Supplementary Information

7.1 Pedestrian Detection

To detect pedestrians in the videos, we used YOLOv5 [8], pre-trained on the COCO 2017 dataset. The model accurately classified pedestrians in the high-resolution 2010 videos, but it performed less accurately on the 1980 videos due to lower resolution and different coloration. To improve accuracy on the 1980 footage, we fine-tuned the model using 260 sampled frames from the historical videos, labeling 4,792 pedestrians in total. This dataset was split into training, validation, and test sets in a $60\% : 20\% : 20\%$ ratio.

Table S1 compares the model's performance on the same test dataset before and after fine-tuning. Fine-tuning improved the model's average precision (IoU = 0.5) from 0.435 to 0.781, precision from 0.673 to 0.902, and recall from 0.84 to 0.93.

To classify pedestrians in groups and measure walking speed, we converted each detected pedestrian's image coordinates to real-world coordinates. First, we manually selected 20 sample points per location, marking each point's image coordinates and corresponding real-world coordinates using QGIS's Georeferencer tool. Using these reference points, we applied OpenCV's warp perspective function [21] to project each pedestrian's bottom-center position to real-world ground coordinates. This transformation accounted for different camera angles and fields of view, ensuring accurate localization of pedestrians in real space. Figure S1 illustrates this projection process.

Finally, since the 2010 videos were time-lapsed, we used a video interpolation algorithm [22] to increase frames per second, making the 1980s and 2010s videos temporally comparable.

Table S1: Comparison of model performance before and after fine-tuning

	mAP(IoU=0.5)	Precision	Recall
Before fine-tuning	0.435	0.673	0.84
After fine-tuning	0.781	0.902	0.93

Note.Table reports the pedestrian object detection results on video frames from 1980 before and after fine-tuning the original YOLOv5s model. IoU is a measure that quantifies the overlap between a predicted bounding box and a ground truth bounding box. Mean average precision (mAP) computes the average area under the precision-recall curve. Precision quantifies the proportion of true positives among all positive predictions. Recall calculates the proportion of true positives among all actual positives.

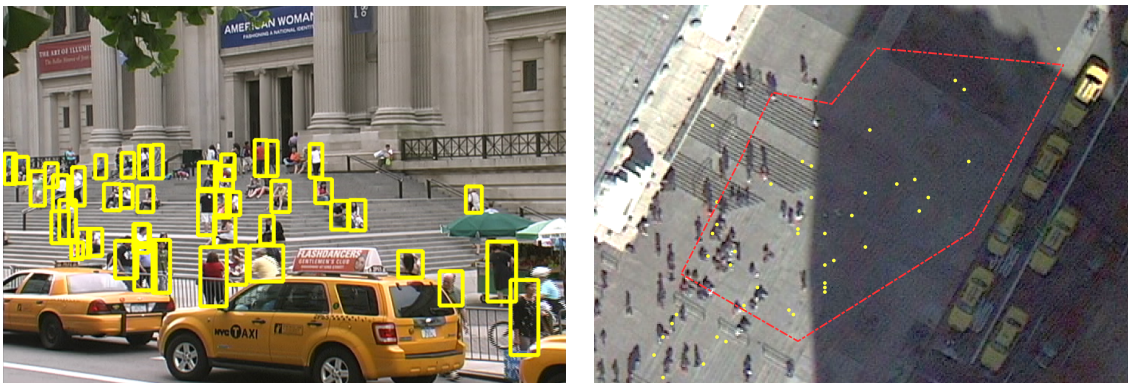

FIGURE S1: Localization of pedestrians in space. Left panel shows the bounding box of each pedestrian detected via the YOLOv5 model[8]. The right panel shows their projected location with yellow points. The red dashed line indicates the observation boundary.

7.2 Group Identification

The group identification process involves four steps:

1) Estimating moving speeds: For each pair of individuals (i, j) identified within the same cluster in a given frame, we estimated their speeds at frame t, denoted as $\vec{V_{it}}$ and $\vec{V_{jt}}$. Here the speed $\|\boldsymbol{v_t}\|$ is calculated as $\frac{Dist(Coords(x_t,y_t),Coords(x_{t-\Delta t},y_{t-\Delta t}))}{\Delta t}$, where $Dist$ represents the great circle distance. We then computed the correlation S_{ijt} between $\vec{V_{it}}$ and $\vec{V_{it}}$ to assess the similarity in movement.

2) Counting co-occurrences in clusters: For each pair of individuals, we counted the number of frames (w) in which they were observed within the same cluster, satisfying the condition of $S_{ijt} > 0$. This ensures that only pairs with correlated movements are considered potential groups.

3) Defining valid groups: To qualify as a group, pairs of pedestrians must appear in the same cluster for at least two seconds, corresponding to $w > 60$ frames for 2010 videos (29.9 fps) and $w > 24$ frames for 1980 videos (11.9 fps). This duration criteria helps distinguish meaningful co-presence from very brief encounters.

4) Classifying pedestrians as moving or lingering: We classified pedestrians as "lingering" if their speed $\|\boldsymbol{v_t}\|$ was below 0.5 at time t. Figure S5 shows the distribution of $\|\boldsymbol{v_t}\|$ across all detected pedestrians.

Fig S2 provides examples of pedestrians detected in groups in both 1980 and 2010, where individuals in groups are marked with colored bounding boxes. Table S2 summarizes the observed group sizes.

To test the impact of varying the minimum required co-presence time (w threshold) on

group identification, we used a series of time thresholds from 0.1 to 10 seconds. Table S3 reports the share of pedestrians observed in groups across sites for each threshold. When the threshold is set to $t = 0.1$ seconds, our are comparable to those of Hampton [23], who identified groups based on single frames. Our method becomes more conservative as w increases, particularly at $t \geq 2$ seconds (equivalent to 60 frames for 2010 and 24 frames for 1980 videos). In most scenarios, singles remain the dominant category, except at the Met, where group identification rates increase with time thresholds lower than 2 seconds.

TABLE S2: Summary statistics on group size distribution

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Chestnut Street		Bryant Park		Downtown Crossing		MET	
	2010s	1980s	2010s	1980s	2010s	1980s	2010s	1980s
Singles(%)	0.79	0.79	0.67	0.82	0.77	0.59	0.48	0.49
	[0.30]	[0.39]	[0.16]	[0.27]	[0.18]	[0.15]	[0.15]	[0.24]
Dyads(%)	0.20	0.20	0.29	0.14	0.20	0.35	0.38	0.42
	[0.30]	[0.39]	[0.15]	[0.24]	[0.17]	[0.14]	[0.14]	[0.23]
Triads(%)	0.01	0.01	0.03	0.03	0.02	0.04	0.08	0.06
	[0.08]	[0.09]	[0.06]	[0.12]	[0.07]	[0.07]	[0.09]	[0.12]
Others(%)	0.00	0.00	0.01	0.01	0.01	0.01	0.05	0.03
	[0.02]	[0.00]	[0.05]	[0.08]	[0.05]	[0.04]	[0.08]	[0.09]
Observations	344081	9346	135785	42570	122476	30450	61691	74250

Note.—Table reports pedestrians within the overlapping study areas between the two decades. Pedestrians in groups is the framed-based measure, considering people can participate or leave a group at different points in time. Standard deviation in square brackets. Others refer to people in large groups (at least 4 people in one group).

TABLE S3: Testing different time thresholds to define groups

Location	Time	Singles	Dyads	Triads	Others	Threshold (s)
Bryant Park	1980s	78.6%	15.9%	4.0%	1.5%	0.1
Bryant Park	2010s	57.7%	33.3%	5.3%	3.8%	0.1
Chestnut Street	1980s	76.4%	22.5%	1.1%	0.0%	0.1
Chestnut Street	2010s	77.1%	21.6%	1.2%	0.1%	0.1
Downtown Crossing	1980s	53.1%	37.6%	6.2%	3.2%	0.1
Downtown Crossing	2010s	63.8%	27.0%	5.4%	3.8%	0.1
MET	1980s	42.7%	43.8%	8.2%	5.3%	0.1
MET	2010s	34.9%	42.2%	12.1%	10.8%	0.1
Bryant Park	1980s	79.9%	15.5%	3.3%	1.3%	1
Bryant Park	2010s	62.0%	31.9%	3.9%	2.1%	1
Chestnut Street	1980s	76.7%	22.2%	1.1%	0.0%	1

TABLE S3: Testing different time thresholds to define groups

Location	Time	Singles	Dyads	Triads	Others	Threshold (s)
Chestnut Street	2010s	77.5%	21.3%	1.1%	0.1%	1
Downtown Crossing	1980s	56.3%	36.5%	5.1%	2.1%	1
Downtown Crossing	2010s	69.3%	25.3%	3.6%	1.7%	1
MET	1980s	45.8%	43.1%	7.2%	3.9%	1
MET	2010s	42.1%	40.4%	10.0%	7.6%	1
Bryant Park	1980s	82.0%	14.2%	2.8%	0.9%	2
Bryant Park	2010s	67.3%	28.7%	2.7%	1.2%	2
Chestnut Street	1980s	79.3%	19.8%	0.8%	0.0%	2
Chestnut Street	2010s	79.0%	20.0%	0.9%	0.1%	2
Downtown Crossing	1980s	59.3%	35.1%	4.1%	1.5%	2
Downtown Crossing	2010s	77.0%	20.5%	1.9%	0.6%	2
MET	1980s	48.6%	41.8%	6.4%	3.2%	2
MET	2010s	48.3%	37.9%	8.4%	5.4%	2
Bryant Park	1980s	80.8%	15.7%	3.1%	0.4%	4
Bryant Park	2010s	70.6%	26.5%	2.0%	0.9%	4
Chestnut Street	1980s	74.9%	23.6%	1.5%	0.0%	4
Chestnut Street	2010s	80.0%	19.3%	0.6%	0.0%	4
Downtown Crossing	1980s	61.3%	34.8%	3.0%	0.9%	4
Downtown Crossing	2010s	82.3%	16.4%	1.1%	0.2%	4
MET	1980s	48.6%	43.2%	5.9%	2.4%	4
MET	2010s	53.5%	35.5%	7.1%	3.9%	4
Bryant Park	1980s	82.5%	14.8%	2.7%	0.0%	6
Bryant Park	2010s	72.7%	25.0%	1.6%	0.7%	6
Chestnut Street	1980s	81.7%	16.4%	1.9%	0.0%	6
Chestnut Street	2010s	80.7%	18.6%	0.6%	0.0%	6
Downtown Crossing	1980s	63.9%	33.2%	2.3%	0.5%	6
Downtown Crossing	2010s	86.2%	13.0%	0.6%	0.2%	6
MET	1980s	50.1%	42.5%	5.5%	1.9%	6
MET	2010s	58.3%	32.5%	6.1%	3.1%	6
Bryant Park	1980s	83.8%	14.4%	1.7%	0.0%	8
Bryant Park	2010s	75.0%	23.6%	1.0%	0.4%	8
Chestnut Street	1980s	86.8%	13.2%	0.0%	0.0%	8

TABLE S3: Testing different time thresholds to define groups

Location	Time	Singles	Dyads	Triads	Others	Threshold (s)
Chestnut Street	2010s	81.8%	17.9%	0.3%	0.0%	8
Downtown Crossing	1980s	67.3%	30.8%	1.6%	0.2%	8
Downtown Crossing	2010s	89.6%	9.9%	0.5%	0.0%	8
MET	1980s	51.2%	42.0%	5.1%	1.7%	8
MET	2010s	62.2%	30.1%	5.6%	2.1%	8
Bryant Park	1980s	84.4%	14.3%	1.3%	0.0%	10
Bryant Park	2010s	76.5%	22.4%	0.8%	0.3%	10
Chestnut Street	1980s	93.1%	6.9%	0.0%	0.0%	10
Chestnut Street	2010s	82.3%	17.4%	0.3%	0.0%	10
Downtown Crossing	1980s	68.8%	29.6%	1.5%	0.1%	10
Downtown Crossing	2010s	92.2%	7.8%	0.0%	0.0%	10
MET	1980s	53.4%	40.8%	4.4%	1.5%	10
MET	2010s	64.9%	28.6%	4.8%	1.7%	10

Note.—Table reports the mean percentage of pedestrians observed in one of the group types using different time thresholds to define groups.

We validated our group detection method by comparing our automated results to manual coding. Table S4 shows the validation metrics, including precision, recall, F1 score, and accuracy, defined as: precision $= TP/(TP + FP)$; recall $= TP/(TP + FN)$; F1 $= 2 \times$ precision $\times$ recall/(precision + recall); accuracy $= (TP + TN)/(TP + FN + TN + FP)$.

Our method achieves accuracy comparable to previous works in a similar setting[18, 24, 25], with an accuracy rate of 89%. We observed higher accuracy in the 1980 videos, likely due to the elevated camera angles in the 1980 footage, which make distances between pedestrians slightly more distinguishable.

7.3 Group Encounters

The frequency of group encounters depends on the selected time threshold used to define encounter. We tested the robustness of this threshold by varying the elapsed time from 1 to 10 seconds and observed how it affected the percentage of encounters. As shown in Table S5, the percentage of detected encounters decreases as the time threshold increases. However, we consistently observed a higher share of group encounters in 1980 compared to 2010 across all thresholds, indicating a general decline in spontaneous interactions over time regardless

TABLE S4: Verify the group detection method with randomly selected frames

	#Frames	True Positive (TP)	False Positive (FP)	True Negative (TN)	False Negative (FN)	Precision	Recall	F1	Accuracy
2010s	40	269	33	284	69	0.891	0.796	0.841	84.43%
1980s	40	246	30	691	73	0.891	0.771	0.827	90.10%
Overall	80	515	63	975	142	0.891	0.784	0.834	87.91%

Note.—Table shows the results of the validation of the group detection algorithm using a manual coder. We randomly selected 80 frames from each location. After applying the algorithm to the images, we color-coded the pedestrians in groups. The researcher reported the following: (1) whether the selected pedestrians identified as being in groups are true positives (TP) or false positives (FP), (2) the number of pedestrians who are also in groups but were not identified by the algorithm, which will be labeled as false negatives (FN), and (3) the remaining pedestrians, which are categorized as true negatives (TN).

FIGURE S2: Examples of people in groups detected in 1980 and 2010. Only people in groups are colored with bounding boxes.

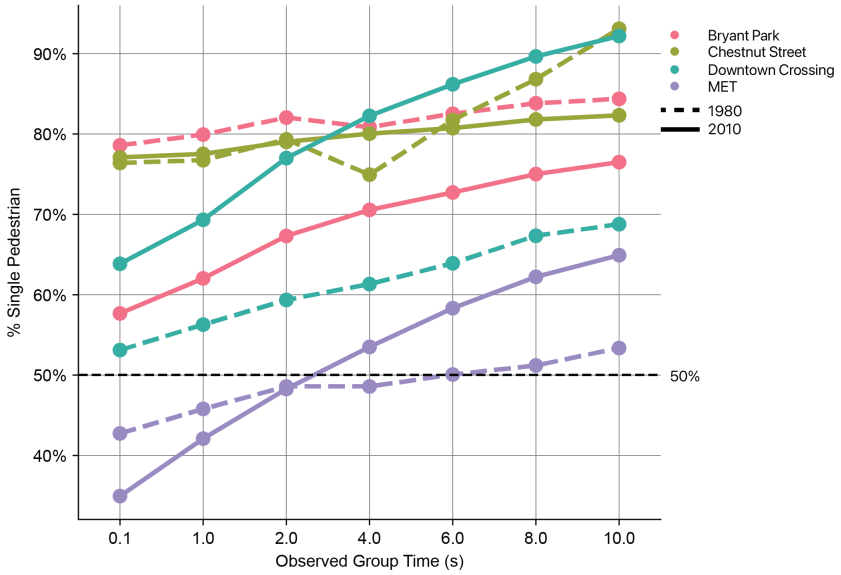

FIGURE S3: Robustness test on group time selection. The figure shows the percentage of single pedestrians as a proportion of all pedestrians per frame across four sites. The horizontal dashed line indicates that 50% of pedestrians are observed as singles.

of the chosen threshold.

TABLE S5: Group encounters using different time thresholds

Threshold (s)	1980s	2010s	2010s-1980s
2	8.78%	5.21%	-3.58%
3	7.44%	3.85%	-3.59%
4	6.03%	2.88%	-3.15%
5	5.40%	2.20%	-3.20%
6	4.62%	1.81%	-2.82%
7	4.14%	1.51%	-2.63%
8	3.68%	1.28%	-2.40%
9	3.34%	1.09%	-2.25%
10	2.89%	0.94%	-1.94%

Note.—Table presents the estimation of group encounters based on various time thresholds. These thresholds indicate the duration from when pedestrians are first observed in the video until a group is formed. A larger threshold indicates that pedestrians remain as individuals for a longer period before forming a group or joining an existing one.

FIGURE S4: Example of detected encounters. Images from left to right show frames captured every 1.5 seconds. The entire sequence illustrates an encounter lasting around 6.5 seconds as two pedestrians approach and meet.

Figure S4 provides an example of an encounter classified using our methodology. The sequence shows two pedestrians approaching each other over a span of approximately 6.5 seconds, with frames captured at intervals of 1.5 seconds.

7.4 Lingerers and Movers

We classified pedestrians as either stationary (lingerers) or moving (movers) based on their speed, calculated from frame-by-frame movement. Pedestrians with speeds below 0.5 m/s were identified as lingerers, while those moving faster were classified as movers.

Figure S5a presents a histogram of pedestrian speeds pooled across all locations. Figure S5b displays the results of a robustness test, where we varied the observation duration to ensure the reliability of the 0.5 m/s threshold. The test confirms that our classification method consistently identifies lingerers across different observation periods.

Figure S5c presents examples of detected lingerers from both 1980 (bottom) and 2010 (top). Only pedestrians within the observation boundaries are annotated.

Table S6 summarizes the percentage of people lingering alone or in groups. The term "% Pedestrian alone lingering" represents the percentage of single pedestrians observed lingering at a site, while "% Pedestrian group lingering" represents the percentage of pedestrians in groups observed lingering at the site.

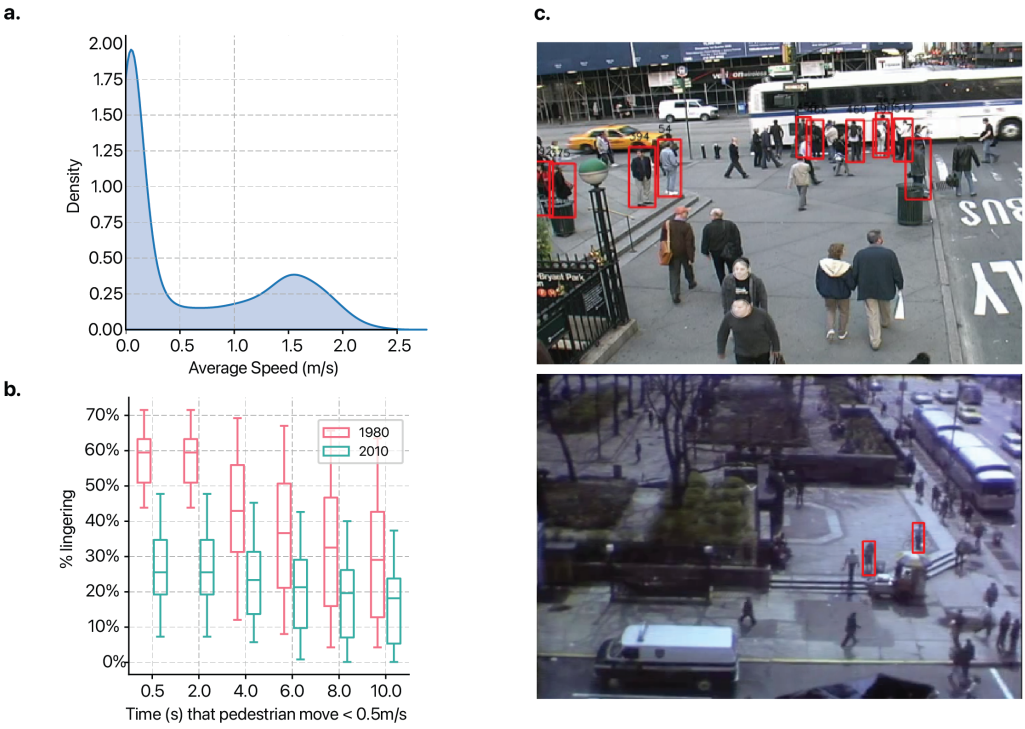

FIGURE S5: a. histogram of speeds pooling all locations. b. The robustness test was conducted by considering the observed duration of pedestrian movement at a speed slower than 0.5 m/s. c. two examples of lingering pedestrians detected via moving speed in 2010 (top) and 1980 (bottom). Only pedestrians within the observation boundaries are annotated.

TABLE S6: Lingering in groups or alone by location

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Chestnut Street		Bryant Park		Downtown Crossing		MET	
	2010s	1980s	2010s	1980s	2010s	1980s	2010s	1980s
# Pedestrian	3.09	1.38	17.21	3.78	10.32	20.67	28.87	11.10
	[1.35]	[0.63]	[5.51]	[2.51]	[2.70]	[6.79]	[7.50]	[6.37]
# Ped. lingering	0.64	0.37	5.50	1.46	1.46	11.26	11.85	6.51
	[0.75]	[0.59]	[3.04]	[1.43]	[1.29]	[4.84]	[4.52]	[4.68]
% Ped. lingering	20.61	24.59	32.09	37.53	13.96	54.23	40.66	56.54
	[25.97]	[39.04]	[14.89]	[31.98]	[12.02]	[14.65]	[12.18]	[22.27]
% Ped. lingering alone	19.92	20.95	29.90	33.99	12.55	51.74	34.38	49.37
	[28.57]	[39.10]	[16.33]	[34.67]	[14.66]	[17.99]	[18.48]	[31.34]
% Ped. lingering in groups	23.54	36.56	34.87	50.14	16.48	56.84	43.92	61.37
	[36.26]	[37.53]	[26.19]	[35.26]	[22.76]	[22.02]	[14.61]	[26.20]
Frames	344081	9346	135785	42570	122476	30450	61691	74250

Note.—Table reports the summary statistics of the pedestrian staying. Only the pedestrians observed in the video for more than 2 seconds are included in the study. Pedestrians are classified as lingering if they have an observed speed below 0.5 m/s for at least t seconds. This table reports the result at t=4. Figure S4 compares the impact of different selections of t.