# Task Redundancy Strategy
## Based on **Volunteers' Credibility** for **Volunteer Thinking Projects**

### Lesandro Ponciano, Francisco Brasileiro, Guilherme Gadelha
### Universidade Federal de Campina Grande

## MOTIVATION OVERVIEW OF THE FIELD

**Volunteer thinking projects:** systems that gather people willing to contribute by executing tasks as a voluntary work.

**Task redundancy (or replication) is usually used to** achieve a satisfactory quality level in the results.
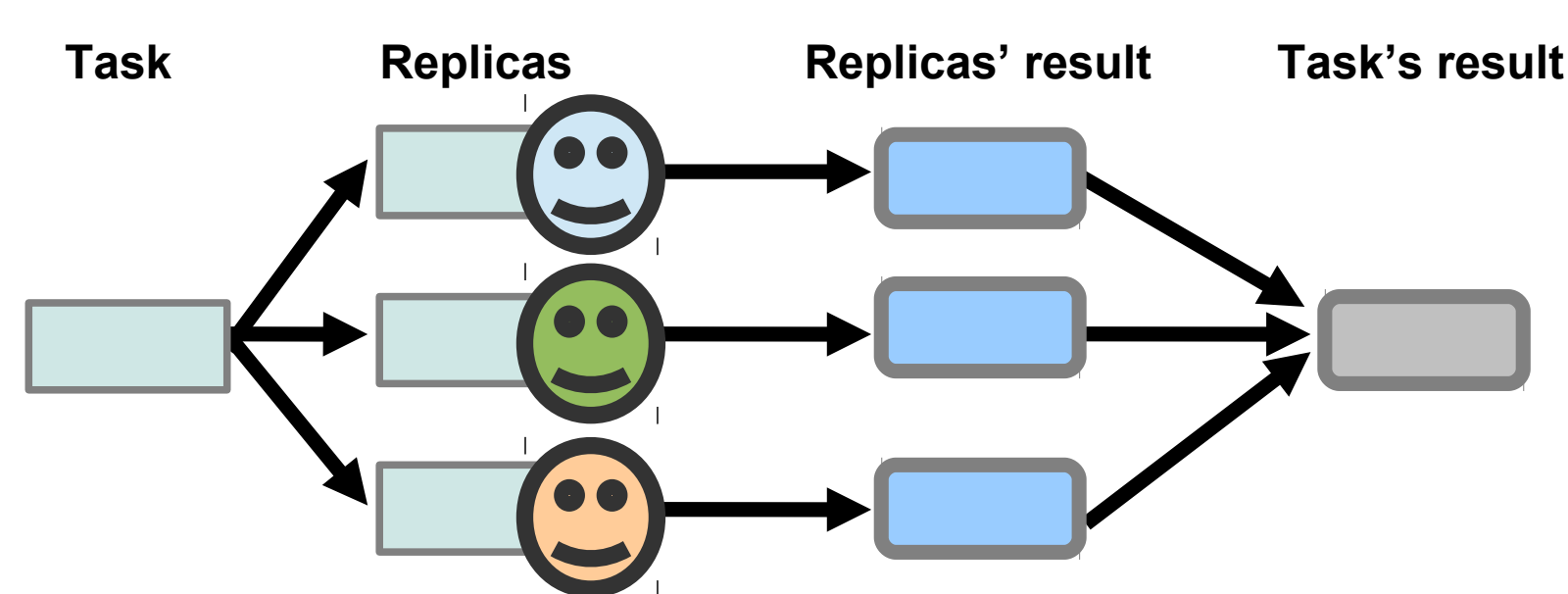


For example, in the first release of Galaxy Zoo each task was executed on average 38 times.

Lintott, C. J. et al. (2008). Galaxy Zoo: morphologies derived from visual inspection of galaxies from the Sloan Digital Sky Survey. MNRAS, 389(3), 1179-1189.

## PROBLEM REDUDANCY DEGREE

**Trade-off** between quality level of the results and the exploitation of volunteers contribution (quality *vs* cost).
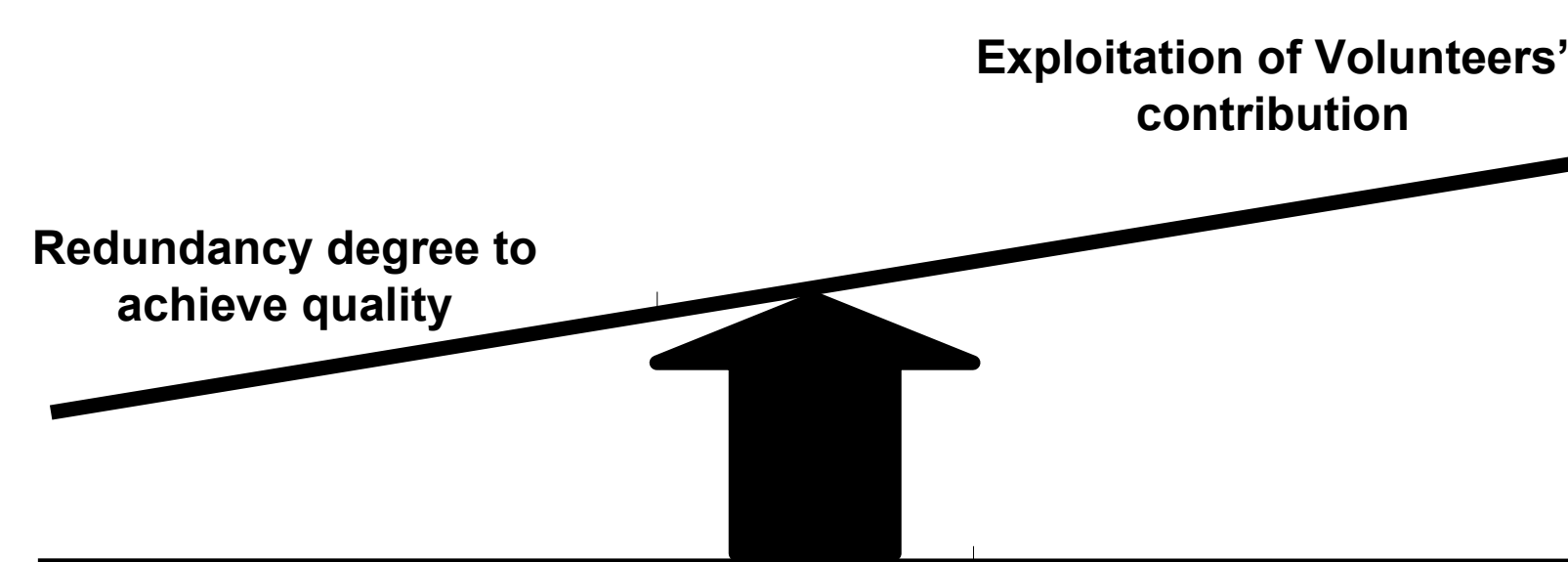


**How to optimize redundancy degree achieving a predefined quality level in the results?**

## CASE STUDY ACCURACY PATTERNS

**Volunteer accuracy over time** in a prototype of a volunteer thinking project: 520 tasks, 3,599 results, 1,098 volunteers.
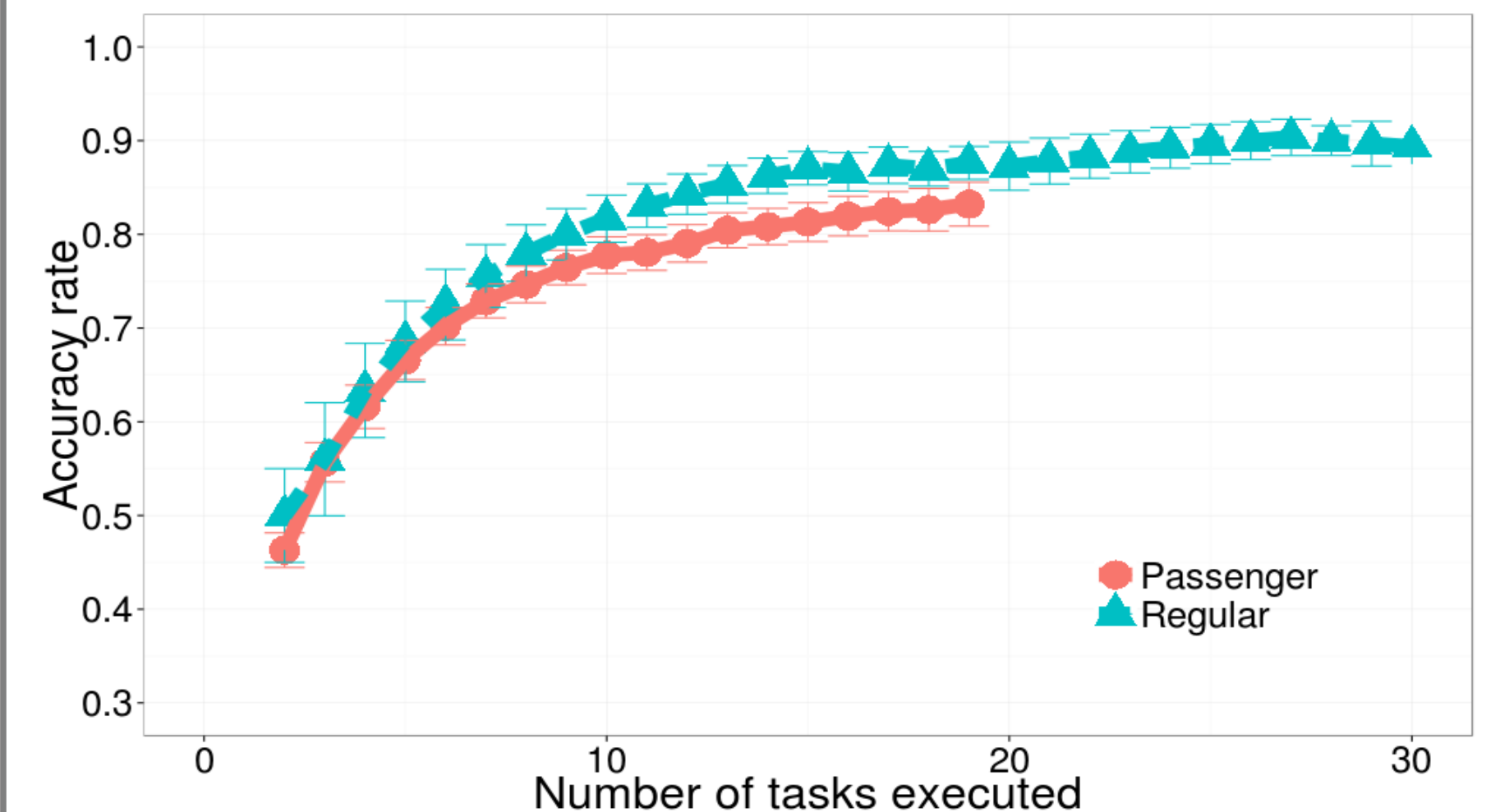


**Volunteers specialize when they execute large quantities of the same type of task.**

## OUR CREDIBILITY-BASED TASK REDUNDANCY STRATEGY

**Main idea:** Take advantage of volunteer accuracy patterns to route tasks in a way to optimize the number of replicas.

### 1 – CREDIBILITY MEASURES

- **Required result credibility** is task owner expected level of credibility;
- **Credibility of a volunteer** is the probability of the volunteer to execute a task "correctly";
- **Credibility of a result** is the credibility of the volunteer that generates it.

Sarmenta, Luis F. G. "Sabotage-tolerance mechanisms for volunteer Computing Systems." Future Gener Comput Syst 18.4 (2002): 561-572.

- **Credibility of a group of results** is the conditional probability of the group of results being correct and the the other groups for the same task being wrong.



| $C(G_A)$ | $C(G_B)$ |
|---|---|
| 0.9992 | 0.0003 |
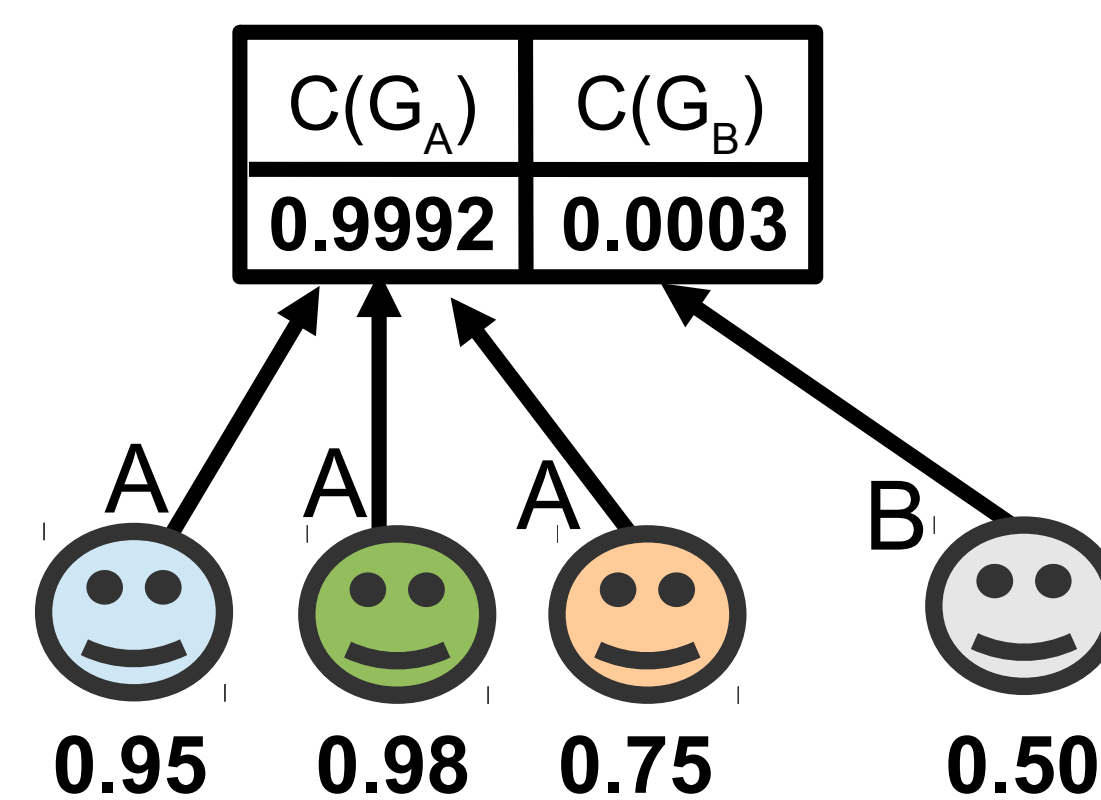
*Assume that answers can be either *A* or *B*:

### 2 – REPLICAS ROUTING

Tasks are routed to the available volunteer that has the highest credibility.

### 3 – ONDEMAND REPLICAS GENERATION

After each task execution, it checks the current task groups to verify whether more task replicas are required.

**Example 1)**
Req. credibility: 0.99
Credibility of $C(G_A)$: 0.00
Credibility of $C(G_B)$: 0.00
Which volunteer? **0.98**
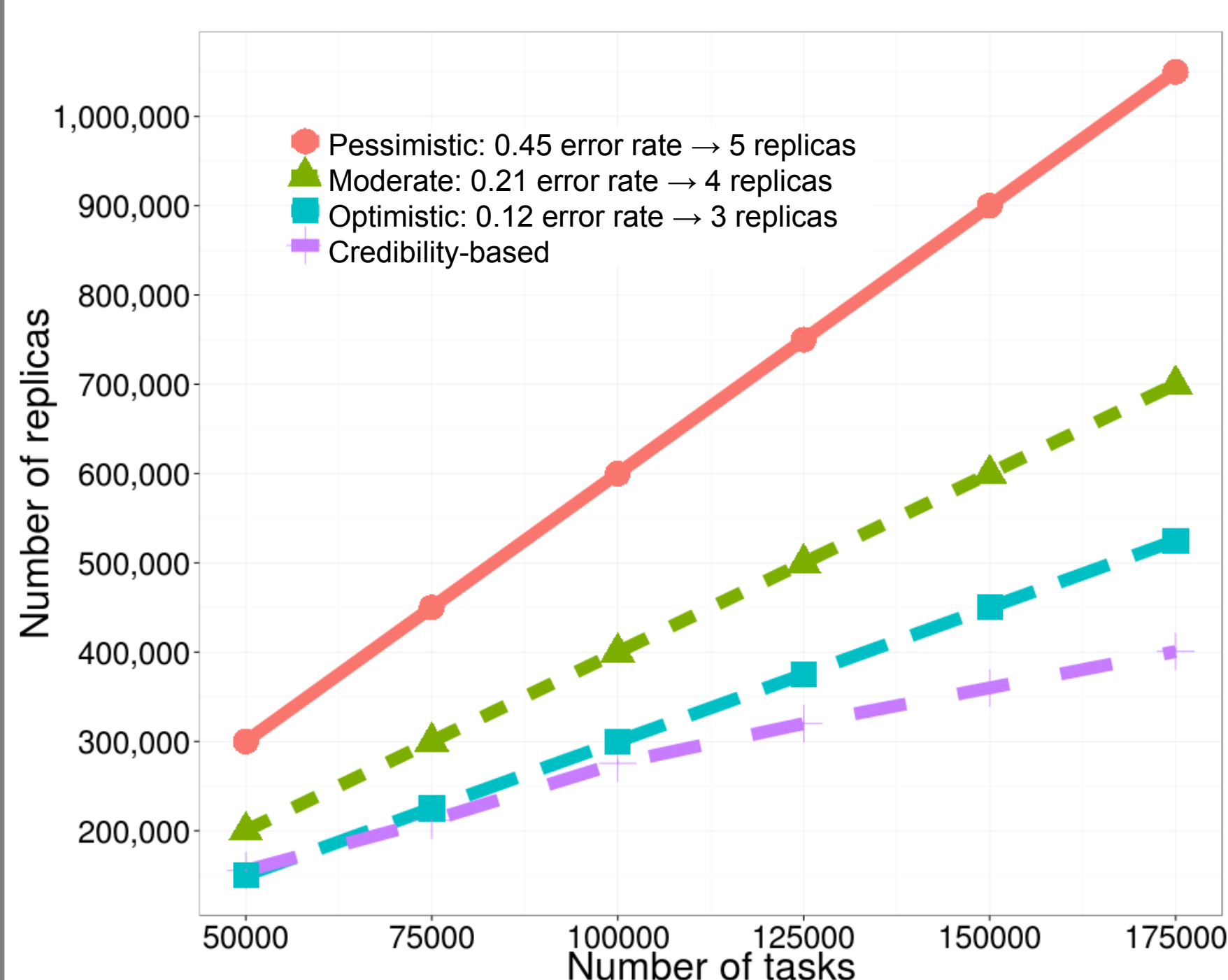More replicas? **Yes**

0.95   0.50
0.75   **0.98**

**Example 2)**
Req. credibility: 0.99
Credibility of $C(G_A)$: 0.93
Credibility of $C(G_B)$: 0.51
Which volunteer? **0.98**
More replicas? **No, if answer is A**

0.95   0.50
0.75   **0.98**

## SIMULATION RESULTS



**Fewer replicas compared to Pessimistic and Moderate. Similar or fewer replicas compared to Optimistic.**

## DISCUSSION and ONGOING WORK

**Problem/optimization:**

- In non-consensual (or very difficult) tasks, the scheduler may generate replicas indefinitely;
- The scheduler may be optimized to choose the volunteer that has credibility high enough to solve the task.

**Ongoing work:**

- Other ways to estimate volunteers' credibility considering individual and collective behavioral information;
- How soon we can accurately estimate volunteers credibility;
- How our strategy can automatically learn and take advantage of others accuracy patterns.

## ADDITIONAL INFORMATION

You can access the whole material used and produced in this study in the QR code below and at: http://goo.gl/mJgnyk

If you have any questions, please, let me know:

lesandrop@lsd.ufcg.edu.br