Probabilistic Adaptive Systems

- *Research groups*
  - Complex Systems Computation Group (CoSCo): Prof. Petri Myllymäki, HIIT Fellow Jorma Rissanen, Dr. Teemu Roos, Dr. Wray Buntine, Prof. Henry Tirri
  - Neuroinformatics: Dr. Aapo Hyvärinen, Dr. Patrik Hoyer, Dr. Jarmo Hurri
  - Statistical Machine Learning and Bioinformatics: Prof. Samuel Kaski, Dr. Janne Nikkilä, Dr. Jaakko Peltonen, Dr. Janne Sinkkonen, Dr. Jarkko Venna

- *Programme Management*
  - Programme Director: Professor Petri Myllymäki
  - Programme Manager: Research Coordinator Tomi Silander
  - Programme Management Group: Prof. Petri Myllymäki, Dr. Aapo Hyvärinen, Prof. Samuel Kaski

PAS Mission

- Computer science is the science of studying how things can be automated. When automating intelligent behaviour, **modeling** plays a central role as an attempt to formalize the properties of processes characterizing learning, inference and intervention (actions).

- Due to the uncertainty and incompleteness of available information in application domains of computer science - the artificial intelligence and machine learning domains in particular - such models are commonly based on **probabilities**.

- The aim of our research is fundamental understanding and development of computationally efficient probabilistic and information- theoretic modeling techniques, and their multi-disciplinary applications from engineering to sciences.

Key Research Lines in 2009

- Minimum Description Length (MDL)
  - Novel theoretical developments
    - Example: J.Rissanen, and T.Roos. Conditional NML Universal Model.
    - Example: P. Kontkanen and P. Myllymäki, A linear-time algorithm for computing the multinomial stochastic complexity.
  - New applications of MDL
    - Example (histogram density estimation): P. Kontkanen and P. Myllymäki, MDL Histogram Density Estimation.
    - Example (clustering): P. Kontkanen, P. Myllymäki, W. Buntine, J. Rissanen, H. Tirri, An MDL Framework for Data Clustering.

...More Key Research Lines in 2009

- Probabilistic graphical models
  - Bayesian network learning
    - Example: T. Silander and P. Myllymäki, A Simple Approach for Finding the Globally Optimal Bayesian Network Structure.
  - Novel dependency models
    - Example: W. Buntine, S. Perttu, V. Tuulos, Using Discrete PCA on Web Pages.
  - New estimation principles for probabilistic models.
    - Example: A. Hyvärinen, Estimation of non-normalized statistical modes by score matching.
  - Causal inference
    - Example: S. Shimizu, P. O. Hoyer, A. Hyvärinen, and A. J. Kerminen, A linear non-gaussian acyclic model for causal discovery

...More Key Research Lines in 2009

- Information visualization
  - Example: J. Venna and S. Kaski. Nonlinear Dimensionality Reduction as Information Retrieval

- Sensor fusion: combination of heterogeneous data sources
  - Example: A. Klami and S. Kaski, Local Dependent Components.
  - Example: combining eye tracking with content-based document search
  - Example: user modelling with location tracking and purchasing history data
  - Example: stemmatological modelling with constraints imposed by domain expert knowledge
  - Example: combining video and text search
  - ...and many more

Key actions

- *Intensive research work*

- Active international collaboration with top researchers in the field, in particular with the HIIT Fellows and their sites

- Active collaboration with other sciences, public agencies and the industry

- Active collaboration within HIIT (both within the PAS programme, and with the other programmes)

- Active role in the Pascal Network of Excellence of EU's FP7 programme: Petri Myllymäki and Sami Kaski in the steering committee

Proactive information retrieval

and

Eye movements as a feedback source
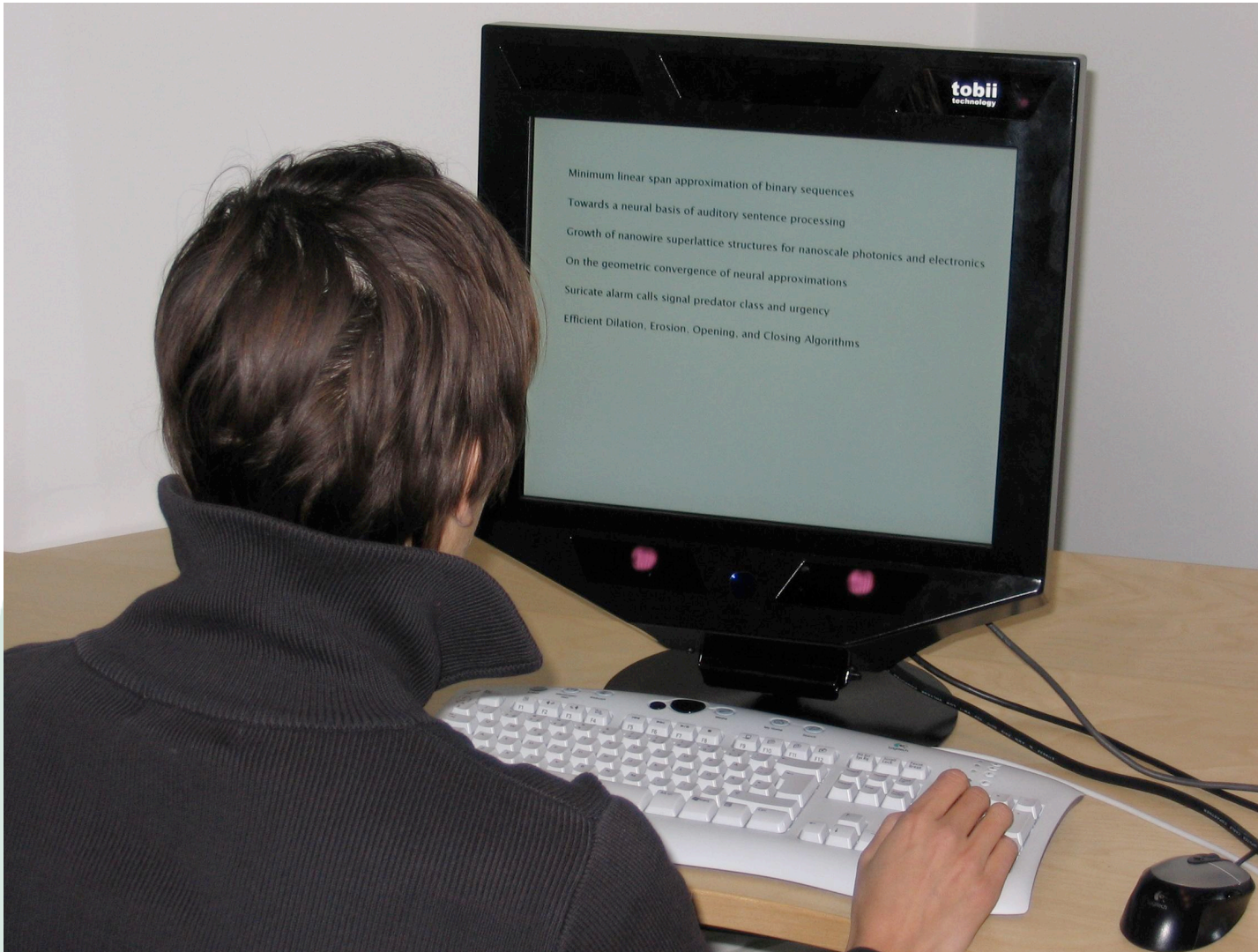
Eyeogle

The Minimum Error Minimax Probability Machine

by Kaizhu Huang, Haiqin Yang, Irwin King, Michael R. Lyu, Laiwan Chan
Journal of Machine Learning Research Vol. 5, pp. 1253–1286, 2004

http://jmlr.csail.mit.edu/papers/v5/huang04a.html   – Cached – Similar pages

Sphere–Packing Bounds for Convolutional Codes

by E.Rosnes and O.Ytrehus
 IEEE Transactions on Information Theory Vol.50(11), pp. 2801–2809, 2004.

ccc.ustc.edu.cn/abstract/rosnes.ps   – Cached – Similar pages

Quantum State Transfer Between Matter and Light

by D. N. Matsukevich and A. Kuzmich
Science vol. 306(5696), 2004.

http://arxiv.org/abs/quant–ph/0410092   – Cached – Similar pages

PAC–Bayesian Stochastic Model Selection

by David A. McAllester
Machine Learning Vol. 51(1), pp. 5–21, 2003.

ttic.uchicago.edu/~dmcallester/posterior01.ps   – Cached – Similar pages

Pictorial and Conceptual Representation of Glimpsed Pictures

by Mary C. Potter, Adrian Staub, and Daniel H. O'Connor
Journal of Experimental Psychology, Human Perception and Performance Vol. 30(3), 2004.

cvcl.mit.edu/IAP05/potterstauboconnor2004.pdf   – Cached – Similar pages

Blink and Shrink: The Effect of the Attentional Blink on Spatial Processing
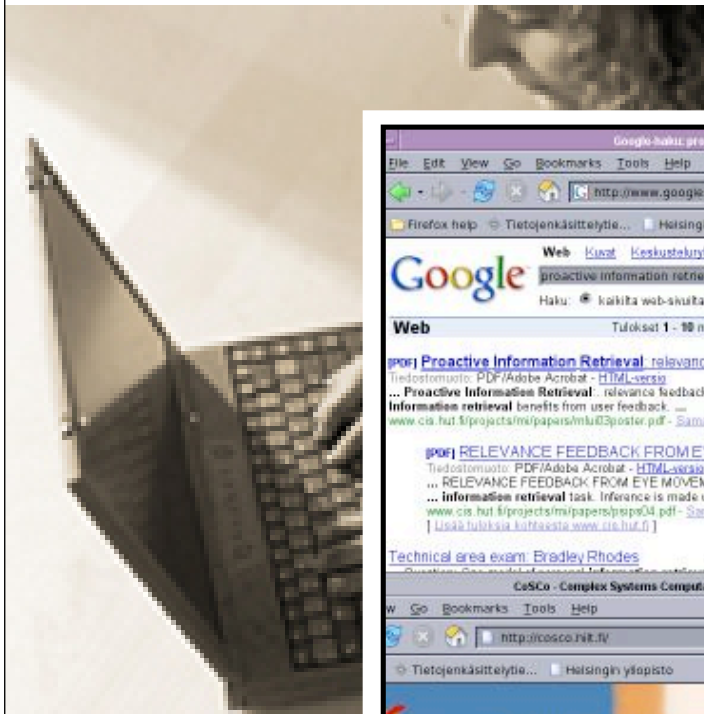
by Christian and N. L. Olivers
Journal of Experimental Psychology, Human Perception and Performance Vol. 30(3), 2004.

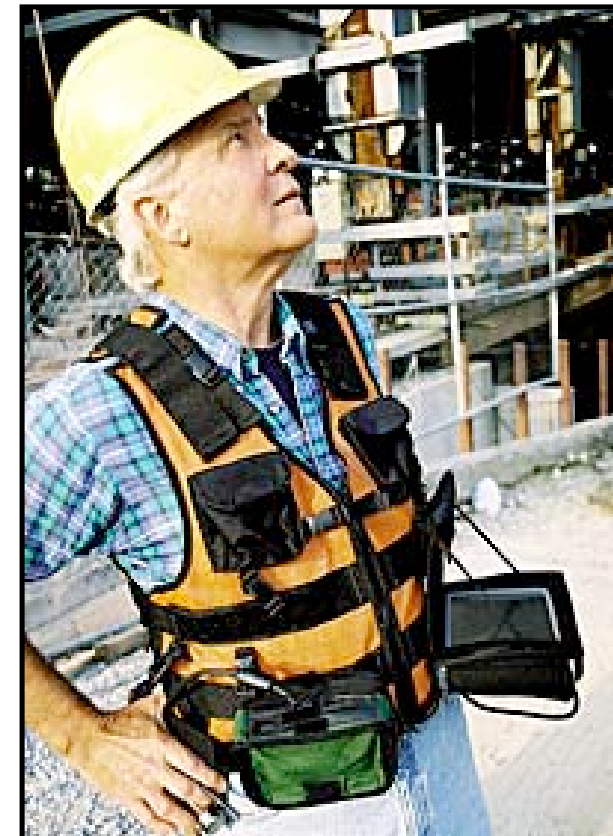http://content.apa.org/journals/xhp/30/3   – Cached – Similar pages

There are other sources of implicit feedback as well

The Minimum Error Minimax Probability Machine

Sphere-Packing Bounds for Convolutional Codes

Quantum State Transfer Between Matter and Light

PAC-Bayesian Stochastic Model Selection

Pictorial and Conceptual Representation of Glimpsed Pictures

Blink and Shrink: The Effect of the Attentional Blink on Spatial Processing

Can we infer the interests of the reader from eye movements?

Two tasks

- 1. Explicit ranking of search results is tedious => automate! (Done to an extent)

- 2. It is difficult to construct good queries => make the whole search proactive! (Hard but being done)

Uses of GnuPG

GPG is stable, production-quality software. It is frequently included in free operating systems, such as FreeBSD, OpenBSD, and NetBSD and nearly all distributions of GNU/Linux.

Although the basic GPG program has a command line interface, there exist various front-ends that provide it with a graphical user interface; for example, it has been integrated into KMail and Evolution, the graphical email clients found in the most popular Linux desktops KDE and GNOME. A plugin known as Enigmail allows GPG to be integrated with Mozilla and Thunderbird, which works on Microsoft Windows as well as Linux and other operating systems. Note that, because the plugin mechanism is not part of GPG itself and not specified by the Open PGP standard, rs were involved in their development, it is possible that en lost as a result of using such auxiliaries.

ac OS X and Windows. For Mac OS X, there is a free port 5 X user interface and its native class definitions. Cross ecause security provisions vary with operating system and pilers should routinely produce executables which will s.

Digital signature
In cryptography, digital signatures are a method of authenticating digital information often treated, sometimes too closely, as analogous to a physical signature on paper. Whilst there are analogies, there are also...

Public-key cryptography
Public-key cryptography is a form of modern cryptography which allows users to communicate securely without previously agreeing on a

GPG encrypts messages using asymmetric keypairs individually generated by GPG users. The resulting public keys can be exchanged with other users in a variety of ways, such as Internet keyservers. They must always be exchanged carefully to prevent identity spoofing by corrupting public key ↔ 'owner' identity correspondences. It is also possible to add a cryptographic digital signature to a message, so the message integrity and sender can be verified, if a particular correspondence relied upon has not been corrupted.

GPG does not use patented or otherwise restricted software or algorithms, including the IDEA encryption algorithm which has been present in PGP almost from the beginning. Instead, it uses a variety of other, non-patented algorithms such as ElGamal, CAST5, Triple DES (3DES), AES and Blowfish. It is still possible to use IDEA in GPG by downloading a plugin for it, however this may require getting a license for some uses in some countries in which IDEA is patented.

GPG is a hybrid encryption software program in that it uses a combination of conventional symmetric-key

Done

Infer *what* was interesting from a reading pattern. Sometimes easy

Xenotarsosaurus ("strange-ankle lizard") is a little-understood theropod of the late

Cretaceous (~83 – 73 mya). It probably weighed 0.7 – 1.0 tons.

The only fossil evidence consists of a small number of vertebrae and leg bones,

retrieved from the Bajo Barreal Formation, Chubut, Argentina. From these samples,

Martinez, Giminez, Rodriguez and Bochatey named the type species, X. bonapartei,

in 1986. It was probably an allosaurid.

...and sometimes hard

The history of film is one of the most rapidly moving of any artistic or communications medium ever, as befits perhaps the first great mass medium of the modern era. Film has gone through a remarkable array of changes and developed a remarkable variety and sophistication in barely more than one hundred years of existence.

Assumptions

1. There is a link between relevance of a word and eye movements

2. There is data for which relevance is known (but for different topics)

3. The link is independent of the actual topic

=> The link can be learned from data collected on a set of topics, and generalized to new topics
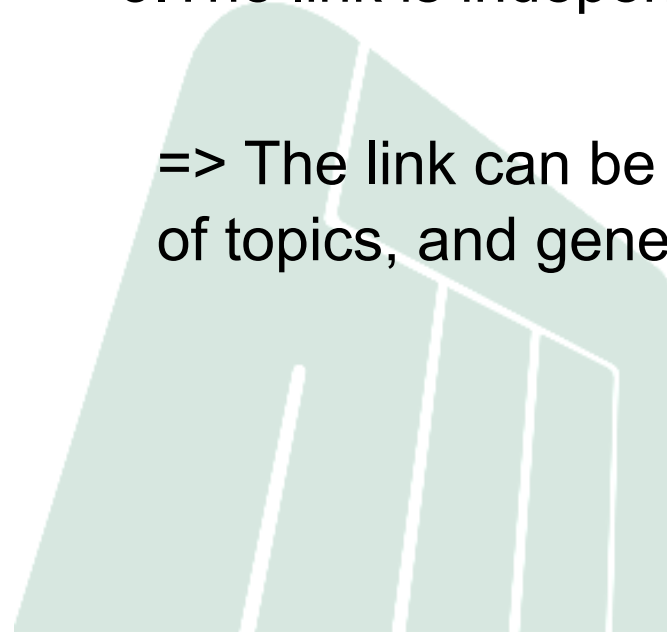
Results

- Given 10 documents with eye movements, infer the implicit query and apply it to 244 new documents.

- Statistically significant improvement from random average precision: 5.11% --> 9.81%

- Usable to supplement other feedback; absolute value is still needs improvement.

	Random	Implicit	Explicit
Space exploration	5.11	**17.74**	62.64
Speeches	5.11	10.55	45.27
Television	5.11	6.78	**29.77**
Transportation	5.11	**16.73**	**34.86**

Models of networks

Models of networks

- Task: Find **hidden structure in large graphs** or networks: modules, components, clusters

- Special in our work: Define as (rapidly computable) generative probabilistic models

- Application areas: Social networks, hypertexts, biological networks, ... More generally: **mining in graphs**

Example: Model that defines modules by clustering links
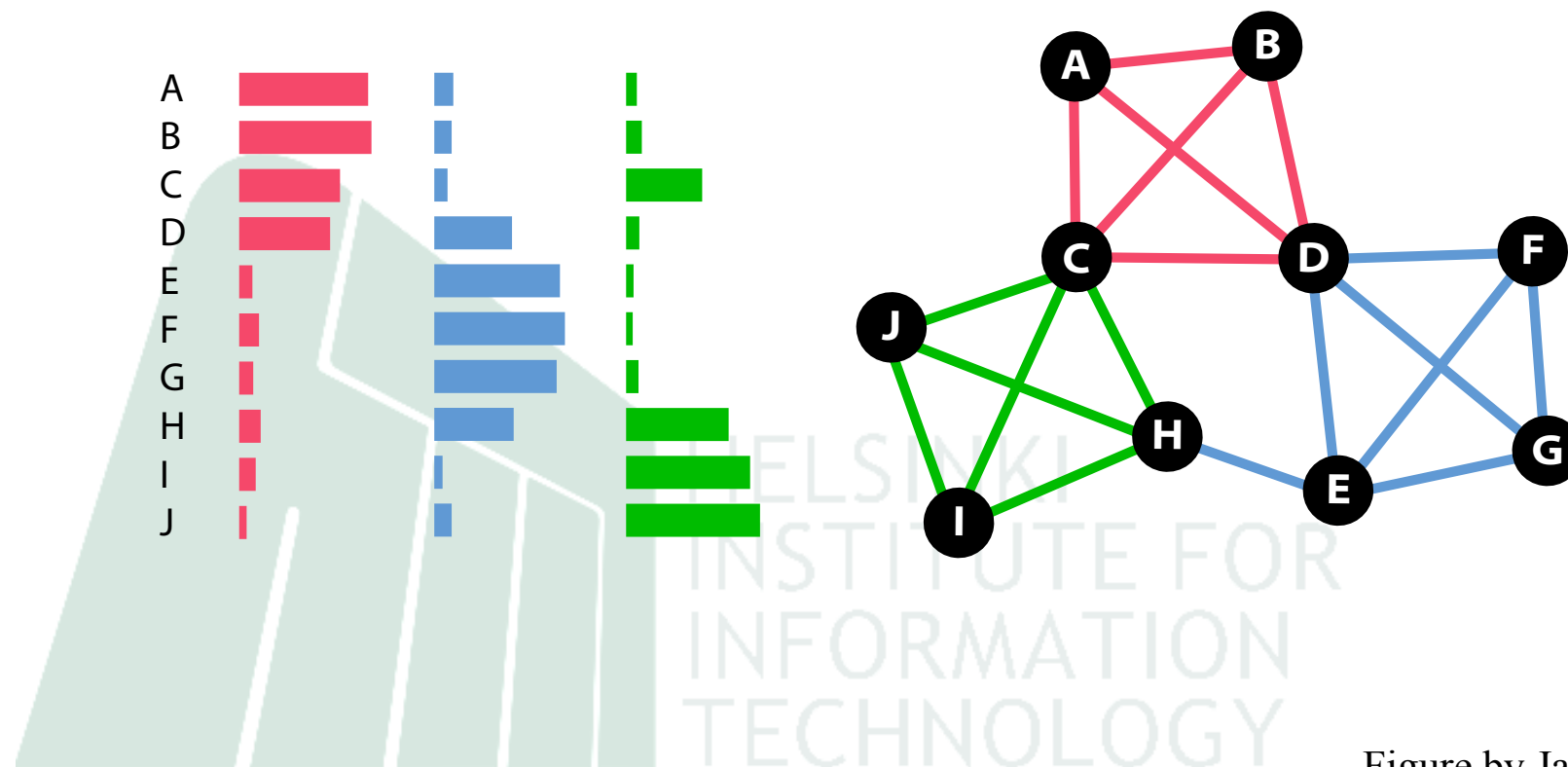
- Extends module or clique search

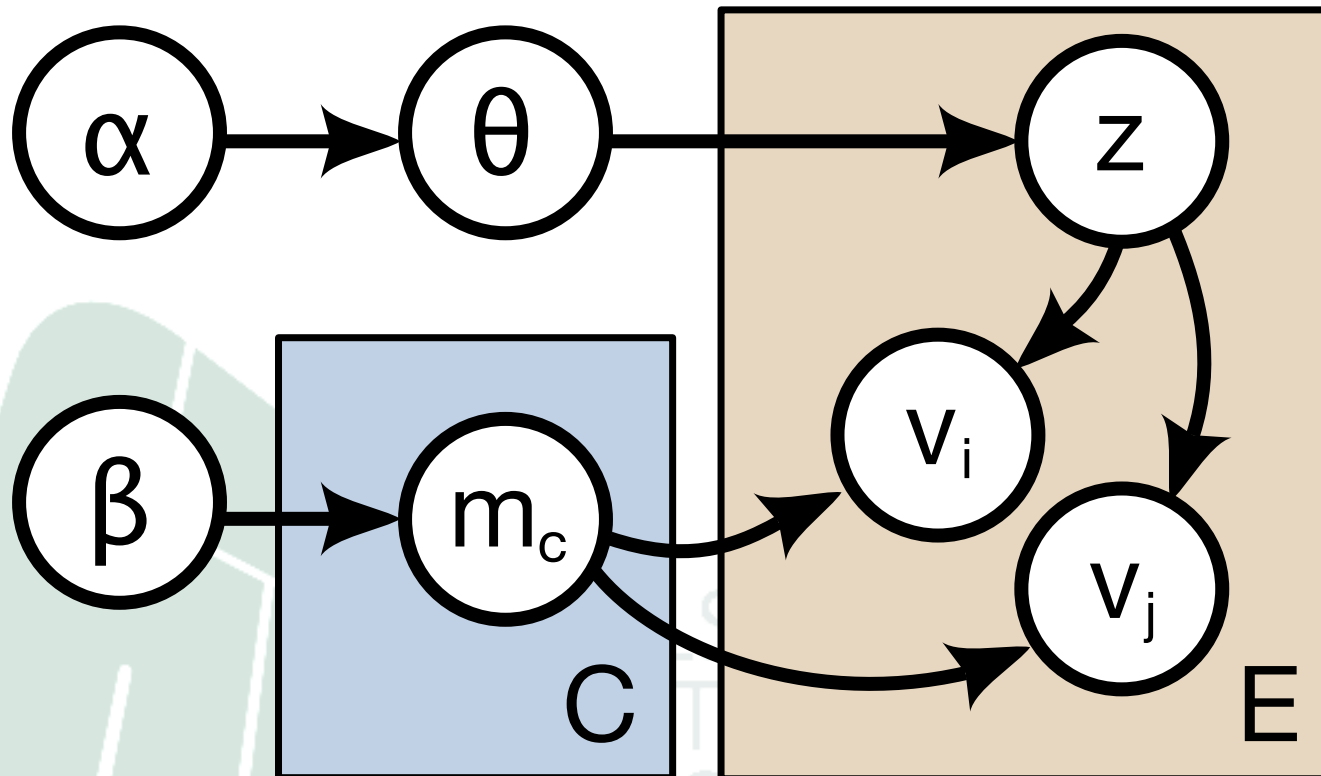
Figure by Janne Aukia

The generative model

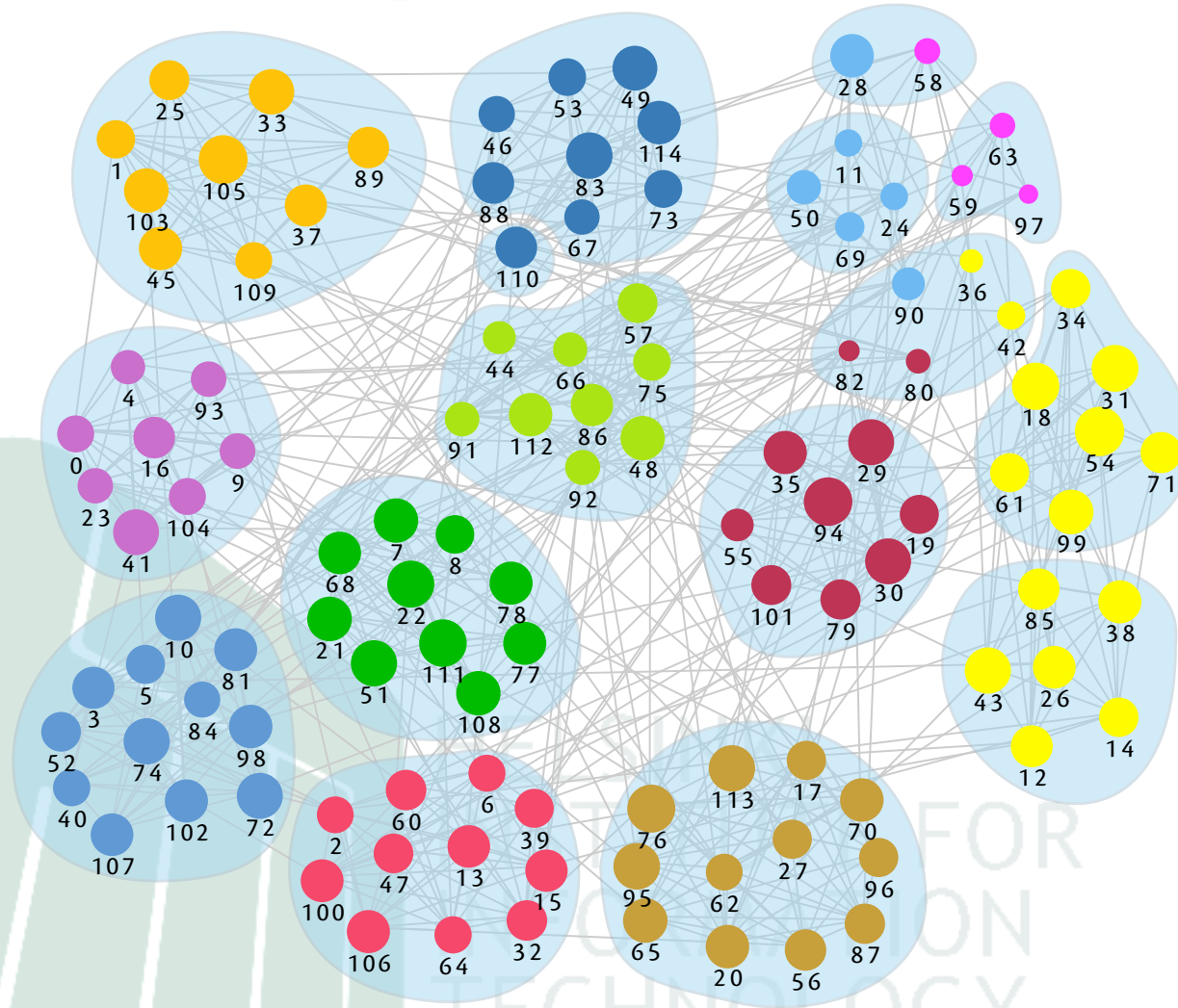Figure by Janne Aukia

Football network: Who plays against whom

Figure by Janne Aukia
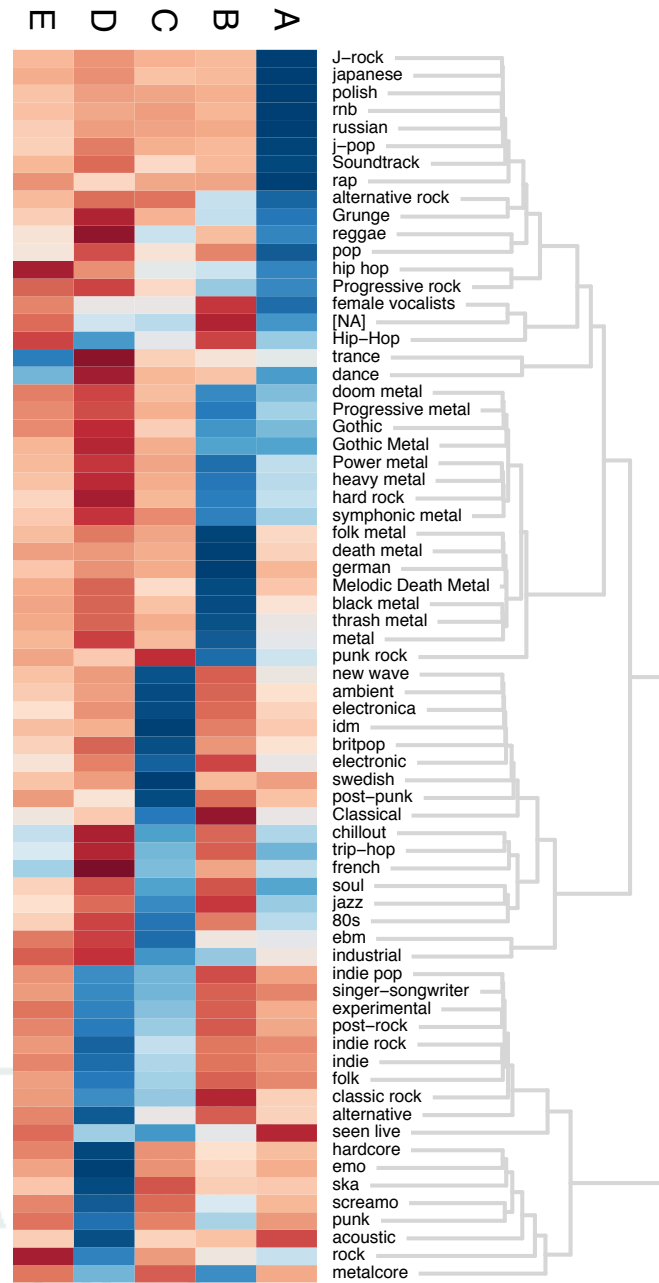
Friendship network: Last.fm internet radio

Figure by Janne Aukia

Figure by Janne Aukia