CAMBRIDGE SYSTEMATICS

Think Forward

It's Only a Matter of Time – Using GTFS in the NY Best Practice Model

presented to

2016 Transport-Tech Summit

presented by

Cambridge Systematics, Inc.

Nikhil Puri

November 15, 2016

Outline

- Background and Motivation
- GTFS Overview
- New York Best Practice Model (NYBPM) Network Update Process
- Conflation of GTFS Feeds with the NYBPM Network
  - » Overview
  - » Methodology and preliminary results
- Next Steps

Background and Motivation

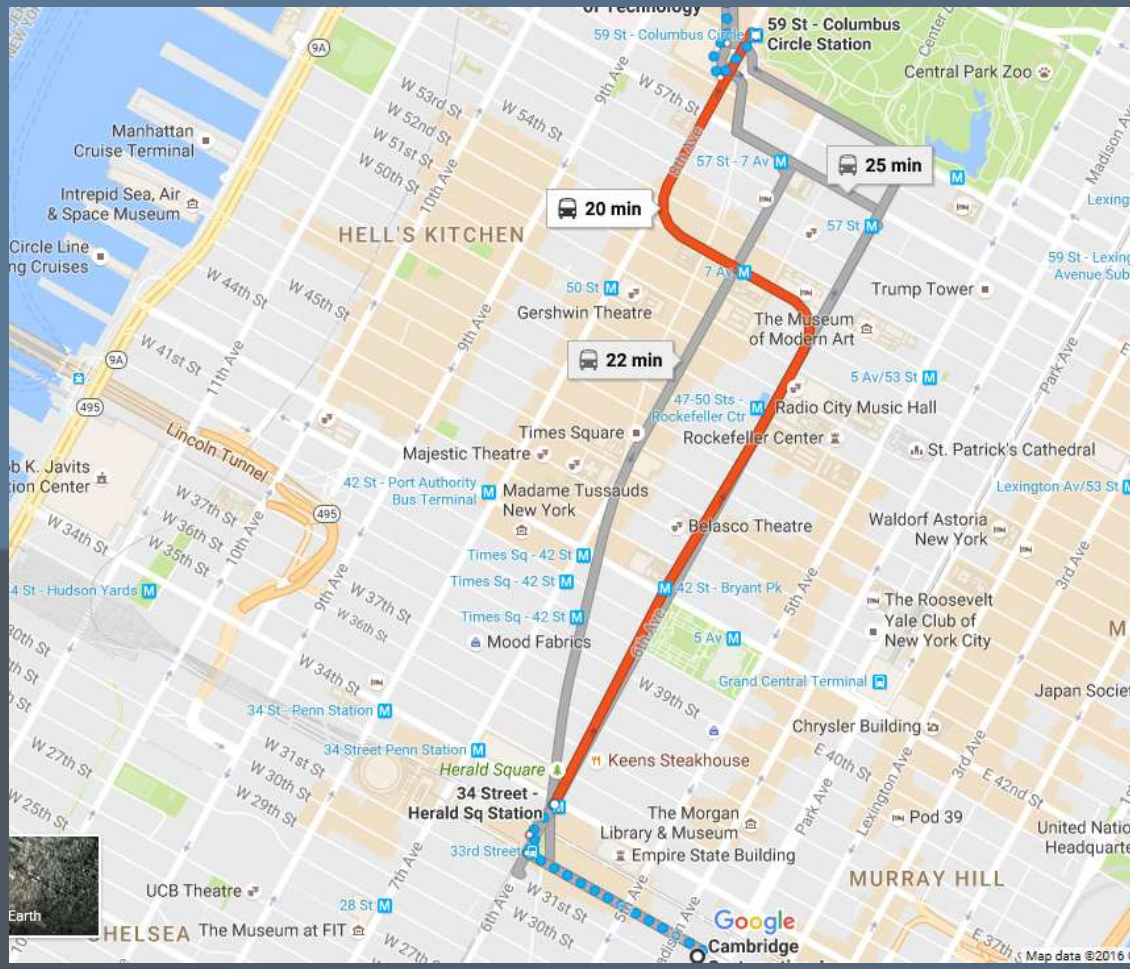
- Improve transit ridership modeling by improving the quality of supply side data in travel demand models
  - » Travel times and transfers
  - » Less manual coding of transit
  - » Highway and transit layer integration
- GTFS conflation proposed as part of the NYBPM network update
- Limited applications at this scale, if any

CAMBRIDGE SYSTEMATICS

GTFS Overview

- GTFS = the General Transit Feed Specification
- An electronic version of paper maps and route schedules
- Has become the de-facto standard among public agencies
- It is a collection of text tables
- GTFS defines a common format for public transportation schedules and associated geographic information. GTFS "feeds" allow public transit agencies to publish their transit data and developers to write applications that consume that data in an interoperable way.
- Common data source for route planning websites and apps

GTFS Example
Real-Time Route Planning



6:40 AM - 7:00 AM (20 min)

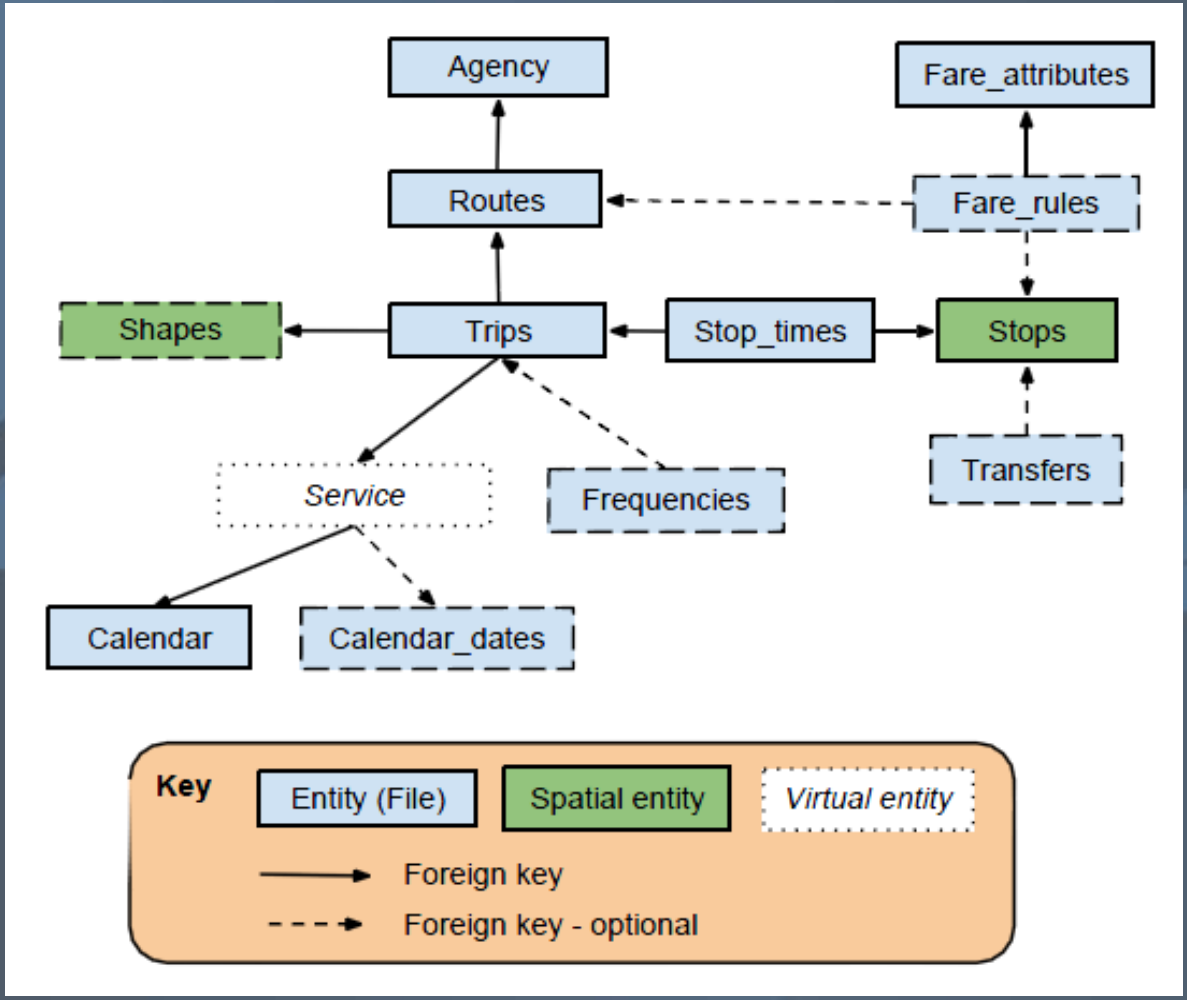
6:50 AM from 34 Street - Herald Sq Station
14 min

SCHEDULE EXPLORER

6:40 AM	**Cambridge Systematics, Inc.** 38 East 32nd Street, New York, NY 10016
	Walk About 10 min , 0.4 mi
6:50 AM	**34 Street - Herald Sq Station**
	Norwood - 205 St 6 min (4 stops) Information
6:56 AM	**59 St - Columbus Circle Station**
	Walk About 4 min , 0.3 mi
7:00 AM	**New York Institute of Technology**

CAMBRIDGE SYSTEMATICS

GTFS Database Structure



Source: Created by Martin Davis, as per blog post Lin.ear th.inking.

GTFS Not Extensively Used in Transportation Planning

- The challenges are only technical and not institutional
- Biggest challenge is network conflation
- Errors and inconsistencies will be encountered while integrating these data sources with other transportation sources
- Limited guidance on Importing GTFS data into planning networks
  - » Route alignments
  - » Stop locations
  - » Headways and frequencies
  - » Transit fares

The NYBPM Context

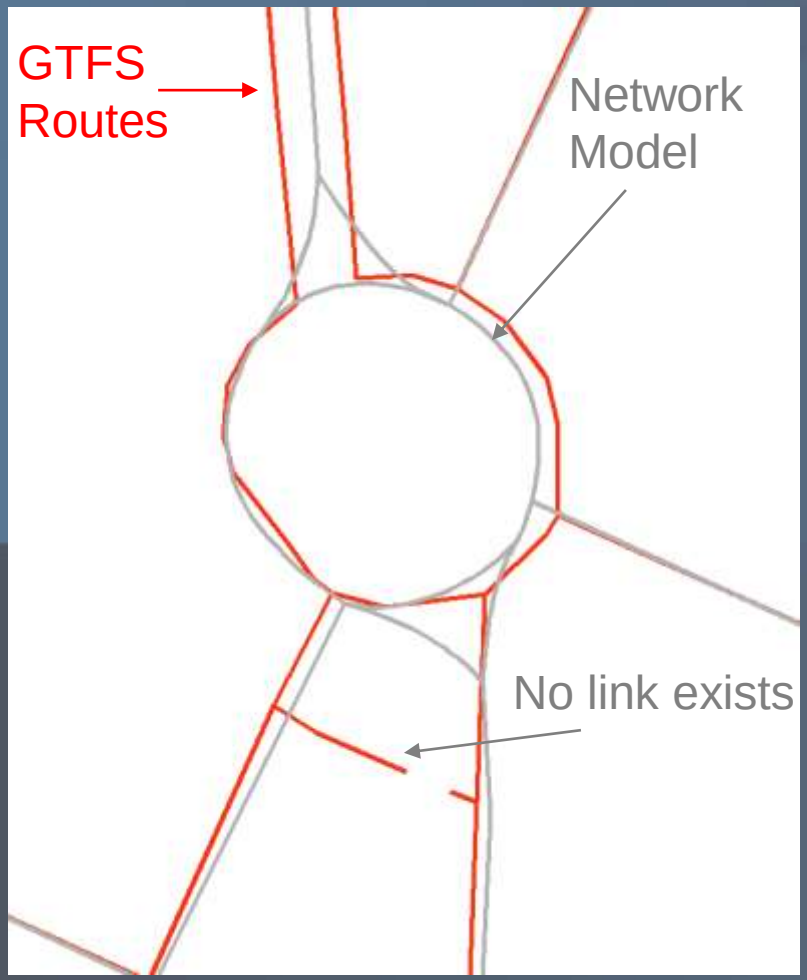
- Complex Transit Network
  - » Path could include a combination of modes
  - » Several transfer opportunities
- Transit Elements
  - » In-vehicle times
  - » Out-of-vehicle times – access, xfer, xfer wait, egress
  - » Fares

CAMBRIDGE SYSTEMATICS

NYBPM Network Update Process

- Identify usable GTFS feeds
- Conflate GTFS feeds with NYBPM highway network using a sophisticated algorithm developed by CS
  - » Off-the-shelf applications in infancy
- Highway network detail where necessary
- Iterative process to improve conflation quality
- Process GTFS data for “skimming”
- Significant amount of QA/QC needed

Network Modeling
Columbus Circle

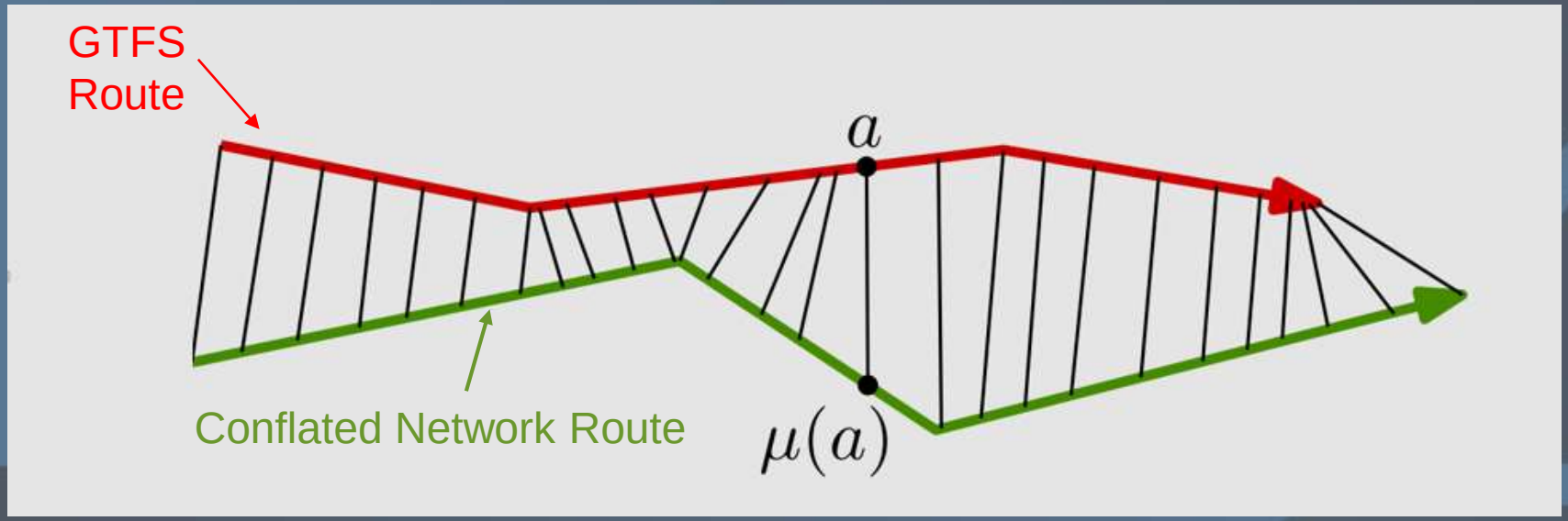


- A model is an imperfect representation of reality
  - » 2D space => nodes, links
- GTFS routes
  - » Alignment errors
  - » Coding errors
- Network model
  - » Approximations
  - » Node and link errors
  - » Network detail is missing

Network Conflation

- Is the task of associating the elements of two networks so that link and node attributes can be transferred
- Is a serious problem in data integration and big data utilization
  - » Mobility data from cell phones or other censors are collected on one network and need to be merged with other data on other networks
- Algorithms for network conflation tested
  - » Snap to closest node – encountered issues
  - » Heuristic rules (if X and Y then do Z) grow out of control
  - » Optimizing a single measure such as the area between the two network links used

Guiding Principle
Minimize Area



CAMBRIDGE SYSTEMATICS

Example
GTFS Route Conflation (M7)

CAMBRIDGE SYSTEMATICS

Conflation Outcome

Missing links in the network model force conflation to parallel paths

Preliminary Conflation Results

- Best results when network links exist
  - » Even if GTFS and the network links are not perfectly aligned
- Algorithm produces reasonable results when links are missing
  - » Finds parallel paths, where exist
  - » However, processing times increase

Preliminary Conflation Results (continued)

- Algorithm implemented in Python
  - » Libraries used – Fiona, shapely, networkX
  - » Python is good for prototyping; good for one-time application (run speed)
  - » Good for multithreading
- Successful conflation achieved
- Route development tested successfully

Next Steps

- Model skim development
- Rigorous testing and validation
- Schedule for Task Completion - May 2017

Contact Information

Nikhil Puri

Phone – 646.364.5491
Email – npuri@camsys.com

CAMBRIDGE SYSTEMATICS