ETH
Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich

D-ITET
Computer Engineering and Networks Laboratory
Communication Systems Group

Social Data Mining on Smartphones

Semester Thesis

Ajita Gupta

Advisor: Sacha Trifunovic
Supervisor: Prof. Dr. Bernhard Plattner

Outline

1. Introduction
2. Related Work
3. Design
4. Implementation
5. Application Deployment
6. Evaluation Results
7. Conclusion

Introduction

Motivation

- Endless number of possibilities to connect with people

  → Applications must exploit heterogeneity and behavior patterns

- Social Data Mining Systems record and analyze user activities

  → User satisfaction is improved

Introduction

Problem Statement

1. Collect social contacts

2. Analyze and correlate social interactions

3. Classify communication patterns

→ Android Application: *SocialMine*

Related Work

Stumbl

- **Aim**
  Define patterns between mobility, social connections and communication of users

- **Methodology**
  Facebook Application

- **Difference to our approach**
  + We capture more details and interaction types
  + Negligible user effort
  - Limited set of users

Related Work

Device Analyzer

- **Aim**
  Improve future smart phones, extract patterns and trends

- **Methodology**
  Android Application

- **Difference to our approach**
  + Focus on more social data (mainly from Facebook)
  - No incentive (e.g. live statistics)

Design

Social Data Mining System

Implementation

Extraction Modules

Contacts-Fetcher	Call-Logger	SMS-Collector	Location-Tracker	Facebook-Miner
- Contact Names - Contact Numbers - Email Addresses - Postal Addresses - Organization - Notes - Websites - Instant Messaging ID's	- Caller - Callee - Timestamps - Call Duration	- Sender - Recipient - Timestamps - SMS Length	- Timestamps - Latitude - Longitude - GSM Cell Location (Cell-ID, LAC) - BSSID's - Signal Strength of AP's	- General Profile - Friends - Messages - Wallposts - Hobbies - Pokes - Likes - Groups - Events

Deployment

Specifications

- **Recruitment**
  Invitation E-Mail → 9 candidates

- **Duration**
  Friday, 10th June – Friday, 17th June, 2011

- **Evaluation**
  Feedback E-Mail → User Survey

Evaluation

1) Contact Graph

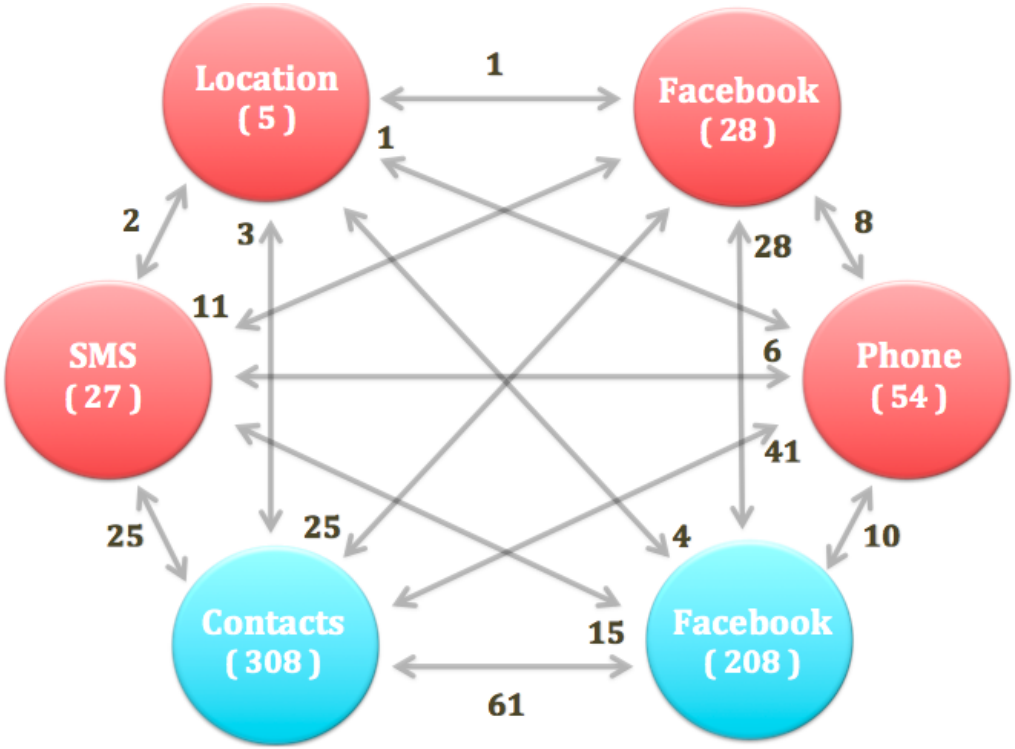
- **Number of Social Ties**
  Contacts list: 308
  Facebook friends: 208 (> 130: Facebook Statistics)
  Overlappings: 61

- **Conclusion**
  High discrepancy between both contact sets

Evaluation

2) Interaction Graph

- Correlations between the contact sets of individual communication mediums



- Conclusion
  - Helpful to classify contacts

Evaluation

3) Communication Patterns Graph

- **Phone**
  - *Call Count: 6.7 calls/day/user*
  - *Call Duration: 1m 54s/call*

- **SMS**
  - *SMS Count: 3 SMS/day/user*
  - *SMS Length: 73 chars/SMS*

- **Location**
  - *User's Daily Time Schedule*
    - ** Workplace: 6h 46min*
    - ** Home: 13h 31min*
    - ** Meals: 1h 24 min*
    - ** Travelling: 2h 17 min*
  - *Interaction Partners: 5.78 colleagues/day/user*
  - *Interaction Duration: 2h 21m/day/user*

Evaluation

3) Communication Patterns Graph

- **Facebook (most recent)**
  - *Message Threads: 13*
  - *Wallposts: 22*
  - *Likes: 8*
  - *Pokes: 0.8 (from and to)*
  - *Groups: 26*
  - *Events: 0.4*

- **Conclusion**
  - *Characterize users (e.g environment)*
  - *Determine the tie strength of relationships*
  - *Derive user preferences*

Conclusion

Summary

- *Basic Social Data Mining System (built upon SocialMine)*
- *Collection of social metadata with five different data extraction modules*
- *Inspection of three different social dimensions*

Conclusion

Outlook

- *Pilot stage of an iterative design process of an efficient Social Data Mining System*

- *Areas for improvement*
  1. *Add User Control*
  2. *Add Data Anonymization Technique*
  3. *Include User Attraction*

→ *Next Step:* **Classification and Profiling of User Behavior**
(SocialMine in the background + User-focused activity in the foreground + large-scale deployment)

Questions?