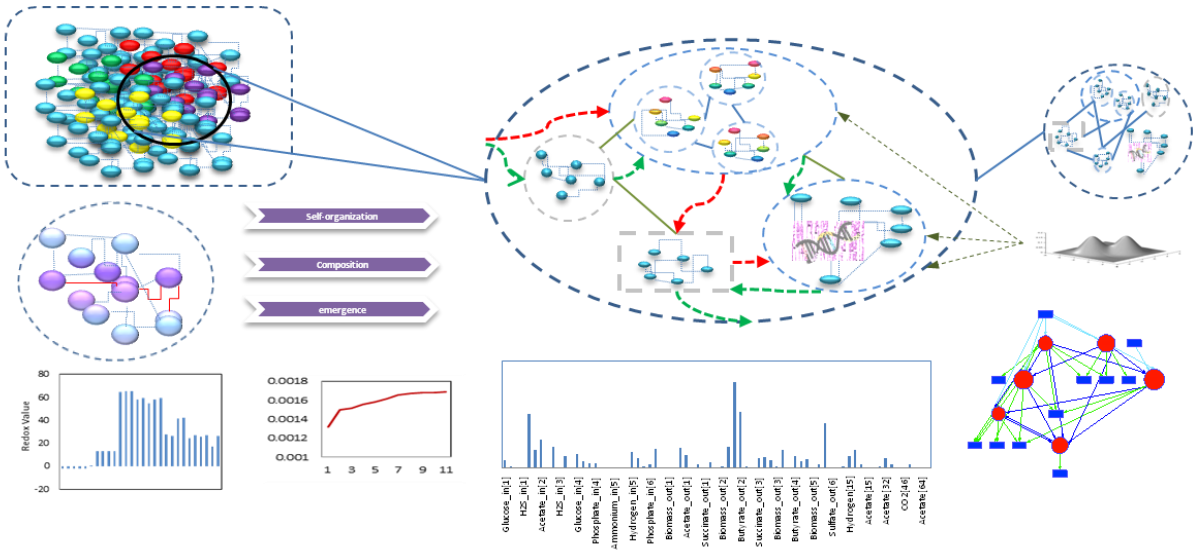
CHALMERS




# Community Systems Modeling

## From Complexity to Optimality

*Master of Science Thesis*

**Pouyan Ghaffari Nouran**

*Systems and Synthetic Biology Division*
*Department of Chemical and Biological Engineering*
*Chalmers University of Technology*
*Gothenburg, Sweden 2013*

***THESIS SUBMITTED***

***FOR THE PARTIAL FULFILMENT***

***OF***

***MASTER OF SCIENCE in COMPLEX ADAPTIVE SYSTEMS***

# COMMUNITY SYSTEMS MODELING
# FROM COMPLEXITY TO OPTIMALITY

POUYAN GHAFFARI NOURAN

Examiner: Prof. Jens Nielsen
Supervisor: Saeed Shoaie

Systems and Synthetic Biology Division
Department of Chemical and Biological Engineering
Chalmers University of Technology
Gothenburg, Sweden 2013

## *Abstract*

Biological systems, whether they are community of species in an ecosystem, circuits of neurons in the brain or molecular mechanisms inside a cell, are inherently dynamic and complex. This complexity is enclosed not only in number and characteristics of constituent components of the system, but rather in the non-uniform and intricate way of connections between the components to form an emergent phenotype. Human microbiota is not an exception and embodies a typical example of complex biological system. The structure, assembly and dynamics of microbiota along with its contribution to the diet, physiology and development of the host are affected by the diversity of species and genome. In addition, interactions between the various species have an important role in the metabolism of ecosystems. Recently genome-scale metabolic models (GEMs) have been used to model the between three predominant species in human gut microbiota and have been explored the interactions between microbes in simplified community. Despite recent progresses, we still have very limited knowledge about the contribution of individual microorganisms within the communities and the interactions between them. This calls for the development of system-level methods and modeling frameworks to discover details of the complex microbial communities. Here, we introduce a multidimensional distributed model to study and analyze the microbiota as whole, encompassing species with interlaced metabolic interactions in different levels. The model is developed as a comprehensive, integrated and predictive framework and can be used for different complex ecosystems with any number of interacting species. We applied our method to analyze six representative species from abundant phyla in human gut microbiota. The outputs of specialized simulations validated based on SCAFs secretion profile and species abundances for lean and obese subjects.

# *Table of contents*

# *1. Background*

## 1.1 Gastrointestinal (GI) tract

Microbes are the oldest and richest form of life on the earth, and accordingly all multi-cellular organisms are adapted to live in microbial abundant environments (Whitman et al., 1998). First evidences for the existence of microorganism in human body were observed late seventeenth by A. V. Leeuwenhoek in mouth and feces of individuals. But the importance of his work was not recognized for a long time(Kovatcheva-Datchary et al., 2013). Ensembles of microorganisms that colonize and reside in an established area referred to microbiota and the collective genome of them is called microbiome. The number of microorganisms exist in different parts of the human body, including skin surface, respiratory tract, oral cavity and gastrointestinal (GI) tract, has been estimated to be up to $10^{14}$, outnumbering the total number of cells in body with factor of 10(Guarner and Malagelada, 2003). Gastrointestinal tract, with largest collection of microorganisms has become the focus of increasing number of research initiatives during past decade. Recently, it has been shown that gut microbiota has a key role in human health and disease by performing several functions for the body including nutritional, developmental, physiological and immunological functions (Claesson et al., 2012; Karlsson et al., 2013; Nicholson, 2006).

### 1.1.1 Diversity and distribution of gut microbiota

Three life domains on the earth, Eukaryote, Archaea and Bacteria, are present in the human gut with dominancy of Bacteria(Kovatcheva-Datchary et al., 2013). Ten bacteria phyla are identified in gut ecosystem which two of them, Bacteroidetes and Firmicutes, consists up to 90% of the phylotypes(Turroni et al., 2008). In Firmicutes phyla the important bacteria groups are *Faecalibacterium prausnitzii* with 5-15% share of total microbiota and butyrate producing groups *Roseburia species* and *Eubacterium rectale* forming 5-10% of the microbiota (Aminov et al., 2006). In *Bactroides phyla*, *Prevotella* and *Bifidobacterium* are important genus comprising gut microbiota (Turroni et al., 2008). Archaea domain is represented in the human gut microbiota with two prevalent methanogens in healthy individuals, *Methanobrevibacter smithii* and *Methanobrevibacter stadmanae* (Salonen et al., 2010).

Human GI tract is a system with specialized compartments which consists of different anatomical areas, from stomach to rectum. These regions are characterized by couple of physicochemical aspects such as pH, redox, luminal content, compounds derived from diets, and host secretions. Stomach and initial part of small intestine have a sparse microbiota with microbial concentration in the range of $10^2$ to $10^4$ per gram of digest. This low abundant microbiota is a result of acid stress, release of pancreatic enzymes and quick flow of nutrients. The concentration of microorganisms increases from middle of small intestine and reaches the highest concentration of $10^{12}$ bacteria per gram in colon(Kovatcheva-Datchary et al., 2013).

### 1.1.2 Effective factors on composition of gut microbiota

Factors that affect the composition of gut microbiota can be explained in three groups: Host, Diet and Treatment. Several studies on identical and fraternal twin subjects have been revealed higher

similarity in intestinal microbiota composition in monozygotic twins than randomly selected unrelated individuals (Turnbaugh et al., 2009). In elder hosts gut is characterized by lower diversity of microbiota, increased concentration of facultative aerobics and decreased anaerobic bacteria(Claesson et al., 2009). In mammals there are complex interactions among the host, microbiota and diet as a result of convolution between microbiota and higher vertebrates during millions of years. Consequently any significant change in diet and lifestyle is likely to affect the eco-physiology of gut and destabilize these established interactions. A good example is the effect of breast milk vs. formula milk in the early stage of life. For the breast feeding Infants gut microbiota are dominated by bifidobacterium, while feeding with formula milk mostly results in more microbial diversity and presence of facultative anaerobic such as streptococci and staphylococci (Palmer et al., 2007). Antibiotics are other group of perturbing factors that cause a long-term reduction in gut microbiota diversity (Dethlefsen and Relman, 2011).

### 1.1.3 Functionality of gut microbiota

In recent studies based on metagenomics it has been indicated that the coding capacity of gut metagenome exceeds the human genome by the factor of 150 folds (Turnbaugh et al., 2009). Gut microbiota have been attributed with couple of important functions such as maturation of immune system, pathogens defense, intestinal microvilli development and nutrition(Kovatcheva-Datchary et al., 2013). The nutritional functions include the fermentation of indigestible components of the dietary fibers and anaerobic process of proteins and peptides, which result in providing more metabolic energy source for the body and maintain the gut ecosystem health (Acheson and Luccioli, 2004).

### 1.1.4 Gut microbiota in host disease

Gut microbiota is mostly occupied by bacteria that establish mutualism or commensalism relationship with the host, but there are group of potentially harmful microorganisms that are classified as pathobionts (Sansonetti 2011; Round and Mazmanian 2009). In healthy individuals the stable condition at the phylum level is maintained through the dominancy of Bacteroidetes and Firmicutes within the gut ecosystem which indicates the high adaptation capability of system and its co-evolution with host. Although the gut microbiota endowed with redundant functionality and tolerance capability to environmental stresses, strong and long lasting perturbations can induce changes into the composition and functionality of the system (Dethlefsen and Relman, 2011). Gut microbiota have a broad contribution to human disease including inflammatory bowel disease (IBD)(Kovatcheva-Datchary et al., 2013; Png et al., 2010), autoimmunity and type 1 diabetes (T1D)(Kovatcheva-Datchary et al., 2013), Celiac disease(Sanz et al., 2011), allergy(Isolauri et al., 2009), autism spectrum disorders(Kovatcheva-Datchary et al., 2013), type 2 diabetes and obesity (Sekirov et al., 2010).

Figure (1.1) illustrates the variations in gut microbiome diversity and functionality along the gastrointestinal track, factors which shape the microbiota composition, and diversity of microbiota through different stages of life.

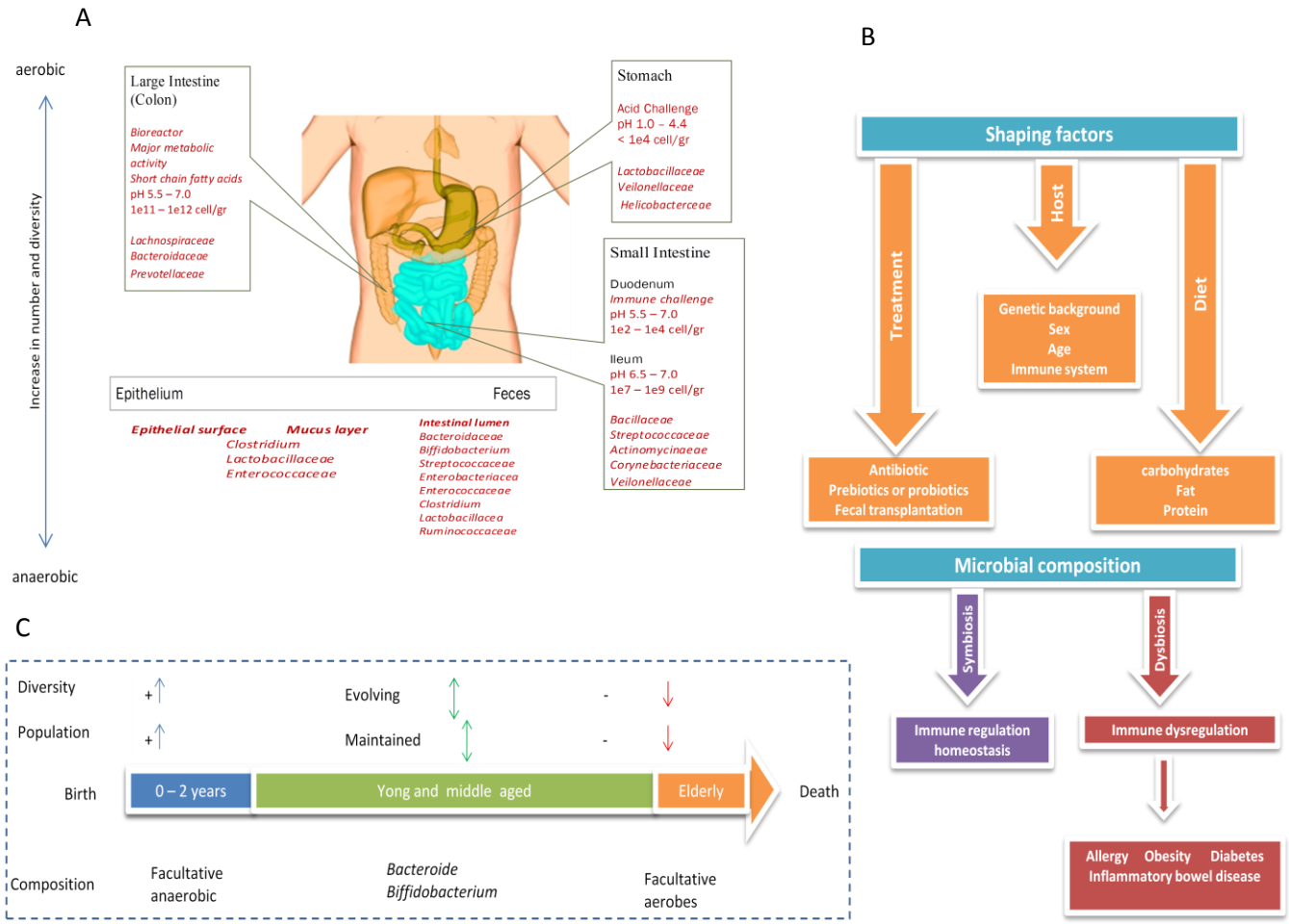



**Figure 1.1 A: Gut microbiota variations along the GI tract. B: microbiota composition shaping factors and diseases. C: temporal variations of gut microbiota**

## 1.2 Systems biology

Systems biology is the quantitative and comprehensive analysis of interactions inside the biological systems to understand the system through a holistic approach. Systems biology body of the knowledge develops and applies concepts from systems theory to capture the complexity of biological system over the integration of experimental data with mathematical modeling and computational simulations. It has been emerged as the limitations of reductionist (conventional) biology appeared in understanding the complexity arising from dynamic interactions between biological components (Dubitzky et al., 2011). Figure 1.2 illustrates the knowledge generation process in systems biology through the phases: data acquisition, computational and mathematical modeling, systematic analysis, technology development, laboratory experiments and knowledge acquisition (Kitano, 2002a, b).

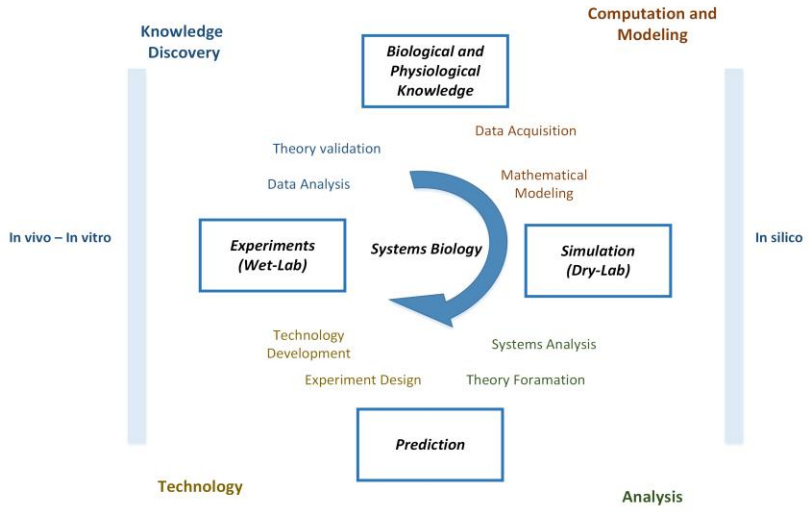



Figure 1.2 Systems Biology Cycle (Kitano, 2002a, b)

## 1.3 Genome-scale modeling

Our information regarding the complex biological phenomena is typically incomplete and building accurate mathematical models is out of access. Consequently it is necessary to find alternative approaches which can simulate these systems' behavior with less detailed knowledge about their inside interactions and in an acceptable range of accuracy. One of the candidate and commonly used methods by system biologists and mathematicians is constraint-based modeling which describes the complex biological system through a set of characterizing constraints.

According to the genome-scale metabolic networks, the constraint-based models are used to find the relationship between genotype and phenotype incorporating the physiochemical and genetic constraints into the model (Feist et al., 2009). Generally speaking this approach is based on three fundamental assumptions and concepts: limiting the phenotype by physiochemical constraints, mathematical definition of the evolutionary pressures and genome-scale perspective of the metabolism. In complex metabolic network three basic constraints are limiting the occurrence of reactions: availability of substrate and necessary enzymes, charge and mass conservation, and thermodynamics(Lewis et al., 2012).

Building genome-scale constraint-base reconstruction networks usually consists of four major stages: gene annotation, manual modification and building mathematical model, model validation using available experimental data, and model improvement through iterations between computational and experimental parts(Oberhardt et al., 2008).

Flux balance analysis (FBA) under steady state condition is the basic and most commonly used method for the genome-scale constraint-based modeling. Mathematically, it means to compute the

basis of corresponding polyhedral cone. In FBA after defining the cellular objective and supplying the metabolites into the metabolic network, linear programming is used to optimize the objective function subject to the imposed metabolic network constraints and uptake rates (Kauffman et al., 2003; Orth et al., 2010). Figure 1.3 illustrates the methodology of FBA applied for a simple metabolic network with three metabolites and seven reactions. The reaction network projected into the stoichiometric matrix and the physicochemical limitations is translated to the constraint equations which produces the feasible cone. Then optimizing the system through the implementation of linear programming for two different objective functions, optimum points are represented on feasible cone.

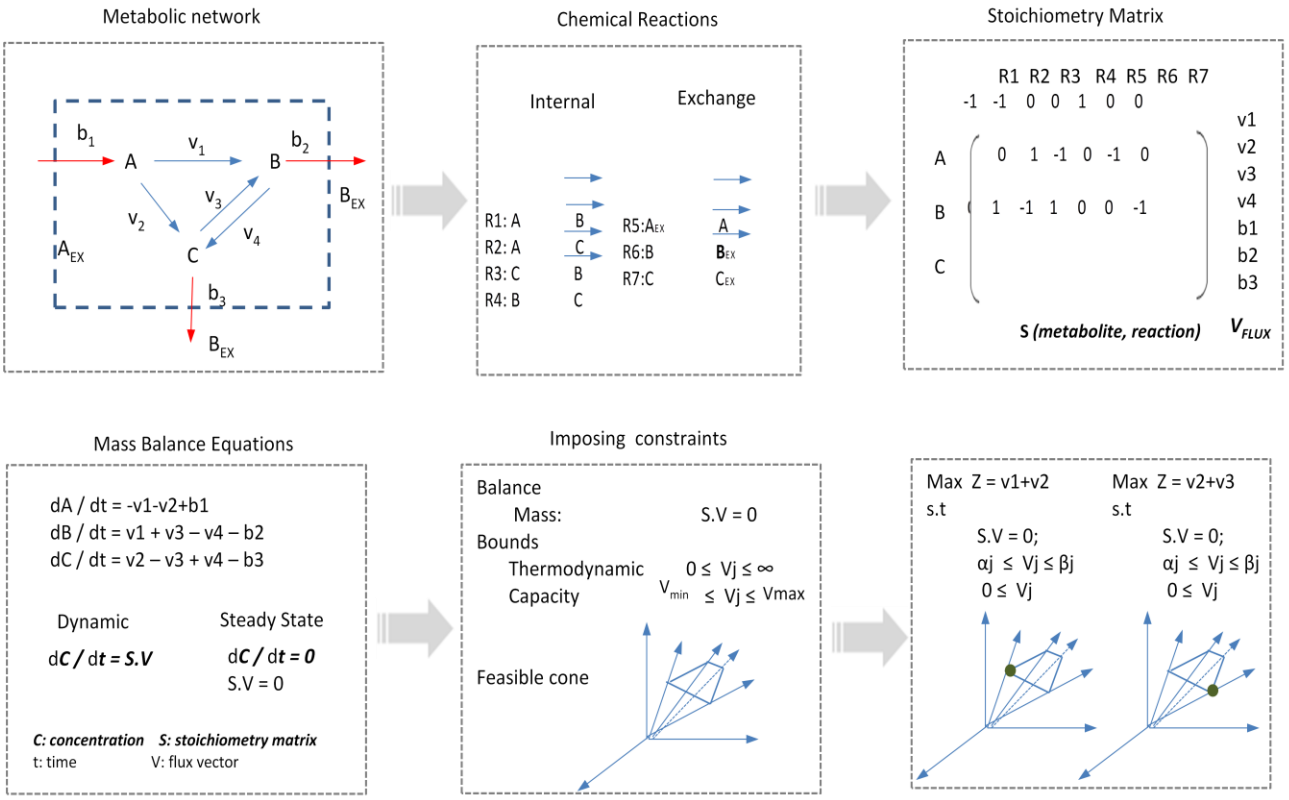



**Figure 1.3 simplified metabolic network and Flux Balance Analysis**

### 1.4 Complex Systems

Most of interesting systems including ecosystems, human brain, immune systems and socio-economic systems are difficult to be described and to be controlled by traditional methods. This difficulty mainly comes from two sources: complexity and adaptability. Complexity refers to nonlinearity of interactions between system components which can lead to an unexpected emergent behavior, a well studied phenomenon in chemical, physical, biological and social systems. Adaptability can be defined as changing the specifications and behaviors of primitive components and evolving of individual agents over time in response to system level perturbations and changes in environmental conditions. According to above mentioned properties, modeling and simulation of

natural eco-systems are challenged by two significant factors, complexity of resembling mathematical models and difficulty of solving these models, which are required to be addressed by any candidate algorithm in a proper way. To have a less ad hoc and more universal solver the following concepts must be considered in developing the model (Grimm et al., 2005):

- Emergence: How to model the individual traits to guaranty the realistic emergence of systems responses?
- Adaptation: Which mechanisms are used by individual agents to adapt in repose to shaping forces imposed by environmental factors?
- Strategy and fitness: What are the appropriate fitness measures to model decision making and navigation procedure both in system level and individual level? Are these measures needed to be modified according to the life state of systems?
- State-based reflection: What are the appropriate assumption to describe the dependency between decision processes and agent-level states?

Figure 1.4 compares the simple and complex systems according to their structure and their response mechanism to environmental perturbations.

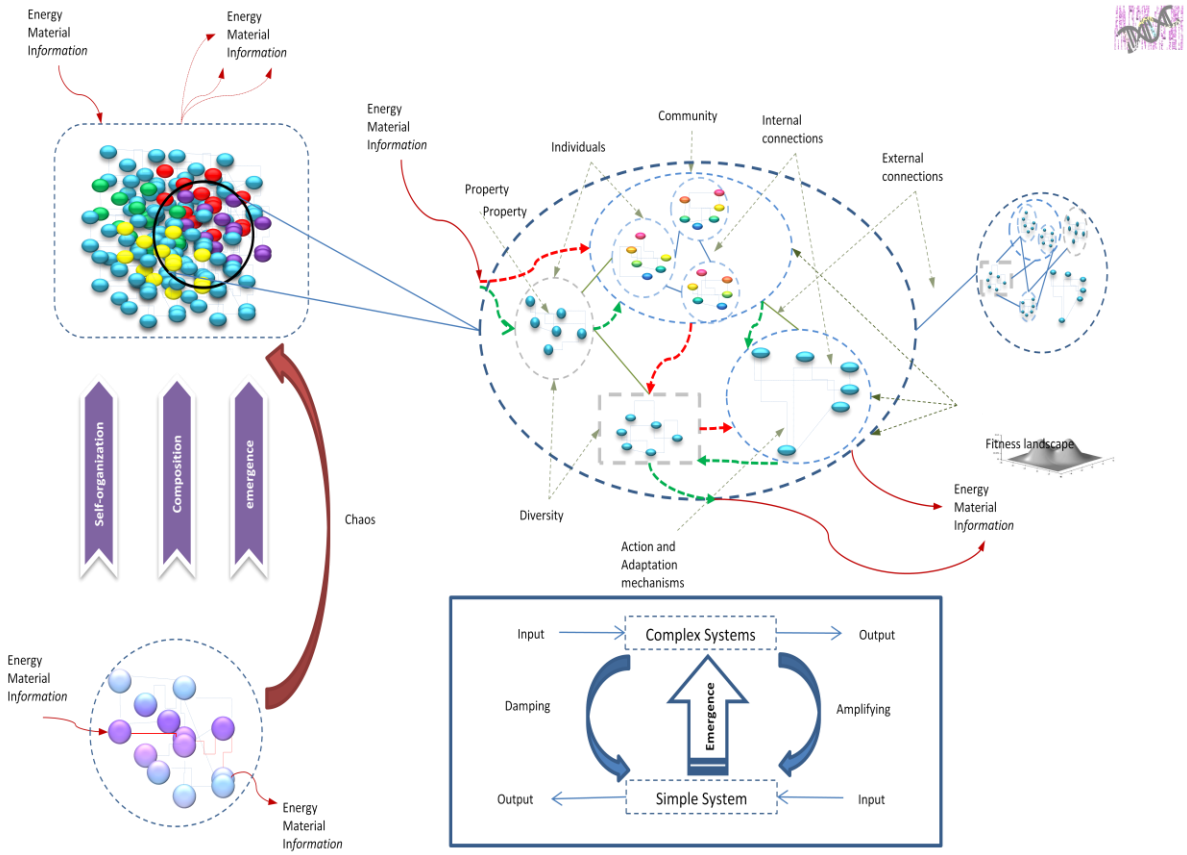



**Figure 1.4, Transition from simple to complex adaptive system and the differences in structure, behavior, response and adaptation processes.**

### 1.5 Optimization

Generally speaking mathematical optimization or alternatively mathematical programming is the procedure of finding best set of parameters from some set of available alternatives regarding some selecting criteria. The mathematical optimization problem can be represented as:

$$Min/Max \quad F(X)$$
$$subject\ to$$
$$g_i(X) \leq 0$$
$$h_i(X) = 0$$

Where

$X = (x_1, \dots, x_n)$ : Optimization variables
$F: R^n \rightarrow R$ : Objective function
$g_i: R^n \rightarrow R, \quad i = 1, \dots, m$ : Inequality constraint functions
$h_i: R^n \rightarrow R, \quad i = 1, \dots, k$ : Equality constraint functions

Feasible points are those $X \in R^n$ which satisfy all constraint functions. Optimal solution $X^*$ is a subset of feasible solution with smallest/largest value of objective function, F(X). Models with m=k=0 are classified as unconstraint problems. Optimization problems are usually difficult to solve and involve some compromises such as long computational time and non-convergent solutions. But there are certain classes of problems: convex optimization, least-squares problems and linear programming problems, which present the valuable properties of efficiency and reliability. Least squares method can be defined as minimizing the sum of squares of residuals or errors made in every individual equation. Its main applications are in data fitting and pattern recognition. Linear programming is most common case of continuous optimization where the objective function is linear and the constraints are affine. Formally linear programming can be defined as:

$$Min\ C^T X$$
$$subject\ to$$
$$a_i^T X + b_i \leq 0 \qquad i = 1, \dots, m$$
$$-x_i \leq 0 \qquad i = 1, \dots, n$$

A problem is defined as convex optimization problem where both objective and constraint functions are convex. Convexity of feasible area guarantees that the local optimum is also the global optimum of the problem.

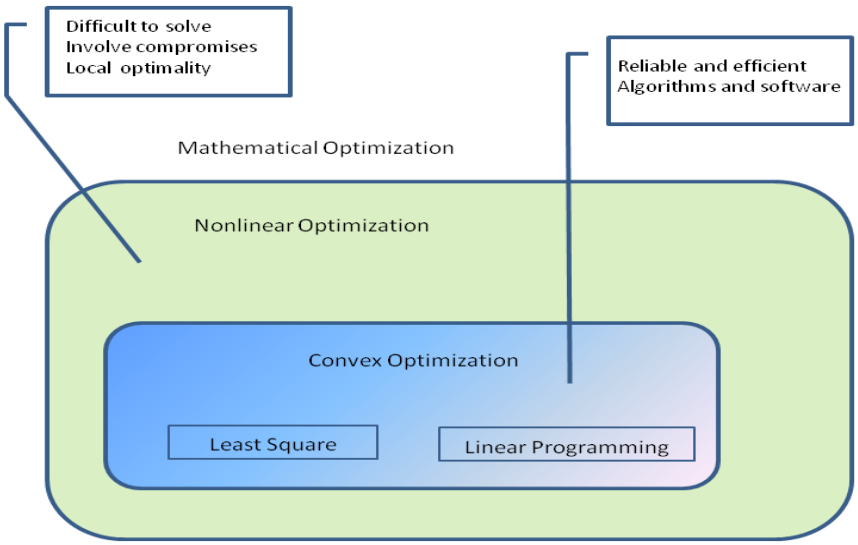



**Figure 1.5 Domain of mathematical programming**

# 2. *Methods*

Simulating the microbial ecosystems is a multifaceted task which needs integration of complexity characteristics of the system with individual-level optimality and community-level tendency to operate in optimal conditions. A multidimensional distributed model was proposed in this project as a mathematical framework and its related toolbox, Community System Modeling and Optimization Structure (COSMOS), to simulate the complex community of microorganisms. The model is general enough to include any complex ecosystem with any number of interacting species. The power of the model comes from its ability to capture the multidimensional optimality through a dynamic feasibility condition and incorporating the complexity of the system as a quantitative factor to guide the solution process. Also the distributed nature of the model makes it possible to simultaneous simulation of a portfolio of ecosystems with enormous number of interactions.

## 2.1 Mathematical Structure

### 2.1.1 Optimization framework

A multi-dimensional collaborative and distributed method has been proposed as the structure of optimizing part of the methodology. The fundamental attitude behind the model developing task was to project the complexity of the real system into the model and at the same time maintaining the mathematical complexity of the model at minimum possible level. So linearity and convexity of the optimization problem was preferred to nonlinearity of objective and constraint functions to avoid the risk of encountering the non-convexity. The eco-system was defined as a System of Systems (SOS) with assemblage of operationally independent sub-systems in a compartmentalized environment. Decomposing the SOS into the compartments enables the model to focus on optimization at lower levels with coordination at higher level which relaxes the system-level computational task by distributing the load all over the sub-systems. After decomposition, implementing collaborative optimization method, the original problem was divided into the optimization at different levels of system, from community-level to species-level, coordinating and solving the inconsistencies between compartments and individuals using a global community-level optimizer. Through this method, each processor in sub-community level was given the control just over its own set of variables, and domain specific constraints are needed to be satisfied just by corresponding sub-system. The proposed optimization method is illustrated in figure (2.2).

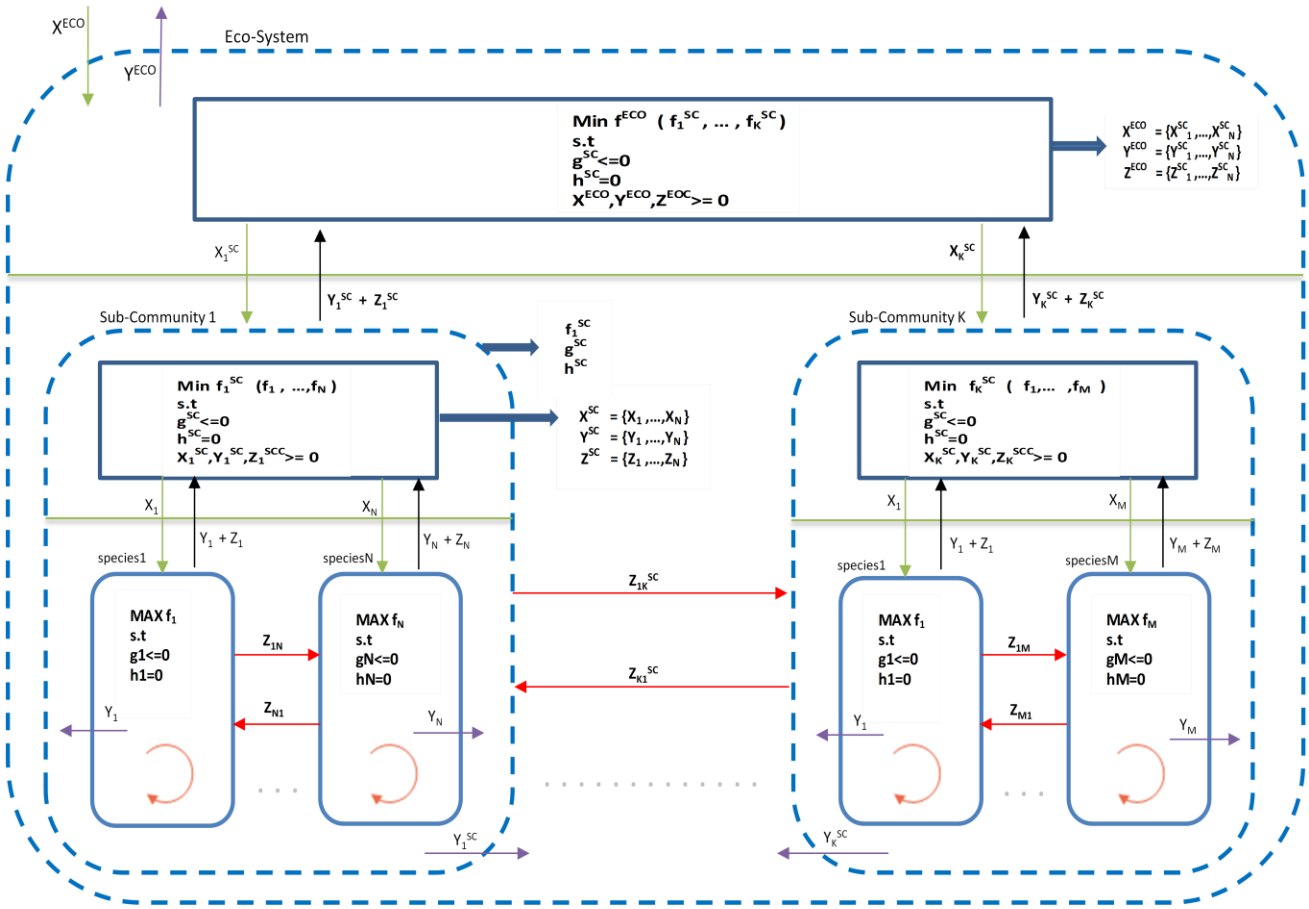



**Figure 2.1, Mathematical programming structure of COSMOS. The framework consists of three compartments: eco-system, community and species. From the system prospective each sub-system is considered as a single processor with inputs and outputs. Design variables are divided in three groups of vectors, (X, Y, Z), representing system-level inputs, system-level outs and connecting parameters. Each sub-system is fed by part of system-level input vector X and connecting vector Z which assigned by system analyzer and produces responses in the form of f, g, and h functions. Cumulated responses in each compartment formulate the domain-specific optimization problem, which will be solved to organize connections inside the domain and also produce proper responses. Optimization in eco-system level simultaneously produces the vector of inputs, outputs and connecting parameters for all species and sub-communities without sharing domain-specific information, which enables the model to be distributed and use heterogeneous computational infrastructures (software and hardware) for each compartment.**

Sub-systems perform on their optimum level, so the model is always locally-optimum and searches through the domain to find the system-level optimality. Constraint equations at system-level are linear with semi-dynamic right-hand-side, which shapes the feasible space as convex polyhedron. Convexity of feasible space is an important property for optimization part of the framework and guaranties to reach a proper optimum point. Partially dynamic feasible space is proposed considering the evolutionary nature of eco-systems. Part of the feasible domain which shaped by system-level inputs was set to be fixed considering the fact that community-level resources are limited and constant. But species and sub-communities are allowed to shape the remaining part of the feasible space in a dynamic way.

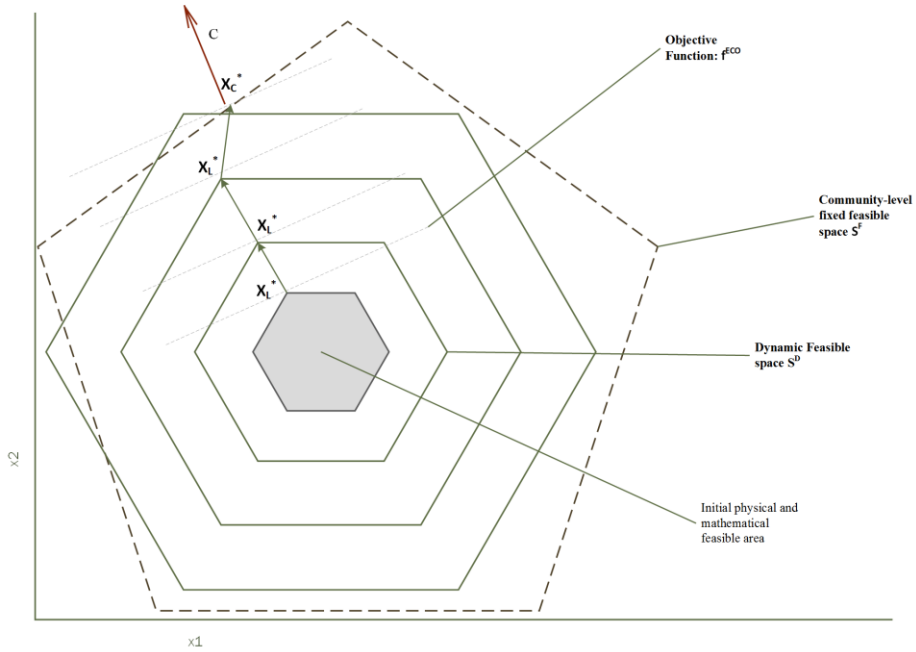



Figure 2.2, Dynamic feasible space

### 2.1.2 Power of Network

Describing and modeling eco-systems without considering their complex network structure and reflecting it into the model could affect the accuracy of the simulation in large scale and produce biased predictions. To avoid this pitfall we developed methods to induce the network effect into the model in three different areas: initialization, optimization and controlling.

Initialization process was affected by network structure through the activation process. According to their position inside the networks nodes were divided into the three groups: sensors, non-primary sensors and non-sensors.

- Primary sensor: node in boundary of system which uses community-level inputs and can grow independently.
- Non-primary sensor: node in boundary of the system which uses community-level inputs but needs at least one connecting input to grow.
- Non-sensor: node which is not in the boundary of the system and depends completely on connecting inputs to grow

In first step primary sensor nodes were activated and the activation diffused into the model based on one of influence or support approaches, which we called the activation cascade and support threshold respectively. In activation cascade each active node tries to activate all the nodes in its first degree connection circle, and the procedure repeats until entire network gets on. In support threshold method nodes were assigned a threshold value related to each input and were activated if the influx passes this threshold. Figure 2.4 represent the activation cascade method implemented for a network of 6 nodes.

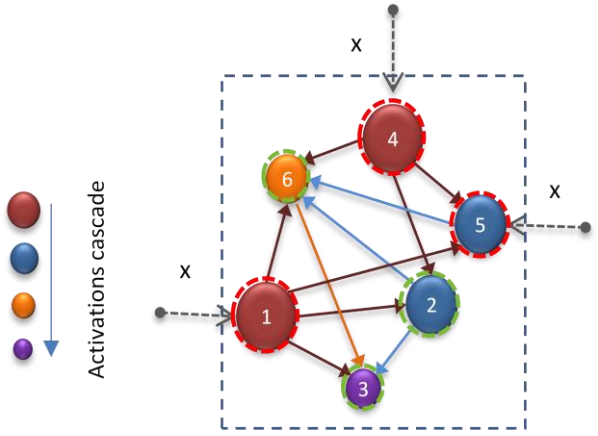



**Figure 2.3 activation cascade procedure. Activation of the network starts by nodes 1 and finishes in node 3 after 4 attempts. Similarity of colors shows the similarity of activation order.**

The optimization framework was integrated with centrality power of nodes (species or sub-systems) and an extra force added as to modify the optimization orientation. Community-level objective function formulated considering two types of steering forces: local force and global force. Local force was described through individuals' internal performance capacity regarding to biomass yield or energy dissipation, but the global force was defined as the impact of network topology and nodes' power on individual's performance and consequently distribution of interactions within the community. Normalized relative centrality degree was used as the measure of complex network effect on optimization, and from available formulations three well-studied ones were selected: Degree centrality, Betweenness centrality(Brandes, 2008) and Eigenvector centrality (Bonacich, 2007).

- Degree Centrality: number of adjacent vertices determines the importance of each vertex.

  $C_D(v_i) = \deg(v_i) = \sum_j A_{ij}$

  Relative Degree centrality: $CR_D(v_i) = \frac{C_D(v)}{n-1}$ n: total number of vertices

- Betweenness Centrality: vertices are categorized with number of shortest paths passed from. Highest Betweenness is important in diffusion and communication.

$$C_B(v_i) = \sum_{v_s \neq v_i \neq v_t \in V, s<t} \frac{\sigma_{st}(v_i)}{\sigma_{st}}$$

$\sigma_{st}(v_i)$: Number of shortest paths between s and t through vertex i
$\sigma_{st}$: Number of shortest paths between s and t

Relative Betweenness centrality:

1. Undirected: $CR_B(v_i) = \frac{C_B(v_i)}{(n-1)(n-2)/2}$
2. Directed : $CR_B(v_i) = \frac{C_B(v_i)}{(n-1)(n-2)}$

- Eigenvector Centrality: vertex importance is determined according to the importance of vertices in its connection circle.

$$C_E(v_i) = \frac{1}{\delta} \sum_{j \in M(i)} C_E(v_j) \qquad \delta: constant \quad M(i): connection\ circle$$

Eigenvector and Degree centralities are used to orient the objective function and the Betweenness centrality in integration with Eigenvector centrality was used to obtain maximum control on the network with minimum number of controlled nodes. As we use the directed network structure to calculate the centrality degrees, Eigenvector centrality mainly represents the node's influence inside the network and the Betweenness centrality can be seen as the effect of support. So the network can be controlled in large extent through the controlling nodes with highest Eigenvector and Betweenness degrees. In Figure 2.5 centrality degrees were calculated for a sample network of a six nodes, the blue network was shaped according to Eigenvector centrality and the green one based on linear combination of Betweenness and Eigenvector.

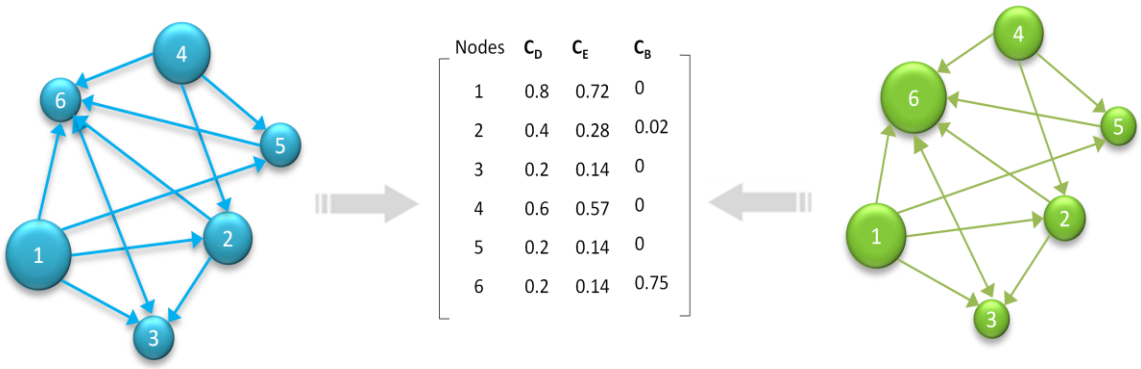



**Figure 2.4 centrality degrees. Size of the nodes represents the value of corresponding Eigenvector centrality degree for network colored blue and the value control factor for network colored green. Control factor defined as linear combination of Eigenvector and Betweenness centralities.**

### 2.1.3 Objective function

Objective functions are formulated based on two approaches: biomass production capacity and energy dissipation. Forces which orient the objective functions are calculated as integrated local and global weights, applying biomass yield of species and their centrality power in complex network of eco-system.

$$Z = \alpha.X + \beta.Z^{in} + \gamma.Y + \delta.Z^{out}$$
$$\varphi_{bms} = [\alpha, \beta, \gamma, \delta]_{1\times(l+k)} \ : Biomass\ coefficients\ matrix$$
$$\varphi_E = [\alpha, \beta, \gamma, \delta]_{1\times(l+k)} \ : Energy\ coefficients\ matrix$$
$$\theta = [X, Y, Z]_{1\times(l+k)} \ : community\ and\ local\ parameters$$
$$\boldsymbol{S} = \boldsymbol{\varphi} \times \boldsymbol{\theta'}$$

φ is a binary matrix with value 1 when a parameter is active in objective equation and 0 otherwise. Accroding to the selected optimization approach, biomass or chemical potential ratios are implemented as weights, , and the resulted matrix is corrected with rational centrality degrees.

$$\boldsymbol{\varphi'} = \boldsymbol{\varphi}.\ \boldsymbol{\omega}\ .\boldsymbol{C_E}$$

$$\omega_{\mathrm{bms}} = \frac{\mathrm{BMS(i)}}{\min(\mathrm{BMS_{PSL}})}$$

PSL: primary sensor list

BMS(i): biomass yield of microorganism i, $\mathrm{i} = 1 \ldots \mathrm{n}$, n: number of PLS members

$\mathrm{BMS_{PSL}}$ : list of biomass yield of primary sensor microorganisms

$$\omega_{\mathrm{E}} = \frac{\mathrm{ChemPot(i)}}{\min(\mathrm{abs}(\mathrm{ChemPot_{PSL}}))}$$

ChemPot(i): chemical potential of microorganism i, $\mathrm{i} = 1 \ldots \mathrm{n}$, n: number of PLS members

$\mathrm{ChemPot_{PSL}}$: list of primary sensors chemical potentail

Figure 2.3 depicts how the objective function was built and implemented.

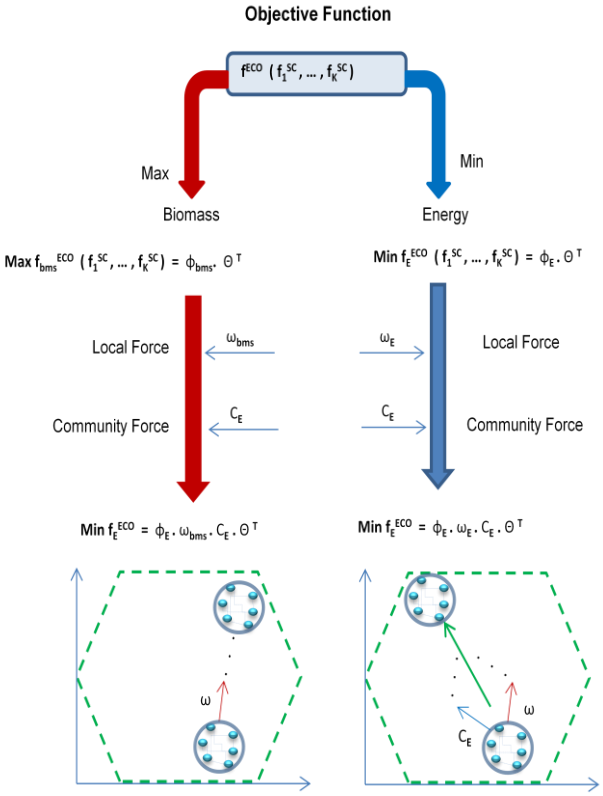



**Figure 2.5 Formulation of objective function integrating local force and community force and its effect on optimization directionality.**

### 2.1.4 Binding the network

To maintain the shape and structure of the network a group of binding equations are introduced considering the fact the input fluxes and output fluxes for each individual must be balanced regarding to the carbon, during all phases of simulation. Applying the binding equations we will keep the network firmly connected without losing the elasticity. Equations defined as:

$$C_i.X_i + C_i.Z_i{}^{in} = C_i.Y_i + C_i.Z_i{}^{out} \qquad \text{i=1…N, N: Number of living units}$$

$C: carbon\ coefficients$
$X: community\ level\ input\ flux \quad X = \{x_1 \dots x_n\}$
$Y: community\ level\ \ output\ flux \quad Y = \{y_1 \dots y_m\}$
$Z^{in}: intra - community\ level\ \ input\ flux \quad Z^{in} = \{z^{in}{}_1 \dots z^{in}{}_l\}$
$Z^{out}: intra - community\ level\ \ output\ flux\ Z^{in} = \{z^{out}{}_1 \dots z^{out}{}_k\}$

## 2.2 Algorithm

The algorithm is consists of two main phases: initialization and simulation.

**Initialization**: is a single step process which prepares the basis for the main task of simulation and optimization. It is in this stage where the microorganisms are connected with each other, the network is shaped and a feasible initial location is formed to be used as the base for optimization.

Algorithm:

1. Identify the primary sensor nodes and build the primary sensor list, PSL.
2. Add the PLS to initially empty network activation list, NAL.
3. Share community level sources (X) equally between members of NAL and let them to grow.
4. Calculate the local orientation force of objective function, ω
5. Use activation cascade method to find the nodes which can be activated by NAL.
6. Add the activated nodes to NAL.
7. Let NAL to grow.
8. Repeat steps 6 to 8 until entire network was activated.
9. Construct community matrix.
10. Calculate the community force of objective fun ction, $\boldsymbol{C_E}$
11. Calculate the network controlling factors.
12. Construct the fixed community level feasible space.

**Simulation:** is a multi-step procedure which continues until the solutions are converged. In this stage, using the data and network structure produced in Initialization phase, the community is optimized and the outputs are visualized. The optimization process is handled through the iterative tradeoff between the local optimality and community optimality and settles to final optimum where both microorganisms and community are in optimum state. In global optimization phase the constructed mathematical model produces predictions of system parameters in community optimality condition and transfers them to individual reconstructed models. Through local optimization phase each individual performs under its optimality condition using inputs from global optimizer. Consequently after convergence there is a vector of solution which obtained by simultaneous local and community level optimality states. Figure 2.4 summarize the solution procedure of algorithm.

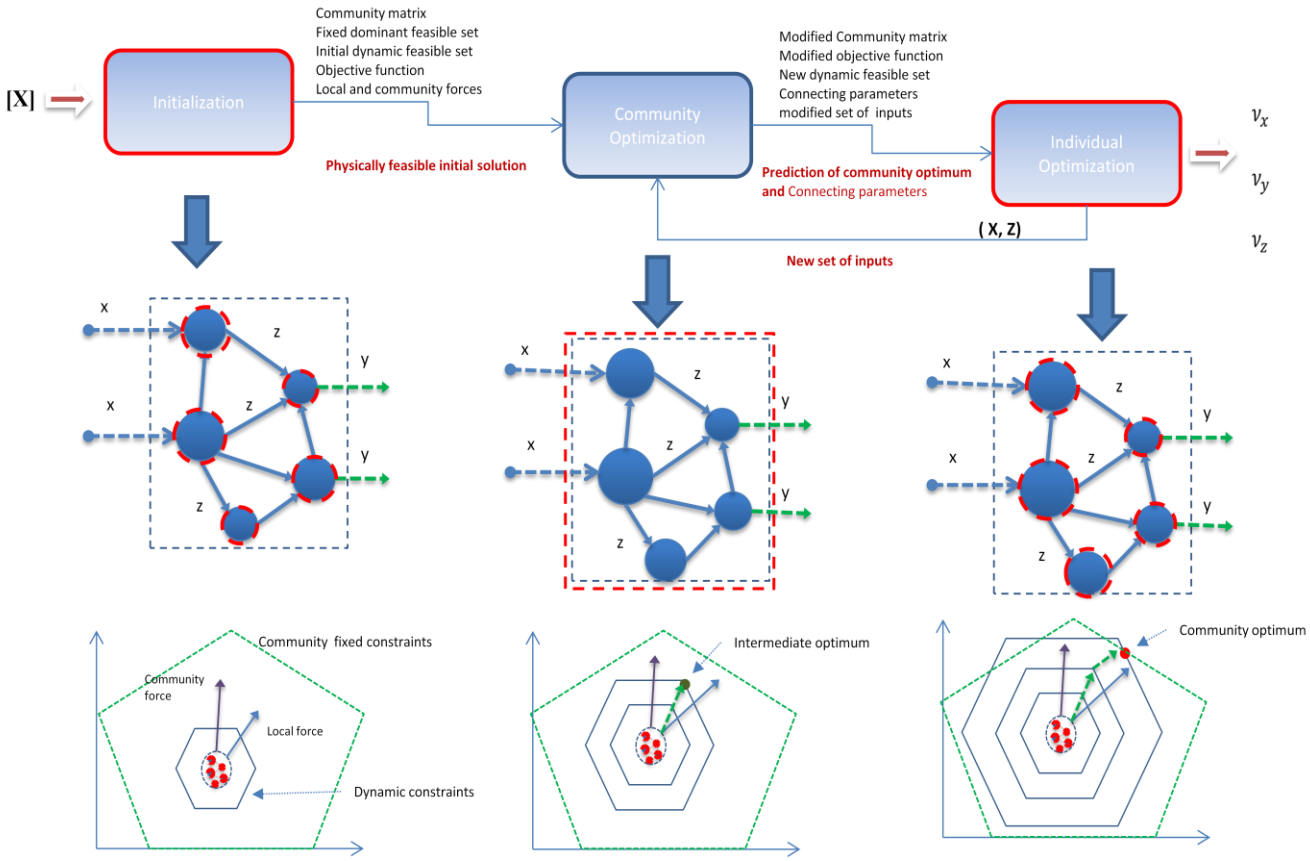



**Figure 2.6 COSMOS algorithm, Schematic presentation of solution procedure, corresponding network structure and migration of optimum solutions from initial state to community optimality.**

# 3. *Results*

Gut microbiota has been used as the case study to explore the capabilities of COSMOS in modeling and analyzing complex communities. Group of six abundant microorganisms representing important phyla were detected based on pubic data and considering their unique function inside the community. Genome-scale models for these six microorganisms (unpublished data) were used as individuals to simulate to community. Available published experimental data, containing microbiota composition and SCFAs profile for 98 lean, overweight and obese subjects (30 lean, 35 overweight and 33 obese)(Schwiertz et al., 2010) compared to the predictions of the personalized simulations conducted by toolbox. Table 3.1 presents an overview of experimental data, related to microbiota composition and secreted SCFAs in average for three classes of subjects according to their BMI value.

**Table 3.1 Fecal microbiota and SCFAs profile (Schwiertz et al., 2010)**

| | **BMI** | | | |
|---|---|---|---|---|
| **Population** | **<25** | **25-30** | **>30** | |
| **Firmicutes** | | | | |
| Clostridium leptum group | 10.4 | 10.3 | 10.2 | *F. prausntzi* |
| Ruminococcus flavefaclens group | 10.3 | 9.9 | 9.9 | *R. bromi* |
| Clistridium coccoides group | 10.3 | 10.3 | 10 | *E. rectale* |
| **Bacteroidetes** | | | | |
| Bacteroides | 10.2 | 10.4 | 10.2 | *B. thta.* |
| **Actinobacteria** | | | | |
| Bifidobacterium | 8.7 | 8.5 | 8.3 | *B. adolesc.* |
| **Archea** | | | | |
| Methanobrevibacter | 8 | 7.3 | 6.2* | *M. smthii* |
| **Acetate** | 50.5 | 56 | 59.8 | |
| **Propionte** | 13.6 | 18.3 | 19.3 | |
| **Butyrate** | 14.1 | 18.5 | 18.1+ | |

* log cells/gr feces + mmol/l

Considering the fact that there was no data related to the composition and amount of food consumed by each subject, to personalize the simulations we fixed the biomass for each microorganism and minimized the model regarding the glucose uptake. In this way we got the

substrate uptake for each subject normalized to glucose. Then applying control mechanism explained in method section, we fixed *R. Bromi and M. Smthii* as control nodes of the community. Finally we optimized the community model according to cumulative biomass production and obtained predictions of substrate uptake, compounds secretion and intra-community metabolite interactions for each microorganism. Figure 3.1 represents the personalization method through minimization of substrate uptake and controlling of the community.

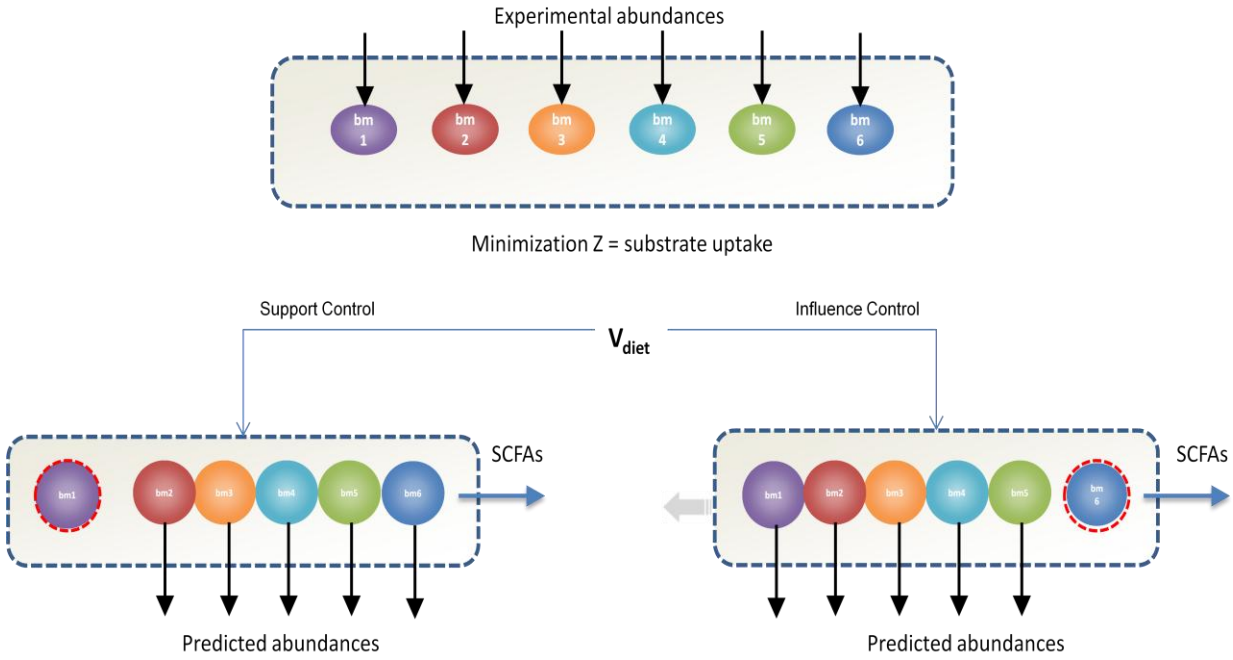
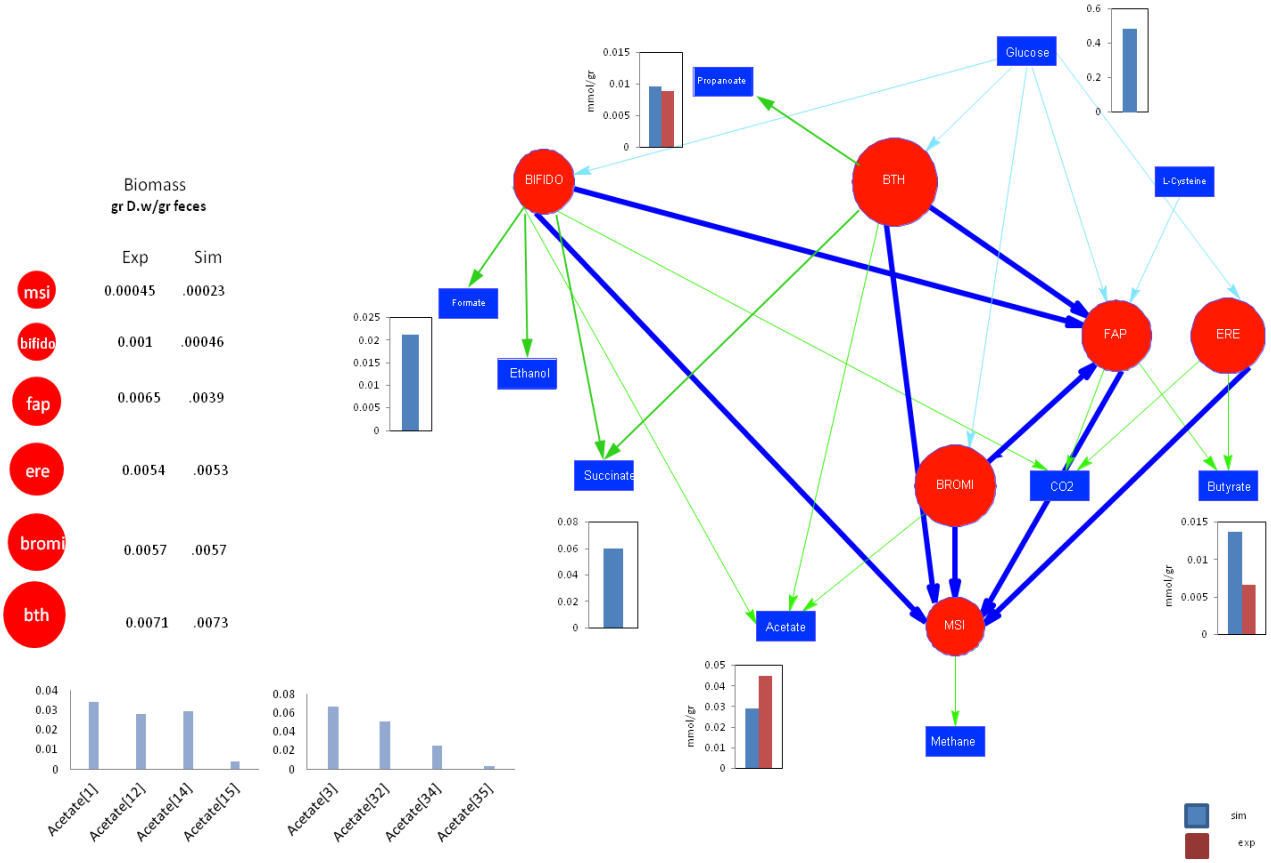
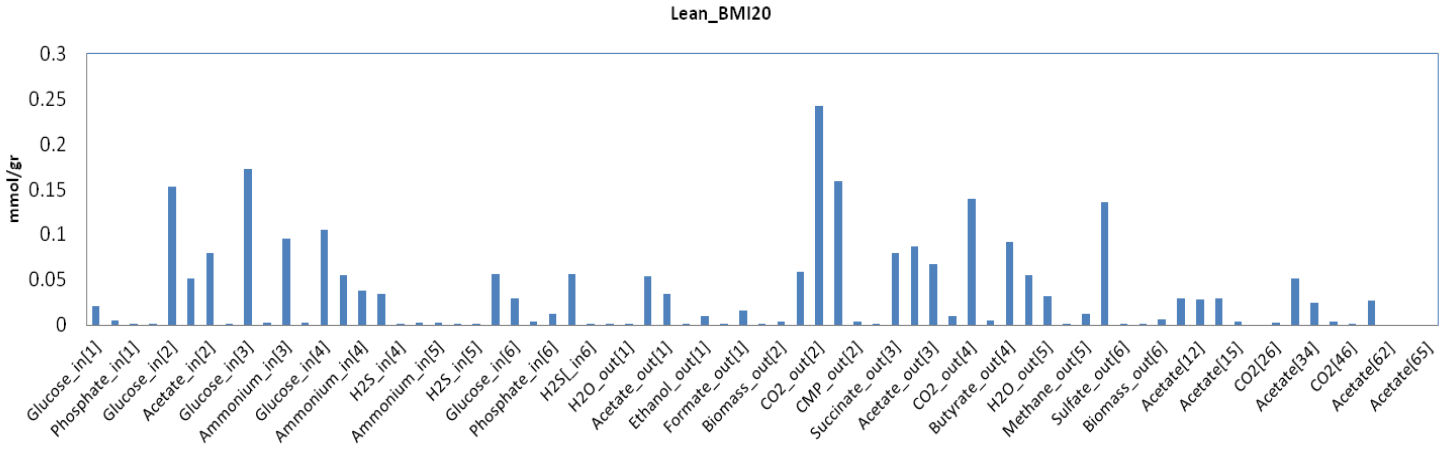


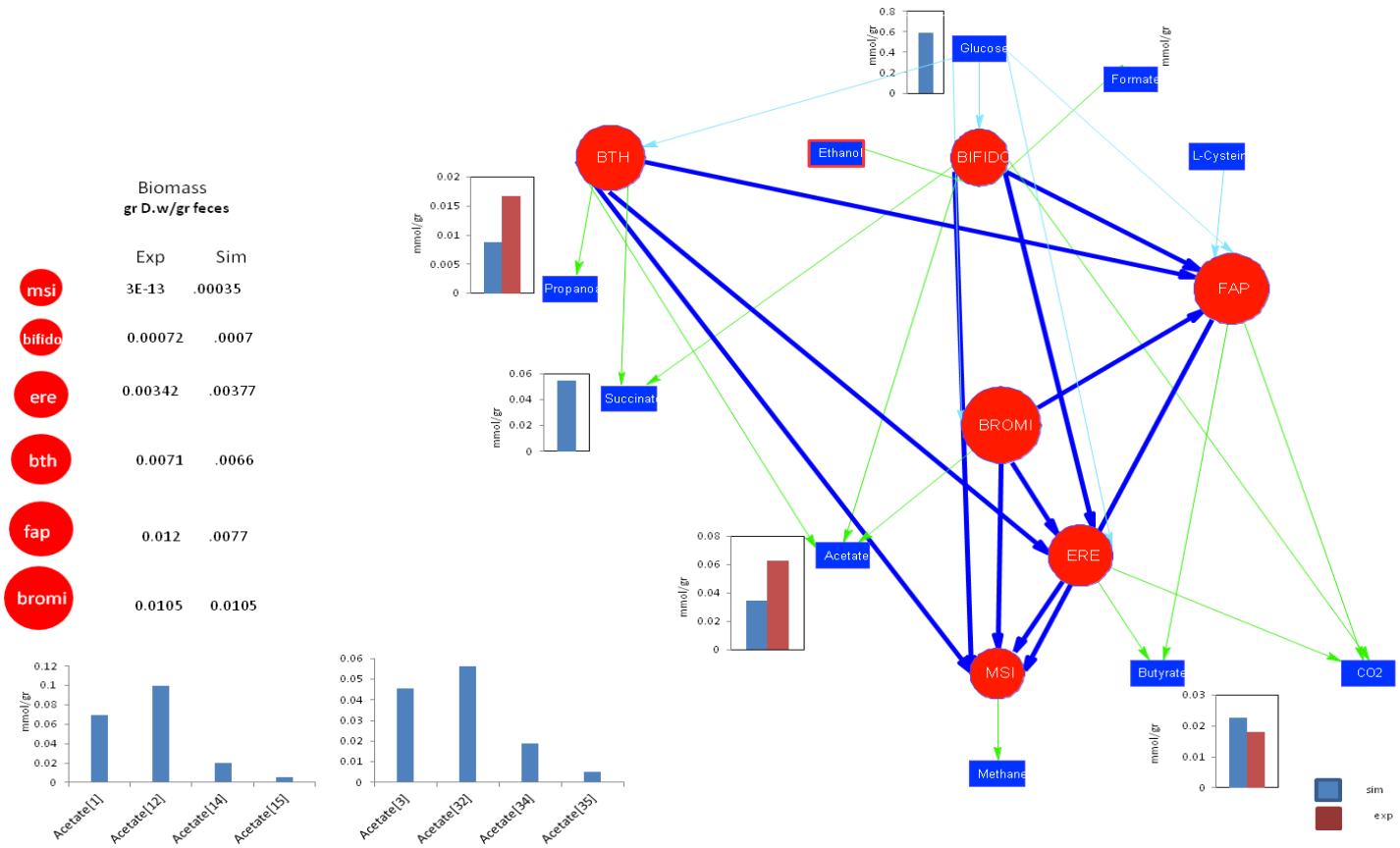

**Figure 3.1 personalizing method**

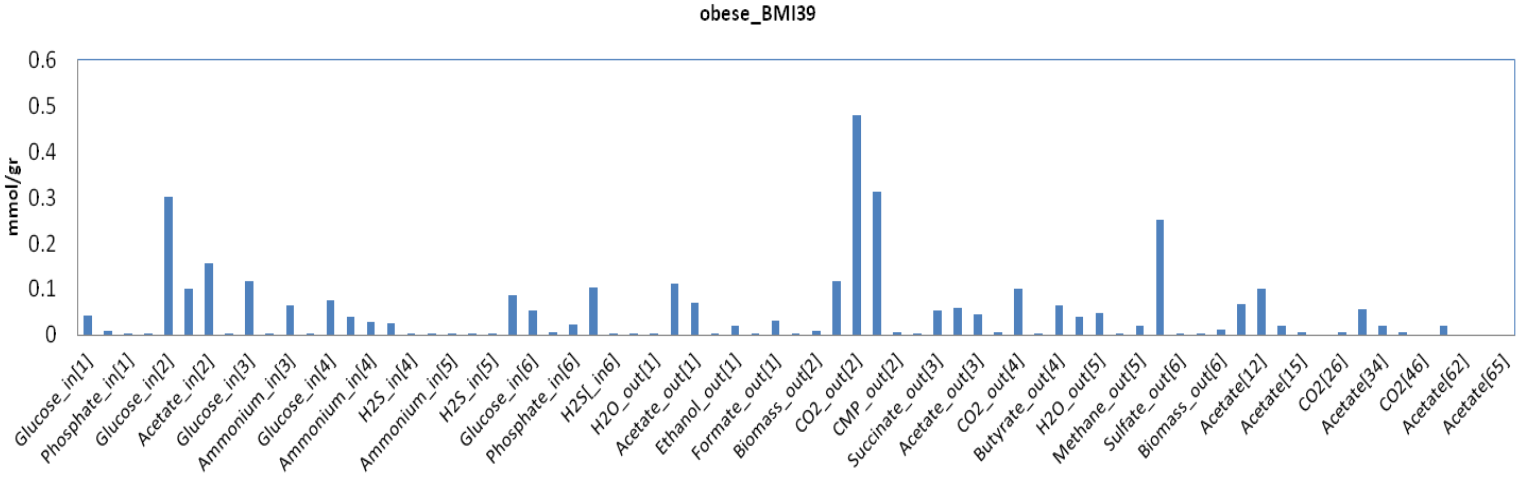
Considering rapid absorption of SCFAs in colon, the secretion rate of 5%-20% was assumed for butyrate, propionate and acetate(Kovatcheva-Datchary et al., 2013). We used the cell dry weight of E. coli as an average and converted the published experimental data for microorganisms' abundances from (cells/gram feces) to (gram DW/gram feces) by multiplying it to 2.8*10^(-13). It has been assumed that all produced hydrogen was consumed within the community. Microorganisms interact Comparing the model outputs with published data we found a good match, 60%-70% between model predictions and experimental data. Taking into account the accuracy of individual reconstructed models and also the minimization and controlling procedures we used to personalize the simulations, 65% agreement of predictions can be considered as a great result. Figures 3.2 and 3.3 depict the outputs of simulation in comparison to the available experimental data for a lean and an obese subject.




**3.2 Lean subject with BMI 20.17 and age 38. To compare experimental data and simulation outputs, an absorption rate of 80% - 95% was assume for butyrate, propionate and acetate. Indices from 1 to 6 represent: bifido, fap, bth, ere, msi and bromi.**




**3.3 Obese subject with BMI 39.5 and age 54. To compare experimental data and simulation outputs, an absorption rate of 80% - 95% was assume for butyrate, propionate and acetate. Indices from 1 to 6 represent: bifido, fap, bth, ere, msi and bromi.**

## *4. Discussion and future perspective*

Here we introduced COSMOS, a multidimensional distributed computational framework for simulating and analyzing of biological ecosystems and communities considering their behavior in a complex web of interactions. We demonstrated that how COSMOS can be used to predict the optimality condition of growth and corresponding metabolic interactions expressed in several dimensions, for different members in a community of microorganisms. Unlike earlier community modeling approaches that rely on experimental data to formulate the community-level objective functions, we developed a new mechanism integrating local and community level orientation forces to from the global objective function. This approach leads to develop a universal methodology to construct the community-specific fitness landscape through the network of tradeoffs between local and global forces. Also semi-dynamic feasible space introduced as a new feature, together with integrated objective function, can project evolutionary nature of biological systems into the model. Community control method, based on influence and support effect of individuals inside the network, was developed to enable the COSMOS to simulate the behavior of complex communities with minimum number of known parameters.

The main goal of studying and analyzing the microbial communities and developing tools like COSMOS, is the targeted manipulation of these systems towards enhancing beneficial functions and suppressing harmful ones. Looking ahead, tremendous implications can be seen for this effort. More effective diagnostics and medical interventions can be obtained by predicting the specific effect of drug treatments, nutritional additives and microbial supplements on microbiome metabolic activities. Detailed view of microbial interactions would increase safety, efficiency and effectiveness of targeted therapies designed to eliminate the diseases by manipulating the microbiome(Greenblum et al., 2013). It can provide In-silico support to more informed design of treatment with prebiotics and probiotics, and to have a better prediction of the stable shift period towards healthier mode(Faust and Raes, 2012). It can be extended as a comprehensive computational framework to model the integrated host and microbiota interactions.

For future, COSMOS can be improved by adding more features of community network structure, integrating energy and biomass objective functions in a multi-objective and multi-dimensional model to form the Pareto front, modifying the community-level force by taking into the account the extra-community factors, and introducing the temporal and spatial dynamicity of biological systems into the model.

Clearly, there is a lot of work ahead to develop a comprehensive model of a complex biological system like human microbiome. However, continuous efforts on system-level modeling of microbiota will take us closer to a fundamental understanding of communities and will enable more rational manipulation of these systems.

## *5. Acknowledgement*

This thesis would not have been possible to be accomplished without the invaluable support of many people. I express my deepest gratitude to Prof. Jens Nielsen who provided me the opportunity to do my thesis in his group. Many thanks to Saeed Shoaie for his guidance and comments all through learning process of my thesis project. I have to appreciate Dr. Sergio Bordel for his ideas and remarks. And finally, I thank the one who supported me with her patience and love, Dilara.

## 6. References


Acheson, D.W., and Luccioli, S. (2004). Microbial-gut interactions in health and disease. Mucosal immune responses. Best practice & research Clinical gastroenterology *18*, 387-404.

Aminov, R.I., Walker, A.W., Duncan, S.H., Harmsen, H.J., Welling, G.W., and Flint, H.J. (2006). Molecular diversity, cultivation, and improved detection by fluorescent in situ hybridization of a dominant group of human gut bacteria related to Roseburia spp. or Eubacterium rectale. Applied and environmental microbiology *72*, 6371-6376.

Bonacich, P. (2007). Some unique properties of eigenvector centrality. Soc Networks *29*, 555-564.

Brandes, U. (2008). On variants of shortest-path betweenness centrality and their generic computation. Soc Networks *30*, 136-145.

Claesson, M.J., Jeffery, I.B., Conde, S., Power, S.E., O'Connor, E.M., Cusack, S., Harris, H.M., Coakley, M., Lakshminarayanan, B., O'Sullivan, O*., et al.* (2012). Gut microbiota composition correlates with diet and health in the elderly. Nature *488*, 178-184.

Claesson, M.J., O'Sullivan, O., Wang, Q., Nikkila, J., Marchesi, J.R., Smidt, H., de Vos, W.M., Ross, R.P., and O'Toole, P.W. (2009). Comparative analysis of pyrosequencing and a phylogenetic microarray for exploring microbial community structures in the human distal intestine. PloS one *4*, e6669.

Dethlefsen, L., and Relman, D.A. (2011). Incomplete recovery and individualized responses of the human distal gut microbiota to repeated antibiotic perturbation. Proceedings of the National Academy of Sciences of the United States of America *108 Suppl 1*, 4554-4561.

Dubitzky, W., Southgate, J., and Fuss, H. (2011). Understanding the dynamics of biological systems : lessons learned from integrative systems biology (New York: Springer).

Faust, K., and Raes, J. (2012). Microbial interactions: from networks to models. Nature reviews Microbiology *10*, 538-550.

Feist, A.M., Herrgard, M.J., Thiele, I., Reed, J.L., and Palsson, B.O. (2009). Reconstruction of biochemical networks in microorganisms. Nature reviews Microbiology *7*, 129-143.

Greenblum, S., Chiu, H.C., Levy, R., Carr, R., and Borenstein, E. (2013). Towards a predictive systems-level model of the human microbiome: progress, challenges, and opportunities. Current opinion in biotechnology.

Grimm, V., Revilla, E., Berger, U., Jeltsch, F., Mooij, W.M., Railsback, S.F., Thulke, H.H., Weiner, J., Wiegand, T., and DeAngelis, D.L. (2005). Pattern-oriented modeling of agent-based complex systems: lessons from ecology. Science *310*, 987-991.

Guarner, F., and Malagelada, J.-R. (2003). Gut flora in health and disease. Lancet (London, England) *361*, 512-519.

Isolauri, E., Kalliomaki, M., Rautava, S., Salminen, S., and Laitinen, K. (2009). Obesity - extending the hygiene hypothesis. Nestle Nutrition workshop series Paediatric programme *64*, 75-85; discussion 85-79, 251-257.

Karlsson, F.H., Tremaroli, V., Nookaew, I., Bergstrom, G., Behre, C.J., Fagerberg, B., Nielsen, J., and Backhed, F. (2013). Gut metagenome in European women with normal, impaired and diabetic glucose control. Nature.

Kauffman, K.J., Prakash, P., and Edwards, J.S. (2003). Advances in flux balance analysis. Current opinion in biotechnology *14*, 491-496.

Kitano, H. (2002a). Computational systems biology. Nature *420*, 206-210.

Kitano, H. (2002b). Systems biology: a brief overview. Science *295*, 1662-1664.

Kovatcheva-Datchary, P., Tremaroli, V., and Bäckhed, F. (2013). The Gut Microbiota. In The Prokaryotes (Springer), pp. 3-24.

Lewis, N.E., Nagarajan, H., and Palsson, B.O. (2012). Constraining the metabolic genotype-phenotype relationship using a phylogeny of in silico methods. Nature reviews Microbiology *10*, 291-305.

Nicholson, J.K. (2006). Global systems biology, personalized medicine and molecular epidemiology. Molecular systems biology *2*, 52.
Oberhardt, M.A., Puchalka, J., Fryer, K.E., Martins dos Santos, V.A., and Papin, J.A. (2008). Genome-scale metabolic network analysis of the opportunistic pathogen Pseudomonas aeruginosa PAO1. Journal of bacteriology *190*, 2790-2803.
Orth, J.D., Thiele, I., and Palsson, B.O. (2010). What is flux balance analysis? Nature biotechnology *28*, 245-248.
Palmer, C., Bik, E.M., DiGiulio, D.B., Relman, D.A., and Brown, P.O. (2007). Development of the human infant intestinal microbiota. PLoS biology *5*, e177.
Png, C.W., Linden, S.K., Gilshenan, K.S., Zoetendal, E.G., McSweeney, C.S., Sly, L.I., McGuckin, M.A., and Florin, T.H. (2010). Mucolytic bacteria with increased prevalence in IBD mucosa augment in vitro utilization of mucin by other bacteria. The American journal of gastroenterology *105*, 2420-2428.
Salonen, A., Nikkila, J., Jalanka-Tuovinen, J., Immonen, O., Rajilic-Stojanovic, M., Kekkonen, R.A., Palva, A., and de Vos, W.M. (2010). Comparative analysis of fecal DNA extraction methods with phylogenetic microarray: effective recovery of bacterial and archaeal DNA using mechanical cell lysis. Journal of microbiological methods *81*, 127-134.
Sanz, Y., De Pama, G., and Laparra, M. (2011). Unraveling the ties between celiac disease and intestinal microbiota. International reviews of immunology *30*, 207-218.
Schwiertz, A., Taras, D., Schafer, K., Beijer, S., Bos, N.A., Donus, C., and Hardt, P.D. (2010). Microbiota and SCFA in Lean and Overweight Healthy Subjects. Obesity *18*, 190-195.
Sekirov, I., Russell, S.L., Antunes, L.C., and Finlay, B.B. (2010). Gut microbiota in health and disease. Physiological reviews *90*, 859-904.
Turnbaugh, P.J., Hamady, M., Yatsunenko, T., Cantarel, B.L., Duncan, A., Ley, R.E., Sogin, M.L., Jones, W.J., Roe, B.A., Affourtit, J.P*., et al.* (2009). A core gut microbiome in obese and lean twins. Nature *457*, 480-484.
Turroni, F., Ribbera, A., Foroni, E., van Sinderen, D., and Ventura, M. (2008). Human gut microbiota and bifidobacteria: from composition to functionality. Antonie van Leeuwenhoek *94*, 35-50.
Whitman, W.B., Coleman, D.C., and Wiebe, W.J. (1998). Prokaryotes: the unseen majority. Proceedings of the National Academy of Sciences of the United States of America *95*, 6578-6583.