A simple clustering of the BioModels database using semanticSBML
Falko Krause and Wolfram Liebermeister
(falko.krause|wolfram.liebermeister)@biologie.hu-berlin.de
www.semanticsbml.org

The BioModels\(^1\) database contains biochemical network models in SBML format, in which the biochemical meaning of elements is specified by MIRIAM-compliant RDF annotations\(^2\). We used these annotations to define a similarity measure for models, scoring the overlap of the biochemical systems described. Based on this score, we used two-way clustering to detect groups of similar models and groups of co-occurring model elements.

To recognize and compare biochemical elements, we used routines from the software semanticSBML\(^3\). A Python script \(b\) extracts all MIRIAM annotations (regardless of their qualifiers) using the semanticSBML annotation classes \(a\). The result is a matrix in which the rows represent the models (e.g. BioModel 001), while the columns represent specific annotations (e.g. urn:miriam:reactome:REACT_15422). A matrix element is set to 1 if an identifier occurs in a model and to 0 otherwise. This matrix was used as an input for a hierarchical clustering algorithm (implemented in Matlab) \(c\) and the clustered matrix was visualized \(d\).

Model clustering allows to detect models describing similar biochemical processes (e.g. glycolysis) and their specific common elements. This may help to find candidate models for completing a given initial model, which could then be merged using semanticSBML \(e\).

---

References
1 Curated Models; released 25th September 2007; http://www.abi.ac.uk/biomodels/
3 Liebermeister W, Krause F, Klipp E. (2008), Merging of systems biology models with semanticSBML.
4 Krause F, Uhlenendorf J., Lubitz T, Klipp E., Liebermeister W. Annotation and merging of SBML models with semanticSBML, submitted