48, 2.4]) For a positive integer n , the **(directed n -dimensional) hypercube** is the directed graph $H = (V, E)$ where the vertex set V consists of all possible 2^n binary strings of length n , i.e.

$$V = \{x = x_1 \dots x_n \mid x_i \in \{0, 1\} \text{ for } i = 1, \dots, n\},$$

and where xy is in E if there exists exactly one value j such that $0 = x_j < y_j = 1$ and $x_i = y_i$ for all $i \neq j$.

In particular, there is an edge from x to y if and only if you can change a single zero in the binary string x to a one and get the binary string y . For example, there is an edge from 1010 to 1110, but not the other way, and not from 1010 to 1111 because you would need to change two zeros. The labels of the vertices in the form of binary strings are precisely the states in the evolutionary pathways. Note that hypercubes are finite, simple (acyclic) and connected. The name *hypercube* comes from the fact that it is the graph corresponding to the 1-skeleton of a hypercube.

A **weighting** on a graph $G = (V, E)$ is a function $w : E \rightarrow \mathbb{R}$, where $w(e) \in \mathbb{R}$ is the **weight** of $e \in E$ and the triple $G = (V, E, w)$ is called a **weighted graph**. In our examples, we will work with discrete-time Markov models. Here, at each timestep, the system occupies a particular vertex, and between timesteps, exactly one outgoing edge from the current vertex is followed to obtain the new state. This edge may in general be a loop, so that the system occupies the same state in two adjacent timesteps. For discrete-time Markov models, the probability that an edge is followed is given by that edge's weight – hence, the weight on each edge is between 0 and 1, and the set of weights on the set of edges leaving a particular node must sum to 1.

Definition 4: A **hypercubic transition graph** $H = (V, E, w)$ is a weighted hypercube $H = (V, E)$ where the sum of the weight of all outgoing edges from a vertex must sum to one, i.e. for all $x \in V$ we have $\sum_{xy \in E} w(xy) = 1$. To fulfil this property we equip the vertex corresponding to the string of all 1s ($x = 111\dots$) with a self-edge of weight 1.

Hypercubic transition graphs defined as above are special cases of Markov transition graphs [1], [2], [3] of evolutionary pathways, where the edge weights are typically inferred by statistical algorithms from real data. The vertex corresponding to the string of all 1s ($x = 111\dots$) acts as an 'absorbing state': once the system reaches this vertex it remains there for all future timesteps. An evolutionary accumulation process involves starting at the vertex corresponding to the string of all 0s ($x = 000\dots$, corresponding to no acquired features) and stepping towards this absorbing state, with each step on a hypercubic edge corresponding to the acquisition of one feature.

In addition to the weight on an edge, which determines the probability that the edge is traversed *given that* the system occupies the source vertex, we are also interested in probability flux, which is the probability that the system *actually* traverses the edge during some specified time window. The flux of an edge is the probability that the system occupies its starting vertex multiplied by the probability of then traversing that edge, integrated over the time window. In our hypercubic cases, we will take this time window to run from 0 to n , the full set of times over which the system evolves. In the hypercubic cases, each edge can only be traversed at exactly one time, corresponding to the number of features in its source vertex.

Of course, the edge probability weight and edge flux weight can differ dramatically, as the source vertex may be very rarely encountered in systems initialised according to some initial condition (for us, $x = 000\dots$).

Definition 5: Let $H = (V, E, w)$ be a hypercubic transition graph, with a initial node v_0 corresponding to the all-zero string. The original probability weight of the edge e_{ij} is denoted by $w(e_{ij}) \in [0, 1]$. We let $P(v_i)$ be the probability of passing through the vertex v_i when starting in v_0 , and we denote the **flux value of the edge** e_{ij} by $f(e_{ij})$. With this notation, the flux value is $f(e_{ij}) = P(v_i) \cdot w(e_{ij})$, and can be calculated recursively by

$$\begin{aligned} P(v_0) &= 1 \\ f(e_{ij}) &= P(v_i)w(e_{ij}) && \text{for } i < j \\ P(v_i) &= \sum_{k < i} f(e_{ki}). \end{aligned} \quad (1)$$

REFERENCES

- [1] F. Kachapova, "Representing Markov chains with transition diagrams," *J. Math. Statist.*, vol. 9, no. 3, pp. 149–154, Mar. 2013.
- [2] T. Richardson and P. Spirtes, "Ancestral graph Markov models," *Ann. Statist.*, vol. 30, no. 4, pp. 962–1030, Aug. 2002, doi: [10.1214/aos/1031689015](https://doi.org/10.1214/aos/1031689015).
- [3] J. D. Boyko and J. M. Beaulieu, "Generalized hidden Markov models for phylogenetic comparative datasets," *Methods Ecol. Evol.*, vol. 12, no. 3, pp. 468–478, Mar. 2021, doi: [10.1111/2041-210x.13534](https://doi.org/10.1111/2041-210x.13534).
- [4] L. J. Allen, *An Introduction to Stochastic Processes With Applications to Biology*. Boca Raton, FL, USA: CRC Press, 2010.
- [5] S. Gómez, A. Arenas, J. Borge-Holthoefer, S. Meloni, and Y. Moreno, "Discrete-time Markov chain approach to contact-based disease spreading in complex networks," *EPL Europhys. Lett.*, vol. 89, no. 3, p. 38009, Feb. 2010.
- [6] T. Michelot, R. Langrock, and T. A. Patterson, "MoveHMM: An R package for the statistical modelling of animal movement data using hidden Markov models," *Methods Ecol. Evol.*, vol. 7, no. 11, pp. 1308–1315, Nov. 2016.
- [7] B.-J. Yoon, "Hidden Markov models and their applications in biological sequence analysis," *Current Genomics*, vol. 10, no. 6, pp. 402–415, Sep. 2009.
- [8] I. G. Johnston and B. P. Williams, "Evolutionary inference across eukaryotes identifies specific pressures favoring mitochondrial gene retention," *Cell Syst.*, vol. 2, no. 2, pp. 101–111, Feb. 2016.
- [9] S. F. Greenbury, M. Barahona, and I. G. Johnston, "HyperTraPS: Inferring probabilistic patterns of trait acquisition in evolutionary and disease progression pathways," *Cell Syst.*, vol. 10, no. 1, pp. 39–51, Jan. 2020. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S2405471219303850>
- [10] R. Diaz-Uriarte and P. Herrera-Nieto, "EvAM-tools: Tools for evolutionary accumulation and cancer progression models," *Bioinformatics*, vol. 38, no. 24, pp. 5457–5459, Dec. 2022.
- [11] R. Diaz-Uriarte, "A picture guide to cancer progression and monotonic accumulation models: Evolutionary assumptions, plausible interpretations, and alternative uses," 2023, *arXiv:2312.06824*.
- [12] T. W. Anderson and L. A. Goodman, "Statistical inference about Markov chains," *Ann. Math. Statist.*, vol. 28, no. 1, pp. 89–110, Mar. 1957.
- [13] P. Billingsley, "Statistical methods in Markov chains," *Ann. Math. Statist.*, vol. 32, no. 1, pp. 12–40, Mar. 1961.
- [14] L. E. Baum, T. Petrie, G. Soules, and N. Weiss, "A maximization technique occurring in the statistical analysis of probabilistic functions of Markov chains," *Ann. Math. Statist.*, vol. 41, no. 1, pp. 164–171, Feb. 1970.
- [15] B. D. McKay, "Graph isomorphism," in *Encyclopedia of Algorithms*, M.-Y. Kao, Ed. Boston, MA, USA: Springer, 2008, pp. 373–376, doi: [10.1007/978-0-387-30162-4](https://doi.org/10.1007/978-0-387-30162-4).
- [16] M. Chellali, T. W. Haynes, S. T. Hedetniemi, and A. McRae, "[1, 2]-sets in graphs," *Discrete Appl. Math.*, vol. 161, no. 18, pp. 2885–2893, 2013. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S0166218X13002904>

- [17] B. Chaluvvaraju, M. Chellali, and K. Vidya, "Perfect k-domination in graphs," *Australas. J. Combinatorics*, vol. 48, no. 10, pp. 175–184, 2010.
- [18] G. Chartrand, G. Kubicki, and M. Schultz, "Graph similarity and distance in graphs," *Aequationes Mathematicae*, vol. 55, nos. 1–2, pp. 129–145, Feb. 1998, doi: [10.1007/s000100050025](https://doi.org/10.1007/s000100050025).
- [19] X. Wang, Q. Zhang, D. Guo, and X. Zhao, "A survey of continuous subgraph matching for dynamic graphs," *Knowl. Inf. Syst.*, vol. 65, no. 3, pp. 945–989, Oct. 2022, doi: [10.1007/s10115-022-01753-x](https://doi.org/10.1007/s10115-022-01753-x).
- [20] M.-L. Fernández and G. Valiente, "A graph distance metric combining maximum common subgraph and minimum common supergraph," *Pattern Recognit. Lett.*, vol. 22, nos. 6–7, pp. 753–758, May 2001. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S0167865501000174>
- [21] A. P. Kartun-Giles and G. Bianconi, "Beyond the clustering coefficient: A topological analysis of node neighbourhoods in complex networks," *Chaos, Solitons Fractals, X*, vol. 1, Mar. 2019, Art. no. 100004. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S259005441930003X>
- [22] B. DasGupta, X. He, T. Jiang, M. Li, J. Tromp, and L. Zhang, "Nearest neighbor interchange and related distances," in *Encyclopedia Algorithms*, M.-Y. Kao, Ed. Boston, MA, USA: Springer, 2008, pp. 573–576, doi: [10.1007/978-0-387-30162-4](https://doi.org/10.1007/978-0-387-30162-4).
- [23] Y. Xu, S. M. Salapaka, and C. L. Beck, "A distance metric between directed weighted graphs," in *Proc. 52nd IEEE Conf. Decis. Control*, Dec. 2013, pp. 6359–6364.
- [24] L. O'Bray, B. Rieck, and K. Borgwardt, "Filtration curves for graph representation," in *Proc. 27th ACM SIGKDD Conf. Knowl. Discovery Data Mining*. New York, NY, USA: ACM, Aug. 2021, pp. 1267–1275, doi: [10.1145/3447548.3467442](https://doi.org/10.1145/3447548.3467442).
- [25] T. Schulz, P. Welke, and S. Wrobel, "Graph filtration kernels," in *Proc. AAAI Conf. Artif. Intell.*, Jun. 2022, vol. 36, no. 8, pp. 8196–8203. [Online]. Available: <https://ojs.aaai.org/index.php/AAAI/article/view/20793>
- [26] P. Wills and F. G. Meyer, "Metrics for graph comparison: A practitioner's guide," *PLoS ONE*, vol. 15, no. 2, Feb. 2020, Art. no. e0228728, doi: [10.1371/journal.pone.0228728](https://doi.org/10.1371/journal.pone.0228728).
- [27] N. D. Monnig and F. G. Meyer, "The resistance perturbation distance: A metric for the analysis of dynamic networks," *Discrete Appl. Math.*, vol. 236, pp. 347–386, Feb. 2018. <https://www.sciencedirect.com/science/article/pii/S0166218X17304626>
- [28] W. Rudin, *Real and Complex Analysis*, 3rd ed. New York, NY, USA: McGraw-Hill, 1987.
- [29] R. Basshuysen and F. Schaefer, "M," in *Odern Engine Technology: From A to Z*, F. E. Szabo, Ed. New York, NY, USA: Academic, 2007, pp. 726–741. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/B9780124095205500205>
- [30] C. Flament, *Applications of graph theory to group structure*. London, U.K.: Prentice-Hall, 1963.
- [31] S. T. Vittadello and M. P. H. Stumpf, "Model comparison via simplicial complexes and persistent homology," *Roy. Soc. Open Sci.*, vol. 8, no. 10, Oct. 2021, Art. no. 211361, doi: [10.1098/rsos.211361](https://doi.org/10.1098/rsos.211361).
- [32] A. Myers, D. Muñoz, F. A. Khasawneh, and E. Munch, "Temporal network analysis using zigzag persistence," *EPJ Data Sci.*, vol. 12, no. 1, p. 6, Mar. 2023, doi: [10.1140/epjds/s13688-023-00379-5](https://doi.org/10.1140/epjds/s13688-023-00379-5).
- [33] G. Carlsson and V. Silva, "Zigzag persistence," *Found. Comput. Math.*, vol. 10, pp. 367–405, Dec. 2008.
- [34] R. D. Olson, "Introducing the bacterial and viral bioinformatics resource center (BV-BRC): A resource combining PATRIC, IRD and ViPR," *Nucleic Acids Res.*, vol. 51, no. D1, pp. D678–D689, Jan. 2023, doi: [10.1093/nar/gkac1003](https://doi.org/10.1093/nar/gkac1003).
- [35] C. R. Harris, "Array programming with NumPy," *Nature*, vol. 585, no. 7825, pp. 357–362, Sep. 2020, doi: [10.1038/s41586-020-2649-2](https://doi.org/10.1038/s41586-020-2649-2).
- [36] *Pandas-Dev/Pandas: Pandas*, The Pandas Development Team, London, U.K., Feb. 2020, doi: [10.5281/zenodo.3509134](https://doi.org/10.5281/zenodo.3509134).
- [37] J. D. Hunter, "Matplotlib: A 2D graphics environment," *Comput. Sci. Eng.*, vol. 9, no. 3, pp. 90–95, 2007.
- [38] M. Waskom, "Seaborn: Statistical data visualization," *J. Open Source Softw.*, vol. 6, no. 60, p. 3021, Apr. 2021, doi: [10.21105/joss.03021](https://doi.org/10.21105/joss.03021).
- [39] H. Wickham, *Ggplot2: Elegant Graphics for Data Analysis*. New York, NY, USA: Springer-Verlag, 2016. [Online]. Available: <https://ggplot2.tidyverse.org>
- [40] T. L. Pedersen. (2022). *Ggraph: An Implementation of Grammar of Graphs for Graphs and Networks*. [Online]. Available: <https://ggraph.dataimaginist.com> and <https://github.com/thomasp85/ggraph>
- [41] M. T. Moen and I. G. Johnston, "HyperHMM: Efficient inference of evolutionary and progressive dynamics on hypercubic transition graphs," *Bioinformatics*, vol. 39, no. 1, Jan. 2023, Art. no. btac803, doi: [10.1093/bioinformatics/btac803](https://doi.org/10.1093/bioinformatics/btac803).
- [42] K. V. Mardia, "Some properties of classical multi-dimensional scaling," *Commun. Statist. Theory Methods*, vol. 7, no. 13, pp. 1233–1241, Jan. 1978, doi: [10.1080/03610927808827707](https://doi.org/10.1080/03610927808827707).
- [43] M. A. A. Cox and T. F. Cox, *Multidimensional Scaling*. Berlin, Germany: Springer, 2008, pp. 315–347, doi: [10.1007/978-3-540-33037-0_14](https://doi.org/10.1007/978-3-540-33037-0_14).
- [44] N. Beerenwinkel and M. Drton, "A mutagenetic tree hidden Markov model for longitudinal clonal HIV sequence data," *Bioinformatics*, vol. 8, no. 1, pp. 53–71, Jan. 2007.
- [45] B.-Y. Li, W.-P. Shi, C.-M. Zhou, Q. Zhao, V. K. Diwan, X.-B. Zheng, Y. Li, S. Hoffner, and B. Xu, "Rising challenge of multidrug-resistant tuberculosis in China: A predictive study using Markov modeling," *Infectious Diseases Poverty*, vol. 9, no. 1, pp. 57–64, Dec. 2020.
- [46] K. Turner, S. Mukherjee, and D. M. Boyer, "Persistent homology transform for modeling shapes and surfaces," *Inf. Inference, J. IMA*, vol. 3, no. 4, pp. 310–344, Dec. 2014, doi: [10.1093/imaia/iau011](https://doi.org/10.1093/imaia/iau011).
- [47] M. Adamaszek and H. Adams, "On Vietoris–Rips complexes of hypercube graphs," Tech. Rep., 2021. [Online]. Available: <https://arxiv.org/abs/2103.01040>
- [48] H. Adams, S. Shukla, and A. Singh, "Čech complexes of hypercube graphs," Tech. Rep., 2022. [Online]. Available: <https://arxiv.org/abs/2212.05871>
- [49] H. Adams and Ž. Virk, "Lower bounds on the homology of Vietoris–Rips complexes of hypercube graphs," Tech. Rep., 2023.
- [50] D. Horak, S. Maletić, and M. Rajković, "Persistent homology of complex networks," *J. Stat. Mech., Theory Exp.*, vol. 2009, no. 3, Mar. 2009, Art. no. P03034, doi: [10.1088/1742-5468/2009/03/P03034](https://doi.org/10.1088/1742-5468/2009/03/P03034).
- [51] M. G. Bergomi, M. Ferri, and L. Zuffi, "Topological graph persistence," *Commun. Appl. Ind. Math.*, vol. 11, pp. 72–87, Dec. 2020, doi: [10.2478/caim-2020-0005](https://doi.org/10.2478/caim-2020-0005).
- [52] H. Edelsbrunner and J. Harer, *Computational Topology—An Introduction*. Providence, RI, USA: American Mathematical Society, 2010.
- [53] U. Bauer, "Ripsper: Efficient computation of Vietoris–Rips persistence barcodes," *J. Appl. Comput. Topol.*, vol. 5, no. 3, pp. 391–423, Jun. 2021, doi: [10.1007/s41468-021-00071-5](https://doi.org/10.1007/s41468-021-00071-5).



BELÉN GARCÍA PASCUAL received the B.S. degree in mathematics from the Complutense University of Madrid, Spain, in 2018, and the M.S. degree in mathematics, topology from the University of Bergen, Norway, in 2020, where she is currently pursuing the Ph.D. degree in biomathematics, funded by the ERC research grant 'EvoConBio' under the supervision of Iain G. Johnston.



LARS M. SALBU received the B.S. degree in physics from the Norwegian University of Science and Technology (NTNU), in 2017, and the M.S. degree in mathematics from the University of Bergen, Norway, in 2019, where he is currently pursuing the Ph.D. degree with the Department of Mathematics.

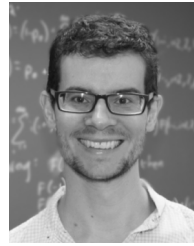


JESSICA RENZ received the B.S. and M.S. degrees in mathematics from the University of Stuttgart, Germany, in 2022. She is currently pursuing the Ph.D. degree in biomathematics with the Department of Mathematics, University of Bergen, Norway, funded through the HyperEvol Project of the Centre for Antimicrobial Resistance in Western Norway (CAMRIA).



KONSTANTINOS GIANNAKIS received the Ph.D. degree in theoretical computer science from the Department of Informatics, Ionian University, Greece, in 2016.

He was a Postdoctoral Researcher with Ionian University, until 2020, and the University of Bergen, Bergen, Norway, until 2023. He is currently a Researcher with the Norwegian Institute of Public Health, Bergen. His research interests include the theory of computation and computational intelligence, statistics, communication networks, and graph structures.



IAIN G. JOHNSTON received the M.Sci. degree in natural sciences from the University of Cambridge, U.K., in 2006, and the D.Phil. degree in theoretical physics from the University of Oxford, U.K., in 2011.

From 2012 to 2015, he was a Research Fellow in biomathematics funded through a Medical Research Council Fellowship at Imperial College London. From 2015 to 2019, he was a Group Leader with the School of Biosciences, University of Birmingham, U.K. Since 2019, he has been with the Department of Mathematics, University of Bergen, Norway, first as an Associate Professor and then as a Full Professor.

• • •