Visualization of origin-destination data

From an information theoretic perspective

Martijn Tennekes

Ukraine Statistics Day, December 4, 2020

NeEDS – Network of European Data Scientists

https://riseneeds.eu/

Martijn Tennekes (CBS) and Min Chen (University of Oxford)
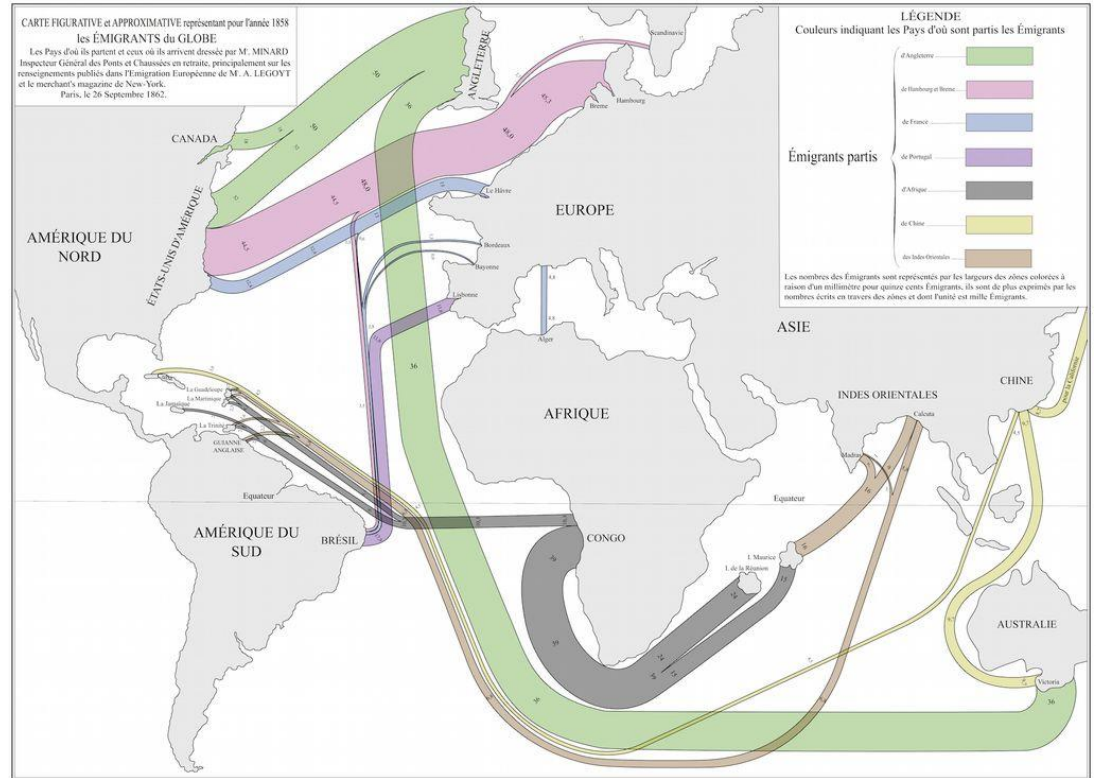
Origin-destination data

- **Origin-destination (OD)** data describe movements from origin to destination.

- Applications: **migration**, **passenger transport**, **export of goods**, **movement of animals**, **spreading of deceases**, etc.

- A *raw* OD dataset describes movements of individuals whereas an *aggregated* OD dataset describes countable flows of individuals.

- Example dataset: commuting flows between 390 Dutch municipalities.

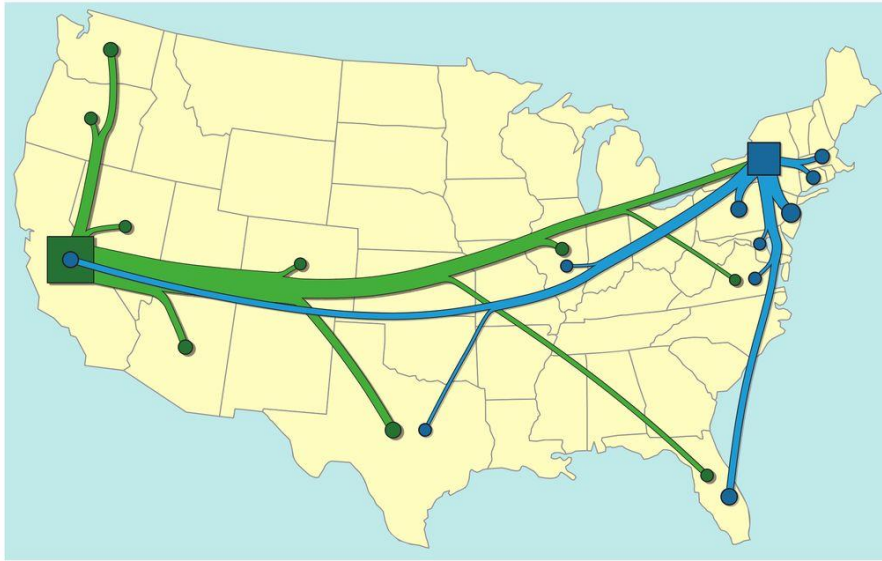# Visualization of origin-destination data
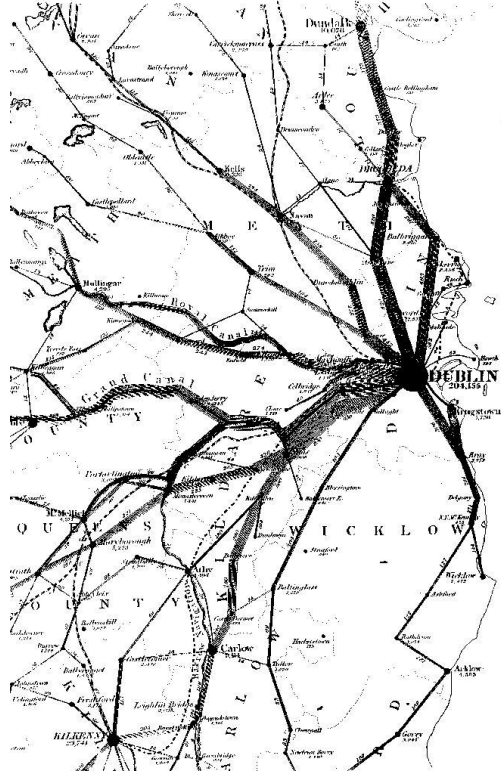
Public transport (Harness, 1837)

World migration (Minard, 1858)

4

Visualization of origin-destination data

(a) Original flows   (b) Filtered flows   (c) Aggregated flows

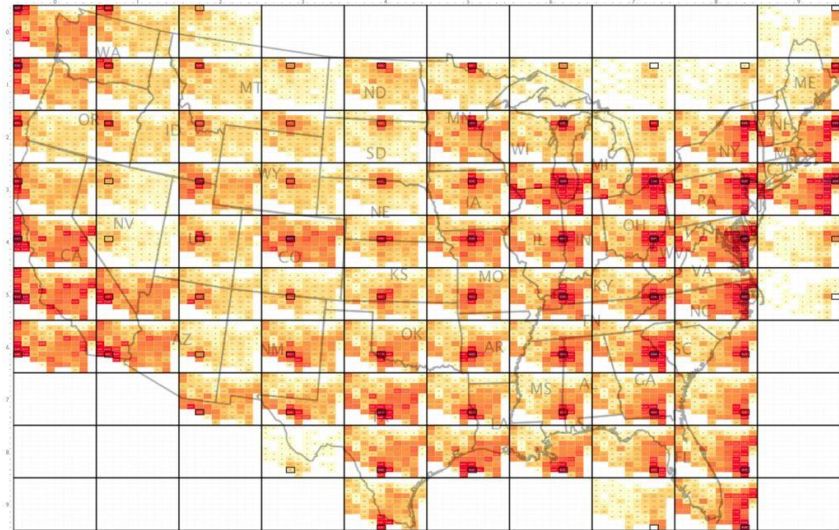Spiral trees (Verbeet et al., 2011)   MobilityGraphs (Landesberger et al., 2016)

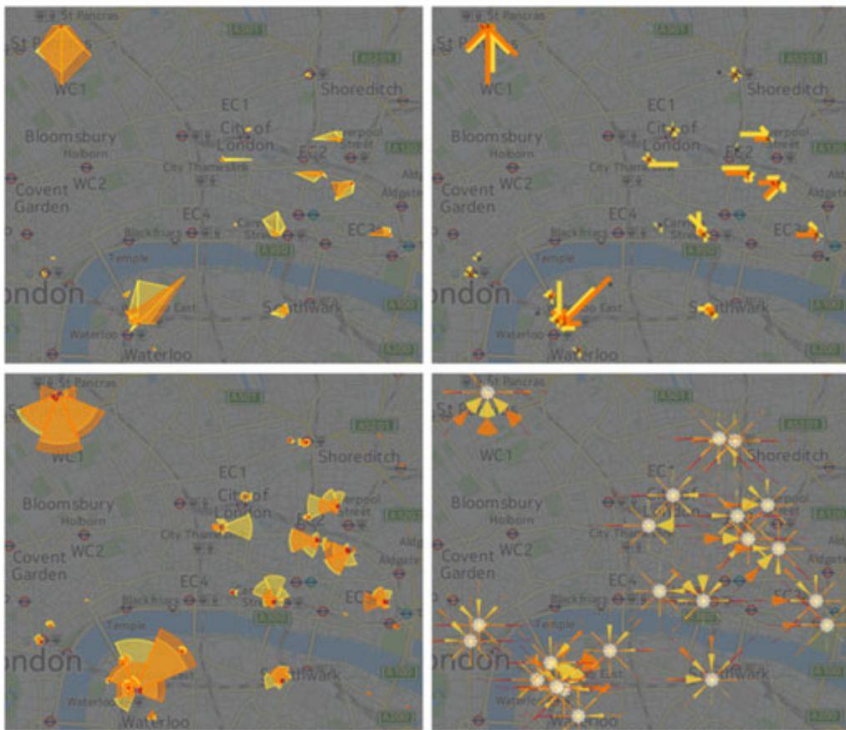# Visualization of origin-destination data
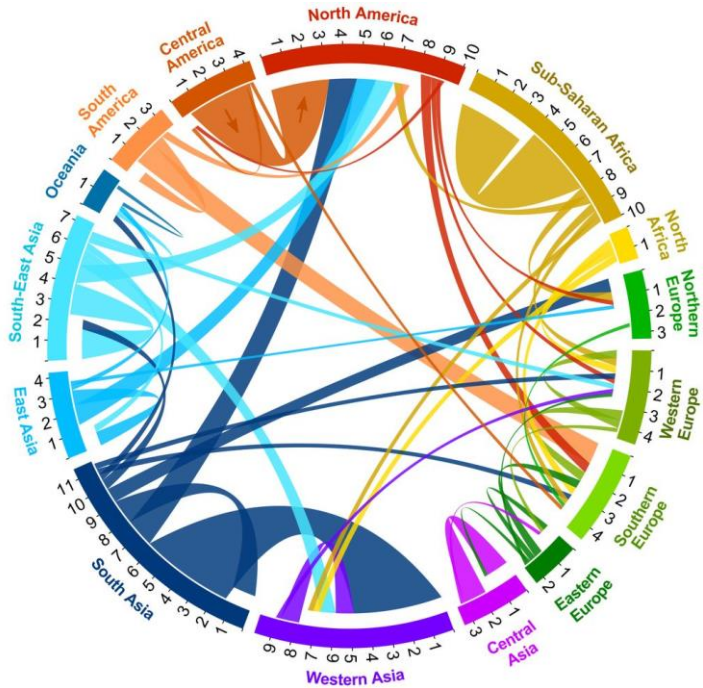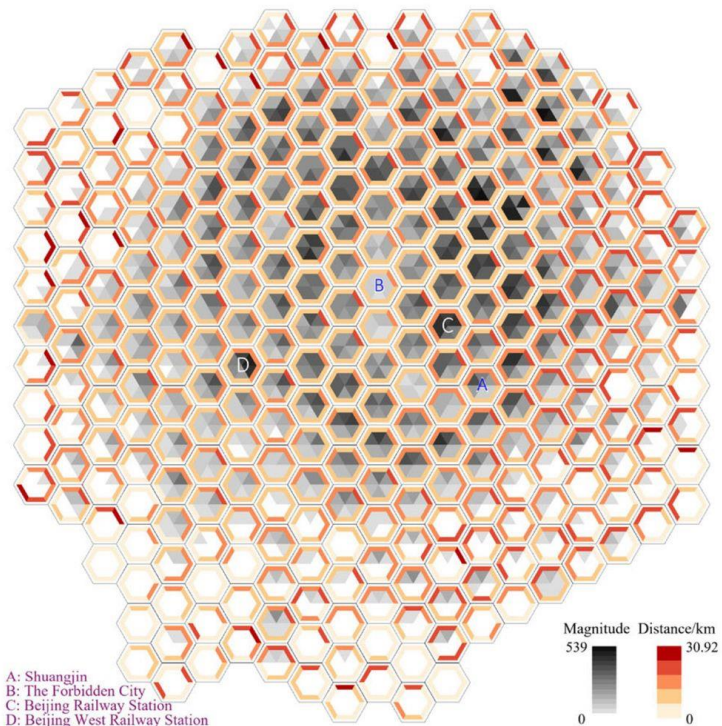

Circular plot (Abel et al., 2014)

OD maps (Wood et al., 2010)

Visualization of origin-destination data

Flow diagrams (Adrienko et al., 2017)

Pattern maps (Yao et al., 2019)

Design space of OD visualization

- Our aim was to **organize** existing OD visualizations systematically and **discover** new OD visualization methods.

- How? By introducing a **design space** which includes all OD visualization methods that are theoretically possible.

- Result: a design space of **four dimensions** using **information theory**, which enabled us to discover a **new OD visualization method**.
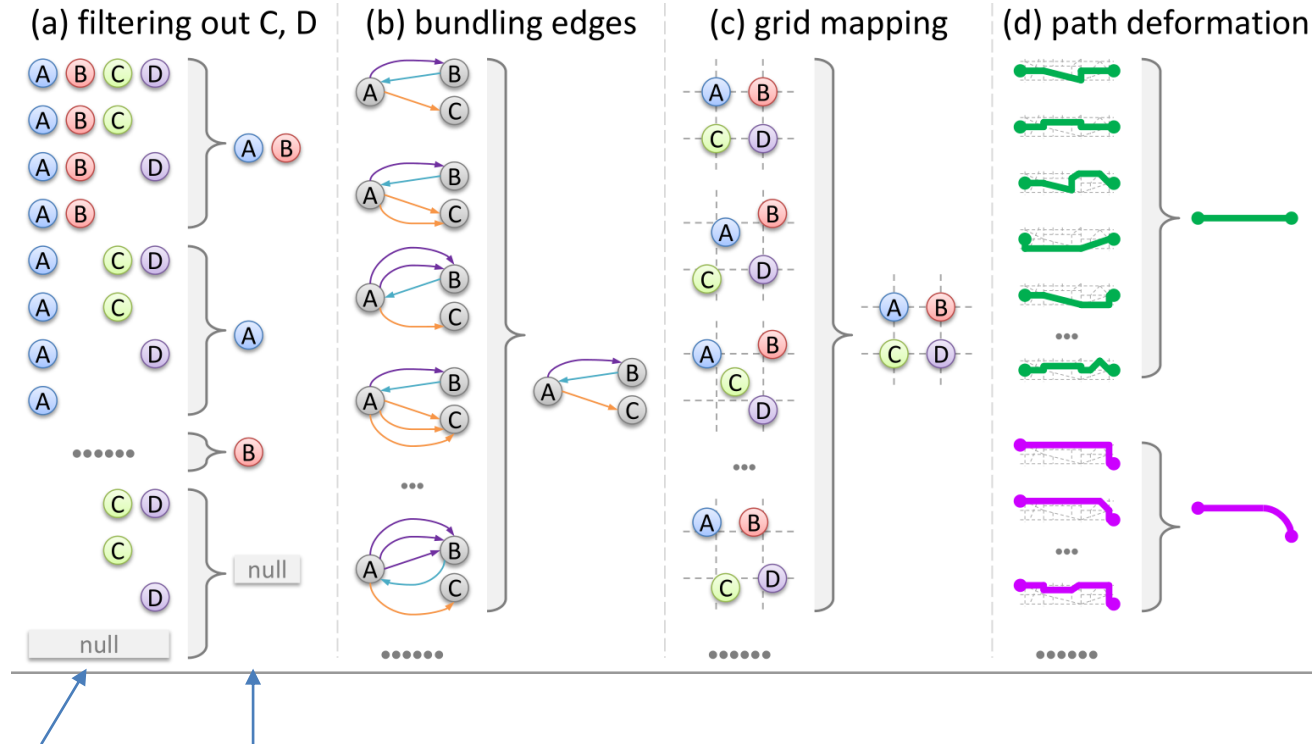

8
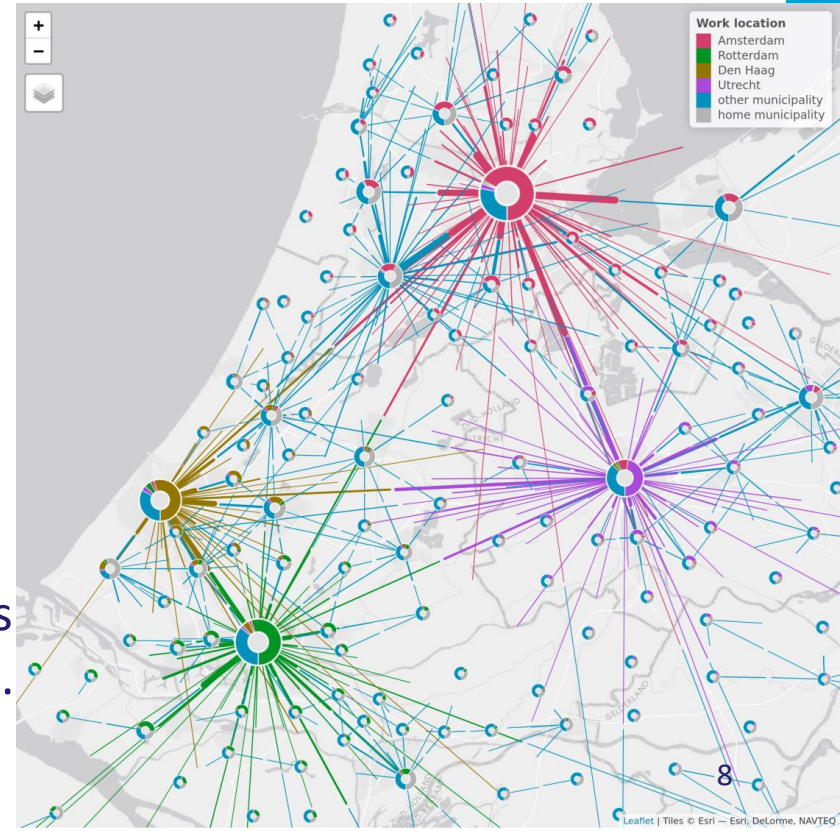
Information Theory

- **Information theory** studies the quantification, storage and communication of information.
- It is (almost) impossible to retrieve the original dataset from a data visualization.
- However, visualization is often much more effective than showing raw data tables. Why?
  - Global overview is often more important than local details.
  - Showing less information helps reducing the time cost and cognitive load.
- **Information loss** (entropy reduction) is key in data visualization. Central questions are: what information can be omitted and how?

Reducing information

(a) filtering out C, D   (b) bundling edges   (c) grid mapping   (d) path deformation

original data
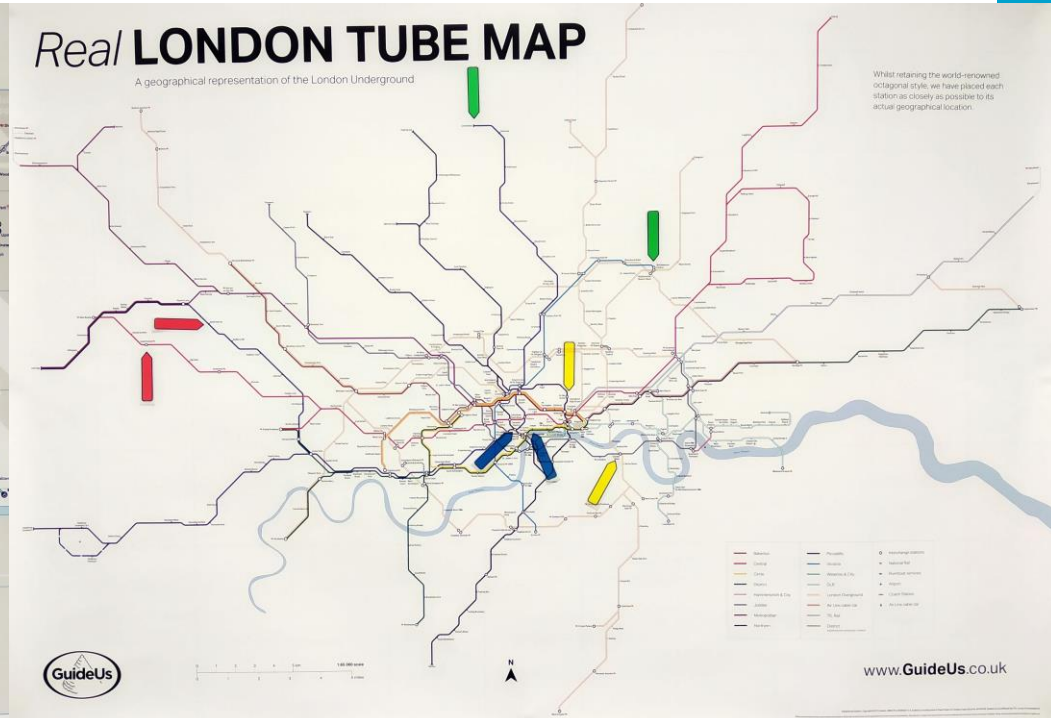(full information)

processed data
(reduced information)

Is realistic always better?
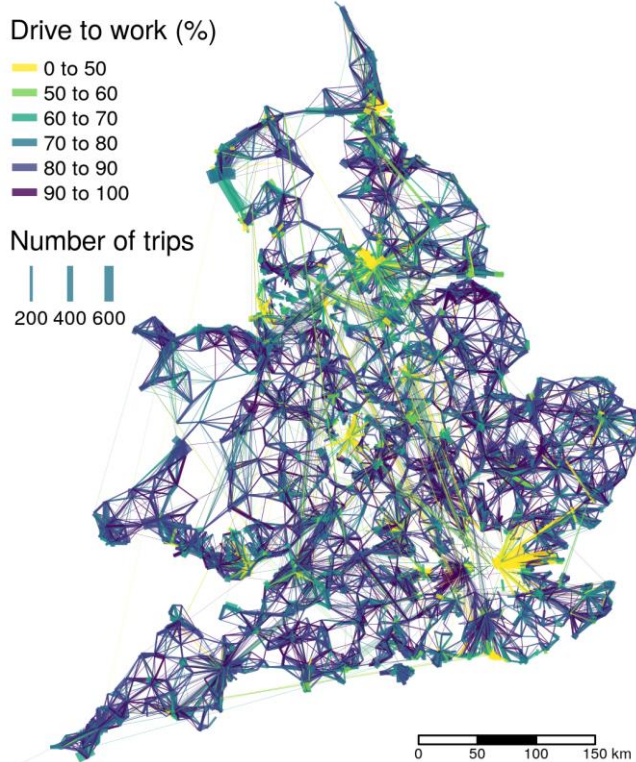

London Underground Map   Realistic locations

Background knowledge matters

Without knowing the locations of the cities, it is hard to use these maps

12

Our design space

- An OD dataset contains of **nodes** and **edges**.
- Our design space consists of four dimensions:
  1. Transformation of the node set
  2. Transformation of the edge set
  3. Transformation of individual nodes
  4. Transformation of individual edges
- Dimensions 1 and 2 process the data structure (e.g. grouping nodes).
- Dimensions 3 and 4 process visual attributes (e.g. coordinates).
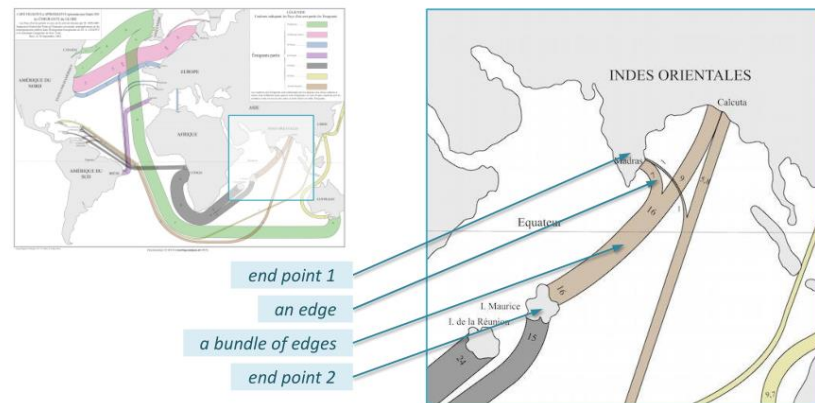
Applied to exisiting OD visualizations
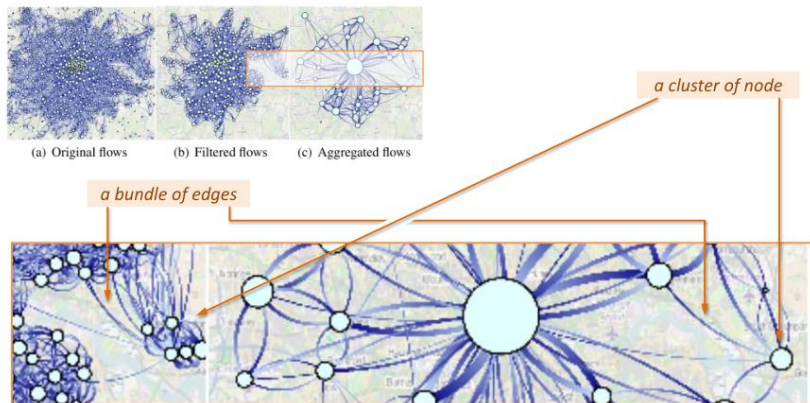
(a) MobilityGraphs
F_1: Group
F_2: Group and filter
F_3: Dimension Enhancement and Attenuation of (x, y)
F_4: Dimension Attenuation w.r.t. ordering,
dimension Attenuation w.r.t. direct path and length

(b) Flow map by Minard (1862)
F_1: Group (same coordinates)
F_2: Group (same nodes)
F_3: Geometric Deformation
F_4: Dimension Enhancement,
dimension Attenuation w.r.t. direct path and length
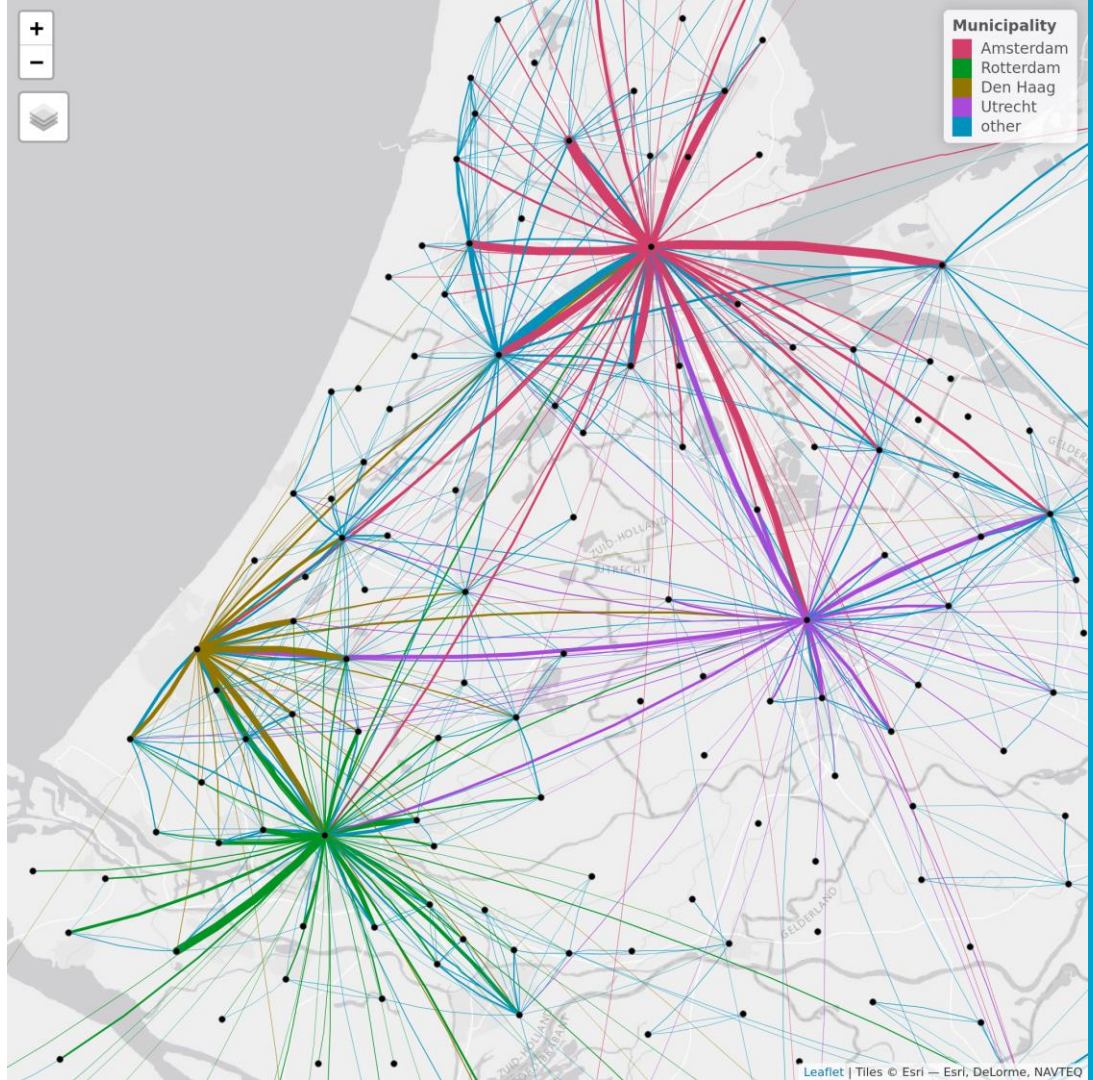
Dutch commuting

Transformations:

- F1: nodes are grouped by municipality

- F2: edges are grouped by same pair of nodes

- F4: missing arrowheads: dimension attenuation

Dutch commuting

Transformations:

- F2: filtering of edges

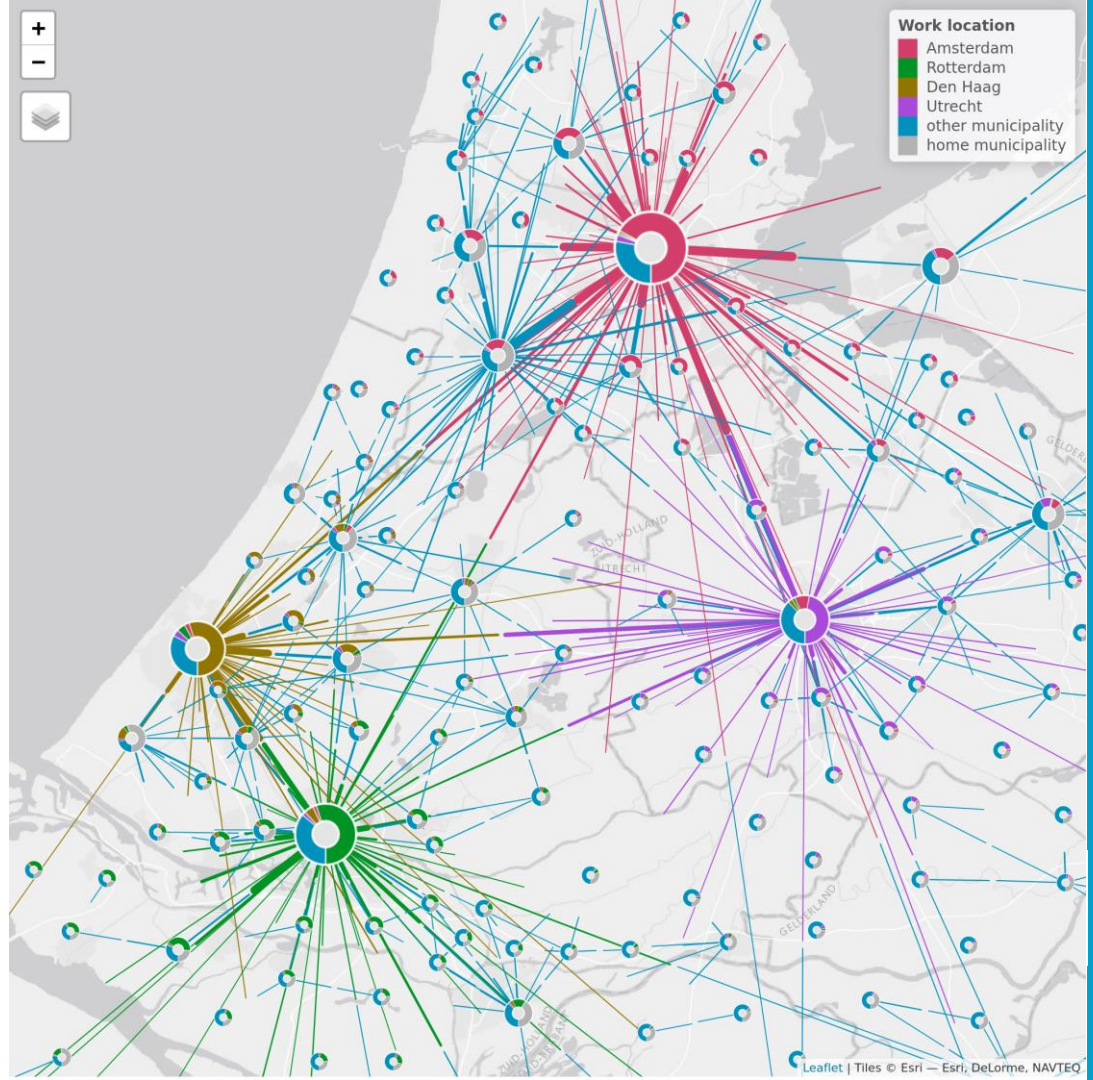- F4: color coding:
  dimension enhancement

Dutch commuting

Transformations:

- F3: doughnuts: dimension enhancement

- F4: half-edges: dimension attenuation (length and direct path)

Link prototype:

http://www.mtennekes.nl/viz/commutingNL

Discussion

- Information Theory has proven to be useful in visualization.
  - "Less is more": reducing information often leads to better visualizations.
  - User and task dependency.
- A design space for OD visualizations enforces users to think about which information to loose and which to show.
- Doughnut map with halfway lines works well for commuting data, but not necessarily for other OD datasets.

Acknowledgements

Dolores Romero Morales
Lili Nottrott

Sofie de Broe
Marko Roos
Chantal Melser

Min Chen
Yu Zhang
Jacob Wilkins