Supplementary Information

Ultrametric distribution of culture vectors in an extended Axelrod model of cultural dissemination

Alex Stivala Garry Robins Yoshihisa Kashima Michael Kirley

Melbourne School of Psychological Sciences, The University of Melbourne, 3010, Australia
Alex Stivala, Garry Robins, Yoshihisa Kashima

Department of Computing and Information Systems, The University of Melbourne, 3010, Australia
Michael Kirley

Corresponding author
Alex Stivala (stivalaa@unimelb.edu.au)

1 Details of the models

As described in the main text, we use two models: the simple Axelrod model, and the extended Axelrod model. Both models use the same definition of cultural similarity.

We will denote element i of the culture vector for agent a by $v_i^{(a)}$. The cultural similarity of two agents, a normalised Hamming similarity, is then

$$c(a,b) = \frac{1}{F}\sum_{i=1}^{F} \delta_{v_i^{(a)},v_i^{(b)}} \tag{1}$$

where $\delta_{x,y}$ is the Kronecker delta function.

The simple Axelrod model is similar to the one used in [1]: an agent can interact with any other agent (the social network is a complete graph). The bounded confidence threshold θ means that interactions between agents with cultural similarity less than θ do not occur.

The extended Axelrod model is described in [2]. The only further extension we make is to incorporate the bounded confidence threshold θ. Our interpretation of bounded confidence is that cultural similarity less than the threshold results in an *unsuccessful* interaction, leading to a decrease in social link weight, and possible migration. (Note that in the simple Axelrod model there is no notion of such an unsuccessful interaction: any interaction results in a change of a cultural trait, and agents with cultural similarity less than θ simply do not interact).

Initially, agents are distributed randomly on the lattice, and all social link weights are 0. In the original model, as is traditional, culture vectors are assigned uniform random values. However, we will also use real and simulated data with different distributions, as described in the main article.

At each step of the simulation, the probability of an interaction between two agents a and b is:

$$I(a,b) = \begin{cases} \frac{1}{Z}\left[d_L(a,b)^{-\beta_s}\max\{s(a,b),k_s\}\right] & \text{if } a \neq b \\ 0 & \text{otherwise.} \end{cases} \tag{2}$$

where $s(a,b)$ is the weight of the social link between agents a and b, $d_L(x,y)$ is the Euclidean distance between two lattice positions x and y, normalised by the diagonal distance $\sqrt{2}L$, Z is a normalisation factor, and $k_s = 0.01$ allows interaction between agents with no social tie (that is, when $s(a,b) = 0$). The parameter β_s controls social mobility: larger values mean more social mobility.

The probability of a *successful* interaction is then

$$I_{\mathrm{S}}(a,b) \quad = \quad \begin{cases} c(a,b) & \text{if } c(a,b) \geq \theta \\ 0 & \text{otherwise.} \end{cases} \tag{3}$$

and so the probability of an unsuccessful interaction is $1-I_{\mathrm{S}}$. On a successful interaction where $c(a,b) < 1$, a feature i such that $v_i^{(a)} \neq v_i^{(b)}$ is chosen randomly and the assignment $v_i^{(a)} \leftarrow v_i^{(b)}$ is made, so that the two agents become more culturally similar. In addition, the strength of the social tie $s(a,b)$ is increased, up to a maximum of 1.

On an unsuccessful interaction between agents a and b, the weight of the social tie is reduced, down to a minimum of 0. In addition, agent a migrates towards one of its social contacts w, with probability equal to the weight of the social link $s(a,w)$. The probability of a migrating to a lattice position t is:

$$M(a,t) \quad = \quad \begin{cases} \frac{1}{Z'} e^{-\beta_p d_L(a,t)} & \text{if } d_L(t,w) < d_L(a,w) \text{ and } t \text{ is unoccupied,} \\ 0 & \text{otherwise.} \end{cases} \tag{4}$$

where Z' is a normalisation factor. The parameter β_p controls physical mobility: larger values mean smaller geographical mobility.

The simulation continues until an absorbing state, in which no more changes are possible, is reached. There are two possible absorbing states: *monocultural* and *multicultural*. In a monocultural absorbing state, all agents have the same culture vector and the social network is a single connected component. $c(a,b) = 1$ and $s(a,b) = 1$ for all agents a, b, and all physical, cultural and social states are fixed.

In a multicultural absorbing state, any two agents a, b with the same culture have $c(a,b) = 1$ and $s(a,b) = 1$. Any two agents from different cultures have $c(a,b) = 0 \vee c(a,b) < \theta$ and $s(a,b) = 0$. Social and cultural states are fixed, however geographical migration is still possible. States that differ only in geographical position of agents are considered equivalent; we consider an absorbing state to be reached when cultural and social states are fixed.

Source code for the Axelrod models used in the paper, as well as scripts and data for generating the figures, is available from `http://munk.csse.unimelb.edu.au/~astivala/ultrametric_axelrod/`.

2 Measuring ultrametricity

As discussed in the main text, it is notable that the cophenetic correlation coefficient largely accords with the intuitive notion, as visualised using dendrograms (for example Fig. S1 and Fig. S2), that real data is more ultrametric than permuted or random data, as described in [1], while ultrametric triangle fraction does not. We therefore use the cophenetic correlation coefficient, rather than a more direct measurement, such as the ultrametric triangle fraction, to measure the degree of ultrametricity.

It might then be contended that we are not actually measuring ultrametricity at all, but rather closeness or amenability to hierarchical clustering. In a sense this is true; perfectly ultrametric data has a hierarchical clustering that induces cophenetic distances identical with those in the data (and hence has cophenetic correlation coefficient 1), while it is not necessarily true that a data set with a higher cophenetic correlation coefficient is literally "more ultrametric" as measured by the ultrametric triangle fraction (such cases are easily seen in Tables 1 and 2 in the main text, as well as Tables S1 and S2, for example). Nevertheless, we will follow the example of Rammal [3, 4] in using such an indirect measurement based on a constructed ultrametric, which also accords with the use of "ultrametricity" in Valori *et al.* [1] in which it is maintained that the real Eurobarometer data is "more ultrametric" than random data.

3 Empirical data and split ballots

Similarly (but not identical) to [1], we used the survey responses for opinions on science and technology from the Eurobarometer survey data. There are 162 such questions. Some sets of questions in the survey have a "split ballot" design, in which two versions (alternative wordings) of the question are given, with 50% of respondents being given one version and the other 50% the other. We handle the split ballot by "merging" the two versions, so that the final result only has one version, and therefore fewer total

opinions. This is done as follows: in a split ballot question, only either version A or version B have real responses, and the other version's responses are coded as "inapplicable". Our merged responses use only version A — in the case that the question uses version B, the responses are converted to corresponding responses to the version A question. This process results in $F = 116$ opinions in the merged data. We note that this is not the same as the procedure used in [1], who handle the split ballot by converting "inapplicable" responses instead to the "most neutral" value (0.5 in a metric response scale of 0 to 1), and not reducing the total number of opinions F.

The original data contains responses from approximately 13 000 individuals; we randomly sample 50 individuals from each of the 12 European countries, giving a total of $N = 600$ individuals. We simply take each integer response as a trait, defining cultural similarities with Equation 1. The largest number of possible responses to any question in this set is 12, so we set $q = 12$. As in [1] we also use random initial conditions, and also shuffled (permuted) real data (each trait's values permuted among individuals so as to destroy correlations between responses by the same individual).

Table S1 shows the mean and standard deviation of inter-vector distances in Eurobarometer data (real, shuffled, random and simulated), as well as the degree of ultrametricity as measured by the cophenetic correlation coefficient, and the ultrametric triangle fraction. Fig. S1 shows the corresponding single-linkage clustering dendrogram. The split ballot questions of the opinion data in this table have been handled similarly to the method used in [1], by converting the inapplicable version responses to the "most neutral" response. This quantifies the assertion in [1] that the real data is "more ultrametric" than the shuffled and random data, when measured using the cophenetic correlation coefficient (but not, as discussed in the main text, when using the ultrametric triangle fraction). Table 1 in the main article shows the Eurobarometer data with split ballot handled according to our method, by merging the different versions, and Fig. S2 shows the corresponding single-linkage clustering dendrogram. Note that this results in a much lower degree of ultrametricity, and in fact the permuted data now has a slightly higher cophenetic correlation coefficient that the real data.

As a second empirical data set for initial opinion vectors, we use the same approach as for the Eurobarometer data, applied to the General Social Survey (GSS) 1993 data [5]. Here we extract 61 opinions on science, the environment, and technology, with $q = 10$. Following the same procedure for merging split ballot questions as with the Eurobarometer data, the 61 original questions are reduced to $F = 58$ opinions. From the approximately 1 600 individuals in the data, we randomly sample $N = 600$ individuals.

4 Generated data

As described in the main article, there are three schemes used to create initial opinion vectors with an ultrametric distribution. The first, "neutral evolution", has a parameter r, the number of traits in each vector to change to a random value. The algorithm is shown as Algorithm 1. The results in the main article use $r = \lfloor \frac{F}{2} \rfloor$, that is, up to 50% of the traits are changed to a random value. Here we also show results for smaller values of r.

The second scheme, "prototype evolution" has a parameter k, the number of initial prototype vectors to use. The algorithm is shown as Algorithm 2. The results in the main article use $k = 3$. Here we show results for other values of k.

The third scheme "trivial ultrametric" has no parameters, and generates data that is perfectly ultrametric, but has constant and very small intervector distance. The algorithm is shown as Algorithm 3.

Intermediate values of the perturbation probability p give intermediate degrees of ultrametricity (measured by cophenetic correlation coefficient), as shown in Table S2 and in Fig. S3 for neutral evolution with $r = 1$, Fig. S4 for neutral evolution with $r = \lfloor \frac{F}{2} \rfloor$ and Fig. S5 for prototype evolution with $k = 3$. Note that, as shown in these figures, ultrametricity (at least as measured by the cophenetic correlation coefficient) becomes more apparent with increasing F; it is known that ultrametricity increases with increasing dimension (F) and with increasing sparsity [4, 6].

These figures also show the ultrametric triangle fraction (the fraction of all triples of points that form either isosceles or equilateral triangles with their intervector distances). Note that, for the three "evolution" schemes, the ultrametric triangle fraction is either approximately constant or actually increases as the perturbation probability p increases (an exception is neutral evolution with $r = 1$ in Fig. S3B, where for $F \geq 50$ the ultrametric triangle fraction is relatively high at $p = 0$, then drops as soon as $p > 0$

Algorithm 1 Neutral evolution. Given an initial culture vector c_0, current list of vectors v (initially empty), integer m and integer r ($1 \leq r \leq F$), the number of traits to change at each step, returns a list of 2^m vectors. F and q, the vector dimension and number of trait values respectively, are global constants. NeutralEvolution is a recursive function that generates 2^m vectors; DoNeutralEvolution given N, the number of vectors, and r as before, generates N vectors using NeutralEvolution.

```
function NeutralEvolution(c_0, v, m, r)
    if m = 0 then
        return concat(v, [c_0])                          ▷ concatenate c_0 to list of vectors v
    else
        c_1 ← c_0
        c_2 ← c_0
        for s ← 1 ... r do
            i ← randomInteger(F)                         ▷ random integer 0 ≤ i ≤ F − 1
            c_1[i] ← randomInteger(q)
        end for
        for s ← 1 ... r do
            i ← randomInteger(F)
            c_2[i] ← randomInteger(q)
        end for
        return concat(NeutralEvolution(c_1, v, m − 1, r), NeutralEvolution(c_2, v, m − 1, r))
    end if
end function

function DoNeutralEvolution(N, r)
    c_0 ← uniformRandomVector(q, F) ▷ vector with F integers uniformly random from 0 ... q − 1
    v ← NeutralEvolution(c_0, [], ⌈log_2 N⌉, r)
    return v[0..N − 1]                                   ▷ first N elements of list v
end function
```

Algorithm 2 Prototype evolution. Given k, the number of prototypes, return a list of N vectors (including the prototypes). F and q, the vector dimension and number of trait values respectively, are global constants.

```
function PrototypeEvolution(k, N)
    v ← []
    for i ← 0 ... k − 1 do
        c_i ← uniformRandomVector(q, F)▷ vector of F integers uniformly random from 0 ... q − 1
        v ← append(v, c_i)                               ▷ append vector c_i to list of vectors v
    end for
    for i ← 0 ... N − k − 1 do
        u_p ← v[randomInteger(k)]                        ▷ a randomly chosen prototype vector
        for s ← 0 ... ⌊F/2⌋ − 1 do
            j ← randomInteger(F)
            u_p[j] ← randomInteger(q)
        end for
        v ← append(v, u_p)
    end for
    return v
end function
```

Algorithm 3 Trivial ultrametric. Return N vectors that are perfectly ultrametric, each differing in exactly two traits from each of the others. F and q, the vector dimension and number of trait values respectively, are global constants.

```
function TrivialUltrametric(N)
    assert N ≤ (q − 1)F
    v ← []
    for i ← 1 . . . q − 1 do
        for j ← 0 . . . F − 1 do
            u_i ← [0]_F                          ▷ zero vector of dimension F
            u_i[j] ← i
            v ← append(v, u_i)
        end for
    end for
    return sample(v, N)          ▷ random sample of N items from v without replacement
end function
```

but then increases again). We can also observe in these figures that the ultrametric triangle fraction is actually smaller for larger values of F.

Fig. S6 shows the statistics for the "trivial ultrametric" scheme. (These figures have fewer values of F and q since the construction of the trivial ultrametric data requires that $N \leq (q-1)F$, so for $N = 125$ there is a limit on how small q or F can be). When $p = 0$, by construction the data is perfectly ultrametric, so the ultrametric triangle fraction is 1.0 (Fig. S6B), as is the cophenetic correlation coefficient (Fig. S6A). The mean and standard deviation of intervector distances are also almost zero when $p = 0$. As p increases, however, the ultrametric triangle fraction drops dramatically as soon as $p > 0$, but then increases, while the cophenetic correlation coefficient smoothly decreases as p increases.

Fig. S7 shows the degree of ultrametricity, as well as mean and standard deviation of intervector distance, for prototype evolution with different value of k. For the values of F and q that we use in the Axelrod model experiments ($F = 100, q = 10$), the cophenetic correlation coefficient remains very high for $1 < k \leq 50$, although standard deviation of intervector distances decreases much more rapidly. The ultrametric triangle fraction, however, remains very low at these parameters. Note that $k = 1$ is an outlier on these graphs, being the situation where there is only a single prototype, and all remaining $N - 1$ initial vectors are clustered around it.

5 Absorbing state of the extended Axelrod model

For all results described here and in the main article, the model was run to an absorbing state 50 times, from identical initial conditions, for each set of parameters, and error bars on graphs represent one standard deviation. Quantities (such as numbers of communities) on the y axis are normalised to lie between 0 and 1 by dividing by the number of agents N. We fix $\beta_p = 10$ and $\beta_s = 1$ in the extended Axelrod model for all results described here and in the main article. Network statistics are computed using the igraph package [7] and graphs produced using the ggplot2 [8] and lattice [9] packages in R [10].

Fig. S8A shows the number of communities in the social network at the absorbing state against the value of the bounded confidence threshold θ with our modified model, and Fig. S8B shows the results for the GSS data. Figures S8C and S8D show the results from the simulated Eurobarometer and GSS data. This data has the same covariance structure as the real data, but, as is evident from Table 1 and Table 2 in the main article, a much lower degree of ultrametricity. The clear distinction between the three initial cultures (real/simulated, permuted, random) at which the value of θ where the number of communities changes from minimum to maximum at the absorbing state is no longer evident for simulated data, with the exception of the simulated Eurobarometer data, in which this value of θ is actually larger, rather than smaller, for the random data compared to simulated and permuted data.

Fig. S9 and Fig. S10 show the number of communities at the absorbing state with different values of θ for different schemes of generating initial culture vectors. We can see that, when $\theta = 0$, only sufficiently small values of F and large values of q result in a multicultural absorbing state. In addition, for all three schemes we can observe that there is a value of θ at which the absorbing state changes from a single

to multiple communities, and this value of θ increases with decreasing initial perturbation probability (the curve for $p = 0$ is shifted right with respect to the curve for $p = 0.2$ and so on). This value of θ corresponds to the value at which the initial culture graph becomes completely disconnected into N cultures each of which has similarity less than θ to all others, and therefore no successful interactions can take place between them in the Axelrod model with bounded confidence threshold θ.

The bounded confidence threshold θ determines the number of connected components in the culture graph at the initial conditions, and influences the number of surviving cultures (and communities in the social network) at the absorbing state. But when $\theta = 0$ multicultural absorbing states are only reached when $F \leq 10$ — all others are monocultural (for q up to 100). The larger the value of F (the higher the culture vector dimension), the less likely it is that two agents have no trait in common (for fixed q). Hence agents are increasingly likely to interact and merge into a single culture with higher values of F. For higher values of q, however, the more likely it is that two agents have no trait in common (for fixed F) and therefore cannot interact. So higher values of q create greater diversity when using uniform random initial culture vectors. When using more structured initial conditions, however, such as the empirical and synthetic data (when $p < 1$) used here, initial diversity can vary (for a fixed value of q) and can be measured by the standard deviation (or variance) of the intervector distances.

Fig. S11, Fig. S12, and Fig. S13 show the number of cultures at the absorbing state as a function of both the number of connected components in the initial culture graph and the cophenetic correlation coefficient of the initial culture vectors. Fig. 5 in the main text, as well as Fig. S9B, Fig. S9D, Fig. S9F, and Fig. S10B show the number of cultures at the absorbing state as a function of the number of connected components in the initial culture graph, using a different curve for each value of the initial perturbation probability p. This has the drawback that we are using p as a proxy for the degree of ultrametricity (defined by the cophenetic correlation coefficient) of the initial culture vectors, based on the behaviour of the cophenetic correlation coefficient as a function of p, rather than using the cophenetic correlation coefficient directly. Figures S11, S12, and S13 allow us to observe the number of cultures at the absorbing state directly as a function the cophenetic correlation coefficient of the initial culture vectors, as well as the number of connected components in the initial culture graph. In these figures, the data points are coloured according to the initial perturbation probability p, from dark blue for no perturbation ($p = 0$) up to very light blue for uniform random vectors ($p = 1$). The data points represent means of the 50 runs from the same initial conditions, but error bars have been omitted for clarity in these plots. We can observe, as already shown in Fig. 3 in the main text, as well as Figures S4, S5, and S6, that in all three schemes the cophenetic correlation coefficient decreases as p increases. This is shown in Figures S11, S12, and S13 by the data points becoming increasingly dark as the cophenetic correlation coefficient (y axis) increases.

Fig. S14 allows us to see more detail for the case of $q = 10$ and $F = 100$ by fitting a surface to the data points. We can see that, for neutral evolution (Fig. S14A and Fig. S14B) and the trivial ultrametric scheme (Fig. S14D), the surface is approximately a plane (equivalent to all the lines being on the diagonal in Fig. 5 in the main text as well as Fig. S9B, Fig. S9D, Fig. S9F, and Fig. S10B), showing that increasing the cophenetic correlation coefficient of the initial culture vectors does not result in a greater number of cultures at the absorbing state for a given number of initial connected components in the culture graph. For the prototype evolution scheme (Fig S14C), however, the surface is not a plane, but rather curves upwards as the cophenetic correlation coefficient increases, showing that a higher cophenetic correlation coefficient does result in a larger number of cultures at the absorbing state for a given number of connected components in the initial culture graph.

Fig. S14B also shows why we cannot entirely rely on the neutral evolution results of Fig. 5 in the main text (to which it corresponds for the neutral evolution case) alone to support our argument that ultrametricity alone is not sufficient for a greater number of cultures at the absorbing state. As can be deduced from Fig. 3 in the main text (as well as Fig. S4 and Table S2), even when $p = 0$, neutral evolution (for $q = 10, F = 100$) does not result in the maximum value of the cophenetic correlation of 1.0, but only approximately 0.72, whereas prototype evolution attains a maximum cophenetic correlation coefficient of approximately 0.97. Therefore, it is possible that prototype evolution results in a greater number of cultures not because of the greater variance of initial intervector distances, but because it allows for greater ultrametricity than neutral evolution.

Neutral evolution with $r = 1$ shows the problem to a much smaller degree (Fig. S14A), with a maximum value of the cophenetic correlation coefficient of approximately 0.91. In addition, neutral evolution ($r = \lfloor \frac{F}{2} \rfloor$) for the same value of F but larger values of q, also shows the problem to a smaller

degree, and not at all for $q = 100$ (Fig. S11). Relying on different values of r and q, however, is not robust evidence of our claim that ultrametricity (as we have measured it) is not sufficient for increased cultural diversity.

Therefore we turn to the "trivial ultrametric" scheme, which covers the full range of cophenetic correlation coefficient from 0 to 1 as p is varied. Fig. S14D shows the surface is close to a plane, with no increase in the number of cultures in the final state as the cophenetic correlation coefficient of the initial vectors increases. Similar results are obtained for different values of q and F (Fig. S13). This unambiguously shows that increasing ultrametricity of the initial culture vectors (while variance of intervector distance remains small and fixed) is not enough to lead to a greater number of cultures in the final state of the model.

Fig. S15 shows the number of cultures at the absorbing state against the number of initial connected cultural components for different values of k in the prototype evolution scheme for generating initial culture vectors. The results are very similar to the value $k = 3$ described in the main article (and Fig. S9F), even the case of $k = 1$, which has a much lower degree of ultrametricity as measured by cophenetic correlation coefficient than $2 \leq k \leq 10$ (Fig. S7).

Fig. S16 shows the number of cultures at the absorbing state against the number of initial connected cultural components for different values of r in the neutral evolution scheme for generating initial culture vectors. The results are similar to the results for $r = \lfloor \frac{F}{2} \rfloor$ described in the main article (and Fig. S9D), and $r = 1$ (Fig. S9B).

6 The number of cultures at the absorbing state is not bounded above by the number of connected components in the initial culture graph

In [1, Supporting Information] it is claimed that the largest possible value of the number of final different cultural domains N_D at the absorbing state is the number of initial connected cultural components N_C for a given value of the bounded confidence threshold θ. In this case, in the graphs of number of cultures at the absorbing state N_D versus initial number of connected culture components N_C (such as, for example, Fig. 1 in the main text of this article, or Fig. S5 in [1, Supporting Information]), a curve on the diagonal represents this maximum level of long-term cultural heterogeneity, so all curves must be on this diagonal or below it, in the lower triangle. It is claimed in [1, Supporting Information] that this maximum value of N_D along the diagonal arises when, in ultrametric data, the initial connected components are cliques and so the agents in the clique will all eventually converge to the same culture vector, as each can interact with any other in the clique. This would give rise to the situation that we have the maximum value of $N_D = N_C$ at the absorbing state since agents cannot interact if they are in different components of the culture graph (of which there are N_C).

However, as shown in the main text of this article, even in the simple Axelrod model as used in [1], situations occur in which curves are in the upper triangle (for ultrametric data generated by the prototype evolution scheme). In the extended model, this situation almost always occurs (the curves are in the upper, not lower, triangle of the graph). Here we show that the argument from [1, Supporting Information] just outlined is not correct, and we can have $N_D > N_C$ at the absorbing state because even agents in a clique in the initial culture graph need not necessarily converge to the same culture at the absorbing state.

We use a small example, to demonstrate how this can occur. Let $q = 2$, $F = 3$, and $N = 3$. With such a small example we can follow the steps of the simple Axelrod model by hand.

We will use the Hamming similarity (without normalisation, for simplicity; this is just Equation 1 without the division by F) for cultural similarity:

$$\kappa(a, b) = \sum_{i=1}^{F} \delta_{v_i^{(a)}, v_i^{(b)}} \tag{5}$$

For convenience we will also define the corresponding (unnormalised) Hamming distance:

$$d_\kappa(a, b) = F - \kappa(a, b) \tag{6}$$

Let the bounded confidence threshold $\theta = 2$, so agents must have similarity $\geq \theta = 2$ to interact, and the culture graph is defined by edges between agents with similarity ≥ 2.

Suppose the initial conditions (with the three agents labelled `A`, `B`, and `C`) are:

```
A: 1 1 0
B: 1 1 1
C: 1 0 1
```

There are three distinct cultures initially (it is always the case that the number of distinct cultures initially is N in the empirical data, and also in random data with sufficiently large values of F it becomes very unlikely that there are any duplicated culture vectors for our relatively small values of N).

Now $\kappa(A, B) = \kappa(B, C) = 2 \geq \theta$ but $\kappa(A, C) = 1 < \theta$, so the graph looks like `[ A-B-C ]` and has a single connected component.

Now in a single step of the simple Axelrod model, agent `B` could interact with `C` (since $\kappa(B, C) = 2 \geq \theta$) and the only trait it has different from `C` would become equal to `C`'s:

```
A: 1 1 0
B: 1 0 1
C: 1 0 1
```

Now `B` and `C` are identical and $\kappa(A, B) = \kappa(A, C) = 1 < \theta$ so the culture graph looks like `[ A B-C ]` so it has two connected components, where there was only one originally. Further, this model is now converged since `B` and `C` are identical and `A` has similarity $< \theta$ with all other agents. And, at convergence, there are two distinct cultures ($N_D = 2$).

In summary, in this example with this value of θ, the original number of connected components is one and the final number of cultures is two — that is, N_D at convergence is greater than N_C in the initial conditions. Hence this shows that it is possible that $N_D > N_C$.

Note however in [1, Supporting Information] the statement about the "largest possible" value of N_D being N_C is made in the context of ultrametric data. Our example above is not ultrametric since we have $d_\kappa(a, b) = 1$, $d_\kappa(b, c) = 1$ and $d_\kappa(a, c) = 2$ so the ultrametric inequality $d_\kappa(a, c) \leq \max\{d_\kappa(a, b), d_\kappa(b, c)\}$ does not hold.

It is stated in [1, Supporting Information], that if the data were ultrametric, the connected component would actually be a completely connected clique and would therefore converge to a single culture eventually. However we now show that the situation just shown above ($N_D > N_C$) can also occur with an ultrametric initial condition.

In this example we will let $\theta = 1$, and it will take two steps of the simple Axelrod model to converge.

Let the initial culture vectors be:

```
A: 0 0 0
B: 0 1 1
C: 1 1 0
```

All vectors have similarity 1 (distance 2) so at $\theta = 1$ we have a single completely connected component, that is, a clique. Note that this set of vectors is ultrametric (forms an equilateral triangle as all distances are equal). In a single step of the Axelrod model ($\theta = 1$) agent `B` can interact with agent `C`, so we now have:

```
A: 0 0 0
B: 0 1 1
C: 1 1 1
```

And now $\kappa(A, C) = 0$ so agents `A` and `C` cannot interact, but there is still a single connected cultural component (the culture graph at $\theta = 1$ is `[ A-B-C ]` since $\kappa(B, C) = 2 \geq \theta$ and $\kappa(A, B) = 1 \geq \theta$ but $\kappa(A, C) = 0$). It is no longer a clique however.

And in a second step of the Axelrod model, now `B` can interact with `C` and we get:

```
A: 0 0 0
B: 1 1 1
C: 1 1 1
```

description	intervector distance		cophenetic	ultrametric
	mean	sd	corr. coeff.	tri. frac.
Real	0.557	0.117	0.711	0.048
Permuted	0.557	0.038	0.413	0.083
Random	0.803	0.032	0.027	0.089
Simulated	0.909	0.027	0.051	0.112
Simulated permuted	0.909	0.024	0.031	0.121

Table S1: Statistics of inter-vector distances, cophenetic correlation coefficients, and ultrametric triangle fractions for Eurobarometer survey data ($N = 600$), with the split ballot question handled by converting the inapplicable version to the "most neutral" response, so that the original $F = 162$ opinions remain.

Now there are two components in the culture graph `[ A B-C ]` at $\theta = 1$ and the model is converged as `B` and `C` are identical and $\kappa(A, C) = \kappa(A, B) = 0$. And there are two distinct cultures ($N_D = 2$), even though there was only one component in the culture graph initially ($N_C = 1$).

So again we have $N_D > N_C$, even when the initial culture vectors were ultrametric.

References

[1] Valori, L., Picciolo, F., Allansdottir, A. & Garlaschelli, D. Reconciling long-term cultural diversity and short-term collective social behavior. *Proc Natl Acad Sci USA* **109**, 1068–1073 (2012).

[2] Pfau, J., Kirley, M. & Kashima, Y. The co-evolution of cultures, social network communities, and agent locations in an extension of Axelrod's model of cultural dissemination. *Physica A* **392**, 381–391 (2013).

[3] Rammal, R., Angles d'Auriac, J. & Doucot, B. On the degree of ultrametricity. *J Phys Lett-Paris* **46**, 945–952 (1985).

[4] Rammal, R., Toulouse, G. & Virasoro, M. A. Ultrametricity for physicists. *Rev Mod Phys* **58**, 765–788 (1986).

[5] Smith, T. W., Marsden, P., Hout, M. & Jibum, K. *General Social Surveys, 1972-2012*. National Opinion Research Center, Chicago. URL `http://publicdata.norc.org/GSS/DOCUMENTS/OTHR/1993_spss.zip`. Accessed 13 May 2013 (2013).

[6] Murtagh, F. The remarkable simplicity of very high dimensional data: application of model-based clustering. *J Classif* **26**, 249–277 (2009).

[7] Csárdi, G. & Nepusz, T. The igraph software package for complex network research. *InterJournal* **Complex Systems**, 1695 (2006).

[8] Wickham, H. *ggplot2: elegant graphics for data analysis* (Springer New York, 2009).

[9] Sarkar, D. *Lattice: Multivariate Data Visualization with R* (Springer, New York, 2008). ISBN 978-0-387-75968-5.

[10] R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. URL `http://www.R-project.org` (2013).

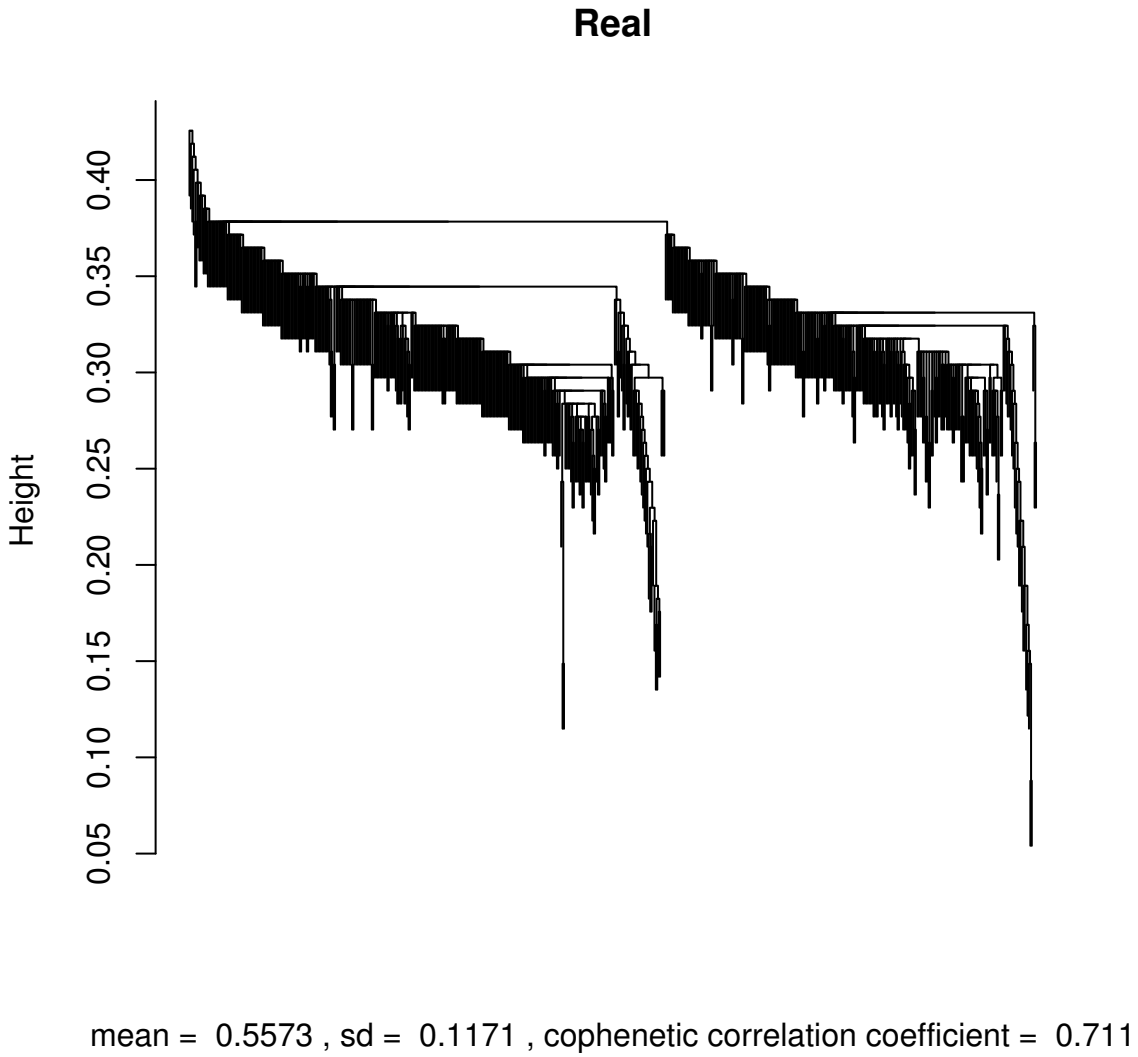


Figure S1: Single-linkage clustering dendrogram for Eurobarometer survey data ($N = 600$), with the split ballot question handled by converting the inapplicable version to the "most neutral" response, so that the original $F = 162$ opinions remain.

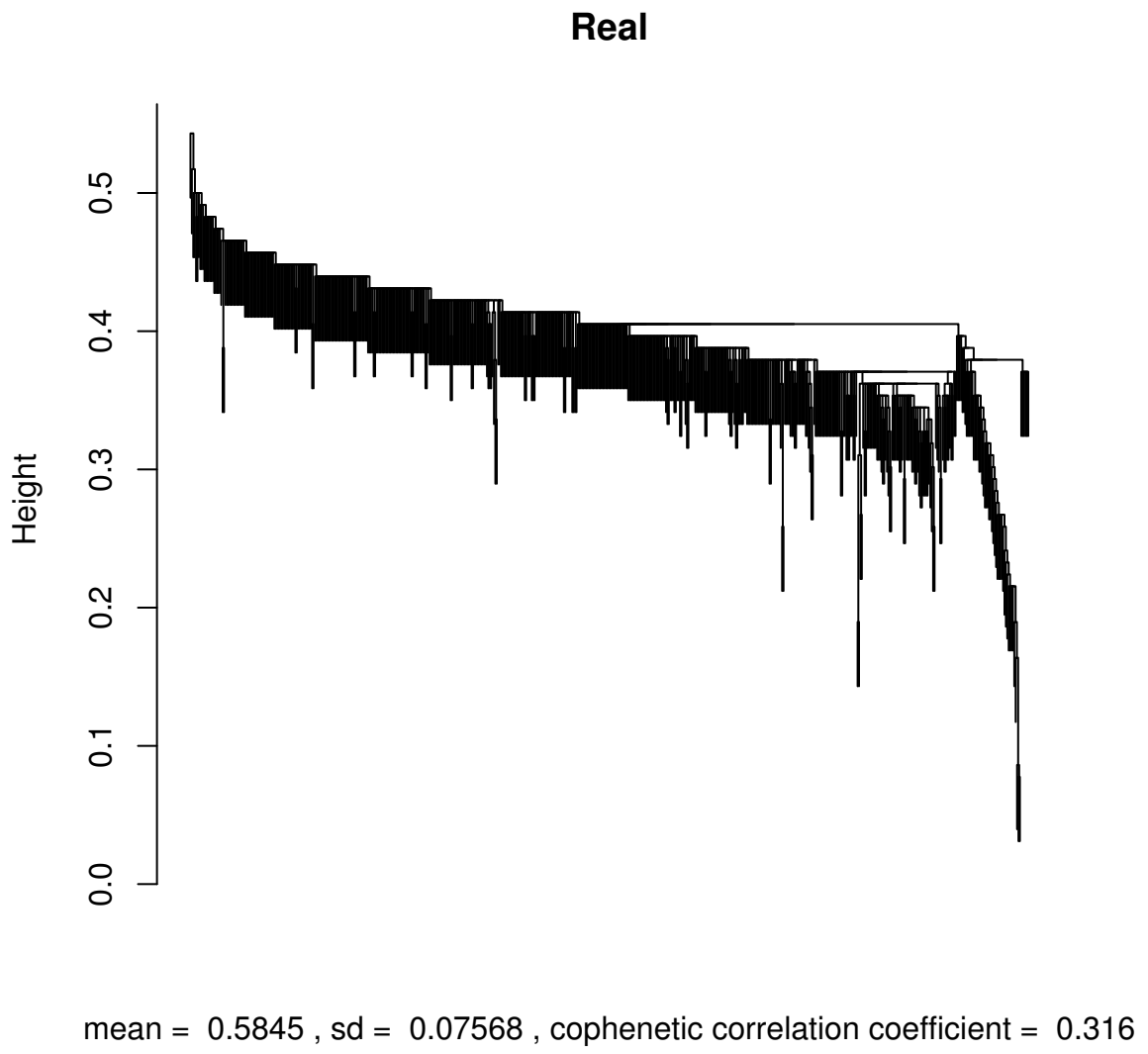


Figure S2: Single-linkage clustering dendrogram for Eurobarometer survey data ($N = 600$), with the split ballot question handled by "merging" the split ballots so that $F = 116$.

description	intervector distance mean	sd	cophenetic corr. coeff.	ultrametric tri. frac.
Neutral evolution $r = \lfloor \frac{F}{2} \rfloor$	0.892	0.045	0.724	0.134
Neutral evolution $r = \lfloor \frac{F}{2} \rfloor$ with perturb prob 0.4	0.897	0.032	0.211	0.139
Neutral evolution $r = \lfloor \frac{F}{2} \rfloor$ with perturb prob 0.6	0.899	0.031	0.112	0.135
Neutral evolution $r = 1$	0.107	0.026	0.908	0.376
Neutral evolution $r = 1$ with perturb prob 0.4	0.617	0.049	0.661	0.102
Neutral evolution $r = 1$ with perturb prob 0.6	0.768	0.043	0.482	0.111
Prototype evolution $k = 3$	0.781	0.160	0.972	0.124
Prototype evolution $k = 3$ with perturb prob 0.4	0.860	0.068	0.765	0.110
Prototype evolution $k = 3$ with perturb prob 0.6	0.881	0.042	0.451	0.118
Trivial ultrametric	0.020	0.001	1.000	1.000
Trivial ultrametric with perturb prob 0.4	0.574	0.047	0.652	0.109
Trivial ultrametric with perturb prob 0.6	0.760	0.041	0.469	0.114
Random	0.901	0.030	0.129	0.141

Table S2: Statistics of inter-vector distances, cophenetic correlation coefficients, and ultrametric triangle fractions for evolved vectors ($N = 125, q = 10, F = 100$).

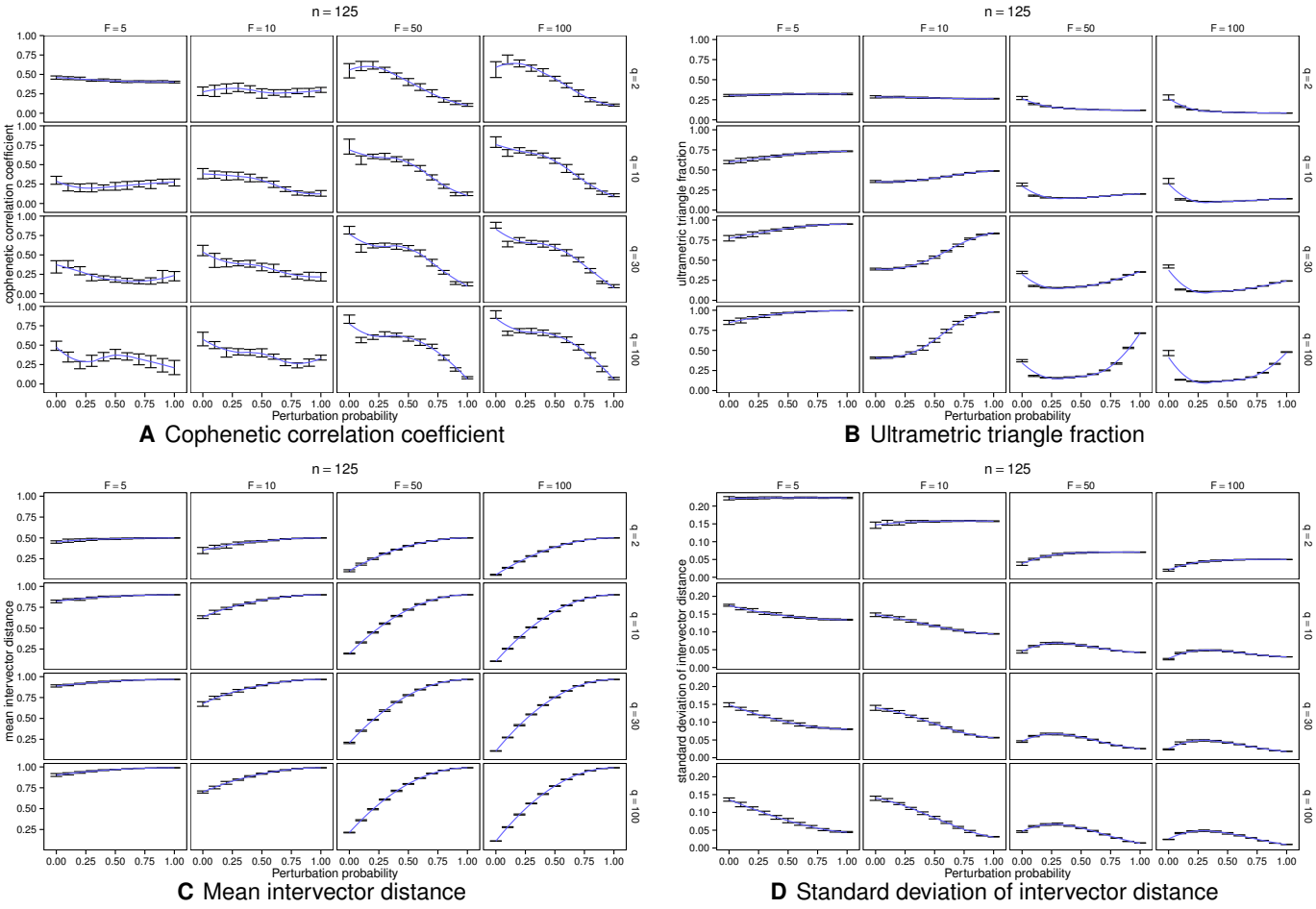


Figure S3: Ultrametricity generated by perturbing neutral evolution ($r = 1$) vectors ($N = 125$). Error bars represent one standard deviation with 15 trials for each point. Curves are locally fitted polynomials fitted by the `loess` function in `R`.

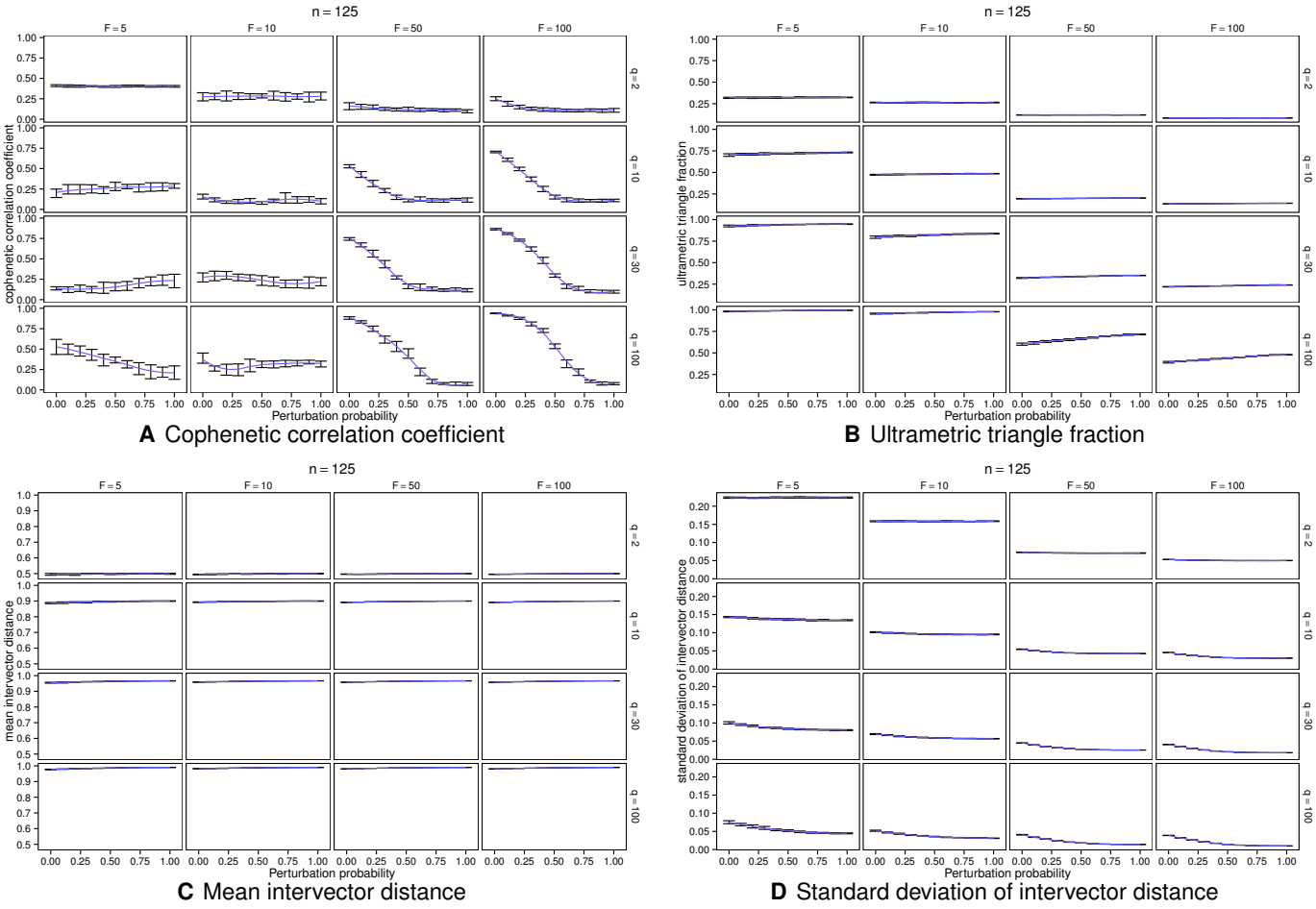


Figure S4: Ultrametricity generated by perturbing neutral evolution ($r = \lfloor \frac{F}{2} \rfloor$) vectors ($N = 125$). Error bars represent one standard deviation with 15 trials for each point. Curves are locally fitted polynomials fitted by the `loess` function in `R`.

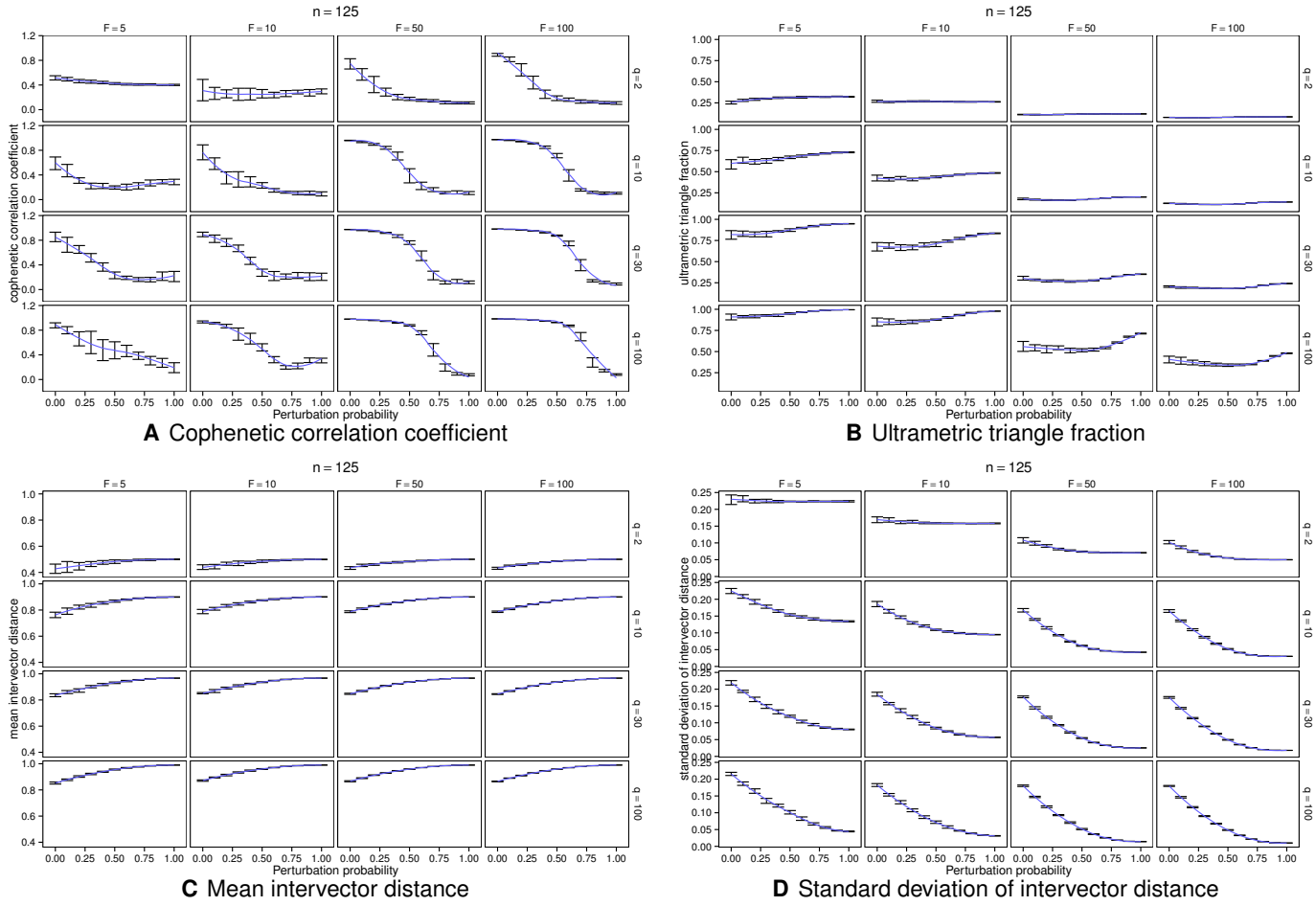


Figure S5: Ultrametricity generated by perturbing prototype evolution ($k = 3$) vectors ($N = 125$). Error bars represent one standard deviation with 15 trials for each point. Curves are locally fitted polynomials fitted by the `loess` function in `R`.

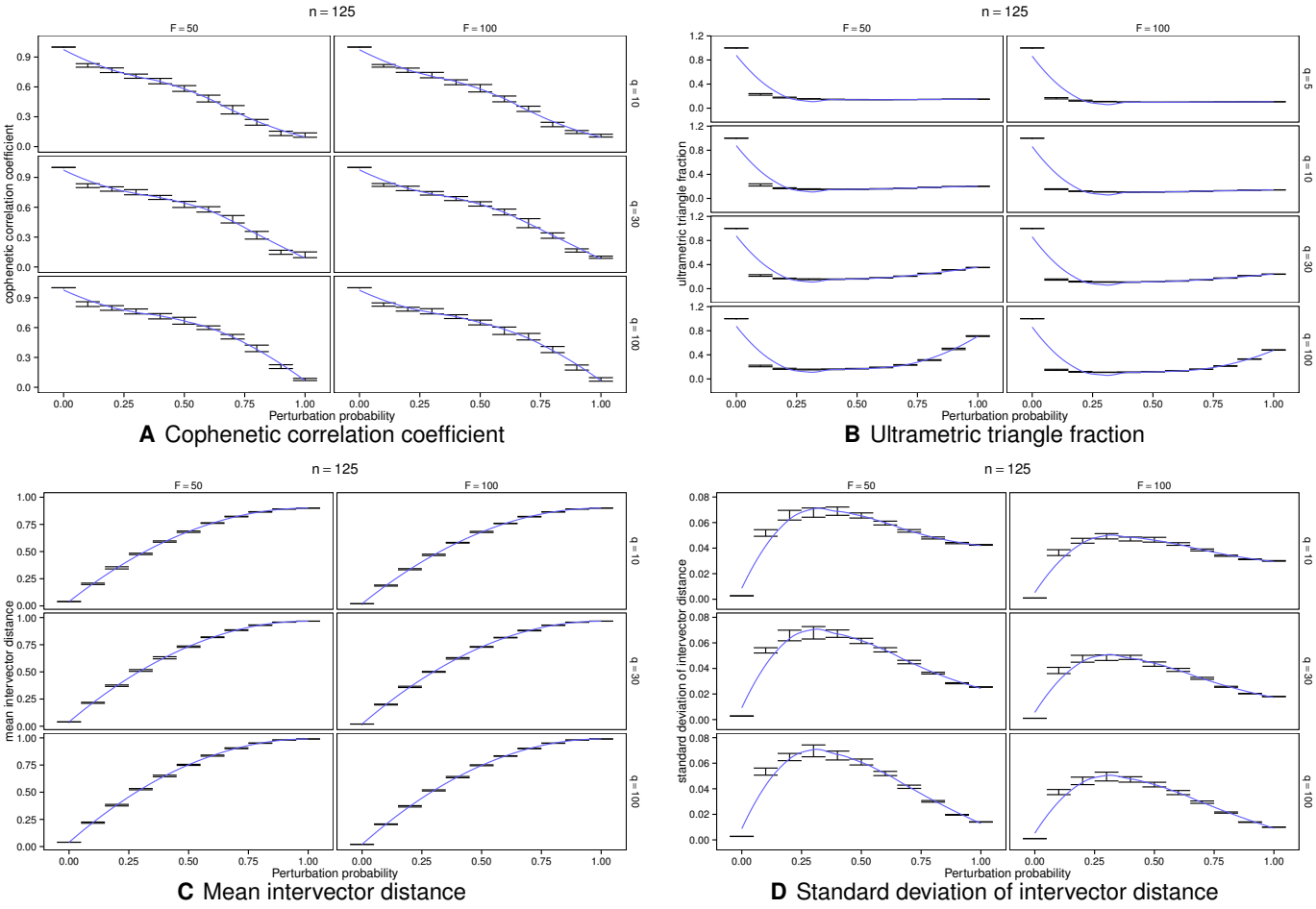


Figure S6: Ultrametricity generated by perturbing the trivial ultrametric vectors ($N = 125$). Error bars represent one standard deviation with 15 trials for each point. Curves are locally fitted polynomials fitted by the `loess` function in `R`.

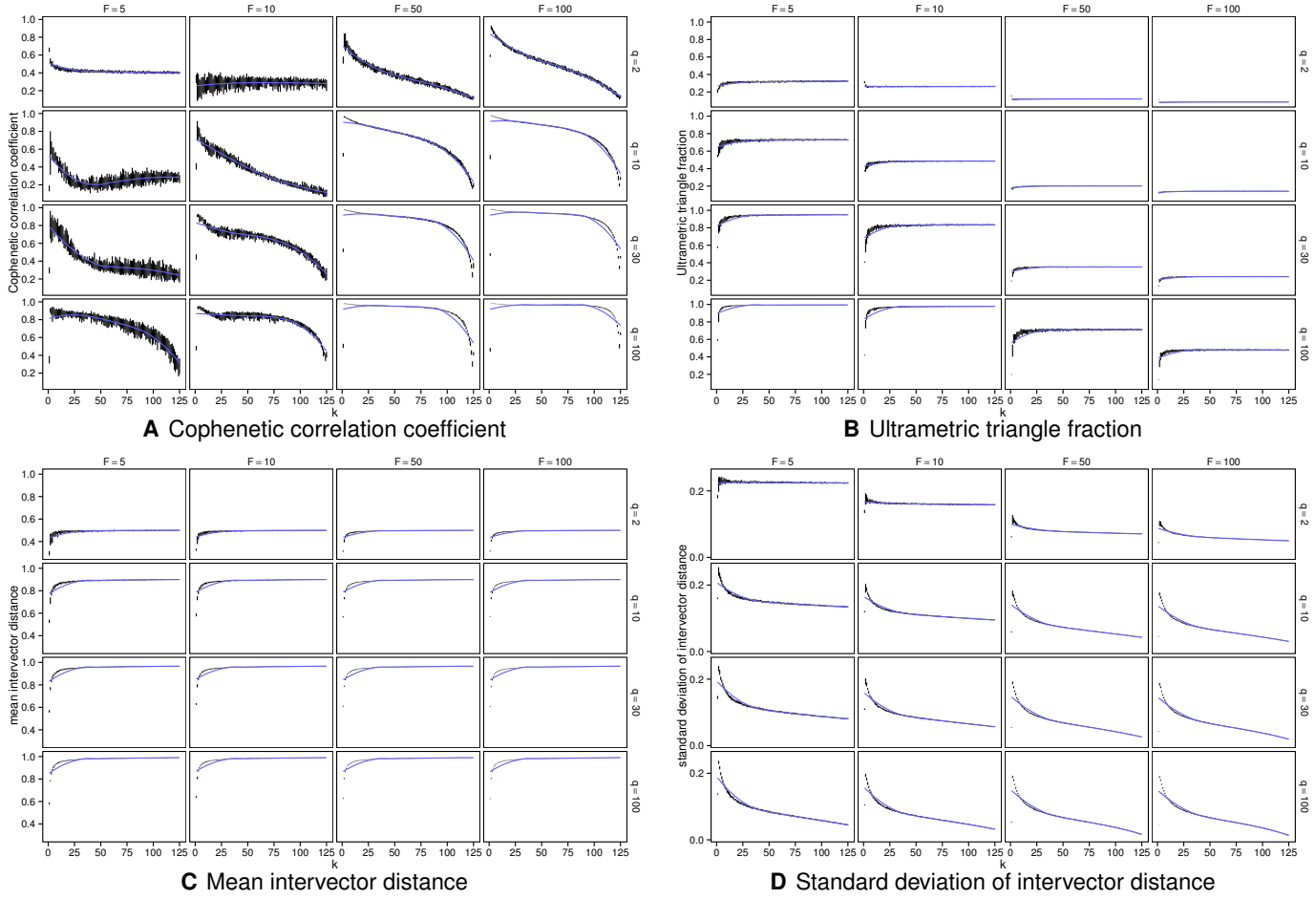


Figure S7: Ultrametricity generated by prototype evolution with different values of k, the number of prototypes ($N = 125$). Error bars represent one standard deviation with 15 trials for each point. Curves are locally fitted polynomials fitted by the `loess` function in R.

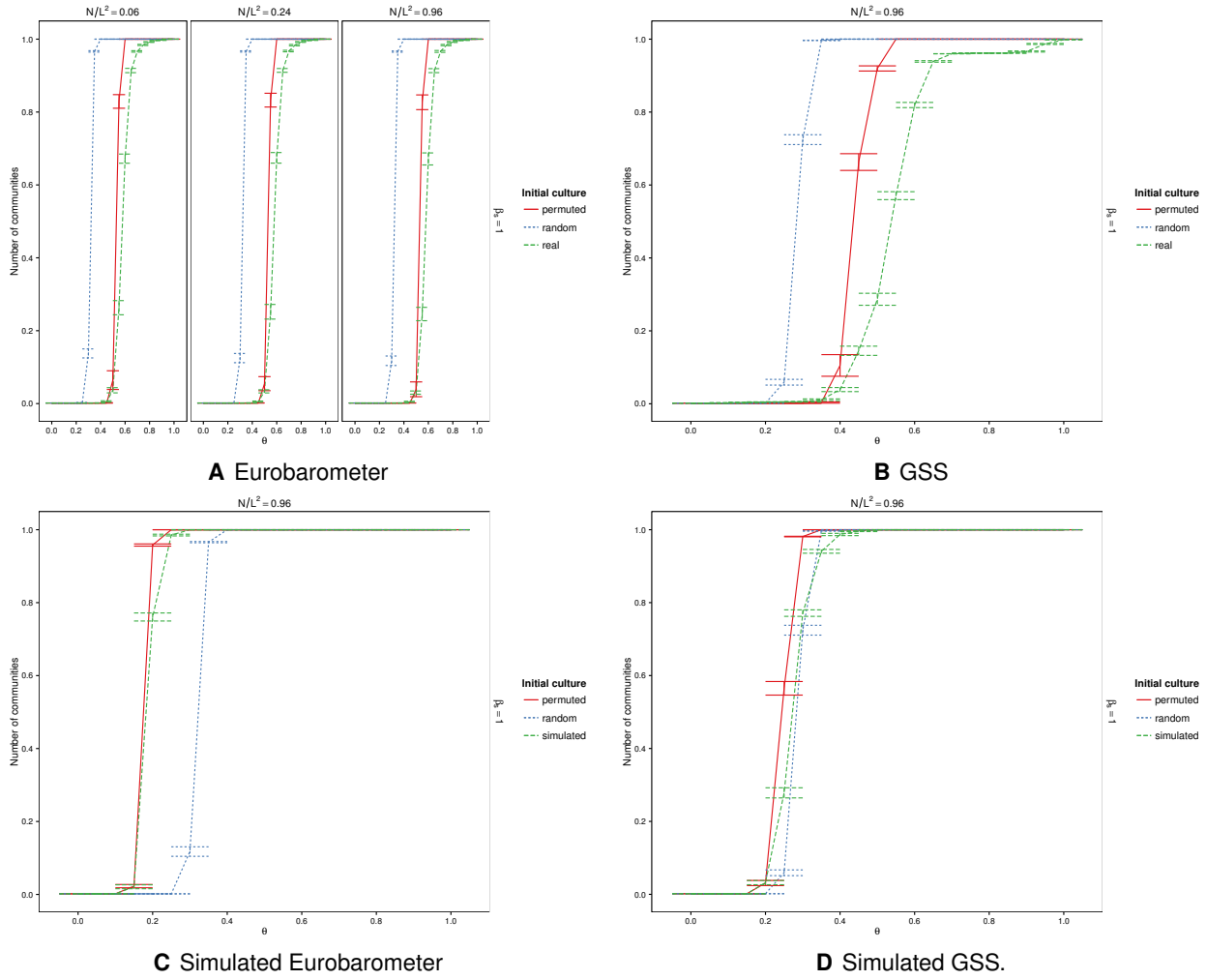
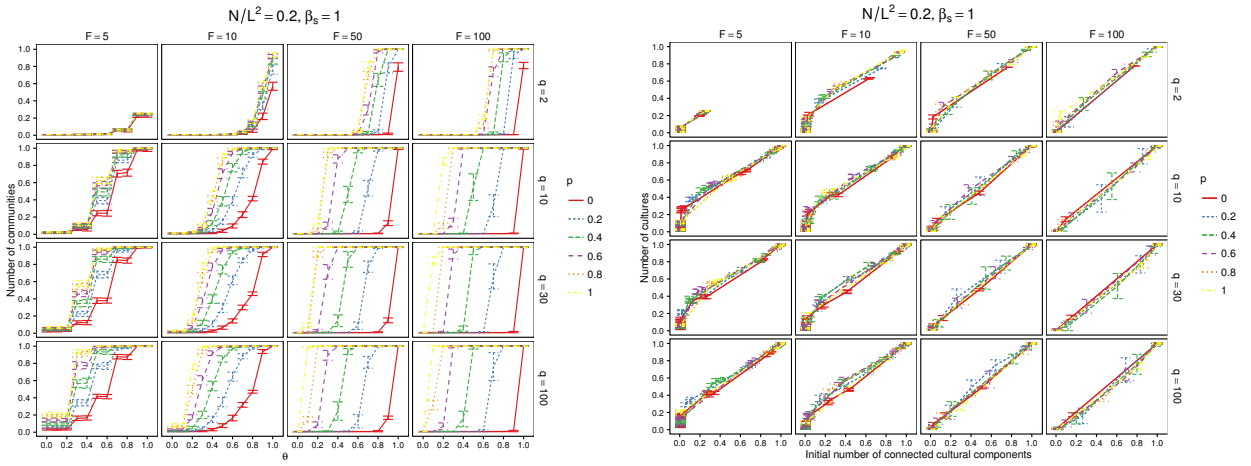
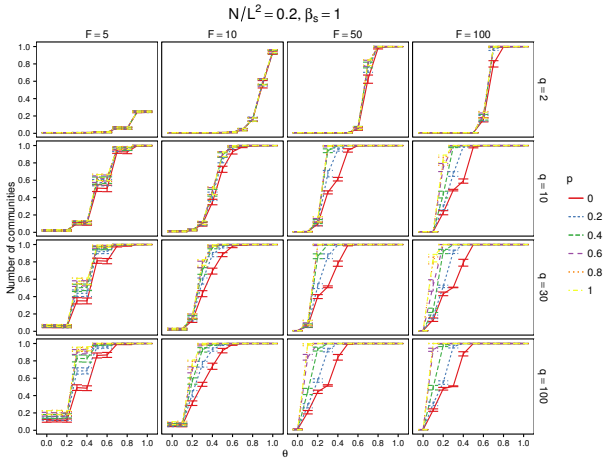


Figure S8: Number of communities at the absorbing state for Eurobarometer and GSS data, both real and simulated using random values with the same covariance structure as the real data.

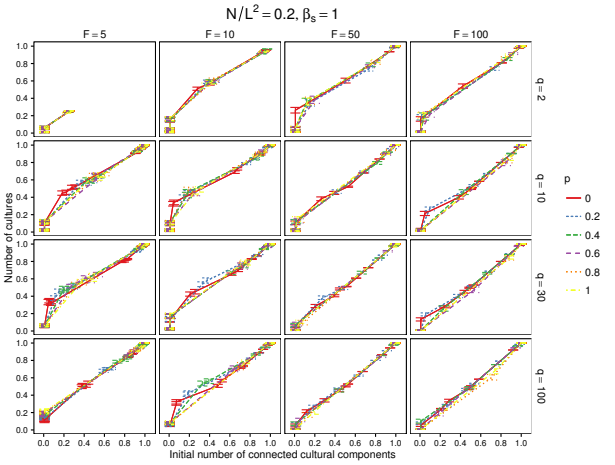


A Neutral evolution ($r = 1$)

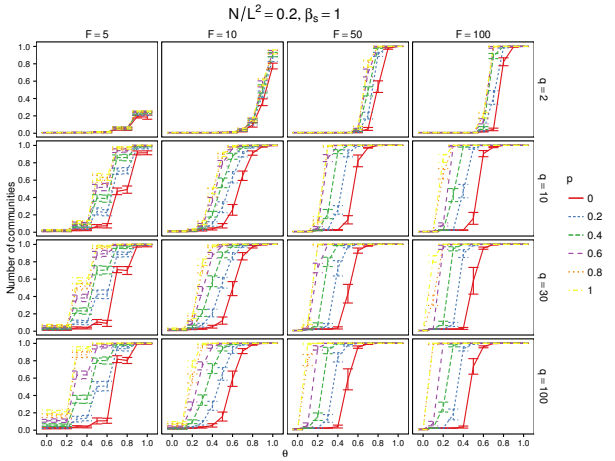
B Neutral evolution ($r = 1$)



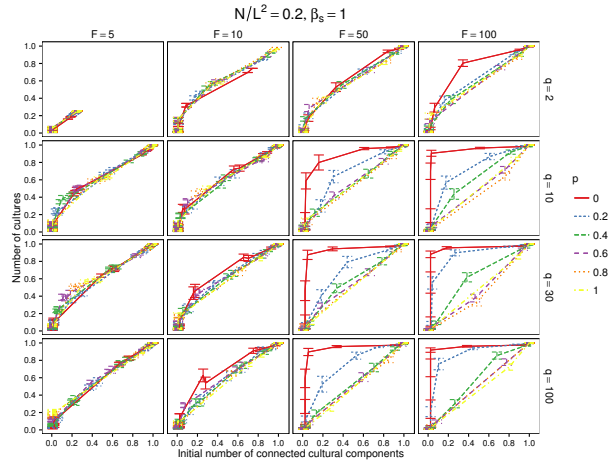
C Neutral evolution ($r = \lfloor \frac{F}{2} \rfloor$)



D Neutral evolution ($r = \lfloor \frac{F}{2} \rfloor$)

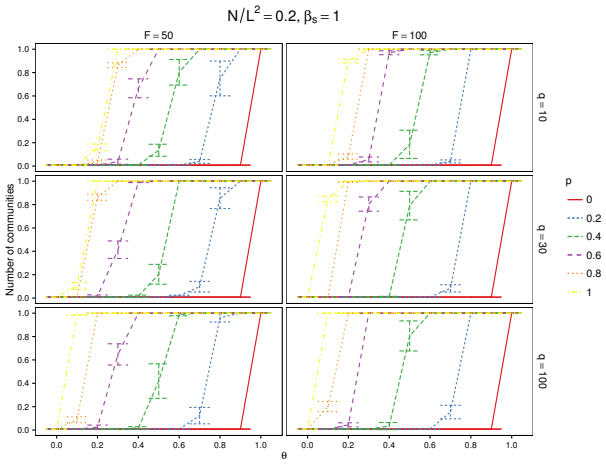


E Prototype evolution ($k = 3$)

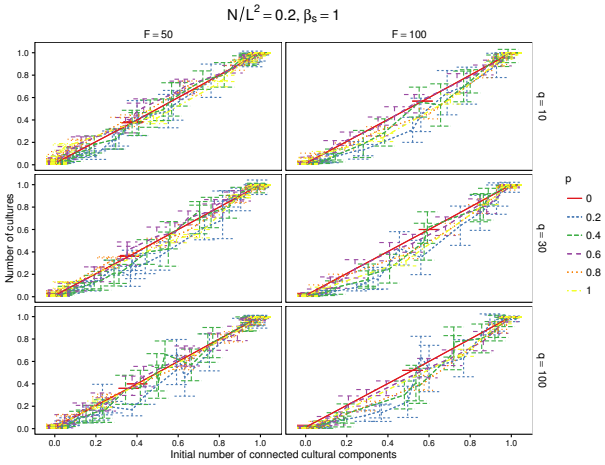


F Prototype evolution ($k = 3$)

Figure S9: Number of communities at the absorbing state plotted against the value of the bounded confidence threshold θ (left column), and number of cultures at the absorbing state plotted against number of connected components in the culture graph of initial conditions (right column) for the "evolved" culture vectors for various initial perturbation probabilities p. In the right column of subfigures, the value of θ is varied to obtain different numbers of initial connected components in the culture graphs along the x axis and the corresponding numbers of cultures at the absorbing state is shown on the y axis. Each row of subfigures uses a different a different scheme for creating the initial opinion vectors.



A Number of communities at equilibrium plotted against the value of the bounded confidence threshold θ.



B Number of cultures at equilibrium plotted against number of connected components in the culture graph of initial conditions.

Figure S10: Number of communities at the absorbing state plotted against the value of the bounded confidence threshold θ, and number of cultures at the absorbing state plotted against number of connected components in the culture graph of initial conditions for the "trivial ultrametric" culture vectors for various initial perturbation probabilities p.

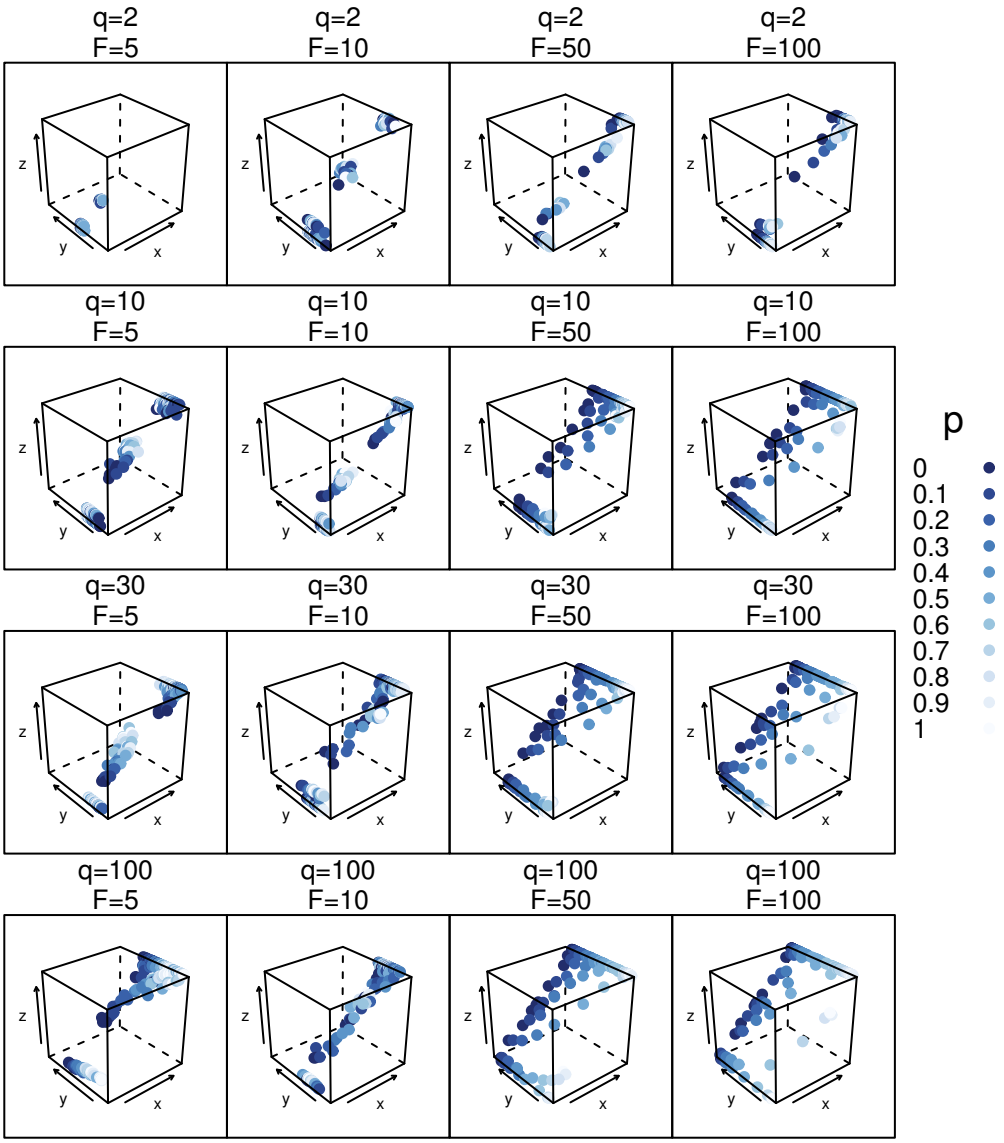


Figure S11: Number of cultures at the absorbing state (z axis) plotted against the number of connected components in the culture graph of the initial culture vectors (x axis) and the cophenetic correlation coefficient of the initial culture vectors (y axis), for initial conditions generated by neutral evolution ($r = \lfloor \frac{F}{2} \rfloor$). The data points are coloured according to the perturbation probability p.

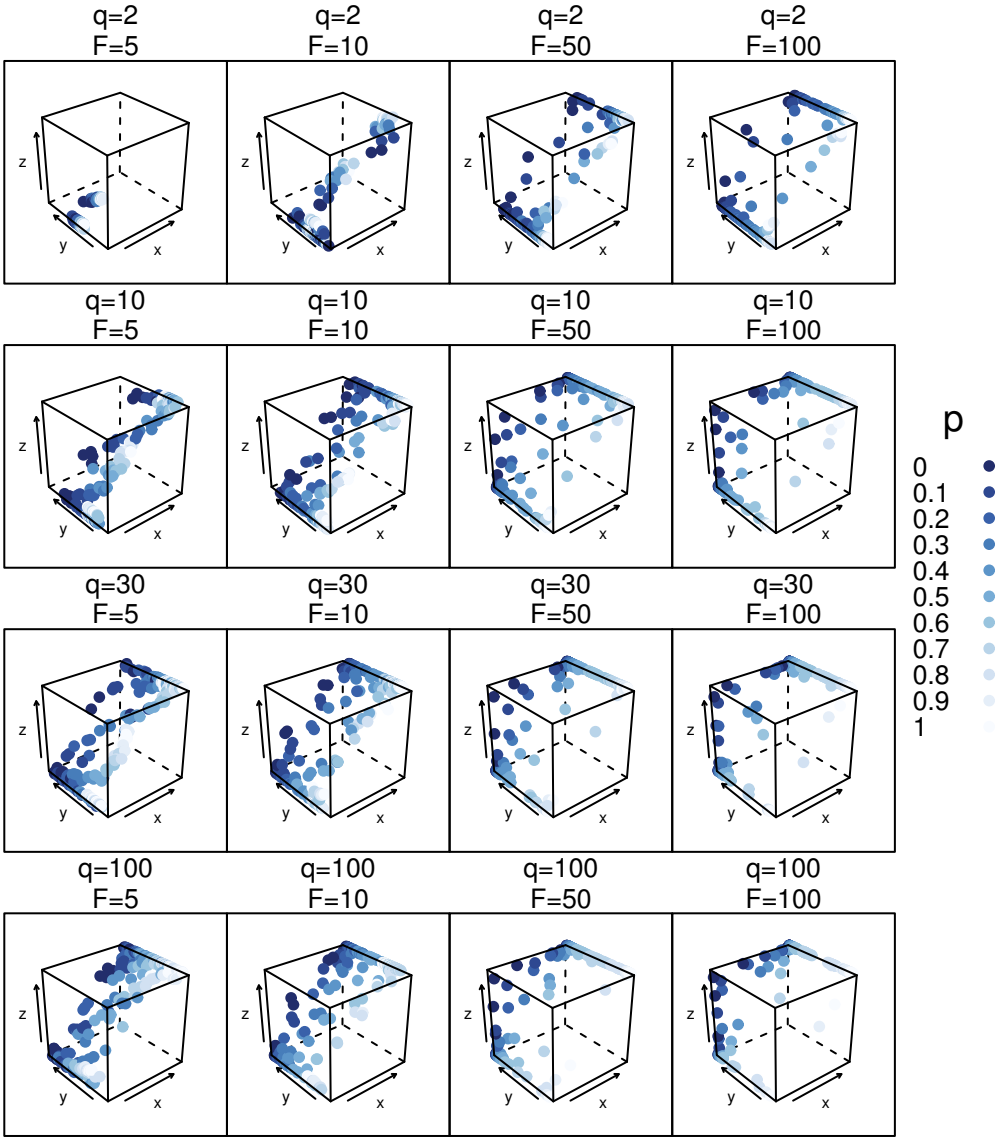


Figure S12: Number of cultures at the absorbing state (z axis) plotted against the number of connected components in the culture graph of the initial culture vectors (x axis) and the cophenetic correlation coefficient of the initial culture vectors (y axis), for initial conditions generated by prototype evolution ($k = 3$). The data points are coloured according to the perturbation probability p.

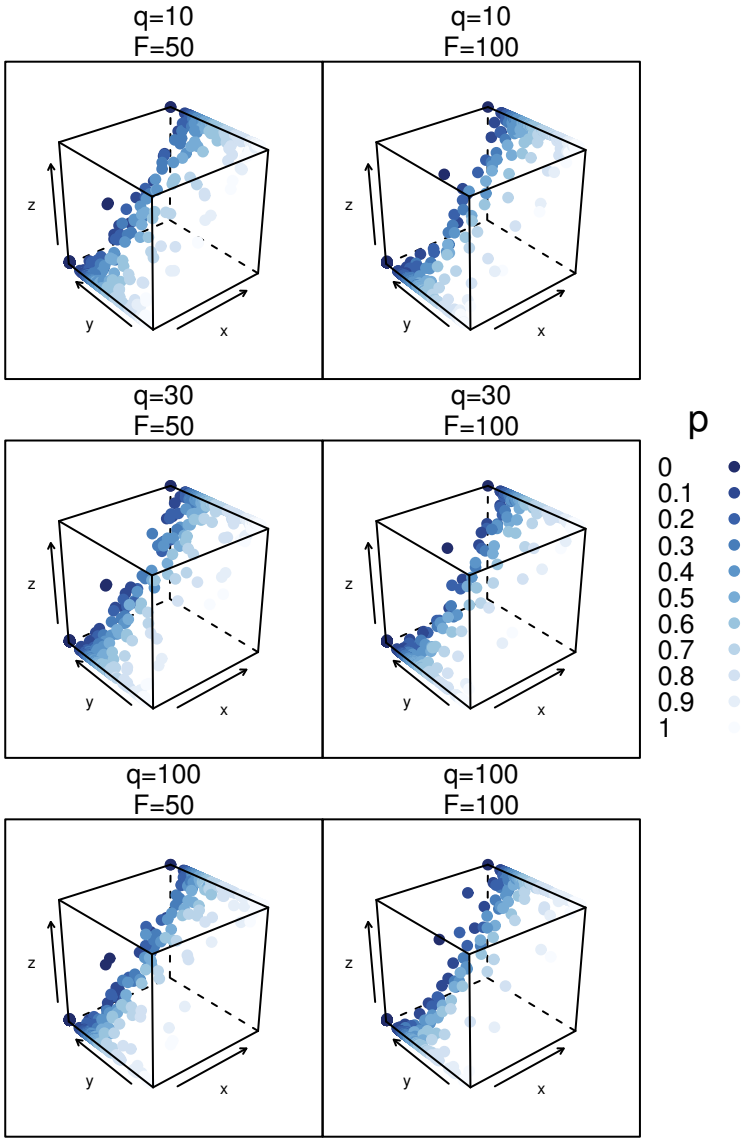
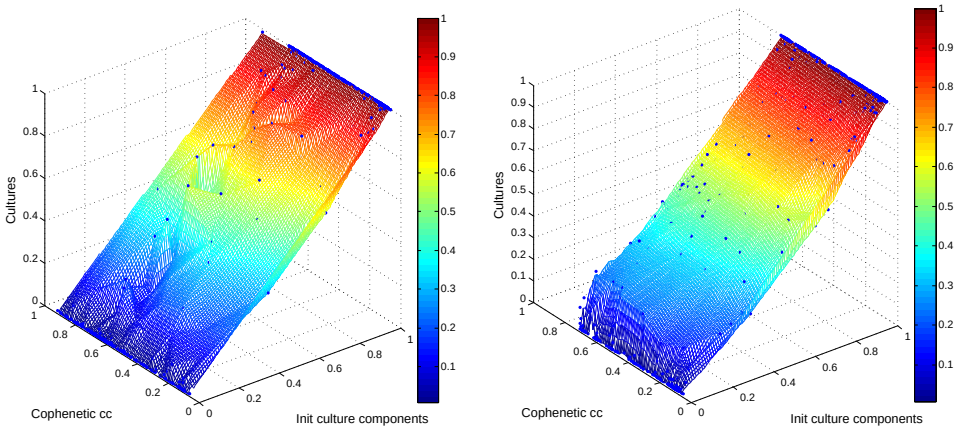
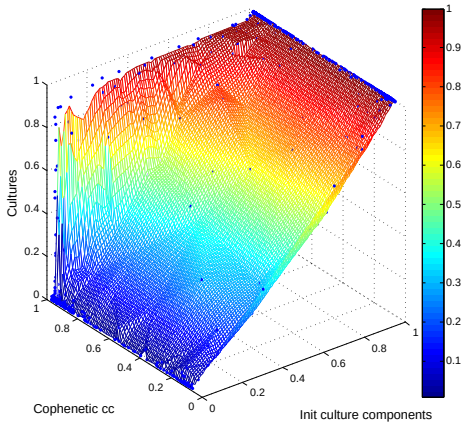


Figure S13: Number of cultures at the absorbing state (z axis) plotted against the number of connected components in the culture graph of the initial culture vectors (x axis) and the cophenetic correlation coefficient of the initial culture vectors (y axis), for initial conditions generated by the "trivial ultrametric" scheme. The data points are coloured according to the perturbation probability p.

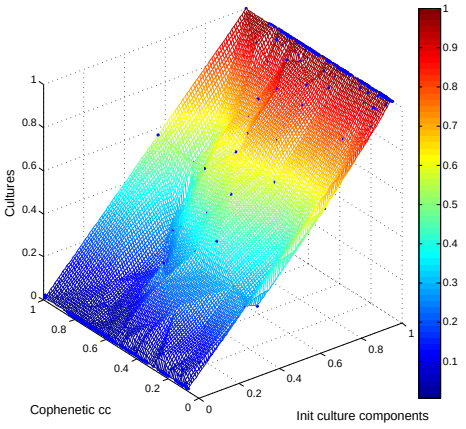


A Neutral evolution ($r = 1$)

B Neutral evolution ($r = \lfloor \frac{F}{2} \rfloor$)



C Prototype evolution ($k = 3$)



D Trivial ultrametric

Figure S14: Number of cultures at the absorbing state plotted against the number of connected components in the culture graph of the initial culture vectors and the cophenetic correlation coefficient of the initial culture vectors ($q = 10, F = 100$). These surface plots were generated by linear interpolation of the data points using MATLAB® (R2014a).

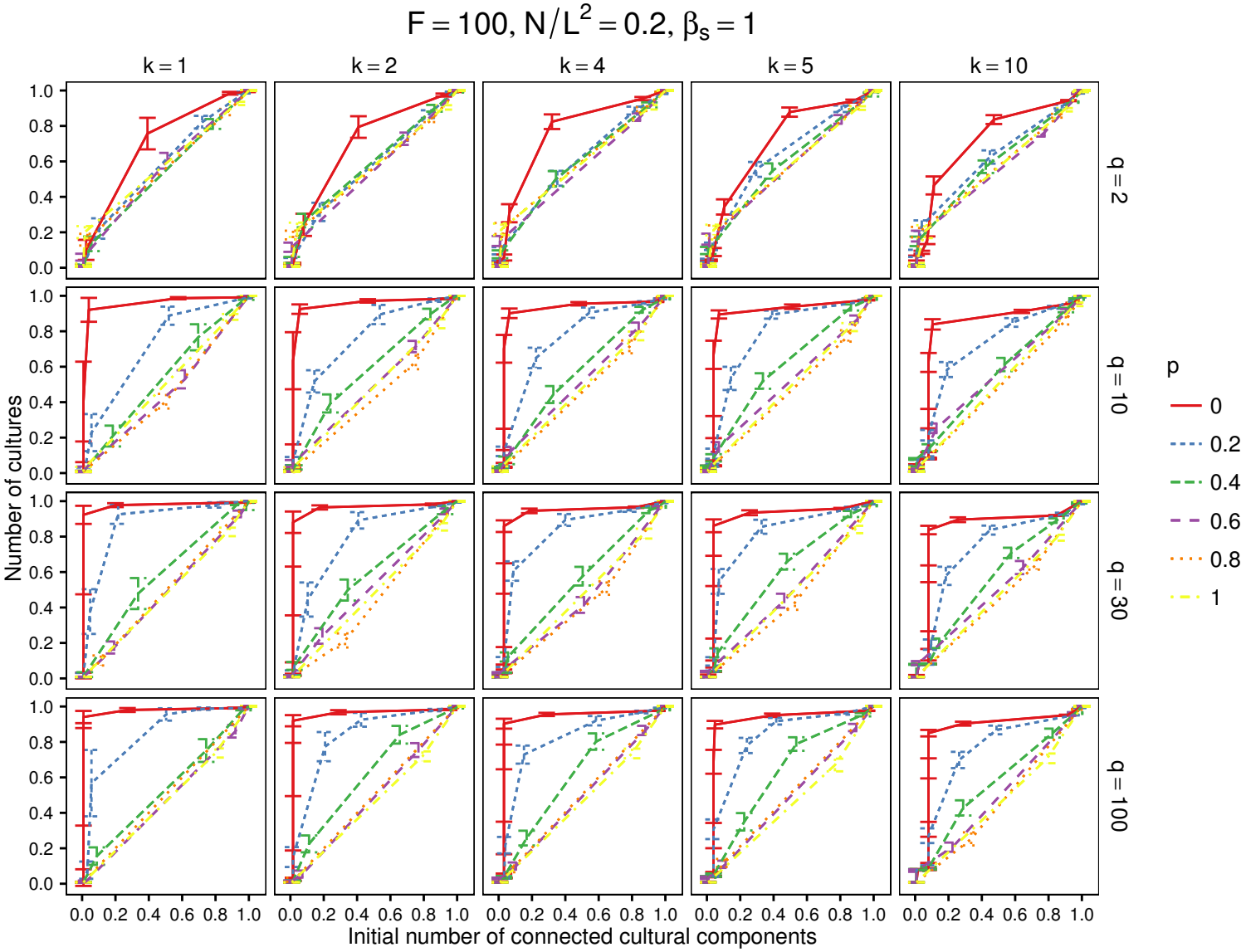


Figure S15: Number of cultures at the absorbing state plotted against number of connected components in the culture graph of initial conditions for the prototype evolved culture vectors with different numbers of prototypes k, for various initial perturbation probabilities p.

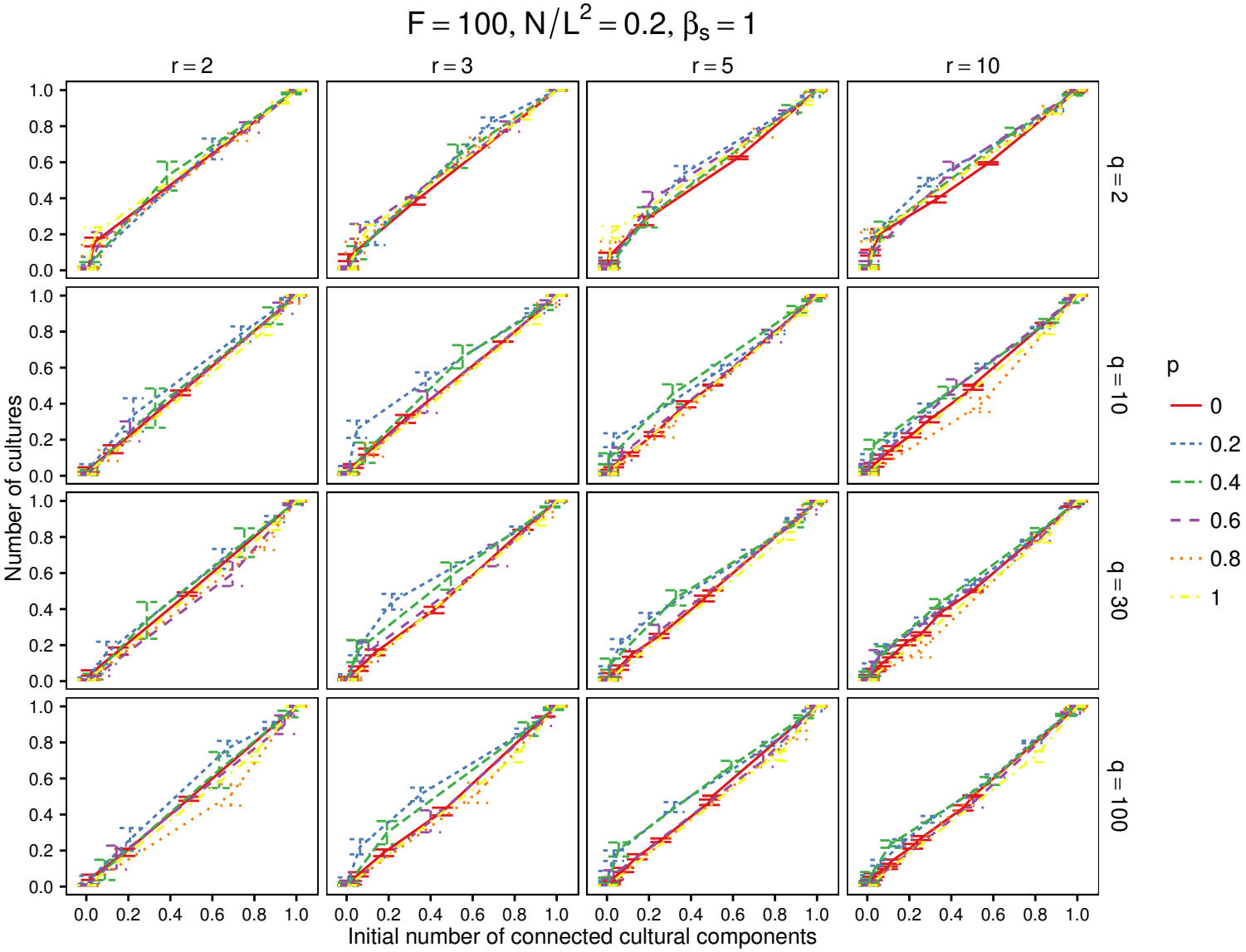


Figure S16: Number of cultures at the absorbing state plotted against number of connected components in the culture graph of initial conditions for the neutral evolution culture vectors with different numbers r of traits changed at each step, for various initial perturbation probabilities p.