Spatial Structures and Dynamics

Spatial Networks and Flows

Keumsook Lee
Department of Geography, Sungshin Women’s University

MooYoung Choi
Department of Physics, Seoul National University
In this lecture, we introduce ideas and methods for analyzing spatial networks and flows, focused on the transportation networks and intra-urban passenger flows.

We utilize two large data sets of urban space (T-card transaction data & building GIS data).

We developed various data mining algorithms to exploring 1) passenger flows between origins and destinations, 2) dynamic edge flows, 3) trip patterns and stay times, 5) commuting behaviors, and land use characteristics, and also developed various dynamic visualization methods for exploring the diurnal passenger flow patterns.

We investigate the structural characteristics of transportation networks and passenger flows and their dynamic complexity.

Final goal of our study is to develop real-time population distribution model by integrating the diurnal passenger flows and land use information for providing urban disaster prevention models accounting the dynamic complexity of urban land uses and mobility of urban people.
OUTLINE

• Introduction

• Study Area and Data

• Analysis
  Structural properties of transportation systems
  Time-space characteristics of passenger flows
  Dynamics of passenger flows
  Complexity in transportation networks

• Concluding Remarks
Many endeavors have been made for analyzing the spatio-temporal characteristics of population distribution & movements.

Time-space analysis of people’s movements

- Hägerstrand (1970)
- Burnett (1978)
- Janelle and Goodchild (1983)
- Kitamura et al., (1990)
- Axhausen and Garling (1992)
- Janelle (1998)
INTRODUCTION

- Intra-urban passenger flows are the product of urban people’s everyday lives on the given transportation systems.

  People visit places where activities are located which are involved in their everyday lives. (work places, shopping places, home… etc.)

- The diurnal patterns of intra-urban passenger flows reveal the life styles and travel behaviors of cities as well as contain the rhythms of urban life and the spatial organization of the city.

- Understanding time-space characteristics of intra-urban passenger flows provide fundamental insights in transportation and urban studies and plannings.
Intra-urban Passenger Flows

Passenger Trips

Origin

Spatial movement

Time flows

Destination

Spacio-temporal phenomena

Intra-urban Passenger Flows

- Various passenger groups/travel purposes/ODs
- Different times / durations

Complex phenomena

Dynamic complexity analysis
STUDY AREA AND DATA

- Metropolitan Seoul
  - Seoul City
  - Incheon City
  - Gyeonggi Do
- The most densely developed area in Korea
Transportation Environments of the Study Area

Urban Characteristics

- Concentration of urban activities
  95% of entrepreneur headquarters
  85% of public administration offices

- Large Population Size
  Population.: 23 million
  49% of national total

- Urban sprawl: Vast expansion of urban area
  Several New Town Developments
  Area: 12,446 sq.km

- Rapid increase of car ownerships

Impact on Transportation

- Frequent passengers & freight flows
  Various functional linkages

- Large volume of traffic generation
  Various trip purposes

- Wide spectrums of trip chains
  Various origin-destination combinations

- Sharp increase in the auto car uses

Severe Traffic Congestion Problem
# Seoul Public Transport System Reform

*July 1, 2004*

<table>
<thead>
<tr>
<th>Objectives</th>
<th>Applied Policies</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Discourage auto car use</td>
<td>• Providing high quality bus services</td>
</tr>
<tr>
<td>• Motivate more people using public transportation</td>
<td>➢ Semi-Public Operation System</td>
</tr>
<tr>
<td>• Increase public transportation Shares</td>
<td>• Public transport network</td>
</tr>
<tr>
<td></td>
<td>➢ “One Seamless System”</td>
</tr>
<tr>
<td></td>
<td>• Real-time control over bus systems</td>
</tr>
<tr>
<td></td>
<td>➢ Unified Fare System</td>
</tr>
<tr>
<td></td>
<td>• Smart Card System</td>
</tr>
<tr>
<td><strong>Bus share</strong></td>
<td></td>
</tr>
<tr>
<td>(26% → 35%)</td>
<td></td>
</tr>
<tr>
<td><strong>Subway share</strong></td>
<td></td>
</tr>
<tr>
<td>(34% → 40%)</td>
<td></td>
</tr>
</tbody>
</table>
T-Card Data

- Smart card systems have been adapted to the transportation systems of Seoul completely since 2004.

- Over 12,000,000 transactions data have been recorded on the databases every day.

- T-card transaction databases hold the whole information of each passenger’s travel trajectory including times/locations of getting ons/ooffs of public transportation systems (bus, subway, and taxi).

- We have analyzed T-card data during the last 10 years.
## Data Structures

<table>
<thead>
<tr>
<th>Column ID</th>
<th>Form</th>
<th>Digit</th>
</tr>
</thead>
<tbody>
<tr>
<td>PCARD_NO</td>
<td>numeric</td>
<td>10</td>
</tr>
<tr>
<td>RIDE_DTIME</td>
<td>char</td>
<td>14</td>
</tr>
<tr>
<td>RIDE_STA_ID</td>
<td>char</td>
<td>7</td>
</tr>
<tr>
<td>TRANS_ID</td>
<td>char</td>
<td>3</td>
</tr>
<tr>
<td>TRAF_FREQ</td>
<td>numeric</td>
<td>3</td>
</tr>
<tr>
<td>TRANSP_METHOD_CD</td>
<td>char</td>
<td>3</td>
</tr>
<tr>
<td>BUS_ROUTE_ID</td>
<td>char</td>
<td>8</td>
</tr>
<tr>
<td>VEH_C_ID</td>
<td>char</td>
<td>9</td>
</tr>
<tr>
<td>RUN_DEPART_DTIME</td>
<td>char</td>
<td>14</td>
</tr>
<tr>
<td>PCARD_USER_CLASS_CD</td>
<td>char</td>
<td>2</td>
</tr>
<tr>
<td>RIDE_TRANSP_BIZR_ID</td>
<td>char</td>
<td>9</td>
</tr>
<tr>
<td>ALIGHT_STA_ID</td>
<td>char</td>
<td>7</td>
</tr>
<tr>
<td>ALIGHT_DT</td>
<td>char</td>
<td>14</td>
</tr>
<tr>
<td>PASGR_NUM</td>
<td>numeric</td>
<td>3</td>
</tr>
<tr>
<td>RIDE_AMT</td>
<td>numeric</td>
<td>10</td>
</tr>
<tr>
<td>RIDE_PENAL_AMT</td>
<td>numeric</td>
<td>10</td>
</tr>
<tr>
<td>ALIGHT_AMT</td>
<td>numeric</td>
<td>10</td>
</tr>
<tr>
<td>ALIGHT_PENAL_AMT</td>
<td>numeric</td>
<td>10</td>
</tr>
<tr>
<td>STAND_D</td>
<td>char</td>
<td>8</td>
</tr>
</tbody>
</table>
Unified Fare System with Smart Card

<table>
<thead>
<tr>
<th>Transaction unit:</th>
<th>A chain of sequential ride records comprising a trip from an origin to a final destination</th>
</tr>
</thead>
<tbody>
<tr>
<td>O₁ → Tₖ → Tₖ → Tₖ → Tₖ → Tₖ → Dⱼ</td>
<td></td>
</tr>
</tbody>
</table>

- **A single combined fare** is applied on ‘a transaction unit’
- **Free Transfer** between Buses within 30 minutes
- **Transfer Discount** between Subways & Buses
- **Discount** with T-cards

Distribute revenue to each Metro Cooperation & Bus Company according to passenger ride records everyday.

- **Require accurate passenger trajectory data**
Modal Shares of Passenger Traffics

  
  Car (26.3%)
  Bus (27.5%)
  Subway (34.8%)
  Taxi (6.5%)
  Walk, etc. (4.9%)

- Public transportation shares over 2/3 of total passenger traffics
ANALYSIS

- **Structural Properties of Transportation Systems**
  - Structural properties of transportation networks
  - Structural properties of passenger flows

- **Time-space Characteristics of Intra-urban Passenger Flows**
  - Flow patterns between origins and destinations
  - Hub structures of passenger flows
  - Classification of passenger trip patterns

- **Dynamic Visualization of Passenger Flows**
  - Various dynamic visualization methods development
  - Dynamic edge flow patterns
  - Diurnal patterns of intra-urban passenger flows
  - Real Time Population Distribution

- **Dynamic Complexity in Passenger Flows**
  - Dynamics of passenger flows
  - Complexity in transportation networks
ANALYSIS

- Public Transportation networks in the Metropolitan Seoul Area
  - Subway systems
  - Bus systems
  - Taxi systems
  - Statistical characteristics
  - Passenger trip behaviors & patterns

- Intra-urban passenger flows in the Metropolitan Seoul Area
  - One day data (Diurnal patterns)
  - One week data (Weekly patterns)
  - Smart Media (panel) data
  - Time-space characteristics
  - Dynamic visualization

- Real –time Population Distribution in the Metropolitan Seoul Area
  - Census data
  - Building information
  - Dynamic complex characteristics
  - Dynamic complexity analysis
Structural Properties of Transportation Systems

- Structural properties of transportation networks
- Structural properties of passenger flows
- Hub structures of passenger flows
- Classification of passenger trip patterns
Statistical Analysis of Complex Network

Small-world behaviors and truncated power-law distributions:
An active research topic in the physics


Numerous real networks observed:
- Biological and social systems
  Brockman (2006)
- Transportation systems
Examples of Complex Network

(a) collaborations in mathematics
(b) citations
(c) World Wide Web
(d) Internet
(e) power grid
(f) protein interactions

(Newman 2003)
Network properties


- Degree distribution
- Clustering coefficient
- Betweenness centrality
- Network efficiency

- Complexity exist in the system:
  Neither totally regular nor totally random

- Power-law distribution appears more frequently than normal distribution in the real world systems
Network of $N$ nodes

characteristic path length $L$, clustering coefficient $C^*$ (excl. terminals)

radius and diameter (min and max eccentricity) $R$ and $D$

**Efficiency of a network** consisting of $N$ nodes

$$\varepsilon \equiv \frac{1}{N(N-1)} \sum_{i \neq j} \frac{1}{r_{ij}}$$

$r_{ij}$: network distance (shortest path length $n_{ij}$, physical distance $d_{ij}$, or time distance $t_{ij}$) between nodes $i$ and $j$

Normalized efficiency with respect to that in the ideal (optimal) case (all-to-all connections, $N(N - 1)/2$ links)

$\rightarrow$ **Network efficiency** $E$ $(0 \leq E \leq 1)$
### Statistical Network Properties of MSS

<table>
<thead>
<tr>
<th>2005 Subway Network</th>
<th>network distance</th>
<th>physical distance</th>
</tr>
</thead>
<tbody>
<tr>
<td>$N$</td>
<td>380</td>
<td>←</td>
</tr>
<tr>
<td>$L$</td>
<td>20.0</td>
<td>27.9 km</td>
</tr>
<tr>
<td>$n_{max}$</td>
<td>62</td>
<td>139 km</td>
</tr>
<tr>
<td>$C^*$</td>
<td>0.00641</td>
<td>←</td>
</tr>
<tr>
<td>$D$</td>
<td>62</td>
<td>139 km</td>
</tr>
<tr>
<td>$R$</td>
<td>31</td>
<td>69.8 km</td>
</tr>
<tr>
<td>$E$</td>
<td>0.0786</td>
<td>0.747</td>
</tr>
</tbody>
</table>
Passenger flows on the subway system

Weight $w_{ij}$ of a link: passenger flows between $(i \rightarrow j$ and $j \rightarrow i)$

Strength $s_i$: departures from & arrivals at the station $i$

$$S_i \equiv \sum_{j=1}^{N-1} w_{ij}$$
Probability distribution of weights $w_{ij}$ - exhibits a **power-law behavior** - passenger flows do not have a characteristic size

Probability distribution of strengths $S_i$ - exhibits **log-normal distribution** - numbers of passengers at single stations have a characteristic size
Time-space Characteristics of Passenger Flows

- Flow patterns between origins and destinations
- Classification of passenger trip patterns
- Commuting behaviors
### Passenger Distribution by Time Zone

#### 3 Time-Zones: Morning / Day-time / Evening

<table>
<thead>
<tr>
<th>Get-off time</th>
<th>Morning</th>
<th>Day</th>
<th>Evening</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Get-on time</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Morning</td>
<td>1585358</td>
<td>86510</td>
<td>23</td>
<td>1671891</td>
</tr>
<tr>
<td>Day</td>
<td>0</td>
<td>995686</td>
<td>132796</td>
<td>1128482</td>
</tr>
<tr>
<td>Evening</td>
<td>0</td>
<td>0</td>
<td>2107168</td>
<td>2107168</td>
</tr>
<tr>
<td>Total</td>
<td>1585358</td>
<td>1082196</td>
<td>2239987</td>
<td>4907541</td>
</tr>
</tbody>
</table>
Distribution of Travel Times

![Graph showing the distribution of travel times with categories for morning, day, evening, and total. The x-axis represents travel time in minutes ranging from 0 to 210, and the y-axis represents the number of passengers ranging from 0 to 1,800,000. Different lines indicate different times of the day, with the total distribution being a combination of these categories.]
3 Time-Zones:

| Morning / Day-time / Evening |

- Weight (Passenger flows): $W_{ij}$

- Strength (Departures & Arrivals): $S_i \equiv \sum_{j=1}^{n-1} W_{ij}$

- Relationships with land use patterns
Passenger Strength Distribution

Departure

Morning

Day

Evening

Arrival
## Hub Cluster Groups

<table>
<thead>
<tr>
<th>Hub group</th>
<th>Inflow</th>
<th>Outflow</th>
<th>Inflow</th>
<th>Outflow</th>
<th>Inflow</th>
<th>Outflow</th>
<th>Hub characteristics</th>
</tr>
</thead>
<tbody>
<tr>
<td>Morning</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Day</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Evening</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CBD</td>
<td>17.67</td>
<td>17.67</td>
<td>13.00</td>
<td>13.00</td>
<td>8.67</td>
<td>16.67</td>
<td></td>
</tr>
<tr>
<td>Business center</td>
<td>12.80</td>
<td>12.80</td>
<td>1.80</td>
<td>1.80</td>
<td>1.20</td>
<td>2.40</td>
<td></td>
</tr>
<tr>
<td>Business(Mixed use)</td>
<td>9.00</td>
<td>9.00</td>
<td>5.00</td>
<td>5.00</td>
<td>6.86</td>
<td>6.14</td>
<td></td>
</tr>
<tr>
<td>Sub-CBD</td>
<td>3.57</td>
<td>3.57</td>
<td>12.00</td>
<td>12.00</td>
<td>12.43</td>
<td>10.43</td>
<td></td>
</tr>
<tr>
<td>Transportation center</td>
<td>2.33</td>
<td>2.33</td>
<td>21.67</td>
<td>21.67</td>
<td>21.00</td>
<td>24.00</td>
<td></td>
</tr>
</tbody>
</table>
Passenger Origins of Hubs

Old CBD

Regional Center

Transportation Center

Travel time:

Red : inflow
Blue : outflow
Passenger Origins of CBDs (Old vs New)
Passenger Trip Patterns

- Classification of passenger trip patterns
- Diurnal trajectory of commuting trip behaviors
- Time-space characteristics of passenger trip behaviors
## Classification of Trip Patterns

<table>
<thead>
<tr>
<th>Pattern</th>
<th>Trip Chain</th>
<th># Trips</th>
<th># Passengers</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pattern I</td>
<td>a ⟹ b</td>
<td>1</td>
<td>1,025,057</td>
<td>37.3</td>
</tr>
<tr>
<td>Pattern II</td>
<td>a ⟹ b, b ⟹ a</td>
<td>2</td>
<td>748,702</td>
<td>27.2</td>
</tr>
<tr>
<td>Pattern III</td>
<td>a ⟹ b, b ⟹ c</td>
<td>2</td>
<td>349,144</td>
<td>12.7</td>
</tr>
<tr>
<td>Pattern IV</td>
<td>a ⟹ b, c ⟹ a</td>
<td>2</td>
<td>151,010</td>
<td>5.4</td>
</tr>
<tr>
<td>Pattern V</td>
<td>a ⟹ b, c ⟹ d</td>
<td>2</td>
<td>123,708</td>
<td>4.5</td>
</tr>
<tr>
<td>Pattern VI</td>
<td>a ⟹ b ... b ⟹ a</td>
<td>≥3</td>
<td>23,030</td>
<td>0.8</td>
</tr>
<tr>
<td>Pattern VII</td>
<td>a ⟹ b ... y ⟹ a</td>
<td>≥3</td>
<td>135,752</td>
<td>4.9</td>
</tr>
<tr>
<td>Pattern VIII</td>
<td>a ⟹ b ... y ⟹ z</td>
<td>≥3</td>
<td>119,370</td>
<td>4.3</td>
</tr>
<tr>
<td>Pattern IX</td>
<td>not belong to pattern I</td>
<td>1</td>
<td>7,130</td>
<td>0.2</td>
</tr>
<tr>
<td>Pattern X</td>
<td>not belong to pattern II, III, IV, V</td>
<td>2</td>
<td>14,079</td>
<td>0.5</td>
</tr>
<tr>
<td>Pattern XI</td>
<td>not belong to pattern VI, VII, VIII</td>
<td>≥3</td>
<td>49,535</td>
<td>1.8</td>
</tr>
</tbody>
</table>
## Passenger Trip Behaviors

<table>
<thead>
<tr>
<th>#ride</th>
<th>#passengers</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1,032,187</td>
<td>37.5817</td>
</tr>
<tr>
<td>2</td>
<td>1,386,643</td>
<td>50.4873</td>
</tr>
<tr>
<td>3</td>
<td>236,197</td>
<td>8.5998</td>
</tr>
<tr>
<td>4</td>
<td>70,292</td>
<td>2.5593</td>
</tr>
<tr>
<td>5</td>
<td>15,623</td>
<td>0.5688</td>
</tr>
<tr>
<td>6</td>
<td>3,986</td>
<td>0.1451</td>
</tr>
<tr>
<td>7</td>
<td>1,053</td>
<td>0.0383</td>
</tr>
<tr>
<td>8</td>
<td>337</td>
<td>0.0122</td>
</tr>
<tr>
<td>9</td>
<td>109</td>
<td>0.0039</td>
</tr>
<tr>
<td>10</td>
<td>42</td>
<td>0.0015</td>
</tr>
</tbody>
</table>
Passenger Distribution over Time
Passenger Distribution over Time

Pattern I

Graph showing the distribution of passengers over time with two patterns: first-departure and first-arrival.
Passenger Distribution over Time

pattern II

- first-departure
- first-arrival
- second-departure
- second-arrival
Durations of Stay

Pattern 2  Pattern 4  Sum

0:00 3:00 6:00 9:00 12:00 15:00 18:00

0 10000 20000 30000 40000 50000 60000 70000
Passenger Distribution by Stay Time

type 1 : a->b  b->a Stay time (unit: 30minutes)
Commuting Time Distribution

![Commuting Time Distribution Graph]

- **Morning**
- **Evening**
## Return Types

<table>
<thead>
<tr>
<th>Return type</th>
<th># passengers</th>
<th>%</th>
<th>Pattern type</th>
</tr>
</thead>
<tbody>
<tr>
<td>One-way</td>
<td>1,617,279</td>
<td>58.8847</td>
<td>pattern 1,3,5,8</td>
</tr>
<tr>
<td>Round-trip</td>
<td>1,058,494</td>
<td>38.5395</td>
<td>pattern 2,4,6,7</td>
</tr>
</tbody>
</table>
Spatial distribution of passenger departures and arrivals

Departures in the morning
main residential areas

Arrivals in the evening
popular entertainment areas

black dots: subway stations
blue lines: ward boundaries in Seoul
Opening and closing a day

(a) Departure time distribution of the first trip
   • similar opening a day, i.e., leaving home at similar times

(b) Arrival time distribution of the last trip in the day
   • ant: single peak with shoulder
   • metrohopper/hybrid: double peaks (second peak ← after-work activities)

not closing yet: other means to home
higher second peak

Sleepless in Seoul!
Numbers of passengers in each category

<table>
<thead>
<tr>
<th>Category</th>
<th>Passenger</th>
<th>Ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 (Ant)</td>
<td>771,935</td>
<td>45.03%</td>
</tr>
<tr>
<td>2 (Metrohopper)</td>
<td>653,046</td>
<td>38.09%</td>
</tr>
<tr>
<td>3 (Hybrid)</td>
<td>289,349</td>
<td>16.88%</td>
</tr>
<tr>
<td>Total</td>
<td>2,746,517</td>
<td>100.0%</td>
</tr>
</tbody>
</table>

- Ant: 771,935 passengers (45.03%)
- Metrohopper: 653,046 passengers (38.09%)
- Hybrid: 289,349 passengers (16.88%)
- Total: 2,746,517 passengers (100.0%)
Dynamic Visualization of Passenger Flows

- Various dynamic visualization methods development
- Dynamic edge flow patterns
- Diurnal patterns of intra-urban passenger flows
- Real Time Population Distribution
## Edge Flow Algorithm

```plaintext
FindPassengerFlowsFromSmartCardTransactions()
{
    BuildSplitSubwayNetwork();
    BuildMergedSubwayNetwork();
    while (smart_card_transaction in the_transaction_file)
        SettingPassengerFlow(smart_card_transaction);
}
```

### SettingPassengerFlow(smart_card_transaction)

1. Extract get-on station ID, alight station ID, get-on time, get-off time from transactions of each subway passenger.
2. Determine indices of the array of station ID on the separated network based on the station ID hash table from the get-on station ID and alight station ID.
3. From the obtained indices, find the sequence of transit stations from the two-dimensional all-to-all shortest path array.
4. Transform the sequence of the transit stations on the separated network to the sequence of stations on the integrated network.
5. Increase of 1 in the slot of time-histogram corresponding to each link of the transit sequence between the get-on station and alight station.
```
Subway Station **Array** (ID, name, line-number, latitude, longitude, transfer-info)

Hash Table for Subway Station ID

Two-Dimensional **Link Matrix**

Link **Time-histogram** Array

Two-Dimensional **Adjacency Matrix** for Subway Station ID

Two-Dimensional Array for All-to-All **Shortest Paths**
Dynamic Visualization of Passenger Flows

- Visualization passenger Flows on MSSs over space and time.
- Edge flows per one minute

- Link Traffic
  - In flow
  - Out flow

- Accumulation
  - Up
  - Down

- Density
  - Color
  - Radius

Operational Function Keys
- Scale: Up/Down
- Speed: Left/Right
- Move: Mouse Dragging
- Stop/Go: Spacebar
- Exit: ESC
Diurnal Passenger Flow patterns of Major Edges

- Seoul stn --> City hall
- City hall --> Seoul stn
- KangNam --> KyoDae
- KyoDae --> KangNam
Stay Times

- Shot time stay (1~4 hours)
- Long time stay (9~12 hours)
CBDs in the Metropolitan Seoul Area

- Old CBD
- New CBD
Passenger Arrivals

- To where?
  - Shot time stay (1~4 hours)
  - Long time stay (9~12 hours)
Passenger Departure

- From where?
  - Shot time stay (1~4 hours)
  - Long time stay (9~12 hours)
Real Time Population Distribution

- Visualization of London Traffic [http://youtu.be/giVWTw76l4k](http://youtu.be/giVWTw76l4k)
Dynamic Visualization Program

Integration Land Use information & Passenger Flow

Building Information

• Mapping on RGB color space
• Red : House
• Green : Work
• Blue : Commercial
• Color = R + G + B
• Brightness : Ratio of Building Area
• Normalized by median

Operation : Mouse
SpaceBar : Stop & Go
↕ : scale (↑large, ↓small)
↔ : speed (← fast, → slow)
Propagation

- Convolution on image over 2 time slot

<p>| | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>+0.03</td>
<td>+0.04</td>
<td>+0.03</td>
</tr>
<tr>
<td>+0.04</td>
<td>-0.28</td>
<td>+0.04</td>
</tr>
<tr>
<td>+0.03</td>
<td>+0.04</td>
<td>+0.03</td>
</tr>
</tbody>
</table>

\[
f(d) = \frac{A}{2} \left[1 - \tanh \frac{d - d_1}{d_2}\right]
\]

Hyewha Station

\[d_1 \approx 460, \quad d_2 \approx 230\]

Passenger Distribution from a Subway Stop
CONCLUDING REMARKS

Purposes of Studies

- To develop the methods and tools for analyzing the ‘T-card transaction data’, and to investigate the spatio-temporal characteristics of passenger flows in the Metropolitan Seoul area.

- To develop various data-mining algorithms for exploring valuable traffic information from the big databases.

- To propose various dynamic visualization methods for analyzing the space-time characteristics of intra-urban passenger flows.

- To analyze the dynamic complexity of the intra-urban passenger flows and land uses.

- To real-time population distribution patterns by integrating passenger flow data, land use data, and smart media data.
Applications

- **Urban transport planning:**
  - operation scheduling by time, by edge, by line
  - preparing sustainable transportation systems based on the diurnal pattern of passenger flows and travel behaviors

- **Land use planning:**
  - location: activities, facilities, land use
  - functional regionalization: functional linkage

- **Urban disaster prevention planning** accounting the dynamic complexity of urban land uses and mobility of urban people
  - Account the socio-economic damage volumes (population, assets)
  - Provide optimal escape/rescue paths
Dynamic Complexity in Transportation Networks

- Dynamics of passenger flows
- Complexity in transportation networks
Dynamics of Passenger Flows
Complexity in Transportation Networks

Transportation networks

Spatial networks in the real world
- Biological systems: Anatomical network vs. Functional network
- Social networks: Router (geographic location) vs. WWW (topological network)

Subway system as a spatial network
- Geographical: Structure of the subway network
- Functional (weighted): Passenger O/D flow

Subway network as an evolving system
- Emergence of skew distributions
  A. actor collaboration
  B. WWW
  C. power grid

Power-law distributions?
E. Bullmore and O. Sporns

Dynamics of Passenger flow

Yule process model for the passenger flow
- Rich-get-richer process or preferential attachment
- Master equation $\rightarrow$ time evolution of passenger distributions

Analysis of the Seoul subway network
- Maximum spanning trees: power-law distributions
- Time evolution and passenger flows: log-normal or Weibull distributions

Complexity in Transportation Networks

Flow decreasing algebraically with (time) distance
- Gravity model modified by Hill function

Analysis of the Seoul bus network
- Power-law correlations $\rightarrow$ criticality
- Scaling and renormalization
Metropolitan Seoul Subway System

8 lines (Seoul) & 2 lines (Satellites), 357 stations (2005)
Minimal Spanning Tree

Prim (1957)
Maximum spanning tree of passenger flows

→ links of largest passenger O/D flows

Morning: many suburban residential areas & a few downtown business districts
Evening: more hubs of less degrees ← diffusion of passengers

Ants and Metrohoppers
Degree distribution of the maximum spanning tree

\[ P(x) \propto x^{-\alpha} \]

Power-law distribution
Subway passenger flow distributions

Weight $w$ of a link bet two stations:
  total number of passengers bet the two as O/D

Strength $s$ of a station:
  total number of passengers using it

Metropolitan Seoul Subway System (2005):
  357 stations, 127092 links
  about $6 \times 10^6$ transactions per day

\[ \chi^2 \text{ and likelihood function} \rightarrow \text{log-normal distributions} \]
Strength graphs fitted better to Weibull distribution.

Weight graphs fitted better to log-normal distribution.

Strength graphs fitted better to log-normal distribution.
How do such skew distributions emerge?

**Master equation approach**

Probability of passenger number configuration \( \{x_1, x_2, \ldots, x_N\} \)

\[
\frac{d}{dt} P(x_1, \ldots, x_N; t) = \sum_j \int dx_j' [\omega(x_j' \to x_j)P(x_1, \ldots, x_j', \ldots, x_N; t) - \omega(x_j \to x_j')P(x_1, \ldots, x_j, \ldots, x_N; t)]
\]

Change of passenger number: \( x \to x' = x + \Delta x \)

- Growth by the amount *proportional to the present number*: \( \Delta x = bx \) **Yule process**
- Transition rate \( \omega(x \to x') = \lambda \delta(x' - x - bx) \)

**Distribution function**

\[
f(x, t) = \frac{1}{N} \int dx_1 \cdots dx_N \sum_i \delta(x_i - x)P(x_1, \ldots, x_N; t)
\]

**Time evolution equation for the distribution function**

- Fixed subway lines or construction of new lines (without redistribution): Self-size production (incl. no production)

\[
\frac{\partial}{\partial t} f(x, t) = -\lambda f(x, t) + \frac{\lambda}{1 + b} f\left(\frac{x}{1 + b}, t\right)
\]

- Redistribution of the passenger flow: **Branching process**

\[
\frac{\partial}{\partial t} f(x, t) = -\lambda_1 f(x, t) + \frac{\lambda_1}{1 + b_1} f\left(\frac{x}{1 + b_1}, t\right) - \lambda_2 f(x, t) + \frac{\lambda_2}{1 + b_2} f\left(\frac{x}{1 + b_2}, t\right)
\]
General Solution

\[ f(x,t) = e^{-rt} \int dx' G_f(x,x';t) f(x',0) + \int_0^t dt' e^{-r(t-t')} \int dx' G_f(x,x';t-t')g(x',t') \]

Green's function

\[ G_f(x,x';t) = \frac{1}{x} G_F(\ln x, \ln x';t) \]

\[ G_F(X,X';t) = \int \frac{dk}{2\pi} \exp \left[ ik(X-X') - \lambda(1-e^{-ika})t \right] \]

\[ = \sum_{n \geq 0} \int \frac{dk}{2\pi} e^{ik(X-X')-\lambda t} \frac{(\lambda t)^n}{n!} e^{-ikan} \]

\[ = \sum_{n \geq 0} \frac{(\lambda t)^n}{n!} e^{-\lambda t} \delta(X-X'-an) \]

Long-time limit

\[ G_f(x,x';t) \xrightarrow{t \to \infty} \frac{1}{\sqrt{2\pi \sigma_t x}} \exp \left[ -\frac{(\ln x - \ln x' - \mu_t)^2}{2\sigma_t^2} \right] \]

\[ \mu_t = a\lambda t, \quad \sigma_t = a\sqrt{\lambda t}, \quad a = \ln(1+b) \]
A. Uniform size production \( g(x,t) = \delta(x-x_0) \)

asymptotic solution \( f_s(x) \equiv f(x,t \to \infty) \)

- From the general solution

\[
f_s(x) = \lim_{t \to \infty} r \int_0^t dt' e^{-r(t-t')} G_f(x,x_0;t-t') = \frac{r}{x} \int \frac{dk}{2\pi} \frac{e^{ik \ln(x/x_0)}}{r + \lambda(1-e^{-ika})}
\]

\[
G_f(x,x';t) = \frac{1}{x} \sum_{n \geq 0} \int \frac{dk}{2\pi} e^{ik \ln(x/x_0) - \lambda t} \frac{(\lambda t)^n}{n!} e^{-ikan}
\]

- Decomposition \( ka = 2\pi n - \theta \)

\[
f_s(x) = \frac{r}{ax} \sum_n e^{2i(\pi/a)n \ln(x/x_0)} \int_0^{2\pi} d\theta \frac{e^{-(i/a)\theta \ln(x/x_0)}}{2\pi} \frac{e^{-i\theta}}{r + \lambda - \lambda e^{i\theta}}
\]

- Dirac comb identity

\[
\sum_{k=-\infty}^{\infty} \delta(t-kT) = \frac{1}{T} \sum_{n=-\infty}^{\infty} e^{i2\pi nt/T}
\]

\[
f_s(x) = \frac{r}{ax} \sum_n \delta\left(n-a^{-1} \ln(x/x_0)\right) \int_0^{2\pi} d\theta \frac{e^{-i\theta}}{2\pi} \frac{e^{-i\theta}}{r + \lambda - \lambda e^{i\theta}}
\]

\( z = e^{i\theta} \)
• With \( x_n = x_0 e^{an} = x_0 (1 + b)^n \)

\[ f_s(x) = \frac{r}{r + \lambda} \sum_{n \geq 0} \left( \frac{x}{x_0} \right)^{-\frac{\ln(1+r/\lambda)}{\ln(1+b)}} \delta(x - x_n) \]

• Stationary solution \( \frac{\partial f}{\partial t} = 0 \), with an ansatz \( f_s(x) \sim x^{-\alpha} \)

\[ \alpha = 1 + \frac{\ln(1 + r / \lambda)}{\ln(1 + b)} \]

• Continuous growth limit \( (u = an) \)

\[ f_s(x) \approx \frac{r}{r + \lambda} \frac{1}{ax} \int_{0}^{\infty} du \left( \frac{x}{x_0} \right)^{1-\alpha} \delta(u - \ln(x / x_0)) = \frac{r}{\lambda bx_0} \left( \frac{x}{x_0} \right)^{-\alpha} \theta(x - x_0) \]

power-law distribution

power law, with exponent \( \alpha \) taking any value (greater or less than two)

\( \leftrightarrow \) usual approach based on Yule process, predicting \( \alpha > 2 \)
B. Self-size production \[ g(x, t) = f(x, t) \]

- Evolution equation
  \[ \frac{\partial}{\partial t} f(x, t) = \lambda f(x, t) + \frac{\lambda}{1 + b} f\left(\frac{x}{1 + b}, t\right) \]

- Effectively the same as the case of no production
  \[ f(x, t) = \int dx' G_f(x, x'; t) f(x', 0) \]

- Initially, all the same size: \[ f(x, t = 0) = \delta(x - x_0) \]

  \[ f(x, t) = G_f(x, x_0; t) = \frac{1}{\sqrt{2\pi\sigma_t x}} \exp\left[\frac{-(\ln x - \mu_t)^2}{2\sigma_t^2}\right] \]

  \[ \mu_t = a\lambda t + x_0, \quad \sigma_t = a\sqrt{\lambda t}, \quad a \equiv \ln(1 + b) \]

  log-normal distribution
Log-normal vs. Weibull distributions

Log-normal distribution  \( f(x, t) = \frac{1}{\sqrt{2\pi\sigma x}} e^{-\frac{(\ln x - \mu)^2}{2\sigma^2}} \)

Weibull distribution  \( f(x, t) = \frac{\gamma}{x} \left( \frac{x}{\eta} \right)^{\gamma-1} e^{-\left(\frac{x}{\eta}\right)^\gamma} \)

Power-law distribution, asymptotically  \( \sim x^{-1} \)

Asymptotic solution of the evolution equation

- No production or Self-size production  \( \rightarrow \)  Log-normal distribution
  \( \mu = a\lambda t, \quad \sigma = a\sqrt{\lambda t} \quad \text{with} \quad a \equiv \ln(1+b) \)

- Branching process  \( \rightarrow \)  Weibull distribution
  \( \gamma = \frac{1}{\sqrt{a_1^2\lambda_1 + a_2^2\lambda_2}} t, \quad \eta = \exp\left[ (a_1\lambda_1 + a_2\lambda_2) t \right] \quad \text{with} \quad a_1 \equiv \ln(1+b_1), \ a_2 \equiv \ln(1+b_2) \)
### Time evolution of the passenger flow

<table>
<thead>
<tr>
<th>Time</th>
<th>Strength</th>
<th>Weight</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Arrival</td>
<td>Departure</td>
</tr>
<tr>
<td>Morning</td>
<td>$\mu$</td>
<td>7.828</td>
</tr>
<tr>
<td></td>
<td>$\sigma$</td>
<td>1.068</td>
</tr>
<tr>
<td></td>
<td>$b$</td>
<td>0.157</td>
</tr>
<tr>
<td></td>
<td>$\lambda t$</td>
<td>53.70</td>
</tr>
<tr>
<td>Afternoon</td>
<td>$\mu$</td>
<td>7.507</td>
</tr>
<tr>
<td></td>
<td>$\sigma$</td>
<td>1.062</td>
</tr>
<tr>
<td></td>
<td>$b$</td>
<td>0.162</td>
</tr>
<tr>
<td></td>
<td>$\lambda t$</td>
<td>49.98</td>
</tr>
<tr>
<td>Evening</td>
<td>$\mu$</td>
<td>8.326</td>
</tr>
<tr>
<td></td>
<td>$\sigma$</td>
<td>0.976</td>
</tr>
<tr>
<td></td>
<td>$b$</td>
<td>0.121</td>
</tr>
<tr>
<td></td>
<td>$\lambda t$</td>
<td>72.70</td>
</tr>
</tbody>
</table>

- **Morning arrival, evening departure: Downtown business districts** …
  - Fewer number of drastic changes
- **Morning departure, evening arrival: Suburban residential areas** …
  - Gradual changes with higher frequency
- **Afternoon:** Stations, terminals, tourist spots …
  - E.g., Express bus terminal, with rather large afternoon weights
- **Evening:** Not reverse of morning
  - *Diffusive* to entertainment districts or part-time work places

Growth factor $b$
Number of events occurred $\lambda t$
Growth due to the influx from outside of the system → Log-normal distribution

- Population influx from other cities, Influx from other transportations

Redistribution of the population → Weibull distribution

- Construction of stations: Change of the catchment areas of the stations
- Passengers using pre-existing stations: Branching process

→ Those still using pre-existing stations + Others using the new station

### Log-normal vs. Weibull distribution again

<table>
<thead>
<tr>
<th></th>
<th>Morning</th>
<th>Daytime</th>
<th>Evening</th>
<th>All day</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$\chi^2$</td>
<td>$\ln L$</td>
<td>$\chi^2$</td>
<td>$\ln L$</td>
</tr>
<tr>
<td></td>
<td>Log-normal</td>
<td>Log-normal</td>
<td>Log-normal</td>
<td>Log-normal</td>
</tr>
<tr>
<td></td>
<td>$2.02 \times 10^{-5}$</td>
<td>$-3324.7$</td>
<td>$4.98 \times 10^{-5}$</td>
<td>$-3208.2$</td>
</tr>
<tr>
<td></td>
<td>$3.25 \times 10^{-5}$</td>
<td>$-3354.0$</td>
<td>$5.00 \times 10^{-5}$</td>
<td>$-3218.9$</td>
</tr>
<tr>
<td></td>
<td>Log-normal</td>
<td>Weibull</td>
<td>Log-normal</td>
<td>Weibull</td>
</tr>
<tr>
<td></td>
<td>$2.95 \times 10^{-5}$</td>
<td>$-3367.9$</td>
<td>$4.61 \times 10^{-5}$</td>
<td>$-3222.3$</td>
</tr>
<tr>
<td></td>
<td>$2.77 \times 10^{-5}$</td>
<td>$-3363.5$</td>
<td>$6.12 \times 10^{-5}$</td>
<td>$-3230.2$</td>
</tr>
<tr>
<td></td>
<td>Log-normal</td>
<td>Weibull</td>
<td>Log-normal</td>
<td>Weibull</td>
</tr>
<tr>
<td></td>
<td>$7.61 \times 10^{-5}$</td>
<td>$-3597.6$</td>
<td>$5.83 \times 10^{-5}$</td>
<td>$-3462.2$</td>
</tr>
<tr>
<td></td>
<td>$8.02 \times 10^{-5}$</td>
<td>$-3597.5$</td>
<td>$7.83 \times 10^{-5}$</td>
<td>$-3472.6$</td>
</tr>
</tbody>
</table>

Morning departure, Evening arrival (suburbs/residential area) → Weibull distributions
Coupling in urban system: Passenger flow vs. (time) distance

\[ F_{ij} = G \frac{M_i M_j}{r_{ij}^\alpha} \quad \text{or} \quad f_{ij} = \frac{F_{ij}}{M_i M_j} = \frac{G}{r_{ij}^\alpha} \to \log f_{ij} = -\alpha \log r_{ij} + C \]

\( F_{ij} \): passenger flow between station \( i \) and \( j \)
\( M_i \): mass (strength) of station \( i \)
\( r_{ij} \): (time) distance between station \( i \) and \( j \)

\( \alpha \)

\( G \)

\( C \)

**gravity model**

![Graphs showing deviations at short distances](image)

Krings et al. (2009)  
Jung et al. (2008)  
deviations at short distances
Passenger flow in Seoul subway system

gravity model \[ \log f_{ij} = -\alpha \log r_{ij} + C \]

- gravity exponent \( \alpha = 1.94 \) except Line 2 (circle)
- curvature around \( \log r_{ij} \approx 6.5 \rightarrow \text{modified gravity model} \)
Reduction at short distances $\rightarrow$ modification by the Hill function

$$f_{ij} = \frac{G}{r_{ij}^\alpha} g(r_{ij}) \eta_{ij} \text{ with } g(r_{ij}) = \frac{r_{ij}^n}{r_{ij}^n + K^n}$$

- gravity exponent $\alpha = 1.94$
- Hill coeff. $n = 3.0$: # of available transportation means (enzyme-substrates)
- time constant $K = 17$ min

characteristic time distance

For all lines (but line 2) universality

$\eta_{ij}$: Yule-type (multiplicative) fluctuations $\rightarrow$ log-normal distribution

$$\mu = 1, \sigma = 0.703$$
• 12986 stops
• 590 bus routes
• About 6,000,000 trips per weekday
  (10~16 April, 2011)
Correlations in the bus system

correlation sum $C$ of bus stops vs distance $R$

$$C(R) = \frac{2}{N(N-1)} \sum_{i=1}^{N} \sum_{j=i+1}^{N} \theta(R-||\vec{s}_i - \vec{s}_j||) \sim R^d$$

Geographical dimension $d = 1.06(1)$ for $R < 200$ m

$1.78(1)$ for $200$ m $< R < 10$ km

finite-size effects $\sim 200$ m
Correlation function

correlations between stop strengths

\[ \Gamma(r) = \langle \tilde{s}_i \tilde{s}_j \rangle - \langle \tilde{s}_i \rangle \langle \tilde{s}_j \rangle = \langle \tilde{s}_i \tilde{s}_j \rangle - 1 \]

where \( \tilde{s}_i \equiv \frac{s_i}{\langle s_i \rangle_t} \) is the normalized strength of stop \( i \)

- Long-range correlations
  \( \Rightarrow \) Range 3 sec ~ 3600 sec (1h)
- Criticality of the bus system
Renormalization

Coarse graining by combining bus stops to form a block stop
box renormalization: partitioning into square boxes of size $a$
node renormalization: taking $p$ neighboring stops

Increase the box/supernode size: scale transformation
 REGARD bus stops in given box/supernode as one block stop (stops = degrees of freedom): reduction in degrees of freedom
Renormalization

Coarse graining by combining bus stops to form a block stop
  box renormalization: partitioning into square boxes of size $a$
  node renormalization: taking $p$ neighboring stops

$a = 0$ m: 12986 stops (bare stops)
  222 m: $\sim 4450$ block stops
  476 m: $\sim 1787$ block stops
  927 m: $\sim 608$ block stops

box renormalization

node renormalization

scale transf. 
reduction in d.o.f. $\rightarrow$ renormalization

box size $a \downarrow$ $\rightarrow$ lattice size $L \uparrow$

thermodynamic limit $\leftrightarrow$ infinite number of bus stops

$p = 32$ or $a = 1394$ m
Correlation function

box renormalization \((a = 476 \text{ m})\)

\[ \Gamma(r) \sim r^{-0.95 \pm 0.07} \]

node renormalization \((p = 6 \text{ or } a = 466 \text{ m})\)

\[ \Gamma(r) \sim r^{-0.91 \pm 0.07} \]
Dimensions and density

Network density

\[ \rho \equiv \frac{2N_w}{N_s(N_s-1)} \propto L^{2(d-d_r)} \]

- # of nodes (stops) \( N_s \propto L^d \)
- # of links (O/D on a route) \( N_w \propto L^{2d_r} \)

System size \( L \sim a^{-1} \) with the effective box size \( a \)

Network density \( \rho \sim a^{1.4} \sim L^{-1.4} \)

Route dimension \( d_r \approx 1.08 \)

Network dimension \( d \approx 1.78 \)

critical region: \( 200 \text{ m} \leq a \leq 1\text{ km} \)
Passenger flow vs. (time) distance: Modified gravity model

Route # 271

\[ \alpha = 1.8(1), \; n = 2.0(1), \; K = 950(90) \text{ sec} \]

Whole system

\[ \alpha = 2.00(1), \; n = 2.50(1), \; K = 291(4) \text{ sec} \]

\( n \): number of transportation means, integer?
Renormalization

Box renormalization ($a = 766$ m)

Node renormalization ($p = 12$ or $a = 759$ m)
Finite-size scaling: Box renormalization

Time constant $K$ vs Box size $a$

Gravity exponent $\alpha$ and Hill coefficient $n$

box size $a \leq 200$ m $\Rightarrow$ finite-size effects

Renormalized values: $\alpha = 1.67$, $n = 2$ (integer) cf. subway: $n = 3$
Finite-size scaling: Node renormalization

Time constant $K$ vs Box size $a$

Gravity exponent $\alpha$ and Hill coefficient $n$

Renormalized values: $\alpha = 1.71$, $n = 2$
## Renormalized parameters

<table>
<thead>
<tr>
<th></th>
<th>$\alpha$</th>
<th>$n$</th>
<th>$K$ (s)</th>
<th>Reduced $\chi^2$</th>
<th>Box renormalization</th>
<th>Node renormalization</th>
</tr>
</thead>
<tbody>
<tr>
<td>Monday</td>
<td>2.00(1)</td>
<td>2.50(1)</td>
<td>291(4)</td>
<td>2.80</td>
<td>1.66(2)</td>
<td>1.69(4)</td>
</tr>
<tr>
<td>Tuesday</td>
<td>2.00(1)</td>
<td>2.51(1)</td>
<td>292(4)</td>
<td>2.78</td>
<td>1.67(2)</td>
<td>1.73(2)</td>
</tr>
<tr>
<td>Wednesday</td>
<td>2.01(1)</td>
<td>2.53(1)</td>
<td>292(4)</td>
<td>2.78</td>
<td>1.68(2)</td>
<td>1.71(3)</td>
</tr>
<tr>
<td>Thursday</td>
<td>2.00(1)</td>
<td>2.52(1)</td>
<td>292(4)</td>
<td>2.79</td>
<td>1.68(2)</td>
<td>1.74(2)</td>
</tr>
<tr>
<td>Friday</td>
<td>2.00(1)</td>
<td>2.51(1)</td>
<td>292(4)</td>
<td>2.78</td>
<td>1.68(2)</td>
<td>1.74(2)</td>
</tr>
<tr>
<td>Saturday</td>
<td>1.94(1)</td>
<td>2.48(1)</td>
<td>282(4)</td>
<td>2.85</td>
<td>1.61(2)</td>
<td>1.68(2)</td>
</tr>
<tr>
<td>Sunday</td>
<td>1.87(1)</td>
<td>2.39(1)</td>
<td>263(4)</td>
<td>2.89</td>
<td>1.52(2)</td>
<td>1.61(2)</td>
</tr>
</tbody>
</table>

Hill coefficients $n \leftarrow$ number of transportation means
Gravity exponent $\alpha \leftarrow$ dimension of the system
Concluding Remarks

Skew distributions
- Power-law distributions of maximum spanning trees
- Log-normal and Weibull distributions of weights and strengths
- Skew distributions manifesting *complexity* in the evolving system

Growth of passenger flows
- Rapid and drastic but infrequent growth ← absorbing populations from other regions
- Gradual and continuous redistribution of population throughout residential areas

Time-zone dependence of flows
- Afternoon: Most drastically evolving distribution
- Evening: *Diffusion* of passengers
Time-distance dependence of flows

- **Power-law** behavior: *universality* for lines of the same topology
- Short-distance reduction ← gravity model modified by **Hill function**

**Bus system**

- **Power-law** correlations between strengths → **criticality**
- **Scaling** behavior → renormalization of exponent and Hill coefficient

Complex system approach applied successfully to **transportation networks and passenger flow**, revealing emergent complexity → complex system science of **social (complex) systems**

Physics viewpoint: statistical mechanics dealing with many-particle *(complex)* systems

- *matter* → condensed matter physics
- *life* → biological physics
- *society* → social physics? cf. A. Comte
Collaboration

Segun Goh, *Center for Theoretical Physics, Seoul National University*

Jong Soo Park, *Department of Information Technology, Sungshin University*

Ho-Sung Kim, *Department of Media Communication, Sungshin University*

Jean-Yves Fortin, *Institut Jean Lamour, H. Poincaré Université Nancy*

THANK YOU!