

Write-up on functional correlations

Jan 25, 2024

I. Structural Connection

There are 68 regions, and so $(68 \times 68 - 68)/2 = 2278$ pairs of regions. The number of pairs of structurally connected regions is 3916, where structural connection is defined by non-zero entries in the structural data matrix.

II. Functional Connection

1. Using all time series data

In the first section, we use the entire fMRI time series data to compute pairwise covariances, unconditional correlations, partial correlations, and first-order conditional correlations. For each type of correlation measurement, we first compute on each session separately to obtain 4 sets of measurements, and then average them across sessions to get an aggregated one.

In addition, after obtaining a matrix for each type of correlation measurement, we sort them according to magnitude and record 20 pairs of regions which have the largest strengths. We will summarize and compare them in section 3.

1.1 Covariance and correlation

The covariance matrix and correlation matrix are computed. For region i, j , the covariance is $\Sigma_{ij} = \frac{1}{N} \sum_{k=1}^N (X_{ik} - \bar{X}_i)(X_{jk} - \bar{X}_j)$, where X_{ik} represents the k -th observation in data X_i . The correlation is $\rho_{ij} = \frac{\Sigma_{ij}}{\sqrt{\Sigma_{ii}\Sigma_{jj}}}$.

The top 20 pairs of regions with greatest covariance and correlation magnitude are found separately.

1.2 Partial correlation

The full partial correlation matrix is computed through $w_{ij} = \frac{-\Sigma_{ij}^{-1}}{\sqrt{\Sigma_{ii}^{-1}\Sigma_{jj}^{-1}}}$.

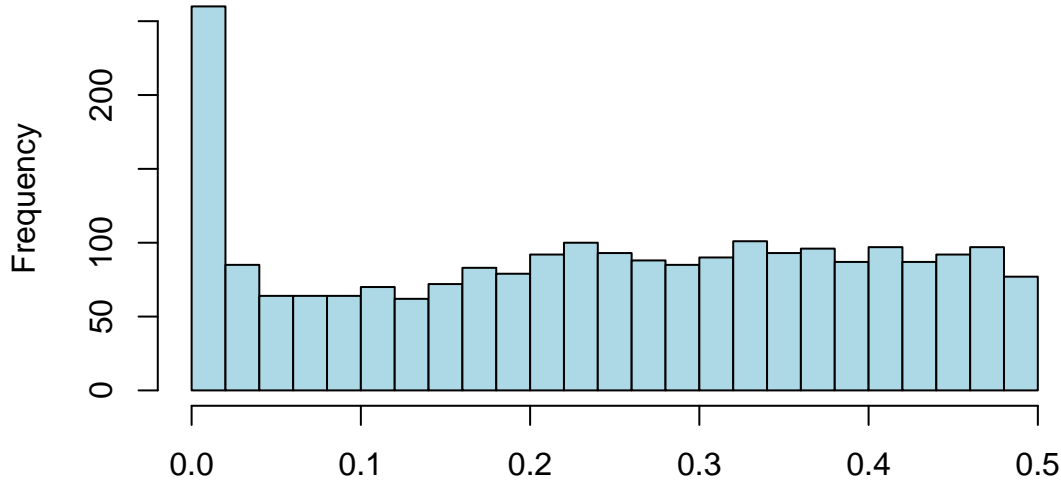
The top 20 pairs of regions with greatest partial correlation magnitude are found.

Fisher transformation By applying Fisher transformation on partial correlations here, which is $\frac{1}{2} \ln\left(\frac{1+w_{ij}}{1-w_{ij}}\right) = \text{atanh}(\rho)$, we have the standard error of $\frac{1}{\sqrt{N-3}} = \frac{1}{\sqrt{1200-3}}$. The test statistic can be calculated through $\frac{\text{atanh}(\rho)}{1/\sqrt{1197}}$, and p-values can be computed through comparing the statistic to $N(0, 1)$.

Below is a summary of the obtained p-values, which are associated with partial correlations in the upper triangular part and diagonals are removed:

```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0000  0.1085  0.2482  0.2424  0.3752  0.5000
```

Histogram of p-val (using entire sample)



In addition, for a significance level of 0.05, there are 372 p-values lower than this level, which is 0.1633 in proportion.

```
##      Count Proportion
## [1,]    372  0.1633011
```

1.3 Low-order conditional dependence

For each fMRI session, the first-order pairwise correlations are computed conditional on each region. This yields $4 \times 68 \times 68 \times 68$ arrays, where the element (i, j, k) in an array represents the conditional correlation between region i, j conditional on region k . Specifically, $\frac{\rho_{ij} - \rho_{ik}\rho_{jk}}{\sqrt{(1-\rho_{ik}^2)(1-\rho_{jk}^2)}}$. Then the 4 arrays are averaged into one.

Next, for each pair (i, j) , we find the correlation with the smallest magnitude conditional on $k = 1, \dots, 68$, and store the values in a new 68×68 matrix, where the (i, j) th element represents the smallest correlation conditional on a specific region k . The k values are stored in a separate matrix as well.

The top 20 pairs of regions with greatest first-order correlation magnitude are found.

2. Subsampled time series data

In consideration of the autocorrelation in time series data, we also try to use a subset of data in each session. In this case, only measurements at the indexes in the form of $6n + 1, n \in \mathbb{N}, 6n + 1 \leq N$ are included. Here we get a subsample containing 200 observations.

2.1 Covariance

The covariance matrix and correlation matrix are computed. The top 20 pairs of regions with greatest covariance and correlation magnitude are found separately.

2.2 Partial correlation

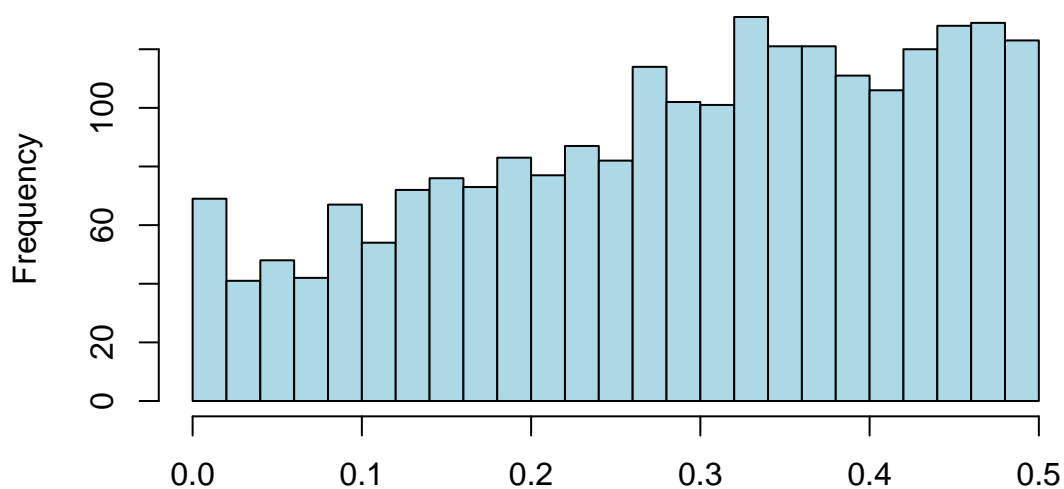
The full partial correlation matrix is computed. The top 20 pairs of regions with greatest partial correlation magnitude are found.

Fisher transformation Following the process described previously, the test statistic can again be calculated through $\frac{\text{atanh}(\rho)}{1/\sqrt{197}}$, and p-values can be computed through comparing the statistic to $N(0, 1)$.

Below is a summary of the obtained p-values, which are associated with partial correlations in the upper triangular part and diagonals are removed:

```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0000  0.1859  0.3105  0.2911  0.4069  0.4999
```

Histogram of p-val (using subsample)



In addition, for a significance level of 0.05, there are 129 p-values lower than this level, which is 0.0566 in proportion.

```
##      Count Proportion
## [1,]    129 0.05662862
```

Note here are more large p-values and fewer small/significant (at $\alpha = 0.05$) p-values, compared to the results from using the entire sample.

2.3 Low-order conditional dependence

The minimal first-order conditional correlations and the regions being conditioned on are computed and stored, following exactly the same process as in the previous section. The top 20 pairs of regions with greatest first-order correlation magnitude are found.

3. Summary and comparison

3.1 Summaries statistics

Summary statistics for different types of correlation measurements when using the entire dataset:

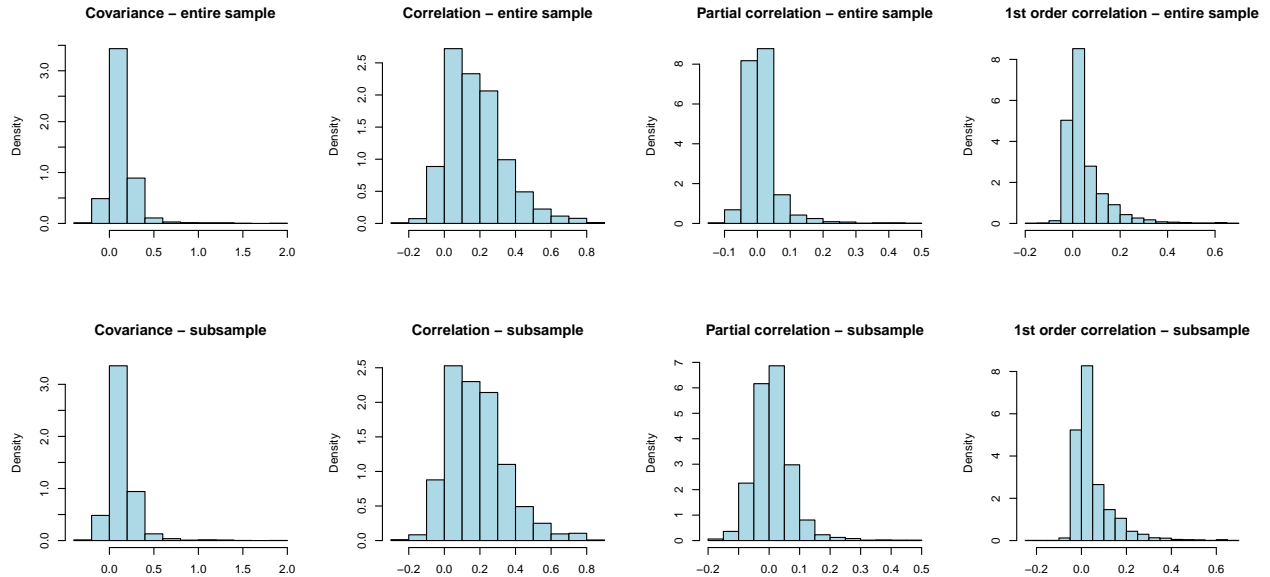
##	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
## covariance	-0.2962	0.0436	0.1015	0.1308	0.1809	1.8420
## correlation	-0.2487	0.0546	0.1573	0.1766	0.2612	0.8401
## partial cor	-0.1197	-0.0143	0.0043	0.0120	0.0251	0.4625
## 1st-order cor	-0.1587	0.0000	0.0109	0.0479	0.0697	0.6803

Summary statistics for different types of correlation measurements when using a subset of the data:

##	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
## covariance	-0.3519	0.0503	0.1076	0.1372	0.1902	1.8477
## correlation	-0.2862	0.0603	0.1626	0.1828	0.2754	0.8335
## partial cor	-0.1712	-0.0277	0.0068	0.0121	0.0427	0.4850
## 1st-order cor	-0.2016	-0.0001	0.0118	0.0491	0.0708	0.6719

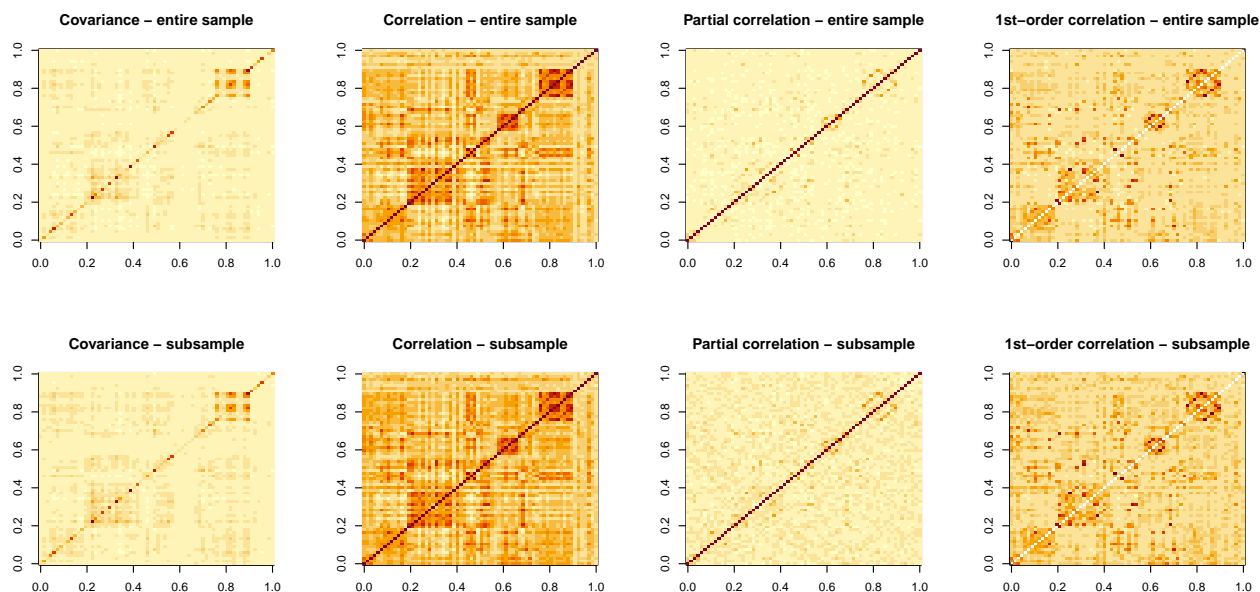
3.2 Histograms and plots

The histograms of measurement are shown below:



From the statistics and histograms, it appears that whether using the entire data or a subsample does not lead to dramatic differences in distributions, under the same type of measurement. While different types of measures are distinct in values, they all display slight right-skewness patterns.

We can take a look at the plots of all four types of correlation measurements, computed using both the entire time-series data and the subsample. Note here regions are rearranged in the order according to their communities. We can observe some clusters of strong correlations from all plots consistently.



3.3 Region connections

We can also find pairs of regions which display greatest magnitudes in covariance/correlation/partial correlation/first-order correlation. Here we show the top 20 pairs.

For the correlation measures using entire data:

##	Covariance	Correlation	Partial cor	1st order cor
## 1	20 54	4 38	28 62	28 62
## 2	38 54	20 54	10 44	24 58
## 3	46 54	24 58	21 55	21 55
## 4	4 54	28 62	20 54	4 38
## 5	9 43	21 55	4 38	20 54
## 6	12 54	38 46	24 58	10 44
## 7	20 38	38 54	55 57	3 7
## 8	4 20	4 54	41 48	37 41
## 9	4 38	37 41	7 14	7 14
## 10	20 46	12 38	37 41	41 48
## 11	12 20	4 20	26 27	17 19
## 12	38 46	4 46	30 64	23 57
## 13	4 46	7 14	7 24	7 24
## 14	12 38	4 12	21 23	30 64
## 15	12 46	7 24	26 60	41 60
## 16	4 12	46 54	41 60	38 46
## 17	9 13	12 46	3 7	7 41
## 18	9 24	41 48	38 54	55 57
## 19	9 58	20 38	16 50	41 58
## 20	9 25	41 60	3 37	14 27

For the correlation measures using a subset of the original data:

##	Covariance	Correlation	Partial cor	1st order cor
## 1	20 54	4 38	21 55	28 62
## 2	38 54	20 54	20 54	4 38

## 3	4 54	28 62	4 38	21 55
## 4	46 54	21 55	28 62	20 54
## 5	12 54	38 46	10 44	24 58
## 6	20 38	24 58	30 64	10 44
## 7	9 43	38 54	55 57	3 7
## 8	4 20	37 41	24 58	7 14
## 9	12 20	12 38	26 27	37 41
## 10	4 38	7 14	3 7	30 64
## 11	20 46	4 54	21 23	7 24
## 12	38 46	12 46	7 14	41 48
## 13	12 38	12 20	37 41	38 46
## 14	12 46	12 54	41 48	23 57
## 15	4 46	3 7	51 53	17 19
## 16	4 12	7 24	26 60	41 60
## 17	9 13	4 20	16 50	7 41
## 18	9 24	30 64	63 67	55 57
## 19	9 58	4 46	38 54	26 27
## 20	18 19	41 60	43 58	14 27

When comparing the same correlation measurement, the results of top 20 pairs using entire data and the subset do not differ too much in general, though appearing in different orders. Specifically, there is a difference of 1 pair in comparison of covariances, 5 pairs in correlation case, 3 pairs in partial correlation case, and 1 pair in 1st order correlation case.

When comparing different correlation measurements using the same set of data, results based on covariance and (unconditional) correlation share 12 common pairs in group1 and 11 in group2. While examining results of covariance and partial/1st order correlation, there are only 3 common pairs in each case. Moreover, when considering comparisons between unconditional correlations, partial correlations, and 1st order correlations, the number of common pairs falls in the range from 10 to 14, which is comparable to the case of covariances vs correlations.

Furthermore, we may investigate whether each top 20 pairs consist of regions within the same region community or not. The following table summarizes the number of pairs in top 20 results which contain regions from different communities, for each correlation measure and for the use of both entire sample and a subsample. It is clear that most pairs contain regions from the same community.

##	Covariance	Correlation	Partial cor	1st order cor
## Entire sample	0	3	5	5
## Subsample	1	3	4	6

Code Appendix

```
knitr::opts_chunk$set(echo = FALSE, warning=F, message=F)
library(readr)
library(ggplot2)
library(glmnet)
library(hdi)
library(magrittr)
library(Matrix)
library(ppcor)
library(tidyr)

n_nodes = 68      # Each subject has 68 nodes
n_sess  = 4       # Each subject has 4 fmri sessions
idx_ctx = c(1:3,5:35,36:38,40:70)
str_conn_raw = read.table(paste0("/Users/wakeup/Downloads/UW_23-25/proj_graph/100307_aparc_connectome/"
                                "100307", '_structural', '.csv'), header=F, sep=' ')
str_conn = as.matrix(sparseMatrix(str_conn_raw$V1, str_conn_raw$V2, x = str_conn_raw$V3,
                                symmetric = TRUE))[idx_ctx,idx_ctx]

# connection represented by a binary matrix
str_conn.m <- ifelse(str_conn>0,1,0) + diag(rep(1,68)) # connecting
#sum(str_conn.m==1)-68
# Helper function #####
get_sqrmat_index <- function(i,p,sym=FALSE) {
  t(sapply(X=i, FUN=function(ind) {
    i.r <- row(diag(1:p))[ind]
    i.c <- col(diag(1:p))[ind]
    if (sym==T) { c(min(i.r,i.c), max(i.r,i.c)) }
    else {c(i.r, i.c)}
  })))
}

get_large_cor <- function(m.cor,n) {
  sort(abs(m.cor),decreasing=T)[(1:n)*2]
}

# communities_tb = read.table('./communities_aparc.csv', header=T, sep=',')
if_same_comm <- function(conn) {
  t(apply(conn,MARGIN=1,FUN=function(x) {
    cbind(identical(communities_tb[communities_tb$id_ROI==x[1],"name_community"],
                    communities_tb[communities_tb$id_ROI==x[2],"name_community"]) ,
          communities_tb[communities_tb$id_ROI==x[1],"name_community"],
          communities_tb[communities_tb$id_ROI==x[2],"name_community"] ) })))
}

get_top_conn <- function(mat,n,p=68,sym=T,tol=1e-6) {
  # remove diagonal elements which are cor/cov with itself
  if (sum(is.na(diag(mat)))==0) { mat <- mat-diag(diag(mat)) }
  val.top <- get_large_cor(mat,n)
  val.index <- sapply(val.top, FUN=function(val) {which(abs(abs(mat)-val)<=tol)})[1,]
  get_sqrmat_index(val.index,p,sym)
}

count_common_pairs <- function(conn1,conn2) {
```

```

sum(apply(conn1,MARGIN=1,function(row1){
  apply(conn2,MARGIN=1,function(row2) {identical(row1,row2)}}) ))
}
load(file="/Users/wakeup/Downloads/UW_23-25/23-1Fall/Research/proj_conn.Rdata")
communities_tb = read.table('/Users/wakeup/Downloads/UW_23-25/proj_graph/100307_aparc_connectome/communities.txt')
idx_ctx = c(1:3,5:35,36:38,40:70)
ts = list() # List of time-series for the 4 available MRI scan sessions
for (id_sess in 1:4)
{
  ts[[id_sess]] = read.table(paste0("/Users/wakeup/Downloads/UW_23-25/proj_graph/100307_aparc_connectome/100307_",
                                   id_sess, ".csv"), header=F, sep=',')
}
cov.avg <- ( cov(ts[[1]])+cov(ts[[2]])+cov(ts[[3]])+cov(ts[[4]]) )/4
cov.top20.conn <- get_top_conn(cov.avg, n=20)
cor.avg <- ( cor(ts[[1]])+cor(ts[[2]])+cor(ts[[3]])+cor(ts[[4]]) )/4
cor.top20.conn <- get_top_conn(cor.avg, n=20)
# partial correlation matrix
pcor.avg <- ( pcor(ts[[1]])$estimate+pcor(ts[[2]])$estimate+
             pcor(ts[[3]])$estimate+pcor(ts[[4]])$estimate )/4
pcor.top20.conn <- get_top_conn(pcor.avg,n=20)
# Fisher transformation (entire sample)
pcor.ft <- atanh(pcor.avg)
# abs(pcor.ft)/ (1/sqrt(1200-3))
# p-values using se=1/(N-3)=1/(1200-3)
pcor.ft.pval <- pnorm(abs(pcor.ft)/ (1/sqrt(1200-3)), lower.tail=F)
summary(pcor.ft.pval[upper.tri(diag(1:68),diag=F)])
hist(pcor.ft.pval[upper.tri(diag(1:68),diag=F)],
     breaks=25,col="lightblue",xlab="",
     main="Histogram of p-val (using entire sample)")
cbind("Count"=sum(pcor.ft.pval[upper.tri(diag(1:68),diag=F)]<0.05),
      "Proportion"=mean(pcor.ft.pval[upper.tri(diag(1:68),diag=F)]<0.05))
pcor.order1.top20.conn <- get_top_conn(pcor.order1.avg.min$m,n=20)
# ts.sub6
cov.sub6.avg <- ( cov(ts.sub6[[1]])+cov(ts.sub6[[2]])+cov(ts.sub6[[3]])+cov(ts.sub6[[4]]) )/4
cor.sub6.avg <- ( cor(ts.sub6[[1]])+cor(ts.sub6[[2]])+cor(ts.sub6[[3]])+cor(ts.sub6[[4]]) )/4
cov.sub6.top20.conn <- get_top_conn(cov.sub6.avg,n=20)
cor.sub6.top20.conn <- get_top_conn(cor.sub6.avg,n=20)
pcor.sub6.avg <- (pcor(ts.sub6[[1]])$estimate+pcor(ts.sub6[[2]])$estimate+
                 pcor(ts.sub6[[3]])$estimate+pcor(ts.sub6[[4]])$estimate)/4
pcor.sub6.top20.conn <- get_top_conn(pcor.sub6.avg,n=20)
# Fisher transformation
pcor.sub6.ft <- atanh(pcor.sub6.avg)
# abs(pcor.ft)/ (1/sqrt(200-3))
# p-values using se=1/(N-3)=1/(200-3)
pcor.sub6.ft.pval <- pnorm(abs(pcor.sub6.ft)/ (1/sqrt(200-3)), lower.tail=F)
summary(pcor.sub6.ft.pval[upper.tri(cov.avg,diag=F)])
hist(pcor.sub6.ft.pval[upper.tri(cov.avg,diag=F)],
     breaks=25,col="lightblue",xlab="",
     main="Histogram of p-val (using subsample)")
cbind("Count"=sum(pcor.sub6.ft.pval[upper.tri(cov.avg,diag=F)]<0.05),
      "Proportion"=mean(pcor.sub6.ft.pval[upper.tri(cov.avg,diag=F)]<0.05))
pcor.sub6.order1.top20.conn <- get_top_conn(pcor.sub6.order1.avg.min$m,n=20)
rbind("covariance"=summary(cov.avg[upper.tri(diag(1:68),diag=F)]),

```



```

    "correlation"=summary(cov.avg[upper.tri(diag(1:68),diag=F)]),
    "partial cor"=summary(pcor.avg[upper.tri(diag(1:68),diag=F)]),
    "1st-order cor"=summary(pcor.order1.avg.min$m[upper.tri(diag(1:68),diag=F)]) %>% round(4)
rbind("covariance"=summary(cov.sub6.avg[upper.tri(cov.sub6.avg,diag=F)]),
    "correlation"=summary(cov.sub6.avg[upper.tri(cov.sub6.avg,diag=F)]),
    "partial cor"=summary(pcor.sub6.avg[upper.tri(pcor.sub6.avg,diag=F)]),
    "1st-order cor"=summary(pcor.sub6.order1.avg.min$m[upper.tri(pcor.sub6.avg,diag=F)]) %>% round(4)
par(mfrow=c(2,4),cex.main=1.3)
hist(cov.avg[upper.tri(diag(1:68),diag=F)],prob=T,
    main="Covariance - entire sample",xlab="",col="lightblue")
hist(cov.sub6.avg[upper.tri(diag(1:68),diag=F)],prob=T,
    main="Covariance - subsample",xlab="",col="lightblue")
hist(cor.avg[upper.tri(diag(1:68),diag=F)],prob=T,
    main="Correlation - entire sample",xlab="",col="lightblue")
hist(cor.sub6.avg[upper.tri(diag(1:68),diag=F)],prob=T,
    main="Correlation - subsample",xlab="",col="lightblue")
hist(pcor.avg[upper.tri(diag(1:68),diag=F)],prob=T,
    main="Partial correlation - entire sample",xlab="",col="lightblue")
hist(pcor.sub6.avg[upper.tri(diag(1:68),diag=F)],prob=T,
    main="Partial correlation - subsample",xlab="",col="lightblue")
hist(pcor.order1.avg.min$m[upper.tri(diag(1:68),diag=F)],prob=T,
    main="1st order correlation - entire sample",xlab="",col="lightblue")
hist(pcor.sub6.order1.avg.min$m[upper.tri(diag(1:68),diag=F)],prob=T,
    main="1st order correlation - subsample",xlab="",col="lightblue")
par(mfrow=c(2,4),cex.main=1.3)
image(cov.avg[communities_tb$id_ROI,communities_tb$id_ROI],
    main="Covariance - entire sample")
image(cov.sub6.avg[communities_tb$id_ROI,communities_tb$id_ROI],
    main="Covariance - subsample")
image(cor.avg[communities_tb$id_ROI,communities_tb$id_ROI],
    main="Correlation - entire sample")
image(cor.sub6.avg[communities_tb$id_ROI,communities_tb$id_ROI],
    main="Correlation - subsample")
image(pcor.avg[communities_tb$id_ROI,communities_tb$id_ROI],
    main="Partial correlation - entire sample")
image(pcor.sub6.avg[communities_tb$id_ROI,communities_tb$id_ROI],
    main="Partial correlation - subsample")
image(pcor.order1.avg.min$m[communities_tb$id_ROI,communities_tb$id_ROI],
    main="1st-order correlation - entire sample")
image(pcor.sub6.order1.avg.min$m[communities_tb$id_ROI,communities_tb$id_ROI],
    main="1st-order correlation - subsample")
tab.top20.conn <- cbind(cov.top20.conn, cor.top20.conn,
    pcor.top20.conn, pcor.order1.top20.conn)
colnames(tab.top20.conn) <- c("Covariance","", "Correlation","",
    "Partial cor","", "1st order cor","")
rownames(tab.top20.conn) <- 1:20
tab.top20.conn
tab.top20.sub6.conn <- cbind(cov.sub6.top20.conn,cor.sub6.top20.conn,
    pcor.sub6.top20.conn,pcor.sub6.order1.top20.conn)
colnames(tab.top20.sub6.conn) <- c("Covariance","", "Correlation","",
    "Partial cor","", "1st order cor","")
rownames(tab.top20.sub6.conn) <- 1:20

```

```

tab.top20.sub6.conn
tab.commdiff <- rbind(c(sum(if_same_comm(cov.top20.conn)[,1]!="TRUE"),
                        sum(if_same_comm(cor.top20.conn)[,1]!="TRUE"),
                        sum(if_same_comm(pcor.top20.conn)[,1]!="TRUE"),
                        sum(if_same_comm(pcor.order1.top20.conn)[,1]!="TRUE))),
                    c(sum(if_same_comm(cov.sub6.top20.conn)[,1]!="TRUE"),
                      sum(if_same_comm(cor.sub6.top20.conn)[,1]!="TRUE"),
                      sum(if_same_comm(pcor.sub6.top20.conn)[,1]!="TRUE"),
                      sum(if_same_comm(pcor.sub6.order1.top20.conn)[,1]!="TRUE")))
colnames(tab.commdiff) <- c("Covariance", "Correlation", "Partial cor", "1st order cor")
rownames(tab.commdiff) <- c("Entire sample", "Subsample")
tab.commdiff

```