Analysis of Network Flow Data

Chapter 9.1 & 9.2
Gravity Models

Lauri Eskor
19. may 2010
Summary

- Introduction to network flow
- Gravity models
- Examples
Routing matrix

\[ B_{e;ij} = \begin{cases} 
1, & \text{if link } e \text{ is traversed in going from } i \text{ to } j, \\
0, & \text{otherwise}.
\end{cases} \]
Routing matrix

- G is tree -> B is binary matrix
- Multiple routes -> $B_{e,ij}$ = fraction of flow
Routing matrix

\[ B_{e;ij} = \]

\[
\begin{array}{c|ccccc}
   & 1 & 2 & 3 & 4 & 5 \\
\hline
AC & 1 \\
AB & 1 \\
AD & 1 & 1 \\
AE & 1 & 1 & 1 \\
AF & 1 & 1 & 1 \\
BD & 1 \\
BE & 1 & 1 \\
BF & 1 & 1 \\
DE & 1 \\
DF & 1 \\
\end{array}
\]
Traffic (or origin-destination) matrix

$Z_{ij} = \text{the total volume of traffic flowing from an origin vertex } i \text{ to a destination vertex } j \text{ in a given period of time}$
Traffic (or origin-destination) matrix

Net in-flow

\[ Z_{i+} = \sum_j Z_{ij} \]

Net out-flow

\[ Z_{+j} = \sum_i Z_{ij} \]
Link totals

\( X_e \) to be the total flow over a given link \( e \in E \)

\[ X = (X_e)_{e \in E} \]

\[ X = BZ \]

traffic matrix written as a vector.
Cost

- \( c \) - associated with paths or links
  - Generalized cost
    - behavior of consumers of transportation resources
  - Quality of Service (QoS)
    - Monitor basic performance characteristics
    - Detect areas of anomaly
    - Validate compliance with service agreements
A snapshot of network can be analyzed
Time dependent traffic matrix $Z^{(t)}$
Routing matrix $B$ is fixed
- changes in routing occur in longer time scales
Analysis of network flow

- Dependant of measurements taken and goal of statistical analysis
- Gravity models
  - Entire traffic matrix $Z$ can be observed
- Traffic matrix estimation (dynamic models)
  - If link totals are known
- Estimation of network flow costs
  - To model and interference network cost parameters
Gravity models

- First known work by Carey in the 1850's
- Formulated by Stewart in 1941
- Developed mainly in social sciences
- For describing aggregate levels of interaction among the people of different populations - prediction of traffic volume
- Interaction among two populations depends on:
  - size of population
  - some measure of separation
General gravity model

\[ E(Z_{ij}) = h_O(i) \, h_D(j) \, h_S(c_{ij}) \]

\[ h_O(i) \, h_D(j) \] origin and destination functions (population size)

\[ h_S(c_{ij}) \] separation function over separation parameters (distance, cost etc)
\[
\log TRD_{ij} = a + DIST_{ij}^b + GDP_{i}^c + GDP_{j}^d + \sum e_j OCEAN + f \cdot BORDER + \sum g_j REGION_{ft} + \sum h_j INTRAREGION + u_{ij}
\]
Gravity models example

- Austrian call data
  - 32 districts
  - throughout a period
  - total $32 \times 31 = 992$ measurements
- $z_{(i,j)} =$ contact intensity over period studied
- $h(i)$ and $h(j) =$ gross regional product (~size of economy)
- $c(i,j) =$ distance between districts
Gravity models example

\[ E(Z_{ij}) = \gamma (\pi_{O,i})^\alpha (\pi_{D,j})^\beta (c_{ij})^{-\theta} \]

- Origin and destination functions are power functions
- 'Demographic gravitational constant' is used
Interaction probabilities

\[ f_{ij} = \mathbb{E}(Z_{ij}/Z_{++} | Z_{++} > 0) \]

\[ Z_{++} = \sum_{i,j} Z_{ij} \]

Expected relative frequency at which interactions are specifically \( ij \)-interactions

\[ f_{ij} = \frac{h_O(i) \ h_D(j) \ h_S(c_{ij})}{\sum_{i' \in \mathcal{I}, j' \in \mathcal{J}} h_O(i') \ h_D(j') \ h_S(c_{i'j'})} \]
Conditional gravity model

All destinations $j$ from a given origin $i$

$$f_{j|i} = \frac{h_D(j) h_S(c_{ij})}{\sum_{j' \in \mathcal{J}} h_D(j') h_S(c_{ij'})}$$
Inference of gravity models

Focus on model

\[ \log \mu_{ij} = \alpha_i + \beta_j + \theta^T c_{ij} \]

\[ \alpha_i = \log h_O(i), \beta_j = \log h_D(j), \text{ and } \theta, c_{ij} \in \mathbb{R}^K \]

\( Z = z \) (IJ)x1 vector of observations of the flows \( Z_{ij} \) that are ordered by origin i, and by destination j

Maximum likelihood estimates for these parameters
Inference of gravity models

\[ \hat{\mu}_{ij} = \hat{\alpha}_i \hat{\beta}_j \exp(\hat{\theta}^T c_{ij}) \]

\[ \hat{\mu}_{i+} = z_{i+}, \text{ for } i \in \mathcal{I} \quad \text{and} \quad \hat{\mu}_{+j} = z_{+j}, \text{ for } j \in \mathcal{J} \]

\[ \sum_{i,j \in \mathcal{I} \times \mathcal{J}} c_{ij;k} \hat{\mu}_{ij} = \sum_{i,j \in \mathcal{I} \times \mathcal{J}} c_{ij;k} z_{ij}, \text{ for } k = 1, \ldots, K \]

where \( \hat{\mu}_{i+} = \sum_{j \in \mathcal{J}} \mu_{ij} \) and \( \hat{\mu}_{+j} = \sum_{i \in \mathcal{I}} \mu_{ij} \)
Example

• Austrian call data
• Standard and general gravity models:

\[ \mathbb{E}(Z_{ij}) = \gamma (\pi_{O,i})^\alpha (\pi_{D,j})^\beta (c_{ij})^{-\theta} \]

\[ \log \mu_{ij} = \alpha_i + \beta_j - \theta \log(\text{Distance}) \]

• Fitted using generic iteratively weighted least-squares method for generalized linear models
• Model arguments are significant at the 0,05 level
Fitted values vs observed values

Standard model

General model
relative errors vs flow volumes

Standard model

General model
The end