

1. A Multiscale Cellular Automata Model for Simulating Complex Transportation Systems

Pawel Topa, Witold Dzwiniel

AGH University of Science and Technology, Institute of Computer Science, Krakow, Poland

We present a new two-level numerical model describing the evolution of transportation network. Two separate but mutually interacting sub-systems are investigated: a "starving" environment and the network. We assume that the slow modes of the environment evolution can be modeled with classical cellular automata (CA) approach. We have coarse-grained the fast modes approximating the transportation network by the graph of cellular automata (GCA). This allows the simulation of transportation systems over larger spatio-temporal scales and scrutinizing global interactions between the network and the environment. We show that the model can mimic the realistic evolution of complex river systems. We also demonstrate how the model can simulate a reverse situation. We conclude that the paradigm of this model can be extended further to a general framework, approximating many realistic multiscale transportation systems in diverse fields such as geology, biology and medicine.

2. Artificial Neural Networks Systems Based on DFT Parameters and Molecular Field Analysis - Computational Tools for Prediction of Ethylbenzene Dehydrogenase Reaction Kinetics

Maciej Szaleniec(1), Małgorzata Witko(1), Johann Heider(2)

(1) Institute of Catalysis and Surface Chemistry, Polish Academy of Sciences, Poland

(2) Institut für Biologie, Technische Universität Darmstadt, Germany

A mono-molybdenum enzyme, Ethylbenzene Dehydrogenase (EBDH), is a key biocatalyst of the anaerobic metabolism in denitrifying bacteria from *Azoarcus* sp. such as EbN1, EB1, PbN1. It catalyzes the oxygen-independent, stereo specific hydroxylation of ethylbenzene to (S)-1-phenylethanol. Our recent research shows that it exhibit significant activity with a wide range of ethylbenzene derivatives. For that reason EBDH promises potential applications in chemical and pharmaceutical industry [1-2]. Therefore, it is very important to find an easy and low cost method for selection of potential substrates. The main aim of our research is to build a theoretical model for screening of EBDH activity with new compounds without need for experimental tests and to get more insight into the mechanistic aspects of the reaction. Experimental measurements have resulted in the identification of 16 substrates and 8 inhibitors. The theoretical DFT-based parameters (partial charges, dipole moment, orbital energies, etc.) and simple topological descriptors are used as input parameters for artificial neural networks (ANNs). The system for classification and prediction (with regression approach) of the biological activity is constructed with the software package Statistica Neural Networks 6.0. For optimization of network architecture SNN Intelligent Problem Solver (IPS) is used with standard back-propagation (100 epochs) and conjunct gradients (1-100 epochs) learning algorithm. Approximately 3000 models were tested by IPS for classification and almost 4000 models for regression. The IPS-based experiments show that Multi-Layer Perceptions architecture with one hidden layer is the most appropriate for solving our problem. We managed to achieve 100% correct classification and very good regression network system (learning error= $8.14 \cdot 10^{-2}$, validation error= $2.4 \cdot 10^{-2}$, testing error= $6.29 \cdot 10^{-2}$) which is characterized by a high square correlation factor of predicted to experimental data ($R^2 = 0.9465$).

In order to determine a substituent influence on substrate reactivity, Molecular Field Analysis (MFA) approach [3] coupled with genetic partial least square algorithm (G/PLS) is applied. This yields the equation ($R^2=0.994$) that binds logarithm of k_{cat} and energies of substrate's interaction with two types of probe molecules: H^+ (for electrostatic interactions) and CH_3 (for steric interactions).

Finally, obtained models are used in laboratory practice for screening of new compounds that are to be used in experimental tests. This external validation method provides the comparison of generalization and extrapolation capabilities of MFA and ANN systems.

[1] O. Kniemeyer,, J. Heider, *J. Biol. Chem.* 276 (2001) 21381-21386.

[2] H.A. Johnson, D.A. Pelletier, A.M. Spormann, *J. Bacteriol.* 183 (2001) 4536-4542.

[3] R. D. III. Cramer, D. E. Patterson, J. D. Bunce, *J. Am. Chem. Soc.* 110 (1988) 5959-5967

3. Chemomentum - Grid Services Based Environment to Enable Innovative Research

Piotr Bala (1,2), Krzysztof Nowinski (1)

1) ICM Warsaw University, Warszawa, Poland

2) Faculty of Mathematics and Computer Science, N. Copernicus University, Torun, Poland

This contribution will present current status of the UNICORE/GS, as well as its extension in the area of computational chemistry.

The UNICORE/GS is a Grid Service infrastructure compliant with the Open Grid Services Architecture (OGSA).

is based on the UNICORE Grid software which becomes important tool to access HPC systems in European HPC Centers. UNICORE is already an integrated Grid solution supporting challenging scientific and industrial applications and complex work flows in production environments that span multiple independent organisations. One of the important enhancements available within UNICORE/GS are generic software components for visualisation and steering of simulations, remote device monitoring and control will be developed to broaden the applicability of Grid computing to new scientific and technological areas. UNICORE/GS allows also for seamless access to the distributed data and databases.

The Chemomentum is newly establish EU project (IST Call 5) which will provide Grid-based solutions for workflow-centric, complex applications, such as risk assessment, toxicity prediction and drug design, It focuses on tools for dealing with data and knowledge in an efficient and reliable manner. The Chemomentum software will enable users to assess the knowledge stored in the system in terms of applicability, range of validity and reliability, and will make it easy to use this knowledge for predictive and decision-making purposes. Leveraging developments in enterprise software systems, such as service oriented architectures and business integration, Chemomentum will achieve these aims by taking up and enhancing European Grid middleware. Grid-enabling relevant applications and building services on the Grid, Chemomentum will provide an easy-to-use Grid system that focuses on the end users, allowing them to use powerful tools in an efficient and transparent way. Intuitive, task-oriented user interfaces will allow the users to focus on their work, keeping any Grid related complexity hidden. Administration of the Grid, for example management of users, software license management and application deployment on the Grid will be greatly simplified.

4. Ensemble of Linear Classifiers for Creating Predictive Models

Tomasz Arodz, Witold Dzwiniel

Institute of Computer Science, AGH University of Science and Technology, Kraków, Poland

Pattern recognition techniques complement the computational methods in attempt to provide understanding of scientific phenomena. One approach to analyse patterns is by computational intelligence methods. Here, we show the results concerning the classification algorithms in applications stemming from biomedical research.

We focus on an ensemble scheme of supervised learning. In this approach, instead of training a sophisticated, monolithic predictive model, a set of simpler rules are trained and their decisions averaged. By using specially tailored classifier combination methods, the generalisation error of the model can be significantly reduced.

The possibility of increasing the accuracy by creating an ensemble depends on the base classifier used. For stable classifiers, such as linear ones, it has been argued that ensembles are not leading to lower error. This has been shown experimentally for classical ensemble creation techniques such as bagging, random subspace method or boosting.

We have created a scheme that allows for constructing accurate ensembles of linear classifiers. To achieve this, we destabilise the procedure for inferring linear decision boundaries. One technique that can be used to this aim is to employ different subsets of features describing the training samples. We have shown theoretically, that such an approach should provide better generalisation abilities than classical ensembles of linear rules. We have also supported this observation experimentally, using a wide set of benchmark problems. We have also shown, that the proposed ensemble scheme can be effectively extended to include a feature selection mechanism into the classifier.

The proposed ensemble framework is competitive in solving decision problems arising in biomedicine. In particular, we have evaluated the proposed classifier on several tasks in predicting the activity or other property of drug-candidate molecules. It allows for creating accurate structure-activity models for use in drug discovery.

The proposed method of feature selection within ensemble classifier has shown its usefulness in applications where the number of features is overwhelming. One such problem is found in diagnostic proteomics, an approach for cancer diagnosis from blood samples using mass spectrometry. The mass spectra may reach thousands of features for describing a single blood sample. On the other hand, the number of available samples is scarce. In such a non-trivial setting, our method has allowed for accurate discrimination between normal and malign samples, while at the same time using only a very minute fraction of all available features.

The proposed methods allow for creation of accurate classifiers for various types of complex patterns. Thus, our techniques can be useful in constructing predictive models for task, where the detailed knowledge on the mechanisms of action and governing processes are lacking or not fully understood.

5. Eve Effect - Variety of Genetic Families in Evolutionary Picture within the Penna Model

A.Z. Maksymowicz and M. Sitarz

Faculty of Physics and Applied Computer Science, AGH University of Science and Technology, Krakow, Poland

Results of simulations of population growth and evolution, using standard Penna model, with individuals

characterized by a string of bits representing a genome containing some possible mutations are presented. Penna model may be seen as important step beyond the simple logistic model when birth rate is balanced against the death toll within the Verhulst scenario, with death rate proportional to current population. This leads to mortality rate which is age-independent. Genetic death introduced by the Penna model, as caused by too many bad mutations, gives the observed exponentially increasing mortality with age. We apply the Penna model to observe the evolution of the population in terms of the genome structure.

In the model presented, individual's genome is represented as a string of bits, where '1' means presence of mutation, and '0' stands for lack of mutation. When individual reaches age a , only the first a bits are exposed. Number of "ones" in that part of genome are considered as active mutations m . When number of only active mutations exceeds threshold T , individual dies. After about 20000 simulation steps, when only a few genetic families are still present from among rich variety of families at the beginning of the simulation game, strong peaks in mutation distribution function are observed.

To distinguish between different genotypes -- we introduced a special marker of individuals. At the beginning of simulation -- each individual inserted into environment, has given unique number. When individual gives birth -- the marker is inherited by progeny. Marker is kept fixed which makes it possible to trace the evolution of different genetic families (by genetic family we mean set of individuals with the same marker). After long time, we usually get from two to five families in a sort of a net balance between these groups which share the same environment. Sometimes we get just a single family. This phenomena is known as "Eve effect". This effect is due to evolution rules with hereditary mechanism.

The birth and death balance in the simulation game also leads to elimination of families specified by different genomes. Number of families versus time follow a power law. Our results show the power coefficient exponent about (-2) for sexual version of the model, which is in contrast with the coefficient about (-1) as claimed in the literature. We proved that this discrepancy may be due to insufficient time scale in some simulations, when only at initial stages the (-1) coefficient is observed. However, for much larger number of simulation steps we get the (-2) power law. We also propose some modification of the model by incorporating the mechanism of birth rate controlled by the current population, instead of the standard Verhulst death factor in which the death rate is controlled by the population size. The effect is a significant plateau in power law for the number of families in short time scale, followed by the (-2) power law.

6. Event Record: Missing Link in the Architectures of Large HEP Collaboration Software Chains

Piotr Golonka
CERN IT/CO-BE

The complexity of High Energy Physics (HEP) experiments prepared nowadays, in particular the ones at the LHC, required the new approach to the organisation of data analysis chains: the place of FORTRAN-based modular code was taken by Object-Oriented/C++ component-based architectures. One of the key elements of these large software chains, event generators (simulations) is however still not well located in the component-based architectures: FORTRAN event generator codes are going to be used (at least for initial few years) for data analysis and processing in the LHC experiments. One of the causes of this situation is the heritage of the architecture for such modules, namely the use of "event record" data structure, used for data exchange between the modules, and lack of a widely adopted and flexible event record standard in C++. In the talk I will discuss the event-record based approach and the requirements that a viable proposal would need to fulfill. Such activity spans far below purely computer-science related data structure analysis, and needs to reach the areas of experimental and theoretical physics. In the examples, the off-line software of the ATLAS experiment will be used.

7. Finite Mesh Representation and Generation

Tomasz Jurczyk and Barbara Glut
Institute of Computer Science, AGH University of Science and Technology, Poland

The efficiency of the process of generation of unstructured anisotropic three-dimensional finite element meshes is considered in this work. The task of the automated mesh generator is to divide the geometrical domain into a finite number of simple elements (e.g. triangles or tetrahedra) conforming to requirements defined by the user with minimum interaction. The meshes are created using incremental Delaunay triangulation with additional procedures of mesh transformation or quality enhancement.

An important part of the generator software is the object oriented mesh representation with the associated set of elementary procedures of mesh accessing and modification. While designing this mesh structure many various (often competing) requirements have to be taken into account, influencing efficiency of meshing process and memory requirements. Introducing full set of adjacency connections between mesh entities (blocks, faces, edges and vertices) will be optimal for mesh accessing operations, but the memory requirements and cost of mesh modifications can be too costly. Similar alternatives have to be considered for various attributes of mesh entities which can be cached or calculated directly.

In most cases (and especially during the mesh adaptation process) an additional structure (control space) is required for storing the data supervising the size and shape (for anisotropic meshes) of the mesh elements throughout the discretized domain. In our approach this sizing data is represented in metric form and is stored in the discrete nodes of the control space structure which may have the form of a quadtree (octree) or a background mesh. The operations of creating and accessing this structure can have significant influence on the overall efficiency of the meshing process. Several methods of optimizing these operations and decreasing their use are inspected.

The process of mesh generation usually consists of a sequence of mesh modification (mostly local), which perform operations like inserting new nodes or replacing some elements with new ones. Depending on the phase of the meshing process the mesh size (number of mesh entities) can grow, shrink or remain roughly constant. Such behavior requires special memory management strategies with aggregate allocation and deallocation of entities, delayed removal of elements at selected phases of the meshing process and prediction of the final mesh size.

Generated meshes and results of practical tests are provided for illustration of these optimization problems for different geometrical models and meshing scenarios.

8. How to Reconstruct the Unknown Physical Quantities Using the Bayesian Approach and Neural Networks

Marcin Wolter

Institute of Nuclear Physics Polish Academy of Sciences, Poland

An application of neural networks to the reconstruction of unknown physical quantities in particle physics is described. As an example the simulated reconstruction of mass of the hypothetical Higgs boson in the typical high energy physics experiment is used.

In the approach described no explicit knowledge of the functional dependence of the unknown Higgs mass estimate on the measured quantities is needed, we rely on simulated data only. Monte Carlo events generated with various Higgs masses are used to determine the probability distribution for each Higgs boson mass. The probability is a function of measured quantities, in this example energies of two jets and the angle between them. The 3-dimensional probability distributions are fitted using neural networks. One of the advantages of a neural network fit is that data are not binned, which improves the quality of the fit while fitting small samples. Also no analytical formula of the fitted function is needed. The complexity of the function shape is determined by the number of links (free parameters) in the neural network.

For each event the trained neural networks return the set of probabilities corresponding to each Higgs mass. To obtain the mass estimate based on a sample of few events the probabilities should be multiplied, the final Higgs mass probability distribution is a product of the single event probability distributions. The mean of such a distribution is taken as a Higgs mass estimate. The miscalibration of the measured quantities is automatically corrected by the probability distributions.

This method is compared with the complementary method based on a neural network trained to return directly the Higgs mass. However, this method is giving highly biased results due to the limited width of the Higgs masses, for which the Monte Carlo samples were generated.

In the example presented here, where the Higgs boson mass is measured, the method gives a mass resolution similar to the one obtained using the standard invariant mass analysis, which leads to the conclusion, that no more information can be extracted from the jet energies and angle between them beyond that encoded in the invariant mass. By adding additional variables it may be possible to improve the power of the presented method.

9. Joint Inversion of Vertical Electrical Sounding and Dipole Induction Sounding with the Use of the Parallel Computing

Justyna Bała, Grzegorz Pszczoła

Department of Geoinformatics and Applied Computer Science, Faculty of Geology, Geophysics and Environmental Protection, AGH University of Science and Technology, Cracow, Poland

In the methods of geophysical set of field data interpretation one of the ways of better information analysis included in the measured data is joining of the interpretation of disparate data sets from the different geophysical methods. In case of geoelectrical methods the example could be the data of apparent resistivity received by the vertical electrical sounding and dipole induction sounding which respond differently to the examined earth structures.

The reason of the ambiguity of the interpretation in case of both methods are i.e. the measurement errors or equivalence phenomena (the equivalence of geological profiles). The use of the joint inversion for both methods can improve the measurement data analysis and geological interpretation of the given area.

The model which is frequently used in case of geological methods is horizontal layered model (so-called one-

dimensional model). One-dimensional (1D) interpretation of vertical electrical sounding and dipole induction sounding data sets is quite simple and fast tool for mapping the vertical variations in the resistivity of the earth's crust. In addition, 1D interpretation results are very useful in constructing initial models for multidimensional interpretations. Therefore a uniqueness study of 1D interpretation is important and joint inversion of vertical electrical sounding and dipole induction sounding gives the possibility of reduction of ambiguity of the 1D interpretation.

A number of interpretation problem solutions are derived using the global inversion technique to study the applicability of joint inversion in various situations. Before applying the joint inversion the separate interpretation were performed so that to see the comparison and estimation effectiveness. The further step is the correct forming the objective function for joint inversion and testing of different optimization methods. The choice of the objective function has the essential meaning for the time of computing and for effectiveness of optimization procedure. The application of local optimization methods not always give satisfactory effects and their effectiveness depends in the large scale from starting model. When the huge number of model parameters are applied, optimization procedure often is stopped by local minimum. The solution of this problem is the choice and use of the suitable global optimization method.

This methods (Monte Carlo group methods, simulated annealing, genetic algorithms etc.) required the huge computing power and the longer computing time. It appears that the use of the single processor is not sufficient enough. The complexity of computing results in too lengthy time of sequence solution of this problem. In addition to this the stores of the single computer (i.e. operating store) are insufficient in this case. One of the ways to overcome this kind of problem is the application of the parallel computing. One of the commonly seen features of performing global optimization algorithms are huge requirements connected with the vast number of computing and insufficient computing stores. But on the other hand, the bigger amount of this algorithms are characterized by natural parallelism and parallel realization may influence greatly on infallibility and effectiveness of algorithm:

10. Linux Clusters in Earth Sciences - Seismic and Geothermal Examples

Tomasz Danek, Anna Pieta

Faculty of Geology, Geophysics and Environmental Protection, Department of Geoinformatics and Applied Computer Science, AGH University of Science and Technology, Krakow, Poland

In the last few years inexpensive Linux clusters became very popular in Earth Sciences. Solutions based on them are used not only in science institutions or universities with limited budgets but also in commercial firms and big companies of oil and gas industry. This popularity is also visible in the amount of software created for this kind of parallel machines.

Both the hardware and software solutions are mostly focused on seismic method - the most important tool in oil exploration nowadays. There are many steps in seismic processing and interpretation procedure which need a lot of computational power but migration and full wave field modeling are two the most time consuming methods. Even now, on modern and reasonably big machines these procedures could last few days or even weeks. In this time scale even a small improvement in algorithm or in system hardware and/or software configuration can result in speed up of computations counted in days.

Our presentation covers some computer science backgrounds and geophysical aspects of seismic wave field modeling and seismic migration. We have analyzed many kinds of algorithms used in our own, free or commercial software. We have tried to answer a question how some popular improvements affect global efficiency of programs. Some hardware solutions, like memory, processors and network switches have been tested. We also have tested some Linux kernels in different configurations, especially in case of Symmetric Multi Processing and Intel Pentium 4 processors with Hyper Threading technology.

But of course Linux clusters in Earth Sciences are not only used for solving commercial oil and gas exploration problems. There are also many purely scientific ones which can be solved much faster with this powerful and inexpensive technology. One of these problems is geothermal modeling. In this case quite simple finite difference method algorithms can be used to solve thermal equation in explicit or implicit manner. But for large scale or global modeling amount of needed resources is enormous. We have tried some of the most popular solutions for both implicit and explicit algorithms, measured time of computations differences between them and choose one for large scale modeling of climate changing and glaciations impact on geothermal gradient. We also did some tests of program efficiency for different granularity of solutions.

All computations for both part of presentation have been performed using our departmental clusters, Geofizyka Krakow Company IBM cluster and some of the biggest European IBM clusters situated at CINECA Computer Center, Bologna during our participation in HPC - Europe program.

11. MC-TESTER: Validation and Testing Tool for HEP Monte Carlo Simulations

*Piotr Golonka
CERN IT/CO-BE*

Monte Carlo simulation codes have been, traditionally, developed in FORTRAN. In the coming era of the LHC experiments, and future Linear Collider experiments, more and more interest is expressed in translation of the existing simulations to C++. To preserve the physical quality and validity of the generators (which was often taken over decades), a tool capable of performing semi-automatic comparison tests would be of value. MC-TESTER was initially created with the purpose of being such auxiliary tool for the TAUOLA Monte Carlo. Its idea is to gather some characteristic distributions of the data produced in the run of the MC generator, then compare the distribution from two runs, and determine (in a visual and quantitative way) whether the distributions are statistically different; as a product, an easy to browse "booklet" with results is produced - it may be used to detect the inconsistency "in a glimpse of an eye" (over the pictures of distributions). The tool was applied recently in the analysis of another Monte Carlo generator, PHOTOS - due to its flexibility important tests were performed, which (in effect of accompanying study and development of the model) led to significant improvement in physical precision tag for PHOTOS... This example proves that the area of applicability of this tool may be much wider than originally expected.

12. New Multidimensional Scaling Algorithm Based on Dissipative Particle Dynamics

*Marcin Kurdziel, Krzysztof Boryczko
Department of Computer Science, University of Science and Technology, Krakow, Poland*

We present a new algorithm for visualization of multidimensional and proximity-based datasets. The algorithm is based on a novel approach to Multidimensional Scaling [1] method, that aims at mapping a given dataset into 2D/3D space in a way preserving inter-point distances. It can be used for revealing dataset's cluster structure, identification of outlier points or estimation of variations in data density. Furthermore, as the method relies only on information on distances (or proximity measures) between data points, it is suited for visualization of sets of objects which do not have well defined dimensional structure. This includes objects such as character strings (e.g. biological sequences), geometrical entities (e.g. protein structures), and other.

Construction of nonlinear mapping in Multidimensional Scaling is performed by minimization of appropriately defined stress function, such as Sammon or Kruskal criterion. However, even for a limited number of data points stress functions are heavily multi-modal, which is a major obstacle in identification of global minimum. Therefore, we propose to apply molecular dynamics paradigm to construction of optimal (or close to optimal) mappings. In particular, we apply Dissipative Particle Dynamics (DPD) [2] to a set of particles representing data points, with conservative forces derived from the gradient of stress function. By virtue of explicit temperature control provided by DPD it is possible to apply a cooling, with a chosen schedule, to the simulated system. When the temperature reaches 0, the particles are frozen in a configuration which corresponds to at least local minimum of the stress function. However, if the cooling schedule is sufficiently slow, an optimal or next to optimal configuration will usually be found. The strength of this approach was validated on a synthetic and real-world datasets that do not have perfect mapping in 3D, i.e. all inter-point distances cannot be simultaneously preserved in 3D. To enable visualization of large datasets, we investigate methods for reducing computational and memory complexity of the stress function estimation. More specifically, we propose to select a subset of the original $O(n^2)$ inter-point interactions for inclusion in the DPD simulation. Two approaches to subset selection were tested, namely inclusion of interactions between k -nearest and k -furthest neighbors of each data point and random selection of interactions driven by histogram of inter-point distances. Experimental results suggest that up to an order of magnitude reduction in computation time can be achieved with insignificant distortions of resultant mapping. For datasets possessing perfect embedding in 3D, larger reduction of computational cost is possible. To further boost the mapping performance, we have developed a parallel implementation of the proposed method. The implementation was done for shared memory multiprocessor systems, using the OpenMP standard. Experimental results on the SGI Altix 3700Bx2 system revealed good scaling for up to 48 processors, on a dataset consisting of approximately 93,000 points.

[1] Cox T.F., Cox A.M., Multidimensional Scaling 2nd ed., Chapman & Hall, 2000

[2] Hoogerbrugge P.J., Koelman J.M., Simulating microscopic hydrodynamic phenomena with dissipative particle dynamics, *Europhys. Lett.*, 19/155, 1992

13. Parallel Multigrid Solver for Unstructured h-Adaptive Finite Element Computations

*Krzysztof Banas
Cracow University of Technology*

The presentation concerns the design and implementation of a parallel iterative solver for systems of linear equations arising in finite element computations. As main solution methods Krylov subspace algorithms, such

CG or GMRES, are used. Multigrid is employed for preconditioning system matrices, using additive Schwarz method combined with local block Gauss-Seidel or ILU(0) iterations as smoothers.

The main algorithmic ideas of the proposed approach consist in using geometric and approximation properties of finite elements to construct prolongation and restriction operators as well as proper blocks for smoothing.

The presented parallel implementation is based on special data structures for message passing to support high latency networks. The choice of overlap in additive Schwarz preconditioning is also designed to minimize communication between processors.

The solver's architecture follows several principles that allow for interoperability of the solver with different types of finite element meshes and approximations. The solver is designed as a separate module and exchanges data with the finite element code using special data structures.

The presentation will include results of parallel solution of systems of equations with several millions of unknowns on a cluster of PCs.

14. Routing in Circulant Graphs

Tomaz Dobravec and Borut Robic

Laboratory for Algorithms and Data Structures, Faculty for Computer and Information Science, University of Ljubljana

In the past few years we have focused our attention on routing problems in special kind of network topologies. Mainly we have studied the so called circulant graphs and their routing capabilities. Roughly speaking we have concentrated on the following topics:

- A) Design of dynamic routing algorithm
- B) Routing simulations
- C) Circulant graphs as a mathematical structure

We designed an optimal dynamic routing algorithm which routes the data using one of the shortest paths connecting the source and destination nodes. The main advantage of our algorithm is in its flexibility, i.e. in ability to switch among different optimal paths according to the link availability. This feature allows our algorithm to dynamically solve or even avoid potential routing problems that may arise from node/link faults (or from congestions if several messages were routed through the network simultaneously).

We designed a special simulator to compare permutation routing algorithms. Using this simulator we have evaluated several properties of static and dynamic routing algorithms. Using the numerical results of simulations we proposed an optimal routing strategy.

Circulant graphs are very interesting mathematical structure with several useful properties. When solving problems such as shortest path or diameter problem in circulant graphs, one can use the integer lattice as a very useful tool. The mentioned problems are then transformed into graphically well presented problems and into mathematical problems such as solving the Diophantine equations with several extra conditions. During our research we have solved the problem of restricted shortest paths and find and prove several properties of packed basis for integer lattice.

15. Semantic Services Grid in Flood-Forecasting Simulations

Babik, M., Habala, O., Hluchy, L., Laclavik, M. and Maliska, M.

Dept. of Parallel and Distributed Computing, Institute of Informatics, Slovak Academy of Sciences, Bratislava Slovakia

Recently, Web service (WS) technologies are gaining importance in the implementation of distributed systems, especially grids. One such example is the Web Service Resource Framework (WSRF), which extends the current WS technologies by modeling the stateful services. Design and development of the service oriented distributed system is quite common and there are several emerging WS initiatives, which tries to automate the process of discovery, composition and invocation of services. The semantic web services are a typical example, showing the potential of how ontological modeling can improve the shortcomings of service oriented computing.

In this paper we will present design and development of the semantic grid services for the flood forecasting simulations. We will highlight the corresponding architecture and the process of service annotation, discovery and composition in the project K-WfGrid. We will describe in detail the challenges of the flood-forecasting application and corresponding design and development of the service oriented model, which is based on the we

known Web Service Resource Framework (WSRF). Semantic descriptions of the WSRF services will be presented as well as the architecture, which exploits semantics in discovery and composition of services. Further, we will demonstrate how experience management solutions can help in the process of discovery and composition. The system provides a unique approach in Semantic Grids by combining the advances of semantic web services and grid architectures.

We will also discuss the semantics of the proposed grid services and provide an overview of the service annotation techniques. Service annotation in this context is a process of generating the semantic descriptions (i.e.OWL-S) of both stateless and stateful services from the web service descriptions. In K-Wf Grid it has become crucial in the process of providing application support and enabling semantics for semantically unaware grid application areas. We will present the issues that we have faced during the design of the ontologies for the stateful services and we will briefly describe the developed annotation tool called WSRF2OWL-S. We will conclude our presentation with a sample flood forecasting use case scenario.

16. Simulation of Hostility and Friendship in a Social Scale

P. Gronek, P. Gawronski and K. Kulakowski

Faculty of Physics and Applied Computer Science, AGH-UST, Cracow, Poland

Recently we have proposed a set of differential equations to describe the time evolution of social relations within a closed community [1-3]. These relations have been reduced to friendly (+1) or hostile (-1) in previous literature [4]. We generalized the model to include a continuous scale of relations. Initial distribution of the relations is assumed to be incoherently random. The time evolution leads the community to two mutually hostile groups. The relations within the groups become exclusively friendly. This final state is called the Heider balance [5]. The driving force of the evolution is a tendency to remove a cognitive dissonance, which is inherent for the human nature. The model has been confirmed by the comparison with sociological data, collected by Zachary from a karate club [3,6]. As a result of a conflict within the club, it was divided into two mutually conflicted groups. The model formulation, proposed by us, allowed to assign the players to two exactly as it happened in reality [3]. Here we report numerical calculations on the character of the social transformation in time. The matrix of relation is assumed to be symmetric; this is justified by the supposition, that mutual relations between two people evolve much faster, than the community as a whole. In the simplest version of the model, the time evolution of the relation $r(i,j)$ between people i and j is given by

$$\frac{dr(i,j)}{dt} = \sum_k^N (r(i,k)r(k,j))$$

where N is the community size. Calculations are performed for N up to 800, what is equivalent to a numerical solution of $800 \times (800-1)/2 = 319600$ differential equations simultaneously.

It appears that the time of the process (from a random initial state to the Heider balance) decreases with the community size N as $N^{-1/2}$. This result, although easy to interpret mathematically [2], is of interest from the sociological point of view, because it means that conflicts develop faster in a larger community.

[1] K. Kulakowski, P. Gawronski and P. Gronek, *Int. J. Mod. Phys. C* 16 (2005) 707.

[2] P. Gawronski, P. Gronek and K. Kulakowski, *Acta Phys. Pol. B* 36 (2005) 2549.

[3] P. Gawronski and K. Kulakowski, *AIP Conf. Proc.* 779, Melville, NY 2005, p. 93 (physics/0503085).

[4] Z. Wang and W. Thorngate, *J. Artificial Societies and Social Simulation*, Vol. 6, No 3 (2003)

(<http://jasss.soc.surrey.ac.uk/6/3/2.html>).

[5] F. Heider, *The Psychology of Interpersonal Relations*, J. Wiley and Sons, NY 1958.

[6] W. W. Zachary, *J. Anthropol. Res.* 33 (1977) 452.

17. Simulation of the Mad Cow Disease Propagation within Cellular Automata Technique

M.S. Magdon-Maksymowicz,

Department of Mathematical Statistics, Agricultural University, Krakow, Poland

Computer simulation of BSE disease propagation, also known as mad cow disease, is discussed. Both vertical (from parents to baby) and horizontal (between neighbors) mechanisms of the disease spread are accounted for. The game takes place on a two-dimensional square lattice of 1000×1000 sites. Each site may be empty or occupied by a cow (or group of cows). The initial population is randomly distributed on the net. The disease may be introduced with the initial population, by a spontaneous development of BSE in an item due to a genetic misprint at a small frequency, either to a newly born baby or among mature individuals. BSE is a very serious and lethal disease, of a very long incubation period, followed by a brief fatal clinical phase. We describe it by a parameter which is zero for not infected items, then in each infected individual this parameter increases with every simulation step. We may put up distinctly some specific critical values such as:

(a) value below which the mild disease is not passed over, or

(b) when it is still unrecognized yet infectious, or

(c) in severe stage when the items are perhaps eliminated, then

(d) threshold value at which death caused by the disease must take place.

The speed of how the disease spreads depends on the probability of the vertical and/or horizontal propagation mechanisms. Obviously it is also vulnerable for local density of the population, the uniform distribution or with tendency for grouping. We did not consider migration explicitly, yet we applied two different scenarios when offsprings are randomly distributed or when they are kept in vicinity of parents. The set of parameters used for the simulation corresponds to the disease among cows. Main results show there is a critical probability of the BSE transmission above which the disease is present in the population - below this critical value the population shows self-recovery after few generations and only sporadic cases may be reported as result of spontaneous outburst of the BSE due to the genetic fault. This value is vulnerable to possible spatial clustering of the population and it also depends on detailed mechanisms responsible for the disease onset, evolution and propagation. A threshold birth rate below which the population is extinct is also seen. Above this threshold, the population is disease free at equilibrium until another birth rate value is reached when the disease is present in population. Actually, for typical model parameters used for the simulation, which correspond to the mad cow disease, we are close to the BSE-free case.

The disease transmission makes agglomerations more open for easy disease distribution. This, however, is not necessarily equivalent to overall higher rate of infected items in the whole population. The dynamics of evolution may give the opposite result if only the infected clusters would die and give room to healthy individuals in the limited environment. The rate of birth plays an important role in the population structure. For the small rate the population vanishes since number of deaths for different reasons is then above the reproductive limit. This (and other) conclusions are not firm, they are rather indicative than decisive. This is so for two reasons. Firstly, the set of model parameters cannot be well established for lack of well grounded experimental data - details of the BSE spread is an open question. Secondly, the dynamics of the system evolution is sensitive to the choice of the model parameters such as the lower limit (about 30 months) for the disease duration before it becomes infectious and may be transmitted. We believe, however, the the overall picture is correct and may represent essential features of the BSE spread.

18. Solid-on-solid Models of Films Growth

P. Kowalczyk, R. Kosturek and K.Malarz

AGH University of Science and Technology, Krakow, Poland

In this lecture several simple models of epitaxial films growth based on cellular automata technique are considered. The results of computer simulation of the film surface roughness dynamics are presented. Simulations of the atomic-scale microscopy of the growth process are included, as well.