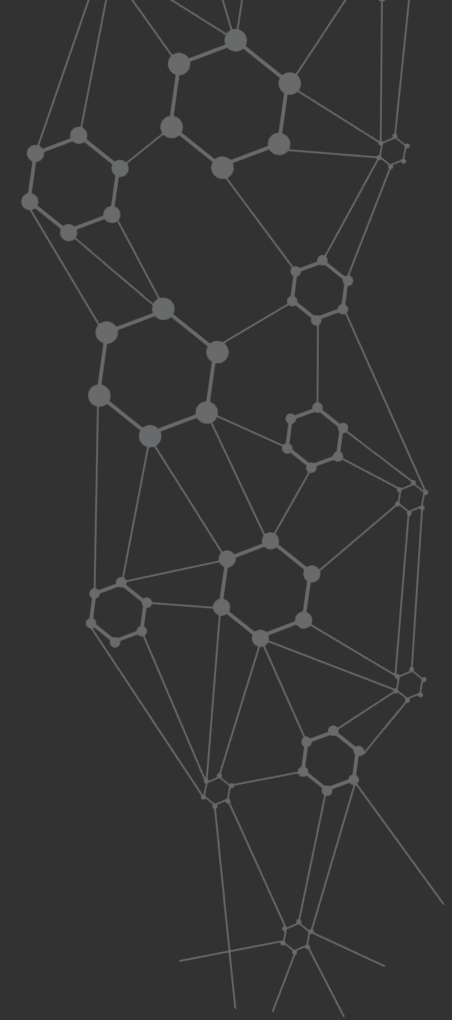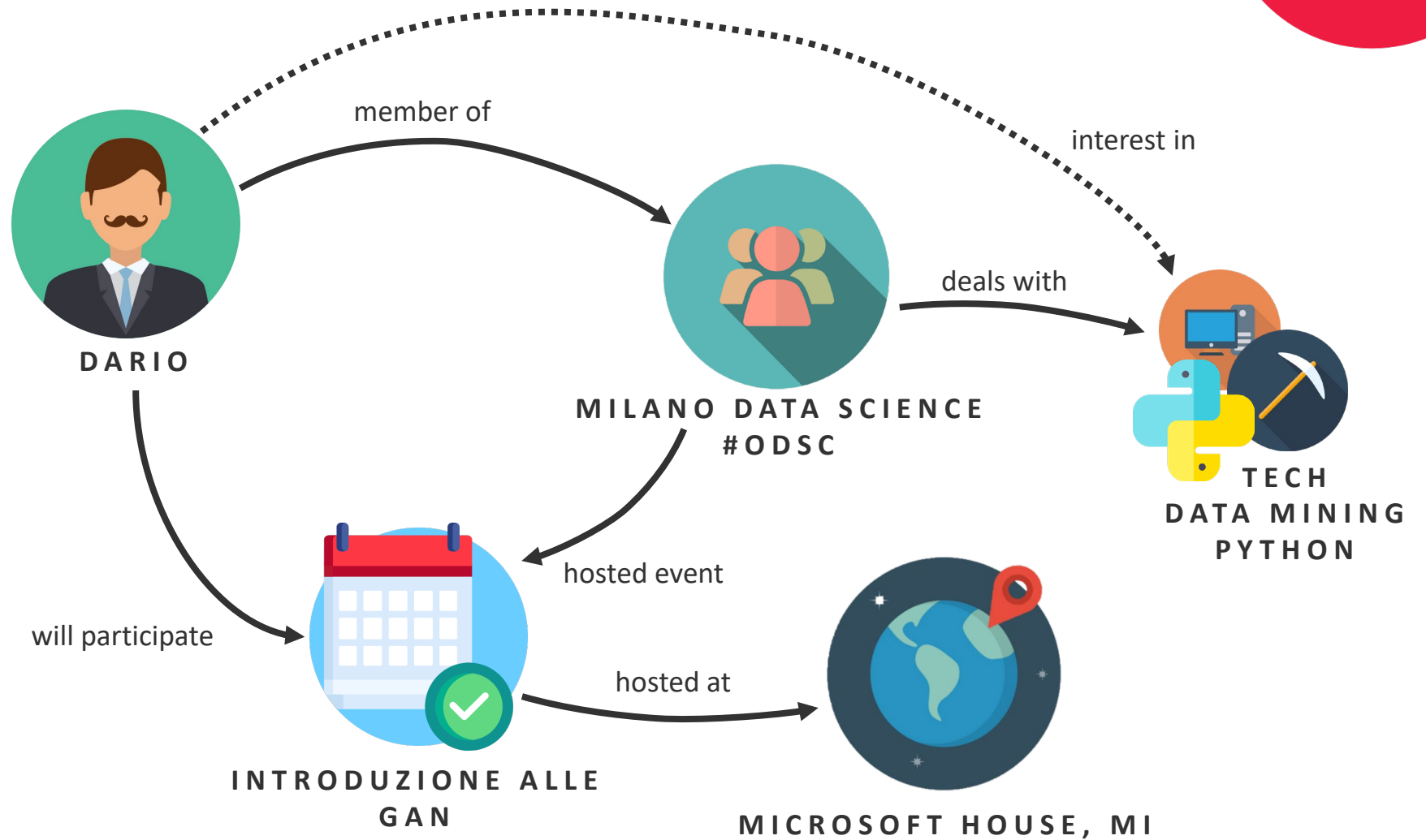# social network analysis

Dario Bertazioli

Fabrizio D'Intinosante

Massimiliano Perletti

introduction

available in **186** countries
40 millions users
320k active groups
12k daily events

member of

interest in

DARIO

MILANO DATA SCIENCE #ODSC

deals with

hosted event

TECH DATA MINING PYTHON

will participate

hosted at

INTRODUZIONE ALLE GAN

MICROSOFT HOUSE, MI

2

goals

quantitative measures

events temporal distribution

events influence area

recommender system efficiency analysis
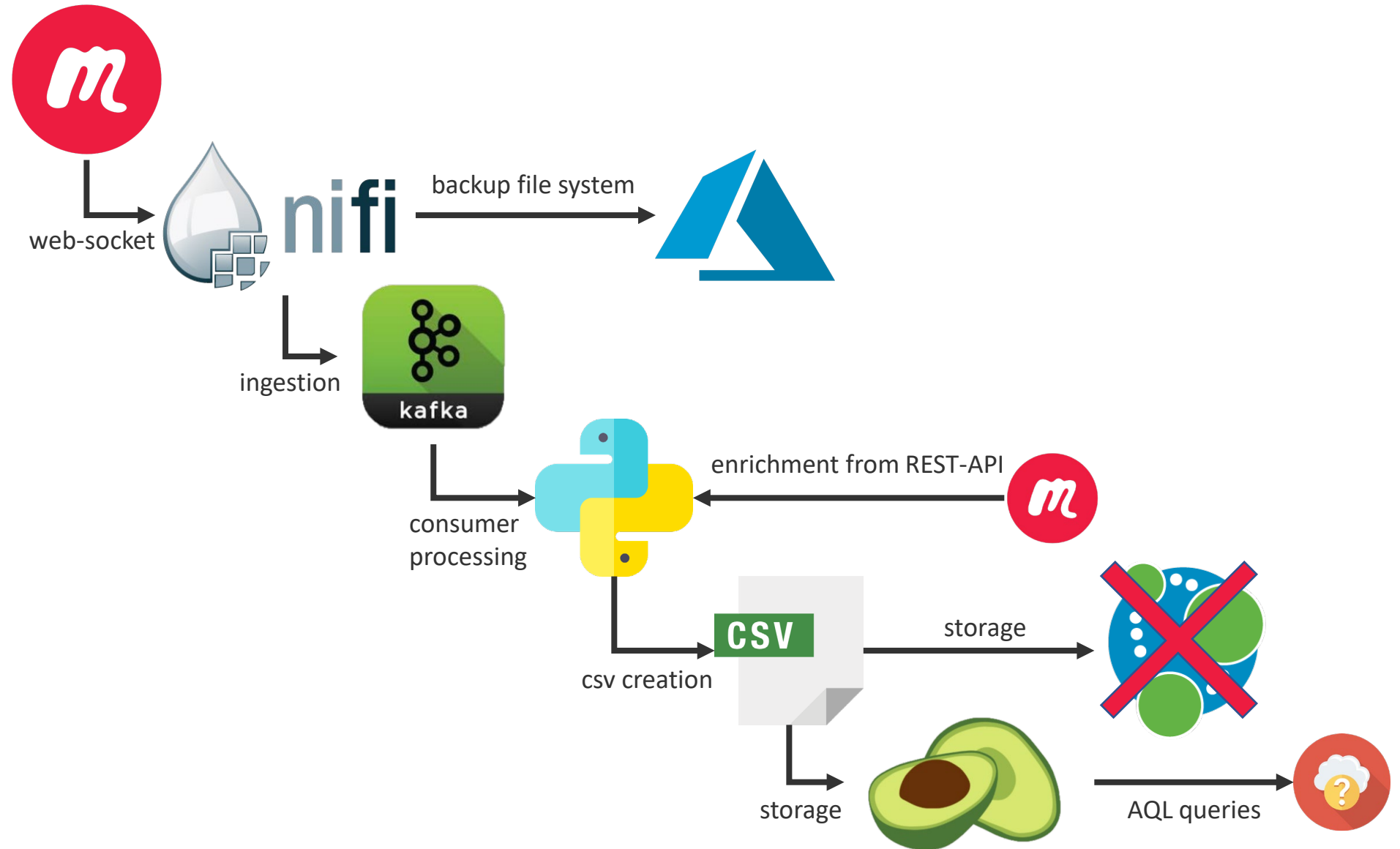
schema

covered points

from Big Data V(s)

volume
more than 2 GB database

velocity
streaming data acquisition

web-socket

backup file system

ingestion

kafka

consumer processing

enrichment from REST-API

csv creation
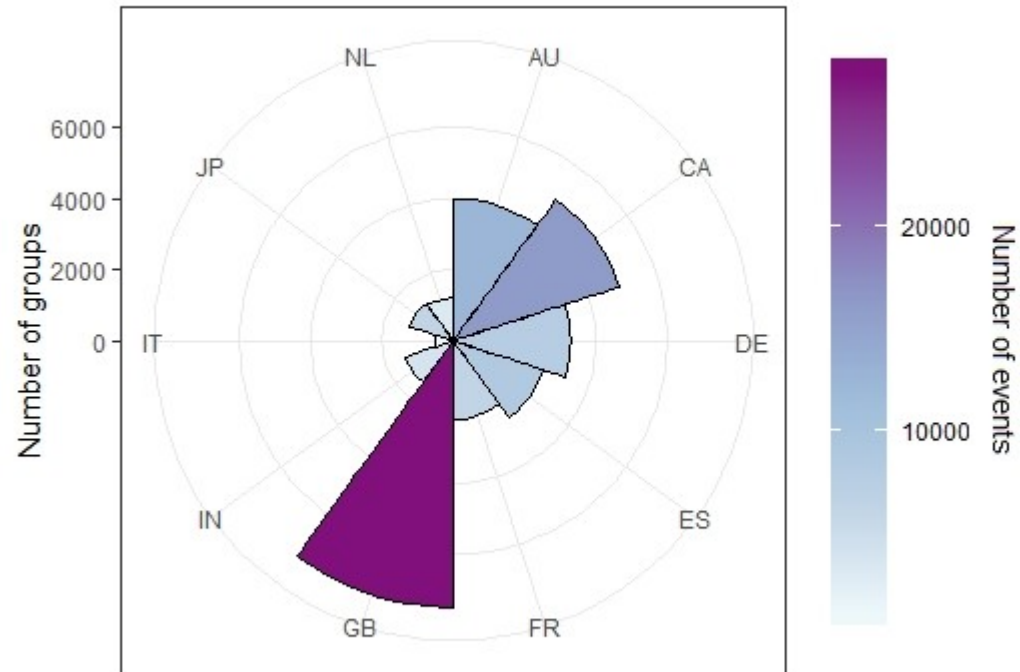
CSV

storage

storage

AQL queries

5

results

quantitative measures

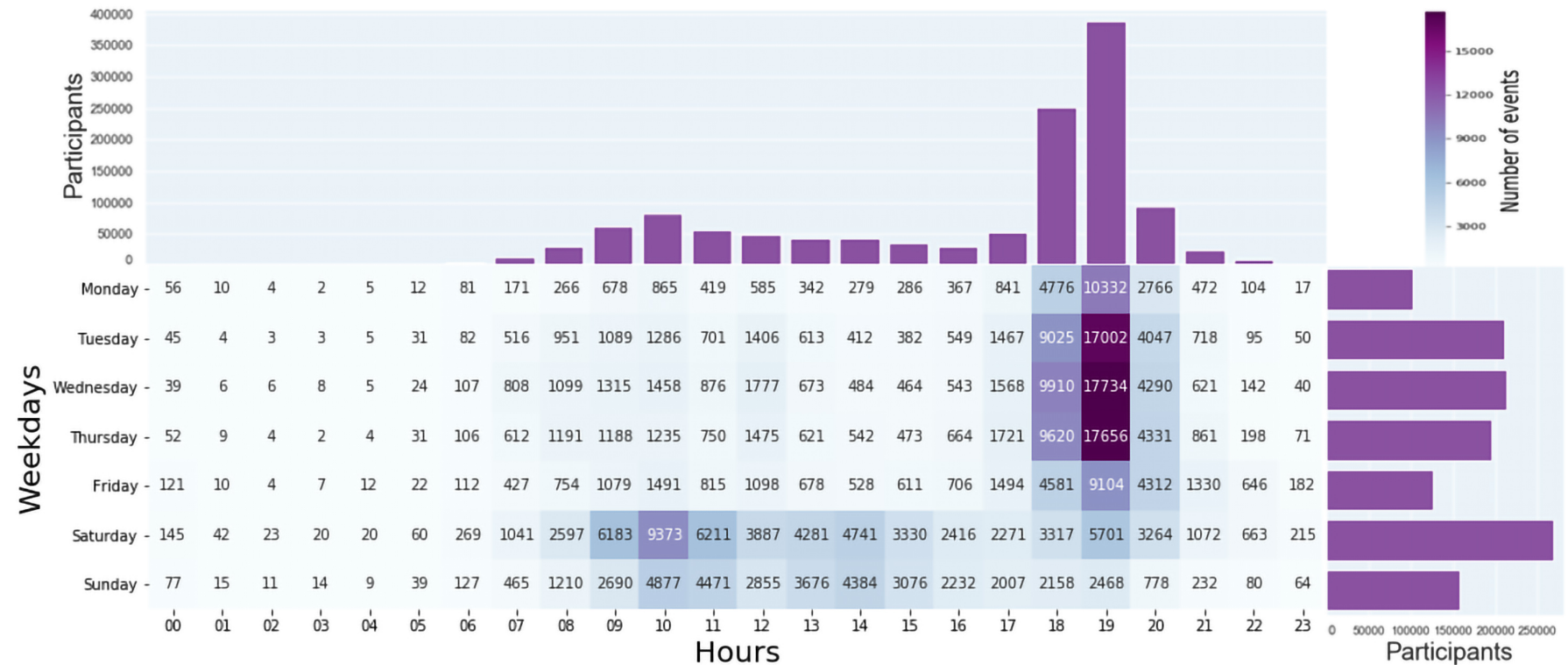Rose plot of quantity of groups and events for country
USA are not shown because completely out of scale

- quantity of events by Country

- quantity of groups by Country

- maximum number of participants by Country

- average number of guests by participant

- trend topic for users
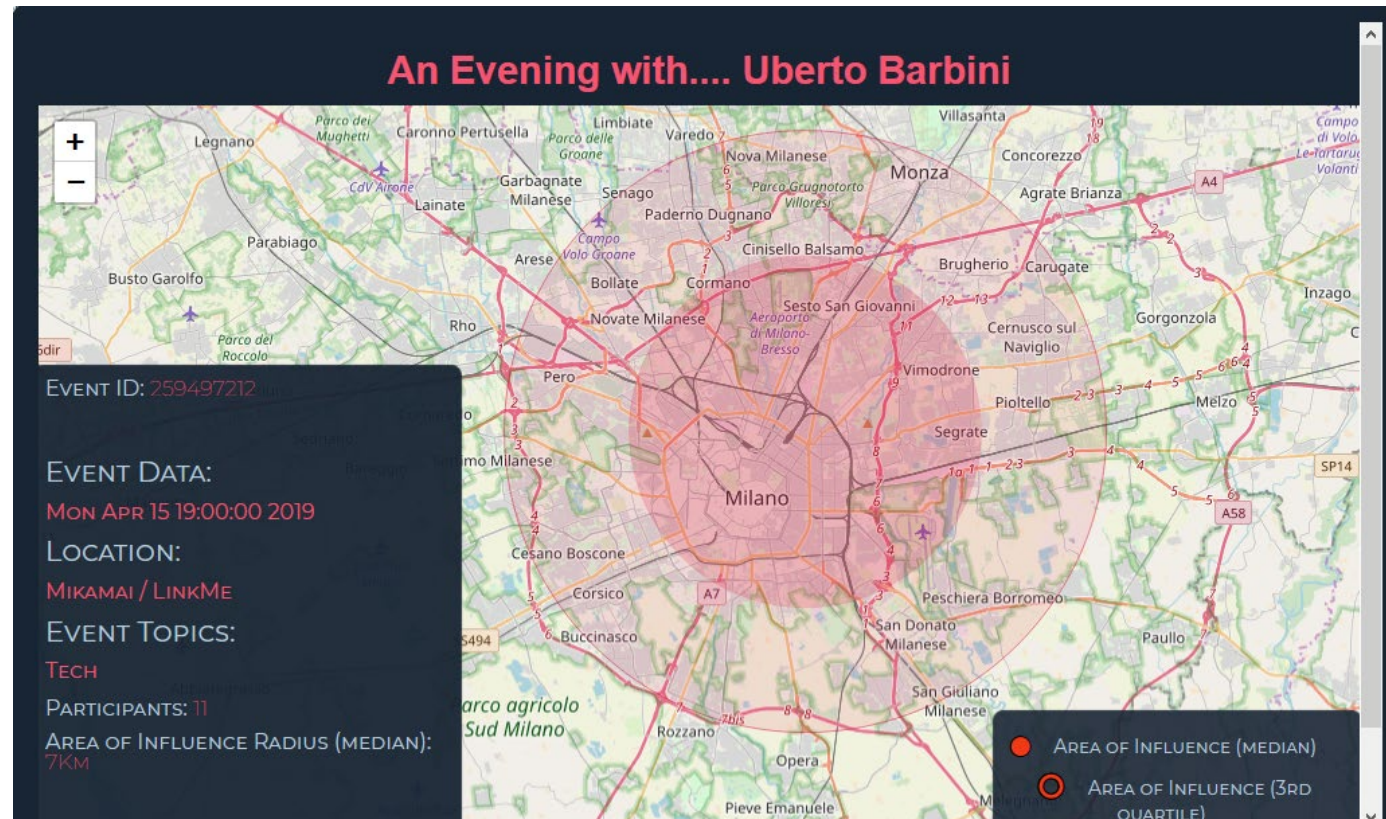
- trend topic for groups

results

events temporal distribution

found the best moment in a day and in a week (worldwide) to organize a meetup
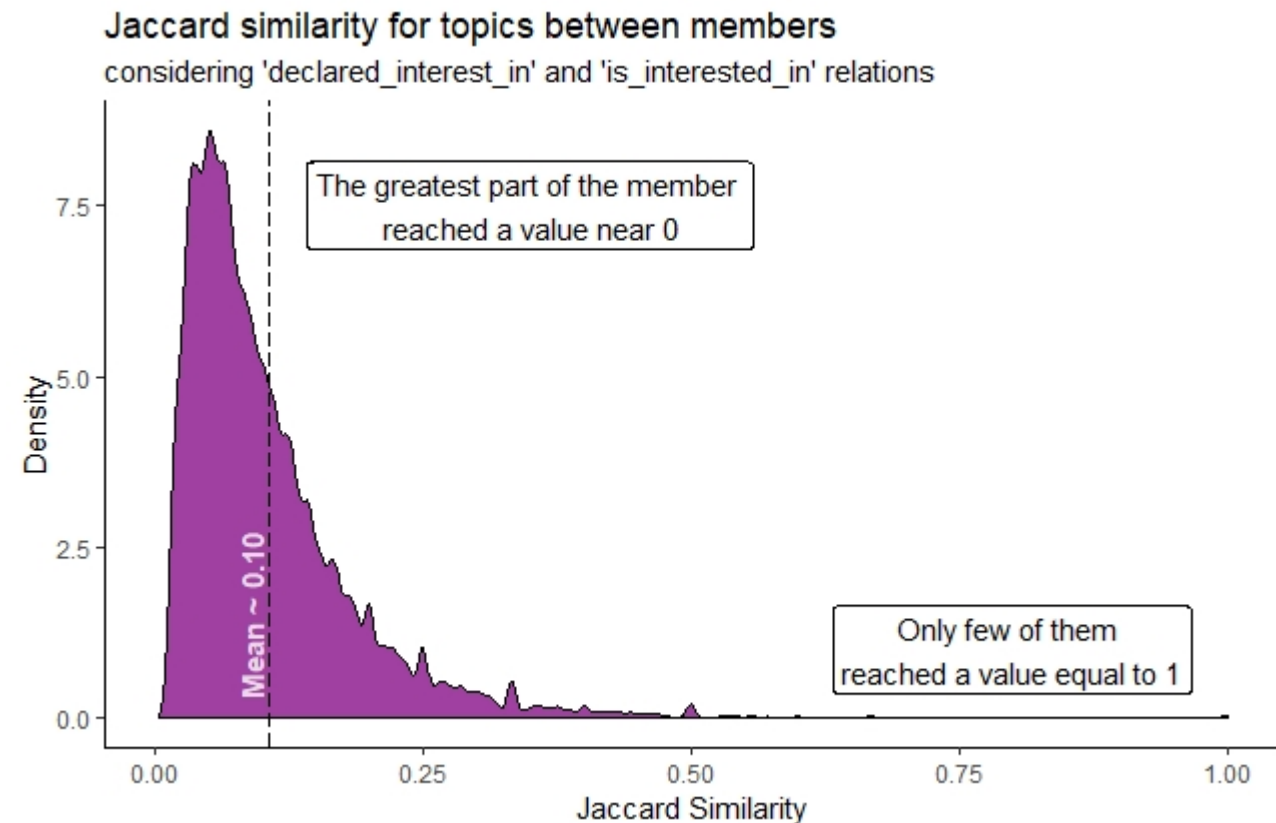
results

events influence area

built an interactive map that display some informations about events including average radius of attraction for users

results

using Jaccard similarity we evaluated the recommender system efficiency exploiting the matches between the group topics and the user topics

recommender system efficiency analysis

Jaccard similarity for topics between members
considering 'declared_interest_in' and 'is_interested_in' relations

The greatest part of the member reached a value near 0

Mean ~ 0.10

Only few of them reached a value equal to 1

interesting challenges

01 | symbolic links

02 | times optimization in messages extraction

03 | import Cypher vs Py2Neo

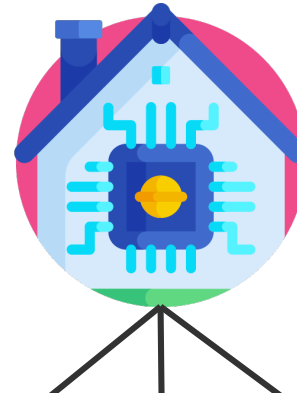04 | import Cypher vs Arangoimp

05 | scalability

06 | temporal metadata conversion

07 | live streaming ingestion
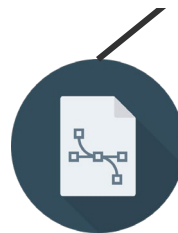
conclusions

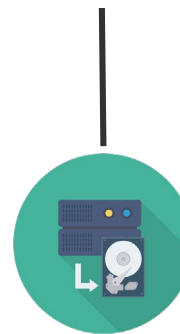answered research
questions successfully

software structure benefits from
these features

improvement of
personal skills

scalability

fault tolerance

consistency

THANK YOU