Identifying Key Nodes and Edges of High-speed Rail Network Based on Minimum Weighted Connected Dominating Set

Wenxia Li, Linzhong Liu, Zhuo Li

***Abstract*—The identification of critical nodes and edges is of great significance for maintaining the normal operation of the high-speed rail network(HSRN). Based on the basic lines of the HSRN and empirical operation data of high-speed trains, this study constructs a weighted service-physical dual-layer network. A heuristic algorithm with immune mechanism and greedy repair strategy is designed based on the constraints of node dominance and connectivity, with the goal of minimizing the network weight ratio. A method for identifying key nodes and edges in the HSRN is proposed, based on the Minimum Weighted Connected Domination Set (MWCDS). Using the Chinese high-speed railway network as a case study, the results show that the backbone network identified by MWCDS consists of key stations and key train flows, and its service scope achieves full coverage of the HSRN. The identified key stations are primarily network hubs, while the key edges exhibit high transportation service efficiency, enabling the simultaneous identification of critical stations and service edges. By leveraging the mapping relationship between the service-physical dual-layer network, trunk railway lines can be derived from an operational perspective. Finally, the effectiveness of MWCDS is verified, demonstrating that the identified backbone network has broad spatial span and coverage, playing a crucial role in maintaining network stability.**

***Index Terms*—High-speed rail network, Minimum Weighted Connected Domination Set, Immune mechanism, Greedy heuristic, Backbone network**

I. INTRODUCTION

AS a critical infrastructure driving national economic development, the high-speed railway network not only reduces intercity travel duration but also significantly enhances regional spatial connectivity. Nevertheless, with continuous network expansion, the system demonstrates increasingly complex, dynamic, and open complex characteristics, rendering it more vulnerable to operational risks and external disruptions. This evolving context makes conducting structural analysis of network components particularly crucial to identify critical components that ensure sustainable long-term operation - a pressing challenge requiring immediate resolution in contemporary railway maintenance and management practices.

The rapid advancement of network science has driven substantial methodological progress in critical component identification. Current approaches are systematically taxonomized into two dominant frameworks: social network analysis [1] and system network analysis [2]. Social network analysis measures the importance of node/edge based on its' topologicalmetrics, with established techniques including network centrality [3], K-shell decomposition [4], and PageRank algorithms [5]. These methods demonstrate robust applicability in transportation networks. Du et al. [6] analyzed spatiotemporal characteristics of urban commuting passenger flow to evaluate topological criticality in urban rail systems. Cheng et al. [7] developed flow-weighted centrality metrics capturing passenger delay propagation. Kim et al. [8] formulated an aviation network model with distance and passenger flow demand as edge weights and identified a set of key nodes. In addition, inspired by traditional social network analysis methods, many researchers have applied methods such as evidence theory [9], structural hole theory [10], and gravity model [11] to identify key components of the network. System network analysis posits that component importance equates to induced functionality degradation. Researchers quantify this through pre/post-removal comparisons of Giant connected component size [12] ,Average path length variance [13] ,Global efficiency reduction [14]. Xin et al. [15] examined the robustness of China's railway transportation network under different attack strategies and identified its backbone network structure. He et al. [16] analyzed the interdependence of multimodal transportation networks and identified key nodes in the Dutch freight network, adopting total travel time during disruptions as the network performance metric.. Li et al. [17] integrated topological characteristics with operational functionality to investigate how station and train failures affect the service efficiency of High-Speed Rail Networks (HSRN).

The above methods evaluate the importance of network components from different perspectives, but they have the following shortcomings: they study nodes and edges separately as independent individuals, thereby losing the correlation between nodes and edges, and are unable to systematically identify the network's backbone structure. More importantly, these methods may overlook components with low centrality that play a crucial role in maintaining the network's structural integrity. The Connected Dominating Set

Manuscript received November 11, 2024; revised April 29, 2025.

This work was supported in part by the National Natural Science Foundation of China (Nos.71361018,71671079).

Wenxia Li is a PhD candidate at School of Traffic and Transportation, Lanzhou Jiaotong University, Lanzhou 730070, China. (e mail: lwx2712196961@126.com).

Linzhong Liu is a professor at School of Traffic and Transportation, Lanzhou Jiaotong University, Lanzhou 730070, China. (Corresponding author, e-mail: liulinzhong@tsinghua.org.cn).

Zhuo Li is a PhD candidate at School of Traffic and Transportation, Lanzhou Jiaotong University, Lanzhou 730070, China. (e mail: zhuo_li1023@163.com).

(CDS) identifies key nodes and edges by integrating dominance and connectivity, thereby offering a novel approach to uncovering backbone structures. Dominance reflects the control ability of nodes over the network, and connectivity ensures the correlation between nodes. However, the structure complexity of network brings about a massive amount of network data, and maintaining a few key components to support the normal operation of the network is of great significance. Theoretically, the Minimum Connected Dominating Set (MCDS) provides a programmatic method for constructing connected network topologies [19]. It achieves full network domination via minimal key nodes while preserving inter-node connectivity. MCDS can efficiently extract backbone structures in large-scale complex networks [20]. Studies indicate that unweighted networks fail to capture real-world system complexities, whereas weighted networks better represent actual network attributes.. To address this, Wang et al. [21] proposed the Minimum Weight Connected Dominating Set (MWCDS) method. The MWCDS identifies backbone structures in weighted networks under dominance and connectivity constraints, thereby yielding effective outcomes in real-network analysis. The successful implementation of MCDS in transportation networks [22] informs the methodological foundation of this study. Therefore, based on previous research, we propose applying MWCDS to identify backbone structures in HSRN. This article analyzes the service and topological characteristics of HSRN, which evaluates inter-station transportation efficiency through train operating frequency and travel time, and constructs a service-physical dual-layer network. To reduce algorithmic search space in large-scale networks, an immune mechanism combined with a greedy heuristic strategy is introduced. An immune greedy heuristic (IGH) algorithm was proposed to develop a problem-solving approach for MWCDS. Taking the Chinese high-speed railway network as a case study, we leverage actual train operation data to identify its backbone structure, simultaneously pinpoint key nodes and edges. The proposed MWCDS method is compared with traditional identification approaches to validate its effectiveness.

II. Methodology

A. MWCDS Basic Theory

As the theoretical foundation for identifying key nodes and edges in weighted complex networks, the MWCDS concept has evolved from its predecessors the MDS and MCDS.

In undirected graph G, V is the set of nodes and E is the set of edges.

Definition 1. Minimum Dominating Set (MDS)

In undirected graph $G=(V,E)$, $S\subseteq V$, $S\neq\varnothing$, if $\forall v\in V-S$ is directly connected to at least one node in S, then S is the dominating set of graph G. If $\forall\widehat{S}\subset S$ does not constitute the dominating set of graph G, then S is the minimum dominating set of graph G[23]. $|S|$ is the minimum number of dominant nodes. As shown in Fig. 1 (a), $\{v_2,v_7,v_{11}\}$ forms dominant relationships with other nodes.

Definition 2. The Minimum Connected Dominating Set (MCDS)

If S is the minimum dominating set of graph $G=(V,E)$ and S cannot form a connected graph, then the minimum node set $V'\subset(V-S)$ is introduced to ensure that S forms a connected graph, then set $C=V'\cup S$ forms the minimum connected dominating set of graph G, and $|C|$ is the minimum number of connected dominating nodes in graph G. As shown in Fig 1 (b), $\{v_2,v_7,v_9,v_{10}\}$ forms a dominant relationship over all nodes and forms a connected subgraph.

The fundamental distinction between the Minimum Connected Dominating Set (MCDS) and Minimum Dominating Set (MDS) resides in the connectivity requirement among dominating nodes. This implies that the MCDS not only identifies node dominance but inherently requires connected paths between these nodes. Specifically, the MCDS constructs a connected subnetwork capable of dominating all nodes in the original network through immediate adjacency. Consequently, this subnetwork achieves full network coverage without requiring multi-hop propagation, ensuring structural stability with minimal node deployment while maintaining continuous network operability.

It should be noted that the MCDS is often applied to unweighted networks, and this application of the MCDS is not unique. As shown in Figure 1 (b), both the nodes set $\{v_2,v_7,v_9,v_{10}\}$ and the nodes set $\{v_2,v_4,v_5,v_7\}$ can form the MCDS of the network. However, the unweighted networks only reflect the connection relationships between various parts of the network, but cannot reflect the strength of the connections. In the weighted networks, nodes or edges are weighted through passenger flow, travel time, traffic capacity, etc., which is closer to the real transportation network. Therefore, Wang et al.[21]proposed the concept of the MWCDS on the basis of the MCDS for the identification of weighted network backbone structure.

Definition 3. The Minimum Weight Connected Dominating Set (MWCDS)

For a given node weighted graph $G_T=(V,E,w)$ where w is a function $w:V\to R^+$, the MWCDS problem is to find a MCDS with the minimum total weight $W_T(V)=\sum_{v_i\in C}w(v_i)$.

As shown in Fig 1 (c), the total weight of nodes in the $\{v_2,v_7,v_9,v_{10}\}$ is 5+7+6+3=21, and the total weight of nodes in the $\{v_2,v_4,v_5,v_7\}$ is 5+5+4+3=17. Therefore, $\{v_2,v_4,v_5,v_7\}$ is the MWCDS of graph G_T. It can be inferred that the MCDS of any graph may not be unique, but the MWCDS is unique.

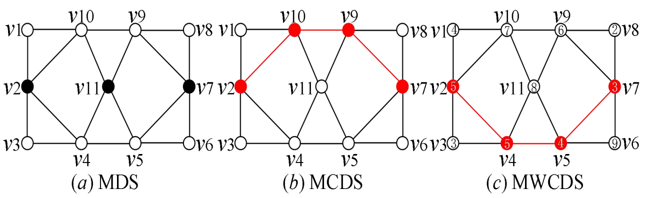


Fig. 1. Diagram of MDS, MCDS and MWCDS

B. Two-layer network model of the HSRN

Transportation service quality is typically measured by timeliness and convenience. Timeliness is represented by travel time $t_{i,j}$, convenience is represented by train operating frequency $f_{i,j}$. The edge weight in the service network is mathematically defined as $w_{i,j} = t_{i,j} / f_{i,j}$, indicating that shorter travel durations and higher service frequencies correspond to lower edge weights – a direct reflection of enhanced service quality. Based on this, the HSRN service network is modeled as a weighted network $G_s = (V, E_s, W)$. Regardless of the presence of multiple high-speed rail stations within a city, only the city represents the network nodes, V is the set of city nodes. If there are direct trains between cities, there are service edges between them, E_s is the set of service edges, and W is the set of service edge weights. The HSRN physical network is modeled as an unweighted network $G_p = (V, E_p)$, where V is the set of city nodes and E_p is the set of railway line edges. If there are railway lines directly connected between adjacent cities, there are edges between city nodes. The HSRN weighted service-physical dual-layer network is shown in Fig 2.

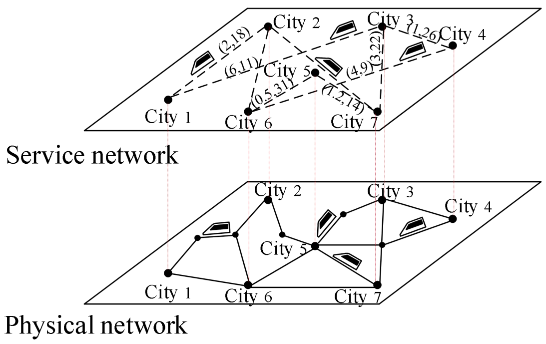


Fig. 2. High-speed railway service-physical two-layer network.

III. ALGORITHM DESIGN

The essence of solving the MWCDS problem is to identify a subset of nodes in the network that satisfies the requirements of dominance and connectivity while minimizing the total node weight. To avoid the impact of urban location advantages and hub functions on MWCDS identification results, this paper excludes node weights and focuses solely on service edge weights. The MWCDS problem is a typical NP-hard problem, and precise algorithms struggle to obtain optimal solution in polynomial time. Due to their low computational complexity and acceptable computation time, heuristic algorithms have gradually become the mainstream approach for large-scale complex networks.

Considering the large scale and high complexity of the HSRN, this paper adopts selection, destruction, repair, and minimization as core mechanisms of population evolution, incorporating an immune mechanism and a greedy repair strategy to propose a IGH algorithm for MWCDS identification.

A. Encoding and population initialization

Encoding of the solution: The essence of the MWCDS problem is the selection of N-dimensional vectors generated through 0-1 encoding to form a feasible solution *X*, where $x_i = 1$ indicates that node i is selected as a dominant node.

Population initialization: Analysis of the HSRN reveals that two types of nodes must be designated as dominating nodes, namely the cut nodes and the neighboring nodes of marginal nodes. Cut nodes play a critical role in maintaining connectivity between subgraphs, while marginal nodes are dominated by their neighbors. These two categories of nodes are fixed as dominating nodes and can serve as immune antibodies to initialize the population, thereby reducing the algorithm's solution search space.

For the cut nodes, first provide a definition and a proposition.

Definition : If v^* is a cut node of an undirected connected graph $G = (V, E)$, then there are two nodes u and w that are different from v^*, so that v^* is on each path of u and w.

Proposition : In undirected connected graph $G = (V, E)$, the set $C \subset V$ constitutes the minimum connected dominant set of graph *G*. If there is a cut node $v^* \in V$ in graph *G*, then the cut node $v^* \in C$.

Proof: Assuming that $C\text{-}v^*$ is a connected graph, there is at least one path between u and w in $C\text{-}v^*$, so there is a path between u and w in *C* that does not pass through v^*. This contradicts definition 1, so $C\text{-}v^*$ is an unconnected graph, which contradicts the hypothesis.

From the above proposition proof, it can be seen that the cut node v^* must be the dominant node.

There are two situations in the network for the neighboring nodes of the marginal nodes:

① as shown in Figure 3 (a), the marginal node v_1 has only one neighboring node v_2, and can only form a dominant relationship with v_1 through v_2. Therefore, v_2 must be the dominant node.

② As shown in Figure 3 (b), marginal node v_1 and its neighboring nodes $\{v_2, v_3, v_4\}$ form a fully connected subgraph, and can only form a dominant relationship with v_1 through one of the neighboring nodes $\{v_2, v_3, v_4\}$. Therefore, based on greedy thinking, v_4 with the highest degree value is selected as the dominant node.

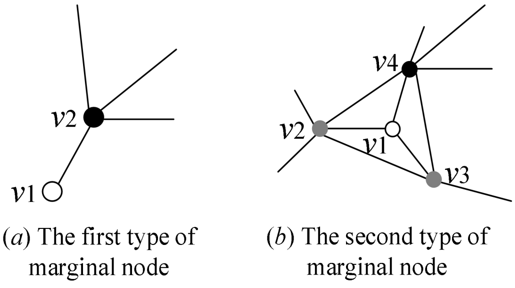


Fig. 3. Two types of marginal nodes and their neighboring nodes.

B. Selection and destruction

Selection: In the population, each individual encodes a feasible solution. The fitness value of an individual is inversely proportional to the size of the CDS derived from decoding its representation(e.g., Fitness=1/$|\text{CDS}|$). Specifically, a smaller CDS size results in a higher fitness value. To select high-quality solutions, we apply tournament selection, where individuals with higher fitness values are chosen as parent candidates for the next generation.

Destruction: Guided by elite individuals, using XOR logic operations to determine the consistency between parent individual X_i and elite individual X_j^* in dimension k, and calculating the similarity $sim(X_i, X_j^*)$ between X_i and X_j^*, finally obtaining the mutation parameter ϕ_k of individual X_i in dimension k. If $rand(0,1) > \phi_k$, a mutation operation is performed on dimension k for the parent individual X_i, traversing each dimension of individual X_i to generate a new offspring individual.

The similarity between individuals is calculated using the following formula:

$$sim(X_i, X_j^*) = \frac{N_{11}}{N_{11} + N_{01} + N_{10}} \qquad (1)$$

In the formula, N_{11} represents the number of dimensions where individuals X_i and X_j^* are the same in their corresponding dimensions and both are 1. Both N_{01} and N_{10} represent the number of dimensions that are different in the corresponding dimension, namely the combination of 0, 1 and 1,0. When the similarity $sim(X_i, X_j^*) = 1$, the individual is reinitialized to ensure the population diversity.

The mutation parameter of an individual on dimension k is calculated using the following formula:

$$\phi_k = sim(X_i, X_j^*) \cdot (x_{i,k} \oplus x_{j,k}) \qquad (2)$$

In the formula, $x_{i,k}$ and $x_{j,k}$ are the values of individual X_i and elite individual X_j^* on dimension k, namely 0 or 1, respectively.

C. Repair and minimization

Repair：The repair mechanism first checks whether the offspring individuals meet the constraints of dominance and connectivity. For offspring violating these constraints, three greedy heuristic-based repair strategies are applied:

Strategy 1: Select the node with the smallest edge weight ratio:

$$\arg\min Gr(v_j) = \frac{w_{i,j}}{\sum w_{j,k}}; \forall i \in S,\ \forall j,k \in (V-S) \quad (3)$$

Strategy 2: Select the node with the smallest edge weight:

$$\arg\min Gw(v_j) = w_{i,j}; \forall i \in S,\ \forall j \in (V-S) \quad (4)$$

Strategy 3: Select the node with the highest number of non dominant nodes in the neighborhood:

$$\arg\max Gn(v_j) = |\chi(v_j)|;\ \forall j \in (V-S) \quad (5)$$

Among the three strategies mentioned above, $w_{i,j}$ is the weight of edge $e_{i,j}$, S is the set of dominant nodes, $(V-S)$ is the set of non dominant nodes, and $\chi(\bullet)$ is the statistical function of the number of non dominant nodes in the node's neighborhood.

Minimization: After repair operations, a connected dominating set containing redundant nodes is constructed. To build the MCDS, we delete redundant nodes to minimize the number of dominating nodes. Experiments show that MCDSs in networks are not unique, exhibiting varying levels of connectivity. Those MCDSs with better connectivity tend to have more edges. However, when obtaining the MWCDS by minimizing the sum of edge weights, MCDS candidates with good connectivity may be excluded. To address this, we propose calculating the network weight ratio for each MCDS candidate. This ratio decreases with both a smaller sum of edge weights and more connected edges, which ultimately leads to higher transportation service levels and improved connectivity in the selected MWCDS. The calculation formula for the network weight ratio is as follows:

$$Wr = \frac{\sum_{i \neq j \in C} w_{i,j}}{|\xi(C)|} \qquad (6)$$

In the formula, C is MCDS, $w_{i,j}$ is the edge weight between node i and node j, and $\xi(\bullet)$ is the statistical function of the number of edges in the MCDS.

The IGH algorithm process is shown in Fig 4.

IV. Simulation analysis

The simulation was conducted on a system equipped with an AMD Ryzen 7 4800U processor (8-core, 1.8GHz base clock) and 16GB RAM. The algorithm implementation was executed in MATLAB R2016b (MathWorks, 2016).

A. Experiments on scale-free artificial networks

To verify the effectiveness of the MWCDS-based identification method, we constructed an undirected scale-free weighted network $G_1 = (V_1, E_1, W_1)$ using Pajek for validation. The test network comprises 100 nodes and 427 edges, with edge weights uniformly randomized in [1,10].. To ensure the accuracy of the solution, set the population size *pop*=100, the maximum number of iterations *Iter*=300. The program was executed 20 times independently, where nodes and edges recurring across results were identified as key elements, yielding one MWCDS set and two MCDS sets. The experimental results are shown in Table I. The MWCDS is shown in Fig 5.

TABLE I
Experimental results of scale-free artificial networks

Type	Node set of backbone network	Number of nodes	Number of edges	Weight ratio *Wr*
MWCDS	{1,2,4,7,8,11,14,17, 20,52,58}	11	18	0.7000
MCDS-1	{1,2,4,5,7,8,11,14,1 7,20,52}	11	18	1.0091
MCDS-2	{1,2,4,7,8,11,14,17, 20,40,52}	11	17	0.8364

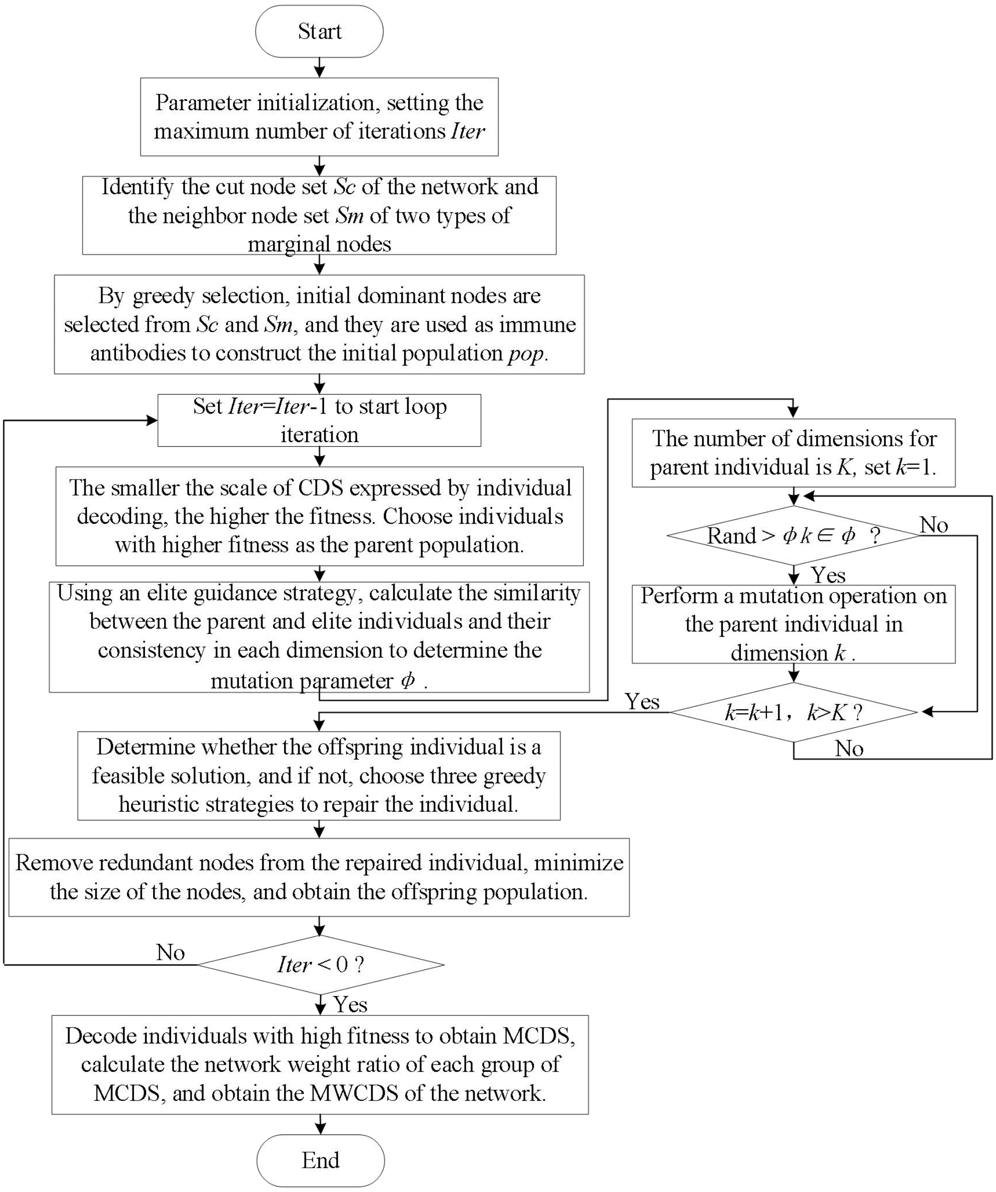


Fig. 4. IGH algorithm flow chart.

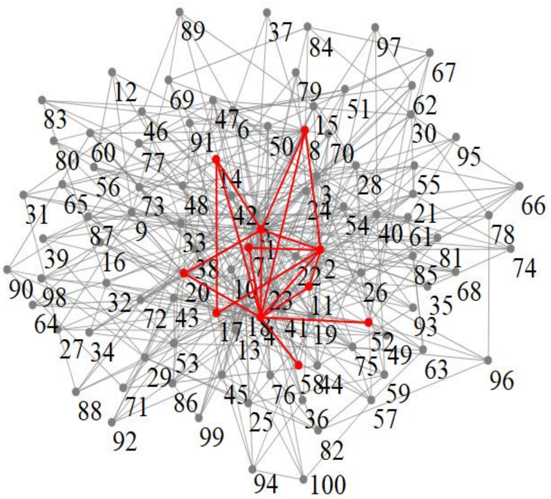


Fig. 5. The MWCDS of scale-free artificial networks.

Scale-free networks have become a fundamental paradigm for modeling real-world complex systems. As shown in Table I and Fig 5, three sets of MCDS were identified in a scale-free artificial network consisting of 100 nodes and 427 edges, among which the MCDS with a network weight ratio of 0.7 was named MWCDS. MWCDS only achieved dominance over the other 89 nodes in the network through a backbone network consisting of 11 nodes and 18 edges. It can be seen that the MWCDS identification method can accurately identify the backbone network of scale-free networks.

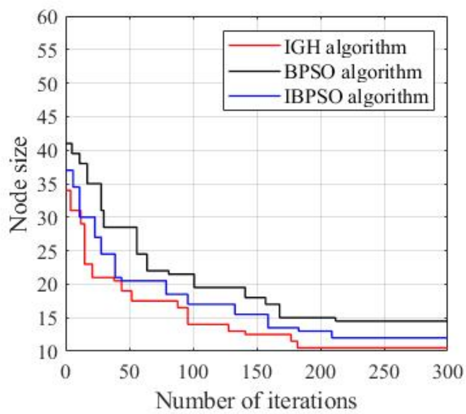


Fig. 6. Algorithm convergence curve.

TABLE II
STATISTICAL RESULTS OF ALGORITHM EXPERIMENTS

Type	Optimal	Average	Average deviation /%	Maximum deviation /%	Average calculation time/s
BPSO	19	20.8	9.47	21.10	11.7
IBPSO	14	15.1	7.85	14.28	8
IGH	11	11.6	5.45	18.18	7.1

We compared the performance of the IGH algorithm with the BPSO and IBPSO algorithms from reference 22, running each 10 times. The convergence curve of the IGH algorithm is shown in Fig 6, and the experimental statistical results are presented in Table II. The average deviation is calculated as (average solution−optimal solution)/optimal solution, while the maximum deviation is defined as (worst solution−optimal solution)/optimal solution. Due to its immune mechanism and three greedy heuristic strategies, the IGH algorithm converges after 182 iterations and obtains an MWCDS with a size of 11 nodes, outperforming both the BPSO and IBPSO algorithms in solution quality and convergence speed. The IGH algorithm achieves superior performance in terms of the optimal solution, average solution, average deviation, and average computation time compared to the other two algorithms, with only the maximum deviation being slightly higher than that of the IBPSO algorithm. The proposed IGH algorithm demonstrates high solution accuracy when applied to large-scale networks.

B. Applications

Data sources

Taking China's high-speed railway network as a study case, the study area is limited to the Chinese Mainland. In addition, since the Lalin Railway in Xizang Autonomous Region is not connected with the high-speed railway network, and the Hainan Roundabout Railway operates independently of the high-speed railway network, therefore the study scope only includes 21 other provinces, 4 autonomous regions and 4 municipalities directly under the Central Government. We select new railway lines with a design speed of 250km/h or higher and upgraded railway lines with a design speed of 160 km/h or higher as network edges, along with 257 major cities along these railways as network nodes, to construct the physical structure of the HSRN. Based on this physical network and the train classification codes from the October 23, 2022 timetable, we extracted operational data for trains prefixed with G, D, and C to build a weighted HSR service network. This resulted in a total of 7,551 service edge pairs. All data were obtained from the China Railway Customer Service Center (http://www.12306.cn/index/).

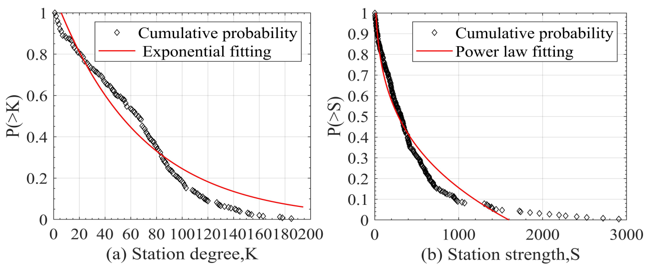


Fig. 7. Cumulative distributions of an HSR service network.

The scale-free property of the network indicates that the failure of a few network components has a significant impact on the overall performance of the network. The scale-free characteristics of the physical and service networks of high-speed railways were studied, and Fig 7 shows the cumulative distribution of station degree and station strength. Obviously, the degree distribution can be fitted with an exponential function $P(>K) \propto 1.093e^{-0.015K}$, while the strength distribution tends towards a power-law distribution $P(>S) \propto -0.275S^{-0.231} + 1.512$. The above results indicate that compared to physical networks, service networks considering train flow have scale-free characteristics, and it is necessary to pay attention to the supporting role of hub stations and critical edges on the overall network. Therefore, this study identifies key nodes and edges through the proposed the MWCDS.

Result analysis

Using the proposed method to identify the backbone network of the HSR service network, four sets of the MCDS are obtained, including one set of the MWCDS. The results are shown in Table III.

TABLE III
EXPERIMENTAL RESULTS OF CHINA'S HIGH SPEED RAILWAY NETWORK

Type	Node set of backbone network	Number of nodes	Weight ratio *Wr*
MWCDS	{Beijing, Changsha, Hangzhou, Xi'an, Chengdu, Zhengzhou, Lanzhou}	7	2.5051
MCDS-1	{Beijing, Changsha, Hangzhou, Xi'an, Chengdu, Zhengzhou, Xining}	7	2.6807
MCDS-2	{Beijing, Changsha, Shenzhen, Xi'an, Chengdu, Zhengzhou, Xining}	7	3.4746
MCDS-3	{Beijing, Changsha, Hefei, Xi'an, Chengdu, Zhengzhou, Lanzhou}	7	2.6617

HSRN exhibit stronger scale-free characteristics than artificially generated scale-free network. As demonstrated in Table III, the backbone of the HSR service network can be formed by merely 7 hub city nodes and their intercity train flows. The 7 nodes form a dominant relationship with other nodes in the network, playing a role in controlling the entire network. The trains running between the 7 nodes achieve inter-city accessibility and ensure the connectivity of the backbone network.

Among the four MCDS, the MWCDS represents an optimal backbone network with superior service quality. Assuming that all nodes can only achieve accessibility through the backbone network, the transportation efficiency from each node to other nodes is calculated by $TE(v_i) = \frac{1}{N-1}\sum_{j,j\neq i} f_{ij}/t_{ij}$, and the transportation efficiency of the network is calculated by $TE = \frac{1}{N(N-1)}\sum_i \sum_{j,j\neq i} f_{ij}/t_{ij}$. The comparison results of transportation efficiency are shown in Fig 8 and Fig 9.

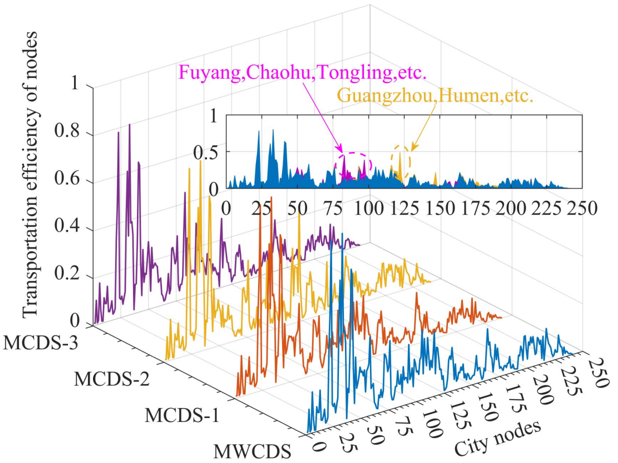


Fig. 8. Transportation efficiency of each station in the network (compare three sets of the MCDS with the MWCDS).

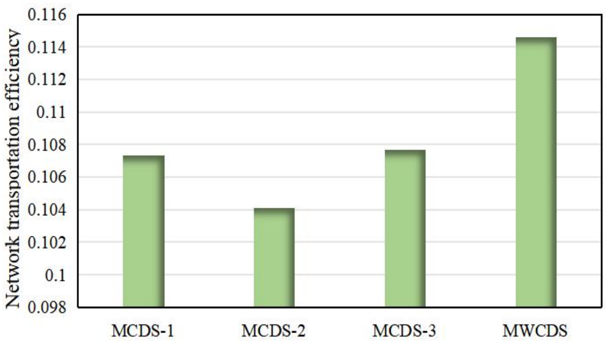


Fig. 9. Network transportation efficiency (comparing three sets of the MCDS with the MWCDS)

In MCDS-1 and MCDS-2, the number of originating and terminating trains at Xining Station is relatively small, and the travel time to other stations is longer. Moreover, Xining only operates direct trains with Xi'an, resulting in lower connectivity of the backbone network and lower network transportation efficiency. As shown in Fig 8, compared with the MWCDS, MCDS-2 has improved the transportation efficiency of some stations closely connected to Shenzhen, such as Guangzhou and Humen. MCDS-3 has improved the transportation efficiency of some stations with Hefei as a transit hub, such as Fuyang, Chaohu, Tongling, etc., but both have led to a decrease in network transportation efficiency. The backbone structure composed of the MWCDS is a complete graph, with direct trains running between any two nodes. The MWCDS meets the requirements of dominance and connectivity while also taking into account the quality of transport services between stations, thus achieving the highest network transportation efficiency. As shown in Fig 9, compared with the three groups of the MCDS, the network transportation efficiency of the MWCDS increased by 10.1%, 6.8%, and 6.4%, respectively.

The MWCDS identifies HSR backbone networks based on upper-level service networks and can determine key nodes and edges from the perspective of transport service functionality. These identification results are crucial for ensuring the normal operation of the railway network. By mapping the MWCDS of the upper service network to the lower physical network, we can identify key railway lines that support HSRN operations. The results are shown in Fig. 10. Key vertical passages include the Beijing-Shanghai Passage, the Beijing-Hong Kong Passage (Beijing-Guangzhou Section), and the Baotou-Haikou Passage (Xi'an-Chengdu Section).Key horizontal passages comprise the Land Bridge Passage (Zhengzhou-Lanzhou Section), the Riverside Passage, and the Shanghai-Kunming Passage.

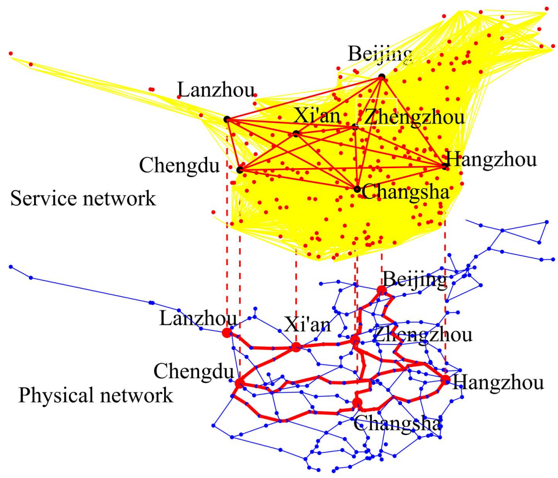


Fig. 10. service-physical two-layer network mapping relationship

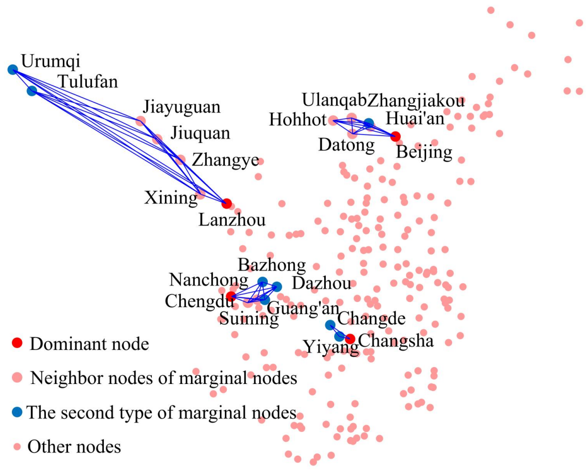


Fig. 11. The community of the second type of marginal nodes and their neighboring nodes

Fig 11 illustrates the fully connected community structure comprising the second type marginal nodes and their neighboring nodes in the HSRN. Beijing connects the nation's primary hub nodes, thereby achieving full coverage in the Northeast and North China regions, while also linking to the marginal node Zhangjiakou. Changsha functions as a central-southern transit hub, connecting the two marginal nodes, Yiyang and Changde. Chengdu achieves full coverage of the southwest region through connections to three marginal nodes: Bazhong, Dazhou, and Guang'an. As a strategic node along the Silk Road, Lanzhou serves Gansu and Qinghai provinces as well as the Xinjiang Uygur Autonomous Region, and is linked to the marginal nodes Urumqi and Turpan. The second type marginal nodes can directly access nodes within their community but require transfers via neighboring nodes to reach external nodes, leading to limited network-wide accessibility. Therefore, identifying these nodes and establishing direct train services between them and external hub nodes would significantly improve their accessibility.

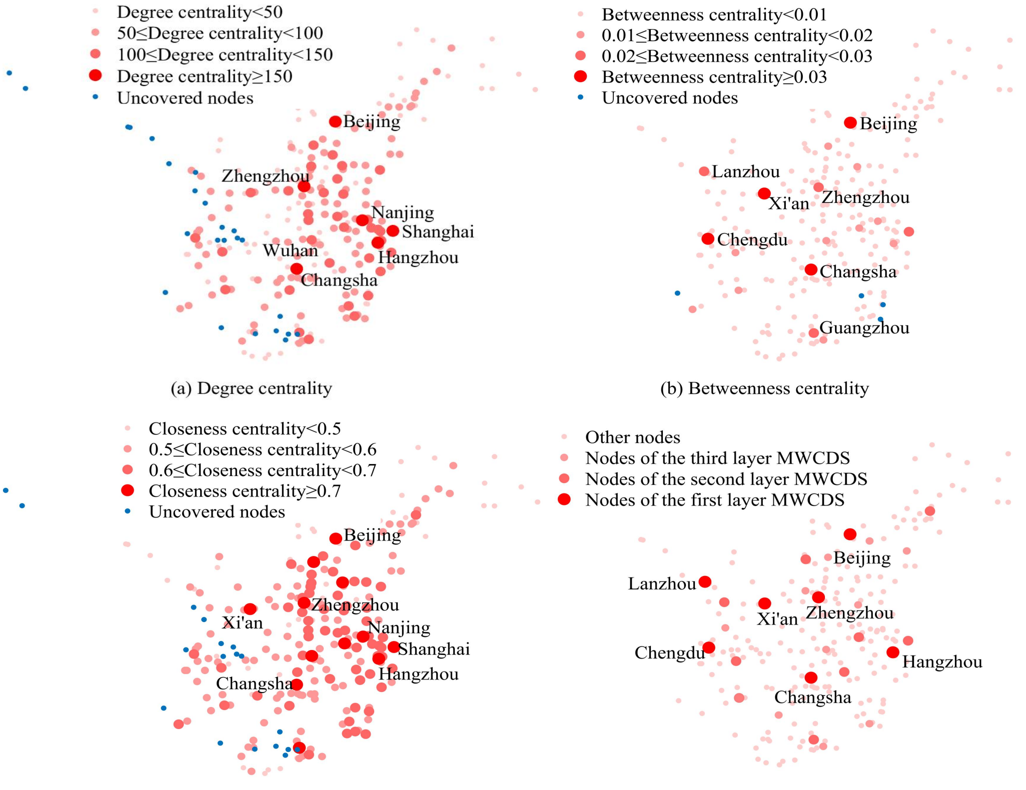
C. *Effectiveness Analysis of MWCDS*

To validate the effectiveness of the MWCDS identification method, three classic indicators in social networks was compared with the MWCDS, namely degree centrality, betweenness centrality and closeness centrality. The comparative results are shown in Table IV.

In Table IV, the key nodes identified by the MWCDS exhibit partial overlap with those derived from other methods, quantitatively validating the effectiveness of the MWCDS in identifying critical nodes within HSRN. The results demonstrate that the findings of the MWCDS and those from betweenness centrality share a higher degree of overlap, reaching 85%. This indicates that the MWCDS incorporates both local and global importance to achieve network dominance and controllability, thereby identifying key nodes have richer connotations.

TABLE IV
COMPARISON RESULTS

No.	Degree centrality	Betweenness centrality	Closeness centrality	MWCDS
1	Shanghai (0.767)	Zhengzhou (0.085)	Zhengzhou (0.811)	Zhengzhou
2	Nanjing (0.717)	Chengdu (0.046)	Changsha (0.779)	Chengdu
3	Zhengzhou (0.696)	Changsha (0.041)	Nanjing (0.767)	Changsha
4	Hangzhou (0.667)	Xi'an (0.032)	Shanghai (0.750)	Xi'an
5	Wuhan (0.642)	Beijing (0.025)	Hangzhou (0.736)	Beijing
6	Changsha (0.633)	Guangzhou (0.023)	Beijing (0.732)	Hangzhou
7	Beijing (0.604)	Lanzhou (0.023)	Xi'an (0.704)	Lanzhou



(a) Degree centrality (b) Betweenness centrality

(c) Closeness centrality (d) MWCDS

Fig. 12. Comparison of key node distribution and coverage.

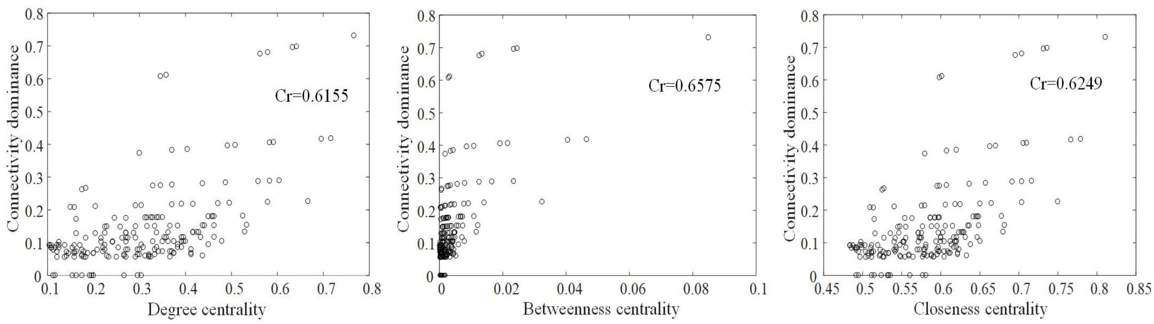


Fig. 13. Correlation analysis between connectivity dominance and three centrality metrics.

TABLE V
COMPARISON OF KEY EDGE IDENTIFICATION RESULTS

Method	Key edges
Edge betweenness	**Beijing-Lanzhou; Beijing-Chengdu; Beijing-Zhengzhou; Changsha-Lanzhou; Beijing-Changsha**; Guangzhou-Lanzhou; Shenyang-Chengdu; **Changsha-Chengdu;** Beijing-Chongqing; Beijing-Guangzhou; Chongqing-Chengdu; Beijing-Guiyang; **Chengdu-Lanzhou; Changsha-Xi'an;** Chengdu-Xining; **Kunming-Xi'an; Chengdu-Xi'an; Zhengzhou-Changsha**; Zhengzhou-Chengdu; **Beijing-Xi'an**
Bridgeness index[24]	**Beijing-Zhengzhou; Zhengzhou-Xi'an; Beijing-Lanzhou; Beijing-Chengdu; Changsha-Lanzhou; Changsha-Xi'an; Beijing-Changsha**; Guangzhou-Lanzhou; Zhengzhou-Guangzhou; Jinan-Nanjing; **Changsha-Chengdu;** Nanjing-Changsha; Zhengzhou-Shijiazhuang; **Chengdu-Lanzhou; Chengdu-Xi'an; Zhengzhou-Lanzhou;** Beijing-Xuzhou; **Zhengzhou-Chengdu; Beijing-Xi'an; Hangzhou-Xi'an**
MWCDS	**Beijing-Lanzhou; Beijing-Chengdu; Beijing-Zhengzhou; Changsha-Lanzhou; Beijing-Changsha; Changsha-Chengdu; Chengdu-Lanzhou; Zhengzhou-Xi'an; Chengdu-Xi'an; Beijing-Xi'an; Hangzhou-Xi'an; Changsha-Xi'an; Zhengzhou-Chengdu;** Hangzhou-Lanzhou; **Zhengzhou-Changsha;** Hangzhou-Chengdu; Xi'an-Lanzhou; Hangzhou-Zhengzhou; Beijing-Hangzhou; Hangzhou-Changsha

Note: Bold indicates coincide edges.

From the spatial distribution of key nodes shown in Fig. 12, those with high degree centrality and closeness centrality are predominantly located in eastern and central regions, forming vertical clusters along the Beijing-Shanghai and Beijing-Guangzhou high-speed railways. In contrast, nodes identified through betweenness centrality and the MWCDS demonstrate more uniform network distribution, primarily situated at hubs where major railway lines intersect. Regarding network coverage, a comparison between the top 7 nodes identified by three centrality measures and those selected by the MWCDS reveals significant differences. As shown in Figures 12 (a), 12 (b), and 12 (c), the top 7 key nodes with the highest degree centrality, closeness centrality, and betweenness centrality achieve network coverage rates of 90%, 92.5%, and 98.3%, respectively, with corresponding counts of unconnected nodes being 24, 18, and 4. Compared with the above three indicators, the MWCDS takes dominance and connectivity as the identification elements, taking into account the local connectivity and global hub of nodes during the identification process. As shown in Fig 12 (d), the 7 key nodes have a wide span and large coverage, achieving the goal of controlling the entire HSRN with a few key nodes.

To validate the scientific rigor of the MWCDS identification method, we applied this approach for hierarchical analysis of all stations in the HSRN, obtaining their connectivity dominance metrics. Pearson correlation coefficient is used to analyze the correlation between connectivity dominance and degree centrality, betweenness centrality, and closeness centrality. From Fig 13, it can be seen that the Pearson correlation coefficients between connectivity dominance and the three centrality evaluation indicators are all over 0.6, with a Pearson correlation coefficient of 0.6575 between connectivity dominance and betweenness centrality. It can be seen that connectivity dominance has a good correlation with all three centrality indicators, reflecting that the MWCDS considers both local and global importance of nodes while ensuring complete control of the network, thereby serving as a comprehensive evaluation method that considers multiple factors.

The 7 key nodes identified by the MWCDS constitute a total of 20 service edges. These were systematically compared with the top-ranked edges derived from edge betweenness centrality and the Bridgeness index. As shown in Table V, among the top 20 key edges obtained by the three methods, the results obtained by the MWCDS and edge betweenness coincide with 12 edges, with a coincidence degree of 60%. The results obtained by the MWCDS and Bridgeness index coincide with 14 edges, with a coincidence degree of 70%. Notably, the top five highest-ranked edges exhibit complete overlap across all three methods, verifying the effectiveness of the MWCDS in identifying key edges.

Based on the four evaluation methods of the MWCDS, degree centrality, betweenness centrality, and closeness centrality, the key nodes identification results are analyzed from the perspective of the impact of node failure on the performance of high-speed rail networks. The node failure methods of random and deliberate attacks were adopted, and the network topology efficiency and network service efficiency were used as the network performance measurement methods. The comparison results are shown in Fig 14.

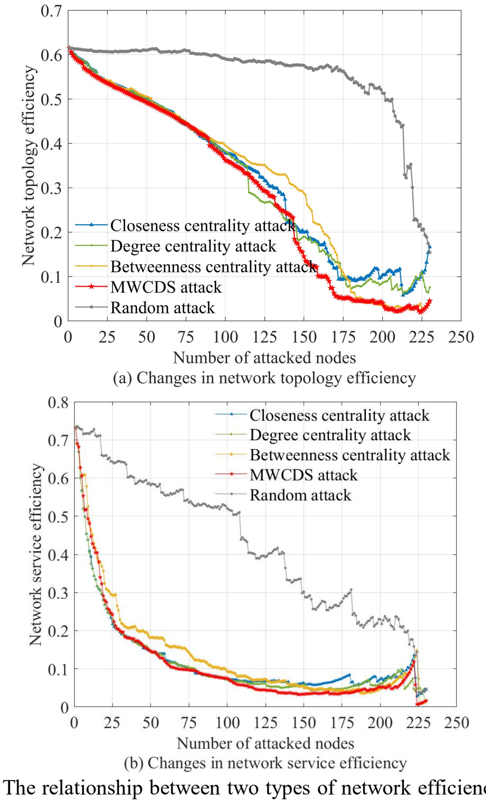


Fig. 14. The relationship between two types of network efficiency and the number of attacked nodes.

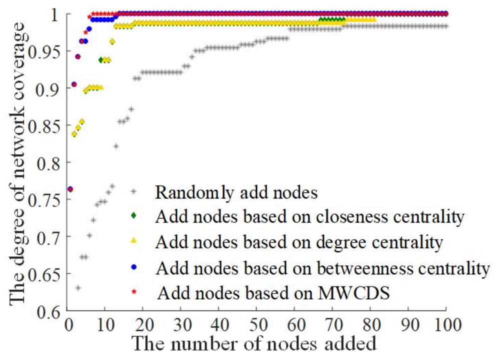


Fig. 15. The relationship between network coverage and the number of added nodes.

Based on four different attack strategies: degree centrality, betweenness centrality, closeness centrality, and the MWCDS, deliberate attacks are carried out on nodes in the network. As shown in Fig 14, the decreasing trend of network efficiency under the four attack strategies remains basically consistent. As the proportion of attacked nodes increases, the network efficiency decreases more significantly when attacked by the MWCDS. It can be seen that the MWCDS not only achieves network controllability but also plays a role in supporting network structure and maintaining normal network operation. In addition, due to the consideration of transportation service quality between nodes in network service efficiency, whether it is random or deliberate attacks, network service efficiency is more sensitive to key node failures. In addition, compare the coverage of key nodes identified by different evaluation methods on the high-speed railway network. As shown in Fig 15, 82 nodes with high degree centrality, 14 nodes with high betweenness centrality, and 74 nodes with high closeness centrality are required to achieve complete coverage of the network. Only 7 key nodes identified by the MWCDS are needed to achieve the same coverage effect on the network, achieving the goal of controlling the entire network with minimal resource cost.

V. Conclusion

This paper introduces the MWCDS theory into the identification of key nodes and edges in the HSRN. The identification process comprehensively considers the dominance, connectivity, and transportation service efficiency between nodes, resulting in a backbone network with a wide span, large coverage, and a more reasonable structure. By establishing multilayer mappings between service networks (train flow) and physical infrastructure (rail lines), our approach enables precise identification of operationally vital railway segments. The research results have important reference significance for railway operation and maintenance.

The proposed IGH algorithm combines the immune mechanisms and three greedy heuristic strategies, enabling efficient identification of network backbone structures within polynomial time complexity. In addition, there is a special type of marginal node in China's high-speed rail network ,which only constitutes a complete subgraph with its neighboring nodes. We propose establishing direct inter-community train services between these nodes and external hubs to enhance network accessibility.

MWCDS was compared with three key node identification indicators and two key edge identification algorithms. The results showed that MWCDS can effectively identify key nodes and edges, not only achieving network controllability with the minimum number of nodes, but also playing an important role in supporting the HSRN structure and maintaining normal operation of the network.

This study did not consider the heterogeneity of train flow at different time periods, and establishing a dynamic network to study the high-speed railway networks will be a future research direction.

References

[1] L. C. Freeman, "Centrality in social networks: Conceptual clarification," Social Network, vol. 1, no. 3, pp. 215-239,1979.

[2] F. Xue, C. L. He, Z. S. Sun, and X. Yu, "Key Node Identification Method of Chengdu Metro Network Based on Comprehensive Assessment, " International Conference on Smart Vehicular Technology, Transportation, Communication and Applications. Springer, Cham, pp. 48-58, 2018.

[3] S. S. Wang, Y. X. Du, and Y. Deng, "A new measure of identifying influential nodes: Efficiency centrality," Communications in Nonlinear Science and Numerical Simulation, vol. 47, pp. 151-163, 2017.

[4] B. H. Rafeeq, R. M. Pir, S. Amir, B. H. Rafiq, and B. Mehri, "A method based on k-shell decomposition to identify influential nodes in complex network," The Journal of Supercomputing, vol. 79, no. 14, pp. 15597-15622, 2023.

[5] L. S. Lv, K. Zhang, T. Zhang, D. Bardou, J. H. Zhang, and Y. Cai, "PageRank centrality for temporal networks," Physics Letters A, vol. 383, no.12, pp. 1215-1222, 2019.

[6] Z. Y. Du, J. J. Tang, Y. Qi, Y. W. Wang, C. Y. Han, and Y. F. Yang, "Identifying critical nodes in metro network considering topological potential: A case study in Shenzhen city — China, " Physica A: Statistical Mechanics and its Applications, vol. 539, Article ID 122926, 2020.

[7] Y. Y. Cheng, R. K. W. Lee, E. P. Lim, and F. D. Zhu, "DelayFlow centrality for identifying critical nodes in transportation networks," Proceedings of the 2013 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining, pp. 1462-1463, 2013.

[8] S. Kim, Y. Yoon, "On node criticality of the Northeast Asian air route network," Journal of Air Transport Management, vol. 80, Article ID 101693, 2019.

[9] J. Zhao, Y. T. Song, and Y. Deng, "A Novel Model to Identify the Influential Nodes: Evidence Theory Centrality," IEEE Access, vol. 8, pp. 46773-46780, 2020.

[10] H. H. Yang, S. An, " Critical Nodes Identification in Complex Networks," Symmetry, vol. 12, no.1, pp. 123-123, 2020.

[11] Q. Y. Zhang, B. Shuai, and M. Lu, " A novel method to identify influential nodes in complex networks based on gravity centrality," Information Sciences: An International Journal, vol. 618, pp. 98-117, 2022.

[12] Y. Z. Cui, X. Y. Wang, and J. Q. Li, " Detecting overlapping communities in networks using the maximal sub-graph and the clustering coefficient, " Physica A: Statistical Mechanics and its Applications, vol. 405, pp. 85-91, 2014.

[13] W. Wang, K. Q. Cai, W. B. Du, X. Wu, L(Carol). Tong, X. Zhu, and X. B. Cao, " Analysis of the Chinese railway system as a complex network," Chaos, Solitons & Fractals, vol. 130, Article ID 109408, 2020.

[14] J. H. Zhang, F. N. Hu, S. L. Wang, Y. Dai, and Y. X. Wang, "Structural vulnerability and intervention of high speed railway networks, " Physica A: Statistical Mechanics and its Applications, vol. 462, pp. 743-751, 2016.

[15] Z. L. Xin, and F. Q. Niu, "Structure and robustness of China's railway transport network," Transportation Letters, vol. 15, no.5, pp. 375-385, 2023.

[16] Z. D. He, K. Navneet, W. V. Dam, and P. V. Mieghem, "Robustness assessment of multimodal freight transport networks, " Reliability Engineering System Safety, vol. 207, Article ID 107315, 2021.

[17] T. Li, and L. L. Rong, "A comprehensive method for the robustness assessment of high-speed rail network with operation data: A case in

China," Transportation Research Part A: Policy and Practice, vol. 132, pp. 666-681, 2020.

[18] M. R. Garey, and D. S. Johnson, Computers and intractability: a guide to the theory of NP-completeness[M]. New York: W. H. Freeman, 1979.

[19] J. Wu, and H. L. Li, "On calculating connected dominating set for efficient routing in ad hoc wireless networks," Proceedings of the 3rd international workshop on Discrete algorithms and methods for mobile computing and communications, pp. 7-14, 1999.

[20] X. G. Chen, L. Sun, and D. X. Ma, "Connected domination critical graphs," Applied Mathematics Letters, vol. 17, no.5, pp. 503-507, 2004.

[21] Y. Wang, W. Z. Wang, and X. Y. Li, "Distributed low-cost backbone formation for wireless ad hoc networks," Proceedings of the 6th ACM international symposium on Mobile ad hoc networking and computing, pp. 2-13, 2005.

[22] J. W. Li, X. X. Wen, M. G. Wu, F. Liu, and S. F. Li, "Identification of key nodes and vital edges in aviation network based on minimum connected dominating set," Physica A: Statistical Mechanics and its Applications, vol.541, pp. 123340-123340, 2019.

[23] L. Shan, H. Li, and Z. Z. Zhang, "Domination number and minimum dominating sets in pseudofractal scale-free web and Sierpiński graph," Theoretical Computer Science, vol.677, pp. 12-30, 2017.

[24] E. Y. Yu, D. B. Chen, and J. Y. Zhao. "Identifying critical edges in complex networks," Scientific reports, vol. 8, pp. 14469, 2018.