



Comment démêlé les Réseaux de Co-occurrence Discordant (RCD/DCN)

Présenté par Guillaume Gauthier

Terminologie de base

- Prenons le temps de définir la terminologie de base :
- Liens **concordants** :
 - Liens entre deux nœuds montrant une relation n'ayant pas changé entre les situations 1 et 2.
- Liens **discordants** :
 - Liens entre deux nœuds montrant une relation ayant changé entre les situations 1 et 2.

Plan de la présentation



Pourquoi utiliser les DCN?



Concept de base



Interprétation de résultat DNC



Exercices



Pop Quiz

Pourquoi utiliser les DCN?

3 grandes questions:

1. Qui interagit avec qui?
2. Est-ce qu'un traitement (ou toute autre perturbation du milieu) affecte la population?
3. Comment est-ce que les interactions sont affectées?

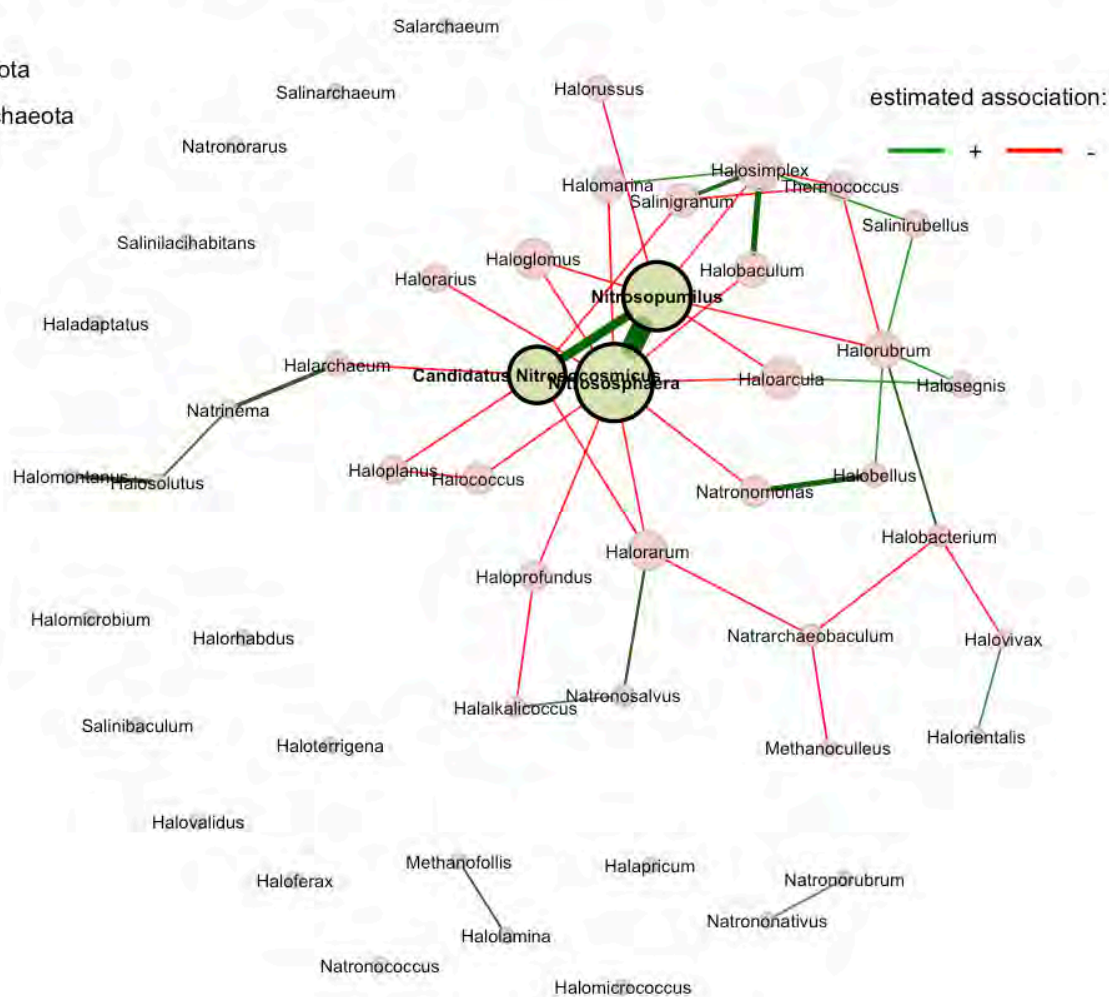


The background features a series of concentric circles in light gray, some solid and some dashed, creating a ripple effect. A large, solid red speech bubble is centered on the page, pointing downwards. The text "Concept de base" is written in white inside the bubble.

Concept de base

Network on genus level with SPRING associations des 50 genus les plus abondants

Qu'est-ce qu'un CN?



C'est une analyse qui permet de visualiser les interactions entre différents composants d'un milieu:

- Les microorganismes composant un microbiome ou microbiote;
- Les étudiants d'une classe;
- Des cultures bactériennes dans un plat de Petri;
- ...

Composé de 2 entités:

1. Les nœuds : représentation des entités (otu, genre, individus, ...).
2. Les liens: lignes connectant les nœuds qui apparaissent souvent ensemble ou dont l'abondance est corrélée.

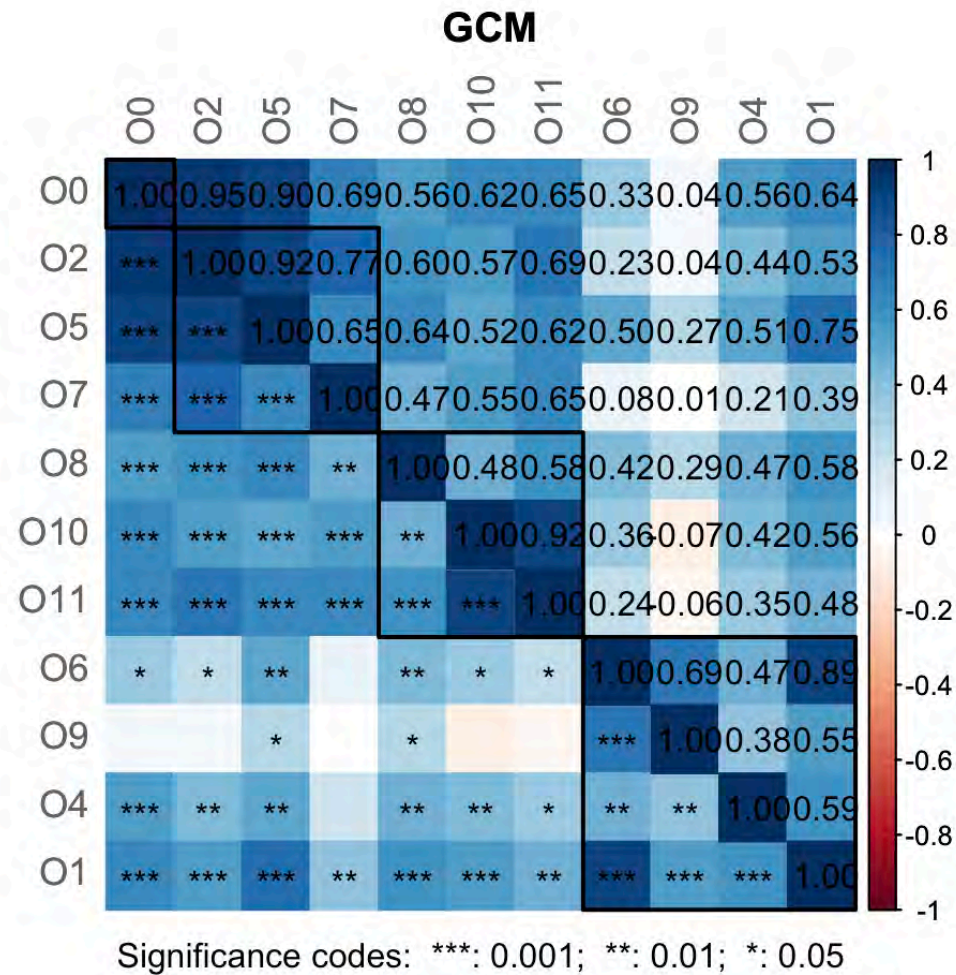
Concept purement mathématique. N'intègre pas de concepts biologiques (métabolisme, ...) ni génétiques/phylogénétiques.



Différence entre DCN et CN?

- On peut faire une analogie entre les métriques de diversité (alpha/bêta) et les CN.
- Les CN n'utilisent ou n'analysent qu'une seule variable, indépendamment des autres. On observe les impacts intratraitement:
 - $\text{Div alpha} \sim \text{CN}$
- Les DCN analysent les variations entre 2 conditions/jeux de données ou 2 variables (*limite de 2^*). Ils permettent donc d'observer des variations intertraitement:
 - $\text{Div bêta} \sim \text{DCN}$

Théoriquement, il est possible de calculer toutes les permutations de paires d'un jeu de données, mais c'est très demandant en termes de ressources



Comment est-ce qu'un CN est construit?

Création d'une matrice d'association ou adjacency;

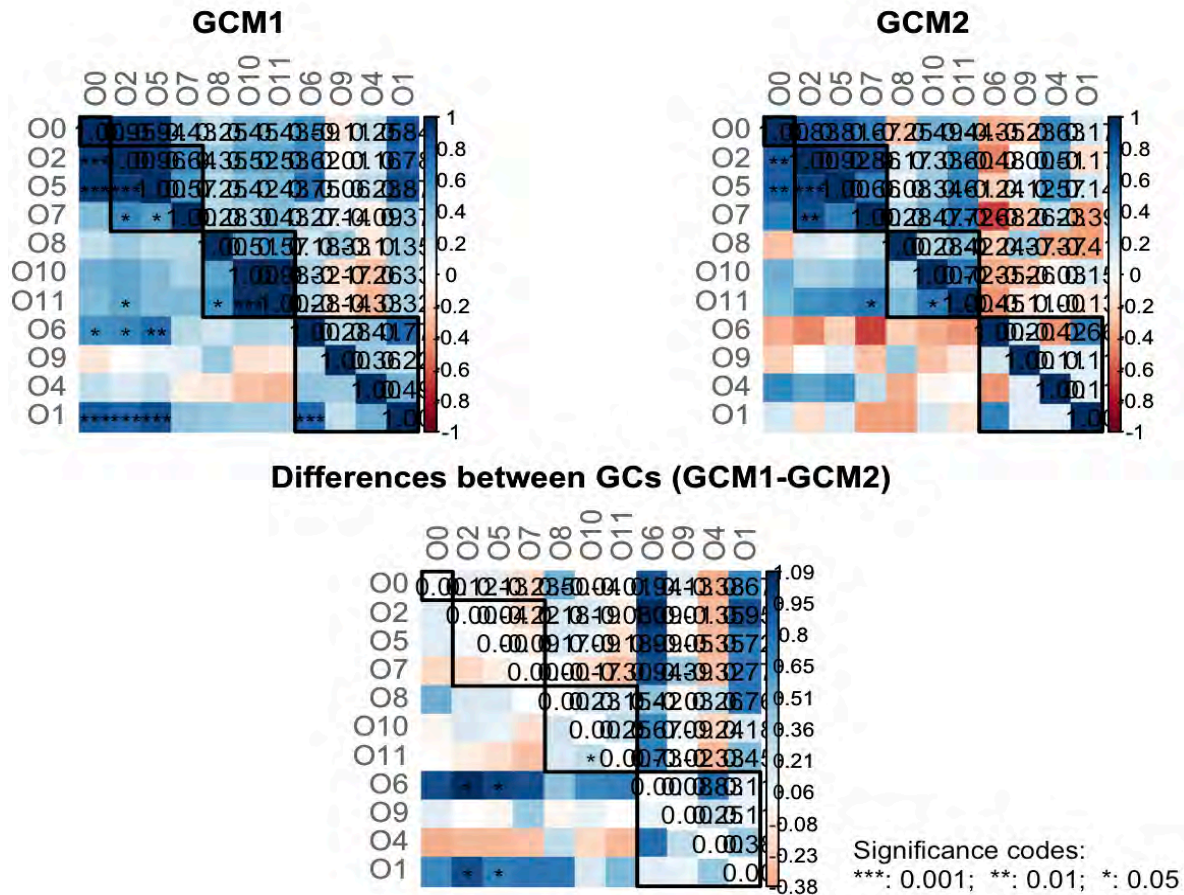
Plusieurs formules et théories existent. Voici 3 qui sont recommandées en microbiologie:

- SparCC (corrélacion)
- CClasso (corrélacion)
- SPRING (dépendance conditionnelle)*

Cela va créer une matrice d'association (Group Connection matrice/GCM) :

- Valeurs avant sparsification.

non compatible avec la méthode discordante



Comment est-ce que les CN sont comparés et comment les DCN sont construits ?

Avec Netcomi 2 étapes:

- Construction des deux matrices d'association.
- C'est à partir de cette matrice que la matrice de comparaison de CN sera construite.

Comparaison des Network:

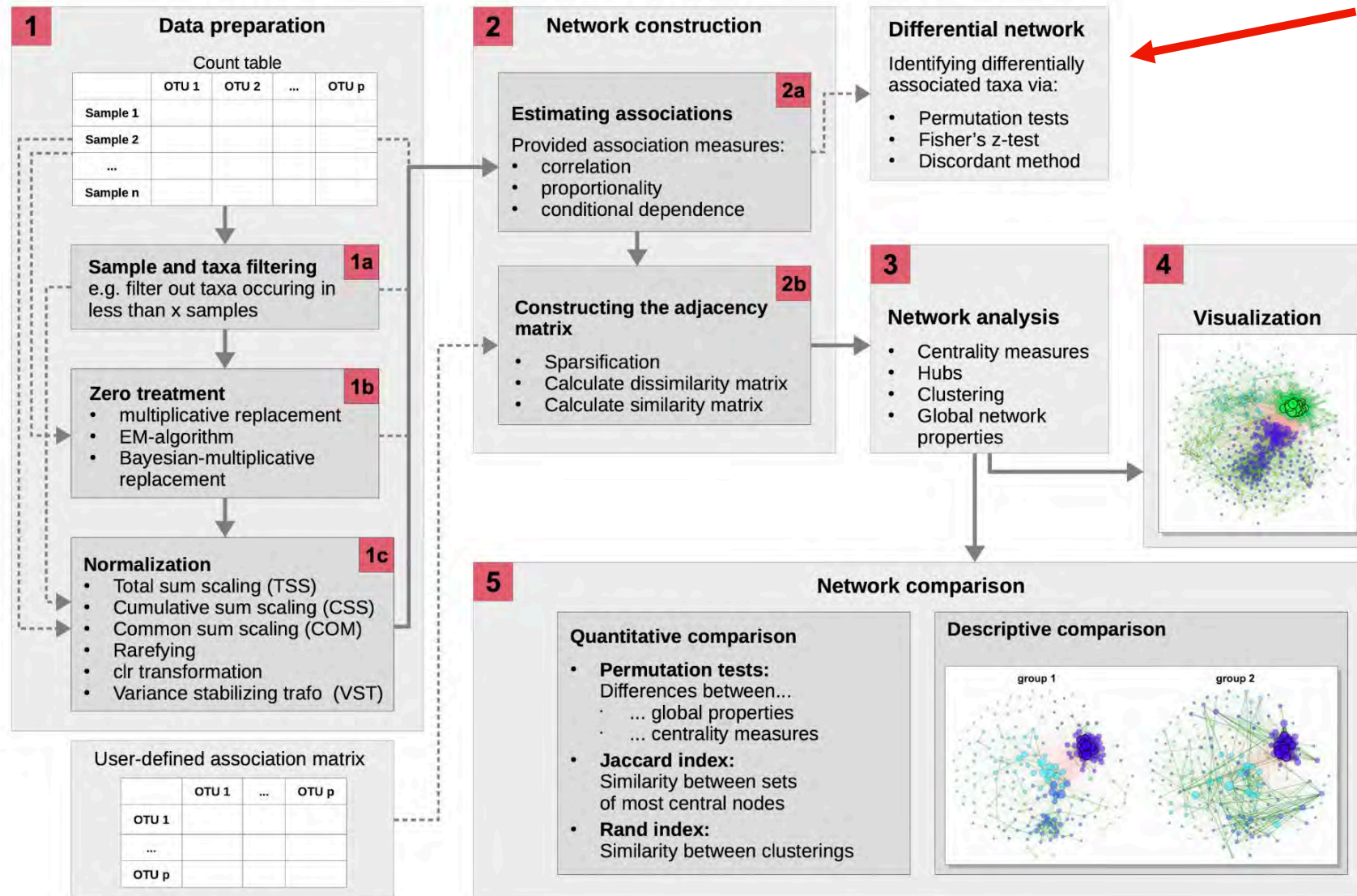
- Discordance et permutation

assoMat1

	Caenorhabditis	Necator	Strongyloides	Loa	Brugia	Trichinella
Caenorhabditis	1.000000e+00	3.512127e-02	-1.344616e-02	6.895716e-03	-1.669490e-02	5.706044e-03
Necator	3.512127e-02	1.000000e+00	2.885280e-03	-1.472533e-04	-2.511525e-03	-7.001424e-04
Strongyloides	-1.344616e-02	2.885280e-03	1.000000e+00	4.034965e-03	-3.591251e-03	1.525047e-02
Loa	6.895716e-03	-1.472533e-04	4.034965e-03	1.000000e+00	-9.920698e-04	5.785333e-03
Brugia	-1.669490e-02	-2.511525e-03	-3.591251e-03	-9.920698e-04	1.000000e+00	-3.437758e-03
Trichinella	5.706044e-03	-7.001424e-04	1.525047e-02	5.785333e-03	-3.437758e-03	1.000000e+00
Fusarium	5.860748e-04	0.000000e+00	-7.276149e-06	-7.603010e-06	8.473536e-06	1.125543e-05
Ilyonectria	-1.243824e-02	-2.486311e-03	-1.330723e-03	6.278358e-06	9.456131e-04	1.710432e-04
Trichoderma	-9.170509e-02	-2.080824e-03	-2.575547e-02	-1.189153e-02	8.875619e-03	-4.098462e-02
Purpureocillium	2.694225e-02	8.042070e-04	1.141627e-03	1.258746e-03	-1.277810e-03	3.750395e-03
Hirsutella	5.502321e-02	1.306966e-02	1.334088e-02	2.214616e-03	-6.536148e-03	7.929011e-03
Drechmeria	9.664770e-02	2.576476e-02	1.712539e-02	-1.610598e-04	-9.829106e-03	-6.859467e-04
Metarhizium	2.337416e-01	-6.116668e-04	-1.173428e-04	-4.180113e-06	2.458862e-04	3.383957e-05
Pochonia	2.427181e-03	1.439468e-04	-1.605039e-05	-1.543627e-04	3.242621e-05	-3.714743e-04
Ustilaginoidea	7.308913e-02	8.439212e-03	7.686199e-03	2.322680e-03	-4.942244e-03	8.055579e-03
Hypocrella	-7.464836e-04	-6.911119e-05	4.562900e-04	1.403088e-04	-5.338741e-05	5.364304e-04
Epichloe	3.506377e-04	1.233134e-06	6.297947e-06	-5.232885e-06	5.040542e-06	7.246295e-06

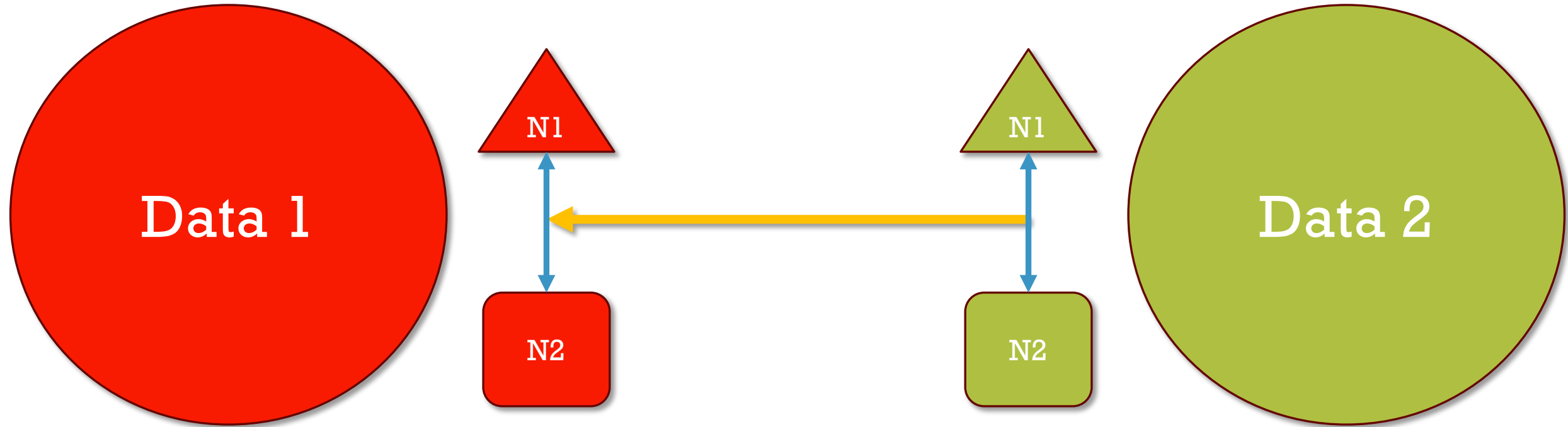
assoMat2

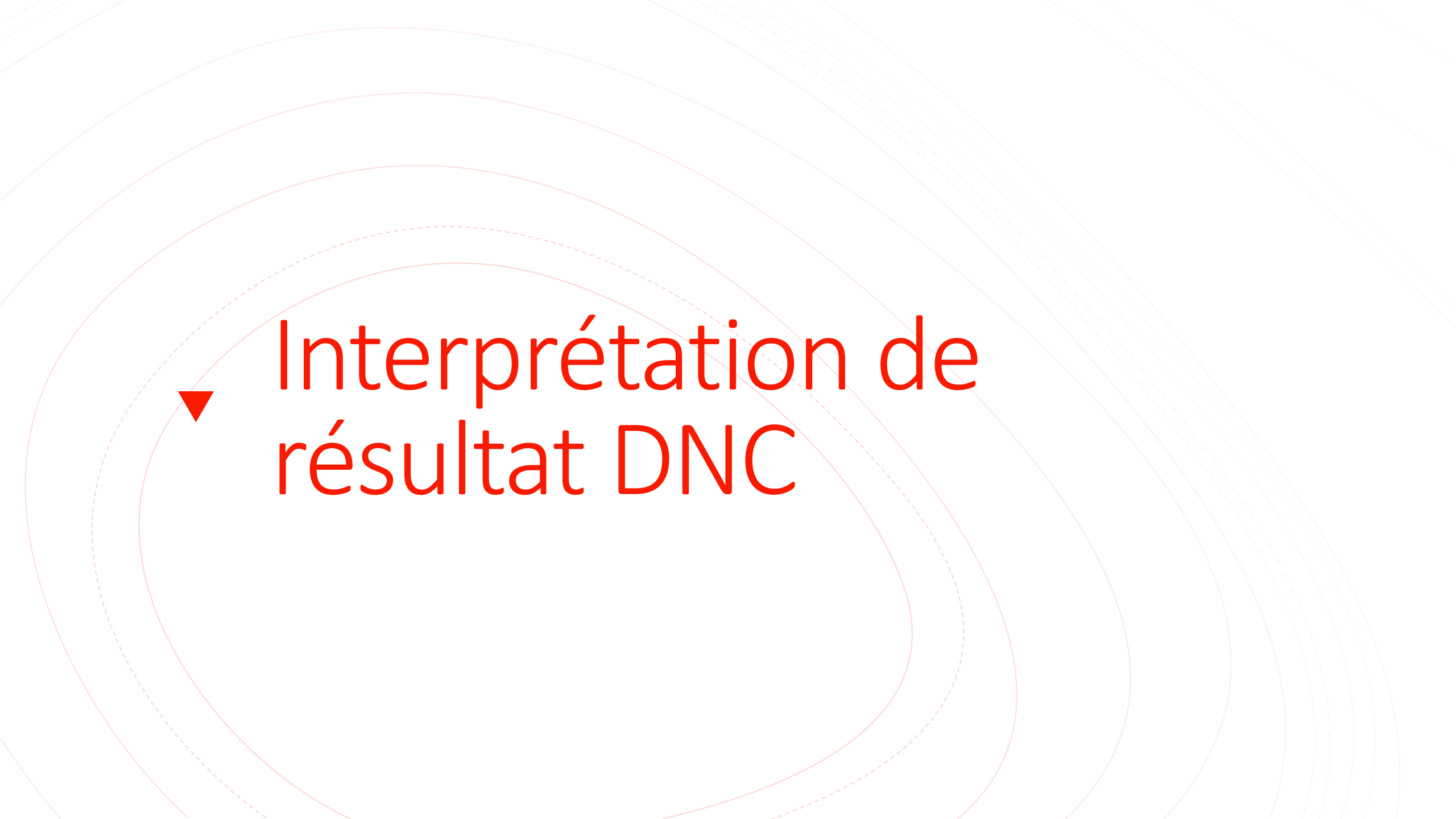
	Caenorhabditis	Necator	Strongyloides	Loa	Brugia	Trichinella
"Necator" abditis	1.0000000000	3.484035e-03	1.792419e-02	2.294418e-02	1.456247e-02	-2.055047e-01
Necator	0.0034840349	1.000000e+00	-2.661144e-03	4.567786e-05	-6.137169e-04	-1.282739e-03
Strongyloides	0.0179241924	-2.661144e-03	1.000000e+00	1.026084e-04	4.339243e-04	6.503436e-03
Loa	0.0229441770	4.567786e-05	1.026084e-04	1.000000e+00	4.421045e-05	3.317499e-05
Brugia	0.0145624708	-6.137169e-04	4.339243e-04	4.421045e-05	1.000000e+00	3.197375e-04
Trichinella	-0.2055046781	-1.282739e-03	6.503436e-03	3.317499e-05	3.197375e-04	1.000000e+00
Fusarium	-0.1764322884	3.961453e-04	2.009323e-04	2.854509e-05	-6.952671e-04	1.879096e-05
Ilyonectria	-0.1176552839	6.809763e-03	-4.033701e-04	1.288039e-04	-1.409280e-02	-1.910115e-03
Trichoderma	-0.0443026320	1.903519e-03	-8.567604e-03	-9.697972e-04	-1.069030e-03	-3.527770e-03
Purpureocillium	-0.2501921492	6.532992e-03	-2.651884e-02	6.882193e-05	-3.309604e-03	-1.436864e-02
Hirsutella	-0.0521203750	1.506699e-03	2.566564e-03	4.994919e-05	-4.797918e-03	9.212428e-04
Drechmeria	-0.0232908470	7.343702e-04	-8.008616e-04	5.315501e-05	-6.066903e-04	-8.168407e-04
Metarhizium	0.0012897831	1.753873e-04	5.614962e-05	2.644924e-05	2.382497e-04	-2.437569e-04
Pochonia	0.0180568269	-1.682642e-04	1.190307e-03	6.186048e-05	1.069617e-04	5.433136e-04
Ustilaginoidea	0.0104758531	1.808523e-04	-3.902497e-04	5.254456e-05	-7.050631e-06	-1.634893e-04
Hypocrella	0.0143386172	-2.418858e-04	1.138556e-03	2.176233e-05	1.254301e-04	5.480606e-04
Epichloe	-0.0003169512	4.960999e-05	3.429104e-05	4.298158e-05	2.612861e-05	5.158402e-06



(Peschel et al., 2021)

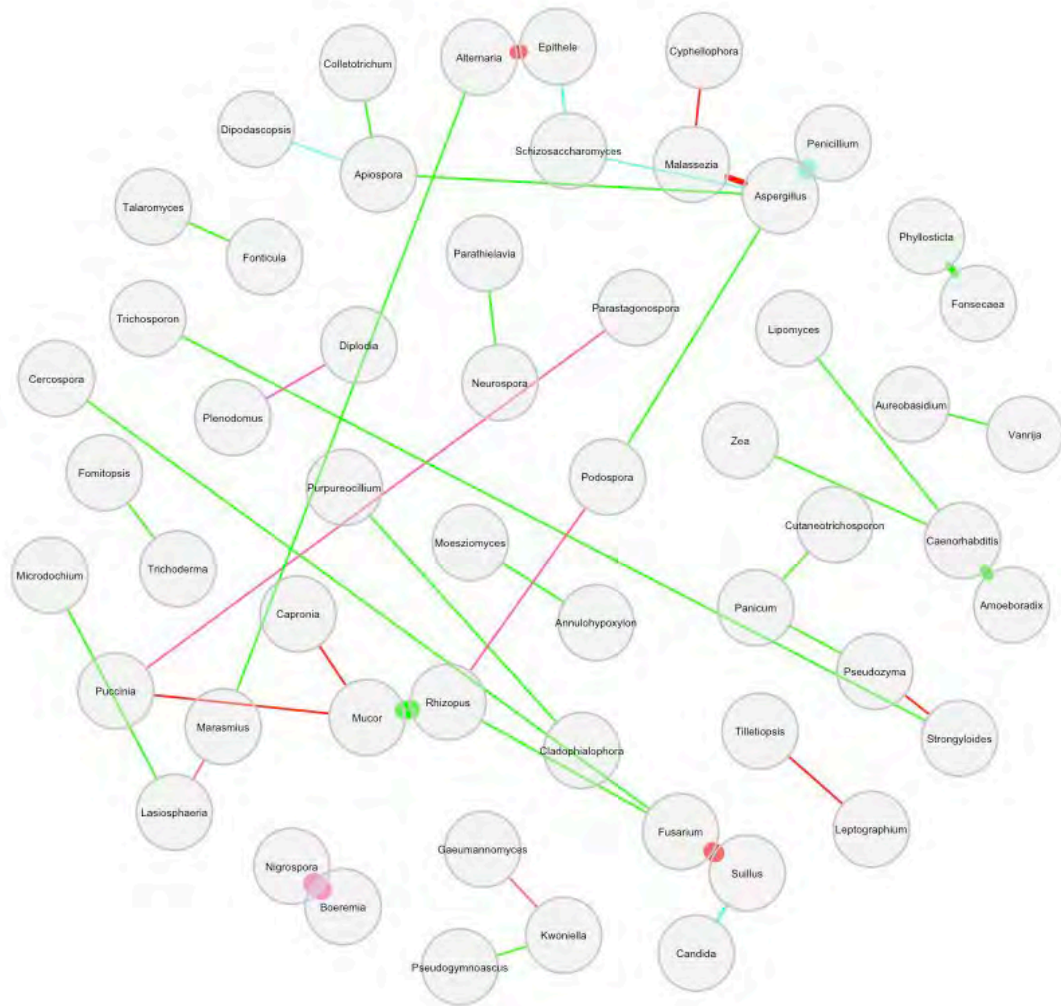
Direction des liens





▼ Interprétation de résultat DNC

Differential network



Associations		
	PPE	SN
—	0	-
—	0	+
—	-	0
—	-	+
—	+	0
—	+	-

Points importants

Discordance vs concordance?

Les 3 liens concordants:

■ ++, -- & 00

Les 6 liens discordants:

■ +-, +0, -+, -0, 0- & 0+

Que comparons-nous?

Est-ce qu'il s'agit d'une analyse directionnelle?

Mise en situation

- Si nous avons un DCN (fertilisé/non fertilisé) et les noeuds Oïdium (n1) et Penicillium (n2) reliés par un lien +-. Que pouvons-nous en tirer?
- Décomposons la situation:
 1. Fertiliser (+): l'abondance d'Oïdium et de Penicillium est plus élevée dans le groupe fertilisé.
 2. Non fertilisé (-): l'abondance d'Oïdium et/ou de Penicillium est plus faible dans le groupe non fertilisé.
 3. Le lien +-: La relation entre les deux a drastiquement changé. Dans le groupe fertilisé, quand Oïdium ou Penicillium est fortement présent, l'autre a tendance à l'être aussi. Et **cette association positive est significativement plus forte, ou de direction opposée, que celle du contrôle.**
 4. Donc leur relation positive est due au traitement.

Situation	Nb. Otu/ traitement	Fertilisé	Contrôle	Différence	Discordance
1	Oïdium	140	100	+40	+-
	Pennicillum	70	30	+40	
2	Oïdium	60	25	+35	+-
	Pennicillum	65	50	+15	
3	Oïdium	80	60	+20	+-
	Pennicillum	80	40	+40	

N'oublié pas!!!

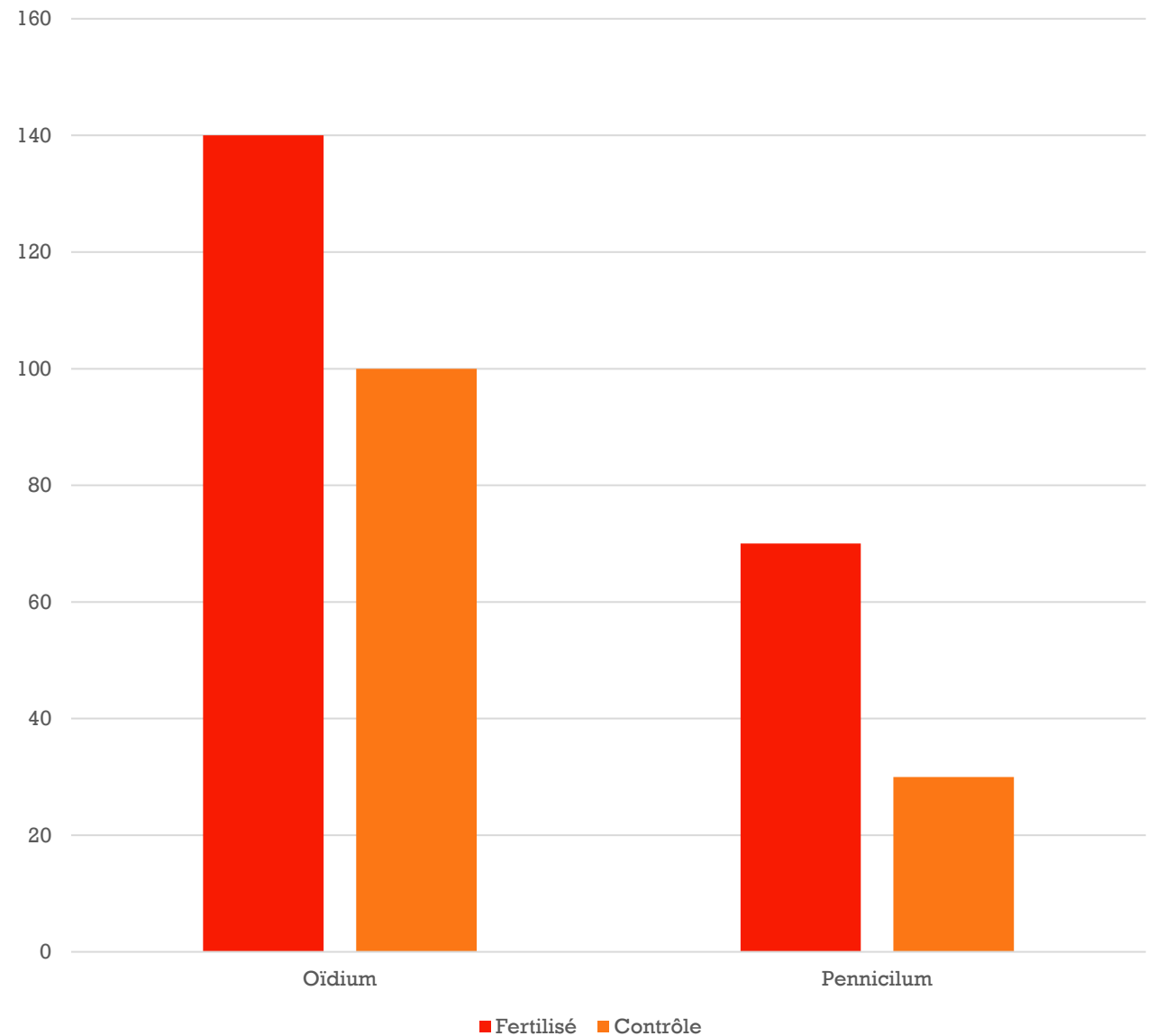
- Les analyses de discordances observent les changements d'une situation « **A** » à une situation « **B** »
- Le poids des liens n'est pas considéré, seulement la variation entre les 2 situations.

Comment vérifier?

La méthode la plus simple =
graphique abondance

Mais ce n'est pas parfait

Variation de l'abondance entre le traitement et le témoin 16



Que pourrait être
les causes de ce
changement?

**Condition du milieu devenue
plus favorable/défavorable?**

**Compétition pour les
ressources?**

**Production d'organite
antifongique?**

...

Faire
attention!!!

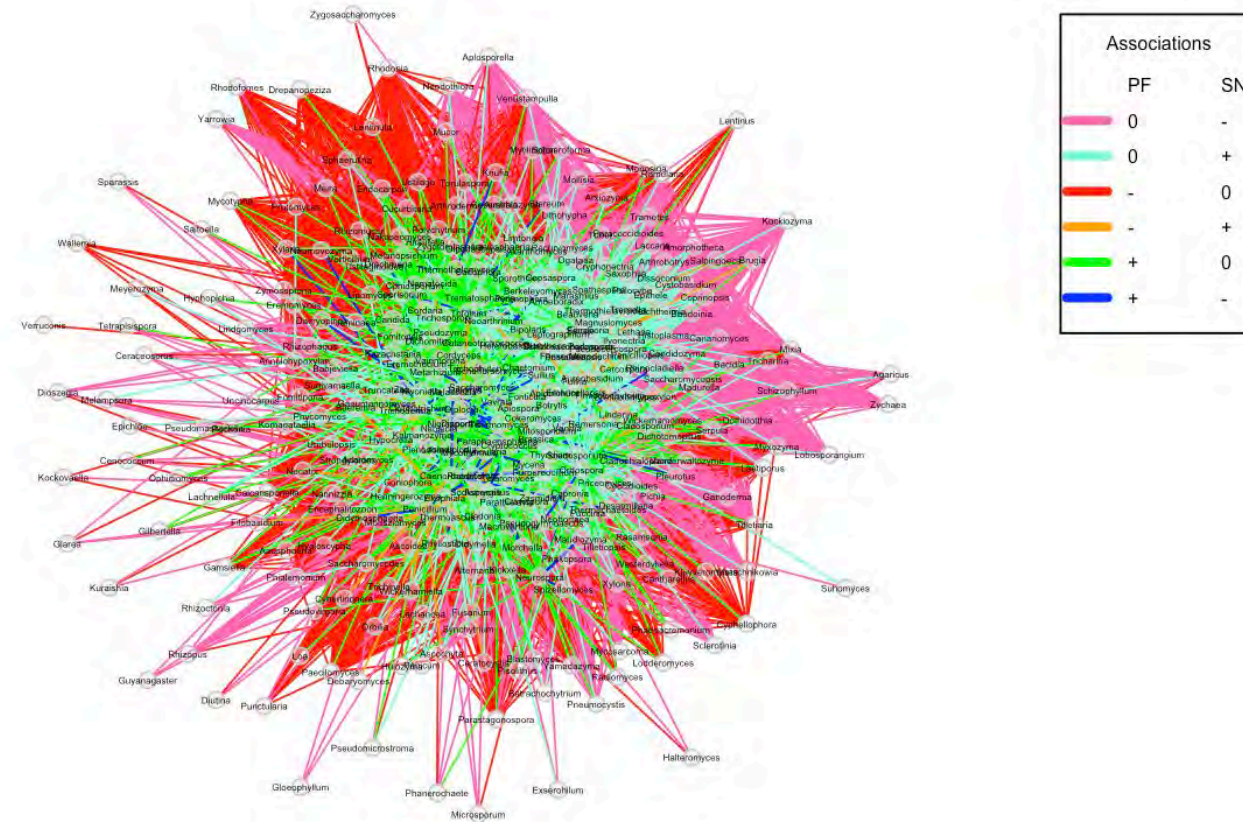
Qu'est-ce que le 0 dans +0 implique?

- 0 ne veut pas dire aucune interaction ni impact! Mais pas d'interaction ni de changement **significatif**.
- Comme pour la capacité tampon du sol, l'effet peut être atténué ou masqué, ou il peut n'y avoir aucun effet.



Exercice

Prenons de mes données en exemple



DCN

Que faire avec ça?

On ne peut rien
faire avec ça
dans sa forme
actuelle

Option?

1. Travailler le réseau avec Gelphi;
2. Sortir les données sous forme de table

Discordance		FM - NS		RM - NS		SM - NS		LM - NS	
		n. edges	%	n. edges	%	n. edges	%	n. edges	%
Fungi	+−	469	3	2718	15	412	3	1277	7
	+0	775	5	5026	27	793	5	4126	23
	−+	523	4	402	2	594	4	416	2
	−0	5247	35	5027	27	6058	39	5255	29
	0+	1153	8	799	4	1006	6	965	5
	0−	6674	45	4689	25	6679	43	6042	33
Total		14841		18661		15542		18081	

Résultat de discordance

Que faire ici?

Choix

- Si on choisi de se concentrer sur le FM-NS et ses liens +-. Quelle serait la prochaine étape selon vous?
- Explorer quelles paires composent les 469 liens?

Edge Type: +-

A tibble: 10 × 6

	edge_type	genus	domain	total_edges	as_node1	as_node2
	<chr>	<chr>	<chr>	<int>	<int>	<int>
1	+-	Caenorhabditis	Nematoda	49	49	0
2	+-	Colletotrichum	Fungi	47	43	4
3	+-	Penicillium	Fungi	33	29	4
4	+-	Pyricularia	Fungi	29	24	5
5	+-	Aspergillus	Fungi	28	18	10
6	+-	Diplodia	Fungi	28	10	18
7	+-	Cryptococcus	Fungi	25	6	19
8	+-	Pseudozyma	Fungi	23	0	23
9	+-	Mycena	Fungi	22	5	17
10	+-	Rhodotorula	Fungi	21	2	19

Voici le top 10 noeuds ayant ce type de lien

Quelles informations peut-on en sortir? Qu'elle est l'autre partis des paires?

	node1	node2	class_code	domain1	domain2	edge_type
	All	All	All	All	All	All
317	Caenorhabditis	Necator		2 Nematoda	Nematoda	-0
633	Caenorhabditis	Strongyloides		5 Nematoda	Nematoda	--
634	Necator	Strongyloides		1 Nematoda	Nematoda	00
949	Caenorhabditis	Loa		5 Nematoda	Nematoda	--
950	Necator	Loa		1 Nematoda	Nematoda	00
951	Strongyloides	Loa		2 Nematoda	Nematoda	-0
1265	Caenorhabditis	Brugia		5 Nematoda	Nematoda	--
1266	Necator	Brugia		1 Nematoda	Nematoda	00
1267	Strongyloides	Brugia		2 Nematoda	Nematoda	-0
1268	Loa	Brugia		1 Nematoda	Nematoda	00

Comment explorer ses données?

Dans R, on peut facilement explorer avec l'option `filter`

	node1	node2	class_code	domain1	domain2	edge_type
	caenor	All	All	All	All	+ -
2529	Caenorhabditis	Trichoderma	6	Nematoda	Fungi	+ -
3161	Caenorhabditis	Hirsutella	6	Nematoda	Fungi	+ -
3477	Caenorhabditis	Drechmeria	6	Nematoda	Fungi	+ -
4425	Caenorhabditis	Ustilaginoidea	6	Nematoda	Fungi	+ -
6637	Caenorhabditis	Geosmithia	6	Nematoda	Fungi	+ -
7901	Caenorhabditis	Berkeleyomyces	6	Nematoda	Fungi	+ -
9165	Caenorhabditis	Thermothelomyces	6	Nematoda	Fungi	+ -
10113	Caenorhabditis	Mycothermus	6	Nematoda	Fungi	+ -
11061	Caenorhabditis	Thermochaetoides	6	Nematoda	Fungi	+ -
11693	Caenorhabditis	Sordaria	6	Nematoda	Fungi	+ -
13273	Caenorhabditis	Pyricularia	6	Nematoda	Fungi	+ -
14537	Caenorhabditis	Sporothrix	6	Nematoda	Fungi	+ -

Ou on peut faire un peu de code

27

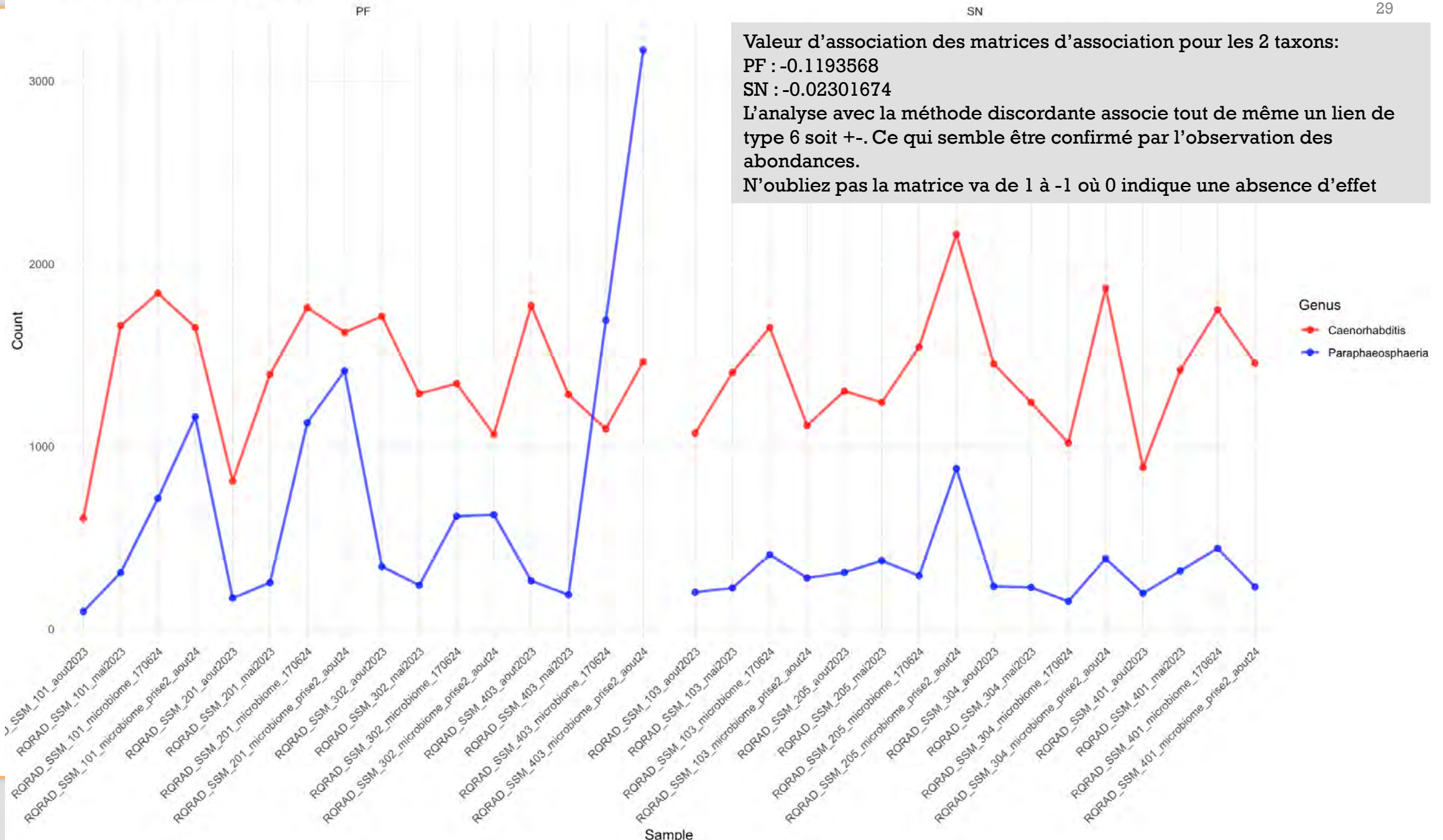
```
caenorhabditis_partners <- results_domain_summary_with_top_genera_PF_SN$edge_data %>%  
  filter(node1 == "Caenorhabditis" & edge_type == "+-") %>%  
  pull(node2) %>%  
  unique()  
# Convert from factor to character  
caenorhabditis_partners <- as.character(caenorhabditis_partners)  
correct_target_genera <- c("Caenorhabditis", caenorhabditis_partners)
```

```
> correct_target_genera  
[1] "Caenorhabditis"      "Trichoderma"      "Hirsutella"      "Drechmeria"      "Ustilaginoidea"  "Geosmithia"  
[7] "Berkeleyomyces"     "Thermothelomyces" "Mycothermus"     "Thermochaetoides" "Sordaria"        "Pyricularia"  
[13] "Sporothrix"         "Cadophora"        "Venustampulla"   "Pseudogymnoascus" "Thermomyces"     "Capronia"  
[19] "Endocarpon"         "Bipolaris"        "Pyrenophora"     "Cucurbitaria"    "Paraphaeosphaeria" "Diplodia"  
[25] "Coniosporium"       "Ogataea"          "Eremothecium"    "Lachancea"       "Nakaseomyces"     "Huiozyma"  
[31] "Ascoidea"           "Dipodascopsis"    "Limtongia"       "Dichomitrus"     "Fomitopsis"       "Heterobasidion"  
[37] "Cantharellus"       "Cryptococcus"     "Cutaneotrichosporon" "Kalmanozyma"     "Melanopsichium"   "Jaminaea"  
[43] "Malassezia"         "Cokeromyces"      "Rhizomucor"      "Polychytrium"    "Nematocida"       "Capsaspora"  
[49] "Sphaeroforma"      "Zea"
```



▼ Comment vérifier les relations?

Valeur d'association des matrices d'association pour les 2 taxons:
PF : -0.1193568
SN : -0.02301674
L'analyse avec la méthode discordante associe tout de même un lien de type 6 soit +-. Ce qui semble être confirmé par l'observation des abondances.
N'oubliez pas la matrice va de 1 à -1 où 0 indique une absence d'effet





Pop quiz

1. On retrouve toujours une paire +/-+, car les noeuds n1 et n2 sont intervertibles ?
 - Non, dans ce type d'analyse les liens ne sont pas dirigés. La position des noeuds n'a pas d'influence sur la direction du lien.
2. Qu'arriverait-il aux résultats si on insère le contrôle en data1 et le traitement en data2?
 - Les directions des liens seront inversées par rapport à l'exemple $n1/n2 = -(n2/n1)$ donc +/- deviendra -+.
3. Travaillons-nous avec des abondances relatives ou absolues ?
 - Relative! Il faudrait quantifier la présence de chaque individu par PCR.
4. Que pouvons-nous interpréter d'un lien négatif (-0) entre Oïdium (n1) et Doryphore adulte (n2) par rapport à un réseau DCN : utilisation d'un pesticide de synthèse (data1) et biocontrôle (data2)?
 - A. L'utilisation du pesticide entraînera une réduction de la population de l'un ou de l'autre des organismes.
 - B. L'effet du biocontrôle sur les populations de ravageurs est le même ou n'a aucun effet. Le biocontrôle attaque peut-être uniquement le stade larvaire du doryphore.
 - C. La réduction de la population d'un des ravageurs entraîne une augmentation de la population de l'autre.
 - D. Il faudrait consulter les tables d'abondance afin de déterminer quelles populations sont affectées et comment.
 - E. ...



Merci de votre écoute

Avez-vous des questions?

Citation

- Peschel, S., Müller, C. L., von Mutius, E., Boulesteix, A.-L., & Depner, M. (2021). NetCoMi: Network construction and comparison for microbiome data in R. *Briefings in Bioinformatics*, 22(4), bbaa290.
<https://doi.org/10.1093/bib/bbaa290>