



University of Cyprus  
Department of Computer  
Science



Laboratory for Internet Computing  
Department of Computer Science  
University of Cyprus  
<http://linc.ucy.ac.cy>

# The Role of Twitter in YouTube Videos Diffusion

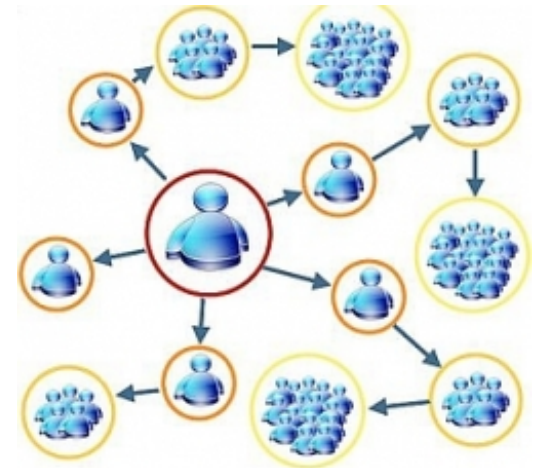
George Christodoulou  
EPFL  
Switzerland

Chryssis Georgiou & George Pallis  
Computer Science Department,  
University of Cyprus

29 November 2012

# Introduction

- Explosive growth of social media: Facebook, Twitter, Digg, etc
  - The primary causes behind the recent increases in HTTP traffic
  - Anyone has the potential of spreading a message
- Growing popularity of streaming media services



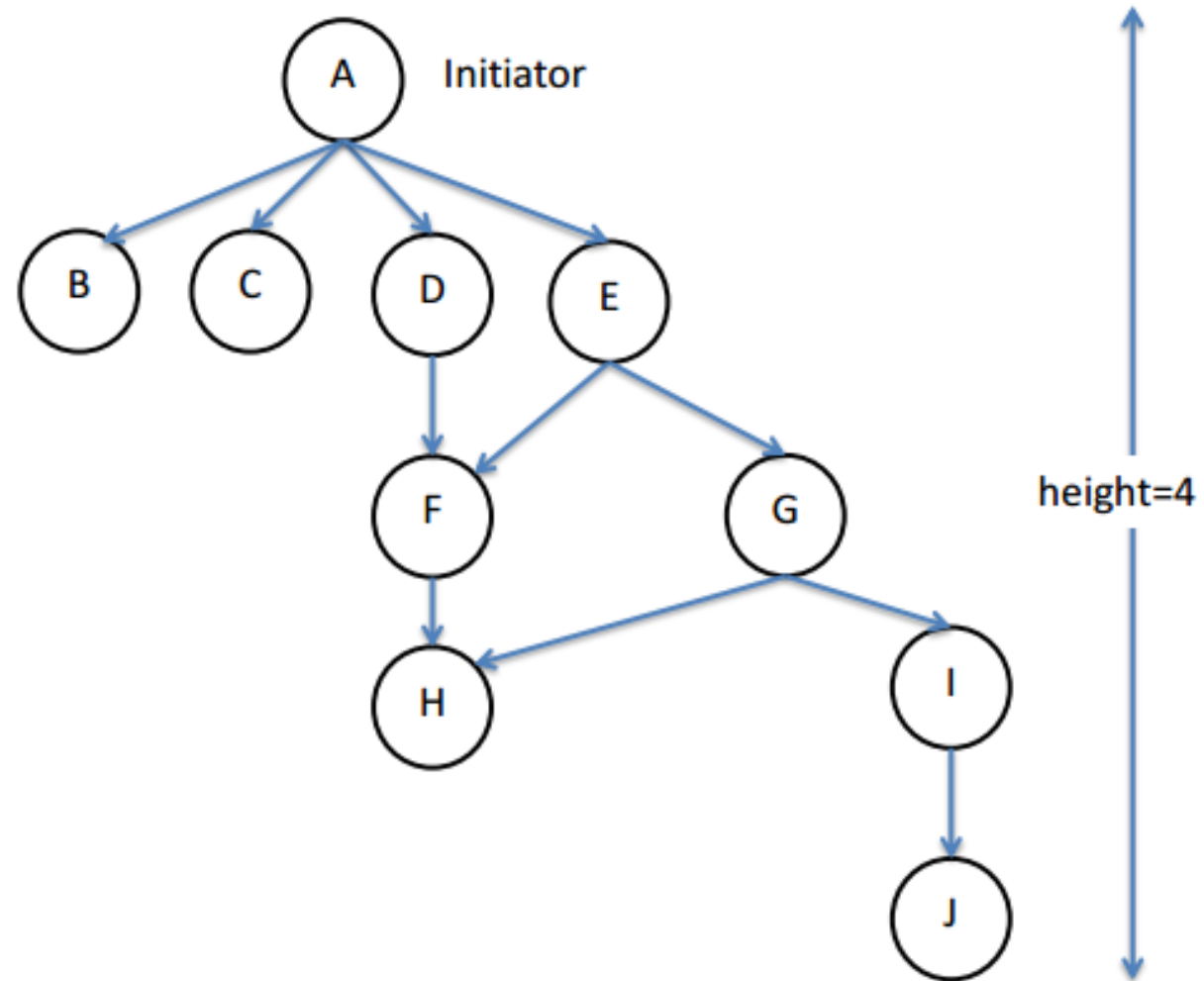
# Spread of information

- *Social network* plays fundamental role in spread of information or influence
- Viral marketing (Word of mouth)
  - An idea gets a sudden widespread popularity
- Examples:
  - GMail achieved wide popularity and the only way to obtain an account was through referral
  - In blogs a piece of information spreads rapidly before eventually picked by mass media
- *Social Cascade* is information spreading through a social network due to *influence* by others

# What is a social cascade?

- We say that user B was reached by a social cascade about content c if and only if:
  - There is another *user A* which posted *content c* and
  - *User A* posted *content c* before *user B* posted it and
  - There was a social connection from *user A* to *user B* when *user A* posted *content c*

# Social Cascade Representation

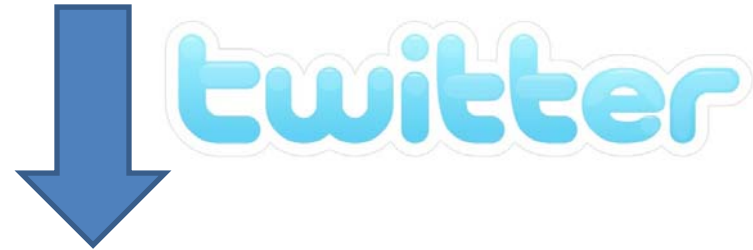


# Background

- Microblogging service
- 200 million users as of March 2011
- User *follow* other users
- Many business, celebrities, news organizations, and others use Twitter
- *Tweet*: A short post of at most 140 characters
- *Retweet*: A tweet originally posted by a user and reposted by another user

# You <sup>TM</sup> Background

- YouTube is the most popular and bandwidth intensive service of today's Internet
  - accounts for 20-35% of the Internet traffic with 35 hours of videos uploaded every minute and more than 700 billion playbacks in 2010
- YouTube has transformed Internet users from video consumers to video producers
- More than 400 tweets per minute with a YouTube Link



affects users' navigation behavior

What is the role of  
retweeting  
in  
YouTube video diffusion?





Serious implications and new challenges  
for Internet services and content providers  
towards improving:

- Caching,
- Content Delivery Networks,
- Searching,
- Content Recommendation

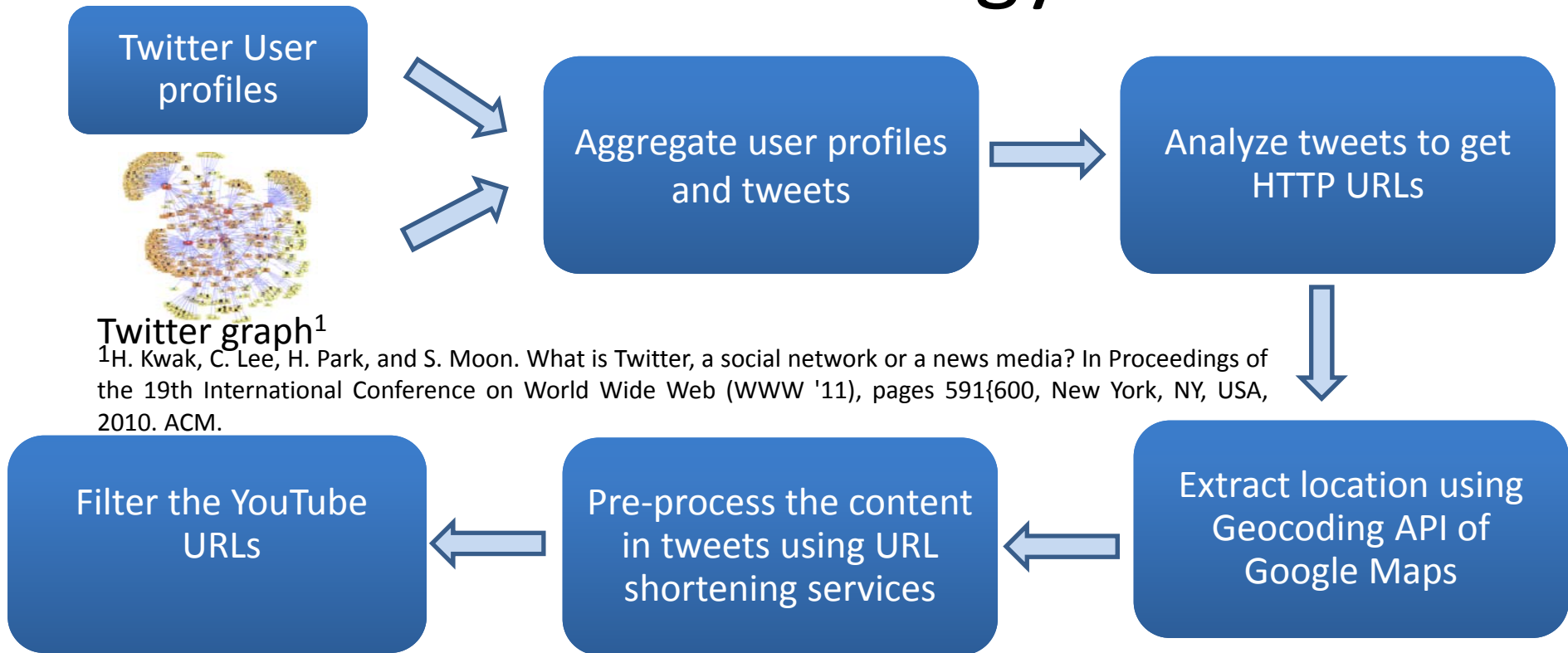
# Contributions

- We present the methodology that we have followed in order to collect the Twitter dataset
  - 1,4 million users
  - 300 million tweets
  - 1,3 millions single messages with a video link
- We examine the role of social cascading in YouTube video diffusion

# Key Challenge: Gathering the data

- Data collection is not a straightforward process
  - Twitter's recently modified policy of limiting the number of search requests per hour from a given IP address
- Data collection took over five months using four Cloud infrastructures
  - Nephelae Cloud - <http://grid.ucy.ac.cy/Nephelae/>
  - Okeanos - <https://cms.okeanos.grnet.gr/about/>
  - Amazon EC2
  - Rackspace

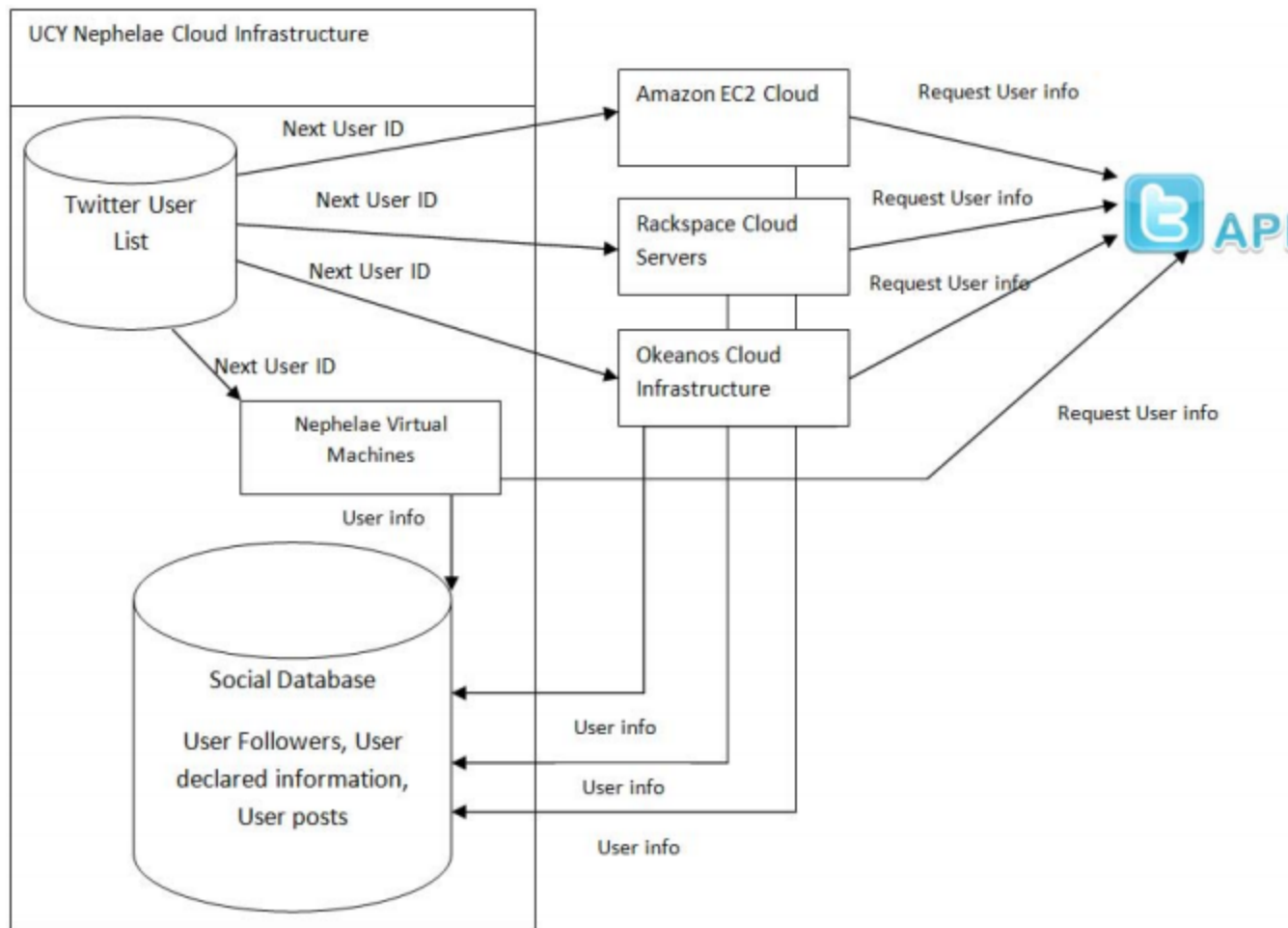
# Methodology



- **1,384,758 Twitter user profiles**
- **247,399,334 directed follower links**
- **299,071,571 public tweets**
- **1,3 millions YouTube URLs**

(December 2011 - April 2012)

# Data Collection System for Twitter



# Data Set Characteristics

Twitter User Profile Information	
Verified	user's identity has been verified by email account
Followers_count	the number of users that follow the user
Protected	user's account is private and only their approved followers can read their tweets or see extended information about them
Listed_count	the number of lists the user is a member of
Friends_count	friends count the number of users the user follows
Location	the location of user
Geo_enabled	if enabled allows applications to send tweets with a geographic location attached
Lang	the language of user
Favourites_count	the number of tweets the user has classified as favorites
Created_at	the date that the account has been created
Time_zone	the time zone of each tweet

# Data Set Characteristics

Tweet Information	
ID	the unique ID of the tweet
Text	the text of tweet (typically up to 140 characters)
Created_at	the date that the tweet has been published
Retweeted	if it is new tweet or a retweet
In_reply_to_status_id	the ID of an existing status that the update is in reply to
In_reply_to_user_id	the user ID that the tweet replies
Urls	the url of the tweet
Retweet_count	the number of times that a tweet has been retweeted

# Effects of Social Cascading

- How can we investigate the role of the retweeting mechanism in YouTube video diffusion?

*video retweet likelihood*

$$\frac{|OutgoingVideos \cap IncomingVideos|}{|IncomingVideos|}$$

- *OutgoingVideos*: the set of videos (over all users) that were retweeted by some user to another user (multiplicities are not counted)
- *IncomingVideos*: all videos that a user received by some other user, as a result of a retweet of the latter



# Effects of Social Cascading



IncomingVideos:



OutgoingVideos:

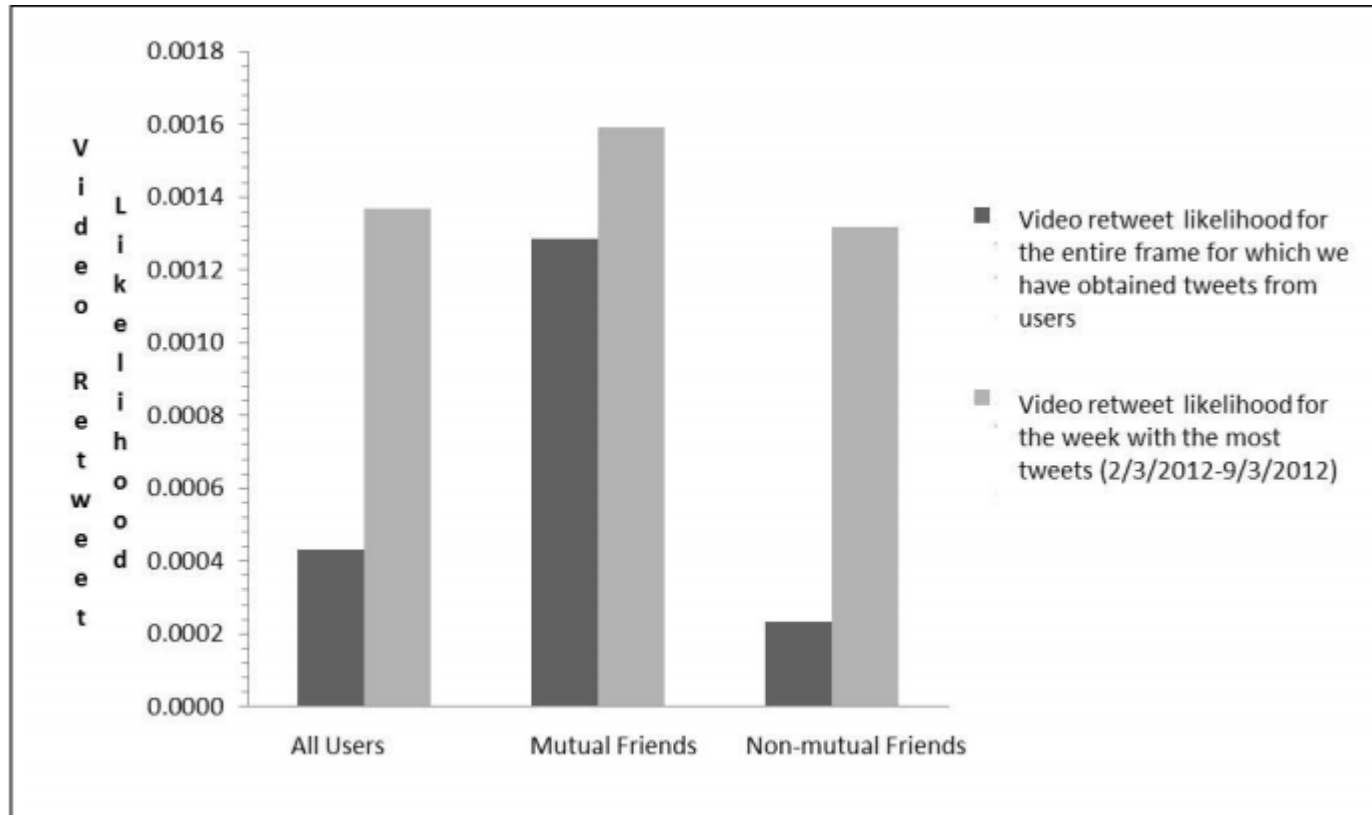


|OutgoingVideos  $\cap$  IncomingVideos|:

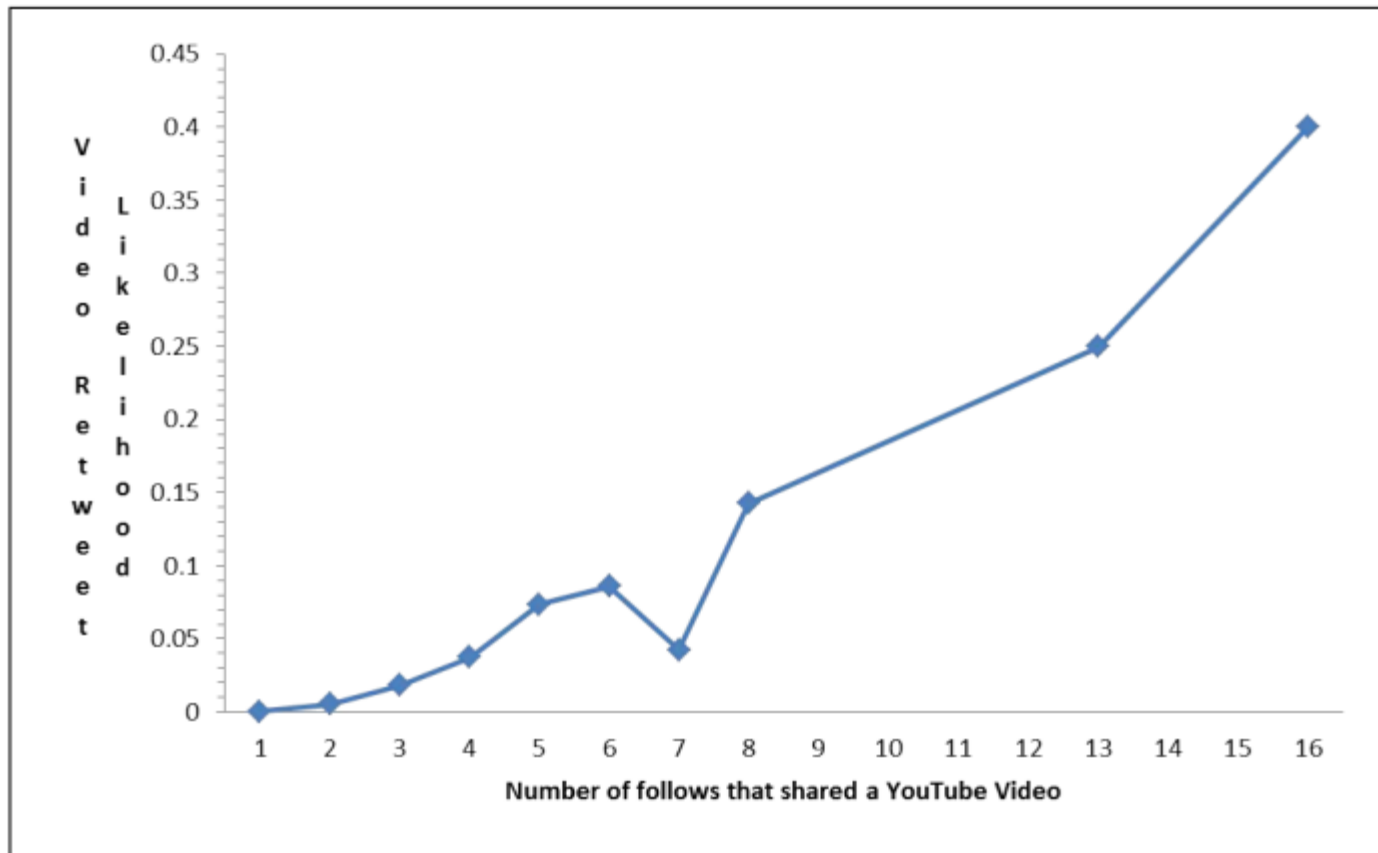


# Evaluation

# Impact of Twitter Users

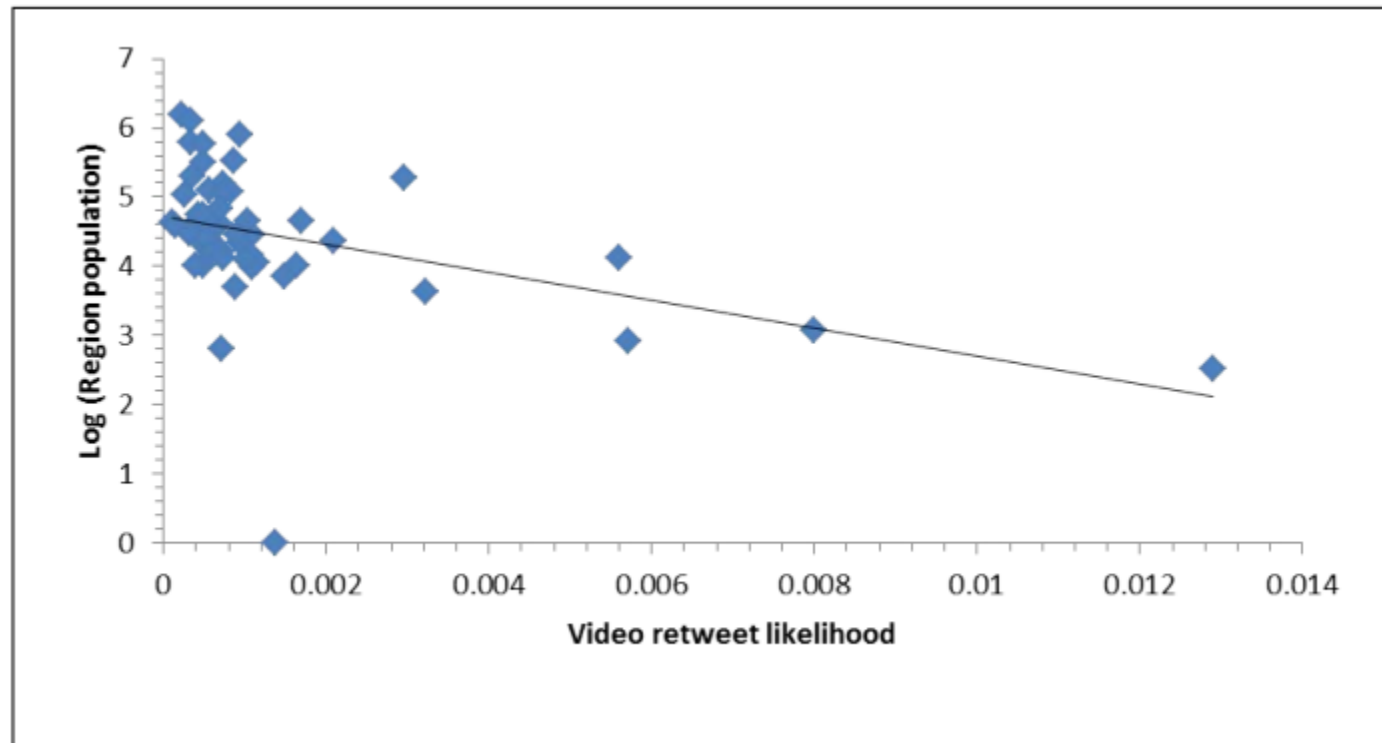


# Impact of Twitter Users



# Impact of Geographic Popularity

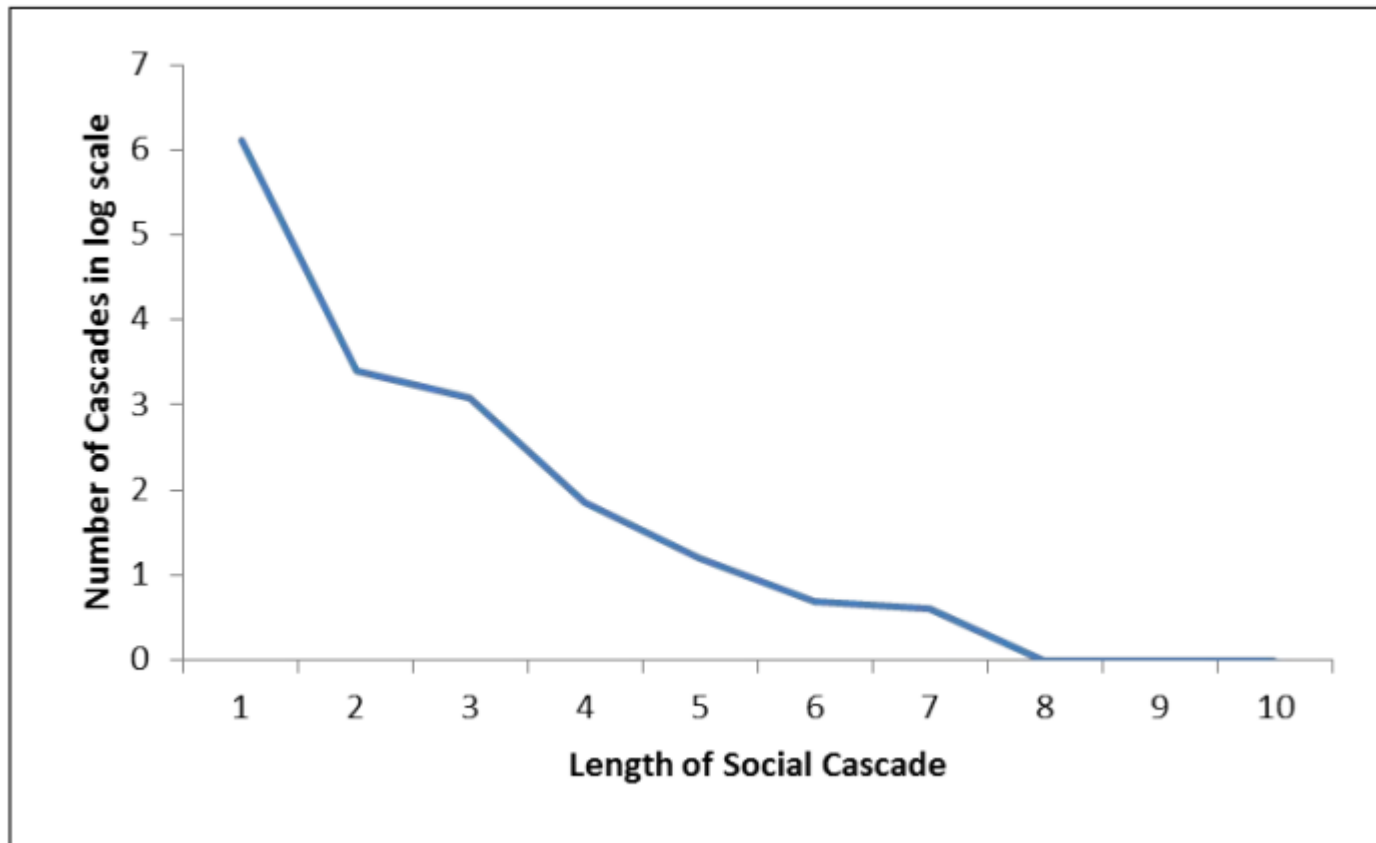
- Twitter uses its own time zone system which divides the globe into *142 zones*



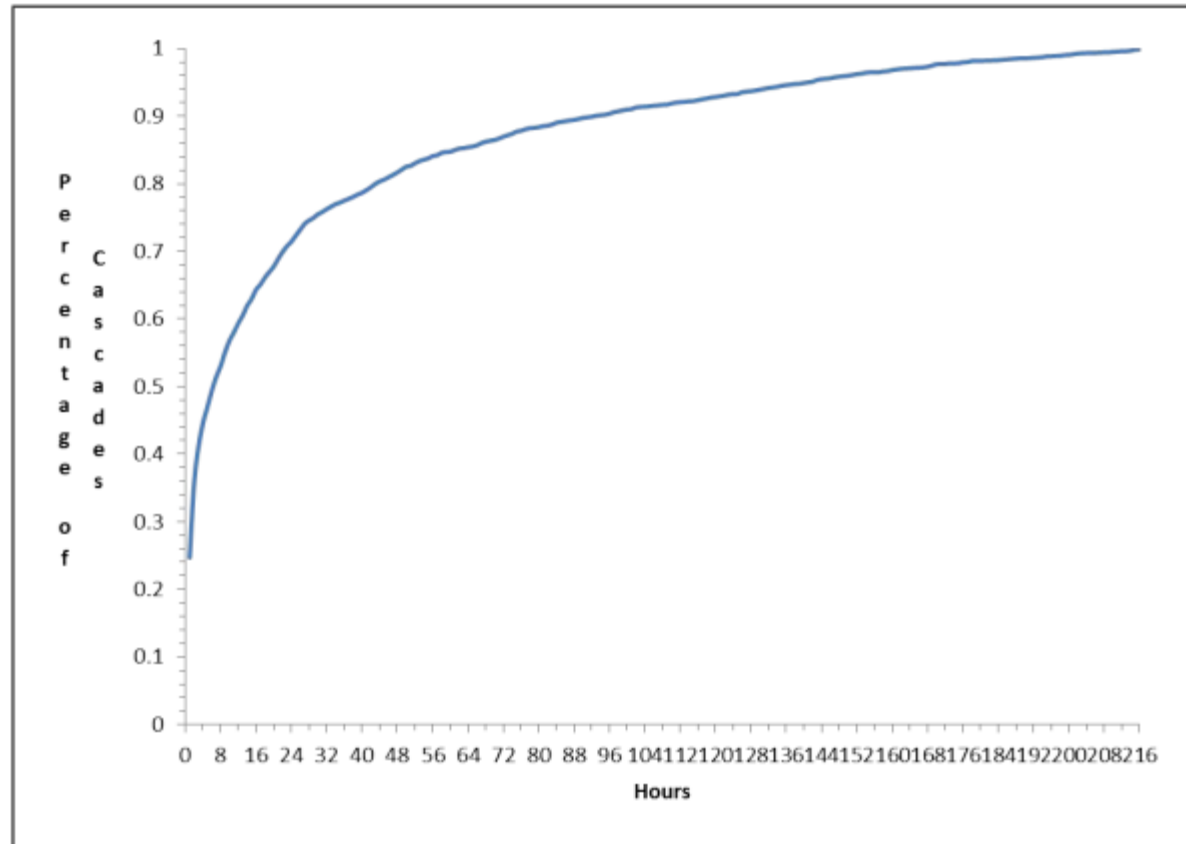
# Impact of Videos Popularity

Views	Number of videos	Number of retweets	Avg. Number of retweets
1000	68655	85252	1.24
5000	37899	45640	1.20
20000	43014	49197	1.14
50000	34855	40515	1.16
200000	53509	68099	1.27
400000	23544	34638	1.47
700000	16010	27011	1.68
1 million	8571	16452	1.92
2 millions	13332	27721	2.08
5 millions	11183	30245	2.70
10 millions	4641	18924	4.07
20 millions	2205	12967	5.88
50 millions	1210	10775	8.90
100 millions	393	10519	26.76
200 millions	21	553	26.33
350 millions	7	176	25.14

# Impact of Social Cascade Length



# Impact of Time





# Conclusions

- We study how the retweeting influences the diffusion of YouTube videos
- *Observations:*
  - Social cascading effect
    - has high impact on a more focused and less diverse set of geographic regions
    - ends within 24 hours
  - Users are influenced more from the follows who also follow them
- *Ongoing/Future work:*
  - study the retweeting influence wrt the popularity dynamics of YouTube videos over the time
  - develop a realistic media workload generator



University of Cyprus  
Department of Computer  
Science

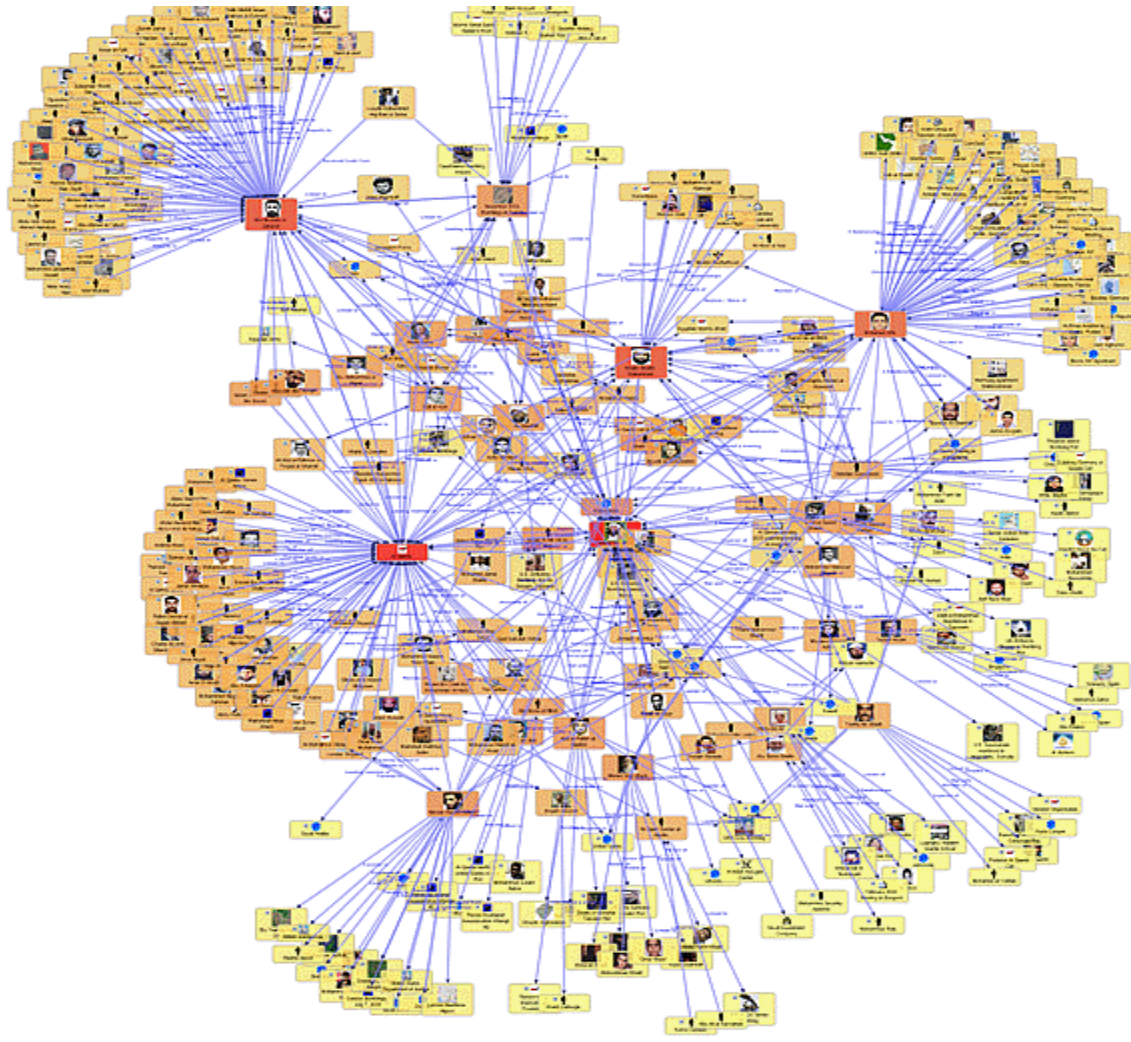


Laboratory for Internet Computing  
Department of Computer Science  
University of Cyprus  
<http://linc.ucy.ac.cy>

# Thank You!

## Questions?

# twitter Background



# Data Collection System for Twitter

