CS 765 Assignment 3
Decentralized App (DApp)

Guramrit Singh
210050061

Indian Institute of Technology, Bombay
April 14, 2024

Contents

1 Design of the DApp

1. The DApp will operate on a blockchain, such as Ethereum, where it functions as a smart contract.

2. Users can view the voting results for a specific article by providing a minimal amount of ETH, which can be adjusted as necessary. In our DApp, we have set this amount to be 0.00001 ETH.

3. To participate as a fact-checker on the DApp, users must deposit some ETH (in our case its 0.001 ETH) using their public key, which serves as their identity. The deposited amount will determine their voting power in computing the voting results. This measure prevents malicious individuals from conducting Sybil attacks, as their voting power remains consistent regardless of the number of accounts they create, provided their total deposit remains unchanged.

4. A registered fact-checker has the flexibility to request de-registration at any time, allowing them to retrieve their updated deposit. If a user opts to register again using the same public key, their trustworthiness score will be recalculated based on their prior voting history and the outcomes of the corresponding articles.

5. A registered fact-checker can vote on any number of articles, selecting either 0 to express disbelief in the news article or 1 to indicate belief in its truthfulness.

6. Since some voters may possess greater expertise in certain subjects, articles are categorized accordingly, with different trustworthiness scores maintained for each category. For simplicity, we consider 10 categories.

7. A voter's stake in a particular category is determined by the ratio of the total number of votes they have cast for that category to the total number of votes they have cast overall. If a voter has not cast any votes, their stake in all categories is considered to be 0.

8. Any user can submit an article to the DApp for fact-checking. Each article is assigned a unique ID calculated using a hash function that incorporates the article's category, title, content, timestamp, and a 10-digit random number.

$$\text{ID} = \text{hash(category)} \;||\; \text{hash}(p_k \;||\; \text{title} \;||\; \text{content} \;||\; \text{timestamp} \;||\; \text{10-digit random number})$$

   The first half of the ID denotes the article's category, while the second half contains general information.

9. The trustworthiness of a voter is calculated and updated based on their previous votes. The formula used is:

$$\text{Trustworthiness}(i, j) = \frac{\text{Total number of votes cast by } i \text{ for category } j \text{ in favor of the result}}{\text{Total number of votes cast by } i \text{ for category } j}$$

   When a person registers on the DApp for the first time, they are assigned a trustworthiness score in every category equal to the current average trustworthiness score in that category.

10. The truth value of a particular article j in a specific category k is computed as follows:

$$V_{0,j} = \sum_{i=0}^{N} \mathbb{1}_{0,j} \cdot \text{stake}(i, k) \cdot \text{trustworthiness}(i, k)$$

$$V_{1,j} = \sum_{i=0}^{N} \mathbb{1}_{1,j} \cdot \text{stake}(i, k) \cdot \text{trustworthiness}(i, k)$$

$$\text{Truth-value}(j) = \begin{cases} 0 & \text{if } V_{0,j} > V_{1,j} \\ 1 & \text{otherwise} \end{cases}$$

11. To incentivize honest participation and discourage malicious behavior, the following measures are put into effect. These measures take effect once a voter has cast votes for at least 10 articles within a specific category. This threshold is chosen because initial votes may not provide enough data to assess a voter's behavior accurately.

    (a) Users who have cast their votes in support of the result for a specific article l and possess a trustworthiness score exceeding 40% in that category k (trustworthiness$(i, k) > 0.4$) will receive a portion of the fees provided by the user who wishes to view the results (F_{total}). The amount received (F_i) by each voter (i) is calculated as follows:

$$M(l) = \text{Vote by } i \text{ for article } j == \text{Majority vote for } l$$

$$P(i) = M(l) \;\&\&\; \text{trustworthiness}(i, k) > 0.4 \;\&\&\; \text{number of votes cast by } i \text{ for category } k \geq 10$$

$$F_i = \frac{\mathbb{1}_{P(i)} \cdot \text{stake}(i, k) \cdot \text{trustworthiness}(i, k)}{\sum_{j=1}^{N} \mathbb{1}_{P(i)} \cdot \text{stake}(j, k) \cdot \text{trustworthiness}(j, k)} \cdot F_{\text{total}}$$

    where stake(i, k) represents the stake of voter i in category k and trustworthiness(i, k) denotes their trustworthiness score in that category.

(b) Voters with a trustworthiness score below 20% in the category k (trustworthiness$(i,k) < 0.2$) will forfeit their stake in that category and won't be allowed to vote any further in that category, also all their previous votes in this category will be declared null and void. The confiscated stake (C_{total}) will be redistributed among all stakeholders in the category, proportionate to their stake and trustworthiness. The amount received (C_i) by each voter (i) is computed as:

$$P(i) = \text{number of votes cast by } i \text{ for category } k \geq 10$$
$$C_i = \frac{\mathbb{1}_{P(i)} \cdot \text{stake}(i,k) \cdot \text{trustworthiness}(i,k)}{\sum_{j=1}^{N} \mathbb{1}_{P(i)} \cdot \text{stake}(j,k) \cdot \text{trustworthiness}(j,k)} \cdot C_{\text{total}}$$

where stake(i,k) represents the stake of voter i in category k and trustworthiness(i,k) denotes their trustworthiness score in that category.

2 Ensuring Security, Integrity and Trust in the DApp

1. **Preventing Sybil Attacks**
   The DApp employs a deposit-based identity system, requiring users to deposit ETH to participate as fact-checkers. This measure ensures that even if a malicious user creates multiple accounts, their influence on the voting results remains limited to their total deposited amount.

2. **Continuous Trust Evaluation**
   Trustworthiness scores are continuously calculated and updated based on fact-checkers' voting behavior. This process assesses their reliability and integrity in evaluating news articles, mitigating the risk of biased or malicious contributions.

3. **Weighted Voting based on Reliability**
   Fact-checkers' votes are weighted according to their trustworthiness scores, giving more weight to those with higher scores. This ensures that more reliable contributors have greater influence on the voting results, enhancing the accuracy and credibility of the fact-checking process. Additionally, the DApp categorizes news items into separate categories, each with its own trustworthiness evaluation and weighting mechanism. This ensures that expertise and trustworthiness are assessed and applied appropriately across different subject areas.

4. **Encouraging Honest Engagement**
   The DApp incentivizes honest participation by rewarding fact-checkers who consistently provide accurate assessments and penalizing those who engage in malicious behavior. This encourages users to contribute truthfully and responsibly, strengthening the integrity of the system.

5. **Efficient Uploading and Verification of NEWS items**
   Users can efficiently upload news items for fact-checking using a unique ID generated through a hashing algorithm. This ensures the integrity and uniqueness of each news item, facilitating accurate evaluation by fact-checkers.

6. **Bootstrapping Trustworthiness**
   Initially, fact-checkers are assigned trustworthiness scores based on the prevailing average. As they participate and contribute to the DApp, their scores are dynamically adjusted, allowing the system to bootstrap trustworthiness from the collective contributions of users over time.

3 Simulation

3.1 Theoretical guarantees through a Case Study

In this section, we explore the theoretical guarantees of our trustworthiness score:

Consider a scenario where "q" fraction of stakes are held by malicious voters. Among the honest voters, let "p" fraction be highly trustworthy, providing the correct vote with a probability of 0.9. The remaining "1-p" fraction provides the correct answer only with a probability of 0.7. Malicious users intentionally choose the wrong answer.

In this context, we demonstrate that the voting results are accurate almost surely if the following condition is met:

$$\frac{2}{7} \cdot (1+p) \geq q \tag{1}$$

Proof: Without loss of generality, let the truth value of the news item be 1, and consider the trustworthiness scores to be initially equal to 1 for everyone.

$$\begin{aligned}
\mathbb{E}[V_{0,j}] &= \mathbb{E}[\sum_{i=0}^{N} \mathbb{1}_{0,j} \cdot \text{stake}(i,k) \cdot \text{trustworthiness}(i,k)] \\
&= q \cdot 1 + p \cdot 0.1 + (1-p-q) \cdot 0.3 \\
&= 0.7q - 0.2p + 0.3 \\
\mathbb{E}[V_{1,j}] &= \mathbb{E}[\sum_{i=0}^{N} \mathbb{1}_{1,j} \cdot \text{stake}(i,k) \cdot \text{trustworthiness}(i,k) \\
&= p \cdot 0.9 + (1-q-p) \cdot 0.7 \\
&= 0.7 + 0.2p - 0.7q
\end{aligned}$$

Hence, for an accurate voting result, we require:

$$\begin{aligned}
0.7 + 0.2p - 0.7q &\geq 0.7q - 0.2p + 0.3 \\
\frac{2}{7} \cdot (1+p) &\geq q
\end{aligned}$$

This condition ensures the reliability of the voting results, given the specified proportions of trustworthy and malicious voters.

3.2 Validation of Case Study Findings through simulation

In this subsection, we present the simulation results based on the following assumptions:

1. We consider a scenario with N voters, each of whom casts a vote on every article.
2. Each voter holds an equal stake in the voting process.
3. The news items are categorized into a single category.
4. As this simulation focuses on evaluating the algorithm used to compute trustworthiness, we do not incorporate any form of ETH, whether for deposits or incentivizing honest voters.

3.2.1 Analysis

1. In accordance with the figures 1, 2, 3, and tables 1, 2, 3, the trustworthiness scores of malicious, trustworthy, and highly trustworthy voters tend towards 0.0, 0.7, and 0.9 respectively. Note that all the values of p and q satisfy condition (1).

2. As observed in figures 4, 5, and tables 4, 5, the trustworthiness scores of malicious, trustworthy, and highly trustworthy voters tend towards 1.0, 0.3, and 0.1 respectively. The chosen values of p and q result in a violation of condition (1). This outcome aligns with expectations, as the stake held by voters fails to meet the requirements of (1), resulting in voting results contrary to the ground truth.

Hence, we conclude that the trustworthiness score computation algorithm demonstrates efficient performance and rapid convergence.

Note: Symbols Used in Figures and Tables

Symbol Key:

- p: Percentage of malicious voters
- q: Percentage of very trust worthy voters
- N: Number of voters in the DApp
- T_{sim}: Simulation time (in ms)

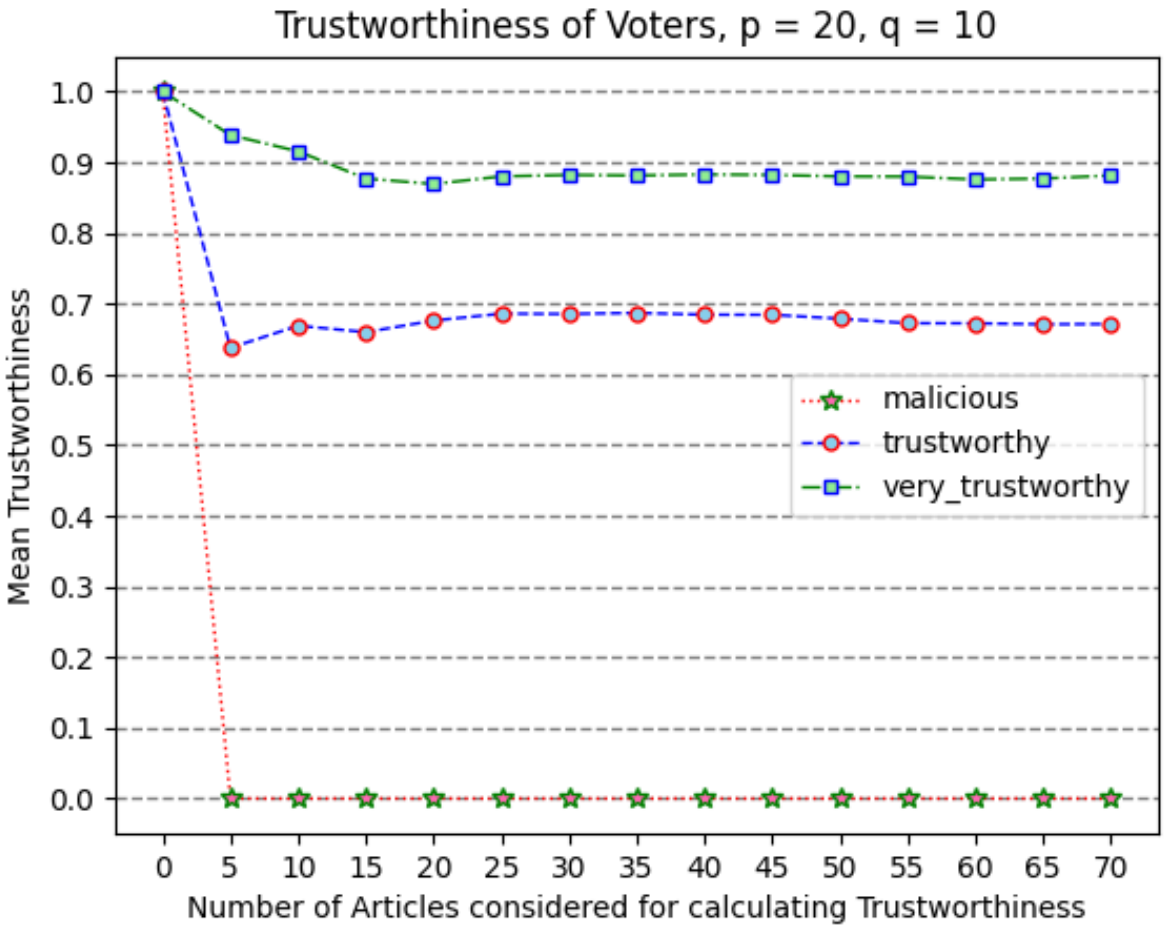

Figure 1: Mean Trust-worthiness score v/s Number of articles: $p = 20$, $q = 10$, $N = 70$, $T_{sim} = 75$ ms

Category	**Number of voters**	**Mean Trust-worthiness score**
Malicious	9	0.0000
Trustworthy	48	0.6711
Very trustworthy	13	0.8813

Table 1: Mean Trust-worthiness score for $p = 20$, $q = 10$, $N = 70$, $T_{sim} = 75$ ms

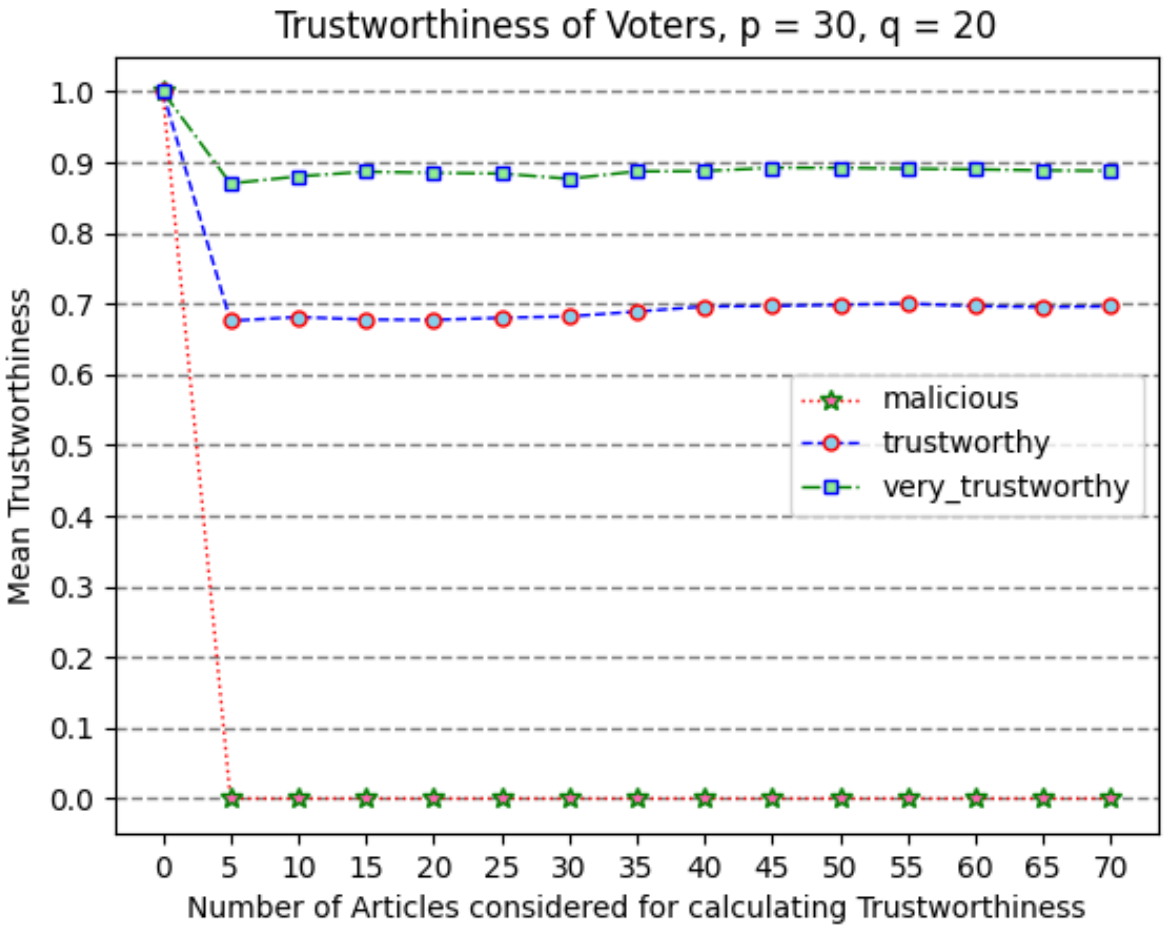

Figure 2: Mean Trust-worthiness score v/s Number of articles: $p = 30$, $q = 20$, $N = 70$, $T_{sim} = 75$ ms

Category	**Number of voters**	**Mean Trust-worthiness score**
Malicious	13	0.0000
Trustworthy	37	0.6961
Very trustworthy	20	0.8879

Table 2: Mean Trust-worthiness score for $p = 30$, $q = 20$, $N = 70$, $T_{sim} = 75$ ms

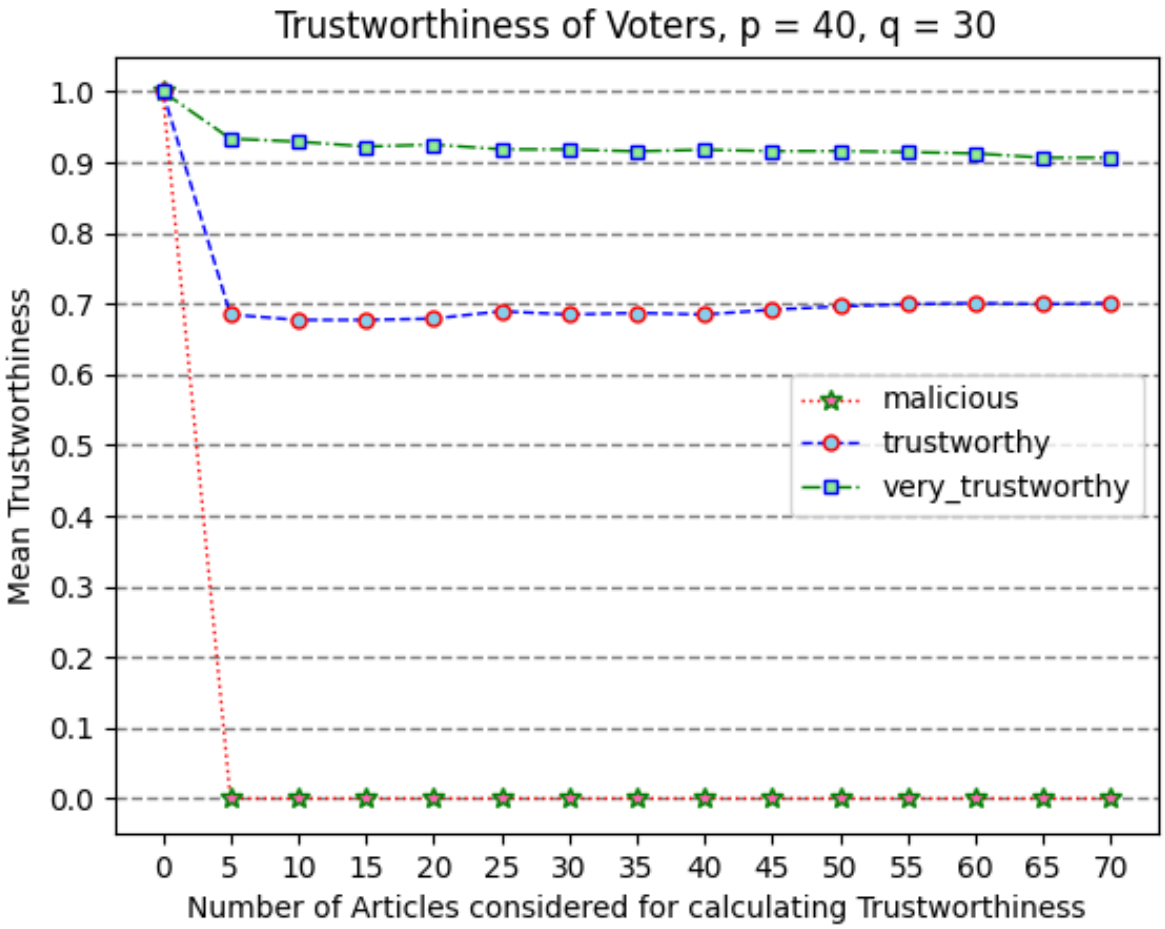

Figure 3: Mean Trust-worthiness score v/s Number of articles: $p = 40$, $q = 30$, $N = 70$, $T_{sim} = 75$ ms

Category	**Number of voters**	**Mean Trust-worthiness score**
Malicious	20	0.0000
Trustworthy	26	0.7005
Very trustworthy	24	0.9065

Table 3: Mean Trust-worthiness score for $p = 40$, $q = 30$, $N = 70$, $T_{sim} = 75$ ms

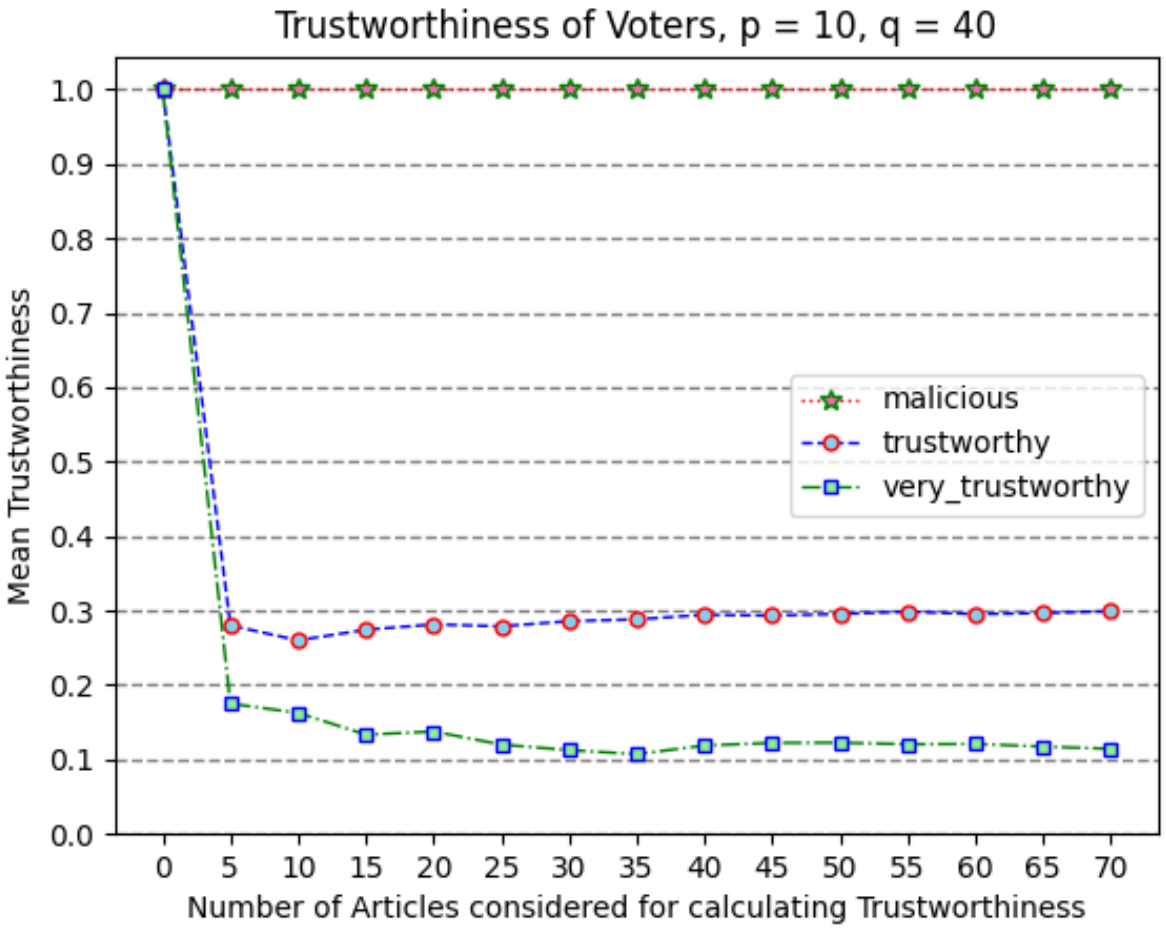

Figure 4: Mean Trust-worthiness score v/s Number of articles: $p = 10$, $q = 40$, $N = 70$, $T_{sim} = 75$ ms

Category	**Number of voters**	**Mean Trust-worthiness score**
Malicious	27	1.0000
Trustworthy	35	0.2992
Very trustworthy	8	0.1143

Table 4: Mean Trust-worthiness score for $p = 10$, $q = 40$, $N = 70$, $T_{sim} = 75$ ms

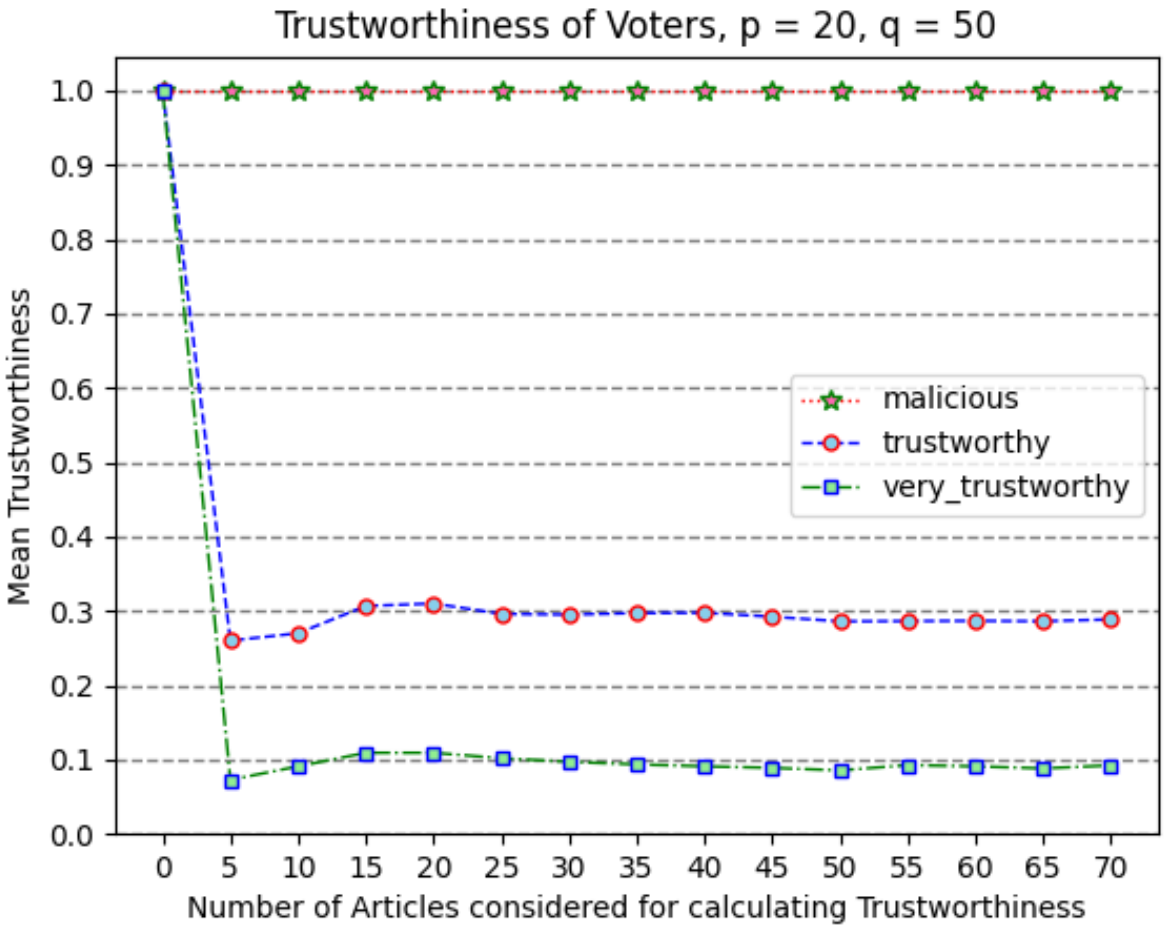

Figure 5: Mean Trust-worthiness score v/s Number of articles: $p = 20$, $q = 50$, $N = 70$, $T_{sim} = 75$ ms

Category	**Number of voters**	**Mean Trust-worthiness score**
Malicious	39	1.0000
Trustworthy	20	0.2886
Very trustworthy	11	0.0922

Table 5: Mean Trust-worthiness score for $p = 20$, $q = 50$, $N = 70$, $T_{sim} = 75$ ms