Mapping Individual Differences in the Topological Landscape of Naturalistic Brain Dynamics

Junxing Xian[#1,2], Yini He[#2,1], Yan Yan[3], Xiaohan Tian[1], Yingjie Peng[2], Jing Lou[1], Xiya Liu[2], Qi Wang[4,5], Tian Gao[2], Qi Wang[6], Yuqing Sun[1], Puze Li[2], Yue Wang[7], Shangzheng Huang[2], Kaixin Li[2], Ke Hu[4,5], Chaoyue Ding[4,5], Dazheng Li[2], Meng Wang[1], Bing Liu*[1], Ang Li*[2]

1. State Key Laboratory of Cognitive Neuroscience and Learning, Beijing Normal University; Beijing 100091, China.

2. State Key Lab of Brain and Cognitive Science, Institute of Biophysics, Chinese Academy of Sciences, Beijing 100101, China.

3. Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences, Shenzhen, Guangdong, 518055, China

4. Brainnetome Center, Institute of Automation, Chinese Academy of Sciences, Beijing, 100190, China.

5. School of Artificial Intelligence, University of Chinese Academy of Sciences, Beijing, 100190, China.

6. College of Science, China Agricultural University, Beijing, 100083, China.

7. Medical School of Tianjin University, Tianjin University, Tianjin 300072, China.

*Corresponding author: al@ibp.ac.cn, bing.liu@bnu.edu.cn.

Abstract

Naturalistic stimuli elicit rich subjective experiences through adaptive neural coordination. However, how inherent behavioral traits shape individual neural dynamics in naturalistic settings remains unclear. Here, we introduce a computational framework, STIM, to systematically capture individual differences in brain dynamics while watching diverse movie stimuli. By leveraging Topological Data Analysis, STIM generates a robust group-level dynamical landscape of brain latent states, mapping individual-specific divergence into global topology and local geometry. Applying STIM to large-sample movie fMRI datasets, we found that inter-individual variation in global topology exhibits a center-periphery gradient in the landscape. This gradient significantly explains individual fluid intelligence from a dual perspective, highlighting the importance of both adaptability and diversity of neural dynamics. At the fine-grained narrative level, individual local geometry attributes are associated with context-specific psychological traits beyond cognition. Furthermore, STIM reveals how the dynamical landscape evolves across neurodevelopment and exhibits abnormalities in psychiatric disorders such as autism. In summary, the STIM framework has the potential to transform rich naturalistic stimuli with brain recording into neural ‘probes’ to measure individual differences in cognition and mental health.

Introduction

Understanding how the brain's functional dynamics relate to behavioral traits and mental disorders is fundamental for neuroscience. Task paradigms, usually highly controlled, aim to activate specific brain circuits[1,2] associated with focused cognitive or emotional functions, and the between-subject variability in brain response patterns could predict task performance and related phenotypic differences[3]. Conversely, the resting-state paradigm uncovers the brain's spontaneous functional organization, providing a broader perspective on interindividual variability in both behavior and brain activation[4]. However, findings from tasks may at times be fragmentary and lacking cohesion, sacrificing generalizability for specificity; whereas the task-free condition may reflect endogenous mental states that are unconstrained and unmeasurable, thus being susceptible to interference from the states during scanning. The emergent naturalistic paradigm, leveraging stimuli like movies or stories, promises a holistic framework for studying individual differences[5]. It merges the specificity of the task paradigm with the broad applicability of resting-state design, eliciting comprehensive cognitive processes and genuine emotional reactions, enriching the stimuli's ecological validity to probe the brain's multifaceted functional system in a high-throughput manner[6].

However, current computational models, which explore individual differences under naturalistic stimulus paradigms, are inadequate in depicting how the brain continuously integrates highly unstructured dynamic inputs at a system-wide level. Consider movie stimuli as an example input: their most distinguishing characteristic is their ability to evoke continuous and irregular shifts that encompass cognitive, emotional, and attentional aspects[7], thereby eliciting participants' real-life-like, integrated subjective experiences, which can be highly specific to individuals, even when the audiovisual stimuli are entirely the same (No two people see the same rainbow). Current computational models, for instance, inter-subject correlation (ISC) and other similar approaches, measure the synchrony between different subjects over the entire duration or a sliding window of time[8–13]. These methods, due to their lack of a flexible timescale, lead to the blurring of brain state transitions. On the other hand, state-based approaches such as the Hidden Markov Model (HMM)[14,15] and Co-activation Patterns (CAPs)[16] are constrained by the number of brain states they can account for (typically, $n < 10$), making them better suited for resting states and presenting challenges in accurately measuring the brain representations of rich stimuli. From a dynamical system perspective, recent studies in different species have suggested that large-scale coordinated neural activity could map to continuous, latent space trajectories (or flows, manifolds), supporting ongoing behavior and cognitive function at the system level[17–20]. Stimuli that are task-oriented, or the global states of the brain, seem to act as anchors or attractor basins that guide low-

dimensional latent dynamics. Here, we hypothesize: by viewing the brain as a dynamical system, the fluctuating narratives of movies can shape the dynamic landscape of brain latent trajectories. These trajectories, very likely stable at the group level, may serve as a neural signature of 'consensus'. 'Divergence' in individual-specific experiences will manifest as individual-specific latent dynamics, deviating either globally or locally, largely depending on the inherent features of the individuals.

To test this hypothesis, here we introduce topological data analysis (TDA) into naturalistic fMRI and develop a new computation framework STIM characterizing individual differences in naturalistic dynamic landscape, for the first time, within a low-dimensional state space. Specifically, the core of STIM is the TDA-based Mapper, a mathematical approach that models high-dimensional datasets as a shape graph to visualize and analyze the topological and geometric information therein[21]. Mapper has been widely used in the study of dynamic processes, including neural activities[22], disease progression[23], transcriptional programs[24], viral evolution[25], and more. We posit that Mapper is particularly well-suited for movie fMRI data for several reasons: i) Movie plots are unstructured and comprise numerous rapid narrative transitions, whereas manual annotation is time-consuming, labor-intensive, and subjective. Mapper has been shown to sensitively capture task-oriented transitions up to ~4-9s faster than traditional methods, without requiring any annotation or assumptions about the analyzed data[22]. ii) Movie plots carry explicit contextual semantics and constitute a story or work with a global structure. Mapper can integrate high-dimensional time series into a combinatorial object while preserving the topological structure of space and the dynamics of time, capable of describing topological properties such as the global shape[21]. iii) Movie plots are information-rich and capable of inducing diversity in higher-order processing among viewers with the same input[26], where Mapper can maximally retain the full spatiotemporal dynamic information at the individual level[22].

Leveraging the strengths of Mapper, the STIM framework effectively aligns individuals to form a group-level dynamic landscape. It quantifies the divergence in individual trajectories at two different scales: overall shape-like global topology and cluster-like local geometry. We apply the STIM framework to two relatively large sample datasets, 170 subjects from the Human Connectome Project (HCP) and 970 subjects from the Healthy Brain Network (HBN) datasets, to explore how brain latent dynamics during movie watching correspond to individual differences across different domains.

Results

Mapper captures rapid and complex dynamics during movie-watching in the individual brain.

In multi-task and resting conditions, Mapper can construct the individual landscape and capture transitions between tasks as well as within spontaneous brain states[27]. Extending previous studies, we expect that the TDA-based Mapper can effectively characterize the topology and dynamics of brain configurations related to complex movie stimuli at the individual level, without necessitating any movie annotations.

Accordingly, we analysed high-quality 7T fMRI data during movie-viewing from the HCP (n=170, 4 runs for each participant), consisting of concatenated movies covering various themes such as science-fiction (sci-fi), drama, and nature scenes (details in Supplementary Tab. 1). Using the classic Mapper approach, we generated a topological landscape at single subject level[28] (Fig. 1a, See Methods). Classic Mapper aims to capture topological and geometric information, serving as a generalized notion of coordinatization for high-dimensional data (i.e., brain activity across fMRI scanning). The Mapper begins with a filter function f to generate a low-dimensional embedding, typically employing nonlinear dimensionality reduction techniques (such as UMAP[29], t-SNE[30]). This approach is crucial to preserve the 'nearness' of data, a desirable property in manifold learning that helps to mitigate the effects of the 'curse of dimensionality'. The low-dimensional filtered data were then encapsulated into overlapping bins, generating a rough image of the data's topology. Within each bin, partial clustering was applied to refine the scale, with each cluster forming a node and the overlaps between nodes forming edges. This process produces a multiresolution image of the data's topology and is less sensitive to the choice of metric. Through this processing, the high-dimensional brain activity data are condensed into a shape graph[21].

Fig.1b displays the topological landscape of global brain dynamics for a representative individual (see more people in Supplementary Fig.1), that is, a Mapper-generated shape graph. This graph intuitively illustrates nodes as clusters of whole-brain volumes with high similarity (with size indicating the number of included time volumes, from 4 to 48 seconds for each node), and nodes sharing overlapping volumes are connected. A distinctive feature of the Mapper method is its embedding of temporal brain activations into the topological structure of the graph, enabling a dynamic, global perspective analysis and visualization of high-dimensional fMRI data. A video version of Fig.1b, available in Supplementary Mov.1, concurrently showcases the dynamic trajectory of the shape graph and the corresponding movie content being watched. The color annotations in Fig.1b represent different movie sources. Interestingly, movies viewed in different runs can form a cohesive topological landscape, and movies with similar themes tend to connect more, such as dramas (*The Social*

Network and *Ocean's 11*) and nature scenes (*Off the Shelf* and *Northwest Passage*), while a theme-rich movie (*Mrs. Meyer's Clean Day*) is dispersed throughout the graph. This suggests that merging multiple movies can establish a global, rich dynamic landscape.

Moreover, Mapper can provide an interactive, flexible visualization to interpret how the brain traverses its dynamical landscape during movie-watching, by annotating the shape graph with different types of information: i) When annotated by time order, it can display the flow and rapidly changing dynamics related to the viewing context (Fig.1c and Supplementary Mov.1); ii) When annotated by focused semantic topics, such as volumes with 'talk' or without, emerging clustering in the individual shape graph becomes visible (Fig.1d); iii) When annotated by network activation, the pie-chart notation of nodes represents the weight of functional networks, revealing interested brain configuration patterns[27] (Fig.1e).

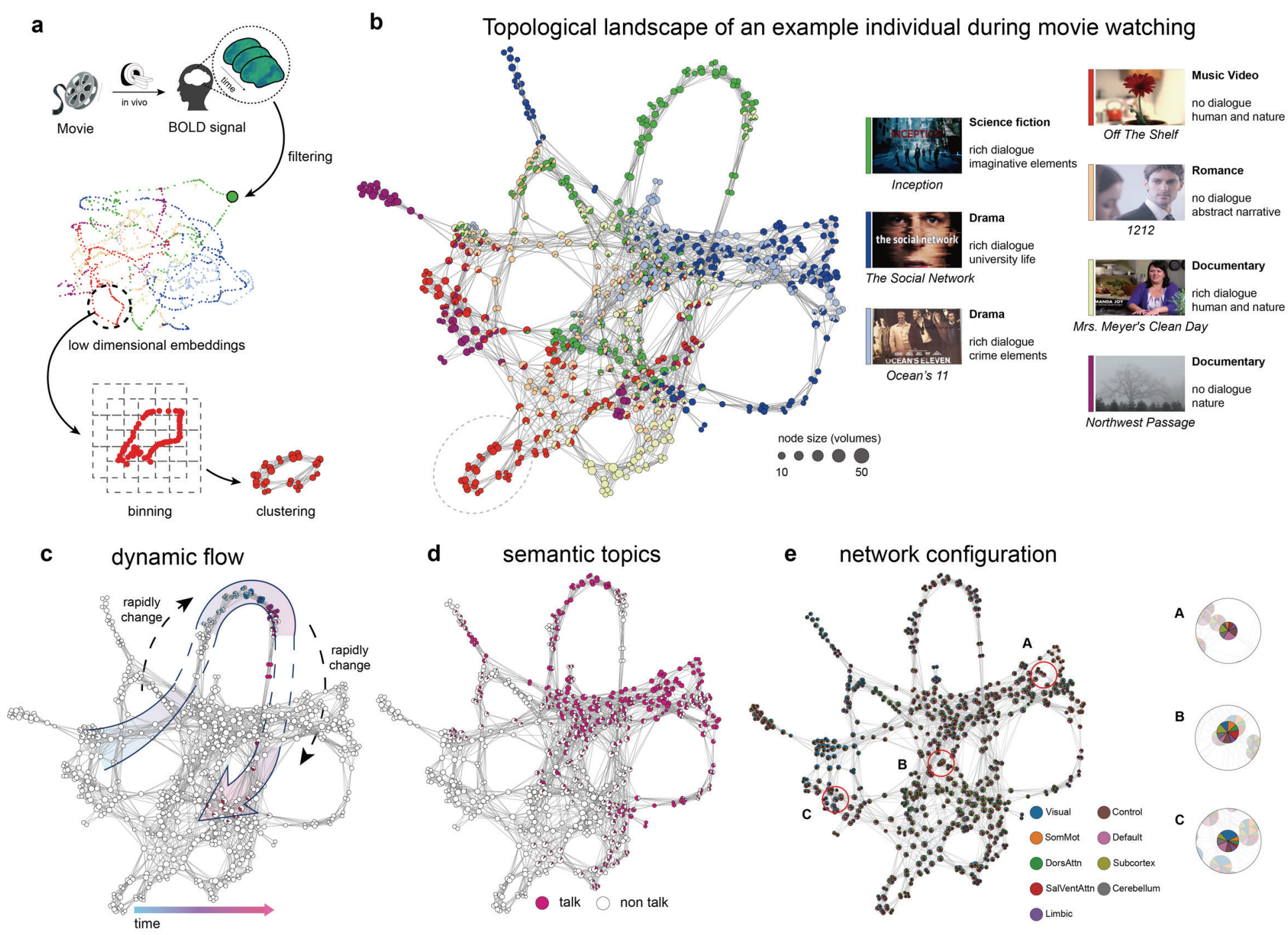


Fig 1 Mapper approach captures the individual dynamic topological landscape. **(a)** A conventional Mapper pipeline for movie-viewing fMRI analysis is outlined as follows: First, the high-dimensional neuroimaging data is embedded into a low-dimensional space using a nonlinear filter function f, here we chose Uniform Manifold Approximation and Projection (UMAP). Second, the low-dimensional embedding is mapped (via binning and clustering) into an overlapping grid, generating a shape graph, where the nodes of the shape graph are determined by bins, and the edges are determined by the overlap between bins. Within the shape graph, closely connected nodes

correspond to highly similar whole-brain time frames. Lastly, annotations of interest are added to the shape graph at the TR scale (for example, time flow) or the nodal scale (for example, node centrality) to further research the topological relationships among different features and to facilitate their visualization. **(b)** The HCP dataset includes various categories of movie stimuli, and 7 of 18 representative clips were used as examples, including sci-fi (*Inception*), drama (*Ocean's Eleven, The Social Network*), music video (*Off the Shelf*), romance (1212), documentary (*Mrs. Meyer's Clean Day*), and nature scene (*Northwest Passage*) (from left to right). **(c, d, e)** Three kinds of mapper independent information were annotated for illustration, including time order (70 volumes), 'talk' semantic label, activation of nine brain network.

STIM framework to capture individual divergence in global topology and local geometry.

While the Mapper approach effectively delineates individual dynamical landscapes, significant challenges persist in characterizing the neural signature of consensus in populations and the divergence specific to individuals. Also, previous studies focused on non-event-related topological descriptors failed to adequately accommodate analysis of specific movie content. Therefore, we introduce the STIM (Synchronized Topological Individual Mapper) framework, aiming to construct a group-level dynamical map of consensus as the reference and quantify individual-specific divergence both in the global topology, corresponding to the overall viewing experience, and in local geometry, corresponding to the experience of specific plot elements. STIM presupposes that individual-specific low-dimensional, latent dynamics, encoded by the Mapper shape graph, reflect subjective experiences during movie viewing, wherein the same stimuli evoke consensus or divergence across viewers. Thus, STIM was developed to align individuals while preserving and quantifying individual-specific topological features.

For a more intuitive demonstration, we chose the segment from the movie *Inception* (Fig. 2a, duration: 3 minutes and 46 seconds) as an example, and manually annotated two scene transitions, both of which feature a significant change, namely a 'jump' from dialogue to sci-fi content. Fig. 2b shows shape graphs of three participants watching *Inception*, with annotations of movie plots (light green and light purple represent the first dialogue and sci-fi, connected by jump-A; deep green and deep purple represent the second dialogue and sci-fi, connected by jump-B). In the low-dimensional topological space, two typical individuals (subject-1 and subject-2 in Fig. 2b) were observed to have dynamical consensus patterns: integrating into two different clusters associated with dialogue and sci-fi, respectively, and exhibited an 'expected' dynamic transition from the dialogue topological region to the sci-fi topological region. Interestingly, some individuals exhibited 'unexpected' transitions towards two distinct topological regions, potentially indicative of their divergence in movie-viewing experiences (such as subject-3 in Fig.2b). The shared and specific features were plausibly observed in individual shape graphs (Fig.2b), however, the shape graphs were independent and thus incomparable. To construct a shared topological space, we concatenated two individuals' data to provide a higher-dimensional representation space and built a combinatorial shape graph of subject-1 and subject-3 (Fig. 2c left), which can represent the synchrony and idiosynchrony: their dynamics cluster at the two dialogue and the second sci-fi contents; however, subject-3 deviated to another topological region while watching the first sci-fi content, 'city folding' scene after jump-A. A time-flow arrow illustrates the temporal dynamics of the two subjects (A full video version is available in Supplementary Mov.2). Therefore, an important procedure in STIM is to concatenate the high-dimensional brain activities across subjects to align brain dynamics and generate a shared low-dimensional topological space using Mapper (see Fig. 2c right). Furthermore, the shared

shape graph can be converted into an adjacency matrix (i.e., temporal connectivity matrix, TCM), in which elements represent the connectivity strength between the two volumes by shared nodes. By calculating the TCM of the combinatorial shape graph, we can quantify the topological similarity between any volumes from different individuals (see Methods for details).

The STIM framework can effectively analyze large-sample movie-viewing data. To characterize the canonical dynamics as group-level consensus for HCP, we concatenated the volumes of all movies and all individuals (2804 volumes×170 subjects, 476,680 volumes) to create a panoramic view of HCP movie-viewing dynamics. We then fed this data into STIM to obtain each subject's TCM relative to the group reference. The group reference was generated by averaging the embeddings across individuals in low-dimensional space (Fig. 2d). We then measured the divergence across individuals by decomposing the individual-group TCM into two distinct scales (Fig. 2d, see Methods):

i) *Global topology similarity*: This metric quantifies the overall shape similarity of topological landscapes cut across movie contents. As an example, Subject-3 might not only deviate in the first sci-fi scene of *Inception* but also exhibit lower similarity with the group reference across various movie scenes. This indicates a lower global topology similarity, reflecting broader deviation in brain dynamics under different cinematic contexts. Global topology similarity is defined as the diagonal value of the TCM (n volumes = 2804 for all HCP contents). STIM filtered low-dimensional embeddings (or trajectories) showed a more consistent population-shared shape of global topology than the original time series (Fig. 2e). Furthermore, trajectories of latent dynamics in the Mapper pipeline showed significantly higher consistency than top ROIs within the visual network (Supplementary Fig. 2), supporting that the latent dynamics of consensus reflects a global integration of the brain during movie-watching beyond merely sensory input. Notably, the individual differences in global topology similarity remained stable even with a small reference sample. For example, when randomly selecting 15 subjects as the group reference, the average Pearson correlation across iterations was 0.98 (Fig. 2f).

ii) *Local geometry similarity*: This metric quantifies the local similarity elicited by specific movie contents. For example, Subject-1 and Subject-3 showed lower similarity in 'city folding' content, while forming a cluster-like structure in the second sci-fi scene (Fig. 2c, 'mirror corridor' in Fig. 2a). Local geometry similarity is defined as the weighted score of cliques detected by a change point detection algorithm[31] (n clusters = 161 across all HCP movie contents, see Methods). To ensure that the algorithm can capture meaningful movie transitions, we compared these topological cliques to the manual segmentations based on semantic labels

manually annotated by the HCP dataset (see Methods). We found that the discrepancy between topological and semantic cliques was significantly smaller compared to a null model ($P < 0.0001$, Fig. 2g), indicating that local topological similarity can capture fine-grained, unstructured transitions in movie plots.

We propose ***the STIM framework*** to effectively align different individuals while preserving individual-specific dynamic topology

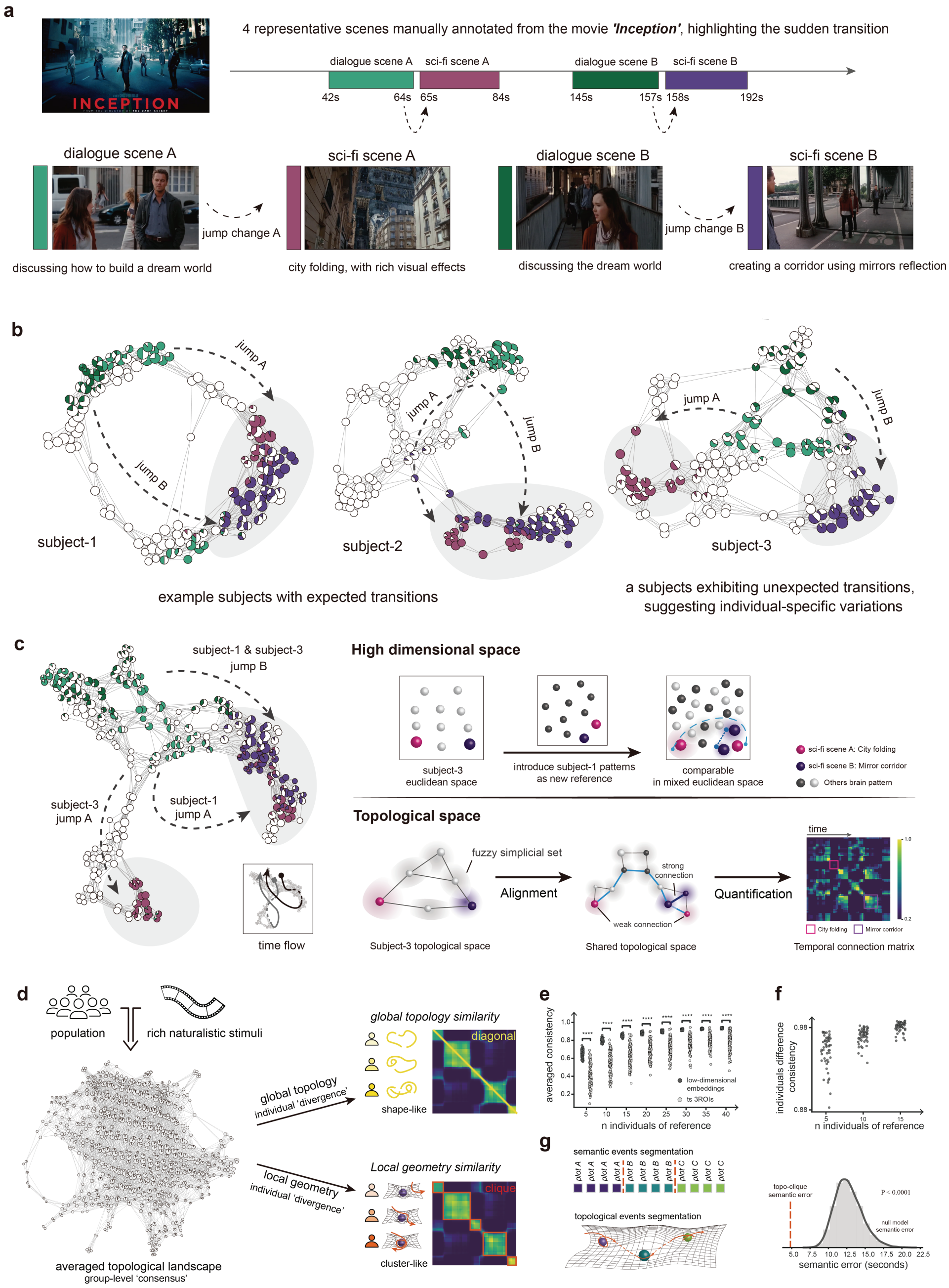


Fig 2 The STIM (Synchronized Topological Individualized Mapper) framework. (a) We manually labeled two representative segments from Inception, each with a continuous sequence, featuring dialogue in the first half and

science fiction scenes in the second half. **(b)** We annotated four parts in the individual shape graphs of the three subjects chosen as examples. According to our assumption, brain dynamics are clustered during dialogue scenes and science fiction scenes at different times. Subject-1 and Subject-2, conforming to the assumption, respectively formed different clusters, while Subject-3 exhibited an "unexpected" transition. **(c)** Aligned mapper approach. The left panel shows the shape graph of alignment for Subject-1 and Subject-3, with annotations of the four scenes. The right panel illustrates the motivation for alignment: introducing a new subject in high-dimensional Euclidean space to provide a comparable combination of TRs. In the topological space, the distances between volumes from different individuals are rescaled through a fuzzy simplicial set (in the context of UMAP). Subsequently, through the filtering and binning steps, a combinatorial shape graph (aligned shape graph) is generated. Finally, the strength between volumes is quantified by TCM. **(d)** Two scales of topological similarity. Global topological similarity is defined as diagonal of subject-group interact TCM (yellow line), represents the overall shape similarity. Local topological similarity is defined as weighted sum of elements in cluster (orange box), represents deviation in specific movie content. **(e)** Group consistency across different sample sizes between mapper and time series: the scatter plot displays the Pearson correlation between two unrelated groups with increasing sample sizes, ranging from 5 to 40 subjects, randomly selected. For a fair comparison, each trial randomly selects time series from 3 ROIs. ****: P<0.0001 **(f)** Pearson correlation between individual global topological similarity using a small sample as reference and the same using the entire group as reference. **(g)** An illustration of the ecological validity of topologically self-organized cliques: The dashed line indicates the error between the topological events segmentation and the semantic events segmentation, accompanied by a gray null distribution, repeated 10,000 times with 161 events segmentations randomly divided in each iteration.

The principal gradient of individual differences in the topological landscape explains cognition from a dual perspective.

Global topology similarity can capture the individual divergence in the shape of the dynamical landscape. We next explore how this divergence in global topology differs between individuals and over time by integrating multi-level information, including topological attributes within the landscape, synchronous modes within the population, brain network configuration, movie contents, and behavioral domains. We performed a principal component analysis (PCA) on the global topological similarity matrix (170 subjects×2804 volumes) across time.

Fig.3a highlights the top 5% positive and negative volumes of the weights in the first principal component (PC1) from the group shape graph, denoted as +VOLs and -VOLs (140 volumes), respectively. At a global level, -VOLs and +VOLs exhibit a trend from the graph's center to its periphery. Further, employing betweenness centrality to measure topological attributes[32], we discovered that + VOLs exhibited significantly lower betweenness centrality compared to - VOLs (t=-13.9, $P < 0.0001$, Fig.3a).

Besides topological attributes, we examined how the + VOLs and - VOLs systematically differed in group synchronous modes. We observed a significant divergence in the dynamics of synchrony between +VOLs and -VOLs (t=18.1, P<0.0001, Fig.3b). Specifically, +VOLs exhibited an attractor-like mode, where subjects tend to converge (stronger consensus); in contrast, -VOLs displayed an unconstrained mode, where subjects are

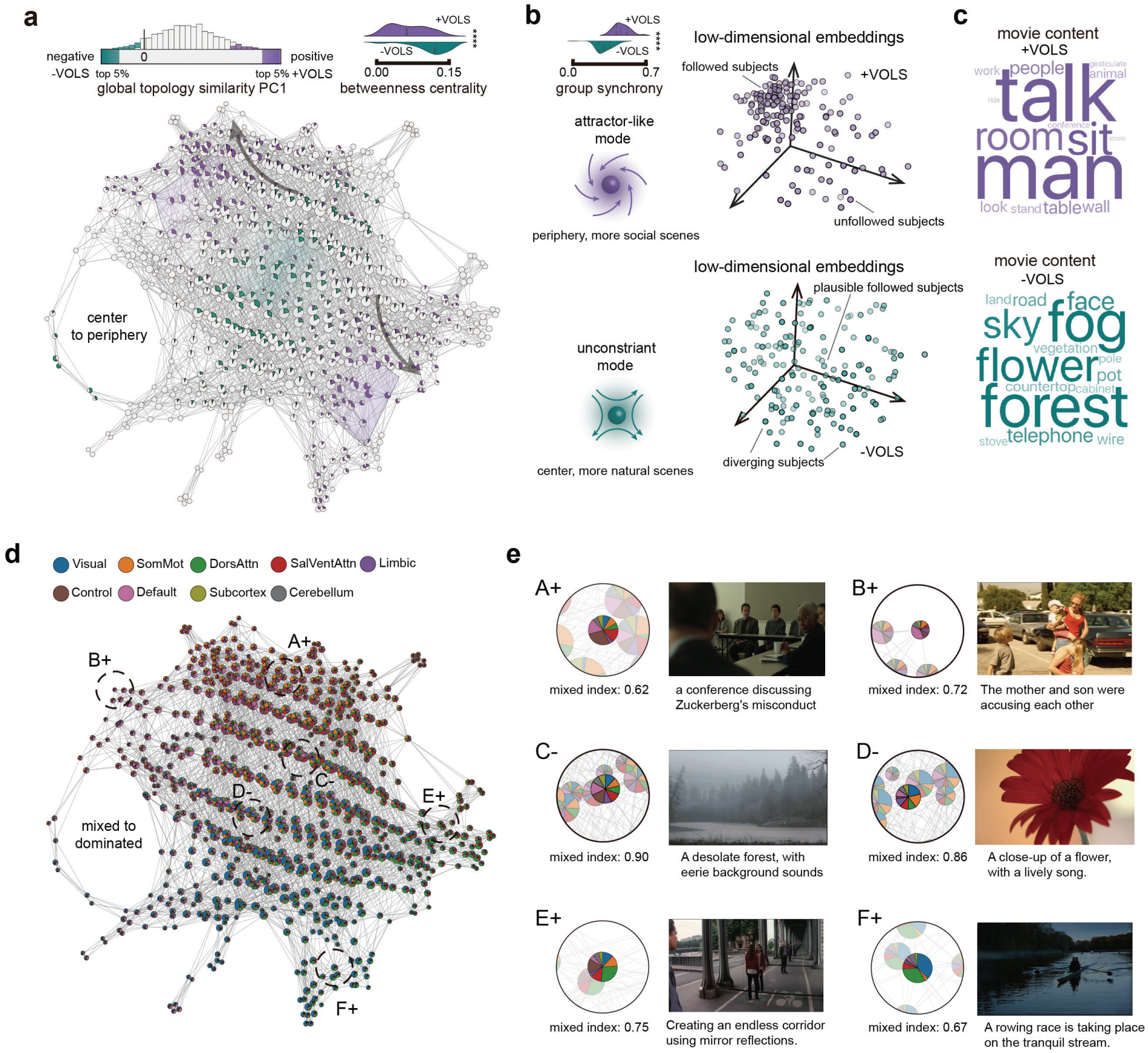
more dispersed (weaker consensus). We presented the low-dimensional embeddings of all subjects' volumes at two representative time points (Fig. 3b). These two modes also exhibited distinct preferences for movie content. We showcased the top 10 semantic labels where +VOLs and -VOLs differ most significantly in their relative proportions (Fig. 3c). We found that +VOLs were more associated with the human-related scenes, predominantly featuring the top entries such as 'man', 'talk', and 'sit'. In contrast, -VOLs were more aligned with nature scenes compared to +VOLs, with the top labels being 'forest', 'fog', and 'flower'.

How is the principal gradient interpreted from the perspective of brain network configuration? Previous work showed a dynamical topographic gradient of resting-state networks[27]. Similarly, we annotated the activation of 9 networks on the group shape graph and qualitatively depicted the network's domination using pie-chart proportions (Fig. 3d). We calculated a mixed index of nodes to describe the overall activation variation across the networks (see Methods). We found that the mixed index for +VOLs was significantly lower than that for -VOLs (group-level across time: t=-11.35, $P < 0.0001$; subject-level across individuals:t=-5.27, $P < 0.0001$), suggesting that the network configuration along the principal gradient shifts from uniform activation to the dominance of one or more networks. Integrating movie content into our analysis, we identified the top six movie segments aligned with the principal gradient, highlighted by magnified nodes. Four of these segments belonged to +VOLs (A+, B+, E+, F+ in Fig. 3e) and were associated with social interactions or human behaviors, engaging networks such as the control network, default mode network, dorsal attention network, and visual network. The remaining two segments belonged to -VOLs (C-, D- in Fig. 3e), which were linked to non-human scenes.

Therefore, individual divergence in group consensus can be interpreted as follows: At the +VOLs, subjects engage in ‘attractor-like’ modes, where higher scores indicate greater flexibility in adapting brain configurations to the diverse cognitive contents (Fig. 3f). At the -VOLs, subjects exhibit ‘unconstrained’ modes linked to nature scenes and mixed network activations, with higher scores suggesting less mind wandering and greater conformity with the group (Fig. 3g). Interestingly, within +VOLs, individuals who align more closely with the group consensus (demonstrating higher global topology similarity) have greater fluid intelligence (Spearman r = 0.30, $P < 0.0001$, Fig.3h); in contrast, within -VOLs, individuals who diverge from the group consensus (exhibiting lower global topology similarity) display increased fluid intelligence (Spearman r = -0.21, $P = 0.0053$, Fig.3i). The results suggest that the principal gradient of individual differences in the topological landscape explains cognition from a dual perspective: in 'attractor-like' modes, stronger dynamic ‘following’ to group latent states correlates with better cognition; in 'unconstrained' modes, greater divergence correlates with better cognition.

Next, we calculated a global principal gradient across all volumes (including +VOLs adherence and -VOLs divergence) for each participant. This gradient is significantly associated with cognitive domains but not with emotional or personality domains (Supplementary Tab. 3). The strongest correlation was observed with fluid intelligence (Spearman r = 0.32, P < 0.0001). This effect is comparable to the predictive capacity of static functional connectivity (sFC, Supplementary Fig.3). To ensure the gradient score as a trait-like marker, we computed principal gradients across two independent days, resulting in a relatively high correlation (Spearman r = 0.54, $P < 0.0001$). Lastly, we examined the contributions of the principal gradient to fluid intelligence at the brain regions level in the population weights. In brief, we calculated the principal variance of individuals for each region of interest (ROI, see Methods for more details). We found that +VOLs were associated with a majority of brain regions (Fig.3j), while significantly fewer regions correlated with -VOLs, notably involving the default mode network and the control network (Fig.3k).

the principal gradient of inter-individual *global topology similarity* exhibits a center-to-periphery mode over time.



the principal gradient of inter-individual *global topology similarity* explains intelligence from a dual perspective

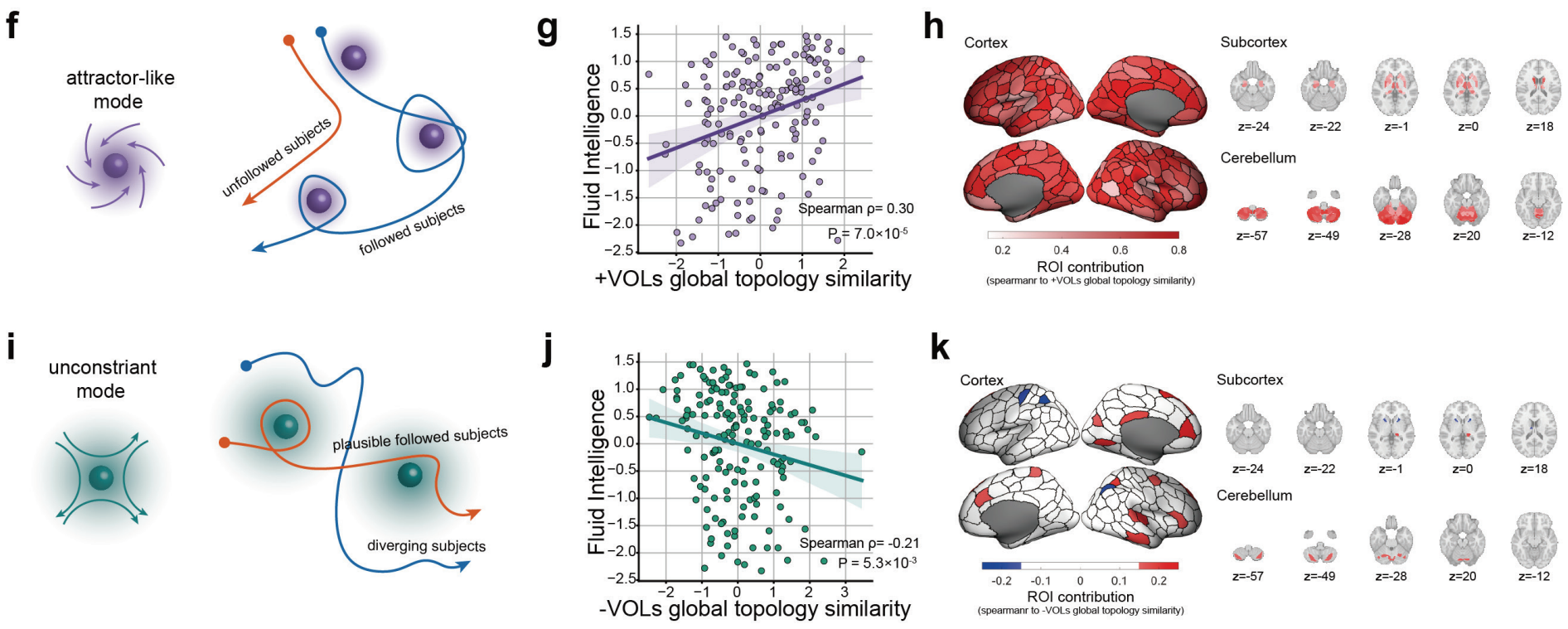


Fig 3 The principal gradient of individual differences in the global topology of the dynamical landscape. (a) We displayed a group shape graph from averaged embeddings and annotated time points at the extremes of PC1

(+VOLs and -VOLs). The distribution plot shows the betweenness centrality for +VOLs and -VOLs. **(b)** Population dynamics of +VOLs and -VOLs. The distribution plot illustrates the population follow for +VOLs and -VOLs. Population follow is defined as the average similarity among individuals, reflecting the degree of group synchrony at that time, corresponding to the illustration of attractor-like and unconstrained modes. UMAP embeddings for all individuals at the most extreme moments are displayed. **(c)** Word clouds depict the 15 terms with the greatest semantic frequency differences between +VOLs and -VOLs. **(d)** Brain network configurations are annotated on the group shape graph, with proportions representing network activation levels. **(e)** Across the most extreme time points in +VOLs and -VOLs, we displayed panoramas of six representative nodes, including network configurations, the mixed index, and corresponding movie content. **(f, g)** We illustrated two distinct dynamic patterns in group dynamics, where orange lines represent individuals not following the group, and blue lines represent those following the group. **(h, i)** Scatter plots show the correlation between topological similarity (follow score) and fluid intelligence for +VOLs and -VOLs. **(j, k)** The brain regions demonstrate the correlation between principal components of ROI-level inter-subject correlation and follow score. DorsAttn, dorsal attention network; SomMot, somatomotor network; SalVentAttn, salience/ventral attention network.

Local geometry of individual differences in topological landscape explains mental health

We hypothesize that the local geometry similarity corresponding to topological cliques can explain multi-dimensional human behaviors in a manner related to the specific movie content. To explore the associations, we included 59 behavioral measurements, including HCP-defined domains in *Cognition*, *Emotion*, *Psychiatric and Life Function* (Fig. 4a, Supplementary Tab.2). Based on the STIM framework, we derived 161 modular topological cliques (average duration = 17.4 seconds) as a range of non-overlapped contents that organized by brain dynamics and semantic information (Fig. 4b). We anticipated that some of the cliques, particularly those rich in cognitive, emotional, and well-being content, could act as 'probes' reflecting corresponding domains of human behavior (Fig. 4c).

We conducted a partial least squares regression (PLSR) to identify such 'probe-like' cliques which are behavioral relevant, resulting 25 cliques after permutation testing ($P < 0.05$). We next used pairwise correlation analysis to link specific behavioral domains to these 'probe-like' cliques (Fig. 4d, see Method). Using fluid intelligence as an example, a few topological cliques show significant correlations. especially a sci-fi scene in 'Inception', which had a comparable effect size to the global principal gradient (Spearman r=0.35, $P<0.0001$, Fig.4d). The *Cognition* domain was discussed in the previous section. Here, we focus on *Emotion* and *Psychiatric and Life Function*, combined as the mental health domain. Some measurements were reversed to align them as uniform 'negative traits' of mental health, such as transforming life satisfaction to life dissatisfaction (see Methods). The local geometry similarity of most scenarios negatively correlated with negative traits, while in a few scenarios was positively correlated with the negative traits (Fig. 4e, significance denoted in the legend, only displaying pairs that are significant and marginally significant: $P < 0.1$). The

contribution of different mental health dimensions was visualized using a word cloud, where font size linearly represents the mean significance rank (Fig.4f).

Among these, the most relevant behavior is *life dissatisfaction*, for which we highlighted the top four most correlated cliques: joyfully discussing gossip while dining with friends (No.58, ρ = -0.29, P = 0.00012); leisurely rowing a boat (No.57, ρ = -0.24, P = 0.0013); a lady happily introducing her farm accompanied by light music (No.108, ρ = -0.24, P = 0.0016); making a joke with a smile during a serious discussion (No.79, ρ = -0.23, P = 0.0028). These scenarios, accompanied by joyful facial expressions or relaxed conversational atmospheres, are positive aspects of life and social interactions. The second correlated dimension is *loneliness*, with the top four related cliques being: a single mother with her children remains silent, their expressions sorrowful, accompanied by sad music (No.131, ρ = -0.25, P=0.00099); the next clique, this single mother with three children seeks help (No.132, ρ = -0.22, P=0.0043); in a somber musical setting, deserted forests and dilapidated buildings are displayed (No.117, ρ = +0.21, P=0.0063) as well as No.58 (ρ = -0.21, P=0.0063). The scenarios associated with loneliness almost invariably involved somber atmospheres and unhappy expressions.

The findings on *life dissatisfaction* and *loneliness* indicate that local brain geometry accounts for individual differences in areas related to specific, explainable movie content. We next identified cliques most closely associated with negative traits of mental health, with the representative examples displayed below (see Methods; additional details in Supplementary Tab. 4): For *DSM (Diagnostic and Statistical Manual of Mental Disorders) Somatic Complaints*, the most related scenarios: In the process of Cobb seeing himself being intimate with his wife, his facial expressions vividly convey the astonished and pain of losing her (No.53, ρ = -0.25, P=0.00092); accompanied by somber music, an interviewee solemnly shows a photo of his deceased mother from his pocket (No.37, ρ = -0.32, P<0.0001). Both scenarios involve intense emotional trauma related to death, which can elevate psychological distress to physical discomfort[33]. Additional cliques-mental health pairs include: being continuously stared at unkindly by strangers, explaining interindividual *ASR (Adult Self Report) Internalizing Score* (No.47, ρ = -0.24, P=0.0021); viewing a child blames himself with a negative face, relating to *Sadness* (No.128, ρ = -0.25, P=0.00088); a montage sequence shows a man and a woman each lying on their sides in bed, corresponding to the *ASR Rule Breaking Behavior Score*. (No.103, ρ = -0.22, P=0.0040). Exposure to hostile expression tends to relate anger dimension: the main character receives a note with 'udick' (*Anger Hostility*, No.70, ρ = -0.18, P=0.021), and a 'damn' word is suddenly used in a dialogue, provoking an angry response from another character (*Anger Affect*, No.78, ρ = -0.19, P=0.013).

Interestingly, the number of cliques negatively correlated with negative mental health significantly exceeds those with a positive correlation (Fig. 4e), suggesting that negative traits primarily disrupt individuals' latent trajectories toward group consensus during specific movie scenes. We hypothesize that a dual perspective may also exist in the mental health domain: in a few 'unconstrained' mode cliques (such as No.117, the deserted forests), negative traits may prevent individual divergence, leading to immersion in a negative emotion. To verify this, we utilized +VOLs and -VOLs from the previous section, analyzing the directional correlation between individual differences and negative emotions (Fig. 4h). We found that individual differences in +VOLs are generally negatively correlated with negative traits (t=-8.8, $P < 0.0001$), whereas -VOLs are positively correlated with negative traits (t=4.7, $P < 0.0001$). This suggests that distinct individual dynamic patterns are linked to negative mental health in 'attractor-like' or 'unconstrained' modes (Fig. 4h).

Fig 4 The individual differences in the local geometry of the dynamical landscape. (a) Multi-dimensional behavior traits used for analysis. A total of 59 behaviors traits were included. **(b)** Multi-dimensional local geometry similarity for 161 cliques, including various movie content. **(c)** Hypothesis of probe-like cliques: the individual local geometry similarity of specific clique may relate to corresponding behavior traits. **(d)** Methodological illustration for identifying probe-like cliques: PLSR with permutation test was conducted to identify probe-like cliques. An example of a probe-like clique related to fluid intelligence: characters create a surreal corridor using mirror reflection. **(e)** Subsequently analysis focused on negative mental health (some behavior traits were reversed). A heatmap of the significant correlation levels between

probe cliques and negative mental health dimensions, arranged by the mean significance. **(f)** A weighted word clouds plot depicting the significance level for various negative mental health. **(g)** Movie contents corresponding to representative cliques significantly associated with mental health dimensions. The selection includes the four most significantly correlated cliques with *loneliness* and *life dissatisfaction*, and the other representative cliques across all probe-like cliques and mental health dimensions (excluding *loneliness* and *life dissatisfaction*). Cliques No.108, No.57, No.128, and No.78 did not pass the permutation test; nevertheless, we include these pairs because of their interpretive value. P-values denote raw Spearman correlations between clique-behavior pairs. **(h)** Dynamic modes of group dynamics in negative emotions. The distribution diagram illustrates the direction of Spearman correlation between the topological similarity in +VOLs and -VOLs (i.e., attractor-like and unconstrained mode in Fig. 3) and individual negative emotions score. The schematic displays the population dynamics of the +VOLs and -VOLs with negative emotions.

Using the STIM framework to characterize developmental progress and conditions

We further investigated whether the STIM framework could offer new insights into the developmental progress and conditions of brain dynamics. To this end, we utilized the Healthy Brain Network (HBN) dataset, which includes fMRI data from 970 participants aged 5-21, who watched a 10-minute clip of the animated film *Despicable Me*. This comedic segment contains numerous instances of social processing. The participants, presenting a wide array of transdiagnostic conditions (e.g., autism spectrum disorder), provide a comprehensive context for investigating developmental trajectories and the spectra of mental disorders (Fig. 5a).

We first focused on the application of STIM in developmental progress. To mitigate the impact of the age distribution of HBN participants, we used an HCP-trained STIM to compute individual topological landscapes (see Methods). We initially examined the consistency of brain dynamics (low-dimensional embeddings after the filter function f) across development. The results showed a significant positive correlation between mean group consistency and age (Fig.5b), indicating that older individuals tend to achieve a more stable consensus in their brain dynamics. We also observed that age differences negatively correlate with global topology similarity between different age-sorted groups ($n_{subjects}$=97 for each group), suggesting that groups closer in age exhibit higher similarities in their consensus (Fig.5c).

We then investigated whether developmental differences were associated with specific movie content. Using the same STIM pipeline, we divided the time volumes into 41 non-overlapping cliques (mean duration = 14.6 seconds). By manual labeling, we annotated the cliques with over 50% screen time featuring kids or minions (little banana-like creatures in the movie) (Fig.5d). Subsequently, we generated the consensus for two extreme age groups (100 subjects for each group) and compared how individuals from intermediate ages deviated from this consensus as the age gap increased (n=770 subjects). The lower part of Figure 5e shows specific cliques

where individuals deviated from the youngest group consensus (in orange) or engaged with the oldest group consensus (in blue) ($P<0.05$ after FDR correction). Interestingly, we found that cliques causing individuals to deviate from the youngest consensus, compared to those deviating from the oldest consensus, featured significantly more scenes with kids or minions when calculated by time volumes (χ^2 (1) = 28.35, $P < 0.0001$, Fig. 5e), indicating that scenes rich in child characters are more likely to induce individual experiential differences among younger viewers during movie watching. Figure 5f synthesizes these observations, illustrating that younger individuals show greater heterogeneity and lower synchronicity, with synchronicity increasing progressively with age, especially in cliques featuring kids or minions. This indicates a continuous maturation in global topology similarity.

We then examined the relationship between individual differences in global topology. First, we replicated the main results found in HCP data. A group-level shape graph was generated and annotated by the PC1 dimension on the global topological similarity matrix over time (top 5%: +VOLs; bottom 5%: -VOLs) (Fig.5g). Consistent with the HCP results, + VOLs exhibited significantly higher peripheral presence in the topological landscape (lower betweenness centrality: t=-4.61, $P<0.0001$) and an attractor-like mode (higher populational synchrony: t=6.28, $P<0.0001$) compared to -VOLs. Given the dialogue-rich content of the *Despicable Me*, we selected listening comprehension scores from the Wechsler Intelligence Scale (WIAT) as the cognitive metric. Across five equal-sized, non-overlapping age groups (Fig.5h), the principal gradient of global topology similarity significantly correlates with cognitive function in four groups (group-1: r=0.31, $P<0.0001$; group-2: r=0.08, $P>0.05$; group-3: r=0.15, P=0.032; group-4: r=0.36, $P<0.0001$; group-5: r=0.26, P=0.00031). In the HCP data, we observed differences between +VOLs and -VOLs in terms of movie content. However, due to the absence of nature scenes, we did not observe the significant opposite direction of cognitive correlation in HBN data.

For local geometry, we focused on the paired correlations between cliques and mental health issues. From the behavioral measurements provided by the HBN dataset, we selected seven cognitive dimensions and twelve psychological health dimensions with complete data, all adjusted to represent negative mental health outcomes (Supplementary Tab. 2). Using partial least squares regression (PLSR), we identified 14 probe-like cliques for behavioral correlation analysis. Consistent with the HCP findings, we discovered that specific movie content could explain the association between local geometry similarity and negative mental health (Fig. 5i). Typically, cliques associated with social mental problems primarily involved scenes of intense social interactions: three kids playing and yelling, refusing to go to bed (No.0,ρ=-0.15, $P<0.0001$), passing a gun to a bystander and

then flying away (No.37, ρ=-0.12, P=0.0007), and a minion vocally protesting to Gru while attempting to escape (No.40, ρ=-0.13, P=0.0004). Anxiety-related issues appeared linked to scenes depicting negative parental reactions, such as a minion's invitation being sternly rejected by Gru (No.34, ρ=-0.11, P=0.002). Emotional problems were potentially linked to extreme emotions: minions suddenly erupting in loud, prolonged laughter during play (No.18 ρ=-0.14, P=0.0002). Additionally, across all probe-like cliques, local geometry similarity was negatively correlated with negative mental health. The average correlation between +VOLs and negative mental health was significantly negative (t=-5.87, P<0.0001).

Given the rich social interactions in *Despicable Me*, we explored the potential application of local geometry in studying autism spectrum disorder (ASD). By conducting PCA on the behavioral measurements associated with ASD, we divided individuals into four groups based on progressively severe ASD-like symptoms (excluding NC individuals, n=165 for each group) We then compared the local geometry of these groups with the NC group (n=63) across all probe-like topological cliques. For further insight, we compared individuals officially diagnosed with ASD to NC individuals. We also annotated the movie content and dialogue for the three cliques most significantly associated with ASD-specific behaviors. Interestingly, these cliques all related to social behaviors (Fig. 5j). Moreover, as ASD-like symptoms intensified, the difference in following behaviors compared to NC individuals progressively increased. This pattern suggests that specific movie scenes could serve as effective probes for identifying particular mental disorders.

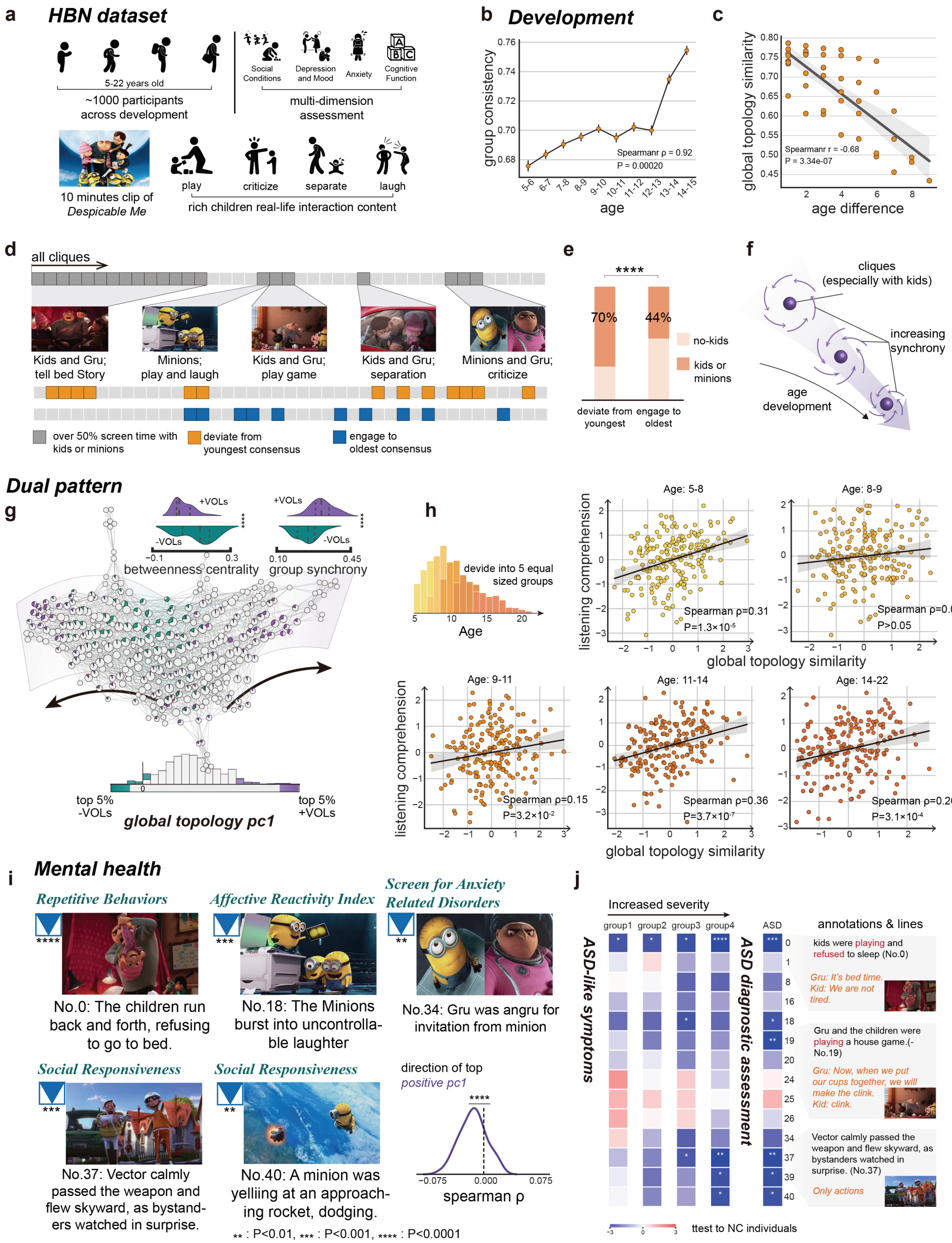


Fig 5 Application in developmental progress and conditions. **(a)** Overview of HBN dataset. **(b)** A point plot of inter-subject correlation within group across different ages, with Spearman's correlation coefficient indicating the relationship between mean group consistency and age, displaying the 95% confidence intervals. **(c)** Scatter plot of age difference and group topological similarity. **(d)** Gray barcodes graph displays cliques where 'kids' or 'minions' appear for over 50% of the screen time, with key characters and times described through snapshots and text. Orange

barcodes graph shows cliques that significantly deviate from the youngest consensus with increasing age. Blue barcodes graph depicts cliques significantly engaging into the oldest consensus as age increases. **(e)** ratio of child appearances in cliques shows significant differences in younger individuals and older individuals. **(f)** An illustration of developmental characteristics in topological properties. **(g)** Shape graph of averaging individuals, with annotations of top 5% and bottom 5% time points of principal gradient. Violin plots show differences of two parts of time points in betweenness centrality and group synchrony (mean global topology similarity across individuals). **(h)** The study population is divided into five age groups, with scatter plots showing the relationship between age and listening comprehension and topological similarity. **(i)** Five representative cliques with description and corresponding mental health. Distribution plot shows the direction of top 5% timepoints in negative mental health. **(j)** Barcodes plot shows relationship between all probe-cliques and ASD disorder (symptoms gradient and diagnostic assessment), with description of top three probe-like cliques.

Discussion

Naturalistic stimuli can elicit a continuous stream of rich subjective experiences through the adaptive coordination of neural systems. While watching a film is a collective experience[8], it also includes individual-specific elements likely determined by the viewer's cognitive and psychological conditions[34,35]. From a dynamical system perspective, we propose that individual divergence from group consensus is reflected in system-wide, low-dimensional brain latent dynamics, allowing for the probing of individual behavioral traits and mental disorders. To this end, we introduce a computational framework, STIM, designed to quantitatively elucidate the topological landscape of brain activity during naturalistic movie viewing. The STIM framework can generate a reliable group-level landscape of brain latent state dynamics and map individualized topological patterns into two distinct scales: global topology and local geometry, corresponding to individual divergence across and within specific movie cliques, respectively. Applying STIM to two large movie fMRI datasets, we demonstrated that the combination of rich naturalistic stimuli with brain recording holds great potential as neural 'probes' to measure individual traits in various behavioral domains. Specifically, global topology describes overall shape similarity and is specifically related to cognition: individuals who closely follow group dynamics in 'attractor-like' modes exhibit higher fluid intelligence, while those showing more diversity in 'unconstrained' modes also exhibit higher fluid intelligence. At the finer movie clique scale, local geometry explains multi-dimensional mental health traits beyond cognition. Lastly, we showed that STIM could capture the topological landscape related to developmental processes, with potential applications in detecting and understanding neurodevelopmental disorders such as autism.

STIM provides a novel method for analyzing brain states in the field of naturalistic fMRI, applicable to longer and more complex movie-viewing data. Compared to HMM, CAPs, and similar state-based methods, the TDA-based mapper can represent a richer array of brain states during movie viewing. Additionally, it automatically and flexibly captures rapid transitions, preventing the blurring of distinct brain states, and provides user-friendly, dynamic visualizations. We observed that using the Mapper technique to construct low-dimensional topological landscapes independently in different subjects exhibits similar rapid transitions in response to the same movie content (Fig. 2b). STIM can further align these low-dimensional spaces across individuals to generate a group template, thereby mapping individual-specific states to behaviors. Remarkably, this group-level topological landscape and the corresponding patterns of individual differences can be reliably established with a relatively small sample size (n=15). Thus, a cost-effective approach for future studies would involve collecting diverse movie clips from a small subset of participants to create a

rich, group-level topological reference[35]. The reference can then be used with plug-and-play movie segments tailored to various applications, guiding downstream scanning in larger populations. Although subjective evaluations during movie viewing are lacking, we speculate that STIM primarily captures the system-wide integration of brain activity during movie watching. This is especially evident considering that the synchronization of latent embeddings (using UMAP as the filter function) is higher than that original time series of visual regions, particularly in scenes requiring cognitive processing (Supplementary Fig. 2).

By applying the STIM framework to large-sample movie fMRI datasets, we have, for the first time, elucidated the individual differences in latent dynamics during watching rich movie contents. In terms of global shape-like topology, our findings reveal that different naturalistic stimuli elicit either common or heterogeneous group consensus. Furthermore, individual divergence is specifically correlated with cognitive abilities, but the direction of this association is context-dependent. The principal divergence in global topology spans multi-level domains, including topological attributes within the brain's landscape (e.g., shape graph), synchronous modes within the population (e.g., interindividual distance), brain configurations (e.g., spatial networks), movie contents (e.g., stimuli categories), and individual traits (e.g., behavioral measures). To enhance interpretability, we decomposed the first component of the global topology metric into two distinct temporal parts: +VOLs, representing top positive contributions, and -VOLs, representing top negative contributions. +VOLs correspond to more complex scenes, including characters and dialogues, aligning with our initial hypothesis of stimuli as anchors or attractor basins guiding shared low-dimensional latent dynamics across individuals. These attractors are distributed in the brain's landscape periphery and correspond to multiple network-dominant brain configurations. During movie-watching, individuals whose landscapes can flexibly 'follow' these +VOLs attractor-like states exhibit higher fluid intelligence. Based on PCA approach, Shine et al demonstrated a common attractor-like space facilitating the execution of multiple cognitive tasks, with individuals having greater fluid intelligence tending to exhibit more effective flow through the space[17]. We speculate that +VOLs may be homologous to this attractor-like space from a dynamical system perspective, but corresponding to more multifaceted, naturalistic cognitive processes. In contrast, -VOLs, which correspond to more natural scenery scenarios, occupy the brain's landscape center and exhibit more mixed network configurations. Seemingly similar patterns have been previously noted in the analysis of dynamic brain states. For instance, Saggar et al. used a TDA-based Mapper to reveal the topological center in individual landscapes during the resting state, suggesting that the uniform state acts as an intermediary to modulate transitions between other network-dominant states[27]. However, differed from previous studies, the -VOLs characterize the overall patterns across the group, thus averaging out the heterogeneous brain activations of many individuals. At the individual level, -VOLs could still encompass

numerous network-dominant volumes, but statistically, the mixed network level is significantly smaller than that of +VOLs. Interestingly, we have discovered for the first time that under specific stimuli (-VOLs), individuals with more divergent traversals from group-level latent dynamics exhibit higher fluid intelligence. This finding suggests the importance of spontaneous divergent thinking, similar to mind wandering, especially when encountering stimuli that lack explicit direction and cognitive load. Considering that the HBN dataset lacks natural-themed content, the first component of the global topology metric consistently has positive weights and can reliably explain individual differences in cognitive dimensions across different age groups.

A conceptual prospect of the naturalistic paradigm is to use designed movie fragments to induce rich brain activation, with the expectation that different fragments can assess various dimensions of behavior or disease symptoms. This potential has been suggested by previous studies, which showed that certain movie stimuli (e.g., independent video clips and the presence of faces) can predict sex[36], cognition[5], or emotion domains[37] to varying degrees. At a finer scale, our results suggest that specific movie scenes can serve as 'probes' to detect corresponding behaviors. This is facilitated by the STIM framework, which can segment complex movie content into local geometry clusters based on brain latent dynamics. Importantly, our reported clique-behavior pairs are largely explainable. For instance, the cliques most correlated with life satisfaction predominantly feature pleasant and peaceful conversations, while the two scenes associated with anger traits both contain unfriendly language, such as a note with 'udick' and a conversation with the word 'damn'. In the HBN dataset, the contagious laughter of the Minions clique corresponds to individual differences in the Affective Reactivity Index, a psychological assessment used to measure emotional reactivity. The clique-behavior explainability highlights that movie scenes can be further strategically designed to target specific dimensions of mental health[35,38], reducing the subjectivity of self-report measures[5,39]. In most cliques, negative mental health issues disrupt individuals from engaging corresponding group-level latent dynamics; within HCP -VOLs, scenes of natural, eerie forests show that individuals with greater divergence from the group tend to have a lower tendency towards negative traits (i.e., loneliness).

Our results suggest the potential of the STIM framework to detect developmental issues or psychiatric disorders and inspire intervention strategies. For the same movie stimuli, groups of similar ages exhibit comparable global topology similarity and become more synchronized within the group as they develop. Different developmental groups show varying latent dynamics, such as cliques relating to different characters. Individuals with autism significantly deviate from healthy controls in latent dynamics, especially in social scenes. Previous works have suggested that the brain activity of individuals with psychiatric

disorders during naturalistic stimuli shows abnormalities compared to healthy controls[40–42]. In the future, constructing landscape references of healthy groups from different age brackets to naturalistic stimuli as a normative model may enable real-time observation of individual divergence from demographic-specific templates using the Mapper visualization advantage. For example, an individual at high risk for autism may specifically deviate in certain social scenes, such as interactions with parents rather than peers. Understanding these deviations could potentially help customize behavioral intervention strategies at the individual level.

We consider several limitations of the current work. First, we used UMAP as the Mapper filter function, but its representational capacity is limited. Trainable deep learning models, such as the recently proposed CEBRA approach[43], could produce higher performance latent spaces. Second, our study is based on whole-brain regions, but for specific subdomains like the emotion domain, selecting key nodes in emotional circuits as inputs may yield better behavioral correlations[44,45]. Third, the STIM framework inherently studies brain activity in low-dimensional representations, which cannot resolve cross-region synchrony. This can be addressed using classical ISC methods. Fourth, the HBN dataset contains a very small proportion of healthy samples and has a wide age distribution, limiting the construction of models for classifying healthy controls and patients. Future studies could build age-matched disease cohorts with movie fMRI to explore the potential for individual-level diagnosis. Fifth, future studies could employ more convenient data collection methods, such as electroencephalograph (EEG) or magnetoencephalography (MEG).

Method

Human Connectome Project (HCP) dataset

Participants

The dataset was sourced from the Human Connectome Project (HCP) 7T release by Washington University–University of Minnesota[46,47], and included movie information, fMRI data, and behavioral data. We analyzed data from 170 healthy participants who completed the movie-watching sessions (average age = 29.4 years, SD = 3.3 years; 105 females). Notably, the dataset contained many twins and siblings, hence participants were from 89 unique families. In our validation tests, we mitigated potential biases from familial relationships. Ethics approval was obtained by the authors of the original studies in accordance with the Washington University Institutional Review Board (IRB), and the Washington University–University of Minnesota (WU-Minn HCP) Consortium ensured full informed consent from all participants.

Movie information

For details of the HCP movie protocol, please refer to the Human Connectome Project protocols for 7T imaging (https://www.humanconnectome.org/hcp-protocols-ya-7t-imaging). Briefly, participants viewed audiovisual movie clips in 4 runs across two separate sessions, with each run consisting of 4-5 distinct movie segments. There was a 20-second rest period between clips during which the word "REST" appeared on the screen. Audio was delivered through Sensimetrics earbuds. In our analysis, we excluded movie segments shorter than 2 minutes, resulting in 13 analyzable clips from various sources including independent films and Hollywood movies, providing a rich array of naturalistic stimuli. These segments covered multiple genres such as sci-fic (Inception), drama (Social Network; Ocean’s Eleven; Home Alone; Erin Brockovich; Empire Strikes Back), mystery (Two Men; 1212), music video (Off The Shelf), documentary (Welcome To Bridgeville), interviews (Mrs Meyers Clean Day; Pockets), and scenic films (Northwest Passage) (for detailed information on the movie segments, see Supplementary Tab. 1). To correlate with the real-time viewing experience, we used manually assigned semantic labels provided by HCP (7T_movie_resources/WordNet_MOVIE1_CC1.txt, etc.). These labels identified objects (e.g., road, men, tree) and human actions (e.g., talk, sit, walk) depicted in the movie scenes, with label frequency aligned to the volumes (i.e., providing an array of labels every second).

Data acquisition

All movie-viewing fMRI(mv-fMRI) data were acquired using a 7 Tesla Siemens Magnetom scanner. Among the four mv-fMRI scans conducted across four runs, runs 1 and 4 were acquired with anterior-to-posterior (AP) phase encoding, while the remaining two runs used posterior-to-anterior (PA) phase encoding. All mv-fMRI scans were collected using a gradient-echo-planar imaging (EPI) sequence with the following uniform parameters: repetition time (TR) = 1000 ms, echo time (TE) = 22.2 ms, flip angle = 45 degrees, field of view (FOV) = 208 × 208 mm (RO x PE), matrix size = 130 × 130 (RO x PE), slice thickness = 1.6 mm, resulting in 1.6 mm × 1.6 mm isotropic voxels with 85 slices per volume. Additional scanning parameters included a multiband factor of 5, an image acceleration factor (iPAT) of 2, partial Fourier sampling of 7/8, echo spacing of 0.64 ms, and a bandwidth of 1924 Hz/Px.

Structural T1-weights scans were acquired using a 3T Siemens Connectome Skyra scanner, with the following parameters: TR=2400ms, TE=2.14ms, TI=1000ms, flip angle=8degrees, FOV=224 × 224 mm, voxel size=0.7mm isotropic, iPAT=2, bandwidth=210 Hz/Px.

Data processing

We initiated our analysis using minimally preprocessed data (Movie Task fMRI 1.6mm/59k Functional Preprocessed)[48]. The HCP minimal processing is a standardized preprocessing pipeline tailored for the HCP dataset designed to minimize information loss. This includes gradient distortion correction, motion correction, fieldmap-based EPI distortion correction, EPI to T1w Registration, non-linear (FNIRT) registration into MNI152 space, and intensity normalization. Additionally, we performed nuisance signal correction by regressing out 24 motion parameters (six parameters of the current volume, six parameters of the preceding volume, and their 12 corresponding squared items), three temporal parameters (constant, linear, and quadratic trends), as well as signals from white matter (WMReg.2.nii.gz), cerebrospinal fluid (CSFReg.2.nii.gz), and the average global signal extracted from the brain mask (brainmask_fs.1.60.nii.gz). Lastly, a temporal band-pass filter (0.01Hz<f<0.08 Hz) was applied to the data.

Brain parcellation

Consistent with our prior work, we selected 271 predefined regions of interest (ROIs) spanning the entire brain. We obtained 200 cortical regions from the Schaefer atlas, 54 subcortical regions from the Melbourne atlas 7T version scale III, and 17 cerebellar regions from the Buckner atlas. Among the cortical ROIs, seven different networks were assigned based on previous work, including the visual network (for short: Visual), somatomotor network (SomMot), dorsal attention network (DorsAttn), ventral attention network (SalVentAttn), limbic network (Limbic), frontoparietal control network (Control), and default mode network (Default). The subcortical and cerebellar ROIs were treated as a whole, without further network subdivision. Considering the hemodynamic response delay, we shifted all time-series signals from movie-viewing by 5 seconds to correspond to 5 volumes.

Healthy Brain Network (HBN) dataset

Participants

The Healthy Brain Network (HBN) dataset is a large-scale multimodal dataset focused on the development and mental health of children and adolescents, collecting neuroimaging and behavioral data from various age groups in the New York area[49]. We utilized the first seven releases of the HBN. Data selection was based on the completeness of imaging and behavioral data, correctness of preprocessing, anomalies in head movement, and fMRI irregularities (see "Data Processing"). This resulted in a subset of 970 participants (ages 5-21, average age = 10.6 years, SD = 3.5 years; 341 females). Among these, 95 individuals had no diagnosis given (NC), while the majority were diagnosed with one or more psychological issues. The Child Mind Institute confirmed that adult participants provided full informed consent, and minors aged 5-17 gave assent, with their parents completing the full informed consent. For more details, please refer to: https://fcon_1000.projects.nitrc.org/indi/cmi_healthy_brain_network/index.html.

Movie information

The HBN dataset provides fMRI scans from two different movie segments watched with simultaneous audio: "The Present" (3 minutes 21 seconds) and a clip from "Despicable Me" (10 minutes). "The Present" narrates the story of a boy with an amputation who receives a three-legged puppy from his mom and eventually warms up to it. The "Despicable Me" clip recounts various events involving three children and the supervillain Gru, including Gru tucking the children into bed, discussing plans to steal the moon with Dr. Nefario, the children being returned to the orphanage, and Gru rocketing into space. Given that "The Present" is shorter and lacks dialogue and plot transitions, we chose to analyze only the data from the "Despicable Me" clip. Building on the work of Camacho, M.C., et al., we utilized publicly available movie annotations, including the presence of spoken words in the scenes. The frequency of these labels was aligned with the TRs (0.8s) of the fMRI scans.

Data acquisition

Neuroanatomical and functional data were collected from two different sites: the CitiGroup Cornell Brain Imaging Center (CBIC) using a Siemens 3T Prisma scanner and the Rutgers University Brain Imaging Center (RUBIC) using a Siemens 3T Tim Trio scanner, both employing a 64-channel head coil with identical scanning parameters. Mv-fMRI scans were conducted with the following parameters: TR=800ms, TE=30ms, flip angle=31 degrees, resolution=2.4 mm × 2.4 mm × 2.4 mm, 60 slices, FOV phase=100%, and a multi-band factor of 6.

T1-weighted scans were collected with: TR=2500ms, TE=3.15ms, TI=1060ms, flip angle=8 degrees, resolution=0.8 mm × 0.8 mm × 0.8 mm, 224 slices, and FOV phase=100%. We did not perform inter-site corrections or analyses during data processing; instead, site information was treated as a covariate and regressed out in subsequent behavioral analyses.

Data preprocessing

Anatomical and functional data preprocessing was conducted using fMRIPrep 20.2.1[50]. The T1-weighted (T1w) image underwent intensity non-uniformity (INU) correction with N4BiasFieldCorrection, provided by ANTs 2.3.3, and it was designated as the reference image. This reference T1w image was then skull-stripped using a Nipype implementation of the antsBrainExtraction.sh script from ANTs, utilizing the OASIS30ANTs template for anatomical accuracy. Following this, segmentation of brain tissues—specifically CSF, WM and GM—was performed on the skull-stripped T1w image using the FAST tool from FSL 5.0.9. The volume was spatially normalized to the MNI152NLin2009cAsym standard space through nonlinear registration using antsRegistration from ANTs 2.3.3, ensuring consistent anatomical alignment across studies.

For each subject, a reference volume and its skull-stripped version were generated. The BOLD reference was aligned (co-registered) to the anatomical T1-weighted (T1w) reference using the FLIRT tool from FSL 5.0.9. Motion correction was applied, and the data were resampled and normalized to the standard space (MNI152NLin2009cAsym). Metrics such as Framewise Displacement (FD), CSF, WM, and global signal were calculated. Noise reduction on the BOLD signals was performed using the CompCor method.

Similar to the process of HCP, we performed nuisance signal correction by regressing out 24 motion parameters, three temporal parameters (constant, linear, and quadratic trends), as well as signals from WM, CSF, and the global signal. Lastly, a temporal band-pass filter (0.008Hz<f<0.15 Hz) was applied to the data. After data processing, only 1,227 subjects had complete imaging data (or were correctly preprocessed) along with basic information (age, gender, disease diagnosis). Among these, we excluded 159 individuals with excessive head motion (mean FD > 1mm). Based on results from the HCP, which demonstrated that movie viewing could induce similar brain activities, we also removed 98 individuals whose data showed a low mean Pearson correlation with the average time series (mean Pearson correlation < 0.1), functionally suggesting that these individuals were not engaged in the movie watching (Supplementary Fig. 4).

Brain parcellation

Consistent with our prior work, we selected 271 predefined regions of interest (ROIs) spanning the entire brain. We obtained 200 cortical regions from the Schaefer atlas[51], 54 subcortical regions from the Melbourne atlas 7T version scale III (3T version scale IV for HBN)[52], and 17 cerebellar regions from the Buckner atlas[53]. Among the cortical ROIs, seven different networks were assigned based on previous work[54], including the visual network (for short: Visual), somatomotor network (SomMot), dorsal attention network (DorsAttn), ventral attention network (SalVentAttn), limbic network (Limbic), frontoparietal control network (Control), and default mode network (Default). The subcortical and cerebellar ROIs were treated as a whole, without further network subdivision. Considering the hemodynamic response delay, we shifted all time-series signals from movie-viewing by 5 seconds to correspond to 5 TRs in HCP. As the HBN movie-viewing protocol does not include resting periods before and after the movie, and thus does not allow for

temporal shifts in the time series, we compensated by delaying the alignment of movie-related labels with the BOLD time series by 5 seconds (equivalent to 6 TRs).

Classic mapper

We utilized the mapper pipeline developed by Geniesse et al. to generate the shape graphs for individuals as shown in Figures 1b and 2b[28]. In this approach, the input data consisted of parcellated time series for each participant; for Figure 1b, we used 7 movie clips with dimensions of $1439\ volumes \times 271\ ROIs$, and for Figure 2b, the Inception clip with dimensions of $226 volumes \times 271 ROIs$. The mapper pipeline outputs a graph object.

The classic mapper pipeline includes the following steps: 1) Dimensionality Reduction: The high-dimensional data are embedded into a lower dimension (d=3) using a filter function. We chose UMAP to preserve the topological features of the data, utilizing PCA for initialization to maintain global structural integrity. 2) Overlapping Binning: This stage involves overlapping three-dimensional binning to encapsulate data points, with a resolution parameter (#bins) set to 12, resulting in a total of $12 \times 12 \times 12 = 1728\ bins$, and a 50% overlap between bins. 3) Partial Clustering: Within each bin, partial clustering groups data points into nodes based on the original high-dimensional information. We utilized density-based spatial clustering of applications with noise (DBSCAN) [55], which does not require a predetermined number of clusters. 4) Shape Graph Generation: The mapper approach treats each cluster within the bins as a node (containing multiple volumes). Overlapping bins cause some nodes to contain identical volumes, thereby generating edges between nodes.

STIM framework

Expanding upon the classic mapper, we introduce the STIM framework, designed to analyze individual differences in response to natural stimulus fMRI data. The primary difference between STIM and the classic mapper is that STIM emphasizes studying individual differences from the perspectives of consensus and divergence. STIM processes the complete time series data from multiple participants, unfolding and aligning individual topologies to establish a consensus at the group level. Then, by comparing the global and local topological differences between individual subjects and the consensus, STIM provides two distinct metrics of individual differences: global topology and local geometry.

Group consensus

Specifically, we modified the dimensionality reduction process and introduced group low-dimensional dynamics as consensus: 1) Aligned Dimensionality Reduction: We stacked the time series data ($170\ subjects \times 2804\ volumes$) of all individuals for dimensionality reduction (dimensions of HCP data: $476{,}680\ volumes \times 271\ ROIs$), creating an aligned low-dimensional space. From this, we extracted each individual's low-dimensional embedding (dimensions: $170\ subjects \times 2804\ volumes \times 3$) and normalized these embeddings using z-score standardization. 2) Introduction of Group Dynamics: We averaged the low-dimensional embeddings to represent the group consensus dynamics. These group dynamics were then processed alongside each individual's data through subsequent steps (dimensions: 5608×3), including Overlapping Binning, Partial Clustering, and Shape Graph Generation. Three common filter functions f were considered, including PCA, t-SNE, and UMAP. By comparing the robustness of group consensus dynamics in different sample size of participants, we select UMAP as filter f (Supplementary Fig. 5).

The shape graphs can be annotated with various meta-information, just as in the classic Mapper pipeline. These annotations provide insights into nodal degrees, activations in brain networks, and other relevant metrics, enabling a comprehensive understanding of the data's underlying neural mechanisms.

Temporal connectivity matrix

Employing the STIM framework, we aligned the complex dynamics across different participants within a shared topological space, abstracting these into a comprehensive shape graph. Employing graph theory techniques, we quantified the topological similarity between time points by utilizing the clustering relationships within nodes and the connections between nodes, resulting in a Temporal Connectivity Matrix (TCM, denoted by T). Each element T_{ij} within T quantifies the connectivity strength between time frames VOL_i and VOL_j.

The connectivity strength, T_{ij}, is calculated as the ratio of the cardinality of the intersection set of nodes shared between time frames VOL_i and VOL_j share to the maximum cardinality of the intersection of nodes involving VOL_i and any other time frame VOL_p across the entire dataset. Symbolically, this is expressed as:

$$T_{ij} = \frac{|V_i \cap V_j|}{\max\limits_{p \in \{1,2,\ldots,n\}} |V_i \cap V_p|}$$

where '$|\cdot|$' indicates the cardinality of a set, V_i and V_j are the sets of nodes involving with time frames VOL_i and VOL_j, respectively, and the denominator is the maximum cardinality obtained by the unite of V_i with every other set V_p across all time points p.

The resulting matrix T is symmetrized by:

$$TCM = \frac{(T + T^{\top})}{2}$$

This matrix provides a measure of topological similarity on a fine temporal scale. Notably, topological similarity is a discrete measure, therefore, we use Spearman rank correlation for subsequent association analyses.

In the classic mapper, the TCM computed for a single subject is dimensioned as $n\ volumes \times\ n\ volumes$ In the adapted mapper, which integrates group dynamics, the resulting TCM dimensions expand to $2n\ volumes \times\ 2n\ volumes$. From this, we extract the intersecting matrix to measure topological similarity between different individuals (Supplementary Fig. 6).

Individual divergence

Features were extracted from the Temporal Connectivity Matrix (TCM) across multiple scales.

At the global scale, we extracted the diagonal of the TCM, defined as global topology similarity (dimensions for HCP data: $170\ subjects\ \times\ 2804\ volumes$). This metric quantifies the differences in shape between two trajectories. In the STIM framework, due to discontinuous overlapping bins, global topology similarity is discrete values. This discrete, or 'fuzzy', metric is particularly appropriate for the analysis of biological signals, which are inherently imprecise. Specifically, within the context of movie viewing, global topology similarity reflects the extent of an individual's alignment with the group consensus.

At the local scale, the TCM matrix visually exhibits a modular structure, suggesting that many individuals remain within an attractor-like clique for short periods. A standard change point detection algorithm, pruned exact linear time (PELT), was employed to segment the averaged TCM. We excluded the 50% of subjects whose data demonstrated lower similarity during the averaging process, to ensure that the segments were not influenced by outliers. The parameters for PELT were selected using a data-driven approach, with the evaluation metric being the degree of clustering modularity, as proposed by Busch, E.L., et al[56]. Using this algorithm, we identified 161 independent cliques in the HCP data (mean duration=16.2s, SD=9.0) and 41 independent cliques in the HBN data (mean duration=13.7s, SD=7.0). Moment of clique separation were mapped back onto individual TCMs, from which submatrices corresponding to distinct cliques were extracted. These submatrices were subsequently integrated with the group average clique through a weighted summation to compute the local geometry for individuals within specific cliques. Local geometry reflects the differences between an individual's experience and the group consensus under a specific movie content.

Reliability of STIM framework

Three analytical methods were employed to assess the reliability of our framework:

1) Stability of Consensus: In the STIM framework, the group reference is established as the averaged low-dimensional dynamics. The stability of this consensus was evaluated by randomly sampling two independent groups of participants and comparing their averaged group references. To ensure a fair comparison, time series data were derived from three randomly selected ROIs. Notably, the individuals in each sampled group had no familial relationships. We tested group references ranging from 5 to 40 individuals and found that the group reference consistency, measured by Pearson correlation, were significantly higher than those of the time series (Fig. 2e).

2) Stability of Individual Differences: Based on the group reference, we calculated the mean global topology under different sample sizes of group references and compared it to the mean global topology across all subject group references using Pearson correlation. We found that with a smaller sample size ($n_{subjects} = 15$), the similarity of individual differences to the group reference formed from all subjects ($n_{subjects} = 170$) averaged 0.98 across multiple samples (Fig. 2f).

3) Integration of whole-brain dynamics: Neural signals from the visual network can effectively reconstruct visual stimuli, as demonstrated in previous studies. To investigate whether the STIM framework integrates flexible transitions across multiple scales of brain dynamics (not just mirroring stimuli), we compared the group consistency using Pearson correlation between whole-brain low-dimensional dynamics and time series from the Visual network. The time series were selected from the three most consistent ROIs within the Visual network. We found that in most movies, the consistency of the group reference obtained through the STIM framework was significantly higher than that of the top 3 ROIs in the Visual network (Supplementary Fig. 2).

Interpretation of local topological cliques

Utilizing the semantic labels, we calculated a semantic temporal connectivity matrix (semantic-TCM), where the elements T_{ij} represent the proportion of overlapping word entries between semantic vectors T_i and T_j. The semantic-TCM was then inputted into the PELT algorithm to segment it into non-overlapping semantic cliques. Subsequently, we minimized the linear error between segments of topological cliques and segments of semantic cliques using a standard algorithm for solving the linear assignment problem[57]. As a null model, we implemented random segmentation along the temporal axis with the same number of segments and repeated this process 10,000 times.

Behavioral items selection

In the HCP dataset, we used 59 behavioral items from three categories: *Cognition*, *Emotion*, *Psychiatric and Life Function*. For a detailed list, please refer to the Supplementary Tab. 2. Subscales under the Substance Use categories were not used due to a high prevalence of missing values or predominantly zero values among the movie-viewing subjects. In the subsequent analysis that pairing cliques with behavior, we focused on divergence in the mental health. Therefore, *Cognition* and *Emotional Recognition* metrics were excluded from this part of study (Fig. 4e-h).

In the HBN data, we selected the Wechsler Individual Achievement Test (WIAT) for its completeness in the cognitive behavior assessment. The WIAT scores include raw scores (_Raw), percentile scores (_P), and

scale scores (_Stnd). We used the scale scores for our analysis. Considering that the movie content primarily consists of dialogue, we chose the listening comprehension subtest (i.e., WIAT_LC_Stnd) from the various subscales. For other cognitive behavior test scales provided in the HBN dataset, please refer to the Supplementary Tab. 3. In terms of mental health, we utilized scales from the HBN dataset that had non-NAN values for more than 500 participants and included total score metrics. Consistently with the HCP data, we selected self-report versions of these scales, excluding participants younger than 8 years from our mental health analysis[49]. We identified 12 relevant items within the mental health dimension, detailed in Supplementary Tab. 2. For autism spectrum disorder (ASD) analysis, we employed four scales: the Autism Spectrum Screening Questionnaire (ASSQ), Social Communication Questionnaire (SCQ), Social Responsiveness Scale (SRS), and Repetitive Behaviors Scale (RBS). Using PCA, we derived the first principal component (PC1) as the primary indicator for ASD-like symptoms. We then divided the participants, excluding those with no diagnosis given (NC), into four groups of approximately equal size.

Global topology

PCA

We conducted Principal Component Analysis (PCA) on global topology similarity of individuals (dimensions for HCP data: $170\ subjects \times 2804\ volumes$) to assess the presence of a dominant topological gradient. In the HBN data, we divided participants into five equally sized age groups (dimensions in HBN: $5 \times 194\ subjects \times 750\ volumes$) and then conducted PCA to calculate the first principal component (PC1) for each group separately, aiming to mitigate the effects of systemic changes in group consensus during developmental processes.

Betweenness centrality

We applied the concept of betweenness centrality from graph theory to estimate the topological attributes of the shape graph. Betweenness centrality measures a node's role in facilitating information flow within the network by quantifying its presence on the shortest paths between other nodes[32].
Mathematically, the betweenness centrality $C_B(v)$ for a node v is defined as:

$$C_B(v) = \sum_{s \neq v \neq t} \frac{\sigma_{st}(v)}{\sigma_{st}}$$

where σ_{st} is the total number of shortest paths from node s to node t and $\sigma_{st}(v)$ represents the number of those paths that pass through v. This measure sums the fraction of all shortest paths that pass through a given node across all pairs of nodes. Due to the potential overlap of volumes across nodes in the shape graph, we calculated the average betweenness centrality of these nodes to determine their corresponding volumes' centrality.

Group synchronous modes

We averaged the global topology similarity across all individuals to identify the population synchrony at different volumes. The synchrony in +VOLs was significantly higher than in -VOLs, indicating two distinct synchronous modes. Additionally, in the HCP data, we observed a linear positive correlation between volumes population synchrony and the weights of PC1 (Pearson $r = 0.48$, $P<0.0001$), indicating that individuals who 'follow' the consensus tend to do so consistently across different volumes. Notably, this characteristic was observed only in older age groups within the HBN data.

Network activation configuration

Adapting methods from Saggar et al. on resting-state networks, we examined network configurations in mv-fMRI[27]. The network configuration is determined by the overall activation level within a shape graph node,

indicated by the proportion of different blocks. We adjusted brain activation by setting negative values in the normalized time series to zero. To determine whether a dominant network configuration exists within nodes, we calculated a node's network mixed index as $\frac{1}{1+S.D.}$, where $S.D.$ represents the standard deviation of the activation levels across different networks. A higher network mixed index signifies more uniform activation across networks.

Network contribution to global individual differences

The contribution of each Region of Interest (ROI) to global topology similarity was separately analyzed for +VOLs and -VOLs. Taking +VOLs as an example, we first concatenated the time series data of +VOLs (the dimension: 170 subjects×271ROIs×140 volumes). For each ROI, we computed the Pearson correlation between subject and group averaged time-series (denoted as ROI similarity). Following this, we calculated the Pearson correlation between ROI similarity and the averaged global topology similarity of +VOLs. This measure represents the ROI's contribution to the individual divergence observed in +VOLs.

Local level analysis

Selection of probe-like cliques

Considering the high-dimensional nature of both cliques and behavioral dimensions, we adopted a two-stage approach to identify statistically significant pairs: 1) Identification of statistically significant cliques with behavior: We performed PCA on 59 behavioral items, using the first five principal components (PCs)that explained 49.6% of the total variance in HCP data (71% in HBN data). Subsequently, we used Partial Least Squares Regression (PLSR) to predict individual local geometry similarities within each clique, based on the behavioral PCs. We then selected cliques that demonstrated statistical significance in permutation tests ($P < 0.05$, n = 10,000 permutations), with each permutation involving a shuffle of subject IDs across local geometry similarities. 2) Behavioral correlation analysis: we calculated the Spearman correlation between individual local topological similarities and single behavioral items, controlling for age and gender as covariates. This step measures how specific cliques explain specific behaviors.

Cliques-Mental health pair wise analysis

After identifying probe-like cliques and their pairwise relationships with specific behaviors, we analyzed the behavioral dimensions of mental health. We assessed the gradients under the mental health dimension during movie watching by averaging the significance of probe-like cliques. The derived weights for different behaviors were visualized using a word cloud. For clarity, we inverted the positive mental health scores, including life satisfaction (*LifeSatisf_Unadj*), meaning and purpose (*MeanPurp_Unadj*), positive affect (*PosAffect_Unadj*), friendship (*Friendship_Unadj*), emotional support (*EmotSupp_Unadj*), instrumental support (*InstruSupp_Unadj*), and self-efficacy (*SelfEff_Unadj)*.

In presentation focusing on clique-mental health pairs, we first concentrated on the two behaviors with the highest weights—*loneliness* and *life dissatisfaction.* We highlighted the four cliques most significantly associated with each of these behaviors. The displayed P-values are the raw Spearman correlation P-values between each behavior and the local geometry similarity of the respective cliques. Among all probe-like cliques, we displayed the other representative clique-behavior pairs; for a complete list, please refer to Supplementary Tab. 4. Notably, we also displayed some significant clique-behavior pairs that did not pass the permutation test, including No.57, No.108, No.78, and No.128.

Dual directions of dynamics modes

To further investigate how negative mental health affects individual divergence, we utilized the global topology similarity (170 subjects × 2804 volumes) from +VOLs and -VOLs, resulting in attractor-like and unconstrained modes for individual topological scores (2 × 170 subjects × 140 volumes). We computed the Spearman correlation for each TR between individual topological scores and the top 10 negative mental health metrics (life satisfaction, loneliness, ASR somatic problems, ASR internalizing, perceived rejection, DSM

somatization, sadness, perceived stress, DSM anxiety, and anger hostility). Finally, we performed a one-sample t-test with a mean of zero on the average Spearman correlations for +VOLs and -VOLs (Fig. 4h).

Developmental analysis

Accommodation of developmental biases

In the HBN data, developmental biases arise from two main factors: 1) Varying cognitive levels across different age groups influence participant responses under movie-watching conditions, resulting in shape graphs that are not directly comparable. 2) The age distribution, particularly the concentration of participants between the ages of 8 and 10, introduces bias in the shared topological space generated from all individuals.

In experiments primarily concerned with developmental issues (Fig. 5b-h), we adjusted the STIM framework to mitigate these biases. Specifically, we employed the shared low-dimensional space trained from the HCP data, and independently transformed individuals from the HBN dataset to establish a bias-free dynamic. We then categorized participants into distinct groups to compute group-specific references.

We hypothesized that strict control for developmental biases is not necessary in analyses related to mental health. Therefore, we utilized data from both HCP and a representative selection of the HBN dataset in the STIM framework to fit a shared topological space, thereby enhancing the specificity of brain dynamics analyses within the HBN data (results in Fig. 5i-j). The selection of representative-HBN time series involved choosing the five volumes with the highest Pearson correlation in each age group at each time volumes (resulting in dimensions of 50×271×750). These time series are compiled from multiple individuals rather than originating from a single individual source. Group references were generated by dividing the participants into five groups, calculating references separately for each, and then aggregating the results.

Topological developmental pattern

To investigate the consistency of low-dimensional dynamics across different age groups, we divided the participants into one-year age intervals. Within each interval, we calculated the consistency of group consensus derived from two different sample subjects (15 subjects for a group). Figure 6b shows the mean value and 95% confidence intervals of the consistency for 1000 times random selection, alongside the Spearman correlation between these mean values and age, illustrating developmental trends.

Further, we explored developmental patterns from the perspectives of consensus and divergence. Considering the uneven distribution of participants across different age groups, we grouped participants into ten non-overlapping groups of equal size ($10 \times 97 subjects$) to ensure a balanced analysis. For consensus analysis, we first calculated the consensus within each group using the STIM framework. Then, we computed and averaged the global topology similarity for each pair of groups.

Context analysis

To analyze the divergence in specific movie content across developmental stages, we established reference groups at the two extremes: the youngest and the oldest age groups, each comprising 100 individuals. We then calculated the correlation between the local geometry similarity of the remaining individuals and their age, using these two reference groups for comparison. We identified cliques specific to developmental stages, with a significance determined at a threshold of $P < 0.05$, adjusted for false discovery rate (FDR).

Subsequently, we manually annotated those cliques prominently featuring scenes with children or minions (constituting over 50% of screen time). A chi-square test was then performed to determine whether scenes prominently featuring kids or minions significantly contribute to developmental divergence.

Behavior association analysis in HBN

For global topology analysis, we divided the participants into five equal-sized groups, ensuring independence in the result. We then calculated the global topology for each participant within these groups ($5 \times 194 subjects \times 750\ volumes$). Subsequently, we computed the Spearman correlation between the principal gradient of the global topology similarity ($5 \times 194 subjects$) and WIAT_LC_Stnd ($5 \times 194 subjects$).

For local geometry analysis, we focused on participants older than 8 years, totaling 725 subjects. While clique segmentation was uniform across the individuals, group references were calculated separately for each of the five groups ($5 \times 145 subjects$). We standardized the local geometry of individuals within each group and aggregated the data ($725 subjects \times 41 cliques$). Following the same process as the HCP data analysis, we performed permutation tests on different cliques to identify probe-like cliques ($725 subjects \times 14\ cliques$) and their paired behavior. Due to the limited number of probe-like cliques identified, we highlighted the top five clique-behavior pairs. Details of all probe-like cliques and their corresponding behaviors with the highest weights are included in Supplementary Tab. 4.

Unlike the HCP data, the fMRI data for HBN were collected from two different sites. Therefore, in all correlation analyses of global topology and local geometry, we controlled for confounding variables by regressing out age, gender, and site information from both the behavioral and topological metrics.

Statistically testing

To assess statistical significance in our analysis, we adopted the following approach: At the global level, we directly calculated the Spearman correlation between the first principal component of global similarity and various behaviors. Consequently, we conducted False Discovery Rate (FDR) cross-validation on all trait-like behaviors provided by HCP (Supplementary Tab. 3). In the HBN data, we applied a similar correction for multiple comparisons across different total score items (Supplementary Tab. 3).

At the local level, we analyzed the statistical significance of overall behaviors with individual cliques. To this end, we generated a null distribution by randomizing local topological similarity (i.e., shuffling participant IDs) to assess statistical significance. This was done by performing 10,000 random permutations.

Parameter perturbation analysis

Previous work has demonstrated the robustness of the mapper approach to parameter perturbations. Once again, we tested the impact of a wide range of Mapper parameters on the main outcomes of individual differences using HCP data. We varied the resolution parameter and the overlap between bins, calculating global and local topological similarities at different resolutions. The results indicate that under most parameter settings, the topological properties and the outcomes related to individual differences remain stable (Supplementary Fig. 7).

Reference

1. Gonzalez-Castillo, J. *et al.* Tracking ongoing cognition in individuals using brief, whole-brain functional connectivity patterns. (2015) doi:10.1073/pnas.1501242112.

2. Dm, B. *et al.* Function in the human connectome: task-fMRI and individual differences in behavior. *NeuroImage* **80**, (2013).

3. Task fMRI paradigms may capture more behaviorally relevant information than resting-state functional connectivity. *NeuroImage* **270**, 119946 (2023).

4. Liégeois, R. *et al.* Resting brain dynamics at different timescales capture distinct aspects of human behavior. *Nat. Commun.* **10**, 1–9 (2019).

5. Es, F. & Pa, B. Movie-watching outperforms rest for functional connectivity-based prediction of behavior. *NeuroImage* **235**, (2021).

6. Naturalistic Stimuli in Neuroscience: Critically Acclaimed. *Trends Cogn. Sci.* **23**, 699–714 (2019).

7. Ip, J., M, S., E, G. & J, A. Movies and narratives as naturalistic stimuli in neuroimaging. *NeuroImage* **224**, (2021).

8. U, H., Y, N., I, L., G, F. & R, M. Intersubject synchronization of cortical activity during natural vision. *Science* **303**, (2004).

9. Kauppi, J.-P., Jääskeläinen, I. P., Sams, M. & Tohka, J. Inter-subject correlation of brain hemodynamic responses during watching a movie: localization in space and frequency. *Front. Neuroinformatics* **4**, 669 (2010).

10. Nastase, S. A., Gazzola, V., Hasson, U. & Keysers, C. Measuring shared responses across subjects using intersubject correlation. *Soc. Cogn. Affect. Neurosci.* **14**, 667–685 (2019).

11. Idiosynchrony: From shared responses to individual differences during naturalistic neuroimaging. *NeuroImage* **215**, 116828 (2020).

12. Intersubject representational similarity analysis reveals individual variations in affective experience when watching erotic movies. *NeuroImage* **216**, 116851 (2020).

13. Simony, E. *et al.* Dynamic reconfiguration of the default mode network during narrative comprehension. *Nat. Commun.*

7, 1–13 (2016).

14. Meer, J. N. van der, Breakspear, M., Chang, L. J., Sonkusare, S. & Cocchi, L. Movie viewing elicits rich and reliable brain state dynamics. *Nat. Commun.* **11**, 1–14 (2020).

15. Song, H., Shim, W. M. & Rosenberg, M. D. Large-scale neural dynamics in a shared low-dimensional state space reflect cognitive and attentional dynamics. *eLife* https://elifesciences.org/articles/85487 (2023) doi:10.7554/eLife.85487.

16. Lga, F. *et al.* Time-resolved effective connectivity in task fMRI: Psychophysiological interactions of Co-Activation patterns. *NeuroImage* **212**, (2020).

17. Shine, J. M. *et al.* Human cognition involves the dynamic integration of neural activity and neuromodulatory systems. *Nat. Neurosci.* **22**, 289–296 (2019).

18. R, C., B, G., B, P., A, P. & I, F. The intrinsic attractor manifold and population dynamics of a canonical cognitive circuit across waking and sleep. *Nat. Neurosci.* **22**, (2019).

19. Ss, K., H, R., S, D. & V, J. Ring attractor dynamics in the Drosophila central brain. *Science* **356**, (2017).

20. Gardner, R. J. *et al.* Toroidal topology of population activity in grid cells. *Nature* **602**, 123–128 (2022).

21. Singh, G., Mémoli, F. & Carlsson, G. Topological Methods for the Analysis of High Dimensional Data Sets and 3D Object Recognition.

22. Saggar, M. *et al.* Towards a new approach to reveal dynamical organization of the brain using topological data analysis. *Nat. Commun.* **9**, 1–14 (2018).

23. Nielson, J. L. *et al.* Topological data analysis for discovery in preclinical spinal cord injury and traumatic brain injury. *Nat. Commun.* **6**, 1–12 (2015).

24. Nicolau, M., Levine, A. J. & Carlsson, G. Topology based data analysis identifies a subgroup of breast cancers with a unique mutational profile and excellent survival. (2011) doi:10.1073/pnas.1102826108.

25. Chen, Y. & Volić, I. Topological data analysis model for the spread of the coronavirus. *PLOS ONE* **16**, e0255584 (2021).

26. Saarimäki, H. Naturalistic Stimuli in Affective Neuroimaging: A Review. *Front. Hum. Neurosci.* **15**, 675068 (2021).

27. Saggar, M., Shine, J. M., Liégeois, R., Dosenbach, N. U. F. & Fair, D. Precision dynamical mapping using topological

data analysis reveals a hub-like transition state at rest. *Nat. Commun.* **13**, 1–19 (2022).

28. Geniesse, C., Sporns, O., Petri, G. & Saggar, M. Generating dynamical neuroimaging spatiotemporal representations (DyNeuSR) using topological data analysis. *Netw. Neurosci.* **3**, 763 (2019).

29. McInnes, L., Healy, J. & Melville, J. UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction. *arXiv.org* https://arxiv.org/abs/1802.03426v3 (2018).

30. Maaten, L. van der & Hinton, G. Visualizing Data using t-SNE. *J. Mach. Learn. Res.* **9**, 2579–2605 (2008).

31. Killick, R., Fearnhead, P. & Eckley, I. A. Optimal Detection of Changepoints With a Linear Computational Cost. *J. Am. Stat. Assoc.* **107**, 1590–1598 (2012).

32. Freeman, L. C. A set of measures of centrality based on betweenness. *Sociometry* 35–41 (1977).

33. Brosschot, J. F. Cognitive-emotional sensitization and somatic health complaints. *Scand. J. Psychol.* **43**, (2002).

34. Zhang, Y., Kim, J.-H., Brang, D. & Liu, Z. Naturalistic Stimuli: A Paradigm for Multi-Scale Functional Characterization of the Human Brain. *Curr. Opin. Biomed. Eng.* **19**, (2021).

35. Towards clinical applications of movie fMRI. *NeuroImage* **217**, 116860 (2020).

36. Li, X., Eickhoff, S. B. & Weis, S. Stimulus selection influences prediction of individual phenotypes in naturalistic conditions. *bioRxiv* 2023.12.07.570273 (2023) doi:10.1101/2023.12.07.570273.

37. Chang, L. J. *et al.* Endogenous variation in ventromedial prefrontal cortex state dynamics during naturalistic viewing reflects affective experience. *Sci. Adv.* (2021) doi:10.1126/sciadv.abf7129.

38. Morgenroth, E. *et al.* Probing neurodynamics of experienced emotions—a Hitchhiker's guide to film fMRI. *Soc. Cogn. Affect. Neurosci.* **18**, (2023).

39. Howard, G. S. Response-Shift Bias. *Eval. Rev.* (1980) doi:10.1177/0193841X8000400105.

40. Examining the relationship between measures of autistic traits and neural synchrony during movies in children with and without autism. *NeuroImage Clin.* **28**, 102477 (2020).

41. Relationships between depressive symptoms and brain responses during emotional movie viewing emerge in adolescence. *NeuroImage* **216**, 116217 (2020).

42. Individualized psychiatric imaging based on inter-subject neural synchronization in movie watching. *NeuroImage* **216**, 116227 (2020).

43. Schneider, S., Lee, J. H. & Mathis, M. W. Learnable latent embeddings for joint behavioural and neural analysis. *Nature* **617**, 360–368 (2023).

44. Camacho, M. C. *et al.* Large-scale encoding of emotion concepts becomes increasingly similar between individuals from childhood to adolescence. *Nat. Neurosci.* **26**, 1256–1266 (2023).

45. Nummenmaa, L. *et al.* Emotions promote social interaction by synchronizing brain activity across individuals. (2012) doi:10.1073/pnas.1206095109.

46. Dc, V. E. *et al.* The WU-Minn Human Connectome Project: an overview. *NeuroImage* **80**, (2013).

47. Dc, V. E. *et al.* The Human Connectome Project: a data acquisition perspective. *NeuroImage* **62**, (2012).

48. Mf, G. *et al.* The minimal preprocessing pipelines for the Human Connectome Project. *NeuroImage* **80**, (2013).

49. Alexander, L. M. *et al.* An open resource for transdiagnostic research in pediatric mental health and learning disorders. *Sci. Data* **4**, 1–26 (2017).

50. Esteban, O. *et al.* fMRIPrep: a robust preprocessing pipeline for functional MRI. *Nat. Methods* **16**, 111–116 (2019).

51. A, S. *et al.* Local-Global Parcellation of the Human Cerebral Cortex from Intrinsic Functional Connectivity MRI. *Cereb. Cortex N. Y. N 1991* **28**, (2018).

52. Y, T., Ds, M., M, B. & A, Z. Topographic organization of the human subcortex unveiled with functional connectivity gradients. *Nat. Neurosci.* **23**, (2020).

53. Rl, B., Fm, K., A, C., Jc, D. & Bt, Y. The organization of the human cerebellum estimated by intrinsic functional connectivity. *J. Neurophysiol.* **106**, (2011).

54. Yeo, B. T. T. *et al.* The organization of the human cerebral cortex estimated by intrinsic functional connectivity. *J. Neurophysiol.* **106**, 1125 (2011).

55. A density-based algorithm for discovering clusters in large spatial databases with noise | Proceedings of the Second International Conference on Knowledge Discovery and Data Mining. *Guide Proceedings*

https://dl.acm.org/doi/10.5555/3001460.3001507 doi:10.5555/3001460.3001507.

56. Busch, E. L. *et al.* Multi-view manifold learning of human brain-state trajectories. *Nat. Comput. Sci.* **3**, 240–253 (2023).

57. Crouse, D. F. On implementing 2D rectangular assignment algorithms. *IEEE Trans. Aerosp. Electron. Syst.* **52**, 1679–1696 (2016).