# Understanding Online Social Network Usage from a Network Perspective

Fabian Schneider*‡

fabian@net.t-labs.tu-berlin.de

Anja Feldmann‡    Balachander Krishnamurthy§    Walter Willinger§

* Work done while at AT&T Labs–Research

‡ Technische Universtität Berlin / Deutsche Telekom Laboratories

§ AT&T Labs–Research

Internet Measurement Conference 2009

# Motivation

- >600,000,000 users on Online Social Networks (OSNs)
  . . . and the number is still growing

- Open questions/challenges
  - Which features are popular among OSN users?
  - How much time do users' spend interacting with OSNs?
  - Is there a correlation between subsequent interactions?

- Relevance of OSN usage
  - ISPs: data transport, connectivity
  - OSN providers: develop and operate scalable systems
  - R&D: Identify trends, suggest improvements, and new designs

# Outline

1. Approach

2. Session Characteristics

3. Feature Popularity

4. Dynamics within Sessions

5. Conclusions

### Sessions
Session = Set of interactions of one user

### Features
Feature = Action a user can perform

# General Approach

① Reconstruct OSN clickstreams from anonymized packet-level traces
- Anonymized HTTP header traces from two large ISPs
- Used Bro[1] to extract HTTP request-response pairs (rr-pairs)

② Map rr-pairs into sessions
- Sessions identified via SessionIDs (from HTTP `Cookie` header)
- Track logins and logouts ⇒ Authenticated or offline state
- Cookies help if login or logout not observed

③ Classify rr-pairs
- Active (rr-pair resulting from user action) or
  Indirect (e.g. followup/embedded via HTTP `Referer` chain)
- Determine user actions, group into 13 categories

---

[1]`www.bro-ids.org`

# OSN Selection

OSN Selection criteria:

- OSNs focussing on profiles (e.g., no YouTube, ...)
- 2 globally popular
- 2 locally popular (well represented at one ISP)

# HTTP Header Traces (anonymized)

- Collected at residential broadband networks of two commercial ISPs
- Each site connects $\geq 20,000$ DSL users
- Endace monitoring cards for packet capture

Table: Overview of anonymized HTTP header traces.

| ID | start date | dur | sites | size | rr-pairs |
|----|-----------|-----|-------|------|----------|
| ISP-A1 | 22 Aug'08 noon | 24h | all | >5 TB | >80 M |
| ISP-A2 | 18 Sep'08 4am | 48h | all | >10 TB | >200 M |
| ISP-A3 | 01 Apr'09 2am | 24h | all | >6 TB | >170 M |
| ISP-B1 | 21 Feb'08 7pm | 25h | OSNs | >15 GB | >2 M |
| ISP-B2 | 14 Jun'08 8pm | 38h | OSNs | >50 GB | >3 M |
| ISP-B3 | 23 Jun'08 10am | >7d | OSNs | >110 GB | >7 M |

# HTTP Header Traces (anonymized)

- Collected at residential broadband networks of two commercial ISPs
- Each site connects $\geq 20,000$ DSL users
- Endace monitoring cards for packet capture

Table: Overview of anonymized HTTP header traces.

| ID | start date | dur | sites | size | rr-pairs |
|---|---|---|---|---|---|
| ISP-A1 | 22 Aug'08 noon | 24h | all | >5 TB | >80 M |
| ISP-A2 | 18 Sep'08 4am | 48h | all | >10 TB | >200 M |
| ISP-A3 | 01 Apr'09 2am | 24h | all | >6 TB | >170 M |
| ISP-B1 | 21 Feb'08 7pm | 25h | OSNs | >15 GB | >2 M |
| ISP-B2 | 14 Jun'08 8pm | 38h | OSNs | >50 GB | >3 M |
| ISP-B3 | 23 Jun'08 10am | >7d | OSNs | >110 GB | >7 M |

# Manual Traces

Data set: Active browsing while monitoring passively

## For customization

- Good faith effort to explore the feature set of the OSN
- Identify site names, relevant cookies, login/logout actions
- Identify URL patterns for action/category classification

## For validation

- Provides ground truth
- 95% of observed actions covered by manual traces
- Remaining actions classified as
  - Guessed (if the URL contains a hint: /ajax/editphoto.php)
  - Unknown

# Category Examples

### Home
All actions on the homepage once authenticated

### Photos
Uploading, tagging, and managing photos

### Profile
Accessing and changing profiles, posting on walls, privacy settings

### Friends
Browsing, inviting, and accepting friends

### Apps
Applications (external and internal), **only** rr-pairs directed towards OSN servers

### Offline
All actions while unauthenticated, e. g., public profile browsing, registering

# Caveats of our Approach

- No automated way for
  - producing the URL patterns or
  - extracting the relevant cookies

- External apps: Not tackled as hosted on different sites
  - Requires customization to all/top external apps
  - Navigation redirects could be leveraged

- Friendship graph: Cannot tell if two users are friends
  - Requires parsing of payload (privacy!)
  - Requires users to actually access their friend lists during observation

# Outline

**1** Approach

**2** Session Characteristics

**3** Feature Popularity

**4** Dynamics within Sessions

**5** Conclusions

# OSN Session Characteristics

## Volume of OSN sessions

- Consistent with a heavy-tailed distribution
- Facebook sessions: 200kB–10MB (StudiVZ: 50kB–5MB)
- Typical Web sessions: 100B–10kB, but heavier tail
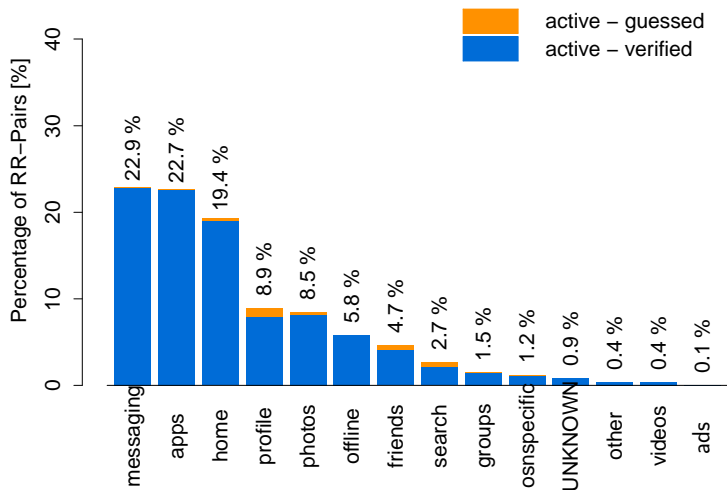
## OSN session durations

- Most sessions are short: 1-5 minutes
- Few lasting for more than an hour (10%–20%)
- Very long (days) sessions observed for 7d trace

# Outline

**1** Approach

**2** Session Characteristics

**3** Feature Popularity
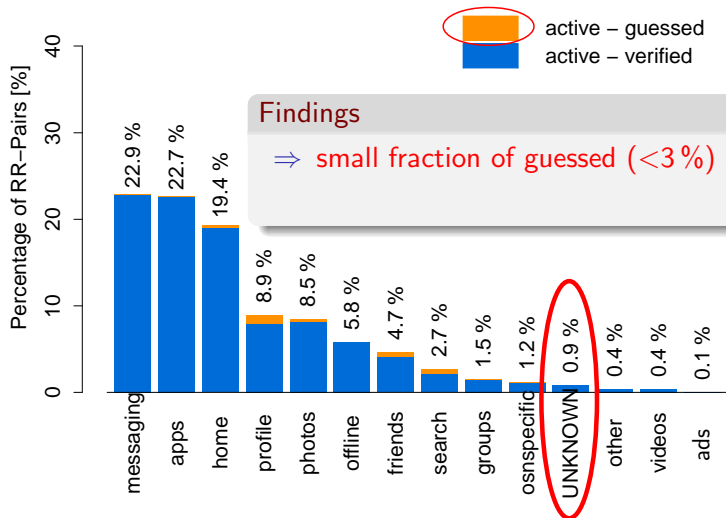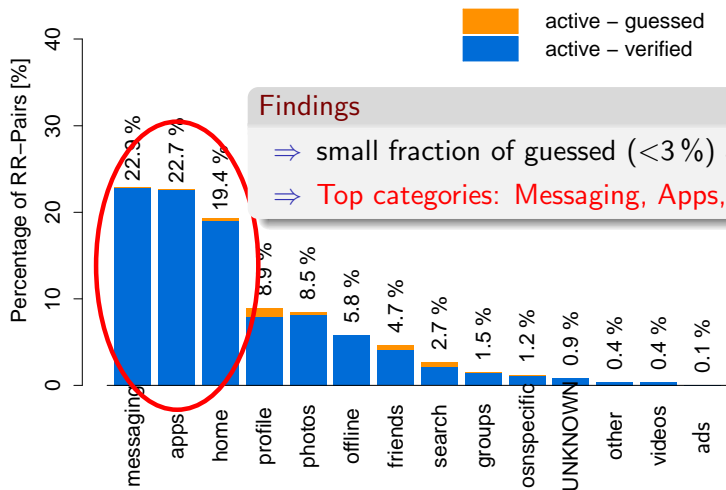
**4** Dynamics within Sessions

**5** Conclusions

# Action Popularity

## Active Facebook rr-pairs by category for ISP-A2

# Action Popularity

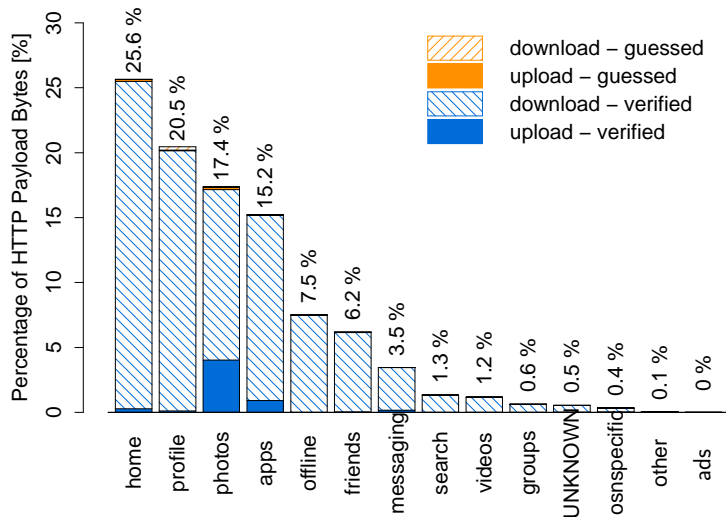Active Facebook rr-pairs by category for ISP-A2



Legend:
- active – guessed
- active – verified

**Findings**

$\Rightarrow$ small fraction of guessed ($<3\,\%$) & UNKNOWN

Chart values (Percentage of RR–Pairs [%]):
- messaging: 22.9 %
- apps: 22.7 %
- home: 19.4 %
- profile: 8.9 %
- photos: 8.5 %
- offline: 5.8 %
- friends: 4.7 %
- search: 2.7 %
- groups: 1.5 %
- osnspecific: 1.2 %
- UNKNOWN: 0.9 %
- other: 0.4 %
- videos: 0.4 %
- ads: 0.1 %

# Action Popularity

Active Facebook rr-pairs by category for ISP-A2



**Findings**

$\Rightarrow$ small fraction of guessed ($<3\%$) & UNKNOWN

$\Rightarrow$ Top categories: Messaging, Apps, Home
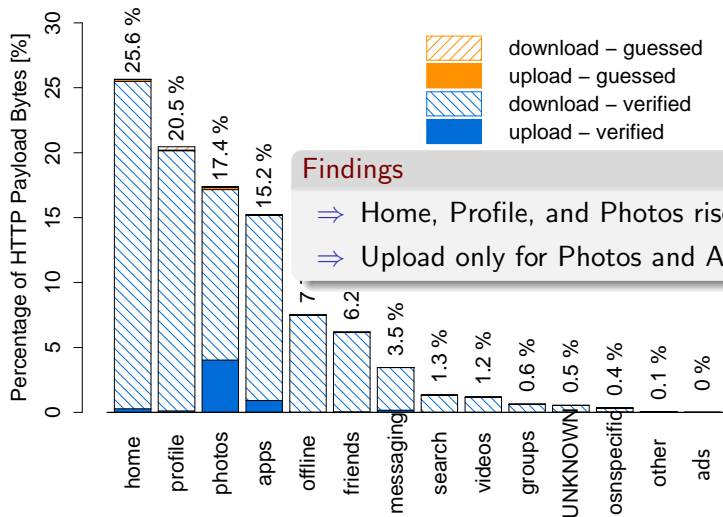
# Volume per Category

Active <u>and indirect</u> Facebook rr-pairs by category for ISP-A2

# Volume per Category

Active <u>and indirect</u> Facebook rr-pairs by category for ISP-A2



Legend:
- download – guessed
- upload – guessed
- download – verified
- upload – verified

Y-axis: Percentage of HTTP Payload Bytes [%]

Bar values:
- home: 25.6 %
- profile: 20.5 %
- photos: 17.4 %
- apps: 15.2 %
- offline: 7.5 %
- friends: 6.2 %
- messaging: 3.5 %
- search: 1.3 %
- videos: 1.2 %
- groups: 0.6 %
- UNKNOWN: 0.5 %
- osnspecific: 0.4 %
- other: 0.1 %
- ads: 0 %

### Findings

⇒ Home, Profile, and Photos rise in importance

⇒ Upload only for Photos and Apps

# Feature Popularity: Observations

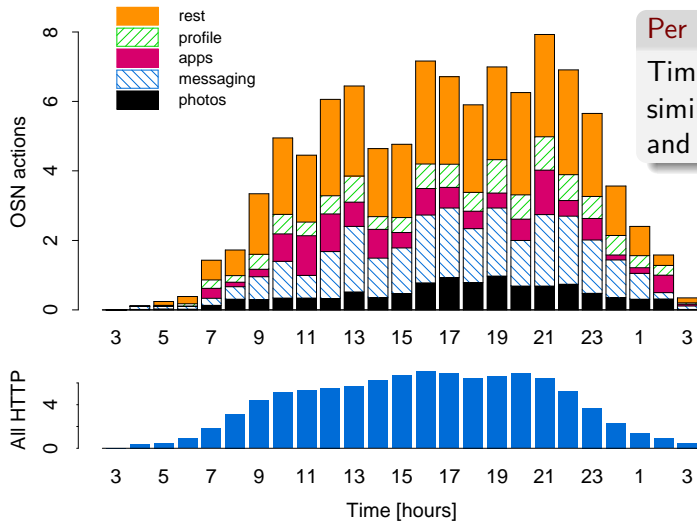Active Facebook rr-pairs per session by category for ISP-A2
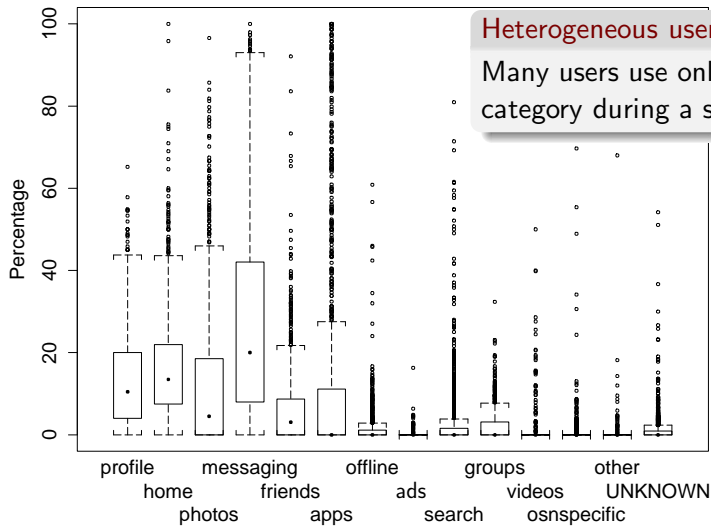


Heterogeneous user base:
Many users use only one feature category during a session.

# Feature Popularity: Observations (cont'd)

## OSN and all HTTP rr-pairs per hour for ISP-A2



Per hour usage:

Time-of-day effects: similar for OSNs and all HTTP

# Requested profiles

Approach:

- Profiles represent a user in an OSN.
  Requests to profiles indicate interest in a user

- We distinguish three types of profiles: own, other, and public

- Method: Count which and how often profiles are requested

## Findings

- Types of profile requests:
  - Majority to profiles of other users, 25-35% to own profile,
  - 12% (22%) to public profiles: Facebook Pages (LinkedIn)
- Profile requests per Facebook session:
  - mean number of requested profiles: 6
  - unique profiles: **only 3**

# Outline

**1** Approach

**2** Session Characteristics

**3** Feature Popularity

**4** Dynamics within Sessions
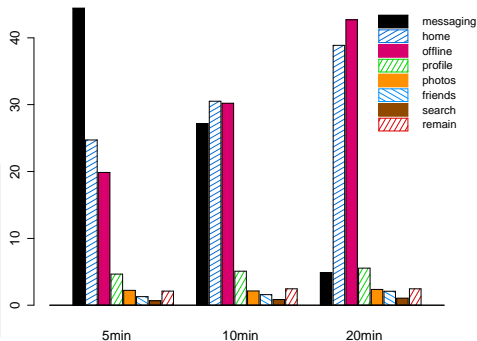
**5** Conclusions

# Activity vs. Inactivity Periods

Apply within session inactivity timeout of 5min:

⇒ Sessions >1min: 50 % of users are active all time

⇒ Sessions >40min: >95 % have inactivity periods
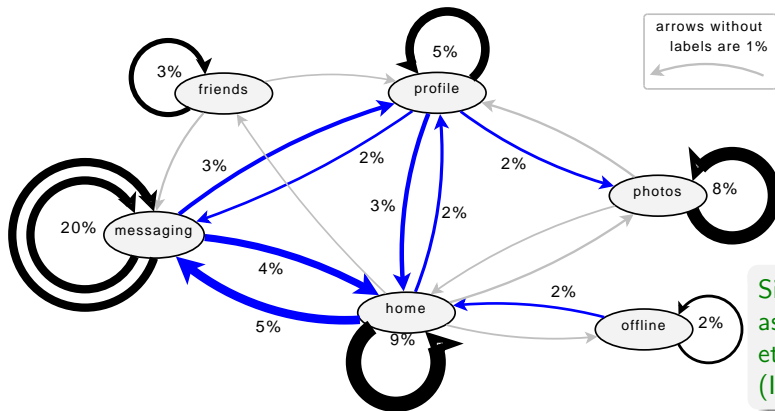
Action after inactivity
- Top categories: Messaging, Home, Offline
- Distribution changes with the length of the pause

Facebook action after inactivity period for ISP-A2

# Feature Sequences

Click sequences of Facebook for ISP-A2: Global transition probabilities



arrows without labels are 1%

Similar findings as Benevenuto et al for Orkut (IMC'09)

## Findings

⇒ Messaging traps users; Home, Photos and Profile attract users to stay

# Summary

Findings:

- Most of the sessions are short (few minutes) and small in terms of volume (several MBytes)
- Long sessions are dominated by inactivity periods
- Top action categories are: Messaging, Apps, Home, Profile, and Photos.
- Facebook users are trapped by Messaging and Photos

### Future Work

- Expand analysis to other OSNs/external apps, and overcome caveats
- Evaluate new OSN designs with OSN user model (e. g., PeerSoN[a])

---

[a]`www.peerson.net`

# Summary

Findings:

- Most of the sessions are short (few minutes) and small in terms of volume (several MBytes)
- Long sessions are dominated by inactivity periods
- Top action categories are:
  Messaging, Apps, Home, Profile, and Photos.
- Facebook users are trapped by Messaging and Photos

**Time for Questions**

**?**

### Future Work

- Expand analysis to other OSNs/external apps, and overcome caveats
- Evaluate new OSN designs with OSN user model (e. g., PeerSoN[a])

[a]www.peerson.net