

Article

Detecting Dynamic Communities in Vehicle Movements Using Ant Colony Optimization

Qiliang Liu , Sancheng Zhu, Meihua Chen * and Wenkai Liu

School of Geosciences and Info-Physics, Central South University, Changsha 410083, China; qiliang.liu@csu.edu.cn (Q.L.); sancheng.zhu@csu.edu.cn (S.Z.); wenkai.liu@csu.edu.cn (W.L.)

* Correspondence: 215007008@csu.edu.cn

Abstract: Detecting dynamic community structure in vehicle movements is helpful for revealing urban structures and human mobility patterns. Despite the fruitful research outcomes of community detection, the discovery of irregular-shaped and statistically significant dynamic communities in vehicle movements is still challenging. To overcome this challenge, we developed an evolutionary ant colony optimization (EACO) method for detecting dynamic communities in vehicle movements. Firstly, a weighted, spatially embedded graph was constructed at each time snapshot. Then, an ant-colony-optimization-based spatial scan statistic was upgraded to identify statistically significant communities at each snapshot by considering the effects of the communities discovered at the previous snapshot. Finally, different rules defined based on the Jaccard coefficient were used to identify the evolution of the communities. Experimental results on both simulated and real-world vehicle movement datasets showed that EACO performs better than three representative dynamic community detection methods: FacetNet (a framework for analyzing communities and evolutions in dynamic networks), DYNMOGA (dynamic multi-objective genetic algorithm), and RWLA (random-walk-based Leiden algorithm). The dynamic communities identified by EACO may be useful for understanding the dynamic organization of urban structures.

Keywords: dynamic community; ant colony optimization; evolutionary clustering; significance test; vehicle movements



Citation: Liu, Q.; Zhu, S.; Chen, M.; Liu, W. Detecting Dynamic Communities in Vehicle Movements Using Ant Colony Optimization. *Appl. Sci.* **2022**, *12*, 7608. <https://doi.org/10.3390/app12157608>

Academic Editor: Yosoon Choi

Received: 29 June 2022

Accepted: 27 July 2022

Published: 28 July 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

In the era of big data, vast amounts of vehicle movement data (e.g., taxi Global Positioning System (GPS) trajectories) provide new opportunities to understand spatiotemporal interactions between different places in a city [1]. The vehicle movements can be transformed into a weighted, spatially embedded network, where the places (e.g., road segments) in a city are regarded as nodes, and the origin and destination (OD) pairs between the places can be considered as weighted edges [2]. The spatially embedded network usually evolves with time, and can be regarded as a kind of dynamic graph. In this dynamic graph, there are usually some subgraphs or dynamic communities, where the movements (e.g., OD pairs) within each community are significantly more than those between that community and other communities [3]. The dynamic communities discovered in vehicle movements play a crucial role in understanding the dynamic organization of urban structures, which can facilitate urban planning and management [4]. Identifying dynamic communities is also useful for modelling traffic correlation patterns in a city, which are helpful for traffic forecasting and traffic control [5].

Detecting dynamic communities in vehicle movements has been paid more attentions in recent years [4]. Existing studies usually divide the time dimension into a series of snapshots. Therefore, a dynamic graph can be constructed. Three classes of dynamic community detection methods are currently available, i.e., instant optimal, temporal trade-off, and cross-time methods [6].

Instant optimal methods first identify communities at each snapshot using existing static community detection methods (e.g., modularity-based methods [7,8] and information-theoretic methods [9,10]), and then match the communities discovered at snapshot t with those discovered at snapshot $t - 1$. Different strategies have been developed to match the communities discovered at continuous snapshots. For example, the iterative similarity-based methods [11] use a quality function (e.g., the Jaccard coefficient [12]) to measure the similarity between the communities at different snapshots, and consider the communities with high similarity in adjacent timesteps as part of the same dynamic community. The iterative core-nodes-based methods [13] define one or several representative node(s) for each community (e.g., the node with the highest centrality metric value), and identify communities containing the same representative node(s) at adjacent snapshots as the same dynamic community. The multistep matching methods [14] not only match the communities between adjacent snapshots, but also match the communities across several snapshots. Instant optimal methods are easy to implement. However, the instability of existing community detection methods (i.e., the same methods usually discover different communities from the same network) may lead to spurious dynamic communities [6]. In addition, the sizes and geometric shapes of communities in vehicle movements are usually varied [5]; however, existing community detection methods usually assume that the shapes of the communities are approximately circular [15].

To alleviate the instability problem of instant optimal methods, temporal trade-off methods have been developed. These methods assume that communities at snapshot t not only depend on the network at snapshot t , but also depend on the communities identified at snapshot $t - 1$. By using the previously identified communities, temporal trade-off methods can identify more stable dynamic communities [6]. Temporal trade-off methods involve two iterative steps: initialization and update. Firstly, the communities are discovered at the initial snapshot t , and then, for the next snapshot $t + 1$, the communities are detected using the network at $t + 1$ and the communities at t . Different update strategies have been developed. For example, the global optimization update strategy [16] first defines a global objective function (e.g., modularity), and then uses the communities found at time $t - 1$ as seeds to optimize the communities at t . The rule-based update strategy [17] defines a series of rules (e.g., addition/deletion of nodes/edges) to determine how the communities evolve when the network changes. The evolutionary clustering strategy [18–20] attempts to balance both partition quality and temporal partition coherence at each snapshot. Temporal trade-off methods do not require a complete knowledge of the network history; therefore, they are useful for real-time community detection [6]. However, without spatial contiguity constraints, existing temporal trade-off methods are usually not robust [21]. In addition, the statistical significance of the identified communities at each snapshot cannot be ensured; therefore, these methods are very likely to identify some spurious communities [22].

Unlike the instant optimal and temporal trade-off methods, cross-time methods aim to find dynamic communities from all snapshots. These kinds of methods are suitable for detecting communities that are coherent over the long-term, but are not suitable for identifying real-time dynamic communities [6]. Cross-time methods first transform networks at all snapshots into a single network, and then identify dynamic communities from the single network using existing static community detection methods. There are two kinds of edges in this single network, i.e., the relationships between nodes at the same snapshot, and the relationships between nodes belonging to adjacent snapshots. Dynamic communities can be discovered using different constraints, i.e., (1) fixed memberships and fixed properties [23], where communities remain the same throughout the studied period; (2) fixed memberships and evolving properties [24], where communities can be non-homogeneous over time, e.g., the interactions among nodes in a community may increase or decrease over time; (3) evolving memberships and fixed properties [25], where node memberships can be varied over time; and (4) evolving memberships and evolving properties [4,26], where during the studied period, communities can appear or disappear, node memberships can be varied, and the densities of communities can also change.

In this study, we aim to identify real-time dynamic communities from vehicle movements. Although existing instant optimal methods and temporal trade-off methods can be used to detect real-time dynamic communities, the discovery of irregular-shaped and statistically significant dynamic communities in vehicle movements is still challenging. To overcome this challenge, we propose an evolutionary ant colony optimization method (EACO) for detecting dynamic communities in vehicle movements. The main contributions of this study include the following three aspects:

- (1) We propose a new pheromone initialization strategy to upgrade an ant-colony-optimization-based spatial scan statistic (ACOScan) [15] to identify dynamic communities in vehicle movements. The proposed method can not only balance both partition quality and temporal partition coherence at each snapshot, but also identify irregular-shaped communities.
- (2) We construct a significance test to evaluate the statistical significance of the identified dynamic communities. Therefore, some weakly interacting or spurious communities can be eliminated.
- (3) The proposed method can be regarded as a spatial extension of evolutionary clustering. Compared with existing spatial community detection methods (e.g., ACOScan), the proposed method is useful for understanding spatiotemporal interactions in a city.

The effectiveness and superiority of EACO were verified using both simulated and Beijing taxi trajectories data. The dynamic communities identified by EACO are of great value for understanding the evolution of urban spatial structure. The rest of this paper is organized as follows: Section 2 describes the proposed method in detail. Section 3 presents the experimental evaluations using both simulated and Beijing taxi trajectories data. Section 4 discusses the experimental results. Section 5 concludes this study and outlines future work.

2. Methods

The EACO framework is displayed in Figure 1. First, the OD points of vehicle movements were mapped onto road network road segments using a nearest-neighbor matching method [27]. Second, at each snapshot, a spatially embedded graph $G = (V, E, W)$ was constructed based on road segments and OD pairs, where V is the set of vertices (each road segment is a vertex), E is the set of edges (two road segments connected by at least one OD pair form an edge), and W is the set of weights (the weight of an edge is the number of OD pairs between two vertices). Third, ACOScan was upgraded to identify statistically significant spatial communities at each snapshot. Finally, different rules defined based on the Jaccard coefficient were employed to find the evolution of the significant spatial communities. The main notations used in this study are listed in Table 1.

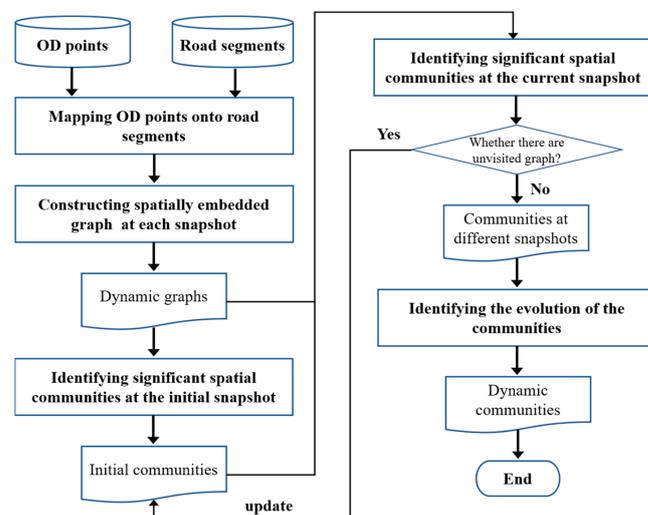


Figure 1. Framework of the proposed method.

Table 1. Main notations used in this study.

Notations	Meaning
V	The set of vertices
E	The set of edges
W	The set of weights
G	A spatially embedded graph
S	The strength sequence of a graph
S_i	The strength of a vertex v_i
Z	A spatial community in G
$S(Z)$	The sum of strengths of Z
V_Z	The set of vertices in Z
p	The probability of the weights of edges in Z
q	The probability of the weights of edges outside Z
$LR(Z)$	The likelihood ratio statistic for a spatial community Z
$L(Z, p, q)$	The likelihood function under the alternative hypothesis
L_0	The likelihood function under the null hypothesis
$W(Z)$	The sum of weights in Z
$W(G)$	The sum of weights in G
$\mu(Z)$	The expected sum of weights of Z under the null hypothesis
$\mu(G)$	The expected sum of weights of graph G under the null hypothesis
W_i	The weight of an edge r_i
N_i	The neighborhood of a vertex v_i
G'	A random graph
p_Z	The p -value of community Z
N_{rep}	The number of random graphs
i_k	An indicator variable in the k th simulation
$LR_{ran}(Z)$	The likelihood ratio value of Z calculated in the random graph
α	The significance level for identifying spatial communities
β	The significance level for identifying candidate road segments
L_N	The likelihood ratio value of C_n at the initial snapshot
L_S	The sum of the likelihood ratio values of all communities at the initial snapshot
Tr	The initial pheromone of each edge
M_{ip}	The pheromone of r_i determined based on the weight of r_i at the current snapshot
γ	The balance coefficient
L	The maximum size of a spatial community
N_{ant}	The number of ants
N_e	The number of elite ants
N_{ite}	The maximum number of iterations
M_{pe}	The pheromone evaporation coefficient
C_t	The community identified at snapshot t

2.1. Spatial Scan Statistic for Spatially Embedded Graphs

In this study, a spatial scan statistic [15] was used for the quantitative assessment of the spatial communities. We denote the strength sequence of a graph G as $S = \{S_1, S_2, \dots, S_N\}$, where the strength S_i of a vertex v_i is defined as the sum of the weights of the edges connected to v_i . For a spatial community $Z \subset G$, the sum of strengths of Z can be defined as $S(Z) = \sum_{v_i \in V_Z} S_i$. If there is a spatial community $Z \subset G$, we assume that the weights of edges in Z and outside Z are generated with probability p and q , respectively. To ensure that Z is a significant spatial community, it needs to prove that the null hypothesis ($H_0: p = q$) is false and the alternative hypothesis ($H_1: p > q$) is true. The likelihood ratio statistic for a spatial community Z can be defined as follows:

$$LR(Z) = \frac{L(Z, p, q)}{L_0} = \begin{cases} \left(\frac{W(Z)}{\mu(Z)}\right)^{W(Z)} \left(\frac{W(G)-W(Z)}{\mu(G)-\mu(Z)}\right)^{W(G)-W(Z)} & \text{if } \frac{W(Z)}{\mu(Z)} > \frac{W(G)-W(Z)}{\mu(G)-\mu(Z)} \\ 1 & \text{otherwise} \end{cases} \quad (1)$$

where $L(\mathbf{Z}, p, q)$ is the likelihood function under H_1 , L_0 is the likelihood function under H_0 , $W(\mathbf{Z})$ is the sum of weights in \mathbf{Z} , $W(\mathbf{G})$ is the sum of weights in \mathbf{G} , $\mu(\mathbf{Z}) = S^2(\mathbf{Z})/4W(\mathbf{G})$ is the expected sum of weights of \mathbf{Z} under H_0 , $\mu(\mathbf{G})$ is the expected sum of weights of graph \mathbf{G} under H_0 , and in this study, $\mu(\mathbf{G}) = W(\mathbf{G})$. Detailed derivation of the likelihood ratio statistic can be found in [15].

Under the null hypothesis, for a spatial community $\mathbf{Z} \subset \mathbf{G}$, the probability that the sum of observed weights of \mathbf{Z} is equal to or greater than W_i is as follows:

$$P(W(\mathbf{Z}) \geq W_i) = 1 - \sum_{j=0}^{W_i-1} \frac{\lambda_{i0}^j e^{-\lambda_{i0}}}{j!} \quad (2)$$

$$\lambda_{i0} = \frac{W(\mathbf{G})}{\mu(\mathbf{G})} \cdot \mu(\mathbf{Z}) \quad (3)$$

To evaluate the significance of the spatial communities, we constructed a random graph with a given sequence of expected strengths [28]. To construct a random graph \mathbf{G}' for \mathbf{G} , an inhomogeneous Poisson process PI (PI has the same number of vertices and the same sequence of expected strengths as \mathbf{G}) was used to generate the weights of the edges in \mathbf{G}' . The probability that two vertices form an edge is determined by PI , proportional to the strength of the vertices of that edge [29]. The p -value of each community can be calculated as follows:

$$p_{\mathbf{Z}} = \frac{\sum_{k=1}^{N_{\text{rep}}} I_k}{N_{\text{rep}}} \quad (4)$$

where N_{rep} is the number of random graphs. I is an indicator variable; in the k th simulation, if $LR_{\text{ran}}(\mathbf{Z}) > LR(\mathbf{Z})$, then $I_k = 1$; otherwise $I_k = 0$ ($LR_{\text{ran}}(\mathbf{Z})$ is the likelihood ratio value of \mathbf{Z} calculated in the random graph). The false discovery rate method [30] was used to deal with the multiple hypothesis testing problems in detecting the spatial communities. If the significance level α was set to 0.05, the p -values of m spatial communities were first sorted in an ascending order ($p_1 \leq p_2 \leq \dots \leq p_m$). Then, starting from p_m , the first p_i that satisfies $p_i \leq (i/m)\alpha$ was identified. Finally, all of the spatial communities whose p -values were less than p_i were identified as the significant spatial communities.

2.2. Evolutionary Ant Colony Optimization Method

In this study, we aimed to detect communities at each snapshot based on the idea of evolutionary clustering. Evolutionary clustering requires that the communities on each timestamp should remain faithful to the current data as much as possible, and should be similar to the communities discovered in the previous snapshot [18]. To balance both partition quality and temporal partition coherence at each snapshot, we developed an evolutionary ant colony optimization method based on ACOScan.

First, the statistically significant spatial communities at the initial snapshot were identified using ACOScan. Ant colony optimization [31] is a typical swarm intelligence method inspired by the foraging behavior of ants; some ants leave chemicals (called pheromones) on the ground to mark paths that are favorable for other colony members to follow [32]. Connecting road segments to form communities can be regarded as the identification of the optimal path found by ants. In ACOScan, the spatial scan statistic described in Section 2.1 was used as the objective function for community detection. To narrow the search space and increase the stability for community detection, we first identified significant road segments as candidate road segments. For each road segment, we calculated $P(W(\mathbf{Z}) \geq W(\mathbf{N}_i))$ using Equations (2) and (3); if the value was less than the significance level β ($\beta = 0.1$), then that road segment was identified as a candidate road segment. The edge connecting two candidate road segments was identified as a candidate edge. The candidate edges were used as basic units for detecting communities. ACO was used as the heuristic method for searching communities.

Second, the communities identified at the initial snapshot were considered to detect communities at the current snapshot. ACO and spatial scan statistic were also used as the heuristic method and the objective function for community detection, respectively. Pheromones play a key role in searching communities using ACO [31,33]. Therefore, we proposed two pheromone reward strategies to consider the effect of communities identified at the initial snapshot on those identified at the current snapshot:

- (1) For each edge in community C_N identified at the initial snapshot, we added an additional pheromone value L_n/L_s at the current snapshot, where L_n is the likelihood ratio value of C_n at the initial snapshot, and L_s is the sum of the likelihood ratio values of all communities at the initial snapshot.
- (2) When we determined whether two edges r_i and r_j should be combined at the current snapshot, a connectivity indicator $\delta(r_i, r_j)$ was defined. If r_i and r_j were in the same community at the initial snapshot, $\delta(r_i, r_j) = 1$; otherwise, $\delta(r_i, r_j) = 0$.

Based on the above two strategies, the initial pheromone (Tr) of each edge at the current snapshot can be defined as follows:

$$Tr = M_{ip} + \frac{L_n}{L_s} + \gamma \times \delta(r_i, r_j) \quad (5)$$

where M_{ip} is the pheromone of each edge determined based on the weight of that edge at the current snapshot, and γ is a balance coefficient. The first term of Equation (5) was used to consider the quality of communities identified at the current snapshot. The last two terms of Equation (5) were used to consider the temporal partition coherence between two consecutive snapshots.

To search communities at the current snapshot, a candidate edge was randomly selected as the starting edge for each ant, and the roulette-wheel selection method [34] was used to select the spatially adjacent candidate edges according to Tr . Each ant stopped walking until the number of candidate edges in the community reached L (L is the maximum size of a spatial community). The paths of each ant formed a community. When all ants had completed the search, the likelihood ratio values of the formed N_{ant} (i.e., the number of ants) communities were sorted in descending order $LR = [LR_1, LR_2, \dots, LR_{N_{ant}}]$. The ants with the top N_e likelihood ratio values in LR were regarded as elite ants. Pheromone volatilization was performed on the candidate edges, and the pheromones on the edges in communities formed by the n th elite ant increased by $2(N_e - n)$ [35]. After N_{ite} iterations, the community with the highest likelihood ratio value was identified as the most likely community. The most likely community was removed from the candidate road segments. The above procedure was iteratively implemented until all of the candidate road segments were grouped into certain communities. The significance of the identified communities was evaluated using the method introduced in Section 2.1. Communities identified at the current snapshot were regarded as the initial communities for detecting communities at the next snapshot. The pseudocode of EACO is described in Algorithm 1.

Algorithm 1: EACOScan (Graphs, α , β , M , parameters of ACO)

```

StaticCommunities= Empty();
InitialCommunities = ACOScan(Graphs.Get(1),  $\beta$ ); //Identify communities at the initial snapshot
StaticCommunities.Add(InitialCommunities);
InitialCommunities = FindSignificantCommunities(PreviousCommunities,  $\alpha$ ,  $M$ ); //Evaluate the
significance of communities
for i = 1:t do
Communities= ACO (InitialCommunities, Graphs.Get(i)); //Identify communities at current snapshot
Communities = FindSignificantCommunities(Communities,  $\alpha$ ,  $M$ );
InitialCommunities = Communities;
StaticCommunities.Add(Communities);
end for
DynamicCommunities = Match(StaticCommunities); //Identify the evolution process of communities

```

The parameters of ACO were set according to the suggestions of [32,35]. The M_{ip} of each edge was set to the normalized value of the weight of that edge. The maximum size (L) of a community was generated by the Gaussian random function $N(\mu, \sigma)$. The initial values of μ and σ were set as the mean and half-standard variance of 20 integers randomly generated from $[1, N_{ant}]$, respectively. After each iteration, μ and σ were updated; μ and σ were set to the mean and half-standard variance of the sizes of the communities identified by elite ants, respectively. Other parameters are listed in Table 2.

Table 2. Explanation of the parameters and their values used for ant colony optimization.

Parameter	Name	Value
Number of ants	N_{ant}	Number of candidate edges
Number of elite ants	N_e	20
Maximum number of iterations	N_{ite}	500
Pheromone evaporation coefficient	M_{pe}	0.1
Significance level for identifying spatial communities	α	0.05
Significance level for identifying candidate road segments	β	0.1
Number of random graphs	N_{rep}	99

2.3. Identification of Dynamic Communities

After the communities at each snapshot were obtained, the evolution of communities needed to be further identified. The similarity between two communities identified at continuous snapshots was measured using the Jaccard coefficient:

$$A(C, C') = \frac{|C_t \cap C'_{t+1}|}{|C_t \cup C'_{t+1}|} \quad (6)$$

where C_t and C_{t+1} represent two communities identified at snapshots t and $t + 1$, respectively, $|C_t \cap C_{t+1}|$ is the number of common nodes in C_t and C_{t+1} , and $|C_t \cup C_{t+1}|$ is the number of nodes in the union of C_t and C_{t+1} .

There are usually seven kinds of community evolution processes [12]: continuation, splitting, merging, growth, contraction, birth, and death (Figure 2). In this study, we defined different rules to identify these community evolution processes.

- (1) Continuation means that a community remains almost unchanged at the next time snapshot. If the similarity between two communities identified at continuous snapshots is greater than 0.9, then the evolution process is identified as continuation.
- (2) Splitting means that a community is split into multiple communities at the next time snapshot. If the sum of the similarity between multiple communities at $t + 1$ and a community at t is greater than 0.9, then the evolution process is identified as splitting.
- (3) Merging indicates that multiple communities merge into one community at the next time snapshot. If the sum of the similarity between a community at t and multiple communities at $t - 1$ is greater than 0.9, then the evolution process is identified as merging.
- (4) Growth indicates that the size of a community increases obviously at the next time snapshot. If the similarity between a community at $t - 1$ (C_{t-1}) and a community at t (C_t) is greater than 0.75, and the size of C_t is larger than that of C_{t-1} , then the evolution process is identified as growth.
- (5) Contraction means that the size of a community decreases obviously at the next time snapshot. If the similarity between C_t and C_{t+1} is greater than 0.75, and the size of C_{t+1} is smaller than that of C_t , then the evolution process is identified as contraction.
- (6) Birth occurs when a new community emerges without a predecessor. The unmatched communities at the next snapshot are identified as birth communities.
- (7) Death occurs when a community disappears without a successor. The unmatched communities at the previous snapshot are identified as death communities.

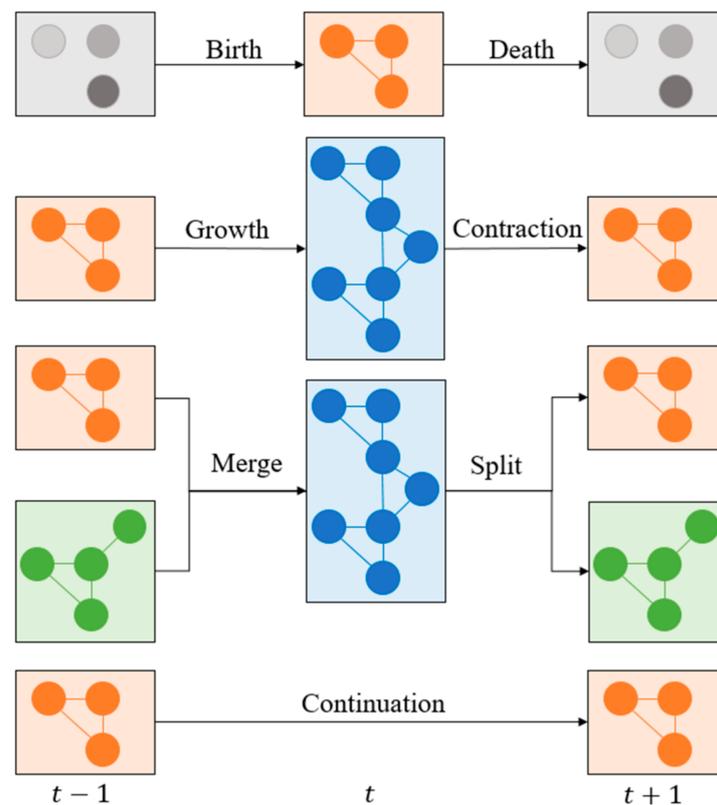


Figure 2. Seven kinds of community evolution processes.

3. Experiments

The performance of EACO was evaluated using both simulated and real-world vehicle movement datasets. EACO was also compared with three representative dynamic community detection methods: FacetNet [19], DYNMOGA [20], and RWLA [4]. FacetNet uses modularity as the objective function, and uses the expectation-maximization method to identify communities. EACO and FacetNet both require a coefficient $\gamma \in (0, 1)$ to balance partition quality and temporal partition coherence at each snapshot. DYNMOGA uses modularity to evaluate the quality of communities identified at the current snapshot, and uses normalized mutual information (NMI) [36] to measure the similarity between the community structure identified at the current snapshot and that identified at the previous one. The parameters of DYNMOGA were set according to the original paper: number of generations = 200, population size = 200, crossover rate = 0.8, and mutation rate = 0.2. RWLA first uses a random walk method [37] to calculate the visiting probabilities of network nodes and layers, then merges the graphs at different time snapshots into a single graph according to the visiting probabilities, and finally uses the Leiden algorithm [38] to identify dynamic communities from the merged graph. Two damping factors α and β were both set to 0.85 according to the suggestion of the original paper. The codes of FacetNet, DYNMOGA, and RWLA were obtained from the authors of the original papers.

3.1. Simulation Experiments

In total, 2378 road segments were selected from the real road network in Beijing to construct the simulated dataset. We generated simulated vehicle movement datasets at three snapshots (Figure 3). At each snapshot, the communities were generated using the procedure developed by Liu et al. [15]. At the first snapshot, we randomly combined spatially adjacent road segments to construct the ranges of m spatial communities (the size of each community $n_i \in [50, 200]$). Then, for each spatial community with n_i road segments, we randomly generated $10n_i$ OD pairs in that community. Finally, we randomly set $\sum 2n_i$ OD pairs as noise in the study area. The seven kinds of community evolution patterns

were predefined in the simulated datasets. From Figure 3, one can see that there were communities with different shapes and sizes in the simulated datasets.



Figure 3. Simulated datasets: (a) road network; (b) 18 spatial communities at snapshot 1; (c) 11 spatial communities at snapshot 2; (d) 13 spatial communities at snapshot 3.

The NMI index was used to quantitatively evaluate how well the communities identified by the two methods matched the predefined communities. The NMI value can be between 0 and 1 (with 1 representing a perfect match). The balance coefficient γ was set from 0.1 to 0.9, with a step size of 0.1. The NMI values of the experimental results obtained by the two methods are listed in Table 3. One can see that EACO performs better than FacetNet. EACO performs best when $\gamma = 0.8$ (Figure 4). FacetNet performs best when $\gamma = 0.9$ (Figure 5). DYNMOGA can automatically determine the balance coefficient. The experimental results of DYNMOGA are displayed in Figure 6. The NMI values of the communities identified at the three snapshots were 0.51, 0.42, and 0.43, respectively. Figure 7 displays the communities identified by RWLA. The NMI values of the communities identified at the three snapshots were 0.64, 0.52, and 0.53, respectively.

Table 3. NMI values of the experimental results obtained by FacetNet and EACO.

Method	Snapshot	$\gamma = 0.1$	$\gamma = 0.2$	$\gamma = 0.3$	$\gamma = 0.4$	$\gamma = 0.5$	$\gamma = 0.6$	$\gamma = 0.7$	$\gamma = 0.8$	$\gamma = 0.9$
FacetNet [19]	1	0.52	0.58	0.60	0.63	0.65	0.66	0.69	0.71	0.73
	2	0.50	0.55	0.57	0.61	0.63	0.64	0.67	0.69	0.71
	3	0.49	0.51	0.53	0.57	0.61	0.63	0.64	0.68	0.70
EACO	1	0.96	0.95	0.97	0.96	0.98	0.97	0.96	0.97	0.97
	2	0.94	0.94	0.94	0.95	0.95	0.96	0.96	0.97	0.96
	3	0.95	0.95	0.96	0.96	0.96	0.97	0.97	0.98	0.98

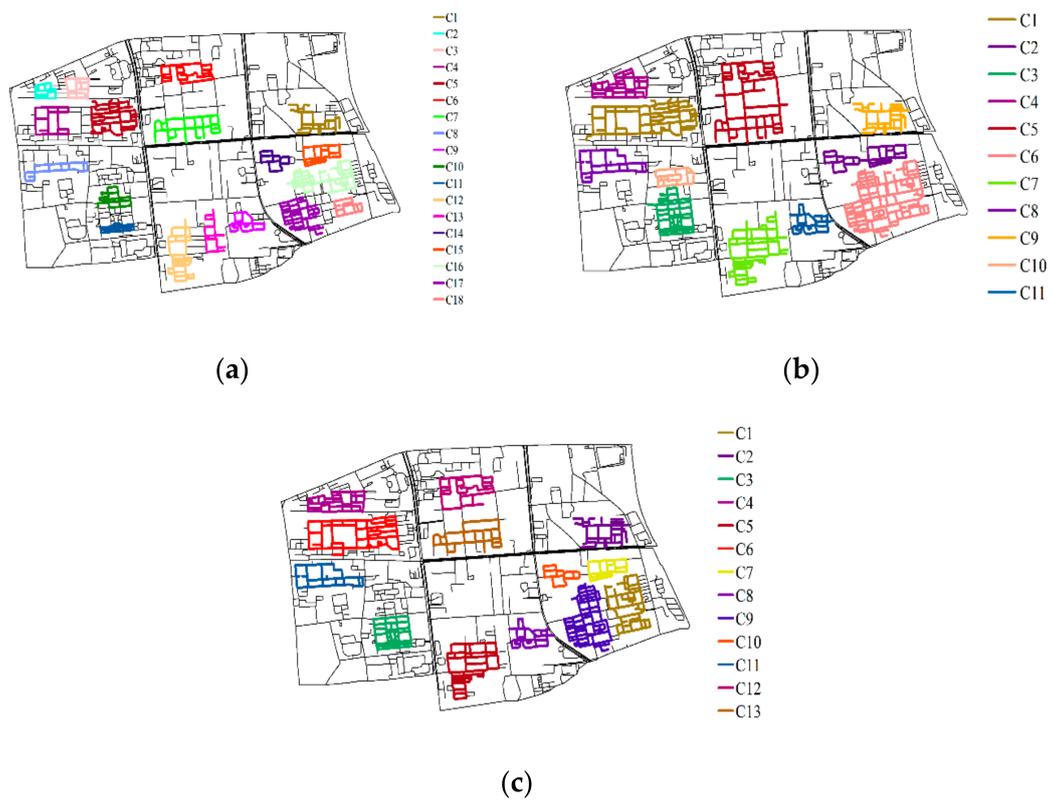


Figure 4. Communities identified by EACO: (a) snapshot 1; (b) snapshot 2; (c) snapshot 3.

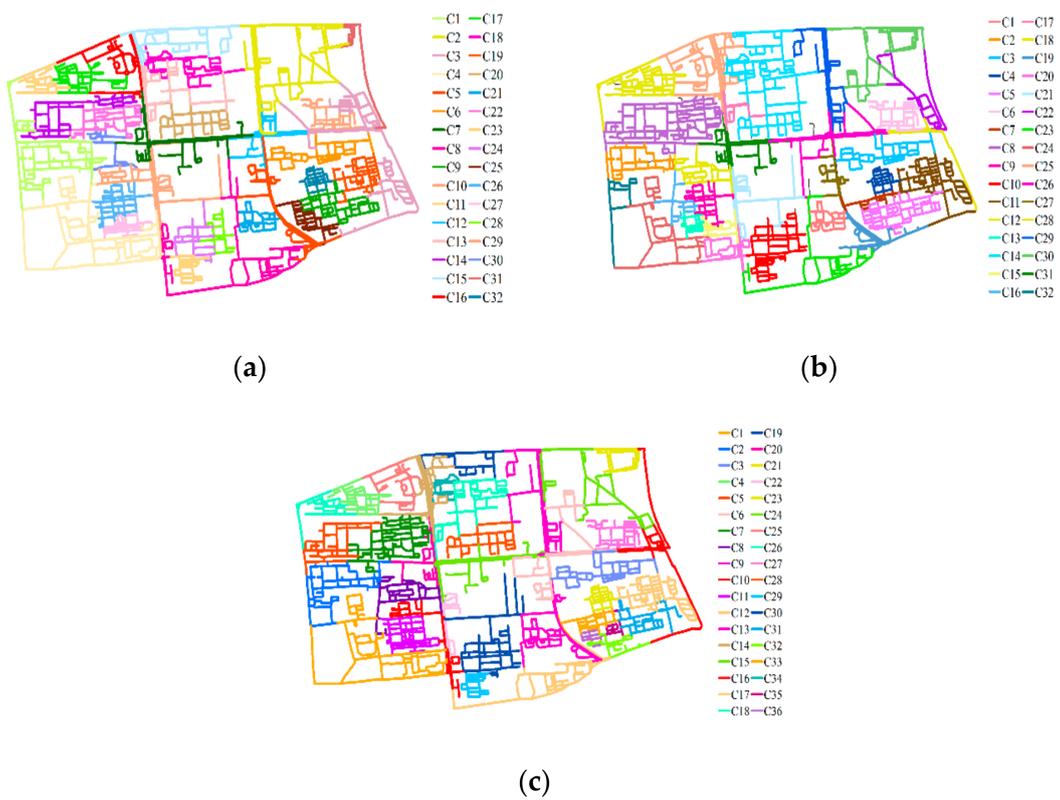


Figure 5. Communities identified by FacetNet: (a) snapshot 1; (b) snapshot 2; (c) snapshot 3.

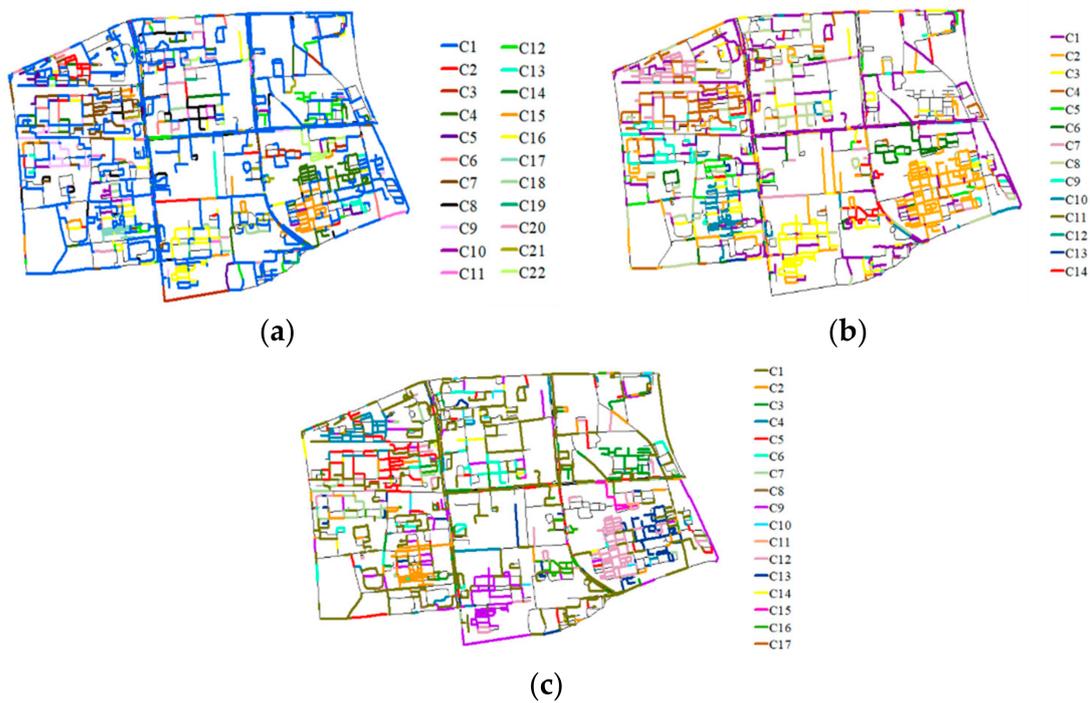


Figure 6. Communities identified by DYNMOGA: (a) snapshot 1; (b) snapshot 2; (c) snapshot 3.



Figure 7. Communities identified by RWLA: (a) snapshot 1; (b) snapshot 2; (c) snapshot 3.

From Figures 5–7, one can see that some predefined communities cannot be accurately identified by the three composition methods. Most of the predefined communities were wrongly segmented. DYNMOGA and RWLA identified some spatially dispersed communities, which were spurious communities formed by noise. In contrast, the predefined communities at different snapshots were well discovered by EACO. In Figure 8, the evolution of the communities identified by EACO is displayed. All seven kinds of community evolution patterns were well revealed by EACO.

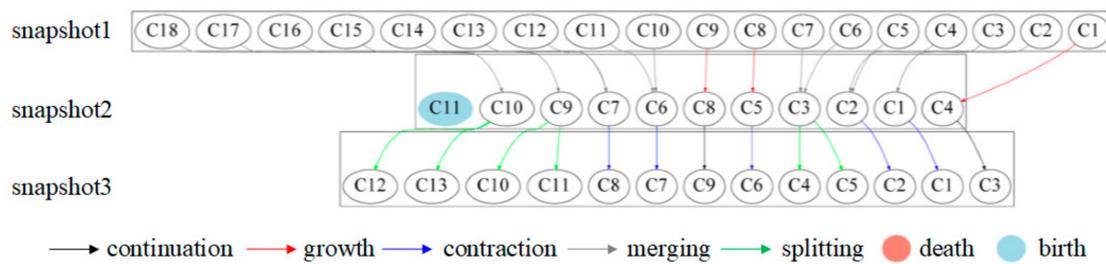


Figure 8. The evolution of the communities identified by EACO in the simulated dataset.

3.2. Case Study

To evaluate the practical value of the proposed method, we applied EACO to detect dynamic communities in the Beijing taxi trajectory dataset. Identifying dynamic communities in human mobility data is useful for understanding the dynamic process of the spatial interactions between different places in a city [4]. These dynamic communities enable data-centric urban planning [39]. The study area of the case study is located within the fifth ring road of Beijing. The selected road network contains 10,919 road segments. 363,323 pairs of OD points were extracted from the trajectories of about 30,000 taxis on 9 May 2016 (from 8:00 to 24:00). For each hour, we constructed a spatially embedded graph. Therefore, a 16-layer snapshot network was obtained. The number of OD pairs per hour is displayed in Figure 9.

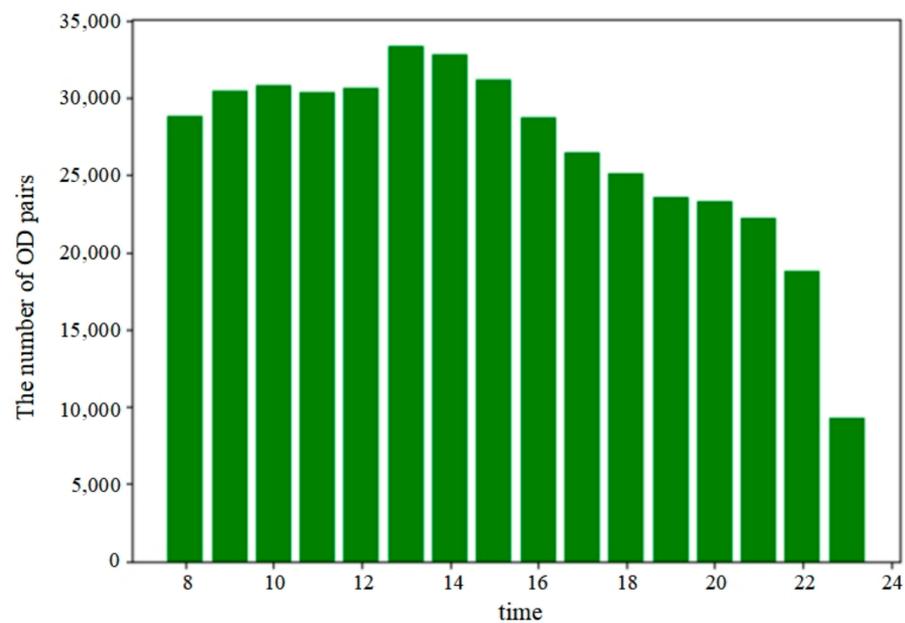


Figure 9. The number of OD pairs per hour.

EACO, FacetNet, DYNMOGA, and RWLA were all applied to the 16-layer snapshot network. Based on the simulation experiments, the balance coefficient for EACO and FacetNet was set to 0.8 and 0.9, respectively. The communities at different snapshots identified by the four methods are displayed in Figures 10–13, respectively. Due to space limitations, we only show the communities identified during morning and evening rush hours, because the communities in these time periods changed more drastically. We found that that EACO can identify communities of irregular shapes and different sizes. The communities identified by EACO were mainly within the fourth ring road, because the interactions between the road segments revealed by taxi trajectories were weak in the suburbs of the city (i.e., regions around the fifth ring road) [40]. In Figures 11–13, we can see that FacetNet, DYNMOGA, and RWLA, respectively, identified communities formed

by some weakly interacting road segments in the suburbs of the city. We can conclude that the proposed method can avoid detecting these weakly interacting communities.

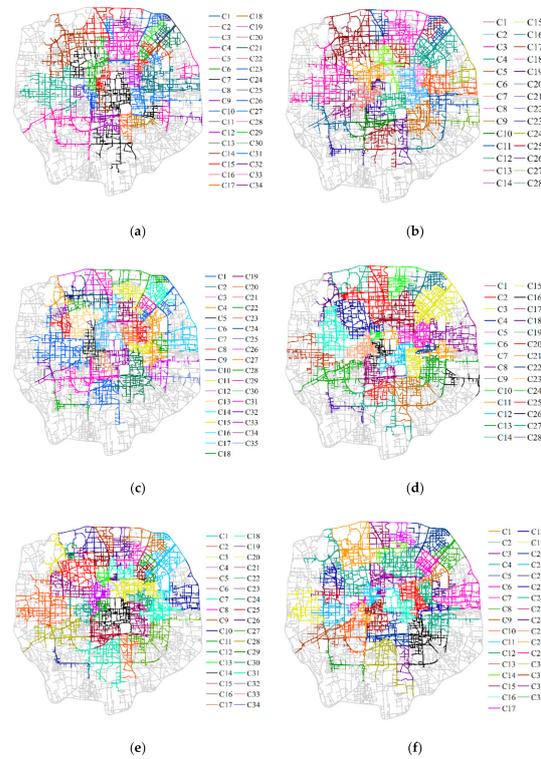


Figure 10. Communities discovered by EACO: (a) 8:00–9:00; (b) 9:00–10:00; (c) 10:00–11:00; (d) 16:00–17:00; (e) 17:00–18:00; (f) 18:00–19:00.

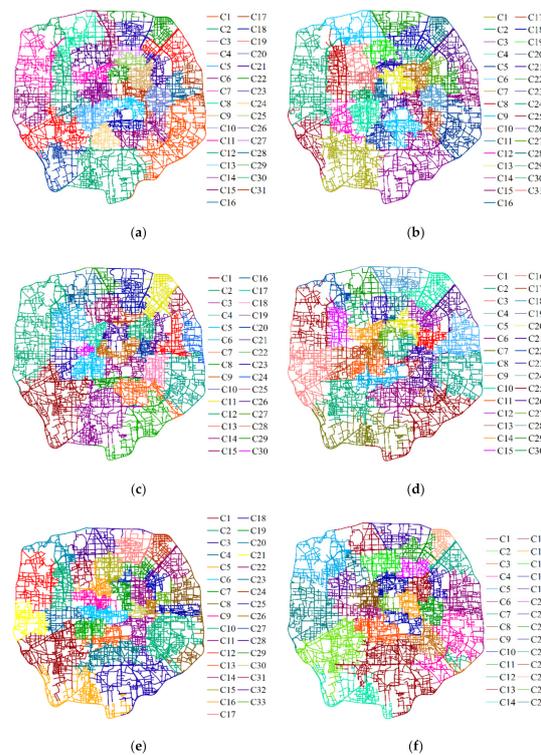


Figure 11. Communities discovered by FacetNet: (a) 8:00–9:00; (b) 9:00–10:00; (c) 10:00–11:00; (d) 16:00–17:00; (e) 17:00–18:00; (f) 18:00–19:00.

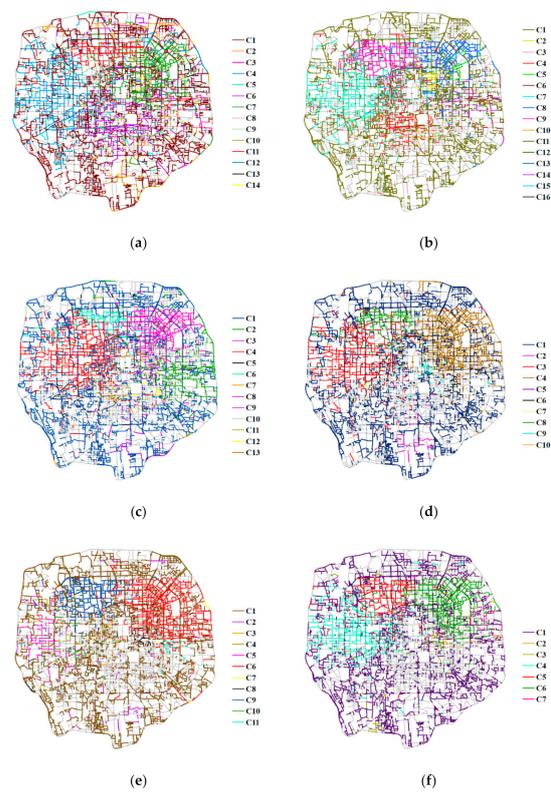


Figure 12. Communities discovered by DYNMOGA: (a) 8:00–9:00; (b) 9:00–10:00; (c) 10:00–11:00; (d) 16:00–17:00; (e) 17:00–18:00; (f) 18:00–19:00.

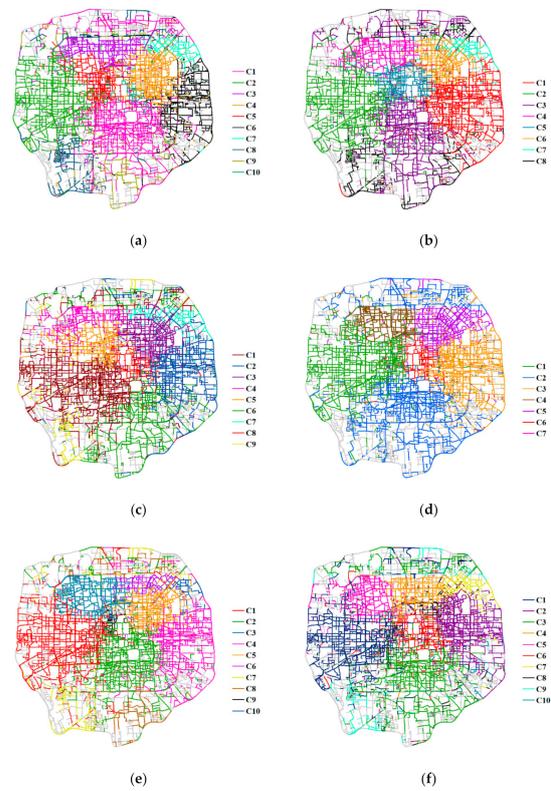


Figure 13. Communities discovered by RWLA: (a) 8:00–9:00; (b) 9:00–10:00; (c) 10:00–11:00; (d) 16:00–17:00; (e) 17:00–18:00; (f) 18:00–19:00.

We further used the average frequency of OD pairs between road segments within the same community [41] to quantitatively evaluate the performance of the three methods. The results in Table 4 show that the average frequencies of OD pairs between road segments within the communities discovered by EACO were higher than those within the communities discovered by the three comparison methods. This indicates that the proposed method performs better in identifying the spatial interaction patterns. In Table 4, we can see that the average frequencies of OD pairs between road segments within the communities discovered by EACO were slightly higher than those within the communities discovered by DYNMOGA and RWLA. Figures 12 and 13 show that the communities identified by DYNMOGA and RWLA were usually spatially dispersed. As Guo et al. [21] noted, these spatially dispersed communities were likely formed by random movements or noise. Therefore, the communities detected by DYNMOGA and RWLA may be not suitable for identifying urban structures and human mobility patterns.

Table 4. Average frequency of OD pairs between road segments within the same community.

Method	Snapshot															
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
EACO	4.6	4.5	4.5	4.3	4.6	4.5	4.4	4.5	4.3	4.3	4.3	4.3	4.3	4.4	4.3	4.2
FacetNet [19]	3.8	2.8	3.1	3.0	3.2	4.0	2.8	3.2	2.6	2.1	2.8	2.2	3.2	2.4	1.7	1.3
DYNMOGA [20]	4.1	4.1	4.1	4.1	4.1	4.1	4.1	4.0	4.0	4.1	4.0	4.1	4.1	4.1	4.0	4.0
RWLA [7]	4.3	4.2	4.2	4.2	4.2	4.3	4.2	4.2	4.2	4.2	4.2	4.2	4.2	4.1	4.1	4.1

In Figure 14, the evolution of communities identified by EACO is displayed. We can see that the communities in vehicle movements usually changed over time. The communities in the morning (8:00–10:00) and evening (16:00–19:00) rush hours changed more drastically because the intensity of human activities was very high. In the following section, the dynamic communities in the morning and evening rush hours are analyzed.

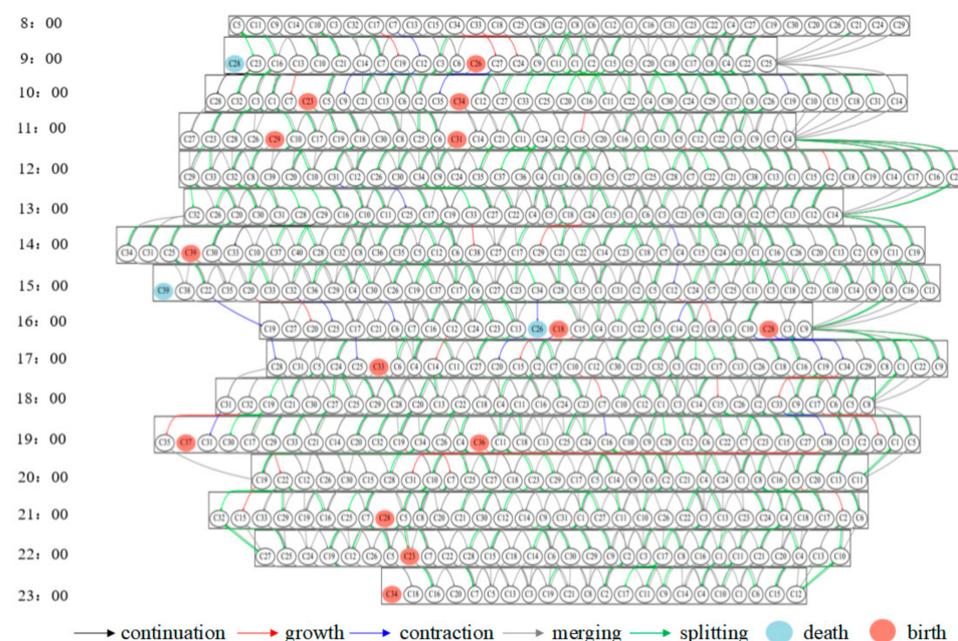


Figure 14. The evolution of the communities identified by EACO in the real-world dataset.

3.2.1. Dynamic Communities during Morning Rush Hours

During the morning rush hour, people purposefully went to work from the residential areas. Therefore, the communities around residential areas and work places usually changed more drastically. Some examples are given in Figure 15.



Figure 15. Some communities identified by EACO: (a) 8:00–9:00; (b) 9:00–10:00; (c) 10:00–11:00.

Community C7 at 8:00–9:00 contains some commercial areas and workplaces (e.g., Chongwenmen business district, Wangfujing business district, Tiananmen Square, and the R&F work area). When people are already working, the intensity of the human activity in C7 decreases. Therefore, C7 in Figure 15a gradually shrinks over time (C12 and C9 in Figure 15b,c). Community C33 at 8:00–9:00 is near the Chaoqing business district. As people continue to come to the business district, Community C33 gradually grows over time (C24 and C16 in Figure 15b,c). Community C17 at 8:00–9:00 is near the residential area of Ding’anli. Since people leave this area to go to the nearby workplaces during the morning rush hour, C17 in Figure 15a grows into C19 in Figure 15b (near Ding’anli residential area and Aegean business district) at 9:00–10:00. At 9:00–10:00, community C23 near Fengtai Science and Technology Park (Figure 15b) is separated from the large community C5 identified at 8:00–9:00. At 8:00–9:00, people go to work in Fengtai Science and Technology Park; therefore, a large community is identified around this place. After people come to their workplaces, the spatial interactions within Fengtai Science and Technology Park become strong. We can see that the dynamic communities identified by EACO can be used to reveal the effect of the morning commuting patterns on the spatial interactions between road segments.

3.2.2. Dynamic Communities during Evening Rush Hours

During the evening rush hours, people gradually left their workplaces after work, and went to the residential areas or entertainment areas. Therefore, the spatial interaction patterns changed due to mass movement of people. Some examples are given in Figure 16.

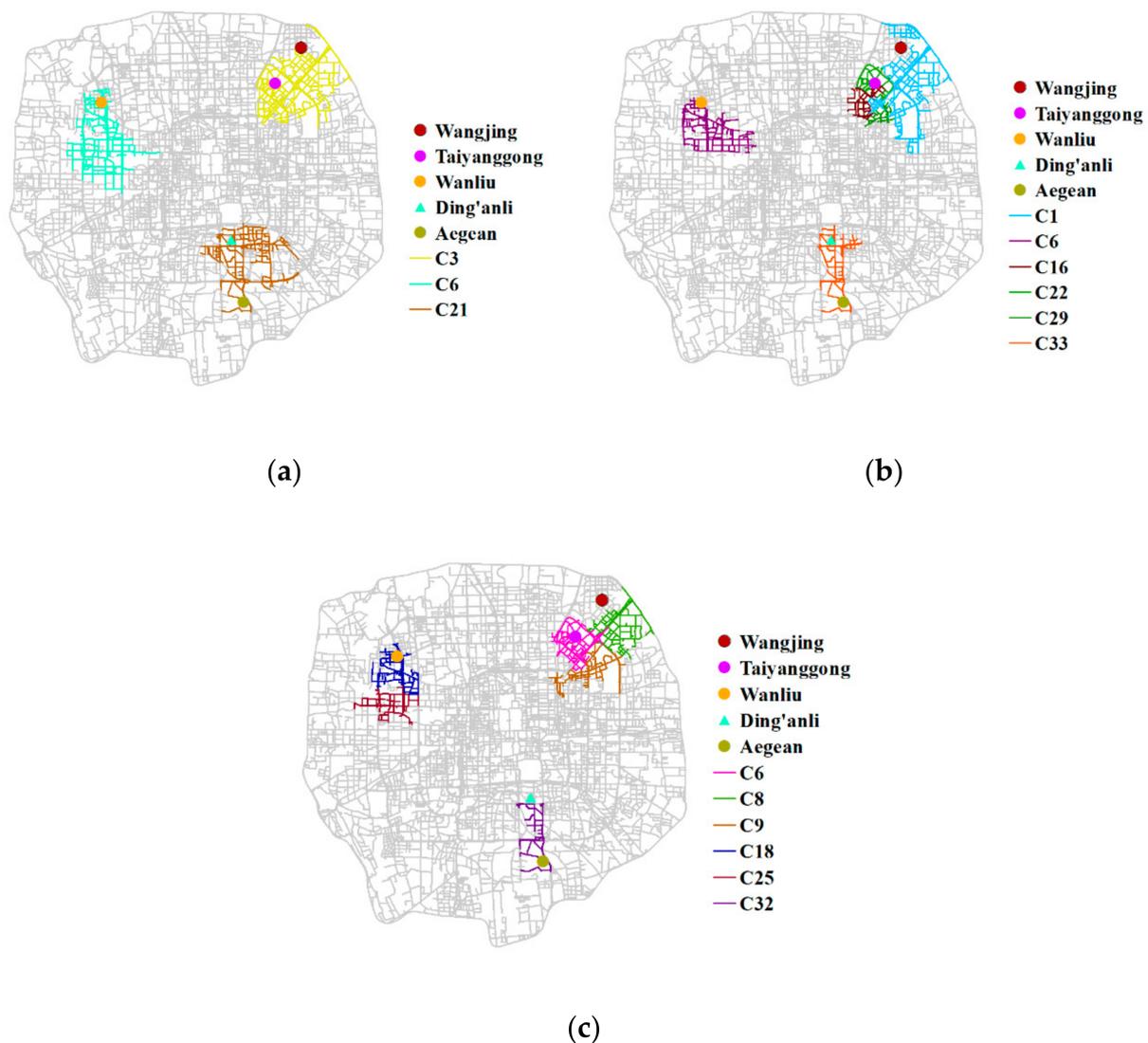


Figure 16. Some communities identified by EACO during evening rush hours: (a) 16:00–17:00; (b) 17:00–18:00; (c) 18:00–19:00.

We can see that community C3 around Wangjing and Taiyanggong (two core central business districts (CBDs)) at 16:00–17:00 gradually splits into several smaller communities at 17:00–18:00 and 18:00–19:00. The reason for this may be that human activities around CBDs become strong when lots of people go to the CBDs for entertainment after work. Community C6 (near an important CBD—Wanliu) at 16:00–17:00 gradually shrinks over time. During the evening rush hours, people try to avoid long-distance activities because of traffic congestion. Therefore, small communities appear around some local commercial or entertainment centers. We also found that the community around Ding'anli residential area and Aegean business district (C33 in Figure 16b and C32 in Figure 16c) reappeared at 17:00–18:00 and 18:00–19:00 (this community was also identified at 8:00–10:00). This community may be formed because of the large numbers of people returning to residential areas. We can infer that commuting activity is an important factor in changing the community structures.

4. Discussion

Based on the simulated and real-world experiments, we can conclude that EACO outperforms FacetNet, DYNMOGA, and RWLA in detecting dynamic communities. The primary reasons for this can be summarized as follows:

- (1) EACO can well consider the temporal partition coherence at each snapshot by using the newly built pheromone initialization method. The ACO-based community searching method used by EACO makes no assumptions about the shapes of communities. The path of an ant can be arbitrary; therefore, EACO is powerful for identifying communities of different sizes and shapes. The expectation-maximization method used by FacetNet and the genetic algorithms used by DYNMOGA actually belong to partitioning clustering methods. The Leiden algorithm used by RWLA is a hierarchical clustering method. It has been proven that partitioning clustering methods and hierarchical clustering methods usually tend to find spherical-shaped communities [42].
- (2) EACO can evaluate the statistical significance of communities. Therefore, some weakly interacting communities can be eliminated. In contrast, FacetNet, DYNMOGA, and RWLA can identify communities in a graph, even when there is not a natural community structure in that graph. For example, these two comparison methods all identified some weakly interacting communities in the suburbs of Beijing, where OD pairs were very sparse.
- (3) EACO is more robust to noise because it adds a spatial contiguity constraint to the optimization process. In Section 3.1, we show that FacetNet, DYNMOGA, and RWLA cannot accurately identify predefined communities from the datasets with noise. As shown in Section 3.2, FacetNet, DYNMOGA, and RWLA identified some spatially dispersed communities, which may be spurious patterns formed by random movements [21].

EACO can be regarded as a spatial extension of evolutionary clustering. Compared with the static communities by existing spatial community detection methods (e.g., ACOScan), the dynamic communities identified by EACO are useful for revealing the evolution of spatial interaction in a city. Spatial community detection methods can only identify the static polycentric structure of a city. In contrast, EACO can reveal how the polycentric structure of a city changes over time. Therefore, the dynamic communities identified by EACO can be used to understand the relationships between human activities and the dynamic organization of urban structure, e.g., the changes in communities caused by urban commuting activities. These dynamic communities may be helpful for understanding how human activities shape the structure of a city, and may provide useful references for human-centered urban planning [37].

5. Conclusions

In this study, we developed an evolutionary ant colony optimization method for detecting dynamic communities in vehicle movements. A multilayer snapshot network was constructed from vehicle movements, with road segments as nodes and OD pairs as weighted edges. With the help of a newly built pheromone initialization strategy, an ACO-based spatial scan statistic was upgraded to identify irregular-shaped and statistically significant dynamic communities in the multilayer snapshot network. Therefore, the identification of communities at each snapshot can balance partition quality and temporal partition coherence. Different rules defined based on the Jaccard coefficient were used to identify the evolution of spatial communities. Compared with the static spatial communities identified in vehicle movements, these dynamic communities can be used to reveal the dynamic organization of urban structures. The effectiveness and superiority of the proposed method were tested using both simulated and real-world vehicle movement datasets. The dynamic communities identified by the proposed method may aid in human-centered urban planning.

In this study, we only used taxi trajectory data as an example to explore the dynamic organization of urban structures. Therefore, the results may be biased. In the future,

multisource human mobility data (e.g., mobile phone logs, or private and public traffic travel data) should be fused to identify dynamic communities. In addition, it is time-consuming to evaluate the statistical significance of the communities using the Monte Carlo simulation method. A parallel computation method should be further developed to improve the efficiency of the proposed method.

Author Contributions: Conceptualization, Q.L. and S.Z.; data curation, S.Z. and M.C.; methodology, Q.L., S.Z. and M.C.; visualization, S.Z. and M.C.; writing—original draft, M.C. and W.L.; writing—review and editing, Q.L. and M.C. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded through support from the National Natural Science Foundation of China (NSFC) (No. 41971353 and 41730105), the Natural Science Foundation of Hunan Province (No. 2021JJ20058 and 2020JJ4695), and the Water Conservancy Science and Technology Project of Guizhou, China (No. KT202110 and KT202002).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Acknowledgments: The authors gratefully acknowledge the comments from the reviewers.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Li, Y.; Sun, Q.; Ji, X.; Xu, L.; Lu, C.; Zhao, Y. Defining the boundaries of urban built-up area based on taxi trajectories: A case study of Beijing. *J. Geovisu. Spat. Anal.* **2020**, *4*, 8. [\[CrossRef\]](#)
2. Zhu, D.; Wang, N.; Wu, L.; Liu, Y. Street as a big geo-data assembly and analysis unit in urban studies: A case study using Beijing taxi data. *Appl. Geogr.* **2017**, *86*, 152–164. [\[CrossRef\]](#)
3. Mucha, P.J.; Richardson, T.; Macon, K.; Porter, M.A.; Onnela, J.-P. Community structure in time-dependent, multiscale, and multiplex networks. *Science* **2010**, *328*, 876–878. [\[CrossRef\]](#) [\[PubMed\]](#)
4. Jia, T.; Cai, C.; Li, X.; Luo, X.; Zhang, Y.; Yu, X. Dynamical community detection and spatiotemporal analysis in multilayer spatial interaction networks using trajectory data. *Int. J. Geogr. Inf. Sci.* **2022**. [\[CrossRef\]](#)
5. Lu, F.; Liu, K.; Duan, Y.; Cheng, S.; Du, F. Modeling the heterogeneous traffic correlations in urban road systems using traffic-enhanced community detection approach. *Phys. A Stat. Mech. Appl.* **2018**, *501*, 227–237. [\[CrossRef\]](#)
6. Rossetti, G.; Cazabet, R. Community discovery in dynamic networks: A survey. *ACM Comput. Surv.* **2018**, *51*, 1–37. [\[CrossRef\]](#)
7. Newman, M.E.J. Modularity and community structure in networks. *Proc. Natl. Acad. Sci. USA* **2006**, *103*, 8577–8582. [\[CrossRef\]](#)
8. Blondel, V.D.; Guillaume, J.L.; Lambiotte, R.; Lefebvre, E. Fast unfolding of communities in large networks. *J. Stat. Mech. Theory Exp.* **2008**, *155*, e168. [\[CrossRef\]](#)
9. Rosvall, M.; Bergstrom, C.T. An information-theoretic framework for resolving community structure in complex networks. *Proc. Natl. Acad. Sci. USA* **2007**, *104*, 7327–7331. [\[CrossRef\]](#)
10. Rosvall, M.; Bergstrom, C.T. Maps of random walks on complex networks reveal community structure. *Proc. Natl. Acad. Sci. USA* **2008**, *105*, 1118–1123. [\[CrossRef\]](#)
11. Bóta, A.; Krész, M.; Pluhár, A. Dynamic communities and their detection. *Acta Cybern.* **2011**, *20*, 35–52. [\[CrossRef\]](#)
12. Palla, G.; Barabási, A.L.; Vicsek, T. Quantifying social group evolution. *Nature* **2007**, *446*, 664–667. [\[CrossRef\]](#)
13. Chen, Z.; Wilson, K.A.; Jin, Y.; Hendrix, W.; Samatova, N.F. Detecting and tracking community dynamics in evolutionary networks. In Proceedings of the IEEE International Conference on Data Mining Workshops, Sydney, Australia, 13 December 2010.
14. Falkowski, T.; Bartelheimer, J.; Spiliopoulou, M. Mining and visualizing the evolution of subgroups in social networks. In Proceedings of the IEEE/WIC/ACM International Conference on Web Intelligence, Hong Kong, China, 18–22 December 2006.
15. Liu, Q.; Zhu, S.; Deng, M.; Liu, W.; Wu, Z. A Spatial Scan Statistic to Detect Spatial Communities of Vehicle Movements on Urban Road Networks. *Geogr. Anal.* **2022**, *54*, 124–148. [\[CrossRef\]](#)
16. Bansal, S.; Bhowmick, S.; Paymal, P. Fast community detection for dynamic complex networks. *Complex Netw.* **2011**, *116*, 196–207.
17. Agarwal, M.K.; Ramamritham, K.; Bhide, M. Real time discovery of dense clusters in highly dynamic graphs: Identifying real world events in highly dynamic environments. *arXiv* **2012**, arXiv:1207.0138. [\[CrossRef\]](#)
18. Chakrabarti, D.; Kumar, R.; Tomkins, A. Evolutionary clustering. In Proceedings of the 12th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Philadelphia, PA, USA, 20–23 August 2006.
19. Lin, Y.R.; Chi, Y.; Zhu, S.; Sundaram, H.; Tseng, B.L. Analyzing communities and their evolutions in dynamic social networks. *ACM Trans. Knowl. Discov. Data* **2009**, *3*, 1–31. [\[CrossRef\]](#)

20. Folino, F.; Pizzuti, C. An evolutionary multiobjective approach for community discovery in dynamic networks. *IEEE Trans. Knowl. Data Eng.* **2014**, *26*, 1838–1852. [[CrossRef](#)]
21. Guo, D.; Jin, H.; Gao, P.; Zhu, X. Detecting spatial community structure in movements. *Int. J. Geogr. Inf. Sci.* **2018**, *32*, 1326–1347. [[CrossRef](#)]
22. Wang, T.C.; Phoa, F.K.H. A scanning method for detecting clustering pattern of both attribute and structure in social networks. *Phys. A Stat. Mech. Its Appl.* **2016**, *445*, 295–309. [[CrossRef](#)]
23. Aynaud, T.; Guillaume, J.L. Multi-step community detection and hierarchical time segmentation in evolving networks. In Proceedings of the 5th Social Network Mining and Analysis Workshop, SNA-KDD workshop, San Diego, CA, USA, 21 August 2011.
24. Gauvin, L.; Panisson, A.; Cattuto, C. Detecting the community structure and activity patterns of temporal networks: A non-negative tensor factorization approach. *PLoS ONE* **2014**, *9*, e86028. [[CrossRef](#)]
25. Ghasemian, A.; Zhang, P.; Clauset, A.; Moore, C.; Peel, L. Detectability thresholds and optimal algorithms for community structure in dynamic networks. *Phys. Rev. X* **2016**, *6*, 031005. [[CrossRef](#)]
26. Viard, T.; Latapy, M.; Magnien, C. Computing maximal cliques in link streams. *Theor. Comput. Sci.* **2016**, *609*, 245–252. [[CrossRef](#)]
27. White, C.E.; Bernstein, D.; Kornhauser, A.L. Some map matching algorithms for personal navigation assistants. *Transp. Res. Part C Emerg. Technol.* **2000**, *8*, 91–108. [[CrossRef](#)]
28. Chung, F.; Lu, L. Connected components in random graphs with given expected degree sequences. *Ann. Comb.* **2002**, *6*, 125–145. [[CrossRef](#)]
29. Wang, B.; Phillips, J.M.; Schreiber, R.; Wilkinson, D.; Mishra, N.; Tarjan, R. Spatial scan statistics for graph clustering. In Proceedings of the 2008 SIAM International Conference on Data Mining, Atlanta, GA, USA, 24–26 April 2008; pp. 727–738.
30. Benjamini, Y.; Yekutieli, D. The control of the false discovery rate in multiple testing under dependency. *Ann. Stat.* **2001**, *29*, 1165–1188. [[CrossRef](#)]
31. Dorigo, M.; Stützle, V. *Ant Colony Optimization*; MIT Press: Cambridge, MA, USA, 2004.
32. Dorigo, M.; Birattari, M.; Stutzle, T. Ant colony optimization. *IEEE Comput. Intell. Mag.* **2006**, *1*, 28–39. [[CrossRef](#)]
33. Zhou, X.; Liu, Y.; Zhang, J.; Liu, T.; Zhang, D. An ant colony based algorithm for overlapping community detection in complex networks. *Phys. A Stat. Mech. Its Appl.* **2015**, *427*, 289–301. [[CrossRef](#)]
34. Lipowski, A.; Lipowska, D. Roulette-wheel selection via stochastic acceptance. *Phys. A Stat. Mech. Its Appl.* **2012**, *391*, 2193–2196. [[CrossRef](#)]
35. Pei, T.; Wan, Y.; Jiang, Y.; Qu, C.; Zhou, C.; Qiao, Y. Detecting arbitrarily shaped clusters using ant colony optimization. *Int. J. Geogr. Inf. Sci.* **2011**, *25*, 1575–1595. [[CrossRef](#)]
36. Chakraborty, T.; Dalmia, A.; Mukherjee, A.; Ganguly, N. Metrics for community analysis: A survey. *ACM Comput. Surv.* **2017**, *50*, 1–37. [[CrossRef](#)]
37. Wu, Z.; Cao, J.; Zhu, G.; Yin, W.; Cuzzocrea, A.; Shi, J. Detecting overlapping communities in poly-relational networks. *World Wide Web* **2015**, *18*, 1373–1390. [[CrossRef](#)]
38. Traag, V.A.; Waltman, L.; van Eck, N.J. From Louvain to Leiden: Guaranteeing well-connected communities. *Sci. Rep.* **2019**, *9*, 5233. [[CrossRef](#)]
39. Ahas, R.; Mark, Ü. Location based services—New challenges for planning and public administration? *Futures* **2005**, *37*, 547–561. [[CrossRef](#)]
40. Yao, X.; Gao, Y.; Zhu, D.; Manley, E.; Wang, J.; Liu, Y. Spatial origin-destination flow imputation using graph convolutional networks. *IEEE Trans. Intell. Transp. Syst.* **2021**, *22*, 7474–7484. [[CrossRef](#)]
41. Liu, K.; Gao, S.; Lu, F. Identifying spatial interaction patterns of vehicle movements on urban road networks by topic modelling. *Comput. Environ. Urban Syst.* **2019**, *74*, 50–61. [[CrossRef](#)]
42. Tan, P.N.; Steinbach, M.; Kumar, V. *Introduction to Data Mining*; Pearson Education India: New Delhi, India, 2016.