# 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 PhD School*, *University* of *Szeged*

# 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. The global efficiency 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 nodes i and j. It is known that complex networks are more vulnerable to targeted attack and less vulnerable to random attack compared to random graph models. We define robustness as the capacity of a network to survive attacks. 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.

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. 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.

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*)

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.

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. We present the bacterial colonies as an interspecies and intraspecies communication networks, 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.

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 several coupled reaction-diffusion equations. Hybrid models use a continuum description for the growth medium as well as for the solutes, and individual descriptions of bacteria. 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.

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, 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.

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.

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.

We 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.

## 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

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 approach is computationally efficient so it allows one to study a wide range of phenomena.

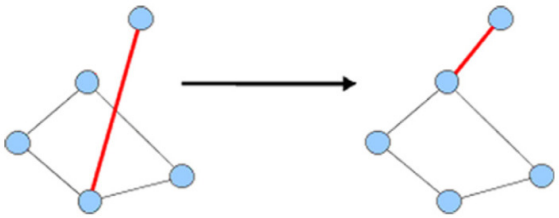

***Figure 1 A simple model of network evolution.*** *At each step an edge is rewired at random. If the new topology is more fitted according to the selected fitness criteria, the new network is used for the next step.*

A second method used to develop optimal solutions was based on a genetic algorithm. 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 with uniform probability across the list of edges and rewiring it. 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, as slow convergence is recommended for genetic algorithms. Due to the fact that mutations are added to the beginning of the list, we can establish the evolutionary history of a graph.

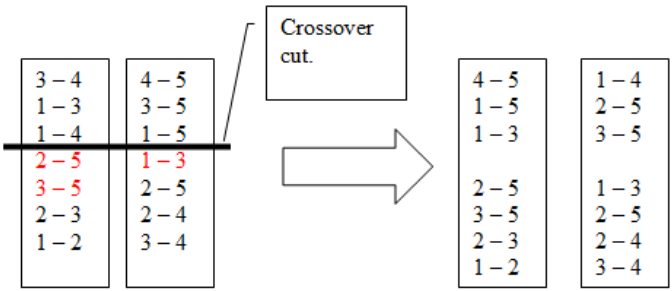


***Figure 2 Crossover example on a graph of 5 nodes and 7 edges****. The edges marked by red under the crossover cut line included in the crossover as well because some of the edges above the cut already belong to the original graph.*

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 S (representing *N*-acyl homoserine lactones, AHL) as well as a diffusible, secreted factors F (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 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.

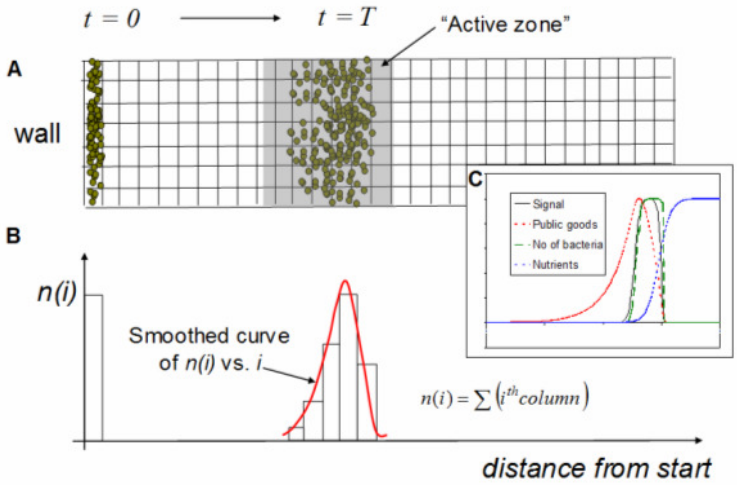


***Figure 3 Model outline.*** *The model describes the movement of cells on a longitudinal segment of the plane, discretized into squares (A). On the longitudinal sides, the track has periodic boundary conditions with respect to cell movement and diffusion. At the beginning (t = 0), the cells are placed 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).*

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 reaction-diffusion equation:

$\frac{du}{dt} = D\nabla^2 u - Ru$ , where $D$ and $R$ are the uniform diffusion and decay constants, respectively.

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. 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 bacteria agents present at the specific timestep and with a number of links dependent on the distance threshold being used.

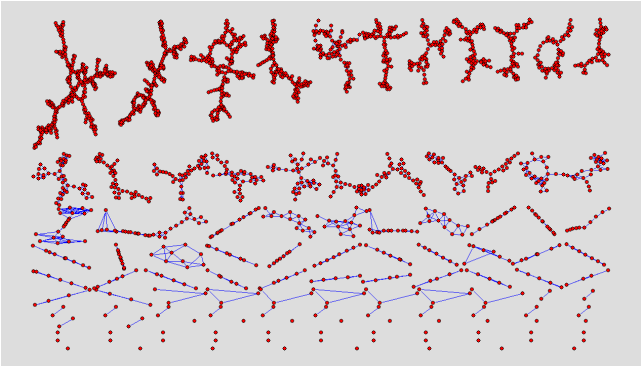

***Figure 4 Example of the bacterial communication network, computed from simulation data, for a small distance threshold.*** *The network is formed by many non-connected subgraphs (connected components). If the threshold is sufficiently increased the number of connected components becomes smaller, and eventually the network becomes fully connected.*

## 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.

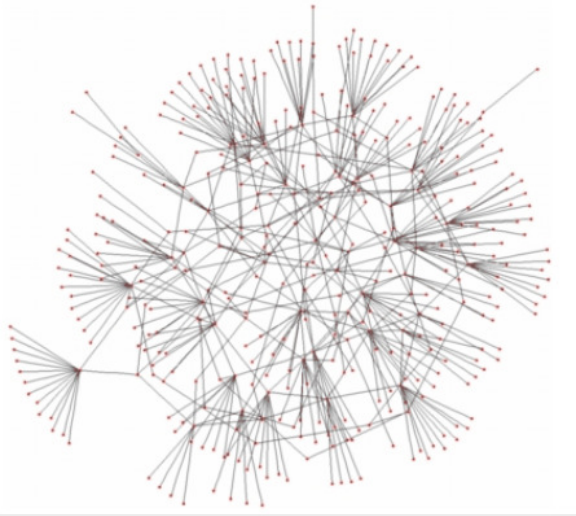

***Figure 5 Example of a 400 nodes / 480 edges evolution experiment.*** *Networks that keep node/edge proportionality have the same outcome in our evolution experiments. For computational purposes we favored smaller networks.*

The main result in this section concerns the core-periphery scale of the evolved networks. Networks evolved with larger emphasis on robustness to multiple attacks will develop a larger core while efficiency favors a larger periphery. Since the two concepts are opposed, complex networks are usually a trade-off between a large core and a large periphery.

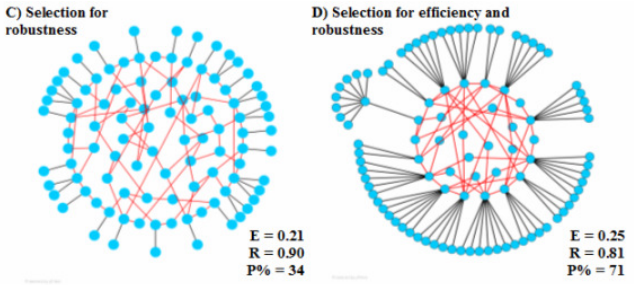


***Figure 6 Schematics of the core-periphery trade-off.*** *Left, selection for robustness favors larger core while selection for both robustness and efficiency (right) increases the periphery.*

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.

The definition of QS is that cells respond to cell density, the model-population in fact acting as a density switch. At a given cell density the cells get activated i.e. they start to produce factors and subsequently the cells also start to swarm. 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 genetic networks underlying QS are considered to act as a two-state switch. 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.

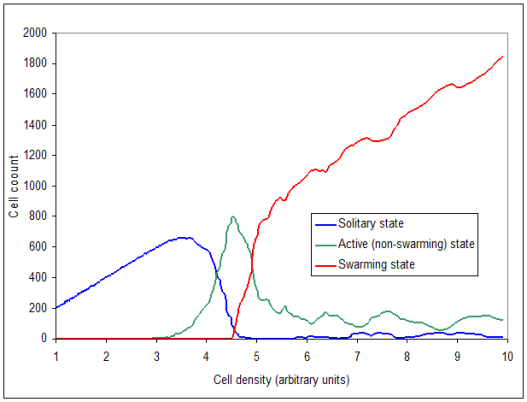


***Figure 7 Dependence of the cell's state on cell's density.*** *Note that beyond a certain cell's density level, nearly all cells switch to the swarming state, i.e. the model acts as a density switch. The panel shows the first steps of a simulation starting from a very small population.*

The behavior of wild-type *P. aeruginosa* PUPa3 as well as its mutants is compared *in vivo* and *in silico***.** 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, and **ii)** the simplified regulatory scheme built into the agent-based model provides a qualitatively adequate description of the events.

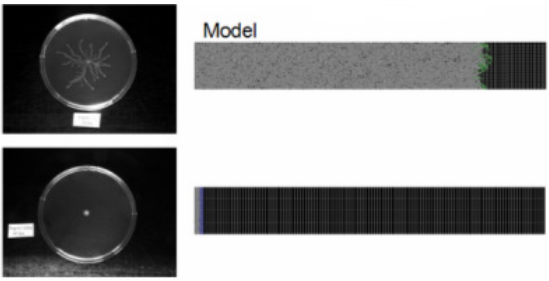


***Figure 8 Example of comparing the simulation model with the wet lab experiments.*** *Mutants that cannot activate the swarming abilities are inhibited from growth.*

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. **A** model of the latter strategy will lead to a collapse of the swarming population.

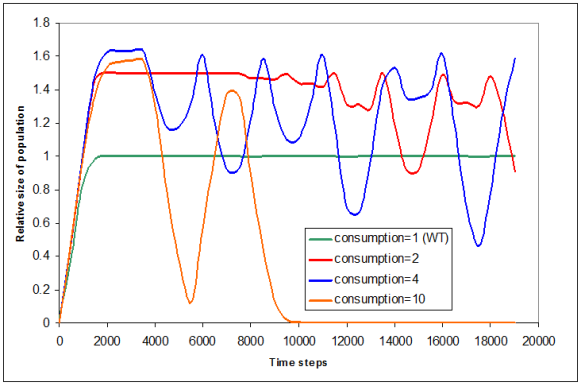


***Figure 9 The effect of overconsumption on the relative size of the swarming population*** *(compared to that of the wild type) A common effect in ecology, over-consumption causes over-division and leads to a large variation in the population number.*

We present the distance based communication bacterial graph constructed from our simulation model and study its evolution and dynamics. We show how deprived but essential bacteria are enforcing their communication network by forming powerful local groups while being sufficiently spread in order to survive, and discuss the implications of studying altruists and cheaters for the evolutionary game dynamics.

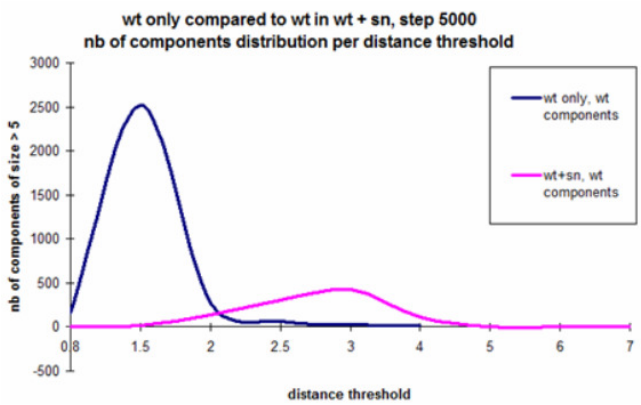


***Figure 10 Connected components vs threshold plot for wild-type compared to a co-swarming population of signal-negative and wild-type bacteria.*** *Left, in red, is the combined WT + SN while right, also in red, only the WT from the mixed experiment. The WT alone experiment is in blue.*

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.

## 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, 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. 2009 Feb 12; 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.
3. Netotea S, Pongor S (2006) Evolution of robust and efficient system topologies. Cell Immunol 244(2):80-83.

**ACKNOWLEDGEMENTS**

I thank the PhD commission and the organizers of my PhD defense.

I thank Prof. Sándor Pongor for giving me the oportunity to fulfill a life dream, for his guidance of my PhD studies and for the chance I had to observe his impressive leadership.

I thank Dr. Zoltán Hegedűs for helping me with the various administrative tasks.

I thank my colleagues Bea, Vilmos and Adam for their friendship and their inspirational research ideas.

I thank the ITC students, we share so many memories, I am sure we can make our own luck whatever path we choose from here.

I thank my wife Otilia with whom I had the chance to live four years full of joyful moments in the beautiful city of Szeged.

I thank the country and people of Hungary for accommodating me during my PhD, for the chance I had to experience their vibrant culture.

I thank my family and friends in Romania for being supportive.