Abstract

Introduction

Data Description

We built a Facebook App[1] to collect information about users, their interactions and preferences. Our dataset contains information about each App user, along with a subset of information about their friends visible to the App. The data collection is performed with full permission from the user and in accordance with an approved Ethics Protocol[2].

Over 200 users installed the Facebook App sometime during the evaluation period. At any time, around 100 users have actively used the App. From these core App users, the App has access to their detailed Facebook profiles and their interactions with a total of 39,850 friends. While we have complete interaction data for the App users with their friends, and profile data (including wall post data) for the App users and friends, we do not have complete interactions for the App users' friends (unless they themselves are App users). Hence in the forthcoming analysis, we limit our evaluation to App users for which we are assured to have full interaction data.

Our App tracks many user (and their friends') details and interactions on Facebook. Interactions that occur through wall posts provide a rich variety of content and interaction data. We distinguish four Facebook items from wall posts: general posts (e.g., status updates, activity updates such as new friends, and interactions such as the user liked these pages), links, photos and videos. Four main interactions on these items are permitted by Facebook: posting an item to a friend's wall, commenting, liking, and tagging[3]. The App does not track deletions of these items and interactions (e.g., unlike) for performance reasons and we found very few deletions during an initial testing stage.

We summarize relevant basic statistics of the data in Table 1-3 below. The tables distinguish the data from the App users and from all App users and friends. Table 1 summarizes the number of records for each item (row) and interaction (column) combination. Table 2 shows some demographics from user profiles[4].

Methodology

Membership feature

$$\mathrm{X}_{uij} = \begin{cases} True & u \in \text{Member(group}_j) \ \& \\ & \exists v \in \text{Member(group}_j)\ likes(v,i) \\ false & otherwise \end{cases}$$

[1]Name and link omitted for anonymity.

[2]Link omitted for anonymity.

[3]Some Facebook interaction features such as liking comments were introduced after App user studies began and so are not tracked.

[4]Note that count of schools are not unique as each user can attend more than one degree of the same type.

App Users	Posts	Tags	Comments	Likes
Wall	27,955	5,256	15,121	11,033
Link	3,974	—	5,757	4,279
Photo	4,147	22,633	8,677	5,938
Video	211	2,105	1,687	710
App Users and Friends	**Posts**	**Tags**	**Comments**	**Likes**
Wall	3,384,740	912,687	2,152,321	1,555,225
Link	514,475	—	693,930	666,631
Photo	1,098,679	8,407,822	2,978,635	1,960,138
Video	56,241	858,054	463,401	308,763

Table 1: Number of records in Items and Interactions Tables. Rows are type of Facebook item and columns are type of Facebook interaction.

Table	#Records (App Users)	#Records (App User and Friends)
Users	103	39,850
Column	**#Non-empty (App Users)**	**#Non-empty (App User and Friends)**
Gender	102	36,401
Birthday	103	27,624
Breakdown	**Count (App Users)**	**Count (App User and Friends)**
Male	73	19,742
Female	29	16,659
High School	104	29,503
College	115	29,223
Graduate School	56	7733

Table 2: App user demographics.

Table	#Records (App Users)	#Records (App User and Friends)
Groups	3,469	373,608
Page Likes	10,771	825,452
Favourites	TBA	TBA

Table 3: Groups of Interests Tables

	Friend recommendation	Non-Friend recommendation
Like	1844	1393
Dislike	1359	2807

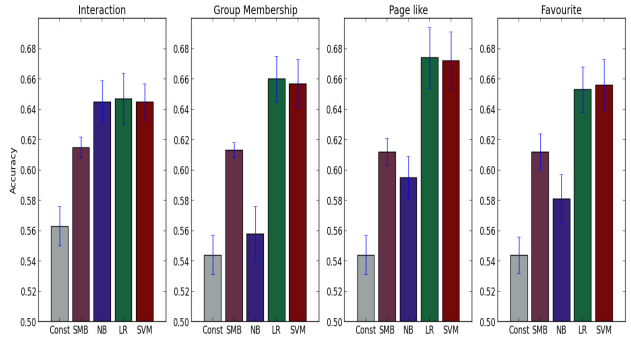

Figure 1: Accuracy plots

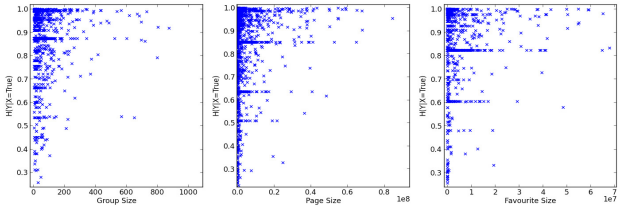

Figure 2: Conditional Entropy vs Size

Evaluation

- Page likes are most predictive followed by group and favourites
- Mutual information vs Size plot shows that the most predictive groups/page/favourites are spread around middle. This signifies that medium sized groups/pages/favourites are more predictive (!!!although not very strong signal)
- Collapsed mutual information ranking
  Video > Photo > Post > Link
  Comments > Like > Tags
  (No significant difference between Incoming and Outgoing)
- For Large social networks encoding user's membership(group/pages/favourites) as features and performing matrix factorization is not scalable (as there are millions of groups/pages/activities). This research shows that the memebership can be directly plugged into existing highly scalable algorithms to achieve better oaccuracy than state of art matrix factorization techniques.

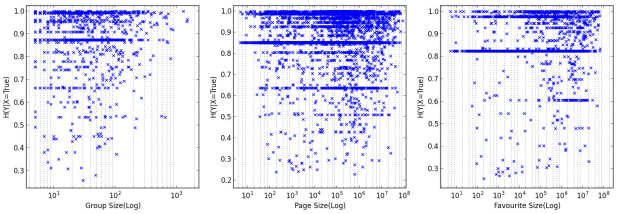

Figure 3: Conditional Entropy vs Size(log)

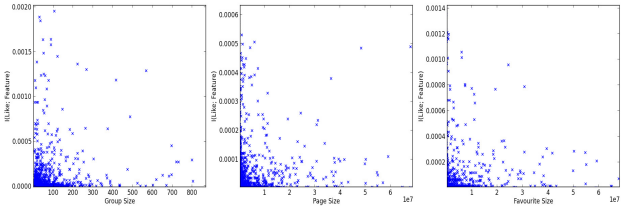

Figure 4: Mutual Information vs Size

Related Work

Conclusions

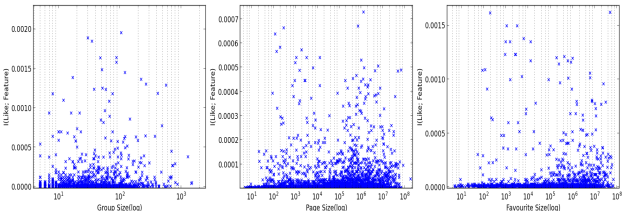

Figure 5: Mutual Information vs Size(log)

	Mutual Information
video	8.0947e-04
photo	1.2251e-04
post	2.4799e-05
link	1.5086e-05

	Conditional Entropy
photo	0.9839
post	0.9872
link	0.9897
video	0.9982

	Mutual Information
comments	6.7656e-05
likes	2.8765e-05
tags	4.6544e-06

	Conditional Entropy
likes	0.9871
comments	0.9879
tags	0.9913

	Mutual Information
Outgoing	5.7508e-05
Incoming	5.7508e-05

	Conditional Entropy
Incoming	0.98696
Outgoing	0.99018

Interaction	Conditional Entropy
FRIENDS	0.9350
VIDEO_TAGS_OUTGOING	0.9425
VIDEO_TAGS_INCOMING	0.9459
VIDEO_LIKES_OUTGOING	0.9755
PHOTO_COMMENTS_INCOMING	0.9796
PHOTO_LIKES_INCOMING	0.9868
POST_COMMENTS_OUTGOING	0.9870
POST_COMMENTS_INCOMING	0.9893
POST_LIKES_INCOMING	0.9906
PHOTO_TAGS_INCOMING	0.9907
LINK_COMMENTS_OUTGOING	0.9914
PHOTO_LIKES_OUTGOING	0.9932
LINK_LIKES_INCOMING	0.9937
POST_LIKES_OUTGOING	0.9942
VIDEO_LIKES_INCOMING	0.9942
VIDEO_COMMENTS_INCOMING	0.9959
POST_TAGS_OUTGOING	0.9972
LINK_LIKES_OUTGOING	0.9973
VIDEO_COMMENTS_OUTGOING	0.9975
POST_TAGS_INCOMING	0.9979
LINK_COMMENTS_INCOMING	0.9984
PHOTO_COMMENTS_OUTGOING	0.9999
PHOTO_TAGS_OUTGOING	0.9999

Interaction	Mutual Information
FRIENDS	4.813e-03
POST_TAGS_INCOMING	8.239e-04
PHOTO_TAGS_OUTGOING	4.347e-04
LINK_COMMENTS_INCOMING	2.281e-04
LINK_LIKES_OUTGOING	2.156e-04
PHOTO_COMMENTS_OUTGOING	1.622e-04
VIDEO_LIKES_OUTGOING	1.597e-04
POST_TAGS_OUTGOING	1.406e-04
VIDEO_LIKES_INCOMING	1.258e-04
POST_LIKES_OUTGOING	8.051e-05
LINK_LIKES_INCOMING	5.741e-05
VIDEO_TAGS_OUTGOING	5.263e-05
VIDEO_TAGS_INCOMING	4.953e-05
PHOTO_COMMENTS_INCOMING	4.083e-05
VIDEO_COMMENTS_OUTGOING	1.378e-05
VIDEO_COMMENTS_INCOMING	1.129e-05
LINK_COMMENTS_OUTGOING	1.097e-05
POST_LIKES_INCOMING	9.909e-06
PHOTO_LIKES_OUTGOING	8.736e-06
PHOTO_TAGS_INCOMING	5.265e-06
POST_COMMENTS_OUTGOING	3.741e-06
PHOTO_LIKES_INCOMING	1.799e-06
POST_COMMENTS_INCOMING	9.797e-07