# MOBILE PHONE ACTIVITY MILAN, ITALY

Peyman Hesami
DSE241 Final Project

MOTIVATIONS

- Mobile phone activities generate massive amount of data

- This can be used in mobility planning, tourist flows, urban structures and interactions, event detection, urban well-being and many others

Development of average data volume per residential customer (postpaid) in the mobile phone network
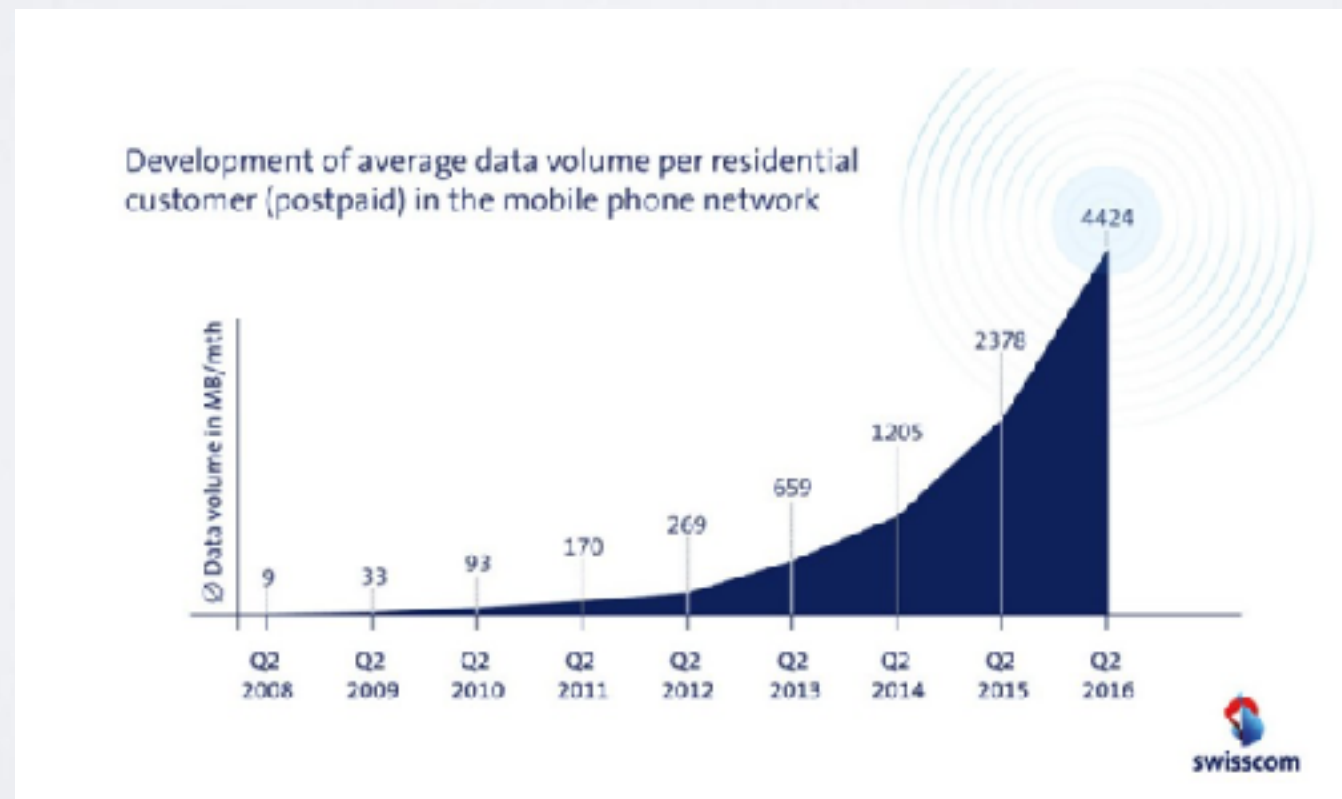
MOTIVATIONS

- *It can also be used for cellular network diagnostics and maintenance*

  - *Finding congested cells/areas*

  - *Finding idle cells/areas*

  - *Finding user's usage pattern*

DATASET

- One week of Call Details Records (CDRs) from the city of Milan and the Province of Trentino (Italy)- 1.5 GB

- Both domestic (Milan to other provinces) and international (Milan to other countries) data

- Third source of data:

    - (lat, long) of countries and provinces of Italy

    - Geojson file of Milan cellular network containing (lat, long) of cells in the city of Milan

datetime	CellID	countrycode	smsin	smsout	callin	callout	internet
2013-11-01 00:00:00	1	0	0.3521			0.0273	
2013-11-01 00:00:00	1	33					0.0261
2013-11-01 00:00:00	1	39	1.7322	1.1047	0.5919	0.402	57.7729
2013-11-01 00:00:00	2	0	0.3581			0.0273	

datetime	CellID	provinceName	cell2Province	Province2cell
2013-11-01 00:00:00	1	MILANO	0.1894	0.0541
2013-11-01 00:00:00	1	PAVIA	0.0273	
2013-11-01 00:00:00	1	TRENTO	0.0261	
2013-11-01 00:00:00	2	MILANO	0.1922	0.0556
2013-11-01 00:00:00	2	PAVIA	0.0273	

DATA WRANGLING

- Converting the raw dataset to two sets of nodes/edges dataset

- Adding label to the nodes based on their type (Milan, domestic, international)

- Removing edges with no 0 values (for sms, call, …)

- Extracting day and hour from date time

- Integrating third source of data:

  - Converting country codes to country names

  - Deriving the coordinates (lat and long) of the cells in Milan, provinces of Italy, and other countries

TASKS

- Goals:

  1. **Diagnostic tool:** Visualization tool to help wireless network engineers in cellular networks diagnostics

  2. **Presentation tool:** User friendly representation of the mobile phone activities for nontechnical presentation

TASKS-DIAGNOSTIC TOOL

- **A node-link diagram with:**

  - Nodes as cells and edges as user activities (width~magnitude channel)

  - Several filters to choose the type of the data; domestic vs international, sms vs call vs Internet

  - Zoom in/out capability

  - Interaction with the graph by selecting nodes and highlighting their connected edge, adding labels to nodes/edges

  - Adding time sliders to choose the time interval within a day, day of the week and animation across time

  - Ability to choose the desired graph layout

  - Ability to show only the most significant edges based on a user input

  - Ability to hide nodes and edges on drag for easy interactions

AUDIENCE-DIAGNOSTIC TOOL

- Cellular network engineers:

    - Identifying the troubled (congested) cells

    - Identifying the idle (inactive) cells

        - Optimize them to serve more users

    - Identifying the cell use pattern across days of the weeks and hours of the day

        - Schedule maintenance time in low traffic time intervals.

    - Identifying the data usage patterns across time and geography

        - Optimize the cells dynamically based on usage
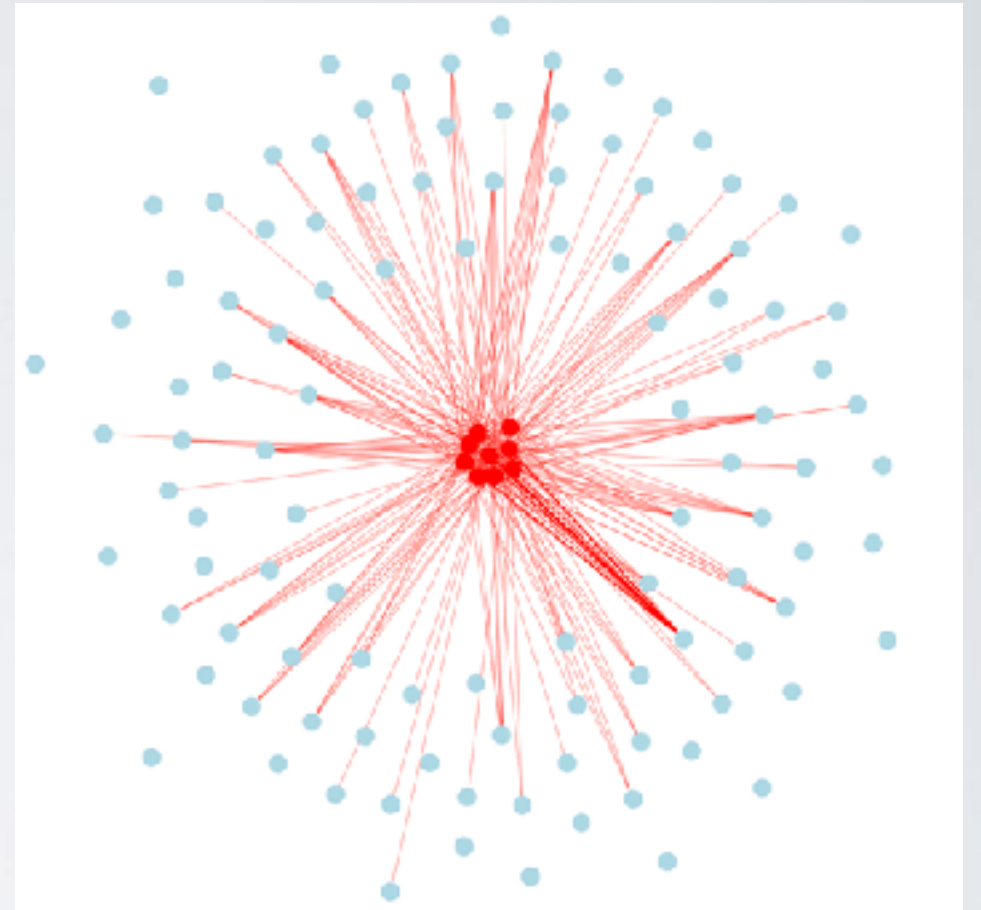
TASKS-PRESENTATION TOOL

- **Great Circle on geo layout with:**

  - points as cells and lines as user activities

  - Several filters to choose the type of the data; domestic vs international, sms vs call vs Internet

  - Adding time sliders to choose the time interval within a day, day of the week and animation across time

  - Ability to show only the most significant edges based on a user input

Great Circles Representation of the International Phone Activity in Milan

AUDIENCE-PRESENTATION TOOL

- Nontechnical users seeking:

  - To find mobile phone user's usage patterns across time and geography.

  - Study the usage pattern alongside other sources of the data (like census data) for socio-technical analysis (like targeted marketing)

SOLUTIONS

- Visualization type

- Data reduction (filtering)

- Data reduction (sampling)

Choose the type of the analysis to display

node-link diagram	▲

node-link diagram

great circles (geo)

Choose the Type of the Data:

- ● Domestic
- ○ International

Sample Size:

| 0.01 | 0.43 | 1 |

0.01 0.11 0.21 0.31 0.41 0.51 0.61 0.71 0.81 0.91 1

SOLUTIONS

- Data Reduction (edge filtering)

Choose the type of the domestic data (edge data):

☑ Incoming Calls

☐ Outgoing Calls

Choose the type of the international data (edge data):

☑ Incoming SMS

☐ Outgoing SMS

☐ Incoming Calls

☐ Outgoing Calls

☐ Internet Connections

Choose the day

☐ Monday

☑ Tuesday

☐ Wednesday

☐ Thursday

☐ Friday

☐ Saturday

☐ Sunday

Time Interval (Hour):

1   9 11   24

1   4   7   10   13   16   19   22   24

▶

- View change over time (static and animation)
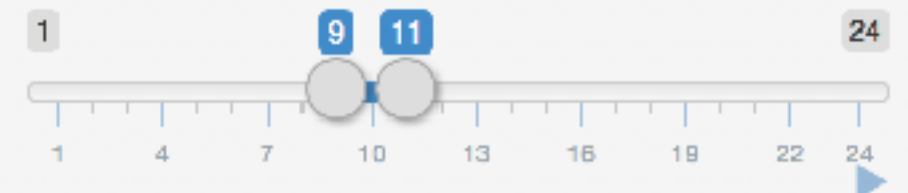
SOLUTIONS

- Data reduction (aggregation)

Cell Aggregation Level:

1 | 1,000 | 10,000

1 1.001 2.001 3.001 4.001 5.001 6.001 7.001 8.001 9.001 10.000

Graph Layout
- ● Force Directed
- ○ Circular
- ○ Spring Forced

- Graph layout

Graph Annotaion
- ☐ Node Label
- ☐ Edge Label
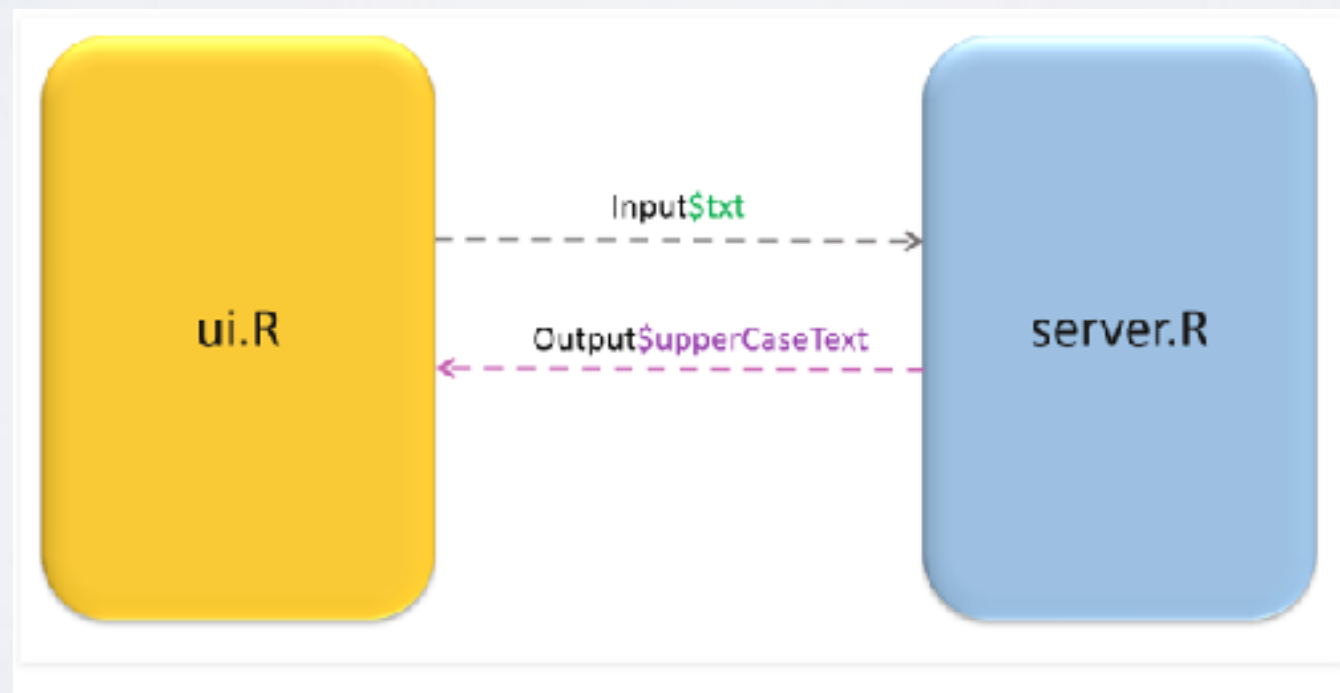- ☐ Highlight Nearest Edge
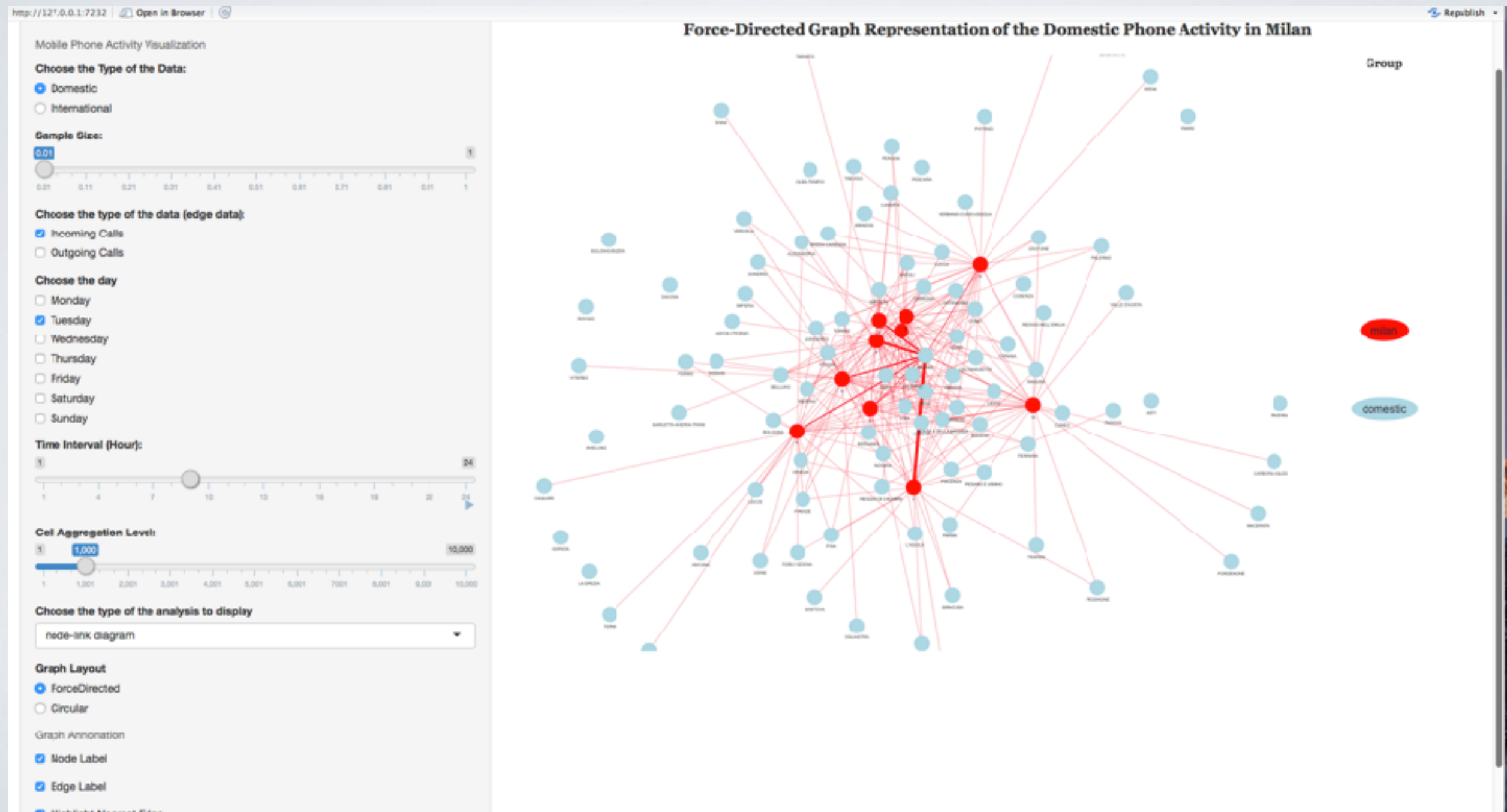- ☐ Hide Nodes on Drag
- ☐ Hide Edges on Drag

- Graph annotations/features

- Data reduction (edge filtering)

Edge Width Threshold (percentile)

0 | 100

0 10 20 30 40 50 60 70 80 90 100
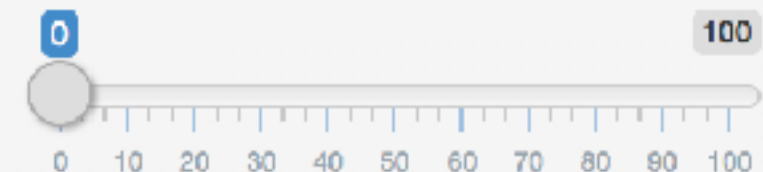
IMPLEMENTATION

- Shiny package in R

- Run ui.R or server.R in RStudio

- Deployed on shiny.io serve: https://peymanshiny.shinyapps.io/
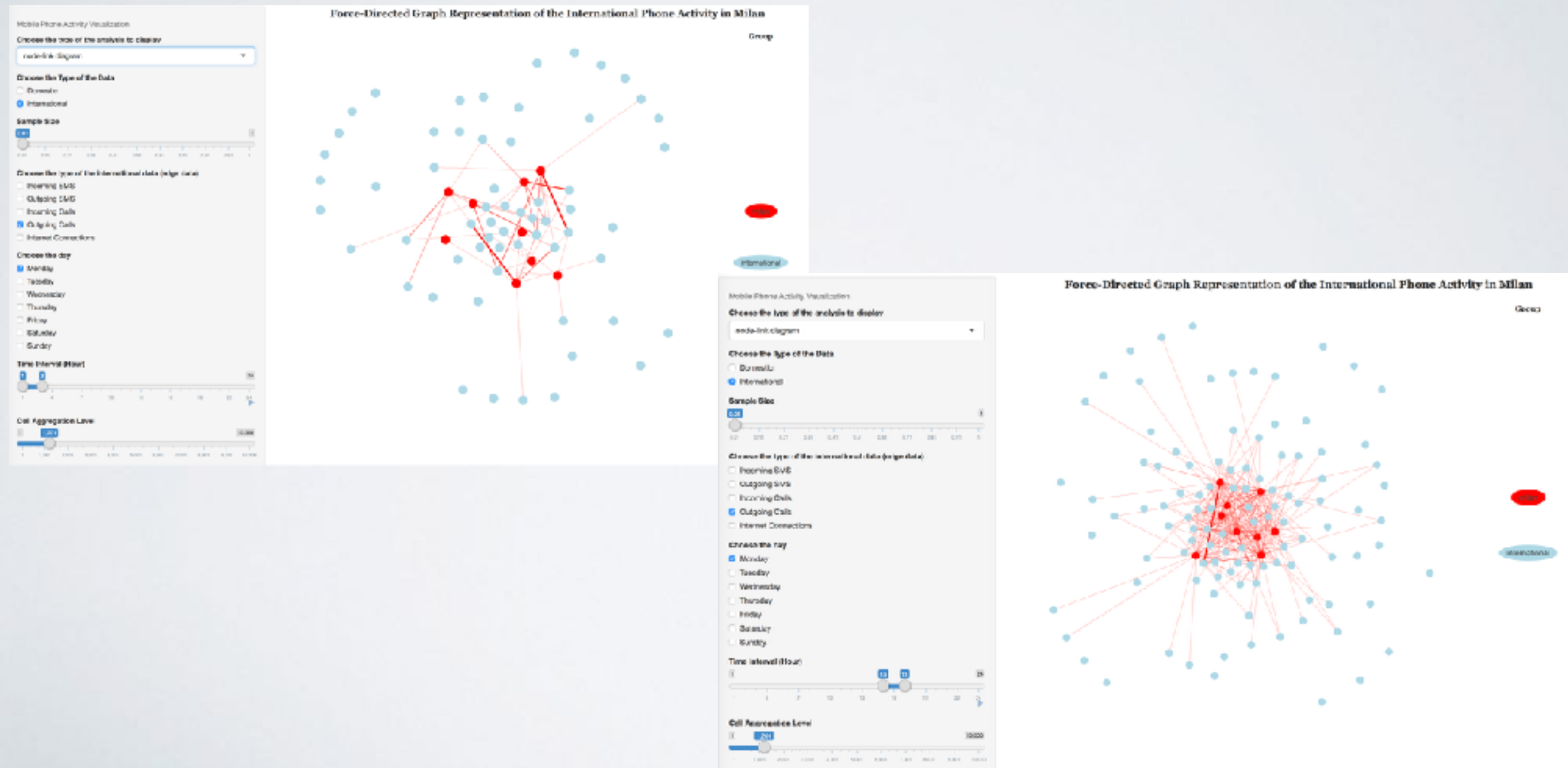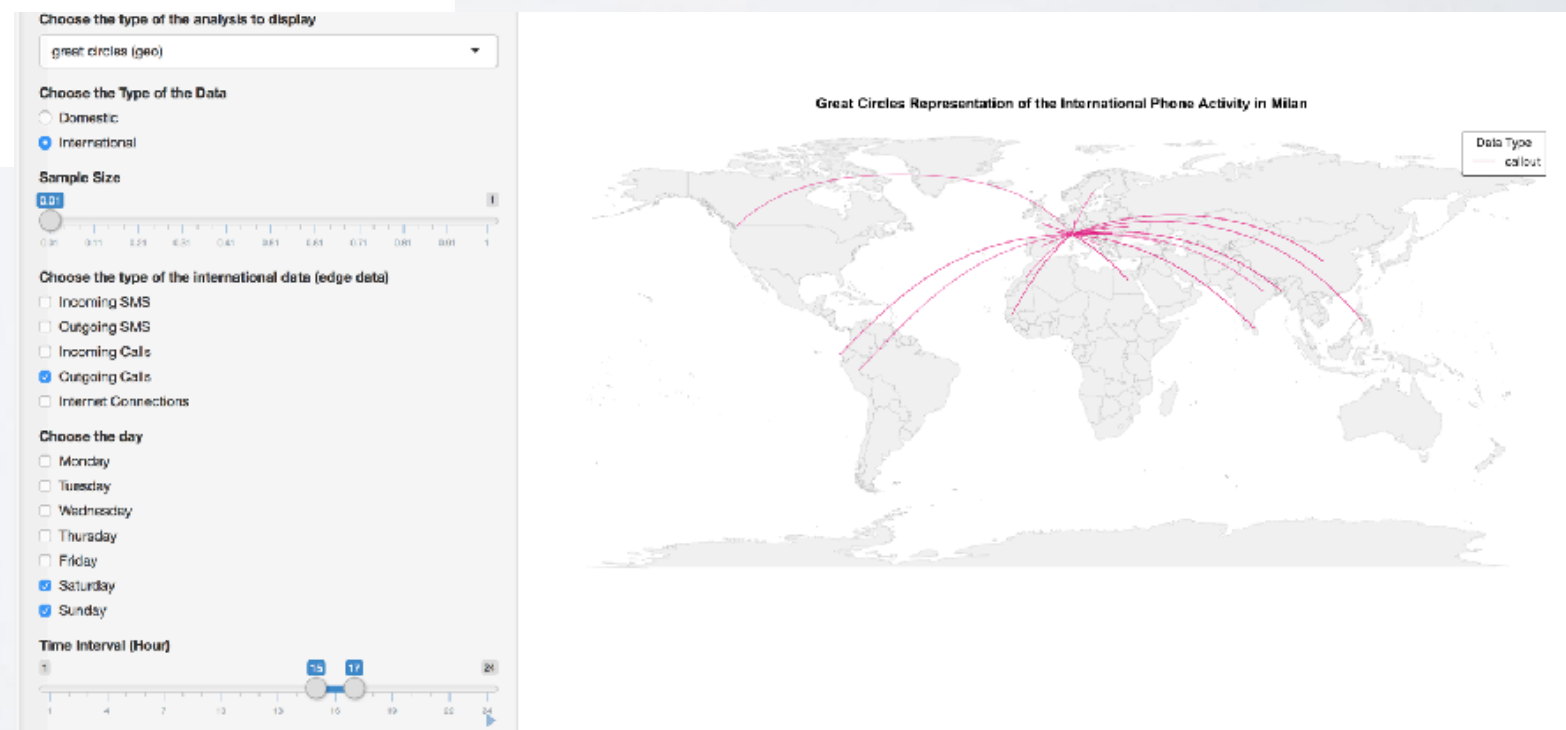 milan_phone_activity_shiny_dse241_peyman_hesami/
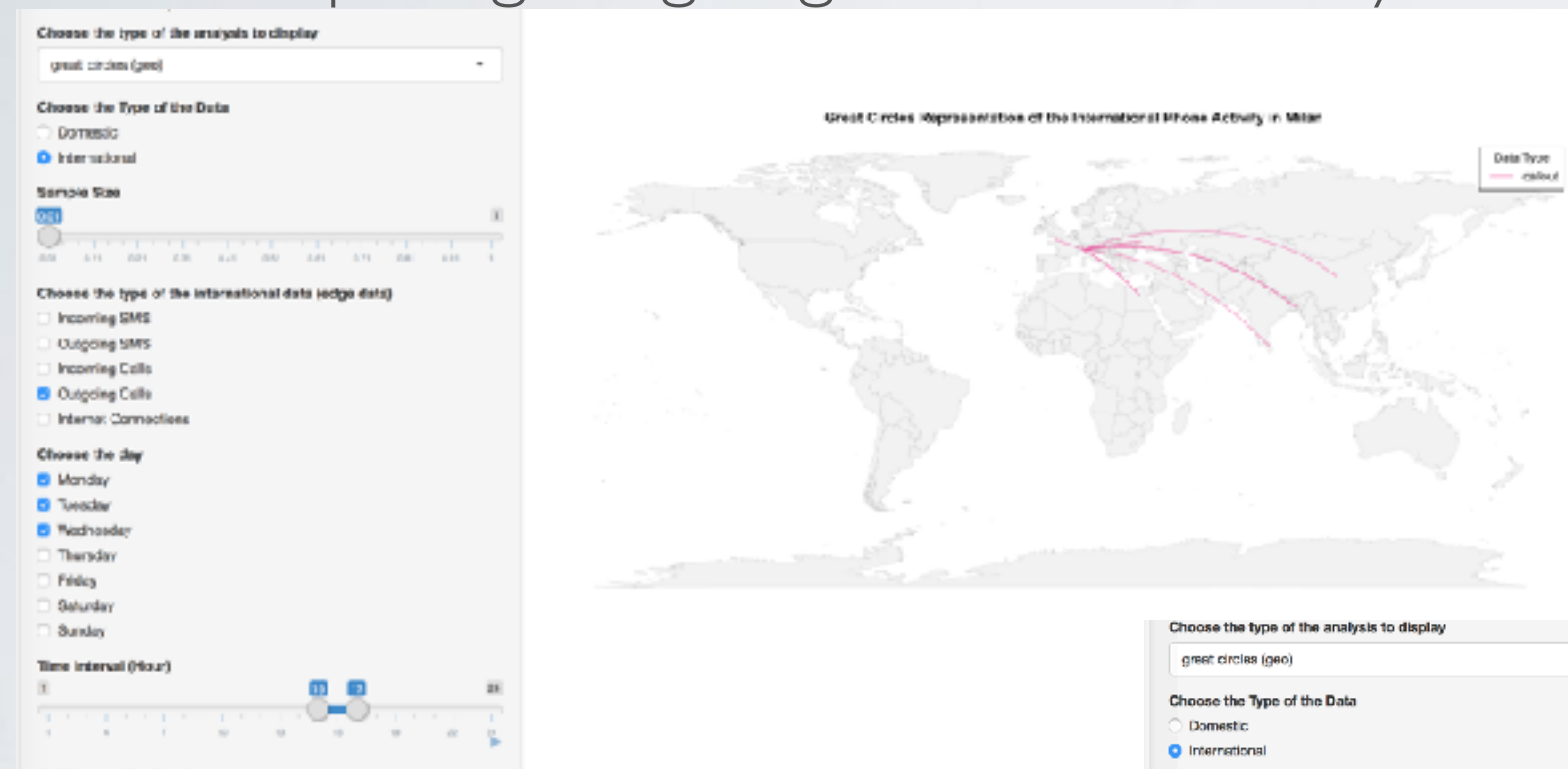
RESULTS

RESULTS

- Comparing network at 1am and 3pm on a Monday

RESULTS

- Comparing outgoing calls on weekdays versus weekend at 3pm

CHALLENGES AND IMPROVEMENTS

- Efficient reading data into memory is required for fast user interactions (multiple libraries tried)

- Hour/Day extraction can be costly (regex)

- Great circle vis is not completely interactive

- Other sources of data (census) can be integrated for more insights