GTFS Bus Stop Mapping to the OSM Network
Introduction
Why bus stop mapping?
Introduction

- Public transport
- Microscopic simulations
- Data requirement
- Difficulties
- New fully automated technique
Data preparation

What do we need?
Data preparation

- GTFS & OSM
Data preparation

- GTFS
  - Optional shapefile

Israel, Kiryat Gat
Data preparation

- GTFS & OSM
  - Convert OSM into a directed network graph
  - Find for each bus stop specified in GTFS a set of candidate network links to attach it
Algorithm

How does it work?
Algorithm 4.1 Determination of optimal assignment of projected stops $S_p$ to GTFS stops $S_G$.

1: $S_p \leftarrow \text{projStops}(S_G)$
2: $\text{fixTrivial}(S_G)$
3: $\langle G_G, G_P \rangle \leftarrow \text{graphFromBusStopSequences}()$
4: repeat
5: \hspace{1em} $\text{removedAtLeastOneCandidate} \leftarrow \text{false}$
6: \hspace{1em} $\text{handleTriples}(\langle G_G, G_P \rangle)$
7: \hspace{1em} $\text{handleNonBifurcatingMaximalSequences}(\langle G_G, G_P \rangle)$
8: \hspace{1em} $\text{handleStars}(\langle G_G, G_P \rangle)$
9: \hspace{1em} $\text{reduceCycleBreakers}(\langle G_G, G_P \rangle)$
10: until $\neg \text{removedAtLeastOneCandidate}$
11: $\text{components} \leftarrow \text{decompose}(\langle G_G, G_P \rangle)$
12: for all $c \in \text{components}$ do
13: \hspace{1em} $\text{assign}(c)$
14: end for

- Assign GTFS stops having a single projection
- Introduces cycleBreakers
- Vertex in the middle has $\text{inDegree} = \text{outDegree} = 1$
- Internal vertices have $\text{inDegree} = \text{outDegree} = 1$
- Removes cycleBreakers that became redundant by fixing some vertices
- Either by explicit discarding or as a consequence of fixing
- Assignment by solution enumeration
Trips
Algorithm 4.1 Determination of optimal assignment of projected stops $S_p$ to GTFS stops $S_G$.

1: $S_p \leftarrow \text{projStops}(S_G)$
2: $\text{fixTrivial}(S_G)$ ▷ Assign GTFS stops having a single projection
3: $\langle G_G, G_P \rangle \leftarrow \text{graphFromBusStopSequences}()$ ▷ Introduces $\text{cycleBreakers}$
4: repeat
5: $\text{removedAtLeastOneCandidate} \leftarrow \text{false}$ ▷ Vertex in the middle has $\text{inDegree} = \text{outDegree} = 1$
6: $\text{handleTriples}(\langle G_G, G_P \rangle)$ ▷ Internal vertices have $\text{inDegree} = \text{outDegree} = 1$
7: $\text{handleNonBifurcatingMaximalSequences}(\langle G_G, G_P \rangle)$ ▷ Removes $\text{cycleBreakers}$ that became redundant by fixing some vertices
8: $\text{handleStars}(\langle G_G, G_P \rangle)$ ▷ Either by explicit discarding or as a consequence of fixing
9: $\text{reduceCycleBreakers}(\langle G_G, G_P \rangle)$
10: until $\neg \text{removedAtLeastOneCandidate}$
11: $\text{components} \leftarrow \text{decompose}(\langle G_G, G_P \rangle)$ ▷ Assignment by solution enumeration
12: for all $c \in \text{components}$ do
13: $\text{assign}(c)$
14: end for
GTFS graph
Algorithm 4.1 Determination of optimal assignment of projected stops $S_P$ to GTFS stops $S_G$.

1: $S_P \leftarrow \text{projStops}(S_G)$
2: $\text{fixTrivial}(S_G)$
3: $(G_G, G_P) \leftarrow \text{graphFromBusStopSequences}()$
4: repeat
5:  \hspace{1em} removedAtLeastOneCandidate \leftarrow \text{false}
6:  \hspace{1em} \text{handleTriples}((G_G, G_P))
7:  \hspace{1em} \text{handleNonBifurcatingMaximalSequences}((G_G, G_P))
8:  \hspace{1em} \text{handleStars}((G_G, G_P))
9:  \hspace{1em} \text{reduceCycleBreakers}((G_G, G_P))
10: until \neg \text{removedAtLeastOneCandidate}
11: components \leftarrow \text{decompose}((G_G, G_P))
12: \text{for all } c \in \text{components} do
13:  \hspace{1em} assign(c)
14: \text{end for}
- Example of a cycle breakers
Algorithm 4.1 Determination of optimal assignment of projected stops $S_p$ to GTFS stops $S_G$.

1: $S_p \leftarrow \text{projStops}(S_G)$
2: $\text{fixTrivial}(S_G)$
3: $(G_G, G_P) \leftarrow \text{graphFromBusStopSequences}()$ \hfill ▶ Assign GTFS stops having a single projection \hfill ▶ Introduces cycleBreakers
4: repeat
5: removedAtLeastOneCandidate $\leftarrow$ false
6: $\text{handleTriples}((G_G, G_P))$ \hfill ▶ Vertex in the middle has inDegree = outDegree = 1
7: $\text{handleNonBifurcatingMaximalSequences}((G_G, G_P))$ \hfill ▶ Internal vertices have inDegree = outDegree = 1
8: $\text{handleStars}((G_G, G_P))$
9: $\text{reduceCycleBreakers}((G_G, G_P))$ \hfill ▶ Removes cycleBreakers that became redundant by fixing some vertices \hfill ▶ Either by explicit discarding or as a consequence of fixing
10: until $\neg$removedAtLeastOneCandidate
11: components $\leftarrow \text{decompose}((G_G, G_P))$
12: for all $c \in$ components do
13: assign($c$) \hfill ▶ Assignment by solution enumeration
14: end for
- Example of a triple
Algorithm 4.1 Determination of optimal assignment of projected stops $S_P$ to GTFS stops $S_G$.

1: $S_P \leftarrow projStops(S_G)$  
2: $fixTrivial(S_G)$  
3: $\langle G_G, G_P \rangle \leftarrow graphFromBusStopSequences()$  
4: repeat  
5: $removedAtLeastOneCandidate \leftarrow false$  
6: $handleTriples(\langle G_G, G_P \rangle)$  
7: $handleNonBifurcatingMaximalSequences(\langle G_G, G_P \rangle)$  
8: $handleStars(\langle G_G, G_P \rangle)$  
9: $reduceCycleBreakers(\langle G_G, G_P \rangle)$  
10: until $\neg removedAtLeastOneCandidate$  
11: $components \leftarrow decompose(\langle G_G, G_P \rangle)$  
12: for all $c \in components$ do  
13: $assign(c)$  
14: end for  

- Assign GTFS stops having a single projection  
- Introduces cycleBreakers  
- Vertex in the middle has $inDegree = outDegree = 1$  
- Internal vertices have $inDegree = outDegree = 1$  
- Removes cycleBreakers that became redundant by fixing some vertices  
- Either by explicit discarding or as a consequence of fixing  
- Assignment by solution enumeration
Example of a maximal non bifurcating sequence
Algorithm 4.1 Determination of optimal assignment of projected stops $S_P$ to GTFS stops $S_G$.

1: $S_P \leftarrow \text{projStops}(S_G)$
2: $\text{fixTrivial}(S_G)$
3: $\langle G_G, G_P \rangle \leftarrow \text{graphFromBusStopSequences}()$
4: repeat
5: $\text{removedAtLeastOneCandidate} \leftarrow \text{false}$
6: $\text{handleTriples}(\langle G_G, G_P \rangle)$
7: $\text{handleNonBifurcatingMaximalSequences}(\langle G_G, G_P \rangle)$
8: $\text{handleStars}(\langle G_G, G_P \rangle)$
9: $\text{reduceCycleBreakers}(\langle G_G, G_P \rangle)$
10: until $\neg \text{removedAtLeastOneCandidate}$
11: $\text{components} \leftarrow \text{decompose}(\langle G_G, G_P \rangle)$
12: for all $c \in \text{components}$ do
13: $\text{assign}(c)$
14: end for

$\triangleright$ Assign GTFS stops having a single projection
$\triangleright$ Introduces cycleBreakers
$\triangleright$ Vertex in the middle has $\text{inDegree} = \text{outDegree} = 1$
$\triangleright$ Internal vertices have $\text{inDegree} = \text{outDegree} = 1$
$\triangleright$ Removes cycleBreakers that became redundant by fixing some vertices
$\triangleright$ Either by explicit discarding or as a consequence of fixing
$\triangleright$ Assignment by solution enumeration
- Example of a star
Algorithm 4.1  Determination of optimal assignment of projected stops \( S_p \) to GTFS stops \( S_G \).

1: \( S_p \leftarrow projStops(S_G) \)  
2: \( fixTrivial(S_G) \)  
3: \( \langle G_G, G_P \rangle \leftarrow graphFromBusStopSequences() \)  
4: repeat
5: \( removedAtLeastOneCandidate \leftarrow false \)  
6: \( handleTriples(\langle G_G, G_P \rangle) \)  
7: \( handleNonBifurcatingMaximalSequences(\langle G_G, G_P \rangle) \)  
8: \( handleStars(\langle G_G, G_P \rangle) \)  
9: \( reduceCycleBreakers(\langle G_G, G_P \rangle) \)  
10: until \( \neg removedAtLeastOneCandidate \)  
11: \( components \leftarrow decompose(\langle G_G, G_P \rangle) \)  
12: for all \( c \in components \) do
13: \( assign(c) \)  
14: end for

\( \triangleright \) Assign GTFS stops having a single projection  
\( \triangleright \) Introduces cycleBreakers

\( \triangleright \) Vertex in the middle has \( inDegree = outDegree = 1 \)

\( \triangleright \) Internal vertices have \( inDegree = outDegree = 1 \)

\( \triangleright \) Removes cycleBreakers that became redundant by fixing some vertices

\( \triangleright \) Either by explicit discarding or as a consequence of fixing

\( \triangleright \) Assignment by solution enumeration
Example of a component
Algorithm 4.1 Determination of optimal assignment of projected stops $S_P$ to GTFS stops $S_G$.

1. $S_P \leftarrow projStops(S_G)$
2. fixTrivial$(S_G)$  
   $\triangleright$ Assign GTFS stops having a single projection
   $\triangleright$ Introduces cycleBreakers
3. $(G_G, G_P) \leftarrow graphFromBusStopSequences()$
4. repeat
5.   removedAtLeastOneCandidate $\leftarrow$ false  
6.   handleTriples$(\langle G_G, G_P \rangle)$  
   $\triangleright$ Vertex in the middle has inDegree $= \text{outDegree} = 1$
7.   handleNonBifurcatingMaximalSequences$(\langle G_G, G_P \rangle)$  
   $\triangleright$ Internal vertices have inDegree $= \text{outDegree} = 1$
8.   handleStars$(\langle G_G, G_P \rangle)$
9.   reduceCycleBreakers$(\langle G_G, G_P \rangle)$  
   $\triangleright$ Removes cycleBreakers that became redundant by fixing some vertices
10. until $\neg$removedAtLeastOneCandidate  
11. components $\leftarrow$ decompose$(\langle G_G, G_P \rangle)$  
    $\triangleright$ Either by explicit discarding or as a consequence of fixing
12. for all $c \in \text{components}$ do
13.   assign$(c)$  
14. end for

Assignment by solution enumeration
Algorithm

- Reconstruct bus trips
  - Shortest path based
Results

Does it work?
Results

- **Network**
  - #nodes: 641 901
  - #links: 1 627 258

- **GTFS**
  - #bus stops: 30 654
  - #unique trips: 6 402

- **Projections**
  - #projected stop: 127 705
  - #average projected stops: 4

- **Algorithm**
  - 28 iterations
  - +/- 24 minutes
Results

- Visual inspection
Results

![Distribution of speed graph with curves labeled: Lognormal(Theta=0 Sigma=0.25 Zeta=3.43) and Gamma(Theta=0 Alpha=16.3 Sigma=1.96).]
The End!
Questions?