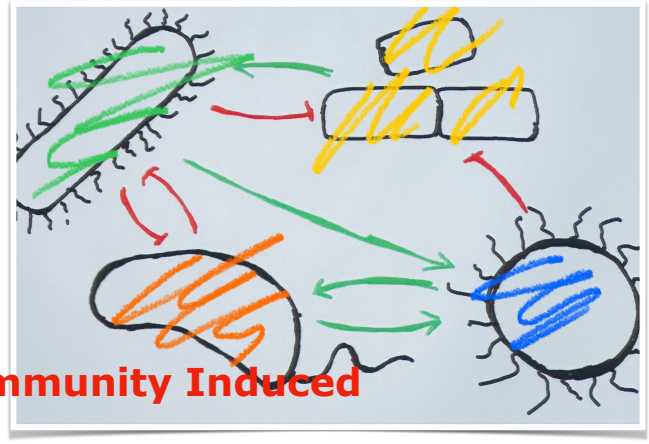
MATOMIC Unlocking network scalability in communities

The coarse graining concept in metabolic modeling

Rupert Tscheliessnig, Branko Ristivojcevic, Xtof Flamm, 12. February 2024

Motivation

MATOMIC



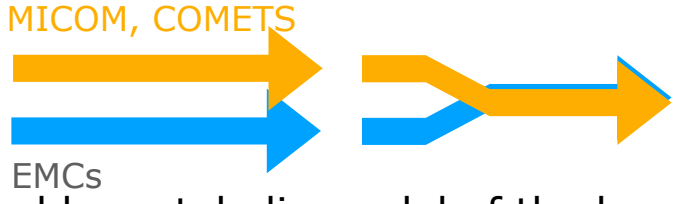
Mathematical Modelling for Microbial Community Induced Metabolic Diseases

Treatment-related interventions that change the structure and composition of gut bacteria among individuals.

Combining metabolic modeling techniques with experimental cultivation of microbiomes of different complexity, to design stable microbial communities for therapeutic use.

First approach

For systematic coarse graining (lump the networks)



MICOM , a customizable metabolic model of the human gut microbiome. COMETS on stoichiometric modeling of individual microbial species, and on a discrete approximation of convection diffusion equations (Cobrapy).

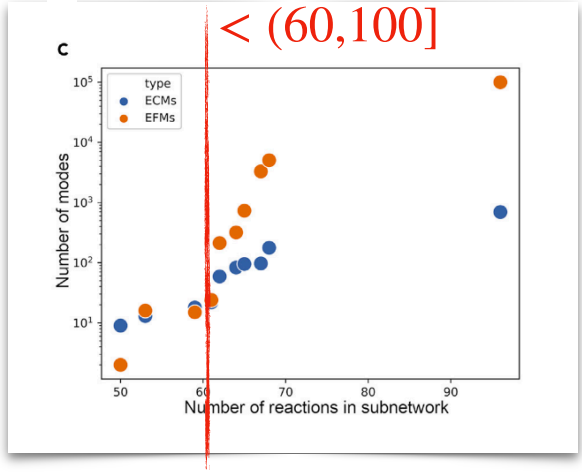
Complement **ecmtool** by the Python Community metabolic Modelling package, **PyCoMo** and benchmark against memory-efficient enumeration of elementary conversion modes.

Why ECMs …

But identifying ECMs is network dependent

Identification of the main routes in metabolic modeling of microbial community interactions. The **current golden standard** ist the computation of **ECMs**. Cells have orders of magnitude fewer ECMs than flux routes (EFMs)

The number of elementary conversion modes in the e_coli_core model37 reduces from 100,274 EFMs to 689 ECMs.



Clement, Tom J., Erik B. Baalhuis, Bas Teusink, Frank J. Bruggeman, Robert Planqué, and Daan H. De Groot. "Unlocking Elementary Conversion Modes: Ecmtool Unveils All Capabilities of Metabolic Networks." *Patterns* 2, no. 1 (January 2021): 100177. https://doi.org/10.1016/j.patter.2020.100177.

Identifying Elementary Conversion Modes (ECM)

Formally, it is a Fourier-Motzkin elimination method

To eliminate a metabolite, we can rewrite the constraints in terms of the metabolite.

$$S_{i,j} \cdot v_j \leq 0$$

$$\begin{cases} 2x - 5y + 4z \leqslant 10 \\ 3x - 6y + 3z \leqslant 9 \\ -x + 5y - 2z \leqslant -7 \\ -3x + 2y + 6z \leqslant 12 \end{cases}$$

$$\begin{cases} x \leqslant \frac{10 + 5y - 4z}{2} \\ x \leqslant \frac{9 + 6y - 3z}{3} \\ x \geqslant 7 + 5y - 2z \\ x \geqslant \frac{-12 + 2y + 6z}{3} \end{cases}$$

$$\begin{cases} 7 + 5y - 2z \leqslant \frac{10 + 5y - 4z}{2} \\ 7 + 5y - 2z \leqslant \frac{9 + 6y - 3z}{3} \\ \frac{-12 + 2y + 6z}{3} \leqslant \frac{10 + 5y - 4z}{2} \\ \frac{-12 + 2y + 6z}{3} \leqslant \frac{9 + 6y - 3z}{3} \end{cases}$$

$$4(n/4)^{2^d}$$

How to decide which metabolite should be taken out as the algorithm produces many unnecessary constraints (constraints that are implied by complement). Constraints can be minimised using FBA.

The annotation problem

8 communities and their missing annotations.

Good and bad

Anaerostipes_caccae_DSM_14662
Anaerostipes_caccae_DSM_14662_NBmod
Bifidobacterium_longum_NCC2705
Blautia_producta_DSM_2950
Clostridium_butyricum_DSM_10702
Clostridium_ramosum_VPI_0427_DSM_1402
Lactobacillus_plantarum_subsp_plantarum_ATCC_14917

From 6 species max. 5 metabolites are not *annotated*, only. However approx. 300 metabolites are annotated by a single reference, poorly

Bacteroides_thetaiotaomicron_VPI_5482
Escherichia_coli_str_K_12_substr_MG1655
Escherichia_coli_str_K_12_substr_MG1655*

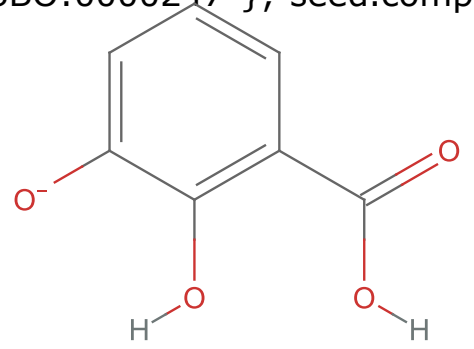
From 2 species 236 metabolites missing, but only 40 of these are not seed or sink metabolites

Solving the annotation issue, web crawling

The annotation of “2,3-dihydroxybenzoate”

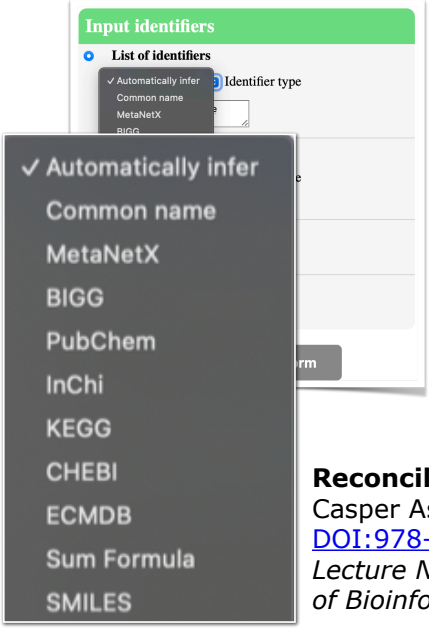
{9, {{"23dhb[c]", "2,3-dihydroxybenzoate"},

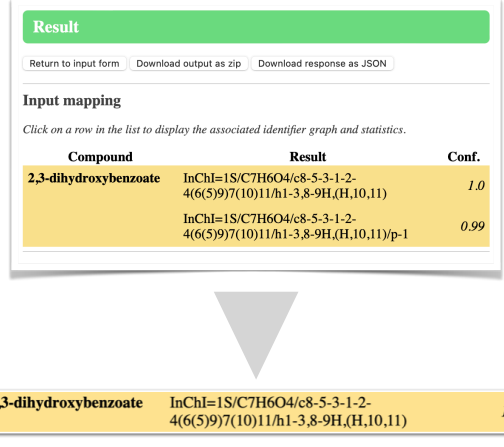
{“bigg.metabolite" -> "23dhb", "chebi" -> "CHEBI:18026", "hmdb" -> "HMDB0000397", **"inchi" -> "InChI=1S/C7H6O4/c8-5-3-1-2-4(6(5)9)7(10)11/h1-3,8-9H,(H,10,11)/p-1"**, "kegg.compound" -> "C00196", "metanetx.chemical" -> "MNXM455", "pubchem.compound" -> "19", "sbo" -> {"SBO:0000247"},"seed.compound" -> "cpd00168"}}}

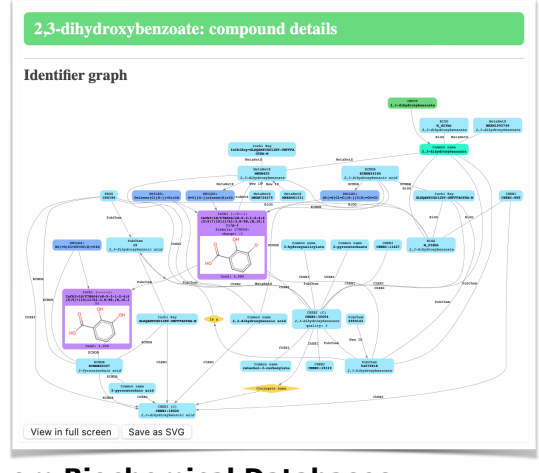


Annotations—structrecon

The metabolite: ”2,3-dihydroxybenzoate”



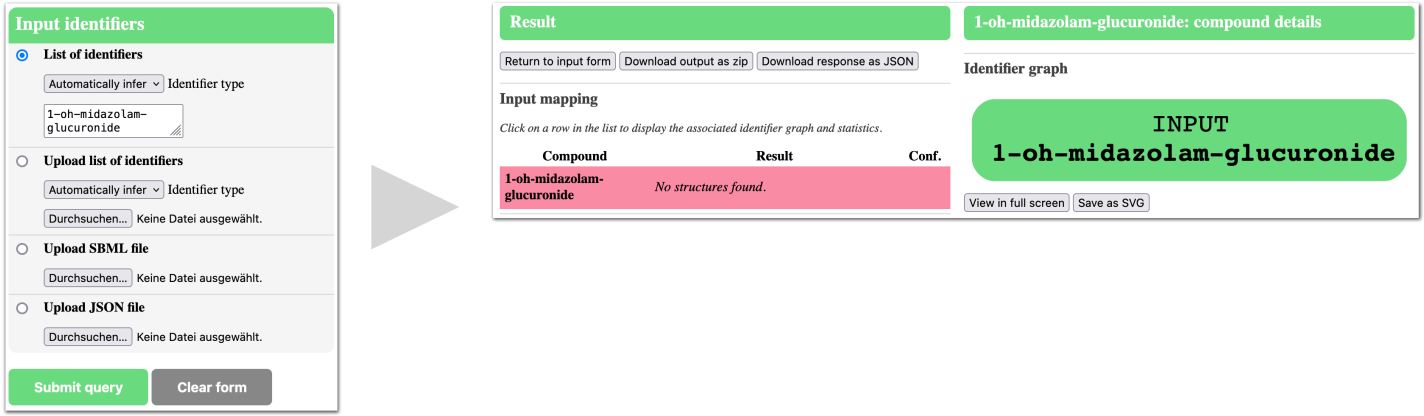




Reconciling Inconsistent Molecular Structures from Biochemical Databases
Casper Asbjørn Eriksen, Jakob Lykke Andersen, Rolf Fagerberg, Daniel Merkle (2023)
DOI:978-981-99-7074-2_5, ArXiV Preprint
Lecture Notes in Computer Science (LNBI, vol 14248), Proceedings of the International symposium of Bioinformatics Research and Applications (ISBRA 2023)

Annotations—structrecon

The metabolite: ”1-oh-midazolam-glucuronide”



Solving the annotation issue, web crawling

Combine Google PubChem for the annotation

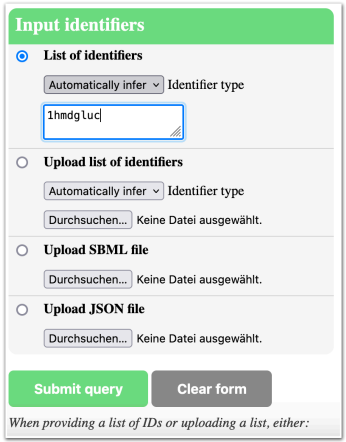
"1-oh-midazolam-glucuronide" // useGooglePubChem[]

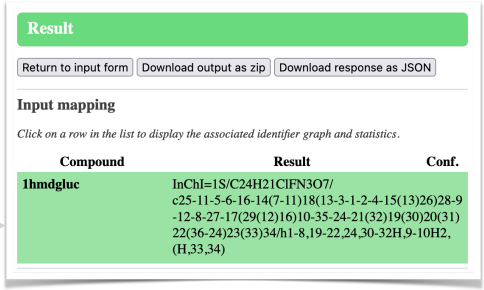
"midazolam-glucuronide", "Mdz-glucuronide"

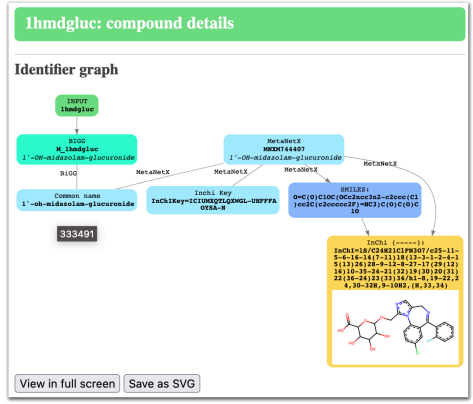
Formula	$C_{24}H_{21}ClFN_3O_7$
PubChemCompoundID	{ PubChem compound 133640 }
PubChemSynonyms	{ …21 }
InChI	InChI=1S/C24H21ClFN3O7/c25-11-5-6-16-14(7-11)18(13-3-1-2-4-15(13)26)28-9-12-8-27-17(29(12
CanonicalSMILES	O=C(O)C1OC(OCc2ncc3n2-c2ccc(Cl)cc2C(c2ccccc2F)=NC3)C(O)C(O)C1O

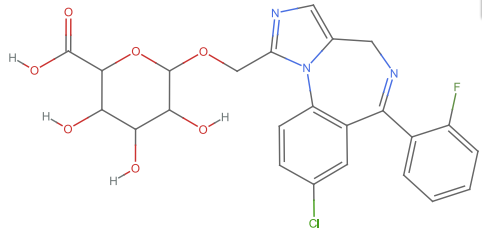
Annotations—structrecon

The metabolite: ”1-oh-midazolam-glucuronide”



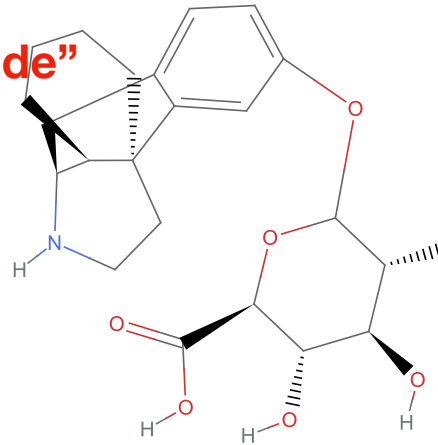






Annotations—structrecon

The metabolite: ”3-hydroxymorphinan o-glucuronide”



Input identifiers

- **List of identifiers**
  - Automatically infer — Identifier type
  - `3oh_mxn_glc`
- **Upload list of identifiers**
  - Automatically infer — Identifier type
  - Durchsuchen... Keine Datei ausgewählt.
- **Upload SBML file**
  - Durchsuchen... Keine Datei ausgewählt.
- **Upload JSON file**
  - Durchsuchen... Keine Datei ausgewählt.

Submit query | Clear form

Result

Return to input form | Download output as zip | Download response as JSON

Input mapping

Click on a row in the list to display the associated identifier graph and statistics.

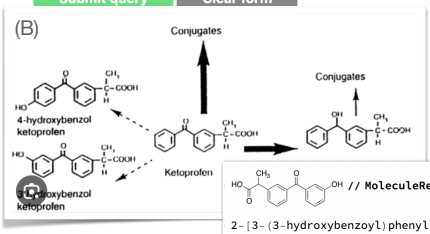
Compound	Result	Conf.
3oh_mxn_glc	*No structures found.*	

(A) "https://go.drugbank.com/structures/metabolites/DBMET01319.inchi"//Import//Molecule

Molecule [Formula: $C_{22}H_{29}NO_7$ Atoms: 59 Bonds: 63]

Molecule [
Formula: $C_{22}H_{29}NO_7$
Atoms: 59 Bonds: 63
SMILES: C1CC[C@@]23CCN[C@@]([H])(Cc4ccc(OC5([H])[C@@](O)([H])[C@](O)([H])[C@@](O)([H])[C@](C(=O)O)([H])O5)cc42)[C@@]3([H])C1
InChIKey: QYVMECDBFFRIEZ-DESFNQSRSA-N]

(B)



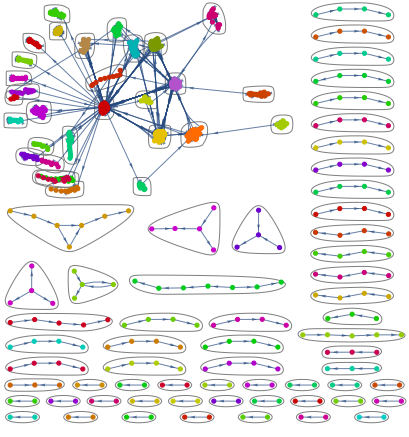
Communities detection

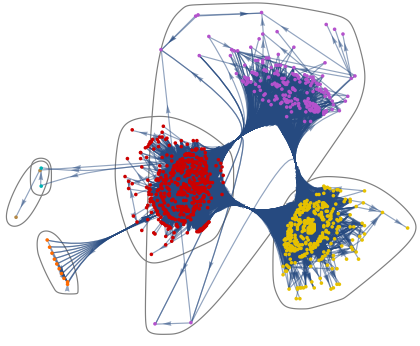
Tagging the reactions

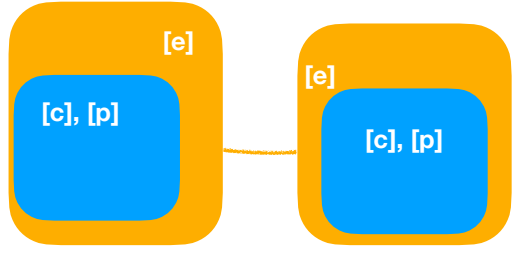
{{23dhmp [c][c]⟶ 3h3mop, 26dap_LL [c][c]⟶ glu_L, 26dap_LL [c][c]⟶ thdp, 2ahbut [c][c]⟶ 3h3mop, tyr_L [c][c]⟶ dad_5, tyr_L [c][c]⟶ imogly, tyr_L [c][c]⟶ met_L,
tyr_L [c][c]⟶ pcresol, 2ins [c][c]⟶ dkdi, 2mbcoa [c][c]⟶ 2mbutACP, 3btcoa [c][c]⟶ b2coa, 3hdecACP [c][c]⟶ tdec2eACP, 10m3hundecACP [c][c]⟶ 10mtundec2eACP,
··· 1174 ···, trp_L [c][c]⟶ biomass, tyr_L [c][c]⟶ PGPm1, tyr_L [c][c]⟶ apoACP, tyr_L [c][c]⟶ biomass, udcpdp [c][c]⟶ PGPm1, udcpdp [c][c]⟶ apoACP, udcpdp [c][c]⟶ biomass,
val_L [c][c]⟶ PGPm1, val_L [c][c]⟶ apoACP, val_L [c][c]⟶ biomass, zn2 [c][c]⟶ PGPm1, zn2 [c][c]⟶ apoACP, zn2 [c][c]⟶ biomass}, {··· 1 ···}, ··· 8 ···, {··· 1 ···}}

"[e]", "[c]", and "[p]" give the **compartmentalization** within biological systems or models, especially in the context of metabolic pathways:**"[e]" extracellular space or environment, "[c]" cytosol or cytoplasmic compartment, and "[p]" periplasmic space**

Anaerostipes_caccae_DSM_14662



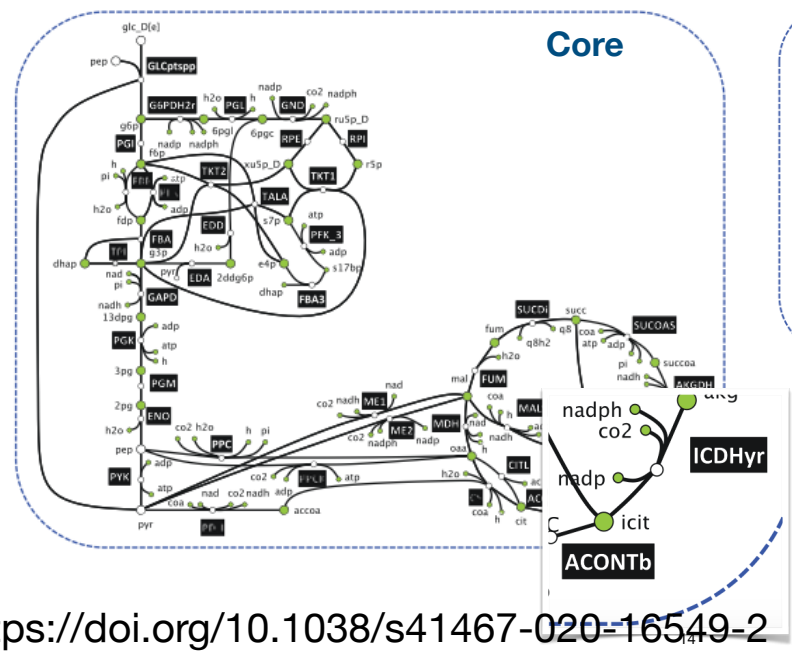




Identify clusters of metabolites (without hub metabolites) form clusters of reactions (lumping) — identify exchange metabolites

Communities detection

Splitting the reactions, workflow by Hatzimanikatis [1]



biomass: amino acids, lipids, cofactors …

https://doi.org/10.1371/journal.pcbi.1005513

[1] https://doi.org/10.1038/s41467-020-16549-2

Communities design

Splitting the reactions, workflow by Hatzimanikatis [1]

Carbohydrate metabolism:

Glycolysis/gluconeogenesis (**Gg**)

Pentose phosphate pathway (**PPP**)

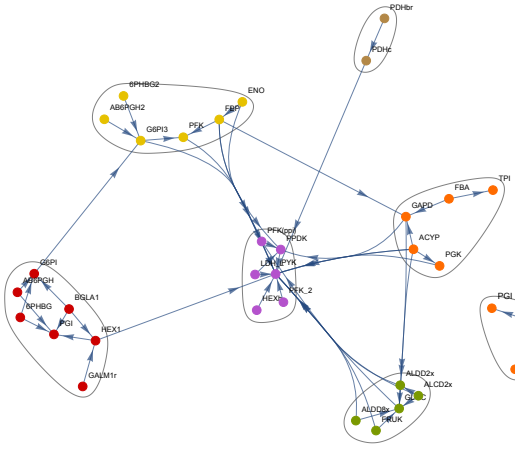
Central metabolism: the titric acid cycle (**TCA**)

Anaplerotic cycles are all MP that replenish the supply of intermediates in TCA cycle, and are crucial for maintaining adequate levels of TCA cycle metabolites. Anaplerotic cycles ensure the continuous operation of the TCA.

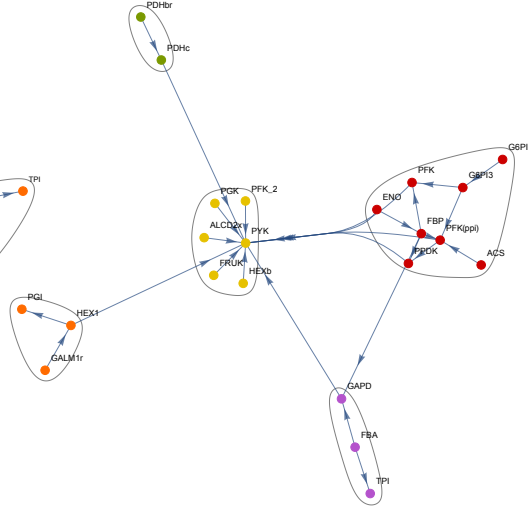
Communities detection

Pentose phosphate pathway (PPP) for 3 species

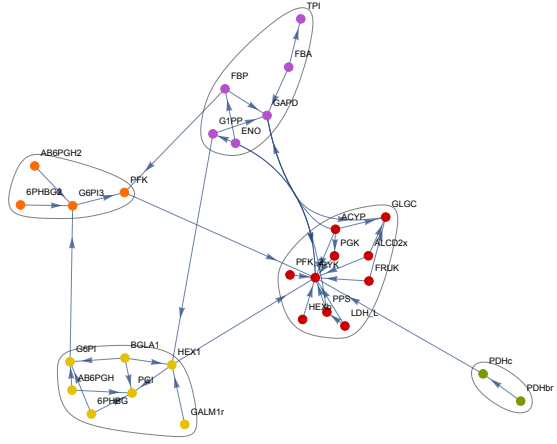
Anaerostipes_caccae_DSM_14662



Bacteroides_thetaiotaomicron_VPI_5482



Escherichia_coli _str_K_12_substr_MG1655



Colors indicate different reactions communities, only!

Outlook

Integration fo the redHUMAN workflow by Hatzimanikatis [1]

Thermodynamic Curation: Estimating Gibbs free energy to define reaction directionality.

Subsystem Selection: Choosing relevant metabolic processes for the study.

Network Expansion: Connecting initial subsystems to form a core metabolic network.

Extracellular Medium Connection: Linking extracellular medium components to the network.

Biosynthetic Reaction Generation: Identifying pathways for biomass building blocks.

Data Integration and Consistency Checks: Integrating experimental data and verifying model consistency.

https://www.yworks.com/products/yed/download

https://fluxer.umbc.edu/

[1] https://doi.org/10.1038/s41467-020-16549-2

Take home message

Begin with simplicity, since complexity will **naturally** evolve on its own.

Acknowldegment

Rupert Tscheliessnig, Branko Ristivojcevic, Xtof Flamm …
the Matomic @, & the TBI