# Finding landmarks – An investigation of viewing behavior during spatial navigation in VR using a graph-theoretical analysis approach

by Jasmin L. Walter[1¶*], Lucas Essmann[1¶], Sabine U. König[1], Peter König[1,2]

1 Institute of Cognitive Science, University of Osnabrück, Osnabrück, NI, Germany
2 Department of Neurophysiology and Pathophysiology, University Medical Center Hamburg-Eppendorf, Hamburg, HH, Germany

* Corresponding author:
jawalter@uni-osnabrueck.de (JLW)

¶ These authors contributed equally to this work.

**Key words: Spatial Navigation, VR, Graph Theory, Eye Tracking, Landmarks**

# Abstract

Vision provides the most important sensory information for spatial navigation. Recent technical advances allow new options to conduct more naturalistic experiments in virtual reality (VR) while additionally gather data of the viewing behavior with eye tracking investigations. Here, we propose a method that allows to quantify characteristics of visual behavior by using graph-theoretical measures to abstract eye tracking data recorded in a 3D virtual urban environment.

The analysis is based on eye tracking data of 20 participants, who freely explored the virtual city Seahaven for 90 minutes with an immersive VR headset with an inbuild eye tracker. To extract what participants looked at, we defined “gaze” events, from which we created gaze graphs. On these, we applied graph-theoretical measures to reveal the underlying structure of visual attention.

Applying graph partitioning, we found that our virtual environment could be treated as one coherent city. To investigate the importance of houses in the city, we applied the node degree centrality measure. Our results revealed that 10 houses had a node degree that exceeded consistently two-sigma distance from the mean node degree of all other houses. The importance of these houses was supported by the hierarchy index, which showed a clear hierarchical structure of the gaze graphs. As these high node degree houses fulfilled several characteristics of landmarks, we named them “gaze-graph-defined landmarks”. Applying the rich club coefficient, we found that these gaze-graph-defined landmarks were preferentially connected to each other and that participants spend the majority of their experiment time in areas where at least two of those houses were visible.

Our findings do not only provide new experimental evidence for the development of spatial knowledge, but also establish a new methodology to identify and assess the function of landmarks in spatial navigation based on eye tracking data.

# Author Summary

The ability to navigate and orient ourselves in an unknown environment is important in everyday life. To better understand how we are able to learn about a new environment, it is important to study our behavior during the process of spatial navigation. New technical advances allow us to conduct studies in naturalistic virtual environments with participants wearing immersive VR-headsets. In addition, we can use eye trackers to observe the participant’s eye movements. This is interesting, because observing eye movements allows us to observe visual attention and therefore important cognitive processes. But, it can be difficult to analyze eye tracking data that was measured in a VR environment, as there is no established algorithm yet. Therefore, we propose a new method to analyze such eye tracking data. In addition, our method allows us to transform the eye tracking data into graphs which we can use to find new patterns in behavior that were not accessible before. Using this methodology, we found that participants who spend 90 min exploring a new virtual town used some houses as orientation anchors which we call gaze-graph-defined landmarks. Our further analysis revealed also new characteristics of those houses that were not yet associated with landmarks.

# Introduction

Having a sense of orientation and being able to navigate in the world that surrounds us is essential in everyday life. Specifically, the awareness of the own position in space combined with the ability to remember key locations to plan mental routes between them (1) is crucial. This enables efficient navigation to a location by using globally accessible knowledge of a new environment or previously acquired knowledge of a known environment. Overall, the ability to remember and use important locations and their relations is essential for spatial navigation.

In classical research, spatial navigation depends on three types of knowledge (2). First, landmarks are characterized by their salience against their surrounding (3). Consequently, they may serve as anchors for localization and are memorized when exploring a new environment, thus forming landmark knowledge (2,4,5). Second, route knowledge refers to the ability to travel along paths and remember routes between landmarks. It is thought to be acquired during active navigation (6), but does not necessarily contain metric distance or direction information (4). Third, survey knowledge is described as a mental map-like representation developed by assembling landmark and route knowledge (4) including spatial relations between multiple landmarks (7,8). Additionally, survey knowledge is thought to contain information about metric properties and the relation to cardinal direction (9). Overall, knowledge of landmarks, routes and cognitive maps have been cornerstones for spatial navigation research for decades.

In spite of their ubiquitous use, the concept of landmarks, routes and cognitive maps does not come without problems. Often, landmark based navigational learning is investigated in environments where single cues are introduced to serve as landmarks in an otherwise undifferentiated environment. For example, the Morris water-maze is a widely used task to study the physiological mechanisms of spatial learning in rodents (10–12). Similarly, adaptations of the Morris water-maze in virtual environments (13–15) and other maze tasks (16,17) have been used to investigate spatial navigation in humans. Moreover, performing eye tracking allows investigation of visual interaction with landmarks (13,16,18,19). Other studies investigating spatial navigation were conducted in real world natural environments using mobile eye tracking systems (20–25). However, with recent technological advances, the virtual environments used in spatial navigation research have become more complex and naturalistic (18,19). Some studies push for even more naturalistic virtual environments in combination with head mounted virtual reality headsets (6,26,27). This development creates the need to clearly define the concepts of landmarks, route knowledge and survey knowledge in such naturalistic environments.

The increasing use of more realistic VR environments with freedom to move, creates new challenges for the analysis of eye tracking data. In classical visual exploration of static 2D images, fixations and saccades dominate. In a complex 3D environment vestibular-ocular reflexes and pursuit movements additionally occur on a regular basis. However, there is no established algorithm to differentiate this expanded set of eye movements in eye tracking data collected in a 3D environment. Thus, to identify characteristics of visual behaviour during free exploration of a virtual village and how they relate to spatial navigation, a new method is needed.

In this study, we propose a procedure to analyze eye tracking data and a data driven method to objectively define and quantify visual behavior with respect to spatial information. We use a graph-theoretical approach to access global navigation characteristics across participants and investigate the occurrence, connectivity, and navigation function of a subset of houses, consistently outstanding in their graph-theoretical properties. The approach is applied to eye tracking data recorded during exploration of a virtual town. Overall, our findings establish a new methodology to identify and assess the function of outstanding houses in spatial navigation based on eye tracking data.

# Results

We collected eye tracking data of 22 participants (11 females, age: M = 22,86, SD = 2,59) during 90 min of free exploration in the virtual city Seahaven. All participants gave their written informed consent to participate. The study was approved by the Ethics Committee of the Osnabrück University following the Institutional and National Research Committees' ethical standards. The participants wore a VR headset with an inbuild eye tracker. They moved using a controller and physically rotated their body on a swivel chair to turn in the virtual world (Fig. 1a). The virtual city was built on an island and comprised 213 houses (Fig. 1b). Colliders, i.e., transparent closely fit box-like structures, surrounded all houses, trees, and roads. We calculated the viewing direction based on the position of the participant in the VR, the rotation of the headset, and information from the eye tracker. After data collection, we cast a virtual ray in the viewing direction until it hit a collider, indicating that the participant viewed the respective object (28). This process was completed 30 times per second. Aggregated over all participants, this led to about 3.500.000 hit points that built the basis for our data analysis.

a. b.

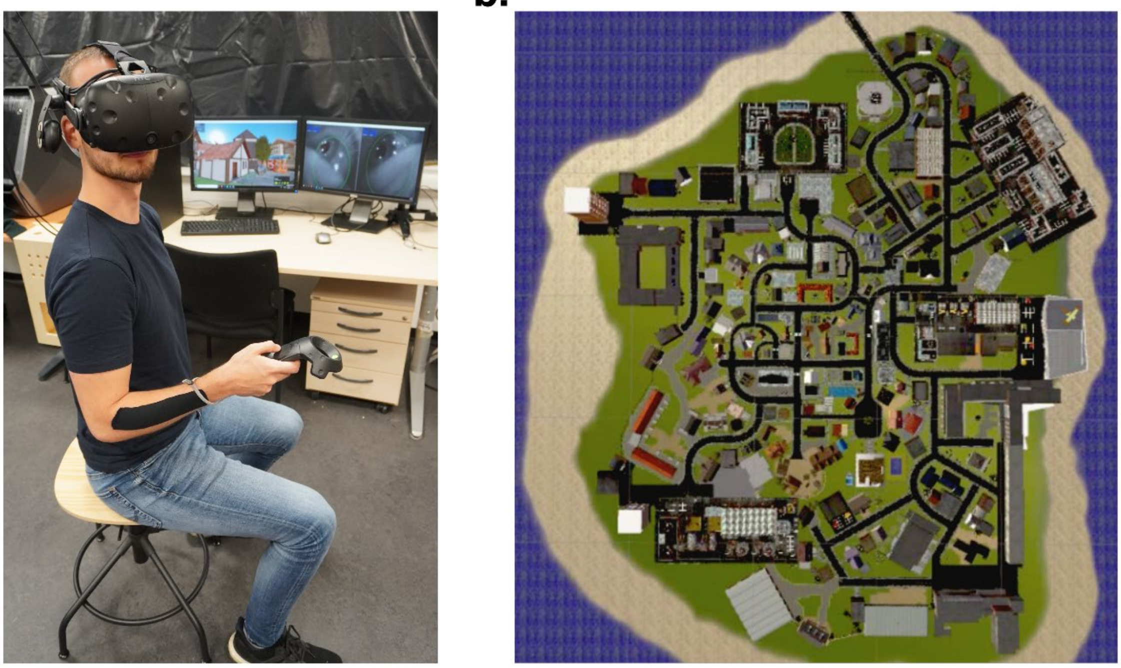

***Figure 1: Experimental Setup and Seahaven (a)****The subject sits on a fully rotatable swivel chair wearing headset and controller. The experimenter's screens in the background allow*

*monitoring the subject's visual field of Seahaven (left) and the pupil labs camera images (right).* ***(b)*** *The island of Seahaven in aerial view.*

## Defining gazes in VR

Eye tracking in an VR environment provides several challenges for data quality. Compared to conventional desktop settings, VR environments allow increased freedom of movement. Under these conditions, subjects perform fixations, saccades, vestibulo-occular reflexes, and smooth pursuit movements. No general algorithm to classify these types of eye movements has been established yet. Thus, the development of appropriate processing algorithms for the eye tracking data is crucial.

As a first step, we use the collider hit points to identify the object the participants focused their gaze on. Each collider corresponds to an entire object; thus, each hit point identifies an entire object. For our analysis, houses form the regions of interests (ROIs). The no house category (NH) summarizes all other collider hits except houses, e.g. grass, roads, trees, and the water. Samples that do not hit a collider, thus hitting the sky or the sun, are identified as "sky" category. We then combine directly consecutive hit points on the same collider to identify clusters. Please note that the clusters do not contain information on where in the participant's visual field the viewed object occurred. This pre-processed data of combined consecutive hit points on the same collider serves as the basis for all following data processing.

As a second step, we address the problem of missing data in individual subjects. We label all data samples that were recorded with a probability of correct pupil detection of less than 50% as "missing data samples". Subsequently, we exclude two participants who had more than 30% of their eye tracking data classified as missing data. All further data analysis is conducted with the data of the remaining 20 participants.

In the following we will need to differentiate between periods where participants could perceive the visual stimuli or not. Given that vestibulo-occular reflexes and smooth pursuit movements stabilize the retinal image in dynamic situations and allow perception in that period similar to fixations (29), here and in all following analysis we subsume these under the general term of fixations. In contrast, participants are blind to visual input during a saccadic suppression (29,30). While classical fixation detection algorithms often differentiate eye movements based on velocity, these eye movements also display a temporal disparity. Specifically, saccades usually range from a duration of 10 to 100 ms, while fixations typically occur from 150 to 600 ms with a mode around 250 ms (29). Therefore, we conjecture that with an appropriate temporal threshold, specifically the time scale of 266 ms which is equivalent to eight data samples, it is possible to identify data clusters containing fixations.

However, since the data still includes a considerable amount of missing data points, we expect a significant number of clusters to be "cut" by missing data points, thus appearing to be of shorter duration. Consequently, using a fixed temporal threshold to identify clusters containing fixations creates the problem of falsely failing to identify these "cut" clusters as clusters containing fixations. Therefore, to counteract this effect, it is crucial to interpolate missing data samples if possible.

Therefore, as a third step in our pre-processing algorithm, we interpolate short intervals of missing data. If seven or fewer data points are missing, it is improbable that subjects had enough time to make a saccade to a different spatial area (ROI), finish a fixation, and make a saccade back to the same area. Consequently, we interpolate data if seven or fewer consecutive data points are missing and only if these occur between two clusters on the same collider (ROI) (Fig. 2a). In case these interpolation criteria apply, the interpolated data points are labelled with the collider name and combined with the two clusters surrounding the short interval of missing data. Larger gaps are not interpolated but treated as missing data. Furthermore, missing data points occurring between clusters of different colliders are not interpolated, independently of the duration of the gap. This procedure ensures to capture most fixations while minimizing false interpretations of missing data.

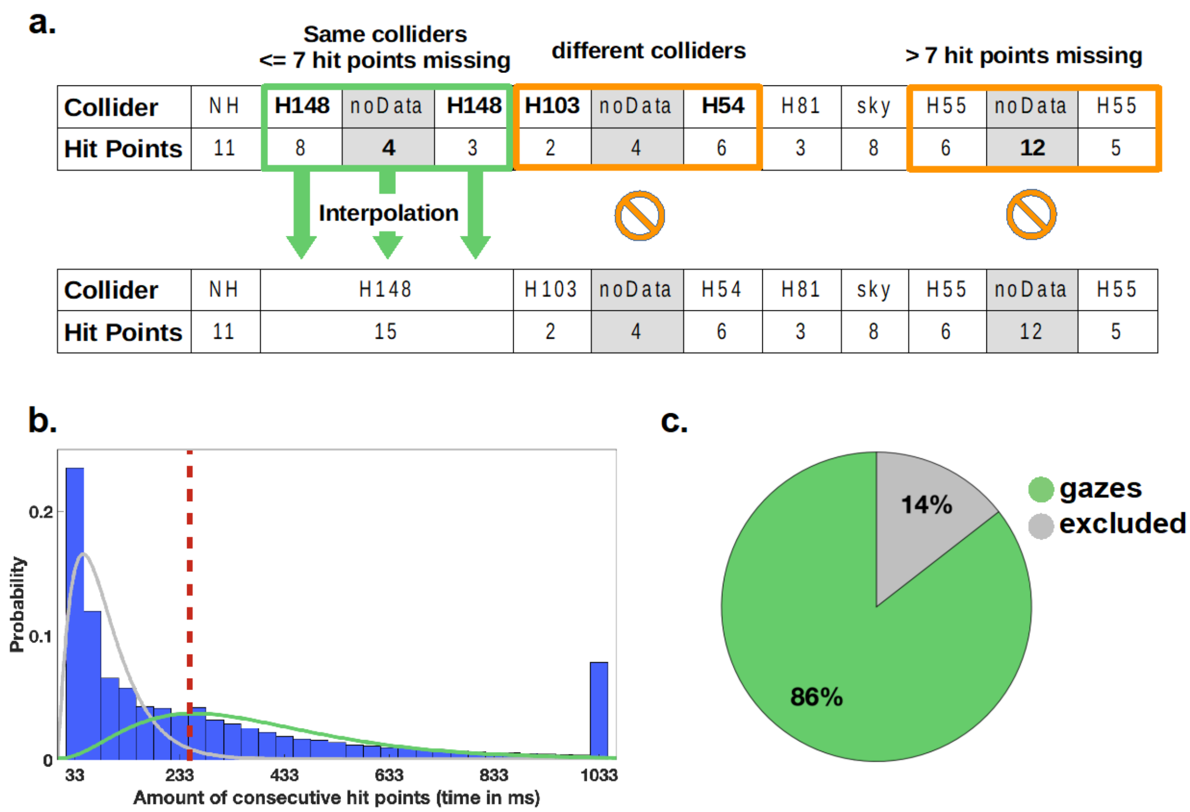


***Figure 2: Defining gazes. (a)*** *We interpolate of missing data only if no more than 7 samples are missing consecutively (pupil detection with less than 50% probability) and if these samples occur between two clusters on the same collider. During the interpolation process, these samples are then unified with the two surrounding clusters to form a new cluster on the same collider. The first row shows three clusters of missing samples (marked noData), while the second row represents the result of the algorithm. In the first cluster (green box), both interpolation conditions apply: there are no more than seven consecutive missing data samples (#4) and they are surrounded by two clusters on the same collider. Consequently, these missing samples are interpolated and combined to a new cluster. In the second cluster (1st orange box) the first interpolation condition applies (#noData samples <= 7) but the cluster occurs between clusters on two different colliders (H103 and H54). Therefore, no interpolation is performed. In the third cluster (2nd orange box) only the second interpolation condition applies. Even though the missing data samples occur between two clusters on the same*

*collider (H55), the first interpolation condition is violated (#noData samples > 7). Consequently, no interpolation is performed.* ***(b)*** *Histogram of hit point cluster length distribution after interpolation. For better visualization all cluster durations longer than 1000 ms are combined in the last bin. The ordinate corresponds to the probability of each duration. Since previous work used gamma distributions to model the distribution of fixation durations or response latencies* (31,32)*, we model the two partly overlapping gamma distributions for visualization only, fitting the distributions of the duration of fixations (green) and non-fixation events (grey). The dashed red line marks the separation threshold for gazes.* ***(c)*** *The pie chart shows the result of the gaze classification across all participants.*

As a fourth step, we finally apply the temporal threshold to identify clusters that contain fixations during which the participants could process the visual input. As described above, we adopt a threshold of eight samples, i.e. 266 ms, to identify clusters that contain at least one fixation (Fig. 2b). Using this approach, shorter clusters likely to be caused by samples during saccades, i.e., periods during which perception was suppressed, would be excluded from further analysis. In the following, we will define these clusters containing fixations as new meaningful “gaze” events and they will form the basis of our further analysis.

This procedure classifies on average 86% of the data samples as belonging to gazes (Fig. 2c). Previous studies reported that humans spent approximately 90% of viewing time on fixations (29). This indicates that our attempt to capture gaze events under dynamic conditions in VR is on the conservative side.

## How to create graphs from gazes

To capture information gathered by the participants during exploration of the virtual town, we create gaze graphs based on the gaze events. In these graphs each node represents a house. Viewing two houses in the VR environment in direct succession gives information on their relative spatial location. Therefore, if anytime during the experiment a gaze event on a house is directly followed by a gaze event on another house the respective nodes are connected by an undirected edge. All gaze graphs are undirected and unweighted. Hence, they do not contain information about the directionality or frequency of their edges (i.e. direction and frequency of succession of gaze events). The completed gaze graphs capture the spatial information obtained by the visual exploration in the virtual town.

The first 30 seconds of one participant exploring Seahaven serve as an example of the process to create a graph (Fig. 3a). The graph contains one node for each house (ROI), but does not consider the spatially unspecific no-house and sky category. The graph creation process starts with an edge between the first house viewed and the second house viewed (Fig. 3b). If the gaze on a house is followed by a missing data cluster, then no edge is created. As stated above, the edges are unweighted, i.e. binary. That is, if the participant looked back to two houses already seen in sequence, an edge between these houses is already in place and the graph is not changed. The process of edge creation is iterated for all gaze transitions. Whenever, a new house is viewed, a new node is created in the graph. Fig. 3c shows a visualization of the result of applying this procedure to the complete data of the example participant.

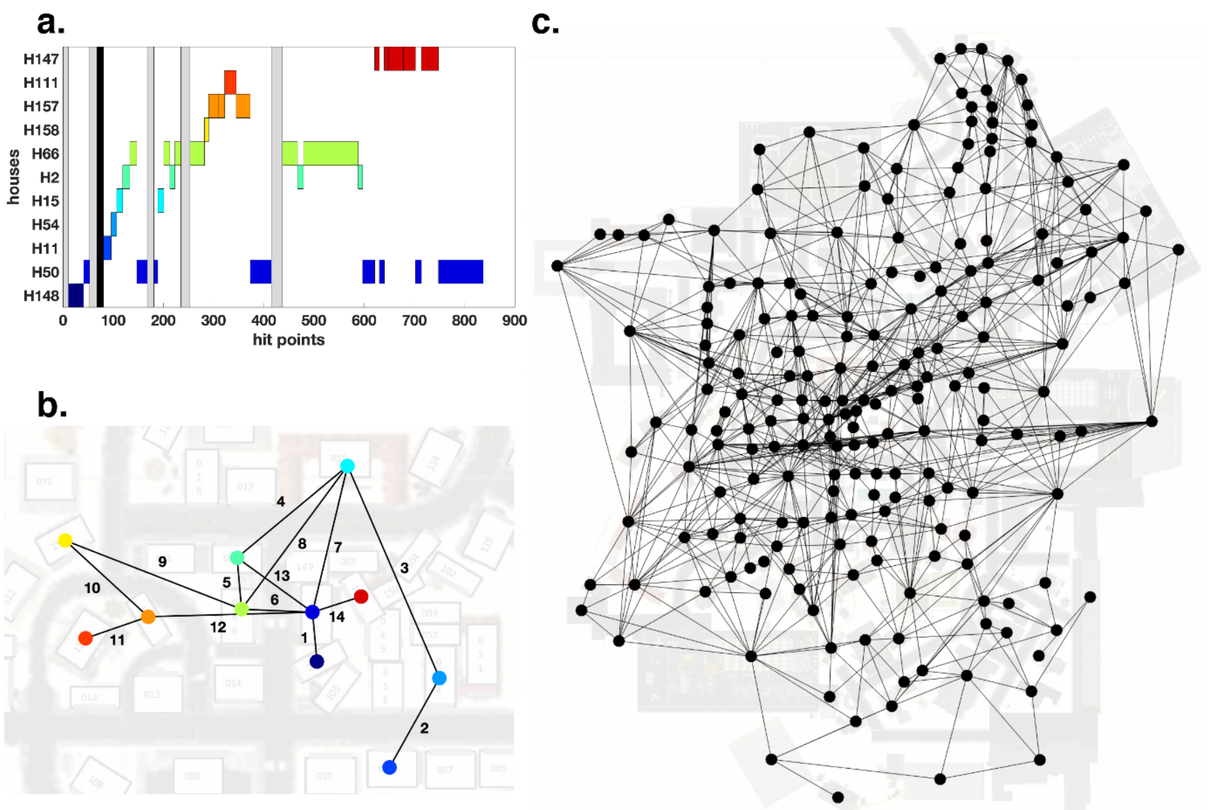


***Figure 3: Graph creation. (a)*** *Time line of gaze events by a participant. The abscissa represents the first 30 seconds (900 hit points) of the recordings. The ordinate contains all viewed houses viewed during that time line. We number houses and name them accordingly, e.g., H148 for house number 148. In this panel each house has a distinct color for visualization only. The grey bars represent clusters of the NH category, which are not considered during graph creation. The black bar identifies a remaining cluster of missing data samples. Therefore, no edge will be created at this moment in the time line.* ***(b)*** *The graph corresponding to the time line of panel A is visualized on top of the map of Seahaven. The colors of the nodes match the colors of the boxes in panel A. Edges are labelled according to the order they were created.* ***(c)*** *The complete graph of a single participant based on all gaze events during 90 min of exploration visualized on top of the 2D map of Seahaven. Note that in this visualization the locations of the nodes correspond to the locations of the respective houses they represent in Seahaven, however, this locational information is not contained in the graph itself.*

## Is it a single city or multiple suburbs? – Graph partitioning

To address questions on spatial cognition, we are interested whether the Seahaven should be treated as a loosely connected set of suburbs or as a coherent single city. The search for distinct clusters in the graph directly relates to the problem of graph partitioning. In the field of graph theory, partitioning is a well investigated problem as it divides a graph into smaller mutually non-overlapping subgraphs. Of specific interest are approaches that maximize within-cluster connections (those that are maintained) and minimize between-cluster connections (those that are cut). Thus, we use graph partitioning to classify Seahaven as a set of suburbs or one coherent city.

For graph partitioning, we employ the spectral graph analysis. This approach includes three steps: the calculation of the graph's Laplacian matrix, finding the second smallest eigenvalue, and splitting the graph based on the corresponding eigenvector. First, we consider the adjacency matrix and the degree matrix of a graph. The former is a binary square matrix where each entry indicates whether an edge connects the two nodes / houses. The latter a square matrix that contains the degree of each node on its diagonal and is zero otherwise. We calculate the Laplacian matrix by subtracting the degree- from the adjacency matrix. Second, according to (33,34), the spectrum of a graph, i.e., the eigenvalues of the Laplacian matrix can be used as a measure of graph connectivity. We calculate these eigenvalues $\mu_i$ and sort them in increasing size. They have the properties that $\mu_1 = 0$ and $\mu_2 \geq 0$. The second property is valid only if the graph is connected. As we eliminate all nodes/houses that were never viewed, this is given in our application ($\mu_2$ = 0.300+/-0.097). Third, we consider the eigenvector with the second smallest eigenvalue and map each node onto the corresponding entry. Sorting the adjacency matrix in this way, a modular structure of the graph would be visible in form of a block structure (Fig. 4a), which, however, is not the case in the present analysis. The partitioning of the graph is performed by splitting of the eigenvector into positive and negative parts (Fig. 4b) and assigning the corresponding nodes to the two separate groups. The larger the gap between positive and negative values, the fewer inter-cluster connections are present. In the present analysis, hardly any gap is visible. Fig. 4c visualizes the results of the graph partitioning by spectral analysis with the nodes of each cluster color coded and plotted onto the city map.

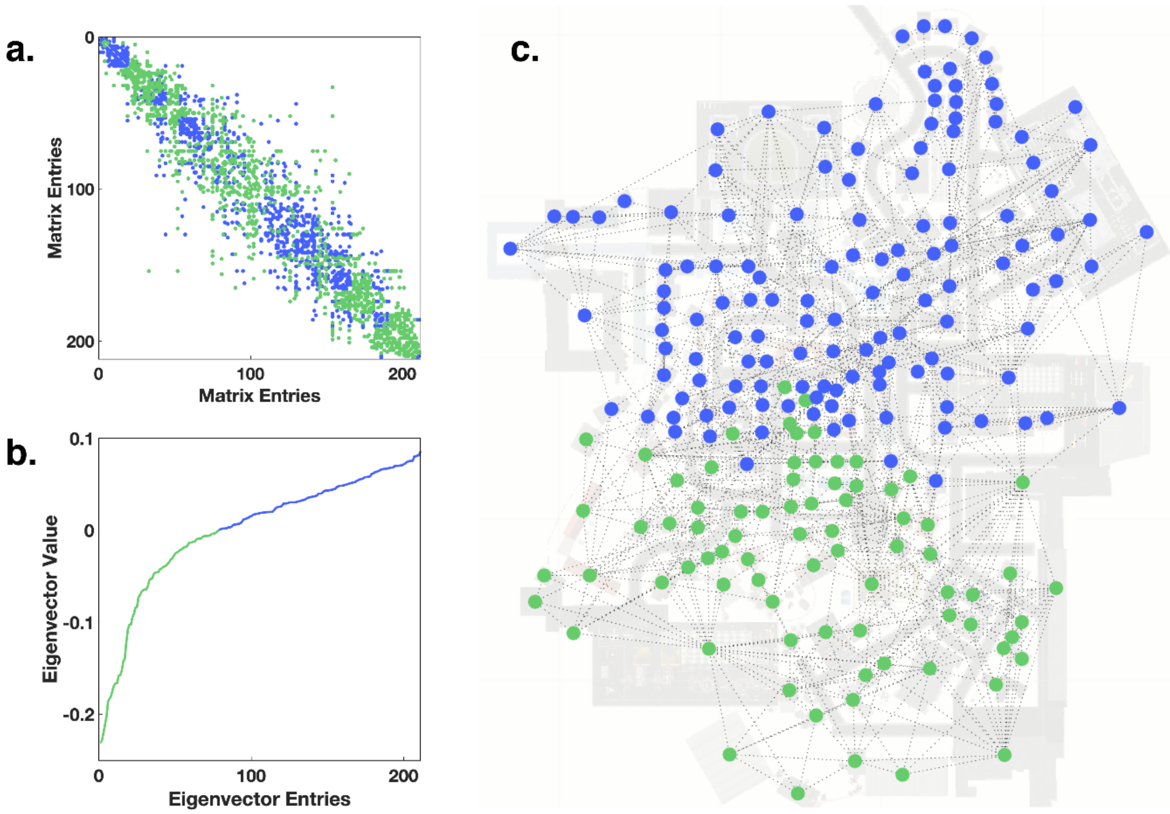


***Figure 4: Graph partitioning. (a)*** *The sparsity pattern of the graph's adjacency matrix sorted by its second smallest eigenvector. Color coded into two clusters obtained by the positive and negative parts of the eigenvector.* ***(b)*** *The second smallest eigenvector of the adjacency matrix*

*is sorted ascendingly and color coded into two clusters.* ***(c)*** *The two clusters are displayed onto the map of subject 35.*

To investigate the quality of the partitioning, we follow the definitions used in Schaeffer (2007). As a measure of cluster goodness, we consider the intra-cluster density and the inter-cluster density. The density of a graph is defined as the ratio of instantiated edges relative to the number of possible edges:

$$\delta(G) = \frac{m}{\binom{n}{2}}$$

with $n = |V|$ as the number of nodes, and $m = |E|$ as the edge count. Furthermore, we define the intra-cluster density as the average of the densities of both clusters and the inter-cluster density as the ratio of inter-cluster edges to the maximum possible inter-cluster edges.

We analyze the graphs of individual participants based on 90 min exploration time. On average, the participants' graphs contain 883 edges. The mean density of the graphs is 0.041+/-0.006. This means that on average 4.1% of all possible edges are instantiated. Furthermore, dividing the graphs into two parts would require on average a cut of at least 9.3% of the edges. Specifically, a sufficient cut has to have 82.6 edges resulting in two clusters with approximately 400 edges each. After partitioning, the mean intra-cluster density was 0.079+/-0.012, while the mean inter-cluster density was 0.0083+/-0.0020. These numbers indicate that the graph cannot be easily partitioned, without cutting a fair number of edges.

In conclusion, our results reveal that the graphs cannot be distinguished into large-scale clusters. That is, the exploration of Seahaven does not show separate city blocks, but was rather well-balanced. Thus, the virtual environment can be treated as one coherent city.

## The distribution of gazes on houses - Node degree centrality

We characterize the role of different houses during visual exploration by indices adapted from graph theory. The node degree centrality is the main and most basic graph-theoretical measure in graph-theoretical research. It is defined as the sum of all edges connecting a node:

$$c(i) = \sum_{j}^{N} x_{ij}$$

with $i$ being the node under consideration and $x_{ij}$ the adjacency matrix. Here, the node degree of a house reflects the number of other houses a participant made a gaze to or from the house under consideration in direct succession. The node degree centrality in the gaze graph thus differentiates houses according to their importance during visual exploration in the virtual town.

First, we use the node degree centrality to investigate the viewing behavior on an individual participant level. Fig. 5a shows the gaze graph of a participant with individual nodes color coded according to the respective node degree. Whereas many nodes have a degree centrality in the single digit range, a few houses reach rather high values. The variance of node degree for each subject ranged between 0 and 33. It is apparent that the range of node degrees is surprisingly large.

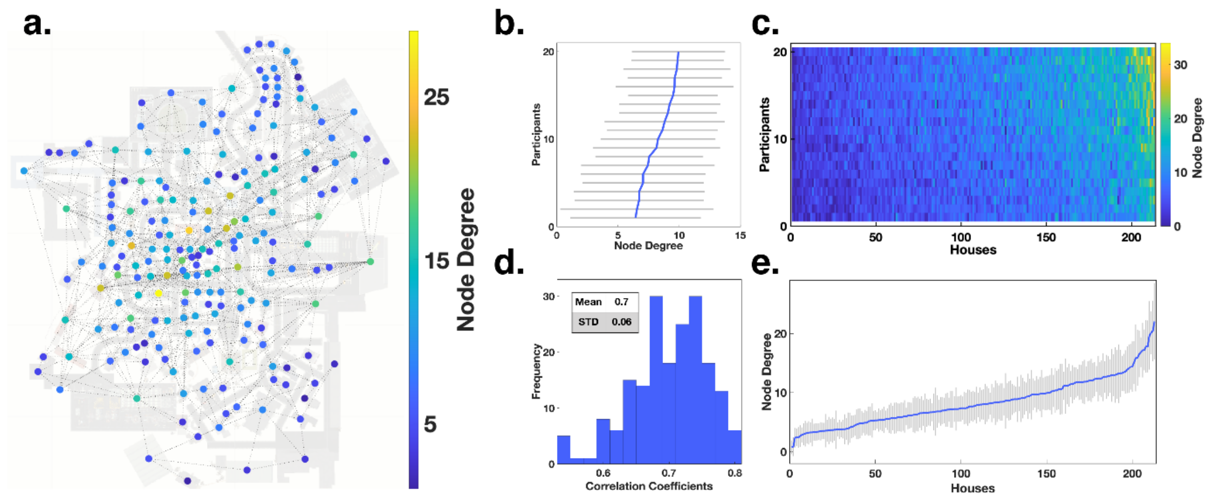


***Figure 5: Node degree centrality. (a)** The graph of one participant is visualized on top of the map of Seahaven. The nodes were colored according to their respective node degree centrality. **(b)** The mean node degree of all subjects (blue line) and their respective standard deviation (grey lines), sorted such that the average node degree increases along the ordinate. **(c)** A pseudo 3D plot color coding the node degree of every house (abscissa) for every subject (ordinate). The houses are sorted so that the average node degree value increases along the abscissa. Similarly, the participants are sorted, so that the average node degree increases along the ordinate. The marginals of this plot result in the panels b and e. **(d)** The distribution of the pairwise inter-subject correlation coefficients of the node degree values of all houses. **(e)** The mean node degree of each house sorted according to the mean node degree along the abscissa (blue line) and their respective standard deviations (grey lines).*

Next, we investigate the similarity of the node degree centrality distribution over subjects. We calculate the average node degree of each participant across all houses (Fig. 5b), which reveals limited variations only. The average node degree of each house across all participants (Fig. 5e) showed a monotonous linear increase. Both aspects are also combined in an image-scale visualization (Fig. 5c) that matches the order of the houses and the order of participants. Sorting the houses according to the average node degree centrality leads to a near monotonous increase on the individual participant level as well. This was quantified by the high correlation of 0.70 (+/- 0.06) of node degree centrality between participants (Fig. 5d). Thus, the node degree centrality in the gaze graph varies considerably while the values of individual houses are rather consistent across participants.

The distribution of node degree centrality over houses reveals interesting aspects. Over a large range of houses the average node degree centrality increases only slowly. Only for the last few houses, we observe a steep increase. These houses stand out from the other houses and are viewed directly before or directly after viewing many other houses. Therefore, the high node degree centrality houses may serve as important reference points.

In summary, the node degree is a simple yet powerful centrality measure that can be used to identify important nodes in the gaze graph, hence important houses in visual behavior. First results indicate that a small number of houses show an especially high node degree across all participants, setting themselves apart from the rest of the city. Interestingly, a small number of houses with subject independent high node degree values, i.e. houses with many visual connections to other houses, would also be the characteristics that we would expect landmarks to display in gaze graphs.

## High node degree centrality houses – Hierarchy index

o further investigate the houses with high node degree in respect to their distribution, we apply the graph-theoretical measurement of the hierarchy index. This index characterizes hierarchical configurations within networks. By applying it to the degree values above the subject's respective median, we focus the index on the upper tail of the distribution of node degree centrality. Here, we fit in a bi-log plot of the node degree centrality with frequency against the node degree values (Fig. 6a), the slope of the regression line starting from the subject's respective median. For the example participant reported already above this results in a slope of -2.6349. Performing this analysis for all participants results in a slope smaller than -2.00 throughout, with a mean of -2.91. Furthermore, the small standard deviation of 0.34 showed, that the hierarchy index is similar across subjects (Fig. 6b). According to (35), networks with strong hierarchical configurations, i.e., with many low degree nodes and few high degree nodes, had a slope of below -2 (or above 2). Therefore, the hierarchy index reveals a clear hierarchical structure of the gaze graphs and emphasizes the importance of the few exceptionally high node degree centrality houses.

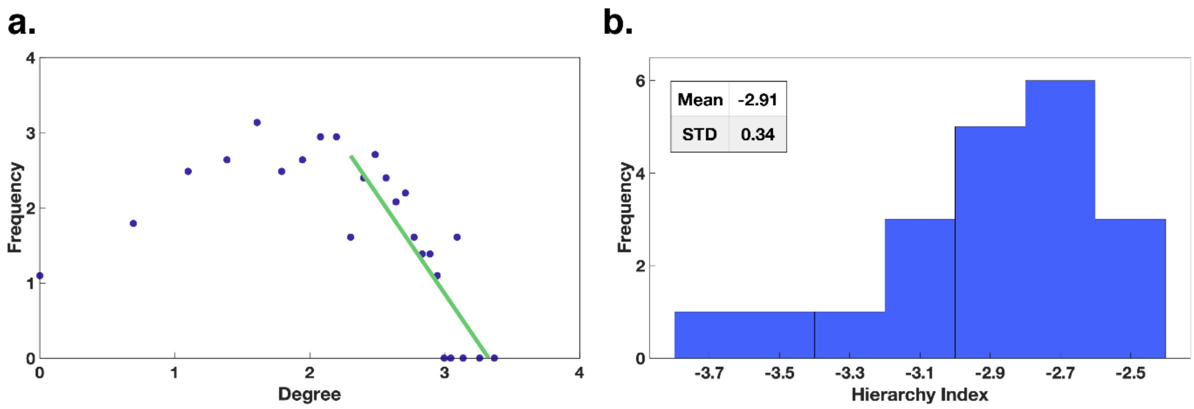


***Figure 6: Hierarchy index. (a)** The frequency of occurrence of the node degree frequency for a single participant. The green line indicates the linear regression starting at the median of the distribution. **(b)** The distribution of the hierarchy index across all subjects.*

This observation triggers a detailed look at the mean node degree distribution over all subjects to identify those special high node degree houses. Plotting the mean node degree of houses averaged across subjects onto the map of Seahaven (Fig. 7c) highlights the scattered subset of high node degree houses. Furthermore, the mean node degree across all houses and all subjects measures 8.3 with a standard deviation of 3.98 (Fig. 7a). We select the value of the 2-sigma distance (16.25) as the threshold to identify the high node degree houses. This results in a set of 10 houses with node degree centrality values exceeding the threshold.

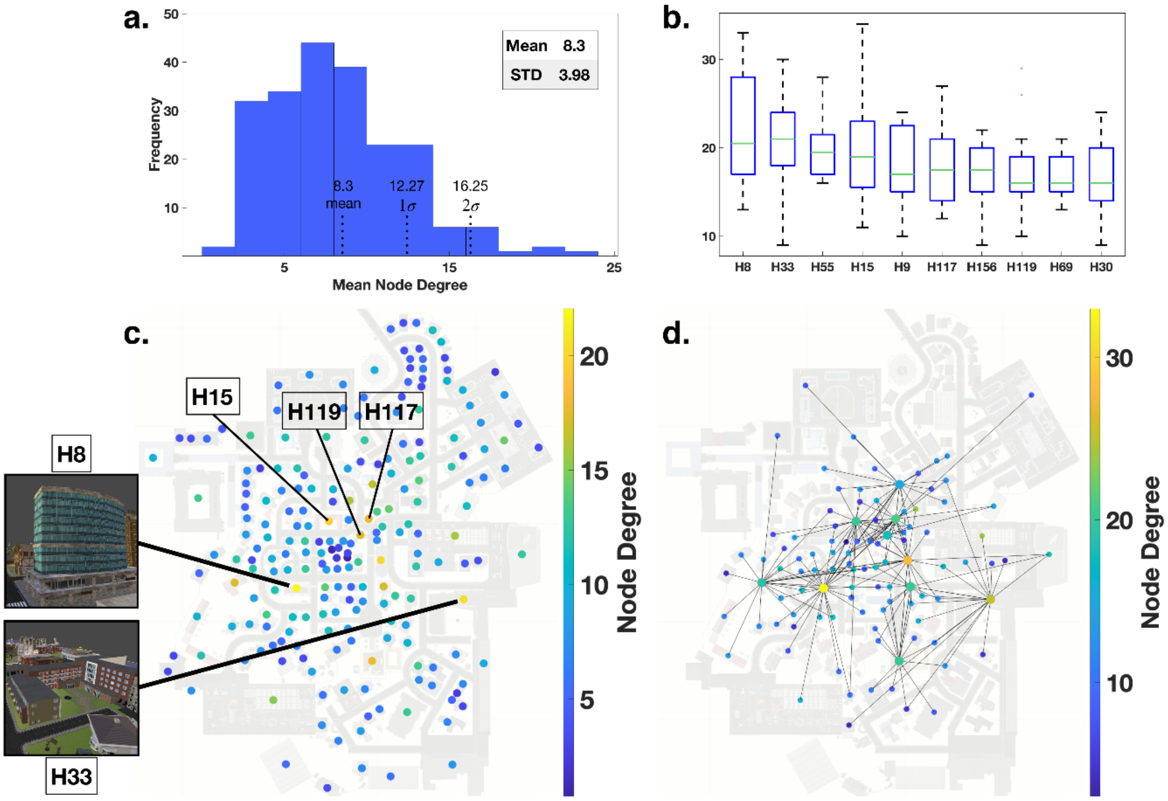


*Figure 7: **The high node degree centrality houses.** **(a)** The mean node degree distribution across all subjects with mean, 1σ- and 2σ-thresholds. **(b)** A box plot of the 10 houses with at least 2σ-distance to the mean node degree. **(c)** A map plot with all nodes color coded with their respective average node degree across all subjects. **(d)** The 10 houses, which exceeded the 2σ-distance to the mean displayed on the map for our example subject with all their connections and color coded with their respective node degree.*

Next, we analyze these houses with respect to their interconnectivity and visibility in the city. The distribution of the node degree centrality over subjects for each of the 10 houses reveals considerable variance (Fig. 7b). However, all 10 houses for all participants have a node degree centrality of 9 or above. Plotting them jointly with all their edges onto the map of Seahaven for our example subject shows that they are located centrally within the city. However, these houses are connected to the outer areas of the city and together cover nearly the whole city (Fig. 7d). All in all, these 10 high node degree houses, are located centrally in the city and their connections reach out into outer areas, i.e. they are viewed from much of the city area.

The characteristics of high visibility of a small number and similar use over participants would be expected in landmarks. Our results show that the node degree distribution is similar across subjects and that only a few houses have exceptionally high node degrees. This is supported by the hierarchical configuration of the network. The 10 houses with an average node degree distribution exceeding the a 2-sigma threshold are more centrally located and had viewing connections into the outer areas covering nearly the whole city. All things considered, these findings suggest the notion that this set of houses is exceptional across multiple domains and displays the characteristics expected from landmarks. Therefore, in the following we will refer to these 10 buildings as "gaze-graph-defined landmarks".

## The connections between the gaze-graph-defined landmarks – The rich club coefficient

In this section we investigate whether the gaze-graph-defined landmarks serve as the core of a network that could be used for navigation in the city. For a quantitative investigation, we applied the concept of the rich club coefficient to our gaze graphs. The rich club coefficient is a frequently practiced graph-theoretical method in network theory and was initially applied in internet network analyses (36). Yet, the rich club coefficient has also been transferred to neuroscientific contexts. The approach has been used to map out both subcortical and neocortical hub regions and to show that those regions with high linkages are also highly connected between each other and, thus, indeed form a rich club (37). In this study, the rich club coefficient allows to quantify in how far gaze-graph-defined landmarks are preferentially connected to each other.

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.

For the interpretation of the rich club coefficient, we need a baseline. For that purpose, we compare generated random graphs with similar statistics of node degrees and calculate their rich club coefficient. For each subject, we generate 1000 random graphs with the same number of nodes as the respective original graph. Of these we select the 10 graphs with the most similar, in terms of the two-sample Kolmogorov Smirnov test, distributions of node degrees to the original distribution. Subsequently, we divide the original rich club coefficient by the 10 random coefficients respectively and averaged. Thus, a value above 1 indicates the existence of a rich club with the respective node degree.

As a final step we investigate the rich club coefficient as a function of the threshold node degree. Due to the strong hierarchy index, the number of nodes decreases drastically with increasing node degree, resulting in increasing uncertainty. Therefore, we cut off the plot at 1σ distance to the mean. With increasing threshold, the interconnectivity of the rich club steadily increases and reaches 1.5 for a node degree of 13 or higher (Fig. 8a). This demonstrates that these nodes are interconnected much more than expected by chance. We calculate the rich club for each subject and specifically mark the top 10 houses with the highest frequency of being part of the rich club across all subjects (Fig 8b). These houses had the averaged highest interconnectivity and were the same houses that were identified as the top 10 node degree houses, i.e., gaze-graph-defined landmarks, earlier. This gives evidence for a highly interconnected network of gaze-graph-defined landmarks in the city, a rich club.

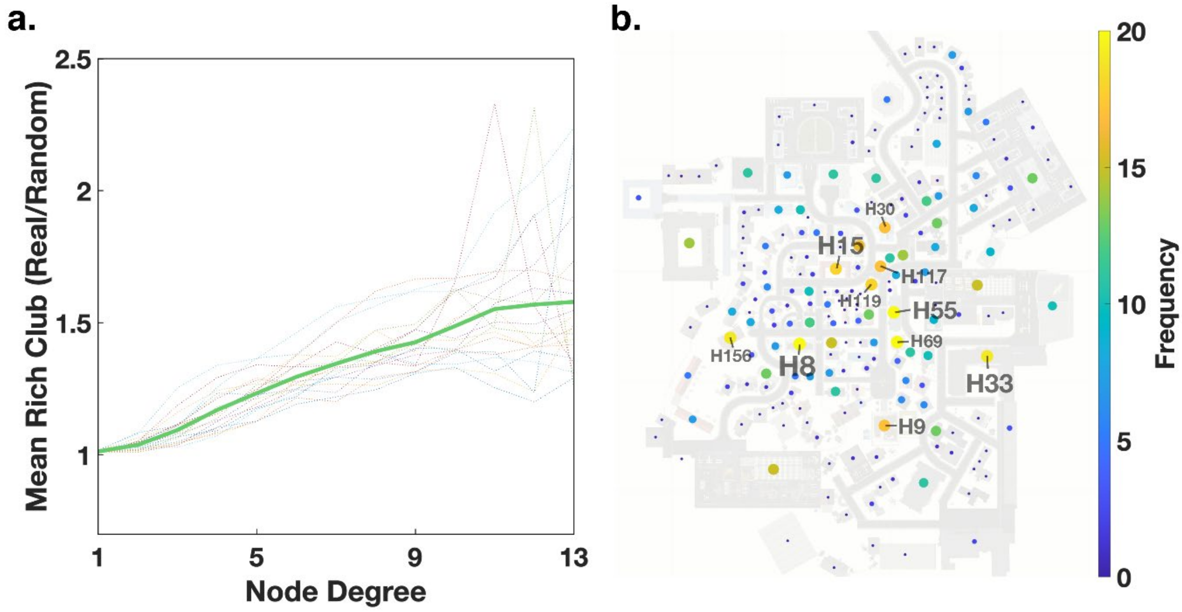


***Figure 8: The rich club coefficient (a)*** *The development of the rich club coefficient with increasing node degree. The dot-lines are the rich club coefficients of individual participants, while the green line is the mean across all subjects.* ***(b)*** *All houses displayed on the map both color coded and size coded according to their frequency of being part of the rich club across subjects.*

## Spatial arrangement of the gaze-graph-defined landmarks - Triangulation

To elucidate the role of the gaze-graph-defined landmarks in spatial navigation, we explore whether they could serve as a basis for triangulation. Triangulation is a method to infer the own location based on the viewing angle in respect to two location anchors. Our analysis has revealed that the gaze-graph-defined landmarks form a highly interconnected rich club. Thus, if the gaze-graph-defined landmarks are visible from most places in the city they could serve as a basis for triangulation.

As a first step we determine the parts of the city where at least one or two of the gaze-graph-defined landmarks were viewed by the subjects. We evaluate the spatial distribution of our participants during exploration of the virtual town and how many gaze-graph-defined landmarks were actually viewed from each location. This analysis is performed with a spatial resolution of 4x4m and an additional smoothing with a 3x3 unity kernel. Please note that this analysis depends on the actual gaze data and, thus, reflect from where participants actually did view the gaze-graph-defined landmarks. Next, we differentiated the locations in three categories: zero/one/two gaze-graph-defined landmark were viewed. The resulting map represents the potential of triangulation based on the gaze-graph-defined landmarks at different locations in the virtual city (Fig. 9). In 39.1% of the city areas that were visited by participants, two or more gaze-graph-defined landmarks were viewed from that location, providing the basis for triangulation. In an additional 32.7% of the city area, exactly one gaze-graph-defined landmark was viewed. Only in 28.1% of the visited city area, none of the gaze-graph-defined landmarks were viewed. Weighting the city areas with the absolute time participants were located in, participants spend an even bigger fraction of the experiment time in areas where the theoretical basis of triangulation was given. Specifically, participants spent 53.2% of the experiment time in areas where at least two or more gaze-graph-defined landmarks were viewed, 19.4% of the time in areas where one gaze-graph-defined landmark was viewed and 27.4% of the time in areas where none was viewed. The latter regions were mostly at the fringes of the city map. Overall, our results indicate that triangulation based on gaze-graph-defined landmarks is possible in most parts of the city and participants spent the majority of their time located in these areas.

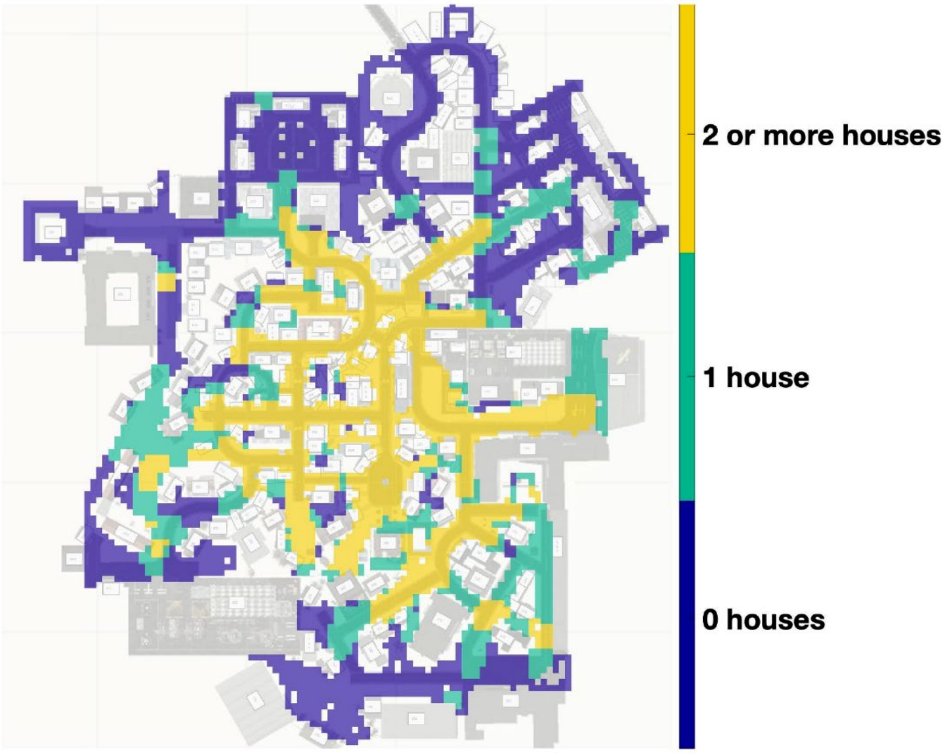


***Figure 9: Triangulation**. Location data of all participants plotted on the map of Seahaven. The color code indicates how many of the gaze-graph-defined landmarks were viewed by participants at each location.*

# Discussion

In this study, we establish a method to quantify characteristics of visual behavior by using graph-theoretical measures to abstract eye tracking data recorded in a 3D virtual urban environment. We define gazes of subjects that freely explored the virtual city Seahaven, and use these to convert the viewing behavior into graphs. In these gaze graphs, nodes represent houses while edges represent their visual connection, i.e., gazes in direct succession on the respective houses. Thus, the gaze graphs capture relevant spatial information gathered during exploration. The node degree centrality graph measure reveals a surprisingly large variance. However, the values of individual houses were rather consistent across subjects, as shown by the high mean correlation. Additionally, we observed that the degree distribution across houses increased steadily, except for only a few high node degree houses. These houses may serve as import reference points, the so-called gaze-graph-defined landmarks. The analysis of the hierarchy index demonstrates that the frequency of houses decreased drastically with increasing node degree, revealing a hierarchical graph structure. The set of identified gaze-graph-defined landmarks were indeed preferentially connected, as demonstrated by the rich club coefficient. Finally, participants spent more than half their exploration time at locations, where at least two of the houses of the rich club were viewed, allowing triangulation for spatial localization. Thus, we presented a graph-theoretical approach for analyzing gaze movements supporting spatial cognition.

In general, spatial navigation is based on multimodal sensory input. In the present study we employ a virtual reality that is more restricted, i.e., limited to visual and vestibular information. Recent literature highlights the importance of idiothetic and sensorimotor information about self-position and about self-motion (38,39). Nevertheless, the dominating sense of spatial perception is vision (40). Only this sense can gather reliable information of space and the environment independently of the physical distance and also allows to perceive topographic characteristics over large distances. Therefore, by observing visual attention and visual behaviour using the method of eye tracking, we could gather important insights about the usage of spatial cues, that indicate cognitive processes related to spatial navigation, specifically landmark usage.

Basing our analysis on eye tracking data allows us to observe and investigate participants visual attention during spatial navigation processing. However, the combination of eye tracking and a head mounted virtual reality headset comes with several challenges. Accuracy of mobile eye tracking systems if often reduced compared to other systems (41). Specifically, in VR experiments, this is often due to the freedom of head movements and weight of the VR headset. Since typical errors due to slippage and head movements increase over time, we conducted a short validation and if necessary, a complete calibration validation procedure every 5 minutes during the experiment (28). Nevertheless, the mean validation error of 1.55° before experiment start and 1.88° during the experiment is rather high compared to classical lab-based eye tracking studies. However, unlike lab-based eye tracking studies, our preprocessing and analysis is based on hit point clusters that fell on the same collider in the VR environment. Thus, we summarized data points that were located in rather close proximity. Though the notion of close proximity must be taken with care since most colliders still corresponded to the size of a complete house, therefore making a small deviation in gaze location due to the validation error less problematic. All in all, considering that the preprocessing is based on the spatial distribution of the accumulated hit point clusters, a minor increase of validation error should not affect our data significantly.

One major factor in the analysis of eye tracking data recorded in an VR environment is the algorithm differentiating the different types of eye movements and in our case the definition and creation of the data form “gaze”. Essentially, the four different types of eye movements expected to occur in such a natural setting, i.e. saccade, fixation, vestibulo-occular reflexes, and smooth pursuit movements, can be separated into two categories of visual input perception. On the one hand, we have fixations as the typical source of visual perception. Since vestibulo-occular reflexes and smooth pursuit movements stabilize the retinal image in dynamic situations, they lead to visual perception similar to the input during fixations (29).

Thus, we can classify fixations, vestibulo-occular reflexes, and smooth pursuit movements as the first category of visual perception. Saccades, on the other hand, render the participant blind to the momentary visual input, hence a hit point sample created during a saccade will not have been perceived by the participant (29,30). Consequently, it is essential to differentiate the data between the first category of visual perception and saccades. Usually, this is done either by velocity or gaze location-based algorithm. However, the virtual environment in Unity3D results in three impeding factors. First, the complexity of the VR environment in Unity3D only allowed for a 30 Hz sampling rate, thus it does not allow a saccade detection based on sudden changes in gaze velocity. Secondly, the gaze location calculated based on the ray cast process is limited to whole colliders, often covering the size of a house, therefore it does not allow to identify small changes in gaze location. And thirdly, depth perception is usually not accounted for in saccade detection algorithms, therefore making it questionable to apply to eye tracking data recorded in a 3D. Consequently, the available eye tracking data recorded in the VR environment did not provide the reliable information necessary in classical fixation detection algorithms that allow to cleanly separate the different eye movements. Therefore, defining the new data form "gaze" allows for a functional method to clean the data. This process did not allow the identification of single eye movements nor did it exclude all saccades from the data. However, by identifying the data clusters that contained at least one fixation and excluded the data clusters that very likely did not contain any fixation, we could clean the data from samples, that were unlikely to be visually processed by the participant. This is further supported by our data, since the process identified 86% of the data as gazes on average across all participants. With approximately 90% of viewing time being expected to be spend on fixations (29), the process identifying gazes appears to be on the conservative side. Overall, with higher sampling rates available in virtual environments, new options to identify saccades might become available. However, we believe that given the data we have available, defining the new data form "gaze" was the best option to differentiate between stray samples unprocessed by the participant and meaningful data carrying important information of the first category of visual perception including fixations, vestibulo-occular reflexes, and smooth pursuit movements.

The process of spatial navigation is abundant in everyday life, therefore investigating the process of gathering spatial navigation under natural conditions is our goal. While there are some examples of studies in the field of spatial cognition conducted in natural environments (42,43), these studies typically rely on behavioral data, and lack physiological data recorded during the spatial navigation itself. With new mobile eye tracking systems available, more studies investigating spatial cognition under natural conditions were conducted in inside and outside environments. For example, Ohm et al. (2014) investigated the selection process of landmarks in large scale indoor environments via the visual attraction measured with mobile eye tracking (21) and evaluated pedestrian navigation systems based on indoor landmarks (22). Kiefer et al. (2014) investigated self-localization based on participants matching maps to the urban environment they were located at (20). Furthermore, Wenczel et al. (2017) found differences in gaze behavior during incidental and intentional route learning in a natural outdoor environment with intentional learning leading to more fixations on landmarks (23). However, studies conducted with mobile eye tracking systems in natural environments are usually challenged with several disadvantages. In addition to the poor accuracy of mobile eye tracking systems, another major issue comes with identifying frames with areas of interest during the data pre-processing. Often, identification of regions of interests and therefore identifying the relevant fractions of the eye tracking data could only be solved by manual detection in each frame (20,21,23). Furthermore, the natural environment allows only limited control of the experimental conditions, especially regarding variances in light affecting the eye tracker systems (20), and variance in the environment due to other people or traffic (24,25). Depending on the system, eye tracking calibration can be distorted for some distances, therefore limiting the valid distance of gazed objects that can be analyzed (20). Consequently, most eye tracking studies struggle to be conducted under natural conditions and thus reduce the ecological validity of gathered information of viewing behavior under natural conditions. Instead, implementing experimental paradigms in VR allows maximal control of experimental conditions while still providing a more naturalistic environment and multisensory experience. Moreover, our analysis method using hit point clusters and gaze events allow a fast and precise analysis option, therefore providing a solution to the problem of detecting ROIs manually as observed in previous studies. Therefore, new experimental paradigms in VR in combination with the eye tracking analysis proposed in this paper provide a new option to investigate spatial navigation under naturalistic conditions.

The graph-theoretical measures used in this study is crucial, since it leads to a large compression and abstraction of the data. Using this method, a few millions of gaze samples were condensed into a few graphs. In general, graph theory is used in many areas to make complex information of pairwise relations accessible (44–46). This process includes several decisions, which might critically influence the later analysis. First, the visibility of the houses depended on the participant's location in the virtual world, which is not necessarily close to one of the respective houses. Accordingly, the participant might view two houses in direct succession leading to a visual connection, even though the houses themselves might not be visible from each other's location. Second, the gaze graph did not contain any information about the order in which the two successive gaze events took place, consequently all gaze graphs were undirected. Third, the graph contained binary, i.e., unweighted, edges. Consequently, the gaze graphs only contained information about whether a visual connection took place at some point during the experiment, and not how often. Fourth, in case the data contained a cluster of missing data points no information about visual connections during this time was available and no edge could be created between nodes. Applying the above considerations for graph creation, our graphs represented the gaze data in a well-defined and meaningful way.

Our results of the graph-theoretical analysis revealed a small subgroup of houses that seemed to correspond to several characteristics we expected landmarks to have in a gaze graph. Landmark knowledge refers to the knowledge of salient objects that serve as orientation anchors and are memorized when exploring a new environment. (2,4,5). Since the node degree is a common measure to investigate the importance of single nodes in the network (47), we expected landmarks to stand out in visual behavior compared to other houses and therefore show high importance in the gaze graph visible in high values of their node degree centrality. The gaze graph represented the visual connections between houses, therefore, the node degree centrality measured how many different houses were visually connected with the house in question. If a house would serve as a landmark, we would expect participants to often create visual connections between the landmark house and other houses while trying to navigate, especially when they try to relate newly learned houses in respect to the landmark house. The results of our node degree centrality analysis showed a clear difference between the subgroup of the rich club compared to the other houses. Moreover, the average node degree values of the houses exceeded the mean node degree over two times the standard deviation. The rarity of these buildings is strengthened by the high mean hierarchy index. Taking all these findings into account, we decided to call this subgroup of houses gaze-graph-defined landmarks, since they fit the characteristics a landmark would display in a gaze graph.

While it is undisputed that landmarks are important for spatial navigation, the details about their functionality are not yet completely understood. Commonly, it is assumed that landmarks form the basic building blocks of landmark knowledge that is then extended to route and survey knowledge by gradually connecting landmarks with routes and achieving knowledge about the relational information of the landmark locations (2). Others have proposed a more continuous and parallel development of landmark, route and survey knowledge (4,42). Nonetheless, especially the mechanisms of how landmarks relate to mental maps, how landmark knowledge is integrated into survey knowledge or which functions landmarks maintain when a map-like survey knowledge is already available remain unclear. By calculating the rich club coefficient of all gaze graphs, we found a causality between increasing node degree and increasing connectivity between the respective nodes. Gaze graph-defined landmarks were above chance level interconnected to each other. Moreover, the interconnected gaze-graph-defined landmarks of the city seem to form a network, i.e. a rich club, covering a large, mostly central part of the city. This could be a first indication that landmarks are not only used as orientation anchors, but could form an underlying network of orientation anchors that span out the framework of a mental map. Consequently, landmarks could anchor the mental map and thus, serve as an important feature of survey knowledge.

Furthermore, our results revealed that participants spent a large fraction of their time in the areas where the theoretical basis of triangulation with gaze-graph-defined landmarks was given. Specifically, participants spend more than 50% or their experiment time in the locations where at least two of the gaze-graph-defined landmarks were visible. Interestingly, these areas were located at the more central regions of the city. This could be explained by several reasons. On the one hand, participants could prefer city areas in which they could triangulate based on gaze-graph-defined landmarks and consequently spend most of their time in these locations. On the other hand, this could also be related to the size of Seahaven. It was shown that within 90 min total exploration time participants explored all areas of Seahaven, central areas a slightly more often than peripheral areas (48). Furthermore, our results indicate that participants created a network of landmarks. This suggests that the subjects deliberately searched for houses that could serve as landmarks. The selection of 10 gaze-graph-defined landmarks constituted to less than 5% of all available houses located in Seahaven. Consequently, our results strongly indicate that participants were not only looking for landmarks, but also quite strategically chose landmarks as orientation anchors that provided a maximum amount of information on the basis of the rather small subset of less than 5% of all city buildings. Therefore, our results seem to support the notion that landmarks were not only selected based on saliency but seemed to follow a specific strategy maximizing the navigational benefits while minimizing processing expenses.

This interpretation of our results was supported by further characteristics of the gaze-graph-defined landmarks. In a navigational context, the term landmark usually refers to any type of object that is highly visible or easily recognizable in the environment and thus serves as a point of reference (49), although the overall saliency also depends on "a unique property of the trilateral relation between the feature itself, the surrounding environment and the observer's point of view" (50). In addition, landmarks are often differentiated into local and global landmarks. Typically, global landmarks can be seen from far distances and provide a reference for directional and more compass-like orientation, whereas local landmarks might only be visible in a local area and are often located at road crossings (51). Looking at the visual appearance of the gaze-graph-defined landmarks, we found typical characteristics of landmarks including visual saliency due to size, color and location in the city. For example, the house with the highest average node degree across all participants is higher than most surrounding buildings and has a large distinct blue window front that sets it aside from its surrounding (Fig. 6b,7b). Additionally, it is located in the very center of the city making it visible from most parts of the city. The house with the second highest average node degree across all participants stood out due to its size regarding its surface area and location next to the main road that connects the most north and most south part of the city, even though it was only visibly in the south-east part of the city. In general, most of the gaze-graph-defined landmarks were located next to crossings of main roads in the city thus fitting a characteristic of local landmarks (51). Our results revealed that the gaze-graph-defined landmarks also had features of global and local landmarks.

While the application of graph theory has enabled us to use a variety of already established graph-theoretical measures to analyze the gaze graphs and resulted in very promising results, only a small amount of the available graph-theoretical measures was applied during our analysis. For example, the node degree centrality is defined as the sum of connections of each node. Generally speaking, nodes with a lot of connections are likely to be important for most networks (45) which is why the node degree centrality is widely used in graph-theoretical analyses and serves as the basis of our gaze graph analysis (52). According to Sporns (2018), the usage of graph theory in neuroscientific studies has increased in the recent years and the node degree centrality can be a useful measure for network analyses. However, these analyses are mostly based on pairwise dyadic approaches and the full potential of graph theory has not yet been applied.

A variety of graph measures is available for analyzing networks and node importance can be defined in different ways, exceeding the node degree-based analysis. A particularly important measure in social network analyses is the so-called betweenness centrality (53).

The measure counts how many shortest paths cross a particular node that is "between" a lot of nodes. In social network analyses, betweenness centrality is beginning to replace the node degree centrality to explain social network dynamics with respect to the importance of nodes with high node betweenness for attracting and strengthening new links (54). Consequently, the betweenness centrality could be a potential candidate for explaining spatial knowledge acquisition. As mentioned, the centrality measure gains importance within the field of social network analysis and is used to characterize the attraction and strengthening of connections. A person, represented by a node in a social network, would have a high betweenness centrality if the person connects a variety of other persons that themselves do not know the connecting person. By transferring this thought to spatial navigation, a house with high betweenness centrality would connect the views to two buildings that are not viewed in direct succession themselves. Thus, this building serves as an anchor point for these two buildings and forms a part of the route between them. Betweenness centrality could serve as a measure for characterizing the gathered route knowledge. Overall, graph theory offers a variety of different measures and has not yet reached its full potential within neuroscientific research.

Overall, our results establish a new methodology to process eye tracking data in complex 3D environments and identify and assess the function of landmarks in spatial navigation. Applying this methodology provides a new and unique insight into behavioral data of visual attention processes during spatial navigation and opens the door for a novel approach to investigate spatial navigation. To fully unlock the potential of graph theory, we propose additional graph-theoretical measures to investigate gaze graphs in the future. Specifically, we consider the betweenness centrality that could help to understand the formation of spatial knowledge beyond landmark knowledge.

# Methods

## The virtual town of Seahaven

The virtual town of Seahaven was built to investigate spatial learning during free exploration (28,48,55). In total, Seahaven contains 213 houses in varying size and shape. Furthermore, the city was designed as a connected urban space of roughly 216,000 m2 (48). The street structure consisted of winding, small and big roads and overall resembled a European city center. The entire virtual environment reflects natural spatial relations, with one Unity unit corresponding to one meter. The virtual height of the participant was set to 2 meters and was unified for every participant. To provide a better frame rate in Unity, a far clipping plane in a distance of 160 meters was introduced. Consequently, no objects were visible to the participants that were located further than 160m away from current participant location.

## Structure of the experiment

In total, 22 participants performed the complete experiment. This included three sessions within at most 10 days. Each session consisted of 5 parts: (1) a brief introduction to the experiment; (2) examples of the spatial tasks to be completed after the last recording session;

(3) preparation of the VR setup, including the adjustment of VR-headset with eye tracker, the calibration and validation of the eye tracker, and in the 1st session a movement training on the virtual island; (4) the main experimental phase, i.e. exploring the virtual town for 30 minutes while movements and eye tracking data were recorded; (5) three spatial tasks performed outside of the VR, which are covered in detail in (48). These data are, however, are outside the scope of the present paper and not further covered here.

### Laboratory setup

Participants wore a head mounted HTC Vive virtual reality headset and were seated on a swivel chair. To prevent limitations on rotations as well as removing the tactile directional feedback due to hanging cables, a vertical cable solution was implemented. Participants moved using the HTC controller at walking speed. To decrease the risk of motion sickness, participants were instructed to only walk forward with the controller. If they wanted to switch directions or turn, they were instructed to stop walking and then rotate their entire body with the chair in the desired direction.

### Eye tracking

A pupil labs eye tracker was directly integrated in the HTC Vive headset (refresh rate 120Hz, gaze accuracy 1,0°, gaze precision 0.08°, visual field 110°) (28). Both calibration and validation were executed in the virtual reality while the participants where still on a separate training island. For each subject, a 17-point calibration and a 9-point validation were conducted until the validation error was below 2° (on average 1.53°). During the experiment, a 1-point validation was performed every 5 minutes ensuring the correct tracking precision. If the error exceeded 2°, a complete 17-point calibration and 9-point validation was performed until the error was below the original threshold of 2° (mean 3.14°, median 2.11° before and mean 1.88°, median 1.28° after a new calibration). 55% of all 1-point validations had a validation error above 2° and had to be recalibrated, hence, highlighting the importance of regular validation control. When the calibration and validation process was performed during the exploration of Seahaven, the display of the virtual city disappeared until the validation was completed.

## Acknowledgments

We gratefully thank Viviane Clay and Nicolas Kuske for their assistance and support. In addition, we gratefully acknowledge the support by the GK Computational Cognition and the Open Access Publishing Fund of Osnabrück University.

## Conflict of Interest Statement

The authors have declared that no competing interests exist.

# Funding

The project was financed by the funds of the research training group Computational Cognition provided by the Deutsche Forschungsgemeinschaft (DFG), Germany (GRK2340).

# Data Availability

All data files the described pre-processing and analysis are based on, are available at the Center for Open Science https://osf.io/aurjk/, DOI 10.17605/OSF.IO/AURJK. All pre-processing, visualization and analysis scripts supporting this publication are available on our Github repository including an extensive documentation to allow full reproducibility: https://github.com/JasminLWalter/FindingLandmarks_a_publication_repository

# References

1. Kelly JW, McNamara TP. Spatial Memory and Spatial Orientation. In: Freksa C, Newcombe NS, Gärdenfors P, Wölfl S, editors. Spatial Cognition VI Learning, Reasoning, and Talking about Space. Berlin, Heidelberg: Springer Berlin Heidelberg; p. 22–38.

2. Siegel AW, White SH. The Development of Spatial Representations of Large-Scale Environments. Adv Child Dev Behav. 10:9–55. doi:10.1016/S0065-2407(08)60007-5

3. Lynch K. The image of the city. Vol. 11, MIT press.

4. Montello DR. A new framework for understanding the acquisition of spatial knowledge in large-scale environments. Spat temporal Reason Geogr Inf Syst. :143–54.

5. Klippel A, Winter S. Structural Salience of Landmarks for Route Directions. In: Cohn AG, Mark DM, editors. Spatial Information Theory. Berlin, Heidelberg: Springer Berlin Heidelberg; p. 347–62. doi:10.1007/11556114_22

6. Meilinger T, Frankenstein J, Bülthoff HH. Learning to navigate: Experience versus maps. Cognition. 129(1):24–30. doi:10.1016/J.COGNITION.2013.05.013

7. O'Keefe J, Nadel L. The hippocampus as a cognitive map. Oxford Univ Press.

8. Sholl MJ. Cognitive Maps as Orienting Schemata. J Exp Psychol Learn Mem Cogn. 13(4):615–28. doi:10.1037/0278-7393.13.4.615

9. Thorndyke PW, Hayes-Roth B. Differences in spatial knowledge acquired from maps and navigation. Cogn Psychol. 14(4):560–89. doi:10.1016/0010-0285(82)90019-6

10. Morris RGM. Spatial localization does not require the presence of local cues. Learn Motiv. 12(2):239–60. doi:10.1016/0023-9690(81)90020-5

11. McDonald RJ, White NM. Parallel information processing in the water maze: Evidence for independent memory systems involving dorsal striatum and hippocampus. Behav Neural Biol. 61(3):260–70. doi:10.1016/S0163-1047(05)80009-3

12. Packard MG, McGaugh JL. Inactivation of hippocampus or caudate nucleus with lidocaine differentially affects expression of place and response learning. Neurobiol Learn Mem. 65(1):65–72. doi:10.1006/NLME.1996.0007

13. Mueller SC, Jackson CPT, Skelton RW. Sex differences in a virtual water maze: An eye tracking and pupillometry study. Behav Brain Res. 193(2):209–15. doi:10.1016/j.bbr.2008.05.017

14. Iaria G, Petrides M, Dagher A, Pike B, Bohbot VD. Cognitive strategies dependent on the hippocampus and caudate nucleus in human navigation: Variability and change with practice. J Neurosci. 23(13):5945–52. doi:10.1523/jneurosci.23-13-05945.2003

15. Hamilton DA, Driscoll I, Sutherland RJ. Human place learning in a virtual Morris water task: Some important constraints on the flexibility of place navigation. Behav Brain Res. 129(1–2):159–70. doi:10.1016/S0166-4328(01)00343-6

16. Andersen NE, Dahmani L, Konishi K, Bohbot VD. Eye tracking, strategies, and sex differences in virtual navigation. Neurobiol Learn Mem. 97(1):81–9. doi:10.1016/j.nlm.2011.09.007

17. Newman EL, Caplan JB, Kirschen MP, Korolev IO, Sekuler R, Kahana MJ. Learning your way around town: How virtual taxicab drivers learn to use both layout and landmark information. Cognition. 104:231–53. doi:10.1016/j.cognition.2006.05.013

18. Farran EK, Formby S, Daniyal F, Holmes T, Van Herwegen J. Route-learning strategies in typical and atypical development; eye tracking reveals atypical landmark selection in Williams syndrome. J Intellect Disabil Res. 60(10):933–44. doi:10.1111/jir.12331

19. Hejtmánek L, Oravcová I, Motýl J, Horáček J, Fajnerová I. Spatial knowledge impairment after GPS guided navigation: Eye-tracking study in a virtual town. Int J Hum Comput Stud. 116(May 2017):15–24. doi:10.1016/j.ijhcs.2018.04.006

20. Kiefer P, Giannopoulos I, Raubal M. Where am i? Investigating map matching during self-localization with mobile eye tracking in an urban environment. Trans GIS. 18(5):660–86. doi:10.1111/tgis.12067

21. Ohm C, Müller M, Ludwig B, Bienk S. Where is the Landmark? Eye tracking studies in large-scale indoor environments. CEUR Workshop Proc. 1241:47–51.

22. Ohm C, Müller M, Ludwig B. Evaluating indoor pedestrian navigation interfaces using mobile eye tracking. Spat Cogn Comput. 17(1–2):89–120. doi:10.1080/13875868.2016.1219913

23. Wenczel F, Hepperle L, von Stülpnagel R. Gaze behavior during incidental and intentional navigation in an outdoor environment. Spat Cogn Comput. 17(1–2):121–42. doi:10.1080/13875868.2016.1226838

24. Kiefer P, Straub F, Raubal M. Location-Aware Mobile Eye Tracking for the Explanation of Wayfinding Behavior. In: Gensel J, Josselin D, Vandenbroucke D, editors. Proceedings of the AGILE'2012 International Conference on Geographic Information Science.

25. Kiefer P, Straub F, Raubal M. Towards location-aware mobile eye tracking. Eye Track Res Appl Symp. :313–6. doi:10.1145/2168556.2168624

26. Kim JY, Kim MJ. Exploring visual perceptions of spatial information for wayfinding in virtual reality environments. Appl Sci. 10(10):1–15. doi:10.3390/app10103461

27. Meilinger T, Frankenstein J, Watanabe K, Bülthoff HH, Hölscher C. Reference frames in learning from maps and navigation. Psychol Res. 79(6):1000–8. doi:10.1007/s00426-014-0629-6

28. Clay V, König P, König SU. Eye tracking in virtual reality. J Eye Mov Res. 12(1). doi:10.16910/jemr.12.1.3

29. Duchowski AT. Eye Tracking Methodology [Internet]. Third Edit. Eye Tracking Methodology. Springer; doi:10.1007/978-3-319-57883-5

30. Breitmeyer BG, Ganz L. Implications of sustained and transient channels for theories of visual pattern masking, saccadic suppression, and information processing. Psychol Rev. 83(1):1–36. doi:10.1037/0033-295X.83.1.1

31. Reichle ED, Pollatsek A, Fisher DL, Rayner K. Toward a Model of Eye Movement Control in Reading. Psychol Rev. 105(1):125–57. doi:10.1037/0033-295X.105.1.125

32. Nuthmann A, Smith TJ, Engbert R, Henderson JM. CRISP: A Computational Model of Fixation Durations in Scene Viewing. Psychol Rev. 117(2):382–405. doi:10.1037/a0018924

33. Fiedler M. Algebraic connectivity of graphs. Czechoslov Math J. 23(2):298–305. doi:10.21136/CMJ.1973.101168

34. Fiedler M. Laplacian of graphs and algebraic connectivity. Vol. 25, Banach Center Publications. p. 57–70. doi:10.4064/-25-1-57-70

35. Rodrigue J-P. The Geography of Transport Systems. 5th ed. Routledge; 480 p.

36. Zhou S, Mondragón RJ. The rich-club phenomenon in the internet topology. IEEE Commun Lett. 8(3):180–2. doi:10.1109/LCOMM.2004.823426

37. Van Den Heuvel MP, Sporns O. Rich-club organization of the human connectome. J Neurosci. 31(44):15775–86. doi:10.1523/JNEUROSCI.3539-11.2011

38. Colombo D, Serino S, Tuena C, Pedroli E, Dakanalis A, Cipresso P, et al. Egocentric and allocentric spatial reference frames in aging: A systematic review [Internet]. Vol. 80, Neuroscience and Biobehavioral Reviews. Pergamon; p. 605–21. doi:10.1016/j.neubiorev.2017.07.012

39. Chrastil ER, Warren WH. Active and passive contributions to spatial learning. Psychon Bull Rev. 19(1):1–23. doi:10.3758/s13423-011-0182-x

40. Ekstrom AD. Why vision is important to how we navigate. Hippocampus. 25(6):731–5. doi:10.1002/hipo.22449

41. Ehinger B V., Groß K, Ibs I, König P. A new comprehensive eye-tracking test battery concurrently evaluating the Pupil Labs glasses and the EyeLink 1000. PeerJ. 2019(7):1–43. doi:10.7717/peerj.7086

42. Ishikawa T, Montello DR. Spatial knowledge acquisition from direct experience in the environment: Individual diVerences in the development of metric knowledge and the integration of separately learned placesIshikawa ¤, T., & Montello, D. R. (2006). Spatial knowledge acquisition f. Cogn Psychol. 52(2):93–129. doi:10.1016/j.cogpsych.2005.08.003

43. Ishikawa T, Fujiwara H, Imai O, Okabe A. Wayfinding with a GPS-based mobile navigation system: A comparison with maps and direct experience. J Environ Psychol. 28(1):74–82. doi:10.1016/J.JENVP.2007.09.002

44. Mondal B, De K. Overview Applications of Graph Theory in Real Field. Int J Sci Res Comput Sci Eng Inf Technol. 2(5):751–9.

45. Sporns O. Graph theory methods: applications in brain networks. Dialogues Clin Neurosci. 20(2):111–20. doi:10.31887/DCNS.2018.20.2/osporns

46. Singh RP. Application of Graph Theory in Computer Science and Engineering. Int J Comput Appl. 104(1):10–3. doi:10.5120/18165-9025

47. Srinivasan S, Hyman JD, O'Malley D, Karra S, Viswanathan HS, Srinivasan G. Chapter Three - Machine learning techniques for fractured media. Moseley B, Krischer L, editors. Vol. 61, Advances in Geophysics. Elsevier; 109–150 p. doi:10.1016/BS.AGPH.2020.08.001

48. König SU, Keshava A, Clay V, Rittershofer K, Kuske N, König P. Embodied Spatial Knowledge Acquisition in Immersive Virtual Reality: Comparison to Map Exploration. Front Virtual Real. 2. doi:10.3389/frvir.2021.625548

49. Presson CC, Montello DR. Points of reference in spatial cognition: Stalking the elusive landmark*. Br J Dev Psychol. 6(4):378–81. doi:10.1111/j.2044-835x.1988.tb01113.x

50. Caduff D, Timpf S. On the assessment of landmark salience for human navigation. Cogn Process. 9(4):249–67. doi:10.1007/s10339-007-0199-2

51. Steck SD, Mallot HA. The role of global and local landmarks in virtual environment navigation. Presence Teleoperators Virtual Environ. 9(1):69–83. doi:10.1162/105474600566628

52. Farahani F V., Karwowski W, Lighthall NR. Application of graph theory for identifying connectivity patterns in human brain networks: A systematic review. Front Neurosci. 13(JUN):1–27. doi:10.3389/fnins.2019.00585

53. Diestel R. Graph Theory [Internet]. Fifth Edit. Axler S, Ribet K, editors. Graduate Texts in Mathematics. Berlin, Heidelberg: Springer Berlin Heidelberg; 49–64 p. (Graduate Texts in Mathematics; vol. 173). doi:10.1007/978-3-662-53622-3

54. Topirceanu A, Udrescu M, Marculescu R. Weighted Betweenness Preferential Attachment: A New Mechanism Explaining Social Network Formation and Evolution. Sci Rep. 8(1):1–14. doi:10.1038/s41598-018-29224-w

55. König SU, Clay V, Nolte D, Duesberg L, Kuske N, König P. Learning of Spatial Properties of a Large-Scale Virtual City With an Interactive Map. Front Hum Neurosci. 13:240. doi:10.3389/FNHUM.2019.00240