GEMNA
Guide User

Version 1.0.2

win7

November 5, 2024

Contents

1 Introduction

The guide is divided into two parts; the first part (Sec. 2) deals with processing and analyzing mass spectrometry data, and the second (Sec. 3) deals with visualizing and finding patterns in the results. For this guide, we use the Mutant dataset containing three phenotypes: WT, pck1, and zwf1.

The source code was divided in backend y fronted. The backend was implemented in Django rest framework, with PyTorch Geometric, PyOD libraries, and it is available at `https://github.com/win7/GEMNA_Backend.git`. On the other hand, the fronted was implemented in Vue.js with Nuxt framework, and it is available at `https://github.com/win7/GEMNA_Frontend.git`.

2 Process and analysis

2.1 Format input

The format of the input file must be in the *.csv* format, and the "|" character delimits the columns. The first four columns are the metadata, and the remaining are the intensity measurements. An example is shown in Figure 1.

	Alignment ID	Average Rt	Average Mz	Metabolite name	WT_3.3.3	WT_5.1.9	WT_5.2.4	WT_5.2.8	WT_3.1.5	WT_3.4.4	...	pck1_1.1.6	pck1_2.2.9	pck1_2.1.2	pck1_1.4.3	zwf1_1.2.7	zwf1_3.2.5	zwf1_3.1.2	zwf1_3.3.6	zwf1_1.4.2
0	1	0	59.0049	Unknown	169	286	575	340	939	410	...	913	640	1078	639	466	502	184	1008	851
1	2	0	59.0137	Unknown	48964	60211	195240	148489	81334	54320	...	69318	162356	135568	118566	51114	84501	66010	99974	129520
2	3	0	59.0291	Unknown	1553	2288	7911	5562	4064	2006	...	3124	6693	5744	4682	2067	3571	2798	4384	5416
3	4	0	59.0370	Unknown	1	257	1262	1012	1247	186	...	616	1549	1205	961	553	913	448	902	1195
4	5	0	59.0453	Unknown	1	112	321	1	634	65	...	217	624	699	399	234	167	212	330	398
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
6240	6241	0	996.5509	Unknown	1745	1280	2179	2526	1805	2533	...	2536	3262	2728	2512	2103	2031	2844	2479	2639
6241	6242	0	996.7096	Unknown	1593	361	979	934	1766	1758	...	2092	1507	2126	2590	2188	3212	3354	4749	2695
6242	6243	0	997.5542	Unknown	1724	982	2727	2481	2318	2253	...	2729	1543	2266	2259	2063	2206	3114	2799	2499
6243	6244	0	997.7131	Unknown	1490	711	968	691	1319	1660	...	1719	216	1971	1449	863	1623	2412	2773	1302
6244	6245	0	998.4845	Unknown	2412	856	2737	2287	1875	0	...	3187	3000	3067	2161	1104	2486	2707	1690	2194

Figure 1: Aligned mass spectrometry data format.

2.2 Process mass spectrometry data

2.2.1 Load your data

The form (see Figure 2) requires an e-mail address, to which you will receive a code to be used later in the visualization section.

2.2.2 Parameters selection

The form (see Figure 3) requires the choice of the following parameters: the method to generate the node embeddings (the best `LVGAE`), network

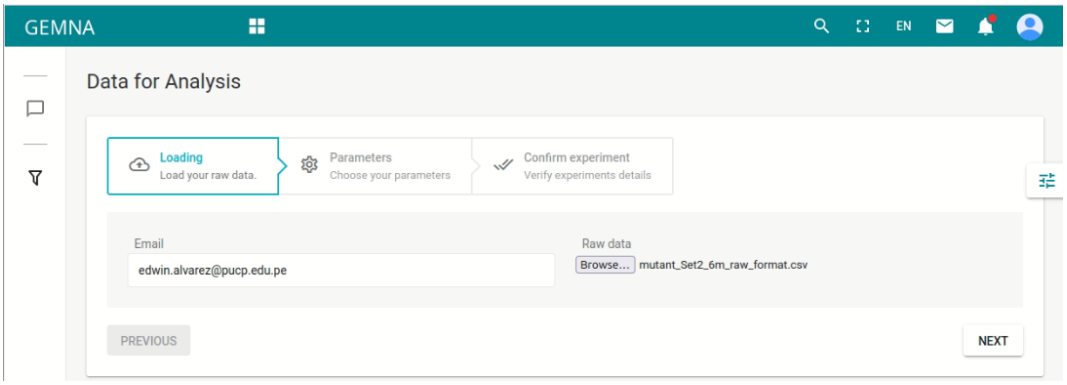

Figure 2: Load your data.

variation (the best many-to-many), dimension of the embeddings (the best 3), control (for this dataset is WT), the data has no transformation, the correlation threshold to generate the networks (default 0.5), and the threshold to filter the significant changes between two metabolomic networks (default 0.05).

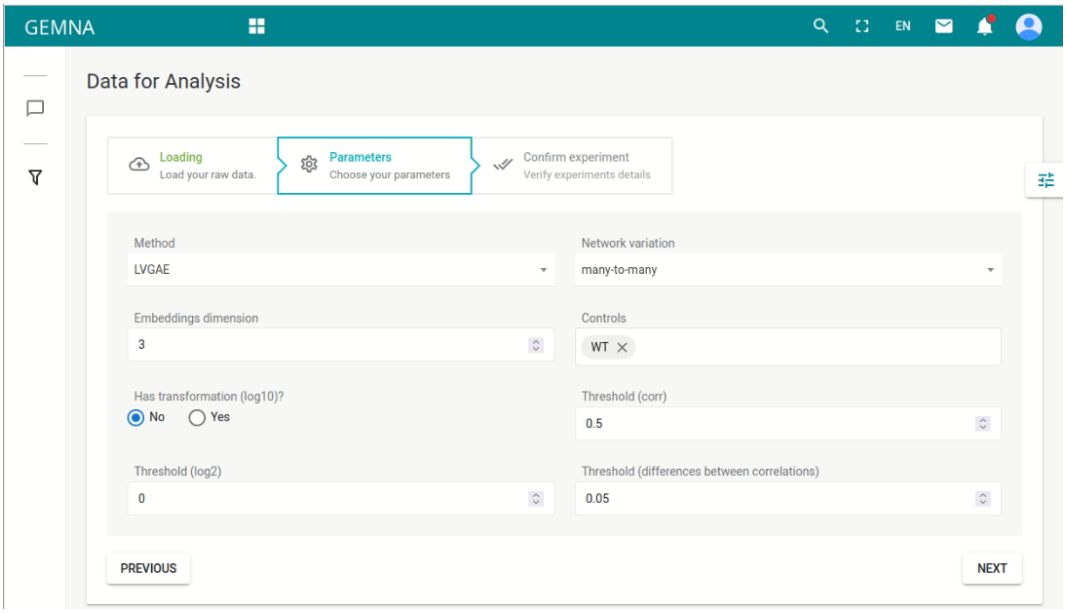

Figure 3: Choose your parameters.

2.2.3 Confirm experiment

In this last step, all selected parameters are displayed (see Figure 4). To start the process, click on the FINISH button. Then, the web application will start processing and analyzing the data. At the end, the user will receive an email

with a code, as shown in Figure 5. Additionally, you will receive .csv files with the correlation changes between the phenotypes. These files are already formatted to upload the results to the `BioCyc` database.

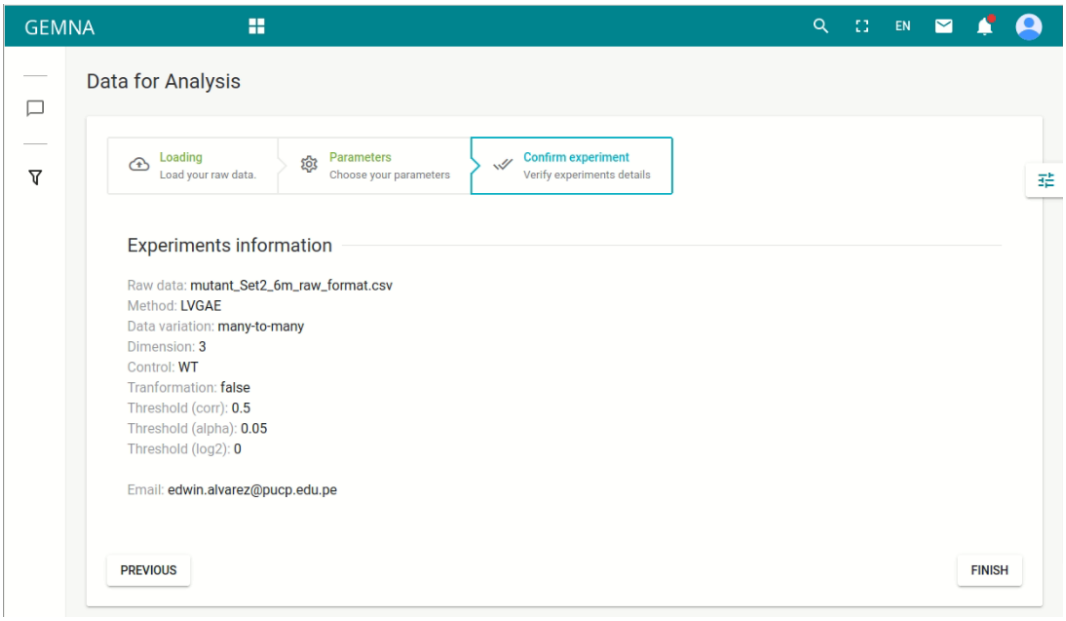

Figure 4: Verify experiment details.

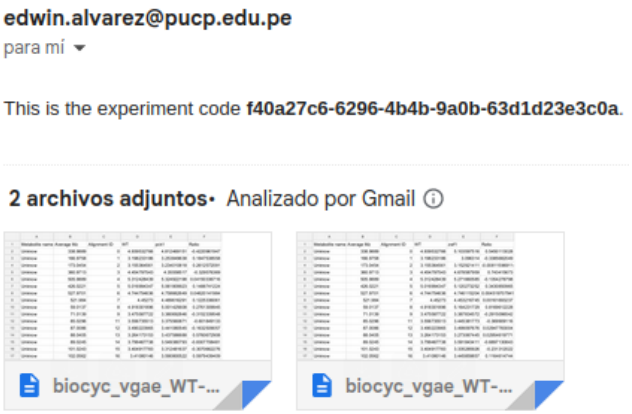

Figure 5: Received e-mail.

3 Visualization

3.1 Load experiment

Type the experiment's code in the form (see Figure 6), and click the LOAD button. After that, two summary tables are displayed. There, you can see the details of the networks and a table with the number of correlation changes, as shown in Figure 7.

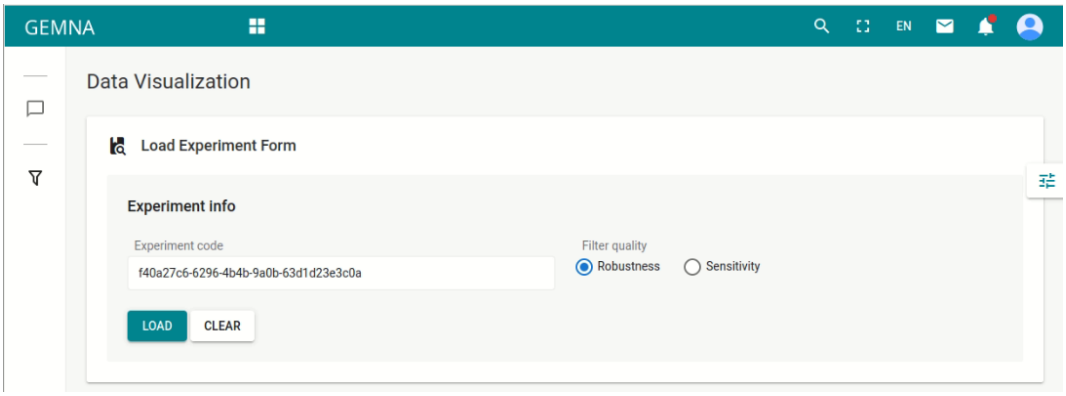

Figure 6: Load experiment form.

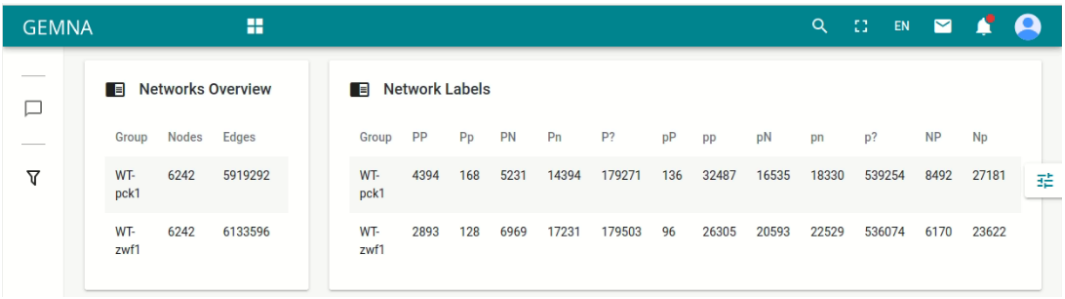

Figure 7: Network overview and network labels.

3.2 Analysis of specific result

After the previous step, a form will be displayed to perform a more specific analysis of the results (see Figure 8). First, you select the groups to compare (Wt-pck1 or WT-zwf1). Second, you select the metabolites of interest, and the search can be done using any of the metadata (ID, Retention time, Average Mz, Metabolite name). Thirdly, the type of visualization selected can be *Correlation nodes*, where the network shows only the correlation between the

selected metabolites. In contrast, the *Correlation + neighbors nodes* option shows the previous network plus all the neighbors of the selected nodes. To see the results, click on the FILTER button.

3.3 Similarity analysis and heatmap

Finally, examine the interaction between the metabolites of interest, Figure 9 (left) shows the change in correlation between the metabolites of interest. On the other hand, Figure 9 (right) shows the heatmap of the correlation changes.

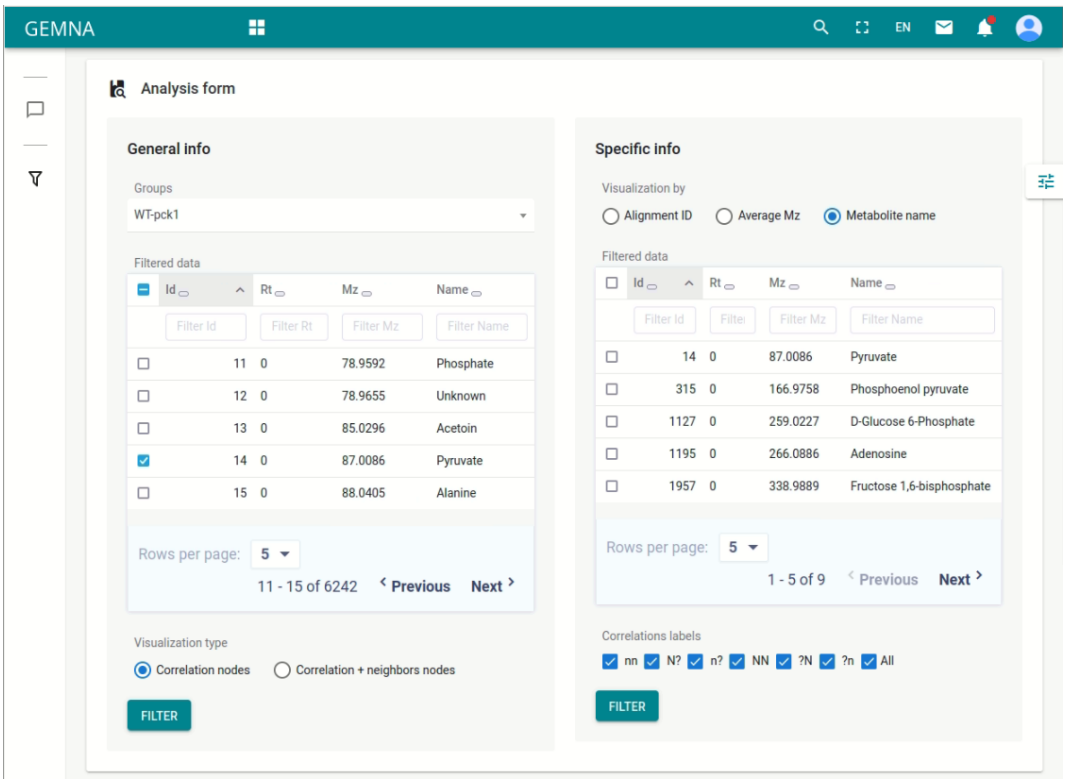

Figure 8: Analysis form.

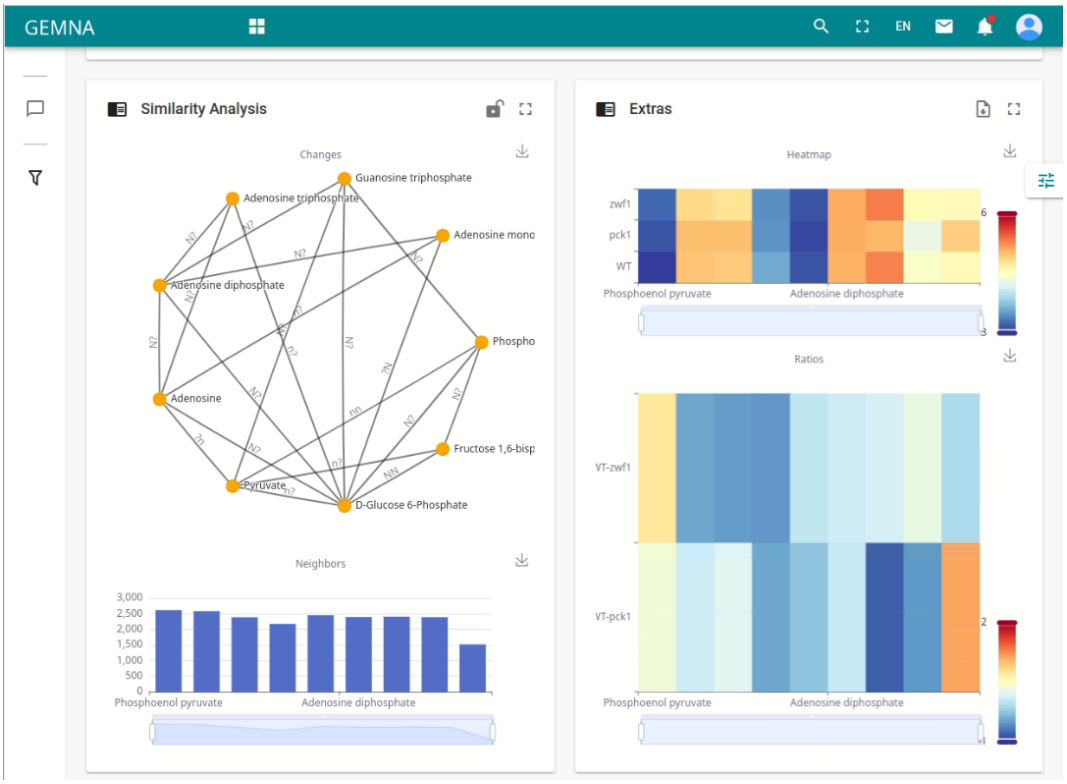

Figure 9: Similarity analysis and heatmap.