

Automatic Identification of Human Subgroups in Time-Dependent Pedestrian Flow Networks

Wenhan Wu, Wenfeng Yi, Jinghai Li, Maoyin Chen *Member, IEEE*, and Xiaoping Zheng

Abstract—The study of identifying human subgroups from videos is a significant topic, which has received a lot of attention in multiple disciplines. So far, however, there has been little consideration about combining it with relevant conceptions in network science. Therefore, this paper proposes a novel method for the automatic identification of human subgroups in dynamic pedestrian flows. The spatial proximity and temporal continuity are combined to calculate the interaction intensity between pedestrians, by which a time-dependent pedestrian flow network is constructed. Based on the objective function of weighted partition density, the optimal threshold is used to determine community structures that correspond to human subgroups in frame images. Numerical experiments demonstrate that our method achieves high identification accuracy under various evaluation datasets, and exhibits better performance than existing methods in terms of different crowd densities, various numbers of subgroup members, and certain levels of trajectory noise. Furthermore, this work provides valuable implications for the understanding of subgroup behaviors and the modeling of subgroup movements.

Index Terms—Video analysis, human subgroups, automatic identification, network science, crowd behavior.

I. INTRODUCTION

NETWORK science covers almost all disciplines and has been widely employed to address systemic problems in the human world [1]. The topological structure plays a key role in determining the function of networks, which is reflected in various networks such as social networks, biological networks, and financial networks [2]. Despite that numerous actual networks have complex topological structures, the hierarchical organization of them is also apparent. As an intermediate layer between the single node and the complete network, the community plays a significant role in the structure and evolution of complex networks [3]. The nodes within the same community are tightly linked and have a high similarity, while those in different communities are loosely connected. With the

Manuscript received June 22, 2022; revised November 29, 2022 and February 10, 2023; accepted March 23, 2023. This work was supported by the National Major Scientific Research Instrument Development Project under Grant 61927804. (Corresponding author: Xiaoping Zheng).

Wenhan Wu, Wenfeng Yi, Maoyin Chen, and Xiaoping Zheng are with the Department of Automation, Beijing National Research Center for Information Science and Technology, Tsinghua University, Beijing 100084, China. (e-mail: wwh19@mails.tsinghua.edu.cn; ywf19@mails.tsinghua.edu.cn; mychen@mail.tsinghua.edu.cn; asean@mail.tsinghua.edu.cn).

Jinghai Li is with the School of Mechanical and Electrical Engineering, Beijing University of Chemical Technology, Beijing 100029, China (e-mail: ljhai725@163.com).

This paper has supplementary downloadable material available at <http://ieeexplore.ieee.org>, provided by the author. The material includes the extraction process of pedestrian trajectories in the self-build dataset. This material is 2.45 MB in size.

in-depth study of network science, community detection has been considered as a significant research topic in this field. The community detection algorithms in existing literature have developed in two main aspects: non-overlapping community detection and overlapping community detection [4]. These techniques are essential for understanding the structural features and functional properties of complex systems.

For non-overlapping community detection algorithms, each node in the network can only be categorized into one community. Traditional clustering-based algorithms [5], [6] were able to discover non-overlapping communities in specific networks. However, these methods may suffer from issues such as insufficient partition accuracy and peripheral nodes being ignored. Modularity [7], as an indicator for measuring partition performance, has been widely used for objective optimization in community detection algorithms [8]. Several studies in recent years have also attempted to maximize modularity by spectral optimization [9] and deep learning [10]. It is notable that modularity-based algorithms have resolution limits [11], requiring additional consideration when detecting communities. In reality, large-scale networks often have overlapping features, and the study of overlapping community detection has become another important direction. The clique percolation method based on the conception of k -cliques was proposed to perform standard component analysis on the clique–clique matrix to discover overlapping communities [12]. The link-based algorithm treated a community as a set of links rather than nodes, and detected overlapping communities by cutting the link dendrogram under the optimal partition density [13]. In addition, mathematical methods (e.g., Bayesian inference [14], non-negative matrix decomposition [15]) were adopted to explore the implicit information of overlapping communities from the aspect of statistical inference or matrix analysis.

It is well known that human activities are common in various public places (e.g., station squares, commercial streets, and subway corridors) [16]. Due to the diversification of building layouts and pedestrian facilities, the complex patterns of crowd movements can be classified into unidirectional, bidirectional, and multidirectional flows [17]. A large number of research findings have been revealed in these dynamic pedestrian flows, involving typical aspects such as behavioral characteristics, fundamental diagrams, and self-organization phenomena [18]. However, most studies are limited in the framework of crowd dynamics and are rarely combined with network science. As a result, we innovatively regard the dynamic pedestrian flow recorded in the surveillance camera as a time-dependent network. This kind of network is composed of network slices at different times, and nodes in a network

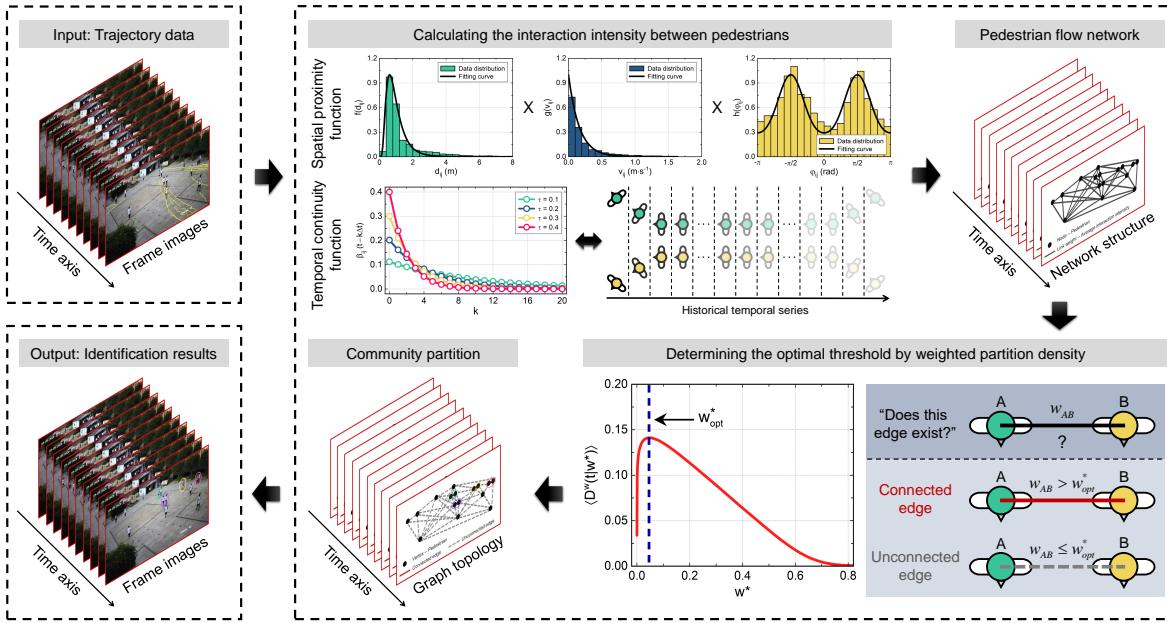


Fig. 1. General framework of the proposed method. First, the trajectory data is treated as the input to calculate the interaction intensity between pedestrians and construct a pedestrian flow network. Then, the optimal threshold is determined by weighted partition density to partition the community structure. Last, the identification results reflected by the community partition are regarded as the output.

slice correspond to pedestrians appearing on the screen at a certain time. Here, subgroups, as an intermediate layer from individuals to human crowds, are prevalent in dynamic pedestrian flows, because those pedestrians with social relationships tend to walk with their peers [19]. Note that there might be different names (e.g., social groups, pedestrian groups, and small groups.) in various publications, which are uniformly called “subgroups” in this paper. Regarding the understanding of human collective behavior, the definitions of subgroups are mainly based on socio-psychological conceptions involving entities [20], interdependence [21], social identity [22], and self-categorization theory [23]. Most previous studies judged subgroups manually based on relevant definitions, however, these methods are very time-consuming, making the automatic identification of subgroups an important issue.

Identifying subgroups from human crowds has received a lot of attention in a wide range of fields. Due to the prevalence of surveillance equipment and mobile cameras, many studies have sought to accomplish subgroup identification through computer vision techniques. Ge et al. [24] and Solera et al. [25] excavated the similarity features among subgroup members and proposed trajectory clustering algorithm-based algorithms to identify small groups. Besides, critical pedestrian features were also incorporated into the task of subgroup identification. Qin et al. [26] provided a general framework for coupling contextual information, social grouping, and computer vision tasks. Li et al. [27] developed a novel framework based on multi-level group descriptors by combining semantic information. The Internet of Things (IoT) and Virtual Reality (VR) are regarded as important supplements to data collection methods [28], and contribute more open ideas for the automatic identification of subgroups. Du et al. [29] presented a mobile device-based identification method for group mobility level

and group structure, with relatively high accuracy in real situations. Zhou et al. [30] adopted VR equipment to extract the social characteristics of individuals and developed a social interaction field model to accurately predict static and dynamic social groups. So far, however, little attention has been paid to echo the subgroup identification in dynamic pedestrian flows with the community detection in network science. Even though it is enormously challenging since the dynamic properties of network structures, this work is still a promising attempt to identify subgroups in realistic environments.

In this paper, we propose a novel method for the automatic identification of subgroups in dynamic pedestrian flows. The subgroups are defined as those who are geographically close to each other and attempt to walk together consistently, whose spatial-temporal characteristics are regarded as the crucial factors in our method. The spatial proximity and temporal continuity are combined to calculate the interaction intensity between pedestrians, which is served as the link weight for constructing a time-dependent pedestrian flow network. After that, the weighted partition density is introduced to determine the optimal threshold, which can be employed to achieve the community partition. Numerical experiments on six (i.e., five public and one self-built) evaluation datasets demonstrate the effectiveness of the proposed method. Compared with existing methods, our method exhibits better identification performance under certain conditions of the crowd density, the number of subgroup members, and the level of trajectory noise. Therefore, this work provides a general framework for identifying pedestrian subgroups automatically and contributes to the understanding of subgroup behaviors.

The rest of this paper is organized as follows. The proposed method of subgroup identification is presented in Section II. Section III provides a series of performance evaluations in

detail. Finally, the main conclusions and further prospects are discussed in Section IV.

II. SUBGROUP IDENTIFICATION METHOD

Fig. 1 shows a general framework of the proposed method. First, the input is the trajectory data extracted from the video, which is used for calculating the interaction intensity between pedestrians to construct a pedestrian flow network. Then, the optimal threshold is determined based on weighted partition density, and the community structure in the network can be identified by this threshold. Last, the output is the identification results corresponding to the community partition. The specific details of each step in this framework are described below.

A. Interaction Intensity Between Pedestrians

Based on the definitions in related socio-psychology literature, we suggest that subgroups consist of pedestrians who are geographically close and attempt to walk together consistently. From the aspect of spatial-temporal dimensions, this involves spatial proximity and temporal continuity. To quantify the interaction between pairwise pedestrians, the interaction intensity is deduced in subsequent contents.

The spatial proximity of subgroups is caused by the social interaction among members, which is mainly reflected in three observables: the relative distance $d_{ij}(t)$ between the positions of pedestrians i and j , the relative speed $v_{ij}(t)$ of pedestrian j with respect to pedestrian i , and the relative angle $\varphi_{ij}(t)$ with which pedestrian j is perceived by pedestrian i . To extract the contributions of these observables to the interaction, a key assumption in general biological considerations is that they can be decoupled from each other and constitute a product form, in analogy to the case of physical particles [31]. This assumption has been demonstrated to characterize the main features of actual interactions [32], thereby the spatial proximity function can be represented explicitly as follows:

$$\alpha_{ij}(t) = f(d_{ij}(t)) g(v_{ij}(t)) h(\varphi_{ij}(t)) \quad (1)$$

in which interaction functions $f(d_{ij})$, $g(v_{ij})$, and $h(\varphi_{ij})$ are formulated from realistic data, rather than being prefabricated [33]. The fact is that these observables tend to remain stable within certain ranges if a pair of pedestrians belong to the same subgroup, otherwise there will be a wide range of fluctuations. Hence, the distributions of observables are first extracted, and then certain mathematical functions are selected to fit them.

The temporal continuity of subgroups depends on the psychological tendency to move towards a common destination, whereby the historical temporal series of spatial proximity is considered. The newer sampled data has a larger weight in judgement, tallying with the short-term memory of human vision. Therefore, we adopt the exponentially weighted moving average (EWMA) method [34] to characterize the temporal continuity function $\beta_{ij}(t - k\Delta t)$, which reflects the influence weight of spatial proximity between pedestrians i and j at time $t - k\Delta t$ on the interaction intensity at the current time t :

$$\beta_{ij}(t - k\Delta t) = \frac{(1 - \tau)^k}{\sum_{x=0}^{\ell_{ij}^t} (1 - \tau)^x} = \frac{\tau(1 - \tau)^k}{1 - (1 - \tau)^{\ell_{ij}^t + 1}} \quad (2)$$

where Δt represents the time step between frame images, and $\ell_{ij}^t = [t - \max(t_i^s, t_j^s)]/\Delta t$ denotes the number of trajectory data points concerning pedestrians i and j until time t , in which t_i^s and t_j^s are the moments when they appear on the screen, respectively. Besides, the smoothing factor τ measures the degree of weight reduction, with a larger τ reflecting a faster weight decay of earlier data.

Based on the above discussion and deduction, the functions of spatial proximity and temporal continuity can be linearly integrated into the interaction intensity, whose mathematical form is expressed by:

$$I_{ij}(t) = \sum_{k=0}^{\ell_{ij}^t} \alpha_{ij}(t - k\Delta t) \beta_{ij}(t - k\Delta t) \quad (3)$$

Here, a larger $I_{ij}(t)$ implies that the likelihood of pedestrians i and j belonging to the same subgroup is higher. In general, $I_{ij}(t)$ and $I_{ji}(t)$ are close but unequal, because the impact of relative angle is not completely symmetrical.

B. Time-Dependent Pedestrian Flow Networks

Existing research has confirmed that most networks in natural and social systems are time-dependent and evolve dynamically [35]. In this case, the dynamic pedestrian flow recorded in the surveillance camera can be defined as a time-dependent network, which is composed of network slices at different times. The weighted adjacency matrix $\mathbf{W}(t)$ is used to describe the structure of the network slice at time t , where nodes and link weights in the network slice are respectively denoted by pedestrians and their average interaction intensities in frame image $\Gamma(t)$, which reads:

$$\mathbf{W}(t) = \begin{bmatrix} w_{11}(t) & w_{12}(t) & \cdots & w_{1N_t}(t) \\ w_{21}(t) & w_{22}(t) & \cdots & w_{2N_t}(t) \\ \vdots & \vdots & \ddots & \vdots \\ w_{N_t 1}(t) & w_{N_t 2}(t) & \cdots & w_{N_t N_t}(t) \end{bmatrix} \quad (4)$$

where N_t stands for the number of pedestrians at time t , the average interaction intensity between pedestrians i and j is defined as $w_{ij}(t) = (I_{ij}(t) + I_{ji}(t))/2$, and therefore, $\mathbf{W}(t)$ is a symmetric matrix with $w_{ij}(t) = w_{ji}(t)$.

The network slices correspond exactly to the frame images segmented by the time step Δt in the video. Considering that pedestrians on the screen often change over time, the weighted adjacency matrix $\mathbf{W}(t)$ at a different time t generally has various dimensions. Here, the time-dependent pedestrian flow network Ω is defined as a finite set, written as follows:

$$\Omega = \{\mathbf{W}(t) | t = t_0, t_0 + \Delta t, \dots, t_0 + (T - 1)\Delta t\} \quad (5)$$

where t_0 is the initial time of the video, and T indicates the total number of network slices. Note that the interslice coupling between nodes at different times is ignored, because the partition of subgroups corresponds to the set of independent partitions on each network slice.

C. Identifying Subgroups by Weighted Partition Density

Regarding the community partition in pedestrian flow networks, a graph partition method is proposed in advance. The network slice at time t can be regarded as an unweighted and undirected graph $G(t) = (V(t), E(t))$, where $V(t)$ and $E(t)$ are the sets of vertices and edges, respectively. The partition process depends on the previously defined average interaction intensity $w_{ij}(t)$ at different levels, a pre-specified threshold $w^* \in [0, 1]$ is introduced to judge whether the edge $e_{ij}(t)$ between pedestrians i and j are retained as below:

$$e_{ij}(t) = \begin{cases} 1, & w_{ij}(t) > w^* \\ 0, & w_{ij}(t) \leq w^* \end{cases} \quad (6)$$

Here, edge $e_{ij}(t)$ exists if the average interaction intensity is relatively high, otherwise it will be removed. As a result, graph $G(t)$ is partitioned into multiple connected subgraphs $\{G_1(t), \dots, G_C(t)\}$, noting that those unconnected vertices are not considered as independent subgraphs. The structures of communities $\{P_1(t), \dots, P_C(t)\}$ exactly correspond to the topologies of subgraphs $\{G_1(t), \dots, G_C(t)\}$, that is, these edges in subgraphs imply those reserved links in communities.

Inspired by the process of cutting the link dendrogram [13], we further consider defining the weighted partition density as an objective function to measure the partition effect of communities. For a community $P_c(t)$ at time t , it has $n_c(t) = |\cup_{w_{ij}(t) \in P_c(t)} \{i, j\}|$ nodes and $m_c(t) = \frac{1}{2} \sum_{\{i,j\} \in P_c(t)} e_{ij}(t)$ present links, and the sum of link weights is calculated as $w_c(t) = \frac{1}{2} \sum_{\{i,j\} \in P_c(t)} w_{ij}(t)$. Thereby, the weighted link density $D_c^w(t)$ is expressed as follows:

$$D_c^w(t) = \frac{w_c(t) - w_{\min}(n_c(t) - 1)}{w_{\max} n_c(t) (n_c(t) - 1)/2 - w_{\min}(n_c(t) - 1)} \quad (7)$$

where the maximum link weight w_{\max} is fixed as 1, and the minimum link weight w_{\min} is equal to w^* . That is, $w_c(t)$ is normalized by two extreme sums of link weights, which correspond to subgraphs on $n_c(t)$ vertices with minimum and maximum edge connectivity, respectively. For a network slice with C communities at time t , the weighted partition density $D^w(t|w^*)$ is defined as the average of $D_c^w(t)$, weighted by the fraction of present links under the condition of w^* :

$$D^w(t|w^*) = \frac{2}{M(t)} \sum_{c=1}^C m_c(t) \frac{w_c(t) - w^*(n_c(t) - 1)}{(n_c(t) - 2w^*)(n_c(t) - 1)} \quad (8)$$

where the sum of present links of C communities is calculated as $M(t) = \sum_{c=1}^C m_c(t)$. The temporal average of $D^w(t|w^*)$ from initial time t_0 to last time t_{\max} is therefore given by:

$$\langle D^w(t|w^*) \rangle_t = \frac{\Delta t}{t_{\max} - t_0 + \Delta t} \sum_{t=t_0}^{t_{\max}} D^w(t|w^*) \quad (9)$$

Here, a higher $\langle D^w(t|w^*) \rangle_t$ implies a better global structure of these partitioned communities.

Determining the optimal value of w^* is important to accurately identify subgroups. The fact is that all nodes in a network slice belong to a large community if w^* is close to 0, whereas each node is treated as an isolated individual if w^* approaches 1. Both of the above extreme situations are unreasonable, with relatively small values of $\langle D^w(t|w^*) \rangle_t$. In

TABLE I
SPECIFIC DETAILS OF EVALUATION DATASETS

Evaluation Dataset	No. P	No. S	No. SFI	Flow Type
Seq_eth	2140	424	361	bidirectional
Seq_hotel	1268	247	286	bidirectional
Crowds_zara01	1221	354	216	bidirectional
Crowds_zara02	1661	443	262	bidirectional
VEG_gall	12189	3201	300	multidirectional
BS_canteen	15502	3711	1158	multidirectional

consequence, we traverse the interval of w^* at a certain spacing Δw^* , and calculate $\langle D^w(t|w^*) \rangle_t$ at each level of w^* . Note that the number of values taken on the interval of w^* is counted as S_w . The optimal threshold w_{opt}^* is deduced as below:

$$\begin{aligned} w_{opt}^* &= \arg \max \langle D^w(t|w^*) \rangle_t \\ s.t. \quad w^* &\in \{0, \Delta w^*, 2\Delta w^*, \dots, 1\} \end{aligned} \quad (10)$$

From this, the best partition of communities would be obtained under the condition of w_{opt}^* . To better understand the effect of threshold selection on the identification performance, please see the examples given in Fig. 5. There may also be meaningful community structures under other thresholds, however, the purpose of finding w_{opt}^* is to make subgroup identification closer to human subjective perception.

III. PERFORMANCE EVALUATION

A. Evaluation Datasets and Metric

Before assessing the performance of identification methods, it is necessary to establish corresponding evaluation datasets. First, we adopted a series of standard procedures to extract 12,326 pedestrian trajectories at Bashu Secondary School in Chongqing, China (see Supplementary Materials). However, the density, speed, and flow of pedestrians are complicated and variable in this area, which makes it unrepresentative to arbitrarily select a video sequence as the evaluation dataset. Thereby, 1,158 frame images were extracted from the video by sampling at equal intervals to ensure that the characteristics of human crowds at various time periods are covered, and 5 participants were invited to independently mark the labels of subgroups. These participants recorded the IDs of subgroup members by perceptual judgements, and made auxiliary judgements using video sequences within a certain range before and after these frame images. The judgement criteria are based on social interactions such as conversation, gestures, laughter, smiles, and play [36]. Those labels of subgroups on which more than half of the participants reached consensus are finally ascertained as correct samples (including duplicate cases), and this self-build evaluation dataset is labeled as BS_canteen in subsequent contents.

Moreover, five public datasets recorded in different natural environments were also selected to enhance the reliability of performance evaluation. The two datasets named Seq_eth and Seq_hotel were collected at 2.5FPS, covering low-density scenarios near a college building and a bus station [37]. The next two datasets called Crowds_zara01 and Crowds_zara02 captured the movement of crowds with a frame rate of 25FPS

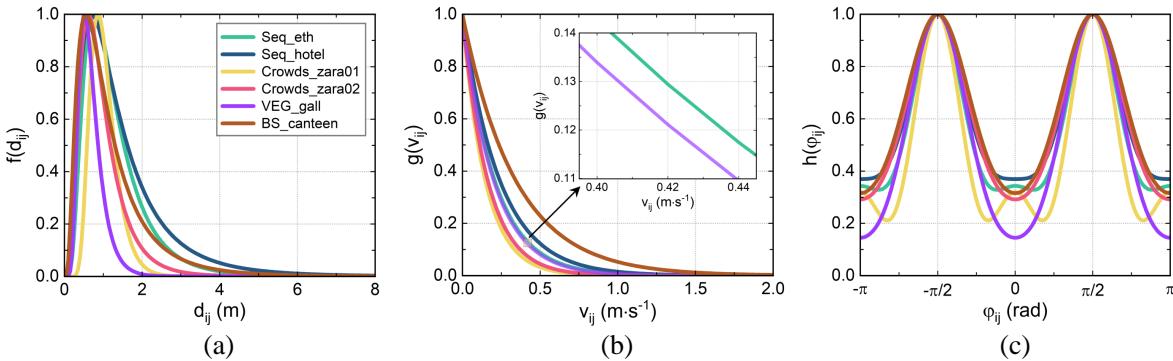


Fig. 2. Interaction functions extracted from various datasets. (a) $f(d_{ij})$ as an interaction function of relative distance d_{ij} . (b) $g(v_{ij})$ as an interaction function of relative speed v_{ij} . (c) $h(\varphi_{ij})$ as an interaction function of relative angle φ_{ij} . The curves with different colors correspond to various datasets.

TABLE II
CONFUSION MATRIX OF THE BINARY CLASSIFICATION

		Ground Truth	
		Positive Class	Negative Class
Judgement Result	Claim Positive	TP	FP
	Claim Negative	FN	TN

[38], which were taken at a low-density shopping street. The last dataset labeled VEG_gall, which was recorded in Vittorio Emanuele II Gallery at 8FPS [39], is a five-minute sequence with rapid and continuous changes in crowd density. We annotated subgroups in these datasets using a similar method as before to create the evaluation datasets, and Table I presents the specific details involving the number of pedestrians (No. P), subgroups (No. S), and sampling frame images (No. SFI), as well as the type of pedestrian flow (Flow Type).

To evaluate the consistency between identification results and human judgements, the F1-score is introduced as an evaluation metric. For binary classification problems, the confusion matrix is a standard representation of accuracy assessment. As shown in Table II, the subgroups correctly identified by relevant methods are regarded as true positives (TP), the subgroups not successfully identified as false negatives (FN), whereas the subgroups that are identified but inconsistent with the correct samples as false positives (FP) [30]. Given that subgroup identification focuses more on positive results, true negatives (TN) are therefore not discussed in our case. From this, the F1-score is defined as the harmonic mean of Precision and Recall, calculated as follows:

$$\text{F1-score} = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \quad (11)$$

where Precision = TP/(TP+FP) is the number of TP results divided by the number of all results declared positive, and Recall = TP/(TP + FN) is the number of TP results divided by the number of all true samples.

It is worth emphasizing that the subgroup labels are marked on sampling frame images, rather than directly providing the subgroup labels for the whole video as in most studies. The reason for this is that the members of partial subgroups change

dynamically in the process of movement [40]. For instance, assuming that pedestrians i and j belong to a subgroup $\{i, j\}$ in the previous period, whereas this subgroup disintegrates $\{i, j\} \rightarrow \{i\}, \{j\}$ or reorganizes $\{i, j\} \rightarrow \{i, j, k\}$ in the later period. In this case, the subgroup label should be adjusted, instead of remaining $\{i, j\}$ throughout the process. In addition, it is unreasonable to compare the judgement results on sampling frame images with the subgroup labels at the video level. Therefore, the subgroup labels at the frame image level are more rigorous, which is also beneficial for a more accurate calculation of the F1-score.

B. Identification Results of the Proposed Method

The first step in our method is to extract the interaction functions for deducing the mathematical form of spatial proximity function. For any individual in a time-dependent pedestrian flow network, the interaction with the nearest neighbor is the most noteworthy. If they belong to a subgroup, the observables of relative distance, relative speed, and relative angle generally will remain stable within certain ranges, which results from a trade-off between walking faster and facilitating social interactions [41]. In contrast, there will be large fluctuations in observables if a pedestrian and the nearest neighbor are not members of a subgroup. That is, those intervals with high frequency in the distributions of observables reveal the interaction preference between neighboring members of subgroups. These distributions are expected to be fitted for determining the interaction functions, which can be further utilized to calculate the spatial proximity between pairwise pedestrians.

By conducting statistical analysis, it is amazing that these variables are found to follow similar distributions in various datasets. First, the distribution of relative distance reflects that the neighboring members of subgroups are more likely to keep a preference distance, which is fitted well by the log-normal distribution function:

$$f(d_{ij}) = \frac{A_d}{\sqrt{2\pi}\sigma d_{ij}} \exp \left[-\left(\ln d_{ij} - \mu \right)^2 / 2\sigma^2 \right] \quad (12)$$

where μ and $\sigma > 0$ are the mean and the standard deviation of logarithmic values, respectively. Second, the distribution of relative speed is consistent with the exponential distribution as

TABLE III
FITTING PARAMETERS IN INTERACTION FUNCTIONS

Evaluation Dataset	$f(d_{ij})$			$g(v_{ij})$		$h(\varphi_{ij})$			
	μ	σ	A_d	λ	A_v	a_0	a_2	a_4	A_φ
Seq_eth	0.079	0.590	1.346	4.868	0.205	0.150	-0.088	0.030	3.734
Seq_hotel	0.154	0.718	1.621	4.199	0.238	0.155	-0.080	0.019	3.941
Crowds_zara01	-0.031	0.338	0.776	6.878	0.145	0.159	-0.114	0.062	2.980
Crowds_zara02	-0.179	0.561	1.006	6.202	0.161	0.152	-0.090	0.012	3.951
VEG_gall	-0.469	0.442	0.630	5.026	0.199	0.154	-0.129	0.018	3.312
BS_canteen	-0.064	0.727	1.313	2.947	0.339	0.151	-0.083	0.009	4.113

the neighboring members of subgroups tend to walk at close speeds, whereby the function is expressed as follows:

$$g(v_{ij}) = A_v \lambda \exp(-\lambda v_{ij}) \quad (13)$$

where $\lambda > 0$ denotes the rate parameter. Third, the distribution of relative angle indicates that the neighboring members of subgroups prefer to stand orthogonal to the walking direction, and the anisotropy of relative angle can be described in the form of a Fourier series expansion [32], with at most two non-zero modes being enough to fit the distribution:

$$h(\varphi_{ij}) = A_\varphi (a_0 + a_2 \cos(2\varphi_{ij}) + a_4 \cos(4\varphi_{ij})) \quad (14)$$

where a_i ($i = 0, 2, 4$) are the non-zero Fourier coefficients. Note that A_d , A_v , and A_φ are scale factors used to control the value domain of interaction functions within $[0, 1]$. Fig. 2 depicts the curves of interaction functions extracted from various datasets, among which these subtle differences are caused by a variety of factors such as personal features [42] and environmental influences [43]. The fitting parameters in interaction functions are listed in Table III, and we also use a t -test to confirm the rationality of each parameter. As a result, if the observables of pairwise individuals are more in accordance with the interaction preference between neighboring members of subgroups, the spatial proximity calculated by these interaction functions will be closer to 1.

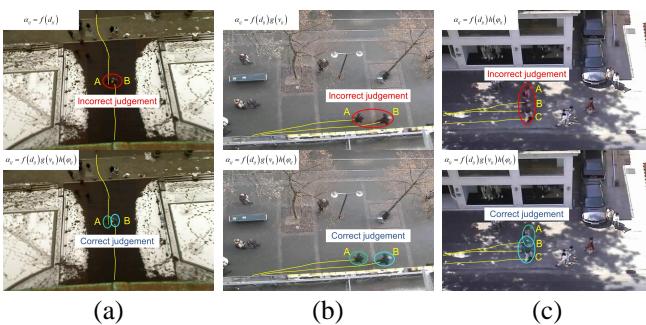


Fig. 3. Effect of interaction functions on the identification performance of the proposed method. (a) Under the conditions of $\alpha_{ij} = f(d_{ij})$. (b) Under the conditions of $\alpha_{ij} = f(d_{ij})g(v_{ij})$. (c) Under the conditions of $\alpha_{ij} = f(d_{ij})h(\varphi_{ij})$. All three cases are compared with the case under the conditions of $\alpha_{ij} = f(d_{ij})g(v_{ij})h(\varphi_{ij})$.

The reason why the three observables involving distance, speed, and angle are selected is worth explaining here. In fact, distance is the primary concern in previous studies, because

many classical clustering algorithms (e.g., K-means) are based on distance metrics. After that, the impact of speed or angle is gradually taken into account due to the walking pattern of subgroups. According to this, the spatial proximity function is assumed to be constructed by the interaction functions of one or two observables. Fig. 3(a) indicates that pedestrians A and B who approach face to face may be incorrectly identified as belonging to a subgroup if only distance is considered. If distance and speed are combined, it is likely to be misjudged as a subgroup in the case of pedestrian A following pedestrian B in Fig. 3(b). If the combination includes distance and angle, as illustrated in Fig. 3(c), pedestrian A with slower speed would be wrongly assigned to a faster subgroup of pedestrians B and C. These results demonstrate that all three observables are indispensable, which are also in line with those mentioned factors in a pioneering research [44].

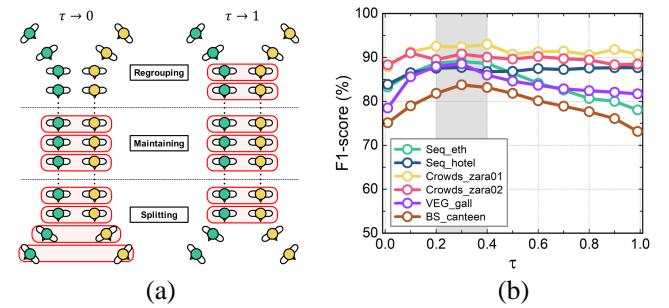


Fig. 4. Effect of smoothing factor τ on the identification performance of the proposed method. (a) A theoretical analysis of identification effects in extreme cases ($\tau \rightarrow 0$ and $\tau \rightarrow 1$). (b) Performance evaluation under different values of smoothing factor τ . The curves with different colors correspond to various evaluation datasets, and the optimal range is highlighted in grey.

The next part is concerned with the temporal continuity function, in which smoothing factor τ has a key impact on the identification performance. A theoretical analysis is performed in Fig. 4(a) to interpret the identification effects in extreme cases. $\tau \rightarrow 0$ denotes that the weights of data in the historical temporal series are almost the same. Due to the over-reliance on earlier data, incorrect judgements may be created when the subgroup members change dynamically. On the contrary, $\tau \rightarrow 1$ means that no data in the historical temporal series will be considered in the current judgement. Even if a small disturbance could affect the identification results, resulting in poor robustness of the proposed method. Fig. 4(b) shows the

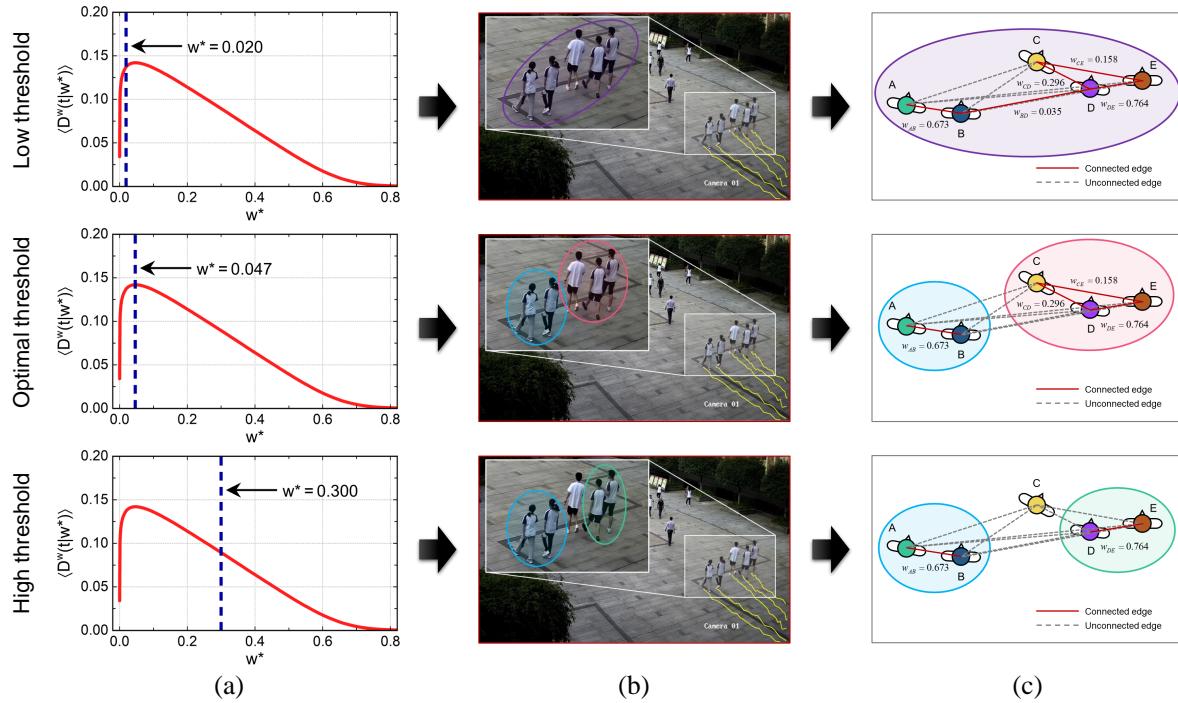


Fig. 5. Effect of threshold w^* on the identification performance of the proposed method. (a) $\langle D^w(t|w^*) \rangle_t$ as a function of threshold w^* , where low, optimal, and high thresholds are selected from top to bottom. (b) Examples of the identification results in frame images. (c) Network topologies corresponding to different cases in (b). The identified subgroups are marked by circles with different colors.

performance evaluation under different values of smoothing factor τ . It is notable that it has little effect on the datasets (Seq_hotel, Crowds_zara01, and Crowds_zara02) where most subgroups remain stable, while the situation is reversed for the datasets (Seq_eth, VEG_gall, and BS_canteen) with more changes in subgroup members. In general, there is an optimal range (in grey) of smoothing factor τ , and $\tau = 0.3$ is chosen as a reasonable value in subsequent simulations.

Once the interaction intensity between pedestrians is calculated, the optimal threshold can be determined by weighted partition density to identify the community structure. Nonetheless, we still wonder about the effect of threshold w^* on the identification performance of the proposed method. As shown in Fig. 5(a), the cases of low, optimal, and high thresholds are selected on the basis of $\langle D^w(t|w^*) \rangle_t$, and examples of the identification results in frame images are given in Fig. 5(b). These divisions are derived from the network topologies in Fig. 5(c), it is apparent that the judgement results under the optimal threshold are most in line with human perception. The analysis reveals the importance of determining an appropriate threshold for identification performance. In this way, the optimal thresholds of various datasets are calculated as below: 0.096 (Seq_eth), 0.038 (Seq_hotel), 0.020 (Crowds_zara01), 0.037 (Crowds_zara02), 0.098 (VEG_gall), and 0.047 (BS_canteen). These optimal thresholds are adopted to realize the community partition in evaluation datasets, and the snapshots of identified subgroups marked by circles with different colors are presented in Fig. 6. These values of the evaluation metric listed below the snapshots achieve satisfactory results, which indicates the effectiveness of the proposed method.

C. Quantitative Comparison with Existing Methods

Turning now to the quantitative comparison with existing methods, we choose one prevalent clustering-based method and three subgroup identification methods proposed in recent years. DBSCAN [45], as one of the most classical density-dependent clustering methods, has been widely used to discover clusters of arbitrary shape without specifying the number of clusters in advance. Ge et al. [24] developed a bottom-up hierarchical clustering method to identify subgroups using a Hausdorff distance defined according to pairwise proximity and velocity. Solera et al. [25] established the learning-based method to capture the affinity between crowd members through a structural SVM framework and specially designed features, and provided a correlation clustering procedure to detect subgroups. Zaki et al. [46] considered the commonality of walking strategies between nearby pedestrians, and proposed an automated analysis method for discovering subgroups. These three methods are almost purely dependent on pedestrian trajectories, which is consistent with the fundamental data used in the proposed method. In the following contents, we conduct several simulations from different aspects to compare the performance of these methods.

Under the conditions of reasonable parameter settings, Table IV presents the detailed performance evaluation (Precision, Recall, and F1-score) of various identification methods. The performance of our method shows a significant improvement over existing methods, which is reflected in these evaluation datasets. From the principles of existing methods, we expect to find some reasons for explaining the performance gaps. First, DBSCAN merely treats individuals clustered in space

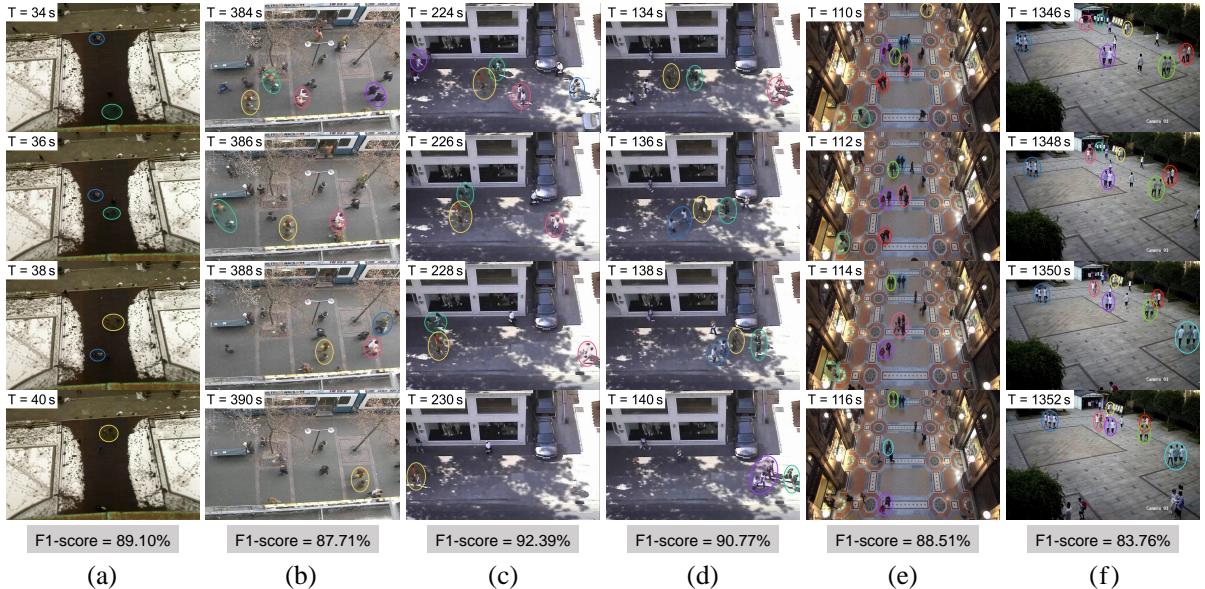


Fig. 6. Performance evaluation of the proposed method for various evaluation datasets. (a) Seq_eth. (b) Seq_hotel. (c) Crowds_zara01. (d) Crowds_zara02. (e) VEG_gall. (f) BS_canteen. The snapshots are selected from various evaluation datasets as demonstration examples, and the identified subgroups are marked by circles with different colors. These values of the evaluation metric are listed below the snapshots in different columns.

TABLE IV
PERFORMANCE EVALUATION OF VARIOUS IDENTIFICATION METHODS (%)

Evaluation Dataset	DBSCAN [45]			Ge et al. [24]			Solera et al. [25]			Zaki et al. [46]			Our Method			
	P	R	F1	P	R	F1	P	R	F1	P	R	F1	P	R	F1	
Seq_eth	58.97	56.60	57.76	73.47	76.42	74.91	77.80	83.49	80.55	75.76	76.65	76.20	88.58	89.62	89.10	
Seq_hotel	68.91	53.85	60.45	94.79	80.97	87.34	81.61	86.23	83.86	84.40	74.49	79.14	82.27	93.93	87.71	
Crowds_zara01	74.83	61.30	67.39	83.81	83.33	83.57	85.39	85.88	85.63	79.66	78.53	79.09	92.13	92.66	92.39	
Crowds_zara02	67.33	68.85	68.08	75.99	86.46	80.89	88.07	86.68	87.37	78.30	83.07	80.61	86.50	95.49	90.77	
VEG_gall	59.49	56.89	58.16	62.60	73.48	67.61	69.11	79.94	74.13	67.10	64.17	65.60	89.34	87.69	88.51	
BS_canteen	60.50	63.51	61.97	72.06	65.18	68.45	75.10	66.80	70.71	77.92	73.24	75.51	84.77	82.78	83.76	

as subgroups, but ignores the implicit common characteristics of subgroup members. It exhibits the worst effects compared to those designed specifically for identifying subgroups, and this method is therefore not further analyzed in subsequent experiments. Next, the two methods proposed by Ge et al. and Zaki et al. lack certain sensitivity when the members within a subgroup change, because the identification results are based on the closeness between complete trajectories. In addition, although the method provided by Solera et al. achieves superior performance on crowds with variable density, it still has information loss due to the neglect of other factors such as speed. These results, while preliminary, show that our method presents better performance than state-of-the-art identification methods, even if the Precision is slightly lower under certain evaluation datasets. The possible reason is that extra subgroup structures deviating from human perception might be produced in the partition process using the optimal threshold, but it has almost no impact on the overall superiority of our method.

The previous research has pointed out that human subgroups become less perceptible as the crowd density increases [47], thereby the impact of crowd density on the identification

performance is worth studying. The method defined in Supplementary Materials is utilized to calculate the crowd density in frame images, whose values are divided into three ranges from low to high for each dataset. Note that the size of crowd density in various datasets is independent of each other, and its degree is only relative to the current dataset. Fig. 7 shows the performance evaluation of various identification methods under different crowd densities. The performance of all methods decreases as the crowd density raises, which echoes the perceptual differences of subgroups in the previous research. However, under most conditions where the crowd densities are close, the F1-score of the proposed method is apparently higher than that of other methods. In particular, the performance advantage is significant at medium and high densities, which might be related to the fact that these interaction functions are able to effectively capture the preference characteristics of subgroup members even in crowded situations. As a result, our method has been demonstrated to cope with the change of crowd density at non-extreme densities and accomplish relatively better identification results.

The proportion of subgroup size has been found to follow

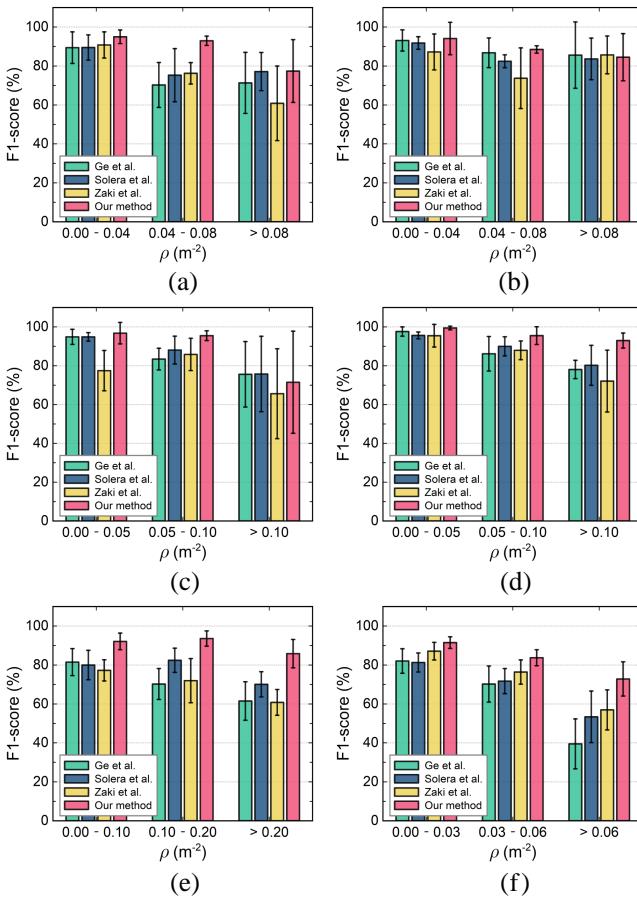


Fig. 7. Performance evaluation of various identification methods under different crowd densities. (a) Seq_eth. (b) Seq_hotel. (c) Crowds_zara01. (d) Crowds_zara02. (e) VEG_gall. (f) BS_canteen. The strips with different colors correspond to various identification methods, and the error bars represent standard deviations.

a (truncated) Poisson distribution by field observations [41], and it was also confirmed to change with the environments in diverse places [48], [49]. We are interested in the identification performance of these methods under different numbers of subgroup members. The previous research has indicated that subgroups composed of 2 to 4 members are common, whereas those with more than 4 members are rare [41]. As illustrated in Fig. 8, the identification accuracy emerges a downward trend with the increase of subgroup size, which is attributed to those small subgroups (i.e., 2 to 4 members) prone to distinguishing. It can be seen from Fig. 8(b)-8(d) that there are no large subgroups (i.e., more than 4 or 5 members) in these evaluation datasets, and certain existing methods even slightly dominate the identification accuracy of subgroups with 3 or 4 members, but this does not influence the superiority of our method in overall performance. Even for other evaluation datasets where large subgroups exist in Fig. 8(a) and 8(e)-8(f), our method also exhibits better performance for different subgroup sizes. Nevertheless, it should be emphasized that the overall performance of these methods in identifying large subgroups is relatively poor, in large part because participants are also more controversial in marking their labels. This suggests the

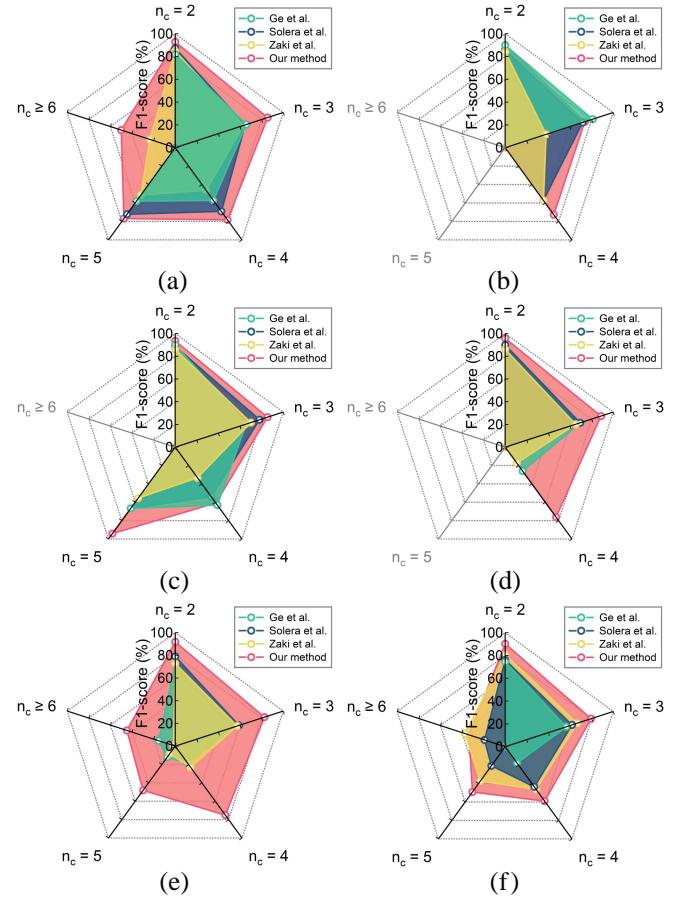


Fig. 8. Performance evaluation of various identification methods under different numbers of subgroup members. (a) Seq_eth. (b) Seq_hotel. (c) Crowds_zara01. (d) Crowds_zara02. (e) VEG_gall. (f) BS_canteen. The areas with different colors in radar maps correspond to various identification methods.

fact that relevant methods necessitate focusing more on the identification performance of small subgroups.

In practice, limited by the conditions of video collection (e.g., device performance, environmental factors), there may be a certain degree of noise in pedestrian trajectory data. Here, it is supposed that the noise follows a Gaussian distribution as $\varepsilon_i(t) \sim \mathcal{N}(0, \sigma_N^2)$, where σ_N is the standard deviation and its values are set as 0 (no noise), 0.05, 0.10, and 0.20. These Gaussian noises with different levels are superimposed to each independent pedestrian trajectory, Fig. 9 displays the corresponding performance of various identification methods. The identification accuracy of existing methods is still hard to surpass our method at low and medium levels of noise ($\sigma_N = 0.05$ and $\sigma_N = 0.10$), whereas our method could be inferior to the methods proposed by Solera et al. and Zaki et al. on certain evaluation datasets at a high level of noise ($\sigma_N = 0.20$). It reveals that our method has a more sensitive tendency to be affected by larger noise, owing to the strong constraint between the interaction functions of observables (i.e., similar to “and” in logical operations). However, it is not a bad thing because those methods with strong robustness are hard to find the fluctuations in trajectory data, which is not conducive to

identifying the changes in subgroup members. Therefore, our method is able to guarantee better performance by controlling pedestrian trajectories to a certain noise level.

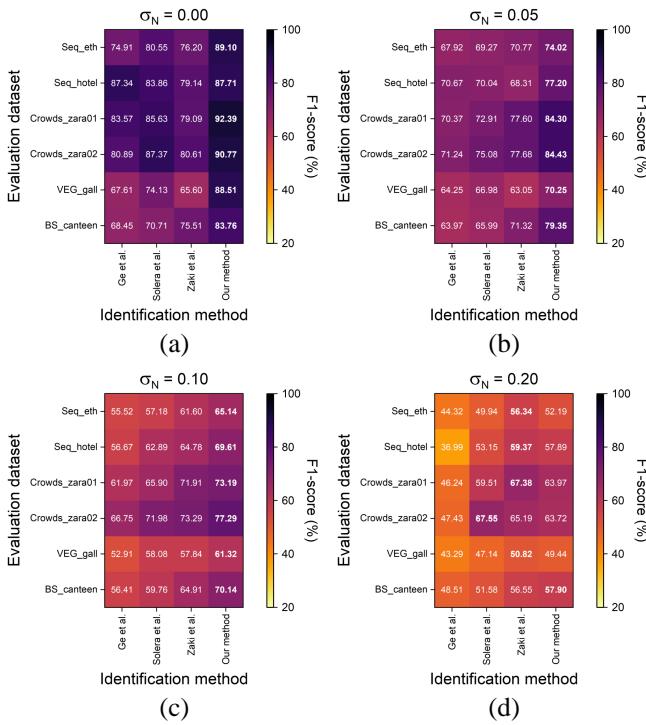


Fig. 9. Performance evaluation of various identification methods under different levels of Gaussian noise. (a) $\sigma_N = 0.00$. (b) $\sigma_N = 0.05$. (c) $\sigma_N = 0.10$. (d) $\sigma_N = 0.20$. The numbers in the squares stand for the precise F1-scores (%).

The final part of this section moves on to discuss the running performance of our method. The overall computational complexity can be approximately written as $\mathcal{O}(S_w T N_t^2)$, which derives from the joint contribution of S_w (the number of values taken on the interval of the threshold), T (the number of frame images or network slices), and N_t (the number of pedestrians in each frame image). It is also worth noting that the former two elements, S_w and T , depend largely on the initialization settings, we therefore focus more on the runtime for each frame image with N_t pedestrians. The code of our method is implemented in MATLAB (Intel Core i7 CPU, 3.0GHz, 16GB RAM) without pre-specified optimization or parallelization. By traversing these datasets, the maximum runtime consumed to calculate the weighted adjacency matrix (i.e., average interaction intensities between all pairwise pedestrians) for a frame image is fewer than 3 s. Besides, it just requires less than 0.005 s to partition subgroups by weighted partition density on a network slice corresponding to the frame image.

IV. CONCLUSION

This paper proposes a novel method for the automatic identification of subgroups in dynamic pedestrian flows. Inspired by community detection in network science, a time-dependent dynamic network is constructed by calculating the interaction intensity between pedestrians from spatial-temporal dimensions. The community structure can be discovered using the

optimal threshold, which is determined by the objective function of weighted partition density. By conducting a series of performance evaluation experiments, several main conclusions are summarized as follows: (1) The observables of relative distance, relative speed, and relative angle between pedestrians and their nearest neighbors follow certain distributions, which can be fitted to form interaction functions for describing the interaction preference of subgroup members. (2) To some extent, the identification results determined by the optimal threshold are more in accordance with those judged by human perception, this is crucial for inferring the community structure and ensuring the identification performance. (3) Our method shows high accuracy of subgroup identification under various evaluation datasets. Compared with existing methods, it also significantly improves the performance in terms of different crowd densities, various numbers of subgroup members, and certain levels of trajectory noise.

It is obvious that automatically inferring subgroups from human crowds from the video is difficult, which poses a huge challenge for relevant research on subgroups. Therefore, this work is beneficial for understanding the potential characteristics of subgroup behaviors. For example, researchers have discovered that subgroups prefer to walk side by side at low density, form “V”-like or “U”-like shapes at medium density [41], and present “river-like” structures at high density [50]. How these spatial configurations of subgroups can be quantitatively explained by these interaction functions is worthy of exploration. In addition, it is expected to facilitate the generation of more realistic models of subgroup movements. The group cohesion force [41], leader-follower principle [51], and local behaviors of subgroups [52] have been incorporated into classical crowd models such as the social force model, cellular automata model, and agent-based model. Nevertheless, subgroup models with high generalizability and verifiability are still lacking, whether these findings on the interaction preference of subgroup members can be utilized to construct such models necessitates further consideration. In a word, the insights gained from this study may be of assistance to further research of human subgroups in crowd dynamics.

Despite that the proposed method has achieved better performance for subgroup identification, it is still necessary to indicate its main limitation. In the process of separating the community structure, the perfect community partition on each network slice may not be determined by the optimal threshold defined at the global level. This fact is likely to cause a few identification results to deviate from human perception, which affects the performance of our method to a certain extent. As a consequence, determining the optimal threshold at the local level to achieve the best community partition in each network slice is expected to be explored in further research. Notwithstanding this limitation, this paper still offers valuable insights for identifying subgroups. For instance, these interaction functions are purely extracted from realistic data, rather than being set as specific similarity functions (e.g., Gaussian kernel function, cosine similarity function) in a plausible way [53]. Because the prefabricated function is hard to capture the preference characteristics of subgroup members in different environments, resulting in less accurate measures of

spatial proximity than those data-driven interaction functions. In summary, this work has important implications for subgroup identification in videos, and also promotes the understanding of crowd dynamics from the perspective of network science.

REFERENCES

- [1] S. H. Strogatz, "Exploring complex networks," *Nature*, vol. 410, no. 6825, pp. 268–276, mar 2001.
- [2] M. E. J. Newman, "The structure and function of complex networks," *SIAM Review*, vol. 45, no. 2, pp. 167–256, jan 2003.
- [3] S. Fortunato, "Community detection in graphs," *Physics Reports*, vol. 486, no. 3-5, pp. 75–174, feb 2010.
- [4] M. A. Javed, M. S. Younis, S. Latif, J. Qadir, and A. Baig, "Community detection in networks: A multidisciplinary review," *Journal of Network and Computer Applications*, vol. 108, pp. 87–111, apr 2018.
- [5] H. Shen, X. Cheng, K. Cai, and M.-B. Hu, "Detect overlapping and hierarchical community structure in networks," *Physica A: Statistical Mechanics and its Applications*, vol. 388, no. 8, pp. 1706–1712, apr 2009.
- [6] J. Shi and J. Malik, "Normalized cuts and image segmentation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 22, no. 8, pp. 888–905, aug 2000.
- [7] M. E. J. Newman and M. Girvan, "Finding and evaluating community structure in networks," *Physical Review E*, vol. 69, no. 2, p. 026113, feb 2004.
- [8] A. Clauset, M. E. J. Newman, and C. Moore, "Finding community structure in very large networks," *Physical Review E*, vol. 70, no. 6, p. 066111, dec 2004.
- [9] M. E. J. Newman, "Spectral methods for community detection and graph partitioning," *Physical Review E*, vol. 88, no. 4, p. 042822, oct 2013.
- [10] L. Wu, Q. Zhang, C.-H. Chen, K. Guo, and D. Wang, "Deep learning techniques for community detection in social networks," *IEEE Access*, vol. 8, pp. 96016–96026, may 2020.
- [11] S. Fortunato and M. Barthélémy, "Resolution limit in community detection," *Proceedings of the National Academy of Sciences*, vol. 104, no. 1, pp. 36–41, jan 2007.
- [12] G. Palla, I. Derényi, I. Farkas, and T. Vicsek, "Uncovering the overlapping community structure of complex networks in nature and society," *Nature*, vol. 435, no. 7043, pp. 814–818, jun 2005.
- [13] Y.-Y. Ahn, J. P. Bagrow, and S. Lehmann, "Link communities reveal multiscale complexity in networks," *Nature*, vol. 466, no. 7307, pp. 761–764, jun 2010.
- [14] T. Yang, Y. Chi, S. Zhu, Y. Gong, and R. Jin, "Detecting communities and their evolutions in dynamic social networks—a bayesian approach," *Machine Learning*, vol. 82, no. 2, pp. 157–189, sep 2010.
- [15] S. Mankad and G. Michailidis, "Structural and functional discovery in dynamic networks with non-negative matrix factorization," *Physical Review E*, vol. 88, no. 4, p. 042812, oct 2013.
- [16] Y. Wu, Y. Ye, C. Zhao, and Z. Shi, "Collective density clustering for coherent motion detection," *IEEE Transactions on Multimedia*, vol. 20, no. 6, pp. 1418–1431, jun 2018.
- [17] X. Shi, Z. Ye, N. Shiawakoti, and O. Grembek, "A state-of-the-art review on empirical data collection for external governed pedestrians complex movement," *Journal of Advanced Transportation*, vol. 2018, pp. 1–42, sep 2018.
- [18] H. Dong, M. Zhou, Q. Wang, X. Yang, and F.-Y. Wang, "State-of-the-art pedestrian and evacuation dynamics," *IEEE Transactions on Intelligent Transportation Systems*, vol. 21, no. 5, pp. 1849–1866, may 2020.
- [19] A. Nicolas and F. H. Hassan, "Social groups in pedestrian crowds: review of their influence on the dynamics and their modelling," *Transportmetrica A: Transport Science*, vol. 19, no. 1, sep 2021.
- [20] D. T. Campbell, "Common fate, similarity, and other indices of the status of aggregates of persons as social entities," *Behavioral Science*, vol. 3, no. 1, pp. 14–25, feb 1958.
- [21] J. DeLamater, "A definition of "group","" *Small Group Behavior*, vol. 5, no. 1, pp. 30–44, feb 1974.
- [22] S. D. Reicher, "The st. paul's riot: An explanation of the limits of crowd action in terms of a social identity model," *European Journal of Social Psychology*, vol. 14, no. 1, pp. 1–21, jan 1984.
- [23] J. C. Turner, "Social categorization and the self concept: A social cognitive theory of group behavior," in *Advances in Group Process*, E. J. Lawler, Ed. Greenwich, CT: JAI, 1985, vol. 2, pp. 77–122.
- [24] W. Ge, R. T. Collins, and R. B. Ruback, "Vision-based analysis of small groups in pedestrian crowds," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 34, no. 5, pp. 1003–1016, may 2012.
- [25] F. Solera, S. Calderara, and R. Cucchiara, "Socially constrained structural learning for groups detection in crowd," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 38, no. 5, pp. 995–1008, may 2016.
- [26] Z. Qin and C. R. Shelton, "Social grouping for multi-target tracking and head pose estimation in video," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 38, no. 10, pp. 2082–2095, oct 2016.
- [27] M. Li, T. Chen, H. Du, N. Ma, and X. Xi, "Social group detection based on multi-level consistent behaviour characteristics," *Transportmetrica A: Transport Science*, vol. 19, no. 1, oct 2021.
- [28] Y. Feng, D. Duives, W. Daamen, and S. Hoogendoorn, "Data collection methods for studying pedestrian behaviour: A systematic review," *Building and Environment*, vol. 187, p. 107329, jan 2021.
- [29] H. Du, Z. Yu, F. Yi, Z. Wang, Q. Han, and B. Guo, "Recognition of group mobility level and group structure with mobile devices," *IEEE Transactions on Mobile Computing*, vol. 17, no. 4, pp. 884–897, apr 2018.
- [30] C. Zhou, M. Han, Q. Liang, Y.-F. Hu, and S.-G. Kuai, "A social interaction field model accurately identifies static and dynamic social groupings," *Nature Human Behaviour*, vol. 3, no. 8, pp. 847–855, jun 2019.
- [31] D. S. Calovi, A. Litchinko, V. Lecheval, U. Lopez, A. P. Escudero, H. Chaté, C. Sire, and G. Theraulaz, "Disentangling and modeling interactions in fish with burst-and-coast swimming reveal distinct alignment and attraction behaviors," *PLOS Computational Biology*, vol. 14, no. 1, p. e1005933, jan 2018.
- [32] R. Escobedo, V. Lecheval, V. Papaspoulos, F. Bonnet, F. Mondada, C. Sire, and G. Theraulaz, "A data-driven method for reconstructing and modelling social interactions in moving animal groups," *Philosophical Transactions of the Royal Society B: Biological Sciences*, vol. 375, no. 1807, p. 20190380, jul 2020.
- [33] Z. Yücel, F. Zanlungo, T. Ikeda, T. Miyashita, and N. Hagita, "Deciphering the crowd: Modeling and identification of pedestrian group motion," *Sensors*, vol. 13, no. 1, pp. 875–897, jan 2013.
- [34] R. G. Brown and R. F. Meyer, "The fundamental theorem of exponential smoothing," *Operations Research*, vol. 9, no. 5, pp. 673–685, oct 1961.
- [35] A. Li, S. P. Cornelius, Y.-Y. Liu, L. Wang, and A.-L. Barabási, "The fundamental advantages of temporal networks," *Science*, vol. 358, no. 6366, pp. 1042–1046, nov 2017.
- [36] J. James, "The distribution of free-forming small group size," *American Sociological Review*, vol. 18, no. 5, p. 569, oct 1953.
- [37] S. Pellegrini, A. Ess, K. Schindler, and L. van Gool, "You'll never walk alone: Modeling social behavior for multi-target tracking," in *2009 IEEE 12th International Conference on Computer Vision*. IEEE, sep 2009.
- [38] A. Lerner, Y. Chrysanthou, and D. Lischinski, "Crowds by example," *Computer Graphics Forum*, vol. 26, no. 3, pp. 655–664, sep 2007.
- [39] S. Bandini, A. Gorrini, and G. Vizzari, "Towards an integrated approach to crowd analysis and crowd synthesis: A case study and first results," *Pattern Recognition Letters*, vol. 44, pp. 16–29, jul 2014.
- [40] T. Do, M. Haghani, and M. Sarvi, "Group and single pedestrian behavior in crowd dynamics," *Transportation Research Record: Journal of the Transportation Research Board*, vol. 2540, no. 1, pp. 13–19, jan 2016.
- [41] M. Moussaïd, N. Perozo, S. Garnier, D. Helbing, and G. Theraulaz, "The walking behaviour of pedestrian social groups and its impact on crowd dynamics," *PLOS ONE*, vol. 5, no. 4, p. e10047, apr 2010.
- [42] F. Zanlungo, Z. Yücel, D. Bršić, T. Kanda, and N. Hagita, "Intrinsic group behaviour: Dependence of pedestrian dyad dynamics on principal social and personal features," *PLOS ONE*, vol. 12, no. 11, p. e0187253, nov 2017.
- [43] F. Zanlungo, D. Bršić, and T. Kanda, "Spatial-size scaling of pedestrian groups under growing density conditions," *Physical Review E*, vol. 91, no. 6, p. 062810, jun 2015.
- [44] C. McPhail and R. T. Wohlstein, "Using film to analyze pedestrian behavior," *Sociological Methods & Research*, vol. 10, no. 3, pp. 347–375, feb 1982.
- [45] M. Ester, H.-P. Kriegel, J. Sander, and X. Xu, "A density-based algorithm for discovering clusters in large spatial databases with noise," in *Proceedings of the Second International Conference on Knowledge Discovery and Data Mining*. AAAI Press, aug 1996.
- [46] M. H. Zaki and T. Sayed, "Automated analysis of pedestrian group behavior in urban settings," *IEEE Transactions on Intelligent Transportation Systems*, vol. 19, no. 6, pp. 1880–1889, jun 2018.
- [47] F. Yang, J. Shabo, A. Qureshi, and C. Peters, "Do you see groups?" in *Proceedings of the 18th International Conference on Intelligent Virtual Agents*. ACM, nov 2018.

- [48] H. Singh, R. Arter, L. Dodd, P. Langston, E. Lester, and J. Drury, "Modelling subgroup behaviour in crowd dynamics DEM simulation," *Applied Mathematical Modelling*, vol. 33, no. 12, pp. 4408–4423, dec 2009.
- [49] L. You, J. Hu, M. Gu, W. Fan, and H. Zhang, "The simulation and analysis of small group effect in crowd evacuation," *Physics Letters A*, vol. 380, no. 41, pp. 3340–3348, oct 2016.
- [50] D. Helbing, L. Buzna, A. Johansson, and T. Werner, "Self-organized pedestrian crowd dynamics: Experiments, simulations, and design solutions," *Transportation Science*, vol. 39, no. 1, pp. 1–24, feb 2005.
- [51] L. Lu, C.-Y. Chan, J. Wang, and W. Wang, "A study of pedestrian group behaviors in crowd evacuation based on an extended floor field cellular automaton model," *Transportation Research Part C: Emerging Technologies*, vol. 81, pp. 317–329, aug 2017.
- [52] I. Karamouzas and M. Overmars, "Simulating and evaluating the local behavior of small pedestrian groups," *IEEE Transactions on Visualization and Computer Graphics*, vol. 18, no. 3, pp. 394–406, mar 2012.
- [53] M. Moussaïd, D. Helbing, S. Garnier, A. Johansson, M. Combe, and G. Theraulaz, "Experimental study of the behavioural mechanisms underlying self-organization in human crowds," *Proceedings of the Royal Society B: Biological Sciences*, vol. 276, no. 1668, pp. 2755–2762, may 2009.



Maoyin Chen (Member, IEEE) received the B.S. degree in mathematics and the M.S. degree in control theory and control engineering from Qufu Normal University, Shandong, China, in 1997 and 2000, respectively, and the Ph.D. degree in control theory and control engineering from Shanghai Jiao Tong University, Shanghai, China, in 2003.

From 2003 to 2005, he was a Postdoctoral Researcher with the Department of Automation, Tsinghua University, Beijing, China. From 2006 to 2008, he visited Potsdam University, Potsdam, Germany, as an Alexander von Humboldt Research Fellow. Since October 2008, he has been an Associate Professor with the Department of Automation, Tsinghua University, Beijing, China. He has authored and coauthored over 100 peer-reviewed international journal papers. He has won the first prize in natural science (2011, ranked first) and the second prize (2019, ranked first) of CAA. His research interests include fault prognosis and complex systems.



Wenhan Wu received the B.S. degree in School of Automation from Central South University, Changsha, China, in 2019. He is currently pursuing the Ph.D. degree in control science and engineering with the Tsinghua University, Beijing, China. His current research interests include collective behavior, emergency evacuation and pedestrian group dynamics.



Wenfeng Yi received the B.S. degree in School of Astronautics from Beihang University, Beijing, China, in 2019. He is currently pursuing the Ph.D. degree in control science and engineering with the Tsinghua University, Beijing, China. His current research interests include collective behavior, emergency evacuation and pedestrian queuing dynamics.



Xiaoping Zheng received the B.S. degree from the Chengdu University of TCM, Chengdu, China, in 1995, and the Ph.D. degree from Sichuan University, Chengdu, China, in 2003.

From 2004 to 2006, he was a Postdoctoral Researcher in School of Management, Fudan University, Shanghai, China. From 2006 to 2013, he was a Professor with the Institute of Safety Management, Beijing University of Chemical Technology, Beijing, China. He is currently a Professor with the Department of Automation, Tsinghua University, Beijing, China. Prof. Zheng was a 973 Chief Scientist in 2011, a recipient of the National Science Fund for Distinguished Young Scholars in 2012, and a Distinguished Professor of the Chang Jiang Scholars Program in 2021. His current research interests include large-scale crowd evacuation, evolutionary game theory and Terahertz technology.



Jinghai Li received the B.S. degree in automation and the M.S. degree in control science and engineering from Tianjin University, Tianjin, China, in 2009 and 2016, respectively, and the Ph.D. degree in control science and engineering from Tsinghua University, Beijing, China, in 2022. He is currently a Postdoctoral Researcher at the School of Mechanical and Electrical Engineering, Beijing University of Chemical Technology, Beijing, China. His current research interests include crowd dynamics, control of robotic systems, and adaptive systems.