Efficient Route Planning on Public Transportation Networks: A Labelling Approach

Sibo Wang†, Wenqing Lin†, Yi Yang§, Xiaokui Xiao†, and Shuigeng Zhou§

1. Problem Definition

- **Public Transportation Network**
  - Each edge (u, v, t₁, t₂, w₁, w₂, b) denotes the bus b departs from station u at timestamp t₁ and arrives at station v at timestamp t₂.
- **Route Planning Queries**
  - Earliest Arrival Path (EAP)
    - Depart from u, arrive at v, how to arrive at v.
  - Latest departure Path (LDP)
    - Depart at t₁, arrive at v, how to depart from u before timestamp t₁.
  - Shortest Duration Path (SDP)
    - Depart from u, arrive at v, shortest travel duration.

2. Timetable Labelling

- **Node order o**
  - A total ordering: for any u and v, we have o(u) > o(v) or o(u) < o(v).
  - u has a higher rank than v if o(u) < o(v).
- **Each node v has an in-label set L_in(v)** (resp. out-label set L_out(v)). Each label is a tuple \( l = (s, t₁, t₂, w₁, w₂, b) \), where
  - s: the starting node (resp. ending node) of a path P from x to v (resp. v to x).
  - P: among all nodes on P, s has the highest rank.
  - \( t₁, t₂, w₁, w₂, b \): path P departs at time t₁ and arrives at time t₂, by taking bus b. If bus transfer is required, b is null.
  - No dominated path P exists such that P has the same starting node and ending node as P with departure time \( t₁ \) and arrival time \( t₂ \), but \( [s, t₁, t₂, w₁, w₂, b] \subseteq [s, t₁, t₂, w₁, w₂, b] \).
- **Index Construction**
  - Forward EAP traversal.
  - From the largest timestamp \( t \) non-dominance guarantee → \( P_t \).
  - Next largest timestamp \( t' \) not dominated by \( P_t \) non-dominance guarantee.
- **Node Ordering**
  - An \( \text{vfm} \)-Approximation Algorithm.
  - Adaptive coverage \( R(v) \) of non-dominated paths not covered by ordered vertices \( v₁, v₂, v₃, \ldots \).
  - Node with maximum \( R(v) \), we have \( o(v) = i, \ldots, v \).
  - Time complexity \( O(n² - m) \), where \( n \) and \( m \) are numbers of nodes and edges respectively.
- **A Heuristic algorithm**
  - Estimate adaptive coverage with sampling.

3. Query Processing

- **Phase 1: Candidate Generation.**
  - Consider the paths from u to v.
  - Given \( l₁ = (w₁, t₁, t₂, v₂) \in L_out(u) \) and \( l₂ = (w₂, t₂, t₃, v₃) \in L_in(v) \) where \( t₂ ≤ t₃ \),
  - Concatenating path pertinent to \( l₁ \) and \( l₂ \), add \((l₁, l₂)\) to a candidate set \( \Sigma \).
- **Pruning Techniques**
  - Exp-1: non-dominance guarantee.
    - Each label is dominated by exactly one label with the same starting node and the same ending node.
    - Given \((v₁, v₂, v₃)\) and \((u₁, u₂, u₃)\),
    - Earliest arrival path: \((v₁, v₂, v₃)\) and \((u₁, u₂, u₃)\) are non-dominated.
    - Latest departure path: \((v₁, v₂, v₃)\) and \((u₁, u₂, u₃)\) are non-dominated.

4. Preprocessing

- **Route-based Compression**
  - Labels with vehicle info
    - Route compression: \( L_{out}(v₄) \rightarrow L_{out}(v₃) \)
  - Time complexity:
    - Exp-3: Preprocessing Time.
      - Concise Representation of Query Results
        - Set \( P = \{ (v₁, v₂, 7, 9, b₁), (v₁, v₂, 9, 10, b₂) \} \)
        - \( P \) can be denoted by a concise path \( P_c \), as follows:
          - \( P_c = (v₁, 7, b₁, v₂, 11, b₂) \)
          - Reduce path unfolding cost.
          - Improving query performance.
    - Exp-4: Compression Ratios.
      - \( A₁ \): Route-based compression.
      - \( A₂ \): Pivot-based compression.
      - \( A₃ \): Route-based and Pivot-based compression.

5. Optimizations and Extensions

- **Label Compression**
  - Route-based compression
    - Labels with vehicle info
      - Compress to \((v₉, 5, 9, b₂, v₂)\) from \((v₉, 10, 13, b₄, v₂)\)
  - Pivot-based compression
    - \( o(u) < o(p) < o(v) \)
      - Consider \( l₁ = (s, t₁, t₂, w₁, w₂, p) \) and \( l₂ = (s, t₂, t₃, p, w₃) \) where \( t₂ ≤ t₃ \),
      - If \( l₁ = (s, t₁, t₂, w₁, w₂, p) \), then \( l₁ \leq l₂ \).
    - Compress to \( L₉₄₃ = \{ u, null, null, null, p \} \)
      - Label dependency can be used to build a weighted dependency graph.
      - NP-hard: maximum-weight independent set

6. Datasets

- **Datasets**
  - City
    - Austin: 27.3K
    - Boston: 44.6K
    - Chicago: 151.8K
    - Los Angeles: 121K
    - New York: 264.2K
    - Miami: 685.0K
    - San Francisco: 36K
    - San Diego: 71.4K

- **Competitors**
  - CSA [Dibbelt et al. SEA’13, Wu et al. VLDB’14]
  - CHT [Geisberger et al. SEA’10]