Visualizing metabolomics data in directed biological networks

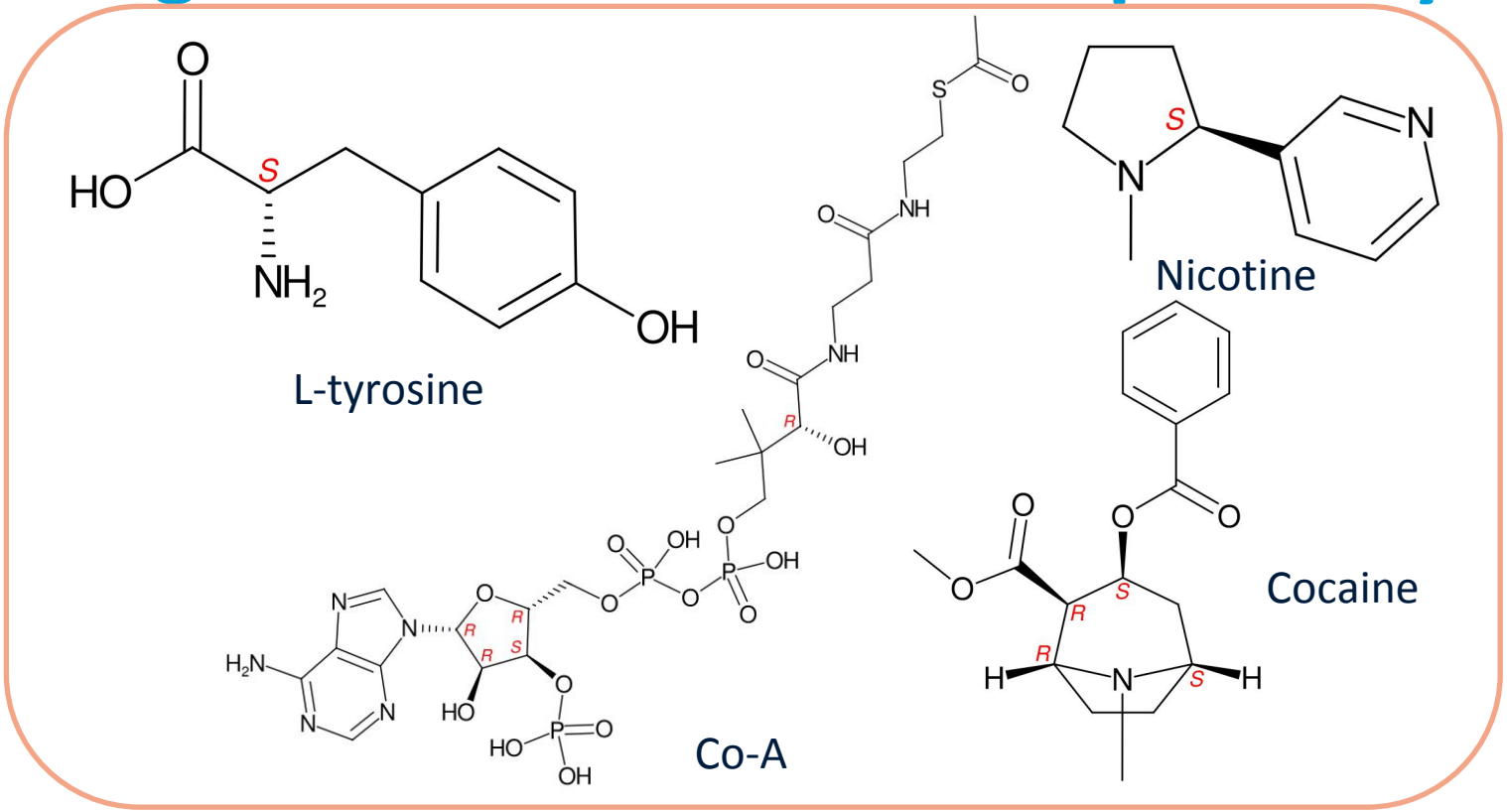
Denise Slenter, Chris Evelo, Egon Willighagen

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

2019-09-25 Master Systems Biology 2019-2020

Maastricht University

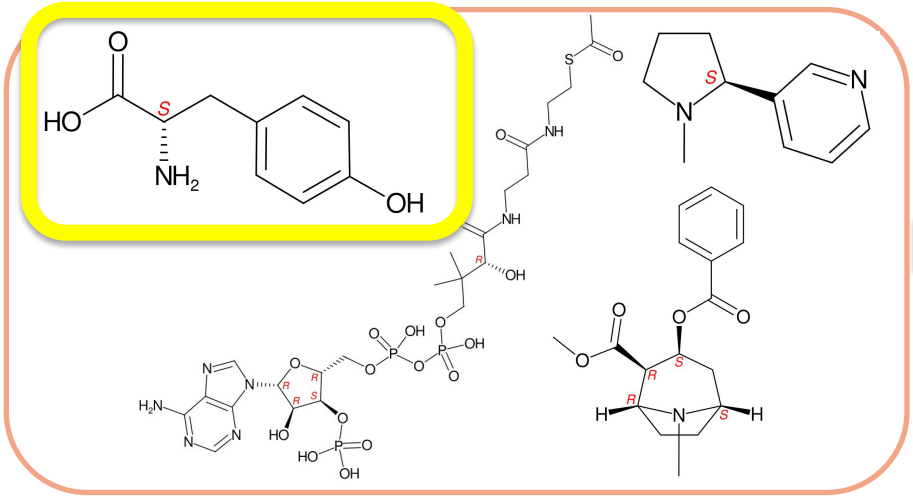
Linking metabolomics data to pathways…

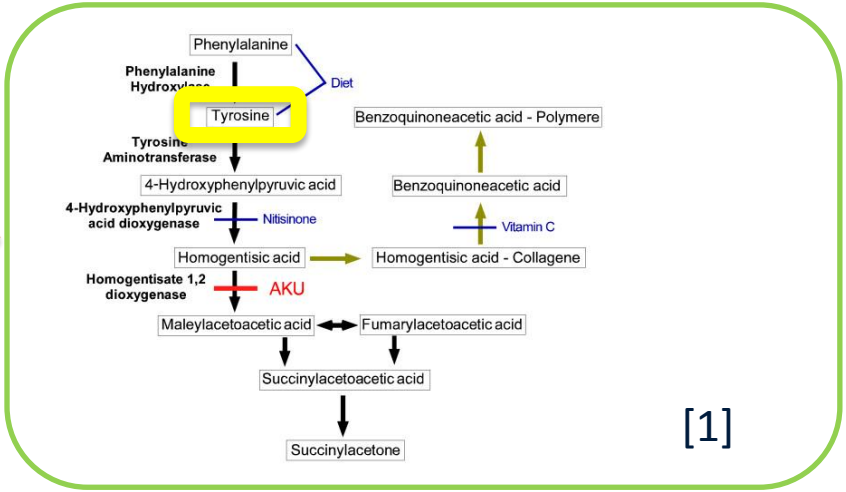


Maastricht University

Structures obtained from Wikidata [2019-06-16] with CDK depict: https://github.com/cdk/depict
Willighagen *et al*. (2017) DOI: 10.1186/s13321-017-0220-4

Linking metabolomics data to pathways…





[1]

Sparseness of Data

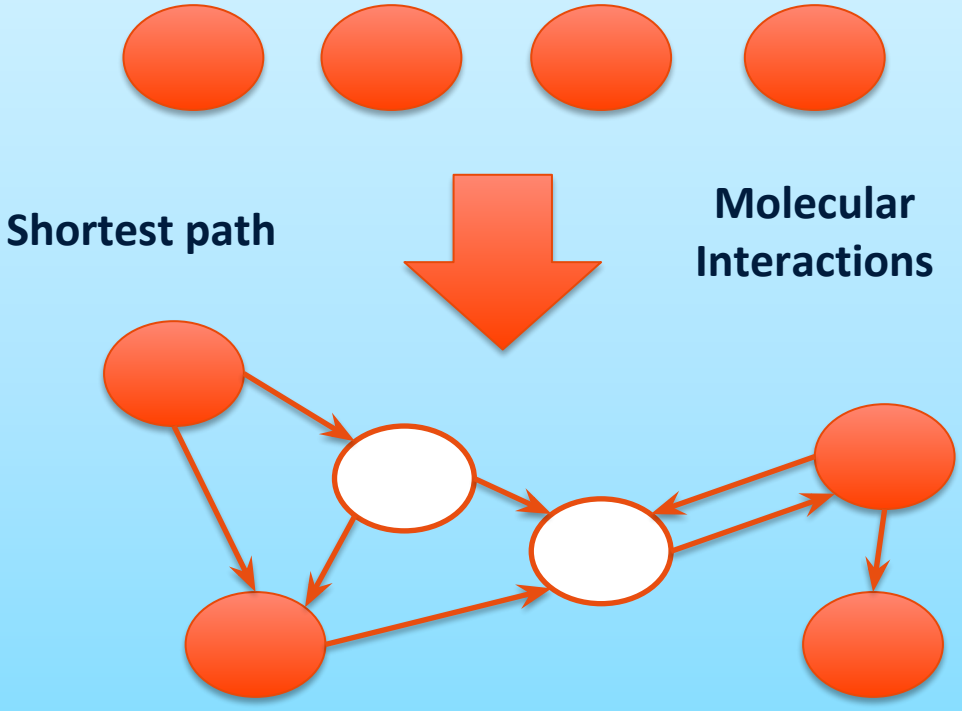
Amount of data

Identifier mapping

Maastricht University

[1] Lindner, et al. BMC Ophthalmology (2014)

Network approach [1]



Maastricht University

[1] DOI: 10.6084/m9.figshare.5234851.v1

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

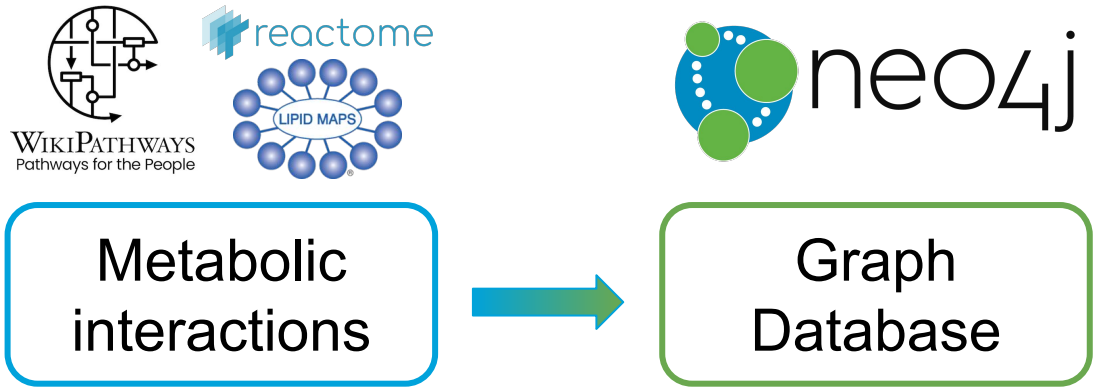
Directed metabolic conversions
3 pathway models
Homo sapiens (Human)

WikiPathways [1] RDF [2]

Maastricht University

[1]WikiPathways: https://www.wikipathways.org, Slenter *et al*. (2018), DOI: 10.1093/nar/gkx1064
[2] SPARQL endpoint: http://sparql.wikipathways.org/, Waagmeester *et al.* (2016) DOI: 10.1371/journal.pcbi.1004989
[3] Neo4j: https://neo4j.com/ [4] Cytoscape: http://cytoscape.org/, Shannon *et al.* (2003) DOI: 10.1101/gr.1239303

WORKFLOW



Directed metabolic conversions
3 pathway models
Homo sapiens (Human)

WikiPathways [1] RDF [2]

Store conversions
Cypher queries
Several available algorithms

Neo4j [3]

[1]WikiPathways: https://www.wikipathways.org, Slenter *et al*. (2018), DOI: 10.1093/nar/gkx1064
[2] SPARQL endpoint: http://sparql.wikipathways.org/, Waagmeester *et al.* (2016) DOI: 10.1371/journal.pcbi.1004989
[3] Neo4j: https://neo4j.com/ [4] Cytoscape: http://cytoscape.org/, Shannon *et al.* (2003) DOI: 10.1101/gr.1239303

WORKFLOW

Metabolic interactions

Directed metabolic conversions
3 pathway models
Homo sapiens (Human)

WikiPathways [1] RDF [2]

Graph Database

Store conversions
Cypher queries
Several available algorithms

Neo4j [3]

Visualisation

Omics data visualisation
Network extendable
Automatisable (REST-API)

Cytoscape [4]

[1]WikiPathways: https://www.wikipathways.org, Slenter *et al*. (2018), DOI: 10.1093/nar/gkx1064
[2] SPARQL endpoint: http://sparql.wikipathways.org/, Waagmeester *et al.* (2016) DOI: 10.1371/journal.pcbi.1004989
[3] Neo4j: https://neo4j.com/ [4] Cytoscape: http://cytoscape.org/, Shannon *et al.* (2003) DOI: 10.1101/gr.1239303

Maastricht University

WORKFLOW

Metabolomics data

Directed metabolic conversions
3 pathway models
Homo sapiens (Human)

WikiPathways [1] RDF [2]

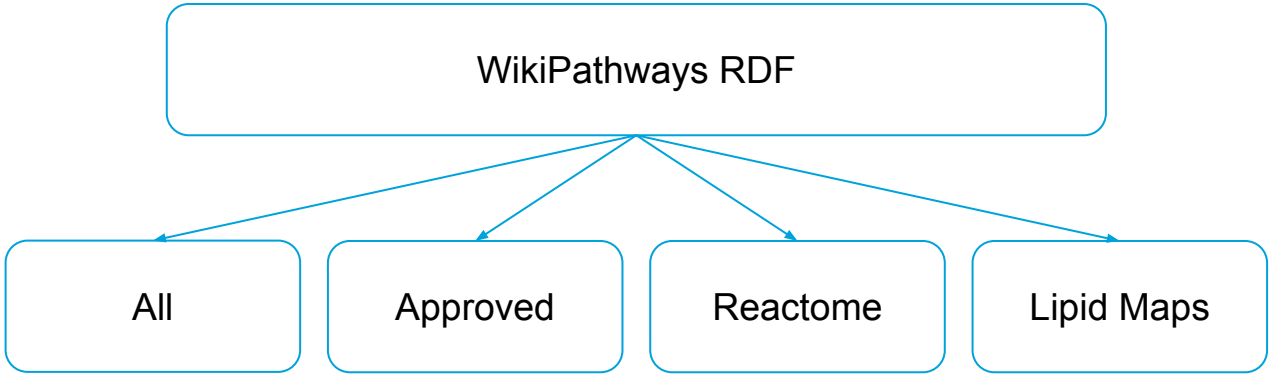
WikiPathways RDF

Maastricht University

[1]WikiPathways: https://www.wikipathways.org, Slenter *et al*. (2018), DOI: 10.1093/nar/gkx1064
[2] SPARQL endpoint: http://sparql.wikipathways.org/, Waagmeester *et al.* (2016) DOI: 10.1371/journal.pcbi.1004989

WORKFLOW

Metabolomics data



Directed metabolic conversions
3 pathway models
Homo sapiens (Human)

WikiPathways [1] RDF [2]

Maastricht University

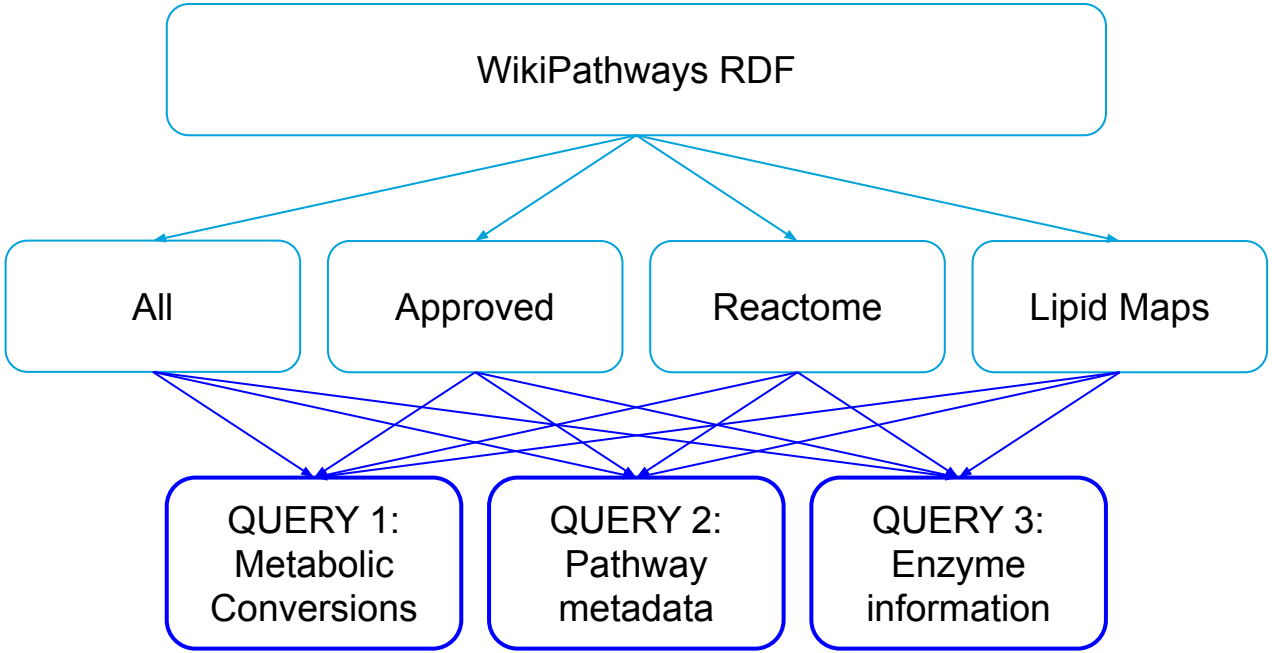
[1]WikiPathways: https://www.wikipathways.org, Slenter *et al*. (2018), DOI: 10.1093/nar/gkx1064
[2] SPARQL endpoint: http://sparql.wikipathways.org/, Waagmeester *et al.* (2016) DOI: 10.1371/journal.pcbi.1004989

WORKFLOW

Metabolomics data

Directed metabolic conversions
3 pathway models
Homo sapiens (Human)

WikiPathways [1] RDF [2]



Maastricht University

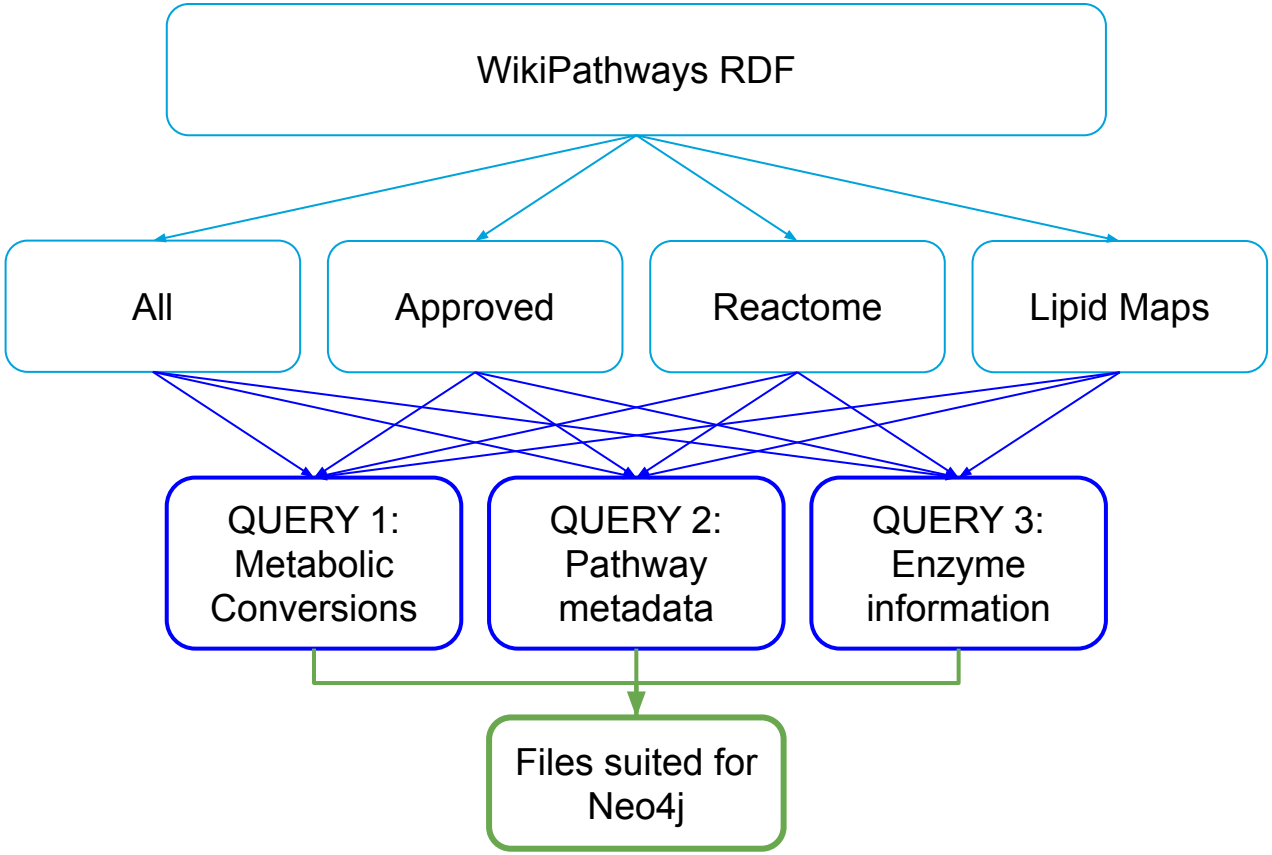
[1]WikiPathways: https://www.wikipathways.org, Slenter *et al*. (2018), DOI: 10.1093/nar/gkx1064
[2] SPARQL endpoint: http://sparql.wikipathways.org/, Waagmeester *et al.* (2016) DOI: 10.1371/journal.pcbi.1004989

WORKFLOW

Metabolomics data

Directed metabolic conversions
3 pathway models
Homo sapiens (Human)

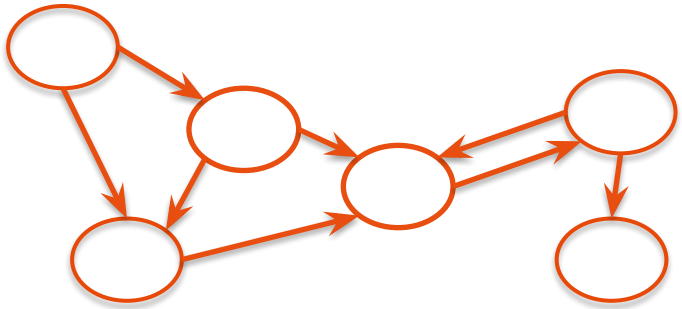
WikiPathways [1] RDF [2]



Maastricht University

[1]WikiPathways: https://www.wikipathways.org, Slenter *et al*. (2018), DOI: 10.1093/nar/gkx1064
[2] SPARQL endpoint: http://sparql.wikipathways.org/, Waagmeester *et al.* (2016) DOI: 10.1371/journal.pcbi.1004989

WORKFLOW



Graph Database

Store conversions
Cypher queries
Several available algorithms

Neo4j [3]

Maastricht University

[3] Neo4j: https://neo4j.com/ [4] Cytoscape: http://cytoscape.org/, Shannon *et al.* (2003) DOI: 10.1101/gr.1239303

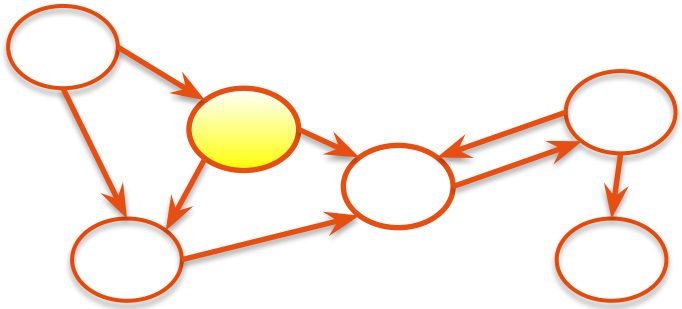
WORKFLOW

neo4j

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)

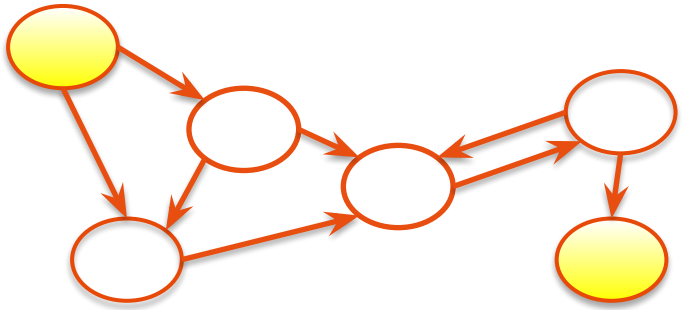
[3] Neo4j: https://neo4j.com/ [4] Cytoscape: http://cytoscape.org/, Shannon *et al.* (2003) DOI: 10.1101/gr.1239303

WORKFLOW

Graph Database

Store conversions
Cypher queries
Several available algorithms

Neo4j [3]



Shortest Path

Maastricht University

[3] Neo4j: https://neo4j.com/ [4] Cytoscape: http://cytoscape.org/, Shannon *et al.* (2003) DOI: 10.1101/gr.1239303

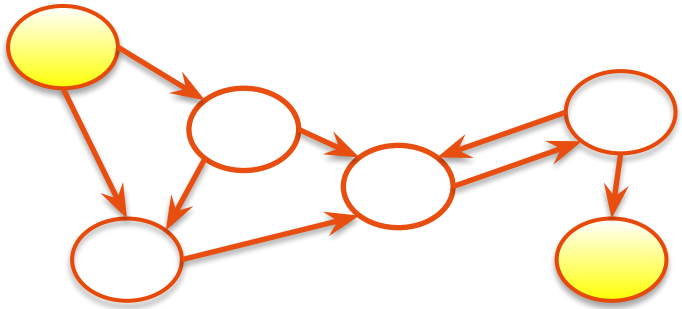
WORKFLOW

neo4j

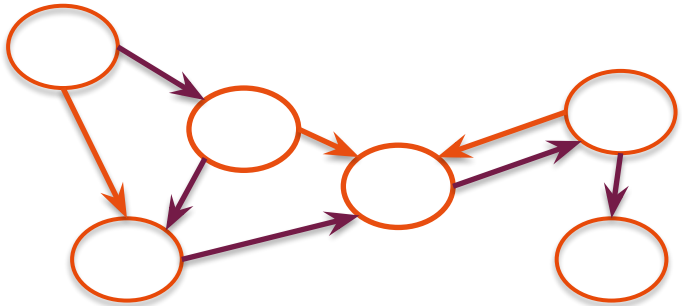
Graph Database

Store conversions
Cypher queries
Several available algorithms

Neo4j [3]



Shortest Path



Steps:

5

Maastricht University

[3] Neo4j: https://neo4j.com/ [4] Cytoscape: http://cytoscape.org/, Shannon *et al.* (2003) DOI: 10.1101/gr.1239303

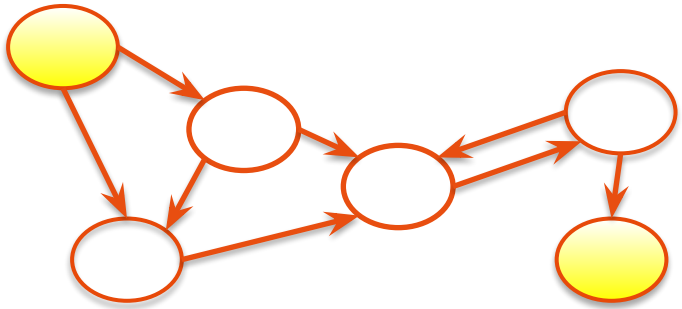
WORKFLOW

neo4j

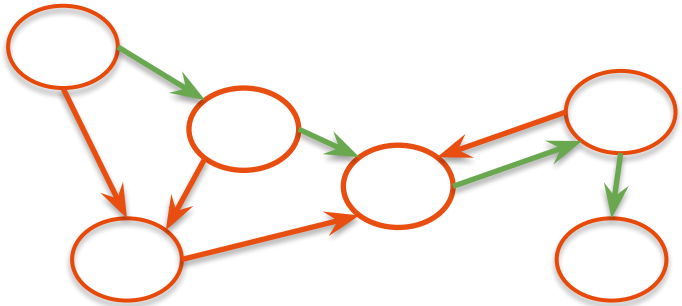
Graph Database

Store conversions
Cypher queries
Several available algorithms

Neo4j [3]



Shortest Path



Steps:

4

Maastricht University

[3] Neo4j: https://neo4j.com/ [4] Cytoscape: http://cytoscape.org/, Shannon *et al.* (2003) DOI: 10.1101/gr.1239303

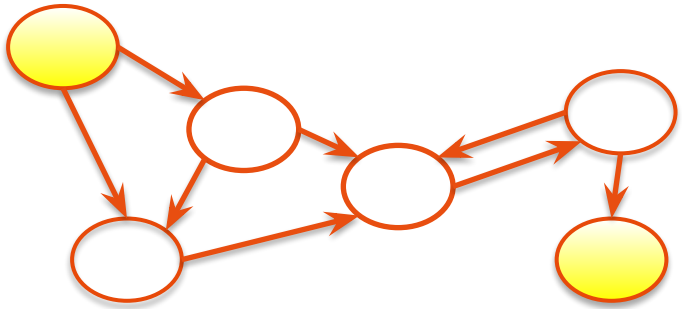
WORKFLOW

neo4j

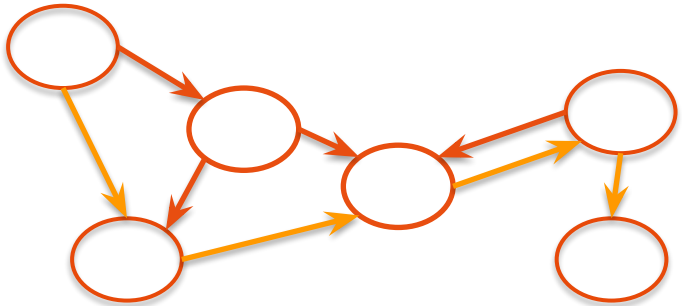
Graph Database

Store conversions
Cypher queries
Several available algorithms

Neo4j [3]



Shortest Path



Steps:

4

Maastricht University

[3] Neo4j: https://neo4j.com/ [4] Cytoscape: http://cytoscape.org/, Shannon *et al.* (2003) DOI: 10.1101/gr.1239303

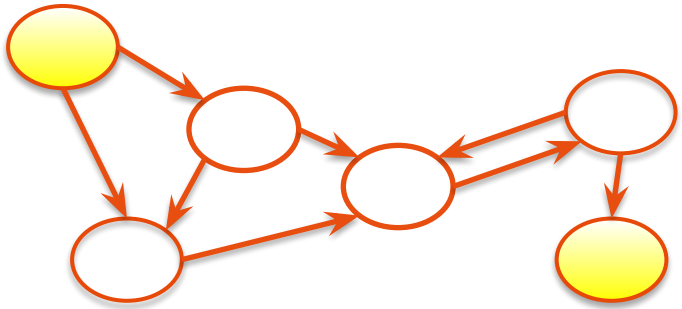
WORKFLOW

neo4j

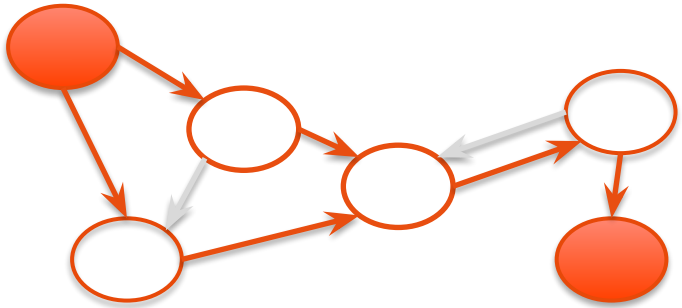
Graph Database

Store conversions
Cypher queries
Several available algorithms

Neo4j [3]



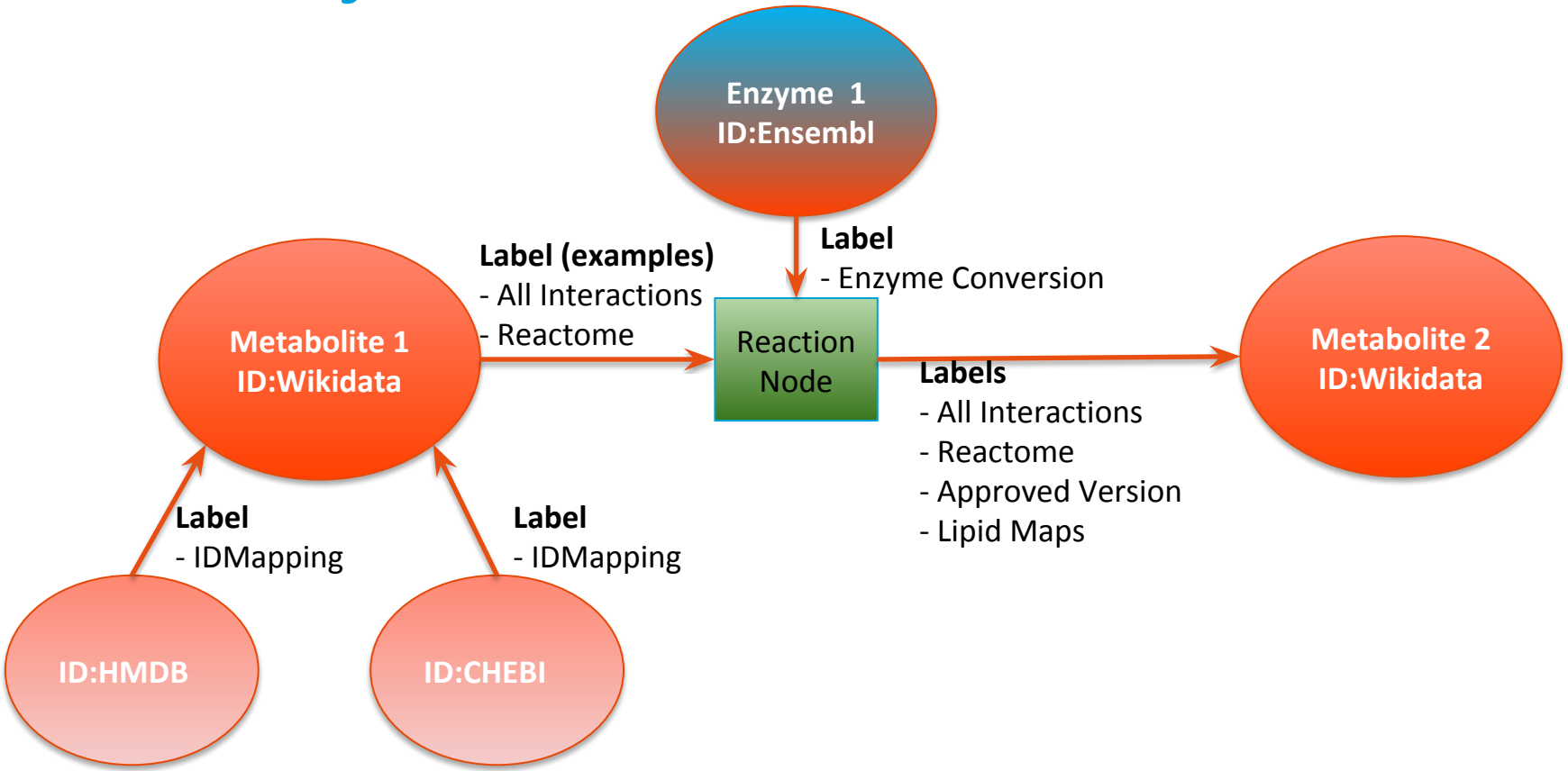
Shortest Path



Maastricht University

[3] Neo4j: https://neo4j.com/ [4] Cytoscape: http://cytoscape.org/, Shannon *et al.* (2003) DOI: 10.1101/gr.1239303

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)

[3] Neo4j: https://neo4j.com/ [4] Cytoscape: http://cytoscape.org/, Shannon *et al.* (2003) DOI: 10.1101/gr.1239303

Datasets used to test visualisation

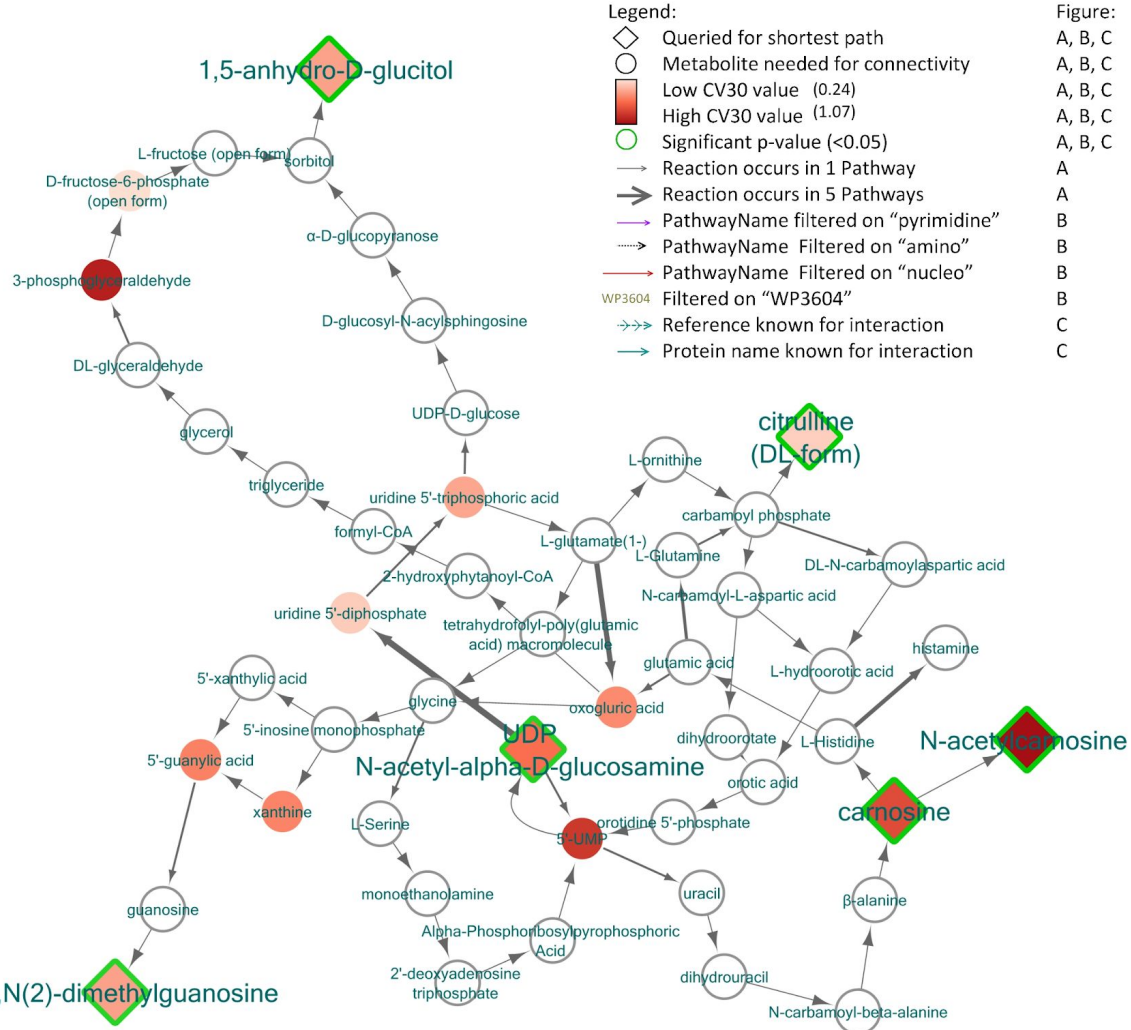
Name	Technique	Matrix	Metabolites of interest detected with:
MTBLS265 [1]	LC-MS	Blood	CV-30, p-value
MTBLS404 [2]	LC-HRMS	Urine	Spearman rank correlation test, Orthogonal partial least-squares (OPLS)
Rist *et al.* [3]	(GCxGC)-MS, targeted GC-MS and LC-MS/MS, 1H-NMR	Blood & Urine	Support Vector Machines (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 *et al.*	172/129	47.5 ± 17.1	400 (plasma), >500 (urine)	8 + 6	6 + 4

[1] Chaleckis, Romanas, *et al*. Proc. Nat. Ac. Sci. (2016). DOI: 10.1073/pnas.1603023113
[2] Thévenot, Etienne A., *et a*l.J. Prot. Res. (2015). DOI: 10.1021/acs.jproteome.5b00354
[3] Rist, Manuela J., *et al.* PloS one (2017). DOI: 10.1371/journal.pone.0183228

MTBLS265

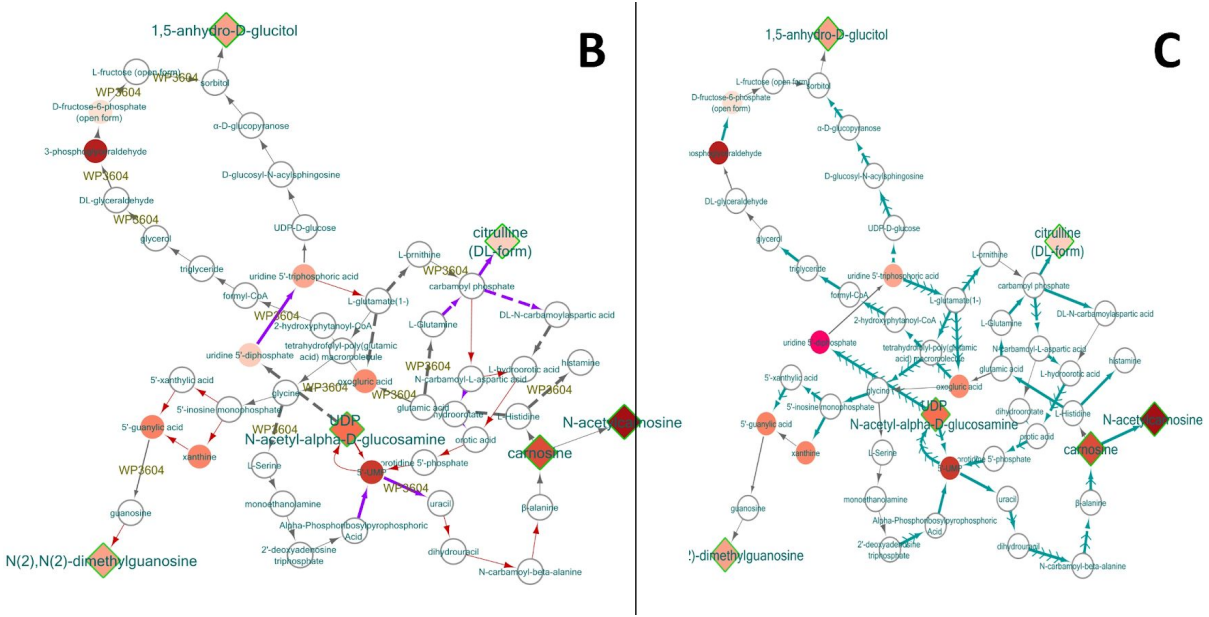
A



Maastricht U

[1] Chaleckis, Romanas, *et al*. Proc. Nat. Ac. Sci. (2016). DOI: 10.1073/pnas.1603023113

MTBLS265

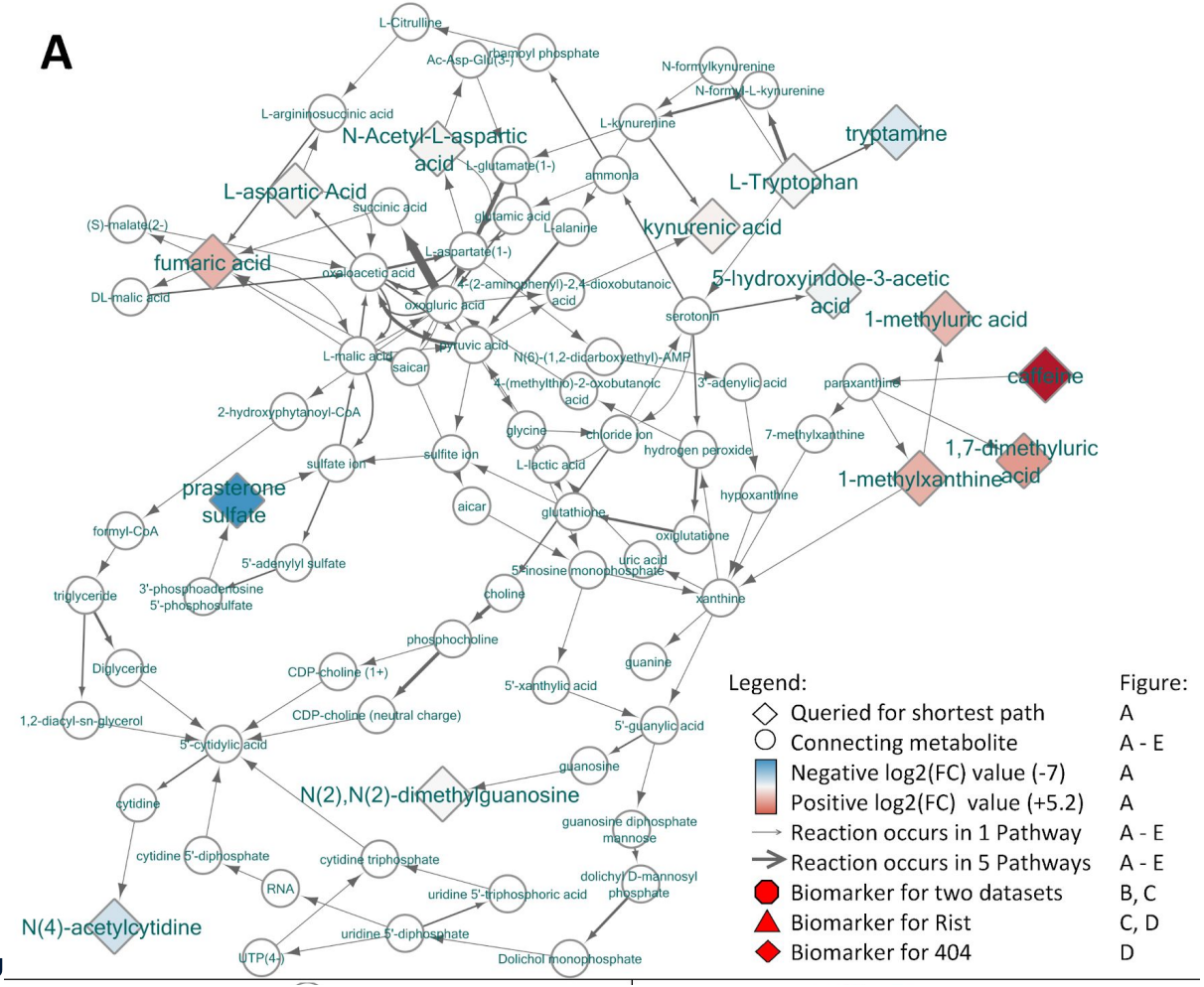


Legend:

	Figure:
Queried for shortest path	A, B, C
Metabolite needed for connectivity	A, B, C
Low CV30 value (0.24)	A, B, C
High CV30 value (1.07)	A, B, C
Significant p-value (<0.05)	A, B, C
Reaction occurs in 1 Pathway	A
Reaction occurs in 5 Pathways	A
PathwayName filtered on “pyrimidine”	B
PathwayName Filtered on “amino”	B
PathwayName Filtered on “nucleo”	B
WP3604 Filtered on “WP3604”	B
Reference known for interaction	C
Protein name known for interaction	C

Maastricht University

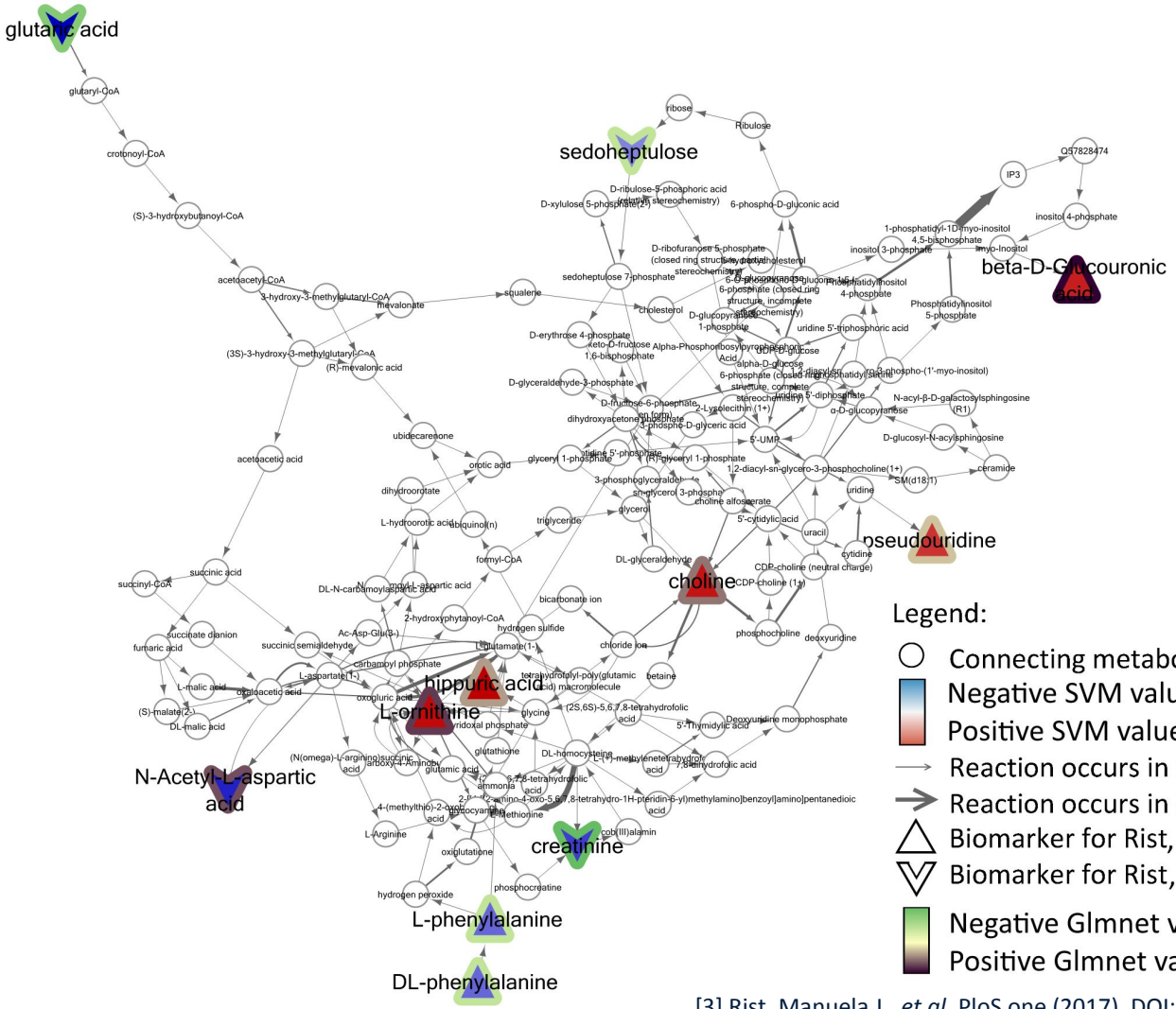
MTBLS404



Maastricht U

[2] Thévenot, Etienne A., *et a*l.J. Prot. Res. (2015). DOI: 10.1021/acs.jproteome.5b00354

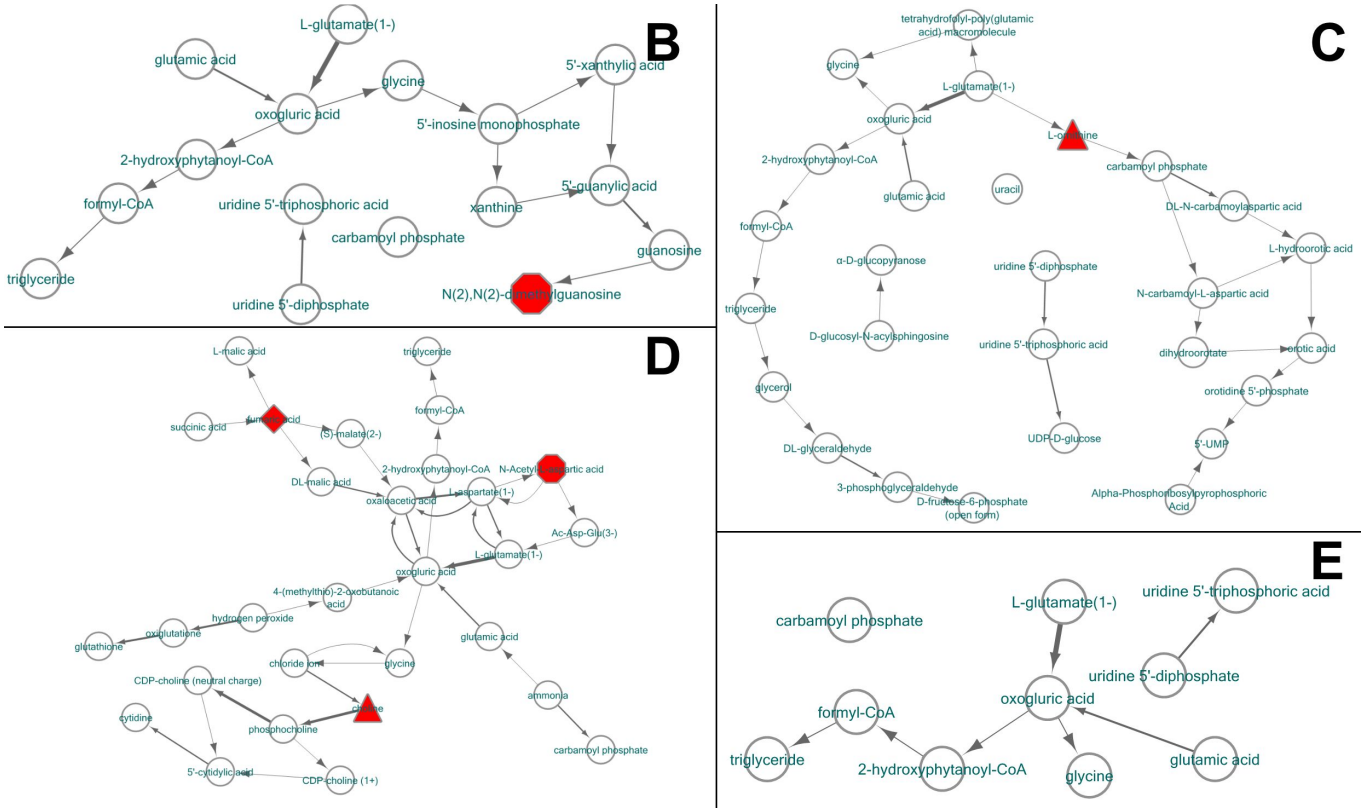
Rist



Maastric

[3] Rist, Manuela J., *et al.* PloS one (2017). DOI: 10.1371/journal.pone.0183228

Comparing all



Legend:	Figure:
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).

Maastricht University

[1] Chaleckis, Romanas, *et al*. Proc. Nat. Ac. Sci. (2016). DOI: 10.1073/pnas.1603023113
[2] Thévenot, Etienne A., *et a*l.J. Prot. Res. (2015). DOI: 10.1021/acs.jproteome.5b00354
[3] Rist, Manuela J., *et al.* PloS one (2017). DOI: 10.1371/journal.pone.0183228

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

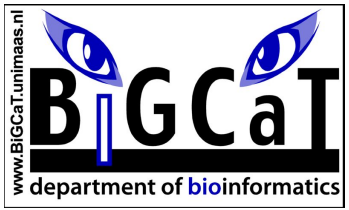
Create complete R-Markdown script for bulk approach

Allow easy integration of other omics data (proteomics and transcriptomics)

Acknowledgements, questions, discussion

- Martina Kutmon
- Jonathan Mélius
- Ryan Miller
- 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

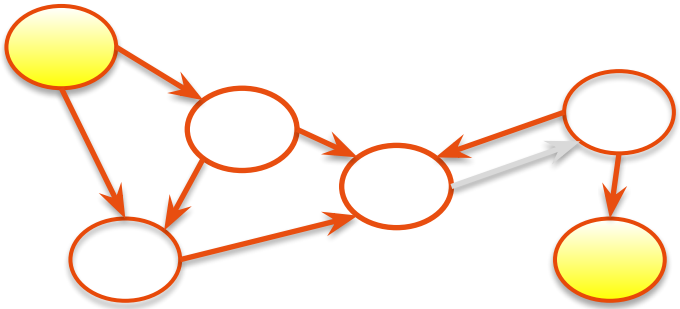
Find this presentation @ DOI: 10.6084/m9.figshare.9901397

Maastricht University

Questions:

- Which models did you consider?
- Why did you changes models?
- What are the limitations to the current model?
- What other algorithms would be interesting?
- What other data could be relevant to add?

WORKFLOW

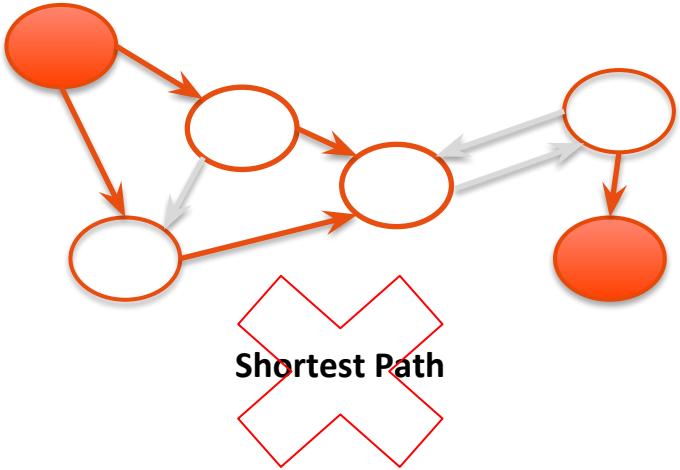


neo4j

Graph Database

Store conversions
Cypher queries
Several available algorithms

Neo4j [3]

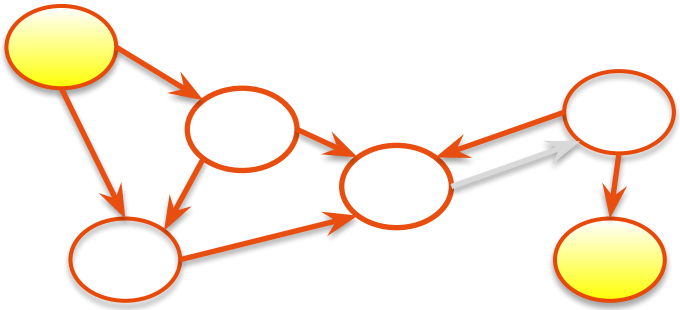


Nearest Neighbour

Maastricht University

[3] Neo4j: https://neo4j.com/ [4] Cytoscape: http://cytoscape.org/, Shannon *et al.* (2003) DOI: 10.1101/gr.1239303

WORKFLOW

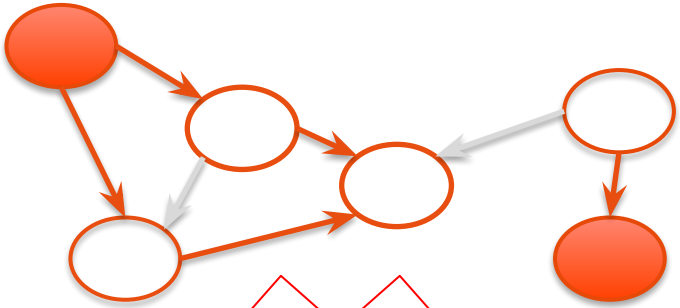


neo4j

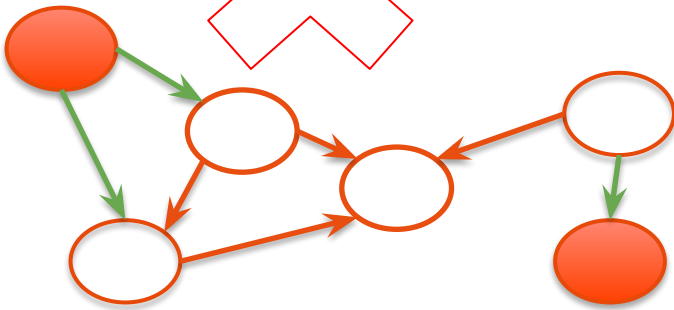
Graph Database

Store conversions
Cypher queries
Several available algorithms

Neo4j [3]



Shortest Path

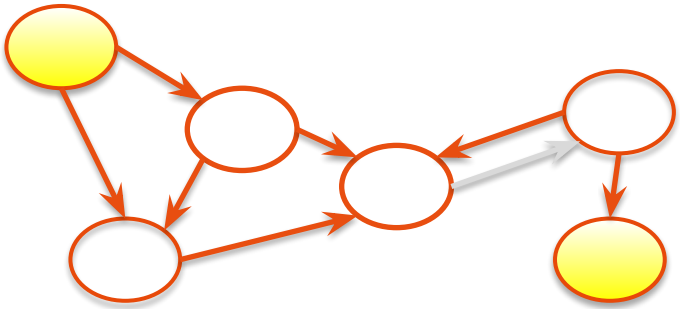


Nearest Neighbour

Maastricht University

[3] Neo4j: https://neo4j.com/ [4] Cytoscape: http://cytoscape.org/, Shannon *et al.* (2003) DOI: 10.1101/gr.1239303

WORKFLOW

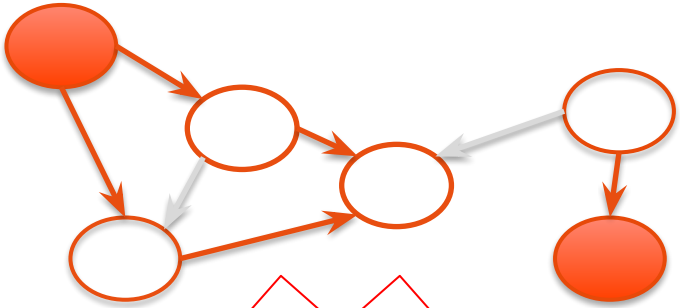


neo4j

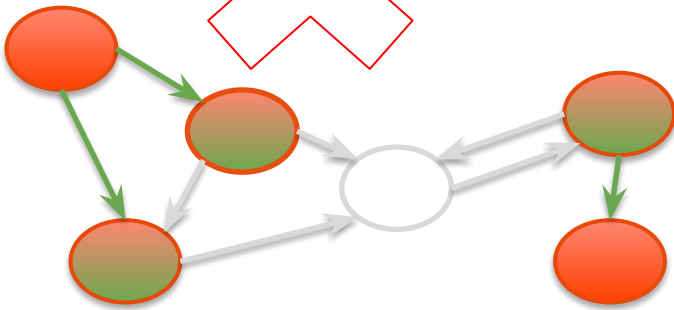
Graph Database

Store conversions
Cypher queries
Several available algorithms

Neo4j [3]



Shortest Path



Nearest Neighbour

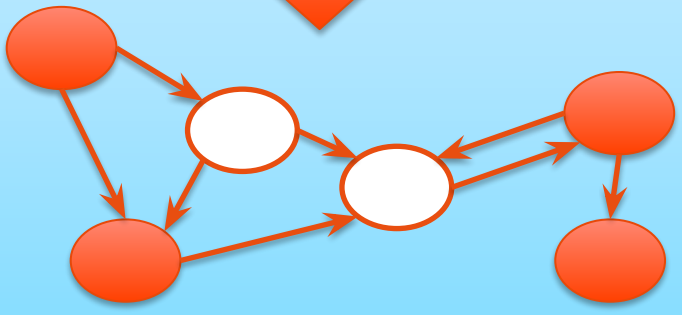
[3] Neo4j: https://neo4j.com/ [4] Cytoscape: http://cytoscape.org/, Shannon *et al.* (2003) DOI: 10.1101/gr.1239303

Two approaches

Network approach [1]

Shortest path

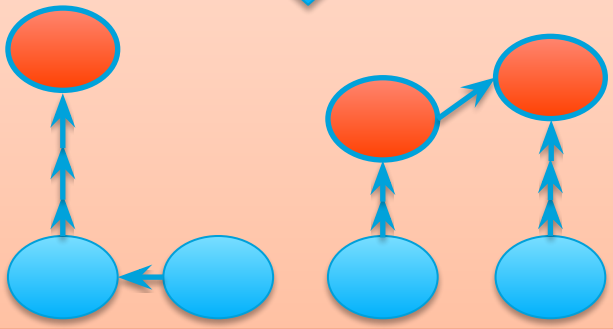
Molecular Interactions



Ontological approach [2]

Ontological Mapping

Molecular Entities



Maastricht University

[1] DOI: 10.6084/m9.figshare.5234851.v1
[2] DOI: 10.6084/m9.figshare.6368921.v1

Biological role					
Electron donor/receiver.		Energy donor/receiver.		Miscellaneous, relevant for various metabolic reactions.	
Identifier	**Name**	**Identifier**	**Name**	**Identifier**	**Name**
Q5203615	O2	Q80863	ATP	Q307434	S-adenosyl-L-homocysteine
Q506710	H+	Q185253	ADP	Q201312	S-adenosyl-L-methioninate
Q20856948	Na+ (redirected to Q3154110)	Q318369	AMP	Q407635	Coenzyme-A
Q3154110	Na+	Q422582	GDP	Q715317	Acetyl coenzyme a
Q283	H2O	Q392227	GTP		
Q1997	CO2	Q26987754	NADP+		
Q177811	PO4 3-	Q26841327	NADPH		
Q411092	Pyrophos-phoric acid	Q26987253	NAD+		
Q190901	ammonium cation	Q26987453	NADH		
		Q27102690	FADH2		

Side metabolites