Définir la distance métabolique

opportunités pour l'interprétation des données omiques et l'exploitation des bases de connaissances

Clément Frainay
CRCN INRAE Toxalim

RÉPUBLIQUE FRANÇAISE
Liberté Égalité Fraternité

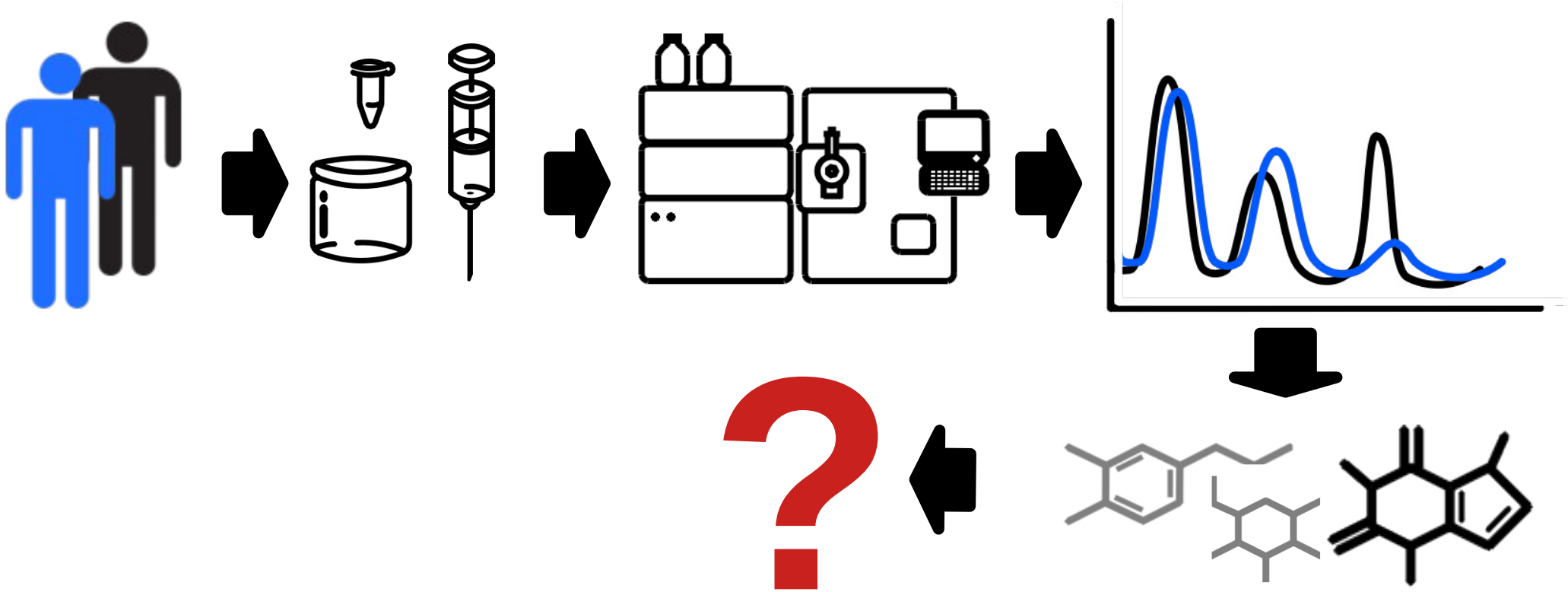
INRAE

Toxalim
Food Toxicology Research Institute

METABOHUB

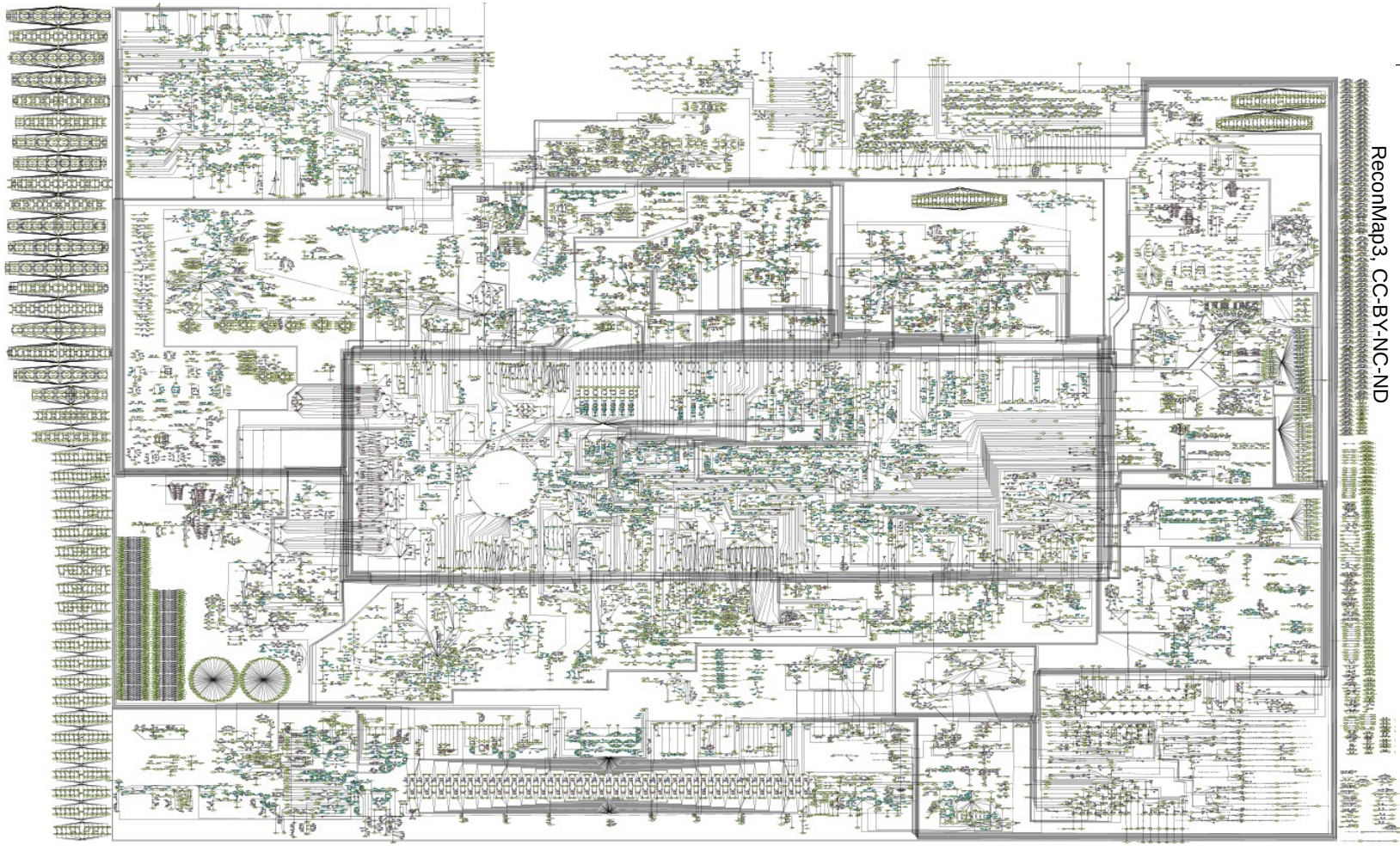
Context

Contextualizing metabolomics data for interpretation

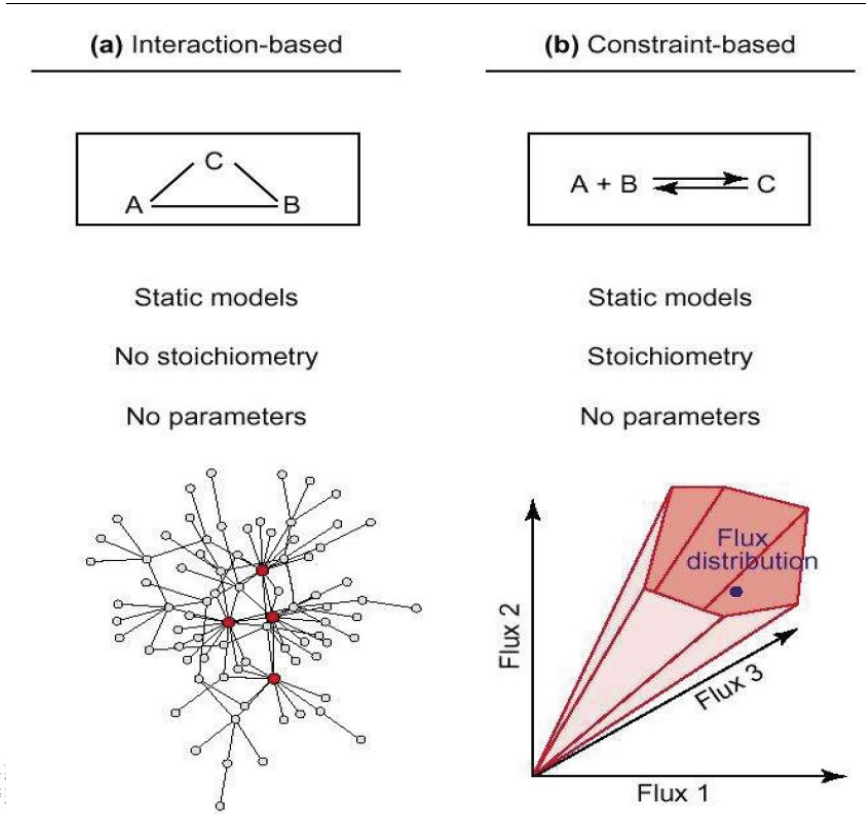


Context

Leveraging metabolic models



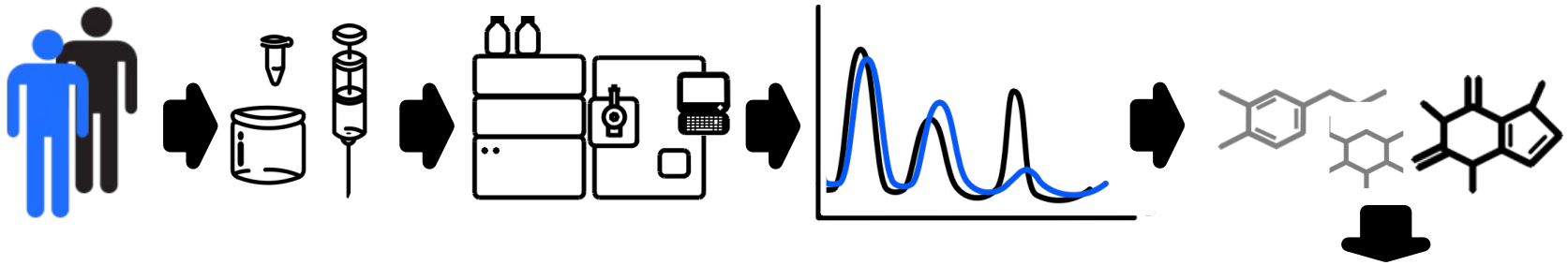
ReconMap3, CC-BY-NC-ND

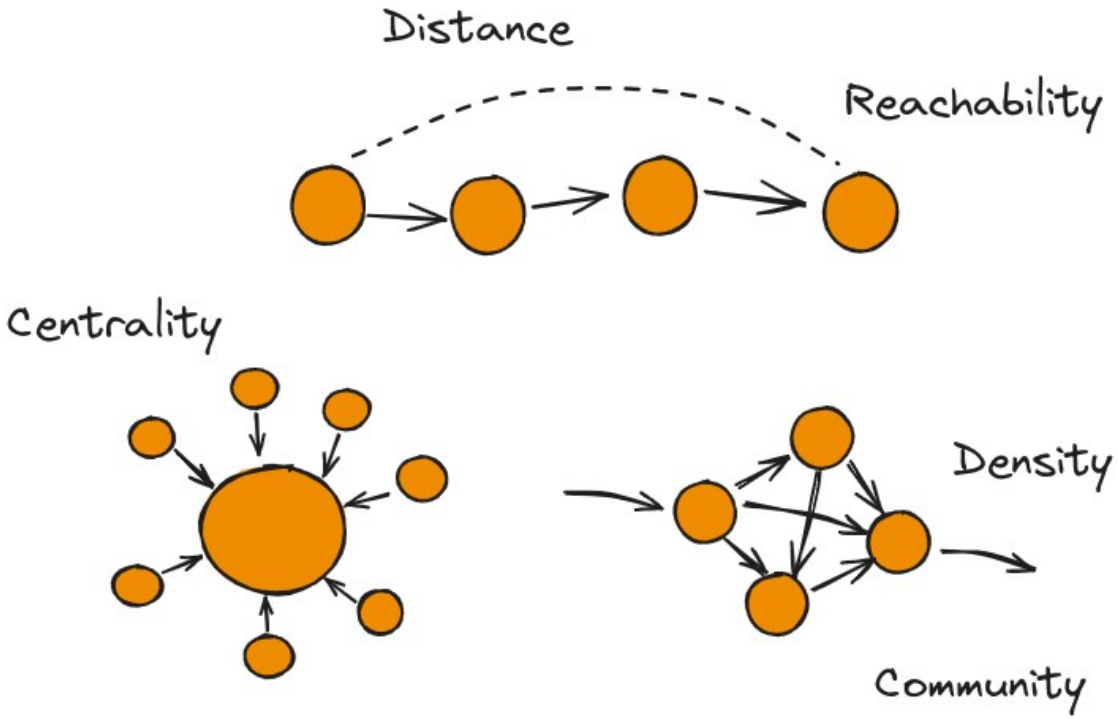


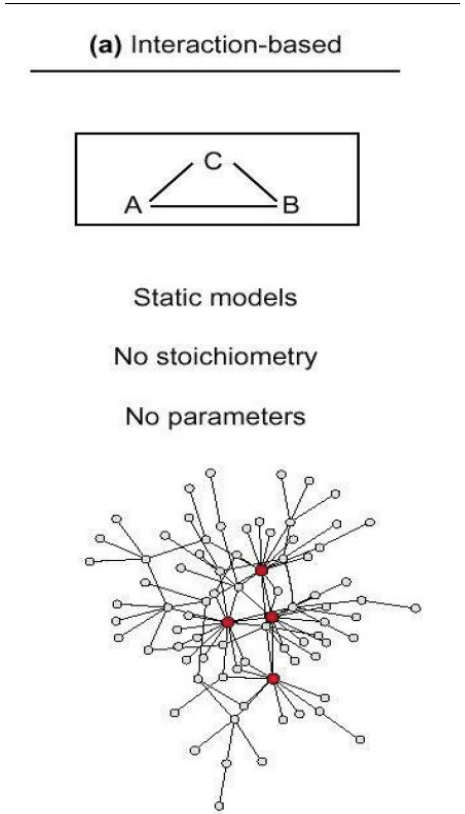
Noronha A. et al. ReconMap: An interactive visualisation of human metabolism. Bioinformatics 2017

Context

Leveraging metabolic models using graph theory



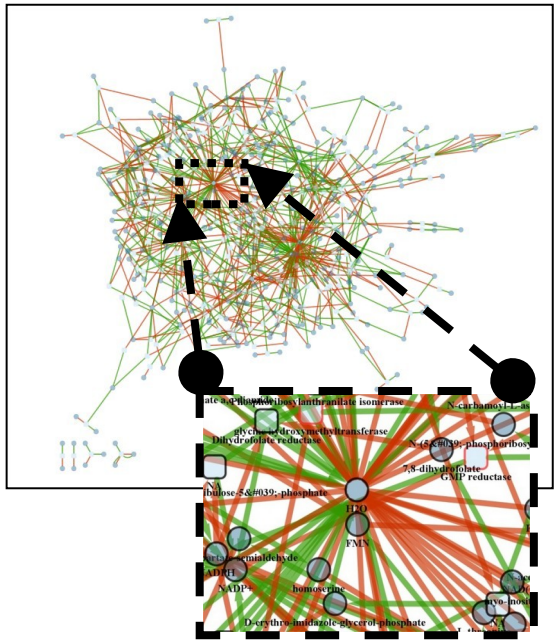




Specific characteristics of metabolic networks

Side compounds hubs

Most connected compounds in the Human metabolic network



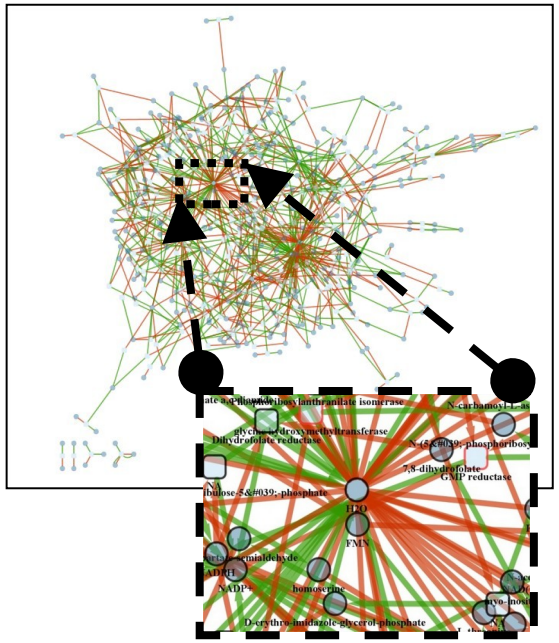
CPD	DEGREE
H+	14959
H2O	13127
ATP	5427
CoA	3961
NADP+	3816
Na+	3448
HCO3-	3129
ADP	3086
Pi	3015
NADPH	2945

Most side compounds are ubiquitous, not specific to any biological process

Specific characteristics of metabolic networks

Side compounds hubs

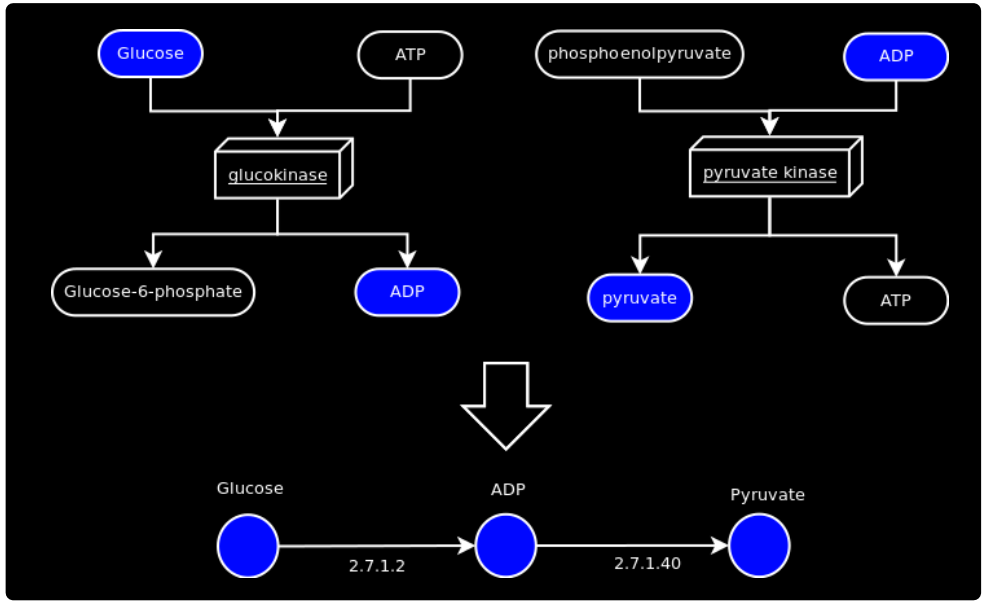
Most connected compounds in the Human metabolic network



CPD	DEGREE
H+	14959
H2O	13127
ATP	5427
CoA	3961
NADP+	3816
Na+	3448
HCO3-	3129
ADP	3086
Pi	3015
NADPH	2945

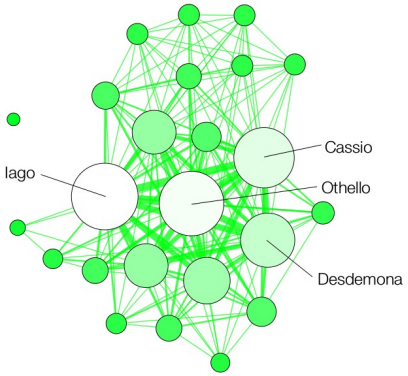
Most side compounds are ubiquitous, not specific to any biological process

Side compounds hubs creates irrelevant shortcuts



From metabolic models to metabolic networks

The side compound problem



ROMEO AND JULIET

Number of characters 41 | 37% Network density

OTHELLO

Number of characters 24 | 55% Network density

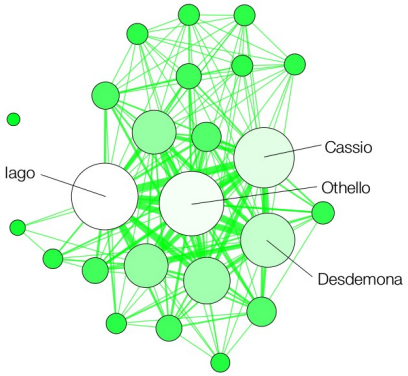
Two characters are connected if they appear in the same scene.
Their size and color intensity are proportional to their (weighted) degree of centrality.
The 'network density' measures how close the graph is to complete. A complete graph (100%) has all possible edges between its nodes.

From metabolic models to metabolic networks

The side compound problem

ROMEO AND JULIET

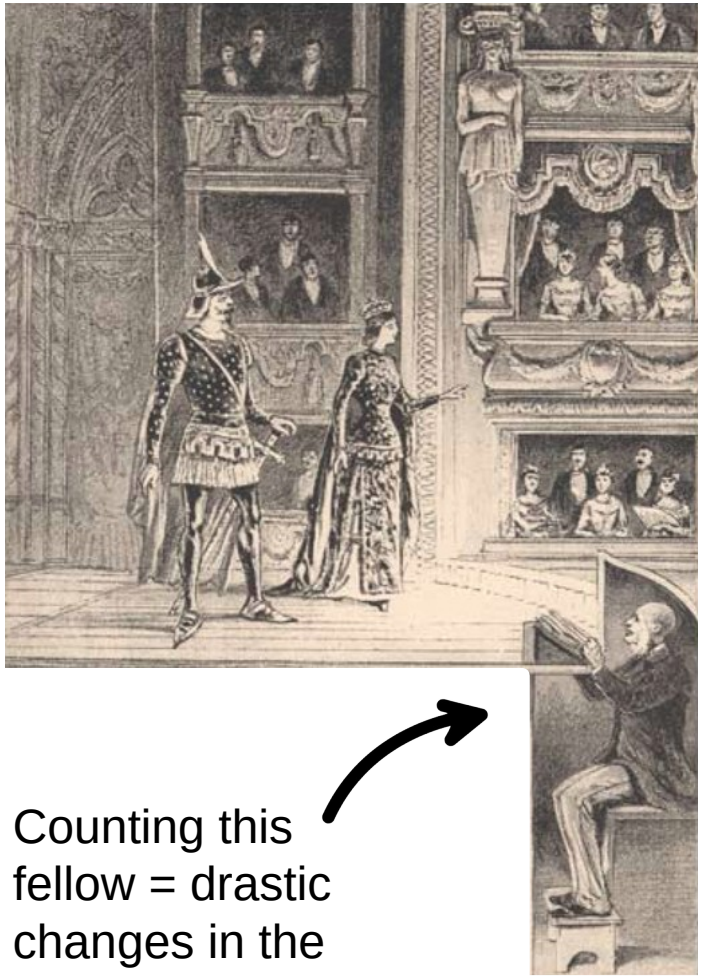
Number of characters 41 | 37% Network density



OTHELLO

Number of characters 24 | 55% Network density

Two characters are connected if they appear in the same scene. Their size and color intensity are proportional to their (weighted) degree of centrality.
The 'network density' measures how close the graph is to complete. A complete graph (100%) has all possible edges between its nodes.



Counting this fellow = drastic changes in the network structure

From metabolic models to metabolic networks

The side compound problem

The large-scale organization of metabolic networks

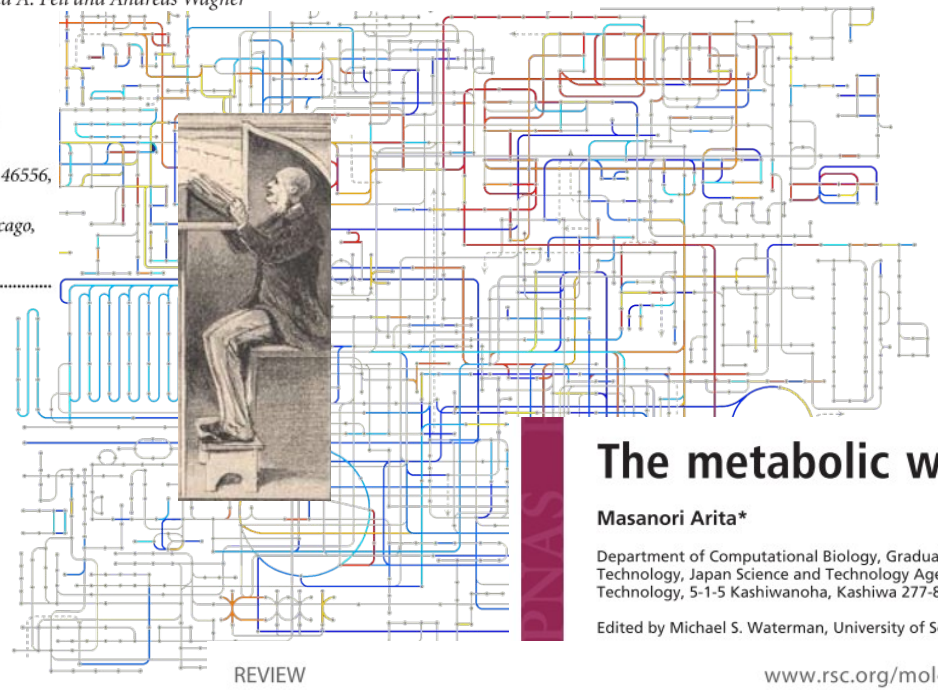
H. Jeong*, B. Tombor†, R. Albert*, Z. N. Oltvai† & A.-L. Barabási*

* Department of Physics, University of Notre Dame, Notre Dame, Indiana 46556, USA
† Department of Pathology, Northwestern University Medical School, Chicago, Illinois 60611, USA

METABOLIC ENGINEERING

The small world of metabolism

David A. Fell and Andreas Wagner



The metabolic world of *Escherichia coli* is not small

Masanori Arita*

Department of Computational Biology, Graduate School of Frontier Sciences, University of Tokyo, Precursory Research for Embryonic Science and Technology, Japan Science and Technology Agency, and Computational Biology Research Center, National Institute of Advanced Industrial Science and Technology, 5-1-5 Kashiwanoha, Kashiwa 277-8561, Japan

Edited by Michael S. Waterman, University of Southern California, Los Angeles, CA, and approved December 9, 2003 (received for review October 7, 2003)

REVIEW www.rsc.org/molecularbiosystems | Molecular BioSystems

The powerful law of the power law and other myths in network biology†

Gipsi Lima-Mendez* and Jacques van Helden*

Received 5th May 2009, Accepted 12th August 2009
First published as an Advance Article on the web 2nd October 2009
DOI: 10.1039/b908681a

Processing practice regarding S.C. caused a lot of disagreements on metabolic networks topology and properties

Defining side compounds: State of the art

- **Using manually curated list**
  - **→ Quite arbitrary**
  - **→ usually poorly reported**

“ Highly connected metabolites (such as ATP, NADH, H_2O etc) were removed ”

Parter, M., Kashtan, N. & Alon, U. Environmental variability and modularity of bacterial metabolic networks. BMC Evol Biol 7, 169 (2007).

Defining side compounds: State of the art

- **Using manually curated list**
  - → **Quite arbitrary**
  - → **usually poorly reported**

- **Using degree (hubs)**
  - → **Not all hubs are Side compounds**
    Not all side compounds are hubs
  - → **User defined threshold**

“ Highly connected metabolites (such as ATP, NADH, H_2O etc) were removed ”

Parter, M., Kashtan, N. & Alon, U. Environmental variability and modularity of bacterial metabolic networks. BMC Evol Biol 7, 169 (2007).

“ adjacency matrix was constructed for the metabolic network. From this adjacency matrix, high degree “hub” nodes were removed. Hub nodes were defined as any node with more than 50 connections and consisted of central metabolites such as hydrogen and water, cofactors such as ATP and NAD+, and the *E. coli* Biomass reactions. Pyruvate was also a hub node but was left in the network due to its role as a carbon source in this analysis. This hub-less network was ”

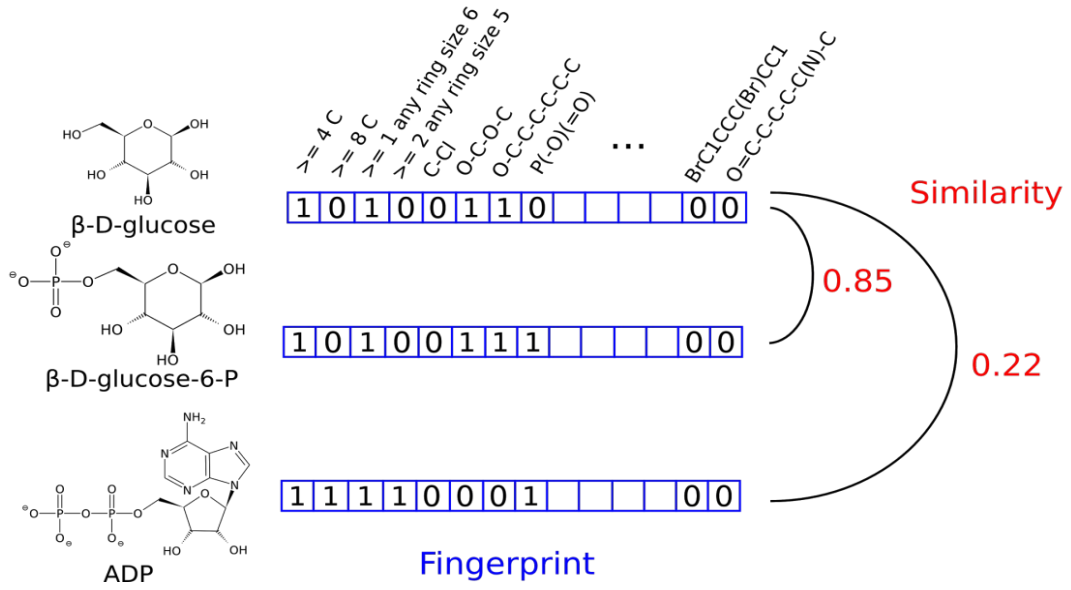
Lahiri, D., Nag, M., Dey, A. et al. Marine bioactive compounds as antibiofilm agent: a metabolomic approach. Arch Microbiol 205, 54 (2023)

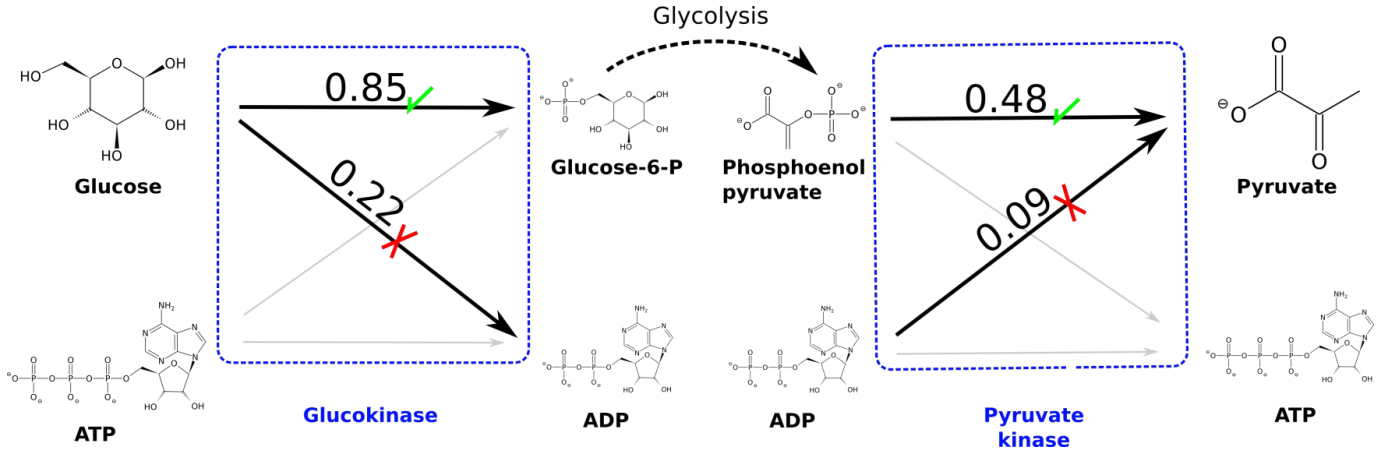
Defining side compounds: State of the art

- **Using manually curated list**
  - → **Quite arbitrary**
  - → **usually poorly reported**

- **Using degree (hubs)**
  - → **Not all hubs are Side compounds**
    Not all side compounds are hubs
  - → **User defined threshold**

- **Using chemical similarity (side transitions)**





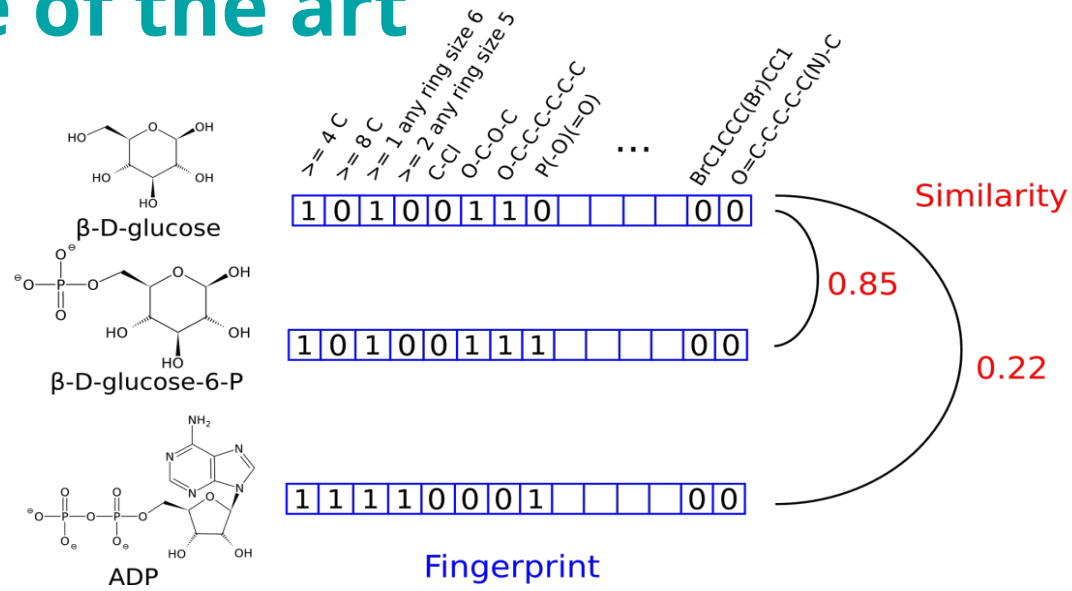
Frainay C. & Jourdan F. Computational methods to identify metabolic sub-networks based on metabolomic profiles. 2016. Briefings in Bioinformatics.

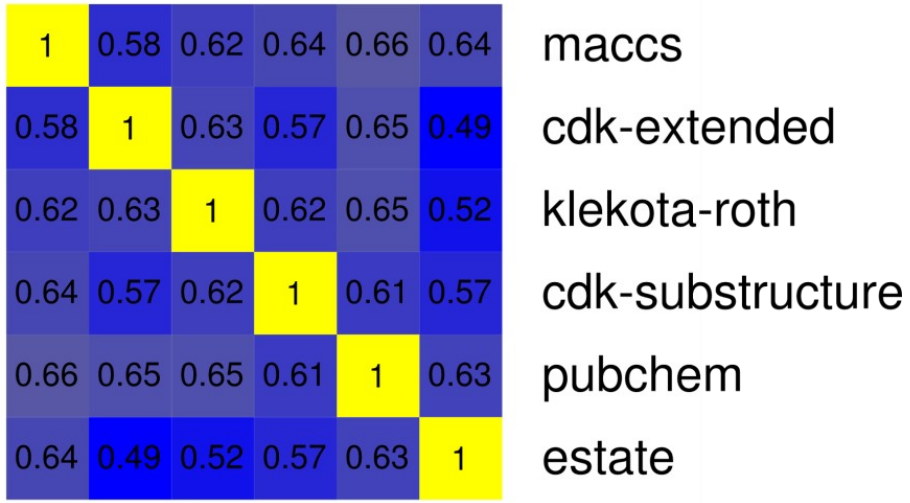
Defining side compounds: State of the art

- **Using manually curated list**
  - → Quite arbitrary
  - → usually poorly reported

- **Using degree (hubs)**
  - → Not all hubs are Side compounds
    Not all side compounds are hubs
  - → User defined threshold

- **Using chemical similarity (side transitions)**
  - → depends on Fingerprint definition
  - → might need thresholding

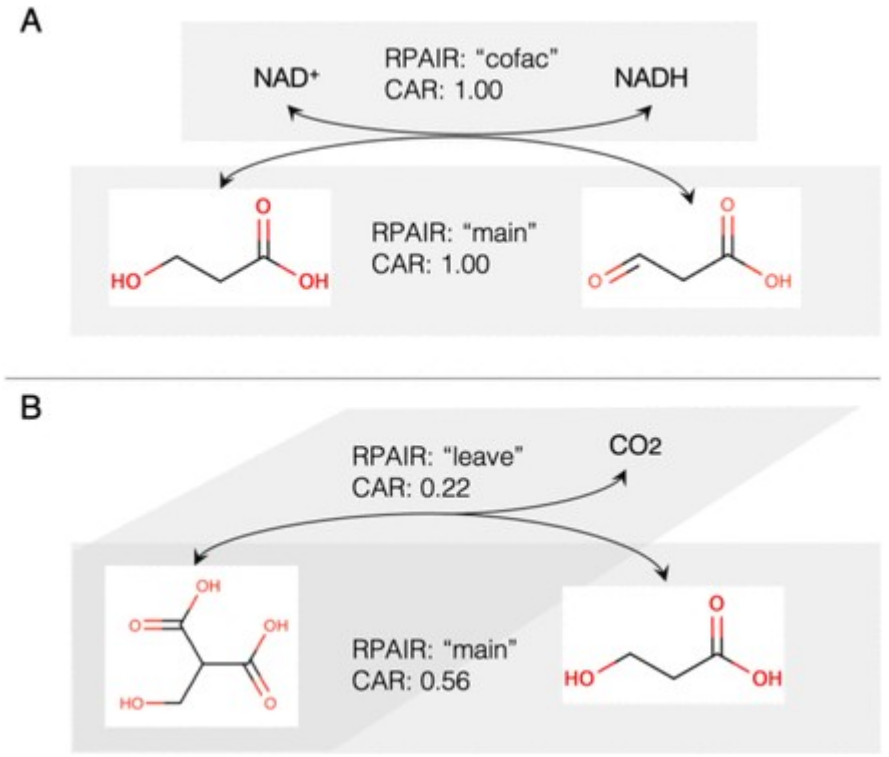




Frainay C. & Jourdan F. Computational methods to identify metabolic sub-networks based on metabolomic profiles. 2016. Briefings in Bioinformatics.

Defining side compounds: State of the art

- **Using RPAIRS (KEGG)**
  - **→ Expert curated**
  - **→ Discontinued (2016)**

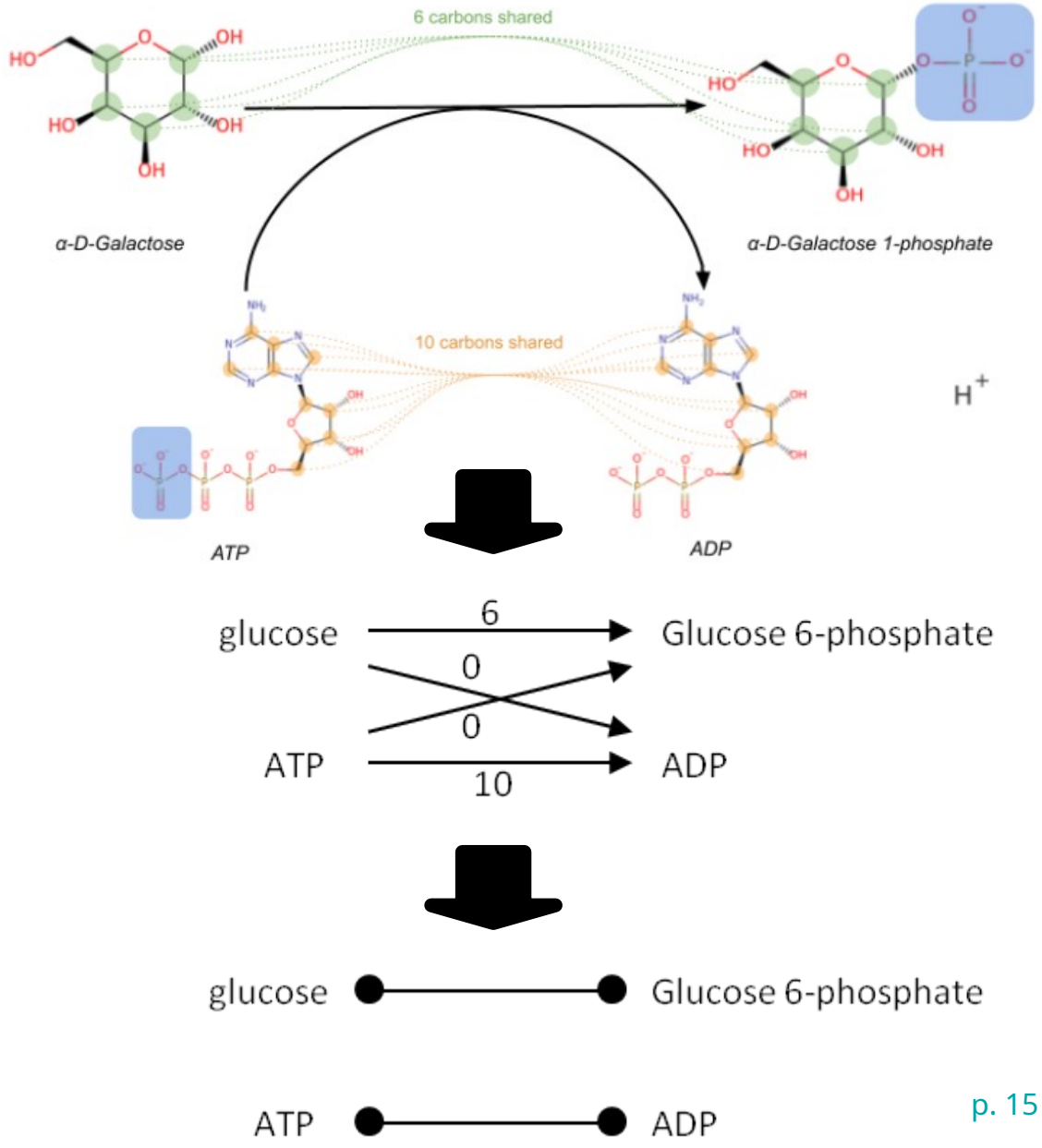


Jasmin Hafner, Vassily Hatzimanikatis, NICEpath: Finding metabolic pathways in large networks through atom-conserving substrate–product pairs, Bioinformatics, Volume 37, Issue 20, October 2021

Defining side compounds: State of the art

- Using RPAIRS (KEGG)
  - → Expert curated
  - → Discontinued (2016)

- Using Atom-Atom Mapping
  - → Step-wise Carbon continuity
  - → Context-dependant side compound
  - → Computationally expensive



Defining side compounds: State of the art

- **Using RPAIRS (KEGG)**
  - → **Expert curated**
  - → **Discontinued (2016)**

- **Using Atom-Atom Mapping**
  - → **Step-wise Carbon continuity**
  - → **Context-dependant side compound**
  - → **Computationally expensive**

	Human-GEM CSN	Human-GEM Raw
nodes	8369	8369
edges	31398	95848
diameter	23	13
avg. path length	7.06	4.04
density	0.00045	0.00137
giant cc. size	4880 (58.31%)	8271 (98.83%)

	Human-GEM CSN		Human-GEM Raw		Human-GEM CSN vs Raw	
	Compound	Neighbors	Compound	Neighbors	Compound	Neighborhood size delta
0	CoA	132	H+	2056	H+	-2055
1	GSH	107	H2O	1447	H2O	-1446
2	CoA	96	ATP	859	ATP	-829
3	acetyl-CoA	89	Pi	710	Pi	-709
4	L-carnitine	73	ADP	691	ADP	-681
5	acetyl-CoA	72	NADPH	518	NADPH	-510
6	CoA	71	H+	486	H+	-485
7	SAM	68	O2	446	O2	-445
8	GSH	55	NADP+	437	NADP+	-426
9	glycine	53	H+	423	H+	-422
10	L-carnitine	50	H+	418	H+	-417

Defining side compounds

challenges

- No methodological consensus, each as pro and cons
- Arbitrarily set parameters and subjective definitions
- Home-brewed closed-source scripts or manual procedure
- Poor reporting practice

→ Reproducibility crisis

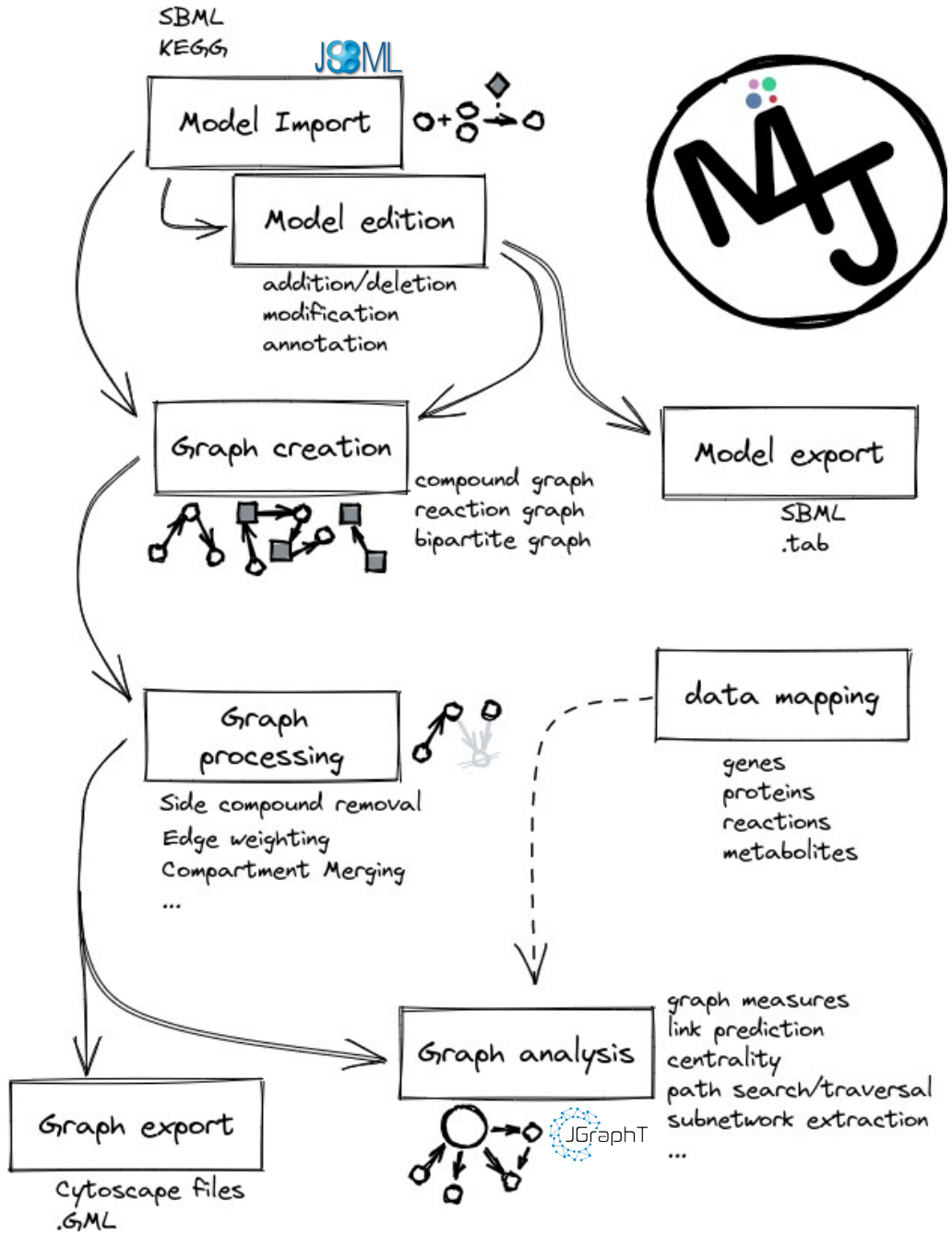
“ The definition of currency metabolites such as water and ATP may also have been different from that in the previous study because the definition has not been clearly described in [10]. ”

Takemoto K (2013) Does Habitat Variability Really Promote Metabolic Network Modularity?. PLOS ONE 8(10)

The Met4J ecosystem

The library

Scripting graph analysis of metabolic models



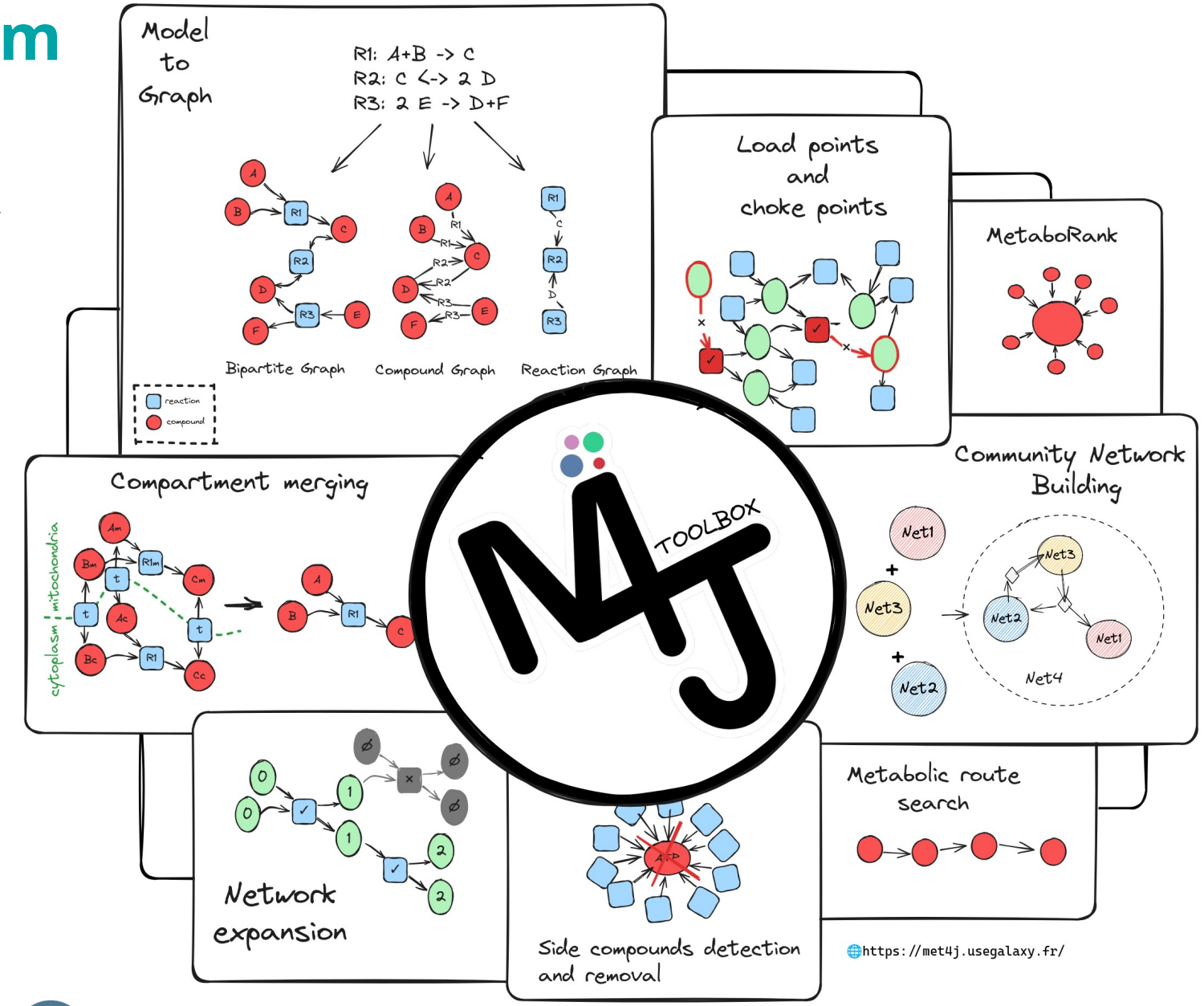
The Met4J ecosystem

The library

Scripting graph analysis of metabolic models

The Toolbox

A registry of CLI tools



45 apps (20 analyses)

Available as executable JAR, Singularity image, Docker image

The Met4J ecosystem

The library

Scripting graph analysis of metabolic models

The Toolbox

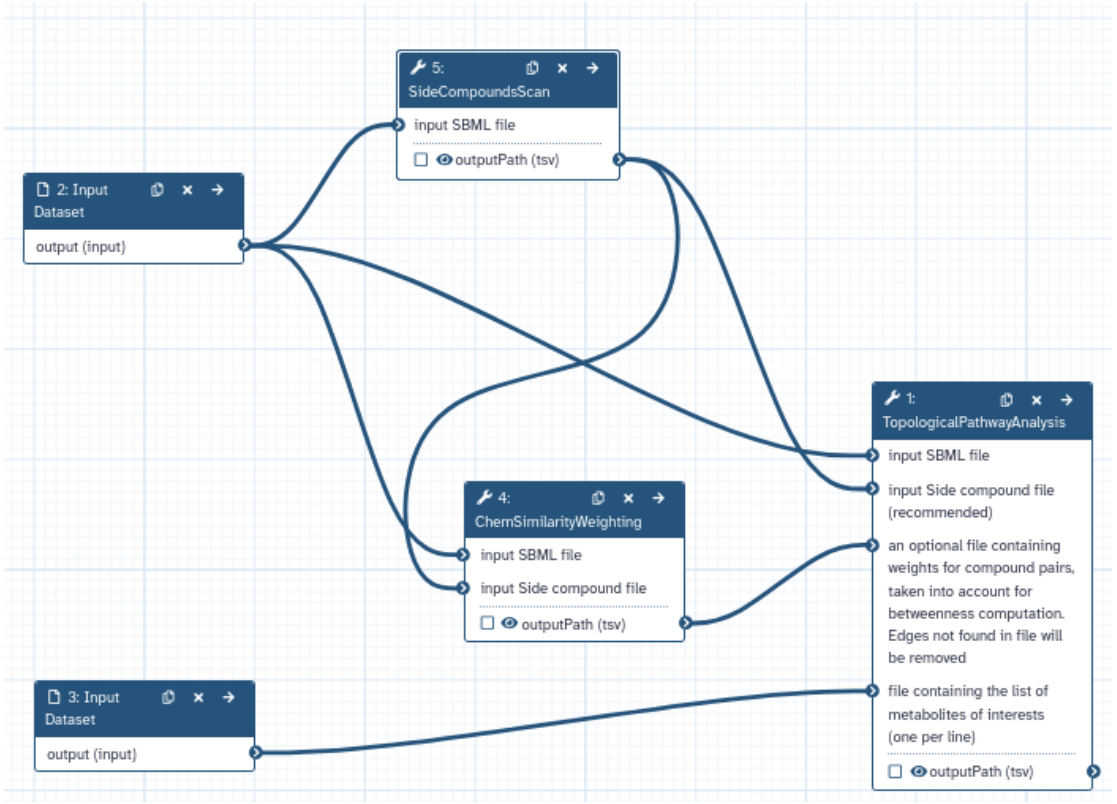
A registry of CLI tools

The workflows

FAIR processing & analysis

- Get Data
- Collection Operations
- **GENERAL TEXT TOOLS**
- Text Manipulation
- Filter and Sort
- Join, Subtract and Group
- **METABOLIC NETWORKS**
- Met4J: Get and Set attributes
- Met4J: Fetch BIGG data
- Met4J: Convert
- Met4J: Map omics data
- Met4J: Topological analyses
- Met4J: Metabolic reconstruction
- **STATISTICS AND VISUALISATION**
- Graph/Display Data
- Interactive Tools

A Galaxy instance to pipe any app from the toolbox with other tools from the community



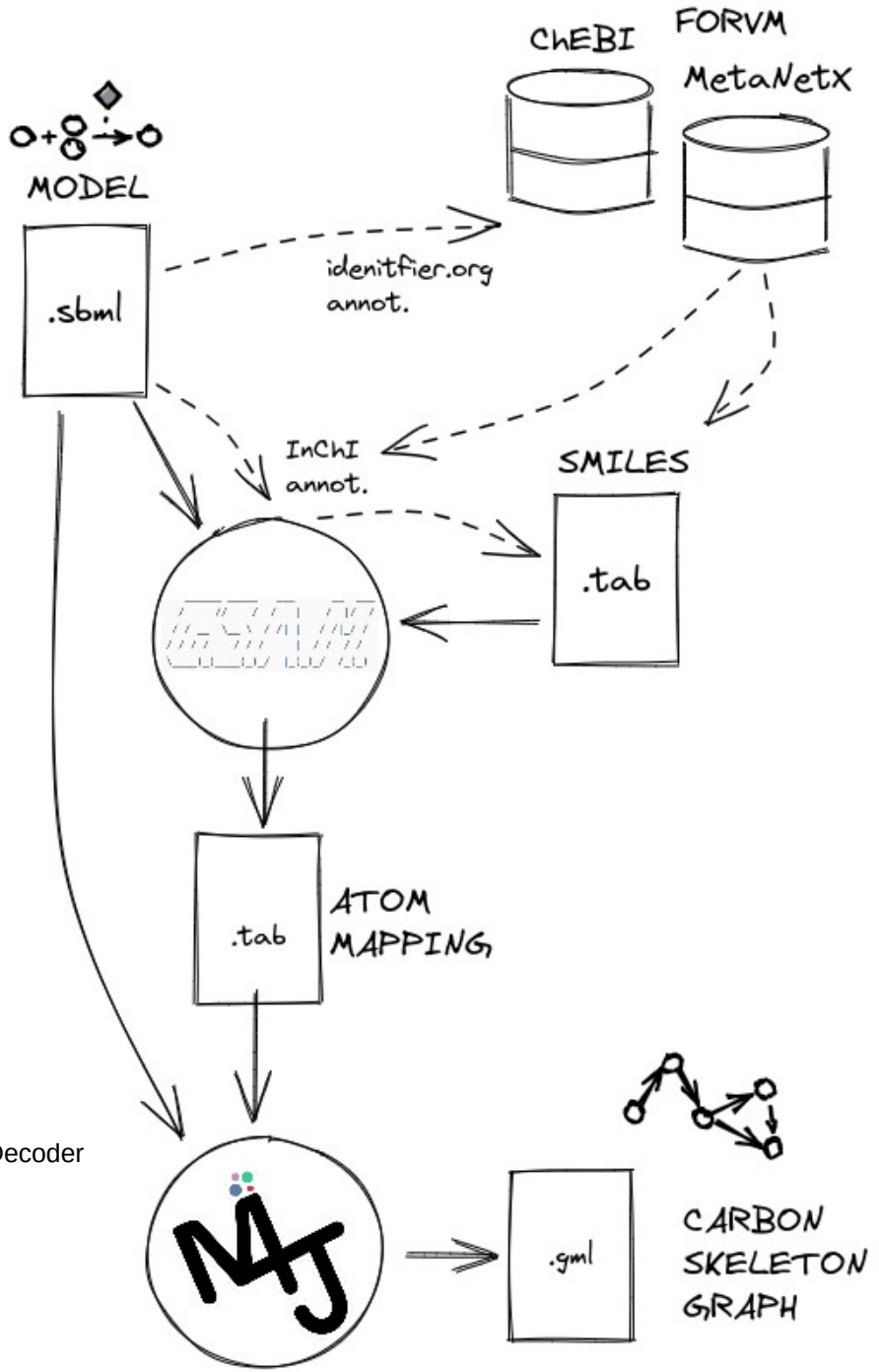
Genome Scale Atom Mapping

Combining RDT[1] with SBMLs

```
   ______ _____   ___      __  ___
  / ____// ___/  /   |    /  |/  /
 / / __  \__ \  / /| |   / /|_/ /
/ /_/ / ___/ / / ___ |_ / /  / /
\____(_)____(_)_/  |_(_)_/  /_(_)
```

```
java -jar GSAM.jar -i path/to/sbml.xml -s path/to/smiles.tab -o path/to/outputDir
```

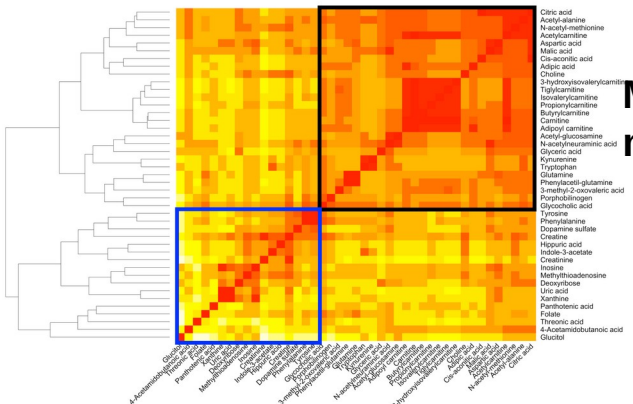
https://forge.inrae.fr/metexplore/gsam



[1]Rahman, S. A., Torrance, G., Baldacci, L., Cuesta, S. M., Fenninger, F., Gopal, N., … Thornton, J. M. (2016). Reaction Decoder Tool (RDT): Extracting features from chemical reactions. Bioinformatics, 32(13), 2065–2066.

Exploiting Metabolic proximity

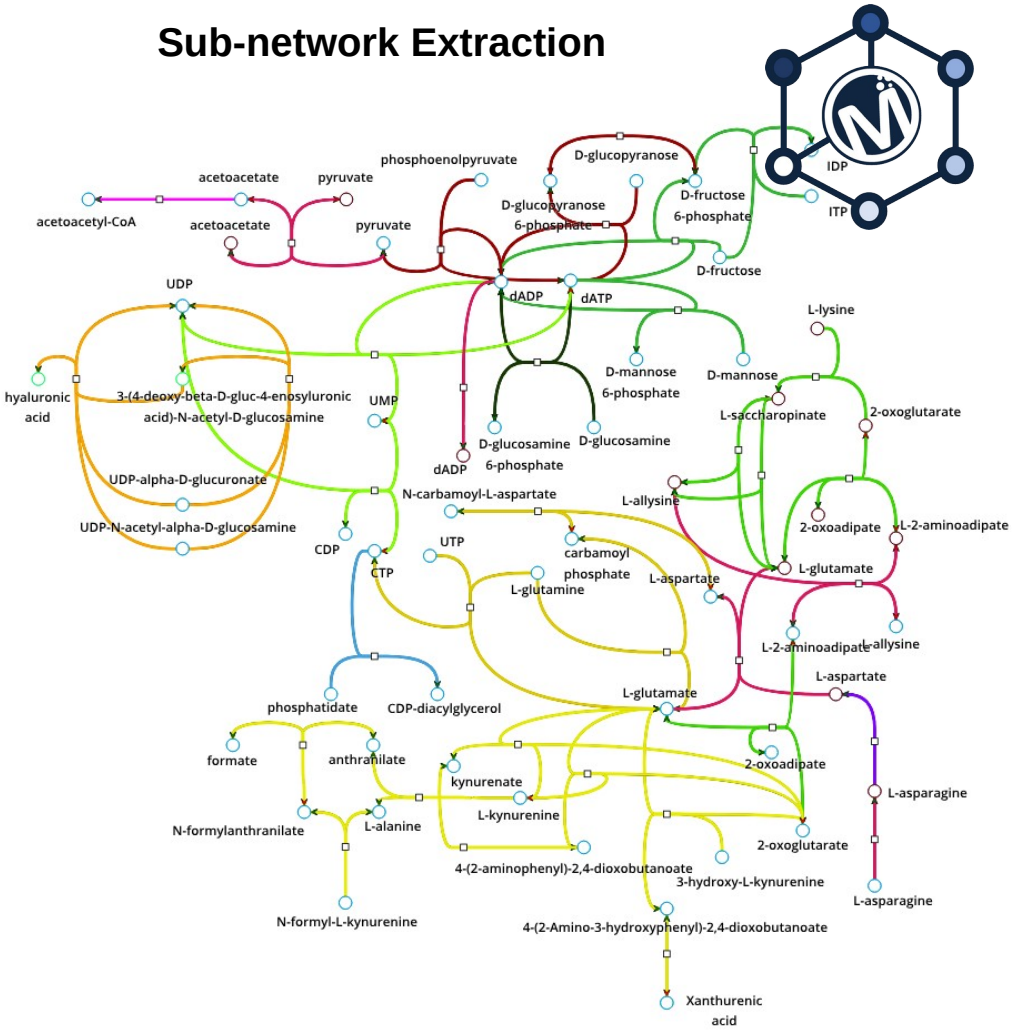
Shortest paths



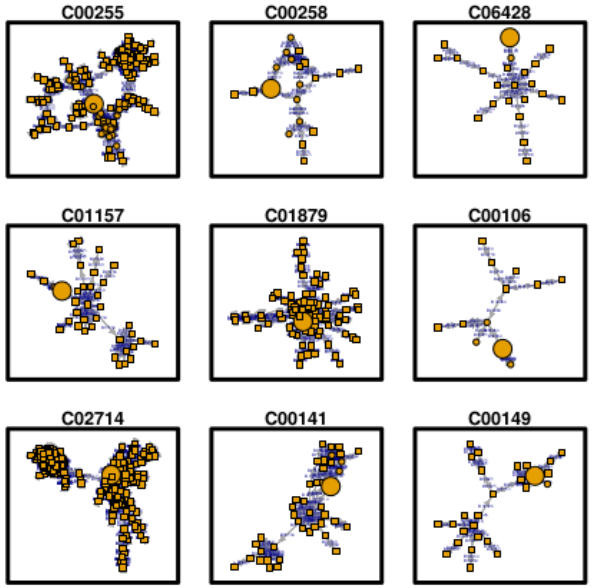
Mochel et al. Journal of Inherited Metabolic Disease. 2018

Metabolic distance matrix

Sub-network Extraction



Network partition



Maxime Chazalviel et al. MetExploreViz: web component for interactive metabolic network visualization. (2017) Bioinformatics

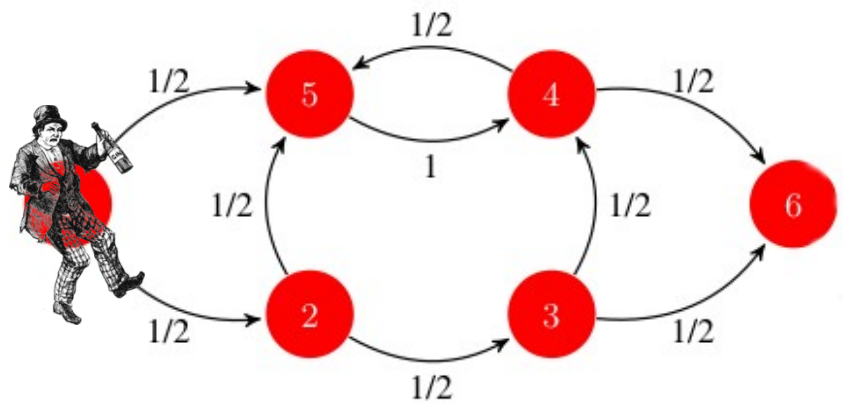
Exploiting Metabolic proximity

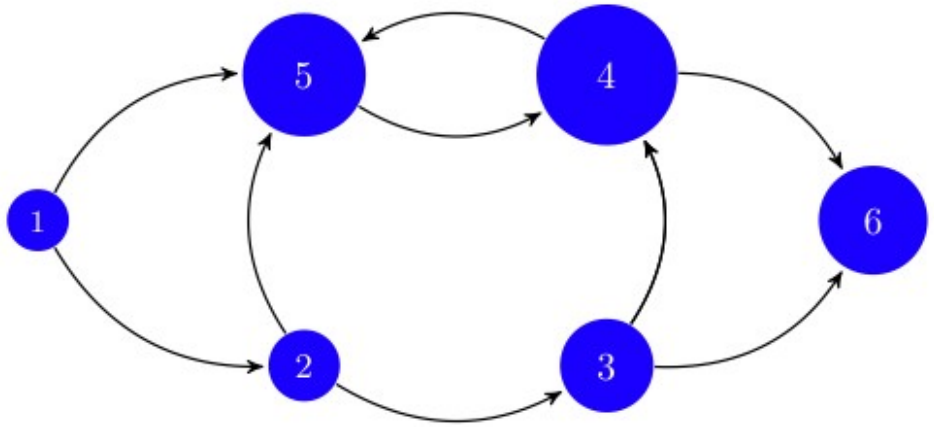
Page Rank: Beyond shortest path

Models the probability of landing on a node after many steps of a random walker

“closer” nodes in terms of link flow get higher scores.

Nodes connected to highly ranked or densely linked regions appear more “proximal”

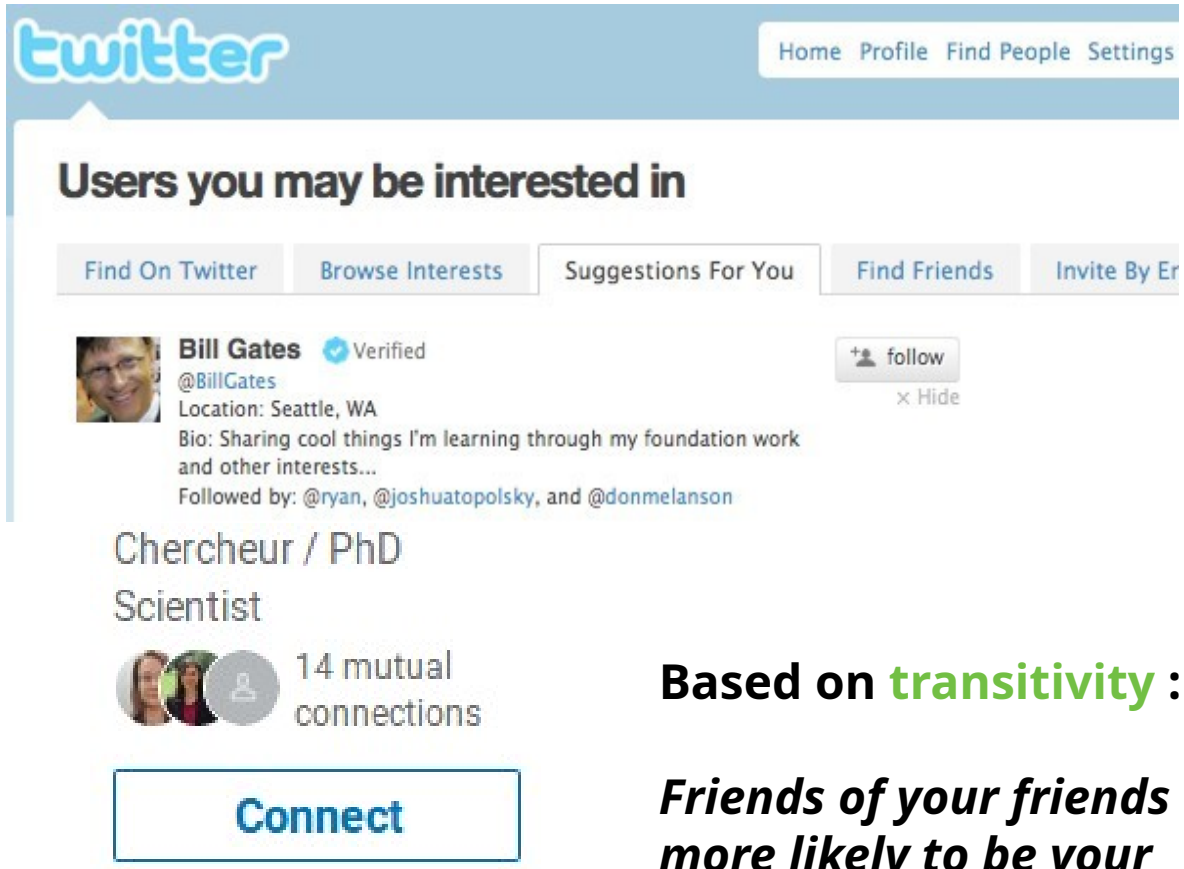




Page L, Brin S, Motwani S, et al. The PageRank Citation Ranking: Bringing Order to the Web. 1999, Stanford, CA, USA: Stanford

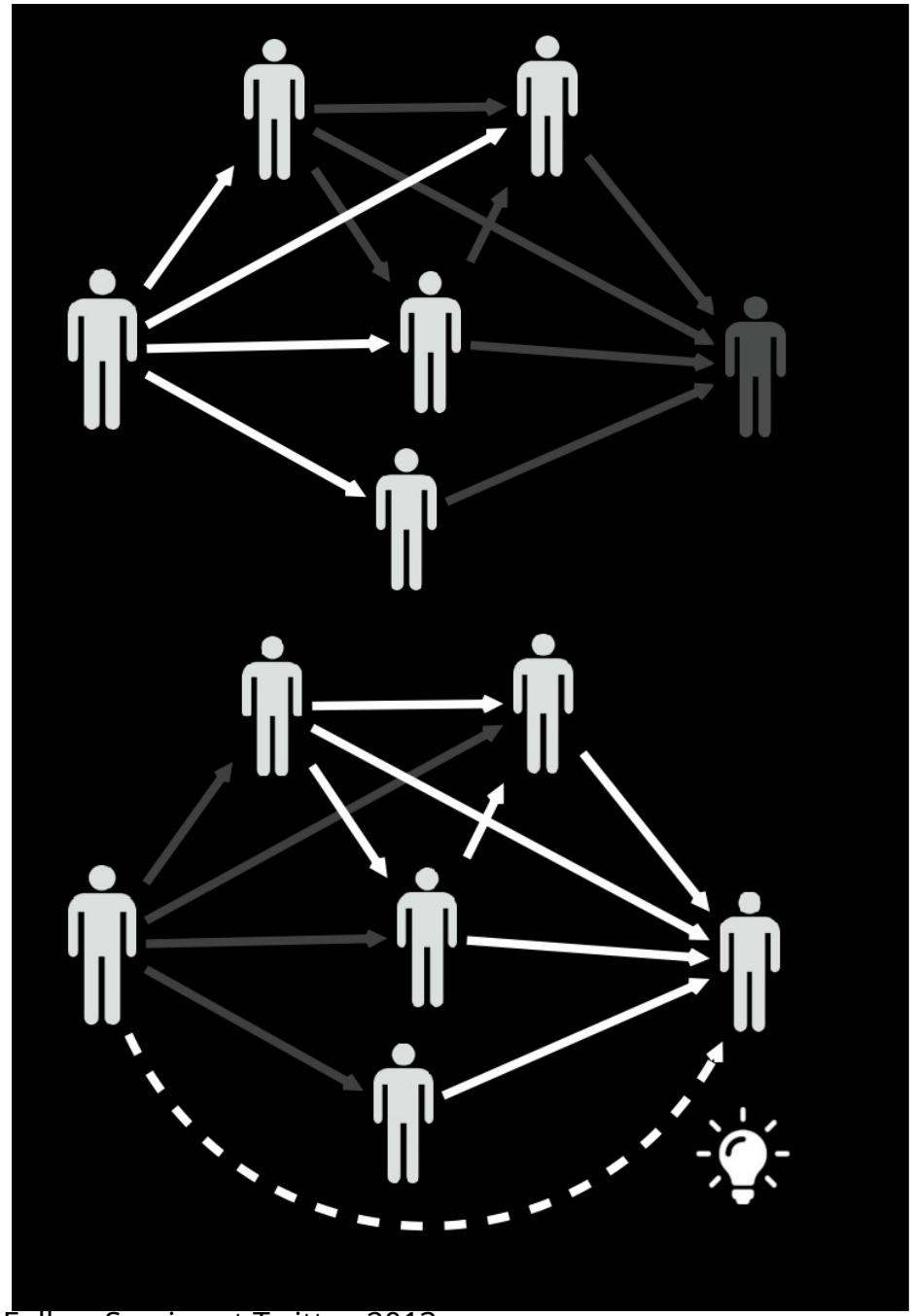
Exploiting Metabolic proximity

Metabolic recommender System



Based on transitivity :

Friends of your friends are more likely to be your friends



Gupta P, Goel A, Lin J, et al. WTF: The Who to Follow Service at Twitter. 2013

Exploiting Metabolic proximity

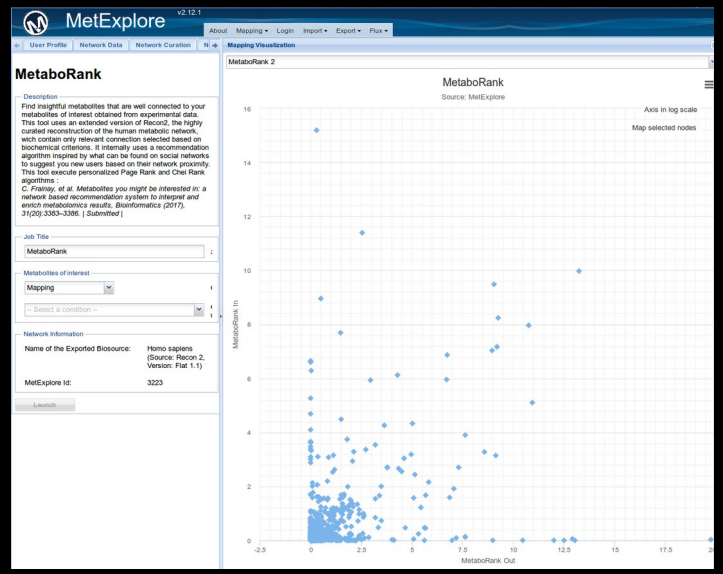
Metabolic recommender System

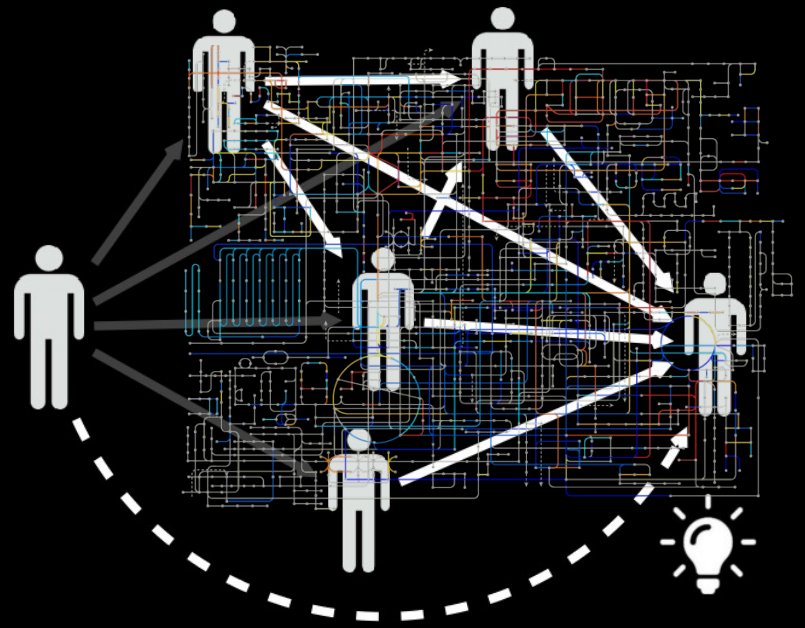
MetaboRank: network-based recommendation system to interpret and enrich metabolomics results

Clément Frainay, Sandrine Aros, Maxime Chazalviel, Thomas Garcia, Florence Vinson, Nicolas Weiss, Benoit Colsch, Frédéric Sedel, Dominique Thabut, Christophe Junot, ... Show more

Bioinformatics, Volume 35, Issue 2, 15 January 2019, Pages 274–283, https://doi.org/10.1093/bioinformatics/bty577

Published: 06 July 2018 **Article history** ▾





Based on transitivity :

Metabolites produced from affected metabolites are more likely to be affected as well

Gupta P, Goel A, Lin J, et al. WTF: The Who to Follow Service at Twitter. 2013

Exploiting Metabolic proximity

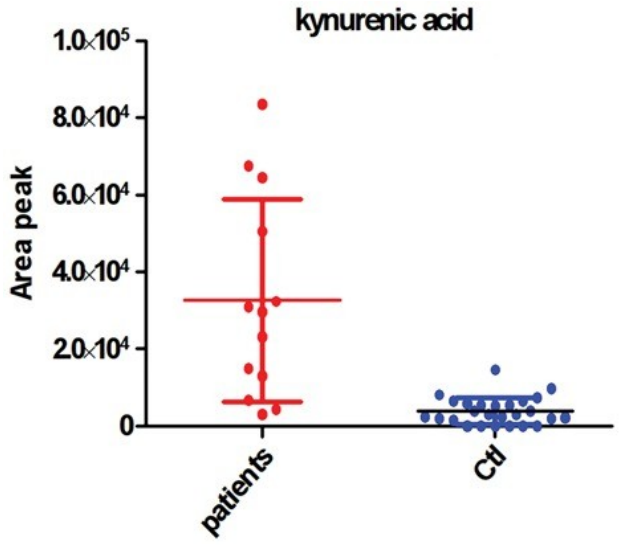
Metabolic recommender System

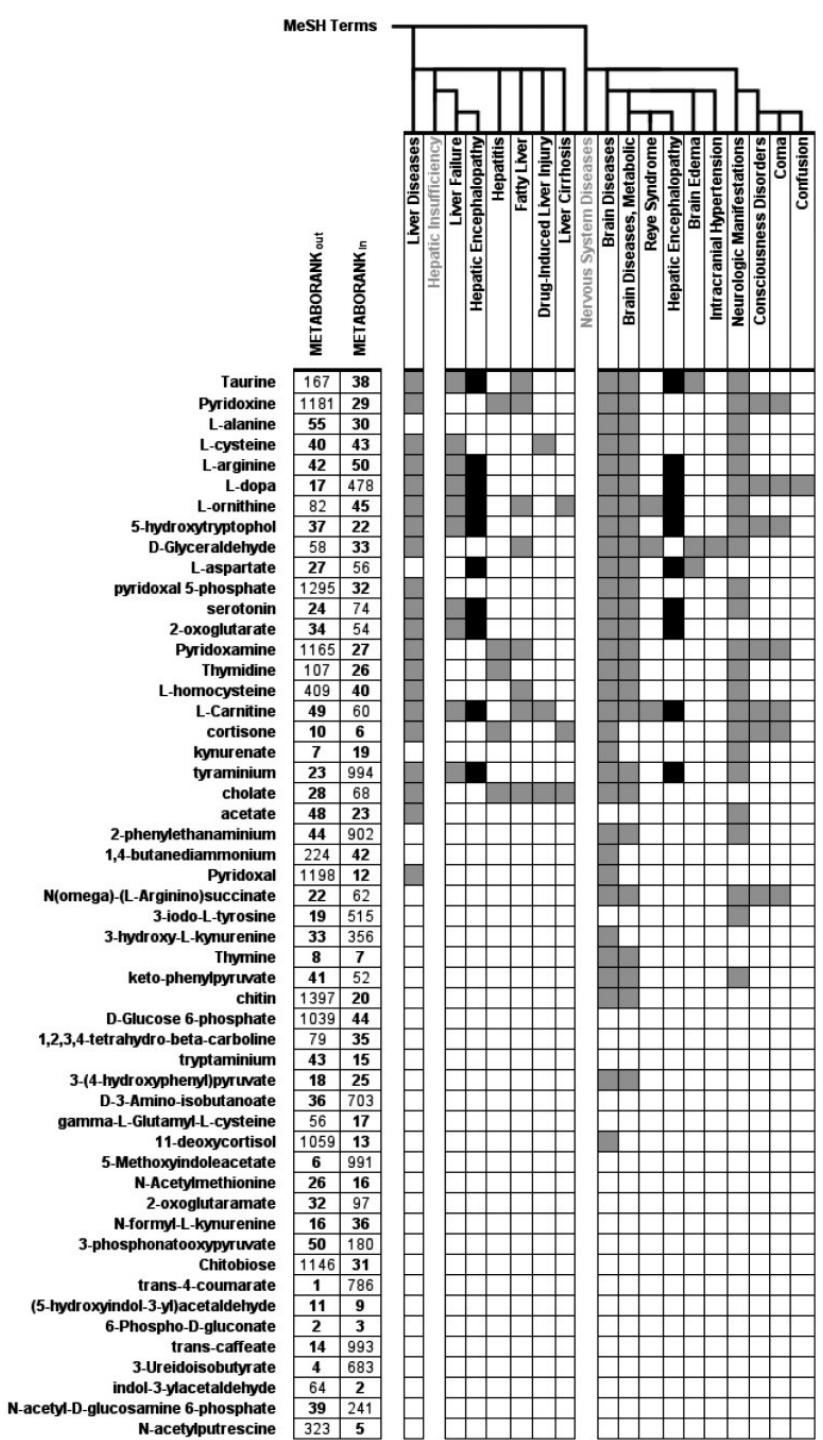
MetaboRank: network-based recommendation system to interpret and enrich metabolomics results

Clément Frainay, Sandrine Aros, Maxime Chazalviel, Thomas Garcia, Florence Vinson, Nicolas Weiss, Benoit Colsch, Frédéric Sedel, Dominique Thabut, Christophe Junot, ... Show more

Bioinformatics, Volume 35, Issue 2, 15 January 2019, Pages 274–283, https://doi.org/10.1093/bioinformatics/bty577

Published: 06 July 2018 **Article history** ▾





Exploiting Metabolic proximity

Metabolic recommender System

MetaboRank: network-based recommendation system to interpret and enrich metabolomics results

Clément Frainay, Sandrine Aros, Maxime Chazalviel, Thomas Garcia, Florence Vinson, Nicolas Weiss, Benoit Colsch, Frédéric Sedel, Dominique Thabut, Christophe Junot, ... Show more

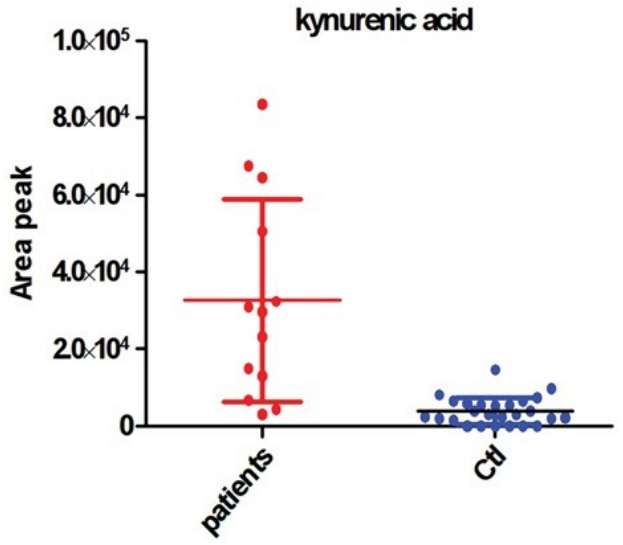
Bioinformatics, Volume 35, Issue 2, 15 January 2019, Pages 274–283, https://doi.org/10.1093/bioinformatics/bty577

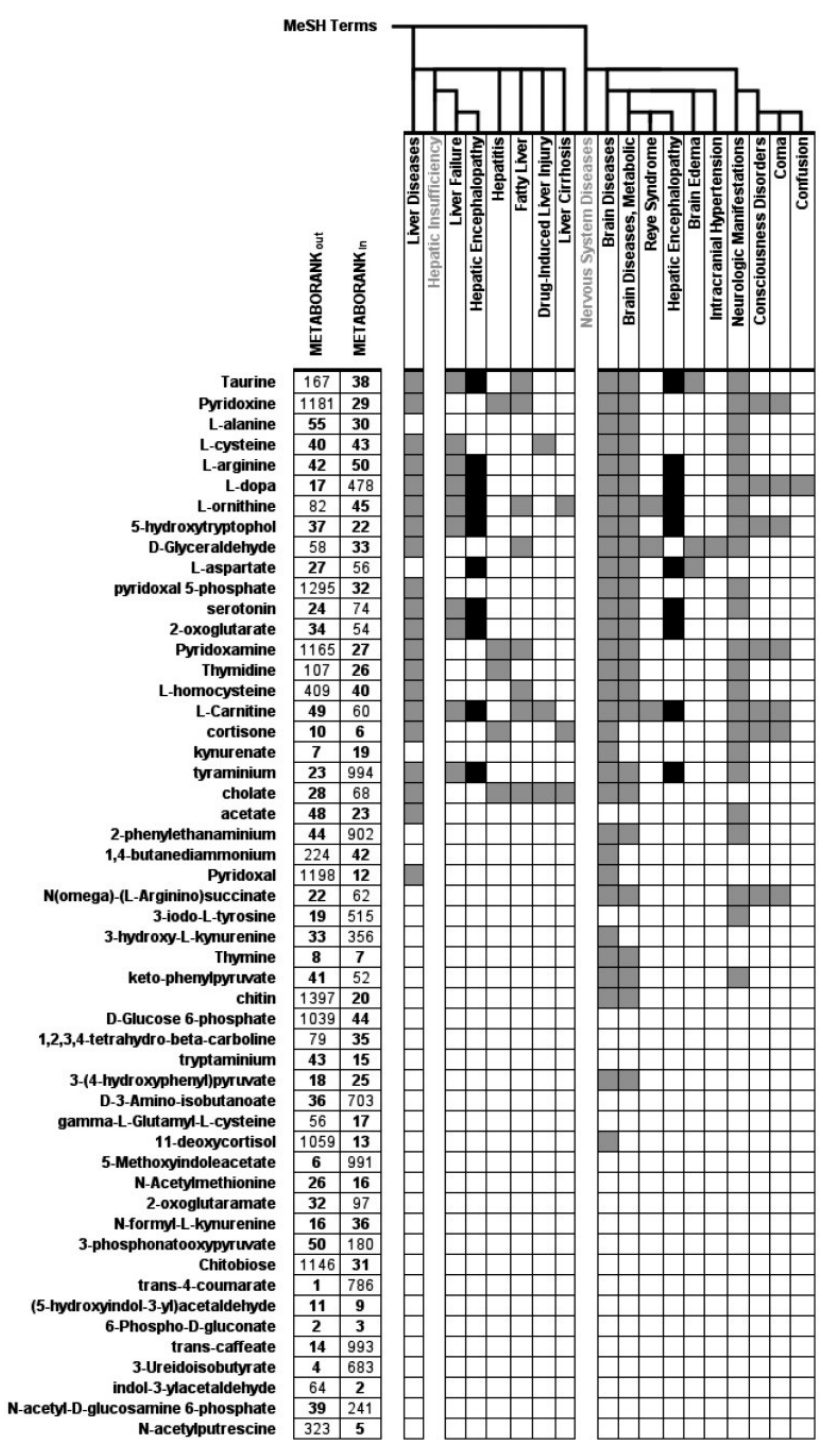
Published: 06 July 2018 **Article history** ▾

Published in final edited form as:
Metab Brain Dis. 2014 December ; 29(4): 991–1006. doi:10.1007/s11011-013-9444-9.

α-Ketoglutaramate: An overlooked metabolite of glutamine and a biomarker for hepatic encephalopathy and inborn errors of the urea cycle

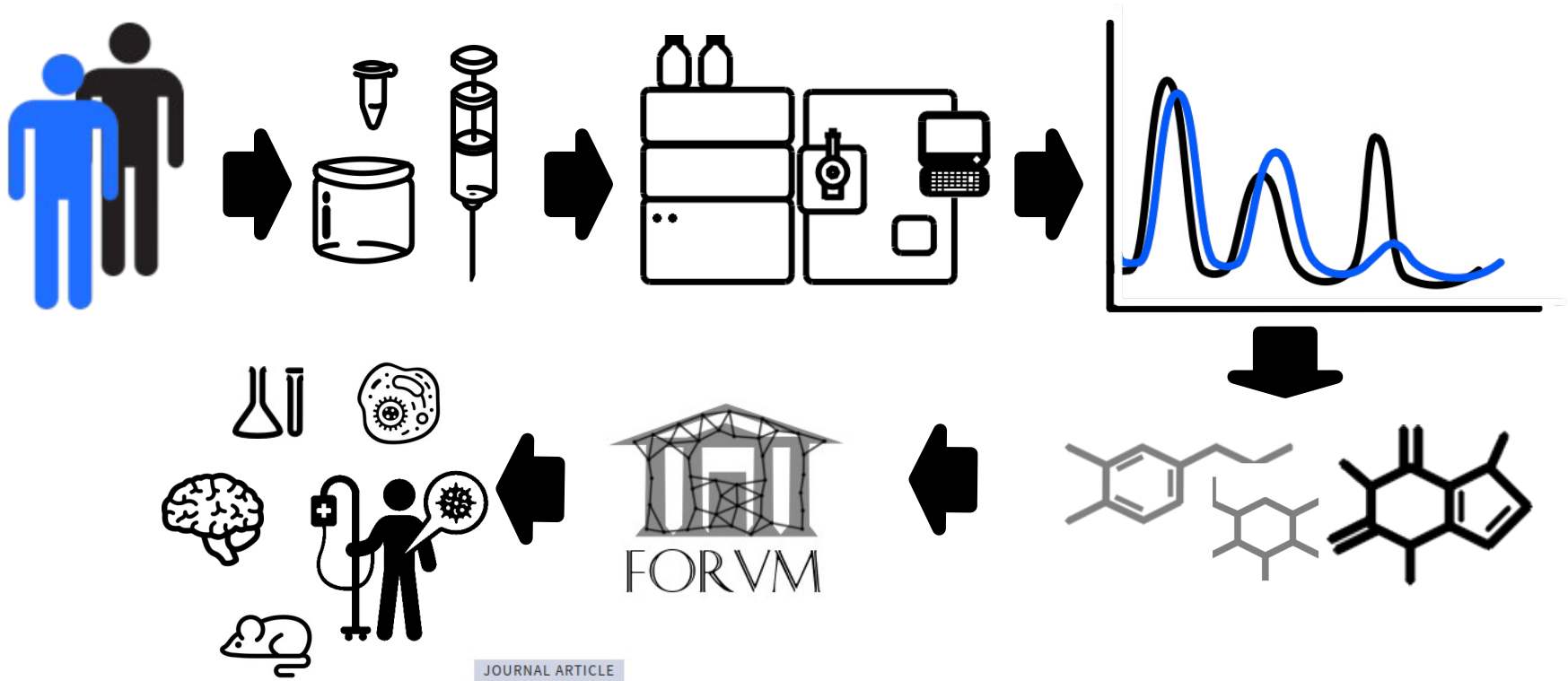
Arthur J. L. Cooper[1] and **Tomiko Kuhara**[2,*]





Beyond metabolism

Proximity within knowledge graphs



JOURNAL ARTICLE

FORUM: building a Knowledge Graph from public databases and scientific literature to extract associations between chemicals and diseases

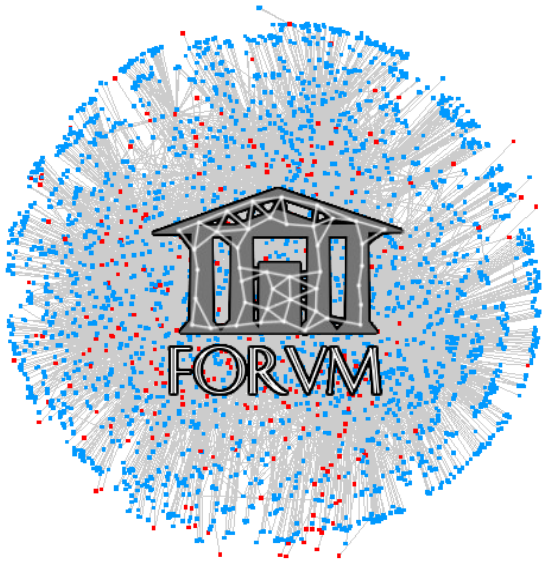
Maxime Delmas, Olivier Filangi, Nils Paulhe, Florence Vinson, Christophe Duperier, William Garrier, Paul-Emeric Saunier, Yoann Pitarch, Fabien Jourdan, Franck Giacomoni ... Show more

Bioinformatics, Volume 37, Issue 21, November 2021, Pages 3896–3904, https://doi.org/10.1093/bioinformatics/btab627

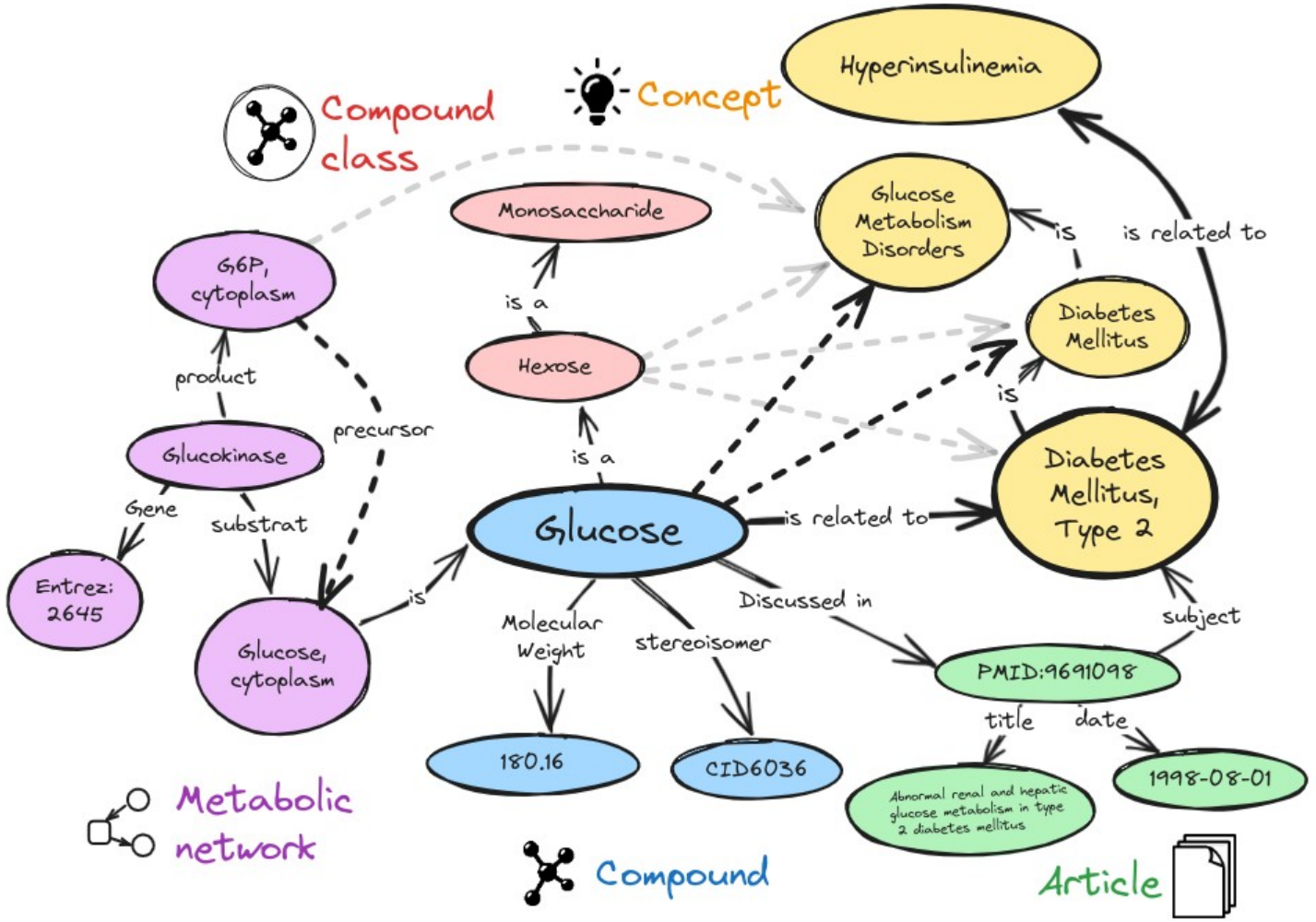
Published: 03 September 2021 **Article history**

Beyond metabolism

Proximity within knowledge graphs



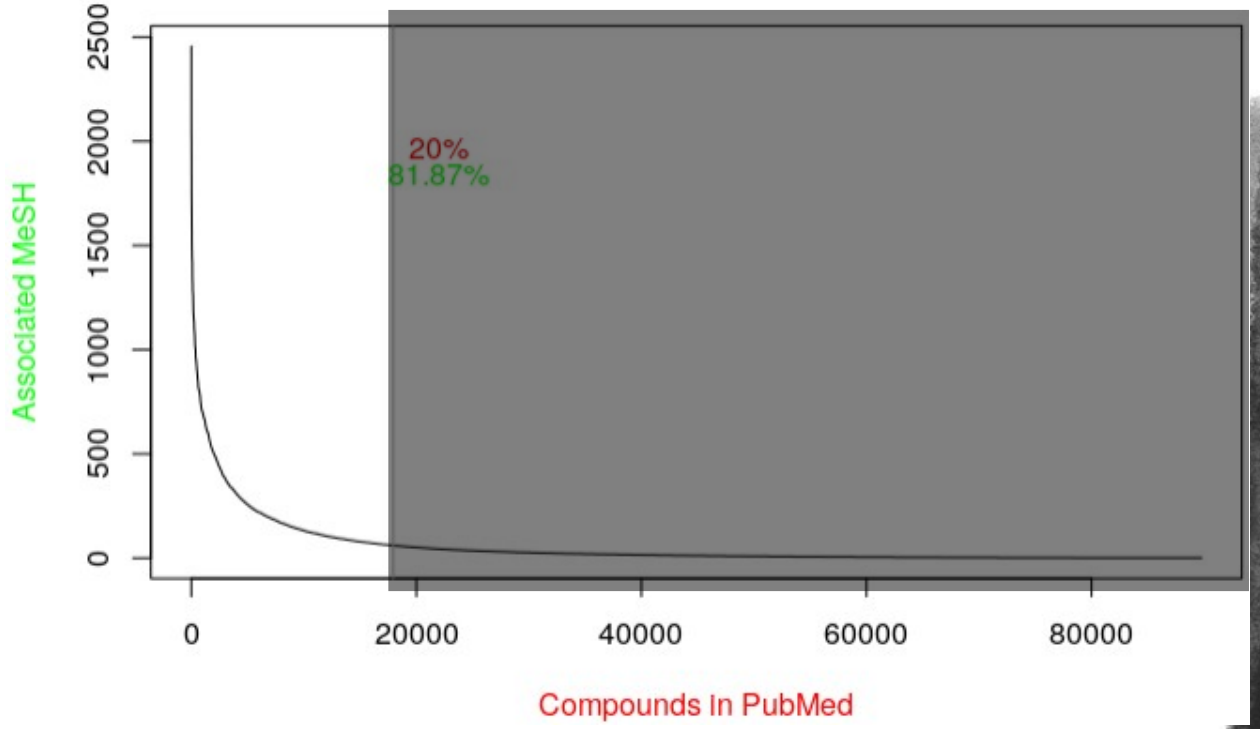
- 13 484 751 scientific articles
- 102 704 476 chemical compounds
- 24 353 MeSH descriptors
- 6 689 ChEBI classes
- **82 248 460 chemical - biomedical concepts links inferred**

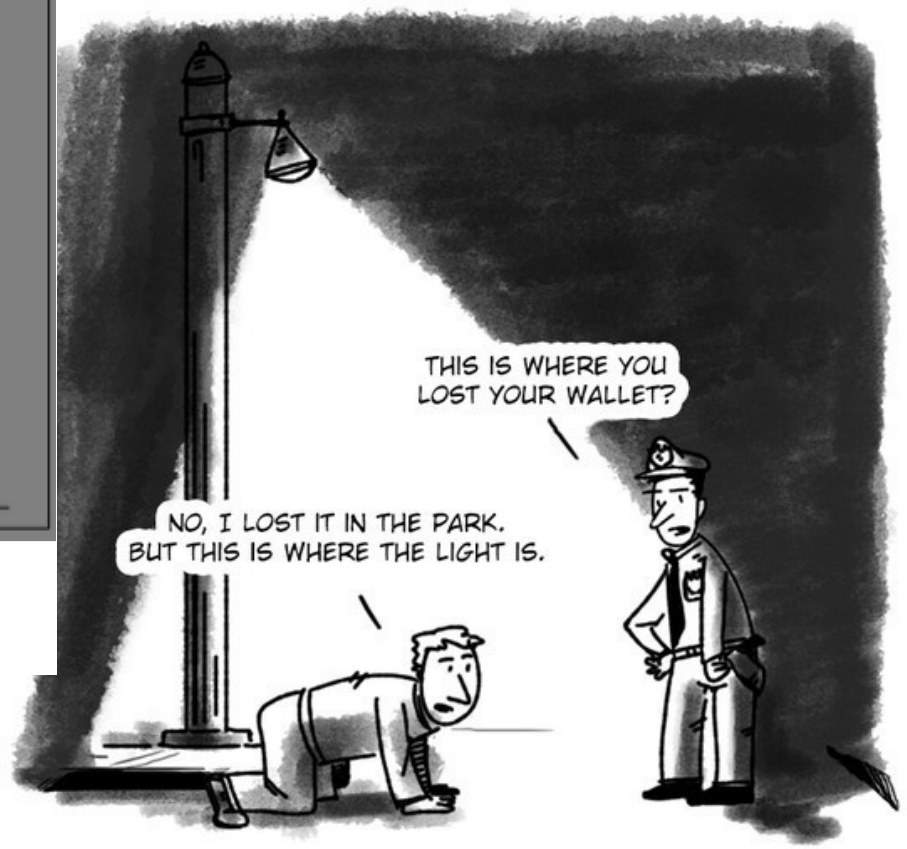


https://forum-webapp.semantic-metabolomics.fr/

Beyond metabolism

Literature bias in metabolism research?





Beyond metabolism

Leveraging metabolic proximity for semantic annotation

JOURNAL ARTICLE

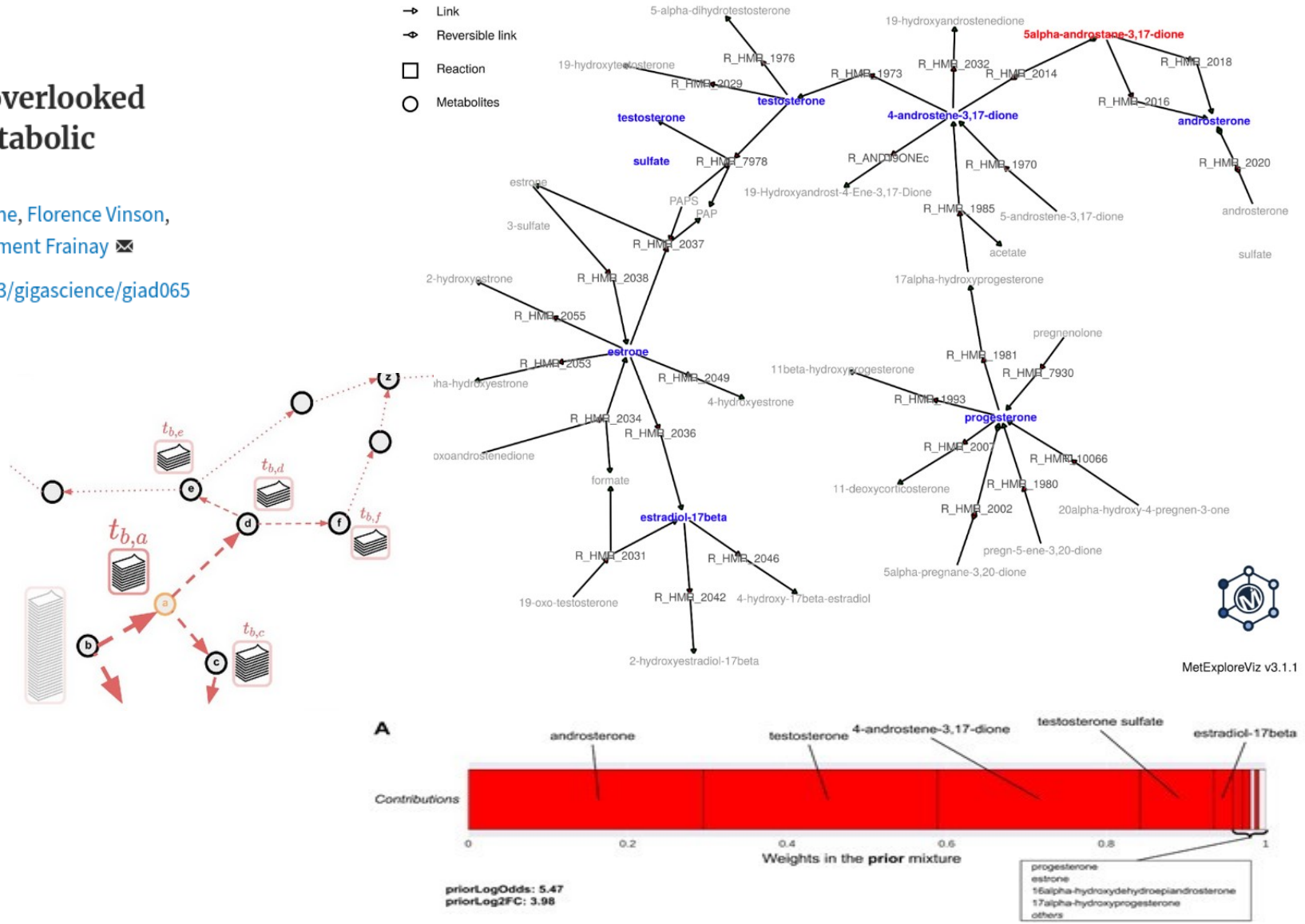
Suggesting disease associations for overlooked metabolites using literature from metabolic neighbors

Maxime Delmas, Olivier Filangi, Christophe Duperier, Nils Paulhe, Florence Vinson, Pablo Rodriguez-Mier, Franck Giacomoni, Fabien Jourdan, Clément Frainay

GigaScience, Volume 12, 2023, giad065, https://doi.org/10.1093/gigascience/giad065

Published: 15 September 2023 **Article history**

Using a Bayesian framework leveraging metabolic proximity, we suggested more than 35,000 associations between 1,047 overlooked metabolites and 3,288 diseases (or disease families)



Beyond metabolism

Leveraging metabolic proximity for semantic annotation

JOURNAL ARTICLE

Suggesting disease associations for overlooked metabolites using literature from metabolic neighbors

Maxime Delmas, Olivier Filangi, Christophe Duperier, Nils Paulhe, Florence Vinson, Pablo Rodriguez-Mier, Franck Giacomoni, Fabien Jourdan, Clément Frainay

GigaScience, Volume 12, 2023, giad065, https://doi.org/10.1093/gigascience/giad065
Published: 15 September 2023 **Article history**

Using a Bayesian framework leveraging metabolic proximity, we suggested more than 35,000 associations between 1,047 overlooked metabolites and 3,288 diseases (or disease families)

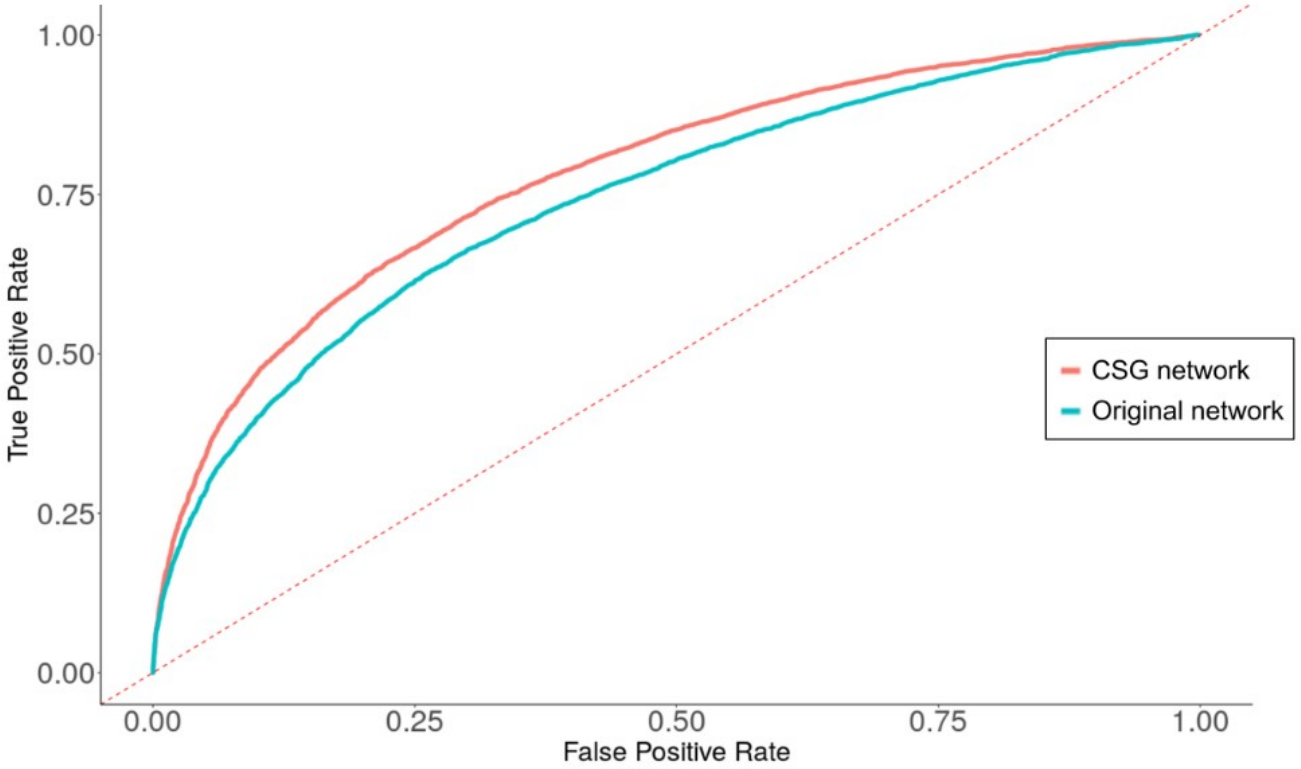


Figure S10: Receiver operating characteristic (ROC) curves using Log_2FC as predictor with the carbon skeleton graph (CSG network) or the original Human1 metabolic network. The AUC are respectively 0.78 and 0.74. The red dotted line corresponds to random strategies.

Thanks!

Maxime Delmas

Meije Mathé

Fabien Jourdan

Equipe Metabolisme & Xenobiotiques

Florence Vinson

Ludovic Cottret

Jean-Clément Gallardo

Louison Fresnais

Toxalim
Food Toxicology Research Institute

Olivier Filangi

Plateau de Profilage Métabolique et Métabolomique

Guillaume Laisney

PlateForme Exploration du Métabolisme

Nils Paulhe

Franck Giacomoni

METABOHUB

DOI:10.1093/gigascience/giad065
DOI:10.1093/bioinformatics/btab627
DOI:10.1093/bioinformatics/bty577
DOI:10.1093/bib/bbv115