1 PROBLEM DEFINITION

We first consider the urban trajectory defined by a continuous mobility model. During a time period T, the user trajectory Γ is composed of a set of temporally continuous records by $\Gamma = \bigcup_{t\in T} < t, \ell(t) >$. $\ell(t)$ denotes the location of the user at time t.

DEFINITION 1. Continuous mobility of a trajectory – *On the continuous segment of the trajectory Γ during a time period $\tau \subseteq T$, denoted as $\gamma = \bigcup_{t\in\tau} < t, \ell(t) >$, we define its mobility by:*

*(a) γ is a **stay** segment if: $|\tau| \geq \Delta T$ and $||\ell(t_i) - \ell(t_j)|| < \Delta S$ ($\forall t_i, t_j \in \tau$);*

*(b) γ is a **travel** segment if: γ does not overlap with any continuous segment satisfying (a).*

Here $|\cdot|$ denotes the length of a time period, $||\cdot||$ is the L_2 norm that computes the spatial distance between two records. ΔT and ΔS are the temporal and spatial parameters in the mobility definition.

As shown by the red curve in Figure 1(a) where the hollow node stands for a long-time stop, Definition 1(a) models the stay segment as a sufficiently long time period ($\geq \Delta T$) when the trajectory is kept within a circular region of radius $\Delta S/2$. This definition is consistent among all the previous literature [?][?][?]. Note that the stay segments by definition can overlap with each other in space and time. Their enclosure is called the maximal stay segment. On the other hand, based on the ground truth that a user can either stay or travel in any time point, the segment not overlapped with any stays is defined as the travel segment (Definition 1(b)).

The continuous mobility model can not be exactly computed in the real world application, as the human trajectory is hardly measured continuously. In most cases, the trajectory is composed of a list of discrete records on certain time points (e.g., $t_1 < \cdots < t_L$): $\Gamma = \bigcup_{t\in\{t_1,\cdots,t_L\}} < t, \ell(t) >$. A discrete mobility model can be defined in analogy to the continuous model.

DEFINITION 2. Discrete mobility of a trajectory – *On the discrete segment of the trajectory Γ in a time series $\omega = \{t_p, \cdots, t_q\}$ ($1 \leq p < q \leq L$), denoted as $\gamma = \bigcup_{t\in\omega} < t, \ell(t) >$, define its mobility by:*

(a) γ is stay if: $t_q - t_p \geq \Delta T$ and $||\ell(t_i) - \ell(t_j)|| < \Delta S$ ($\forall t_i, t_j \in \omega$);

(b) γ is travel if: γ does not overlap with any discrete segment satisfying (a).

The discrete mobility model can be optimally computed by an exact algorithm (Algorithm 1). Nevertheless, the resulting mobility is not always equivalent to that of the continuous model with the full trajectory information. For example, in Figure 1(b), the stay and travel segments detected on a densely sampled trajectory by the discrete mobility model generally echo those by the continuous model (Figure 1(a)). In comparison, the detected segments shown by Figure 1(c) on the same but sparsely sampled trajectory turn out to be erroneous and largely deviate from the continuous model. We discover a theorem that reveals the relationship of the two models.

THEOREM 1. Intrinsic linkage between discrete and continuous mobility of a trajectory – *Consider a discrete segment γ of the trajectory. Let ϵ be the maximal time interval between the consecutive records of γ, v_{max} be the maximal movement speed in γ:*

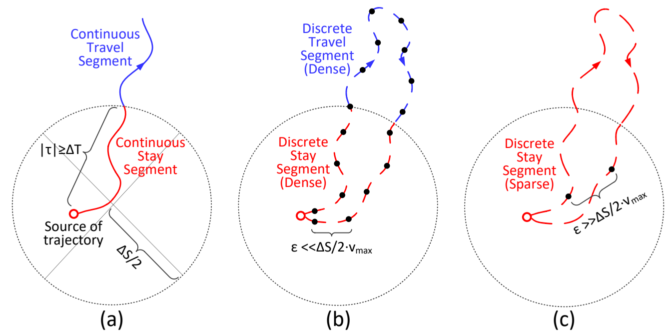

Figure 1: Illustrative examples of Definition 1 and Definition 2: (a) continuous stay/travel segments; discrete stay/travel segments on (b) dense trajectory; (c) sparse trajectory.

Algorithm 1: The exact algorithm on dense trajectories.

Input : $\Gamma = \bigcup_{i\in[1,L]} < t_i, \ell(t_i) >, t_1 < \cdots < t_L$ (dense trajectory), ΔT, ΔS (the space and time parameters)
Output : $I_{S/T}(t_i), \forall i \in [1, L]$ (the mobility of each record)

begin
  for $head \leftarrow [1, L-1]$ **do**
    for $cursor \leftarrow [head+1, L]$ **do**
      /* iterate all the candidate stay segments */
      if $t_{cursor} - t_{head} \geq \Delta T$ **then**
        for $i \leftarrow [head, cursor-1]$ **do**
          for $j \leftarrow [i+1, cursor]$ **do**
            if $||\ell(t_i) - \ell(t_j)|| \geq \Delta S$ **then**
              $Stay \leftarrow$ False, **Break**
        if $Stay! = False$ **then**
          for $i \leftarrow [head, cursor]$ **do**
            $I_{S/T}(t_i) \leftarrow$ S
  /* the remaining records are travel trips */
  for $i \leftarrow [1, L]$ **do**
    if $I_{S/T}(t_i)! = S$ **then**
      $I_{S/T}(t_i) \leftarrow$ T
  return $I_{S/T}(t_i), i = [1, L]$

(a) γ satisfying Definition 2(a) under the parameters of ΔS and ΔT is also a stay segment by Definition 1(a) in the continuous model under the parameters of $\Delta' S = \Delta S + 2\cdot\epsilon\cdot v_{max}$ and $\Delta' T = \Delta T$;

(b) γ satisfying Definition 2(b) under the parameters of ΔS and ΔT is also a travel segment by Definition 1(b) in the continuous model under the parameters of $\Delta' S = \Delta S$ and $\Delta' T = \Delta T + 2\cdot\epsilon$.

The proof is given in Appendix A. By Theorem 1, for the discrete trajectory satisfying $\epsilon << \min(\frac{\Delta S}{2\cdot v_{max}}, \frac{\Delta T}{2})$, i.e., having a dense sampling rate, the discrete mobility of the trajectory computed by the exact algorithm can approximate its continuous mobility with tiny parameter changes. Unfortunately, the measurement of human trajectories in big cities is often extremely sparse over time for pragmatic constraints such as the power consumption and the user privacy

This work studies the inference of the continuous mobility from the sparse trajectory, which can not be approximated by Theorem 1.

MOBILITY INFERENCE PROBLEM ON SPARSE TRAJECTORIES
Given: (1) *a set of urban users*; (2) *each user's sparse trajectory* $\Gamma = \bigcup_{i\in[1,L]} < t_i, \ell(t_i) >$ *that* $\exists j \in [1, L), ||t_j - t_{j+1}|| > \min(\frac{\Delta S}{2\cdot v_{max}}, \frac{\Delta T}{2})$; (3) *the parameters of* ΔS *and* ΔT *that define the mobility of the trajectory*.
Infer: *the continuous mobility of the sparse trajectory at the time of each record, which is denoted by* $I_{S/T}(t_i), \forall i \in [1, L]$.

Note that the parameters of ΔS and ΔT determine the spatiotemporal scale of the mobility. Unless otherwise noted, we use $\Delta S = 800\ m$ and $\Delta T = 30\ min$ to study the intra-city mobility. The selection process of these parameters is detailed in Appendix B.

2 SINGLE TRAJECTORY INFERENCE

We propose the mobility inference algorithm on the single trajectory. The main idea is to leverage the long-tailed sparsity pattern discovered in our trajectory data. Though the average record interval in a trajectory is too large to apply Theorem 1, each trajectory can be decomposed into multiple densely sampled segments, on which the continuous mobility can be confidently inferred.

DEFINITION 3. Dense stay segment – *Any discrete segment γ of the trajectory Γ defined in the time series $\omega = \{t_p, \cdots, t_q\}$ $(1 \le p < q \le L)$ is a **dense stay segment** of Γ if:*

(a) γ is a stay segment of Γ by Definition 2(a);

(b) any consecutive time interval of γ is small enough: $\forall p \le i < q,\ t_{i+1} - t_i \le \Delta T$. ΔT is the parameter used in Definition 2(a).

OBSERVATION 1. Continuous stay assumption – *Consider a dense stay segment γ detected from the long-tailed sparse trajectory, which is defined in the time series $\omega = \{t_p, \cdots, t_q\}$ $(1 \le p < q \le L)$. For any unobserved time point $t \in (t_i, t_{i+1}), \forall p \le i < q$, we hypothesize that $||\ell(t) - \ell(t_i)|| < \Delta S$ and $||\ell(t) - \ell(t_{i+1})|| < \Delta S$.*

Observation 1 states that if a user is observed frequently in a region of diameter ΔS, any intermediate location between observations is also within a similarly constrained region. We empirically validate this observation by the experiment in Appendix B on our trajectory data set. The probability of violating the observation is below 10^{-5} in most cases. When Observation 1 holds, we can develop two theorems that characterize the continuous mobility of stay and travel on long-tailed sparse trajectories.

THEOREM 2. Continuous mobility of dense stay segments – *In the long-tailed sparse trajectory Γ:*

(a) any dense stay segment γ satisfying Definition 3 under the parameters of $\Delta S/3$ and ΔT is also the stay segment by Definition 1(a) in the continuous model under the parameters of ΔS and ΔT;

(b) the continuous mobility of any discrete segment γ in the time period $\tau \in [t_p, t_q]$ can be inferred as stay by Definition 1(a) under the parameters of ΔS and ΔT only if γ defined in $\omega = \{t_p, \cdots, t_q\}$ is the dense stay segment under the same parameters.

THEOREM 3. Continuous mobility of travel records – *Consider a discrete trajectory Γ defined in the time series $\omega = \{t_1, \cdots, t_L\}$:*

(a) any record at time t_i $(1 < i < L)$ is in the travel segment by the continuous model of Definition 1(b) under the parameters of ΔS and ΔT if only there exist $1 \le p < i < q \le L$ that: 1) $||\ell(t_i) - \ell(t_p)|| \ge \Delta S$; 2) $||\ell(t_i) - \ell(t_q)|| \ge \Delta S$; 3) $t_q - t_p \le \Delta T$;

(b) any record at time t_i can be inferred as in the travel segment by Definition 1(b) under the parameters of ΔS and ΔT only if there exist $1 \le p < i < q \le L$ that: 1) $||\ell(t_i) - \ell(t_p)|| \ge \Delta S/2$; 2) $||\ell(t_i) - \ell(t_q)|| \ge \Delta S/2$; 3) $t_q - t_p \le \Delta T$.

The proofs are given in Appendix A.

A PROOFS AND THE EXACT ALGORITHM

THEOREM 1: INTRINSIC LINKAGE BETWEEN DISCRETE AND CONTINUOUS MOBILITY OF A TRAJECTORY.
Proof. Theorem 1(a). For the discrete stay segment γ in the time series $\omega = \{t_p, \cdots, t_q\}$ $(t_q - t_p \ge \Delta T)$, consider its corresponding continuous segment γ' in the time period $\tau = [t_p, t_q]$. γ' satisfies $|\tau| = t_q - t_p \ge \Delta T$. $\forall t_i, t_j \in \tau$, we have $||\ell(t_i) - \ell(t_j)|| \le ||\ell(t_i) - \ell(t'_i)|| + ||\ell(t'_i) - \ell(t'_j)|| + ||\ell(t'_j) - \ell(t_j)||$, given that the straightline is the shortest distance between $\ell(t_i)$ and $\ell(t_j)$. Here t'_i and t'_j are the closest time point in ω to t_i and t_j respectively. Because $||\ell(t'_i) - \ell(t'_j)|| < \Delta S$, $||\ell(t_i) - \ell(t'_i)|| \le \epsilon \cdot v_{max}$, $||\ell(t'_j) - \ell(t_j)|| \le \epsilon \cdot v_{max}$, we have $||\ell(t_i) - \ell(t_j)|| < \Delta S + 2 \cdot \epsilon \cdot v_{max}$. That is, γ' is a stay segment by Definition 1(a) under the parameters of $\Delta' S = \Delta S + 2 \cdot \epsilon \cdot v_{max}$ and $\Delta' T = \Delta T$.

Theorem 1(b). For the discrete travel trip γ in the time series $\omega = \{t_p, \cdots, t_q\}$, by definition, we have $\forall t'_i, t'_j \in \omega$ $(t'_j - t'_i \ge \Delta T)$, there exist two time points $t'_m, t'_n \in \omega$ $(t'_i \le t'_m < t'_n \le t'_j)$ satisfying $||\ell(t'_m) - \ell(t'_n)|| \ge \Delta S$. Consider the corresponding continuous segment γ' in the time period $\tau = [t_p, t_q]$, $\forall t_i, t_j \in \tau$ $(t_j - t_i \ge \Delta T + 2 \cdot \epsilon)$, we can find t'_i (the closest time point in ω no smaller than t_i) and t'_j (the closest time point in ω no larger than t_j), having $t'_j - t'_i \ge \Delta T$. There exist two time points $t'_m, t'_n \in \omega$ $(t_i \le t'_i \le t'_m < t'_n \le t'_j \le t_j)$ satisfying $||\ell(t'_m) - \ell(t'_n)|| \ge \Delta S$. That is, γ' is a travel trip by Definition 1(b) under the parameters of $\Delta' S = \Delta S$ and $\Delta' T = \Delta T + 2 \cdot \epsilon$. ■

THEOREM 2: CONTINUOUS MOBILITY OF DENSE STAY SEGMENTS.
Proof. Theorem 2(a). For the dense stay segment γ defined in the time series $\omega = \{t_p, \cdots, t_q\}$, consider its corresponding continuous segment γ' in the time period $\tau = [t_p, t_q]$. We have $|\tau| = t_q - t_p \ge \Delta T$ because γ is the dense stay segment. For any two time points $t, t' \in [t_p, t_q]$, denote the closest time points in the time series of ω to t and t' as t_i and t_j $(p \le i \le q,\ p \le j \le q)$. We have $||\ell(t) - \ell(t')|| \le ||\ell(t) - \ell(t_i)|| + ||\ell(t_i) - \ell(t_j)|| + ||\ell(t_j) - \ell(t')|| < \Delta S/3 + \Delta S/3 + \Delta S/3 = \Delta S$ by Observation 1. The conditions for the continuous model of the stay segment in Definition 1(a) then hold.

Theorem 2(b). For the discrete segment γ defined in the time series $\omega = \{t_p, \cdots, t_q\}$, consider its corresponding continuous segment γ' in the time period $\tau = [t_p, t_q]$. If $\exists p \le i < q,\ t_{i+1} - t_i > \Delta T$, i.e., the unobserved time period of (t_i, t_{i+1}) has a duration longer than ΔT. Observing a time period $\tau' \subset (t_i, t_{i+1})$ with $|\tau'| = \Delta T$ can detect a different stay segment from the other part of the segment in (t_i, t_{i+1}). Then there can be travel trips surrounding the segment in τ' to connect the trajectory. This possibility can not be validated or

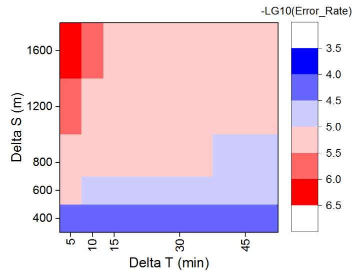

Figure 2: The probability for violating Observation 1, under different ΔS and ΔT, mapped by the $-\lg_{10}$ operator.

rejected given the information of the discrete segment γ only. Therefore, the corresponding continuous segment γ' can not be inferred as stays, unless $\forall p \le i < q,\ t_{i+1} - t_i \le \Delta T$.

On the dense segment γ, if the corresponding continuous segment γ' is the stay segment, by Definition 1(a), $\forall t_i, t_j \in \omega \subset \tau$, $||\ell(t_i) - \ell(t_j)|| < \Delta S$. Therefore, γ must be a dense stay segment. ■

THEOREM 3: CONTINUOUS MOBILITY OF TRAVEL RECORDS.

Proof. Theorem 3(a). For the record at time t_i, consider any time period $\tau = [t, t']$ satisfying $|\tau| \ge \Delta T$ and $t_i \in \tau$. If the three conditions hold, the time period of $\tau' = [t_p, t_q]$ satisfies $|\tau'| \le \Delta T$ and $t_i \in \tau'$. We have $t_p \in \tau$ or $t_q \in \tau$. Otherwise, we will have $t_p < t$ and $t_q > t'$, which leads to the contradiction of $|\tau'| = t_q - t_p > t' - t = |\tau| \ge \Delta T$. For the either case of $t_p \in \tau$ or $t_q \in \tau$, we have $||\ell(t_i) - \ell(t_p)|| \ge \Delta S$ and $||\ell(t_i) - \ell(t_q)|| \ge \Delta S$. This contradicts to Definition 1(a). Therefore, the record at time t_i can not be in any stay segment, and it must be in a travel trip by Definition 1(b).

Theorem 3(b). For the record at time t_i $(1 < i < L)$, if the condition does not hold, $\forall 1 \le p < i < q \le L$ satisfying $t_q - t_p \le \Delta T$, we have $||\ell(t_i) - \ell(t_p)|| < \Delta S/2$ or $||\ell(t_i) - \ell(t_q)|| < \Delta S/2$.

Consider the smallest time point t_j satisfying $t_j > t_i$ and $||\ell(t_i) - \ell(t_j)|| \ge \Delta S/2$. We should have $t_j - t_i \le \Delta T$ because otherwise $\forall t_i < t_k < t_j$, $||\ell(t_i) - \ell(t_k)|| < \Delta S/2$. There exists a time period of $\tau' = [t_i, t_j)$, for all observed $t_k \in \tau'$, $||\ell(t_i) - \ell(t_k)|| < \Delta S/2$. Then $\forall t_k, t_{k'} \in \tau'$, $||\ell(t_k) - \ell(t_{k'})|| \le ||\ell(t_i) - \ell(t_k)|| + ||\ell(t_i) - \ell(t_{k'})|| < \Delta S$, t_i will be possibly in a stay segment, without the information to reject the possibility.

Having $t_j - t_i \le \Delta T$ and $||\ell(t_i) - \ell(t_j)|| \ge \Delta S/2$, using the proof by contradiction, we have $\forall k < i$ satisfying $t_j - t_k \le \Delta T$, we have $||\ell(t_i) - \ell(t_k)|| < \Delta S/2$. Consider the largest time point $t_{j'}$ satisfying $t_j - t_{j'} > \Delta T$, we can construct a time period of $(t_{j'}, t_j)$, for all the observed time point of t_k having $j' < k < j$, we have $||\ell(t_i) - \ell(t_k)|| < \Delta S/2$. The distance between these observed time points is below ΔS. Then there can be a continuous segment in the time period of $\tau' \subset (t_{j'}, t_j)$ satisfying $|\tau'| = \Delta T$. We do not have any information to reject the inference of stays on this segment. Therefore, the record at $t_i \in \tau'$ can not be in any travel trip. ■

B MATERIAL FOR THE SDS ALGORITHM

To validate Observation 1, we conduct an experiment on the full data set in Beijing. For each record in a trajectory, we explicitly remove the record and detect all the dense stay segments from the remaining trajectory. If the record is within a dense stay segment, we check whether the record, as time t, violates Observation 1. As shown in Figure 2, among 10-billion potential <record, interval> pairs for each parameter setting, the probability of violating Observation 1 is below 10^{-5} if $\Delta S \ge 800m$ and $\Delta T \le 30min$.