Department of Theoretical and Computational Physics

Babeş–Bolyai University

Ph.D thesis summary

STATISTICAL PHYSICS MODELS FOR BIOLOGICAL AND SOCIOLOGICAL PHENOMENA

Derzsi Aranka

Scientific advisor: Prof. Dr. Néda Zoltán

Cluj–Napoca

2012

Keywords: complex systems, statistical physics, macroecological patterns, neutral theory, neutral community models, spatially explicit models, networks, social networks, random network models

Contents

Introduction

Statistical physics, employing mathematical tools of probability theory and statistics to describe large ensembles of interacting units, traditionally has been dealing with problems related to the field of physics. The first application of statistical methods in physics aimed the explanation of the observed macroscopic properties of materials from the microscopic characteristic of individual atoms and molecules. Nowadays, reference to statistical physics is of common occurrence in fields that seem – at first sight – have nothing to do with physics. Economics and finance, biology and ecology, psychology and medicine, sociology and music are all fields, among others, where tools of statistical physics have found applications. In the frame of the foresaid fields, one can easily find many examples for systems composed of a large number of elements in interaction with each other that recall "self-organization", "critical behavior" or "phase transition" for instance, terms well-known from statistical physics. From this point of view, cross-disciplinary use of the statistical physics models, adapted to given problems in various fields, may appear less peculiar. This thesis follows this line by addressing the study of complex systems from the field of biology and sociology. A simple spatially explicit two-parameter model and models based on random diffusion of seeds are introduced to describe diversity patterns observed in tropical forests, ecosystems comprising a large number of individuals from ecologically similar species, referred to as neutral ecological communities. These models are studied by Monte Carlo simulations. As sociological system, study of the collaboration network generated by the Erasmus student mobilities is considered, different projections of the network are analyzed and modeled.

Structure of the thesis

The thesis is structured into two main parts: (I) Modeling ecological patterns and (II) Modeling social networks.

In the first part of the thesis on modeling ecological communities an introductory chapter is dedicated to clarify the basic concepts in ecology, to define the ecological patterns of interest and to introduce the main theoretical approaches to describe ecosystems. The second chapter reviews the evolution of theoretical description and modeling of ecosystems, from simple statistical models to complex approaches with biological processes included, and reveals the establishment of neutral models in ecology. The third chapter presents a natural neutral like ecological community and the corresponding experimental data sets which serve as a reference for testing the performance of the models. In the fourth and fifth chapters original neutral models for tropical tree diversity patterns are introduced and studied, a simple spatially explicit neutral model and models with dynamics based on seed diffusion.

The second part of the thesis, consisting of two chapters, is reserved for network studies: a general introduction on networks in chapter six is followed by the presentation of a social network of student mobility and modeling of this network in chapter seven.

At the end of the thesis, a summary of the thesis and a list of publications is given.

Review of Part I – Modeling ecological patterns

Ecology is a sub-discipline of biology that deals with the investigation of ecological systems. *Ecological systems* are spatial units of the Earth, interacting building blocks of the environment, including all the living organisms (biotic component such as competitors, predators, decomposers, etc.) and components of the surroundings (abiotic component such as soil humidity, soil nutrients, sunlight intensity, air temperature, wind speed, etc.) within their boundaries. Ecological systems exhibit various *ecological patterns*, such as the abundances of organisms and species, spatial distribution of species, species diversity. Features of the observed ecological patterns depend on the temporal and spatial scales at which the given ecological system is investigated. *Macroecology* [1] studies the relationships between organisms and their environment at large spatial scales in order to characterize and explain the patterns of abundance, distribution, and diversity [2]. Single trophic-level communities, containing a large number of ecologically similar species competing only for the available resources, define *neutral* systems. Such communities are especially suitable to be approached by statistical models, based on simple hypothesis. As original contribution in the first part of the thesis spatially explicit neutral models are presented – a simple two-parameter neutral model [3] and models based on seed diffusion [4] – aimed to describe the macroecological patterns observed in a natural neutral-like ecological community, the well-mapped tropical tree plot of the Barro Colorado Island (BCI) [5].

Macroecological patterns

The most important macroecological patterns are the biological diversity the species-area relationship, the spatial distribution of species and the relative species abundances.

Biological diversity or *biodiversity* is the variety of life forms in a particular area.

Over spatial scales, there are three diversity indices commonly used for measuring biodiversity [6]: α-*diversity* (usually given as the number of different species of a given ecosystem), β-*diversity* (compares biodiversity of different ecosystems, it is the number of species unique to each of the studied ecosystems) and γ-*diversity* (describes the species diversity at geographic scales).

The *species-area relationship* (SAR), characteristic of a given region, is the relationship between the area of a habitat and the number of different species found in that area, is one of the most documented patterns in macroecology [7]. The slope of the species-area curve is a measure of biodiversity in the geographical region investigated.

The *spatial distribution of species* is the ecological pattern defined by the spatial positions of individual organisms of a given species. This pattern can be characterized by the *spatial autocorrelation function* (SAF) [3]. Spatial distribution patterns are generally not static, their spatial arrangement can change in response to changes of the environmental conditions (seasonal changes, change in the level of available resources, etc.) and also depend on the spatial and temporal scales at which they are investigated.

Among the most studied patterns in macroecology, *relative species abundance* (RSA) is the one which attracted exceptional attention. This pattern indicates how rare or common a species is relative to the other species in the studied ecosystem [8]. The RSA characterizes the frequency of species having a given abundance. In general, most species are represented by only one individual in the sample and there are only a few species with large number of individuals. There are different ways to present the RSA by the construction of relative species abundance distribution plots: (i) *frequency histogram* or *Preston-plot* [9], (ii) *rank-abundance plot* or *Whittaker-plot* [10], (iii) *probability density function* for finding a species with a given abundance.

Approaches to describe ecosystems

There are two main but conflicting perspectives on the nature of ecological communities: the *niche-assembly perspective* [1] and the *dispersal-assembly perspective* [8].

The term *niche* in ecology describes the role of a species in its environment, it can be considered as a collection of all the biotic and abiotic factors that define the life of organisms of a species. The classical niche theory of ecology is based on the principle of competitive exclusion. This principle, often referred as the Gause's law of competitive exclusion [11], states that two species living in the same habitat (with constant conditions) and competing for the same resources cannot coexist for a long time – the niche partitioning of the system is realized and each species is the best competitor in its own ecological niche. According to the niche-assembly perspective, ecological systems are closed, stable systems, they are found in the state of interactive equilibrium of their species, where all the possible niches are occupied by single species. Under such conditions its is hard to disturb the equilibrium, the appearance of a new species is not favored.

The neutral theory in ecology, introduced by Stephen Hubbell [8], received much attention due to the unusual assumptions it has as a base. This approach presumes the functional equivalence of species and considers that individuals of an ecosystem, composed of organisms of a single trophic level, obey exactly the same dynamical rules regardless of the species they belong to. Contrary to the niche-theory, which has in the focus the genetic fitness of species, the neutral theory attributes the key role to randomness in the development of the observed macroecological patterns. This theory allows many species to occupy the same ecological niche for a long time, new species are permitted to show up in the system. In this view, also called dispersal-assembly perspective, the ecological communities are not in equilibrium, they are open, continuously changing complex systems governed by random speciation, extinction, dispersal and ecological drift [8]. It is emphasized that the principle of neutrality is applicable in case of single trophic-level ecosystems (for instance plankton communities, tropical forests, coral reefs) that contain very similar species competing only for the available resources. The models aimed to describe such neutral communities based on the assumption of functional equivalence of species, are called *neutral models*.

Theories of species abundance and neutral modeling in ecology

Most of the theoretical approaches to describe ecosystems focus on the reproduction and explanation of the observed patterns of species abundance distribution in various ecosystems. Early attempts of the study of species abundance consisted in inductive approaches: mathematical models were fit to the observed distribution of species abundance, or rules for the partitioning of available resources among species were proposed, without attempts to explain the origin of the observed patterns and to reveal the underlying biological processes. Opposed to the statistical models of species abundance, the dynamic community models in ecology incorporate demographic processes and offer description of the community dynamics by containing the fundamental processes of birth, death and dispersal. The neutral models also belong to the group of dynamic community models. Despite their amazing simplicity, neutral models produced reasonable predictions for some taxa [8, 12], however, the assumption of functional equivalence of individuals regardless of species in an ecological community generated serious debates among ecologists. In these days, the neutral approach is more and more received as a null model of community dynamics and attempts are made to bridge it to the niche-assembly theory [13].

In the past decade a series of neutral models based on Hubbell's simple model [8] have been developed [14]. In these models, called *spatially implicit* or *mean-field type* neutral models, the spatiality of the individuals is lost, their exact spatial location in the local community is not taken into account. Recently, *spatially explicit* neutral models have been introduced [3, 4, 15, 16]. These models replace the implicit local community of mean-field type neutral models with a system where the spatial position of each individual is known explicitly. Such models can take into account distance-dependent processes and incorporate more realistic assumptions on the dynamics of the system than it was possible in case of spatially implicit approaches.

A simple spatially explicit neutral model

A simple spatially explicit two-parameter neutral model has been developed [3], aimed to describe the experimental ecological patterns of the BCI tropical tree census. This model is inspired by the classical voter model [17] and the single-parameter spatially explicit neutral model of Zillio [16].

In this model a local community is considered in connection with an external species pool, the metacommunity. The space of the local community is modeled by considering a uniform square lattice, individual organisms are represented by grid points of this predefined lattice. It is assumed that the lattice is always completely filled up, the total number of individuals, N is constant in the local community.

Initially, a randomly chosen Potts state is assigned to each lattice site. The number of possible states is W, the number of species in the metacommunity. At each discrete time step a lattice site is randomly chosen and its state is changed. This process models the death of an individual. The vacancy created by death is immediately occupied by an offspring of an individual from inside or from outside the local community. The p and q parameters govern the selection of the species of the new individual as follows:

- with probability p, the species of the new individual is chosen to be the same as the species of a randomly selected individual from among the remaining $N-1$ individuals in the lattice;

- with probability q, a randomly selected species from the ensemble of possible W species is assigned to the chosen site;
- with probability $1-p-q$, the state of the chosen site takes the species of one of its 8 Moore-neighbors.

The steps above are repeated until a dynamic equilibrium of the system is reached.

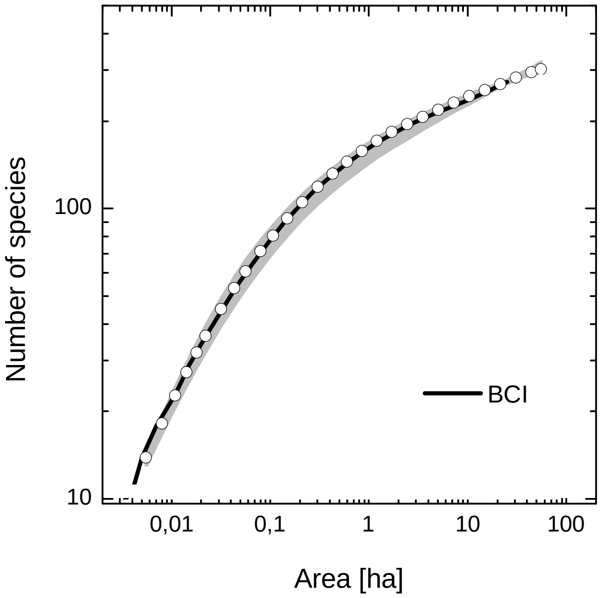

Figure 1: *Species-area curves: comparison of simulation (grey shaded area) and experimental (black line) results. The shaded area corresponds to a set of 50 simulations, and represents the mean value ± 2 standard deviations. The white circles represent the simulation result that fits the experimental data best. Simulation parameters:* $p = 0.3$ *and* $q = 1.3 \times 10^{-4}$. *Experimental data: BCI census (census year: 1995).*

As experimental data the eastern half part of the 50 ha BCI plot is considered [5]. This 500 m × 500 m region contains in total 112543 trees and shrubs with diameter at least 1 cm at breast hight in the census for the year 1995. The total number of species in the modeled ecological community is about 270. Another characteristic of this community is the power-law type behavior of the species-area curve in the limit of large areas with exponents between 0.14 and 0.16 for different census years. Further, the spatial distribution of conspecific individuals is characterized by a spatial autocorrelation function exhibiting a power-law type decay.

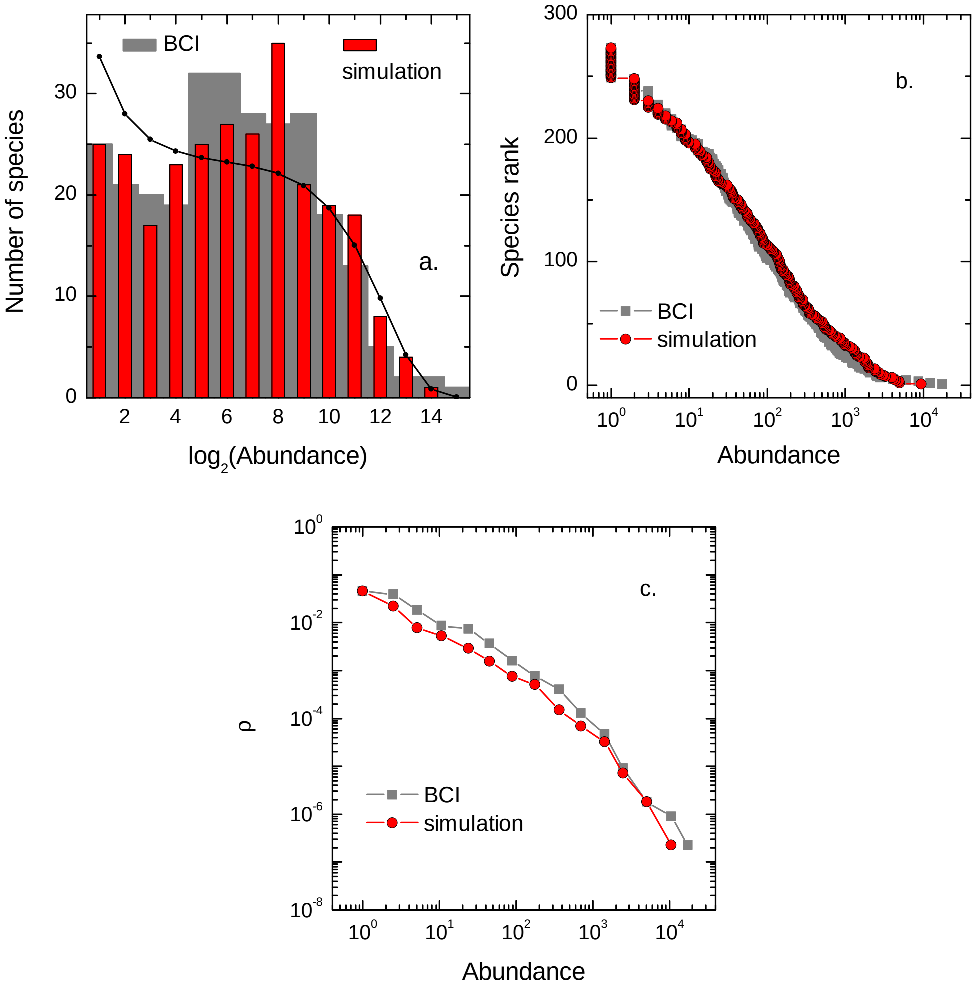

Figure 2: *Simulation results (red) for the RSA curve compared with the measurement data (light grey), and the predictions of the Fisher log series for* $\alpha = 33.64$ *(solid black lines). Panel (a) shows the Preston plot, (b) the rank-abundance plot, and (c) the probability density function. Simulation parameters:* $p = 0.3$ *and* $q = 1.3 \times 10^{-4}$*. Experimental data: BCI census (census year: 1995).*

Large scale MC simulations have been performed on the model. We were looking for those values of the p and q parameters that produce species number, relative species abundances (RSA) distribution, species-area relationship (SAR) and spatial distribution of individuals of a given species in accordance with the experimental data on the BCI community.

For reasonable values of the p and q parameters (optimal values were found to be 0.3 and $1.3{\times}10^{-4}$ respectively), the model reproduces the shape of the species-area curve (Figure 1) and RSA curve (Figure 2), as measured in the BCI dataset. The species-area function is reproduced over the full range of areas. The model fails to reproduce the experimentally observed power-law decay of the spatial auto-correlation function of individuals within a species as a function of the separation distance. This result suggests that in reality long-range interactions at the local scale might be more important than it was considered in this simple approach.

Despite its simplicity, the present model is successful in reproducing the majority of the experimental macroecological measures of interest for the studied real ecological community. More complex models are needed in order to fully understand all aspects of such communities and to reproduce all the macroecological patterns characterizing these systems.

Spatially explicit neutral models based on seed diffusion

Seed diffusion (seed dispersal) is the travel of seeds, both in time and space, away from their parent individual. In case of plant communities, diffusion of seeds ensures migration of species and makes possible for them to reach locations favorable for their survival and reproduction, or to colonize vacant places in a community. Therefore, this process has important role in maintaining species diversity. In real ecological communities – where the transport of seeds can be performed by wind or gravity, through water, or by animals –, seeds can travel long distances starting from the position of their parent before they shoot or become completely unable to grow and die.

We have proposed a model to describe the diffusion of seeds in a given ecological community. Considering a community occupying an area of $L \times L$ and containing a number of individuals of different species, it is assumed that time to time individuals in the community emit certain number of seeds which disperse in the 2D space. Further, it is assumed that (i) each individual emits the same number of seeds at each time step, which randomly diffuse in the 2D space, (ii) the seeds survive W time steps (survival period), after which they become unable to grow and die and (iii) all the seeds have the same diffusion constant and the same W survival period.

The spatial region of size $L \times L$ is divided into small regions by defining a uniform square lattice of size $N \times N$ on it. The schematic picture of the community with an assigned grid is shown in Figure 3. Each cell contains certain number of individuals, $K_{i,j}$ stands for the number of individuals in cell (i, j), while $S_{i,j}$ gives the number of

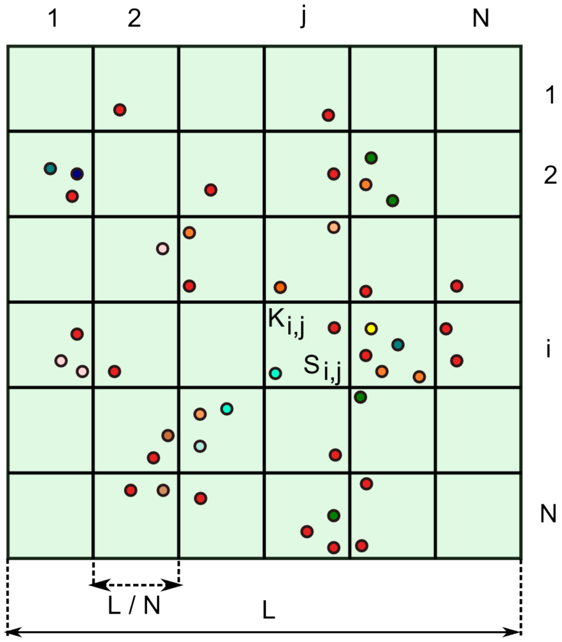

Figure 3: *Schematic picture of the community of size $L \times L$ with a uniform square lattice of size $N \times N$ assigned to it. The filled circles correspond to positions of individuals. Each cell of the grid, labeled by (i, j), contains a certain number of individuals ($K_{i,j}$) belonging to $S_{i,j}$ different species.*

different species at that site. In each cell there can be seeds originating from other cells. $Q_{ij}(k, l)$ denotes the number of seeds in cell (i, j) originating from the cell (k, l).

A statistical description of the spatial distribution of seeds is considered. A simple two-dimensional unbiased random walk is assumed for the motion of the seeds, in unit time a seed can make a unit step defined as the grid constant L/N, randomly moving to a neighboring cell. The $Q_{ij}(k, l)$ numbers are computed for each pair of cells, according to the basic statistical properties of simple random walks.

A neutral model based on seed diffusion (1)

A spatially explicit neutral model based on the diffusion process of seeds has been elaborated in order to reproduce the observed macroecological measures for the BCI forest plot. In this model a local community is considered, embedded in a metacommunity with U species. Two main processes are assumed in the model, the dispersal of seeds capable of growth and speciation, the parameters governing these processes are W and q, respectively.

Individuals of the local community are spatially arranged according to the posi-

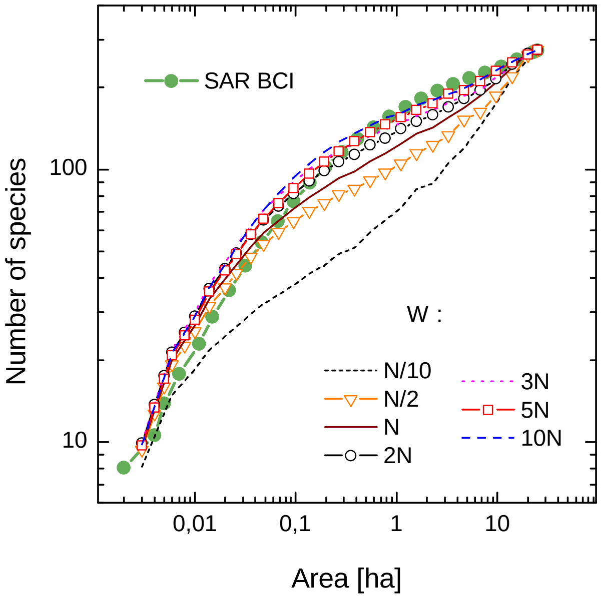

Figure 4: *Species-area curves. Simulation results (different lines and lines with open symbols) are presented for various values of the W seed survival period with fixed size of the lattice (N=100) and speciation rate ($q = 3.1 \times 10^{-4}$). The total number of species is 273. The continuous line with filled circles shows the species-area relationship for the BCI census data (census year: 1995).*

tion of individuals in the experimental plot. A uniform $N \times N$ lattice is constructed over the local community and the $K_{i,j}$ and $Q_{i,j}(k,l)$ quantities are defined for all the $(i,j),(k,l) \in N \times N$ cell pairs.

It is assumed that time to time one individual dies in the community and its place is taken by a new individual. Our basic hypothesis is that the origin of this new individual is a seed found in the neighborhood of the vacant place. There can be many seeds in the neighborhood of the vacancy, released by individuals at different places of the local community, therefore, it is assumed that the seed which originates the new individual is a randomly selected one from the set of seeds accumulated in the neighborhood of the empty place. The vacancy can be also filled, with smaller chance, by an individual originating from a seed of an individual from outside the local community (vacancy filled by immigration).

The dynamics of the system is defined by the following rules:

- in each time step one individual is randomly selected; this individual is considered death, a vacancy at its spatial position is created, in cell (i,j);
- with probability q, the vacancy at (i,j) is occupied by a randomly selected species

from the reservoir U;

- with probability $1-q$, the vacancy at (i,j) is occupied by the descendent of an individual already present in the community.

In the simulations, the steps of the model are repeated until a dynamic equilibrium of the system is reached. For a speciation rate of $q = 3.1 \times 10^{-4}$, simulation of the model converges to a species number conforming to the experimental one. Assuming a maximum survival distance of about 2500 m, the model produces results in agreement with the experimental ecological measures: the species-area scaling and species abundance distribution are reproduced (Figure 4 and Figure 5), the autocorrelation function of the most abundant species shows the power-law type decay in the limit of small distances.

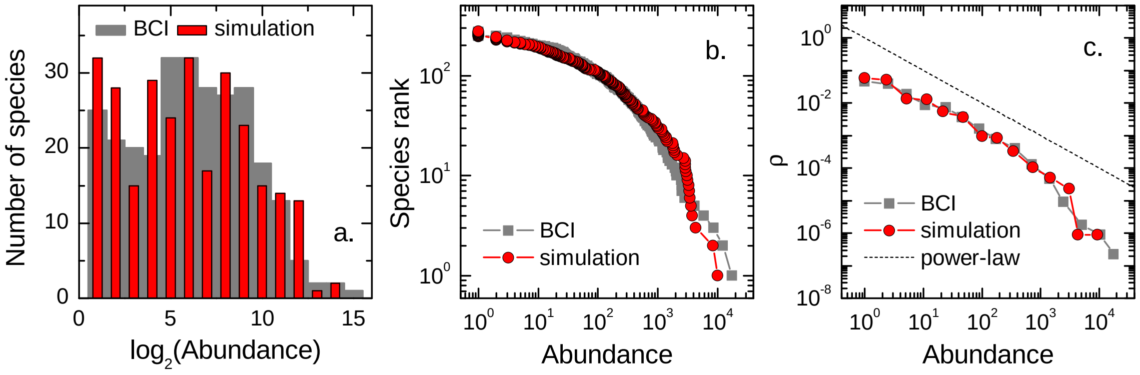

Figure 5: *Relative species abundances. The RSA distribution is plotted in different ways: (a) Preston-plot, (b) probability density function and (c) rank-abundance plot. Simulation results are presented for $N = 100$ lattice size, $W = 500$ seed surviving period and $q = 3.1 \times 10^{-4}$ speciation rate. Experimental data obtained from the BCI dataset is also plotted (census year: 1995). For guiding the eye in panel (C) a power-law with exponent -1 is also indicated with a dashed line.*

Obviously, the model neglects many existing ecological processes, however, the results obtained suggest that the primary forces responsible for the observed patterns are successfully captured in this model.

A neutral model based on seed diffusion (2)

Another spatially explicit neutral model based on the diffusion process of seeds has been introduced [4]. This model has three parameters: the maximum diffusion distance of seeds (W), the probability of speciation in the system (q) and the size of the metacommunity (M). Using this model to describe the macroecological patterns of the tropical tree community of BCI, two parameters of the model (the speciation rate and the size of the metacommunity) are fixed to values characteristic of the studied real ecological community, thus obtaining a model with only one adjustable parameter. Therefore, all the relevant macroecological measures are fitted by one parameter, the survival period of seeds.

A difference compared to the spatially explicit neutral model based on seed diffusion presented before, is that in this model individuals of the local community are arranged on a uniform grid of size $N \times N$, each point of the grid represents a single individual organism. Further, the metacommunity is also defined in a spatially explicit way, due to the backward-in-time simulation technique [18].

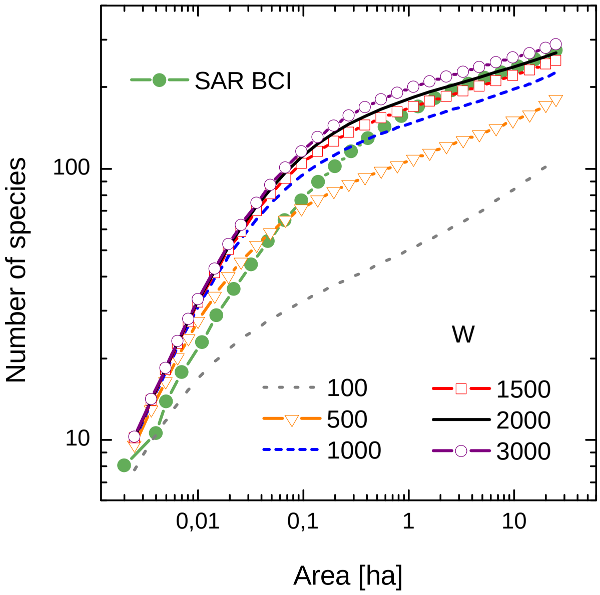

Figure 6: *Species-area curves. Simulation results (different lines and lines with open symbols) are presented for different values of the W seed surviving period with fixed speciation rate ($q = 1.0 \times 10^{-5}$). The continuous line with filled circles shows the species-area relation for the Barro Colorado Island census data (census year: 1995).*

The dynamics of the model is defined by the following rules:

- in each step one individual is randomly selected from all the lineages being traced at that moment (active lineages); the selected individual is the one that was most recently born – the origin of this individual has to be defined;
- with probability q, speciation happens in the lineage – if speciation takes place,

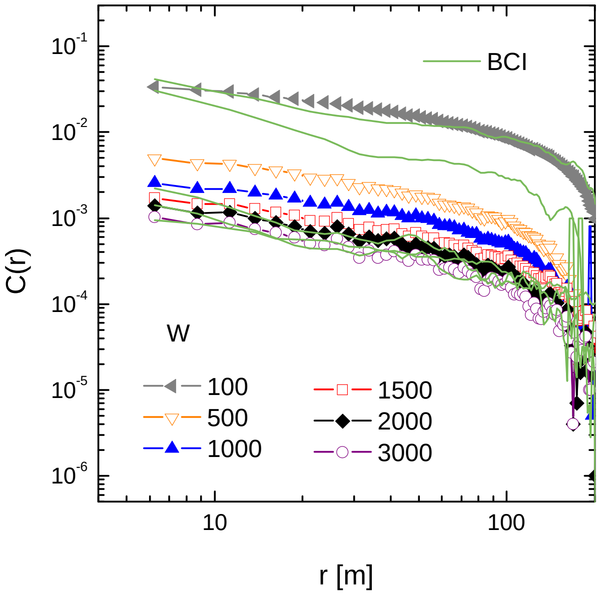

Figure 7: *Autocorrelation functions of the most abundant species. Simulation results (lines with symbols - averaged results) are presented for different values of the W seed survival period with fixed speciation rate ($q = 1.0 \times 10^{-5}$). Results obtained for the five most abundant species in the BCI census data are shown with lines (census year: 1995).*

tracing of the lineage is stopped (finalized lineage) and the selected individual is considered to be the root of a new species;

- with probability $1 - q$, a parent individual is defined for the selected individual – the position of the parent is chosen according to the dispersal kernel of the model.

The parent selection mechanism of the last rule is done by respecting the weights given by the $Q_{i,j}(k, l)$ values, defining the dispersal kernel of the model.

The results obtained for W=1700 and q=1.0 × 10^{-5} with a metacommunity-size corresponding to the area of the whole Barro Colorado island, are acceptable for the ecological measures considered here (Figure 6, Figure 7, Figure 8). Naturally, none of the fits for the relevant ecological measures is perfect, one cannot expect this from a simple one-parameter neutral model. The strength of the model is that, with only one freely adjustable parameter, we are able to qualitatively approach a quite complex phenomenon. Neutral approaches are more and more viewed as null models of complex community dynamics, niche-assembly elements are also needed to obtain a better description of the systems. Results of this simple approach confirms this view.

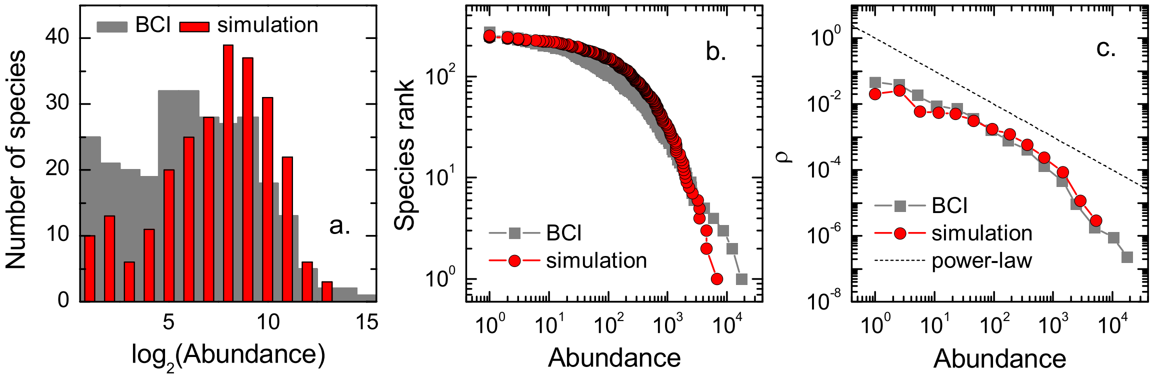

Figure 8: *Relative species abundances. The RSA is plotted in different ways: (a) Preston-plot, (b) rank-abundance plot and (c) probability density function. Simulation results are presented for* $W = 1700$ *steps seed survival period and* $q = 1.0 \times 10^{-5}$ *speciation rate. Experimental data obtained from the BCI dataset is also plotted (census year: 1995). In panel (c) a power-law with exponent* -1 *is indicated with a dashed line.*

Review of Part II – Modeling social networks

Networks, also called *graphs*, are systems of connected items. Development of the computational infrastructure and availability of large digitalized datasets on many social, natural and artificial systems in the latest decades, give the field of network research a boost: networks are found to be the backbone of all complex systems ranging from atomic levels to systems with global extent.

In *social networks* the vertices (agents) correspond to individuals and edges (ties) are formed by the various interactions between them. Investigation of social networks typically focus on the study of how individuals are connected to one another (connectivity patterns) and how well individuals are connected (centrality patterns) in real-world networks. While early works on this topic have been realized in the frame of social sciences, nowadays, investigation of the structure and dynamics of human contacts is the subject of interdisciplinary research projects. During the past decades, the scale of the investigated networks has been also shifted from small groups of people to large-scale social systems.

As original contribution to this topic, in the second part of the thesis a large-scale social network defined by international student mobilities in frame of the Erasmus program is analyzed, the patterns of student mobility for the year 2003 are investigated and simple models are proposed to reproduce the observed characteristics [19].

Study of networks

Study of networks focuses on the analysis and modeling of the structure, evolution and dynamics of such systems in order the reveal the interesting statistical properties and provide models to predict their future behavior. The main measures used to qualitatively describe the patterns of interconnections, the topology of networks, are the average path length, the clustering coefficient and the degree distribution. The investigation of these measures in real-world networks and the desire to find the mechanisms responsible for the development of the observed structures resulted in modeling approaches that can be organized in three groups: (i) random graphs (Erdős–Rényi model [20]), (ii) small-world networks (Watts–Strogatz model [21]), and (iii) scale-free graphs (Barabási–Albert model [22, 23]). While models falling into the group of random networks and small-world networks focus on the reproduction of the network topology considering constant number of nodes and random attachment rules for the nodes, the scale-free type network models reckon with the evolution of networks in time by considering growth of the system and assuming preferential attachment rules for the nodes.

A social network of student mobility

We have studied a large-scale social network, the Erasmus student mobility (ESM) network comprising student exchange pathways between European universities: nodes of the network are the universities involved in the Erasmus student exchange and the links are student mobility between them. This network also reflects the graph of professional connections built up between professors and researchers working at European universities. The database we studied contains information on 134330 student mobility within the Erasmus program, involving 2333 European universities in the year 2003.

The non-directed and non-weighted ESM network has been analyzed (NNESM network): the nodes of the network are the universities and there is a link between two universities in case of student exchange in either direction between them. The NNESM network has $N = 2333$ nodes and $L = 37492$ links. In Figure 9 a part of the whole structure of this network is shown.

Contrary to the expectations, it is found that the NNESM network is not a scale-free one, and the tail of its degree distribution can be fitted by an exponential function, not typical for social networks. Large-scale social networks are usually scale-free, exhibiting power-law type tail of the degree distribution [22]. The NNESM network is thus an interesting example of a social network where this observation is not true. This exponential degree distribution can be explained by assuming that the universities have exponential size-distribution and their degree and sizes are linearly proportional. Empirical data for some randomly selected European universities and the top UK and US universities confirmed the exponential nature of this size distribution and a simple

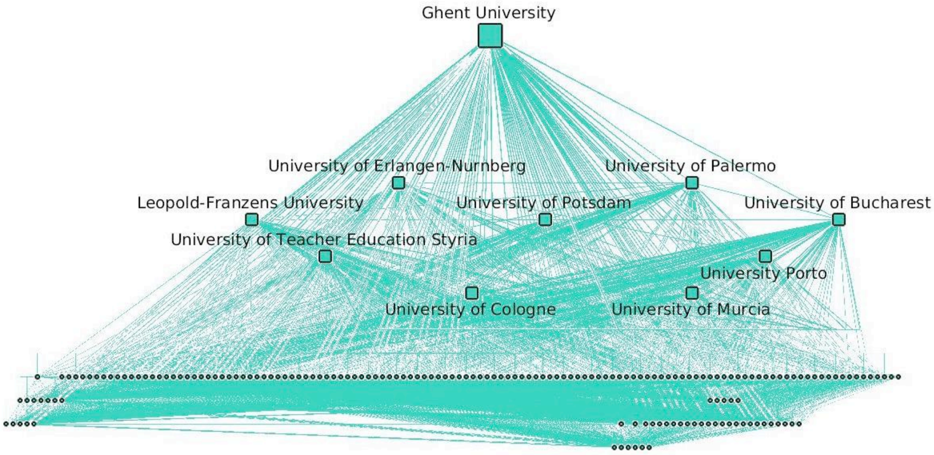

Figure 9: *The k-degree component of the NNESM network for k* = 55*. The 180 nodes of this component are presented in hierarchical arrangement. The most highly connected 10 hubs are indicated by their names.*

maximum entropy assumption leads also to this conclusion.

The global clustering coefficient of the NNESM network is 0.183, the local clustering coefficient is 0.292. The NNESM network is not a one-component graph: it exhibits several connected components, one of them being a giant component (containing 99% of all the nodes) characterized by an average distance of 2.91 between its nodes. The relatively high clustering coefficients and the non-Poissonian degree distribution suggest that the NNESM network is not an Erdős-Rényi type simple random graph.

The presence of selective linking (assortative mixing) in the NNESM network – characteristic of many social networks – has been investigated. In Figure 10 the mean degree of the neighbors of a node and the local clustering coefficients for the nodes are plotted as a function of the node degrees. The plots suggest a rather random attachment rule in establishing the connections in the NNESM network, despite the fact that the obtained degree distribution is not characteristic for a simple Erdős-Rényi type random network. These results suggest that assortative linking, here consisting of the interconnectivity of universities with high number of links, is not present in the NNESM network. The network has been approached by a configuration model as well [24], leading to the same statement.

The directed and weighted version of the ESM network has been also investigated. The properties of this network can be determined in terms of IN degree and OUT degree of the nodes.

The directed and non-weighted structure of the ESM network could be described by

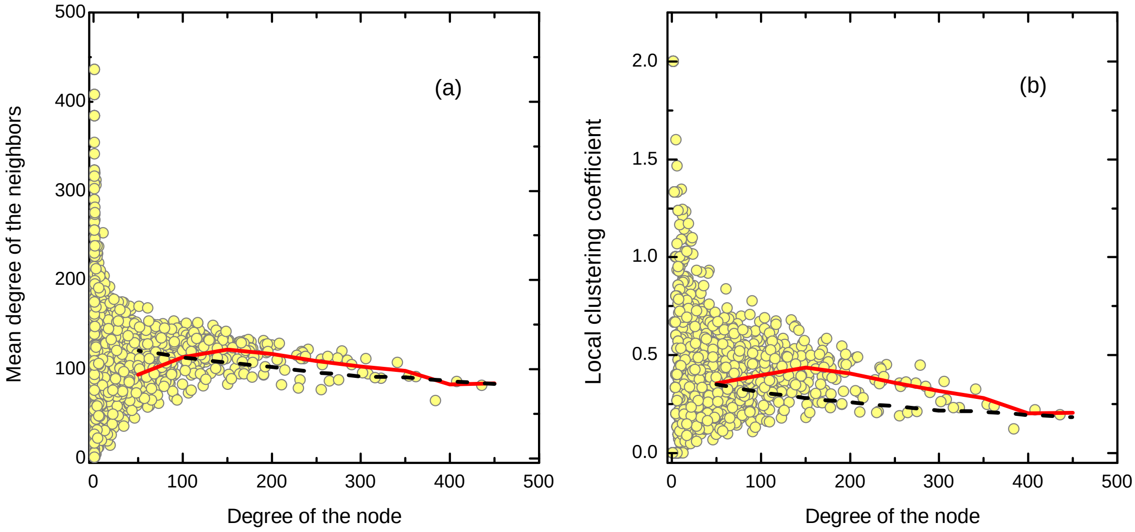

Figure 10: *Degree of a node neighbors (a) and the local clustering coefficients of the nodes (b) plotted as a function of the degree of a node (filled circles). The continuous line on both panels indicate averages of the points taken with a moving average of length 50 on the horizontal axes. Averages calculated for the network obtained by using the configuration model appear with dashed black lines.*

randomly directing the links in the network obtained by the configuration model and constructing with 1/3 probability bidirectional links. The weight distribution of the ESM network is approximated by considering that (i) the number of outgoing students from one university scales with an exponent 1.17 as a function of the number of OUT connections of the given university and (ii) there is a simple linear preferential selection rule in occupying the links: links where more students are already assigned are more probably chosen. Results obtained for the weight distribution of the network in case of the random and uncorrelated allocation of students on the OUT links is presented in Figure 11. This figure also shows the weight distribution of the original ESM network and the weight distribution obtained in case of the preferential selection rule.

The ESM network is an interesting example of a complex large-scale social network which is not scale-free and where simple random connection models are helpful in understanding the development of their topology. Although correlations proved not to be important for understanding the structure of the non-weighted ESM network, for describing the weights of the links a correlated partitioning of the students is necessary.

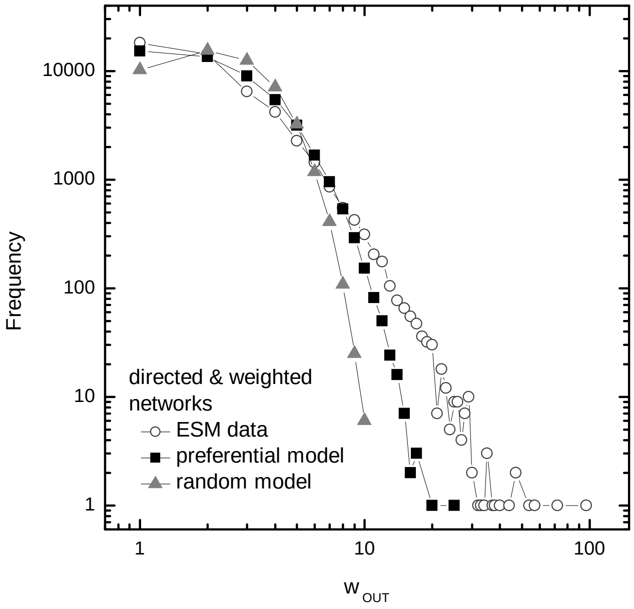

Figure 11: *Comparison of weight distributions for the directed and weighted networks: ESM network (open circles), the directed version of the network obtained by the configuration model with random (filled triangles) and preferential (filled squares) generation of weights.*

Summary

The present thesis aimed the study of biological and sociological systems by means of methods and models of statistical physics.

As biological system, the study of a neutral ecological community composed of a large number of individuals from ecologically similar species has been addressed. The diversity patterns exhibited by a well-mapped forest plot have been analyzed and approached by neutral models. A simple two-parameter spatially explicit neutral model has been introduced to capture these patterns [3]. The parameters of this model characterize the probability of speciation and the global dispersal rate of species. Despite its simplicity, the model is successful in reproducing several macroecological measures: the relative species abundance distribution and the species-area scaling for the BCI tropical tree community.

In order to reproduce another macroecological measure, the spatial distribution of species, a theoretical description of the seed dispersal process has been proposed and models based on this approach have been developed [4]. The spatially explicit neutral model, which takes into account the spatial position of trees in the real ecological community, has two parameters: the maximal survival time of seeds and the rate of immigration of species from the external metacommunity into the local community. For reasonable values of the parameters, the model produces results in agreement with the experimental ecological measures.

Another spatially explicit neutral model based on the diffusion process of seeds has been introduced with three parameters: the average survival time of the seeds, the speciation rate, and the size of the metacommunity. Fixing the two latter parameters of the model to values characteristic of the studied real ecological community, all the relevant macroecological measures are fitted by only one parameter, the survival period of seeds.

As sociological system, the collaboration network generated by the Erasmus student mobilities has been considered, different projections of this network has been analyzed

and modeled [19]. Contrary to expectations, the non-directed and non-weighted projection of this network does not exhibit the scale-free nature characteristic of social networks, but shows an exponential degree distribution, a relatively high clustering coefficient and a small radius. This network has been modeled by using a configuration model and the origin of the exponential degree distribution has been also revealed. The weighted and directed version of the network has been approached by a simple random network model.

Selected bibliography

[1] Brown JH (2005) Macroecology. Univ. of Chicago Press.

[2] May RM (1975) Patterns of species abundance and diversity. in: ML Cody, JM Diamond (Eds), Ecology and Evolution of Communities .

[3] Horvát S, Derzsi A, Néda Z, Balog A (2010) A spatially explicit model for tropical tree diversity patterns. Journal of Theoretical Biology 265: 517–523.

[4] Derzsi A, Néda Z (2012) A seed-diffusion model for tropical tree diversity patterns. arXiv:12030844v1 [q-bioPE] .

[5] Hubbell SP, Condit R, Foster RB (2005). Barro colorado island forest census plot data. `http://www.ctfs.si/edu/datasets/bci`.

[6] Whittaker RH (1972) Evolution and measurement of species diversity. Taxon 21: 213–251.

[7] Rosenzweig M (1995) Species Diversity in Space and Time. Cambridge: Cambridge University Press.

[8] Hubbell SP (2001) The Unified Neutral Theory of Biodiversity and Biogeography. Princeton University Press.

[9] Preston FW (1948) The commonness and rarity of species. Ecology 29: 254–283.

[10] Whittaker RH (1965) Dominance and diversity in land plant communities. Science 147: 250-260.

[11] Gause G (1934) The struggle for existence. Baltimore, MD: Williams and Wilkins.

[12] Néda Z, Ravasz M, Balog A, Derzsi A (2005) The species abundances distribution in a neutral community model. Studia Universitatis Babeş-Bolyai, Physica, L 2.

[13] Gravel D, Canham CD, Beaudet M, Messier C (2006) Reconciling niche and neutrality: the continuum hypothesis. Ecology Letters 9: 399–409.

[14] Chave J, Muller-Landau HC, Levin SA (2002) Comparing classical community models: Theoretical consequences for patterns of diversity. Am Nat 159: 1-23.

[15] Derzsi A (2007) Complex spatial model for macroecological patterns. Studia Universitatis Babeş-Bolyai, Physica, LII 2.

[16] Zillio T, Volkov I, Banavar JR, Hubbell SP, Maritan A (2005) Spatial scaling in model plant communities. Phys Rev Lett 95: 098101–4.

[17] Holley RA, Liggett TM (1975) Ergodic theorems for weakly interacting infinite systems and the voter model. Ann Probab 3: 643663.

[18] Rosindell JL, Wong Y, Etienne RS (2008) A coalescence approach to spatial neutral ecology. Ecological Informatics 3: 259–271.

[19] Derzsi A, Derzsy N, Káptalan E, Néda Z (2011) Topology of the erasmus student mobility network. Physica A: Statistical Mechanics and its Applications 390: 2601–2610.

[20] Erdős P, Rényi A (1959) On random graphs. Publicationes Mathematicae Debrecen 6: 290–297.

[21] Watts D, Strogatz S (1998) Collective dynamics of 'small-world' networks. Nature 393: 440–442.

[22] Barabási AL, Albert R (1999) Emergence of scaling in random networks. Science 286: 509–512.

[23] Barabási AL, Albert R, Jeong H (1999) Mean-field theory for scale-free random networks. Physica A 272: 173–187.

[24] Newman M, Park J (2003) Why social networks are different from other types of networks. Physcial Review E 68: 036122.

Publication list

Publications related to the thesis

1. Sz. Horvát, **A. Derzsi**, Z. Néda, A. Balog, A spatially explicit model for tropical tree diversity patterns, *Journal of Theoretical Biology*, **265**, 517–523 (2010).
2. **A. Derzsi**, N. Derzsy, E. Káptalan, Z. Néda, Topology of the Erasmus student mobility network, *Physica A Statistical Mechanics and its Applications* **390** 2601–2610 (2011).
3. **A. Derzsi**, Z. Néda, A seed-diffusion model for tropical tree diversity patterns, *Physica A Statistical Mechanics and its Applications* doi:10.1016/j.physa.2012.05.008 (2012).
4. Z. Néda, M. Ravasz, A. Balog, **A. Derzsi**, The species abundances distribution in a neutral community model, *Studia Universitatis Babeş–Bolyai, Physica*, L **2** (2005).
5. **A. Derzsi**, Complex spatial model for macroecological patterns, *Studia Universitatis Babeş–Bolyai, Physica*, LII **2** (2007).

Other publications in the Ph.D period

1. S. Biri, **A. Derzsi**, É. Fekete, I. Iván, Upgraded TrapCAD code, *High Energy Physics and Nuclear Physics - Chinese Edition Supplement* **31** 156–158 (2007).
2. L. Kenéz, J. Karácsony, **A. Derzsi**, S. Biri, Theoretical model for study of the voltage-current curve of a Langmuir-probe used in the hot region of the ECR plasma, *Physics Letters A* **372** 4927–4931 (2008).

3. **A. Derzsi**, P. Hartmann, I. Korolov, J. Karácsony, G. Bánó, Z. Donkó, On the accuracy and limitations of fluid models of the cathode region of dc glow discharges, *Journal of Physics D: Applied Physics* **42** 225204 (2009).

4. V. Efimova, **A. Derzsi**, A. Zlotorowicz, V. Hoffmann, Z. Donkó, J. Eckert, Influence of the anode material on the characteristics of an analytical glow discharge cell, *Spectrochimica Acta Part B* **65** 311–315 (2010).

5. **A. Derzsi**, Z. Donkó, Effect of the external electrical circuit on the ignition of the glow discharge in a Grimm-type cell, *Journal of Analytical Atomic Spectrometry* **26** 792–797 (2011).

6. I. Korolov, G. Bánó, Z. Donkó, **A. Derzsi**, P. Hartmann, Experimental study of the asymmetric charge transfer reaction between Ar^+ ions and Fe atoms, *Journal of Chemical Physics* **134** 064308 (2011).

7. J. Schulze, **A. Derzsi**, Z. Donkó, Electron heating and the electrical asymmetry effect in dual-frequency capacitive CF_4 discharges, *Plasma Sources Science and Technology* **20** 045008 (2011).

8. J. Schulze, **A. Derzsi**, K. Dittmann, T. Hemke, J. Meichsner, Z. Donkó, Ionization by drift and ambipolar electric fields in electronegative capacitive radio frequency plasmas, *Physical Review Letters* **107** 275001 (2011).

9. S. Biri, R. Rácz, J. Imrek, **A. Derzsi**, Zs. Lécz, Imaging of ECR plasma by computer simulation, *IEEE Transactions on Plasma Science* **39** 2474 (2011).

10. Á. Budai, **A. Derzsi**, P. Hartmann, Z. Donkó, Shear Viscosity of Liquid-Phase Yukawa Plasmas from Molecular Dynamics Simulations on Graphics Processing Units, *Contributions to Plasma Physics* **52** 194–198 (2012).