



ARTICLE

<https://doi.org/10.1057/s41599-019-0295-9>

OPEN

Modelling rapid online cultural transmission: evaluating neutral models on Twitter data with approximate Bayesian computation

Simon Carrignon ¹, R. Alexander Bentley² & Damian Ruck²

ABSTRACT As social media technologies alter the variation, transmission and sorting of online information, short-term cultural evolution is transformed. In these media contexts, cultural evolution is an intra-generational process with much ‘horizontal’ transmission. As a pertinent case study, here we test variations of culture-evolutionary neutral models on recently-available Twitter data documenting the spread of true and false information. Using Approximate Bayesian Computation to resolve the full joint probability distribution of models with different social learning biases, emphasizing context versus content, we explore the dynamics of online information cascades: Are they driven by the intrinsic content of the message, or the extrinsic value (e.g., as a social badge) whose intrinsic value is arbitrary? Despite the obvious relevance of specific learning biases at the individual level, our tests at the online population scale indicate that unbiased learning model performs better at modelling information cascades whether true or false.

¹ Universitat Pompeu Fabra (UPF), Barcelona, Spain. ² University of Tennessee, Knoxville, Tennessee, USA. Correspondence and requests for materials should be addressed to S.C. (email: simonpaul.carrignon01@estudiant.upf.edu) or (email: simon.carrignon@gmail.com)

Introduction

Cultural evolution is undoubtedly altered by social media technologies, which impose new, often algorithmic, biases on social learning at an accelerated tempo on a vast virtual landscape of interaction. Unlike traditional societies that share in person (Danvers et al., 2019; Smith et al., 2019), sharing on social media is often not primarily kin or need-based. Important evolved psychologies, for example, such as shame and social exclusion (Robertson et al., 2018), or the visibility of social interactions involving others (Barakzai and Shaw, 2018), can be greatly altered in online social networks. One way they affect social learning is by the prominent display of social metrics (likes, shares, followers, etc) that feed biases toward popularity and often novelty; digital social data far exceed what a human Social Brain can track without technological assistance (Crone and Konijn, 2018; Dunbar and Shultz, 2007; Falk and Bassett, 2017; Hidalgo, 2015). Additionally, homophily is facilitated as people interact within their “social media feeds, surrounded by people who look like [them] and share the same political outlook,” as Barack Obama (2017) put it.

Reshaping social learning in diverse ways, social media platforms provide an opportunity to observe how their algorithmic constraints and biases on social learning affect how information spreads and evolves. Indeed, as the popularity among those platforms changes substantially from year to year (Lenhart, 2015), the cultural evolutionary process itself may be changing on an intra-generational time scale.

Since social learning strategies are culturally variable (Molleman, Gächter, 2018), they should also vary by social media platform. While Facebook, for example, may bias social learning towards recent and popular material among clustered groups of friends, Twitter often involves larger groups of followers who may largely be strangers to each other and to the people (or bots) they are following (Ruck et al., 2019).

Two major categories of social learning strategies are content-biased learning and context-biased learning (Kendal et al., 2018); either bias can be manipulated by a social media platform. Content-bias can be imposed, for example, by algorithmic filters that customize social media feeds to individual users. Context biases are also routinely imposed by social media platforms, which often prioritize popularity and recentness. Since strategies such as “copy recent success” are most competitive in fast-changing social landscapes (Rendell et al., 2010; Mesoudi et al., 2015), we might expect these context biases to flourish among social media (Bentley and O’Brien, 2017; Acerbi and Mesoudi, 2015; Kendal et al., 2018). Until recently, context-bias in online media was often underestimated. The hosts of Google Flu, for example, over-predicted influenza rates for 2013 (Lazer et al., 2014) by not accounting for context-biased learning about flu from other Internet users rather than individuals’ own symptoms (Bentley and Ormerod, 2010; Ormerod et al., 2014).

With over 300 million users worldwide, Twitter makes many social learning parameters explicit, including the numbers of followers, re-tweets and likes of users and their messages. Aggregated Twitter content has previously been used for counting the frequencies of specific words across online populations, which can reveal mundane cycles of daily life (Golder and Macy, 2011), the risk of heart disease (Eichstaedt et al., 2015) and numerous other phenomena.

Subsequently, more work has been done on the dynamics of information flow online. Vosoughi et al. (2018) documented how false news travel “farther and faster” than true information among Twitter users. Here we begin to explore these dynamics by focusing on the sizes of Twitter “cascades” measured by Vosoughi et al. (2018), in terms of total number of re-tweets of messages that had been independently classified as true or false. Here we

refer to “true” versus “false” rumors in their data, which correspond to confirmed fact-checked rumors versus rumors that were debunked, respectively—see Vosoughi et al. (2018).

Established in cultural evolution research (Acerbi and Bentley, 2014; Bentley et al., 2011; Neiman, 1995; Premo, 2014; Reali and Griffiths, 2010), our null model is that Twitter messages are “neutral”, i.e., copied randomly with negligible bias, until proven otherwise. Models of unbiased (neutral) copying have also been applied to social media (Gleeson et al., 2014). In this approach, we assume that the probability of a message being re-tweeted depends only on the current frequency of the message and not on its content. Content bias would be identified when the model is falsified (Acerbi et al., 2009; Acerbi and Bentley, 2014; Bentley et al., 2004; Bentley et al., 2007; Bentley et al., 2014; Kandler and Shennan, 2013; Mesoudi and Lycett, 2009; Neiman, 1995).

Unbiased copying models have been calibrated empirically against real data sets that represent easily copied variants, such as ancient pottery designs (Bentley and Shennan, 2003; Crema et al., 2016; Eerkens and Lipo, 2007; Neiman, 1995; Premo and Scholnick, 2011; Shennan and Wilkinson, 2001; Steele et al., 2010), bird songs (Byers et al., 2010; Lachlan and Slater, 2003), English word frequencies since 1700 (Ruck et al., 2017), baby names (Hahn and Bentley, 2003), and Facebook app downloads (Gleeson et al., 2014). The time scales of these studies range from centuries to decades, months or days.

The two most important parameters of unbiased copying models are population size, N , and the probability, μ , of inventing a new variant (Hahn and Bentley, 2003; Neiman, 1995).

In order to compare different models in explaining the data, models need to be generalized, through multiple parameters, to generate as many outcomes as possible, taking into account sampling and different possible biases in cultural transmission (Kandler and Powell, 2018). The goal is to estimate probability distributions of those parameters to explain the set of posterior distributions.

To model the dynamics of Twitter cascades, we test several different models of context-biased learning. These models can be compared based on their ability to replicate the data while minimizing the number of model parameters. Once we have determined the best model, we then estimate a probabilistic range of each parameter values to best fit each data set. The goal is to compare parameter ranges between different data sets.

Kandler and Powell (2018) advocate the use of Approximate Bayesian Computation (ABC), which can produce a probabilistic representation of parameter space that shows how likely the parameters are to explain the data. ABC allows models to be compared using Bayesian Inference to estimate the probability that a model explains the data (posteriors) given existing knowledge of the system (priors); models are often compared using likelihood ratios.

Using this approach on the Twitter data explored by Vosoughi et al. (2018), we can select the model of social transmission that best reproduce the observation. Moreover, as prior information on Twitter users is available, we can determine with precision the distribution of biases at the individual level in the population of Twitter users. This opens the possibility of explaining *how* the observed differences (Vosoughi et al., 2018) appear.

Models of context-based re-tweeting

Here we consider a model of unbiased social learning, also known as the Neutral model. Consider a population of Twitter users in a fully-connected network. The number of modelled Twitter users, N , is kept constant, as we will assume the modelled time period

(days or weeks) is short compared to any growth in number of users.

Each Twitter user observes N randomly-selected users (self plus $N-1$ others) in each time step. In this population, users either tweet something unique of their own or else re-tweet another message. At time t , each of the re-tweeting agents chooses randomly among the N agents, and either re-tweets that agent's message, with probability $(1-\mu)$, or else composes an original new tweet, with probability μ . We run the model until reaching a steady state (for $\tau = 4\mu^{-1}$ time steps (Evans and Giometto, 2011)). In this basic unbiased copying model, a re-tweet is chosen from among N Twitter users, as opposed to choosing from the different Tweet messages themselves. This means some/all of the N Twitter users will be re-Tweeting the same message. The number of different messages, k , observed by each user is typically much less than N .

Next, we modify this unbiased model to introduce context-biases through three different forms of popularity bias. The first is a frequency bias, where the probability of a message being copied increases with frequency above the inherent frequency-dependent probability of the neutral model itself.

As social media feeds often highlight "trending" messages in some form, the other two versions represent "toplist" biases, in that Twitter users are biased towards the top y (where y is the size of a "trending" list) most popular messages (Acerbi and Bentley, 2014).

The first context-biased model derives from a more general model of discrete choice with social interactions (Brock and Durlauf, 2001; Bentley et al., 2014; Brock et al., 2014; Caiado et al., 2016). A parameter β represents the overall magnitude of social learning biases. Another parameter, J , represents context-bias, specifically popularity bias here. For the population with global β and J , a simple representation (Bentley et al., 2014; Bentley et al., 2014; Caiado et al., 2016) for the probability, P_i , that a Twitter user re-tweets message i is:

$$P_i = \frac{e^{\beta[U_i + Jp_i]}}{\sum_{i=1}^k e^{\beta[U_i + Jp_i]}} \tag{1}$$

Here the term U_i would denote the intrinsic payoff to choice i , which here could be the 'attraction' (Acerbi, 2019) of message i for (re-)Tweeting. While future models can explore this parameter, due to the computational cost to the ABC (see Discussion section), here we simply assume the messages have no intrinsic utility, i.e., $U_i = 0$ for all messages, i . This yields:

$$P_i = \frac{e^{\beta J p_i}}{\sum_{i=1}^k e^{\beta J p_i}} \tag{2}$$

In this case the context-bias, Jp_i is based upon the popularity, p_i . Note that both context- and content-bias are, respectively, homogeneous for all agents. We could, in a more advanced model, have heterogeneous distributions of J and U across all agents, but this becomes unwieldy, as the parameter space becomes too large to be explored by our current ABC algorithm.

Note also that when $\beta = 0$ and/or $J = 0$, the model reduces to a random guess model, where each choice has equal probability regardless of its frequency, i.e., $P_i = 1/k$ for all choices, i . By contrast, under the neutral (a.k.a. random copying) model, the expected frequency of each future choice is predicted by its previous frequency. Equations 1 and 2 do not reduce to the basic neutral model in a simple way; the copying is neutral in this sense only for particular momentary combinations of β and $p_i(t)$.

Next, in our "Top threshold" model, Tweets are exhibited in a "top list", such that a parameter C determines the fraction of individuals that will re-tweet a message from this list of the top y trending Tweets in the population (Acerbi and Bentley, 2014).

The other $1-C$ fraction of the population will re-tweet something else at random, per the Neutral model.

Our "Top Alberto" model, named for its inventor (Acerbi and Bentley, 2014), is a slight modification. At each time step in the Top Alberto model, a fraction, $C(1-\mu)$, of the Twitter users compare the rank of their own tweet, to y most popular tweets, and the user only re-tweets something else if their tweet is not already on the top y list. The remaining fraction of individuals $(1-C)(1-\mu)$ re-tweet as in the basic neutral model. Note that a new set of 'conformist' agents, represented as a fraction $C(1-\mu)$ of the population, are randomly selected at each time step.

In summary, our models for re-Tweeting activity, which are detailed in the Appendix, are

- Unbiased copying (Neutral model)
- Conformist copying
- Top Threshold copying
- Top Alberto copying

All the algorithmic description of the model are made available in the supplementary materials. Then to summarize the variables and parameters in these models, we have:

- i : index of message i
- p_i : popularity of message i
- β : intensity parameter
- μ : probability of writing a new tweet (as opposed to re-tweeting)
- k : number of different messages among the N users
- N : number of different Twitter users
- U_i : intrinsic payoff of message i ($U_i = 0$ here, we will explore $U_i > 0$ in the future)
- J : context bias, specifically frequency bias (universal among all agents)

While they do not span the space of all possible models, even these four models require a rigorous means of discrimination when compared to the data. In calibrating these models to Twitter cascade size, we use Approximate Bayesian Computation (Kandler and Powell, 2018).

Approximate Bayesian computation

Here we use Approximate Bayesian Computation (ABC) to calibrate our models against Twitter data. The aim is to find the distribution of parameters of each model knowing data distribution, i.e. the posterior distributions of the model. To do so one usually use Bayes equation:

$$P(\theta|D) = \frac{P(D|\theta)P(\theta)}{P(D)} \tag{3}$$

where θ is the vector of parameters for a given model, $P(\theta)$ is the prior distribution of the parameters, $P(D)$ the data and $P(D|\theta)$ the likelihood.

Finding the likelihood of our models is not an easy task, however, as there is no direct way to do so. An indirect method approximates the likelihood by simulating the models and rejecting the parameter ranges that yield results to far from the data distribution (Kandler and Powell, 2018).

ABC requires a definition of a distance between model and the data, which allows approximation of the likelihood distribution of different model parameters. Here we adapt an ABC version by Crema et al. (2016), which randomly samples the parameter space and computes a distance between the simulations and the data

using the following distance function:

$$\sqrt{\sum_{i=1}^{100} [Q_i(\log(S)) - Q_i(\log(D))]^2}, \tag{4}$$

where $Q_i(X)$ is the i th percentile of the sample X , S is the sample generated by the simulation, and D the data.

In this simple version of ABC, called the rejection algorithm, a huge number of simulations are run and only the parameters of a small percentage of simulations yielding the lowest distance to the data are kept to draw the approximate posterior distributions.

Data

Vosoughi et al. (2018) analyzed a set of 126,000 news stories distributed on Twitter from 2006 to 2017. These stories were (re-) tweeted 4.5 million times by approximately 3 million Twitter users. These news stories were classified as true and false using multiple independent fact-checking organizations, which were over 95% consistent with each other and further confirmed by undergraduate students who examined a sample of these determinations (Vosoughi et al., 2018).

For each different message in this dataset, Vosoughi et al. (2018) measured how many Twitter users re-tweeted the message, called the “cascade size”, which we model here. Tweet cascades started by bots were not a significant factor in these data Vosoughi et al. (2018). Vosoughi et al. (2018) also measured other dimensions of these Twitter cascades, included the “depth” (maximum number of successive re-tweets from the origin tweet), “breadth” (maximum number of parallel re-tweets at the same time), and time elapsed over the cascade.

Since Vosoughi et al. (2018) counted multiple cascades for certain messages, we binned identical message cascades together, yielding a group of cascade distribution for each message (Fig. 1). This reduces the dimension of the data set to one unique distribution, avoids the need to keep the full structure of each cascade and allows to directly compare our model to the data collected by Vosoughi et al. (2018).

Figure 1a shows the distribution of cascade sizes when each cascade is taken separately, Fig. 1b shows aggregate cascade sizes where we have aggregated the number of re-tweets for cascades of identical messages.

Results

Model Selection. To formally select between the different models we can use the posterior distributions of the different models given the data. The Bayes equation described by the Eq. 3 becomes:

$$P(m|D) = \frac{P(D|m)P(m)}{P(D)} \tag{5}$$

where $P(m)$ is the prior and the likelihood $P(D|m)$ is estimated through Approximate Bayesian Computation (Toni et al. 2009, Toni and Stumpf, 2010), and the probability of the data, $P(D)$, in the denominator cancels out when we compare models to each other. To calculate $P(m|D)$, we define a level of acceptance, λ , that determines the number of simulations we will accept (ie. accept the λ best simulations). Then we calculate how many simulations of each model are below this acceptance level. Table 1 summarizes this distribution of m for different levels of $\lambda \in [500, 5000, 50000]$.

We note that the Top Threshold is by far the least likely model to explain the data. The best models in Table 1 are the Unbiased and Top Alberto models. Since the Bayes factors do not change much even if we divide the level λ by 100, the accepted simulations appear to be a good approximation of the real distribution.

To compare models more formally, having used uniform prior probability distributions for all models, we can compute the Bayes

Table 1 Bayes factor table for different acceptance ratio, for distribution of true (top) and false (bottom) rumors

	$P(m D, \lambda = 500)$	$P(m D, \lambda = 5000)$	$P(m D, \lambda = 50000)$
<i>m</i> vs True			
Unbiased (1)	0.307	0.315	0.352
Conformist (2)	0.298	0.303	0.273
Top	0.048	0.047	0.110
Threshold (3)			
Top Alberto (4)	0.302	0.335	0.311
<i>m</i> vs False			
Unbiased (1)	0.384	0.351	0.323
Conformist (2)	0.198	0.277	0.270
Top	0.006	0.018	0.086
Threshold (3)			
Top Alberto (4)	0.412	0.354	0.321

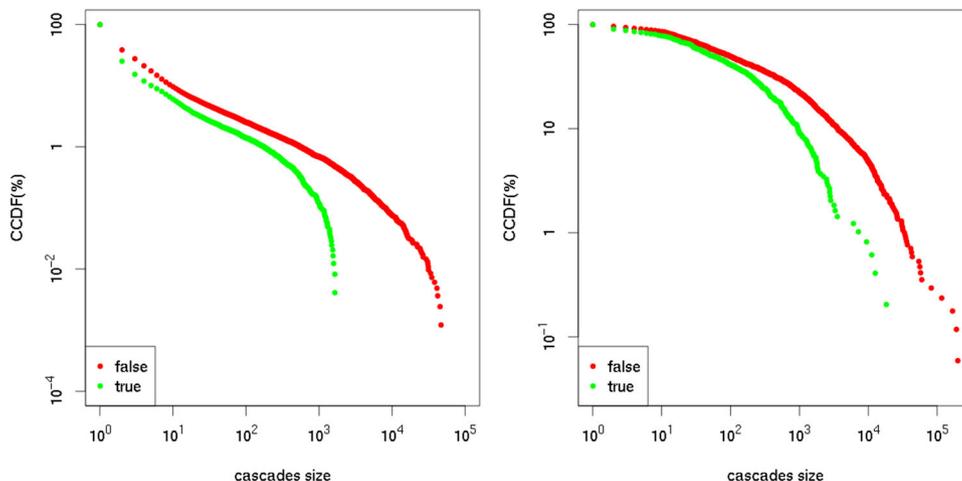


Fig. 1 Complementary Cumulative Distribution Functions (CCDFs) of (a) non aggregated cascades vs (b) aggregated cascades with same rumors. Those CCDFs represent the percentage of respectively (a) cascades and (b) rumors that have reached a given number of re-tweets between 2007 and 2017

Factor K_{m_A, m_B} between pairs of models as follows:

$$K_{(m=A, m=B)} = \frac{P(m = A|D)}{P(m = B|D)} \tag{6}$$

For the smallest level of acceptance, $\lambda = 500$, and when modelling the true twitter cascades, the Bayes factors show the Unbiased model to be slightly better than the Conformist model ($K = 1.18$) or the Top Alberto model ($K = 1.19$), while the Conformist and Top Alberto models are equivalent ($K = 1.01$) and the Top Threshold model is highly unlikely compared to the other three models ($K < 0.16$).

Similarly, when modelling the cascades of false tweets, the Top Threshold model is highly unlikely ($K < 0.02$) compared to any of the other three models. For false tweets, Unbiased and Top Alberto models are equally good ($K = 1.07$) and do better than the Conformist model ($K > 2.0$).

The number of parameters is, implicitly, taken into account in the Bayes factor: To approximate the likelihood while doing the ABC we randomly sample the same number of data points from the prior distribution, thus if the number of parameters for one model is higher, the parameter space is bigger and the sample size drawn from the prior will cover a smaller fraction of the total space, yielding a lower probability to find good simulations that fall under our λ threshold.

To calculate something comparable to AIC, we use the raw values from Table 1 divided by the total number of simulations. This would give us the approximated likelihood, L , for each model. Then AIC is $-2 \times \ln L + 2p$ with L the likelihood and p the number of parameters. This gives a set of “corrected” Bayes factors as in Table 2, in which the Unbiased (basic Neutral) model (1) is the best for both sets of data.

Posterior distributions. The ABC algorithm allows us not only to select between the models but also to look at the posterior distribution of the parameters that yield to simulations reproducing the data. The idea is then to explore the result of the Eq. 3, once the likelihood $P(D|\theta)$ has been approximated by the ABC.

As the Unbiased model is the most likely, we present only its posterior probability distributions in Fig. 2. Each panel of Fig. 2 compares the posteriors of the ABC done with the true tweets (in green) against the ABC done with the false tweets (in red) and the prior used in both case (in grey).

Figure 2 indicates the effect of the invention rate, μ , population size, N , and run-in time, τ . The most well-defined probability peak related to false tweet cascades is the invention rate, μ . There are fewer time steps required in modelling the true tweets compared to the false tweets. This is consistent with false tweet cascades persisting longer (Vosoughi et al. 2018). Similar figures for the three context-dependent models are shown in the supplementary material.

From those posterior distributions it is possible to determine the parameter range with the highest probability: the highest density region (HDR) (Hyndman, 1996) – often also called the Highest Posterior Density Region (HPD) when they are calculated on posterior distributions, as it is the case here. We

use the R package HDRCDE (Hyndman, 2018) to calculate those HPDs. The resulting intervals and modes are given in the Tables 3 and 4.

Posterior checks. For the ABC, since we could not store the full results of all simulations, we saved only the parameters used together with the distance to the data. Keeping this information for the 9 million simulations we ran for each model yielded about 1.7 TB of data. Thus, to check the adequacy between the simulations selected through ABC and the observed distribution, we

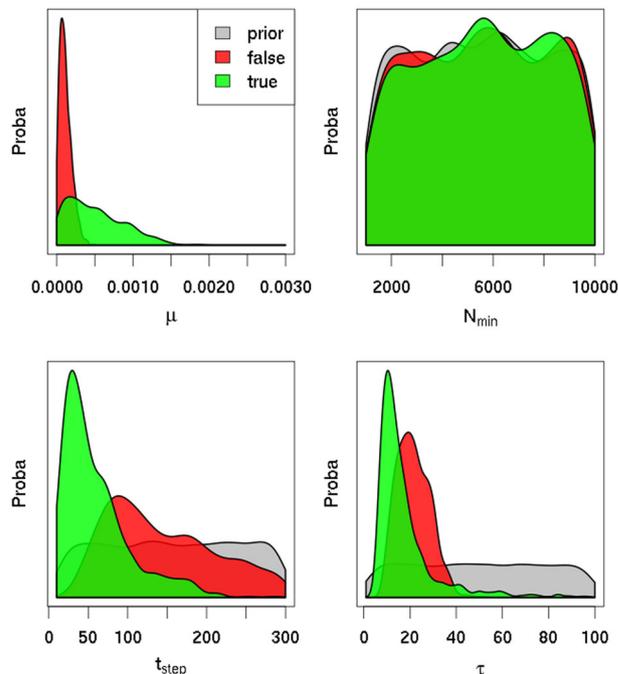


Fig. 2 ABC posterior probability estimations for several parameters— invention rate μ , population size N , run time t_{step} and run-in time τ —of the unbiased (random) copying model, when the model is fit to the cascade size distributions of true (green) and false (red) tweets, respectively. The grey curve represents the prior distributions for each parameters

Table 3 Mode and 95% interval of the High Posterior Density region for the parameters of the unbiased model with respect to the distribution of true tweets

Parameter	mode	95% HDP
μ	2.48×10^{-4}	$[1.04 \times 10^{-5}; 0.12]$
N	5532	$[1050; 9579]$
t_{step}	23	$[5; 158]$
τ	9	$[2.7; 43]$

Table 4 Mode and 95% interval of the high posterior density region for the parameters of the unbiased model with respect to the distribution of false tweets

Parameter	mode	95% HDP
μ	1.226×10^{-3}	$[6 \times 10^{-7}; 2.6 \times 10^{-2}]$
N	1528	$[1091; 9599]$
t_{step}	81	$[32; 274]$
τ	17	$[8; 34]$

Table 2 Corrected Bayes factor table for different acceptance ratio

Model	vs True	vs False
Unbiased (1)	25.38	25.21
Conformist (2)	27.72	28.54
Top Threshold (3)	33.37	37.53
Top Alberto (4)	29.69	29.07

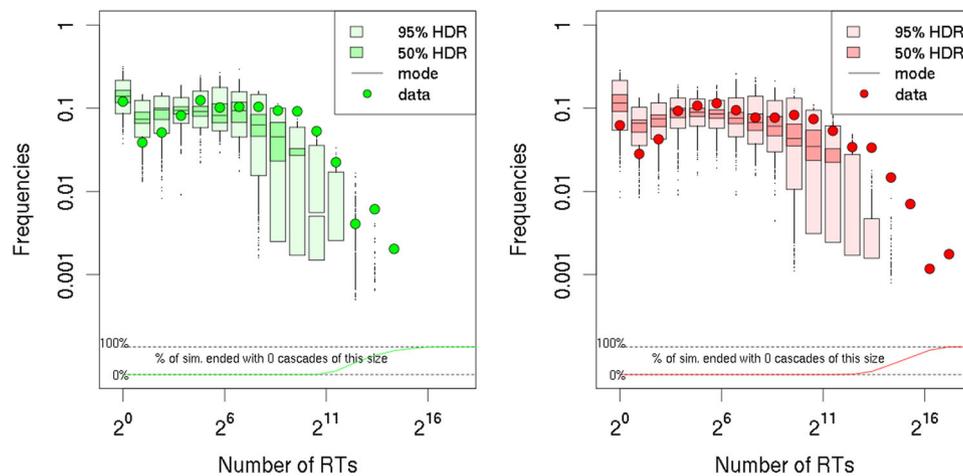


Fig. 3 Posterior check of distributions of aggregated cascade sizes for the Unbiased neutral model versus data from true rumors at left (in green) and false rumors at right (in red). Each plot represents the percentage of rumors for which the accumulate number of RT falls within 18 bins of logarithmically growing size. The frequency of rumors within each bin is represented by a colored dot for data set, versus the mode and High Density Regions for the 10,000 posterior checks of the model. The curve at the bottom of each plot shows the percentage of simulations where zero rumors felt within the given bin. Note Figs 4–6 use this same format

run a new set of simulations. For every model, we re-run 10,000 simulations, sampling the parameters from the selected posteriors distributions, for both the posteriors obtained with true as well as false messages. This is also often called posterior predictive check (Gelman and Hill, 2007).

We present the results of the new simulations as distributions of cascade sizes. For a better visualization, we binned the cascades with similar size within logarithmic bins. The High Density Regions for all bins and models are represented in Figs 3 to 6. The colored dots represent the data from Vosoughi et al. (2018). The raw data (i.e., without the binning and the HDRs) are given in Figs 4–7 of the supplementary material.

Figures 3–6 show how the respective models span a range of size distributions of aggregated cascade sizes. In each case, the simulation results are compared to the actual data of Vosoughi et al. (2018), for both the true messages and the false messages. The simulations and models fit well, except for underestimating the sizes of the top four or five largest tweet cascades, in both false and true categories (Figs 3–6). The models, particularly the Unbiased (a.k.a. random copying) model, otherwise predict the rest of the distribution of cascade sizes, albeit with different goodness of fit, which we consider below.

Discussion

In calibrating neutral model variations against distributions of Twitter cascade sizes, we find that the Unbiased neutral model applies well to re-tweeting activity, and better than models with added conformity bias. Among our advances here is the use of ABC to resolve the joint probability distribution of neutral model variations (Kandler and Powell, 2018).

Here we have applied the same set of models to both data sets, with the expectation that the different models, and the probability distributions on the parameter values for each model, would differ meaningfully between the false and true tweets. As Fig. 2 and Tables 1–4 show, we find that the Unbiased (and Top Alberto) model fits the distribution of both true and false Tweets, but the invention rate, μ , for the false Tweets (0.00123) is about five times higher than for the true Tweets (0.00025).

The modal value of the invention rate, μ , returned from our model runs was 0.00025 for the true tweets and 0.00123 for the false tweets (Tables 3 and 4). These values agree well with

estimates we can make from the data from Vosoughi et al. (2018, S2.3), which comprise 2448 different rumors re-tweeted about 4.5 M times, about two thirds of which (~3M) were re-tweeting the 1699 false rumors and the other third (~1.5 M) re-tweeting the 490 true rumors. This implies that about 0.00032 of the true tweets and 0.00057 of the false tweets were original, both of which are well within the High Posterior Density region for μ of the respective models (Tables 3 and 4).

In our tests, the most important parameter was the invention rate, μ , particularly in modelling the distribution of false tweet cascades. Another important parameter was the transmission bias, such that neutrality ($\beta = 0$) and positive-frequency bias ($\beta > 0$) can be evaluated in terms of likelihood of explaining the data (Kandler and Powell, 2018). The biased models performed worse than the Unbiased ($\beta = 0$) model, as the biased models failed to generate the largest cascades (Figs 3–6). This is not due to limits on modelled population size; if it were, we would expect the posterior distribution for N to be more right-skewed than it was. Indeed the value of N had little effect on outcome (Supplementary material).

Replicating the largest cascades was a challenge for all models. This suggests a few ways forward for future modifications. One is to allow for interdependence between the distribution of true and false messages, rather than as separate data sets, since real Twitter users are exposed to both kind of messages. Another is to specify the network of connections among agents (Lieberman et al., 2005; Ormerod et al., 2012).

Heterogeneous networks may help resolve a discrepancy between our findings—that conformist biases were unhelpful in modelling Twitter data—and the highly-skewed nature of influence on social media. In a recent study of social media (Grinberg et al., 2019), “Only 1% of individuals accounted for 80% of fake news source exposures, and 0.1% accounted for nearly 80% of fake news sources shared.” This phenomenon is not unique to social media; in order to fit the Neutral model to evolving English word frequencies over 300 years of books, Ruck et al. (2017) needed to assume that most of the copying was directed to a relatively small corpus of books, or “canon”, within the larger population of millions of books.

Testing such models will require more granular data, including the content and word counts from tweeted messages, than we had access to in this study. If counts of specific words through time are available, then additional diagnostic signatures include both the Zipf law of ranked word frequencies and turnover within “top y”

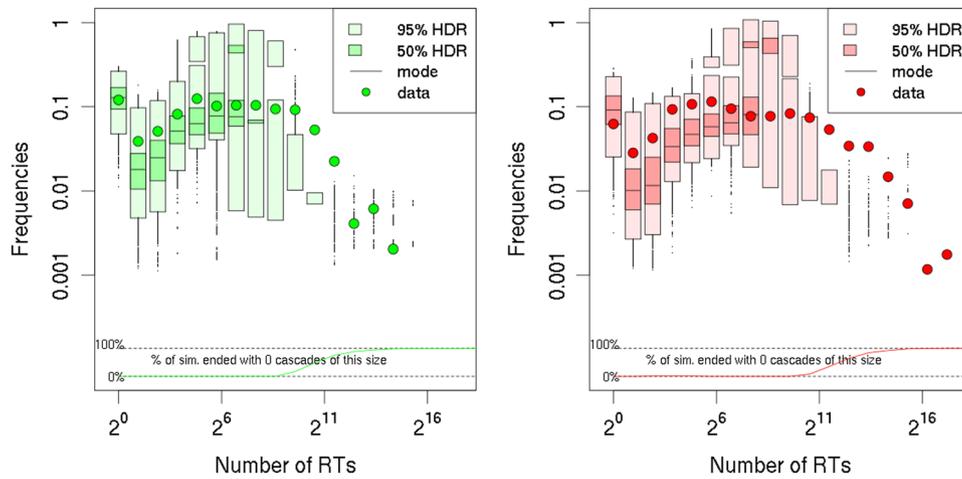


Fig. 4 Posterior check of distributions of aggregated cascade sizes for the Conformist model, versus the data for true rumors (in green) on the left and false rumors (red) on the right. Plots were generated as described in the caption of Fig. 3

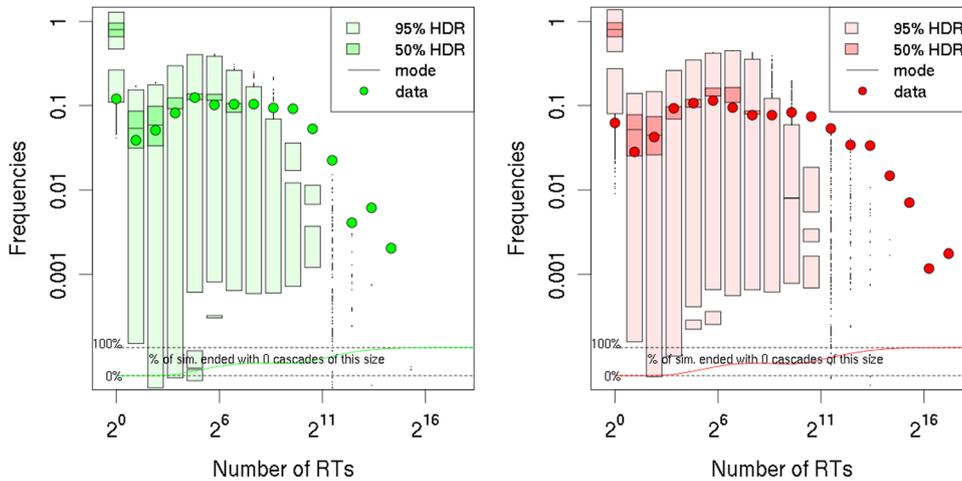


Fig. 5 Posterior check of distributions of aggregated cascade sizes for the Top Threshold model, versus the data for true rumors (in green) on the left and false rumors (red) on the right. Plots were generated as described in the caption of Fig. 3

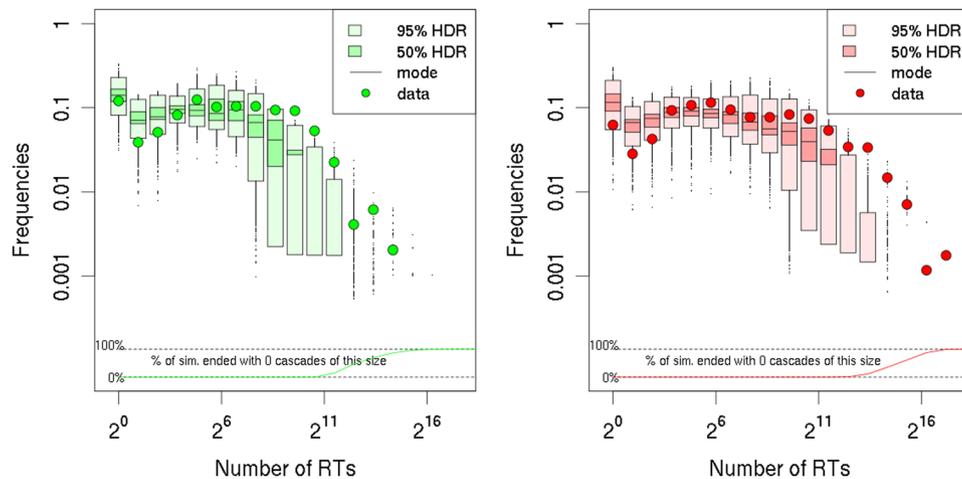


Fig. 6 Posterior check of distributions of aggregated cascade sizes for the Top Alberto model, versus the data for true rumors (in green) on the left and false rumors (red) on the right. Plots were generated as described in the caption of Fig. 3

lists of those words (Acerbi and Bentley, 2014; Bentley et al., 2007; Ruck et al., 2017). While we opted for parsimony here, more granular data would justify testing more complicated Neutral model modifications, such as non-equilibrium assumptions (Crema et al., 2016), variable “memory” (Bentley et al., 2011; Gleeson et al., 2014) and isolation by distance effects (Bentley et al., 2014).

For this study, we needed to assume messages had no intrinsic payoff, i.e., $U_i = 0$ for all messages, i . This was necessary because the additional parameters needed to represent heterogeneous population of individual would require running on the order of 10^{14} simulations to explore the content under our ABC conditions. In the future, we can use more advanced ABC algorithms (Beaumont et al., 2009), which optimize the exploration of the parameter space, to explore the content-bias model in Eq. 1. The goal would be to estimate, for an assumed distribution of β and J among the audience, the mean utility U_i from a source of tweets, from the statistical patterns of the avalanches. Alternatively, we might attempt to derive, for a given mean payoff U_i of the source’s tweets, the full distribution of β and J across all the agents in the source’s audience.

Conclusion

Here we have tested variations of culture-evolutionary neutral models on aggregated Twitter data documenting the spread of true and false information. We used Approximate Bayesian Computation to resolve the full joint probability distribution of models with different social learning biases, emphasizing context-biased versus content-biased learning. Our study here begins to address how online social learning dynamics can be modelled through the tools of cultural evolutionary theory.

Data availability

The raw data used to calculate the distribution of cascades size and aggregated size are available upon demand using the form: <https://docs.google.com/forms/d/e/1FAIpQLSdVL9q8w3MG6myI48FI5X45SmnRzGoOEDBROe-BoNni5lbfKw/viewform>, The code used to generate the simulated data is available as a R-package at: <https://github.com/simoncarrignon/spreadrt>. Examples on how to use this code and regenerate all the data used in this paper are provided with the package.

Received: 31 March 2019 Accepted: 11 July 2019

Published online: 30 July 2019

References

- Acerbi A (2019) Cognitive attraction and online misinformation. *Palgrave Commun* 5(1):15
- Acerbi A, Bentley RA (2014) Biases in cultural transmission shape the turnover of popular traits. *Evol Hum Behav* 35(3):228–236
- Acerbi A, Enquist M, Ghirlanda S (2009) Cultural evolution and individual development of openness and conservatism. *Proc Natl Acad Sci USA* 106(45):18931–18935
- Acerbi A, Mesoudi A (2015) If we are all cultural Darwinians what’s the fuss about? Clarifying recent disagreements in the field of cultural evolution. *Biol Philos* 30(4):481–503
- Barakzai A, Shaw A (2018) Friends without benefits: When we react negatively to helpful and generous friends. *Evol Hum Behav* 39(5):529–537
- Beaumont MA, Cornuet J-M, Marin J-M, Robert CP (2009) Adaptive approximate Bayesian computation. *Biometrika* 96(4):983–990
- Bentley RA, Caiado CCS, Ormerod P (2014) Effects of memory on spatial heterogeneity in neutrally transmitted culture. *Evol Hum Behav* 35:257–263
- Bentley RA, Hahn MW, Shennan SJ (2004) Random drift and culture change. *Proc B* 271:1443–1450
- Bentley RA, Lipo CP, Herzog HA, Hahn MW (2007) Regular rates of popular culture change reflect random copying. *Evol Hum Behav* 28:151–158
- Bentley RA, O’Brien MJ (2017) The acceleration of cultural change: from ancestors to algorithms. M.I.T. Press, Cambridge, MA
- Bentley RA, O’Brien MJ, Brock WA (2014) Mapping collective behavior in the big-data era. *Behav Brain Sci* 37:63–119
- Bentley RA, Ormerod P (2010) A rapid method for assessing social versus independent interest in health issues: A case study of “bird flu” and “swine flu”. *Soc Sci Med* 71:482–485
- Bentley RA, Ormerod P, Batty M (2011) Evolving social influence in large populations. *Behav Ecol Sociobiol* 65:537–546
- Bentley RA, Shennan SJ (2003) Cultural transmission and stochastic network growth. *Am Antiq* 68:459–485
- Brock WA, Bentley RA, O’Brien MJ, Caiado CCS (2014) Estimating a path through a map of decision making. *PLoS ONE* 9(11):e111022
- Brock WA, Durlauf SN (2001) Discrete choice with social interactions. *Rev Econ Stud* 68:229–272
- Byers BE, Belinsky KL, Bentley RA (2010) Independent cultural evolution of two song traditions in the chestnut-sided warbler. *Am Nat* 176:476–489
- Caiado CCS, Brock WA, Bentley RA, O’Brien MJ (2016) Fitness landscapes among many options under social influence. *J Theor Biol* 405:5–16
- Crema ER, Kandler A, Shennan SJ (2016) Revealing patterns of cultural transmission from frequency data: equilibrium and non-equilibrium assumptions. *Sci Rep* 6:39122
- Crone EA, Konijn EA (2018) Media use and brain development during adolescence. *Nat Commun* 9:Article 588
- Danvers AF, Hackman JV, Hruschka DJ (2019) The amplifying role of need in giving decisions. *Evol Human Behav* 40(2):188–193
- Dunbar RIM, Shultz S (2007) Evolution in the social brain. *Science* 317:1344–1347
- Eerkens JW, Lipo CP (2007) Cultural transmission theory and the archaeological record: providing context to understanding variation and temporal changes in material culture. *J Archaeol Res* 15:239–274
- Eichstaedt JC, Schwartz HA, Kern ML, Park G, Labarthe DR, Merchant RM, Jha S, Agrawal M, Dziurzynski LA, Sap M, Weeg C, Larson EE, Ungar LH, Seligman MEP (2015) Psychological language on Twitter predicts county-Level heart disease mortality. *Psychol Sci* 26:159–169
- Evans TS, Giometto A (2011) Turnover rate of popularity charts in neutral models. Preprint at <https://arxiv.org/abs/1105.4044>
- Falk EB, Bassett DS (2017) Brain and social networks: Fundamental building blocks of human experience. *Trends Cogn Sci* 21(9):674–690
- Gelman A, Hill J (2007) Data analysis using regression and hierarchical/multilevel models. Cambridge University Press, New York, NY
- Gleeson JP, Cellai D, Onnela J-P, Porter MA, Reed-Tsochias F (2014) A simple generative model of collective online behavior. *PNAS* 111:10411–10415
- Golder SA, Macy MW (2011) Diurnal and seasonal mood vary with work, sleep, and daylight across diverse cultures. *Science* 333:1878–1881
- Grinberg N, Joseph K, Friedland L, Swire-Thompson B, Lazer D (2019) Fake news on Twitter during the 2016 U.S. presidential election. *Science* 363:374–378
- Hahn MW, Bentley RA (2003) Drift as a mechanism for cultural change: an example from baby names. *Proc B* 270:S1–S4
- Hidalgo C (2015) Why information grows: the evolution of order, from atoms to economies. Basic Books, New York
- Hyndman RJ (1996) Computing and graphing highest density regions. *Am Stat* 50(2):120–126
- Hyndman RJ (2018) *hdcrc: highest density regions and conditional density estimation*. R package version 3.3. <https://cran.r-project.org/web/packages/hdcrc/index.html>
- Kandler A, Powell A (2018) Generative inference for cultural evolution. *Philos Trans Roy Soc B* 373(1743):20170056
- Kandler A, Shennan SJ (2013) A non-equilibrium neutral model for analysing cultural change. *J Theor Biol* 330:18–25
- Kendal RL, Boogert NJ, Rendell L, Laland KN, Webster M, Jones PL (2018) Social learning strategies: Bridge-building between fields. *Trends Cogn Sci* 22(7):651–665
- Lachlan RF, Slater PJB (2003) Song learning by chaffinches: how accurately and from where? A simulation analysis of patterns of geographical variation. *Anim Behav* 65:957–969
- Lazer D, Kennedy R, King G, Vespignani A (2014) The parable of Google Flu: traps in big data analysis. *Science* 343:1203–1205
- Lenhart A (2015) Teens, social media and technology overview 2015. [online] Pew Research Center. <https://www.pewinternet.org/2015/04/09/teens-social-media-technology-2015>. Accessed 6 Nov 2018
- Lieberman E, Hauert C, Nowak MA (2005) Evolutionary dynamics on graphs. *Nature* 433:312–316
- Mesoudi A, Chang L, Murray K, Lu HJ (2015) Higher frequency of social learning in China than in the West shows cultural variation in the dynamics of cultural evolution. *Proc R Soc B* 282:20142209
- Mesoudi A, Lycett SJ (2009) Random copying, frequency-dependent copying and culture change. *Evol Hum Behav* 30(1):41–48

- Molleman L, Gächter S (2018) Societal background influences social learning in cooperative decision making. *Evol Hum Behav* 39(5):547–555
- Neiman FD (1995) Stylistic variation in evolutionary perspective. *Am Antiq* 60:7–36
- Obama B (2017) Farewell address, Jan 10, 2017. <https://obamawhitehouse.archives.gov/farewell>. Accessed 22 Oct 2018
- Ormerod P, Nyman R, Bentley RA (2014) Nowcasting economic and social data: when and why search engine data fails, an illustration using Google Flu Trends. Preprint at <https://arxiv.org/abs/1408.0699>
- Ormerod P, Tarbush B, Bentley RA (2012) Social network markets: the influence of network structure when consumers face decisions over many similar choices. Preprint at <https://arxiv.org/abs/1210.1646>
- Premo LS (2014) Cultural transmission and diversity in time-averaged assemblages. *Curr Anthropol* 55(1):105–114
- Premo LS, Scholnick J (2011) The spatial scale of social learning affects cultural diversity. *Am Antiq* 76(1):163–176
- Realí F, Griffiths TL (2010) Words as alleles: connecting language evolution with Bayesian learners to models of genetic drift. *Proc B* 277:429–436
- Rendell L, Boyd R, Cownden D, Enquist M, Eriksson K, Feldman MW, Fogarty L, Ghirlanda S, Lillcrap T, Laland KN (2010) Why copy others? Insights from the social learning strategies tournament. *Science* 328:208–213
- Robertson TE, Sznycer D, Delton AW, Tooby J, Cosmedes L (2018) The true trigger of shame: social devaluation is sufficient, wrongdoing is unnecessary. *Evol Hum Behav* 39(5):566–573
- Ruck DJ, Bentley RA, Acerbi A, Garnett P, Hruschka D (2017) Role of neutral evolution in word turnover during centuries of english word popularity. *Adv Complex Syst* 20:1750012
- Ruck DJ, Rice NM, Borycz J, Bentley RA (2019) Internet research agency Twitter activity predicted 2016 U.S. election polls. First Monday, July 1, 2019. <https://journals.uic.edu/ojs/index.php/fm/article/view/10107>
- Shennan SJ, Wilkinson JR (2001) Ceramic style change and neutral evolution: A case study from Neolithic Europe. *Am Antiq* 66(4):577–593
- Smith D, Dyble M, Major K, Page AE, Chaudhary N, Salali GD, Thompson J, Vinicius L, Migliano AB, Mace R (2019) A friend in need is a friend indeed: Need-based sharing, rather than cooperative assortment, predicts experimental resource transfers among Agta hunter-gatherers. *Evol Human Behav* 40(1):82–89
- Steele J, Glatz C, Kandler A (2010) Ceramic diversity, random copying, and tests for selectivity in ceramic production. *J Archaeol Sci* 37:1348–1358
- Toni T, Stumpf MPH (2010) Simulation-based model selection for dynamical systems in systems and population biology. *Bioinformatics* 26(1):104–110
- Toni T, Welch D, Strelkova N, Ipsen A, Stumpf MPH (2009) Approximate Bayesian computation scheme for parameter inference and model selection in dynamical systems. *J R Soc Interface* 6(31):187–202
- Vosoughi S, Roy D, Aral S (2018) The spread of true and false news online. *Science* 359:1146–1151

Acknowledgements

Funding for this work was provided by the ERC Advanced Grant EPNet (340828). Computational power was made available by the Barcelona Supercomputing Center (BSC). We thank the National Institute for Mathematical and Biological Sciences, University of Tennessee, for support to SC in summer 2018 and the Severo Ochoa Mobility Program for funding his stay. DJR was supported by grants from the University of Tennessee and the National Science Foundation (DataONE project).

Additional information

The online version of this article (<https://doi.org/10.1057/s41599-019-0295-9>) contains supplementary material, which is available to authorized users.

Competing interests: The authors declare that they have no conflict of interest.

Reprints and permission information is available online at <http://www.nature.com/reprints>

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2019