**Summary of “POISED: Spotting Twitter Spam Off the Beaten Paths”**

**Name:** Avineet Kumar Singh

**ID#:** M12135865

**Abstract.** Online social networks are used by attackers to spread spam and malicious content by using fake accounts, creating bot nets, sending unsolicited messages, through phishing and many more. Existing techniques detect such spam by predicting the trustworthiness of accounts and analyzing the content of these messages. The attackers can still successfully evade these defenses existing spam detection systems detect the way in which malicious accounts infiltrate the network and build connections, rather than the way in which malicious messages spread across the network in comparison to legitimate ones. This paper exploits the personal connections or common interests of people to form communities. This networked community share some topics of interest and can be recognized by their structure in the underlying connection graph, as they form strongly connected subgraphs. Moreover, content shared on these social networks tend to propagate according to the interests of people.

A spam detection system called POISED (Parties of Interest Semantic Extraction and Discovery) is introduced that detects networked communities in Twitter by partitioning its social graph then identifies topics of interest in these communities it then tracks the dissemination of similar messages through communities and constructs a probabilistic model of the parties of interest through which these messages are normally disseminated. Finally, leveraging this model, a classifier detects malicious content by identifying the messages that do not follow these expected parties of interest.

Community detection is done using the infomap algorithm. Latent Dirichlet Allocation (LDA) is used to detect topics of interest of communities. It employs supervised machine learning to classify messages as spam or benign. Evaluation of POISED classifier on a dataset of 1.3M tweets collected from 64K users how that the approach is effective in detecting malicious messages, reaching 91% precision and 93% recall outperforming the existing spam detection systems.

**Contributions.** The paper introduces a concept of networked community in Twitter which shares some topics of common interest by partitioning its social graph and constructs a probabilistic model of the parties of interest.

It also introduces POISED (Parties of Interest Semantic Extraction and Discovery), that calculates and maps the most probably networked communities on Twitter, then singles out those messages in communities which deviate from the norm. POISED outperforms the other state-of-the-art detection systems like SpamDetector, Compa, and BotOrNot.

**Weaknesses.** The weakness of the paper is that the POISED method usesfour-gram analysis to identify and group similar messages before applying the probabilistic model to detect spam and if the attacker is aware of the analysis method it could evade it.

This method also requires certain number of users and messages to form communities of interest which eventually forms a social graph. This creates an escape point for small group of malicious connections.