

# Learning a Macroscopic Model of Cultural Dynamics

Aris Anagnostopoulos  
Sapienza University of Rome  
aris@dis.uniroma1.it

Mara Sorella  
Sapienza University of Rome  
sorella@dis.uniroma1.it

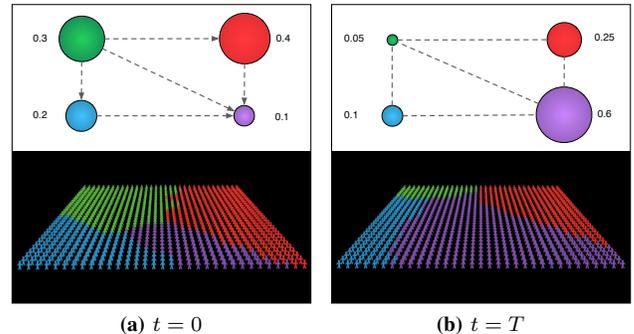
**Abstract**—A fundamental open question that has been studied by sociologists since the 70s and recently started being addressed by the computer-science community is the understanding of the role that influence and selection play in shaping the evolution of socio-cultural systems. Quantifying these forces in real settings is still a big challenge, especially in the large-scale case in which the entire social network between the users may not be known, and only longitudinal data in terms of masses of cultural groups (e.g., political affiliation, product adoption, market share, cultural tastes) may be available. We propose an influence and selection model encompassing an explicit characterization of the feature space for the different cultural groups in the form of a natural equation-based macroscopic model, following the approach of Kempe et al. [EC 2013]. Our main goal is to estimate edge influence strengths and selection parameters from an observed time series. To do an experimental evaluation on real data, we perform learning on real datasets from Last.FM and Wikipedia.

## I. INTRODUCTION

The evolution of human culture has quite naturally been the study of sociologists and anthropologists, since as early as the mid 20th century. Research postulates that cultural traits, such as language in use, religion, music preferences, and so on, evolve to some extent because of influence exerted by groups of the population to other groups [1]–[5] and in the last decade this topic has also entered the field of study of economists and physicists [6]–[8].

The idea behind the growing effort in developing mathematical models for explaining socio-cultural evolution is grounded by the fact that, in describing many contexts, qualitative and even some quantitative properties of large-scale phenomena depend to a large extent on high-level features, such as symmetries, dimensionality, or conservation laws [8].

In 2012, Kempe et al. [9] introduced the problem to the computer-science community, by presenting a family of models for cultural evolution based on the concepts of influence and selection, and studying graph-theoretic properties of the equilibria reached by these models. Following previous work [6], in these graph-theoretic models (*macroscopic models* [10]), nodes in the graph do not model the interactions between individual agents, as is typical in social-network models (*microscopic models*); instead, each node corresponds to a *cultural type* (or *group*) and it is associated with a mass, which captures the percentage of the population of this type. An edge between two nodes represents a possible flow



**Fig. 1:** Example of cultural dynamics unfolding in a setting with 4 distinct cultural groups and 2 subsequent observations in time. The macroscopic model (above) abstracts the microscopic interactions that occur at the microscopic level (below) among the individuals. Between the two time steps, there is transfer of mass from the green type to the other tree, and from the red and blue types to the purple one.

from one type to another<sup>1</sup>. For instance, a node’s mass may represent the english-speaking population of a given society, whereas another may represent a spanish-speaking one, and an edge between them corresponds to the potential of interaction between the two population groups. The models are dynamic, with continuous or discrete time steps, and the difference between two consecutive time steps captures the transfer of mass flow among the masses (see Figure 1).

Surprisingly, to our knowledge, there has not been any work to validate these macroscopic models for culture evolution: the literature on social dynamics exhibits a striking imbalance between theoretical modelization and empirical evidence, in favor of the former [8]. The work of Kempe et al. places its emphasis on the theoretical properties of the introduced models. Prior works, were more oriented towards modeling theories developed by sociologists at a theoretical level. The main goal of this work is to initiate a validation of such models for cultural dynamics and verify what information they can provide on the interdependence of different cultures. Inspired by the models of Kempe et al. we propose two models and fit them to large scale datasets that we have collected, of two different kinds: music (what genres people listen to), and wiki article editing (used as a proxy for users’ topical interests).

This research was partially supported by the Google Focused Research Award “Algorithms for Large-Scale Data Analysis”, and by the EU FET projects MULTIPLEX 317532 and SIMPOL 610704.

<sup>1</sup>Actually the model is more complicated with the presence of two graphs, but for this informal discussion this description suffices; we describe the complete model in Section III.

To summarize our contributions:

- We define new generic models for modeling cultural dynamics in the presence of both influence and selection, which can account for the vicinity of different culture types and allow for multiple features (Section III).
- We perform learning (Section IV) and validation of these models on real datasets (Sections V and VI), and we show that they can be useful for predicting future states.

## II. RELATED WORK

A number of models have been proposed in the literature, and may be broadly divided in two categories, namely *microscopic* and *macroscopic* models [10]. We start by mentioning some work performed by sociologists in the area of cultural dynamics, proceeding by describing mathematical models, which can be divided into *microscopic* and *macroscopic* models [10].

### A. Sociological Studies on Cultural Dynamics

Sociologists have studied the topic of cultural dynamics for several decades. Even the definition of what is meant by *culture* has been the topic of extensive discussion [11]. Kashima [10] treats the definition of culture, as “a set of non-genetic information that is available (i.e., information exists), accessible (i.e., information can be acquired), and applicable (i.e., information is usable) to a group of people”. His work surveys the diverse existing methodologies for research on cultural dynamics (longitudinal surveys, formal models and computer simulation) comparing their ability to explain short-, medium-, and long-term links between micro-level mechanisms and macro-level dynamics.

### B. Microscopic Models

Microscopic models constitute the vast majority of the proposed work in the field: they represent a social system as graph-type fixed structure where agents are initially associated with an opinion that iteratively changes in time as long as they interact with their neighborhood [2]–[4]. The concept of bounded confidence is introduced by Hegelsmann et al. [12], in this model an agent updates its opinion considering only the neighbors whose difference in opinion lies inside a confidence bound. Axelrod [1] proposed a cultural dissemination model where the probability of interaction between two agents is proportional to the number of dimensions in which they agree, and after an interaction the agent copies the value of a random differing dimension.

### C. Macroscopic Models

This class of models mainly derives from the field of statistical physics, whose aim is to study collective phenomena emerging from the interactions of individuals as elementary units in social structures. The models discussed in this work belong to this category. The Abrams-Strogatz model [6] presented a two-language competition model to explain historical data on the decline of endangered languages, proposing that the attractiveness of a language increases with the number of speakers and with its perceived status, foreseeing the extinction of languages with “lower social statuses”. In a generalization of this work, Patriarca et al. [7] introduce spatial dependence with a reaction–diffusion equation allowing survival of both

languages under the assumption that they are spoken in different zones and can interact only in a narrow transition region. More recently, Kempe et al. [9] concentrated on the interplay between selection and influence, envisioning a cultural system represented by a graph of cultural types which may interact and influence each other. The authors study the steady state behavior of the system characterizing the set of the stable equilibria.

## III. MODELING CULTURAL DYNAMICS

Kempe et al. [9] propose a model of cultural dynamics to characterize the evolution of networks under the combined effect of selection and influence. These two forces have distinguishably different causal mechanisms and determine opposing results: selection is the tendency of people to interact (and then form ties) with similar peers, thus leading to fragmentation; influence, spread through social ties, instead fosters homogeneity. More specifically, Kempe et al. consider the population as divided into a set of types  $V$  (where the types may represent opinions, cultural choices, language, etc.). The population is represented as a continuum such that the relative mass of type  $u \in V$  at time  $t$  is described by the real value  $x_u(t)$ . The authors introduce two different undirected graphs representing the relationships between types, namely the *interaction graph*  $\mathcal{S}$  (types that can interact with each other) and the *influence graph*  $\mathcal{T}$  (types that can influence one another), with  $\mathcal{T} \subseteq \mathcal{S}$ . In particular, there is an edge  $(u,v) \in \mathcal{T}$  if the two types are sufficiently *close* in a cultural sense. The process unfolds in discrete time steps, thus determining a discrete-time dynamical system where people can interact, provided that interaction is affected by *selection*: each person is more likely to choose an interaction partner of its type rather than another type, for a fixed parameter  $\alpha \geq 1$ . When a person of type  $u$  interacts with a person of type  $v$ , then if  $(u,v) \in \mathcal{T}$ , with fixed probability  $p \in (0,1]$  she may switch to type  $v$ . The quantity  $\mathcal{M}_u(t) = \alpha x_u(t) + \sum_{v \in V \setminus \{u\}} x_v(t)$  represents the *interaction mass* of type  $u$ , roughly speaking, the quantity of mass with which type  $u$  interacts; later we will generalize this expression for our models.

Given an influence graph  $\mathcal{T}$ , Kempe et al. study three different cases, progressively restricting the interaction graph  $\mathcal{S} \supseteq \mathcal{T}$ : (1) the *global* model where  $\mathcal{S}$  is the complete graph, thus all types are allowed to interact with each other; (2) the *general* model (where  $\mathcal{S}$  is an arbitrary superset of  $\mathcal{T}$ ); and (3) the *local* model (where  $\mathcal{S} = \mathcal{T}$ ). In the three settings, they aim at characterizing the (stable) equilibria: they prove that the global model converges to an equilibrium<sup>2</sup> for any initial node masses, and find that stable<sup>3</sup> equilibria are the ones in which the nodes with positive mass form an independent set. For the other two models they prove convergence results for particular instances.

However, as we will see, these models are not sufficient to capture the interactions that we are interested in. Thus, in the next section, building on it, we detail two derived models, which we will subsequently subject to the parameter-learning process.

<sup>2</sup>An equilibrium occurs when the masses of the types do not change as the system evolves.

<sup>3</sup>The notion of stability is the *Lyapunov stability*; see [9].

### A. Our Models

The goal of our work is to fit models of cultural-dynamics on real-life data as a means to study the dynamics and observe how different culture types interact with each other. Thus we need to characterize more precisely the types, how they relate to each other, and how the corresponding population interacts. Therefore we introduce some modifications to the models in [9]. First we define exactly what we mean by types, and then we present the first model (*generalized global model*), which is a generalization of the global model in [9]. Subsequently we modify it by introducing the *hypercube model*, which can capture scenarios not possible to be captured by the other one.

We start by characterizing explicitly the cultural types in terms of the different *cultural traits* (features) each type may have. Furthermore, whereas the interaction and influence graphs  $\mathcal{S}$  and  $\mathcal{T}$  are supposed to be known, this information is not easy to obtain apriori. We will therefore need to contemporaneously identify the network structure and the influence and selection parameters: to do so, we allow interactions between any two types (as with the global model), while allowing the probability of influence between any two types to be 0.

We consider a set  $V$  of  $n$  cultural types or groups). Furthermore, each type  $u$  in  $V$  is characterized by an  $m$ -dimensional *cultural feature vector*  $F_u = (F_u^1, F_u^2, \dots, F_u^m)$ , where each feature  $F_u^i$  can take binary values. Feature vectors represent configurations of cultural traits that span a domain (e.g., language, music tastes, political affiliation). For instance, point  $(1, 1, 0, 1, \dots)$  may represent the type of population listening to music of type “blues,” “jazz,” “not pop,” “rock,” and so on; however, the feature vectors can be generalized to span multiple domains [1], [9].

To model the evolution of the system, we assume that each person in the population belongs at each time step to one of the types. At each time  $t$ , for each type  $u$ , we have a mass  $x_u(t)$ , that is, we model the system as a stochastic process  $\{\mathbf{x}(t)\}_{t=0,1,\dots}$ , where  $\mathbf{x}(t) = (x_1(t), x_2(t), \dots, x_n(t))$ . We assume that we have conservation of mass, so for each  $t$  we have  $\sum_u x_u(t) = B$ , for a constant  $B$  that represents the total population mass.

For two types  $u$  and  $v$  we can define their *cultural distance*  $D(u,v)$  as the number of features in which they differ:  $D(u,v) = |\{i : F_u^i \neq F_v^i\}|$ . In other words,  $D(u,v)$  corresponds to the *Hamming* distance between  $F_u$  and  $F_v$ .

Given that we don’t have any information about which types may interact with each other, the types are assumed to be connected forming an undirected clique over  $V$  (i.e.,  $\mathcal{S} = K_n$ ).

The influence graph  $\mathcal{T} = (V, E_{\mathcal{T}}, p(\cdot, \cdot))$ , differently from [9], is a directed edge-weighted graph representing the influence relationships that exist over the set of cultural types. Given two types  $u$  and  $v$ , the existence of an edge  $e = (u,v) \in \mathcal{T}$  implies that type  $u$  can be influenced by type  $v$ , and the influence probability is given by the weight  $p(u,v) = p_{uv}$ . As with the interaction graph, this graph is a full clique, although the influence between two groups may be zero. This model generalizes the global model of [9] introducing different influence strengths between each (directed) pair of types.

Likewise, the expression of the *interaction mass* is similar as in [9], with the generalization of having a different selection parameter  $\alpha_u$  for each node (representing the unique extent to

which each cultural combination may give rise to selection):

$$\mathcal{M}_u(t) = \alpha_u x_u(t) + \sum_{v \in V \setminus \{u\}} x_v(t) = (\alpha_u - 1)x_u(t) + B. \quad (1)$$

The probability for a person of type  $u$  of interacting with another person in  $u$  at time  $t$  is therefore  $\frac{\alpha_u x_u(t)}{\mathcal{M}_u(t)}$  and the probability of interacting with another type  $v$  is  $\frac{x_v(t)}{\mathcal{M}_u(t)}$ .

Next we describe in detail our models, which differ on how the masses are being updated.

1) *Generalized Global Model*: Before stating the update rule for the *generalized global model* it is useful to introduce the notion of flow, which is instrumental to identify the contribution of each type–type relationship to the overall dynamics of the system. Given two types  $u$  and  $v$  with masses at time  $t$   $x_u(t)$  and  $x_v(t)$ , respectively, their interaction is affected by selection and by influence (captured by parameter  $p_{uv}$ ) leading to a (directed) flow of mass  $f_{uv}(t)$  from type  $u$  to type  $v$ :

$$f_{uv}(t) = x_u(t) \frac{x_v(t)}{\mathcal{M}_u(t)} p_{uv}. \quad (2)$$

Intuitively, the flow  $f_{uv}(t)$  represents the *expected* transfer of mass from type  $u$  to type  $v$ . It is proportional to the masses and to the influence probability  $p_{uv}$ . Then, for each  $v \in V$  we have the update rule:

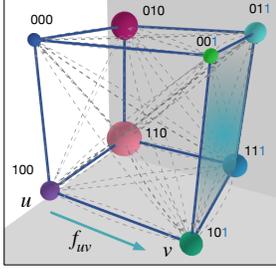
$$x_u(t+1) = x_u(t) + \sum_{v \in V \setminus \{u\}} f_{vu}(t) - f_{uv}(t). \quad (3)$$

2) *Cultural Hypercube Model*: As we stated previously, the global model is a direct generalization of the model in [9], with multiple parameters for influence and selection. However, in these models, when a person of cultural type  $u$  interacts and is influenced by a person of type  $v$ , at the next time step she will be part of group  $v$ : this means that she must change *all* her features to turn to the ones of  $v$ . Even though such a modeling choice may be generally appropriate for issues such as religion, in several real-life setting this is not realistic [1]. Consider the music scenario where the feature of each type corresponds to a music type. A person who enjoys rock music (type “rock”), may be influenced towards jazz by a person of type “jazz and country,” and as a result start listening to jazz as well, becoming of type “rock and jazz”; this cannot be captured by the existing models.

For this reason, along the lines of the model of Axelrod [1], we introduce the *cultural hypercube model*: we still allow all the types to interact and potentially influence each other; yet, for each timestamp there will be an actual flow of mass only between types that are neighbors in the *culture transition graph*  $\mathcal{C}$ , an undirected graph where there is an edge only between types at cultural distance  $D(u,v) = 1$  (see Figure 2). (It can be generalized to be connected to nodes of larger distance, but  $D(u,v) = 1$  is the simplest and most natural model.)

The interaction and influence graph  $\mathcal{S}$  and  $\mathcal{T}$  are unchanged, whereas the culture transition graph  $\mathcal{C}$  has a hypercube structure whose dimension is equal to the number of features that we consider.  $\mathcal{C}$  represents the possible flow exchanges: given two types  $u$  and  $v$  there is a possibility of having flow from node  $u$  to node  $v$  at a given time  $t$  if and only if  $D(u,v) = 1$ .

Let us now see how this flow between culturally adjacent types is created. Let  $i$  be the feature in which  $u$  and  $v$  differ.



**Fig. 2:** Example of cultural flow from node  $u$  to node  $v$  ( $m = 3$ ,  $n = 8$ ). The blue links forming the edges of the 3-dimensional hypercube are the edges of the cultural transition graph  $\mathcal{C}$ . Notice that  $u$ 's feature vector (100) and  $v$ 's (101) differ on the third feature. The flow from  $u$  to  $v$  originates from the interaction of  $u$  with all nodes on the highlighted 2-dimensional hypercube (i.e., nodes whose third feature is 1). The interaction of  $u$  with  $v$  leads to direct flow, and the interaction of  $u$  with the others, to indirect one.

The flow can be decomposed into two components: first, a *direct flow* of mass from node  $u$  to  $v$ , is the result of interaction between types  $u$  and  $v$ , as in the previous models. In addition, we also have an *indirect flow* from  $u$  to  $v$ , which originates from the interaction of the mass of type  $u$  with the mass of each type  $z \notin \mathcal{N}_C(u)$ <sup>4</sup> that belongs to the  $m - 1$  dimensional hypercube such that  $F_v^i = F_z^i$ . Informally, we have indirect flow from  $u$  to  $v$  by the interaction of  $u$  with the nodes  $z$  that share with  $v$  the feature  $i$  where  $u$  and  $v$  differ (see Figure 2).

Let us now formalize the previous discussion. The flow between nodes  $u$  and  $v$ , which differ in feature  $i$ , is the sum of the direct and indirect flows:

$$f_{uv}(t) = x_u(t) \frac{x_v(t)}{\mathcal{M}_u(t)} \frac{p_{uv}}{m} + \sum_{z \in C_v^u \setminus v} x_u(t) \frac{x_z(t)}{\mathcal{M}_u(t)} \frac{p_{uz}}{m} \quad (4)$$

where  $C_v^u = \{z \in V : (u, z) \in E_{\mathcal{T}} \wedge F_v^i = F_z^i\}$ . Thus:

$$f_{uv}(t) = x_u(t) \sum_{z \in C_v^u} \frac{x_z(t)}{\mathcal{M}_u(t)} \frac{p_{uz}}{m}. \quad (5)$$

The intuition of this expression is (as before) that the probability of interaction of  $u$  and  $v$  is proportional to their masses. Furthermore, when  $u$  and  $v$  interact, given that we have  $m$  features, we assume that the probability that they interact on feature  $i$  is  $1/m$ . Note that this equation holds both for attaining a feature as well as losing it; even though this assumption may sound unnatural, we decided to make it, in line with the symmetric treatment in previous work [1].

The update rule for type  $u$  can thus be written in the form of a difference equation, where the mass for type  $u$  at time  $t+1$  is given by its mass at time  $t$  plus the incoming cultural flows, from neighboring cultural types, minus the outgoing flow:

$$x_u(t+1) = x_u(t) + \sum_{v: u \in \mathcal{N}_C(v)} f_{vu} - \sum_{v \in \mathcal{N}_C(u)} f_{uv}. \quad (6)$$

The global mass update rule is a vector valued function where each dimension ( $x_i(t)$ ) is determined by the update rule (6) of each type.

<sup>4</sup>We use the standard notation  $\mathcal{N}_X(u)$  to indicate a neighboring node of  $u$  in graph  $X$ .

## IV. LEARNING

Now we explain how we can learn the parameters  $\alpha_u$  and  $p_{uv}$  of the models described in Section III-A. Our starting point is a time series of observed group masses (in terms of number of people belonging to each group).

### A. Large-Scale Nonlinear System Identification

The update rules of the two models described in Section III-A have a closed form expressed by Equations 3 and 6: these equations relate the mass-vector  $\mathbf{x}(t)$  of the different cultural masses to the previous  $\mathbf{x}(t-1)$  through the (unknown) parameters of selection and influence. Thus, they both identify a discrete-time dynamical system in the form of a multivariate nonlinear autoregressive process or order 1 (Markovian). Note that these systems are nonlinear both in the state  $\mathbf{x}(t)$  and in the parameters  $\theta = \begin{pmatrix} \theta_\alpha \\ \theta_p \end{pmatrix}$ , where  $\theta$  is a static (time-independent) vector of parameters:  $\theta_\alpha = (\alpha_1, \dots, \alpha_n) \in \mathbb{R}^n$  with  $\alpha_i \geq 1$ , is a vector encoding the values  $\alpha_u$  for every type  $u \in V$  and  $\theta_p$  contains all the influence probabilities  $p_{ij} \in [0, 1]$  over the edges of the influence graph  $\mathcal{T}$ . Recalling that the number of types is  $n$ , let  $\mathcal{O}$  be an  $n \times (T+1)$  matrix where each row  $\mathcal{O}_t = [x_1^{obs,t}, x_2^{obs,t}, \dots, x_n^{obs,t}]$  contains the  $n$ -dimensional observations of the cultural masses for every time step  $t \in \{0, 1, \dots, T\}$ .

For each model  $\mathcal{M}$  we want to find the parameters  $\theta$  that best fit the observed data under  $\mathcal{M}$ .

We choose an estimator that takes into consideration the cumulative error between the real observed trajectory  $\mathcal{O}$  and a simulation of the model from the initial time instant *over the whole timespan*. As a cost function it is standard to choose a *least-squares* formulation; to avoid overfitting, we perform regularization using two different regularization coefficients,  $\lambda_\alpha$  for the parameters  $\lambda_p$  for the parameters  $\alpha_u$  and  $p_{uv}$  respectively, in an  $\ell_1$  norm penalization scheme. We use two regularizers because the parameter space for the  $\alpha_u$  and  $p_{uv}$  are very different. We use the  $\ell_1$  norm to enforce sparsity of the learnt parameter vector: this choice is motivated by a high number of parameters that are likely to be zero; for example, pairs of types for which there might be observations of no influence among them. Thus, let  $\tilde{x}_u^{M\theta} = \text{update\_rule}(\mathcal{O}_0, \theta, T) = (x_u^{obs,1}, x_u^{obs,2}, \dots, x_u^{obs,T})$  be a vector representing the ordered outcomes of  $T$  recursive applications of the update rule (Equation (3) or (6) depending on the model) of node  $u$  for model  $\mathcal{M}$ .

The regularized least squares estimator is:

$$\hat{\theta}_{\mathcal{M}} = \underset{\theta}{\operatorname{argmin}} \sum_{t=1}^T \sum_{u=1}^n (x_u^{obs,t} - \tilde{x}_u^{M\theta,t})^2 + \lambda_\alpha \|\theta_\alpha\|_1 + \lambda_p \|\theta_p\|_1. \quad (7)$$

We tested several learning approaches (line-search, simulated annealing, genetic algorithms) and we ended up selecting a specific kind of population-based algorithm which is *particle swarm* [13]. We exploited an efficient implementation in MATLAB<sup>5</sup>, which is able to parallelize the evaluation of the objective function of the particles in a convenient way (given the very high number of parameters), achieving good performances in terms of execution time.

<sup>5</sup><http://www.mathworks.com/help/gads/particle-swarm.html>

Dataset	Period	Duration	# Users	# actions
Wikipedia	2004/10-2007/10	3 years	2573	2876974
Last.FM	2009/05-2014/05	5 years	38213	627138188

**TABLE I:** Summary statistics on the datasets used for experimental validation.

## V. DATASETS

To study the performance of our models on real data, we take two different case studies: the first is the famous music social network Last.FM (we observed users’ listening history) and the second comes from Wikipedia (we observed users’ editing history). As we have discussed, our models deal with groups of users having given features. For both cases, we use datasets<sup>6</sup> that contain complete information about songs (articles, for Wikipedia) listened (modified) by a set of users in a period of time. Furthermore, the processing of the history for a user accounts for a behavior model, in a way to derive, for each time instant considered a feature vector encoding that user’s activities (the cultural group he belongs to). We then consider the time series of the resulting cultural group populations to fit our models.

In Sections V-A and V-B we describe in detail the collection and preprocessing phases performed on the datasets, and in Section V-C we detail the user model.

### A. Last.FM

Motivated by Lewis et al. [5], our first dataset captures the music preferences of users over time. To obtain our user set, we used the Two Million Last.FM User Profiles dataset [14], containing the profile information of 1,840,647 users. We then extracted a total of 44,154 users with more than 10k and fewer than 40k listens. For these users we downloaded the full listening history for the 5-years period from May 2009 to May 2014, obtaining an amount of 721 million listenings and around 4.6 millions unique tracks (see Table I). We then assigned each song to a set of predefined music genres, which form the features of our types. We omit the details of this step for paucity of space. To limit the number of parameters to learn, we grouped the genres into 5 macro genres: “Pop and Rock,” “Hip-Hop and Electronic,” “Jazz,” “World Music,” and “Classical”. Given that “Classical” was highly underrepresented in our dataset ( $< 0.01\%$ ) we removed it from the classification and divided “Hip-Hop/Electronic” into “Hip-Hop/RnB” and “Electronic.” We then assigned each song to a set of predefined music genres, which form the features of our types. We omit the details of this step for paucity of space. To limit the number of parameters to learn, we grouped the genres into 5 macro genres: “Pop and Rock”, “Hip-Hop and Electronic”, “Jazz”, “World Music”, and “Classical”. Given that “Classical” was highly underrepresented in our dataset ( $< 0.01\%$ ) we removed it from the classification and divided “Hip-Hop/Electronic” into “Hip-Hop/RnB” and “Electronic.” Note that even 5 features result in  $n = 32$  cultural types and a total of 1024 parameters ( $|\theta_\alpha| = 32$ ,  $|\theta_p| = 992$ ) to be learnt.

### B. Wikipedia

For Wikipedia, we take edits to articles as expressions of interests on specific topics. We started from a dataset containing the full English Wikipedia edit history (up to 2008) [15]:

timestamped metadata about all article edits (modification of pages) made by users. We limited our study to *Active Wikipedians*<sup>7</sup>, i.e., registered users with more than 5 edits per month on average. After this filtering phase, we selected the three years period from 24 October 2004 to 24 October 2007. This led to a set of 2576 users (see Table I). Using the Wikipedia taxonomy<sup>8</sup> we categorized each page into one of five categories, which define the features for our cultural types: “Science and Technology”, “Politics, Society, Religion and Philosophy”, “History and Events”, “Arts, Culture, Literature and Music”, and “Geography and Environment”. The details of how exactly we performed the categorization will appear in an extended version of this paper.

### C. User Model

We now need a way to identify the group (a collection of feature values) to which each user belongs, at each examined point in time. We want to use the user activity information to determine a (possibly changing over time) profile that ideally is able to capture the moment in which a user’s activity features (her currently preferred genres/interest topics) change in time, while accounting for more long-term interests in the domain, in a way to mimic the natural “forgetting” behavior of the user. To this end, we adopt an exponential decay function to weigh each sequence of actions according to its position in the user timeline. More formally, we segment the activity history of each user  $u$  into time frames, and for each time frame  $t$  we keep an interest score vector  $S_t^u$ , which stores the normalized score of actions performed by the user relative to genres listened or topics of articles edited in that time frame. We then keep a user profile using a time-sensitive exponential decay weighting scheme; that is,  $u$ ’s profile at time  $\tau$  is:  $\mathcal{P}_\tau^u = \sum_{t=0}^{\tau} S_t^u \cdot e^{-\lambda_f(\tau-t)}$ . The parameter  $\lambda_f$  represents the forgetting decay parameter. The resulting vector is then normalized to have all components sum to 1 and with a thresholding mechanism the user is assigned to the corresponding cultural group for each timestamp.

## VI. EXPERIMENTAL RESULTS

In this section, we study, compare and validate the models proposed in section III on our datasets.

### A. Experimental Setup

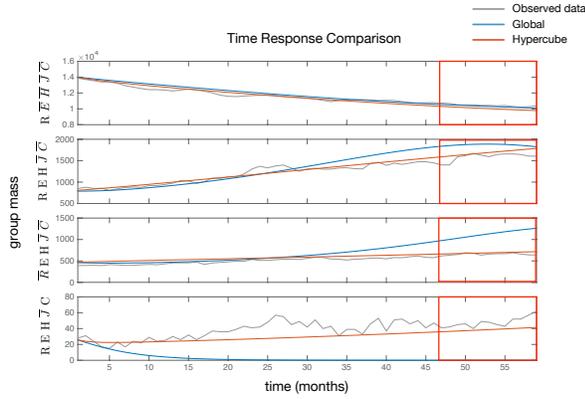
We use the Last.FM and Wikipedia datasets described in Section V where we have access to observed masses for the different cultural groups for, respectively, periods of 5 and 3 years.

For both datasets, we have  $m = 5$  features,  $n = 2^m = 32$  cultural groups. The sampling period is 1 month and we set the forgetting threshold  $\lambda_f$  of the user model (Section V-C) to the value 0.5, which after the exponential decay gives a very small weight to scores after about 6 months. We divide all the observed periods in training and test portions in the following way: 80%-20% (Last.FM) and 83%-17% (Wikipedia) and evaluate on both portions the performance of the generalized global model and the cultural hypercube model (subsequently, respectively, *global* and *hypercube*). Let  $T$  be the length of the observed group masses time series (for either the training or test portion) and  $\epsilon_t^2 = \sum_{u=1}^n (x_u^{\text{obs},t} - \hat{x}_u^{M\theta,t})^2$  be

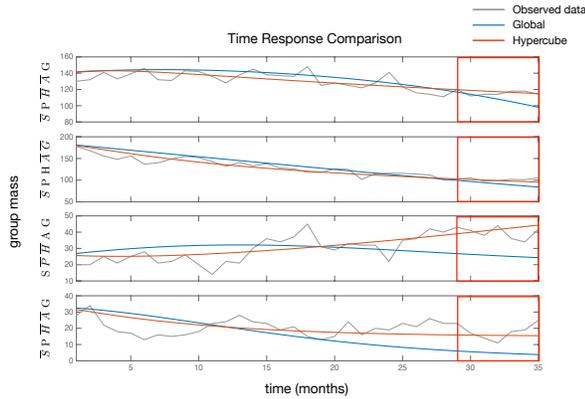
<sup>6</sup>All the datasets created by us (or resulting from a processing of existing ones) are available at <http://wadam-data.dis.uniroma1.it/wadam-datasets/cultural-dynamics/index.html>.

<sup>7</sup><http://stats.wikimedia.org/EN/Tables/WikipediansEditsGt5.htm>

<sup>8</sup><http://wikimedia.wansec.com/archive/enwiki/20080103/>



(a) Last.FM



(b) Wikipedia

**Fig. 3:** Comparison between observed data and a simulation of the global and hypercube models using the learnt parameters  $\theta_{\text{opt}}$  for some cultural groups. The red block highlights the test portion.

the absolute squared error of the predicted mass values with respect to the observed data for timestamp  $t$ . A measure of performance of the fitted models is the RMSE (Root Mean Squared Error):  $\text{RMSE} = \sqrt{\frac{1}{T} \sum_{t=1}^T \epsilon_t^2}$ . however, this error metric is sensitive to the scale of the input data. Hence we use the normalized version:  $\text{NRMSE} = \text{RMSE} / |x_{\text{max}}^{\text{obs}} - x_{\text{min}}^{\text{obs}}|$ , where  $x_{\text{max}}^{\text{obs}} = \max_{u,t}(x_u^{\text{obs}}, t)$  and  $x_{\text{min}}^{\text{obs}} = \min_{u,t}(x_u^{\text{obs}}, t)$  are the maximum and minimum values of all observed group masses.

Figures 3a and 3b show a comparison between observed data (test and training) and predictions made by the models simulated using the fitted  $\theta_{\text{opt}}$  parameters: due to space constraints we selected 4 out of the total 32 cultural groups for both datasets. These 4 groups were chosen on the basis of the average mass they exhibit during the examined periods, listed in decreasing order (first: high average mass, last: low average mass). Quite surprisingly, we observe that we have a good fit even in the test data and predict future (short-term) trends. The performance of the fitted models on Last.FM are consistently better than the ones on Wikipedia, mainly because of the higher amount of training data and the bigger population sample. For both datasets, the hypercube model is found to have consistently higher performance both over training and over prediction data.

## VII. CONCLUSION

To the best of our knowledge we are the first to do an experimental validation of a nontrivial macroscopic model for cultural dynamics. ([6] performed some experiments on a toy example consisting of two nodes). To do this we introduced new models, modifying the model of [9]. Fitting the models requires learning a very large number of parameters. We discovered that we obtain a good fit on the training data, and rather surprisingly (given that we are dealing with systems that in real life are not closed, as the models here assume) we discovered that our models are able to follow closely actual real data. These findings indicate that our model is able to characterize to some extent the evolution of some cultural traits, complementing the long line of work of sociologists, anthropologists, psychologists, and computer scientists who have proposed such models for cultural dynamics.

## VIII. ACKNOWLEDGEMENTS

We would like to thank Emanuele Petagna for his help in preprocessing the datasets, Laura Palagi for providing us with recommendation about the optimization, and Giorgio Grisetti and Luca Iocchi for useful discussions. We also thank the anonymous reviewers for their comments. Finally, we thank Google for awarding us with Google Cloud Credits, providing us the necessary infrastructure for running our resource-demanding algorithms.

## REFERENCES

- [1] R. Axelrod, "The dissemination of culture a model with local convergence and global polarization," *Journal of conflict resolution*, vol. 41, no. 2, 1997.
- [2] P. Clifford and A. Sudbury, "A model for spatial conflict," *Biometrika*, vol. 60, no. 3, 1973.
- [3] M. H. DeGroot, "Reaching a consensus," *Journal of the American Statistical Association*, vol. 69, no. 345, 1974.
- [4] N. E. Friedkin and E. C. Johnsen, "Social influence and opinions," *Journal of Mathematical Sociology*, vol. 15, no. 3-4, 1990.
- [5] K. Lewis, M. Gonzalez, and J. Kaufman, "Social selection and peer influence in an online social network," *Proceedings of the National Academy of Sciences*, vol. 109, no. 1, 2012.
- [6] D. M. Abrams and S. H. Strogatz, "Linguistics: Modelling the dynamics of language death," *Nature*, vol. 424, no. 6951, 2003.
- [7] M. Patriarca and T. Leppänen, "Modeling language competition," *Physica A: Statistical Mechanics and its Applications*, vol. 338, no. 1, 2004.
- [8] C. Castellano, S. Fortunato, and V. Loreto, "Statistical physics of social dynamics," *Reviews of modern physics*, vol. 81, no. 2, 2009.
- [9] D. Kempe, J. Kleinberg, S. Oren, and A. Slivkins, "Selection and influence in cultural dynamics," in *Proceedings of the Fourteenth ACM Conference on Electronic Commerce*, 2013.
- [10] Y. Kashima, "How can you capture cultural dynamics?" *Frontiers in psychology*, vol. 5, 2014.
- [11] J. R. Baldwin, S. L. Faulkner, M. L. Hecht, and S. L. Lindsley, *Redefining culture: Perspectives across the disciplines*, 2006.
- [12] R. Hegselmann and U. Krause, "Opinion dynamics and bounded confidence models, analysis, and simulation," *Journal of Artificial Societies and Social Simulation*, vol. 5, no. 3, 2002.
- [13] J. Kennedy, "Particle swarm optimization," in *Encyclopedia of Machine Learning*, 2010.
- [14] Socrata. (2012) The two million lastfm users dataset. [Online]. Available: <https://opendata.socrata.com/Business/Two-Million-LastFM-User-Profiles/5vvd-truf>
- [15] G. Kossinets. (2012) Processed wikipedia edit history. stanford large network dataset collection (SNAP). [Online]. Available: <http://snap.stanford.edu/data/bigdata/wikipedia08/enwiki-20080103.main.bz2>