# Assortativity measures

To understand the difference between measures of community structure, I wrote R code that calculates assortativity (Newman 2003) and modularity (Sah 2014) using equations from the text and compared these values with igraph's functions which calculate these values automatically.

```
### Modular graph (Q~0.4) to compare assortativity and modularity
library(igraph)
```

```
##
## Attaching package: 'igraph'
```

```
## The following objects are masked from 'package:stats':
##
##     decompose, spectrum
```

```
## The following object is masked from 'package:base':
##
##     union
```

```
get_e_type <- function(elRow, types){
  ifelse(diff(types[elRow])==0, "w", "b")
}

setwd("~/Documents/phd/research-projects/miller-tb-assortativity/analysis/epidemic-simulations")
g <- read_graph("scalefree_Q0.4_N1000_d10_m2.graphml", format = "graphml")

plot(g, vertex.label="", vertex.size=1.5, vertex.color=V(g)$module)
```

*Assortativity* is calculated as a ratio of the number of edges within-group to the number of edges that should occur within-group by random chance.

$$r = \frac{\sum_i e_{ii} - \sum_i \left( a(i)b(i) \right)}{1 - \sum_i \left( 1 - a(i)b(i) \right)} = \frac{Tr\mathbf{E} - ||E^2||}{1 - ||E^2||}$$

where $e_{ij}$ is the fraction of edges connecting vertices of type i and j, $a(i) = \sum_j e_{(ij)}$ and $b(j) = \sum_i e_{(ij)}$.

```
## CALCULATE ASSORTATIVITY
## EXTRACT EDGE TYPES
V(g)$id <- 1:1000
el <- as_edgelist(g) # edgelist
el1 <- V(g)$module[el[, 1]]
el2 <- V(g)$module[el[, 2]]

elj <- cbind(el1, el2)
foo <- apply(elj, 1, function(x){paste0(x[1], x[2])})
table(foo)
```

```
## foo
##   00   01   11
## 2254  486 2252
```

```
## CONSTRUCT EDGE MATRIX
Eij <- matrix(c(2254, 486/2, 486/2, 2252), nrow=2, byrow = TRUE)
eij <- Eij/sum(Eij)
```
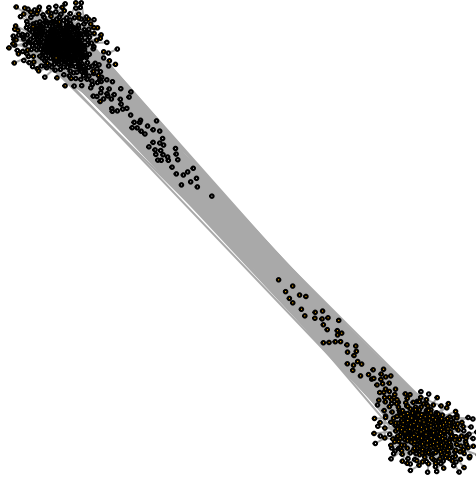
Figure 1: Modular graph (Q=0.4) generated by Sah et al. 2014 algorithm.

```
sum(eij)
```

```
## [1] 1
```

```
ai = rowSums(eij)
ai
```

```
## [1] 0.5002003 0.4997997
```

```
bi = colSums(eij)
bi
```

```
## [1] 0.5002003 0.4997997
```

```
r <- (sum(diag(eij)) - sum(ai * bi))/(1 - sum(ai * bi))
r
```

```
## [1] 0.8052884
```

```
## CHECK IF CALCULATION EQUALS IGRAPH FUNCITON
r==assortativity_nominal(g, types=(V(g)$module+1), directed=FALSE)
```

```
## [1] TRUE
```

*Modularity* is the sum of the difference in the number of within-group edges from the number of between-group edges. In other words, it is not divided by the number of edges expected to occur between groups by random chance.

```
## EQUATION 1 from SAH 2014
V(g)$id <- 1:1000
el <- as_edgelist(g) # edgelist
el1 <- V(g)$module[el[, 1]]
el2 <- V(g)$module[el[, 2]]

elj <- cbind(el1, el2)
foo <- apply(elj, 1, function(x){paste0(x[1], x[2])})
table(foo)
```

```
## foo
##   00   01   11
```

```
## 2254  486 2252
```

```r
m <- length(E(g))

# k=0  # group 0 nodes
ekk1 <- 2254/m
ak21 <- sum(degree(g, v=V(g)$id[V(g)$module==0]))/(2*m)
k1 <- ekk1 - ak21^2

# k=0  # group 1 nodes
ekk2 <- 2252/m
ak21 <- sum(degree(g, v=V(g)$id[V(g)$module==1]))/(2*m)
k2 <- ekk1 - ak21^2

# Q is the sum of these values
Q <- k1 + k2
Q
```

```
## [1] 0.4030448
```

```r
# igraph's built in function very closely matches this value
# perhaps a rounding error in my code?
modularity(g, (V(g)$module + 1))
```

```
## [1] 0.4026442
```

```r
# igraph's modularity is exactly equal to assortativity /2
r/2
```

```
## [1] 0.4026442
```

Here, we see an exact relationship between assortativity and modularity multiplied by 2. We expected modularity and assortativity to be related such that

modularity = assortativity / (1 - expected.prop.between.edges)

where the expected proportion of between edges is calculated by taking the proportion of edges "touching" each subgroup, squaring it, and summing them all up.

Since groups have equal size, you would expect 1/2 to occur between groups by chance. Thus, assortativity gets divided by while modularity does not.