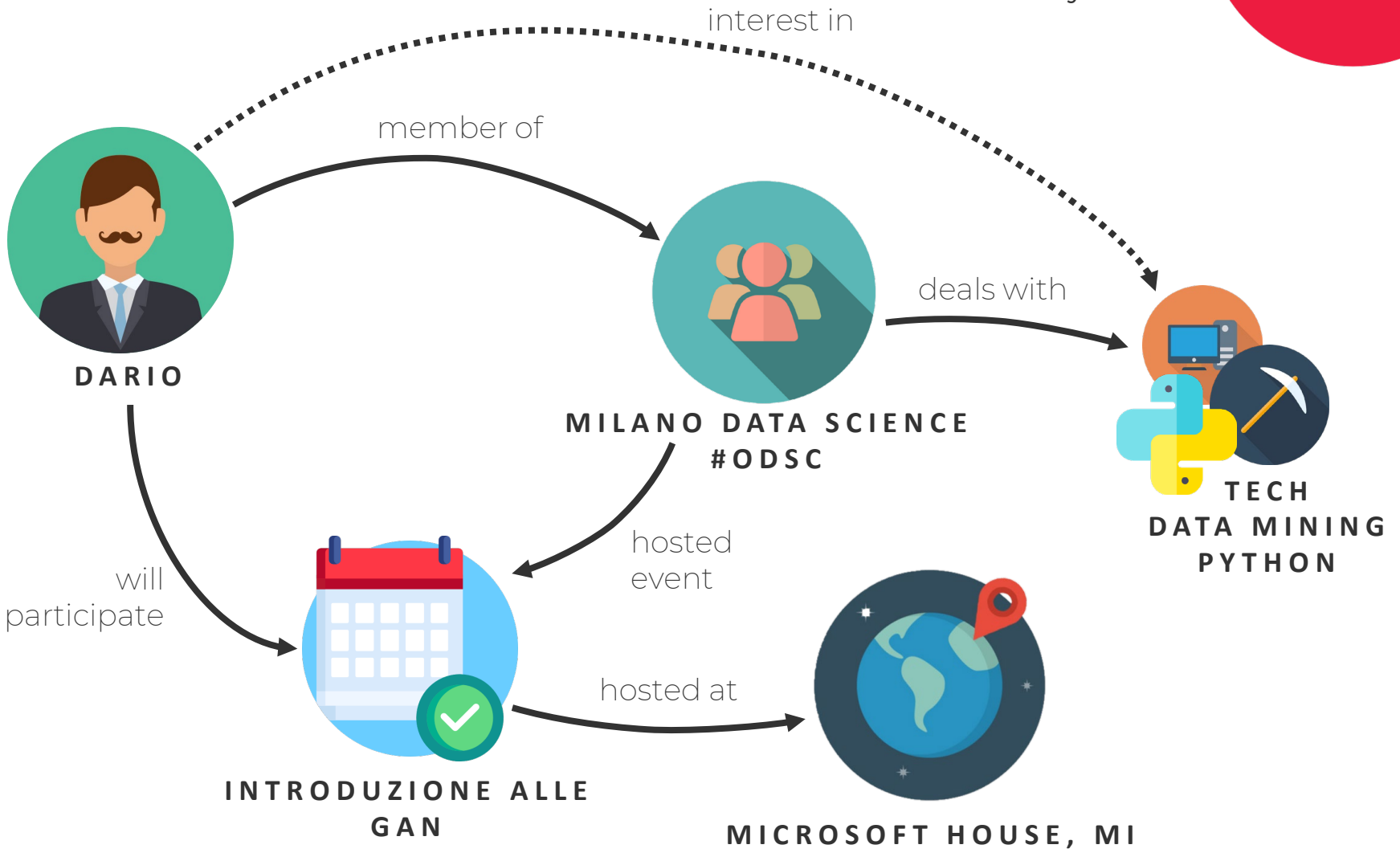
UNIVERSITÀ DEGLI STUDI DI MILANO BICOCCA

social network analysis

Dario Bertazioli
Fabrizio D'Intinosante
Massimiliano Perletti

introduction

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



goals

quantitative measures

temporal distribution of the events

event area of influence

recommender system efficiency analysis

covered points

from Big Data V(s)

volume
more than 2
GB database

velocity
streaming data
acquisition

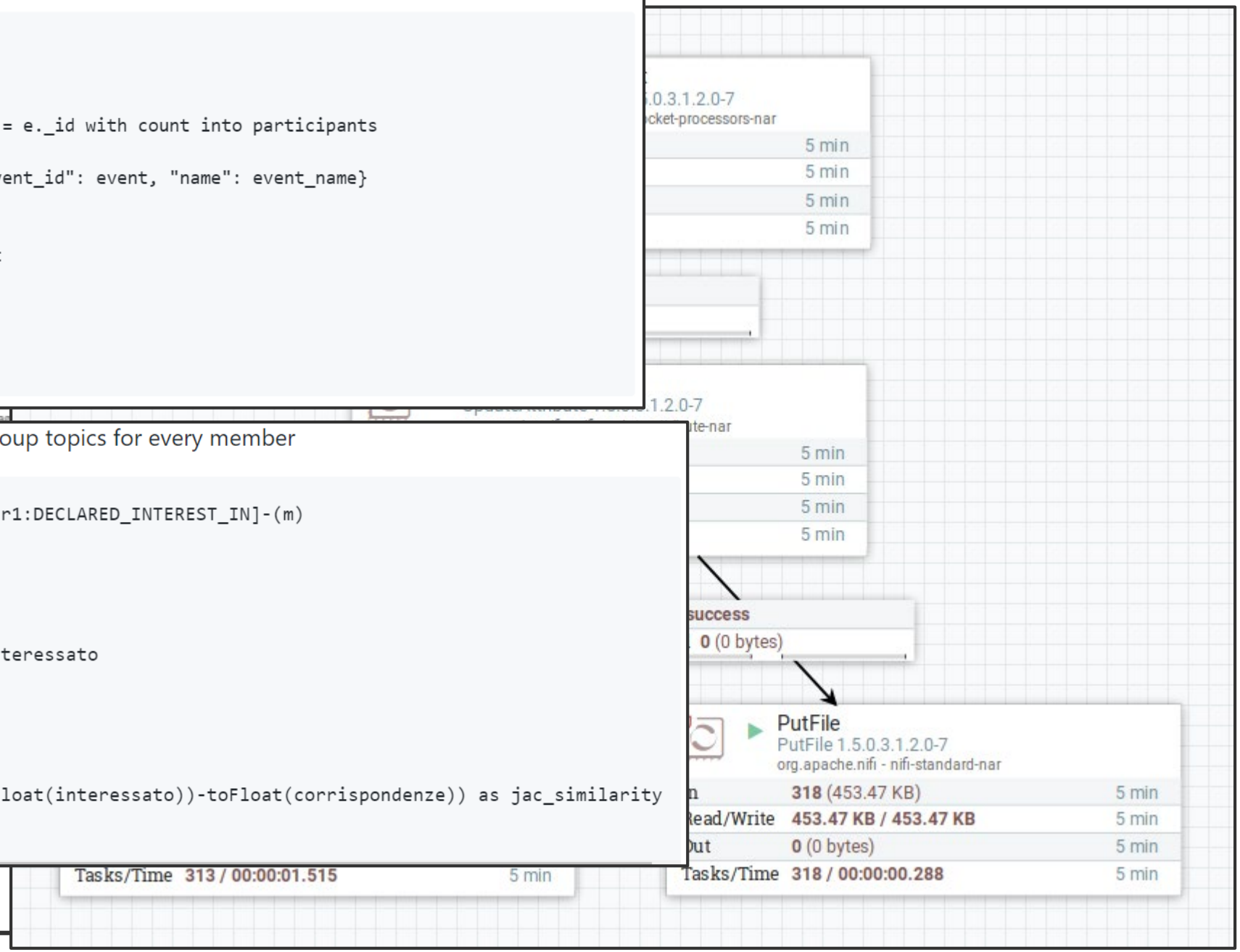
architecture

- Massimo numero di partecipanti ad un evento per Paese

```
let eventsWithPart = (
    for m in member
        for e in 1..1 outbound m will_partecipates
            collect event_name = e.event_name, event = e._id with count into participants
            sort participants desc
            return {"participants": participants, "event_id": event, "name": event_name}
)
for event in eventsWithPart
    for g in 1..1 inbound event.event_id hosted_event
    collect country = g.group_country
    aggregate max_part = max(event.participants)
    sort max_part desc
    return {country, max_part}
```

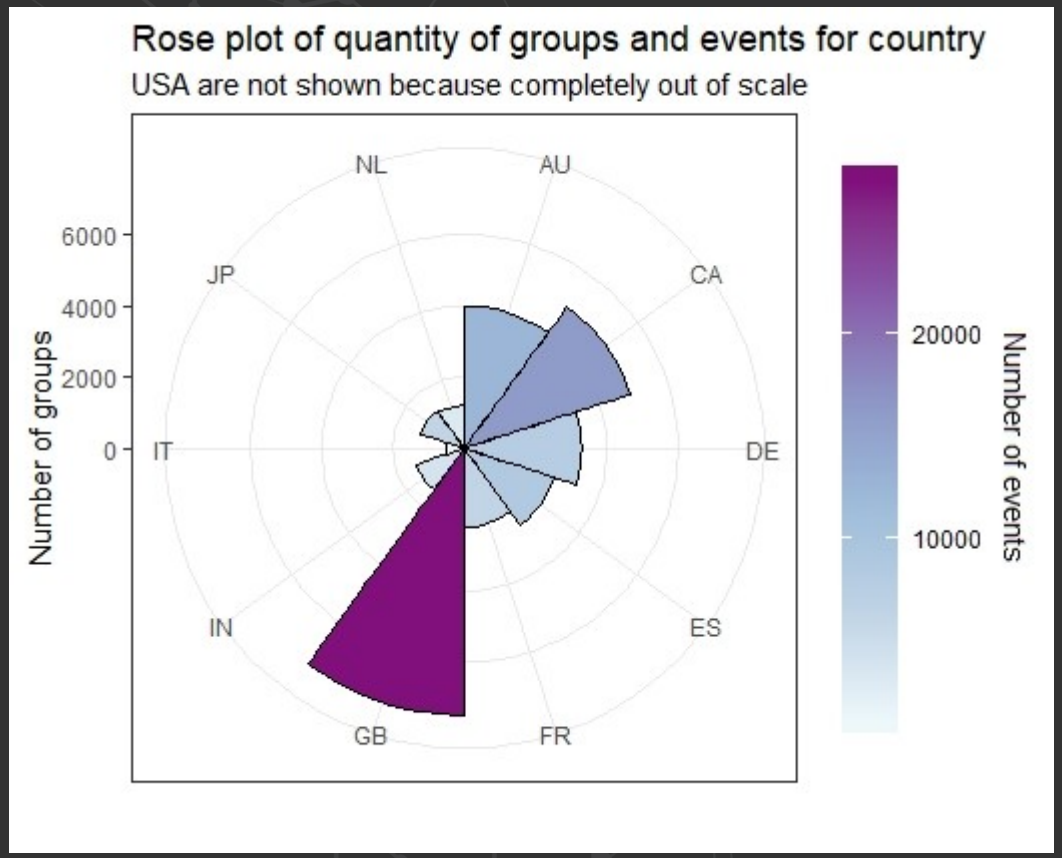
- Jaccard similarity measure between declared and group topics for every member

```
match (m:Member)-[r:IS_INTERESTED_IN]->(t:Topic)<-[r1:DECLARED_INTEREST_IN]-(m)
with count(t) as corrispondenze, m
match (m)-[r:DECLARED_INTEREST_IN]->(t:Topic)
with count(r) as dichiarati, corrispondenze, m
match (m)-[r1:IS_INTERESTED_IN]-(t:Topic)
with m, corrispondenze, dichiarati, count(r1) as interessato
return m.id as ID,
corrispondenze,
dichiarati,
interessato,
(toFloat(corrispondenze))/((toFloat(dichiarati)+toFloat(interessato))-toFloat(corrispondenze)) as jac_similarity
order by jac_similarity desc
```



relations_members-groups-topics.py

results



quantitative measures

quantity of events by Country

quantity of groups by Country

maximum number of participants by Country

average number of guests for each participant

trend topic among users

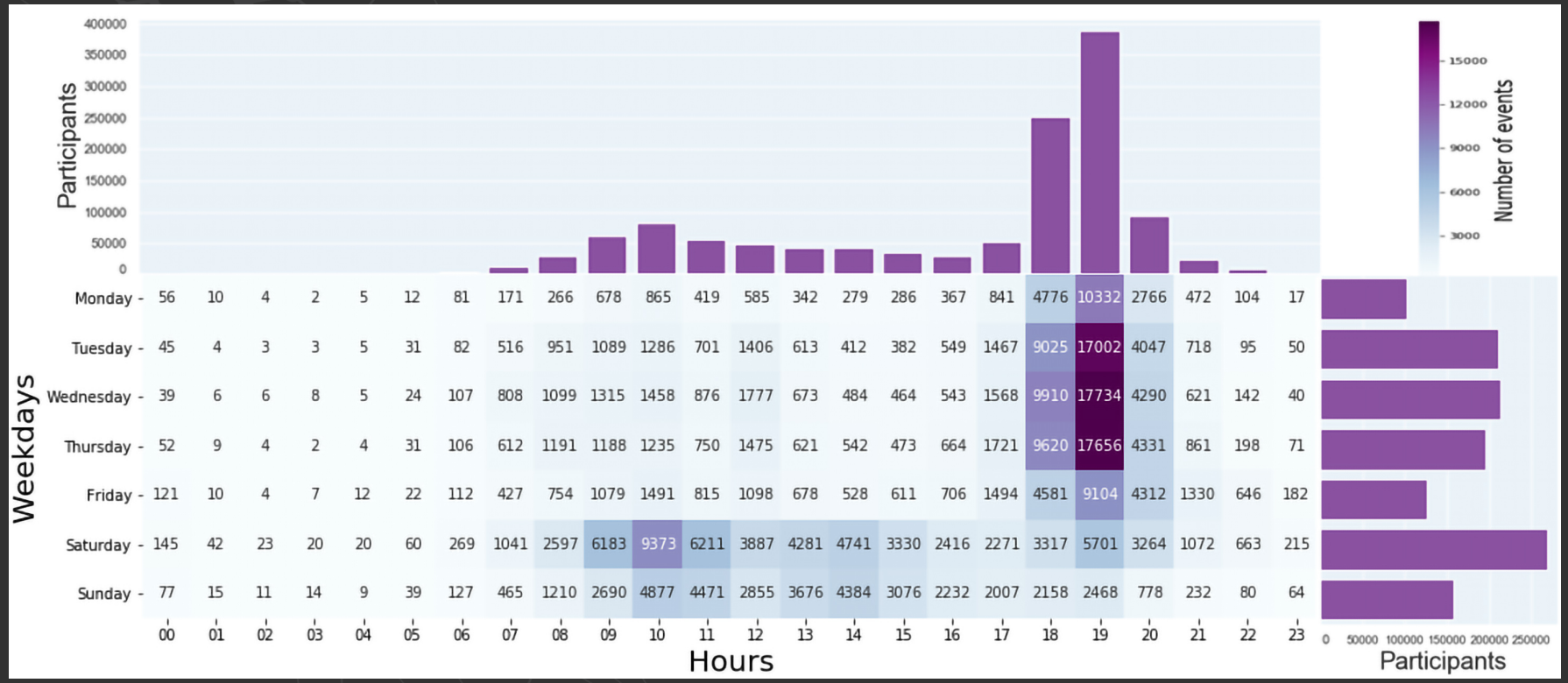
trend topic for groups

temporal distribution of the events

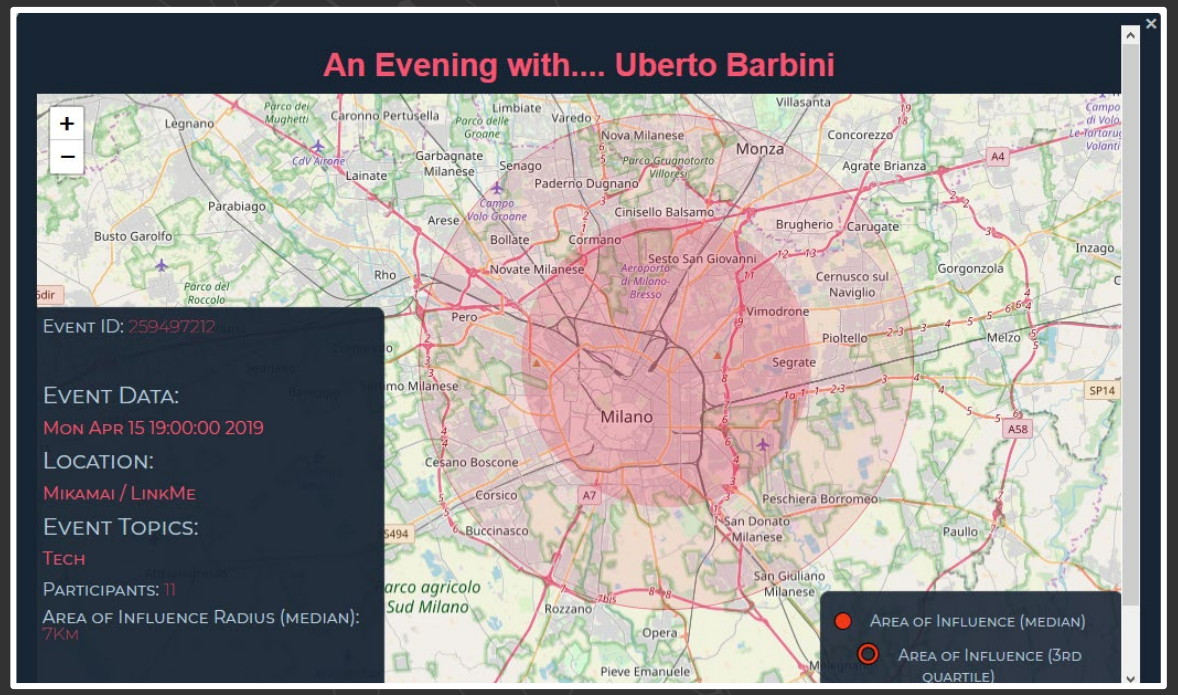
best moment for a meetup

- in a day
- during the week

worldwide vs locally



area of influence (event)



an interactive map that

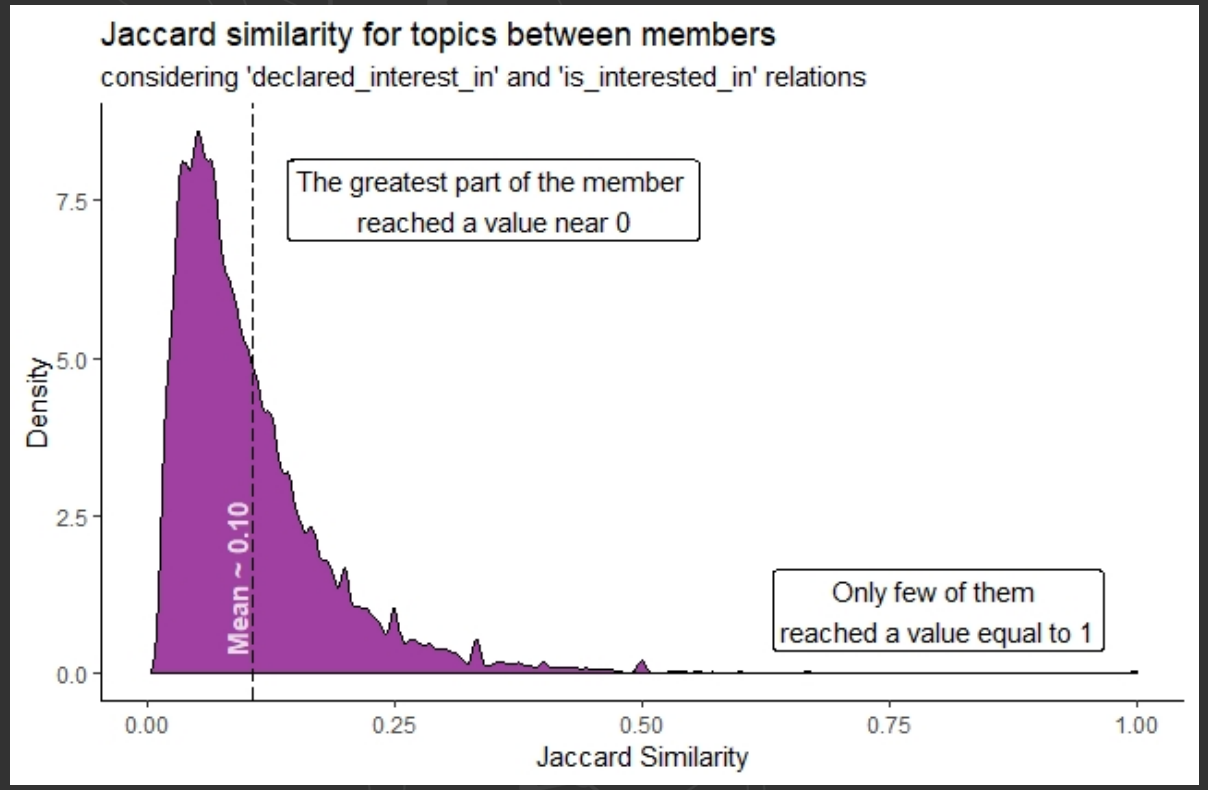
displays useful informations

about events, including a

measure of the area of

influence radius for each event

recommender system analysis



evaluate the recommender system

efficiency investigating similarity between

the group topics and those topics the user

is interested in

it is more like an idea!

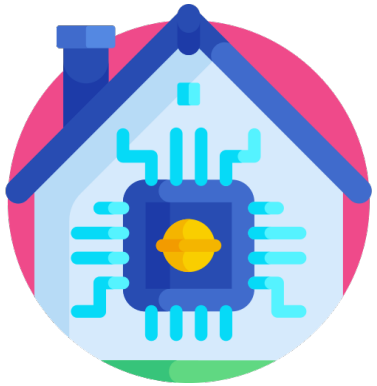
challenges


1 symbolic links
2 import Cypher vs Py2Neo
3 temporal metadata conversion
4 optimization in message extraction
5 import Cypher vs Arangoimp
6 scalability
7 live streaming ingestion

conclusions

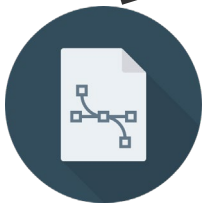
KEY POINTS

answered research questions successfully



software structure benefits from these features

improvement of personal skills



scalability

fault tolerance

consistency

THANK YOU