

INTRODUCTION Many online social networks, including Facebook and Google+, enforce a Real-Name policy, requiring users to use their real names when creating accounts [3, 2]. The cited reasons for the Real-Name policy include that it improves the quality of the content and the service (helping decrease spam, bullying, and hacking), increases accountability, and helps people to find each other. The Real-Name policy, however, also enables the social networks to tie user interests—as reflected from their use of the online services—with their true names, generating a treasure trove of consumer data. This has resulted in many debates [13] and petitions [6], with privacy advocates claiming that Real-Name policy erodes online freedom [31]. Privacy-conscious users have started finding ways to bypass the policy, hiding their real identity while continuing to use these social networks [22]. Twitter, on the other hand, does not impose strict rules for users to provide their real names, although it does require them to register with and employ unique pseudonyms. Taking advantage of this lack of Real-Name policy, many Twitter users choose to employ pseudonyms that have no relation to their real names. Some users choose such a pseudonym only because they enjoy being associated with a particular fun or interesting pseudonym. But many users likely choose pseudonyms with no relation to their real names because they want to be anonymous on Twitter. For example some users may desire the ability to tweet messages without revealing their actual identities. Other users may desire to follow sensitive and controversial accounts without exposing their real identities. The lack of Real-Name policy enforcement has turned Twitter into a popular information exchange portal where users share and access information without being identifiable—as is evident by Twitter’s role in Egyptian revolution [25] and for reporting news in Mexico [34]. However, there is a meaningful debate about the pros and cons of online anonymity, as it allows people to more easily spread false rumours [14], defame individuals [12], attack organizations [33], and even spread spam [41, 17]. In this work we use Twitter to study the prevalence and behavior of Identifiable users (those disclosing their full name) and Anonymous users (those disclosing neither their first nor last name). Although both on-line and off-line anonymity has been considered by researchers in psychology and sociology, as discussed in Section 7, these studies have generally been carried out with small data sets and surveys. There have also been a few data-driven studies of anonymity in blogs and postings to Web sites [16, 5, 36]. To our knowledge, this paper is the first to conduct a large-scale data-driven analysis of user anonymity in online social networks. The potential benefits of such a study include: (i) a deeper understanding of the importance and role of anonymity in our society; (ii) guidance for the incorporation of privacy and anonymity features in existing and future online social networks; (iii) and as we shall discuss in the body of the paper, the discovery of illegal (such as child-porn and terrorism) or controversial (such as ethnic or religious hate) activities