

Identifying Differentially Connected Genes of Breast Invasive Carcinoma in Co-expression Networks

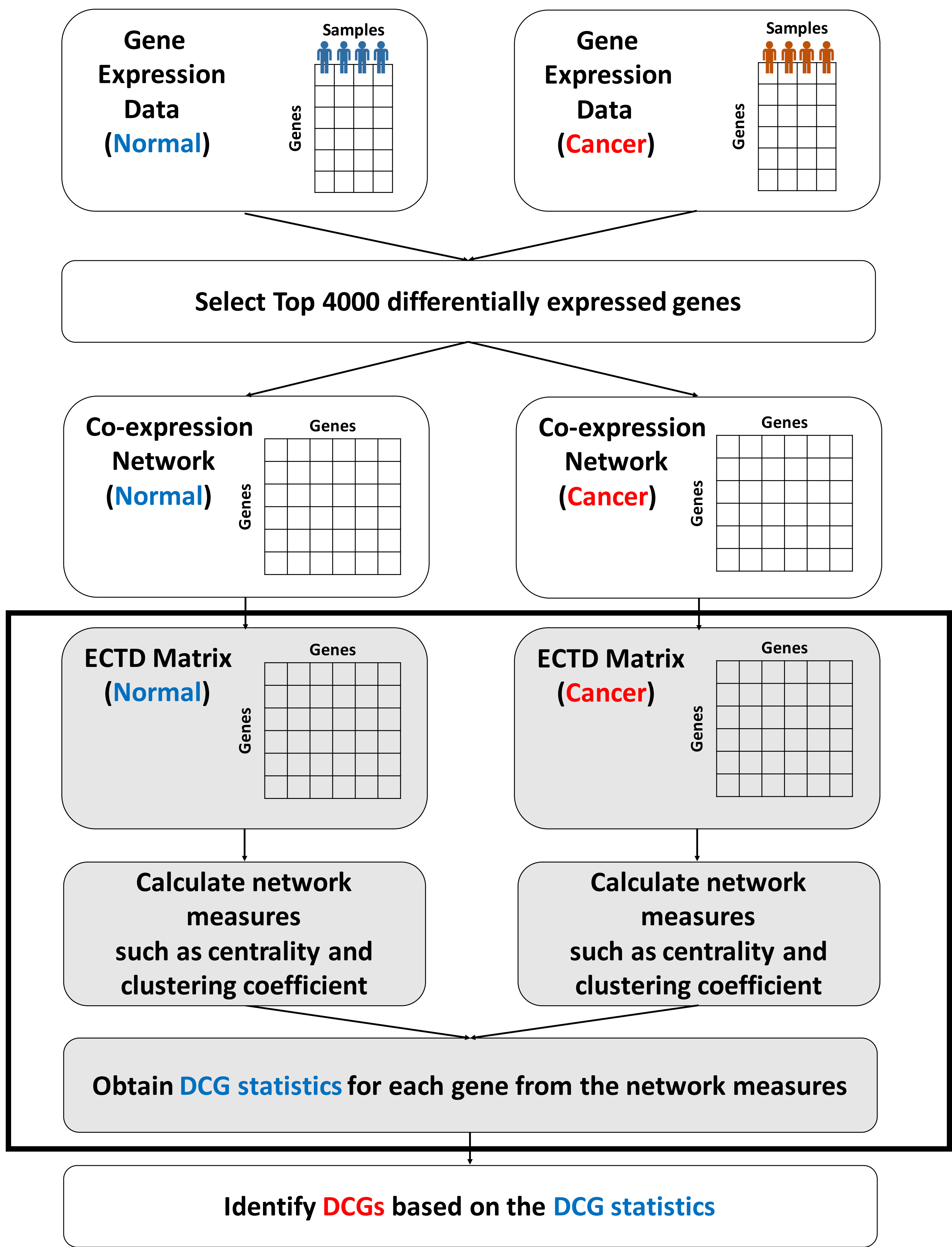
Fen Pei and Seo-Jin Bang

Joint Carnegie Mellon and University of Pittsburgh PhD Program in Computational Biology

MOTIVATION/GOAL

Topological changes of a gene network across different conditions such as normal versus cancer can provide hints regarding the disrupted regulatory relationships or affected regulatory sub-networks specific to a phenotype of interest. This work aims to identify **differentially connected genes (DCGs)** that cause significant topological changes between two gene co-expression networks: breast invasive carcinoma versus normal samples.

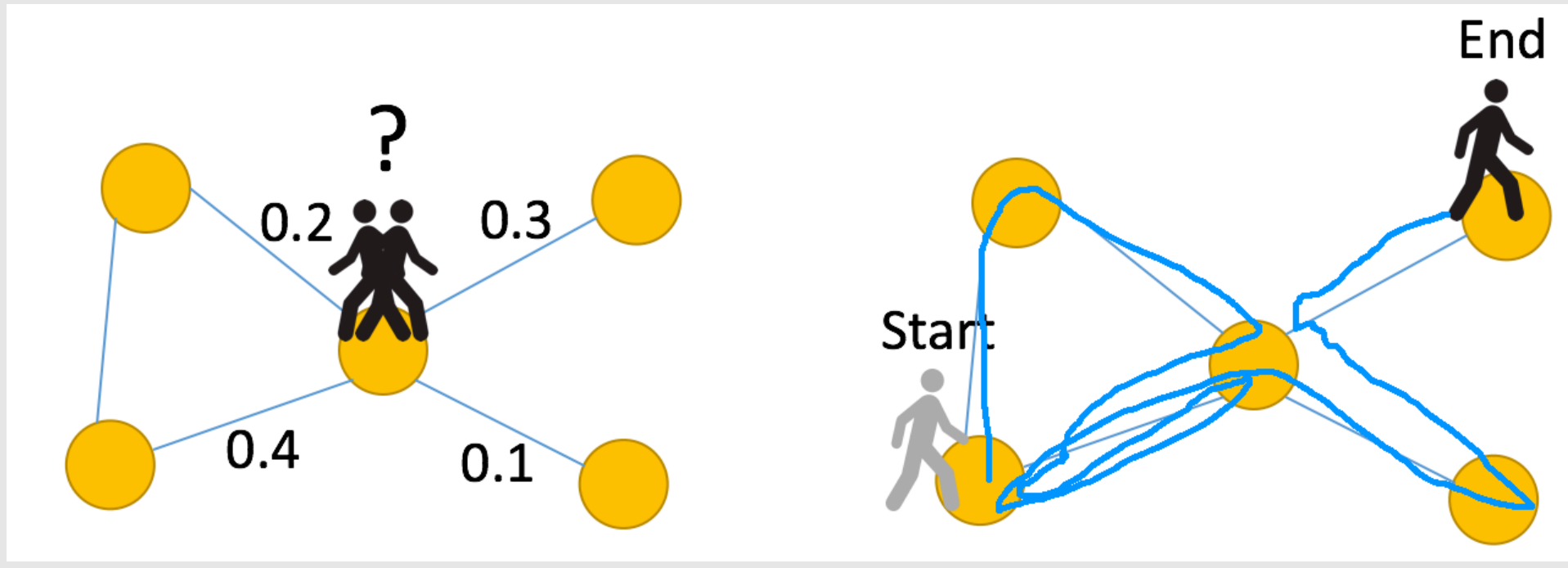
METHOD



DATA

- **Cancer Genome Atlas (TCGA)**, <https://tcga-data.nci.nih.gov/tcga/>
- **Expression profiles of 17,814 genes**
- **531 Tumor and 62 Normal samples**

Euelidean Commute Time Distance (ECTD) Matrix



Network Measures

Degree Centrality ($Mean_{group}$)	sum of weights to its edges
Betweenness Centrality ($Betweenness_{group}$)	the number of times a node acts as a bridge along the shortest path between two other nodes
Closeness Centrality ($Closeness_{group}$)	the inverse sum of shortest distances to all other nodes from a focal node
Clustering Coefficient ($ClusteringCoefficient_{group}$)	the ratio between the total number (e_i) of the edges actually connecting its nearest neighbors to the with node and the total number of all possible edges between all these nearest neighbors

DCG Statistics

Mean Degree Statistic	$\frac{ Mean_{cancer}(i) - Mean_{normal}(i) }{Mean_{cancer}(i) + Mean_{normal}(i)}$
Betweenness Statistic	$ Betweenness_{cancer}(i) - Betweenness_{normal}(i) $
Closeness Statistic	$ (Closeness_{cancer}(i) - median(Closeness_{cancer})) - (Closeness_{normal}(i) - median(Closeness_{normal})) $
Clustering Coefficient Statistic	$\left \log \left \frac{ClusteringCoefficient_{normal}(i) \cdot Mean_{normal}(i)}{ClusteringCoefficient_{cancer}(i) \cdot Mean_{cancer}(i)} \right \right $

* i is the index of each gene

RESULT

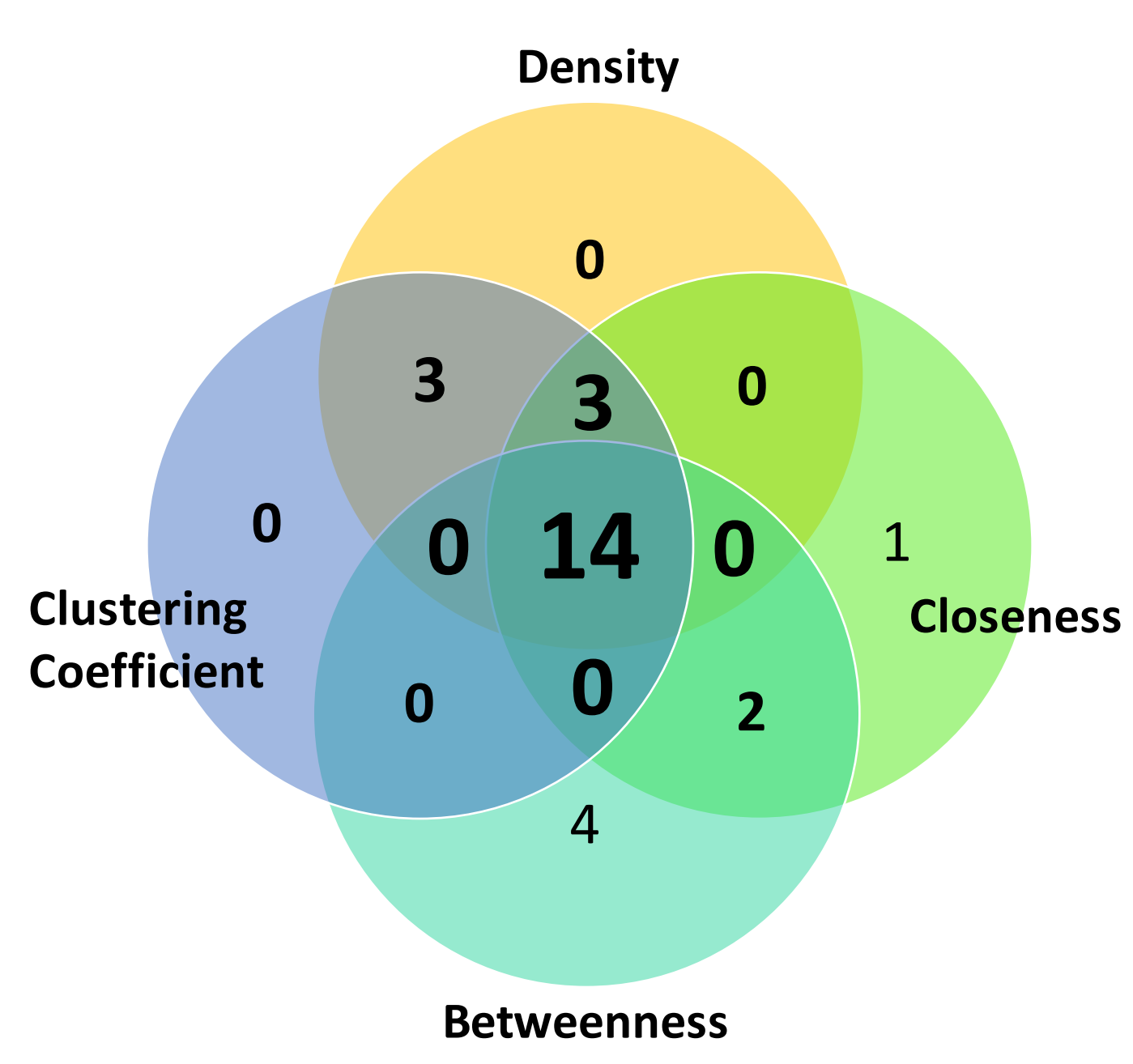


Figure 1. Overlapped genes between the measures of DCG. Top 20 genes in each method. Total 27 genes

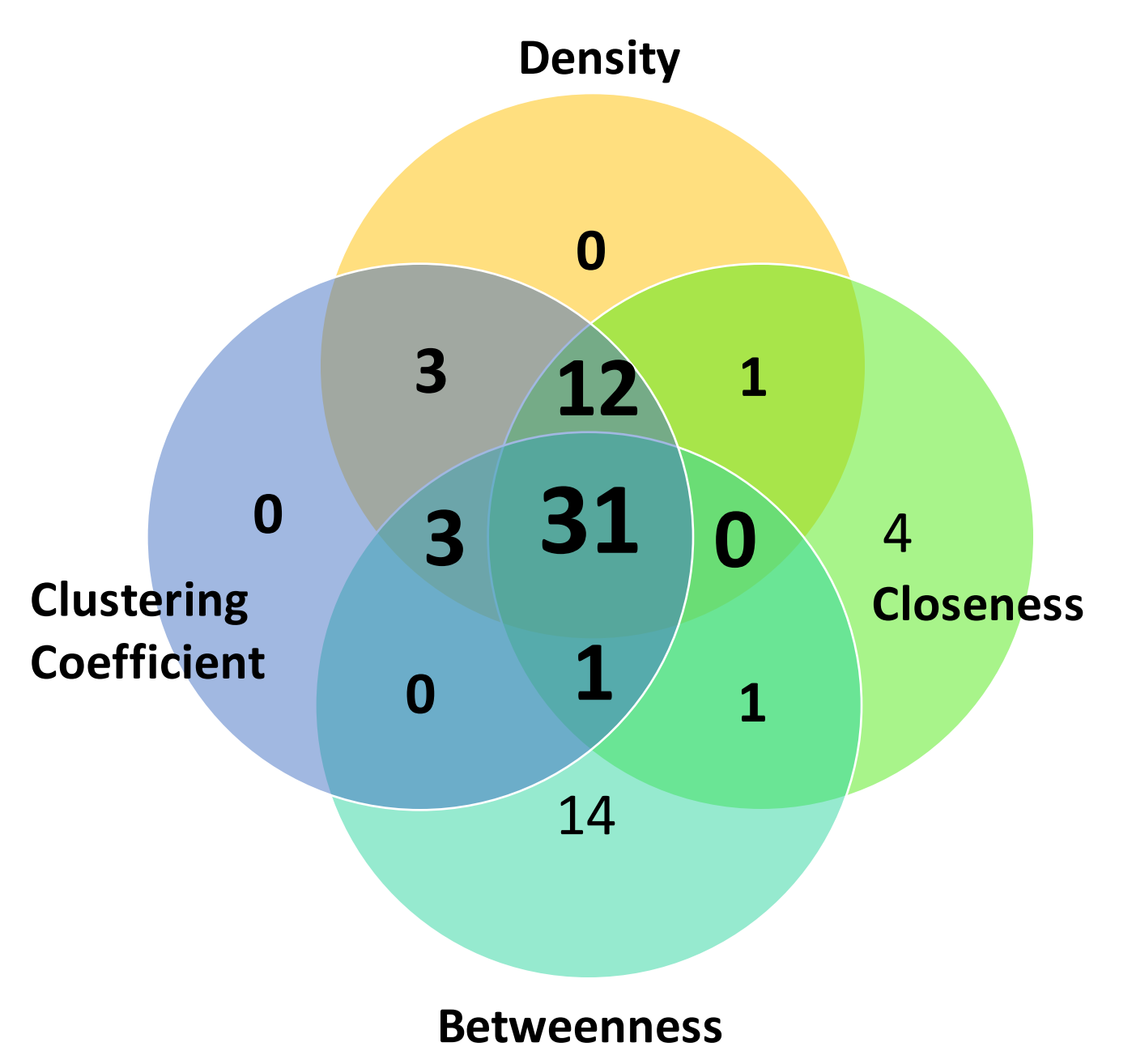


Figure 2. Overlapped genes between the measures of DCG. Top 50 genes in each method. Total 70 genes

Gene	DEG rank (out of 4000)
APBB1	2069
ATAD1	3881
C2orf37	647
CAPRIN1	3790
CKAP2	205
KIAA0859	1609
LOC388284	2405
LRFN3	3751
NUP155	1151
PCGF5	1611
RBJ	1685
RFWD3	1040
STAU1	2889
TBRG4	2780

Table1. Rank of DEG analysis for a list of 14 genes **completely overlapped** among all the measures of DCG. Top 20 genes in each method.

Gene	DEG rank (out of 400)
ATE1	3139
ATP13A1	1895
BRMS1	1777
C20orf121	2425
FAM10A5	2953
FKSG24	1559
GGPS1	2571
GON4L	3543
GSTP1	1049
NSMCE4A	4030
RXFP4	3592
WDR8	2787
WSB1	1235

Table2. Rank of DEG analysis for a list of 13 genes **partially overlapped** among the measures of DCG. Top 20 genes in each method.

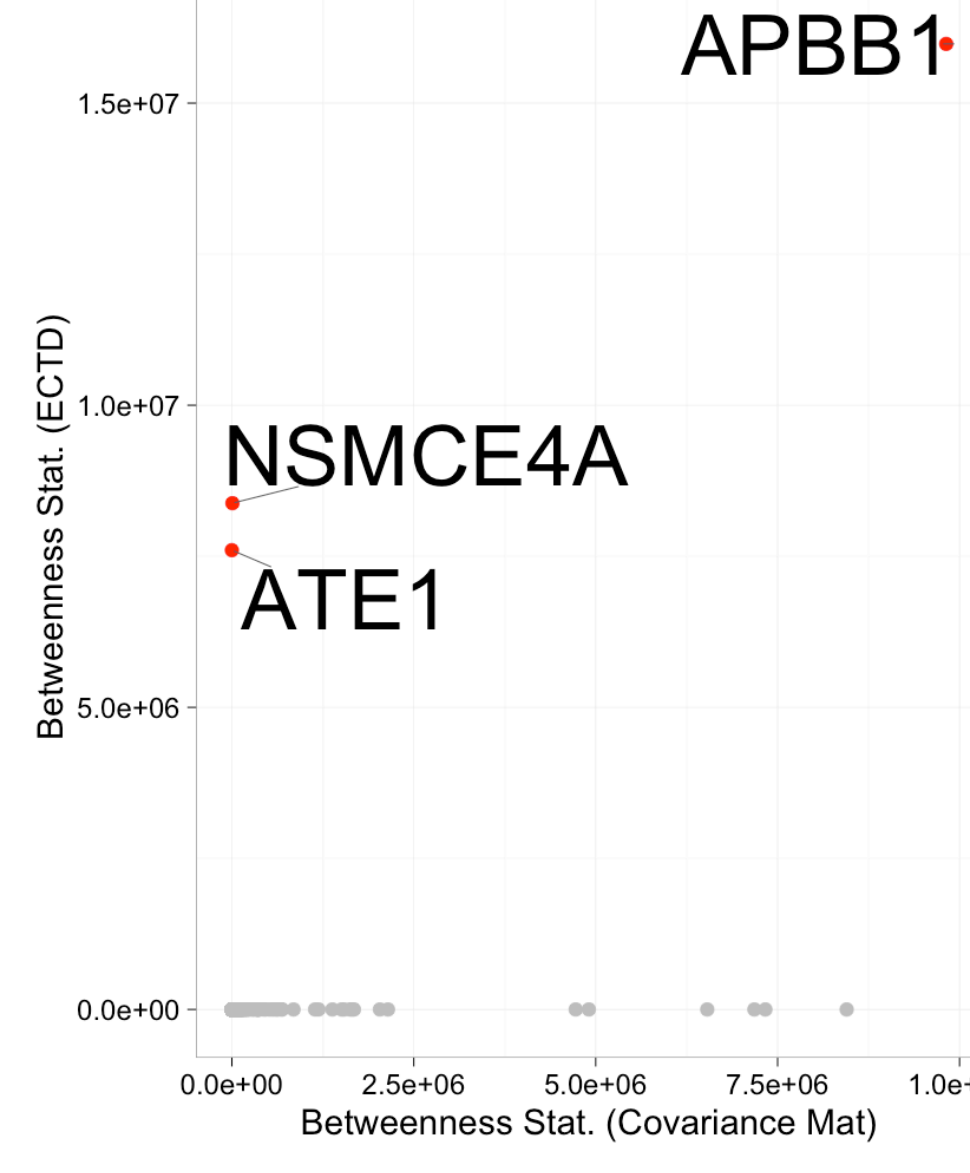


Figure3. Three interesting genes identified by comparing the two betweenness statistics: ECTD versus co-expression network.

FUTURE WORK

- **Biological Evaluation:** identify enriched biological GO terms and pathways of the DCGs.