# Handling Bias in Large Urban Datasets **Using Smart Fare Card Data for Transit Planning** Mark Hallenbeck, Mayuree Binjolkar, Daniel Dylewsky, Andrew Ju, Wenonah Zhang, Michael Wolf

# **CASCADIA URBAN** ANALYTICS COOPERATIVE

# **Origin-Destination Data from Orca is Biased Against Short Duration Trips**



Example of ORCA Financial Transfers

## **ORCA** can combine Two Trips into one **Trip A + Trip B = long trip with transfer**

#### **Research Task:**

•To determine if a transfer was real or financial •If transfer is financial then create new origin – destination pairs

### **Chose Semi- Supervised Based Machine** Learning Approach for Classification of **Types of Transfers**

Labels learned from Label Spreading(KNN) Headway

Real Transfer
Financial Transfer

Outcome of Transfer Classification

**Transfer Classification Results by Stop** 



Number and Type of Transfer Activity

THE UNIVERSITY OF BRITISH COLUMBIA

#### **University of Washington**



0.1	MORE REAL
0.2	
0.3	
0.4	
0.5	
0.6	
0.7	
0.8	

**TRANSFERS** 

MORE FALSE

TRANSFERS



- to ORCA data



APC (Automatic Passenger Count) and ORCA Factors by Traffic Analysis Zones

### **Future Work**

- statistical bias analysis

# **Sponsoring Agencies**









# **STATISTICAL BIAS ANALYSIS**

Not all transit users are ORCA users **Need to account for cash-paying** customers for business planning

**Compare limited passenger count data** 

Total

**ORCA = Total** 

**ORCA Overestimates** Total

**Rebuilding origin-destination data files** using transfer analysis results

**Deploy Hidden Markov Models for** 



King County METRO



Puget Sound Regional Council

### **UBC** Data Science Institute