

Data Mining: Theory and Applications

Kai Puolamäki
28 October 2011

Current people

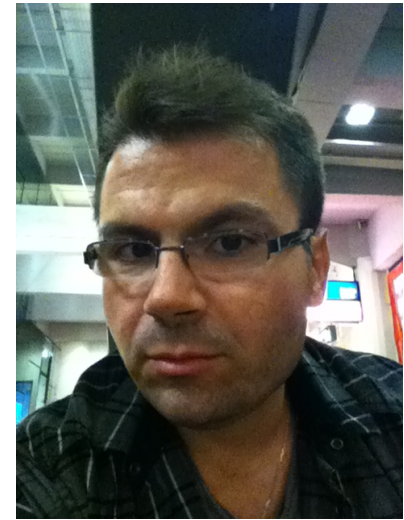
Heikki Mannila, group leader emeritus



Kai Puolamäki, acting group leader



Panagiotis Papapetrou, postdoc



Jefrey Lijffijt, doctoral student



Aleksi Kallio, part-time doctoral student



Effective staff 2+x persons

Other people

Recent Alumni:

Markus Ojala, doctoral student (-1/2011, defense 12 Nov 2011)

Sami Hanhijärvi, doctoral student (-2010, defense soon)

Niko Vuokko, doctoral student (-4/2011, defense soon)

Adjunct alumni:

Esa Junttila, UH doctoral student (-2/2011, defense Aug 2011)

Prolific co-authors in Algodan:

Petteri Kaski and Jaakko Hollmén.

Prolific co-authors in Algodan or FDK alumni:

Gemma Garriga, Aristides Gionis, and Evimaria Terzi.

Group's publications in 2010 (1/5)

J. T. Eronen, K. **Puolamäki**, L. Liu, K. Lintulaakso, J. Damuth, C. Janis, and M. Fortelius. Precipitation and large herbivorous mammals, part I: Estimates from present-day communities. **Evolutionary Ecology Research**, 12(2):217–233, 2010.

J. T. Eronen, K. **Puolamäki**, L. Liu, K. Lintulaakso, J. Damuth, C. Janis, and M. Fortelius. Precipitation and large herbivorous mammals, part II: Application to fossil data. **Evolutionary Ecology Research**, 12(2):235–248, 2010.

Markus Ojala and Gemma *Garriga*. Permutation Tests for Studying Classifier Performance. **Journal of Machine Learning Research**, 11:1833-1863, 2010.

T. Vesala et al. with Petteri Kaski and Heikki **Mannila**. Autumn temperature and carbon balance of a boreal Scots pine forest in Southern Finland. **Biogeosciences**, 7(1): 163, 2010.

Theodoros Lappas, Evimaria *Terzi*, Dimitrios Gunopulos, and Heikki **Mannila**. Finding effectors in social networks. In Proc 16th ACM SIGKDD, 1059-1068, 2010.

Niko Vuokko and Petteri Kaski. Testing the Significance of Patterns in Data with Cluster Structure. In Proc 10th IEEE ICDM, 1097-1102, 2010.

Group's publications in 2010 (2/5)

Markus Ojala. Assessing Data Mining Results on Matrices with Randomization. In Proc 10th IEEE ICDM, 959-964, 2010.

Niko Vuokko. Consecutive Ones Property and Spectral Ordering. In Proc 10th SDM, 350-360, 2010.

Niko Vuokko and Evimaria *Terzi*. Reconstructing randomized graphs. In Proc 10th SDM, 49-59, 2010.

Markus Ojala, Gemma *Garriga*, Aristides *Gionis*, and Heikki **Mannila**. Evaluating Query Result Significance in Databases via Randomizations. In Proc 10th SDM, 906-917, 2010.

Panu Luosto, Jyrki Kivinen, and Heikki **Mannila**. Gaussian Clusters and Noise: An Approach Based on the Minimum Description Length Principle. In Proc DS, LCNS 6332/2010, 251-265, 2010.

Plus a book chapter and peer-reviewed workshop papers.

Group's publications in 2011 (3/5)

Aleksi **Kallio**, Kai **Puolamäki**, Mikael Fortelius, and Heikki **Mannila**. Correlations and co-occurrences of taxa: the role of temporal, geographic and taxonomic restrictions. **Palaeontologia Electronica**, 14(1), 2011.

Antti Ajanki, Mark Billingham, Hannes Gamper, Toni Järvenpää, Melih Kandemir, Samuel Kaski, Markus Koskela, Mikko Kurimo, Jorma Laaksonen, Kai **Puolamäki**, Teemu Ruokolainen, and Timo Tossavainen. An augmented reality interface to contextual information. **The International Journal of Virtual Reality**, 15:161–173, 2011.

Jussi T. Eronen, Kai **Puolamäki**, Hannes *Heikinheimo*, Heikki Lokki, Ari Venäläinen, Heikki **Mannila**, and Mikael Fortelius. The effect of scale, climate and environment on species richness and spatial distribution of Finnish birds. **Annales Zoologici Fennici**, 48(5):257–274, 2011.

Panagiotis **Papapetrou**, Vassilis Athitsos, Michalis Potamias, George Kollios, and Dimitrios Gunopulos. Embedding-based Subsequence Matching in Time Series Databases. In **ACM TODS**, 36(3): 17, 2011.

Group's publications in 2011 (4/5)

Panagiotis Papapetrou, Gary Benson, and George Kollios. Mining Poly-regions in DNA Sequences. **International Journal of Data Mining and Bioinformatics**, 2011 (to appear).

Alexi Kallio, Niko Vuokko, Markus Ojala, Niina Haiminen, and Heikki Mannila. Randomization Techniques for Assessing the Significance of Gene Periodicity Results. **BMC Bioinformatics**, 12:330, 2011.

Gemma Garriga, Esa Junttila, and Heikki Mannila. Banded structure in binary matrices. **Knowledge and Information Systems**, 28(1): 197-226, 2011.

Terttu Nevalainen, Helena Raumolin-Brunberga, and Heikki Mannila. The diffusion of language change in real time: Progressive and conservative individuals and the time depth of change. **Language Variation and Change**, 23: 1-43, 2011.

Alexios Kotsifakos, Panagiotis Papapetrou, Jaakko Hollmén, and Dimitrios Gunopulos. A Subsequence Matching with Gaps-Range-Tolerances Framework: A Query-By-Humming Application. In Proc **PVLDB**, 4(11): 761-771, 2011.

Group's publications in 2011 (5/5)

Jefrey Lijffijt, Panagiotis Papapetrou, Kai Puolamäki, and Heikki Mannila. Analyzing word frequencies in large text corpora using inter-arrival times and bootstrapping. In Proc ECML PKDD, 2011.

Panagiotis Papapetrou, Aristides Gionis, and Heikki Mannila. A Shapley-value Approach for Influence Attribution. In Proc ECML PKDD, 549-564, 2011.

Orestis Kostakis, Panagiotis Papapetrou, and Jaakko Hollmén. Artemis: Assessing the Similarity of Event-Interval Sequences. In Proc ECML PKDD, 229-244, 2011.

Niko Vuokko and Petteri Kaski. Significance of Patterns in Time Series Collections. In Proc 11th SDM, 676-686, 2011.

Esa Junttila and Petteri Kaski. Segmented nestedness in binary data. In Proc 11th SDM, 235-246, 2011.

Sami Hanhijärvi. Multiple Hypothesis Testing in Pattern Discovery. In Proc DS, LCNS 6926/2011, 122-134, 2011.

Plus peer-reviewed workshop papers.



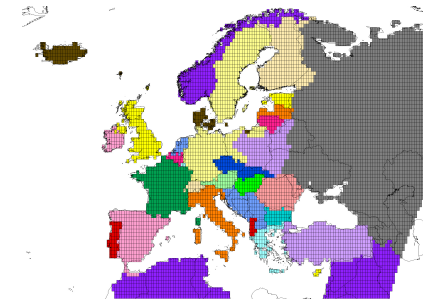
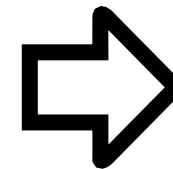
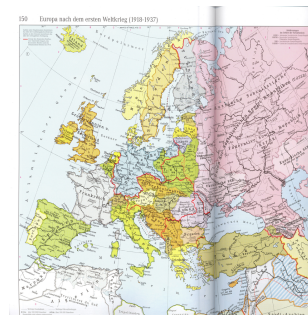
Aalto University
School of Science

0-1 structures in ecology

Computational methods to do good (natural) science (ecology, paleontology, biology).

Data:

0-1 occurrences of species (mammals, plants, birds) in map grid or point locations (fossil sites) with auxiliary information (Europe on 50x50 km grids) such as
climate,
soil,
distribution of genotypes,
historic political maps,
population,
topography etc.



Collaboration (induced by Rajapaja)

Heikki Mannila

Kai Puolamäki

Emilia Oikarinen

Pauli Miettinen

(Hannes Heikinheimo)

(Antti Ukkonen)

Greger Lindén

Mikael Fortelius

Jussi Eronen

Marianne Fred

Tuuli Toivonen

Maria Salonen

Markku Löytönen

Antti Sajantila

Jukka Palo

Antti Leino

Raino Lampinen

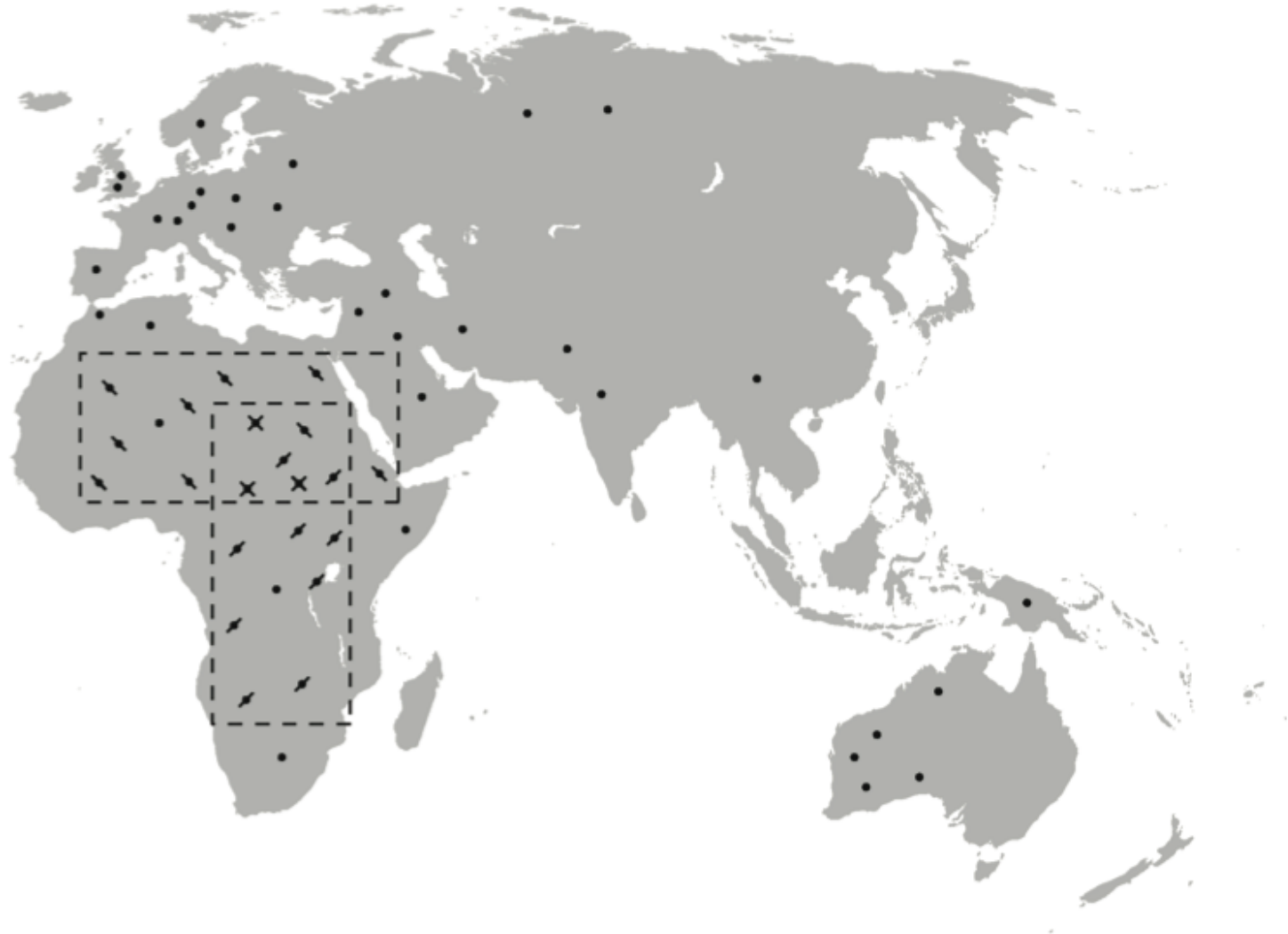
Arto Kurtto

Lena and Larry Huldén

Max Engman

Correlations and co-occurrences

KALLIO ET AL.: CORRELATIONS AND CO-OCCURENCES



Correlations and co-occurrences (cont.)

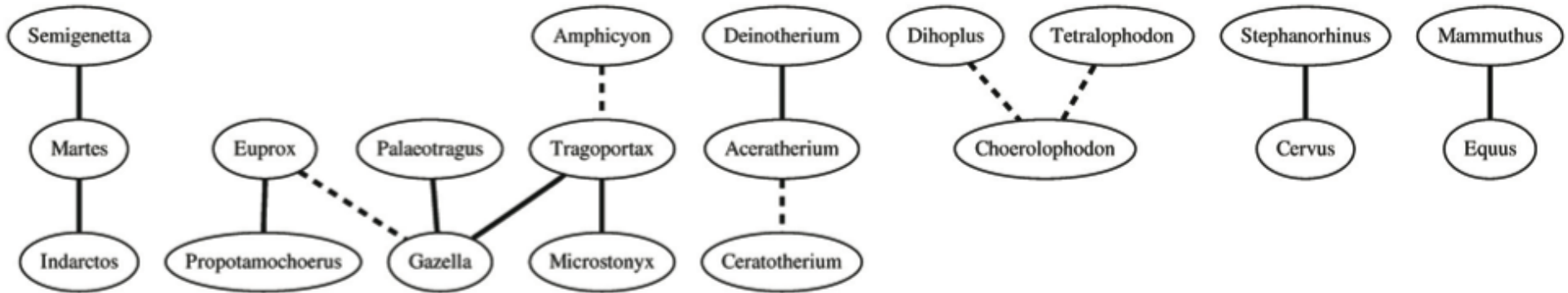
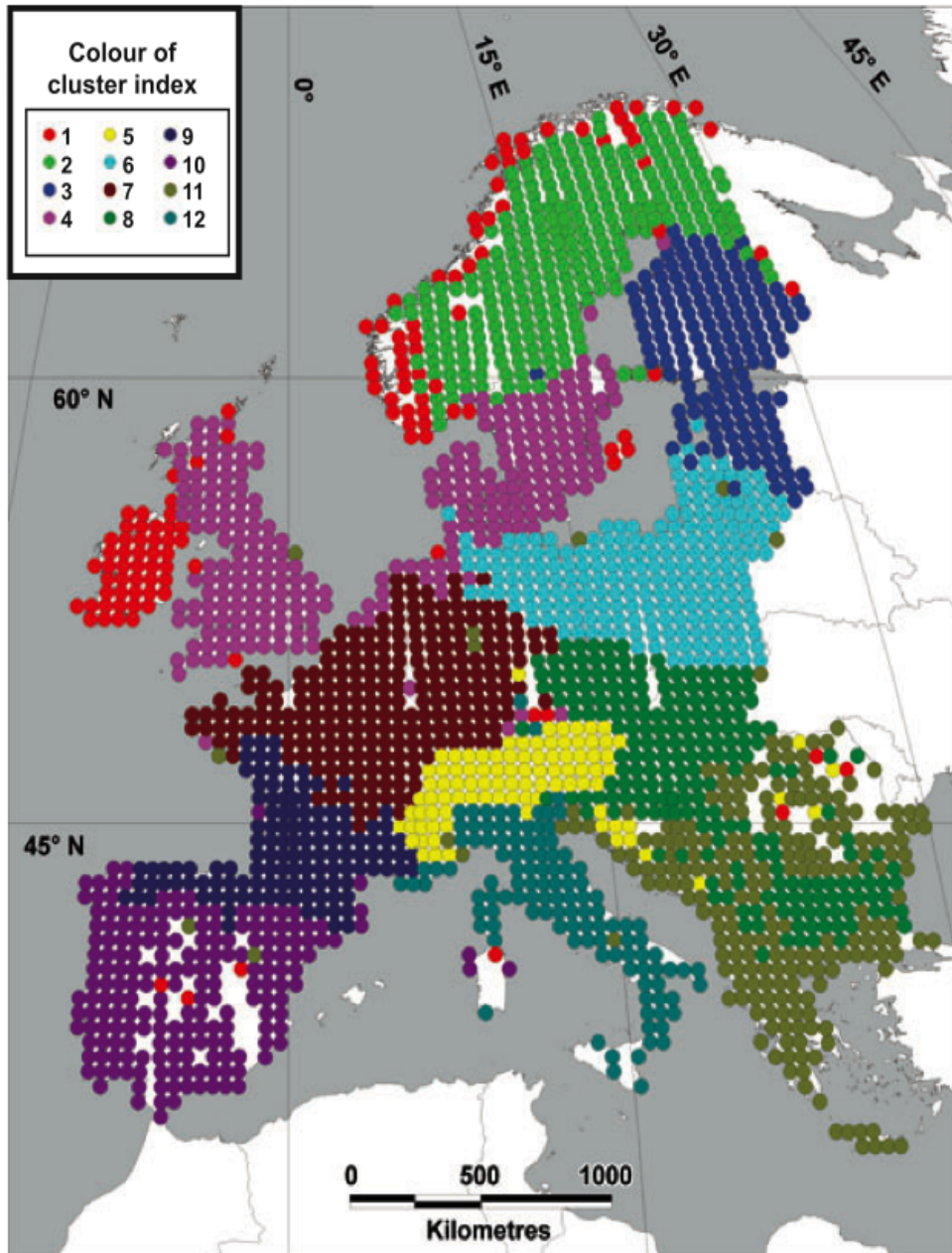


FIGURE 2. Correlations between genera with geographic filtering (**GeoUnion**). Genera are presented as ellipses, statistically significant positive correlations as bold lines, and negative correlations as dashed lines.

Alexi Kallio, Kai Puolamäki, Mikael Fortelius, and Heikki Mannila. Correlations and co-occurrences of taxa: the role of temporal, geographic and taxonomic restrictions. *Palaeontologia Electronica*, 14(1), 2011.

Spatial patterns



Hannes Heikinheimo, Mikael Fortelius, Jussi Eronen, and Heikki Mannila. Biogeography of European land mammals shows environmentally distinct and spatially coherent clusters. *Journal of Biogeography*, 34(6): 1053-1064, 2007.

Spatial patterns - ongoing work



Aalto University
School of Science