

Visualizing metabolomics data in directed biological networks

Denise Slenter, Chris Evelo, Egon Willighagen

Twitter: @SMaLLCaT4Sci and @BiGCaT_UM

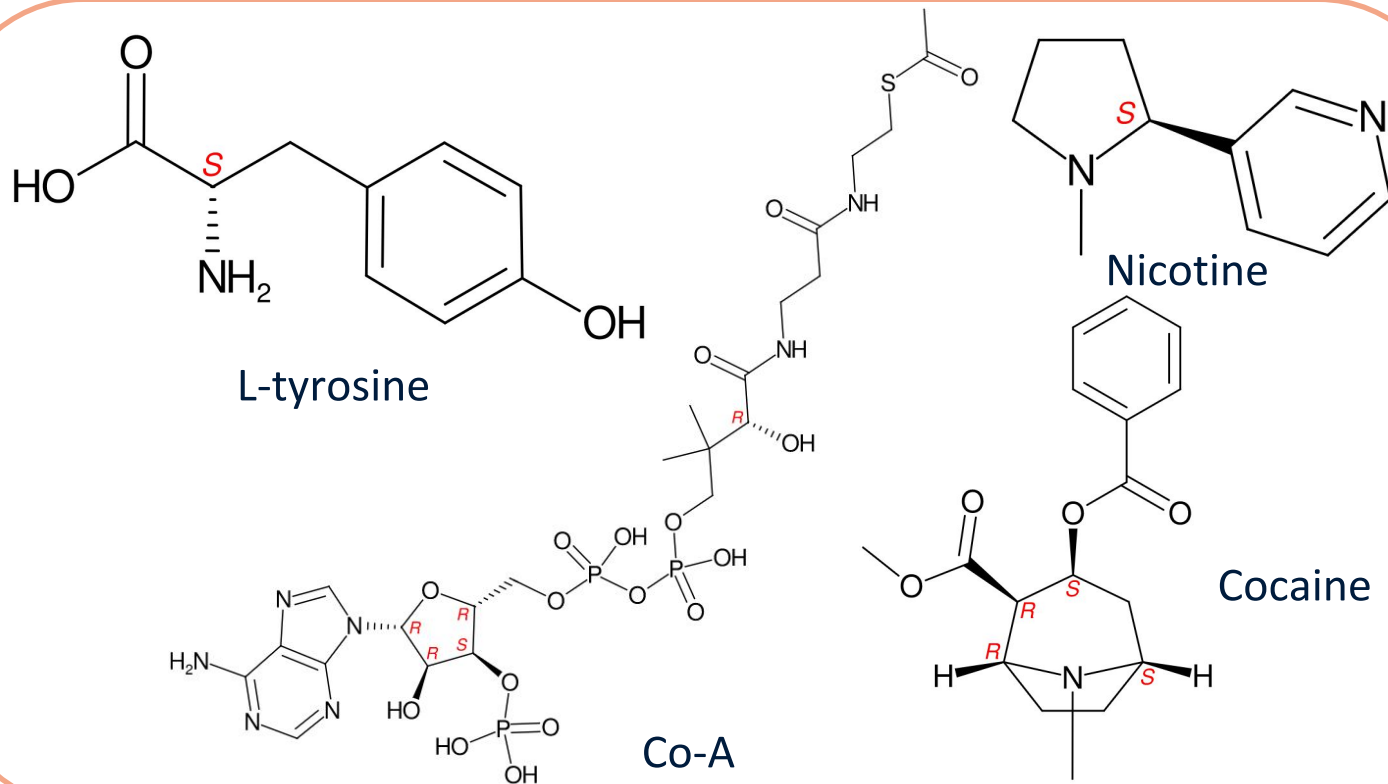
Blog: <http://smallcats4science.blogspot.nl>

ORCID: 0000-0001-8449-1318

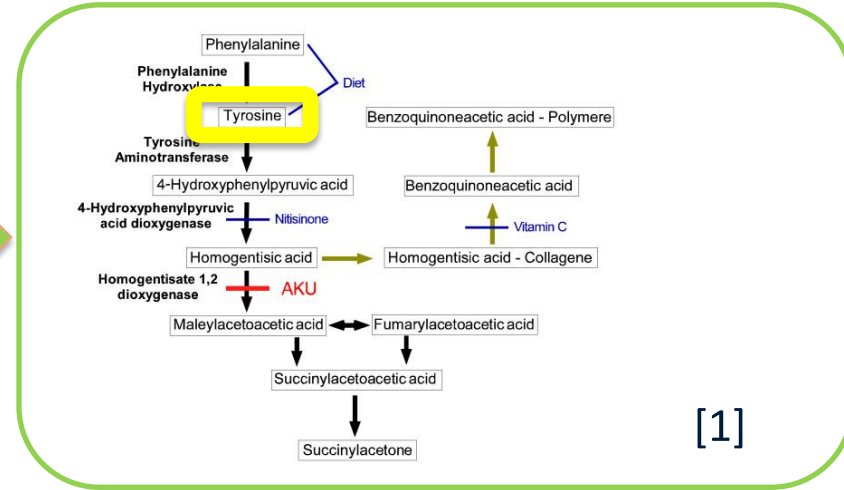
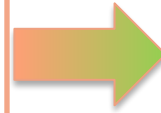
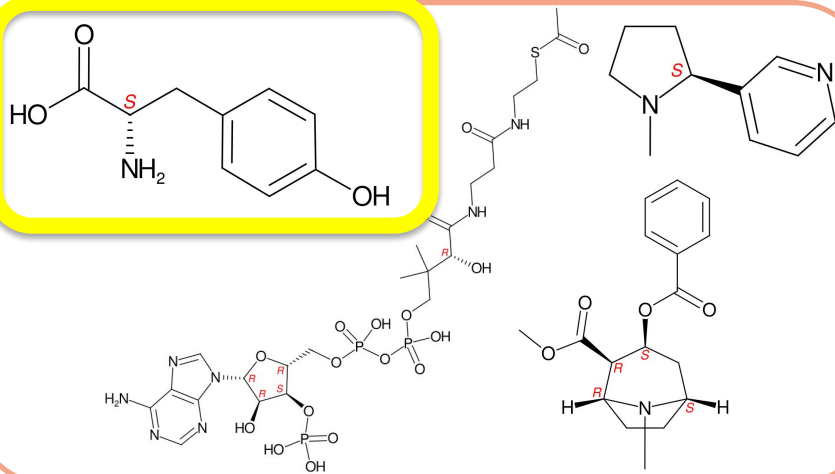
2019-11-14 IDS-BiGCaT, Data Science Research Seminar Series



Linking metabolomics data to pathways...



Linking metabolomics data to pathways...



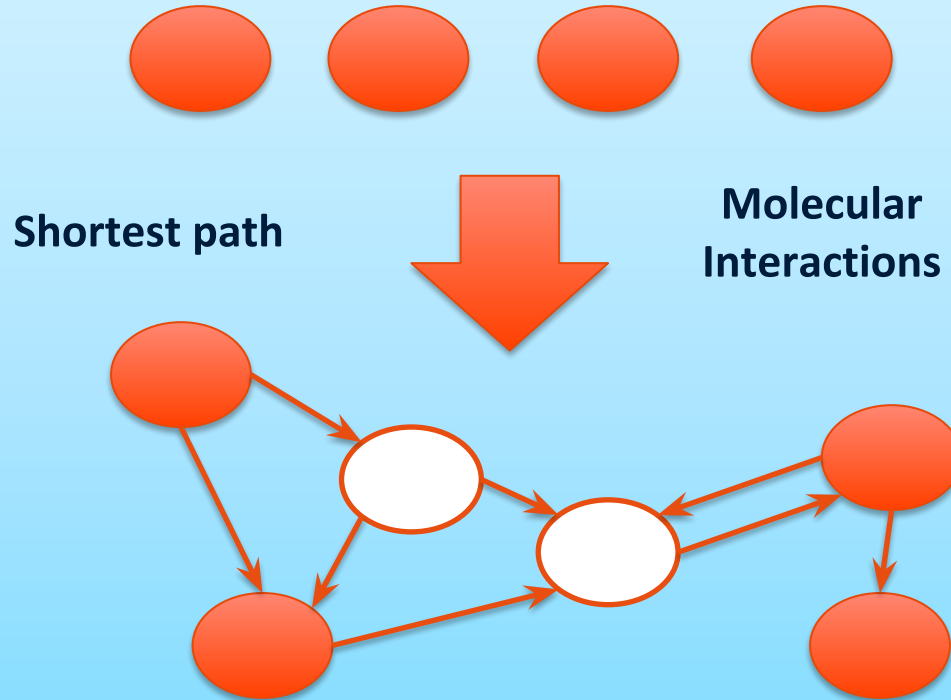
[1]

Sparseness of Data

Amount of data

Identifier mapping

Network approach [1]



Aim of Network Approach

- Directed network of metabolites from pathway knowledge bases

Aim of Network Approach

- Directed network of metabolites from pathway knowledge bases
- Calculate sub-network between active metabolites

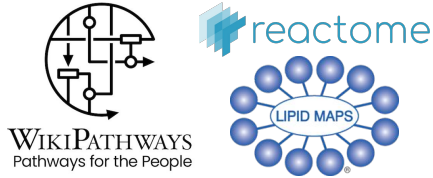
Aim of Network Approach

- Directed network of metabolites from pathway knowledge bases
- Calculate sub-network between active metabolites
- Visualise directed paths

Aim of Network Approach

- Directed network of metabolites from pathway knowledge bases
- Calculate sub-network between active metabolites
- Visualise directed paths
- Interpret metabolomics datasets

WORKFLOW



Metabolic interactions

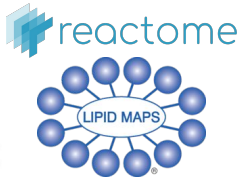
Directed metabolic conversions

3 pathway models

Homo sapiens (Human)

WikiPathways [1] RDF [2]

WORKFLOW



Metabolic
interactions



Graph
Database

Directed metabolic conversions

3 pathway models

Homo sapiens (Human)

WikiPathways [1] RDF [2]

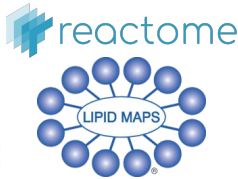
Store conversions

Cypher queries

Several available
algorithms

Neo4j [3]

WORKFLOW



Metabolic
interactions



Graph
Database



Visualisation

Directed metabolic conversions
3 pathway models
Homo sapiens (Human)

WikiPathways [1] RDF [2]

Store conversions
Cypher queries
Several available
algorithms

Neo4j [3]

Omics data visualisation
Network extendable
Automatisable (REST-API)

Cytoscape [4]

WORKFLOW



WikiPathways RDF

Metabolomics
data

Directed metabolic conversions

3 pathway models

Homo sapiens (Human)

WikiPathways [1] RDF [2]

WORKFLOW



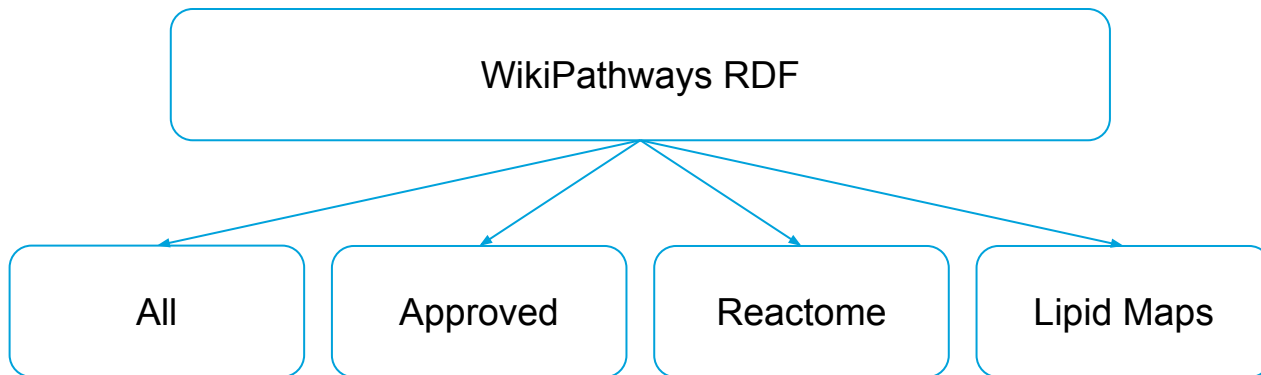
Metabolomics
data

Directed metabolic conversions

3 pathway models

Homo sapiens (Human)

WikiPathways [1] RDF [2]



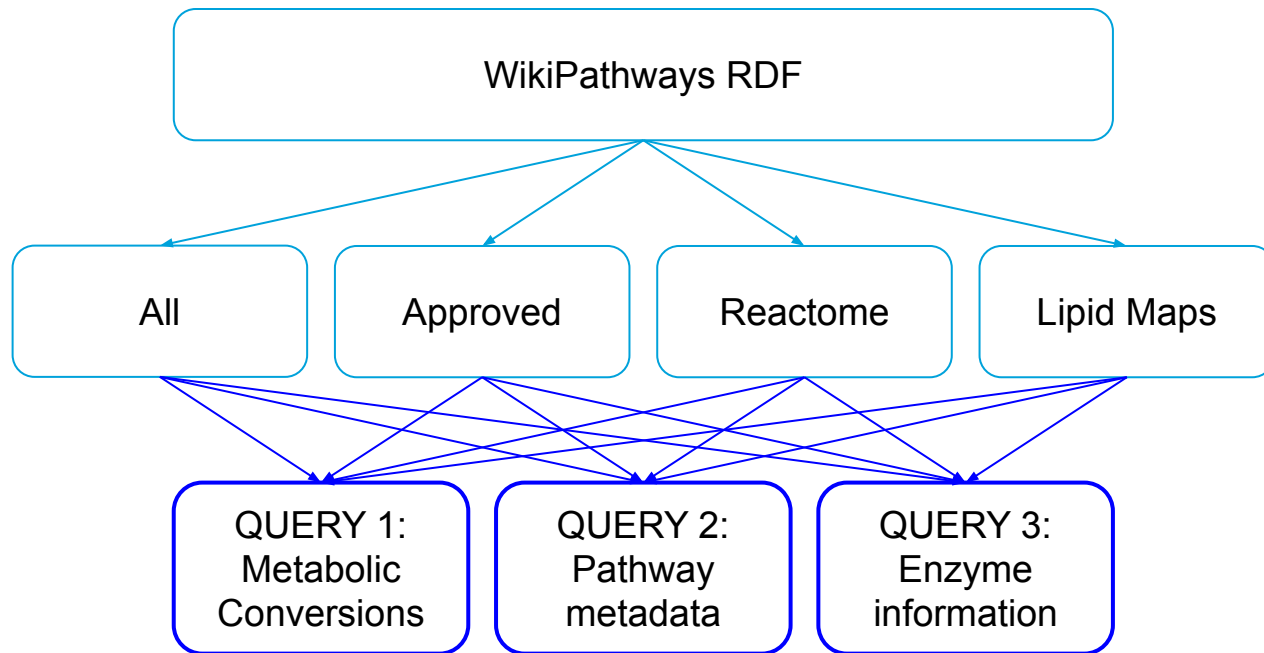
WORKFLOW



Metabolomics
data

Directed metabolic conversions
3 pathway models
Homo sapiens (Human)

WikiPathways [1] RDF [2]



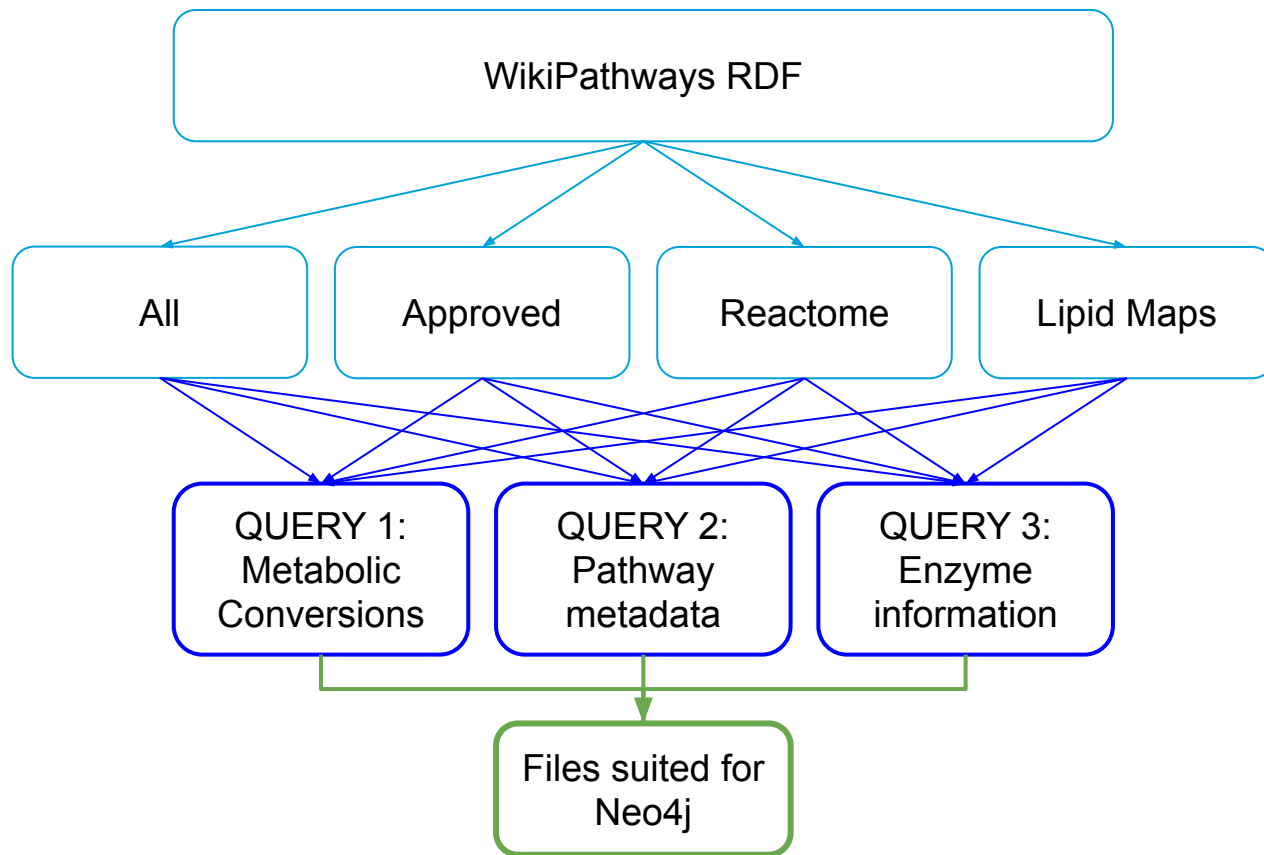
WORKFLOW



Metabolomics data

Directed metabolic conversions
3 pathway models
Homo sapiens (Human)

WikiPathways [1] RDF [2]



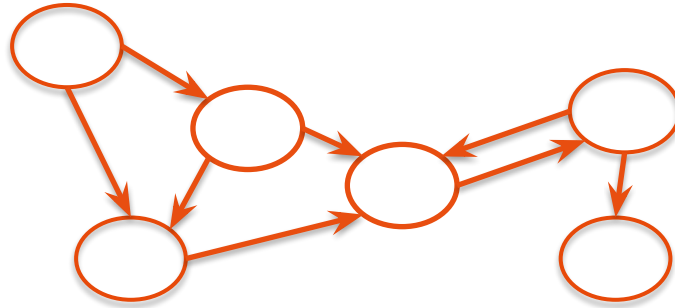
WORKFLOW



Graph
Database

Store conversions
Cypher queries
Several available
algorithms

Neo4j [3]



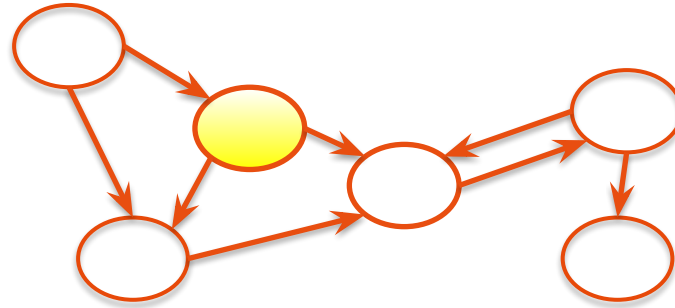
WORKFLOW



Graph Database

Store conversions
Cypher queries
Several available
algorithms

Neo4j [3]



Property	Value	Identifier
Source	L-Tyrosine	Q188017 (Wikidata)
Target	L-Dopa	Q300989 (Wikidata)
Occurrence	2	WP4220, WP4156
Enzyme	Tyrosine 3-mono oxygenase	P07101 (Uniprot-TrEMBL)

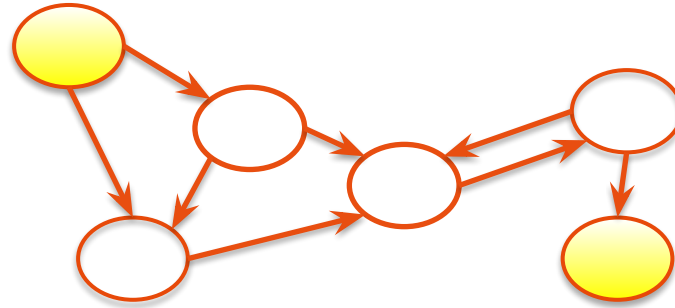
WORKFLOW



Graph
Database

Store conversions
Cypher queries
Several available
algorithms

Neo4j [3]



Shortest Path

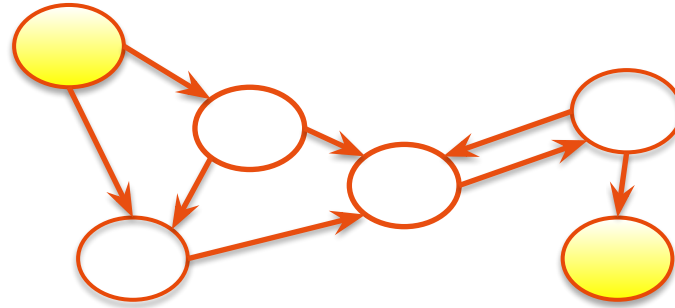
WORKFLOW



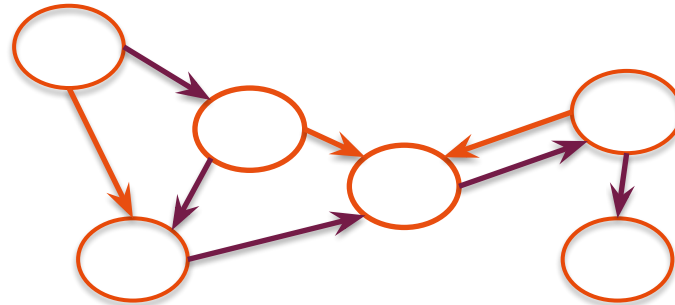
Graph
Database

Store conversions
Cypher queries
Several available
algorithms

Neo4j [3]



Shortest Path



Steps:

5

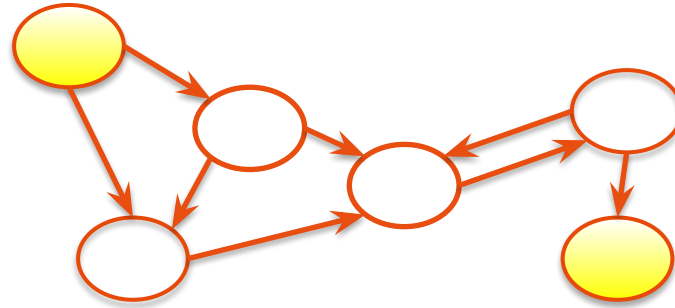
WORKFLOW



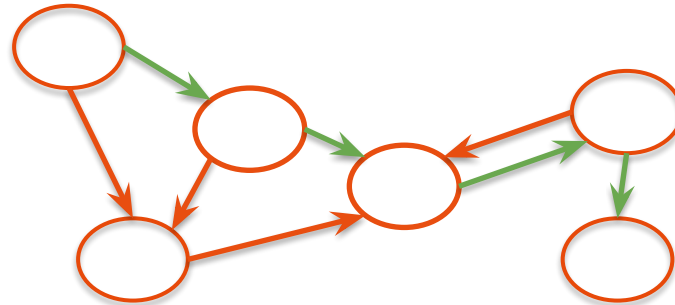
Graph
Database

Store conversions
Cypher queries
Several available
algorithms

Neo4j [3]



Shortest Path



Steps:

4

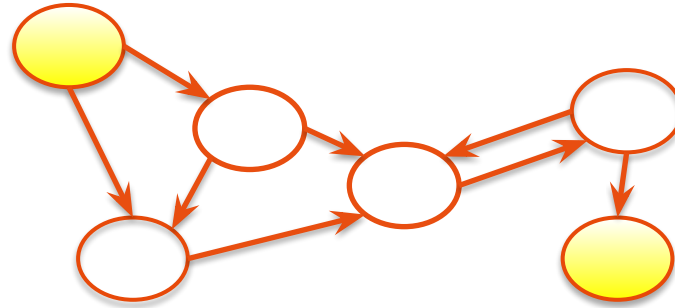
WORKFLOW



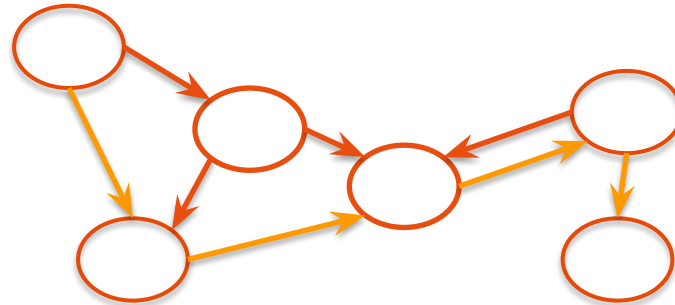
Graph
Database

Store conversions
Cypher queries
Several available
algorithms

Neo4j [3]



Shortest Path



Steps:

4

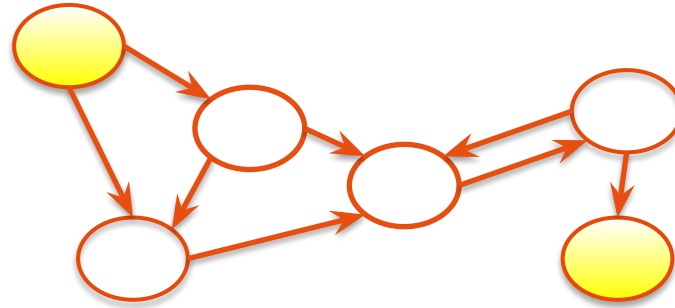
WORKFLOW



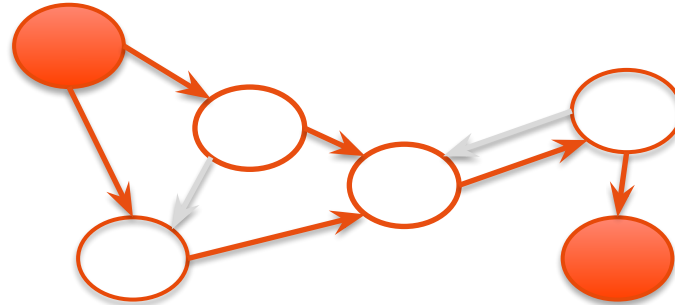
Graph
Database

Store conversions
Cypher queries
Several available
algorithms

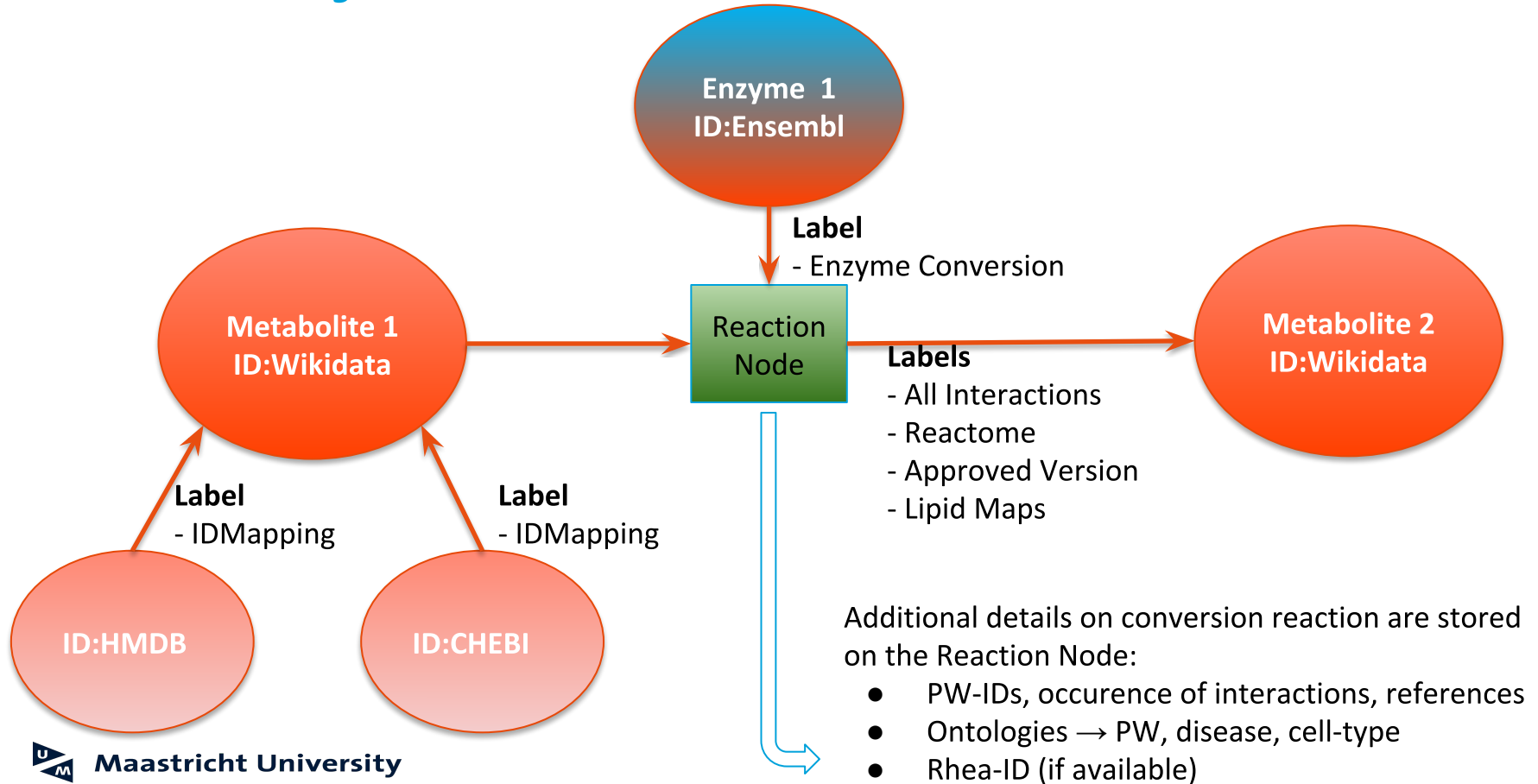
Neo4j [3]



Shortest Path



Neo4j database details: “final” model



WORKFLOW



Visualisation

Omics data visualisation
Network extendable
Automatisable (REST-API)

Cytoscape [4]

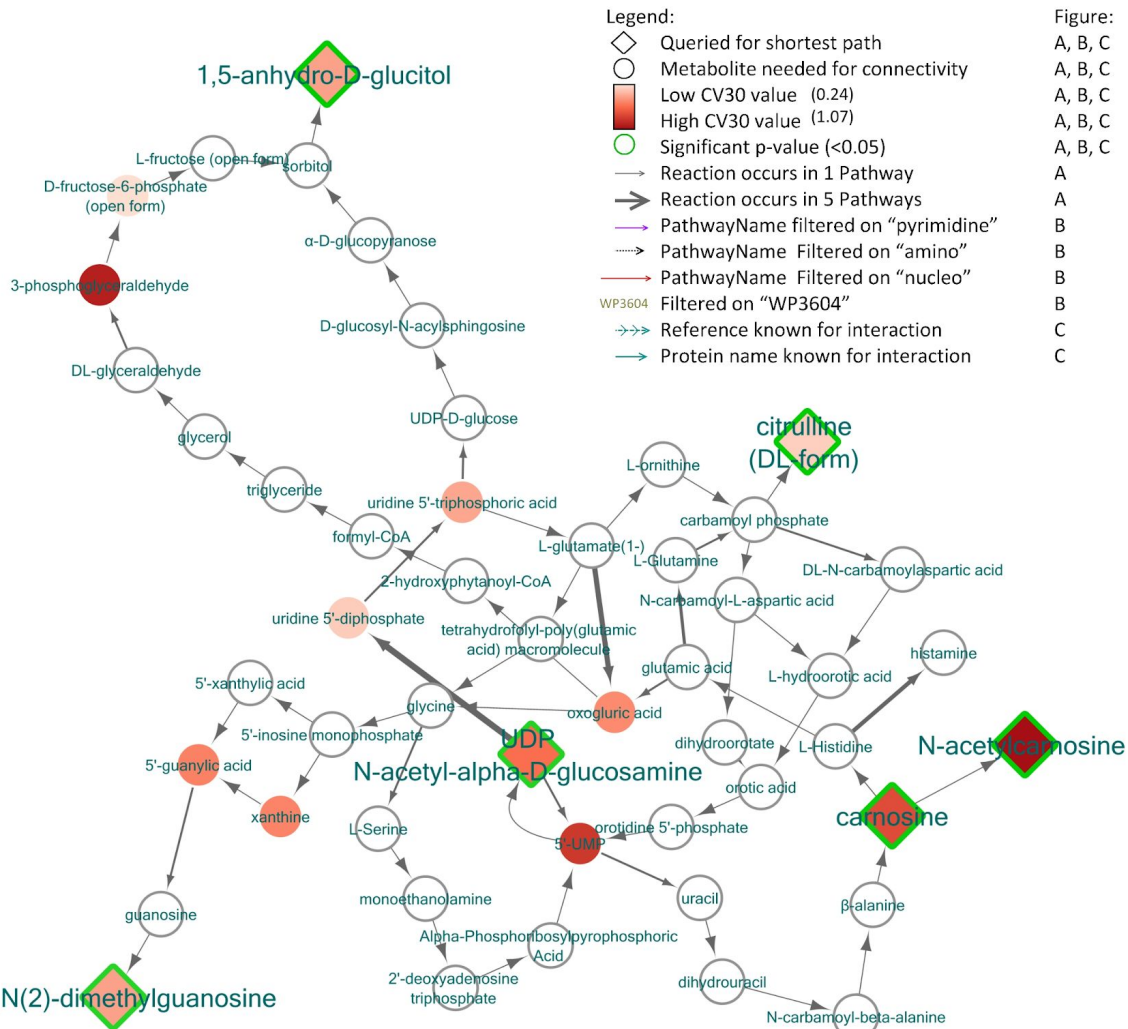
1. Used in biological network community
2. Free to use, source code available online
3. Connectable through CyNeo4j app
4. Automatisable through REST-API, from R and Python
5. Possibility to integrate OMICS data
6. Extendable with CyTargetLinker app (TF, drug-target, SNPs)

Datasets used to test visualisation

Name	Technique	Matrix	Metabolites of interest detected with:
MTBLS265 [1]	LC-MS	Blood	CV-30, p-value
MTBLS404 [2]	LC-HRMS	Urine	Spearman rank correlation test, Orthogonal partial least-squares (OPLS)
Rist <i>et al.</i> [3]	(GCxGC)-MS, targeted GC-MS and LC-MS/MS, 1H-NMR	Blood & Urine	Support Vector Machines (SVM, linear kernel), generalized linear model net (glmnet), Partial least squares (PLS)

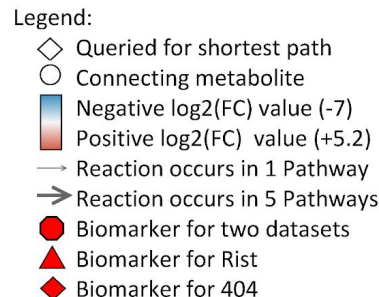
Name	Male/Female	Age range (y)	# Identified metabolites	# age related	Compounds linked through shortest path
MTBLS265	14/16	Young: 29 ± 4 Old: 81 ± 7	126	14	6
MTBLS404	100/83	40.9 ± 10.3	120	30	14
Rist <i>et al.</i>	172/129	47.5 ± 17.1	400 (plasma), >500 (urine)	8 + 6	6 + 4

A





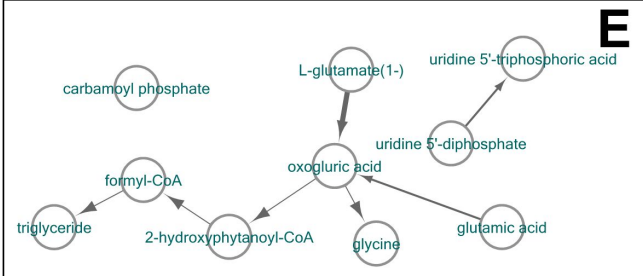
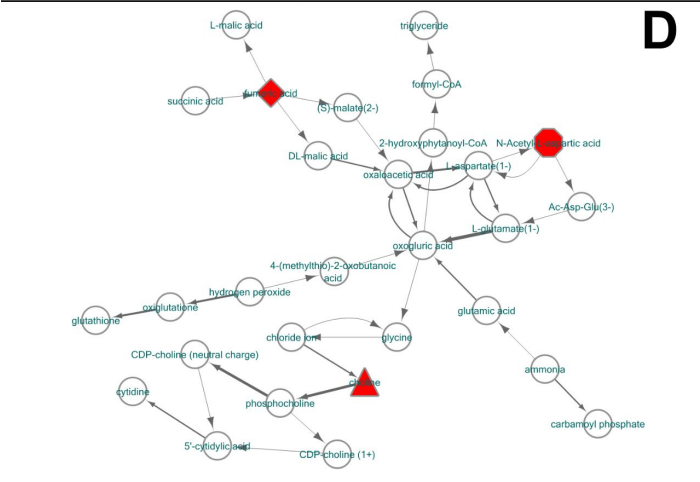
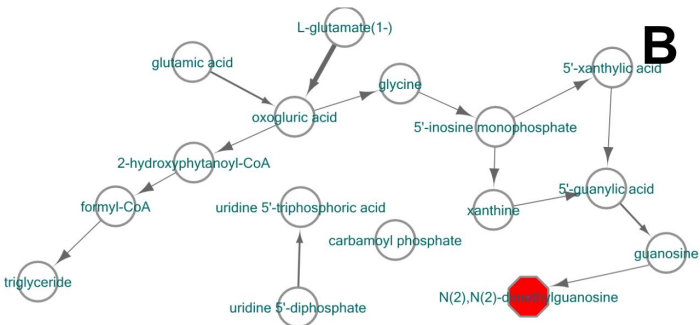
- A, B, C
A, B, C
A, B, C
A, B, C
A
A
B
B
B
B
C
C



A
A - E
A
A
A - E
A - E
B, C
C, D
D



Comparing all



◇ Queried for shortest path	A
○ Connecting metabolite	A - E
█ Negative log2(FC) value (-7)	A
█ Positive log2(FC) value (+5.2)	A
→ Reaction occurs in 1 Pathway	A - E
➔ Reaction occurs in 5 Pathways	A - E
● Biomarker for two datasets	B, C
▲ Biomarker for Rist	C, D
◆ Biomarker for 404	D

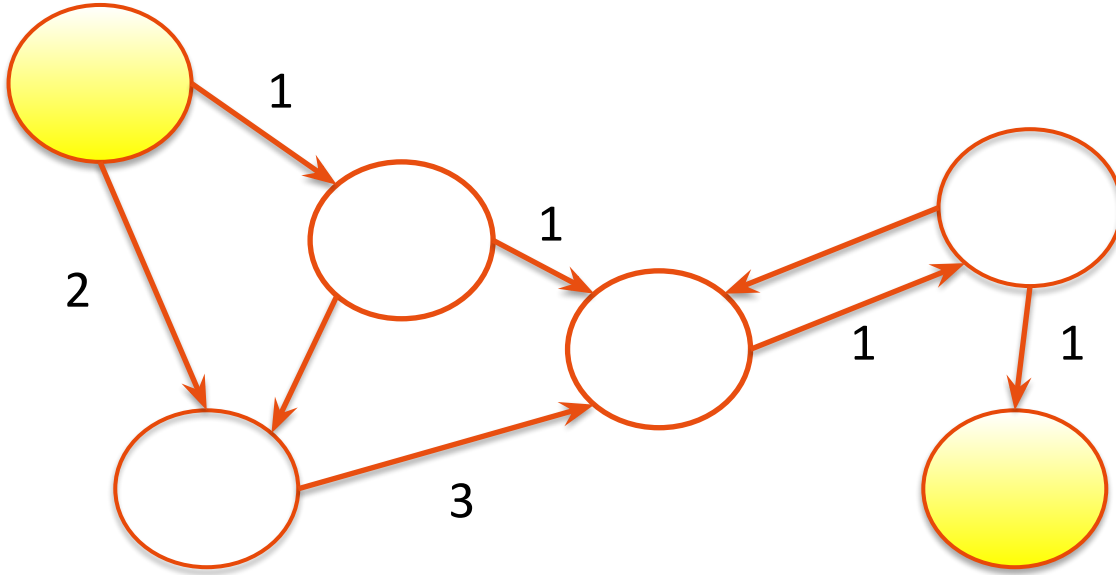
B. MTBLS265(blood) vs MTBLS404(urine);

C. MTBLS265(blood) vs Rist(blood and urine); Only overlaps on 1 blood metabolite

D. MTBLS404(urine) vs Rist(blood and urine); Only overlaps on 1 urine metabolite

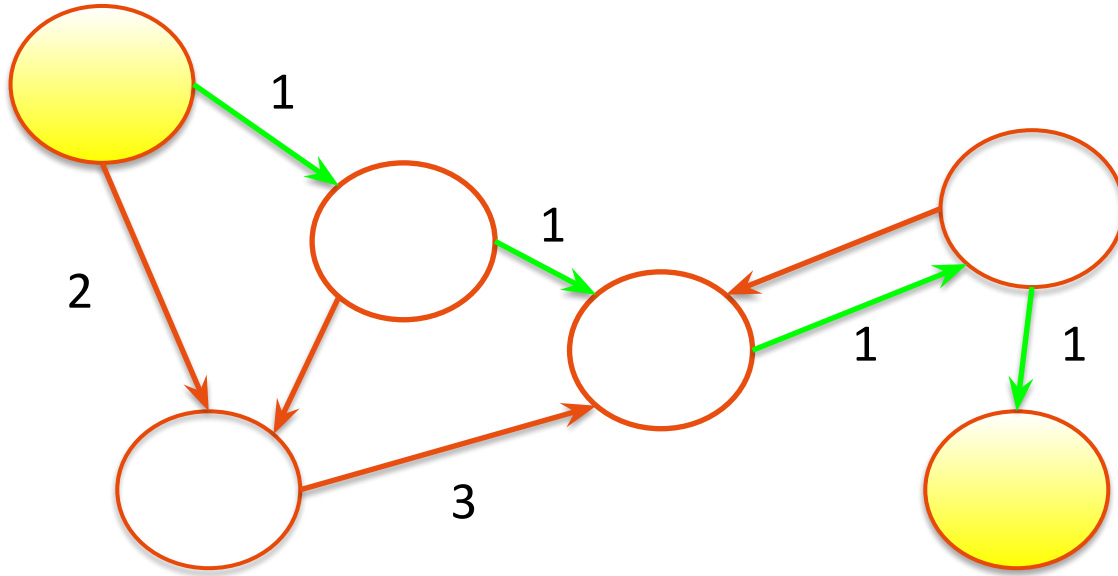
E. Comparison of all three datasets (MTBLS265, MTBLS404 and Rist).

Including Kinetics



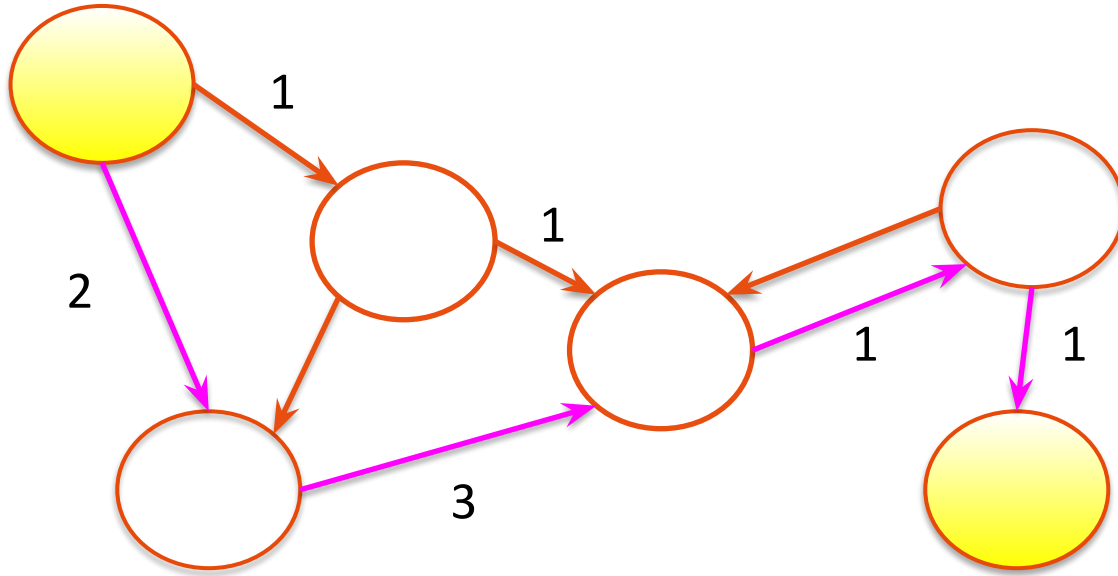
- Weighted directed network
- “Fastest” biological track (!= most relevant)
- Finetunes final model

Including Kinetics



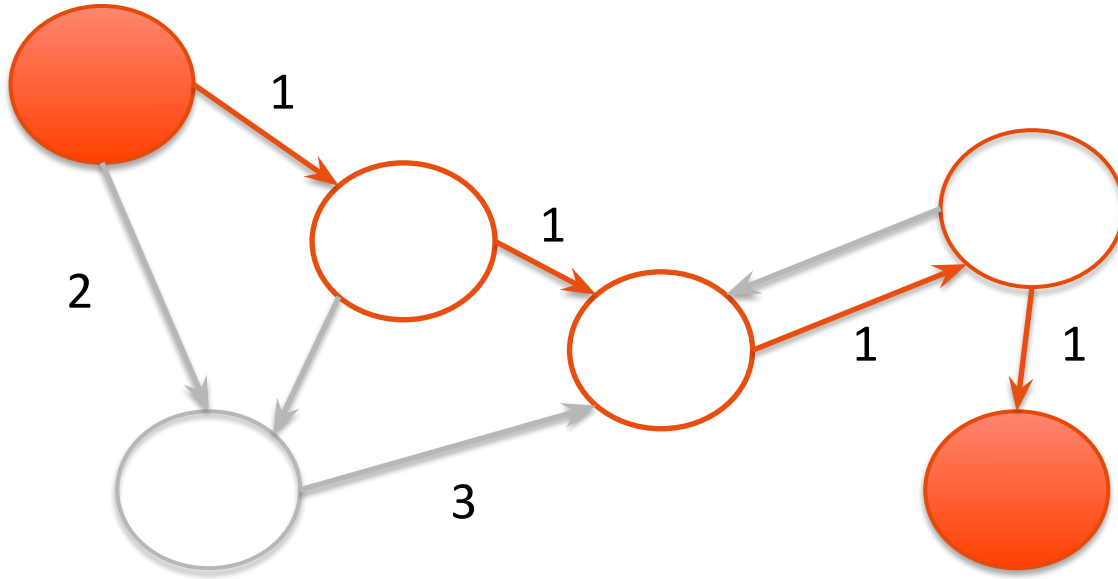
- Green:
- $1+1+1+1 = 4$

Including Kinetics



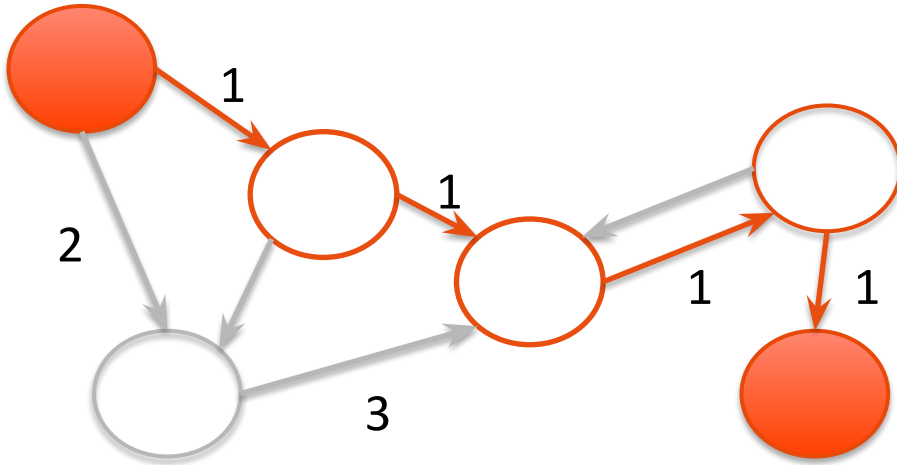
- Green:
- $1+1+1+1 = 4$
- Purple:
- $2+3+1+1 = 7$
- Green path is shorter than purple path

Including Kinetics

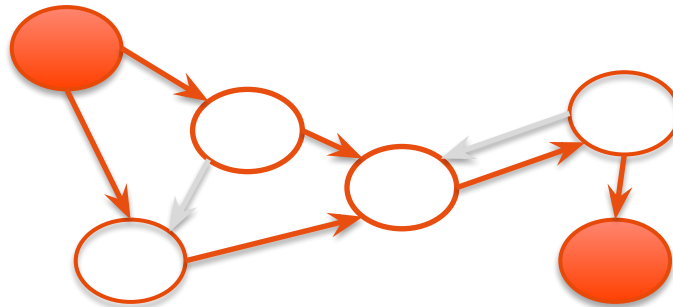


- Green path is actual shortest path
- More interactions are left from final model
- Metabolite bottom left is also not present in final model

Including Kinetics



- With kinetic info



- Without kinetic info

Using machine learning to learn from experts?

- Under certain conditions (diseased state, treatment), different paths could be more biologically relevant.
 - How should these paths be calculated?
 - How do we keep this process transparent and reproducible?

Using machine learning to learn from experts?

- Under certain conditions (diseased state, treatment), different paths could be more biologically relevant.
 - How should these paths be calculated?
 - How do we keep this process transparent and reproducible?
- Dynamics within biological networks may change over time.
 - How to model this behavior in (static) graph/pathway databases?

Using machine learning to learn from experts?

- Under certain conditions (diseased state, treatment), different paths could be more biologically relevant.
 - How should these paths be calculated?
 - How do we keep this process transparent and reproducible?
- Dynamics within biological networks may change over time.
 - How to model this behavior in (static) graph/pathway databases?
- Irrelevant interactions may be skipped by users.
 - How do we capture this information, to improve existing models?

Take home messages

Calculation of directed subnetwork connecting active metabolites is possible with presented workflow

Pathway analysis of (untargeted) metabolomics data becomes easier, if you use network approaches and graph databases

Workflow is easy for novel users, and adaptable for skilled users, integratable with the data you have and customizable for the analysis you need

Work ahead

Add more pathway knowledge bases → larger coverage of metabolic reactions

Test app in larger audience → first tests have been run

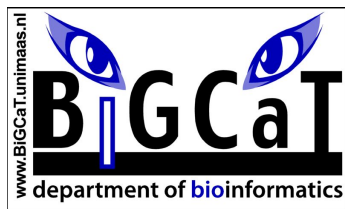
Create complete R-Markdown script for bulk approach



Acknowledgements, questions, discussion

- Martina Kutmon
- Jonathan Mélius
- Georg Summer
- Chris T Evelo
- Egon L Willighagen

- WikiPathways: team and curators
- Reactome
- Lipid Maps
- Metabolights
- Elixir



- Twitter: @SMaLLCaT4Sci and @BiGCaT_UM
- Blog: <http://smallcats4science.blogspot.nl>
- ORCID: 0000-0001-8449-1318

Find the original poster @ DOI: [10.6084/m9.figshare.5234851.v1](https://doi.org/10.6084/m9.figshare.5234851.v1)



Maastricht University

Find this presentation @ DOI:

2nd UM Metabolomics Meeting

Date: 12-12-2019

- Labtour
- Talks by Nightingale and BioCrates
- Talks by several researchers from UM

More details coming soon through Metabolomics mailing list!

Email to :

- denise.slenter or
- egon.willighagen
@maastrichtuniversity to be added to the list

We're hiring:

3 PostDocs

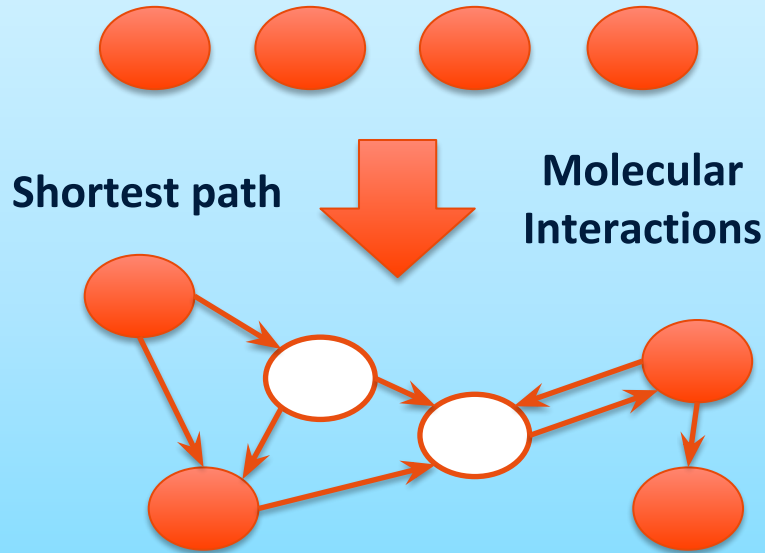
Contact: Freddie Ehrhart (rare diseases)
friederike.ehrhart@maastrichtuniversity.nl

Contact: Egon Willighagen (nanomaterials+FAIR data)
egon.willighagen@maastrichtuniversity.nl

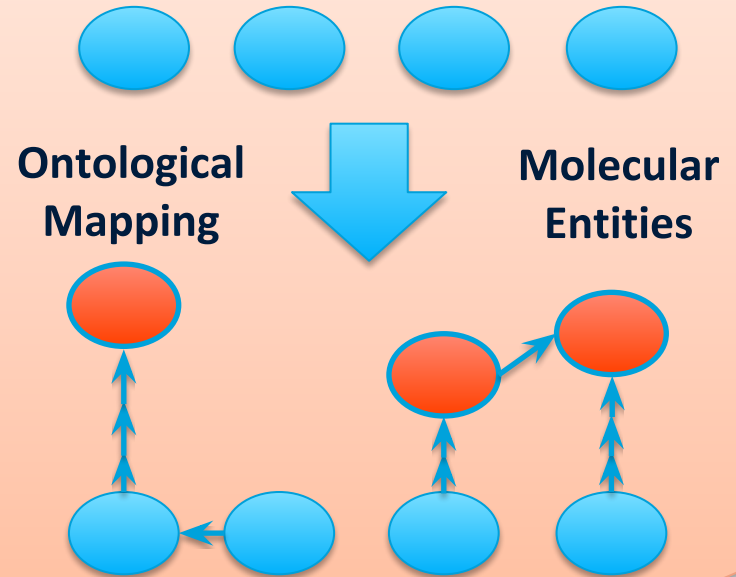
Contact: Susan Steinbusch-Coort (nutrition)
susan.coort@maastrichtuniversity.nl

Two approaches

Network approach [1]

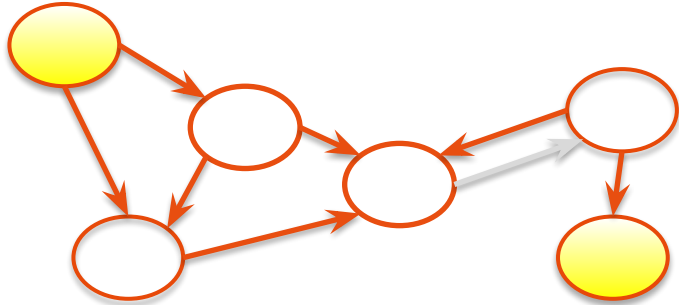


Ontological approach [2]



Biological role					
Electron donor/receiver.		Energy donor/receiver.		Miscellaneous, relevant for various metabolic reactions.	
Identifier	Name	Identifier	Name	Identifier	Name
Q5203615	O ₂	Q80863	ATP	Q307434	S-adenosyl-L-homocysteine
Q506710	H ⁺	Q185253	ADP	Q201312	S-adenosyl-L-methionine
Q20856948	Na ⁺ (redirected to Q3154110)	Q318369	AMP	Q407635	Coenzyme-A
Q3154110	Na ⁺	Q422582	GDP	Q715317	Acetyl coenzyme a
Q283	H ₂ O	Q392227	GTP	Side metabolites	
Q1997	CO ₂	Q26987754	NADP ⁺		
Q177811	PO ₄ 3-	Q26841327	NADPH		
Q411092	Pyrophosphoric acid	Q26987253	NAD ⁺		
Q190901	ammonium cation	Q26987453	NADH		
		Q27102690	FADH ₂		

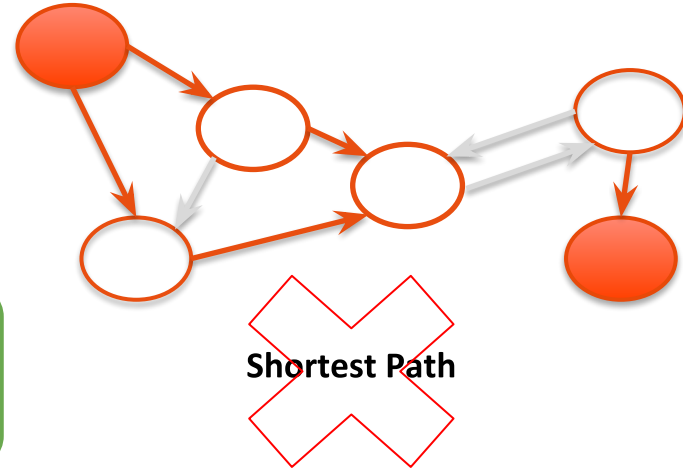
Additional algorithms



Graph
Database

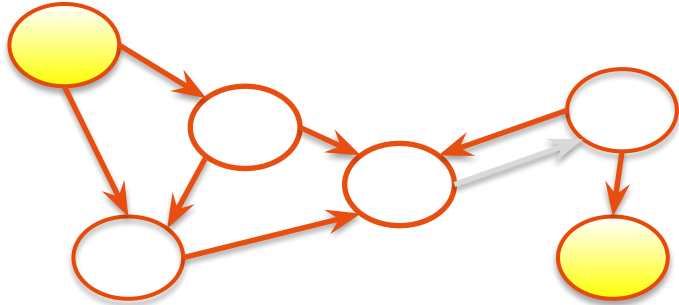
Store conversions
Cypher queries
Several available
algorithms

Neo4j [3]



Nearest Neighbour

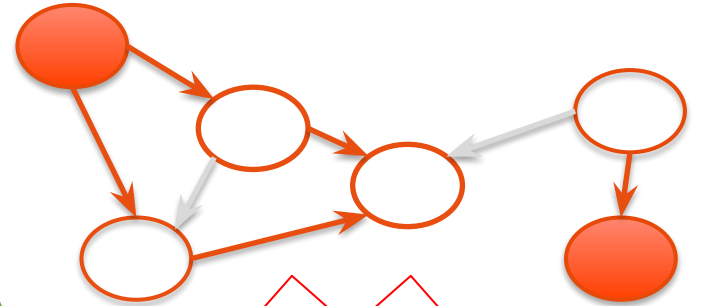
Additional algorithms



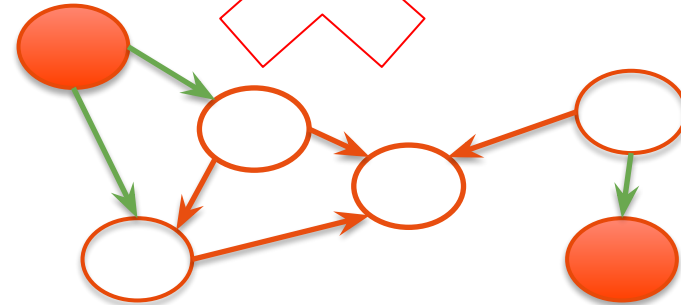
Graph
Database

Store conversions
Cypher queries
Several available
algorithms

Neo4j [3]

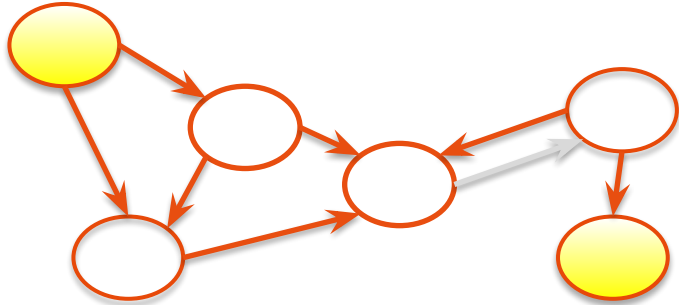


Shortest Path



Nearest Neighbour

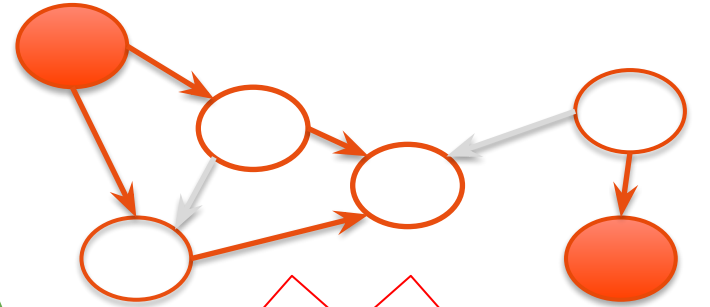
Additional algorithms



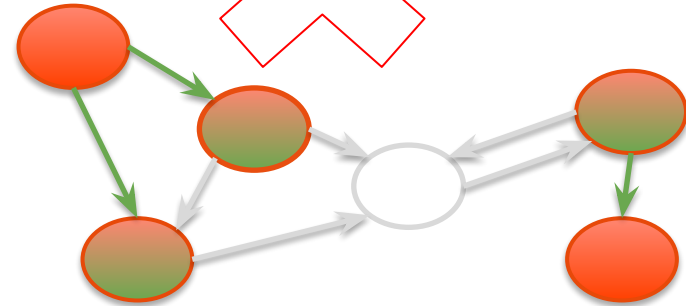
Graph
Database

Store conversions
Cypher queries
Several available
algorithms

Neo4j [3]



Shortest Path



Nearest Neighbour