Documentation

Documentation corresponding to the repository "FindingLandmarks_a_publication_repository". In the publication "Finding landmarks – an investigation of viewing behavior during spatial navigation in VR using a graph-theoretical analysis approach" by Jasmin L. Walter, Lucas Essmann, Sabine U. König, and Peter König, a new pre-processing pipeline and graph theoretical analysis pipeline is proposed to process eye tracking data recorded in an immersive virtual reality environment with head mounted headsets. This repository contains all the scripts used in this pre-processing pipeline and graph-theoretical analysis. This documentation describes the individual script function, input files, output files, and the variables necessary to customize within the script before running it.

All data files used in the publication are available at the Center for Open Science https://osf.io/aurjk/, DOI 10.17605/OSF.IO/AURJK.

In case of questions, feel free to contact the corresponding author Jasmin L. Walter (jawalter@uos.de)

Content

Pre-processing pipeline:

step1_condenseRawData_V3.m

Description:

First script to run in pre-processing pipeline. Reads in raw Raycast3.0 files and condenses the data into collider clusters. Essentially, the script goes through the data and combines all consecutive data rows with the same hit collider into one row (hit point cluster). The information of the other columns is preserved in arrays accordingly.

Input:

- Raycast3.0_VP*[participantNumber]*.txt
  - File containing the list of hit points obtained after the ray casting process in Unity.
  - All data files are available at the Center for Open Science https://osf.io/aurjk/, DOI 10.17605/OSF.IO/AURJK. For more information about the data format and column content, please refer to the READ_ME.pdf file located at the location of the Raycast3.0 files.

Output:

- *[participantNumber]*_condensedColliders_V3.mat
  - new data files with each row containing the data of a hit point cluster
- OverviewAnalysis.mat
  - summary of number and percentage of data rows with noData (= missing data samples) for each participant
- Missing_Participant_Files.mat
  - contains all participant numbers where the data file could not be loaded when running the script

Variables necessary to adjust:

- savepath
  - should contain the path to the folder where the new data files should be saved
- cd (Matlab identifier to change current folder)
  - should contain the path to the current folder where the 'Raycast3.0_VP*[participantNumber]*.txt' data files are stored
- PartList
  - Should contain the list of participant numbers that should be processed (in cell format)

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S
	TimeStamp	Collider	Distance	hitPointX	hitPointY	hitPointZ	PosX	PosY	PosZ	PosRX	PosRY	PosRZ	PosTimeStamp	PupilLTimeStamp	VectorX	VectorY	VectorZ	eye2Dx	eye2Dy
	0	NH	9.849162	445.4368	0.4	736.6464	455.06	2.15	735.49	2.18	275.82	358.49	0	199674.1	297.1493	-26.56636	754.4654	0.521727	0.358967
	0.03	NH	9.850833	445.4351	0.4	736.6466	455.06	2.15	735.49	2.18	275.82	358.49	0	199674.1	297.1493	-26.56133	754.4654	0.521727	0.358967
	0.07	NH	9.850845	445.4351	0.4	736.6466	455.06	2.15	735.49	2.18	275.82	358.49	0	199674.1	297.1493	-26.56129	754.4654	0.521727	0.358967
	0.1	NH	9.850845	445.4351	0.4	736.6465	455.06	2.15	735.49	2.18	275.82	358.49	0	199674.1	297.1493	-26.56129	754.4653	0.521727	0.358967
	0.13	NH	9.850849	445.4351	0.4	736.6465	455.06	2.15	735.49	2.18	275.82	358.49	0	199674.1	297.1493	-26.56128	754.4653	0.521727	0.358967
	0.16	NH	9.850849	445.4351	0.4	736.6465	455.06	2.15	735.49	2.18	275.82	358.49	0	199674.1	297.1493	-26.56128	754.4653	0.521727	0.358967

Data file: Raycast3.0_VP*[participantNumber]*.txt

Data file: *[participantNumber]* condensedColliders V3.mat

TimeStamp	Collider	Samples	Distance	hitPointX	hitPointY	hitPointZ	PosX	PosY	PosZ	PosRX	PosRY	PosRZ	PosTimeStamp	PupilLTimeStamp	VectorX	VectorY	VectorZ	eye2Dx	eye2Dy
[0 0.0300 0.07...	'NH'	6	[18.7698 1...	[454.8322 4...	[0.4000 0.4...	[753.6976 7...	[455.0600 ...	[2.1400 2.1...	[735.0100 ...	[357.8900 ...	[359.0300 ...	[358.4300 ...	[0 0.0400 0.0800 ...	[4228426 4228426 4...	[453.1021 ...	[-12.8178 -...	[895.6561 ...	[0.5084 0.5...	[0.369
0.2000	'noData'	1	0	454.7729	0.4000	753.5194	455.0600	2.1400	735.0100	357.8100	359.0200	358.4600	0.0800	4228426	0	0	0	0	0
1x17 double	'NH'	17	1x17 double	1x17 double	1x17 double	1x17 double	1x17 double	1x17 double	1x17 double	1x17 double	1x17 double	1x17 double	1x17 double	1x17 double	1x17 double	1x17 double	1x17 double	1x17 double	1x17 d
[0.7900 0.8200]	'noData'	2	[0 0]	[454.4888 4...	[0.4000 0.4...	[757.3000 7...	[455.0600 ...	[2.1400 2.1...	[735.0100 ...	[357.9700 ...	[359.0700 ...	[358.3700 ...	[0.7600 0.8100]	[4228426 4228427]	[0 0]	[0 0]	[0 0]	[0 0]	[0 0]
0.8600	'NH'	1	52.7839	454.4037	0.4000	787.7611	455.0600	2.1400	735.0100	358	359.0900	358.3500	0.8600	4228427	453.0659	-3.1464	895.2769	0.5053	0.432
0.8900	'noData'	1	0	454.4037	0.4000	787.7611	455.0600	2.1400	735.0100	358	359.0900	358.3500	0.8600	4228427	0	0	0	0	0
[0.9200 0.960...	'NH'	5	[52.6859 5...	[454.2784 4...	[0.4000 0.4...	[787.6614 7...	[455.0600 ...	[2.1400 2.1...	[735.0100 ...	[358.0300 ...	[359.0700 ...	[358.3600 ...	[0.9000 0.9400 0...	[4228427 4228427 4...	[452.6809 ...	[-3.1560 -3...	[895.2624 ...	[0.5038 0.5...	[0.432
1x12 double	'NH'	12	1x12 double	1x12 double	1x12 double	1x12 double	1x12 double	1x12 double	1x12 double	1x12 double	1x12 double	1x12 double	1x12 double	1x12 double	1x12 double	1x12 double	1x12 double	1x12 double	1x12 d
1.4800	'NH'	1	46.6262	454.1320	0.4000	781.5945	455.0600	2.1400	735.0100	358.1200	358.9500	358.4700	1.4600	4228427	451.8674	-3.8462	895.2764	0.4997	0.4280
[1.5200 1.550...	'noData'	3	[0 0]	[454.1320 4...	[0.4000 0.4...	[781.5945 7...	[455.0600 ...	[2.1400 2.1...	[735.0100 ...	[358.0200 ...	[358.9800 ...	[358.4700 ...	[1.5000 1.5400 1...	[4228427 4228427 4...	[0 0]	[0 0]	[0 0]	[0 0]	[0 0 0]
1.6200	'NH'	1	53.1448	454.2080	0.4000	788.1195	455.0600	2.1400	735.0100	357.8600	359.1100	358.5100	1.5900	4228427	452.4887	-3.1112	895.2921	0.5001	0.4299
[1.6500 1.6800]	'noData'	2	[0 0]	[454.2080 4...	[0.4000 0.4...	[788.1195 7...	[455.0600 ...	[2.1400 2.1...	[735.0100 ...	[357.8700 ...	[359.1800 ...	[358.5000 ...	[1.6300 1.6800]	[4228427 4228427	[0 0]	[0 0]	[0 0]	[0 0]	[0 0

step2_clean_Participants_V3.m

Description:

Step 2 in pre-processing pipeline. The script identifies the participants that have more than 30% of their data labeled missing data (noData), thus, data samples during which pupil recognition was below 50%. It also creates a new participant list.

Input:

- overviewAnalysis.mat
  - Overview of how much data samples were missing for each participant. The file is created when running script step1_condenseRawData_V3.m

Output:

- newParticipantList
  - file containing the cleaned participant list
- NewDataOverview
  - data overview of the participants who are in the new list
- discartedDataOverview
  - data overview of the removed participants

Variables necessary to adjust:

- savepath
  - should contain the path to the folder where the new data files should be saved
- cd (Matlab identifier to change current folder)
  - should contain the path to the current folder where the condensedColliders_V3.mat data files are stored
- PartList
  - Should contain the list of participant numbers that should be processed (in cell format)

step2_optional_join3SessionsVR_V3.m

Description:

Optional step after step 2 in pre-processing pipeline (only necessary if data acquisition was completed in several sessions). The script combines the condensedColliders files of different VR sessions into one file. As it is implemented currently, it combines the three 30 min sessions of one participant into one file.

Input:

- *[participantNumber]_*condensedColliders_V3.mat
  - files created when running script step1_condenseRawData_V3.m
  - Is a list that contains the participant numbers belonging together sorted into the different sessions. Here the list "combinedSessions_newPartNumbers.csv" is used. It is uploaded with the other experiment data.

Output:

- combinedSessions_newPartNumbers.csv
  - list matching the different numbers of each session to the respective participant (uploaded at https://osf.io/aurjk/)
- *[participantNumber]*_condensedColliders3S.mat
  - file for each participant containing the three recording sessions in correct order

Variables necessary to adjust:

- savepath
  - should contain the path to the folder where the new data files should be saved
- cd (Matlab identifier to change current folder)
  - should contain the path to the current folder where the condensedColliders_V3.mat data files are stored
- listpath
  - should contain path and name of list that contains the participant numbers belonging together sorted into the different sessions.

step3_interpolateLostData_V3.m

Description:

Third script to run in pre-processing pipeline applies the interpolation of lost data samples (noData), but only if the interpolation conditions apply, meaning it interpolates clusters only if the noData cluster is smaller than 8 samples and occurs between the same collider.

Input:

- *[participantNumber]*_condensedColliders_3Sessions_V3.mat
  - the condesedColliders file after combining all 3 session in the script "step2_optional_join3SessionsVR_V3"

Output:

- *[participantNumber]*_interpolatedColliders_3Sessions_V3.mat
  - the newly interpolated data file
- Missing_Participant_Files.mat
  - contains all participant numbers where the data file could not be loaded when running the script

Variables necessary to adjust:

- savepath
  - should contain the path to the folder where the new data files should be saved
- cd (Matlab identifier to change current folder)
  - should contain the path to the current folder where the condensedColliders_V3.mat (or the data files condensedColliders_3Sessions_V3.mat) are stored
- PartList
  - Should contain the list of participant numbers that should be processed (in cell format)

step4_gazes_vs_noise_V3.m

Description:

Fourth step in the pre-processing pipeline. The script divides the interpolatedCollider data based on the gaze threshold into gazes and noisy samples (excluded data), i.e. it identifies the gaze events.

Input:

- *[participantNumber]_*interpolatedColliders_3Sessions_V3.mat
  - the interpolated data file created when running step3_interpolatedLostData_V3.m

Output:

- *[participantNumber]_*gazes_data_V3.mat
  - a new data file containing all gazes
- *[participantNumber]_*noisy_data_V3.mat
  - all excluded data
- Overview_Gazes.mat
  - overview of the amount of gazes and excluded data for each participant
- Missing_Participant_Files.mat
  - contains all participant numbers where the data file could not be loaded when running the script

Variables necessary to adjust:

- savepath
  - should contain the path to the folder where the new data files should be saved
- cd (Matlab identifier to change current folder)
  - should contain the path to the current folder where the interpolatedColliders_3Sessions_V3.mat data files are stored
- PartList
  - Should contain the list of participant numbers that should be processed (in cell format)

step5_optional_create_Graphs_V3.m

Description:

The fith and last step of pre-preocessing pipeline. The script creates the gaze graphs from the gaze events. This step is unnecessary if analysis does not involve graphs. The script creates unweighted and binary graph objects the gaze events. To achieve this, it removes all repetition and self references from the graphs. It also removes the noData node after creation of graph.

Input:

- *[participantNumber]._*gazes_data_V3.mat
  - a new data file containing all gazes created when running step4_gazes_vs_noise_V3.m

Output:

- *[participantNumber]*_Graph_V3.mat
  - the gaze graph object for every participant
- Missing_Participant_Files.mat
  - contains all participant numbers where the data file could not be loaded when running the script

Variables necessary to adjust:

- savepath
  - should contain the path to the folder where the new data files should be saved
- cd (Matlab identifier to change current folder)
  - should contain the path to the current folder where the gazes_data_V3.mat data files are stored
- PartList
  - Should contain the list of participant numbers that should be processed (in cell format)

create_gazesCSV.m

Description:

The script creates csv versions of the Matlab gaze data files, created during running the script step4_gazes_vs_noise_V3.m of the preprocessing pipeline.

Input:

- *[participantNumber]*_gazes_data_V3.mat
  - the data file containing all gazes

Output:

- *[participantNumber]*_gazes_data.csv
  - the same file in csv format
- Missing_Participant_Files.mat
  - contains all participant numbers where the data file could not be loaded when running the script

Variables necessary to adjust:

- savepath
  - should contain the path to the folder where the new data files should be saved
- cd (Matlab identifier to change current folder)
  - should contain the path to the current folder where the gazes_data_V3.mat data files are stored
- PartList
  - Should contain the list of participant numbers that should be processed (in cell format)

create_noisyDataCSV.m

Description:

The script creates csv versions of the Matlab noisy data files (excluded data), created during running the script step4_gazes_vs_noise_V3.m of the preprocessing pipeline.

Input:

- *[participantNumber]_* noisy_data_V3.mat
  - the data file containing all excluded / noisy data

Output:

- *[participantNumber]_noisy_data*.csv
  - the same file in csv format
- Missing_Participant_Files.mat
  - contains all participant numbers where the data file could not be loaded when running the script

Variables necessary to adjust:

- savepath
  - should contain the path to the folder where the new data files should be saved
- cd (Matlab identifier to change current folder)
  - should contain the path to the current folder where the noisy_data_V3.mat data files are stored
- PartList
  - Should contain the list of participant numbers that should be processed (in cell format)

Analysis and visualization scripts:

The analysis and visualization scripts are sorted corresponding to the order of the result section in the publication. If the pre-processed data is run through the scripts in the order listed here, a full reproduction of the results in the publication will be produced.

analysis_gazes_clusterLength.m

Description:

Combines all participant data after interpolation. In addition, it creates the pie plot comparing the number of gazes with the number of excluded data (Fig. 2c in the paper). Thirdly, it creates the histograms of the distribution of cluster durations (Fig. 2b in the paper)

Input:

- interpolatedColliders_3Sessions_V3
    - the interpolated data files with all hit point clusters (not gazes!)
    - created when running the step3_interpolateLostData_V3.m

Output:

- mean_gazes_noise_distr.png
    - pie plot of gazes and excluded data (fig. 2c)
- viewing_duration_all.png
    - histogram of hit point cluster distribution
- viewing_duration_bigCombined.png
    - histogram of hit point cluster
- distribution where all clusters
    - bigger than 1000ms are combined in the last bin (fig. 2b)
- OverviewGazes.mat
    - summary of number and percentage of data rows with noData (= missing data samples) for each participant
- allInterpolatedData.mat
    - all the loaded data from all participants combined in one file
- allSamples_int.mat
    - all the samples from all the participants combined in one file

Variables necessary to adjust:

- savepath
    - should contain the path to the folder where the new data files should be saved
- cd (Matlab identifier to change current folder)
    - should contain the path to the current folder where the interpolatedColliders_3Sessions_V3.mat data files are stored
- PartList
    - Should contain the list of participant numbers that should be processed (in cell format)

drawGraphMap.m

Description:

Creates a visualization of the graph objects on top of the map of Seahaven for each participant. (Creates figure 3c of the paper)

Input:

- *[participantNumber]_*Graph_V3.mat
  - the gaze graph object for every participant
- Map_Houses_SW2.png
  - image of the map of Seahaven in black and white
  - located in the additional files folder of the repository
- CoordinateListNew.txt
  - csv list of the house names and x,y coordinates corresponding to the map of Seahaven
  - located in the additional files folder of the repository

Output:

- *[participantNumber]_*graphVisualizationSeahaven.png
  - image of the graph visualization on top of the map for each participant
- Missing_Participant_Files.mat
  - contains all participant numbers where the data file could not be loaded

Variables necessary to adjust:

- savepath
  - should contain the path to the folder where the new data files should be saved
- imagepath
  - should contain the path to the map image location
- clistpath
  - should contain the path to the coordinate list location
- cd (Matlab identifier to change current folder)
  - should contain the path to the current folder where the noisy_data_V3.mat data files are stored
- PartList
  - Should contain the list of participant numbers that should be processed (in cell format)

plot_graph_creation.m

Description:

Creates a visualization of the graph creation process (Fig. 3a, 3b). Note that for a good visualization of the corresponding graph on top of the map (Fig. 3b), it is necessary to manually zoom into the figure produced by the script.

Input:

- *[participantNumber]*_gazes_data_V3.mat
  - data file containing all gazes
- Map_Houses_New.png
  - image of the map of Seahaven
  - located in the additional files folder of the repository
- CoordinateListNew.txt
  - csv list of the house names and x,y coordinates corresponding to the map of Seahaven
  - located in the additional files folder of the repository

Output:

- Figure 1 (needs to be saved manually)
  - Timeline of gaze events within specified scope (Fig. 4a) – color coded
- Figure 2 (needs to be saved manually)
  - Timeline of gaze events (Fig. 4a) – only blue marked gazes
- Figure 3 (needs to be zoomed in and saved manually)
  - Graph on top of the map corresponding to the time line and specified by the scope variable. The edges are numbered according to their creation order

Variables necessary to adjust:

- savepath
  - should contain the path to the folder where the new data files should be saved
- imagepath
  - should contain the path to the map image location
- clistpath
  - should contain the path to the coordinate list location
- cd (Matlab identifier to change current folder)
  - should contain the path to the current folder where the gazes_data_V3.mat data files are stored
- PartList
  - Should contain the list of participant numbers that should be processed (in cell format)
- scope
  - Identifies the length of the time line – how much of the data should be visualized

SpectralPartitioning.m

Description:

The spectral partitioning of the graphs, i.e. finding two clusters with maximized within-cluster and minimized between-cluster connections, consists of three steps:

1. Pre-processing: Calculating the Laplacian Matrix of the graph
2. Decomposition: Computing the eigenvalues and corresponding eigenvectors of the matrix. The eigenvector entries are corresponding to the respective node of the graph. The smallest eigenvalue is always zero since the matrix is linearly independent. If the second smallest

eigenvalue is not zero, continue. Otherwise, the graph is not fully connected and either the node has to be removed or the third smallest eigenvalue has to be used.

3. Grouping: Sorting the eigenvector of the second smallest eigenvalue in ascending order and splitting the entries (corresponding to the graph nodes) into negative and positive component to acquire two clusters.

Input:

- *[participantNumber]*_gazes_data_V3.mat
  - data file containing all gazes

Output:

- 2ndSmallestEigenvector.png
  - The second smallest eigenvector of the Laplacian matrix is sorted ascendingly and color coded into two clusters. (Fig 4c in the paper)
- Cluster.png
  - cluster visualization of the graph
- Spy_AdjacencyMatrix.png
  - The sparsity pattern of the graph's adjacency matrix sorted by the entries in second smallest eigenvector. Color coded into edges between nodes of one cluster (green), edges between nodes of the other cluster (red), edges between nodes of the two clusters (black) and a distinction between the clusters (yellow) (Fig 4a in the paper)
- Histogram_2nd_Smallest_EigentvectorL
  - Histogram of the distribution of second smallest eigenvector for each participant
- eig_neg.mat
  - negative part of the eigenvector
- eig_pos.mat
  - positive part of the eigenvector
- EigenvalueSpectrumL
  - all Eigenvalues
- SpectralDocumentation.mat
  - overview of Eigenvalue statistics over all participants

Variables necessary to adjust:

- Savepath and directory path
  - currently bound to the script location, needs to be changed at will
- participant list
  - Currently created using the naming convention of the graph files, i.e. partNum_graph.mat with a 4 integer long participant number

drawGraphMap_centrality_V3.m

Description:

Creates a visualization of the graph on top of the map of Seahaven while color coding the nodes according to their node degree value.

Input:

- *[participantNumber]_*Graph_V3.mat
    - o the gaze graph object for every participant
- Map_Houses_New.png
    - o image of the map of Seahaven
    - o located in the additional files folder of the repository
- CoordinateListNew.txt
    - o csv list of the house names and x,y coordinates corresponding to the map of Seahaven
    - o located in the additional files folder of the repository

Output:

- Graph_nodeDegree_visualizationMap.png
    - o image of the graph visualization with nodes color coded according to their node degree centrality on top of the map for each participant
- Missing_Participant_Files.mat
    - o contains all participant numbers where the data file could not be loaded

Variables necessary to adjust:

- savepath
    - o should contain the path to the folder where the new data files should be saved
- imagepath
    - o should contain the path to the map image location
- clistpath
    - o should contain the path to the coordinate list location
- cd (Matlab identifier to change current folder)
    - o should contain the path to the current folder where the graph files are stored
- PartList
    - o Should contain the list of participant numbers that should be processed (in cell format)
- nodecolor
    - o specifies the color map, can be adjusted if desired

nodeDegree_createOverview_V3.m

Description:

Creates the node degree overview file used in other analysis scripts. It reads in the graphs from all participants, calculates the node degree centrality values for every house and saves it in the overview.

Input:

- *[participantNumber]_*Graph_V3.mat
    - o the gaze graph object for every participant
- CoordinateListNew.txt
    - o csv list of the house names and x,y coordinates corresponding to the map of Seahaven

- located in the additional files folder of the repository

Output:

- Overview_NodeDegree.mat
  - table consisting of all node degree values for all participants
- Missing_Participant_Files.mat
  - contains all participant numbers where the data file could not be loaded

Variables necessary to adjust:

- savepath
  - should contain the path to the folder where the new data files should be saved
- clistpath
  - should contain the path to the coordinate list location
- cd (Matlab identifier to change current folder)
  - should contain the path to the current folder where the Graph_V3.mat data files are stored

node_degree_centrality_analysis.m

Description:

Creates an image scale (pseudo 3D plot) color coding the node degree centrality values for every house and every participant (Fig. 5c). Also creates corresponding box plots with error bars: the individual mean node degrees of all subjects (Fig. 5a) and the individual mean node degree of each house (Fig. 5e)

Input:

- Overview_NodeDegree.mat
  - table consisting of all node degree values for all participants
  - Created when running the script nodeDegree_createOverview_V3.m

Output:

- nodeDegree_imageScale.png
  - pseudo 3D plot color coding the node degree centrality values for every house and every participant (Fig. 5c)
- nodeDegree_mean_std_allHouses.png
  - error bar plot of mean and std for all houses (Fig. 5e)
- nodeDegree_mean_std_allParticipants.png
  - error bar plot of mean and std for each participant (Fig. 5b)

Variables necessary to adjust:

- savepath
  - should contain the path to the folder where the new data files should be saved
- cd (Matlab identifier to change current folder)
  - should contain the path to the current folder where the overviewNodeDegree file is stored

Correlation_betweenSubjects.m

Description:

Calculating the correlation coefficients between all subjects. Including a fisher z transformation to normalize the values.

Input:

- CentralityOverview table consisting of all node degree values (n) of all subjects (m) of shape mxn

Output:

- List of all correlation coefficients and the average coefficient.

Variables necessary to adjust:

- Savepath and directory path
    - currently bound to the script location, needs to be changed at will

HierarchyIndex.m

Description:

Calculating the hierarchy index consists of 2 steps:

1. Plotting all occurring logarithms of the degree values against their frequency
2. Fitting a linear function to the data (a*x+b) considering only the median of the node degree distribution.

The hierarchy index is -b. The higher the index, the stronger the hierarchical configuration of the network. Values below 1 are considered as no hierarchical effect.

Input:

- Unweighted, undirected graphs

Output:

- A documentation of the Hierarchy index values
- A scatter plot with the fitted line
- A histogram of the subject's hierarchy indices

Variables necessary to adjust:

- Savepath and directory path
    - currently bound to the script location, needs to be changed at will
- participant list
    - Currently created using the naming convention of the graph files, i.e. partNum_graph.mat with a 4 integer long participant number

RichClub.m

Description:

We calculated the connectivity between nodes with a specific degree value using

$$RC(k) = \frac{2E_{\geq k}}{N_{\geq k}(N_{\geq k} - 1)}$$

with k as the set node degree of the rich club, $E_{\geq k}$ as the number of edges between nodes with degree larger or equal to k, and $N_{\geq k}$ as the number of nodes with degree larger or equal to k. Thus, the rich club coefficient is the fraction of edges instantiated between nodes of degree k or larger and the total number of edges possible between nodes the same degree or larger.

But since the rich club coefficient is an abstract measure, the script creates a random graph based on the degree distribution of the original graph and calculates the respective rich club of this random graph. Afterwards, it divides the 'RealCoefficient' by the 'RandomCoefficient'. Therefore, a value above 1 would indicate that high node degree nodes are connected to other high node degree nodes above chance level.

Input:

- *[participantNumber]*_Graph_V3.mat
    - the gaze graph object for every participant
- Map_Houses_New.png
    - image of the map of Seahaven
    - located in the additional files folder of the repository
- CoordinateListNew.txt
    - csv list of the house names and x,y coordinates corresponding to the map of Seahaven
    - located in the additional files folder of the repository

Output:

- Figure 1
    - All houses displayed on the map both color coded and size coded according to their frequency of being part of the rich club across participants (Fig 8b in paper)
- Figure 2 saved as MeanRichClub.png
    - The development of the rich club coefficient with increasing node degree. The dot-lines are the rich club coefficients of individual participants, while the green line is the mean across all participants (Fig 8a in paper)
- RichClub_AllSubs.mat
    - overview of all rich club values as a function of the threshold node degree for each participant. Note: that the columns correspond to the selection of the houses included in the rich club based on the node degree - so column 2 corresponds to the first threshold 1 = all nodes of a degree 1 or larger are considered. Consequently, column 3 corresponds to the threshold: all nodes of degree 2 and larger etc.
- Mean_RichClub.mat
    - overview of the mean rich club values as a function of the threshold node degree averaged over all participants. Here, column 1 corresponds to the threshold of all

nodes of a degree of 1 and larger and column 2 corresponds to the threshold of all nodes of a degree of 2 and larger etc.

- List_RichClub_Frequency_ofallHouses.mat
  - overview of all houses and the frequency of the houses appearing in a rich club across participants

Variables necessary to adjust:

- savepath
  - should contain the path to the folder where the new data files should be saved
- imagepath
  - should contain the path to the map image location
- clistpath
  - should contain the path to the coordinate list location
- cd (Matlab identifier to change current folder)
  - should contain the path to the current folder where the graph_V3.mat data files are stored

gazes_allParticipants_V3.m

Description:

Script combines all gazes data from all participants into one file. Script also combines all interpolated data files from all participants into one file.

Input:

- *[participantNumber]*_gazes_data_V3.mat
  - gaze data files created when running step4_gazes_vs_noise_V3.m
- *[participantNumber]*_interpolatedColliders_3Sessions_V3.mat
  - interpolated data files created when running step3_interpolateLostData_V3.m

Output:

- gazes_allParticipants.mat
  - combined gaze files from all participants
- interpolData_allParticipants.mat
  - combined interpolated data files from all participants
- Missing_Participant_Files.mat
  - contains all participant numbers where the data file could not be loaded
  - NOTE: only checks the gazes files for missing participants

Variables necessary to adjust:

- savepath
  - should contain the path to the folder where the new data files should be saved
- gazePath
  - should contain the path to the folder where the gaze files are saved
- interpolPath
  - should contain the path to the folder where the interpolatedCollider files are saved
- PartList
  - Should contain the list of participant numbers that should be processed (in cell format)

triangulation_analysis_V3.m

Description:

Script analyses how many gaze-graph-defined landmarks were viewed from each location participants visited in the city. In addition, it analyses how much of the total experiment time participants spend in these areas where the theoretical basis for triangulation would be given. The analysis is performed with a spatial resolution of 4x4m and an additional smoothing with a 3x3 unity kernel.

Input:

- gazes_allParticipants.mat
  - data file containing all gazes
  - created when running script gazes_allParticipants_V3.m
- interpolData_allParticipants.mat
  - data file containing all interpolated data from all participants
  - created when running script gazes_allParticipants_V3.m
- Overview_NodeDegree.mat
  - table consisting of all node degree values for all participants (alternatively the list of the rich club count for all houses)
  - created when running script nodeDegree_createOverview_V3
- Map_Houses_New.png
  - image of the map of Seahaven
  - located in the additional files folder of the repository
- CoordinateListNew.txt
  - csv list of the house names and x,y coordinates corresponding to the map of Seahaven
  - located in the additional files folder of the repository

Output:

- Figure 1: visibility of top 10 houses - rich club and node degree
  - map plot color coded for all top 10 houses
- Figure 2: Visibility of top 10 houses - rich club & node degree'
  - like figure 1, but here the map is only color coded to differentiate areas where 0 top 10 houses, 1 top 10 house, and 2 or more top 10 houses were viewed (Fig. 9 of the paper)
- Figure 3: grid size_vibility plots.png
  - visualization of the 4x4 grid dividing the city
- Figure 20: Percentage of possibility to triangulate in walked area
  - pie plot of the percentages of the different city areas participants were located in
- Figure 21: Percentage of times triangulation was possible
  - pie plot of the percentages of experiment time participants spend in triangulation areas (same visualization as Figure 20, but different data)
- table_percentage_triangulation.mat
  - table listing the percentages of the areas in the city where participant viewed 0, 1, 2 or more houses
- table_times_triangulation_possible.mat
  - table listing the percentage of experiment time participants spend in areas where 0, 1, 2 or more houses where visible

Variables necessary to adjust:

- savepath
  - should contain the path to the folder where the new data files should be saved
- imagepath
  - should contain the path to the map image location
- clistpath
  - should contain the path to the coordinate list location
- overviewNDpath
  - should contain the path to the overviewNodeDegree file location
- cd (Matlab identifier to change current folder)
  - should contain the path to the current folder where the gazes_allParticipants.mat and interpolData_allParticipants.mat files are stored

ToyExamplesGraphMeasures.ipynb

Description:

This script includes the creation of the following figures:

- Node degree - toy example
- Graph partitioning - toy example
- Hierarchy index - toy example
- Rich club coefficient - toy example

Output:

- Toy_NodeDegree.png
  - Figure 11 in the publication
- Toy_Partitioning.png
  - Figure 12 in the publication
- Toy_Hierarchy.png
  - Figure 13 in the publication
- RichClub.csv
  - Summary of all rich club values of the toy example
- Toy_RichClub.png
  - Figure 14 in the publication

Variables necessary to adjust:

- DATA_PATH
  - should contain the path to the folder where the new data files should be saved

Additional files in the repository

Files in the folder additional_files located in the repository.

- Map_Houses_New.png
    - image of the map of Seahaven
- Map_Houses_SW2.png
    - image of the map of Seahaven in black and white
- CoordinateListNew.txt
    - csv list of the house names and x,y coordinate corresponding to the map of Seahaven