

INFLUENCE ANALYSIS ON TWITTER

Master Project in Information and Software
Engineering

70336 – Hugo Manuel Antunes Lopes

Advisor – Prof. Helena Sofia Andrade Nunes Pereira Pinto
Co-Advisor – Prof. Alexandre Paulo Lourenço Francisco

CONTENTS



Hugo Lopes
@hugomalopes

1. Social Networks
 2. Twitter
 3. Hypothesis
 4. Related Work
 5. Our Dataset
 6. Our Goals
 7. Evaluation
 8. Schedule
 9. Conclusions
- #influenceontwitter
#masterproject

1. SOCIAL NETWORKS

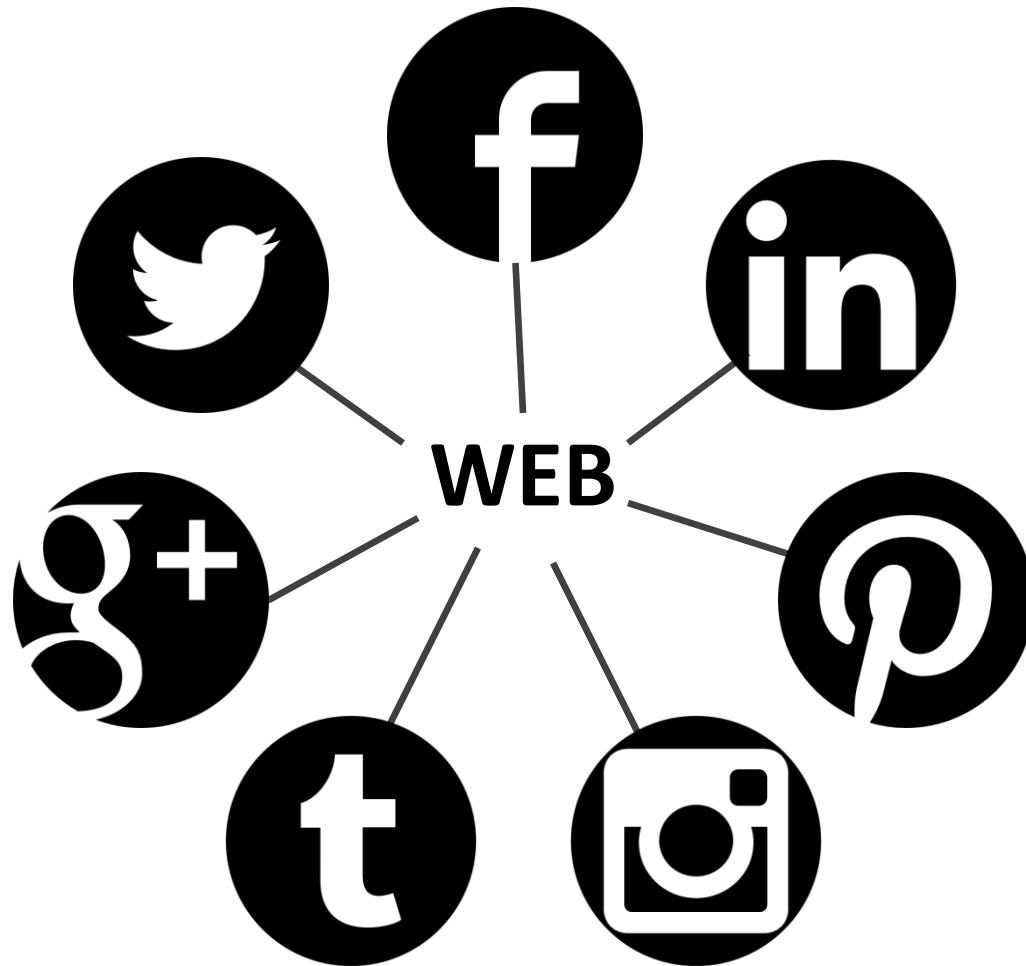
- People are **connected** to each other since they **interact** and **communicate**
- All these ties, that connect Humans, together form **Social Networks**

1. SOCIAL NETWORKS

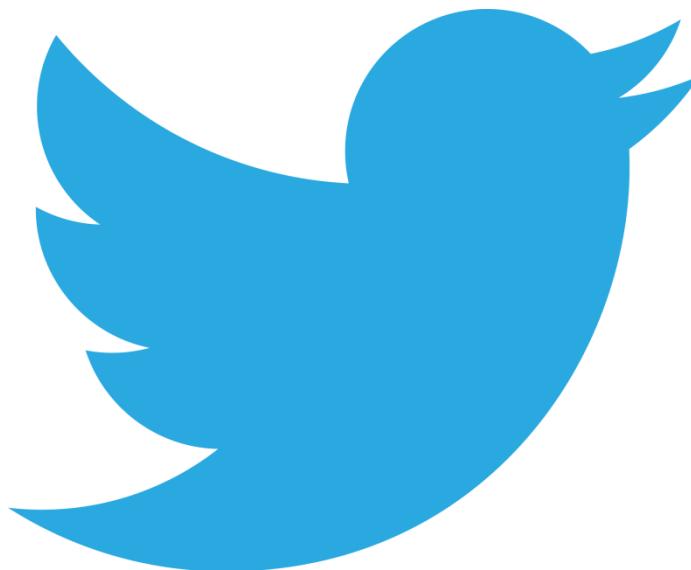
1950's



1. SOCIAL NETWORKS



2. TWITTER



140 characters – Microblogging

2. TWITTER

"RT @zatu_gaku_kun: 【衝撃】童貞の前で美女が裸になった結果ww\n\nいきなり服を脱ぎだす美女が\n\n口すぎると話題にw\n\n「童貞じゃなくとも興奮するww」\n「可愛くてエロいとか反則w」\n\n詳細を見る\n\nhttp://t.co/QuYqQkLuMy http://t..."

"The Game □"

"RT @LifeHelps: A friend is someone who understands your past, believes in your future, and accepts you today just the way you are."

"RT @irvkenM: 【ボッキ注意】ドックダディの娘のおっぱいww\n\nいつのおっぱいデカすぎワロタww\n\n爆乳すぎワロタww」\n「吸い付きてえww」\n「エロすぎ抜いたわww」\n\n詳細を見る\n\nhttp://t.co/QtBW1n3 http://t.co/yUj..."

"@SaRa_xry0he1ngx 行ったんですが治りそうで治らずの繰り返しで..."

"@TayylorWheeler also depends on my work schedule"

"RT @SorteosEnLinea: Increible sorteo de este iphone 6\nIngresá aqui y participa>> http:// .co/IXZOQE2Mb4 http://t.co/iJJJan2Q511"

"Jonathan Emile Ft Kendrick Lamar - Heaven Help Dem http://t.co/7KwwKJL1cK"

"RT @DaddyMoore_: @luxurylaurn should let me have her dog"

"@rokouni ok dalee"

"RT @rapunzel2001: @"napoleonbravo: Escucha a Francisco http://t.co/BiSWMfIR3i""

"RT @zatu_gaku_kun: 【衝撃】童貞の前で美女が裸になった結果ww\n\nいきなり服を脱ぎだす美女が\n\n口すぎると話題にw\n\n「童貞じゃなくとも興奮するww」\n「可愛くてエロいとか反則w」\n\n詳細を見る\n\nhttp://t.co/QuYqQkLuMy http://t..."

"RT @lutepazecatu: 【恐怖映像】森のくまさんに襲われる\n\n自転車でクマから逃げる男性の動画が\n\n怖すぎてやばいと話題にww\n\n「これ絶対ちびるw」\n「怖すぎw」\n\n結構グロいので\n\n覚悟してご覧ください\n\nhttp://t.co/TUgdPNodJv http://..."

"RT @lutepazecatu: 【恐怖映像】森のくまさんに襲われる\n\n自転車でクマから逃げる男性の動画が\n\n怖すぎてやばいと話題にww\n\n「これ絶対ちびるw」\n「怖すぎw」\n\n結構グロいので\n\n覚悟してご覧ください\n\nhttp://t.co/TUgdPNodJv http://..."

"@ziggyu So gracious of you to say, Linda. I hope our paths cross in real life again! Until then, we'll always have the twitterverse."

"@_loco4life phoooooo ヽ(=^・^=)ノ"

"@PlaylistLive Jc or Kian"

100 1203k 0 1203k 0 0 199k 0 ---:---:--- 0:00:06 ---:---:--- 240k"NO DEJES DE APROVECHA
NUESTRAS PROMOS!!!"

"@Y14NN1 @ClarkyHeraldSun yes please!"

2. TWITTER



 **FIFAWorldCup**  @FIFAWorldCup

KICK-OFF: Brazil & Croatia are underway!
The wait is over. #joinin #WorldCup
#BRACRO – fifa.to/1q8Dt2F

RETWEETS 2,444 FAVOURITES 1,959

9:03 pm - 12 Jun 2014



2. TWITTER



A screenshot of a Twitter post from the official account @FIFAWorldCup. The post features a red arrow pointing to the 'Follow' button, which is highlighted with a red circle. The tweet reads: "KICK-OFF: Brazil & Croatia are underway! The wait is over. #joinin #WorldCup #BRACRO – fifa.to/1q8Dt2F". Below the tweet is a photograph of two women cheering at a soccer match. At the bottom, there are engagement metrics: 2,444 retweets and 1,959 favorites, along with a row of profile pictures.

FIFAWorldCup @FIFAWorldCup

KICK-OFF: Brazil & Croatia are underway!
The wait is over. #joinin #WorldCup
#BRACRO – fifa.to/1q8Dt2F

RETWEETS 2,444 FAVOURITES 1,959

9:03 pm - 12 Jun 2014

2. TWITTER



FIFAWorldCup  @FIFAWorldCup

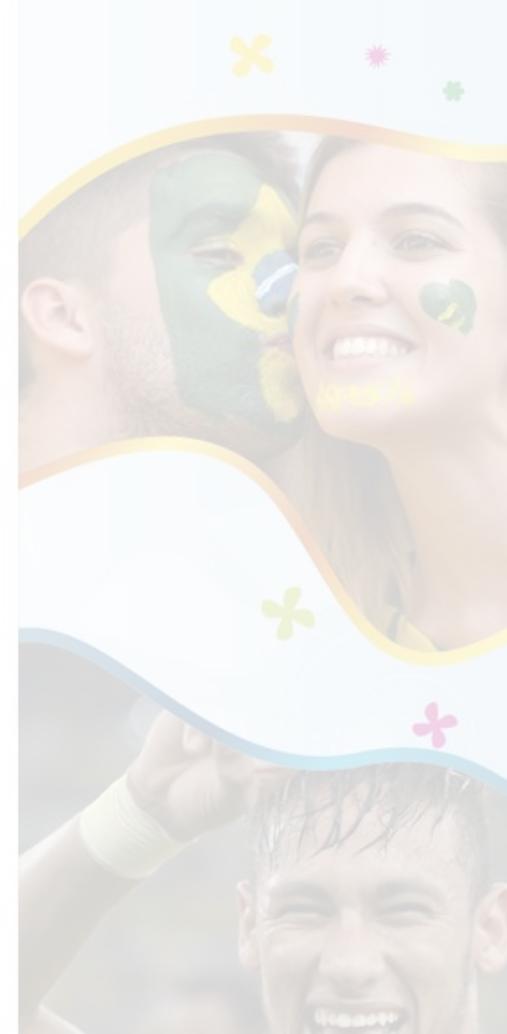
KICK-OFF: Brazil & Croatia are underway!
The wait is over. #joinin #WorldCup
#BRACRO – fifa.to/1q8Dt2F

123 ...

RETWEETS 2,444 FAVORITES 1,959

9:01 pm - 12 Jun 2014

The screenshot shows a tweet from the official FIFA World Cup Twitter account (@FIFAWorldCup). The tweet is about the start of the tournament between Brazil and Croatia. It includes a link to a dedicated landing page. Two specific metrics are highlighted with red circles: the number of retweets (2,444) and favorites (1,959). The timestamp indicates the tweet was posted at 9:01 pm on June 12, 2014.



2. TWITTER



A screenshot of a Twitter post from the official account @FIFAWorldCup. The post is a tweet with a red circle highlighting the profile picture and handle. The text reads: "KICK-OFF: Brazil & Croatia are underway! The wait is over. #joinin #WorldCup #BRACRO – fifa.to/1q8Dt2F". Below the tweet is a photograph of two women smiling and holding up a large Brazilian flag. At the bottom, there are engagement metrics: 2,444 retweets and 1,959 favorites, along with a row of small profile pictures.

FIFAWorldCup
@FIFAWorldCup

KICK-OFF: Brazil & Croatia are underway!
The wait is over. #joinin #WorldCup
#BRACRO – fifa.to/1q8Dt2F

RETWEETS 2,444 FAVOURITES 1,959

9:03 pm - 12 Jun 2014



2. TWITTER



FIFAWorldCup  @FIFAWorldCup

KICK-OFF: Brazil & Croatia are underway!
The wait is over. #joinin #WorldCup
#BRACRO – fifa.to/1q8Dt2F

RETWEETS 2,444 FAVOURITES 1,959

9:03 pm - 12 Jun 2014



2. TWITTER



FIFAWorldCup  @FIFAWorldCup

KICK-OFF: Brazil & Croatia are underway!
The wait is over. Join in! #WorldCup
#BRACRO - fifa.to/1q8Dt2F



RETWEETS 2,444 FAVOURITES 1,959

9:03 pm - 12 Jun 2014



2. TWITTER



FIFAWorldCup  @FIFAWorldCup

KICK-OFF: Brazil & Croatia are underway!
The wait is over. #joinin #WorldCup
#BRACRO - fifa.to/1q8Dt2F



RETWEETS 2,444 FAVOURITES 1,959

9:03 pm - 12 Jun 2014

This block contains a screenshot of a Twitter post from the official FIFA World Cup account (@FIFAWorldCup). The post includes a message about the kick-off between Brazil and Croatia, a call to action with hashtags #joinin and #WorldCup, and a specific hashtag #BRACRO with a link. The image attached to the tweet shows two Brazilian women cheering with a flag. Below the tweet are engagement metrics: 2,444 retweets and 1,959 favorites. The timestamp is 9:03 pm on June 12, 2014.

2. TWITTER



Empire State Bldg @EmpireStateBldg

Congratulations, #GER! Tonight, we'll glow gold, red & black to honor their #WorldCup win. Photo by @ABC.

pic.twitter.com/5M4bQDNM0y

13 Jul



Lukas-Podolski.com

@Podolski10

Selfie 2...:-)

12:00 AM - 14 Jul 2014

[Follow](#)



Alexander Gerst @Astro_Alex · Jul 13

Gratulation vom #ISS an @DFB_Team für Topleistung! Als Experten in Sachen Sterne haben wir schon mal einen besorgt...



2.8K



4.4K



[Expand](#)



Sergio Ramos @SergioRamos · Jul 14

Congratulations bro (@MesutOzil1088) and Khedira for The World Cup champions. Enjoy it ...!



6.8K



6.1K



Kobe Bryant @kobebryant · Jul 13

Congrats #Germany well deserved! Amazing goal to win the #WorldCup #respect



8.5K



8.6K



Elissa @elissakh · Jul 13

Congrats to all German fans! That's the fair play spirit tonight. Mabrouk @CKmusicpro



138



247



Hugh Jackman @RealHughJackman · Jul 13

Congratulations to World Cup 2014 champions @DFB_Team_EN!!! Well deserved. HJ

2. TWITTER

Something is technically wrong.

Thanks for noticing—we're going to fix it up and have things back to normal soon.

Bahasa Indonesia Bahasa Melayu Deutsch English Español Filipino Français Italiano Nederlands

Português Türkçe Русский हिन्दी 日本語 简体中文 繁體中文 한국어

© 2012 Twitter About Help Status

Problems



3. HYPOTHESIS

SENTIMENT & INFLUENCE

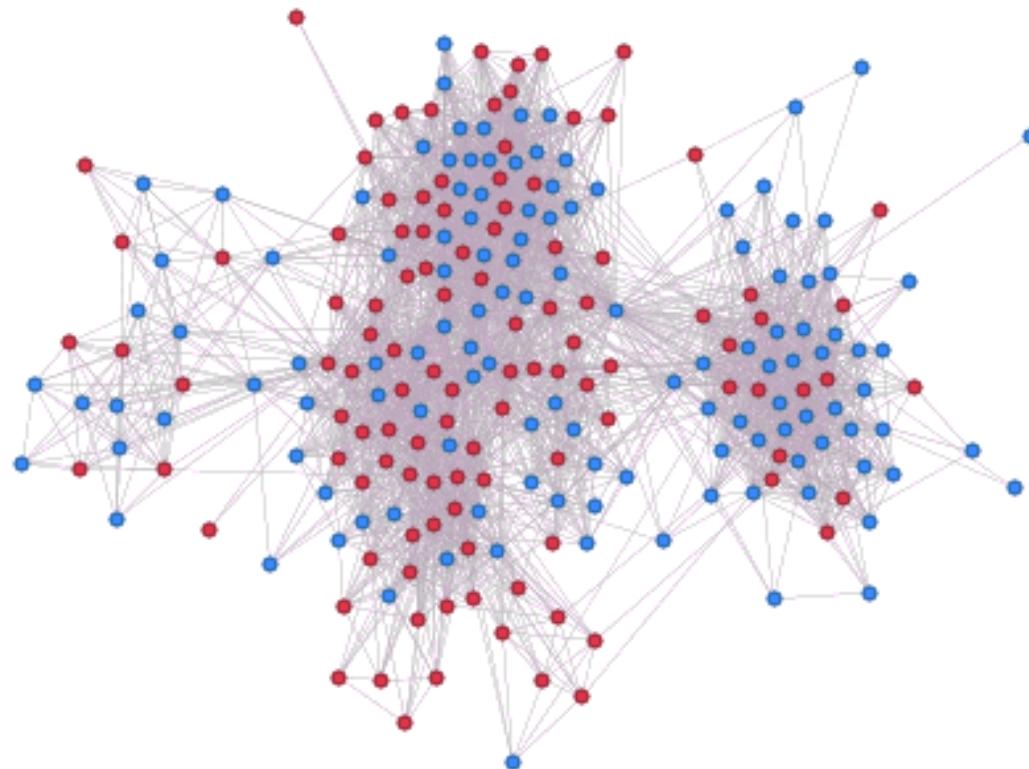
- Can we infer the overall sentiment of a social circle?
- Is people's sentiment influenced by their peers' mood?

4. RELATED WORK

- Easley David and Kleinberg Jon. Networks, Crowds, and Markets: Reasoning About a Highly Connected World.
- S. Wasserman and K. Faust. Social Network Analysis: Methods and Applications. Structural Analysis in the Social Sciences.
- + 66 papers

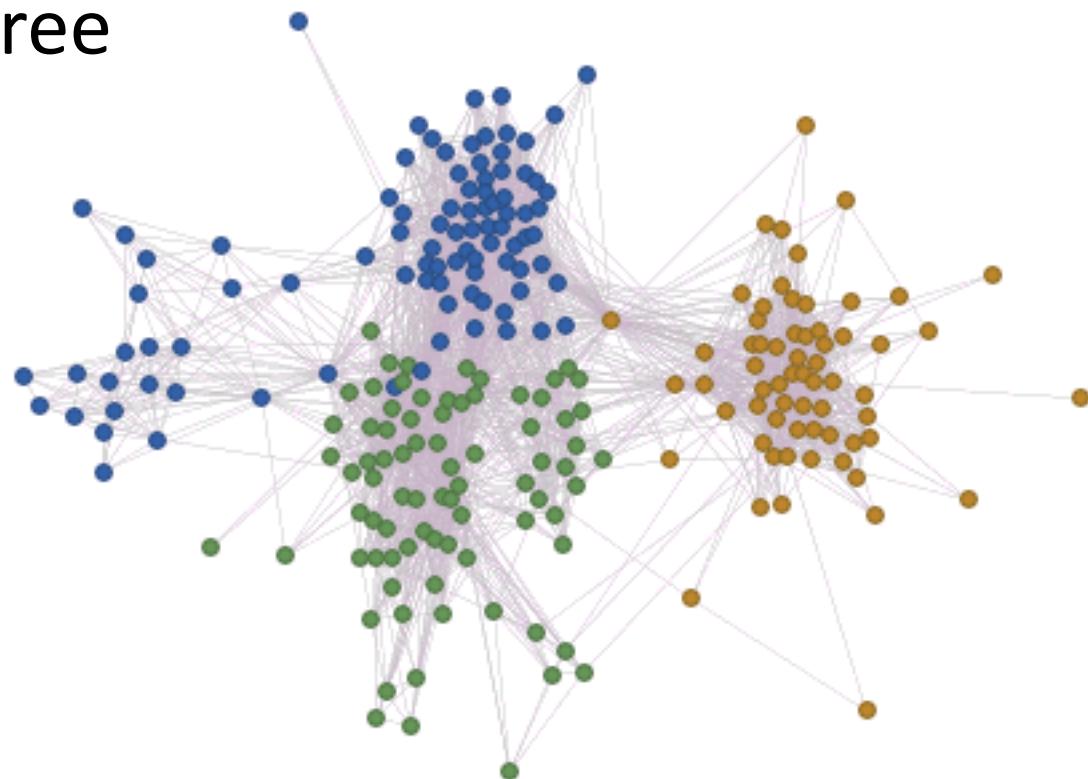
4. GRAPHS

- Nodes
- Links



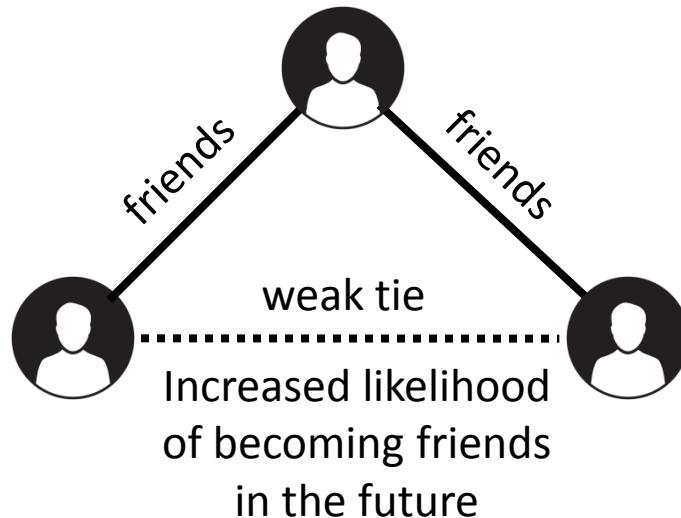
4. GRAPHS

- Connected graphs
- Centrality: degree
- Gatekeeping

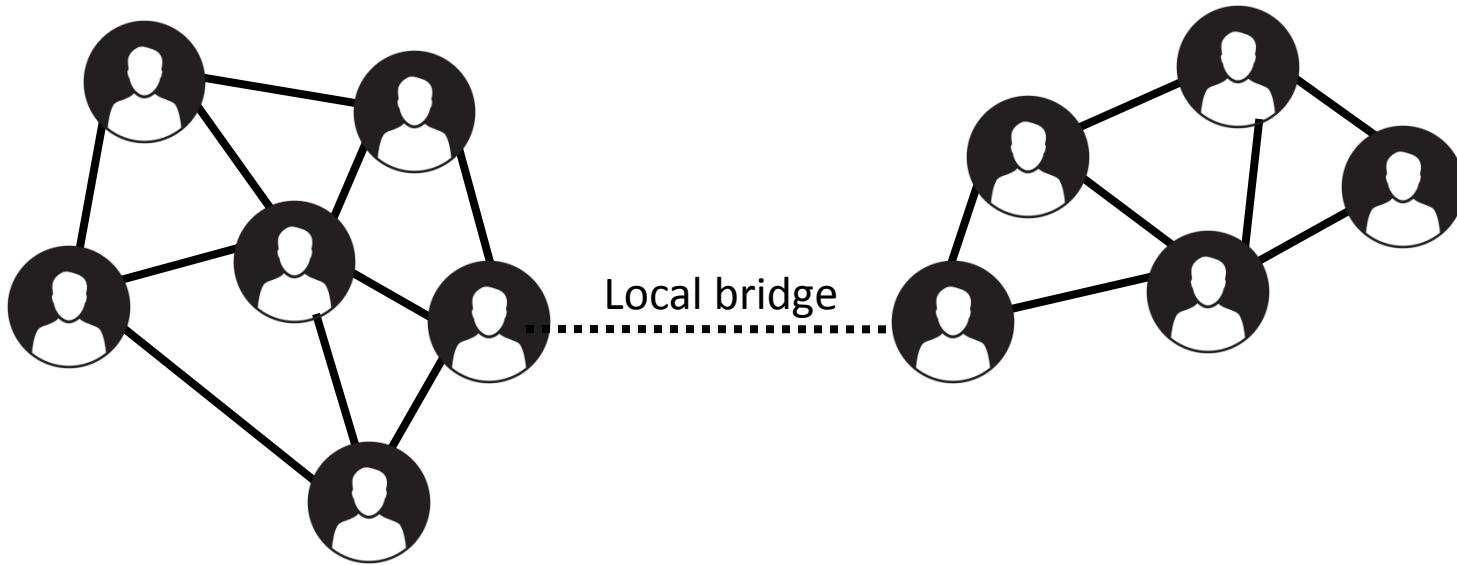


4. STRONG & WEAK TIES

- Friends and acquaintances
- **Triadic Closure**

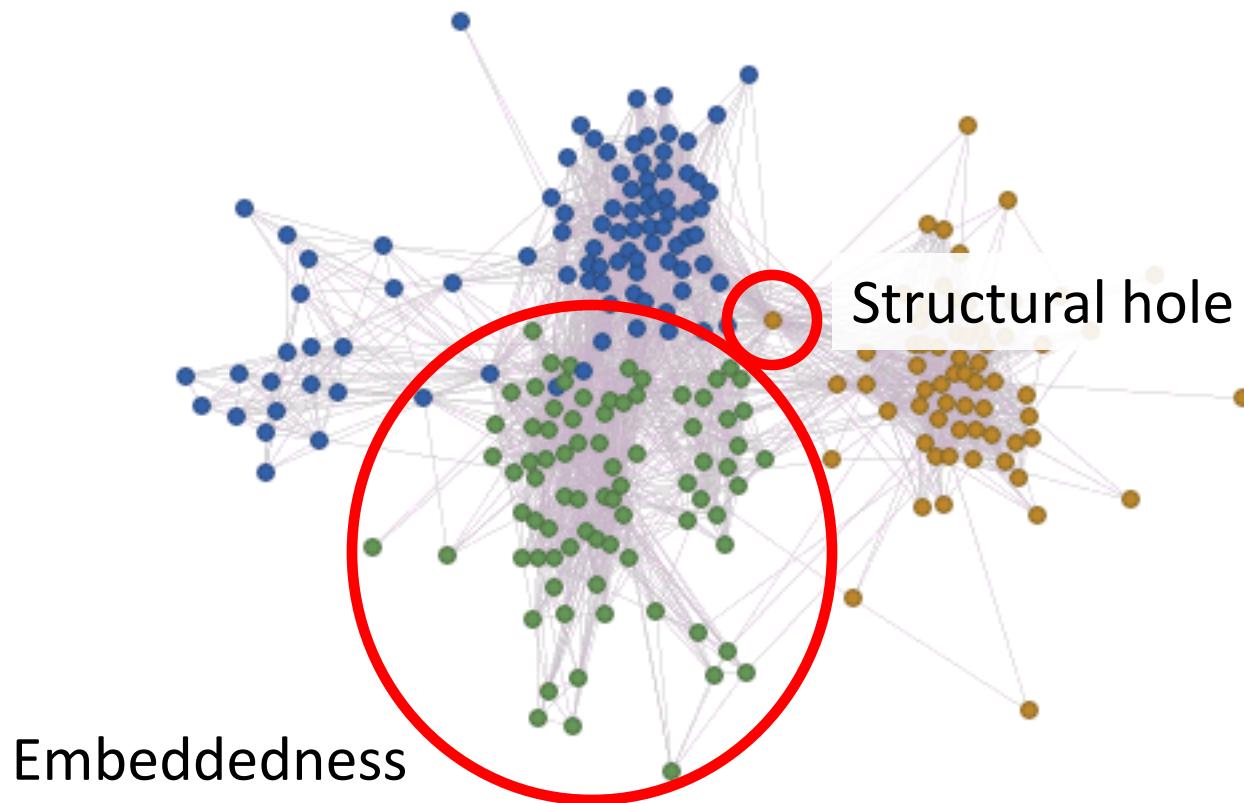


4. STRENGTH OF WEAK TIES



$$\text{neighborhood overlap of edge}_{AB} = \frac{n^o \text{ nodes neighbors of both } A \text{ and } B}{n^o \text{ nodes neighbors of at least 1 of } A \text{ and } B}$$

4. STRUCTURAL HOLE & EMBEDDEDNESS

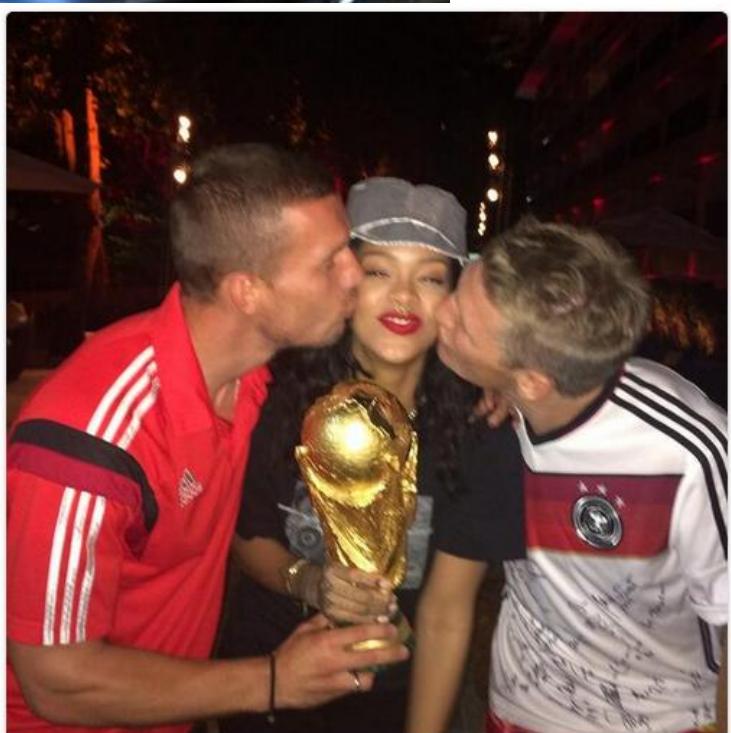


4. HOMOPHILY

- We tend to be similar to our friends



4. SOCIAL INFLUENCE & SELECTION



 **Rihanna** 
@rihanna

Podolski x RiRi x Shweinsteiger

7:21 AM - 14 Jul 2014

52,417 RETWEETS 68,187 FAVORITES

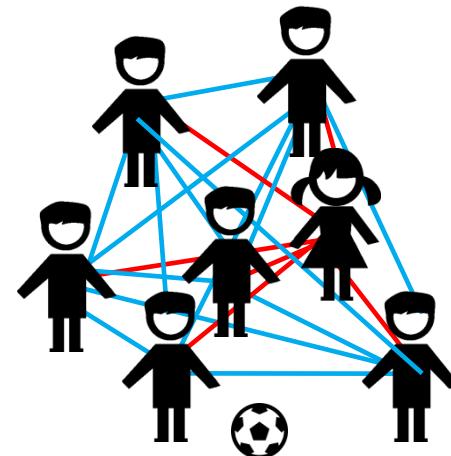
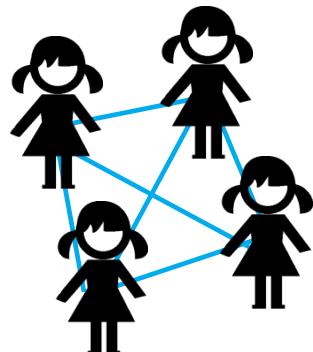
4. HOMOPHILY TEST

Elementary-school class – Gender Homophily

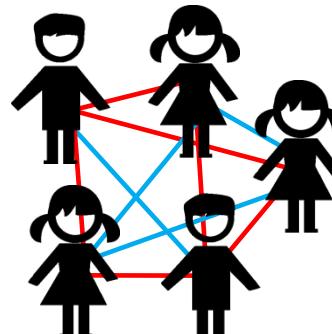
- 16 children
- $P_m = P_f = \frac{8}{16}$
- cross-gender ratio = $2 * P_m * P_f = \frac{1}{2}$

$$\frac{12}{37} \ll \frac{1}{2}$$

Homophily

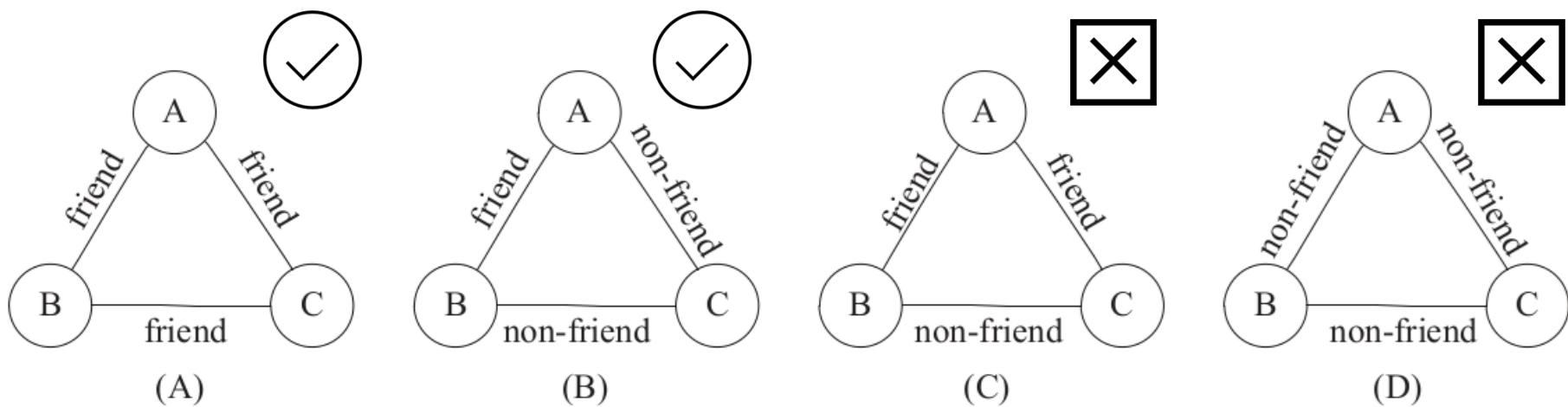


same-gender
cross-gender



4. POSITIVE & NEGATIVE RELATIONSHIPS

- **Structural Balance:**



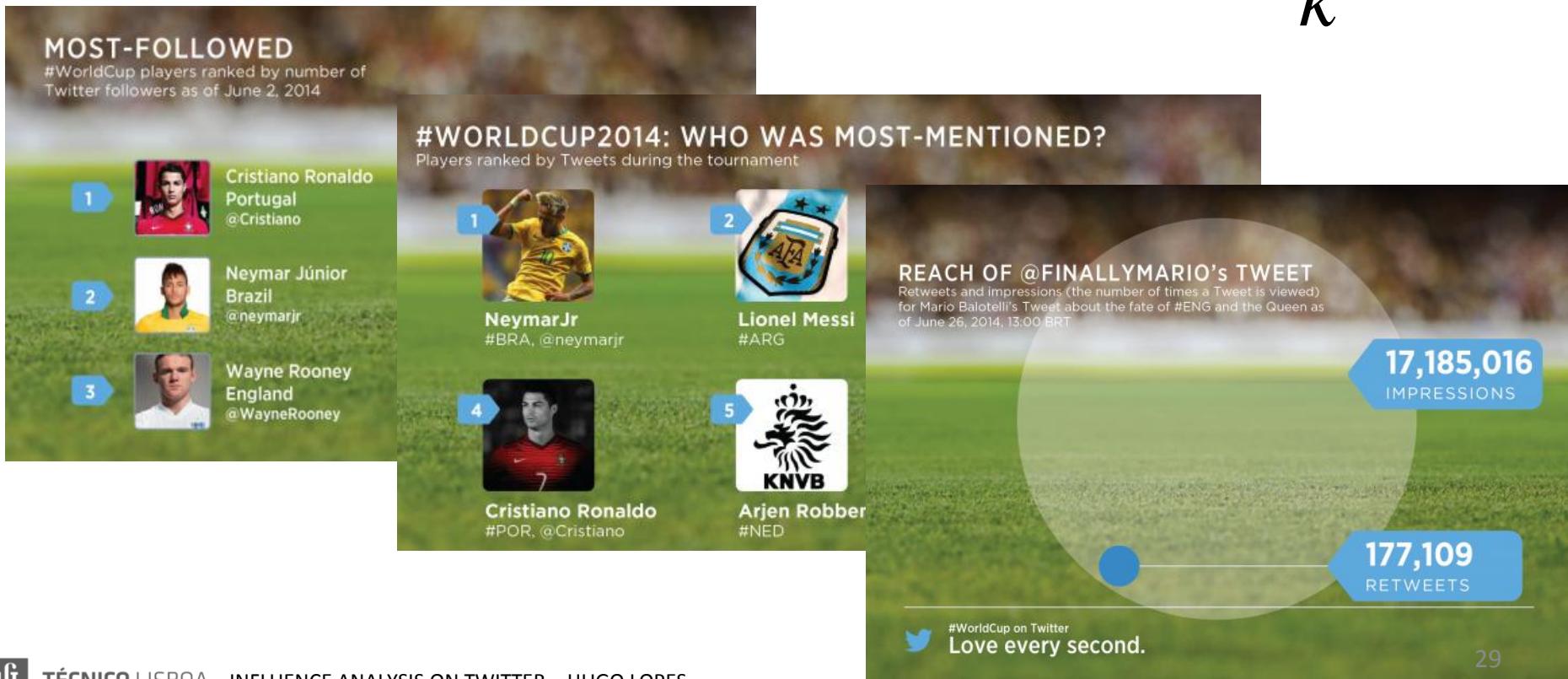
4. INFLUENCE AND CASCADES

- Influence: **imitating the choices of others**
- Information cascades: **abandoning own information** in favor of **inferences** based on earlier actions from others
- Cascading behavior: **direct payoffs from copying the decisions of others**

4. POPULARITY

- Number of in-links
- *Power-law* distribution

$$f(k) = \frac{a}{k^c}$$



5. OUR DATASET

- “672M tweets”, *Twitter Blog*
- Topic – 2014 FIFA World Cup in Brazil
- From March 13 to July 15 of 2014
- 340 707 851 tweets
- 419 zipped files
- 166GB

6. OUR GOALS

- Infer User-User interactions' Network



"WORLD CUP!!!"



"RT @kevinbridges86: Brazil need a goal in the first 10 minutes or the whole country gets torched. Pressure game."



"@World Let the football talk."



"I want those Messi cleats"



"Because you showed us patriotism doesn't need background



music:\nÔ brigado brasil! It was beautiful #WorldCup2014 #brazil
#Brasil2014"



"@jacobruiiz I thought you hated soccer?"



"the perfect month long distraction #WorldCup2014"



"Como están los croatas papaaaaa!!! #Brasil2014 #BRA #CRO"



"Watching the World Cup at work"



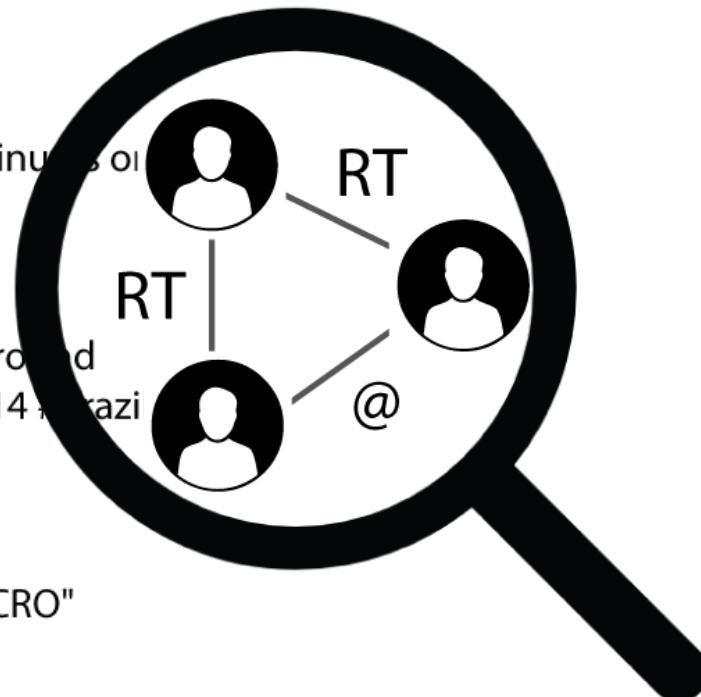
"Come on Brazil.. #WC2014 #BRAvsCRO"

.....

6. OUR GOALS

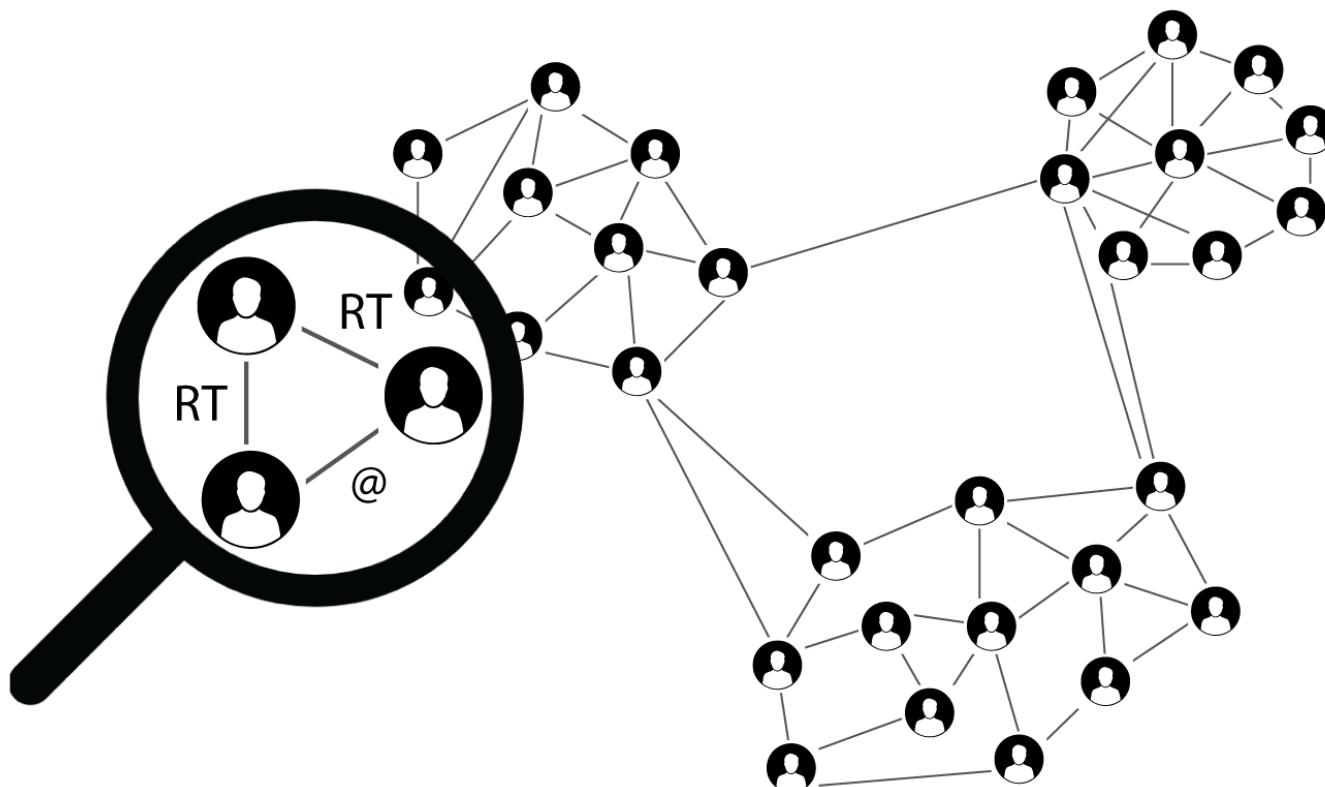
- Infer User-User interactions' Network

- "WORLD CUP!!!"
- "RT @kevinbridges86: Brazil need a goal in the first 10 minutes or the whole country gets torched. Pressure game."
- "@World Let the football talk."
- "I want those Messi cleats"
- "Because you showed us patriotism doesn't need background music:\nÔ brigado brasil! It was beautiful #WorldCup2014 #Brasil #Brasil2014"
- "@jacobruiiz I thought you hated soccer?"
- "the perfect month long distraction #WorldCup2014"
- "Como están los croatas papaaaaa!!! #Brasil2014 #BRA #CRO"
- "Watching the World Cup at work"
- "Come on Brazil.. #WC2014 #BRAvsCRO"
-



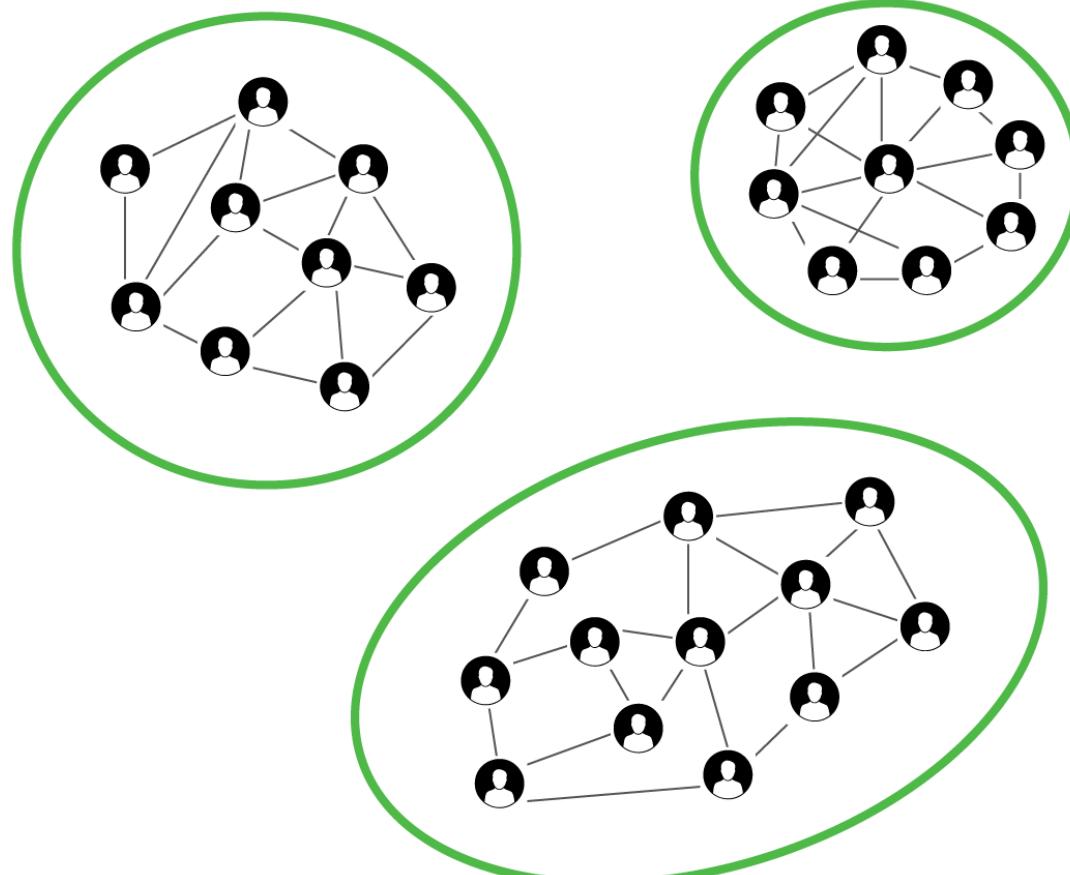
6. OUR GOALS

- Infer User-User interactions' Network



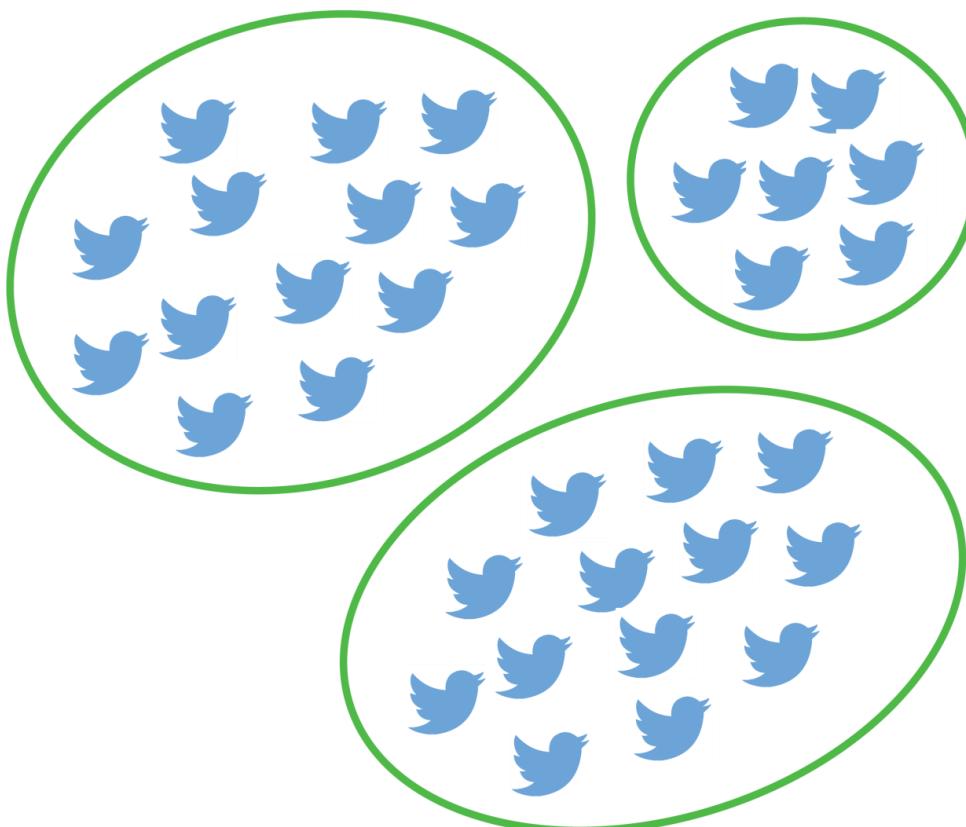
6. OUR GOALS

- Find Communities' Clusters



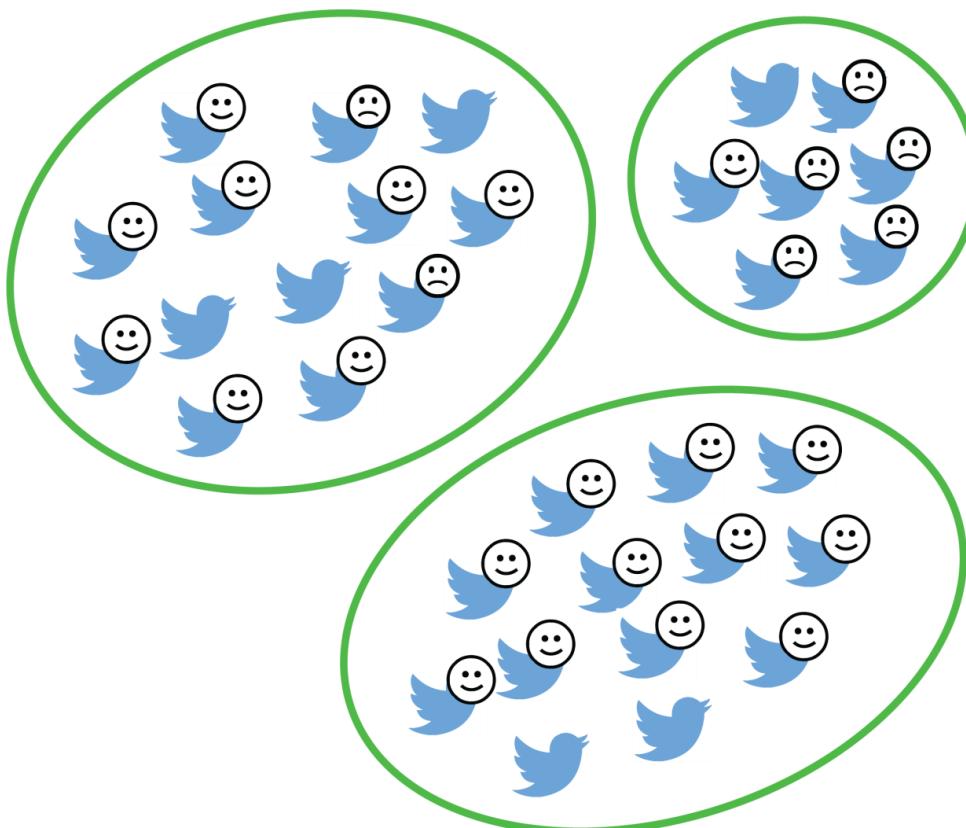
6. OUR GOALS

- Group tweets by Cluster



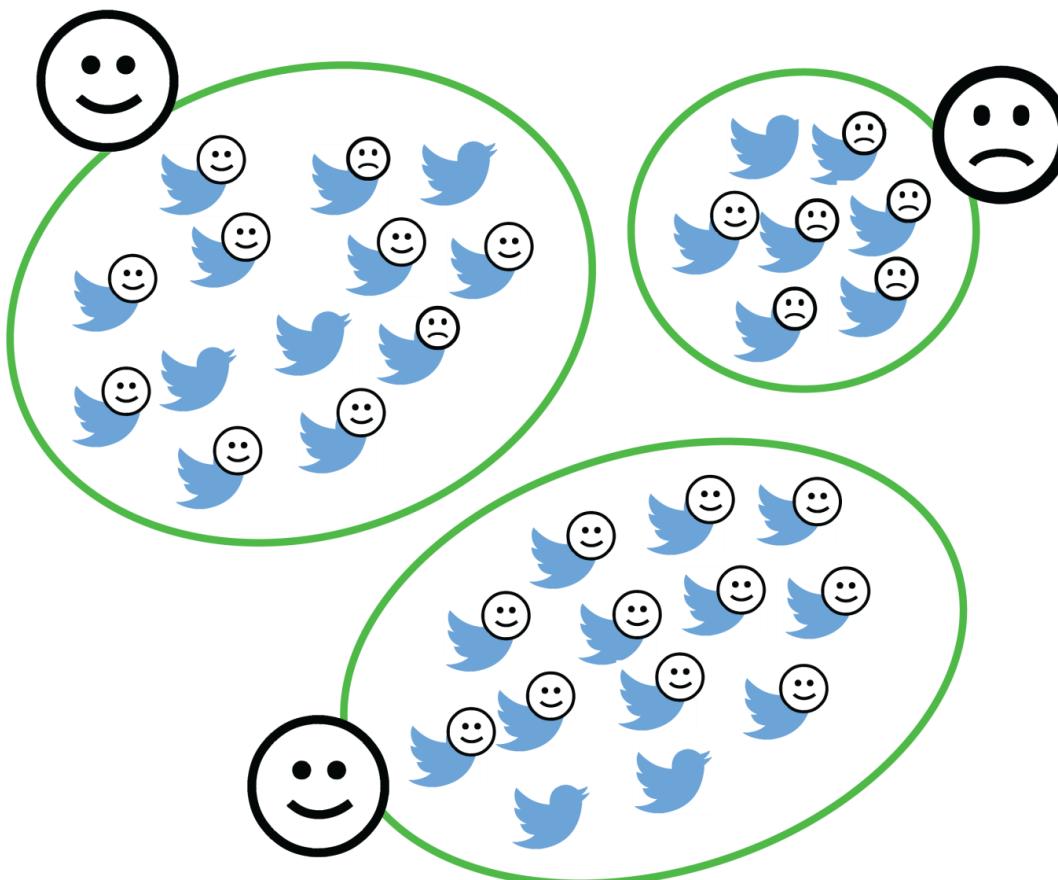
6. OUR GOALS

- Classify Clusters' Sentiment: tweet-by-tweet



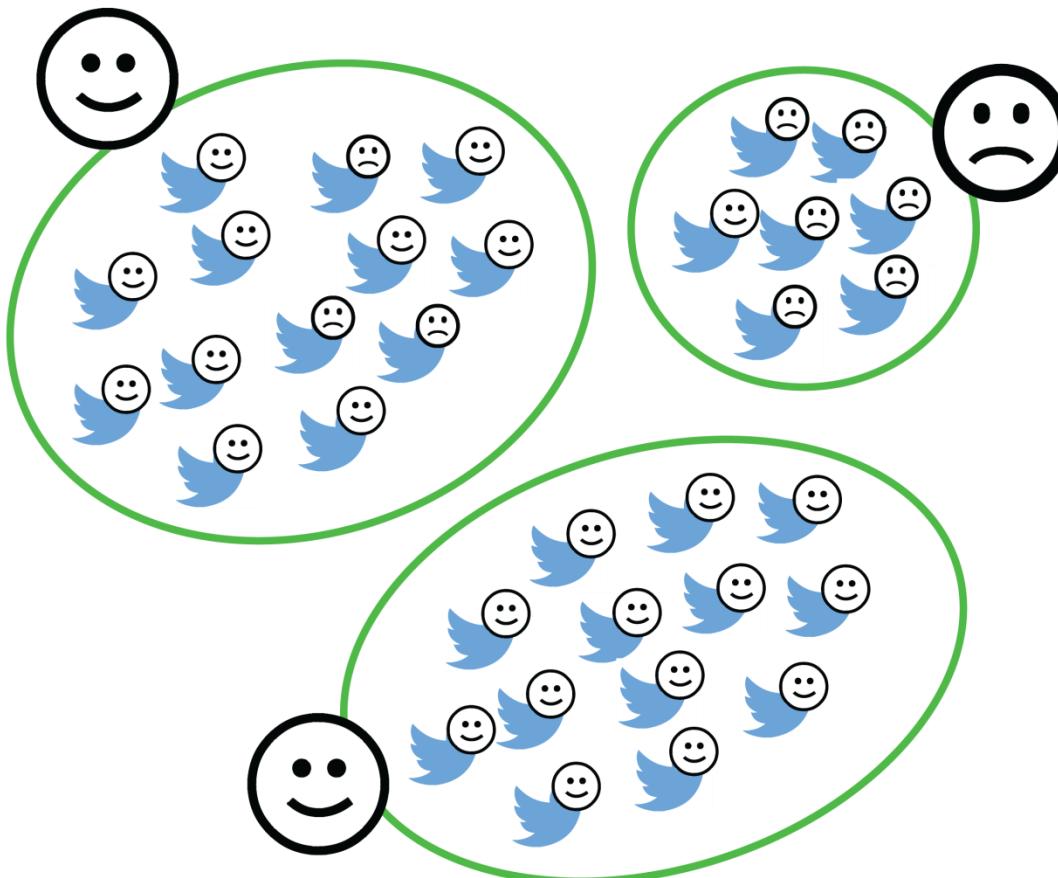
6. OUR GOALS

- Classify Clusters' Sentiment: tweet-by-tweet



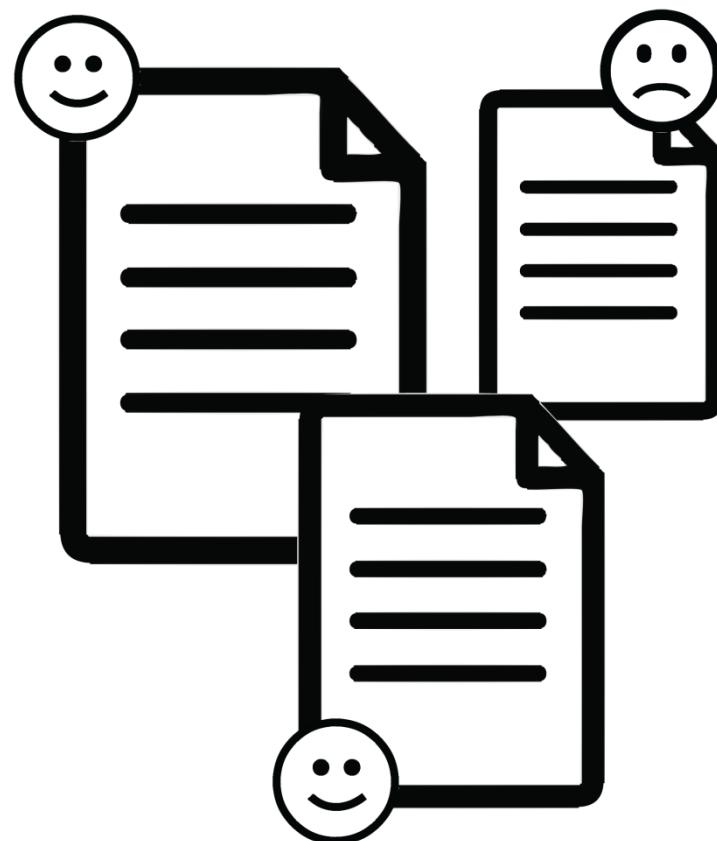
6. OUR GOALS

- Extrapolate Clusters' Sentiment



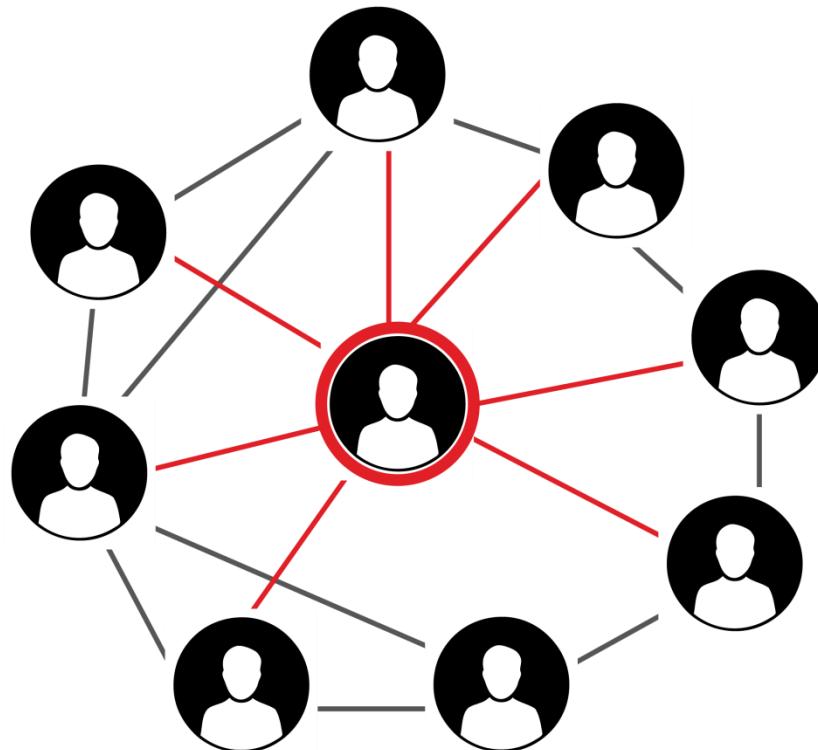
6. OUR GOALS

- Classify Clusters' Sentiment: documents



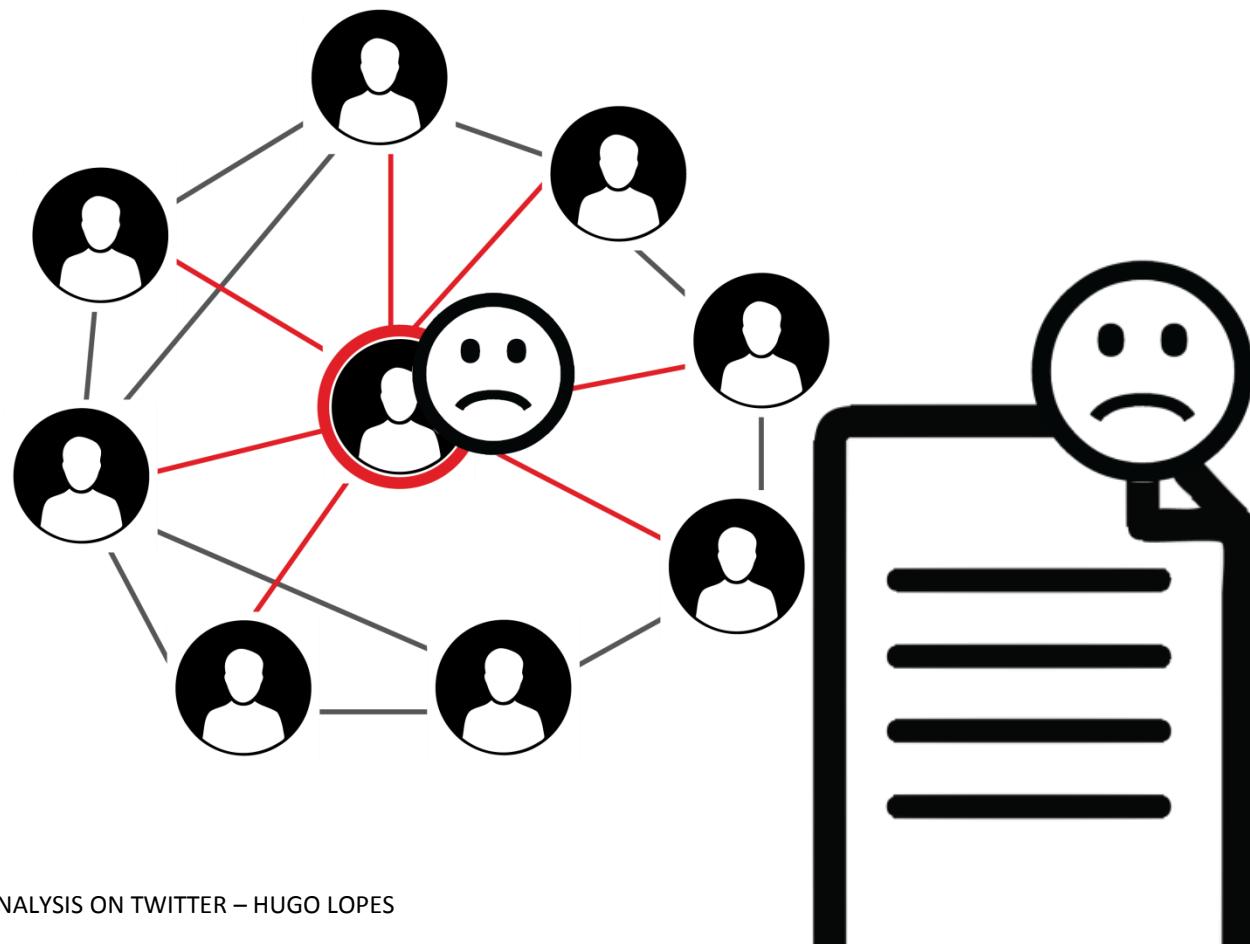
6. OUR GOALS

- Find Influent Actors



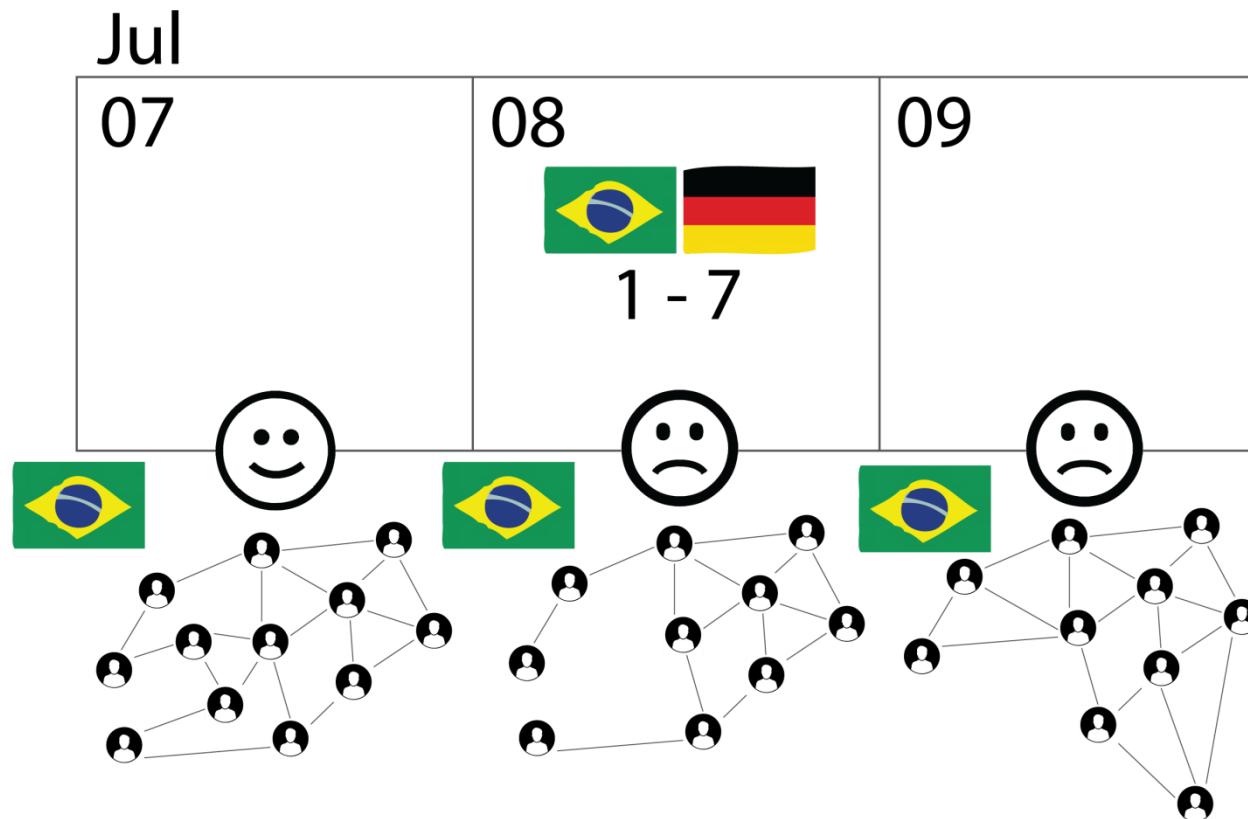
6. OUR GOALS

- Find Influent Actors and their Sentiment



6. OUR GOALS

- Dynamic Network and Sentiment Changes



7. EVALUATION

- Manual Validation
 - n random samples
 - 3 human evaluators VS tool under test, per sample
- K-fold Cross-Validation
 - k subsets, k iterations
 - $k-1$ classify
 - 1 validates

8. PRELIMINARY ANALYSIS

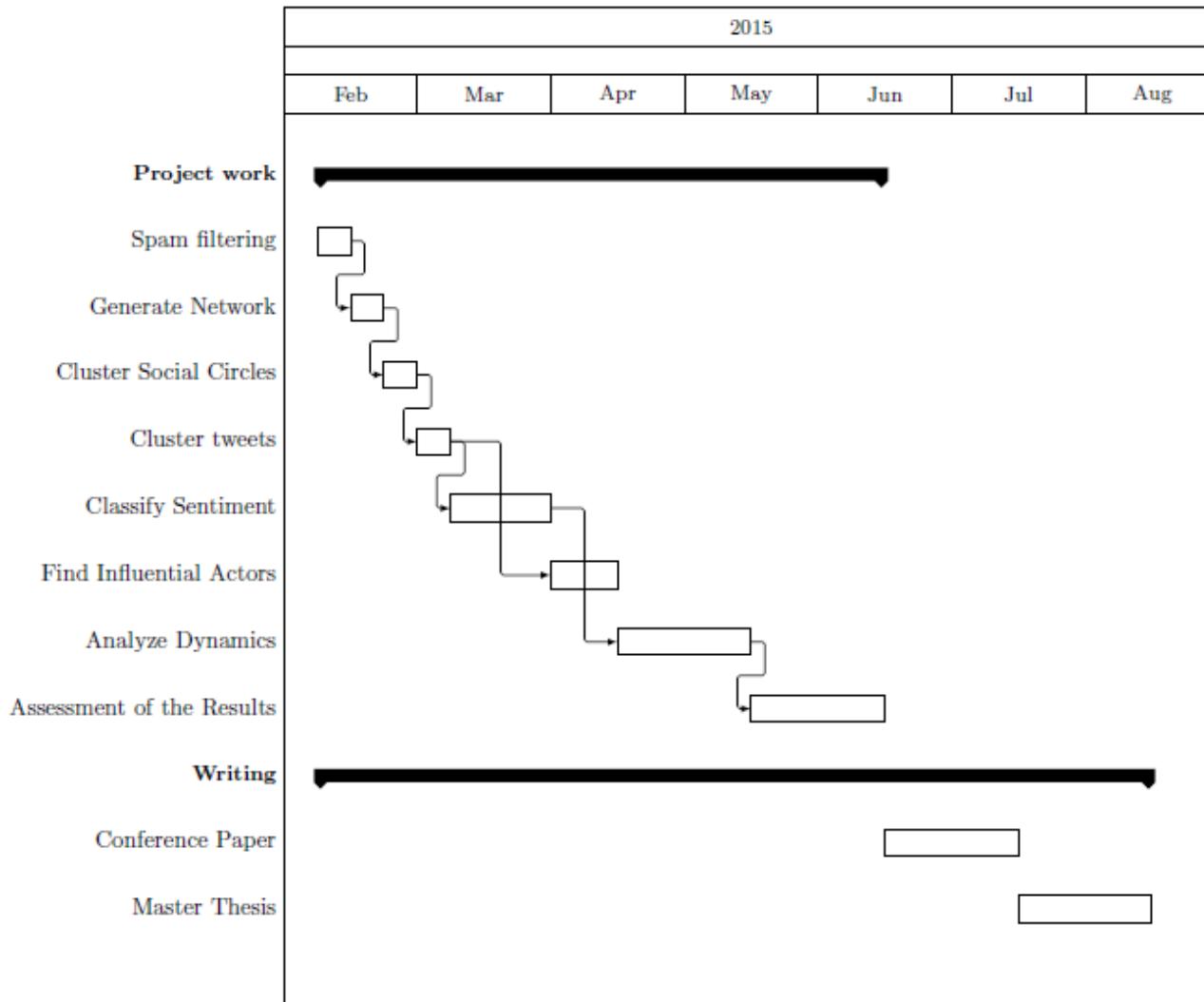
Total: 340 707 851

- Simple: 41%
- RT: 51%
- RE: 7%
- URL: 28%
- Urls only: **9 tweets!** Almost no spam

Languages:

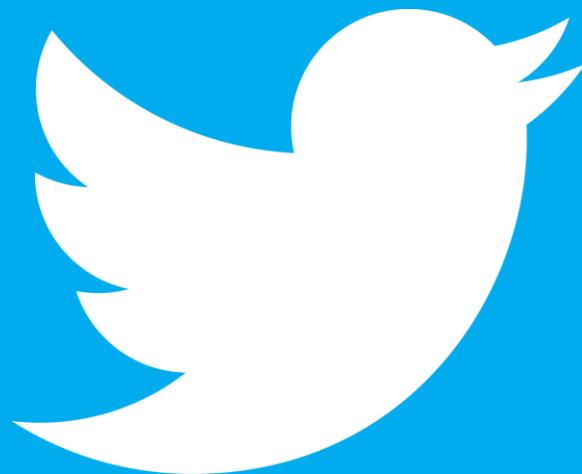
- 191 538 627 - en
- 75 174 800 - es
- 18 295 251 - pt
- 15 208 450 - fr
- 10 476 957 - id
- 6 430 021 - ja

9. SCHEDULE



10. CONCLUSIONS

- Obtain high-level sentiment of communities from aggregated information of individuals
 - Homophily, Structural Balance and Influence
- Understand polarity changes throughout the World Cup



THANK YOU!

PAPERS & TOPICS

PAPER	NETWORK STRUCTURE	FINDING COMMUNITIES	SENTIMENT ANALYSIS	INFLUENCE	DYNAMIC ANALYSIS
Jiliang Tang, Yi Chang, and Huan Liu. Mining social media with social theories: A survey.	X				
Meeyoung Cha, Hamed Haddadi, Fabricio Benevenuto, and Krishna P. Gummadi. Measuring user influence in twitter: The million follower fallacy	X			X	
Justin Cheng, Lada Adamic, P. Alex Dow, Jon Michael Kleinberg, and Jure Leskovec. Can cascades be predicted?	X				
P. Boldi and S. Vigna. The webgraph framework I: Compression techniques.	X				
Paolo Boldi, Massimo Santini, and Sebastiano Vigna. Pagerank: Functional dependencies.		X			
Amine Abou-Rjeili and George Karypis. Multilevel algorithms for partitioning power-law graphs.		X			
Santo Fortunato. Community detection in graphs.		X			
Mike Thelwall and Kevan Buckley. Topic-based sentiment analysis for the social web: The role of mood and issue-related words.			X		
Mike Thelwall, Kevan Buckley, and Georgios Paltoglou. Sentiment in twitter events.			X		
Xia Hu, Lei Tang, Jiliang Tang, and Huan Liu. Exploiting social relations for sentiment analysis in microblogging.		X	X		
Jun Ito, Takahide Hoshida, Hiroyuki Toda, Tadasu Uchiyama, and Kyosuke Nishida. What is he/she like?: Estimating twitter user attributes from contents and social neighbors.				X	
Jeffrey Nichols, Jalal Mahmud, and Clemens Drews. Summarizing sporting events using twitter					X
Adam Bermingham, Maura Conway, Lisa McInerney, Neil O'Hare, and Alan F. Smeaton. Combining social network analysis and sentiment analysis to explore the potential for online radicalisation	X		X		X
Xia Hu, Lei Tang, Jiliang Tang, and Huan Liu. Exploiting social relations for sentiment analysis in microblogging		X	X		
Jun Ito, Takahide Hoshida, Hiroyuki Toda, Tadasu Uchiyama, and Kyosuke Nishida. What is he/she like?: Estimating twitter user attributes from contents and social neighbors		X		X	
VA Traag and Jeroen Bruggeman. Community detection in networks with positive and negative links.	X	X			

RESEARCH ON TWITTER

- *Sentiment Analysis*
- *Spam Filtering*
- *Geo-location*
- Network Structure and Finding Communities
- Event Detection
- Event Prediction
- Information Flow
- Influence