Network Concepts and their geometric Interpretation

'mini' R Tutorial

Computation of fundamental and eigengene-based network concepts

Jun Dong, Steve Horvath

Correspondence: shorvath@mednet.ucla.edu, http://www.ph.ucla.edu/biostat/people/horvath.htm

This is a self-contained R software tutorial that illustrates how to compute network concepts and their eigengene based analogues. This tutorial shows how we arrived at Figure 1 and Table 3 in Horvath and Dong (2008).
The microarray data set corresponds to gene expression measurements in the mouse liver tissues of female mice of an F2 mouse cross. Some familiarity with the R software is desirable but the document is fairly self-contained.
To cite the methods and results of this article, please use the following references

- *Horvath S, Dong J (2008) Geometric Interpretation of Gene Coexpression Network Analysis. PLoS Comput Biol 4(8): e1000117*
- *Dong J, Horvath S (2007) Understanding Network Concepts in Modules, BMC Systems Biology 2007, 1:24*
- *The WGCNA R package is described in: Langfelder P, Horvath S (2008) WGCNA: an R package for weighted correlation network analysis. BMC Bioinformatics, 2008 9:559*

The tutorial, R functions, data etc can be found at the following webpage
http://www.genetics.ucla.edu/labs/horvath/ModuleConformity/GeometricInterpretation/

Please also download the **WGCNA R library** from
http://www.genetics.ucla.edu/labs/horvath/CoexpressionNetwork/Softwares/WGCNA/
To install it in R use the command "Install package(s) from local zip file" which can be found in the tab "Packages" The WGCNA package contains our most recent R software code and accompanying software tutorials.

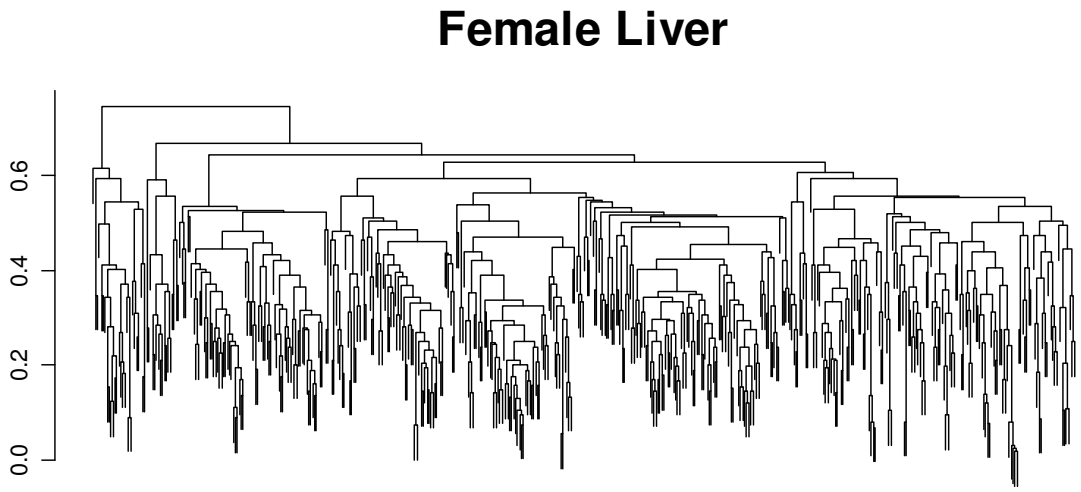
Disclaimer: Absolutely no warranty on the code. Please contact Steve Horvath (shorvath@mednet.ucla.edu), Peter Langfelder (peter.langfelder@gmail.com) or Jun Dong with suggestions.

Start of R codes

```
rm(list=ls())

setwd("C:/Documents and Settings/Steve Horvath/My
Documents/ADAG/JunDong/WebpageAugust2009")
# installation instructions for the following library are given above
library(WGCNA)
# install the following R packages from the cran webpage using the packages tab on R
library(sma)
library(impute)

# Load the data, the first row are samples names; the second row is a vector of trait, and
# the other rows are gene expression data, with each row for a gene.
```

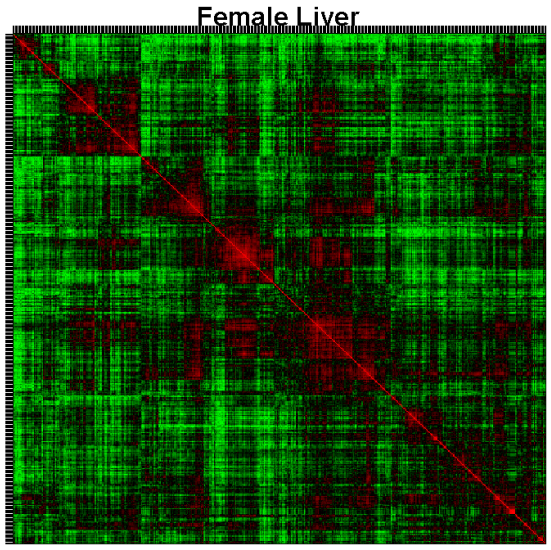
```
dat1=read.table("http://www.genetics.ucla.edu/labs/horvath/ModuleConformity/GeometricInterpretation/MouseWeightBlueFemaleLiver.xls", sep="\t", quote="", header=T);
dim(dat1); dat1[1:5,1:5];

# this is the power used for constructing the weighted network
power1=1

# Now we compute the network concepts for the Female Liver dataset
datalab= "Female Liver"
datExpr=t(dat1[-1,]); dim(datExpr);
Trait=as.numeric(dat1[1,] ) ### The first column for weight
colorh1=rep("blue", dim(datExpr)[2])
colorlevel1=levels(factor(colorh1))

# the following computes network concepts for each of the 8 co-expression networks
NC1= networkConcepts(datExpr, trait=Trait)
# ADJ hierarchical plots based on Adjacency matrix
ADJ1 = adjacency(datExpr, power=power1)
hierADJ1 = hclust(as.dist(1-ADJ1),method="average")

# This reports network concepts that only depend on the adjacency matrix
# type help(networkConcepts) to learn more details
NC1$Summary
> NC1$Summary
                  Fundamental Eigengene-based Conformity-Based
Density             0.3864613       0.3859529        0.3861584
Centralization      0.1876239       0.1914378        0.1899320
Heterogeneity       0.1803588       0.1925771        0.1846080
Mean ClusterCoef    0.4166819       0.4142770        0.4129817
Mean Connectivity 192.0712906     191.8186050      191.9207178
                  Approximate Conformity-based
Density                             0.3869620
Centralization                      0.1908601
Heterogeneity                       0.1849447
Mean ClusterCoef                    0.4130553
Mean Connectivity                 192.3201120

#this reports network concepts that depend on a trait based gene significance measure
NC1$Significance
> NC1$Significance
                       Fundamental Eigengene-based
ModuleSignificance       0.3946043       0.3932602
HubGeneSignificance      0.5872317       0.5875391
EigengeneSignificance    0.6336498              NA

## Factorizability and VarExplained by ME
c(NC1$Factorizability, NC1$VarExplained[1])
 [1] 0.9151752 0.6404968
```
Message: This correlation network has very high factorizability. The eigengene explains 64% of the variation.

```
# Now we visualize the average linkage cluster trees for each network
windows(width=8, height=4);
par(mar=c(1,2.5,5,1))
plot(hierADJ1, main= datalab, labels=F, xlab="", sub="", cex.main=2);
```



Caption: This is Figures 1A in (Horvath and Dong 2008). The Figure depicts an average linkage hierarchical cluster tree of the genes.

```
# Now we create the heatmap plots
windows();
par(mar=c(1,1, 2.5,1))
plot.mat(ADJ1[hierADJ1$order, hierADJ1$order],main= datalab, cex.main=2)
```



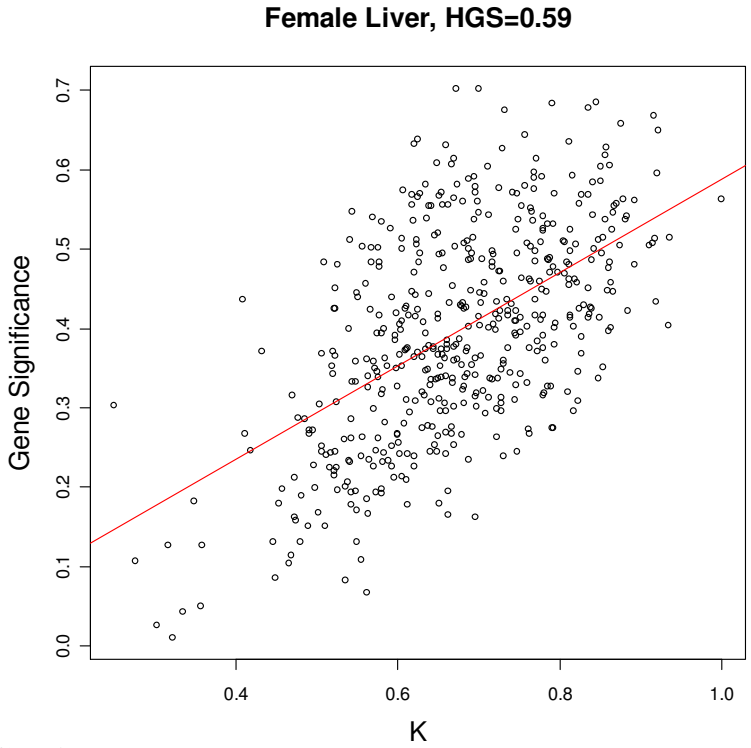
Caption: This is Figure 1B shows the corresponding heat maps which color-code the absolute pair-wise correlations a_{ij} : red

and green in the heat map indicate high and low absolute correlation, respectively. The genes in the rows and columns of each heat map are sorted by the corresponding cluster tree.

Body weight based gene significance and hub gene significance

Numerous network articles have pointed out that highly connected "hub" nodes are central to the network architecture but several authors have pointed out that hub genes may not always be biologically significant. Network theorists have long studied the relationship between gene significance and connectivity. We define a gene significance measure as the absolute correlation between the gene expression profile and body weight. The higher the gene significance, the higher is its absolute correlation with body weight.

```
par(mar=c(4,5,5,1))
K=NC1$Connectivity/max(NC1$Connectivity)
GS=NC1$GS
plot(K, NC1$GS, main= paste(datalab, ", HGS=", signif(NC1$Significance[2,1],2),
sep=""), xlab="K", ylab="Gene Significance", sub="", cex.main=1.5, cex.lab=1.5);
abline(0, NC1$Significance[2,1], col=2)
```



Caption: This is Figure 1C in our paper. The figure shows the relationship between this gene significance measure and connectivity in the different female liver network. We find a high correlation between gene significance and connectivity in the female ($r = 0.59$).

THE END