China Data Institute | Center for Geographic Analysis, Harvard University | 社会地理计算联合研究中心 Geocomputation Center for Social Science | NSF Spatiotemporal Innovation Center

计算时间地理学研究

Computational Time Geographic Studies in the Era of Big Data

陈碧宇

武汉大学
测绘遥感信息工程国家重点实验室
社会地理计算联合研究中心

Background Time geographic analysis

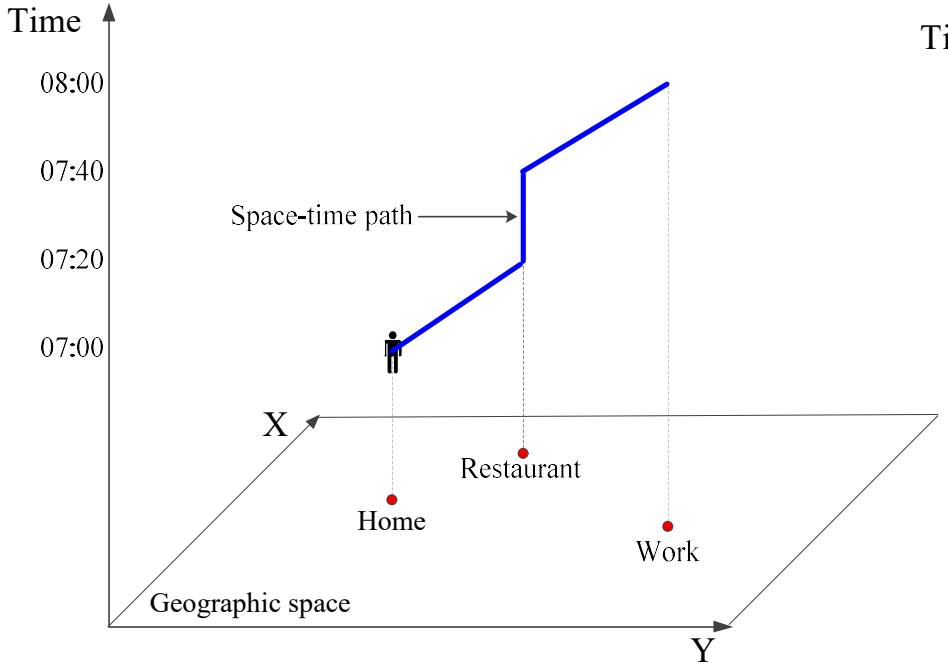
- Time geography
  - Introduced by Torsten Hägerstrand, one of the most important geographers in the 20th century
  - The need to study the individual in order to understand social and group practices
  - A link between space and time that had previously been poorly developed.
  - Elegant notation system for analyzing human behaviors

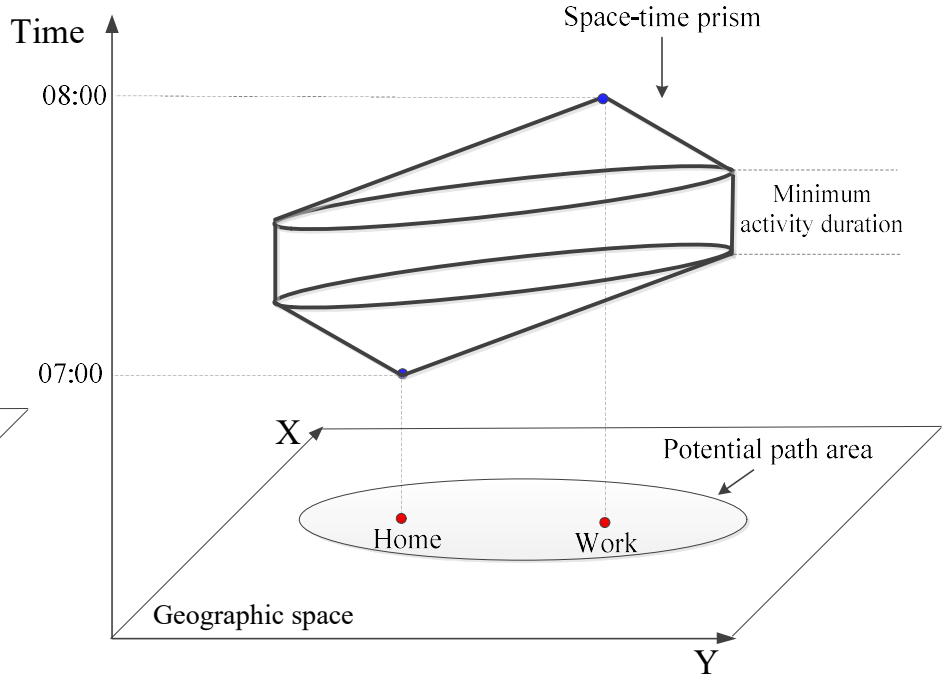
Torsten Hägerstrand
(1916 - 2004)

Hägerstrand, T., 1970, What about people in regional science? Papers in Regional Science, 24, pp. 7-24.

Background Time geographic analysis

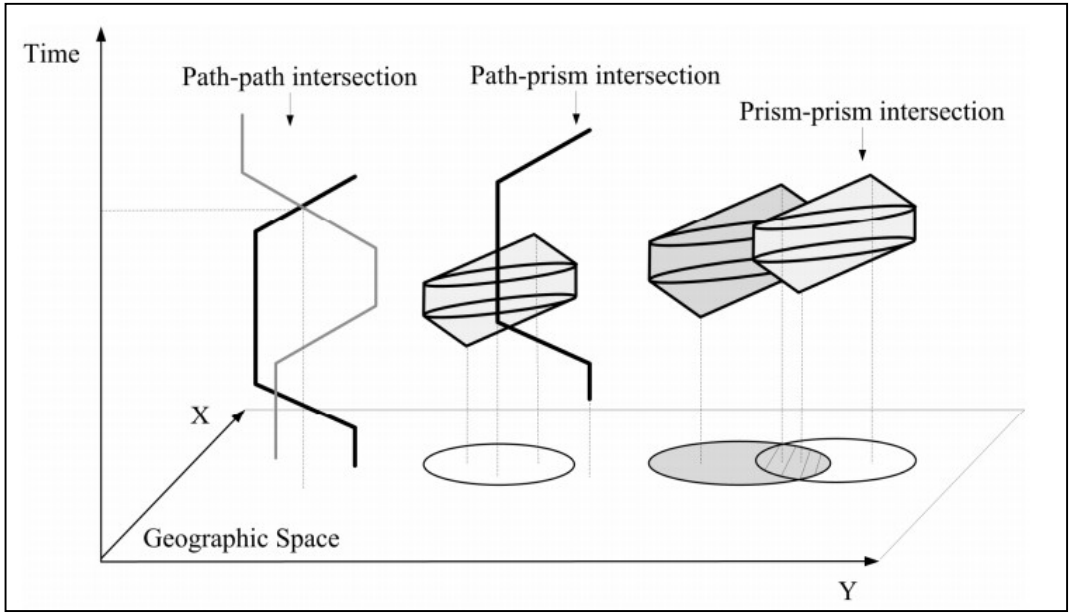
- Elegant notation system for analyzing human behaviors
  - Space-time entities: space-time path, space-time prism, etc.

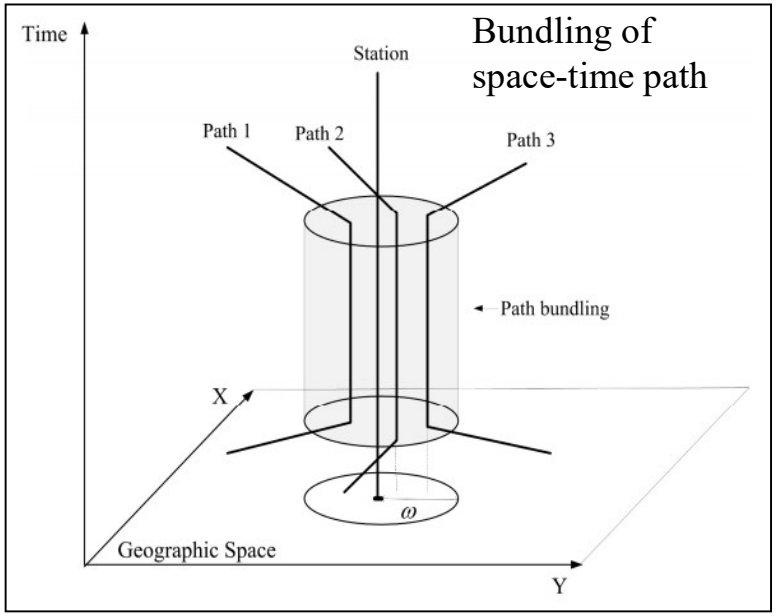




Background Time geographic analysis

- Elegant notation system for analyzing human behaviors
  - Space-time entities: space-time path, space-time prism, etc.
  - Space-time relationships: space-time intersection, bundling



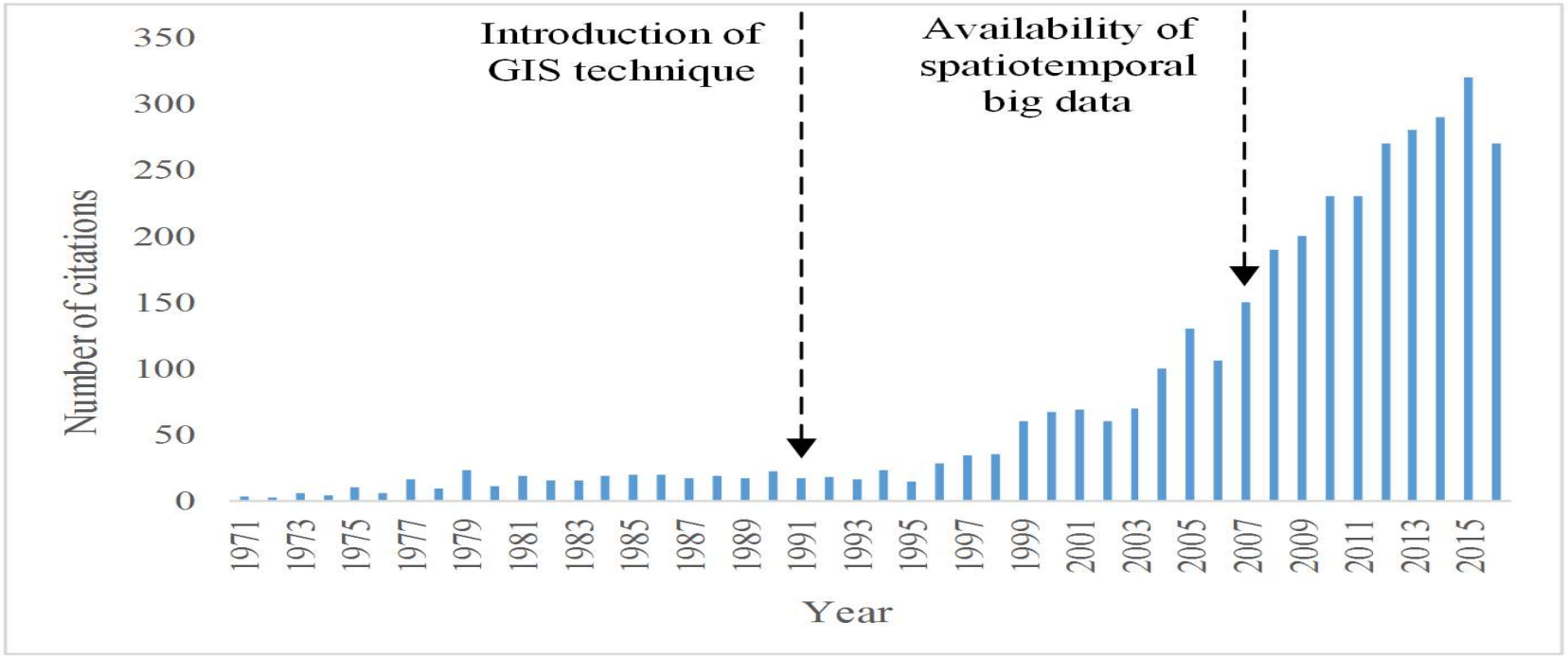


Background Time geographic analysis

- Space-time path analysis
  - Moving objects database
  - Trajectory clustering, classification, pattern recognition, etc.
- Space-time prism analysis
  - Accessibility
  - Social exclusion
  - Social inequity
  - Social segregation
  - Etc.

Background Time geographic analysis

- Conceptual time geographic studies (1970s – 1980s)
- Quantitative time geographic studies (1990s –)
- Computational time geographic studies in the era of big data?



Google scholar citations to the paper of Hägerstrand (1970)

Hägerstrand, T., 1970, What about people in regional science? Papers in Regional Science

Background Big data challenges

- Emerging spatiotemporal big data
  - Floating car data (e.g, 14,000 taxi cabs in Shenzhen)
  - Mobile phone records (e.g., 10 millions of phone users in Shenzhen)
  - Smart car data (e.g., 12,000 buses in Shenzhen), etc.
- Opportunities for time geographic analysis
  - Move from place-based to people-based studies
  - Toward data-driven geographic studies
- Big data challenges for time geographic analysis
  - How to represent, store, manage, index and query time geographic entities (both space-time path and prism) and relationships for knowledge discovery?

Literature Review Data models in GIS

- Spatiotemporal data models in GIS
  - Snapshots model (Armstrong, 1988)
  - Base state with amendments data model (Chen, 1995)
  - Spatiotemporal composite model (Langran and Chrisman, 1988)
  - Three-domain data model (Yuan et al., 1999)
  - Geo-atom model (Goodchild et al., 2007)
  - ……

These spatiotemporal data models are not tailored to the representation of network time geographic entities and relations

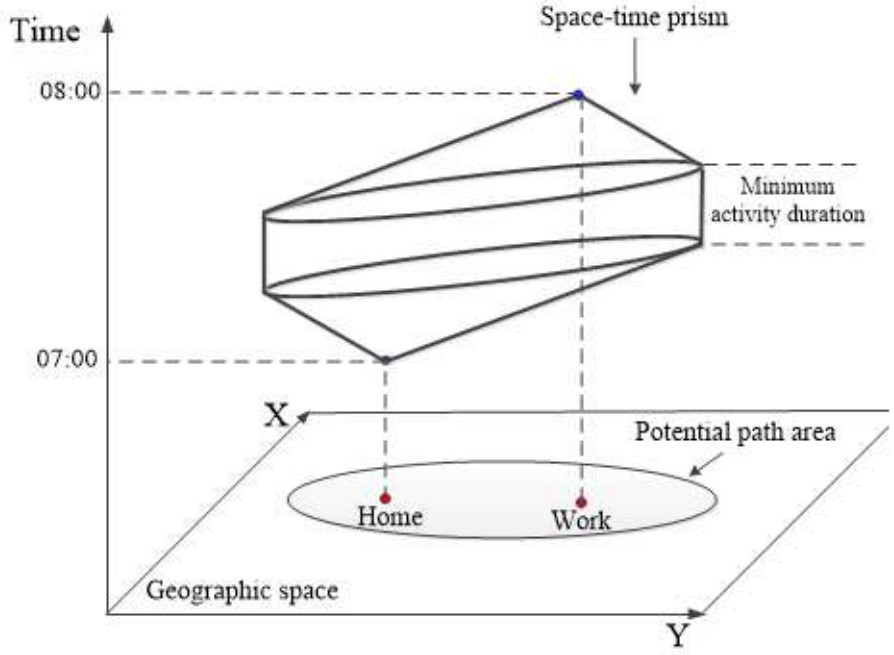
Literature Review Moving objects models

- Moving objects data models
  - Vazirgiannis's model (2001)
  - FNR-tree model (Frentzos, 2003)
  - Mon-tree model (Almeida and Güting, 2005; Güting, 2006)
  - ……
- Limitations
  - Only model space-time path, but ignore space-time prism, etc.
  - Seldom integrate with the GIS platforms

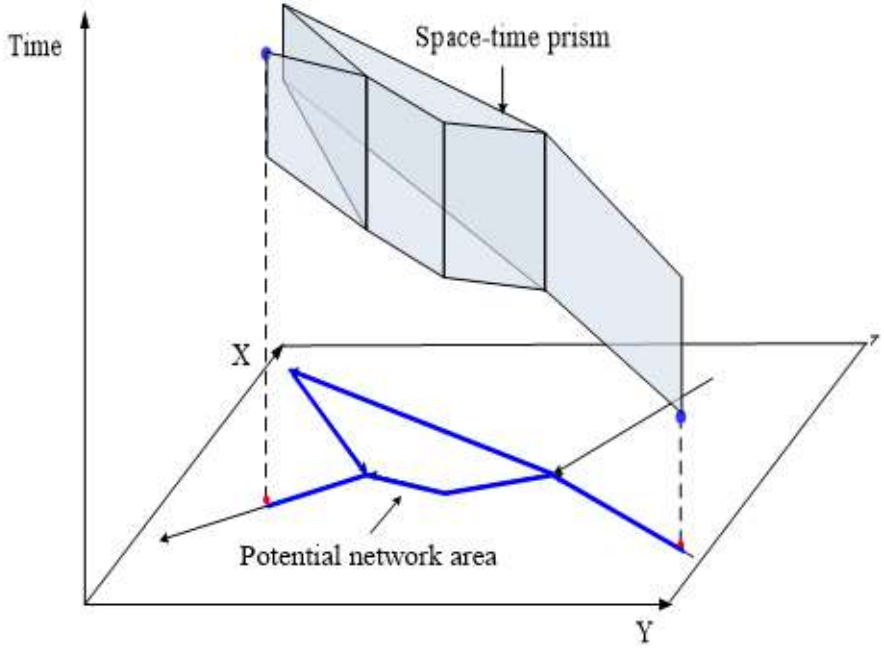
Research Framework

- Part 1. Spatiotemporal data model
  - Represent, store, index and query time geographic entities
  - Analyze space-time intersection relationships
- Part 2. Spatiotemporal index structure
- Part 3. Spatiotemporal proximity analysis
- Part 4. Case study

Part 1 Network-constrained time geographic entities



Space-time prism in the planar space

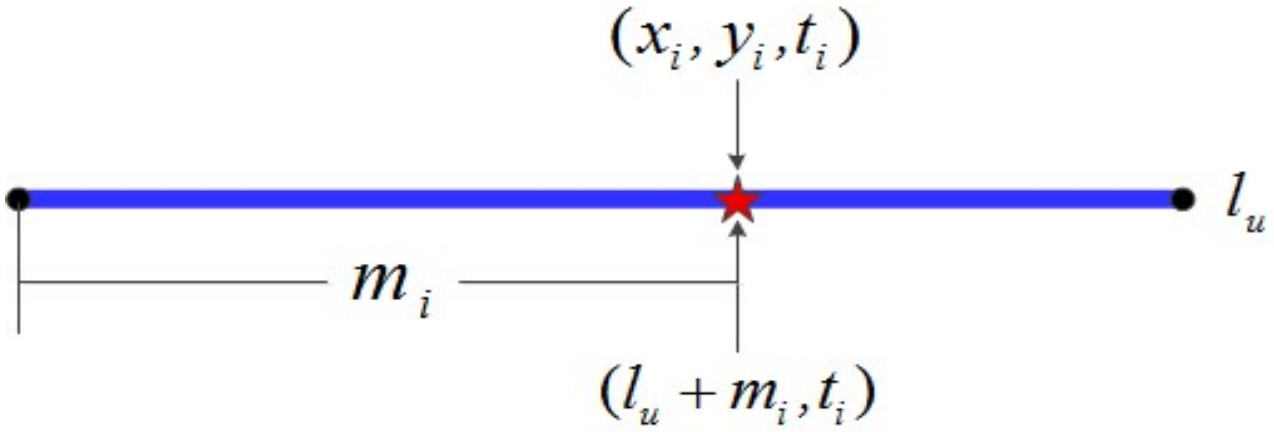


Network space-time prism in the (x, y, t) space

Time geographic entities are 3 dimensional entities cannot be stored and managed in existing GIS platforms.

Part 1 Compressed linear referencing

Linear referencing system (LRS), is a method of spatial referencing, in which the locations of features are described in terms of measurements along a linear element.



Compressed linear referencing (CLR)

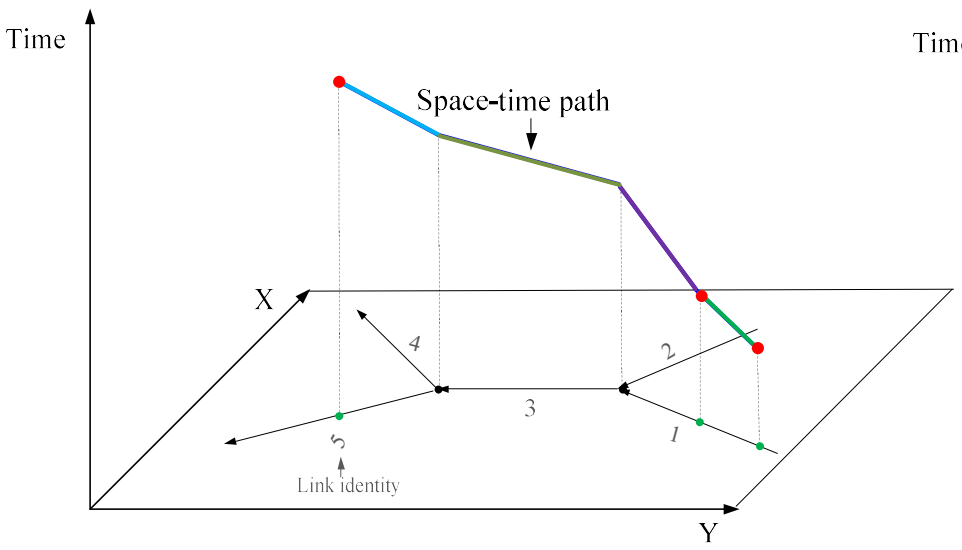
$$(x_i, y_i, t_i) \Longrightarrow (l_u + m_i, t_i)$$

3D

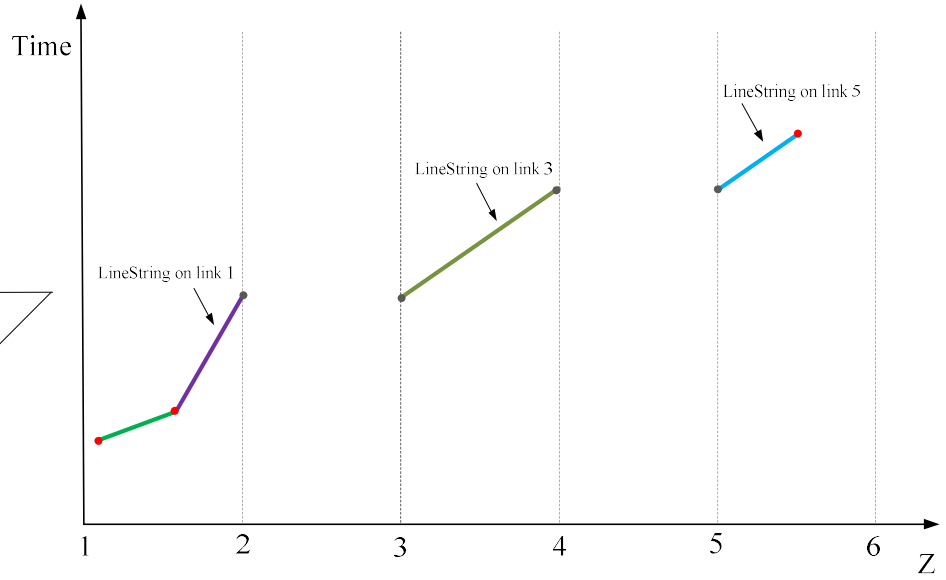
2D

Part 1 Space-time path in the CLR space

Proposition 1. The space-time path $\tilde{P}^q$ in CLR space is an equivalent representation of space-time path P^q in (x, y, t) space.



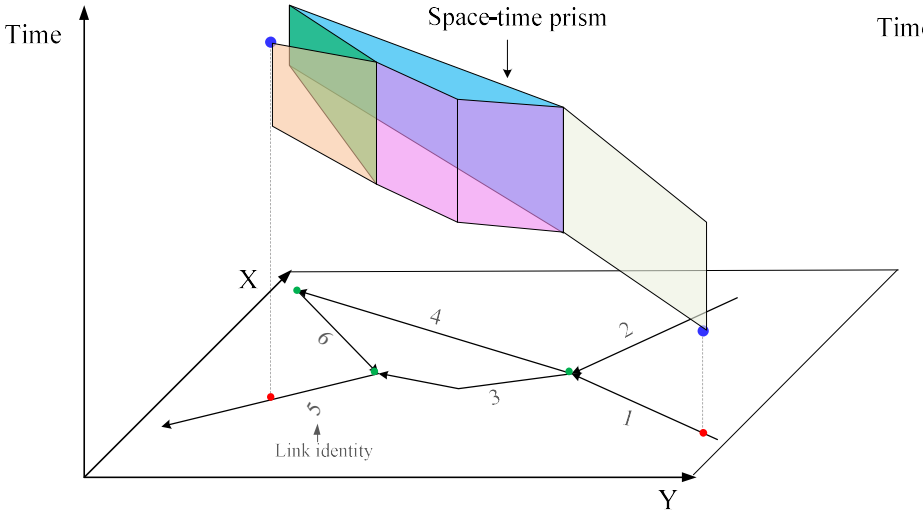
Network space-time path in the (x, y, t) space



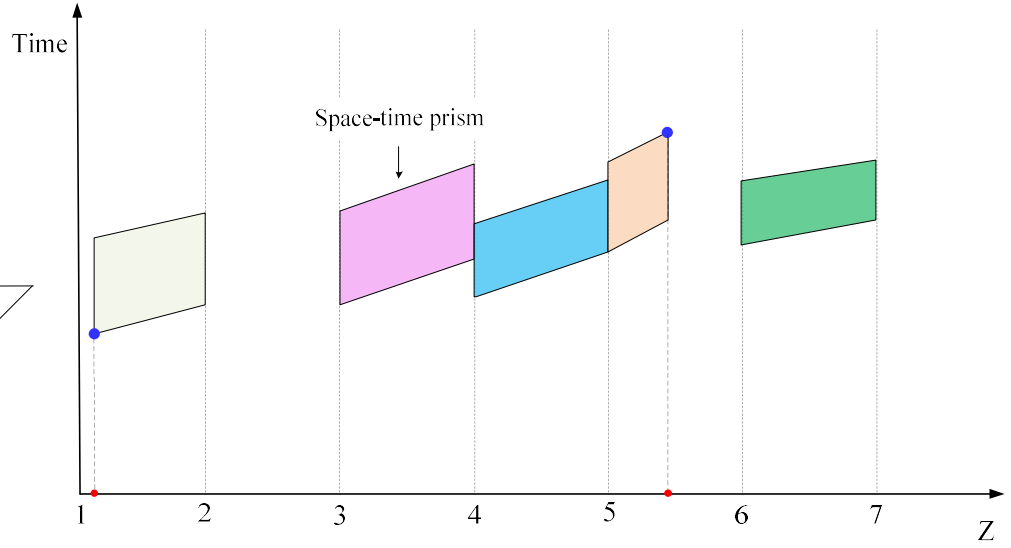
Network space-time path in the CLR space

Part 1 Space-time prism in the CLR space

Proposition 2. The network space-time prism $\widetilde{STP}_{od}$ in CLR space is an equivalent representation of the network space-time prism STP_{od} in (x, y, t) space.



Network space-time prism in the (x, y, t) space



Network space-time prism in the CLR space

Part 1 Intersection relations in the CLR space

Relations	Path-Path	Path-Prism	Prism-Prism
A Contains B			
A Within B			
A Touches B			
A Overlaps B		-	
A Crosses B			

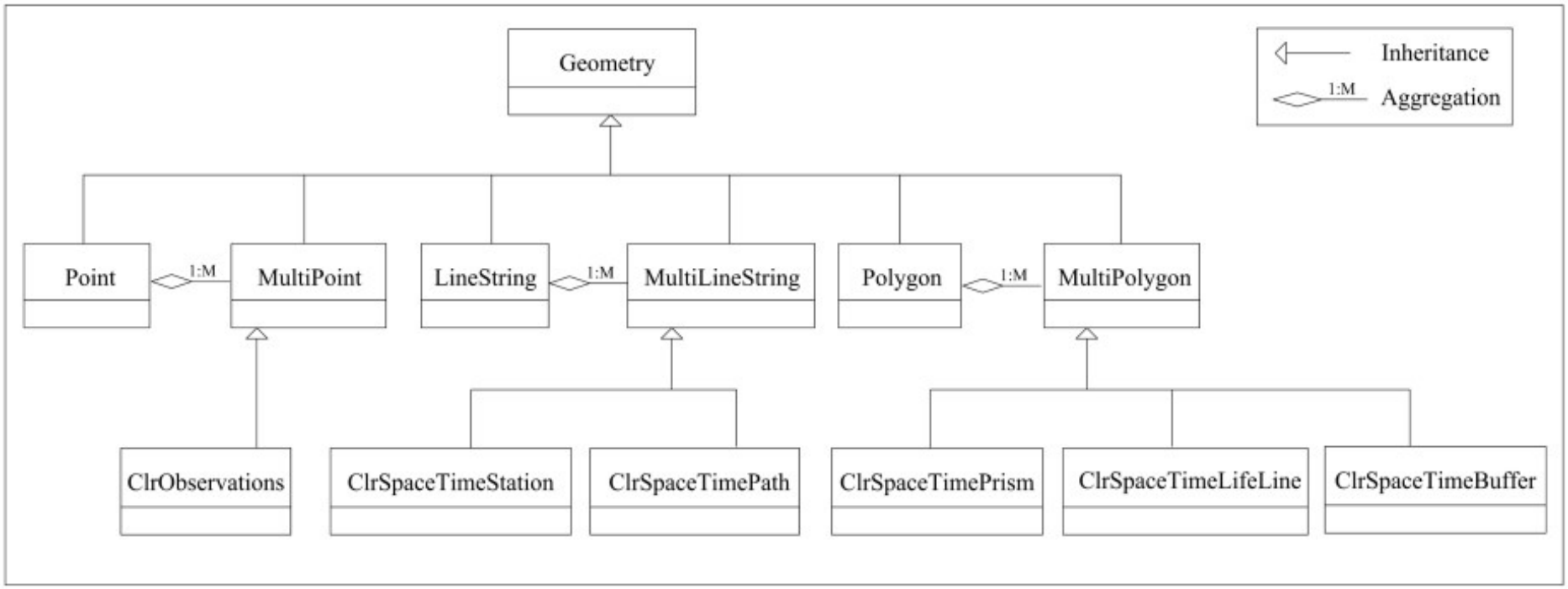
3D

Spatiotemporal relations

2D

Spatial relations

Part 1 Spatiotemporal data model



Spatiotemporal data model for network time geographic entities in the CLR space

3D
spatiotemporal entities

2D
spatial geometries

Part 1 Spatiotemporal database design

Storage

- Classical spatial databases (i.e. ESRI Geodatabase , Oracle Spatial)
- Space-time entities (Feature layer)
  - Space-time paths (Polyline feature layer)
  - Space-time prisms (Polygon feature layer)

Spatial operations

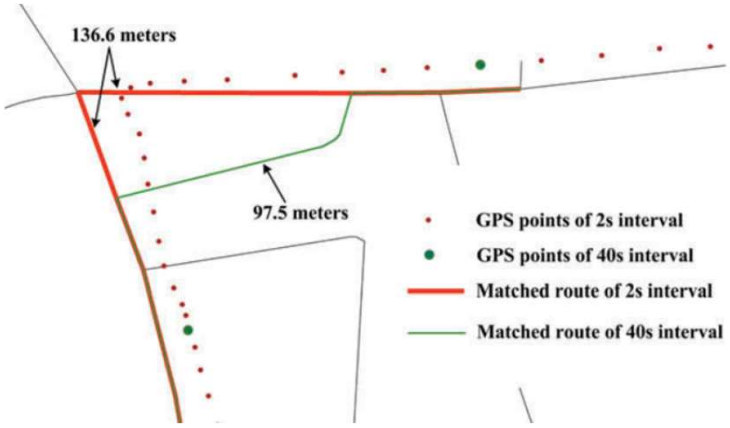
- Spatial operations in the current GIS platforms can be directly utilized
- Intersection operation (i.e., IRelationalOperator, ITopologicalOperator)

Index

- Grid-file index (i.e., ArcGIS)
- R-tree Index (i.e., ArcGIS)

Part 1 Computational experiments

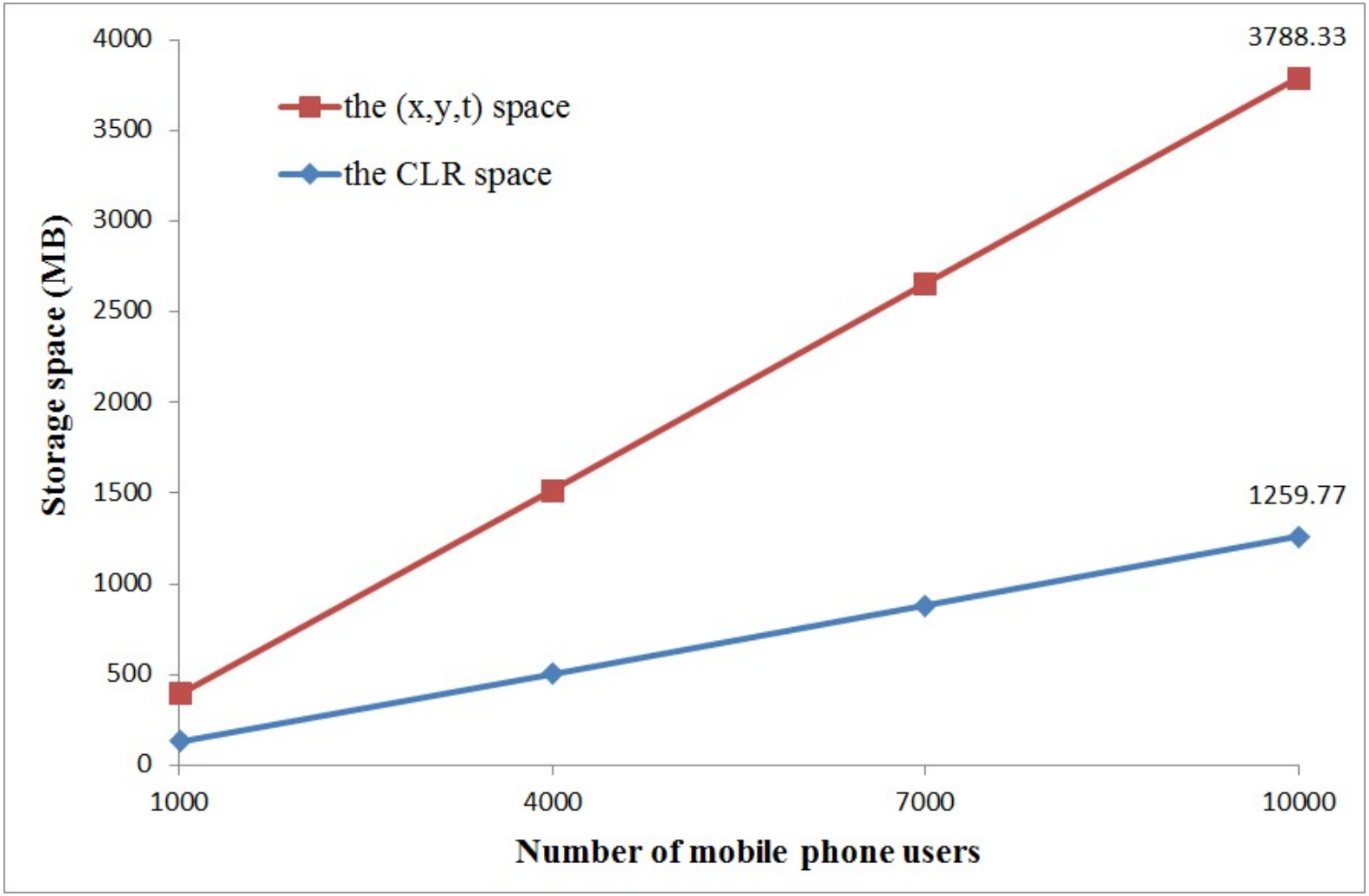
- Database platforms
  - Spatial database: Oracle database 11g R2
  - SDE: ArcSDE 10.2
  - 3D Visualization: ArcScene 10.2
- Datasets
  - Shenzhen Network: 69,198 nodes, 192,582 links
  - Space-time paths: 13,455 taxi trajectories
  - Space-time prisms: hourly locations of 10,000 mobile users



Chen, B.Y.*, et al., 2014, Map matching algorithm for large-scale low-frequency floating car data. *International Journal of Geographical Information Science*, 28(1), 22-38.

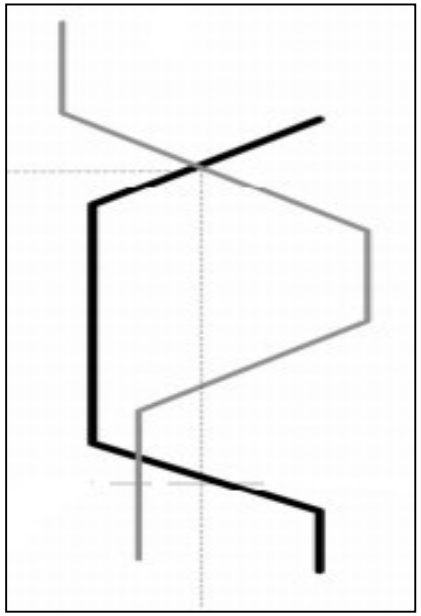
Part 1 Storage performance

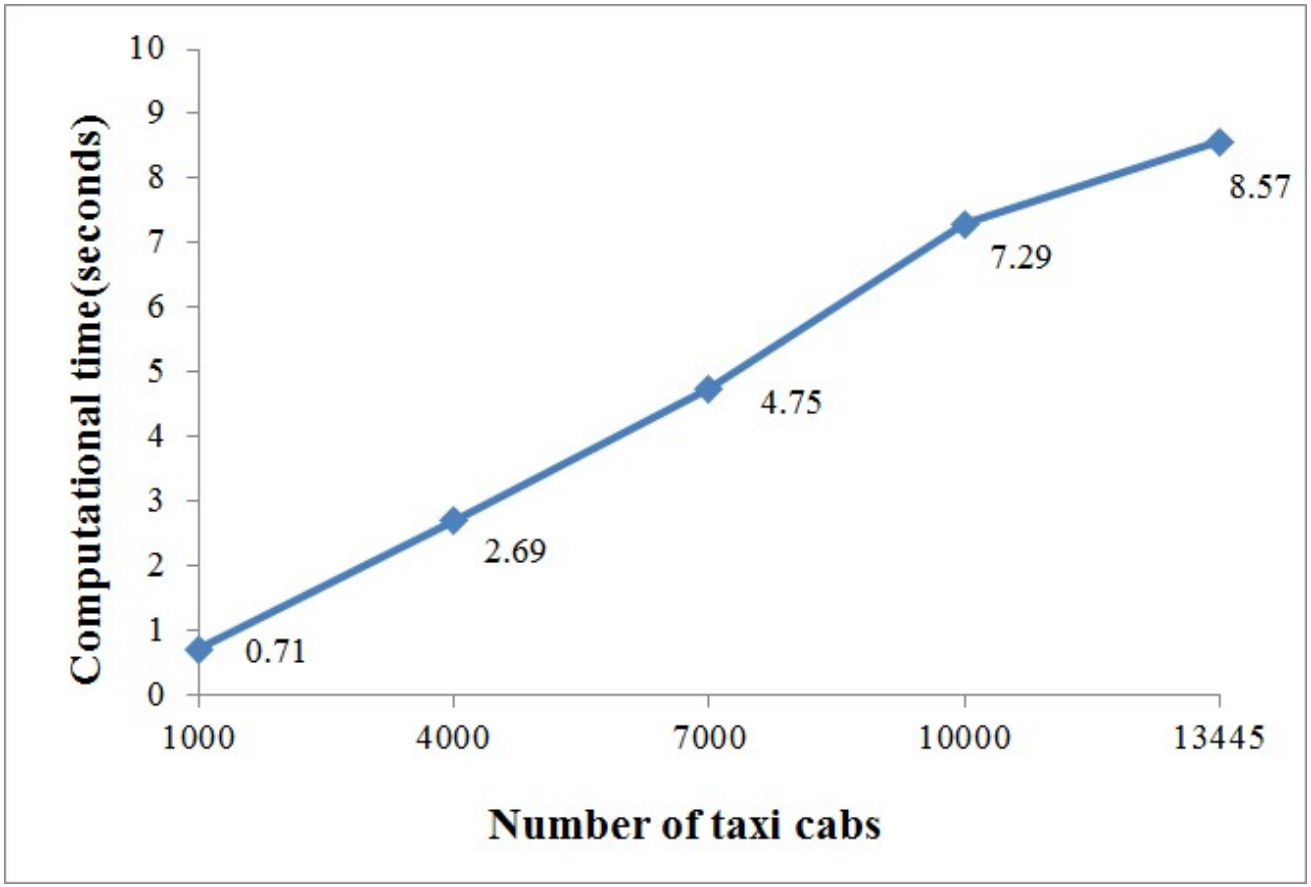
CLR space saved 67% of storage space for storing space-time prisms



Part 1 Path-path intersection query performance

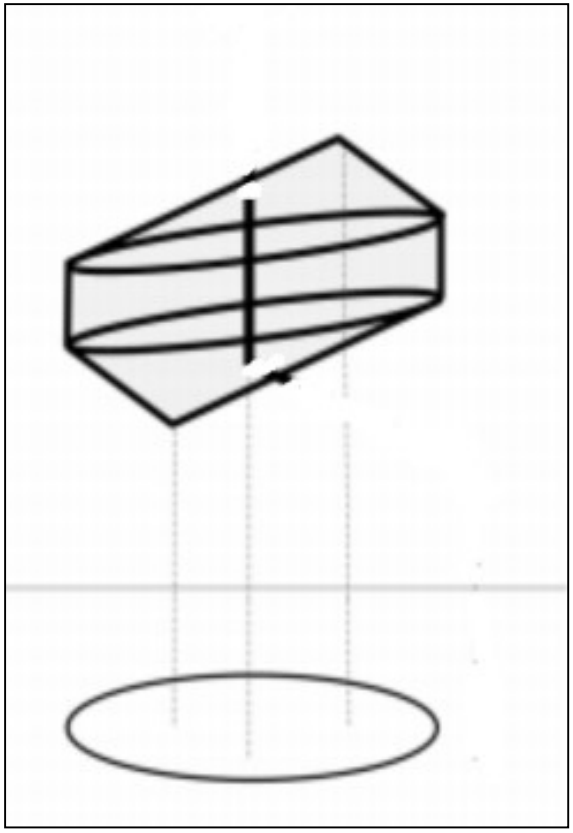
Path-path intersection query. Given a space-time path, $\tilde{P}^q$, of an individual, find all space-time paths intersecting the path.

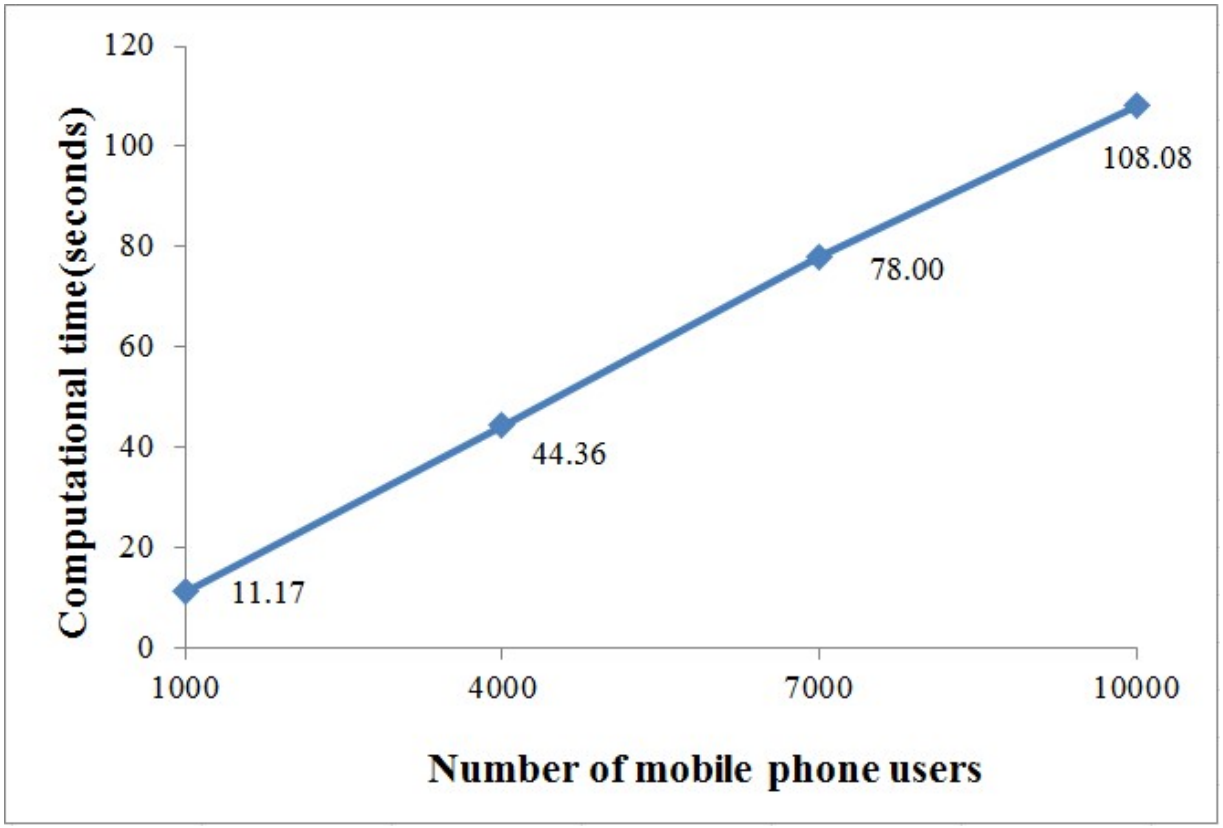




Part 1 Station-prism intersection query performance

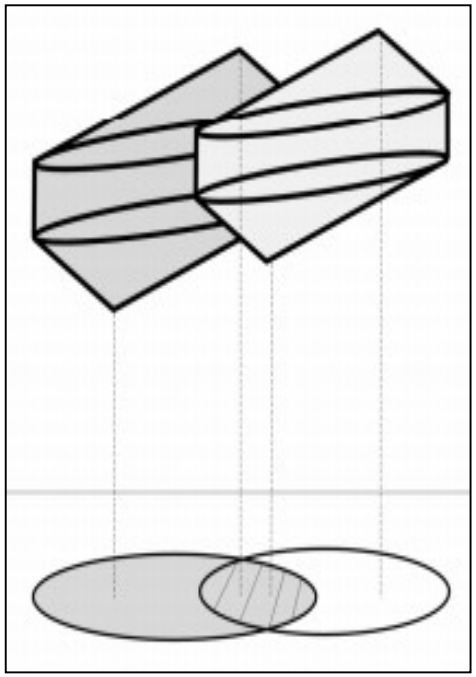
Station-prism within query. Given a location, z_i^u, and a time period, $[t_o,t_d]$, find all space-time prisms containing the location z_i^u for the time period $[t_o,t_d]$.

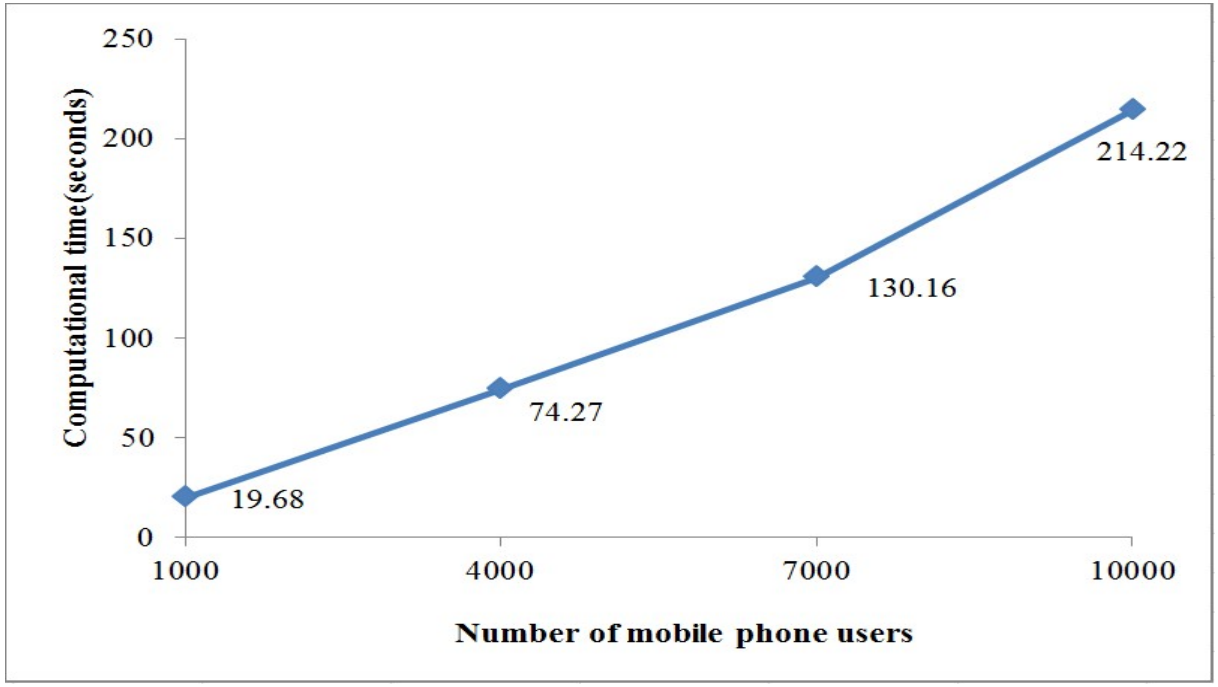




Part 1 Prism-prism intersection query performance

Prism-prism intersection query. Given a space-time prism, STP_{od}, of an individual, find all space-time prisms that intersect with the given prism $\widetilde{STP}_{od}$.



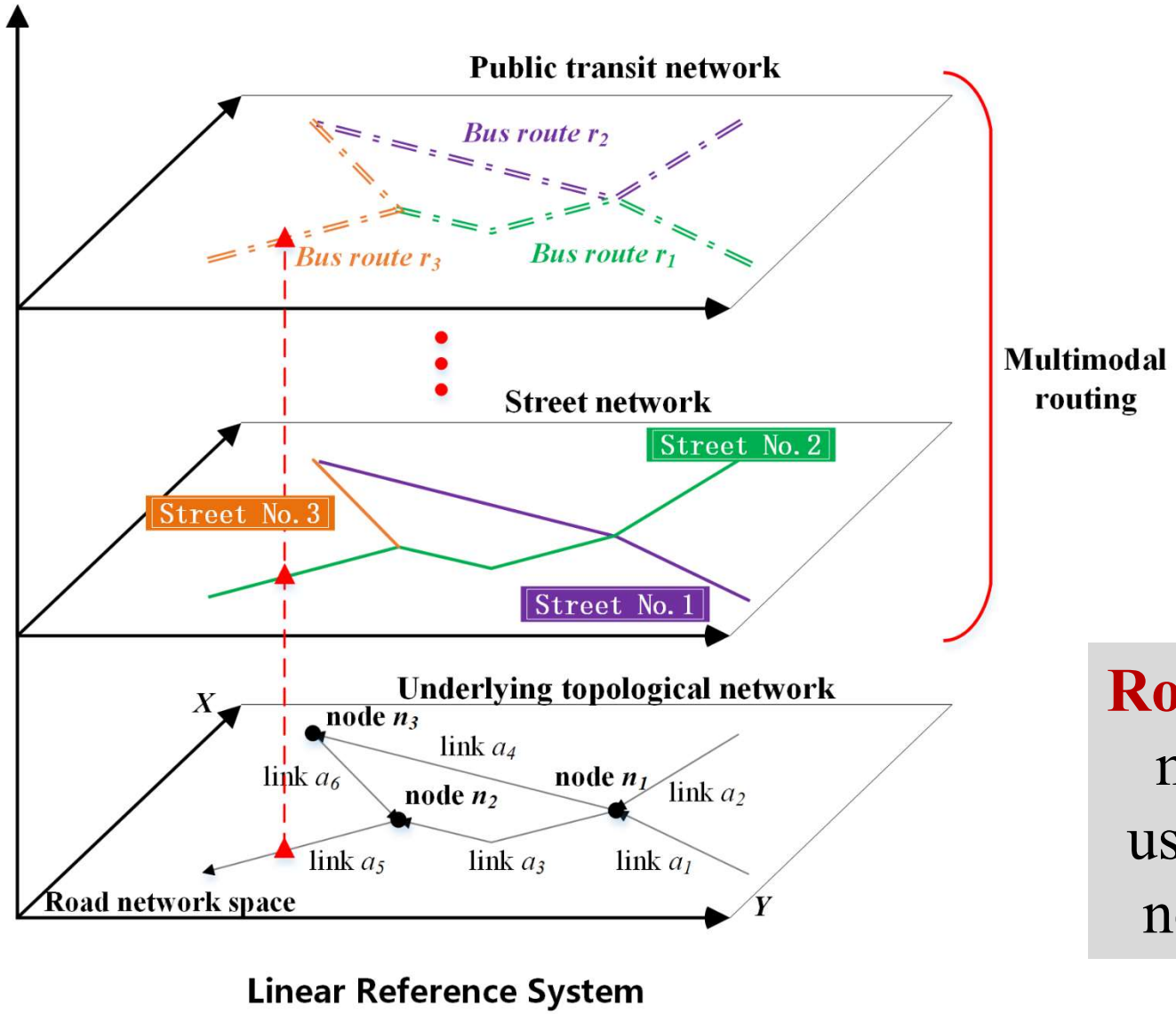


Chen, B.Y.*, et al., 2016, Spatiotemporal data model for network time geographic analysis in the era of big data. *International Journal of Geographical Information Science*, 30, pp. 1041-1071.

Research Framework

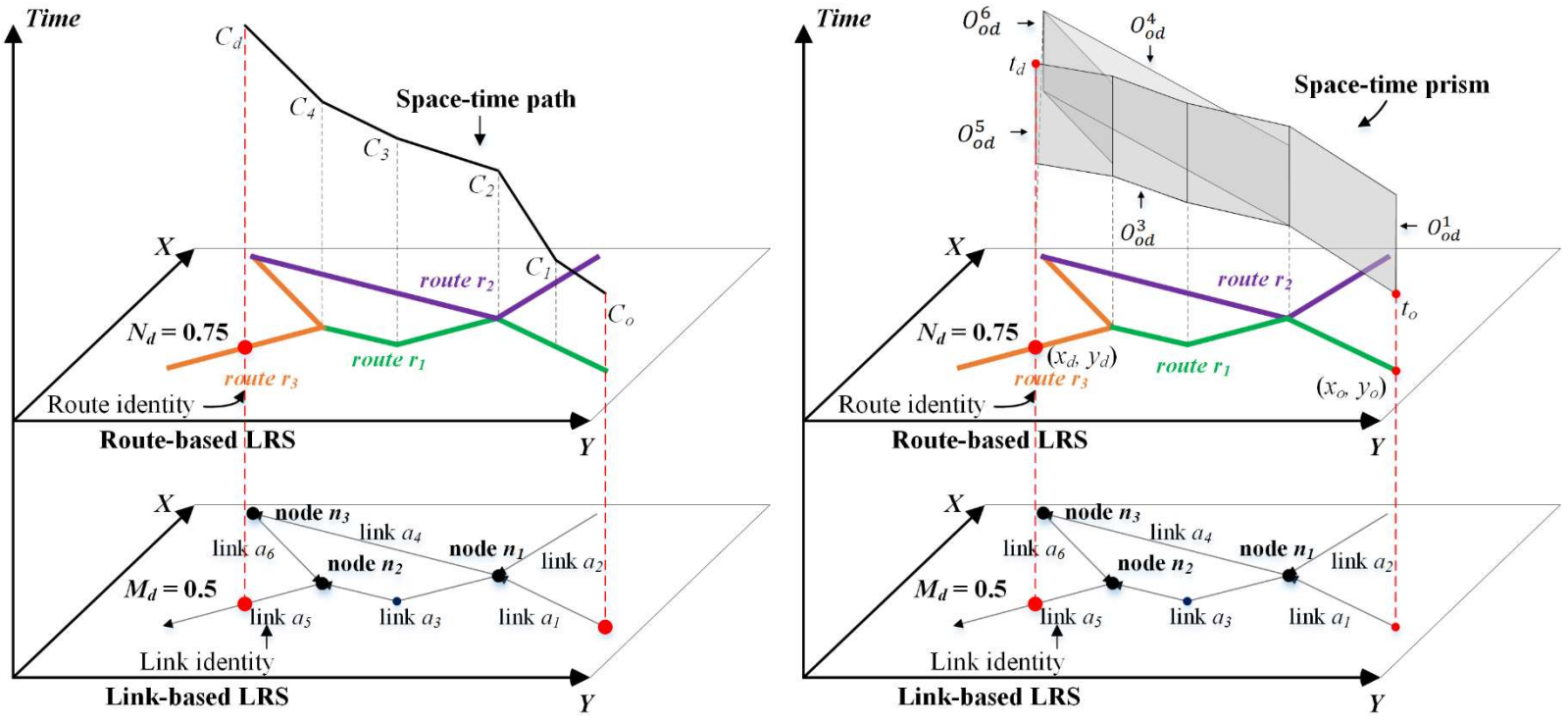
- Part 1. Spatiotemporal data model
- **Part 2. Spatiotemporal index structure**
  - **Route-based CLR model**
  - **LRS-based spatiotemporal index structure**
- Part 3. Spatiotemporal proximity analysis
- Part 4. Trajectory clustering in road network
- Part 5. Case study

Part 2 Route-based linear reference system



Route-based LRS is more commonly used to reference a network location.

Part 2 Route-based network-constrained entities

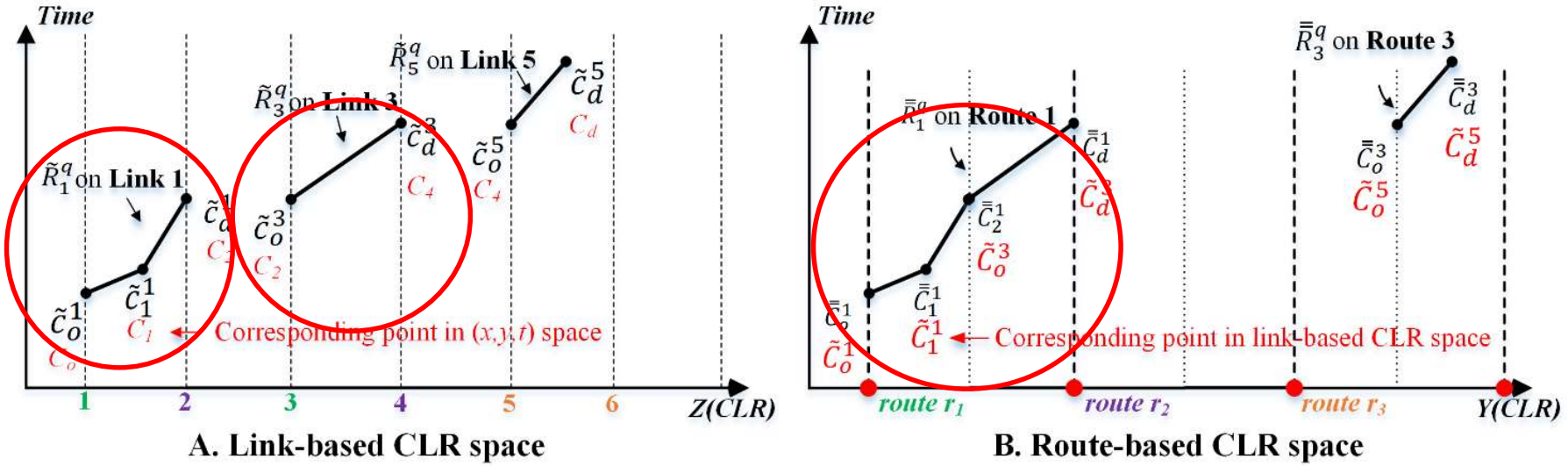


A. Route-based network-constrained space-time path **B. Route-based network-constrained space-time prism**

Route-based LRS(**s**) for locating time geographic entities.
Link-based LRS is formulated as the underlying linear datum.

Part 2 Path in the route-based CLR space

Proposition 1. The space-time path $\bar{\bar{P}}^q$ in route-based CLR space is an equivalent representation of space-time path $\tilde{P}^q$ in link-based CLR space.



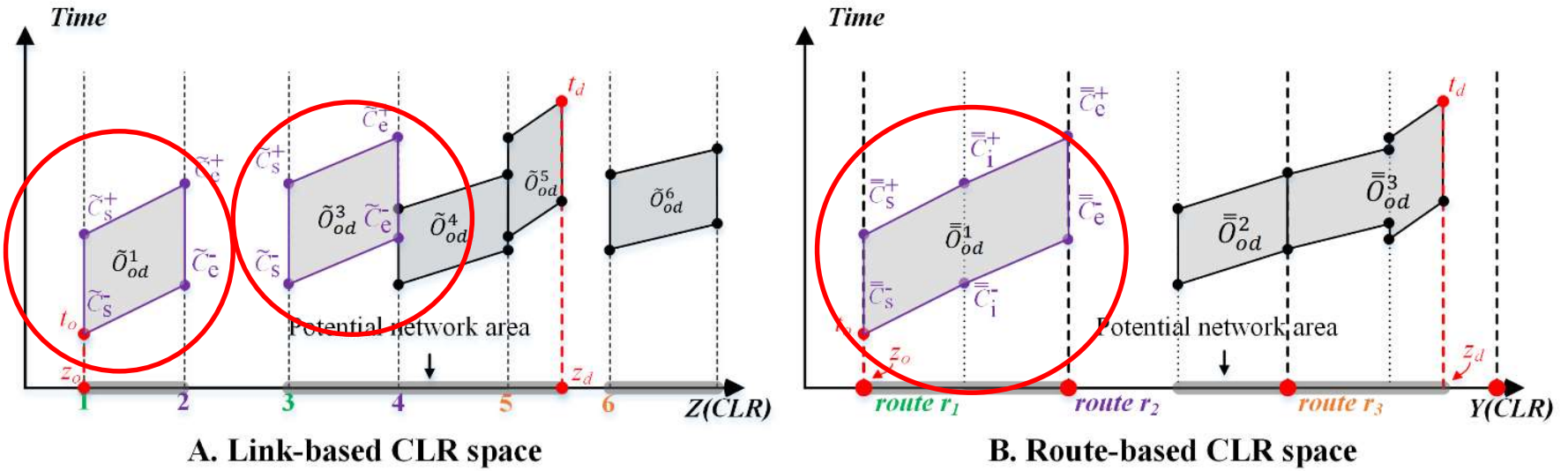
Network space-time path in link-based CLR space

Network space-time path in route-based CLR space

Route-based CLR model can reduce the number of sub-entities.

Part 2 Prism in the route-based CLR space

Proposition 2. The network space-time prism $\overline{\overline{STP}}_{od}$ in route-based CLR space is an equivalent representation of the network space-time prism $\widetilde{STP}_{od}$ in link-based CLR space.

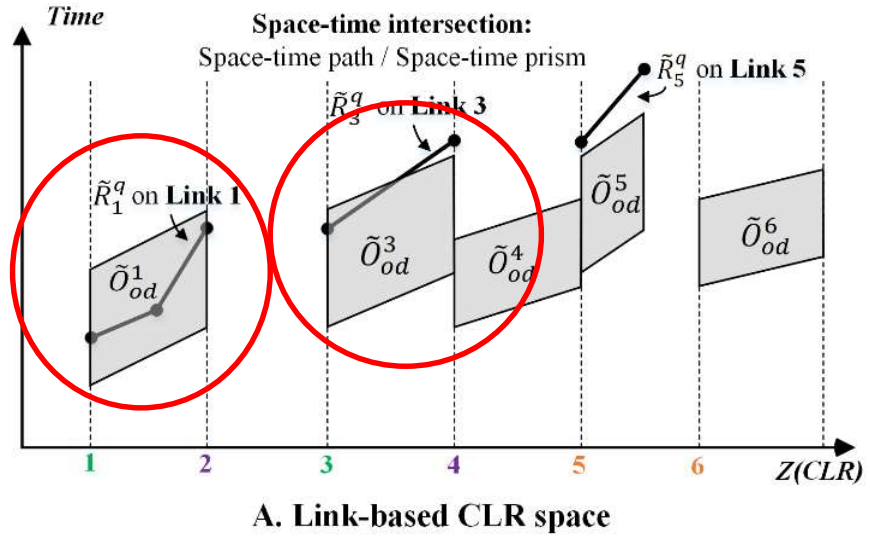


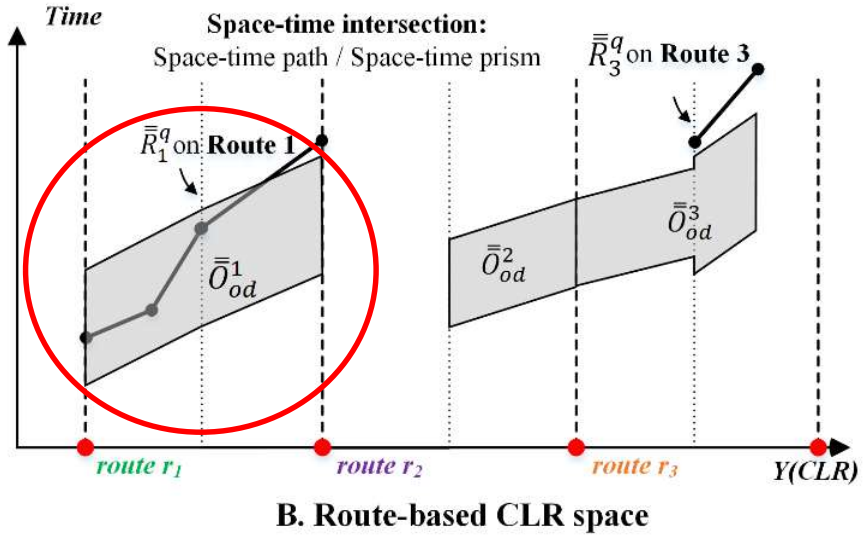
Network space-time prism in link-based CLR space Network space-time prism in route-based CLR space

Route-based CLR model can reduce the number of sub-entities.

Part 2 Intersection in the CLR space

The intersection results in the route-based CLR space are consistent with those in the link-based CLR space



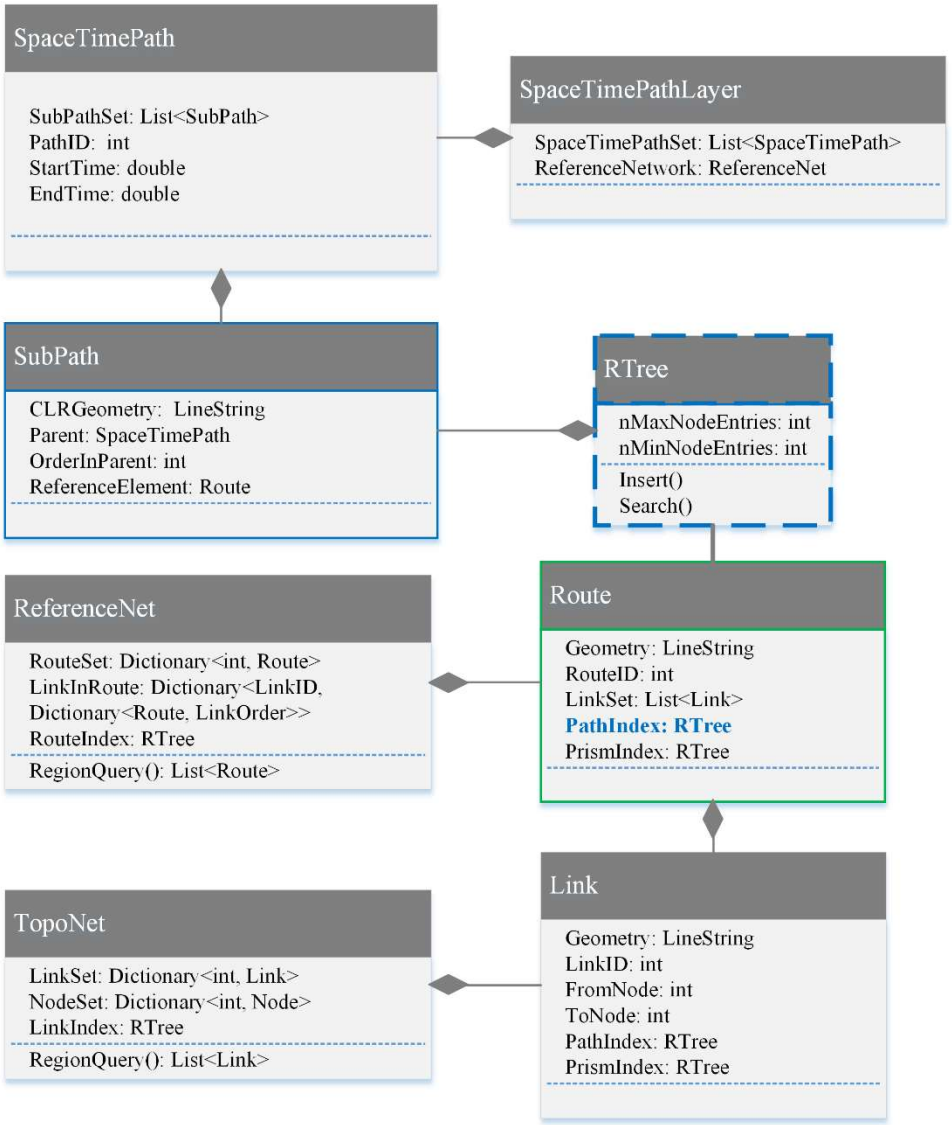


Space-time intersection in link-based CLR space

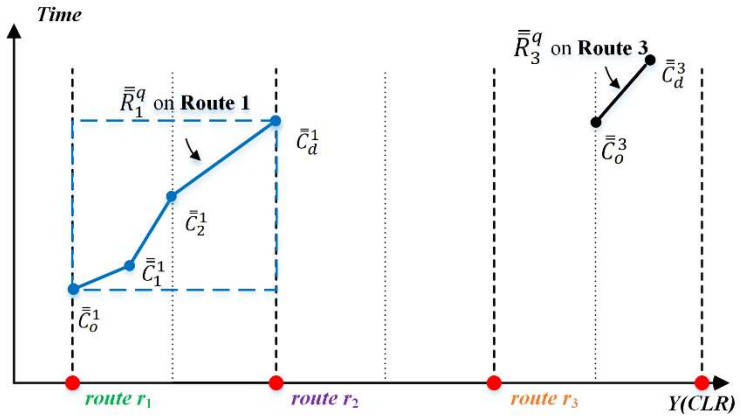
Space-time intersection in route-based CLR space

Route-based CLR model can improve the query performance.

Part 2 LRS-based spatiotemporal index



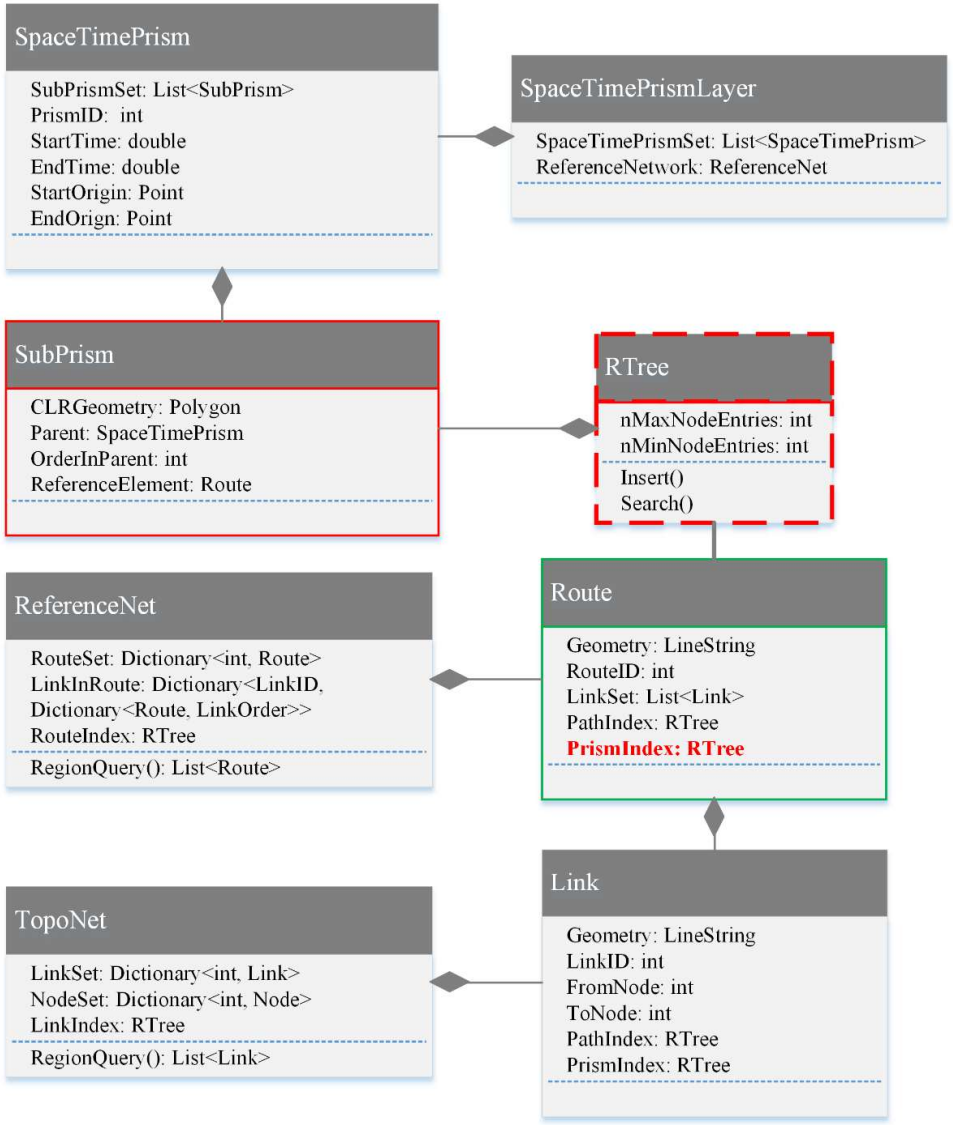
Class diagram



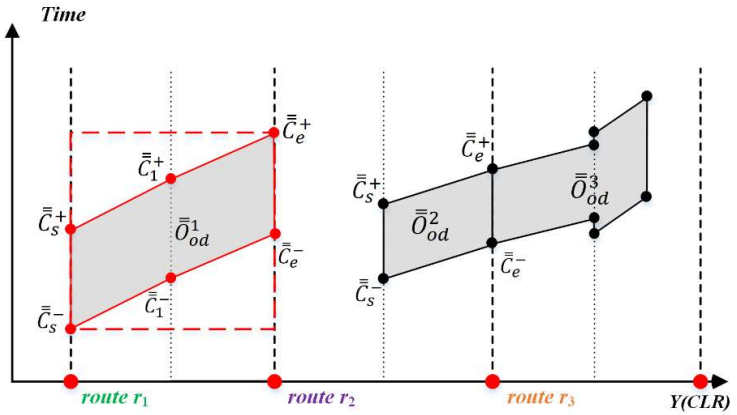
Network space-time path in route-based CLR space

R-Tree to index all sub-entities passing through the specific **route**

Part 2 LRS-based spatiotemporal index



Class diagram



Network space-time prism in route-based CLR space

R-Tree to index all sub-entities passing through the specific **route**

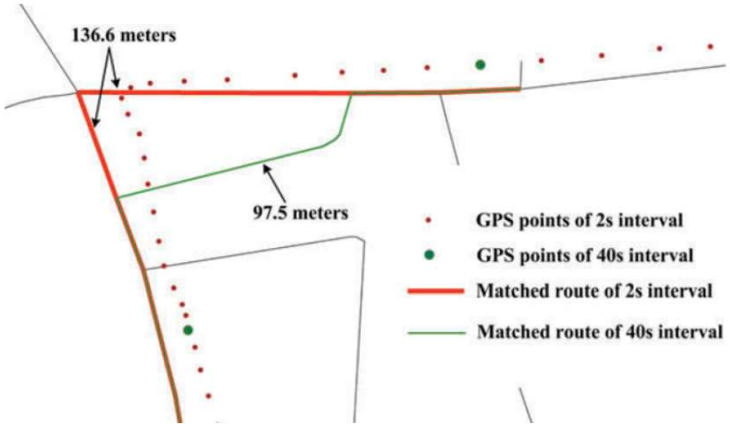
Part 2 Computational experiments

Database platforms

- Spatial database: Oracle database 11g R2
- SDE: ArcSDE 10.2

Datasets

- Beijing Network: 111,543 nodes, 304,434 links
- Space-time paths: 7,000 taxi trajectories (T-Drive dataset)
- Space-time prisms: 7000 artificially synthesized O-D pairs

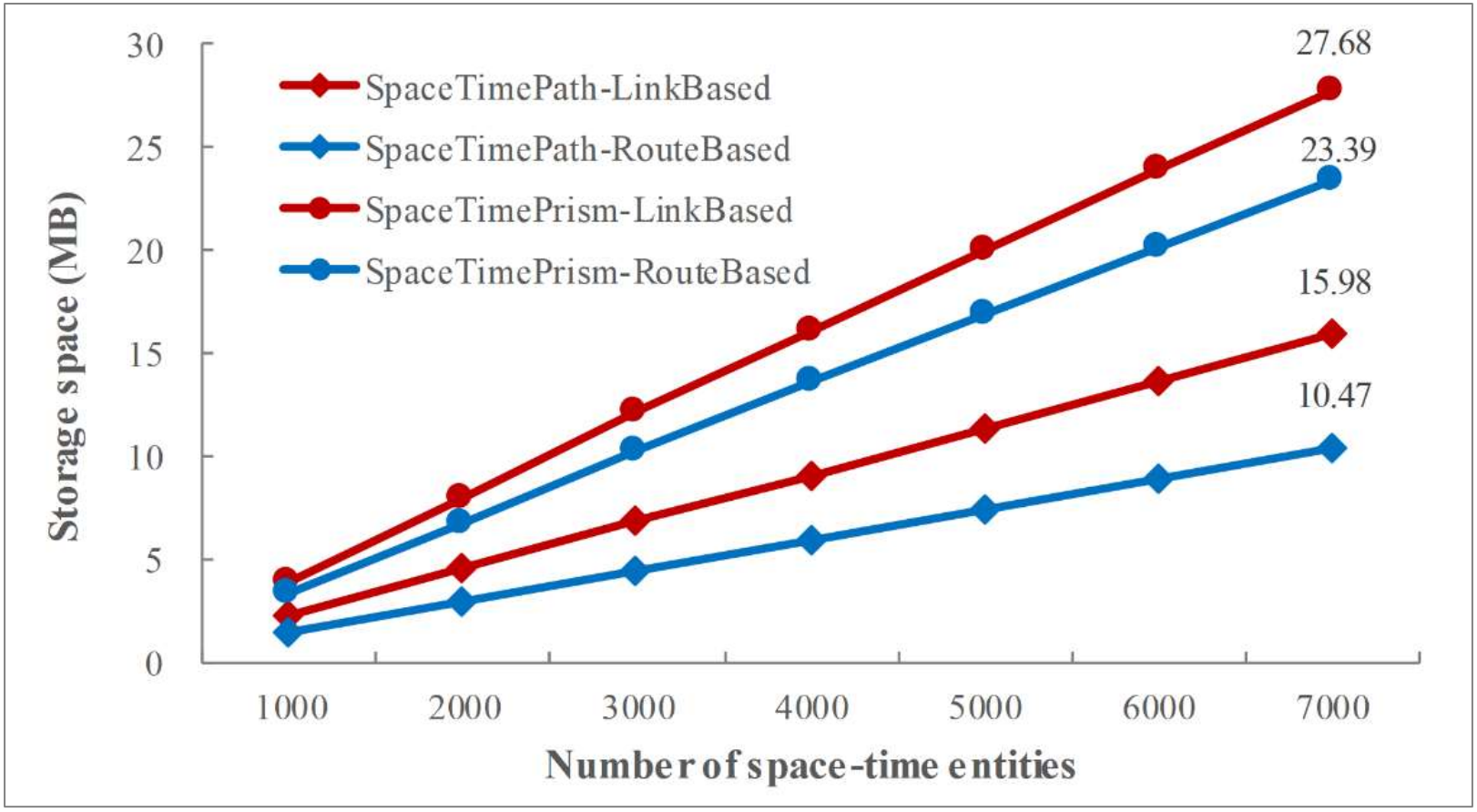


Chen, B.Y.*, et al., 2014, Map matching algorithm for large-scale low-frequency floating car data. *International Journal of Geographical Information Science*, 28(1), 22-38.

Part 2 Storage performance

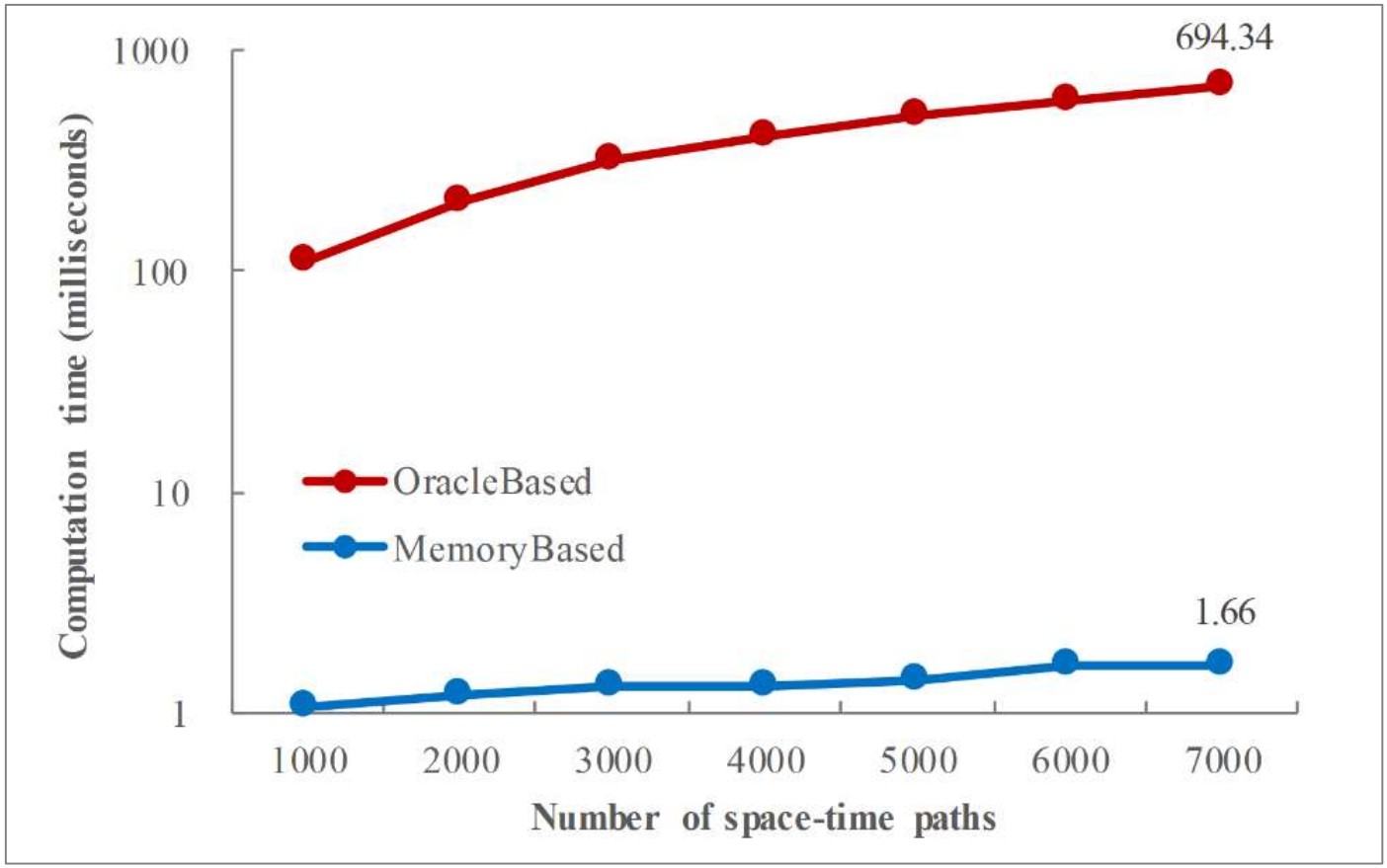
Route-based CLR space saved 35% of storage space for storing space-time paths

Route-based CLR space saved 16% of storage space for storing space-time prisms



Part 2 Query performance in memory

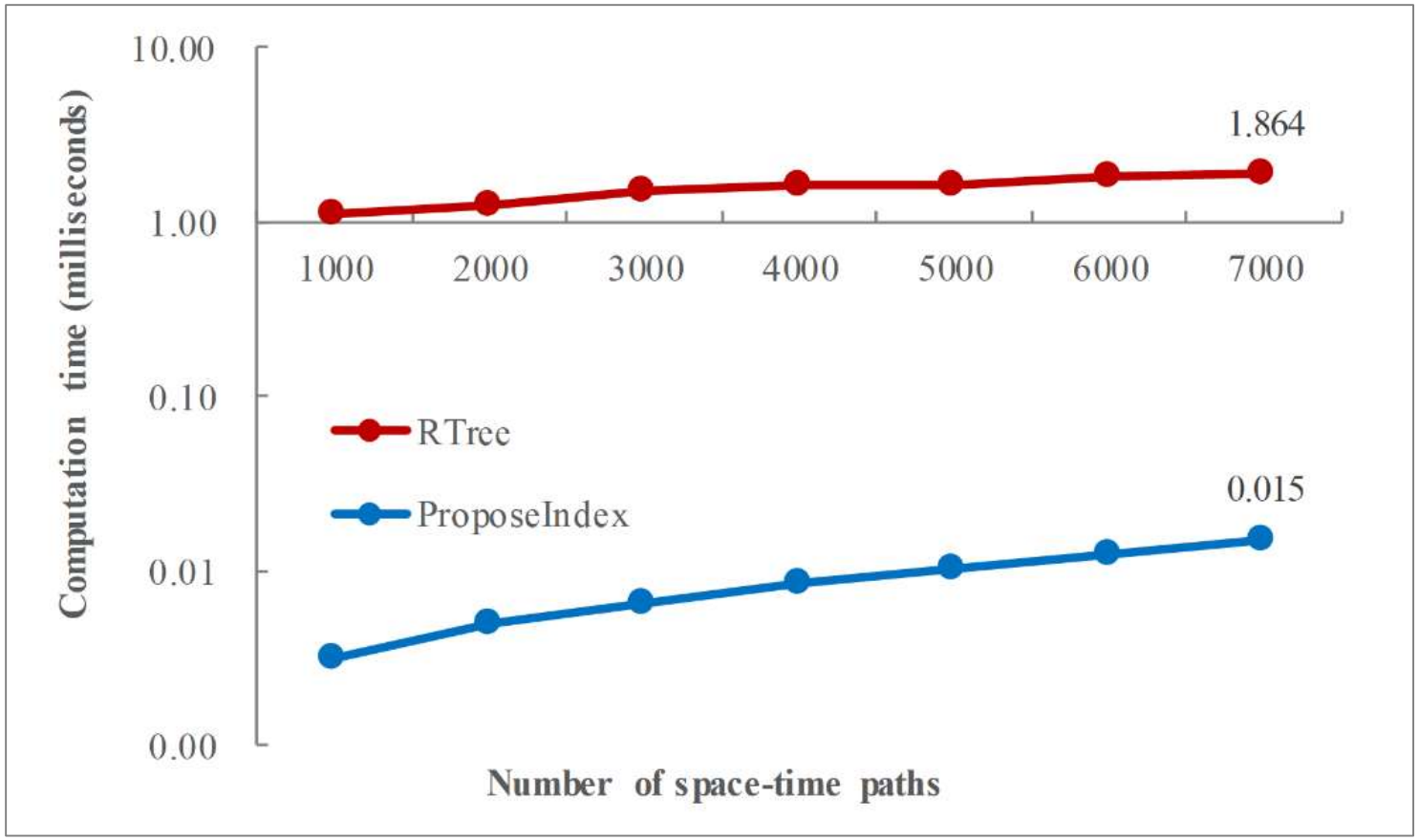
Query in memory significantly improves query performance
(The results are plotted in integer-logarithmic axis.)



Path-path intersection query. Given a space-time path, P_q, of an individual, find all space-time paths intersecting the path P_q.

Part 2 Query performance with proposed index

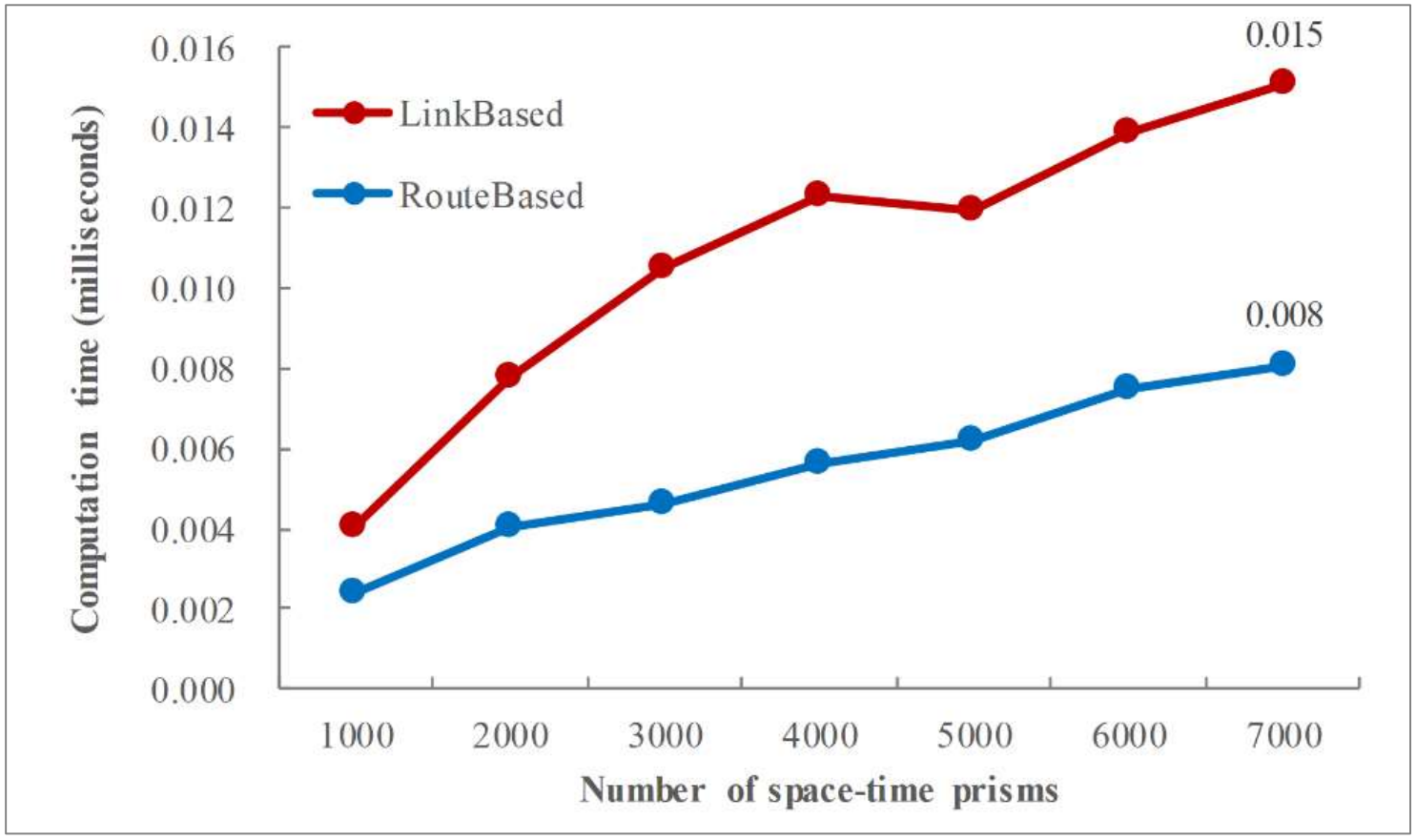
Query with LRS-based spatiotemporal index significantly improves query performance. (The results are plotted in integer-logarithmic axis.)



Prism-path intersection query. Given a space-time prism, STP_q, of an individual, find all space-time paths intersecting the prism STP_q.

Part 2 Query performance with route-based model

Query with route-based CLR model enhance the prism-prism intersection query performance by up to 47% compared to the previous link-based CLR model



Prism-prism intersection query. Given a space-time prism, STP_q, of an individual, find all space-time prisms intersecting the prism STP_q.

Research Framework

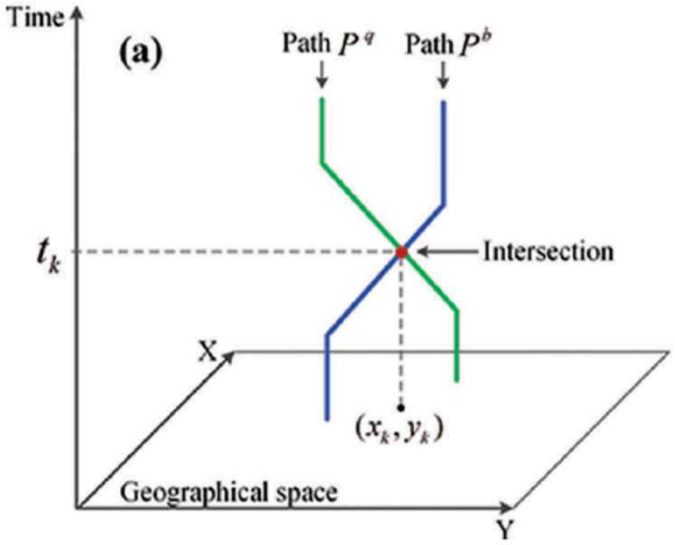
- Part 1. Spatiotemporal data model
- Part 2. Spatiotemporal index structure
- **Part 3. Spatiotemporal proximity analysis**
  - **Spatiotemporal buffering and overlapping**
  - **Network constrained buffering and overlapping**
- Part 4. Case study

Part 3 Spatiotemporal proximity analysis

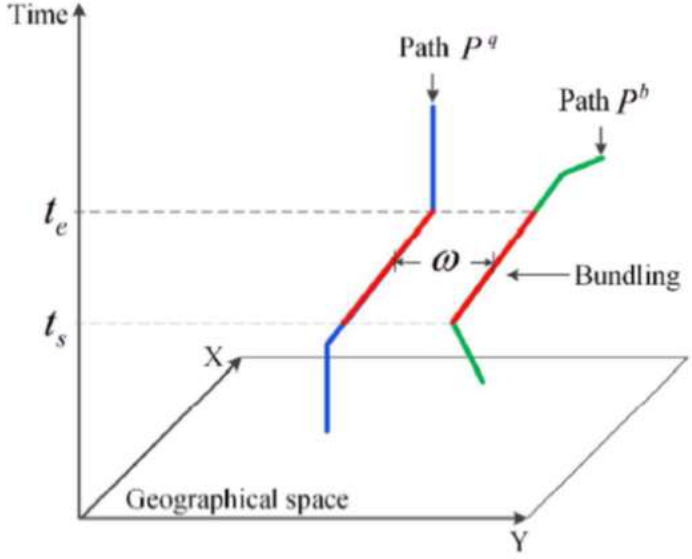
- Spatiotemporal proximity analysis
  - to retrieve all paths which are spatiotemporal proximal to a given path
- A critical step for various space-time path analysis
  - Trajectory clustering
  - Trajectory classification
  - Individual-group patterns
  - Etc.

Part 3 Spatiotemporal proximity analysis

- Different spatiotemporal proximity metrics
  - Time geographic relationships: intersection and bundling
  - Space-time similarity metrics



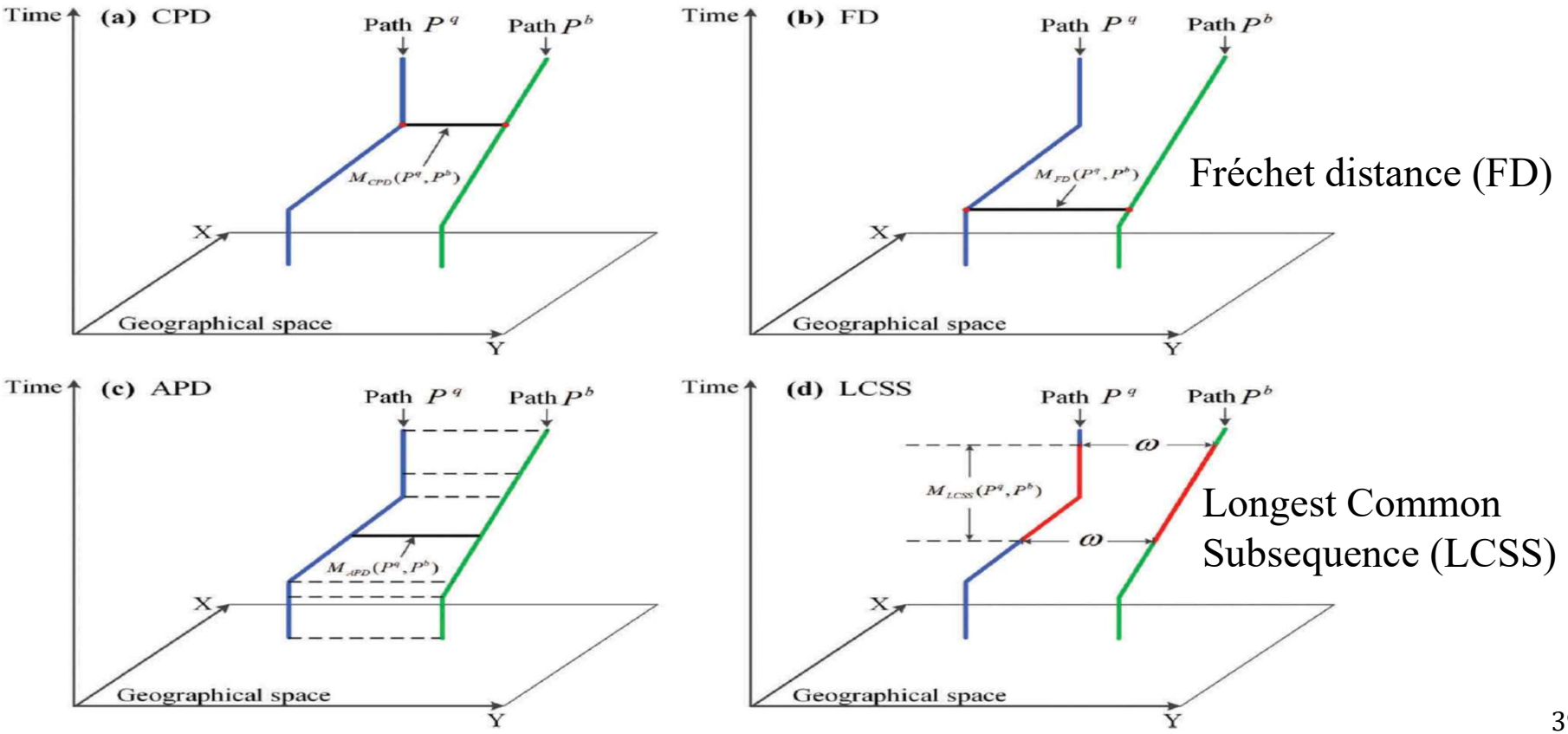
Space-time intersection



Space-time bundling

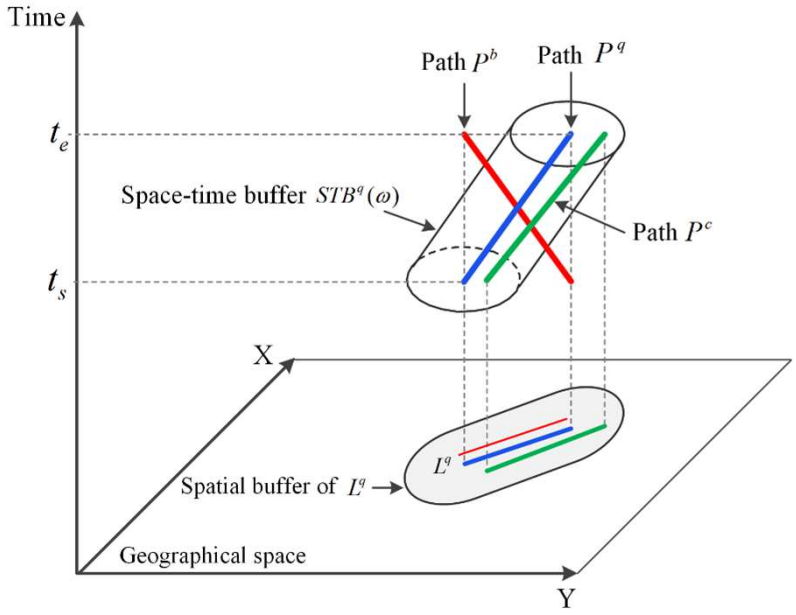
Part 3 Spatiotemporal proximity analysis

- Space-time proximity metrics
  - Time geographic relationships: intersection and bundling
  - Space-time similarity metrics:



Part 3 Spatiotemporal proximity analysis

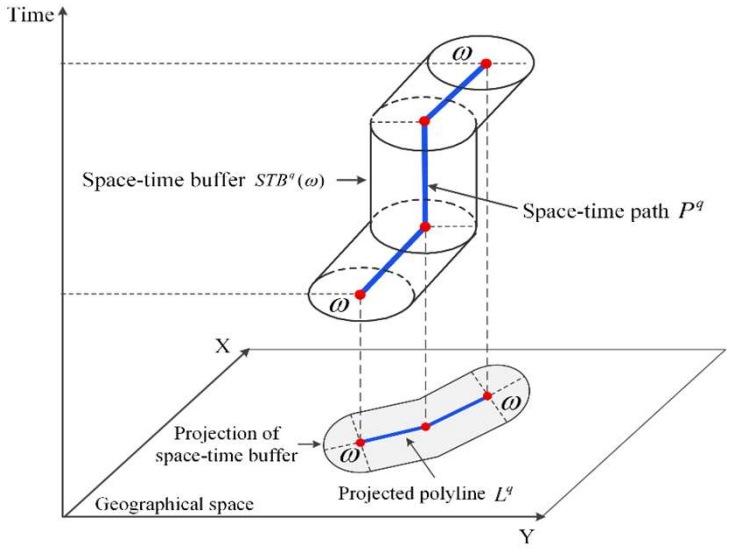
- Existing analysis approaches
  - Brute force approach (i.e., calculate space-time distance metrics from all space-time paths to the given path)
  - Space-time separated approach (i.e., search spatial proximal paths, then check with the time dimension)

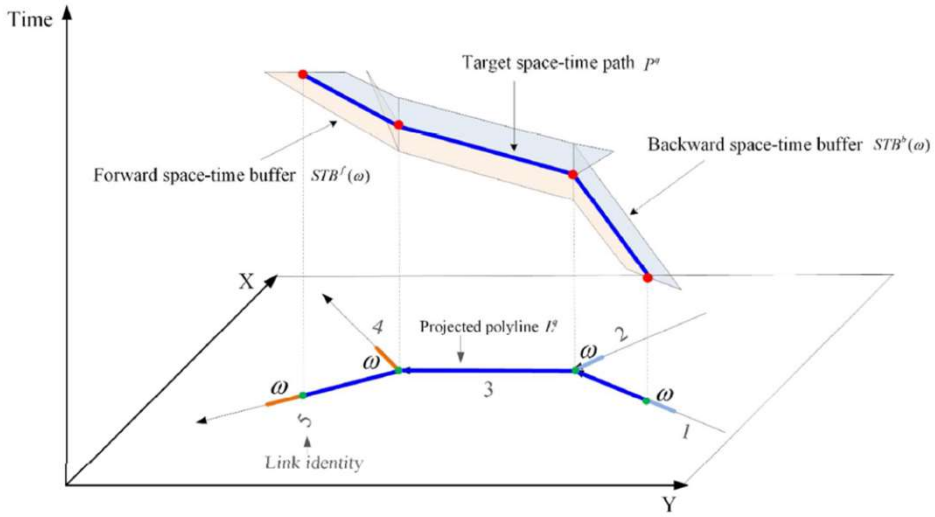


Space-time separated approaches are only necessary rather than sufficient conditions, and computational intensive for big data analytics

Part 3 Proposed space-time integrated approach

Space-time buffering operation: to continuously construct spatial buffers for any space-time location along the trajectory



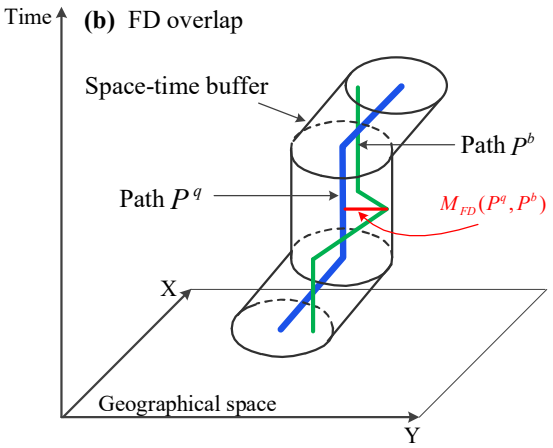


Definition 1. Given a space-time path P^q and a spatial tolerance ω, the space-time buffer $STB^q(\omega)$ delimits all possible space-time points (x_k, y_k, t_k) satisfying

$$STB^q(\omega) = \{(x_k, y_k, t_k) \mid D\left((x_k, y_k), P^q(t_k)\right) \le \omega, t_s^q \le t_k \le t_e^q\}$$

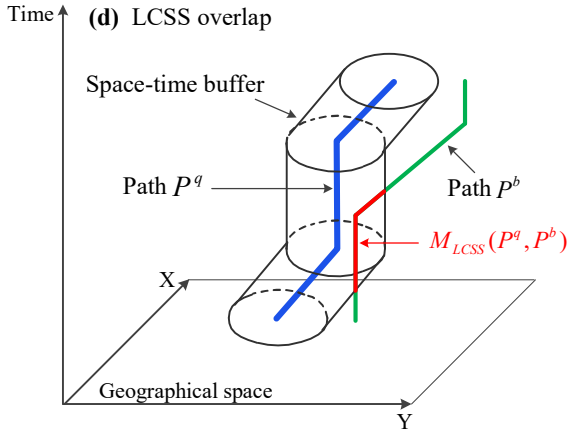
Part 3 Proposed space-time integrated approach

Proposition 3. The FD overlap can retrieve all space-time paths with M_{FD} to $P^q \le \omega$



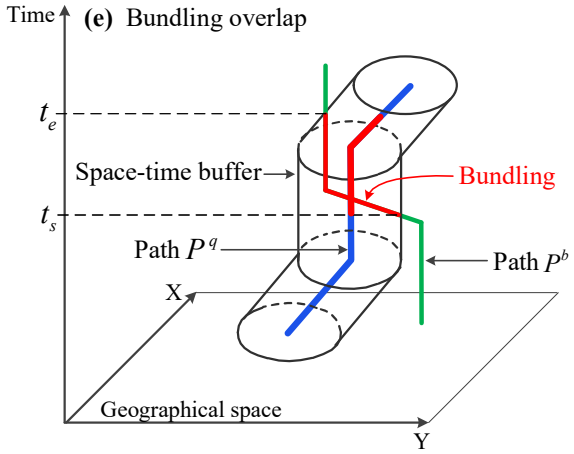
(1)Find all space-time paths contained by the space-time buffer

Proposition 5. The LCSS overlap retrieves all space-time paths with M_{LCSS} to $P^q \ge 0$



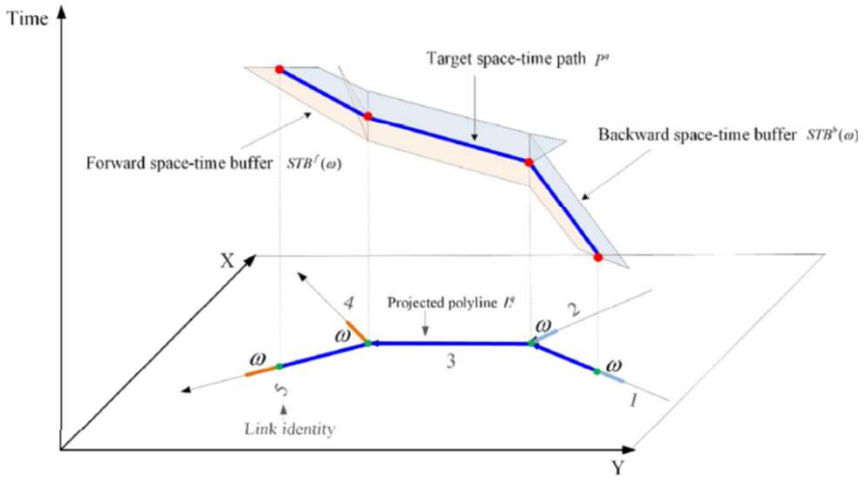
(1) Find all space-time paths intersected with the space-time buffer
(2) Calculate M_{LCSS} for each found path

Proposition 6. The bundling overlap retrieves all space-time paths bundled with P^q

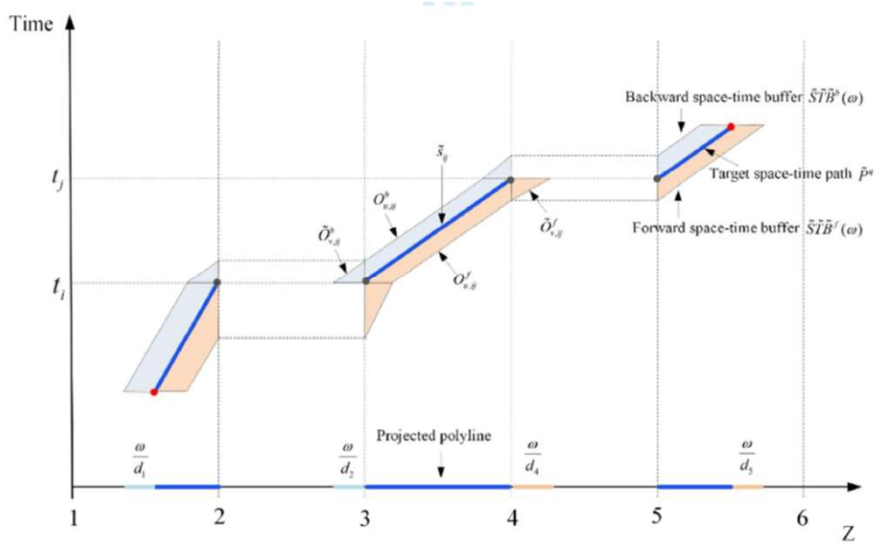


(1) Find all space-time paths intersected with the space-time buffer
(2) Clip each found path to obtain its sub-path within buffer, and check temporal coverage of sub-path

Part 3 Implementations in CLR space



Network space-time in (x, y, t) space



Network space-time in CLR space

Proposition 8. The network space-time buffer $\tilde{STB}^q(\omega)$ in CLR space is an equivalent representation of $STB^q(\omega)$ in (x, y, t) space.

3D
spatiotemporal entities

2D
spatial geometries

Part 3 Computational times

Network space-time buffer construction

Spatial tolerance (m)	Target path length (number of links)	Proposed algorithm (ms)
100	100	3
200	100	4
300	100	5
400	100	6
500	100	7
500	150	12
500	200	21
500	250	27
500	300	36

80km

Space-time overlapping

Spatial tolerance (m)	Target path length (number of links)	CPD overlapping (s)	FD overlapping (s)	APD overlapping (s)	LCSS overlapping (s)	Bundling (s)
100	100	0.752	0.763	0.784	0.808	0.818
200	100	0.967	0.997	1.020	1.069	1.081
300	100	1.338	1.392	1.445	1.534	1.596
400	100	3.913	4.053	4.115	4.213	4.289
500	100	5.771	6.010	6.121	6.268	6.381
500	150	15.836	16.531	18.818	19.189	19.330
500	200	21.292	22.383	24.858	25.245	25.517
500	250	30.628	31.684	34.527	35.343	35.692
500	300	45.140	47.532	56.099	57.856	58.279

Part 3 Compared to different approaches

Table 3. Computational times for three space-time proximity analysis approaches in seconds

Number of space-time paths in database	Space-time integrated approach	Space-time separated approach (Step 1 + Step 2)	Brute force approach
1,000	0.286	9.163 (0.009 + 9.154)	121.863
5,000	0.575	47.831 (0.010 + 47.821)	681.096
10,000	0.752	80.129 (0.011 + 80.118)	1371.524

Table 4. Space-time paths obtained by each step of the separated space-time approach

Number of space-time paths in database	Step 1	Step 2
1,000	180	12
5,000	824	46
10,000	1,417	85

Space-time separated approaches are only necessary rather than sufficient conditions, and computational intensive for space-time path analysis

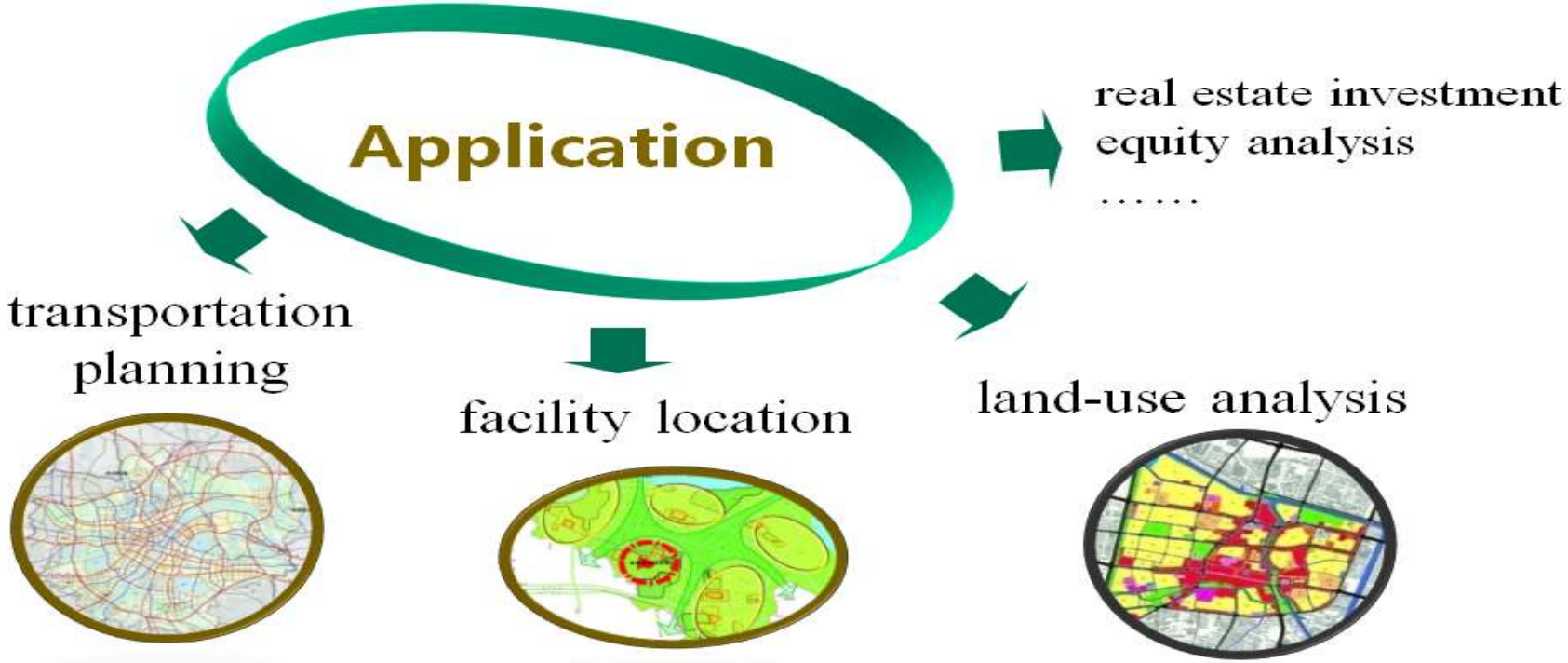
Yuan, H., **Chen, B.Y.***, et al., 2018, Towards space-time buffering for spatiotemporal proximity analysis of movement data. *International Journal of Geographical Information Science*, 32, pp. 1211-1246.

Research Framework

- Part 1. Spatiotemporal data model
- Part 2. Spatiotemporal index structure
- Part 3. Spatiotemporal proximity analysis
- **Part 4. Case study**
  - **Individual accessibility using mobile phone tracking**
  - **Impacts of human mobility on accessibility**

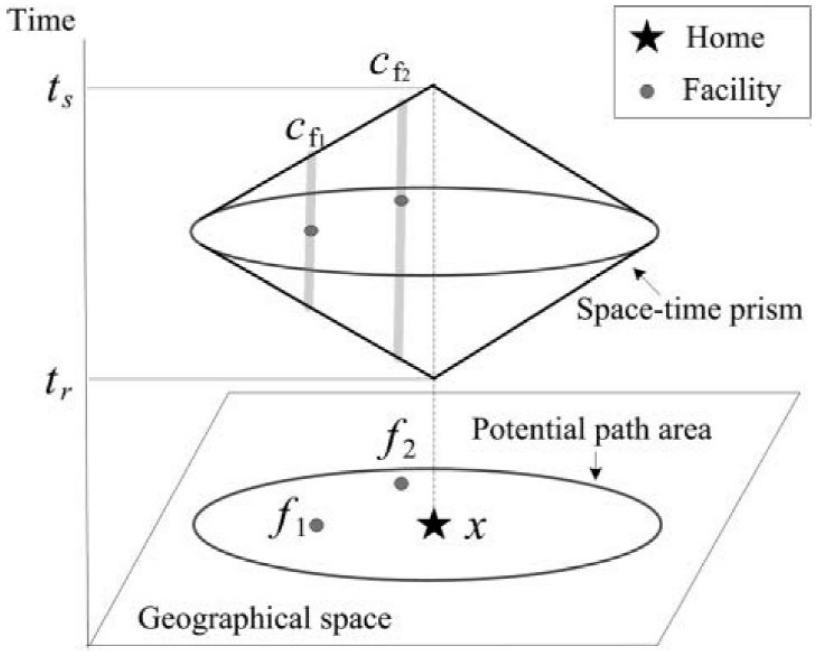
Part 4 Accessibility studies

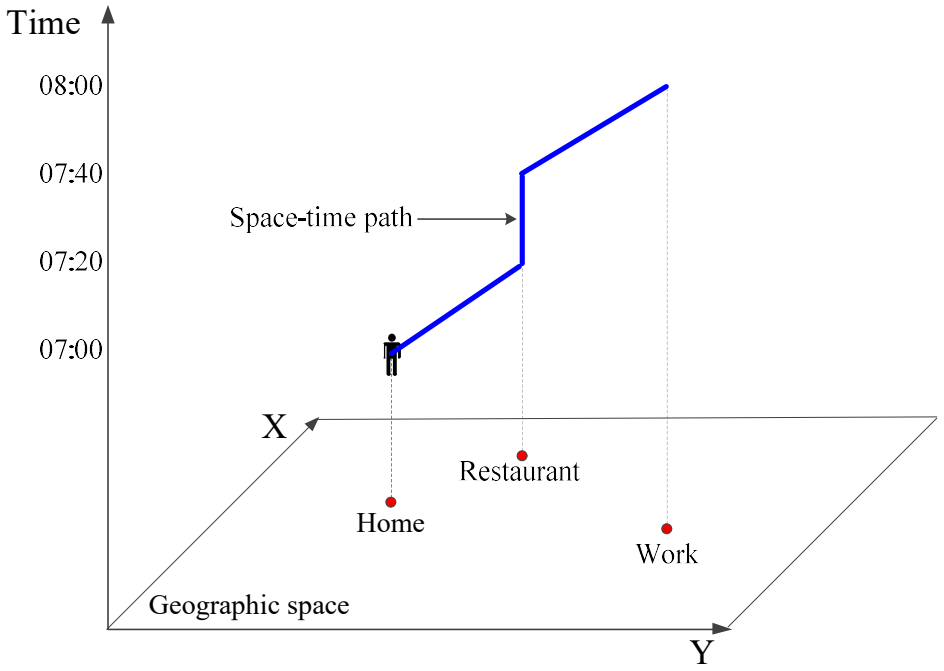
- Accessibility concept
  - the ease with which urban services can be reached from a particular location or by individuals at that location
- High accessibility level relates to high quality of life



Part 4 Accessibility studies

- Conventional place-based accessibility measures
  - Lack of effective methods to collect human mobility data
  - Single anchor point ignoring human mobility



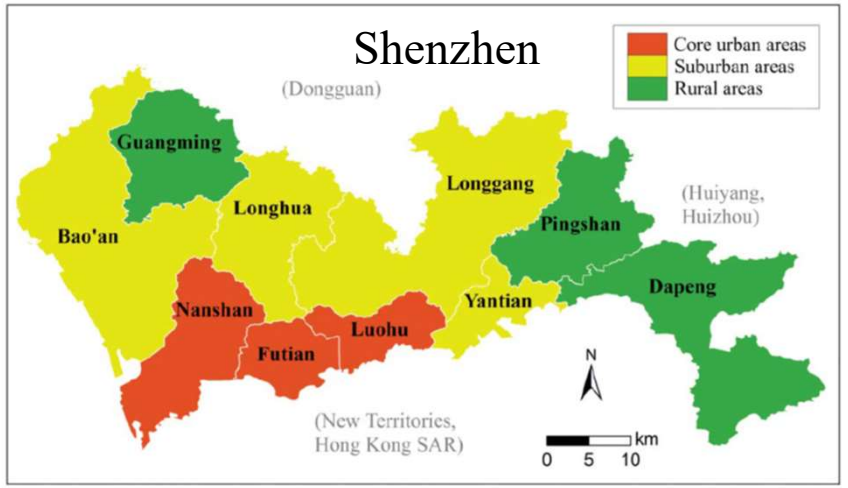


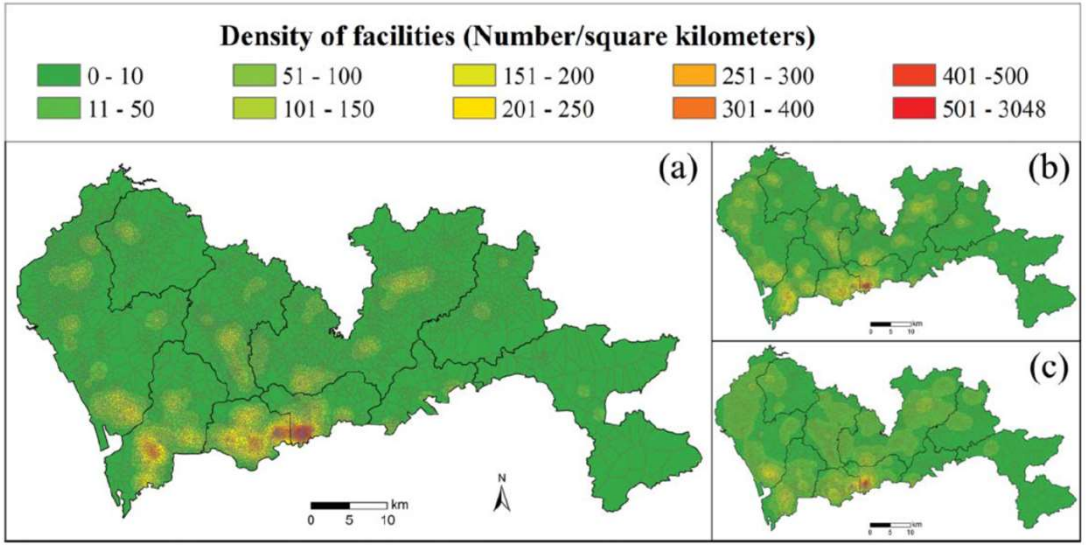
Part 4 Human mobility impacts on accessibility

- Mobile phone data is promising to collect human mobility information
  - Large number of individuals across the entire region
- What are impacts of human mobility on accessibility?
  - Methodological Perspective
    - To quantify the bias of conventional place-based accessibility measures ignoring human mobility
  - Theoretical Perspective
    - To better understand the relationships between transport, land use and human behaviors to facilitate transport and urban planning

Part 4 Study area and data collection

- Collected service facilities
  - 6,899 Food services (a)
  - 8,807 Recreational services (b)
  - 3,705 Shopping facilities (c)

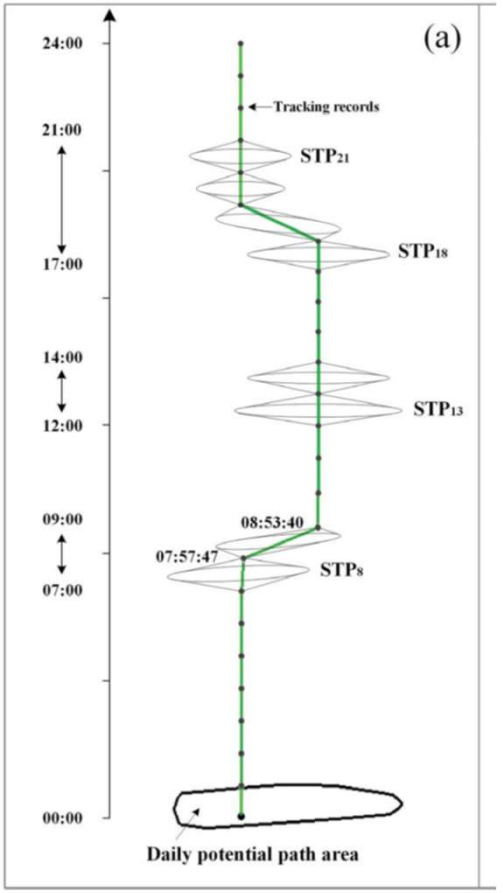




Facility density in Shenzhen

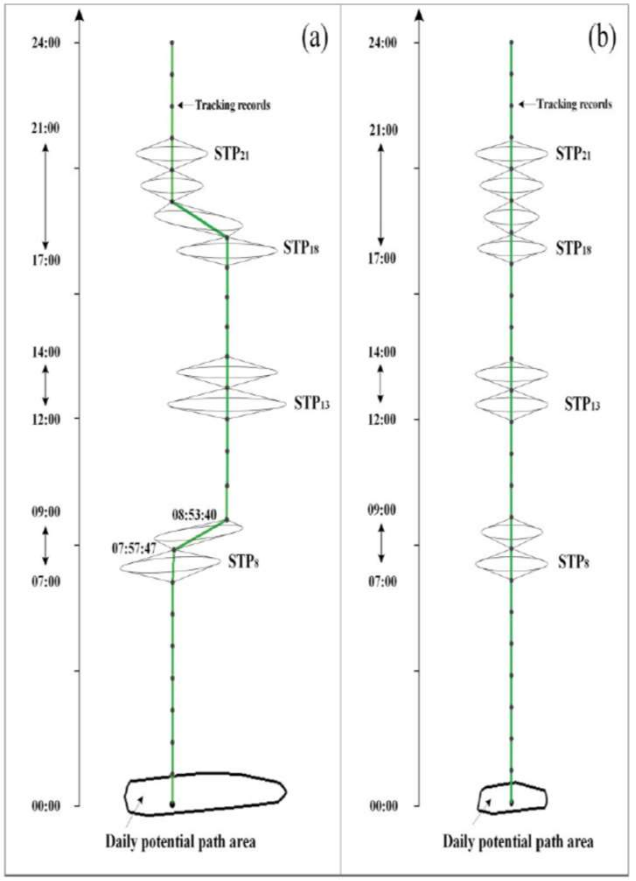
Part 4 Analysis method

- Collected phone users > 6 millions
- 24 hourly location records of each user
- 4 steps
  - 1. Identify home locations (22:00-06:00)
  - 2. Estimate individual accessibility
    - Daily potential path area (2D)
    - Daily space-time prism (3D)
    - Excluding sleeping/in-home activity hours (21:00–07:00) and working hours (09:00–12:00 and 14:00–17:00)
  - 3. Calculate individual accessibility
    - CUM: Cumulative number of facilities within 2D activity space
    - DUR: Cumulative activity durations at facilities within 3D activity space



Part 4 Analysis method

- Collected phone users > 6 millions
- 24 hourly location records of each user
- 4 steps
  - 4. Mobility impact on accessibility
    - **Ratio > 1**
      Positive impact:
      Enhanced accessibility
    - **Ratio = 1**
      No impact
    - **Ratio < 1**
      Negative impact:
      Reduced accessibility



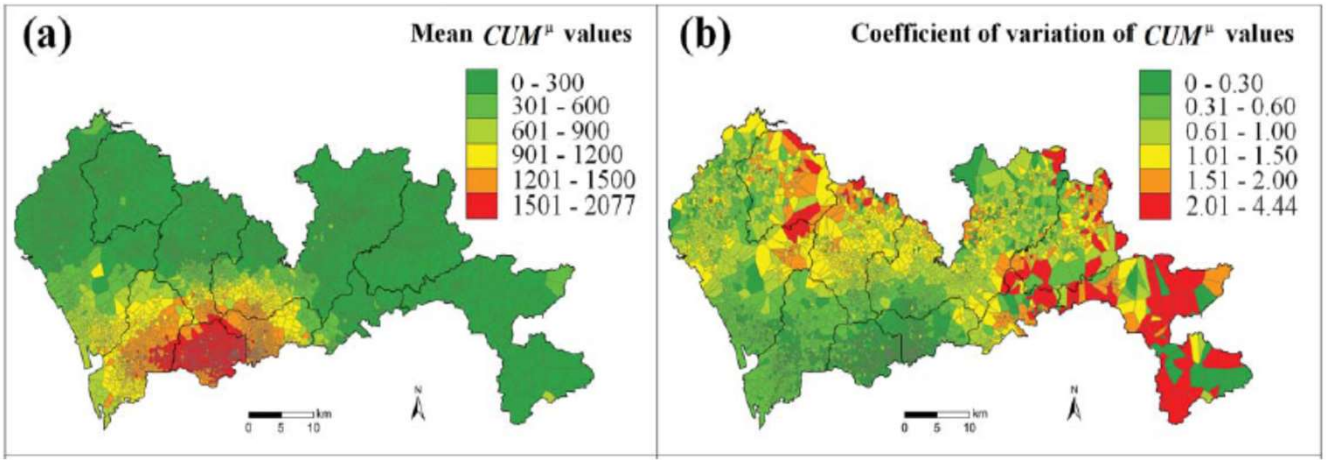
Actual user

Virtual stationary user

Part 4 Results

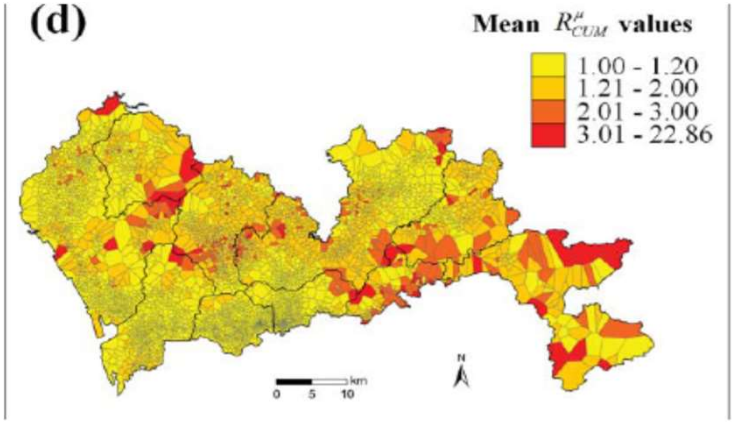
Individual accessibility of CUM without time dimension

- Capture inter-person variation in the same home location, due to distinctive mobility patterns



Mobility impact on accessibility in terms of CUM

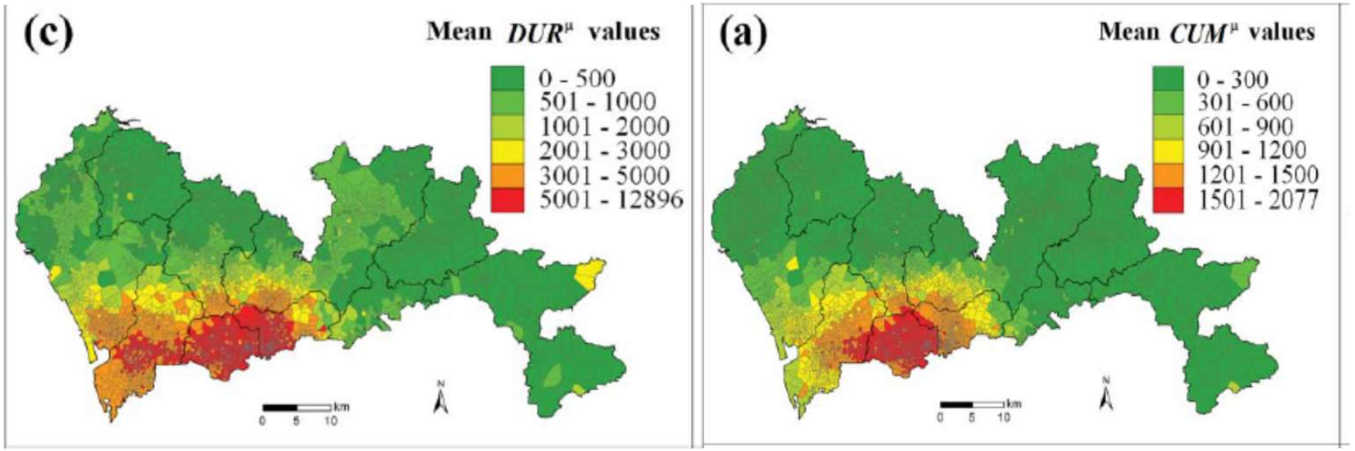
- Human mobility has positive impacts for improving accessibility of people living in all city regions
- Policy implication: Build more roads -> Enhance human mobility -> Improve individual accessibility
- Traffic congestions in core urban areas



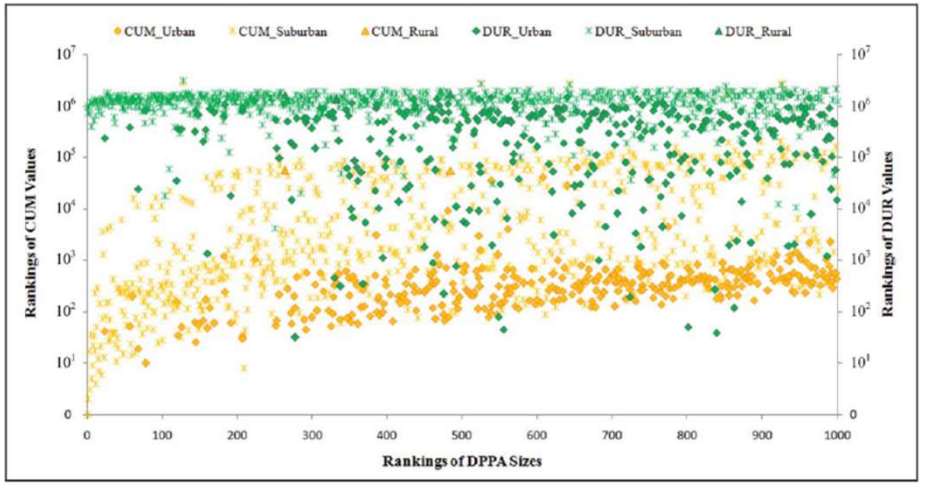
Part 4 Results

Individual accessibility of DUR time dimension

- Spearman's rank correlation coefficient = 0.9



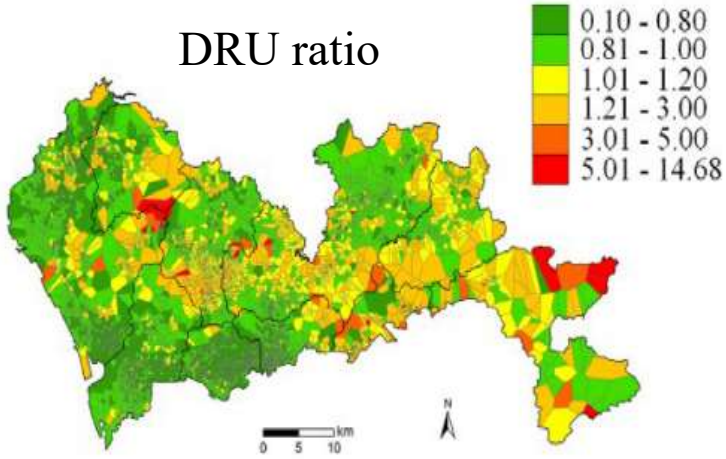
- Huge difference of accessibility ranking of CUM and DUR measures for phone users with a long distance travel (i.e., large 2D activity space)



Part 4 Results and discussions

Mobility impacts on individual accessibility of DUR

- Negative impacts on people in core urban areas with rich services
- Positive impacts on people in sub-urban and rural areas with poor services
- Mitigate spatial inequity of accessibility for people living in different regions



Policy implications

- Rich-service regions: Mix land use -> Reduce mobility-> Improve accessibility
- Poor-service regions: Improve transport infrastructure -> Enhance mobility-> Improve accessibility

Gini coefficients for different service types

Measure service type	DUR^{μ}	DUR^{h}
Foodservice	0.50	0.52
Recreational service	0.46	0.48
Shopping service	0.40	0.42

Methodological implications: incorporate mobility and time dimension

Chen, B.Y.*, et al., 2018, Understanding the impacts of human mobility on accessibility using massive mobile phone tracking data. *Annals of the American Association of Geographers*, 108, pp. 1115-1133.

Conclusions and further studies

- Emerging spatiotemporal big data for time geographic analysis
  - Provide golden opportunities
  - Post significant challenges
- Part 1. Proposed spatiotemporal data model
- Part 2. Spatiotemporal index structure
- Part 3. Proposed spatiotemporal proximity analysis
- Part 4. Individual accessibility using mobile phone big data
- Further studies
  - Advanced models, spatiotemporal operations, and data mining algorithms are required to support computational time geography in the ear of big data

Thanks for your attentions !

Wuhan University **Prof. Bi Yu Chen**; Email: chen.biyu@whu.edu.cn