Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Optimization of Community-Consideration Centrality for Breast Cancer Comorbidity Using Various Community Detection Algorithms

Gilbert Evan Tilung
University of Multimedia Nusantara
Angga Aditya Permana
angga.permana@umn.ac.id
University of Multimedia Nusantara

Research Article

Keywords: Breast Cancer, Community Detection, Comorbidity, Optimization

Posted Date: October 26th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11316239/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Optimization of Community-Consideration Centrality for Breast Cancer Comorbidity Using Various Community Detection Algorithms

Gilbert Evan Tilung[1] and Angga Aditya Permana[2*†]

[1,2]Informatics, University of Multimedia Nusantara, Tangerang, 15811, Tangerang, Indonesia.

*Corresponding author(s). E-mail(s): angga.permana@umn.ac.id;
Contributing authors: gilbert.evan@student.umn.ac.id;
†These authors contributed equally to this work.

Abstract

Breast cancer is one of the most prevalent diseases worldwide and is often accompanied by various comorbidities. Comorbidity network analysis can be used to identify diseases that play important roles within the network; however, conventional centrality measures do not consider the underlying community structure. This study aims to optimize the Community-Consideration Centrality (CCC) method by comparing several community detection algorithms and evaluating the effect of different $\boldsymbol{\alpha}$ parameter values. A comorbidity network was constructed using Disease Ontology Identifier (DOID) data and partitioned using the Leiden, Girvan-Newman, Belief Propagation, and Spinglass algorithms. Community-Consideration Degree, Closeness, and Betweenness Centrality were then calculated. The results show that the Leiden algorithm achieved the highest modularity, indicating the best community structure. The CCC method consistently identified central diseases, including DOID.863, DOID.77, DOID.331, and DOID.150. Furthermore, Jackknife evaluation demonstrated that $\boldsymbol{\alpha = 0.9}$ provided the most stable performance. These findings indicate that integrating community detection with CCC is effective for identifying important diseases in breast cancer comorbidity networks.

Keywords: Breast Cancer, Community Detection, Comorbidity, Optimization

1 Introduction

Breast cancer is one of the most significant global health challenges and continues to be a major concern in cancer control [1, 2]. According to the *Global Cancer Statistics 2020*, breast cancer surpassed lung cancer as the most commonly diagnosed cancer worldwide, accounting for approximately 2.3 million new cases or 11.7% of all newly diagnosed cancer cases in that year [3]. This high incidence is also reflected in many countries, including the United States, where more than 250,000 new cases of invasive breast cancer are reported annually [1]. In addition to its high incidence, breast cancer remains one of the leading causes of cancer-related mortality among women worldwide [1, 2]. Although advances in early detection and treatment have significantly improved patient survival, disparities in healthcare access, delayed diagnosis, and limited availability of effective treatment continue to contribute to high mortality rates, particularly in developing countries [2]. These conditions demonstrate that breast cancer is not only an individual health issue but also a substantial global health burden requiring continuous attention and intervention [2–4].

In addition to the characteristics of the cancer itself, comorbidities represent another important factor influencing disease progression in patients with breast cancer. Numerous studies have shown that breast cancer patients frequently experience comorbid conditions such as hypertension, diabetes mellitus, coronary heart disease, chronic respiratory diseases, and stroke [5]. In a prospective cohort study involving 4,664 breast cancer patients, hypertension was identified as the most prevalent comorbidity (22.4%), followed by chronic gastritis (14.3%), diabetes mellitus (6.2%), chronic bronchitis or asthma (5.8%), coronary heart disease (5.0%), and stroke (2.2%) [5]. These comorbid conditions have been shown to influence clinical outcomes, with diabetes being associated with increased mortality risk, while a history of stroke is linked to a higher risk of non–breast cancer-related mortality [5]. Similar findings were reported in a study of breast cancer survivors in Australia, where 88.7% of participants had at least one additional chronic health condition [6]. Beyond increasing the complexity of patient management, a high burden of comorbidities has also been associated with an elevated risk of developing second primary cancers among breast cancer survivors [7]. Therefore, understanding comorbidities and the relationships among them is essential for improving prognosis, quality of life, and the overall effectiveness of breast cancer management [5–7].

Comorbidity refers to the presence of one or more additional diseases or medical conditions occurring alongside the primary disease that is the main focus of clinical attention [8, 9]. The presence of comorbidities has become an important concern in healthcare because it affects multiple aspects of disease management, including diagnosis, treatment selection, and patient outcomes [9]. Among cancer patients, comorbidities increase treatment complexity, limit available therapeutic options, hinder the completion of optimal treatment regimens, and negatively affect prognosis and overall survival [9]. Furthermore, previous studies have demonstrated that comorbidities are associated with increased mortality, reduced quality of life, and greater healthcare utilization [5, 6]. The high prevalence of comorbidities among cancer

patients indicates that this is not an uncommon phenomenon but rather a critical factor that should be considered when understanding disease progression and improving patient management [6, 10].

Although previous studies have successfully identified various comorbid diseases associated with breast cancer and their prevalence, most have relied on conventional statistical approaches that are limited to identifying pairwise relationships between diseases [11]. Such approaches cannot adequately control for potential confounding diseases, making the identified associations susceptible to being indirect or superficial [12]. Moreover, conventional methods fail to provide a comprehensive understanding of the complex relationships among multiple comorbidities simultaneously [13]. To address these limitations, network analysis has become increasingly popular in comorbidity research. In this approach, diseases are represented as nodes, while associations between diseases are represented as edges, with edge weights reflecting the strength of their associations [13]. Network analysis has been successfully applied to various diseases, including cardiovascular disease, heart failure, type 2 diabetes, depression, and chronic obstructive pulmonary disease (COPD), revealing complex relationships that cannot be detected using traditional statistical methods [11, 14]. Furthermore, network analysis enables the identification of central nodes that may serve as effective intervention targets, as well as bridge symptoms that contribute to the progression of comorbidities [15].

Within comorbidity network analysis, two complementary concepts are community detection and centrality analysis. Community detection aims to identify groups of densely connected nodes (diseases), thereby revealing clusters of diseases that frequently co-occur or share common underlying risk factors in clinical practice [11, 16]. Several algorithms have been developed for community detection, including *Leiden*, *Spinglass*, *Girvan-Newman*, and *Belief* [16–20]. Meanwhile, centrality measures evaluate the importance or influence of individual nodes within the network. Three commonly used centrality measures are *degree centrality*, *closeness centrality*, and *betweenness centrality*. Degree centrality quantifies the number of direct connections a node has, reflecting how actively a disease is associated with other diseases. Closeness centrality measures how close a node is to all other nodes in the network, indicating how rapidly a disease can influence other diseases through the shortest paths. Betweenness centrality measures how frequently a node lies on the shortest paths connecting other nodes, reflecting its role as a bridge facilitating interactions between comorbid diseases [14, 15, 21]. In the *Community-Consideration Centrality* (CCC) method proposed by [16], these traditional centrality measures are extended by incorporating community structure through the weighting parameter α, which balances the importance of community-level information and the overall network structure.

Although the *Community-Consideration Centrality* (CCC) method has been proposed as an extension of traditional centrality measures that incorporates community structure [16], its application remains limited. Previous research evaluated the method using a lung cancer *Protein-Protein Interaction* (PPI) network, where nodes represent proteins and edges represent biological interactions [16]. This network structure differs fundamentally from disease comorbidity networks, in which nodes represent diseases

and edges represent clinical associations between diseases. Consequently, the applicability of CCC to disease comorbidity networks, particularly breast cancer comorbidity networks, requires further investigation.

Furthermore, the original study constructed community structures using the *Disassembly Greedy Modularity* (DGM) algorithm [16]. Meanwhile, various community detection algorithms, including *Leiden*, *Spinglass*, *Girvan-Newman*, and *Belief*, have been widely employed in network analysis. However, to the best of the author's knowledge, no previous study has systematically compared these algorithms for constructing community structures in breast cancer comorbidity networks to support the identification of central diseases using the CCC method. Consequently, the most suitable community detection algorithm for generating meaningful community structures in breast cancer comorbidity networks remains unknown.

Moreover, the CCC method utilizes the parameter α to balance the contribution of community-level and global network information [16]. While this parameter has been evaluated using lung cancer PPI networks, studies investigating the influence of different α values in disease comorbidity networks remain limited. In addition, research integrating the comparison of multiple community detection algorithms with the CCC method in *Disease Ontology Identifier* (DOID)-based breast cancer comorbidity networks is still scarce. Therefore, further research is required to compare different community detection algorithms, evaluate the effect of varying the parameter α, and identify central diseases in DOID-based breast cancer comorbidity networks using the CCC method.

2 Methodology

The overall research workflow is presented in Figure 1.

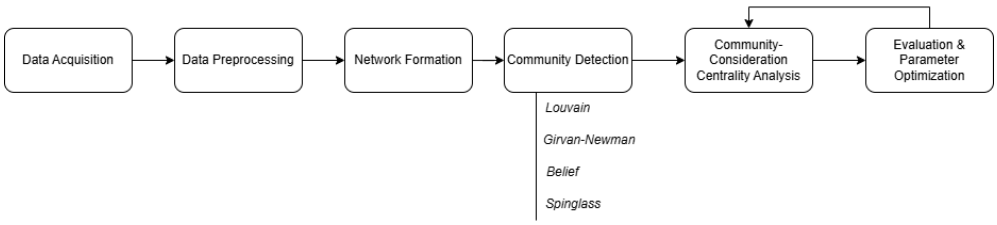


Fig. 1 Research framework of the proposed methodology.

The detail of each research steps as follow:

- **Data Acquisition.** The dataset used in this study was adopted from the breast cancer comorbidity network research conducted by [22]. The original dataset was constructed by mapping breast cancer comorbidities to their corresponding *Disease Ontology Identifiers* (DOIDs) and computing pairwise semantic similarity using the *Disease Ontology* framework. Among the available similarity matrices, the Wang

similarity matrix was selected because it provides a more stable similarity distribution and achieves a better balance between community modularity and false positive rates. The resulting similarity matrix serves as the primary dataset for constructing the breast cancer comorbidity network in this study.

- **Data Preprocessing.** The Wang similarity matrix is preprocessed by applying a similarity threshold of 0.5, where similarity values below the threshold are discarded to eliminate weak disease associations. Subsequently, the diagonal elements of the matrix are set to zero to remove self-loops, resulting in an adjacency matrix that serves as the basis for constructing the breast cancer comorbidity network.
- **Network Formation.** The preprocessed adjacency matrix is transformed into an undirected weighted comorbidity network, where each node represents a disease identified by its *Disease Ontology Identifier* (DOID), and each edge represents the semantic similarity between a pair of diseases. To ensure a connected and representative network for subsequent analyses, only the largest connected component is retained. The resulting network provides the structural foundation for community detection and centrality analyses.
- **Community Detection.** The constructed comorbidity network is partitioned into communities using four community detection algorithms: Leiden [17], Girvan–Newman [18], Belief Propagation [20], and Spinglass [19]. These algorithms employ different optimization strategies to identify groups of densely connected diseases within the network. Since the resulting community structures may vary depending on the algorithm, multiple community detection methods are evaluated to investigate their influence on the subsequent *Community-Consideration Centrality* analysis and to determine the most suitable community structure for identifying important diseases in the breast cancer comorbidity network.
- **Community-Consideration Centrality Analysis.** The community structures identified in the previous stage are incorporated into the *Community-Consideration Centrality* (CCC) framework proposed by Rustamaji *et al.* [16]. Unlike conventional centrality measures, which evaluate node importance solely based on the global network topology, CCC integrates both global and community-level information to identify diseases that are influential across the entire comorbidity network as well as within their respective communities. In this study, the CCC framework is applied to three centrality measures, namely degree, closeness, and betweenness centrality. These measures respectively evaluate the connectivity, accessibility, and bridging role of each disease while simultaneously accounting for both global and local community structures. The corresponding formulations are presented as follows.

$$CCDC(v) = \frac{d_C(v) + \alpha(d_G(v) - d_C(v))}{|c| + \alpha(|V| - |c|) - 1}, \tag{1}$$

$$CCCC(v) = \frac{|c| + \alpha(|V| - |c|) - 1}{d_C(v) + \alpha(d_G(v) - d_C(v))}, \tag{2}$$

$$CCBC(v) = \frac{BC_C(v) + \alpha\left(BC_G(v) - BC_C(v)\right)}{P_C + \alpha(P_G - P_C)}. \tag{3}$$

In the above equations, $d_G(v)$ and $d_C(v)$ denote the degree or total shortest-path distance of node v computed over the global network and its corresponding community, respectively, depending on the centrality measure being evaluated. Similarly, $BC_G(v)$ and $BC_C(v)$ represent the global and community betweenness centrality of node v. The terms $|V|$ and $|c|$ denote the total number of nodes in the network and the number of nodes within the corresponding community, respectively, while P_G and P_C represent the normalization factors for the global network and the local community. The balancing parameter α determines the relative influence of global and community-level information, where larger values emphasize the global network structure and smaller values emphasize the local community structure.

- **Performance Evaluation and Parameter Optimization**. To determine the most effective configuration of the proposed framework, the *Community-Consideration Centrality* method is evaluated using different values of the balancing parameter α in combination with each community detection algorithm. The resulting disease rankings are subsequently compared against known breast cancer hallmark genes to assess the ability of each configuration to identify biologically relevant diseases. This evaluation facilitates the selection of the optimal combination of community detection algorithm and parameter setting for breast cancer comorbidity network analysis.

3 Results

3.1 Data Acquisition

The dataset used in this study was adopted from the breast cancer comorbidity network developed by Permana *et al.*, in which breast cancer comorbidities were mapped to their corresponding *Disease Ontology Identifiers* (DOIDs) and pairwise semantic similarity was computed using the *DOSE* package. After data cleaning, 1,006 unique diseases were retained. The Wang similarity matrix was selected as the input dataset because it provides a stable similarity distribution and a favorable balance between community modularity and false positive rates.

3.2 Data Preprocessing

The Wang semantic similarity matrix containing 1,006 *Disease Ontology Identifiers* (DOIDs) was preprocessed by applying a similarity threshold of 0.5. Pairwise similarity values below this threshold were discarded to eliminate weak disease associations, while the diagonal elements were set to zero to remove self-loops. Following preprocessing, 192 DOIDs became isolated, indicating that they no longer shared semantic similarity of at least 0.5 with any other disease. Consequently, these isolated nodes were excluded from subsequent network analysis, resulting in an adjacency matrix containing 814 interconnected DOIDs.

Table **??** presents a representative portion of the resulting adjacency matrix after preprocessing. The retained similarity values represent disease pairs with semantic similarity of at least 0.5, whereas weaker associations have been removed. This adjacency

matrix serves as the basis for constructing the breast cancer comorbidity network used in the subsequent analyses.

Table 1 Summary of preprocessing results.

Metric	Value
Original DOIDs	1,006
Similarity threshold	0.5
Isolated DOIDs removed	192
Remaining DOIDs	814
Output	Sparse adjacency matrix

3.3 Network Formation

The threshold Wang semantic similarity matrix was transformed into an undirected weighted network, where each node represents a disease identified by its corresponding *Disease Ontology Identifier* (DOID) and each edge represents a semantic similarity relationship with a value of at least 0.5. The resulting *raw network* consists of 1,006 nodes and 3,155 edges, partitioned into 222 connected components. The large number of connected components indicates that many diseases form small isolated sub-networks after weak semantic associations are removed, while only a subset of diseases remains highly interconnected.

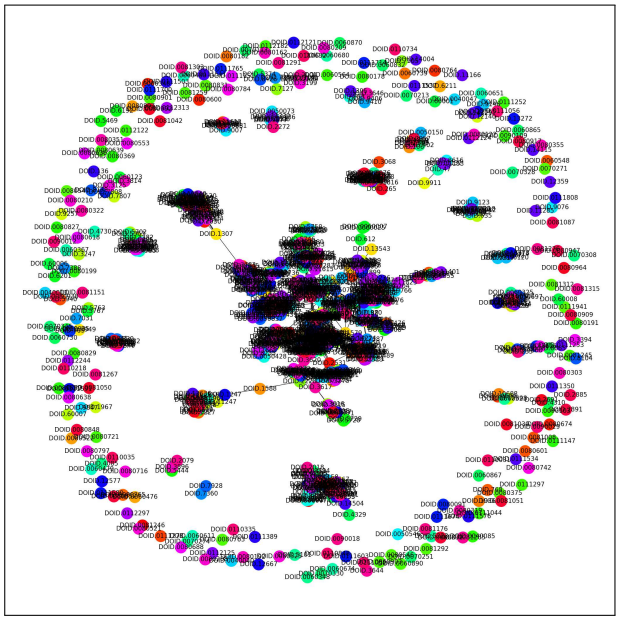


Fig. 2 Raw breast cancer comorbidity network obtained after applying a similarity threshold of 0.5.

As illustrated in Figure 2, one connected component clearly dominates the network, whereas the remaining components consist of relatively small disconnected sub-networks. Since community detection algorithms and shortest-path-based centrality measures are most meaningful when applied to a connected graph, only the largest connected component, referred to as the *Giant Component*, was retained for subsequent analysis. The extracted Giant Component contains 323 nodes and 1,078 edges, while 683 nodes and 2,077 edges belonging to the remaining 221 connected components were excluded. This Giant Component represents the principal structure of the breast cancer comorbidity network and serves as the input for all subsequent community detection and *Community-Consideration Centrality* analysis.

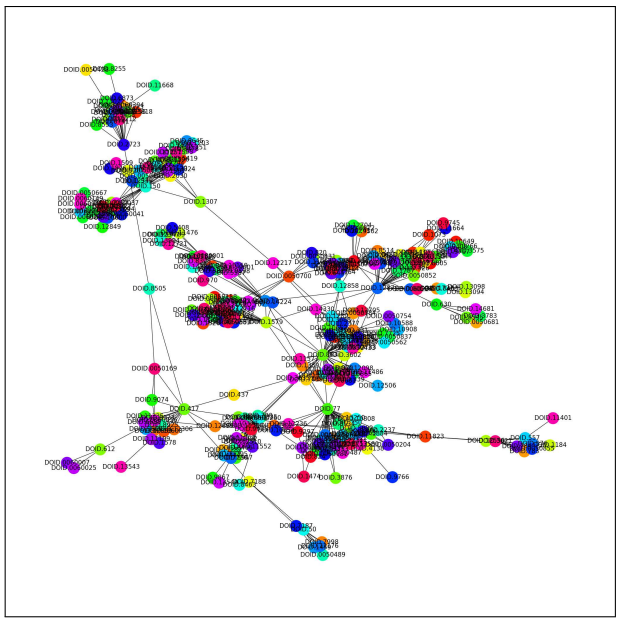


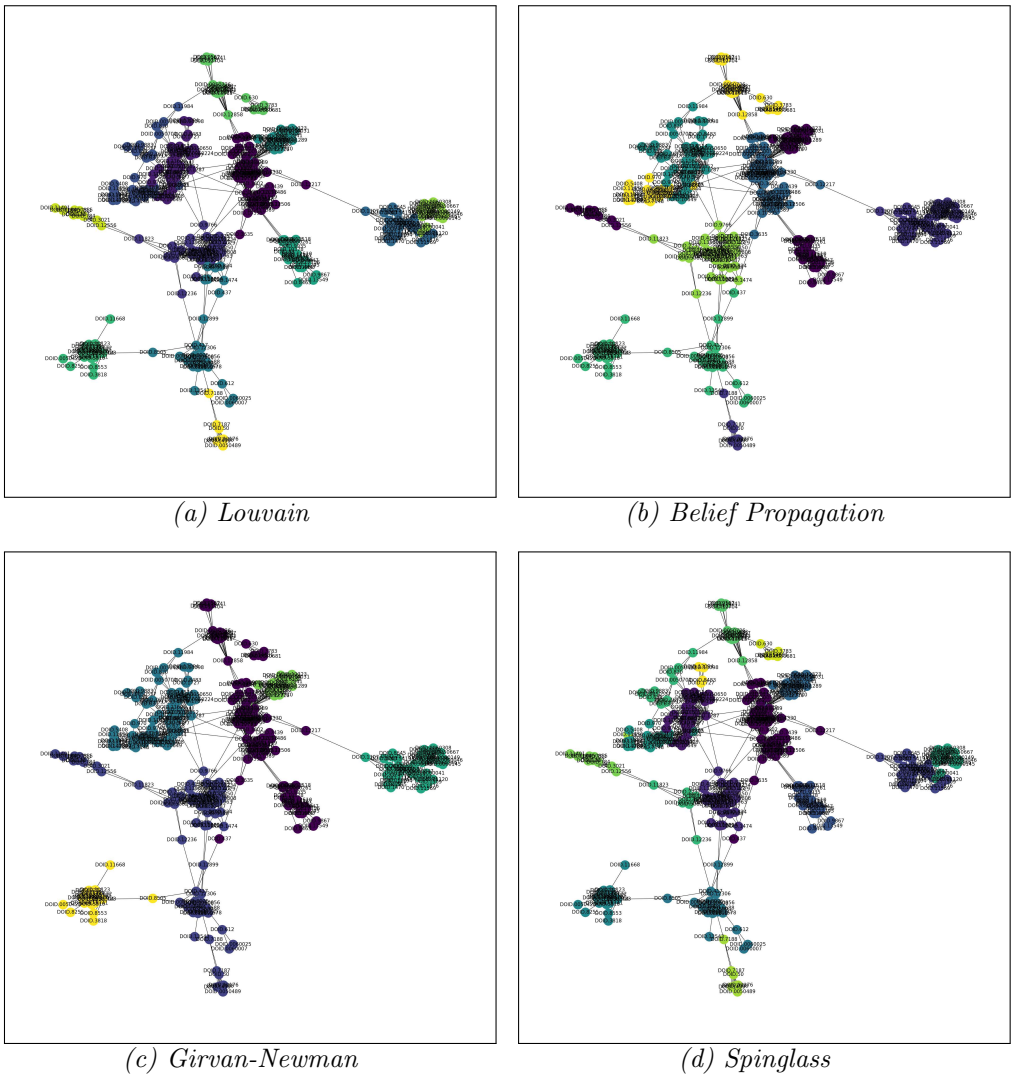
Fig. 3 Giant Component extracted from the threshold breast cancer comorbidity network.

Figure 3 illustrates the resulting network after the Giant Component extraction process. As shown in the figure, the previously disconnected components have been removed, leaving only the largest connected structure of the threshold breast cancer comorbidity network. The resulting network consists of 323 nodes and 1,078 edges, representing the interconnected disease relationships that are retained for the subsequent community detection and *Community-Consideration Centrality* analysis.

3.4 Community Detection

Community detection was performed on the Giant Component using four algorithms implemented through the *CDlib* library: *Louvain*, *Belief Propagation*, *Girvan-Newman*, and *Spinglass*. The algorithms produced 12, 5, 6, and 17 communities, respectively, as summarized in Table 2. The different numbers of communities reflect

the distinct characteristics of each algorithm in partitioning the network. The resulting community structures are visualized in Figure 4.



(a) Louvain *(b) Belief Propagation*

(c) Girvan-Newman *(d) Spinglass*

Fig. 4 Community detection results using the four community detection algorithms.

Table 2 Number of communities generated by each algorithm

Algorithm	Number of Communities
Louvain	12
Belief Propagation	5
Girvan-Newman	6
Spinglass	17

To evaluate the agreement between the community detection results, a *Consensus Similarity Matrix* was constructed. The matrix measures the proportion of algorithms that assign each pair of diseases to the same community, with values ranging from 0 to 1. A value of 1 indicates that a pair is consistently grouped together by all four algorithms, while values of 0.25, 0.50, and 0.75 indicate agreement by one, two, and three algorithms, respectively. An example of the resulting matrix is presented in Table 3. For instance, DOID.678 and DOID.4724 have a similarity value of 1, whereas DOID.678 and DOID.10123 have a value of 0.

Table 3 Example of the *Consensus Similarity Matrix* from community detection results.

DOID	**DOID.0050681**	**DOID.678**	**DOID.11282**	**DOID.4724**	**DOID.10123**
DOID.0050681	1.00	0.25	0.25	0.25	0.25
DOID.678	0.25	1.00	0.25	1.00	0.00
DOID.11282	0.25	0.25	1.00	0.25	0.00
DOID.4724	0.25	1.00	0.25	1.00	0.00
DOID.10123	0.25	0.00	0.00	0.00	1.00

The *Consensus Similarity Matrix* was subsequently used to construct a consensus network by retaining only disease pairs with $S_{ij} = 1$. As shown in Figure 5, each connected component represents a group of diseases that were consistently assigned to the same community by all four algorithms. A total of 23 connected components were identified. The largest components contained 34, 33, and 29 diseases, while four components consisted of single diseases: DOID.437, DOID.11668, DOID.8505, and DOID.11486. These consensus community structures were then used as the basis for the subsequent *Community-Consideration Centrality* analysis.

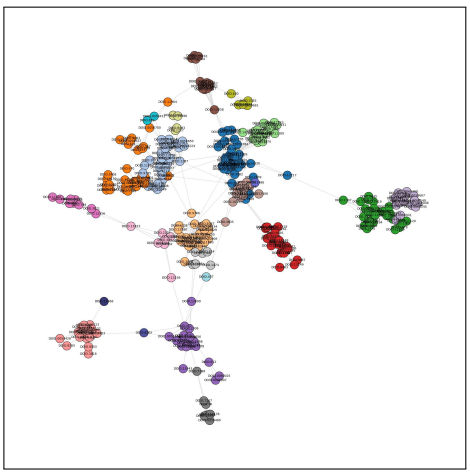


Fig. 5 Consensus network constructed from the community detection algorithms based on $S_{ij} = 1$.

3.5 Community-Consideration Centrality

After obtaining the community structure through the community detection process, the next stage is to calculate the *Community-Consideration Centrality* (CCC) values. This calculation aims to identify diseases that play important roles not only at the global network level but also within the communities to which they belong. By considering both aspects, the CCC method is expected to provide a more representative ranking than conventional centrality measures that rely solely on global network information.

In this study, three variants of *Community-Consideration Centrality* are used: *Community-Consideration Degree Centrality* (CCDC), *Community-Consideration Closeness Centrality* (CCCC), and *Community-Consideration Betweenness Centrality* (CCBC). All calculations are performed using the *Python* programming language by utilizing the previously constructed graph structure and the community detection results from each algorithm. The parameter α is used to control the balance between the contribution of global network information and local community information. Following the study by Rustamaji et al. [23], α is varied from 0 to 1 at intervals of 0.1 to evaluate its effect on the identification of central diseases.

3.5.1 Louvain

After obtaining the community structure using the *Louvain* algorithm, *Community-Consideration Centrality* (CCC) was calculated to identify diseases with important roles in the breast cancer comorbidity network. The analysis uses *Community-Consideration Degree Centrality* (CCDC), *Community-Consideration Closeness Centrality* (CCCC), and *Community-Consideration Betweenness Centrality* (CCBC), with α varied from 0 to 1. Tables present the top ten diseases for $\alpha = 0$, $\alpha = 0.1$, $\alpha = 0.9$, and $\alpha = 1.0$.

For *Community-Consideration Degree Centrality*, Figure 6 and Table 4 show a substantial ranking change when α increases from 0 to 0.1. At $\alpha = 0$, DOID.50 ranks

first because only within-community connectivity is considered. When global network information is introduced at $\alpha = 0.1$, DOID.863 becomes the highest-ranked disease and remains in first place through $\alpha = 1.0$. DOID.77 also remains consistently near the top, while DOID.331, DOID.1287, and DOID.150 remain among the top five at higher α values. In contrast, DOID.7998, DOID.1459, and DOID.12176 disappear from the top ten as global connectivity becomes more influential, indicating that their centrality is more strongly associated with local community structure.

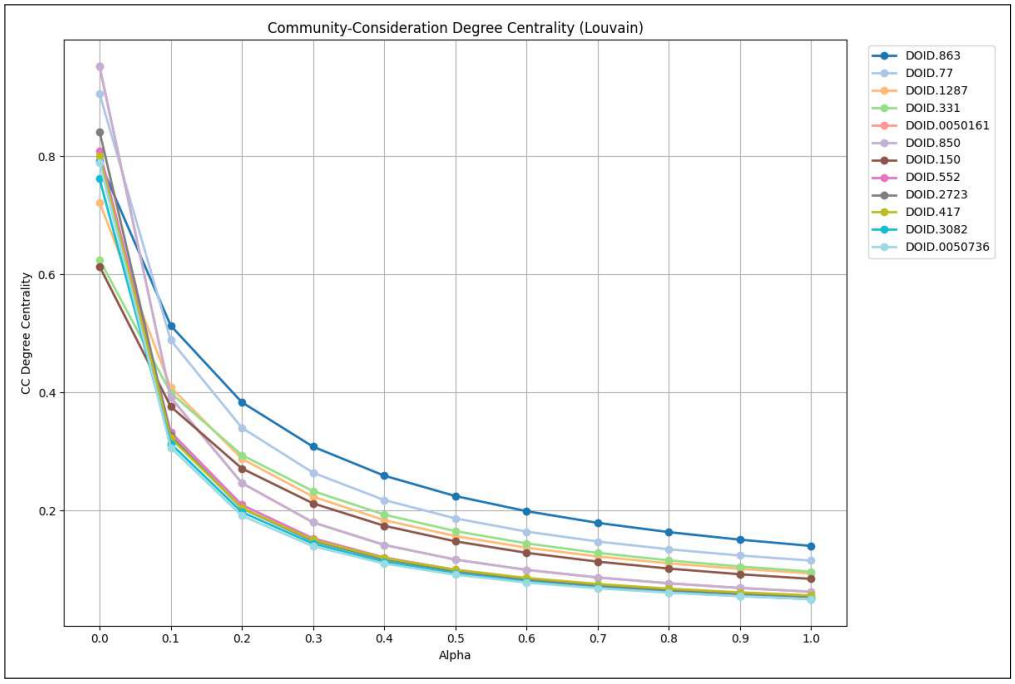


Fig. 6 Top ten diseases by *Community-Consideration Degree Centrality* using the Louvain algorithm at different α values.

For *Community-Consideration Closeness Centrality*, Figure 7 and Table 5 show a similar transition from local to global network structure. At $\alpha = 0$, DOID.50, DOID.0050161, and DOID.850 occupy the highest positions, reflecting their strong closeness within their respective communities. When α increases to 0.1, DOID.863 becomes the top-ranked disease and remains first through $\alpha = 1.0$, while DOID.77 consistently ranks second. Diseases such as DOID.1287, DOID.17, DOID.1579, DOID.14330, DOID.3688, and DOID.3635 appear among the top ten at higher α values, whereas DOID.7998, DOID.1459, DOID.12176, and DOID.7187 no longer appear. This indicates that increasing α shifts the ranking toward diseases with greater closeness across the overall network.

For *Community-Consideration Betweenness Centrality*, Figure 8 and Table 6 show a more pronounced ranking change. At $\alpha = 0$, DOID.77 ranks first, followed by

Table 4 Top ten diseases by *Community-Consideration Degree Centrality* using the Louvain algorithm.

Rank	$\alpha = 0.0$	$\alpha = 0.1$	$\cdots$	$\alpha = 0.9$	$\alpha = 1.0$
1	DOID.50	DOID.863	$\cdots$	DOID.863	DOID.863
2	DOID.0050161	DOID.77	$\cdots$	DOID.77	DOID.77
3	DOID.850	DOID.1287	$\cdots$	DOID.331	DOID.331
4	DOID.77	DOID.331	$\cdots$	DOID.1287	DOID.1287
5	DOID.2723	DOID.0050161	$\cdots$	DOID.150	DOID.150
6	DOID.7998	DOID.850	$\cdots$	DOID.0060037	DOID.0060037
7	DOID.1459	DOID.150	$\cdots$	DOID.0050161	DOID.0050161
8	DOID.12176	DOID.552	$\cdots$	DOID.1561	DOID.850
9	DOID.7187	DOID.2723	$\cdots$	DOID.850	DOID.1561
10	DOID.552	DOID.417	$\cdots$	DOID.1289	DOID.1289

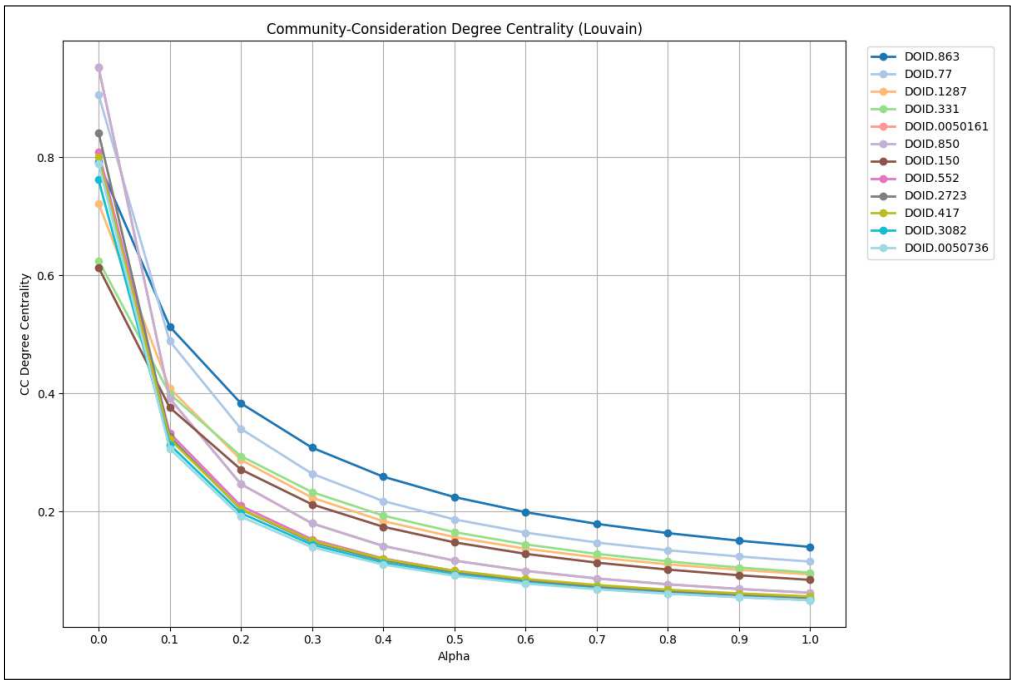


Fig. 7 Top ten diseases by *Community-Consideration Closeness Centrality* using the Louvain algorithm at different α values.

DOID.0050736 and DOID.1287, reflecting their importance as intermediary nodes within local communities. At $\alpha = 0.1$, DOID.863 becomes the highest-ranked disease and remains first through $\alpha = 1.0$, while DOID.77 remains second. DOID.12217, DOID.1307, and DOID.417 also rise as global network information becomes more influential, indicating their roles as bridge nodes between communities. Conversely, DOID.0050736, DOID.11823, DOID.0060056, and DOID.0080001 appear mainly at

Table 5 Top ten diseases by *Community-Consideration Closeness Centrality* using the Louvain algorithm.

Rank	$\alpha = 0.0$	$\alpha = 0.1$	$\cdots$	$\alpha = 0.9$	$\alpha = 1.0$
1	DOID.863	DOID.863	$\cdots$	DOID.863	DOID.863
2	DOID.9335	DOID.331	$\cdots$	DOID.77	DOID.77
3	DOID.790	DOID.9335	$\cdots$	DOID.1287	DOID.1287
4	DOID.0060260	DOID.790	$\cdots$	DOID.1289	DOID.17
5	DOID.331	DOID.0060260	$\cdots$	DOID.1579	DOID.1579
6	DOID.1289	DOID.1289	$\cdots$	DOID.17	DOID.1289
7	DOID.870	DOID.870	$\cdots$	DOID.331	DOID.331
8	DOID.1826	DOID.1826	$\cdots$	DOID.14330	DOID.14330
9	DOID.0050850	DOID.0050850	$\cdots$	DOID.3688	DOID.3688
10	DOID.4724	DOID.4724	$\cdots$	DOID.3635	DOID.3635

lower α values. The resulting ranking indicates that local intermediary roles do not necessarily correspond to the same importance at the global network level.

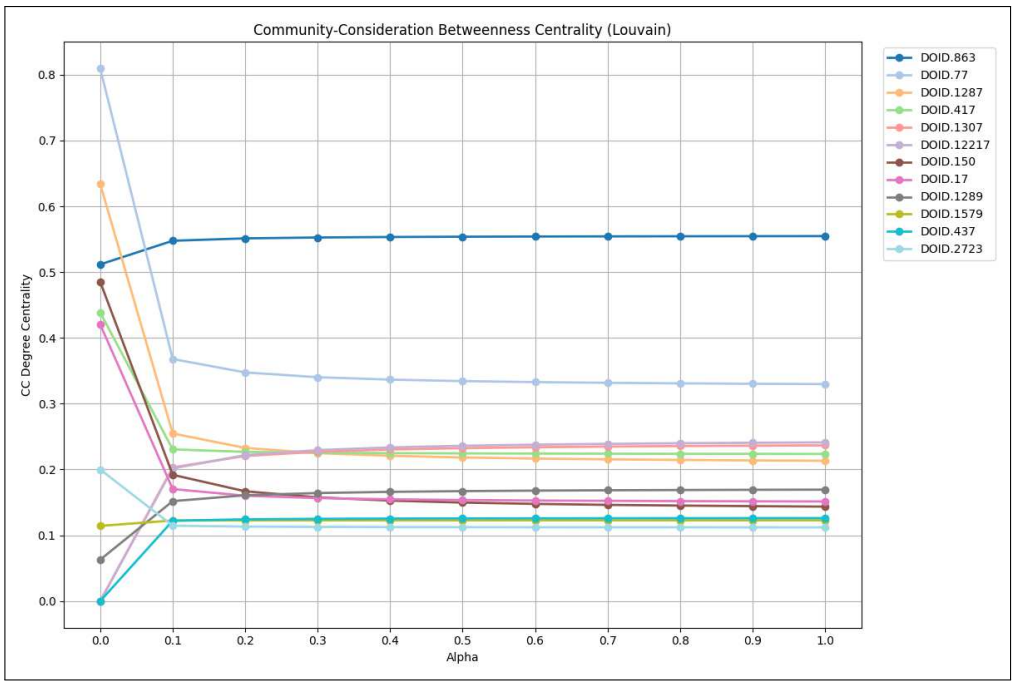


Fig. 8 Top ten diseases by *Community-Consideration Betweenness Centrality* using the Louvain algorithm at different α values.

Table 6 Top ten diseases by *Community-Consideration Betweenness Centrality* using the Louvain algorithm.

Rank	$\alpha = 0.0$	$\alpha = 0.1$	$\cdots$	$\alpha = 0.9$	$\alpha = 1.0$
1	DOID.77	DOID.863	$\cdots$	DOID.863	DOID.863
2	DOID.0050736	DOID.77	$\cdots$	DOID.77	DOID.77
3	DOID.1287	DOID.1287	$\cdots$	DOID.12217	DOID.12217
4	DOID.11823	DOID.417	$\cdots$	DOID.1307	DOID.1307
5	DOID.863	DOID.1307	$\cdots$	DOID.417	DOID.417
6	DOID.150	DOID.12217	$\cdots$	DOID.1287	DOID.1287
7	DOID.0060056	DOID.150	$\cdots$	DOID.1289	DOID.1289
8	DOID.417	DOID.17	$\cdots$	DOID.17	DOID.17
9	DOID.17	DOID.1289	$\cdots$	DOID.150	DOID.150
10	DOID.0080001	DOID.1579	$\cdots$	DOID.3635	DOID.3635

3.5.2 Girvan-Newman

After obtaining the community structure using the *Girvan-Newman* algorithm, *Community-Consideration Centrality* (CCC) was calculated using *Community-Consideration Degree Centrality* (CCDC), *Community-Consideration Closeness Centrality* (CCCC), and *Community-Consideration Betweenness Centrality* (CCBC). The *Girvan-Newman* algorithm forms communities divisively by progressively removing edges with the highest *edge betweenness*. The results are evaluated at $\alpha = 0.0$, $\alpha = 0.1$, $\alpha = 0.9$, and $\alpha = 1.0$ to observe the transition from local community structure to global network structure.

For *Community-Consideration Degree Centrality*, Figure 9 and Table 7 show that DOID.0050161 ranks first at $\alpha = 0.0$ and remains first at $\alpha = 0.1$, while DOID.863 gradually rises and becomes first at higher α values. At $\alpha = 1.0$, DOID.863 and DOID.77 occupy the first and second positions, followed by DOID.331, DOID.1287, and DOID.150. Meanwhile, diseases such as DOID.3082, DOID.10123, DOID.1584, DOID.1558, and DOID.8893 mainly appear at lower α values, whereas DOID.0060037, DOID.1561, and DOID.1289 emerge at higher values. This indicates a gradual transition from local community connectivity toward global network connectivity.

For *Community-Consideration Closeness Centrality*, Figure 10 and Table 8 show that DOID.0050161 ranks first at $\alpha = 0.0$, followed by DOID.850 and DOID.2723. At $\alpha = 0.1$, DOID.863 becomes first and remains there through $\alpha = 1.0$, while DOID.1287 initially ranks second before DOID.77 takes that position at higher α values. DOID.17, DOID.1579, DOID.331, DOID.1289, DOID.14330, DOID.3688, and DOID.3635 remain among the top-ranked diseases as the global network contribution increases. In contrast, DOID.3082, DOID.10123, DOID.1584, DOID.1558, DOID.8893, and DOID.0050426 are mainly present at $\alpha = 0.0$. These changes indicate a gradual shift from local closeness toward closeness across the overall network.

For *Community-Consideration Betweenness Centrality*, Figure 11 and Table 9 show a more dynamic ranking pattern. At $\alpha = 0.0$, DOID.1287 ranks first, followed by DOID.863 and DOID.77, reflecting their importance as intermediary nodes within the local communities. When α increases to 0.1, DOID.863 becomes first and remains

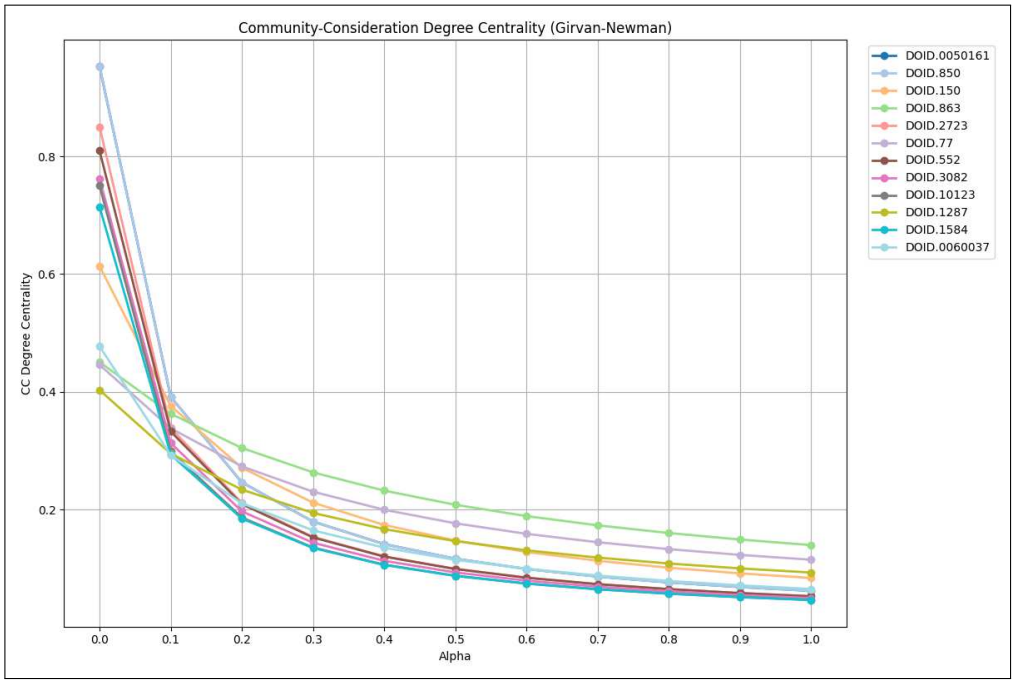


Fig. 9 Top ten diseases by *Community-Consideration Degree Centrality* using the Girvan-Newman algorithm at different α values.

Table 7 Top ten diseases by *Community-Consideration Degree Centrality* using the Girvan-Newman algorithm.

Rank	$\alpha = 0.0$	$\alpha = 0.1$	$\cdots$	$\alpha = 0.9$	$\alpha = 1.0$
1	DOID.0050161	DOID.0050161	$\cdots$	DOID.863	DOID.863
2	DOID.850	DOID.850	$\cdots$	DOID.77	DOID.77
3	DOID.2723	DOID.150	$\cdots$	DOID.331	DOID.331
4	DOID.552	DOID.863	$\cdots$	DOID.1287	DOID.1287
5	DOID.3082	DOID.2723	$\cdots$	DOID.150	DOID.150
6	DOID.10123	DOID.77	$\cdots$	DOID.0060037	DOID.0060037
7	DOID.1584	DOID.552	$\cdots$	DOID.0050161	DOID.0050161
8	DOID.1558	DOID.3082	$\cdots$	DOID.850	DOID.1561
9	DOID.8893	DOID.10123	$\cdots$	DOID.1561	DOID.850
10	DOID.0050426	DOID.1287	$\cdots$	DOID.1289	DOID.1289

there through $\alpha = 1.0$, while DOID.77 consistently ranks second. DOID.12217 and DOID.1307 also rise into the top five at higher α values, indicating stronger roles as bridges between communities. DOID.417, DOID.1287, DOID.1289, DOID.17, and DOID.150 remain in the top ten, whereas DOID.12858, DOID.331, DOID.2723, and DOID.0060037 are mainly present at lower α values. At higher α, the ranking converges

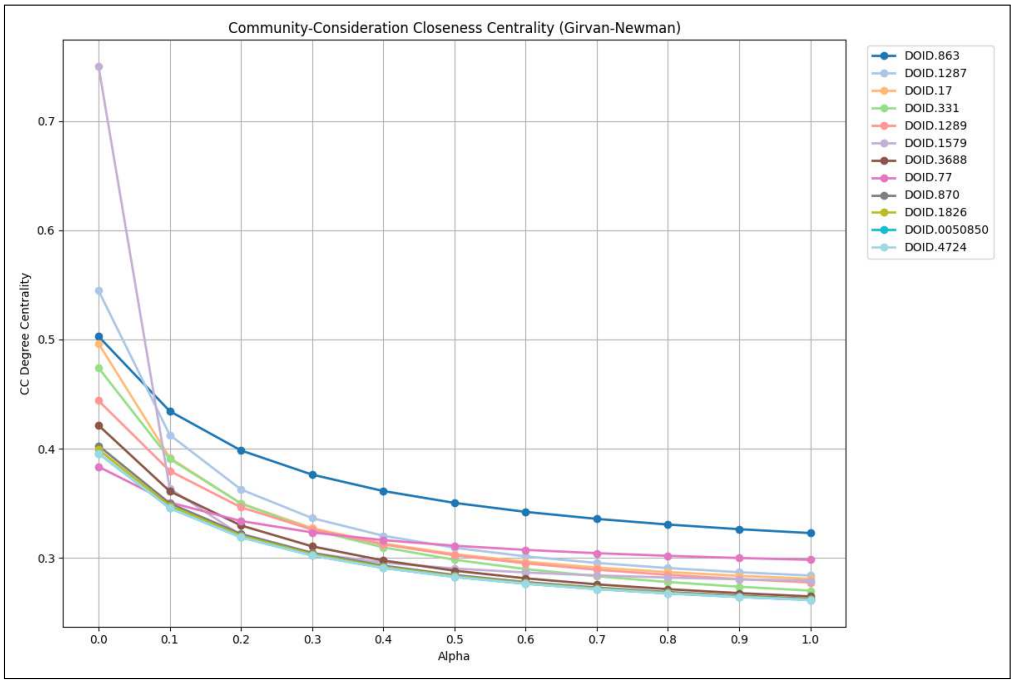


Fig. 10 Top ten diseases by *Community-Consideration Closeness Centrality* using the Girvan-Newman algorithm at different α values.

Table 8 Top ten diseases by *Community-Consideration Closeness Centrality* using the Girvan-Newman algorithm.

Rank	$\alpha = 0.0$	$\alpha = 0.1$	$\cdots$	$\alpha = 0.9$	$\alpha = 1.0$
1	DOID.0050161	DOID.863	$\cdots$	DOID.863	DOID.863
2	DOID.850	DOID.1287	$\cdots$	DOID.77	DOID.77
3	DOID.2723	DOID.17	$\cdots$	DOID.1287	DOID.1287
4	DOID.552	DOID.331	$\cdots$	DOID.17	DOID.17
5	DOID.3082	DOID.1289	$\cdots$	DOID.1579	DOID.1579
6	DOID.10123	DOID.1579	$\cdots$	DOID.331	DOID.1289
7	DOID.1584	DOID.3688	$\cdots$	DOID.331	DOID.331
8	DOID.1558	DOID.77	$\cdots$	DOID.14330	DOID.14330
9	DOID.8893	DOID.870	$\cdots$	DOID.3688	DOID.3688
10	DOID.0050426	DOID.1826	$\cdots$	DOID.3635	DOID.3635

toward DOID.863, DOID.77, DOID.12217, DOID.1307, and DOID.417 as the main intermediary diseases in the network.

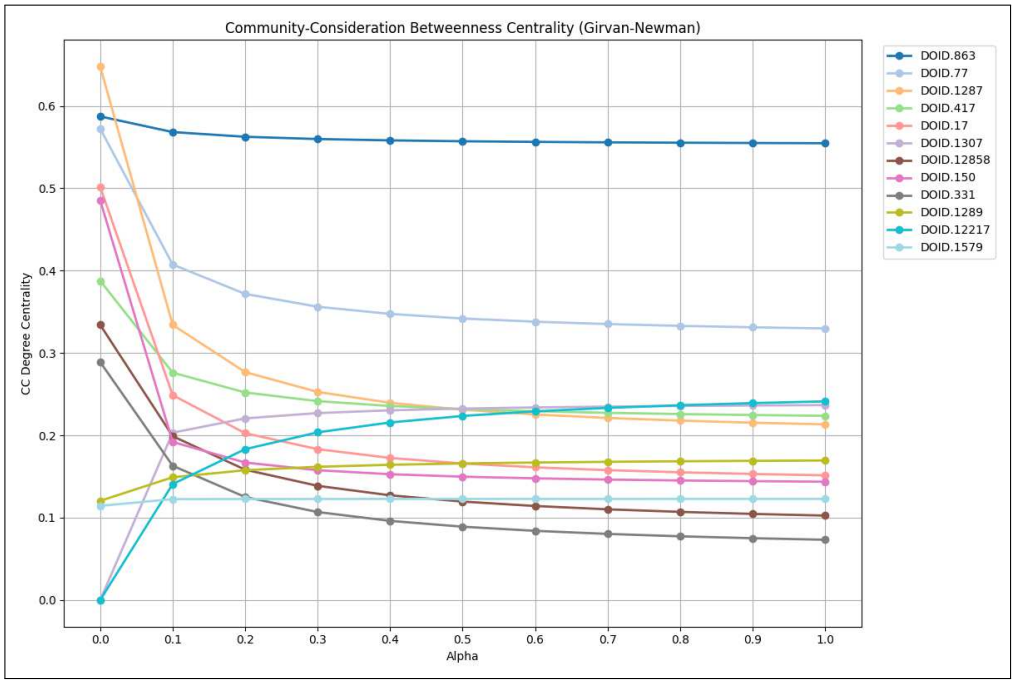


Fig. 11 Top ten diseases by *Community-Consideration Betweenness Centrality* using the Girvan-Newman algorithm at different α values.

Table 9 Top ten diseases by *Community-Consideration Betweenness Centrality* using the Girvan-Newman algorithm.

Rank	$\alpha = 0.0$	$\alpha = 0.1$	$\cdots$	$\alpha = 0.9$	$\alpha = 1.0$
1	DOID.1287	DOID.863	$\cdots$	DOID.863	DOID.863
2	DOID.863	DOID.77	$\cdots$	DOID.77	DOID.77
3	DOID.77	DOID.1287	$\cdots$	DOID.12217	DOID.12217
4	DOID.17	DOID.417	$\cdots$	DOID.1307	DOID.1307
5	DOID.150	DOID.17	$\cdots$	DOID.417	DOID.417
6	DOID.417	DOID.1307	$\cdots$	DOID.1287	DOID.1287
7	DOID.12858	DOID.12858	$\cdots$	DOID.1289	DOID.1289
8	DOID.331	DOID.150	$\cdots$	DOID.17	DOID.17
9	DOID.2723	DOID.331	$\cdots$	DOID.150	DOID.150
10	DOID.0060037	DOID.1289	$\cdots$	DOID.3635	DOID.3635

3.5.3 Belief

After obtaining the community structure using the *Belief Propagation* algorithm, *Community-Consideration Centrality* (CCC) was calculated using *Community-Consideration Degree Centrality* (CCDC), *Community-Consideration Closeness Centrality* (CCCC), and *Community-Consideration Betweenness Centrality* (CCBC).

Unlike *Girvan-Newman*, which forms communities by progressively removing edges, *Belief Propagation* uses a *message-passing* approach to estimate the probability of each node belonging to a community. The results are evaluated at $\alpha = 0.0$, $\alpha = 0.1$, $\alpha = 0.9$, and $\alpha = 1.0$ to examine the transition between local and global network contributions.

For *Community-Consideration Degree Centrality*, Figure 12 and Table 10 show a relatively rapid ranking convergence. At $\alpha = 0.0$, DOID.0050161 ranks first, followed by DOID.850 and DOID.863. At $\alpha = 0.1$, DOID.863 becomes the top-ranked disease, while DOID.0050161 remains second. At higher α values, DOID.863, DOID.77, DOID.331, DOID.1287, and DOID.150 consistently occupy the top five, while DOID.0060037, DOID.0050161, DOID.1561, DOID.850, and DOID.1289 remain within the top ten. In contrast, DOID.3082, DOID.1584, DOID.1579, and DOID.552 mainly appear at lower α values. This indicates that the ranking converges relatively quickly as global network information becomes more influential.

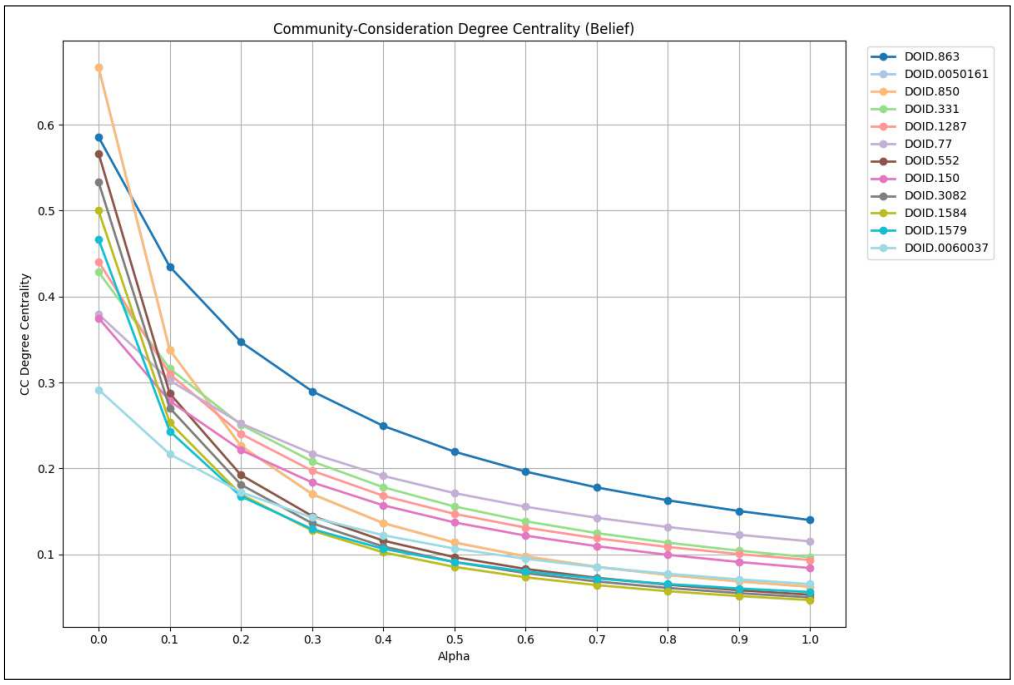


Fig. 12 Top ten diseases by *Community-Consideration Degree Centrality* using the Belief Propagation algorithm at different α values.

For *Community-Consideration Closeness Centrality*, Figure 13 and Table 11 show a highly stable ranking across α values. DOID.863 remains first from $\alpha = 0.0$ through $\alpha = 1.0$, while DOID.1287 and DOID.17 remain among the top three at lower α values. As α increases, DOID.77 rises to second place, followed by DOID.1287. DOID.1579, DOID.1289, DOID.331, DOID.14330, DOID.3688, and DOID.3635 also remain in the

Table 10 Top ten diseases by *Community-Consideration Degree Centrality* using the Belief Propagation algorithm.

Rank	$\alpha = 0.0$	$\alpha = 0.1$	$\cdots$	$\alpha = 0.9$	$\alpha = 1.0$
1	DOID.0050161	DOID.863	$\cdots$	DOID.863	DOID.863
2	DOID.850	DOID.0050161	$\cdots$	DOID.77	DOID.77
3	DOID.863	DOID.850	$\cdots$	DOID.331	DOID.331
4	DOID.552	DOID.331	$\cdots$	DOID.1287	DOID.1287
5	DOID.3082	DOID.1287	$\cdots$	DOID.150	DOID.150
6	DOID.1584	DOID.77	$\cdots$	DOID.0060037	DOID.0060037
7	DOID.1579	DOID.552	$\cdots$	DOID.0050161	DOID.0050161
8	DOID.1287	DOID.150	$\cdots$	DOID.850	DOID.1561
9	DOID.331	DOID.3082	$\cdots$	DOID.1561	DOID.850
10	DOID.77	DOID.1584	$\cdots$	DOID.1289	DOID.1289

top ten at higher α values. Only minor positional changes occur, indicating that the *Belief Propagation* community structure produces stable closeness rankings with limited sensitivity to changes in α.

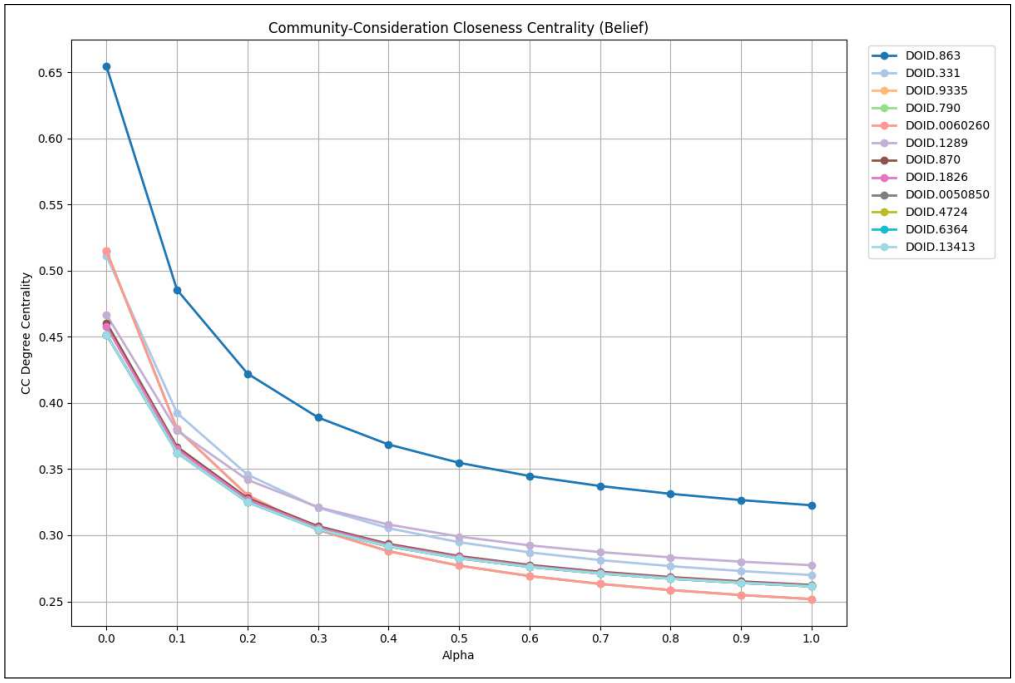


Fig. 13 Top ten diseases by *Community-Consideration Closeness Centrality* using the Belief Propagation algorithm at different α values.

Table 11 Top ten diseases by *Community-Consideration Closeness Centrality* using the Belief Propagation algorithm.

Rank	$\alpha = 0.0$	$\alpha = 0.1$	$\cdots$	$\alpha = 0.9$	$\alpha = 1.0$
1	DOID.863	DOID.863	$\cdots$	DOID.863	DOID.863
2	DOID.1287	DOID.1287	$\cdots$	DOID.77	DOID.77
3	DOID.17	DOID.17	$\cdots$	DOID.1287	DOID.1287
4	DOID.331	DOID.331	$\cdots$	DOID.17	DOID.17
5	DOID.1289	DOID.1289	$\cdots$	DOID.1579	DOID.1579
6	DOID.1579	DOID.1579	$\cdots$	DOID.1289	DOID.1289
7	DOID.3688	DOID.3688	$\cdots$	DOID.331	DOID.331
8	DOID.77	DOID.77	$\cdots$	DOID.14330	DOID.14330
9	DOID.870	DOID.870	$\cdots$	DOID.3688	DOID.3688
10	DOID.1826	DOID.1826	$\cdots$	DOID.3635	DOID.3635

For *Community-Consideration Betweenness Centrality*, Figure 14 and Table 12 show the most stable ranking among the three measures. DOID.863 and DOID.77 remain first and second from $\alpha = 0.0$ through $\alpha = 1.0$, while DOID.12217 and DOID.1307 rise into the top five as the global contribution increases. At higher α values, DOID.863, DOID.77, DOID.12217, DOID.1307, and DOID.417 consistently occupy the top five, followed by DOID.1287, DOID.1289, DOID.17, DOID.150, and DOID.3635 within the top ten. The limited ranking changes indicate that the main bridge nodes remain relatively stable across different α values.

Table 12 Top ten diseases by *Community-Consideration Betweenness Centrality* using the Belief Propagation algorithm.

Rank	$\alpha = 0.0$	$\alpha = 0.1$	$\cdots$	$\alpha = 0.9$	$\alpha = 1.0$
1	DOID.863	DOID.863	$\cdots$	DOID.863	DOID.863
2	DOID.77	DOID.77	$\cdots$	DOID.77	DOID.77
3	DOID.1287	DOID.12217	$\cdots$	DOID.12217	DOID.12217
4	DOID.417	DOID.1307	$\cdots$	DOID.1307	DOID.1307
5	DOID.17	DOID.417	$\cdots$	DOID.417	DOID.417
6	DOID.1307	DOID.1287	$\cdots$	DOID.1287	DOID.1287
7	DOID.12858	DOID.12858	$\cdots$	DOID.1289	DOID.1289
8	DOID.150	DOID.150	$\cdots$	DOID.17	DOID.17
9	DOID.331	DOID.331	$\cdots$	DOID.150	DOID.150
10	DOID.1289	DOID.1289	$\cdots$	DOID.3635	DOID.3635

3.5.4 Spinglass

After obtaining the community structure using the *Spinglass* algorithm, *Community-Consideration Centrality* (CCC) was calculated to identify diseases with important roles in the breast cancer comorbidity network. Unlike *Louvain*, *Girvan-Newman*, and *Belief Propagation*, *Spinglass* forms communities using a statistical-physics approach

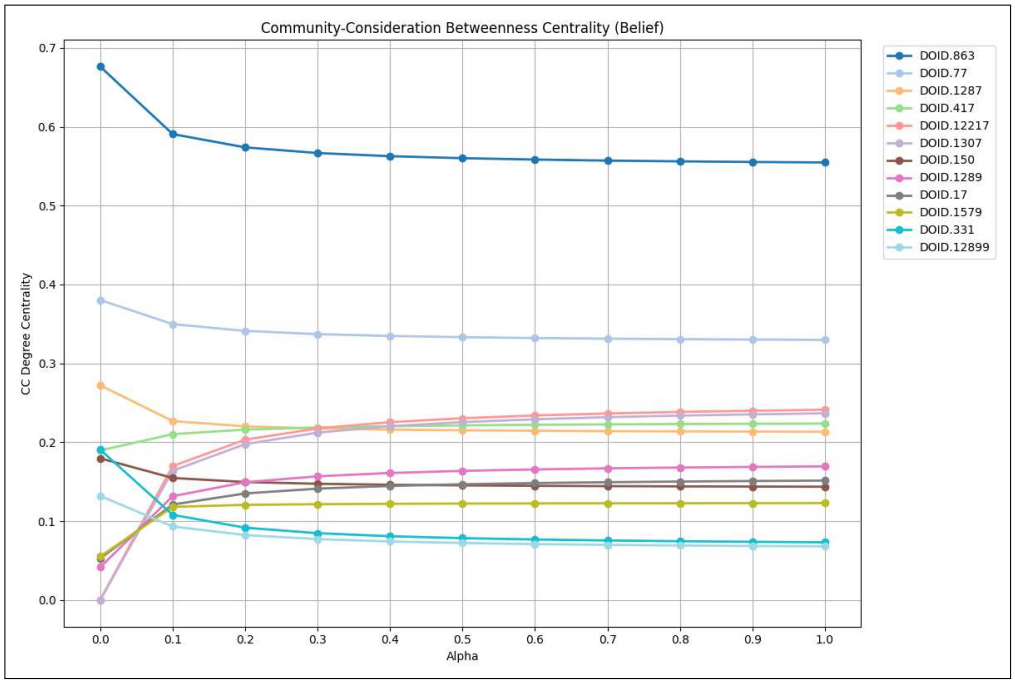


Fig. 14 Top ten diseases by *Community-Consideration Betweenness Centrality* using the Belief Propagation algorithm at different α values.

based on Hamiltonian energy optimization. The resulting structure was used to calculate *Community-Consideration Degree Centrality* (CCDC), *Community-Consideration Closeness Centrality* (CCCC), and *Community-Consideration Betweenness Centrality* (CCBC). Results are presented for $\alpha = 0.0$, $\alpha = 0.1$, $\alpha = 0.9$, and $\alpha = 1.0$.

For *Community-Consideration Degree Centrality*, Figure 15 and Table 13 show a distinct ranking pattern at low α. At $\alpha = 0.0$, DOID.0050736, DOID.1727, DOID.13094, DOID.13098, DOID.630, and DOID.8483 dominate the top ten, reflecting the strong influence of the local *Spinglass* community structure. At $\alpha = 0.1$, DOID.77 becomes the highest-ranked disease, followed by DOID.331 and DOID.863. As α increases, the ranking gradually converges, with DOID.863 becoming first and DOID.77, DOID.331, DOID.1287, and DOID.150 following at $\alpha = 1.0$. DOID.0060037, DOID.0050161, DOID.850, DOID.1561, and DOID.1289 also remain among the top ten at high α, while the diseases dominant at $\alpha = 0.0$ disappear. This indicates a gradual transition from community-level connectivity toward global network connectivity.

For *Community-Consideration Closeness Centrality*, Figure 16 and Table 14 show the largest ranking change among the community detection algorithms as α increases. At $\alpha = 0.0$, the top ten is dominated by DOID.0050736, DOID.1727, DOID.13094, DOID.13098, DOID.630, and DOID.8483, indicating strong local closeness within the *Spinglass* communities. At $\alpha = 0.1$, DOID.863 becomes first, followed by

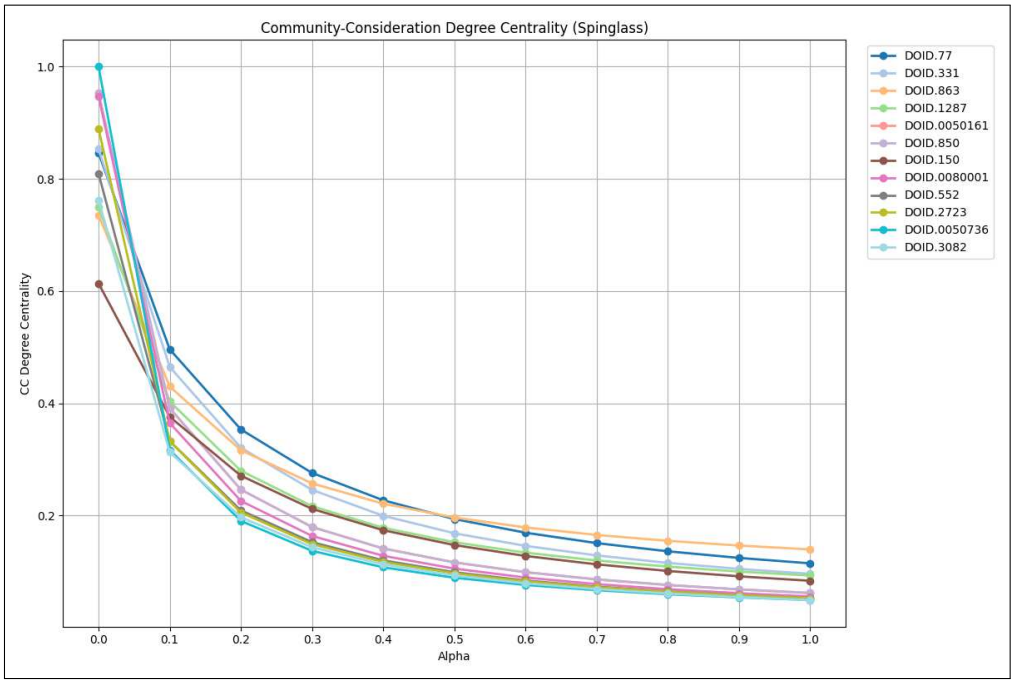


Fig. 15 Top ten diseases by *Community-Consideration Degree Centrality* using the Spinglass algorithm at different α values.

Table 13 Top ten diseases by *Community-Consideration Degree Centrality* using the Spinglass algorithm.

Rank	$\alpha = 0.0$	$\alpha = 0.1$	$\cdots$	$\alpha = 0.9$	$\alpha = 1.0$
1	DOID.0050736	DOID.77	$\cdots$	DOID.863	DOID.863
2	DOID.1727	DOID.331	$\cdots$	DOID.77	DOID.77
3	DOID.13094	DOID.863	$\cdots$	DOID.331	DOID.331
4	DOID.13098	DOID.1287	$\cdots$	DOID.1287	DOID.1287
5	DOID.630	DOID.0050161	$\cdots$	DOID.150	DOID.150
6	DOID.8483	DOID.850	$\cdots$	DOID.0060037	DOID.0060037
7	DOID.50	DOID.0060037	$\cdots$	DOID.0050161	DOID.0050161
8	DOID.0060037	DOID.0080001	$\cdots$	DOID.850	DOID.1561
9	DOID.0050161	DOID.150	$\cdots$	DOID.1561	DOID.850
10	DOID.850	DOID.1561	$\cdots$	DOID.1289	DOID.1289

DOID.77, DOID.1287, DOID.331, and DOID.1289, while DOID.1826, DOID.0050850, DOID.4724, DOID.6364, and DOID.13413 enter the top ten. At higher α values, DOID.863, DOID.77, and DOID.1287 consistently occupy the top three, followed by DOID.17, DOID.1579, DOID.1289, DOID.331, DOID.14330, DOID.3688, and DOID.3635. The results indicate a strong transition from local community closeness toward global network closeness.

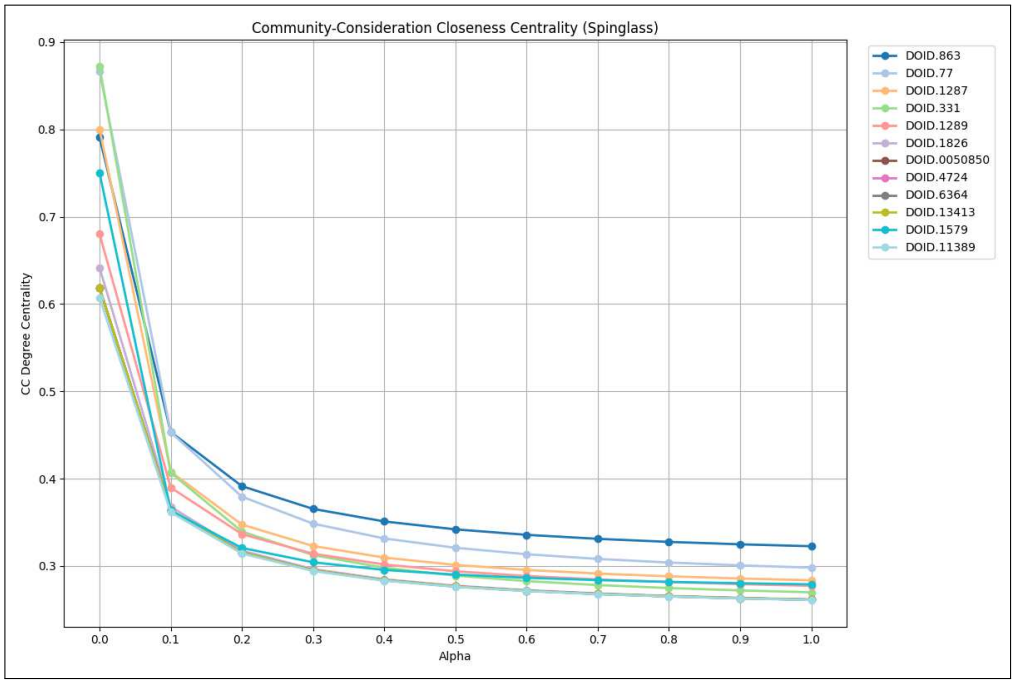


Fig. 16 Top ten diseases by *Community-Consideration Closeness Centrality* using the Spinglass algorithm at different α values.

Table 14 Top ten diseases by *Community-Consideration Closeness Centrality* using the Spinglass algorithm.

Rank	$\alpha = 0.0$	$\alpha = 0.1$	$\cdots$	$\alpha = 0.9$	$\alpha = 1.0$
1	DOID.0050736	DOID.863	$\cdots$	DOID.863	DOID.863
2	DOID.1727	DOID.77	$\cdots$	DOID.77	DOID.77
3	DOID.13094	DOID.1287	$\cdots$	DOID.1287	DOID.1287
4	DOID.13098	DOID.331	$\cdots$	DOID.17	DOID.17
5	DOID.630	DOID.1289	$\cdots$	DOID.1579	DOID.1579
6	DOID.8483	DOID.1826	$\cdots$	DOID.1289	DOID.1289
7	DOID.50	DOID.0050850	$\cdots$	DOID.331	DOID.331
8	DOID.0060037	DOID.4724	$\cdots$	DOID.14330	DOID.14330
9	DOID.0050161	DOID.6364	$\cdots$	DOID.3688	DOID.3688
10	DOID.850	DOID.13413	$\cdots$	DOID.3635	DOID.3635

For *Community-Consideration Betweenness Centrality*, Figure 17 and Table 15 show a substantial ranking change between $\alpha = 0.0$ and $\alpha = 0.1$. At $\alpha = 0.0$, DOID.630, DOID.77, DOID.0080001, DOID.870, and DOID.1287 occupy the top five, reflecting bridge roles within the local *Spinglass* communities. At $\alpha = 0.1$, DOID.863 becomes first, followed by DOID.77, DOID.1287, DOID.417, and DOID.1307, while

DOID.12217, DOID.1289, DOID.150, DOID.17, and DOID.3635 enter the top ten. At higher α values, DOID.863 and DOID.77 remain first and second, followed by DOID.12217, DOID.1307, DOID.417, and DOID.1287. DOID.1289, DOID.17, DOID.150, and DOID.3635 also remain in the top ten. Meanwhile, diseases such as DOID.630, DOID.0080001, DOID.870, DOID.11823, DOID.0050700, DOID.0060056, and DOID.0050736 disappear as global network information becomes dominant. Thus, the results show a strong transition from local bridge roles toward globally important intermediary diseases.

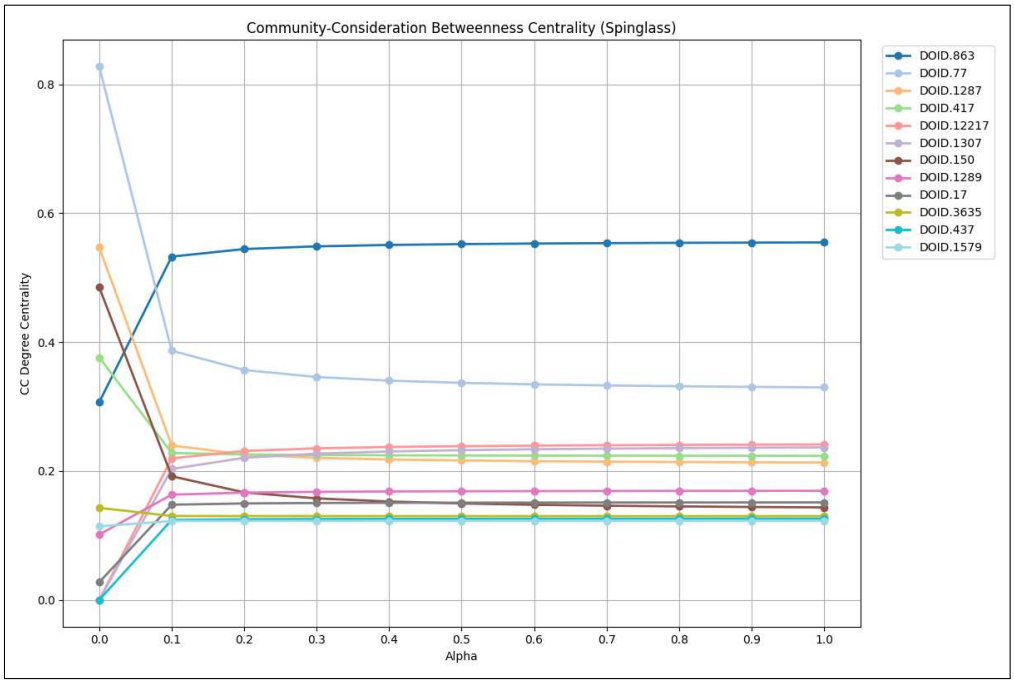


Fig. 17 Top ten diseases by *Community-Consideration Betweenness Centrality* using the Spinglass algorithm at different α values.

3.6 Performance Evaluation and Parameter Optimization

The effect of α was evaluated using the *Jackknife* method following the approach of Permana *et al.* [22]. For each α from 0.0 to 1.0 with an interval of 0.1, diseases were ranked according to their *Community-Consideration Centrality* and compared with the confirmed breast cancer comorbidities used as the *ground truth*. The number of confirmed diseases identified within the top-k rankings was used to evaluate the performance of each parameter. A higher and more consistent Jackknife curve indicates better identification performance.

For *Louvain*, the selected parameters were $\alpha = 0.3$, $\alpha = 0.0$, and $\alpha = 0.0$ for Degree, Closeness, and Betweenness Centrality, respectively. Degree Centrality showed

Table 15 Top ten diseases by *Community-Consideration Betweenness Centrality* using the Spinglass algorithm.

Rank	$\alpha = 0.0$	$\alpha = 0.1$	$\cdots$	$\alpha = 0.9$	$\alpha = 1.0$
1	DOID.630	DOID.863	$\cdots$	DOID.863	DOID.863
2	DOID.77	DOID.77	$\cdots$	DOID.77	DOID.77
3	DOID.0080001	DOID.1287	$\cdots$	DOID.12217	DOID.12217
4	DOID.870	DOID.417	$\cdots$	DOID.1307	DOID.1307
5	DOID.1287	DOID.1307	$\cdots$	DOID.417	DOID.417
6	DOID.11823	DOID.12217	$\cdots$	DOID.1287	DOID.1287
7	DOID.150	DOID.1289	$\cdots$	DOID.1289	DOID.1289
8	DOID.0050700	DOID.150	$\cdots$	DOID.17	DOID.17
9	DOID.0060056	DOID.17	$\cdots$	DOID.150	DOID.150
10	DOID.0050736	DOID.3635	$\cdots$	DOID.3635	DOID.3635

relatively similar performance across parameters, although $\alpha = 0.0$ reached the highest number of confirmed diseases at larger top-k values. However, $\alpha = 0.3$ provided a more stable performance throughout the evaluation. Closeness and Betweenness Centrality both favored $\alpha = 0.0$, which consistently produced higher Jackknife curves than the other parameters. The corresponding results are shown in Figure 18, Figure 19, and Figure 20.

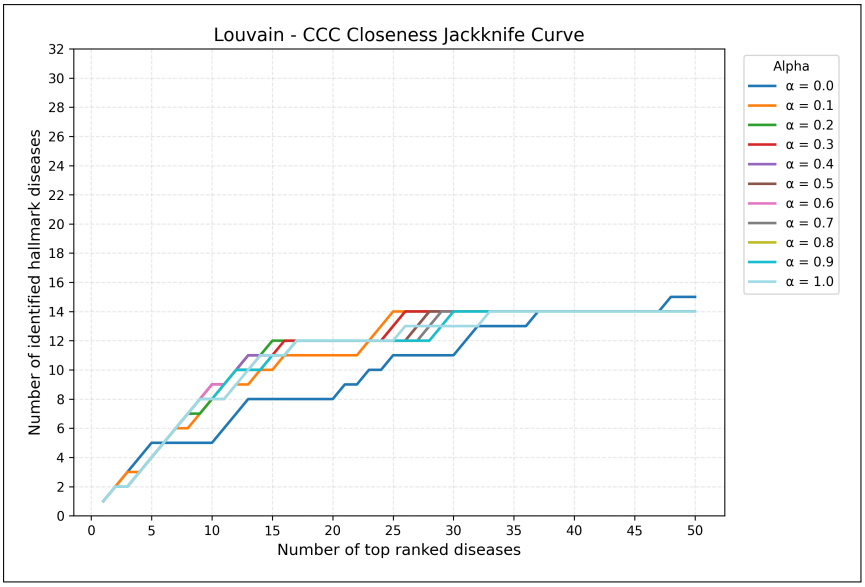


Fig. 18 Jackknife curve for *Community-Consideration Degree Centrality* using the Louvain algorithm.

For *Girvan-Newman*, the selected parameters were $\alpha = 1.0$, $\alpha = 0.0$, and $\alpha = 0.0$ for Degree, Closeness, and Betweenness Centrality, respectively. Degree Centrality

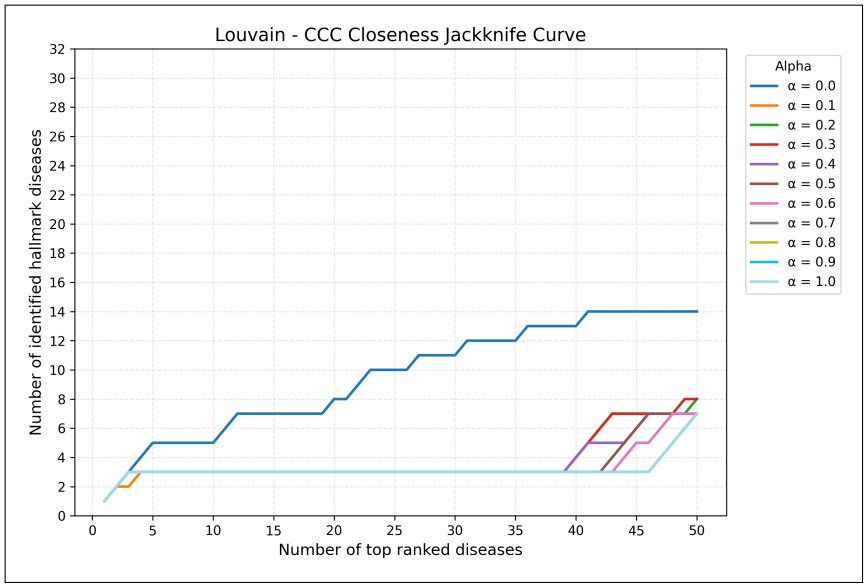


Fig. 19 Jackknife curve for *Community-Consideration Closeness Centrality* using the Louvain algorithm.

showed improved performance at higher α values, with $\alpha = 1.0$ producing the highest Jackknife curve. In contrast, Closeness and Betweenness Centrality consistently favored $\alpha = 0.0$, indicating stronger identification performance when community-level information was emphasized. The results are presented in Figure 21, Figure 22, and Figure 23.

For *Belief Propagation*, the selected parameters were $\alpha = 1.0$, $\alpha = 0.1$, and $\alpha = 0.1$ for Degree, Closeness, and Betweenness Centrality, respectively. Degree Centrality favored higher α values, with $\alpha = 1.0$ providing the highest and most stable performance. Closeness Centrality showed relatively small differences between parameters but achieved its highest performance at $\alpha = 0.1$. Betweenness Centrality also favored $\alpha = 0.1$, which reached the highest performance earlier than the other parameters. These results are shown in Figure 24, Figure 25, and Figure 26.

For *Spinglass*, $\alpha = 0.0$ was selected for all three centrality measures. Degree and Closeness Centrality showed a clear advantage for $\alpha = 0.0$, reaching substantially higher numbers of confirmed diseases than the other parameters. Betweenness Centrality also favored $\alpha = 0.0$, although the difference was smaller. The results indicate that the local community structure generated by *Spinglass* provides the strongest contribution to disease identification across the three centrality measures. The corresponding curves are shown in Figure 27, Figure 28, and Figure 29.

The selected parameters are summarized as follows: *Louvain* uses $\alpha = (0.3, 0.0, 0.0)$, *Girvan-Newman* uses $\alpha = (1.0, 0.0, 0.0)$, *Belief Propagation* uses $\alpha = (1.0, 0.1, 0.1)$, and *Spinglass* uses $\alpha = (0.0, 0.0, 0.0)$ for Degree, Closeness, and Betweenness Centrality, respectively.

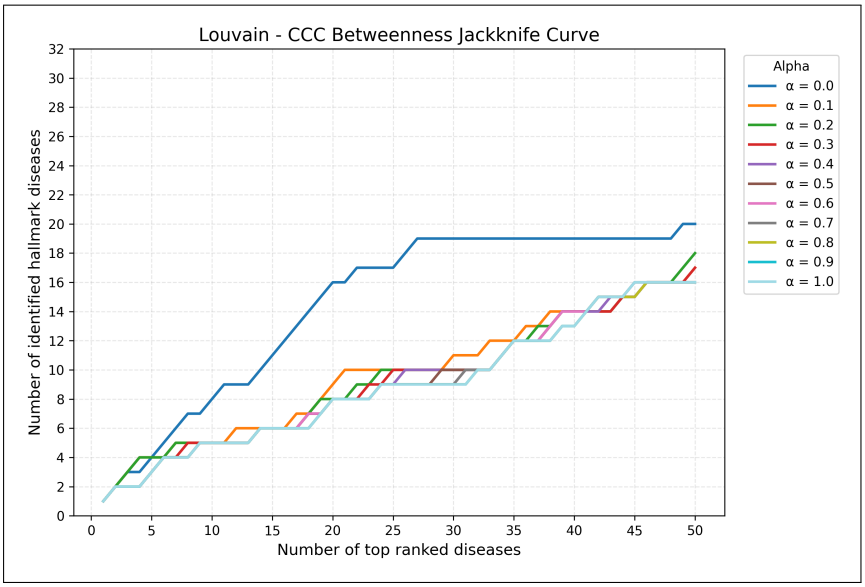


Fig. 20 Jackknife curve for *Community-Consideration Betweenness Centrality* using the Louvain algorithm.

3.7 Disease Name Mapping from DOID

Based on the rankings obtained from all combinations of community detection algorithms, centrality measures, and α values, several diseases repeatedly appeared among the top-ten results. Each disease is represented by its *Disease Ontology Identifier* (DOID), allowing consistent identification across configurations. From the 132 tested combinations, 52 unique diseases appeared in the top-ten rankings. The complete list is presented in Table 16.

Table 16: Diseases that appeared in the top-ten rankings of *Community-Consideration Centrality*.

No.	DOID	Disease Name
1	DOID.77	*Gastrointestinal System Disease*
2	DOID.863	*Nervous System Disease*
3	DOID.1287	*Cardiovascular System Disease*
4	DOID.1289	*Neurodegenerative Disease*
5	DOID.331	*Central Nervous System Disease*
6	DOID.150	*Disease of Mental Health*
7	DOID.17	*Musculoskeletal System Disease*
8	DOID.0050161	*Lower Respiratory Tract Disease*
9	DOID.850	*Lung Disease*
10	DOID.0060037	*Developmental Disorder of Mental Health*
11	DOID.417	*Autoimmune Disease*

No.	DOID	Disease Name
12	DOID.1579	*Respiratory System Disease*
13	DOID.1307	*Dementia*
14	DOID.12217	*Lewy Body Dementia*
15	DOID.3635	*Congenital Myasthenic Syndrome*
16	DOID.1561	*Cognitive Disorder*
17	DOID.14330	*Parkinson's Disease*
18	DOID.3688	*Plexopathy*
19	DOID.870	*Neuropathy*
20	DOID.1826	*Epilepsy*
21	DOID.552	*Pneumonia*
22	DOID.0080001	*Bone Disease*
23	DOID.2723	*Dermatitis*
24	DOID.4724	*Brain Edema*
25	DOID.0050850	*Diabetic Encephalopathy*
26	DOID.3082	*Interstitial Lung Disease*
27	DOID.12858	*Huntington's Disease*
28	DOID.0050736	*Autosomal Dominant Disease*
29	DOID.0060260	*Ptosis*
30	DOID.9335	*Scotoma*
31	DOID.790	*Ocular Hypotension*
32	DOID.50	*Thyroid Gland Disease*
33	DOID.1584	*Acute Chest Syndrome*
34	DOID.630	*Genetic Disease*
35	DOID.557	*Kidney Disease*
36	DOID.10123	*Pigmentation Disease*
37	DOID.0060056	*Hypersensitivity Reaction Disease*
38	DOID.437	*Myasthenia Gravis*
39	DOID.1459	*Hypothyroidism*
40	DOID.7998	*Hyperthyroidism*
41	DOID.6364	*Migraine*
42	DOID.0060038	*Specific Developmental Disorder*
43	DOID.8893	*Psoriasis*
44	DOID.1558	*Angioedema*
45	DOID.0050426	*Stevens–Johnson Syndrome*
46	DOID.11823	*Hepatorenal Syndrome*
47	DOID.1727	*Retinal Vein Occlusion*
48	DOID.13094	*Branch Retinal Artery Occlusion*
49	DOID.8483	*Retinal Artery Occlusion*
50	DOID.13098	*Central Retinal Artery Occlusion*
51	DOID.0050700	*Cardiomyopathy*
52	DOID.13413	*Hepatic Encephalopathy*

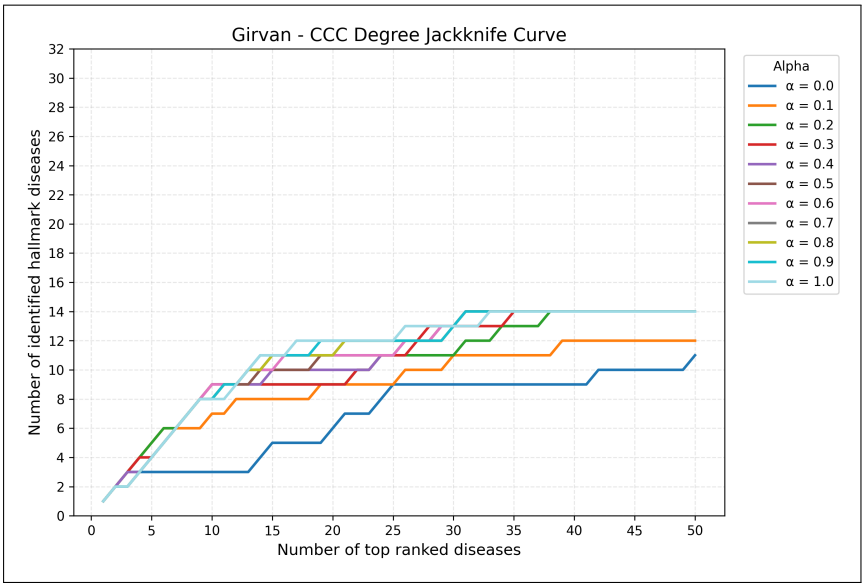


Fig. 21 Jackknife curve for *Community-Consideration Degree Centrality* using the Girvan-Newman algorithm.

4 Conclusion

Community detection using four algorithms produced different community structures: 12 communities for *Louvain*, six for *Girvan-Newman*, five for *Belief Propagation*, and 17 for *Spinglass*. Based on modularity, *Louvain* produced the best community structure, with stronger internal connectivity and clearer separation between communities.

The *Community-Consideration Centrality* (CCC) method successfully identified diseases with central roles by considering both community structure and the overall network. The three CCC measures, namely *Degree*, *Closeness*, and *Betweenness Centrality*, showed that diseases such as DOID.863 (*Nervous System Disease*), DOID.77 (*Gastrointestinal System Disease*), DOID.331 (*Central Nervous System Disease*), and DOID.1287 (*Cardiovascular System Disease*) frequently appeared among the top-ranked diseases.

The value of α affected the balance between local community information and global network structure. Lower α values were more influenced by community structure, while higher values produced more consistent results across algorithms. Based on the *Jackknife* evaluation, $\alpha = 0.9$ provided stable performance across most combinations of community detection algorithms and centrality measures, making it a suitable parameter for CCC analysis on the breast cancer comorbidity network.

Acknowledgements. The authors would like to thank the support of the All facility of Universitas Multimedia Nusantara, Tangerang, Indonesia, used for this research.

Author Contributions. Conceptualization, AAP, GET; Methodology, AAP, GET; Software, GET; Validation, AAP, GET; Formal analysis, AAP, GET; Investigation,

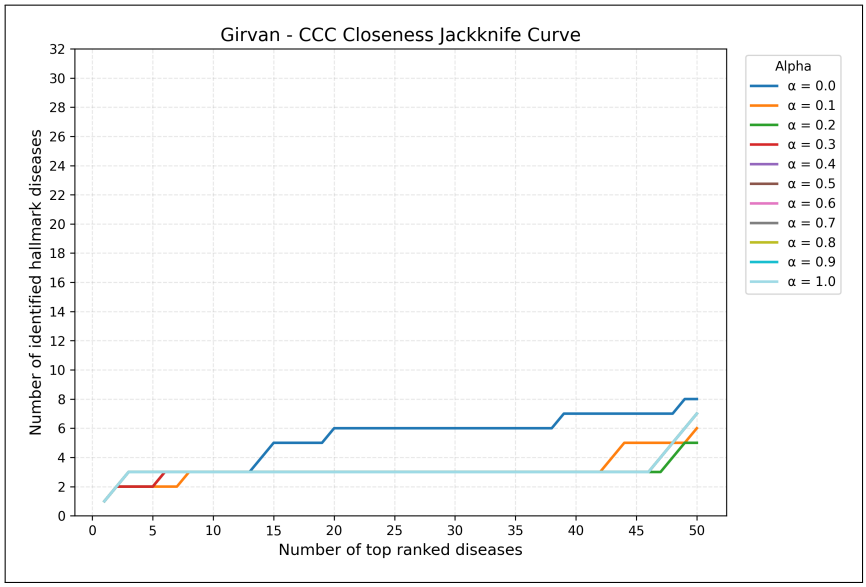


Fig. 22 Jackknife curve for *Community-Consideration Closeness Centrality* using the Girvan-Newman algorithm.

AAP, GET; Data curation, AAP, GET; Visualization, AAP, GET; Writing-original draft, GET; Writing-review and editing, AAP; Supervision, AAP; Project administration, AAP, GET. All authors made substantial contributions to the conception and design of the study, interpretation of the results, and preparation of the manuscript. All authors critically reviewed the manuscript, approved the final version, and agreed to be accountable for all aspects of the work.

References

[1] Watkins, E.J.: Overview of breast cancer. Journal of the American Academy of Physician Assistants **32**(10), 13–17 (2019) https://doi.org/10.1097/01.JAA.0000580524.95733.3d

[2] Heer, E., Harper, A., Escandor, N., Sung, H., McCormack, V., Fidler-Benaoudia, M.M.: Global burden and challenges of breast cancer. The British Journal of Radiology **95**(1138), 20211033 (2022) https://doi.org/10.1259/bjr.20211033

[3] Sung, H., Ferlay, J., Siegel, R.L., Laversanne, M., Soerjomataram, I., Jemal, A., Bray, F.: Global cancer statistics 2020: Globocan estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA: A Cancer Journal for Clinicians **71**(3), 209–249 (2021) https://doi.org/10.3322/caac.21660

[4] Arnold, M., Morgan, E., Rumgay, H., Mafra, A., Singh, D., Laversanne, M., Vignat, J., Gralow, J.R., Cardoso, F., Siesling, S., Soerjomataram, I.: Global burden of female breast cancer in 2022 and projections to 2050. Nature Medicine

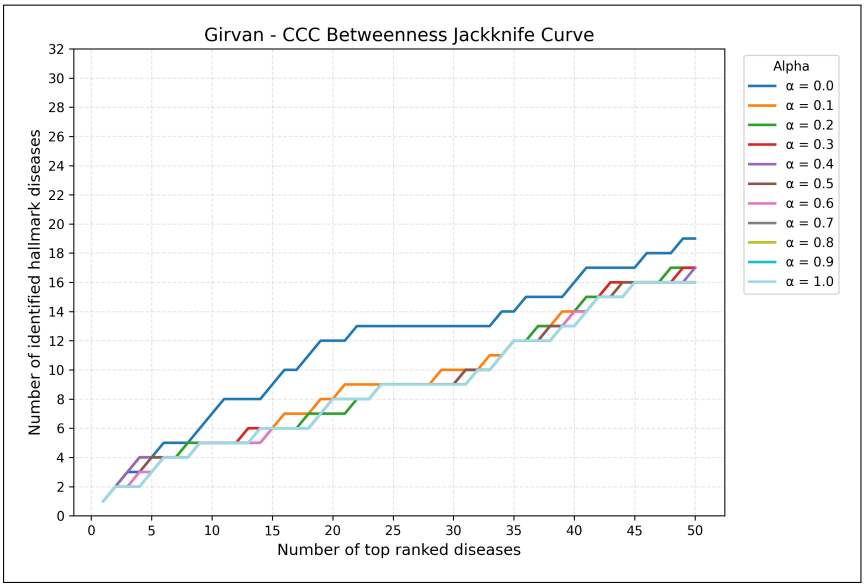


Fig. 23 Jackknife curve for *Community-Consideration Betweenness Centrality* using the Girvan-Newman algorithm.

31, 1574–1581 (2025) https://doi.org/10.1038/s41591-025-03502-3

[5] Pei, Z., *et al.*: Comorbidity and breast cancer survival: a prospective cohort study of breast cancer survivors. Breast Cancer Research and Treatment **145**(2), 411–419 (2014) https://doi.org/10.1007/s10549-013-2521-2

[6] Young, J.M., *et al.*: Comorbidity burden and management among breast cancer survivors. Supportive Care in Cancer (2023) https://doi.org/10.1007/s00520-023-07678-7

[7] Liu, X., *et al.*: Comorbidity burden and risk of second primary non-breast cancer in breast cancer survivors. Cancer Epidemiology (2025) https://doi.org/10.1016/j.canep.2025.102867

[8] Liu, Y., et al.: Comorbidities in multiple sclerosis: A review of modifiable risk factors and their impact on disease outcomes. Therapeutic Advances in Neurological Disorders **18** (2025) https://doi.org/10.1177/17562864251404455

[9] Unknown, A.: Patterns of comorbidity and survival in non-small cell lung cancer. Cancers **17**(9), 1577 (2025) https://doi.org/10.3390/cancers17091577

[10] Peng, T., Chen, S., He, X., *et al.*: Cancer comorbidities: Clinical impacts, mechanistic insights, and management strategies. MedComm – Oncology **5**, 1–20 (2025) https://doi.org/10.1002/mog2.70052

[11] Names, A.: Comorbidity patterns in hospitalized patients with chronic obstructive

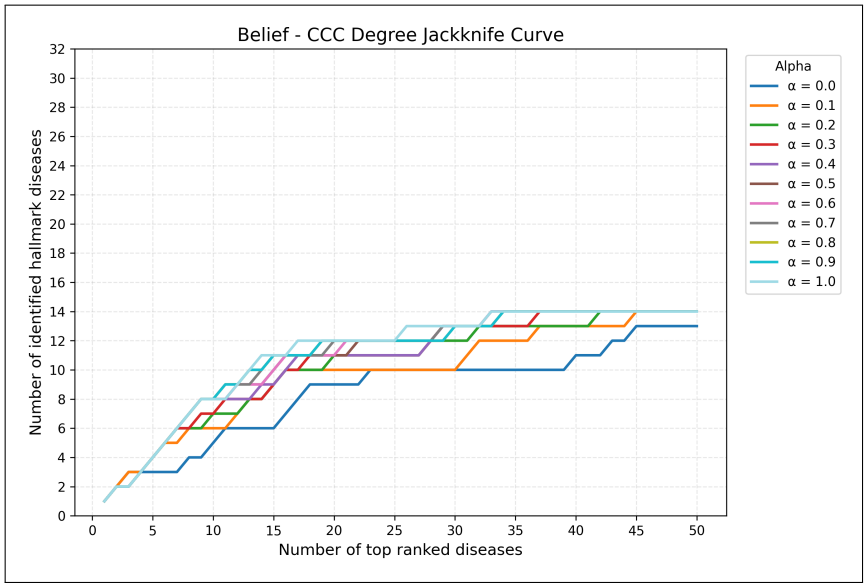


Fig. 24 Jackknife curve for *Community-Consideration Degree Centrality* using the Belief Propagation algorithm.

pulmonary disease in southwest china: a network-based study. Scientific Reports **15**(1), 1–12 (2025) https://doi.org/10.1038/s41598-025-12580-9

[12] Zhao, B., Huepenbecker, S., Zhu, G., Rajan, S.S., Fujimoto, K., Luo, X.: Comorbidity network analysis using graphical models for electronic health records. Frontiers in Big Data **6** (2023) https://doi.org/10.3389/fdata.2023.846202

[13] Chen, C., Wu, H., Yang, L., Kan, K., Zhang, X., Zhang, S., Jia, R., Li, X.: Network analysis of chronic disease among middle-aged and older adults in china: a nationwide survey. Frontiers in Public Health **13** (2025) https://doi.org/10.3389/fpubh.2025.1551034

[14] Lanzer, J.D., Levinson, R.T.: Comorbidity networks of heart failure and heart disease. Current Heart Failure Reports **21**, 1–12 (2024) https://doi.org/10.1007/s11897-024-00693-7

[15] Jin, Y., Sha, S., Tian, T., Wang, Q., Liang, S., Wang, Z., Liu, Y., Cheung, T., Su, Z., Ng, C.H., Xiang, Y.T.: Bridge symptoms and central symptoms in the comorbidity of depression and anxiety: A network analysis. Journal of Affective Disorders **311**, (2022) https://doi.org/10.1016/j.jad.2022.06.051

[16] Rustamaji, H., Kusuma, W., Nurdiati, S., Batubara, I.: Community-consideration centrality, a case study of lung cancer proteins. IEEE Access **PP**, 1–1 (2023) https://doi.org/10.1109/ACCESS.2023.3337375

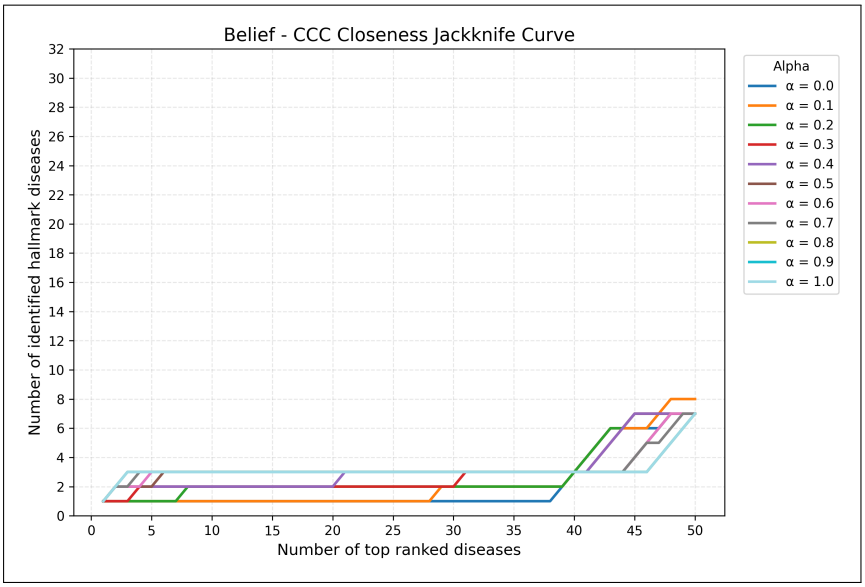


Fig. 25 Jackknife curve for *Community-Consideration Closeness Centrality* using the Belief Propagation algorithm.

[17] Traag, V.A., Waltman, L., Eck, N.J.: From louvain to leiden: guaranteeing well-connected communities. Scientific Reports **9**(1) (2019) https://doi.org/10.1038/s41598-019-41695-z

[18] Girvan, M., Newman, M.E.J.: Community structure in social and biological networks. Physical Review E **69**(2), 026113 (2004) https://doi.org/10.1103/PhysRevE.69.026113

[19] Reichardt, J., Bornholdt, S.: Statistical mechanics of community detection. Physical Review E **74**(1), 016110 (2006) https://doi.org/10.1103/PhysRevE.74.016110

[20] Zhang, P., Moore, C.: Scalable detection of statistically significant communities and hierarchies, using message passing for modularity. Proceedings of the National Academy of Sciences **111**(51), 18144–18149 (2014) https://doi.org/10.1073/pnas.1409770111

[21] Batool, Z., Junaid, M., Naeem, M., Ahmed, M., Shah, L., Saeed, Y., Jehangiri, A., Khan, F.: Identifying disease comorbidity patterns using centrality measures in computing. Revista Gestão Inovação e Tecnologias **11**, 2964–2975 (2021) https://doi.org/10.47059/revistageintec.v11i4.2332

[22] Permana, A.A., Yaputra, R.M.: Analyzing breast cancer comorbidities: a network approach using community detection algorithms. Applied Network Science **9**(1) (2024) https://doi.org/10.1007/s41109-024-00644-0

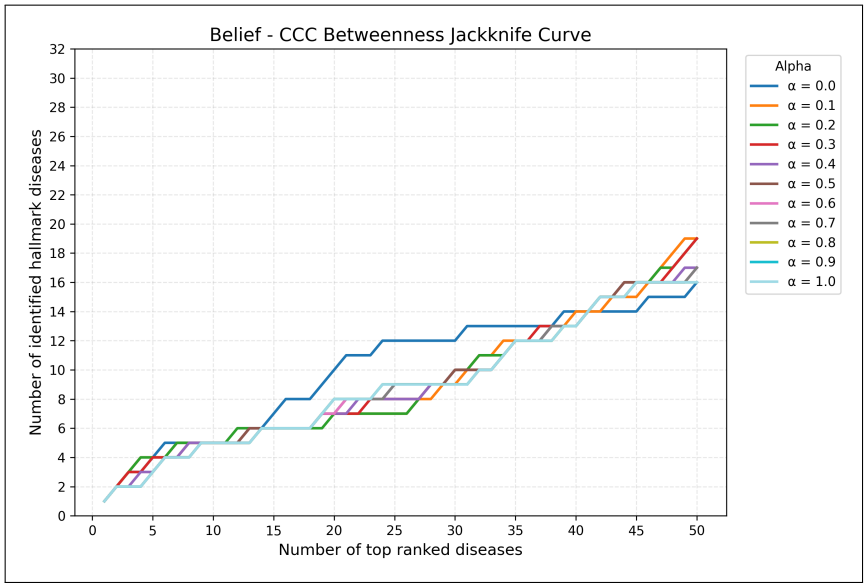


Fig. 26 Jackknife curve for *Community-Consideration Betweenness Centrality* using the Belief Propagation algorithm.

[23] Rustamaji, H.C., Kusuma, W.A., Nurdiati, S., Batubara, I.: Community detection with greedy modularity disassembly strategy (2023) https://doi.org/10.21203/rs.3.rs-2954140/v1

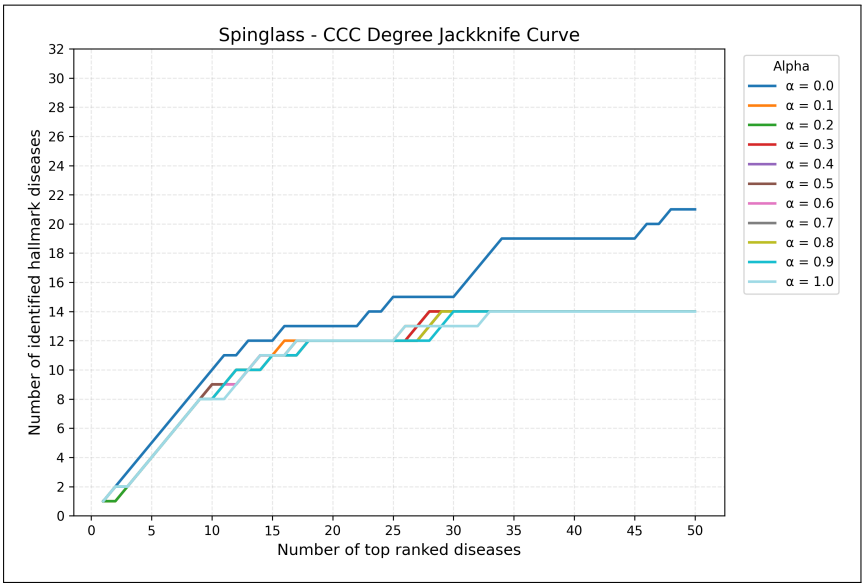


Fig. 27 Jackknife curve for *Community-Consideration Degree Centrality* using the Spinglass algorithm.

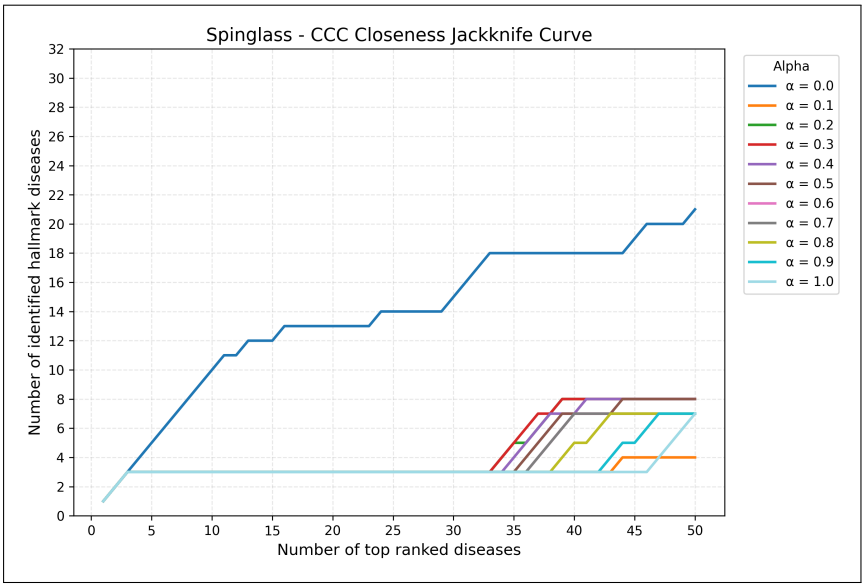


Fig. 28 Jackknife curve for *Community-Consideration Closeness Centrality* using the Spinglass algorithm.

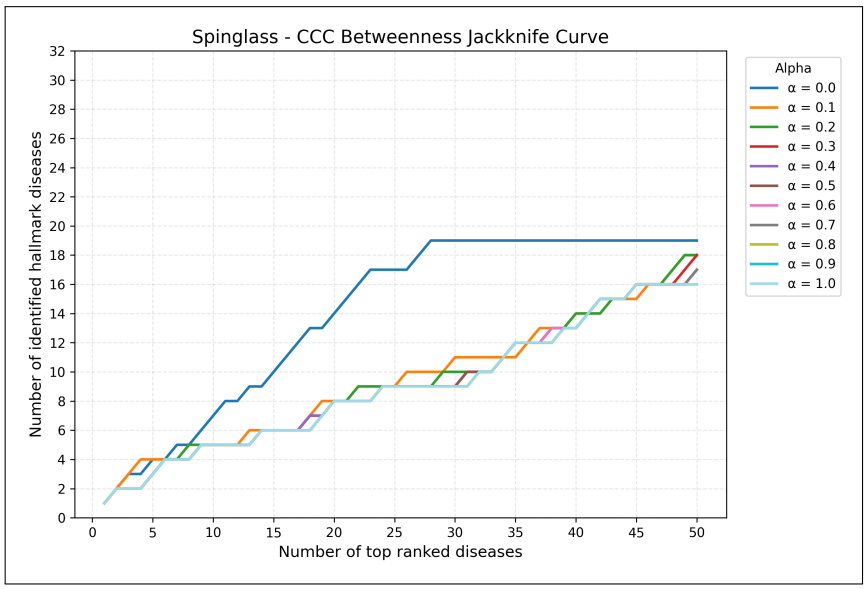


Fig. 29 Jackknife curve for *Community-Consideration Betweenness Centrality* using the Spinglass algorithm.