

Open Data for Open Transit

William Qi, Lap-Tak Chu, Saeid Allahdadian, Mina Park

CASCADIA URBAN ANALYTICS COOPERATIVE



Introduction

Building an open and inclusive public transportation network is a difficult task. Far too often, those who depend most on frequent and reliable transit live in the least accessible regions. With high costs and long turnaround times associated with traditional methods of data collection, many transit agencies just don't have the resources to effectively target their services.

We set out to tackle this problem by creating more economical and responsive methods of data approximation that depend only on widely available sources. These methods can be reproduced in almost any region, allowing for improvement of open transit with open data.

Why Open Data?

- Widely accessible to transit planners, no data storage policies or special permissions required
- Generally free of cost, a major advantage for budget-conscious transit agencies
- Commonly available open data sources include:
 - General Transit Feed Specification (GTFS) data
 - Regional census profiles
 - Social media chatter from Twitter and Facebook



Above: Example visualization of the transit network within the City of Surrey created using only open data sources. Coloured circles represent bus stops with varying levels of service (green is less, red is more) and green lines represent routes (thicker is higher capacity). The intensity of colour in the underlying map is proportional to population within each census tract.

How Can Graph Models Help?

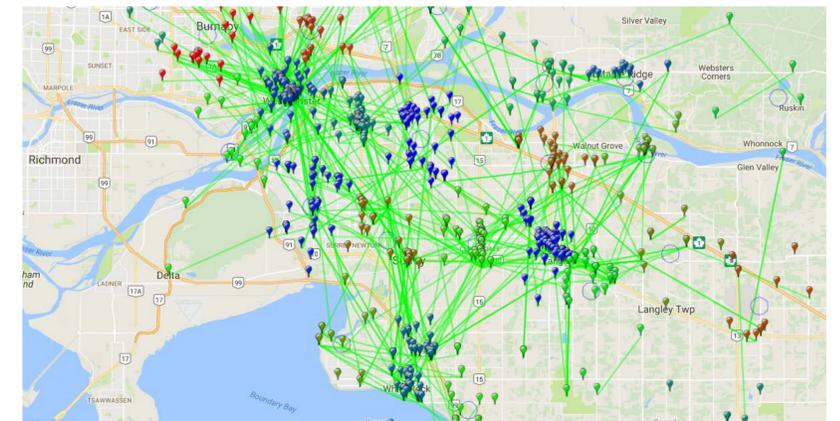
- Graph models are a natural representation for transportation networks as the system can be viewed as a collection of nodes (stops) connected by edges (route segments)
- Mathematical characterization of graphs can expose areas of concern such as underserved regions and potential bottlenecks
- We developed a tool to transform open GTFS data from any region into a corresponding graph model and computes significant metrics including:
 - Measures of centrality
 - Network connectivity
 - Network complexity
 - Clustering coefficients

Travel Patterns From Social Media

- Twitter is the most active "open" social network in North America, with over 1 million Tweets sent every week in Metro Vancouver alone
- Approximately 1% of these Tweets are geo-tagged
- Geo-tagged Tweets can be used as a proxy for location, allowing for characterization of travel patterns, as well as estimation of travel demand on each route
- Automated keyword extraction from responses directed at Twitter accounts of transit agencies can also be used to identify problematic stops/routes in the network
- Social media provides a promising solution for cheaper and more responsive analysis of network utilization

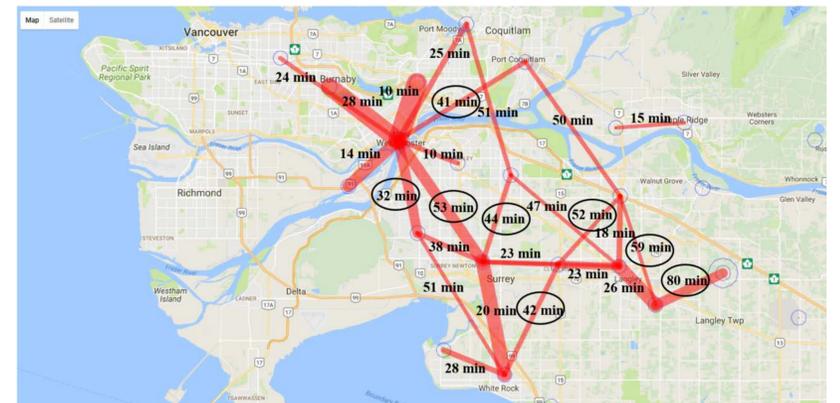
Case Study – City of Surrey

- In a case study using open data from the City of Surrey, we were able to correctly identify the 5 most important stops and the least well-served areas using graph metrics
- By extracting travel patterns from local Tweets and comparing the average speed of public transit on common commutes, we were also able to identify a candidate route for upgrades



Above: Visualization of travel between popular origins and destinations in the Surrey region. Markers are coloured based on sub-region and green lines indicate connections between origin-destination pairs.

Below: Visualization of relative travel demand on popular commuter routes to/from the City of Surrey (thicker indicates more interaction). Travel times on public transit are indicated for each of the routes with particularly slow segments circled.



Future Direction

- Our work with graph models has focused primarily on open data, but functionality could be significantly enhanced with flow data
- Analysis of social media trends becomes more accurate as additional data is collected; it would be worthwhile to revisit the case study in the future with more data