

# Chapter 9: Subgroup GIMME and perturbR

*kmg*

6/25/2021

## Prepare environment

```
library(gimme)
library(perturbR)
```

## Data generating function

```
genData <- function(A = NULL,
                   Phi = NULL,
                   obs = 200,
                   n = 10 ){
  data2ls <- list()
  for (p in 1:n)
  {
    negA1 <- diag(5)-A

    time <- matrix(0, nrow = 5, ncol = obs+400)

    time1 <- matrix(0, nrow = 5, ncol = obs+400)

    noise <- solve(negA1, matrix(rnorm(5*(obs+400)),0,1), nrow = 5, ncol = obs+400))

    time[,1] <- noise[,1]

    time1[,1] <- solve(negA1, Phi) %*% time[,1] + noise[,1]

    time[,2] <- time1[,1]

    for (i in 2:(obs+400)){
      time1[,i] <- solve(negA1, Phi) %*% time[,i] + noise[,i]
      if (i<(obs+400))
        time[,i+1] <- time1[,i]
    }

    data2ls[[p]] <- t(time[,400:600])
    names(data2ls)[p] <- paste0('ind', p)
  }
  return(gen_data = data2ls)
}
```

## Data Generation

Let's use the same data generation patterns we used in Chapter 6 on model building. Here, half of the individuals have one pattern of relations and the other half have a couple of differences in their patterns.

Specifically, pattern A1 has variable 1 predict variable 4 contemporaneously and A2 variable 3 predict variable 5 contemporaneously. All other relations are the same between the two, with the exception that variable 5 predicted by 4 is positive for on subset of the data (A1) and negative for the other (see A2).

```
# Generate first pattern #
A1 <- matrix(
  c(0, 0, 0, 0, 0,
    .7, 0, 0, 0, 0,
    0, .7, 0, 0, 0,
    .7, 0, 0, 0, 0,
    0, 0, 0, .7, 0), nrow = 5, ncol = 5, byrow = TRUE)

Phi1 <- matrix(
  c(.5, 0, 0, 0, 0,
    0, .5, 0, 0, 0,
    0, 0, .5, 0, 0,
    0, 0, 0, .5, 0,
    0, 0, 0, 0, .5), nrow = 5, ncol = 5, byrow = TRUE)

Data1 <- genData(A = A1, Phi = Phi1, obs = 200)

# Generate second pattern #
A2 <- matrix(
  c(0, 0, 0, 0, 0,
    .7, 0, 0, 0, 0,
    0, .7, 0, 0, 0,
    0, 0, 0, 0, 0,
    0, 0, .7, -.7, 0), nrow = 5, ncol = 5, byrow = TRUE)

Phi2 <- matrix(
  c(.5, 0, 0, 0, 0,
    0, .5, 0, 0, 0,
    0, 0, .5, 0, 0,
    0, 0, 0, .5, 0,
    0, 0, 0, 0, .5), nrow = 5, ncol = 5, byrow = TRUE)

Data2 <- genData(A = A2, Phi = Phi2, obs = 200)

# combine data into one list #
Data_all <- append(Data1, Data2)
# make sure each individual has a unique name
names(Data_all) <- paste0('ind', seq(1,length(Data_all)))
```

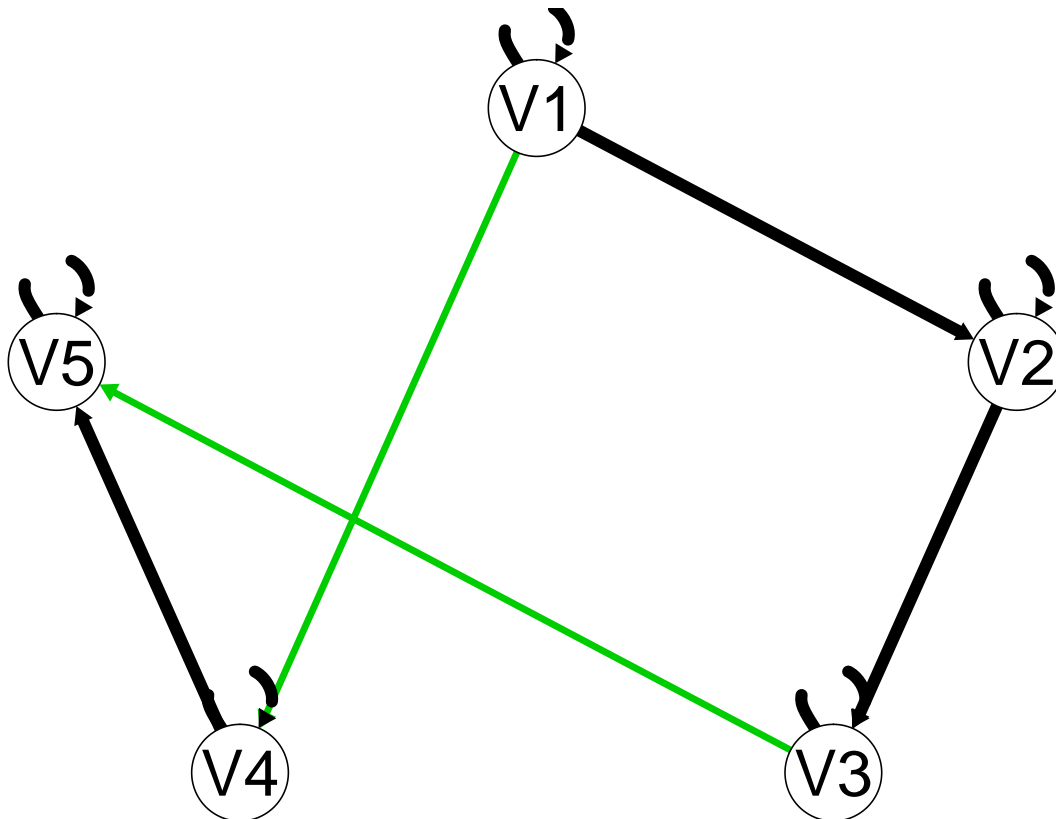
## Subgrouping GIMME

We know there are two subgroups in our data; can subgrouping GIMME find them?

```
subgroup_out <- gimme(data = Data_all,  
                      subgroup = TRUE)
```

```
plot(subgroup_out)
```

```
## Please specify a file id for individual plots. Otherwise, summary plot is presented.
```



Phew, we got 2

subgroups, just as expected. We also see that the subgroup-level paths (green) are the ones that do in fact differ according to subgroups in our data-generation code.

Let's take a look at the composition of these subgroups to make sure people are placed in the right one.

```
new <- cbind(subgroup_out$fit$file, subgroup_out$fit$sub_membership)  
new <- new[order(new[,2]),]  
new
```

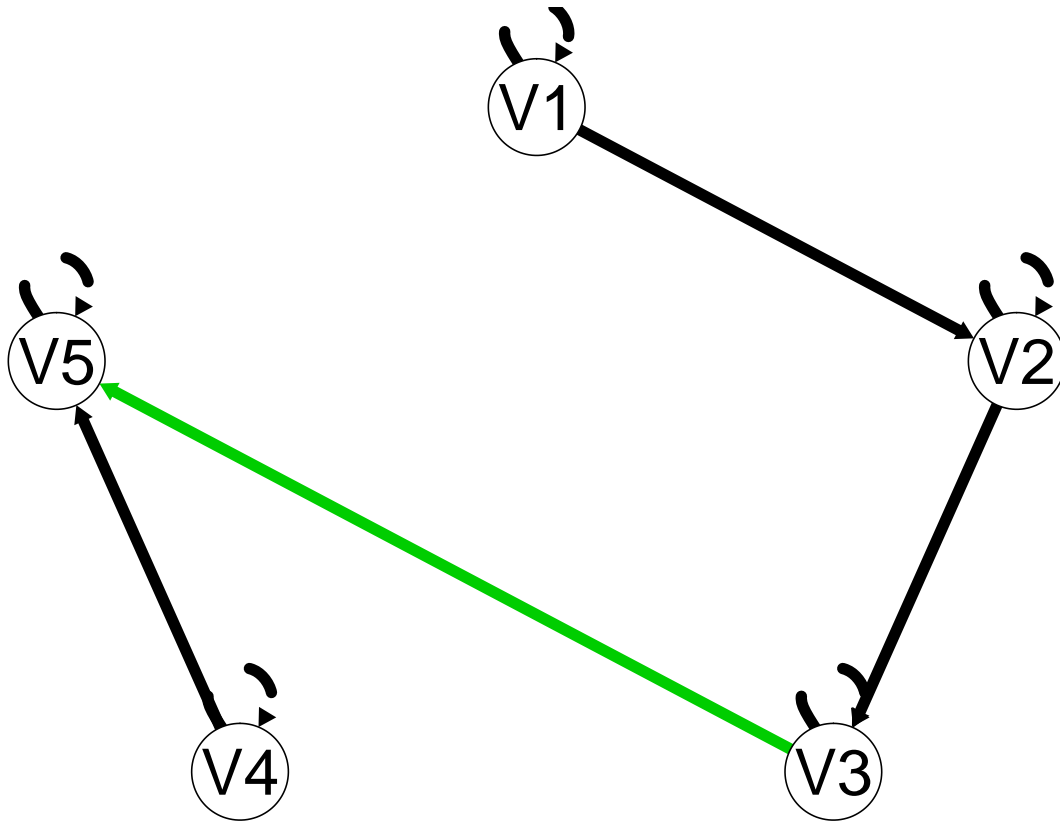
```
##      [,1]  [,2]  
## [1,] "ind11" "1"  
## [2,] "ind12" "1"  
## [3,] "ind13" "1"  
## [4,] "ind14" "1"  
## [5,] "ind15" "1"  
## [6,] "ind16" "1"  
## [7,] "ind17" "1"  
## [8,] "ind18" "1"  
## [9,] "ind19" "1"  
## [10,] "ind20" "1"
```

```
## [11,] "ind1" "2"  
## [12,] "ind10" "2"  
## [13,] "ind2" "2"  
## [14,] "ind3" "2"  
## [15,] "ind4" "2"  
## [16,] "ind5" "2"  
## [17,] "ind6" "2"  
## [18,] "ind7" "2"  
## [19,] "ind8" "2"  
## [20,] "ind9" "2"
```

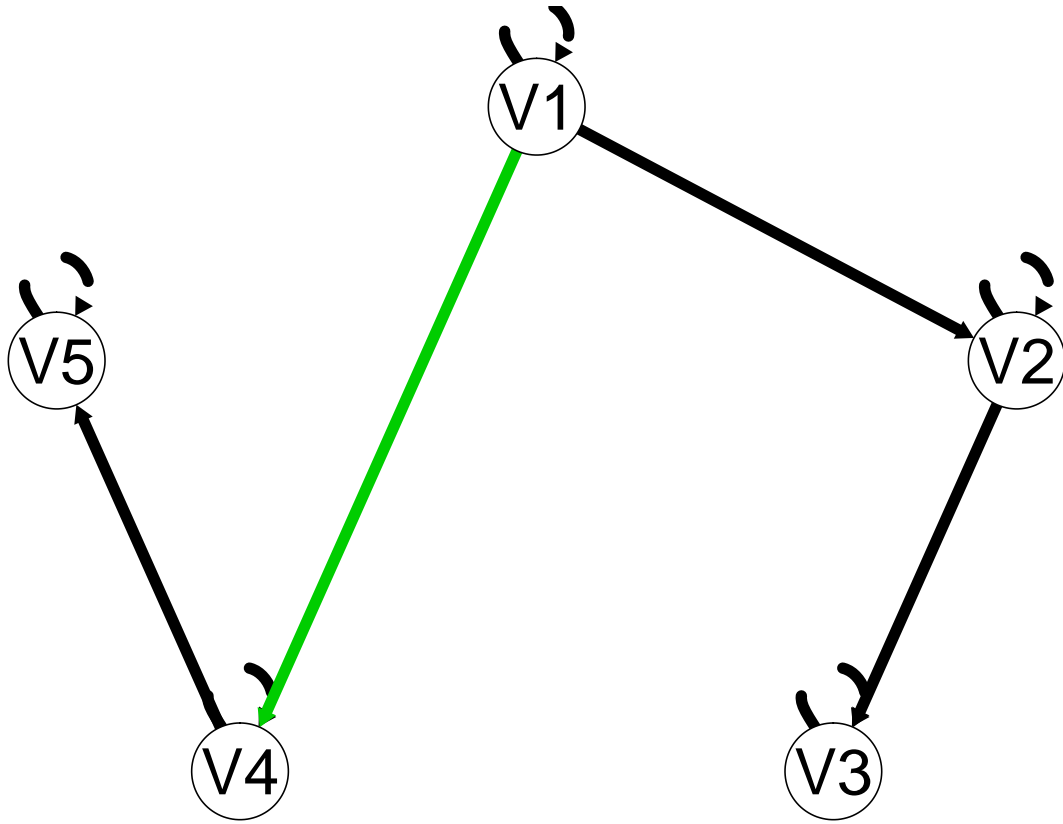
Perfect!

Let's plot the subgroups to see the paths a bit clearer.

```
plot(subgroup_out$sub_plots_paths[[1]])
```



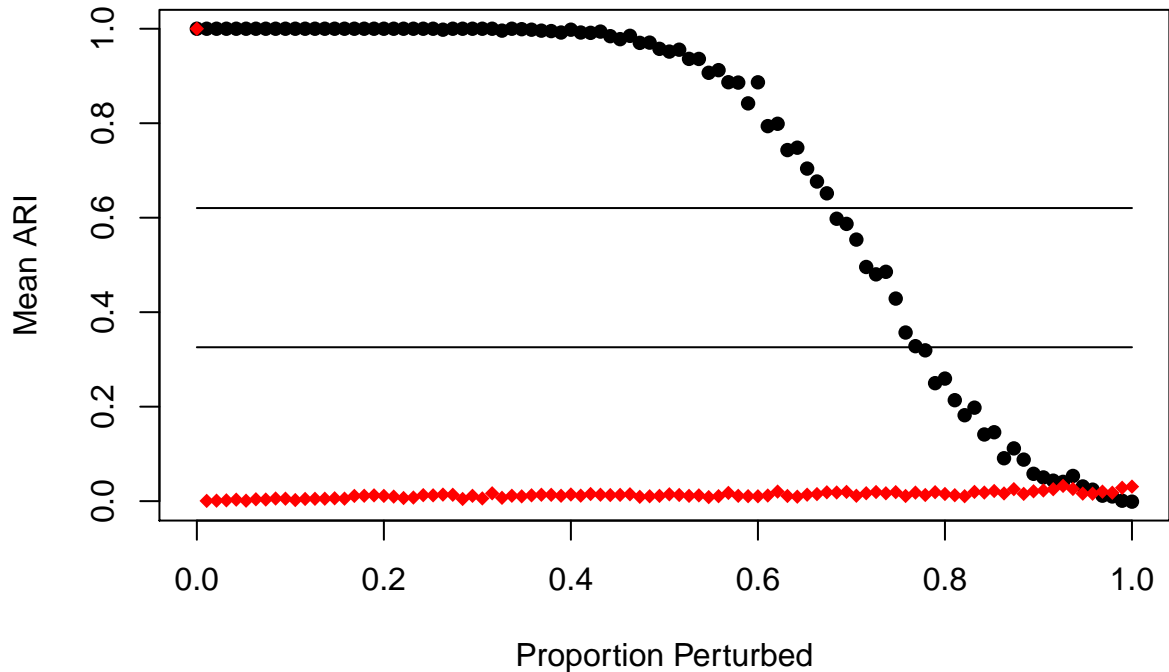
```
plot(subgroup_out$sub_plots_paths[[2]])
```



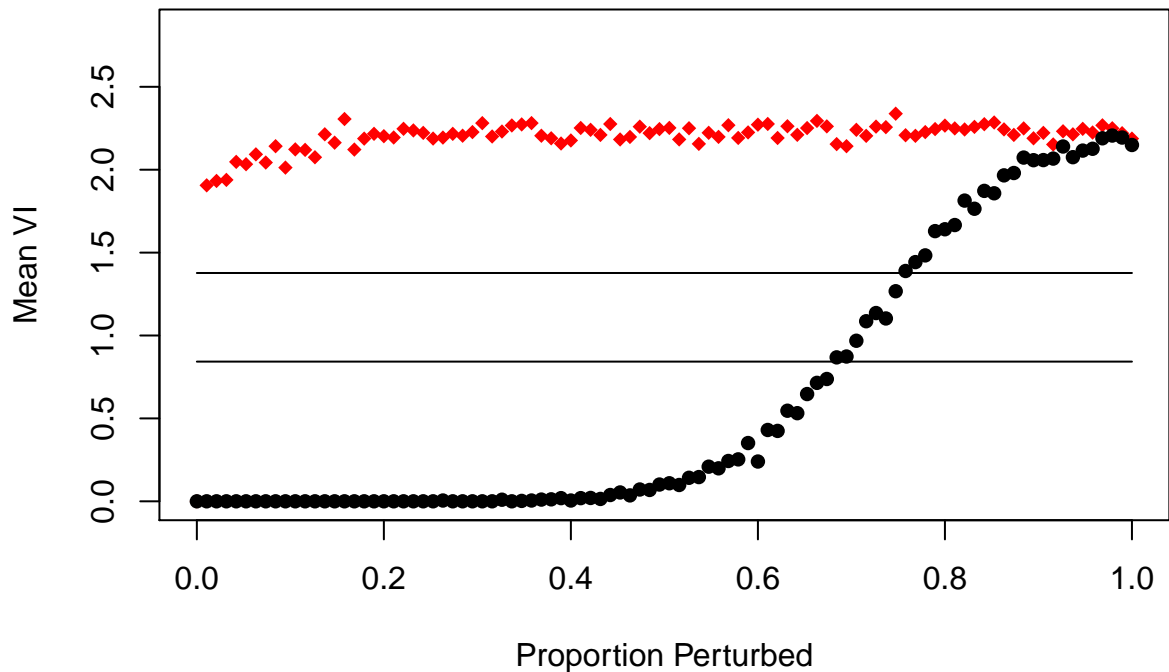
```
# Robustness of subgroup classification Let's check to see if our subgroups are robust.
```

```
test <- perturbR(subgroup_out$sim_matrix)
```

### Comparison of original result against perturbed graphs: ARI



### Comparison of original result against perturbed graphs: VI



Wow, this solution is robust. The black dots above indicate the similarity between the original subgroup solution (i.e., which individuals were in which subgroup) with the solution obtained by increasingly perturbing the matrix used to subgroup.

Robust results are ones where the black dots take a long time - at least past the  $\alpha = 0.20$  point - to intersect with the horizontal lines.

It must be noted that some solutions may not evidence robustness but still may be interesting. It is important to keep in mind your goal through subgrouping - do you want to (1) arrive at distinct subsets of people who differ in their patterns of relations, or (2) identify patterns of relations that tend to exist in some people? For the latter goal, it is not very necessary that the solutions be robust - the goal of detecting similar patterns across some subset of people can still aid in interpretation of the individual-level nuances.