

CrossOSN-U: Event

This sub-dataset consists of two sql tables for Twitter users and three sql tables for YouTube users. The overlapped users are identified from the 'googlePlusUserId' attribute in the sql table.

➤ *Twitter tables:*

Table 1 – users' historical tweeting behavior table (*twitteruseractivityinfo.sql*)

This table contains the tweeting behaviors of all the 8,540 overlapped users in our dataset (this table is not provided publicly due to privacy reason). A snapshot of this table is presented as below:

tweetId	twitterUserId	text	created_at	hashtags	user_mentions	eventIndex
160905925698404352	13056112	Going to Santa Monica tonight.	Sun Jan 22 02:05:36 +0000 2012			11
160905946535702528	9520692	Photo: http://t.co/pMOH11NO	Sun Jan 22 02:05:41 +0000 2012			(Null)
160905953577943041	82359540	I liked a @YouTube video http://t.co/pMOH11NO	Sun Jan 22 02:05:43 +0000 2012		10228272	(Null)
160906017440415746	132167822	He publicado una foto nueva e	Sun Jan 22 02:05:58 +0000 2012			(Null)
160906034972602371	33043947	RT @CNMLive: CNN projects th	Sun Jan 22 02:06:02 +0000 2012	NewtGingrich,SCPrimary	9245812	(Null)
160906143772848128	3549531	@leeunkrich When are we goir	Sun Jan 22 02:06:28 +0000 2012		13236602	(Null)
160906225012322304	10007212	RT @tcarmody: Here, I briefly	Sun Jan 22 02:06:48 +0000 2012		14439168,18751963,1110717	(Null)
160906229290504192	7532972	@redlipslove Awesome! Let me	Sun Jan 22 02:06:49 +0000 2012		14886693	(Null)
160906245258231810	177243612	RT @chikasplus: Partida De La	Sun Jan 22 02:06:53 +0000 2012		362882220	(Null)
160906246990479362	262116170	R.I.P. Coach RT @BreakingNe	Sun Jan 22 02:06:53 +0000 2012		6017542	(Null)
160906262646241280	1037311	This thing is crazy...40-26 Ginc	Sun Jan 22 02:07:01 +0000 2012			1
160906267671029761	33043947	Wish I could see the US vs Cer	Sun Jan 22 02:07:03 +0000 2012			(Null)
160906340422791169	22376946	Wow! Not sure if it's true, but	Sun Jan 22 02:07:15 +0000 2012		87791277	(Null)

Figure 1: A snapshot of Twitter users' tweeting behavior table

where 'tweetId' is the unique tweet identifier, 'text' is the textual content of this tweet, 'hashtags' is kind of symbol to mark keywords or topics in a tweet (start by #, and also separated by the ',' separator), 'user_mentions' indicates who this user has ever mentioned in his/her tweets. For the 'eventIndex', we adopted the event extraction method in our iciap paper to identify the event for each tweet, and give the index (i.e. 1-20) of the event for each tweet.

Table 2 – the involved Twitter user list for each event (*involvedTwitterUserlistFor EachEvent.sql*)

For each considered event (i.e. 1-20), we listed all the involved users who have ever paid attention to this event on Twitter. A snapshot of this table is presented:

EventIndex	InvolvedUserlist_TwitterId	InvolvedUserlist_GooglePlusId	InvolvedUserNum
1	50771752,26417366,1501731	105830943954405391380,111	2908
2	9505692,26417366,15017316	110605428111103638712,10E	3847
3	8381062,26417366,9505692,;	111028119742567900608,10E	2347
4	172398772,34111776,166725	111028119742567900608,10E	1367
5	10138412,20466813,2732411	115289562690092131459,11C	1061
6	26417366,15017316,5622042	105830943954405391380,10E	2377
7	26417366,15017316,5622042	105830943954405391380,111	2238
8	8381062,14181778,5622042,t	110605428111103638712,11E	846

Figure 2: A snapshot of involved Twitter user list for each event

where 'InvolvedUserlist_TwitterId' and 'InvolvedUserlist_GooglePlusId' indicate all the involved Twitter user ids and their corresponding google plus user ids for each event, respectively. Note that the twitter user ids and google plus user ids in this table are also separated by ',' separator.

➤ **YouTube tables:**

Table 3 – users’ historical video behavior table (*youtubeuseractivityinfo.sql*)

This table contains the historical YouTube video behaviors of all the 8,540 overlapped users in our dataset. A snapshot of this table is presented as below:

Id	googlePlusUserId	youtubeUserName	behavior	videoid	date	videotags	eventIndex
706030	110763343369464388505	zymgy	RATED	3Pu-bWrdgY	2012-12-16 23:58	Frosty The Snowman, Jimmy D	11
706031	110763343369464388505	zymgy	RATED	4AnMIQNW1M8	2012-12-16 23:50	Alphabet Photography, Musica	11
706032	110763343369464388505	zymgy	RATED	4HLV1Nt04M	2013-04-09 21:40	Mark, Ronson, featuring, Amy	(N/A)
706033	110763343369464388505	zymgy	RATED	52_8kalDj_8	2013-04-18 15:42	Moon, Entwistle, Daltry, Town	(N/A)
706034	110763343369464388505	zymgy	RATED	5A12kbW4vPQ	2013-04-18 15:38	Led, Zeppelin, Communication,	(N/A)
706035	110763343369464388505	zymgy	RATED	5Ksl-v775SQ	2013-03-06 18:22	orvillecopter, cat turned into h	(N/A)
706036	110763343369464388505	zymgy	RATED	5m-a00sKzY	2013-03-19 15:53	Blues, Etta, James, Jimmy, Re	(N/A)
706037	110763343369464388505	zymgy	RATED	9wGLPW9-cj0	2013-04-09 21:20	Chet, Atkins, Jethro, Burns, H	(N/A)
706038	110763343369464388505	zymgy	RATED	9yZyCtJqIQ	2013-03-31 18:23	Something, Stupid, Original son	(N/A)
706039	110763343369464388505	zymgy	RATED	ae-e8fay3CE	2011-12-22 15:03	mathematics, multiplication	(N/A)
706040	110763343369464388505	zymgy	RATED	bg5MCzq8	2013-02-08 23:08	Grinderman, Nick Cave, Warr	(N/A)
706041	110763343369464388505	zymgy	RATED	BINOUZzh6Ac	2012-11-28 23:41	the sonics, 1960s music hits, g	(N/A)
706042	110763343369464388505	zymgy	RATED	bly6w_Jub5s	2012-11-28 17:20	dumb ways to die, parody, cur	(N/A)
706043	110763343369464388505	zymgy	RATED	CMILc-rSL64	2012-12-06 13:01	Christmas	11
706044	110763343369464388505	zymgy	RATED	dcoWKLFB8x8	2013-04-18 16:19	, ORIANTHI, WOODOO, CHILE	(N/A)

Figure 3: A snapshot of YouTube users’ video behavior table

which includes user’s five kinds of typical video behaviors with timestamp: rated, favorited, subscribed, commented and uploaded. We also label the event index of each video behavior using the same method just as in the Twitter table 1.

Table 4 – the involved YouTube user list for each event (*involvedYouTubeUserlistFor EachEvent.sql*)

For each considered event (i.e. 1-20), we listed all the involved users who have ever paid attention to this event on YouTube. A snapshot of this table is presented:

EventIndex	InvolvedUserlist_YouTubeName	InvolvedUserlist_GooglePlusId	InvolvedUserNum
1	Fantt,arconna,scolford,gianpadano,	107574652535084661746,1138689	1304
2	atotic,xXSIMPXx,shaunmoss,tinker4f	109835747955449700376,1010353	1401
3	mikenfer,LuekeJohannes,flashmegar	117884007449109471406,1029083	965
4	madjo80,mikenfer,bicyclefeathers,fc	110605428111103638712,1099790	356
5	biarnylr,miniwanRusso11,kebernet,a	107000112862313197264,1008477	645
6	ErdemZen,kebernet,darrenmaharaj,l	109835747955449700376,1178840	1952
7	arconna,alexm7577,robq2251,aworr	109835747955449700376,1075746	933
8	wulfjr,CSerio117,Fantt,TheSholto,bc	115387193683563872497,1070001	802

Figure 4: A snapshot of involved YouTube user list for each event

where ‘InvolvedUserlist_YouTubeName’ and ‘InvolvedUserlist_GooglePlusId’ indicate all the involved YouTube user names and their corresponding google plus user ids for each event, respectively. Note that the YouTube user names and google plus user ids in this table are also separated by ‘,’ separator.

Table 3 –YouTube video metadata table (*youtubevideoinfo.sql*)

This table contains all the textual metadata of the related videos for the users in table 1, including video title, category, tags, description, uploaded time, etc. The data format is the same as the video metadata table of the sub-dataset [CrossOSN-U: Hetero](#).