Embedded Algorithms for Transit Performance Measurement

Timothy Becker
General Transit Feed Specification:

- defines a common format for public transportation schedules

- intended for source-destination path finding

GTFS: https://maps.google.com/transit
GTFS: https://maps.google.com/transit
Overview: System Analysis

GTFS can be used for system analysis:

- unique stop id (lat, lon)
- unique trip id (sequence of stop ids)
- unique route id (set of trip ids)

Stop ids contained inside FIPS spaces allow association from GTFS data to ACS5YR and LEHD data.
Database Import Workflow

ACSIYR annual

LEHD annual

GTFS asynchronous

ACSIYR2 MSSQL mssql_conn acsfiveyr_files

LEHD2 MSSQL mssql_conn lehd_files

GTFS2 MSSQL mssql_conn gfts_files

DB

FIPS

ACS

LEHD

GTFS

EDIT

EDIT
transformed and processed GTFS representation that leverages system level spatial-temporal context
Analysis Tools: Segmentation?

- summarize important parts of a route
- standardize route-based analysis
- facilitate what if analysis
Hierarchical segmentation by:

- vertex sequence similarity (Edit Distance)
- vertex nearest neighbor set similarity  
  (Jaccard Index)
Vertex Sequence Similarity

Edit Distance - minimum number of edit operations that make one sequence into another
Vertex NN Similarity

Jaccard Index – set intersection magnitude over set union magnitude

\[ \frac{|X \cap Y|}{|X \cup Y|} = \frac{3}{4} = 0.75 \]

\( X = \{b, c, f\} \)
\( Y = \{b, c, f, e\} \)

\[ |X \cap Y| = 3 \]
\[ |X \cup Y| = 4 \]

\[ \frac{3}{4} = 0.75 \]
Curtis Denton, MS
Geography PhD Student
University of Connecticut