**Blockchain Social Media Bot Identification**

Name: Amulya Geereddy

Date: 07.30.2021

Master’s Degree

Business Analytics and Information Systems

Professors Balaji Padmanabhan and Kaushik Dutta

Contents:

Introduction…………………………………………………………...1

Analysis for identifying the bots……………………………………...2

Conclusion…………………………………………………………….5

**Introduction:**

Social media platforms have always been an entry point for establishing social communications for billions of people, and at present, platforms based on Blockchain technology are attracting a lot of public attention worldwide. ​Blockchain Social Media are nothing but decentralized platforms that allow the development of applications and smart contracts. The major benefit of such platforms is that they offer end-to-end encryptions for every interaction enabling individuals to have more privacy and control over their information.​

A decentralized or blockchain-based social media addresses the first and the most serious issue with today’s social networks – centralization. When there is only one company storing all the data of the users and it is not accountable to anyone for what it does with that data, it can do whatever it wants without ever being answerable to anyone.

Blockchain social network decentralized the control through its distributed ledger. The data of the users is not saved on one server but many servers that are owned by people who are not connected or related to each other. That’s the best way to prevent rule from going in the hands of one entity. When you divide the rule, the Facebook-like monopoly is out of question.

Steemit is a social media platform built on the blockchain that pays its content creators and curators in cryptocurrency Steem is a complete ecosystem that has the potential to fix a lot of issues with the current social media platforms. Steem is decentralized, free from censorship and pays you for your time. The tokens, STEEM and steem Dollars, reward you for creating, commenting, and voting and content.

**Analysis for identifying the bots**

As per the analysis of the underlying cryptocurrency transfer network on the blockchain reveals that more than 16% transfers of cryptocurrency in Steemit are sent to curators suspected to be bots and finds the existence of an underlying supply network for the bots, Concretely, if there is a suspicious vote cast by a curator to an author through a vote operation, there should also be a suspicious fund transferred to the curator from the author through a transfer operation that happens before the vote operation is performed by the curator. Specifically, we consider a transfer operation is suspicious if the ‘memo’ area (allowing sender to leave a message) of the transfer operation only contains a link pointing to a recent post created by the sender. If in addition, the recipient of the transfer operation, after receiving the fund from the sender, votes for the post matching that link within the 7-days’ time window after the post creation time, we consider it as a suspicious trade between a post author and a voting bot.

As we have data for 45 months for transfer and vote operations in steemit. We can do monthly analysis.

Step 1:

Loading the files from two aspects, Votes and Transfers, parallelly while iterating through all the 45 months. The files are handled on monthly basis. Processing Votes and Transfers files simultaneously for each month.

Step 2:

Loading each file into a Dataframe with appropriate columns and headers in a low memory format to avoid any abruptions.

We are loading a URL extractor to obtain the links from comments tab of the Transfers file. This is the scenario of bot identification from the paper discussed. (2019)

Step 3:

Applying a Lambda function on the comments column to screen for URLs, which specifies the Sender’s most recent post and if this is validated, this is a potential bot’s behavior.

Step 4:

Defined a function to check if the sender’s URL has Sender’s name. We are verifying if the URL contains sender’s name to validate the behavior. On validation we are setting the comments column to True.

Step 5:

We are eliminating all the rows with False values to avoid the entries without URL.

Step6:

Grouping all the rows based on ‘From’ and ‘To’ columns with ‘TransferCount’ and changed the column names to ‘Voter’ and ‘Author’. This indicates the number of transactions between a particular combination of Voter and Author.

Step7:

Processing the Votes dataframes with grouping by based on the columns, ‘Author’ and ‘Voter’ so as to have a combination of all the senders and receivers which have to and fro transactions among then with in a month.

Step8:

Merging the transfer table ‘transfercount’ column and voter table ‘votercount’. Which helps us to get the records in which the transfer of money was done in transfer table between ‘from’ and ‘to’ and matching the ‘to’ column with ‘voter’ and ‘from’ with ‘author’ and find the records and saving those records into a new data frame with the ‘votercount’ and ‘author count’.

Step9:

Checking if the ‘votercount’ and ‘transfercount’ is same for each record and saving in a new column with TRUE if they both match.

Step12:

Filtering the data with the true and saving all the records with true in a new data frame. Now we have a list of all possible suspects in a dataframe with their count of possible fraudulent transactions

Step13:

The monthly analysis files are combined into a single dataframe with a column consisting of unique Senders and Recievers.

The final dataframe has a list of all the potential fraudulent users with their STEEMIT profile links to verify their accounts with ease.

**Conclusion**

The purpose of this research was to identify effective strategies for Identifying the bots Steemit transactions data. Based on the analysis conveyed, it can be concluded that there are multiple behavior modification therapies important for the improvement of this behavior. From the actual data we could identify Approximately 52000 records which are possible suspicious bots. This procedure helped to reduce huge data to very effective number. This procedure helps to reduce the data to almost 1% of the original data.