iSocial

Investigating the complete corpus of Referendum and Elections tweets

Despoina Antonakaki

despoina@ics.forth.gr, antonakd@gmail.com

FORTH Institute of Computer Science

MARIE CURIE ACTIONS PEOPLE

The Events / political background

- New "anti-austerity" government of the SYRIZA party is elected at 25th January (36.3%)
- They start a long negotiation about debt reconstruction with Eurogroup
- Until June 2015 no visible progress has been achieved. The pressure from inside and outside Greece to "get over this" was immense
- SYRIZA decides to throw a referendum at 5th of July so that the Greek people decide whether to accept or not the current austerity measures proposed by Eurogroup.
- Capital controls are enforced in Greece
- The result of the referendum was "do not accept" with 61.3%.

The Events / political background

- Eurogroup does not accept the result of the referendum as a bargaining tool.
- Under extreme pressure the government decides to accept the proposed measures.
- Internal disagreement from SYRIZA party threat not to support or vote the measures.
- Prime Minister decides to expel disagreeing Members of the Parliament and other members and announces new legislative elections at 20th of September.
- SYRIZA wins again (35.5%) and the party formed from disagreeing members (LAE) does not enter the parliament (minimum percentage required is 3%)

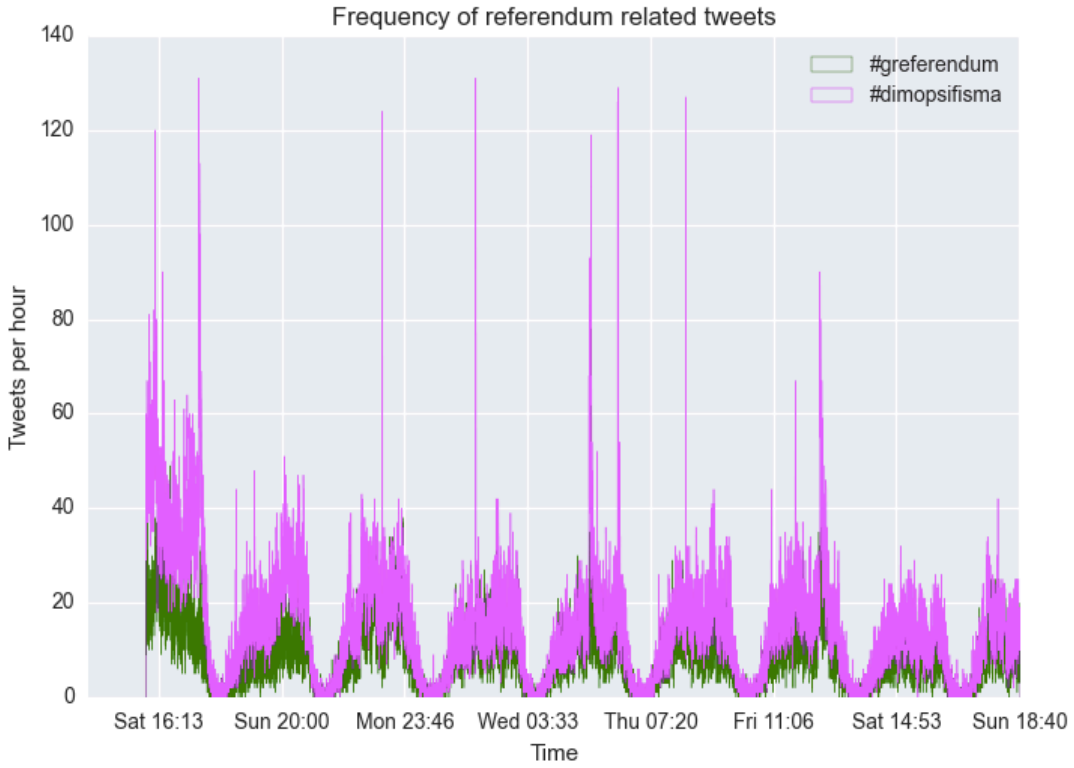
Referendum data collection

Collect tweets:

- Collection period:
  - from: 27th June (announcement)
  - to: 5th July (referendum date)
- #dimopsifisma: 289.516 tweets (237.155 at least one greek letter)
- #greferendum: 435.982 tweets (105.970 at least one greek letter) - mainly used by international users
- Common tweets: 300.893

Frequency of tweets per hour, with at least one greek character

- Day/night patterns
- A decline over time of tweets frequency



YES / NO hastags

Manually assign all strong associated hashtags to YES/NO

ναι
yes
menoume_evropi
menoume_eurwph
menoumeevropi
syriza_xeftiles
syriza_apateones
Nai
team_nai
pasok
πασοκ
πασόκ
Nd
NeaDimokratia
νδpotami
topotami
ποτάμι
ποταμι
τοποταμι
τοποτάμι

← YES (ναι)

NO (όχι) →

όχι
οχι
oxi
oxi2015
no
team_oxi
lemeoxi
πεςοχι
Noway
noeurope
isayno
notroika
no2015
notoausterity
οχι2015
nointhenameofdemocracy
nomeansno
saynotoausterity
say_no
vote_no
nobanks
noeu
noeuro
όχιευχαριστώ
voteno
vote_for_no
voteforno
teamοχι
noausterity
greecesaysno
λέμεόχι
sayno
non
sayingno
no_team
λέμεοχι
team_no
team_οχι

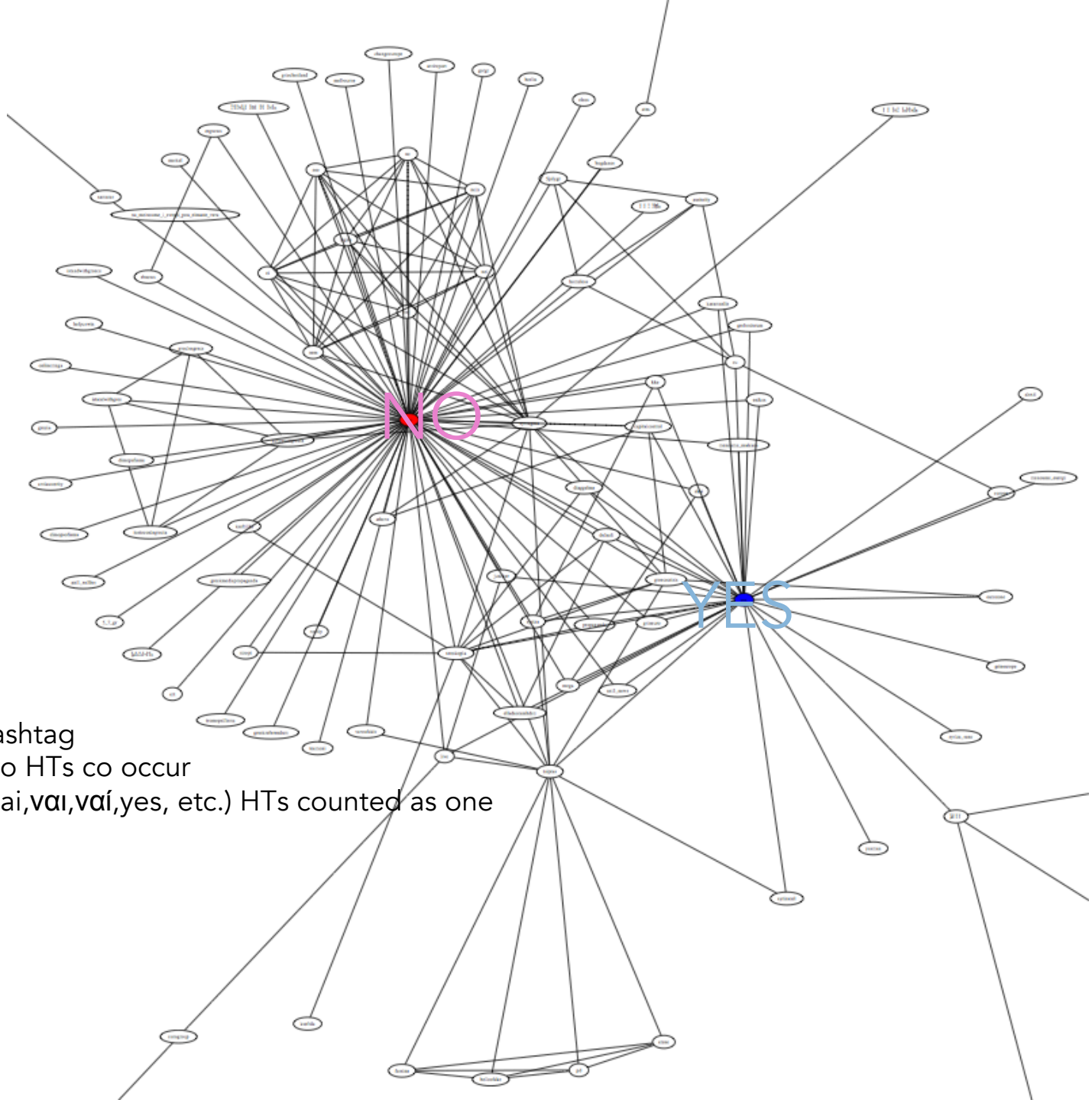
Although the number of NO hashtags are more, but many of them are actually contained in a very few tweets

Hashtags co-occurrence

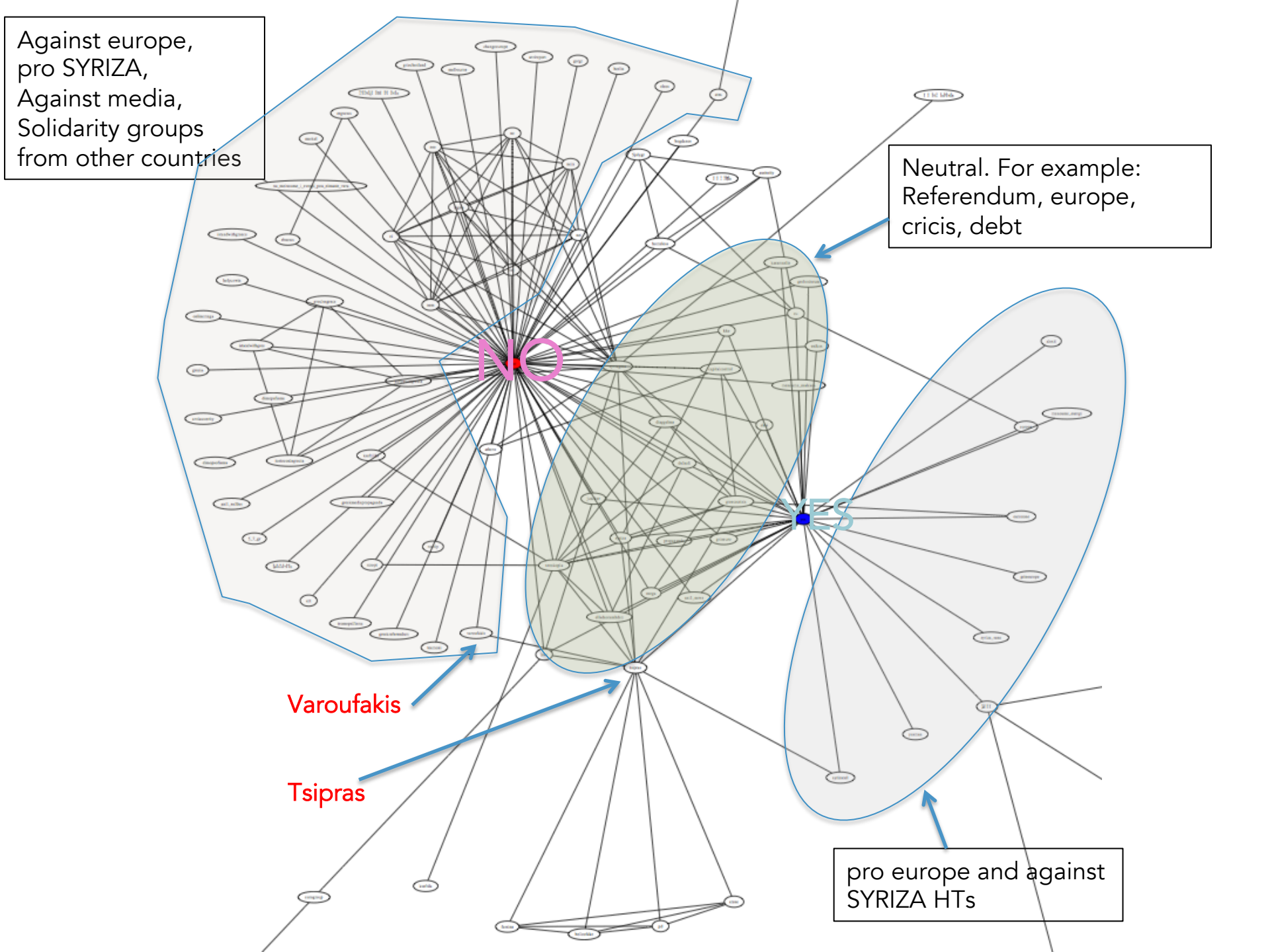
- Two hashtags are co-occurring if there exist at least one tweet that contains both hashtags
- Define distance between hashtags as:

$$d = \log(10 + c_{max} - c)$$

- C is the number of tweets that contains a specific pair of hashtags
- C_{max} is the maximum C (max co-occurrence)
- Apply the "neato" visualization method (graphviz) that emulates *spring link attractive forces* between nodes.

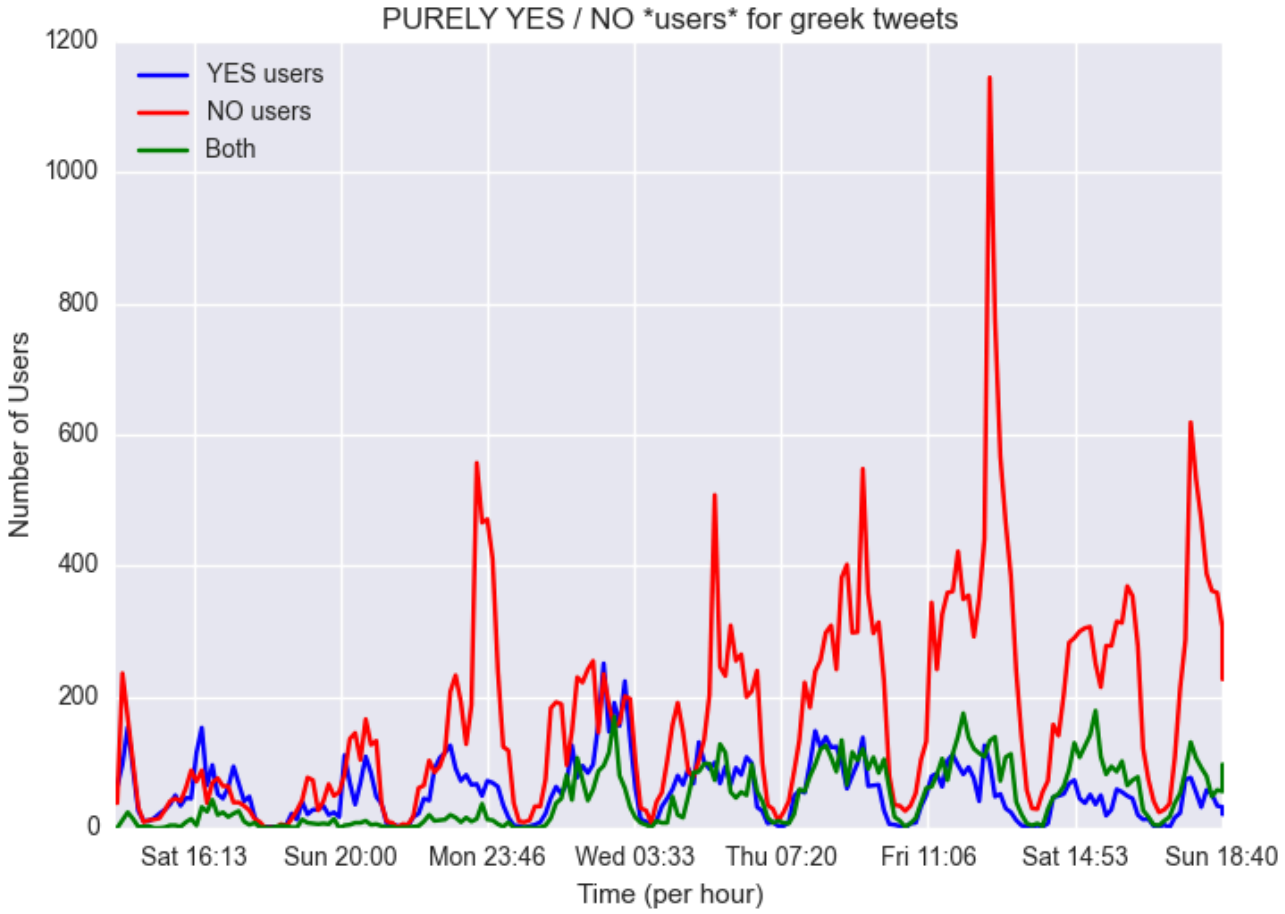


- Node: hashtag
- Edge: two HTs co occur
- All YES(nai,ναι,ναί,yes, etc.) HTs counted as one


Against europe,
pro SYRIZA,
Against media,
Solidarity groups
from other countries
Neutral. For example:
Referendum, europe,
cricis, debt
NO
YES
Varoufakis
Tsipras
pro europe and against
SYRIZA HTs

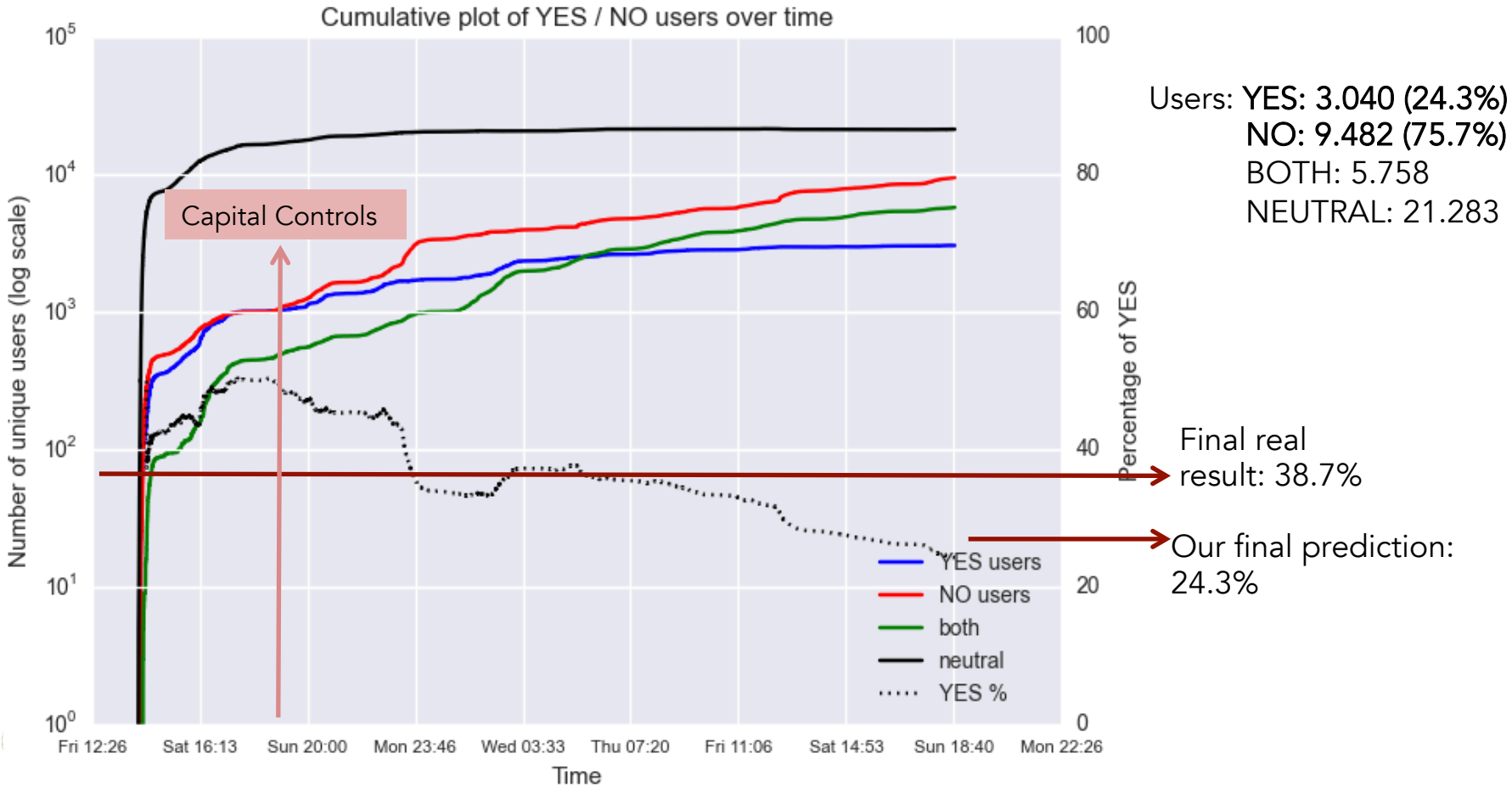
User YES/NO inclination

- For every hour measure the number of users that have posted ONLY YES and ONLY NO tweets
- The “NO” gain momentum the last days
- The spikes are mainly RTs of Tsipras (user of #NO hashtag)



Progress of YES/NO tendency

- For each hour measure only NEW users that have posted **only** YES or NO tweets. Measure cumulative percentage.
- Users of NO are continuously increasing. Users of YES after Monday are almost the same
- The YES percentage flatlined from Tuesday
- The YES percentage is plundering the last days
- Announcement of Capital Controls seems to have increased the YES percentages



Users: **YES: 3.040 (24.3%)**
NO: 9.482 (75.7%)
BOTH: 5.758
NEUTRAL: 21.283

Final real result: 38.7%

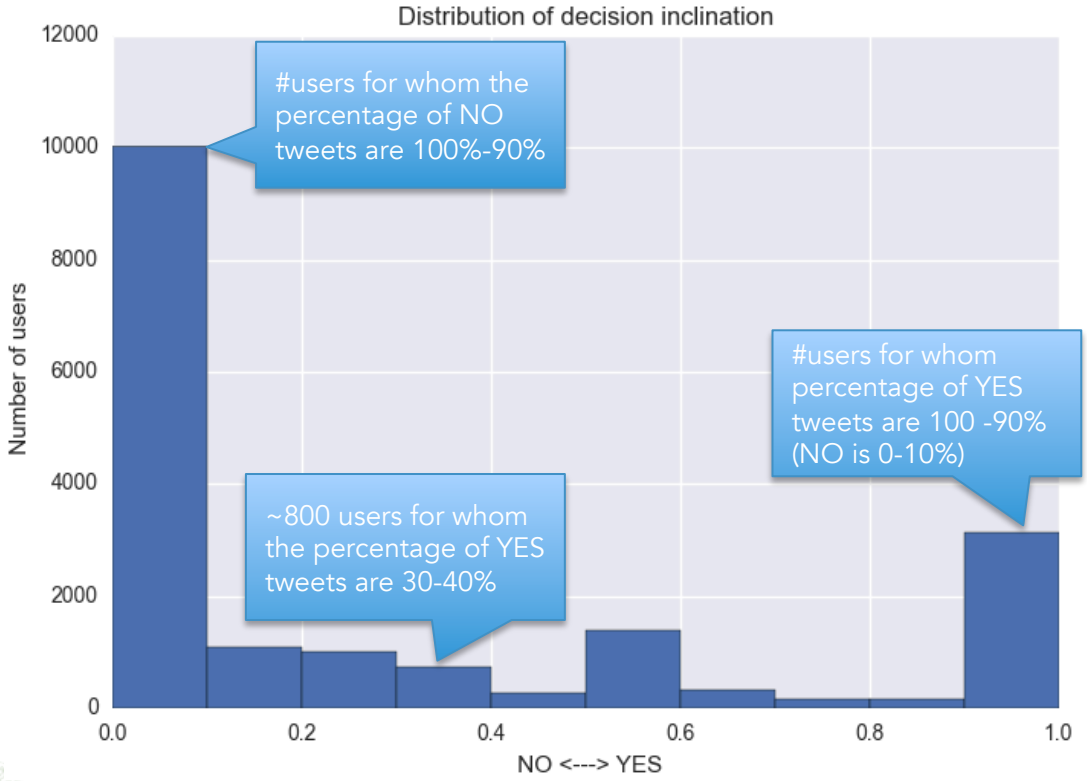
Our final prediction: 24.3%

Most users are opinionated

For every user that has posted at least one "YES" or "NO" tweet, measure the inclination towards "YES" or "NO" as:

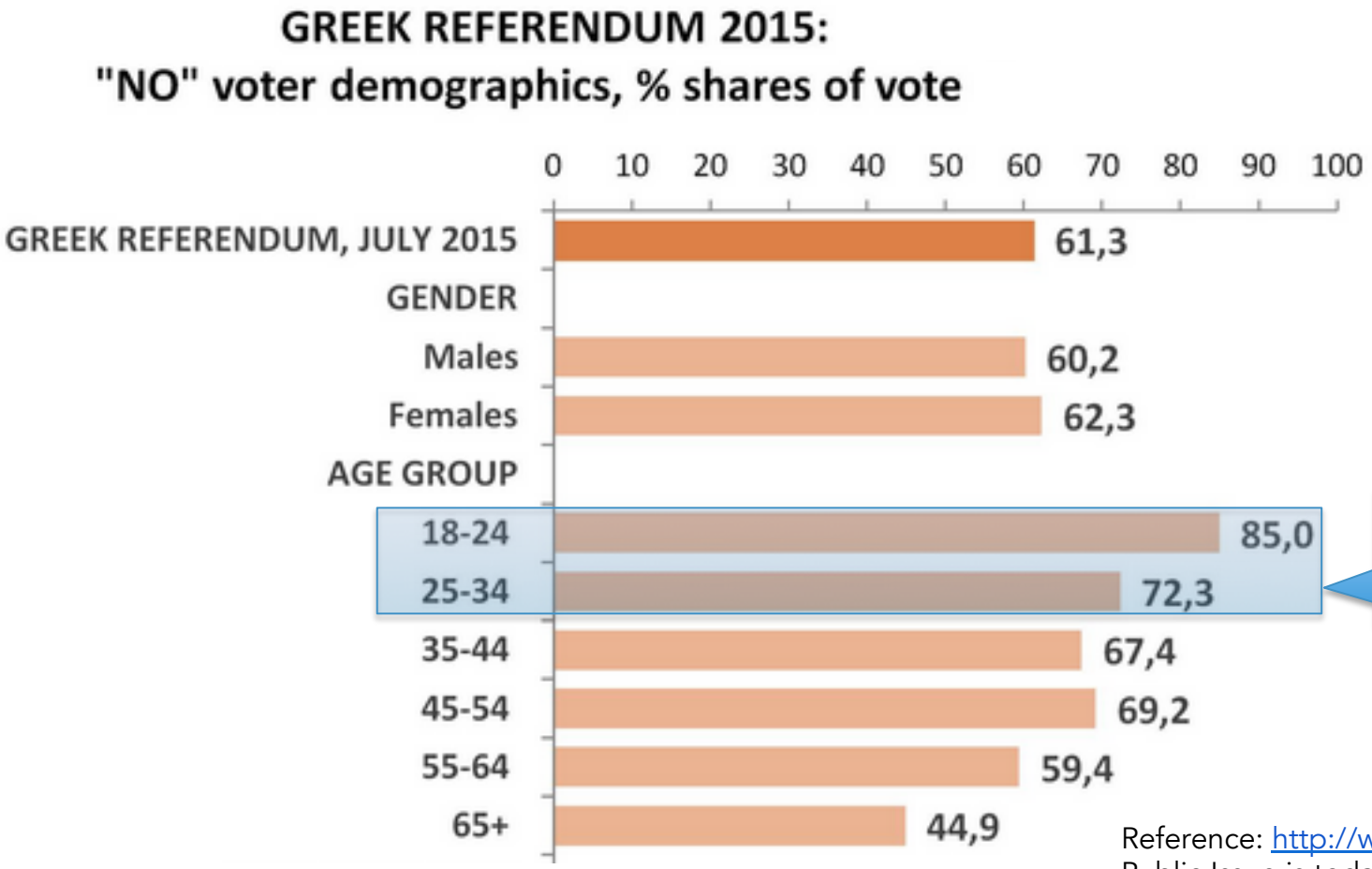
$$d = YES_{tweets} / (YES_{tweets} + NO_{twees})$$

(0 = All tweets are NO, 1= All tweets are YES)



Histogram with 10 bins

Referendum demographics

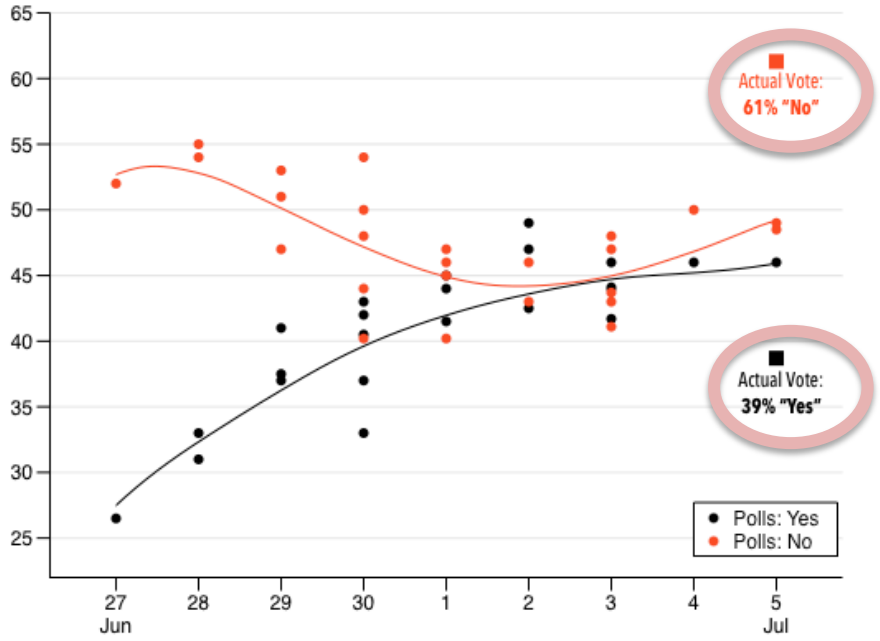


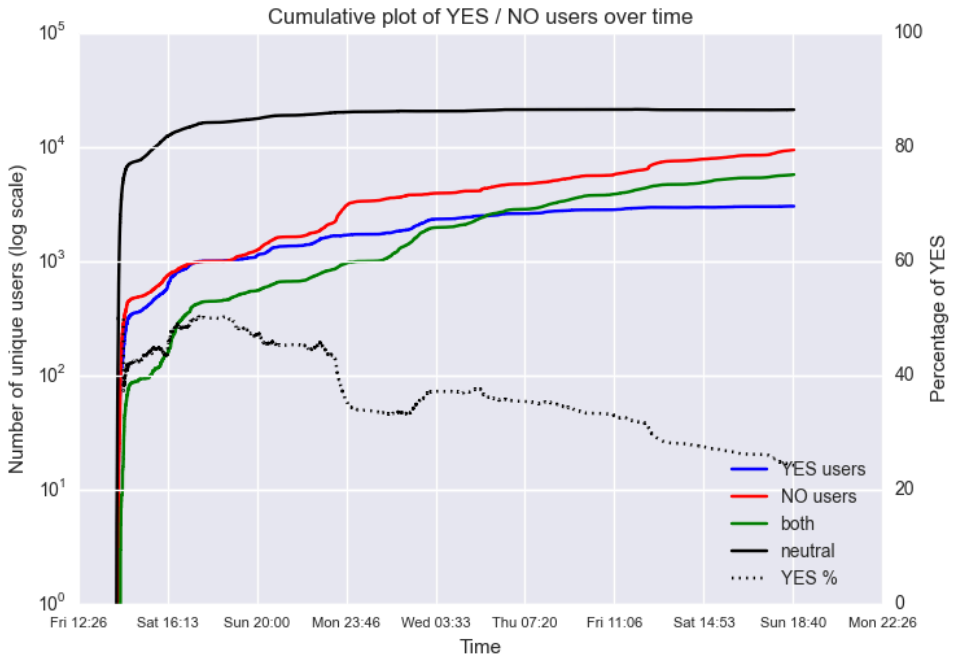
Our prediction: 75.7%

Reference: http://www.publicissue.gr/en/2837/
Public Issue is today one of the leading opinion polling companies in Greece.

- These results were published post referendum
- Our results are inline with Twitter demographics assuming that the average Greek Twitter users are between 18 and 34

Reference: http://www.huffingtonpost.com/2015/07/08/greece-polls-wrong_n_7754874.html





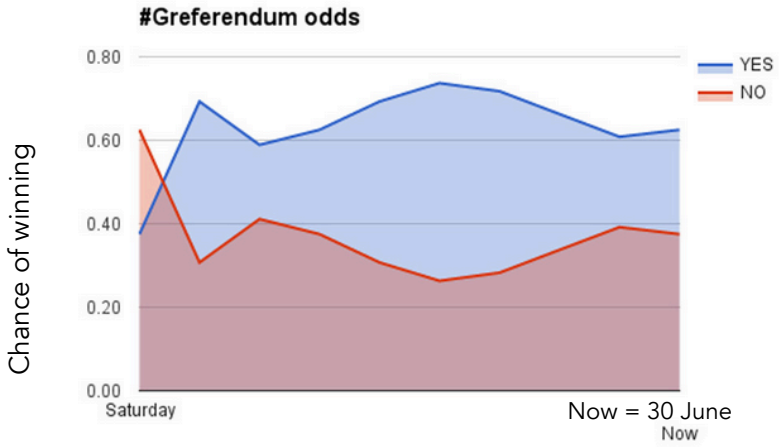
Opinion polls

- Every dot is a poll (red: NO, black: YES)
- Greek opinion polls failed to make accurate predictions:
- Late opinion polls measured the YES vote close to 45% (actually got 39%)
- Opinion poll error 45-39 = 6%
- Our error: 39-24 = 15%
- Researchers estimate that there was a **"late NO trend"**
- Our estimation shows upward trend for NO, polls showed upward trend for YES
- It was captured in Twitter data but not in opinion polls

Online betting got it wrong as well!

Online betting odds:

Ladbrokes



The chance of winning if you bet on YES was higher

Betfair

Majority Winner						Matched: £138,669
Going In-Play			Back	Lay		
Yes to proposal	1.45 £179	1.46 £1,043	1.48 £6,505	1.49 £2,134	1.5 £1,765	1.52 £87
No to proposal	2.88 £1,197	2.92 £45	3 £1,943	3.1 £3,105	3.2 £476	3.25 £79

Yes vote was favorite at 4/11(bet £11 to profit by £4), but offering 2/1 that they will reject the plan

William Hill

William Hill.....Greece to vote 'Yes' to the EU/IMF & ECB bailout plan- 4/11. Greece to vote 'No' to the EU/IMF & ECB bailout plan- 2/1 (bets void if Referendum does not take place during July2015)

Sources:
Ladbrokes: http://news.ladbrokes.com/politics/greece-set-do-thing-numbers-add.html
William-hill: http://www.williamhillplc.com/media/newsroom/media-releases/2015/greeks-21-to-reject-bailout-plan-in-referendum/
Betfair: http://www.bettingmarket.com/betfairgreece.html

Conclusions on referendum tweets

- Twitter data achieved to:
  - Capture the downward trend of “YES”
  - Estimate the percentage of young (<35) users.

- Twitter data fail to:
  - Make better overall estimations than opinion polls.
    - Polls’ error was towards the wrong direction, making “YES” a likely outcome (45% against real final: 38.7%)

http://hellenicreferendum.blogspot.gr

#ThisIsACoup

Background:

- After the referendum (5th of July), a Eurogroup meeting took place at 13th of July where the "No" verdict was greatly disregarded as a negotiation medium forcing the Greek government to accept more harsh austerity measures than anticipated.
- This generated a large backlash that was iconized with the #ThisIsACoup hashtag making it a world wide trend for a few days.
- We collected all #ThisIsACoup tweets from 13th of July until 16th of July (580.00 tweets)

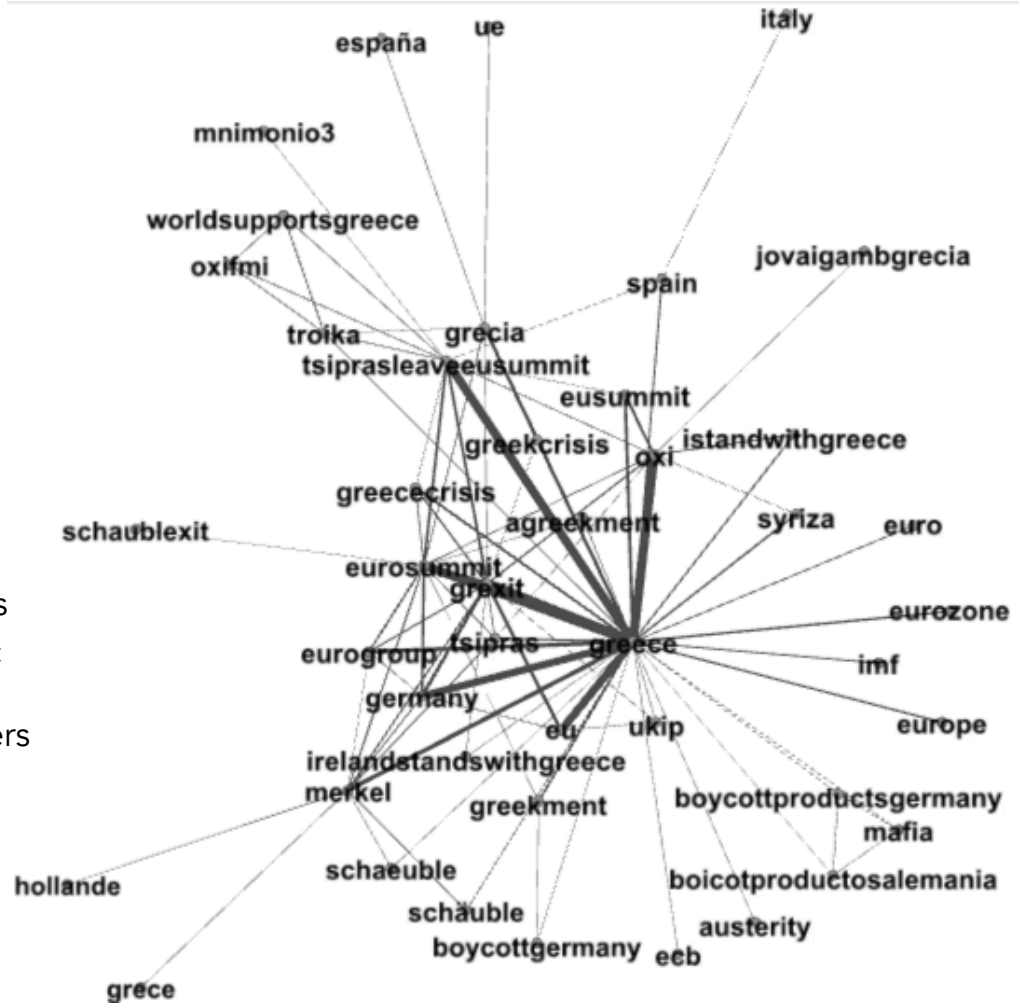
#ThisIsACoup Hashtag co-occurrence

Measure all pairs of co-occurring hashtags:
For example:
Tweet_1 = "#thisisacoup #A #B"
Tweet_2 = "#thisisacoup #A #B #C"
Then:

co-occurrence(A,B) = 2
co-occurrence(A,C) = 1
co-occurrence(B,C) = 1

Plot all co-occurrences > 500

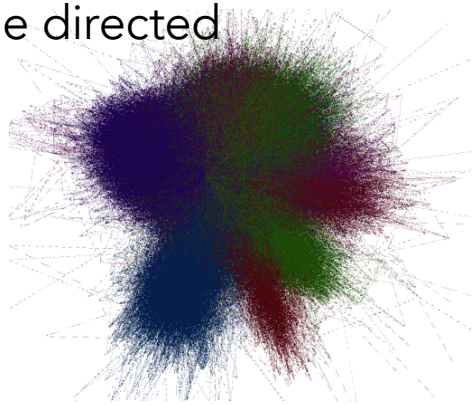
- The edge **width** is proportional to co-occurences
- Maybe a tag-cloud/HASHCLOUD of euroskeptic concepts!
- Here we have only international opinionated users raging for NO



We use a different method to measure the co-occurrence (not matplotlib but gephi)

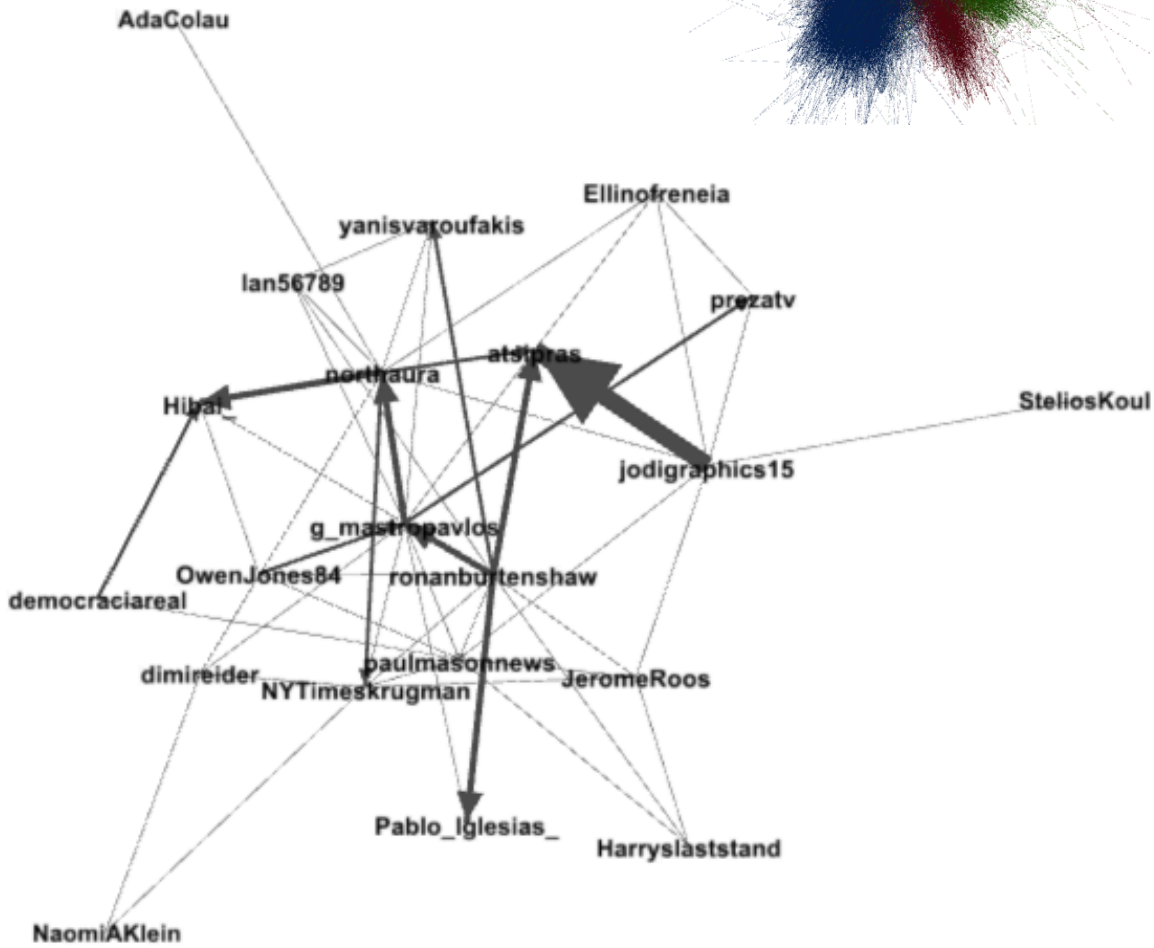
#ThisIsACoup user mentions

- If a user A, posts a tweet that mention user B, then we have the directed edge: A →B
- The complete graph contains: 120.000 nodes, 448.000 edges
- Plot the 20 most mentioned (or mentioning) users.
- Arrows' width is proportion to the number of mentions.



It contains mainly "celebrities":

- Politicians: Tsipras, Varoufakis, Ada Colau (Mayor of Barcelona), Pablo Iglesias (Podemos).
- Journalists: Owen Jones, Jerome Roos, Stelios Kouloglou, Paul Mason, Krugman
- Writers: Naomi Klein



Greek legislative elections of 20th September

- Hashtags:
  - #ekloges2015_round2, 141.504 tweets
  - #ekloges2015, 45.858 tweets
- Collection period: 20th August (announcement of elections) – 19th September (night before the elections)
- Merge all tweets and remove duplicates.
  - Final dataset: 181.815 unique tweets.
- Collect all hashtags that had more than 100 tweets and were strongly affiliated to a specific party.
- We assigned users to parties according to these hashtags.

Correction according to referendum results

- We have noticed that Greek Twitter users are not an unbiased representation of the Greek electorate, mainly due to age demographics
- So… we applied a correction on the estimated statistics based on the official position towards YES or NO of the parties. We estimated our bias as:
  - $NO_{correction} = REAL_{NO} / PREDICTION_{NO} = 0.613 / 0.757 = 0.8$
  - $YES_{correction} = REAL_{YES} / PREDICTION_{YES} = 0.387 / 0.243 = 1.6$
- So every vote to a NO party counts as a 0.8 of a vote
- Every vote to a YES party counts as a 1.6 of a vote
  - YES Parties: PASOK, ND, POTAMI
  - NO Parties: SYRIZA, ANEL, LAE
  - Neutral parties: XA, EN.KETROON, KKE

From tweets to votes

- Which tweets should we take into account? Two methods:
  1. When it contains at least one "party" hashtag
  2. Same as 1 but remove tweets that contain **more** than one "party" hashtags.

- What should we do to users that have posted different tweets with different parties? Three methods:
  1. Cast one vote distributed equally to parties.
  2. Cast vote to the majority of tweets. If there is a tie then exclude user.
  3. Exclude user.

In total we have 2 X 3 = 6 measurement methods

Final Results:

PASOK:	6.3%
POTAMI:	4.1%
LAE:	2.9%
EN.KENTROON:	3.4%
XA:	7.0%
ND:	28.1%
SYRIZA:	35.5%
KKE:	5.6%
AN.EL.:	3.7

- Succeeded percentages estimation of two rival parties (SYRIZA, ND)
- Failed at POTAMI, LAE (assume very strong young voting base)
- Colored boxes: Subtract prediction for every party & sum absolute error
- Vote fluctuations according to method but result almost same
- Many tweets with negative sentiment towards these parties. Future work: correct with sentiment analysis
- All methods: 5.700 - 9.800 votes. Far higher than a regular opinion poll sample size (from 1500 to 2000) samples

All party tweets

Proportion distribution of multiple party tweets

Party	Percentage	Votes
PASOK	6.675463	651
POTAMI	14.484178	1414
LAE	8.736517	853
EN.KENTROON	2.603625	254
XA	6.300000	615
ND	26.673852	2604
SYRIZA	25.724607	2511
KKE	3.919712	382
AN.EL	4.882047	476
Total	100.000000	9764

32.2

The vote goes to the majority party. Ties are excluded

Party	Percentage	Votes
PASOK	4.776805	377
POTAMI	14.088550	1114
LAE	8.188908	647
EN.KENTROON	1.922083	152
XA	6.300000	498
ND	24.579381	1943
SYRIZA	30.583756	2418
KKE	4.464632	353
AN.EL.	5.095884	402
Total	100.000000	7908

30.0

Remove users that have different tweets to different parties

Party	Percentage	Votes
PASOK	5.136447	293
POTAMI	15.701660	896
LAE	8.264429	471
EN.KENTROON	2.346344	134
XA	6.300000	359
ND	24.930561	1423
SYRIZA	29.660894	1693
KKE	3.399048	194
AN.EL.	4.260616	243
Total	100.000000	5711

31.7

Exclude tweets with more than one party hashtag

Party	Percentage	Votes
PASOK	6.321747	584
POTAMI	15.290248	1414
LAE	8.576233	793
EN.KENTROON	2.864989	265
XA	6.300000	582
ND	25.351229	2345
SYRIZA	26.330763	2436
KKE	4.101461	379
AN.EL	4.863330	449
Total	100.000000	9251

32.7

Party	Percentage	Votes
PASOK	4.920304	391
POTAMI	14.340887	1141
LAE	8.210179	653
EN.KENTROON	2.085061	166
XA	6.300000	501
ND	24.541517	1953
SYRIZA	30.022923	2390
KKE	4.488650	357
AN.EL.	5.090479	405
Total	100.000000	7961

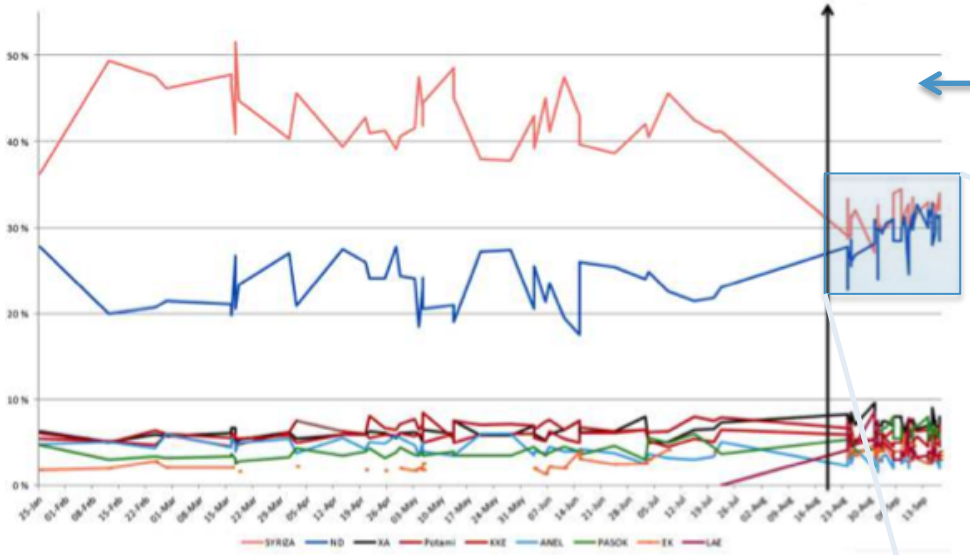
30.5

Party	Percentage	Votes
PASOK	5.028733	307
POTAMI	15.592971	953
LAE	8.237824	503
EN.KENTROON	2.453371	150
XA	6.300000	385
ND	25.026719	1530
SYRIZA	29.725905	1817
KKE	3.580944	218
AN.EL.	4.053533	247
Total	100.000000	6114

31.0

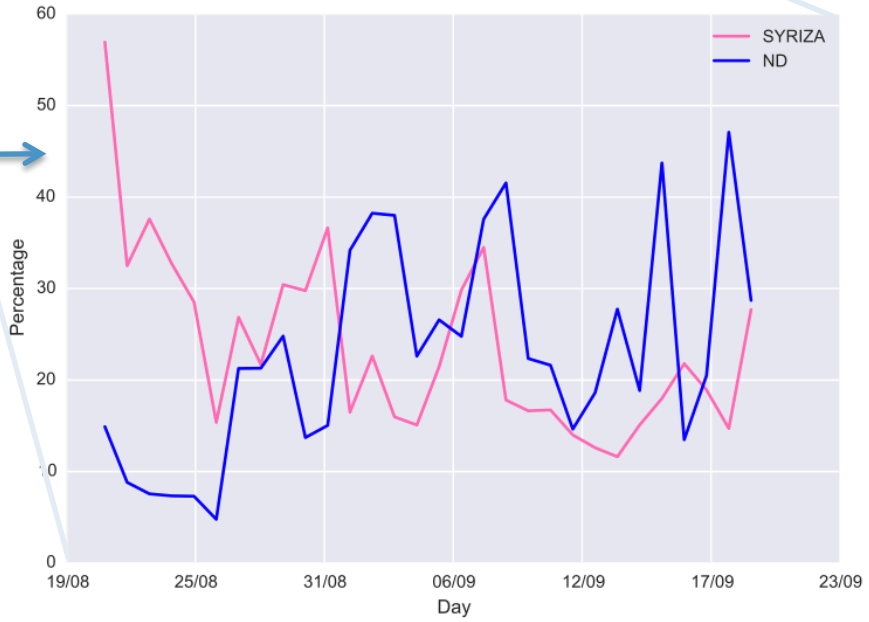
Comparison with opinion polls

Start of our collection time (20 Aug) – elections are announced



Opinion polls prediction.
Source wikipedia:
https://en.wikipedia.org/wiki/Greek_legislative_election,_September_2015

Prediction day by day according to the best measurement method for the two rival parties (SYRIZA, ND)



Data from opinion polls after the announcement of the elections are almost in concordance with estimation from Twitter

Latent Dirichlet Allocation

- An automatic way to discover topics in a coprus
- How? LDA represents docs as mixtures of topics that spit out words with certain probabilities **(unsupervised clustering)**
- Saliency & distinctiveness:
  - Measures how much information a term conveys about a topic

$$saliency(w) = P(w) \times distinctiveness(w)$$

distinctiveness weighted by the term's overall frequency

$$distinctiveness(w) = \sum_T P(T|w) \log \frac{P(T|w)}{P(T)}$$

computes the KL divergence between the distribution of topics given a term and the marginal distribution of topics

- Applied LDA on elections dataset:
- demo

https://pyldavis.readthedocs.org/en/latest/

Discussion

- The predictive ability of Twitter is questionable mainly because of
  - demographics..
  - trolls… (e.g on election dataset: over representation of #ND. Tweets on #ND not necessarily from ND voters...according to content)
- Yet it is a powerful tool to pinpoint concepts, emerging tendencies and key figures on a social discourse.
- Future work includes:
  - Sentiment analysis
  - How did YES/NO (referendum) users vote at elections?
  - Build an exit poll-like dataset! (a highly predictive subset of users)

- Questions?

despoina@ics.forth.gr, antonakd@gmail.com