Evolutionary dynamics of structured genetic algorithms

Working paper SFI Complex Systems Summer School 2010

Felix J. H. Hol * † § Xin Wang * ‡ §¶

* these authors contributed equally to this work. † Department of Biomanoscience, Kavli Institute of Nanoscience, Delft University of Technology, Lorentzweg 1, 2628 CJ Delft, The Netherlands, and §¶ Academy of Mathematics and Systems Science, Chinese Academy of Sciences, Beijing 100190, China

Populations are shaped by the spatial structure of their environment, it gives rise to a contact structure (interactions between individuals) and a population structure (interactions between (sub)populations). The topology of structures at both levels can have a profound influence on the evolutionary dynamics of a population. Here, we present a simple computational model in which genetic algorithms evolve cellular automaton rules to perform a classification task with which we investigate the evolutionary dynamics of (meta)populations consisting of sexually reproducing individuals that crossbreed within several distinct contact structures. We find that contact structures — modeled as graphs with various topologies — that limit the number of breeding partners show a higher evolvability and produce individuals with a higher fitness than populations with contact structures in which individuals have a higher number of possible partners. Furthermore, we show that well–mixed populations greatly benefit from being subdivided into (coupled) subpopulations. Within a metapopulation setting well–mixed contact structures can attain a fitness comparable to populations with a structured contact structure.

Any naturally occurring population is shaped by the spatial structure of its environment. This structure can influence interactions between individuals as well as the interplay between (sub)populations. Interactions on both the individual and population level can have a profound influence on the evolutionary dynamics of a population. Individuals in a well–mixed population, as for example bacteria or yeast in a stirred flask, compete globally for resources; this generally leads to the dominance of a single species which will be entirely replaced upon the invasion of a fitter mutant. In the presence of spatial structure however, a contact structure (interactions between individuals) and/or population structure (interactions between subpopulations) emerges, causing interactions to remain local and facilitating coexistence.

An attractive way of representing the topology of biological interactions is by using graphs. Graphs have been successfully used to model systems ranging from the molecular level (metabolic pathways [1]) to organism (the spreading of disease [2] and species levels (food webs [3]). Graph theory has also been applied to study the evolutionary dynamics of populations with various social structures, in this respect evolutionary games on graphs have received a great deal of attention [4, 5] (for a review see ref. [6]). These studies however, are limited to only one generator of biodiversity: mutation, and do not address recombination. Sexual contacts between humans [7, 8] and other animals [9] in various communities however, have been shown to be very well described by complex networks making the interplay between contact structure and the evolution of sexually reproducing organisms an interesting and relevant topic of study.

On top of its contact structure, a population may also be subdivided in distinct but coupled subpopulations. The presence of such an additional level of structure can also influence the evolutionary dynamics of the population. Sewall Wright hypothesized that a collection of interacting and interbreeding subpopulations would efficiently navigate an adaptive landscape by exploring different evolutionary pathways in parallel and, through occasional crossbreeding, make ‘adaptive hops’ that span large distances in genotypic space [10]. A larger single population on the other hand, is more likely to quickly converge to a nearby local optimum, subsequently lose genetic diversity through inbreeding, and, therefore, remain at this fitness peak without exploring paths to potentially higher peaks.

In this paper we use simple evolutionary models (genetic algorithms) to investigate the evolutionary dynamics of (meta)populations having contact structures with different topologies and population structures with varying degrees of connectedness. A population is made up of cellular automaton (CA) rules which are possibly the simplest system with which sexual reproduction can be modeled.

### The model

CA rules are convenient individuals to use in evolutionary computation because they can be coded as bit strings making mutation and crossbreeding easy. We implement different contact structures by defining the geometry in which crossbreeding takes place. CA individuals can either live in a well–mixed population or inhabit graphs with different topologies.

**CA basics.** This model utilizes one–dimensional, binary–state CA rules adapting to perform a density classification task. In the density classification task CAs need to decide whether the majority of the initial condition (IC) consists of 0’s or 1’s (let p denote the fraction, or density, of 1’s). The CA correctly performs this task by quickly converging to all 0’s (1’s) when the majority of the IC is 0 (1). The starting point of our model is analogous to work by Mitchell, Crutchfield and colleagues in which a genetic algorithm (GA) is used to evolve CAs with neighborhood size (radius) r = 3 [11]. The radius of the CA severely restricts the information available to a CA. When the length of an IC l >> r the CA needs to make a global decision by using only local information, this makes the density classification task non trivial. The look–up table defining a single CA rule can be coded as a bit string of length 2r+1, in the case of r = 3 this gives rise to 2128 ≈ 1038 possible CA rules. This rule space is vastly too large to do an exhaustive search for the CA which performs the density classification task the best. Using a GA however, previous studies have shown that
good (80–90% accuracy) CA rules can evolve from populations initiated by randomly generated individuals [12].

**GA basics.** The initial population of CA individuals is made by randomly generating 128–long bit strings, if λ denotes the fraction of 1’s in a CA rule, the initial population is uniformly distributed over λ ∈ [0, 1]. All CAs in a population are subsequently tested on 100 ICs with length l = 149 and density ρ (uniformly distributed ρ ∈ [0, 1]). Fitness is defined as the fraction of correctly classified ICs in less than 300 time steps. CAs within a population are ranked according to fitness and crossbred. Crossbreeding proceeds along the specified contact structure by selecting two individuals and producing a random number u uniformly distributed between 0–128. The individuals are subsequently crossed over at the position specified by u. Newly bred individuals are mutated (flipping of a bit) at every position with probability pm = 0.001. This finalizes the genetic process after which fitness can be assessed again.

**Contact structure — well–mixed population.** GAs are usually modeled as well–mixed populations in which an ‘elite’ (top 20% of the population) crossbreeds with the less fit individuals in the population. In our model this breeding is proportional to fitness: individuals that are not a member of the elite pick a crossbreeding partner from the elite according to: P(crossbreed with elite member i) = \frac{f_i}{\sum_{j=1}^{m} f_j}, where f_i denotes the fitness of individual i and the sum extends over the entire elite. Elite Individuals are mutated but do not crossbreed.

**Contact structure — graph topologies.** A node on a graph represents a CA individual which can only crossbreed with individuals to which it links; when an individual is fitter than all its neighbors it does not crossbreed (it does mutate). All graphs in this study are undirected and connected (i.e. there is a path from any node to any other node), in this way we make sure that there are no completely separated subpopulations. We investigate regular graphs, random graphs, and scale-free graphs. Regular graphs are modeled as lattices with either a von Neumann or Moore neighborhood giving connected 4– and 8–regular graphs. Connected random graphs are generated following the Erdős–Rényi model in which nodes are connected randomly with probability p, with the constraint that the graph has to be connected (which is generally the case when the degree k >> \ln(N)) [13]. Scale–free graphs are created by the Barabási–Albert model of preferential attachment: to an initially small connected random graph nodes are added one by one. Each new node brings in m links which are attached to existing nodes proportional to their degree k. This model gives rise to a graph that has a power–law degree distribution \(P(k) \sim k^{-\gamma}\) with an exponent 2 < \(\gamma < 3\) independent of m [14]. All graphs are updated (crossbred and mutated) synchronously.

**Population structure.** Population structure is implemented by generating a metapopulation of n subpopulations with similar contact structure. More specific, a metapopulation for example consists of n well–mixed subpopulations or n regular graphs. These subpopulations evolve separately for t generations after which mixing occurs. During mixing, individuals are placed into a random location of another subpopulation while the contact structure of each subpopulation is preserved.

**Results and Discussion**

We find that both contact and population structure have a profound effect on the evolutionary dynamics of the CA populations. Here, we focus on the evolutionary dynamics of populations of 500 individuals during 200 generations, a time frame in which most populations seem to reach a stable fitness level after which the genotype of the fittest individual becomes fixed in the population. We will assess the capacity of a population to evolve and innovate by analyzing the fittest population to evolve and innovate by analyzing the fittest individual present in a well–mixed population at generation t is shown in figure 2 A; multiple of these evolutionary runs can be averaged (figure 2 B) to yield the typical evolutionary performance of a specific contact/population structure. Individual traces show that the evolution of the CAs proceeds in a punctuated equilibrium fashion: periods of stasis are separated by large jumps in fitness. The slight increase in fitness that can be seen dur-
The influence of contact structure. We find that almost any contact structure leads to a better evolutionary outcome (i.e. higher fitness) when compared to an unstructured, that is, a well–mixed population. A well–mixed population produces on average an individual with a fitness of 90±2% whereas structured populations reach a fitness of up to 96±2% after 200 generations. Figure 3A shows the mean evolutionary performance of a well–mixed population and of various random graphs. Different graph structures can be compared by their efficiency given by:

\[ E = \frac{2}{N(N-1)} \sum_{i<j} s_{ij} \]  

where \( N \) is the number of nodes and \( s_{ij} \) denotes the shortest path between any pair of nodes \((i,j)\). It can be clearly seen that graphs with lower efficiency (< 0.6) attain a higher fitness and graphs with high efficiency converge to lower fitness. The large size of the error bars for graphs with a high efficiency shows that the fitness these contact structures achieve differs strongly between populations, whereas the small error bars for low \( E \) graphs indicate that these populations reliably produce high fitness individuals. Regular versus random contact structures do not differ significantly in evolutionary performance: both 4– and 8– regular graphs reach a fitness level comparable to random graphs with low efficiency (see figure 3B).

The evolutionary dynamics of scale–free contact structures are more difficult to characterize (figure 3B). Scale–free graphs with a rather low average degree (\( \langle k \rangle = 2–5 \)) give rather poor results, furthermore, simulations with scale–free graphs with an average degree of 10 yield less fit individuals than random graphs with the same average degree indicating that scale–free contact structures are less successful in utilizing their evolutionary potential. This can be understood by realizing that the presence of hubs (nodes with very high \( k \)) causes biodiversity to shrink rapidly: within the first couple of generations hubs become the fittest individuals in the population and subsequently crossbreed with most other individuals and therefore genetically homogenize the population; thereafter evolution becomes mutation limited. It has been noted before that the properties that give scale–free topologies their interesting and robust characteristics also make them fragile in other respects, as for example targeted attacks [15]. Only the most connected scale–free graph (\( \langle k \rangle = 39 \)) shows evolutionary dynamics comparable to a random graph with similar \( \langle k \rangle \), possibly due to the presence and competition between multiple connected hubs.

It is interesting to note that the evolutionary rate (\( \frac{df}{dt} \), where \( f \) denotes fitness) is independent of graph structure. Partly due to the existence of a maximum fitness in our model (a 100% score on the classification task) the evolutionary rate decays exponentially; a fit of \( \frac{df}{dt} \) with an exponential function results in exponents that are not significantly different (within 95% confidence bounds) for most contact structures. Only the 4–regular graph and scale–free graphs with low average degree have slightly lower evolutionary rates. This suggests that structured populations facilitate the generation of high fitness individuals without compromising the pace at which evolution proceeds.

The influence of population structure. By incorporating only the most rudimentary characteristics of a metapopulation, we find a strong influence on the evolutionary dynamics of CA populations. We find that a metapopulation of occasionally crossbreeding well–mixed subpopulations systematically finds better CA rules than a single well–mixed population. This effect is exhibited by a range of metapopulations consisting of a different number of subpopulations and crossbreeding at various intervals. Figure 4 shows that quickly after mixing of the
subpopulations, much fitter individuals evolve. This is probably due to crossbreeding between fit individuals that evolved in separate subpopulations and meet after mixing. The magenta curve in figure 4 shows a metapopulation consisting of 10 well-mixed subpopulations of only 50 individuals each. As can be seen from the curve at $t < 40$ this population size severely impedes evolution whereas subpopulations with $N = 100, 250$ or 500 show a faster evolutionary rate. It is quite striking however, that even this slowly evolving metapopulation jumps to a fitness level comparable to the other metapopulations right after mixing at $t = 40$. The metapopulation at the other extreme of our range ($N = 2 \times 250$) initially shows dynamics comparable to other metapopulations ($N = 5 \times 100$, various mixing intervals) but sustains this rate for a shorter period and thus converges to a lower fitness, on average.

Remarkably, the implementation of metapopulation structure in populations already having a ‘structured’ contact structure has no implications for the evolutionary dynamics of such a population. Populations with a regular or low efficiency random graph contact structure for example, produce individuals with the same fitness regardless of population structure.

**Conclusion**

The presence of contact as well as population structure clearly facilitates evolution. We have taken a very simple approximation to the ‘occasionally crossbreeding local races’ first proposed by Wright. Still, the elementary aspects of a metapopulation that we implement show a radical impact on the evolutionary path a (meta)population takes. Populations benefit greatly from being subdivided in subpopulations that are loosely coupled. Even when subpopulations are so small that evolution of the individual subpopulations is hampered (our $N = 10 \times 50$ simulations) a metapopulation structure can counter this negative effect and utilize the potential present in the initial population to a larger extent than a single ($N = 500$) population.

Regarding the majority of the contact structures presented here, we can conclude that the rate of evolution remains unaffected yet evolution is sustained for a longer period resulting in a higher fitness in the long term. We hypothesize that this prolonged evolution is caused by the persistence of biodiversity in structured populations. In well-mixed populations, fit individuals spread their genome rapidly through the entire population thereby genetically homogenizing the population without realizing the full potential present in the initial population. Contact structures with low efficiency (graphs with long average shortest path length) succeed in preserving diversity for a prolonged period and therefore produce individuals with higher fitness (close to the maximum possible in our model). Remarkably, this gain in fitness is hardly compromised by a decrease in the rate of evolution. It would be interesting to see whether sexually reproducing organisms have evolved social ties that lie within the range in which both evolutionary rate as well as performance are optimized.

**ACKNOWLEDGEMENTS**. The authors would like to thank the Santa Fe Institute for providing a very stimulating environment in which part of this work was carried out.

![Fig. 4. Average fitness of the best individual present in various well-mixed metapopulations through time. A single population consisting of 500 individuals shows the worst long term fitness (blue). The performance of 2 subpopulations evolving in parallel (cyan: $N = 2 \times 250$, mixing every 40th generation) is considerably better. Populations with a more pronounced metapopulation character (red: $N = 5 \times 100$, mixing every 40th generation; green: $N = 5 \times 100$, mixing every 60th generation; black: $N = 5 \times 100$, mixing every 80th generation) attain even higher fitness levels. Even when the size of separate subpopulations seems to impede evolution (magenta: $N = 10 \times 50$, mixing every 40th generation) occasional mixing within a metapopulation setting allows these subpopulations to reach a higher fitness level than the single population.](image-url)