Key Pathway Mining - Network-based data mining in biological data (Jan Baumbach)

Recent advances in systems biology provided us with massive pathway data that model the interplay of biological entities, i.e. genes, proteins and metabolites. The emerging biological networks are modeled as graphs. Besides, we measure the entities' activity (expression) by means of OMICS technology, for example microarrays, RNA-Seq and ion mobility spectrometry. Here, we seek to gain deeper insights into the expression data by combining it with the network data. We will introduce KeyPathwayMiner. Given a biological network modelled as graph and a set of expression studies, it efficiently finds and visualizes connected sub-networks where most entities are expressed in most cases. KeyPathwayMiner finds all maximal connected sub-networks where all nodes but K exceptions are expressed in all experimental studies but at least L exceptions. We demonstrate the power of the approach by studying Huntingon's disease gene expression data sets together with the human protein-protein interaction network. In the second part of the talk, we will briefly investigate statistically how much (or how little?) we actually know about the biological networks that we use for our studies.

Jan Baumbach studied computer science in the natural sciences at Bielefeld University and Rothamsted Research (UK) where he received a Diploma degree with distinction in 2005. During his PhD studies at the Center for Biotechnology in Bielefeld, he was also a visiting research scholar at the University of California at San Francisco. After graduating with distinction in 2008, he worked at the University of California at Berkeley in the Algorithms group of Richard Karp. Since 2010 he is the head of the Computational Systems Biology research group at the Max Planck Institute for Informatics, Saarbrücken, Germany. His main research interests are Data Mining in biomedical data, clustering algorithms for large data sets, gene regulatory network analysis and reconstruction.