ORTEGA: an object-oriented time-geographic analytical approach to trace space-time contact patterns in movement data

Somayeh Dodge[a,*], Rongxiang Su[a], Jasper Johnson[b], Achara Simcharoen[c], Konstadinos Goulias[a], James L.D. Smith[d], Sean C Ahearn[e]

[a]*Department of Geography, University of California Santa Barbara, USA*
[b]*Department of Geography, University of Minnesota, Twin Cities, USA*
[c]*Conservation Ecology Program, King Mongkut's University of Technology, Thailand*
[d]*Department of Fisheries, Wildlife & Conservation Biology, University of Minnesota, Twin Cities, USA*
[e]*Hunter College – CUNY, New York City*

Abstract

This paper uses movement as a marker to study interactions in humans and animals to better understand their collective behaviors. Interaction is an important driving force in social and ecological systems. It can also play a significant role in the transmission of infectious diseases and viruses as witnessed during the ongoing COVID-19 pandemic. Although a number of approaches have been developed to analyze interaction using movement data sets, these methods mainly capture concurrent and dyadic interaction (i.e. when two individuals have direct contact or move synchronously in the spatial proximity of each other). Less work has been done on tracing interaction between multi-

[*]Corresponding author: S. Dodge sdodge@ucsb.edu

Email addresses: sdodge@ucsb.edu (Somayeh Dodge), rongxiangsu@ucsb.edu (Rongxiang Su), joh10891@umn.edu (Jasper Johnson), simtom@windowslive.com (Achara Simcharoen), kgoulias@ucsb.edu (Konstadinos Goulias), smith017@umn.edu (James L.D. Smith), sahearn@hunter.cuny.edu (Sean C Ahearn)

URL: https://somayehdodge.info/ (Somayeh Dodge)

ple individuals, especially when the interaction occurs with a delay or via indirect contact (i.e. when individuals visit the same location asynchronously). This paper introduces a new Object-oRiented Time-Geographic Analytical approach (ORTEGA) to extract concurrent and delayed interaction patterns between individuals in space and time. The method leverages the time-geography framework to incorporate the effects of uncertainty and gaps in movement data in the analysis of interaction and tracing contact patterns. Using two different case studies and real GPS tracking data, the method is evaluated in (1) detecting patterns of dyadic, intra and interspecific interactions between two apex predators, tigers and leopards in Thailand; and (2) tracing potential contacts between a large group of individuals of the same and different households in San Jose, California. The results indicate that tigers and leopards have an awareness of each other and their interaction is mainly indirect and delayed. In the human context, the results show that while individuals of the same household have more concurrent interaction, members of different households follow similar patterns asynchronously exhibiting delayed interaction. The delayed interactions and potential asynchronous contacts are often underestimated by the common digital contact tracing technologies. With this study we show how a generic method can be used to identify interesting movement patterns across the human and animal divide.

Keywords: Time geography, contact tracing, interaction analysis, wildlife encounter analysis, delayed interaction, tiger leopard interaction

1. Introduction

Interaction and contact between individuals are important driving factors of many social and ecological systems. Collective behaviors of animals and humans result in complex social dynamics which can be observed through movement of individuals (Potts et al., 2014; Laube, 2014; Dodge et al., 2008). Movement patterns shape urban and natural ecosystem dynamics, structure human and wildlife social networks, and are essential to understanding human and wildlife interactions. There have been advances in the analysis of spatial interactions, but less research has focused on the temporal aspects of interaction between moving individuals. Specifically, approaches to analyzing temporal delays in spatial proximity are lacking. These delayed interactions are critical in virus transmissions and exposure to airborne pathogens. As the decade turned to the 2020s, we witnessed the widespread transmission of Coronavirus disease 2019 (COVID-19) which led to a set of unprecedented non-pharmaceutical interventions (NPIs) implemented by the governments across the globe to mitigate the spread of SARS-CoV-2. Examples of these NPIs include: policies for shelter-in-place, physical distancing, and contact tracing (Flaxman et al., 2020; Ferretti et al., 2020). Similar to influenza viruses, COVID-19 spreads via close contacts and through respiratory droplets that can stay in the air for some time (Centers for Disease Control and Prevention, 2020). Similarly, spatial proximity but delayed temporal response in animals can range up to weeks for scent marks to months for anthrax transmission. Therefore, our ability to analyze delayed interaction and trace contacts in human and animal social networks through their movements is critical to understanding social dynamics (Oliver et al., 2020;

Hoover et al., 2020).

Movement data, whether obtained from wearable devices equipped with Global Positioning Systems (GPS) such as animal collars, smart watches, activity loggers, smart phones, or other location-aware technologies including Radio Frequency Identification (RFID) tags, motion sensors, Wi-Fi sensors, card readers, Bluetooth sensors, can be used to study the interaction between individuals and their space-time contact patterns. Recently, Apple and Google joined their efforts to include contact tracing functionalities using the Bluetooth technology in their smart phones (Sainz, 2020). Similarly, many other companies have built digital contact tracing apps to track proximity between individuals in space and times and inform if a risky contact with an infected individual has occurred (Kitchin, 2020). However, most of these efforts focus on detecting synchronous interactions between individuals. These technologies rely on the concurrent proximity between moving individuals to detect whether the individuals come within a certain distance of each other in space and at the same time. The proximity is identified either based on the intersection of Bluetooth signals of mobile devices carried by the individuals or the synchronous distance between individuals. These methods are not well suited to detect delayed interactions when spatial proximity occurs asynchronously. This requires techniques capable of estimating the potential paths of individuals and retaining information on their previous locations. Developing computational methods to detect delayed interactions can contribute to animal behavioral studies such as leadership and species competition, as well as research on estimating human exposure risks to airborne pathogens or hazardous agents.

Recent advances in tracking technologies and quantitative techniques have enabled scientists to analyze more complex patterns of animal and human movement in relation to environmental and geographic contexts across space and time (Dodge, 2016; Long et al., 2018; Miller et al., 2019). Among these techniques, measuring and quantifying interaction and contact between moving individuals have become a major interest in the areas of movement ecology, epidemiology, geographic information science (GIScience), computer science, and related disciplines (Potts et al., 2014; Joo et al., 2018). Arguing for an integrated science of movement, Miller et al. (2019) highlighted the importance of developing new computational approaches to analyzing the spatial and temporal patterns of movements that are critical to gaining an understanding of the collective movement behavior across the human and animal divide. This paper responds to this methodological challenge by developing and evaluating an analytical approach to trace critical space-time contacts in the social networks of humans and animals. While existing approaches provide valuable metrics to measure static interaction in space (i.e. the spatial overlap between activity spaces of individuals), our methods to quantify dynamic and temporally delayed movement interaction in space and time are less evolved (Miller, 2015). Most dynamic measures focus on quantifying *dyadic* interactions which occur synchronously (i.e. the proximity between *two* individuals in space over a time window) (Miller, 2015; Long et al., 2015; Joo et al., 2018). In a recent study, Hoover et al. (2020) developed a method to identify asynchronous interactions in animal dyads given a predefined time lag. As movement data increasingly become available in large volumes and heterogeneous forms, there is a need for more effective computational approaches to extract and model dynamic and temporally delayed interaction patterns in movements of multiple individuals.

This study focuses on two types of dynamic interaction: *concurrent interaction* and *delayed interaction*. Concurrent interaction occurs between individuals when they move synchronously in spatial proximity of each other in a shared space and at the same time. In concurrent interaction, individuals can potentially come to a close or direct contact (i.e. interaction at close distance at the same time). This is a key driver in the transmission of viruses or shaping social networks in humans and animals. Delayed interaction happens when individuals visit the same locations in space however asynchronously. This type of interaction occurs when individuals indirectly interact via a shared space however with a time lag. It can contribute to virus transmission, for example, when individuals visit the same infected location with a time delay. In this work, we introduce and evaluate a novel Object-oRiented TimE-Geographic Analytical approach (ORTEGA) to trace concurrent and delayed interaction patterns between individuals in space and time through their movement trajectories. Our approach builds upon the time-geography model (Miller, 2005; Long et al., 2015) to measure the probability of contact along the trajectories of two moving entities. We use an object-oriented scheme to make the time-geography method 'smart' by modeling trajectories as objects with properties and behaviors (actions) which can memorize the previous locations, potential contacts, and times of interactions with other individuals along their movement paths. Our main contributions include: (1) an object-oriented approach which can be used to trace concurrent and delayed interaction patterns between dyads or multiple individuals; (2) a spatial and temporal indexing technique to speed up the computation process; and (3) an approach to tracing temporally delayed contacts in the context of human contact tracing. Using GPS tracking data, we show how the proposed methodology can be applied to identify key interaction patterns among individuals across the human and animal divide. This is important, as despite the common interest for studying spatial behavior in movement ecology and human mobility, there has been little scientific cross-fertilization across these domains (Miller et al., 2019). We use two case studies to evaluate the proposed methodology: The first study uses tracking data of tigers and leopards to demonstrate how our method can capture *dyadic* tiger-tiger (intraspecific) and leopard-tiger (interspecific) interactions. The second study uses human tracking data to identify concurrent and delayed interactions among *multiple* individuals of the same and different households in the context of contact tracing. The case studies also demonstrate how ORTEGA compares to the proximity-based approaches across different temporal scales.

2. Movement Interaction

Studying dyadic movement interaction has been a major interest in movement ecology. There are two different types of dyadic interactions: *static interaction* and *dynamic interaction*. Static interaction is when the space usage (i.e. activity space or home range) of two individuals intersects. Dynamic interaction occurs when two individuals move in a shared geographic space or within the proximity of each other over a certain time interval (Potts et al., 2014; Miller, 2015).

A common technique to quantify interaction spatially is to measure the percentage of the overlap between the activity spaces of the entities (Benhamou et al., 2014). The activity space in human context is often measured using convex hulls or time-geography measures as discussed later in Section 3.1 (Miller, 2005; Long et al., 2015). The activity space in movement ecology context is computed using home range estimation measures such as convex hulls, Brownian Bridges, Kernel density estimation (Worton, 1987; Horne et al., 2007; Powell and Mitchell, 2012; Long et al., 2015). Another static measure is to compute the proportion of observations (e.g. occurrence or presence of two species) recorded within the spatial proximity of each other over the entire sampled locations regardless of observation time (Cole, 1949).

Various measures have been offered to analyze dyadic interaction using trajectories. Long et al. (2014); Miller (2015); Joo et al. (2018) provide extensive reviews on these measures, including: *the proximity index* (Bertrand et al., 1996; Doncaster, 1990)—the frequency at which the two entities come near each other within a certain distance threshold, *the coefficient of association* (Cole, 1949)—the proportion of proximal fixes, *the coefficient of sociality* (Kenward et al., 1993)—the ratio of average raw distances between simultaneous fixes (i.e. locations recorded at the same timestamps) against the average distance between non-simultaneous fixes, *the coefficient of interaction*—synchronous use or avoidance of a reference area, *the joint potential path area* (Long and Nelson, 2013)—the relative size of the potential encounter area, and *the correlation indices* (Konzack et al., 2017)—correlation between movement parameters of the two individuals such as speed, direction, step length, turn angle.

Spatial proximity is the most common metric used in the interaction measures described above. It is often computed using a kernel and a distance metric such as Euclidean distance. The *proximity* measure indicates the average number of times where the two entities meet within a spatial buffer of a certain threshold δ. Another way to measure spatial proximity is to compute the average distance between the entities over a kernel using a similarity measure (i.e. distance metric) such as the average Euclidean distance over a kernel, Edit distance, Frèchet distance, Hausdorff distance (Alt and Godau, 1995; Guibas et al., 2000; Konzack et al., 2017).

Existing dynamic measures quantify the closeness of moving entities in space at the same time using user defined spatial and temporal thresholds (Miller, 2015). This approach involves using a spatial distance buffer around the individuals' location for a certain time window, and if the buffers intersect in space and time, it can reasonably be assumed that those two individuals possibly interacted or came into contact given the uncertainty in their actual locations during the time window. A similar proximity-based approach is also applied in current Bluetooth-based contact tracing apps (Sainz, 2020). These technologies capture if individuals move in close proximity of each other or come to contacts based on the synchronous intersection between their Bluetooth signal range areas (i.e. which can be modeled as a spatial buffer). Long and Nelson (2013) enhances the definition of dynamic interaction by incorporating movement direction (i.e. azimuth) in addition to the distance. In their definition, two entities interact when their azimuth shows similarity over time. These measures require that the movement of individuals to be recorded simultaneously at the same sampling rates. Konzack et al. (2017) incorporated a similar measure to identify delayed interaction in trajectories that follow similar paths with a time delay.

Dynamic interaction measures which are based on a similarity or distance metric can be less effective when interacting entities do not follow a similar path simultaneously. These measures are less reliable when movement data are collected at different sampling rates or include gaps due to imperfect tracking or signal loss. Additionally, these approaches have limited capability to identify delayed interactions. In contrast, the potential path area (PPA) using the time-geography framework (Miller, 2005) provides a more robust method for discovering potential interaction between individuals at different time scales (Long et al., 2015; Hoover et al., 2020). The PPA represents possible locations that could be occupied by the individual between known timestamps given its maximum speed and a time budget. Long et al. (2015) showed the intersection of the potential path areas (called "joint PPA") can identify potential locations for interaction between two entities in space and time. Hoover et al. (2020) further extended the joint PPA measure to identify delayed dyadic interaction (called "temporally asynchronous-joint potential path area (ta-jPPA)") in animal pairs with a user-defined time-distance window. In this paper, we enhance the PPA measure to quantify concurrent and delayed interactions between dyads and multiple tracked individuals in space, and provide a more flexible search approach to trace interaction patterns.

3. Methods

3.1. The time-geography framework

This study uses the time-geography framework (Hägerstrand, 1970) to identify potential concurrent and delayed interactions between moving individuals (hereafter referred to as 'entities' in the description of the methodology). In time geography, the *activity space* of a moving entity (i.e. the locations that are accessible to the entity) is mapped with a *space-time prism* in a three-dimensional space-time cube (Figure 1a). The prism is anchored at two fixed locations P_i and P_j (e.g. representing a pair of origin and destination locations or the locations of two fixed activities). The shape of the prism is a function of the entity's maximum speed capacity v_{max} and the time budget ($\Delta t = t_j - t_i$) to travel between the two locations (Miller, 2005). The *potential path area (PPA)* is the projection of the space-time prism on a two-dimensional geographic or Euclidean space. The PPA_{ij} ellipse delineates the spatial locations that are accessible to the moving entity during the interval $[t_i, t_j]$. Figure 1 illustrates the space-time prism and the potential path area of a moving entity between two locations P_i and P_j along its trajectory at the time interval $[t_i, t_j]$.

Following Long and Nelson (2015), we modify the space-time prism by dynamically varying the maximum speed at different time intervals instead of using a fixed v_{max} for the entire trajectory. Here, v_{max} for a given time interval $[t_i, t_j]$ is estimated using the speed values v_{ij} during the preceding time intervals in the data. The premise is that an entity does not always move at its maximum speed and the speed varies over time according to the entity's activity type (i.e. commuting, leisure activity, foraging, hunting) or its transport mode (i.e. walking, biking, driving, etc.). Given a trajectory, $T = \{(x_0, y_0, t_0), (x_1, y_1, t_1), ..., (x_i, y_i, t_i), ..., (x_n, y_n, t_n)\}$, the maximum speed capacity v_{max} for the time interval $[t_i, t_j]$ is estimated by a floating average of speed over an exponential kernel (Gijbels et al., 1999) of size m, a

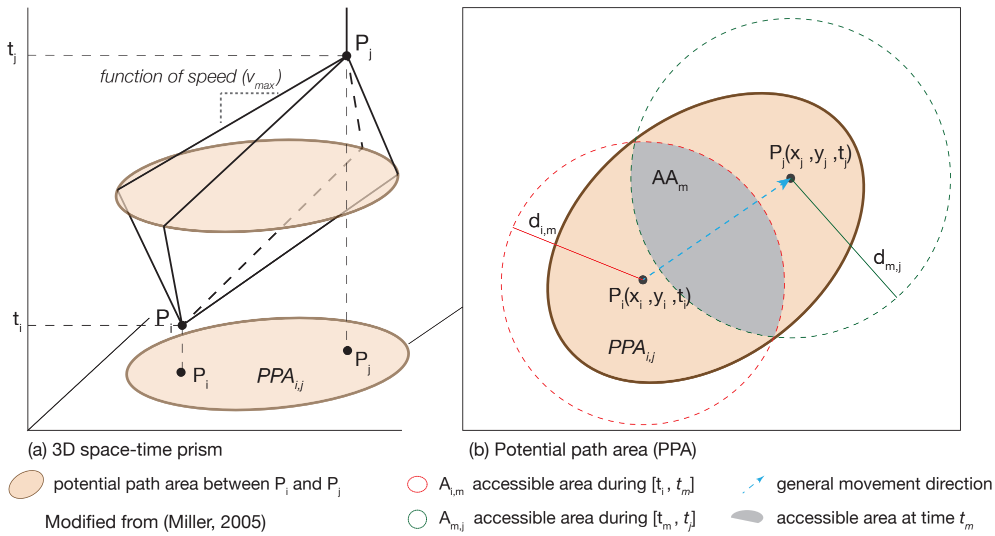

Figure 1: Illustration of (a) the space-time prism in a 3D space-time cube; (b) the computation of PPA_{ij} as the potential path area in a 2D Euclidean space between tracking points $P_i(x_i, y_i, t_i)$ and $P_j(x_j, y_j, t_j)$, where $1 \leq i < j \leq n$ and $i < m < j$. Modified from Miller (2005).

smoothing constant of α, and an additional error term β using Equation (1). In this equation, m is the size of the exponential kernel, i is the index of the current point P_i in the trajectory, k is the location in the kernel. The speed values of m previous trajectory points contribute to the floating average of the speed of the current point s_i. However, their contribution is decreased by a smoothing constant α so that the contributions of the points further from the current location in the trajectory are suppressed multiplicatively in calculating v_{ij}. If $\alpha = 1$ only the speed value of the current point (v_i) is considered. The error term β controls the uncertainty of speed in the estimation of v_{max} enabling the entity to move faster than the previous speed. The parameter β can be set as the maximum rate of speed change (i.e. deviation from average speed) for a given behavior. For example, $\beta = 1.25$ allows 25%

deviation from the current average speed.

$$s_i = \alpha \sum_{k=0}^{m-1} (1-\alpha)^k v_{i-k} \quad \text{with} \quad 0 < \alpha \leq 1$$
$$v_{max} = \beta * s_i \tag{1}$$

The potential path area between t_i and t_j, PPA_{ij}, can then be computed using the estimated v_{max} and the time budget as shown in Figure 1 and Equation (2), following Miller (2005); Long et al. (2015):

$$\begin{aligned} d_{i,m} &= (t_m - t_i) * v_{max} \quad \textit{radius of the accessible area during } [t_i, t_m] \\ d_{m,j} &= (t_j - t_m) * v_{max} \quad \textit{radius of the accessible area during } [t_m, t_j] \\ AA_m &= A_{i,m} \cap A_{m,j} \quad \textit{accessible area at time } t_m \\ PPA_{ij} &= \cup AA_m \end{aligned} \tag{2}$$

3.2. Mapping interaction as the intersection between space-time prisms

Given a trajectory T of length n, a series of space-time prisms between each pair of consecutive tracking points (i.e. $P_i(x_i, y_i, t_i)$ and $P_{i+1}(x_{i+1}, y_{i+1}, t_{i+1})$, where $1 \leq i \leq n$) can be used to estimate the locations that are accessible to the entity along its trajectory. The space-time prisms map the uncertainty in the activity spaces of the entities between known GPS locations. The intersection between the space-time prisms along the paths of the two moving entities can then be used to map the spatiotemporal overlap between their activity spaces as potential locations for dyadic interaction in space and time

(Long et al., 2015). In a two-dimensional Euclidean space, the intersection of the projected pairs of $PPAs$ of the two entities delimits the potential area for a dyadic interaction during the time period over which the $PPAs$ are calculated.

Formally, let $PPA_{ij}^{e_1}$ be the potential path area between two consecutive GPS points $P_i(x_i, y_i, t_i)$ and $P_j(x_j, y_j, t_j)$ of entity e_1, and $PPA_{kl}^{e_2}$ be the potential path area between two consecutive GPS points $P_k(x_k, y_k, t_k)$ and $P_l(x_l, y_l, t_l)$ of entity e_2. Then, $PPA_{intersect}$ as the potential locations for interaction is quantified following Long et al. (2015) using Equation (3). If the time intervals of the two $PPAs$ are the same or overlapped (i.e. $[t_i, t_j] \cap [t_l, t_k] \neq \phi$), then the $PPA_{intersect}$ (Equation 4) indicates a potential direct contact or *concurrent interaction*. It is important to note that, the $PPA_{intersect}$ in this case represents all the locations that the two individuals could potentially come to close contact between the period $[t_i, t_j] \cap [t_l, t_k]$ and does not necessarily mean that the individuals intentionally or directly interacted. However, if the time constraint is relaxed (i.e. $[t_i, t_j] \cap [t_l, t_k] = \phi$ or $[t_i, t_j] \cap [t_l, t_k] \geq t_{lapse}$), the $PPA_{intersect}$ at different time intervals can identify potential *delayed interactions* via indirect contact after a time lapse t_{lapse} (e.g. e_2 goes to the same geographic area where e_1 visited with a delay t_{lapse}).

$$PPA_{intersect} = PPA_{ij}^{e_1} \cap PPA_{kl}^{e_2} \tag{3}$$

$$contact = \begin{cases} \text{no contact}, & \text{if } PPA_{intersect} = \phi \\ \text{concurrent}, & \text{if } PPA_{intersect} \neq \phi \ \& \ [t_i, t_j] \cap [t_l, t_k] \neq \phi \\ \text{delayed}, & \text{otherwise} \end{cases} \tag{4}$$

Although closely related, our method to identify delayed interactions is different from the one presented in Hoover et al. (2020) in two ways: (1) Hoover et al. (2020) consider two additional predefined parameters to set a time window tw (time before and after the tracking point) for a delay (these can be set by the domain expert) to intersect the PPA of entity e_1 at time T_1 ($t_i < T_1 < t_{i+1}$) and the PPA of entity e_2 at time T_2 ($t_i + tw_i < T_2 < t_{i+1} + tw_j$), (2) both Hoover et al. (2020) and Long et al. (2015) consider a finer resolution time-slicing approach to compute the accessible space at each time slice T ($t_i < T < t_{i+1}$) for each individual. However, our method considers the intersection of the PPAs in space over the entire time interval between the consecutive tracking points (e.g. $[t_i, t_{i+1}]$) and does not limit the accessible space to a specific timestamp T. Although the time-slicing approach makes $PPA_{intersect}$ more explicit to a specific timestamp T, it poses several limitations: (a) it requires tracking data of both entities to be of a regular and perhaps an equal sampling rate, so that the interval can be sliced to a set of equal time increments δ (a predefined parameter); (b) it increases the computation cost of recalculating $PPA_{intersect}$ at much higher frequencies per increment δ (i.e. $t_i < T + k\delta < t_{i+1}$ where $k = \frac{t_{i+1}-t_i}{\delta}$) per entity, while ORTEGA only computes one PPA per interval $[t_i, t_{i+1}]$ per entity; and (c) it assumes that the entities always move at constant speed V_{max} between the two consecutive tracking points and therefore the highest probability of interaction always occurs along the beeline between the two consecutive tracking points of each entity. Moreover, in contrast to Hoover et al. (2020), our approach does not require predefined time-distance tw thresholds to identify delayed interaction over a fixed time lag period.

As explained later in Section 3.3, we develop a more flexible approach for tracing delayed interaction by incorporating an object-oriented scheme and a space-time index-based search algorithm to store and retrieve calculated PPAs and their intersections regardless of intersection time and without a need for frequent recalculation of $PPA_{intersect}$ on the fly for each specified time lag.

3.3. ORTEGA methodology to trace contact patterns in space and time

In order to trace contact patterns in the forms of concurrent and delayed interactions in movement data sets, we propose an Object-oRiented TimE-Geographic Analytical method (ORTEGA). First, we apply an object-oriented scheme which constructs trajectories as a set of interacting moving objects (or agents) to facilitate tracing possible PPA intersections (see Section 3.3.1). Then, the interaction analysis is applied using the $PPA_{intersect}$ measure and a spatiotemporal indexing approach to optimize the extraction of concurrent and delayed interactions via potential direct and indirect contacts between trajectories (see Section 3.3.2). The object-oriented model is used as the building block of our method to enhance the capacities of time-geography for a more complex and flexible interaction analysis, especially when more than two entities are involved. For example, it enables storing spatial and temporal characteristics of the PPAs and their intersections in the history of the data and using the information on demand. It can store information on possible interactions between PPAs of multiple agents as their properties (e.g. the time and location of first interaction, duration and number of interactions, etc.). The model is flexible, for example, to incorporate 'behaviors' for the PPA ellipses to take actions when an interaction is detected or when the time lag is too long for a meaningful interaction.

3.3.1. The object-oriented scheme

The object-oriented model is an important aspect of ORTEGA. Using a set of Class and Object definitions, the model enables moving entities to act like smart agents as self-contained objects with a set of properties and behaviors that can interact with each other. In this model, a $MovingObject$ class (Güting et al., 2000) is used to represent the moving entities. The UML diagram of the proposed object-oriented scheme is provided in Figure 2. Each $MovingObject$ can have a $Trajectory$ object which itself is a series of $MovingPoint$ objects. The proposed scheme models a space-time prism object ($STPO$) as an object class constructed from sequential pairs of $MovingPoint$ objects. Each $STPO$ has a unique identifier and records the intersections with the $STPO$ of other $Trajectory$ objects as an array of intersected $STPOs$ (or in short $iSTPO$). The array $iSTPO$ stores all possible intersected PPA ellipses for each $STPO$ along the $Trajectory$ of each $MovingObject$ in the data set. This way, the information about the location and time of intersected $PPAs$ are stored and can be retrieved when needed.

3.3.2. Interaction analysis algorithm

Figure 3 presents the algorithm for tracing possible contacts in movement data of multiple entities to identify concurrent and delayed interactions. Given a database of GPS tracking data including M moving objects, the $Trajectory$ objects are first constructed using the object-oriented model described in Section 3.3.1, following a trajectory preprocessing step (Dodge

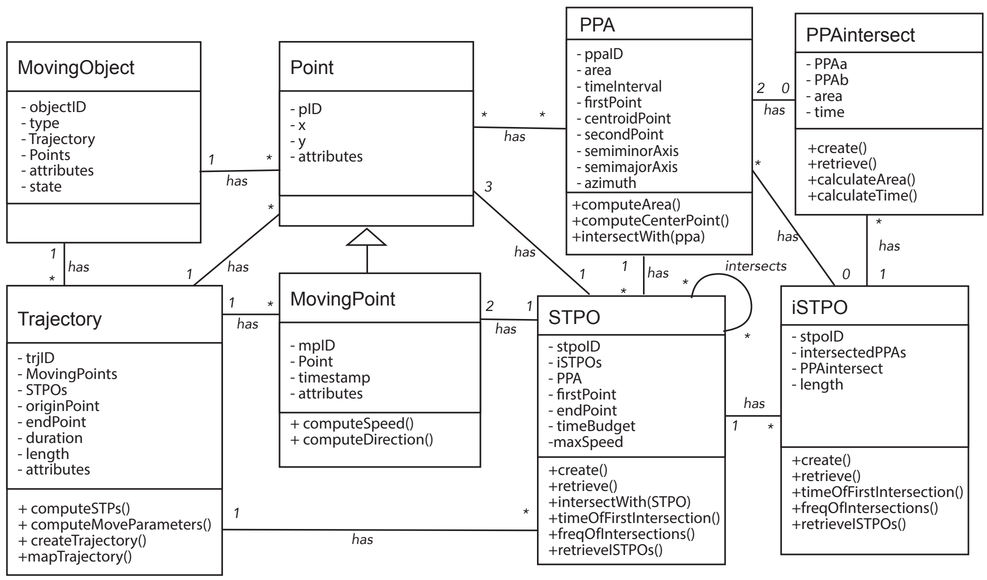

Figure 2: The object-oriented model applied in ORTEGA for moving object interaction analysis.

et al., 2009) to eliminate outliers and erroneous data. Next, the space-time prisms and their corresponding potential path areas are calculated along the $Trajectory$ of each $MovingObject$. These are stored as arrays of $STPO$ and PPA objects embedded in $Trajectory$ objects. In order to optimize the computation in large tracking data sets, a spatial and temporal indexing method, a 'compressed kd-tree' or CKD-tree (Caro et al., 2016; Bentley, 1975), is applied based on the centroid points of the calculated $STPOs$. This way, only the PPA objects that are closer in space and within a given temporal interval are considered for interaction analysis.

In order to trace contacts between M moving entities, one entity is considered as a reference $MovingObject$ (e.g. MO^r) (Figure 3). Then using the CKD-tree indexing, potential $STPOs$ that are proximate to the $STPOs$ of the reference MO^r are retrieved. Once potential $STPOs$ are retrieved, their

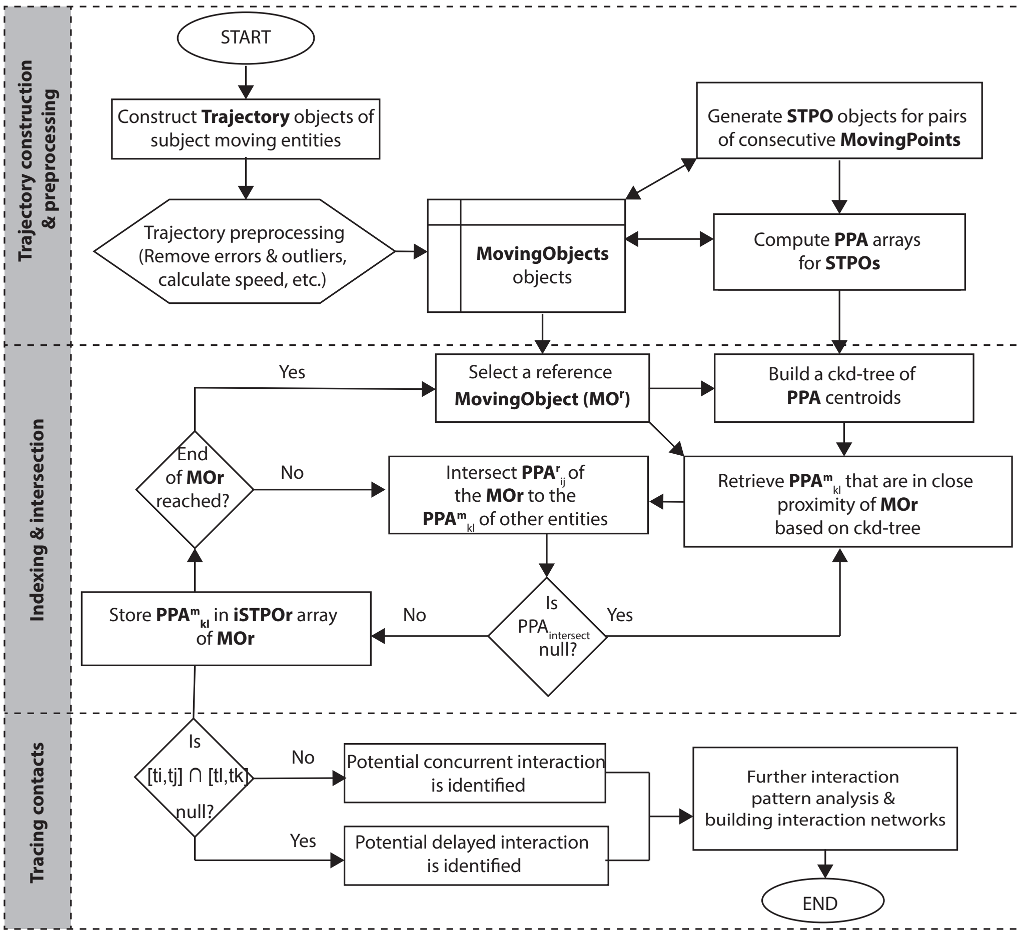

Figure 3: Workflow of the interaction analysis algorithm used in ORTEGA.

PPA ellipses are intersected with the PPA ellipses of the references MO^r. For any pair of PPA^r_{ij} (i.e. the PPA of the reference MO^r at interval $[t_i, t_j]$) and PPA^m_{kl} (i.e. the PPA of the m^{th} entity at interval $[t_k, t_l]$) that their intersection is not *null* ($PPA_{intersect} \neq \phi$), the intersecting PPA^m_{kl} is stored in the $iSTPO$ array for further analysis. The advantage of this approach as compared to the existing dyadic interaction analysis techniques is that, it can be applied simultaneously to multiple entities and enhanced using parallel computing. Since the identifiers, time and location of the intersected areas (i.e. $PPA_{intersect}$) are stored in the $iSTPO$ array as properties of a $STPO$ object, the approach is flexible enough to retrieve all possible intersections among a group of moving entities. This is also useful for the detection of different types of movement patterns such as leader and follower, avoidance, or divergence.

To detect concurrent interaction, the intersected $PPAs$ for which $[t_i, t_j] \cap [t_k, t_l] \neq \phi$ are selected. These can be further analyzed to calculate the length of contact, the number of $MovingObjects$ that the individual came into close contact, and the frequency of contacts. Delayed interaction can be extracted from the intersected $PPAs$ stored in $iSTPO$ using a time window. For example, we can retrieve all $MovingObjects$ that visited the same location within a certain time of each other.

4. Case Study I: Analysis of Dyadic Interactions in Animal Movement

We evaluate the proposed performance of the methodology on mining dyadic interaction patterns between tigers and leopards in the Western Forest Complex in Thailand. The purpose of this case study is twofold: (1) to evaluate ORTEGA in analyzing dyadic movement interaction in a Euclidean space, (2) to conduct a comparative evaluation of ORTEGA and the proximity-based approach on GPS tracking data that are collected at relatively coarse frequencies and different sampling rates (15 min to 1 hour).

4.1. Animal GPS tracking data set

Four tigers and one leopard were captured and fitted with VECTRONIC Aerospace GmbH collars (Simcharoen et al., 2014). Location data were acquired at 1 hour or 15 minute intervals over a period of four months as summarized in Table 1.

Table 1: The duration and temporal intervals of tiger tracking data.

animal ID	**No. points**	**temporal resolution**	**start date**	**end date**	**home range area (km^2)**
tiger 20080	1798	1 hour	2016-04-30	2016-09-16	57.12
tiger 20083	2751	1 hour	2016-04-30	2016-09-17	56.35
tiger 22901	4529	variable	2018-09-27	2019-02-02	49.68
tiger 22904	1998	1 hour	2018-09-27	2019-02-02	60.67
leopard 31898	7170	15 min	2018-09-27	2019-02-02	40.57

4.2. Tiger-tiger interaction

A delayed interaction occurs when a tiger scent marks a tree and that mark is subsequently inspected hours or days later by another tiger. Tigers have large home ranges and scent marks serve as message boards to alert other tigers of territorial boundaries, and in the case of a female, her reproductive status. A female tiger scent marks intensively starting a week before she comes into estrous (Smith et al., 1989). A male patrolling a large territory encompassing several females is alerted that the female will soon be receptive. When the female becomes receptive, she reduces scent marking and begins repeated calling. The male response then brings them together (Ahearn et al., 2001). Scent marks that are detected and often over-sprayed up to three weeks later serve to demarcate territorial boundaries while reducing the likelihood of direct encounters that can result in serious injuries to both animals and sometimes mortal injuries to the loser.

Here, we demonstrate the applicability of the method to study the interaction between two female tigers (IDs: 20080 and 20083) sharing a boundary along one side of their home ranges. Figure 4a illustrates the GPS data and home ranges of the two tigers during the tracking period between April and September 2016. The home ranges are calculated as the convex hulls of 95% of GPS points (Worton, 1987). As seen in the 3D space-time cube representation of the tracking data in Figures 4a-b, the two tigers patrol their shared boundary often asynchronously, and rarely have direct encounters. These asynchronous messages reduced the likelihood of costly aggressive encounters (Smith et al., 1989). We applied our method to extract concurrent and delayed interactions between the two tigers. The yellow PPAs in Figure 4c illustrate the interactions between the two tigers allowing three hours. Figure 4d visualizes the frequency of detected interactions over a range of time lags from 30 minutes up to five weeks. To be specific, each bar represents the frequency of the first visit of one tiger to the same location visited by the other tiger (i.e. first ellipse intersection) with a certain time delay. There are 53 close encounters within 3 hours out of a total of 6756 detected spatially intersected PPA pairs (Table 2). As the histogram shows, most of the interactions between the two tigers occur within one or two days of each other when the tiger scent appear to be the strongest. Few yellow marks in Figure 4c and low frequencies in Figure 4d within three hours indicate the small likelihood of tigers directly encountering each other. The visits to the same locations increase after one week and last until week three. Our findings here support field observations in Thailand using camera traps and by human observers which indicate that tigers appear to detect scent marks only at a close spatial range (several meters at the most). In our field observations, we can detect a mark for up to 21 days. However, the olfactory acuity of tigers is clearly much higher than humans; therefore, we see visits to the same locations up to even five weeks.

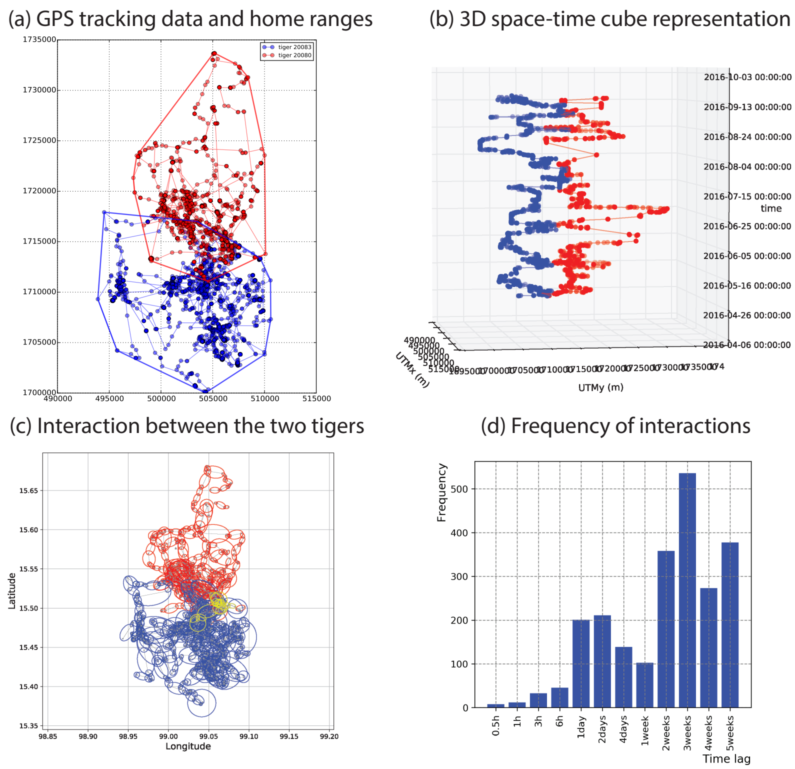

Figure 4: Interaction analysis of two female tigers (IDs: 20080 shown in blue and 20083 shown in red) sharing a boundary: (a) 2D representation of GPS tracking data and home ranges; (c) 3D space-time cube representation of the tigers' tracking data; (c) Intersected PPAs highlighted in yellow indicates interaction allowing three hours of delay; and (d) the frequency of interactions (first visits to the same locations) detected between the two tigers over a range of time lags.

Table 2: The total number of detected PPA intersections (spatial) among tigers and leopards and the number of close encounters (interaction allowing three hours).

animal 1	**animal 2**	**total PPA intersections**	**close encounters**
tiger 20080	tiger 20083	6756	53 (0.784%)
tiger 22901	tiger 22904	40260	31 (0.077%)
leopard 31898	tiger 22901	41410	2 (0.005%)
leopard 31898	tiger 22904	41255	107 (0.259%)

4.3. Tiger-leopard interaction

This experiment uses GPS tracking data of two tigers 22901 (young male) and 22904 (female) and a leopard 31898 (male) over the period between September 2018 and February 2019. Figure 5a (in the middle) illustrates the tracking data and home ranges of the three carnivores. The two tigers, a resident, breeding female and a subadult, non-reproductive male, share a portion of their home ranges. The leopard's home range overlaps with both tigers, but is completely within the home range of tiger 22904. The two tigers were originally tracked with a sampling interval of one hour; however, the tracking frequency of tiger 22901 was increased later to 15 minutes. These data help to analyze the impact of temporal granularity (i.e. sampling rates) of tracking on the analysis of animal interaction. With this experiment, we demonstrate that in contrast to a simple proximity-based approach, in which tracking data must be of the same sampling rate and collected synchronously, our method is capable of handling tracking data of variable sampling rates. The results of interaction analysis using a three hour window between tigers 22901 and 22904 indicate that these two animals only came into close contact 31 out of 40260 times when their spatial path crossed (Figure 5b). This limited direct interaction over a period of four months may be explained by the fact that there is no territorial competition between tigers of the opposite sex. Also, the male is not yet of breeding age and the female is raising young so she is not receptive. In contrast, females 20080 and 20083 are territorial, breeding females each defending a territorial boundary. They show a much higher rate of potential concurrent interactions, as they patrolled and likely marked their common boundary (see Section 4.2 and Table 2).

The interaction analysis applied on this data set suggests that the leopard tends to avoid the two tigers in the area. The number of delayed interactions is much higher between the leopard and the two tigers with very few close encounters detected (Figure 5c-d). Out of the total of 41410 spatially intersecting PPA pairs, only two incidences of close contacts (i.e. encounter in space and time allowing three hours) were detected between tiger 22901 and the leopard (Figure 5c). Figure 5d highlights 107 close contacts within a three hour window which occurred in the shared home range areas of tiger 22904 and the leopard, out of the total of 41255 spatially intersecting PPA pairs. It is interesting to note that the number of near-concurrent interactions of the leopard with tiger 22904 is relatively higher than with tiger 22901, as the leopard's home range is contained within the home range of tiger 22904. Therefore, the larger shared space increases the probability of having close encounters between the two animals. The histograms in Figure 5b-c represent the frequency of intersecting $PPAs$ over a range of time lags from 30 minutes to 24 hours. The histogram suggests that the animals happen to visit the locations visited by the other animal with a delay of 24 hours, while they keep distance for at least a few hours of each other. Based on biological observations delayed interaction between tigers and leopards after

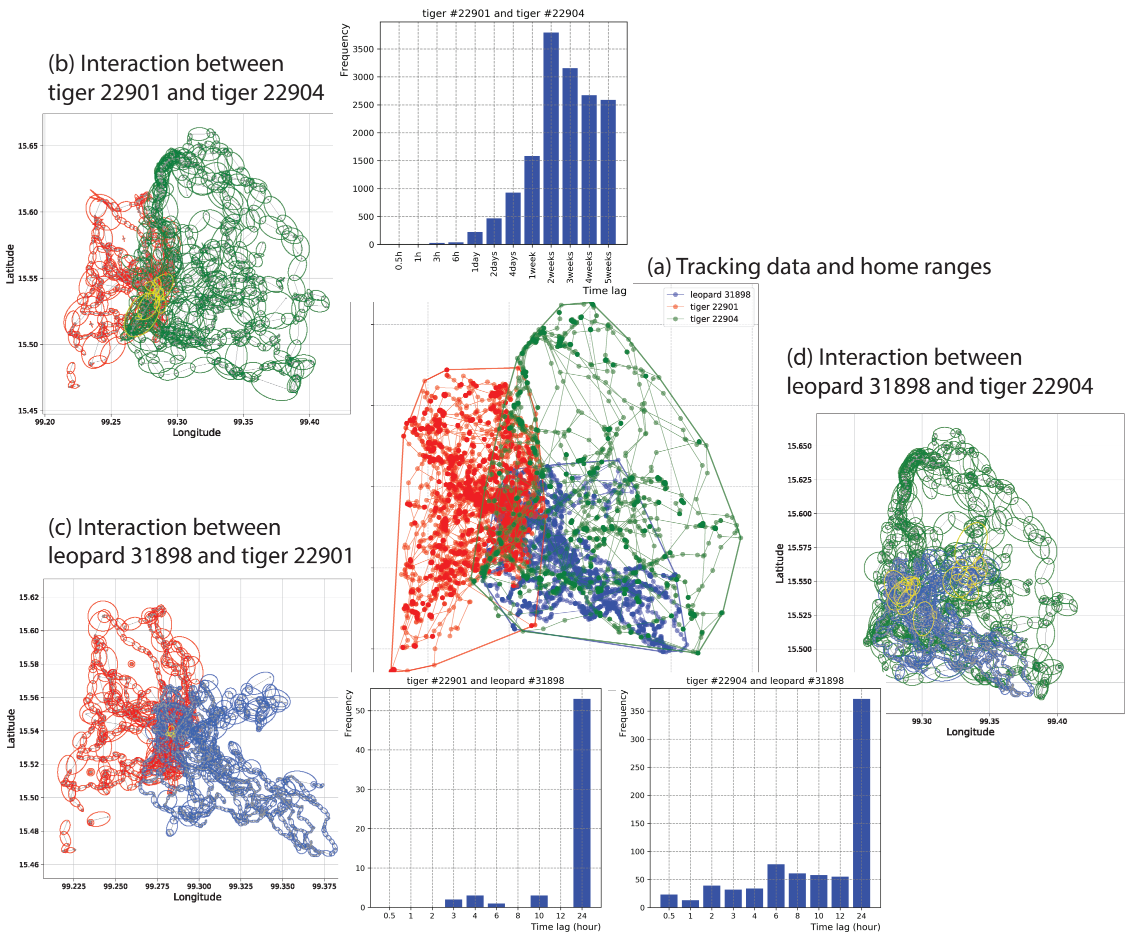

Figure 5: Results of tiger-leopard interaction analysis using (a) GPS tracking data of two tigers 22901 (in red) and 22904 (in green) and one leopard 31898 (in blue). Map and the frequency of interactions over a range of time lag (b) between tigers 22901 and 22904; (c) between tiger 22901 and leopard 31898; and (d) between tiger 22904 and leopard 31898. The PPA intersections allowing three hours of delay are highlighted in yellow.

more than several hours does not have a significant meaning. In contrast, avoidance, quantified as little to no concurrent interaction or close contact within few hours, which is also confirmed in our interaction analysis, is key to the survival of the leopards when sharing the same geographic space with tigers.

As a simple experiment to compare ORTEGA to the proximity-based approach and their sensitivity to data granularity, we applied both techniques on the 15-min data set of tiger 22901 and leopard 31898 and re-sampled the data to reduce its granularity (see Table 3). As the results suggest, the proximity-based approach is more sensitive to the sampling rate, while both approaches result in more false negatives as the sampling rate decreases.

Table 3: Delayed interaction between tiger 22901 and leopard 31898 within 4 hours over different sampling rates. A buffer size of 428 meters is used for the proximity-based approach to detect the same number of interactions as in the ORTEGA approach using the original data.

Data completeness (data granularity)	**ORTEGA**	**proximity**
100% (15-min)	54	54
50% (30-min)	32	8
25% (1 hour)	4	0

5. Case study II: Tracing Contacts Among Multiple Individuals

In this case study, we evaluate the performance of the methodology on tracing contacts between a group of people from the same or different households using fine resolution GPS tracking data in a network space. We further investigate the impact of varying temporal scales on the outcomes and evaluate the results in comparison to the proximity-based approach. As compared to the previous case study in which dyadic interactions of only two individuals were analyzed at the time, here we evaluate the method in identifying concurrent and delayed interactions in a larger network of people. This can be useful in contact tracing applications when detection of possible encounters between individuals is critical to monitor and estimate infection exposure.

5.1. Human GPS tracking data set

This study uses the GPS component of the 2012-13 California Household Travel Survey (CHTS) (NuStats, 2013), which includes human movement tracking data over a three-day period at a temporal resolution of three seconds. The data were collected using GlobalSat GPS Data Loggers that can be worn on the waist, clipped to a purse or backpack, or dropped in a pocket. From the CHTS data set, which covers most of California, we only used the data of respondents from San Jose as a test case. This subset contains GPS traces of 402 persons from 176 households and spans from February 3, 2012 to January 31, 2013 (total of 75,770 GPS tracking points). Each GPS record contains information including: a person's anonymous ID, location in longitude and latitude format, and local time. We only considered the GPS data from 5 am to 23:59 pm because humans mainly stay at home during the night and we were interested in day-time interactions. The original data are in much finer resolution than what is needed for the purposes of interaction analysis (i.e. human movement over three seconds might not be significant). Therefore, we down-sampled the data into 1-minute intervals. The original 3-second sampling rate data result in very narrow $PPAs$ which are basically equivalent of using a beeline between the consecutive GPS points. Such fine resolution GPS tracking data can perhaps improve the performance of the proximity-based interaction analysis methods, but it is very expensive (computationally and financially) to collect tracking data at this very fine sampling rate. Our method does not require such fine sampling rate as the potential path area inherently incorporates the uncertainty of positioning and the accessible locations at times that the GPS data are not recorded.

Table 4, summarizes the results of applying interaction analysis on this data set. The results are described in the following sections. The first three experiments present the outcomes of ORTEGA on tracing concurrent interaction (defined as contacts within 5 minutes) as compared to a proximity-based approach (see Section 5.2). The last four experiments summarize the outcomes of delayed interaction analysis using ORTEGA for a range of time lags from 30 minutes to 3 hours (see Section 5.3).

Table 4: Number of concurrent and delayed interactions detected among individuals of the same and different households. As a reference for comparison to ORTEGA, the outcomes of the proximity-based approach for concurrent interaction analysis using two buffer sizes are provided.

Exp.	**method**	**type**	**parameters**	**within household**	**outside household**	**total**
1	proximity	concurrent	5 min, 100 m	187	12	190
2	proximity	concurrent	5 min, 500 m	198	59	257
3	ORTEGA	concurrent	5 min	202	149	351
4	ORTEGA	delayed	30 min	250	376	626
5	ORTEGA	delayed	1 hour	279	542	821
6	ORTEGA	delayed	2 hours	299	708	1007
7	ORTEGA	delayed	3 hours	306	820	1126

5.2. Tracing concurrent interactions and comparison to the proximity-based approach

The goal of this experiment is to compare ORTEGA to the proximity-based approach in detecting concurrent interactions between the 402 participants in the data set, and investigate the influence of temporal scale (i.e. sampling rate) on both methods. The proximity-based approach is implemented by intersecting spatial buffers of a certain distance threshold around synchronous GPS tracking points. In practice, often a user-defined time threshold is considered to relax the restriction of requiring synchronous fixes which can be hard to achieve in real tracking data. Here, a five-minute time lag is considered to extract concurrent interactions using both approaches. In these experiments, two buffer sizes (i.e. 100 meters in Exp. 1 and 500 meters in Exp. 2) are considered (see Table 4). These buffer sizes are considered to account for average distance traveled by intermittent or continuous walking over a five-minute interval. The average distance traveled by walking is 100 meters over one minute. Our proposed ORTEGA approach is not reliant on a buffer threshold or a set time window. Although in this experiment we considered the same 5-min time window to make it comparable to the proximity-based approach.

Figure 6 represents two networks generated using Gephi (Bastian et al., 2009) based on the detected concurrent interaction (close contacts allowing 5 minutes delay) among all individuals using (a) the proximity-based approach (with a buffer size of 100 m), and (b) the ORTEGA approach. The networks include 402 nodes (i.e. each node represents one person). The edges represent the interactions between every pair of two persons. The lighter pink to beige represent less to no interactions, while more saturated pink to purple colors represent higher number of interactions. A comparison between the two networks (Figure 6 and Table 4) suggests that our approach is capable of finding more potential direct contacts between individuals, while a higher degree of concurrent interaction is detected among individuals of the same household. The network generated using ORTEGA unveils some clusters grouping people who interacted more with their own household members, while they had less close contacts with people outside their households. This may indicate households with children who mainly traveled together and had fewer encounters with other individuals in the data set. A closer look into the generated network using ORTEGA reveals more details on the frequency of interactions between individuals within the same households and their connections to individuals of other households. For example, the zoomed-in network of person H01P01 (from household H01) in Figure 6 illustrates that she/he interacted closely with three of her/his household members and came into close contacts with four other persons from different households over the course of the three days tracking period (see also Table 5). The histograms in Figure 6 indicate that the proximity-based approach in this case missed most of the close contacts among people outside households as compared to our approach. Therefore, it resulted in a less structured and more homogeneous network with more isolated nodes. It is worth mentioning that the proximity-based approach resulted in a higher number of individuals with no direct contact as compared to our approach (isolated nodes in the middle of the networks). However, based on the CHTS survey data we know that most of these individuals lived and interacted with at least one other person. The data include 154 (out of 176) households with at least two persons. According to the histograms in Figure 6, ORTEGA also identified a higher number of possible close contacts between individuals of different households. This result indicates a higher chance of encountering more people from other households in a shared location at the same time (e.g. a grocery shop). However, the proximity-based approach could only identify close contacts between a handful of outside household individuals.

The results indicate that in general the proximity-based approach detects

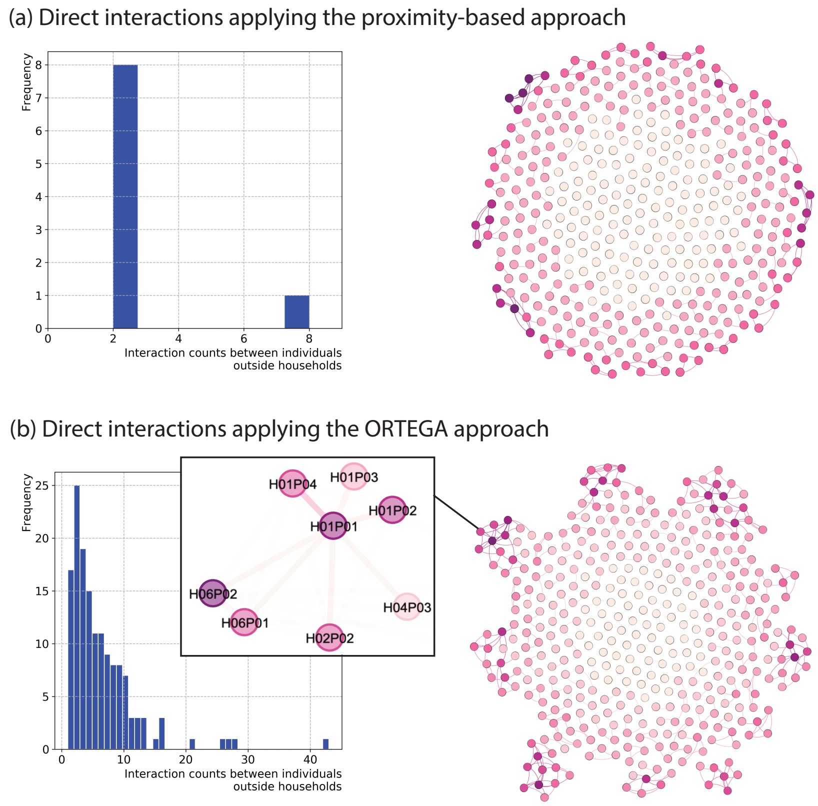

Figure 6: Illustration of the networks of individuals who came into close contacts (allowing five minutes time lag) extracted applying (a) the proximity-based approach (using a buffer size of 100 m) (b) the ORTEGA approach. The histograms show the frequency of the concurrent interaction incidences detected between people of different households. ORTEGA detects more concurrent interactions between people of different households as compared to the proximity-based approach. The zoomed-in network shows that person H01P01 interacted concurrently the most with person H01P04 from their own household. They also came into close contact with four other people from three different households.

a smaller number of interactions as compared to our approach. The detected number of interactions increases when a larger buffer size is used. For example, with a buffer size of 500 meters (which is quite large for a meaningful human interaction), the proximity-based approach can detect almost the same number of interactions as ORTEGA did in terms of concurrent contacts among individuals of the same households. However, household members tend to stay in close proximity to each other over a period of time when they travel together. These outcomes show that the proximity-based approach can significantly under-estimate the number of contacts given the buffer size. However, ORTEGA detects more potential concurrent interactions as it considers possible accessible locations to the moving entities. This comparison indicates that ORTEGA may be better suited to detect potential concurrent interactions (or close encounters) between individuals of different households as it extracts all possible cases for potential interactions. This observation is especially important in the context of contact tracing for infectious diseases and risk exposure in which an over-estimation might be more desired than the under-estimation of potential risky contacts. ORTEGA enables detection of possible contacts between individuals who may not travel together over an extended period of time and only come into contact for a short period of time. For example, when two individuals from different households happen to be in the same location (e.g. a grocery store or a gas station) at the same time for a few minutes. This is a key advantage of using a time-geographic approach over the proximity-based approach, which can under-estimate contacts given the selected buffer size and time threshold. The proximity-based approach might miss contacts when the locations of individuals have not been recorded at exact time when they happened to be close to each other or when a small buffer distance is used to represent the proximity (see Section 6 for more information). Using a larger temporal threshold and bigger buffer size may alleviate this problem but increases the uncertainty in contact tracing. The difference between the two approaches becomes even more pronounced when the GPS sampling rate decreases (i.e. coarser temporal granularity). Figure 7 demonstrates the influence of temporal scales on concurrent interaction analysis (allowing 5 minutes) using both approaches for sampling rates of 1 min, 5 min, 10 min, 20 min, and 30 min. The proximity-based approach seems to be more prone to data granularity for outside household interaction when individuals are not tracked synchronously. The number of direct contacts identified by the proximity-based approach slightly decreases within households (Pearson's $R = -0.77, P - value = 0.1$) and drops to zero outside households (Pearson's $R = -0.58, P - value = 0.3$) as the sampling rate decreases. In contrast, the potential path area used in ORTEGA takes into account the potential locations accessible to the individuals between known GPS recordings. And therefore, while the number of identified contacts stays the same for people within households using ORTEGA (Pearson's $R = 0.05, P - value = 0.9$), the chance of identifying potential interactions between individuals of different households becomes higher, although not significantly, as the sampling rate decreases due to larger PPAs (Pearson's $R = 0.79, P - value = 0.1$).

5.3. Tracing delayed interaction through indirect contacts

Using ORTEGA and a range of time lags, the delayed interactions were computed in Table 4. The networks representing delayed interactions are provided in Figure 8. As the time lag increases, more distinct clusters are detected and the networks become more fragmented. These clusters represent individuals of both inside and outside households who share the same spatial patterns (i.e. spatial interaction) but may not necessarily encounter at the same time. The longer tails in the histograms resulting from longer time lags

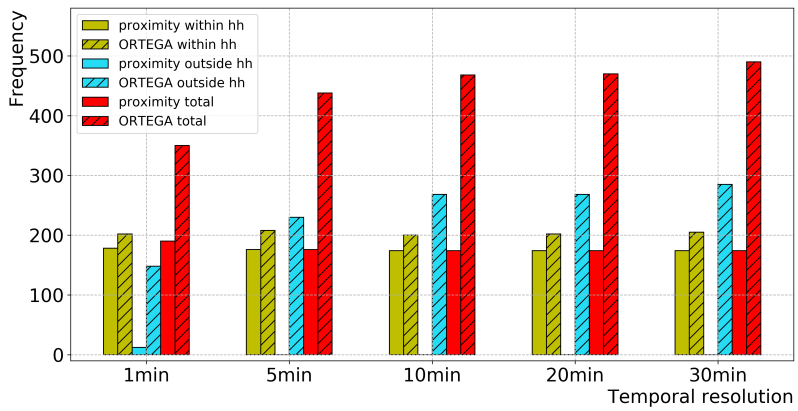

Figure 7: The influence of temporal scale on identifying concurrent interactions (allowing 5 min) using ORTEGA and the proximity-based approach (of 100 m buffer) between individuals of the same household and outside households.

represent the higher number of interactions detected among individuals of different households. This result indicates individuals of different households might visit similar locations over longer delays. That is, the chance of using the same space by more individuals becomes higher over a longer time period. The clusters in the fragmented network may indicate the households from the same neighborhoods who tend to use the same geographic space for their daily activities. With these clusters, we can detect a set of individuals that have similar spatial patterns but not necessarily following the same schedule. For example, people who go to the same gym or the same grocery shop but at different times of the day. This is the powerful aspect of our method which is capable of tracing delayed spatial interactions.

A closer look into the network of person H01P01 (as shown in Figure 6) can inform us about how many individuals this person came into contact synchronously or visited the same location as other individuals asynchronously

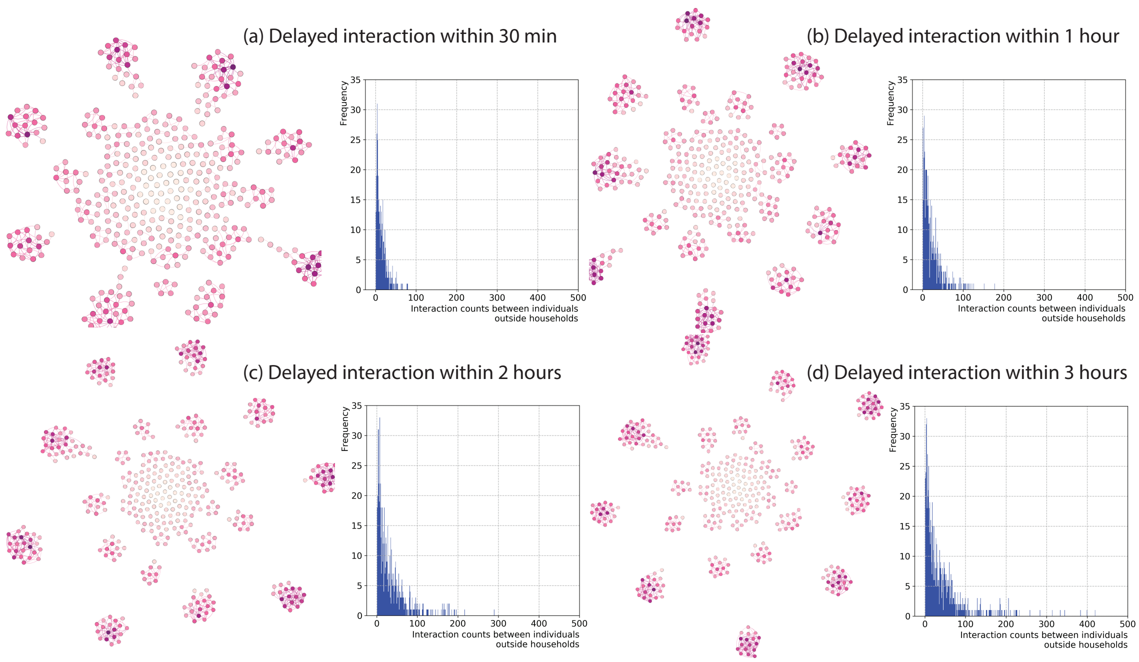

Figure 8: Illustration of the networks of individuals who had delayed interaction (i.e. visited the same location) over a range of time lags (30 min, 1 hour, 2 hours, 3 hours). The histogram shows the frequency of different number of delayed contacts between individuals outside households.

(Table 5). The outcomes indicate that this person interacted mostly with person H01P04 from the same household and person H02P02 from household H02. Overall, person H01P01 was in close contact (met with or visited the same locations at the same time) with four individuals outside the household (from households H02, H04, and H06) over the course of a three day tracking period. She/he visited the same locations as 10 other individuals from six other households after a temporal lag of 30 minutes to three hours. This is significant for finding possibility of exposure to viruses or other hazardous conditions that may last in the air for a period of time. In total, this person had concurrent or delayed contacts with 13 members of seven different households including their own household.

Table 5: Cumulative number of concurrent and delayed interactions detected between person H01P01 and other individuals within a range of time lags. The three individuals above the horizontal line are from the same household as person H01P01, while the others are from six different households (H02-H07).

interacted with	**within 5min**	**within 30min**	**within 1hour**	**within 2hours**	**within 3hours**
H01P02	29	142	190	280	327
H01P03	9	40	69	173	303
H01P04	408	755	1059	1372	1476
H02P01	0	22	40	44	56
H02P02	15	19	63	104	137
H03P02	0	15	21	34	41
H04P01	0	0	25	46	48
H04P03	6	12	15	28	28
H04P04	0	0	11	11	15
H05P01	0	4	4	7	12
H06P01	5	8	27	34	49
H06P02	3	6	56	61	64
H07P01	0	9	9	17	17

6. Discussion

In this section, we discuss the performance and efficacy of the developed methods in light of the results described above. The strengths and weaknesses of the approach are presented in comparison with the classic proximity-based interaction analysis approach which is commonly used in the contexts of animal interaction and human contact tracing.

6.1. Method parametrization

Computation of the potential path area requires two parameters which can be derived from data: the time budget Δt (which is typically calculated from the timestamps of known GPS points) and the maximum speed capacity V_{max}. Our proposed approach calculates the maximum speed parameter based on the data itself by applying a floating average over an exponential kernel instead of using a predefined fixed value for maximum speed. Therefore, the PPAs are computed with the actual (data-driven) speed capacity of the individual for different behavioral modes and at each given time. The maximum variation from the average speed can also be calculated from the data itself. We allowed 25% variation from the computed average speed. This avoids having unreasonably large PPAs by setting a fixed large V_{max} for the entire data set, which may result in more false positives in tracing potential contacts.

The proximity-based approach requires setting a predefined distance threshold as the buffer size, as well as a user-defined time threshold for the search window to relax the requirement of having synchronous fixes. Setting larger distance and time thresholds result in a higher number of contacts. This is a major limitation of the proximity-based approach. In contrast, ORTEGA is not reliant on a distance threshold or a set time window. Although in the experiments we considered the same five minutes time window for concurrent interactions to make ORTEGA comparable to the proximity-based approach. Figure 9 schematically demonstrates the difference between buffer intersection in the proximity-based approach versus the PPA intersection used in ORTEGA in terms of sensitivity to buffer size and data granularity. Figure 9a shows a smaller buffer size might lead to more false negatives and under-estimation of contacts in the proximity-based approach.

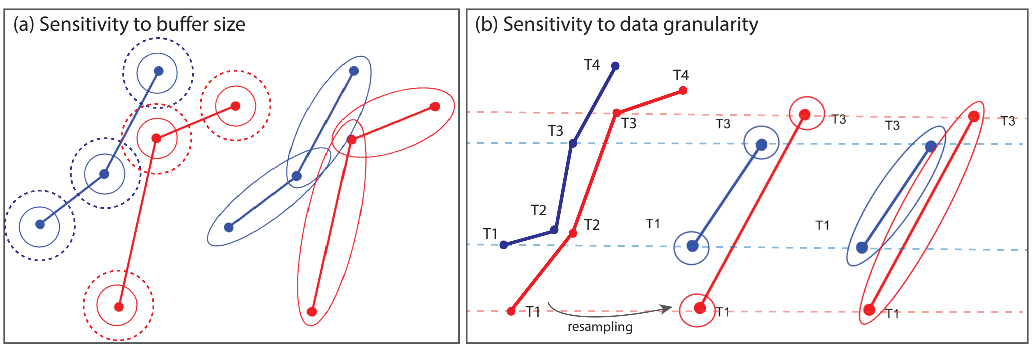

Figure 9: Illustration of method sensitivity to (a) buffer size and (b) data granularity. The proximity-based approach is highly sensitive to the buffer size and temporal resolution of the data. The ORTEGA approach does not rely on a buffer size and it outperforms the proximity-based approach when the data is collected at lower frequencies.

In comparison to previous time-geographic approaches (Hoover et al., 2020; Long et al., 2015), ORTEGA is flexible to retrieve delayed interactions using any time window in the history of the data and does not depend on several time slicing and time window thresholds to compute concurrent and delayed interactions.

6.2. Data granularity and temporal scale considerations

The original proximity-based approach is only able to detect proximate points collected at synchronous intervals unless this restriction is relaxed through a search window or kernel. Depending on the search window size, it might miss interactions that happened between known GPS points. For example, it might miss the cases where the individuals were close to each other but moved away from each other between two GPS fixes or when individuals move at different rates. In contrast, ORTEGA does not require synchronous sampling. The overlap of potential path areas of two individuals indicates their potential interaction between consecutive GPS recordings.

Overall, both approaches are sensitive to the temporal resolution used in data collection (Figure 7). However, the proximity-based approach results in more underestimation or false negatives when data of coarser sampling rate is used. In Figure 9b notice the missed proximate points at time T_2 after re-sampling using the proximity-based approach.

A weakness of ORTEGA is that for data of coarser sampling rates (e.g. 1 hour) it generates larger $PPAs$ as compared to higher-resolution tracking data (e.g. 15 min) (Figure 5d). Although it is possible to intersect PPA pairs of different resolutions–which itself is a strength when compared to the proximity-based approach, their intersections results in a higher level of uncertainty in the detection of interaction. This situation may lead to more overestimation or false positive cases for interaction as explained in Section 5.2. For example, the larger PPA of one hour duration for tiger 22904 might contain or intersect with several smaller 15-min $PPAs$ of the leopard and it might not be possible to determine actual interaction time over the one hour period.

6.3. Computation consideration

The efficiency of ORTEGA lies in lowering the computation cost that is associated with the retrieval of potential concurrent and delayed interactions among multiple individuals. This is incorporated via two key elements of the methodology: (1) ORTEGA applies an object-oriented model to create $MovingObjects$ as 'smart agents' which can retain information about the PPAs and their intersections along the trajectories of individuals. This way, the PPA polygons of $MovingObjects$ only need to be computed and intersected between different tracks once. The $PPA_{intersects}$ and their associated time intervals are also stored as properties of PPAs for efficient retrieval of delayed intersections. (2) Using a CKD-tree indexing technique, ORTEGA applies space-time indexing to limit the search area for potential interactions to smaller regions and restricted time windows when needed, so the search does not incorporate the entire trajectories of all $MovingObjects$. This offers a more efficient approach as compared to the method proposed in Hoover et al. (2020), by reducing the need for on-the-fly and redundant computation of pairwise PPA intersections over long trajectories at finer time increments and for different time windows to identify potential delayed interactions. ORTEGA retains all possible spatial intersections between PPAs, and therefore, potential delayed interactions can be retrieved on demand using an optimized search through the CKD-tree indexing.

In comparison with the proximity-based approach which relies on a distance threshold comparison, the time-geographic interaction analysis approaches (as in ORTEGA) are overall more computationally intensive as they rely on PPA ellipse computation and polygon intersection. To give a sense of the computation speed of both approaches, the interaction analysis between two GPS tracks (total of 265 points: track #1 with 183 GPS points and track #2 with 82 points) takes 3.88 seconds using ORTEGA (allowing 5 min delay) and 423 milliseconds using the proximity-based approach (allowing 5 min delay, 100 m buffer) to run using a Macbook Pro laptop with 2.3 GHz 8-Core Intel Core i9 processor, 16 GB RAM. In practice, the computation speed of the proximity-based versus PPA approaches can greatly vary based on the indexing methods used and the distribution of GPS points in the data. Though spatial and temporal indexing on larger data sets are necessary for both proximity-based approaches and PPA approaches to avoid exponential runtime growth, the indexing of simple radius proximity-based approaches is much simpler than index creation and index querying of multi-sized PPA polygons.

For both ORTEGA and proximity-based approaches, first the GPS tracking data sets need to be preprocessed to remove erroneous points and outliers. The proximity-based approach involves a search through the list of points to pick the proximate points with a time difference of less than the minimum search window threshold. The search can be improved using an indexing approach, especially for larger data sets. The ORTEGA approach involves creating PPA ellipses for consecutive GPS points in the path of each trajectory. These collections of PPAs are then intersected while the computation is optimized by taking advantage of the characteristics of object-oriented programming and the CKD-tree indexing as described in the methodology. In our experiments both the proximity-based and ORTEGA approaches were conducted in an object-oriented fashion to retain information about the points that interacted in the history of the data. This approach provides advantages and flexibility over a non-object-oriented approach, when dealing with delayed interactions.

6.4. Application considerations and the context of interaction

As shown, ORTEGA is a powerful tool to trace not only concurrent interactions, but also delayed interactions via indirect contacts in movement data of both animals and humans. However, it is important to note that the outcomes require a careful interpretation and perhaps consultation with domain experts. For example, it is important in any case study to contemplate a meaningful set of the conditions or criteria for concurrent and delayed interactions prior to the analysis. Here, we used domain expertise of a tiger biologist and an expert in travel behavior and transportation to set meaningful parameters and interpret the results. Moreover, we used field observations to back up the selected parameters. For example, we knew how long a tiger scent might last in nature for a meaningful delayed interaction.

It is important to note that depending on the setting of the case study and data granularity, the detected concurrent interactions or close contacts might not actually mean that the individuals met or socially interacted. For example, if there is a busy coffee shop and two individuals arrive at the coffee shop within 5 minutes of each other, stay for an hour or so, then leave at the same time, it can reasonably be assumed that they interacted because they planned a meeting with each other for a set time. This is in contrast to two individuals who may have been in a coffee shop together for an hour or so of overlapping time, but arrived and left at totally different times, indicating they were strangers who coincidentally were in the same coffee shop and did not interact. Other points to consider when analyzing interaction or tracing contacts are the impact of mode of transportation, physical or natural barriers, and 3D spaces in detecting meaningful contacts. For example, the interaction might not be meaningful when movement happens in separate cars on the same road or when moving individuals are separated by walls or different floors in multistory buildings or when animals are separated by a natural barrier such as a river.

7. Conclusion and Future work

This study presented a new object-oriented time-geographic analytical approach (ORTEGA) to trace space-time contact patterns in movement data. The method is capable of detecting direct and indirect contacts to identify concurrent and delayed interactions between humans or animals in space and time. In contrast to existing approaches which are limited to the interaction analysis of two individuals, ORTEGA enables tracing interaction patterns among a group of moving individuals. Our approach uses the potential path area between GPS fixes to measure potential exposures that might have been missed due to small data gaps or irregular sampling rates. These are major problems in the proximity-based approaches which are employed in most existing contact tracking technologies using Bluetooth or GPS in cellphones. We applied and tested the proposed method on two different case studies using real GPS tracking data of animals (tigers and leopards) and humans (people of the same and different households) of different resolutions. The results showed that the proposed ORTEGA method performs better than the classic proximity-based approach in tracing concurrent and delayed contact patterns in movement data, although at a higher computation cost. The outcomes suggest that the proximity-based approach underestimates contacts when individuals do not move together or are not tracked synchronously. As compared to the proximity-based approach, ORTEGA requires less parameterization and is less sensitive to data granularity. By incorporating the time-geography framework, ORTEGA incorporates movement data uncertainty and potential accessible areas between known locations, and hence it is a more powerful approach. The proximity-based approach is more sensitive to the selected distance buffer size, tracking frequency, and the search time window. For future extensions, ORTEGA can be strengthened by embedding information on the context of movement and incorporating a probabilistic method to better represent movement across the potential path areas. Currently, following the traditional time-geography model, ORTEGA assumes that the entire potential path area is accessible to the moving entity. To further extend the methodology it will be important to identify the duration of contacts which is a critical factor in the analysis of social interactions for both humans and animals. ORTEGA also opens the possibility of adding context-awareness (Ahearn et al., 2017) to interaction analysis by incorporating behavioral, environmental, and geographic parameters that influence movement interaction patterns. Using ORTEGA and sample weights expansion it is possible to estimate the total number of persons that interact in a specific place. This approach can be used not only for disease transmission but also congestion and crowding management.

Acknowledgment

This work was supported in part by the National Science Foundation Award No.1853681, the University of California Santa Barbara Office of Vice Chancellor for Research COVID-19 Seed Grants. J. Johnson's contribution to this work was supported by the University of Minnesota Office of the Vice Provost for Research's Grant-in-Aid of Research, Artistry, Scholarship Program. J.L.D. Smith's contribution to this research was supported by the USDA National Institute of Food and Agriculture.

References

Ahearn, S.C., Dodge, S., Simcharoen, A., Xavier, G., Smith, J.L.D., 2017. A context-sensitive correlated random walk: a new simulation model for movement. International Journal of Geographical Information Science 31, 867–883. URL: http://dx.doi.org/10.1080/13658816.2016.1224887, doi:10.1080/13658816.2016.1224887.

Ahearn, S.C., Smith, J.L., Joshi, A.R., Ding, J., 2001. TIGMOD: An individual-based spatially explicit model for simulating tiger/human interaction in multiple use forests. Ecological Modelling 140, 81–97. doi:10.1016/S0304-3800(01)00258-7.

Alt, H., Godau, M., 1995. Computing the Fréchet distance between two polygonal curves. International Journal of Computational Geometry & Applications 05, 75–91. doi:10.1142/s0218195995000064.

Bastian, M., Heymann, S., Jacomy, M., 2009. Gephi: an open source software for exploring and manipulating networks, in: Proceedings of the International AAAI Conference on Web and Social Media.

Benhamou, S., Valeix, M., Chamaillé-Jammes, S., Macdonald, D.W., Loveridge, A.J., 2014. Movement-based analysis of interactions in african lions. Animal Behaviour 90, 171–180.

Bentley, J.L., 1975. Multidimensional binary search trees used for associative searching. Communications of the ACM 18, 509–517.

Bertrand, M.R., DeNicola, A.J., Beissinger, S.R., Swihart, R.K., 1996. Effects of parturition on home ranges and social affiliations of female white-tailed deer. The Journal of Wildlife Management 60, 899–909.

Caro, D., Rodríguez, M.A., Brisaboa, N.R., Fariña, A., 2016. Compressed kd-tree for temporal graphs. Knowledge and Information Systems 49, 553–595. doi:10.1007/s10115-015-0908-6.

Centers for Disease Control and Prevention, 2020. How Coronavirus Spreads. URL: https://www.cdc.gov/coronavirus/2019-ncov/prevent-getting-sick/how-covid-spreads.html.

Cole, L.C., 1949. The measurement of interspecific association. Ecology 30, 411–424.

Dodge, S., 2016. From Observation to Prediction: The Trajectory of Movement Research in GIScience., in: Onsrud, H., Kuhn, W. (Eds.), Advancing Geographic Information Science: The Past and Next Twenty Years. GSDI Association Press. chapter 9, pp. 123 – 136.

Dodge, S., Weibel, R., Forootan, E., 2009. Revealing the physics of movement: Comparing the similarity of movement characteristics of different types of moving objects. Computers, Environment and Urban Systems 33, 419–434. doi:10.1016/j.compenvurbsys.2009.07.008.

Dodge, S., Weibel, R., Lautenschütz, A.K., 2008. Towards a taxonomy of movement patterns. Information Visualization 7, 240–252. doi:10.1057/palgrave.ivs.9500182.

Doncaster, C.P., 1990. Non-parametric estimates of interaction from radio-tracking data. Journal of Theoretical Biology 143, 431–443.

Ferretti, L., Wymant, C., Kendall, M., Zhao, L., Nurtay, A., Abeler-Dörner, L., Parker, M., Bonsall, D., Fraser, C., 2020. Quantifying SARS-CoV-2 transmission suggests epidemic control with digital contact tracing. Science 368. doi:10.1126/science.abb6936.

Flaxman, S., Mishra, S., Gandy, A., Unwin, H.J.T., Mellan, T.A., Coupland, H., Whittaker, C., Zhu, H., Berah, T., Eaton, J.W., Monod, M., Perez-Guzman, P.N., Schmit, N., Cilloni, L., Ainslie, K.E., Baguelin, M., Boonyasiri, A., Boyd, O., Cattarino, L., Cooper, L.V., Cucunubá, Z., Cuomo-Dannenburg, G., Dighe, A., Djaafara, B., Dorigatti, I., van Elsland, S.L., FitzJohn, R.G., Gaythorpe, K.A., Geidelberg, L., Grassly, N.C., Green, W.D., Hallett, T., Hamlet, A., Hinsley, W., Jeffrey, B., Knock, E., Laydon, D.J., Nedjati-Gilani, G., Nouvellet, P., Parag, K.V., Siveroni, I., Thompson, H.A., Verity, R., Volz, E., Walters, C.E., Wang, H., Wang, Y., Watson, O.J., Winskill, P., Xi, X., Walker, P.G., Ghani, A.C., Donnelly, C.A., Riley, S.M., Vollmer, M.A., Ferguson, N.M., Okell, L.C., Bhatt, S., 2020. Estimating the effects of non-pharmaceutical interventions on COVID-19 in Europe. Nature 584, 257–261. doi:10.1038/s41586-020-2405-7.

Gijbels, I., Pope, A., Wand, M.P., 1999. Understanding exponential smoothing via kernel regression. Journal of the Royal Statistical Society. Series B: Statistical Methodology 61, 39–50. doi:10.1111/1467-9868.00161.

Guibas, L.J., Alt, H., Guibas, L.J., 2000. Discrete Geometric Shapes: Matching, Interpolation, and Approximation, in: Sack, J.R., Urrutia, J. (Eds.), Handbook of Computational Geometry. North-Holland, Amsterdam, pp. 121–153. doi:10.1016/B978-044482537-7/50004-8.

Güting, R.H., Böhlen, M.H., Erwig, M., Jensen, C.S., Lorentzos, N.a., Schneider, M., Vazirgiannis, M., 2000. A foundation for representing and querying moving objects. ACM Transactions on Database Systems 25, 1–42. doi:10.1145/352958.352963.

Hägerstrand, T., 1970. What about people in Regional Science? Papers of the Regional Science Association 24, 6–21. doi:10.1007/BF01936872.

Hoover, B.A., Miller, J.A., Long, J., 2020. Mapping areas of asynchronous-temporal interaction in animal-telemetry data. Transactions in GIS 24, 573–586. doi:10.1111/tgis.12622.

Horne, J.S., Garton, E.O., Krone, S.M., Lewis, J.S., 2007. Analyzing animal movements using Brownian bridges. Ecology 88, 2354–2363. doi:10.1890/06-0957.1.

Joo, R., Etienne, M.p., Bez, N., Mahévas, S., 2018. Metrics for describing dyadic movement : a review. Movement Ecology 6, 1–17. doi:https://doi.org/10.1186/s40462-018-0144-2.

Kenward, R.E., Marcström, V., Karlbom, M., 1993. Post-nestling behaviour in goshawks, *Accipiter gentilis*: II. sex differences in sociality and nest-switching. Animal Behaviour 46, 371–378.

Kitchin, R., 2020. Using digital technologies to tackle the spread of the coronavirus: Panacea or folly? The Programmable City Working Paper 44, 1–24.

Konzack, M., McKetterick, T., Ophelders, T., Buchin, M., Giuggioli, L., Long, J., Nelson, T., Westenberg, M.A., Buchin, K., 2017. Visual analytics of delays and interaction in movement data. International Journal of Geographical Information Science 31, 320–345. doi:10.1080/13658816.2016.1199806.

Laube, P., 2014. Computational Movement Analysis. SpringerBriefs in Computer Science. 1 ed., Springer International Publishing. doi:10.1007/978-3-319-10268-9.

Long, J., Nelson, T., 2015. Home sange and habitat analysis using dynamic time geography. Journal of Wildlife Management 79, 481–490. doi:10.1002/jwmg.845.

Long, J.A., Nelson, T.A., 2013. Measuring Dynamic Interaction in Movement Data. Transactions in GIS 17, 62–77. doi:10.1111/j.1467-9671.2012.01353.x.

Long, J.A., Nelson, T.A., Webb, S.L., Gee, K.L., 2014. A critical examination of indices of dynamic interaction for wildlife telemetry studies. Journal of Animal Ecology 83, 1216–1233. doi:10.1111/1365-2656.12198.

Long, J.A., Webb, S.L., Nelson, T.A., Gee, K.L., 2015. Mapping areas of spatial-temporal overlap from wildlife tracking data. Movement Ecology 3, 1–14. doi:10.1186/s40462-015-0064-3.

Long, J.A., Weibel, R., Dodge, S., Laube, P., 2018. Moving ahead with computational movement analysis. International Journal of Geographical Information Science 32, 1275–1281. doi:10.1080/13658816.2018.1442974.

Miller, H.J., 2005. A Measurement Theory for Time Geography. Geographical Analysis 37, 17–45. doi:10.1111/j.1538-4632.2005.00575.x.

Miller, H.J., Dodge, S., Miller, J., Bohrer, G., 2019. Towards an integrated science of movement: converging research on animal movement ecology and human mobility science. International Journal of Geographical Information Science 33, 855–876. doi:10.1080/13658816.2018.1564317.

Miller, J.A., 2015. Towards a Better Understanding of Dynamic Interaction Metrics for Wildlife: a Null Model Approach. Transactions in GIS 19, 342–361.

NuStats, 2013. 2010-2012 California Household Travel Survey Final Report. Technical Report. California Department of Transportation.

Oliver, N., Lepri, B., Sterly, H., Lambiotte, R., Delataille, S., De Nadai, M., Letouzé, E., Salah, A.A., Benjamins, R., Cattuto, C., Colizza, V., de Cordes, N., Fraiberger, S.P., Koebe, T., Lehmann, S., Murillo, J., Pentland, A., Pham, P.N., Pivetta, F., Saramäki, J., Scarpino, S.V., Tizzoni, M., Verhulst, S., Vinck, P., 2020. Mobile phone data for informing public health actions across the COVID-19 pandemic life cycle. Science Advances 6. doi:10.1126/sciadv.abc0764.

Potts, J.R., Mokross, K., Lewis, M.A., Dina, P., 2014. A unifying framework for quantifying the nature of animal interactions. Journal of the Royal Society Interface 11. doi:10.1098/rsif.2014.0333, arXiv:1402.1802.

Powell, R.A., Mitchell, M.S., 2012. What is a home range? Journal of Mammalogy 93, 948–958. doi:10.1644/11-MAMM-S-177.1.

Sainz, F., 2020. Apple and Google partner on COVID-19 contact tracing technology. URL: https://nr.apple.com/d2s6h070n0.

Simcharoen, A., Savini, T., Gale, G.A., Simcharoen, S., Duangchantrasiri, S., Pakpien, S., Smith, J.L., 2014. Female tiger panthera tigris home range size and prey abundance: important metrics for management. Oryx 48, 370–377.

Smith, D.J.L., McDougal, C., Miquelle, D., 1989. Scent marking in free-ranging tigers, Panthera tigris. Animal behavior 37, 1–10.

Worton, B.J., 1987. A review of models of home range for animal movement. Ecological Modelling 38, 277–298. doi:10.1016/0304-3800(87)90101-3.