AutoNetCan Webserver User Manual

Automated Biomolecular Network Construction for Translational Cancer Systems Biology

(*Version 1.0*)

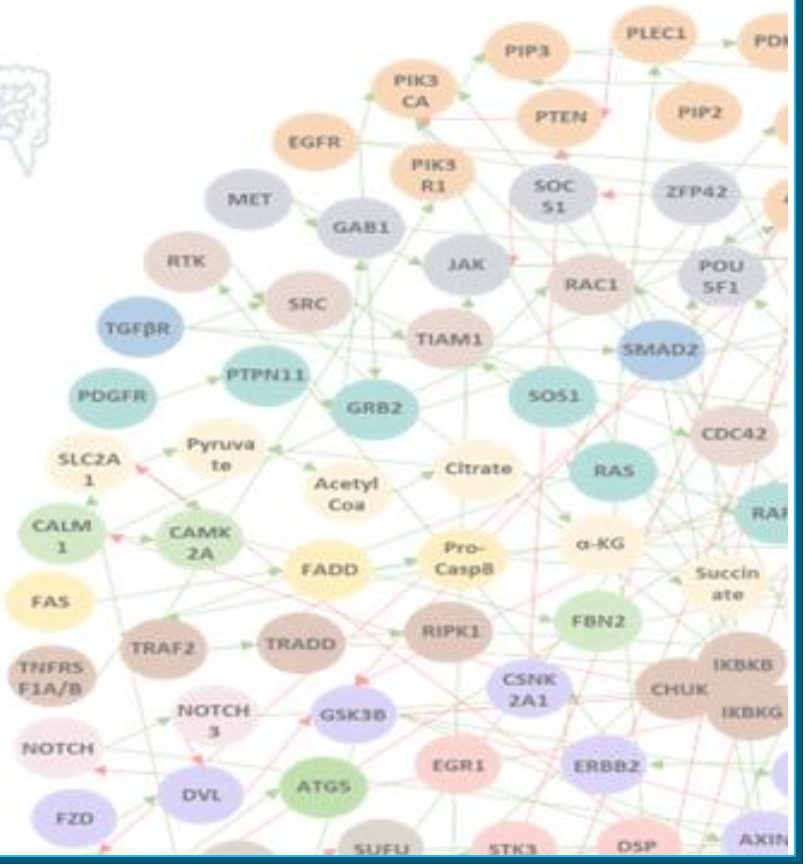

Table of Contents

Getting Started: Select a Cancer Type

AutoNetCan is designed to facilitate the automated construction of cancer-type-specific biomolecular networks. The process begins with the selection of a target cancer type using a drop-down menu in the platform interface (**Figure 1**).

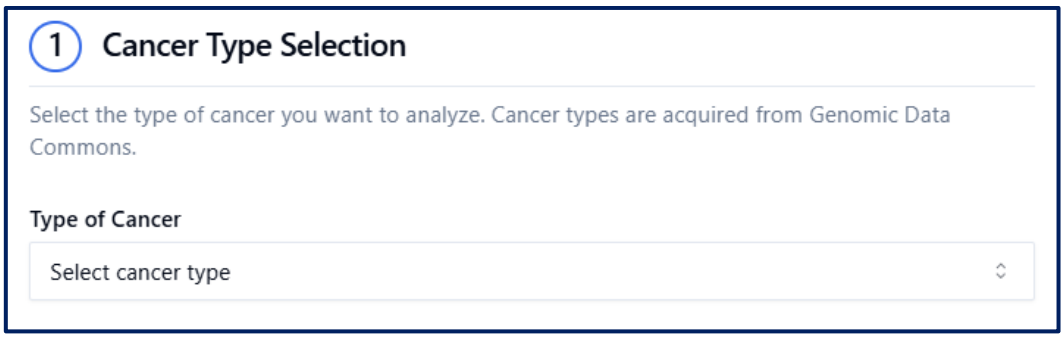

Figure 1: Cancer Type Selection Interface

The user selects from a list of 28 predefined cancer types sourced from the Genomic Data Commons (GDC). Once the cancer type is selected, AutoNetCan guides the user through a structured four-step workflow to build a network:

1. **Node Set Acquisition** – Assemble biologically relevant genes from curated datasets.
2. **Node Set Enrichment** – Expand the node list by identifying enriched pathways and gene sets.
3. **Connectivity Mapping** – Define molecular interactions by integrating known regulatory relationships.
4. **Logical Rules Modeling** – Infer logical rules governing gene regulation using transcriptomic data.

Each step is outlined in detail in the sections below, with clear guidance on user input, available options, and output files generated at every stage.

1. Node Set Acquisition

This is the first stage of network construction, where AutoNetCan gathers an initial node set (genes/nodes) relevant to the selected cancer type. Nodes are acquired from multiple high-quality sources to ensure a comprehensive starting network. These sources include:

- **Differentially Expressed Genes (DEGs)**: Genes significantly up- or down-regulated in the cancer tissue compared to normal tissue.
- **Frequently Mutated Nodes**: Genes with high mutation frequency in the cancer cohort.
- **Therapeutically Targetable Nodes**: Genes that are targets of known cancer therapies (approved drugs, clinical trial drugs, or experimental compounds), including DNA repair genes.
- **Oncogenes, Tumor Suppressors, and Driver Genes**: Known cancer-associated genes curated from literature.
- **Cancer-Specific Gene Panels**: Genes from diagnostic or prognostic panels specific to certain cancers.

Each category above can be selected to include the corresponding nodes in the network. The web interface will guide you through each in sequence.

Tip: At each step of node acquisition, you may enable the *"Download detailed file(s) for this step"* option to obtain a comprehensive output file for that step. By default, AutoNetCan will always produce a summary list of nodes for each step, but enabling this option provides full details for advanced analysis.

1.1 Differentially Expressed Genes (DEGs)

The cancer type selected determines which data (expression, mutations, etc.) will be used from the Genomic Data Commons. After choosing the cancer type, users have the option to specify parameters for differential gene expression analysis (**Figure 2**).

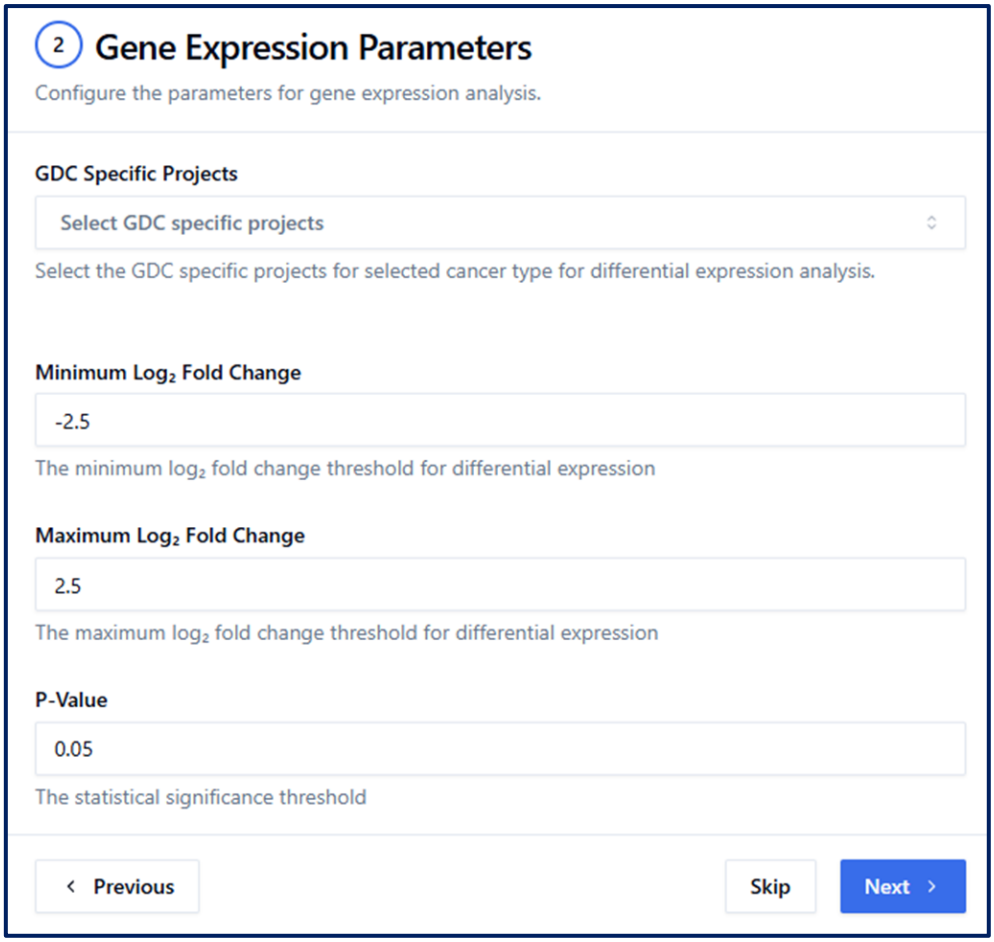

Figure 2: Differential Gene Expression Parameters

Users can provide thresholds to identify significant DEGs by setting the minimum and maximum $\log_2$ fold-change and a p-value cutoff for significance. Once parameters are entered (or defaults are used), click "*Next*" to proceed.

Note: Some cancer types (e.g., bone, eye, lymph nodes, nervous system, ovary, pleura, testis) do not have normal (healthy) tissue samples in Genomic Data Commons (GDC). In those cases, the DEGs step is skipped because differential expression cannot be computed for those cancers.

Input Parameters

- **GDC specific projects:** A list of cancer projects available from GDC used for differential gene expression analysis.
- **Minimum $\log_2$ Fold Change:** The lower threshold for gene expression changes. Genes with a $\log_2$ fold change below this value will be included (default –2.5).
- **Maximum $\log_2$ Fold Change:** The upper threshold for gene expression changes. Genes with a $\log_2$ fold change above this value will be included (default +2.5).

- ***p*-Value Threshold:** The statistical significance cutoff for identifying differentially expressed genes. Only genes with p-values below this threshold are considered significant (default 0.05).

Note: User can only select one project at a time for constructing the network.

Output

- **Detailed DEGs File** – Comprehensive list of all genes tested, each with its $\log_2$ fold-change, p-value, adjusted p-value, etc.
- **Summary DEGs File** – A summary list of significantly up-regulated and down-regulated genes (nodes) meeting the thresholds. This provides the set of DEGs that will be included in the network.

DESeq2 Analysis

For the selected cancer type, differential gene expression analysis is carried out using DESeq2, a widely used R package for analyzing RNA-Seq data [1]. The platform automatically initiates the DESeq2 analysis once the differential expression parameters are defined and the necessary input files, including gene expression counts and sample metadata, are provided (at backend). DESeq2 identifies differentially expressed genes by comparing tumor and normal samples. For most cancer types, *Primary Tumor* samples are treated as tumor samples, while *Solid Tissue Normal* samples are considered healthy controls. However, for bone marrow and blood cancers, *Primary Blood Derived Cancer – Bone Marrow* and *Primary Blood Derived Cancer – Peripheral Blood* are classified as tumor samples, whereas *Blood Derived Normal* and *Bone Marrow Normal* are treated as healthy controls. To ensure relevance and consistency, the analysis is restricted to a unique set of protein-coding genes.

1.2 Frequently Mutated Genes

AutoNetCan allows users to optionally include genes that are frequently mutated in the selected cancer type (**Figure 3**). When this option is enabled, the platform automatically queries the Genomic Data Commons (GDC) API to retrieve mutation frequency data based on the specified primary site and tissue or organ of origin. All genes reported by GDC as frequently mutated in the selected cancer type are included, resulting in a node set that typically comprises approximately 700 genes.

To include these genes, simply tick the checkbox labeled "*Acquire nodes based on their mutational frequency*" No additional parameters are required — AutoNetCan will fetch

the relevant genes using default criteria via the GDC API. Once selected, click "*Next*" to continue.

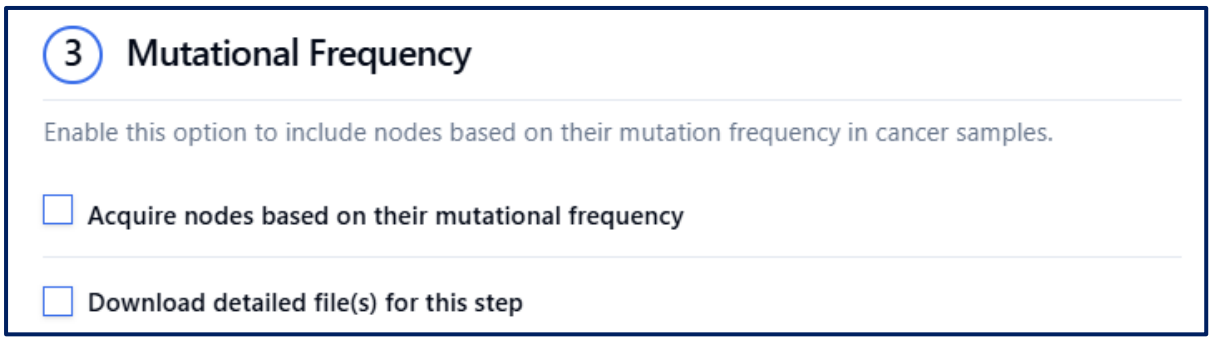

Figure 3: Mutational Frequency Option

Input Parameters

- **Include Mutated Genes:** Checkbox to include genes with high mutation frequency in the selected cancer cohort (optional).

Output

- **Detailed Mutational Frequency File** – List of genes with their mutation statistics (e.g., number of tumor samples with mutations, copy number variations, etc.) in the cohort, along with relevant annotations.
- **Summary Mutational Frequency File** – A unique list of high-frequency mutated genes identified for the cancer. These genes will be added as nodes in the network.

1.3 Therapeutically Targetable Nodes

In this step, you can expand the node set to include genes that are targeted by various cancer therapies. AutoNetCan allows you to incorporate: **(a)** targets of FDA-approved drugs, **(b)** targets of drugs in clinical trials, **(c)** targets of experimental compounds (e.g., from cell-line screens), and **(d)** genes from selected DNA repair pathways. Users can choose any combination of these to augment the network with clinically relevant nodes.

1.3.1. FDA-Approved Therapies

These are nodes targeted by drugs already approved for cancer treatment (chemotherapies, targeted therapies, immunotherapies, hormonal therapies, etc.). AutoNetCan curates drug–target relationships from public cancer therapy database.

Figure 4: FDA-Approved Therapy Targets

Select one or more categories of FDA-approved treatments whose target genes you want to include. For example, you may check *Chemotherapy*, *Targeted Therapy*, *Immunotherapy*, *Hormonal Therapy*, and/or *Unclassified* to include drug targets from those categories (**Figure 4**).

Note: "Unclassified" refers to therapies not explicitly categorized in databases as chemotherapy, targeted therapy, immunotherapy, etc.

Once you have selected the desired therapy categories, click "*Next*" to proceed.

Input Parameters

- **Therapy Category Selection:** Check one or more drug categories to include their target genes (e.g., chemotherapy, immunotherapy).

Output

- **Detailed Drug Targets File** – Includes drug names, approval status, therapy category, and corresponding target genes.
- **Summary Drug Targets File** – Unique list of therapy-associated nodes added to the network.

1.3.2. Clinical Trial-Based Therapies

Nodes targeted by drugs that are currently in clinical trials, retrieved from ClinicalTrials.gov data, can also be added.

Figure 5: Clinical Trial Therapy Selection

If desired, include targets of trial-phase drugs by selecting the trial phases. For instance, check *Phase I*, *Phase II*, *Phase III*, etc., to include gene targets of therapies in those trial stages (**Figure 5**). You can select any or all phases. After selecting, click "*Next*" to continue.

Input Parameters

- **Trial Phases:** Select one or more phases (Early Phase I, Phase I–IV) to include associated drug targets.

Output

- **Detailed Clinical Trial Targets File** – A list of drugs, trial phases, and their associated gene targets.
- **Summary Clinical Trial Targets File** – Unique list of genes associated with selected trial-phase therapies.

1.3.3. Experimental Therapies (Cell Line-Based)

To include experimental drug target nodes, provide a Z-score threshold (**Figure 6**). Z-score reflects sensitivity — the lower (more negative) the score, the more sensitive the cell line is to the drug. Adjust this cutoff as needed and click "*Next*".

Figure 6: Experimental Therapy z-score filter

Input Parameters

- **Z-Score Threshold:** Provide a z-score value (default: 0). Negative values indicate higher sensitivity.

Output

- **Detailed Experimental Targets File** – Includes drug name, z-score, dataset type (GDSC1/GDSC2), and target genes.
- **Summary Experimental Targets File** – Unique list of high-sensitivity experimental drug targets.

1.3.4. DNA Repair Pathway Genes

You may also enrich the network with genes from DNA damage repair pathways, as these are often crucial in cancer. AutoNetCan includes predefined sets of DNA repair genes from the REPAIRtoire database (covering pathways like

DNA Damage Signaling, Base Excision Repair, Mismatch Repair, etc.) (

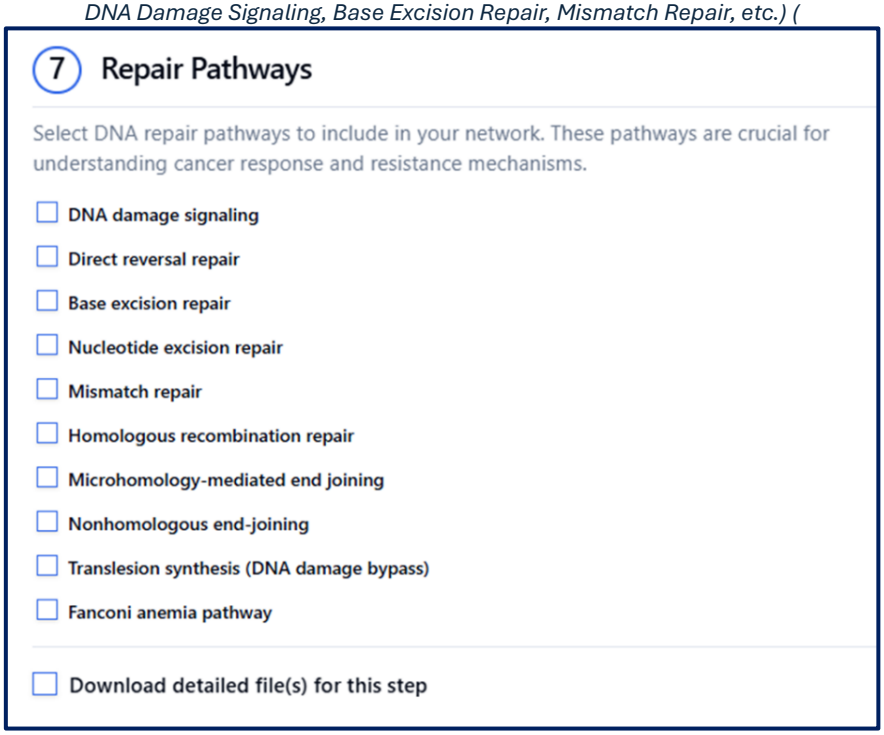

Figure 7).

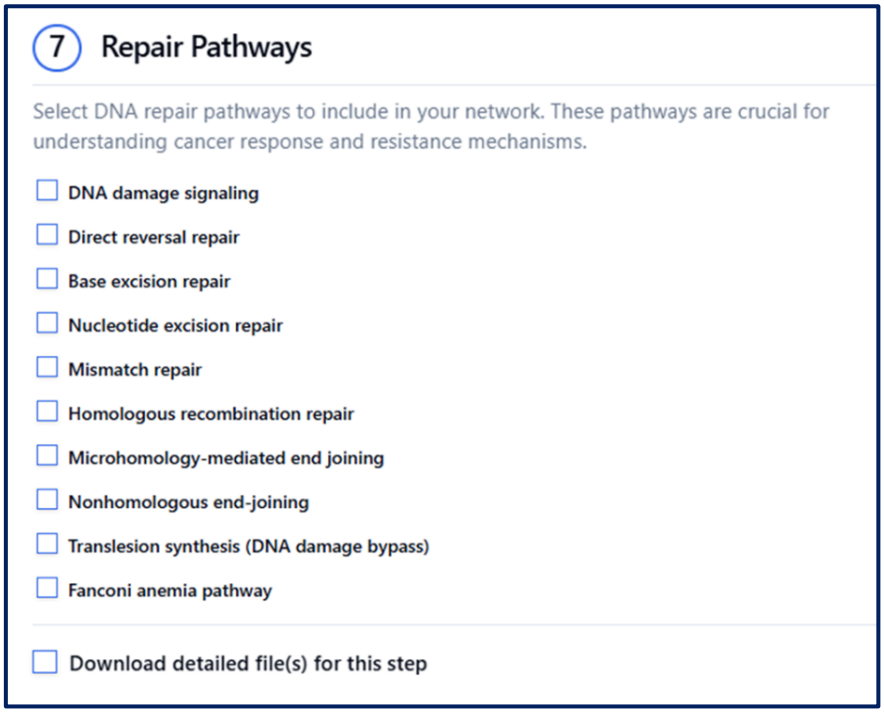

Figure 7: DNA Repair Pathway Selection

Check any option for DNA repair pathway by clicking the checkbox to add all genes in that pathway as nodes. For example, choosing Mismatch Repair (MMR) will include known MMR genes in the network. After selection, click "*Next*". s

Input Parameters

- **Repair Pathway:** Tick the checkbox to include genes from DNA repair pathways. One or more predefined pathways can be selected.

Output

- **Repair Pathway Summary File** – Unique list of genes from the selected repair pathway.

1.4 Cancer Signature Genes

In this step, users can enrich the network by adding genes that are recognized as being significantly associated with cancer. AutoNetCan allows for the inclusion of genes from four key categories: *Tumor Suppressor Genes* (TSG), *Oncogenes*, *Driver Genes*, and genes from *Cancer-Specific Diagnostic Panels* (**Figure 8**). These genes play crucial roles in cancer development and progression, making them valuable additions to the biomolecular network.
Users can select one or more gene categories, depending on their research needs, and proceed to include them in the network construction. After making selections, users can click *"Next"* to proceed.

These categories can be selected individually or in combination. Once selected, click "*Next*" to proceed.

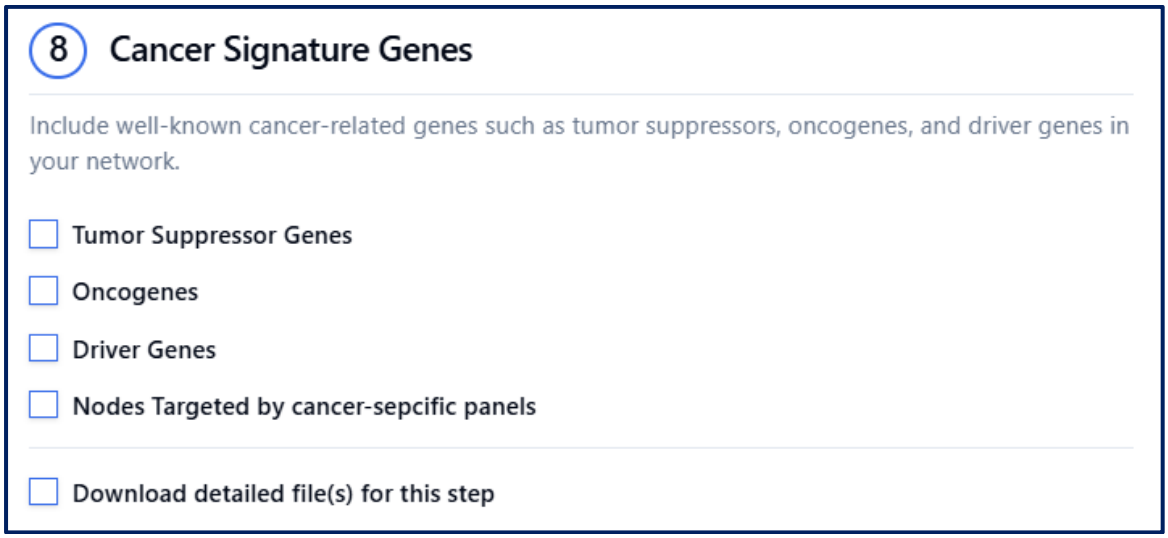

Figure 8: Cancer Signature Genes

Input Parameters

- **Cancer Signature Gene Types:** Select one or more of the following gene types to include:
    - Tumor Suppressor Genes
    - Oncogenes
    - Driver Genes
    - Cancer Panel Nodes

Output

- **Cancer Gene Detailed File** – Detailed information for each included cancer-related gene, including its gene ID, name, its classification (oncogene, TSG, or

driver), and reference information from CancerMine (e.g. the cancer context in which it's reported).

- **Cancer Panel Detailed File** (if the "Cancer Panel Nodes" checkbox is marked) – A list of all genes from cancer-specific panel(s) relevant to the chosen cancer, with details such as the panel name, provider (company or institution), the cancer type the panel is designed for, and the date of the panel data.
- **Cancer Gene Summary File** – A unique list of all oncogenes/TSGs/drivers added to the network, with an indication of their source category.

2. Node Set Enrichment

In this step, AutoNetCan allows you to expand your existing node list by identifying additional genes that are biologically enriched in known pathways. Using Over-Representation Analysis (ORA), AutoNetCan compares your current gene list against established pathway libraries and detects significant overlap. Pathways with a significant overlap will have their genes added to your node list, thus expanding the network and potentially revealing new biological insights.

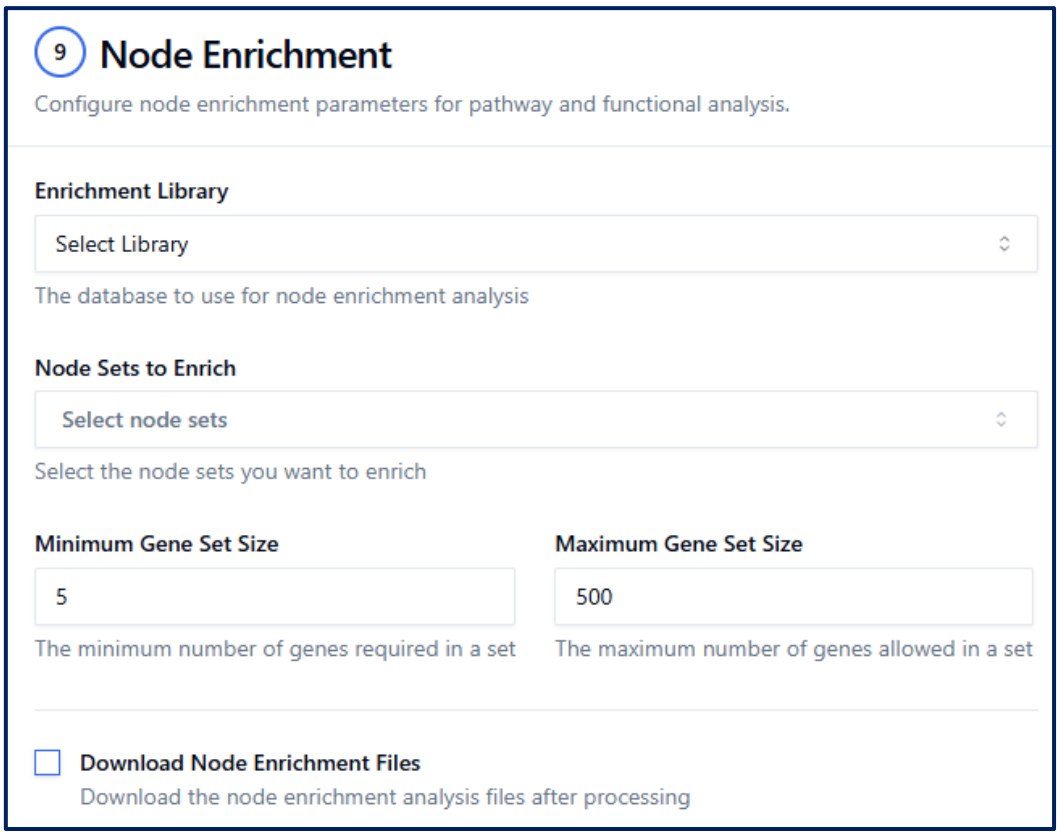

Figure 9: Node Set Enrichment

To run enrichment, select one of the available gene set libraries — *MSigDB* (Molecular Signatures Database) or *Enrichr*. After selecting the library, choose the specific library files you want to include in the analysis. Set the minimum and maximum gene set size to filter out very small or very large sets (default: 5–500 genes) (**Figure 9**). Click "*Next*" to begin. AutoNetCan will compare your current node list against the selected gene sets and identify those with statistically significant overlap. Once the enriched gene sets are identified, AutoNetCan will add unique genes from the enriched sets to your node list. This expansion enhances the network, potentially identifying new biomarkers or relevant genes associated with the cancer type under study.

Note: AutoNetCan supports the inclusion of various node sets for enrichment, including differentially expressed genes (upregulated and downregulated genes separately), frequently mutated genes, therapeutically targetable genes (including FDA-approved

drug targets, clinical trial-based targets, experimental therapy targets, and DNA repair pathway genes – each selectable independently), and cancer signature genes (tumor suppressor genes, oncogenes, and driver genes). Nodes that are part of predefined cancer-associated panel are excluded from the enrichment analysis.

Input Parameters

- **Enrichment Library:** Select the source of gene sets for ORA. Options: *MSigDB* or *Enrichr*.
- **Enrichment Library Files:** Select the files from *MSigDB* (39 library files) or *Enrichr* (more than 150 files) for enrichment.
- **Minimum Gene Set Size:** Smallest gene set size to consider (default 5; gene sets with fewer genes than this will be ignored).
- **Maximum Gene Set Size:** Largest gene set size to consider (default 500; gene sets larger than this will be ignored).
- **p-adjusted threshold**: Significance cutoff for enrichment analysis (default 0.05).

Output

- **Enrichment Detailed File** – Detailed results for each enriched gene set/pathway: including the name of the gene set, the library/source, number of genes from your list that overlap with it, the total genes in that set, *p*-value, adjusted *p*-value, and other statistics.
- **Pathway Information File** – A summary of each significant enriched pathway/gene set (with its name and source library) and the list of node genes associated with that pathway that have been added to the network.
- **Summary File** – A compiled list of all unique new nodes added in this enrichment step, along with their source annotation.

Tip: Selecting MSigDB or Enrichr provides broad coverage of pathways and gene sets, which can be valuable for discovery. However, if your initial node list is large, consider narrowing the library selection or adjusting the gene set size filters. This helps maintain a more focused and interpretable network.

Note: For more information on the enrichment library files, please visit Enrichr and MSigDB databases.

3. Connectivity Mapping

In this step, AutoNetCan automatically defines regulatory connections (edges) between nodes in the constructed network. The platform integrates known regulatory interactions—such as activation or inhibition—by default, using curated data from established databases including TRRUST, INDRA, SIGNOR, and OmniPath (**Figure 10**).

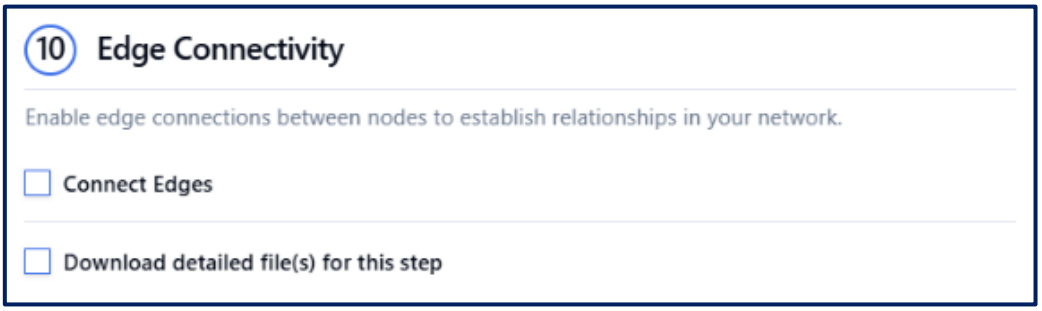

Figure 10: Edge Connectivity Option

Input Parameters

- **Connect Edges**: Boolean option (checkbox) that is permanently enabled by default and integrates known regulatory interactions among the input genes. All constructed networks automatically include predefined edges based on curated data from external databases.

Output

- **Mapped Edges File** – A comprehensive list of regulatory interactions added to the network. Each entry contains the source node, target node, interaction type (activation/inhibition), description (if available), and supporting source (e.g., database or publication reference).
- **Mapping Statistics** – A summary table that reports the total number of edges mapped from each source database and overall edge count statistics.
- **Summary File** – A node-level summary that provides connectivity statistics such as the number of incoming and outgoing edges (in-degree and out-degree) per node. These metrics facilitate the identification of hub genes or key regulators within the network.

4. Omics Data-based Logical Rules Modeling

In this step, AutoNetCan allows users to infer logical rules that describe how each node in the network is regulated based on expression data. Two modeling approaches are available (**Figure 11**), and the user must select only one for the analysis:

- **Expectation-Maximization (EM)** – A probabilistic modeling approach recommended for nodes with a relatively small number of regulators (≤15).
- **Boolean Rules** – A deterministic method based on binary expression patterns using truth tables, applicable regardless of network complexity.

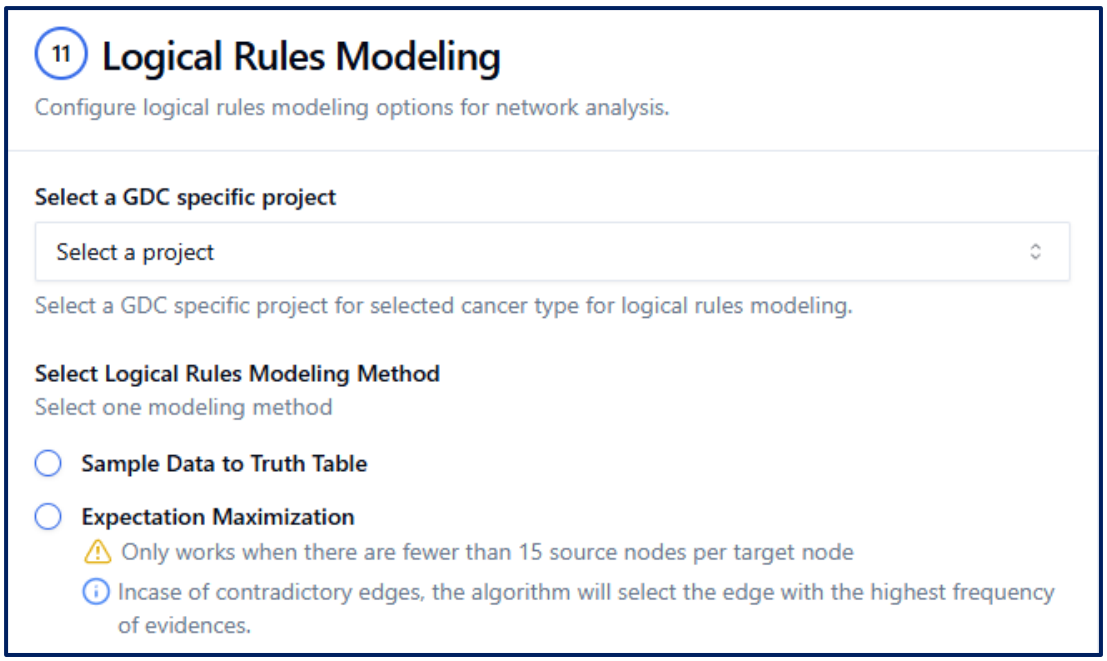

Figure 11: Boolean Modeling Option

After selecting a modeling approach, users must specify a cancer project from which RNA-Seq expression data will be used to infer the logical rules. Only one project can be selected per run. Click *"Next"* to initiate the modeling process.

Note: The EM-based method may require longer computation times when nodes have many regulators. To maintain performance and reliability, AutoNetCan limits EM inference to nodes with approximately 15 or fewer regulators.

Upon successful completion of the modeling process, users will receive an email containing download links for all output files, along with a dedicated link to visualize the network directly in the AutoNetCan Canvas interface. This web-based canvas provides an interactive environment to explore node connectivity, logic rules, and edge types within the inferred network.

Input Parameters

- **Logical Modeling Method:** Select one of the following options for rule inference:
    - *Expectation-Maximization (Probabilistic)* – Learns logic rules by fitting a probabilistic model to RNA expression data. Interaction types (e.g., activation, inhibition) from databases are used. In cases of contradictory annotations, the interaction type with the highest frequency is selected. If frequencies are equal, the type is chosen randomly.
    - *Boolean (Deterministic)* – Uses a truth-table–based approach to infer logic rules from binarized expression patterns. Interaction types from databases are not used in this method.
- **Expression Data (Project):** Select a single cancer cohort (e.g., TCGA-BRCA) to serve as the input expression dataset for rule inference.

Output

- **Network Rules File (Cytoscape format):** A complete network with nodes, edges, and inferred logical rules, compatible with Cytoscape for visualization and statistical analysis.
- **Network Rules File (TISON format):** A file containing the same network and rules, formatted for simulation and perturbation analysis in the TISON platform.
- **Network Object (JSON format):** A structured JSON file containing the entire network and rule set for advanced integration or programmatic processing.

Note: After downloading output files, users may visualize and explore the network using:

- AutoNetCan Canvas (link included in the result email)
- Cytoscape for network editing and advanced graph analysis
- TISON for dynamic simulation and perturbation modeling

Detailed steps for loading and interacting with the network in Cytoscape and TISON are provided in the **Section 6**.

5. Final Review and Submission

As one of the final steps in the workflow, you will be asked to provide your email address within the form (**Figure 12**). AutoNetCan will use this to notify you when your network construction is complete and send a link to download the results.

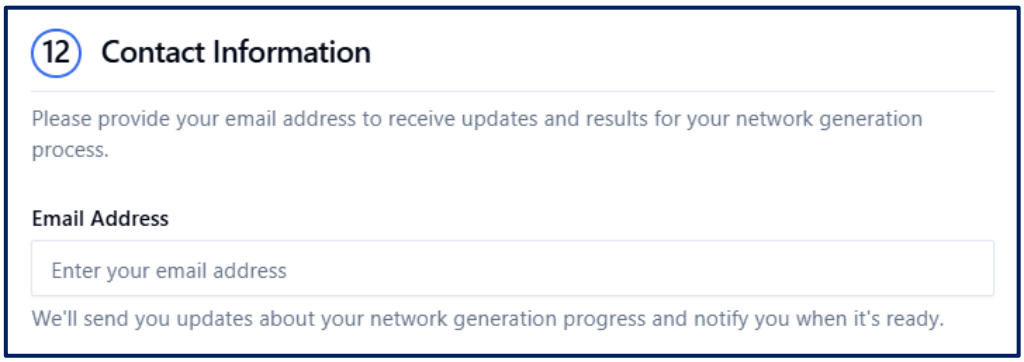

Figure 12: Contact Information

After providing your email, click "*Review*" to continue. You will then see a preview page that summarizes all your selections from each step — including cancer type, thresholds, selected categories, and enrichment settings (**Figure 13**).

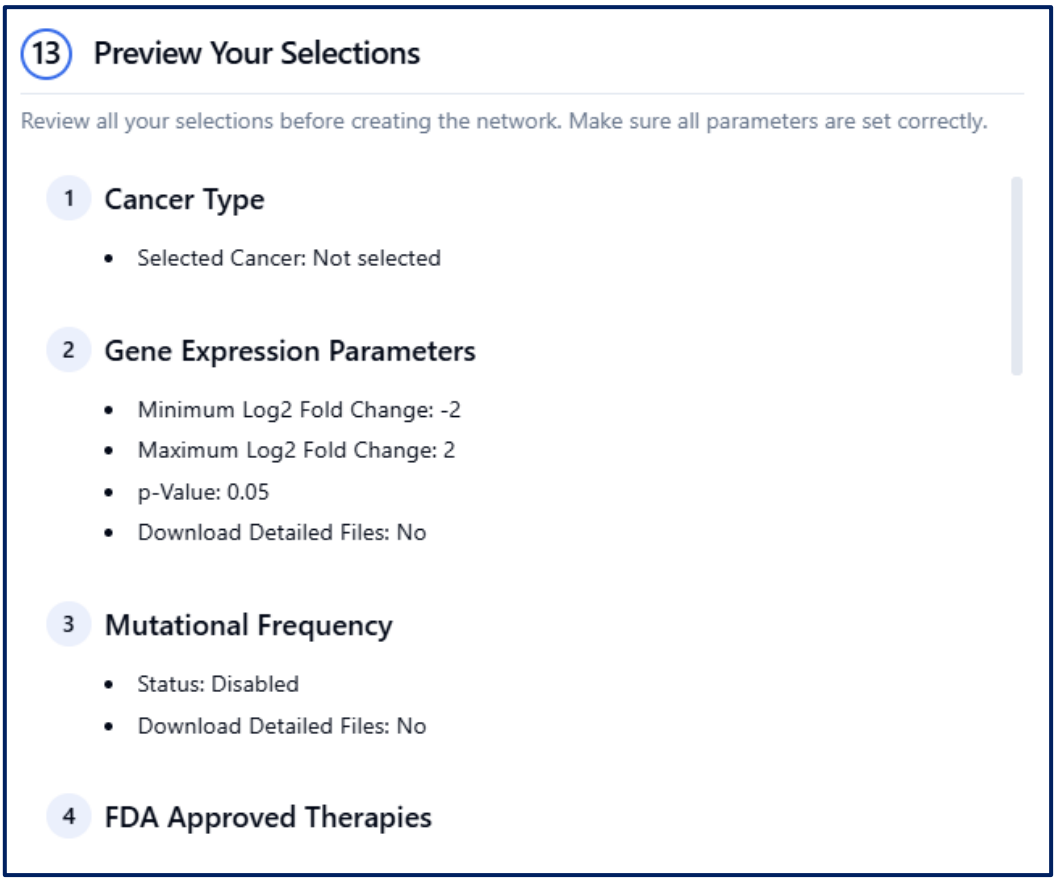

Figure 13: Preview Your Selections

Carefully review all parameters, if something needs to be changed, go back and edit it. Once everything looks correct, click "*Submit*" to start the network construction job. AutoNetCan will begin processing and notify you by email once your results are ready.

Each job-completion e-mail contains a download link to a zipped results package. Inside, separate sub-folders correspond to every analysis step selected on the canvas, including (where applicable): differential-expression tables, mutation-frequency summaries, FDA-approved and investigational drug matches, DNA-repair pathway outputs, cancer-signature gene sets, node-mapping and enrichment reports, edge-connectivity files, Boolean rule models, and global / node-level topology metrics. The message also provides a direct link for interactive network visualisation on the AutoNetCan canvas.

Warning: Please download the package within 24 hours; the download and visualization links will expire automatically thereafter.

Note: Node-mapping and topology-metric steps are executed with default settings and do not require user-defined parameters.

6. Network Visualization

The constructed network can be visualized using several platforms, including AutoNetCan, Cytoscape, and TISON.

6.1 Network Visualization in AutoNetCan

After the network construction is complete, users receive an email notification. This email includes a link to download the results via the "*Download Results*" option, as well as a "*Visualize Network*" button that redirects to the AutoNetCan page (**Figure 14**).

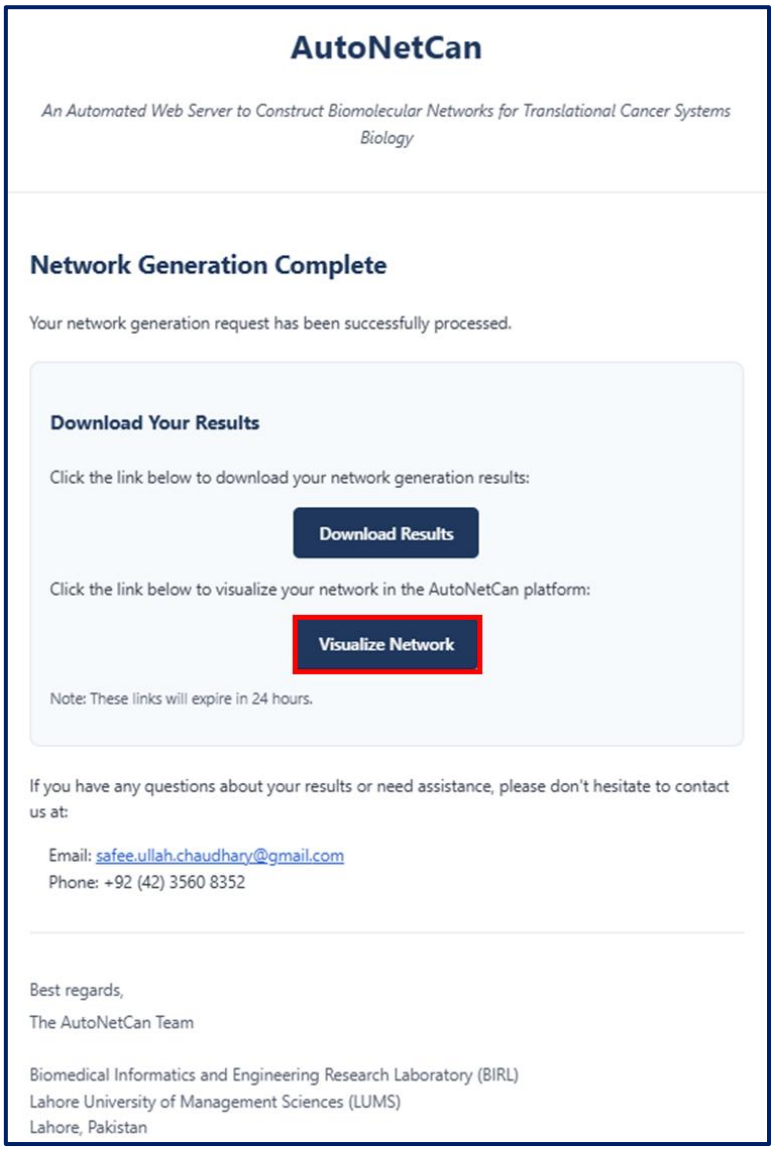

Figure 14: Email notification containing links to download result files and visualize the network.

On the AutoNetCan platform, the network can be viewed using various layout options such as circular, grid, or force-directed layouts (**Figure 15**). Users can interact with the network by searching for specific nodes through the search bar or by clicking directly on nodes to explore their incoming and outgoing interactions (**Figure 16**).

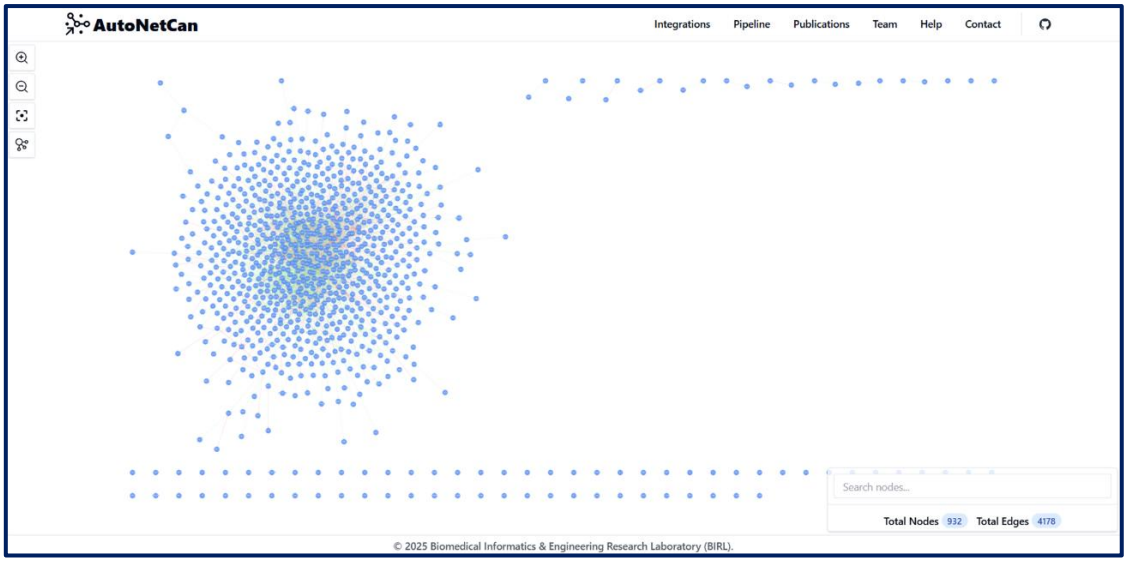

Figure 15: Network visualization interface in AutoNetCan.

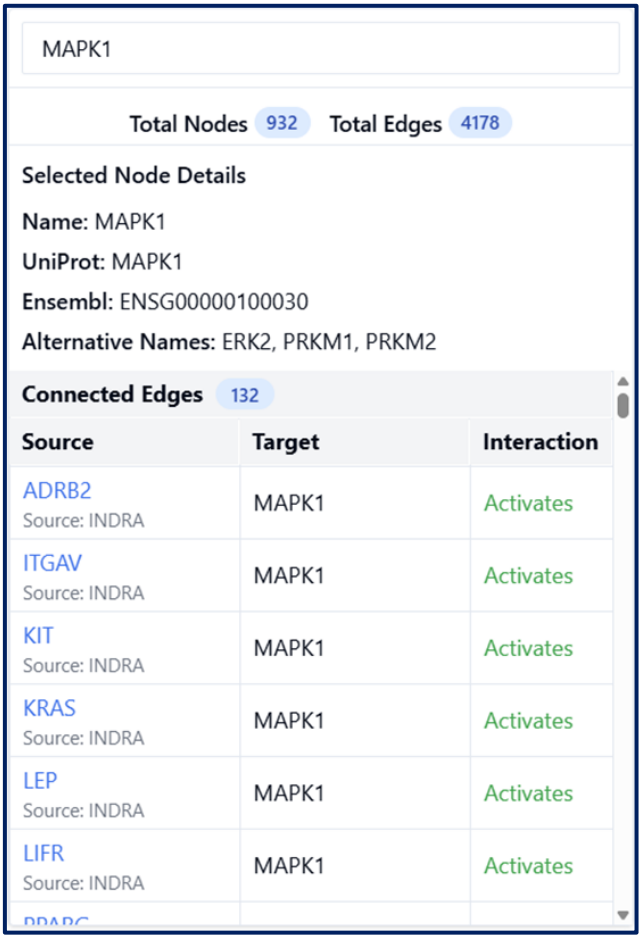

Figure 16: Visualization of the node search feature in AutoNetCan, highlighting MAPK1 with 132 total incoming and outgoing interactions.

6.2 Network Visualization in Cytoscape

AutoNetCan generates a .csv file with a structured format that can be seamlessly imported into Cytoscape for network visualization. Follow these steps:

1. Launch Cytoscape (desktop version 3.10.3). Go to File → Import → Network from File (**Figure 17**).
2. In the Network file to load window, locate and select the .csv file generated by AutoNetCan (**Figure 18**).
3. The Import Network from Table window will open. Column headers will be automatically mapped to:
    - Source Node
    - Target Node
    - Interaction Type
    - Edge Attribute
    - Edge Color

    Click OK to import the network (**Figure 19**).
4. Once the network is loaded (**Figure 20**), go to the Style tab, then select the Edge sub-tab (**Figure 21**).
5. Under Stroke Color, click the dropdown arrow, set the column to Edge Color, and choose Passthrough Mapping as the mapping type (**Figure 22**).
6. For Target Arrow Shape, click the dropdown, select Edge Color as the column, and again choose Passthrough Mapping (**Figure 23**).
7. The network will now appear on the canvas with directed and color-coded edges representing the type and attributes of interactions (**Figure 24**).

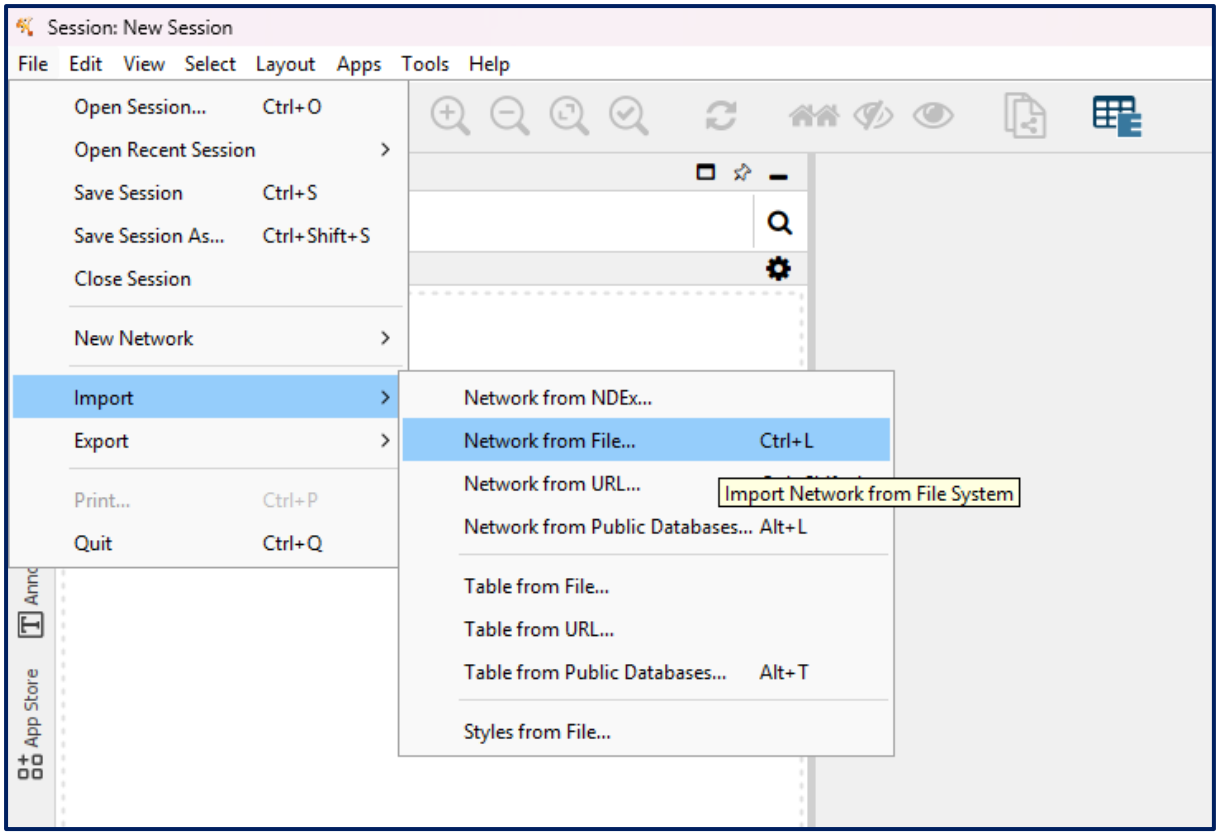

Figure 17: Import network in Cytoscape desktop window through "Network from File" option.

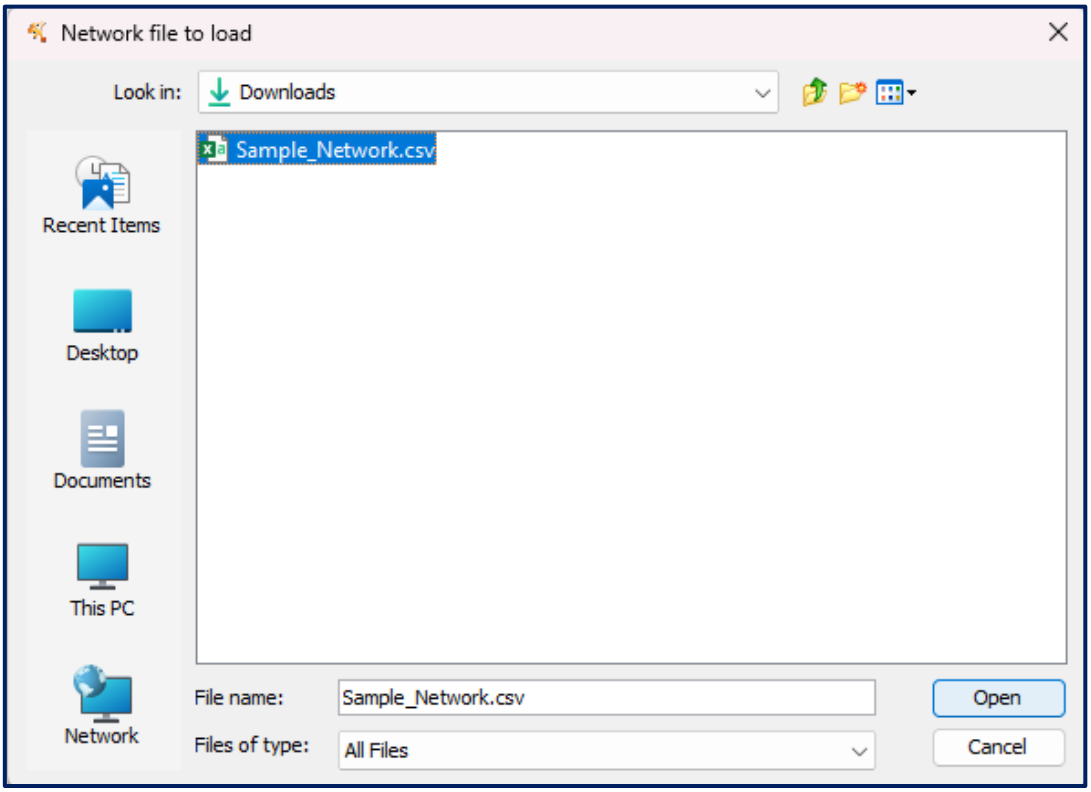

Figure 18: Sample Network upload window

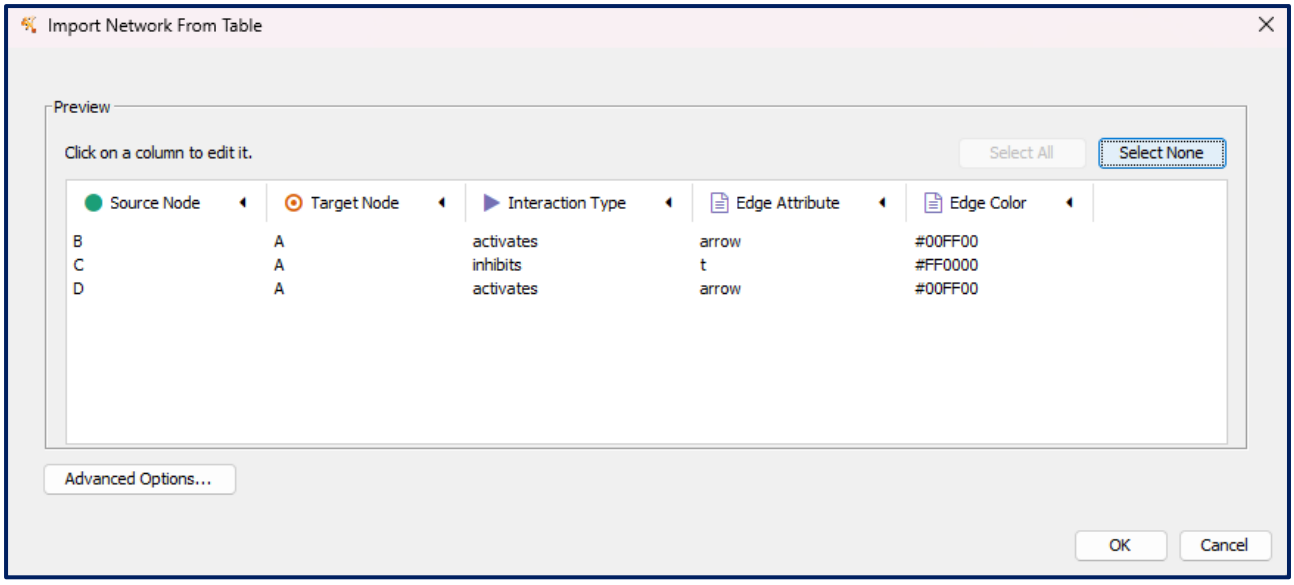

Figure 19: Import Network from Table window and assignment of network variables.

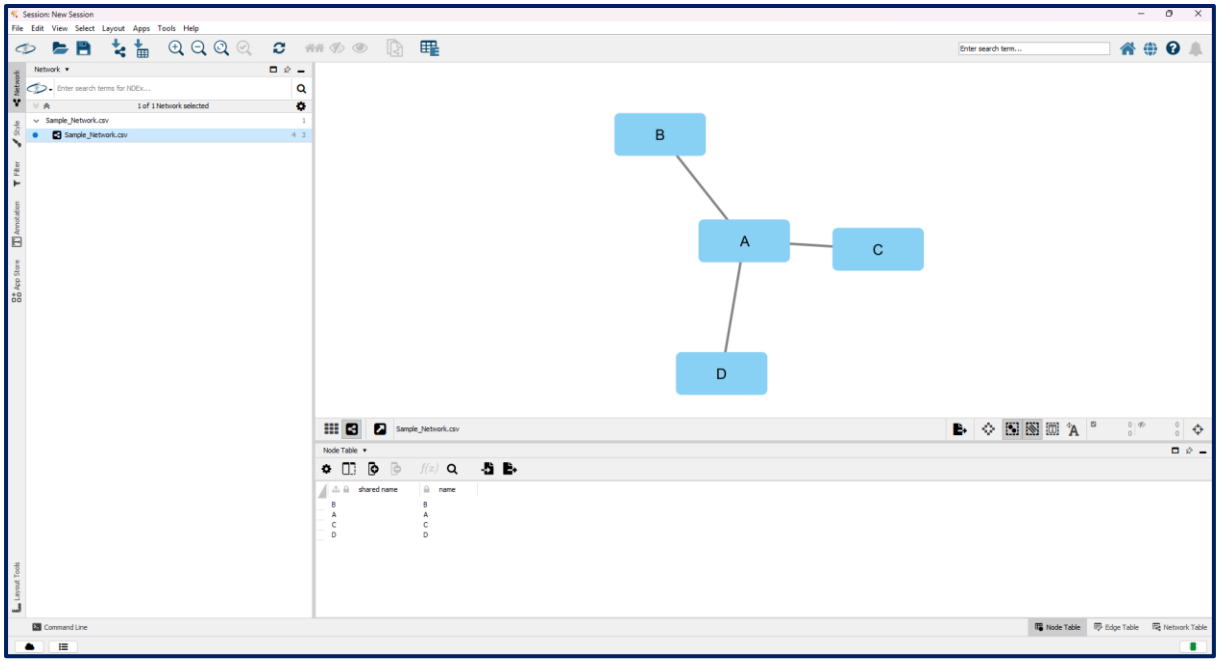

Figure 20: Visualization of nodes and undirected edges in Cytoscape.

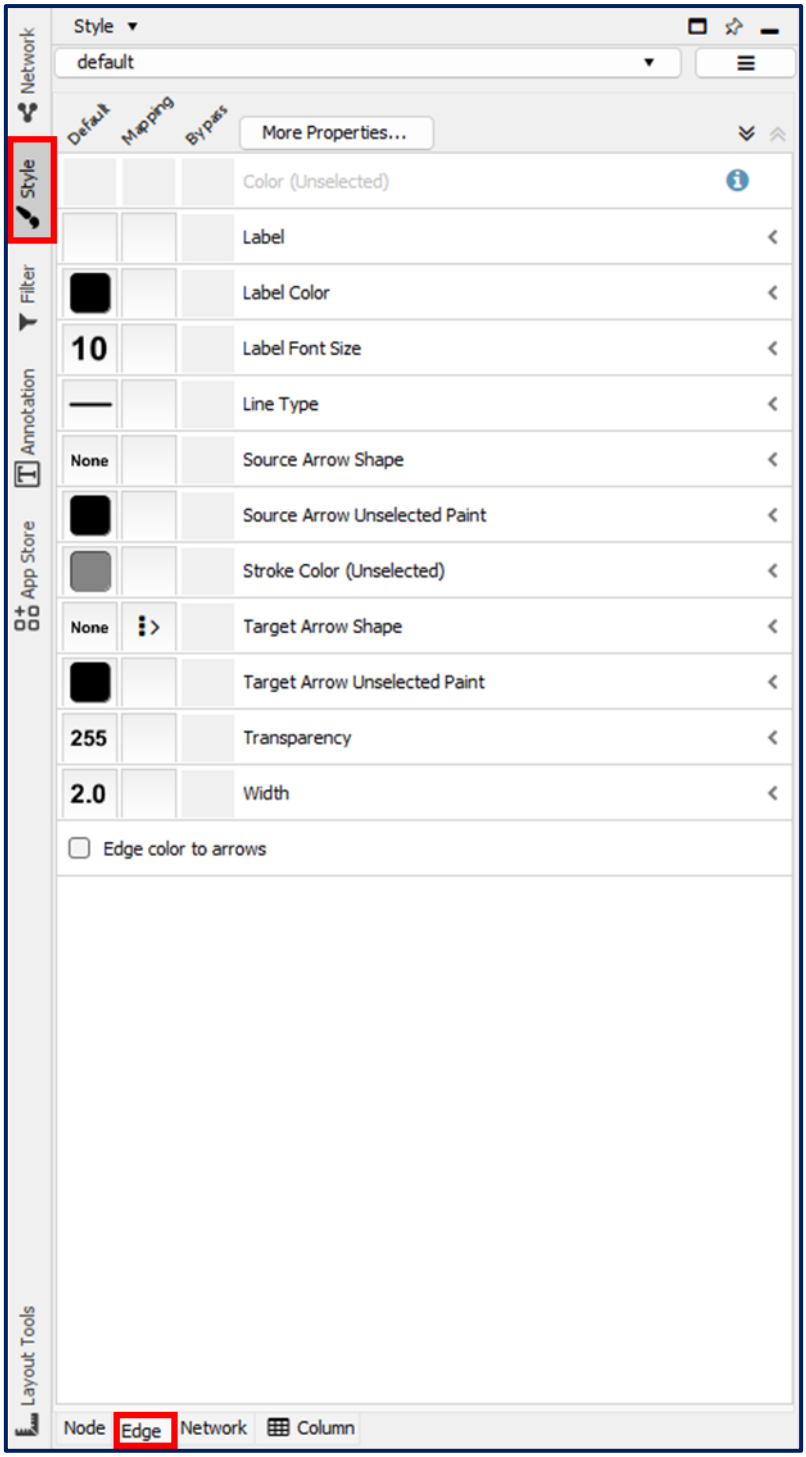

Figure 21: Network tab and Edge sub-tab on the left panel of window.

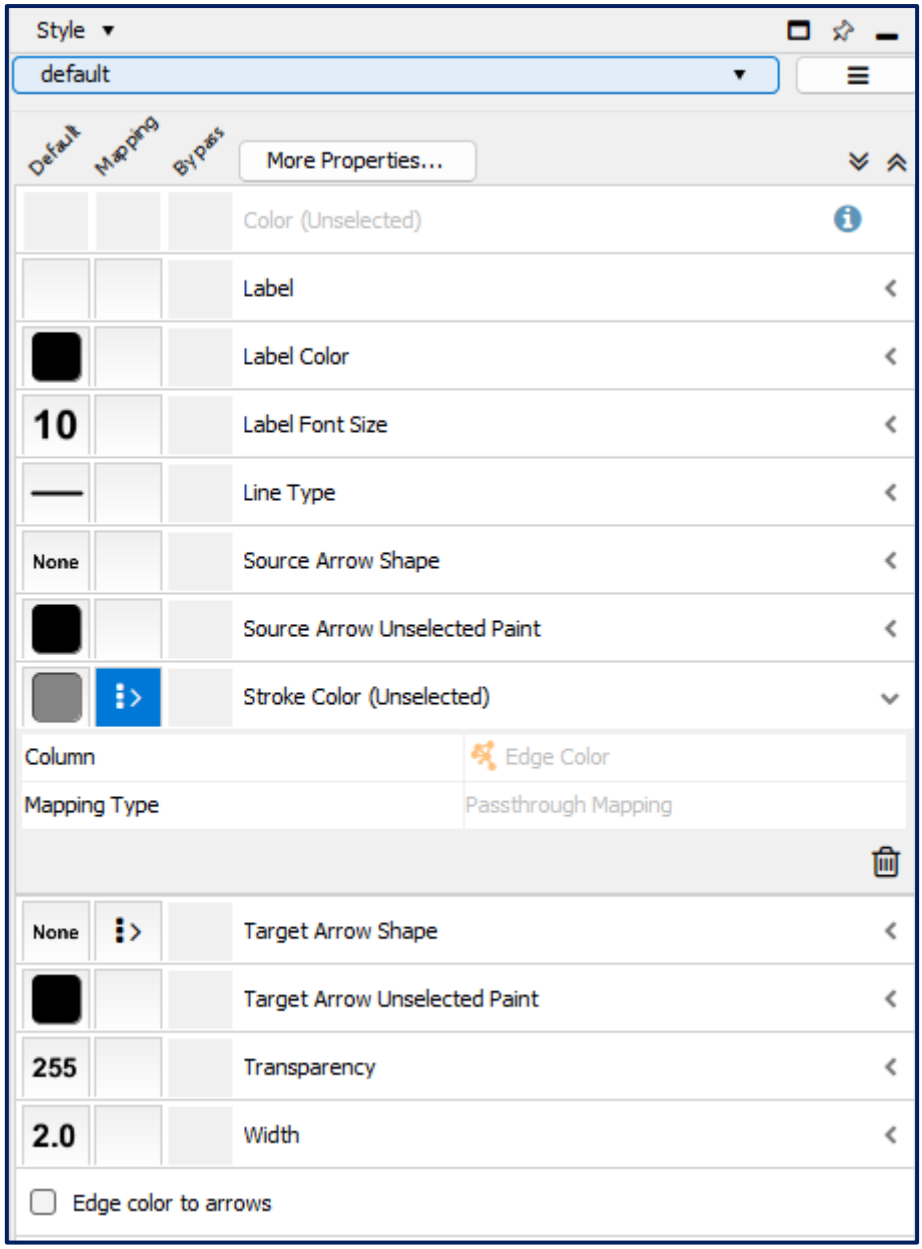

Figure 22: Assignment of edge color

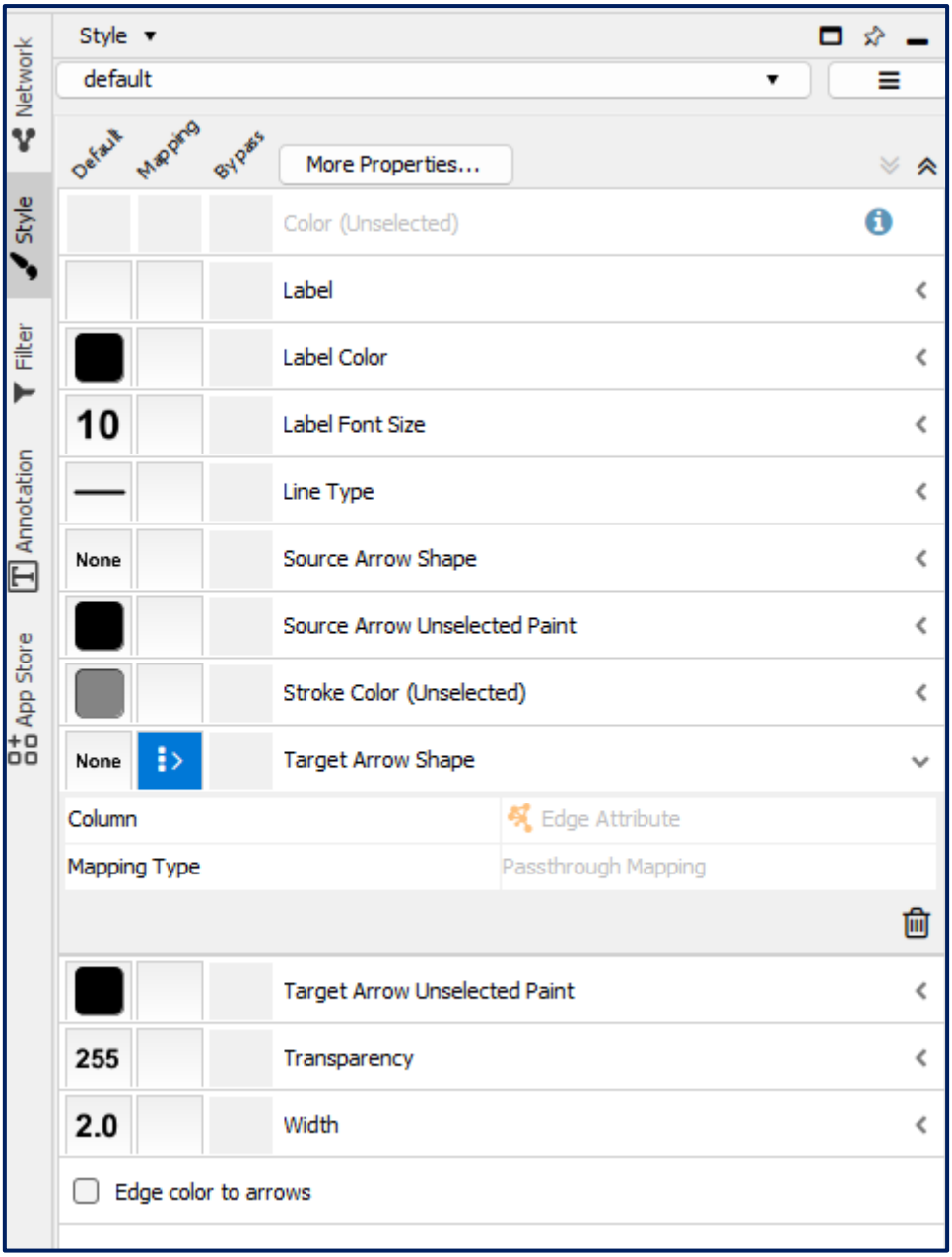

Figure 23: Assignment of target arrow shape

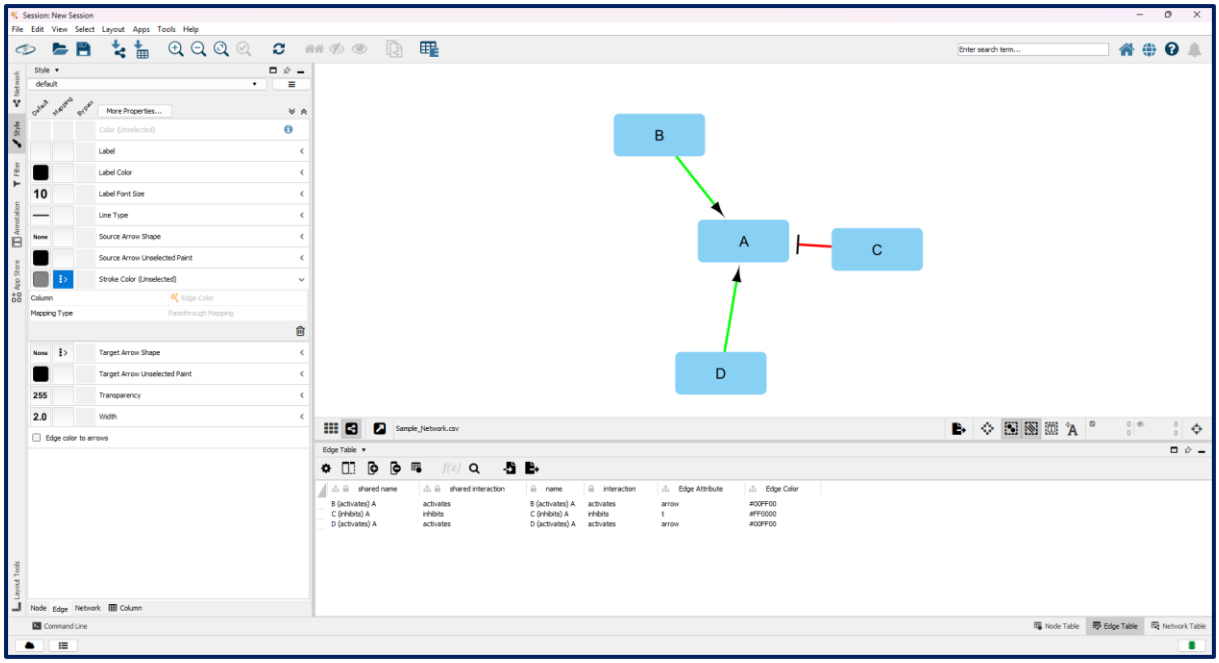

Figure 24: Network visualization with directed and colored edges in Cytoscape.

6.3 Network Visualization in TISON

AutoNetCan generates .txt file which can be uploaded in TISON by following these simple steps:

1. Visit https://tison.lums.edu.pk and click on "*Project Explorer*" (**Figure 25**).
2. Click "*Create a New Project*" and assign a name (e.g., CaseStudy) (**Figure 26** - **Figure 28**).
3. Open the newly created project and navigate to the "*Networks*" editor (**Figure 29**).
4. Click the "*Create Network*" button (the **'+'** icon) (**Figure 30**).
5. In the pop-up window, select "*Write Rules*" (**Figure 31**).
6. Paste the .txt file generated by AutoNetCan's logical modeling step and click "*Create Network*" (**Figure 32**).
7. Network with nodes and edge count visible on the canvas in TISON (**Figure 33**).

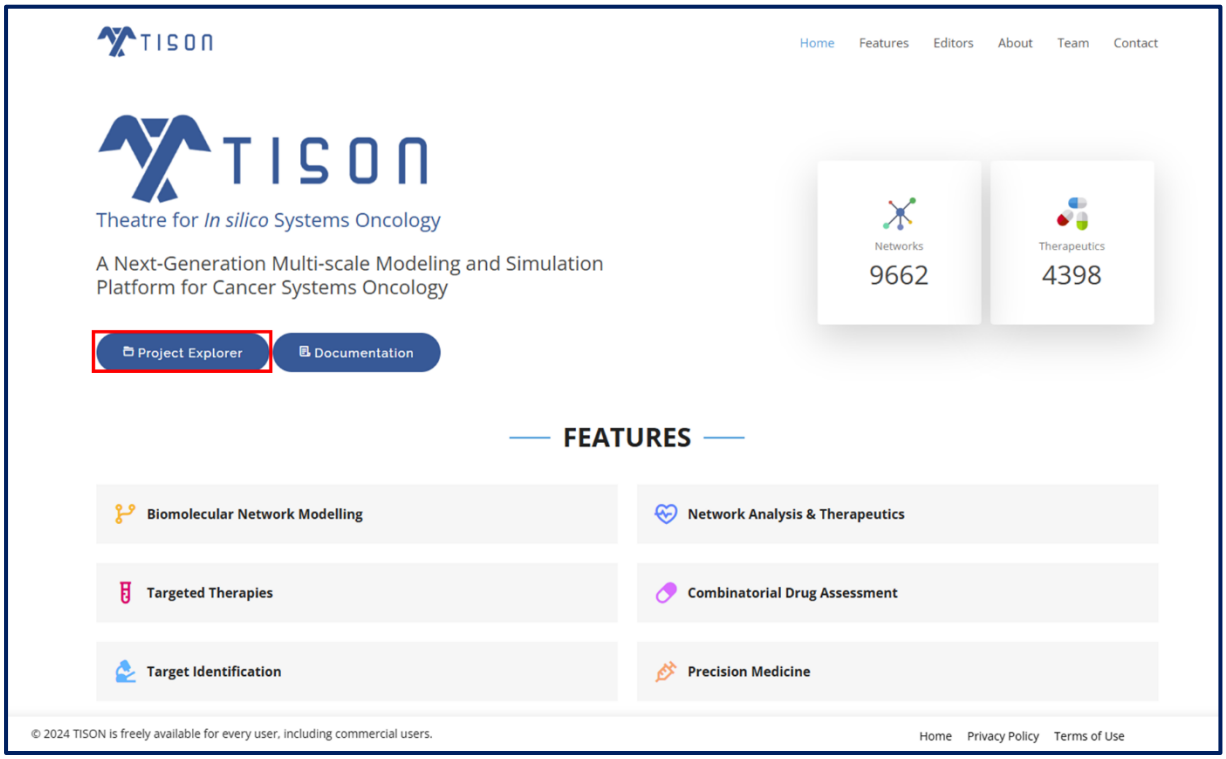

Figure 25: TISON homepage displaying the Project Explorer tab used to create projects.

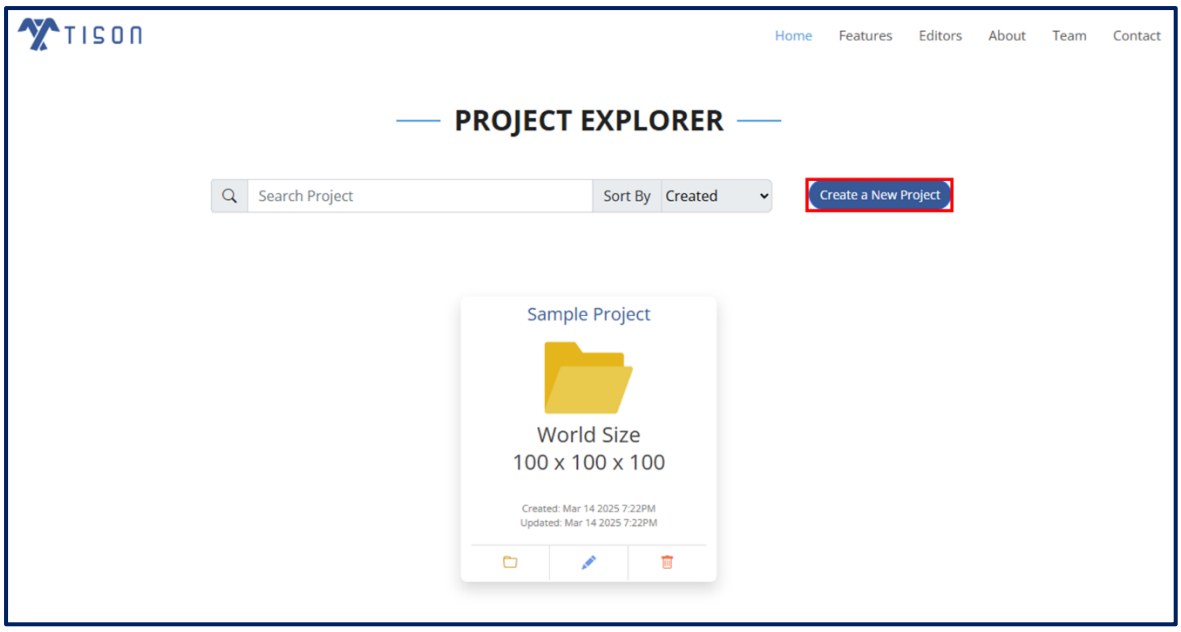

Figure 26: Create a New Project tab in Project Explorer

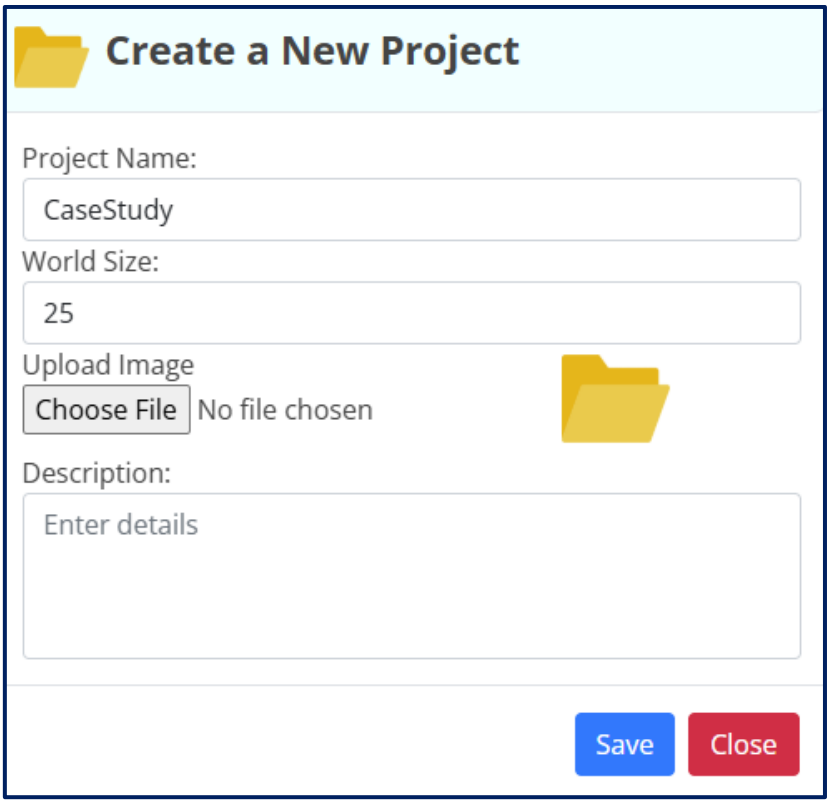

Figure 27: Details required by TISON for creating a new project

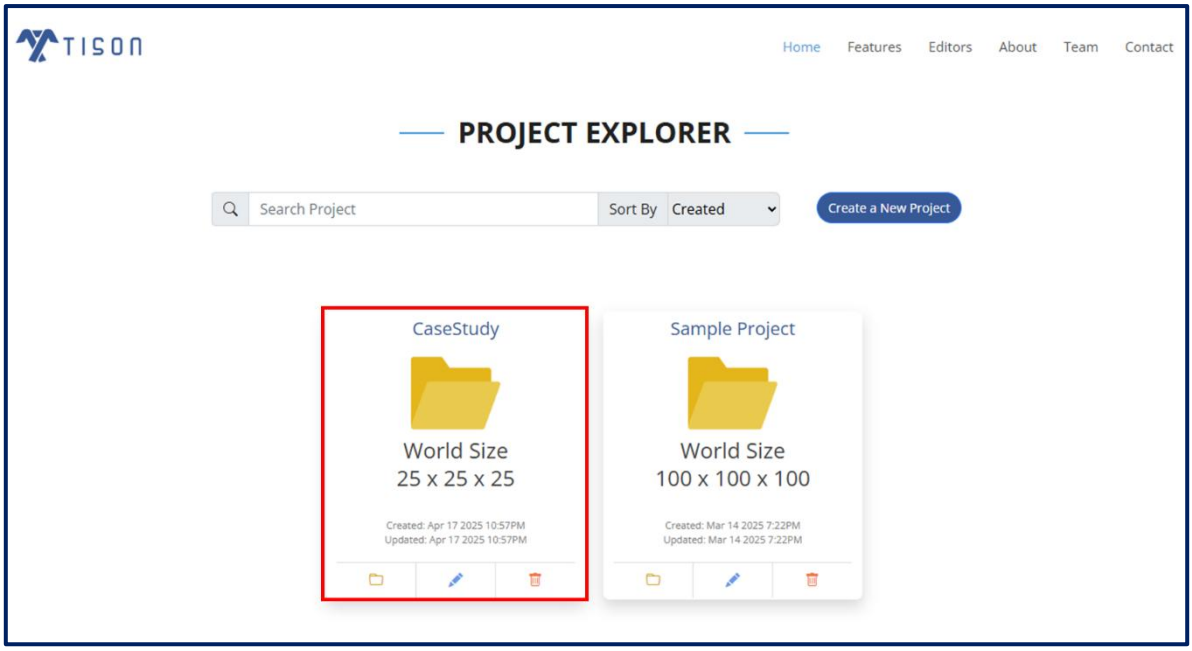

Figure 28: Case Study Project

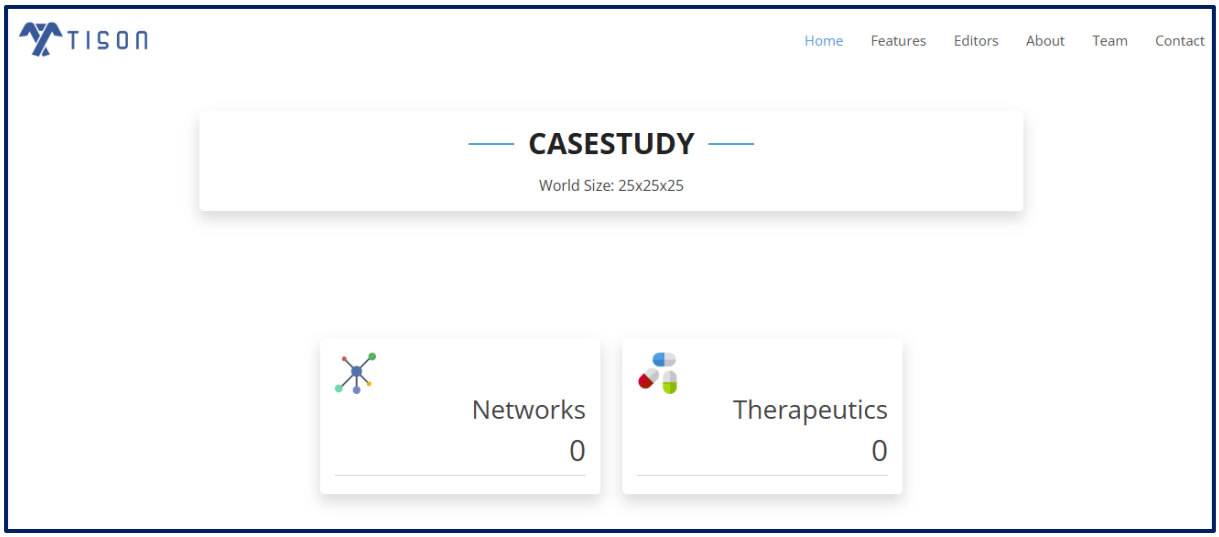

Figure 29: Navigate to Networks Editor

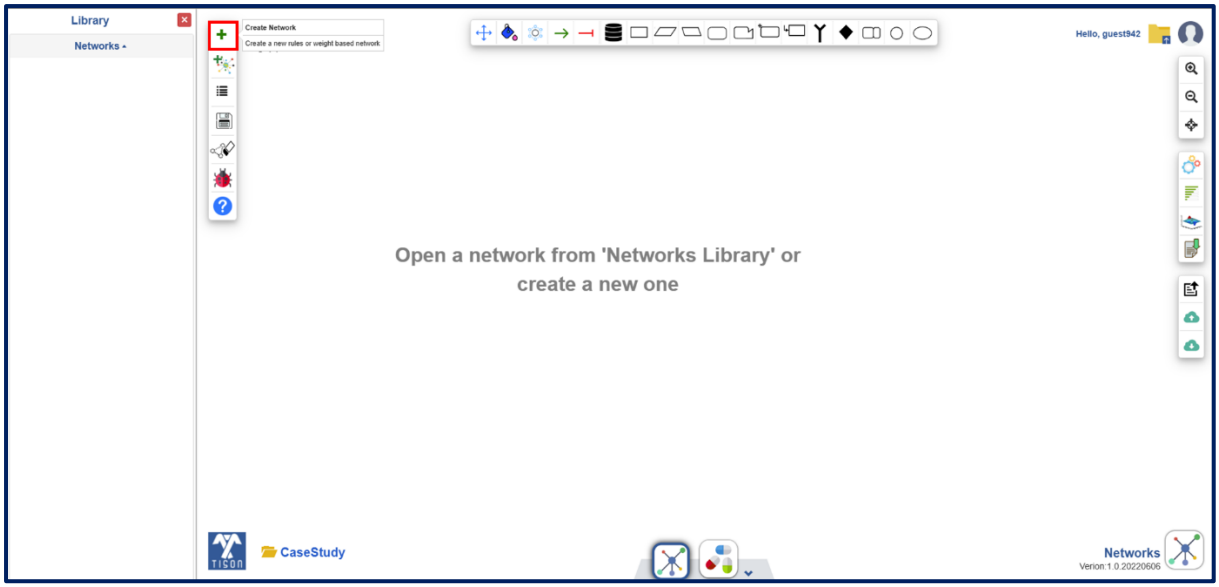

Figure 30: Click the "Create Network" button

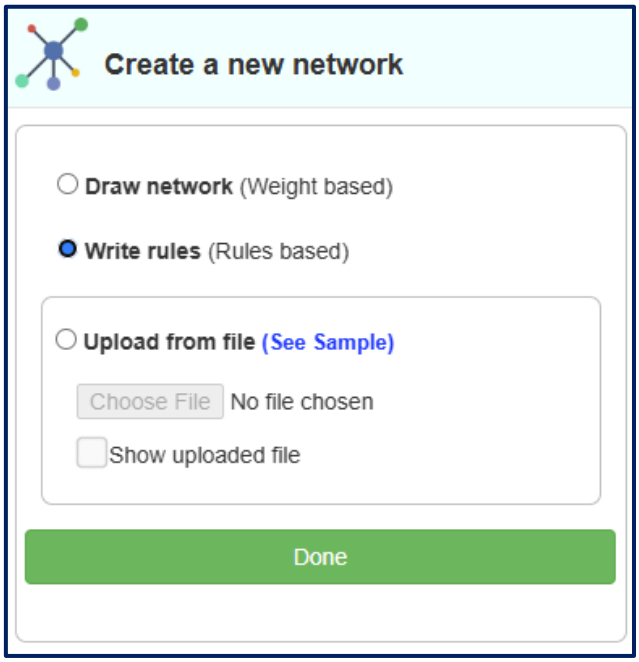

Figure 31: Select "Write rules" option to create network

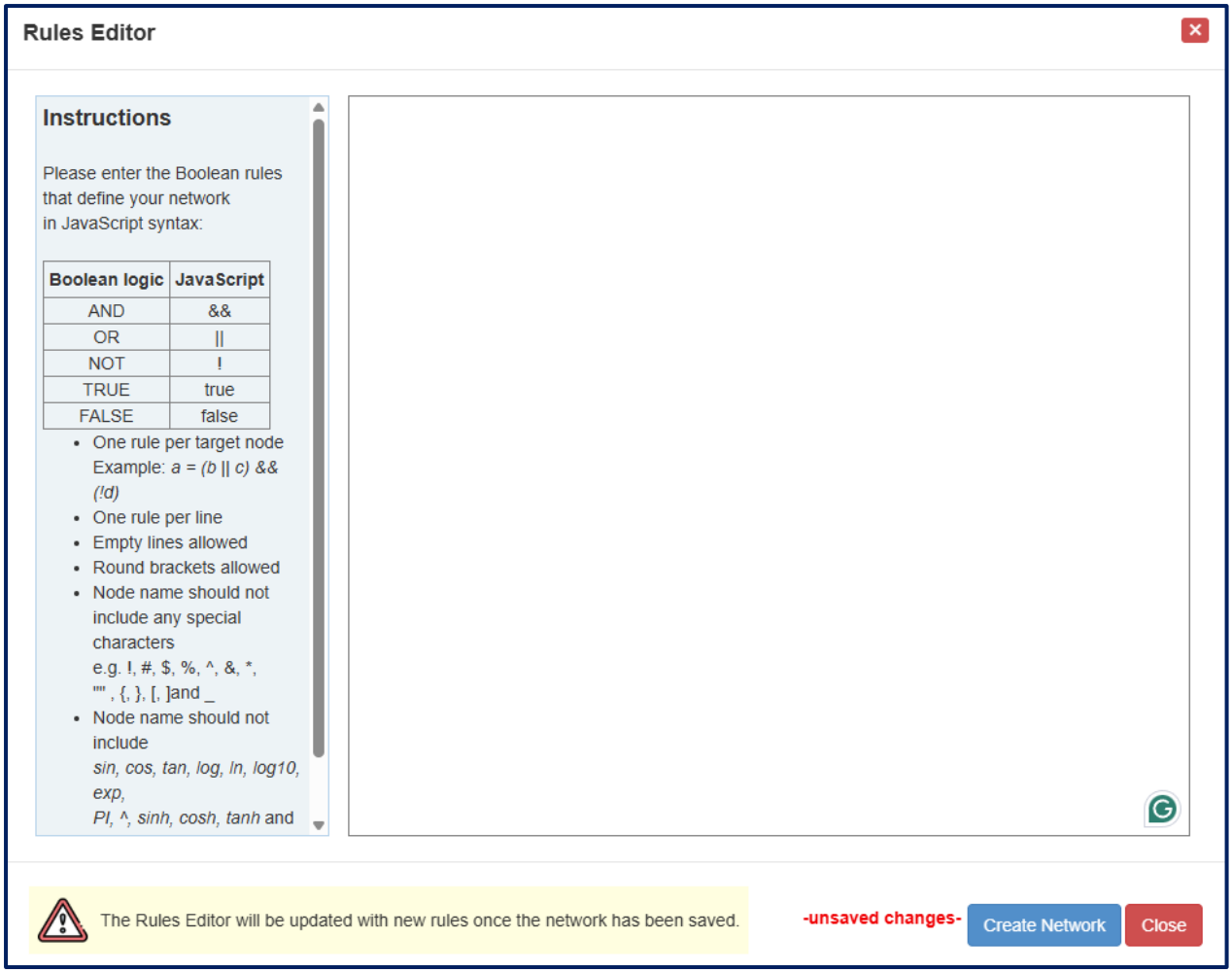

Figure 32: Rules editor window

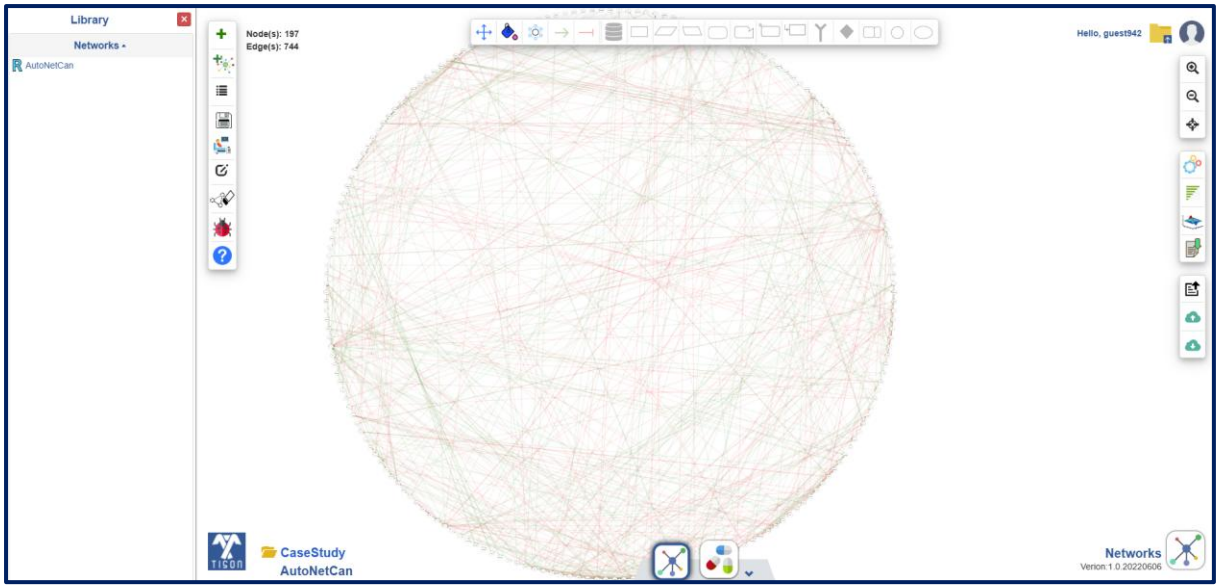

Figure 33: AutoNetCan generated network visualization in TISON