

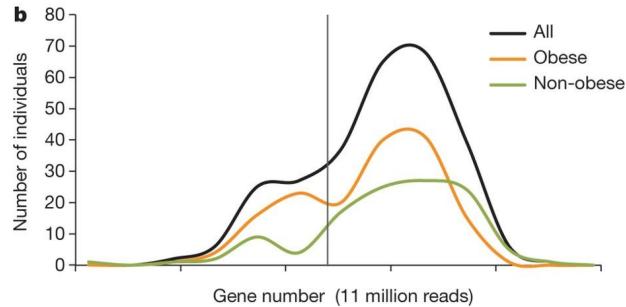
Functional role of gut microbiota

Metabolic network modelling via multi-omics analysis, and applications in metabolic diseases

Some slides and results are provided by A. Weber (PhD in Bioinformatics,
co-supervised with H. Soula)

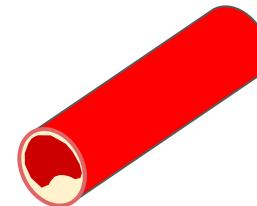
Importance of gut microbiota

- **Microbiota and metabolic diseases** : Influence in obesity and comorbidities



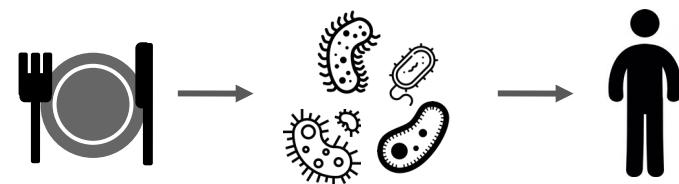
Obesity and microbiota gene richness

Le Chatelier, Nielsen, Qin, Prifti et al, 2013 : *Figure 1b*
Dao and Clément, 2018



Atherosclerosis and microbiota

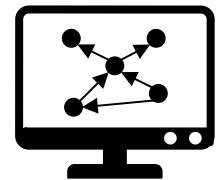
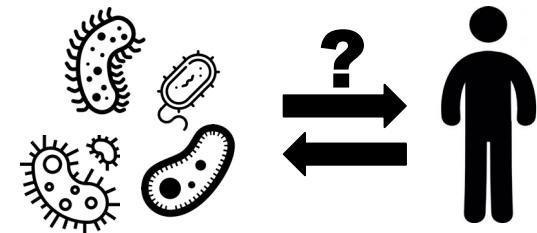
Karlsson et al, 2012



- Importance of **short chain fatty acids**
Morrison and Preston, 2016

Some biomedical challenges

- Functional properties mostly **unknown**
 - essentially microbiota/host **correlations**
- Description of functional roles
 - **focus on metabolism**
 - ◆ Computer science description of metabolism :
Metabolic networks
 - ◆ Systems biology : Study network properties

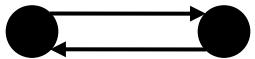


Acides gras à chaîne courte

- **Exemples :** Butyrate, acétate, propionate...
- Dérivées des **fibres alimentaires** (sucres complexes non digestibles par l'Homme) par le microbiote par fermentation
- **Fonctions pour l'Homme :** Substrats majeurs colonocytes (source d'énergie), déclenchement de cascades de signalisation (différenciation, prolifération, régulation synthèse cholestérol...), ...
- Retrouvé dans **nombreux tissus**

Metabolic networks

Networks : Objects (**nodes**) linked by links (**edges**).

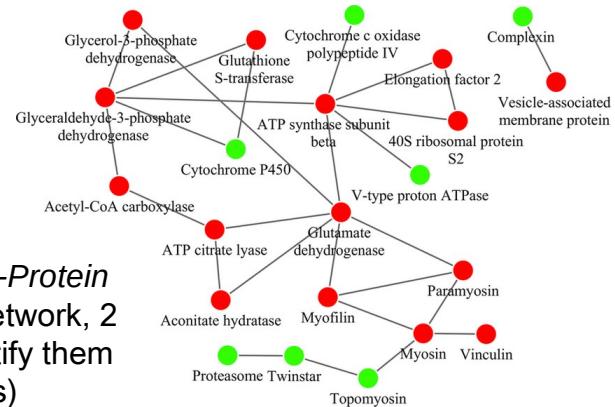


Directed



Undirected

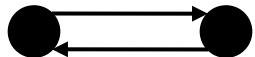
Ex : Protein-Protein
Interaction network, 2
ways to identify them
(colors)



Wang, 2015

Metabolic networks

Networks : Objects (**nodes**) linked by links (**edges**).

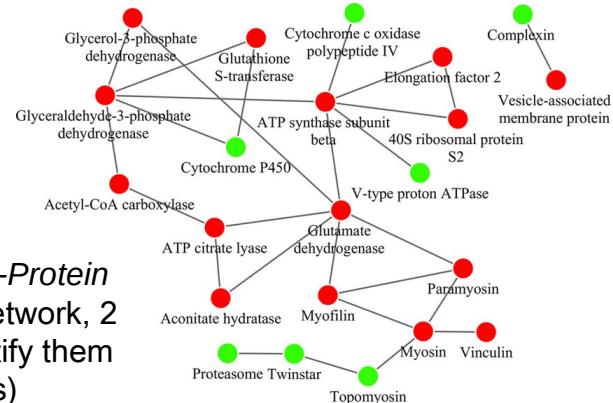


Directed



Undirected

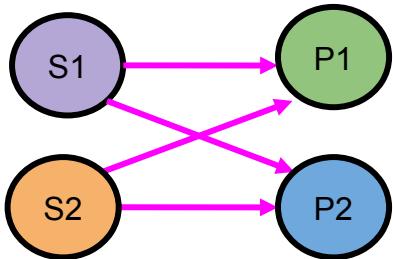
Ex : Protein-Protein Interaction network, 2 ways to identify them (colors)



Wang, 2015

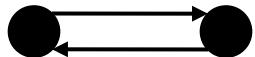
Metabolic networks : Representation of all chemical reactions

Eg : Molecules (objects), chemical reactions (links)



Metabolic networks

Networks : Objects (**nodes**) linked by links (**edges**).

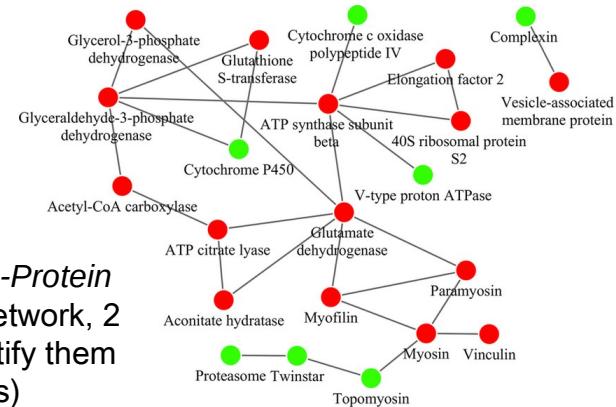


Directed



Undirected

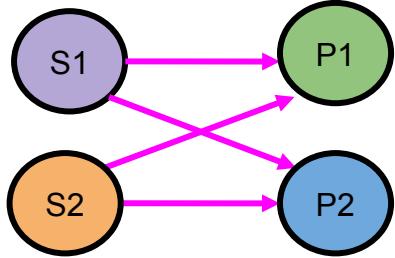
Ex : Protein-Protein Interaction network, 2 ways to identify them (colors)



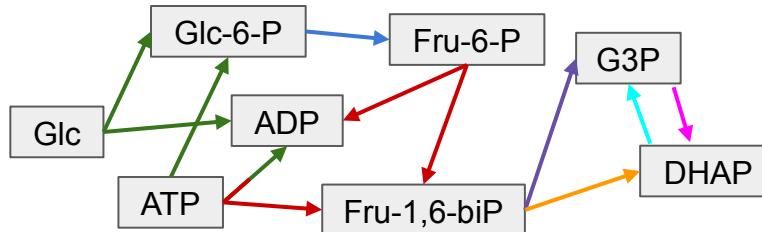
Wang, 2015

Metabolic networks : Representation of all chemical reactions

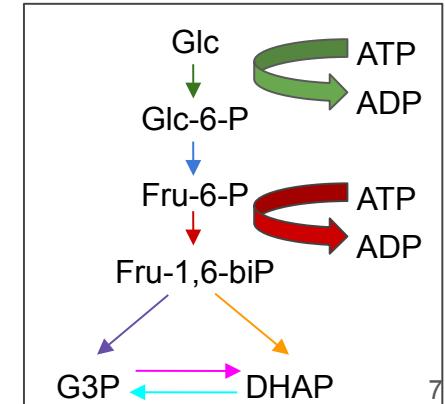
Eg : Molecules (objects), chemical reactions (links)



$S1 + S2 \xrightarrow{\text{Enzyme}} P1 + P2$



Glycolysis phase 1



Metabolic network reconstruction

b0001
b0002
...


Gene names
(in cDNA fasta files from
GenBank or Ensembl)

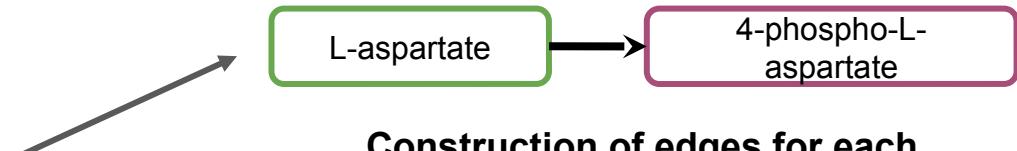
Escherichia coli K-12 MG1655; b0001		
 Escherichia coli K-12 MG1655; b0002		
Entry	b0002	CDS T00007
Gene name	thrA	
Definition	(RefSeq) Bifunctional aspartokinase/homoserine dehydrogenase 1	
KO	K12524 bifunctional aspartokinase / homoserine dehydrogenase 1 [EC:2.7.2.4 1.1.1.3]	
Organism	eco Escherichia coli K-12 MG1655	
Pathway	eco00260 Glycine, serine and threonine metabolism eco00261 Monobactam biosynthesis	

* Kyoto Encyclopedia of Genes and Genomes

ENZYME: 1.1.1.3	
 ENZYME: 2.7.2.4	
Entry	EC 2.7.2.4 Enzyme
Name	aspartate kinase; aspartokinase; AK; beta-aspartokinase; aspartic kinase
Class	Transferases; Transferring phosphorus-containing groups; Phosphotransferases with a carboxy group as acceptor BRITE hierarchy
Sysname	ATP:L-aspartate 4-phosphotransferase
Reaction(IUBMB)	ATP + L-aspartate = ADP + 4-phospho-L-aspartate [RN:R00480]
Reaction(KEGG)	R00480
Substrate	ATP [CPD:C00002]; L-aspartate [CPD:C00049]
Product	ADP [CPD:C00008]; 4-phospho-L-aspartate [CPD:C03082]

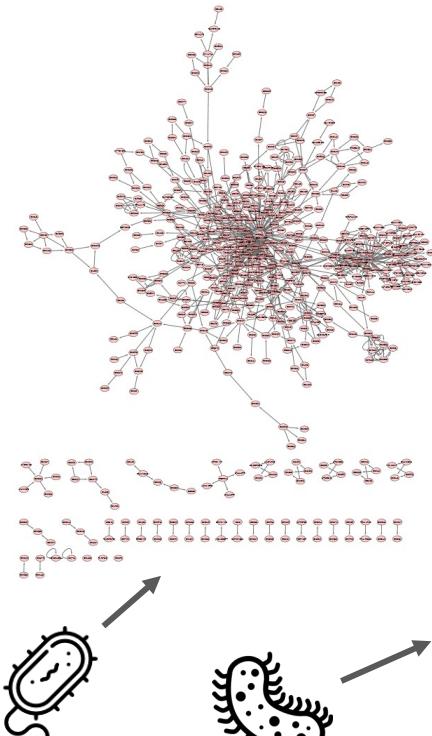
**Enzyme entries in KEGG* database
(associated to ECs)**

Gene entries in KEGG* database

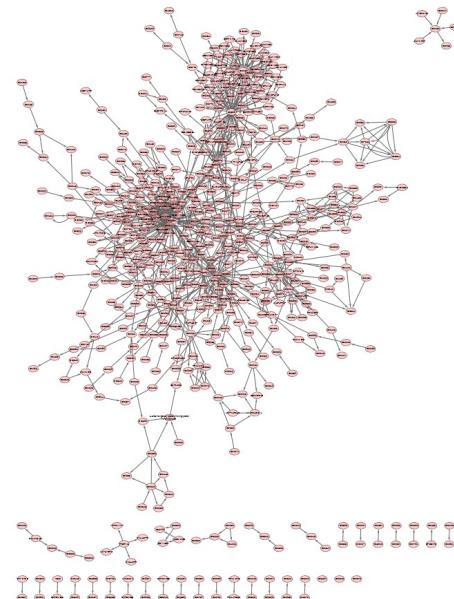


Construction of edges for each reaction

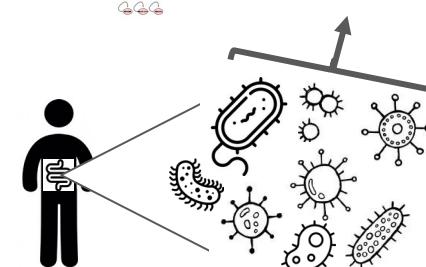
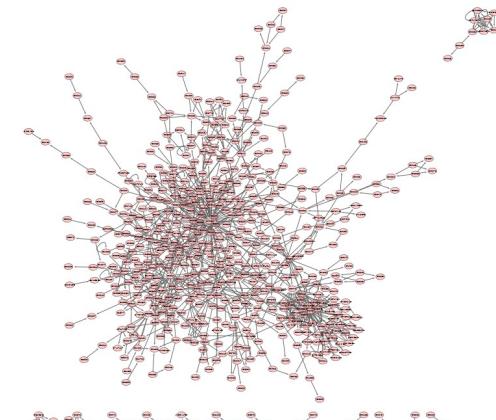
Metabolic network reconstruction



~ mean of 700-800 nodes
~ mean of 1000-1500 edges



Per species

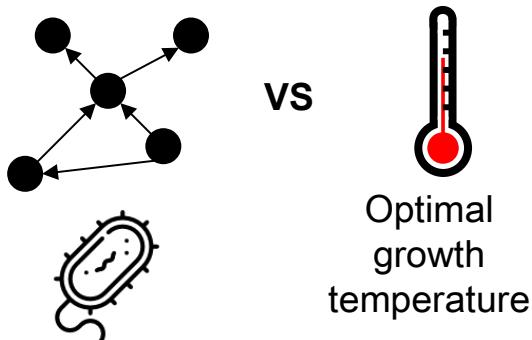


Per
microbiota

➡ Complex mathematical objects

Some state-of-the-art results

- **Aim** : Link network structural properties with environmental phenotypes
- **Proof of concept**, simple and strong phenotype



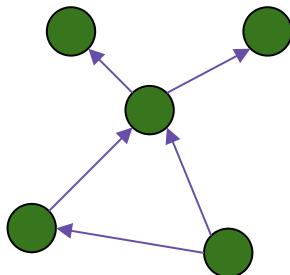
STEPS :

1. Network reconstruction
2. Measure metabolic network properties
3. Analyse relations between properties and temperature

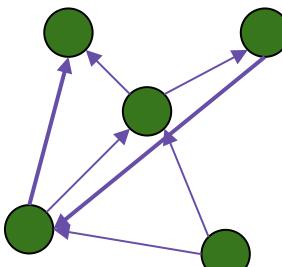
Examples of structural network properties

Edge density =

$$\frac{2 \# \text{ edges}}{\# \text{ nodes}}$$



$$\frac{2 \times 5}{5} = 2$$



$$\frac{2 \times 7}{5} = 2.8$$

Mean number of edges
per node



Mean number of
reactions per
metabolite

Short summary of Takemoto et al., 2007

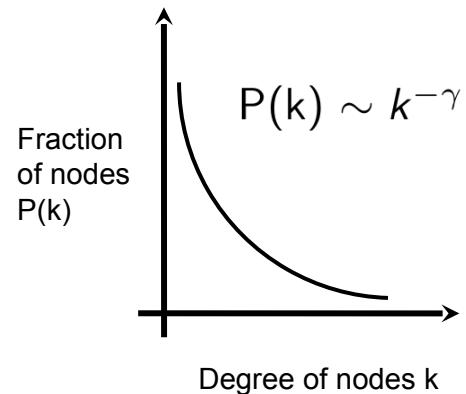
- 113 prokaryotic species
- Network properties **positively** or **negatively correlated** to optimal growth temperature:

- Edge density : Edges ÷ Nodes
- Maximum likelihood estimate of degree exponent :

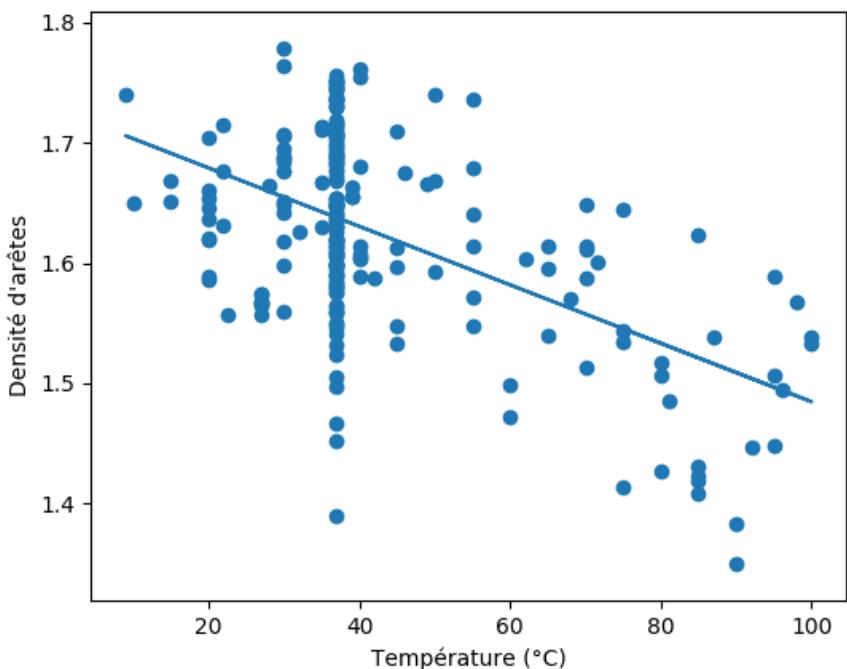
$$\gamma = 1 + N \times \left[\sum_{i=1}^N \ln \frac{k_i}{k_{min}} \right]^{-1}$$

- Average clustering coefficient :

$$\frac{1}{N} \sum_{i=1}^N \frac{M_i}{M_{possible}}$$



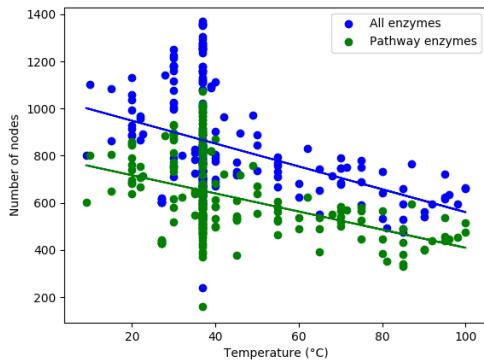
Topological structure and phenotypes



Pearson's $r = -0.54$, $p\text{-value} < 10^{-17}$

- Example of correlation with growth temperature for **228 prokaryotes from different environments**
 - Mean number of reactions per metabolite **decreases** with temperature
- **The state-of-the-art results validated**

Other network properties and temperature



Number of nodes vs Temperature

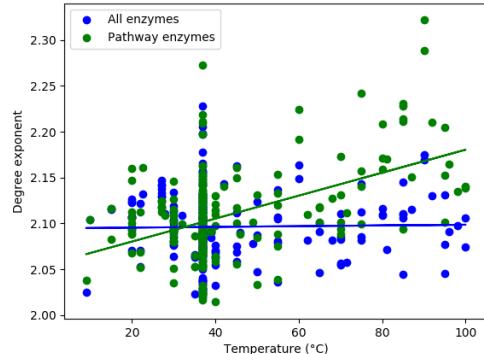
All enzymes
 $r = -0.40, p < 10^{-9}$

Pathway enzymes
 $r = -0.41, p < 10^{-10}$

All enzymes
 $r = -0.07, p \sim 0.30$

Pathway enzymes
 $r = -0.54, p < 10^{-17}$

Pearson's r through linear regression



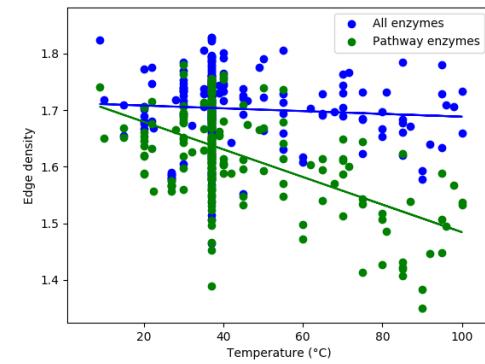
Degree exponent vs Temperature

All enzymes
 $r = 0.02, p \sim 0.76$

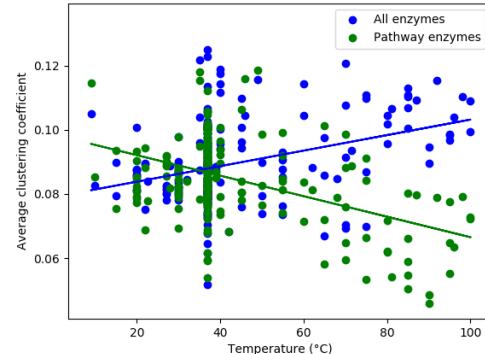
Pathway enzymes
 $r = 0.47, p < 10^{-13}$

All enzymes
 $r = 0.36, p < 10^{-7}$

Pathway enzymes
 $r = -0.46, p < 10^{-12}$



Edge density vs Temperature

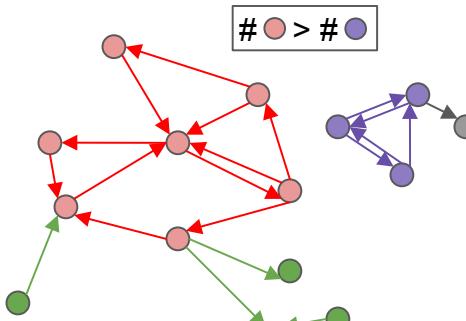


Average clustering coefficient vs Temperature

Directed network properties proposition

Strongly connected components:

group of nodes that can access all of the others in the group through *directed* edges (arrows) . &



Weakly connected components:

group of nodes that can access all of the others in the group through *undirected* edges (lines) .



Node normalised size of largest strongly connected component:

$$\frac{\# \text{ red}}{\# \text{ red} + \# \text{ green} + \# \text{ purple} + \# \text{ grey}}$$

Node normalised flow hierarchy:

$$\frac{(\# \text{ red} + \# \text{ green}) - \# \text{ red}}{\# \text{ red} + \# \text{ green} + \# \text{ purple} + \# \text{ grey}}$$

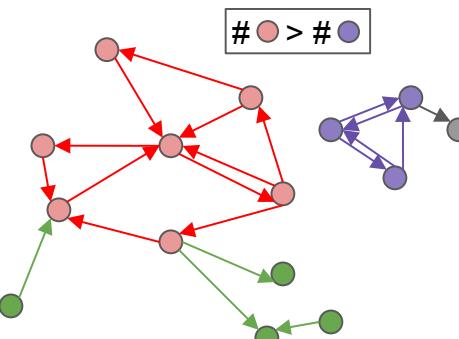
Directed network properties proposition

Strongly connected components:

group of nodes that can access all of the others in the group through *directed* edges (arrows) . &



$$\# \text{ (pink)} > \# \text{ (purple)}$$

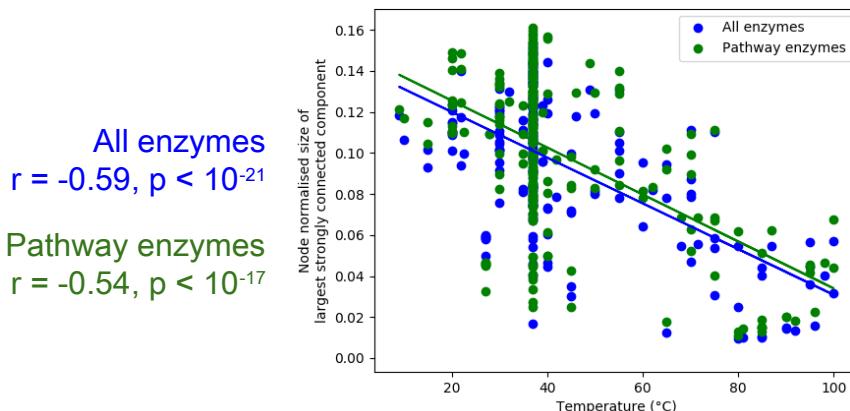


Weakly connected components:

group of nodes that can access all of the others in the group through *undirected* edges (lines) .



$$\# \text{ (green)} + \# \text{ (pink)} \quad \& \quad \# \text{ (purple)} + \# \text{ (grey)}$$



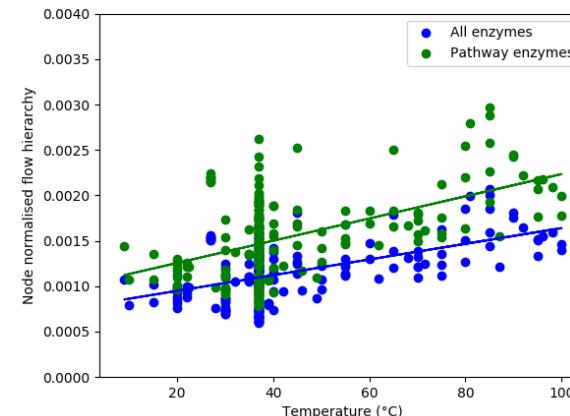
Node normalised size of largest strongly connected component vs Temperature

Node normalised size of largest strongly connected component:

$$\frac{\# \text{ (pink)}}{\# \text{ (pink)} + \# \text{ (green)} + \# \text{ (purple)} + \# \text{ (grey)}}$$

Node normalised flow hierarchy:

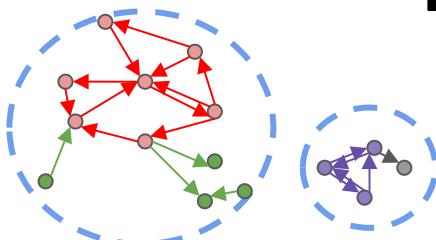
$$\frac{(\# \text{ (pink)} + \# \text{ (green)}) - \# \text{ (pink)}}{\# \text{ (pink)} + \# \text{ (green)} + \# \text{ (purple)} + \# \text{ (grey)}}$$



Node normalised flow hierarchy vs Temperature

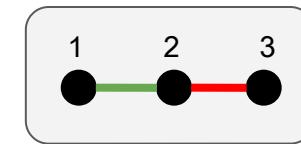
Laplacian spectrum : what is it?

- Networks can be represented as **adjacency matrices**
- **Spectrum** (= set of all **eigenvalues**) of **Laplacian matrix** (~adjacency matrix)
- # eigenvalues = # nodes
- Eigenvalues sorted in **decreasing order**
- # weakly connected components = # (eigenvalues = 0)



Idea of **flow** in network

2 weakly connected components

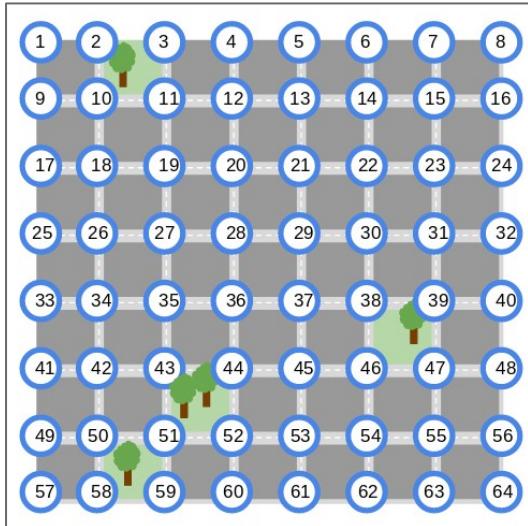
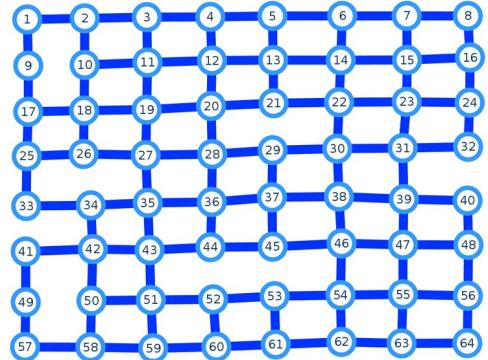


Nodes			
1	2	3	
1	0	1	0
2	1	0	1
3	0	1	0

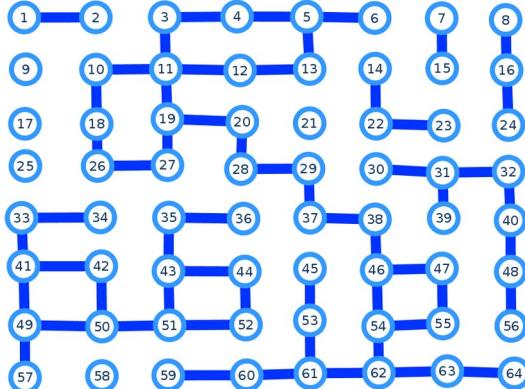
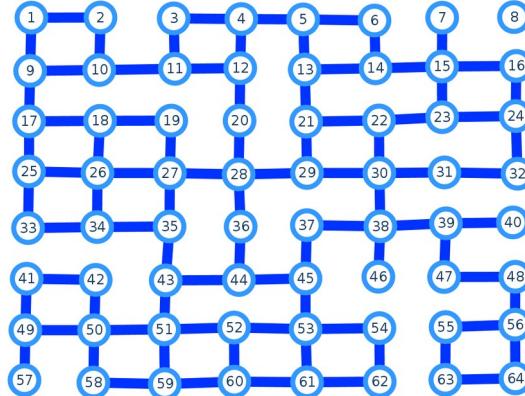
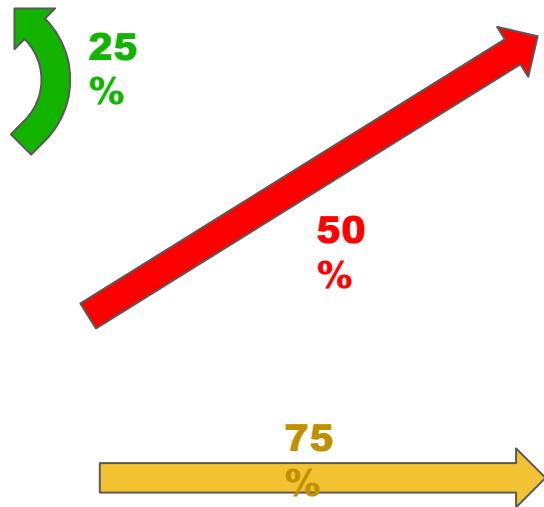
There is a link between 1 and 2

Adjacency matrix

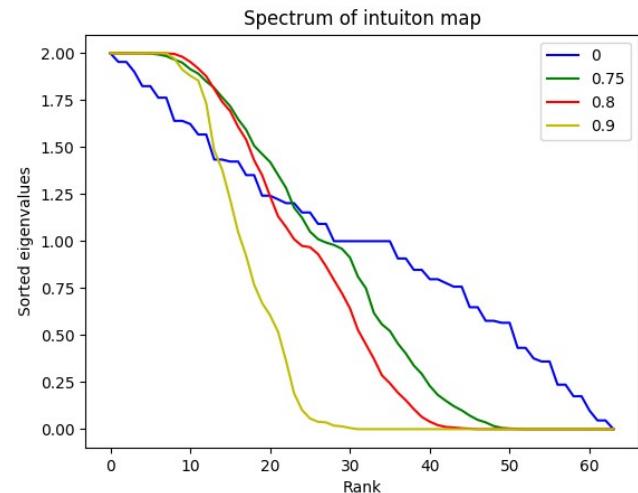
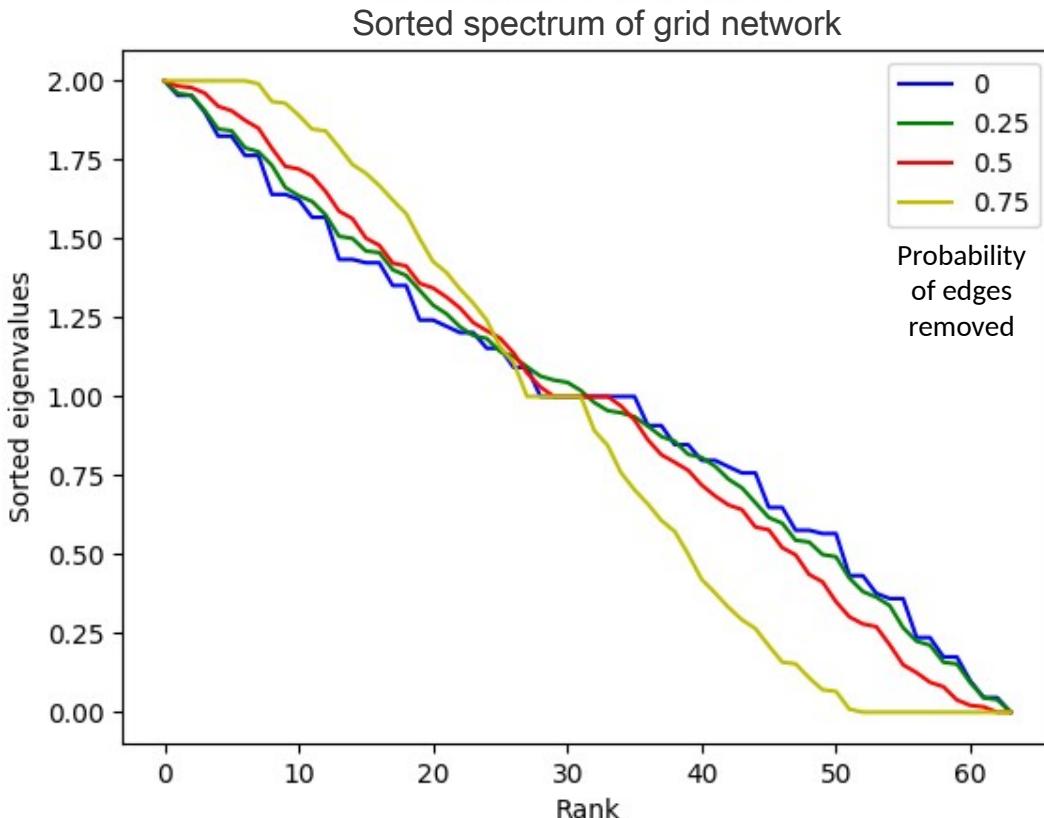
Spectrum intuition : grid network (1)



Probability of disappearance of random edges (roads) :



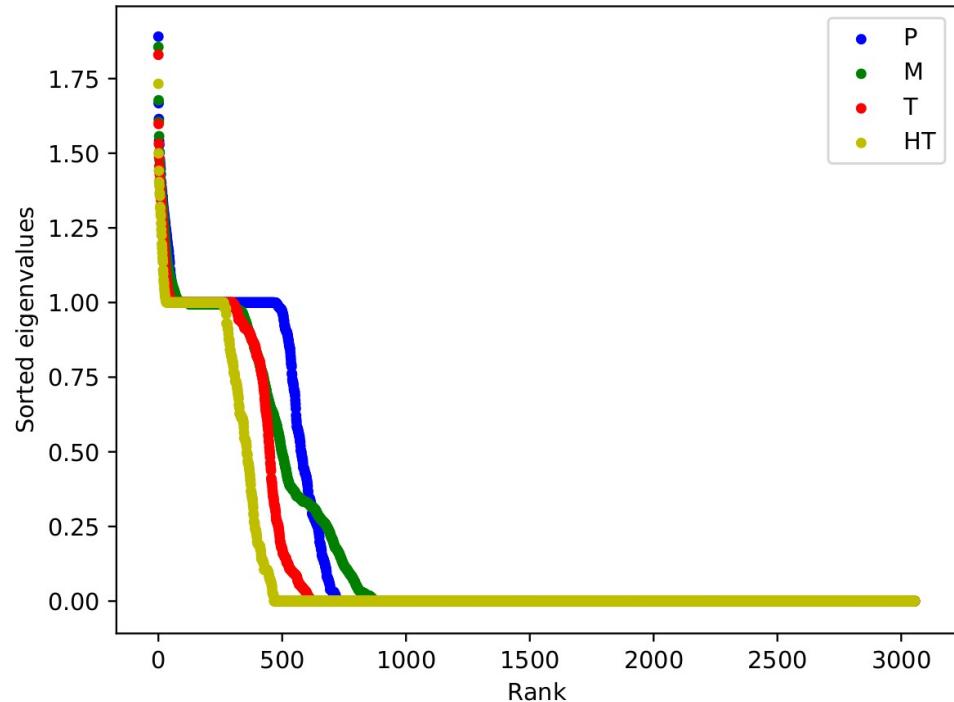
Spectrum intuition : grid network (2)



Laplacian spectrum on our 228 prokaryotes

- On our 228 prokaryotes from different environments
- Mean spectrum per temperature class

→ ↗ temperature
↔
quicker the decrease to 0

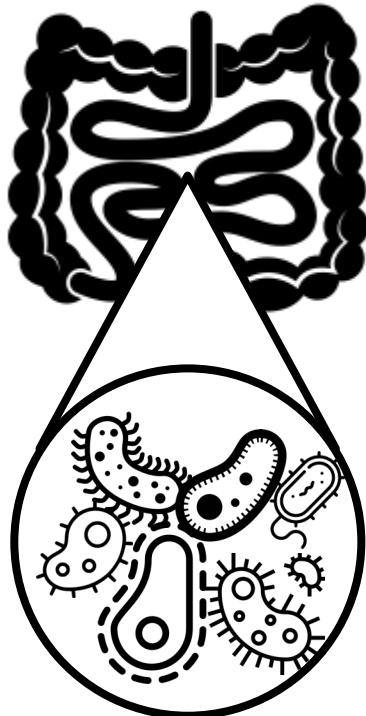


Microbaria

- 61 patientes à obésité sévère, candidats de chirurgie bariatrique (anneau gastrique modulable AGB, n = 20, ou bypass gastrique Roux-en-Y RYGB , n = 41)
- 24 patients suivis 1, 3 et 12 mois post-chirurgie (14 RYGB, 10 AGB).
- Variables cliniques, anthropométriques et biologiques ont été évaluées.
- Microbiote intestinal et métabolome sérique analysés par métagénomique *shotgun* et chromatographie liquide spectrométrie de masse.

Effectifs	Âge (ans)	Masse (kg)	BMI (kg/m²)	% femmes
61	36.9±9.86	123±18	45.6±5.23	100

Microbiote intestinal



- **Microbiote intestinal** : Micro-organismes colonisant les voies gastro-intestinales
- **Fonction** : Maintien d'une homéostasie immunitaire et métabolique, protection contre pathogènes



Nombre cellules humaines

Nombre bactéries



- Composé de bactéries, champignons, archées, virus, protozoaires
- Composition exacte hôte-spécifique, évoluant le long de la vie

Metabolic network reconstruction : Microbaria (1)

	KO1	KO2	KO3	KO4	KO5
Patients	0,0	...			
	...				

KO abundance matrix
(microbiota sequencing)

*KO = KEGG Orthology
= orthology genes*

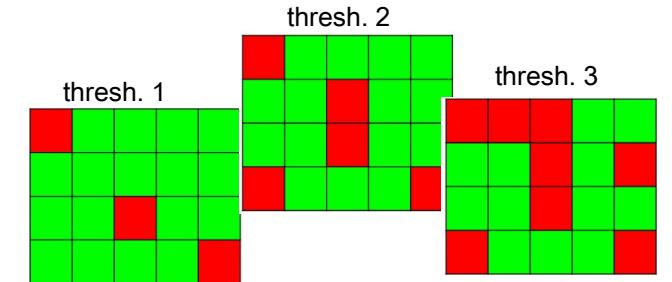
Metabolic network reconstruction : Microbaria (1)

	KO1	KO2	KO3	KO4	KO5
Patients	0,0	...			
1	Smiley				
2	Smiley				
3	Smiley				
4	Smiley				
5					

KO abundance matrix

*KO = KEGG Orthology
= orthology genes*

→
Threshold



Matrix of KO presence/absence in patients
(above threshold, below threshold)

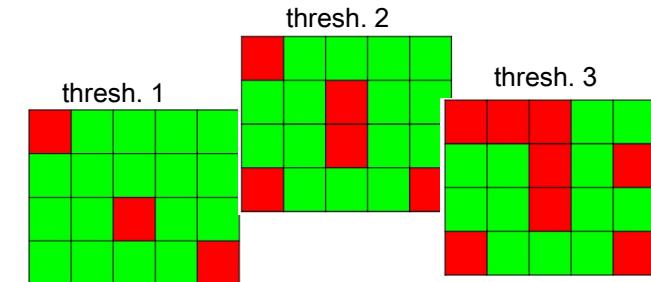
Metabolic network reconstruction : Microbaria (1)

	KO1	KO2	KO3	KO4	KO5
Patients	0,0	...			

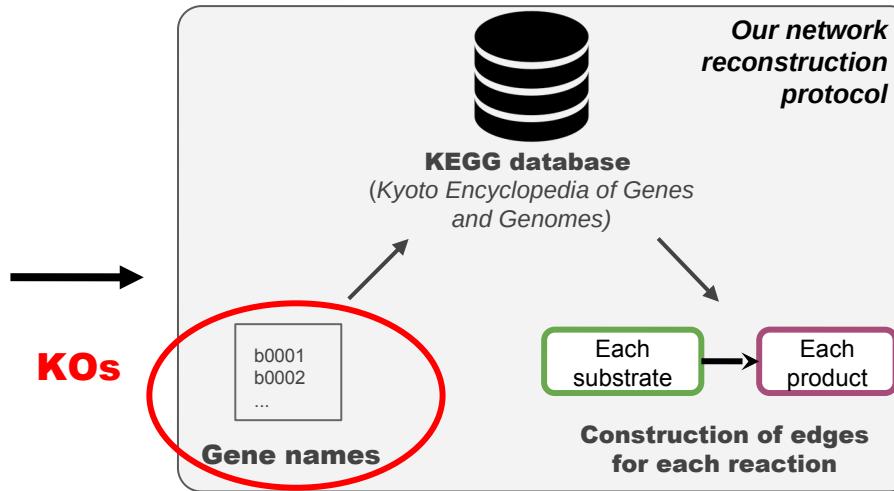
KO abundance matrix

*KO = KEGG Orthology
= orthology genes*

Threshold



Matrix of KO presence/absence in patients
(above threshold, below threshold)



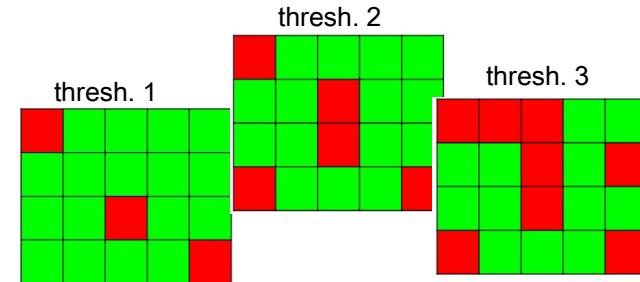
Metabolic network reconstruction : Microbaria (1)

	KO1	KO2	KO3	KO4	KO5
Patients	0,0	...			
1	Smiley				
2	Smiley				
3	Smiley				
4	Smiley				
5					
6					
7					
8					
9					

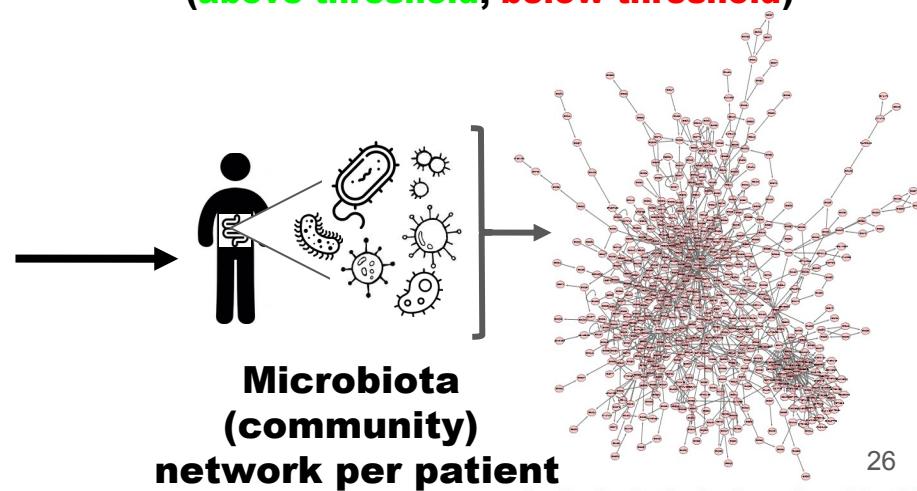
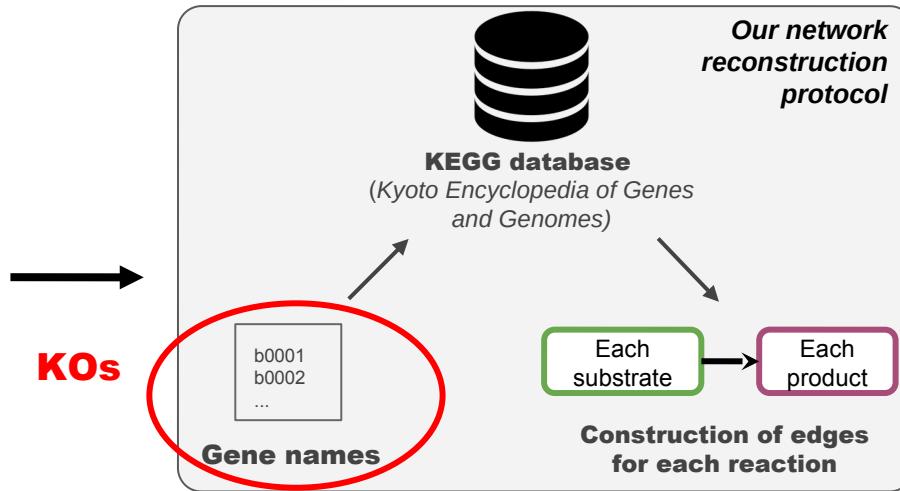
KO abundance matrix

KO = KEGG Orthology
= orthology genes

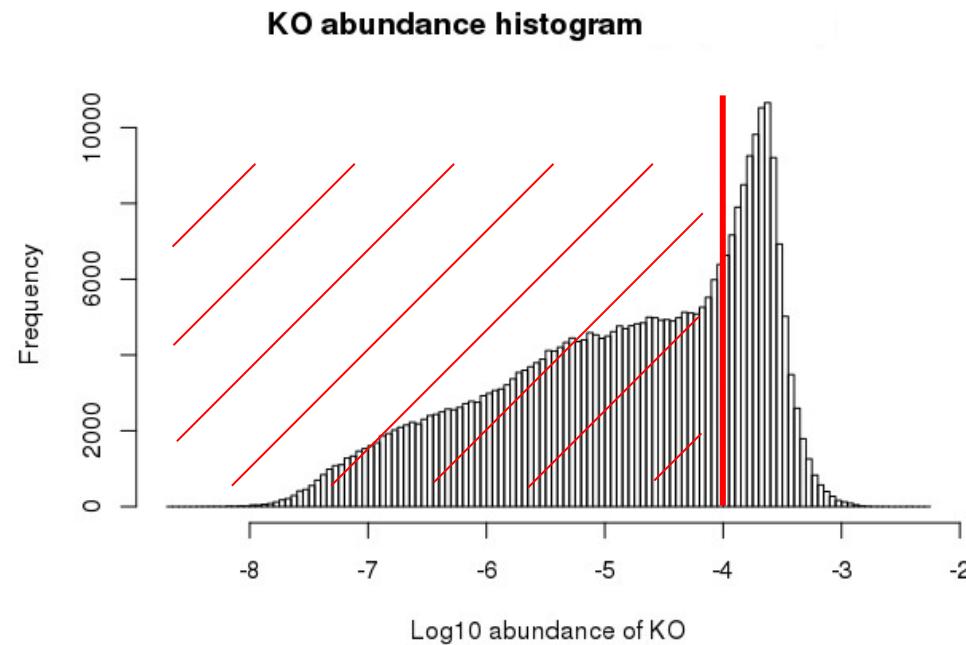
Threshold



Matrix of KO presence/absence in patients
(above threshold, below threshold)



The challenge of thresholds (1)

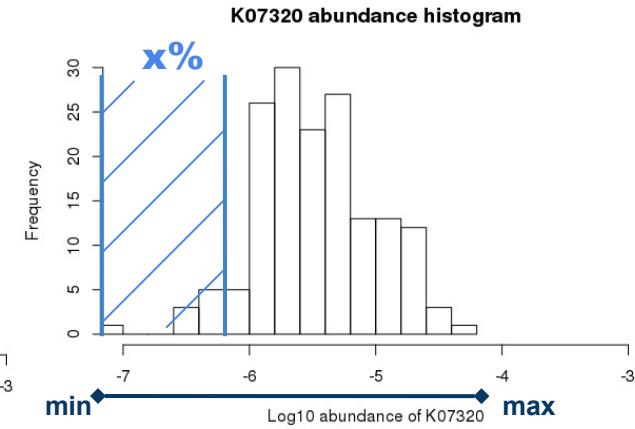
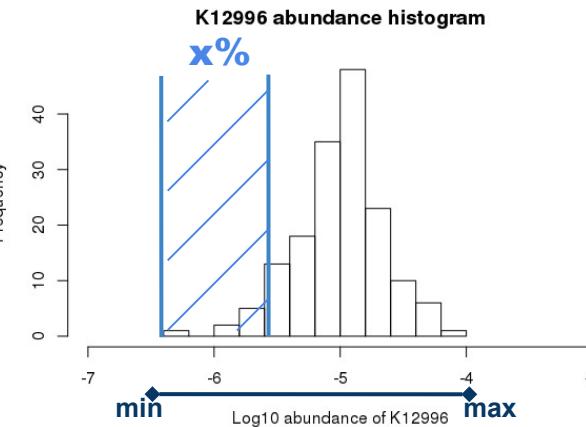
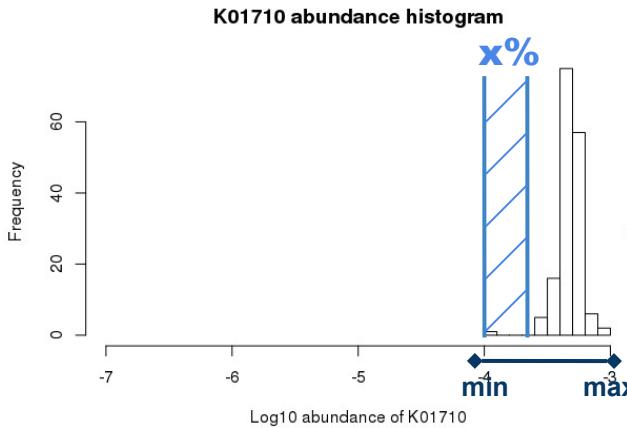


Absolute threshold :

All KOs included

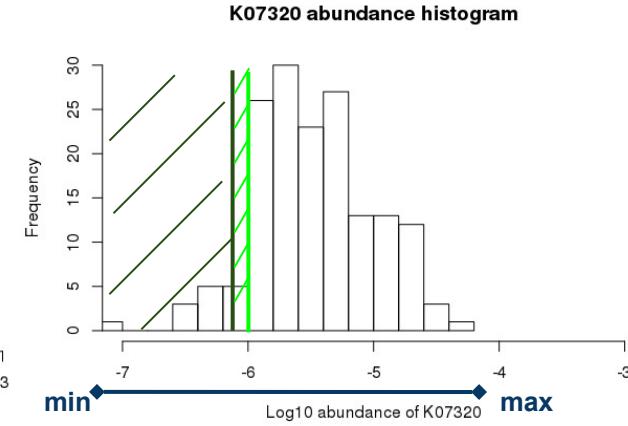
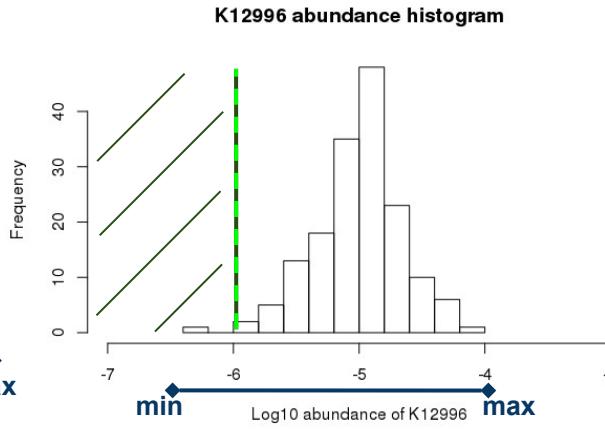
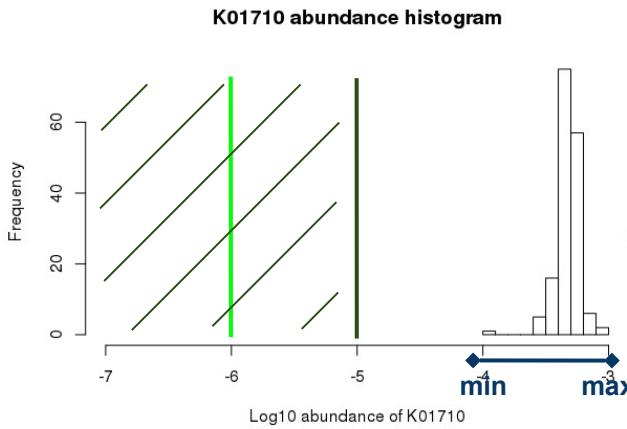
Remove KOs under threshold

The challenge of thresholds (2)



Specific threshold :
remove x last percent of span (ex: 30%)

The challenge of thresholds (3)



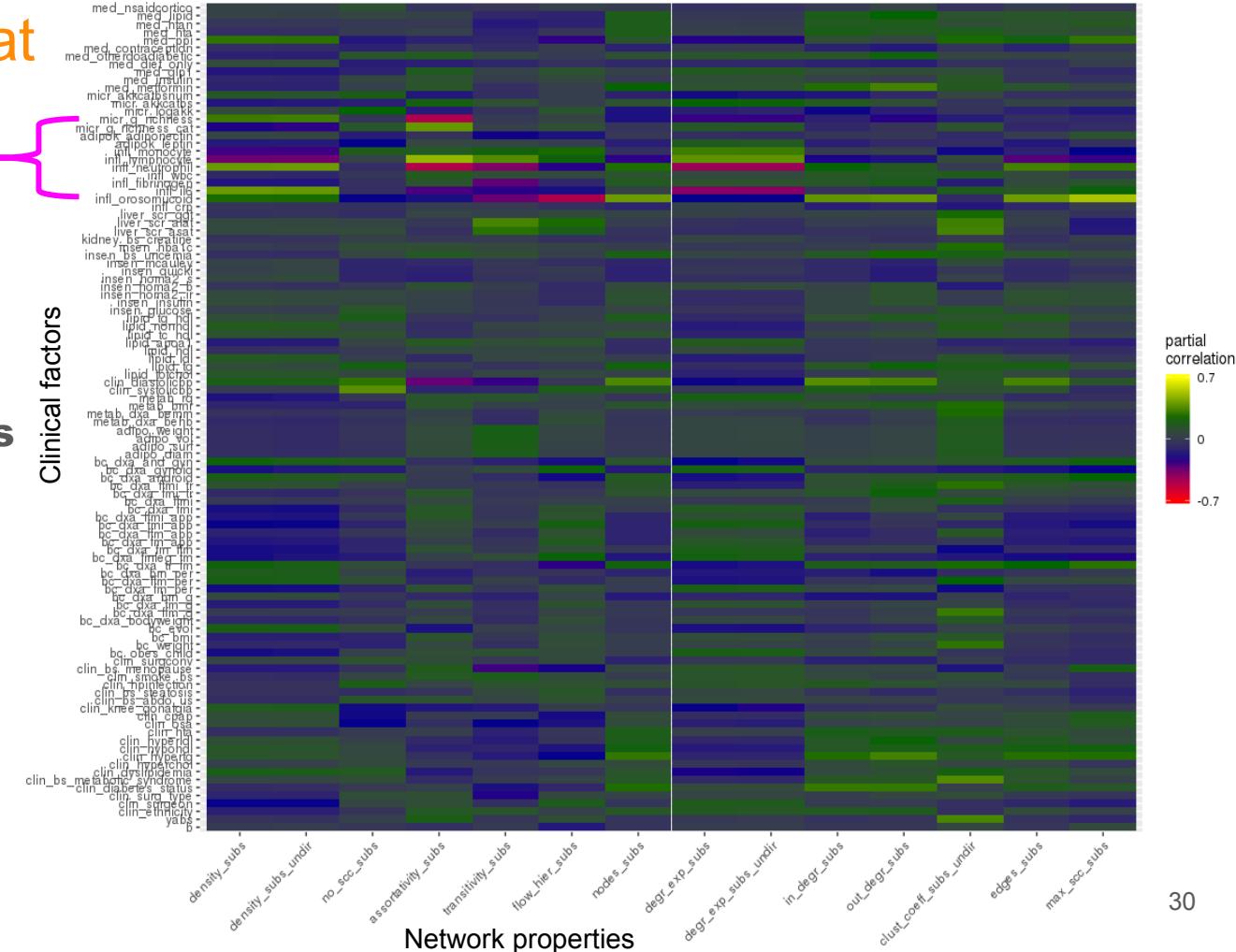
Threshold on span :

Remove :

- Absolute threshold : **KO < -6**
- KO < **max - value** (ex: **max - 2**)

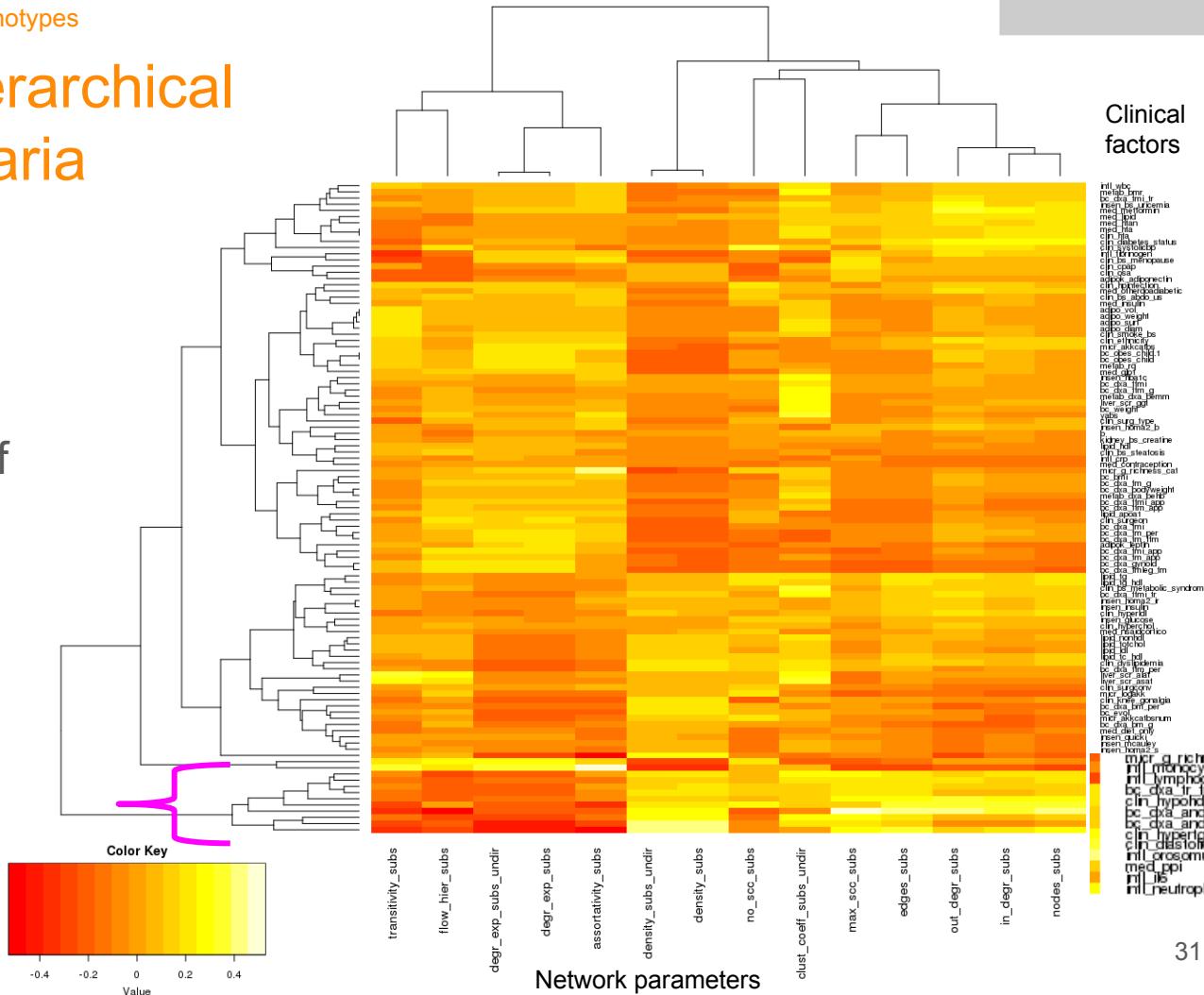
Partial correlation at
T0 :
Microbaria

- Patients **before** bariatric surgery (T0)
 - **Partial correlations**: clinical factors VS network properties
 - Adjusted for **age**
 - Specific threshold of **40%**



Correlation and hierarchical clustering : Microbaria

- **Hierarchical clustering** of network parameters and clinical factors of **partial correlation matrix (at T0)**



Including time : Linear mixed-effect model

- Allow adjusting for **time**, **bmi** and **age**, per **patient**
- Between **size of largest strongly connected component** against **all clinical factors**
- 57 patients, 155 observations
- Wald Chi-squared test to evaluate significance

Predictor variables	Estimate	Std. Error	t value
(Intercept)	86.2220663	20.4909329	4.207816
clin_hypertg	12.2886393	5.1925692	2.366582
month	-1.0299962	0.5574967	-1.847538
bc_bmi	-0.4485863	0.3829464	-1.171408
clin_age	-0.3433104	0.2422156	-1.417375

Table 1: **Linear mixed-effects model fixed effects.** clin_hypertg - hypertriglyceridemia, month - month of data set, bc_bmi - body mass index, clin_age - patient age.

Predictor variables	Chisq	Df	p-value
clin_hypertg	5.6007	1	0.01795
month	3.4134	1	0.06467
bc_bmi	1.3722	1	0.24144
clin_age	2.0090	1	0.15637

Table 2: **Wald test results.** Chisq - Wald χ^2 value ; Df - degrees of freedom ; p-value

Reconstruction de réseaux par espèce

b0001
b0002
...



Liste des noms de gènes
(dans fichier fasta d'ADNc de GenBank ou Ensembl)

KEGG

Escherichia coli K-12 MG1655; b0001

Entry b0002 CDS T00007

Gene name thrA

Definition (RefSeq) Bifunctional aspartokinase/homoserine dehydrogenase 1

KO K12524 bifunctional aspartokinase / homoserine dehydrogenase 1 [EC:2.7.2.4 1.1.1.3]

Organism eco Escherichia coli K-12 MG1655

Pathway eco00260 Glycine, serine and threonine metabolism
eco00261 Monobactam biosynthesis

Enzyme Codes (EC)

* Kyoto Encyclopedia of Genes and Genomes

Entrées de ces gènes sur la base de données KEGG*

KEGG ENZYME: 1.1.1.3

ENZYME: 2.7.2.4

Entry EC 2.7.2.4 Enzyme

Name aspartate kinase;
aspartokinase;
AK;
beta-aspartokinase;
aspartic kinase

Class Transferases;
Transferring phosphorus-containing groups;
Phosphotransferases with a carboxy group as acceptor
[BRITIE hierarchy](#)

Sysname ATP:L-aspartate 4-phosphotransferase

Reaction(IUBMB) ATP + L-aspartate = ADP + 4-phospho-L-aspartate [RN:R00480]

Reaction(KEGG) R00480

Substrate ATP [CPD:C00002];
L-aspartate [CPD:C00049]

Product ADP [CPD:C00008];
4-phospho-L-aspartate [CPD:C03082]

Substrats **Produits**



Construction des arêtes pour chaque réaction

Entrées des enzymes associées aux EC sur la base de données KEGG*

Reconstruction de réseaux par microbiote



ORTHOLOGY: K00161
ORTHOLOGY: K12524

Entry	K12524	K0
Name	thrA	
Definition	bifunctional aspartokinase / homoserine dehydrogenase 1 [EC:2.7.2.4 1.1.1.3]	Enzyme Codes (EC)
Pathway	ko00260 Glycine, serine and threonine metabolism ko00261 Monobactam biosynthesis	

Entrées de ces KO sur la base de données KEGG*

* Kyoto Encyclopedia of Genes and Genomes

ENZYME: 1.1.1.3
ENZYME: 2.7.2.4

Entry	EC 2.7.2.4	Enzyme
Name	aspartate kinase; aspartokinase; AK; beta-aspartokinase; aspartic kinase	
Class	Transferases; Transferring phosphorus-containing groups; Phosphotransferases with a carboxy group as acceptor BRITÉ hierarchy	
Sysname	ATP:L-aspartate 4-phosphotransferase	
Reaction(IUBMB)	ATP + L-aspartate = ADP + 4-phospho-L-aspartate [RN:R00480]	
Reaction(KEGG)	R00480 Reaction	
Substrate	ATP [CPD:C00002]; L-aspartate [CPD:C00049]	Substrats
Product	ADP [CPD:C00008]; 4-phospho-L-aspartate [CPD:C03082]	Produits



Construction des arêtes pour chaque réaction

Entrées des enzymes associées aux EC sur la base de données KEGG*