Writeup on comparing structural and functional connectivity (May28)

Contents

1 Methods

1.1 Network construction

Procedure for most networks: Construct networks for every individual, aggregate them and compute the rate of being selected for every edge. This gives a matrix where entry (i, j) represents the proportion of drawing an edge between region i and j. Based on this proportion matrix, I set 136 largest entries to 1 and leave the rest 0.

Note1: A few networks cannot reach exactly 136 edges due to ties in proportions.

Note2: All networks are constructed based on 957 subjects.

Structural:

Global thresholding: *[sum_and_thre] [thre_and_sum]*

1. sum and threshold, 2. threshold and sum

(Note related to (1): in previous analysis I tried normalization on each subject before sum and threshold, and the results differ only in 2 edges. To keep things simple I omit this normalization step.)

Local thresholding/disparity filter : *[sum_and_thre] [thre_and_sum]*

1. sum and threshold, 2. threshold and average

Sign test: *[thre_dens_p035] [thre_count_p027]*

1. Fix a count threshold = 32500 for every subject, binarize for every individual based on the threshold, and sum over the networks. Perform binomial sign test using prob=0.027 and get p-values, set entries $\leq \alpha$ to be 1 and others 0 (Bonferroni corrected $\alpha = 0.05/2278$).
Note: The probability parameter comes from the following procedure: if we binarize every subject's structural connectomic data by 32500, the mean and median of network density are both around 0.027 (equivalently, estimate $P(\text{connectomic count} \geq 32500) = 0.027$).

2. Fix network density = $80/2278 \approx 0.035$ for every subject, binarize for every individual at the density, and sum over the networks. Perform binomial sign test using prob= 0.035 and get p-values, set entries $\leq \alpha$ to be 1 and others 0 (Bonferroni corrected $\alpha = 0.05/2278$).

(Note: This method does not involve proportions of edge selection. These two networks are constructed based on the cutoff values and probabilities different from previous versions. These values are chosen to make the final aggregated network have approximately 136 edges, though the second one can only achieve 138 as the closest.)

Functional:

Below describes approaches to compute edge selection proportion matrix. Specifically, magnitude-based means individual networks are constructed based on 136 *greatest connection strength, significance-based means individual networks are constructed based on whether p-values are below* $\alpha = 0.05/2278$. *1200/200 time points indicates number of fMRI observations used.*

Marginal correlation: *(mcor)*

{Magnitude-based, Significance-based} × {1200 time points, 200 time points}

Partial correlation: *(pcor)*

{Magnitude-based, Significance-based} × {1200 time points, 200 time points}

Low-order correlation: *(pcor_ord1)*

{Magnitude-based, Significance-based} × {1200 time points, 200 time points}

Graphical lasso: *(gl)*

{1200 time points, 200 timepoints}

1.2 Network comparison

- Within the same type of connectivity matrices (structural/functional) - summaries between different approaches
- Structural vs Functional - edit distance between binary connection networks constructed from edge selection proportion matrices (each ≈ 136 edges)
- Structural vs Functional - compute matrices containing shortest path lengths between every pair of regions using Dijkstra's algorithm. Focus on FC at positions where SC is absent - referred to as false positive (FP) in later sections. *(Absence/FP will be defined either through the network obtained from proportion matrices, or through the shortest path length matrices.)*

2 Network summary

2.1 Structural connectivity summary

Connection by proportions and shortest paths

A summary table of edges in structural connectivity networks. *bin_nw* represents binary networks constructed based on edge selection proportion matrices. *path_nw* represents connection based on shortest path length matrices, where two regions are considered as connected if their shortest path length is nonzero.

	global_thre		**local_thre**			**sign_test**	
	bin_nw	*path_nw*	*bin_nw*	*path_nw*		*bin_nw*	*path_nw*
sum_and_thre	136	1540	136	2278	*thre_count_p027*	136	1431
thre_and_sum	136	1540	136	2278	*thre_dens_p035*	138	1540

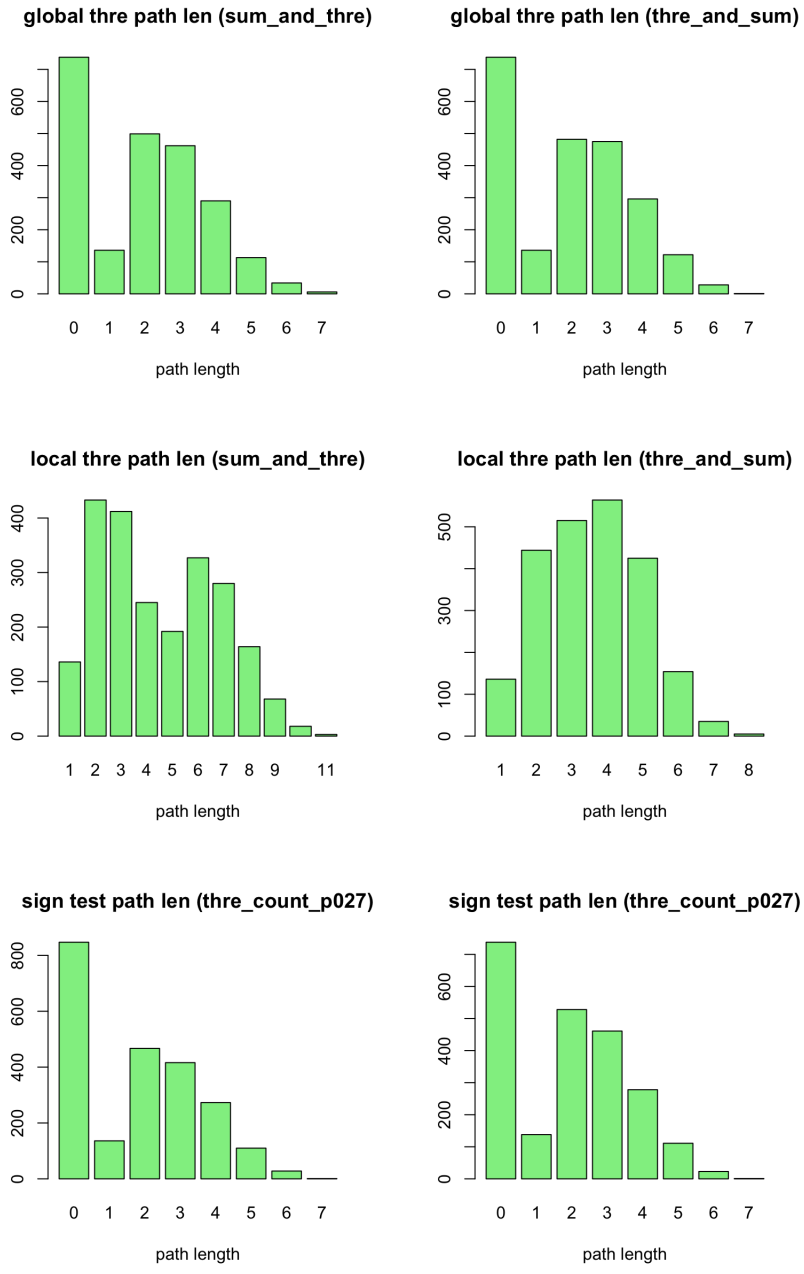
Table 1: SC number of edges/nonzero pairs

Regions in the results from local thresholding approaches are pairwisely connected by at least one path.

For *path_nw* in the above table, all 1540's have the same positions of nonzeros (two global thresholding results and sign test *thre_dens_p035*).

Shortest path lengths

Distributions of shortest path lengths are shown in the barplots, all upper-triangular parts in matrices.



Two approaches in local thresholding show distinct patterns.

2.2 Functional connectivity summary

Connection by proportions and shortest paths

	mar_cor		**par_cor**		**pcor_ord1**		**glasso**	
	bin_nw	*path_nw*	*bin_nw*	*path_nw*	*bin_nw*	*path_nw*	*bin_nw*	*path_nw*
mag_all	136	990	136	1771	136	1434	136	1035
mag_sub6	136	990	136	1771	136	1381	136	1081
sig_all	137	1005	136	1771	136	1830	-	-
sig_sub6	136	867	138	1655	136	1381	-	-

Table 2: FC number of edges/nonzero pairs

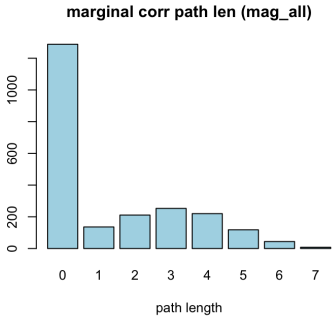
Interpretation of the table: *e.g. the [mag_all, mar_cor] part: Consider the functional connectivity network constructed using* 1200 *time points with edges corresponding to greatest marginal correlations in magnitude. The binary network based on proportion matrix has* 136 *edges, meaning* 136 *region pairs are connected, i.e. entry=1. If we compute a matrix containing pairwise shortest path lengths,* 990 *pairs are connected through at least one path, i.e. nonzero/positive entry.*

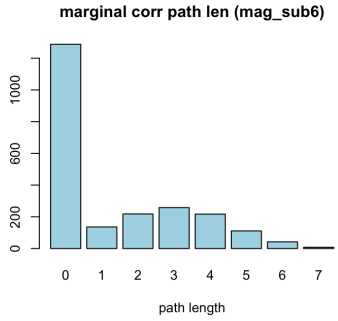
Same locations of nonzero entries in shortest path networks *(path_nw)*:

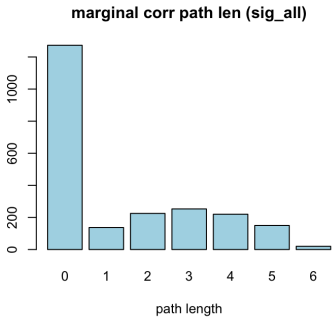
- mcor *mag_all* and *mag_sub6* are identical in nonzero positions, different in path length distribution though (990 nonzero entries)
- pcor *mag_all*, *mag_sub6*, and *sig_all* are identical in nonzero positions, different in path length distribution though (1771 nonzero entries)
- pcor_order1 *mag_sub6* and *sig_sub6* are identical in nonzero positions, different in path lengths though (1381 nonzero entries)

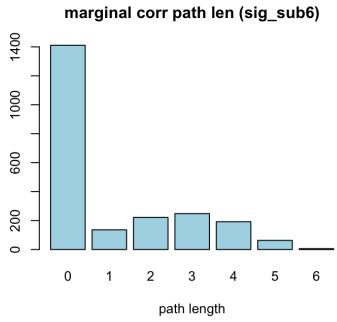
Shortest path lengths

Marginal correlation-based

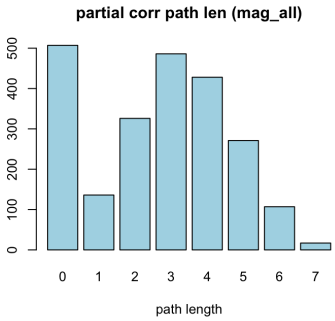


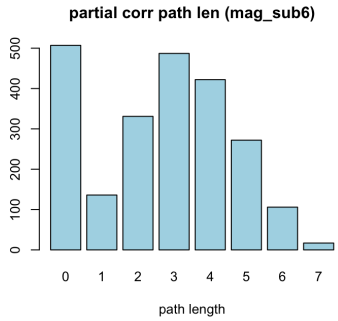


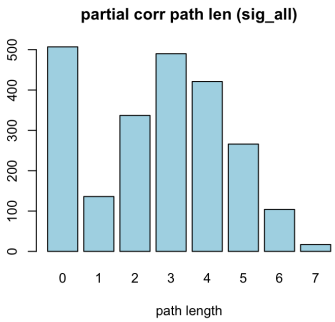


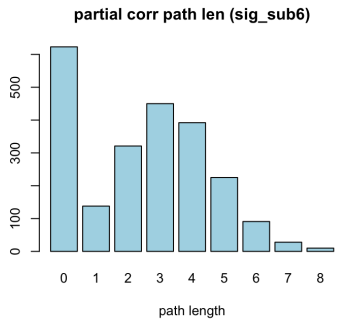


Partial correlation-based

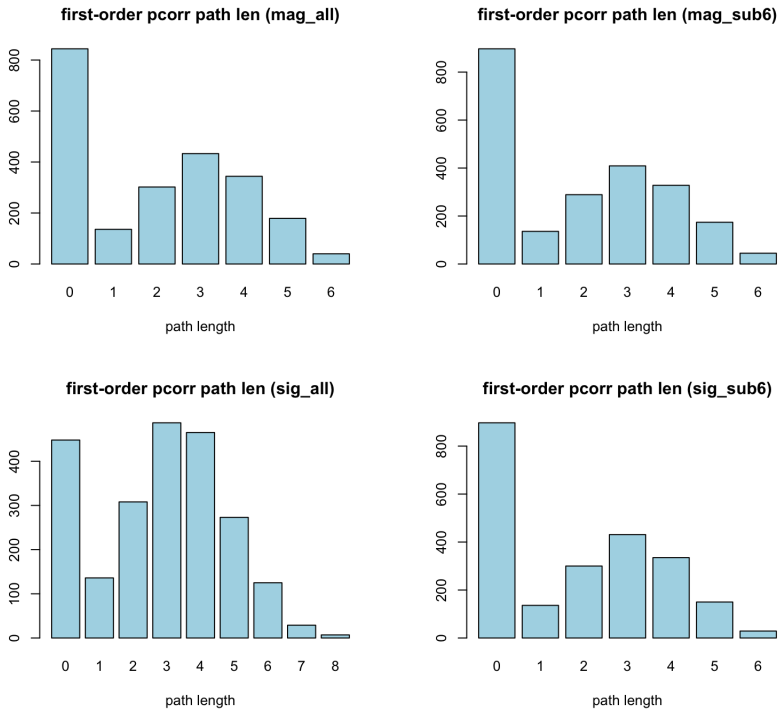




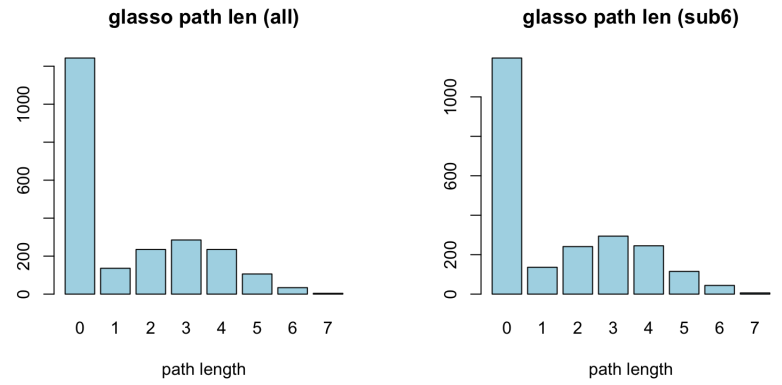




First-order partial correlation-based



Graphical lasso-based



3 Network comparison

3.1 About edge selection proportions

In most functional connectivity network construction methods, (28,62) is selected in functional connectivity, by every subject in every session. While this pair is also selected by global thresholding and sign test, it is not selected by local thresholding (surprisingly).

In marginal correlation significance-based network, using all 1200 time points, this edge is not drawn as well. However, the edges between (62,57) and (57,28) are always selected by every subject every fMRI session, using this method.

While I have not thoroughly looked into this aspect, this may provide some additional interesting information.

3.2 Edit distance

Edit distance between every binary structural and binary functional network, using the upper triangular part only. Two separate tables are shown for 1200 time points and 200 time points, with an attempt of visualization.

Local thresholding approaches give more deviated results compared to other SC construction methods.

		Global threshold		**Local threshold**		**Sign test**	
		sum_thre	*thre_sum*	*sum_thre*	*thre_sum*	*count_p027*	*dens_p035*
mcor	*mag*	248	246	262	256	246	248
	sig	243	241	255	249	241	241
pcor	*mag*	246	244	252	246	244	244
	sig	246	244	250	244	244	244
pcor_ord1	*mag*	248	246	250	246	246	246
	sig	248	246	254	250	246	246
glasso	*glasso*	246	244	260	254	244	246

Table 3: Edit distance (upper triangle, 1200 time points)

		Global threshold		Local threshold		Sign test	
		sum_thre	*thre_sum*	*sum_thre*	*thre_sum*	*count_p027*	*dens_p035*
mcor	*mag*	248	246	262	256	246	248
	sig	248	246	258	252	246	246
pcor	*mag*	244	242	250	244	242	242
	sig	248	246	252	248	246	246
pcor_ord1	*mag*	248	246	250	246	246	246
	sig	250	248	252	248	248	248
glasso	*glasso*	246	244	260	254	244	246

Table 4: Edit distance (upper triangle, 200 time points)

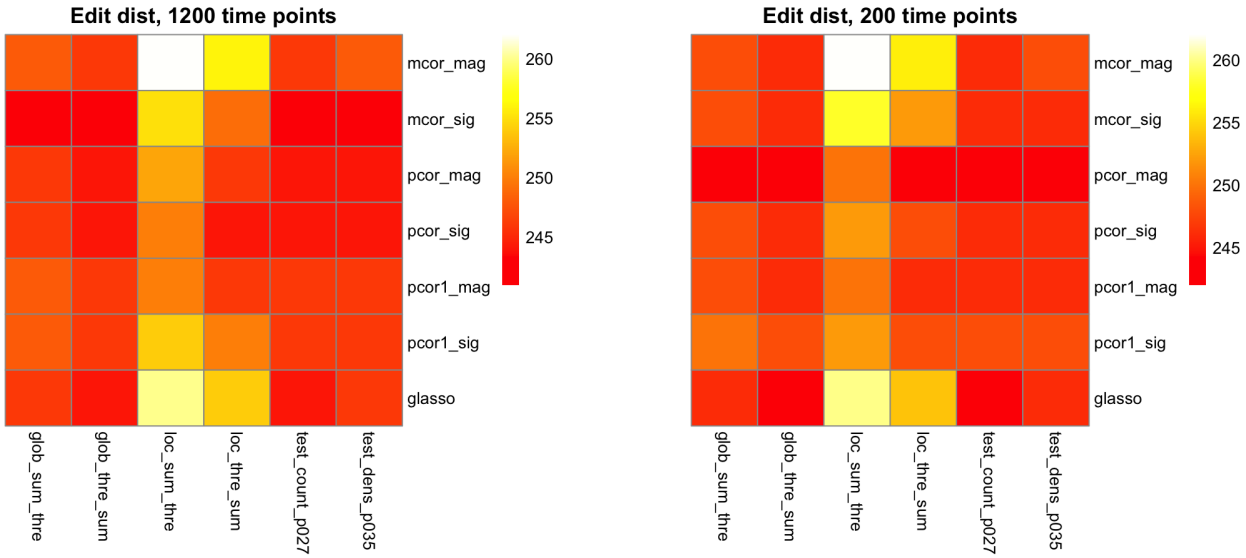
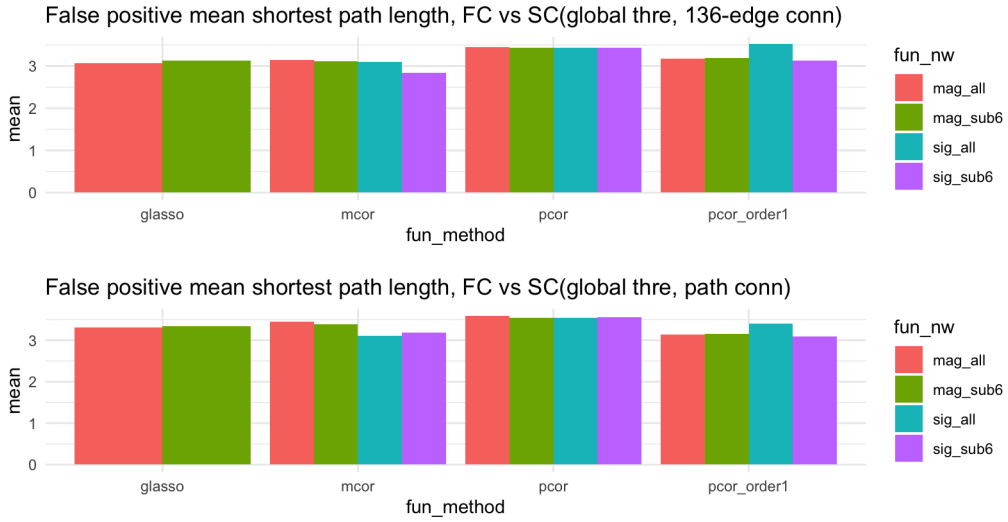


Figure 1: A visualization of edit distance matrices

3.3 False positive pairwise shortest path length

Here we focus on false positive edges in FC networks using the SC network as a reference. We use *false positive* (FP) to represent the case that an edge exists in FC while absent in SC. This can be defined with respect to either the 136-edge SC network based on proportion matrix, or the shortest path matrix. In the former case, FP occurs between regions that are not connected in SC; in the latter case, FP occurs when there is not path connecting the two regions.

Mean of FP shortest path lengths are mostly around 3, regardless of which SC network is used. Here are two examples of using global thresholded SC and all FCs. Figures of SC using local thresholding and sign test can be found at the end of this write-up.



3.4 False positive edges

3.4.1 SC based on edge selection proportion

Number of FP edges are summarized in tables below. Here the SC networks all contain 136 edges based on edge selection proportion matrices. FPs are calculated using upper triangular parts only. Two separate tables are shown for FCs using 1200 and 200 time points.

		Global threshold		**Local threshold**		**Sign test**	
		sum_thre	*thre_sum*	*sum_thre*	*thre_sum*	*count_p027*	*dens_p035*
mcor	*mag*	124	123	131	128	123	123
	sig	122	121	128	125	121	120
pcor	*mag*	123	122	126	123	122	121
	sig	123	122	125	122	122	121
pcor_order1	*mag*	124	123	125	123	123	122
	sig	124	123	127	125	123	122
glasso	*glasso*	123	122	130	127	122	122

Table 5: False positives (upper triangle, 1200 time points)

		Global threshold		**Local threshold**		**Sign test**	
		sum_thre	*thre_sum*	*sum_thre*	*thre_sum*	*count_p027*	*dens_p035*
mcor	*mag*	124	123	131	128	123	123
	sig	124	123	129	126	123	122
pcor	*mag*	122	121	125	122	121	120
	sig	125	124	127	125	124	123
pcor_order1	*mag*	124	123	125	123	123	122
	sig	125	124	126	124	124	123
glasso	*glasso*	123	122	130	127	122	122

Table 6: False positives (upper triangle, 200 time points)

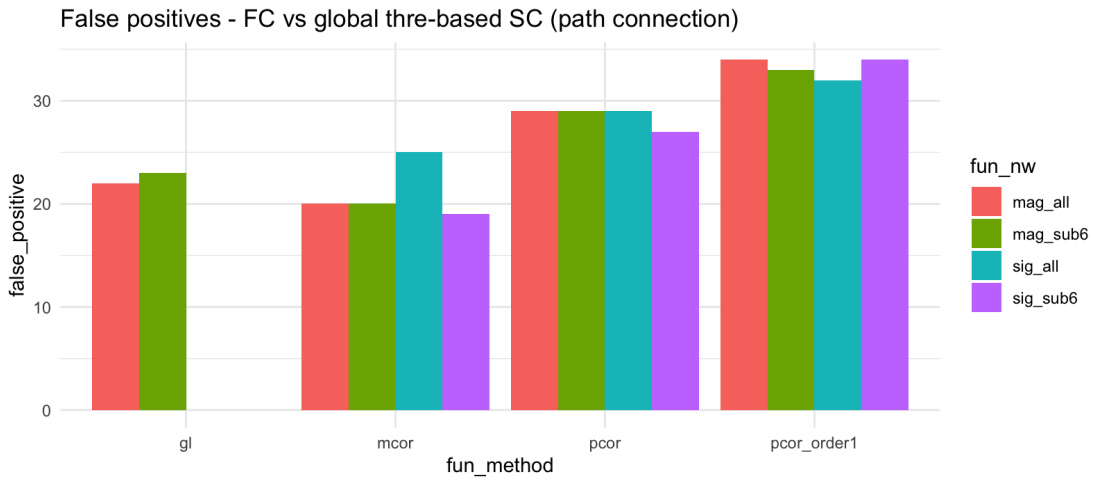
All numbers are approximately half of the edit distances.

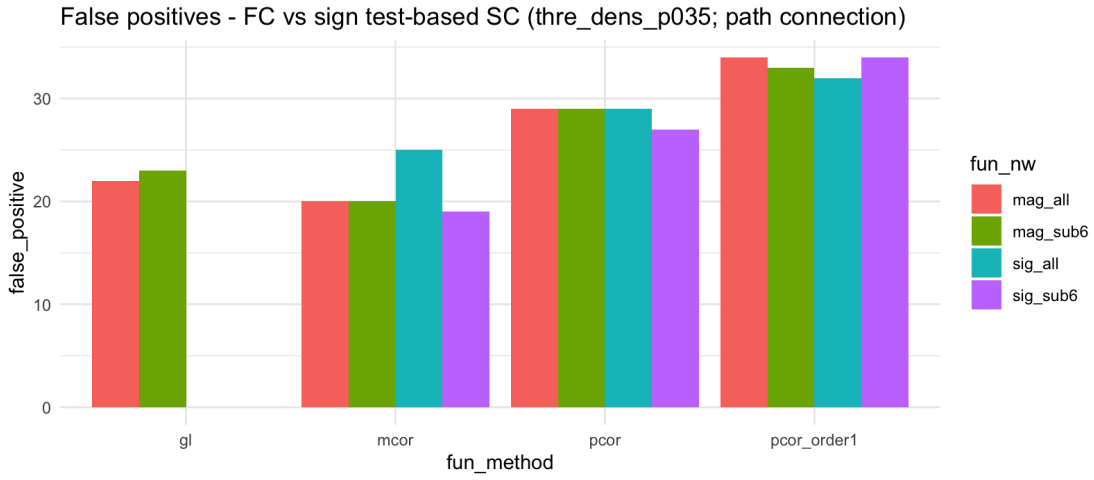
3.4.2 SC based on path connection

Now SC are defined in terms of whether there exists a path between two regions. As before, FPs are calculated using upper triangular parts only, and two separate tables are shown for FCs using 1200 and 200 time points.

		Global thre	**Local thre**	**Sign test**(*count_p027*)	**Sign test**(*dens_p035*)
mcor	*mag_all*	20	-	30	20
	mag_sub6	20	-	31	20
	sig_all	25	-	36	25
	sig_sub6	19	-	31	19
pcor	*mag_all*	29	-	45	29
	mag_sub6	29	-	45	29
	sig_all	29	-	45	29
	sig_sub6	27	-	43	27
pcor_ord1	mag_all	34	-	49	34
	mag_sub6	33	-	47	33
	sig_all	32	-	45	32
	sig_sub6	34	-	47	34
glasso	*all*	22	-	33	22
	sub6	23	-	34	23

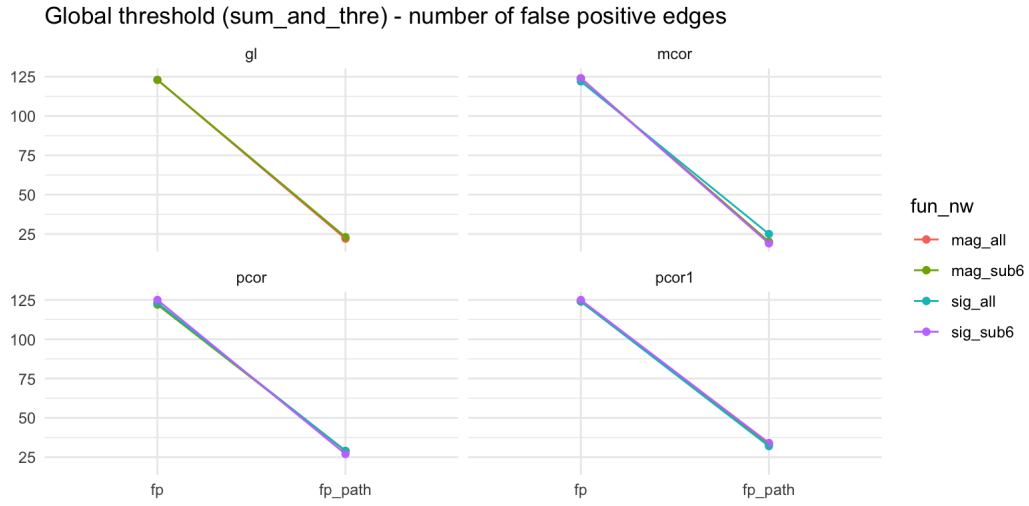
The following figures aim to visualize these FP counts using barplots. Each figure contains all FC networks and one SC method. In terms of nonzero positions, two global thresholding approaches give the same result, so only one is shown. Additionally, local thresholding is not plotted since all pairs of nodes are connected by some path (as shown in a previous section).



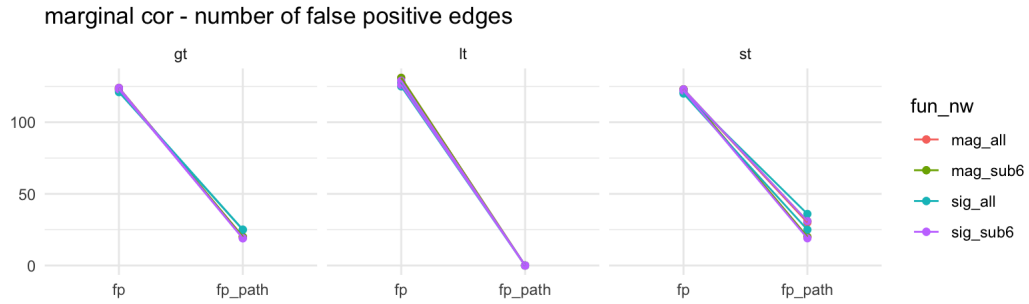


I also tried to plot the changes in FP, comparing using proportion-based SC and path connection-based SC, (1) for every SC method (6 plots), (2) for every FC method (4 plots). Example figures:

Global threshold SC vs all FCs.



Marginal corr FC vs all SCs.



Other figures can be found at the end of this write-up.

4 Aside

1. Global thresholding with no disconnected region

For global thresholding SC, currently using 136 as a cutoff leads to disconnected regions. Consider relaxing the restriction on number of edges to make all nodes have a nonzero degree: suppose we sum over 957 subjects' structural connectivity matrices and perform a global thresholding on it to ensure every brain region is connected with at least one other region. Then the thresholded network will contain at least 601 edges (density $\geq 601/2278 \approx 0.264$). Not every pair in this SC network (with 601 edges) is connected by some path.

2. Performance of local thresholding

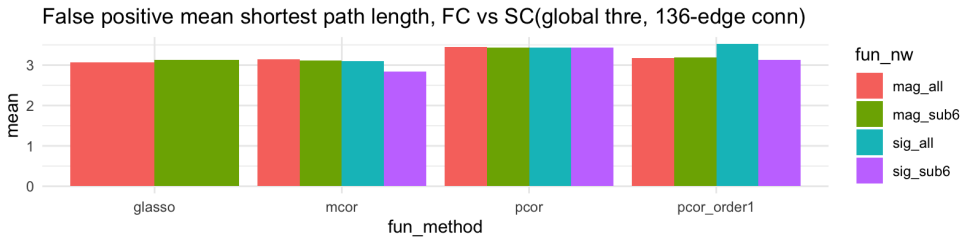
Local thresholding guarantees that every node has a nonzero degree. Taking a look at path length-related analyses, local thresholding seem to give relatively deviated behaviors. While this may be attributed to the fact that global thresholding and sign test care more about global values and structures, which is different from how local thresholding works, we should notice it does not capture the connectivity between $(28, 62)$ which is commonly selected by other SC and FC methods (section 3.1).

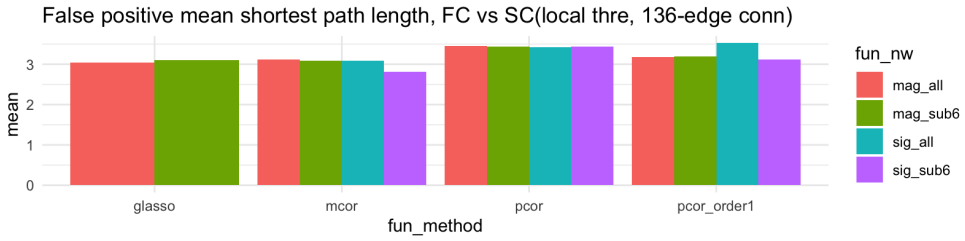
3. Other aspects of comparison

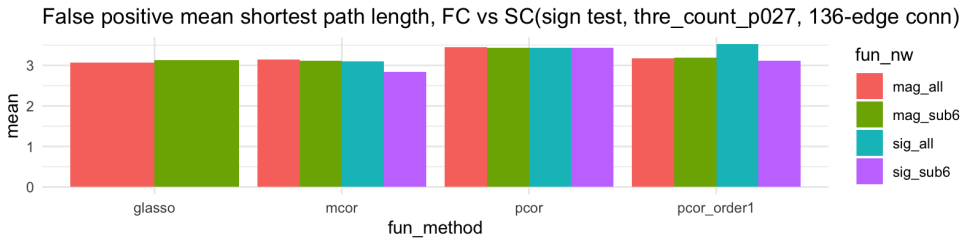
This write-up mainly focuses on false positives with respect to different types of SC. Other types of analyses may provide additional insights, e.g. proportions of edge selection for different methods.

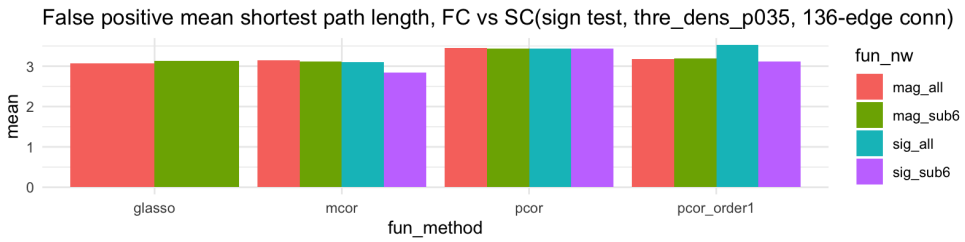
Additional figures

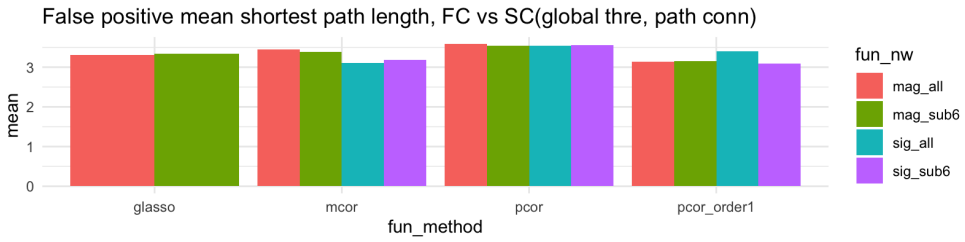
Figures for mean of FP shortest path lengths.

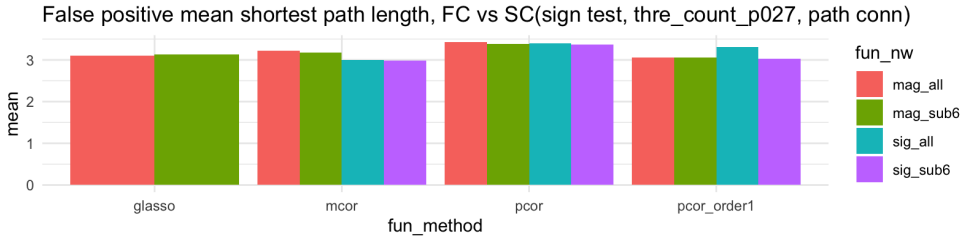


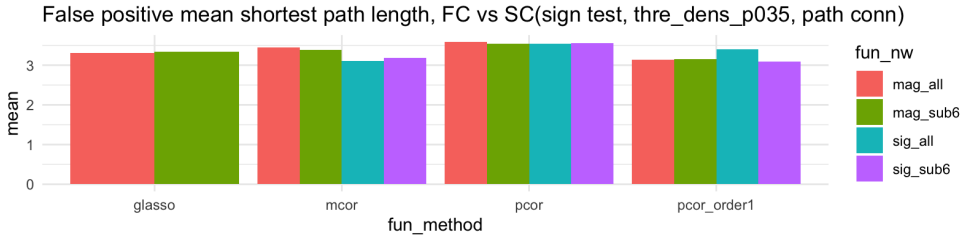




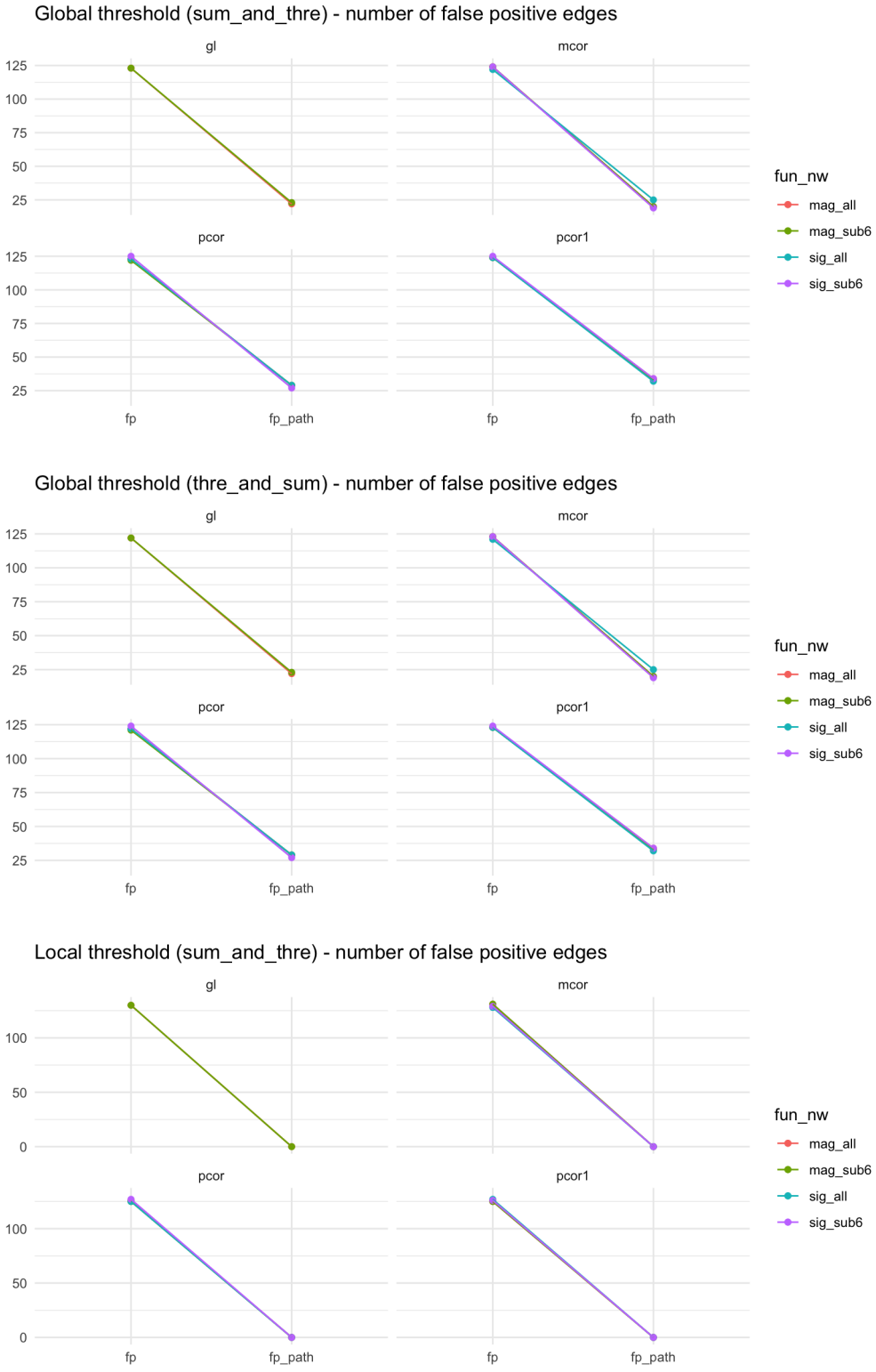


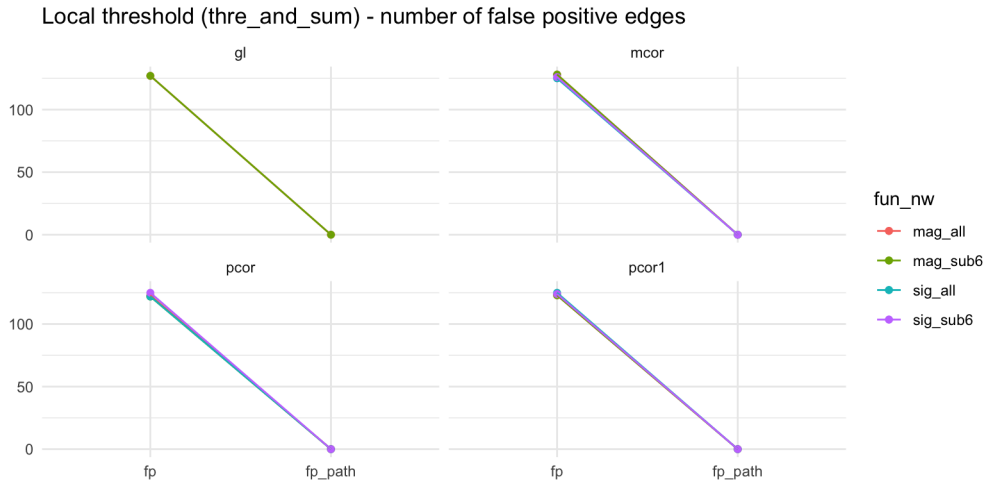


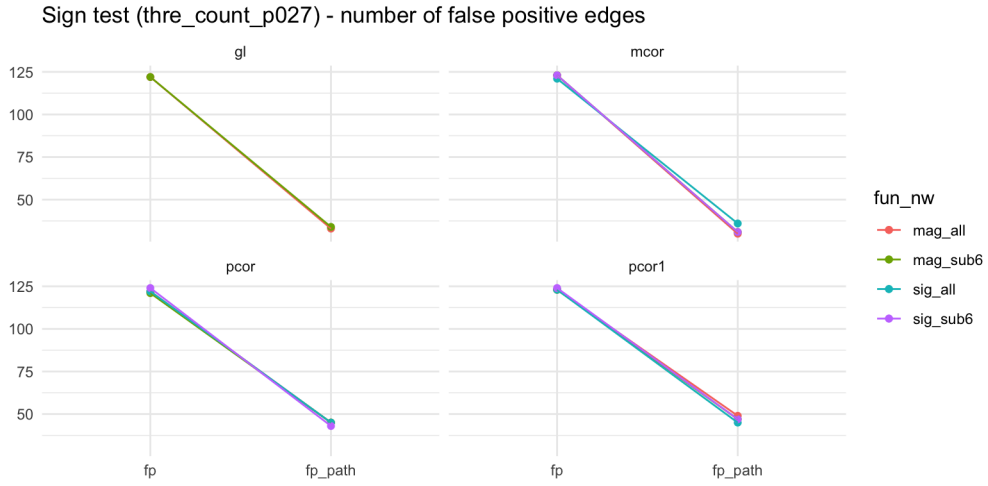

False positive mean shortest path length, FC vs SC(sign test, thre_count_p027, path conn)
mean
0
1
2
3
glasso
mcor
pcor
pcor_order1
fun_method
fun_nw
mag_all
mag_sub6
sig_all
sig_sub6

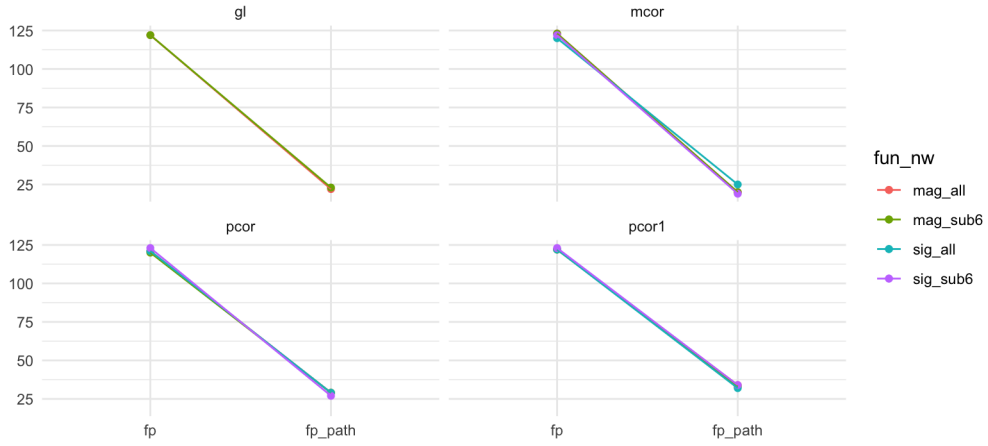

False positive mean shortest path length, FC vs SC(sign test, thre_dens_p035, path conn)
mean
0
1
2
3
glasso
mcor
pcor
pcor_order1
fun_method
fun_nw
mag_all
mag_sub6
sig_all
sig_sub6

One SC vs all FCs.




Local threshold (thre_and_sum) - number of false positive edges
gl
mcor
pcor
pcor1
100
50
0
fp
fp_path
fun_nw
mag_all
mag_sub6
sig_all
sig_sub6


Sign test (thre_count_p027) - number of false positive edges
gl
mcor
pcor
pcor1
125
100
75
50
fp
fp_path
fun_nw
mag_all
mag_sub6
sig_all
sig_sub6


Sign test (thre_dens_p035) - number of false positive edges
gl
mcor
pcor
pcor1
125
100
75
50
25
fp
fp_path
fun_nw
mag_all
mag_sub6
sig_all
sig_sub6

One FC vs all SCs.