



Using a multiview data integration approach for active module detection

Giresse TCHOTANEU & Claude PASQUIER

I3S Laboratory

June 22 2023



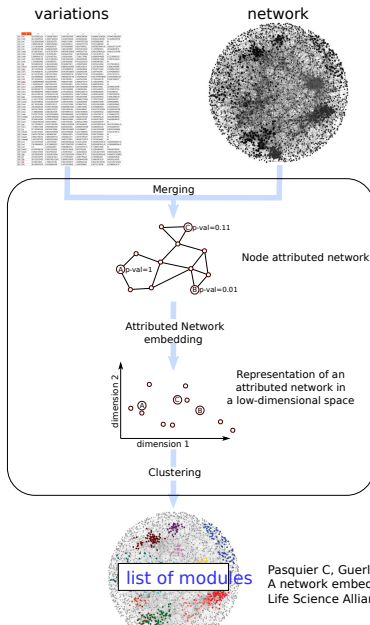
Active module detection with AMINE

I
N
P
U
T

A
M
I
N
E

W
O
R
K
F
L
O
W

O
U
T
P
U
T



<http://amine.i3s.unice.fr/>

The screenshot shows the AMINE web interface with the following sections:

- Header:** Includes the URL <http://amine.i3s.unice.fr/> and navigation links (Home, About, Contact).
- Introduction:** Describes the AMINE algorithm as a greedy approach for active module detection based on network embedding.
- Execution:** Provides instructions on how to use the tool, including a note about the input file format (CSV or Excel) and a link to the GitHub repository.
- Module detection form:** Contains fields for:
 - Job file (text input)
 - Differential expression data (button: [Browse...](#))
 - Column with gene names (dropdown menu)
 - Column with p-values (dropdown menu)
 - Column with log2 fold change values (dropdown menu)
 - Species (dropdown menu: [Homo sapiens](#))
 - Favourite genes (optional) (button: [Browse...](#))
 - E-mail (optional) (text input)
 - [Submit](#) button
- Footer:** Displays the date 05/2020.

<https://github.com/claudepasquier/amine>

Pasquier C, Guerlais V, Pallez D, Rapetti-Mauss R & Soriani O (2023).
A network embedding approach to identify active modules in biological interaction networks.
Life Science Alliance. Volume 6, No. 9. September 2023 .

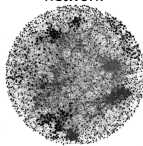
Internship

I
N
P
U
T

variations



network



variations 2



A
M
I
N
E

W
O
R
K
F
L
O
W

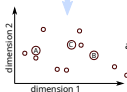
O
U
T
P
U
T

Merging



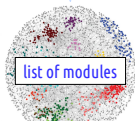
Node attributed network

Attributed Network
embedding



Representation of an
attributed network in
a low-dimensional space

Clustering



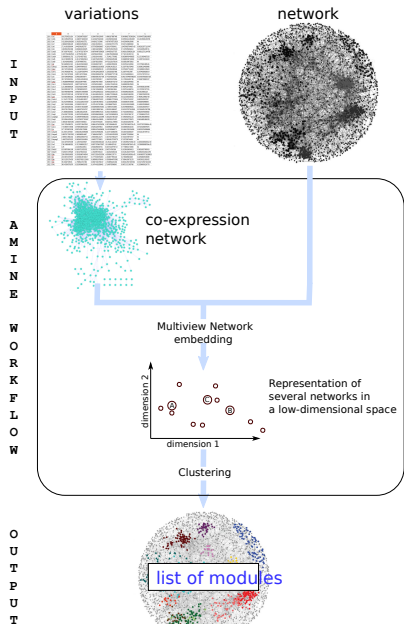
Internship

Development of an active module identification method
using gene expression and protein abundance (1)

- developing a multiview data integration approach
(multiview clustering, multiview embedding)
- Combining the multiview approach
with an active module identification method
- Evaluation of the method, on artificial and public data sets
- Release of the method to the community

processing of data generated by the biology trainee

Work in progress



Work in progress

- Transformation of gene expressions into a co-expression network
- Evaluating various libraries for multiview network embedding
- ➔ - Assessing the quality of the resulting embeddings
- Performing module detection on the combined embedding
- Evaluation of the method on artificial and public data sets
- Diffusing the method to the community

- Céline used the application and analyzed the results.
- She executed the method on several types of data.
 - ▶ Full RNA-Seq data.
 - ▶ Filtered RNA-Seq data, specifically keeping only genes that show significant variation (associated with a $p\text{-value} < 0.05$).
- This enabled us to identify differences in the results, which may provide clues for using and improving the algorithm
 - ▶ The modules found with filtered data are larger and show more significant enrichments.
 - ▶ Using all the data enable us to identify genes within certain modules that exhibit minimal variation (but may nevertheless be of interest).
 - ▶ The reason is that when genes are missing from the AMINE input, they are associated with a $p\text{-value}$ of 1, which is very (too?) penalizing.
 - ▶ Using filtered data seems to provides a quick idea of which modules are active, while using all the data gives a more precise understanding of what is going on.