



**European Centre
for Soft Computing**

Research Unit

**Intelligent Data Analysis
and Graphical Models**

Intelligent Data Analysis and Graphical Models

- Principal Researcher
Christian Borgelt (Germany)
- Associate Researcher
Wolfgang Trutschnig (Austria)
- Postdoctoral Researchers
Marc Segond (France)
David Picado Muño (Spain, from April 1, 2011)
- Predoctoral Researchers
Ana-Belén Ramos Guajardo (Spain)
Anett Hoppe (Germany, until February 28, 2011)

- **Frequent Pattern Mining**
(item sets, sequences, subgraphs, molecular fragments etc.)
- **Neurobiology: Analysis of Parallel Spike Trains**
- **Knowledge Representation and Discovery with Graphs**
- **Learning Graphical Models from Data**
- **Automatic Construction of Graphical Models**
- **Statistics with Fuzzy Data** (esp. hypothesis testing)
- **Fuzzy Regression Models**



- **Analysis of Molecular Databases**

Example: NCI DTP HIV Antiviral Screen Data Set

- **Objective**

Structure-Activity Relationship Discovery/Modeling

Identify (at least hints) what substructures/fragments are responsible for the activity of a compound.

- **General Approach**

Frequent Subgraph Mining:

Find discriminative molecular fragments that are

- frequent in the active compounds
- rare in the inactive compounds



- **MoSS — Molecular Substructure Miner**
 - Java program for Frequent Subgraph Mining (free software, Gnu Lesser (Library) public license)
 - Strong support for molecular fragment mining: works with SMILES, SLN, SDfile/Ctab etc.
 - Special feature: ring mining (treat rings as units instead of sets of bonds)
- **Applications/Users**
 - Konstanz Information Miner (plug in)
 - Pharmaceutical Companies: Nycomed, Pfizer
 - Biochemical Research (public research institutions)



- **Current/Ongoing Work**

- Exploration of alternative canonical forms
- Exploration of alternative extension strategies
- Incorporation of configuration/stereometry

- **Challenges/Future Work**

- Extension of ring mining to other building blocks (idea/guide: Wiswesser line notation)
- Consider activity levels instead of categories
- Take conformation into account (very challenging task!)

- **BISON** (EU FP7 FET Open Project)
Bisociation Networks for Creative Information Discovery
Explore bisociative discovery on the basis of graph-based data mining and develop a conceptual and theoretical framework.
(8 partners, 6 countries)
- **COST Action IC0702**
Combining Soft Computing Techniques and Statistical Methods to Improve Data Analysis Solutions
(22 countries, chair: Christian Borgelt, grant holder: ECSC)
- **Alimerka**
Sales Prediction, especial w.r.t. the effect of special offers.



Main Current Collaborations

- **Michael Berthold**
Nycomed Chair for Bioinformatics, University of Konstanz
- **Sonja Grün**
RIKEN Brain Science Institute, Tokyo
- **Alberto Pascual-Montano**
Functional Bioinformatics Group, CSIC, Madrid
- **Gyözö Gidófalvi**
KTH Royal Institute of Technology, Stockholm
- **Andreas Nürnberger**
Data & Knowledge Engineering, University of Magdeburg
- **Rudolf Kruse**
Computational Intelligence, University of Magdeburg

Web: <http://www.softcomputing.es/>
<http://www.borgelt.net/>

Email: christian.borgelt@softcomputing.es
christian@borgelt.net

Phone: +34 985 45 65 45

Fax: +34 985 45 66 99

Address: **Intelligent Data Analysis and Graphical Models**
European Centre for Soft Computing
Edificio Científico-Tecnológico
c/ Gonzalo Gutiérrez Quirós s/n
33600 Mieres, Asturias, Spain