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

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.

---

[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	36,359	7,711	22,388	15,999
Link	5,304	—	7,483	6,566
Photo	4,933	28,341	10,976	8,612
Video	245	2,525	1,970	843
App Users and Friends	**Posts**	**Tags**	**Comments**	**Likes**
Wall	4,301,306	1,215,382	3,122,019	1,887,497
Link	678,612	—	891,986	995,214
Photo	1,268,816	9,620,708	3,431,321	2,469,859
Video	59,244	904,604	486,677	332,619

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	119	38,378
Column	**#Non-empty (App Users)**	**#Non-empty (App User and Friends)**
Gender	118	37,872
Birthday	118	28,633
Breakdown	**Count (App Users)**	**Count (App User and Friends)**
Male	85	20,840
Female	33	17,032

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	4,284	892,820

Table 3: Groups of Interests Tables

	Friend recommendation	Non-Friend recommendation
Like	1392	1127
Dislike	895	2111
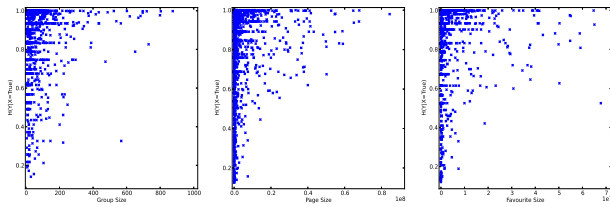
Figure 5: Conditional Entropy vs Size

Figure 7: Mutual Information vs Size

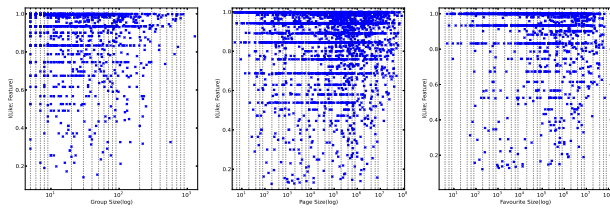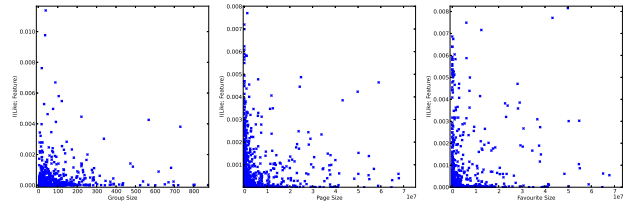Figure 6: Conditional Entropy vs Size(log)

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 accuracy than state of art matrix factorization techniques.

Figure 8: Mutual Information vs Size(log)

Figure 1: Accuracy plots

Figure 2: Mutual Information vs Group Size

Figure 3: Mutual Information vs Page Size

Figure 4: Mutual Information vs Favourite Size

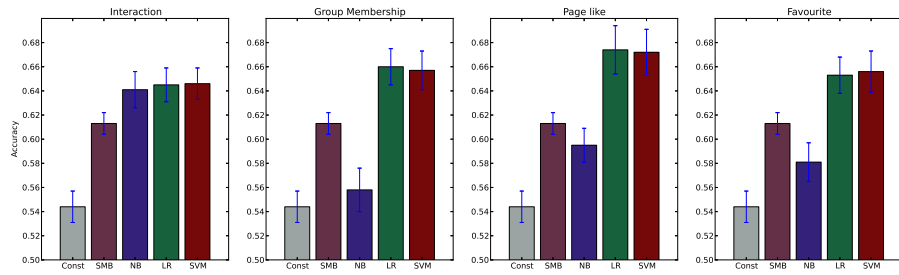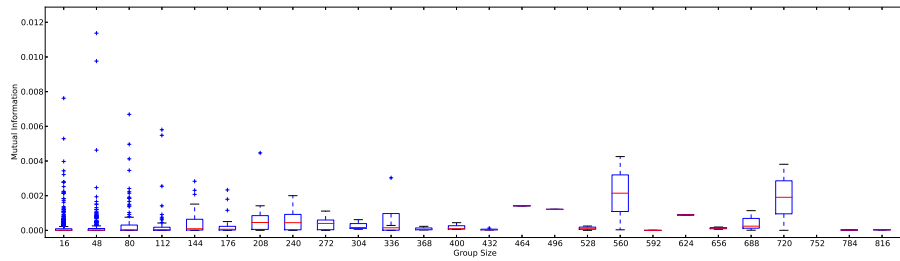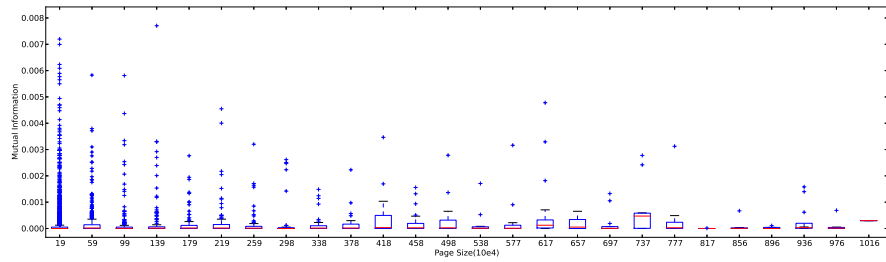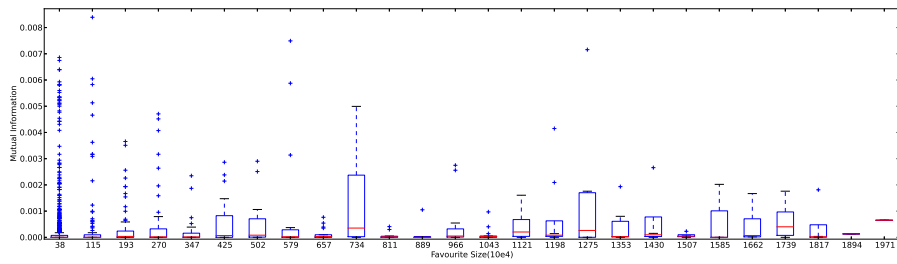| Modality	ConditionalEntropy	(Like,True)	(Dislike,True)	(Like,False)	(Dislike,False)	P(like—True)
video	0.850116919421	117	44	2402	2962	0.7267
link	0.914700608855	989	486	1530	2520	0.6705
post	0.918160390299	1154	576	1365	2430	0.6671
photo	0.9259513318	675	349	1844	2657	0.6591
Modality	MutualInformation	(Like,True)	(Dislike,True)	(Like,False)	(Dislike,False)	P(like—True)
post	0.0594688018722	1154	576	1365	2430	0.6671
link	0.0490133947836	989	486	1530	2520	0.6705
photo	0.0273700186713	675	349	1844	2657	0.6591
video	0.0064494273283	117	44	2402	2962	0.7267
Type	Conditional Entropy	(Like,True)	(Dislike,True)	(Like,False)	(Dislike,False)	P(like—True)
Tags	0.919971597996	809	407	1710	2599	0.6653
Comments	0.921373670369	1009	511	1510	2495	0.66382
Likes	0.924414408934	1135	583	1384	2423	0.6607
Type	Mutual Information	(Like,True)	(Dislike,True)	(Like,False)	(Dislike,False)	P(like—True)
Likes	0.0553328447681	1135	583	1384	2423	0.6607
Comments	0.0479360954681	1009	511	1510	2495	0.66382
Tags	0.0361031572103	809	407	1710	2599	0.6653
Direction	Mutual Information	(Like,True)	(Dislike,True)	(Like,False)	(Dislike,False)	P(like—True)
Outgoing	0.049470651248	1074	562	1445	2444	0.6565
Incoming	0.0470981200647	1081	584	1438	2422	0.64924
Direction	Conditional Entropy	(Like,True)	(Dislike,True)	(Like,False)	(Dislike,False)	P(like—True)
Outgoing	0.928353525673	1074	562	1445	2444	0.6565
Incoming	0.934921690705	1081	584	1438	2422	0.64924

Interaction	Conditional Entropy	(Like,True)	(Dislike,True)	(Like,False)	(Dislike,False)	P(like—True)
TAGS_OUTGOING	0.885214556947	661	287	1858	2719	0.6973
LIKES_OUTGOING	0.885367017942	911	396	1608	2610	0.6970
TAGS_INCOMING	0.899959283697	653	301	1866	2705	0.6845
LIKES_INCOMING	0.901653997367	986	458	1533	2548	0.6828
COMMENTS_OUTGOING	0.908080562132	790	377	1729	2629	0.6769
COMMENTS_INCOMING	0.911567201865	840	407	1679	2599	0.6736

LIKES_INCOMING	0.0532717033854	986	458	1533	2548	0.6828
LIKES_OUTGOING	0.0527680578968	911	396	1608	2610	0.6970
COMMENTS_INCOMING	0.0402962708309	840	407	1679	2599	0.6736
COMMENTS_OUTGOING	0.0381700399797	790	377	1729	2629	0.6769
TAGS_OUTGOING	0.035303106185	661	287	1858	2719	0.6973
TAGS_INCOMING	0.0318329570229	653	301	1866	2705	0.6845

Interaction	Conditional Entropy	(Like,True)	(Dislike,True)	(Like,False)	(Dislike,False)	P(like—True)
PHOTO_OUTGOING	0.856939074179	521	203	1998	2803	0.7196
VIDEO_OUTGOING	0.863258948786	69	27	2450	2979	0.7188
LINK_OUTGOING	0.895377102746	809	366	1710	2640	0.6885
LINK_INCOMING	0.895672876337	797	361	1722	2645	0.6882
POST_INCOMING	0.901803982795	1007	468	1512	2538	0.6827
POST_OUTGOING	0.905555107759	1006	475	1513	2531	0.6793
VIDEO_INCOMING	0.915141339321	68	33	2451	2973	0.8395
PHOTO_INCOMING	0.920746179598	563	284	1956	2722	0.6647

Interaction	Mutual Information	(Like,True)	(Dislike,True)	(Like,False)	(Dislike,False)	P(like—True)
POST_INCOMING	0.0547156102171	1007	468	1512	2538	0.6827
POST_OUTGOING	0.0533330284575	1006	475	1513	2531	0.6793
LINK_OUTGOING	0.0426702641841	809	366	1710	2640	0.6885
LINK_INCOMING	0.0417931651624	797	361	1722	2645	0.6883
PHOTO_OUTGOING	0.0307864066352	521	203	1998	2803	0.7196
PHOTO_INCOMING	0.0229340503984	563	284	1956	2722	0.6647
VIDEO_OUTGOING	0.0035452127605	69	27	2450	2979	0.7188
VIDEO_INCOMING	0.00253337061407	68	33	2451	2973	0.8395

Interaction	Conditional Entropy	(Like,True)	(Dislike,True)	(Like,False)	(Dislike,False)	P(like—True)
VIDEO_LIKES_OUTGOING	0.722397472066	31	7	2488	2999	0.8158
VIDEO_LIKES_INCOMING	0.775086171837	43	12	2476	2994	0.7818
POST_TAGS_INCOMING	0.802855788214	444	143	2075	2863	0.7564
PHOTO_COMMENTS_OUTGOING	0.816036600382	135	45	2384	2961	0.75
PHOTO_TAGS_OUTGOING	0.818540125915	427	145	2092	2861	0.7465
PHOTO_LIKES_INCOMING	0.853009353418	277	106	2242	2900	0.7232
LINK_LIKES_OUTGOING	0.855368864256	728	282	1791	2724	0.7208
POST_TAGS_OUTGOING	0.856163233945	505	196	2014	2810	0.7204
LINK_COMMENTS_INCOMING	0.865533720447	530	213	1989	2793	0.7133
POST_LIKES_OUTGOING	0.870618130159	834	342	1685	2664	0.7092
VIDEO_COMMENTS_OUTGOING	0.883596675997	43	18	2476	2988	0.7049
LINK_LIKES_INCOMING	0.888406184542	742	326	1777	2680	0.6948
POST_LIKES_INCOMING	0.889444522106	929	410	1590	2596	0.6938
POST_COMMENTS_INCOMING	0.890449138325	763	338	1756	2668	0.6930
PHOTO_TAGS_INCOMING	0.890856994354	485	215	2034	2791	0.6928
POST_COMMENTS_OUTGOING	0.894129257133	694	312	1825	2694	0.6899
LINK_COMMENTS_OUTGOING	0.895134128697	543	245	1976	2761	0.6891
PHOTO_LIKES_OUTGOING	0.89759612767	161	73	2358	2933	0.6880
PHOTO_COMMENTS_INCOMING	0.906541518549	290	137	2229	2869	0.6792
VIDEO_COMMENTS_INCOMING	0.92652449307	53	27	2466	2979	0.6625
FRIENDS	0.965942963751	1392	895	1127	2111	0.6087
VIDEO_TAGS_INCOMING	0.999999611934	17	17	2502	2989	0.5
VIDEO_TAGS_OUTGOING	0.999999611934	16	16	2503	2990	0.5

Interaction	Mutual Information	(Like,True)	(Dislike,True)	(Like,False)	(Dislike,False)	P(like—True)
POST_LIKES_INCOMING	0.0524561740233	929	410	1590	2596	0.6938
POST_LIKES_OUTGOING	0.050407410911	834	342	1685	2664	0.7092
FRIENDS	0.0473539909731	1392	895	1127	2111	0.6087
LINK_LIKES_OUTGOING	0.0457244144031	728	282	1791	2724	0.7208
POST_COMMENTS_INCOMING	0.0405155210996	763	338	1756	2668	0.6930
LINK_LIKES_INCOMING	0.0396004717225	742	326	1777	2680	0.6948
POST_COMMENTS_OUTGOING	0.0352609340207	694	312	1825	2694	0.6899
POST_TAGS_INCOMING	0.0315770377754	444	143	2075	2863	0.7564
LINK_COMMENTS_INCOMING	0.0299391900579	530	213	1989	2793	0.7133
POST_TAGS_OUTGOING	0.0296108833783	505	196	2014	2810	0.7204
PHOTO_TAGS_OUTGOING	0.0286060075004	427	145	2092	2861	0.7465
LINK_COMMENTS_OUTGOING	0.0261662648974	543	245	1976	2761	0.6891
PHOTO_TAGS_INCOMING	0.0235803030398	485	215	2034	2791	0.6928
PHOTO_LIKES_INCOMING	0.0155060725343	277	106	2242	2900	0.7232
PHOTO_COMMENTS_INCOMING	0.0120447345005	290	137	2229	2869	0.6792
PHOTO_COMMENTS_OUTGOING	0.00851951888057	135	45	2384	2961	0.75
PHOTO_LIKES_OUTGOING	0.00687365820792	161	73	2358	2933	0.6880
VIDEO_LIKES_INCOMING	0.00309679794205	43	12	2476	2994	0.7818
VIDEO_LIKES_OUTGOING	0.00259896191026	31	7	2488	2999	0.8158
VIDEO_COMMENTS_OUTGOING	0.00197259388552	43	18	2476	2988	0.7049
VIDEO_COMMENTS_INCOMING	0.0017837362825	53	27	2466	2979	0.6625
VIDEO_TAGS_INCOMING	3.598735115e-05	17	17	2502	2989	0.5
VIDEO_TAGS_OUTGOING	3.39757297236e-05	16	16	2503	2990	0.5

	id	name	size	egosize	Cond Entropy	Mutual Information
1620	22934684677	The Big Bang Theory	24382711	2339	0.898472	0.018366
558	29534858696	The Simpsons	59144063	1439	0.786530	0.017291
1546	9588466619	Futurama	24939188	1131	0.818993	0.017171
630	494637910144	The IT Crowd	937445	565	0.790827	0.016082
1678	24609282673	Family Guy	49865912	2036	0.804842	0.015698
37	6708787004	South Park	43243922	1444	0.855054	0.015199
1425	29112278285	Scrubs	12543319	1532	0.861083	0.014135
24	28627688223	MythBusters	6060917	597	0.460559	0.013485
1409	163542213701619	Avascular Necrosis	195	36	0.114687	0.012040
1304	21570237112	Blind Guardian	422327	54	0.328430	0.012004
328	292772493767	Tortured	1082	32	0.120693	0.011223
685	115344841825821	Elysian	3195	36	0.120693	0.011223
1687	76613428183	Community	1369762	304	0.760992	0.010984
230	7496603409	Opeth	997062	155	0.260887	0.010700
1372	19058419696	Pantera	6125926	220	0.353390	0.010469
39	149992358386897	Anno Domini	6306	21	0.127431	0.010408
745	108511629178228	Hellbringer	1533	25	0.129247	0.010204
1388	170724093400	Johnny Roadkill	973	25	0.129247	0.010204
1445	17222896745	Darker Half	12964	25	0.129247	0.010204
848	186487216175	Robert The Bruce	220	69	0.539975	0.010013

Table 4: Top 20 Favourite ranked by Mutual Information

	id	name	size	egosize	Cond Entropy	Mutual Information
1057	130815873610129	ANU StalkerSpace	1292	1292	0.861615	0.030603
778	239607379457024	ANU CSSA	38	38	0.745879	0.023007
424	2397818353	CSSA	35	35	0.600649	0.019016
1422	215525841848413	Australian National Uni AI & ML	18	18	0.785032	0.015477
720	324570326277	ANU Engineering Students' Association (ANUESA)...	88	88	0.471934	0.012753
602	271451673295	Canberra Rock Gigs	105	105	0.200642	0.010805
96	11891033953	Ekta - Indian Subcontinent Students' Associati...	121	121	0.420844	0.010377
693	219745416919	Heavy Metal - Australian Capital Territory, ...	30	30	0.133053	0.009798
810	213225102082	The Great Australian Internet Blackout	76	76	0.491275	0.009466
865	208417795185	Stephen Conroy Should Not Filter Our Internet!	48	48	0.146109	0.008583
1232	69858518886	Our Hero: Clem Baker-Finch	91	91	0.852441	0.008526
707	2259054917	i feel my phone vibrate when it doesn't.	222	222	0.301406	0.008358
272	121627904865	Lift ACT ban on fireworks	568	568	0.309571	0.007971
174	2336048220	I grew up in Australia in the 90s	731	731	0.803557	0.007696
1399	130385273640777	Silicone Stripper	17	17	0.162342	0.007374
914	105714109273	Canberra Music	67	67	0.276221	0.006458
387	243732165652389	iDiscount ANU	338	338	0.887847	0.006378
603	300607567325	Hardcore dancing is not moshing	8	8	0.179274	0.006372
394	206485584139	Metal bands come to Canberra cause I'm sick of...	12	12	0.187194	0.005973
90	19689998305	1,000,000 Strong Against High School Musical	146	146	0.309571	0.005299

Table 5: Top 20 Group ranked by Mutual Information

	id	name	size	egosize	Cond Entropy	Mutual Information
1091	148551616293	The Australian National University	10522	917	0.926747	0.026323
2268	399805615144	Australian National University Students' Assoc...	3059	389	0.788451	0.026124
4050	192286270282	Humans vs Zombies @ ANU	554	212	0.841817	0.024231
3706	174936642528981	ANU Engineering Students' Association 2011 (AN...	285	141	0.504845	0.024180
128	192078840839006	ANU Stalkerspace	5085	1065	0.948173	0.022915
3047	207140035973127	ANU Computer Science Students' Association (AN...	165	85	0.873524	0.022711
4456	105586819475691	Australian National University	11622	454	0.783520	0.021732
3202	159352257471339	ANU ducks	1445	372	0.858136	0.021594
2858	254743647903977	Robogals Canberra	135	32	0.501452	0.020362
3240	293434554026394	ANU O-Week 2012: Escape to the East	1483	240	0.829381	0.018705
3055	22934684677	The Big Bang Theory	24355106	2552	0.899177	0.018333
4083	9588466619	Futurama	24899564	1213	0.811316	0.018158
3780	171648066192062	ANU XSA	421	123	0.858345	0.017345
1828	157065757691727	Angie To Photography	203	63	0.646059	0.016147
1046	494637910144	The IT Crowd	936956	578	0.793751	0.016003
63	28627688223	MythBusters	6059381	648	0.429793	0.015709
3429	290539813359	Trust Me, I'm an "Engineer"	2139099	283	0.538777	0.015233
1736	264476006941706	Potential	116	37	0.214530	0.014676
3768	221086387944225	ANU/Canberra Python Social Club	21	9	0.741611	0.013916
2855	273323547059	ANUSA Euro O-Week 2010	1453	428	0.546198	0.013857

Table 6: Top 20 Page ranked by Mutual Information

Related Work
Conclusions