Network evolution models and bacterial communication

PhD Thesis
Sergiu Netotea

Supervisor: **Dr. Sándor Pongor**

Biological Research Center of the Hungarian Academy of Sciences

Institute of Biophysics

Laboratory of Bioinformatics

Szeged, Hungary, 2010

Biology Ph.D. School, *University* of *Szeged*

CONTENTS

1. Introduction

We first investigate models of network evolution. The network evolution models we propose are based on the rearrangement of links (''rewiring'') and follow the traditions of evolutionary modeling, i.e. optimize a fitness function that combines various factors into one numerical index. Naturally, there are many ways to formulate and combine the components of the fitness function and testing the possibilities makes the process computationally expensive.

We describe two algorithmic approaches in which all parameters are treated essentially as constraints: a mutation is selected if all of its parameters exceed or at least reach the corresponding values of the previous state, so there are no tunable parameters. This approach is computationally efficient so it allows one to study a wide range of phenomena. The other approach is a genetic algorithm that is used to discover fitness-optimized network structures by avoiding local maxima.

Our first goal was to study the evolution of robust yet efficient network topologies and to see if selecting mutations only for efficiency or only for attack tolerance (robustness) will influence network topology. We show that concomitant selection for efficiency and robustness influences the fundamental topological properties of the network, and that evolution under multiple attacks leads to distinct topologies.

Next we propose a model for bacterial colony dynamics that is used to explain experimental data related to the onset of swarming in environmental *Pseudomonas aeruginosa* PUPa3. The process was described with a simplified computational model in which cells in random motion communicate via a diffusible signal (representing *N*-acyl homoserine lactones, AHL) as well as a diffusible, secreted factors (enzymes, biosurfactans, i.e. "public goods") that regulate the intensity of movement and metabolism in a threshold-dependent manner. As a result, an "activation zone" emerges in which nutrients and other public goods are present in sufficient quantities, and swarming is the spontaneous displacement of this high cell-density zone towards nutrients and/or exogenous signals. The model correctly predicts the behavior of genomic knockout mutants in which the QS genes responsible either for the synthesis (*lasI, rhlI*) or the sensing (*lasR, rhlR*) of AHL signals were inactivated.

We next study the interaction of the bacteria based on inter-species distances and study the evolution dynamics of the graph of interacting bacterial agents.

1.1 Networks as models of communication

One of the most important and straightforward ways to model communication is to represent the communicating agents as nodes in a network. If the interaction between two nodes overpasses a certain threshold, we will say the two nodes are linked. The network will be defined by its set of nodes and links. We will define a few network properties and metrics used in this study, and we will describe several well knows biological networks and their topology from the point of view of information transfer.

1.1.1 Network properties

Let V be a non-empty set called the set of *vertices* or *nodes*, and let E be a set of unordered pairs of vertices (symmetric binary relation) on V. G = (E, V) is called an *undirected graph* [1] *or network* and E is the set of *edges* or *links* of G. In *directed graphs*, edges have a *direction* i.e. they go from one vertex to another and the pairs of vertices are ordered pairs. In the text we will use the term graph when we will use the terms network, nodes and links for natural, social or biological models of network and the terms graph, vertices and edges when we are making mathematical or algorithmic considerations on the network structure.

When two vertices of an undirected graph form an edge they are called *adjacent* or *neighbors*. A *path* from vertex u to vertex v in a graph G is a sequence of edges that are traversed going from u to v with no edge traversed more than once. The *length of a path* is the number of edges composing it. The *shortest path* between two vertices u and v is the path with the smallest length joining u to v.

A graph is said to be *connected* if a path exists between any two vertices. Complete graphs have any pair of vertices connected by an edge. The complete graph of order n has n(n-1)/2 edges. Graphs with a number of edges much lower than this maximum are called *sparse graphs*, while those graphs with the number of edges closed to the maximum are called *dense graphs*.

A *k-regular* graph is a graph in which each vertex has the degree k. A *strong regular graph* is a regular graph in which each two adjacent vertices have the same number of neighbors n in common and each non-adjacent pair of vertices has the same number of neighbors m in common. Graph regularity can be expressed in terms of distances as well. *Distance regular graphs* are graphs in

which for any two vertices v and w at distance i, the number of vertices adjacent to w and at distance j from v is the same.

1.1.2 Network metrics

The study of various social and biological networks uses a set of classic measurements for investigating network properties [2]. Some of these measurements refer to the global network, other address the properties of a single node or a pair of nodes in the network, but can be extended to the whole network by averaging. Node centrality measurements are important for our study, thus we will present a few centrality definitions. Other network measures that are not related to centrality are defined as follows:

Clustering coefficient

Clustering coefficient [3] is a measure of the likelihood that two associates of a node are associates themselves. For a vertex v_i, the clustering coefficient is defined as the proportion of edges within its neighborhood divided by the number of edges that could exist between them. In the case of an undirected graph, the number of edges possible is $\dfrac{k_i(k_i-1)}{2}$. So the formula for computing the clustering coefficient of vertex v_i is:

$$C_i = \frac{2N_i}{k_i(k_i-1)}$$

, where N_i is the number of links between the neighbors of v_i. The clustering coefficient of the whole network is the average of C_i for all the vertexes of the graph.

Path length – based measurements

Average path-length is the average of the path length between all pairs of vertices.

The *eccentricity* of the vertex *v* is the maximum distance from *v* to any vertex.

The *radius* of *G* is the minimum eccentricity among the vertices of *G*.

The *diameter* of *G* is the maximum eccentricity among the vertices of *G*. Thus, *diameter(G)=max{e(v):v in V(G)}*.

6

The *periphery* is the set of vertices with eccentricity equal to the diameter.

The *center* is the set of vertices with eccentricity equal to radius.

Centrality measures

These measures give a rough indication of the power and influence of a node estimating how central that node is in the network. Among the most important measures of centrality are degree centrality, closeness centrality, betweenness centrality, bridging centrality and eigenvector centrality. In our studies of network evolution we focused on the use of betweeness centrality as an indicator of central nodes, but we are detailing a few other measurements to show that the notion of centrality is extensive to many network properties.

Degree centrality

The *degree* k of a vertex is the number of neighbors. The *average degree* <k> is the average of all the vertex degrees in G. Nodes with a high degrees are usually more central in a network, but even if situated in the periphery they influence at least their direct neighbors. Degree distribution plots can tell us information about the overall topology of a network. In the case of directed networks, an in-degree and an out-degree is measured, for the links convergent on the node or divergent from it, respectively.

Mathematically the degree centrality of a node v, belonging to a graph of order n is expressed as:

$$C_D(v) = \frac{\deg(v)}{n-1}$$

Betweenness centrality

This measures the extent to which a node lies between other nodes of the network. Nodes that connect large clusters will thus have a higher centrality. This measure can also be applied to links. Nodes or links with high betweeness usually occurs if there are many shortest paths traversing them, therefore the betweeness is a measure of information flow.

For a vertex v of a graph with the set of vertices V, the *betweeness centrality* [2] is given by the formula:

$$C_B(v) = \sum_{\substack{s \neq v \neq t \in V \\ s \neq t}} \frac{\sigma_{st}(v)}{\sigma_{st}}$$

, where $\sigma_{st}(v)$ is the number of shortest paths from s to t that pass through a vertex v and σ_{st} is the number of shortest paths from s to t.
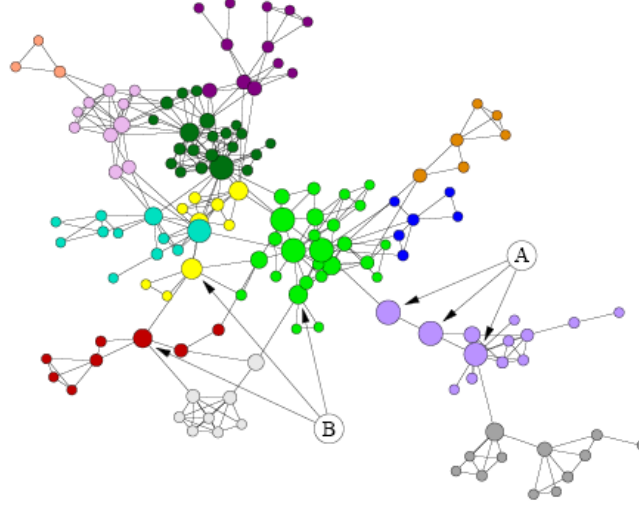
Figure 1 Schematic representation of scientific coauthorship network, bigger nodes have higher betweenness. One can observe that betweeness centrality is different than other measures of centrality, like degree centrality. Many nodes with high betweenness have small degree. Image adapted from [4].

Closeness centrality

According to this measure, a node is more central if it is more near to all the other nodes of the network. Thus for a node v, closeness defined as the inverse of the sum of the shortest distances between each individual and every other person in the network:

$$C_C(v) = \frac{1}{\sum_{t \in V \backslash v} d_G(v,t)}$$

Bridging Centrality

An edge is said to be a *bridge* if deleting it would cause its endpoints to lie in different components of a graph. For a vertex v, the bridging centrality [5] is defined as:

$$BC(v) = \frac{d(v)^{-1}}{\sum_{i \in N(v)} \frac{1}{d(i)}}$$, where d(v) is the degree of node v.

Just like in the case of betweenness centrality, bridging vertices do not lie necessarily on nodes with high degree centrality.
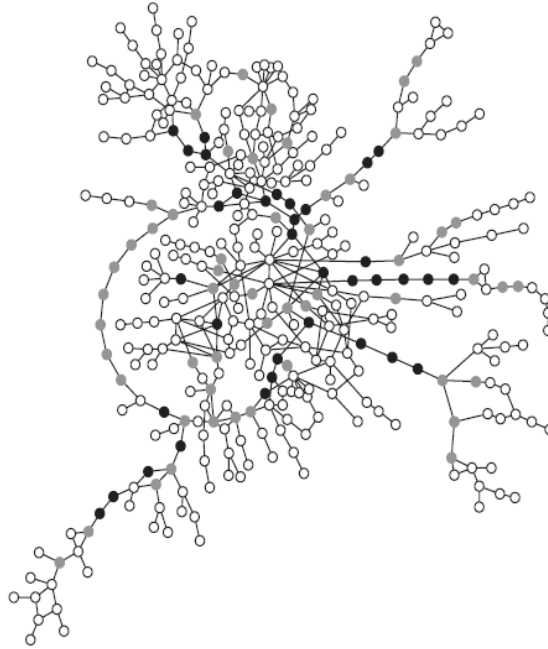
Figure 2 **Schematic representation of the yeast metabolic network with high bridging nodes outlined in black.** *Bridging nodes do not lie necessarily on nodes with high degree centrality.[5]*

Eigenvector centrality

Eigenvector centrality [6] is measuring the importance of a node by the number of important nodes the node links to. To all nodes centrality scores are assigned. A certain node will have a higher score if it is connected to other high score nodes. A famous variant of this centrality measure is Google's *PageRank* algorithm [7].

To define this measure we need to introduce the *adjacency matrix* of the graph $A = (a_{ij})_{0 \le i, j < n}$, having $a_{ij} = 1$ if vertex i and j are adjacent, and $a_{ij} = 0$ otherwise. Let x_i be the centrality score of node i.

The centrality of node i must be proportional to the sum of the centralities of the adjacent nodes, therefore:

$$\lambda x_i = \sum_{0 \le j < n} a_{ij} x_j$$

This can be rewritten as the *eigenvector equation*:

$$Ax = \lambda x$$

Because the *eigenvector* x entries must be positive only the greatest *eigenvalue* λ can be used to compute the desired centralities. The components of the eigenvector will hold the centralities of the graph.

1.1.3 Network topologies

A complex network is usually defined in opposition to a regular or a random graph, having certain features that are neither regular nor random. Among such features we number a heavy tail in the degree distribution, high clustering coefficient, assortativity among the vertices, community structure and hierarchical structure.

Naturally occurring networks have properties that distinguish them from mathematical models of graphs [8][9]. The ability to artificially construct or evolve networks with abilities similar to real world networks is a challenge that receives a lot of interest from the scientific community. Among the constructed graphs there are several important models that aim to reproduce properties observed in natural networks: random graphs, scale free graphs, small worlds. We will present each model together with its major properties.

Network topology describes the arrangement of the links and nodes in a network. Networks can be classified topologically in several ways, depending on the field of study. From the point of view of information transfer networks can be grouped in a few major topologies:

1. Bus topology. All nodes are connected to a major information highway.
2. Star topology. All nodes are connected to a central node thru which all data is transmitted.

3. Ring topology. Each node is connected to two other nodes in a circular fashion.
4. Meshed network. Each node has the same number of neighbors, forming a regular graph. If each node is connected to all other nodes, the network becomes fully connected.
5. Tree or hierarchical network. A central root node (or set of nodes) situated at the top level of the hierarchy is connected to other nodes (or set of nodes) at the next level of the hierarchy.
6. Line topology. Same as the star topology only that information is not flowing in a circular fashion.
7. Fully connected. Each two nodes have direct connection.
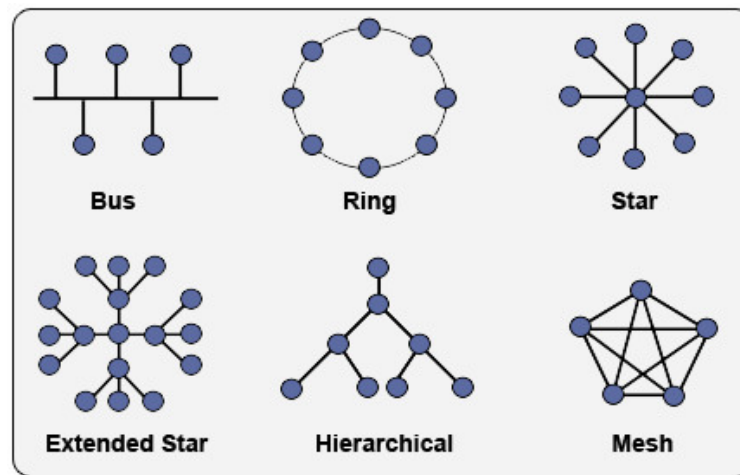
Figure 3 Different network topologies. Real networks, especially natural ones usually combine together these topologies. Image from Wikimedia's public domain.

Growing versus evolving network topologies

There are 2 ways of constructing a network with desired topology [10]: growing a network and evolving a network. Although the terms are used sometimes without distinction, it is clear what makes the difference.

In a *growing network* new nodes or links are added according to probabilistic or algorithmic patterns. In an *evolving network*, new nodes and links are not added, while links are rewired between other pairs of nodes according to probabilistic or algorithmic patterns. The evolution of a biological network is different from the biological theory of evolution, although there are clear similarities, like mutation and selection (but they have an algorithmic foundation rather than a biological one). Evolution in our studies is best understood as physical rather than biological, meaning *change in time*, the dynamics of network transformation under selective pressures.

Each network construction process has its own importance, modeling an important feature of the real world networks. Relevant to the topology of a network is not only the degree distribution, but other centrality plots as well, as well as several local and global measures of efficiency, compactness, clusterization, etc [11].

Random graphs

The *Erdős–Rényi model*, named for Paul Erdős and Alfréd Rényi, is a model for generating *random graphs* [12]. This model is usually defined in 2 ways:

1. The G(n,e) model, where a graph is chosen uniformly from the set of graphs of n vertices and e edges.
2. The G(n,p) model, where a number of n vertices are connected with a given probability p.

The second definition is used for constructing random graphs computationally, while the first definition is the standard mathematical concept of a random graph. According to the probabilistic definition, edges are independent and each edge is equally likely. As a result the total number of edges in a random graph is a random variable whose expectation value is $p\binom{n}{2}$, where $\binom{n}{2} = \frac{n(n-1)}{2}$ is the maximum number of edges possible. The graphs with n vertices and e edges have equal probability:

$$p^e (1-p)^{\binom{n}{2}-e}$$

The degree distribution is binomial, given by the law:

$$\Pr\big(\deg(v) = k\big) = \binom{n-1}{k} p^k (1-p)^{n-1-k}$$, and such a graph has on average $\binom{n}{2} p$ edges.

There are many other models of random graphs, some of whom have other degree distributions, including scale free. Random graphs can be constructed with a specified degree distribution [13]. Usually if the process involved in constructing a graph has a random choice of

edges in it, the graph will be termed random. In this work we use the original random graph concept, meaning uniform probability of choosing the graph edges, but we add the additional constraint that the graph must be connected. It was shown [12] that connected random graphs converge to random graphs in the limit of large graph size.
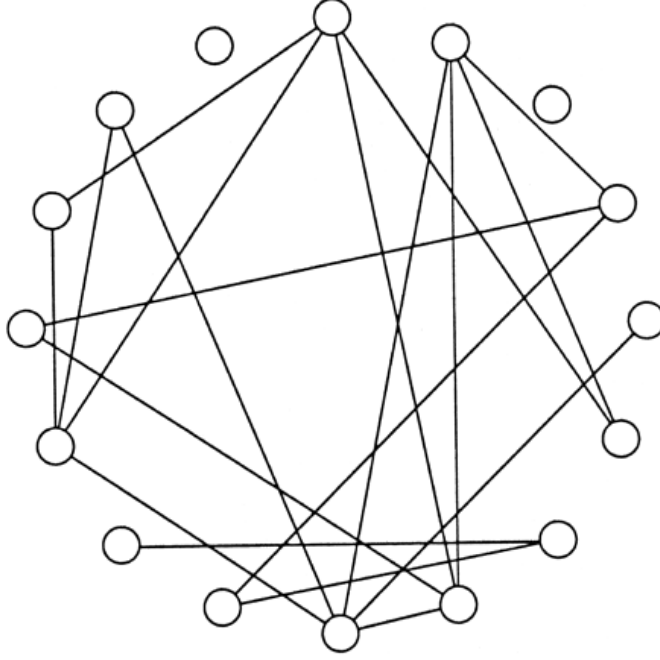
Figure 4 An example of a standard random graph of the type first discussed by Erdős and Rényi.
In this case, the number of vertices N is 16 and the probability p of an edge is 1/7. The example here explains the simple process of constructing a random graph. The vertices are arranged in a circular fashion and edges are wired randomly. [14]

Scale free networks

Random graphs constructed using the Erdős–Rényi model have a characteristic binomial degree distribution, but it was observed that many important real networks have a distribution that follows the law:

$\Pr(\deg(v) = k) = k^{-\gamma}$, where γ is a constant that for most naturally occurring networks has the range $2 < \gamma < 3$.

The most famous generative model for scale-free networks is the *Barabasi-Albert model* [15]. This model assumes:

1. Growth: the number of vertices in the graph increases over time.

2. Preferential attachment [16]: each new vertex added over time is connected to an existing node with a probability that is proportional to the number of nodes that the existing node has. Thus the probability p_i that a new node is connected to node i is: $p_i = \dfrac{k_i}{\sum\limits_{j} k_j}$, where k_i is the degree of node i.

Scale free network distribution appears as a roughly straight line on a log plot, due the low number of highly connected nodes, phenomenon termed "fat tailed distribution". Most communication networks as well as most biological networks exhibit this phenomenon.
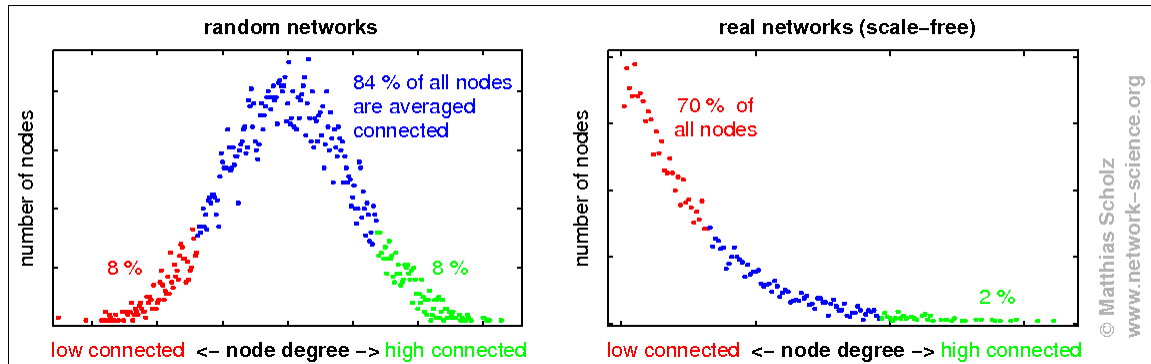
Figure 5 Comparative degree distribution of random versus scale-free networks. While the quarter of nodes of lowest degrees is shown in red, the quarter of highest degrees is shown in green. The 50 percent of nodes having medium degrees is shown in blue. It clearly reveals that in traditionally assumed random networks the majority of nodes have medium node degrees, while in real networks the majority of nodes has only a small number of connections. Note, the given percentages can vary slightly by using different data. Image from network-science.org.

Small worlds

A small world is a graph in which most vertices are neighbors and close to one another. This translates in mathematical terms into two properties: small average shortest path and small clustering coefficient.

The random graphs generated by the Erdős–Rényi model have a small average shortest path but they do not have a small clustering coefficient. The *Wattz-Strongatz model* [3] is the most famous algorithm for generating a small-world graph. The model consists in two steps:

1. Construct a regular ring lattice. Let K be the average degree. Connect each vertex to K nearest neighboring vertices (K/2 on each side).

2. For each vertex rewire every adjacent edge with a probability $0 < p < 1$, choosing a new vertex uniformly, but avoiding loops and edge duplication.

It is considered that most natural networks combine the properties of the scale-free and small world models.

Biological networks

The biological networks are inferred based on expression data for the genes, proteins, metabolites, communication signals, etc [17]. Biological networks were partly responsible for the development of the scale-free and small-world network models [18], as well as many studies concerning their community structure [19]. A lot of the current research is directed into network motifs and network homology studies, dynamics of networks and evolution patterns. Next we will describe the major biological network types and their most important properties.

Transcriptional regulatory network

The nodes of a regulatory network [20] are *genes* (that transcribe mRNA which in turn will make specific proteins). The network is directed, genes serving as the source of a directed regulatory activation or inhibition of a target gene through the release of *transcription factors*. The transcription factors are proteins or mRNA. Proteins bind to the promoter region of other genes, activating them in a network of regulatory cascades. Transcription factors can be activators or inhibitors of gene expression [21].

Metabolic network

A metabolic network [22] [23] [24] is the network of metabolic pathways of a cell, comprising all the chemical reactions that occur between the different metabolites of the cell. It contains both the chemical reactions between metabolites and the regulatory interactions that guide these reactions. The various metabolites are nodes and the links are directed, the target metabolite being a direct reaction product of the source metabolite.

Signal transduction network

The signal transduction network [25] [26] contains the entire set of cell changes due to receptors being activated by signals, called *signaling pathways*. The nodes consist in the set of signaling agents and the proteins involved in signaling, and the network is usually directed.

Protein-protein interaction network

The networks comprising the interaction of all proteins in a cell are extensive and it is sometimes difficult to visualize and compute global parameters on it. A protein-protein interaction network [27] [28] [29] is usually undirected, having as nodes the various proteins in the cell. A link is established if the two proteins interact by temporarily binding to each other, or if the interaction has strength by forming protein complexes.

1.1.4 Communication metrics in networks

Network efficiency

The global efficiency [30] [31] [32] of a network is defined as the average of the shortest paths inverses:

$$E = \frac{1}{n(n-1)} \sum_{i \neq j} \frac{1}{d_{ij}}$$

, where n is the number of nodes, and d_{ij} is the shortest path length between the nodes i and j. The use of the inverse pathlength instead of the pathlength itself is due to situations in which the graph might not be connected, in which case it will add zero instead of infinity, so the formula can be used for unconnected graphs as well. In the case of an unweighted graph, the efficiency is a measure of information flow across a network, and by naming $S(G) = \sum_{i \neq j} \frac{1}{d_{ij}}$ it can me formulated in the following way:

$$E(G) = \frac{S(G)}{S(G_{ideal})}$$

, where G_{ideal} is the complete graph that maximizes sum S.
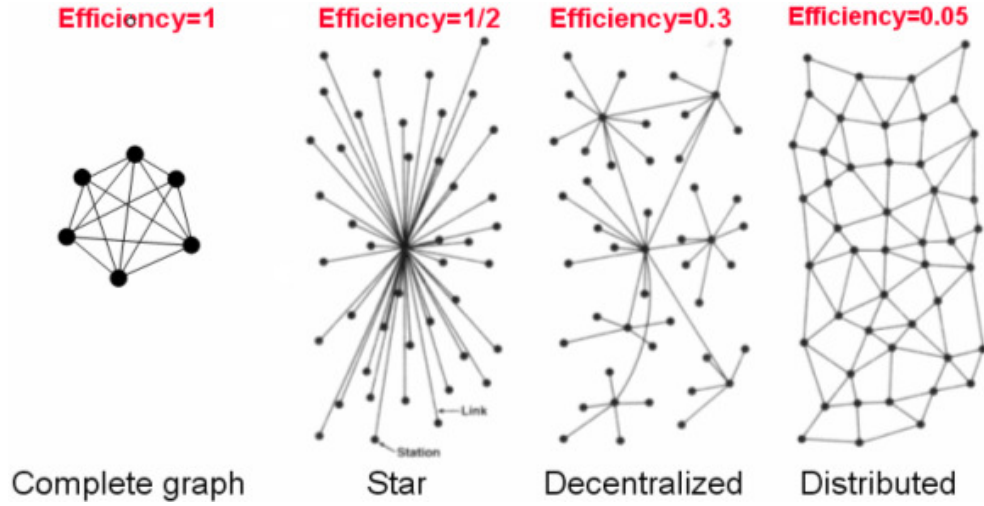
Figure 6 Network communication efficiency for various networks. The most efficient structure in terms of communication is a complete graph where all nodes communicate directly, of course a complete graph is also the most costly network structure. The second most efficient network structure is a star graph, and most biological networks but also modern communication networks have a decentralized structure with a few central nodes and many peripheral nodes. Networks distributed according to uniform probability laws as well as lattice graphs are among the least efficient models of communication.

Besides the global measure of efficiency, characteristic node efficiency can also be computed. The local efficiency of node i is defined as follows:

$E_i = \dfrac{E(G_i)}{E(G_i^{ideal})}$, where G_i is the subgraph of the neighbors of i and G_i^{ideal} is the ideal graph containing all the other possible edges of G_i. The local efficiency can tell us how efficient is the communication between the direct neighbors of a node, therefore how fault tolerant are its neighbors.

Network robustness

Network robustness [33] [34] has received a lot of attention from researchers. Much of this attention was due to the fundamental role that networks play in sustaining the society: communication, power grids, etc. In recent years however, concepts of network robustness were used also for the study of biological networks [35] [36]. The robustness of a biological network stems partly from its selective fitness and it is best understood if viewed in an evolutionary context.

A good example for the importance of network robustness is a 2005 study of our group [37], where it is shown that even though robust systems, like the molecular networks of living cells are often resistant to single hits such as those caused by high-specificity drugs, partial weakening of the *E. coli* and *S. cerevisiae* transcriptional regulatory networks at a small number (3–5) selected nodes can have a greater impact than the complete elimination of a single selected node. The attack strategies were the elimination of all links adjacent to a node, called *complete knockout,* the elimination of half of the nodes called *partial knockout*, and giving higher weight to some links called *attenuation*. In all cases, the targeted nodes have the greatest possible impact. The results suggest that multiple but partial attacks mimic well a number of in vivo scenarios and may be useful in the efficient modification of other complex systems. For example, in some cases broad specificity compounds or multitarget drug therapies may be more effective than individual high-affinity, high-specificity ones.
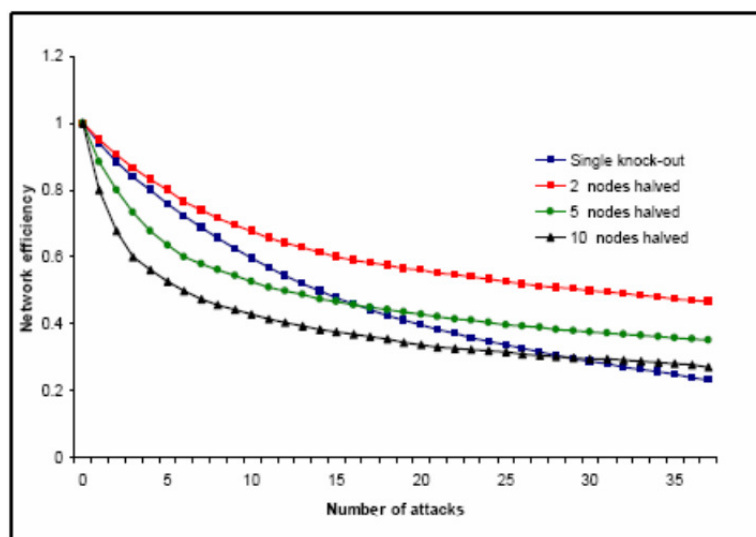
Figure 7 *Plot illustrating the work of Ágoston et al. [37] studying the behavior of E.coli and S. cerevisiae under single knock-out and attenuation experiments. Attenuating a few central nodes has similar effects over network efficiency.*

Although atacking links has its importance, in our thesis, we will restrain our analysis to attacking only nodes. A network node can be attacked in two important ways [38]:

a) *targeted attack*, in which the nodes that are more central are removed

b) *random attack*, in which nodes are removed with a uniform probability

It was found that complex networks are more vulnerable to targeted attack and less vulnerable to random attack compared to random graph models [39].

18

Vitality measures are important for the determination of important nodes in a network. We measure the *vitality* of a certain node by the amount of change it does to a network if that node is removed [2]. For example, if we use the efficiency to measure the change that the removal of node v does to a network, we have the following formula for the efficiency vitality of node v:

$$V(v) = E - E_v$$

, where E_v is the efficiency of the network after the removal of node v and its adjacent links and E is the efficiency of the original network.

We define robustness as the capacity of a network to survive attacks. Similarly to the attack types, we can define two types of robustness: robustness to targeted attack and robustness to random failures.

To measure robustness we extend the notion of robustness to multiple attacks: we attack a number of k most vital nodes of a network and then measure the efficiency E_k of the remaining network. We will define robustness as:

$$R = \frac{E_k}{E} = 1 - \frac{V_k}{E}$$

, where V_k is the vitality of the k most vital nodes.

As the equation suggests, the robustness has its values in the interval $[0,1)$ and the higher the robustness, the less effect have the most vital nodes to the efficiency of the network. This type of robustness is therefore robustness to targeted attack.

The mode of targeted attack itself is important: attacking all the k most vital nodes at once has not the same result as successive attack, where each time the most vital node is computed and attacked. The attack of a vital node changes the network structure profoundly and sometimes after one attack other nodes that were less vital before the attack appear as more vital after the attack.

We mention that node vitality and network robustness based on efficiency is just one of several ways to measure the evolutionary fitness of a network in terms of robustness. Network can be robust in many ways, in this work however we are interested to attack robustness [40].

1.1.5 Evolution in network structures

The concept of network evolution has several meanings considering the field of study. If network physics are studied, the evolution means the dynamics of network transformation due to growth and internal change [41]. If the evolution of a network is viewed biologically, it is usually defined as the optimization of a network-dependent fitness, across a landscape of network structures [42] [25] [24].

The biological details for network evolution were studied more extensively on protein-protein interaction networks and gene regulatory networks. There are two major processes of evolution involved: *duplication* and *divergence*. The first process called *duplication* is based on the mutations suffered by a gene which will cause the proteins to duplicate and in time to interact differently by having new connections to other proteins or loosing some of the existing connections (link attachment and detachment or *divergence*) [43] [44].
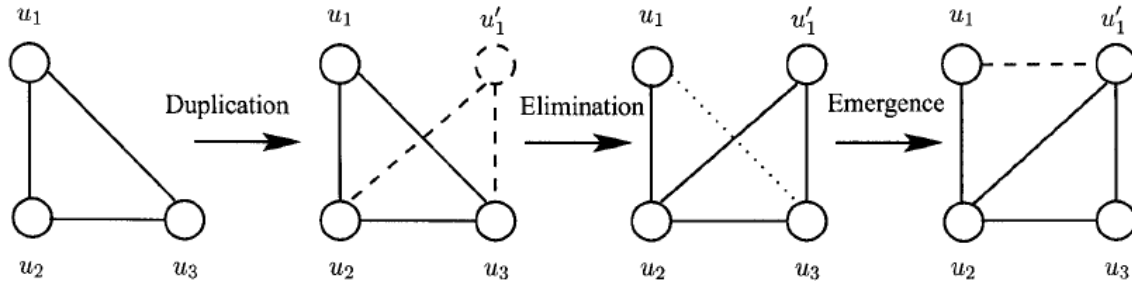
Figure 8 Duplication/divergence model for evolution of PPI networks. Starting with three interactions between three proteins, protein u_1 is duplicated to add u_1' into the network together with its interactions Then, u_1 loses its interaction with u_3. Finally, an interaction between u_1 and u_1' is added to the network. [45]

The natural constraints that evolve a biological network are various and any model is only a simplification of an evolutionary process. The simple assumption is that there are two ways in which a network evolves: *growth* and *rewiring*. The growth will add or remove nodes and links so that the overall network structure is more fitted to the environment pressure. Rewiring is the process of deleting a link and placing it between another pair of nodes. In dynamic networks, the process of rewiring and growth becomes more complex, taking into account changes in the load of the nodes affecting the flow patterns [46] [47].

Network alignment

Since biological networks evolve, networks can be aligned similarly to sequence alignment and basic homology studies can be done. Network alignment is particularly developed for protein-protein interaction networks and gene regulatory networks. Various programs have been developed that identify the homologous substructures, called network motifs, one such example being *PathBLAST* [48].

The basic alignment is pairwise network alignment [45], where two networks are compared and searched for homologous pair of interactions. For protein-protein interaction networks this method is used for identifying conserved clusters indicating protein complexes. The straightforward method for network alignment is the generation of a network alignment graph which is a merged representation of the two networks being compared, having as notes sets of molecules with good sequence alignment scores while the links are conserved molecular interaction across the two networks.

An extra improvement is using a scoring function that weighs the density of a given subgraph to the likelihood of discovering the same patterns at random [49]. For regulatory networks, conserved paths correspond to important signaling pathways [50].

Network motifs

An important result of network evolution is the identification of conserved subgraphs called *motifs*. Several research groups have devised algorithms for the identification of network motifs [51] [27] [52] [27]. These algorithms search for the relevant subgraphs of n nodes by comparing the appearance of these subgraphs in random networks and selecting only patterns that appear in real networks in numbers that are significantly higher. The identification of large network motifs remains a computationally hard problem.

Other models of graph evolution

Several other algorithms for evolving networks have been developed in parallel with our studies. In [53], Venkatasubramanian et al used a genetic algorithm to grow a network of variable number of nodes and edges using a fitness function that maximizes efficiency, robustness and gives a cost to the

addition of new edges and nodes. They showed that selection of efficient and robust networks leads to certain patterns which they termed as star, ring and hub. In [54] a genetic algorithm is used to evolve a network for resilience against random failures and the resulted topologies reveal a high level of clustering, modularity and long path lengths.

Another study [55] used a neural network of boolean components and used processes of random mutation and selection to train and evolve its topology towards a target fitness function.

1.2 Bacterial colonies as a communication network

We will present the bacterial colonies as an interspecies and intraspecies communication network, releasing and consuming various signals and external substances to modulate their metabolism. We will present the basic morphology and behavior of a bacterial colony, describe the phenomenon of quorum sensing and make an inventory of the *in silico* models available so far for simulating the behavior of a bacterial colony.

1.2.1 Bacterial colonies

Bacteria [56] are unicellular organisms, typically a few micrometers in size. They can survive in extreme conditions, and form much of the world biomass. Most bacteria are either spherical or rod shape (called *bacilli*). Many bacterial species are curved, spiraled, coiled or even rod-shaped. Their morphological diversity is dictated by the need to adapt their environment and extract nutrients with ease, and also by the need to sustain a thick cell wall and *cytoskeleton* which can give bacteria advantages in nutrient-poor environments.
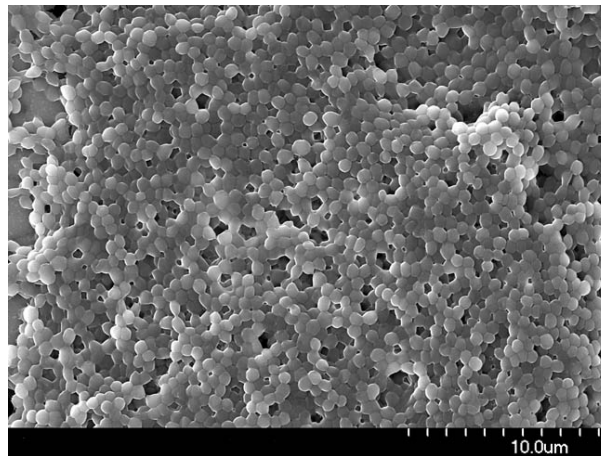
Figure 9 Image of a bacterial colony. From www.bacteria-world.com.

Some bacteria exist as simple cells, other associate in characteristic *colonies*: *Neisseria* form pairs, *Streptococcus* form chains, and *Staphylococcus* form clusters, *Actinobacteria* form filaments. Bacteria often attach to surfaces and form dense aggregations called *biofilms* ranging in thickness from a few micrometers to up to half a meter, and may contain multiple species of bacteria. Bacteria living in biofilms display a complex arrangement of cells and extracellular components, with networks of channels to enable better diffusion of nutrients and specialized levels of metabolic activity.
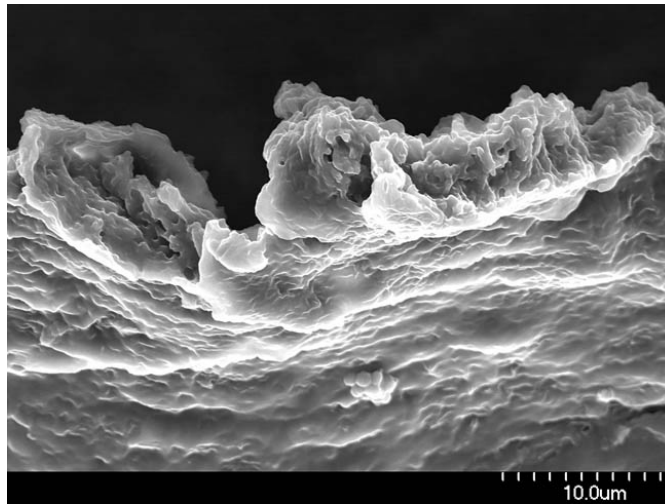
Figure 10 Scanning Electron Microscopy micrograph of a biofilm. *The biofilm was created by a bacterial colony which formed on a silicon device. A small colony of spherical Cocci type bacteria can be seen forming on the surface of this well developed biofilm. The biofilm is an adhesive matrix excreted by the colony of bacteria living on the device. The film is a polymeric compound called EPS (Extracellular Polymeric Substance). At the edge of the biofilm, one can see spherical Cocci type bacteria. Some researchers suggest that a polymeric matrix such as this one can play a role in the antibiotic resistant nature of some bacterial infections. From www.bacteria-world.com.*

In bacteria, cell division is directly linked to the increase in size of bacteria. The time required for a normal population of bacteria to divide through a process called *binary fission* until it doubles its size is several minutes, under normal growth conditions. The growth of bacterial colonies is usually studied in laboratory using solid or liquid media. Solid growth media such as *agar plates* are used to isolate a specific bacterial strain. Liquid growth mediums are used when large volumes of cells are required.

The *agar plate* is usually a *Petri dish* with a gelatinous substance derived from seaweed, called *agar*, used as a solid nutrient substrate for bacteria culture. Individual organisms are usually isolated and placed in such a plate to grow into bigger colonies, each individual being a clone of the initial organism (excepting a few naturally occurring mutations).
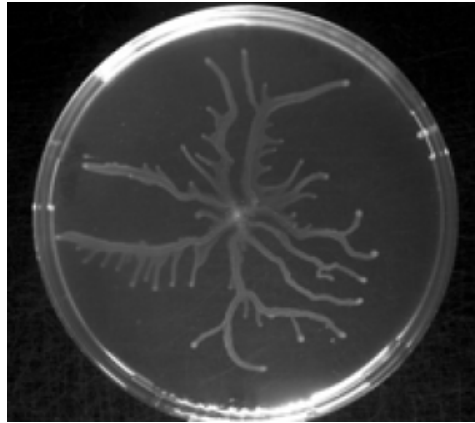
Figure 11 Picture of Pseudomonas aeruginosa bacteria swarming colony on an agar plate.

Natural growth environments usually differ very much form the laboratory mediums, having a limited quantity of nutrients and adverse life conditions. This made bacteria adapt in several ways. Some grow rapidly when nutrient is available; others inhibit their growth when nutrients are scarce. The natural occurring biofilm state usually offers increased nutrient availability and protection from environment stress.

The process of bacterial growth [57] usually has three phases. First is a *lag phase* in which the bacteria adapt their environment and are preparing the internal and external environment required for fast growth. This phase is characterized by high metabolic activity as the protein necessary for rapid growth are synthesized. The second phase is known as *logarithmic or exponential*, the growth rate is highest, and the time required for the population to double is known as the generation time. When one of the necessary nutrients was exhausted form the medium the development of the colony enters a third phase of *stationary limited growth*. In this last stage bacteria are usually characterized by stress response and an increase in the expression of genes involved in DNA repair, nutrient transport and antioxidant metabolism.

Pseudomonas aeruginosa

Pseudomonas is a common bacterium found in soil, water, and most man-made environments. It can thrive in environments deprived of oxygen, it uses a wide range of organic material for food; in animals this versatility enables the organism to infect damaged tissues or people with reduced immunity. If such colonization does occur in critical body organs such as the lungs, the urinary tract, and kidneys, the results can be fatal.

The title of opportunistic pathogens that *P. aeruginosa* have is supported by the heavy use of quorum sensing to coordinate swarming motility, exopolysaccharide production, cell aggregation and ultimately biofilm formation. The bacteria usually grow within a host until a critical concentration without harming it; after reaching the critical concentration threshold the bacteria become aggressive forming biofilm and overcoming the host immune system [58]. The *quorum sensing* process is what drives this aggressive behavior and it is hoped that by enzymatic degradation of the signaling molecules this process can be inhibited, phenomenon named *quorum quenching* [59] .

1.2.2 Quorum sensing

The term "*quorum sensing*" [60] describes one particular type of intercellular communication that takes place when bacteria are present at high cell population densities. Bacteria can achieve quorum sensing thanks to the synthesis, release, detection and response to threshold concentrations of signal molecules, originally called *autoinducers*. These molecules can be of different chemical nature and can be found among both gram-positive and gram-negative group of bacteria.

Quorum sensing in gram-negative bacteria: the model system of Vibrio fischeri.

The most intensely studied quorum sensing system in gram-negative bacteria is that of the bioluminescent marine bacterium *Vibrio fischeri*, which forms a symbiotic relationship with some marine fishes and squids. The best example of such a symbiosis is in *Euprymna scolopes*, where *V. fischeri* can reach extremely high densities in the light organ of this squid. The bioluminescent phenotype is exploited by the squid in order to perform a behavioral phenomenon called *counterillumination*, in which the squid camouflages itself from predators residing below. Bioluminescence in this bacterium is controlled by the quorum sensing system, which is composed by two regulatory genes, *luxI* and *luxR*, coding for *LuxI* and *LuxR*, respectively. *LuxI* is the

autoinducer synthase responsible for the production of the signal molecule N- (3-oxohexanoyl)-homoserine lactone (*C6 3-oxo HSL*), and *LuxR* is the regulatory protein that binds the signal molecule and activate the transcription of the light production operon. This operon is composed by the *luxCDABEG* genes, where the *luxAB* genes encode the enzyme *luciferase* and *luxCDE* encode proteins required for biosynthesis of the aldehyde substrate, used by *luciferase*. *LuxG* encodes for a probable flavin reductase, and is followed by a transcriptional termination site; however its function is still unknown [61]. When threshold concentration of the freely diffusible signal molecule *C6-3-oxo HSL* is reached, the *LuxR* protein binds the signal molecule and interacts with the lux-box promoter sequence, thus activating the transcription of the downstream operon. This results in higher transcription (many thousand-fold) of *luxI* and light production genes. At the same time, there is a reduction of the rate of transcription of *luxR*, dependent upon the presence of a lux box type element located within *luxD* [62] , as a way of compensation of this positive feedback.
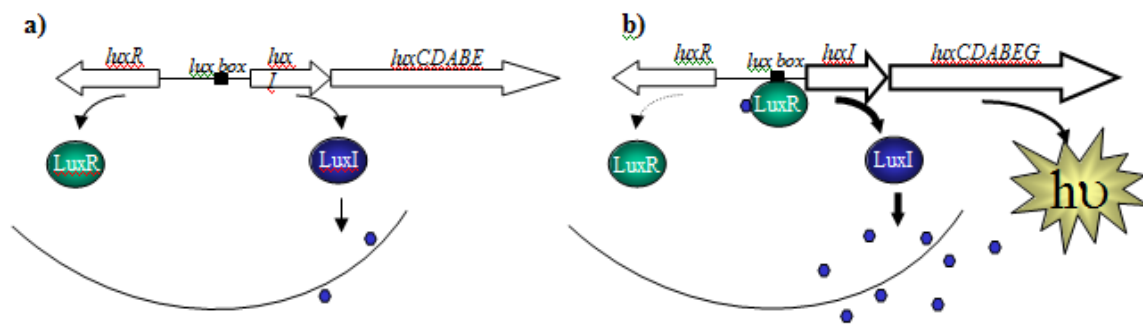
Figure 12 The Vibrio fischeri LuxIR quorum sensing system. *a) The system is not active and there is basal transcription of luxR, luxI and luxCDABE. While the cell density increases, the freely diffusible signal molecule C6-3-oxo HSL accumulates until it reaches a threshold concentration that allows its interaction with LuxR. b) LuxR is bound to C6 3-oxo HSL, thus driving the transcription of the downstream operon while it decreases the rate of transcription of luxR. Picture adapted from [61].*

Quorum sensing in gram-negative bacteria: the model system of *Pseudomonas aeruginosa*.

In *Pseudomonas aeruginosa*, as in most other gram negative bacteria, the agent for cell communication is a small diffusible molecule called *N-acylated homoserine lactone (acyl-HSL)*. These signals are produced by the *LuxI* type signal synthases and accumulate as the population

density increases. At reaching a threshold concentration they will bind to *LuxR* receptors that will activate the expression of different genes.

In *Pseudomonas aeruginosa* there are two signaling systems using acyl-HSL [63], called *las* and *rhl*. The las system contains the signal synthasse *LasI* producing *N-3-oxo-dodecanoyl-homoserine lactone (3OC12-HSL)* and the signal receptor *LasR*. The second system called *rhl* consists of the signal synthase *RhlI*, generating *N-butanoyl-homoserine lactone (C4-HSL)*, and the signal receptor *RhlR*, inducing gene expression when complexed with *C4-HSL. LasR* and *RhlR* also induce the transcription of their cognate synthase genes, thus a positive feedback loop is created. The two quorum sensing systems are arranged in a hierarchical fashion as the *LasR–LasI* system activates the *RhlR–RhlI* system. Specifically, *LasR-3OC12-HSL* activates transcription of *rhlR* and *rhlI*. The genes responsible either for the synthesis (*lasI, rhlI*) or the sensing (*lasR, rhlR*) of AHL signals are important in our studies, as they will be inactivated both experimentally and computationally to test several hypotheses regarding quorum sensing.

Genome-scale studies reveal that *Pseudomonas aeruginosa* quorum sensing system is regulatory for many different cellular functions. Studies showed that several quorum-controlled genes exhibit a delayed response to exogenously added acyl-HSL signals until the stationary phase of growth. These observations suggested that the activation of most quorum-controlled genes is not solely triggered by the accumulation of signal, and seems to require additional factors. In accordance with this hypothesis, several regulatory systems have been identified that affect the las or rhl mediated quorum system of *Pseudomonas aeruginosa*.

The regulatory network of the quorum system has two distinct levels within high signal integration potential. *LasR* represents a central checkpoint with the highest degree of interconnection in the network. The integration of several signaling pathways at this level makes sense given the fact that quorum sensing regulates hundreds of genes, thus a quorum switch is a big commitment for the bacterial cell that should be tightly controlled, not unlike a terminal differentiation process such as bacterial spore formation. A second level of signal integration appears to occur at the promoters of many quorum sensing target genes. The *hcnABC* operon for example, has a complex promoter architecture, being directly regulated by *LasR*, *RhlR*, *ANR*, and *GacAS/RsmAZ* [64] .

These quorum sensing target genes, along with the factors that directly control them, resemble a motif that is commonly found in transcription networks, termed a dense overlapping *regulon* [65].

This topology of multiple signaling pathways regulating overlapping sets of genes allows for specific responses to various growth conditions, which can help explain the exceptional environmental versatility and adaptability of *Pseudomonas aeruginosa*. Several groups have developed mathematical approaches to model the effects of pharmacological intervention on quorum sensing. These models usually output the concentration of active, signal-bound receptors, which allows important conclusions about the efficacy of targeting the signal-receptor interaction for quorum inhibition [63].
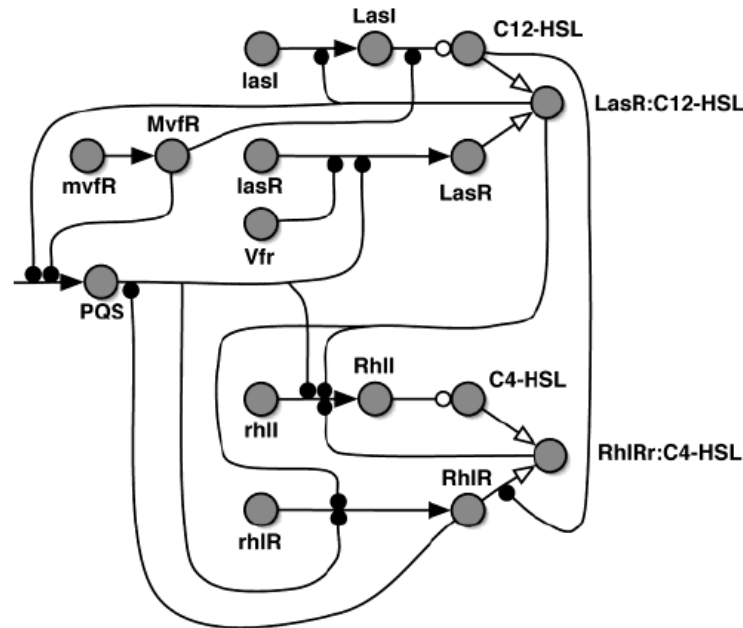
Figure 13 Schematic representation of P. aeruginosa quorum sensing network. Each type of process is indicated by different arrows: transcription and translation: filled arrows, complex formation: open arrows, modulation: filled circles, enzymatic synthesis: open circles. [63]

Several research groups performed a system analysis of a prototypical QS network, using experimentally proven kinetic rates for the transcription of the various factors involved in the system and a minimal regulatory kinetic model [66] [63]. Using this model they investigated how the positive feedback loops effect the switch-like behavior of the network, as well as the stability of its on and off states under molecular noise.

Swarming in Pseudomonas

An important phenotype regulated by AHL QS in *P. aeruginosa* is the swarming motility, which is a community phenomenon involving the fast movement of a bacterial population on a semi-solid,

28

viscous surface. Swarming by *P. aeruginosa* is characterized by a dendritic community appearance and has been shown to require the expression of several loci including flagella, pili, and rhamnolipid-encoding genes. AHL QS is pivotal for the swarming of *P. aeruginosa*, as it is involved in the regulation of many of the genes required for this community behavior.

Research suggested the importance of type IV retractable pili for the movement of Pseudomonas. The thin filaments composing them are essential for cell twitching motility [67]. An infecting phage was given to a variety of Gram-negative bacteria and studies of electron microscopy have revealed how type IV pili generate movement. The infecting phage is pulled to the cell surface where productive infection occur. It was demonstrated that Type IV pili can retract with considerable force [68]. Non-motile mutants were pili-less or had non-retractile pili; thus pilus retraction powers twitching motility. The pili were not observed directly; instead pilus function was assayed by monitoring the position of a latex bead attached to pili with antibodies. Later Skerker & Berg, (2000) have developed a technique to label pili using an amino-reactive Cy3 and observed directly pilus extension, pilus retraction, and retraction-mediated cell movement.

Studies of electron microscopy have shown that cells with mutation to the las signaling system have diminished swarming behavior, while the mutants defective in rhl are completely unable to swarm. Evidence suggests that a variety of rhamnolipids are the key surfactant involved in the motility of swarming [69].

Cell-cell communication molecules biosynthesis

A general mechanism proposed for HSLs biosynthesis is depicted in *Figure 14*, in which a *HSL synthase* uses the substrates *S-adenosyl-L-methionine* (SAM) and *acylated acyl* carrier protein (acyl-ACP) in a proposed "bi-ter" sequentially ordered reaction [70]. In this reaction, the acyl chain is presented to the *HSL synthase* as a thioester of the ACP phosphopantetheine prosthetic group, which results in nucleophilic attack on the 1-carbonyl carbon by the amine of SAM in the acylation reaction. Lactonization occurs by nucleophilic attack on the 1-carbonyl carbon of SAM by its own carboxylate oxygen to produce the HSL product. The *N-acylation reaction*, involving an enzyme-acyl-SAM intermediate, is thought to occur first since *butyryl-SAM* acts as both a substrate and as an inhibitor for the P. aeruginosa HSL synthase, RhlI, to produce *C4-HSL*. A unique aspect of the HSL synthesis mechanism is that the substrates adopt roles that differ quite dramatically from their normal cellular functions. SAM usually acts as a methyl donor, whereas acyl-ACPs are components of the fatty acid

biosynthetic pathway and had not been implicated in cell-cell communication until their discovery as acyl chain donors in HSL synthesis.
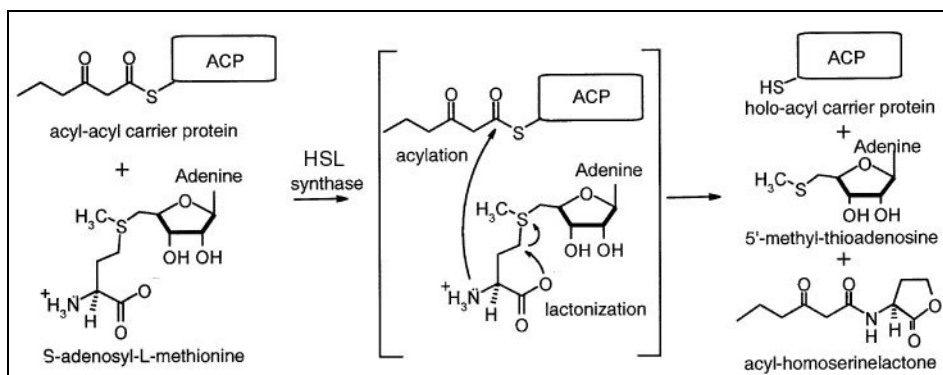
Figure 14 General features of the HSL biosynthesis reaction. Two substrates: acyl-ACP and SAM, bind to the enzyme. After the acylation and lactonization reactions, the product HSL and byproducts holo-ACP and 5'-methylthioadenosine are released. Picture adapted from [71].

Non-HSL cell-cell communication molecules

In addition to HSL, other alternative gram-negative cell density dependent signaling molecules can be found among gram-negative bacteria. Some examples are in *Ralstonia solanacearum*, a phytopathogen that produces *3-hydroxypalmitic acid methyl ester* as a novel signaling molecule together with *C8-HSL* [72]; *Xanthomonas campestrs*, a cabbage pathogen that produces a diffusible signal factor (DSF) of which structure seems not to be a HSL and has yet to be fully characterized; *Pseudomonas aeruginosa*, which produces a *2-heptyl-3-hydroxy-4-quinolone* together with *C4-HSL* and *C12-3-oxo-HSL* [73]. Moreover, butyrolactones have been isolated from *Pseudomonas aureofaciens* cell-free culture supernatants, and a novel family of signaling compounds identified as *diketopiperazines* (DKPs) were isolated from cell-free culture supernatants of *Pseudomonas putida, Pseudomonas aeruginosa, Pseudomonas fluorescens, Pseudomonas alcaligenes, Enterobacter agglomerans, Proteus mirabilis and Citrobacter freundi*. Although these non-HSL molecules can modulate the activity of a number of LuxR-based quorum sensing systems, the concentrations required are much greater than those of the cognate HSL are. On other hand, some of the DKPs act antagonistically, reducing C6-3-oxo-HSL-mediated bioluminescence and suggesting that they may be able to compete for LuxR binding [74].

Interestingly, *AI-2* (a molecule structurally different from acyl homoserine lactone) synthesised by *Vibrio harvey*, has been proposed as a universal signaling molecule, which would facilitate inter-

30

species communication [75]. The structure of AI-2 has been shown to be a furanosyl borate diester. LuxS, the protein that drives the synthesis of AI-2, has been showed to exist in over 30 species of both gram-negative and gram-positive bacteria, thus defining a new family of proteins involved in cell-cell communication molecules production.

Different phenotypes are modulated by Quorum sensing

One single bacterium cannot evade the barriers that the colonization of a target organism represents. Quorum sensing is the genetic tool that bacteria use to coordinately take advantage of the number of individuals in order to establish such efficient colonization. The benefits of a unified response may include improved access to complex nutrients or environmental niches, collective defense against other competitive micro-organisms or eukaryotic host defense mechanisms and optimization of population survival by differentiation into morphological forms better adapted to a given environmental threat [76].

As mentioned above, one of the first and best-described phenotypes to be controlled by quorum sensing is bioluminescence in the marine bacterium *V. fischeri*. Since then, the fact that expression of certain genes is correlated to high cell density has become more the rule than a curious and isolated phenomenon performed by an esoteric light-emitting marine bacterial species. In fact, the list of organisms for which a quorum sensing system has been identified is in constant growth. It is now evident that diverse gram-negative bacteria produce HSLs. Quorum sensing modulates a variety of physiological processes, including bioluminescence, swarming, swimming and twitching motility, antibiotic biosynthesis, biofilm differentiation and conjugation. In *Table 1*, some example organisms are described by their quorum sensing genetic elements and the phenotypes modulated by them, where known. As can be observed, some opportunistic human pathogens such as Pseudomonas aeruginosa, Aeromonas hydrophila or Chromobacterium violaceum produce HSLs that are used to modulate the production of virulence determinants as exoenzymes or proteases. These are capable of contributing to virulence by causing tissue damage. Interestingly, obligate human pathogens such as Haemophilus influenzae, Neisseria meningitidis, and Neisseria gonorrhoeae do not seem to make HSLs.

Table 1 *Example of microorganisms, their respective quorum sensing systems, and the phenotypes modulated in each case, where known. Table used with permission of V. Venturi [77]*

BACTERIUM	LUX I/R HOMOLOGUES	MAJOR AHL	MODULATED PHENOTYPE	REFERENCE
Aeromonas hydrophila	AhyR, AhyI	C4-HSL	Extracellular protease, biofilm formation	(Swift *et al*., 1999) (Lynch *et al*., 2002)
Aeromonas salmonicida	AsaR, AsaI	C4-HSL	Extracellular protease	(Swift *et al*., 1997)
Agrobacterium tumefaciens	TraR, TraI	3-oxo-C8-HSL	Conjugation	(Fuqua & Winans, 1994; Piper *et al*., 1999)
Chromobacterium violaceum	CviR, CviI	C6-HSL	Antibiotics, violacein, exoenzymes, cyanide	(Chernin *et al*., 1998; McClean *et al*., 1997)
Erwinia carotovora	CarR, ExpR CarI (or ExpI)	3-oxo-C6-HSL	Carbapenem antibiotic, exoenzymes, exopolysaccharide	(Andersson *et al*., 2000; Bainton *et al*., 1992; Pirhonen *et al*., 1993; Swift *et al*., 1993)
Escherichia coli	SdiA	Unknown	Cell division	(Sitnikov *et al*., 1996)
Pantoea stewartii	EsaR, EsaI	3-oxo-C6-HSL	Exopolysaccharide	(Beck von Bodman & Farrand, 1995)
Pseudomonas aeruginosa	LasR, LasI	3-oxo-C12-HSL	Exoenzymes, Xcp, biofilm formation, RhlR, cell-cell spacing.	(Chapon-Herve *et al*., 1997; Gambello & Iglewski, 1991; Glessner *et al*., 1999; Passador *et al*., 1993)

	RhlR, RhlI (VsmR, VsmI)	C4-HSL	Exoenzymes, cyanide, RpoS, lectins, pyocyanin, rhamnolipid, type 4 pili.	(Glessner, *et al*., 1999; Latifi *et al*., 1996; Latifi *et al*., 1995; Pearson *et al*., 1997; Winson *et al*., 1995).
Pseudomonas aureofaciens	PhzR, PhzI	C6-HSL	Phenazine antibiotic	(Pierson *et al*., 1994; Wood *et al*., 1997)
Pseudomonas fluorescens	PhzR, PhzI HdtS	C6-HSL; C10-HSL; 3OH-7-*cis*-C14-HSL	Phenazine antibiotic	(Laue, *et al*., 2000; Shaw *et al*., 1997)
Ralstonia solanacearum	SolR, SolI	C8-HSL	Unknown	(Flavier *et al*., 1997b)
Rhizobium leguminosarum	RhiR	3OH-7-*cis*-C14-HSL	Nodulation, bacteriocin, stationary phase survival	(Rodelas *et al*., 1999; Thorne & Williams, 1999)
Rhodobacter sphaeroides	CerR, CerI	7-*cis*-C14-HSL	Community escape	(Puskas *et al*., 1997)
Serratia liquefaciens	SwrR, SwrI	C4-HSL	Swarming, protease	(Eberl *et al*., 1996; Givskov *et al*., 1997; Lindum *et al*., 1998)
Vibrio anguillarum	VanR, VanI	3-oxo-C10-HSL	Unknown	(Milton *et al*., 1997)
	VanN, VanT	3-OH-C6-HSL; C6-HSL	EmpA expression, pigment production, and biofilm formation.	(Croxatto *et al*., 2002)
Vibrio fischeri	LuxR, LuxI	3-oxo-C6-HSL	Bioluminescence	(Engebrecht & Silverman, 1987)

The stationary phase sigma factor RpoS.

Another very effective mechanism employed by bacteria to bring about a major switch in gene expression at high cell densities, often resulting to a stationary phase of growth, is the use of the RpoS alternative sigma factors that alters RNA polymerase core specificity. The RNA polymerase holoenzyme is composed of the core enzyme, with the catalytic activity of RNA polymerization, and one of the several different species of sigma (factors present in bacteria, that bind reversibly to the core. Each sigma plays a determining role in the specificity of transcription initiation, being released in the elongation of the RNA chain [78].

The fact that sometimes the high cell density in a bacterial population turns out to be the inducing signal for the stationary phase-regulated genes is enough to think of a possible correlation between RpoS and quorum sensing. Nevertheless, although there is some evidence that highlights the importance of RpoS in quorum sensing and vice-versa, there is conflicting data regarding this reciprocal regulation. In studies performed in R. solanacearum, indirect evidence show that RpoS regulates the HSL production by the negative modulation of solR and solI expression, and that rpoS is not regulated by the quorum sensing system [79].

On other hand, in P. aeruginosa quorum sensing was shown to modulate the expression of rpoS [80]. However, a recent study demonstrated that is actually RpoS that regulates rhlI, in other words, quorum sensing [81].

Alternative hypotheses to quorum sensing

The cooperative behavior of the bacteria involved in common sensing is being challenged due to its evolutionary game dynamics implications. The hypothesis of quorum sensing assumes that bacteria respond to high autoinducer concentration related to the cell density in order to make a coordinated group response once the cell density is big enough. It is suggested that the evolutionary benefit consist in maximizing the group fitness. Given the fact that many non-producing or over-producing cheaters are able to thrive in their environments, the dynamics of bacterial colonies contradicts the mathematical laws governing these types of systems, and an alternative explanation to cooperative behavior has been sought, called *diffusion sensing*.

Diffusion sensing [82] is based on the observation that the 'decision' of the bacteria to alter behavior when a quorum has been reached is not based on perfect information. The autoinducer concentration that would normally function as an estimate of cell density is altered by many factors, including diffusion and advection, spatial distribution, degradation and the production of the same autoinducer by third parties, whether intentionally or by chance. According to diffusion sensing, the function of secreted autoinducers is to determine whether secreted factors would rapidly diffuse away from the cell, thereby allowing bacteria to detect situations in which the disappearance of effectors happens because of a low diffusivity which makes the release of factors efficient. The evolutionary benefit consists in individual fitness maximization as opposed to the altruistic group benefit found in quorum sensing.

Quorum sensing and diffusion sensing were later viewed as extreme cases of *efficiency sensing* [83]. Efficiency sensing is based on the idea that the cells use autoinducers to measure a combination of cell density, spatial distribution and mass transfer limitations due to diffusion and advection. The purpose of releasing autoinducers is testing the efficiency of releasing more costly diffusible factors. What is relevant for the fitness of a cell is not cell density but the concentration of autoinducer and therefore of factors that is available in the immediate vicinity.

The adaptive implication of quorum sensing as a promoter of swarming and increased bacterial communication

One of the main results of the process of quorum sensing is an increase in the metabolic activity of the cells, which translates into a changing pattern of movement for large groups of bacteria. This physical process of synchronized motion due to collective or individual forces is called swarming. While swarming, bacteria form veritable communication networks based on cell signaling. The adaptive power of such network is apparent in experiments where a change in external conditions favors rapid adaptive mutation in bacteria. The fast response to selective pressure is suggesting that the colony behaves as a network, not just as randomly mutating bacteria.

The way the phenomenon of swarming and adaptive communication among bacteria is usually analyzed is by using the theory of evolutionary game dynamics [84] [85] [86]. Evolutionary game dynamics deals with the drift of independent strategies in populations. It is originating in the mathematical game theory and was developed by John Maynard Smith. The initial observation that leads to the development of evolutionary game dynamics is that frequency dependent fitness introduces a game strategy aspect to evolution. The evolutionary stable strategy, once adopted by a

population of players cannot be invaded by an alternative strategy that is initially rare [87]. In game theory the typical solution of game is called Nash equilibrium. Players are aware of the structure of the game and attempt to maximize their payoff by predicting the moves of the opponents. It is assumed that players have knowledge of their fellow player movements. In evolutionary game dynamics the equilibrium strategies are formulated differently. The players have biologically encoded and heritable strategies. Individuals have no conscious control over the play and cannot predict the movements of their opponents. Individuals reproduce according to the laws of natural selection and the payoffs are in the form of biological fitness. Via mutations, alternative strategies occur, and in order for a strategy to be called evolutionary stable, it must present resistance to such mutations.

Evolutionary game dynamics considerations for bacterial cooperation

From an evolutionary game perspective, bacteria can be classified as altruists or cheaters. Sometimes the signal sender is favored when pretending that the population density is higher since this way it can manipulate the receptor into producing costly factors in the absence of quorum or before the quorum has been reached. In such signaling systems the evolutionary problem of honesty is raised. The recipient has no means of punishing the sender for the quality of the information received. It is good to view the cooperative signal production and cooperative signal response as two separate traits that need to maintain equilibrium in order for the system not to collapse. Signal – deficient and response – deficient engineered mutants are both considered cheaters, but under some assumptions they can thrive not because they cheat but because they are more adapted to the environment. For example the signal – blind lasR mutant of Pseudomonas aeruginosa is more resistant to cell lysis and cell death than the wild type in stationary phase [88].

Can network theory provide a different perspective on studying bacterial cooperation?

In this study we will try to offer a different method for the study of bacterial cooperation, by means of network theory. If two bacteria are close enough to signal to each other, we say the two are linked. Linking together bacteria based on a proximity threshold to the scale of the entire colony will form a graph of bacterial communication. Network theory does not assume *a priori* principles of bacterial ethics, there are not cheaters and altruists, while the dynamics of a colony stem from the continuously evolving network of communicating bacteria.

1.2.3 Bacterial colony models

Biofilm and agar plate models

Bacterial growth can be modeled with a variety of techniques, most of which are directed to colony morphology. The underlying methods fall into several broad categories. The first distinction is drawn between *biofilm models* and *agar plate models*. The biofilm models differ from the agar plate ones due to the fact that layers of bacteria can grow on each other vertically, thus nutrient diffusion becomes essential for survival and shapes the morphology and functioning of the entire colony. One such successful model for biofilm formation uses individual based modeling. In individual based modeling biofilm growth is due to the processes of diffusion, reaction and growth (including biomass growth, division and spreading). Each bacterium is a spherical cell in continuous space and has variable growth parameters. Spreading of biomass occurred by shoving of cells to minimize overlap between cells. The model was used to investigate the co-evolution of several bacterial species, investigating phenomena of cooperation and conflict [89].
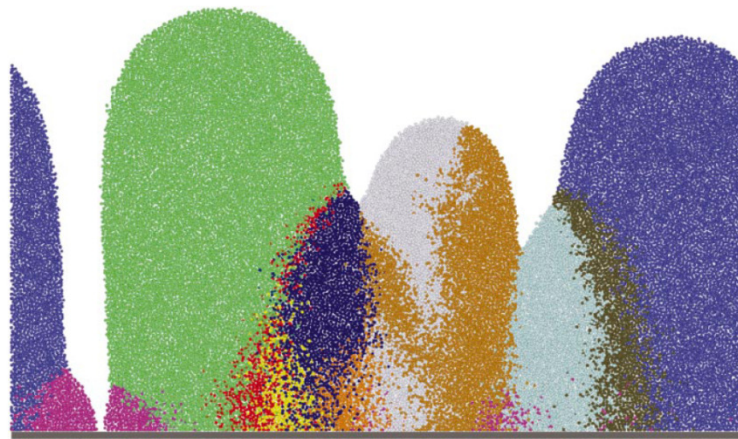
Figure 15 Growth of the biofilm modeled using an individual based model. Each bacterium is a spherical cell in continuous space and has variable growth parameters.[89]

Agar plate models are essentially two-dimensional, there are no vertical layers of bacteria and the expansion is limited by a combination of agar softness, nutrient availability and swarming phenomena. A considerable part of computational efforts has been devoted to the simulation of branched colony patterns, which various microbes are known to produce under harsh growth conditions [90]. Even though the formation of bacterial colonies is a complex process that requires a

variety of mechanisms (including flagella, pili, secretion of surfactants, siderophores, and enzymes), both continuum models and particle-based hybrid models can reproduce the fractal-like branched patterns characteristic of mature bacterial colonies.

The available models of bacterial colony fall into two main classes: continuous models and discrete models. Continuum models treat bacterial colonies as a continuous material that diffuses and expands in an environment of other continuous materials in a process described by reaction-diffusion equations [91] [92] [93] [94] .

We will present two such models, with the observation that most models approach the modeling of bacterial colonies using roughly the same procedure.

The continuous models of Kawasaki and Cohen.

We consider a system consisting of bacterial cells and nutrients in two dimensions. Both cells and nutrient undergo diffusion while cells proliferate by consuming the nutrient. Let us denote the population density of the cells at time t and spatial position p=(x,y) by b(p,t) and the concentration of the nutrient by n(p,t). Then b and n are in general governed by the following equations:

$$\frac{\partial n}{\partial t} = D_n \nabla^2 n - f(n,b)$$

$$\frac{\partial b}{\partial t} = \nabla \cdot \left\{ D_b \nabla^2 b \right\} + \theta f(n,b)$$

where Db and Dn are the diffusion coefficients of the bacterial cells and nutrient respectively. We assume that the diffusion coefficient of the nutrient D_n is constant while the diffusion coefficient D_b of the bacterial cells depends on both the bacterial density and the nutrient concentration. The term $f(n,b)$ represents the consumption rate of nutrient by the cells and $\theta f(n,b)$ the growth rate of bacteria, where θ is the conversion rate "growth yield" of consumed nutrient to bacterial growth. The equations above express the observation that growth rate of bacteria as well as the consumption rate is dependent on both the concentration of bacteria and nutrient. A simplified version of the above equations would be:

nutrient rate = diffusion - consumption

bacteria rate = movement + reproduction

When numerically solved, the equations will yield a slow fractal growth pattern at small values of nutrient and high values of agar concentration (corresponding to slow bacterial diffusivity).
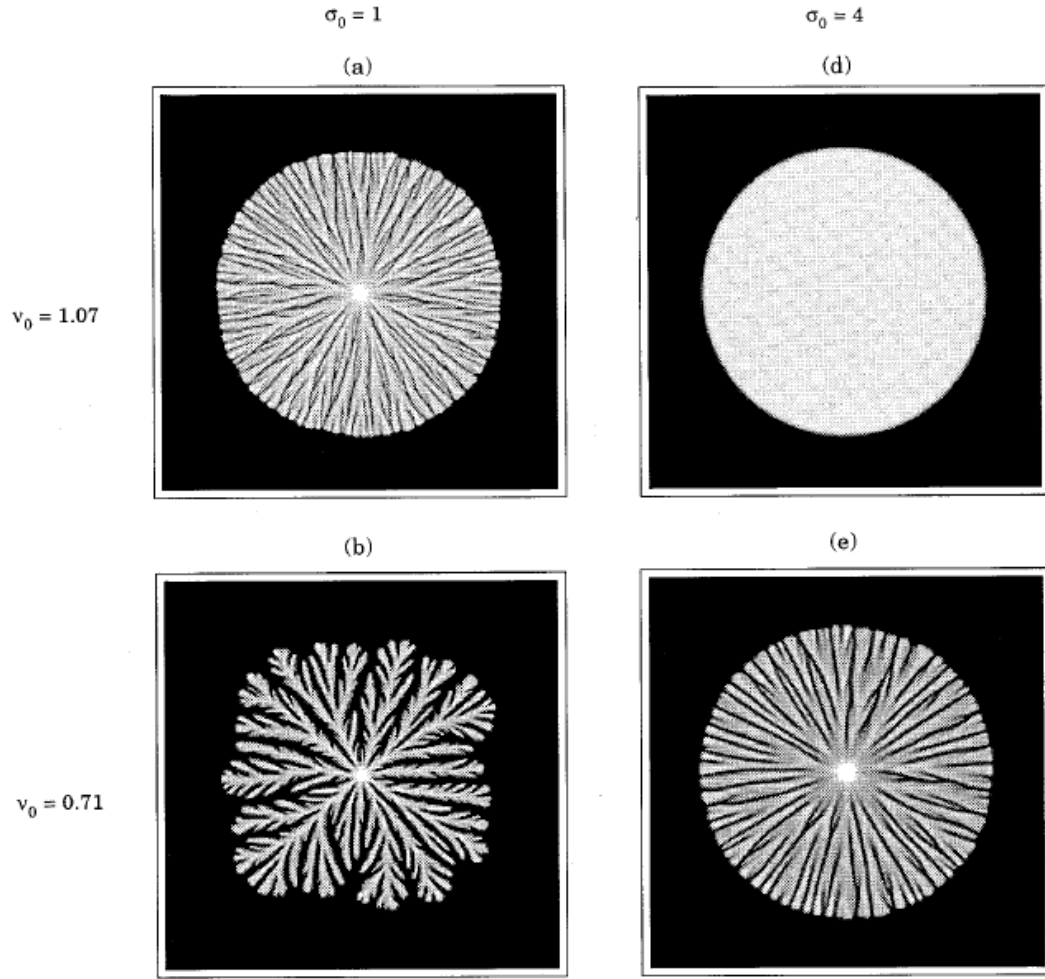
Figure 16 Morphological changes of colonies predicted from the diffusion-reaction model with varying nutrient concentration υ_0 and bacterial diffusivity σ_0. The time for the colony to reach the size as in figure is (a)396 (b)2828 (d)127 (e)566. Thus fractal colonies form slowly and are a result of diffusion limited growth [95].

Another continuous model developed by Cohen et al, assumes four coupled fields [96] [93]. One field describes the bacterial density, the second describe the height of lubrication layer in which the bacteria swim, third field describes the nutrients and the fourth field is the stationary bacteria that "freeze" and begin to sporulate. Additional to this the diffusion coefficient is made nonlinear and proportional to the bacterial density and 3 types of chemotaxis fields are added: food chemotaxis, a short range repulsive chemotaxis and a long range attractive chemotaxis.

Figure 17 Phase diagram of pattern changes in the colony of B. subtilis as a function of the concentration of nutrient Cn and the solidity of agar medium expressed as 1/Ca where Ca is the agar concentration. One can see that branched patterns are achi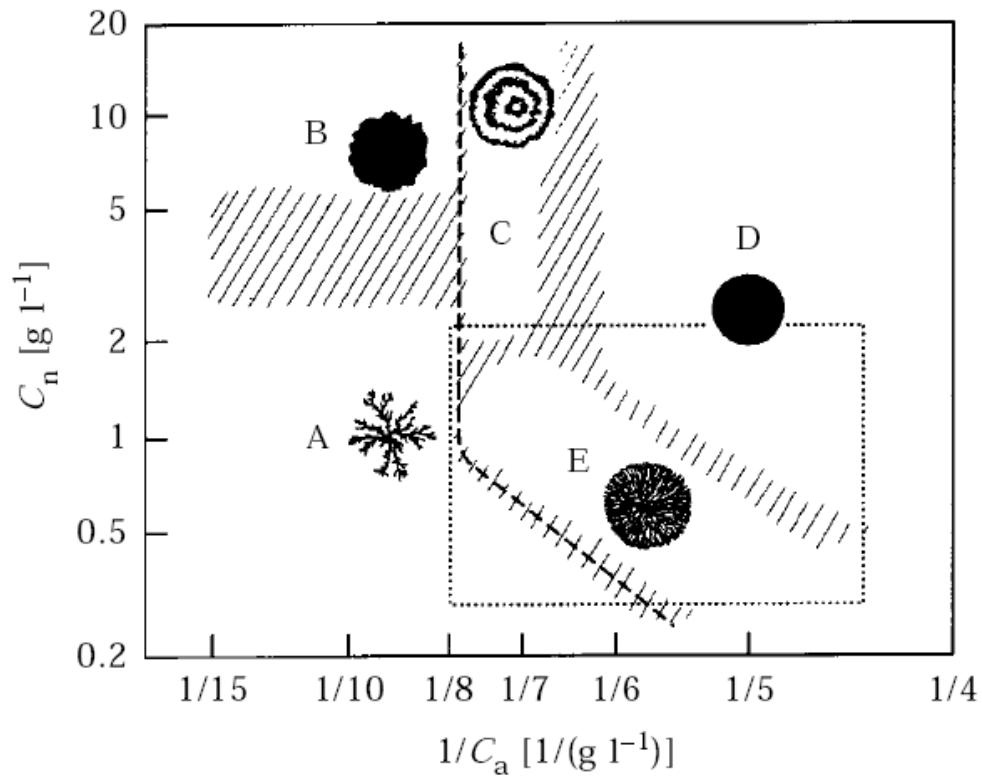eved in regions of high agar concentration and low nutrient. As one can see the development of a branched colony is a typical diffusion limited phenomenon. If other conditions become limiting, for instance by using a higher nutrient concentration, or by lowering the agar level, the cell colony tends towards a circular shape. Other processes, like swarming and chemotaxis also affect the overall shape of a bacterial colony [95].

Hybrid models

Hybrid models use a continuum description for the growth medium as well as for the solutes, and individual descriptions of bacteria. One of the best known models is based on cell clusters (groups of cells consisting of up to 10^4 cells) that have their own rules for division, growth, and interaction, and orient their movement according to various concentration gradients (nutrients, chemotactic signals, and so on) within the medium. In a recent study, Gerlee and Anderson presented yet a different kind of hybrid model in which individual cells were represented by cellular automata fixed to regularly spaced locations representing the culture medium [97].

The usual way bacteria are modeled using the hybrid method models is based on autonomous agents. Individual bacteria or sometimes groups of bacteria are moving independently according to simple principles. These rules of movement can be diverse, from a simple random motion to general laws of attraction, repulsion and alignment. The result of such a motion of discrete autonomous agents termed *bots* [98] is called swarming and it is characterized by the emergence of clear movement patterns. These type of swarming principles can be applied not only to bacteria but they were used to model the flight of birds, the synchronized swimming of fish schools, pedestrian traffic, large scale character animation and even in swarming intelligence algorithms [99] [100] [90].

Figure 18 A swarm of autonomous agents following Reynolds general rules of attraction, repulsion and alignment. [98]

Bacteria approximated as self-propelled autonomous agents abide the general rules that govern any swarm of particles of their type [101], The patterns of their movement can be explained using the laws of hydrodynamics [102]. Several phenomena can be explained using basic laws of hydrodynamics and the dynamics of swarms, like a state transition to turbulent vortexes observed in some swarming bacteria or the formation of chiral patterns [103] in chemotaxis regulated colonies.

41
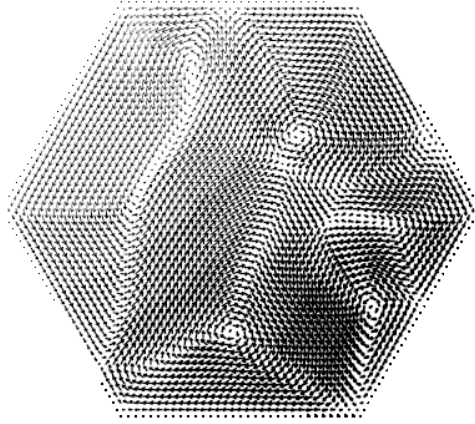
Figure 19 The emergence of turbulent motion in a colony of bacteria simulated solving numerically the hydrodynamic equations of a system of self propelled particles approximated as rods sliding on a 2D surface.[103]

The vortexes formed by swarming bacteria were used by Cohen et. al to justify another colony model termed *communicating walker*, in which vortexes of swarming bacteria are approximated as swarming agents [104]. The model uses random walkers (agents) that represent aggregates of bacteria moving in response to gradients in nutrient concentration and communicating with each other by means of chemotactic 'feedback'. The use of chemotaxis changes the pattern of the dendritic arms of the colony, making them denser.
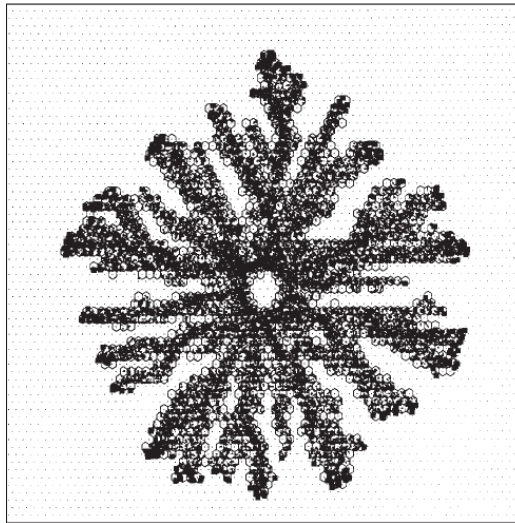
Figure 20 Image of the colony modeled by Cohen's communicating walker method at low nutrient concentrations. Bacteria expand in a certain direction if the direction is visited for a number of times walkers.[93]

2. Aims of this study

In our work we wanted to investigate the effect of rewiring on the global communication fitness of a network. Our first goal was to study the evolution of robust yet efficient network topologies and to see if selecting mutations only for efficiency or only for attack tolerance (robustness) will influence network topology. We also wanted to study how efficient and robust biological network behave, and if multiple attack has an outcome on the overall topology. We wanted to model the onset of swarming in *Pseudomonas aeruginosa* with a simplified agent-based model that could allow us to study the properties of the emergent behavior of the colony. We wanted to predict the experimental behavior of genomic knockout mutants in which the QS genes responsible either for the synthesis (*lasI, rhlI*) or the sensing (*lasR, rhlR*) of AHL signals were inactivated. We next wanted to study the interaction of bacteria by modeling their spatial dynamics as an evolving graph of interacting bacterial agents.

3. Materials and methods

3.1 Network evolution model

The network evolution models we propose are based on the rearrangement of links (''rewiring'') and follows the traditions of evolutionary modeling, i.e. optimizes a fitness function that combines various factors into one numerical index. Naturally, there are many ways to formulate and combine the components of the fitness function and testing the possibilities makes the process computationally expensive.

We describe two algorithmic approaches in which all parameters are treated essentially as constraints: a mutation is selected if all of its parameters exceed or at least reach the corresponding values of the previous state, so there are no tunable parameters. This approach is computationally efficient so it allows one to study a wide range of phenomena. The other approach is a genetic algorithm, that is used to discover fitness-optimized network structures by avoiding local maxima.

Our main goal was to study the evolution of robust yet efficient network topologies and to see if selecting mutations only for efficiency or only for attack tolerance (robustness) will influence network topology. We show that concomitant selection for efficiency and robustness influences the fundamental topological properties of the network, and that evolution under multiple attacks leads to distinct topologies [105].

As shown before, our previous studies on the attack tolerance of the sparse regulatory networks of *E. coli* and *S. cerevisiae* found that partially weakening a few central nodes has the same effect with knocking out the most central node. The natural questions that arise is how have the networks evolved mechanisms of protection to single and multiple attacks and what are the structural differences between a network that is evolving under single attack compared to the networks evolved under multiple attacks?

To answer these questions we developed two evolution strategies and investigated the outcome and dynamics of evolution starting from random graphs. In our algorithms we imposed several constraints. We use undirected graphs with no growth. Thus the evolving graphs have a specified number of nodes and edges. The graphs are sparse, the number of edges is only slightly higher than the number of nodes, and by this we mimic most biological and other naturally occurring networks. The first

method is a random evolution, in which a rewiring (mutation) is accepted only if it has the same or slightly higher fitness. This is a greedy method of evolution that will eventually trap the network structures in local minima (but as we will see further this method gives good approximation to optimal structures, suggesting that the fitness landscape funnels the solutions close to the optimal ones). A second method based on a genetic algorithm was used to develop optimal solutions, and we will describe the optimal graphs that we obtained, which we call k-rings.

3.1.1 Random selection algorithm

The random selection algorithm [105] is as follows:

1. Generate a random undirected non-weighted graph with n nodes and e edges as an input graph.
2. Measure the efficiency E and robustness R of the input graph.
3. Choose an edge with uniform probability and assign a different node to it.
4. Measure the efficiency E_t and robustness R_t of the resulting graph.
5. Compute the fitness of the new graph, $f(E_t, R_t)$. If $f(E_t, R_t) \geq f(E, R)$ accept the rewiring and assign to E and R the values of E_t and R_t respectively.
6. If there is no accepted rewiring for a big number of repeats stop, else go to step 3.

In the random selection algorithm a random graph of a specified number of nodes and edges is generated as input graph. The random rewiring process consists choosing an edge with uniform probability and linking it to a different end-node picked uniformly from all the network nodes. Only one of the adjacent nodes is rewired, this is to better simulate the evolution of biological networks. Simulations were done however also with both end-nodes rewired at the same time with the same results. Given the uniformity of choosing an edge for rewiring and the non-existence of additional costs for rewiring to distant nodes, the two alternatives are of similar outcome.
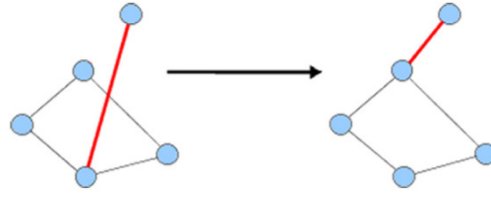
Figure 21 A simple model of network evolution. At each step an edge is rewired at random. If the new topology is more fitted according to the selected fitness criteria, the new network is used for the next step.

The choice of fitness function

The fitness function f is a function of the efficiency and robustness. Several fitness functions were used and their outcome varies in the speed and the capability of avoidance of getting the evolution process stuck into local maxima.

Neutral evolution fitness function:

We will say that $f(E_t, R_t) \geq f(E, R)$ if and only if $E_t \geq E$ and $R_t \geq R$.

Affine evolution fitness function

$f(E, R) = \alpha E + (1 - \alpha)R$, where $\alpha \in [0,1]$ is a real valued parameter.

Probabilistic evolution fitness function

Somewhat similar to the simulated annealing with variable temperature method of optimization, this fitness function is designed to better avoid trapping the topology in local maxima. However since there is no cost penalty for rewiring to distant nodes, this method cannot be classified as simulated annealing.

A rewiring will be accepted if and only if:

- $E_t \geq E$ and $R_t \geq R$

or,

- will be accepted with a probability proportional to:

46

$$\frac{(E - E_t)(R - R_t)t}{K}$$, where t is the number of timesteps (number of 3-6 cycles in the algorithm) since the last accepted rewiring and K is a proportionality constant tweaked to take into account the maximum number of steps allowed without rewiring and the maximal differences $E - E_t$ and $R - R_t$ over which a rewiring is not allowed anymore.

Distance based fitness function

This method computes E_t, R_t, and then computes the distance to an optimal point $P(E_{max}, R_{max})$:
$$f(E_t, R_t) = dist(P'(E_t, R_t), P(E_{max}, R_{max})).$$

This method allows the search space to better avoid local minima. The outcome of the evolution is dependent on choosing the convergence point P.

3.1.2 Genetic algorithm

The genetic algorithm [105] was used to discover optimized graph topologies. Its main outline is as follows:

1. Start with a population of random graphs encoded as a chromosome describing each graph topology.
2. A number of graphs ''individuals'' are selected for crossover and mutation. Crossover is done by exchanging a number of relevant subgraphs. The mutation means rewiring an edge at random. The crossover and mutation rates are set heuristically to insure better convergence.
3. A new generation is accepted if the graph passes the efficiency/robustness criteria as stated in the fitness function.
4. Repeat steps 2 and 3 until the appearance rate of the most fitted individuals significantly decreases.

Given the fact that the number of nodes and edges is fixed, the graphs were encoded as lists of pairs of nodes describing the graph's edges. A mutation means choosing an edge (n_1, n_2) with uniform probability across the list of edges, choosing a random node n_3, adding the new edge (n_1, n_3) to the beginning of the list and deleting the (n_1, n_2) pair from the list. Crossover is done by choosing a crossover point and exchanging the graph edges. The crossover cut is usually small to ensure that the graph population is diverse enough. Also, the mutation and crossover rate are slow, particularly the

mutation as mutation can dramatically change the structure of the graph. This is generally recommended for genetic algorithms; the slower the evolution works, the more optimized will be the result of the evolution.

Due to the fact that mutations are added to the beginning of the list, we can establish the evolutionary history of a graph, and it is also important for the mechanism of crossover. Thus crossover will not modify the graph structure too dramatically, while also allowing the crossover to spread a new successful mutation fast enough.
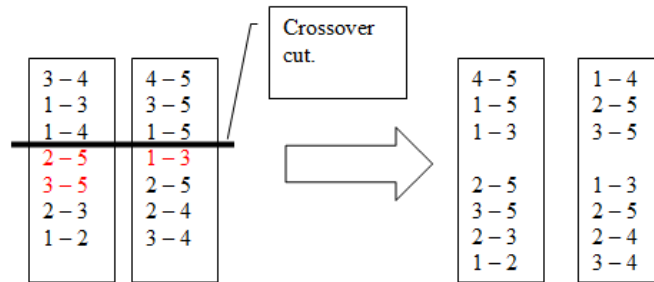

Figure 22 *Crossover example on a graph of 5 nodes and 7 edges. The edges marked by red under the crossover cut line included in the crossover as well because some of the edges above the cut already belong to the original graph.*

A list of best individuals is kept and a random individual is artificially added to the population if large periods of time with no selective success appear.

Convergence discussion, software and algorithmic details

Concerning the convergence of these two models, the random selection algorithm was found to lead to quite optimized structures and the speed of the simulation makes is useful for heavy computational purposes. Networks of 400 nodes and 480 links take a few hours to evolve to an optimized network structure on an average capacity desktop computer. The genetic algorithm is slow and tedious in adjusting the parameters. The simulation was running in both cases until for a specified amount of time there was no further evolution. The results of the simulations will be discussed separately.

The simulations were performed on a Linux cluster, using 10 independent processes at a time in the case of the random selection algorithm. The core program was written in C++ using the Boost Graph Library.

3.2 Bacterial strains and growth experiments

The first step towards modeling the colonization abilities of *P. aeruginosa* is defining the biological, experimental, and computational frameworks. The biological question we seek to answer is the contribution of QS to the onset of swarming. We define swarming as the concerted movement of a bacterial community in a given direction, for example, towards nutrients and/or other exogenous signals. In accordance with previous molecular studies, we base the model on the regulation of a few key genes. The experimental framework is based on the growth properties of bacteria on agar plates. We used the so-called swarming agar plates that allow the growth of activated bacteria, but not of non-activated ones.

The experimental work was done by our research collaborators in a microbiology lab (ICGEB, Trieste, Italy). The *P. aeruginosa* strain PUPa3 used in our studies is an environmental rice rhizosphere isolate from India. In order to construct the signal negative (SN) mutant of strain PUPa3, both *lasI* and *rhlI* were inactivated via a two-step insertional inactivation using suicide plasmids [106]. Similarly, the signal blind (SB) mutant was constructed by inactivating both the *lasR* and *rhlR* genes in strain PUPa3 by insertional inactivation using suicide plasmids.

Swarming assays were performed using M8 medium plates (M9 salts without NH_4Cl) [69] containing 0.5% agar and supplemented with 0.2% glucose and 0.05% glutamate [107]. The inoculation was performed with a sterile toothpick dipped in a bacterial suspension of OD_{600} 2.7. Next, plates were incubated at 30°C overnight, followed by room temperature incubation for additional 48 hours. AHLs were either acquired from Fluka-Sigma-Aldrich or from P. Williams (University of Nottingham, UK) and added exogenously to swarming plates to a final concentration of 2 M. *P. aeruginosa* was also grown in LB rich media with 0.5 w/v of agar.

3.3 Bacterial model

The onset of swarming in environmental *P. aeruginosa* PUPa3 was described with a simplified computational model in which cells in random motion communicate via a diffusible signal (representing *N*-acyl homoserine lactones, AHL) as well as a diffusible, secreted factors (enzymes,

biosurfactans, i.e. "public goods") that regulate the intensity of movement and metabolism in a threshold-dependent manner. As a result, an "activation zone" emerges in which nutrients and other public goods are present in sufficient quantities, and swarming is the spontaneous displacement of this high cell-density zone towards nutrients and/or exogenous signals. The model correctly predicts the behavior of genomic knockout mutants in which the QS genes responsible either for the synthesis (*lasI, rhlI*) or the sensing (*lasR, rhlR*) of AHL signals were inactivated. For wild type cells the model predicts sustained colony growth that can however be collapsed by the overconsumption of nutrients.
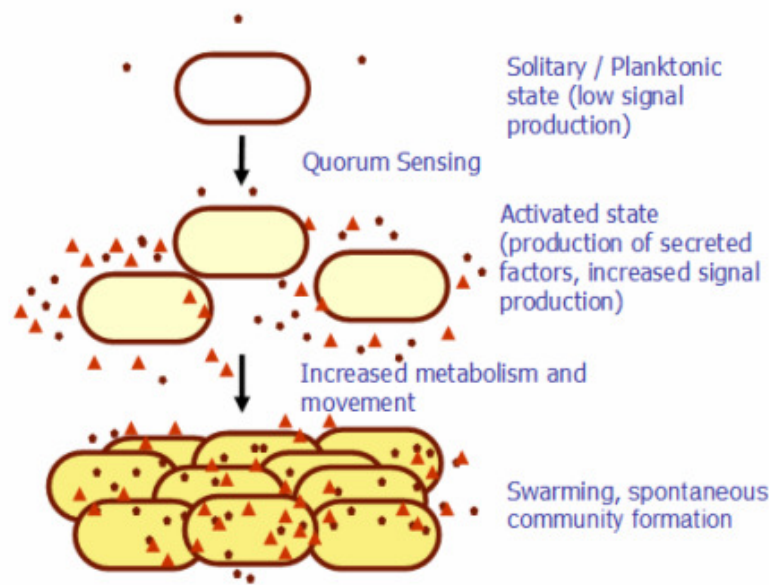

Figure 23 The principle of QS-mediated swarming in P. aeruginosa

We address the onset of quorum sensing-mediated swarming in *P. aeruginosa.* We consider the initial phase of swarming as being controlled by threshold levels of AHL signals and secreted factors (public goods), under the dual control of regulatory proteins and signal synthases of the LasI/R and RhlI/R AHL QS systems. We compare the swarming behavior of the wild-type strain and quorum sensing knock-out mutants in the presence and absence of exogenous AHL signal molecules. We present a simplified, agent-based model for describing the onset of QS, based on a threshold-dependent representation of the early regulatory events and demonstrate that our simplified model provides a qualitatively correct description of the swarming response, in addition to agreement with other models that are more concerned with the overall pattern of colony formation.
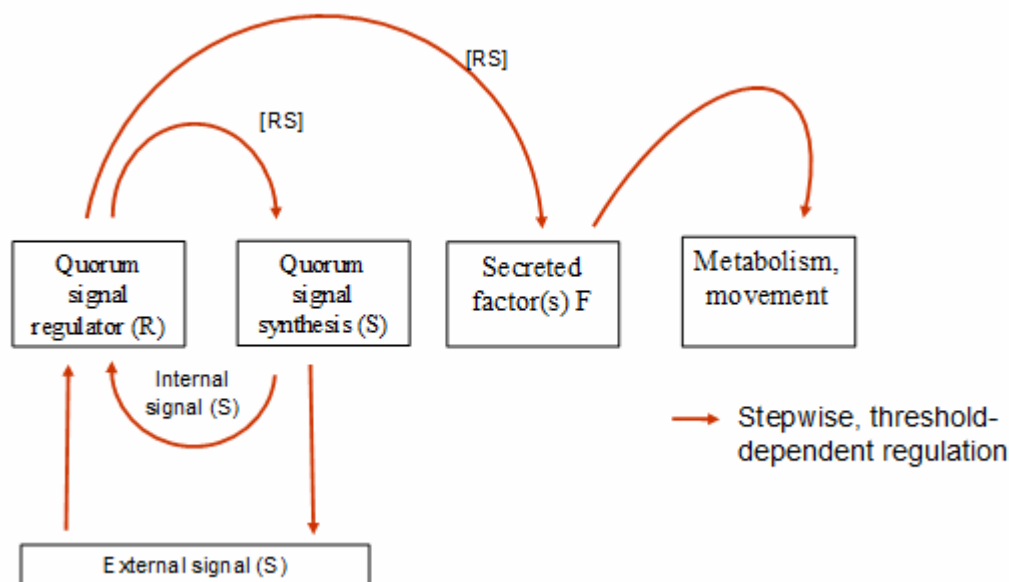
Figure 24 Simplified regulatory framework. The system has a single QS signal S (that in vivo corresponds to C12-3-oxo-AHL and C4-AHL). If the level of S exceeds a certain threshold level, the cell becomes activated. Production of S is increased by a positive feedback loop, and production of a factor F starts. F corresponds to all secreted factors (i.e. "public goods") such as surfactants, enzymes, siderophores, and so on, that the cells secrete into the environment. If the concentration of F exceeds a threshold, the cells start to swarm: they increase their movement, nutrient intake, and metabolism.

Computational model

We constructed a simplified logical framework that incorporates the salient features of the early regulatory events. In this scheme, there is only one signal molecule S, which corresponds to C4-AHL and C12-3-oxo-AHL of *P. aeruginosa*. The cellular concentration of this signal is in equilibrium with the environment. *In vivo*, AHL signals activate the synthesis of various secreted factors, such as the biosurfactant rhamnolipid, which is necessary for the cells to move on the surface, enzymes, such as proteases that digest macromolecular nutrients in the environment, antibiotic compounds that fend off competition by other bacteria, siderophores that help collect metal ions from the environment, and so on. In our simplified model, all public goods are included in a generalized secreted factor F. This factor is in equilibrium with the environment and stimulates the cells' metabolism and movement. If the concentration of F is greater than a given threshold, the cells move and divide faster and consume more nutrients, that is, they initiate swarming. The model has

51

three states: **i)** the solitary or planktonic state, **ii)** the activated state, and **iii)** the swarming state. In the solitary or planktonic state, cells produce low levels of signal molecules, and have low rates of movement and metabolism. Once the level of S reaches a threshold, the cells enter the activated phase in which **a)** the signal production increases and **b)** production of secreted factors ("public goods") starts.

Table 2 Genotype and expected phenotype of *P. aeruginosa PUPa3* and its knock-out mutants

Cells	Genotype				Expected phenotype					
					Signal		Factor		Swarming	
	LasR	*LasI*	*RhlR*	*RhlI*	Pro-duces	Res-ponds	Pro-duces	Res-ponds	Alone	With added signal
WT	+	+	+	+	+	+	+	+	+	+
SN	+	-	+	-	-	+	+	+	-	+
SB	-	+	-	+	(+)	-	-	+	-	-

Algorithmic model

We chose an agent-based scenario to simulate the movement of cells on a 2D agar surface. During each time interval, the cells move to a new location, consume nutrients, and produce AHL signals. The cells make steps of equal length in a randomly selected direction, and if they get into a region with insufficient supply of nutrients, they enter a stationary phase. If nutrients are available locally, the cells ingest them in terms of "energy", and if the stored energy exceeds a certain level, they divide. This is a highly simplified scenario, in which the cells do not sense the nutrients' concentration or gradient, and/or the AHL signals, and they do not orient their movements as a function of these gradients. Rather, they simply switch on and off their genes in a threshold-dependent manner. At the beginning of the simulation, the cells are placed at the starting line of a longitudinal track representing the agar plate

t = 0 ⟶ t = T "Active zone"

A

wall

B

n(i)

Smoothed curve of n(i) vs. i

C
— Signal
··· Public goods
— No of bacteria
··· Nutrients

$$n(i) = \sum \left(i^{th} column \right)$$
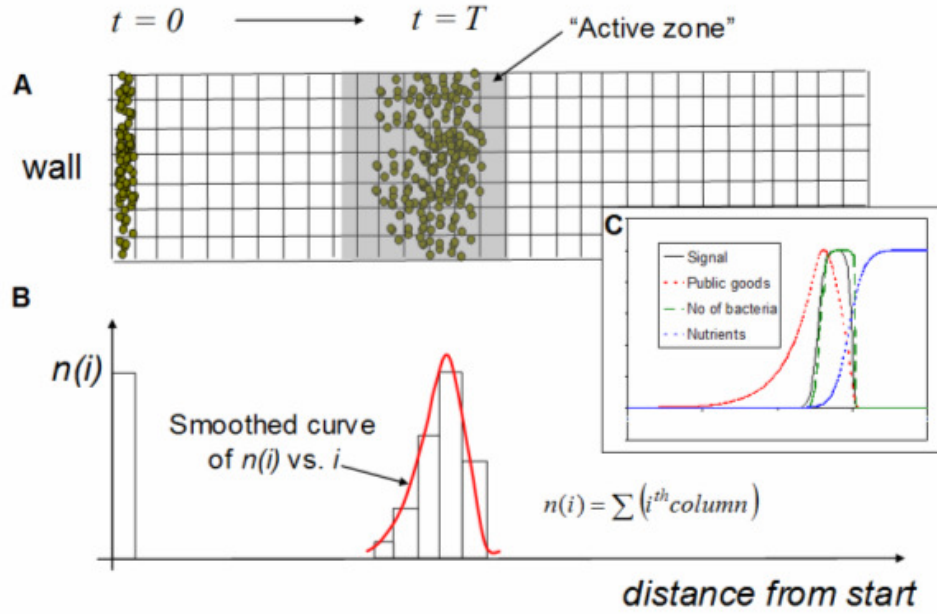
distance from start

Figure 25 Model outline. The model describes the movement of cells on a longitudinal segment of the plane, discretized into squares (A). On the longitudinal sides, the track has periodic boundary conditions with respect to cell movement and diffusion. At the beginning (t = 0), the cells are placed at the starting point at random positions. At each time point, the cells carry out the algorithm prescribed. As a result, the cells form an advancing front, and at each time T, the distribution of cell density as well as signal concentration is determined. The distributions found are irregular and asymmetrical and were scaled to the same upper value (inset C).

In contrast to full colony morphology models, this setting includes only a small portion of the colony. During the simulation, the randomly moving cells spontaneously form a front or "activation zone" in which the level of public goods is sufficient for keeping the cells in an activated state. This zone then spontaneously moves in one direction, i.e., towards the nutrient-rich region. At given intervals during the run, the cell density is calculated by counting the cells within selected areas or the race track. In addition to the moving cell agents, the model includes diffusible materials (nutrients, AHL signals) that are allowed to diffuse at each time point. The density of the cells and the concentration of the AHL signals show irregular bell-shaped curves. In the above model system, all quantities are defined in arbitrary units, and only a few "realistic" choices are made. For instance, we assume that AHL signal production substantially increases as the cells become activated [108]. Also, we assume that the production of the AHL signal requires relatively little energy, while the production of factor *F* is much more energy-expensive. This is based on the well-known fact that swarming requires massive quantities of secreted factors such as enzymes, siderophores, and surfactants produced by a large number of genes, as opposed to the relatively few signaling genes [109] [64].

We designed an agent-based model for representing the cells of *P. aeruginosa*. In this model, each cell is an autonomous agent that regulates its own behavior depending on the concentration of nutrients as well as AHL signals (S, F) found in its environment. The cells perform random movements on the 2D plane, and interact with each other via AHL diffusible signals. Each parameter in this system is defined in arbitrary units, with typical examples provided in the parameter tables (**Table 3** - **Table 5**).

Each autonomous agent carries out a simple algorithm. The functions performed by the cell are regulated in a threshold-based manner according to the regulatory scheme shown. In the solitary or planktonic state (**P**), there is a baseline level of signal *S* production. As the environmental concentration of *S* exceeds the threshold, the production of *S* increases 5-15-fold, and the production of secreted factor (*F*) starts. This is the activated state (**A**). As soon as the concentration of *F* in the environment surpasses a threshold, the cells increase their nutrient intake and move faster, resulting in the swarming state (**SW**). It can be conceived that the level of S falls below threshold while F is still above it. In this case, the cells move and metabolize at the rate of the swarming state, but production of S falls back to the lower level.
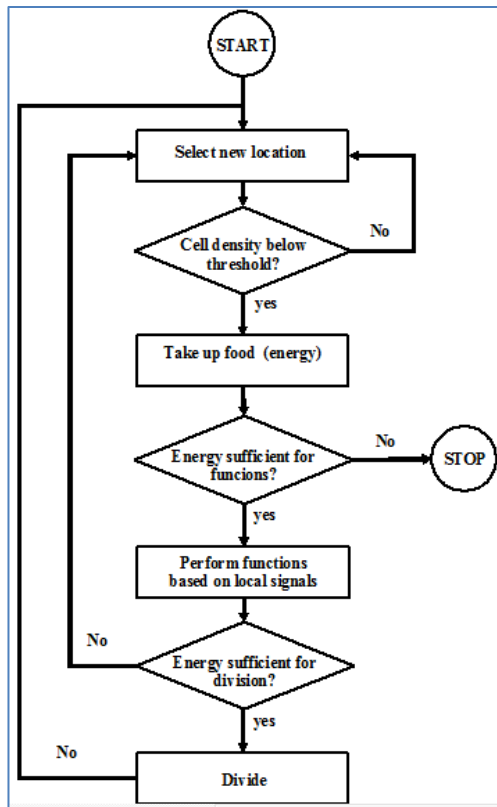

Figure 26 Properties of cell agents. Description of the algorithm carried out by the agents at each time point. At 'STOP', the cells irreversibly enter a metabolically inactive stationary state. The functions depend on the state of activation of each mutant and these are regulated in a threshold-based manner by the signal levels.

The process is governed by the energy balance of the cells. At each step, the cells take up a certain amount of nutrients, defined in arbitrary energy units. The energy is spent on maintenance ("metabolism"), production of S and F, while the remainder is stored. A cell's stored energy can thus be described as

$$E(t+1) = E(t) + E(food) - E(S) - E(F) - E(metabolism) \ [1]$$

, where $E(t)$ is the stored energy at time t, while the other terms represent the energy expenditure corresponding to nutrient intake, AHL signal production, nutrient production and metabolism, respectively. If the stored energy exceeds a threshold, the cell divides. If the stored energy is not sufficient to cover the expenditures, the cell will enter into a stationary phase i.e. it irreversibly ceases to function. The agents proceed via random steps, where a step of a given length is taken in a randomly chosen direction. In the swarming state the cells move approximately 3 times faster than in other (solitary or activated) states.
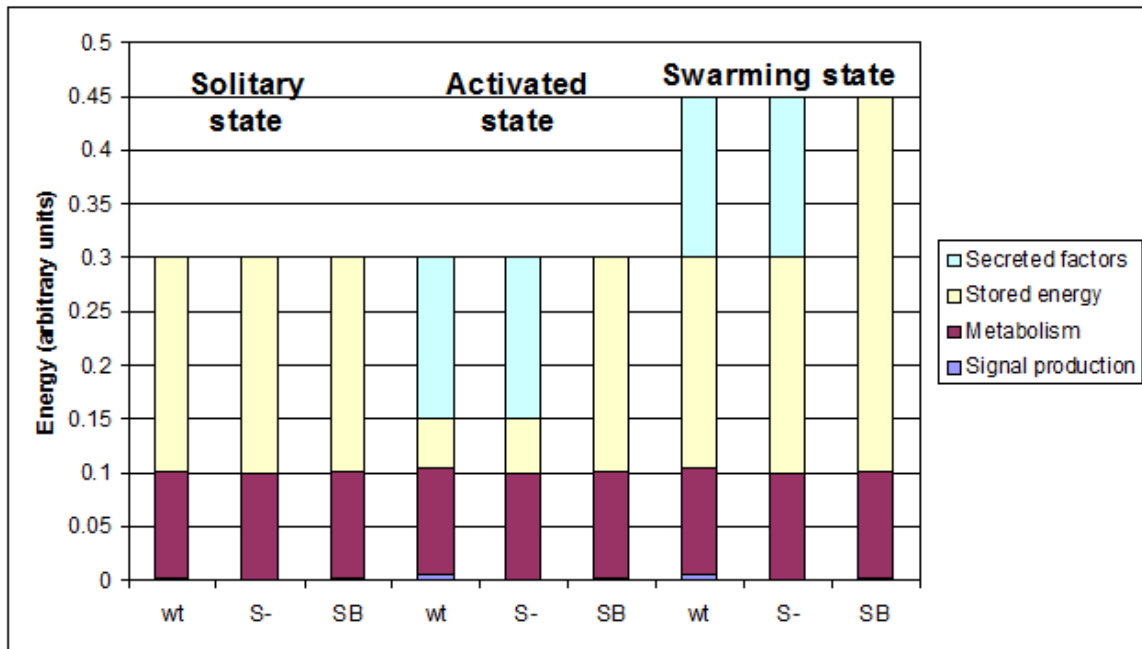

Figure 27 Energy balance of the various mutants studied. The numerical values are shown in Table 1 and described in the Appendix. "Swarming state" includes two states (with S production "on" [swarming level] or "off" [solitary level]).

Diffusing materials

Initially, the environment is represented in terms of a single diffusible material N, denoting all nutrients. In the process of the simulation, cells will produce other diffusible materials, such as signal S and factor F. The concentration of such a component u is described by the equation:

$$\frac{du}{dt} = D\nabla^2 u - Ru , \quad [2]$$

, where D and R are the uniform diffusion and decay constants, respectively. We assume that nutrients diffuse but do not decay (R[nutrients] = 0), but S and F both diffuse and decay. Equation [2] is a typical reaction-diffusion equation that is solved independently for nutrients and signals on a rectangular grid with periodic boundary conditions using an explicit finite difference method, at each time point of the simulation. When compared to models designed to describe colonial patterns, our model is highly simplified, since it is threshold-regulated and does not make the reactive terms proportional to the different signal concentration gradients.

The environment is represented as a 2D longitudinal track with periodic boundary conditions on the longitudinal sides. The plane is discretized into squares, and the concentration of diffusible materials is considered constant within the square. This setup corresponds to a longitudinal cylindrical surface, starting with an impenetrable "wall" at the beginning of the longitudinal "racetrack". At the beginning of the run, the cells have a randomly chosen amount of stored energy, and an equal number of such cells are placed to randomly chosen locations in each square along the starting line. The colony boundary is represented by a line separating the cell colony from the environment. Initially, this separating line will be parallel with the starting line, one square away from it. As the simulation progresses, the cells will move randomly within the boundary while both F and S diffuse outside the boundary. The advancement of the boundary was modeled according to a modified principle adapted from Cohen [96], that is, the cells' escape attempts were counted for each square of the outer square adjacent to the boundary, by incrementing a boundary advancement counter BI with a value of $1+ (k * F)$, where F is the concentration of the factor and k is a constant of proportionality. If BI reaches a threshold, the border moves past the square in question. As a result, the border advances from square to square.

Parameters used in the computational model

Table 3 Global parameters

Parameter	Typical value
Medium size X	250[a]
Medium size Y	2000[a]
Lattice square size	5[a]
Lattice size X	400 squares
Lattice size Y	50 squares
Initial nutrient per square	400[b]
Initial signal per square	0[c]
Maximum number of bacteria per square	10

Values given in arbitrary units, [a]length units, [b]Nutrient concentration units, [a]signal concentration units

Table 4 Movement parameters

Parameter	Typical value
Cell movement/time point	
Solitary and activated states	1[a]
Swarming states	5[a]
Border movement	
Border advancement threshold	120
Border advancement constant k	1

Table 5 Energy expenditure per time point

	Solitary state (S off, F off)			Activated state (S on, F off)			Swarming state (1) (S on, F on)			Swarming state (2) (S off, F on)		
	WT	SN	SB	WT	SN	SB	WT	SN	SB	WT	SN	SB
Signal production	0.001	0.000	0.001	0.005	0	0.001	0.005	0.000	0.001	0.001	0.000	0.001
Metabolism	0.100	0.100	0.100	0.100	0.100	0.100	0.100	0.100	0.100	0.100	0.100	0.100
Stored energy	0.199	0.2	0.199	0.045	0.05	0.199	0.195	0.200	0.349	0.199	0.200	0.349
Secreted factors	0.000	0.000	0.000	0.150	0.150	0.000	0.150	0.150	0.000	0.150	0.150	0.000
Nutrient intake	0.300	0.300	0.300	0.300	0.300	0.300	0.450	0.450	0.450	0.450	0.450	0.450
Total energy consumed	0.101	0.100	0.101	0.255	0.250	0.101	0.255	0.250	0.101	0.251	0.250	0.101

Values are expressed in arbitrary energy units and indicate typical parameter settings for each time point. The threshold for division was expressed in the same units and typically set to 6.0.

Table 6 Diffusion and decay parameters

Name	Signal S	Factors F	Nutrients N
Diffusion coefficient D	0.020	0.020	0.030
Decay rate R	0.001	0.001	0.000

All values are given in arbitrary units

3.4 Computing the bacterial communication network

To compute the global communication network is computationally expensive, if done for many time-steps. Each individual agent position is stored for certain times steps. To compute the communication network, the distances from each bacterium to all the other bacteria are computed and those bacteria falling under a proximity threshold are linked to the bacterium. Thus a very large network can develop, with a number of nodes equal to the number of bacterium agents present the simulation the specific timestep and with a number of links dependent on the distance threshold being used. On the resulting network, various parameters can be computed, as mentioned in the previous chapters.
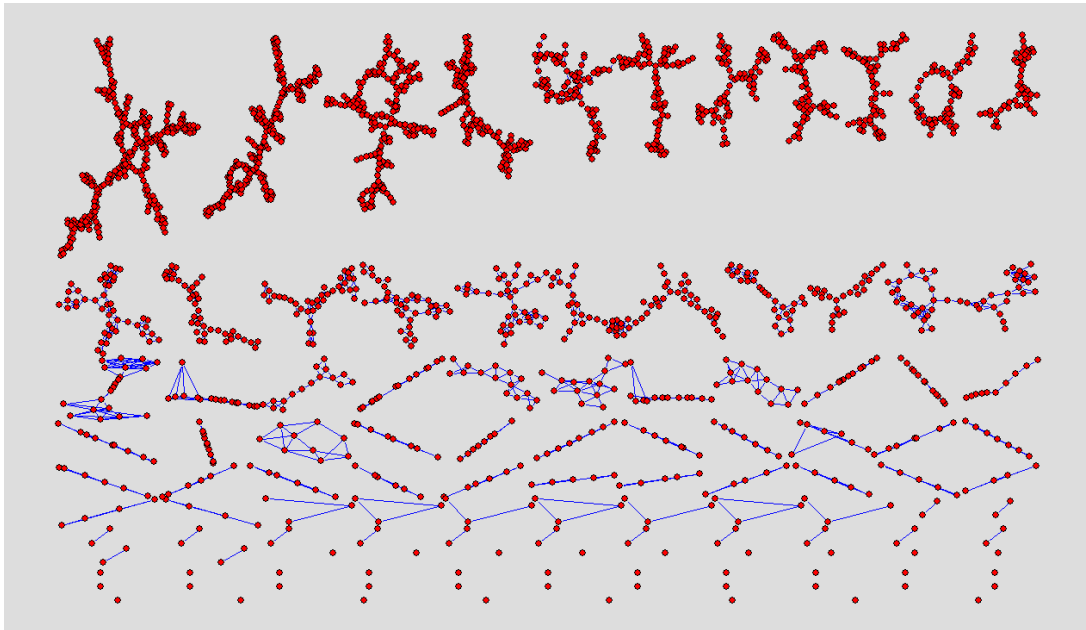

Figure 28 Example of the bacterial communication network, computed from simulation data, for a small distance threshold (1.83). The network is formed by many non-connected subgraphs. If the threshold is sufficiently increased the number of connected components becomes smaller, and eventually the network becomes fully connected.

Good mixing among the different species in the simulation can be essential for the survival of swarming. To measure mixing one can look at the bacterial colony as a communicating network and employ specific measures of mixing. Example of that is the mixing coefficient (which measures what is the average number of neighbors that belong to a certain species) or the average nearest-neighbor distance, from within the same species or between one species and another. The

data that will be presented in the result section will show how important is good communication for the survival of the colony.

4. Results and discussion

4.1 Results of the network evolution simulations

The emergence of self organized network structures is one of the key aspects of molecular and cellular interactions. In order to remain stable in time and resistant to attack, biological networks must evolve in a particular way. One of the fundamental models proposed for network evolution is that of Barabási's that suggests a link between the growth of a network, and its topology as well as robustness. Other models, including the one discussed here, rearrange a given size network and look at the resulting topologies. Rearrangement ("rewiring") models follow the traditions of evolutionary modeling, i.e. they optimize a fitness function that combines various factors into one numerical index. Naturally, there are many ways to formulate and combine the components of the fitness function and testing the possibilities makes the process computationally expensive.

Our main goal was to study the evolution of robust yet efficient network topologies and to see if selecting mutations only for efficiency or only for attack tolerance (robustness) will influence network topology. We will show that concomitant selection for efficiency and robustness influences the fundamental topological properties of the network, and that evolution under multiple attacks leads to distinct topologies.

The choice of network size

The network evolution on smaller graphs can lead to some topologies different from those developed by large networks, depending on the choice of nodes/edges proportionality.

The variation in the number of nodes and edges doesn't change the optimal topologies if the proportions are kept for nodes and edges. We used a number of edges proportional to the number of nodes:

$\frac{e}{n} = k$. In our study we used for k a value of 0.024 that will give 120 edges for a network of 100 nodes. We modeled the evolution of sparse graphs with 100 nodes and 120 edges, which roughly corresponds to the node/edge ratio of gene-regulation networks.

The proportionality of edges can also be to the maximum number of edges allowed for a given number of nodes:

$$\frac{e}{\frac{n(n-1)}{2}} = k \ .$$

But this type of proportionality makes the evolution experiments differ for different network sizes.
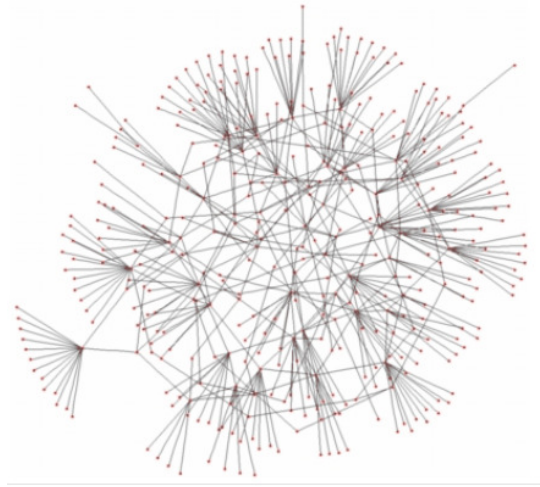

Figure 29 Example of a 400 nodes / 480 edges evolution experiment. Networks that keep node/edge proportionality have the same outcome in our evolution experiments. For computational purposes we favored smaller networks.

Results of the simulation of the random selection algorithm.

The first observation regarding the random selection algorithm is that any of the fitness functions described will make the graph converge to the same structure, since the graph measures being optimized are always the same: efficiency and robustness. The degree of optimality and the convergence speed however differs with the choice of fitness function.

We choose the neutral evolution fitness function ($f(E_t, R_t) \geq f(E, R)$ if and only if $E_t \geq E$ and $R_t \geq R$) because it is straightforward and it doesn't make assumption on the importance of one evolutionary trait (efficiency or robustness) over another. The structures evolved using this

greedy approach are generally less optimized, getting stuck in a local minima, but the structures are sufficiently evolved to see a general pattern emerging.

Another observation is that the efficiency and robustness are not the sole possible principles governing network evolution; rather they are only two of the evolutionary traits of a network. To test this we can measure the level of efficiency, robustness and redundancy in networks evolved using either the efficiency or robustness criteria. The redundancy measure uses an algorithm for computing the k-shortest paths between any two nodes, then counting the average over k of the number of k-shortest paths scaled by the largest number of paths possible (n(n-1)/2):

$$R(G) = \frac{2 < n_k >_k}{n(n-1)} \ .$$

We see (*Figure 31*) that generally an increase in efficiency decreases the robustness of a network, while redundancy is mostly unaffected. The drastic decrease in robustness is due to the formation of high degree central nodes, which if attacked will decrease the efficiency of the network. Evolving the network using an efficiency-only rule leads to the well-known star topology where the center is formed by one (or a few) nodes and all other nodes are peripheral. Evolution under a robustness-only regime results in a large core and a small periphery, with the nodes apparently preserving the small degrees seen in the initial network. It thus appears that robustness strengthens the core, while efficiency increases the periphery.
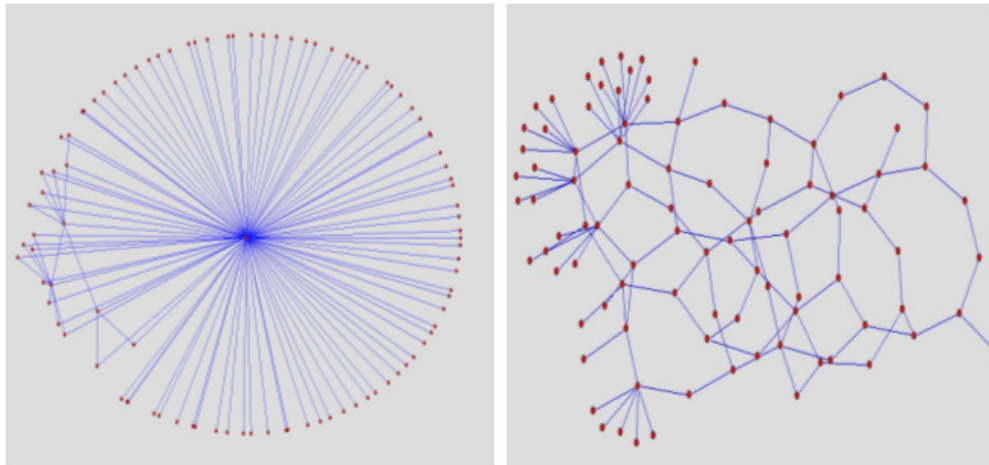

Figure 30 The different results of evolving network optimized in terms of efficiency and robustness. *Left side of the figure the network is optimized for efficiency only, while on the right the network is optimized for 5 attack robustness only.*
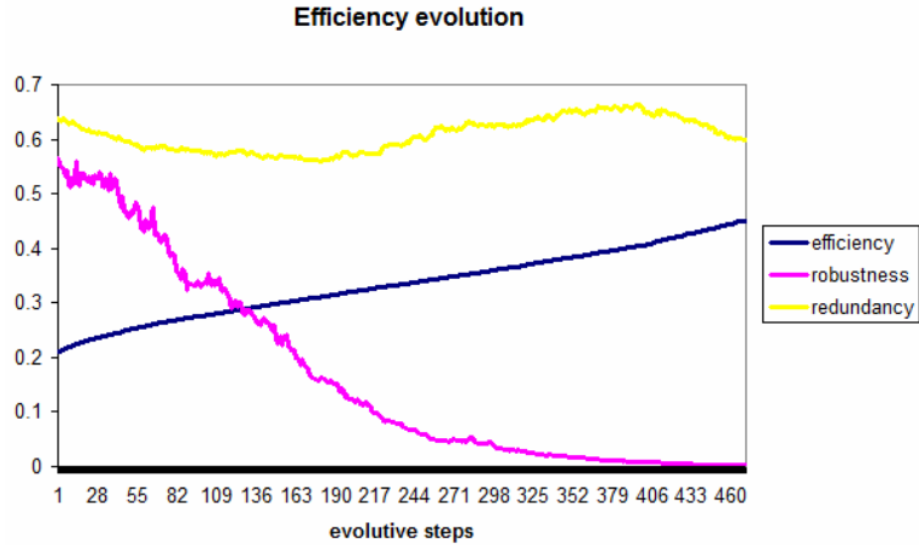
Efficiency evolution

Figure 31 Time plot showing the relation between efficiency and robustness. An increase in efficiency will decrease the robustness while having no strong influence on redundancy. The network order/size is 100/120, the robustness is measured for 5 consecutive attacks.
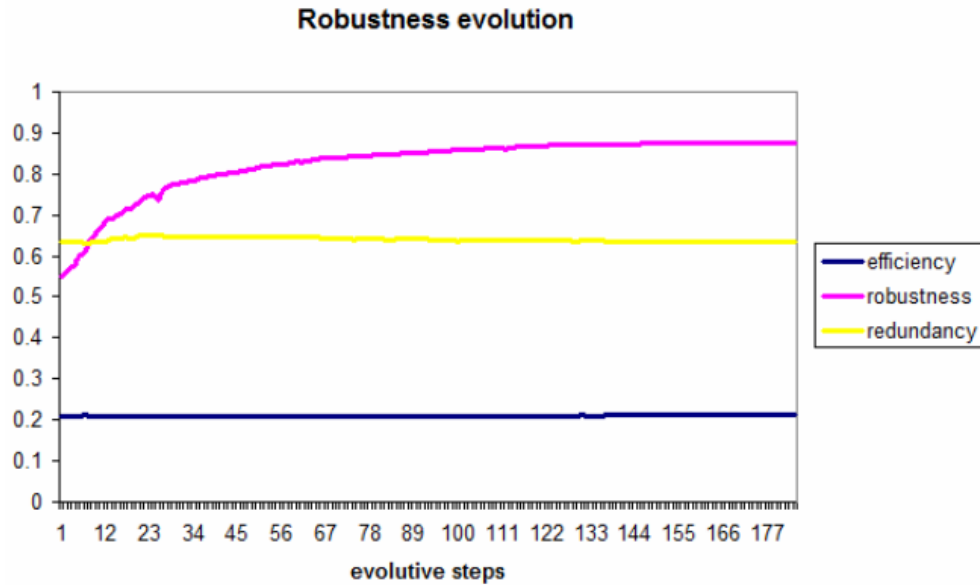

Robustness evolution

Figure 32 Time plot showing the relation between efficiency and robustness. An increase in robustness has no effect on efficiency or redundancy. The network order/size is 100/120, the robustness is measured for 5 consecutive attacks.

The result of attacking multiple nodes

The typical outcome of an evolution experiment that tries to maximize evolution and successive attack robustness is presented in *Figure 33*. In general single attacks lead to a network with several high degree nodes while multiple attacks lead to a more evenly distributed network.
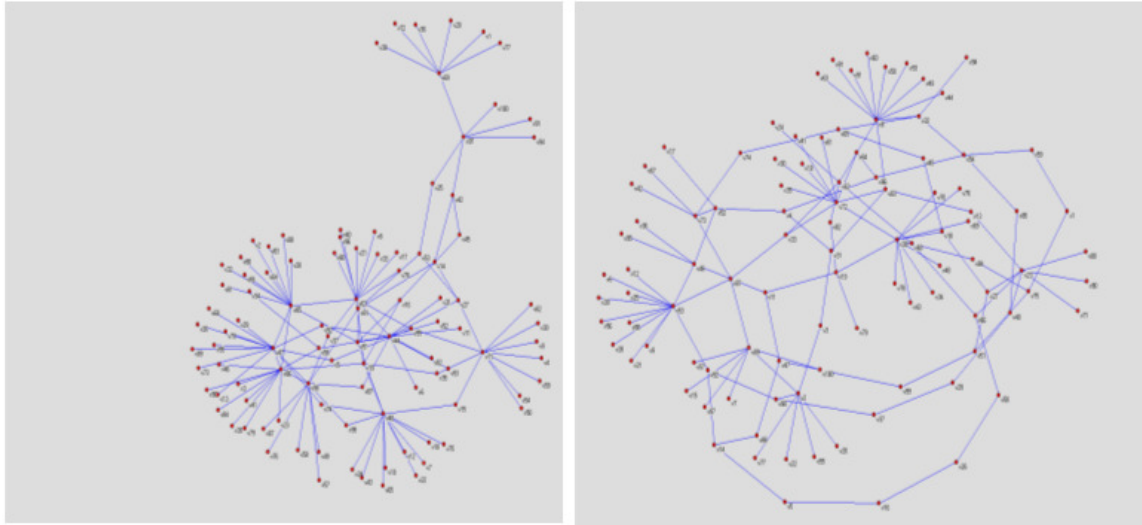

Figure 33 The result of attacking multiple nodes. Left side a network evolved for both efficiency and single attack robustness, right side a network evolved for efficiency and multiple attack robustness. It can be seen how multiple attacks lead to a more evenly distributed network.

This effect can be quantified by measuring the core to periphery ratio of the network. The core is formed by the nodes with degree centrality higher than 1, while the periphery is formed by the end-nodes of the graph, having degree one. The core defined in this manner is complementary to periphery, but other definitions of the core are possible based on different graph centrality measurements. As shown in Table 7**,** selection for multiple attacks robustness greatly increases the core while decreasing the periphery.
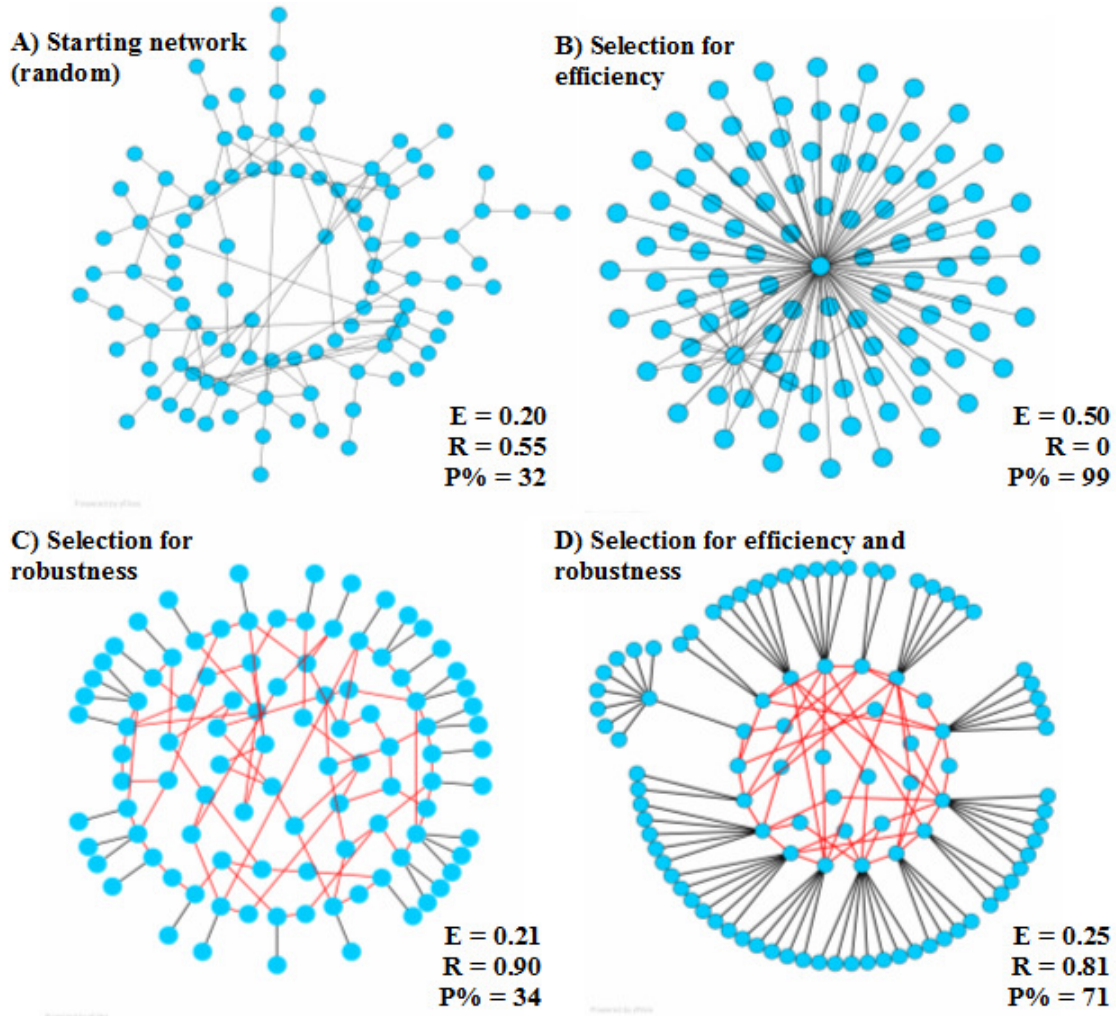
Figure 34 Example of network topologies. *A random network of 100 nodes and 120 edges (A) was subjected to selection under various regimes (B-D) E=efficiency; R=robustness; P=periphery. Note that the parameter P% of D) is between B) and C)*

We measure the robustness to successively attacking the node with highest betweenness centrality. However, a node with high betweenness centrality doesn't guarantee the highest impact over network efficiency if attacked. We used betweenness centrality for computational speed purposes, the alternative being to attack all network nodes and see which of them has the highest impact over network efficiency. Such a centrality measure is generally referred to as efficiency vitality. As we can see from the table, the efficiency vitality based robustness also increases with the number of attacks.

The average nearest neighbor degree can tell us if a network is mixed, i.e. if high degrees neighbors are in the vicinity of every node. We can see that robustness tends to decrease mixing while efficiency increases it.

The clustering value is mostly left at minimum, this is due to the fact that clusters are bad in terms of communication efficiency, and also in terms of robustness (the removal of cluster nodes greatly impacts the overall information flow. Thus we see a general decrease in clustering, except for the case of an overevolved network efficiency, that favors a few highly interconnected nodes.

The effects of evolving efficiency in reducing the graph distances can be seen the next set of measures, the diameter and the radius increases with increased efficiency and decreases with increased robustness, while efficient networks develop a greater periphery (this is a pathlength based periphery different from the simple denomination we used in our core to periphery study). An interesting observation is that the center value tends to decrease with the number of attacks; this is due to the fact that attack robustness favors more eccentricity.

Table 7 Table showing the correlation between several global graph measures for several selection strategies. The values are the average of 20 simulations. Core% - fraction of nodes that are degree > 1. Efficiency vitality robustness n nodes attack - we each time attack the node whose removal is making the resulted network the most damage in terms of efficiency. Redundancy – t he average number of redundant paths between any two nodes. Vertex disjoint paths redundancy - the average number of vertex disjoint redundant paths between any two nodes. Clustering coefficient - the likelihood that two associates of a node are associates themselves. The *eccentricity* of the vertex *v* is the maximum distance from *v* to any vertex. The *radius* of *G* is the minimum eccentricity among the vertices of *G*. The *diameter* of *G* is the maximum eccentricity among the vertices of *G*. Thus, *diameter(G)=*max*{e(v):v* in *V(G)}*. The periphery is the set of nodes with eccentricity equal to the diameter. The center is the set of nodes with eccentricity equal to radius.

Parameter	Initial random graph	Criterion of selection			
		Robustness only	Efficiency + 5attack robustness	Efficiency + 1attack robustness	Efficiency only
efficiency	0.20	0.21	0.23	0.28	0.50
robustness 5 attack	0.55	0.90	0.88	0.56	0.00
robustness 1 attack	0.91	0.98	0.99	0.99	0.04
core%	0.68	0.72	0.59	0.34	0.1
efficiency vitality based robustness 5 attack	0.56	0.65	0.50	0.47	0.00
redundancy	0.64	0.59	0.64	0.57	0.41
disjoint paths redundancy	0.40	0.41	0.40	0.30	0.26
average nndeg	3.21	3.56	4.94	7.69	86.86
average clustering	0.06	0.00	0.01	0.00	0.19
diameter	14	14	14	8	2
periphery	2	6	11	18	99
radius	7	8	7	5	1
center	1	8	8	13	1

We conclude that the topology of the networks changes when the networks adapt to an environment where both efficiency and robustness are required, and the presence of multiple attacks bring about further topology changes as compared to single attacks.

Degree correlations in evolved networks

A secondary type of correlations we can see if we watch how several graph details vary with degree. In *Figure 35* we can see the degree distributions for graphs evolved under multiple attacks differ, the core of high degree and highly connected central nodes becoming more evenly distributed as the number of attacks increases.
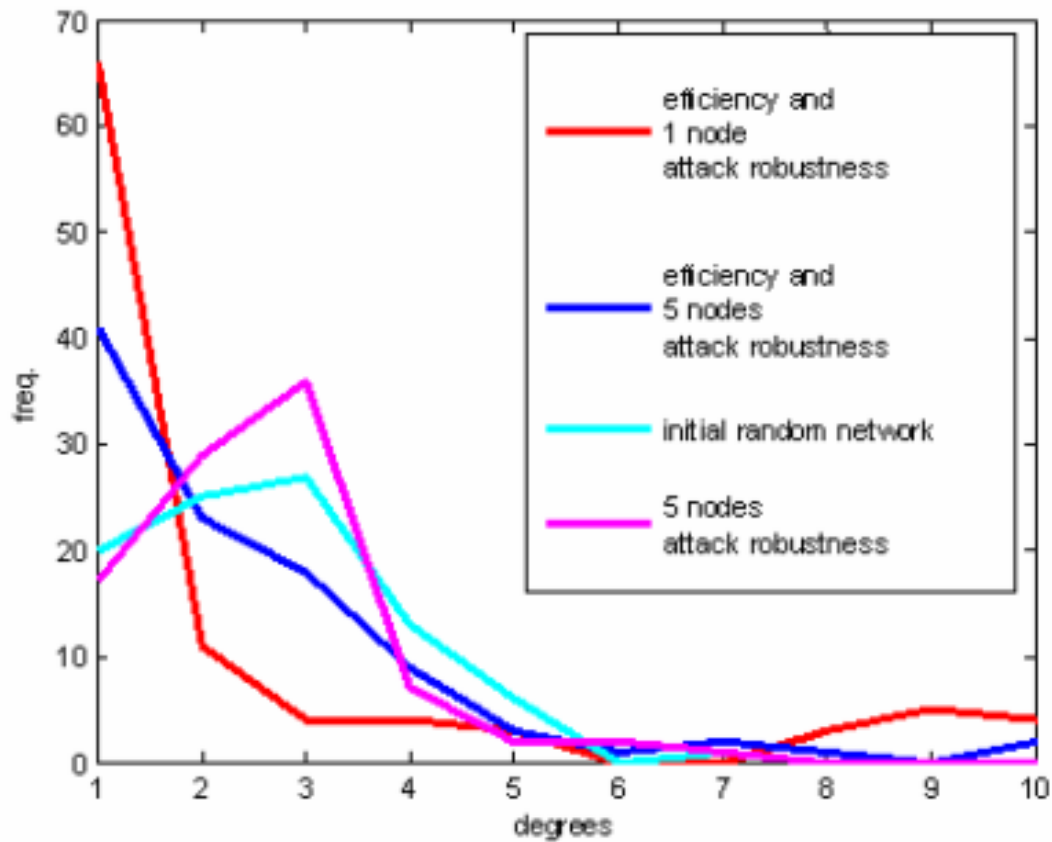
Figure 35 Degree frequencies of networks evolved under several attack types. From the plots we can see how selection for increased robustness results in a higher core, while selection for increased efficiency results in higher periphery.

The degree frequency has a characteristic fat-tailed distribution characteristic to many biological networks, with a slope varying around -2.2 (we know that the slope varies between -2 and -3 for most natural scale-free networks [15]). Characteristic to highly evolved networks is the 'hump' that develops due to many nodes having about the same degree. This is not happening usually in the natural networks, but we have to consider the fact that our model has two details that natural networks are lacking: it doesn't grow in size and it evolves highly optimized network structures (while most natural networks are in a continuous process of mutation and selection, therefore they are not highly optimized structures).
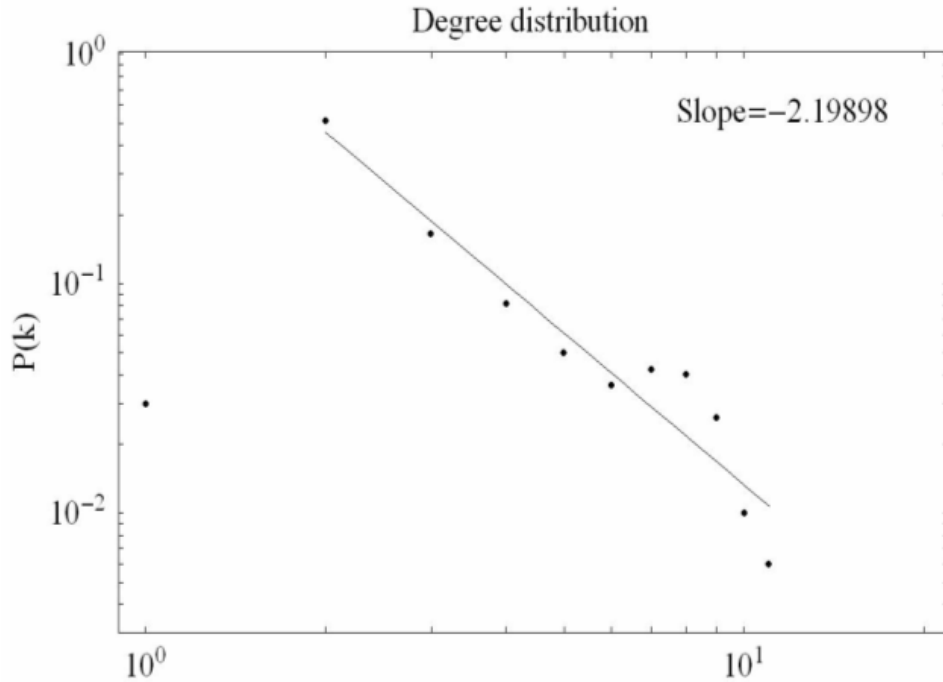
Figure 36 Characteristic degree distribution of a highly optimized 100 nodes/120 degrees network using selection for efficiency and 1 attack robustness. The distribution has a characteristic 'hump' formed by its central nodes.

Other network properties are also correlated with the degree, nodes with high degree having smaller clustering coefficient (involvement into a cluster would make a high impact on network robustness in case of a targeted attack) and a smaller nearest neighbor degree which suggests disassortative networks (disassortativity is another adaptation to increased targeted attack robustness, targeting a node with degree would affect its high degree neighbors as well leading to a downfall in robustness).
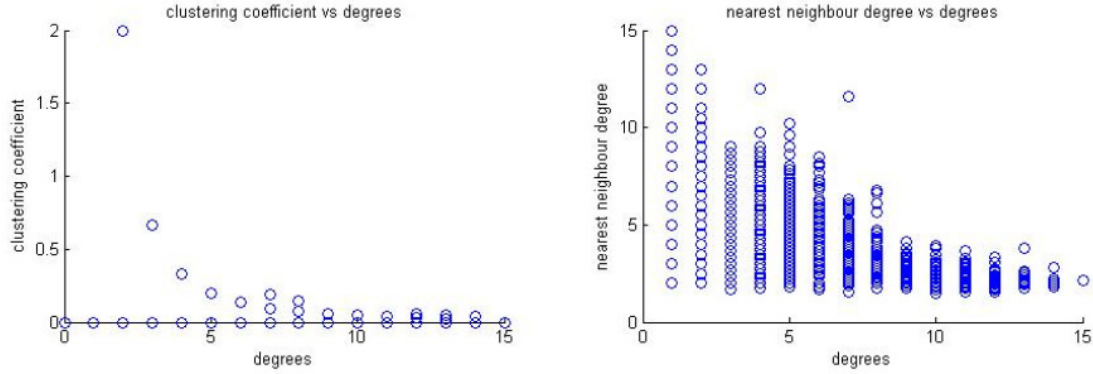
Figure 37 Degree correlations for networks 100 nodes/ 120 links networks, evolved for efficiency and 5 attack robustness. The data from 20 experiments is plotted to make the dispersion of the values more clear. Left: clustering coefficient vs. degrees. Right: nearest neighbor degree vs degree.

The local efficiency also increases with degree, and this is to be expected from a robust network, affecting a high degree node should have small impact over its neighbors. The betweeness centrality also tends to increase with the degree, so the most connected nodes are also in the middle of many paths connecting all graph nodes.
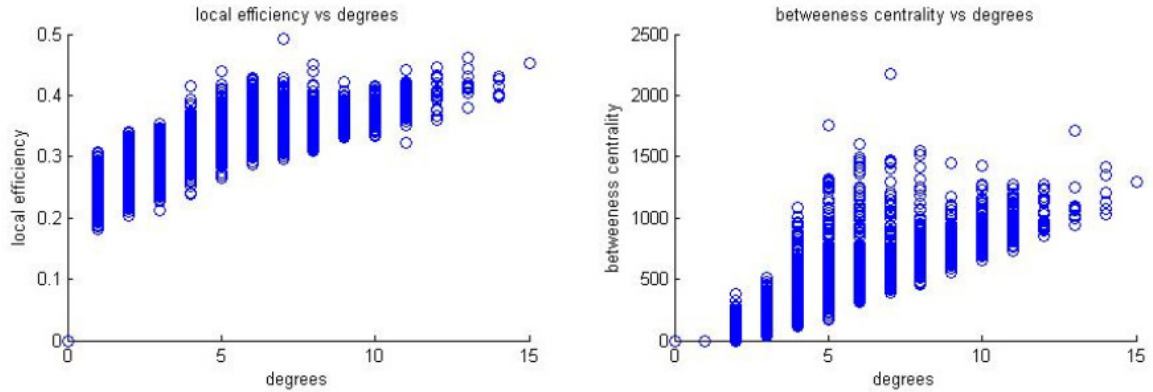

Figure 38 Degree correlations for networks 100 nodes/ 120 links networks, evolved for efficiency and 5 attack robustness. The data from 20 experiments is plotted to make the dispersion of the values more clear. Left: local efficiency vs. degrees. Right: betweenness centrality vs degree.

71

Motifs and path correlations in evolved networks

Motif analysis on evolved network structures reveal the absence of closed loops and the prevalence of bridging pathways connecting high degree nodes. The analysis was done with several graph motif analysis programs, the results displayed in *Figure 39* were computed with Fanmod.
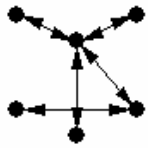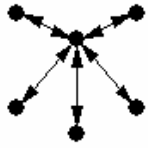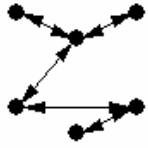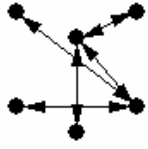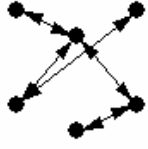
Adj	Frequency [Original]	Mean-Freq [Random]	Standard-Dev [Random]
	30.888%	28.62%	0.0052202
	22.844%	5.9967%	0.0034895
	18.764%	16.303%	0.0068415
	17.247%	18.061%	0.0043183
	7.1144%	17.315%	0.0050955

Figure 39 Graph motif search display for 6 nodes motifs. The search was done using Fanmod, using as input a 100 nodes/120 edges network evolved for efficiency and 5 nodes attack robustness. The most frequent motifs are compared to the frequencies found in random networks. We can observe the appearance of star motifs and a slight improvement or conservation in the number of bridging patterns.

72

Another observations related to the pathways of evolved network is that they become less widely distributed and get shorter on average. We have seen in Table 7 that selection for robustness favors mode redundancy, what we can also see is that the redundant pathways also tend to generally decrease in length, as shown in *Figure 40*.
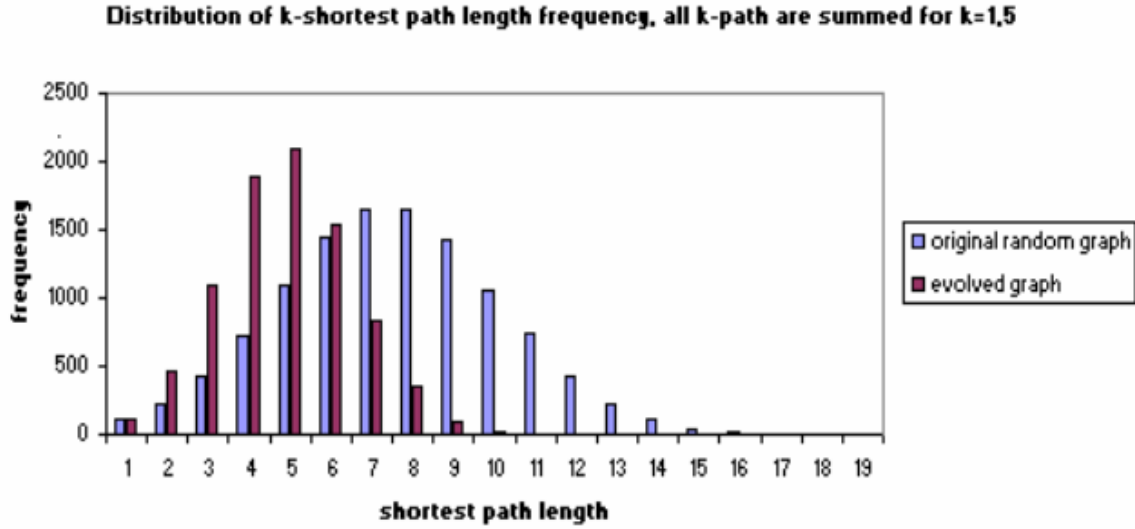
Figure 40 Distribution of k-shortest path length frequency, all k-paths are summed for k=1,5
We used a 100 nodes/120 edges network evolved for efficiency and 5 nodes attack robustness.

The convergence to highly optimized structures

Networks evolved using selection for both efficiency and attack robustness using the random selection algorithm have generally a good convergence to optimized structures. This is demonstrated by the scatter plot in *Figure 41*. One can see a clear distinction between the first part of the evolution experiment and the second part, therefore random evolution has a good convergence to optimized structures.

Figure 41 Scatter plot of the time dynamic of a 100 nodes/120 links random selection evolution experiment, selecting for efficiency and 5 nodes attack robustness. Blue dots mark values for time steps before the middle of the evolution experiment, while red squares mark values after the middle of the evolution. One can see a clear distinction between the two regime, therefore random evolution has a good convergence to optimized structures.
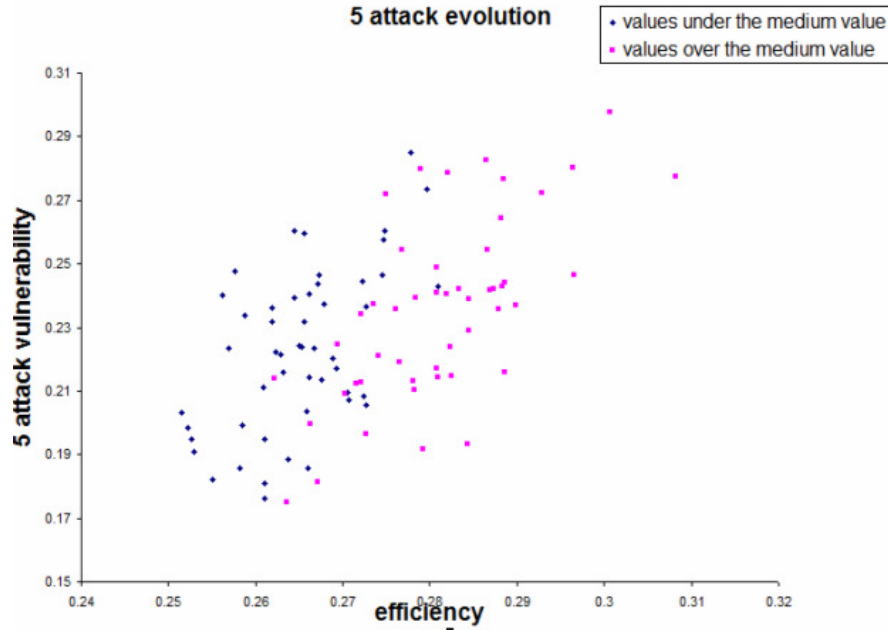
The genetic algorithm described before was used to better use the search space of network structures, avoiding entrapment in local maxima to achieve highly evolved graph structures. The simulation time is much longer even for a small population of 10 graphs, but the results allowed us to formulate an optimal random graph model.

The **(R,P,k)-ring** model contains three parts:

1. A ring graph of order R.
2. A number of P edges will be equally distributed as end-edges attached to each of the ring vertices.
3. The remaining edges will connect the closest vertices in the ring.
4. Randomly selecting with uniform probability a proportion k among the edges previously established at step 3 and rewiring them to link the most distant of the vertices of the ring.

The model is similar to Wattz-Strongatz small world networks, but while a small world can be scale-free, the **(R,P,k)-ring** has many nodes with an equally high degree. The number of attacks decides the choice of R and P. Generally, the more attacks for measuring robustness, the higher the value or R and the lesser the value of P due to the fact that evolution in this case favors a large core and a smaller periphery.
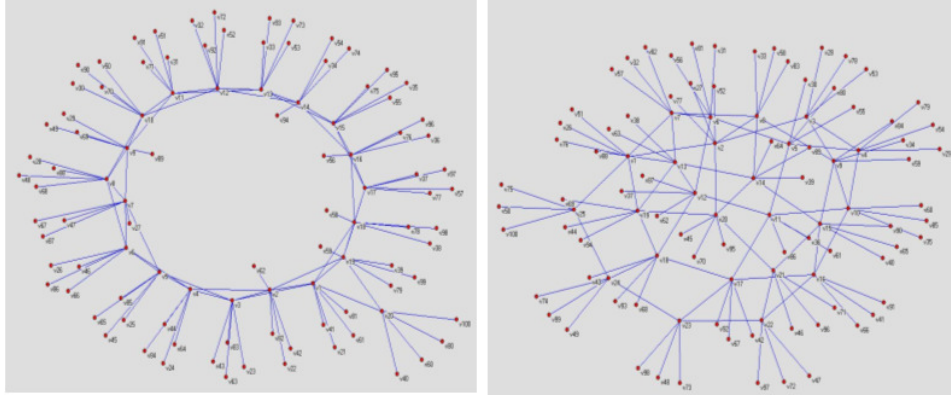

Figure 42 Picture of the process of constructing the (R,P,k) ring model. Left the image of the graph after step 1,2 and 3. Right the result of rewiring the same edges between the most distant nodes of the ring. The values of efficiency and robustness vary with the selection of R, P and k parameter.

Alternative evolution strategies

Our principles for network evolution were based strictly on maximizing efficiency and robustness. Natural networks usually evolve under many other types of constraints. In gene regulatory networks for example, the overall communication efficiency is less important as local robustness in the dynamics of gene regulation, which favors several motives like the feed-forward loop, that are not encountered in our evolved networks.

The same evolution experiments can be done for more than two selection factors, by using complex functions that optimize several network traits that have more relevance for the specific network in question. An example is offered in *Figure 43*, where the network was optimized for efficiency and redundancy instead of robustness. The lack of robustness observance favored the appearance of a dense center, which attacked could destabilize the efficiency of the whole

network. However, due to our selection for redundancy the network exhibits many closed cycles, therefore displaying many of the motifs of a natural gene regulatory network.
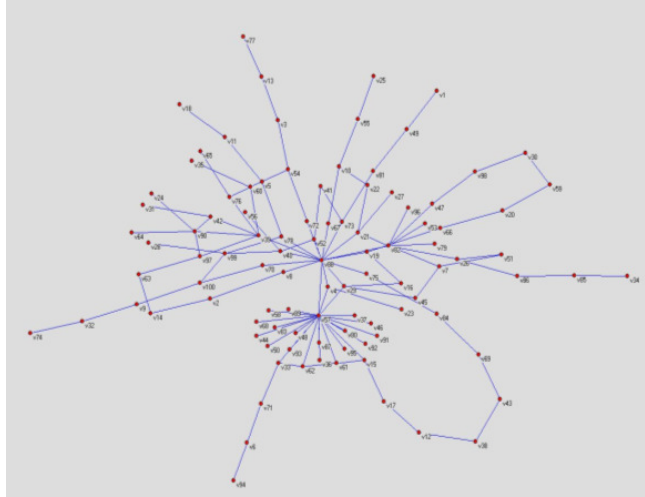

Figure 43 The result of a different evolution principle. *The network was evolved for increased efficiency and redundancy, but no attack robustness. The lack of observance for robustness and the drive for increased efficiency favored a network with a visible central node, which if attacked can greatly damage the efficiency of the network.*

This work has only investigated non-directed networks, but the same analysis was repeated for directed networks and can also be applied for network dynamics. Although simple and straightforward, our network evolution principles can help explain many phenomena observed in biological networks.

4.2 Results of the bacterial colony model

Basic properties of the in silico model

Typical simulation snapshots are shown in *Figure 44*. The actively swarming cells are shown in green and the active zone, that is, the zone in which the cells are active and swarming, is clearly visible throughout the simulation.

Swarming and non-swarming models show distinctly different pictures in this scenario. Swarming cells form colonies that move relatively fast, the number of cells present in the advancing front increases and reaches a plateau that corresponds to the maximal cell density

allowed. On the other hand, non-swarming colonies can not move, so their number can increase only as long as nutrients are locally available. As nutrients are depleted, the number decreases to a baseline level that can be supported by the diffusion of nutrients (*Figure 45*).
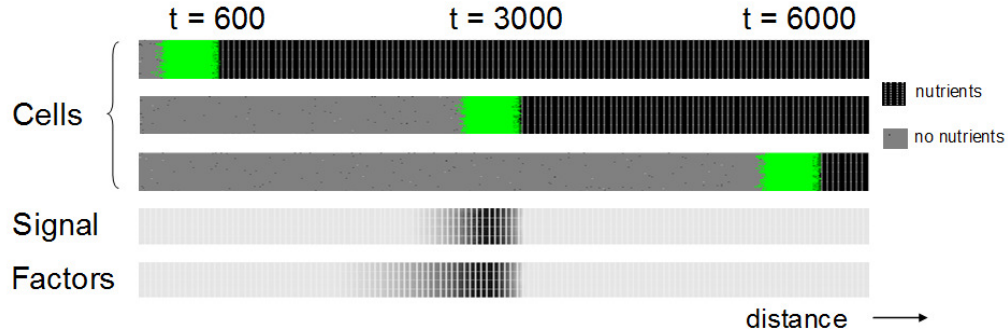
Figure 44 Model behavior. *Snapshots of the advancing front. The green area shows living wild-type cells in the swarming state. The concentration of signals and factors follows the movement of bacterial cells.*
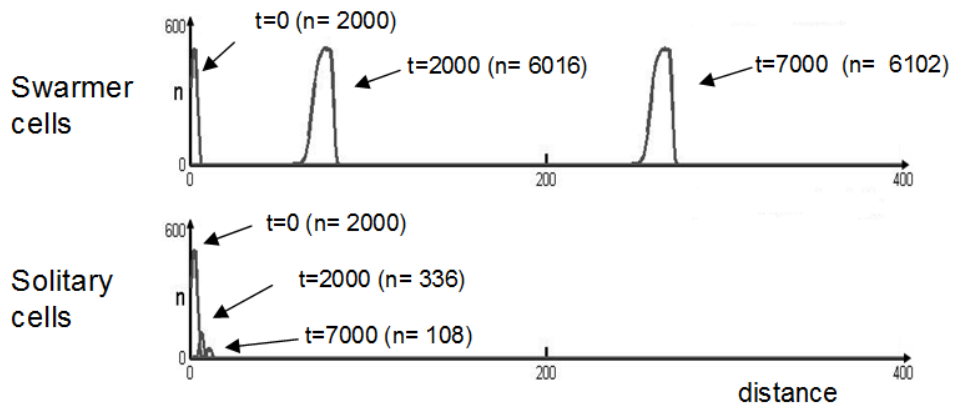
Figure 45 Typical cell density profiles taken at various time points for cells in the swarming and solitary states. *The time is measured in arbitrary units (time steps) and n denotes the number of cells. Noteworthy, the number of swarming cells usually reaches a plateau corresponding to the maximum cell density, while the number of solitary cells unable to swarm decreases rapidly.*

The definition of QS is that cells respond to cell density. *Figure 46* and *Figure 47* show that the model-population in fact acts as such a density switch. At a given cell density the cells get activated i.e. they start to produce factors (green line), and subsequently the cells also start to swarm (red line). The model assumes a threshold signal concentration to activate the quorum sensing response, then another threshold for the level of factors for the onset of swarming. The

77

genetic networks underlying QS are considered to act as a two-state switch [66]. It is worth to note that the starting population is random (both in terms of locations and in terms of metabolic states). Nevertheless, this random population shows a coordinated behavior as it switches from solitary to swarming state.
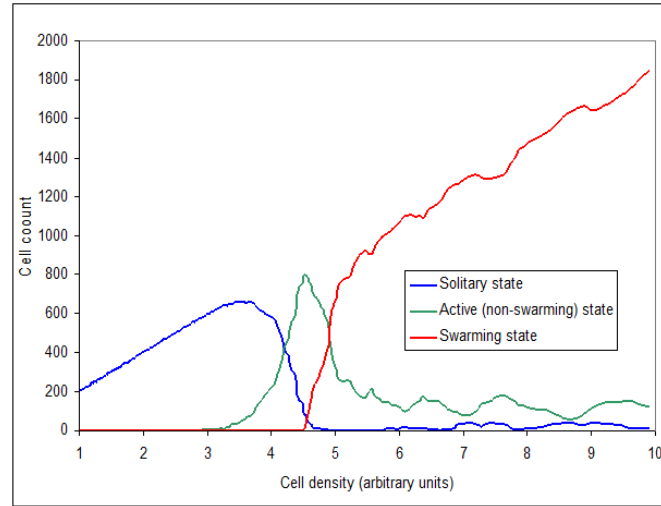
Figure 46 Dependence of the cell's state on cell's density. Note that beyond a certain cell's density level, nearly all cells switch to the swarming state, i.e. the model acts as a density switch. The panel shows the first steps of a simulation starting from a very small population (200 cells).
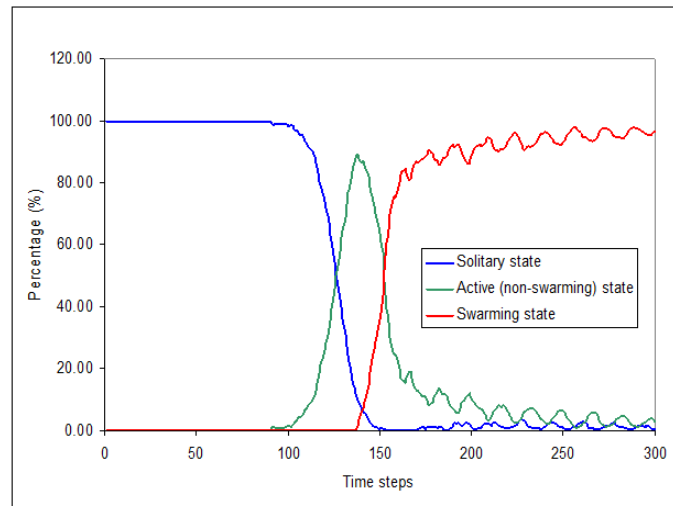
Figure 47 Percentage of solitary, signal producing but not yet swarming (green) and swarming (red) cells in the population. Note that after a certain time practically all cells are in swarming state. The panel shows the first steps of a simulation starting from a very small population (200 cells).

In the modeling experiments presented so far, the bacterial front followed the availability of nutrients. On the other hand, nutrients and signals are both required for movement, which implies that the cell agents are in principle also capable of following a trail of exogenous signals. The example in *Figure 48* illustrates this property of the model. Thus it seems that a simple gene-activation model is sufficient to explain the response of bacteria not only to cell density, but also to exogenous cues that are known to play roles in host/symbiont and plant/pathogen interactions.

t = 600

t = 3000

t = 5000
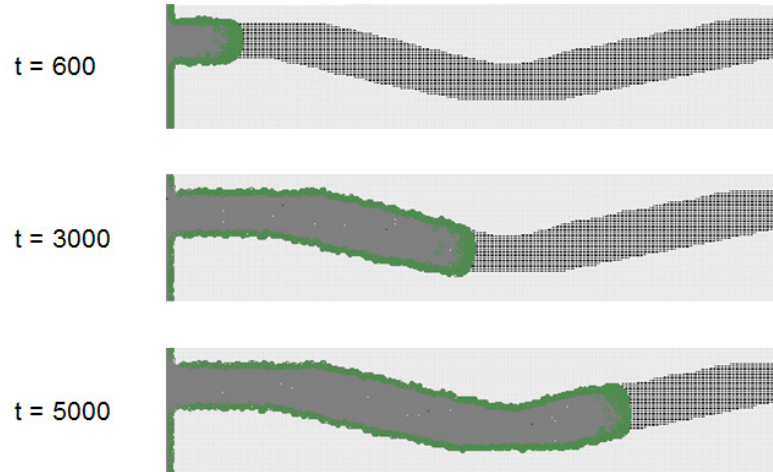
Figure 48 Activation of cell agents by an exogenous signal. In this experiment, a non-diffusing and non-decaying signal S was provided in the form of an irregular trail. The cellular agents were of the SM type that are unable to produce the signal but can respond to it. Note that swarming occurs only along the signal trail. (Screenshots taken at different times t, t expressed in arbitrary units)

Swarming of P. aeruginosa in vivo and in silico

The behavior of wild-type *P. aeruginosa* PUPa3 as well as its mutants is compared in *Figure 49.* In the absence of exogenous AHL signal, only the wild-type cells swarm. If the exogenous AHL signal is added to the plates, the SN mutants will also swarm, both *in vivo* and *in silico*, yet the SB mutants will not. These results show that **i)** the genetic modifications produced the expected phenotypes (Table 2), and **ii)** the simplified regulatory scheme built into the agent-based model provides a qualitatively adequate description of the events.
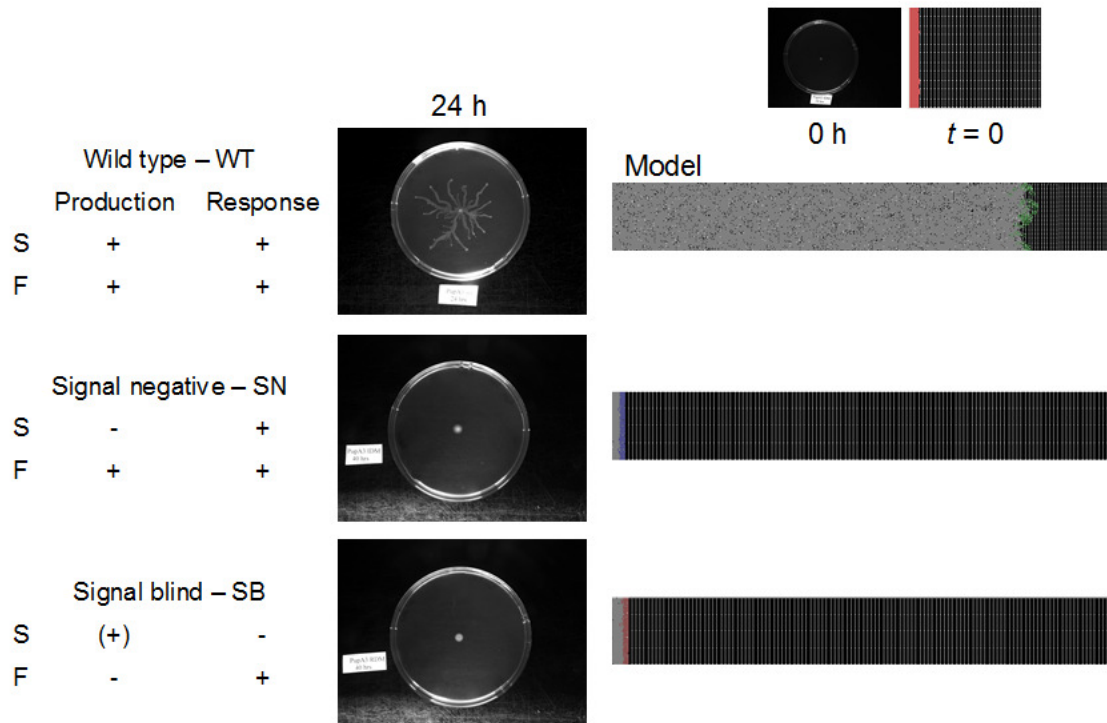
Figure 49 Swarming of P. aeruginosa in vivo and in silico. *The cells' phenotype is indicated on the left. The swarming plates (center) were developed as previously described. The computer model (left) shows the behavior of the in silico model, which is in agreement with the lab. Green color was used for the wild-type, blue for SN, and red for SB cells. (+) indicates that the cells have a basal (solitary) level of production of signal S, that is preserved in all states but does not increase with the onset of swarming.*
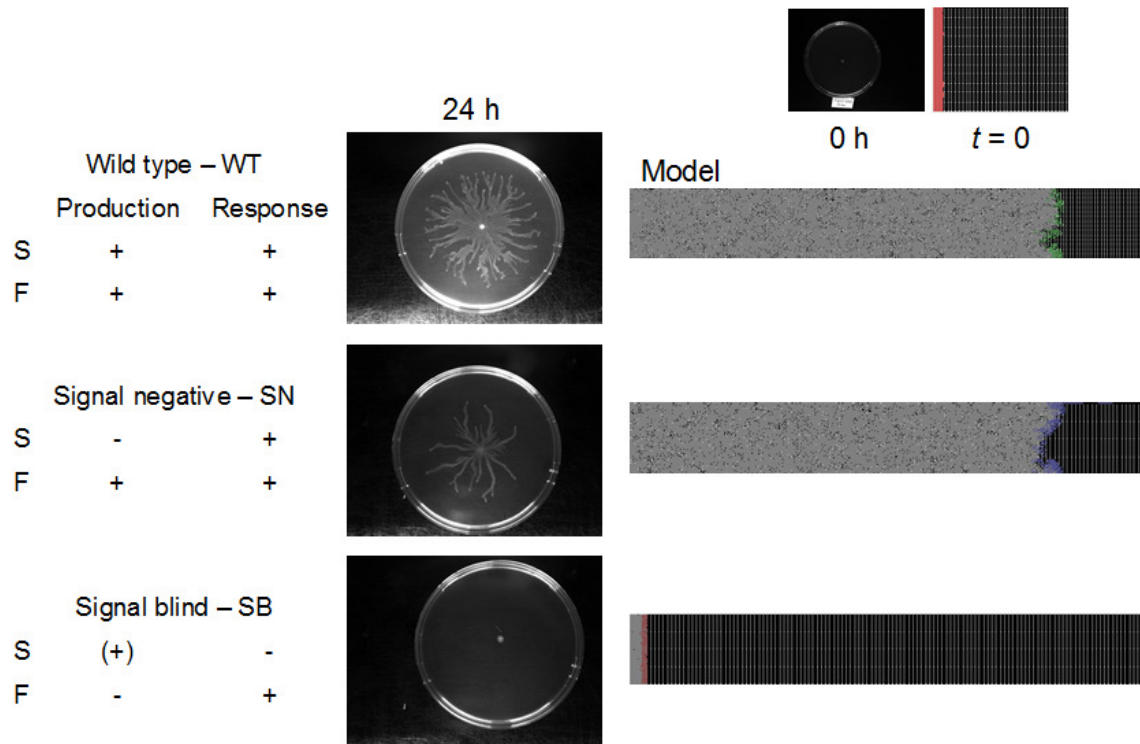
Figure 50 Swarming in the presence of exogenous signals. *The outline of the experiments is the same, except that the colonies were left to develop in the presence of exogenous AHL signals. (In the lab experiments, AHL signals were added to swarming plates, to a final concentration of 2M, whereas in the computer simulation, the signal concentration was simply kept above the threshold level). (+) indicates that the cells have a basal (solitary) level of production of signal S, that is preserved in all states but does not increase with the onset of swarming. Note that SN swarms in response to the signal, while SB does not. The behavior of the wild-type was approximately similar to that observed without exogenous signal.*

Kinetics in silico. Dynamics of the bacteria colony model.

The *in silico* model makes it possible to follow the kinetics of cell populations during the simulation. Swarming experiments show typical saturation-type kinetics that can be described in terms of approximate initial, transient and steady-state phases.

We evaluate the results in qualitative terms, and whenever possible, on a comparative basis (e.g. in comparison with the wild type cells. We show a plot of and population size (*Figure 51*) and

speed (*Figure 52*) as a function of the simulation time (arbitrary units). By way of analogy with reaction kinetics, it is convenient to divide the curves into approximate initial, transient and steady-state phases. As swarming occurs in space, we use the terms "short-range" for the initial phase and "long-range" for the steady-state phase, respectively. The population size and the speed observed in the steady state are independent of the size of the starting population. It is noted that the steady state does not always appear, some model populations (such as the very small populations, or models with inefficient metabolism) die out after a transient swarming phase.
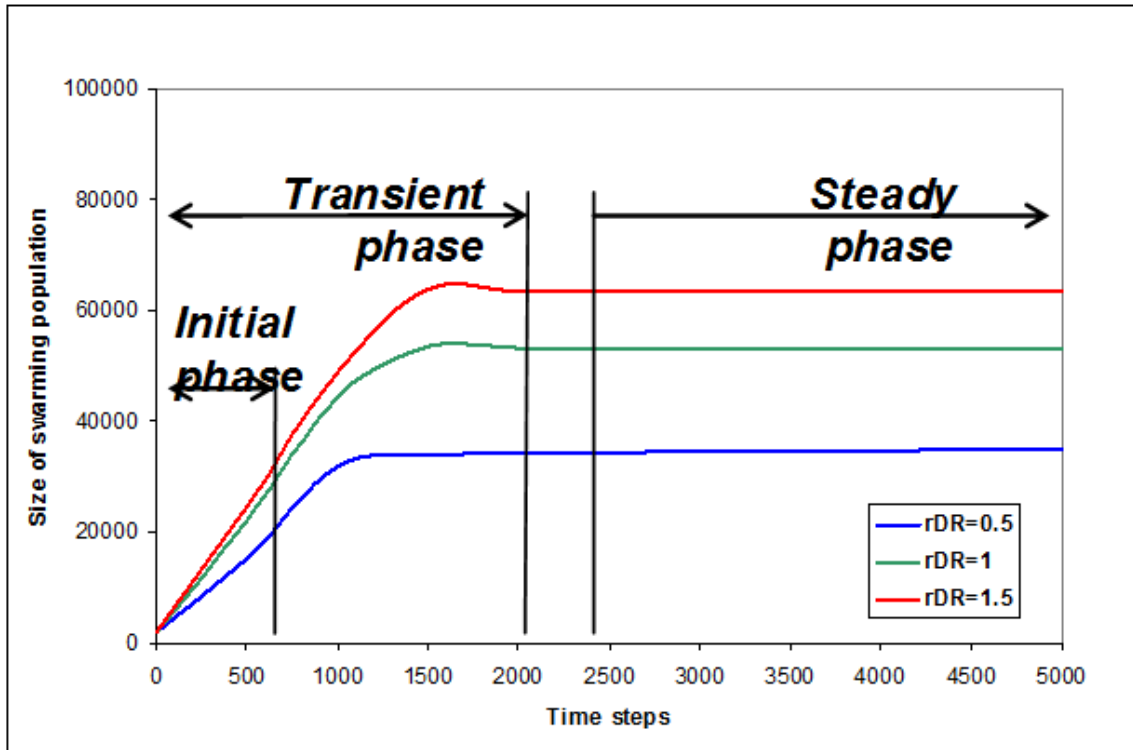
Figure 51 The dynamics of population size. *The initial phase corresponds to the time lapse from the solitary state to the activated state (when signal and production is increased and factors are released, but there is no swarming yet), the transient phase holds from the activated state until the onset of the swarming and the steady phase corresponds to a formed swarming front. One can see that the steady phase is stable. The plots are done for different relative division rates, as compared to the wild type.*

82
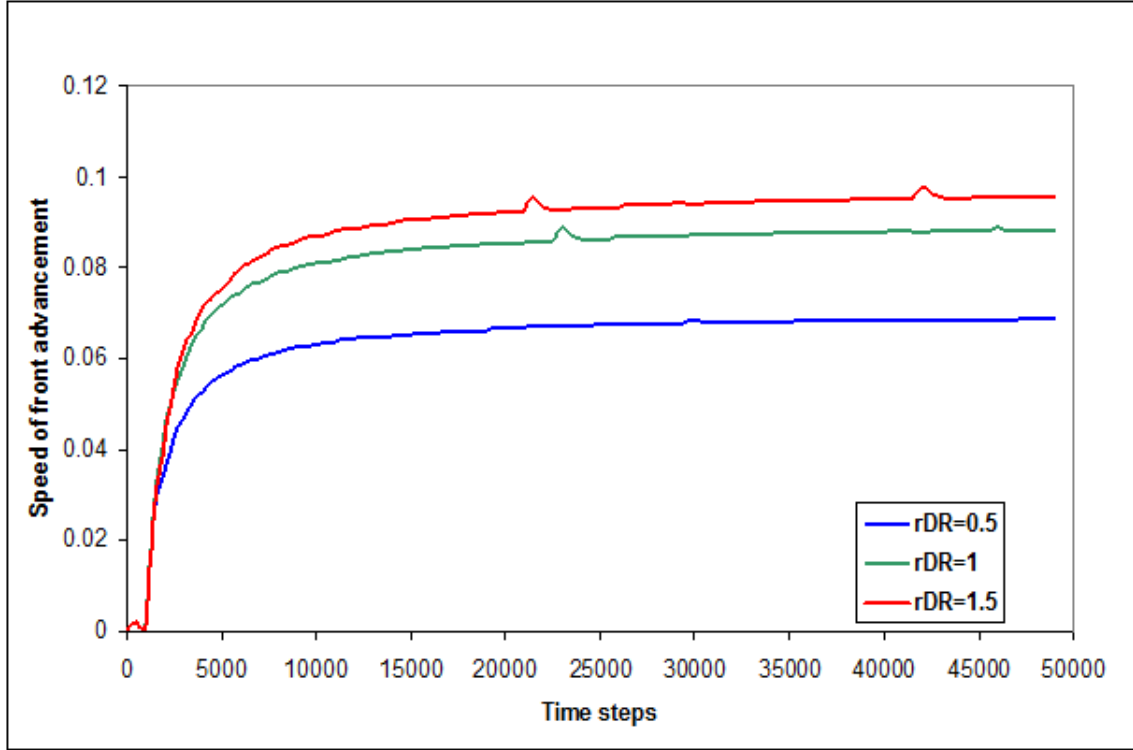
Figure 52 Dynamics of the speed of front advancement. *The speed of front is also becoming stable in time. The plots are done for different relative division rates, as compared to the wild type.*

Using the numerical values extracted from the modelling experiments, we can define the swarming fitness of a bacterium. For the characterization of swarming ability, we define swarming fitness as a measure of how efficiently a cell reaches a certain location in space. The swarming fitness of a cell type is proportional to the population size *p* and to the speed *v* observed in the steady state. We can then define the relative swarming fitness of a mutant as

$$SF^{rel} = \frac{p_m \times v_m}{p_{wt} \times v_{wt}}$$

Equation 3 The long range fitness

where the subscripts *m* and *wt* refer to mutant and wild type respectively. For the steady state, p and v values can read from the curves, e.g. we can read averaged values calculated for a longer period of time. The resulting SF value will characterize the mutant's ability to reach a long-range destination. In principle, SF_{long_range} is independent of the starting population size, nevertheless we

83

routinely calculated it by using equal starting populations for mutant and wt. The short range swarming fitness, on the other hand, refers to the ability of a cell to reach a destination in an early stage of swarming. Since we can approximately say that the movement of any colony at the onset is very small, so we use the approximation $v_m \approx v_{wt}$ .so the short-range relative swarming fitness can be calculated as

$$SF^{rel}_{short\_range} = \frac{p_m}{p_{wt}}$$

Equation 4 The short range fitness

, where p_m and p_{wt} indicate the swarming population size taken at a very early time point such as 200 time steps after the onset of swarming. As the short range fitness depends on the initial population sizes so we determined it using strictly identical initial starting populations, typically 1000 cells of 2 different models, distributed randomly. Examples of calculated values are shown in the inset in *Figure 54*.

Figure 54 shows examples of mutant models that differ in their relative division rates. In our model, the speed of the front movement is mainly regulated by the rate of division, which is, in turn, dependent on the amount of energy that the cells are capable of storing (saving) at each step. Therefore, cells that spend less energy (have less metabolic costs) will be more viable and compete out the cells that spend more energy, as shown by the example in *Figure 53*. In these calculations, the rate of division (average division per cell per time point) can be determined directly for the entire experiment by counting the divisions for a given population. The relative division rate, R, was then calculated by dividing the division rate of a mutant with that of the wild type. The energy consumption was calculated as $E(S) + E(F) + E(metabolism)$, and the relative energy consumption E was calculated by dividing the energy consumption of the mutant with that of the wild type. The energy consumption was then altered so as to produce mutant models with different relative swarming fitness values. The asymmetrical distribution of the two species within the front (insets 1 and 2 in *Figure 53*) shows that the less successful species is pushed towards the regions containing fewer nutrients.

The swarming fitness of a bacterium is proportional to its population size and with the speed of the front advancement. As both of these quantities are in arbitrary units of the model, it is more appropriate to calculate a relative swarming fitness in comparison with a reference, such as the

wild type. Using this relative fitness measure, one can construct models that grow faster or slower than the wild type. *Figure 53* shows the competition of such models. As can be expected, the fitter (faster growing) model simply outcompetes the less fit (slower growing) model.
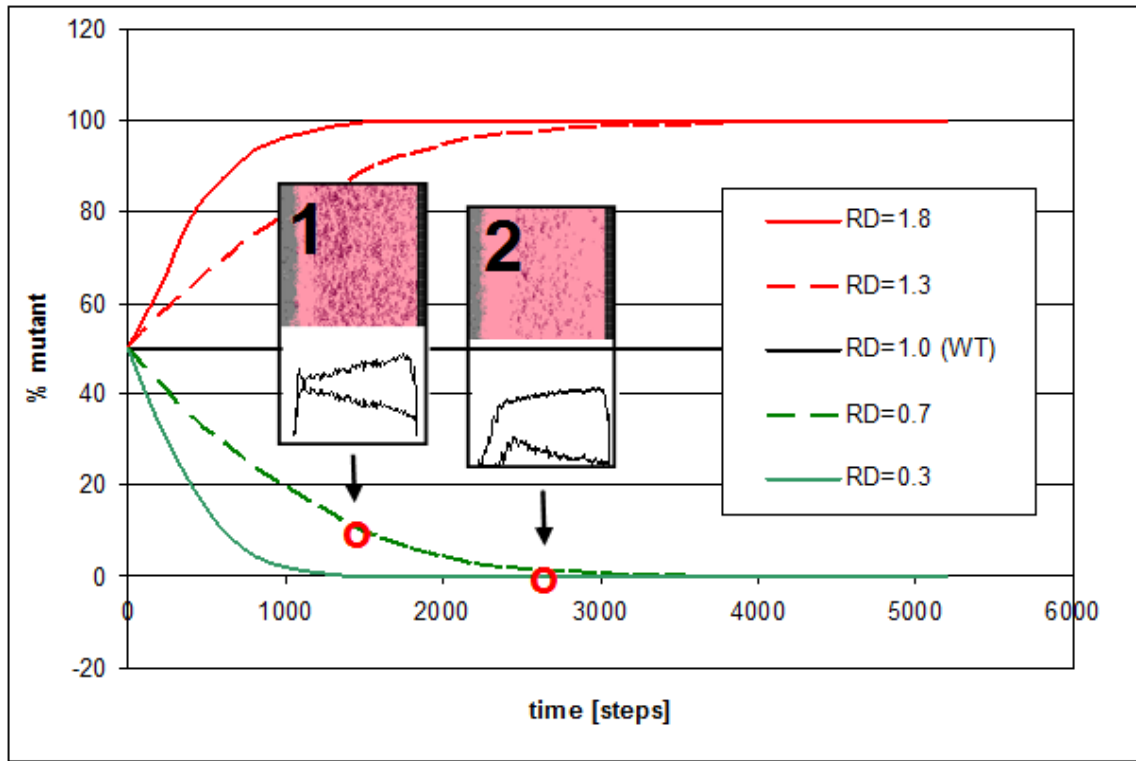
Figure 53 The initial phase of population growth. *The figure shows the simulated behavior of cells with different rates of division as compared to the wild-type P. aeruginosa. The results show the behavior of a population containing, at the start of the simulation, 50% wild type and 50% mutant cells. The mutant models are similar to the wild type except that the relative division rate is altered, as compared to the wild type. RD is the relative division rate, as compared to the wild type. Note that faster dividing cells compete out the slower dividing cells which are gradually are lost from the population. Two snapshots (insets A and B) show that the wild type (red) cells compete out a slower dividing mutant (blue). The distribution of the two populations (lower part of the insets) are different from each other, the faster dividing cells (top curves within the inset) are more abundant near the front, i.e. in the regions where more nutrients are available.*

85
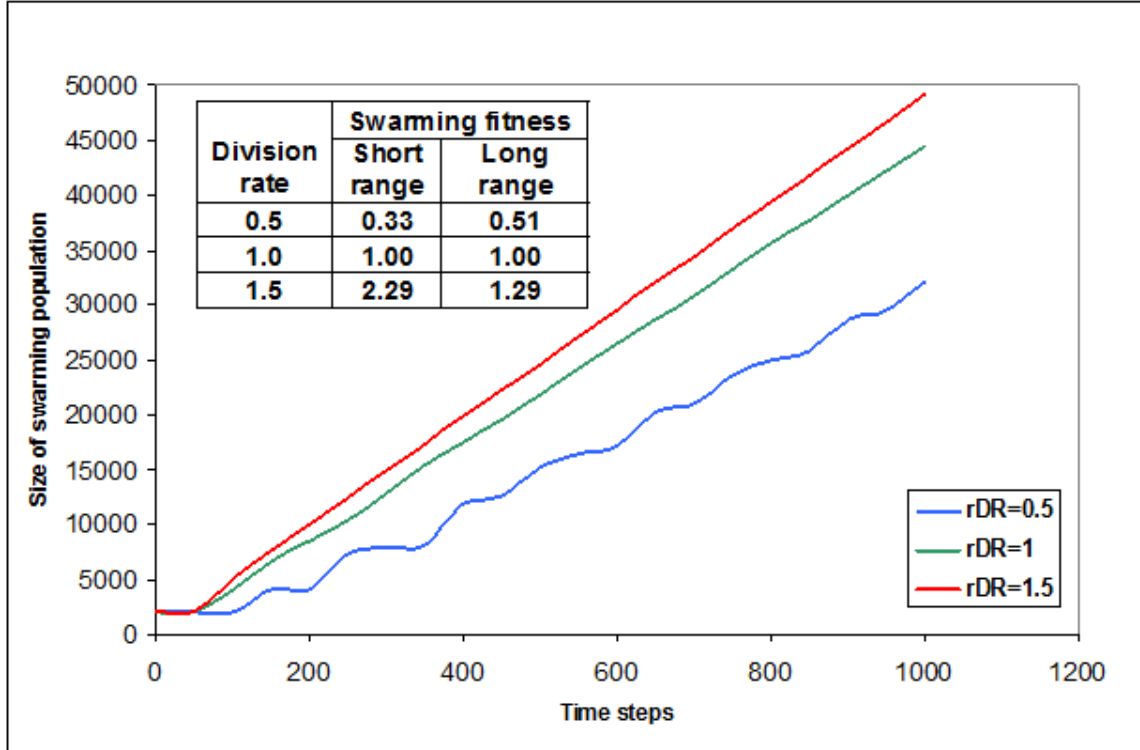
Figure 54 The phases of a simulation experiment. *The sketch shows the 1000 steps swarming fitness for different relative division rates as compared to the wild type. Both the short range fitness and the long range fitness increase with division rate, but in the case of high division rates there is an higher increase in short range fitness compared to the long range fitness.*

It is also worthwhile to note that nutrients, signals (information) and secreted factors (public goods) are asymmetrically distributed within the activation zone (*Figure 53*, insets) , i.e. some parts of the activation zone will be less favorable for growth than others. In accordance with this, we see the less fit cells accumulating in regions less abundant in nutrients and public goods (*Figure 53*).

Competition *in silico*. Overconsumption and evolutionary game dynamics of the bacteria colony model.

In our model system the cells are maintained by a flux of nutrients provided by diffusion. In other terms, their survival depends on a balance between nutrient consumption and diffusion. We can break this balance in two different ways: **a)** by decreasing the flux of nutrients (i.e. decreasing the nutrient concentration or decreasing the diffusion constant of the nutrients), or **b)** making the

cells over-consume nutrients. *Figure 55* shows a model of the latter strategy that leads to a collapse of the swarming population. As we increase nutrient intake to 10 fold as compared to the WT model, population size decreases 4 orders of magnitude, and the migration slows down. The kinetics shows increasing fluctuations – also seen in a variety of other, non-biological model systems – that finally leads to collapse. The long-range swarming fitness of the population (Equation 3) first increases with over-consumption but after a limit it decreases to zero (*Figure 56*, inset). It is important to note that the steady population in our longitudinal model corresponds to a colony that steadily grows in two dimensions. In other terms, the model predicts a steady, i.e. sustainable colony growth that can however be collapsed by over-consumption.
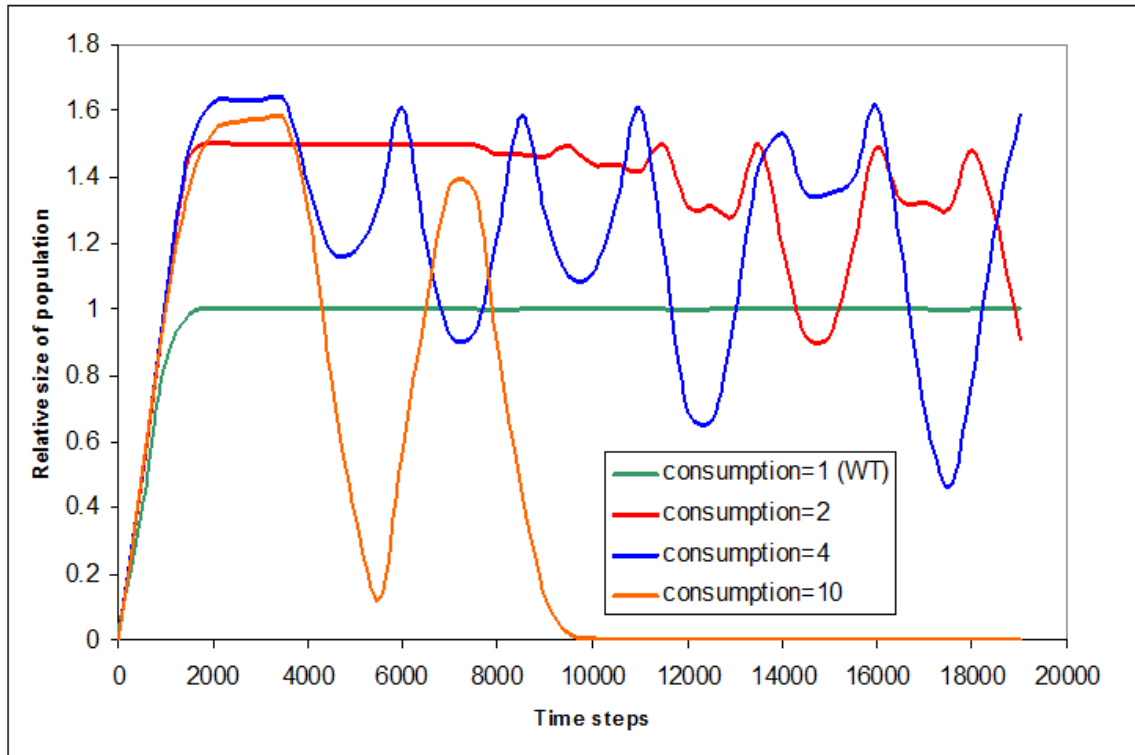
Figure 55 The effect of overconsumption on the relative size of the swarming population *(compared to that of the wild type).*
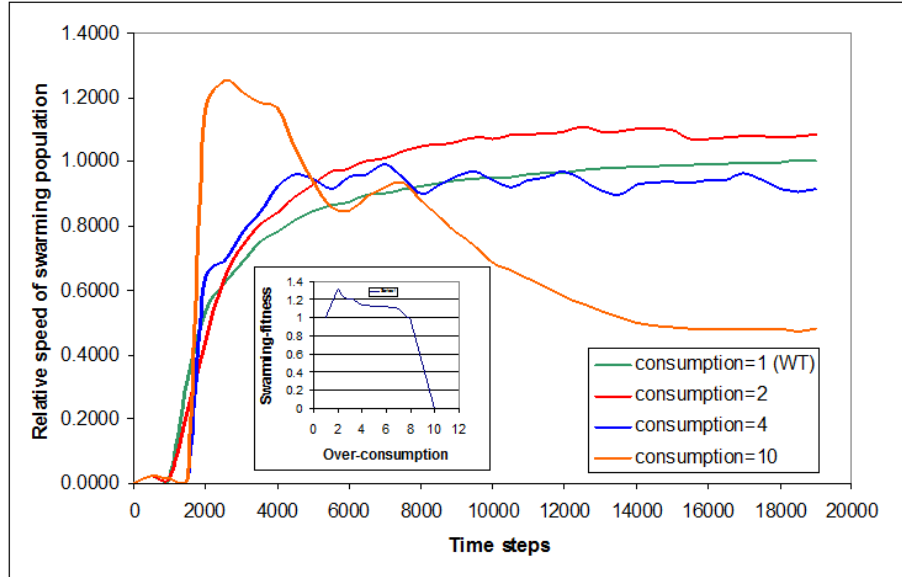
Figure 56 The effect of overconsumption on the relative speed of the swarming front (compared to that of the wild type). Overconsumption expressed as the nutrient intake divided by that of the WT. In the small inset, the long range swarming fitness is plotted as a function of overconsumption.

This model with a relatively simple set up explains certain behaviours as emergent without assuming any kind of in-built cell-cell interaction. But is this reflective of the natural situation in a proteobacterium with such a large genome? Cells potentially behave differently towards kin as against non-kin. More competition experiments are possible between wild type and mutants and even between different species and they are the subject of our current studies.

Agreement with other continuum or hybrid models

The simulations are performed in a 2D longitudinal track, which is almost in 1D. This approach was preffered for measuring the advancement of the active zone. However, if the model is correct, it should also reproduce the dendritic colony morphologies when extended to two dimensions. Using the computational framework such an extension is straightforward, just the modelled area should be increased. Dendritic morphologies develop because cell fronts move randomly and the initiated advancing fronts deplete the nutrients around them, in a typical reaction-diffusion mechanism. The realism of such dendritic patterns is questionable, however, since, for computational reasons, we can not model large populations of cells such as present on a

single agar plate. Moreover, we can suppose that not only a few genes, but a large part of the genome may contribute to such patterns. This is why we thought about limiting our interests and our interpretations to a smaller context, i.e. the onset of swarming.
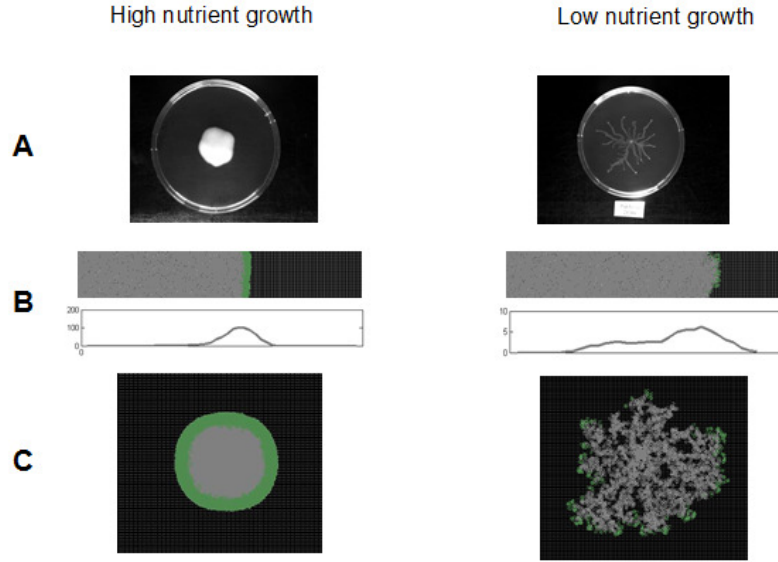
Figure 57 Growth at high and low nutrient concentrations. *A) Agar plates with high and low nutrient concentrations (see Methods section for details). The cells are wild-type P. aeruginosa PUPa3 cells. B) Longitudinal and C) 2D models at high (50 units) and low (10 unit) nutrient concentrations. The green dots indicate the living, swarming wild type cells. Note that the number of living cells is much lower in the irregular front-line found at low nutrient concentrations.*

Hybrid models can be analogous to continuum models described by reaction-diffusion equations. Since in our model the nutrient, signals and factor diffusion are independent of cell density, the immediate conversion of our model into a continuous model would be:

$$\frac{du_i}{dt} = D_i \nabla^2 u_i - R_i u_i$$

$$\frac{db}{dt} = D_i \nabla^2 b + f(\theta_i, u_i),$$

where f is a function that expresses the internal growth of the colony due to nutrient intake and factor and signal availability (its form could be as simple as $f = \sum_i \theta_i u_i$), $u_i, \theta_i, i = \overline{1,3}$ are the concentration of nutrients, signal and factors and the conversion rate of nutrients, signal and

factors into bacterial growth respectively. Additional layers of sophistication are the inclusion of concentration dependent conversion rates $\theta_i(b, u_i)$ or reaction terms $R_i(b, u_i)$.

However, our goal was to set up a hybrid model in which the behaviour of individual cells can be studied and compared with agar plate experiments, and do this without making further suppositions of self-orienting capabilities in bacteria. The construction and characterization of a continuum model for this system worth being studied as a future project.

The role of chemotaxis

One of the important findings is that under certain conditions it is possbile to obtain a moving „activation zone" of swarming cells that moves towards nutrients without invoking a directional response, such as chemotaxis. In the model the activation zone moves in one direction, because the nutrients are depleted behind the front. The extent of such depletion will be influenced by the diffusion constant of the nutrients. If nutrient diffusion is fast with respect to cell division rates and swarming rates, then the activation zone will not move towards the source of nutrients and the advancing front will die. i.e. its survival depends on a balance between nutrient consumption and diffusion, as it can be seen in *Figure 55*.

We purposefully disregarded chemotaxis because it mainly influences the shape of the dentritic patterns forming the so called chirallic patterns to a colony. Of course the addiction of a chemotaxis flux would be interesting in itself, and Cohen et al have studied how the addition of two chemotaxis-like principles modifies the overall shape of hte colony, using a reversed Lennard-Jones type of potential, with a short range attractive force and a long range repulsive force. In the studies of Cohen et al [96], chemotaxis was apparently necessary to obtain realistic colony morphologies, while, on the other hand, our model showed that salient features of swarming colonies can be reproduced without chemotaxis.

4.3 Bacterial colonies as adaptive spatial networks of interacting agents

To depict differences in the spatial network of interacting agents, three types of simulations were carried out, one in which the wild-type bacteria is alone and two in which it is co-swarming with mutants. The mutants are more or less dependent on the wild-type to initiate swarming, and the amount of dependence dictates the success of swarming. For one mutant, called signal-negative,

swarming is successful, while for the other, called signal-blind, swarming has collapsed and happens in occasional bursts, when the concentration of wild-type prepares the medium well enough.

Comparing the wild-type population to a random population

In the wild-type alone experiments the bacterial front is the thickest, since there are no competing mutants to benefit from the resourcefulness of the wild-type and to impede its swarming activation. As a consequence, the bacteria don't have much reason to form an efficient communication network. To observe this several types of plots were made by studying the properties of the communication network of the wild type at an arbitrary timestep (t=5000 was used in this case) and comparing them to network drawn form the random placement of bacteria. The random population has the same amount of individuals as the wild-type and spreads the same area as the wild-type.

The first plot shows a comparative zoomed in snapshot of the colony, compared to a snapshot with the same dimensions. We can observe that both populations display the same random pattern.
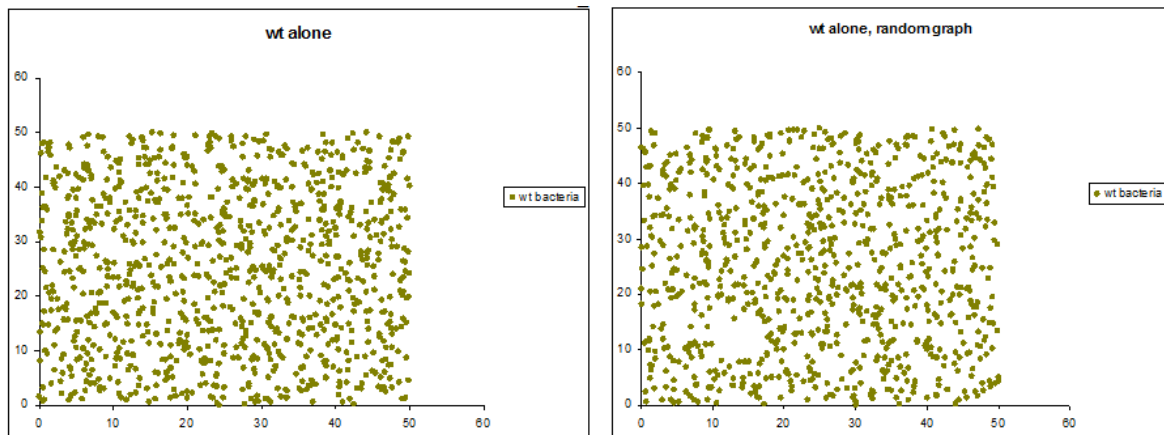
Figure 58 Zoomed in spatial snapshot of a population of wild type (WT) (left) and a random population of bacteria (SN) (right) bacteria, having the same amount of individuals as the wild-type and spreading the same area as the wild-type. *An agent based framework allows us to measure properties of the spatial network formed by bacteria.*

Next we display a table showing the mixing of the individuals. The populations have similar properties, the same average distance among individuals and a relatively high average density per square (we allowed a maximum of 10 cells per square in our experiments).

Table 8 Mixing results for wild-type compared to a random population.

Simulation type	Average distances	Average density (cells per square)
wt only	0.85	8.61
wt only, random wt	0.86	8.62

Finally we compute the number of connected components of the wt network and compare it to that of the random network. The plot is done for different distance thresholds. We can see the difference is not great; the two networks behave the same while increasing the distance thresholds. For small thresholds, there are few components of size higher than 5 (5 connected nodes). Increasing the threshold, more components of size greater than 5 are forming until reaching a critical threshold from which onward the graph becomes less fragmented and giant components are forming up. Eventually for very high thresholds the whole network graph becomes connected. At its peak, the number of components is slightly smaller in the simulation case compared to the random case. This is probably due to the fact that simulated swarming bacteria need room to move, so it is a side effect of swarming.
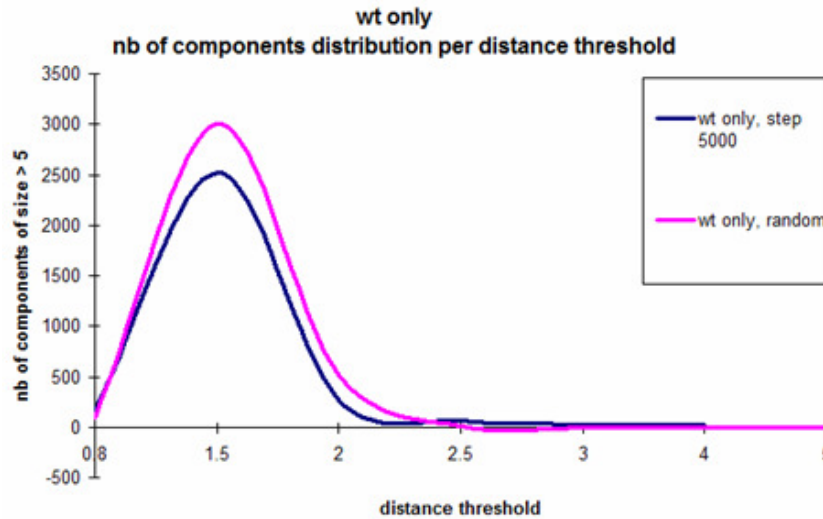
Figure 59 Connected components vs threshold plot for wild-type compared to a random population.

Comparing the wild-type population to a mixed population containing signal-negative mutants

Our next purpose was comparing the wild-type population to another simulated population containing the wild-type and a signal-negative mutant. The signal-negative releases the factors ('public goods' necessary for the swarming processes) as the wild-type does, but needs the signal released by the wild-type in order to activate a superior metabolic rate. Although the total population of wild-type and signal-negative bacteria is almost in the same amount as in the wild-type alone case, the coexistence of the wild-type with the dependent mutant forces the bacterial communication network to change properties.

Looking at the zoomed-in snapshot, we can see that both mutant and wild-type are well mixed. This is required by the signal consuming mutant and is not forced by our model, but it is an assembled consequence of co-swarming: without a good mixing, the signal released by the wild-type would not reach the mutant and would not trigger the superior metabolic rate, therefore the release of factors required for swarming.
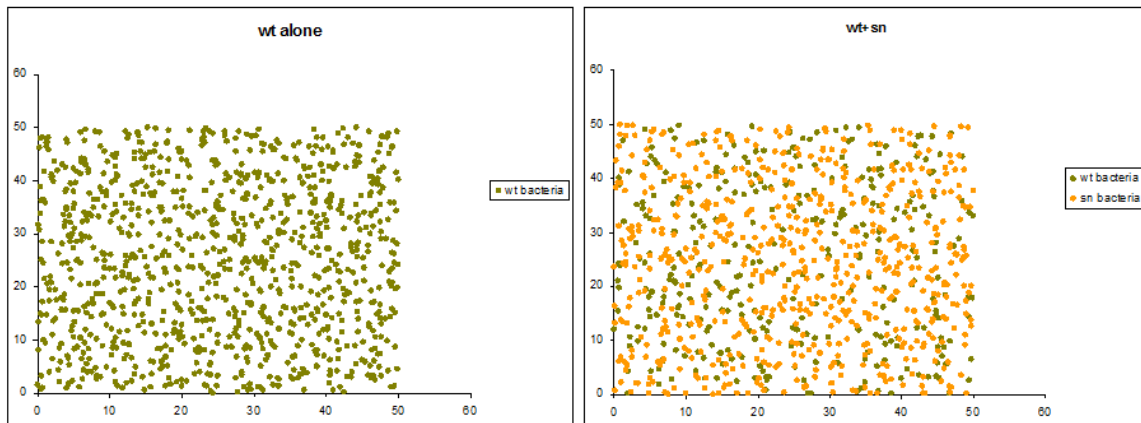
Figure 60 Zoomed in spatial snapshot of a population of wild type (WT) (left) and a co-swarming population of signal negative (SN) and wild-type bacteria (right). The signal-negative bacteria swarm only if they are well mixed with the wild-type.

From the mixing table the fitness of the wild-type is apparent from the small average distance relative to its very small density. As we can see, the total wild-type and signal-negative mixing is

closed to that of wild-type alone, so the front is thriving almost as well as that of the wild-type alone.

Table 9 Mixing results for wild-type compared to a co-swarming population of signal-negative and wild-type bacteria.

Simulation type	Average distances	Average density (cells per square)
wt only	0.85	8.61
wt+sn, wt only	1.47	2.82
wt+sn, all	1.01	7.95

To see the reliance of the signal-negative on the amount of signal around them, we devised a demonstrative simulation experiment, where a strain of signal negative (SN) bacteria competes with the wild-type. The signal negative mutant is dependent on the signal released by the wild-type in order to activate the process of swarming. If the quantity of available signal decreases then the mutant SN will not switch off from swarming into a non-active state and that will have adverse effects on its surviving fitness leading to a population decline.
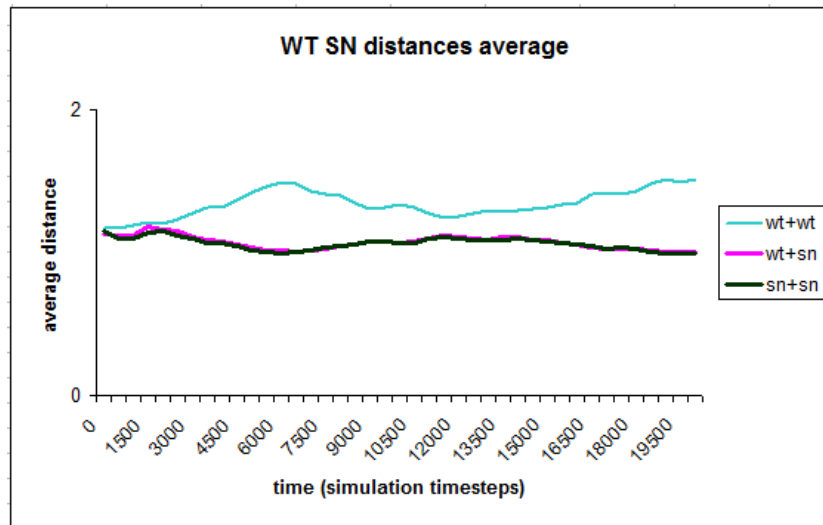
Figure 61 Time plot of the average nearest-neighbor distance for a competing colony of wild type (WT) and signal negative (SN) cells. When the average distances wt+sn is decreasing, thus the two species are well mixes, the sn takes advantage of its superior metabolic function and remains well mixed with its own kin (sn+sn follows closely the dynamics of wt+sn). The result is even more displacement for WT. However this will translate in decreased signal thus the swarming intensity decreases and WT is again favored.

94

The number of wt components in the mixed case is greater for higher thresholds, while if we take into account both the mutant and the wild-type, it follows the same distribution with the wt only. This suggests that the wild-type sub-network becomes more efficient, while being at the same time more sparse, to allow the development of the mutant.
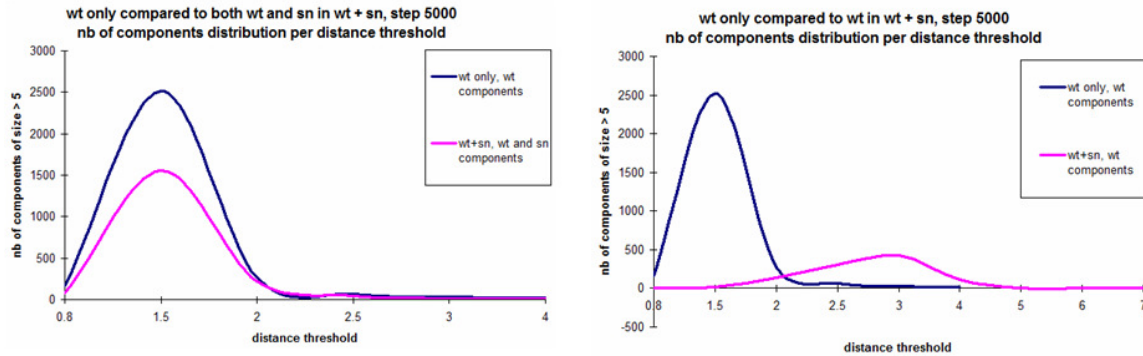
Figure 62 Connected components vs threshold plot for wild-type compared to a co-swarming population of signal-negative and wild-type bacteria. Left, in red, is the combined WT + SN while right, also in red, only the WT from the mixed experiment. The WT alone experiment is in blue.

Comparing a mixed population containing wild-type and signal-blind mutants with a random wild-type population

As previously mentioned, signal-blind mutants do not produce signal and do not produce factors. Therefore they are totally dependent on the wild-type to enter the swarming state, while on the same time being metabolically more fit (since they do not spend metabolic energy on releasing signals and factors). This translates into a thinner advancement front with local 'bursts' of signal-blind swarming bacteria. As one can see in the snapshot below, the wild type is much less uniformly distributed this time.
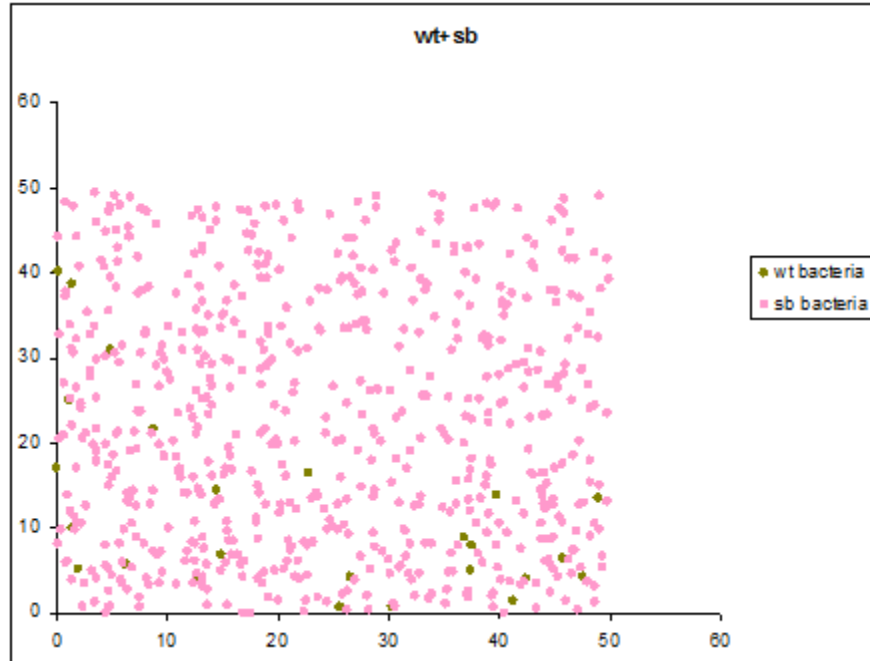
Figure 63 Zoomed in spatial snapshot of a co-swarming population of wild type (WT) and signal-blind (SB) bacteria. The wild type is much less uniformly distributed this time.

Compared to a random wild-type population of the same size and covering approximately the same surface, the non-uniform distribution of wild-type bacteria in the co-swarming experiment is clear. It is remarkable that, while having a very small average density, the wild-type still keep a low average distance.

Table 10 Mixing results for a co-swarming population of signal-blind and wild-type bacteria compared to a random population of wild-type.

Simulation type	Average distances	Average density (cells per square)
wt + sb, wt only	3.06	0.12
wt + sb, random wt	2.01	1.8

The number of components plot shows a clear peak at relatively high threshold, compared to the fuzziness of the random wild-type population (the fuzziness is due to the small number of individuals and the large area they cover).
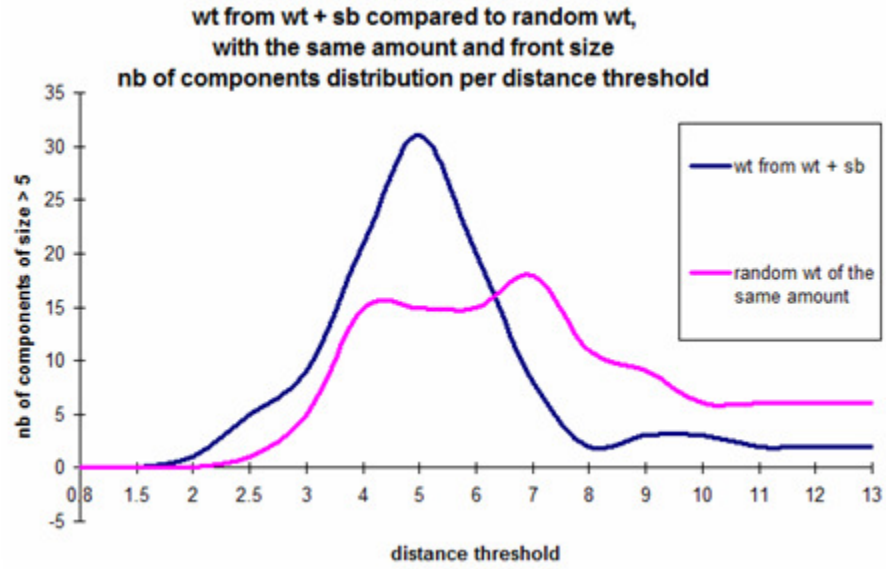
Figure 64 Connected components vs threshold plot for a co-swarming population of signal-blind and wild-type bacteria compared to a random population of wild-type. The wild-type from the co-swarming population has a clear peak and a smooth distribution by threshold level.

5. Conclusions

In this work we analyzed how networks evolve under the constraints of efficiency and robustness. We show that concomitant selection for efficiency and robustness influences the fundamental topological properties of the network, and that evolution under multiple attacks leads to distinct topologies.

Next we developed a model for the study of the swarming process in *Pseudomonas aeruginosa*. The model correctly predicts the behavior of genomic knockout mutants in which the QS genes responsible either for the synthesis (*lasI, rhlI*) or the sensing (*lasR, rhlR*) of AHL signals were inactivated.

An agent based model makes it possible to study how the signaling network kinetics influences the dynamic of a colony for single or co-swarming species, while also allowing for the study of the evolving communication network of the spatial conformation of individual bacteria.

We made experiments and described the various dynamic properties of a simulation experiment, computed the communication network of spatially interacting bacteria and studied its behavior. The use of multiple species makes our model an alternative to the evolutionary game dynamics methods used today to describe bacterial cooperation.

6. Acknowledgements

I thank my PhD coordinator, prof. Sandor Pongor for guiding my thesis; I thank my scientific collaborators and co-authors; I thank my BRC colleagues and ITC students for the joyful memories and I thank my wife Otilia.

7. Bibliography

[1]   R. Diestel, *Graph Theory (Graduate Texts in Mathematics)*, Springer, 2005.

[2]   U. Brandes and T. Erlebach, *Network Analysis Methodological Foundations*, Springer-Verlag Berlin Heidelberg, 2005.

[3]   D. Watts and S. Strogatz, "Collective dynamics of small-world networks," *Nature*, vol. 393, 1998, pp. 440-442.

[4]   M.E. Newman, *A measure of betweenness centrality based on random walks*, 2003.

[5]   S.P. Borgatti, "Centrality and network flow," *Social Networks*, vol. 27, 2005, pp. 55-71.

[6]   B. Ruhnau, "Eigenvector-centrality -- a node-centrality?," *Social Networks*, vol. 22, 2000, pp. 357-365.

[7]   S. Brin, "The Anatomy of a Large-Scale Hypertextual Web Search Engine," *Computer Networks and ISDN Systems*, 1998, pp. 107-117.

[8]   N. Przulj, D.A. Wigle, and I. Jurisica, "Functional topology in a network of protein interactions.," *Bioinformatics*, vol. 20, 2004, pp. 340-348.

[9]   D.S. Lee, J. Park, K.A. Kay, N.A. Christakis, Z.N. Oltvai, and A.L. Barabási, "The implications of human metabolic network topology for disease comorbidity.," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 105, 2008, pp. 9880-9885.

[10]   R. Albert and A. Barabasi, *Topology of evolving networks: local events and universality*, 2000.

[11]   P. Mahadevan, D. Krioukov, K. Fall, and A. Vahdat, *A Basis for Systematic Analysis of Network Topologies*, 2006.

[12]   B. Bollobas, *Random Graphs*, Cambridge University Press, 2001.

[13]   M.E. Newman, S.H. Strogatz, and D.J. Watts, "Random graphs with arbitrary degree distributions and their applications.," *Phys Rev E Stat Nonlin Soft Matter Phys*, vol. 64, 2001.

[14]   M.E. Newman, D.J. Watts, and S.H. Strogatz, "Random graph models of social networks," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 99, 2002, pp. 2566-2572.

[15]   A. Barabasi, "Scale-Free Networks: A Decade and Beyond," *Science*, vol. 325, 2009, pp. 412-413.

[16]   A. Vazquez, *Growing networks with local rules: preferential attachment, clustering hierarchy and degree correlations*, 2003.

[17]   W. Lee and K. Yang, "Applying Intelligent Computing Techniques to Modeling Biological Networks from Expression Data," *Genomics, Proteomics \& Bioinformatics*, vol. 6, 2008, pp. 111-120.

[18]   E. Alm and A.P. Arkin, "Biological networks," *Current Opinion in Structural Biology*, vol. 13, 2003, pp. 193-202.

[19]   M. Girvan and M.E. Newman, *Community structure in social and biological networks*, Proc. Natl. Acad. Sci. USA 99, 2001.

[20]   A. Blais and B.D. Dynlacht, "Constructing transcriptional regulatory networks," *Genes \& Development*, vol. 19, 2005, pp. 1499-1511.

[21]   Z. Hu, P.J. Killion, and V.R. Iyer, "Genetic reconstruction of a functional transcriptional regulatory network.," *Nat Genet*, 2007.

[22]   H. Jeong, B. Tombor, R. Albert, Z.N. Oltvai, and A.L. Barabási, "The large-scale organization of metabolic networks," *Nature*, vol. 407, 2000, pp. 651-654.

[23]   J. Förster, I. Famili, P. Fu, B.Ø. Palsson, and J. Nielsen, "Genome-scale reconstruction of the Saccharomyces cerevisiae metabolic network.," *Genome Res*, vol. 13, 2003, pp. 244-253.

[24]   V. Spirin, M.S. Gelfand, A.A. Mironov, and L.A. Mirny, "A metabolic network in the evolutionary context: Multiscale structure and modularity.," *Proc Natl Acad Sci U S A*, 2006.

[25]   O.S. Soyer, T. Pfeiffer, and S. Bonhoeffer, "Simulating the evolution of signal transduction pathways.," *J Theor Biol*, vol. 241, 2006, pp. 223-232.

[26]   H.A. Kestler, C. Wawra, B. Kracher, and M. Kühl, "Network modeling of signal transduction: establishing the global view.," *BioEssays : news and reviews in molecular, cellular and developmental biology*, vol. 30, 2008, pp. 1110-1125.

[27]   E. Yeger-Lotem, S. Sattath, N. Kashtan, S. Itzkovitz, R. Milo, R.Y. Pinter, U. Alon, and H. Margalit, "Network motifs in integrated cellular networks of transcription-regulation and protein-protein interaction.," *Proc Natl Acad Sci U S A*, vol. 101, 2004, pp. 5934-5939.

[28]   L. Yang, T. Vondriska, Z. Han, R. Maclellan, J. Weiss, and Z. Qu, "Deducing topology of protein-protein interaction networks from experimentally measured sub-networks," *BMC Bioinformatics*, vol. 9, 2008.

[29]   C. Lin, H. Juan, J. Hsiang, Y. Hwang, H. Mori, and H. Huang, "Essential Core of Protein−Protein Interaction Network in Escherichia coli," *Journal of Proteome Research*, vol. 8, 2009, p. 1925–1931.

[30]   V. Latora and M. Marchiori, "Economic small-world behavior in weighted networks," *The European Physical Journal B - Condensed Matter*, vol. 32, 2003, pp. 249-263.

[31]   V. Latora and M. Marchiori, "A measure of centrality based on network efficiency," *New J. Phys.*, vol. 9, 2007, p. 188+.

[32]   V. Latora and M. Marchiori, "Efficient Behavior of Small-World Networks," *Physical Review Letters*, vol. 87, 2001, p. 198701+.

[33]   B. Shargel, H. Sayama, I.R. Epstein, and Y. Bar-Yam, "Optimization of robustness and connectivity in complex networks.," *Physical review letters*, vol. 90, 2003.

[34]   J. Zhao and K. Xu, "Enhancing the robustness of scale-free networks," *J. Phys. A: Math. Theor.*, vol. 42, 2009.

[35]   L. Ma and P. Iglesias, "Quantifying robustness of biochemical network models," *BMC Bioinformatics*, vol. 3, 2002.

[36]   H. Kitano, "Biological robustness.," *Nat Rev Genet*, vol. 5, 2004, pp. 826-837.

[37]   V. Ágoston, P. Csermely, and S. Pongor, "Multiple weak hits confuse complex systems: A transcriptional regulatory network as an example," *Physical Review E*, vol. 71, 2005, p. 051909+.

[38]   J. Liu, Z. Wang, and Y. Dang, "Optimization of robustness of scale-free network to random and targeted attacks," *Modern Physics Letters B*, 2007.

[39]   R. Albert, H. Jeong, and A. Barabasi, "Error and attack tolerance of complex networks," *Nature*, vol. 406, 2000, pp. 378-382.

[40]   P. Holme, B.J. Kim, C.N. Yoon, and S.K. Han, "Attack vulnerability of complex networks," *Physical Review E*, vol. 65, 2002, p. 056109+.

[41]   S.N. Dorogovtsev and J.F. Mendes, "Evolution of networks," *Advances in Physics*, vol. 51, 2002, pp. 1079-1187.

[42]   T. Pfeiffer, O.S. Soyer, and S. Bonhoeffer, "The evolution of connectivity in metabolic networks.," *PLoS Biol*, vol. 3, 2005, p. 0.

[43] S.A. Teichmann and M.M. Babu, "Gene regulatory network growth by duplication.," *Nat Genet*, vol. 36, 2004, pp. 492-496.

[44] H. Bersini, T. Lenaerts, and F.C. Santos, "Growing biological networks: Beyond the gene-duplication model," *Journal of Theoretical Biology*, vol. 241, 2006, pp. 488-505.

[45] M. Koyutürk, Y. Kim, U. Topkara, S. Subramaniam, W. Szpankowski, and A. Grama, "Pairwise alignment of protein interaction networks.," *J Comput Biol*, vol. 13, 2006, pp. 182-199.

[46] S. Jones and J.M. Thornton, "Principles of protein-protein interactions.," *Proc Natl Acad Sci U S A*, vol. 93, 1996, pp. 13-20.

[47] J. Berg, M. Lässig, and A. Wagner, "Structure and evolution of protein interaction networks: a statistical model for link dynamics and gene duplications.," *BMC evolutionary biology*, vol. 4, 2004.

[48] B.P. Kelley, B. Yuan, F. Lewitter, R. Sharan, B.R. Stockwell, and T. Ideker, "PathBLAST: a tool for alignment of protein interaction networks.," *Nucleic Acids Res*, vol. 32, 2004, p. 0.

[49] M. Koyutürk, Y. Kim, S. Subramaniam, W. Szpankowski, and A. Grama, "Detecting conserved interaction patterns in biological networks.," *J Comput Biol*, vol. 13, 2006, pp. 1299-1322.

[50] R. Sharan and T. Ideker, "Modeling cellular machinery through biological network comparison.," *Nature biotechnology*, vol. 24, 2006, pp. 427-433.

[51] R. Milo, S. Shen-Orr, S. Itzkovitz, N. Kashtan, D. Chklovskii, and U. Alon, "Network Motifs: Simple Building Blocks of Complex Networks," *Science*, vol. 298, 2002, pp. 824-827.

[52] H.D. Kim, T. Shay, E.K. O'Shea, and A. Regev, "Transcriptional regulatory circuits: predicting numbers from alphabets.," *Science (New York, N.Y.)*, vol. 325, 2009, pp. 429-432.

[53] V. Venkatasubramanian, S. Katare, P.R. Patkar, and F. Mu, "Spontaneous Emergence of Complex Optimal Networks through Evolutionary Adaptation," *Computers and Chemical Engineering*, vol. 28, 2004, p. 1789–1798.

[54] J. Ash and D. Newth, "Optimizing complex networks for resilience against cascading failure," *Physica A: Statistical Mechanics and its Applications*, vol. 380, 2007, pp. 673-683.

[55] P. Oikonomou and P. Cluzel, "Effects of topology on network evolution," *Nat Phys*, vol. 2, 2006, pp. 532-536.

[56] J.C. Pommerville, *Alcamo's Fundamentals of Microbiology, 8th Edition*, Jones \& Bartlett Publishers, 2006.

[57] M.H. Zwietering, I. Jongenburger, F.M. Rombouts, and K. van 't Riet, "Modeling of the Bacterial Growth Curve.," *Appl Environ Microbiol*, vol. 56, 1990, pp. 1875-1881.

[58] K. Sauer, A.K. Camper, G.D. Ehrlich, J.W. Costerton, and D.G. Davies, "Pseudomonas aeruginosa Displays Multiple Phenotypes during Development as a Biofilm," *J. Bacteriol.*, vol. 184, 2002, pp. 1140-1154.

[59] Y.H. Dong and L.H. Zhang, "Quorum sensing and quorum-quenching enzymes.," *J Microbiol*, vol. 43 Spec No, 2005, pp. 101-109.

[60] W.C. Fuqua, S.C. Winans, and E.P. Greenberg, "Quorum sensing in bacteria: the LuxR-LuxI family of cell density-responsive transcriptional regulators.," *J Bacteriol*, vol. 176, 1994, pp. 269-275.

[61] N.A. Whitehead, A.M. Barnard, H. Slater, N.J. Simpson, and G.P. Salmond, "Quorum-sensing in Gram-negative bacteria," *FEMS Microbiology Reviews*, vol. 25, 2001, pp. 365-404.

[62] G.S. Shadel and T.O. Baldwin, "Identification of a distantly located regulatory element in the luxD gene required for negative autoregulation of the Vibrio fischeri luxR gene.," *The Journal of biological chemistry*, vol. 267, 1992, pp. 7690-7695.

[63] A.U. Viretta and M. Fussenegger, "Modeling the quorum sensing regulatory network of human-pathogenic Pseudomonas aeruginosa.," *Biotechnology progress*, vol. 20, 2004, pp. 670-678.

[64] M. Schuster and Peter, "A network of networks: Quorum-sensing gene regulation in Pseudomonas aeruginosa," *International Journal of Medical Microbiology*, vol. 296, 2006, pp. 73-81.

[65] S. Shen-Orr, R. Milo, S. Mangan, and U. Alon, "Network motifs in the transcriptional regulation network of Escherichia coli," *Nat Genet*, vol. 31, 2002, pp. 64-68.

[66] A. Goryachev, D. Toh, and T. Lee, "Systems analysis of a quorum sensing network: Design constraints imposed by the functional requirements, network topology and kinetic constants," *Biosystems*, vol. 83, 2006, pp. 178-187.

[67] J.M. Skerker and H.C. Berg, "Direct observation of extension and retraction of type IV pili," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 98, 2001, pp. 6901-6904.

[68] A.J. Merz, M. So, and M.P. Sheetz, "Pilus retraction powers bacterial twitching motility," *Nature*, vol. 407, 2000, pp. 98-102.

[69] T. Köhler, L.K. Curty, F. Barja, C. van Delden, and J.C. Pechère, "Swarming of Pseudomonas aeruginosa is dependent on cell-to-cell signaling and requires flagella and pili.," *Journal of bacteriology*, vol. 182, 2000, pp. 5990-5996.

[70] M.R. Parsek, D.L. Val, B.L. Hanzelka, J.E. Cronan, and E.P. Greenberg, "Acyl homoserine-lactone quorum-sensing signal generation," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 96, 1999, pp. 4360-4365.

[71] M.L. Watson, "Chemokines linking receptors to response," *Immunology*, vol. 105, 2002, pp. 121-124.

[72] A.B. Flavier, S.J. Clough, M.A. Schell, and T.P. Denny, "Identification of 3-hydroxypalmitic acid methyl ester as a novel autoregulator controlling virulence in Ralstonia solanacearum.," *Molecular microbiology*, vol. 26, 1997, pp. 251-259.

[73] S.L. McKnight, B.H. Iglewski, and E.C. Pesci, "The Pseudomonas quinolone signal regulates rhl quorum sensing in Pseudomonas aeruginosa.," *Journal of bacteriology*, vol. 182, 2000, pp. 2702-2708.

[74] M.T. Holden, S.R. Chhabra, R. de Nys, P. Stead, N.J. Bainton, P.J. Hill, M. Manefield, N. Kumar, M. Labatte, D. England, S. Rice, M. Givskov, G.P. Salmond, G.S. Stewart, B.W. Bycroft, S. Kjelleberg, and P. Williams, "Quorum-sensing cross talk: isolation and chemical characterization of cyclic dipeptides from *Pseudomonas aeruginosa* and other Gram-negative bacteria," *Molecular Microbiology*, vol. 33, 1999, pp. 1254-1266.

[75] M.G. Surette, M.B. Miller, and B.L. Bassler, "Quorum sensing in Escherichia coli, Salmonella typhimurium, and Vibrio harveyi: a new family of genes responsible for autoinducer production.," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 96, 1999, pp. 1639-1644.

[76] P. Williams, M. Camara, A. Hardman, S. Swift, D. Milton, V.J. Hope, K. Winzer, B. Middleton, D.I. Pritchard, and B.W. Bycroft, "Quorum sensing and the population-dependent control of virulence.," *Philosophical transactions of the Royal Society of London. Series B, Biological sciences*, vol. 355, 2000, pp. 667-680.

[77]   S. Netotea, I. Bertani, L. Steindler, A. Kerenyi, V. Venturi, and S. Pongor, "A simple model for the early events of quorum sensing in Pseudomonas aeruginosa: modeling bacterial swarming as the movement of an "activation zone,"" *Biology Direct*, vol. 4, 2009, p. 6+.

[78]   J.D. Helmann, F.R. Masiarz, and M.J. Chamberlin, "Isolation and characterization of the Bacillus subtilis sigma 28 factor.," *Journal of bacteriology*, vol. 170, 1988, pp. 1560-1567.

[79]   A.B. Flavier, M.A. Schell, and T.P. Denny, "An RpoS (sigmaS) homologue regulates acylhomoserine lactone-dependent autoinduction in Ralstonia solanacearum.," *Molecular microbiology*, vol. 28, 1998, pp. 475-486.

[80]   A. Latifi, M. Foglino, K. Tanaka, P. Williams, and A. Lazdunski, "A hierarchical quorum-sensing cascade in Pseudomonas aeruginosa links the transcriptional activators LasR and RhIR (VsmR) to expression of the stationary-phase sigma factor RpoS.," *Molecular microbiology*, vol. 21, 1996, pp. 1137-1146.

[81]   M. Whiteley, M.R. Parsek, and E.P. Greenberg, "Regulation of Quorum Sensing by RpoS in Pseudomonas aeruginosa," *J. Bacteriol.*, vol. 182, 2000, pp. 4356-4360.

[82]   R. Redfield, "Is quorum sensing a side effect of diffusion sensing?," *Trends in Microbiology*, vol. 10, 2002, pp. 365-370.

[83]   B. Hense, C. Kuttler, J. Müller, M. Rothballer, A. Hartmann, and J. Kreft, "Does efficiency sensing unify diffusion and quorum sensing?," *Nature Reviews Microbiology*, vol. 5, 2007, pp. 230-239.

[84]   M. Travisano and G.J. Velicer, "Strategies of microbial cheater control.," *Trends in microbiology*, vol. 12, 2004, pp. 72-78.

[85]   J. Kreft and S. Bonhoeffer, "The evolution of groups of cooperating bacteria and the growth rate versus yield trade-off," *Microbiology*, vol. 151, 2005, pp. 637-641.

[86]   A.S. Griffin, S.A. West, and A. Buckling, "Cooperation and competition in pathogenic bacteria.," *Nature*, vol. 430, 2004, pp. 1024-1027.

[87]   J. Maynard-Smith, *Evolution and the theory of games*, Cambridge University Press, 1982.

[88]   K. Heurlier, V. Denervaud, and D. Haas, "Impact of quorum sensing on fitness of Pseudomonas aeruginosa," *International Journal of Medical Microbiology*, vol. 296, 2006, pp. 93-102.

[89]   J.U. Kreft, C. Picioreanu, J.W. Wimpenny, and M.C. van Loosdrecht, "Individual-based modelling of biofilms.," *Microbiology (Reading, England)*, vol. 147, 2001, pp. 2897-2912.

[90]   M. Mimura, H. Sakaguchi, and M. Matsushita, "Reaction-diffusion modelling of bacterial colony patterns," *Physica A: Statistical Mechanics and its Applications*, vol. 282, 2000, pp. 283-303.

[91]   E. Ben-Jacob, I. Cohen, and H. Levine, "Cooperative self-organization of microorganisms," *Advances in Physics*, vol. 49, 2000, pp. 395-554.

[92]   E. Ben-Jacob, I. Cohen, O. Shochet, I. Aranson, H. Levine, and L. Tsimring, "Complex bacterial patterns.," *Nature*, vol. 373, 1995, pp. 566-567.

[93]   E. Ben-Jacob, O. Schochet, A. Tenenbaum, I. Cohen, A. Czirok, and T. Vicsek, "Generic modelling of cooperative growth patterns in bacterial colonies," *Nature*, vol. 368, 1994, pp. 46-49.

[94]   E. Ben-Jacob and H. Levine, "Self-engineering capabilities of bacteria.," *Journal of the Royal Society, Interface / the Royal Society*, vol. 3, 2006, pp. 197-214.

[95]   K. Kawasaki, "Modeling Spatio-Temporal Patterns Generated byBacillus subtilis," *Journal of Theoretical Biology*, vol. 188, 1997, pp. 177-185.

[96]   I. Cohen, E. Ben-Jacob, I. Golding, and Y. Kozlovsky, "Continuous And Discrete Models Of Cooperation In Complex Bacterial Colonies," *Fractals*, vol. 7, 1998, pp. 235-247.

[97]   P. Gerlee and A.R. Anderson, "A hybrid cellular automaton model of clonal evolution in cancer: The emergence of the glycolytic phenotype," *Journal of Theoretical Biology*, vol. 250, 2008, pp. 705-722.

[98]   C.W. Reynolds, "Flocks, Herds, and Schools: A Distributed Behavioral Model," *Computer Graphics*, 1987, pp. 25-34.

[99]   J.A. Shapiro, "The significances of bacterial colony patterns," *BioEssays*, vol. 17, 1995, pp. 597-607.

[100]   S. Childress, *Mechanics of Swimming and Flying (Cambridge Studies in Mathematical Biology)*, Cambridge University Press, 1981.

[101]   T. Vicsek, A. Czirók, E.B. Jacob, I. Cohen, and O. Shochet, "Novel Type of Phase Transition in a System of Self-Driven Particles," *Physical Review Letters*, vol. 75, 1995, pp. 1226-1229.

[102]   J. Lega and T. Passot, "Hydrodynamics of bacterial colonies: a model.," *Phys Rev E Stat Nonlin Soft Matter Phys*, vol. 67, 2003.

[103]   Z. Csahok and A. Czirok, "Hydrodynamics of bacterial motion," *PHYSICA A*, vol. 243, 1997, p. 304.

[104]   A. Czirók, E.B. Jacob, I. Cohen, and T. Vicsek, "Formation of complex bacterial colonies via self-generated vortices," *Physical Review E*, vol. 54, 1996, pp. 1791-1801.

[105]   S. Netotea and S. Pongor, "Evolution of robust and efficient system topologies," *Cellular Immunology*, vol. 244, 2006, pp. 80-83.

[106]   R.S. Kumar, N. Ayyadurai, P. Pandiaraja, A.V. Reddy, Y. Venkateswarlu, O. Prakash, and N. Sakthivel, "Characterization of antifungal metabolite produced by a new strain *Pseudomonas aeruginosa* PUPa3 that exhibits broad-spectrum antifungal activity and biofertilizing traits," *Journal of Applied Microbiology*, vol. 98, 2005, pp. 145-154.

[107]   T.S. Murray and B.I. Kazmierczak, "FlhF is required for swimming and swarming in Pseudomonas aeruginosa.," *Journal of bacteriology*, vol. 188, 2006, pp. 6995-7004.

[108]   P.C. Seed, L. Passador, and B.H. Iglewski, "Activation of the Pseudomonas aeruginosa lasI gene by LasR and the Pseudomonas autoinducer PAI: an autoinduction regulatory hierarchy," *J. Bacteriol.*, vol. 177, 1995, pp. 654-659.

[109]   C. Fuqua and P.E. Greenberg, "Listening in on bacteria: acyl-homoserine lactone signalling," *Nat Rev Mol Cell Biol*, vol. 3, 2002, pp. 685-695.

8. ABSTRACT OF THE THESIS

Network evolution models and bacterial communication

PhD. Thesis

Sergiu Netotea

Supervisor: **Dr. Sándor Pongor**

Biological Research Center of the Hungarian Academy of Sciences

Institute of Biophysics

Laboratory of Bioinformatics

INTRODUCTION

In this study we investigate models of network evolution. The network evolution models we propose are based on the rearrangement of links (''rewiring'') and follow the traditions of evolutionary modeling, i.e. optimizes a fitness function that combines various factors into one numerical index. Naturally, there are many ways to formulate and combine the components of the fitness function and testing the possibilities makes the process computationally expensive.

We start by defining the main computational measures by which most biological networks are analyzed and also present the main classes of network topologies. Next we describe the concepts of network efficiency and robustness and their application, and present the main results of the efforts done so far in the study of network evolution.

We next study the network evolution of a graph of interacting bacterial agents. We present the swarming colonies of environmental *Pseudomonas aeruginosa* PUPa3, present the quorum sensing network of this species, the process of swarming, and detail the main models that were used to study the colony dynamics.

AIMS OF THE STUDY

In our work we wanted to investigate the effect of rewiring on the global communication fitness of a network. Our first goal was to study the evolution of robust yet efficient network

topologies and to see if selecting mutations only for efficiency or only for attack tolerance (robustness) will influence network topology. We also wanted to study how efficient and robust biological network behave, and if multiple attack has an outcome on the overall topology. We wanted to model the onset of swarming in *Pseudomonas aeruginosa* by a simplified agent-based model that could allow us to study the properties of the emergent behavior of the colony. We wanted to predict the experimental behavior of genomic knockout mutants in which the QS genes responsible either for the synthesis (*lasI, rhlI*) or the sensing (*lasR, rhlR*) of AHL signals were inactivated. We next wanted to study the interaction of bacteria by modeling their spatial dynamics as an evolving graph of interacting bacterial agents.

METHODS

To perform studies of network evolution we describe two algorithmic approaches in which all parameters are treated essentially as constraints. In the first approach, termed the random selection algorithm, a mutation is selected if all of its parameters exceed or at least reach the corresponding values of the previous state, so there are no tunable parameters. This approach is computationally efficient so it allows one to study a wide range of phenomena. The other approach is a genetic algorithm that is used to discover fitness-optimized network structures by avoiding local maxima.

Next we propose a model for bacterial colony dynamics that is used to explain experimental data related to the onset of swarming in environmental *Pseudomonas aeruginosa* PUPa3. The process was described with a simplified computational model in which cells in random motion communicate via a diffusible signal (representing *N*-acyl homoserine lactones, AHL) as well as a diffusible, secreted factors (enzymes, biosurfactans, i.e. "public goods") that regulate the intensity of movement and metabolism in a threshold-dependent manner. As a result, an "activation zone" emerges in which nutrients and other public goods are present in sufficient quantities, and swarming is the spontaneous displacement of this high cell-density zone towards nutrients and/or exogenous signals.

We next study the interaction of the bacteria based on inter-species distances and study the evolution dynamics of the graph of interacting bacterial agents.

RESULTS AND DISCUSSION

We explain how the choice of network size influences the resulting topologies; we describe the result of the random selection algorithm by measuring several network parameters. We examine how multiple node attacks changes the dynamics of the network evolution and its outcome. We show there are correlations between several node properties and the degree and explain why that happens. We make motifs and path correlation analysis and study the convergence to highly optimized structures.

Next we discuss the basic properties of the *in silico* agent-based model we have proposed for the bacterial colony of *P. aeruginosa* and compare its swarming *in vivo* and *in silico*. We make several dynamic measurements to demonstrate the capabilities of our model, and then we address the issue of competition among different mutants in which the QS genes responsible either for the synthesis (*lasI, rhlI*) or the sensing (*lasR, rhlR*) of AHL signals were inactivated, and we compare our results to laboratory results. We also discuss the agreement with other continuum or hybrid models and the avoidance of chemotaxis in studying the colonial dynamics.

We present the distance based communication bacterial graph constructed from our simulation model and study its evolution and dynamics.

CONCLUSIONS

We show that concomitant selection for efficiency and robustness influences the fundamental topological properties of the network, and that evolution under multiple attacks leads to distinct topologies.

The model correctly predicts the behavior of genomic knockout mutants in which the QS genes responsible either for the synthesis (*lasI, rhlI*) or the sensing (*lasR, rhlR*) of AHL signals were inactivated.

An agent based model makes it possible to study how the signaling network kinetics influences the dynamic of a colony, while also allowing for the study of the evolving communication network of the spatial conformation of individual bacteria.

LIST OF PUBLICATIONS

1. Netotea S, Bertani I, Steindler L, Kerényi Á, Venturi V, Pongor S, (2009) A simple model for the early events of quorum sensing in Pseudomonas aeruginosa: modeling bacterial swarming as the movement of an "activation zone", *Biol Direct.* 4(1):6.

2. Kertész-Farkas A, Dhir S, Sonego P, Pacurar M, Netotea S, Nijveen H, Kuzniar A, Leunissen J, Kocsor A, Pongor S (2007) Benchmarking protein classification algorithms via supervised cross-validation, *J Biochem Biophys Methods* 35:1215-1223.

3. Netotea S, Pongor S (2006) Evolution of robust and efficient system topologies. *Cell Immunol* 244(2):80-83.

9. A DISSZERTÁCIÓ MAGYAR NYELVŰ ÖSSZEFOGLALÁSA

Hálózati evolúciós modellek és bakteriális kommunikáció

PhD. tézis

Sergiu Netotea

Témavezető: **Dr. Sándor Pongor**

MTA Szegedi Biológiai Központ

Biofizikai Intézet

Bioinformatika Csoport

BEVEZETÉS

Jelen dolgozatban a hálózat-evolúció modelljeit tanulmányoztuk. Az általunk javasolt modellek az élek átrendeződésén, ún. áthuzalozáson alapulnak és az evolúciós modellezés alapelemeit követik, azaz egy fitness függvényt optimalizálnak. A fitness függvény az alkalmazás elemeinek numerikus jellemzése. Ennek a függvénynek a megszerkesztésére sok lehetőség van és a tesztelési lehetőségek nagy száma miatt az evolúciós modellezés rendkívül nagy számításigényű folyamat.

A dolgozat első részében a biológiai hálózatokban leggyakrabban előforduló mértékeket és hálózat topológiákat mutatjuk be. Ezt követően a hálózati hatékonyság és robusztusság (támadás tolerancia) fogalmait, és azok alkalmazásait ismertetjük, valamint bemutatjuk a hálózat evolúció legfontosabb eredményeit.

A dolgozat második részében kölcsönható baktérium-ágensek gráfjának evolúcióját vizsgáljuk. Bemutatjuk a *Pseudomonas aeruginosa* PUPa3 baktériumfaj rajzó kolóniáit, valamint ismertetjük a fajra jellemző quorum sensing (QS) hálózatot és a rajzás folyamatát. Végül a kolónia dinamikájának legfőbb modelljeit ismertetjük.

CÉLKITŰZÉSEK

Munkánkban, tanulmányoztuk az áthuzalozás hatását a hálózat globális kommunikációs fitneszére. Elsődleges célunk a robusztus, mégis hatékony hálózat-topológiák evolúciójának tanulmányozása volt, valamint annak a vizsgálata, hogy egy adott mutáció hogyan befolyásolja a hálózat kommunikációs hatékonyságát vagy robusztusságát. Továbbá, azt is tanulmányoztuk, hogyan viselkedik egy hatékony és robusztus biológiai hálózat, és milyen hatása van az összetett támadásnak a hálózat topológiára nézve.

A rajzás folyamatára egy modellt hoztunk létre a *Pseudomonas aeruginosa* baktérium esetén egy egyszerűsített ágens alapú modellel, mely alkalmas volt a kolónia emergens tulajdonságainak a tanulmányozására. Célunk volt megjósolni knockout mutáns kolóniák viselkedését, melyekben a szintetizáló (lasI, rhlI), ill. az AHL jelmolekulák érzékeléséért felelős QS gének voltak inaktiválva. A következőkben a baktérium interakciókat tanulmányoztuk azok térbeli modellezésével. A modellünk: egymással kölcsönhatásban levő baktérium ágensek evolváló gráfja volt.

MÓDSZEREK

A hálózat evolúció vizsgálata érdekében két algoritmikus megközelítést mutatunk be, melyekben minden paraméternek megszorító szerepe van. Első megközelítésünkben, melyet „véletlen szelekciós" algoritmusnak neveztünk el, egy mutációt akkor választunk ki, ha minden paraméterének értéke legalább eléri az előző állapotbeli értéket, tehát nincs hangolandó paraméter a modellben. Ez az algoritmus hatékony, így széleskörben alkalmas jelenségek tanulmányozására. Második megközelítésben egy genetikus algoritmust vizsgáltunk, mellyel fitness-re optimalizált hálózatokat kerestünk elkerülve a lokális maximumokat.

A dolgozat második részében baktérium kolóniák dinamikájának leírására javaslunk egy modellt, melyet *Pseudomonas aeruginosa* PUPa3 rajzásával kapcsolatos kísérleti adatok megmagyarázására használunk. A rajzást egy egyszerű számítási modellel írtuk le, melyben a sejtek véletlen mozgást végeznek. A sejtek közti kommunikációt diffundáló jelek (*N*-acil homoszerin lakton, AHL molekulák megfelelői) és diffundáló, termelt faktorok (enzimek, felület-aktív anyagok megfelelői) segítségével valósítjuk meg. A jelek és faktorok küszöbszerűen befolyásolják a mozgás intenzitását és a metabolizmust. Ennek eredménye az ún. „aktivációs zóna" kialakulása, melyben a tápanyagok, jelek, faktorok elégséges mennyiségben vannak jelen, és a rajzás, mely során ez a nagy sejtsűrűségű zóna spontán elmozdul a táplálék, illetve az exogén

jelek irányába. A következőkben a baktérium interakciót tanulmányoztuk. A baktérium interakciót a fajok közötti távolságok alapján határozzuk meg, majd az így keletkezett hálózat evolúciós dinamikáját tanulmányozzuk.

EREDMÉNYEK ÉS ÖSSZEFOGLALÁS

Megmutattuk, hogyan befolyásolja a háló méretének megválasztása az eredményül keletkező struktúrákat, és bemutatjuk a véletlen szelekció algoritmusának eredményeit több háló-paramétert vizsgálva. Továbbá megvizsgáltuk, hogyan változtatja az összetett támadás a hálózat evolúció dinamikáját és az eredményt. Majd kimutattuk, hogy korreláció van több csomópont tulajdonság és a fokszám között, és megmagyarázzuk, miért történik ez. Motívumokat és út-korrelációt elemeztünk és vizsgáltuk a túloptimalizált hálók struktúrájának konvergenciáját.

A következőkben bemutatjuk az *in silico* ágens alapú modellek alaptulajdonságait, melyet a *Pseudomonas aureginosa* baktérium kolónia modellezésére javasoltunk, majd összehasonlítjuk az *in vivo* és *in silico* rajzást. Több dinamikus mérést végeztünk, hogy demonstráljuk a modellünk képességeit, majd megvizsgáljuk, hogyan befolyásolja a QS jelek szintéziséért felelős gének (*lasI, rhlI*) illetve a a jelek érzékeléséért felelős gének (*lasI, rhlI*) inaktiválása a modellek viselkedését, majd összevetjük eredményeinket a laboratóriumi eredményekkel. Továbbá, összehasonlítottuk a modellünket más kontinuum/hibrid modellekkel, illetve tanulmányoztuk a kolónia dinamikáját a kemotaxis jelensége nélkül.

Egy távolság alapú baktérium kommunikációs gráfot építünk a szimulációs modellünkre és vizsgáljuk annak evolúcióját és dinamikáját.

KÖVETKEZTETÉSEK

Megmutattuk, hogy a hatékonysággal és robusztussággal járó szelekció befolyásolja a háló alapvető topológiai tulajdonságait, és az evolúció összetett támadás esetén különböző topológiákhoz vezet.

A modell helyesen becsli a génkiütéses mutánsok viselkedését, melyben a rajzásért felelős géneknek az AHL molekulák szintéziséért (*lasI, rhlI*) vagy érzékeléséért (*lasR, rhlR*) felelős részét inaktiváltuk.

Egy ágens alapú modell lehetővé teszi a hálózati kinetika a kolónia dinamikájára való hatásának tanulmányozását, továbbá megengedi a baktériumok térbeli alakzatából kialakuló kommunikációs hálózat tanulmányozását is.

PUBLIKÁCIÓK

1. Netotea S, Bertani I, Steindler L, Kerényi Á, Venturi V, Pongor S, (2009) A simple model for the early events of quorum sensing in Pseudomonas aeruginosa: modeling bacterial swarming as the movement of an "activation zone", *Biol Direct.* 4(1):6.

2. Kertész-Farkas A, Dhir S, Sonego P, Pacurar M, Netotea S, Nijveen H, Kuzniar A, Leunissen J, Kocsor A, Pongor S (2007) Benchmarking protein classification algorithms via supervised cross-validation, *J Biochem Biophys Methods* 35:1215-1223.

3. Netotea S, Pongor S (2006) Evolution of robust and efficient system topologies. *Cell Immunol* 244(2):80-83.