INTRODUCTION

**T**WITTER, a microblogging service, is considered a popular online social network (OSN) with a large user base and is attracting users from different walks of life and age groups. OSNs enable users to keep in touch with friends,relatives, family members, and people with similar interests, profession, and objectives. In addition, they allow users to interact with one another and form communities. A user can become a member of an OSN by registering and providing details, such as name, birthday, gender, and other contact information. Although a large number of OSNs exist on the web, Facebook and Twitter are among the most popular OSNs and are included in the list of the top 10 websites1 around the worldwide.

*A. OSN and the Social Spam Problem*

Twitter, which was founded in 2006, allows its users to post their views, express their thoughts, and share news and other information in the form of tweets that are restricted to280 characters. Twitter allows the users to follow their favourite politicians, athletes, celebrities, and news channels, and to subscribe to their content without any hindrance. Through *following* activity, a follower can receive status updates of subscribed account. Although Twitter and other OSNs are mainly used for various benign purposes, their open nature, huge user base, and real-time message proliferation have made them lucrative targets for cyber criminals and social bots. OSNs have been proven to be incubators for a new breed of complex and sophisticated attacks and threats, such as cyberbullying, misinformation diffusion, stalking, identity deception, radicalization, and other illicit activities, in addition to classical cyber attacks, such as spamming, phishing, and drive by download [1], [2]. Over the years, classical attacks have evolved into sophisticated attacks to evade detection mechanisms. A report2submitted to the US Securities and Exchange Commission in August 2014 indicates that approximately 14% of Twitter accounts are actually spambots and approximately 9.3% of all tweets are spam. In social networks, spambots are also known as socialbots that mimic human behaviour to gain trust in a network and then exploit it for malicious activities [3]. Such reports and findings demonstrate the extent of cyber crimes committed by spambots and how OSNs are proving to be a heaven for these bots. Although spammers are less than benign users, they are capable of affecting network structure and trust for various illicit purposes.

The main contributions of this study can be summarized as follows.

• A novel study that uses community-based features with other feature categories, including *metadata*, *content*, and *interaction*, for detecting automated spammers.

• Six new features are introduced and two existing features are redefined to design a feature set with improved discriminative power for segregating benign users and spammers. Among the six new features, one is content based, three are interaction-based, and the remaining two are community-based. Meanwhile, both redefined features are content-based. When defining interaction based features, focus should be on the *followers* of a user, rather than on the ones he/she is *followings*.

• A detailed analysis of the working behavior of automated spammers and benign users with respect to newly defined features. In addition, two-tailed *Z*-test statistical significance analysis is performed to answer the following question: “*is the difference between the working behavior of spammers and benign users in terms of newly defined features a random chance?*”

• A thorough analysis of the discriminating power of each feature category in segregating automated spammers from benign users.