APPENDIX A
QUESTIONS AND ANSWERS IN THE EXPERT INTERVIEW

To clarify the research problem of brain network comparison in our clinical and neuroscience context, we conducted pilot studies with three domain experts. The first expert was a doctor from the department of neurology in a hospital. He has both clinical and research responsibility and focuses on white matter diseases. The second expert was a neuroscientist whose doctoral thesis studied the analysis of white matter tracts from neuroimaging data. The third expert was a computer scientist whose research interests include brain network visualization and analysis. All three experts have 10+ years of experience in relevant areas of neuroscience or computer science. Below, we report on the responses of three experts to the research question.

A. Doctor

1. **Basic information:** *a neurology doctor with more than 10 years of experience in neurological diseases.*

2. **Familiarity with brain imaging and neurological diseases:** *more than 5 years of experience in analyzing MRI/DTI image, and the association between neurological diseases and brain imaging.*

3. **Questions and answers in the expert interview**

Q1: Is there any clinical evidence or literature on the association of brain fiber connections with neurological diseases?

A: Yes, mostly diseases that affect white matter (nerve fibers). Brain diseases include cortical involvement and white matter involvement. Alzheimer's disease (AD), for example, is predominantly cortical involvement, so most imaging studies have focused on measuring cortical atrophy. Brain tumors can affect nerve fiber deformation, so DTI is applied more in the study of brain tumors.

Q2: If the answer of Q1 is yes, which part of brain region is connected with what neurological diseases? How does it relate?

A: Tumor: evaluate whether the tumor compresses nerve fibers (benign) or directly invades nerve fibers (malignant).

Motor neurone disease: Motor neurone disease mainly affects the pyramidal tract. DTI observed the involvement of pyramidal tract.

Multiple sclerosis (MS) : White matter involvement.

AD: It may affect some fiber connections. Please refer to the literature for details.

Cerebral infarction: Wallerian degeneration was assessed.

Development: A child is developing myelin, the insulating layer that surrounds nerve fibers in the brain.

Q3: Is it possible that the geometric features of the brain fibers mentioned above have medical significance? Is there a link to neurological disease? Do you have relevant literature or clinical evidence?

A: Yes, they all have.

Q4: Existing work suggests that the above characteristics may predict uncertainty of brain fiber data generated during imaging (is it actually the uncertainty of the method of brain fiber generation based on brain images?). Do you agree or disagree with the results of the calculation of the strength of connections between brain regions?

A: How to calculate? DTI mainly looks at two diagrams, ADC and FA, and can also reconstruct fiber bundle deformation. It is generally evaluated by these parameters. As to uncertainty, can we compare patients with normal controls to assess whether it's uncertainty or whether it's related to the disease?

Q5: If a geometric feature is found to be highly correlated with disease in subsequent studies, do you think it is more likely to be due to its (clinical) medical significance or imaging uncertainty (the uncertainty of the analysis method)?

A: If it does differ from a normal control, and the geometry can be explained by some mechanism, it is still considered medically relevant.

Q6: Could it be medically meaningful to study the areas of white matter (or gray matter) through which brain fibers travel? Or is it just the brain regions where the fibers start and end that matter?

A: DTI mainly reflects the condition of nerve fibers. Through the white matter/gray matter area? Do you want to express the nucleus through the basal ganglia?

Q7: Are there any other medically meaningful grouping methods (e.g. based on the above characteristics) other than by disease, age and gender?

A: The control group should be matched for age and gender. The disease itself can also be classified, such as the severity of the disease and whether the disease has subtypes (motor neurone disease, for example, has several subtypes, some involving the limbs first, others involving swallowing and breathing first).

Q8: How do medical experts/scholars compare the differences and similarities of brain fiber tracts between two samples?

A: We generally looked at the ADC and FA measures on the graph. The fiber bundle reconstruction is also a most intuitive way. You can try to conduct statistical analysis on these measure, (because I'm not a data scientist), I haven't done these data analyses before.

Q11: In addition to healthy controls and patients, is there a connection between brain fiber tracts and early/late cognitive impairment (MCI) in intermediate-stage subjects (such as in AD)?

A: This analysis can be done in conjunction with clinical practices, such as the degree of dementia. In addition, gray matter conditions (cortical atrophy measurement, hippocampal volume measurement) can also be combined for analysis.

Q12: If the brain connectivity bio-markers associated with neurological diseases are identified, how can they be applied clinically?

A: 1. Diagnosis of disease by whether there are specific markers of diseases. However, it may not be enough to compare AD and normal control alone. Other types of dementia, such as frontotemporal dementia and vascular dementia, can be added to see if there are markers specific to AD. (Note:

different types of dementia are sometimes clinically difficult to distinguish.)

2. Assessment of illness and prognosis because prognosis is known information.

B. Neuroscientist

1. **Basic information:** *a neuroscientist with more than 10 years of work experience.*

2. **Familiarity with brain imaging and neurological diseases:** *more than 5 years of experience in analyzing MRI/DTI images, neurological diseases, comparative study of brain imaging, and research on the association between neurological diseases and brain imaging.*

3. **Questions and answers in the expert interview:**

Q1: Is there any clinical evidence or literature on the association of brain fiber connections with neurological diseases?

A: Yes.

Q2: If the answer of Q1 is yes, which part of brain region is connected with what neurological diseases? How does it relate?

A: Whole brain connections are associated with Alzheimer's disease. By investigating the changes of various parameters derived from DTI in white matter, the changes of whole brain connectivity caused by Alzheimer's disease can be investigated.

Q3: Is it possible that the geometric features of the brain fibers mentioned above have medical significance? Is there a link to neurological disease? Do you have relevant literature or clinical evidence?

A: Yes. Please refer to Y. Jin, et al., "3D Retract -specific local and global analysis of White matter Integrity in Alzheimer's Disease," Human Brain Mapping 38(3), 1191–1207, 2017. This paper studied the correlation between Brain fiber connectivity and Alzheimer's disease.

Q4: Existing work suggests that the above characteristics may predict uncertainty of brain fiber data generated during imaging (is it actually the uncertainty of the method of brain fiber generation based on brain images?). Do you agree or disagree with the results of the calculation of the strength of connections between brain regions?

A: Yes.

Q5: If a geometric feature is found to be highly correlated with disease in subsequent studies, do you think it is more likely to be due to its (clinical) medical significance or imaging uncertainty (the uncertainty of the analysis method)?

A: It needs to be analyzed on a case-by-case basis.

Q6: Are there any other medically meaningful grouping methods (e.g. based on the above characteristics) other than by disease, age and gender?

A: Yes. For example, by whether a gene has a mutation or not.

Q7: How do medical experts/scholars compare the differences and similarities of brain fiber tracts between two samples?

A: Statistical analysis (ANOVA, t-test) was used to compare the differences between the groups of the extracted features. MD (ADC) and FA are key measures for the study of brain network connectivity difference. You can read my paper for more details on the previous study: Y. Jin, et al., "3D Retract -specific local and global analysis of White matter Integrity in Alzheimer's Disease", Human Brain Mapping 38(3), 1191-1207, 2017.

Q8: If the cerebral fiber plexus is mapped against a set of samples (multi-brain images), does the following description have medical significance?

A) The density of fibers passing through an area

B) The fiber orientation distribution of different samples in a certain region (whether the fiber orientation of different people in this region is the same)

C) Shape distribution of different samples in a certain region (whether the fibers have similar shapes in this region for different people)

A: Very likely, but further research is needed.

Q9: In addition to healthy control, is there a connection between brain fiber connections and early/late cognitive impairment (MCI) in intermediate stage patients such as Alzheimer's disease?

A: Yes, please refer to Y. Jin, et al., "3D Retract -specific local and global analysis of White matter Integrity in Alzheimer's Disease", Human Brain Mapping 38(3), 1191-1207, 2017.

Q10: If the brain connection regions associated with neurological diseases are identified through data analysis, how will they be used clinically?

A: Because imaging is non-invasive, regular follow-ups can be done to extract brain connections as a bio-marker for early diagnosis of Alzheimer's disease.

C. Computer Scientist

1. **Basic information:** *a computer scientist with more than 10 years of work experience.*

2. **Familiarity with brain imaging and neurological diseases:** *2–5 years of experience in MRI/DTI image, comparative study of brain imaging, and research on the association between neurological diseases and brain imaging.*

3. **Questions and answers in the expert interview:**

Q1: Is there any clinical evidence or literature on the association of brain fiber connections with neurological diseases?

A: Yes. The brain network study for disease early detection and diagnosis have been a while in the research community of computational neuroscience.

Q2: If the answer of Q1 is yes, which part of brain region is connected with what neurological diseases? How does it relate?

A: We have previously surveyed related literature on the damage of brain white matter in neurological diseases. Dating back to more than 100 years ago, researchers have found that disconnection syndromes can lead to aphasia, e.g. the fiber tracts between Broca's and Wernicke'area. Recently, more research were conducted on the correlation between the change of brain networks and other diseases such as alzheimer's

disease, schizophrenia, etc. AD patients tends to have certain break-downs in their brain networks.

Q3: Is it possible that the geometric features of the brain fibers mentioned above have medical significance? Is there a link to neurological disease? Do you have relevant literature or clinical evidence?

A: As we are computer scientist, we do not know.

Q4: Existing work suggests that the above characteristics may predict uncertainty of brain fiber data generated during imaging (is it actually the uncertainty of the method of brain fiber generation based on brain images?). Do you agree or disagree with the results of the calculation of the strength of connections between brain regions?

A: Yes, we have previously seen much variation on the same connection of different subjects. The same pattern happens in several data sets acquired from multiple sources.

Q5: If a geometric feature is found to be highly correlated with disease in subsequent studies, do you think it is more likely to be due to its (clinical) medical significance or imaging uncertainty (the uncertainty of the analysis method)?

A: As we are computer scientist, we do not know.

Q6: Are there any other medically meaningful grouping methods (e.g. based on the above characteristics) other than by disease, age and gender?

A: We have seen in the public and private data sets some other attributes. For example, IQ measure, years of education, weight, etc. Also, some data sets are longitudinal study. The year from the onset of AD or other diseases may be a useful criterion for comparison.

Q7: How do medical experts/scholars compare the differences and similarities of brain fiber tracts between two samples?

A: As a computer scientist, I have compared brain networks, but have not tried the comparison of a single brain fiber. Visualization of both fibers in the same interface might be a solution.

Q8: If the cerebral fiber plexus is mapped against a set of samples (multi-brain images), does the following description have medical significance?

A) The density of fibers passing through an area

B) The fiber orientation distribution of different samples in a certain region (whether the fiber orientation of different people in this region is the same)

C) Shape distribution of different samples in a certain region (whether the fibers have similar shapes in this region for different people)

A: As we are computer scientist, we do not know.

Q9: In addition to healthy control, is there a connection between brain fiber connections and early/late cognitive impairment (MCI) in intermediate stage patients such as Alzheimer's disease?

A: Yes, as we can tell from the data, MCI subjects have networks between AD patients and controls.

Q10: If the brain connection regions associated with neurological diseases are identified through data analysis, how will they be used clinically?

A: As a computer scientist, I can only guess. I think the biomarkers (trends in the longitudinal study or the absolute degree of damage) can be used as either an early sign or indicator of severity for the disease.

APPENDIX B
LAYOUT ALGORITHM OF BRAIN NECKLACE VISUALIZATION

A key issue of brain necklace visualization in the composite view is the layout of feature pearls as they can heavily overlap with each other when there are a large number of edges in the composite view. We propose a distributed layout algorithm that can be computed very fast to meet the interactivity requirement in online visualization. The algorithm is based on a principle of cost minimization in placing feature pearls. In details, the cost contains two components: the overlapping cost (*cost_over*) which describes the degree of a pearl overlapping with other edges/pearls, and the spacing cost (*cost_space*) which describes the compactness of pearl placement from its previous pearl on the same edge. The costs are defined by

$$cost\_over(d_j) = \begin{cases} 0 & d_j \geq d_{max} \\ 1 & d_j < d_{max} \end{cases}, cost\_space(l_j) = (l_j - l_{min})/l_{min} \tag{1}$$

d_j denotes the perpendicular distance from the jth pearl (MD feature in the figure) to the closest edge other than its current edge. When this distance is large enough, i.e., small probability of overlapping, the cost is zero; otherwise the cost is one. On the other hand, the spacing cost of the jth pearl increases linearly with the interval from the previous pearl on the same edge, denoted as l_j, excluding the minimal interval l_{min} between adjacent pearls to avoid overlaps.

The layout algorithm tries to optimize a weighted sum of the two cost components on all n pearls of an edge.

$$Minimize \quad \sum_{j=1}^{n} [(1-\alpha) \cdot cost\_over(d_j) + \alpha \cdot cost\_space(l_j)] \tag{2}$$

where α is the parameter to balance the two types of cost.

As there are infinite layout positions on an edge, we design a feasible algorithm by discretizing layout solutions. We assume the pearls can only be placed in m candidate positions with a fixed interval of l_{min} from each other where $m = \lfloor L/l_{min} \rfloor$ and L is the length of the edge. The algorithm exploits a pattern in the spacing cost that the cost of all pearls in an edge only depends on the position of the last pearl, which is described by

$$\sum_{j=1}^{n} cost\_space(l_j) = \sum_{j=1}^{n} l_j/l_{min} - n \tag{3}$$

Knowing that the overlapping cost of each candidate position can be pre-computed, the algorithm becomes an iteration of the position of the last pearl on the edge. For each feasible position, the overlapping cost of all pearls as well as their overall cost can be directly computed. The pseudocode of the algorithm is given in Algorithm 1. Except for the pre-computing of overlapping costs, the algorithm has a linear

complexity of $O(m-n)$ for each edge. Note that when an edge can not admit n features ($m < n$), the pearls are drawn as hollow circles in the minimal size and placed uniformly on the edge.

To focus on one ROI in the composite view, users can click the corresponding node to highlight all connections of the ROI, as well as feature pearls on these edges. The other unselected edges will fade out in the background. When there are many nodes/edges in the composite view, the distributed layout result can also include severe visual clutter caused by overlapping. To reveal feature pattern clearly on the selected ROI, we introduce a distortion method to the edges connected to the selected ROI. The basic observation is that a majority of visual clutter are caused by overly small angles between the adjacent edges or overly close adjacent ROIs to the selected ROI on the other endpoint of the edge. The distortion method enforce a minimal angle between adjacent edges connecting to the same ROI and a minimal length for each edge, so that all the feature pearls can be spaced uniformly for effective comparison. The distorted edges will be connected back to the original destination through B*é*zier curves. The selected visualization result by the distortion method in comparison to the original distributed layout result is illustrated in Figure 1. The method considerably reduces the visual clutter and improves the efficiency of pattern discovery from our visualization.

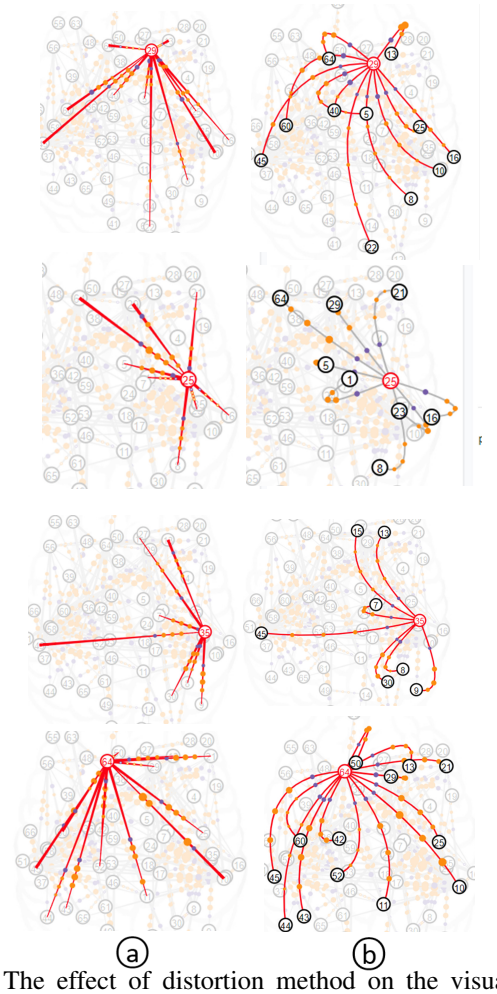

Fig. 1: The effect of distortion method on the visualization of selected ROI edges: (a) undistorted picture; (b) distorted picture.

APPENDIX C
NOTES ON FEATURE QUALITY FOR BRAIN NETWORK COMPARISON

The comparison among brain networks of separate subject groups generally focuses on individual or a set of brain connections which differ significantly across the groups of networks. The differences can take place on geometric features of the brain connection (length, curvature, entropy, etc.), diffusion features (FA, MD, RD, etc.), or both. After these features are measured, extracted, and computed through the neuroimaging process, the brain network comparison problem is mostly treated as a statistical or machine learning problem: how to identify the most different connectivity features from all network connections among subject groups under comparison?

The formulation of this statistical problem without incorporating the domain knowledge of human brain network analysis can be incomplete or even misleading. First, the human brain is known to be densely interconnected by brain white matter. Yet, it is not a fully connected network. As reported in Ref. [1], their structural brain networks defined on 70 ROIs have a density of 0.2$\sim$0.45, which are constructed over the same data with the work here. More than a half of connections under the current brain network definition are null. This also corresponds well with the visualization result in Figure 5 (same with Figure 4(a)(b) in the main paper), where more than a half of connectivity features in the subject $\times$ connectivity matrix are invalid (grey color). Second, the measurement of brain connectivity features can be uncertain caused by quite a few factors. Most notably, the same node (ROI) of the 70-ROI brain network can refer to quite different cerebral cortex region in separate subjects. This is because the structure of cerebral cortex can be at least slightly diversified across subjects. The registration of an individual subject into the same brain atlas often leads to deviated cortex-node mapping from the template brain cortex. Meanwhile, other factors such as the parameter and method choice of fiber tracing algorithms in brain tractography also add to the uncertainty of measured brain connectivity features.

The domain characteristics on human brain networks bring extra complexity to the network comparison problem studied here. In theory, it is assumed that, on each connectivity feature, the two subject groups under comparison will have two feature value distributions. When the two distributions are significantly different as identified by statistical tests, the underlying connectivity feature will be a potential bio-marker that distinguishes the two subject groups under comparison. Nevertheless in practice, this comparison process works for some but not all connectivity features. Take the diffusion feature of FA as an example, as shown in Figure 2, there are at least two types of cases. In the first type, a representative feature value distribution is given in Figure 2(a) where null

Algorithm 1: The Pearl Layout Algorithm on an Edge.

Input : e (the edge to layout), n (#pearls on an edge), E (the set of edges on the network), d_{max} (max. pearl size), l_{min} (min. pearl interval), α (cost parameter)

Output: P (position of all pearls on the edge)

begin
- $m = \lfloor Length(e)/l_{min} \rfloor$
- **if** $m < n$ **then**
  - /* use the uniform-interval layout on short edges */
  - **return** uniform_layout(e, n)
- /* pre-compute overlapping cost on each candidate position */
- **for** $j \leftarrow [1, m]$ **do**
  - $p_j = start(e) + j \cdot \frac{l_{min}}{Length(e)} \cdot \hat{e}$
  - **for** *remote_edge in* $E \backslash e$ **do**
    - $d_j = min(d_j, distance(p_j, remote\_edge))$
  - **if** $d_j < d_{max}$ **then**
    - $c_j = 1$
  - **else**
    - $c_j = 0$
- /* iterate over feasible positions of the last pearl */
- **for** $i \leftarrow [n, m]$ **do**
  - $cost_s = i - n$
  - $cost_o = n - 1 - min(\sum_{j=1}^{i-1}(1 - c_j), n-1) + c_i$
  - **if** $cost_{all} > (1-\alpha) \cdot cost_o + \alpha \cdot cost_s$ **then**
    - $cost_{all} = (1-\alpha) \cdot cost_o + \alpha \cdot cost_s$, $last\_pos = i$
- /* compute and return pearl positions with smallest possible overall cost given the position of the last pearl */
- **return** best_position(n, $last\_pos$, $\{c_j\}_{j=1}^{last\_pos}$)

features predominate the distribution. This indicates that the corresponding edge of the brain network is without fiber connection anatomically. This and similar features should clearly be discarded before the comparison. The first type of connectivity feature is made complicated in the case of Figure 2(b) where a small portion of feature values on the same connection are non-zero due to the uncertainty of measurement. Probably, we should not compare on these connectivity features as zero values still dominate the distribution. In the second type of cases, there are physical fiber connection on the measured brain network edge. A typical feature distribution on FA is given in Figure 2(c) which largely follows normal distribution as indicated by the Kolmogorov-Smirnov normality test. This is the ideal feature distribution for group-level comparison. Similarly, complexity shows up when the same connection is not detected on some subjects, or the feature value is unusually low (Figure 2(d)). These abnormal subjects should be removed in the comparison of the connectivity feature. Figure 3 depicts typical value distributions of the geometric feature of Entropy on the same data. The classification of two types of cases and their variations are similar.

The above analysis shows that, because of the existence of null connection in the brain network and the occasional measurement deviation, there is no clear-cut answer on which connectivity feature should be included in the brain network comparison. We propose an interactive method to allow users, who are mostly domain experts, to decide on the scope of feature comparison for bio-marker detection, based on computational evidence from the feature analysis. In the first step of our method, we define a measure of feature quality for brain network comparison, namely Quality for Comparison (QoC) in short. The zero feature values are assigned the zero quality, as we will not compare null connection features[1] or mis-measured features. For non-zero feature values, those features statistically deviated from the feature value distribution are assigned low qualities. In the second step of the method, we provide an online QoC filter whose threshold is set by users manually. The feature values below the quality threshold will be removed from the comparison. Note that, we provide another brain connection filter which sets the minimal percentage of remaining subjects for the valid comparison of underlying brain connection. For example, when the percentage threshold is set to 50%, the connectivity features having low quality in more than a half subjects will be excluded from the network comparison. This allows to distinguish between null connection cases and valid connection cases as shown in Figure 2(b)(d). The interactive data wrangling design is also described in Section 5.2 of the main paper (feature heatmaps).

According to the definition of QoC measure, we apply anomaly detection algorithms to assign zero quality to outliers in the feature value distribution, and set largely deviated feature values to low quality. We have considered three anomaly detection algorithms. The cluster-based anomaly detection algorithm by DBSCAN is first excluded due to the high computational complexity and the high sensitivity to parameters (i.e., minPts and neighborhood radius). The two statistical anomaly detection algorithms, Grubbs's test and Extreme Value Theory (EVT), can both isolate outliers and compute quality measures for the others very fast. We choose the Grubbs's test as it is parameter-free. Also, the precondition for the Grubbs's test holds well. For all connectivity features, the value distributions largely follow normal distribution after removing zero values. Figure 4 gives the distribution of p-values from the Kolmogorov-Smirnov normality test on representative geometric and diffusion features. To focus on valid brain connections, the features having more than a half values being zero are excluded from the KS normality test. From the figure, it can be found that more than 96% features on valid connections have p-values above 0.05, i.e., follow normal distribution after removing zero values. In fact, under appropriate parameters, the set of outliers detected by EVT overlap significantly with the output of Grubbs's test. The overlapping ratio is as high as 90% on all features. In this

[1]In this work, we only consider the neurological diseases or pathological development on the brain network that change the network connectivity. They do not thoroughly remove the connectivity. For applications that demand zero vs. non-zero feature comparison, the definition of QoC should be reconsidered.

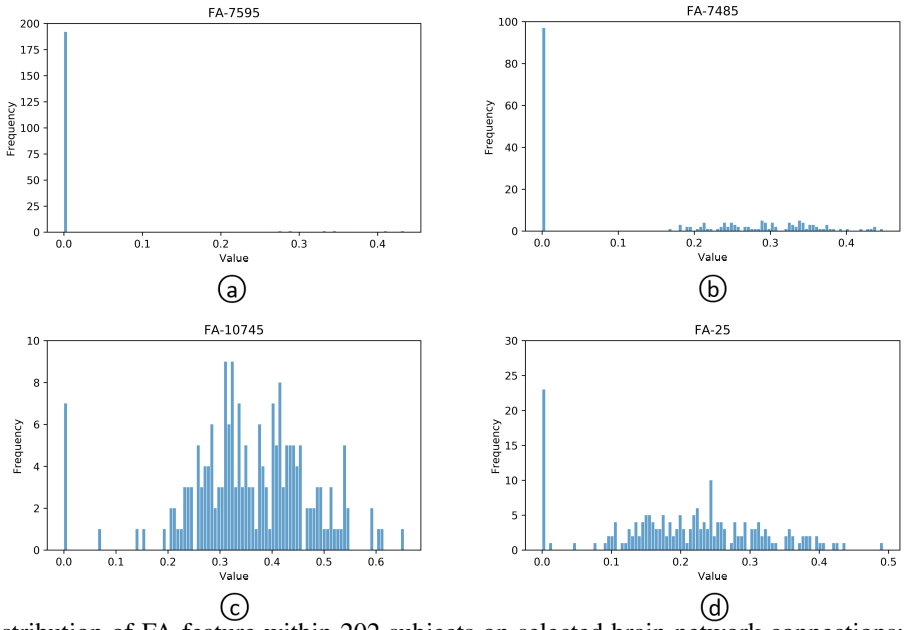

Fig. 2: The value distribution of FA feature within 202 subjects on selected brain network connections: (a) the null connection with almost all zero values; (b) the probable null connection with a few nonzero feature values; (c) the valid connection with its feature value distribution close to the normal distribution; (d) the valid connection with more zero feature values, probably due to measurement deviations.

sense, the EVT algorithm can also be applied. Changing its algorithm parameter actually re-scales the distribution of the quality measure. The users can re-adjust their quality filter to achieve the same filtering outcome.

To visually explain the QoC measure to users and illustrate the necessity to filter low-quality features, we also design a detailed quality visualization panel on the heatmap view of MV^{2}Net. As shown in Figure 5, by mouse hovering of any cell on the heatmap, the quality panel is shown on-demand at the upper-right corner of each heatmap view. In the panel, the probability density distribution (PDF) of all feature values within the hovered cell is displayed in green bar charts (Figure 5(a)). In the background of the quality panel, the average PDF of the same connectivity features on two comparing subject groups are drawn in purple and orange line charts respectively. They correspond to the upper purple part and the lower orange part in the same column with the hovered cell. The shaded purple/orange contours centered on the purple/orange line charts represent the variation (75% CI) of feature distributions across all subjects in the same group. The visualization result in Figure 5(b) on a low-quality cell (grey color) indicates that the distribution of low-quality feature values, as shown by the green bar charts close to zero, is significantly different from the high-quality features in both subject groups, as shown by the line charts. The difference between the high vs. low quality features is much larger than the difference between the features in comparing subject groups.

REFERENCES

[1] X. Yang, L. Shi, M. Daianu, H. Tong, Q. Liu, and P. Thompson, "Blockwise human brain network visual comparison using nodetrix representation," *IEEE Transactions on Visualization and Computer Graphics*, vol. 23, no. 1, pp. 181–190, 2017.

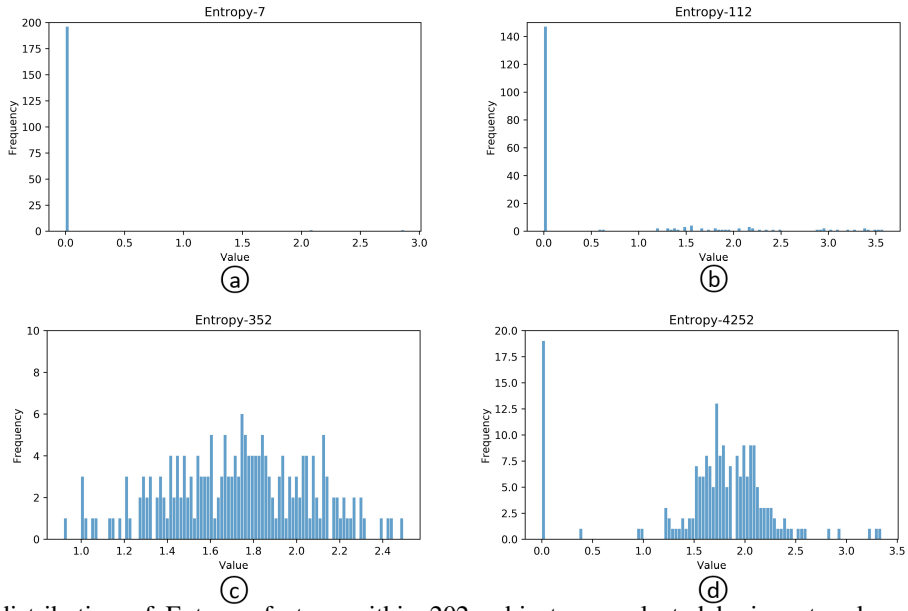

Fig. 3: The value distribution of Entropy feature within 202 subjects on selected brain network connections: (a) the null connection with almost all zero values; (b) the probable null connection with a few nonzero feature values; (c) the valid connection with its feature value distribution close to the normal distribution; (d) the valid connection with more zero feature values, probably due to measurement deviations.

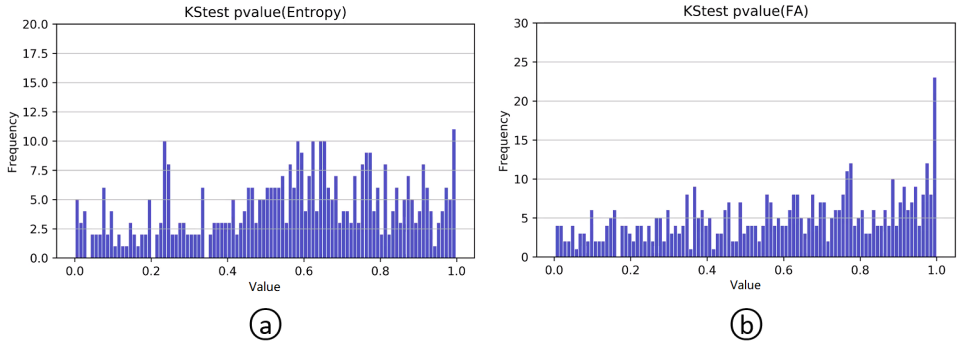

Fig. 4: The distribution of Kolmogorov-Smirnov normality test's p-value on certain feature distribution in 202 subjects. Zero feature values are removed, features with more than a half being zero values are excluded. (a) the entropy feature (geometric); (c) the FA feature (diffusion).

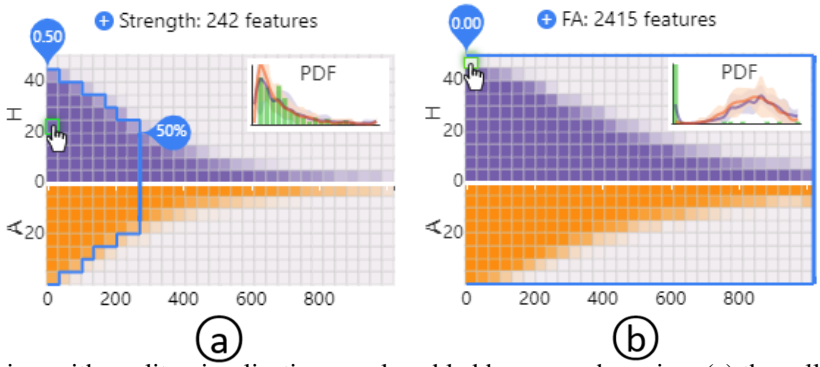

Fig. 5: Feature heatmap view with quality visualization panel enabled by mouse hovering: (a) the cell with high quality features (purple color) where the feature value distribution within the cell as shown by green bar charts is similar to the distribution in the same column (purple and orange line charts); (b) the cell with low quality features (grey color) where the feature value distribution as shown by green bar charts peaks at zero value and is quite different from the high quality features in the same column (purple and orange line charts).