

Examples of network usage in Bioinformatics

W. Evan Durno

Initiation to microbial interaction network inference

- **See:** Faust, Karoline, and Jeroen Raes. "Microbial Interactions: From Networks to Models." *Nature Reviews Microbiology* 10, no. 8 (August 2012): 538–50. doi:10.1038/nrmicro2832.
- **Topic:** Strategies for microbial community graph construction
- **Motivation:** Make models to prediction how an entire community will function
- **Suggested methods:**
 - 16S Binary co-occurrence or correlation
 - Differential or difference equations
 - (Multiple) regression (with environmental regressors)
 - Boolean functions
 - Flux Balance Analysis: Differential equations modelling metabolism. Known to be very approximate.

Example with human microbiome

- **See:** Faust, Karoline, J. Fah Sathirapongsasuti, Jacques Izard, Nicola Segata, Dirk Gevers, Jeroen Raes, and Curtis Huttenhower. "Microbial Co-Occurrence Relationships in the Human Microbiome." PLoS Comput Biol 8, no. 7 (July 12, 2012): e1002606. doi:10.1371/journal.pcbi.1002606.
- **Goal: Exploration**
- **Criticism: Regressors were not used correctly, instead analysis was done both between and within regressor categories.**
- **Inference methods:**
 - Pearson correlation
 - Spearman correlation (on counts)
 - Generalized boosted linear model

Gene regulatory network inference model overview

- **See:** Hekker et al. "Gene regulatory network inference: Data integration in dynamic models—A review." *BioSystems* 96 (2009) 86–103
- Inference methods referenced & not described
- Data: RNA-seq counts
- Popular models described:
 - Correlation networks
 - Boolean logic networks
 - (Non-linear) differential or difference systems
 - Bayesian networks

A cultural icon: Barabasi

- **See:** Jeong, H., S. P. Mason, A.-L. Barabási, and Z. N. Oltvai. "Lethality and Centrality in Protein Networks." *Nature* 411, no. 6833 (May 3, 2001): 41–42. doi:10.1038/35075138.
- **Topic:** Protein correlation networks in yeast
- **Inference:** Pearson correlation, significantly positive.
- **Conclusions:**
 - Most connected nodes are vitally essential in yeast
 - Power law exists in degree distribution

Manual curation is also popular

- **See:** Schumm, Phillip, Caterina Scoglio, Qian Zhang, and Duygu Balcan. "Global Epidemic Invasion Thresholds in Directed Cattle Subpopulation Networks Having Source, Sink, and Transit Nodes." *Journal of Theoretical Biology* 367 (February 21, 2015): 203–21. doi:10.1016/j.jtbi.2014.12.007.
- **Topic:** Epidemics in cattle populations
- **Conclusion:** Resistance to infection and movement topology effects epidemic robustness.
- **See:** Varshney, Lav R., Beth L. Chen, Eric Paniagua, David H. Hall, and Dmitri B. Chklovskii. "Structural Properties of the *Caenorhabditis Elegans* Neuronal Network." *PLoS Comput Biol* 7, no. 2 (February 3, 2011): e1001066. doi:10.1371/journal.pcbi.1001066.
- **Topic:** *C. elegans* neural cell topology
- **Reported graph descriptive statistics.**

An example

- **See:** M. Akimoto et al. "Targeting of GFP to newborn rods by Nrl promoter and temporal expression profiling of flow-sorted photoreceptors." Proc Natl Acad Sci U S A. 2006 Mar 7;103(10):3890-5. Epub 2006 Feb 27.
- **Data:** log-Illumination values, microarray
 - 39 samples, 29949 dimensions
 - 2 categorical regressors exist
- **Illustrated:**
 - Correlation distribution before & after regression
 - Node degree before & after regression

Code

```
y = read.table("GSE4051_data.tsv",sep="\t")
x = read.table("GSE4051_design.tsv",sep="\t",header=T)

sampleCorrns = function(data=y,n=1000)
{
  J = sample( 1:nrow(data) , n )
  cr = cor( t(data[J,]) )
  as.array(cr)[ as.array(upper.tri(cr)) ]
}

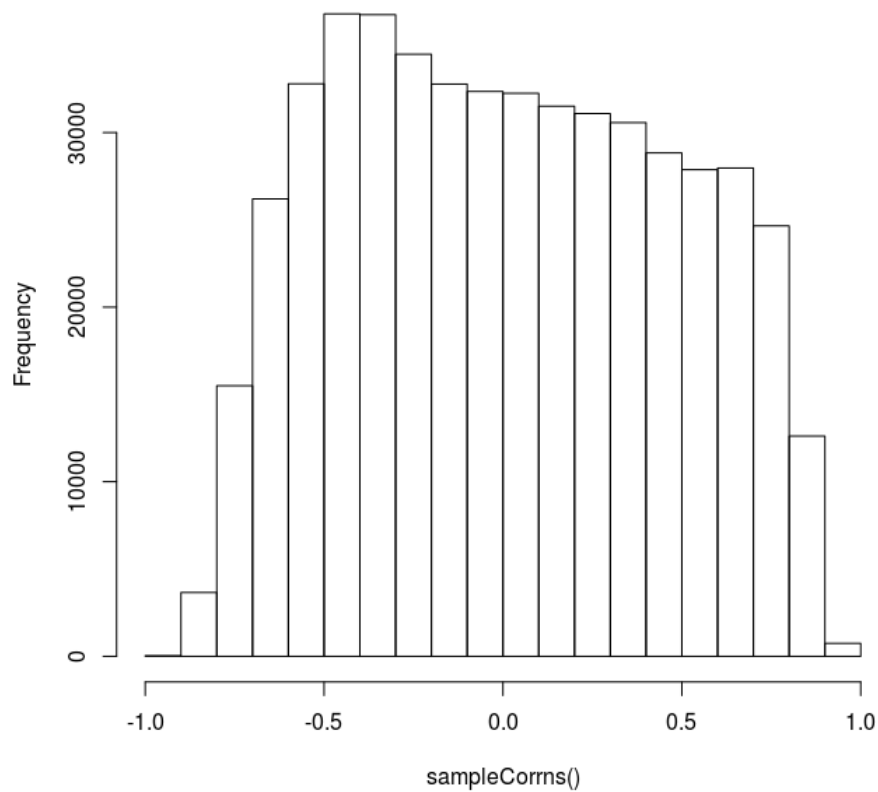
sampleDegree = function(data=y,n=1000,cut=0.25)
{
  J = sample( 1:nrow(data) , n )
  cr = cor( t(data[J,]) )
  colSums( abs(cr) > cut )
}

sampleResiduals = function(n=1000,formula= ~x[,3] + x[,4] )
{
  mat = model.matrix( formula )
  J = sample( 1:nrow(y) , n )
  lm.fit( mat , t(y[J,]) )$residuals
}

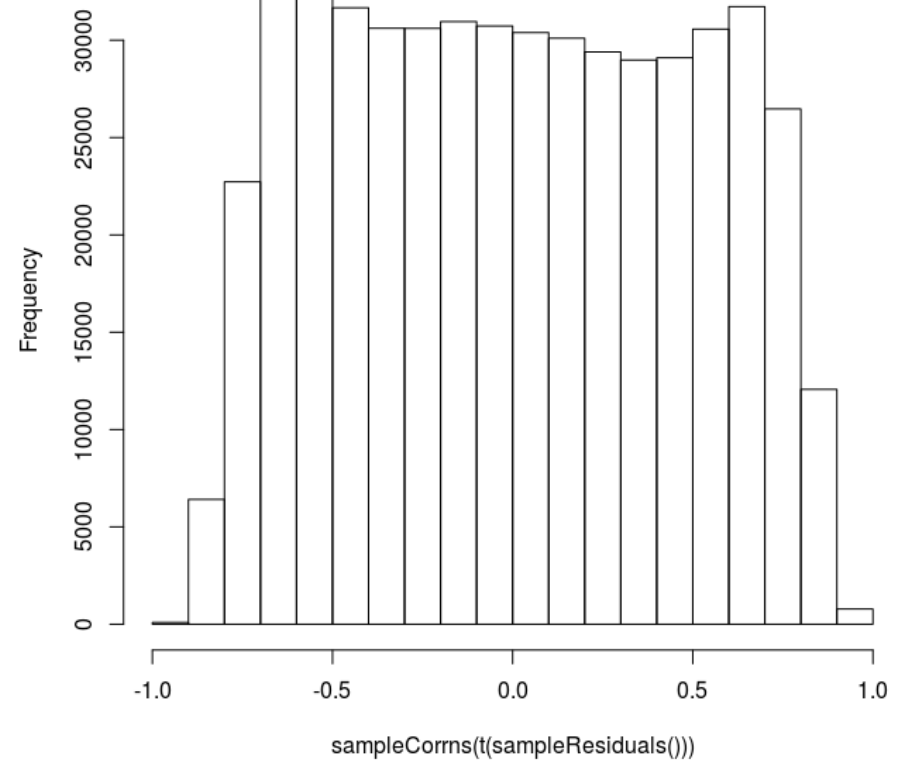
# hist( sampleCorrns() , main="Correlations before regressing" )
# hist( sampleDegree() , main="Node degree before regressing" )
# hist( sampleCorrns( t( sampleResiduals() ) ) , main="Correlations after regressing" )
# hist( sampleDegree( t( sampleResiduals() ) ) , main="Node degree after regressing" )
```


Correlations

Correlations before regressing

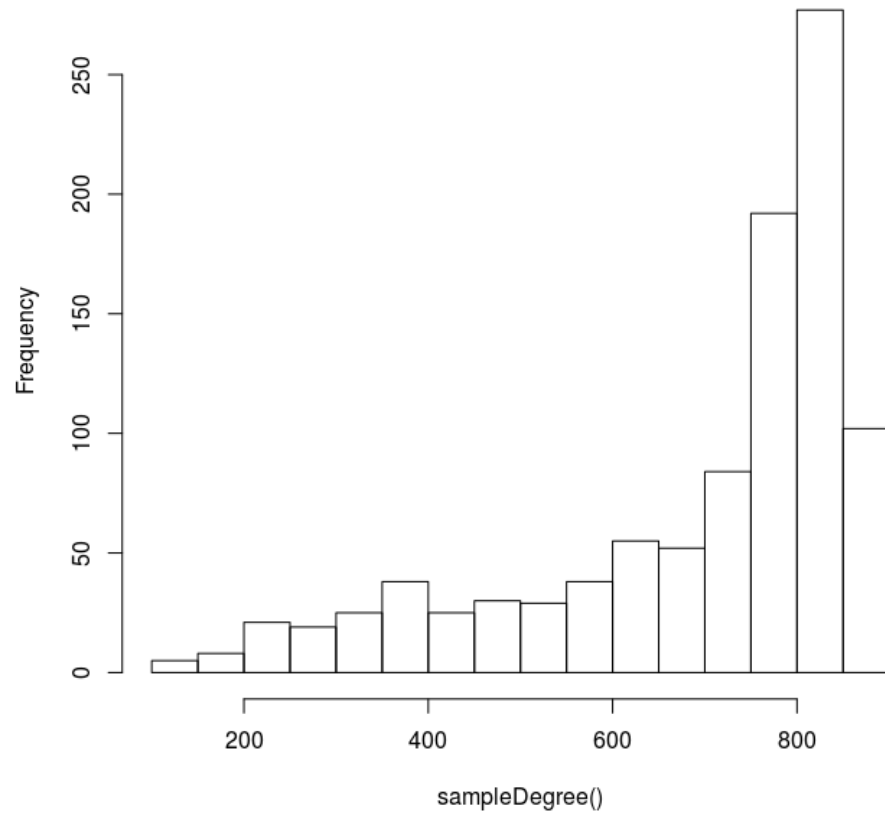


Correlations after regressing

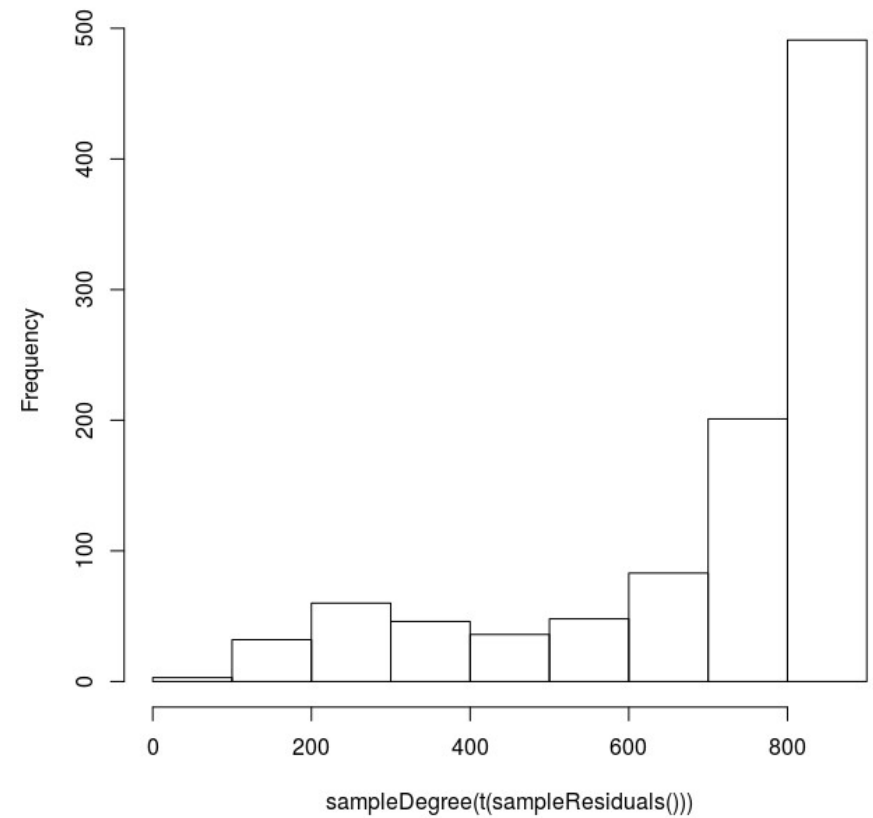


Node degree

Node degree before regressing



Node degree after regressing



Appendix

```
> # Demonstrate limitation of (p-2)-partial correlations
> f = function(n)
+ {
+   # data matrix
+   x = matrix( rnorm(5*n) , ncol=5)
+   x[,2] = 2*x[,1] + x[,2]
+   x[,3] = x[,1] + 0.5*x[,3]
+   x[,4] = 2*x[,2] + 3*x[,3] + x[,4]
+   x[,5] = 1.5*x[,5]
+
+   # partial correlations
+   # x = cov2cor( solve( cor(x) ) )
+   x = -solve( cov(x) )
+   diag(x) = -diag(x)
+   x = cov2cor(x)
+
+   # z scores
+   z = function(x) sqrt( n - 5+2-3 ) * abs( 0.5*log((1+x)/(1-x)) )
+
+   # return p-values
+   pnorm( z(x) , lower.tail=F )
+ }
> f(10000000)
      [,1]      [,2]      [,3]      [,4]      [,5]
[1,] 0.00000000 0.00000000 0.00000000 0.06748976 0.3333610
[2,] 0.00000000 0.00000000 0.00000000 0.00000000 0.1799229
[3,] 0.00000000 0.00000000 0.00000000 0.00000000 0.2469775
[4,] 0.06748976 0.00000000 0.00000000 0.00000000 0.1481506
[5,] 0.33336096 0.1799229 0.2469775 0.14815056 0.00000000
>
```