Information Processing and Management 63 (2026) 104441

Contents lists available at ScienceDirect

Information Processing and Management

journal homepage: www.elsevier.com/locate/infoproman

Surpassing probabilistic based community detection in flow-based mobility networks

Yanzhong Yin [a,b,c,d], Qunyong Wu [a,b,c,d,*]

[a] Key Lab of Spatial Data Mining and Information Sharing of Ministry of Education, Fuzhou University, Fuzhou, Fujian 350108, China
[b] College of Computer and Data Science, Fuzhou University, Fuzhou, Fujian 350108, China
[c] National & Local Joint Engineering Research Center of Satellite Geospatial Information Technology, Fuzhou, Fujian 350108, China
[d] The Academy of Digital China (Fujian), Fuzhou, Fujian 350108, China

ARTICLE INFO

Keywords:
Community detection
Urban integration
Complex network analysis
Element flow

ABSTRACT

Understanding community structures in flow-based mobility networks is critical for analysing regional integration patterns, yet existing methods face two key limitations: (1) current modularity optimization algorithms struggle with resolution limits, and (2) failure to combine both local and global community detection method in flow-based mobility networks. To address these gaps, this study develops a novel framework integrating surpassing probability theory with community detection. The surpassing probability-based Leiden method (SPBL) first reshuffles flow weights to overcome resolution limits in the Leiden algorithm, enabling identification of macro-communities. Next, the two-phase surpassing probability community detection (TPSPCD) algorithm systematically decomposes these communities into granular sub-communities while preserving critical anchor relationships. The framework further introduces an Aggregate Surpassing Degree (ASD) metric to quantify the relative strength of internal versus external community connections. Our results revealed distinct core-periphery patterns within flow-based mobility networks, with strong community cohesion around key node centres. This study concludes that the proposed community detection method effectively captures localized interactions in flow-based mobility networks. This work advances both the theory and application of community detection in flow-based mobility networks, offering planners actionable tools for regional development.

1. Introduction

Mobility flows such as traffic flows, human mobility, and logistics are pervasive in the real world, forming the backbone of flow networks that reflect spatial and temporal dynamics (Cattaneo et al., 2024; Yin, Wu, Lin & Zhao, 2024; M. Zhou, Yang & Chen, 2023). These flow-based networks inherently exhibit multi-level and multi-centric community structures, capturing the complex, hierarchical interactions of various nodes and flows (T. Liu et al., 2024; Yangtianzheng Zhao & Gao, 2024). However, accurately and efficiently identifying such structures within mobility networks has remained a persistent methodological challenge.

Community detection is a powerful tool for analysing mobility flows, enabling the partitioning of network nodes into groups (communities) where nodes within the same group share strong connections, while inter-group connections are sparse (Newman & Girvan, 2004). Researchers have proposed multiple types of community detection methods. Different community detection methods

* Corresponding author.
E-mail address: qywu@fzu.edu.cn (Q. Wu).

https://doi.org/10.1016/j.ipm.2025.104441
Received 27 June 2025; Received in revised form 2 September 2025; Accepted 9 October 2025
Available online 16 October 2025
0306-4573/

are usually designed based on the characteristics of different networks (Molnár, Márton, Horvát & Ercsey-Ravasz, 2024). For example, community detection methods have been widely adopted in social network analysis (Li, Sun & Zia, 2020; Ni, Guo, Wu & Wang, 2023; Ni, Guo, Wu & Wang, 2022; Sayari, Harounabadi & Banirostam, 2025). The community result can help conduct threat analysis (Wang et al., 2024). According to the information they take into account, these methods can be broadly divided into two categories: global methods and local methods.

Global methods utilize information about the entire network structure, such as modularity optimization (Blondel, Guillaume, Lambiotte & Lefebvre, 2008; Traag, Waltman & Van Eck, 2019) or spectral partitioning (Danila, 2016), to identify communities. The Louvain algorithm (Blondel et al., 2008) is a prominent method for identifying high-quality communities by optimizing modularity, a metric quantifying the strength of community partitions. However, the Leiden algorithm (Traag et al., 2019) addresses a key limitation of Louvain by ensuring well-connected communities, thereby improving modularity optimization (Rostami, Oussalah, Berahmand & Farrahi, 2023). Both algorithms rely on modularity maximization, which has become a standard approach for community detection in mobility networks.

Applying modularity as an objective function to detect spatial community directly may lead to a resolution limit problem. The resolution limits of existing approaches hinder the discovery of these intricate patterns. The method of utilizing short cycles to identify intercommunity edges is applied to alleviating the resolution limit problem (W. Liu et al., 2024). However, it only pays attention to the topological relationship of the network and ignores the importance of flow weight in the flow-based mobility network.

In contrast, local methods focus on localized information to discover communities without requiring knowledge of the entire network. The main idea behind local methods is defining local metrics and forming local communities by adding nodes and expanding the local communities. Some methods are based on discovering core nodes and expanding the community from inner to outer side. Recent studies further refine community detection by identify local leaders. The study rethinks the concept of a community from the perspective of local dominance and the core idea is that nodes in a network will determine their own affiliation based on the connection degree of their neighbouring nodes (Shi et al., 2024). To solve the problem of single perspective, a dual-perspective framework (GLC) is proposed (Ruan, Liu, Tang, Guo & Yu, 2025). It balances global and local views to identify influential node. Furthermore, flow volume and flow direction are taken into consideration (Yin, Wu, Zhao & Chen, 2025). It offers a high degree of interpretability.

However, most studies focus on the selection of core nodes. It is important to recognize pairs of strongly interconnected nodes that act as critical anchors within their communities. For instance, in the Guangdong-Hong Kong-Macao Greater Bay Area, cities like Guangzhou-Foshan and Shenzhen-Dongguan exhibit significant reciprocal flows that define their regional influence (Yin, Wu & Li, 2023). By identifying these central edges, we can gain a better understanding of the flow dynamics within the network and the role of specific nodes or edges in maintaining network coherence.

With the development of artificial intelligence, community detection has evolved from classical methods to advanced deep learning techniques, driven by the need to handle large-scale and high-dimensional network data with a large series of attributes. A comprehensive taxonomy of community detection with deep learning is provided (Su et al., 2022). This study categorizes them into deep neural networks, deep nonnegative matrix factorization, and deep sparse filtering. Their survey highlights superiority of deep learning in capturing non-linear relationships but notes challenges in interpretability and computational cost. Community detection methods with deep learning operate as "black boxes," while flow-based mobility networks require transparent, actionable insights for policymaking. Urban planners need to justify why region A and region B are grouped, not rely on unexplainable neural activations.

Over all, the purpose of our paper is to overcome resolution limits in the Leiden algorithm using surpassing probabilistic, identify core node pairs, and locally extend the community based on core node pairs. The framework offers an effective and highly interpretable framework for community detection in flow-based networks, addressing the shortcoming of low interpretability in existing deep learning-based methods, combining global and local methods.

2. Research significance and objectives

2.1. The significance of this research

With the advancement of technologies, spatiotemporal travel data such as navigation data, mobile phone data, and taxi trajectories have provided valuable insights for urban community research. By leveraging these data, we can model networks and mine mobile communities within cities, allowing us to map the travel network structure to urban spatial structures (Hao et al., 2024; Xu, Peng, Lu & Claramunt, 2024). This approach helps in evaluating the rationality and effectiveness of urban spatial boundaries (Jin et al., 2021). Recent researches have utilized navigation services to simulate travel times and employed multi-year time series data as constraints to delineate multi-scale urban communities (Hong et al., 2024; J. Zhang, Liu & Senousi, 2021).

The network with a distinct community structure is shown in Fig. 1(a), but the flow-based mobility networks have the following two types of characteristics. First, due to the intricate flow of elements within these networks, it is rare for there to be a complete absence of flow between communities. As shown in Fig. 1(b), it would result in the formation of short cycles. For instance, in the context of intercity population mobility in China, there is almost always some degree of population flow between provinces, making the formation of short cycles nearly inevitable. Second, in flight-based mobility networks, the flow of airline often occurs between specific countries, as shown in Fig. 1(c). It is difficult to form short rings, resulting in the sparse network. Reliable community detection methods need to be designed based on the characteristics of flow-based mobility networks.

Accurately identifying the community structure in the population mobility network is conducive to deeply grasping the regional development trends and social operation laws, providing an important basis for scientific decision-making. Community detection can reveal the spatial correlation patterns hidden behind the mobility flow data, enabling decision-makers to grasp the overall pattern of

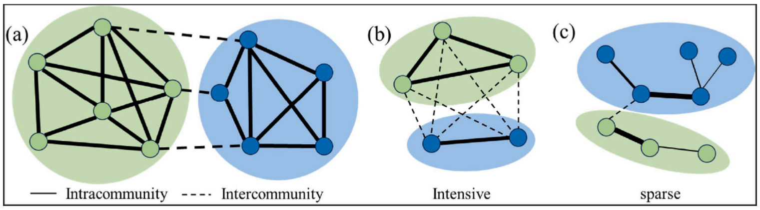

Fig. 1. Examples of different flow-based mobility networks. (a) Obvious community structure. (b) Intensive mobility network. (c) Sparse mobility network.

population migration at the macro scale and identify the flow characteristics of specific groups at the micro level. By analysing the intensity and direction of population interaction among different regions, the functional connections within urban agglomerations can be perceived. For instance, the "daytime commuting circle" centred on the core city or the "weekend leisure circle" across administrative regions can be identified. These findings will directly affect the priority ranking of regional transportation planning and infrastructure construction. In the field of public security, the identification of community structure makes epidemic prevention and control more precise and efficient. When abnormal mobility occurs in a specific community, the risk transmission path can be quickly identified and targeted measures can be taken.

2.2. The objective of this research

In order to identify the precise community structure, reliable community detection methods need to be designed based on the characteristics of flow-based mobility networks mentioned in Section 2.1. The problem of resolution limitation is always faced. Given a flow-based mobility network $G(V,\ E,\ W)$, V is the set of nodes, and E is the set of edges. W is the weight flow of each edge. Modularity (Q) is a widely used quality function for evaluating community partitions in networks. It is defined as:

$$Q = \frac{1}{2m}\sum_{i,j}\left[w_{ij} - \frac{k_i^w k_j^w}{2m}\right]\delta(c_i, c_j) \tag{1}$$

Where, m is the total flow weight, k_i^w is the flow weighted degree of node i, k_j^w is the flow weighted degree of node j , c_i denotes the community of node i, c_j denotes the community of node j, and $\delta(c_i, c_j) = 1$ if i and j belong to the same community, else 0.

The term $\frac{k_i^w k_j^w}{2m}$ represents the expected weight between nodes i and j in a configuration null model, which assumes edges are rewired while preserving each node's weighted degree. Modularity maximization seeks partitions where the actual intracommunity weight $\Sigma_{i,j\in c} w_{ij}$ exceeds the null-model expectation.

Consider a small community c with $k^w(c) = \sum_{i\in c} k_i^w$ (total weighted degree of the community) and internal weight $e^w(c) = \frac{1}{2}\sum_{i,j\in c} w_{ij}$. The modularity contribution of c is:

$$\Delta Q(c) = \frac{1}{2m}\left[e^w(c) - \frac{(k^w(c))^2}{2m}\right] \tag{2}$$

For c to be detected as a community,

$$\Delta Q(c) > 0, e^w(c) > \frac{(k^w(c))^2}{2m} \tag{3}$$

The resolution limit emerges because this condition becomes too lenient for small communities. Specifically, when the community's total weighted degree $k^w(c)$ is small relative to $\sqrt{2m}$, the null-model expectation $\frac{(k^w(c))^2}{2m}$ becomes smaller than the minimal possible intracommunity weight (e.g., a single edge with weight w_{min}), making it impossible to distinguish meaningful communities from random fluctuations.

To address this problem, the key solution is to reduce the flow weight of intercommunity edges. As per prior studies (Gao, Yu & Zhang, 2022; W. Liu et al., 2024), intracommunity edges are tended to form short cycles within their neighbours. However, the method of utilizing short cycles to identify intercommunity edges may not be applicable in flow-based mobility networks, as described in Section 2.1. In this paper, we tend to introduce a surpassing probabilistic-based method to reduce resolution limit problem on modeling community structures in flow-based mobility networks.

In flow-based networks, central edges represent critical pathways through which information, resources, or influence is transmitted. By identifying these central edges, we can gain a better understanding of the flow dynamics within the network and the role of specific nodes or edges in maintaining network coherence. The main objectives of this paper are listed below:

- Introducing a surpassing probabilistic-based method to reduce resolution limit problem on modeling community structures in flow-based mobility networks.
- Proposing a flow-based surpassing probability for accurate identification of flow-based core node pairs.

- Enhancing the interpretability of modeling community structures in flow-based mobility networks, combining global and local methods.

To address the aforementioned objectives, our paper is structured as follows: Section 1 presents the introduction, while Section 2 outlines our research objectives and significance. Section 3 provides related concepts. Section 4 introduces the research methodology and give computation complexity analysis. Section 5 explains the proposed measurement with a numerical example. Section 6 validates the model by comparing them with other models. Section 7 illustrates potential application scenarios using a typical case in China. Finally, Section 8 presents the discussions and conclusions.

3. Related concepts

In this section, we tend to describe several concepts related to our method and input data. These concepts include flow-based mobility network, flow-based mobility community, flow-based core node pairs and flow-based surpassing probability.

Definition 1. (Flow-based mobility network): A flow-based mobility network refers to a transportation or communication network where the movement of entities such as vehicles, elements or people is optimized based on the flow of traffic through various interconnected nodes or routes. Flow refers to the rate or volume of movement across the network. It contains both complex population flows such as large−scale movements populated areas and sparse flows such as less frequent or less dense movement or less trafficked regions.

Definition 2. (Flow-based mobility community): Flow is a key feature of flow-based mobility network. Community has no universally accepted formal definition (Sperlí, 2019). Therefore, considering the characteristics of flow-based mobility network, we define a flow-based mobility community as follows: A flow-based mobility community refers to a set of nodes where the probability that intracommunity flow intensities exceed intercommunity flow intensities is statistically dominant– a characteristic that defines it as a cohesive unit with internally reinforced mobility patterns. This definition emphasizes the importance of flow dynamics in determining community boundaries, where the strength of interactions within a community is notably higher than between communities.

Definition 3. (Flow-based core node pairs): In a flow-based mobility network, when a connection is established between node A and node B, it is observed that A frequently interacts with B, and conversely, B also frequently interacts with A. This mutual frequency of interaction indicates that A and B can be considered a core node pair.

Such a relationship not only demonstrates a high volume of interactions but also reflects a significant level of stability in the flow between them. This means that the probabilities of successful flows from A to B and from B to A both exceed the established threshold τ (default 1). This bidirectional high mobility suggests that the relationship between A and B is robust, potentially driven by shared interests, frequent engagements, or other mobility network dynamics.

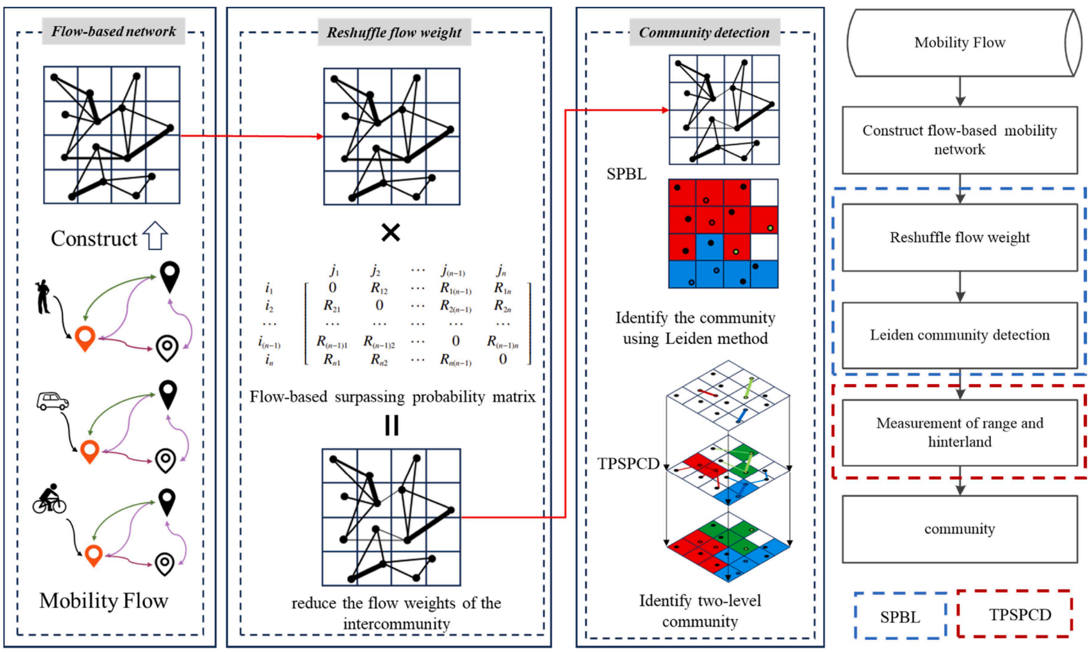

Fig. 3. Workflow of this research.

Definition 4. (Flow-based surpassing probability): The flow weight of edges is a key feature in flow-based mobility network, so that the related importance of flow edges must be taken into consideration (Yin et al., 2025). Therefore, we define the flow-driven surpassing probability to reshuffle the flow weight. The surpassing probability is the probability that the weight of the edge from node h to node i is greater than the weights of other edges originating from u in the same direction. It was applied in identifying key node in flow-based complex network.

The calculation of flow-based surpassing probability is as blows:

$$T_{ij} = P(w_{ij} > w_{jt},\ \forall t \in N(j)) \tag{4}$$

$$T_{ji} = P(w_{ji} > w_{it},\ \forall t \in N(i)) \tag{5}$$

Where, $N(i)$ is the node set of neighbors of node i, excluding node i. $N(j)$ is the node set of neighbors of node j, excluding node j. T_{ij} is the transcendence index of edge e_{ij}, which represents the probability that the flow along edge e_{ij} exceeds the flow along other edges connected to node j. T_{ij} is a numerical value. The range of this value is from 0 % to 100 %.

4. Research methodology

First of all, we construct a flow-based mobility network from diverse mobility flow datasets to represent spatial interactions, where nodes denote spatial units and weighted edges indicate the intensity of mobility flow between them. The framework of this study contains two main methods, as illustrated in the Fig. 3. For one thing, we propose surpassing probability-based Leiden method (SPBL) to overcome resolution limits in the Leiden algorithm, enabling identification of macro-communities. The main idea of this method is that a flow-weight reshuffling process is introduced to enhance the network's structural clarity and reduce the weights of intercommunity connections to highlight internal connectivity patterns. Detailed information can be found in Section 4.1. Secondly, we propose a two-phase surpassing probability community detection algorithm (TPSPCD) to identify hinterland. More information can be found in Section 4.2. Our workflow combines both global and local algorithms. The identified communities are analysed to reveal distinct spatial interaction patterns and mobility-driven regional structures.

4.1. *Surpassing probability-based Leiden method*

To address the resolution limit problem in community detection, we propose the surpassing probability as a metric to reshuffle the flow weights within the network. The resolution limit problem arises when community detection algorithms struggle to identify smaller communities or differentiate between communities that are interconnected, often due to insufficient or imbalanced edge weights. Traditional methods may fail to accurately capture the subtleties in network structures, especially when the flow distribution varies significantly across edges.

By leveraging surpassing probability, we aim to enhance the discriminatory power of the flow weights. The surpassing probability quantifies the likelihood that the flow between two nodes exceeds the flow between other nodes in the same direction, which helps to prioritize edges that carry more substantial flow. This reshuffling process effectively redistributes flow weights, enabling the detection of finer community structures and overcoming the resolution limit that would otherwise blur the boundaries between communities. We offer the example of process of SPBL method, as shown in Fig. 4. The surpassing probability-based Leiden method (SPBL) involves several steps to calculate the surpassing probability and then use it to enhance community detection in flow-based complex networks. Here's a detailed step-by-step process:

Reshuffle the Flow Weights:

For each edge (u, v), select the maximum of the two surpassing probabilities P_{uv} and P_{vu}. Let's denote this maximum surpassing probability as $P_{max}(u, v)$.

Update the weight of each edge (u, v) by multiplying the original weight w_{uv} with $P_{max}(u, v)$. The new weight w'_{uv} is given by:

$$w'_{uv} = w_{uv} \times P_{max}(u, v)^4 \tag{6}$$

Apply the Leiden Community Detection Algorithm:

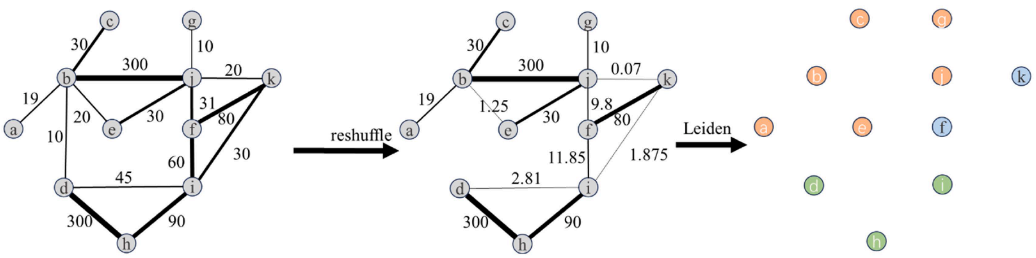

Fig. 4. Example of applying surpassing probability-based Leiden method.

Use the Leiden algorithm on the reshuffled network with the updated edge weights w'_{uv} to detect communities. The Leiden algorithm is an iterative method that optimizes the modularity of the partition by moving nodes between communities and merging communities.

4.2. Two-phase surpassing probability community detection algorithm

4.2.1. Identification of core node pairs

The algorithm starts by iterating through all nodes in the graph, evaluating each node's neighbors to calculate the surpassing probability. This probability determines whether the edge between node and its neighbor is stronger than the edges connected to other neighboring nodes. If the surpassing probability exceeds a predefined threshold and the reverse edge exists in the list, the algorithm adds the edge to a set of first-level community edges. After processing all edges, the algorithm finds connected components in the graph based on these edges, identifying first-level communities.

4.2.2. Community extension

The community expansion phase, also known as Phase 2, further optimizes the community structure. During this phase, the list of edges not selected in Phase 1 is first re-evaluated. Subsequently, the algorithm used in Phase 1 is applied to calculate the surpassing probability of edges with a probability value exceeding the preset threshold are then classified into the second set of edges. After that, the algorithm identifies connected components in the graph once again. Unlike Phase 1, however, this identification process is based on the refined edge set, ultimately enabling the recognition of more localized second-level communities.

The reason for designing the second phase is to further expand the community structure based on the results from the core node pair identification in the first phase. In this phase, nodes that are most closely connected to the core node pairs are considered to be part of the same community. Unlike the first phase, which focuses on mutual interactions between node pairs, the second phase shifts its attention to the importance of unidirectional edges.

4.2.3. The pseudocode and computational complexity analysis

The pseudocode is as follows:

```
Input: G(V, E, W): Graph with nodes, edges, and weights; τ: threshold
Output: C1: First-level communities; C2: Second-level communities
Initialize an empty list T
for each edge (u, v) ∈ E do
    Compute the flow-based surpassing probability T(u, v) for nodes u and v
    Append (u, v, T(u, v)) to T
end
Stage 1: Identification of core node pairs
Initialize an empty set E1
for each (u, v, T(u, v)) ∈ T do
    if T(u, v) ≥ τ then
        Add edge (u, v) to E1
    end
end
Find connected components in G(V, E1)
Store connected components as first-level communities C1
Stage 2: Community extension
Initialize an empty set E2
for each (u, v, T(u, v)) ∈ T do
    if T(u, v) ≥ τ and (u, v) ∉ E1 then
        Add edge (u, v) to E2
    end
end
Find connected components in G(V, E2)
Store connected components as second-level communities C2
return C1 (First-level communities), C2 (Second-level communities)
```

The computational complexity of the TPSPCD algorithm is primarily determined by the calculation of the transcendence index and the subsequent community detection process. For each edge in the network, the algorithm computes the transcendence index $T(u,v)$, which requires comparing the edge weight with the weights of all neighbouring edges of the corresponding nodes. Assuming the average degree of the graph is $\overline{d}$, this step incurs a time complexity of $O(|E|\overline{d})$. The first stage of the algorithm iterates over the set of edges once more to filter and construct the candidate edge set E_1, followed by identifying connected components, which can be accomplished in nearly linear time $O(|V| + |E|)$. The second stage repeats a similar procedure to build the extended edge set E_2 and detect additional connected components, also within $O(|V| + |E|)$. Therefore, the overall computational complexity of TPSPCD can be expressed as $O(|E|\overline{d} + |V| + |E|)$, which reduces to $O(|E|\overline{d})$ in general cases.

4.3. Evaluation metric

4.3.1. Normalized mutual information (NMI)

In community detection, the Normalized Mutual Information (NMI) is a widely used measure to evaluate the similarity between two community partitions of a network. It helps in comparing the results of different community detection algorithms or assessing how well a detected partition matches a ground-truth partition.

$$NMI(X, Y) = \frac{MI(X, Y)}{\sqrt{H(X) \cdot H(Y)}} \tag{7}$$

$$MI(X, Y) = \sum_{x \in X} \sum_{y \in Y} p(x, y) \log \frac{p(x, y)}{p(x)p(y)} \tag{8}$$

$$H(X) = -\sum_{x \in X} p(x) \log p(x) \tag{9}$$

$$H(Y) = -\sum_{y \in Y} p(y) \log p(y) \tag{10}$$

where MI(X, Y) is the mutual information between two random variables X and Y, and $H(X)$ and $H(Y)$ are the entropies of the random variables X and Y, respectively.

4.3.2. Aggregate surpassing degree (ASD)

The Aggregate Surpassing Degree (ASD) of a community is a novel measure of how strong the internal edges of the community are compared to the external edges.

For each edge within a community, the surpassing degree is calculated as the probability that the weight of the edge is greater than the weights of the edges connecting the same node to nodes outside the community. The ASD is the average surpassing degree of all internal edges within a community. Then, we introduce community-sized penalties and use logarithmic penalties. The final ASD is the weighted average of the revised ASD for all communities, with the weight being the number of internal edges in each community.

Formally,

$$ASD(C) = \frac{\sum_{(u,v) \epsilon E_C} P(W(u, v) > W(u, w)),\ \forall w \notin C}{|E_C| \cdot \log(|C|)} \tag{11}$$

$$ASD = \frac{\sum_{C \in Com} ASD(C) \cdot |E_C|}{\sum_{C \in Com} |E_C|} \tag{12}$$

Where, E_C is the set of internal edges within community C. $P(W(u, v) > W(u, w))$ is the probability that the weight of internal edge (u, v) is great than the weight of edges connecting node u to nodes outside the community w. $|C|$ is the number of nodes in community C ($C \geq 2$, single-node community has no internal edge, directly excluded). Com is a collection of all communities.

Furthermore, we construct a flow-based complex network (Fig. 5(a)) to demonstrate the error on identifying community using Leiden method and explain the advantage of the ASD metric. Given that mobility networks are inherently flow-imbalanced, the ASD metric provides more meaningful evaluation than modularity. Community detection methods optimizing for directional flow dominance are essential for mobility patterns, whereas traditional modularity-based approaches overlook the characteristics of flow-based mobility networks. Leiden community detection method (Traag et al., 2019) has been widely applied in complex networks. The result is shown in Fig. 5(b). Edges between nodes in different communities are typically classified as intercommunity edges. However, considering the flow weight of mobility, an edge may appear to connect two communities but, upon the flow weight of connections, may not be an appropriate representative of an intercommunity relationship. As illustrated in Fig. 5(c), the edge marked in red connects a node in the green community with a node in the blue community. The flow weight between node h and node j is 90. $90 > 60$ and $90 > 30$. According to common sense, node i should belong to community B. The flow (i, f) and (i, k) should belong to intercommunity flow.

We applied the value of ASD to evaluate the quality of community result. The ASD value calculated using SPBL and TPSPCD was 0.638, while the ASD value obtained through the Leiden method was 0.577. Considering modularity, the modularity value calculated using SPBL and TPSPCD was 0.469, and the modularity value obtained through the Leiden method was 0.470. As described above, considering the characteristic of flow-based mobility network, SPBL and TPSPCD better capture directional flow imbalances in the flow-based mobility network, where internal movements more strongly dominate external movements compared to Leiden. This aligns

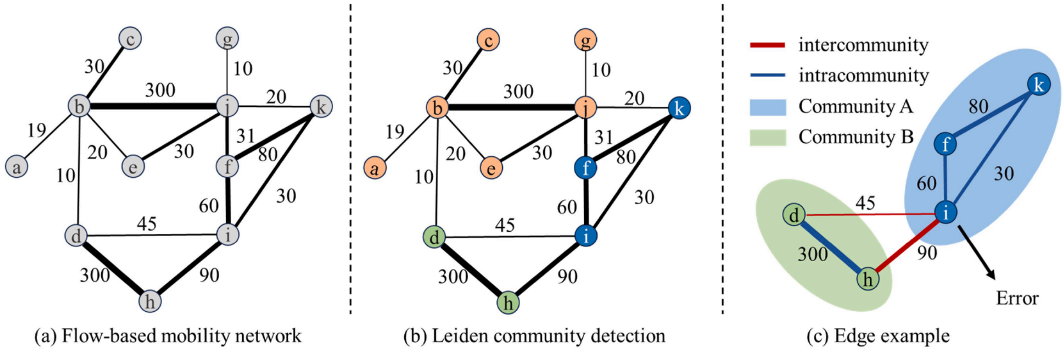

Fig. 5. Examples of error on identifying community. (a) Flow-based mobility network. (b) Leiden community detection result. (c) Edge example.

perfectly with ASD's design purpose of measuring directional dominance.

Unlike modularity, which aggregates edge weights symmetrically and ignores directional flow weight significance, ASD explicitly quantifies the imbalance between internal and external edge strengths. For each internal edge, ASD computes the probability that its weight surpasses those of edges bridging to external nodes, thereby detecting communities where internal flows dominantly outweigh external interactions. This makes ASD particularly suited for directed or flow-imbalanced systems where modularity may overlook community boundaries. A high ASD value indicates that internal edges are consistently stronger than external connections, reflecting functional cohesion. This aligns with flow-based community definitions, where communities emerge from directional flow patterns rather than mere structural clustering.

5. Numerical simulation demonstration

In this section, we use the example mentioned in Section 4.3.2 to demonstrate how to calculate flow-based surpassing probability. For example, as shown in Fig. 6, considering the flow value 90:

- $90 < 300$ (one comparison is not satisfied),
- $90 > 60$, $90 > 30$ and $90 > 45$ (three comparisons are satisfied).

Thus, the surpassing probability for the edge with flow 90 is 100 %.

We present the process of community detection based on the example described in Section 4.3.2. First, we constructed a flow-based mobility network. Then, we reshuffled the flow weights of each edge. The Leiden method was applied to detect the large community, referred to as SPBL. Subsequently, we employed the TPSPCD algorithm to identify the hinterland of the flow-based mobility network. The overall process is illustrated in Fig. 7.

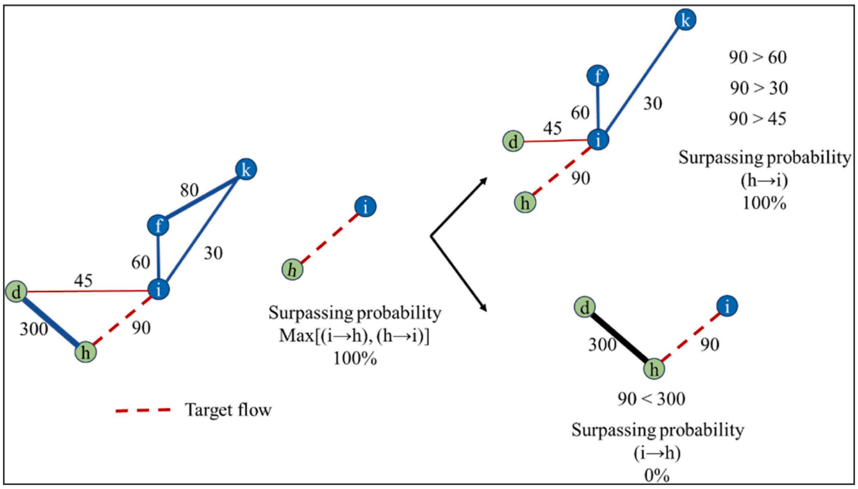

Fig. 6. Examples of calculating surpassing probability.

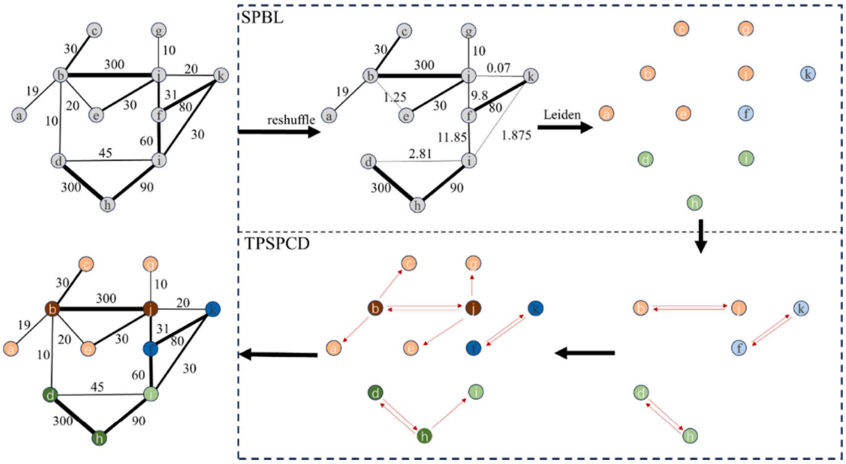

Fig. 7. Examples of SPBL and TPSPCD.

6. Validation experiments

In this section, we first validate the proposed method on eight flow-based mobility datasets. To strengthen the objectivity of the assessment, we furthermore use a series of synthetic benchmark networks, applying Normalized Mutual Information (NMI). The community detection methods evaluated include Louvain (Newman & Girvan, 2004), Greedy-modularity (Clauset, Newman & Moore, 2004), LPA (Raghavan, Albert & Kumara, 2007), pycombo (Sobolevsky, Campari, Belyi & Ratti, 2014), pair (Bonald, Charpentier, Galland & Hollocou, 2018), Leiden (Traag et al., 2019) and Short cycle approach (LSSCL) (W. Liu et al., 2024). The reason is that these methods are widely used in community detection. In addition, our method is designed by the characteristic of flow-based mobility network to solve the limitation of LSSCL.

6.1. *Comparison in real-world datasets*

6.1.1. *Real-world dataset description*

Table 1 and Table 2 provide the description of eight flow-based mobility networks, containing different types of flow-based mobility networks. The types of flow-based mobility networks are as followed.

6.1.2. *Parameter sensitivity analysis*

Prior to executing the TPSPCD algorithm, the threshold τ should be predefined. In this section, we conduct several experiments, employing diverse threshold τ, to evaluate the sensitivity of this parameter, as shown in Fig. 8. k is the number of communities. We set the default threshold $\tau = 1$ for definition 3 and justify this choice from both empirical and domain-specific perspectives.

For one thing, the number of communities exhibits a pronounced downward trend as τ decreases. These experiments indicate that lowering the threshold introduces many weak edges that progressively merge smaller, well-defined communities into larger, loosely connected components. This merging process reduces the total number of detected communities, as evidenced by the pronounced downward trend.

For another, in flow-based mobility networks, a core node pair is expected to capture a tightly-bound origin–destination pair that acts as a single functional unit. Setting $\tau = 1$ is therefore intrinsically consistent with Definition 3. In other words, setting $\tau = 1$ enforces

Table 1
The description of different flow-based mobility networks.

Type	Dataset	Description
Airline flow network	USAir 97	The USAir 97 dataset contains a list of all American Airlines flights in 1997
	Flight	The dataset has been collected from FlightAware (Diop, Cherifi, Diallo & Cherifi, 2021, 2023). The dataset spans six days, specifically from May 17th to May 22nd, 2018.
Vehicular flow network	Car	The network was constructed using electronic toll collection (ETC) data from Taiwan Province (collected on November 15, 2024).
Population mobility network	Xiamen	The mobile phone positioning data used in this research was purchased from third-party data service providers, covering data from about 1.4 million mobile phone users in Xiamen City from March 6 to March 12, 2023. By identifying where individuals stay, we build an OD matrix based on movement frequency and finally get flow-based mobility networks.
Vessel mobility network	Container Bulk Oil GC	we compiled a dataset from vessel traffic tracking platforms (https://www.shipxy.com/). The vessels were categorized into four primary types: Bulk Carriers (Bulk), General Cargo Ships (GC), Chemical/Oil Tankers (Oil), and Container Ships (Container), observed from March 26, 2025, to April 26, 2025.

Table 2
The information of different flow-based mobility networks. C is average clustering coefficient. AverD is average degree. AverWD is average weighted degree.

Datasets	Nodes	Edges	C	AverD	AverWD
USAir 97	332	2126	0.625	12	0.924
Flight	2735	16,665	0.464	12	3314
Car	335	21,183	0.607	126	15,803
Xiamen	4770	683,366	0.447	286	4076
Container	168	1124	0.623	13	274
Bulk	155	1388	0.598	18	220
Oil	182	1422	0.559	15	235
GC	182	213	0.538	14	213

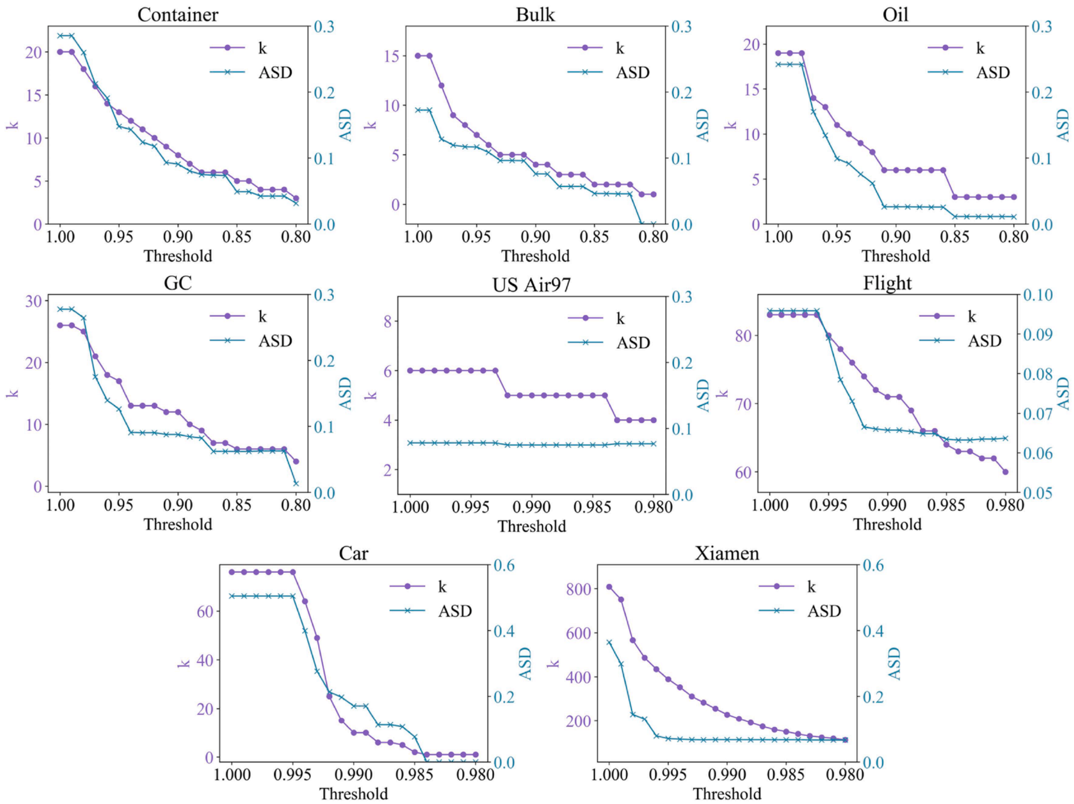

Fig. 8. Parameter sensitivity analysis. The corresponding experiments of eight datasets are shown in the figure, respectively.

that the surpassing probability in both directions equals 1, which implies that each of the two nodes simultaneously outranks every other neighbour of the other node. Consequently, the resulting core pair is strictly limited to exactly two nodes—no third node can satisfy this simultaneous bilateral-dominance condition.

6.1.3. *Effectiveness analysis of reshuffling strategy*

In this section, we will demonstrate the advantage of our methods on reshuffling flow weight in flow-based mobility networks. We conduct a set of experiments that cover (1) a head-to-head running-time comparison against the competitive and similar re-shuffling strategy, (2) An ablation study that disentangles the contribution of this design strategy.

Firstly, we compared the runtime of two reshuffle methods across eight datasets. The result is shown in Table 3. Across all datasets, the "Runtime (SPBL)" is significantly shorter than the "Runtime (LSSCL)". On the US Air 97 dataset, the short cycle approach takes 2.0 s

Table 3
The runtime of different reshuffle methods.

Datasets	Runtime	
	LSSCL	SPBL
US Air 97	2 s	**0.6 s**
Flight	25.1 s	**21.12 s**
Car	175.05 s	**25.25 s**
Xiamen	–	**1502.06 s**
Container	0.99 s	**0.81 s**
Bulk	1.5 s	**0.99 s**
Oil	1.42 s	**0.98 s**
GC	1.38 s	**1.1 s**

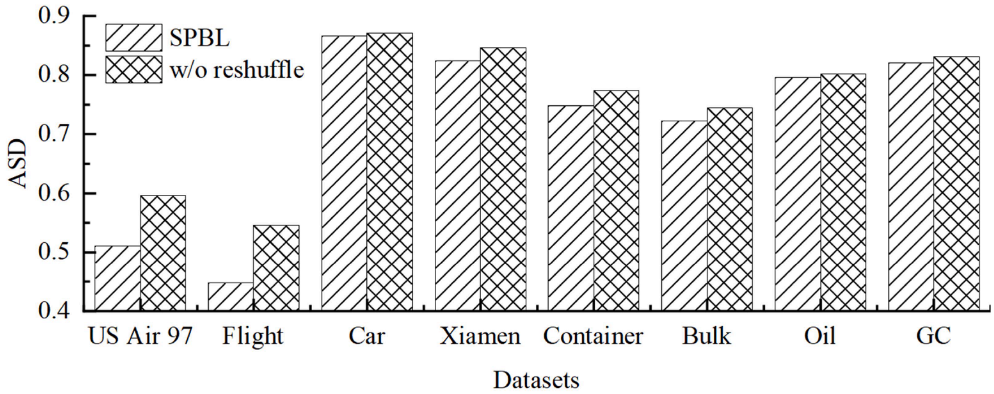

Fig. 9. The contribution of reshuffle edge weights.

compared to the surpassing probability approach's 0.6 s, a reduction of 1.4 s. As dataset complexity escalates, like with the Car dataset, the difference is even more stark: the short cycle approach requires 175.05 s while the surpassing probability approach needs only 25.25 s. This clearly demonstrates that the surpassing probability approach is more computationally efficient.

To evaluate the contribution of reshuffle edge weights in our proposed method, we construct several experiments by systematically removing individual steps. As shown in Fig. 9, the reshuffle step enhanced the value of Aggregate Surpassing Degree (ASD) across all eight datasets, confirming that the reshuffling of edge weights is the dominant contributor to the observed performance gain. The value could also be found in Table 5.

6.1.4. Effectiveness analysis in real-world datasets

To strengthen the objectivity of the assessment, we firstly calculated the modularity value of different method in eight flow-based mobility networks. Since TPSPCD is a local method to identify more refined communities, and modularity is a measure of the overall division effect. It is inappropriate to calculate its modularity. We will explain in detail the practical significance of TPSPCD in the Case study, where we mainly use SPBL for comparison. The results were shown in Table 4. Modularity is basically the same as other methods. The additional comparisons are given in Section 6.2.

As we described in Section 4.3.2, ASD explicitly quantifies the imbalance between internal and external edge strengths. Then, we use ASD to illustrate the effectiveness of SPBL and TPSPCD method. As shown in Table 5. TPSPCD is more effective at detecting communities with stronger internal connections relative to external ones. Table 5 presents a comparison of Aggregate Surpassing Degree (ASD) values across various community detection methods, applied to different networks.

Across all networks, TPSPCD consistently outperforms other methods, yielding the highest ASD values. Furthermore, we present a series of synthetic benchmark networks generated using the Lancichinetti-Fortunato-Radicchi (LFR) model and use NMI to demonstrate the effectiveness of our methods.

6.2. Comparison in simulated datasets

6.2.1. Simulated dataset description

In this section. we present a series of synthetic benchmark networks generated using the Lancichinetti-Fortunato-Radicchi (LFR) model, designed to evaluate community detection algorithms under conditions of increasingly weak community structure. Each network consists of 1000 nodes with power-law distributed degrees (exponent $\tau_1 = 2.0$, minimum degree 10, maximum degree 50) and heterogeneous community sizes following a power law (exponent $\tau_2 = 3$, minimum community size 20, maximum 100). The networks span a high mixing parameter range ($\mu = 0.7$ to 0.95 in 0.05 increments), systematically controlling the fraction of inter-community

Table 4
The modularity of different flow-based mobility networks.

Algorithm	SPBL	LSSCL	Pair	Leiden	Greedy	Louvain	pycombo	LPA
GC	0.58	0.59	0.56	0.59	0.60	0.58	0.60	0.58
Container	0.56	0.57	0.51	0.58	0.59	0.59	0.59	0.51
Bulk	0.41	0.42	0.42	0.43	0.44	0.43	0.44	0.35
Oil	0.58	0.60	0.57	0.61	0.59	0.60	0.61	0.56
US Air97	0.16	0.18	0.14	0.19	0.19	0.20	0.21	0.00
Flight	0.45	0.47	0.00	0.46	0.46	0.47	0.47	0.08
Car	0.68	0.69	0.61	0.68	0.70	0.70	0.70	0.64
Xiamen	0.57	–	0.00	0.59	0.63	0.63	0.63	0.52

Table 5
The comparison of different flow-based mobility networks using ASD. **bold**: best, underline: second best.

Datasets	US Air 97	Flight	Car	Xiamen	Container	Bulk	Oil	GC
LPA	0.003	0.041	0.268	0.078	0.214	0.127	0.177	0.205
pycombo	0.067	0.045	0.225	0.092	0.178	0.130	0.162	0.180
Louvain	0.079	0.047	0.224	0.092	0.178	0.124	0.160	0.179
Greedy	0.078	0.047	0.231	0.095	0.179	0.132	0.153	0.179
Leiden	0.094	0.050	0.263	0.123	0.205	0.140	0.166	0.185
Pair	0.049	0.001	0.152	0.000	0.132	0.130	0.137	0.145
LSSCL	0.116	0.065	0.255	–	0.205	0.141	0.168	0.186
SPBL	**0.132**	0.067	0.269	0.137	0.245	0.162	0.190	0.215
TPSPCD	0.090	**0.097**	**0.504**	**0.365**	**0.287**	**0.176**	**0.246**	**0.277**

edges to simulate challenging scenarios where nodes form stronger connections outside their ground-truth communities than within them. All networks are guaranteed to be connected and incorporate uniformly distributed edge weights (100–10,000) to model real-world interaction strengths. This dataset provides weighted adjacency matrices alongside precise community assignments, enabling rigorous testing of algorithm robustness, overlapping community detection methods, and fundamental limits of community identifiability in noisy regimes while maintaining realistic scale-free properties and weighted interactions characteristic of complex systems.

6.2.2. Comparison and analysis

Fig. 10 illustrates the performance of seven community detection algorithms in terms of the Normalized Mutual Information (NMI) as the Mixing Parameter (μ) varies from 0.70 to 0.99. In general, as the mixing parameter increases, which means the community structure becomes weaker, all algorithms except TPSPCD show a limited performance in terms of NMI. TPSPCD stands out for its stability and relatively high NMI values across the entire range of μ. SPBL and LSSCL have moderate performance with some trends related to μ, while Leiden and Louvain struggle significantly, especially Louvain which has consistently low NMI values.

7. Case study

In this section, we conduct case studies using intercity population flow data. We apply the proposed method to identify community structures within the datasets. The identified communities are then analysed and interpreted, providing further validation of the rationality and interpretability of the proposed approach.

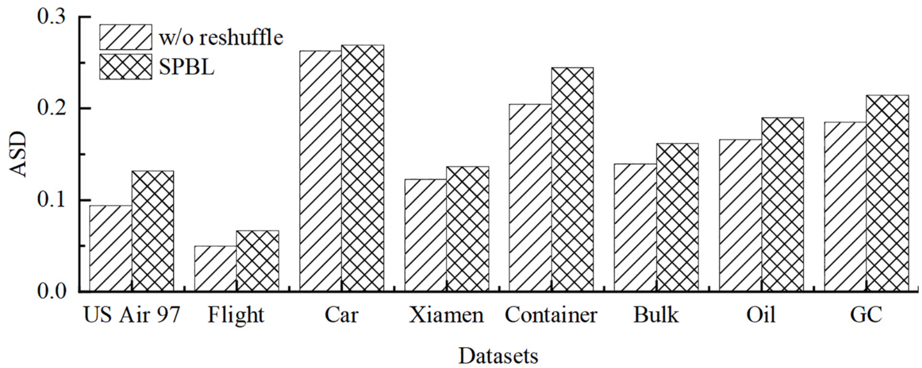

Fig. 10. NMI values of the experimental results obtained by the nine methods.

Table 6
Description of the intercity population mobility flow-based mobility networks.

Name	Node	Edge	Type
Inter-city population flow-based network	366	65,702	Intercity

7.1. Case dataset description

The Baidu migration dataset captures hundreds of millions of population mobility trajectories within China, which serves as a reflection of the dynamic patterns in population movement (Liu et al., 2023; Wei & Wang, 2020; Q. Zhou & Qi, 2023). The description of this dataset is shown in Table 6. Source: https://qianxi.baidu.com/. Time: 2022/01/01–2022/12/31.

7.2. Community detection results

7.2.1. The result of SPBL in intercity mobility networks

The community results are shown in Fig. 11. This increased resolution is particularly evident in regions with dense and complex population flows, such as the Yangtze River Delta, Pearl River Delta, and parts of Central China. Where Louvain tends to merge multiple provinces into a single community, SPBL differentiates sub-regional clusters that better reflect the actual inter-city mobility structure. SPBL is more sensitive to subtle variations in flow intensity and direction, enabling the identification of peripheral or semi-central regions that may be functionally significant but are often masked in coarser methods.

7.2.2. The result of TPSPCD in intercity mobility networks

Based on the TPSPCD results, as shown in Fig. 12, the TPSPCD method effectively identifies urban pairs that exhibit strong integration and high levels of connectivity, reflecting the process of urban agglomeration and regional integration. The detected city pairs represent key areas of "same-city" development, where human mobility and flow intensity are significantly higher compared to other regions.

Then, we explain the advantage of TPSPCD and the high interpretability. As shown in Fig. 12, at the first level of community detection, the proposed method effectively identifies city pairs with a high degree of urban integration, reflecting strong functional and spatial linkages. These city pairs often exhibit intensive population flows, frequent economic exchanges, and high levels of infrastructure connectivity, all of which are indicative of an advanced stage of inter-city integration or ongoing metropolitan coordination.

The second-level communities capture finer regional divisions that reflect local economies, transportation networks, and administrative boundaries. The Yangtze River Delta and Pearl River Delta regions, key economic hubs, are identified as more distinct second-level communities due to their dense intraregional connections. Meanwhile, areas in the northern and western regions show smaller, more isolated second-level communities, consistent with their lower population densities and less developed infrastructure.

7.3. Results validation

In addition to the broader implications of our methodology, the communities identified by our methods exhibit a significant overlap with existing urban agglomeration plans, highlighting the practical utility of our findings, as shown in Fig. 13. The TPSPCD method can identify communities that closely match the predefined urban agglomeration plans is a significant finding. This alignment suggests that our method is not only effective in detecting community structures within flow-based mobility networks but also that it can provide valuable insights that are consistent with current urban planning strategies.

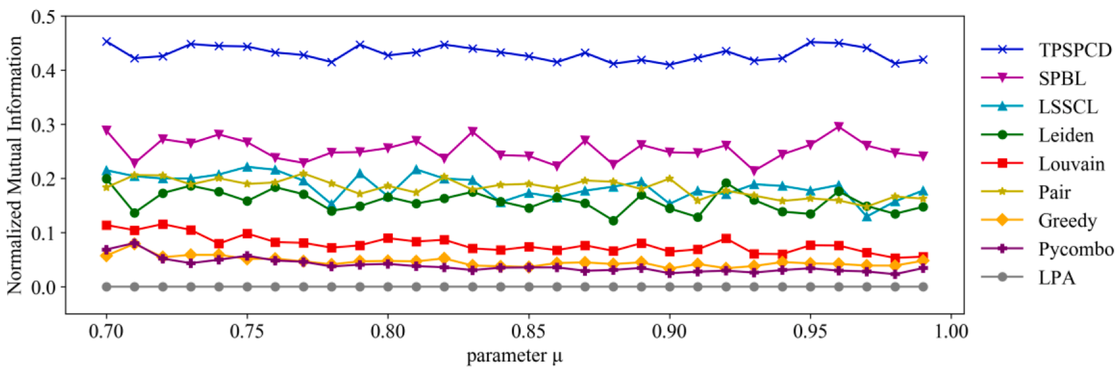

Fig. 11. Results of Louvain and SPBL method. (a) The results of Louvain method. (b) The results of SPBL method.

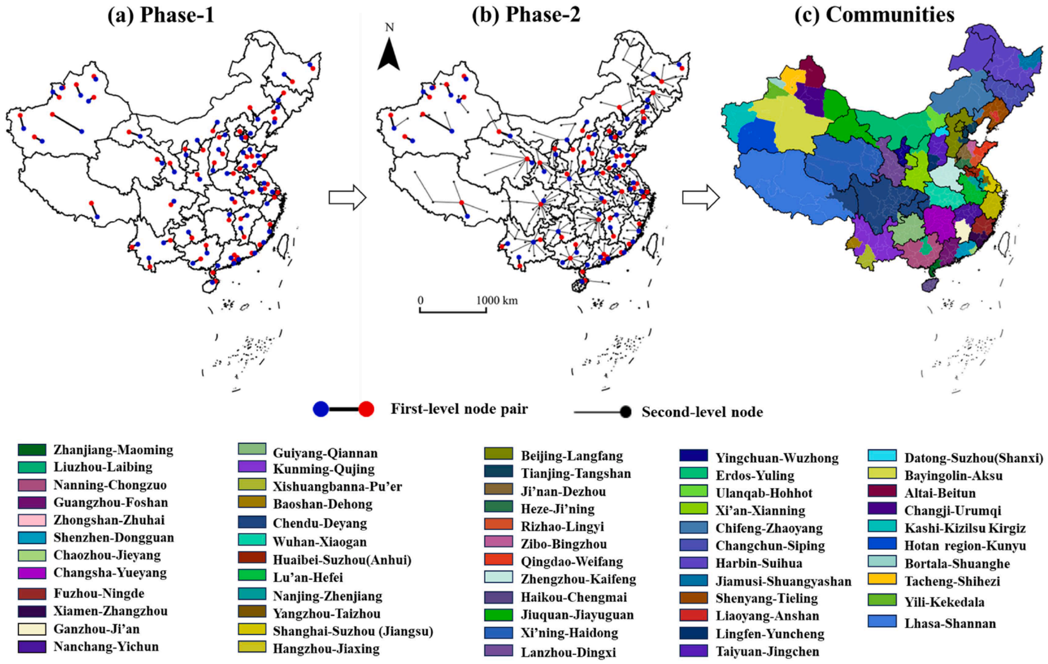

Fig. 12. Application of TPSPCD strategy in intercity mobility network.

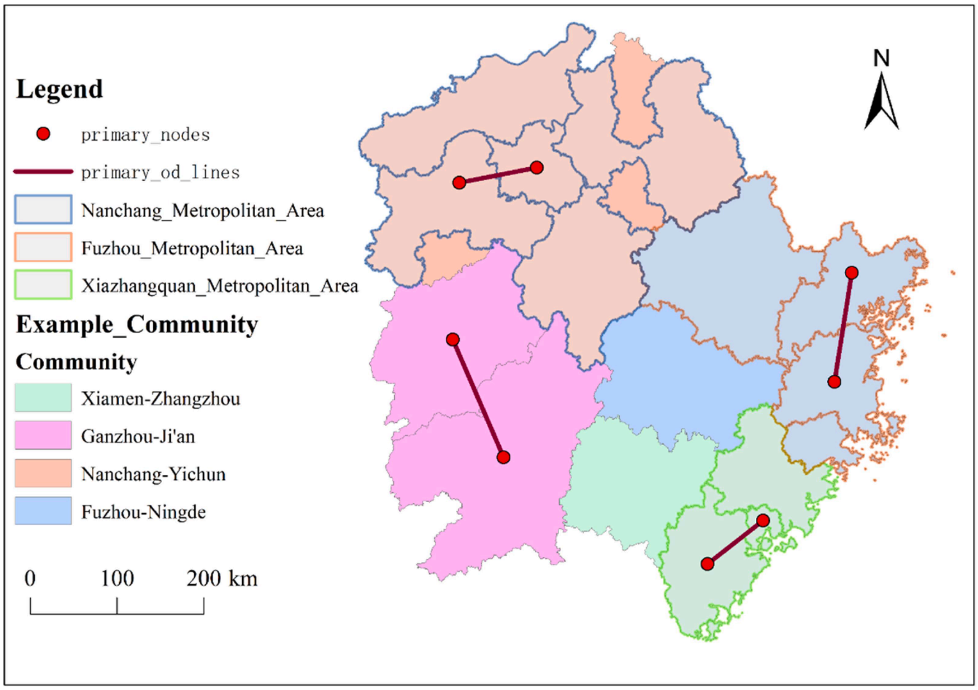

Fig. 13. Community Structures identified by TPSPCD method and the alignment with urban agglomeration plans in China.

Human activities have daily regularity (Yuhui Zhao, Chen, Gao & Zhu, 2023). Similar to the studies (L. Zhang et al., 2023; Yangtianzheng Zhao & Gao, 2024), these identified communities reflect areas where urban centres and surrounding regions show significant levels of interconnectedness, suggesting ongoing processes of regional collaboration and integration. The method was able to capture key spatial relationships between cities, thus contributing to a more nuanced understanding of urban dynamics.

8. Discussion and conclusion

8.1. Discussion

In this study, we model community structures in flow-based mobility networks, using surpassing probabilistic-based approach. Our findings indicate that the proposed workflow effectively identifies community structures within the flow-based mobility network. This highlights the value of incorporating community detection in understanding urban mobility patterns, offering a clearer understanding of urban spatial organization and its potential evolution. To evaluate the rationality of the identified communities, we introduced the Aggregate Surpassing Degree (ASD) to capture the interaction strength within communities. Our methods are validated through experiments on a series of synthetic benchmark networks and eight real-world flow-based mobility networks, demonstrating its effectiveness in accurately identifying community structures among three community detection methods.

Moreover, the case study shows that the eastern coastal areas, including Shanghai, Jiangsu, Zhejiang, and Guangdong, are clustered together, reflecting their high degree of economic integration and dense human mobility networks. The results demonstrate strong intraregional connections driven by industrial collaboration and urbanization. The northern provinces, such as Beijing, Tianjin, Hebei, and Shandong, form a distinct community. This clustering highlights the central role of the Beijing-Tianjin-Hebei metropolitan region, where intercity commuting and coordinated economic activities foster strong regional cohesion. Shandong's inclusion reflects its geographic and economic linkages to the northern cluster, particularly along the Bohai Rim Economic Zone. The Tibetan Plateau, including Tibet and neighbouring regions, is relatively isolated, forming a distinct community with weaker external ties. This reflects the region's unique geographic challenges and limited human mobility compared to other regions.

Looking forward, potential improvements could focus on the merger of communities between core cities, further refining the analysis of central urban hubs and their interactions, which could enhance our understanding of the evolving patterns of urban growth and inter-city integration. The current approach could be further refined by incorporating temporal dimensions into the analysis, allowing for the examination of how community structures evolve over time. This would provide a more comprehensive understanding of dynamic flow networks. Furthermore, the integration of additional data sources and the exploration of multi-layered flow networks could enhance the depth and breadth of community detection in flow-based mobility networks.

9. Conclusion

This study has made contributions in addressing the shortcomings identified in the literature regarding community structure modeling in flow-based mobility networks. The research has successfully developed a novel workflow that effectively identifies community structures within these flow-based networks, overcoming the limitations of traditional methods. The following conclusions can be drawn:

(1) By replacing modularity with the Aggregate Surpassing Degree, the work explicitly incorporates directionality and magnitude of flows into the very definition of community quality.
(2) The Surpassing-Probability-based Leiden (SPBL) algorithm resolves the long-standing small-community identification problem in flow-based mobility networks.
(3) The surpassing probability-based Leiden method (SPBL) and the Two-Phase Surpassing Probability Community Detection (TPSPCD) algorithm not only enhances the granularity of community detection but also improve the accuracy and efficiency of identifying core node pairs and their associated community structures.
(4) The applicability of the method and principles proposed in this paper can be seamlessly extended to various flow network, such as social networks, logistics networks, trade networks, and information networks.

CRediT authorship contribution statement

Yanzhong Yin: Writing – review & editing, Writing – original draft, Visualization, Methodology, Formal analysis, Conceptualization. **Qunyong Wu:** Writing – review & editing, Supervision, Conceptualization, Formal analysis.

Acknowledgments

This work was supported by Project of the Collaborative Innovation Platform of the National Independent Innovation Demonstration Zone of Fuzhou, Xiamen and Quanzhou (2025E3003) and the National Natural Science Foundation of China under Grant No 42201500.

Data availability

Data will be made available on request.

References

Blondel, V. D., Guillaume, J.-L., Lambiotte, R., & Lefebvre, E. (2008). Fast unfolding of communities in large networks. *Journal of Statistical Mechanics: Theory and Experiment, 2008*, Article P10008.

Bonald, T., Charpentier, B., Galland, A., & Hollocou, A. (2018). Hierarchical graph clustering using node pair sampling. *arXiv preprint arXiv:1806.01664*.

Cattaneo, A., Girgin, S., de By, R., McMenomy, T., Nelson, A., & Vaz, S. (2024). Worldwide delineation of multi-tier city–Regions. *Nature Cities, 1*, 469–479.

Clauset, A., Newman, M. E., & Moore, C. (2004). Finding community structure in very large networks. *Physical Review E—Statistical, Nonlinear, and Soft Matter Physics, 70*, Article 066111.

Danila, B. (2016). Comprehensive spectral approach for community structure analysis on complex networks. *Physical Review E, 93*, Article 022301.

Diop, I. M., Cherifi, C., Diallo, C., & Cherifi, H. (2021). Revealing the component structure of the world air transportation network. *Applied Network Science, 6*, 92.

Diop, I. M., Diallo, C., Cherifi, C., & Cherifi, H. (2023). Exploring the regional structure of the worldwide air traffic and route networks. *Complexity, 2023*, Article 3411799.

Gao, Y., Yu, X., & Zhang, H. (2022). Graph clustering using triangle-aware measures in large networks. *Information Sciences, 584*, 618–632.

Hao, H., Yao, E., Yang, Y., Liu, S., Pan, L., & Wang, Y. (2024). How to build vibrant communities by utilizing functional zones? A community-detection-based approach for revealing the association between land use and community vibrancy. *Cities, 155*, Article 105431.

Hong, W., Li, J., He, B., Guo, R., Li, X., Chen, Y., & Wang, Z. (2024). Observing the dynamic community structure of urban travel networks based on navigation data. *International Journal of Geographical Information Science*, 1–23.

Jin, M., Gong, L., Cao, Y., Zhang, P., Gong, Y., & Liu, Y. (2021). Identifying borders of activity spaces and quantifying border effects on intra-urban travel through spatial interaction network. *Computers, Environment and Urban Systems, 87*, Article 101625.

Li, X., Sun, C., & Zia, M. A. (2020). Social influence based community detection in event-based social networks. *Information Processing & Management, 57*, Article 102353.

Liu, H., Wang, J., Liu, J., Ge, Y., Wang, X., Zhang, C., Cleary, E., Ruktanonchai, N. W., Ruktanonchai, C. W., & Yao, Y. (2023). Combined and delayed impacts of epidemics and extreme weather on urban mobility recovery. *Sustainable Cities and Society, 99*, Article 104872.

Liu, T., Pei, T., Fang, Z., Wu, M., Liu, X., Yan, X., Song, C., Jiang, J., Jiang, L., & Chen, J. (2024a). Spatiotemporal mobility network of global scientists, 1970–2020. *International Journal of Geographical Information Science, 38*, 1991–2018.

Liu, W., Cai, H., Xing, H., Hu, S., Tan, Z., & Song, C. (2024b). Alleviating the resolution limit problem in spatial community detection: A local network structure-based method. *International Journal of Geographical Information Science*, 1–23.

Molnár, B., Márton, I.-B., Horvát, S., & Ercsey-Ravasz, M. (2024). Community detection in directed weighted networks using Voronoi partitioning. *Scientific Reports, 14*, 8124.

Newman, M. E., & Girvan, M. (2004). Finding and evaluating community structure in networks. *Physical Review E, 69*, Article 026113.

Ni, Q., Guo, J., Wu, W., & Wang, H. (2023). Influence-based community partition with Sandwich method for social networks. *IEEE Transactions on Computational Social Systems, 10*, 819–830.

Ni, Q., Guo, J., Wu, W., Wang, H., & Wu, J. (2022). Continuous influence-based community partition for social networks. *IEEE Transactions on Network Science and Engineering, 9*, 1187–1197.

Raghavan, U. N., Albert, R., & Kumara, S. (2007). Near linear time algorithm to detect community structures in large-scale networks. *Physical Review E—Statistical, Nonlinear, and Soft Matter Physics, 76*, Article 036106.

Rostami, M., Oussalah, M., Berahmand, K., & Farrahi, V. (2023). Community detection algorithms in healthcare applications: A systematic review. *IEEE Access, 11*, 30247–30272.

Ruan, Y., Liu, S., Tang, J., Guo, Y., & Yu, T. (2025). GLC: A dual-perspective approach for identifying influential nodes in complex networks. *Expert Systems with Applications, 268*, Article 126292.

Sayari, S., Harounabadi, A., & Banirostam, T. (2025). Robust community detection based on improved user interaction, enhanced local path index and pattern mining in social networks. *Information Processing & Management, 62*, Article 104008.

Shi, D., Shang, F., Chen, B., Expert, P., Lü, L., Stanley, H. E., Lambiotte, R., Evans, T. S., & Li, R. (2024). Local dominance unveils clusters in networks. *Communications Physics, 7*, 170.

Sobolevsky, S., Campari, R., Belyi, A., & Ratti, C. (2014). General optimization technique for high-quality community detection in complex networks. *Physical Review E, 90*, Article 012811.

Sperlí, G. (2019). A deep learning based community detection approach. In *Proceedings of the Proceedings of the 34th ACM/SIGAPP Symposium on Applied Computing* (pp. 1107–1110). Limassol, Cyprus: Association for Computing Machinery.

Su, X., Xue, S., Liu, F., Wu, J., Yang, J., Zhou, C., Hu, W., Paris, C., Nepal, S., & Jin, D. (2022). A comprehensive survey on community detection with deep learning. *IEEE Transactions on Neural Networks and Learning Systems, 35*, 4682–4702.

Traag, V. A., Waltman, L., & Van Eck, N. J. (2019). From Louvain to Leiden: Guaranteeing well-connected communities. *Scientific Reports, 9*, 1–12.

Wang, Z., Hu, Y., Li, S., Wang, W., Zhang, L., Zhang, B., Wen, Y., & Meng, D. (2024). DyCom: A dynamic community partitioning technique for system audit logs. In *2024 IEEE 23rd International Conference on Trust, Security and Privacy in Computing and Communications (TrustCom)* (pp. 1486–1493).

Wei, S., & Wang, L. (2020). Examining the population flow network in China and its implications for epidemic control based on Baidu migration data. *Humanities and Social Sciences Communications, 7*, 145.

Xu, Y., Peng, P., Lu, F., & Claramunt, C. (2024). Uncovering the multiplex network of global container shipping: Insights from shipping companies. *Journal of Transport Geography, 120*, Article 103991.

Yin, Y., Wu, Q., & Li, M. (2023). Characterizing intercity mobility patterns for the Greater Bay Area in China. *ISPRS International Journal of Geo-Information, 12*, 5.

Yin, Y., Wu, Q., Lin, H., & Zhao, Z. (2024). Analysis of urban centrality and community patterns from the perspective of "intercity mobility flow" in China. *Journal of Geo-information Science, 26*, 666–678.

Yin, Y., Wu, Q., Zhao, Z., & Chen, X. (2025). Quantifying centrality using a novel flow-based measure: Implications for sustainable urban development. *Computers, Environment and Urban Systems, 116*, Article 102217.

Zhang, J., Liu, X., & Senousi, A. M. (2021). A multilayer mobility network approach to inferring urban structures using shared mobility and taxi data. *Transactions in GIS, 25*, 2840–2865.

Zhang, L., Zuo, X., Wu, Z., Chen, C., Pan, Z., & Hu, X. (2023). The spatial structure and driving mechanisms of multi-source networks in the Chengdu–Chongqing economic circle of China. *ISPRS International Journal of Geo-Information, 12*, 411.

Zhao, Y., Chen, B. Y., Gao, F., & Zhu, X. (2023). Dynamic community detection considering daily rhythms of human mobility. *Travel Behaviour and Society, 31*, 209–222.

Zhao, Y., & Gao, Y. (2024). Spatial patterns and trends of inter-city population mobility in China—Based on Baidu migration big data. *Cities, 151*, Article 105124.

Zhou, M., Yang, M., & Chen, Z. (2023). Flow colocation quotient: Measuring bivariate spatial association for flow data. *Computers, Environment and Urban Systems, 99*, Article 101916.

Zhou, Q., & Qi, Z. (2023). Urban economic resilience and human capital: An exploration of heterogeneity and mechanism in the context of spatial population mobility. *Sustainable Cities and Society, 99*, Article 104983.