

Tweets world during Champions League



Damian Kluba, Łukasz Opióła



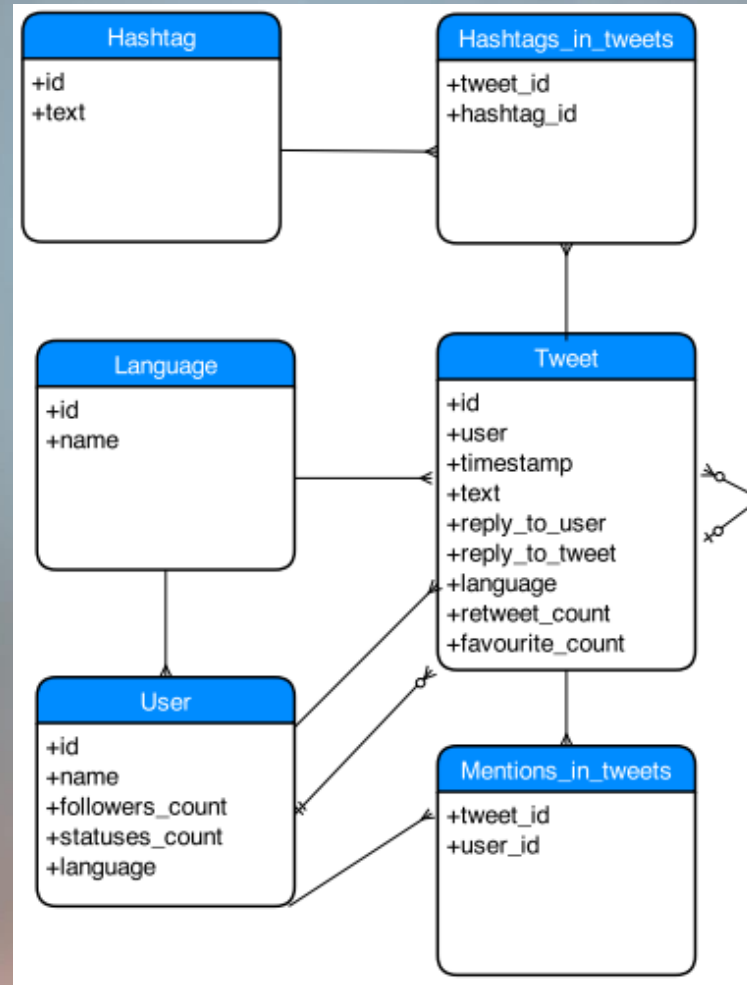
Goals of the project

- Find proper events,
- Prepare hashtags,
- Implement scripts which are using Twitter Streaming API,
- Choose interesting matches,
- Query database for interesting data,
- Create graphs and charts depicting significant features of analyzed data.

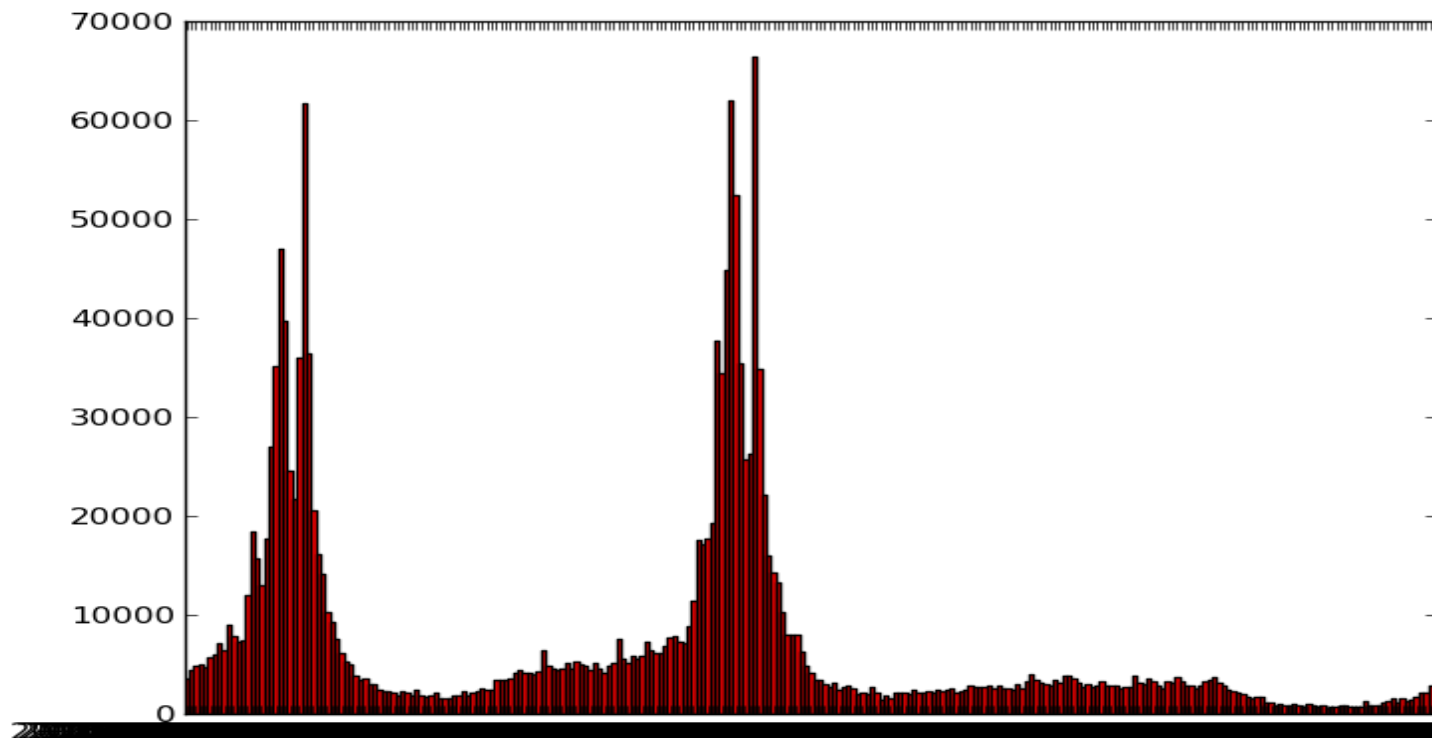
Gathering data

- Twitter Streaming API,
- Prepared 44 hashtags related to champions league and matches,
- Collecting data for almost 3 days (21.10 - 23.10),
- Got **1,784,265** tweets from **674,460** users,
- **2,739,238** hashtags used in those tweets,
- **53** languages used.

DB Schema



Tweets over time

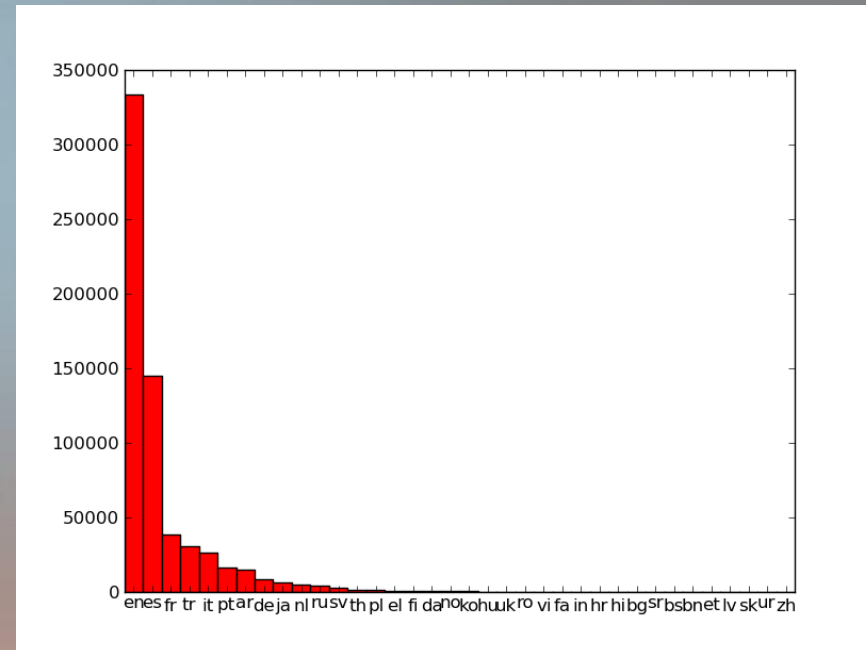


Statistics

most popular languages

- en: 333,249
- es: 145,123
- fr: 38,499
- tr: 30,207
- it: 25,995
- pt: 15,981
- ar: 15,104
- de: 8,726
- ja: 6,528
- nl: 5,151

languages histogram

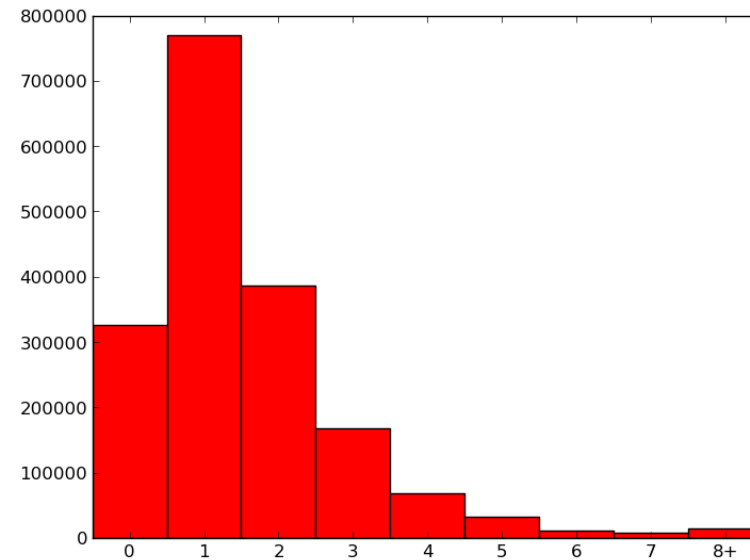


Statistics

most popular hashtags

- #UCL: 399,384
- #HalaMadrid : 267,256
- #LFC : 159,313
- #ChampionsLeague : 104,690
- #RomaBayern : 78,728
- #RealMadrid : 48,884
- #Arsenal : 44,879
- #PSG : 40,224
- #LIVvsRealMadrid : 37,812
- #Roma : 31,317

hashtags histogram

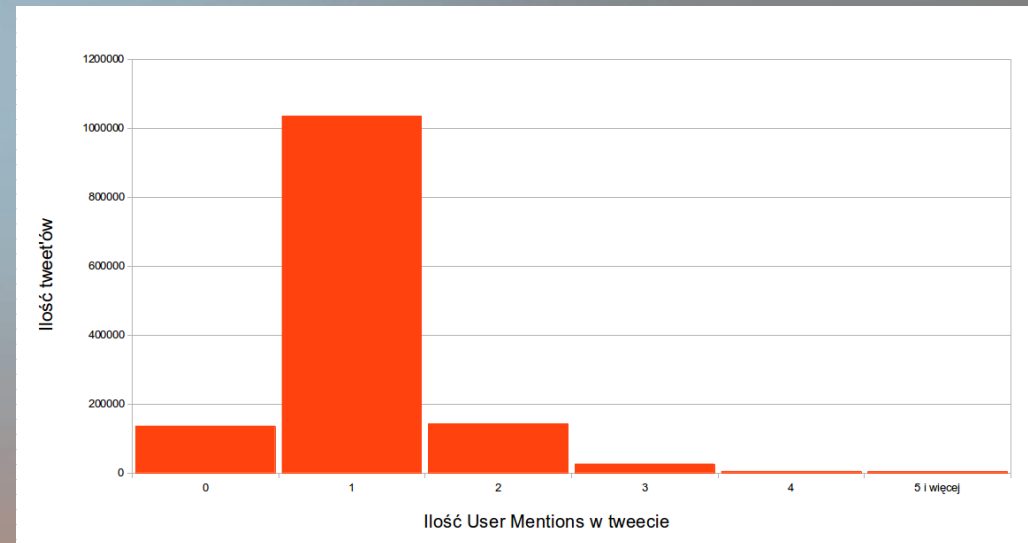


Statistics

most mentioned users

- Champions League: 99,393
- Real Madrid C. F: 80,870
- Liverpool FC: 48,971
- Cristiano Ronald: 26,170
- Metsut Ozil: 24,591
- MisterChip(Alex: 23,835
- Real Madrid C.F.: 18,406
- FC Barcelona: 18,395
- Chelsea FC: 17,464
- Gareth Bale: 16,994

user mentions histogram



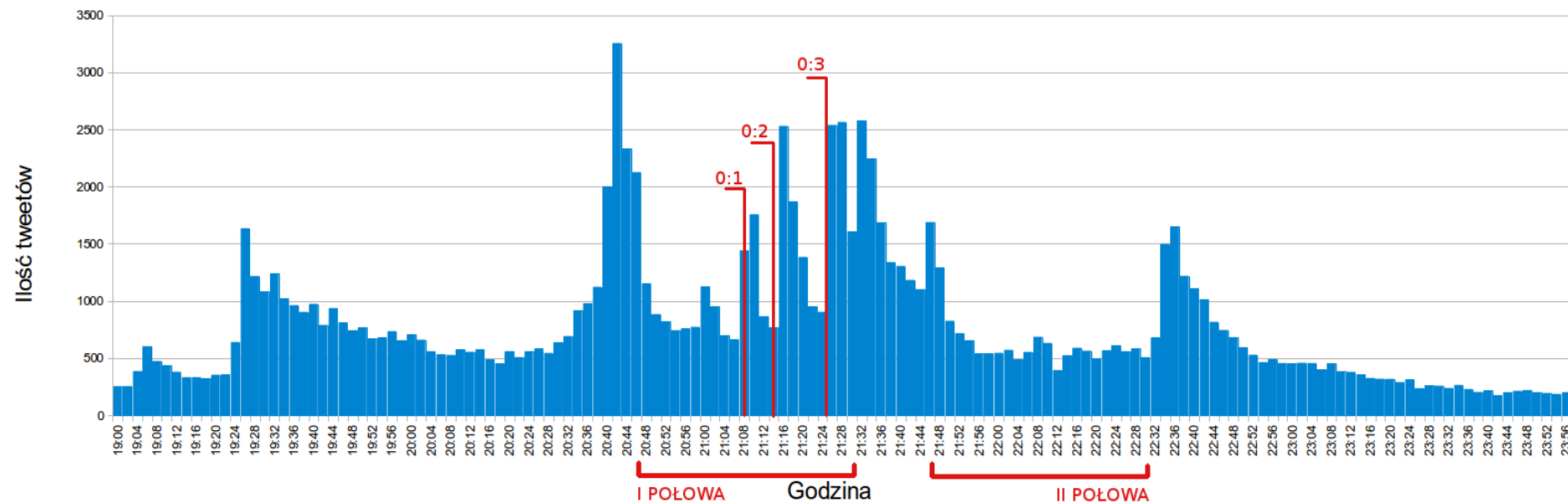
Data analysis

- Picked two most optimistic matches – AS Roma vs FC Bayern and FC Liverpool vs Real Madrid,
- Analyzed user mentions,
- Analyzed retweets (standard numbers plus groups in Gephi and CFinder),
- Analyzed chosen groups.

Liverpool : Real

Liverpool vs. Real Madrid - #lfc & #realmadrid

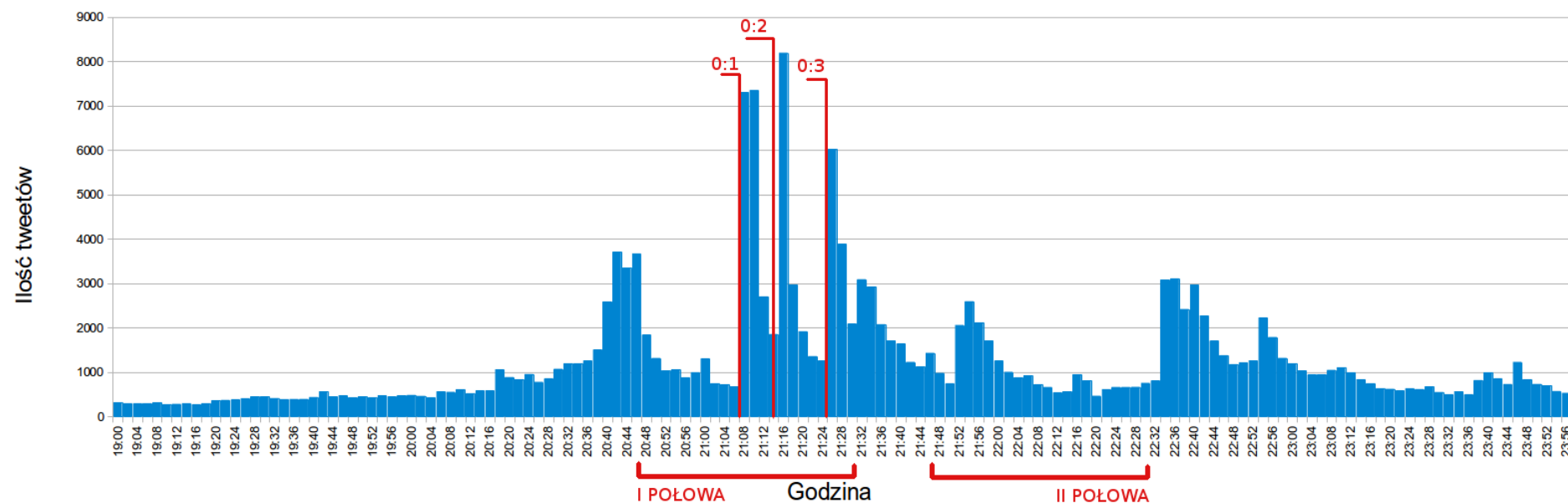
Środa, 22.10.2014



Liverpool : Real

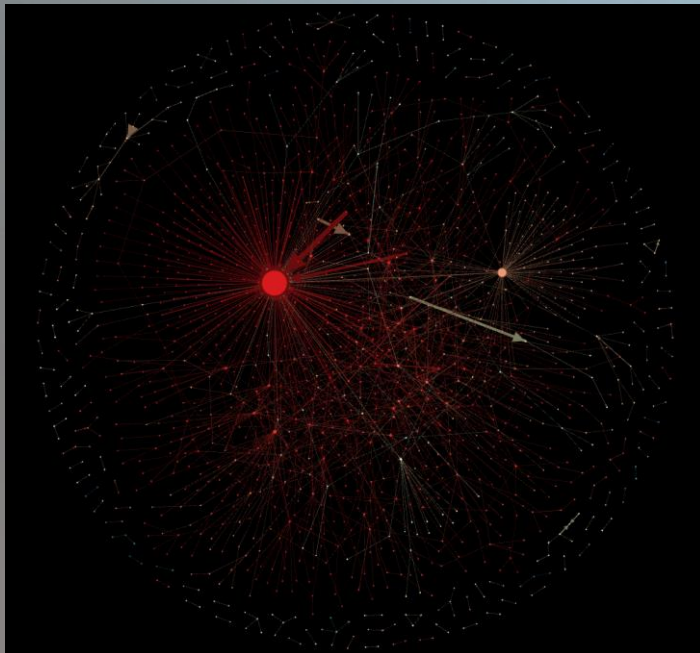
Liverpool vs. Real Madrid - #cr7 & #realmadrid & #halamadrid

Środa, 22.10.2014



Liverpool : Real

retweets

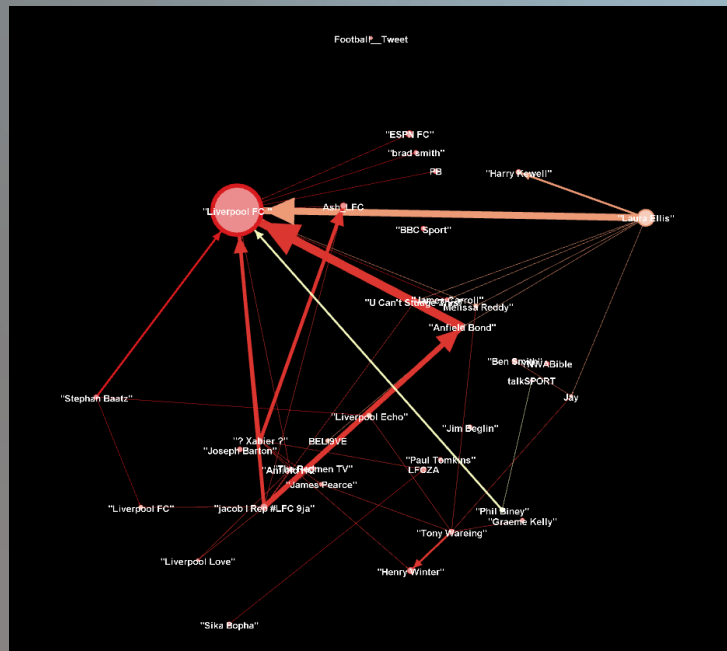


user mentions

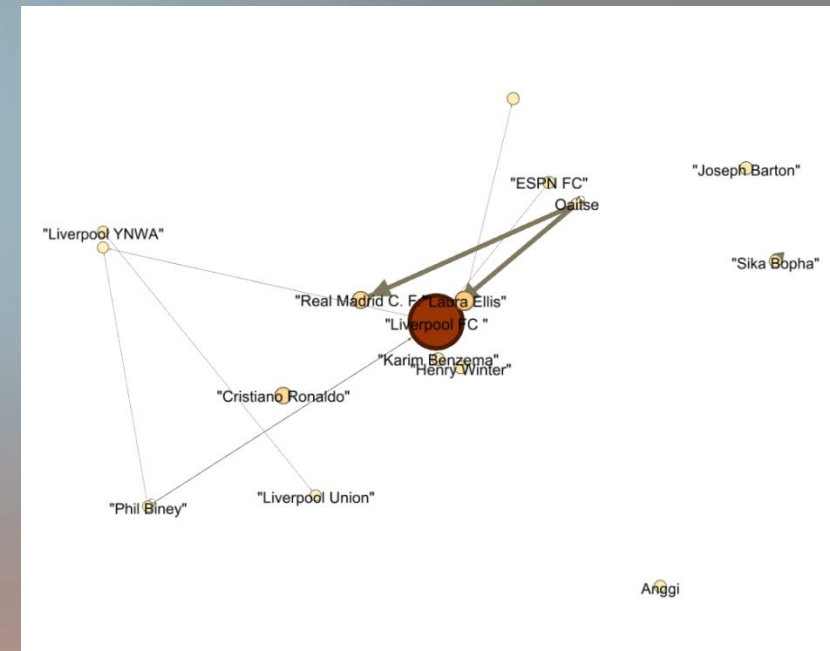


Liverpool : Real

retweets (k>10)



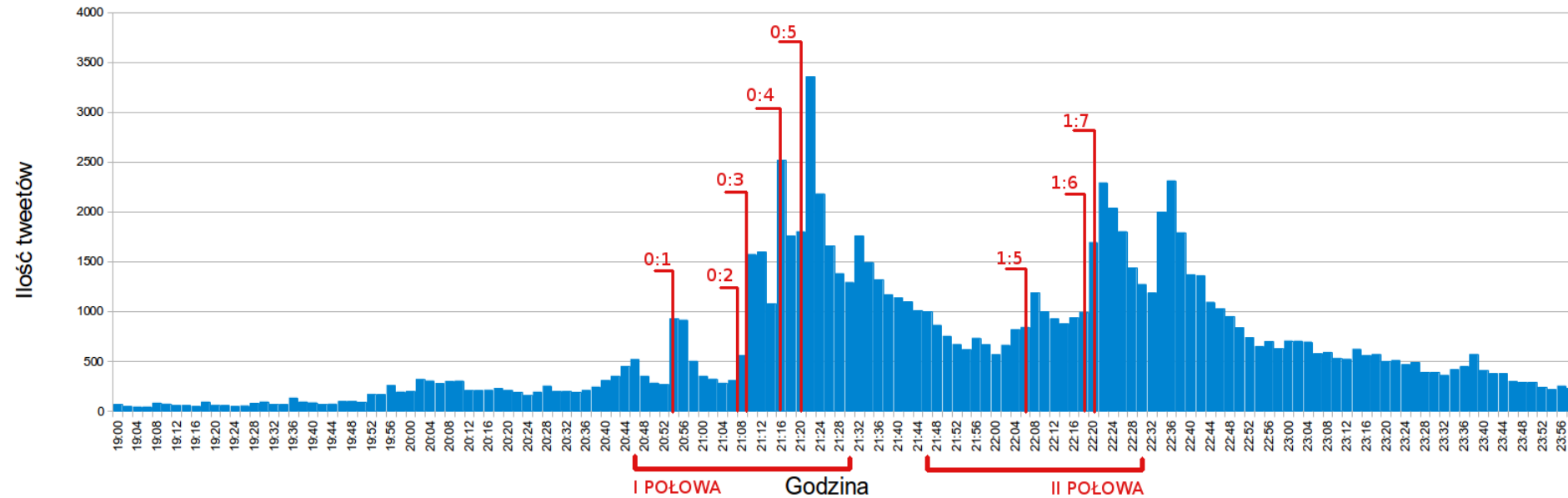
user mentions (k>20)



Roma : Bayern

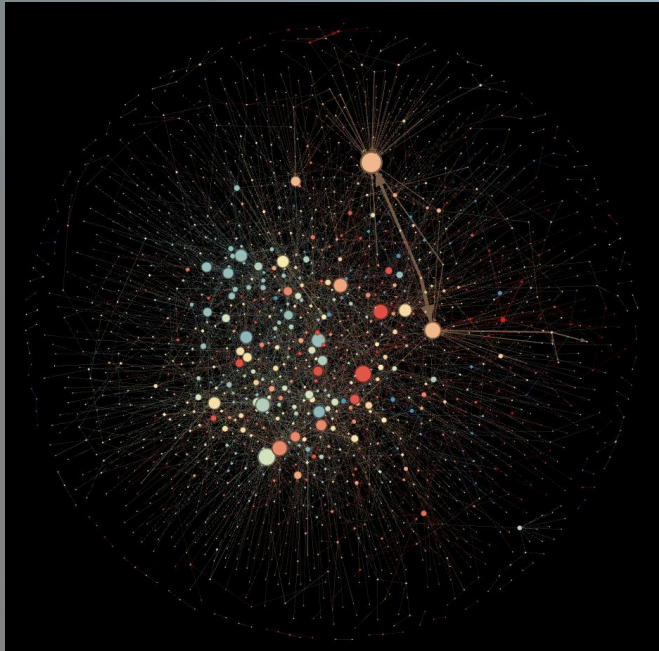
Roma vs. Bayern - wszystkie tweety

Wtorek, 21.10.2014

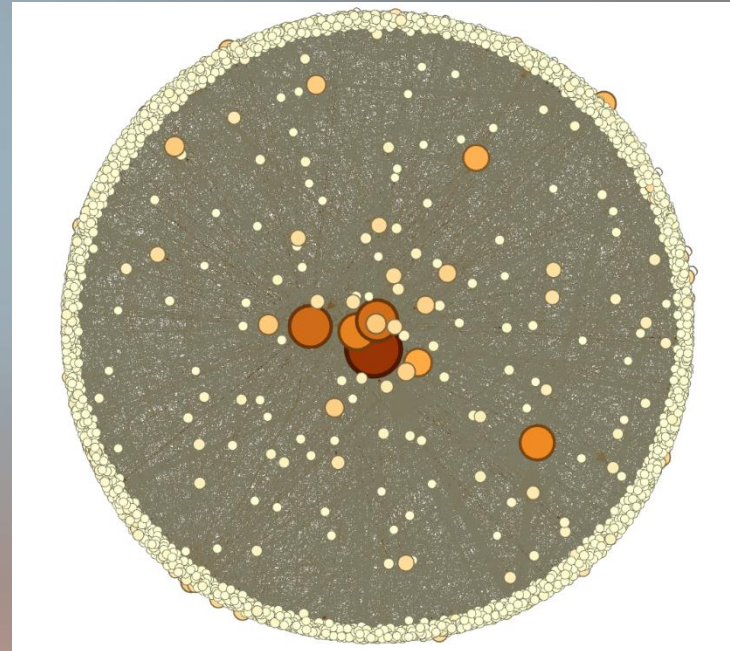


Roma : Bayern

retweets

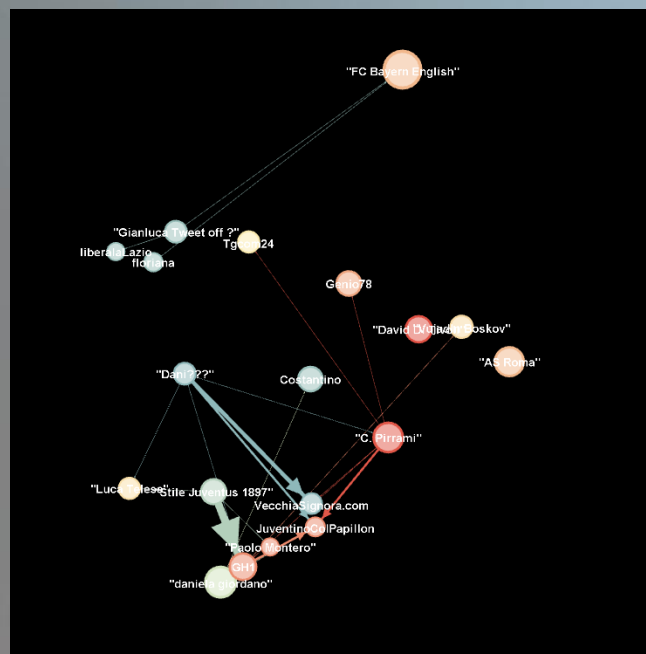


user mentions

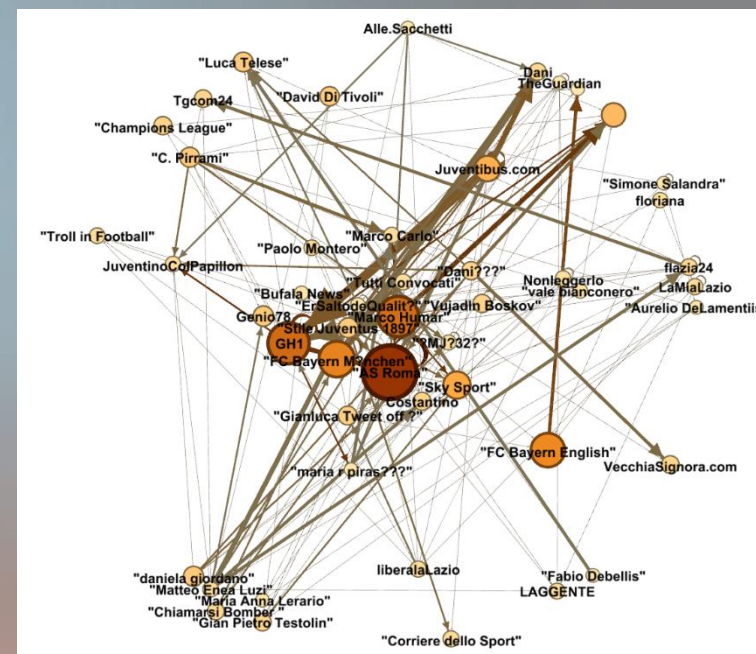


Roma : Bayern

retweets (k>20)



user mentions (k>25)



Conclusions

- Hashtags and user mentions are very popular on twitter platform (1/6 of tweets didn't have hashtag, 30% of tweets didn't mention anybody),
- Users quickly responds on twitter after some events,
- „reply to” option is rarely used on twitter in comparison to user mention,
- Social networks between italians are bigger than germans, english or spanish,
- It's really important to choose proper keywords to search for data like this.

What can be done

- Download additional data through REST API (for example, the database misses source tweets of some retweets),
- Search for dependencies between matches,
- Similar research on data from final match to see how the prestige of the event affects twitter's traffic.