

# Quantifying Cross-platform Engagement through Large-scale User Alignment

Jiejun Xu, Tsai-Ching Lu, Ryan Compton, David Allen  
HRL Laboratories  
3011 Malibu Canyon Road  
Malibu, CA 90265  
{jxu, tlu, rcompton, dallen}@hrl.com

## ABSTRACT

As online social media becomes prevalent as part of our daily life, it is increasingly common for a user to have accounts on multiple social media platforms. In this work, we present our findings on quantifying the extent of user engagement of different platforms as well as their correlations. The study is conducted based on a large-scale user alignment on 6 major social media platforms. Specifically, we identify both explicit and implicit mentions of social media accounts from the Twitter Decahose stream over a period of 22 months. During the process, we have aligned a total of 21,456,808 Twitter users to their alternative accounts on different platforms. Subsequently, we extract the number of overlapping users between any combination of these social media platforms exhaustively.

## Categories and Subject Descriptors

H.4 [Information Systems Applications]: Miscellaneous;  
J.4 [Computer Applications]: Social And Behavioral Sciences

## General Terms

Measurement, Human Behavior

## 1. INTRODUCTION

Online social media has gained tremendous popularity over the past decade. Each social media platform has different characteristic and usually focuses on different aspect of the user needs. For instance, Twitter and Facebook focus more on communication, while Tumblr and Instagram focus more on photo sharing. As a results, it is common for a user to have multiple accounts on different social media platforms to accommodate his needs. In this work, we present a study on quantifying the level of user engagement on different combinations of the social media platforms. Specifically, we conduct a large-scale user alignment across 6 major social media platforms to extract the number of overlapping

users. These platforms include Twitter, Tumblr, Wordpress, Blogger, Instagram, and Facebook.

Several recent works have started to investigate the cross-platform alignment problem for user modeling, personalization and recommendation. For instance, Abisheva et al. [1] combined user data from Twitter and YouTube to provide a descriptive analysis of the demographics and behavioral features on the two online communities. Vu et al. [7] proposed a general and extensive system architecture for aggregating and integrating users' social profiles into collaborative systems. Tiroshi et al. [6] modeled online users based on features extracted from different social media sites. They showed that aggregated user profile enhanced personalization service effectively. Similar study has been conducted by Deng et al. [2] to improve personalized video recommendation. There are also studies focused on developing methodologies to accurately map users across platforms. Zafarani et al. [9] introduced a behavioral-modeling based approach to connect individual across social media sites. Other recent works on user mapping utilized network attributes [3] and a variety of content features [5] to de-anonymize users across sites.

Most prior work focused on aligning users with relatively confined datasets. In contrast, the main contribution of our effort is the cross-platform user alignment *at scale*, using a large real-world data stream to derive representative results for major social media platforms. To the best of our knowledge, this is the first attempt to quantify overlapping social media usage from a variety of platforms in such a scale.

## 2. USER ALIGNMENT

We take a simple extractive-based approach to identify mentions of social media user accounts from the Twitter datastream, which consists of 10% sample of public tweets

Tumblr	<code>http://[www.]*[a-zA-Z0-9- _]+.tumblr.com</code> <code>http://tumblr.co/(\S{4,20})\</code>
Wordpress	<code>http://[www.]*[a-zA-Z0-9- _]+.wordpress.com</code> <code>http://.*wp[.]me.*</code>
Blogger	<code>http://[www.]*[a-zA-Z0-9- _]+.blogspot.com</code> N/A
Instagram	<code>http://[www.]*instagram.com/[a-zA-Z0-9- _]+</code> <code>http://instagr\{3,7\}/p/\S+?/\</code>
Facebook	<code>http://www.facebook.com/(\S+)</code> N/A

Figure 1: Regex patterns used for identifying social media accounts from the Twitter data. Note that the matched Short URLs are resolved to full account address in post processing.

Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage, and that copies bear this notice and the full citation on the first page. Copyrights for third-party components of this work must be honored. For all other uses, contact the owner/author(s). Copyright is held by the author/owner(s).  
WebSci'14, June 23–26, 2014, Bloomington, IN, USA.  
ACM 978-1-4503-2622-3/14/06.  
<http://dx.doi.org/10.1145/2615569.2615662>.

