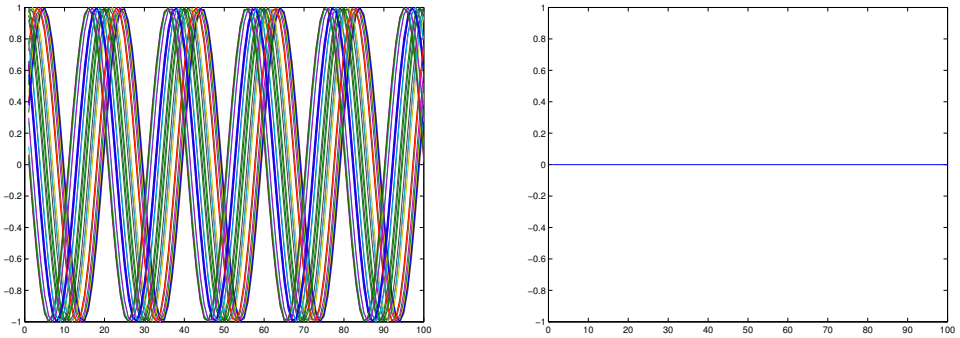


Figure 1: Left: Sample data created from sine waves shifted randomly at most 10 time steps apart (no added noise). Right: Comparison of the permutation returned by the algorithm with that of the actual data set for 100 separately generated data sets.

1 1/24/16

1.1 Permutation testing

In order to test how the algorithm returns the permutation results, I set up sample matrices using `create_toy.m` and recorded the distance between the result returned by the algorithm (`eig_perm`) and that of the actual input data (`ss`).

Listing 1: Compare Permutations

```
rois = 33;
noise_mag = 0;
trace_type = 'sines';
shift = 10;
y = zeros(1,100);
for i = 1:length(y)
        [Z, ss] = create_toy(trace_type, 'rois', rois, ...
                'noisemag', noise_mag, ...
                'shift', shift);
        [eig_phases,eig_perm,slm,evals] = cyclic_analysis(Z);
        y(i) = cyclic_distance(ss,eig_perm);
end
plot(y)
```

The function `create_toy` produces a set of `rois` identical sine waves randomly spread over a `shift` time step period (i.e., each copy of the trace is shifted along the x axis so that all copies are within `shift` time steps of each other). The variable `ss` shows the ordering of the traces produced. Using the setting above, for example, a data set like that in on the left of Figure 1 is produced. The figure on the right shows that the algorithm returns the identical permutation each time. The distance is measured using `cyclic_distance(V1,V2)` which simply counts the how many steps V2 is from V1 (let V1 be (1:5), then [2,1,3,4,5] is a distance of 1 from V1 and [2,3,4,5,1] is a distance of 5 from V1.) This doesn't really count "cyclic distance" but works for this application since we would like to see if we are picking up the correct starting point and cycling in the right

direction. Increasing `shift` to 18 (the period for the sine wave is 20) starts to mess up the permutation since the algorithm can't determine the proper starting point.

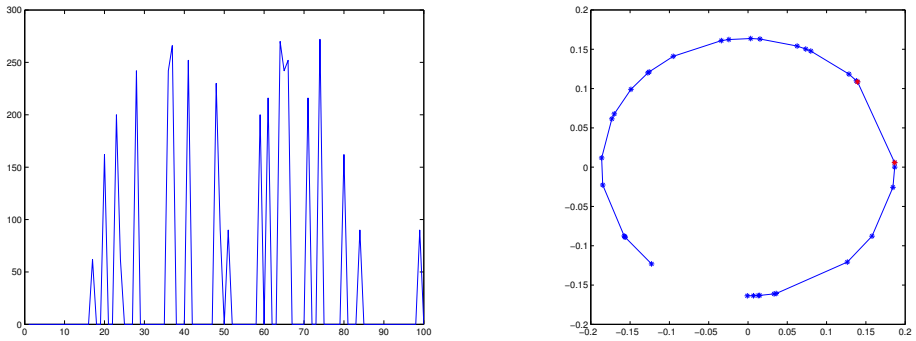


Figure 2: Left: Algorithm performance when shift=18. Right: Plot of phases for a permutation which the algorithm had shifted by 11 places (distance = 242). The red asterisks highlight the 11th and 12th points in the cycle.

The algorithm determines the starting point by finding the largest gap between points in the phase plot (greatest difference in angle). Once the phases are too spread out, the largest gap can easily be between a pair of middle points. Note, however, that still the algorithm cycles in the correct direction in each case.

2 Listings

Listing 2: cyclic_distance

```
function [dist] = cyclic_distance(V1, V2)
[~, ss] = sort([V1(:), V2(:)]);
dist = sum(abs(diff(ss,1,2)))/2;
```

2.1 Movies

2.1.1 Data Treatment

In the original data set, there are gaps in the data for some genres and not all genres began at the same time so some initial data treatment was required. First, years in which there were gaps in the data were inpainted using linear interpolation. Next, the genre counts were taken on a log scale to account for the fact that movie production has grown exponentially. Finally, only years in which genres had data points were considered (1916-2015). The film-noir genre had to be excluded since the production years were so limited. The following table shows the results of `analyze_cyclicity` on the movie data (processed as described above).

quad results	zscore results
musical	musical
music	music
family	family
thriller	thriller
mystery	mystery
sci fi	sci fi
drama	drama
crime	crime
adventure	romance
action	adventure
fantasy	history
comedy	comedy
romance	action
animation	western
history	fantasy
biography	documentary
documentary	animation
horror	biography
war	horror
sport	sport
western	war

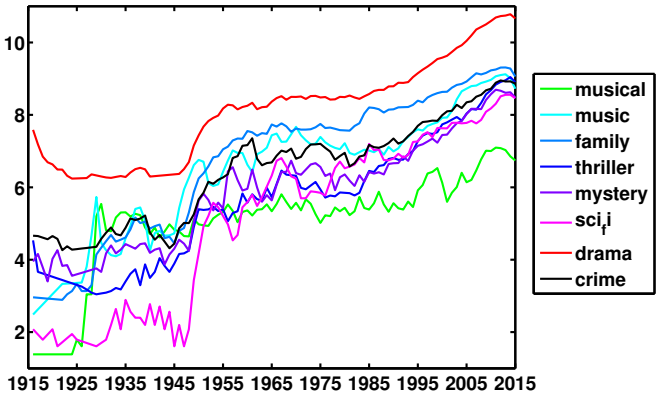


Figure 3: Traces from the first eight genres.

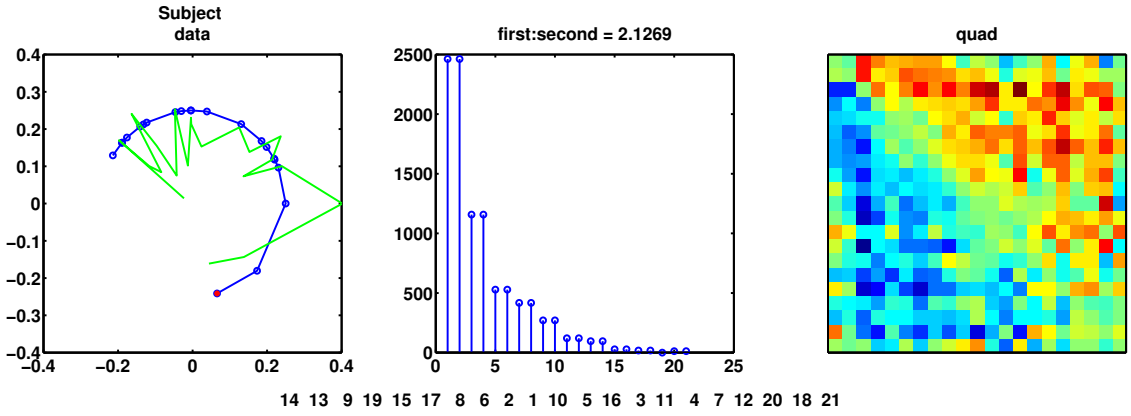


Figure 4: Cyclicity results using quadratic variation normalization

Note that the Musicals and Music genres look very similar to each other, but dissimilar to everything else and show a drastic jump around 1925. Since this is such a long time ago, it seemed worth considering more recent trends in the data.

quad results	zscore results
romance	romance
family	family
adventure	adventure
documentary	documentary
thriller	musical
musical	history
history	thriller
drama	drama
sport	comedy
comedy	sport
music	music
mystery	mystery
crime	crime
western	western
biography	animation
horror	biography
war	horror
action	war
fantasy	action
animation	fantasy
sci-fi	sci-fi

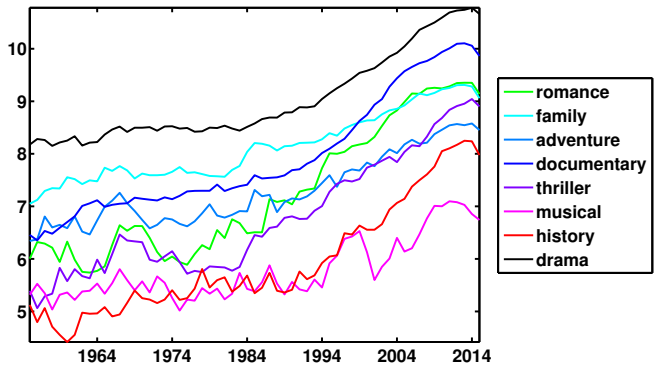


Figure 5: Traces from the first eight genres using only years 1955-2015 in the cyclicity calculation.

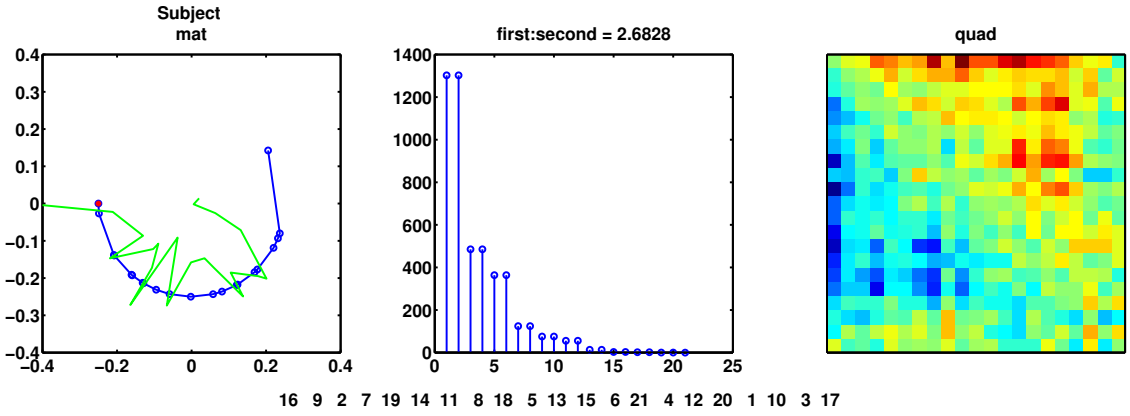


Figure 6: Cyclicity results from 1955-2015 movie data using quadratic variation normalization

3 2/2/16

3.1 Movie trends with words counts from New York Times

We added some word trends to the movie data, taking search counts for selected words on the New York Times website. The search was filtered by year so that we could get word counts for each year that we have movie data for (1916-2015). Word counts were pulled

using `times_scrape.py` and looping through the years in a bash script (see scraping folder in Movies). The words searched were searched

crisis	terror	war	attack	expansion
growth	invasion	prosperity	shooting	

and the following results obtained:

perm results
EXPANSION
WAR
musical
ATTACK
music
INVASION
family
western
mystery
thriller

perm results
sci fi
drama
crime
adventure
action
fantasy
comedy
animation
biography
horror

perm results
war
history
romance
documentary
sport
TERROR
CRISIS
SHOOTING
GROWTH
PROSPERITY

Figure 7: Resulting permutation from full movie and word analysis

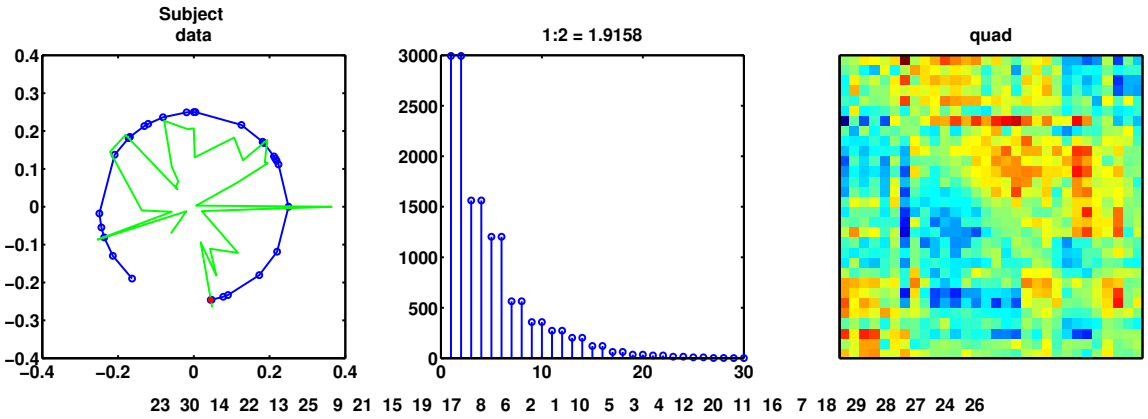


Figure 8: Results of full movie and word analysis

3.2 Refining Movie-word results

Next, we wish to consider smaller subsets of traces using the magnitude of the phase associated with a trace as an indicator of agreement with a particular ordering. For each set of figures, `p` indicates which eigenvector was used to generate the ordering , `n` indicates how many categories were used in the smaller group analysis and `group` is which subset of the full set. So, for example, if `p=1`, `n=8`, `group=1` then the first eigenvector was used to generate an ordering of full set of categories and then the eight categories with the highest magnitude of associated phase vector were pulled out and reanalyzed using the cyclicity algorithm to yield the ordering shown in the legend. If `group=2` the group 1 categories

were pulled out of the original data set and then the same procedure was followed, ordering the categories associated with the top eight phases of this smaller subset. When `p=2`, the second eigenvector is used in the cyclicity algorithm to generate the ordering.

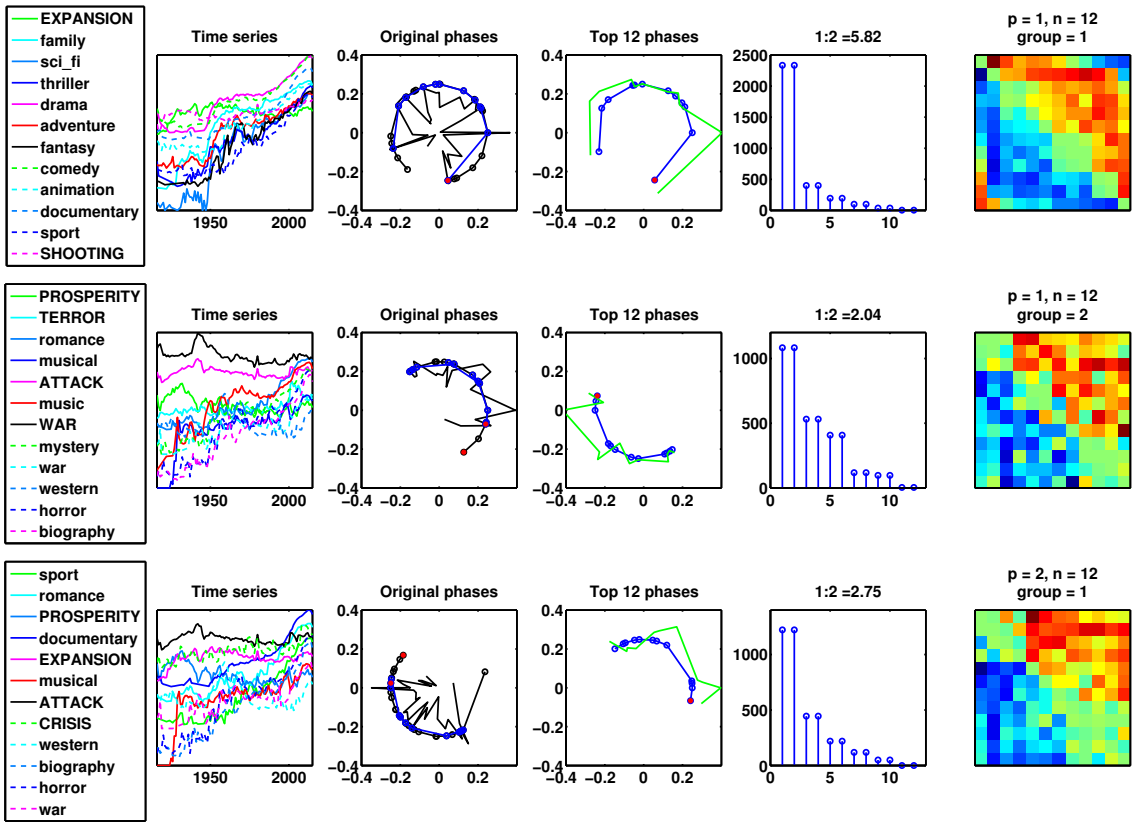


Figure 9: Results with n=12.

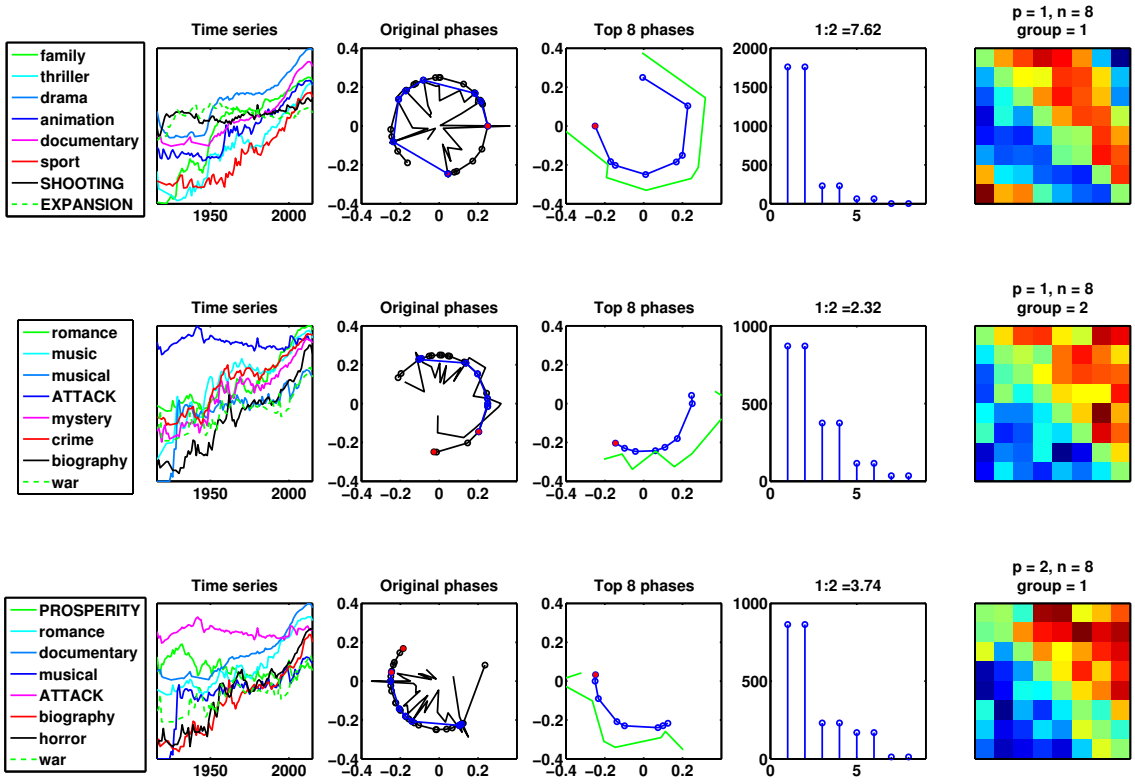


Figure 10: Results with n=8.

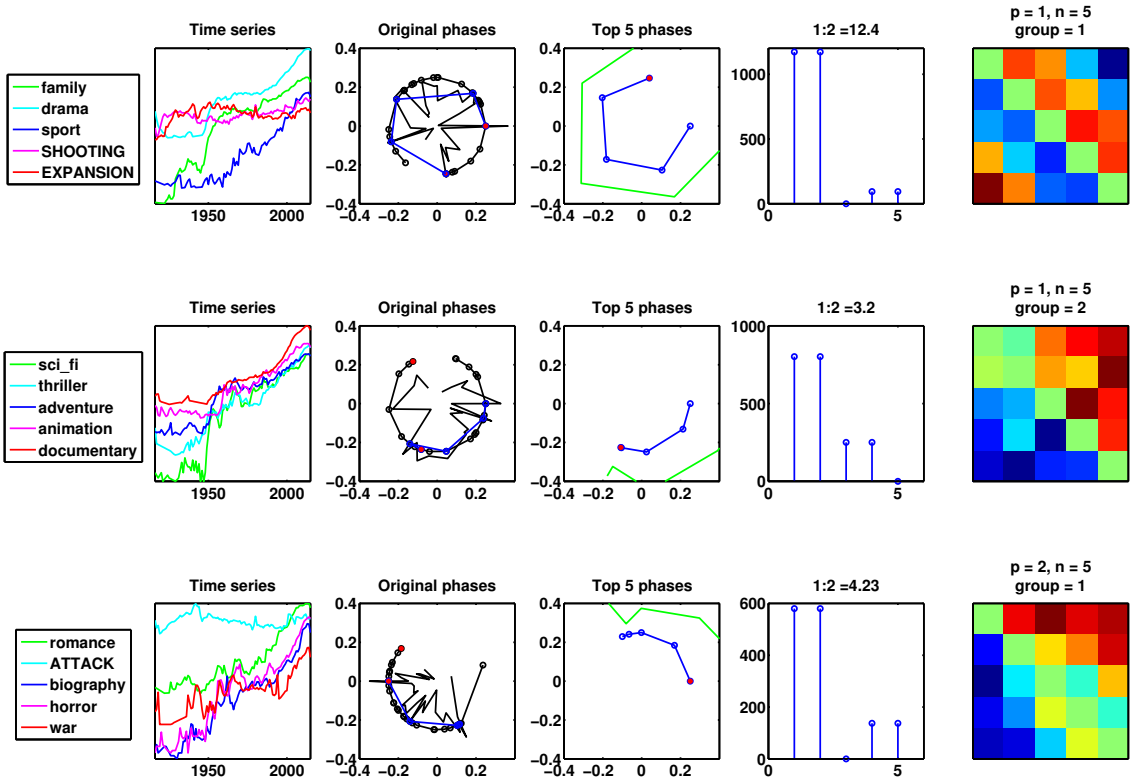


Figure 11: Results with n=5.

In each figure, the first subplot on the left shows the original traces (log scale) of the n categories. The next subplot shows the original phase plot in black with the final n categories highlighted in blue with the new ordering. Mostly the ordering from the initial analysis is preserved, but there are a few places (`p=1`, `n=5`, `group=2`, for example) where if two categories were initially close together, they get exchanged in the final ordering. The final three plots are the same thing that we have been looking at all along for the subset of n categories. Note that for the eigenvector ratio label on subplot 4, I the label says 1:2, which is actually 1:3 since the eigenvalues are complex conjugates so 1 and 2 have the same absolute value. If the third eigenvector was equal to 0 (odd man out), then I took the ratio of the first to the next non-zero (first:fourth technically).

3.3 Impressions

- I like the set of five. It is easy to digest and has high e1:e2 ratios. Until I have some reason to do otherwise, I'll look at sets of 5.
- Group 2 is not fantastic. Of the three analyses shown, it has the lowest e1:e2 ratio. I will probably run it anyway when we look at the brain permutations.
- I would like to cut off the first 40 years as before and see what those trends look like.

4 2/6/16

4.1 Recent movie data (1955-2015)

As before, I reduced the movie data to only more recent trends. I learned today that "talkies" started in the 1920's with the first feature length film with synchronized dialogue was released in 1927. This probably explains the leap in production of music and musical films in the late 1920's. It is interesting, however, that in analyzing only more recent trends,

- Clusters become quite obvious by looking at the lead matrix;
- First eigenvalue, group 1 analysis does not wrap around the full cycle anymore;
- Second eigenvalue ordering has lower e1:e2 ratio with a fuzzier lead matrix and now wraps around a full cycle.

Observe the following results on only the recent data.

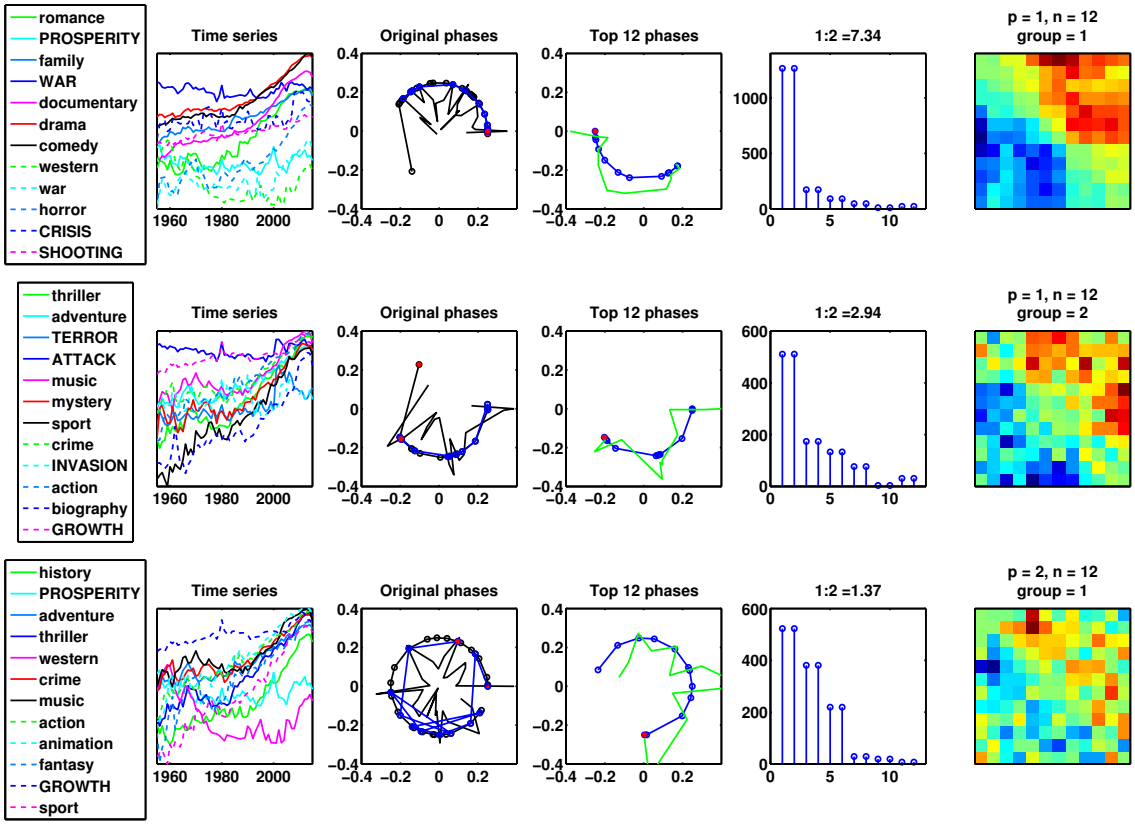


Figure 12: Results with n=12.

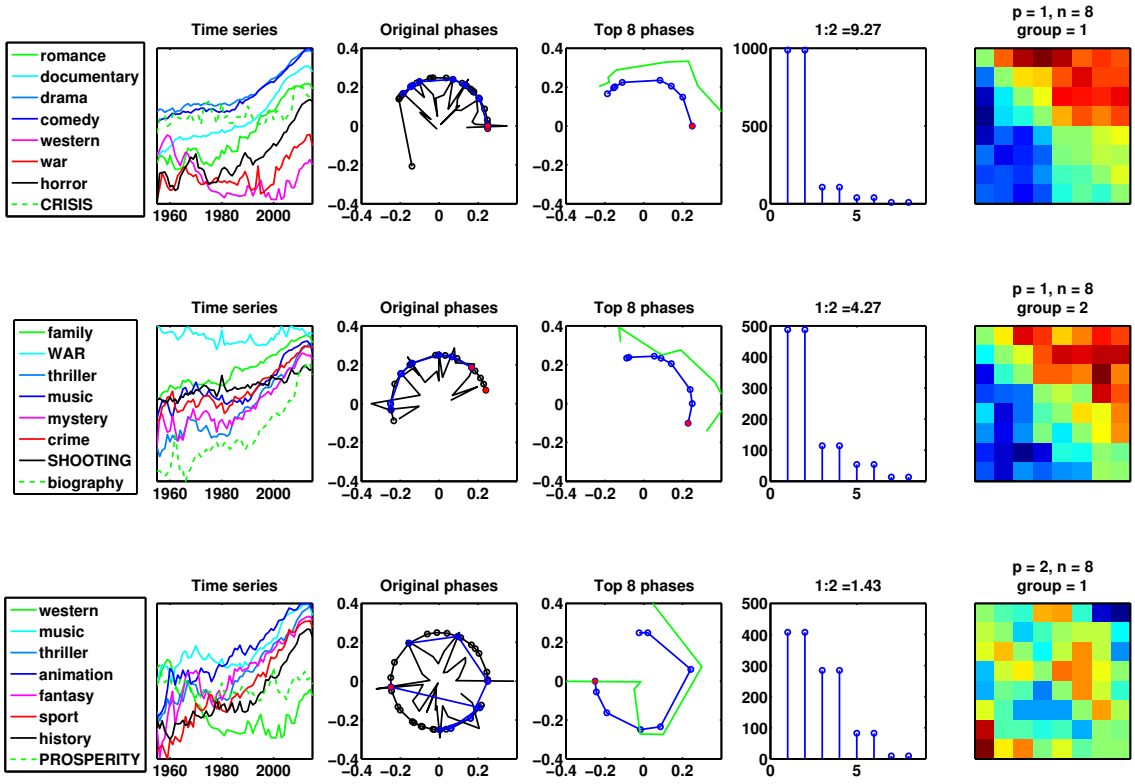


Figure 13: Results with n=12.

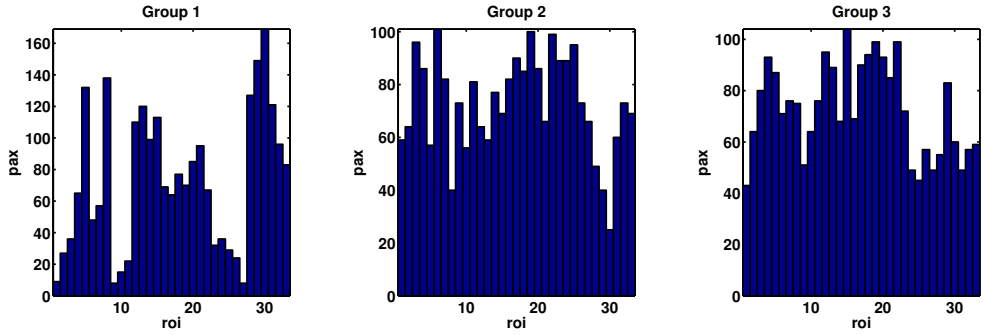


Figure 15: Histogram of top 12 regions (highest phase magnitude) for 200 HCP subjects. Group 1 shows the results for the top 12 regions using the first eigenvector; Group 2 shows the results for the top 12 regions using first eigenvector from the subset of all regions minus those in Group 1; Group 3 shows the results for the top 12 regions using the second eigenvector.

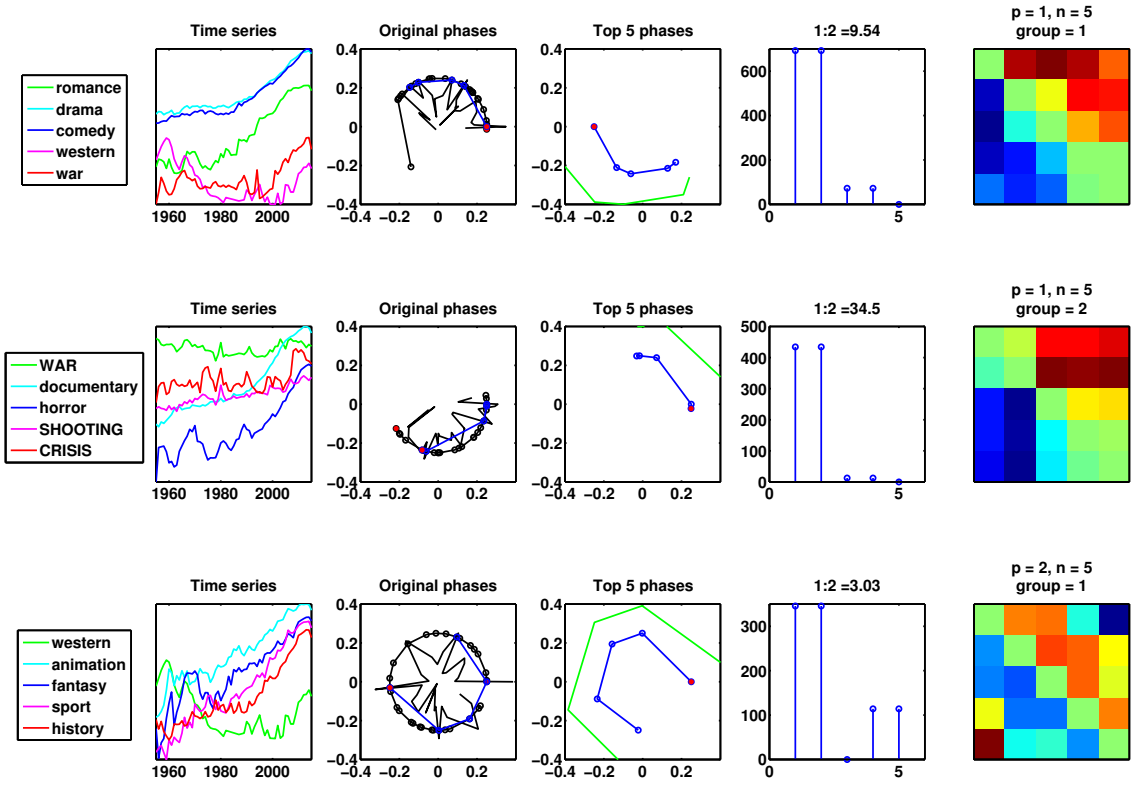


Figure 14: Results with n=12.

5 2/16/16

This week we analyzed the data from the 200 HCP subjects and compared the top n regions in each subject ($n \in \{5, 8, 12\}$). We then summarized the results by histogramming the number of appearances of each roi for all 200 of the subjects. The results for $n = 12$

are shown in 5. These results were then used to generate the following groups of 10 "champions."

Group 1	
30	l cuneus
29	l primary visual cortex
8	l primary auditory cortex
5	precuneus*
28	r primary visual cortex
31	r cuneus
13	l ventral intraparietal sulcus**
15	l posterior intraparietal sulc**
12	r ventral intraparietal sulcus**
14	r posterior intraparietal sulc**

Group 3	
15	l posterior intraparietal sulc**
19	l superior temporal junction*
22	r mid frontal gyrus*
12	r ventral intraparietal sulcus**
18	r superior temporal junction*
4	posterior cingulate cortex*
20	r superior temporal sulcus*
17	r inferior partietal lobe**
13	l ventral intraparietal sulcus**
5	precuneus*

Definitely need to do more reading or talk to Fatima and Sara about how these regions are related. Wikipedia gives the following functions (for the less obvious regions):

- Ventral intraparietal sulcus (VIPS): visual attention and saccadic eye movement
- Posterior (caudal) IPS: perception of depth from stereopsis
- Cuneus: involved in basic visual processing
- Precuneus: involved in loads of stuff including default network and visuospatial processing
- Superior temporal junction (temporoparietal junction): information processing and perception (right and left do not do exactly the same thing - right is more associated with attention and left with language processing)
- Middle frontal gyrus: part of prefrontal cortex - executive function including attention and working memory
- Posterior cingulate: central node in DMN
- Superior temporal sulcus: social perception - responds more to human stimuli rather than environmental/other object
- Inferior parietal lobe: perception of emotion and faces, interpretation of sensory input

Regions marked with '*' are those which are indicated in the DMN. Those marked with '**' surround the angular gyrus, which is also included in the DMN, but is not one of the regions on our list. The acutal DMN regions listed in Wikipedia include

- PCC & Precuneus
- mPFC
- Angular gyrus

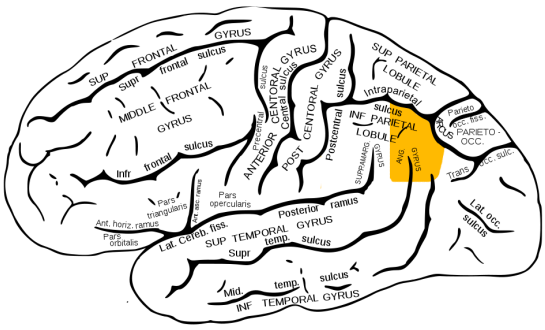


Figure 16: Angular gyrus

- TPJ (Dorsomedial subsystem)
- Lateral temporal cortex (Dorsomedial subsystem)

It might be worth looking at the functional
regions from Stanford's lab to see if we can reproduce their network associations using the regions they provide.

One interesting observation is the total dominance of the left primary auditory cortex (region 8) in Group 1 compared to the right primary auditory cortex (region 9) which hardly shows up at all. Many other symmetric regions seem to show up in pairs so this contrast is striking.

5.1 Triples

5.1.1 Group 1

Regions here are numbered as follows:

1	l cuneus
2	l primary visual cortex
3	l primary auditory cortex
4	precuneus
5	r primary visual cortex
6	r cuneus
7	l ventral intraparietal sulcus
8	l posterior intraparietal sulc
9	r ventral intraparietal sulcus
10	r posterior intraparietal sulc

Analyzing the triples for Group 1 (without correcting for cyclicity) yields the following cycles which are present in at least a third of the subjects (if the cycles are random we might expect any given cycle to be present in 1/6 of the subjects):

(3,2,1)	84	(3,6,4)	109	(3,2,7)	90
(3,4,1)	70	(9,6,4)	71	(3,4,7)	80
(3,5,1)	99	(3,7,4)	75	(3,5,7)	100
(3,6,1)	96	(3,8,4)	109	(3,6,7)	105
(3,7,1)	87	(6,8,4)	73	(3,8,7)	91
(9,7,1)	67	(9,8,4)	76	(3,9,7)	116
(3,8,1)	85	(3,9,4)	105	(3,10,7)	96
(3,9,1)	94	(3,10,4)	110	(3,2,8)	71
(3,10,1)	89	(9,10,4)	68	(3,5,8)	72
(3,5,2)	83	(3,2,5)	75	(3,6,8)	78
(3,6,2)	87	(3,6,5)	86	(3,9,8)	81
(3,8,2)	68	(3,8,5)	79	(3,10,8)	72
(3,9,2)	79	(3,9,5)	88	(3,5,10)	71
(3,10,2)	78	(3,10,5)	86	(3,6,10)	76
(3,1,4)	72	(3,9,6)	73	(3,8,10)	72
(3,2,4)	98	(3,10,6)	67	(3,9,10)	82
(3,5,4)	105				

Worth noting from this is the fact that region 3 (lPAC) nearly always shows up first. After that, however, the order of the subsequent regions seems to be nearly split. By taking cyclicity into account (reordering cycles so that the lowest index always appears first) we can filter out these cycles and we get the following results that are present in at least 60% of subjects (again, thinking that if cycles are random then any given cycle should show up in 50% of subjects):

(3,5,4)	127	(1,3,5)	133	(1,9,7)	133
(3,6,4)	130	(1,3,6)	129	(3,9,7)	137
(3,8,4)	127	(1,3,7)	120	(1,3,9)	124
(3,9,4)	123	(3,5,7)	122	(1,3,10)	123
(3,10,4)	127	(3,6,7)	125		

Again, the lPAC shows up in every cycle except for one. Is this an outlier or are the fMRI sounds triggering this whole cycle?

5.1.2 Group 3

Regions in this group are numbered as follows:

1	l posterior intraparietal sulc
2	l superior temporal junction
3	r mid frontal gyrus
4	r ventral intraparietal sulcus
5	r superior temporal junction
6	posterior cingulate cortex
7	r superior temporal sulcus
8	r inferior partietal lobe
9	l ventral intraparietal sulcus
10	precuneus

First, without normalizing for cyclicity:

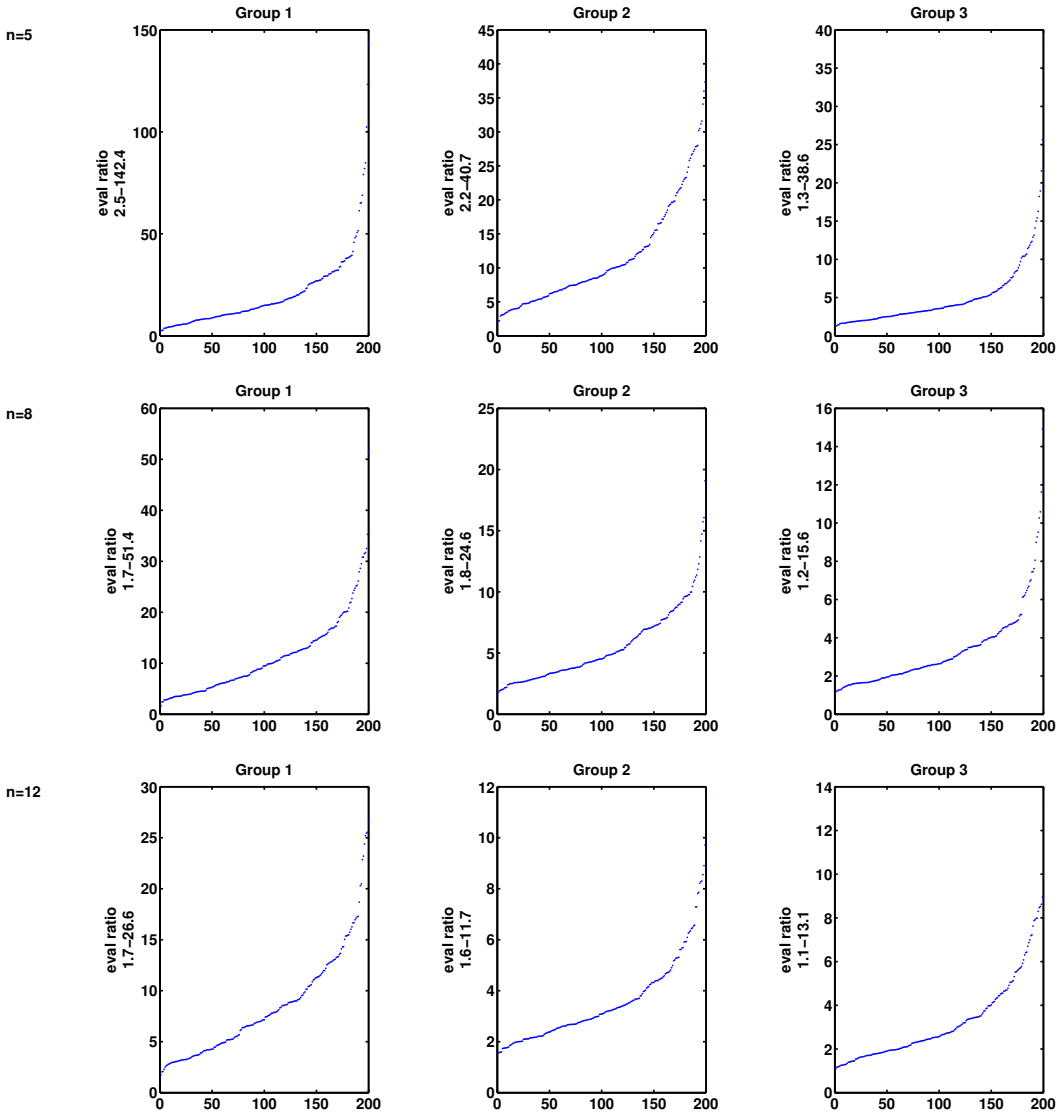
(3,2,1)	69	(8,1,9)	72	(7,8,9)	75	(3,4,10)	81
(3,5,1)	72	(3,2,9)	91	(2,10,9)	79	(5,4,10)	73
(3,7,1)	73	(5,2,9)	85	(3,10,9)	82	(7,4,10)	74
(3,8,1)	67	(6,2,9)	75	(5,10,9)	88	(3,5,10)	101
(5,8,1)	69	(7,2,9)	70	(6,10,9)	81	(6,5,10)	73
(3,1,4)	75	(2,4,9)	78	(7,10,9)	85	(7,5,10)	78
(5,1,4)	69	(3,4,9)	94	(8,10,9)	81	(2,6,10)	68
(3,2,4)	70	(5,4,9)	89	(2,1,10)	80	(3,6,10)	100
(5,2,4)	67	(6,4,9)	81	(3,1,10)	109	(5,6,10)	78
(3,5,4)	68	(7,4,9)	92	(4,1,10)	67	(7,6,10)	83
(3,7,4)	69	(8,4,9)	70	(5,1,10)	92	(8,6,10)	69
(3,2,8)	70	(3,5,9)	86	(6,1,10)	76	(3,7,10)	94
(3,5,8)	75	(3,6,9)	81	(7,1,10)	94	(5,7,10)	80
(2,1,9)	78	(3,7,9)	81	(3,2,10)	104	(2,8,10)	93
(3,1,9)	102	(5,7,9)	74	(5,2,10)	89	(3,8,10)	112
(5,1,9)	93	(2,8,9)	68	(6,2,10)	80	(5,8,10)	109
(6,1,9)	78	(3,8,9)	86	(7,2,10)	84	(6,8,10)	75
(7,1,9)	90	(5,8,9)	90			(7,8,10)	98

The most notable feature in this case is that regions 9 and 10 (lVIPS, precuneus) show up last in most of the cycles.

Taking cyclicity into account:

(1,4,3)	120	(4,9,6)	129	(8,10,9)	134
(1,9,3)	130	(1,9,7)	121	(3,5,10)	125
(1,10,3)	131	(4,9,7)	125	(3,6,10)	123
(2,10,3)	130	(4,9,8)	120	(2,8,10)	120
(1,9,5)	123	(3,4,9)	122	(3,8,10)	135
(4,9,5)	125	(2,10,9)	122	(4,8,10)	124
(2,8,6)	127	(5,10,9)	123	(5,8,10)	126
(5,8,6)	126	(6,10,9)	129	(7,8,10)	120
(1,9,6)	122	(7,10,9)	120		

5.1.3 Eigenvalue Ratios



6 3/8/16

6.1 Comparison of champions: triples v. phases

Out of curiosity, I wanted to compare the regions that show up the most in the top triples (those that are present in at least 70% of subjects) to the regions that have the highest phase magnitudes. I was expecting the regions to more or less match up since those with higher phase magnitudes are considered more certain, but they often don't, and in some cases, a region which is very low on the phase rating, is very high in the triples rating. Note, in particular, regions 92, 10, and 6 in the first, second, and third plots, respectively, in figure (6).

Something that may also be informative, but less surprising, is a closer look at the eigenvalue ratios. Figure (6) shows the eigenvalue ratios for several regions or groups of

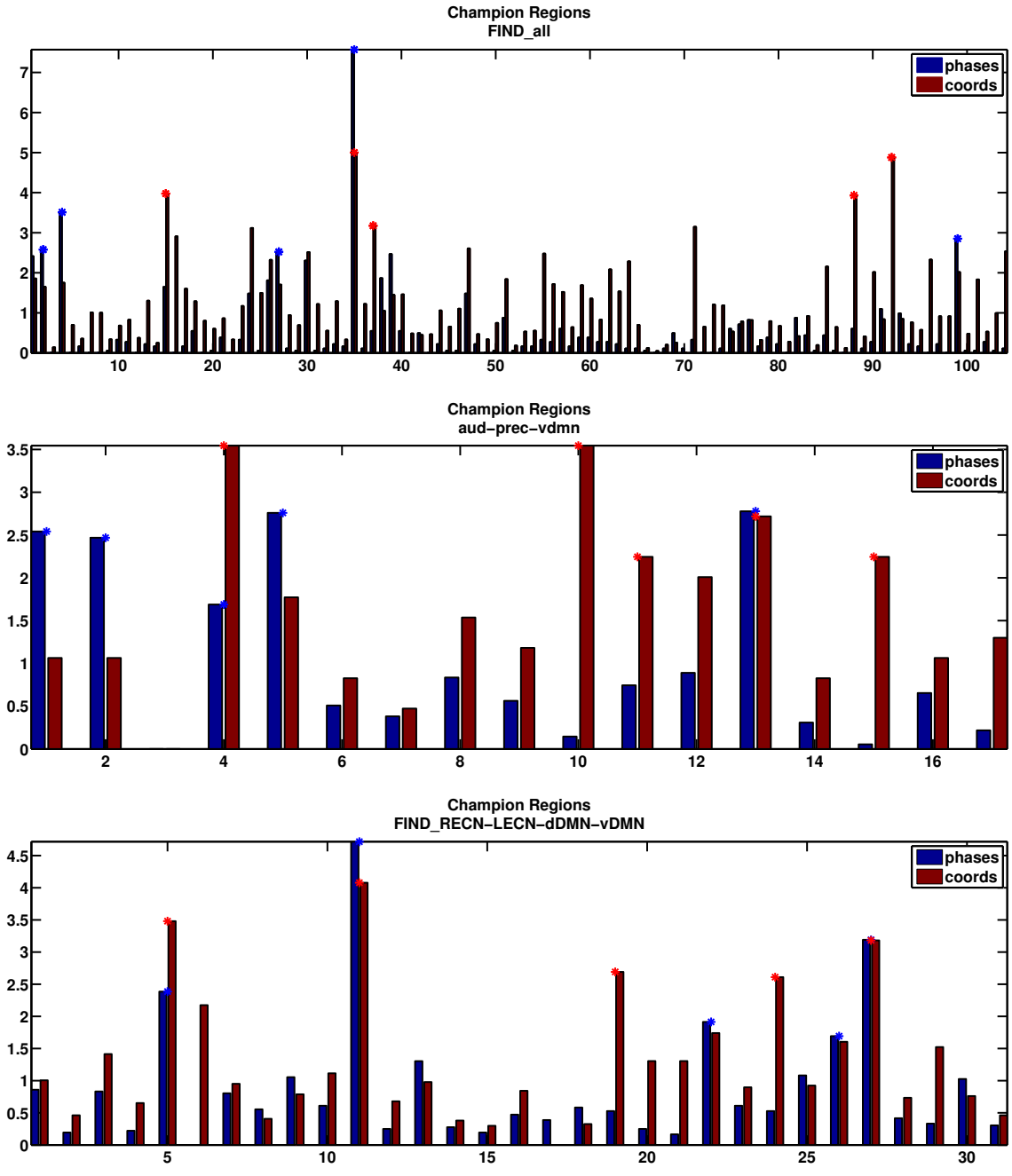

Champion Regions
FIND_all
phases
coords
Champion Regions
aud−prec−vdmn
phases
coords
Champion Regions
FIND_RECN−LECN−dDMN−vDMN
phases
coords

Figure 17:

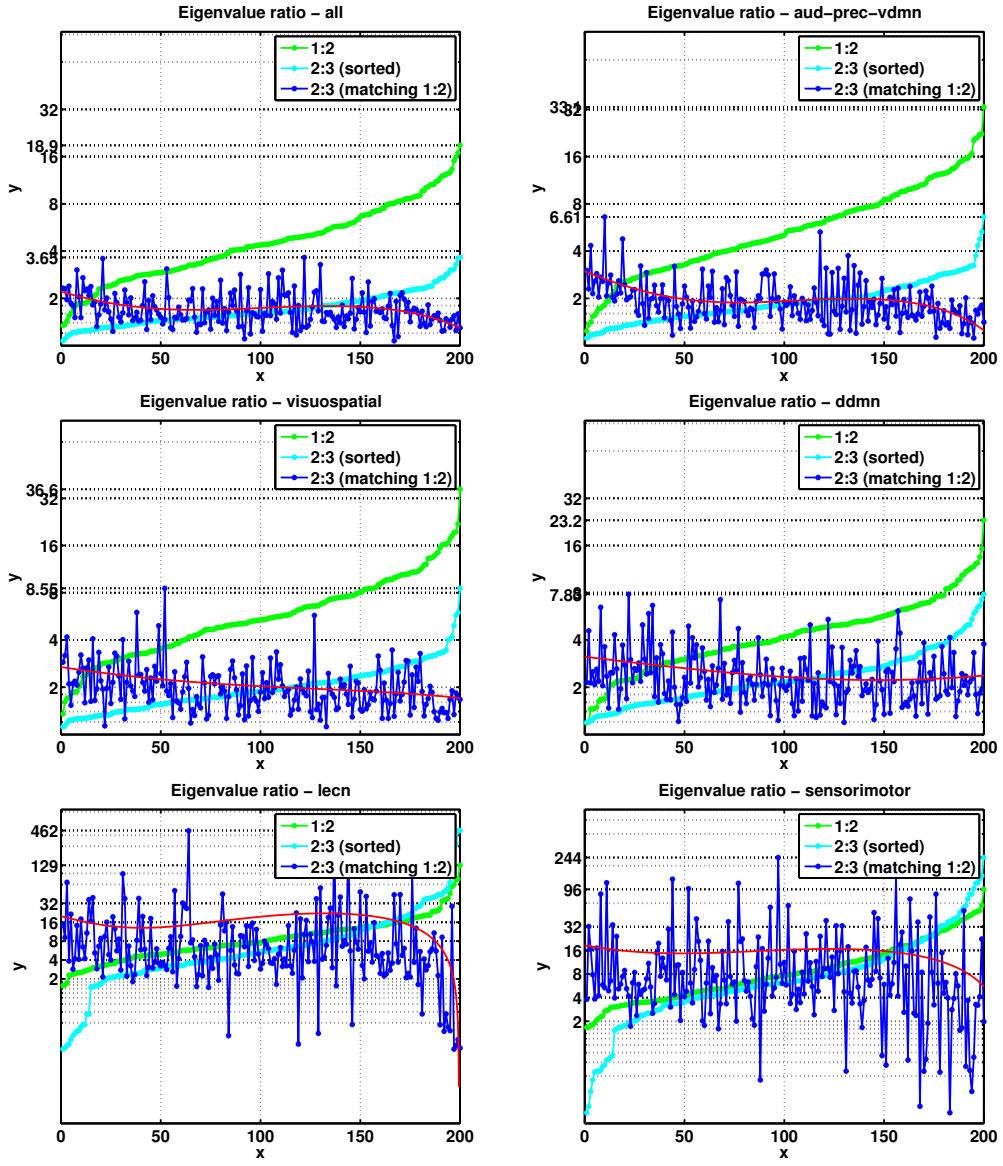
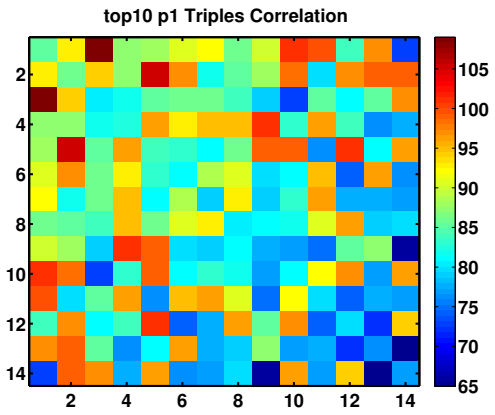

Eigenvalue ratio – all
Eigenvalue ratio – aud–prec–vdmn
Eigenvalue ratio – visuospatial
Eigenvalue ratio – ddmn
Eigenvalue ratio – lecn
Eigenvalue ratio – sensorimotor
1:2
2:3 (sorted)
2:3 (matching 1:2)
x
y

Figure 18:



1	030907
2	010305
3	010907
4	030604
5	010306
6	030504
7	030804
8	031004
9	030607
10	010309
11	030904
12	010310
13	030507
14	010307

Figure 19:

regions. The top two plots show eigenvalue ratios for larger sets of regions - the plot on the left shows all regions and the plot on the right shows those in the auditory, precuneus and ventral DMN functional groups. The next row of plots shows the eigenvalue ratios for single functional groups which we would expect to be coherent in a resting state - visuospatial and dorsal DMN. The plots on the bottom row show the ratios for functional groups which we expect to be unrelated to a resting state - left executive control and sensorimotor.

A few differences that stand out in the bottom plots are the fact that the dark blue line jumps around more, the light blue line crosses over the green, and the maximal values are much higher (greater than 200 vs. less than 50).

The green represents the ratio of $\lambda_1 : \lambda_2$ sorted in ascending order. The light blue line is the ratio $\lambda_2 : \lambda_3$ (also sorted). The dark blue is $\lambda_2 : \lambda_3$, but this time sorted using the ordering of the sorted $\lambda_1 : \lambda_2$ trace so that the ratios can be directly compared. The red line shows a cubic fit of the blue line (other fits might be more informative... exp2 fits the ordered ratios pretty well).

6.2 Correlations

Figure (6.2) shows the correlation between Group 1 cyclic-normalized triples. For each subject, the algorithm generates triples from the permutation and then for each triple in the set of champions, the algorithm counts which other champion triples are present and then adds one to the corresponding place in the correlation matrix. The algorithm is way too slow right now, but I wanted to make sure this is what we are trying to look at before optimizing.