String_vs_sgd_vs_tflink

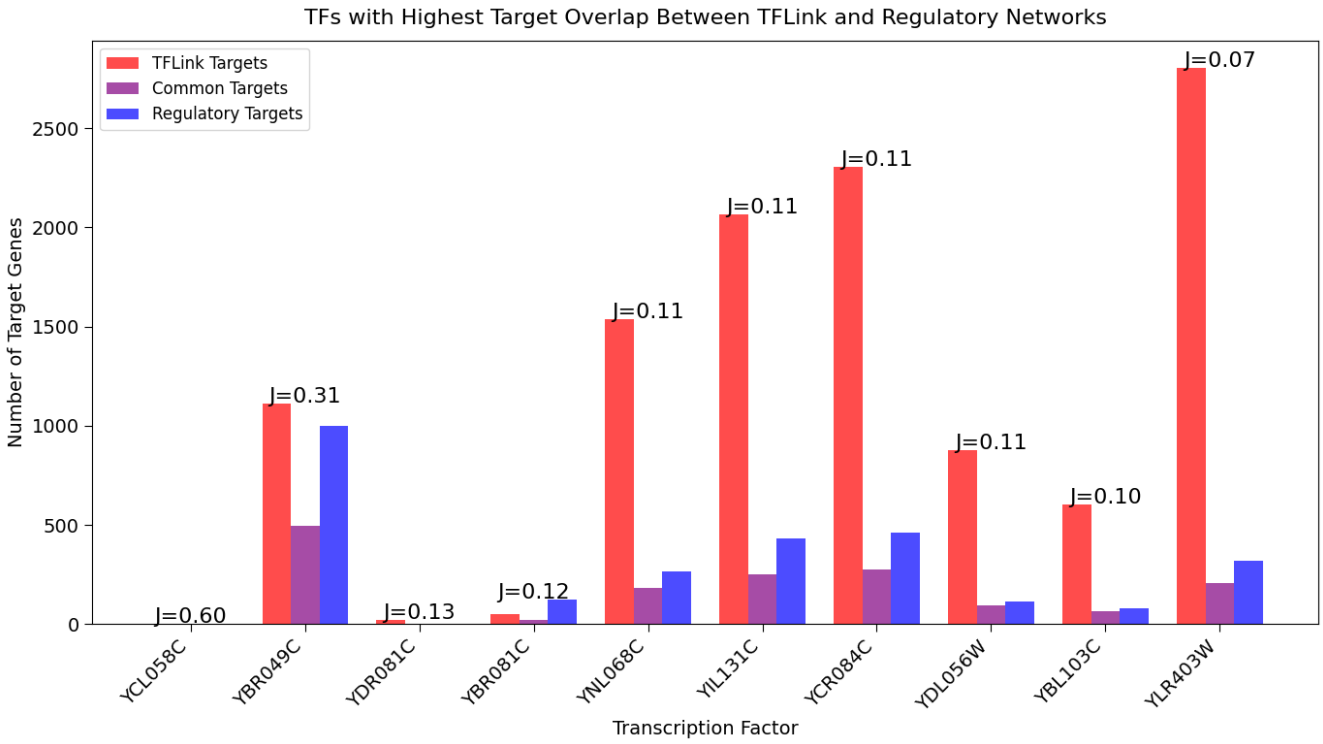
```
michaelvolk@M1-MV torchcell % /Users/michaelvolk/opt/miniconda3/envs/torchcell/bin/python /Users/michaelvol
-dmi-tmi/scripts/string_vs_sgd_vs_reg.py
/Users/michaelvolk/opt/miniconda3/envs/torchcell/lib/python3.11/site-packages/torch_geometric/typing.py:68
  Referenced from: <B4DF21CE-3AD4-3ED1-8E22-0F66900D55D2> /Users/michaelvolk/opt/miniconda3/envs/torchcell
  Reason: tried: '/Library/Frameworks/Python.framework/Versions/3.11/Python' (no such file), '/System/Volu
  warnings.warn(f"An issue occurred while importing 'pyg-lib'. "
/Users/michaelvolk/opt/miniconda3/envs/torchcell/lib/python3.11/site-packages/torch_geometric/typing.py:12
  Referenced from: <B4DF21CE-3AD4-3ED1-8E22-0F66900D55D2> /Users/michaelvolk/opt/miniconda3/envs/torchcell
  Reason: tried: '/Library/Frameworks/Python.framework/Versions/3.11/Python' (no such file), '/System/Volu
  warnings.warn(f"An issue occurred while importing 'torch-sparse'. "
Asset images will be saved to: /Users/michaelvolk/Documents/projects/torchcell/notes/assets/images
/Users/michaelvolk/Documents/projects/torchcell/data/go/go.obo: fmt(1.2) rel(2024-11-03) 43,983 Terms

===== Network Statistics =====
Regulatory: 3632 nodes, 9753 edges
Physical: 5721 nodes, 139463 edges
TFLink: 5092 nodes, 201898 edges
STRING12_Neighborhood: 2204 nodes, 147874 edges
STRING12_Fusion: 3095 nodes, 11810 edges
STRING12_Cooccurence: 2615 nodes, 11115 edges
STRING12_Coexpression: 6503 nodes, 1002538 edges
STRING12_Experimental: 6036 nodes, 825101 edges
STRING12_Database: 4044 nodes, 73818 edges

===== Pairwise Comparison Results =====

Physical vs STRING12_Coexpression:
  Network 1 edges: 139463
  Network 2 edges: 1002538
  Shared edges: 69946
  Jaccard similarity: 0.0652

Physical vs STRING12_Cooccurence:
  Network 1 edges: 139463
  Network 2 edges: 11115
  Shared edges: 1005
  Jaccard similarity: 0.0067

Physical vs STRING12_Database:
  Network 1 edges: 139463
  Network 2 edges: 73818
  Shared edges: 11230
  Jaccard similarity: 0.0556

Physical vs STRING12_Experimental:
  Network 1 edges: 139463
  Network 2 edges: 825101
  Shared edges: 105034
```

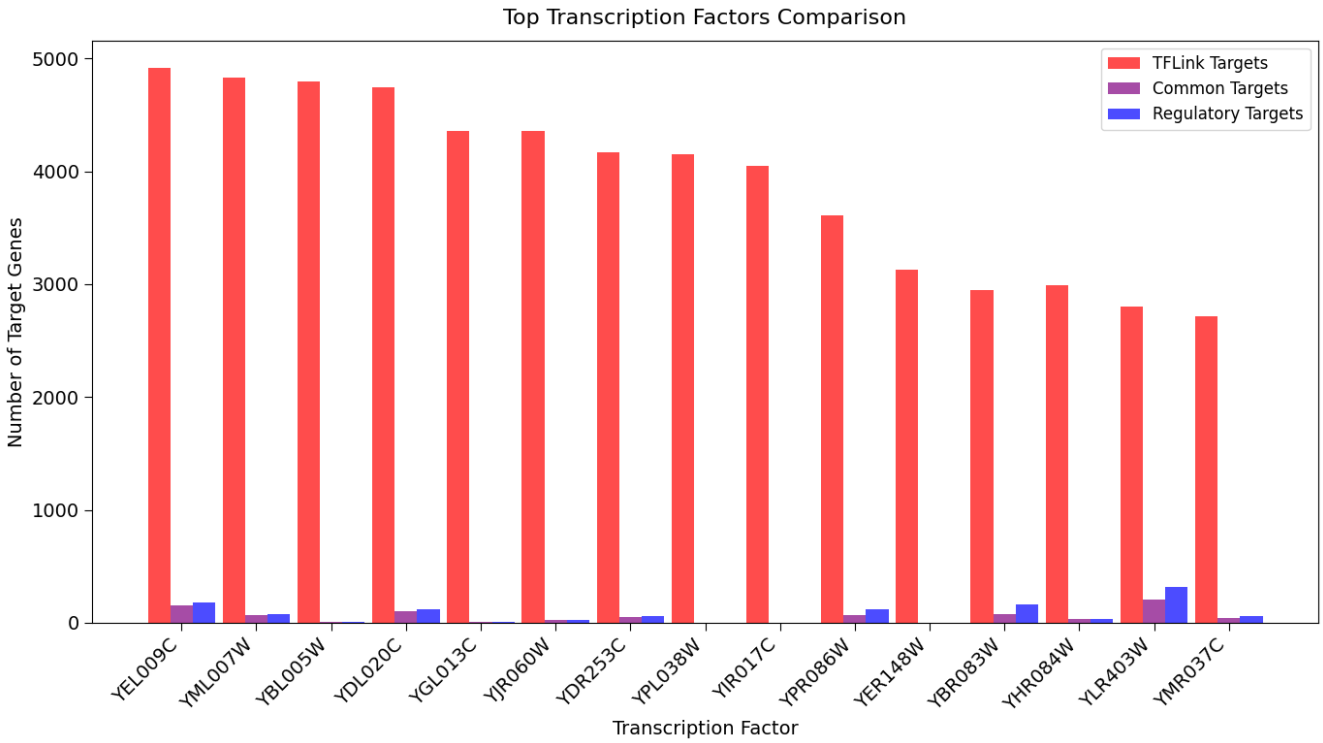
```
  Jaccard similarity: 0.1222

Physical vs STRING12_Fusion:
  Network 1 edges: 139463
  Network 2 edges: 11810
  Shared edges: 788
  Jaccard similarity: 0.0052

Physical vs STRING12_Neighborhood:
  Network 1 edges: 139463
  Network 2 edges: 147874
  Shared edges: 4578
  Jaccard similarity: 0.0162

Physical vs TFLink:
  Network 1 edges: 139463
  Network 2 edges: 200803
  Shared edges: 2013
  Jaccard similarity: 0.0060

Regulatory vs Physical:
  Network 1 edges: 9745
  Network 2 edges: 139463
  Shared edges: 403
  Jaccard similarity: 0.0027

Regulatory vs STRING12_Coexpression:
  Network 1 edges: 9745
  Network 2 edges: 1002538
  Shared edges: 1050
  Jaccard similarity: 0.0010

Regulatory vs STRING12_Cooccurence:
  Network 1 edges: 9745
  Network 2 edges: 11115
  Shared edges: 37
  Jaccard similarity: 0.0018

Regulatory vs STRING12_Database:
  Network 1 edges: 9745
  Network 2 edges: 73818
  Shared edges: 65
  Jaccard similarity: 0.0008

Regulatory vs STRING12_Experimental:
  Network 1 edges: 9745
  Network 2 edges: 825101
  Shared edges: 1244
  Jaccard similarity: 0.0015

Regulatory vs STRING12_Fusion:
  Network 1 edges: 9745
  Network 2 edges: 11810
  Shared edges: 11
  Jaccard similarity: 0.0005

Regulatory vs STRING12_Neighborhood:
  Network 1 edges: 9745
```

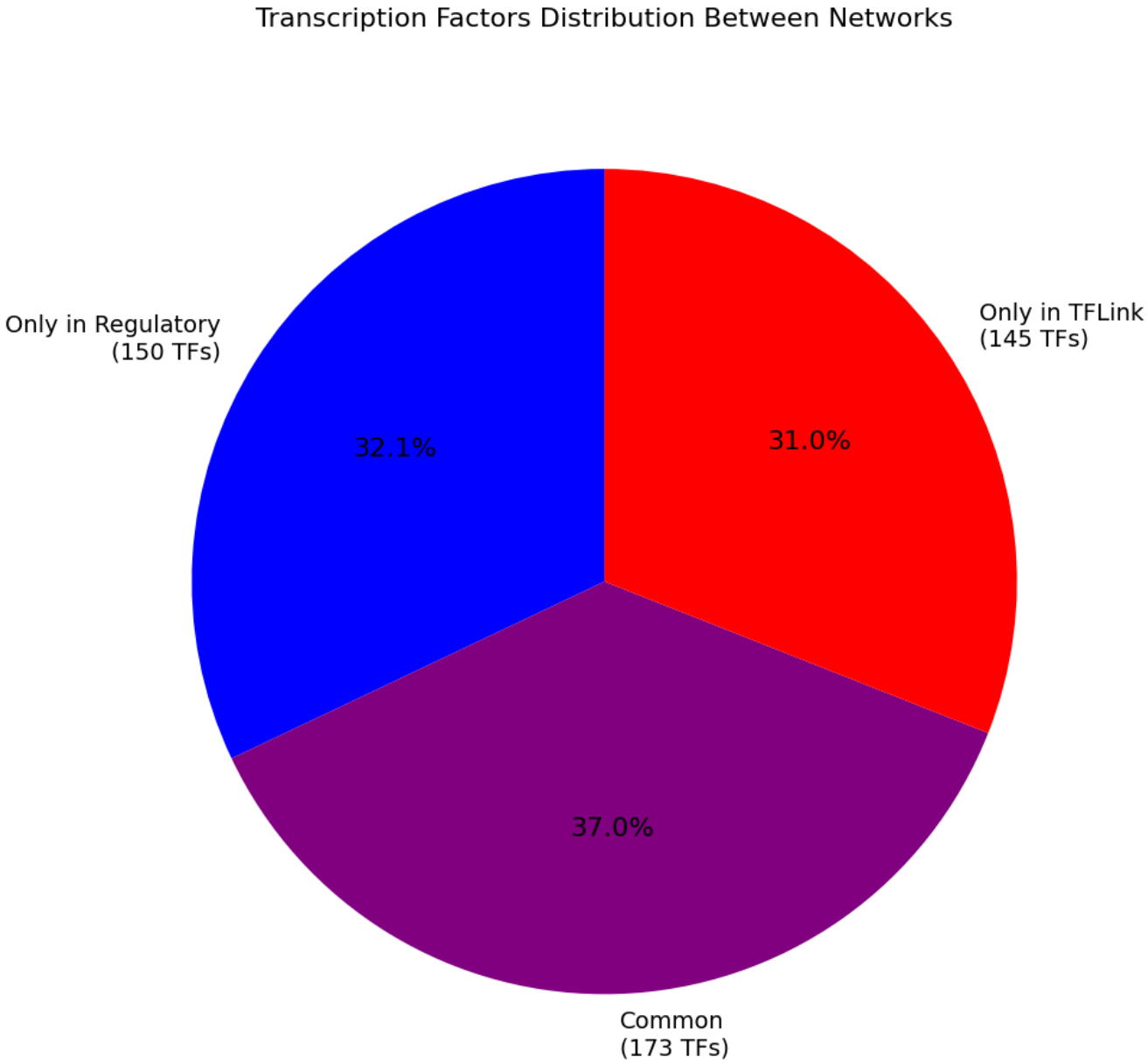
```
  Network 2 edges: 147874
  Shared edges: 27
  Jaccard similarity: 0.0002

Regulatory vs TFLink:
  Network 1 edges: 9745
  Network 2 edges: 200803
  Shared edges: 3577
  Jaccard similarity: 0.0173

STRING12_Coexpression vs STRING12_Database:
  Network 1 edges: 1002538
  Network 2 edges: 73818
  Shared edges: 57388
  Jaccard similarity: 0.0563

STRING12_Coexpression vs STRING12_Experimental:
  Network 1 edges: 1002538
  Network 2 edges: 825101
  Shared edges: 557632
  Jaccard similarity: 0.4391

STRING12_Cooccurence vs STRING12_Coexpression:
  Network 1 edges: 11115
  Network 2 edges: 1002538
  Shared edges: 7119
  Jaccard similarity: 0.0071

STRING12_Cooccurence vs STRING12_Database:
  Network 1 edges: 11115
  Network 2 edges: 73818
  Shared edges: 1184
  Jaccard similarity: 0.0141

STRING12_Cooccurence vs STRING12_Experimental:
  Network 1 edges: 11115
  Network 2 edges: 825101
  Shared edges: 5660
  Jaccard similarity: 0.0068

STRING12_Experimental vs STRING12_Database:
  Network 1 edges: 825101
  Network 2 edges: 73818
  Shared edges: 40076
  Jaccard similarity: 0.0467

STRING12_Fusion vs STRING12_Coexpression:
  Network 1 edges: 11810
  Network 2 edges: 1002538
  Shared edges: 8601
  Jaccard similarity: 0.0086

STRING12_Fusion vs STRING12_Cooccurence:
  Network 1 edges: 11810
  Network 2 edges: 11115
  Shared edges: 318
  Jaccard similarity: 0.0141
```

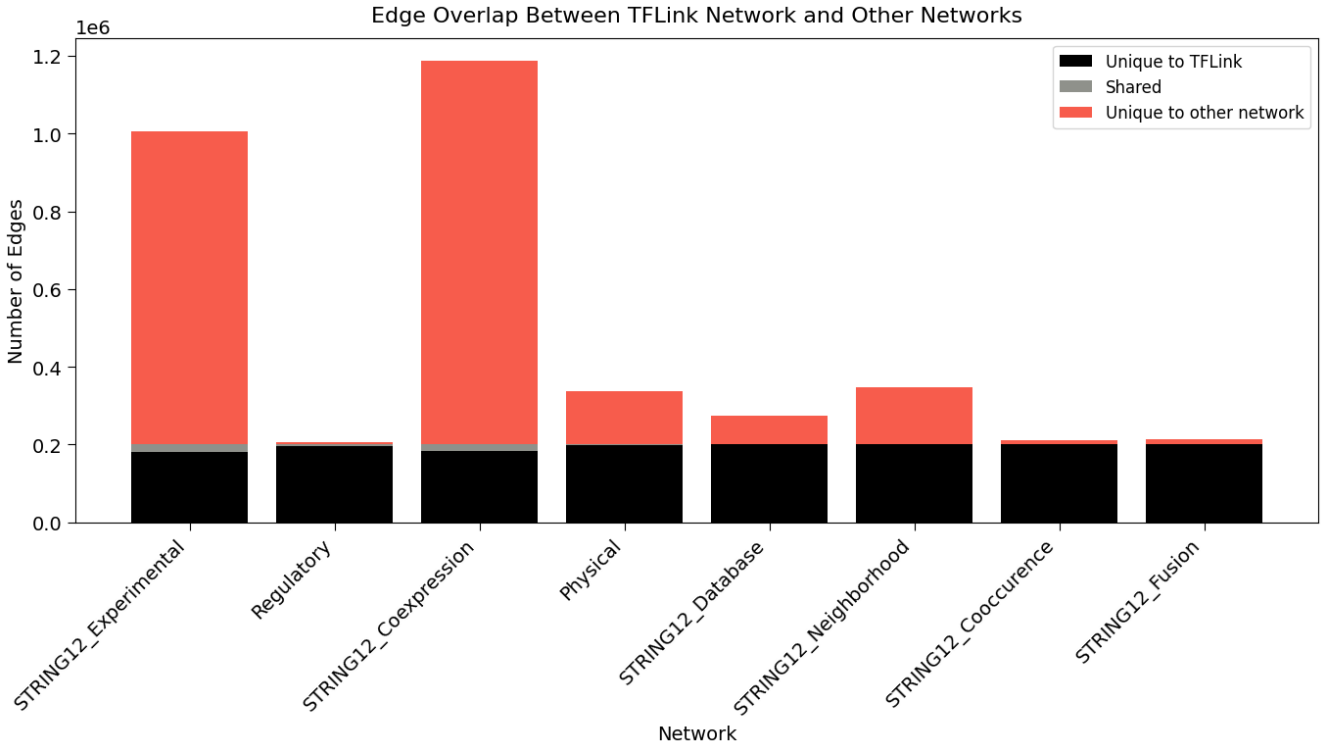
```
STRING12_Fusion vs STRING12_Database:
  Network 1 edges: 11810
  Network 2 edges: 73818
  Shared edges: 876
  Jaccard similarity: 0.0103

STRING12_Fusion vs STRING12_Experimental:
  Network 1 edges: 11810
  Network 2 edges: 825101
  Shared edges: 5960
  Jaccard similarity: 0.0072

STRING12_Neighborhood vs STRING12_Coexpression:
  Network 1 edges: 147874
  Network 2 edges: 1002538
  Shared edges: 120681
  Jaccard similarity: 0.1172

STRING12_Neighborhood vs STRING12_Cooccurence:
  Network 1 edges: 147874
  Network 2 edges: 11115
  Shared edges: 287
  Jaccard similarity: 0.0018

STRING12_Neighborhood vs STRING12_Database:
  Network 1 edges: 147874
  Network 2 edges: 73818
  Shared edges: 22212
  Jaccard similarity: 0.1113

STRING12_Neighborhood vs STRING12_Experimental:
  Network 1 edges: 147874
  Network 2 edges: 825101
  Shared edges: 61904
  Jaccard similarity: 0.0679

STRING12_Neighborhood vs STRING12_Fusion:
  Network 1 edges: 147874
  Network 2 edges: 11810
  Shared edges: 1678
  Jaccard similarity: 0.0106

TFLink vs STRING12_Coexpression:
  Network 1 edges: 200803
  Network 2 edges: 1002538
  Shared edges: 16708
  Jaccard similarity: 0.0141

TFLink vs STRING12_Cooccurence:
  Network 1 edges: 200803
  Network 2 edges: 11115
  Shared edges: 120
  Jaccard similarity: 0.0006

TFLink vs STRING12_Database:
  Network 1 edges: 200803
  Network 2 edges: 73818
  Shared edges: 576
```

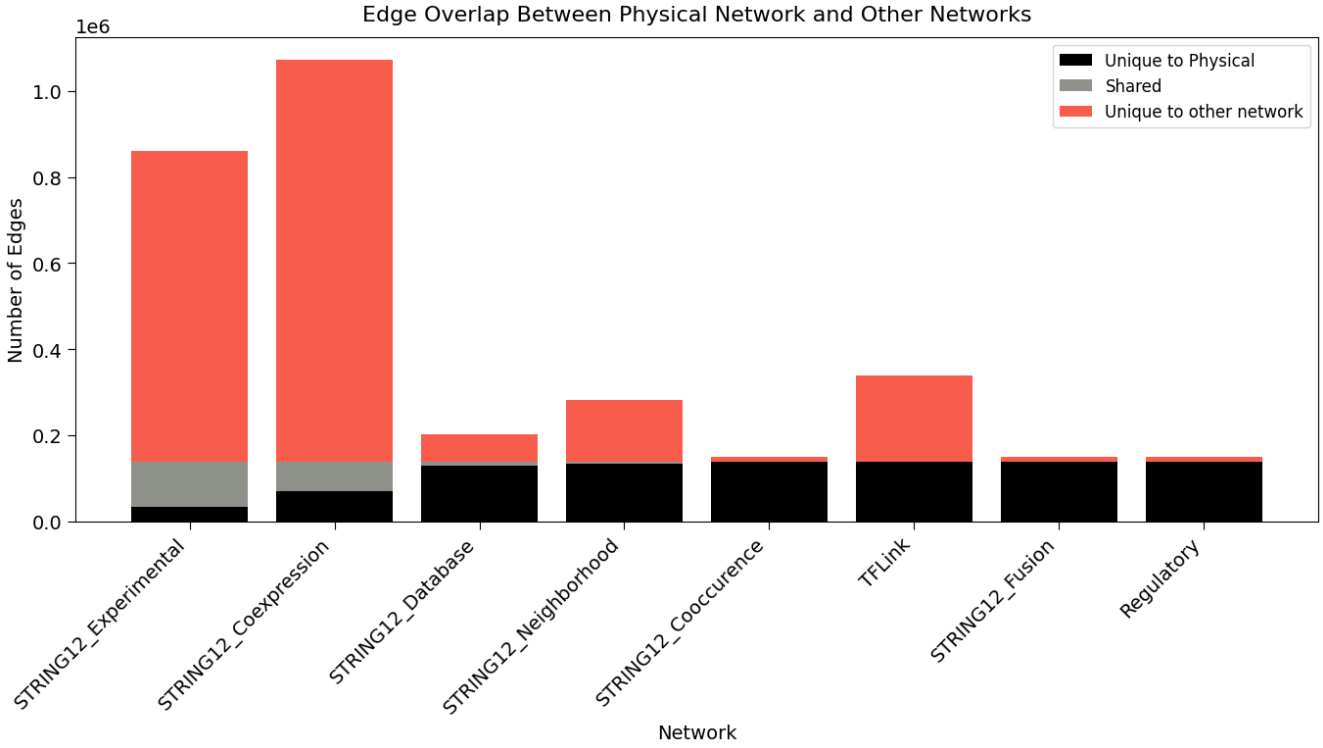
```
  Jaccard similarity: 0.0021

TFLink vs STRING12_Experimental:
  Network 1 edges: 200803
  Network 2 edges: 825101
  Shared edges: 20035
  Jaccard similarity: 0.0199

TFLink vs STRING12_Fusion:
  Network 1 edges: 200803
  Network 2 edges: 11810
  Shared edges: 60
  Jaccard similarity: 0.0003

TFLink vs STRING12_Neighborhood:
  Network 1 edges: 200803
  Network 2 edges: 147874
  Shared edges: 414
  Jaccard similarity: 0.0012

===== Generating Visualizations =====
 Network sizes chart created
 Jaccard similarity heatmap created
 Edge overlap matrices created
 Network overlap charts created
 TF regulatory comparison visualizations created

All visualizations saved to /Users/michaelvolk/Documents/projects/torchcell/notes/assets/images
```

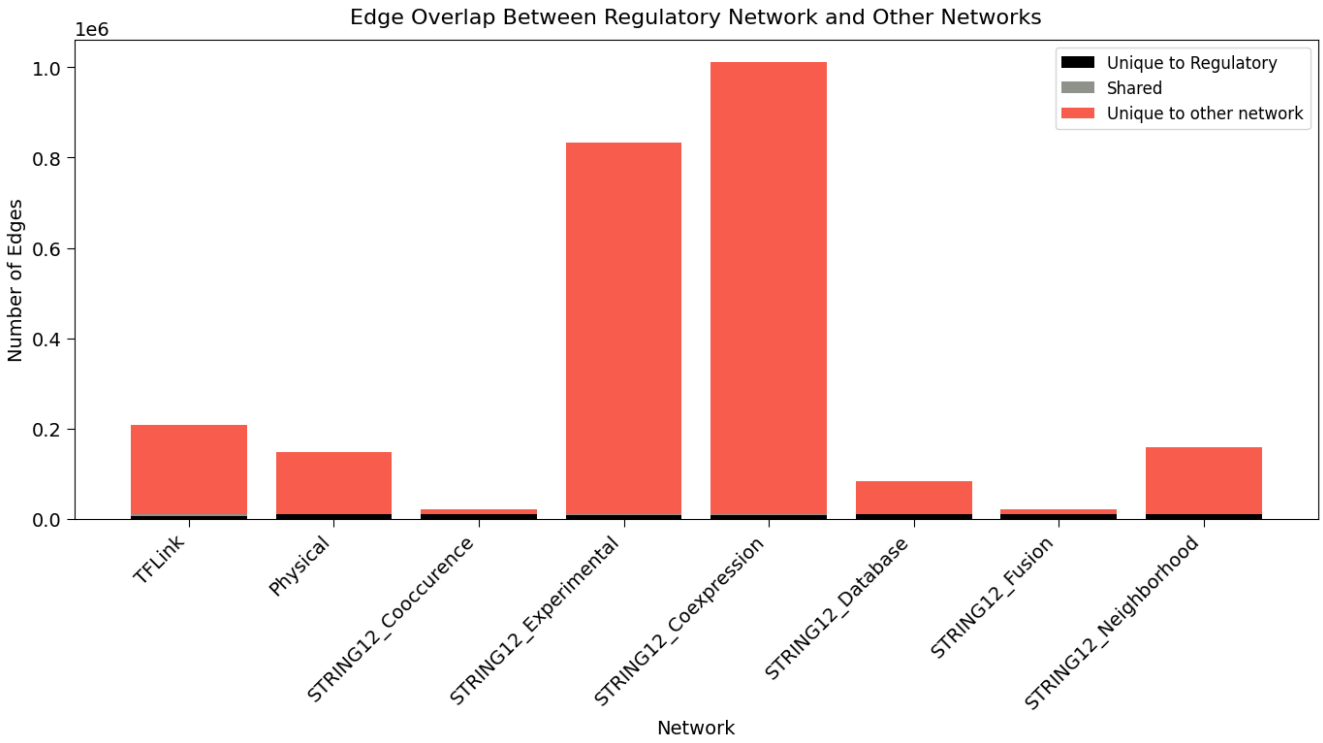


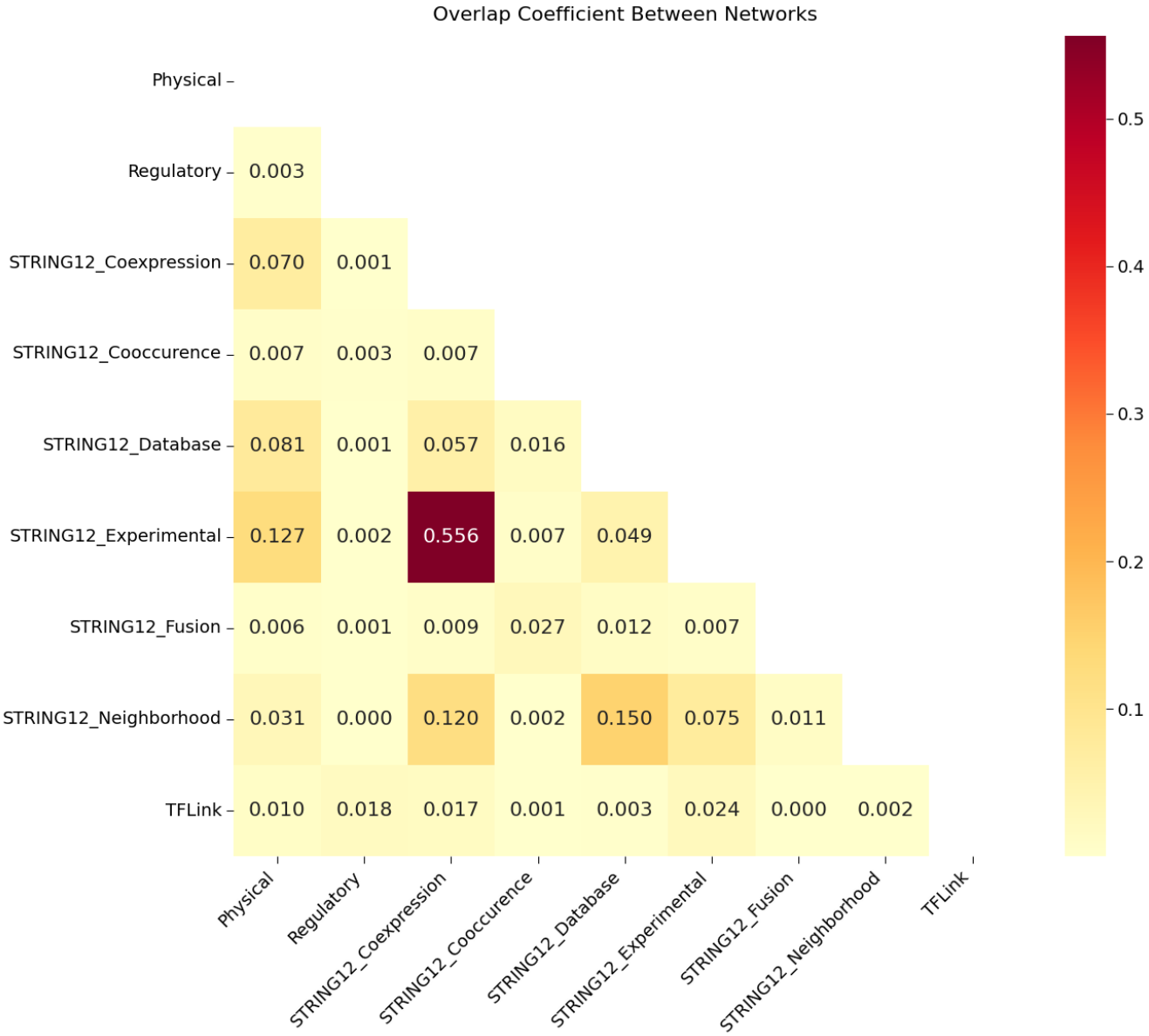

Top Transcription Factors Comparison
TFLink Targets
Common Targets
Regulatory Targets
Number of Target Genes
5000
4000
3000
2000
1000
0
YEL009C
YML007W
YBL005W
YDL020C
YGL013C
YJR060W
YDR253C
YPL038W
YIR017C
YPR086W
YER148W
YBR083W
YHR084W
YLR403W
YMR037C
Transcription Factor

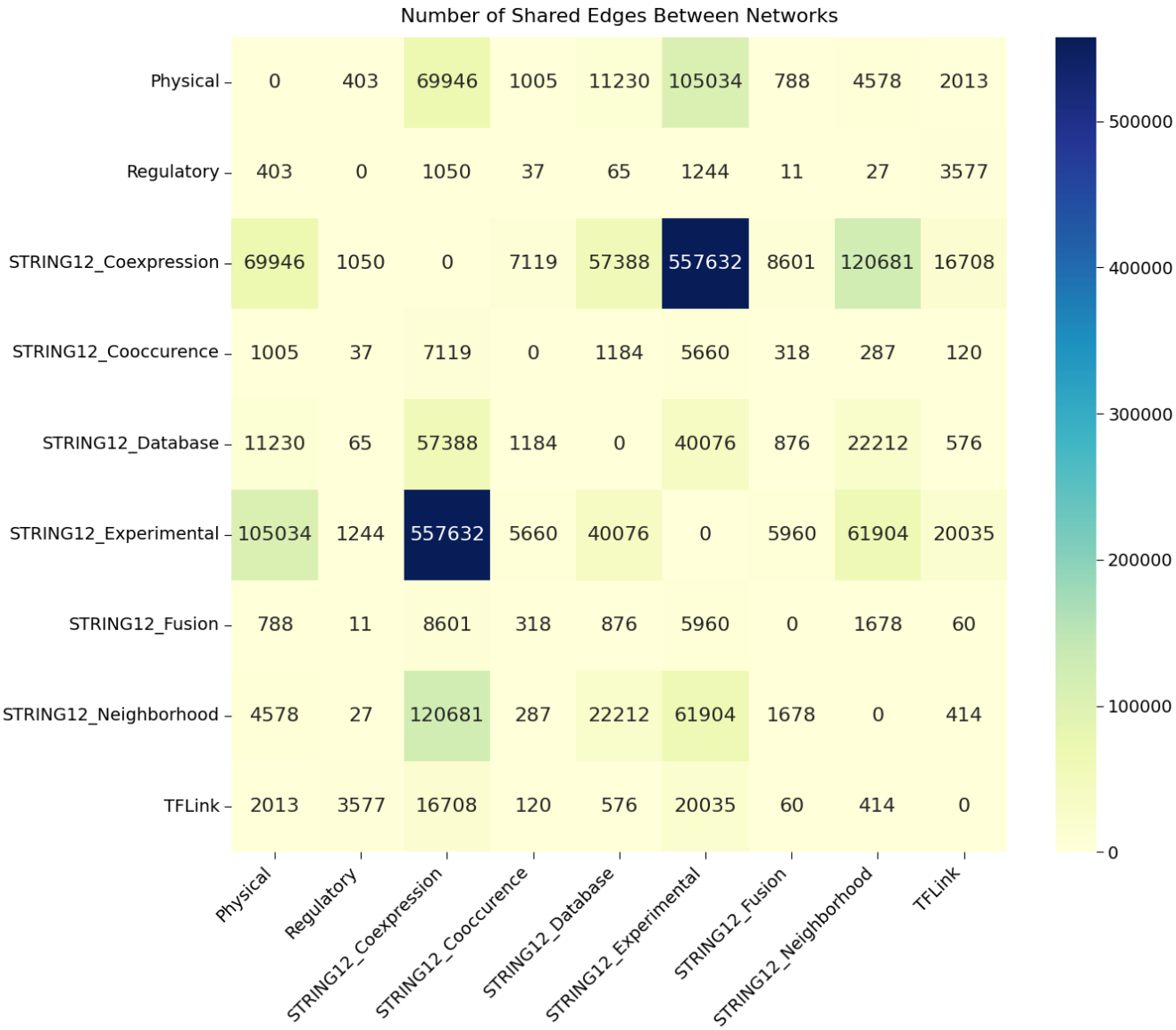

Transcription Factors Distribution Between Networks
Only in Regulatory
(150 TFs)
32.1%
Only in TFLink
(145 TFs)
31.0%
37.0%
Common
(173 TFs)

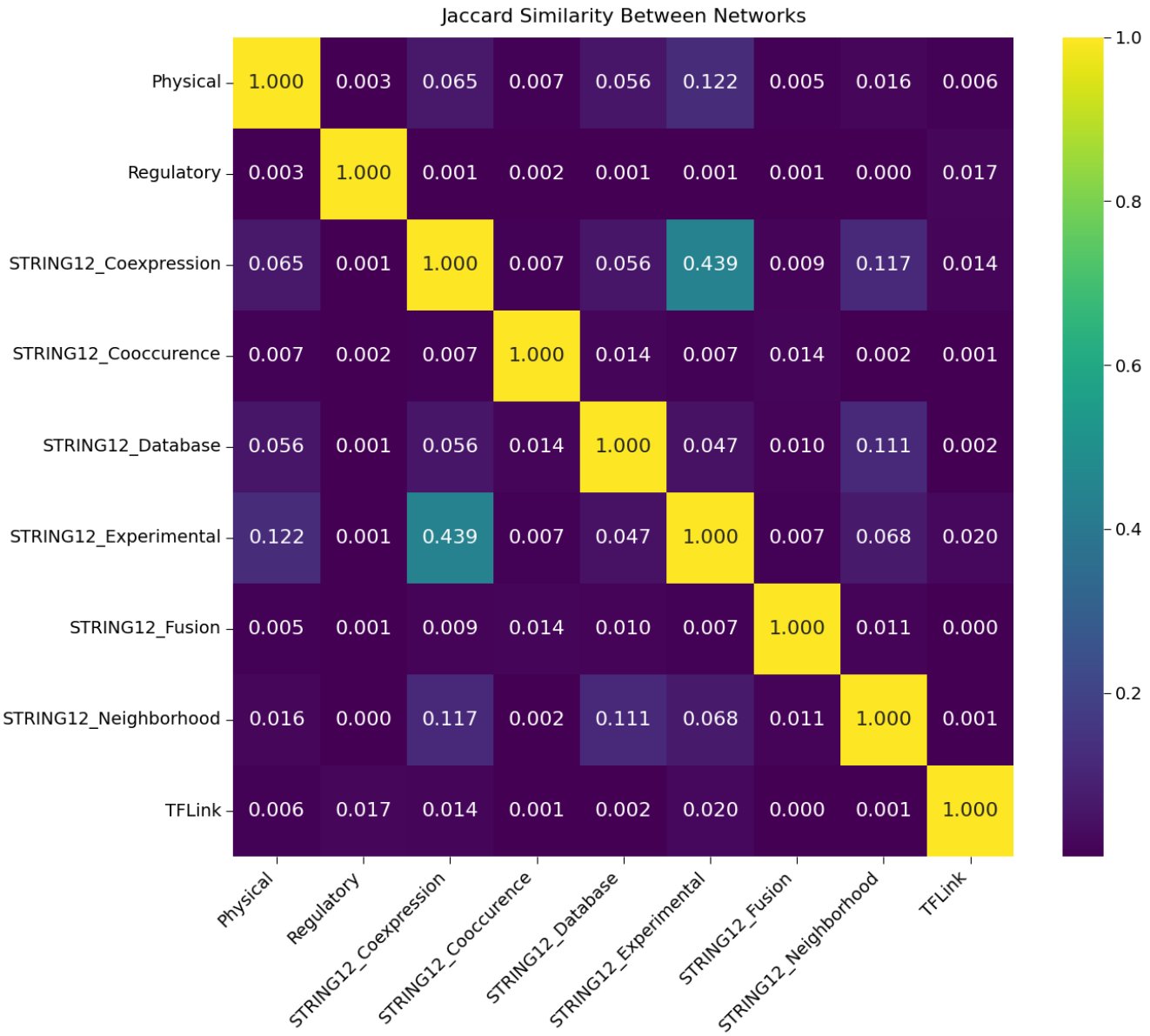

Edge Overlap Between TFLink Network and Other Networks
1e6
1.2
1.0
0.8
0.6
0.4
0.2
0.0
Number of Edges
Unique to TFLink
Shared
Unique to other network
STRING12_Experimental
Regulatory
STRING12_Coexpression
Physical
STRING12_Database
STRING12_Neighborhood
STRING12_Cooccurence
STRING12_Fusion
Network

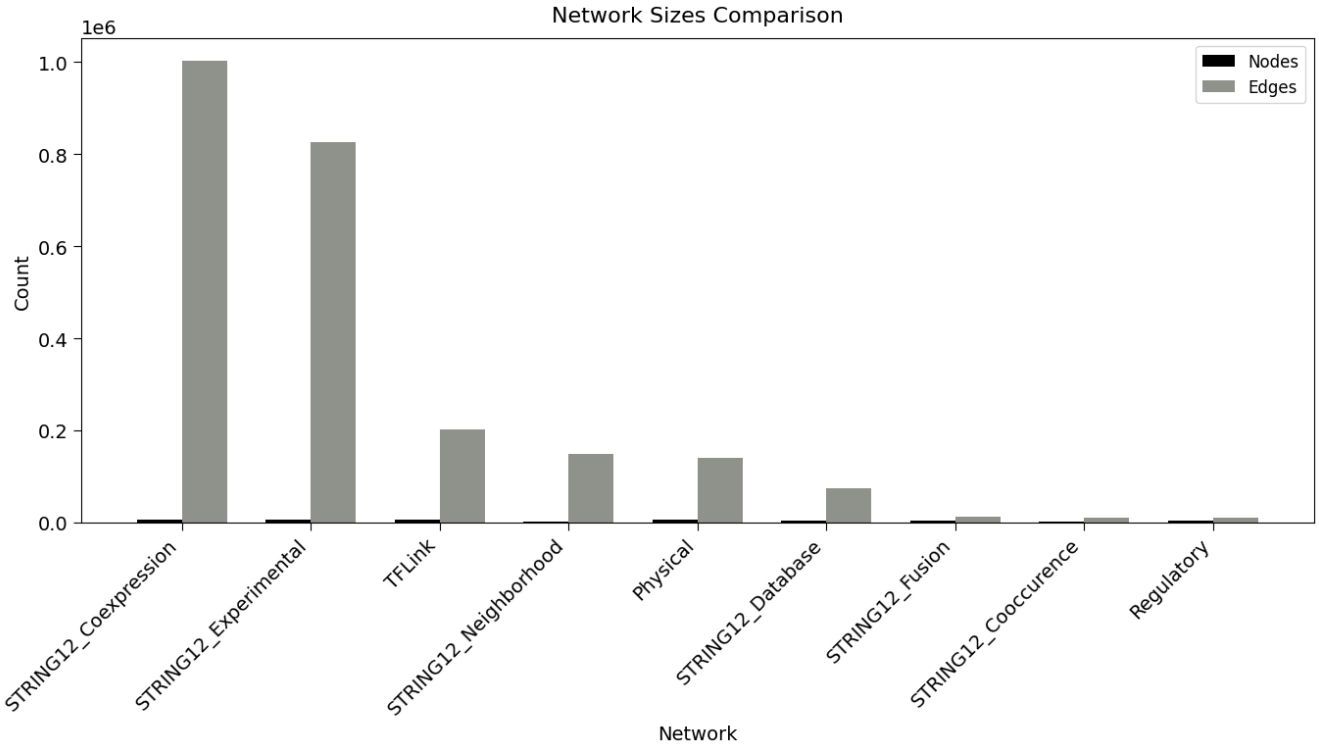

Edge Overlap Between Physical Network and Other Networks
1e6
1.0
0.8
0.6
0.4
0.2
0.0
Number of Edges
Unique to Physical
Shared
Unique to other network
STRING12_Experimental
STRING12_Coexpression
STRING12_Database
STRING12_Neighborhood
STRING12_Cooccurence
TFLink
STRING12_Fusion
Regulatory
Network


Edge Overlap Between Regulatory Network and Other Networks
1e6
1.0
0.8
0.6
0.4
0.2
0.0
Number of Edges
Unique to Regulatory
Shared
Unique to other network
TFLink
Physical
STRING12_Cooccurence
STRING12_Experimental
STRING12_Coexpression
STRING12_Database
STRING12_Fusion
STRING12_Neighborhood
Network


Overlap Coefficient Between Networks
Physical
Regulatory
STRING12_Coexpression
STRING12_Cooccurence
STRING12_Database
STRING12_Experimental
STRING12_Fusion
STRING12_Neighborhood
TFLink
0.003
0.070 0.001
0.007 0.003 0.007
0.081 0.001 0.057 0.016
0.127 0.002 0.556 0.007 0.049
0.006 0.001 0.009 0.027 0.012 0.007
0.031 0.000 0.120 0.002 0.150 0.075 0.011
0.010 0.018 0.017 0.001 0.003 0.024 0.000 0.002
0.5
0.4
0.3
0.2
0.1


Number of Shared Edges Between Networks
Physical 0 403 69946 1005 11230 105034 788 4578 2013
Regulatory 403 0 1050 37 65 1244 11 27 3577
STRING12_Coexpression 69946 1050 0 7119 57388 557632 8601 120681 16708
STRING12_Cooccurence 1005 37 7119 0 1184 5660 318 287 120
STRING12_Database 11230 65 57388 1184 0 40076 876 22212 576
STRING12_Experimental 105034 1244 557632 5660 40076 0 5960 61904 20035
STRING12_Fusion 788 11 8601 318 876 5960 0 1678 60
STRING12_Neighborhood 4578 27 120681 287 22212 61904 1678 0 414
TFLink 2013 3577 16708 120 576 20035 60 414 0
Physical
Regulatory
STRING12_Coexpression
STRING12_Cooccurence
STRING12_Database
STRING12_Experimental
STRING12_Fusion
STRING12_Neighborhood
TFLink
500000
400000
300000
200000
100000
0


Jaccard Similarity Between Networks
Physical 1.000 0.003 0.065 0.007 0.056 0.122 0.005 0.016 0.006
Regulatory 0.003 1.000 0.001 0.002 0.001 0.001 0.001 0.000 0.017
STRING12_Coexpression 0.065 0.001 1.000 0.007 0.056 0.439 0.009 0.117 0.014
STRING12_Cooccurence 0.007 0.002 0.007 1.000 0.014 0.007 0.014 0.002 0.001
STRING12_Database 0.056 0.001 0.056 0.014 1.000 0.047 0.010 0.111 0.002
STRING12_Experimental 0.122 0.001 0.439 0.007 0.047 1.000 0.007 0.068 0.020
STRING12_Fusion 0.005 0.001 0.009 0.014 0.010 0.007 1.000 0.011 0.000
STRING12_Neighborhood 0.016 0.000 0.117 0.002 0.111 0.068 0.011 1.000 0.001
TFLink 0.006 0.017 0.014 0.001 0.002 0.020 0.000 0.001 1.000
Physical
Regulatory
STRING12_Coexpression
STRING12_Cooccurence
STRING12_Database
STRING12_Experimental
STRING12_Fusion
STRING12_Neighborhood
TFLink
1.0
0.8
0.6
0.4
0.2



Network	Nodes	Edges
Regulatory	3632	9753
Physical	5721	139463
TFLink	5092	201898
STRING12_Neighborhood	2204	147874
STRING12_Fusion	3095	11810
STRING12_Cooccurence	2615	11115
STRING12_Coexpression	6503	1002538
STRING12_Experimental	6036	825101
STRING12_Database	4044	73818
Sum	**38942**	**2423370**