



# Identification and Quantification of Node Criticality through EWM–TOPSIS: A Study of Hong Kong’s MTR System

Yifan Zhang<sup>1</sup> · S. Thomas Ng<sup>2</sup>

Received: 26 August 2020 / Revised: 26 December 2020 / Accepted: 8 July 2021 / Published online: 3 August 2021  
© The Author(s) 2021

**Abstract** Public transport networks (PTNs) are critical in populated and rapidly densifying cities such as Hong Kong, Beijing, Shanghai, Mumbai, and Tokyo. Public transportation plays an indispensable role in urban resilience with an integrated, complex, and dynamically changeable network structure. Consequently, identifying and quantifying node criticality in complex PTNs is of great practical significance to improve network robustness from damage. Despite the proposition of various node criticality criteria to address this problem, few succeeded in more comprehensive aspects. Therefore, this paper presents an efficient and thorough ranking method, that is, entropy weight method (EWM)–technology for order preference by similarity to an ideal solution (TOPSIS), named EWM–TOPSIS, to evaluate node criticality by taking into account various node features in complex networks. Then we demonstrate it on the Mass Transit Railway (MTR) in Hong Kong by removing and recovering the top  $k$  critical nodes in descending order to compare the effectiveness of degree centrality (DC), betweenness centrality (BC), closeness centrality (CC), and the proposed EWM–TOPSIS method. Four evaluation indicators, that is, the frequency of nodes with the same ranking ( $F$ ), the global network efficiency ( $E$ ), the size of the largest connected component (LCC), and the average path length (APL), are computed to

compare the performance of the four methods and measure network robustness under different designed attack and recovery strategies. The results demonstrate that the EWM–TOPSIS method has more obvious advantages than the others, especially in the early stage.

**Keywords** Public transport network · Network analysis · Node criticality · Entropy weighting method · Network connectivity

## 1 Introduction

Public transport networks (PTNs), including buses, trolleybuses, trams or light rail, rapid transit (i.e., metro, subway, underground), and ferries, are the backbone and central pillars for urbanization. Urbanization, especially in coastal regions, has exacerbated PTNs’ vulnerabilities to disruptions. Various disruptions, ranging from natural hazards such as super typhoons or flooding to man-made events such as accidents, terrorism, and social unrest, will induce cascading failures in PTNs. Cascading failure of highly connected and interrelated components can cause significant damages to the whole network. As a result, identifying vulnerabilities of each component of PTNs has recently generated considerable research interest in the impacts of disruptive events.

Critical nodes can affect the structure and function of complex networks more significantly than the others [1]. Meanwhile, it reflects its control over structural connectivity and contributions to the functional operability of the system. Consequently, identifying the critical nodes is of considerable significance in analyzing the vulnerability and fragility of a tiny fraction of critical nodes against disruptive events [2–5]. Current methods, including degree

✉ S. Thomas Ng  
thomasng@cityu.edu.hk

<sup>1</sup> Department of Civil Engineering, The University of Hong Kong, Hongkong, China

<sup>2</sup> Department of Architecture and Civil Engineering, City University of Hong Kong, 83Tat Chee Avenue, Kowloon Tong, Hong Kong, China

Communicated by Xuesong Zhou.

centrality (DC), betweenness centrality (BC), closeness centrality (CC), eccentricity, page rank, and subgraph centrality, have been widely used to measure node criticality in complex networks. For example, Yin et al. [6] addressed the importance evaluation problem by establishing the betweenness-based indexes of stations, edges, and lines based on passenger flow, shortest paths, and  $k$  shortest paths. By conducting empirical studies on Shanghai's subway system, Zhang et al. [7] depicted that the system is more vulnerable and susceptible to targeted attacks, and the high betweenness node-based removal would result in further damage to the whole network. Psaltoglou and Calle [8] provided a novel methodology for identifying the critical nodes in urban PTNs over time by comparing the degree centrality and betweenness centrality. Kanwar et al. [9] compared the Delhi Metro (DMop) network and its extension (DMext) by performing targeted and random attacks on high-degree, high-betweenness-based node attacks, and edge-based attacks to study the network vulnerability.

Although many efforts have been made on using a single criterion, such as degree centrality (DC), betweenness centrality (BC) [10], and closeness centrality (CC) to evaluate node criticality, these only exhibit the single-sided aspect of node criticality. DC [11] represents the number of neighbor nodes connected to a node, which implies its influence on the network structure. However, as a simple centrality measure, it only calculates the number of nodes connected to a certain node and does not account for how critical those neighbors are, which indicates the computation precision is not high enough. Moreover, the nodes with the same degree play different roles in a complex network [10]. Meanwhile, BC only considers nodes that belong to the shortest path of other node pairs [12]. At the same time, CC would fail when there is a disconnected component in a network. Both centrality measures have high computational complexity.

Considering the various node features, this paper presents a comprehensive method that can effectively identify the critical nodes in PTNs. This method incorporated the entropy weight method (EWM) [13] and the technology for order preference by similarity to an ideal solution (TOPSIS) that are widely used in other research areas. EWM is a commonly used weighting method to measure the structural complexity and characterize the size of information in decision-making [14]. On the other hand, TOPSIS is a multicriteria decision-making approach proposed by Hwang and Yoon in 1981 and further developed by Yoon in 1987 and Hwang et al. in 1993 [15]. It is based on the concept that an ideal alternative should have the shortest geometric distance to a positive ideal solution (PIS)  $A^+$  and the farthest distance to a negative ideal solution (NIS)  $A^-$ .

TOPSIS has been widely used to identify the critical nodes in different research fields such as the coal industry [16], infrastructure resources [17], security risk [18], social network [19], and public transportation systems [20]. It accredits equal weights to each criterion, whereby the different roles various criteria would play during the procedure are ignored [21].

In this paper, EWM is used to calculate the weight of each criterion and reduce the disadvantages of TOPSIS, which adopts equal weights. This method addresses the problem of inaccurate assessment and avoids the single-sided effects of using a single criterion. Additionally, structural data from the Mass Transit Railway (MTR) network in Hong Kong are used to illustrate the effectiveness of the proposed method. Four evaluation indicators, that is, the frequency of nodes with the same ranking ( $F$ ), global network efficiency ( $E$ ), the size of the largest connected component (LCC), and the average path length (APL), are adopted to measure network robustness before and after removing or recovering the top  $k$  ranked nodes by different ranking methods. The results indicate the superiority of the proposed method.

The remainder of this paper is structured as follows. Section 2 describes the main elements of the node properties, existing centrality criteria, and the proposed method. An empirical analysis is conducted on the MTR network in Hong Kong. Section 3 quantifies the effectiveness of the proposed method by evaluating the network robustness after node removal and the network recovery speed based on different designed sequences. Section 4 presents a detailed discussion, conclusion, and further work.

## 2 Materials and Methods

### 2.1 Methods

Considering the problems of using a single criterion to evaluate the node criticality, this paper adopts a weighting method, namely EWM, and a multicriteria decision analysis method, namely TOPSIS, which incorporate different centrality measures for node criticality evaluation.

#### 2.1.1 Centrality Measures

A network can be represented as a graph that contains nodes and edges with attributes [22]. PTNs can be denoted as an undirected network  $G = (V, E, W, A)$ , where  $V = \{v_1, v_2, \dots, v_n\}$  stands for the set of stations, and  $E = \{e_{ij}\}$ , ( $i \neq j = 1, 2, 3, \dots, n$ ) denotes the connection between the stations, respectively [23].  $W$  denotes the weight set, which equals the travel time between the

stations. In practice, from the passenger's point of view, it is better to spend the minimum time to travel in daily life. Hence, in this paper, the shortest path length  $d_{ij}$  between the nodes is defined as the minimum travel time from  $v_i$  to  $v_j$ . It is expressed as:

$$w_{ij} = \begin{bmatrix} 0 & t_{12} & t_{13} & t_{14} & \cdots & t_{1n} \\ t_{21} & 0 & t_{23} & t_{24} & \cdots & t_{2n} \\ t_{31} & t_{32} & 0 & t_{34} & \cdots & t_{3n} \\ t_{41} & t_{42} & t_{43} & 0 & \cdots & t_{4n} \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ t_{n1} & t_{n2} & t_{n3} & t_{n4} & \cdots & 0 \end{bmatrix} \quad (1)$$

$A = \{a_{ij}\}_{m \times n}$  is the adjacency matrix, where  $a_{ij} = 1$  if there is a connection between node  $i$  and node  $j$ , and 0 otherwise. By calculating the sum of  $a_{ij}$ ,  $k_i = \sum_j a_{ij}$  can be expressed as the degree of node  $i$ . Hence the degree centrality can be denoted as

$$DC(i) = \frac{k_i}{n-1} \quad (2)$$

Generally speaking, a node's influence is significantly associated with its capacity and its surrounding neighbors [24]. Therefore, degree centrality (DC) is used to evaluate the ability of a node to transmit information to others. The greater the value, the more critical the node is.

Betweenness centrality (BC) is a measurement of node criticality proposed over the years to evaluate the node or edge criticality by measuring the structural centrality [25]. It reflects the ability of a node to control the whole network flow, such as passenger flow passing through the shortest path in PTNs. For undirected network, it can be expressed as follows:

$$BC(i) = \frac{2B_i}{(n-1)(n-2)} \quad (3)$$

$B_i$  denotes as the fraction of the shortest paths between nodes.

In a connected PTN, closeness centrality (CC) of a node is calculated as the reciprocal of the sum of the length or links of the shortest paths between two nodes [26]. It is regarded as a measure of how long it will take passengers to travel from a given node to other reachable nodes in the network [27]. It can be represented as in the following equation:

$$CC(i) = \frac{n-1}{\sum_{j=1}^n d_{ij}} \quad (4)$$

where  $d_{ij}$  is the shortest paths between nodes  $i$  and  $j$ . It can be regarded as a measurement of the average travel time from  $v_i$  to  $v_j$ . The greater the value, the higher the efficiency of the node.

Although these criteria have been used to evaluate node criticality from different perspectives, they only reflect a single-side characteristic of the nodes. Hence, there should

be a unified criterion to evaluate the node criticality from diverse perspectives. To illustrate this problem, this paper proposes a novel method using the centrality criterion illustrated above, i.e., TOPSIS and EWM, to assess the node criticality in undirected networks such as PTNs.

## 2.1.2 Proposed Method

### Step 1: Initialize the original matrix

If there is a set of nodes in a graph-based network, whose nodes are denoted as  $V = \{v_1, v_2, \dots, v_n\}$  and another three centrality measures  $j = (1, 2, 3)$ , which represent DC, BC, and CC respectively, then the  $v_i(c_j)$  ( $i = 1, 2, \dots, n; j = 1, 2, 3$ ) represents the value of  $j$ th centrality measures for the  $i$ th node. The original matrix would be decided as

$$D = \begin{bmatrix} v_{11} & v_{12} & v_{13} \\ v_{21} & v_{22} & v_{23} \\ \vdots & \vdots & \vdots \\ v_{n1} & v_{n2} & v_{n3} \end{bmatrix} \quad (5)$$

To eliminate dimensional differences among the centrality measures and standardize the original matrix, the centrality measures can be standardized into benefit criteria denoting that the higher the measure, the more important the node is. It is expressed as

$$r_{ij} = \frac{v_{ij} - v_{ij \min}}{v_{ij \max} - v_{ij \min}} \quad (6)$$

where  $v_{ij \min} = \min\{v_{ij}, 1 \leq i \leq n\}$ ,  $v_{ij \max} = \max\{v_{ij}, 1 \leq i \leq n\}$ . Therefore, the standardized matrix can be denoted as

$$R = \begin{bmatrix} r_{11} & r_{12} & r_{13} \\ r_{21} & r_{22} & r_{23} \\ \vdots & \vdots & \vdots \\ r_{n1} & r_{n2} & r_{n3} \end{bmatrix} \quad (7)$$

### Step 2: Calculate the objective weight of each criterion

According to Shannon's entropy calculation procedure, there is a calculation of the ratio of each criterion and the sum of all criteria, where  $p_{ij} = r_{ij} / \sum_{k=1}^n r_{kj}$ .

EWM is used to calculate each weight of each criterion, which decides the weight according to the different roles it plays. Therefore, the entropy is defined as

$$E_j = -K \sum_{i=1}^n p_{ij} \ln p_{ij}, \quad j = 1, 2, \dots, n \quad (p_{ij} \neq 0) E_j = 0, \\ j = 1, 2, \dots, n \quad (p_{ij} = 0) \quad (8)$$

where  $K = 1 / \ln n$ .

When  $p_{ij} = 0$ ,  $E_j = 0$ . Then the weighting coefficient of  $j$ th criterion will be calculated through

$$\omega_j = \frac{1 - E_j}{\sum_{k=1}^3 (1 - E_k)} \quad (9)$$

*Step 3: Normalize the standardized matrix*

Multiplying the columns of the standardized matrix by the corresponding weights yields the weighted matrix, which can be expressed as

$$Y = [y] = (\omega_j r_{ij}) = \begin{bmatrix} \omega_1 r_{11} & \omega_2 r_{12} & \omega_3 r_{13} \\ \omega_1 r_{21} & \omega_2 r_{22} & \omega_3 r_{23} \\ \vdots & \vdots & \vdots \\ \omega_1 r_{n1} & \omega_2 r_{n2} & \omega_3 r_{n3} \end{bmatrix} \quad (10)$$

*Step 4: Calculate the distance to the ideal solution*

Based on each standardized criterion, the positive ideal solution (PIS)  $Y^+$  and the negative ideal solution (NIS)  $Y^-$  can be denoted as

$$\begin{aligned} Y^+ &= \left\{ \max_{i \in \{1, 2, \dots, n\}} (y_{i1}, y_{i2}, y_{i3}) \right\} = \{y_1^+, y_2^+, y_3^+\} \\ Y^- &= \left\{ \min_{i \in \{1, 2, \dots, n\}} (y_{i1}, y_{i2}, y_{i3}) \right\} = \{y_1^-, y_2^-, y_3^-\} \end{aligned} \quad (11)$$

Therefore, the distance between each standardized criterion and PIS, and NIS, can be calculated through the following equations:

$$\begin{aligned} S_i^+ &= \left[ \sum_{j=1}^3 (y_{ij} - y_j^+)^2 \right]^{1/2} \\ S_i^- &= \left[ \sum_{j=1}^3 (y_{ij} - y_j^-)^2 \right]^{1/2} \end{aligned} \quad (12)$$

*Step 5: Calculate the relative closeness degree*

The closer to the PIS and the farther away from the NIS, the more importance the node is. Accordingly, the relative closeness degree to the ideal solution is the measurement of a node's importance and criticality, which can be calculated as follows:

$$C_i = \frac{S_i^-}{S_i^- + S_i^+}, (i = 1, 2, \dots, n) \quad (13)$$

*Step 6: Ranking node importance*

The vector of node importance can be denoted as

$$Q_e = [C_1, C_2, \dots, C_n] \quad (14)$$

Then  $Q_e$  is ranked in descending order based on each centrality measure of each node; it can be acquired through

$$\begin{aligned} Q'_e &= \text{sort}[Q_e] = \text{sort}([C_1, C_2, \dots, C_i, \dots, C_n]) \\ &= [Q'_{e1}, \dots, Q'_{ej}, \dots, Q'_{en}] \end{aligned} \quad (15)$$

On the basis of the theoretical analysis, the algorithm to rank the node importance in a graph-based network is shown in Algorithm 1. All the analyses are executed through MATLAB.

---

#### Algorithm 1. Rank the most important nodes

---

**Input:** Decision matrix  $D$  with  $|V|=n$

**Output:** Most important nodes with the result of ranking  $Q'_e = \text{sort}[Q_e] = \text{sort}([C_1, C_2, \dots, C_i, \dots, C_n]) = [Q'_{e1}, \dots, Q'_{ej}, \dots, Q'_{en}]$

---

(1) Constitute the weighted decision matrix

(i) For each  $v_i$ , calculate standardized decision matrix  $R$  through Eq. (5).

(ii) The weights of each centrality measure are determined through Eq. (8) on the basis of EWM.

(iii) Based on each weight, the weighted decision matrix  $Y$  is generated.

(2) Calculate the ideal solution (PIS)  $Y^+$  and NIS  $Y^-$  and the distance ( $S_i^+$  and  $S_i^-$ ) to  $Y^+$  and  $Y^-$  through Eq. (10) and (11), respectively

(3) Set and rank the node importance in descending order through Eq. (14), and obtain the results

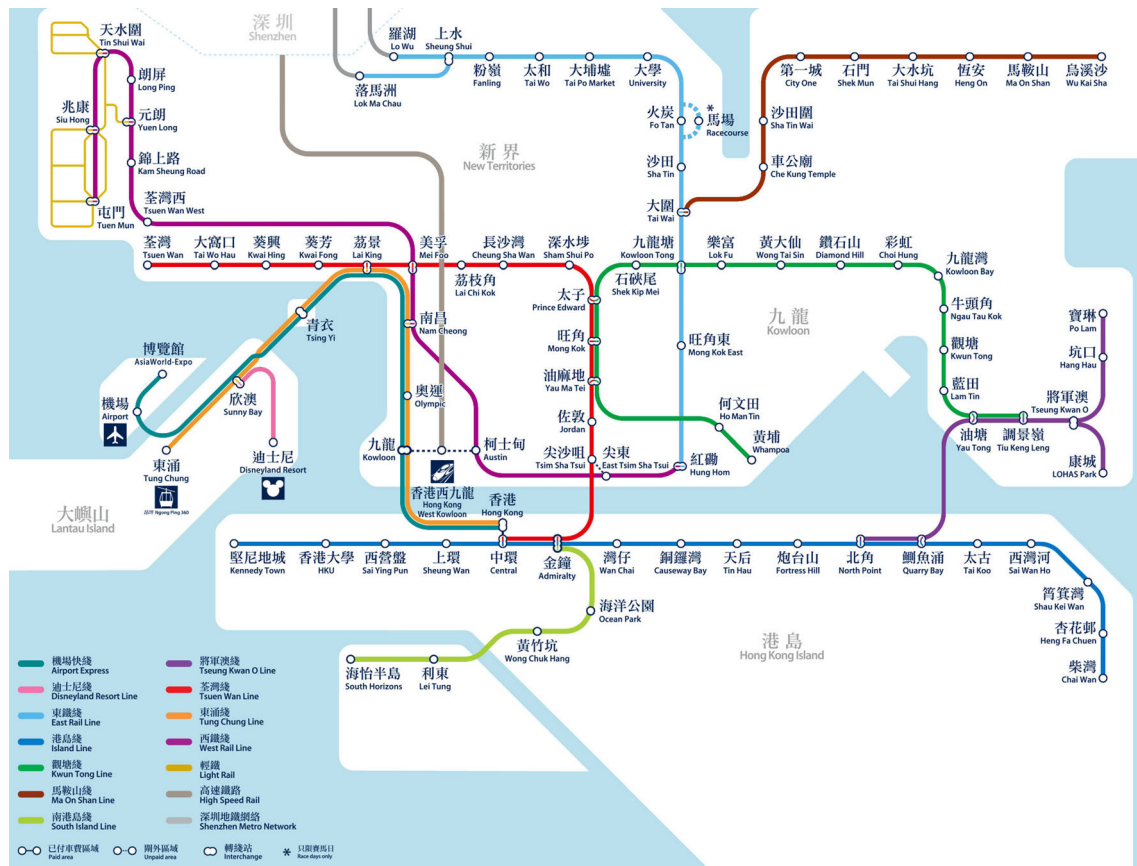
---

## 2.2 Materials

To validate the feasibility and effectiveness of the proposed method, a case study is demonstrated on the MTR network in Hong Kong. MTR, as a rapid transit system in Hong Kong, consists of 11 lines serving the urbanized areas of Hong Kong Island, Kowloon, and the New Territories (Fig. 1): Airport Express; Tseung Kwan O Line; Disneyland Resort Line; Tsuen Wan Line; East Rail Line; Tung Chung Line; Island Line; West Rail Line; Kwun Tong Line; Ma On Shan Line; South Island Line. Each station is regarded as a node, and each link between stations is regarded as an edge in a graph-based network. The dataset is collected from MTR's official website.

This paper computed the basic topological features of the Hong Kong MTR network, which contains 95 nodes and 102 edges. The number of nodes, number of edges, maximum degree ( $D_{\max}$ ), minimum degree ( $D_{\min}$ ), average degree ( $D_{\text{avg}}$ ), clustering coefficient ( $C$ ), and average betweenness (AB) are presented in Table 1. The node betweenness is defined as the number of the shortest paths passing through nodes within the whole network. AB is an attribute to evaluate the average frequency of shortest paths passing a node, which indicates the average transport capacity of nodes.

The degree distribution of the Hong Kong MTR network is shown in Fig. 2. It depicts that when the degree is equal to 2, the probability of degree distribution of the Hong Kong MTR network will exceed 0.6. When the degree is equal to 4, the probability of degree distribution is less than 0.1. The maximum degree is 4, while the minimum degree is 1. Particularly, underground pedestrian passages serve as



**Fig. 1** Hong Kong Mass Transit Railway (MTR) network

**Table 1.** Basic topological statistical parameters of Hong Kong MTR network

Parameter	Hong Kong MTR network
Node number	95
Edge number	102
$D_{\max}$	4
$D_{\min}$	1
$D_{\text{avg}}$	2.15
AB	408.11

the connection between Hong Kong Station and Central Station and East Tsim Sha Tsui Station and Tsim Sha Tsui Station. Hence, in this paper, we assume the connection between the aforementioned stations to be a weighted edge.

### 3 Quantify the Effectiveness of the Proposed Method

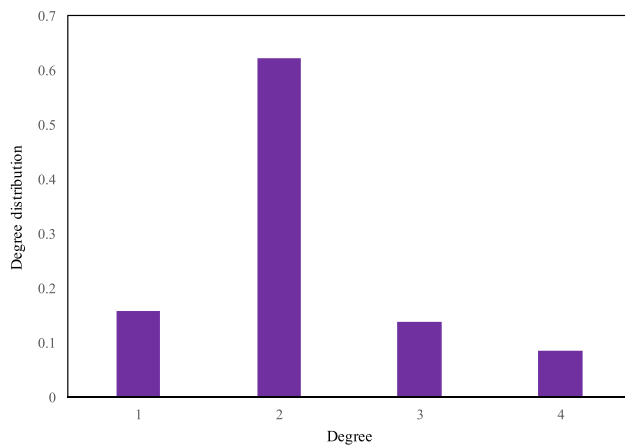
Demonstrated on the real and synthetic PTNs, we examine the effectiveness of the proposed method in the case of random, targeted attacks and recovery speed. For targeted attacks, we remove the nodes in descending order ranked

by degree centrality (DC), betweenness centrality (BC), closeness centrality (CC), and the proposed EWM–TOPSIS method, respectively, while network recovery refers to the recovery speed based on the above-mentioned methods after random attacks in the removal of top 30 nodes. The random attack is used as a benchmark. Then, four evaluation indicators, that is, the frequency of nodes with the same ranking ( $F$ ) initially, the global network efficiency ( $E$ ), the size of the largest connected component (LCC), and the average path length (APL), are calculated to measure the performance of the proposed method.

#### 3.1 Evaluation Indicators

**Global efficiency ( $E$ )**—Global efficiency, developed by Latora and Marchiori [28], is used to eliminate the disadvantage of the size of giant component theory and can be calculated through Eq. (16). It measures how efficient a network performs before and after disruptions, which is expressed as follows:





**Fig. 2** Degree distribution of MTR network in Hong Kong

$$E = \frac{1}{n(n-1)} \sum_{i,j \in G(i \neq j)} \frac{1}{d_{ij}} \quad (16)$$

where  $d_{ij}$  is the shortest path length between node  $i$  and  $j$  in  $G$  and  $n$  is the total number of nodes. The distance  $d_{ij}$  should be infinite if nodes are disconnected. The higher the value of  $E$ , the higher the efficiency of the networks. The ratio of network efficiency before and after node removal  $\mu$  is adopted to evaluate the effect of removing nodes on network connectivity, which is expressed as

$$\mu(k) = \frac{E_k}{E_0} \quad (17)$$

where  $E_0$  denotes the network efficiency before removing the top  $k$  critical nodes, whereas  $E_k$  denotes the efficiency after removing the top  $k$  nodes at  $k$ th time step. The faster the decrease in the value of  $\mu$ , the more critical the removed nodes and the more efficient the methods.

The size of the largest connected component (LCC)—In complex graph theory, a connected component is a

subgraph in which edges connect any two nodes. The largest connected component is the maximal connected subgraph. When a fraction of the nodes is removed, the network will be split into several connected components. The size of the largest connected component (LCC) is used as a robustness indicator to evaluate network connectivity and compare the performance of the proposed method with other methods. The ratio of the size of the largest connected components before and after node removal  $S(k)$  is expressed as

$$S(k) = \frac{LCC_k}{LCC_0} \quad (18)$$

where  $LCC_k$  denotes the final size of LCC after removing the top  $k$  nodes, and  $LCC_0$  denotes the initial size of LCC. The smaller the  $S$ , the better the ranking method is.

Average path length (AOL)—The average path length is a structural robustness indicator to measure how interconnected a network is. It is calculated by determining the average value of the length of the shortest path between all pairs of nodes in the weighted network. Smaller APL reflects strong connectivity between the nodes. When a fraction of nodes in the network fails, the value of APL will change. It is calculated as follows:

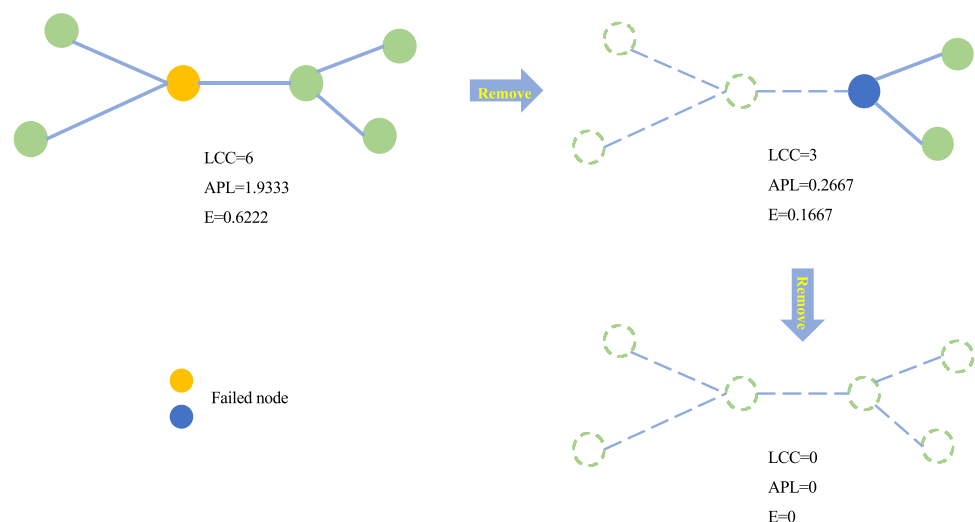
$$l = \frac{1}{n \cdot (n-1)} \cdot \sum_{i,j=1(i \neq j)}^n d_{ij} \quad (19)$$

The ratio of the average path length before and after removing the top  $k$  critical nodes  $L(k)$  is denoted as

$$L(k) = \frac{l_k}{l_0} \quad (20)$$

where  $l_k$  is the value of APL after removing the top  $k$  critical nodes, and  $l_0$  is the original value of APL.

**Fig. 3** The removal process in the representative network with six nodes



### 3.2 Attack Strategies

Different attack strategies are introduced in this section. In this paper, we continuously attack the network by removing the nodes according to the attack protocols designed as follows:

- (1) The highest degree-based attacks
- (2) The highest betweenness-based attacks
- (3) The highest closeness-based attacks
- (4) The highest EWM–TOPSIS-based attacks
- (5) Random attacks serve as a benchmark, which refers to removing the nodes randomly, while 20 independent runs are conducted and the ensemble mean values for the LCC, APL, and  $E$  are plotted
- (6) Recovery speed after random attacks in the early stage.

This paper adopts a dynamic attack manner because it performs better than the static attack for the same attack sequence [29]. After an attack, the topological features are recalculated, ensuring that the most critical node in the current network is removed in each round of attack. To increase statistical significance, nodes with the same ranking will fail randomly, and 20 independent runs are considered. Figure 3 shows the removal process in an undirected network with six nodes and the change of network connectivity under attacks.

Nodes in yellow and blue are selected for targeted removal. When a certain node is removed, it will lead to the failure of the edges that connect it with other neighboring nodes. Furthermore, the neighbored nodes with degree 1 will get dysfunctional. The dashed nodes and edges indicate that they have been detached from the network and, hence, incapacitated, resulting in the deduction of LCC, APL, and  $E$ .

### 3.3 Results

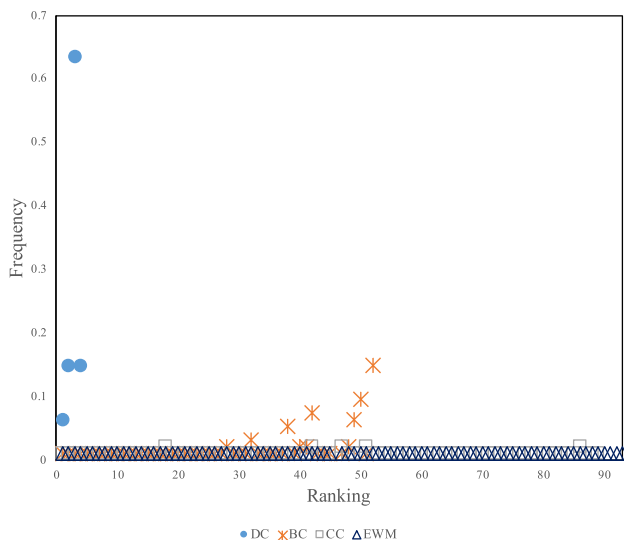
#### 3.3.1 Frequency of Nodes with the Same Ranking ( $F$ )

Table 2 presents the results for comparison, which indicates the 20 top-ranked nodes from original state on the basis of DC, BC, CC, and EWM–TOPSIS by descending order.

A higher frequency of the same ranking will lead to difficulty in ranking node criticality effectively. According to the overall ranking results, the frequency of nodes with the same ranking sorted by the DC method is the highest, reaching 62.11%. Nodes ranked by DC are divided into four categories according to the degree's value; for example, the top eight nodes share the same ranking with four adjacent nodes. This is due to it only considering the number of adjacent nodes from a local perspective and exhibiting less structural information from a global perspective. In contrast, the BC and CC methods have much a

**Table 2** Top 20 nodes ranked by DC, BC, CC, and EWM–TOPSIS

Rank	DC	BC	CC	EWM–TOPSIS
1	Diamond Hill	Tai Wai	East Tsim Sha Tsui	Kowloon Tong
2	Kowloon Tong	Kowloon Tong	Tsim Sha Tsui	Tai Wai
3	Tai Wai	Admiralty	Prince Edward	Admiralty
4	Lai King	Tsim Sha Tsui	Kowloon Tong	Mei Foo
5	Mei Foo	Mei Foo	Shek Kip Mei	Tsim Sha Tsui
6	Tsing Yi	Prince Edward	Hung Hom	Prince Edward
7	Nam Cheong	East Tsim Sha Tsui	Jordan	East Tsim Sha Tsui
8	Admiralty	Shek Kip Mei	Mong Kok	Diamond Hill
9	Yau Tong	Quarry Bay	Admiralty	Quarry Bay
10	Prince Edward	Wan Chai	Yau Ma Tei	Lai King
11	Yau Ma Tei	Sham Shui Po	Sham Shui Po	Nam Cheong
12	Tsim Sha Tsui	Causeway Bay	Austin	Yau Tong
13	East Tsim Sha Tsui	Yau Tong	Lok Fu	Sha Tin
14	Central	Cheung Sha Wan	Mong Kok East	Shek Kip Mei
15	Quarry Bay	Sha Tin	Tai Wai	Central
16	Tseung Kwan O	Diamond Hill	Mei Foo	Tsing Yi
17	Kowloon	Lai Chi Kok	Cheung Sha Wan	Wan Chai
18	Sunny Bay	Tin Hau	Central	Sham Shui Po
19	Sha Tin	Fortress Hill	Wan Chai	Causeway Bay
20	University	North Point	Lai Chi Kok	Cheung Sha Wan



**Fig. 4** Frequency of nodes with the same ranking using the four methods

lower frequency at 16.84% and 2.11%, respectively. However, the frequency of nodes with the same ranking by the EWM–TOPSIS method has reached zero. Figure 4 shows the comparison of the frequency of nodes with the same ranking by applying the four methods. It is concluded that the EWM–TOPSIS method is much more accurate and efficient than the other centrality measures with the lowest  $F$ .

Although the EWM–TOPSIS method is calculated based on DC, BC, and CC, it shows a better performance of ranking critical nodes and of higher resolution than DC, BC, and CC, themselves. The ranking results will be used as a baseline to quantify the network robustness under target attacks. Kowloon Tong, Tai Wai, Admiralty, Mei Foo, and Tsim Sha Tsui, as the five top-ranked nodes ranked by EWM–TOPSIS, highlight their criticality in terms of network connectivity and transport capacity. Further discussion will be portrayed in the next section.

### 3.3.2 Connectivity Analysis

In this section, we examine the effectiveness of the proposed method in terms of network robustness and recovery through connectivity measures, including  $\mu(k)$ ,  $S(k)$ , and  $L(k)$ . The robustness of a complex network refers to its capability to maintain network connectivity given the failure of a fraction of nodes or edges. Firstly, nodes are removed from the network in descending order according to designed attack protocols 1–5. Then, the value of  $\mu(k)$  is recalculated through formula (17) separately in each round of attack. The lower the value of  $\mu(k)$ , the greater the decrease in network efficiency, indicating a higher-ranking accuracy. A fraction of critical nodes is removed according

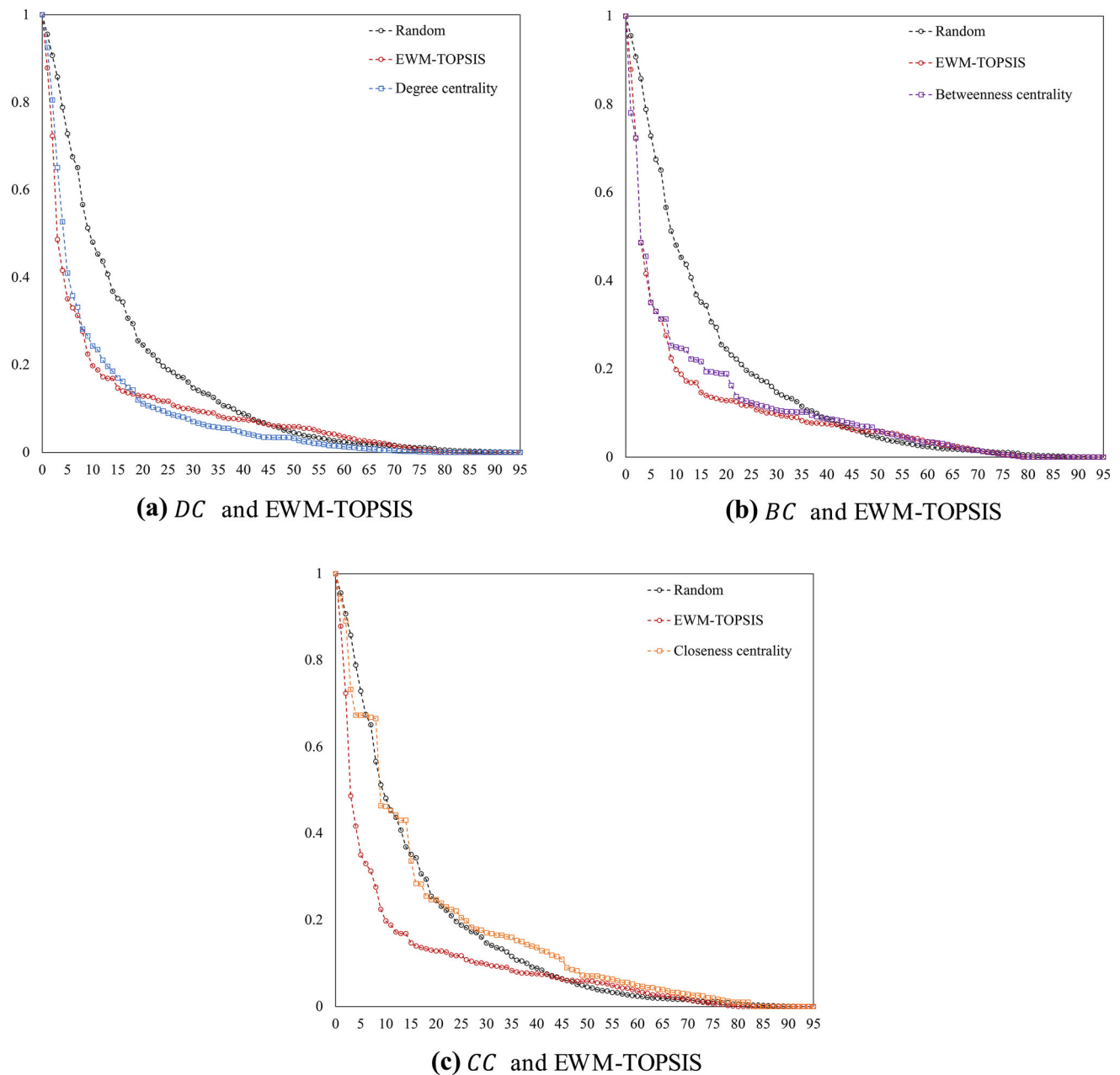
to the ranking results in a dynamic manner. The process is conducted continuously until the value of  $\mu(k)$  reaches the threshold limit value (TLV). The resulting  $\mu(k)$  values are depicted in a line graph, as shown in Fig. 6 to provide a more in-depth insight into the quantitative change in network efficiency.

As can be seen from Fig. 5, the decrease of  $\mu(k)$  in the case of EWM–TOPSIS is more significant than the others, which has shown its higher accuracy and efficiency, especially in the initial stage. The EWM–TOPSIS method corresponds to the most significant decline in global network efficiency. It has declined from 0.5999 to 0.5269 when “Kowloon Tong” is removed, which shows a significant (12.17%) fall in  $\mu(k)$ . Simultaneously, there is a sudden decrease from 0.5999 to 0.2918 (i.e., 51.36%) when “Kowloon Tong,” “Tai Wai,” and “Admiralty” are removed, while the corresponding numbers of DC, BC, and CC are 34.92%, 51.36%, and 26.82%. BC shows the same performance level as EWM–TOPSIS in the early stage. However, EWM–TOPSIS is superior to BC attack protocols during the latter stages. When 69 nodes ranked by EWM–TOPSIS are removed, the value of  $\mu(k)$  has declined to less than 0.01, which implies the most significant connectivity loss and network paralysis. Meanwhile, in the case of node criticality ranked by DC, BC, and CC, the change of network efficiency follows a similar downward trend.

Figure 6 displays the decrease of the size of the largest connected component ( $S(k)$ ) subject with four attack protocols. As can be seen from Fig. 7b, it drastically decreases to less than half of its original size when only removing the top three nodes under EWM–TOPSIS and BC attack protocols, whereas it requires the removal of five and nine nodes to achieve the same performance level in regard to nodes ranked by DC and CC, respectively. It can be inferred that CC corresponds to the least accuracy and effectiveness when quantifying node criticality. Since the EWM–TOPSIS attack strategy can increase the decline speed of  $S$  fastest when removing a small fraction of nodes and ensure continuously high strike effectiveness when removing a large fraction of nodes, the proposed EWM–TOPSIS method performs the best to quantify node criticality further.

Furthermore, we explore the decline speed of average path length (APL) subjected to different attack strategies. As shown in Fig. 7, the variation of APL shows an “upward-downward” trend, which refers to the fact that the removal of the most critical nodes can result in the failure of edges that are critical to network connectivity. The variation exhibits a descending trend when top  $k$  nodes are removed at  $k$ th time step. The smaller the ratio of the average path length before and after removing the top  $k$  critical nodes ( $L$ ), the more accurate the ranking method is.





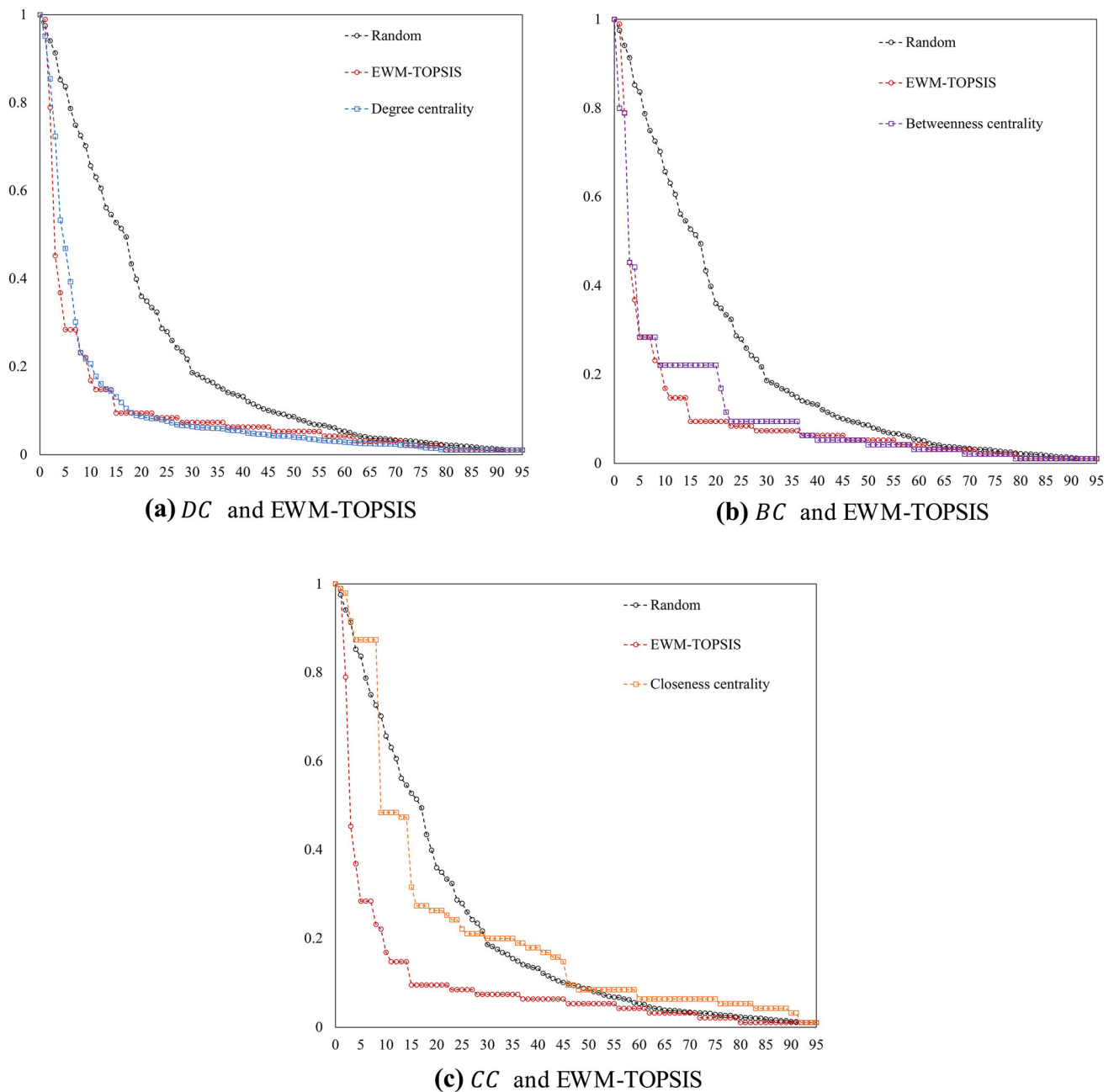
**Fig. 5** The decline rate of global network efficiency ( $\mu$ ) with different attack protocols

It can be seen from the figure that the EWM-TOPSIS method proposed in this paper has the most significant strike effect on network fragmentation, and hence the effectiveness of the proposed method is further quantified by different topological features.

### 3.3.3 Recovery Speed

To further examine the effectiveness of the proposed method, recovery sequences are generated by different methods after 30% random attacks. Additionally, 20

random recovery sequences are generated to be regarded as a baseline for comparison. In this case, the recovery speed is measured by computing the area between the Y-axis and the recovery curve, which is defined as the impact area (IA) [24]. A larger IA indicates a less efficient recovery strategy. Observing the results, as shown in Fig. 8, we can find that the recovery sequence generated by EWM-TOPSIS has an approximately 10% smaller IA than the corresponding value based on the DC-based sequence, which shows that the proposed EWM-TOPSIS can make the



**Fig. 6** The decline rate of the size of the largest connected component ( $S$ ) with different attack protocols

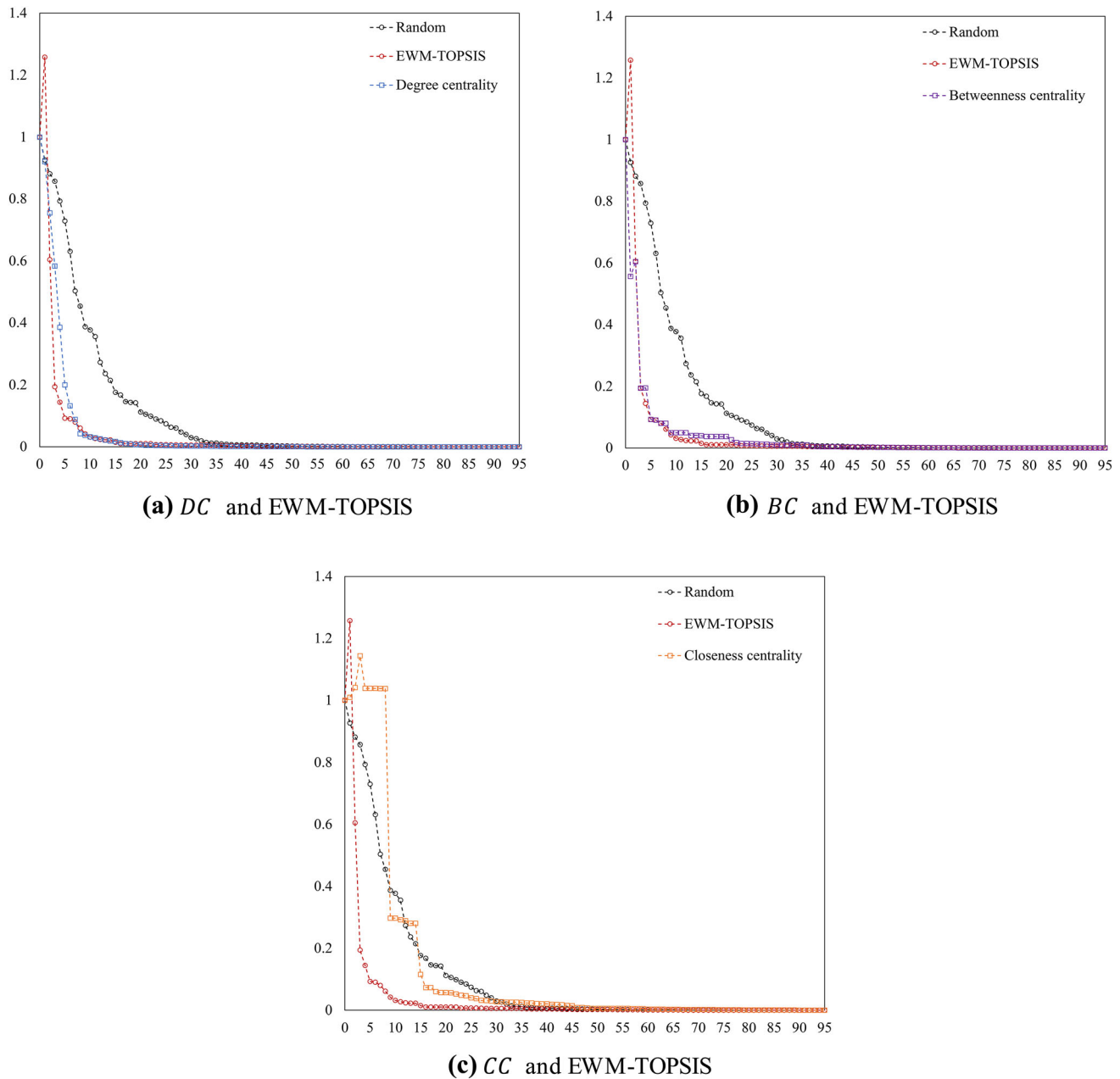
network recover from random attacks more efficiently and indicates its superiority for ranking node criticality.

In real-world transportation systems, node criticality is complicatedly correlated with different centrality measures. However, a purely single centrality-based measure is incapable of satisfying the demand for the fastest restoration. In summary, the results indicate that the proposed EWM-TOPSIS emerges as a better method for examining network robustness and, hence, evaluating node criticality than degree centrality (DC), betweenness centrality (BC), and closeness centrality (CC) in a dynamic attack manner,

which has more evident advantages in the initial rounds of attack, indicating its highest application value. Comprehensively, BC performs the second-best in evaluating node criticality, whereas CC is the worst.

#### 4 Discussion and Conclusions

Prior work has utilized a single criterion, for example, BC, CC, DC, page rank, etc., to evaluate node criticality in complex networks. However, these studies only reflect the



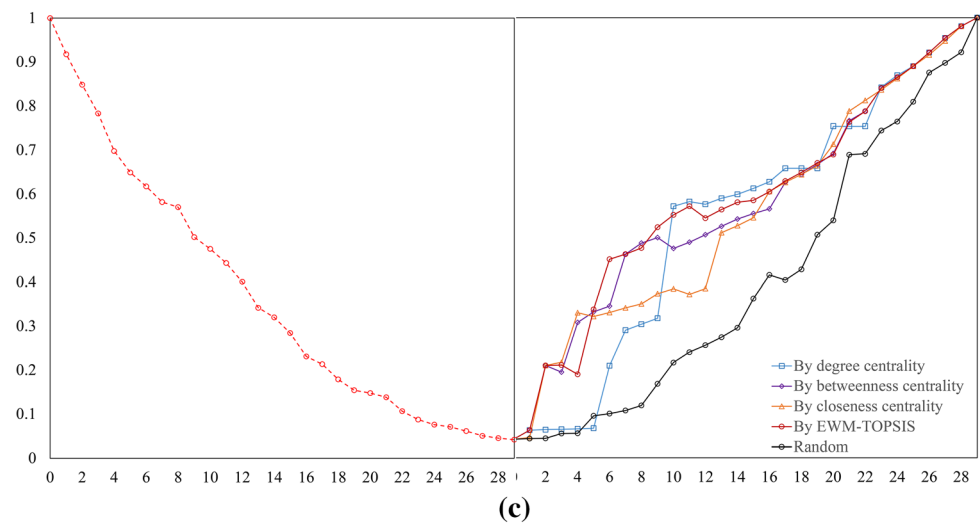
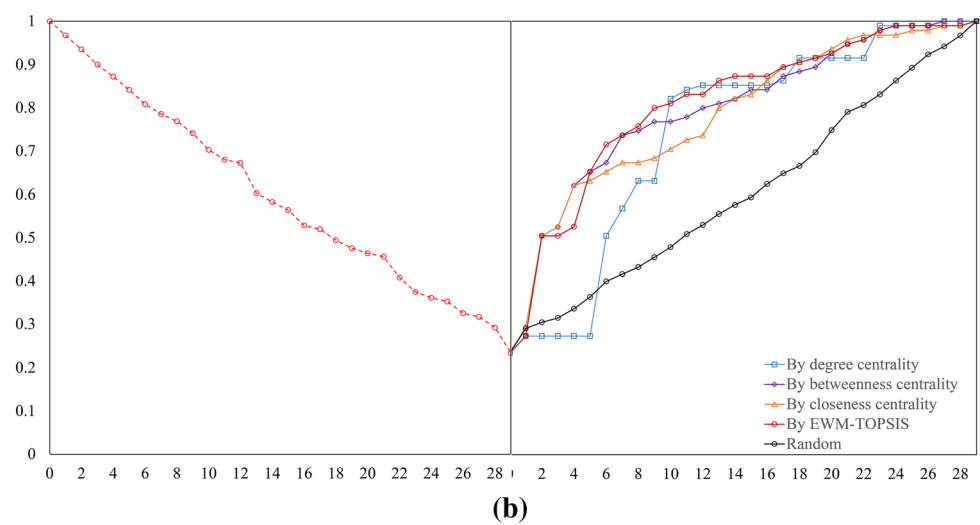
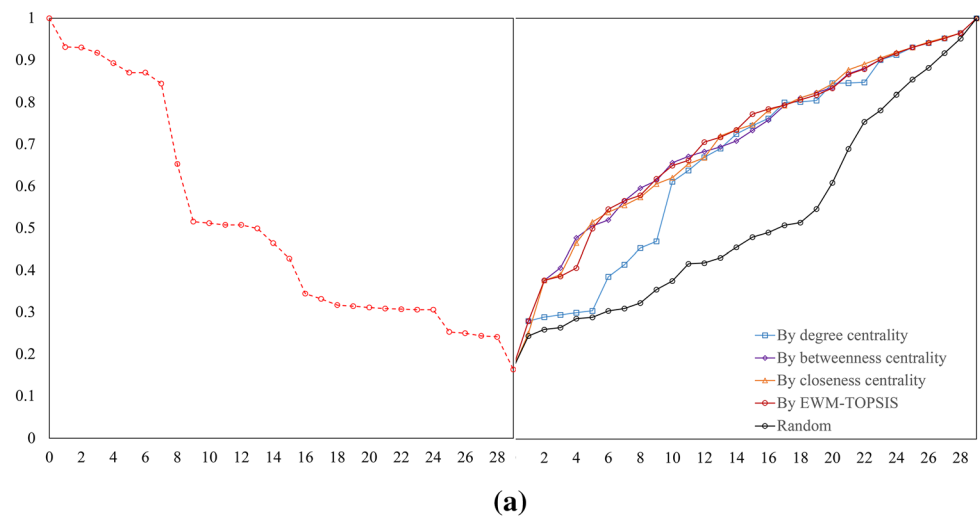
**Fig. 7** The decline rate of the size of the largest connected component ( $L$ ) with different attack protocols

single-sided characteristic of node criticality. To eliminate this deficiency, we propose a node criticality identification method based on EWM and TOPSIS. The network robustness and recovery simulation experiments on real PTNs show that the proposed method synthesizes the advantages of three single criteria (DC, BC, and CC), and gains more accurate and effective assessment results.

An empirical analysis is conducted on MTR in Hong Kong to warrant the effectiveness of the proposed method. Four evaluation indicators have been used to quantify the effectiveness of different ranking methods, including the frequency of nodes with the same ranking ( $F$ ), the global

network efficiency ( $E$ ), the size of the largest connected component (LCC), and the average path length (APL). The network connectivity decreases sharply after removing the top three critical nodes in descending order ranked by DC, BC, CC, and EWM-TOPSIS. Since the EWM-TOPSIS method corresponds to the most significant decline in the value of  $E$ , LCC, and APL after removing the top  $k$  critical nodes in a dynamic attack manner, it shows superiority over the other three centrality measures in the context of accurately evaluating node criticality. Its superiority over the other three measures is more evident when a smaller fraction of nodes is removed. The network gets close to

**Fig. 8** The change of ( $\mu$ ), ( $S$ ), and ( $L$ ) subject with different recovery strategies after random attacks



dysfunctionality when only less than 10% of critical nodes are removed in the case of all ranking methods. To further certify the application value of the proposed method, we compare the performance of multiple station recovery sequences generated from different methods after randomly removing 30% nodes. In this case, EWM–TOPSIS tends to generate the most efficient recovery sequences and bring the network back to full functionality most effectively.

This paper has proposed a comprehensive method of evaluating node criticality. It takes a multicriteria approach and recalculates the weights of each centrality criterion to overcome the deficiencies of using a single criterion, which can provide support for the planning of highly efficient transport systems. However, some limitations are worth noting. The proposed method in this paper is for the undirected networks. In real PTNs, many of them are directed networks where the connections between nodes are not bidirectional.

Additionally, although the proposed method has superior performance in evaluating node criticality and providing necessary analysis results for assessing network topology, further optimization still exists, including considering the impact of dynamic features (i.e., passenger flow). Future work will, therefore, consider the effect of passenger flow through different stations over time. Furthermore, more tests on real networks will be carried out for generalization into other fields.

**Acknowledgements** The authors would like to thank the Research Grants Council of the HKSAR Government for financially supporting this research through the General Research Fund (Grant Nos.: 17248216 and 17204017).

**Author contributions** Y.Z.—data collection, data analysis, paper writing; T.N.—results interpretation, paper editing, funding.

**Funding** Funding was provided by Research Grants Council of the HKSAR Government through the General Research Fund (Grant Nos. 17248216 and 17204017).

**Data availability** The raw data about the network structure and operation schedule of MTR in Hong Kong are extracted from [http://www.mtr.com.hk/en/customer/services/system\\_map.html](http://www.mtr.com.hk/en/customer/services/system_map.html).

**Code availability** The original code is written by the authors and therefore cannot be shared. The authors would be willing to provide the code to interested parties upon request.

**Declarations**

**Conflict of interest** The authors declare that they have no conflicts of interest.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as

long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

## References

1. Ren X, Lu L (2014) Review of ranking nodes in complex networks. *Chin Sci Bull* 59(13):1175–1197
2. Sienkiewicz J, Hołyst JA (2005) Statistical analysis of 22 public transport networks in Poland. *Phys Rev E* 72(4):046127
3. von Ferber C, Holovatch T, Holovatch Y, Palchykov V (2007) Network harness: metropolis public transport. *Phys A* 380:585–591
4. Louf R, Roth C, Barthélemy M (2014) Scaling in transportation networks. *PLoS ONE* 9(7):102007
5. de Regt R, von Ferber C, Holovatch Y, Lebovka M (2019) Public transportation in Great Britain viewed as a complex network. *Transp A Transp Sci* 15(2):722–748
6. Yin H, Han B, Li D, Wang Y (2016) Evaluating disruption in rail transit network: a case study of Beijing subway. *Procedia Eng* 137:49–58. <https://doi.org/10.1016/j.proeng.2016.01.233>
7. Zhang J, Wang S, Wang X (2018) Comparison analysis on vulnerability of metro networks based on complex network. *Phys A* 496:72–78
8. Psaltoglou A, Calle E (2018) Enhanced connectivity index—a new measure for identifying critical points in urban public transportation networks. *Int J Crit Infrastruct Prot* 21:22–32
9. Kanwar K, Kumar H, Kaushal S (2019) Complex network based comparative analysis of Delhi metro network and its extension. *Physica A Stat Mech Appl* 526:120991
10. Freeman LC (1977) A set of measures of centrality based on betweenness. *Sociometry* 40(1):35–41
11. Albert R, Jeong H, Barabási AL (1999) Internet: diameter of the world-wide web. *Nature* 401(6749):130
12. Testa AC, Furtado MN, Alipour A (2015) Resilience of coastal transportation networks faced with extreme climatic events. *Transp Res Rec* 2532(1):29–36
13. Ai X (2017) Node importance ranking of complex networks with entropy variation. *Entropy* 19(7):303
14. Buldyrev SV, Parshani R, Paul G, Stanley HE, Havlin S (2010) Catastrophic cascade of failures in interdependent networks. *Nature* 472:1025–1028
15. Debnath A, Roy J (2018) Integrated fuzzy AHP–TOPSIS model for optimization of national defense management based on inclusive growth drivers using SWOT analysis. In: *Handbook of research on military expenditure on economic and political resources*. IGI Global, pp 81–105
16. Li X, Wang K, Liu L, Xin J, Yang H, Gao C (2011) Application of the entropy weight and TOPSIS method in safety evaluation of coal mines. *Procedia Eng* 26:2085–2091
17. Shoosharian L, Safaei F (2019) A maximally robustness embedding algorithm in virtual data centers with multi-attribute node ranking based on TOPSIS. *J Supercomput* 75:8059–8093
18. Zhang P, Li H, Ni Y, Gong F, Li M, Wang F (2019) Security aware virtual network embedding algorithm using information entropy TOPSIS. *J Netw Syst Manag* 28:35–57



19. Kong Y, Dankyi AB, Ankomah-Asare ET, Addo AA (2019) An application of TOPSIS approach in determination of spread influencers in a competitive industrial space: evidence from the banking network of Ghana. *J Hum Resour Sustain Stud* 7(2):93239
20. Demircan, M.L. and Tunc, S. (2019) A proposed service level improvement methodology for public transportation using Interval Type-2 Fuzzy EDAS based on customer satisfaction data. In: *Proceedings: international conference on intelligent and fuzzy systems*, July 23–25, Istanbul, Turkey, pp 1351–1359
21. Yang Y, Yu L, Zhou Z, Chen Y, Kou T (2019) Node importance ranking in complex networks based on multicriteria decision making. *Math Probl Eng* 2019:9728742
22. Barabási AL (2016) *Network science*. Cambridge University Press
23. Derrible S, Kennedy C (2011) Applications of graph theory and network science to transit network design. *Transp Rev* 31(4):495–519
24. Lü L, Chen D, Ren XL, Zhang QM, Zhang YC, Zhou T (2016) Vital nodes identification in complex networks. *Phys Rep* 650:1–63
25. Barthelemy M (2004) Betweenness centrality in large complex networks. *Eur Phys J B* 38(2):163–168
26. Sabidussi G (1966) The centrality index of a graph. *Psychometrika* 31(4):581–603
27. Bavelas A (1950) Communication patterns in task-oriented groups. *J Acoust Soc Am* 22(6):725–730
28. Latora V, Marchiori M (2001) Efficient behavior of small-world networks. *Phys Rev Lett* 87(19):198701
29. Bhatia U, Kumar D, Kodra E, Ganguly AR (2015) Network science-based quantification of resilience demonstrated on the Indian Railways Network. *PLoS ONE* 10(11):e014189