User Manual of integrated Network Analysis Pipeline (iNAP) in Denglab

http://mem.rcees.ac.cn:8081

Updated

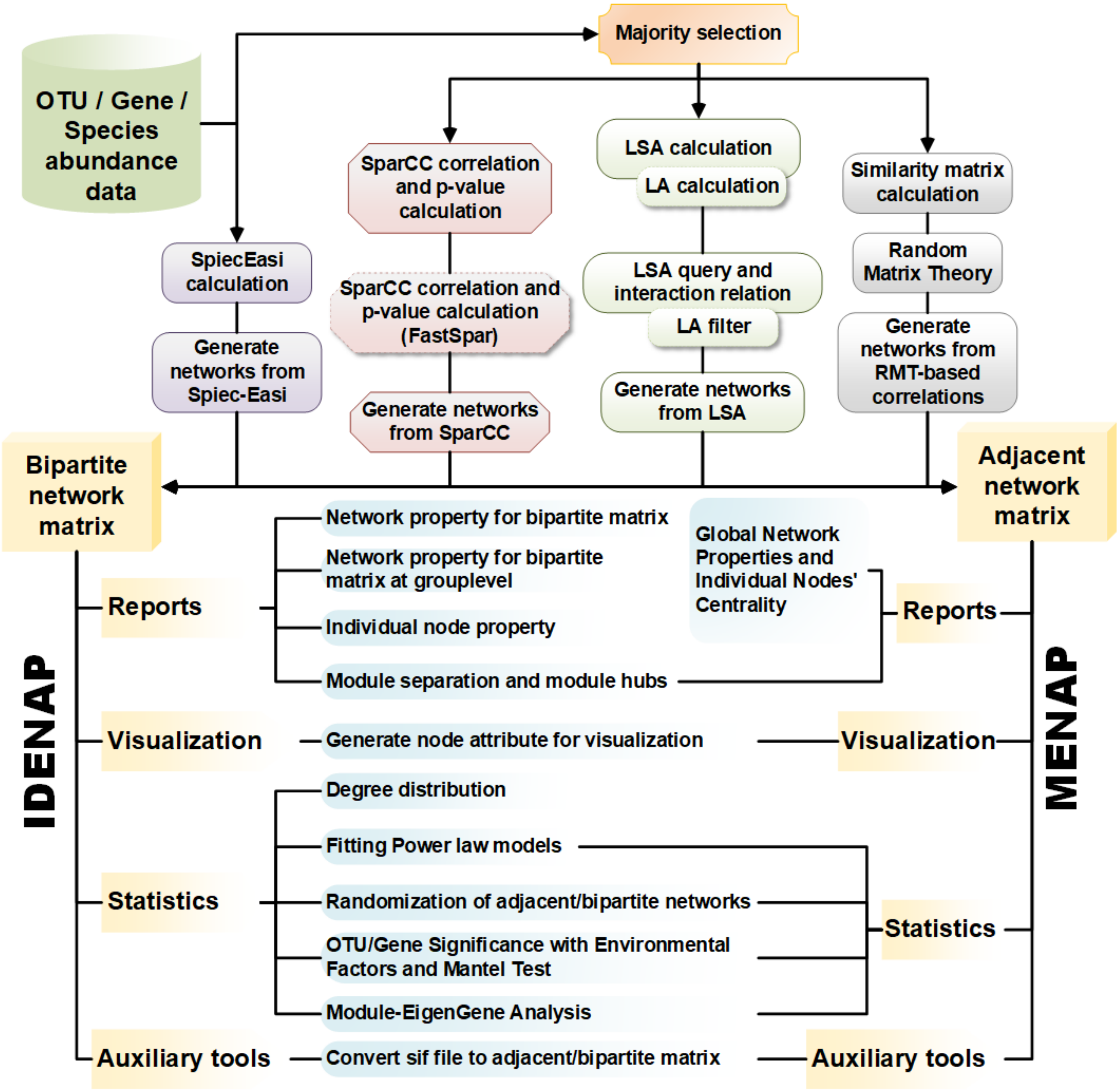
Jan. 2022

@ Denglab

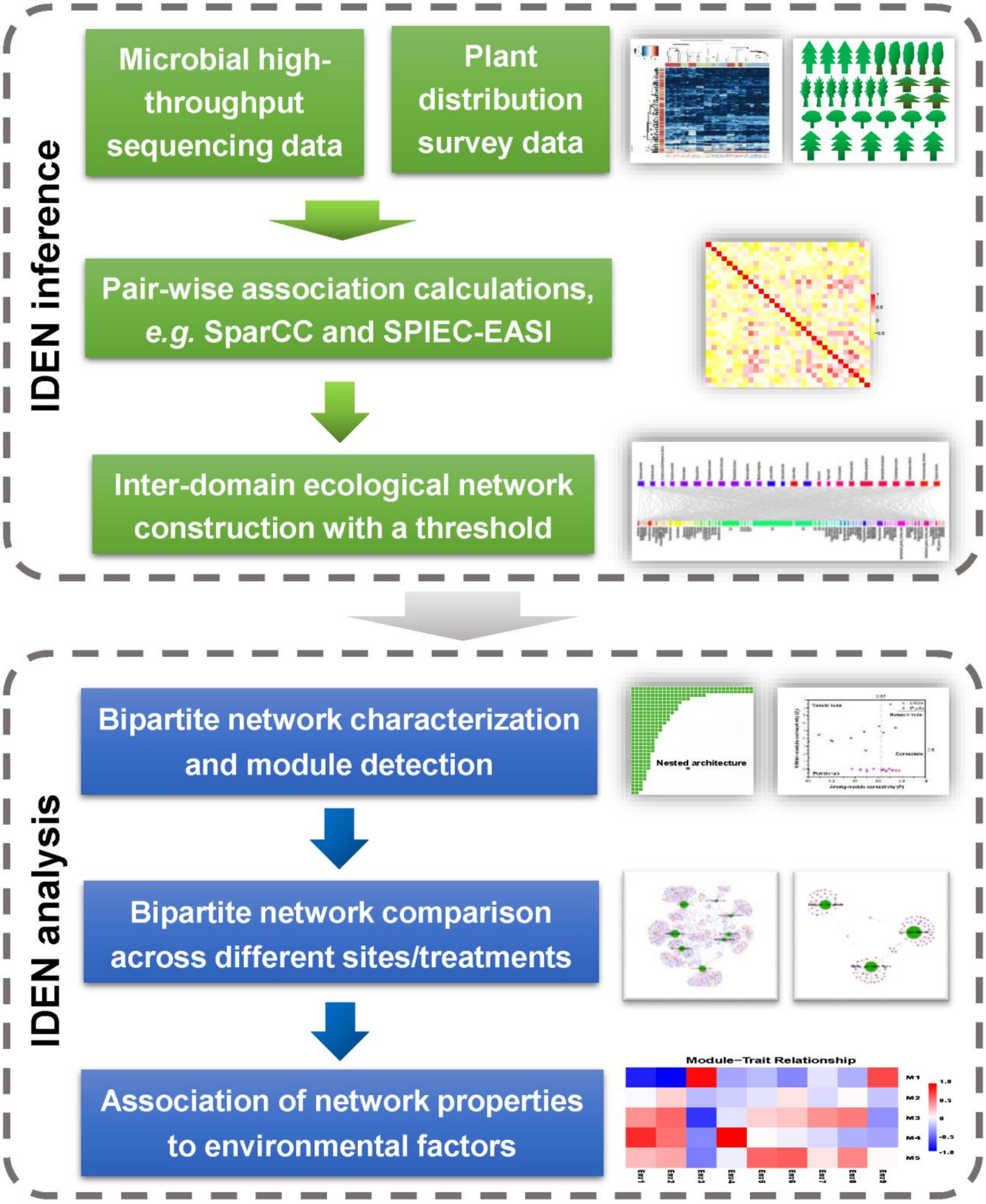
Metagenomics for Environmental Microbiology (MEM)

Research Center for Eco-Environmental Sciences, CAS

Workflow of iNAP



Steps of InterDomain Ecological Network Analysis Pipeline (IDENAP)



IDEN inference

Microbial high-throughput sequencing data

Plant distribution survey data

Pair-wise association calculations, *e.g.* SparCC and SPIEC-EASI

Inter-domain ecological network construction with a threshold

IDEN analysis

Bipartite network characterization and module detection

Bipartite network comparison across different sites/treatments

Association of network properties to environmental factors

Content

If you have questions and is willing to make a contribution to this pipeline, please feel free to contact Prof. Ye Deng (yedeng@rcees.ac.cn).

Pipeline maintainer: Kai Feng (kaifeng@rcees.ac.cn)

Please cite: Kai Feng, Yuguang Zhang, Zhili He, Daliang Ning, Ye Deng. Interdomain ecological networks between plants and microbes. Molecular Ecology Resources. 2019,19 (6):1565-1567. **doi:10.1111/1755-0998.13081**

To users

1. **This pipeline is aimed to explore the inter-domain ecological networks, ignoring the associations between intra-domain species.**
2. **Users are allowed to register account. Anonymous users are also allowed to operate in this pipeline.**
3. **Since the limitation of server memory and the ability of association methods to handle large dataset, we strongly recommend to reduce the No. of species/OTUs, i.e., not exceed 2000 OTUs/species, otherwise the program will be disrupted or take a long time (weeks to months).**
4. **Generally, we recommend SparCC and SPIEC-EASI to construct IDEN for beginners. The other two methods, eLSA and RMT-based correlations, can also be used if you are clear about these methods.**
5. **This pipeline is for research only, and please do not use for commercial activities.**
6. **Since our original goal of this pipeline is focusing on plant-microbe networks, we will add a mark in each of names of species or OTUs to distinguish the two domain species, *i.e.* "_P" to indicate plant and "_M" to indicate microbe. The two marks will be removed during the construction of IDEN.**
7. **In order to reduce the potential errors, please do not include special characters in species names and sample names, e.g. blank character, "-", "#", "$" and etc. And do not use pure numbers in samples names or number started characters, i.e., "11", or "1A" and etc.**
8. **For new users, we provided a test dataset in the Shared library/Test data section.**

Instructions

A. Approaches Network Construction (IDEN)

1. SparCC

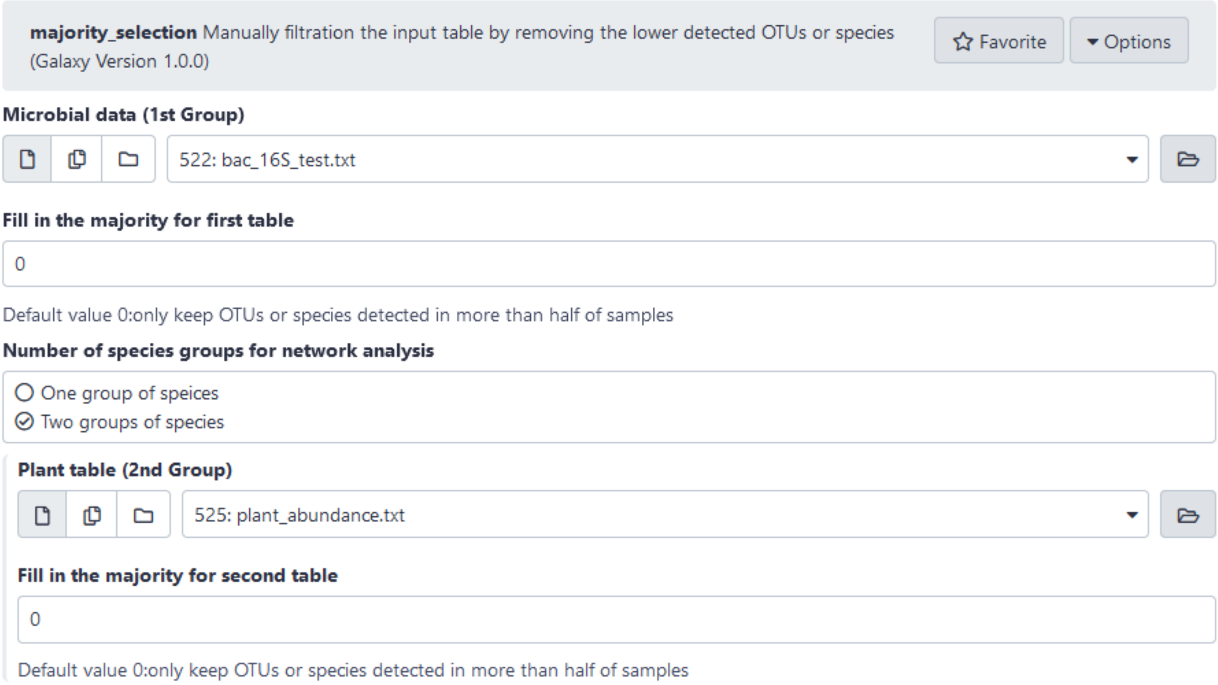
https://bitbucket.org/yonatanf/sparcc/src/default/

Genomic survey data, such as those obtained from 16S rRNA gene sequencing, are subject to underappreciated mathematical difficulties that can undermine standard data analysis techniques. We show that these effects can lead to erroneous correlations among taxa within the human microbiome despite the statistical significance of the associations. To overcome these difficulties, we developed SparCC; a novel procedure, tailored to the properties of genomic survey data, that allow inference of correlations between genes or species. We use SparCC to elucidate networks of interaction among microbial species living in or on the human body.

1.1 Majority selection

The majority is to filter the species or OTUs which were less detected among all samples. The integer of this value is recommended as the 80% of the sample numbers.

Inputs:



majority_selection Manually filtration the input table by removing the lower detected OTUs or species (Galaxy Version 1.0.0)

Microbial data (1st Group)

522: bac_16S_test.txt

Fill in the majority for first table

0

Default value 0:only keep OTUs or species detected in more than half of samples

Number of species groups for network analysis

○ One group of speices
● Two groups of species

Plant table (2nd Group)

525: plant_abundance.txt

Fill in the majority for second table

0

Default value 0:only keep OTUs or species detected in more than half of samples

Parameters:

Input Parameter	Value	Note for rerun
Microbial data	bac_16S_test.txt	OTU/Gene/Species abundance table for microbial community
Fill in the majority for first table	0	selected accordingly

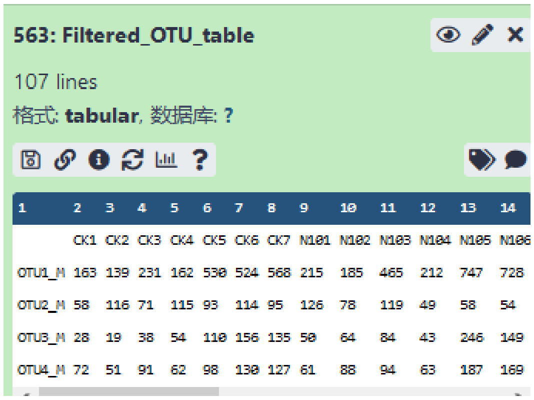
Plant table	plant_abundance.txt	OTU/Gene/Species abundance table for plant community
Fill in the majority for second table	0	selected accordingly

Attention:

Do not use any symbols like ",","(",")","#","-" in your sample names

Do not start your sample name with numbers, e.g. "1A". Please rename it like "A1".

Outputs:



563: Filtered_OTU_table

107 lines

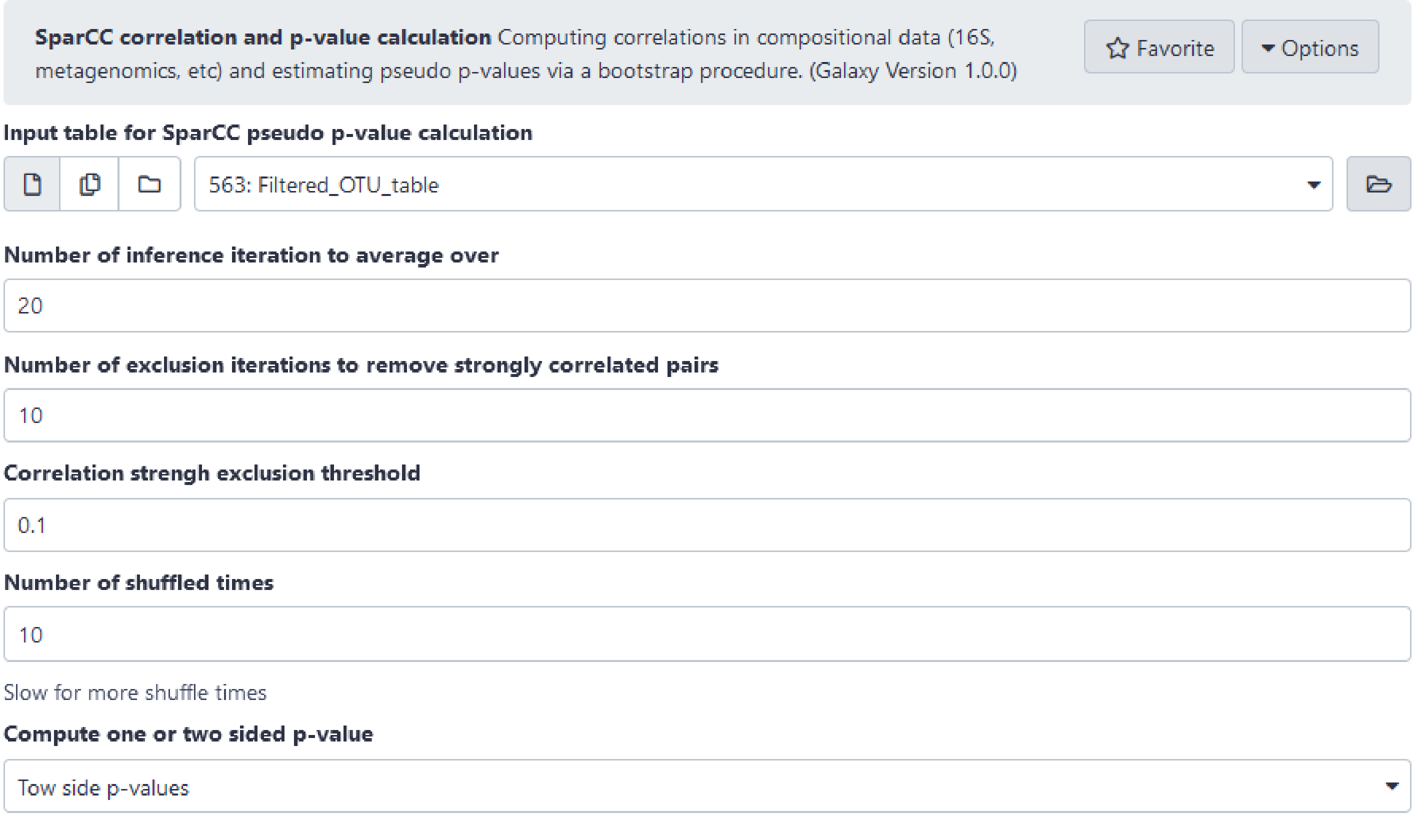
格式: **tabular**, 数据库: ?

1	2	3	4	5	6	7	8	9	10	11	12	13	14
	CK1	CK2	CK3	CK4	CK5	CK6	CK7	N101	N102	N103	N104	N105	N106
OTU1_M	163	139	231	162	530	524	568	215	185	465	212	747	728
OTU2_M	58	116	71	115	93	114	95	126	78	119	49	58	54
OTU3_M	28	19	38	54	110	156	135	50	64	84	43	246	149
OTU4_M	72	51	91	62	98	130	127	61	88	94	63	187	169

One tabular file contained filtered plants and microbes, named with their majorities.

1.2 SparCC correlation and p-value calculation

Inputs:



SparCC correlation and p-value calculation Computing correlations in compositional data (16S, metagenomics, etc) and estimating pseudo p-values via a bootstrap procedure. (Galaxy Version 1.0.0)

Input table for SparCC pseudo p-value calculation

563: Filtered_OTU_table

Number of inference iteration to average over

20

Number of exclusion iterations to remove strongly correlated pairs

10

Correlation strengh exclusion threshold

0.1

Number of shuffled times

10

Slow for more shuffle times

Compute one or two sided p-value

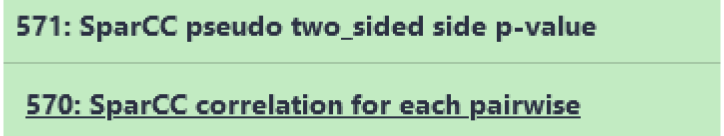
Tow side p-values

Parameters:

Input Parameter	Value	Note for rerun
Input table for SparCC pseudo p-value calculation	Filtered_OTU_table	Filtered OTU/Gene/Species table with majority
Number of inference iteration to average over	20	
Number of exclusion iterations to remove strongly correlated pairs	10	
Correlation strengh exclusion threshold	0.1	
Number of shuffled times	100	Slow for more shuffle times
Compute one or two sided p-value	Tow side p-values	

Outputs:

Two tabular files: one is the correlation matrix, another is the p-value matrix.



571: SparCC pseudo two_sided side p-value

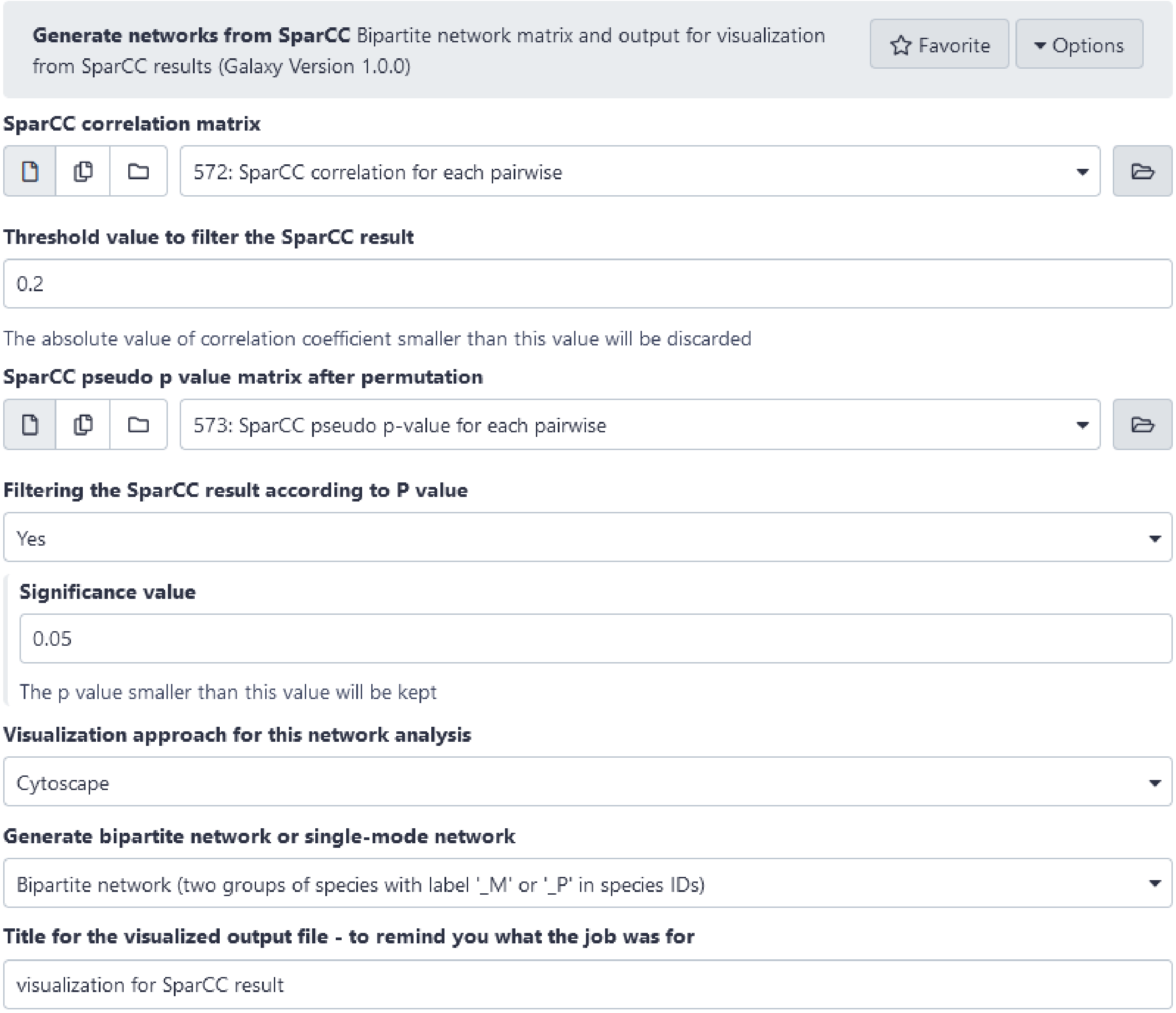
570: SparCC correlation for each pairwise

1.3 SparCC correlation and p-value calculation (FastSpar)

The step FastSpar method to calculate these outputs for the SparCC method, which is an efficient alternative way to handle larger datasets (more than 1,000 species), similar to 1.2.

1.4 Output file from SparCC result

Inputs:



Generate networks from SparCC Bipartite network matrix and output for visualization from SparCC results (Galaxy Version 1.0.0)

SparCC correlation matrix

572: SparCC correlation for each pairwise

Threshold value to filter the SparCC result

0.2

The absolute value of correlation coefficient smaller than this value will be discarded

SparCC pseudo p value matrix after permutation

573: SparCC pseudo p-value for each pairwise

Filtering the SparCC result according to P value

Yes

Significance value

0.05

The p value smaller than this value will be kept

Visualization approach for this network analysis

Cytoscape

Generate bipartite network or single-mode network

Bipartite network (two groups of species with label '_M' or '_P' in species IDs)

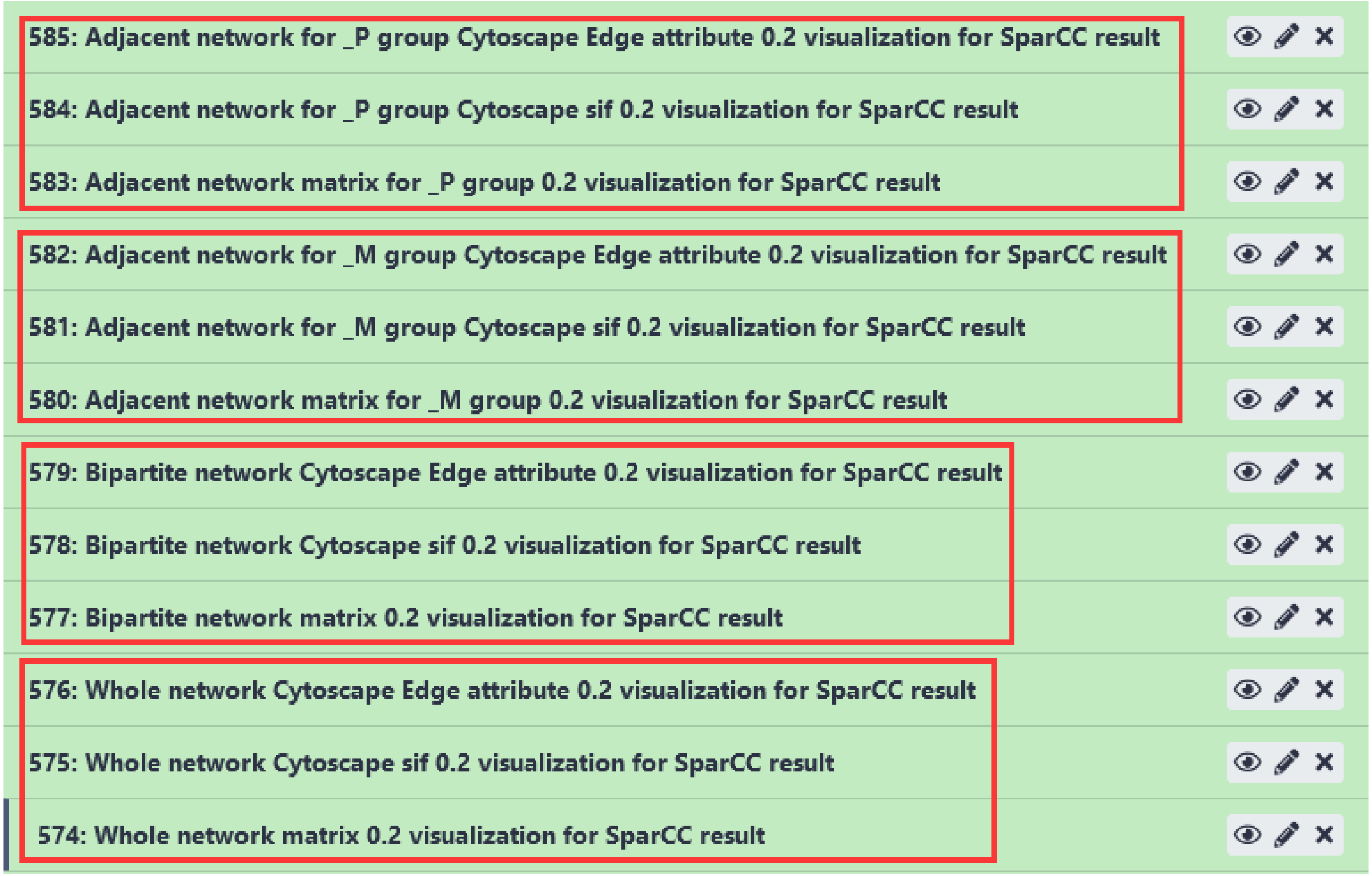
Title for the visualized output file - to remind you what the job was for

visualization for SparCC result

Parameters:

Input Parameter	Value	Note for rerun
SparCC correlation matrix	SparCC correlation for each pairwise	Correlation matrix
Threshold value to filter the SparCC result	0.3	Default value from SparCC
SparCC pseudo p value matrix after permutation	SparCC pseudo p-value for each pairwise	P-value matrix
Filtering the SparCC result according to P value	yes	Default
Significance value	0.05	
Visualization approach for this network analysis	Cytoscape	Cytoscape or Gephi
Generate bipartite network or single-mode network	Bipartite network (two groups of species with label '_M' or '_P' in species IDs)	Bipartite or Adjacent
Title for the visualized output file - to remind you what the job was for	visualization for SparCC result	Put anything you want to remember

Outputs:



585: Adjacent network for _P group Cytoscape Edge attribute 0.2 visualization for SparCC result

584: Adjacent network for _P group Cytoscape sif 0.2 visualization for SparCC result

583: Adjacent network matrix for _P group 0.2 visualization for SparCC result

582: Adjacent network for _M group Cytoscape Edge attribute 0.2 visualization for SparCC result

581: Adjacent network for _M group Cytoscape sif 0.2 visualization for SparCC result

580: Adjacent network matrix for _M group 0.2 visualization for SparCC result

579: Bipartite network Cytoscape Edge attribute 0.2 visualization for SparCC result

578: Bipartite network Cytoscape sif 0.2 visualization for SparCC result

577: Bipartite network matrix 0.2 visualization for SparCC result

576: Whole network Cytoscape Edge attribute 0.2 visualization for SparCC result

575: Whole network Cytoscape sif 0.2 visualization for SparCC result

574: Whole network matrix 0.2 visualization for SparCC result

Twelve output files (Cytoscape) for one whole network, one bipartite network and two adjacent networks. For each matrix, there are a ‘sif’ file to importing into Cytoscape and an edge attribute file.

Eight output files (Gephi) for one whole network, one bipartite network and two adjacent networks. For each matrix, there is a ‘csv’ file to importing into Gephi as edge file.

2. eLSA

The Local Similarity Analysis (LSA) technique is unique to capture the time-dependent associations (possibly time-shifted) between microbes and between microbe and environmental factors (Ruan et al., 2006). Significant LSA associations can be interpreted as a partially directed association network for further network-based analysis. A similar approach called Local Trend Analysis (LTA) has also been developed for the state change series, where a relative change threshold is applied to convert the original time series data into up-change, no-change and down-change state series (Xia et al 2015). Many advanced network analysis tools (including ELSA) have been analyzed in a benchmark paper published in the ISME Journal (Weiss et al. 2016). The more introduction can be found in the home page (https://bitbucket.org/charade/elsa/wiki/Home).

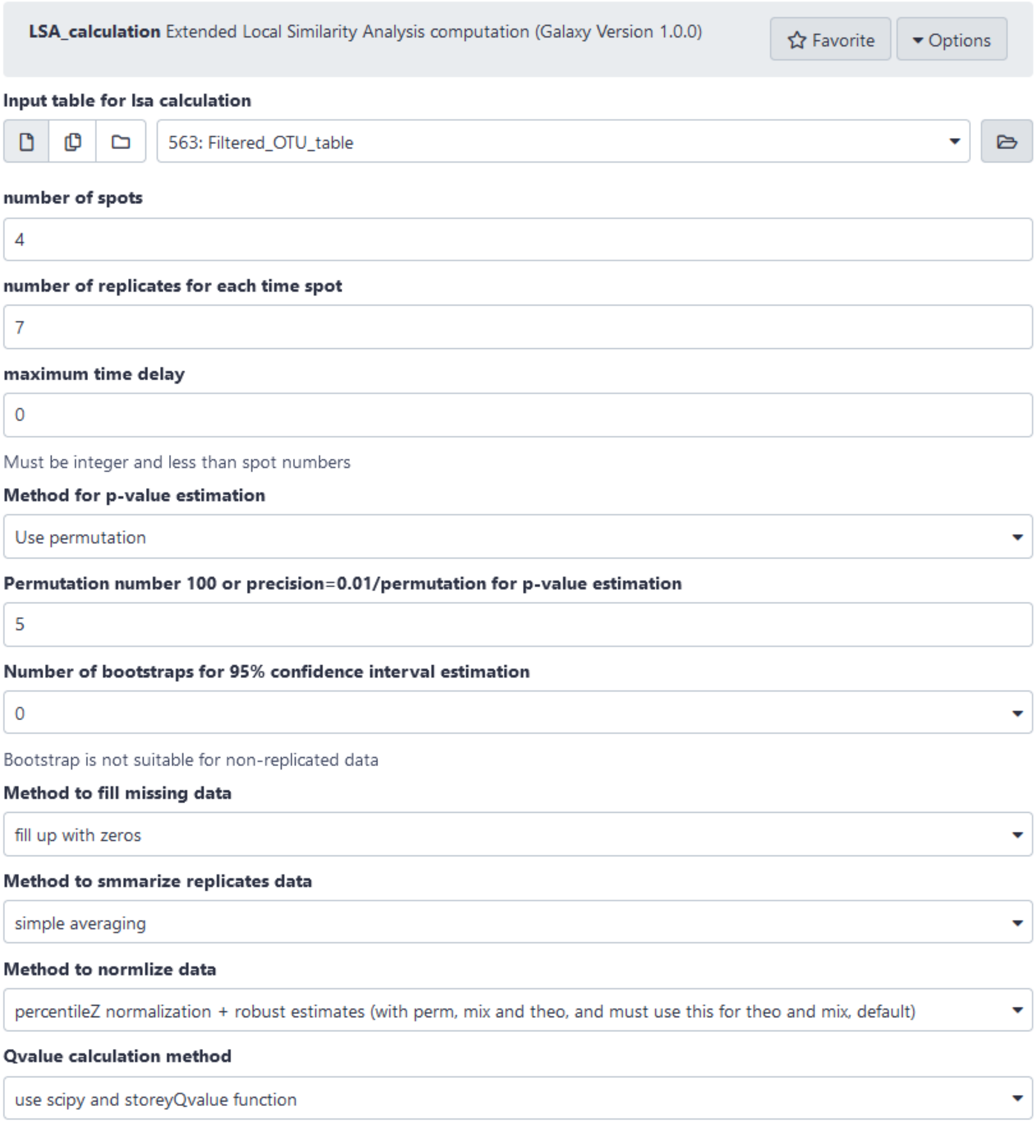
2.1 Majority selection

Similar steps to **1.1**.

2.2 LSA calculation

This step is aimed to calculate the LS for each pairwise association.

Inputs:



LSA_calculation Extended Local Similarity Analysis computation (Galaxy Version 1.0.0)

Input table for lsa calculation

563: Filtered_OTU_table

number of spots

4

number of replicates for each time spot

7

maximum time delay

0

Must be integer and less than spot numbers

Method for p-value estimation

Use permutation

Permutation number 100 or precision=0.01/permutation for p-value estimation

5

Number of bootstraps for 95% confidence interval estimation

0

Bootstrap is not suitable for non-replicated data

Method to fill missing data

fill up with zeros

Method to smmarize replicates data

simple averaging

Method to normlize data

percentileZ normalization + robust estimates (with perm, mix and theo, and must use this for theo and mix, default)

Qvalue calculation method

use scipy and storeyQvalue function

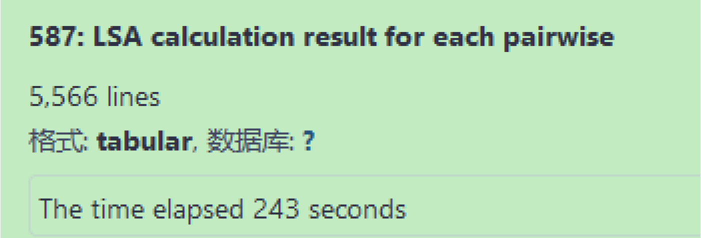
Parameters:

Input Parameter	Value	Note for rerun
Input table for lsa calculation	Filtered_OTU_table	
number of spots	4	
number of replicates for each time spot	7	
maximum time delay	0	Integer and less than spot numbers
Method for p-value estimation	Use permutation	
Permutation number 100 or precision=0.01/permutation for p-value estimation	100	

Number of bootstraps for 95% confidence interval estimation	0	Bootstrap is not suitable for non-replicated data
Method to fill missing data	fill up with zeros	
Method to smmarize replicates data	simple averaging	
Method to normlize data	percentileZ normalization + robust estimates (with perm, mix and theo, and must use this for theo and mix, default)	
Qvalue calculation method	R's qvalue package	

Outputs:

One tabular file contained local similarity score (LS) and other relevant scores.



587: LSA calculation result for each pairwise

5,566 lines

格式: **tabular**, 数据库: **?**

The time elapsed 243 seconds

Output explaination for each term

-X: factor name X

-Y: factor name Y

-LS: Local Similarity Score

-low/upCI: low or up 95% CI for LS

-Xs: align starts position in X

-Ys: align starts position in Y

-Len: align length

-Delay: calculated delay for align, Xs-Ys

-P,Q: p/q-value for LS

-PCC,Ppcc,Qpcc: Pearson's Correlation Coefficient, p/q-value for PCC

-SCC,Pscc,Qscc: Spearman's Correlation Coefficient, p/q-value for SCC

-SPCC,Pspcc,Qspcc,Dspcc: delay-Shifted Pearson's Correlation Coefficient, p/q-value, delay size for SPCC

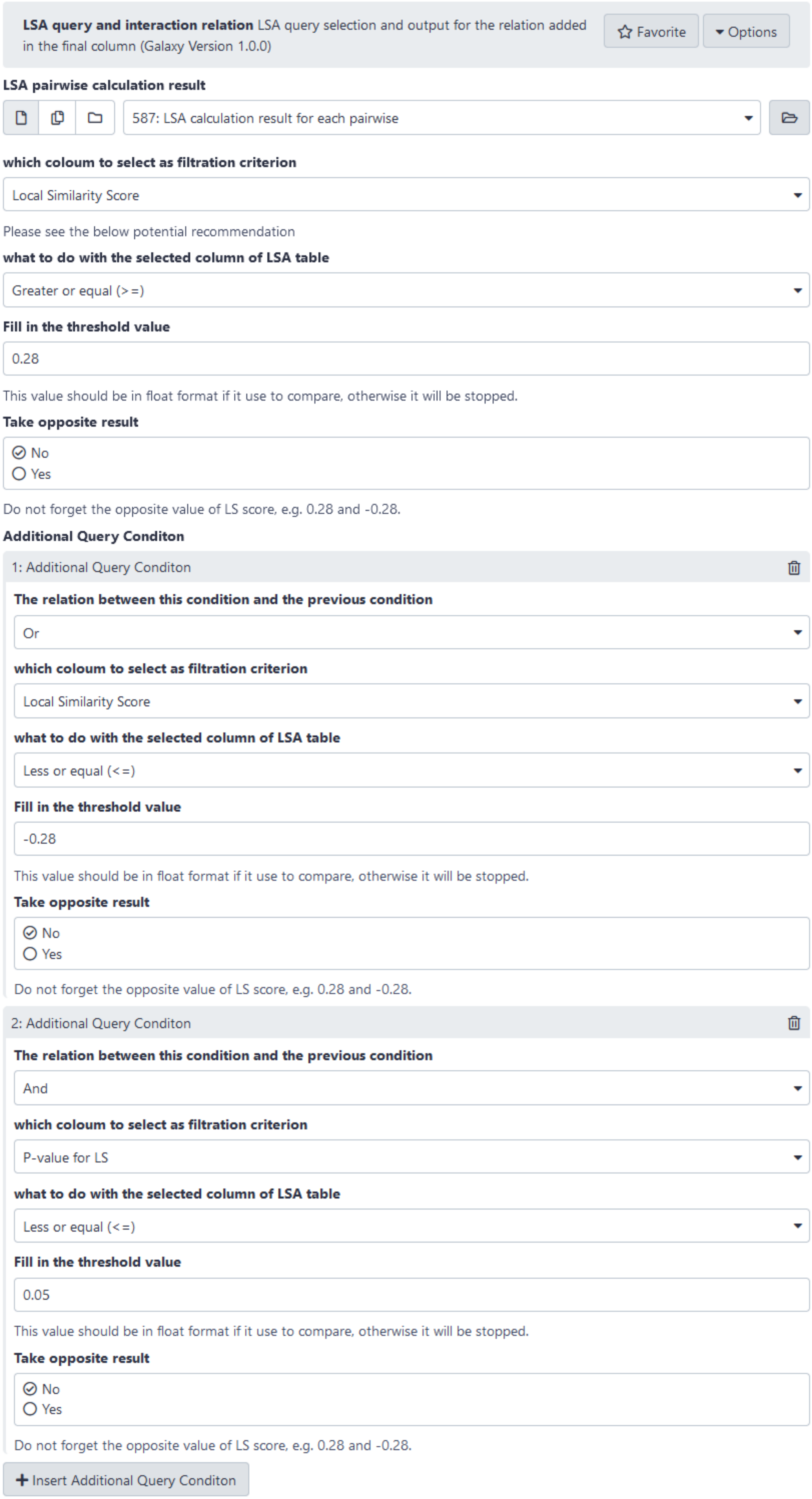
-SSCC,Psscc,Qsscc,Dsscc: delay-Shifted Spearman's Correlation Coefficient, p/q-value, delay size for SSCC

2.3 LSA query and interaction relation

This step is to filter LS according to your own criterion.

The matching pattern for LSA query should be filled correctly according to the recommended format. [!]Key1[>,<,>=,< =,==,!=]V1[|,&][!]Key2[>,<,>=,<=,==,!=]V2[|,&]... and any groupings using '(' and ')' e.g.'(!P>0.01)&($Q<0.01)'. You could choose the parameter provided to make a query formula showing previously. In this step, we provide all the possible selections to fulfill the requirement for your query choice.

Inputs:



LSA query and interaction relation LSA query selection and output for the relation added in the final column (Galaxy Version 1.0.0)

LSA pairwise calculation result

587: LSA calculation result for each pairwise

which coloum to select as filtration criterion

Local Similarity Score

Please see the below potential recommendation

what to do with the selected column of LSA table

Greater or equal (>=)

Fill in the threshold value

0.28

This value should be in float format if it use to compare, otherwise it will be stopped.

Take opposite result

● No
○ Yes

Do not forget the opposite value of LS score, e.g. 0.28 and -0.28.

Additional Query Conditon

1: Additional Query Conditon

The relation between this condition and the previous condition

Or

which coloum to select as filtration criterion

Local Similarity Score

what to do with the selected column of LSA table

Less or equal (<=)

Fill in the threshold value

-0.28

This value should be in float format if it use to compare, otherwise it will be stopped.

Take opposite result

● No
○ Yes

Do not forget the opposite value of LS score, e.g. 0.28 and -0.28.

2: Additional Query Conditon

The relation between this condition and the previous condition

And

which coloum to select as filtration criterion

P-value for LS

what to do with the selected column of LSA table

Less or equal (<=)

Fill in the threshold value

0.05

This value should be in float format if it use to compare, otherwise it will be stopped.

Take opposite result

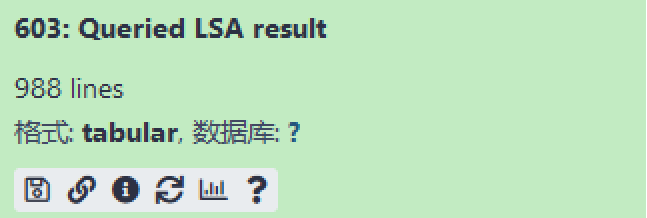
● No
○ Yes

Do not forget the opposite value of LS score, e.g. 0.28 and -0.28.

Parameters:

Input Parameter	Value	Note for rerun
LSA pairwise calculation result	LSA calculation result for each pairwise	
which coloum to select as filtration criterion	Local Similarity Score	1st condition
what to do with the selected column of LSA table	Greater or equal (>=)	
Fill in the threshold value	0.8	
Take opposite result	No	
The relation between this condition and the previous condition	Or	2nd condition
which coloum to select as filtration criterion	Local Similarity Score	
what to do with the selected column of LSA table	Less or equal (<=)	
Fill in the threshold value	-0.8	
Take opposite result	No	
The relation between this condition and the previous condition	And	3rd condition
which coloum to select as filtration criterion	P-value for LS	
what to do with the selected column of LSA table	Less or equal (<=)	
Fill in the threshold value	0.05	
Take opposite result	No	

Outputs:



603: Queried LSA result

988 lines

格式: **tabular**, 数据库: **?**

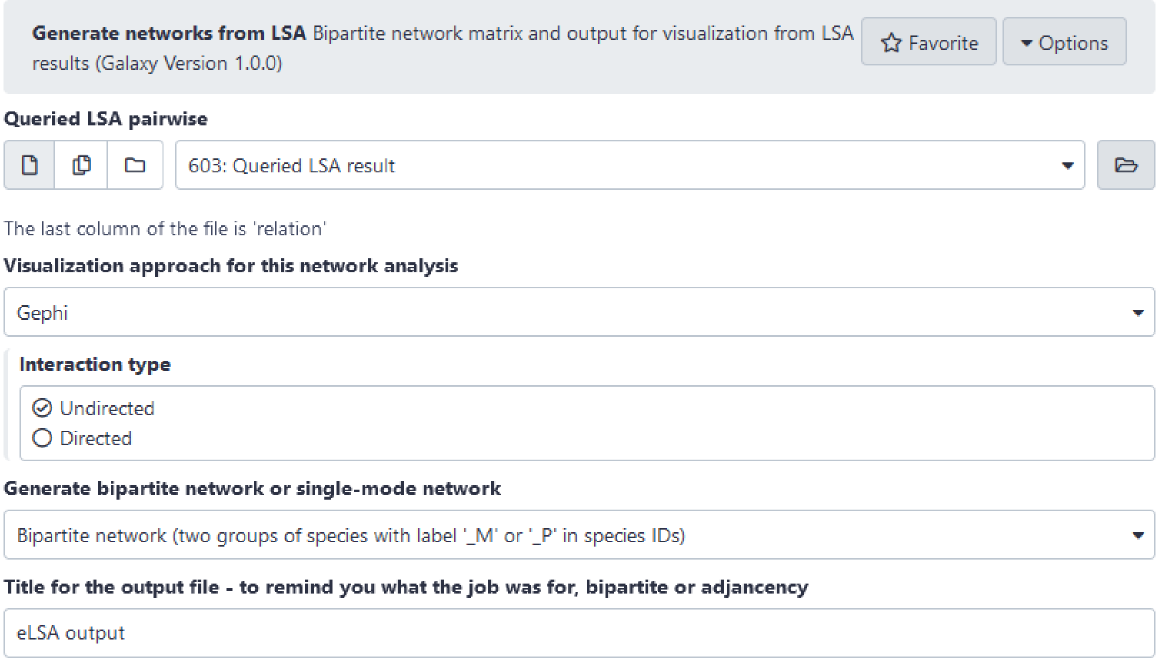
About the interaction type

- pu: positive undirected
- nu: negative undirected
- pdl: positive directed lead (X lead Y)
- ndl: negative directed lead (X lead Y)
- pdr: positive directed retard (X retard Y)
- ndr: negative directed retard (X retard Y)

2.4 Output file from LSA result

This step will generate three output files, including bipartite network matrix (or adjancency matrix), visualization input (for Cytoscape or Gephi) and associated edge attributes if needed.

Inputs:



Generate networks from LSA Bipartite network matrix and output for visualization from LSA results (Galaxy Version 1.0.0)

Queried LSA pairwise

603: Queried LSA result

The last column of the file is 'relation'

Visualization approach for this network analysis

Gephi

Interaction type

☑ Undirected
○ Directed

Generate bipartite network or single-mode network

Bipartite network (two groups of species with label '_M' or '_P' in species IDs)

Title for the output file - to remind you what the job was for, bipartite or adjancency

eLSA output

Parameters:

Input Parameter	Value	Note for rerun
Queried LSA pairwise	81: Queried LSA result	
Visualization approach for this network analysis	Gephi	Cytoscape / Gephi
Interaction type	Undirected	Only for Gephi
Generate bipartite network or single-mode network	Bipartite network (two groups of species with label '_M' or '_P' in species IDs)	Bipartite or Adjacent
Title for the output file - to remind you what the job was for, bipartite or adjancency	Bipartite network matrix of LSA	Put anything you want

Outputs:

639: Adjacent network for _P group Gephi Edge attribute eLSA output

638: Adjacent network matrix for _P group eLSA output

637: Adjacent network for _M group Gephi Edge attribute eLSA output

636: Adjacent network matrix for _M group eLSA output

635: Bipartite network Gephi Edge attribute eLSA output

634: Bipartite network matrix eLSA output

633: Whole network Gephi Edge attribute eLSA output

632: Whole network matrix eLSA output

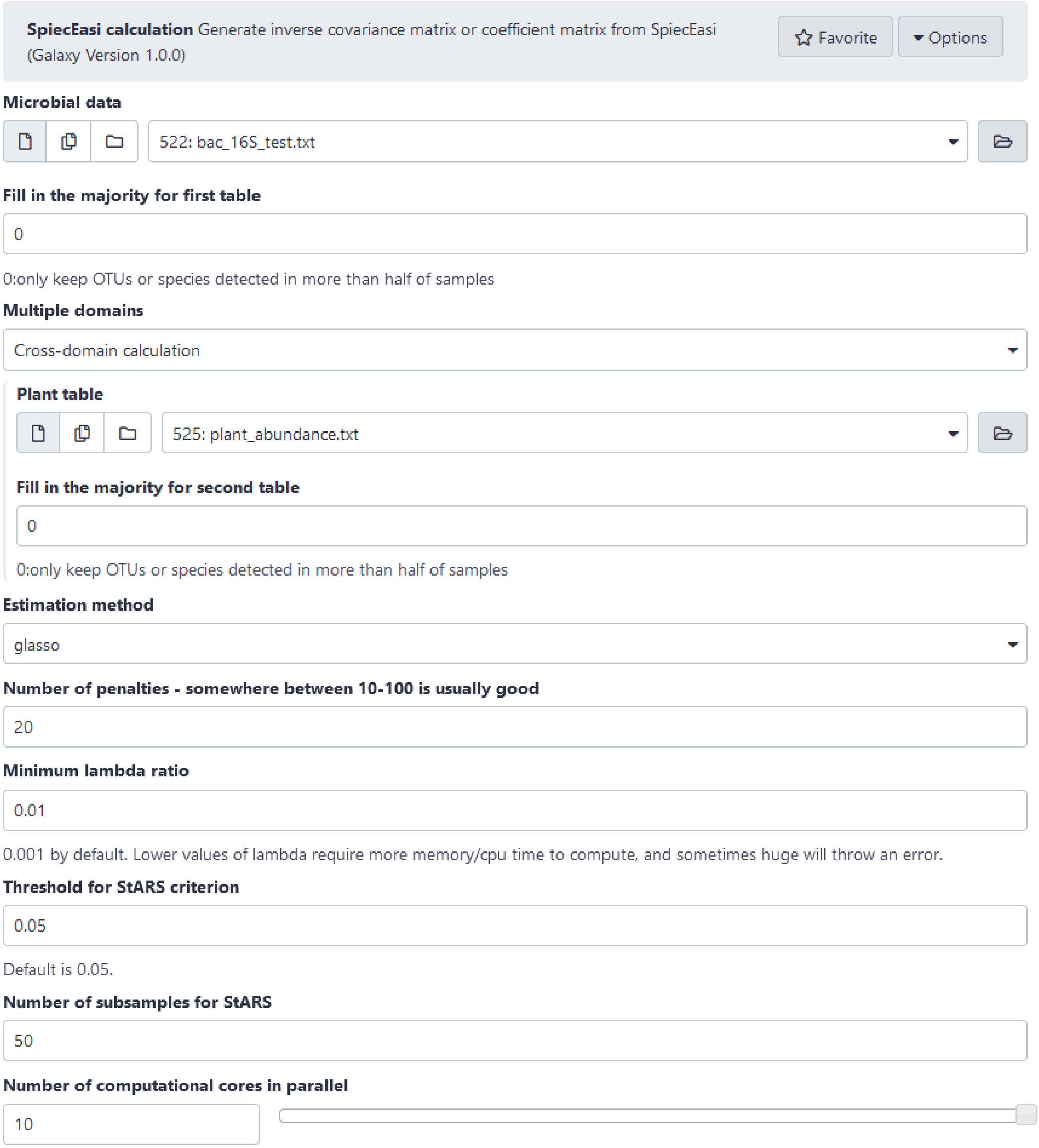
Files similar to 1.4

3. SPIEC-EASI

Sparse InversE Covariance estimation for Ecological Association and Statistical Inference. Please see more in https://github.com/zdk123/SpiecEasi#cross-domain-interactions.

3.1 SpiecEasi calculation

Inputs:



SpiecEasi calculation Generate inverse covariance matrix or coefficient matrix from SpiecEasi (Galaxy Version 1.0.0)

Microbial data

522: bac_16S_test.txt

Fill in the majority for first table

0

0:only keep OTUs or species detected in more than half of samples

Multiple domains

Cross-domain calculation

Plant table

525: plant_abundance.txt

Fill in the majority for second table

0

0:only keep OTUs or species detected in more than half of samples

Estimation method

glasso

Number of penalties - somewhere between 10-100 is usually good

20

Minimum lambda ratio

0.01

0.001 by default. Lower values of lambda require more memory/cpu time to compute, and sometimes huge will throw an error.

Threshold for StARS criterion

0.05

Default is 0.05.

Number of subsamples for StARS

50

Number of computational cores in parallel

10

Parameters:

Input Parameter	Value	Note for rerun
Microbial data	bac_16S_test.txt	
Fill in the majority for first table	8	
Multiple domains	T	Inter-domain or intra-domain

Plant table	plant_abundance.txt	
Fill in the majority for second table	6	
Estimation method	glasso	Glasso/MB selection
Number of penalties - somewhere between 10-100 is usually good	50	
Threshold for StARS criterion	0.05	Default
Minimum lambda	0.1	Default: 0.01
Number of subsamples for StARS	10	According to no. of samples
Number of computational cores in parallel	2	Not too large

Outputs:

136: Filtered matrix with majority

135: SpiecEasi report

134: SpiecEasi matrix

SpiecEasi matrix: A matrix for inverse covariance, used to infer associations.

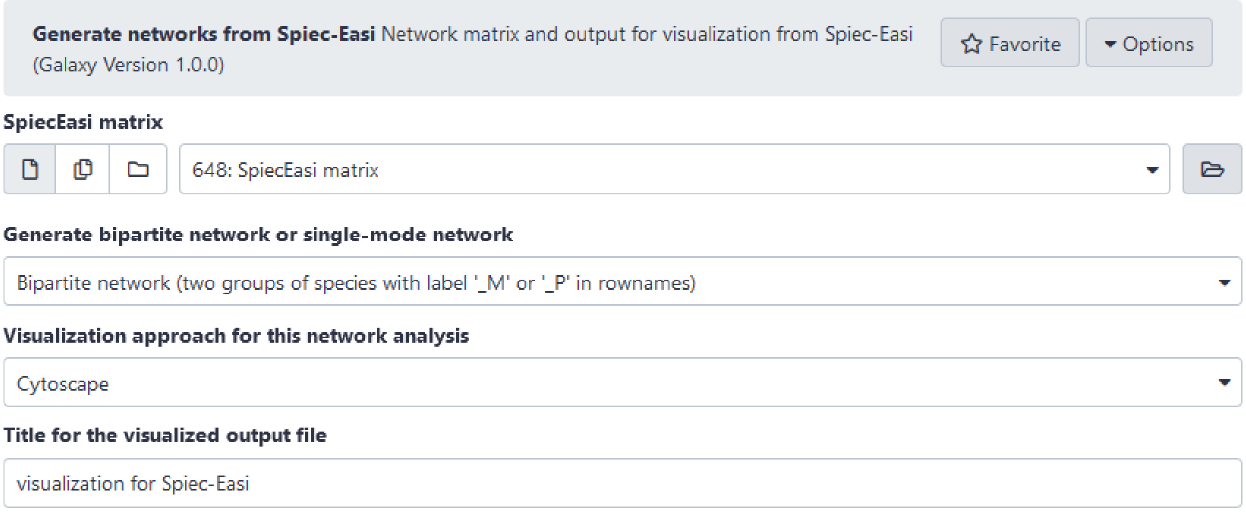
SpiecEasi report: A report for SPIEC-EASI processing.

Filtered matrix with majority: OTU/Gene/Species table after majority selection

3.2 Output file from Spiec-Easi

This step will generate three output files, including bipartite network matrix, visualization input (for Cytoscape or Gephi) and associated edge attributes if needed.

Inputs:



Generate networks from Spiec-Easi Network matrix and output for visualization from Spiec-Easi (Galaxy Version 1.0.0)

SpiecEasi matrix
648: SpiecEasi matrix

Generate bipartite network or single-mode network
Bipartite network (two groups of species with label '_M' or '_P' in rownames)

Visualization approach for this network analysis
Cytoscape

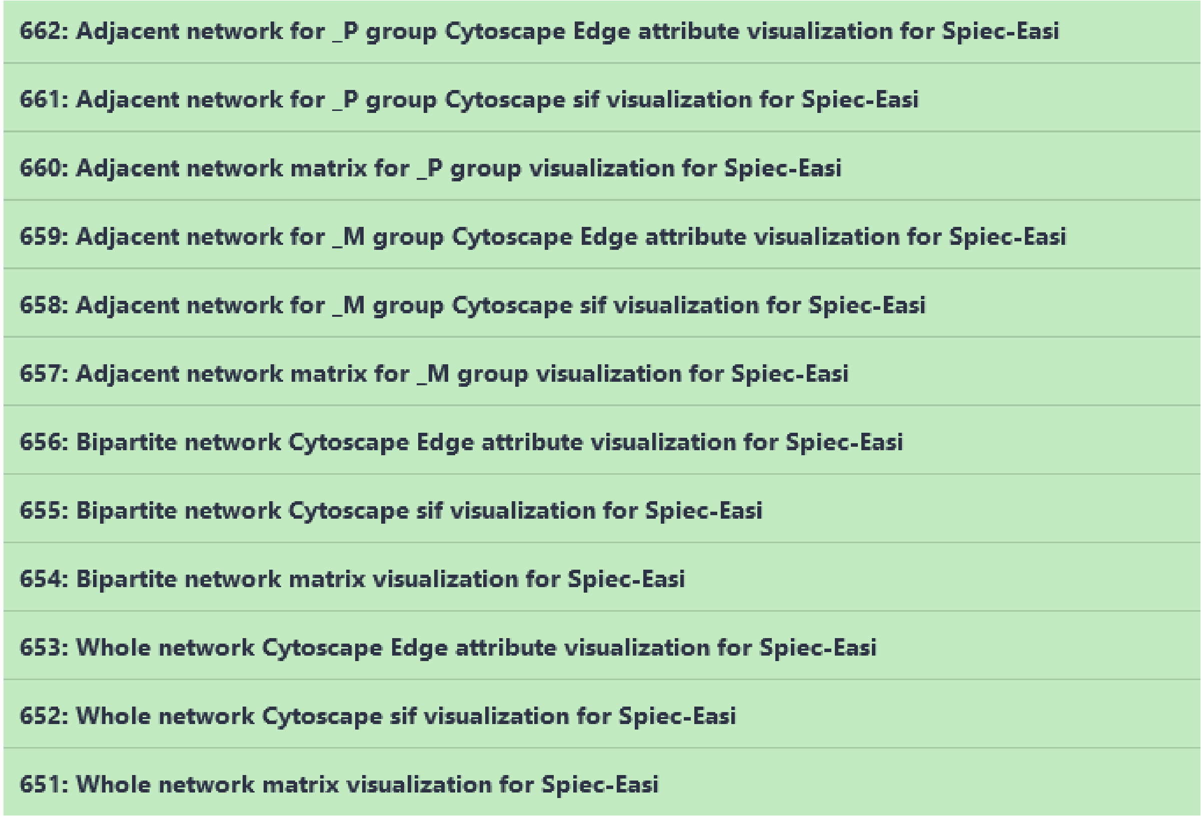
Title for the visualized output file
visualization for Spiec-Easi

Parameters:

Input Parameter	Value	Note for rerun
SpiecEasi matrix	SpiecEasi matrix	

Generate bipartite network or single-mode network	Bipartite network (two groups of species with label '_M' or '_P' in species IDs)	Bipartite or Adjacent
Generate bipartite network or single-mode network	Bipartite network	Bipartite network or single-mode
Visualization approach for this network analysis	Cytoscape	Cytoscape/Gephi
Title for the visualized output file	visualization for Spiec-Easi	Put anything you want

Outputs:



662: Adjacent network for _P group Cytoscape Edge attribute visualization for Spiec-Easi

661: Adjacent network for _P group Cytoscape sif visualization for Spiec-Easi

660: Adjacent network matrix for _P group visualization for Spiec-Easi

659: Adjacent network for _M group Cytoscape Edge attribute visualization for Spiec-Easi

658: Adjacent network for _M group Cytoscape sif visualization for Spiec-Easi

657: Adjacent network matrix for _M group visualization for Spiec-Easi

656: Bipartite network Cytoscape Edge attribute visualization for Spiec-Easi

655: Bipartite network Cytoscape sif visualization for Spiec-Easi

654: Bipartite network matrix visualization for Spiec-Easi

653: Whole network Cytoscape Edge attribute visualization for Spiec-Easi

652: Whole network Cytoscape sif visualization for Spiec-Easi

651: Whole network matrix visualization for Spiec-Easi

Twelve output files (Cytoscape) for one whole network, one bipartite network and two adjacent networks. For each matrix, there are a 'sif' file to importing into Cytoscape and an edge attribute file.

Eight output files (Gephi) for one whole network, one bipartite network and two adjacent networks. For each matrix, there is a 'csv' file to importing into Gephi as edge file.

4. RMT-based method for Pearson/Spearman correlations

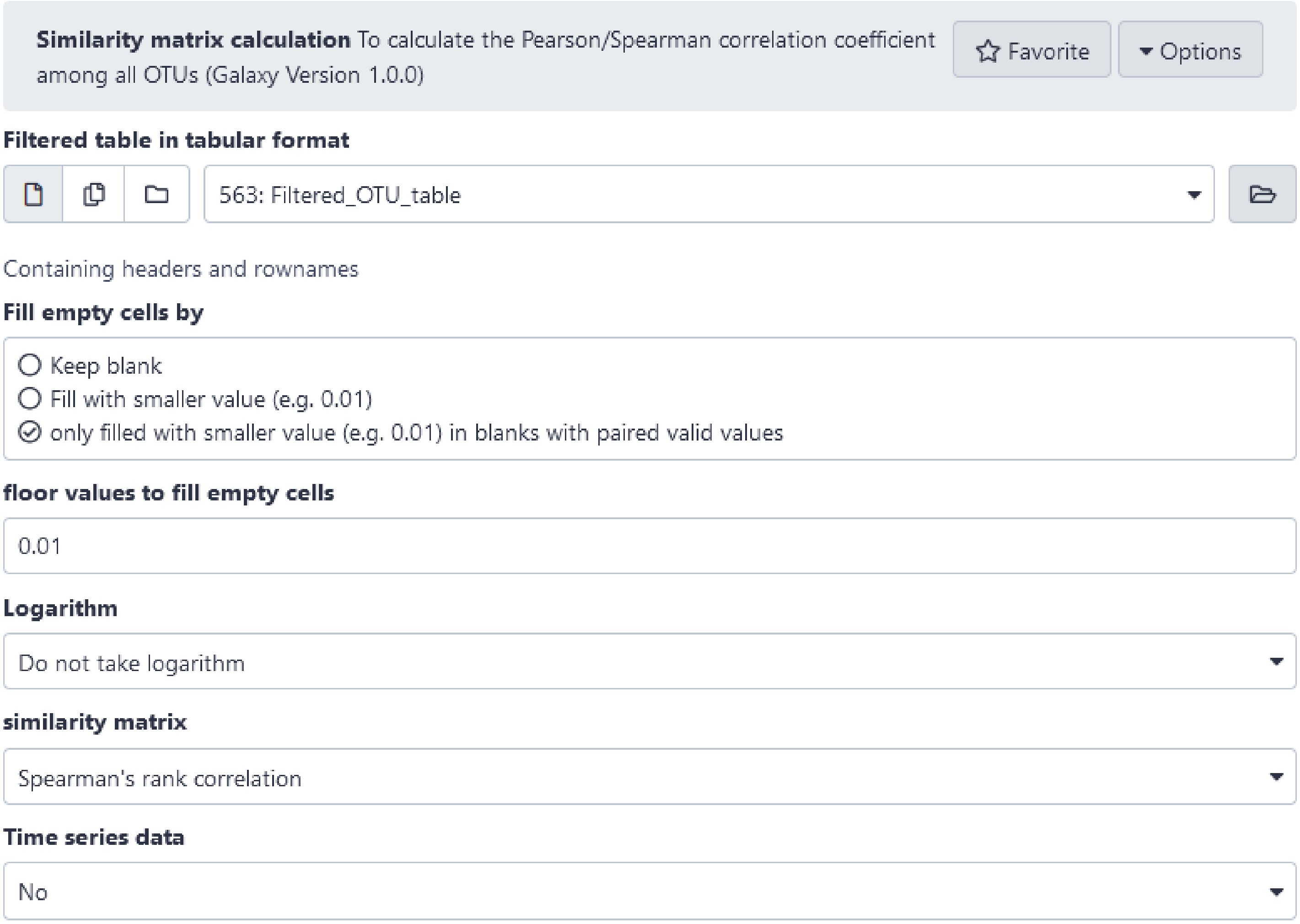
Generate bipartite networks using Pearson/Spearman correlations with RMT-based cutoff. (Coming soon)

4.1 Majority selection

Similar steps to 1.1.

4.2 Similarity matrix calculation

Input:



Similarity matrix calculation To calculate the Pearson/Spearman correlation coefficient among all OTUs (Galaxy Version 1.0.0)

Filtered table in tabular format

563: Filtered_OTU_table

Containing headers and rownames

Fill empty cells by

○ Keep blank
○ Fill with smaller value (e.g. 0.01)
☑ only filled with smaller value (e.g. 0.01) in blanks with paired valid values

floor values to fill empty cells

0.01

Logarithm

Do not take logarithm

similarity matrix

Spearman's rank correlation

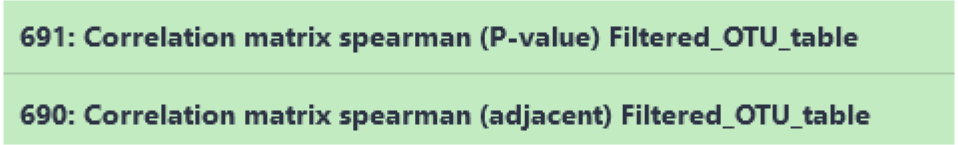
Time series data

No

Parameters:

Input Parameter	Value	Note for rerun
Filtered table in tabular format	Filtered_OTU_table	
Fill empty cells by	only filled with smaller value (e.g. 0.01) in blanks with paired valid values	
floor values to fill empty cells	0.01	Small value
Logarithm	Do not take logarithm	
similarity matrix	Spearman's rank correlation	Spearman/Pearson
Time series data	no	Temporal dataset

Outputs:



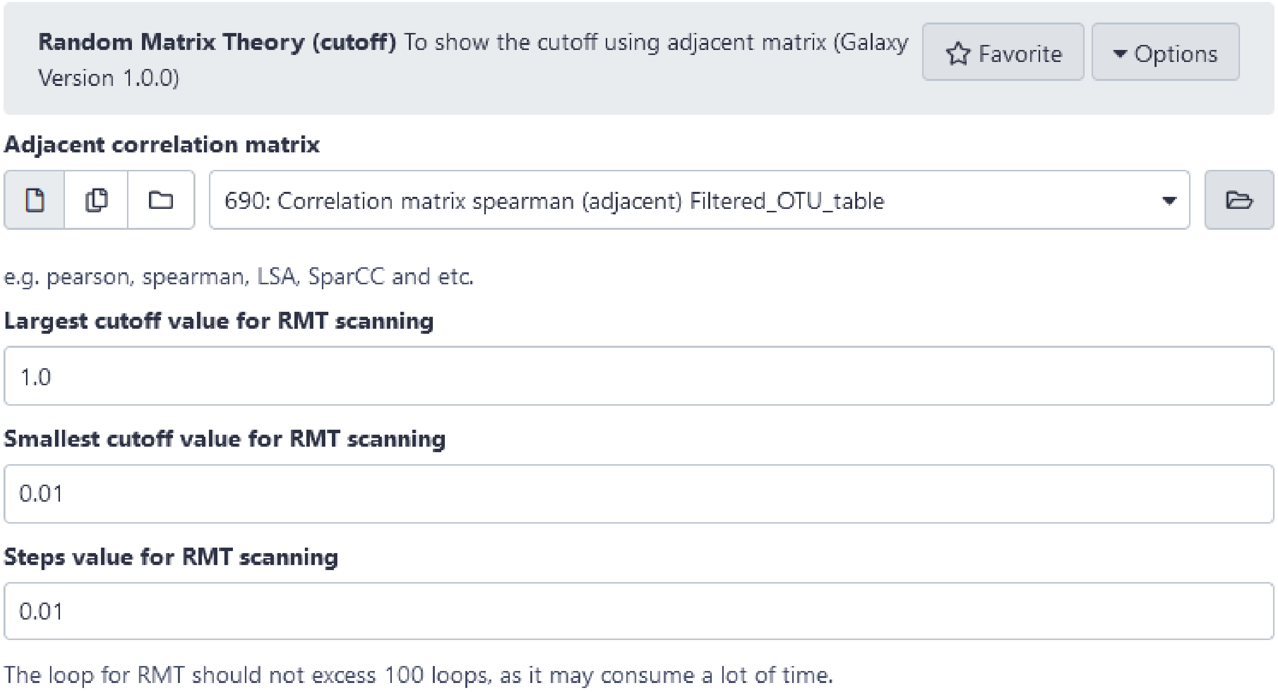
691: Correlation matrix spearman (P-value) Filtered_OTU_table

690: Correlation matrix spearman (adjacent) Filtered_OTU_table

The output file contains an adjacent correlation coefficient matrix and the corresponding P values.

4.3 Random Matrix Theory (cutoff)

Input:



Random Matrix Theory (cutoff) To show the cutoff using adjacent matrix (Galaxy Version 1.0.0)

Adjacent correlation matrix

690: Correlation matrix spearman (adjacent) Filtered_OTU_table

e.g. pearson, spearman, LSA, SparCC and etc.

Largest cutoff value for RMT scanning

1.0

Smallest cutoff value for RMT scanning

0.01

Steps value for RMT scanning

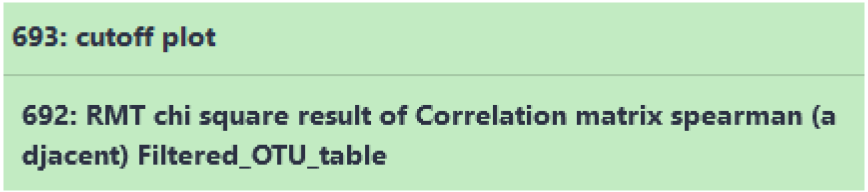
0.01

The loop for RMT should not excess 100 loops, as it may consume a lot of time.

Parameters:

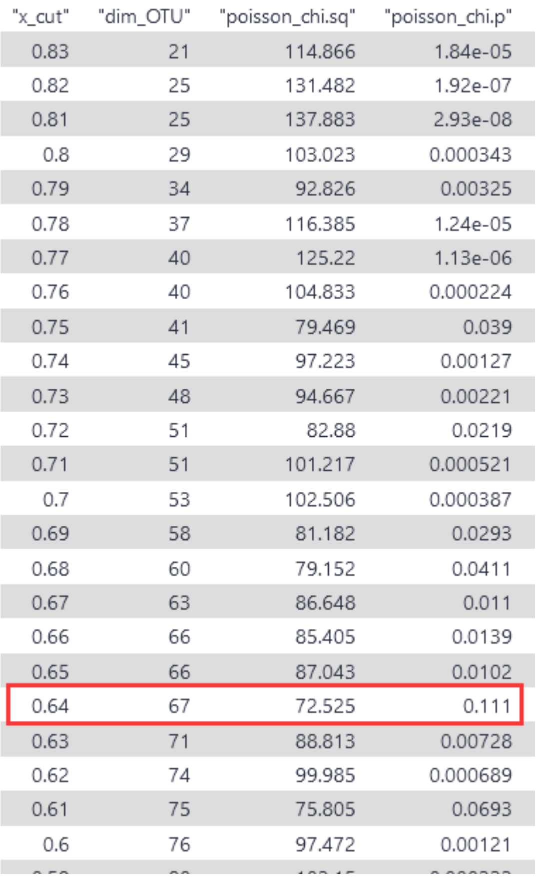
Input Parameter	Value	Note for rerun
Adjacent correlation matrix	Correlation matrix spearman (adjacent) Filtered_OTU_table	
Largest cutoff value for RMT scanning	1.0	
Smallest cutoff value for RMT scanning	0.01	
Steps value for RMT scanning	0.01	

Output:

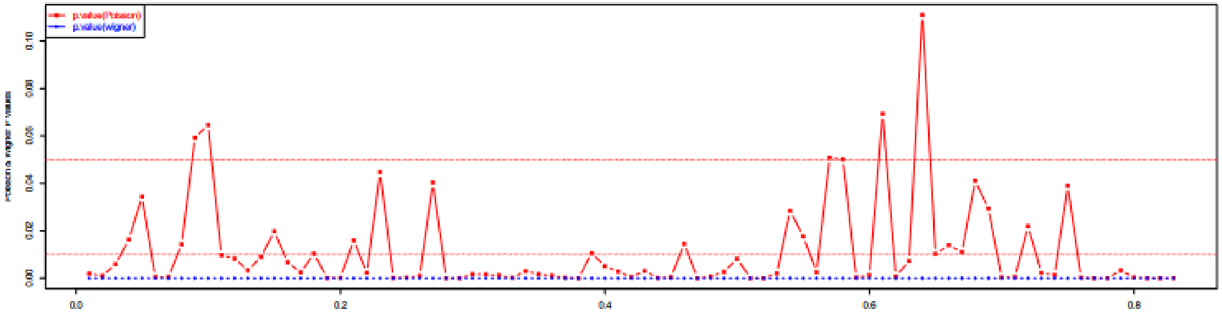


693: cutoff plot

692: RMT chi square result of Correlation matrix spearman (adjacent) Filtered_OTU_table



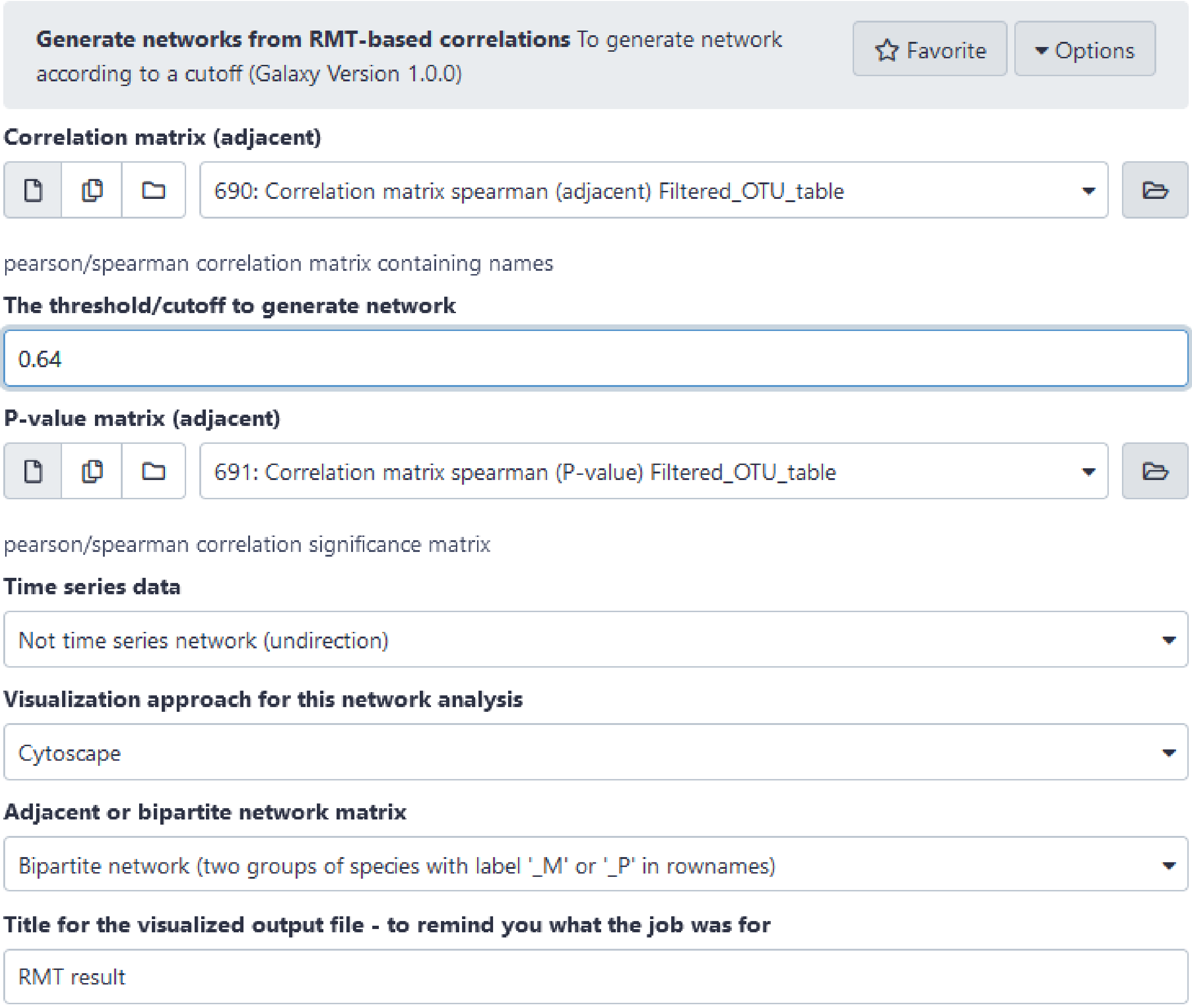
"x_cut"	"dim_OTU"	"poisson_chi.sq"	"poisson_chi.p"
0.83	21	114.866	1.84e-05
0.82	25	131.482	1.92e-07
0.81	25	137.883	2.93e-08
0.8	29	103.023	0.000343
0.79	34	92.826	0.00325
0.78	37	116.385	1.24e-05
0.77	40	125.22	1.13e-06
0.76	40	104.833	0.000224
0.75	41	79.469	0.039
0.74	45	97.223	0.00127
0.73	48	94.667	0.00221
0.72	51	82.88	0.0219
0.71	51	101.217	0.000521
0.7	53	102.506	0.000387
0.69	58	81.182	0.0293
0.68	60	79.152	0.0411
0.67	63	86.648	0.011
0.66	66	85.405	0.0139
0.65	66	87.043	0.0102
0.64	67	72.525	0.111
0.63	71	88.813	0.00728
0.62	74	99.985	0.000689
0.61	75	75.805	0.0693
0.6	76	97.472	0.00121



The output matrix lists the results of cutoff and corresponding goodness-of-fit using χ^2. It is suggested that the recommended cutoff is the value at which the statistics of χ^2 rejects the hypothesis. Another output file shows the cutoff distributions at significance levels of 0.05 and 0.001, and the candidate cutoffs are above the significance lines.

4.4 Generate networks from RMT-based correlations

Input:



Generate networks from RMT-based correlations To generate network according to a cutoff (Galaxy Version 1.0.0)

Correlation matrix (adjacent)

690: Correlation matrix spearman (adjacent) Filtered_OTU_table

pearson/spearman correlation matrix containing names

The threshold/cutoff to generate network

0.64

P-value matrix (adjacent)

691: Correlation matrix spearman (P-value) Filtered_OTU_table

pearson/spearman correlation significance matrix

Time series data

Not time series network (undirection)

Visualization approach for this network analysis

Cytoscape

Adjacent or bipartite network matrix

Bipartite network (two groups of species with label '_M' or '_P' in rownames)

Title for the visualized output file - to remind you what the job was for

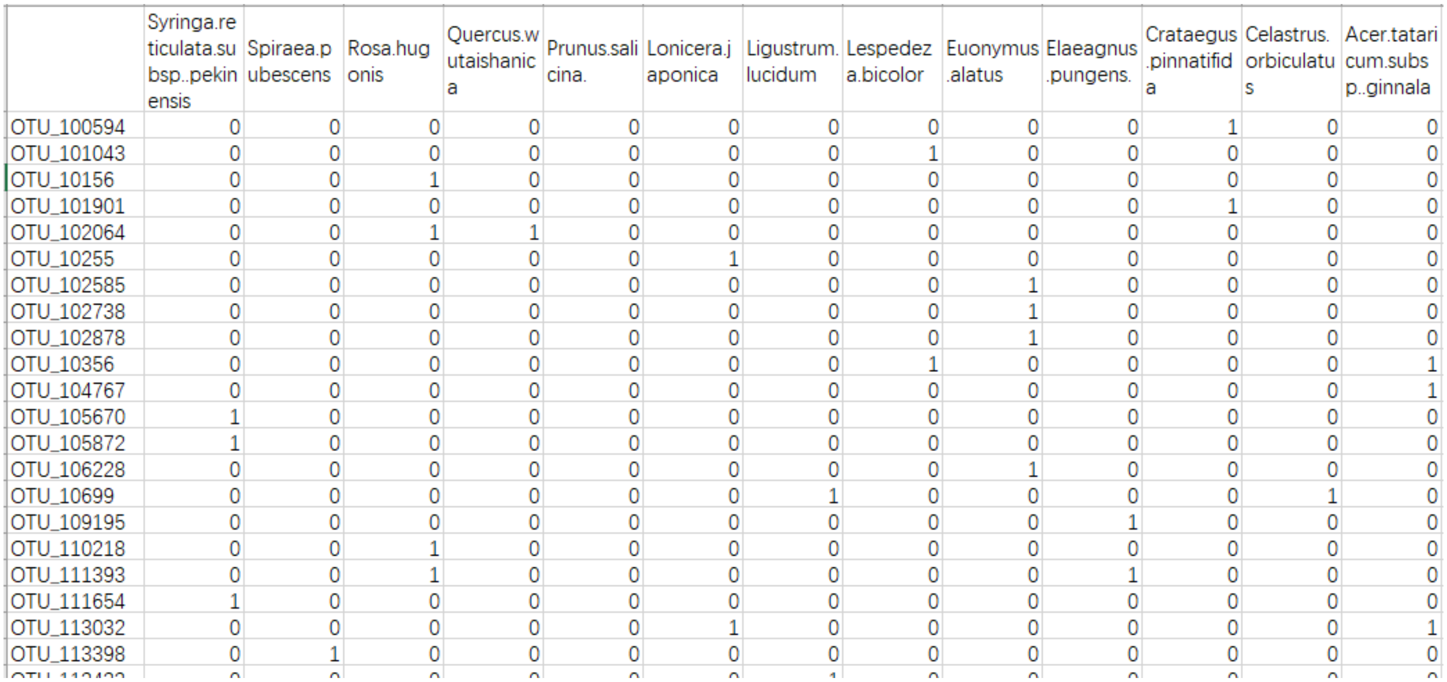
RMT result

Parameters and Outputs:

Network output files would be exported accordingly containing the correlation values and associated P values.

B. Network analysis (IDEN)

All statistics analysis methods were based on bipartite network matrix. The simple format for this matrix is illustrated as following:

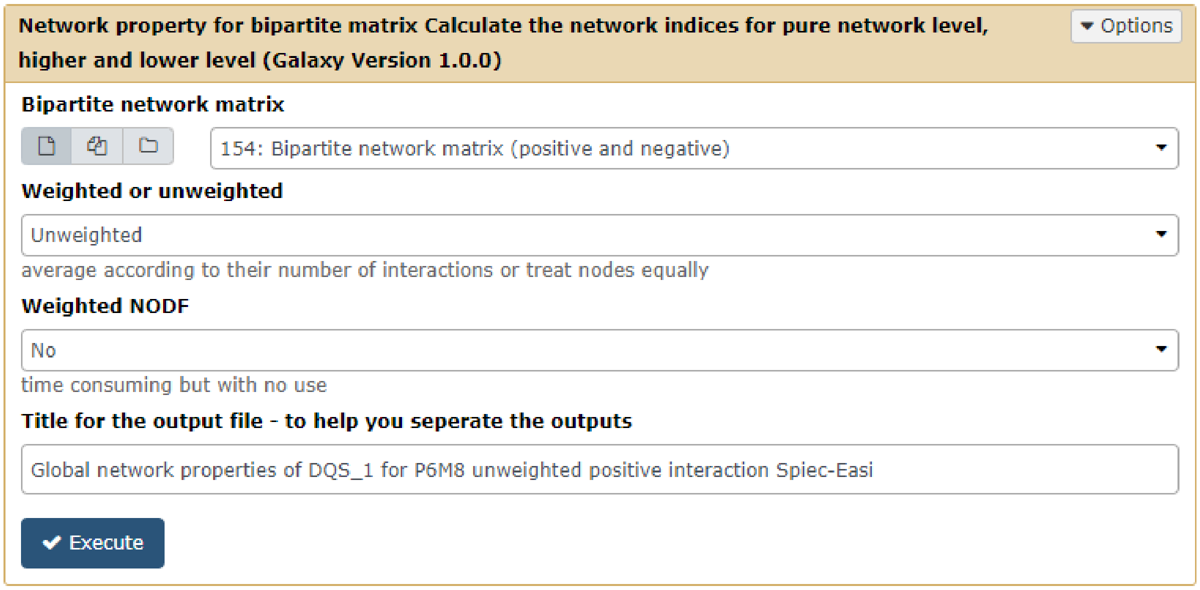


	Syringa.reticulata.subsp..pekinensis	Spiraea.pubescens	Rosa.hugonis	Quercus.wutaishanica	Prunus.salicina.	Lonicera.japonica	Ligustrum.lucidum	Lespedeza.bicolor	Euonymus.alatus	Elaeagnus.pungens.	Crataegus.pinnatifida	Celastrus.orbiculatus	Acer.tataricum.subsp..ginnala
OTU_100594	0	0	0	0	0	0	0	0	0	0	1	0	0
OTU_101043	0	0	0	0	0	0	0	1	0	0	0	0	0
OTU_10156	0	0	1	0	0	0	0	0	0	0	0	0	0
OTU_101901	0	0	0	0	0	0	0	0	0	0	1	0	0
OTU_102064	0	0	1	1	0	0	0	0	0	0	0	0	0
OTU_10255	0	0	0	0	0	1	0	0	0	0	0	0	0
OTU_102585	0	0	0	0	0	0	0	0	1	0	0	0	0
OTU_102738	0	0	0	0	0	0	0	0	1	0	0	0	0
OTU_102878	0	0	0	0	0	0	0	0	1	0	0	0	0
OTU_10356	0	0	0	0	0	0	0	1	0	0	0	0	1
OTU_104767	0	0	0	0	0	0	0	0	0	0	0	0	1
OTU_105670	1	0	0	0	0	0	0	0	0	0	0	0	0
OTU_105872	1	0	0	0	0	0	0	0	0	0	0	0	0
OTU_106228	0	0	0	0	0	0	0	0	1	0	0	0	0
OTU_10699	0	0	0	0	0	0	1	0	0	0	0	1	0
OTU_109195	0	0	0	0	0	0	0	0	0	1	0	0	0
OTU_110218	0	0	1	0	0	0	0	0	0	0	0	0	0
OTU_111393	0	0	1	0	0	0	0	0	0	1	0	0	0
OTU_111654	1	0	0	0	0	0	0	0	0	0	0	0	0
OTU_113032	0	0	0	0	0	1	0	0	0	0	0	0	1
OTU_113398	0	1	0	0	0	0	0	0	0	0	0	0	0

1. Bipartite reports

1.1 Network property for bipartite matrix

Inputs:



Network property for bipartite matrix Calculate the network indices for pure network level, higher and lower level (Galaxy Version 1.0.0)

Bipartite network matrix

154: Bipartite network matrix (positive and negative)

Weighted or unweighted

Unweighted

average according to their number of interactions or treat nodes equally

Weighted NODF

No

time consuming but with no use

Title for the output file - to help you seperate the outputs

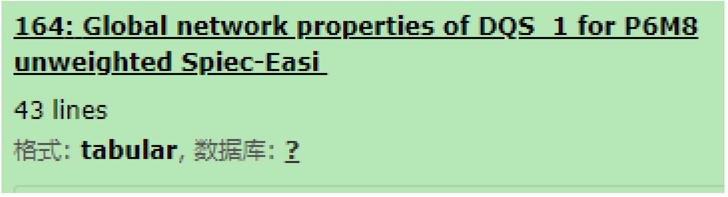
Global network properties of DQS_1 for P6M8 unweighted positive interaction Spiec-Easi

Parameters:

Weighted or Unweighted: unweighted

Weighted NODF: an index to indicate nestedness.

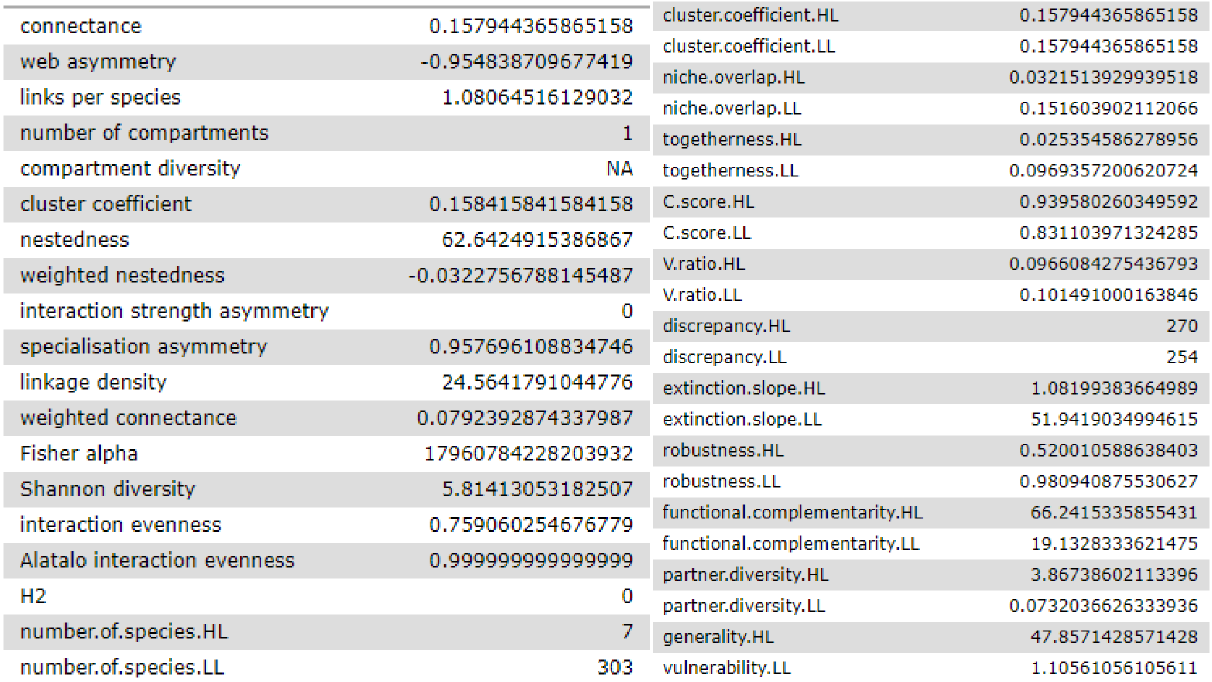
Outputs:



164: Global network properties of DQS_1 for P6M8 unweighted Spiec-Easi

43 lines

格式: **tabular**, 数据库: ?



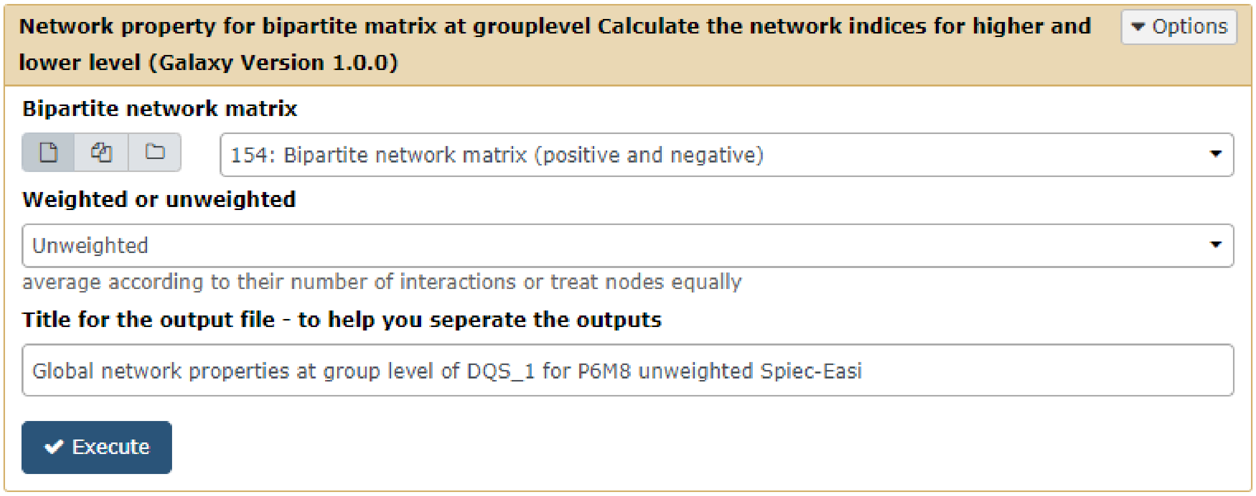
connectance	0.157944365865158
web asymmetry	-0.954838709677419
links per species	1.08064516129032
number of compartments	1
compartment diversity	NA
cluster coefficient	0.158415841584158
nestedness	62.6424915386867
weighted nestedness	-0.0322756788145487
interaction strength asymmetry	0
specialisation asymmetry	0.957696108834746
linkage density	24.5641791044776
weighted connectance	0.0792392874337987
Fisher alpha	17960784228203932
Shannon diversity	5.81413053182507
interaction evenness	0.759060254676779
Alatalo interaction evenness	0.999999999999999
H2	0
number.of.species.HL	7
number.of.species.LL	303

cluster.coefficient.HL	0.157944365865158
cluster.coefficient.LL	0.157944365865158
niche.overlap.HL	0.0321513929939518
niche.overlap.LL	0.151603902112066
togetherness.HL	0.025354586278956
togetherness.LL	0.0969357200620724
C.score.HL	0.939580260349592
C.score.LL	0.831103971324285
V.ratio.HL	0.0966084275436793
V.ratio.LL	0.101491000163846
discrepancy.HL	270
discrepancy.LL	254
extinction.slope.HL	1.08199383664989
extinction.slope.LL	51.9419034994615
robustness.HL	0.520010588638403
robustness.LL	0.980940875530627
functional.complementarity.HL	66.2415335855431
functional.complementarity.LL	19.1328333621475
partner.diversity.HL	3.86738602113396
partner.diversity.LL	0.073203662633936
generality.HL	47.8571428571428
vulnerability.LL	1.1056105610561

1.2 Network property for bipartite matrix at grouplevel

This step calculates the network properties for the two groups of species shown in the bipartite network matrix, e.g. higher lever for column species (plants in this pipeline), and lower level for row species (microbial data in this pipeline).

Inputs:



Network property for bipartite matrix at grouplevel Calculate the network indices for higher and lower level (Galaxy Version 1.0.0)

Bipartite network matrix

154: Bipartite network matrix (positive and negative)

Weighted or unweighted

Unweighted

average according to their number of interactions or treat nodes equally

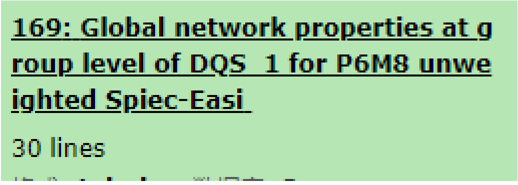
Title for the output file - to help you seperate the outputs

Global network properties at group level of DQS_1 for P6M8 unweighted Spiec-Easi

Parameters:

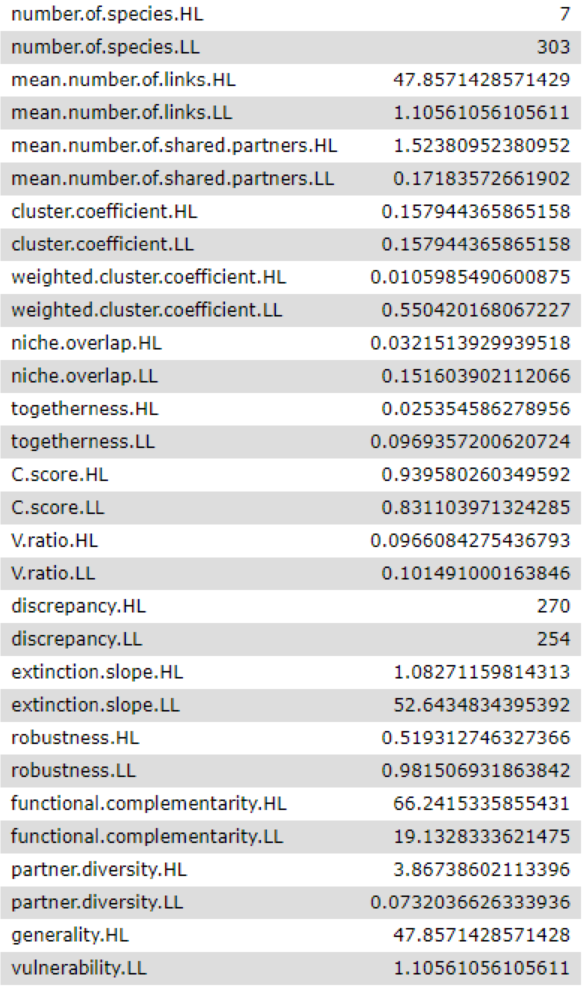
Weighted or Unweighted: unweighted

Outputs:



169: Global network properties at group level of DQS_1 for P6M8 unweighted Spiec-Easi

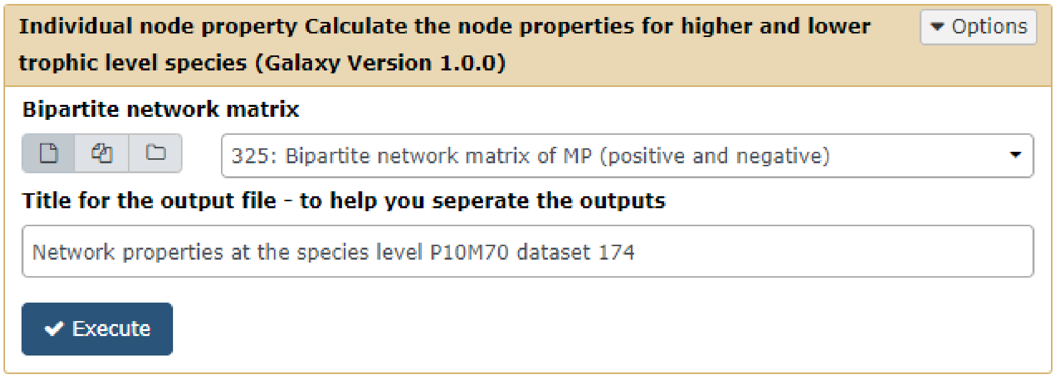
30 lines



number.of.species.HL	7
number.of.species.LL	303
mean.number.of.links.HL	47.8571428571429
mean.number.of.links.LL	1.10561056105611
mean.number.of.shared.partners.HL	1.52380952380952
mean.number.of.shared.partners.LL	0.17183572661902
cluster.coefficient.HL	0.157944365865158
cluster.coefficient.LL	0.157944365865158
weighted.cluster.coefficient.HL	0.0105985490600875
weighted.cluster.coefficient.LL	0.550420168067227
niche.overlap.HL	0.0321513929939518
niche.overlap.LL	0.151603902112066
togetherness.HL	0.025354586278956
togetherness.LL	0.0969357200620724
C.score.HL	0.939580260349592
C.score.LL	0.831103971324285
V.ratio.HL	0.0966084275436793
V.ratio.LL	0.101491000163846
discrepancy.HL	270
discrepancy.LL	254
extinction.slope.HL	1.08271159814313
extinction.slope.LL	52.6434834395392
robustness.HL	0.519312746327366
robustness.LL	0.981506931863842
functional.complementarity.HL	66.2415335855431
functional.complementarity.LL	19.1328333621475
partner.diversity.HL	3.86738602113396
partner.diversity.LL	0.0732036626333936
generality.HL	47.8571428571428
vulnerability.LL	1.10561056105611

1.3 Individual node property

Inputs:



Individual node property Calculate the node properties for higher and lower trophic level species (Galaxy Version 1.0.0)

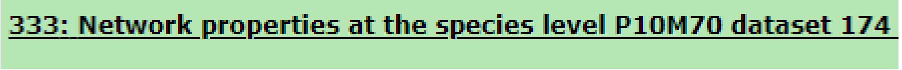
Bipartite network matrix

325: Bipartite network matrix of MP (positive and negative)

Title for the output file - to help you seperate the outputs

Network properties at the species level P10M70 dataset 174

Outputs:

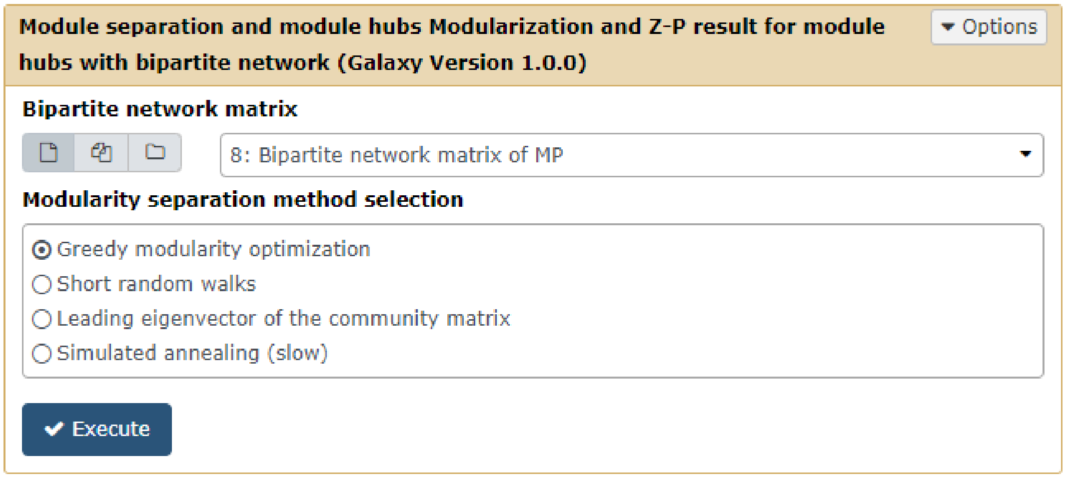


333: Network properties at the species level P10M70 dataset 174

1.4 Module separation and module hubs

There are four methods provided in this pipeline for modularization: greedy modularity optimization, short random walks and leading eigenvector of community matrix. Besides, the Z-P result can provide information of module hubs.

Inputs:



Module separation and module hubs Modularization and Z-P result for module hubs with bipartite network (Galaxy Version 1.0.0)

Bipartite network matrix

8: Bipartite network matrix of MP

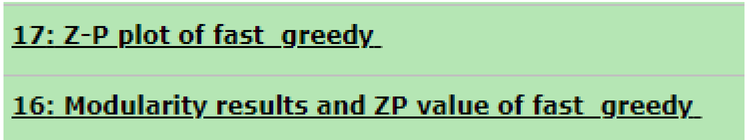
Modularity separation method selection

● Greedy modularity optimization
○ Short random walks
○ Leading eigenvector of the community matrix
○ Simulated annealing (slow)

Parameters:

Modularity method: Greedy modularity optimization / Short random walks / Leading eigenvector of the community matrix / Simulated annealing (slow). Select according to previous choice.

Outputs:



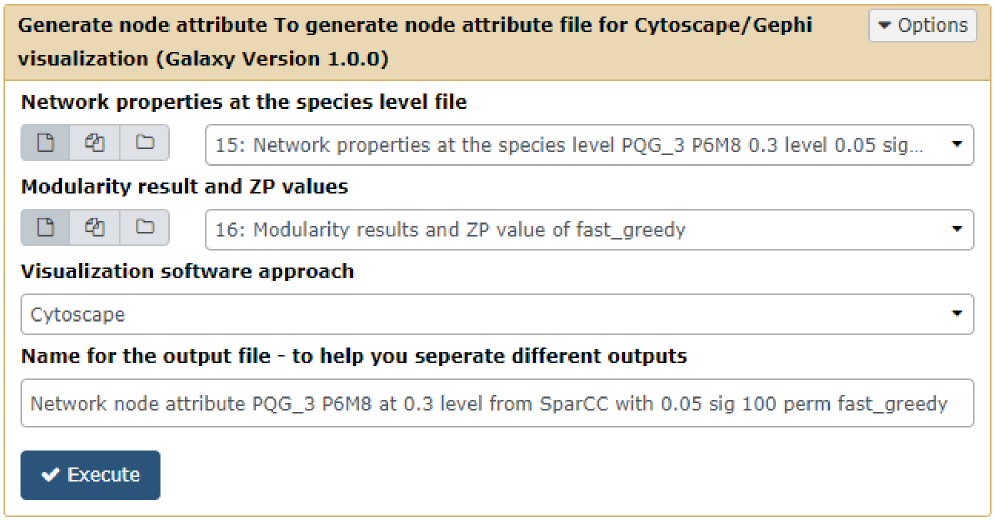
17: Z-P plot of fast_greedy_

16: Modularity results and ZP value of fast_greedy_

2. Bipartite visualization

2.1 Generate node attribute

Inputs:



Generate node attribute To generate node attribute file for Cytoscape/Gephi visualization (Galaxy Version 1.0.0)

Network properties at the species level file

15: Network properties at the species level PQG_3 P6M8 0.3 level 0.05 sig...

Modularity result and ZP values

16: Modularity results and ZP value of fast_greedy

Visualization software approach

Cytoscape

Name for the output file - to help you seperate different outputs

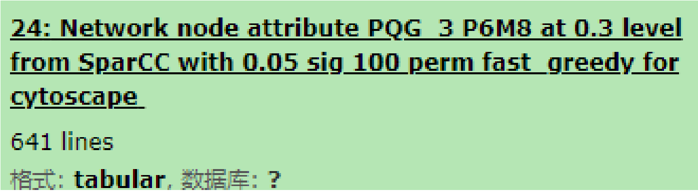
Network node attribute PQG_3 P6M8 at 0.3 level from SparCC with 0.05 sig 100 perm fast_greedy

Parameters options:

Visualization software approach: Cytoscape / Gephi

Outputs:

A tabular file for Cytoscape or a csv file for Gephi.



24: Network node attribute PQG_3 P6M8 at 0.3 level from SparCC with 0.05 sig 100 perm fast_greedy for cytoscape

641 lines

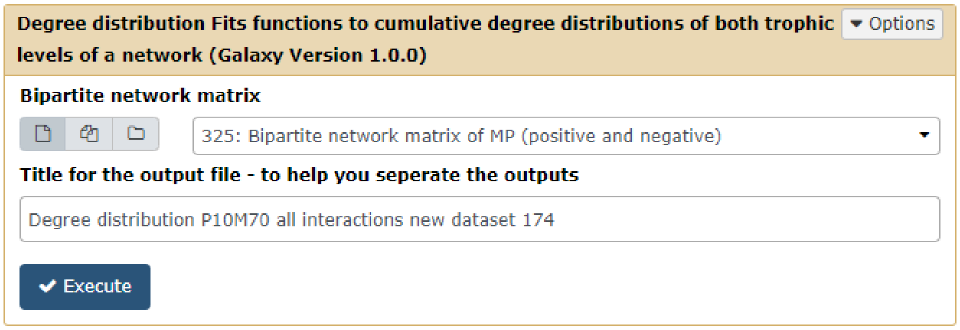
格式: **tabular**, 数据库: **?**

3. Bipartite statistics

3.1 Degree distribution

Fits functions to cumulative degree distributions of both trophic levels of a network. This program is mainly for cumulative distribution.

Inputs:



Degree distribution Fits functions to cumulative degree distributions of both trophic levels of a network (Galaxy Version 1.0.0)

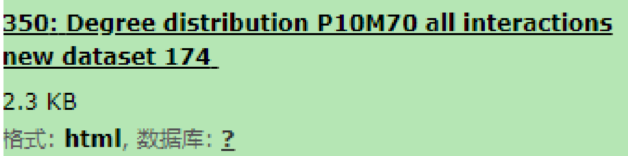
Bipartite network matrix

325: Bipartite network matrix of MP (positive and negative)

Title for the output file - to help you seperate the outputs

Degree distribution P10M70 all interactions new dataset 174

Outputs:



350: Degree distribution P10M70 all interactions new dataset 174

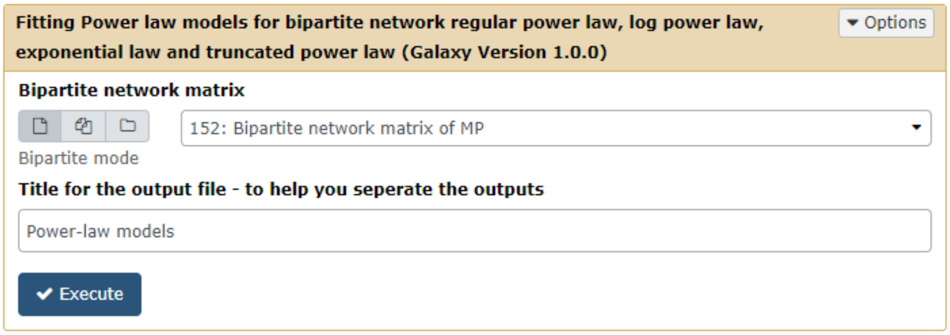
2.3 KB

格式: **html**, 数据库: **?**

3.2 Fitting Power law models for bipartite network

Fitting Power law models for bipartite network regular power law, log power law, exponential law and truncated power law. This fitting is mainly to use node connectivity or node degree as response.

Inputs:



Fitting Power law models for bipartite network regular power law, log power law, exponential law and truncated power law (Galaxy Version 1.0.0)

Bipartite network matrix

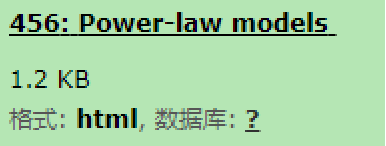
152: Bipartite network matrix of MP

Bipartite mode

Title for the output file - to help you seperate the outputs

Power-law models

Outputs:



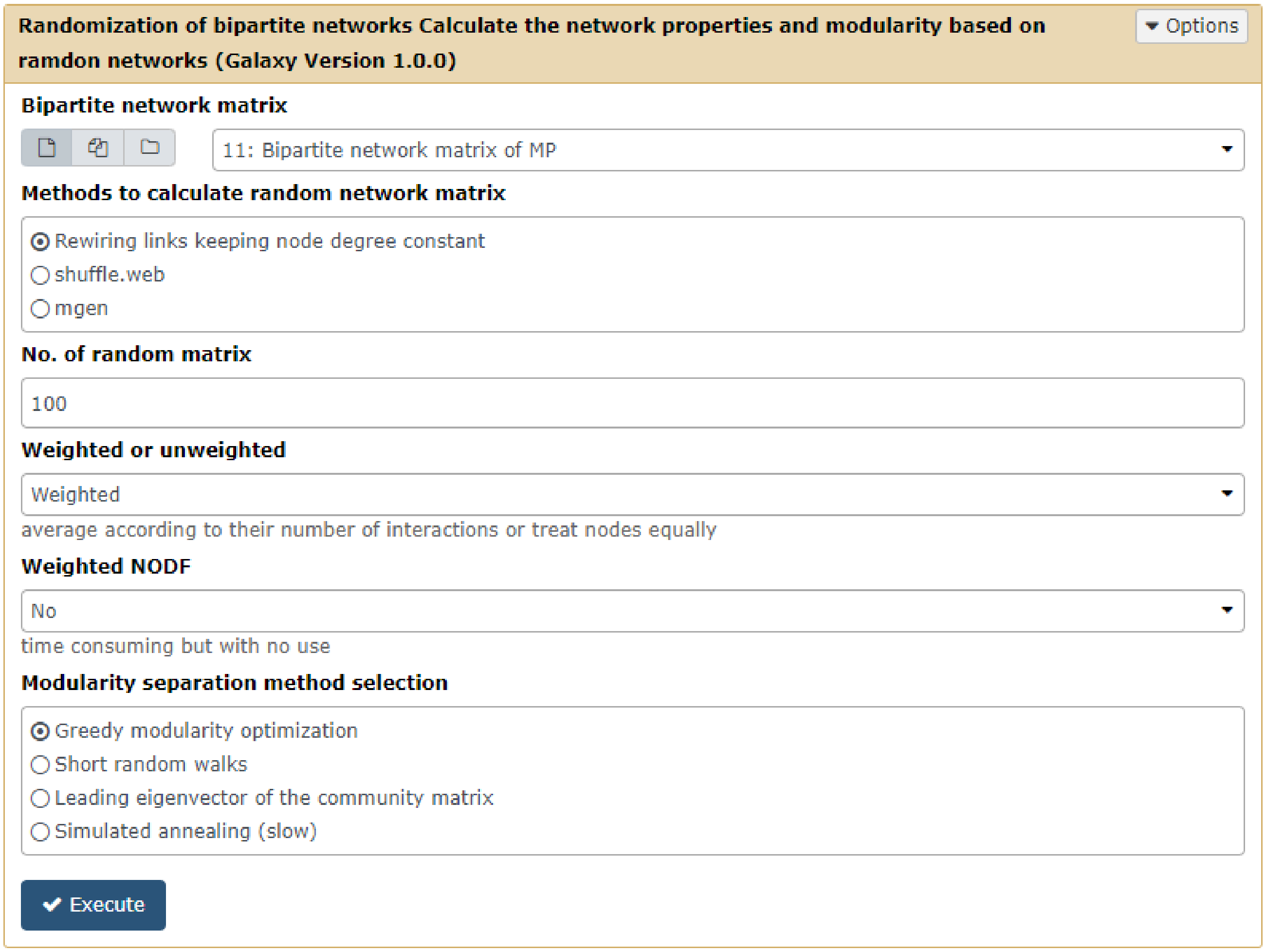
456: Power-law models

1.2 KB

格式: **html**, 数据库: ?

3.3 Randomization of bipartite networks

Inputs:



Randomization of bipartite networks Calculate the network properties and modularity based on ramdon networks (Galaxy Version 1.0.0)

Bipartite network matrix

11: Bipartite network matrix of MP

Methods to calculate random network matrix

● Rewiring links keeping node degree constant
○ shuffle.web
○ mgen

No. of random matrix

100

Weighted or unweighted

Weighted

average according to their number of interactions or treat nodes equally

Weighted NODF

No

time consuming but with no use

Modularity separation method selection

● Greedy modularity optimization
○ Short random walks
○ Leading eigenvector of the community matrix
○ Simulated annealing (slow)

Parameter options:

Methods:

- **Rewiring links keeping node degree constant:** rewiring the links between the randomly selected two links.
- **Shuffle.web:** It implements a method where matrix is first filled honouring row and column totals, but with integers that may be larger than one. Then the method inspects random 2x2 matrices and performs a quasiswap on them. It is similar to ordinary swap, but it also can reduce numbers above one to ones maintaining marginal totals.

- **Mgen:** This is a generic function to build null models for mutualistic networks, used by Vázquez et al. (2009). It is general in the sense that it allows any type of probability matrix to be used for constructing the simulated matrices. It does not, however, constrain rown and column totals, nor does it constrain connectance.

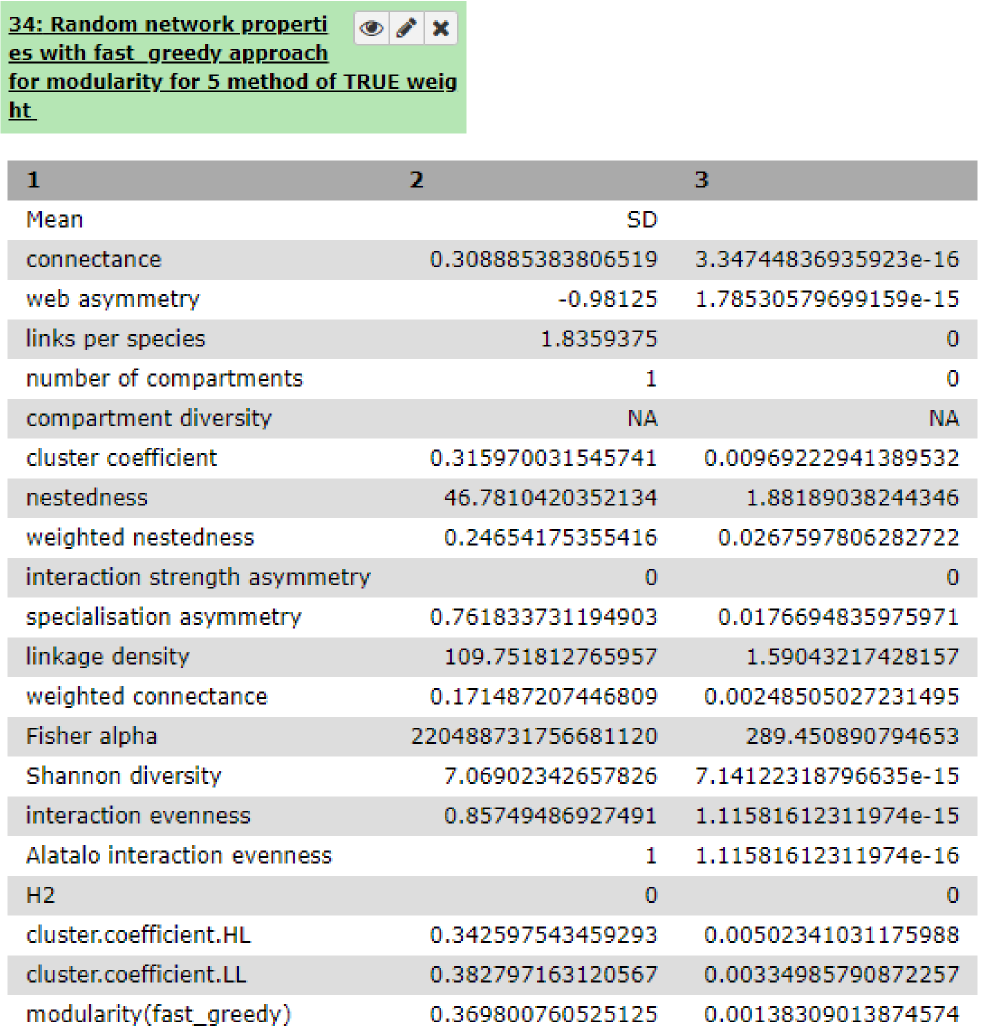
No. of random matrix: 100

Weighted or unweighted: Weighted. Average according to their number of interactions or treat nodes equally

Weighted NODF: No

Modularity separation method selection: Greedy modularity optimization / Short random walks / Leading eigenvector of the community matrix / Simulated annealing (slow). Select according to previous choice.

Outputs:



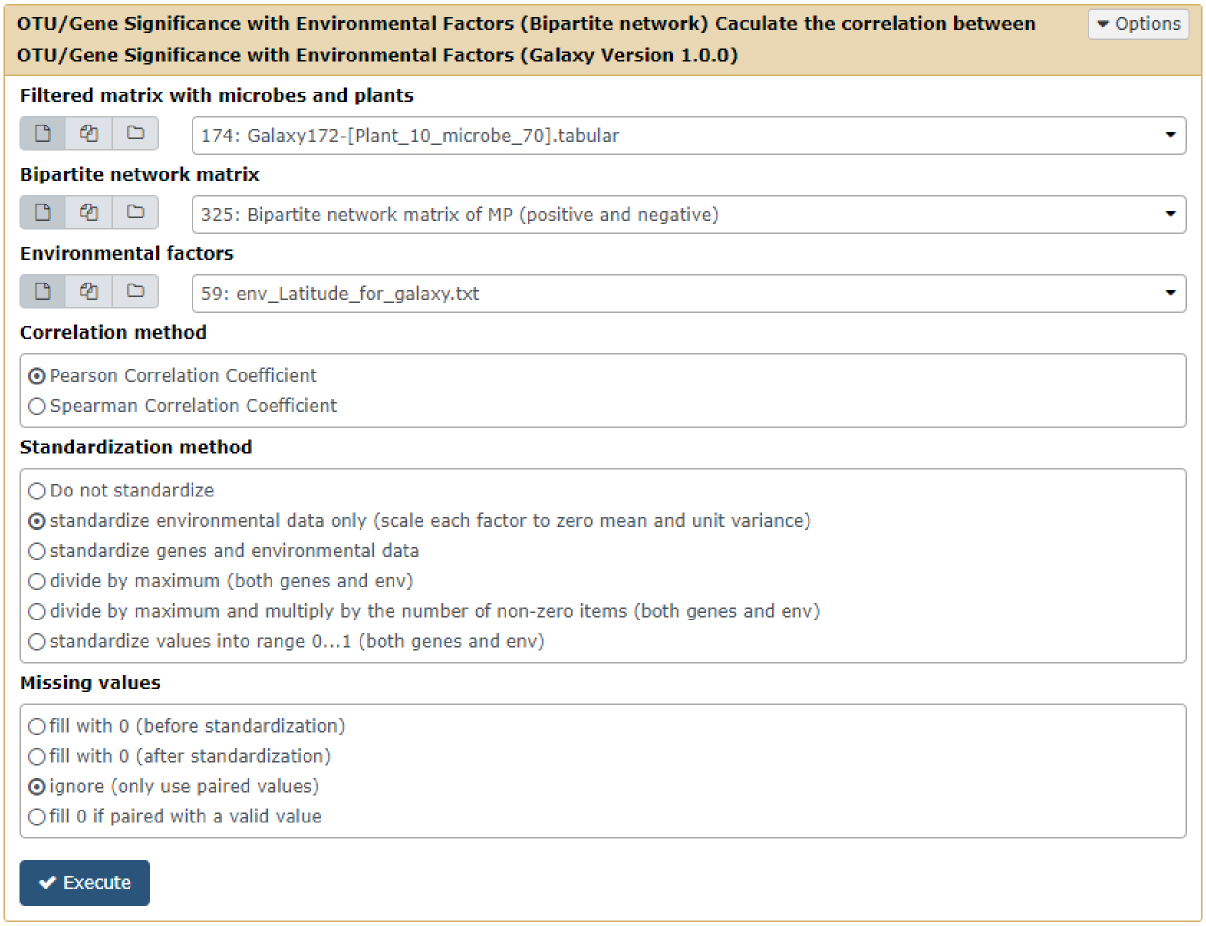
34: Random network properties with fast_greedy approach for modularity for 5 method of TRUE weight

1	2	3
Mean	SD	
connectance	0.308885383806519	3.34744836935923e-16
web asymmetry	-0.98125	1.78530579699159e-15
links per species	1.8359375	0
number of compartments	1	0
compartment diversity	NA	NA
cluster coefficient	0.315970031545741	0.00969222941389532
nestedness	46.7810420352134	1.88189038244346
weighted nestedness	0.24654175355416	0.0267597806282722
interaction strength asymmetry	0	0
specialisation asymmetry	0.761833731194903	0.0176694835975971
linkage density	109.751812765957	1.59043217428157
weighted connectance	0.171487207446809	0.00248505027231495
Fisher alpha	220488731756681120	289.450890794653
Shannon diversity	7.06902342657826	7.14122318796635e-15
interaction evenness	0.85749486927491	1.11581612311974e-15
Alatalo interaction evenness	1	1.11581612311974e-16
H2	0	0
cluster.coefficient.HL	0.342597543459293	0.00502341031175988
cluster.coefficient.LL	0.382797163120567	0.00334985790872257
modularity(fast_greedy)	0.369800760525125	0.00138309013874574

3.4 OTU/Gene Significance with Environmental Factors (Bipartite network)

The output of this program can be used for significance test using Mantel test for further analysis.

Inputs:



OTU/Gene Significance with Environmental Factors (Bipartite network) Caculate the correlation between OTU/Gene Significance with Environmental Factors (Galaxy Version 1.0.0)

Filtered matrix with microbes and plants

174: Galaxy172-[Plant_10_microbe_70].tabular

Bipartite network matrix

325: Bipartite network matrix of MP (positive and negative)

Environmental factors

59: env_Latitude_for_galaxy.txt

Correlation method

- ● Pearson Correlation Coefficient
- ○ Spearman Correlation Coefficient

Standardization method

- ○ Do not standardize
- ● standardize environmental data only (scale each factor to zero mean and unit variance)
- ○ standardize genes and environmental data
- ○ divide by maximum (both genes and env)
- ○ divide by maximum and multiply by the number of non-zero items (both genes and env)
- ○ standardize values into range 0...1 (both genes and env)

Missing values

- ○ fill with 0 (before standardization)
- ○ fill with 0 (after standardization)
- ● ignore (only use paired values)
- ○ fill 0 if paired with a valid value

Parameter options:

Filtered matrix with microbes and plants

Bipartite network matrix

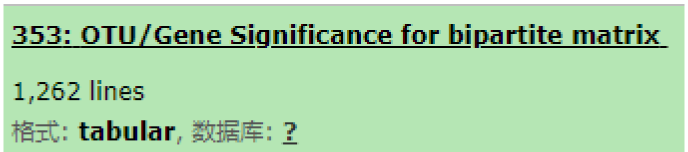
Environmental factors

Correlation method: Pearson Correlation Coefficient / Spearman Correlation Coefficient

Standardization method: standardize environmental data only (scale each factor to zero mean and unit variance) or other choice

Missing values: ignore (only use paired values)

Outputs:



353: OTU/Gene Significance for bipartite matrix

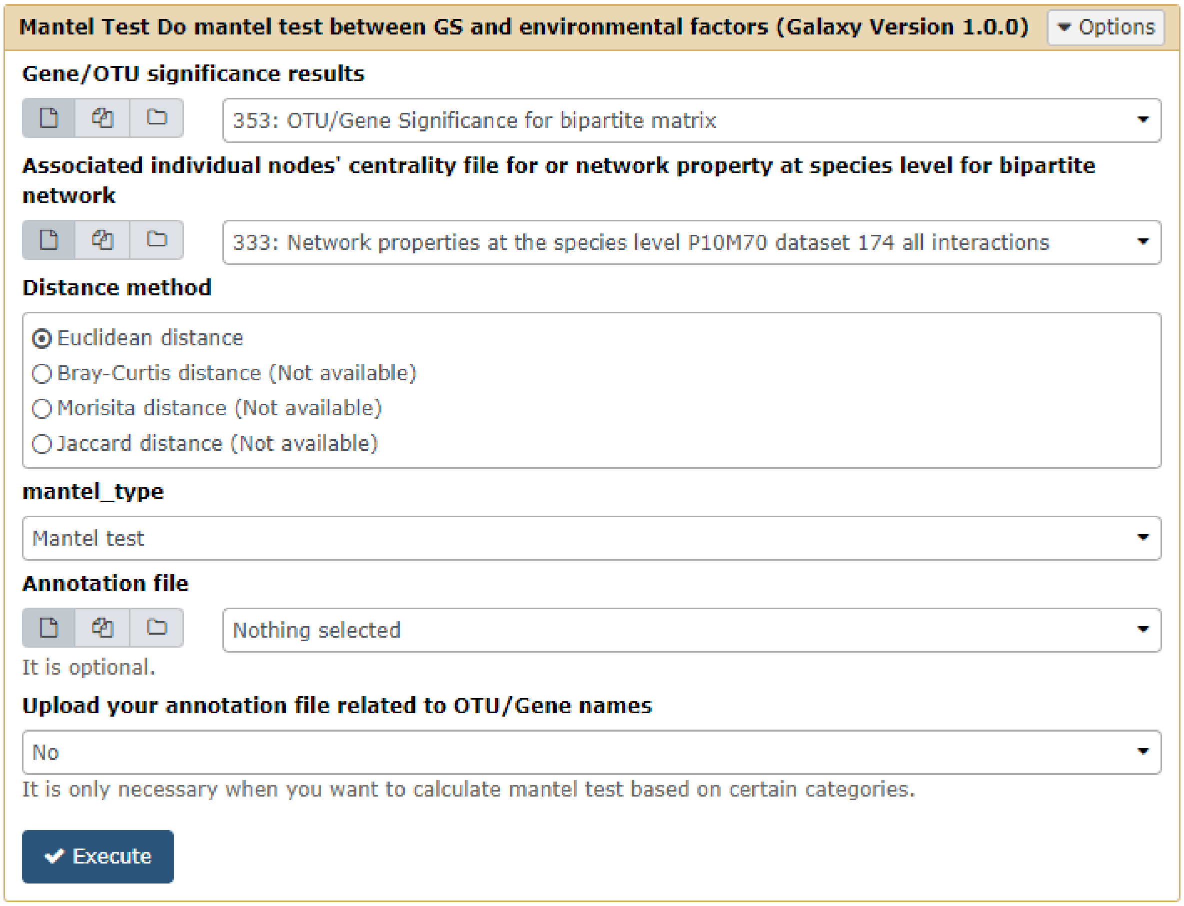
1,262 lines

格式: **tabular**, 数据库: ?

3.5 Mantel Test

For mantel test:

Inputs:



Mantel Test Do mantel test between GS and environmental factors (Galaxy Version 1.0.0)

Gene/OTU significance results

353: OTU/Gene Significance for bipartite matrix

Associated individual nodes' centrality file for or network property at species level for bipartite network

333: Network properties at the species level P10M70 dataset 174 all interactions

Distance method

● Euclidean distance
○ Bray-Curtis distance (Not available)
○ Morisita distance (Not available)
○ Jaccard distance (Not available)

mantel_type

Mantel test

Annotation file

Nothing selected

It is optional.

Upload your annotation file related to OTU/Gene names

No

It is only necessary when you want to calculate mantel test based on certain categories.

Parameter options:

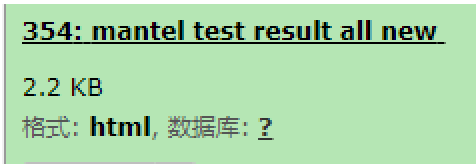
Select the corresponding files according to the labels shown in the pipeline.

Mantel type: Mantel test or Partial mantel test

Annotation file: (Optional)

Upload your annotation file related to OTU/Gene names: No / Yes (if you have uploaded)

Outputs:



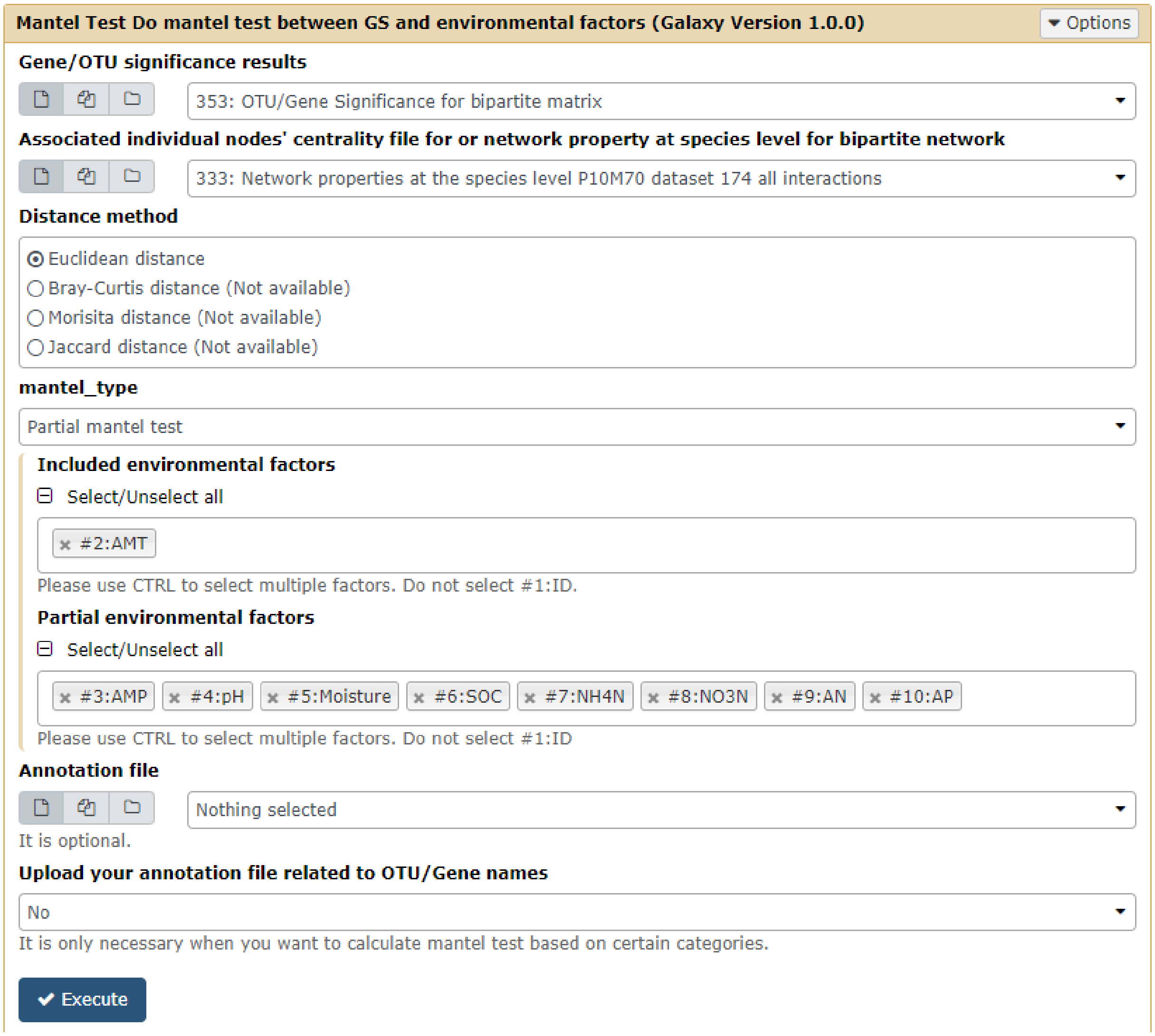
354: mantel test result all new

2.2 KB

格式: **html**, 数据库: ?

For partial mantel test:

Inputs:



Mantel Test Do mantel test between GS and environmental factors (Galaxy Version 1.0.0)

Gene/OTU significance results

353: OTU/Gene Significance for bipartite matrix

Associated individual nodes' centrality file for or network property at species level for bipartite network

333: Network properties at the species level P10M70 dataset 174 all interactions

Distance method

● Euclidean distance
○ Bray-Curtis distance (Not available)
○ Morisita distance (Not available)
○ Jaccard distance (Not available)

mantel_type

Partial mantel test

Included environmental factors

⊟ Select/Unselect all

× #2:AMT

Please use CTRL to select multiple factors. Do not select #1:ID.

Partial environmental factors

⊟ Select/Unselect all

× #3:AMP × #4:pH × #5:Moisture × #6:SOC × #7:NH4N × #8:NO3N × #9:AN × #10:AP

Please use CTRL to select multiple factors. Do not select #1:ID

Annotation file

Nothing selected

It is optional.

Upload your annotation file related to OTU/Gene names

No

It is only necessary when you want to calculate mantel test based on certain categories.

Parameter options:

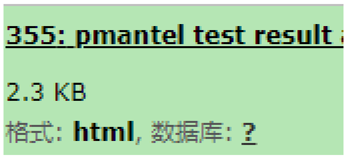
Select the corresponding files according to the labels shown in the pipeline.

Mantel type: Mantel test or Partial mantel test

Annotation file: (Optional)

Upload your annotation file related to OTU/Gene names: No / Yes (if you have uploaded)

Outputs:



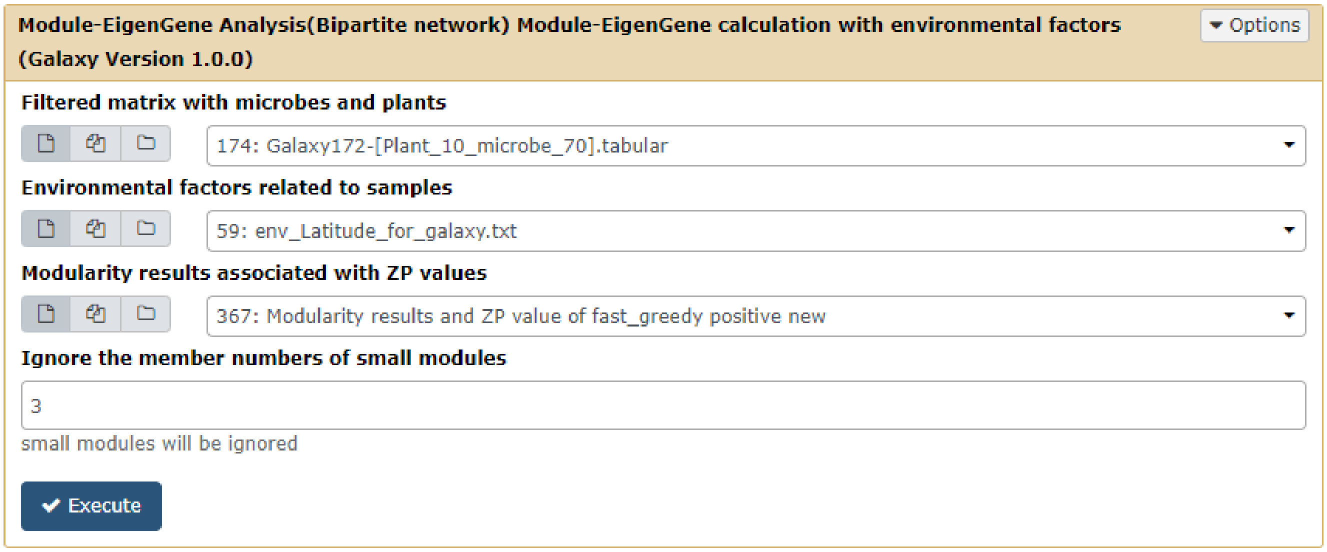
355: pmantel test result

2.3 KB

格式: **html**, 数据库: ?

3.6 Module-EigenGene Analysis (Bipartite network)

Inputs:



Module-EigenGene Analysis(Bipartite network) Module-EigenGene calculation with environmental factors (Galaxy Version 1.0.0)

Filtered matrix with microbes and plants

174: Galaxy172-[Plant_10_microbe_70].tabular

Environmental factors related to samples

59: env_Latitude_for_galaxy.txt

Modularity results associated with ZP values

367: Modularity results and ZP value of fast_greedy positive new

Ignore the member numbers of small modules

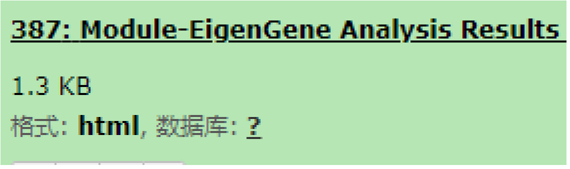
3

small modules will be ignored

Parameter options:

Select the corresponding files according to the labels shown in the pipeline.

Outputs:



387: Module-EigenGene Analysis Results

1.3 KB

格式: **html**, 数据库: ?

C. Network Construction (MENAP)

For the molecular ecological network analyses pipeline (MENAP, Figure 2), due to the adjacent network consisting of a single group, the parameters during network construction should shift to “one group of species” in “majority selection” step. The following steps should take care of this accordingly.

Please see details above.

D. Network Analyses (MENAP)

After network generation, there is one adjacent matrix to show the potential interactions between all species with presence/absence data and the computations of further topological features.

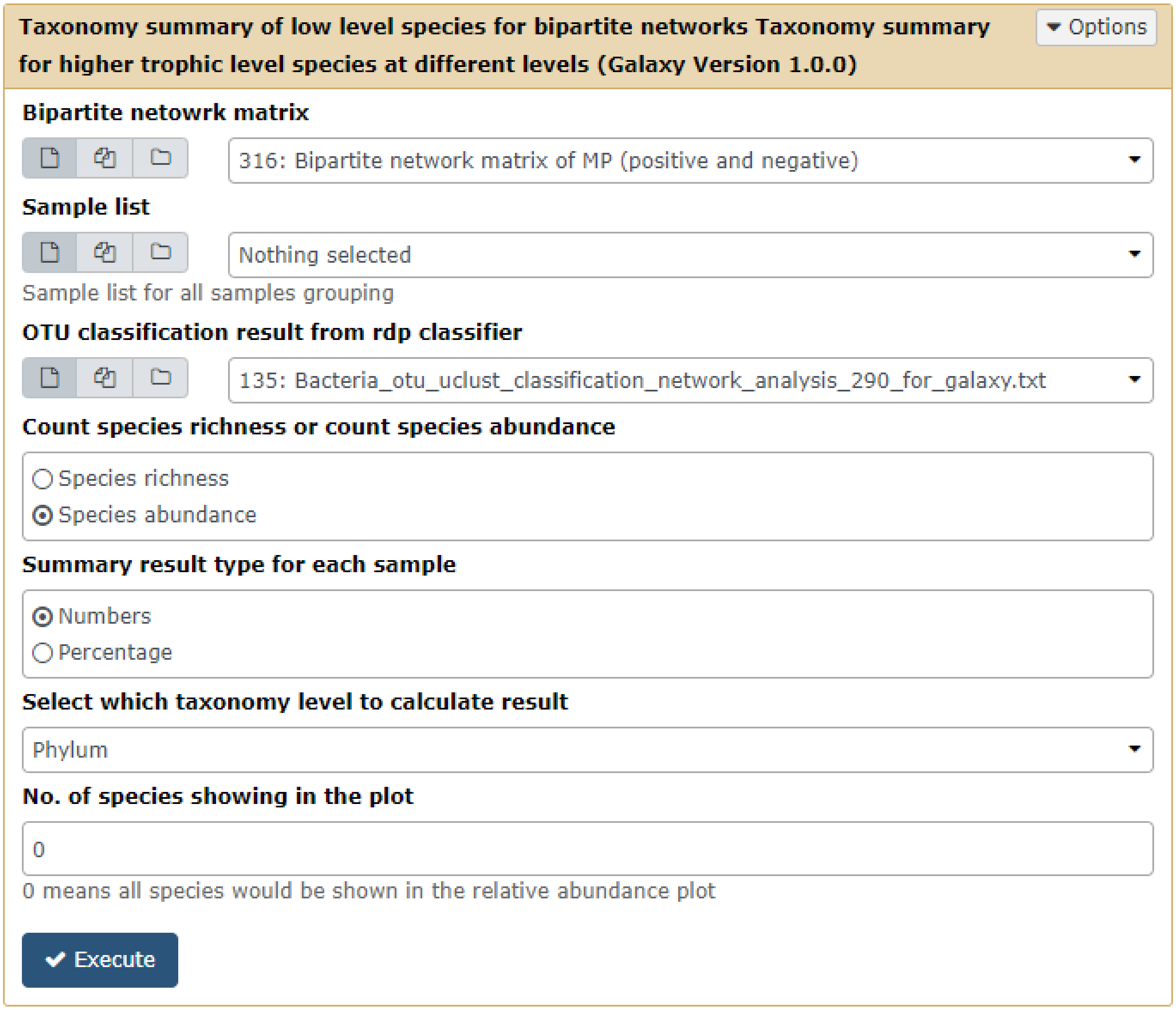
Please refer to above details and import the correct adjacent files within the parameters.

E. Auxiliary tools in miscellaneous section

1. Taxonomy summary of low level species for bipartite networks

This tool is mainly used to assign different OTUs or species into higher taxonomic level, e.g. phylum and genus, and then generate sub-graph matrix for each group at the specific level.

Inputs:



Taxonomy summary of low level species for bipartite networks Taxonomy summary for higher trophic level species at different levels (Galaxy Version 1.0.0)

Bipartite netowrk matrix

316: Bipartite network matrix of MP (positive and negative)

Sample list

Nothing selected

Sample list for all samples grouping

OTU classification result from rdp classifier

135: Bacteria_otu_uclust_classification_network_analysis_290_for_galaxy.txt

Count species richness or count species abundance

○ Species richness
● Species abundance

Summary result type for each sample

● Numbers
○ Percentage

Select which taxonomy level to calculate result

Phylum

No. of species showing in the plot

0

0 means all species would be shown in the relative abundance plot

Parameters:

Bipartite network matrix: matrix of bipartite graph

Sample list: Not useful at this stage

OTU classification result from RDP classifier: OTU/Gene/Species classification file

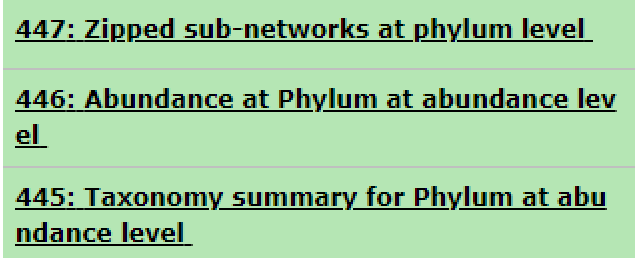
Count species richness or count species abundance: Species abundance / Species richness

Summary result type for each sample: Numbers / Percentage

Select which taxonomy level to calculate result: Phylum (select from classification file)

No. of species showing in the plot: 0

Output:



447: Zipped sub-networks at phylum level

446: Abundance at Phylum at abundance level

445: Taxonomy summary for Phylum at abundance level

One file contains the bipartite network matrix at the specific taxonomic level and another plot is the summary for this level.

After download the zipped file to local directory, you need to unzip this file twice. For the first step of unzipping process, you can easily unzip it. For the second step of unzipping process, you need to rename the extension file type to ".zip" or ".gz" and thereafter you could to unzip this file. After the two steps of unzipping, you can see the separated txt files.

2. Convert sif file to bipartite matrix

This step aims to generate a bipartite matrix from a sif file exported from Cytoscape software. The two groups of species can be separated according to the provided unique label, e.g. "B_OTU" and "P_OTU". The output can be used for bipartite network matrix analysis.

Input:

Convert sif file to bipartite matrix (Galaxy Version 1.0.0)

Sif file from Cytoscape or own generated file containing three columns

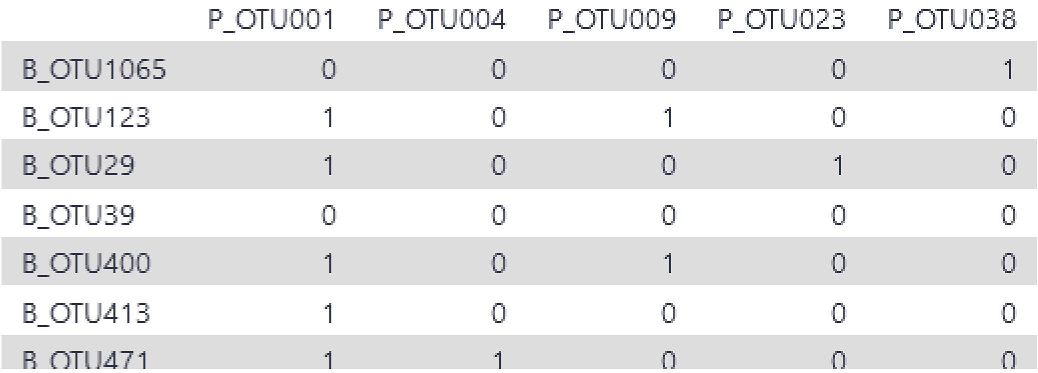
496: bac_prot_full_loam_kai_20200122 0.710 bipartite.sif

Unique label for lower trophic species, e.g. bacteria with 'B_'

B_

Please change this accordingly.

Output:

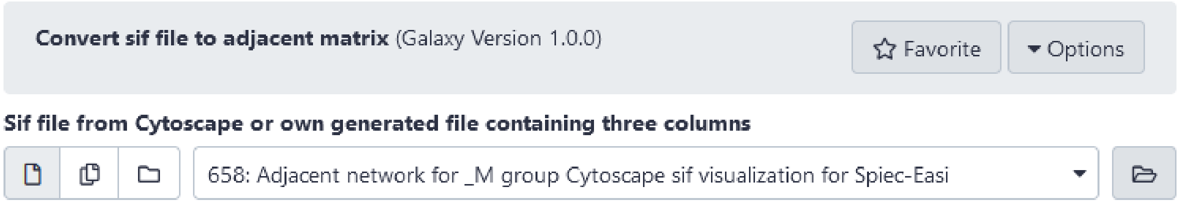


	P_OTU001	P_OTU004	P_OTU009	P_OTU023	P_OTU038
B_OTU1065	0	0	0	0	1
B_OTU123	1	0	1	0	0
B_OTU29	1	0	0	1	0
B_OTU39	0	0	0	0	0
B_OTU400	1	0	1	0	0
B_OTU413	1	0	0	0	0
B_OTU471	1	1	0	0	0

3. Convert sif file to adjacent matrix

Any exported sif file from Cytoscape software can be imported here and converted to an adjacent network matrix for network analysis

Input:

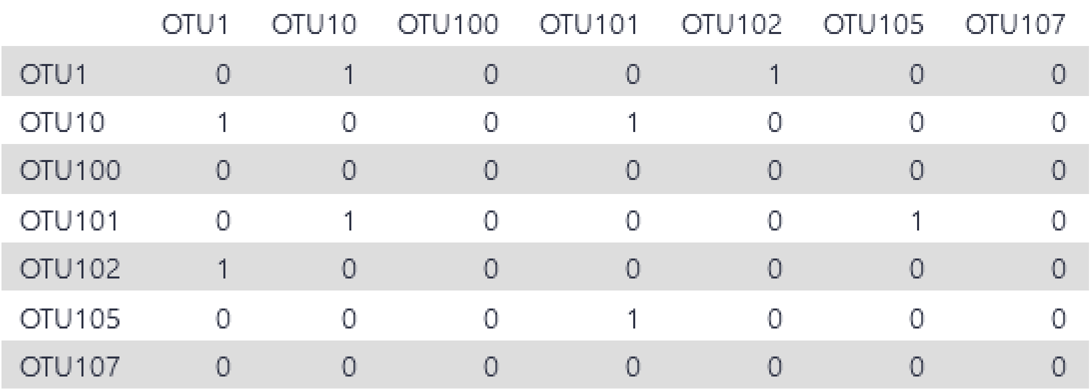


Convert sif file to adjacent matrix (Galaxy Version 1.0.0)

Sif file from Cytoscape or own generated file containing three columns

658: Adjacent network for _M group Cytoscape sif visualization for Spiec-Easi

Output:

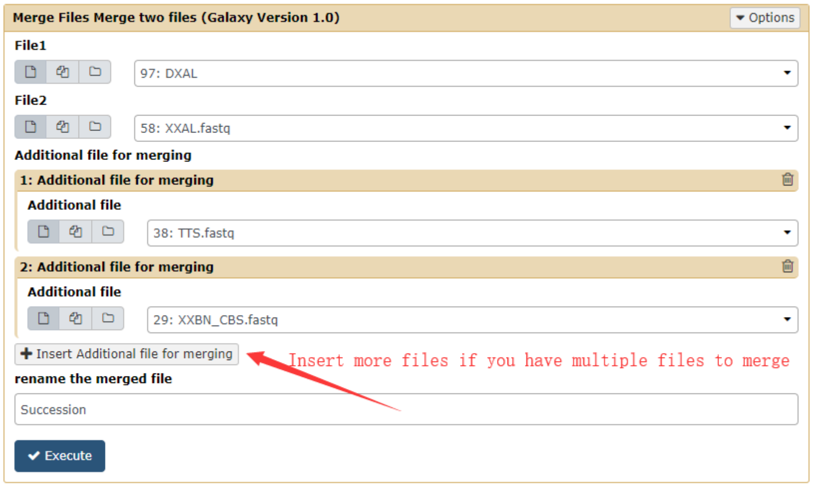


	OTU1	OTU10	OTU100	OTU101	OTU102	OTU105	OTU107
OTU1	0	1	0	0	1	0	0
OTU10	1	0	0	1	0	0	0
OTU100	0	0	0	0	0	0	0
OTU101	0	1	0	0	0	1	0
OTU102	1	0	0	0	0	0	0
OTU105	0	0	0	1	0	0	0
OTU107	0	0	0	0	0	0	0

4. Merge files

This tool is mainly used to merge multiple files into one file.

Input:



Merge Files Merge two files (Galaxy Version 1.0)

File1

97: DXAL

File2

58: XXAL.fastq

Additional file for merging

1: Additional file for merging

Additional file

38: TTS.fastq

2: Additional file for merging

Additional file

29: XXBN_CBS.fastq

Insert more files if you have multiple files to merge

rename the merged file

Succession

Output:

- Succession (If you put another name in the “rename the merged file”, it will show what you have fill in.)

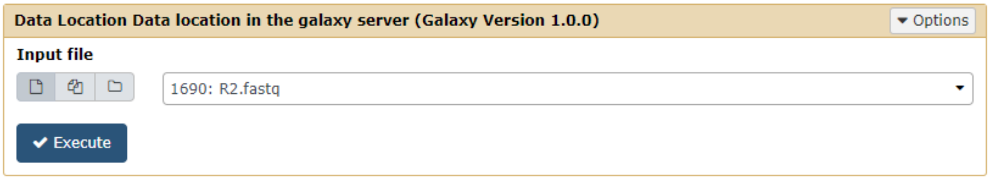
5. File Compression

Compress file into smaller size.

6. Data location

This tool is mainly used to find the data location for certain dataset in the server. The data location is helpful to find the dataset for Galaxy administrators when you have problems.

Input:



Data Location Data location in the galaxy server (Galaxy Version 1.0.0)

Input file: 1690: R2.fastq

Output:

- **file_location.txt**

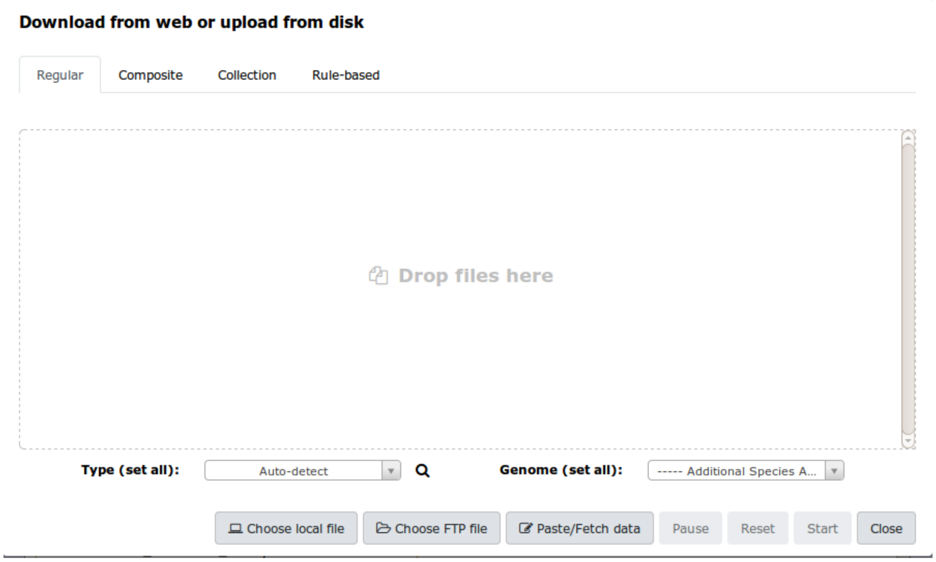
```
/newdatabase1/galaxy/user_data/datasets/000/105/dataset_105483.dat
```

F. Operation tricks and common problem solutions

1. Upload

Upload the all the OTU tables or environmental variable datasets to selected history.

Upload button（multiple files are available）



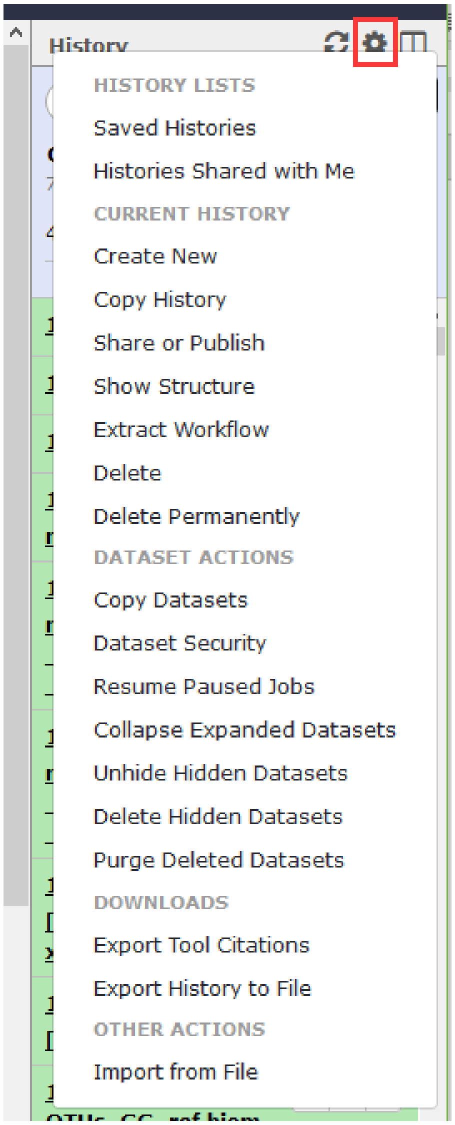
Type (set all): Auto-detect

Genome (set all): ----- Additional Species A...

Required files:

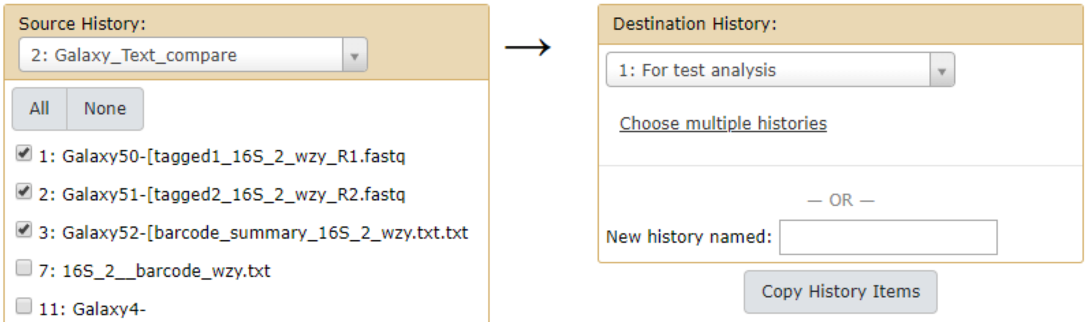
You can find following test data from the “shared library/test data” directory and import these three files there.

2. Basic operations in Galaxy



Please remember to choose “choose permanently” if you want to erase your history permanently, otherwise it will store into a temporary place and your quota will not decrease. See the below introduction for how to find temporarily deleted history.

Copy datasets:



Source History: 2: Galaxy_Text_compare

☑ 1: Galaxy50-[tagged1_16S_2_wzy_R1.fastq
☑ 2: Galaxy51-[tagged2_16S_2_wzy_R2.fastq
☑ 3: Galaxy52-[barcode_summary_16S_2_wzy.txt.txt
☐ 7: 16S_2__barcode_wzy.txt
☐ 11: Galaxy4-

→

Destination History: 1: For test analysis

3. Dataset deletions

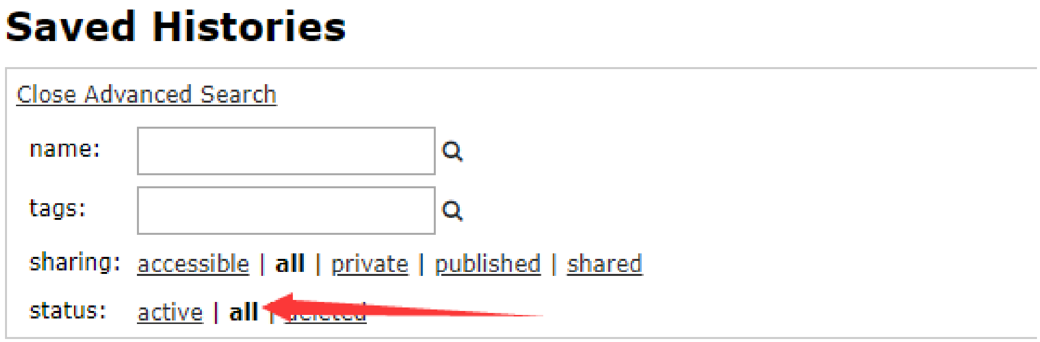
Select “saved history” and further choose “Advanced Search” button:

Saved Histories

search history names and tags

Advanced Search

Choose “all” button to show all history that you have created in your account. And select the deleted history to further erase or retrieve.



Saved Histories

Close Advanced Search

name:

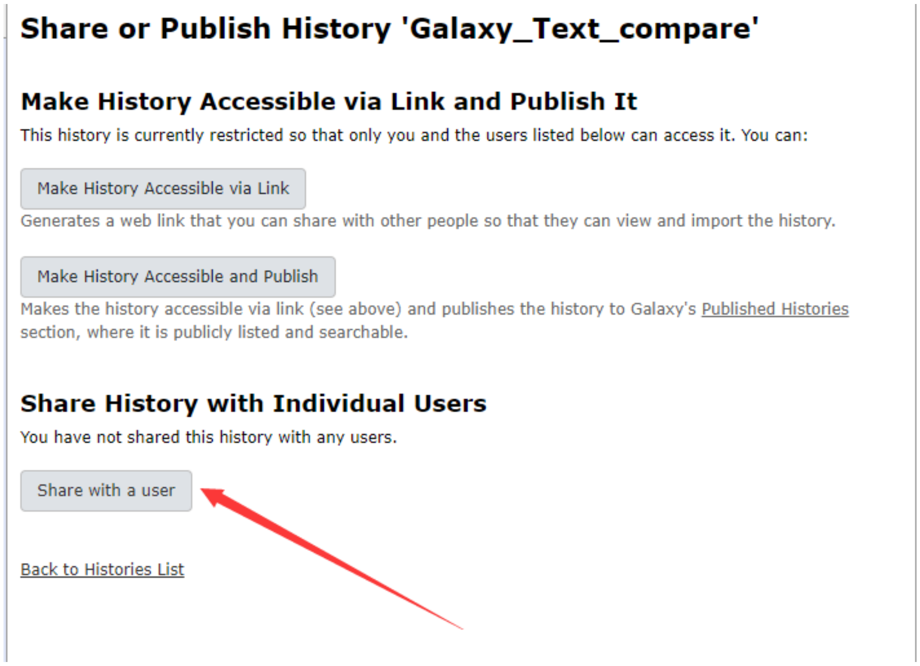
tags:

sharing: accessible | **all** | private | published | shared

status: active | **all** | deleted

4. Share historys to other users

Select “share or publish” of a certain history, then fill in the individual users:



Share or Publish History 'Galaxy_Text_compare'

Make History Accessible via Link and Publish It

This history is currently restricted so that only you and the users listed below can access it. You can:

Make History Accessible via Link

Generates a web link that you can share with other people so that they can view and import the history.

Make History Accessible and Publish

Makes the history accessible via link (see above) and publishes the history to Galaxy's Published Histories section, where it is publicly listed and searchable.

Share History with Individual Users

You have not shared this history with any users.

Share with a user

Back to Histories List