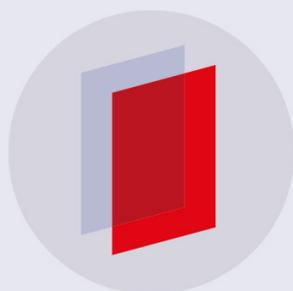**PAPER • OPEN ACCESS**

Ultrametricity increases the predictability of cultural dynamics

To cite this article: Alexandru-Ionu Bbeanu *et al* 2018 *New J. Phys.* **20** 103026

View the article online for updates and enhancements.

IOP ebooks™

Bringing you innovative digital publishing with leading voices to create your essential collection of books in STEM research.

Start exploring the collection - download the first chapter of every title for free.

New Journal of Physics

The open access journal at the forefront of physics

Deutsche Physikalische Gesellschaft **DPG**

IOP Institute of Physics

Published in partnership with: Deutsche Physikalische Gesellschaft and the Institute of Physics

CrossMark

OPEN ACCESS

PAPER

Ultrametricity increases the predictability of cultural dynamics

Alexandru-Ionuţ Băbeanu[1] , Jorinde van de Vis[1,2] and Diego Garlaschelli[1,3]

1   Lorentz Institute for Theoretical Physics, Leiden University, The Netherlands
2   Dutch National Institute for Subatomic Physics, The Netherlands
3   IMT School for Advanced Studies, Lucca, Italy

E-mail: a.i.babeanu@gmail.com

Abstract

A quantitative understanding of societies requires useful combinations of empirical data and mathematical models. Models of cultural dynamics aim at explaining the emergence of culturally homogeneous groups through social influence. Traditionally, the initial cultural traits of individuals are chosen uniformly at random, the emphasis being on characterizing the model outcomes that are independent of these ('annealed') initial conditions. Here, motivated by an increasing interest in forecasting social behavior in the real world, we reverse the point of view and focus on the effect of specific ('quenched') initial conditions, including those obtained from real data, on the final cultural state. We study the predictability, rigorously defined in an information-theoretic sense, of the *social content* of the final cultural groups (i.e. who ends up in which group) from the knowledge of the initial cultural traits. We find that, as compared to random and shuffled initial conditions, the hierarchical ultrametric-like organization of empirical cultural states significantly increases the predictability of the final social content by largely confining cultural convergence within the lower levels of the hierarchy. Moreover, predictability correlates with the compatibility of short-term social coordination and long-term cultural diversity, a property that has been recently found to be strong and robust in empirical data. We also introduce a null model generating initial conditions that retain the ultrametric representation of real data. Using this ultrametric model, predictability is highly enhanced with respect to the random and shuffled cases, confirming the usefulness of the empirical hierarchical organization of culture for forecasting the outcome of social influence models. These results appear to be highly independent of the empirical data source.

1. Introduction

Understanding the self-organization and emergence of large-scale patterns in real societies is one of the most fascinating, yet extremely challenging problems of modern social science [1]. A prominent field of research studies the spontaneous emergence of groups of culturally homogeneous individuals. One of the mechanisms that are believed to play a key role in this process is *social influence*, i.e. the gradual convergence of the cultural traits, attitudes and opinions of individuals subject to mutual social interactions—this is a restricted definition that is implicit in this study and in previous work that this study builds on; see [2] for a more generic definition. Stylized models of cultural dynamics under social influence have attracted the interest of an interdisciplinary community of sociologists, computational social scientists and statistical physicists [3].

One of the prototypical models in this context is the popular Axelrod model [4], which has been studied in many variants over the last two decades [5–13]. The model is multi-agent, with a cultural vector associated to each agent. One cultural vector is a sequence of subjective cultural traits (opinions, preferences, beliefs) that each agent possesses, with respect to a predefined set of features (variables, topics, issues). The dynamics is driven by social influence, which iteratively increases the similarity of the cultural vectors of pairs of interacting individuals. However, interactions are only allowed among pairs of individuals whose vectors are already closer than a certain (implicit or explicit) threshold distance, a mechanism known as *bounded confidence* and having its

origins in the so-called 'assimilation-contrast theory' [14] in social science. The intuition behind the model, successfully confirmed via numerical simulations and analytic calculations, is that social influence increases cultural similarity, yet full convergence is precluded by bounded confidence. The net result is the emergence of a certain number of *cultural domains*, each containing several individuals with identical cultural vectors and mutually separated by a distance larger than the bounded confidence threshold, thus no longer interacting with each other. The value of the model is the identification of a viable, decentralized mechanism according to which cultural diversity can persist at a global (inter-domain) scale, even if it vanishes at a local (intra-domain) scale.

Given the focus on the qualitative aspect of such an emergent pattern, the Axelrod model has been traditionally studied by specifying uniformly random initial conditions for the cultural vectors of all individuals, i.e. by drawing each cultural trait independently from a probability distribution that is flat over the set of possible realizations. Consistently with this uninformative (and deliberately unrealistic) choice, the focus of many studies has been the characterization of the outcomes of the model that are robust upon averaging over multiple realizations of the initial randomness. Since the cultural dynamics evolving the initial state is also stochastic, a second average over the dynamics is also required. We may therefore say that this is the 'annealed' version of the model. Examples of quantities that are stable across multiple realizations of uniformly random initial conditions are the expected *number* and expected *size* of final cultural domains. An obvious counter-example is the *values* of the vectors ending up in such domains: as follows from the complete symmetry in cultural space implied by the uniformity of the initial randomness, such values are by construction maximally unpredictable.

On the other hand, recent studies have investigated the model starting from different classes of initial conditions, beyond the uniformly random one. In particular, emphasis has been put on using initial conditions constructed from empirical data [15–17] and their randomized, trait-shuffled counterparts—obtained by randomly shuffling, for each component of the cultural vectors, the empirical values (traits) of all individuals in the sample. These studies have emphasized a strong dependence of the final outcome on the initial conditions. For instance, certain model outcomes that have an interesting interpretation in terms of enabling the coexistence of short-term social collective behavior and long-term cultural diversity [15] (more details are provided later in this paper) are found to vary significantly across the classes of empirical, trait-shuffled, and uniformly random initial conditions, while remaining largely stable when considering different instances belonging to the same class. This stability implies that empirical cultural data share certain remarkably universal properties, independent of the specific sample considered and at the same time significantly different from those exhibited by random and randomized data [17]. This has stimulated the introduction of stochastic, structural models aimed at capturing the essential properties of the empirical cultural data [16, 18].

Strong dependence of cultural dynamics on the initial conditions might be a useful property to exploit in the light of the increasing interest towards forecasting social and cultural behavior in the real world. Examples include the predictability of certain aspects of political elections, public campaigns, spreading of (fake) news, financial bubbles and crashes, and commercial success of new items. If interest is shifted towards the predictability of future long-term outcomes given certain initial conditions, then a corresponding change of perspective is implied at the level of modeling. In particular, the aforementioned 'annealed' framework, where the outcome of models of cultural dynamics is averaged over multiple realizations of the initial randomness, becomes less relevant. On the contrary, if a specific (e.g. empirical) initial condition is known, it becomes natural to use it as the single initial specification of the heterogeneity of the system. Obviously, averaging with respect to different random trajectories of the social influence dynamics, all starting from the same initial cultural state, remains important and necessary. We may therefore call this the 'quenched' version of the model.

In this work we focus for the first time on the predictability of the *social content* of the cultural domains in the final state of the Axelrod model, given a certain initial state. By social content we mean the composition of the different domains in terms of individuals, i.e. we are interested in forecasting 'who ends up in which cultural domain'. It should be noted that the social content is one of those properties that, just like the values of the final cultural vectors, is maximally unpredictable when considering the usual annealed model under uniformly random initial conditions. By contrast, we consider the quenched scenario starting from specific initial conditions sampled from empirical, shuffled, random, and an additional 'ultrametric' class of initial conditions.

We find that, remarkably, empirical and random initial conditions are associated with the highest and, respectively, lowest degree of predictability, which we rigorously define in an information-theoretic sense. This means that, as compared with the usual uniform specification of the initial conditions of the model, empirical data allow for a much more reliable forecast of the identity of the individuals forming the final cultural domains. We find that this result follows from the fact that the hierarchical, ultrametric-like organization of empirical cultural vectors, when coupled with bounded confidence, largely confines cultural convergence within the lower levels of the hierarchy. This result is confirmed using surrogate data that, while retaining only the ultrametric representation of real data, are also found to be associated with a higher predictability with respect to the shuffled and random conditions. The predictability associated to random and randomized cultural vectors is lower because it is difficult to identify a meaningful and robust hierarchical structure within the lower levels of which

social influence remains confined. The analysis gives similar results for all the empirical datasets considered here, pointing out the generic nature of these findings.
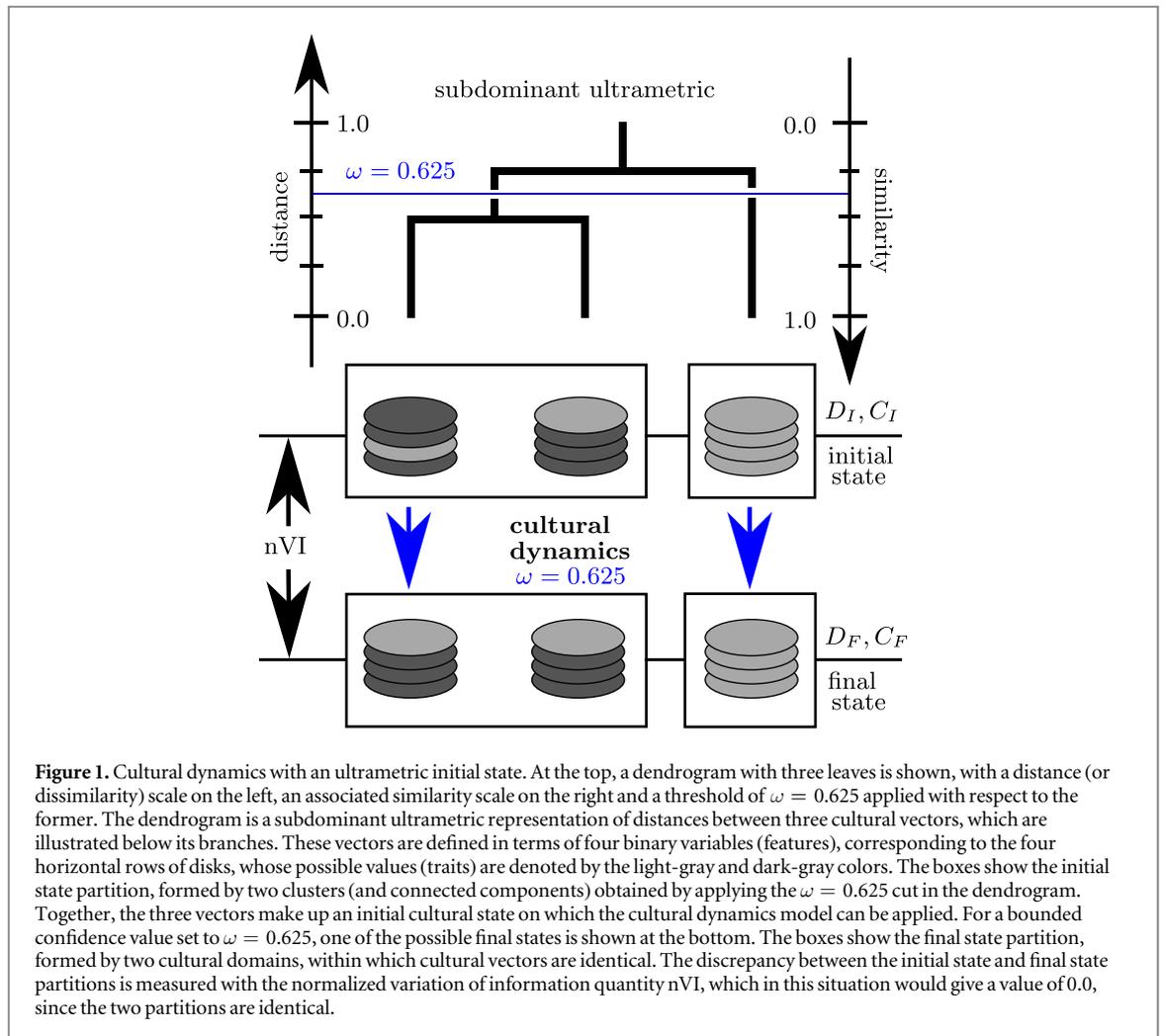
Even if we do not perform an explicit analysis of the *cultural content* of the final domains (the cultural traits that are perfectly shared by the individuals within every final group), the finding that their social content is predictable (the set of individuals within every final group), coupled with the fact that the initial cultural vectors of all individuals are known, implies that each final cultural vector will be a mixture of the traits of the initial vectors of the individuals ending up in the same cultural domain. This means that, the higher the predictability of the social content, the higher that of the cultural content as well. The take-home message is that the empirical hierarchical organization of culture and its ultrametric representation are very informative and useful for forecasting the outcome of models of cultural dynamics.

2. Ultrametricity and culture

The notion of ultrametricity refers to sets of objects that are hierarchically organized in certain abstract spaces, with applications in various fields, including mathematics (*p*-adic numbers), evolutionary biology (phylogenetic trees) and statistical physics (spin glasses) [19]. In practice, an ultrametric representation can be produced as the output of a hierarchical clustering algorithm applied to a matrix of pairwise distances between objects [19]. For the purpose of this work, these objects are the cultural vectors, whose pairwise cultural distances are computed in the same manner as in [15–18], based on a combination between the Hamming distance and the Manhattan distance, which are used in association with nominal and ordinal cultural features, respectively—also see equation (A2) and the associated description for more details. The following explanations concerning ultrametricity are mostly restricted to cultural vectors, although many of the concepts have a wide range of applicability. The ultrametric representation of N cultural vectors can be visualized as a dendrogram (a binary hierarchical tree; see the top of figure 1) with N leaves (one for each vector) and $N-1$ branching points (often referred to as 'branchings', for simplicity), sorted by $N-1$ real numbers that are attached to them. These numbers can be defined in two equivalent ways: on a distance scale (top-left axis) or on a similarity scale (top-right axis)—both quantities take values between 0.0 and 1.0, while adding up to 1.0. Each number is an approximation for distances between leaves that are first merged at the respective branching point. These $N-1$ numbers and the topology of the dendrogram retain part of the information inherent in the cultural distance matrix (which is specified by $N(N-1)/2$ numbers), so the dendrogram is an approximation of this matrix. The approximation is exact and algorithm-independent only when the original distances are perfectly ultrametric: a stronger version of the triangle inequality is satisfied for all triplets of distinct objects [19]. A cut can be performed at a certain height ω in the dendrogram, providing an ω-dependent partition of the N cultural vectors (see figure 1)—most of the results shown in this study involve a systematic exploration of the meaningful ω interval. For a dendrogram obtained via the single-linkage hierarchical clustering algorithm (see [20] and references therein), the ω-dependent partition is the same as that encoding the connected components obtained by applying an ω-threshold to the initial matrix of distances.

Reference [15] pointed out that a dendrogram approximating an empirical cultural state shows a clearer hierarchical organization than dendrograms approximating the shuffled or random counterparts, suggesting that the ultrametric representation is better suited for empirical data than for shuffled or random data. In addition, cultural dynamics (with a built-in threshold) applied to the empirical cultural state appeared to mostly induce convergence within the groups of the ω-dependent partition, if ω is equal to the bounded confidence threshold used in the cultural dynamics model (see below), where an identification is made between this threshold and the cut on the dendrogram. These observations were made in a qualitative way, by visually inspecting dendrograms obtained with the average-linkage hierarchical clustering algorithm [21, 22]. Instead, we perform here a systematic, quantitative comparison between ω-dependent partitions of initial cultural states and associated partitions of final states resulting from cultural dynamics, for different classes of initial cultural states—the 'variation of information' quantity is used for this purpose, as explained below. The initial-state ω-dependent partitions are always extracted from the dendrogram provided by the single-linkage algorithm [20], rather than the average-linkage one, since it provides the subdominant ultrametric, which is the 'closest below' the original distances and unique [23], while also being equivalent to the hierarchical connected-component representation, as mentioned above. This choice is also common for the purpose of evaluating measures of ultrametricity, like the cophenetic correlation coefficient, which is done in [16].

In this study we also propose a new class of 'ultrametric' initial states, based on a stochastic generation procedure that enforces the ultrametric representation of a given empirical state. Specifically, this procedure provides, for every run, a set of N cultural vectors whose pairwise distances reproduce, on average, the pairwise distances encoded by the subdominant ultrametric representation of an empirical set of cultural vectors of the same N. This is achieved using an extension of a method originally proposed in [24], in the context of DNA

Figure 1. Cultural dynamics with an ultrametric initial state. At the top, a dendrogram with three leaves is shown, with a distance (or dissimilarity) scale on the left, an associated similarity scale on the right and a threshold of $\omega = 0.625$ applied with respect to the former. The dendrogram is a subdominant ultrametric representation of distances between three cultural vectors, which are illustrated below its branches. These vectors are defined in terms of four binary variables (features), corresponding to the four horizontal rows of disks, whose possible values (traits) are denoted by the light-gray and dark-gray colors. The boxes show the initial state partition, formed by two clusters (and connected components) obtained by applying the $\omega = 0.625$ cut in the dendrogram. Together, the three vectors make up an initial cultural state on which the cultural dynamics model can be applied. For a bounded confidence value set to $\omega = 0.625$, one of the possible final states is shown at the bottom. The boxes show the final state partition, formed by two cultural domains, within which cultural vectors are identical. The discrepancy between the initial state and final state partitions is measured with the normalized variation of information quantity nVI, which in this situation would give a value of 0.0, since the two partitions are identical.

sequences. The generalization introduced here allows the method to work with combinations of features of different ranges and types, where the range stands for the number of traits and the type indicates whether the feature is treated as ordinal or nominal. The method is described in detail in appendix A.

Figure 1 illustrates the concepts that are most relevant for this study and the relationships between them. At the center, the figure shows an initial cultural state with 3 vectors, defined in terms of 4 binary features, with possible traits (values) denoted by the two shades of gray. Each of the three vectors is matched to a branch of the dendrogram drawn at the top, which encodes the subdominant ultrametric representation of the initial cultural state. For this specific case, the distance between the first two vectors is 0.5, while the distances between any of these two and the third are 0.75, which together make up a perfectly ultrametric discrete space, thus exactly matching the distances encoded by the dendrogram. The horizontal line denotes a possible ω-cut that can be applied to the dendrogram, which induces a splitting into two (in the example shown) branches and two associated subsets of vectors, which together form an ω-dependent partition (or clustering) of the initial set. This partition is the same as that induced by the set of connected cultural components of the ω-threshold cultural graph. At the bottom, the figure shows one possible final state resulting from the cultural dynamics process, for a bounded confidence threshold set to the same ω value as the dendrogram cut. The groups of identical vectors constitute another, ω-dependent partition characterizing the cultural state, which exactly matches, in this case, the initial state partition. Other final configurations are possible, due to the stochastic nature of cultural dynamics. It is even possible, although unlikely, that by a succession of convenient interactions the second vector 'migrates' from the group on the left to the one on the right during the dynamics. The abundance of such deviations is quantitatively studied below, for several classes of initial conditions.

3. Cultural dynamics and partition-specific quantities

Every cultural dynamics process simulated in this study starts with an initial cultural state, consisting of a set of N cultural vectors, each associated to one of the N agents in the model—see center of figure 1. Four classes of initial

cultural states are used in this study, which have already been mentioned above: empirical, ultrametric, shuffled and random cultural states. However, any ultrametric, shuffled or random state is generated in a stochastic way, conditionally on a given empirical state, so one could say that these three classes are composite, each one being a collection of statistical ensembles, one for each empirical state. Most of this study focuses on an empirical cultural state constructed from Eurobarometer 38.1 [25, 26] data (collected via face-to-face interviews with people in the EU), formatted according to the procedure in [17], whose cultural features are associated to survey questions dealing with opinions on various topics concerning science, technology, the environment and the European community. The associated ultrametric cultural state is generated using the new procedure mentioned in section 2 and explained in detail in appendix A, which retains, in a certain sense, the empirical ultrametric structure. The associated shuffled cultural state is obtained by randomly and independently permuting the empirical cultural traits among vectors, with respect to every feature, thus exactly enforcing all the empirical trait frequencies. Finally, the associated random state is obtained by drawing each trait at random, from a uniform probability distribution with respect to every feature, while only retaining the empirical data format—the number of features, together with the range and type of each feature—and thus the associated cultural space, which is also retained by the ultrametric and the shuffled states. Part of this study makes use of three other empirical states (constructed from other datasets) and of the associated ultrametric, shuffled and random states—see section 4 and appendix B.

Cultural dynamics is simulated here using a simple, Axelrod-type model, without any underlying geometry for a social network or a geographical-physical space: essentially, all N agents are connected to each other. Instead, an explicit bounded-confidence threshold ω is present, which defines the maximum cultural distance for which social influence interactions can successfully occur—further convergence occurs only if there is already some level of overlap. At each simulation step, two agents are randomly picked. If the distance d_{ij} between their cultural vectors is smaller than ω and if these vectors are different with respect to at least one feature, then an interaction successfully occurs with probability $1 - d_{ij}$: one of the agents switches its trait to match the trait of the other agent, with respect to one of the features that differentiates between them. This is exactly the model used in [15, 17, 18] and partly in [16]. As anticipated in sections 1 and 2, this model converges to a random final, absorbing state, one that consists of groups (cultural domains) of internally identical and externally non-interacting cultural vectors—distances within such groups are zero, while distances across are larger or equal to ω, as illustrated at the bottom of figure 1.
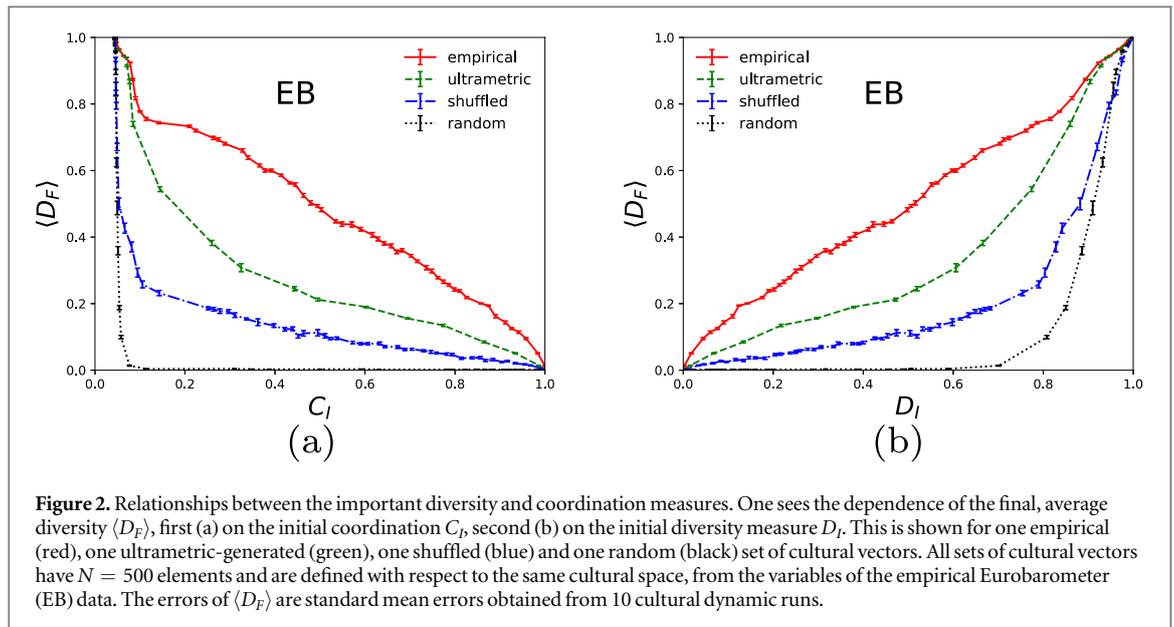
All calculations performed in this study are heavily based on the partitions characterizing the initial and final cultural states, consisting of initial dendrogram-based clusters (the connected components) and final groups of identical vectors, respectively, as explained in section 2. As illustrated in figure 1, each type of partition is characterized by two types of quantities, denoted by (D_I, C_I) for initial partitions and by (D_F, C_F) for final partitions. These quantities are referred to as the coordination (C_I and C_F) and the diversity measures (D_I and D_F). They are computed according to the following formulas:

$$D_a(\omega) = \frac{N_C^a(\omega)}{N}, \quad C_a(\omega) = \sqrt{\sum_A \left(\frac{S_A^a}{N}\right)^2}_\omega, \quad (1)$$

where $a \in \{I, F\}$ distinguishes between 'initial' and 'final', N_C^a is the number of groups (connected components if $a = I$, domains of identical vectors if $a = F$), and S_A^a is the size of group A for the given ω value. Note that D_a is a measure of diversification, while C_a is a measure of non-homogeneity inherent in the respective partition. Moreover, since cultural dynamics is a stochastic process, it is meaningful to talk about averages over final state partitions (over multiple dynamical runs), which is particularly useful for the final diversity measure $\langle D_F(\omega) \rangle$.

The $\langle D_F(\omega) \rangle$ quantity has been interpreted as a measure of propensity to long-term cultural diversity, while the $C_I(\omega)$ has been interpreted as a measure of propensity to short-term collective behavior [15, 17]. Through their common dependence on ω, the correspondence between the two quantities is graphically illustrated in figure 2(a). Along each curve, different points correspond to different ω-values, while different curves correspond to different classes of initial conditions. It is clear that the empirical cultural state allows for much more compatibility between the aspects measured by the two quantities than the shuffled and the random cultural state, as pointed out in [15]. In fact, this is the analysis used in [15] to highlight the structure of empirical cultural data and in [17] to emphasize the universality of this structure, except for the 'ultrametric' scenario, explained in section 2, which is first introduced here. Note that the ultrametric cultural state comes closer to the empirical behavior than the shuffled cultural state, suggesting that empirical ultrametric is better than empirical trait frequencies at explaining the generic empirical structure.

For the same four sets of cultural vectors used in figure 2(a), the average final diversity $\langle D_F(\omega) \rangle$ is plotted against the initial diversity $D_I(\omega)$ in figure 2(b). This visualization, previously used [15, 16] without the ultrametric scenario, illustrates the extent to which cultural dynamics preserves the number of groups when going from the initial to the final partition. As observed before, the number of groups is well preserved by

Figure 2. Relationships between the important diversity and coordination measures. One sees the dependence of the final, average diversity $\langle D_F \rangle$, first (a) on the initial coordination C_I, second (b) on the initial diversity measure D_I. This is shown for one empirical (red), one ultrametric-generated (green), one shuffled (blue) and one random (black) set of cultural vectors. All sets of cultural vectors have $N = 500$ elements and are defined with respect to the same cultural space, from the variables of the empirical Eurobarometer (EB) data. The errors of $\langle D_F \rangle$ are standard mean errors obtained from 10 cultural dynamic runs.
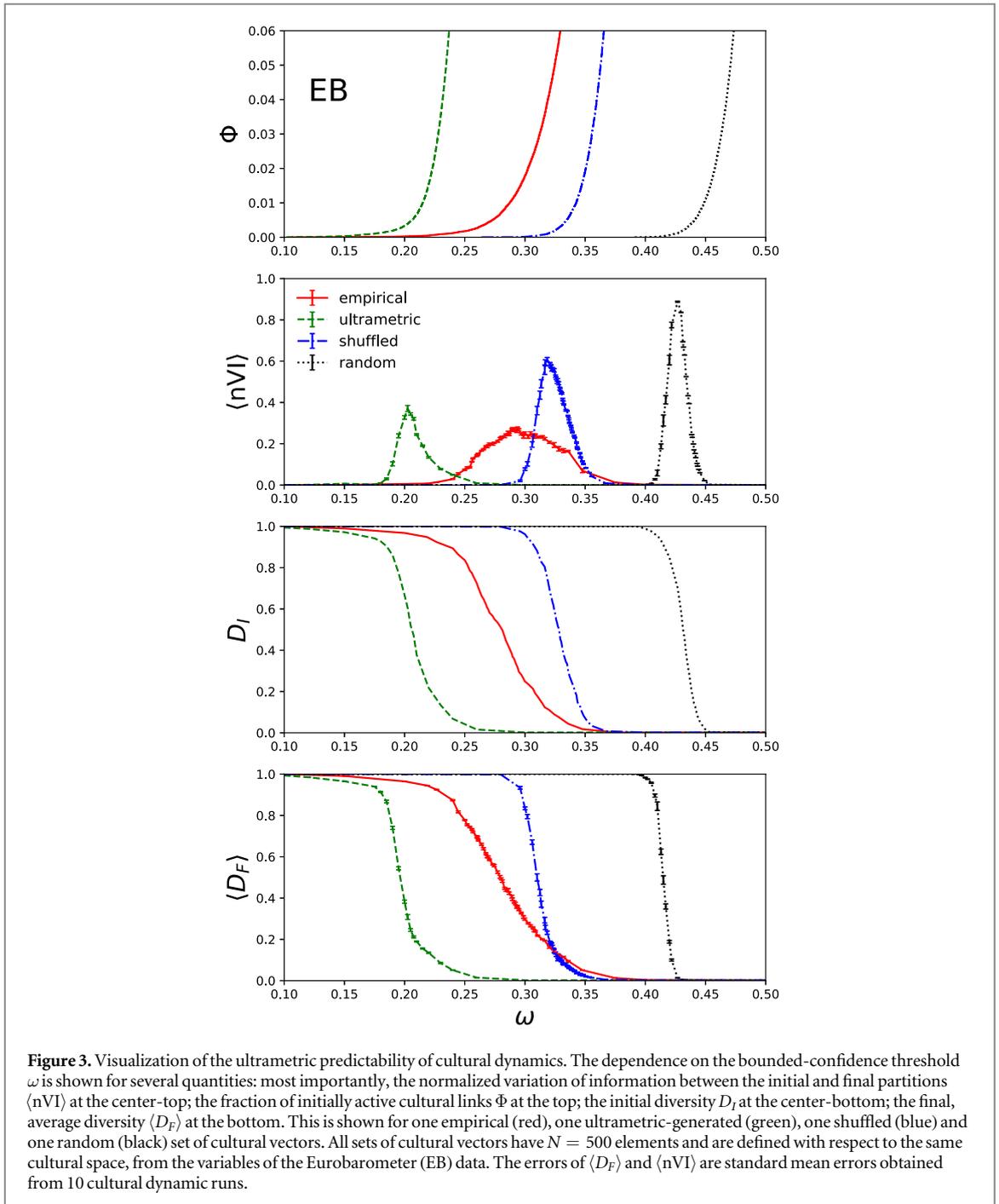
cultural dynamics acting on empirical data, which happens much less for shuffled data and even less for random data. This goes along with the idea that the final partition can be predicted from the initial partition if empirical data is used for specifying the latter. Note that, like in figure 2(a), ultrametric-generated data lies in between the empirical and shuffled scenarios, confirming that the subdominant ultrametric information, which is directly related to the sequence of ω-dependent initial partitions, is rather robust with respect to cultural dynamics.

Although informative, the comparison between the $\langle D_F(\omega) \rangle$ and $D_I(\omega)$ is incomplete as a way of assessing the predictability of the final partition from the initial partition: two partitions might have the same number of groups, but the sizes and/or contents of these groups might be very different. In order to take all this into account in a consistent way, the discrepancy between the initial and final state partitions is evaluated using the variation of information measure VI, as a function of ω. This is an information-theoretic measure that acts as a metric distance within the space of possible partitions of a set of N elements, which has been shown to have a multitude of advantages compared to other possible measures [27]. It is convenient to work with the normalized version of this quantity $nVI(\omega) = VI(\omega)/\log(N)$ (also mentioned in figure 1), which retains the meaning and metricity of the original quantity, as long as N remains the same ($N = 500$ for all cultural states studied here). This quantity is very important for section 4.

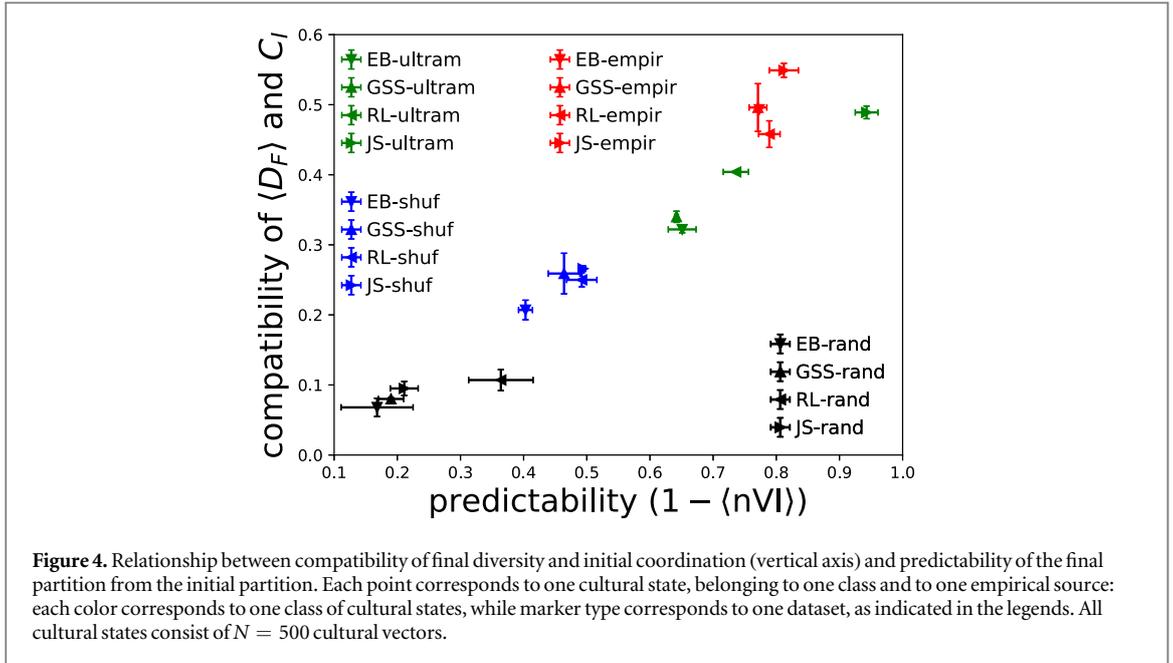## 4. Predictability of the final state

This section focuses on evaluating the predictability of the final state partition from the initial state partition. This is done using the (normalized) variation of information quantity $\langle nVI \rangle$ mentioned above, which measures the discrepancy between the two partitions: predictability is higher when $\langle nVI \rangle$ is lower. The dependence of $\langle nVI \rangle$ on ω is shown in the second panel of figure 3, for the same 4 cultural states used in figure 2, where the averaging is performed over multiple dynamical runs, like for the $\langle D_F \rangle$ quantity. The empirical state shows the lowest maximal $\langle nVI \rangle$ value, followed by the ultrametric, the shuffled and the random states. This shows that the outcome of cultural dynamics can be predicted relatively well based on the initial state, if this is constructed from empirical data and comparably well if this is constructed based on the empirical ultrametric information. On the other hand, shuffled and random data exhibit lower predictability. Note that, for either scenario, $\langle nVI \rangle$ vanishes for the low-ω and the high-ω regions, which is where both the initial and final partitions consist of N single-object groups and of one N-objects group, respectively. This can be understood by looking at the dependence of the D_I and $\langle D_F \rangle$ quantities on ω shown in the third and fourth panels: the ω region for which $\langle nVI \rangle$ is significantly larger than 0.0, thus signaling some discrepancy between the initial and final partitions, is roughly the region where either D_I or $\langle D_F \rangle$ is substantially different from 1.0 or 0.0.

In parallel, the first panel of figure 3 shows the ω-dependence of the fraction of initially active cultural links Φ: the fraction of pairs (i, j) of cultural vectors whose distance $d_{ij} < \omega$ in the initial state. This shows that the ω interval that is non-trivial with respect to D_I, $\langle D_F \rangle$ and $\langle nVI \rangle$ seems to be largely determined by the shape of Φ, which is nothing else than the cumulative distribution of intervector distances. The properties of this distribution—average lower for empirical data than for random data, standard deviation higher for empirical data than for either shuffled or random data—have been studied before [15, 16] and are recognizable in the first

Figure 3. Visualization of the ultrametric predictability of cultural dynamics. The dependence on the bounded-confidence threshold ω is shown for several quantities: most importantly, the normalized variation of information between the initial and final partitions $\langle nVI \rangle$ at the center-top; the fraction of initially active cultural links Φ at the top; the initial diversity D_I at the center-bottom; the final, average diversity $\langle D_F \rangle$ at the bottom. This is shown for one empirical (red), one ultrametric-generated (green), one shuffled (blue) and one random (black) set of cultural vectors. All sets of cultural vectors have $N = 500$ elements and are defined with respect to the same cultural space, from the variables of the Eurobarometer (EB) data. The errors of $\langle D_F \rangle$ and $\langle nVI \rangle$ are standard mean errors obtained from 10 cultural dynamic runs.

panel of figure 3. Note that, for the ultrametric scenario, the interesting ω region and the Φ profile are compressed in a lower-ω region compared to empirical data. This means that the branchings in the dendrogram obtained from ultrametric-generated data occur at lower ω-values than those in the dendrogram obtained from the original, empirical data. In turn, this is due to the distances between the ultrametric-generated cultural vectors reproducing, on average, the subdominant ultrametric empirical distances, rather than the original empirical distances, while the former are known to systematically underestimate the latter, particularly for higher distance values, as long as the empirical vectors are not perfectly ultrametric—in practice they are never perfectly ultrametric.

There is another aspect that can be noted when comparing, for either scenario, the shape of $\Phi(\omega)$ in the first panel with the shape of $D_I(\omega)$ in the third panel of figure 3: as ω is decreased, most of the cultural links need to be eliminated in order to reach the abrupt region of the $D_I(\omega)$ transition, for which the number of groups in the initial partition becomes comparable to N. This is not surprising on general grounds. For instance, the Erdős– Rényi model of random graphs [28] exhibits a critical link density of $1/N$, at which a giant connected component is present, if N is the number of nodes in the graph, instead of the number of cultural vectors. Still, this analogy

Figure 4. Relationship between compatibility of final diversity and initial coordination (vertical axis) and predictability of the final partition from the initial partition. Each point corresponds to one cultural state, belonging to one class and to one empirical source: each color corresponds to one class of cultural states, while marker type corresponds to one dataset, as indicated in the legends. All cultural states consist of $N = 500$ cultural vectors.

should not be taken too far. The random graph interpretation is closest to the random cultural state scenario used here, since the expected pairwise distance entailed by the latter is the same for any pair of cultural vectors, just like the connection probability entailed by the former is the same for any pair of nodes. However, even the random scenario has an underlying metric structure, due to how cultural spaces are defined [17], which should introduce more triangles than expected otherwise, while the shuffled and empirical scenarios are additionally affected by inhomogeneities in their cultural space distributions.

The analysis presented in figures 2 and 3 was repeated for three other empirical datasets—based on each dataset, one empirical, one ultrametric, one shuffled and one random cultural state are constructed—with very similar results. These additional datasets are: the General Social Survey (GSS) [29] 1993, recording opinions on a variety of topics from people in the US, via face-to-face interviews; Jester [30], recording online ratings of jokes; the Religious Landscape (RL) [31], recording opinions on several religious but also political topics from people in the US, via telephone interviews. The details concerning the formatting of these three datasets are also present in [17]. For all four datasets, the results are presented in a joint, compact manner by means of figure 4, while more detailed results are shown in appendix B. Each of the points in the figure corresponds to a combination of one dataset and one scenario. The vertical axis corresponds to a measure of compatibility between long-term cultural diversity $\langle D_F \rangle$ and short-term collective behavior C_I, namely a measure of the overall departure of the $\langle D_F \rangle$ versus C_I curve from the lower-left corner in figure 2(a). The horizontal axis corresponds to a measure of predictability of the final state from the initial state, namely an inverse measure of the overall departure of the $\langle nVI \rangle$ versus ω from the horizontal axis in the second panel of figure 3.

For both measures, simple definitions are employed: rather than integrating information from every ω value for which some departure is present, both definitions conceptually rely only on one, representative ω^* point, for which both departures are relatively high. Specifically, ω^* is defined by intersecting the $\langle D_F \rangle$ versus C_I curve with the main diagonal $\langle D_F \rangle = C_I$. In practice, since just a finite number of ω-values are available for any combination of dataset and scenario, one uses instead the two ω-values that are closest to the main diagonal of the $\langle D_F \rangle$ versus C_I plot from either of the two sides. These two values, labeled as ω_L and ω_R, 'bracket' ω^* from the left and right, respectively: $\omega_L < \omega^* < \omega_R$. The ω^* itself is never explicitly calculated, but is conceptually useful for the explanations below.

The compatibility approximates the distance between the $(\langle D_F(\omega^*) \rangle$ versus $C_I(\omega^*))$ point and the $(\langle D_F \rangle = 0, C_I = 0)$ point, normalized by the length of the main diagonal of the $\langle D_F \rangle$ versus C_I plot. In practice, this is evaluated in terms of ω_L and ω_R according to:

$$\frac{\sqrt{\langle D_F(\omega_L) \rangle^2 + C_I^2(\omega_L)} + \sqrt{\langle D_F(\omega_R) \rangle^2 + C_I^2(\omega_R)}}{2\sqrt{2}},$$

while the associated error is evaluated as:

$$\frac{\sqrt{\langle D_F(\omega_L)\rangle^2 + C_I^2(\omega_L)} - \sqrt{\langle D_F(\omega_R)\rangle^2 + C_I^2(\omega_R)}}{2\sqrt{2}}.$$

The predictability approximates the distance between the $(\omega^*, \langle \mathrm{nVI}(\omega^*)\rangle)$ point and the $\langle \mathrm{nVI}\rangle = 1$ line. In practice, this is evaluated as:

$$1 - \frac{\langle \mathrm{nVI}(\omega_L)\rangle + \langle \mathrm{nVI}(\omega_R)\rangle}{2},$$

while the associated error is evaluated as:

$$\frac{|\langle \mathrm{nVI}(\omega_L)\rangle - \langle \mathrm{nVI}(\omega_R)\rangle|}{2}.$$

Note that compatibility increases with predictability in a roughly linear way, at least for the cultural states considered here. Moreover, cultural states belonging to the same class tend to cluster together in the compatibility-predictability space. A notable exception is ultrametric-Jester, which is significantly outside the ultrametric class in terms of predictability, showing higher predictability than any of the empirical states. Still, it is clear that cultural states that are closer to the universal $\langle D_F\rangle$ versus C_I empirical behavior also allow for better estimates of the final partition from the initial one.

The observed increase of compatibility with predictability provides some insights about the nature of empirical data, or at least about the shape of an empirical-like dendrogram characteristic for the upper-right corner of figure 4. This can be understood by realizing that the ultrametric and empirical states approach an ideal, limiting situation of perfect predictability, for which the initial and final partitions are identical irrespective of ω. This implies that $\langle D_F(\omega)\rangle = D_I(\omega)$ and consequently that the $\langle D_F\rangle$ versus C_I curve is essentially the D_I versus C_I curve and thus controlled by the geometry of the subdominant ultrametric dendrogram. One can then show—see appendix C—that this geometry needs to be highly 'unbalanced' in order to explain the close-to-linear $\langle D_F\rangle \approx 1 - C_I$ empirical behavior in figure 2(a) and the compatibility values of approximately 0.5 following from it. For a perfectly-unbalanced geometry, the kth highest dendrogram branching separates only one leaf from the remaining $N - k$, for all $k \in \{1, \ldots, N - 1\}$. By contrast, a perfectly-balanced geometry entails a splitting into two, equal clusters for each dendrogram branching, which would induce an inverse square $\langle D_F\rangle \propto C_I^{-2}$ behavior—see appendix C—closer to that of shuffled and random cultural states, with a lower compatibility value. Thus, while going from the random to the empirical class, by enforcing more and better empirical information, the increasing level of compatibility becomes more suggestive of an unbalanced dendrogram geometry, while the increasing level of predictability increases the reliability of this geometric interpretation.

5. Conclusion

This study focused on the ultrametric representation of sets of cultural vectors used for specifying the initial state of cultural dynamics models. On one hand, it introduced another procedure for randomly generating initial conditions based on the subdominant ultrametric information of empirical data. On the other hand, it examined the extent to which the subdominant ultrametric representation can be used for predicting the final state of cultural dynamics in a simple theoretical setting. The bounded-confidence threshold parameterizing the dynamical model was used to extract an initial-state partition from the ultrametric representation. This was systematically compared, in terms of variation of information, with the corresponding final state partition consisting of groups of identical cultural vectors. The comparison showed that the predictive power of the ultrametric is relatively high for empirical cultural states, which are closely followed by ultrametric-generated states, which are followed by the shuffled and then by the random states. Moreover, higher predictability appears to go hand in hand with higher compatibility between a propensity to long-term cultural diversity and a propensity to short-term collective behavior, which was previously shown to be a hallmark of empirical structure. This means that ultrametric information is better than trait-frequency information at explaining this structure. These results further advance the understanding of the relationship between ultrametricity and cultural dynamics. Moreover, it is tempting to speculate that, for the purpose of forecasting the dynamics of culture in the real world, knowledge about the current distribution of individuals in cultural space might be sufficient, with little or no need for running simulations, at least if one assumes that consensus-favoring social influence is the essential driving force of this dynamics. The importance of these findings is further enhanced by two aspects: first, the results are highly robust across different empirical sources; second, the empirical data used here is entirely independent of assumptions about opinion-changing interactions between people, which only come into play at the level of dynamical models using such data for their initial conditions.

Appendix A. Ultrametric-generation method

This section explains the method for generating sets of cultural vectors belonging to the 'ultrametric' class. The method is an extension of that developed in [24], so the description is also somewhat similar, although the nomenclature specific to cultural vectors is used, instead of that specific to genetic sequences.

The method takes as input a dendrogram, as well as a target cultural space—the number of cultural features F, together with the range (number of traits) q and type (nominal or ordinal) of each feature. This information is taken from empirical data and the single-linkage hierarchical clustering algorithm is employed for constructing the dendrogram whenever the method is used in this study. Upon every use, the method generates, in a stochastic way, a set of N cultural vectors associated to the N leaves of the dendrogram, such that, on average, the pairwise similarities between cultural vectors match the similarities encoded by the dendrogram.

More precisely, for each cultural feature in the target space, the method enforces:

$$E[s_{ij}^q] = \rho_{\alpha_{ij}}, \tag{A1}$$

where $E[\ldots]$ stands for 'expectation value', α_{ij} is the lowest branching in the dendrogram joining leaves i and j, $\rho_{\alpha_{ij}}$ is the similarity encoded by this branching and s_{ij}^q is the partial contribution to the similarity between cultural vectors i and j of a feature of range q, which is computed according to the following formula:

$$s_{ij}^q = \begin{cases} \delta(x_i^k, x_j^k) & \text{if nominal,} \\ 1 - \dfrac{|x_i^k - x_j^k|}{q^k - 1} & \text{if ordinal,} \end{cases} \tag{A2}$$

which depends on whether the feature is nominal or ordinal, where δ stands for the Kroneker delta function, x_i^k and x_j^k are the traits of vectors i and j with respect to feature k with range q^k—for ordinal features k, the traits are marked with integers between 1 to q^k. Equation (A2) is consistent with the cultural distance definition in [15–18] (as mentioned above: similarity $= 1.0 -$ distance).

In equation (A1), the expectation $E[\ldots]$ implies averaging over multiple runs of the method, for the same dendrogram and the same cultural feature. Although in practice the method is used only once (and independently) for each feature, the fact that a large number F of features are present makes this approach sensible: the expectation $E[s_{ij}]$ of the complete similarity s_{ij} will also match $\rho_{\alpha_{ij}}$ (since the complete similarity is the arithmetic average of the feature-level similarities), while the fluctuations of s_{ij} around $\rho_{\alpha_{ij}}$ will decrease with F. In other words, as pointed out in [24], the expectation in equation (A1) can be interpreted in two idealized ways: averaging over infinitely many runs or averaging over infinitely many features.

In order to enforce equation (A1) for every pair (i, j), the method controls for the extent to which the traits of different vectors are chosen independently of each other. For every feature, all the N chosen cultural traits originate in independent random draws from a uniform probability distribution, but the number of draws is smaller or equal to N. Thus, the traits of vectors i and j either originate in the same draw, with probability P_{ij}, or originate in different draws, with probability $1 - P_{ij}$. In the former case the two traits are identical, with a well-determined feature-level similarity $s_{ij}^q = 1$. In the latter case, the two traits may be identical or different, so that s_{ij}^q fluctuates around an expectation value $f(q)$. Taking both cases into account, the expectation value of s_{ij}^q is:

$$E[s_{ij}^q] = P_{ij} + [1 - P_{ij}]f(q), \tag{A3}$$

where the expectation for different draws $f(q)$ reads:

$$f(q) = \begin{cases} \dfrac{1}{q} & \text{if nominal,} \\ \dfrac{2q - 1}{3q} & \text{if ordinal,} \end{cases} \tag{A4}$$

which is the expression of the expected, feature-level similarity between two traits drawn at random from a uniform probability distribution, obtained analytically from equation (A2) for either type of features. The choices of traits and the associated random draws are managed by the stochastic-algorithmic part of the method (briefly explained at the end of this section), which is designed to ensure that:

$$P_{ij} = \rho^I_{\alpha_{ij}} \tag{A5}$$

is satisfied, where $\rho^I_{\alpha_{ij}}$ is a corrected version of the similarity $\rho_{\alpha_{ij}}$ implicit in the α_{ij} branching:

$$\rho^I_{\alpha_{ij}} = \rho_{\alpha_{ij}} - h(\rho_{\alpha_{ij}}, q), \tag{A6}$$

where h is a correction function chosen such that equation (A1) holds, subject to (A3) and (A5). Specifically, by combining equation (A5) with equation (A3) and then with equation (A1), one obtains:

$$\rho^I_{\alpha_{ij}} + [1 - \rho^I_{\alpha_{ij}}]f(q) = \rho_{\alpha_{ij}}. \tag{A7}$$

By inserting equation (A6) in equation (A7) and further manipulations, one obtains the following expression for the correction function:

$$h(\rho_\alpha, q) = \frac{1 - \rho_\alpha}{1 - f(q)}f(q). \tag{A8}$$

Note that equation (A5) identifies $\rho^I_{\alpha_{ij}}$ with a probability, meaning that $\rho^I_\alpha > 0$ should be satisfied for all branchings α. This implies, given equations (A6) and (A8), that $\rho_\alpha > f(q)$ for all branchings α of the given dendrogram and for all features in the target space. This condition needs to be satisfied in order for this method to be valid and is actually satisfied by all four empirical dendrograms used in this study. Also note that the method in [24] is recovered as a special case of the above, by restricting to nominal features of constant q via equation (A4).

Finally, it is worth describing the stochastic-algorithmic part of the method. For each of the F features in the target space, the following steps are carried out:

- the dendrogram is recursively explored starting with the root branching; for every branching α reached by this exploration, one of the following two things happens:

  one of the q traits is randomly chosen, according to a uniform distribution and assigned to all cultural vectors corresponding to leaves under branching α, without further exploring any branching below α;

  the exploration is continued with each of the two branches emerging from α, if that branch leads to another branching, instead of leading to a leaf;

  with probability Q_α for the former and probability $1 - Q_\alpha$ for the latter, where:

$$Q_\alpha = \frac{\rho^I_\alpha - \rho^I_{p(\alpha)}}{1 - \rho^I_{p(\alpha)}}, \tag{A9}$$

where $p(\alpha)$ is the parent branching of α, if α is not the root, while $\rho^I_{p(\alpha)} = 0$ if α is the root.
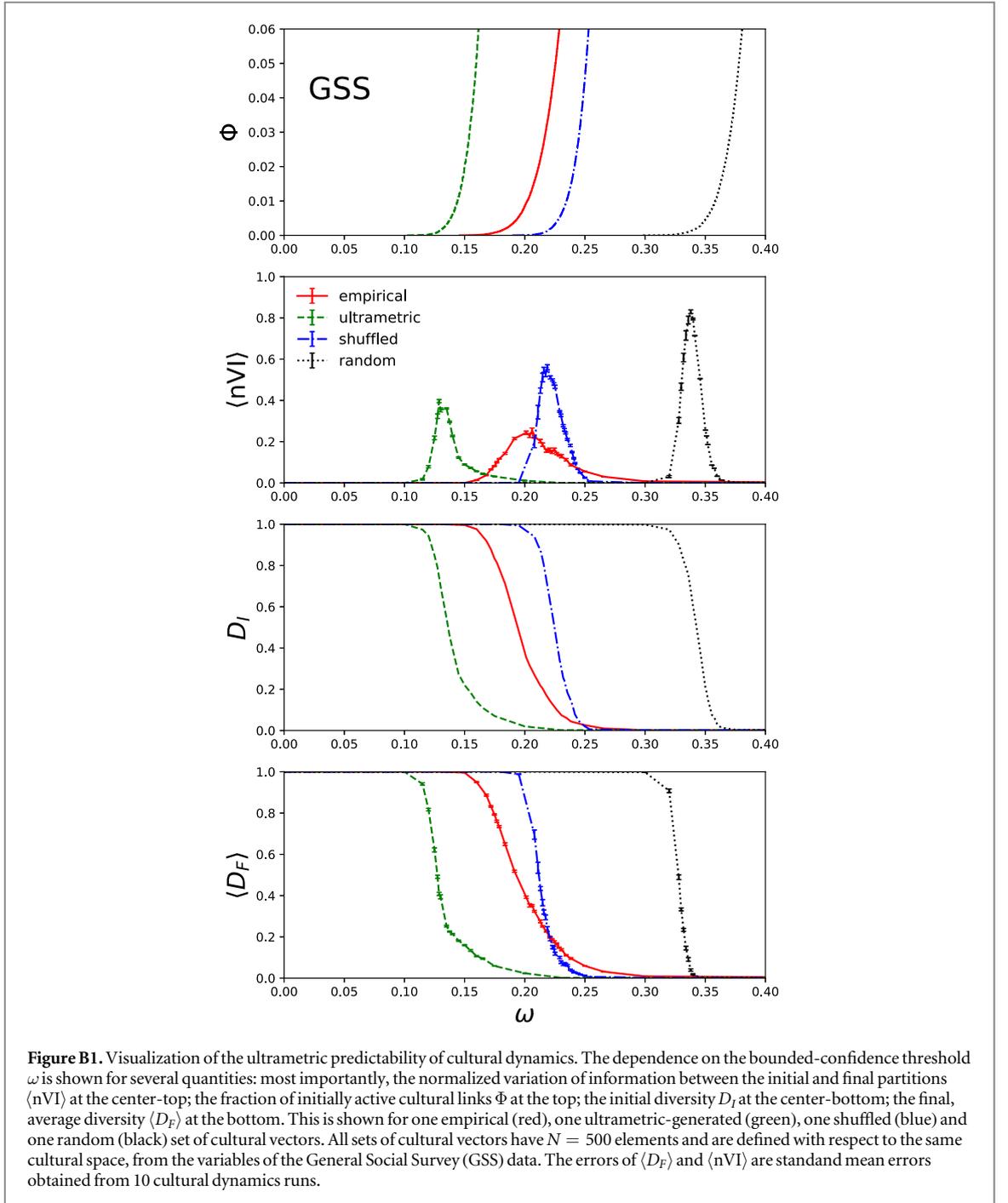
- for each of the leaves whose traits are not assigned during the above step, one of the q traits is randomly chosen, according to a uniform distribution and assigned to the respective cultural vector.

This algorithmic procedure ensures that equation (A5) holds, for reasons that are fully explained in [24].

It is worth noting that the ultrametric-generation method described in this section makes use of all the information inherent in the geometry of the dendrogram that it receives as input—both the topology and the similarities ρ encoded by the branching points of the dendrograms are used. However, the generated sets of cultural vectors will in general not be precisely ultrametric, in the strict mathematical sense [19] (unless it is applied in the limit of F being much larger than N). Still, they are generated based on the empirical ultrametric information and are arguably as close as they can be to reproducing the ultrametric set of pairwise distances.

Appendix B. Detailed results

This section shows the complete results concerning the ω-dependence of relevant quantities, for the other three datasets that are used in this study in addition to the Eurobarometer (EB [25, 26]): the GSS [29] data in figure B1, the RL [31] data in figure B2 and the Jester (JS [30]) data in figure B3. Each of these three figures follows the format of figure 3 above, with four panels and four scenarios. Although, for each type of scenario, there is a certain variability in the width and location of the non-trivial ω-interval, the results are qualitatively similar to those obtained for EB data, with a notable exception visible for the analysis of Jester data in figure B3: the second panel shows that the discrepancy between the initial and the final partition, as measured by $\langle nVI \rangle$, is clearly smaller for the ultrametric cultural state than for the empirical cultural state, so the overall predictability is higher. This is in agreement with the observation made in relation to figure 4 about the relatively high predictability value of the Jester-ultrametric point.

Figure B1. Visualization of the ultrametric predictability of cultural dynamics. The dependence on the bounded-confidence threshold ω is shown for several quantities: most importantly, the normalized variation of information between the initial and final partitions $\langle nVI \rangle$ at the center-top; the fraction of initially active cultural links Φ at the top; the initial diversity D_I at the center-bottom; the final, average diversity $\langle D_F \rangle$ at the bottom. This is shown for one empirical (red), one ultrametric-generated (green), one shuffled (blue) and one random (black) set of cultural vectors. All sets of cultural vectors have $N = 500$ elements and are defined with respect to the same cultural space, from the variables of the General Social Survey (GSS) data. The errors of $\langle D_F \rangle$ and $\langle nVI \rangle$ are standand mean errors obtained from 10 cultural dynamics runs.
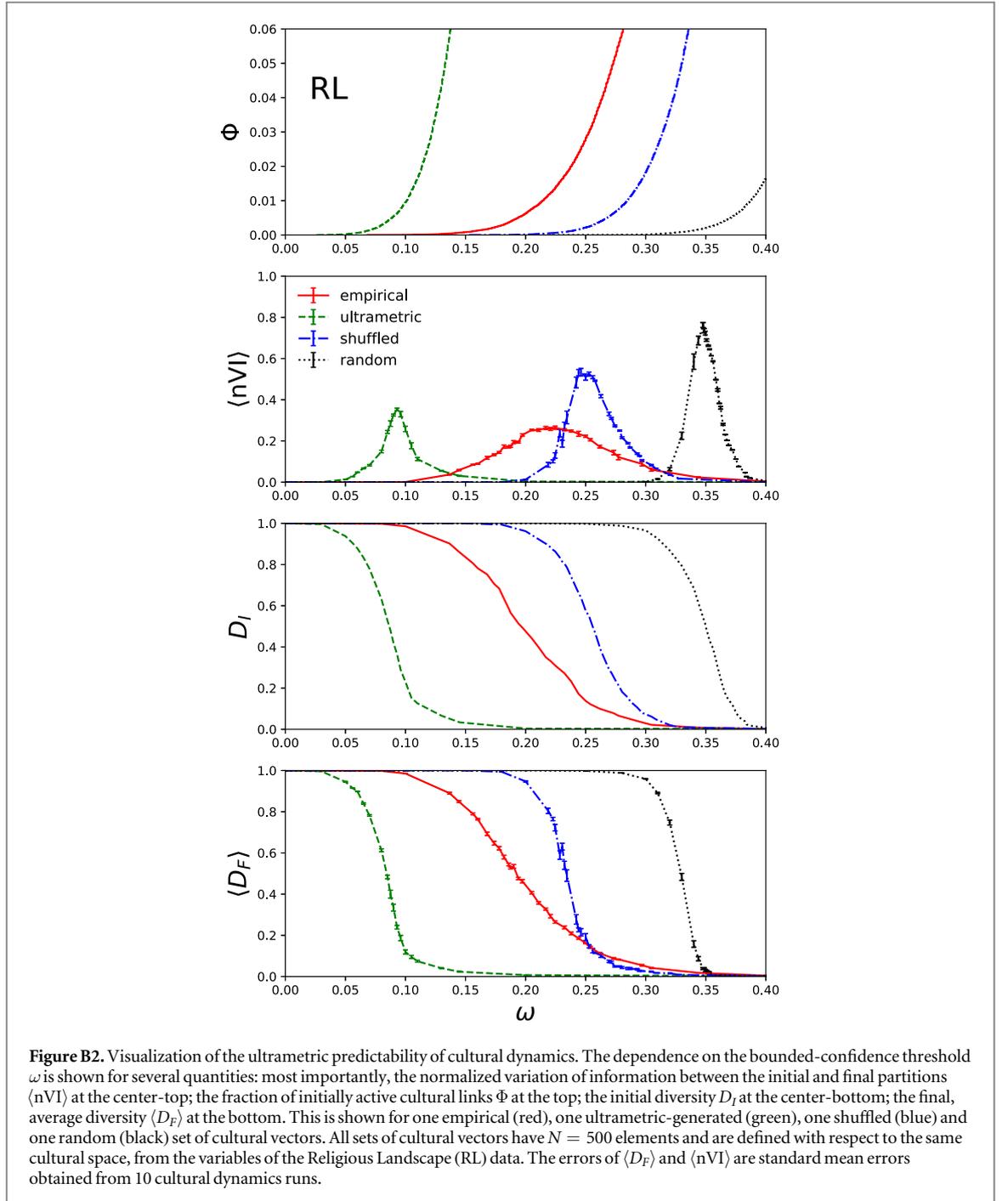
Appendix C. Dendrogram geometry

This section gives some analytical insight on how the dendrogram geometry is related to the behavior of the two measures of initial diversity D_I and initial coordination C_I. As functions of ω, the two measures only change (in steps) when ω crosses the distance value associated to any of the branchings of the dendrogram. Thus, one can replace the dependence of D_I and C_I on ω with a dependence on k, which counts the number of dendrogram branchings above a given ω, in terms of their associated distance values—k increases from 0 to $N - 1$ as ω decreases from 1.0 to 0.0. Based on equation (1), one can thus write:

$$D_I(k) = \frac{N_C^I(k)}{N}, \quad C_I(k) = \sqrt{\sum_A \left(\frac{S_A^I}{N} \right)_k^2}. \tag{C1}$$

There are two extreme types of dendrogram geometries that are worth considering, the 'perfectly-unbalanced geometry' and the 'perfectly-balanced geometry'. These are illustrated in figure C1.

Figure B2. Visualization of the ultrametric predictability of cultural dynamics. The dependence on the bounded-confidence threshold ω is shown for several quantities: most importantly, the normalized variation of information between the initial and final partitions $\langle \mathrm{nVI} \rangle$ at the center-top; the fraction of initially active cultural links Φ at the top; the initial diversity D_I at the center-bottom; the final, average diversity $\langle D_F \rangle$ at the bottom. This is shown for one empirical (red), one ultrametric-generated (green), one shuffled (blue) and one random (black) set of cultural vectors. All sets of cultural vectors have $N = 500$ elements and are defined with respect to the same cultural space, from the variables of the Religious Landscape (RL) data. The errors of $\langle D_F \rangle$ and $\langle \mathrm{nVI} \rangle$ are standard mean errors obtained from 10 cultural dynamics runs.

For the perfectly-unbalanced geometry, shown on the left side of figure C1, the number of connected components is:
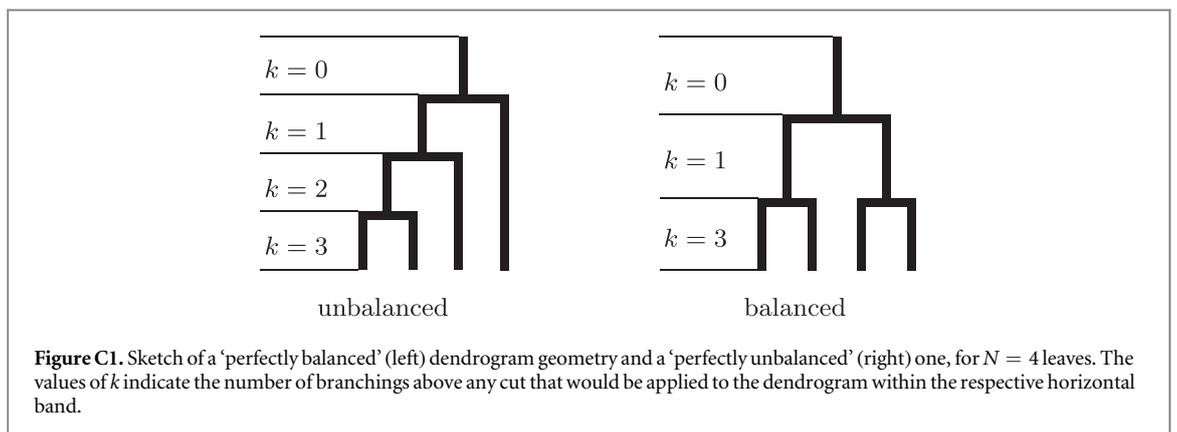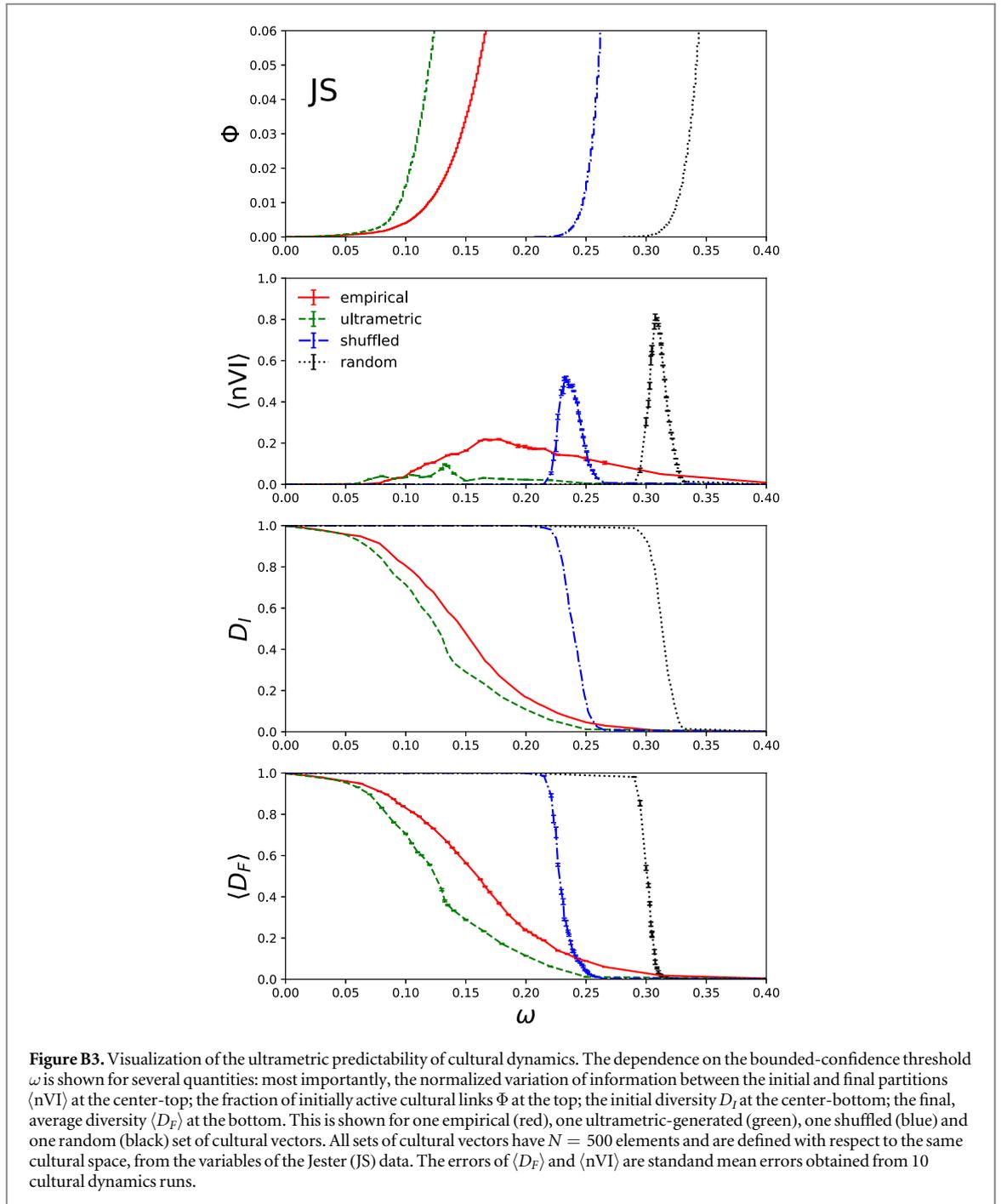
$$N_C^I(k) = k + 1, \tag{C2}$$

while the sizes of the connected component are:

$$S_A^I(k) = \begin{cases} N - k, & \text{if } A = 1 \\ 1, & \text{if } A \in \{2, 3, \ldots, k + 1\} \end{cases}. \tag{C3}$$

From equations (C1) and (C2), one obtains the behavior of the initial diversity measure:

$$D_I(k) = \frac{k + 1}{N}, \tag{C4}$$

Figure B3. Visualization of the ultrametric predictability of cultural dynamics. The dependence on the bounded-confidence threshold ω is shown for several quantities: most importantly, the normalized variation of information between the initial and final partitions $\langle \text{nVI} \rangle$ at the center-top; the fraction of initially active cultural links Φ at the top; the initial diversity D_I at the center-bottom; the final, average diversity $\langle D_F \rangle$ at the bottom. This is shown for one empirical (red), one ultrametric-generated (green), one shuffled (blue) and one random (black) set of cultural vectors. All sets of cultural vectors have $N = 500$ elements and are defined with respect to the same cultural space, from the variables of the Jester (JS) data. The errors of $\langle D_F \rangle$ and $\langle \text{nVI} \rangle$ are standand mean errors obtained from 10 cultural dynamics runs.

Figure C1. Sketch of a 'perfectly balanced' (left) dendrogram geometry and a 'perfectly unbalanced' (right) one, for $N = 4$ leaves. The values of k indicate the number of branchings above any cut that would be applied to the dendrogram within the respective horizontal band.

while from equations (C1) and (C3) one obtains the behavior of the initial coordination measure:

$$C_I(k) = \sqrt{\left(\frac{N-k}{N}\right)^2 + k\left(\frac{1}{N}\right)^2},$$ (C5)

from which it follows that:

$$C_I(k) = \sqrt{1 - 2\frac{k}{N} + \frac{k^2}{N^2} + \frac{k}{N^2}},$$ (C6)

where one can neglect the $\frac{k}{N^2}$ term in the limit of large N, thus obtaining:

$$C_I(k) \approx 1 - \frac{k}{N}.$$ (C7)

From equations (C4) and (C7) it follows that:

$$C_I(k) \approx 1 - D_I(k) - \frac{1}{N},$$ (C8)

which can be rephrased, after neglecting the $\frac{1}{N}$ term in the limit of large N, to:

$$D_I(k) \approx 1 - C_I(k),$$ (C9)

which describes the second-diagonal empirical behavior of figure 2(a), under the assumption that $D_F(k) = D_I(k)$, $\forall k$.

For a perfectly-balanced geometry, shown on the right side of figure C1, the only relevant values of k (those corresponding to non-vanishing ω-intervals) are $k = \sum_{i=0}^{l-1} 2^i$, with $l \in \{0, 1, 2, \ldots, \log_2 N\}$. For these values of k, the number of connected components, like in the unbalanced case, is described by equation (C2), while the sizes of the connected components are:

$$S_A^I(k) = N/(k+1), \forall A \in \{1, 2, \ldots, k+1\},$$ (C10)

from which it follows that the initial coordination measure is:

$$C_I(k) = \sqrt{(k+1)\left(\frac{1}{k+1}\right)^2} = \frac{1}{\sqrt{k+1}}.$$ (C11)

Since the k-dependence of the initial diversity measure D_I, like in the unbalanced case, is described by equation (C4), it follows that:

$$D_I(k) = \frac{1}{NC_I^2(k)},$$ (C12)

which, under the assumption that $D_F(k) = D_I(k)$, $\forall k$, entails a curve more similar to that of the shuffled or random curves of figure 2(a), than to that of the empirical curve. Moreover, this curve comes arbitrarily close to the lower-left corner as N increases.

To sum up, the above reasoning shows that, as long as $D_F(\omega) = D_I(\omega)$, $\forall \omega$, an unbalanced dendrogram geometry fits the empirical $D_F(C_I)$ behavior very well, while a balanced dendrogram geometry does not. Although the latter entails a $D_F \propto C_I^{-2}$ behavior quite similar to that observed for shuffled or random data, one cannot say that a balanced geometry is a good description for either of these two cases, since the assumption that $D_F = D_I$ is false for both these cases, for the interesting ω-intervals.

ORCID iDs

Alexandru-Ionuț Băbeanu ● https://orcid.org/0000-0001-6274-4871
Diego Garlaschelli ● https://orcid.org/0000-0001-6035-1783

References

[1] Buchanan M 2007 *The Social Atom* (New York: Bloomsbury)
[2] Turner J C 1991 *Social Influence* (Buckingham: Open University Press)
[3] Castellano C, Fortunato S and Loreto V 2009 Statistical physics of social dynamics *Rev. Mod. Phys.* **81** 591–646
[4] Axelrod R 1997 The dissemination of culture *J. Conflict Resolution* **41** 203–26
[5] Klemm K, Eguíluz V M, Toral R and Miguel M S 2003 Global culture: a noise-induced transition in finite systems *Phys. Rev.* E **67** 045101
[6] Klemm K, Eguíluz V M, Toral R and Miguel M S 2003 Nonequilibrium transitions in complex networks: a model of social interaction *Phys. Rev.* E **67** 026120
[7] Kuperman M N 2006 Cultural propagation on social networks *Phys. Rev.* E **73** 046139
[8] Flache A and Macy M W 2007 Local convergence and global diversity: the robustness of cultural homophily arXiv:physics/0701333

[9] González-Avella J C, Cosenza M G and Tucci K 2005 Nonequilibrium transition induced by mass media in a model for social influence *Phys. Rev.* E **72** 065102

[10] Centola D, González-Avella J C, Eguíluz V M and Miguel M S 2007 Homophily, cultural drift, and the co-evolution of cultural groups *J. Conflict Resolution* **51** 905–29

[11] Pfau J, Kirley M and Kashima Y 2013 The co-evolution of cultures, social network communities, and agent locations in an extension of Axelrod's model of cultural dissemination *Physica* A **392** 381–91

[12] Battiston F, Nicosia V, Latora V and Miguel M S 2017 Robust multiculturality emerges from layered social influence *Sci. Rep.* **7** 1809

[13] Stivala A, Kashima Y and Kirley M 2016 Culture and cooperation in a spatial public goods game *Phys. Rev.* E **94** 032303

[14] Sherif M and Hovland C I 1961 *Social Judgment: Assimilation and Contrast Effects in Communication and Attitude Change* (New Haven, CT: Yale University Press)

[15] Valori L, Picciolo F, Allansdottir A and Garlaschelli D 2012 Reconciling long-term cultural diversity and short-term collective social behavior *Proc. Natl Acad. Sci.* **109** 1068–73

[16] Stivala A, Robins G, Kashima Y and Kirley M 2014 Ultrametric distribution of culture vectors in an extended Axelrod model of cultural dissemination *Sci. Rep.* **4** 4870

[17] Băbeanu A-I, Talman L and Garlaschelli D 2017 Signs of universality in the structure of culture *Eur. Phys. J.* B **90** 237

[18] Băbeanu A-I and Garlaschelli D 2017 Evidence for mixed rationalities in preference formation *Complexity* **2018** 3615476

[19] Rammal R, Toulouse G and Virasoro M A 1986 Ultrametricity for physicists *Rev. Mod. Phys.* **58** 765–88

[20] Sibson R 1973 Slink: an optimally efficient algorithm for the single-link cluster method *Comput. J.* **16** 30

[21] Anderberg M R 1973 Hierarchical clustering methods *Cluster Analysis for Applications* (*Probability and Mathematical Statistics: A Series of Monographs and Textbooks*) ed M R Anderberg (New York: Academic) ch 6, pp 131–55

[22] Sokal R R and Michener C D 1958 A statistical method for evaluating systematic relationships *Univ. Kansas Sci. Bull.* **28** 1409–38

[23] Rammal R, Angles d'Auriac J C and Doucot B 1985 On the degree of ultrametricity *J. Phys. Lett.* **46** 945–52

[24] Tumminello M, Lillo F and Mantegna R N 2008 Generation of hierarchically correlated multivariate symbolic sequences *Eur. Phys. J.* B **65** 333–40

[25] Reif K and Melich A 1995 Euro-barometer 38.1:Consumer protection and perceptions of science and technology, November 1992 http://doi.org/10.3886/ICPSR06045.v2

[26] Commission of the European Communities 1992 Euro-barometer: public opinion in the European community, December 1992

[27] Meilă M 2007 Comparing clusterings—an information based distance *J. Multivariate Anal.* **98** 895

[28] Erdős P and Rényi A 1959 On random graphs: I *Publicationes Math.* **6** 290–7

[29] Smith T W, Marsden P and Hout M 1993 General social surveys 1972–2014

[30] Goldberg K, Roeder T, Gupta D and Perkins C 2001 Eigentaste: a constant time collaborative filtering algorithm *Inf. Retr.* **4** 133–51

[31] Lugo L *et al* 2008 US religious landscape survey. Religious beliefs and practices: diverse and politically relevant