Spatially Structured Evolutionary Algorithms

Marco Tomassini
University of Lausanne, Switzerland
marco.tomassini@hec.unil.ch

Why Topology Matters

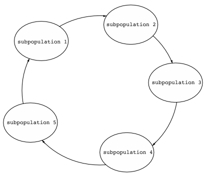
The spatial structure of a population will be called its *topology*

- Population topology has a marked influence on the dynamical processes taking place in the population
- To some extent, the dynamics can be controlled by using the appropriate topology
- Population topology can be mathematically characterized using the tools of graph theory

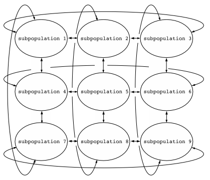
Main Categories of Population Topologies

- Multiple Populations, also called *island* models (each node of the graph is a population in itself)
- Cellular Populations (each node of the graph is a single individual)
- There are many possible *hybrid* models, such as islands of cellular populations, or islands that themselves contain other islands etc.

Examples of Island Population Topologies

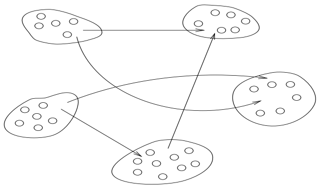


ring cellular structure



grid cellular structure

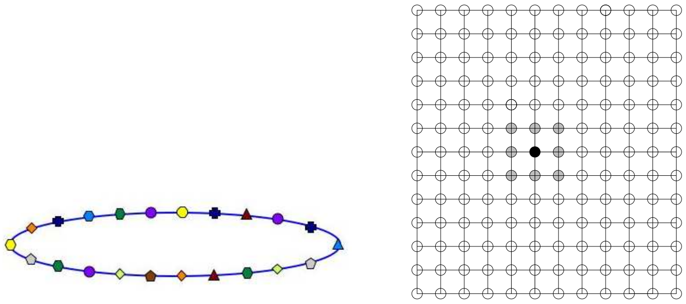
Mesh and Ring Topologies. Each circle represents a panmictic population.



Random Topology

Cellular or Lattice Topologies

Each individual occupies a cell in a 1-D, 2-D or 3-D lattice, or another graph structure



ring cellular structure grid cellular structure

Evolutionary Algorithms in Structured Populations

Island Models

- The whole population is subdivided into a number of subpopulations
- Subpopulations are loosely coupled: they evolve independently for a while
- A topological pattern of communication is established among the islands
- From time to time selected individuals are exchanged between populations and replace local individuals

A number of parameter values must be determined somehow: number of islands (subpopulation size), topology of communication, frequency of migration, individual replacement policy... Some of those might even change during the run

Island Models

- Island models have been often used as they are just a small departure with respect to standard panmictic EAs
- Empirically, they have been found to nearly always outperform the panmictic population model
- The most complete modeling and analysis has been done by Cantú-Paz for GAs [3]

Multi-Population Genetic Programming

A number of parameters must be considered:

- The number of islands (subpopulations)
- The size of the subpopulations
- The communication topology
- The number and type of migrating individuals
- The frequency of migration

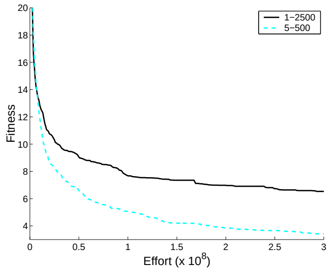
These parameters have been empirically investigated in Fernández et al. [5], on standard and real-life problems. Details of the test problems and results can be found there

Multi-Population Genetic Programming: Results

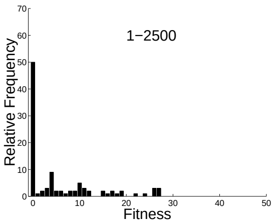
- In general, multi-population GP is more efficient than standard panmictic GP on those problems: **better results with the same computational effort**
- For a given total population size, there is a preferred interval for subpopulation size which is problem-dependent
- If the subpopulations are **too small**, island GP does not perform well
- The “optimal” number of individuals to exchange is about 10% of the subpopulation size; the frequency of exchange should be between 5 to 10 generations independent of the problem
- The influence of inter-island communication topology is comparatively less important

---

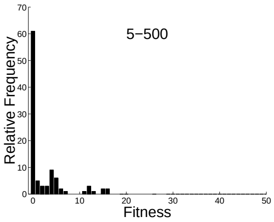
Experimental Results: Ant Problem



average effort



relative frequency of solutions panmictic GP

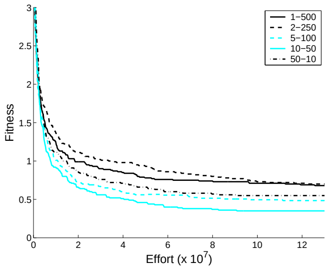


	E = 1×10^8	E = 2×10^8	E = 3×10^8
1- 2500	43	47	50 *(σ = 5.000)*
5- 500	53	60	61 *(σ = 4.877)*

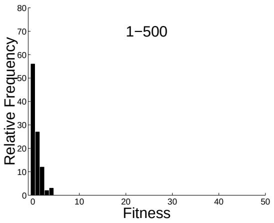
hit rate/100 runs

relative frequency of solutions multipop GP

Experimental Results: Even Parity Four



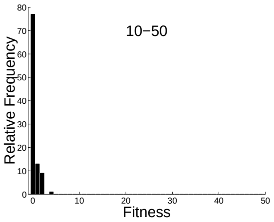
average effort



relative frequency of solutions panmictic GP

	E = 8x10^7	E = 10x10^7	E = 12x10^7
1- 500	53	54	56 *(σ =4.964)*
2- 250	55	60	60 *(σ =4.899)*
5- 100	64	65	65 *(σ =4.770)*
10- 50	76	77	77 *(σ =4.208)*
50- 10	62	64	65 *(σ =4.770)*

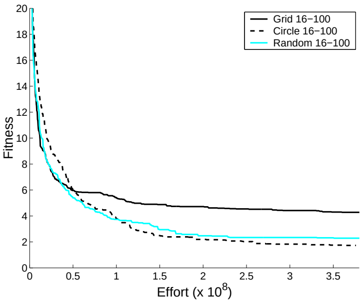
hit rate/100 runs



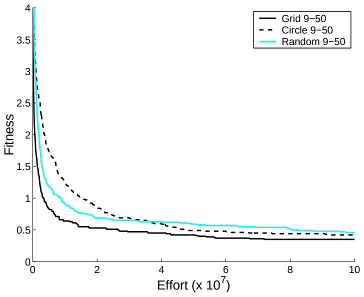
relative frequency of solutions multipop GP

Comparing Topologies

The empirical result is that, for island models, the precise migration topology is relatively unimportant, at least for the cases studied here. This is reasonable, given that evolution is still mainly panmictic



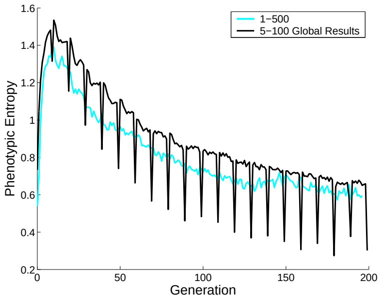
Ant problem



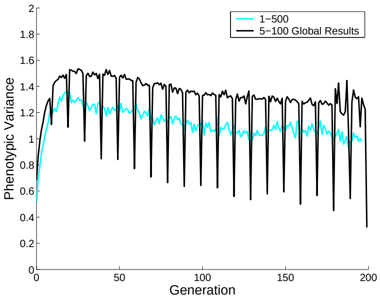
Even Parity 4 problem

Maintaining Diversity in Island GP

A better global phenotypic diversity during the run seems to be correlated with the good results obtained with multi-population GP



Even Parity - Entropy



Even Parity - Variance

Effectivity of Multi-Population EAs

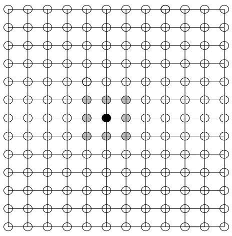
Summarizing, and extending to other island EAs for which many results exist:

- Most empirical results tend to show that island EAs are more efficient than panmictic EAs
- The effectivity of multi-population EAs seems to depend on the nature of the problem
- Overall population diversity is better maintained in a multi-population setting
- Separable problems and problems with multiple solution paths seem to be more suitable for the distributed approach

Evolutionary Algorithms in Structured Populations

Cellular Models

- Each individual occupies a cell in a regular lattice or a more general graph
- Genetic operators are local. Selection, mutation and recombination take place only within a small neighborhood.
- After selection and variation, each cell is replaced, e.g., by the best individual in the neighborhood



Selection Pressure in Lattice Cellular EAs

It is a good case study because:

- The effects of topology are most easily seen in lattice cellular EAs
- Selection pressure is a fundamental aspect of EAs
- Variation operators do not interfere with the dynamics
- Mathematical analysis is possible in some cases

Selection Pressure and Takeover Times

Takeover Time is the time it takes for a single best individual to take over the whole population

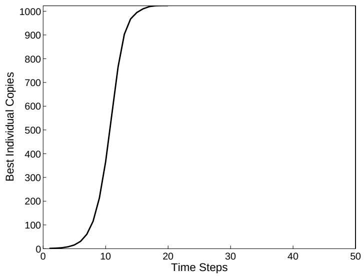
No variation operators: only **selection** is active with a probability p_s that depends on the selection method

Long takeover times mean less intense selection and viceversa for short TT

Selection intensity is related to the **explorative** or **exploitative** character of an EA: the stronger the selection the more exploitative the EA

Growth Curves in Panmictic Populations

In mixing populations the best individual propagates under selection following a *Logistic Curve.*



Analytical and experimental results indicate that, among the usual selection methods, (μ, λ), tournament, and linear ranking induce a stronger selection pressure than fitness proportionate selection

The Origins of Logistic Growth

Logistic growth occurs in situations where the growth is exponential at first but then it flattens out being limited by diminishing “resources”. In our case, it means that, as time goes by, less and less individuals remain to be “conquered”

Thus, the growth rate is not simply proportional to the current amount N, but rather to a maximum possible “capacity” θ, minus the current amount (Verhulst):

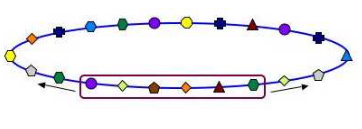
$$\frac{dN}{dt} = \alpha N(1 - \frac{N}{\theta})$$

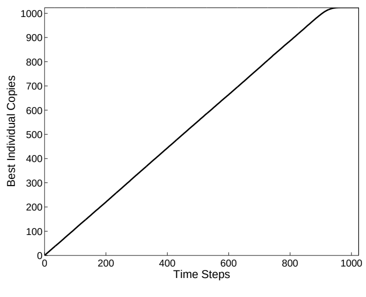
which has the solution:

$$N(t) = \frac{\theta}{1 + (\frac{(\theta - N_0}{N_0})e^{-\alpha t}}$$

Growth Curves in Rings

In rings the best individual can only grow at a *linear* rate:

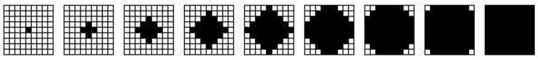


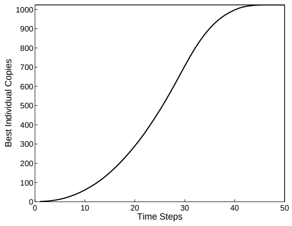


The frontier of the growing region can only expand, at best, to the next two individuals on the next time step

Growth Curves in Two-Dimensional Lattices

In grids the best individual can only grow at most at a *quadratic* rate:



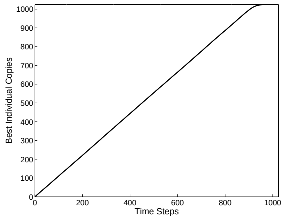


The *diameter* of the expanding region grows at a linear rate, and thus the whole area, which is proportional to the population size, grows at *quadratic* rate

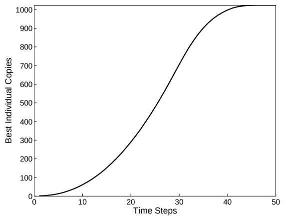
---

Growth Curves and Topology

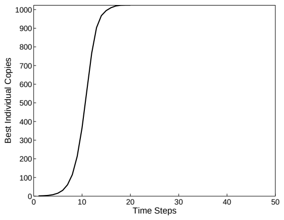
The influence of the population structure is clearly seen:



ring



torus



panmictic

The growth rate is much slower in rings than it is in 2-D grids, which is in turn slower than the mixing population

Mathematical Models for Growth Curves I

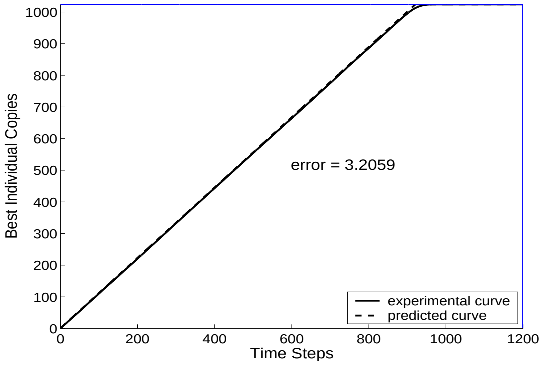
- The models are based on probabilistic difference equations giving the *expected value* $E[N(t)]$ of $N(t)$, the number of best individuals at time t (see [8] for the mathematical details)
- For *rings* with a neighborhood of three individuals, the solution is:

  $$E[N(t)] = 2p_s t + 1$$

  where the actual probability p_s should be inserted for different selection methods
- The equation can easily be checked for the deterministic case $p_s = 1$ in which $N(t)$ is no longer an expectation (i.e. a random variable)

Comparing Theory and Experiments

As expected, for the ring case the agreement between theory and experiment is excellent. The experimental curve (black) is the average of 100 runs. Selection method: binary tournament. Population size is 1024.



Mathematical Models for Growth Curves II

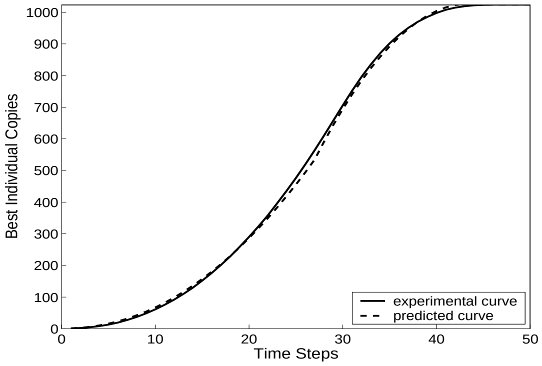
- For the synchronous growth curve in a 2-D torus, assuming a 5 cell (NWCES) neighborhood we get:

$$\begin{cases} N(0) = 1 \\ N(t) = N(t-1) + 4p_2 \frac{\sqrt{N(t-1)}}{\sqrt{2}} & , \quad for\ N(t) \le \frac{n}{2} \\ N(t) = N(t-1) + 4p_2 \sqrt{n - N(t-1)} & , \quad for\ N(t) > \frac{n}{2} \end{cases}$$

- The approximation is geometrical and is based on the growth of a closed planar shape that contains the region of interest (a 45 degrees rotated square). p_2 is the selection-dependent probability of selecting the best individual when there are two copies of it in the neighborhood [8]

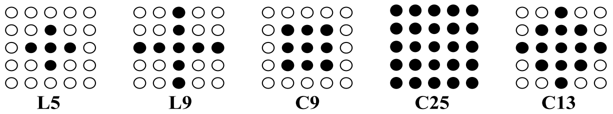
Comparing Theory and Experiments

For the torus the agreement between theory and experiment is still very good, in spite of the approximations in the model. The experimental curve (full) is the average of 100 runs. Selection is by binary tournament. Population size is 1024.



What About the Neighborhood?

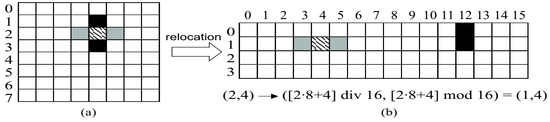
- What happens if the neighborhood's *size* and/or *shape* change?
- It is easy to modify the model to take that into account (see []). However, the effects had already been empirically studied by Sarma and De Jong for 2-D grids [11,12]
- Their conclusion: propagation times, and thus selection pressure, are closely related to the neighborhood's size. *Larger* neighborhoods imply *stronger* selection pressure



- Also: neighboorhoods having the same "linear extension" such as L9 and C13 induce a similar selection pressure; thus, neighborhood's *shape* matters too

Neighborhood Size and Shape: the Ratio

- Sarma and De Jong were able to characterize the global induced selection pressure by a single parameter: the *ratio* r
- The ratio is, in essence, the radius of a circle centered on the mean center $(\bar{x}, \bar{y})$ of a neighborhood pattern of n points
- Under this measure $r(L9) = 1.49$ and $r(C13) = 1.47$, which explains why the selection pressure is similar
- As the ratio → size of the grid, selection pressure → panmictic
- Alba and Troya later extended the concept of ratio to take into account the *whole grid shape*
- Selection pressure *decreases* as the grid *flattens*



The Time Dimension

- Up to now, only “space” in the form of topological population structures has entered into the picture
- Time has been considered *synchronous*; i.e., all the individuals act simultaneously at the ticks of a global clock
- But does this global synchronization make sense or is it only a useful abstraction?

Asynchronous Evolution

- Synchronous evolution is simple and can be used in artificial systems, where no physical limitation exists
- Asynchronous evolution is more complex but it is more faithful to Nature. No global clock. Signals can only travel at finite speed in physical and biological systems
- Since there can be many different sequential update orders for a cellular system, asynchronous evolution gives another degree of freedom to play with

Asynchronous Evolution: the Models

Three asynchronous evolution models will be used: *Line Sweep*, *Uniform Choice*, and *Random Sweep*

- In Line sweep (LS), the n cells are updated sequentially from left to right and line after line starting from the upper left corner cell.
- In Fixed Random Sweep (FRS), the next cell to be updated is chosen with uniform probability without replacement; this will produce a certain update sequence $(c_1^j, c_2^k, \ldots, c_n^m)$, where c_q^p means that cell number p is updated at time q and $(j, k, \ldots, m)$ is a permutation of the n cells. The same permutation is then used for all update cycles.

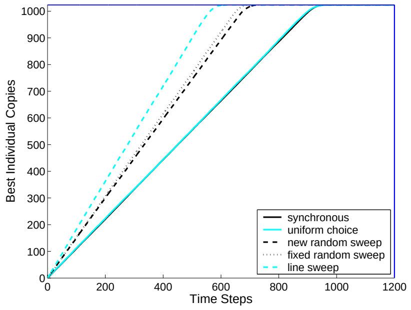
Asynchronous Evolution: the Models II

- The New Random Sweep method (NRS) works like FRS, except that a new random cell permutation is used for each sweep through the array.
- In *uniform choice* (UC), the next cell to be updated is chosen at random with uniform probability and with replacement. This corresponds to a binomial distribution for the updating probability.

A *Time Step* is defined as updating n times sequentially, which corresponds to updating *all* the n cells in the grid for LS, FRS and NRS, and possibly less than n different cells in the uniform choice method, since some cells might be updated more than once

Asynchronous Evolution: Results for Rings

Takeover Times results for **rings** for various update methods

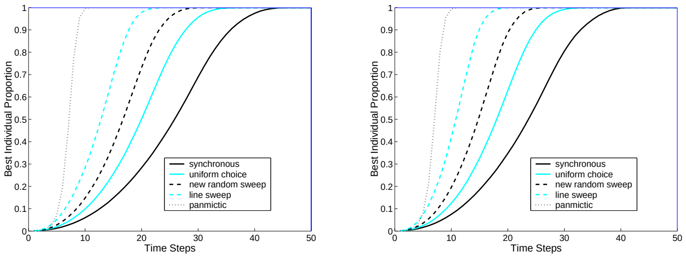


Takeover times with binary tournament selection: mean values over 100 runs. The vertical axis represents the number of copies of the best individual as a function of the time step

---

Asynchronous Evolution: Results for Thorus

Takeover Times results for **tori** for various update methods



(a) (b)

Takeover times with (a) binary tournament selection, and (b) linear ranking. Mean values over 100 runs. The vertical axis represents the number of copies $N(t)$ of the best individual in each population as a function of the time step t

Asynchronous Evolution: Results

Results can be summarized as follows [8]:

- As in the synchronous case, asynchronous evolution in lattices produces a selection pressure that is lower than the panmictic case. The ranking does not change, with selection being more intense in mixing populations than in grids, which is in turn more intense than rings
- Selection intensity using asynchronous evolution is slightly stronger than for the synchronous case for the same topological parameters. Uniform choice is close to synchronous
- In a given topology, different asynchronous update methods give rise to different global induced selection pressures
- Thus, selection intensity in cellular populations can be changed, even *dynamically*, by using different cell update methods, different grid or neighborhood ratios, or both

---

What About "Real" Cellular EAs?

- Typical benchmarks have been used, both continuous and discrete (The problems and the experiments are described in [4])
  - massively multimodal deceptive problems (MMDP)
  - satisfiability (SAT) problems
  - multimodal problem generator (P-PEAKS)
  - maximum cut of a graph (MAXCUT)
  - scheduling problems (MTTP)
  - continuous functions such as: Frequency Modulation Sounds (FMS), Ackley, Rastrigin etc.
- Those cover most classes of problems found in practice and should give an indication as to the observed tendencies

Many other problems have been studied, especially by Mühlenbein. Gorges-Schleuter, Rudolph, and coworkers

Summary of Results

- Synchronous CEAs with various grid axes ratios have been compared with asynchronous CEAs
- Results broadly confirm the influence of selection pressure: asynchronous CEAs converge faster than synchronous ones for nearly all problems
- On the other hand, synchronous CEAs with “flatter” ratios show an increased hit rate for most problems
- In general, although selection pressure plays a key role, it appears that the particular fitness landscape, and the operators used to traverse it are very important too

---

Random and Irregular Population Structures

Let’s begin with the *random graph* population structure

- A random graph with n vertices can be constructed by taking all possible pairs of vertices and connecting each pair with probability q, or not connecting it with probability $1 - q$
- A panmictic population can be seen as a completely connected graph or, equivalently, as a random graph with probability $q = 1$ of having an edge between any pair of vertices; such a graph has thus $\frac{1}{2}n(n-1)$ edges.

Random Networks

- In the completely connected graph (i.e. panmictic population), the number of neighbors of any individual is $n - 1$
- The random graph case is difficult to solve, since the number of neighbors (i.e. vertex degree) of a given vertex is a binomially distributed random variable. However, the *mean degree* is a constant equal to $q(n - 1)$. We thus use the *mean-field hypothesis*, taking for all individuals the same average number of neighbors
- We only consider *connected* RGs. Disconnected components do not make sense here

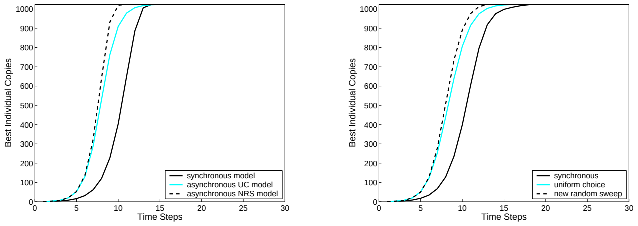
---

Using the Mean-Field Approximation

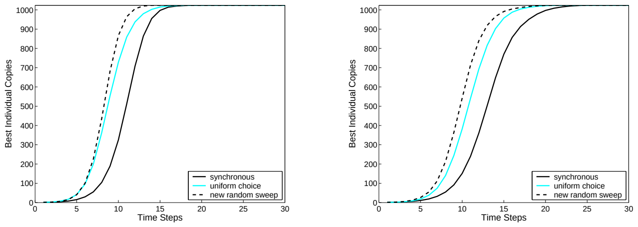
With the mean-field approximation, it turns out that both the *panmictic* and *random graph* topologies obey the *same* growth equation. The growth is obviously logistic in form, and is given as a discrete recurrence:

$$\begin{cases} N(0) = 1 \\ E[N(t)] = E[N(t-1)] + (n - E[N(t-1)])\frac{E[N(t-1)]}{n}, \end{cases}$$

Comparing Theory and Experiments



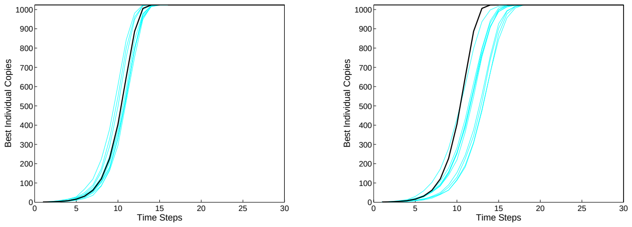
theoretical panmictic



random graph q=0.05 random graph q=0.01

Comparing Theory and Experiments

The agreement between theory (full curve) and experiment (light curves) is very good for the random graph with $q = 0.1$:



q = 0.05 q = 0.01

The fit is bad for small q. This is due to the mean-field approximation: for $n = 1024$ the average number of neighbors is ~ 50 for $q = 0.05$, while it is ~ 10 for $q = 0.01$. The σ is thus ~ 7 and ~ 3 respectively. Thus, many nodes will have very few edges for $q = 0.01$, slowing down the propagation

What Is in-between?

Regular Lattices ? Random Graphs

Some Graph Statistics

- The *average path length* $\langle L \rangle$ of a graph G is the mean of all the shortest paths from all vertices to all other vertices
- The *clustering coefficient* C of G is (informally) the likelihood, averaged over all nodes in G, that nodes that are connected to a given node are also connected between them. The higher this probability, the higher is C
- The *degree distribution* function $P(k)$ of G is the probability that a given node has exactly k neighbors
- The *average degree* $\langle k \rangle$ of G is the average number of neighbors of each node

Small-World Networks I

Small-world graphs are networks in which the average path length is short ($\langle L \rangle = O(logN)$, where N is the number of vertices). Thus, one can travel from any vertex to any other vertex in comparatively few steps, even in large graphs

This is also the case for standard random graphs. However, small worlds with the same number of vertices have a larger clustering coefficient. In other words, while random graphs are homogeneous in the average ($C = q = \langle k \rangle / N$), small-world networks have more local structure

Small-World Networks II

- two important kinds of small-world networks are the *Watts-Strogatz* model and the *Scale-Free* model
- the latter is much more typical of real networks, while the former is a mathematical convenience that can be used in *artificial systems*, where there are no hard constraints on the topology
- The following are useful references to start with: [6],[7],[10]

Small-World Networks III

The Model of Watts and Strogatz

One obtains a small-world graph by starting from a regular ring and successively “rewiring” edges with a certain probability β

Even very low values of β around 10^{-2} are sufficient to keep a high clustering coefficient C, while causing the mean path length $\langle L \rangle$ to tend to the low values typical of standard random graphs

These phenomena are due to the appearance of *shortcuts i.e.*, edges (links) that join distant parts of the graph

Small-World Networks IV

The Scale-Free Model

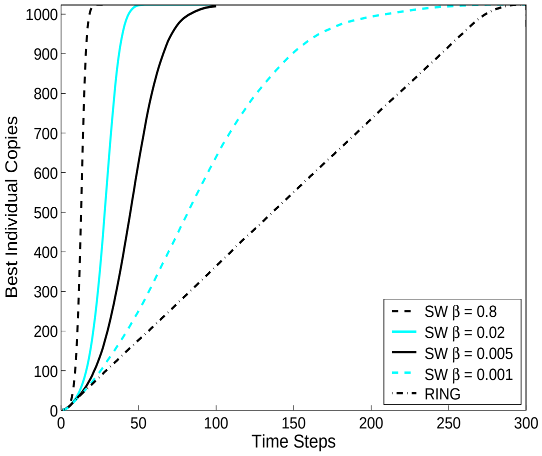
Scale-Free graphs are also small-world (high clustering, low average path length) but they are characterized by a degree distribution function $P(k)$ of the power-law form:

$$P(k) = c\,k^{-\gamma},$$

with c and γ positive constants, whereas random graphs and, to some extent, the Watts-Strogatz small-world model have a binomial $P(k)$

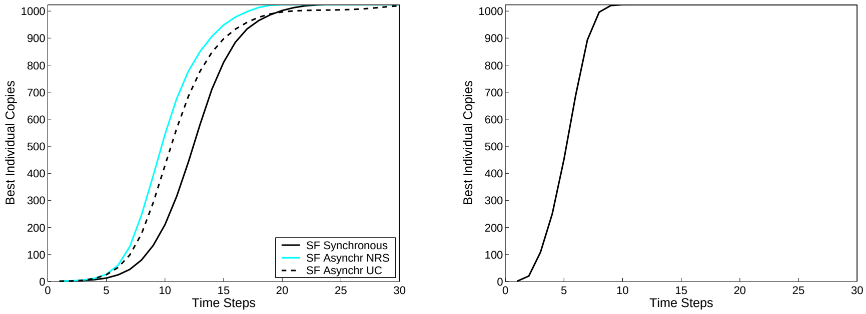
This form has been found in many real-world networks such as the Internet, the WWW, some biological networks, citation and collaboration networks and several others

Growth Curves on WS Small-Worlds



Synchronous update in WS graphs. Rewiring probabilities β grow from left ($\beta = 0$: ring) to right ($\beta = 0.8$: almost a random graph). Population size is 1024.

Growth Curves In Scale-Free Graphs



Left part: Growth curves for synchronous, and two asynchronous update policies. Initial best individual uniformly distributed among the nodes.

Right: Growth curve for synchronous evolution when the initial best individual is placed on a highly connected node.

Summary of Results on Small-World Networks

- The growth curves in Watts-Strogatz graphs are similar to those of random graphs and panmictic populations. Thus, just a few shortcuts allow very fast information flow through the network
- Scale-free graphs behave in the same manner when the initial individual is placed randomly among the nodes. When the initial individual is in a "hub" the takeover times are even shorter
- This suggests that information flow, and thus selection pressure, can be controlled by choosing β in WS graphs, or through the highly connected nodes in a SF graph

To Know More

[1] E. Alba and M. Tomassini. *Parallelism and evolutionary algorithms.* IEEE Transactions on Evolutionary Computation, 6(5):443–462, October 2002.

[2] D. Andre and J. R. Koza. *Parallel genetic programming: A scalable implementation using the transputer network architecture.* In P. Angeline and K. Kinnear, editors, Advances in Genetic Programming 2, pages 317–337, Cambridge, MA, 1996. The MIT Press.

[3] E. Cantú-Paz. *Efficient and Accurate Parallel Genetic Algorithms.* Kluwer Academic Publishers, Boston, 2000.

[4] B. Dorronsoro, A. Alba, M. Giacobini, and M. Tomassini. *The Influence of Grid Shape and Asynchronicity on Cellular Evolutionary Algorithms.* In Proceedings of CEC '04, IEEE Press, 2004

[5] F. Fernández, M. Tomassini, and L. Vanneschi. *An empirical study of multipopulation genetic programming.* Genetic Programming and Evolvable Machines, 4:21–51, 2003.

[6] R. Albert and A.-L. Barabasi. Statistical mechanics of complex networks. *Reviews of Modern Physics*, 74:47–97, 2002.

[7] M. E. J. Newman. The structure and function of complex networks. *SIAM Review*, 45:167–256, 2003.

[8] M. Giacobini, M. Tomassini, and A. Tettamanzi. *Selection intensity in cellular evolutionary algorithms for regular lattices.* IEEE Transactions on Evolutionary Computation, To appear.

[9] B. Manderick and P. Spiessens. *Fine-grained parallel genetic algorithms.* In J. D. Schaffer, editor, *Proceedings of the Third International Conference on Genetic Algorithms*, pages 428–433. Morgan Kaufmann, 1989.

[10] D. J. Watts. *Small worlds: The Dynamics of Networks between Order and Randomness.* Princeton University Press, Princeton NJ, 1999.

[11] J. Sarma and K. A. De Jong. *An analysis of the effect of the neighborhood size and shape on local selection algorithms.* In H. M. Voigt et al. editors, Parallel Problem Solving from Nature (PPSN IV), Lecture Notes in Computer Science vol. 1141, pages 236–244. Springer-Verlag, Berlin, 1996.

[12] J. Sarma and K. A. De Jong. *An analysis of local selection algorithms in a spatially structured evolutionary algorithm.* In T. Bäck, editor, Proceedings of the Seventh International Conference on Genetic Algorithms, pages 181–186. Morgan Kaufmann, 1997.

[13]D. Whitley, S. Rana, and R. B. Heckendorn. *Island model genetic algorithms and linearly separable problems.* In D. Corne and J. L. Shapiro, editors, Evolutionary Computing: Proceedings of the AISB Workshop, Lecture notes in computer science, vol. 1305, pages 109–125. Springer-Verlag, Berlin, 1997.