Identification of Real-World Daily Patterns based on Interactions in Online Social Networks

Hariton Efstathiades, Demetris Antoniades, George Pallis, Marios D. Dikaiakos
Laboratory for Internet Computing (LInC), Computer Science Department, University of Cyprus
{h.efstathiades, danton, gpallis, mdd} @ cs.ucy.ac.cy

Introduction - Motivation

Increasing availability of Internet enabled devices and massive use of OSNs
- Creates a continues stream of information

People tend to publish information from several locations and devices at different times of the day [2]
- This information contains meta-data (e.g. timestamp or location)

Online Social Networking interactions can be used to address real world challenges [1]
- Analysis of real-world patterns based on users' home or work areas
- Real-world incidents/events identification based on anomaly detection

Users interaction in Online Social Networks is being influenced by different factors [3]

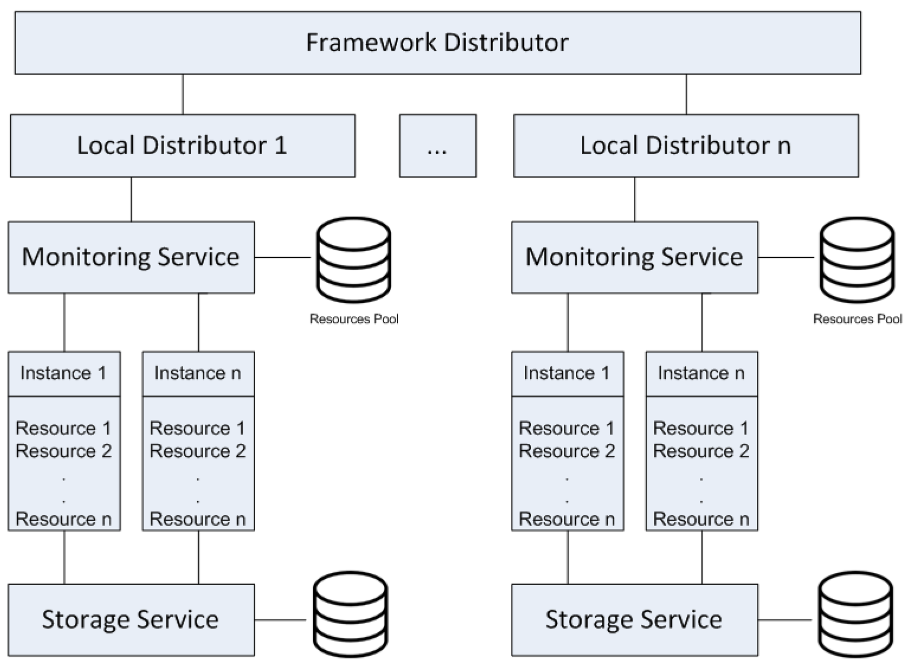
We aim in investigating if users' key locations influence their behavior in OSNs

Research Questions:

1. Based on the differences of users' behavior in Online Social Networks during different time-frames, is it possible to identify their key locations by analyzing the meta-data of the information that they publish?
2. Does users' Online Social Networking behavior have a relation with the geographic areas that they live in or visit? Are there differences in social graphs' structure regarding the different characteristics of areas?
3. Do users' key geographic areas influence their real-world patterns?

Dataset Collection Framework

Intelligent and efficient Twitter Data Collection



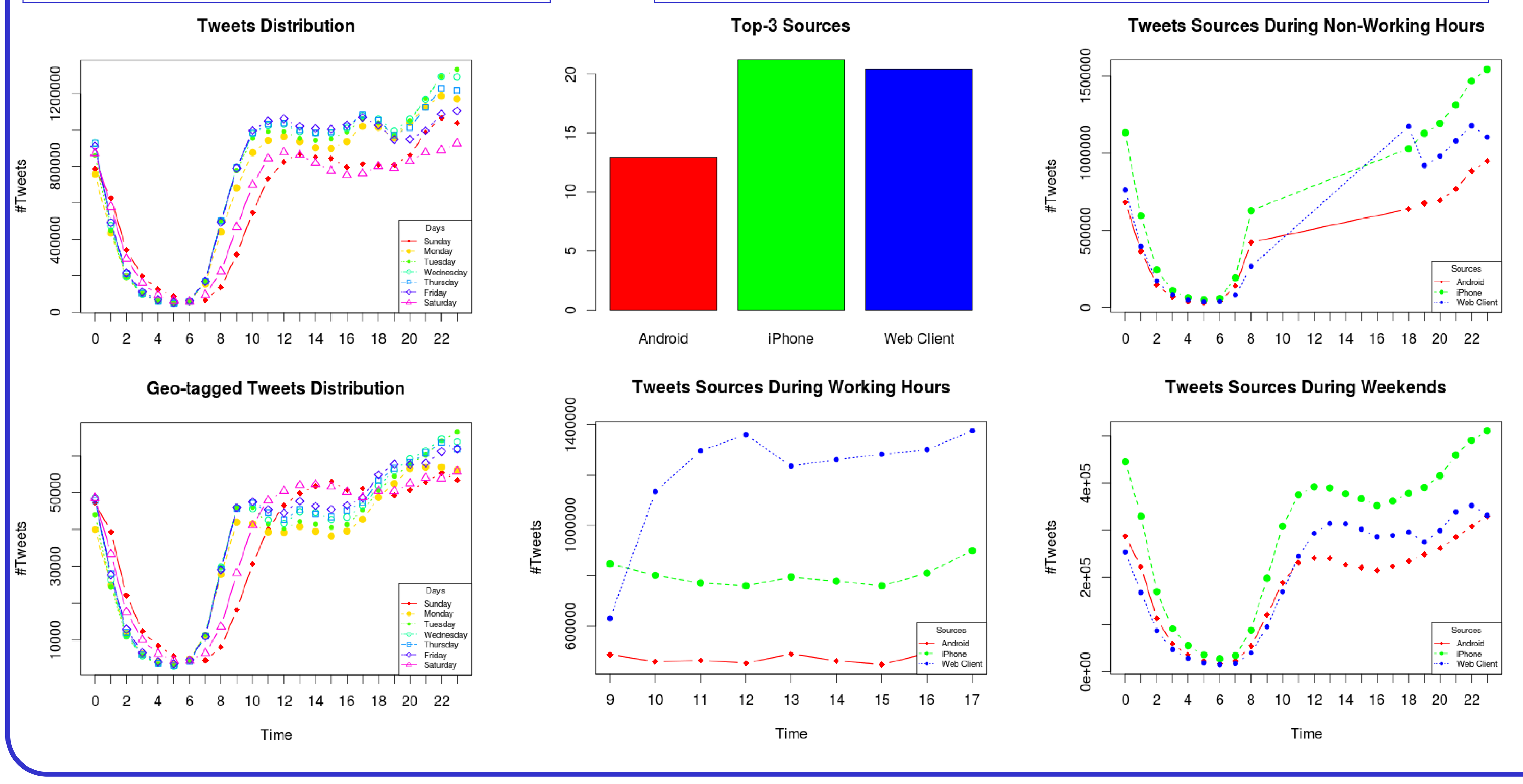
Given as input a list of user_ids or screen_names:
- **Global Workload** is distributed based on the number of Local Distributors
- **Local Workload** is distributed in different instances based on availability of local resources
- **Each instance** is able to run forever as monitoring service adds or removes resources based on instance needs

Throughput: 3000 – 3200 users/hour per Local Distributor

Dataset Description - Analysis

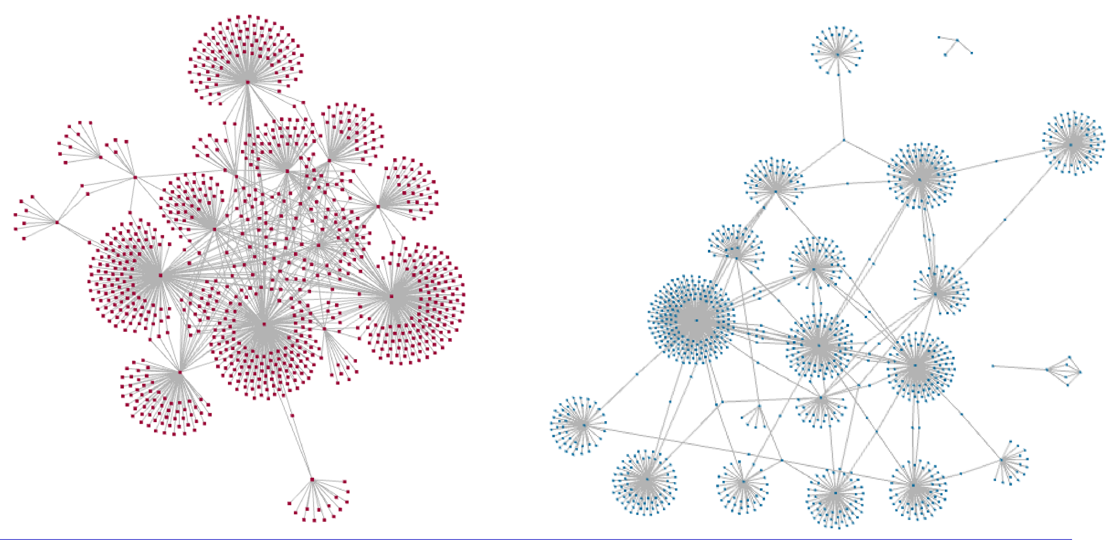
Online Social Network Data	
OSN	Twitter
#Users	99,950
#Tweets	118,012,958
#Geo-tagged Tweets	6,161,290

Ground Truth Data	
Source	Twitter
Description	• Official geographical boundaries of the country of Netherlands. In total 1024 different areas • Population and number of Employees for each area



Zwolle is a municipality and the capital city of the province of Overijssel, Netherlands [Wikipedia]
- Population: about 125,000
- Its habitants are mostly locals

Amstelveen is a municipality in the province of North Holland, Netherlands [Wikipedia]
- Population: about 85,000
- A large percentage of its habitants are students, as VU Amsterdam is located in this area



Habitants In-Area Activity		
	Zwolle	Amstelveen
Non-Working Hours	80%	76%
Working Hours	33%	31%

Other Habitants Areas	
Zwolle	Amstelveen
ABROAD, LEISURE AREAS IN UTRECT, LEISURE AREAS IN AMSTERDAM	ABROAD, SCHIPHOL INTL AIRPORT, HOLLAND SPORT BOAT CENTER

Future Plan

- Model and Simulate social graph's evolution based on users' key locations
- Model and Simulate Online Social Networking behavior based on users' key locations

References

[1] Abel Fabian, Claudia Hauff, Geert-Jan Houben, Richard Stronkman, Ke Tao. "Semantics+ filtering+ search= twitcident. exploring information in social web streams." *Proceedings of the 23rd ACM conference on Hypertext and social media*. ACM, 2012.

[2] Katragadda Satya, Miao Jin, Vijay Raghavan. "An Unsupervised Approach to Identify Location Based on the Content of User's Tweet History." *Active Media Technology*. Springer International Publishing, 2014. 311-323.

[3] Sadilek Adam, Henry Kautz, Jeffrey P. Bigham. "Finding your friends and following them to where you are." *Proceedings of the fifth ACM international conference on Web search and data mining*. ACM, 2012.

MARIE CURIE ACTIONS
SEVENTH FRAMEWORK PROGRAMME