Observing real world on Twitter

Ossi Karkulahti
Joint work with Jussi Kangasharju and Lasse Nordgren
Department of Computer Science
University of Helsinki

Outline

- Introduction

- Tale of two cities

- Topical results

- Conclusion & future

Introduction

- ## Twitter

  - Twitter is a real-time information network powered by people all around the world that lets you share and discover what's happening now. Twitter asks "what's happening" and makes the answer spread across the globe to millions, immediately.

- ## Tweet

  - A tweet is a post or status update on Twitter. The maximum size of a tweet is 140 characters.

Introduction

- Motivation

  - To understand better the reasons of users to create tweets, and see if the reasons correspond to real-life situations

  - Adaptive content distribution

Introduction

- We have collected more than 5 million different tweets, by using two different methods:

  – **Topical keywords**, such as "H1N1" and "Olympics"

  – Based on the **location** of the users, e.g. Liverpool, Madrid, Rome etc.

    - Indicated in the profile or by geotag

Tale of two cities

- During Jan - April 2010 we have gathered tweets from two cities:

  - Liverpool, UK (~3.2 million tweets)

  - Madrid, the capital of Spain (~3.4)
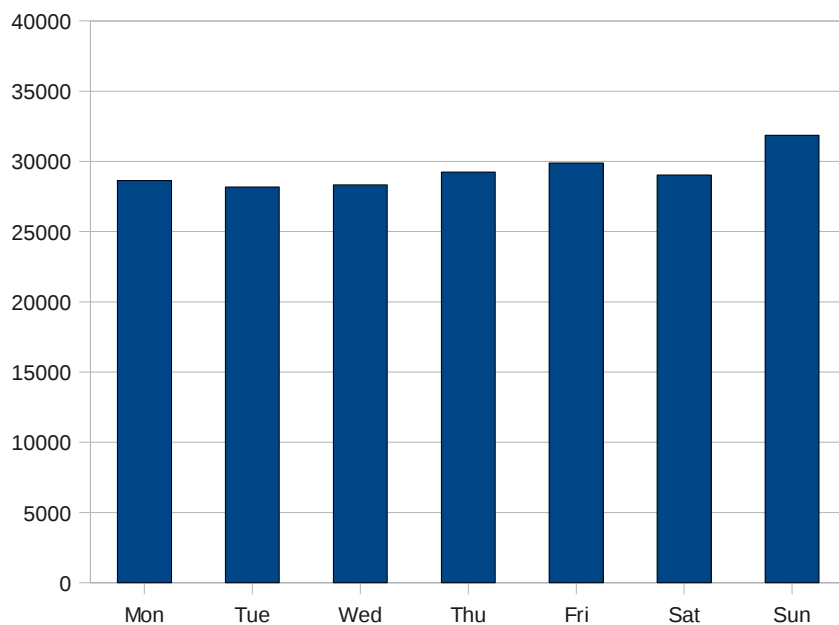
- Location-based method

Tale of two cities

- Results:
  - Daily pattern
  - Hourly pattern
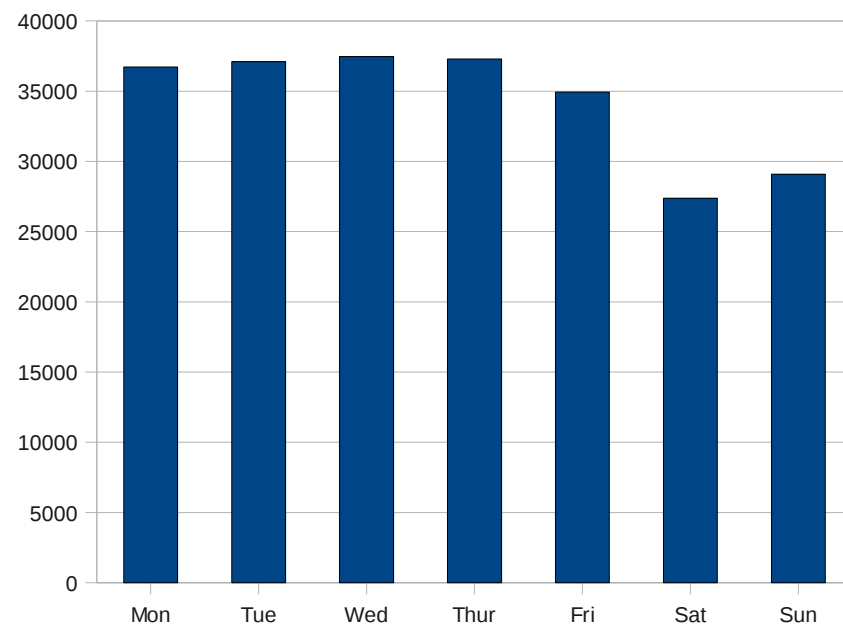 - Incidents
  - Statistics
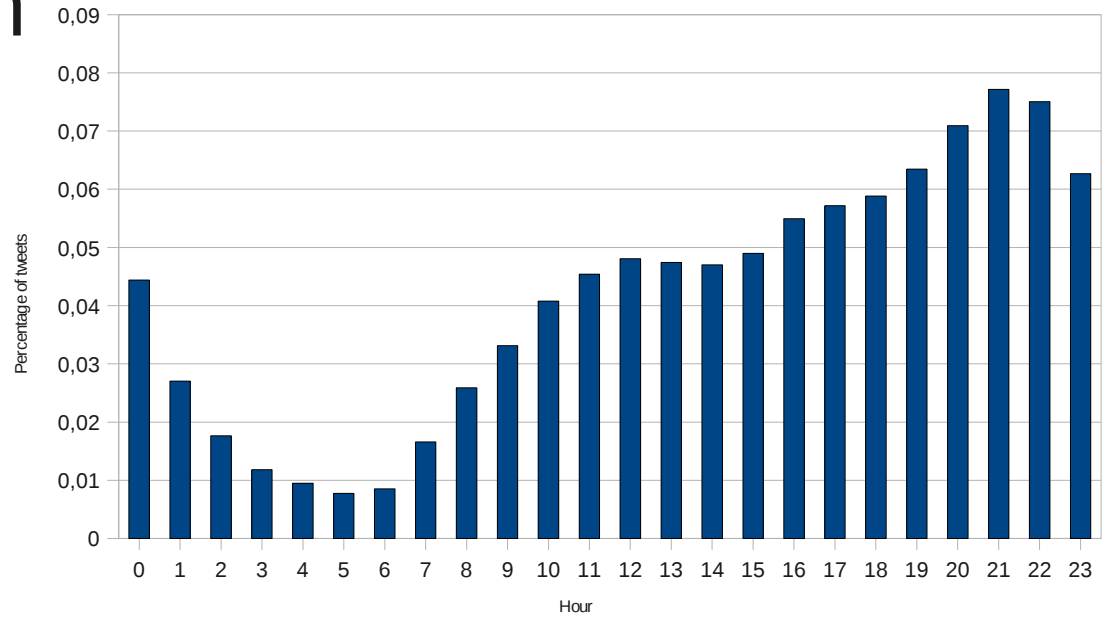
Daily Pattern

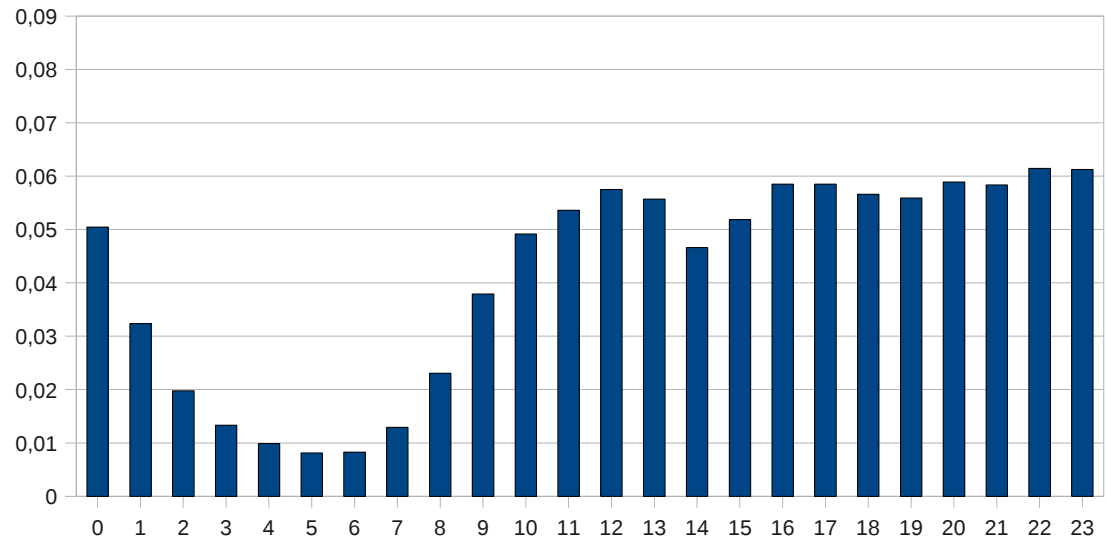## Liverpool

Average per day 29307

Madrid

Average per day 34281

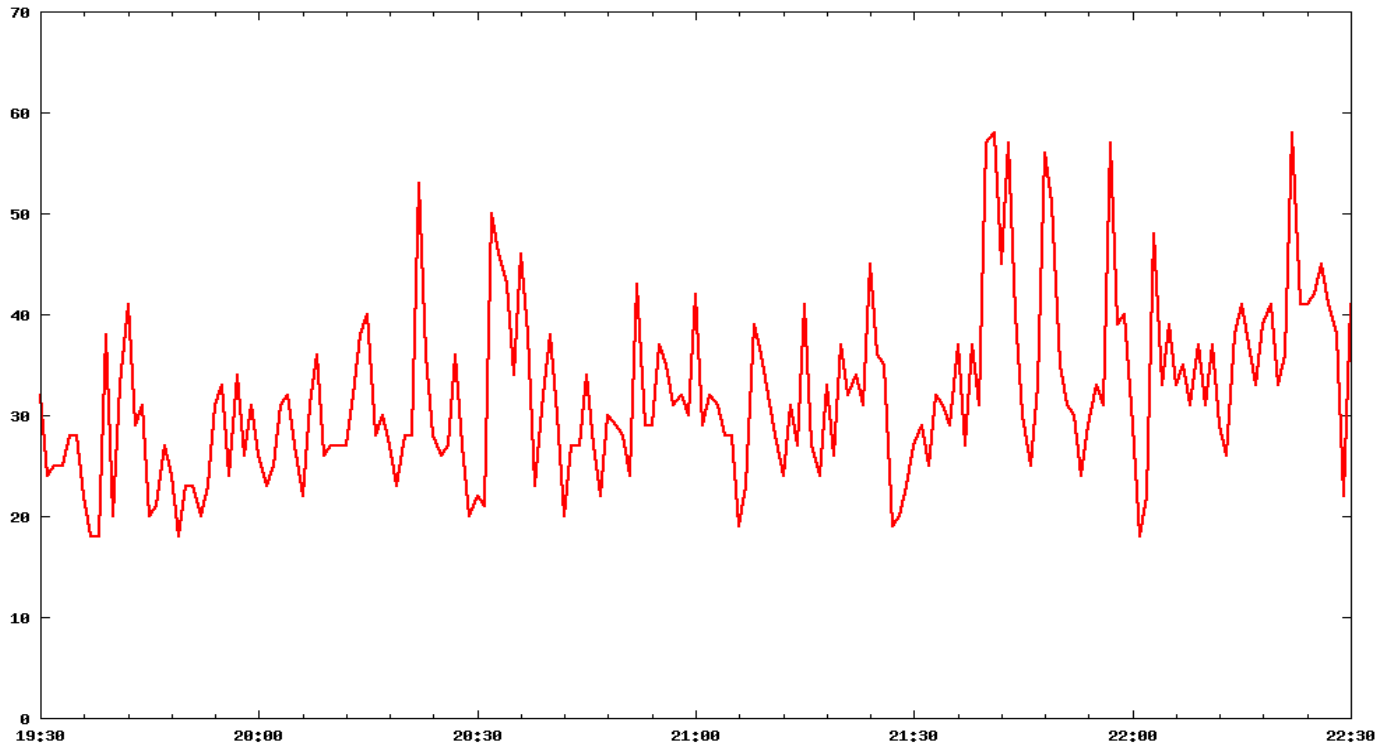# Hourly Pattern

Liverpool

Madrid

Liverpool

Liverpool-Reading 1-2, FA Cup, 13th January 2010

1945 GMT: kick-off
2031: GOAL 1-0
2033: Half-time 1-0
~2050: 2nd half kick-off
2139: Penalty to Reading
2140: GOAL 1-1
2142: End of 90 mins 1-1
2147: Extra-time begins
2156: GOAL 1-2
2203: Half-time in extra-time 1-2
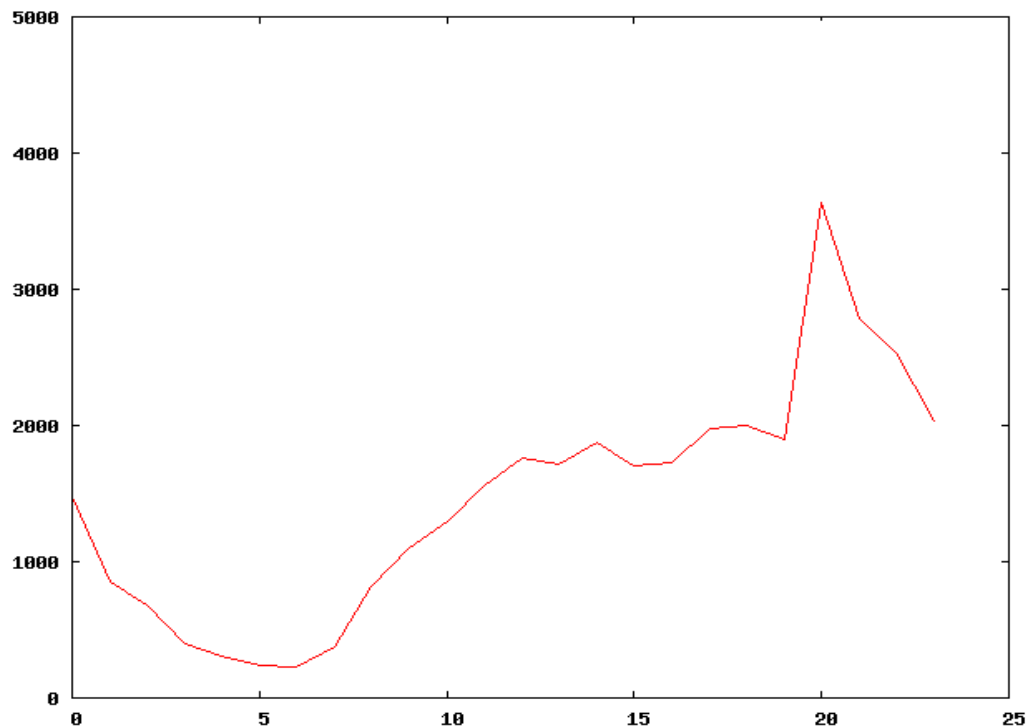2221: Full-time in extra-time 1-2
2240: Interviews

Liverpool

The Brit Awards, February 16th, 8pm ->

Most common words during 20-23 pm:

Rank		Count
2	brits	919
17	gaga	336
20	peter	286
21	robbie	285
36	cheryl	216
37	brit	213
45	lady	186
46	awards	183
56	liam	160
61	award	147
69	florence	139
78	music	123
89	williams	112
106	cole	98
108	gallagher	94
120	alicia	88
129	kasabian	84
137	ladygaga	79
152	dizzee	70

Madrid

2000 GMT: kick-off
2032: GOAL 0-1
2047: Half-time 0-1
~2102: 2nd half kick-off
2112: GOAL 0-2
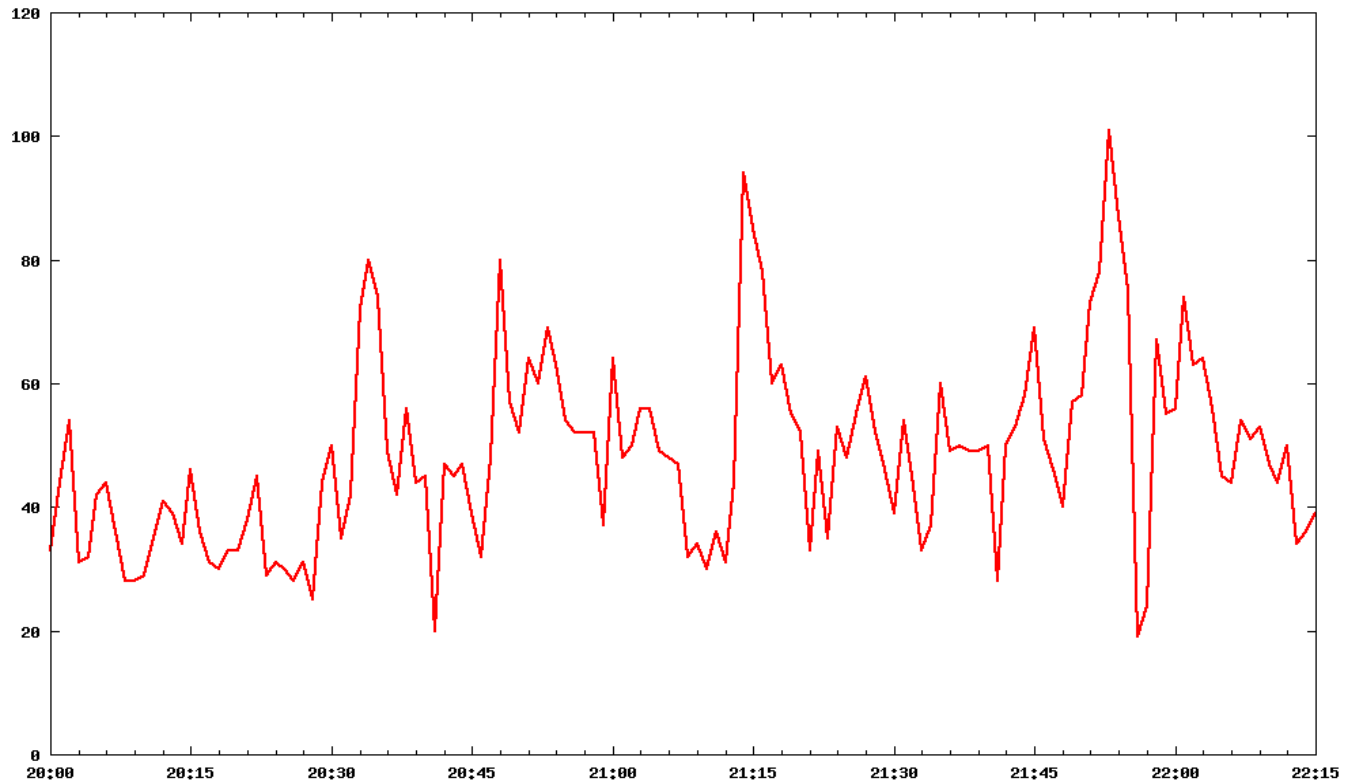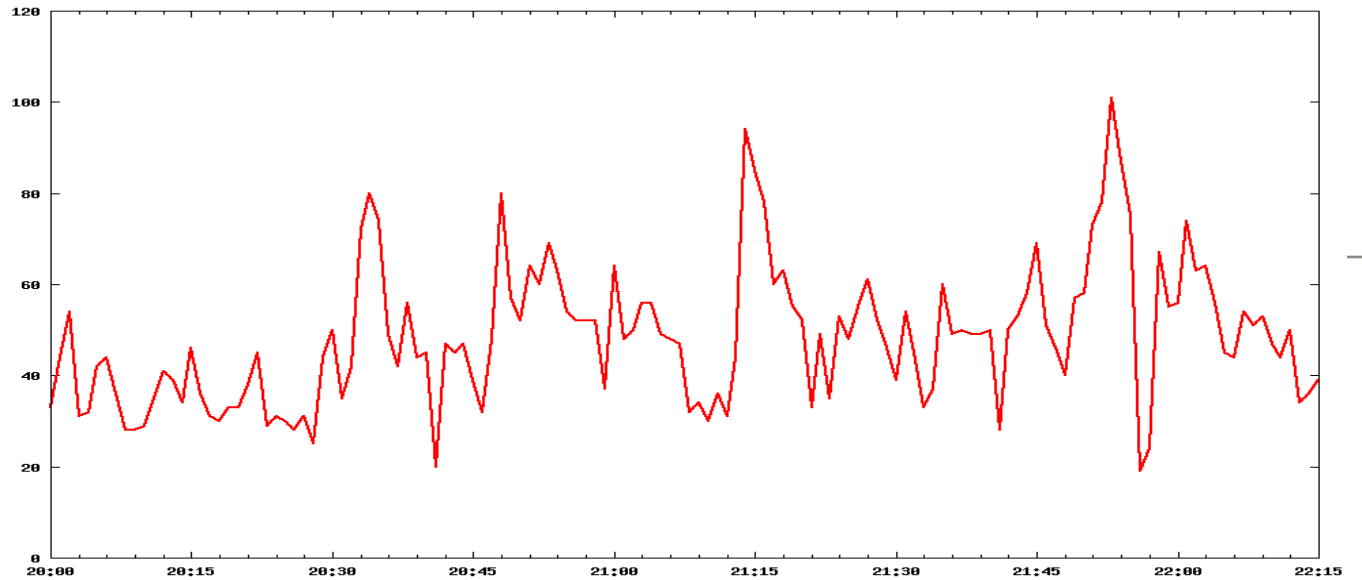2150: Full-time 0-2

Real Madrid-Barcelona 0-2, La Liga, April 10th

Madrid:

2000 GMT: kick-off
2032: GOAL 0-1
2047: Half-time 0-1
~2102: 2nd half kick-off
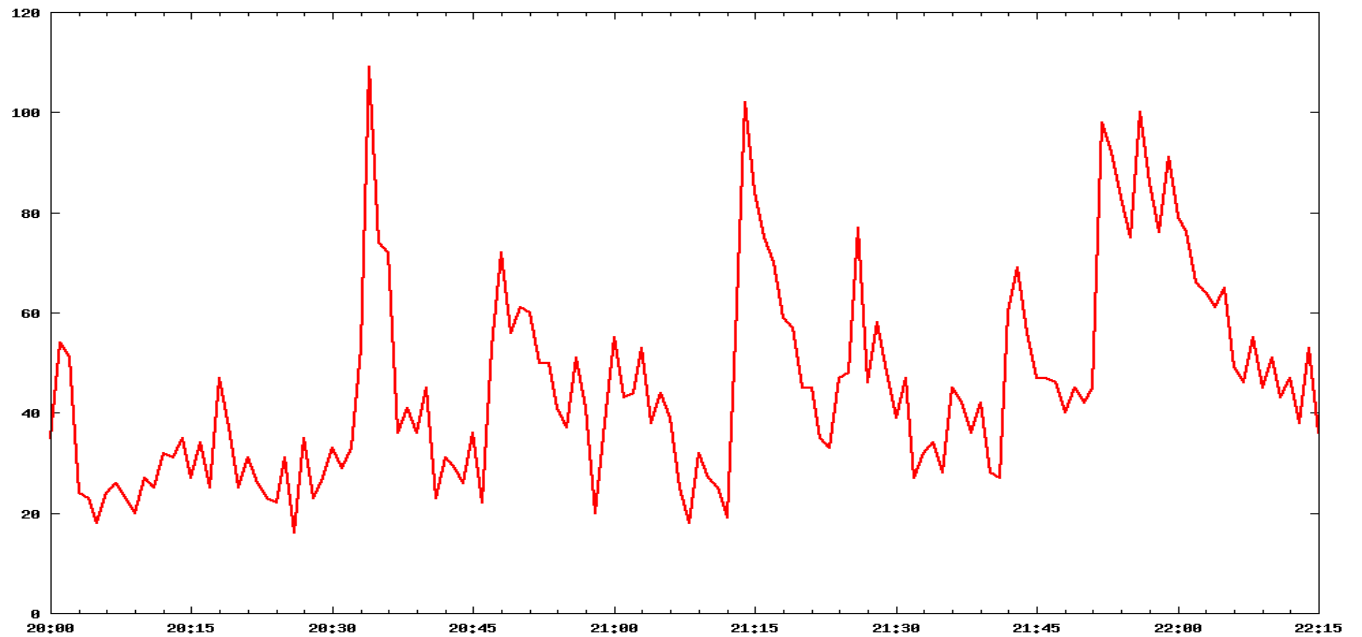2112: GOAL 0-2
2150: Full-time 0-2

Barcelona:

Links & Retweets

- Liverpool

  - On average every sixth tweet has at least one link

  - Every 15th tweet is a retweet

- Madrid

  - On average every third tweet has at least one link

  - Every tenth tweet is a retweet

Users

- Liverpool, in total 87761 users

  - Users with 1000+ tweets: 802 = 1 %

    - But 50 % of all tweets

  - Users with 51-1000 tweets: 6222 = 7 %

    - 26 % of all tweets

  - Users with 2-50 tweets: 36105 = 41 %

    - 8 % of all tweets

  - Users with only one tweet: 44632 = 51 %

    - 16 % of all tweets

Users

- Madrid, in total 103 632 users

    - Users with 1000+ tweets: 813 = 0.8 %

        - But 48 % of all tweets

    - Users with 51-1000 tweets: 7573 = 7 %

        - 28 % of all tweets

    - Users with 2-50 tweets: 44713 = 43 %

        - 9 % of all tweets

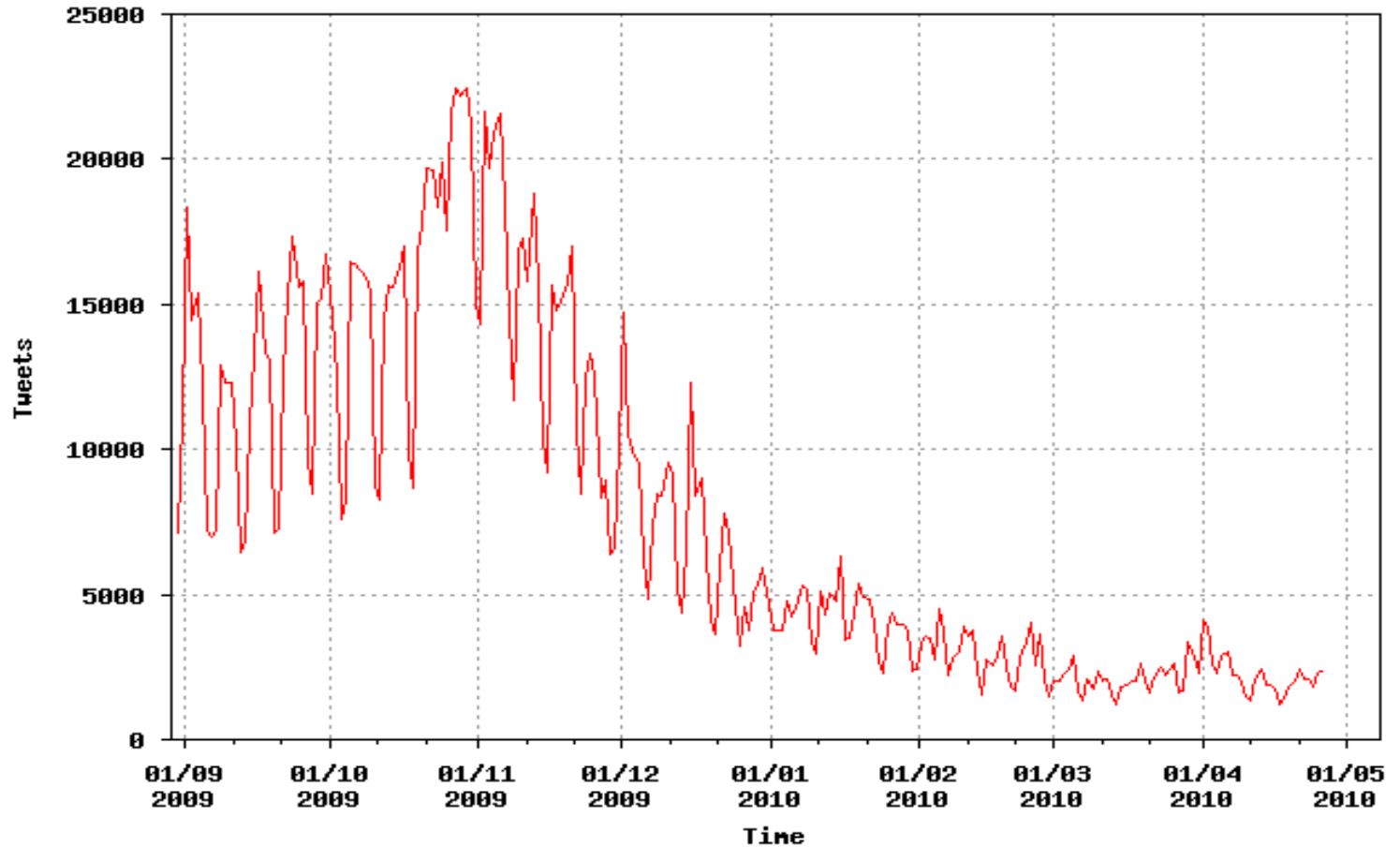   - Users with only one tweet: 50533 = 49 %

        - 16 % of all tweets

Topical Results

- We have collected tweets related to the swine flu outburst and the 2010 Winter Olympics with such keywords as:

  – H1N1, swineflu, and swine flu

  – Vancouver, olympic, olympics, and olympic games
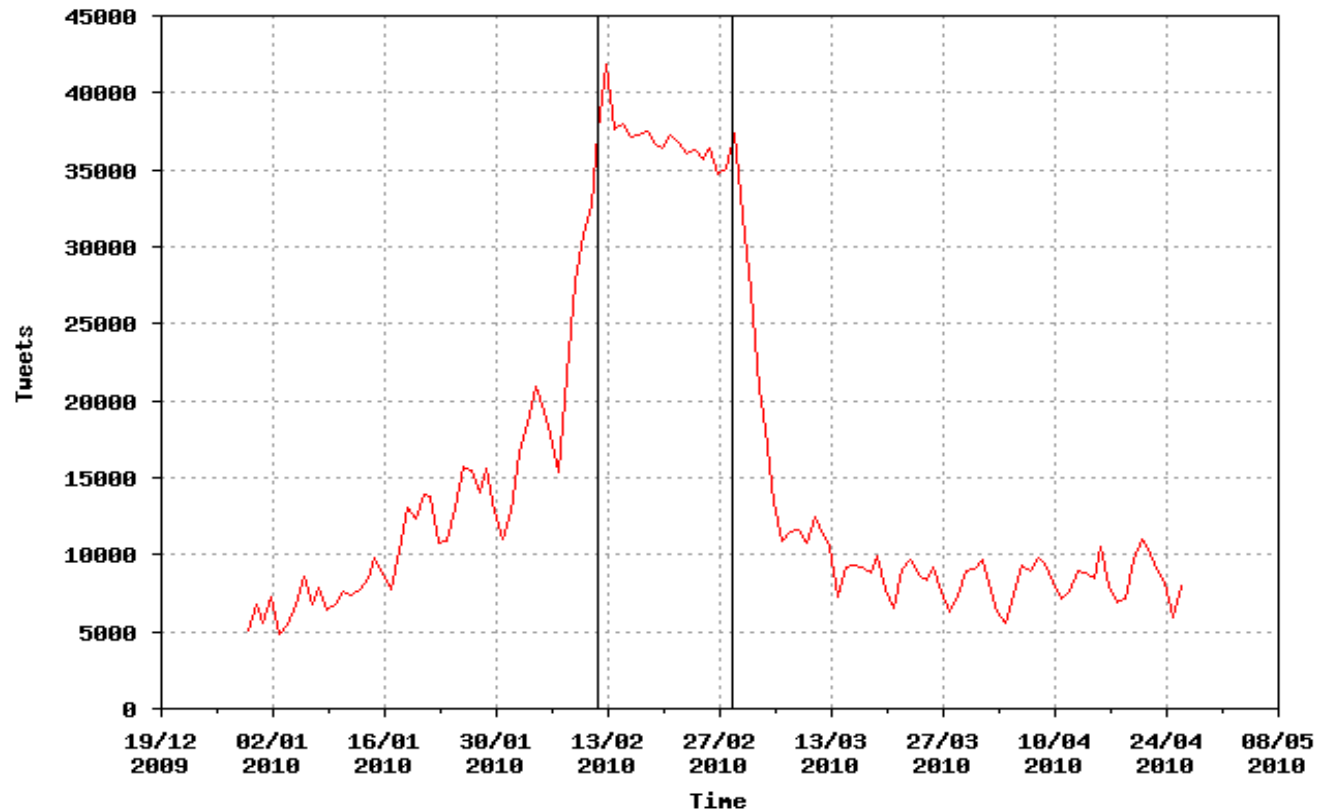
Topical Results

Twitter results with keywords H1N1, swineflu, and swine flu
Tweets per 24 hours
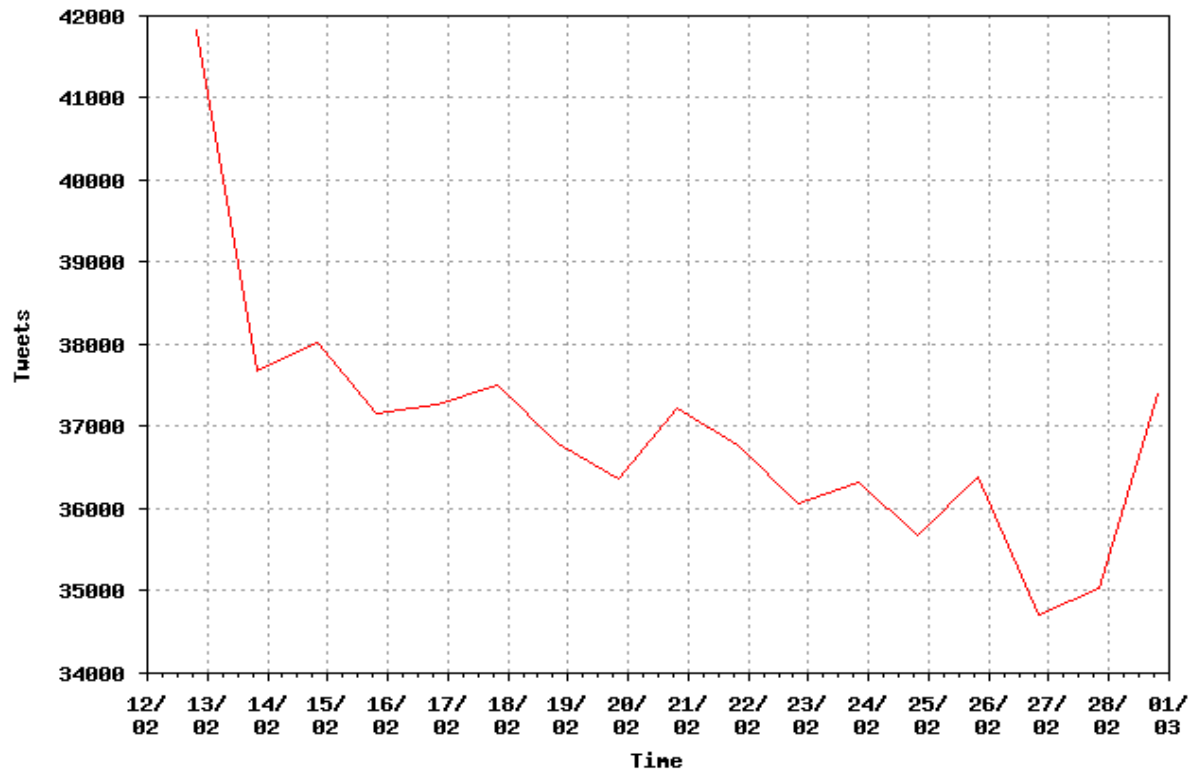
Topical Results

Twitter results with keywords vancouver, olympic, olympics and "olympic games"
Tweets per 24 hours

Topical Results

Twitter results with keywords vancouver, olympic, olympics and "olympic games"
Tweets per 24 hours

Conclusion

- The results indicate that
  - There are regional and cultural differences
  - The users are tweeting about current events, such as sporting events, awards shows, and topical situations
  - The user are willing to express both their positive and negative thoughts

Future

- More cities

- Natural language analysis

- World Cup 2010

- Comparison against other social media services