**Supplemental Text S5. Frequent subgraph mining and feature selection methods for the certain network.**

Definition 1(Certain graph) Certain graph is an undirected and deterministic graph, which is represented as a set of nodes and a set of deterministic edges .

Definition 2 (frequency) is a certain graph and the number of subgraphs is n, where the frequency of subgraphs is defined as the following formula:



**Method for mining frequent subgraphs (gSpan)**

The detailed algorithm is shown in Table 1.

|  |
| --- |
| Input: |
| certain network set  and support degree . |
| 1. A lexicographic order among graphs is constructed by using the gSpan algorithm *s* and employing a unique minimum depth-first search (DFS) code as the canonical label. |
| 2. Judgment of whether the k-subgraph candidate mode is a frequent subgraph. From an initial one-edge frequent subgraph mode, gradually increase the number of edges and determine if this mode is a frequent subgraph. |
| 3. If the k+1-subgraph is non-frequent, delete it from the candidate subgraph. |
| Output: |
| the set of frequent subgraphs. |

**Frequent-scoring feature selection.**

The discriminant subgraph features can be used for classification. However, the subgraph obtained by the gSpan algorithm is not discriminative. In particular, some frequent subgraphs have weak discriminative ability, and the resulting classification performance is not good. Therefore, it is necessary to define the discriminant and use it as an index to select the most discriminative subgraphs as features. This method is called frequent-scoring feature selection.

Given the positive sample set  and the negative sample set , the discrimination score is defined as



the difference between its positive frequency and negative frequency. A larger score reflects a larger difference between the two groups.

In this paper, this method involved choosing the same number of frequent subgraphs from both the NC and the MDD groups, calculating and sorting the discrimination scores of the frequent subgraphs, and then selecting top *k* subgraphs. As a result, a total of *2k* subgraphs were selected.