

基于动态客流的城市轨道交通关键站点识别

高超^{1,2}, 蒋世洪², 王震^{1*}, 邓越², 范懿², 李学龙^{1*}

1. 西北工业大学光电与智能研究院, 西安 710072

2. 西南大学计算机与信息科学学院, 重庆 400715

* 通信作者. E-mail: zhenwang0@gmail.com, li@nwpu.edu.cn

收稿日期: 2020-10-03; 修回日期: 2020-12-06; 接受日期: 2021-01-14; 网络出版日期: 2021-09-17

国家科技部重点研发计划(批准号: 2019YFB2102304)、国家自然科学基金(批准号: 61976181, U1803263, 11931015)、重庆市自然科学基金(批准号: cstc2018jcyjAX0274)和中央高校基本科研业务费专项资金(批准号: D5000210738)资助项目

摘要 定量评估城市轨道交通站点的重要性有助于优化城市轨道交通网络, 提升针对突发事件的应急管理能力。现有工作常根据轨道拓扑结构或静态客流的分布来识别关键站点, 然而, 由于居民日常出行行为展现时空变化特征, 它们对关键站点的识别也有重要影响。为此, 本文提出一种结合轨道网络拓扑结构和动态客流的拓扑-客流中心性指标来动态识别轨道交通关键站点。首先, 将轨道交通网络拓扑结构抽象为节点负载网络, 利用节点负载刻画客流时变特征。其次, 利用级联失效模型对比拓扑-客流中心性指标与其他中心性指标对网络平均效率、极大连通系数和损失客流的影响。大量实验表明所提指标能有效识别轨道交通网络中的关键站点。同时, 关键站点会随客流演变展现出动态变化特征, 特别是当客流量波动剧烈时最为显著。

关键词 交通网络, 关键站点, 中心性指标, 动态客流, 级联故障

1 引言

城市轨道交通是公共交通系统的重要组成部分^[1], 能够缓解城市交通拥堵^[2,3]。识别城市轨道交通网络中的关键站点有助于提升轨道交通服务水平和应对突发事件的能力。例如, 识别站点重要性的时序变化特征, 有助于及时设置导流围栏、提前规划安检人员排班日程^[4], 并提高轨道交通网络的抗毁性^[5]。此外, 关键站点的识别能为交通规划提供决策支持^[6]。例如, 将高关键程度站点设置为新建交通线路的换乘站, 可进一步优化网络结构, 提高运力。

现有研究中, 关键站点度量方法主要从拓扑结构和静态客流两个方面展开。基于拓扑结构的方法主要关注轨道网络的拓扑特征, 基于静态客流的方法则结合了节点的网络结构属性以及静态客流分布。例如, Yang 等^[7]提出一种耦合节点度和介数的中心性度量指标。实验表明, 该方法比单独使用度中心

引用格式: 高超, 蒋世洪, 王震, 等. 基于动态客流的城市轨道交通关键站点识别. 中国科学: 信息科学, 2021, 51: 1490–1506, doi: 10.1360/SSI-2020-0303
Gao C, Jiang S H, Wang Z, et al. A novel method to identify influential stations based on dynamic passenger flows (in Chinese). Sci Sin Inform, 2021, 51: 1490–1506, doi: 10.1360/SSI-2020-0303

性或介数中心性能更有效地识别北京地铁枢纽站点. Liu 等^[8]认为地铁站点重要性由站点所处位置、连通程度以及邻居站点的重要性决定, 他们据此提出的方法可以在地铁网络中有效区分度中心性相同的站点. Wu 等^[9]基于站点间的最短路径和最小转移路径确定站点重要性. 大量实验表明, 该方法和度中心性紧密相关, 并且比介数中心性更适合地铁网络. Zhao 等^[10]提出的方案通过融合度中心性、介数中心性、接近度中心性和特征向量中心性来评估上海地铁站点的重要程度, 并用层次分析法计算各中心性指标的权重. 他们发现介数中心性在该方案中发挥了关键作用.

作为一个开放系统, 一个站点在轨道交通中是否重要, 不仅由网络拓扑结构决定, 还与站点的客流特征有关. 例如, Xia 等^[11]将多种节点中心性指标和日平均客流相结合, 设计了站点重要性排序算法来探究上海地铁网络的关键站点. 实验表明, 静态客流分布在评估过程中扮演了关键角色. 同时, 度中心性比集聚系数和接近中心性更能反映拓扑结构对站点重要性的影响. Du 等^[4]首先利用深圳市工作日或休息日的日平均客流构建有向加权网络, 然后聚合多种中心性指标评估地铁站点重要性. 实验表明, 客流分布特征有助于提高关键站点的识别能力.

综上所述, 当前轨道交通关键站点识别方法大都基于网络拓扑结构或结合客流的静态分布. 由于居民日常地铁出行展现的时空演变规律^[12,13], 本文首先基于 L-Space^[14]方法将轨道交通网络的拓扑结构抽象为节点负载网络, 利用节点负载描述客流的时变特征; 其次, 提出一种结合网络拓扑结构和动态客流的拓扑–客流中心性指标 (topology-flow centrality, TFC) 来识别轨道交通系统中的关键站点. 该指标不仅通过聚合度中心性、介数中心性和接近度中心性来度量轨道交通网络拓扑结构对站点重要性的影响, 而且结合站点客流来度量站点重要性变化. 最后, 利用网络平均效率和极大连通系数衡量 TFC 指标的性能. 此外, 本文还定义了损失客流指标进一步量化因站点故障而无法正常出行的乘客人数.

2 模型与方法

2.1 网络模型构建

为刻画轨道交通网络拓扑结构和时变客流对站点重要性的影响, 本文基于真实轨道交通网络拓扑结构和其上的时变客流信息构建节点负载网络 (node load network, NLN). t 时刻的 NLN 可形式化为具有 N 个节点的无向网络 $G_t = \langle V, E, H \rangle$, 其中节点集合 $V = \{v_i | i \in [1, N]\}$ 和边集合 $E = \{e_{ij} | i \neq j\}$ 分别对应轨道交通网络中站点以及站点之间的轨道. $H = \{h_i(t) | i \in [1, N]\}$ 是 t 时刻的节点负载集合, 用以刻画站点 v_i 的客流时变特征. L-Space^[14]由于较好地保留了站点之间的地理特性, 被广泛用于交通系统的网络化表示过程中, 它用网络节点表示交通系统中的站点, 用边连接两个地理位置相邻且位于同一交通线路上的站点. 综上, NLN 的具体构建可分为两步. 首先, 采用 L-Space 方法将轨道交通网络的站点和相邻站点之间的轨道转化为 NLN 的节点和边. 其次, 利用 $f_{ij}(t)$ 量化 t 时刻以站点 v_i 为出发地并以站点 v_j 为目的地的乘客数量, 进而站点 v_i 在 t 时刻的客流可表示为 $h_i(t) = \sum_{j=1}^N f_{ij}(t) + \sum_{j=1}^N f_{ji}(t)$. 其中, $\sum_{j=1}^N f_{ij}(t)$ 和 $\sum_{j=1}^N f_{ji}(t)$ 分别表示在 t 时刻以站点 v_i 为出发地和目的地的总客流, N 表示站点总数. 图 1 展现了 NLN 演化过程及节点负载变化.

2.2 拓扑–客流中心性指标

在交通网络中, 居民更关注站点故障是否会导致出行效率降低. 因此, Xu 等^[15], Yang 等^[7]和 Du 等^[4]在探究交通网络的关键节点时均使用网络鲁棒性指标度量站点的重要程度. 换言之, 交通网络的关键节点被广泛认为是能对网络结构和功能造成最大破坏的节点. 首先, 网络中心性指标被广泛用于

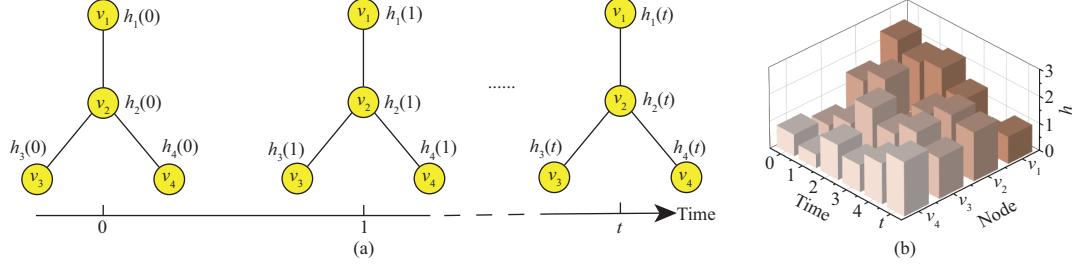


图 1 (网络版彩图) (a) 节点负载网络 (NLN) 演化过程. NLN 是轨道交通网络抽象图, 其网络结构由网络拓扑结构决定, 节点负载来自乘客出行数据的统计分析. (b) 节点负载变化示意图, 代表站点客流的时序变化

Figure 1 (Color online) (a) The evolution process of NLN. NLN is an abstraction of a subway network, whose structure depends on the topology of the subway network. The load of nodes comes from the statistical analysis of passenger travel data. (b) The schematic diagram of changes in the load of nodes, representing the time-varying changes of passenger flows in stations

度量节点关键程度^[16, 17]. 例如, 度中心性 (degree centrality, DC)^[18] 从网络结构局部信息角度识别关键节点; 介数中心性 (betweenness centrality, BC)^[19] 和紧密度中心性 (closeness centrality, CC)^[20] 均从网络结构全局信息的角度刻画节点重要性. 鉴于这些指标仅具备单一视角, 难以全面揭示网络拓扑结构的作用, 本文综合多种中心性指标来评估网络拓扑结构在关键节点识别过程中的影响. 其次, 由于客流信息有助于识别对网络结构有重要影响的站点^[4], 并且最大客流站点在一定条件下触发的级联失效将导致更多站点发生故障^[21], 因此, 具备时变特征的客流信息是刻画关键节点的另一因素. 为此, 在识别轨道交通关键站点时, 本文结合网络拓扑结构和站点客流信息提出 TFC 指标来识别站点在具有时变客流特征的轨道网络中的重要性, 具体定义为

$$TFC_i = T(i)F(h'_i(t)), \quad (1)$$

其中, $T(i)$ 和 $F(h'_i(t))$ 分别为网络结构和节点负载对站点 v_i 的重要性影响, 将分别在 2.2.1 和 2.2.2 小节阐述.

2.2.1 网络拓扑结构影响

为避免落入单一结构视角, 本文使用逼近理想解排序法 (technique for order preference by similarity to ideal solution, TOPSIS)^[22] 融合多种中心性指标来计算网络拓扑结构对站点重要性的影响. 为使结果更合理, 由熵权法^[23] 确定各中心性指标的权重, 并利用相对熵^[24] 提升 TOPSIS 的性能^[25]. 对于轨道交通网络, 本文选取 DC, BC 以及 CC 作为 TOPSIS 的融合对象. 结合 3 种中心性指标的 $T(i)$ 计算如下.

首先, 构建决策矩阵 $D = \{x_{ij}\}_{N \times M}$. 其中, x_{ij} 表示节点 v_i 的第 j 种中心性指标值, N 表示 NLN 的节点总数, M 表示结合的中心性指标数. 由于本文结合了度中心性、介数中心性及紧密度中心性 3 种中心性指标, 故 $M = 3$. 将决策矩阵标准化为 $D' = \{r'_{ij}\}_{N \times M}$.

其次, 利用熵权法计算各中心性指标的权值. 第 j 种中心性指标的信息熵为 $En_j = -\frac{1}{\ln N} \sum_{i=1}^N r'_{ij} \ln r'_{ij}$, 其中, $r'_{ij} = \frac{r_{ij}}{\sum_{k=1}^N r_{kj}}$, 当 $r'_{ij} = 0$ 时, 令 $r'_{ij} \ln r'_{ij} = 0$. 于是, 第 j 种中心性指标的权重为 $c_j = \frac{1-En_j}{\sum_{k=1}^M (1-En_k)}$.

然后, 定义 $v_{ij} = c_j \times r_{ij}$, 并获取加权决策矩阵 $W = \{v_{ij}\}_{N \times M}$. 而后, 定义 $v_j^+ = \max\{v_{ij} | 1 \leq i \leq N, 1 \leq j \leq M\}$

$N\}$ 和 $v_j^- = \min \{v_{ij} | 1 \leq i \leq N\}$, 并获取正理想解 $A^+ = \{v_1^+, \dots, v_M^+\}$ 和负理想解 $A^- = \{v_1^-, \dots, v_M^-\}$.

接着, 基于相对熵计算节点 v_i 与 A^+ 的距离 $R_i^+ = \sum_{j=1}^M \{v_j^+ \log \frac{v_j^+}{v_{ij}} + (1-v_j^+) \log \frac{1-v_j^+}{1-v_{ij}}\}$ 以及与 A^- 的距离 $R_i^- = \sum_{j=1}^M \{v_j^- \log \frac{v_j^-}{v_{ij}} + (1-v_j^-) \log \frac{1-v_j^-}{1-v_{ij}}\}$.

最后确定 $T(i)$:

$$T(i) = \frac{R_i^-}{(R_i^+ + R_i^-)}. \quad (2)$$

2.2.2 负载客流影响

本文利用 $F(h'_i(t))$ 刻画节点负载对站点 v_i 的重要性影响. 由于站点故障会迫使以此站点为目的地或出发地的乘客改变出行计划, 所以拥有大量客流的站点发生故障将导致更为深远的影响. 例如, 客流较大的站点发生故障将导致本站乘客流失, 同时由于乘客重新规划出行路线, 可能增加周边站点的运输压力, 甚至触发级联故障^[21]. 因此, 客流在一定程度上反映了站点的重要程度, 是站点重要性评估过程中的关键因素. Zhao 等^[26] 和 Chen 等^[27] 认为聚集系数对节点重要性的影响可根据其指数函数表达, 受此启发, 定义

$$F(h'_i(t)) = \theta^{h'_i(t)}, \quad (3)$$

其中, 大于 1 的可调参数 θ 使 TFC 指标更为灵活. θ 越大, 表示越重视节点负载的作用, 本文取 $\theta = 2$. 参数 θ 对 TFC 指标性能的影响将在 3.6 小节讨论. $h'_j(t)$ 是 t 时刻节点 v_j 归一化后的节点负载, 定义为

$$h'_i(t) = \frac{h_i(t) - h_{\min}}{h_{\max} - h_{\min}}, \quad (4)$$

其中, h_{\max} 和 h_{\min} 分别表示在不同时刻不同节点的最大和最小节点负载.

图 2 展现了 TFC 指标如何结合网络拓扑结构和节点负载一起度量节点重要的过程. 首先, 图 2(b) 根据图 2(a) 展示的网络拓扑结构获取节点的 3 类中心性指标值. 该过程中, DC, BC 和 CC 度量的最重要节点分别为 v_8 , v_3 和 v_2 . 其次, 图 2(c) 采用熵权法计算各中心性指标的权重, 进而由图 2(d) 根据式 (2) 计算网络拓扑结构对节点重要性的综合影响. 此时, v_3 被认为是最重要的节点. 同时, 图 2(f) 根据图 2(e) 统计的站点客流提取节点负载. 进而由图 2(g) 根据式 (3) 计算负载对节点的重要性影响. 其中, v_5 和 v_2 分别为 $t = 0$ 和 $t = 5$ 时受负载影响最大的节点. 最后, 图 2(h) 利用式 (1) 结合网络拓扑结构和节点负载计算节点重要性. 结果表明 v_8 和 v_2 分别以 1.29 和 1.99 的最大值成为 $t = 0$ 和 $t = 5$ 时的最重要节点. 综上, TFC 指标因同时关注网络拓扑结构和节点负载, 所以能获取更全面的结果, 并能捕捉站点重要性的动态变化.

2.3 负载容量模型

本文以站点故障的影响范围来评估关键站点. 该影响范围包含拓扑结构和客流两方面, 其值越大, 表明站点越重要. 虽然很多级联失效模型 (如耦合印象格子模型^[28]、负载容量模型^[29]、线性阈值模型^[30]) 均可模拟故障在网络中的传播过程, 进而分析故障的影响范围. 但由于轨道交通网络中受初始故障站点影响的其他站点依然具备运输乘客的能力, 这与负载容量模型^[31] 在故障扩散中保留故障节点部分功能的性质相符, 所以本文采用负载容量模型刻画轨道交通网络中的故障传播过程.

在负载容量模型^[31] 中, 边 e_{ij} 的权值 b_{ij} 表示沿此边通信的效率^[31], 在 $t = 0$ 时, $b_{ij} = 1$, 否则 $b_{ij} \in (0, 1]$. 同时, 节点 v_i 具有两种属性 (负载、容量), 其在 t 时刻的负载 $L_i(t)$ 为通过该节点的效率最优路径^[32] 的条数, 其容量定义为 $C_i = \alpha L_i(0)$, $i = 1, 2, \dots, N$. 式中 $\alpha > 1$. 当某节点在 t 时刻因

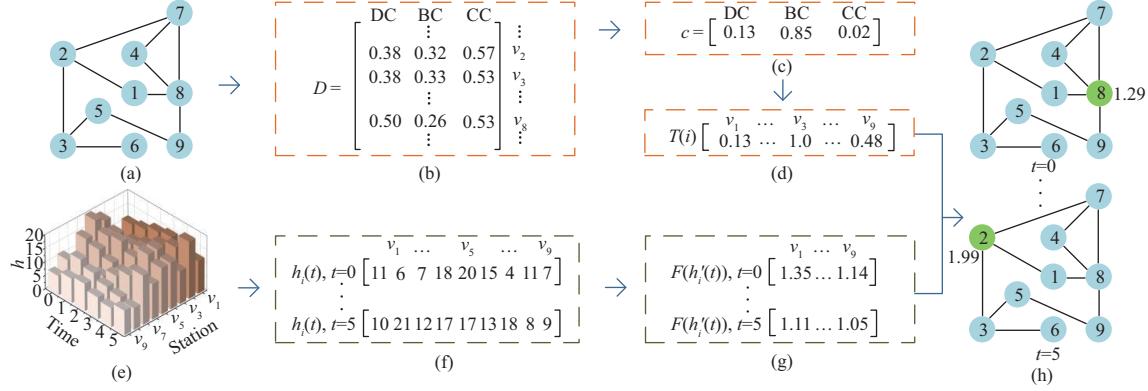


图 2 (网络版彩图) 基于 TFC 指标识别关键站点. 首先, 根据 (a) 网络拓扑结构获取 (b) 节点的 3 种中心性指标值, 并使用熵权法确定 (c) 中心性指标权重, 进而由式 (2) 计算 (d) 网络拓扑结构对站点重要性的影响. 同时, 根据实证数据 (e) 中的站点客流初始化 (f) 节点负载, 进而按式 (3) 计算 (g) 节点负载对站点重要性的影响. 最后, 由式 (1) 结合网络拓扑结构和节点负载确定 (h) 站点重要性. 该例表明 TFC 指标具备识别关键站点动态变化的能力

Figure 2 (Color online) The evaluation of the critical node based on the TFC criterion. First, based on the network topology in (a), we obtain three values of centrality criteria for nodes in (b). And then the entropy weighting method is used to determine the weight of each centrality criterion in (c). The effect of the network topology on each node is calculated by (2), displayed in (d). Simultaneously, based on the passenger flows of stations in the empirical data as shown in (e), the load of nodes in (f) is initialized, and its influence on each node can be calculated by (3) in (g). Finally, the critical node is determined by (1) combining the network topology and the load of nodes in (h). This example shows that the TFC criterion has the capacity of identifying dynamic changes of the influential station

故障被移出网络后, 需根据负载容量模型^[31] 提供的公式

$$b_{ij}(t+1) = \begin{cases} b_{ij}(0) \frac{C_i}{L_i(t)}, & \text{if } L_i(t) > C_i, \\ b_{ij}(0), & \text{if } L_i(t) \leq C_i \end{cases} \quad (5)$$

重新计算各边权值, 并将负载大于容量的节点视为新故障节点. 该过程重复多次, 最终导致级联失效. 需要明确的是, 只有初始故障节点会被移出网络, 而后出现的新故障节点依然保留在网络中. 因为对负载容量模型而言, 故障的级联传播是通过改变与故障节点相连的边的权值来实现, 而非移除故障节点. 在后续实验中, 首先根据不同关键站点评估指标确定引发级联失效的初始故障站点, 待故障的级联传播稳定后, 再利用评估指标度量初始故障站点的重要性.

为度量各种关键节点识别指标的性能, 本文以广泛使用的网络平均效率^[31] 和极大连通系数^[33] 作为评估指标, 它们通过量化网络拓扑结构的性能来评估故障节点的重要性. 同时, 本文进一步从客流视角提出损失客流指标用以度量关键站点. 相关指标定义如下.

(1) 网络平均效率. 该指标用于衡量级联失效对网络的破坏^[32], 是网络中各节点对之间最大效率之和的平均值. 其值越小, 表明破坏规模越大, 对应的初始故障节点越重要. 其表达式为

$$E = \frac{1}{N(N-1)} \sum_{i \neq j} \epsilon_{ij}, \quad (6)$$

其中, N 为网络节点数. ϵ_{ij} 是节点 v_i 和 v_j 之间所有路径的调和效率的最大值^[32], 若不存在路径, 则 $\epsilon_{ij} = 0$.

(2) 最大连通系数. 该指标基于网络规模评估网络鲁棒性, 是故障前后网络最大连通子图的比值.

表 1 上海地铁智能交通卡存储的数据图例

Table 1 Data samples stored in the intelligent transportation card of Shanghai subway

Record	User ID	Date	In_Time	In_Station ID	Out_Time	Out_Station ID
1	100018830	2015-4-13	8:19	v_{112}	8:52	v_{123}
2	100021809	2015-4-14	17:24	v_{45}	17:57	v_{36}
3	100026571	2015-4-15	13:02	v_{252}	13:56	v_{240}

其值越小, 表明网络鲁棒性越差, 对应的初始故障节点越重要. 其表达式为

$$S = \frac{N'}{N}, \quad (7)$$

其中, N 为网络节点总数, N' 表示部分节点发生故障后最大连通子图的节点数.

(3) 损失客流指标 (the lost passenger flow, LPF). 该指标用于量化由于站点故障而放弃乘坐地铁的乘客人数. 边权 b_{ij} 降低意味着乘客沿轨道 e_{ij} 出行的难度增大, 导致 β 比例的乘客放弃乘坐地铁, 使得站点的负载客流 $h_i(t)$ 减少. 同时, 重要站点引发的级联失效使边权下降更多, 进而对负载客流的破坏更严重. 因此, 损失客流指标能够度量节点重要性, 其值越大, 对应的初始故障节点越重要. 由于节点 v_i 在 t 时刻引发级联失效后, 损失的节点负载由两部分组成: 节点负载 $h_i(t)$ 和边权降低导致的节点负载损失, 因而损失客流指标定义为

$$\text{LPF} = h_i(t) + \beta \sum_{j=1}^N h_j(t) \frac{\text{eff}_j - \text{eff}'_j}{\text{eff}_j}, \quad (8)$$

其中, eff_j 和 eff'_j 分别表示级联失效发生前后与节点 v_j 相连边的权值之和. $\beta \in (0, 1)$ 是可调参数, 根据 3.6 小节的参数分析, 本文设 $\beta = 0.3$.

3 实验分析

3.1 数据集与数据分析

3.1.1 数据集

为证明所提 TFC 指标能识别关键站点并探究站点的重要性演变规律, 本文以上海地铁 2015 年 4 月 13 日至 19 日每天 t_6 (6:00~6:59) 至 t_{23} (23:00~23:59) 地铁刷卡记录为基础, 结合轨道拓扑数据度量地铁站点重要性. 其中, 轨道拓扑数据包含上海 14 条交通线路 L_i ($i \in [1, 13]$) 和 L_{16} 以及 289 个站点 v_i ($i \in [1, 289]$). 乘客乘车数据来自上海开放数据创新应用大赛组委会公布的智能交通卡数据¹⁾, 其包含约 1100 万条乘客个人信息和约 1.23 亿次出行记录. 经清洗和处理后, 表 1 列出了该数据的数据图例.

3.1.2 客流统计分析

图 3(a) 刻画了一周内任意两天总客流的 Pearson 相关系数^[34]. 其中, 工作日内每两天的相关系数大于 0.96, 休息日内每两天的相关系数为 0.74. 然而, 工作日和休息日的相关系数较小. 例如, 周一和周六的相关系数仅为 0.23. 图 3(b) 描述了各站点在 2015 年 4 月 14 日的客流情况. 由图可知, 大

1) <http://data.sh.gov.cn/>.

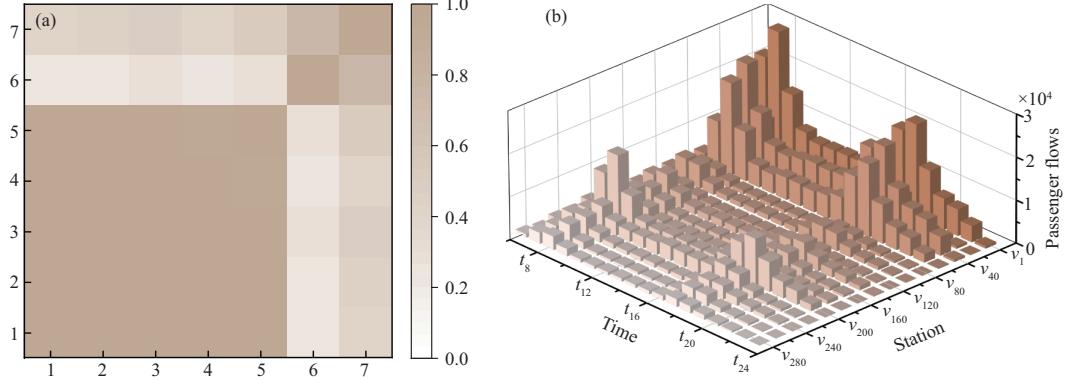


图 3 (网络版彩图) (a) 一周内任意两天的 Pearson 相关系数。横轴和纵轴表示时间。例如, 1 代表星期一。颜色刻画了相关系数的值, 其颜色越深, 表示相关系数越大。由图可知, 工作日或休息日的任意两天非常相似, 但工作日和休息日有显著差异。(b) 各站点在不同时间的负载客流。结果表明, 大多数站点的客流将在 t_8 或 t_{18} 达到最大值, 并且同一时间下不同站点的客流存在差异。这说明客流具有时空演化特征

Figure 3 (Color online) (a) Pearson correlation coefficients for any two days in a week. The horizontal and vertical axes indicate time (i.e., 1 stands for Monday). The color depicts the value of the correlation coefficient. The darker the color, the greater the correlation coefficient. There is a conclusion that any two workdays or two days on the weekend are very similar. However, one workday and one day on the weekend own a dramatic difference. (b) The passenger flows of stations at different time. The results show that most stations will reach the maximum passenger flow at t_8 or t_{18} . Moreover, the passenger flows differ at the same time in different stations. Therefore, the passenger flows have spatiotemporal characteristics

多数站点的客流在 t_8 或 t_{18} 迎来最大值, 但在 t_{10} 到 t_{16} 期间没有显著变化并且数量较少。例如, 站点 v_1 (莘庄) 的客流在 t_8 和 t_{18} 分别为 28770 人和 21662 人, 而在 t_{10} 到 t_{16} 期间的平均值仅为 9836 人。同时, 不同站点在同一时段的客流存在差异。例如, 站点 v_{40} (南京东路) 在 t_8 的客流为 24335 人, 而站点 v_{240} (嘉定新城) 在 t_8 的客流仅为 3230 人。综上, 客流具有显著的时空演变特征。因此, 在识别轨道交通关键站点时, 要充分考虑客流时空演变特征的影响。此外, 由于 t_8 , t_{12} 和 t_{18} 是最能体现上海地铁客流特征的 3 个时段, 所以在后续实验中将以这些时段为数据源。

3.2 识别关键站点

为验证所提 TFC 指标识别关键站点的能力, 选择半局部中心性 (semi-local centrality, SLC)^[35]、加权混合指标 (weighted mixed index, WMI)^[7] 以及相对熵改进的 TOPSIS (TOPSIS with relative entropy, TRE)^[25] 为对比算法。这些指标均采用多视角分析网络拓扑结构对节点中心性的影响。其中, SLC 是度中心性的扩展, WMI 融合了度中心性和介数中心性, TRE 进一步结合了紧密度中心性, 它们均比单一中心性指标 (度中心性、介数中心性、紧密度中心性) 有效^[7, 25, 35]。

由于交通网络的关键节点被广泛认为是能对网络结构和功能造成最大破坏的节点, 所以不同关键节点度量方法的对比分析可由蓄意攻击下的网络鲁棒性实验完成。首先, 选定初始失效节点。即, 采用某种关键节点度量指标量化站点重要性后, 根据排序结果选择 p 比例排名靠前的站点作为初始失效节点。其次, 利用负载容量模型刻画初始失效站点引发的故障传播过程。最后, 通过评估指标 (如网络平均效率 E 、极大连通系数 S 以及损失客流指标 LPF) 刻画初始失效站点的重要性。其中, 网络平均效率越小、极大连通系数越小或损失客流指标越大, 表明初始失效站点越重要。根据 3.1.2 小节的分析, 工作日 t_8 , t_{12} 以及 t_{18} 最能体现上海轨道交通网络的客流特征, 所以本小节实验均在此展开。

为从网络拓扑结构视角验证 TFC 指标性能, 选择网络平均效率 (E) 和极大连通系数 (S) 作为评估指标展开仿真实验。图 4 表明各重要性指标对应的网络平均效率和极大连通系数均随 p 的上升而降

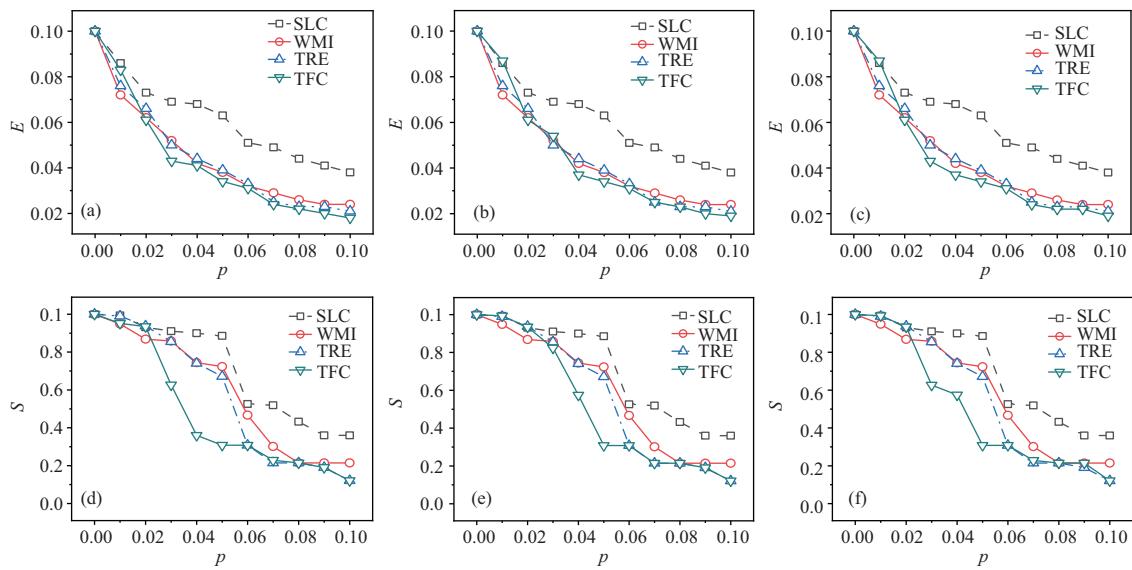


图 4 (网络版彩图) 各重要性指标在不同时段 (t_8, t_{12}, t_{18}) 对应的网络平均效率 (E) 和极大连通系数 (S). p 表示初始失效节点的比例. 结果表明 TFC 指标拥有显著的下降趋势, 可有效识别关键站点. 这源于 TFC 指标能从客流角度审视站点重要性, 有助于发现由网络拓扑结构难以识别的重要站点

Figure 4 (Color online) The network average efficiency (E) and the relative size of giant component (S) corresponding to various importance criteria during different time periods (i.e., t_8, t_{12} and t_{18}). p represents the proportion of initial failure nodes. The results illustrate that the TFC criterion has a significant downward trend, which can effectively identify the influential station. The reason is that the TFC criterion can examine the station's importance from the perspective of passenger flows, helpful in finding important stations which are difficult to be identified by the network topology. (a) E during t_8 ; (b) E during t_{12} ; (c) E during t_{18} ; (d) S during t_8 ; (e) S during t_{12} ; (f) S during t_{18}

低. 其中, SLC 指标的下降趋势最为缓慢, 原因在于它只具备单一结构视角, 难以同时从网络拓扑结构全局信息和局部信息的角度度量节点重要性. 对比 WMI 指标可知, TFC 指标在初始失效节点比例较大时 (如 $p > 0.02$) 总能取得更优异的表现, 这一现象在图 4(d)~(f) 中更为显著. 该结果不仅因为 TFC 指标使用基于熵权法和相对熵的 TOPSIS 整合网络拓扑结构的 3 种中心性指标, 而且源于 TFC 指标将节点负载视为度量关键节点的关键因素. 相较 TRE 指标而言, 虽然 TFC 指标的表现 $p < 0.06$ 时更为优异, 但在 $p \geq 0.06$ 时与 TRE 指标近似. 这是因为 TFC 指标选定的关键站点虽然兼顾网络拓扑结构和节点负载的双重影响, 但是网络平均效率 (E) 和极大连通系数 (S) 仅以网络拓扑结构为衡量标准, 故而难以体现 TFC 指标的优越性. 综上, 当 TFC 指标选定的关键站点发生故障时, 地铁网络将被分割为更小的子网络, 并且站点间的通信效率会严重下降. 因此, 网络拓扑结构和站点客流的组合使得 TFC 指标能精确有效地识别轨道交通网络中的关键站点.

为从客流视角验证 TFC 指标性能, 图 5 进一步分析了损失客流指标 (LPF) 随 p 的变化过程. 结果表明所有重要性指标对应的损失客流均具备时变特征, 在 t_8 和 t_{18} 的值远大于在 t_{12} 的值. 这表明同一站点在不同时段故障后对负载的影响不同, 该现象源于站点客流本身的时变特征. 此外, TFC 指标的结果随 p 的增加而增大并且几乎总能达到最优表现. 这表明由 TFC 指标选定的关键站点发生故障时将损失更多客流, 原因在于它兼顾网络拓扑结构和节点负载的双重影响, 能更全面合理地识别关键站点, 特别是对客流意义重大的站点.

为从全局时间的视角进一步分析所提方法的性能, 本文根据 TFC 指标计算了站点在各时间段的重要性, 并对比了其均值. 图 6 表明 TFC 指标在整体比较中依然占优, 特别是在最大连通系数和损失

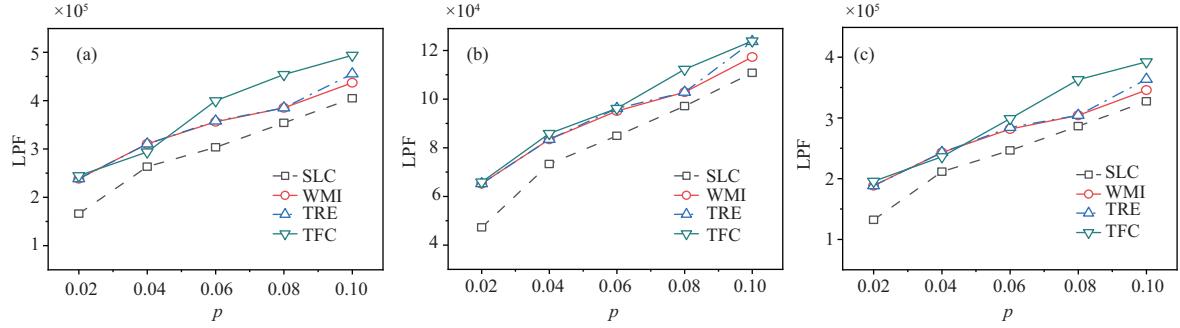


图 5 (网络版彩图) 各重要性指标在 (a) t_8 , (b) t_{12} , (c) t_{18} 关于损失客流指标 (LPF) 的对比结果. p 为初始故障节点的比例. 由图可知, 站点在 t_8 和 t_{12} 发生故障的影响比在 t_{18} 更为严重, 因而站点重要性具备动态变化特征. 同时, TFC 指标因结合网络拓扑结构和节点负载而取得了比其他指标更好的表现

Figure 5 (Color online) The comparison of various importance criteria regarding the LPF during (a) t_8 , (b) t_{12} and (c) t_{18} . p represents the proportion of initial failure nodes. The figure indicates that the failures are more serious during t_8 or t_{18} than t_{12} , revealing the dynamic change characteristic of the influential station. Simultaneously, the TFC criterion has achieved better performance than the compared criteria, because it is the combination of the network topology and the load of nodes

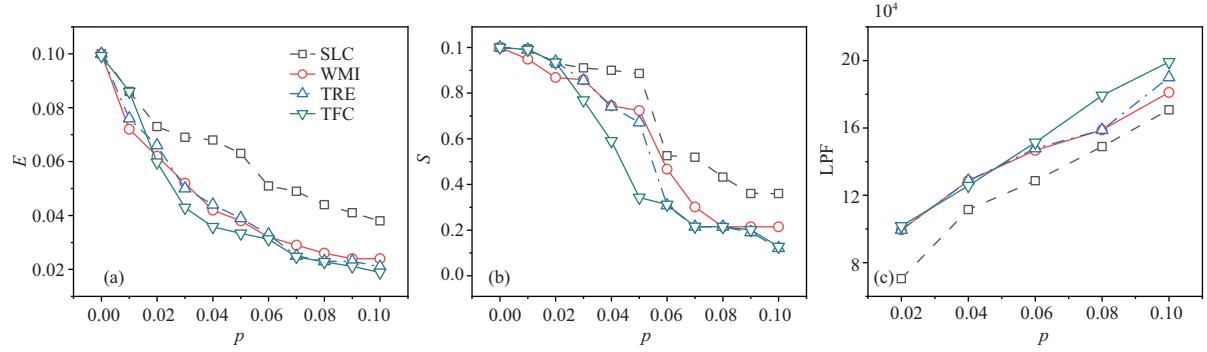


图 6 (网络版彩图) 各重要性指标在全局时间的视角下关于 (a) 网络平均效率 (E)、(b) 最大连通子图 (S) 以及 (c) 损失客流 (LPF) 的对比结果. 其中, 站点关于 TFC 指标的重要性由各时段重要性的均值确定. 结果表明 TFC 指标在各时间点的整体比较中依然占有优势, 这表明网络拓扑结构和客流的结合使得 TFC 指标能够更有效地识别关键站点

Figure 6 (Color online) The comparison results of various importance criteria in terms of (a) the average network efficiency (E), (b) the relative size of giant component (S), and (c) the LPF from the perspective of global time. The importance of stations in the TFC criterion is determined by the average values of the importance of each time period. The results show that the TFC criterion still achieved better performance because it combines the network topology and the passenger flows

客流方面优势显著. 因此, TFC 指标能从全局时间的角度有效识别轨道交通网络的关键站点, 这源于它将站点客流和轨道拓扑结构相结合考虑站点的综合重要性.

3.3 相关性分析

本小节利用相关性分析进一步揭示 TFC 指标优势. 首先, 选择一个站点作为初始故障节点. 而后, 利用负载容量模型刻画该站点引发的故障传播过程. 最后, 使用网络平均效率 (E)、极大连通系数 (S)、损失客流指标 (LPF) 度量该站点造成的影响. 当各站点都被选为初始故障节点后, 通过肯德尔相关系数 (Kendall's correlation coefficient, τ) [36] 量化各评估指标与各重要性指标的相关性. 该系数

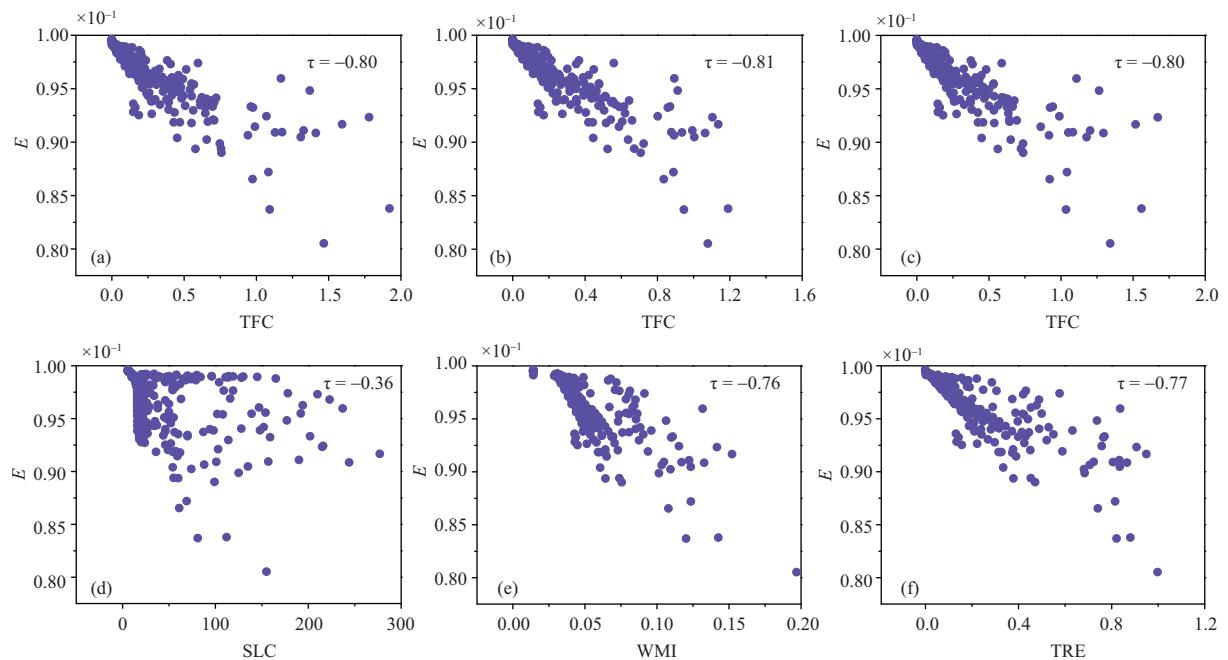


图 7 (网络版彩图) 各 TFC 指标在时段 (a) t_8 , (b) t_{12} , (c) t_{18} 与网络平均效率 (E) 的肯德尔相关系数 (τ), 以及 (d) SLC, (e) WMI, (f) TRE 与网络平均效率 (E) 的肯德尔相关系数 (τ). 图中一个点代表一个站点, 横轴和纵轴分别对应各重要性指标与网络平均效率. 结果表明 TFC 指标因结合了站点客流而能有效识别关键站点, 并且其与网络平均效率的相关性呈现动态变化, 这表明 TFC 指标能揭示关键站点重要性的动态变化特征

Figure 7 (Color online) The Kendall's correlation coefficient (τ) between TFC criterion and E during (a) t_8 , (b) t_{12} and (c) t_{18} , and the τ between E and (d) SLC, (e) WMI and (f) TRE. A point in the figure represents a station, and x -axes and y -axes stand for the importance criterion and the network average efficiency, respectively. The results illustrate that the TFC criterion can effectively identify the influential station owing to the consideration of passenger flows in stations, and its correlation with the network average efficiency shows dynamic changes, which concludes that the TFC criterion can reveal the dynamic change characteristic of the influential station

反映了两种变量的相关程度, 适用于序列间的相关性分析, 同时被广泛用于验证关键节点识别算法的效率^[37~39], 因此本文仍用其评估 TFC 指标性能. 其绝对值越大, 表示相关性越强, 对应的重要性指标越有效. 此外, 为充分展现轨道交通网络的客流特征, 本小节在 t_8 , t_{12} 和 t_{18} 展开实验.

图 7 和表 2 分别揭示了各重要性指标与网络平均效率以及极大连通系数在不同时段的相关性. 结果表明, TFC 指标与网络平均效率在任意时段的相关性 $\tau \geq 0.80$, 该结果略优于 WMI 和 TRE 指标, 远胜于 SLC 指标. 同时, 虽然 4 种重要性指标均未与极大连通系数展现出较强相关性, 但 TFC 指标依然略优于 WMI 和 TRE 指标, 显著优于 SLC 指标. 这不仅说明仅依靠网络拓扑结构局部信息的 SLC 指标在挖掘关键节点时缺乏精准性, 而且说明蕴含节点负载的 TFC 指标比仅使用网络拓扑结构的 WMI 和 TRE 指标更有效. 此外, TFC 指标将节点负载作为关注焦点, 导致它与网络平均效率和极大连通系数的相关性呈现出时变特征. 因此, TFC 指标具备识别站点重要性动态变化特征的能力. 图 8 进一步呈现了损失客流指标与各重要性指标在不同时段的相关性. 其中, TFC 指标与损失客流在任意时段的相关性均为最强, 这表明基于地铁网络拓扑结构和站点客流的 TFC 指标能更全面有效地评估站点重要性.

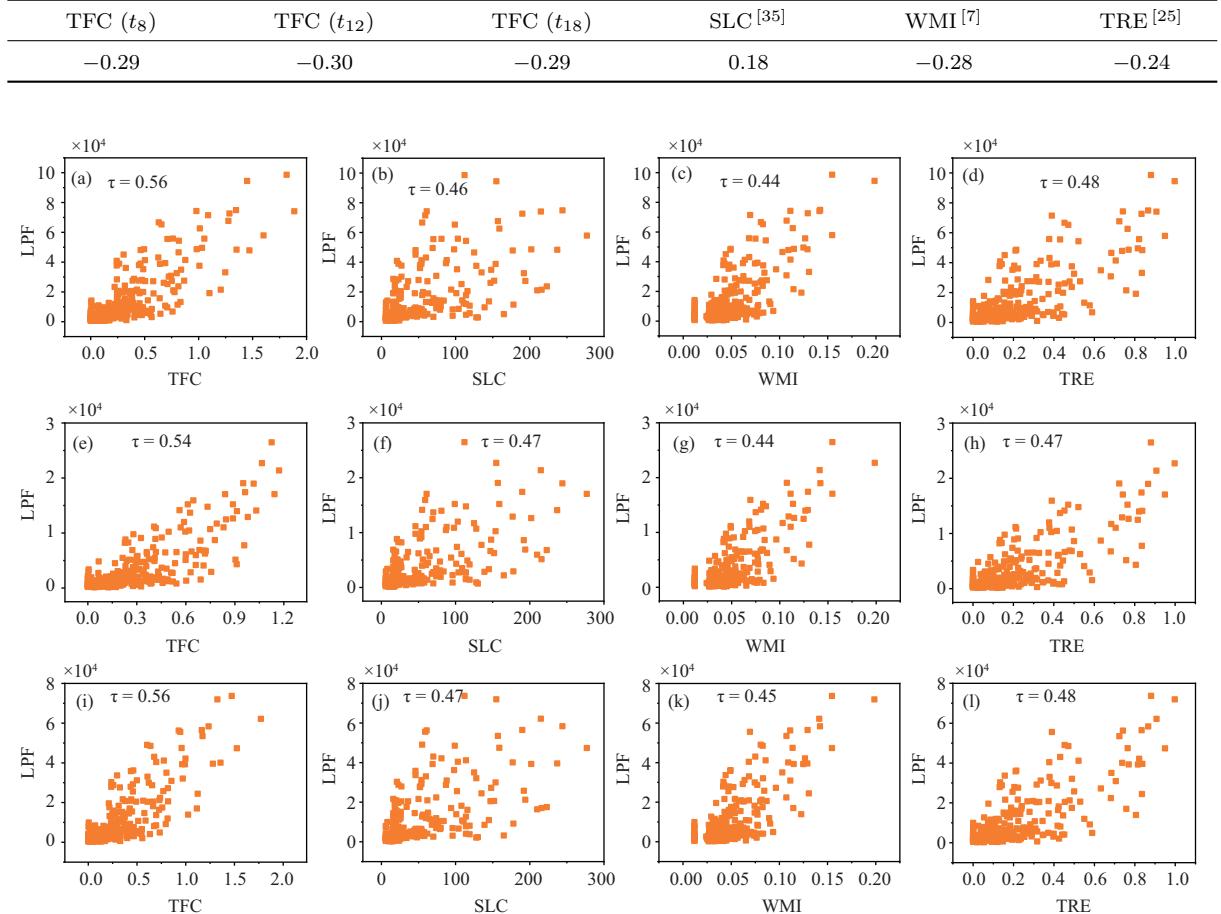
表 2 各重要性指标与极大连通系数 (S) 的肯德尔相关系数 (τ)Table 2 Kendall's correlation coefficient (τ) between various importance criteria and S 

图 8 (网络版彩图) 四类重要性指标在 (a)~(d) t_8 , (e)~(h) t_{12} , (i)~(l) t_{18} 与损失客流指标 (LPF) 的肯德尔相关系数 (τ). 图中一个点代表一个站点, 横轴和纵轴分别是各重要性指标与损失客流指标 (LPF). 由图可知, TFC 指标在任意时段均能取得最佳结果, 这源于该指标兼顾了网络拓扑结构和节点负载的双重影响

Figure 8 (Color online) The Kendall's correlation coefficient (τ) between LPF and four kinds of importance criteria during time periods (a)~(d) t_8 , (e)~(h) t_{12} , and (i)~(l) t_{18} . A point in the figure means a station, and x-axes and y-axes refer to the importance criterion and the LPF, respectively. It can be seen from the figure that the TFC criterion can achieve the best results during all the three time periods. This is because the criterion takes into account the dual effects of the network topology and the load of nodes

3.4 动态变化特征分析

客流的时变特征导致轨道交通网络关键站点的重要性也随时间变化. 为进一步揭示关键站点的重要性随时间的演变规律, 本小节对关键站点重要性的动态变化特征作进一步探究. 首先, 使用 TFC 指标度量站点在工作日和休息日各时间段的重要性. 其次, 在各时段 (time) 按重要性对站点降序排列以获取其排名 (rank). 最后, 以工作日和休息日各时段排名前 20 的站点的并集为研究对象.

图 9 表明, 站点排名在工作日的 t_8 和 t_{18} 会发生对称变化, 然而这些变化的站点在这两个时段以外并无明显变化. 例如在图 9(a) 中, 站点 v_{67} (宝山路) 和 v_{40} (南京东路) 的排名在 t_8 急剧下降或上升后将在 t_{18} 迎来同样变化, 然而它们在其他时间内无显著变化. 站点排名在休息日没有突增或突降的

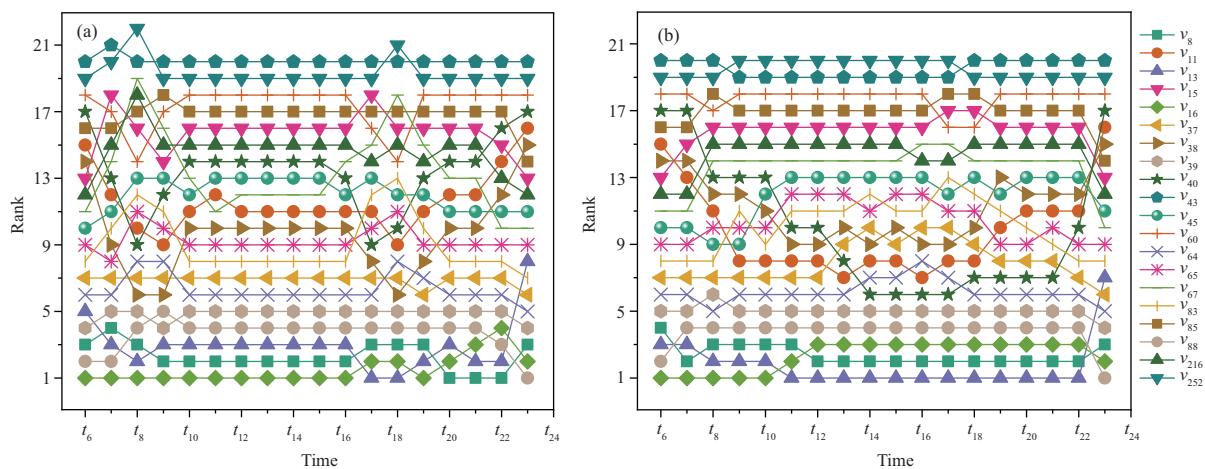


图 9 (网络版彩图)以上海轨道交通网络各时段排名前 20 的站点的交集为研究对象,揭示站点排名 (rank) 随时段 (time) 的演变规律。结果表明关键站点的重要性在客流波动剧烈的时段 (如工作日 t_8 和 t_{18}) 会发生显著变化

Figure 9 (Color online) The evolution of station ranking (rank) with time period (time) is revealed by the intersection of the top 20 stations in each period of the subway network of Shanghai. The results show that the station's importance will change significantly during the periods of violent passenger flow fluctuations (e.g., t_8 and t_{18} on workdays). (a) Workdays; (b) weekend

情况,并且缺乏规律性变化。例如在图 9(b) 中,站点 v_{40} 的排名随时间先上升再降低,而站点 v_{83} (海伦路) 的排名随时间呈现出不规律变化。这些现象和上海轨道交通网络总客流的演变过程相吻合,例如在工作日的 t_8 和 t_{18} ,客流突增的同时部分站点的重要性显著上升。因此,客流演变是站点重要性展现动态变化特征的原因。即,站点重要性会随客流演变展现动态性,特别是当客流波动剧烈时最为显著。此外,实验结果还表明站点客流能弥补站点在地铁网络拓扑结构中的劣势。如图 9(a) 所示,在客流相对稀少的 t_{10} 到 t_{16} ,尽管站点 v_{65} (镇坪路) 的客流比站点 v_{40} 少,但它的排名依然高于站点 v_{40} 。这是因为根据式 (3) 可知,它们仅就地铁网络拓扑结构而言的重要性分别为 0.6 和 0.53。然而在 t_8 和 t_{18} ,站点 v_{40} 的排名会快速上升并超过站点 v_{65} 。这源于在工作日的 t_8 和 t_{18} ,站点 v_{40} 的客流分别为 24335 人和 19535 人,而站点 v_{65} 的客流分别为 13465 人和 9405 人。由此可知,客流是刻画站点重要性的关键考量因素。

客流的时变特征使得轨道站点的重要程度随时间变化,为进一步获取站点的综合重要程度,本文首先采用站点一周客流的平均值消除客流波动性,然后按照 TFC 指标度量站点重要程度。表 3 表明 v_{13} (人民广场) 的 TFC 指标最高,但在其他指标排名中并未取得最优值。在实际运营中,管理者应采用更全面的评估方法获取站点的重要程度,以避免单一结构性指标的不足。同时,应当为重要站点配备充足的安保人员以提升轨道网络安全系数,并为其地面交通增设出租车停靠点或公交换乘站,方便乘客出行。

3.5 实证分析

本小节基于实证分析说明 TFC 指标的性能。首先,通过分析站点的实际数据验证 TFC 指标的正确性。其次,解读站点排名在不同重要性指标上的差异。最后,探究新增枢纽站点在 TFC 指标下的排名。

表 4 表明 TFC 指标在早高峰评定的前两名站点分别为 v_{13} (人民广场) 和 v_{16} (上海火车站),它们不仅是 3 条线路的换乘站,而且拥有大量客流。其中, v_{13} 和 v_{16} 分别是早高峰客流最多和介数中心

表 3 TFC 指标基于平均客流获取的前 5 名站点

Table 3 The top 5 stations obtained by the TFC criterion based on average passenger flows

Station	Rank	TFC	DC	BC	CC
v_{13} (People's Square)	1	1.780	0.021	0.157	0.105
v_{16} (Shanghai Railway Station)	2	1.682	0.013	0.239	0.102
v_8 (Xujiahui)	3	1.555	0.208	0.186	0.098
v_{88} (Century Avenue)	4	1.291	0.028	0.232	0.093
v_{39} (West Nanjing Road)	5	1.287	0.139	0.211	0.106

性最大的站点。因此, 站点 v_{13} 和 v_{16} 对结构和客流均具有重要意义。此外, 虽然站点 v_{39} (南京西路) 在结构和客流上均不占优, 但由于它作为 v_{13} 的邻居站点, 需承载大量通往 v_{13} 的乘客, 因而在 TFC 指标中位列第 7。

根据表 4, 基于 TFC, WMI 以及 TRE 指标的排序结果具有较强的相似性, 但部分站点的排名存在较大差异。站点 v_{38} (静安寺) 分别在 TFC, WMI 和 SLC 指标中排第 4, 第 17 和第 13 位。由于 v_{38} 在地铁系统中早高峰客流量排名第 4, 并且在轨道结构中紧密度中心性排名第 2, 所以 TFC 指标赋予其更高的排位。此外, 站点 v_{41} (陆家嘴) 在 TFC 指标中排名第 13, 在其他指标中均位于 40 名以后。但因为 v_{41} 位于陆家嘴金融贸易区的核心位置, 拥有较高的客流量, 故 TFC 指标赋予其更高的排名。由此可知, TFC 指标聚焦于地铁网络拓扑信息和动态客流信息, 因此能够挖掘出被传统指标忽略的关键站点。

本文探究了新增枢纽站点在 TFC 指标下的排名。2015 年 5 月至 2020 年 10 月期间, TFC 指标排名前 10 的站点中有 4 个成为了新增枢纽站点。例如, 在 TFC 指标中位列第 4 的 v_{38} (静安寺) 被选为新建 14 号线与 2 号线及 7 号线的换乘站、位列第 6 的 v_{11} (陕西南路) 被选为 12 号线新建路段与 1 号线及 10 号线的换乘站。此外, 虽然排在第 1 的 v_{13} (人民广场) 因施工困难未能成为新建计划的换乘站, 但其邻居节点 v_{12} (黄陂南路) 和 v_{172} (大世界) 均被选为新建 14 号线的换乘站。由于这些枢纽站点是管理者结合多方因素论证得出, 因此, 它们在 TFC 指标中如能获得较高排名则表明 TFC 指标具备挖掘潜在关键站点的能力。

3.6 参数分析

为分析参数 θ 对 TFC 指标的性能影响, 本文探究了不同 θ 值时, TFC 指标与网络平均效率之间相关性增量的绝对值 $|\Delta E|$ 。图 10(a) 表明 $|\Delta E|$ 随 θ 的增大而逐步降低并在 $\theta > 2$ 时接近于 0。换言之, 虽然 TFC 指标与网络平均效率之间的相关性会随着 θ 的增大而降低, 但降低幅度极小, 因此本文设 $\theta = 2$ 。图 10(b) 揭示了基于不同 β 值的损失客流指标与关键站点识别方法的相关性。结果表明该相关性随参数 β 的增大而逐渐提升, 并在 $\beta > 0.3$ 时趋于稳定, 因此实验时设为 $\beta = 0.3$ 。

4 结论

本文结合轨道交通网络拓扑结构和站点客流提出了一种新的轨道交通网络关键站点识别指标 TFC。基于实际轨道交通数据的实验分析表明: (1) TFC 能更有效识别关键站点并揭示其重要性动态变化特征; (2) 站点重要性的动态变化在工作日的早高峰 (t_8) 和晚高峰 (t_{18}) 最为显著; (3) 客流激增使得一些在轨道交通网络拓扑结构上并不占优的站点变得更关键。由以上结论知, TFC 能从轨道交通网络拓扑结构和动态客流两个角度识别关键站点, 从而有助于轨道交通运营方制定最优决策, 如根

表 4 根据 TFC 指标获取的前 13 名站点在其他重要性指标中的排名

Table 4 The ranking of the top 13 stations obtained by the TFC criterion when using other importance indicators

Station	TFC (t_8)	SLC [35]	WMI [7]	TRE [25]
v_{13} (People's Square)	1	7	5	3
v_{16} (Shanghai Railway Station)	2	37	2	4
v_8 (Xujiahui)	3	1	3	2
v_{38} (Jing'an Temple)	4	13	17	13
v_{88} (Century Avenue)	5	17	1	1
v_{11} (South Shaanxi Road)	6	3	8	6
v_{39} (West Nanjing Road)	7	2	4	5
v_{37} (Jiangsu Road)	8	11	7	7
v_{40} (East Nanjing Road)	9	16	18	14
v_{64} (Caoyang Road)	10	24	6	8
v_{60} (Yishan Road)	11	5	13	15
v_{65} (Zhenping Road)	12	43	11	10
v_{41} (Lujiazui)	13	72	49	40

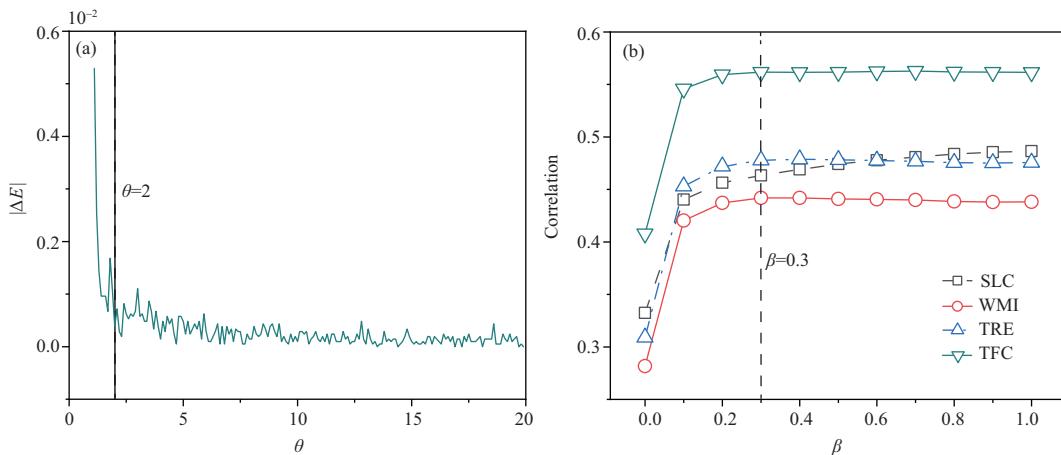
图 10 (网络版彩图) (a) 基于不同参数 (θ) 的 TFC 指标与网络平均效率 (E) 之间相关性增量的绝对值 $|\Delta E|$. (b) 基于不同 β 值的损失客流指标与各重要性指标的相关性

Figure 10 (Color online) (a) The absolute value of the increment of the correlation $|\Delta E|$ between the TFC criterion based on different parameters (θ) and the average network efficiency (E). (b) The correlation between the LPF based on different parameters (β) and various importance criteria

据站点的重要性动态调整安保资源、设计更有效的保护措施以提高地铁系统对恶意攻击或偶然故障的抵抗力.

参考文献

- Phang S Y. Urban rail transit PPPs: survey and risk assessment of recent strategies. *Transp Policy*, 2007, 14: 214–231
- Smyth R, Mishra V, Qian X L. The environment and well-being in urban China. *Ecol Economics*, 2008, 68: 547–555
- de Jong M, Mu R, Stead D, et al. Introducing public-private partnerships for metropolitan subways in China: what is the evidence? *J Transp Geogr*, 2010, 18: 301–313

- 4 Du Z Y, Tang J J, Qi Y, et al. Identifying critical nodes in metro network considering topological potential: a case study in Shenzhen city-China. *Phys A-Stat Mech Its Appl*, 2020, 539: 122926
- 5 Sun L S, Huang Y C, Chen Y Y, et al. Vulnerability assessment of urban rail transit based on multi-static weighted method in Beijing, China. *Transpation Res Part A-Policy Practice*, 2018, 108: 12–24
- 6 Wang J Z, Kong X J, Rahim A, et al. IS2Fun: identification of subway station functions using massive urban data. *IEEE Access*, 2017, 5: 27103–27113
- 7 Yang Y H, Liu Y X, Zhou M X, et al. Robustness assessment of urban rail transit based on complex network theory: a case study of the Beijing subway. *Saf Sci*, 2015, 79: 149–162
- 8 Liu Y J, Tan Y. Complexity modeling and stability analysis of urban subway network: Wuhan city case study. *Procedia Social Behaval Sci*, 2013, 96: 1611–1621
- 9 Wu X T, Dong H R, Tse C K, et al. Analysis of metro network performance from a complex network perspective. *Phys A-Stat Mech Its Appl*, 2018, 492: 553–563
- 10 Zhao L J, Li H Y, Li M C, et al. Location selection of intra-city distribution hubs in the metro-integrated logistics system. *Tunnelling Underground Space Tech*, 2018, 80: 246–256
- 11 Xia F, Wang J Z, Kong X J, et al. Ranking station importance with human mobility patterns using subway network datasets. *IEEE Trans Intell Transp Syst*, 2020, 21: 2840–2852
- 12 Li X H, Guo J Y, Gao C, et al. Network-based transportation system analysis: a case study in a mountain city. *Chaos Solitons Fractals*, 2018, 107: 256–265
- 13 Zhong C, Manley E, Arisona S M, et al. Measuring variability of mobility patterns from multiday smart-card data. *J Comput Sci*, 2015, 9: 125–130
- 14 Sienkiewicz J, Holyst J A. Statistical analysis of 22 public transport networks in Poland. *Phys Rev E*, 2005, 72: 046127
- 15 Xu M, Wu J P, Liu M Q, et al. Discovery of critical nodes in road networks through mining from vehicle trajectories. *IEEE Trans Intell Transp Syst*, 2019, 20: 583–593
- 16 Zhou M Y, Wu X Y, Cao Y, et al. A novel method to identify multiple influential nodes in complex networks. *Sci Sin Inform*, 2019, 49: 1333–1342 [周明洋, 吴向阳, 曹扬, 等. 基于群体影响力的网络传播关键节点选择策略. 中国科学: 信息科学, 2019, 49: 1333–1342]
- 17 Ren X L, Lv L Y. Review of ranking nodes in complex networks. *Chin Sci Bull*, 2014, 59: 1175–1197 [任晓龙, 吕琳媛. 网络重要节点排序方法综述. 科学通报, 2014, 59: 1175–1197]
- 18 Gao C, Su Z, Liu J M, et al. Even central users do not always drive information diffusion. *Commun ACM*, 2019, 62: 61–67
- 19 Newman M E J. A measure of betweenness centrality based on random walks. *Soc Networks*, 2005, 27: 39–54
- 20 Freeman L C. Centrality in social networks conceptual clarification. *Soc Networks*, 1978, 1: 215–239
- 21 Gao C, Fan Y, Jiang S H, et al. Dynamic robustness analysis of a two-layer rail transit network model. *IEEE Trans Intell Transp Syst*, 2021. doi: 10.1109/TITS.2021.3058185
- 22 Du Y X, Gao C, Hu Y, et al. A new method of identifying influential nodes in complex networks based on TOPSIS. *Phys A-Stat Mech Its Appl*, 2014, 399: 57–69
- 23 Guiaşu S. Weighted entropy. *Rep Math Phys*, 1971, 2: 165–179
- 24 Kullback S, Leibler R A. On information and sufficiency. *Ann Math Statist*, 1951, 22: 79–86
- 25 Fei L G, Deng Y. A new method to identify influential nodes based on relative entropy. *Chaos Solitons Fractals*, 2017, 104: 257–267
- 26 Zhao X H, Liu F A, Wang J L, et al. Evaluating influential nodes in social networks by local centrality with a coefficient. *Int J Geo-Inform*, 2017, 6: 35
- 27 Chen D B, Gao H, Lü L, et al. Identifying influential nodes in large-scale directed networks: the role of clustering. *Plos One*, 2013, 8: e77455
- 28 Wang X F, Xu J. Cascading failures in coupled map lattices. *Phys Rev E*, 2004, 70: 056113
- 29 Zhang Z H, Song Y R, Xia L L, et al. A novel load capacity model with a tunable proportion of load redistribution against cascading failures. *Secur Commun Networks*, 2018, 2018: 1–7
- 30 Fan Y, Zhang F, Jiang S H, et al. Dynamic robustness analysis for subway network with spatiotemporal characteristic of passenger flow. *IEEE Access*, 2020, 8: 45544–45555
- 31 Crucitti P, Latora V, Marchiori M. Model for cascading failures in complex networks. *Phys Rev E*, 2004, 69: 045104

- 32 Wang X F, Li X, Chen G R. Network Science: An Introduction. Beijing: Tsinghua University Press, 2006. 270–275
[汪小帆, 李翔, 陈关荣. 复杂网络理论及其应用. 北京: 清华大学出版社, 2006. 270–275]
- 33 Ruan Y R, Lao S Y, Wang J D, et al. Node importance measurement based on neighborhood similarity in complex network. *Acta Phys Sin*, 2017, 66: 038902 [阮逸润, 老松杨, 王竣德, 等. 基于领域相似度的复杂网络节点重要度评估算法. 物理学报, 2017, 66: 038902]
- 34 Bonett D G, Wright T A. Sample size requirements for estimating pearson, kendall and spearman correlations. *Psychometrika*, 2000, 65: 23–28
- 35 Chen D B, Lü L Y, Shang M S, et al. Identifying influential nodes in complex networks. *Phys A-Stat Mech Its Appl*, 2012, 391: 1777–1787
- 36 Kendall M G. A new measure of rank correlation. *Biometrika*, 1938, 30: 81–93
- 37 Kumar S, Panda B S. Identifying influential nodes in social networks: neighborhood coreness based voting approach. *Phys A-Stat Mech Its Appl*, 2020, 553: 124215
- 38 Gao C, Liu J M. Uncovering spatiotemporal characteristics of human online behaviors during extreme events. *Plos One*, 2015, 10: e0138673
- 39 Lv Z W, Zhao N, Xiong F, et al. A novel measure of identifying influential nodes in complex networks. *Phys A-Stat Mech Its Appl*, 2019, 523: 488–497

A novel method to identify influential stations based on dynamic passenger flows

Chao GAO^{1,2}, Shihong JIANG², Zhen WANG^{1*}, Yue DENG², Yi FAN² & Xuelong LI^{1*}

1. School of Artificial Intelligence, Optics and Electronics (iOPEN), Northwestern Polytechnical University, Xi'an 710072, China;

2. College of Computer and Information Science, Southwest University, Chongqing 400715, China

* Corresponding author. E-mail: zhenwang0@gmail.com, li@nwpu.edu.cn

Abstract Quantitative evaluation of station importance in subway networks can help optimize the urban rail transit networks and enhance the management capability for emergencies. Several existing studies have analyzed the important stations based on the rail structure or the static distribution of passenger flows. However, the spatiotemporal characteristics of residents in daily travel also play a crucial role when evaluating the station's importance. Therefore, this paper proposes a novel evaluation method named topology-flow centrality for identifying the important stations by combining the topology of the subway networks and dynamic passenger flows. To begin with, the topology of a rail transit network is abstracted as a node load network, and the load of nodes is used to describe the time-varying characteristics of passenger flows. Then, according to cascading failure, this paper compares the topology-flow centrality and other centrality criteria on the impact of the average network efficiency and the lost passenger flow. Experimental results demonstrate that the proposed criterion can effectively identify the important stations in subway networks. Moreover, the importance of stations displays dynamism with the evolution of passenger flows, especially when the passenger flows fluctuate sharply.

Keywords traffic networks, influential station, centrality measures, dynamic passenger flows, cascading failure



Chao GAO was born in 1980. He received his Ph.D. degree in computer science from Beijing University of Technology, in 2010. He is currently a professor at the School of Artificial Intelligence, Optics and Electronics (iOPEN), Northwestern Polytechnical University. His research interests include complex social networks analysis, nature-inspired computing, and data-driven complex systems modeling.



Shihong JIANG was born in 1996. He is currently pursuing a master's degree at the College of Computer and Information Science, Southwest University, Chongqing, China. His research interests include statistical analysis and data-driven modeling.



Zhen WANG was born in 1986. He received his Ph.D. degree from Hong Kong Baptist University, Hong Kong, in 2014. He is currently a professor at the School of Artificial Intelligence, Optics and Electronics (iOPEN), Northwestern Polytechnical University. His current research interests include complex networks, evolutionary game, and data science.



Xuelong LI is currently a professor at the School of Artificial Intelligence, Optics and Electronics (iOPEN), Northwestern Polytechnical University. His current research interests include artificial intelligence and data analysis.