

# Identifying pattern in reaction networks of computational models

Fabienne Lambusch, Rebekka Alm, Ron Henkel, Dagmar Waltemath

October 2015 | COMBINE, Salt Lake City, US



**SYSTEMS BIOLOGY  
BIOINFORMATICS  
ROSTOCK**

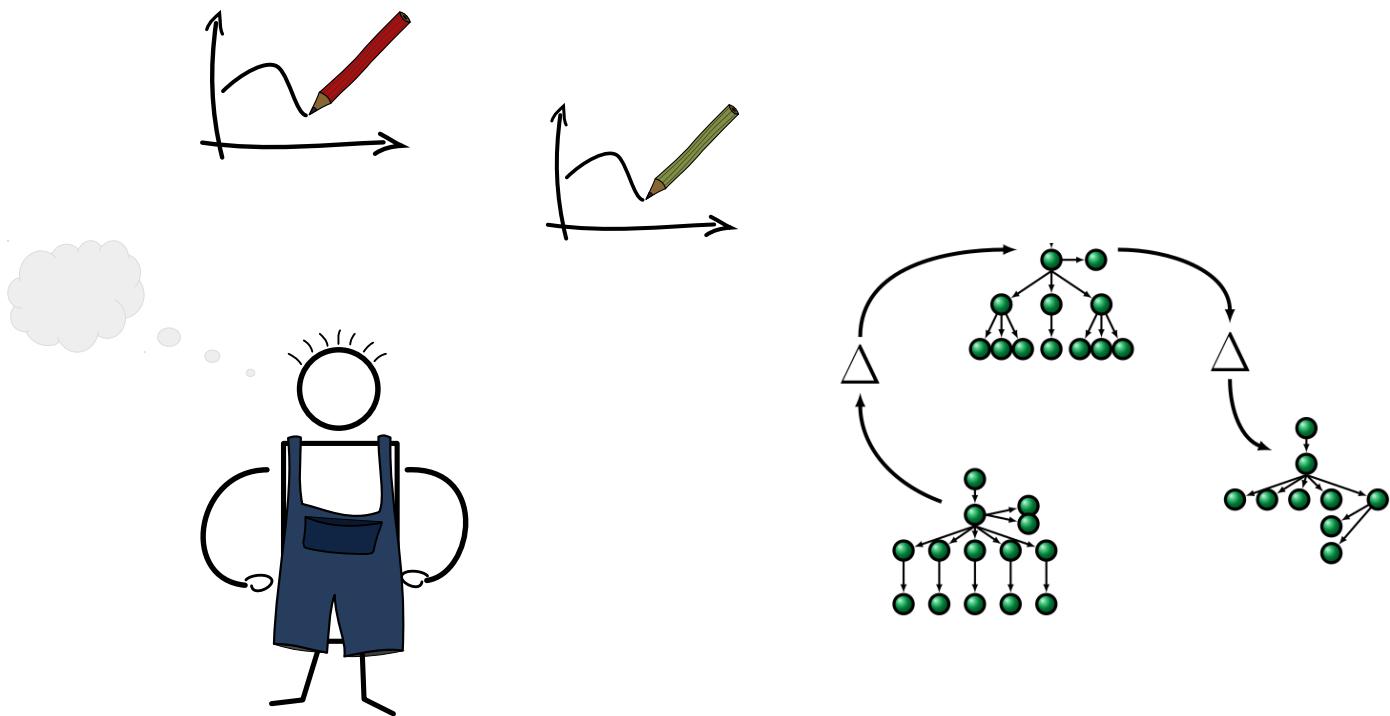
**Universität  
Rostock**



Traditio et Innovatio

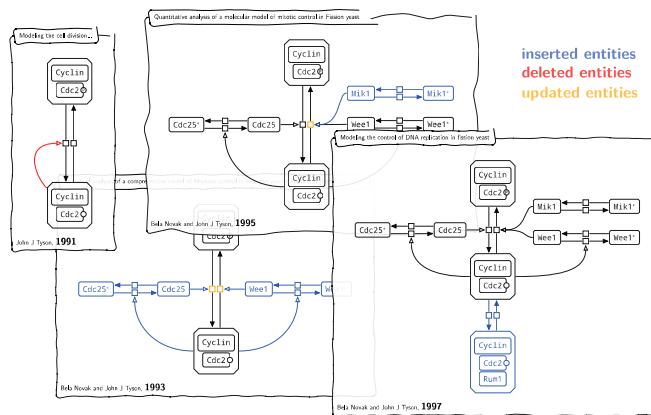
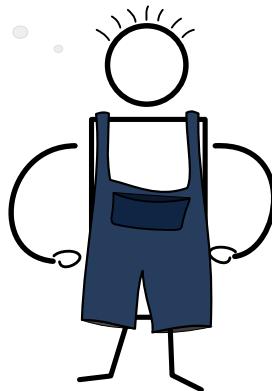
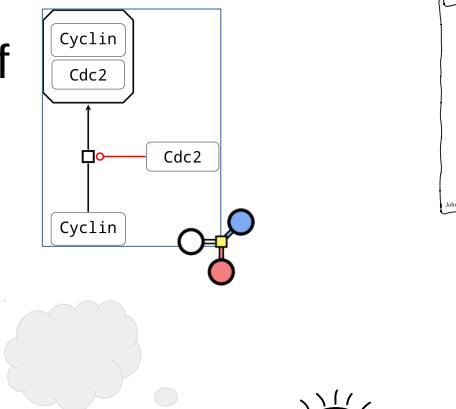


## How can we calculate the similarity between biological simulation models?



# How can we calculate the similarity between biological simulation models?

versions of  
models



thematic sets  
of models

arbitrary models

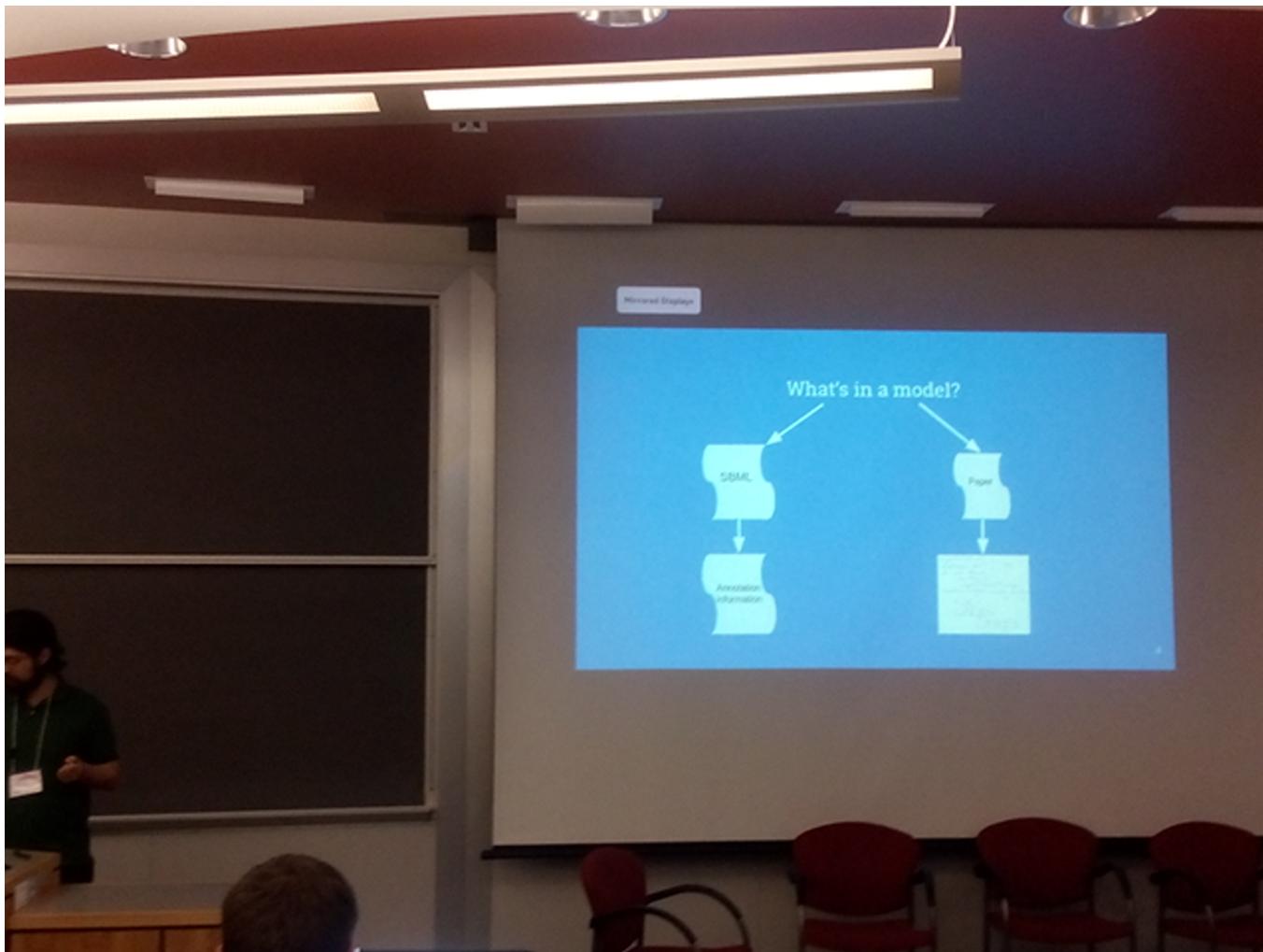
Query: Calcium Signalling [all]

The search returned 174 models.

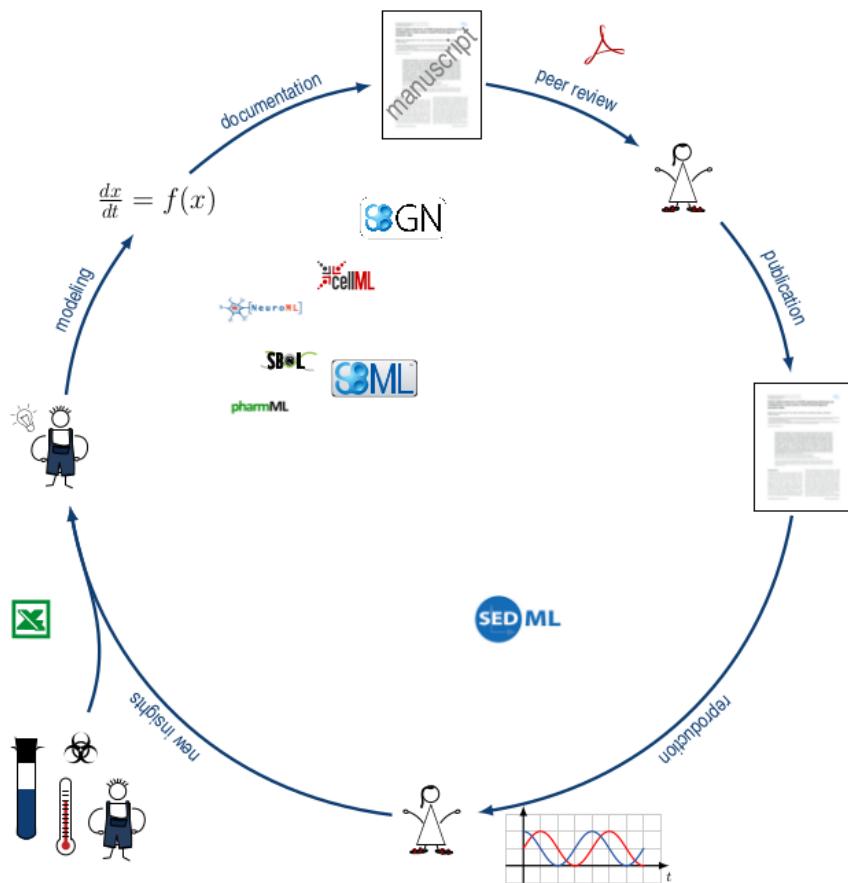
[New Search](#)

39 Curated models returned:

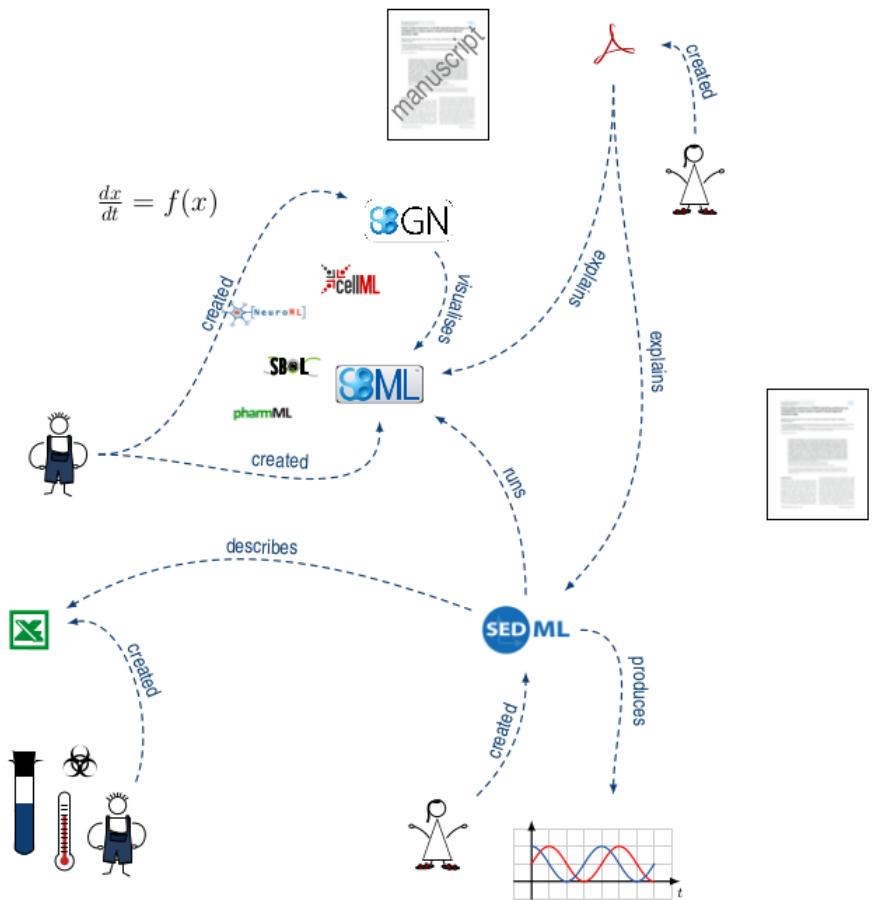
BioModels ID	Name
<a href="#">BIOMD0000000043</a>	Borghans1997 - Calcium Oscillation - Model 1
<a href="#">BIOMD0000000044</a>	Borghans1997 - Calcium Oscillation - Model 2
<a href="#">BIOMD0000000045</a>	Borghans1997 - Calcium Oscillation - Model 3
<a href="#">BIOMD0000000058</a>	Bindschadler2001_coupled_Ca_oscillators
<a href="#">BIOMD0000000059</a>	Fridlyand2003_Calcium_flux



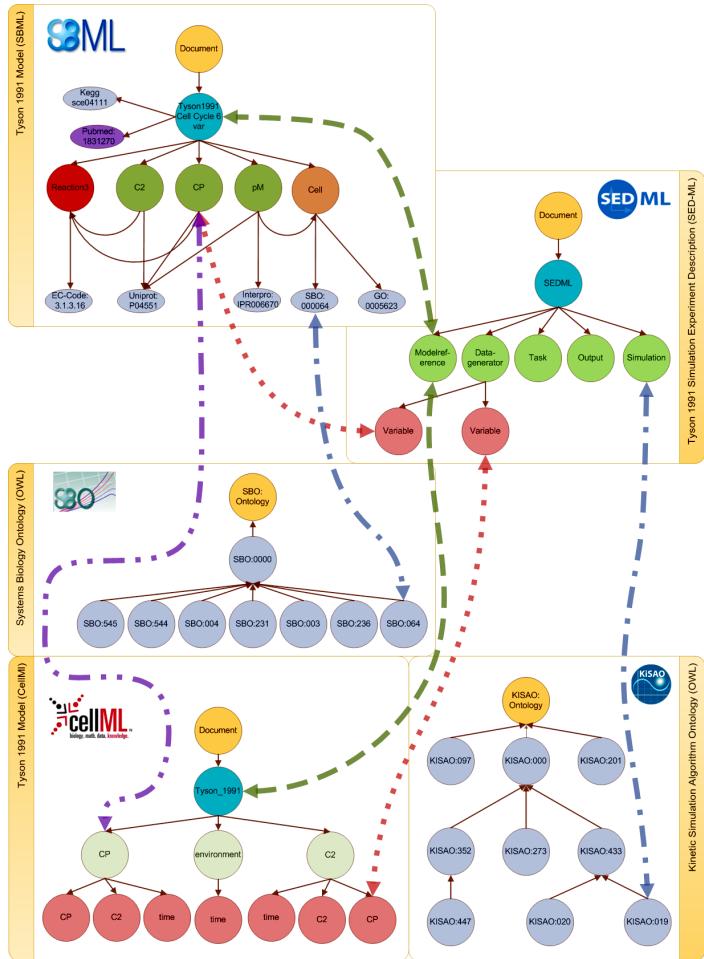
Jose-Juan Tapia @ COMBINE2015



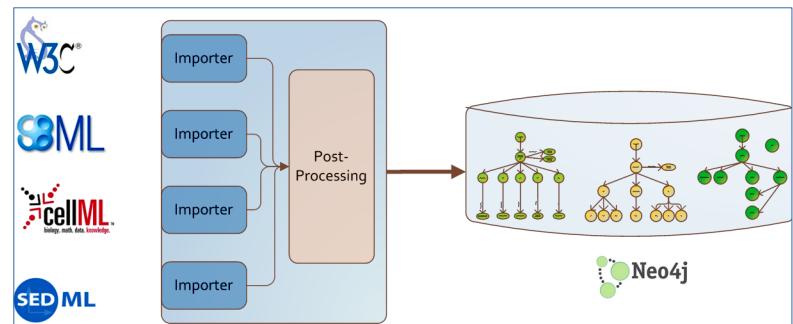
- What constitutes a simulation study in computational biology?
  - Publication
  - Experimental data
  - Model file(s)
  - Simulation setup(s)
  - Simulation data
  - Visualisations
  - Model assumptions
  - Operating Procedures



- What constitutes a simulation study in computational biology?
- How are model-related data linked amongst each other?



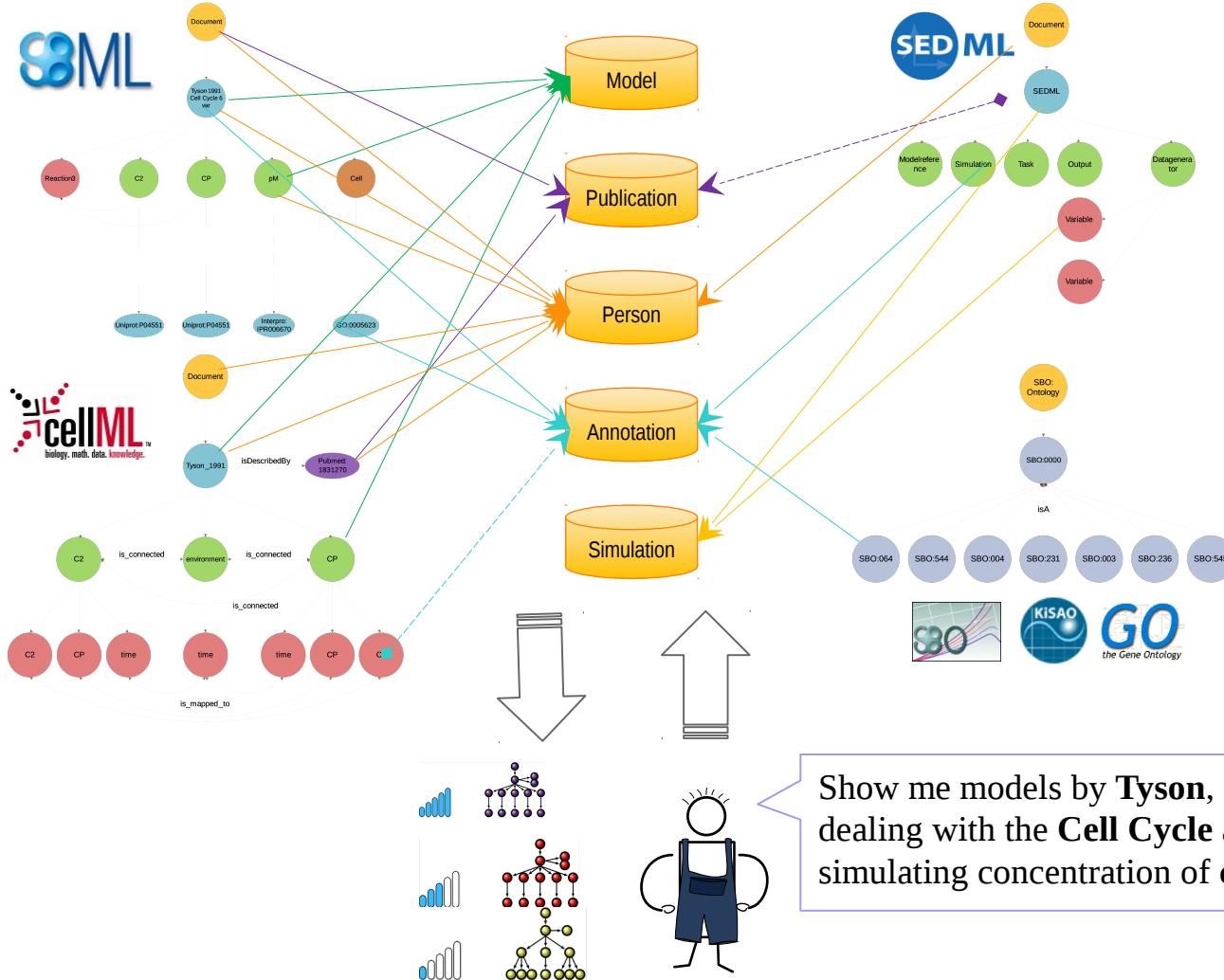
- What constitutes a simulation study in computational biology?
- How are model-related data linked amongst each other?
- How can we represent these links on the storage layer?





# Model similarity

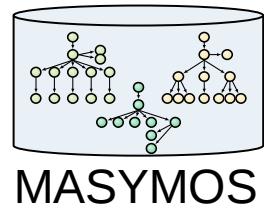
## Annotation-based ranked retrieval



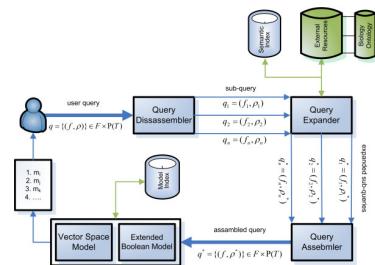
Ron Henkel

Henkel (2010)

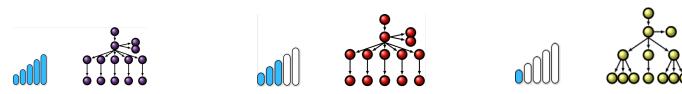
*More descriptive model features from annotations*



Ranking



Henkel (2010)



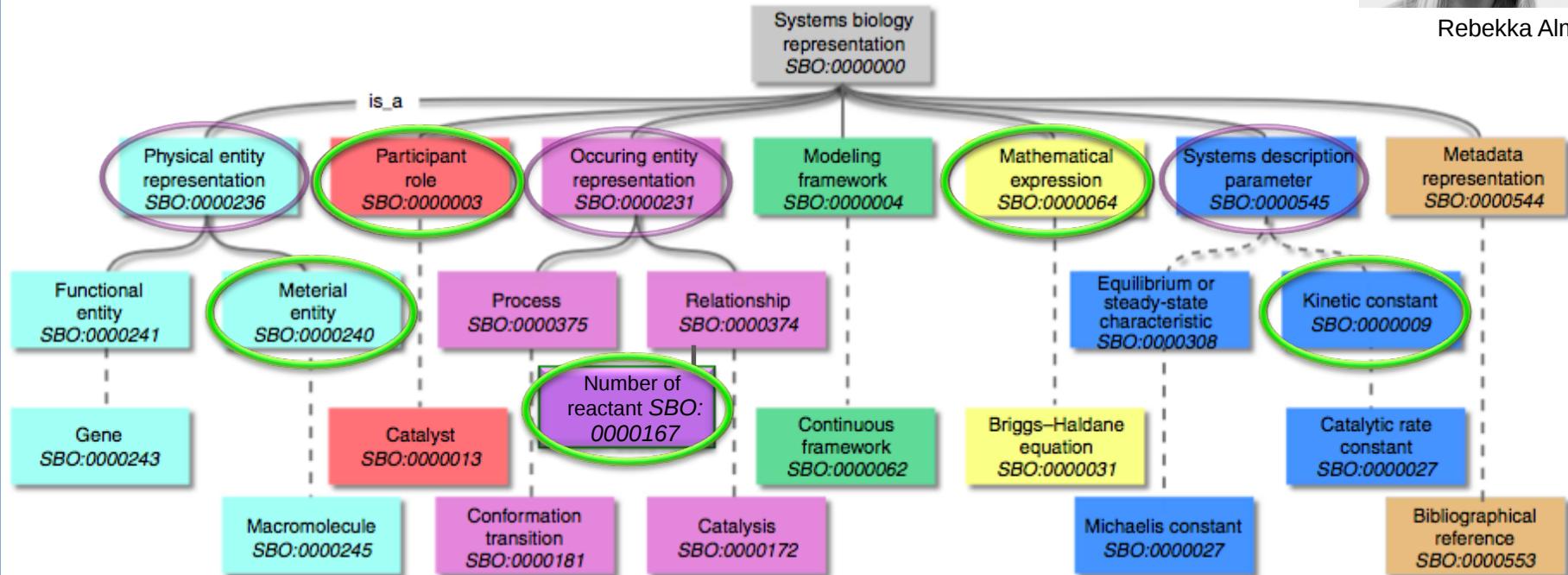
*Structure-based similarity scores*

### SBML models from BioModels Database (R25)

GO (~11.000 references) | SBO (~13.000) | ChEBI (6.000)



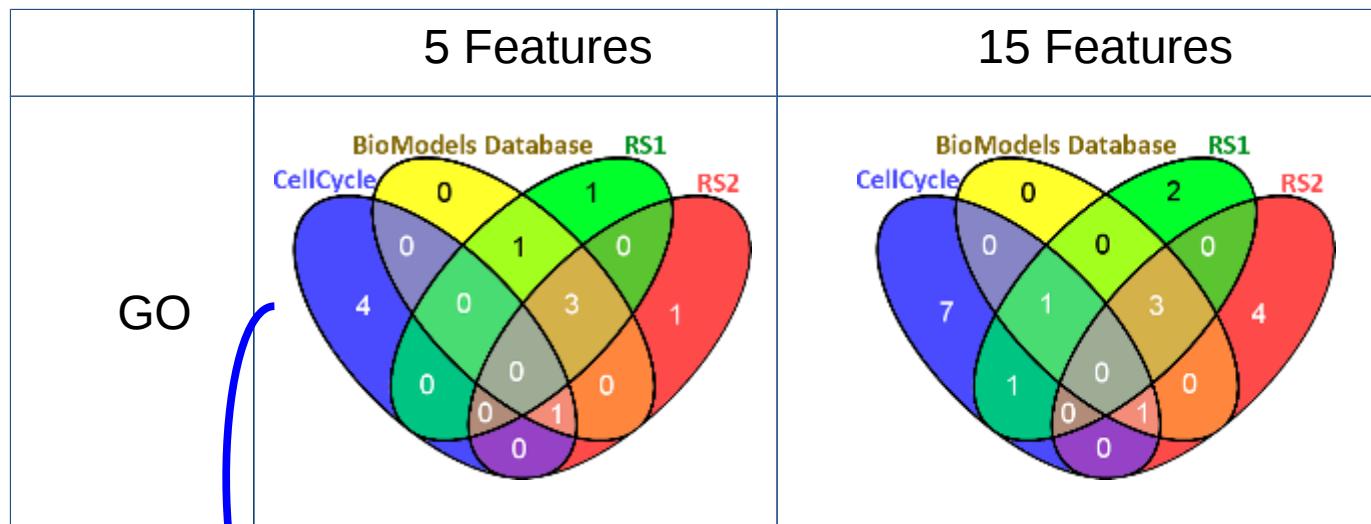
Rebekka Alm



Adapted from: Courtot et al.: Controlled vocabularies and semantics in systems biology. In: Molecular Systems Biology 7 (2011), Nr. 1. Alm (2015) JBMS

### SBML models from BioModels Database (R25)

**GO (~11.000 references) | SBO (~13.000) | ChEBI (6.000)**



GO:0022411 (depth 4) cellular component disassembly

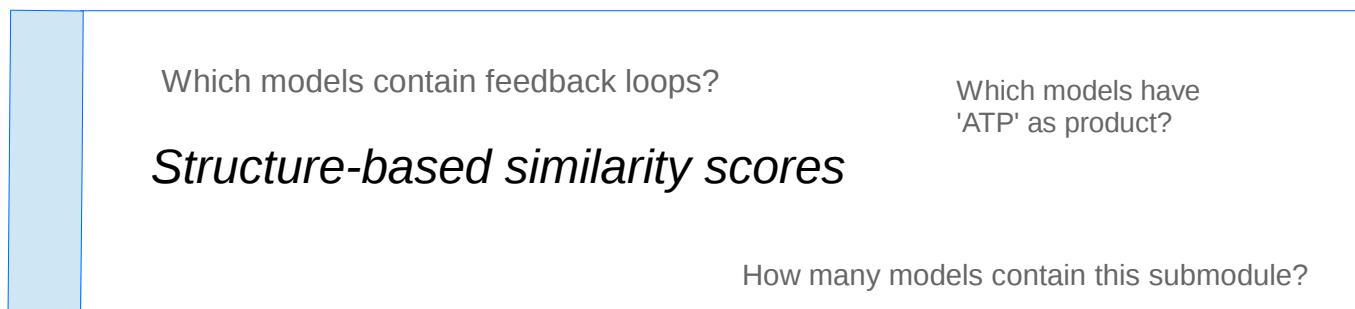
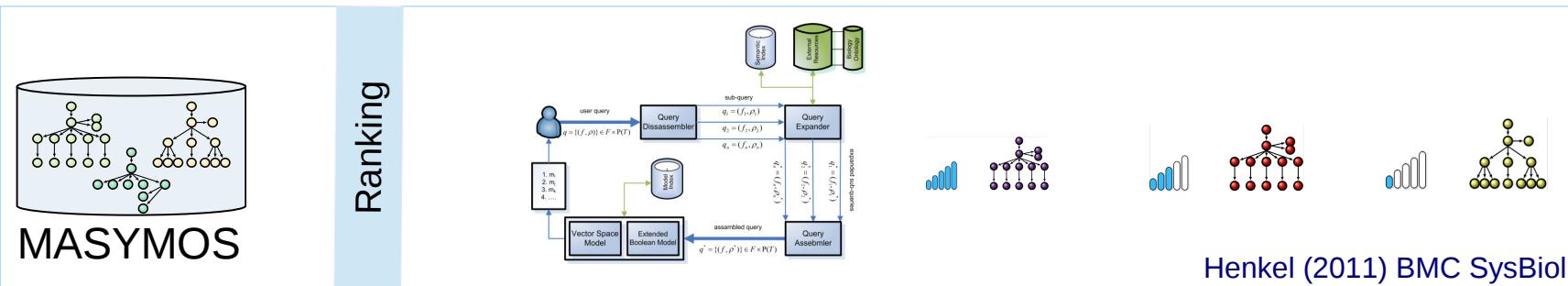
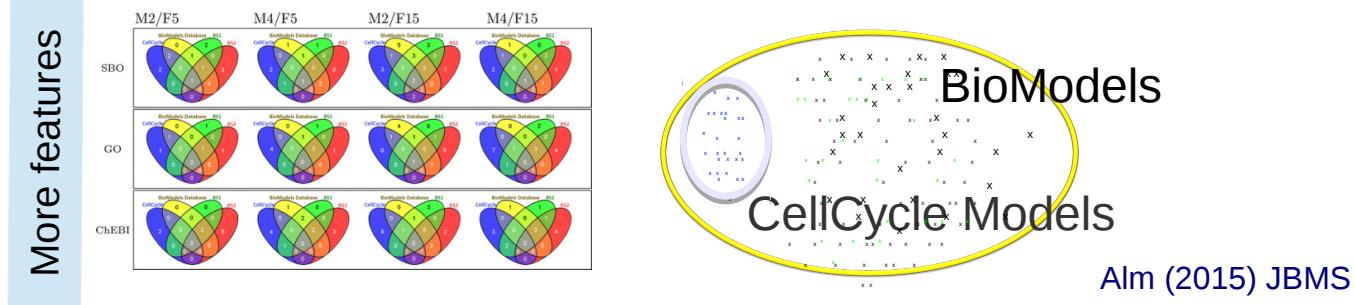
GO:0030163 (depth 5) protein catabolic process

GO:0051726 (depth 5) **regulation of cell cycle**

GO:0065009 (depth 3) regulation of molecular function

GO:0071822 (depth 5) protein complex subunit organization

Adapted from: Alm (2015) JBMS



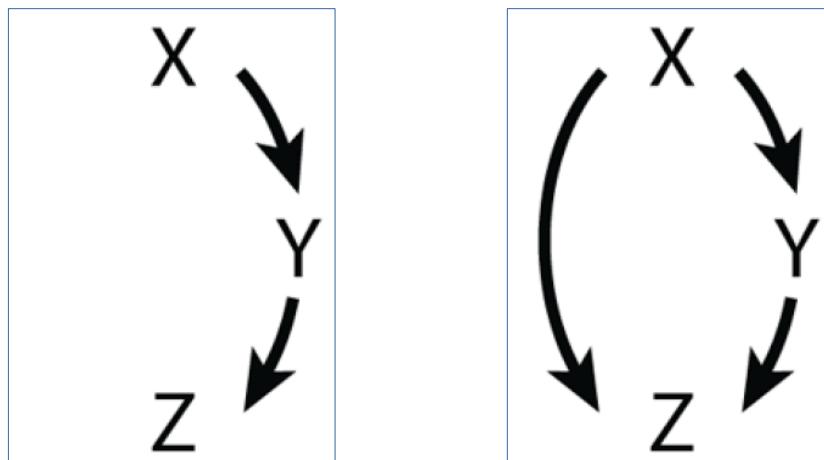


# Structure-based similarity

Models are similar, if they share (many) common subgraphs / patterns / motifs



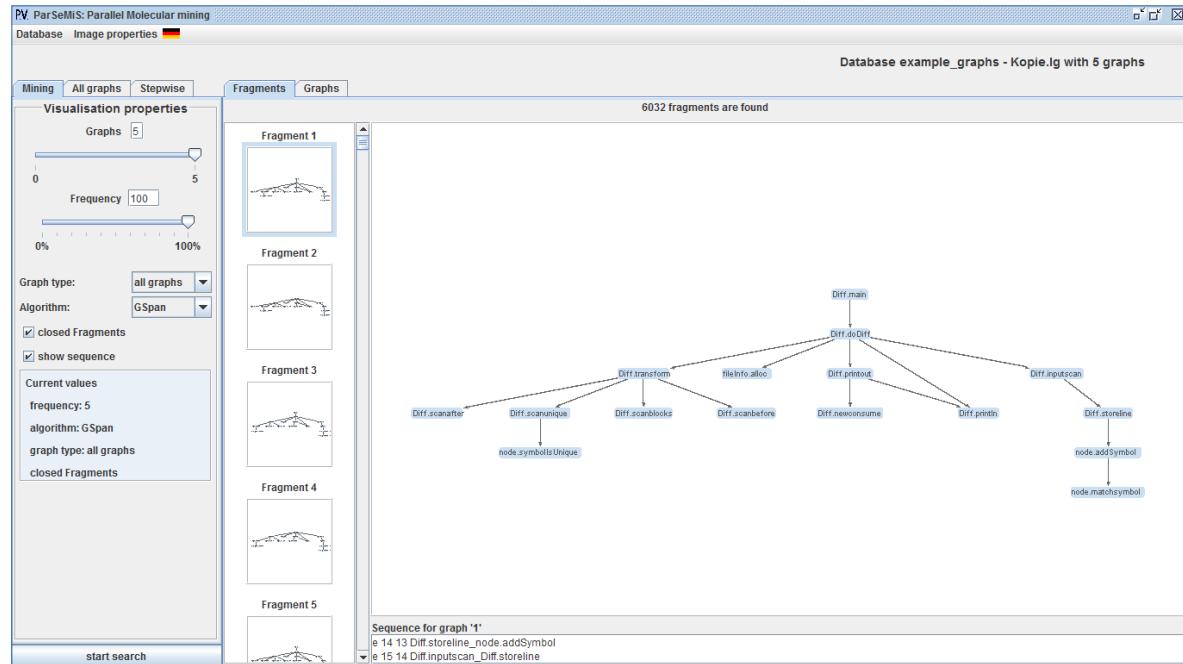
Fabienne Lambusch



## Parallel and Sequential Mining Suite (ParSeMiS)

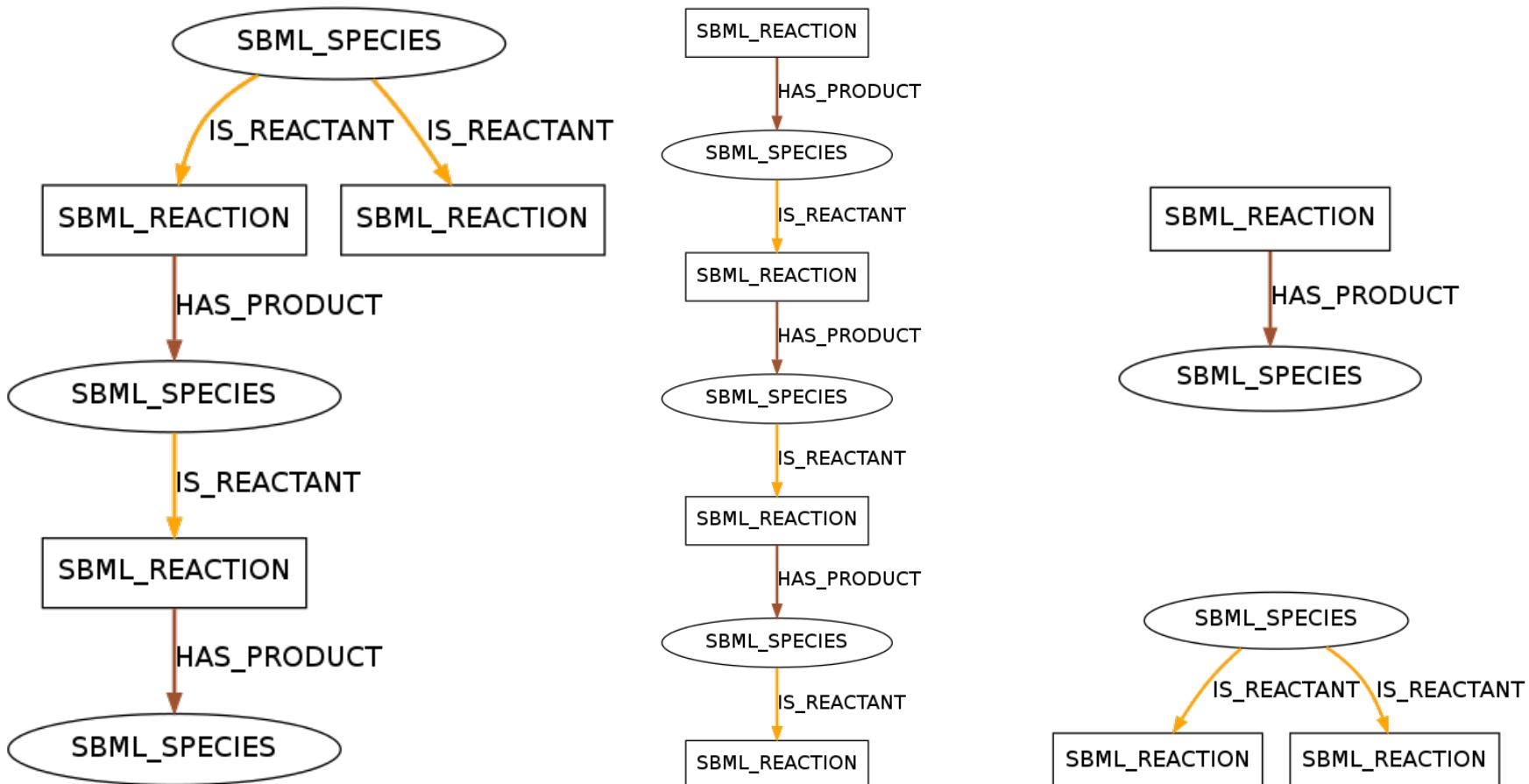
Find patterns  
which occur in at  
least X of the  
models

- Input: SBML  
models from  
MASYSMOS
- Output: List of  
patterns and  
frequencies



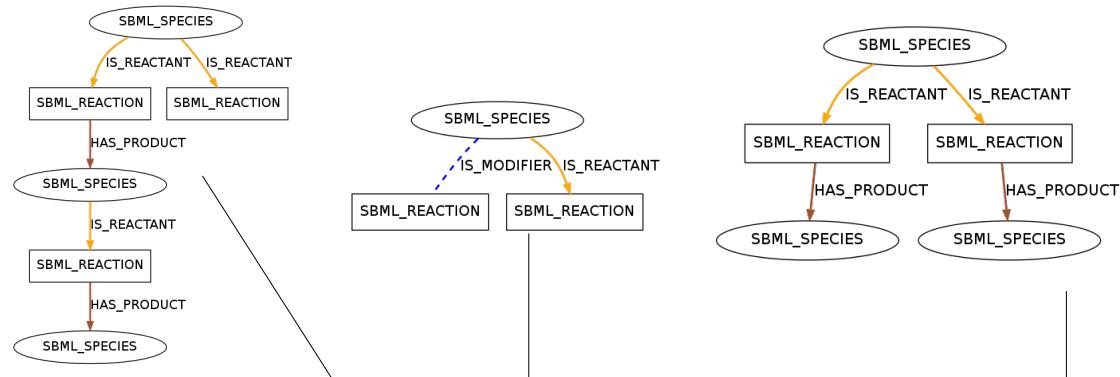
ParSeMiS: <https://github.com/timtadh/parsemis>

# Structure-based similarity Pattern



Frequency of occurrence (560 models, threshold 350 models)

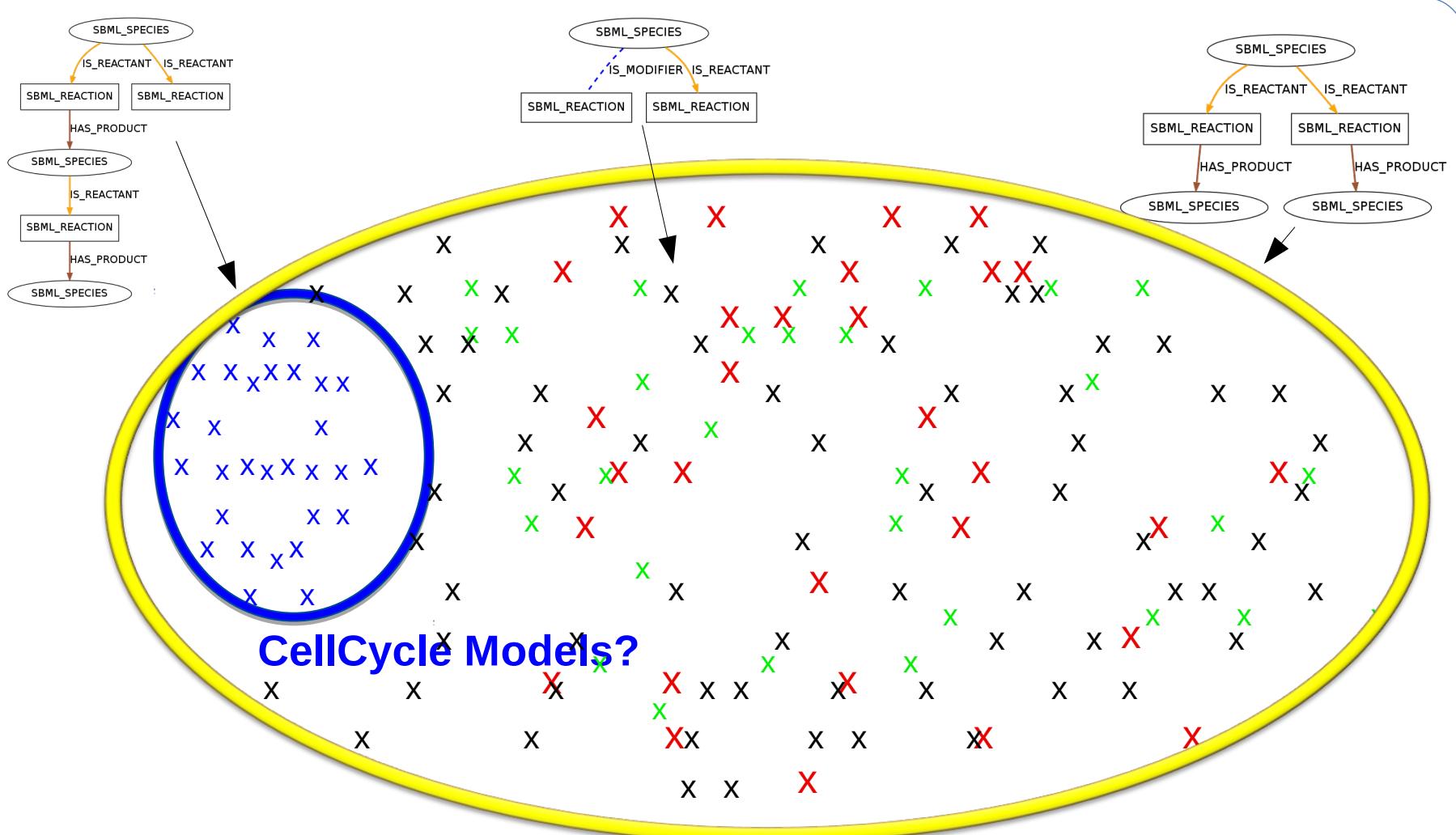
# Structure-based similarity Pattern frequencies



	A	B	C	D	E	F	G	H	I	J
1	m.NAME	350-30	351-07	351-18	351-19	354-17	355-13	359-06	361-23	364-05
2	Goldbeter1991 - Min Mit Oscil	0	1	0	0	0	0	0	0	0
3	Tyson1991 - Cell Cycle 6 var	4	0	8	4	4	17	6	2	8
4	Goldbeter1991 - Min Mit Oscil; Expl Inact	0	1	0	0	0	0	0	0	0
5	Edelstein1996 - EPSP ACh species	85	0	85	28	20	20	14	0	68
6	Levchenko2000_MAPK_noScaffold	98	0	156	212	194	314	288	0	80
7	Kholodenko2000 - MAPK feedback	0	0	0	0	0	0	0	0	0
8	Huang1996 - MAPK ultrasens	0	0	0	0	0	0	0	0	0
9	Gardner1998 - Cell Cycle Goldbeter	6	1	15	12	12	15	12	0	6
10	Tyson1991 - Cell Cycle 2 var	0	0	0	0	0	0	0	0	0
11	Novak1997 - Cell Cycle	41	0	58	154	108	98	98	14	18
12	Curto1998_purineMetabol	0	0	0	0	0	0	0	0	0
13	Poolman2004_CalvinCycle	100	34	129	196	118	145	102	36	116
14	Elowitz2000_Repressilator	0	0	0	0	0	0	0	0	0
15	Levchenko2000_MAPK_Scaffold	22577	0	24132	206239	217357	187133	59990	0	3998
16	Edelstein1996 - EPSP ACh event	19	0	19	20	20	20	14	0	12
17	Markevich2004_MAPK_orderedMM	0	3	2	2	2	2	4	10	2
18	Markevich2004_MAPK_orderedElementary	7	0	6	10	10	30	12	0	6
19	Smolen2002_CircClock	0	0	2	2	2	2	4	2	2
20	Markevich2004_MAPK_phosphoRandomMM	5	22	10	8	12	16	10	32	6
21	Markevich2004_MAPK_phosphoRandomElementary	28	0	24	64	64	103	76	0	24
22	Vilar2002_Oscillator	41	0	50	66	69	74	56	0	30
23	Smolen2004_CircClock	4	4	0	8	0	8	0	4	0
24	Ueda2001_CircClock	235	24	201	1584	2521	609	2292	27	58
25	Scheper1999_CircClock	0	2	2	2	2	4	2	2	2



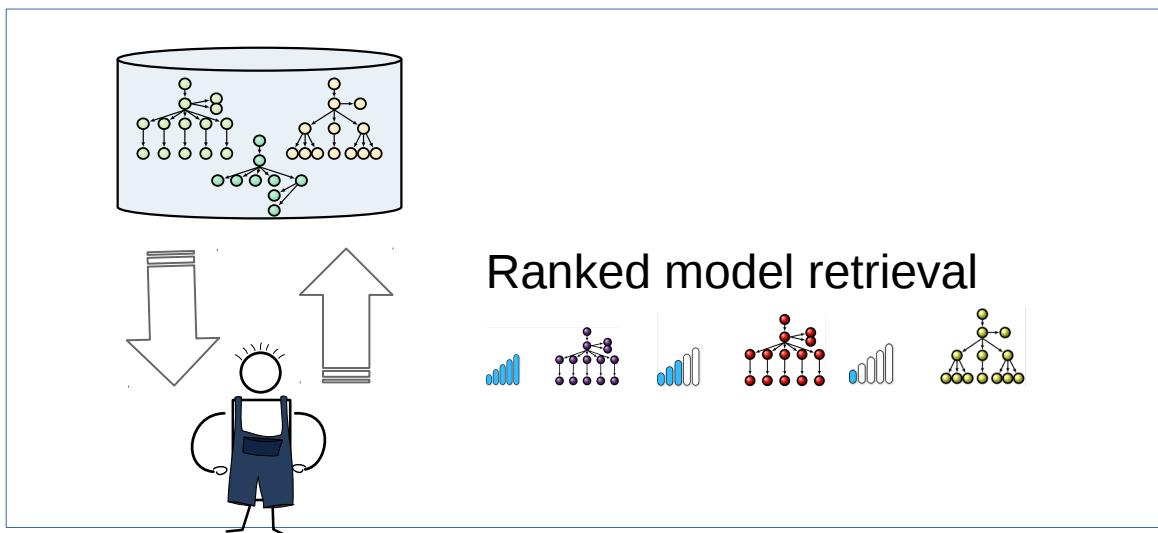
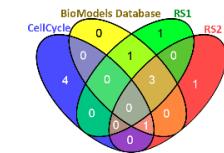
# Structure-based similarity Clustering



## Available pattern for search



## Set-dependent features



## Ranked model retrieval



Thank you for your attention.



2013

Fabienne Lambusch | Martin Scharm | Dagmar Waltemath | Mariam Nassar  
Tom Gebhardt | Martin Peters\* (all SEMS) | Vasundra Touré (SBGN-ED+)  
Ron Henkel (de.NBI-SYSBIO)



2015





## Additional slide: Resources

### MASYMOS

Paper: <http://database.oxfordjournals.org/content/2015/bau130>

Software demo: <http://graphdb-review.sems.uni-rostock.de:17474/browser/>

Homepage: <https://sems.uni-rostock.de/projects/masymos/>

### Annotation-based feature extraction

Paper: <http://www.jbiomedsem.com/content/6/1/20>

Code: <https://bitbucket.org/ronhenkel/masymos>

Homepage: [https://sems.uni-rostock.de/projects/masymos/model\\_feature\\_extraction](https://sems.uni-rostock.de/projects/masymos/model_feature_extraction)

### Structure-based similarity

Paper: on the way...

Homepage: [https://sems.uni-rostock.de/projects/masymos/model\\_feature\\_extraction](https://sems.uni-rostock.de/projects/masymos/model_feature_extraction)