Situation-awareness in social overlays

Hariton Efstathiades
Research Focus

- **Situation-aware overlay:** Each individual is connected with the source of information, according to the current situation.
Scenario - Vision

User

Andres Ledesma
Researcher at University of Cyprus
Cyprus | Computer Software

LinkedIn Analysis

Andres Ledesma @wizardseal
See you in #stockholm for #iSocial summer school (31st of May - 5th of June)

Twitter Analysis

Andres Ledesma’s Skills & Expertise
- C
- C++
- Java
- Python
- Programming
- Computer Science
- Software Engineering
- Eclipse
- Linux
- Algorithms
- LaTeX
- C#
- SQL
- JavaScript
- Software Architectural Design

Social web data

Nick Thibieroz @NThibieroz · May 6
AMD and Microsoft are presenting on advanced graphics topics in a free Stockholm developer event on Monday 2 June: eventbrite.com/e/amd-microsoft

Nick Thibieroz @NThibieroz · May 27
Hey crowd! I’ve got two ticket for Aerosmith in the 1st of June and John Mayer on 12th of June in Stockholm, giving out half-priced, DM


Practical Issues

- Limited attention

- User is not able to follow everyone
- Not interested in the entire social stream of the publisher
Practical Issues

- Limited discovery

Only community can discover published info
Practical Issues

- Fast data

- Information gets obsolete fast
- User’s environment change fast
Today’s presentation
Today’s presentation

Environment - Situation

Dataset Collection

Sample Verification

Analysis - Observations

Linkedin

twitter

Dataset Collection

Sample Verification

Analysis - Observations

Data Aggregation

Analysis - Observations
LinkedIn analysis

Business-market knowledge extraction from online social networks
LinkedIn Analysis: Problems

- Existing APIs have many limitations
  - LinkedIn API limitations
    - User oriented: Need user to give permissions
  - Without users’ permissions
    - Access to basic profile fields (name, headline, community size)
    - Throttle limits
      - Search for user
      - Access resource
      - Access company profile

- The public information that we can find on the website is much more than the information that we can retrieve from API!
LinkedIn Analysis: Solution

- Development of LinkedIn crawler
  - “Web Scraping” method
- Simulates visitor’s behavior
- Reads the tags and extracts the data
LinkedIn Analysis: Crawler

- Two parts:
  - Directories collector
  - Users collector
Dataset Collection tool

- Directories collector
  - OSNs have directories with the public URLs of their users
  - The tool is able to visit and retrieve LinkedIn’s directories
    - Country-wise
    - Global directory
Dataset Collection tool

- Users collector
  - Each user has a public profile URL
  - The tool is able to retrieve user’s public profile information
    - Input: Public URL
    - Output:
      - Information about the User, experience, education, interests

- Tested on cy.linkedin.com
  - 119,817 profiles

- Collection in progress
  - Netherlands, Portugal
Data Description
Data analysis

**Purpose**
- Infer implicit connections between users
- Identify market-business trends
  - Collaboration with Department of Management, University of Cyprus
- Anomalies detection for early warning system
- Personalized warnings based on user’s background

<table>
<thead>
<tr>
<th>Introduction</th>
<th>LinkedIn Analysis</th>
<th>Twitter Analysis</th>
<th>Summary</th>
<th>Next steps</th>
</tr>
</thead>
<tbody>
<tr>
<td>● ● ● ● ● ●</td>
<td>● ● ● ● ● ● ● o o</td>
<td>o o o o o o o</td>
<td>o</td>
<td>o</td>
</tr>
</tbody>
</table>
Early results
Banking sector:
- Increasing trend in both charts!
- Cyprus is a special case from March of 2013
  - Employees obligatory left from Laiki Bank and joined Bank of Cyprus

Telecommunications:
- Decreasing “leaving” trend from 2009
- Increasing “hiring” trend from 2012
- Many foreign big companies (Amdocs, TSys, UTX etc.) are still growing their operations in Cyprus
Industry Leaves for 2013 (per month)

- **Banking sector:**
  - About 30% of total year leaves left in August
  - Deadline to get a reimbursement if they decide to leave

- **Education:**
  - **Government Policy:**
    - Temporal school teachers’ contracts expire in June.
    - Renew in September
Identification of transportation patterns based on online social networks’ meta-data
Dataset Collection

- Problems:
  - Twitter throttle limits
  - We can retrieve only 1% of total Twitter stream

- Solution:
  - Development of Twitter stream listeners
  - Distribute different listeners in the areas that we want to collect
### Dataset Collection

- **Collection of Twitter users from city of Amsterdam**
  - Tweets history ($\leq$3200 tweets) for each user
  - Meta-data of each Tweet
    - Timestamp
    - Location
- **Dataset:**
  - ~10,000 users
  - ~6 millions tweets
    - ~1,5 millions geo-tagged
Dataset Collection

- Ground truth data from the city of Amsterdam
- Amsterdam zones in polygons
- Information about each zone:
  - habitants
  - jobs
Data analysis
# Data analysis

## Purpose (Steps in progress...)
- Infer implicit connections between users from meta-data
  - E.g. same working/leisure area
- Identification of transportation patterns
- Analysis of the influence of home or working location to the transportation pattern of the user
- Events detection based on meta-data

<table>
<thead>
<tr>
<th>Introduction</th>
<th>LinkedIn Analysis</th>
<th>Twitter Analysis</th>
<th>Summary</th>
<th>Next steps</th>
</tr>
</thead>
<tbody>
<tr>
<td>● ● ● ● ● ● ●</td>
<td>● ● ● ● ● ● ● ● ● ●</td>
<td>● ● ● ● ○ ○ ○ •</td>
<td>○ ● ○ ○ ○</td>
<td>○ ○ ○ ○</td>
</tr>
</tbody>
</table>
Early results

One example
**Data analysis**

- **Identification of Home/Working Location**
  - Divide day-hours in parts (case of Netherlands)
    - Working: 09:00 – 17:30
      - Most popular area: marked as Job area
    - Home: 22:00 – 09:00
      - Most popular area
  - Leisure:
    - Most popular area except Home
KING’S DAY
AMSTERDAM
...the craziest day of the year!!!
Summary

Environment - Situation

Linkedin

Dataset Collection

Sample Verification

Analysis - Observations

Twitter

Dataset Collection

Sample Verification

Analysis - Observations

Data Aggregation

Analysis - Observations
Next steps
Next steps

- Extend datasets (In progress)
  - LinkedIn: Different countries
  - Twitter: More users
  - Other SNs
- Enrich users’ profiles based on knowledge from social networks analysis (e.g. Unemployed, based on transportation pattern, working location etc.) (In progress)
- Identify the same user in different platforms (Early stage)
- Investigate the correlation between data that user publishes (Twitter) with his background (LinkedIn) during the reported period
- Investigate if we can model the behavior of the user based on the situation
Next step focus

Environment - Situation

Dataset Collection
  Sample Verification
  Analysis - Observations
  Data Aggregation
  Analysis - Observations
Decentralized Vision
Questions?