



GENE SET PATHWAY MERGING AND ANALYSIS FOR ZTTK

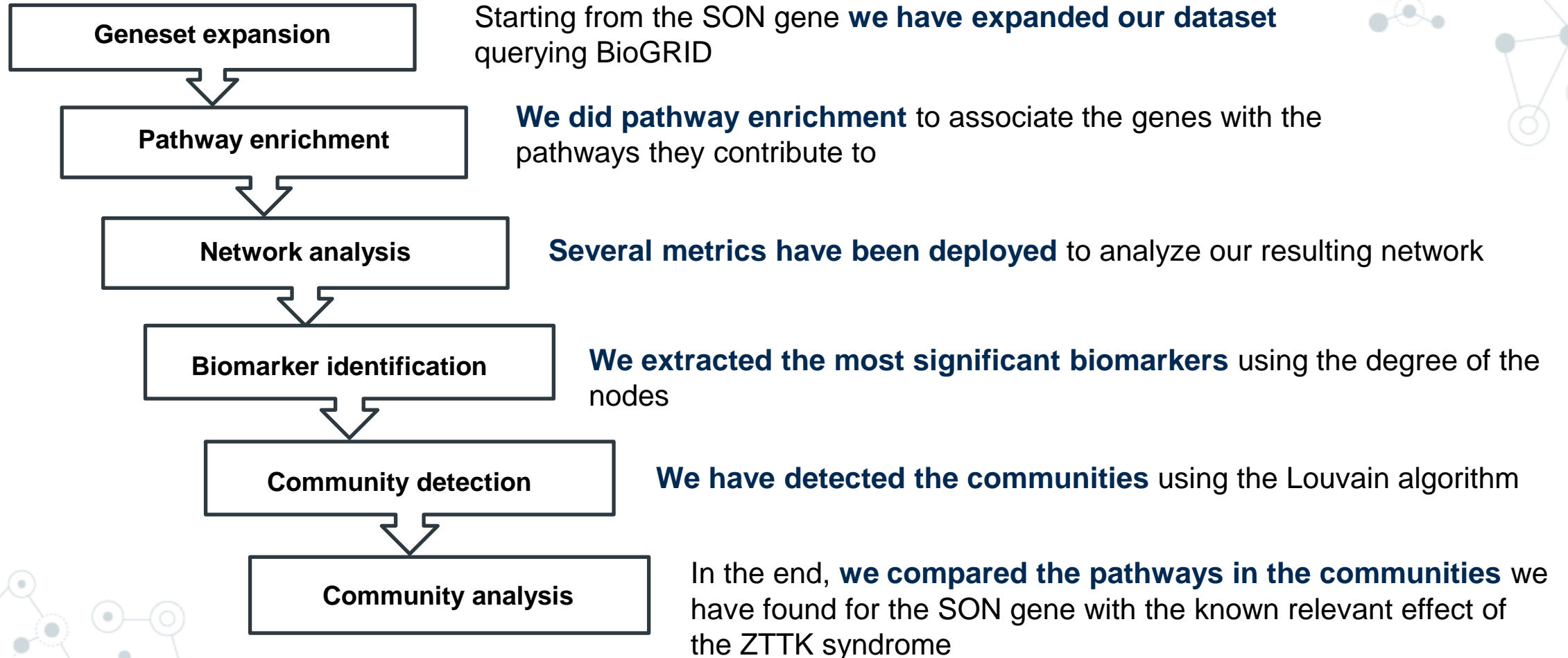
Niko Dalla Noce, Alessandro Ristori, Andrea Zuppolini

Computational Heath Laboratory Project

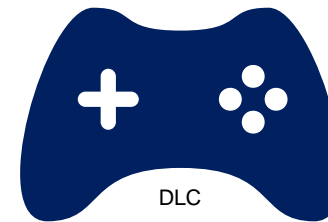
16/05/2022



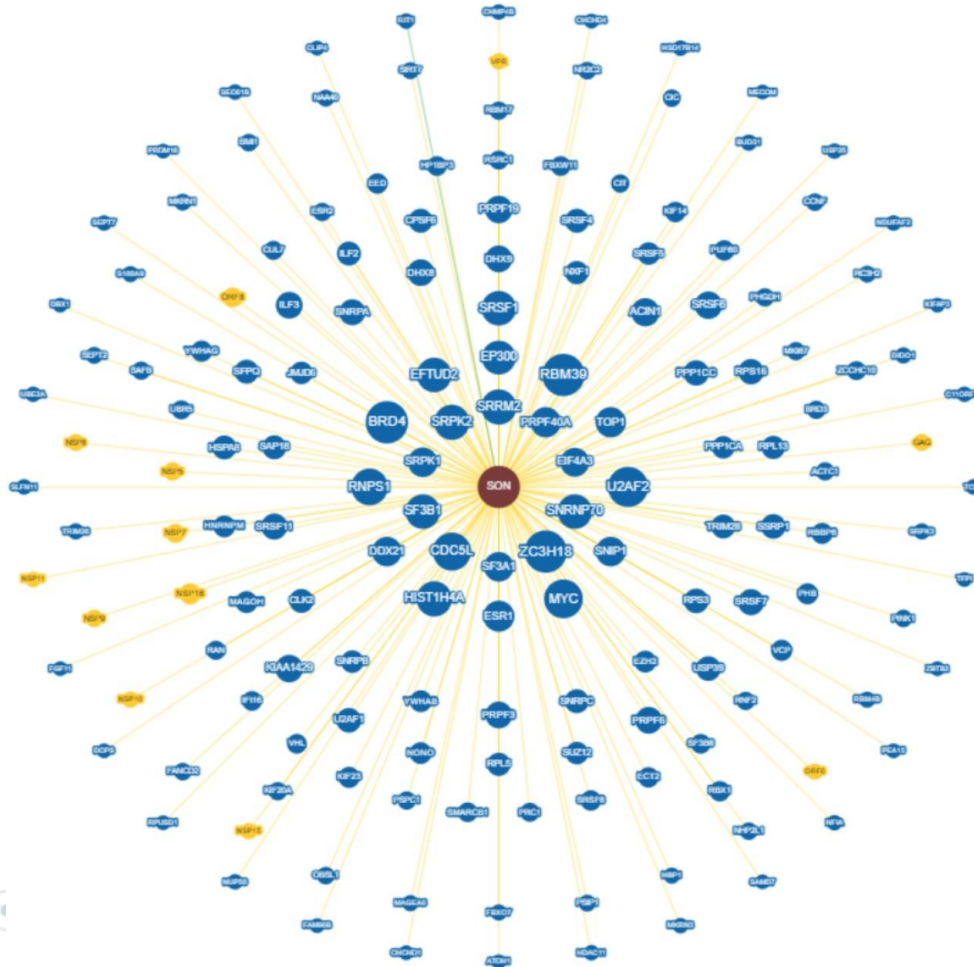
Development keypoints



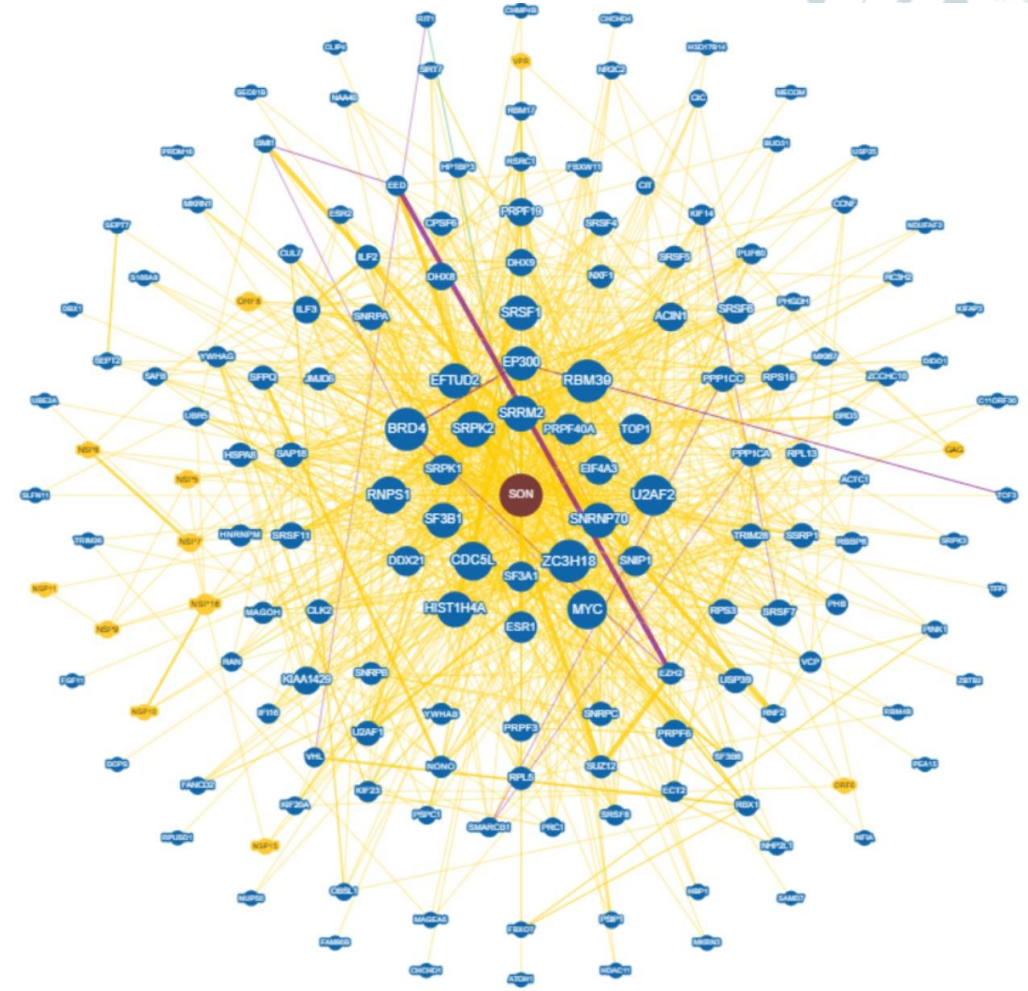
GENESET EXPANSION



Geneset expansion

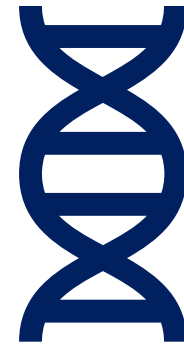


The 146 starting interactions







The interactions at the first order

PATHWAY ENRICHMENT



Pathway enrichment

- Starting from the interactions dataset obtained at the end of the geneset expansion phase, **we exploit the GSEapy package to perform pathway enrichment** on the nodes linked to it:
 - First, we tried with the  and  reactome human datasets with no success since we would have to manually remove all those pathways not related to any disease;
 - then we found the  dataset, which satisfied our needs.
- Using the  dataset, **we retrieved and filtered the disease pathways** by keeping those having a p-value lower than 0.1, totaling 589 pathways.

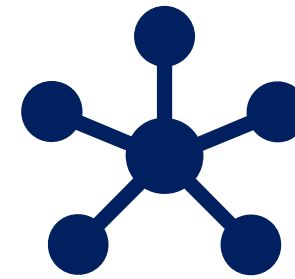
Disgenet

- The **GSEAPy** package, no matter the dataset, always returns a dataframe.

	Term	Overlap	P-value	Adjusted P-value	Genes
584	Chronic otitis media	55/69	0.005896	0.098950	IGHM;CD81;WIPF1;FMR1;DOCK8;CHD7;JMJD1C;COMT;GT...
585	Inadequate arch length for tooth size	47/58	0.005953	0.099228	AMER1;SETD5;NOTCH3;TRIO;RPL10;SATB2;GNAI3;PLOD...
586	Tooth Crowding	47/58	0.005953	0.099228	AMER1;SETD5;NOTCH3;TRIO;RPL10;SATB2;GNAI3;PLOD...
587	Tooth mass arch size discrepancy	47/58	0.005953	0.099228	AMER1;SETD5;NOTCH3;TRIO;RPL10;SATB2;GNAI3;PLOD...
588	Tooth size discrepancy	47/58	0.005953	0.099228	AMER1;SETD5;NOTCH3;TRIO;RPL10;SATB2;GNAI3;PLOD...

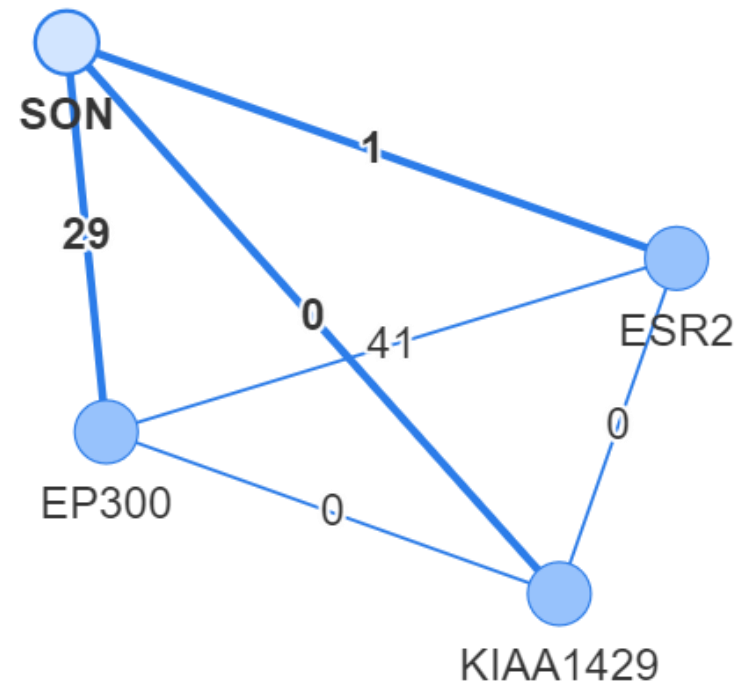
An example of disease pathways retrieved by using the GSEAPy package.

NETWORK ANALYSIS



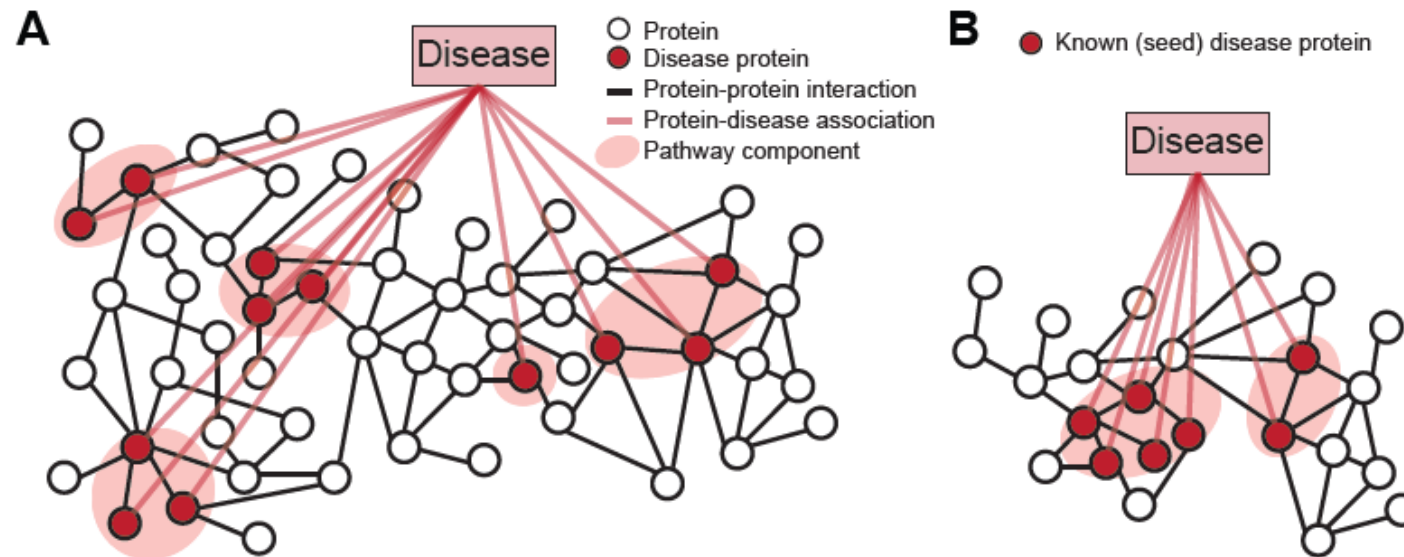
Network building

- Starting from the interaction dataset built after the geneset expansion **we developed our protein-to-protein network** thanks to  **NetworkX**, a Python package for network analysis.



Protein network insight

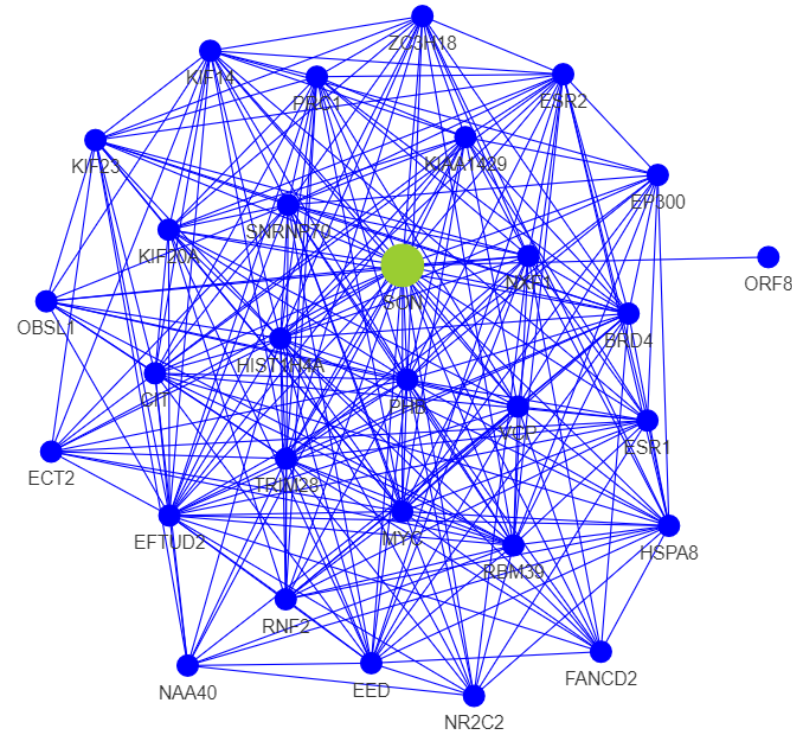
- A relevant number of nodes/proteins does not belong to any disease, more precisely 5101 of the 13010 proteins;
- **We decided to keep them** since we did not want to end with possibly a non connected graph;
- **We observed that 45000 edge had no shared diseases** between the couple of nodes they connect.



Biomarker identification

- By using a **centrality algorithm** from **networkx** based on the nodes' degree **we extracted the biomarkers**.
- The centrality is the **ratio between the node degree over the entire number of nodes** that compose the graph.

	centrality
KIAA1429	0.223922
ESR2	0.175801
ESR1	0.174879

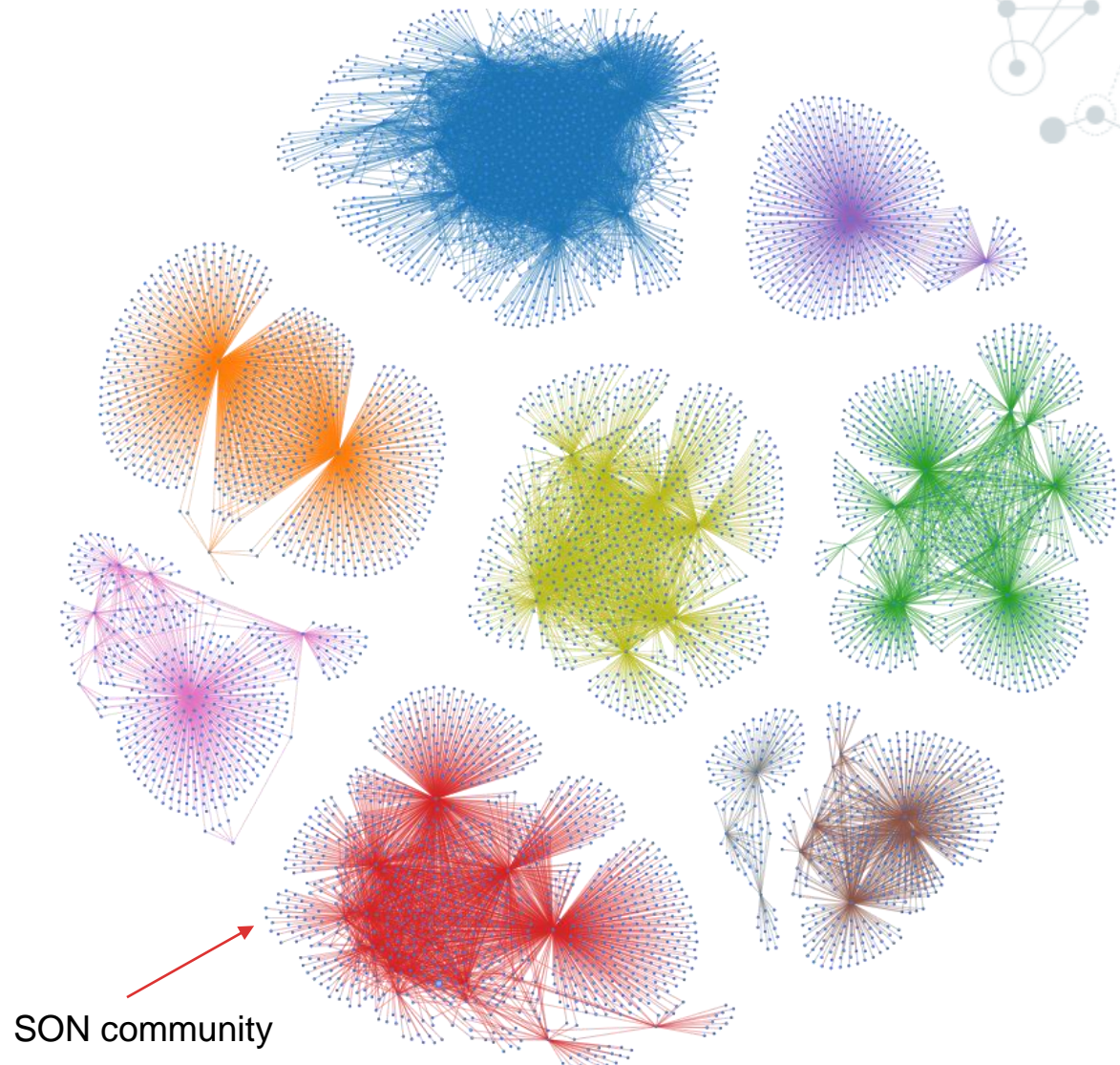


COMMUNITY ANALYSIS

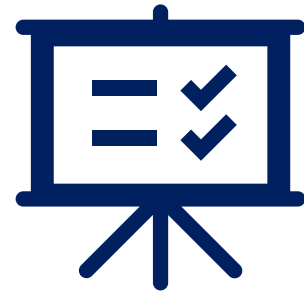


COMMUNITY DETECTION

- In order to find communities in our protein-protein interaction graph, **we exploit the Louvain community detection algorithm** which runs in $O(n * \log(n))$
- **We drop the communities having less than two nodes;**
- After the pruning, **from 8 to 11 communities are left.**



RESULTS



Community evaluation: metrics

- **Ratio disease:** measures the ratio between the genes of the given community which participate to a disease pathway over the whole size of that pathway, formally $R_d = \frac{n_c}{n_d}$
- **Ratio community:** a ratio between the number of genes in a disease pathway inside in a community over the size of the community defined as $R_c = \frac{n_c}{|V_c|}$
- **Relevance:** it combines the absolute contribution of the pathway with the community size and is obtained multiplying the previous 2 metrics, formally $Rel = R_d * R_c$

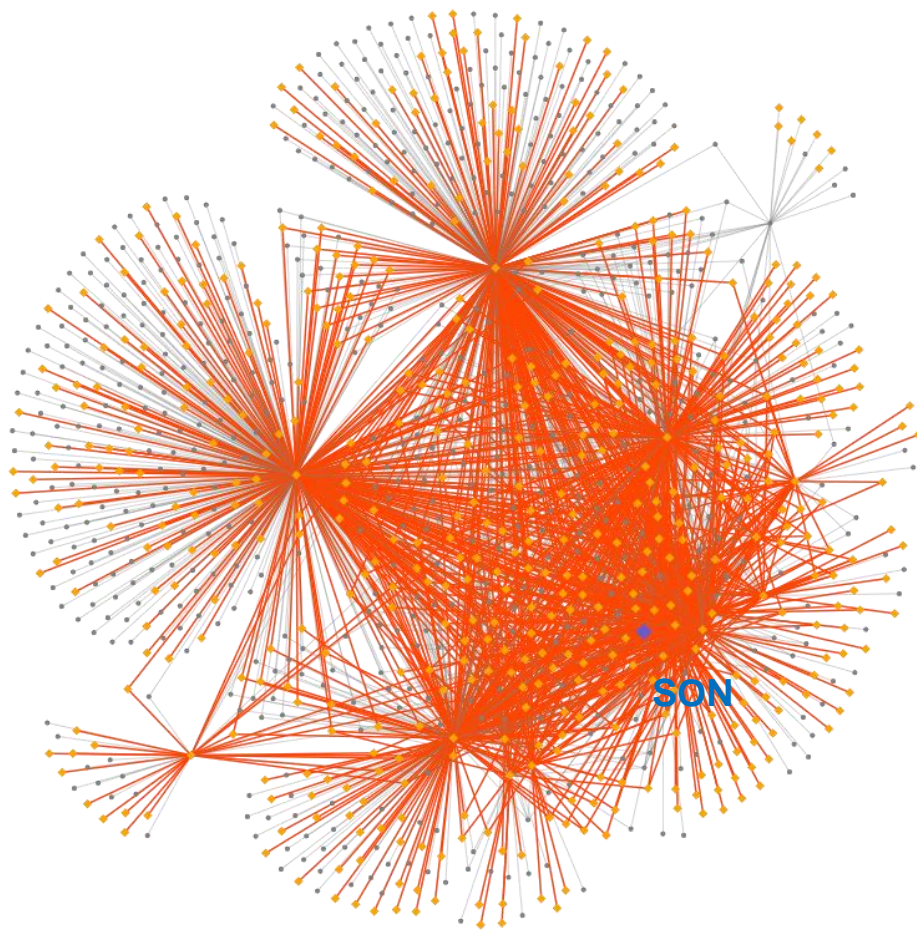
	Community	Disease	Shared genes	Disease genes	Community size	Ratio disease	Ratio community	Relevance
566	0	Tooth size discrepancy	5	47	1084	0.106383	0.004613	0.000491
1084	1	Tooth size discrepancy	2	47	510	0.042553	0.003922	0.000167
1630	2	Tooth size discrepancy	3	47	894	0.063830	0.003356	0.000214

Relevant diseases in SON community

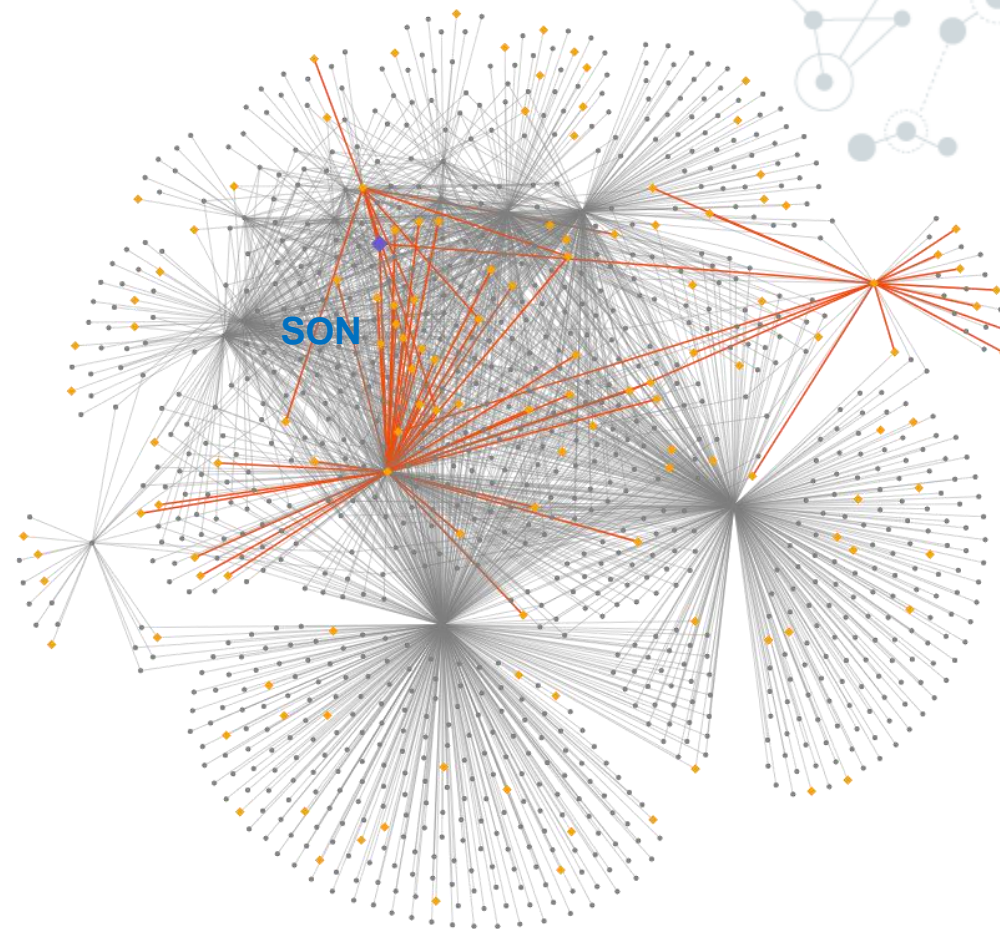
- Diseases found in the SON community, ordered by Relevance.
- Each disease **contains SON** in its pathway.

Disease	Ratio disease	Ratio community	Relevance ^
Intellectual Disability	0.231748	0.514742	0.119290
Mental and motor retardation	0.296758	0.292383	0.086767
Mental Retardation	0.283688	0.294840	0.083643
Poor school performance	0.302294	0.275184	0.083187
Cognitive delay	0.298153	0.277641	0.082780
Mental deficiency	0.294194	0.280098	0.082403
Global developmental delay	0.279859	0.293612	0.082170
Dull intelligence	0.304775	0.266585	0.081248
Low intelligence	0.304775	0.266585	0.081248
Short stature	0.279310	0.199017	0.055588
Generalized hypotonia	0.261053	0.152334	0.039767
Strabismus	0.312693	0.124079	0.038799
Failure to gain weight	0.269136	0.133907	0.036039
Pediatric failure to thrive	0.268473	0.133907	0.035950
Undergrowth	0.270202	0.131450	0.035518
Genetic Diseases, Inborn	0.256684	0.117936	0.030272
Acquired scoliosis	0.275000	0.108108	0.029730
Curvature of spine	0.271565	0.104423	0.028358
Low set ears	0.274262	0.079853	0.021900
Dilated ventricles (finding)	0.329268	0.066339	0.021843
Cerebellar Hypoplasia	0.361345	0.052826	0.019088
Feeding difficulties	0.266667	0.063882	0.017035

Relevant diseases in SON community



Intellectual disability disease



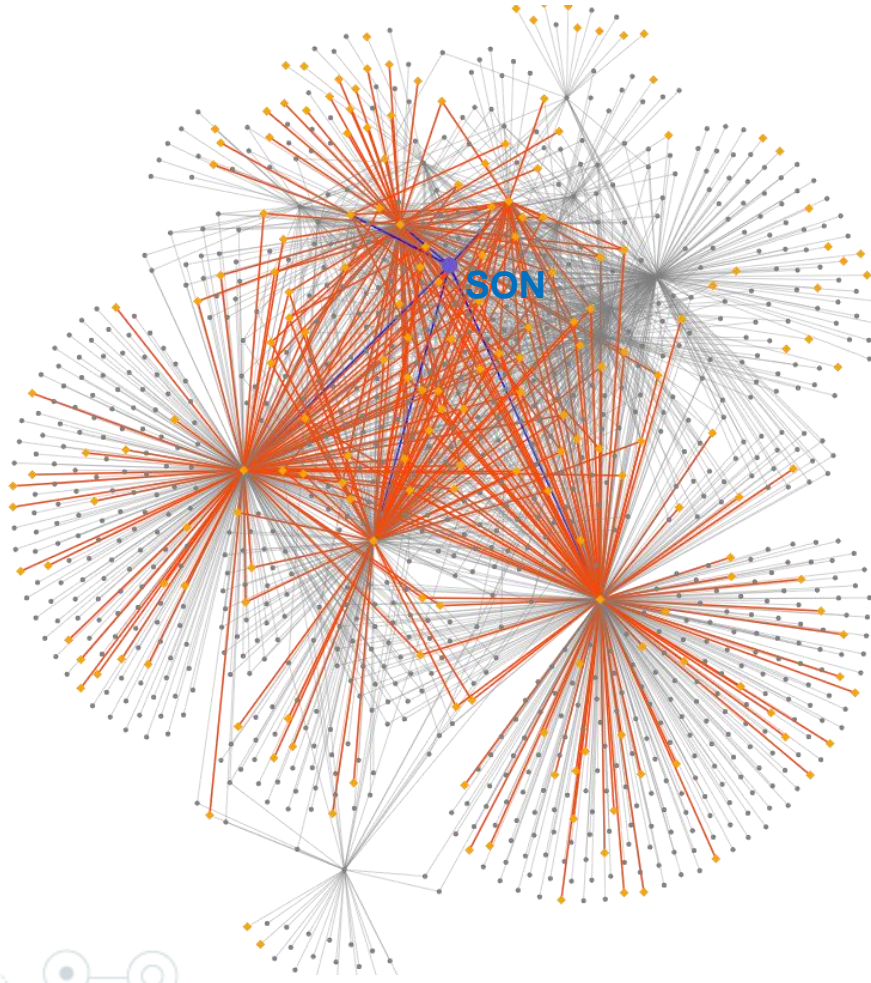
Undergrowth disease

Relevant diseases in SON community

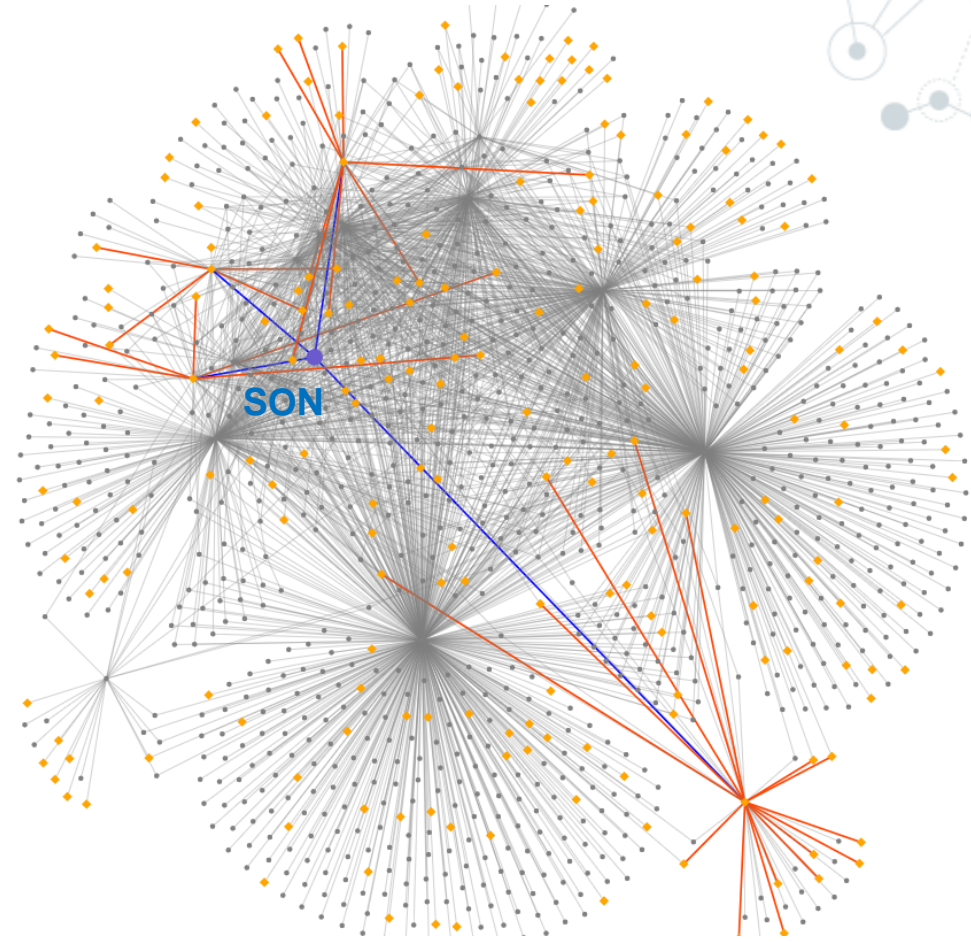
- **Diseases found in the SON community**, ordered by Relevance.
- **SON is not inside any of those diseases**, but some genes of their pathways interact with SON.

Disease	Ratio disease	Ratio community	Relevance
Small head	0.417021	0.178020	0.074238
Seizures	0.265931	0.197094	0.052413
Malignant neoplasm of breast	0.125584	0.390554	0.049047
Epilepsy	0.249395	0.187103	0.046662
Breast Carcinoma	0.118969	0.360581	0.042898
Aplasia/Hypoplasia of the mandible	0.355556	0.101726	0.036169
Hypoplastic mandible condyle	0.354633	0.100817	0.035753
Retrusion of lower jaw	0.354633	0.100817	0.035753
Decreased projection of lower jaw	0.354633	0.100817	0.035753
Decreased projection of mandible	0.354633	0.100817	0.035753
Micrognathism	0.349057	0.100817	0.035191
Hyperreflexia	0.407725	0.086285	0.035181
Primary microcephaly	0.596154	0.056312	0.033571
Carcinogenesis	0.109986	0.276113	0.030368
Microcephaly	0.407035	0.073569	0.029945
Mitochondrial Diseases	0.339223	0.087193	0.029578
Epileptic encephalopathy	0.299720	0.097184	0.029128
Muscle Spasticity	0.329710	0.082652	0.027251
Fetal Growth Retardation	0.346774	0.078111	0.027087
Cryptorchidism	0.303797	0.087193	0.026489

Relevant diseases in SON community

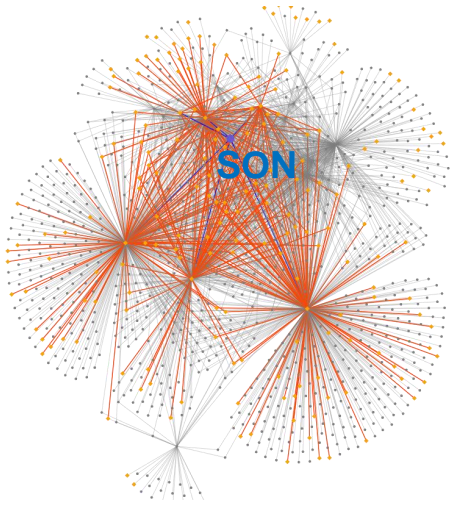


Small head disease



Epilepsy disease

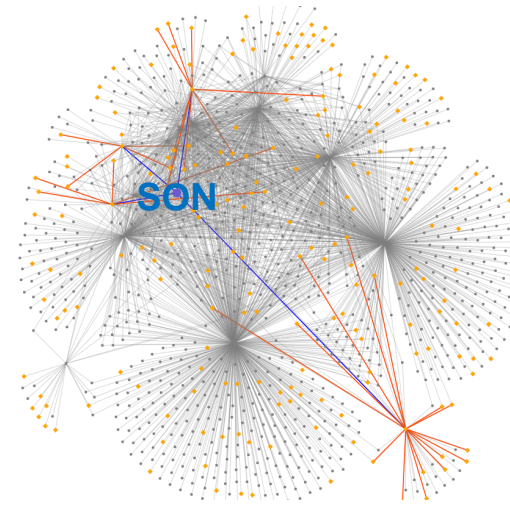
Relevant diseases in SON community



Small head disease

Neighbouring genes:

- Citron Rho-Interacting Serine/Threonine Kinase
- Elongation Factor Tu GTP Binding Domain Containing 2
- Histidine Triad Nucleotide-Binding Protein 5
- Fanconi Anemia Complementation Group D2



Epilepsy disease

Neighbouring genes:

- Histone-Lysine N-Methyltransferase
- Ubiquitin-Conjugating Enzyme E2A
- Myc-Induced Mitochondrial Protein

A decorative network diagram in the top-left corner, featuring a complex web of interconnected nodes and lines. Some nodes are highlighted with blue circles, and others with solid blue dots.

THANKS FOR THE ATTENTION

A decorative network diagram in the bottom-right corner, featuring a complex web of interconnected nodes and lines. Some nodes are highlighted with blue circles, and others with solid blue dots.