BioNet Compare

User Manual & Research-oriented Documentation

BioNet Compare is a Streamlit-based application for comparative analysis of multiple biological networks. It supports common exchange formats (edge lists, SIF, and GraphML) and provides a reproducible workflow for quantifying network-level topology, node-level centrality, and overlap across networks.

Version: 1.1 • Document date: 15 Feb 2026

What it does	Compares two or more networks using global metrics, local (node-level) metrics, node/edge overlap (Jaccard; Venn for 2–3 networks), and distance to the union network.
Who it is for	Biologists and bioinformaticians who need fast, interpretable network comparisons for exploratory analysis, QC, method benchmarking, or multi-condition studies.
Supported inputs	Edge lists (.txt/.tsv/.csv), SIF (.sif), GraphML (.graphml/.xml). Optional directionality and edge weights.
Outputs	Sortable tables, interactive plots, and one-click exports (CSV/TSV/HTML/PNG/JPG/PDF/SVG).

Note: To avoid StreamlitDuplicateElementId errors, all download buttons in the app should include a unique key.

Contents

1. Scientific motivation and scope

Biological systems are often best represented as networks: proteins interact to form complexes, genes regulate other genes, metabolites participate in biochemical reactions, and cells communicate through signaling pathways. Across experimental conditions (e.g., disease vs control), across data sources (e.g., curated pathways vs inferred networks), or across methods (e.g., different network inference algorithms), researchers frequently need to quantify how network structure changes and which nodes become more central. BioNet Compare provides a consistent, repeatable framework to: (i) summarize network topology with global metrics, (ii) identify key nodes using local centrality measures, (iii) quantify overlap of nodes and edges, and (iv) contextualize each network relative to a combined union reference.

How this helps biologists

- Quality control: detect unexpected sparsity/density, fragmentation, or changes in connectedness between runs.
- Mechanistic insight: identify hubs/bottlenecks that may represent regulators, essential proteins, or pathway cross-talk points.
- Condition comparison: quantify rewiring by comparing overlap and distance across conditions or time points.
- Method benchmarking: compare inferred networks against curated references, or across inference pipelines.

2. Typical use cases in network biology

Differential network analysis: Compare networks built from transcriptomics in disease vs control to identify changes in hub genes, bottlenecks, and overall connectedness.

Pathway curation vs inference: Compare a curated pathway network to an inferred network to quantify coverage (node/edge overlap) and identify missing or spurious interactions.

Multi-omics integration: Compare networks derived from proteomics vs phosphoproteomics to assess whether signaling edges appear as expected and which nodes become central in each modality.

Time-course rewiring: Compare networks at successive time points (e.g., infection progression) to quantify gradual changes in overlap and distance from the union.

3. Installation & running the app

BioNet Compare is distributed as a Streamlit project folder. Installation is one-time per computer (or per environment).

macOS / Linux (recommended):

```
bash install.sh source .venv/bin/activate streamlit run app.py
```

Windows (PowerShell):

```
.\install_windows.ps1 .\.venv\Scripts\Activate.ps1 streamlit run app.py
```

Updating:

If you receive an updated app.py (e.g., bug fixes), you can typically replace only app.py and restart Streamlit. Re-run installation only if requirements.txt changes.

4. Input formats and preprocessing assumptions

BioNet Compare supports the following formats. Node identifiers are treated as strings; ensure consistent naming across files (e.g., gene symbols vs Ensembl IDs).

4.1 Edge lists (.txt / .tsv / .csv)

Minimum: two columns (source, target). Optional weight column. Choose delimiter in the UI.

```
sourcetarget sourcetargetweight
```

4.2 SIF (.sif)

Each line: source interaction target1 target2 ... (expanded into edges). Weights are typically absent and treated as 1.0 if enabled.

```
TP53 pp MDM2 TP53 pp BAX BBC3
```

4.3 GraphML (.graphml / .xml)

GraphML may include node/edge attributes. If “Has weight” is enabled, the app reads edge attribute “weight” when present.

Preprocessing assumptions

- Undirected edge sets are normalized so that (A,B) equals (B,A) for overlap calculations.
- Some metrics (e.g., clustering, connected components) are computed on the undirected projection for comparability.
- Missing weights (if enabled) are treated as 1.0.

5. Workflow on the landing page

Upload networks: Upload one or more files. Each file becomes a network entry.

Configure: Optionally rename networks; set directed/undirected and weight parsing; choose delimiter for edge lists.

Sanity check: Parse and validate each file. Networks with parsing errors are disabled for analysis.

Select networks: Choose the subset to include in the comparison.

Perform analysis: Precompute all results for the four modules with progress bars and an analysis log.

6. Module 1: Global property comparison

This module summarizes network topology using global properties. These measures help evaluate overall network organization and can serve as QC indicators (e.g., unexpected density shifts) or biological signals (e.g., increased connectedness due to pathway activation).

Key metrics and biological interpretation

Number of nodes/edges: Reflects coverage and interaction richness; changes may indicate filtering thresholds or condition-specific interactions.

Density: Higher density can reflect more interactions per node (e.g., broad co-expression), but can also indicate noise in inferred networks.

Connected components / largest component fraction: Fragmentation can indicate modularity, incomplete coverage, or disconnected pathway sets.

Average clustering: High clustering suggests triadic closure and local modular structure, common in functional modules or protein complexes.

Outputs

A sortable table and grouped bar/radar visualizations. Export tables and figures for supplementary materials.

7. Module 2: Local property comparison

This module computes node-level properties (centrality and clustering) for each network, then aligns them across networks using the union set of nodes. It supports discovery of condition-specific key nodes and comparison of node rankings across networks.

Centrality measures in biological context

Degree / hubness: High-degree nodes often correspond to hubs (e.g., essential proteins, master regulators) but can also reflect annotation bias in curated networks.

Betweenness (bottlenecks): Nodes that lie on many shortest paths can mediate information flow between modules (e.g., pathway cross-talk points).

Closeness: Nodes with short average distance to others may be efficient broadcasters/receivers of signals in connected networks.

Eigenvector centrality: Highlights nodes connected to other highly connected nodes; useful for identifying influential nodes in dense cores.

PageRank (directed): Models signal propagation or influence in directed graphs; can help rank upstream regulators in regulatory networks.

Top-k and search workflows

For each property within each network, the app highlights the top-k nodes (default k=10). Use the search box (regex supported) to focus on gene families, pathways, or user-defined lists. The heatmap summarizes patterns across networks and properties.

8. Module 3: Node & edge overlap (Jaccard / Venn)

This module quantifies similarity between networks by comparing node sets and edge sets. Node overlap reflects shared entities (genes/proteins), while edge overlap reflects shared relationships (interactions/regulations).

Biological utility

- Assess reproducibility: compare networks inferred from replicates or alternative pipelines.
- Compare coverage: curated reference vs inferred networks, or pathway databases vs literature-derived interactions.
- Condition-specific rewiring: edges may change even when nodes overlap strongly.

Interpretation tips

High node Jaccard with low edge Jaccard is a common signature of network rewiring: the same genes are present, but the inferred interactions differ. Conversely, high edge overlap suggests stable interaction structure.

9. Module 4: Comparison with respect to the union network

This module builds a union network (all nodes and edges observed across the selected networks) and then quantifies each network's distance to the union using node and edge Jaccard distances. The union acts as a reference "superset" summarizing the combined evidence.

Distance definitions

```
node_distance = 1 − Jaccard(nodes(N), nodes(U)) edge_distance = 1 − Jaccard(edges(N), edges(U))
```

Biological utility

- Identify outlier networks: a network far from the union may reflect distinct condition biology or data quality issues.
- Summarize multi-condition studies: distances can quantify progression over time or treatment response.
- Contextualize reference coverage: curated networks close to the union may better capture consensus interactions.

Visualization

Radar plots display node and edge distances to the union across networks, enabling quick comparison across multiple conditions.

10. Exporting results for reports and publications

Tables can be exported as CSV/TSV for further analysis or inclusion as supplementary data. Plots can be exported as interactive HTML or static images (PNG/JPG), vector graphics (SVG), or PDF for publication-quality figures.

11. Performance notes and reproducibility

Some centrality measures can be computationally expensive on large networks. If analysis is slow, reduce the number of local properties selected, analyze smaller subsets, or consider approximate methods. For reproducibility, record input files, preprocessing choices (direction/weights/delimiter), and selected networks.

12. Troubleshooting & FAQs

Q: Sanity check shows 0 nodes/edges
A: Verify delimiter and confirm each line has at least two columns.

Q: Weights are incorrect
A: Confirm weight column index (0-based) and numeric formatting.

Q: PNG/PDF export fails
A: Ensure kaleido is installed in the environment.

Q: A tab shows no output
A: Check for earlier exceptions; Streamlit stops rendering after an error.

13. Glossary of network biology terms (contextual)

Node: An entity in the network (e.g., gene, protein, metabolite, cell type).

Edge: A relationship between two nodes (e.g., interaction, regulation, co-expression, reaction).

Directed edge: An asymmetric relationship (TF → gene, kinase → substrate).

Undirected edge: A symmetric relationship (physical interaction, correlation).

Hub: A node with high degree; often essential or multifunctional, but may also reflect annotation bias.

Bottleneck: A node with high betweenness centrality; may connect functional modules or mediate cross-talk.

Module / community: A set of nodes with dense internal connections, often corresponding to pathways or complexes.

Clustering coefficient: Quantifies triadic closure around a node; high values often indicate locally cohesive functional groups.

Connected component: A subset where each node is reachable from others (undirected sense). Many components indicate fragmentation.

Network rewiring: A change in edges/relationships across conditions even when nodes remain largely the same.

Jaccard index: Set similarity measure used here to quantify overlap of nodes or edges.

Union network: A combined network containing all observed nodes and edges across selected networks.

End of document.