#mytweet via Instagram: Exploring User Behaviour across Multiple Social Networks

Bang Hui Lim[1] Dongyuan Lu[1] Tao Chen[1] Min-Yen Kan[1,2]
[1]Web IR/NLP Group (WING), School of Computing, National University of Singapore
[2]NUS Interactive and Digital Media Institute, Singapore
limbanghui@u.nus.edu.sg {ludy, taochen, kanmy}@comp.nus.edu.sg

***Abstract*—We study how users of multiple online social networks (OSNs) employ and share information by studying a common user pool that use six OSNs – Flickr, Google+, Instagram, Tumblr, Twitter, and YouTube. We analyze the temporal and topical signature of users' sharing behaviour, showing how they exhibit distinct behaviorial patterns on different networks. We also examine *cross-sharing* (*i.e.*, the act of user broadcasting their activity to multiple OSNs near-simultaneously), a previously-unstudied behaviour and demonstrate how certain OSNs play the roles of originating source and destination sinks.**

***Keywords*—Online Social Networks, cross-sharing, user behaviour**

I. Introduction

Reading and posting from social networks has become a staple daily activity for many. The Pew Internet Project's January 2014 research reported that 74% of online adults use social networking sites[1]. More recently, the Global Web Index's January 2015 Social Report monitored engagement with close to 50 online social networks (OSN) observed an average of 5.54 social media accounts per user, with 2.82 being used actively[2].

Even with so much social network activity and the wide variety of networks, the published literature reports little about their usage. We know precious little about what users actually do on OSNs, aside from our own individual use. Even less is known about how individuals interact across multiple OSNs. Many functionalities across networks are similar. So why do people find themselves using more than one? Are what individuals do on one network the same as their behavior on another? Does participation in one network impact their activity on another?

These questions are important, but yet remain unaddressed by existing literature. With many of today's users being engaged on multiple platforms, do studies limited to individual social network platforms provide a good picture of user behaviour in general? To date, there has not been a definitive answer to this question.

The above are all questions about single networks, but users can use several OSNs simultaneously. One artifact of this is *cross-network sharing* (akin cross-posting in mailing lists) – when people post about their activities on more than one network. To our knowledge, why and how do people do such cross-sharing has not previously been investigated.

We aim to address both issues of multi- and cross-OSN behavior in part in this work. We present an exploratory user-centric study on a large sample of users that participate in multiple OSNs. In particular, we analyse over 15,000 individuals that participate and link their accounts on six platforms – *Flickr, Google+, Instagram, Tumblr, Twitter*, and *YouTube* – through their publicly-available profile descriptions and public sharing activity on these networks. Through our study of this dataset on macro-, meso- and micro-scale analyses, we conclude that single network analysis is limited and does not yield representative holistic patterns. Even though we only study public data, we find support for the claim that each network is different and has a particular social networking niche to fill.

We first review related work in Section II and then describe our dataset in detail in Section III. We then analyse user profiles, posting time, post topic and professions with respect to *multiple networks* in Section IV. Finally, in Section V we turn to *cross-network* interactions, discovering a clear topology of source and sink network relationships among public OSNs.

II. Related Work

Structure, content and user behavior are three major aspects that can be said to characterize research on social networks. Leveraging the user relationship graph, early studies [1], [2] investigated the structural properties among networks such as Flickr, YouTube and Myspace. They confirmed two empirical observations true of other networks; namely, that they follow a power law distribution and exhibit both small-world and scale-free properties. More recent work examined how the user generated content can be leveraged for knowledge discovery by examining networks over time and topic evolution. For example, Althoff *et al.* presented a comprehensive study about the evolution of topics across three online media streams [3].

The third aspect, user behavior, plays a key role. Understanding user behavior is a key modeling problem as it affects the social network structure as well as attempts to best model users themselves. For example, Lerman *et al.* conducted empirical analyses of user activities on Digg and Twitter to assess how it affected dissemination patterns [4]. Other user activities, such as social connection [5], content generation [6],

This research is supported by the Singapore National Research Foundation under its International Research Centre @ Singapore Funding Initiative and administered by the IDM Programme Office.

[1]http://www.pewinternet.org/fact-sheets/social-networking-fact-sheet

[2]http://www.globalwebindex.net/blog/internet-users-have-average-of-5-social-media-accounts

participation in conversation [7], have also been studied to gain better understanding of OSNs.

However, a major shortcoming of these works is that they are limited to macroscopic analyses from a network (graph) perspective. Most works have neglected any analysis of micro- (individuals) or meso- (small aggregrates) levels. In contrast, we perform a user-centric analysis: following the same users across their multiple social networks to uncover how and why users participate in and interact among their multiple social networks.

To understand how user interacts with multiple OSNs, early studies exploited user clickstream data from passively monitored network traffic [8] or a social network aggregator [9]. However, clickstream data is difficult to obtain for cross-network analysis due to privacy reasons, rendering such techniques difficult to execute in practice. While user-generated content (UGC) in part overcomes these problems (unlike web search history, UGC in OSNs is often created by users voluntarily and for public consumption, alleviating privacy concerns), the additional problem of user linkage needs to be solved. Automated user linkage aims to link user accounts among different social networks that belong to the same person in the real world [10], [11].

Given the associated user accounts and collected UGC among different networks, works can then address the subsequent analysis and build downstream application. For example, Kumar *et al.* performed pioneering analyses on users' migration patterns across seven OSNs, providing guidelines to encourage or deter social media traffic [12]. Chen *et al.* examined the extent of personal information revealed by users across multiple OSNs, and found the amount of information revealed in user profiles correlated with occupation and pseudonym use [13].

There are relatively few studies on cross-posting. Many OSNs provide a cross-site linking functionality – linking two accounts on different platforms so that information can be (automatically) shared between them. This is often a mandatory step for users to perform cross-posting. Chen *et al.*, using a data-driven approach, investigated how users cross-linked their Foursquare accounts to other OSNs, namely Facebook and Twitter [14]. Most relevant to cross-posting is work by Ottoni *et al.* [15]. They studied the correspondence and discrepancy in user activities across Twitter and Pinterest. Their key findings were that 1) users often generate content on Pinterest and then share them to Twitter, and 2) users exhibit more focused interests on Pinterest than Twitter. In contrast, our research is based on six OSNs that span a more diverse set of media, and correlates these with analyses from other perspectives (*i.e.*, temporal and professional aspects).

III. DATASET

A user-centric study of cross-OSN behavior requires a collective study of many individual users, each of whom use multiple OSNs. We purposefully sidestep the issue of user linkage to focus our attention on user behavior.

We leverage an OSN aggregation service called *About.me*[3], which enables people to easily create a public online identity

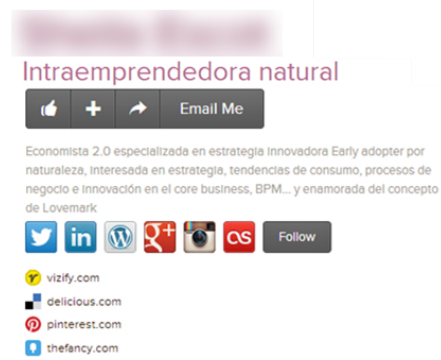


Fig. 1. A sample About.me profile linking to 10 of the user's websites.

that unifies a self-described short biography with prominent links to the person's other OSN accounts and personal websites. According to web analytics provided by Alexa Internet[4], About.me is ranked as the 1,731th most popular website worldwide, which is more popular with female users, users completing graduate education, and with a worldwide user base with a prominent number of visitors in the U.S., India, Spain and Canada, among others. In our own informal analysis of About.me users, we find that they are often people who may benefit from having a stable, open, public and visual presence on the Web for professional reasons: creative, freelancing and marketing types are common job profiles of About.me users.

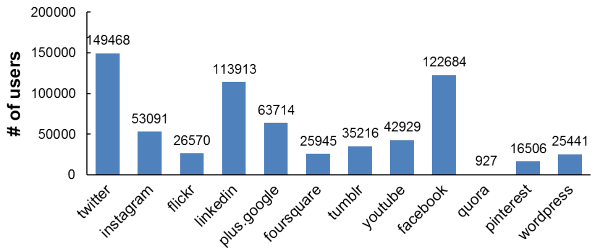
While About.me users are clearly atypical Web users, they do represent an important subset of OSN users that use multiple OSNs and benefit from having a central point to aggregate their activities and publish an overarching biographical sketch of themselves. We argue that such users are important to study as they represent key aggregators and disseminators of information on OSNs, by virtue of needing a service like About.me to manage their distributed activities and identity. Aside from solving the user linkage problem manually for us, About.me importantly 1) offers an application programming interface (API) and 2) ensures that the data captured by their site is publicly available, which are both considerations for programmatic and reproducible analyses.

Using the About.me API, we collected a set of more than $180,000$ registered user profiles. Fig. 1 shows the profile of Jane Doe (not her real name), which links to her OSN accounts and websites (including Twitter, Linked.In and Wordpress). From the cumulative distribution of the number of linked OSN accounts per user, we see that the a slight majority of users list four or more accounts (Fig. 2(b), red line).

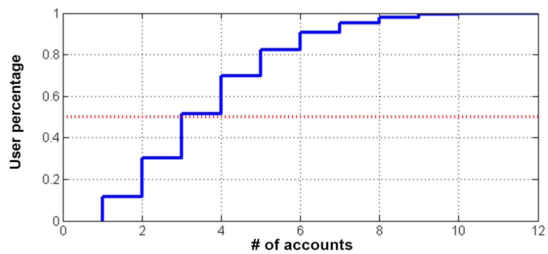
We further limit the dataset used in this paper to users that link to certain OSNs meeting three criteria for inclusion. To ensure our results are representative, we limit our study to the twelve most frequently-linked OSNs (Fig. 2(a)). We further require that the OSN expose most user information publicly through an API, so that we could retrieve user activity; and that the final OSNs chosen represent the breadth of media and functionality common in today's Web 2.0 ecology. With these selection criteria, we selected the $15,595$ users that linked at least the average ($n = 4$, $\bar{n} = 3.7$) number of OSNs from

[3]http://about.me

[4]http://www.alexa.com/siteinfo/about.me, retrieved on 27 April 2015.



(a) # of users per OSN



(b) CDF of # of accounts per user

Fig. 2. # of About.me users linked to the 12 most frequently-linked OSN.

TABLE I. DEMOGRAPHICS OF OUR COMPILED ABOUT.ME DATASET.

OSN	# Users	Activity Type	# Activities
Twitter (Twi)	15,103	microblog	43,042,857
Google+ (G+)	12,445	post	2,522,873
Instagram (Ins)	11,922	photo (upload)	1,054,047
Tumblr (Tum)	10,259	post	8,171,592
Flickr (Fli)	10,139	photo (upload)	11,266,954
YouTube (YT)	8,883	feed (upload)	180,618

the set of six OSNs: *Flickr*, *Google+*, *Instagram*, *Tumblr*, *Twitter* and *YouTube*. Here, we note that *Flickr*, *Instagram* and *YouTube* focus on photo and video sharing, *Twitter* and *Tumblr* are microblogging providers, and *Google+* is a typical social networking site. To compile the dataset used in this study, we crawl each user's publicly accessible activities via the respective APIs on 15 August 2013. Since all of the data we have obtained is public, and since we believe that the compiled dataset is a valuable resource for studying multiple OSN behavior, we have released our dataset for others to conduct further study[5]. Table I gives statistics on the resultant dataset.

A. Statistics

We first calculate the degree of overlap of users of one OSN with the others in our dataset using Jaccard similarity (Table II). Our figures are largely consistent with the previous Pew Internet study that was performed over a global sample of online adults [16]. We see that Twitter has the largest overlap, followed by Google+, revealing their popularities among active social media users who interact with multiple OSNs. Among pairs, we see that YouTube shares 84.1% users with Google+, likely explained by their unique affiliation to Google and easy cross-sharing mechanism; and that overlaps between Instagram and Tumblr are also high (78.8% and 68.5%), which validates previous survey work [17] demonstrating that users prefer OSNs that better support visual media.

[5] http://wing.comp.nus.edu.sg/downloads/aboutme

TABLE II. % OF USERS OF ONE OSN WHO ALSO PARTICIPATE IN ANOTHER OSN IN OUR ABOUT.ME DATASET.

	also use					
	Twi	G+	Ins	Tum	Fli	YT
% of Twi	–	79.4	76.4	65.2	64.4	56.2
% of G+	96.4	–	73.5	61.7	61.0	65.0
% of Ins	96.7	76.8	–	68.5	60.4	51.0
% of Tum	96.0	74.9	78.8	–	59.4	49.2
% of Fli	96.0	74.8	71.0	60.1	–	53.3
% of YT	95.5	84.1	68.4	56.6	60.9	–

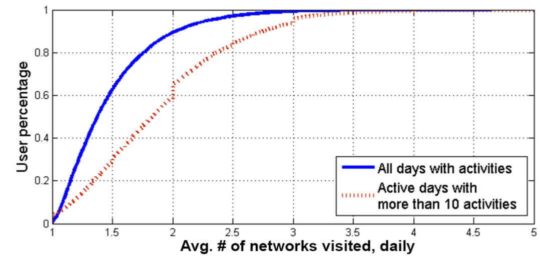


Fig. 3. Cumulative distribution of users' average daily OSN visits (excluding zeros).

Merely linking an OSN account does not necessarily imply that the user actively participates. To gain deeper insight, we filter away dates where users did not use one of the examined OSN. We plot the average number of networks visited daily as a cumulative distribution function (Fig. 3), where a single user's different days of use each contribute one data point. We see that involvement on multiple networks is quite common – $\sim 40\%$ of all users in our dataset are active on any given day, interacting with an average of 1.5 networks. We further examine that on highly active days – *e.g.*, the days with over 10 activities – the average number of networks utilized is also correspondingly larger (dashed line in Fig. 3).

IV. MULTI-NETWORK PARTICIPATION ANALYSIS

Many online social network platforms have similar functions. Being able to follow individuals, post media and comments are pretty much ubiquitous activities that all OSN expose to users. Given this common functionality, why do users choose to use different OSNs? Is it due to their personal social networks (sharing things with different people who serve different social roles), or due to differences in functionality? Our dataset also records users' self-reported profile description (on each network and About.me as well), in addition to historical activity logs of the users' interaction with each OSN. These two data sources allow the analyses to be done on each source, which can be used to triangulate support for conclusions and yield complementary perspectives. Our results from both source corroborate that different OSN platforms are utilised differently, and serve distinct purposes collectively.

A. User Profiles

Online social networks often encourage users to maintain and complete their profiles, possibly to increase users' vested interest in using their network. These optional profiles are typically short sentences (*e.g.*, a tagline) or paragraphs and are publicly accessible, making them a good source of free text demographic information. Here we seek to answer the

TABLE III. USER PROFILES FOR TWO USERS WHO PARTICIPATE IN ALL SIX OSNS. NOUNS USED IN SIMILARITY CALCULATION ARE BOLDED.

User 1	
Twitter	I'm a Digital **Media Specialist** passionate about self **education**, lifelong learning...
Google+	**Knowledge** is **freedom**. I run a **website** called DIY **Genius** that helps young **people** self **education**.
Instagram	Explore **Dream** Create.
Tumblr	I'm interested in digital **media**, **adventure sports**, and **mountains**.
Flickr	All my **photographs** are posted under the creative **commons** non commercial **attribution**...
YouTube	A **collection** of **videos** I've filmed on my **iPhone** while hiking **skiing** and **biking** in the **mountains**.
User 2	
All 6 networks	**Web Geek**?, Senior Digital **Strategist** in **Melbourne**, social **media maven**, **Google**-aholic, **simplicity** and UX **advocate**...

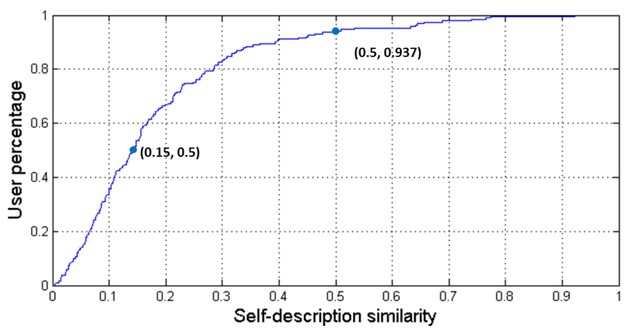


Fig. 4. Cumulative user distribution for self-description similarity. 7% of users have similar profiles across networks (top dot), while the majority describe themselves quite differently on different OSNs (left dot).

question: *How do profile descriptions differ across multiple OSNs?*

The user profile function is a common feature in OSNs, which is usually optional (as in the case of the studied six OSNs). We expect that some users reuse the same textual description over all of their OSN accounts. This is observed in our dataset (*e.g.*, User 2 in Table III) but actually makes up a relatively small percentage of cases; of the users that filled all six OSNs profiles, 3% used the same descriptions throughout. More common were user profiles that hinted about the user's identity with respect to the common functionality of the OSN. User 1 illustrates this case where each profile description is different, and customizes it towards the main functionality of the OSN; disclosing their job title in Google+ and Twitter, introducing personal interests in Tumblr, and summarizing their media contributions in Flickr and YouTube.

To examine this issue further, we collected all users who had populated profiles in all six networks (inclusive of Users 1 and 2). We further filtered out users whose descriptions were too short (less than 5 words), to lessen the impact of word sparsity. On this final set of 190 users, we first preprocessed their profile descriptions in each network by retaining only frequent non-stopword nouns (as bolded in Table III). We employ the standard Jaccard coefficient over the two sets of remaining single-word nouns from the pairs of the user's profile descriptions, to calculate the average pairwise similarity of each user's profiles.

We plot the cumulative distribution of the average pairwise Jaccard profile description similarity in Fig. 4. We see most users have an average pairwise similarity significantly less than 0.5 ($\sim$ 95%, top labeled dot), which is considered a common threshold for sentential similarity in prior work [18]. In fact, the majority users' profile similarity score is lower than 0.15 (left labeled dot), indicating that most users describe themselves very differently across different OSNs. Our follow-up manual sample analysis confirms this hypothesis. We thus posit that social networks are distinct to users based on their functionality, and that the functional difference manifests itself in the users' self-description in the user profiles.

B. Post Time

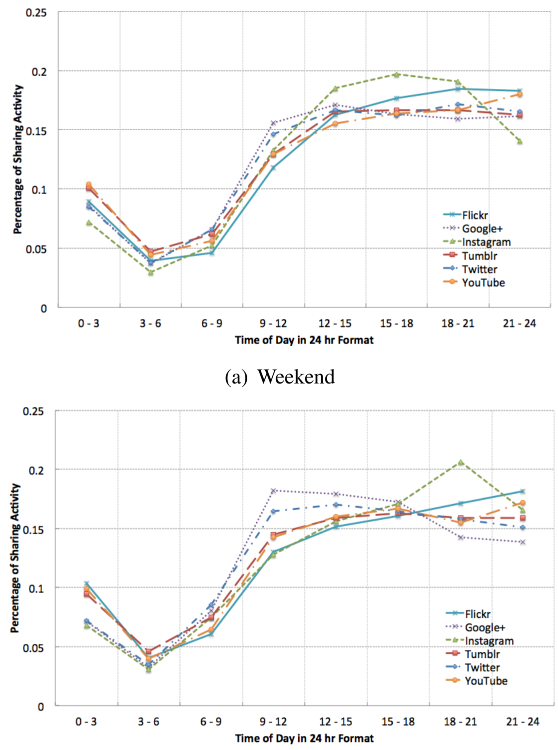
Being habitual beings, timing describes the way we behave in our daily lives: activities such as eating, sleeping and working. Correspondingly, post timing gives us a way to ascertain how users behave on each OSN. In this subsection, we analyse user behaviour by looking at temporal metadata of all posts shared by the user. We study a larger pool of 27K About.me users that provide their locale information, from which we are able to find out the local time of each post. Here, we answer the question: *How does user sharing behaviour vary with time?*

1) Time of Day Analysis: To begin, we analyse the time of day users are most active in sharing content on social networks. We first divide the hours of a day into 8 intervals of 3 hours each and distinguishing between weekdays and weekends. For each user, we aggregate his/her sharing activity on each social network into a distribution over the 8 intervals. After summing and normalizing the distributions for all the users, we plot the resultant aggregated distributions as seen in Fig. 5.

Macroscopically, the results show that activity across all six social networks follow a similar general trend. Low levels of activity are observed in the wee hours of the night, and higher levels of activity during the waking hours of the day and evening.

Taking a closer look, we notice that Google+ exhibits higher levels of activity during working hours (09:00-18:00) on weekdays; but exhibits lower and more evenly distributed activity during and after working hours on weekends. Interestingly, Instagram shows an opposite trend, peaking in activity after working hours (>18:00) on weekdays, and showing a decline after working hours on weekends. This hints, albeit subtly, at the contrasting nature of these two social networks – Instagram is a platform more frequently used during non-working hours, while Google+ is used more during working hours.

2) Day of Week Analysis: We see a clearer duality between OSN's usage during working times and non-working times when we perform the same analysis over days of the week. Fig. 6 shows the distribution of sharing activity over days of the week and provides us with several key insights. Image based social networks (*i.e.*, Flickr and Instagram) show a different trend throughout the week, peaking on the weekends, in comparison with the other video based, text and mixed media social networks that peak during the middle of the week



(a) Weekend

(b) Weekday

Fig. 5. Distribution of user sharing activity over 24 hours for each social network.

– Wednesday. One possible reason for this is that users are less able to take photos due to work on weekdays.

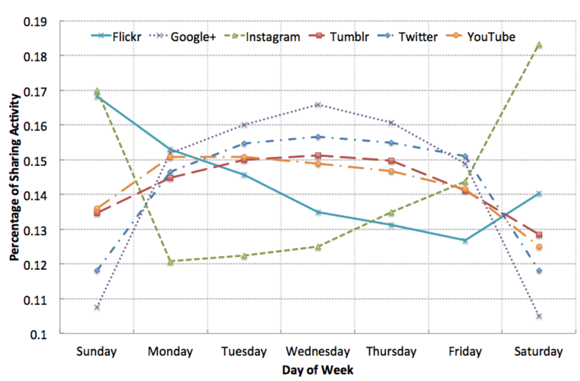


Fig. 6. Distribution of user sharing activity over days of the week for each social network.

To quantify the temporal similarity of the social networks in sharing activity, we employ Kullback-Leibler (KL) divergence to measure the difference between their posting day-of-week distribution. KL divergence is an asymmetric measure of difference between two probability distributions, defined by the information loss when using one distribution to approximate another:

$$D_{kl}(P||Q) = \sum_i P(i) ln \frac{P(i)}{Q(i)} \tag{1}$$

We chart a similarity matrix as seen in Fig. IV using scores derived from Equation 1 (KL divergence for discrete distributions). We note how dissimilar Instagram and Google+ are. This is consistent with what we observe from Fig. 6 – Instagram users are extremely active during weekends, while Google+ users are mostly active during the middle of the week.

TABLE IV. KL DIVERGENCE SCORES BETWEEN THE DISTRIBUTIONS OF OSN SHARING ACTIVITY OVER DAYS OF THE WEEK.

$Q \backslash P$	KL Divergence Scores					
	Twi	G+	Ins	Tum	Fli	YT
Twi	0	0.0021	0.0365	0.0018	0.0172	0.0023
G+	0.0021	0	0.0226	0.0070	0.0075	0.0002
Ins	0.0340	0.0217	0	0.0505	0.0131	0.0233
Tum	0.0019	0.0074	0.0553	0	0.0282	0.0072
Fli	0.0164	0.0073	0.0132	0.0263	0	0.0067
YT	0.0022	0.0002	0.0244	0.0069	0.0069	0

TABLE V. PROFESSION DESCRIPTION EXAMPLES.

Profession	Self-description
Marketing Expert	I am making new improvements new innovative solutions and new marketing tactics for e commerce affiliate marketing and vertical e commerce in Turkey. I want to make e commerce pie bigger. Also I am making consulting for e commerce and dotcom projects. If I believe in something, I have to do it.
Producer	I'm a Delhi based Electronic musician creative artist. I DJ produce minimal techno tracks as [...]. Also, I love robots.
Developer	I'm a fifty something Scottish born software developer living in strand near cape town South Africa and working for [...] in Stellenbosch. I enjoy photography and am a committee member of [...] photographic society photography.

C. Profession

As described in Section II, About.me is a platform for users to advertise themselves professionally. Leveraging on this, we seek to answer the following question: *Do users favour certain social networks for sharing personal and work related content?*

The separation of work related and personal posts are user dependent, *e.g.*, photography is one's hobby but another's profession. Therefore, to answer the above question accurately, we need to group users into professions. We start with a set of seed keywords that are related to ten representative professions in About.me (i.e. Marketing Expert, Producer, Developer, Photographer, Realtor, F&B, Healthcare, Blogger/Writer, Automobile, Graphic Designer/Artist). Then we search users' About.me profiles by these keywords and manually validate if each keyword adequately describes the profession. We prune and add more keywords and repeat this process until we are left with a sizable set of keywords. Finally, we use these keywords to identify user's profession and select users for our profession group analysis.

Due to limited space, we limit our analyses here to three professions, namely, *Marketing Expert*, *Producer* (*e.g.*, Musician, Podcaster, Filmmaker), and *Developer*. We select these three professions for two reasons: 1) they are typical professions in About.me and have a large user base, and 2) they are easily distinguishable and thus help to simplify the identification and validation process. In Table V, we provide example self-description for the three professions.

Our sample used in this analysis consists of 200 users from each of the three selected professions. We determine if a user's post is work related or personal by drawing a relation between the post's topic and the user's profession. To automate the process of inferencing topics from posts, we employ statistical topic modelling in the form of Latent Dirichlet Allocation

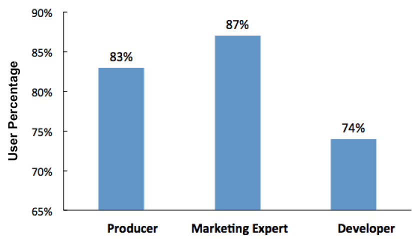


Fig. 7. Percentage of users who are workaholic – primarily using social networks for professional purpose.

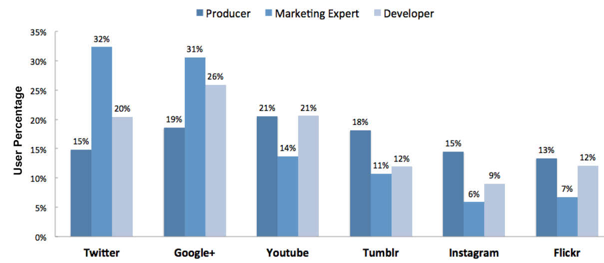


Fig. 8. User's social network preference to share work related content.

(LDA) [19]. LDA models documents and their vocabulary in the same space, clustering similar documents and words together based on co-occurrence. Social Media posts being very short documents face the issue of insufficient verbosity for LDA to assign topics accurately. We tackle this issue by author-time pooling of posts as described in [20]: Instead of a post being a single document, we treat a document as a collection of posts that have been shared within a time interval of 4 hours by the user. To construct our LDA model, we select a group of 2000 random users and train our model on a collection of $1,441,987$ author-time pooled documents from their posts from all six social networks, with a predefined number of topics set to 50 by experimentation. We manually assign a label to each topic by looking at its top keywords, and further associate relevant topics to their respective profession groups. Some examples of the topic–profession associations are: Search Engine Optimisation (SEO) topics to marketing professionals, technology and gadgetry topics to developers, and music and video making topics to producers. We aggregate the frequency of posts with the same labelled topic given by LDA to find out the top topics for each user.

We start our analysis by asking: *For each profession group, what percentage of users are "workaholics", those who frequently post about their work in OSNs?* To be specific, we deem a user a workaholic if any of his/her top two topics in any of the six social networks matches the topics related to his/her profession. Fig. 7 shows the percentage of workaholics in the three professions.

Developers do not frequently use social networks for professional purpose as compared to the other two professions. This phenomenon may be explained by the fact that the other two professions have more reasons to use OSNs as means of reaching out to their target audience: producers of content may use OSNs to publish their works and likewise, marketing experts use them as means of promoting their cause. However, there may be less motivation for developers to do the same, hence a less sharing of work-related content is observed for them.

A natural follow up question is: *For those workaholics, which social networks do they prefer?* Fig. 8 shows that Google+ is consistently a popular choice amongst all three groups for sharing content related to work. YouTube is a video based OSN and is the primary choice for producer, as it is a suitable medium for them to distribute and publish their work. Twitter is the favourite choice amongst marketing experts possibly because of the high propensity for Twitter users to browse tweets via hashtags or trending topics, thus making it easier for unrelated users to discover each other – a great marketing outreach medium.

For each profession group, we further investigate the most popular topics posted on each OSN for both hourly and daily intervals. We observe that the most popular topics remain consistent over time, *e.g.*, the most popular topic for developers on Twitter is *technology* for any day of the week and any time of the day. This shows each platform, to a user, has its own dedicated function, *i.e.*, a platform used mostly for work will unlikely be used often for personal (non-work related) purpose.

Our analysis in this subsection concurs with results obtained from our temporal analysis. Google+ sharing activity reaches its peak during working days and similarly, is the network of choice for professions that do not deal with media (music, video, images). Also, Instagram consistently ranks one of the lowest for all three professions supporting our earlier postulation of it being a network for personal rather than professional use. We can conclude from our analysis that some OSNs are used more frequently for work related purposes, whereas others are preferred for personal use.

V. Cross-network Interaction Analysis

Informally, *cross OSN sharing* is when a user multicasts his/her activity over multiple networks. For example, photos shared from Instagram to Flickr are automatically tagged as "uploaded:by=instagram", indicating their originality. Like the Instagram/Flickr case, cross sharing often has a source OSN platform (shared from) and a sink (shared to).

To examine this user behavior, we build timelines for each user's activities and identify cross-sharing activities. An example timeline fragment of one user is shown in Table VI, where we observe cross-sharing. Cross-sharing can be enabled by third party software, which may broadcast the content simultaneously to target OSNs. We identify cross-shared (aggregate) activities as consecutive activities on different OSNs, where common timing and content define the bounds of the cross-shared activity (bolded instances in Table VI). In our dataset, we programmatically identified $111,431$ cross-sharing activities.

Formally, we define cross-sharing as a 4-tuple:

$$CS := < U, P, N^{source}, N^{sink} >$$

where U is the user identification, P is the post being cross-shared, N^{source} indicates the network where the post is originally created (or *source*), and N^{sink} indicates the destination network (or *sink*). Note that we only assign source and sink

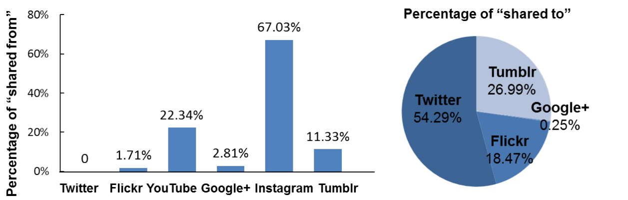


Fig. 9. Cross-sharing per OSN. (l) percentage of posts *shared from* an OSN; (r) percentage of all identified cross-shared activities in our dataset *shared to* a particular OSN.

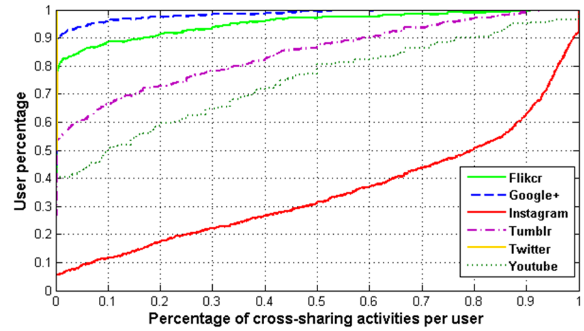


Fig. 10. Cumulative distribution of users with respect to their percentage of cross-sharing activities.

roles only when evidence of the direction of sharing is present; *e.g.*, in Table VI, the tweet's inlined shortened URL resolves to Instagram, and the photo on Flickr contains the machine generated tag "uploaded:by=instagram".

TABLE VI. A FRAGMENT OF A SAMPLE USER'S ACTIVITY TIMELINE, THAT ILLUSTRATES A CROSS-SHARED ACTIVITY (BOLDED).

Time	Network	Content
26/05/2013 18:36:46	Twitter	my poolside jam... http://t.co/f2480xhw5u
27/05/2013 03:00:42	**Instagram**	I miss baby snuggles when the kids are away... :) I'm so blessed. #momlife
27/05/2013 03:00:44	**Flickr**	I miss baby snuggles when the kids are away... :) I'm so blessed. #momlife
27/05/2013 03:00:44	**Twitter**	I miss baby snuggles when the kids are away... :) I'm so blessed. #momlife @ home tweet home http://t.co/WioRNR6BjA
27/05/2013 21:02:02	Twitter	you just never know who or what's going to show up at a wedding ... ;)

Given the identification of these cross-sharing activities, we characterize cross-sharing behavior in terms of platform considerations in the following.

We map the dissemination flow of user content in the six OSNs by aggregating the respective N^{source} and N^{sink} identities over all detected cross-sharing activities. Fig. 9 (left) shows the percentage of all posts in each OSN that is cross-shared to another OSN. Instagram and YouTube hold the highest *shared from* percentage and serve as the most significant source OSNs. We also plot (in Fig. 10) the CDF of users with respect to their cross-sharing ratio; *i.e.*, the percentage of the user's own content that are also disseminated to another OSN. Cross-sharing turns out to be platform-sensitive. Instagram serves as a popular source; over 90% of Instagram users have shared their Instagram posts to another network. For Flickr and Google+, fewer than 20% of users took their original platform content and shared it to another OSN.

We graph each platform's share of all 111K+ identified cross-shared activities as a destination (sink) in Fig. 9 (right). Twitter dominates, being the dominant destination for 54% of cross-sharing activities. We believe that OSNs examined play largely different roles in source/sink discrimination. Source and sink networks for cross-sharing activity are markedly different.

Instagram user content originates from its mobile application that places sharing as a central theme, manifesting in the embedded "Share" button to route the post to various social media networks. As shown in Fig. 11, the central placement of the sharing functionality and its usability (dedicated buttons for

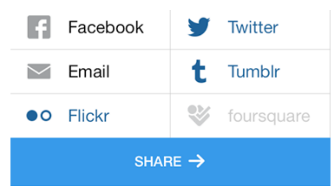


Fig. 11. Sharing functionality in the Instagram mobile application.

different OSNs) drives the cross-sharing behavior we observe. This suggests that platform functionality strongly drives cross-sharing. We examined the current (*ca.* April 2015) state of embedded cross-sharing options for each of the six networks, both in their website versions as well as mobile apps for the iOS and Android mobile platforms, presented in Table VII. We find that Instagram and YouTube support the most sharing destinations; Tumblr and Flickr support sharing to certain sinks, while Twitter and Google+ do not support sharing. Aside from in-platform support, cross-sharing can be done through third-party plug-ins (*e.g.*, friendplus.me for Google+) or manually done by the user. Fig. 12 illustrates our understanding of the information flow among the networks examined, where the horizontal axis represents the tendency of a platform to serve as source or sink, and the vertical axis represents the level of explicit cross-sharing interaction. With some caveats, we see that OSNs that focus more on visual media tend to be sources, whereas platforms focusing textual media serve more as sinks. Tumblr, Flickr and Google+, positioned in the middle, play both roles.

We make several additional observations surrounding cross-sharing source (source support). Although Instagram cut off its sharing to Google+ after being acquired by Facebook (April 2012), some of the About.me users in our dataset still manage to cross-share their Instagram photos to Google+. Also, while Google+ positions itself as a sink network, and does not support any cross-sharing, users have found means to do so through other workflows. We believe both source and sink characteristics have their roles within the OSN ecology: "sourceness" acknowledges original content and possibly the value add of using a particular platform (artistic filters in Instagram may be an example), while "sinkness" promotes an aggregator effect that enables downstream analytics to gain more complete pictures.

TABLE VII. CROSS-SHARING SUPPORT BY PLATFORM.

	Supported destination network
Twitter	–
Google+	–
Instagram	Twitter, Flickr, Tumblr
Tumblr	*depends on the user's personalized setting*
Flickr	Twitter, Tumblr
YouTube	Twitter, Tumblr, Google+

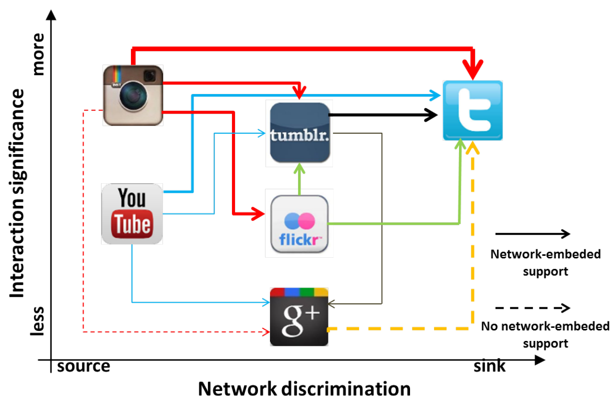


Fig. 12. User generated content flow among OSN platforms. Content flows in the directions indicated by arrows and line thickness represents flow volume. Arrow color correspond to a specific OSN.

VI. CONCLUSION

We studied multiple online social networks (OSNs) from a user-centric perspective, with the aim of discovering behavioral patterns in their multiple network use. We analyzed how users participate and interact among six popular OSNs in our crawled dataset which we have made publicly available. Our study validates the hypothesis that users exhibit varied behavior on different OSNs, accounting in part for the OSN's primary media type. In our multi-network analysis of single OSNs, we initially showed how the majority of users portray themselves differently across OSNs, suggesting differences in use. By examining users' sharing activities, we uncovered a dichotomy between usage for professional and personal reasons. In our cross-network sharing analyses, we mapped how users post from one source network to a sink network. By plotting the source–sink directionality of cross-sharing, we labeled the media-centric OSNs of YouTube and Instagram to be sources, and the lowest common denominator Twitter OSN to be the common sink.

Our study has examined the public face of OSNs, uncovering just the surface of the vibrant and varied ecology that is today's totality of social media. While our study covers a much larger scale than previous works that have largely confined themselves to the analyses of one or two individual networks, a key limitation of our work is that we have only studied large networks, and their users' public sharing activities. In the future, we plan to harvest data from private social networks where possible, to gain further insight on the differentiation between public and private OSN use. We believe such work may benefit social media applications.

REFERENCES

[1] Y.-Y. Ahn, S. Han, H. Kwak, S. Moon, and H. Jeong, "Analysis of topological characteristics of huge online social networking services," in *Proceedings of International Conference on World Wide Web*, 2007.

[2] A. Mislove, M. Marcon, K. P. Gummadi, P. Druschel, and B. Bhattacharjee, "Measurement and analysis of online social networks," in *Proceedings of ACM SIGCOMM Conference on Internet Measurement*, 2007.

[3] T. Althoff, D. Borth, J. Hees, and A. Dengel, "Analysis and forecasting of trending topics in online media streams," in *Proceedings of ACM International Conference on Multimedia*, 2013.

[4] K. Lerman and R. Ghosh, "Information contagion: An empirical study of the spread of news on digg and twitter social networks," in *The International Conference on Weblogs and Social Media*, 2010.

[5] J. Leskovec, L. Backstrom, R. Kumar, and A. Tomkins, "Microscopic evolution of social networks," in *Proceedings of ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 2008.

[6] L. Guo, E. Tan, S. Chen, X. Zhang, and Y. E. Zhao, "Analyzing patterns of user content generation in online social networks," in *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 2009.

[7] M. De Choudhury and H. Sundaram, "Why do we converse on social media?: an analysis of intrinsic and extrinsic network factors," in *Proceedings of ACM SIGMM International Workshop on Social Media*, 2011.

[8] F. Schneider, A. Feldmann, B. Krishnamurthy, and W. Willinger, "Understanding online social network usage from a network perspective," in *Proceedings of ACM SIGCOMM Conference on Internet Measurement Conference*, 2009.

[9] F. Benevenuto, T. Rodrigues, M. Cha, and V. Almeida, "Characterizing user behavior in online social networks," in *Proceedings of ACM SIGCOMM conference on Internet measurement conference*, 2009.

[10] R. Zafarani and H. Liu, "Connecting users across social media sites: A behavioral-modeling approach," in *Proceedings of the 19th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 2013.

[11] H. Zhang, M.-Y. Kan, Y. Liu, and S. Ma, "Online social network profile linkage," in *Proceedings of Asian Information Retrieval Symposium*, 2014.

[12] S. Kumar, R. Zafarani, and H. Liu, "Understanding user migration patterns in social media." in *AAAI Conference on Artificial Intelligence*. Citeseer, 2011.

[13] T. Chen, M. A. Kaafar, A. Friedman, and R. Boreli, "Is more always merrier?: A deep dive into online social footprints," in *Proceedings of the 2012 ACM Workshop on Online Social Networks*, 2012.

[14] Y. Chen, C. Zhuang, Q. Cao, and P. Hui, "Understanding cross-site linking in online social networks," in *Proceedings of Workshop on Social Network Mining and Analysis*, 2014.

[15] R. Ottoni, D. Las Casas, J. P. Pesce, W. Meira Jr, C. Wilson, A. Mislove, and V. Almeida, "Of pins and tweets: Investigating how users behave across image-and text-based social networks," *International AAAI Conference on Weblogs and Social Media*, 2014.

[16] M. Duggan and A. Smith, "Social media update 2013," *Pew Research Center report*, 2013.

[17] L. Rainie, J. Brenner, and K. Purcell, "Photos and videos as social currency online," *Pew Internet & American Life Project*, September 2012.

[18] P. Achananuparp, X. Hu, and X. Shen, "The evaluation of sentence similarity measures," in *Proceedings of the 10th International Conference on Data Warehousing and Knowledge Discovery*, 2008.

[19] D. M. Blei, A. Y. Ng, and M. I. Jordan, "Latent dirichlet allocation," *the Journal of machine Learning research*, vol. 3, pp. 993–1022, 2003.

[20] R. Mehrotra, S. Sanner, W. Buntine, and L. Xie, "Improving lda topic models for microblogs via tweet pooling and automatic labeling," in *Proceedings of International ACM SIGIR Conference on Research and Development in Information Retrieval*, 2013.