Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich

Distributed
Computing

Strategyproof Cardinal Peer Grading

Semster Thesis

Felix Rüsli

fruessli@student.ethz.ch

Distributed Computing Group
Computer Engineering and Networks Laboratory
ETH Zürich

Supervisors:
Diana Ghinea, Jakub Sliwinski
Prof. Dr. Roger Wattenhofer

January 22, 2021

Acknowledgements

I would like to thank Diana Ghinea and Jakub Sliwinski for their proficiency and their valuable insights during our weekly discussions.

Abstract

This work is about cardinal peer grading. An agent's grade is split up into accuracy and score. The accuracy serves as an incentive for agents to grade to participate truthfully in the peer grading, by giving them rewards for accurate gradings. The score reflects how its peers rate an agent. For these two parts a number of properties are studied. This works further includes a proof on the equality of independence and bipartite graphs, and a proof on the feasibility of score monotonicity.

Contents

Chapter 1

Introduction

As we have seen in the last few months, during the COVID-19 pandemic, online classes become more and more important, whether it be from the local primary schools or for whole online courses at a university. This can for example be a MOOC (massive open online course) which can have hundreds of enrolled students. Grading all of these students is extremely time-consuming for the instructors. This is where peer grading comes into play: students grade each other. The instructors then only have to do spot-checking if they choose to do so.

There are multiple modes for peer grading. The most common ones are cardinal, ordinal or selective peer grading. In cardinal peer grading the grades are in form of a cardinal number that the students assign to the work of their peers. In ordinal peer grading the students either compare the works against one another, e.g., paper A is better than paper B but worse than paper C, or the students make a ranked list of the works, e.g., the best one is essay D, then comes essay F, etc. Finally, in the selection problem the students choose the k best works amongst the given works. This last mode is for example used in scientific conferences.

This work is about cardinal peer grading. An agent's grade is split up into accuracy and score. For those two a number of properties are studied. The accuracy serves as an incentive for agents to grade to participate truthfully in the peer grading, by giving them rewards for accurate gradings. And the score reflects how its peers rate an agent. For all of this a fixed point approach is taken.

In chapter 2, one can find some background information on peer grading. Chapter 3 contains most of the terms and variables used in this work for look-up purposes. Chapter 4 discusses the properties used, some in more detail in the following subchapters. Chapter 5 argues the approaches and models used. In chapter 6 we construct a family of functions that calculates the accuracy and with that the incentive for the students to participate in the peer grading honestly. In chapter 7 we put everything together and give an idea how to calculate the final grade. Chapter 8 is about the previously conducted work in this area, especially about Toby Walsh's PeerRank algorithm [1]. Finally, chapter 9 concludes and summarizes what has been achieved.

Chapter 2

Background Information

Peer grading is a system where subjects which are all at the same level assess each other's work, for example students which give feedback on the essays of their classmates. The main advantage of such methods is the time saved by the teacher or instructor, and has also been to shown improve the students understanding of the studied material [2]. It is generally not known who these peers are, whether they are students or scientists, the peers are called agents. So, one or multiple agents grade the work of another agent.

2.1 Incentive

The assessments agents give can however not be trusted since they are assumed to be selfish, meaning their main goal is to have the highest grade possible with as little effort as possible. For this reason a mechanism has to be used in order to convince the agents to grade honestly. This mechanism is usually an incentive. The incentive is a reward or a penalty in the grade the agent receives for the accuracy of their grading [3, 4].

2.2 The PeerRank Algorithm

Toby Walsh's PeerRank Algorithm [1] is an elegant and simple solution for the peer grading problem. He assumes that agents with a high grade have a good knowledge of the subject and therefore are also best fit to assess the work of other agents. The PeerRank rule as a fixed point function over multiple iterations.

The algorithm fulfills the following properties: The grade always stays in the given domain, an unanimous decision of the grading agents lead to that decision as the received grade, every grade in the domain is possible, and all given grades have an impact on the end result [1]. However, according to the author, the PeerRank method is not monotonic [1].

In [1] m denotes the number of participating agents. $A_{i,j}$ is the grade agent i receives from agent j. X_i^n is the final grade agent i receives after n iterations of the algorithm. α indicates how much the weighted average of the received grades counts towards the grade and β how much the incentive counts towards the grade. The grades, α and β are normalized to be in the domain $[0, 1]$.

The PeerRank Algorithm is the following:

$X_i^0 = \frac{1}{m}\Sigma_j A_{i,j}$

$X_i^{n+1} = (1-\alpha) * X_i^n + \frac{\alpha}{\Sigma_j X_j^n}\Sigma_j X_j^n * A_{i,j}$

However, this mechanism doesn't have an incentive like mentioned above. For this reason Mr. Walsh expanded his PeerRank rule to the Generalized PeerRank algorithm.

$X_i^0 = \frac{1}{m}\Sigma_j A_{i,j}$

$X_i^{n+1} = (1-\alpha-\beta) * X_i^n + \frac{\alpha}{\Sigma_j X_j^n}\Sigma_j X_j^n * A_{i,j} + \frac{\beta}{m} * \Sigma_j 1 - |A_{j,i} - X_j^n|$

Chapter 3

Naming

The terms listed below are used for this semester project. The lists below serve as a lookup-table on what the terms mean and what variables are used for them throughout the project.

3.1 Meaning

1. *score*: The *score* is the evaluation of the work an agent did. For example S_{22} is the joined grade of agent 22.
2. *accuracy*, incentive: The *accuracy* evaluates the ability of an agent to grade its peers. In other works this is sometimes called incentive.
3. *total grade*: The *total grade* is the final evaluation the agents receives. It is partly dependent on the *score* and the *accuracy*.
4. *individual grade*: The *individual grades* are the evaluations the agents give to and receive from their peers. For example '$g_{6,2} = 13$' means agent 6 received an individual grade of 13 from agent 2.
5. *fixed point*: When an agent gives an *individual grade* which coincides with its *fixed point*, this agent receives the highest possible accuracy. The *fixed point* depends on the other agents that grade the same peer.
6. $u(.)$: $u(.)$ is a function with an *individual grade* as input. The function is mainly used in the section about shift invariance.
7. *weight-accuracy-function*: The *weight-accuracy-function* calculates the weight from the accuracy of an agent in a weighted average.
8. *agent*: The peers in the peer grading problem are called *agents*.

3.2 Variables in the n^{th} iteration

A fixed point approach was taken. Thus, the mechanism uses iterations.

1. *score*: S_i^n
2. *accuracy*: A_j^n
3. *total grade*: X_j^n, G_j^n
4. *individual grade*: $g_{i,j}$
5. *fixed point*: $F_{i,j}^n$
6. $u(.)$
7. *weight-accuracy-function*: $t(.)$

Where some terms like the *scores* or *accuracies* change over iterations the *individual grades* stay constant. Sometimes for accuracy and score the iteration will not be written for brevity.

Chapter 4

Preliminaries

This chapter lists a handful of useful properties we aim to achieve. Some of them are taken from the PeerRank method by Toby Walsh [1] and marked with *(PR)* as such and some are taken from [5] and marked with *(Sel)*.

The total grade, the score and also the accuracy should stay in a given domain, e.g., [0,1], [1,10] or [1,6].

Definition 4.1 (Domain (PR))**.** The total grade, the score and the accuracy stay in their given domain. The domains may differ between the total grade, the score and the accuracy.

All scores and total grades within their given domain are possible.

Definition 4.2 (Non-Imposition (Sel)/No Discrimination (PR))**.** All sets of scores and total grades within their respective domain are reachable.

All individual grades should have an influence on the score of the agent they are evaluating.

Definition 4.3 (No Score Dummy (PR))**.** If agent j changes their evaluation of agent i and all other individual grades stay the same, i's score will change.

Definition 4.4 (No Dictators)**.** A single agent can not solely decide the score or the total grade of another agent.

If all agents that grade agent i have a unanimous decision on the grade of agent i, then that grade should be the score. However, the total grade is still depending on i's accuracy.

Definition 4.5 (Unanimity (PR))**.** If all agents give an agent the individual grade k, then the *score* given to this agent will be k.

Definition 4.6 (Continuity)**.** A function is continuous if for all $\delta > 0$, there is an $\epsilon > 0$, such that, if you change the argument by less than ϵ, the value changes by less than δ.

In the continuity property, for us an argument is an individual grade. The value can be the score, the accuracy or even the total grade.

Definition 4.7 (Surjectivity)**.** A function f from a set X to a set Y is surjective, if for every element y in the codomain Y of f, there is at least one element x in the domain X of f such that f(x) = y. [7]

The influence and the grade an agent has should not be depending on their name.

Definition 4.8 (Anonymity (Sel))**.** Permutation of agents leads to the same permutation of total grades.

Definition 4.9 (Strategyproofness)**.** An agent fares not worse by grading the other agents truthfully.

Definition 4.10 (Impartiality (Sel))**.** Agents can't influence their own score.

Definition 4.11 (Shift invariance)**.** If all agents shift their individual grade for agent j by $+a$, then the score of agent j also shifts by $+a$.

Definition 4.12 (Score Monotonicity)**.** If agent j increases their evaluation of agent i and all other individual grades stay the same, i's score will

- increase if $A_i > 0$, or
- decrease if $A_i < 0$,

where A_i is i's accuracy.

Definition 4.13 (Accuracy Monotonicity)**.** If agent i grades agent j 'better' and the other agents grade agent j the same way, agent i's accuracy should increase. 'Better' means that i's individual grade is closer to its fixed point.

Definition 4.14 (Score Independence)**.** Changing $g_{j,i}$ can not change the score of agent i, $\forall(i,j), i \neq j$.

Definition 4.15 (Accuracy Independence)**.** Changing $g_{j,i}$ can not change the accuracy of agent j, $\forall(i,j), i \neq j$.

Definition 4.16 (Symmetry)**.** If all individual grades are flipped, then all the scores are also flipped. And by flipped is meant, the maximum grade minus the individual grade and the maximum grade minus the score, e.i., $g_{max} - grade + g_{min}$, respectively $g_{max} - score + g_{min}$.

In the following we take a closer look at some of the previously mentioned properties.

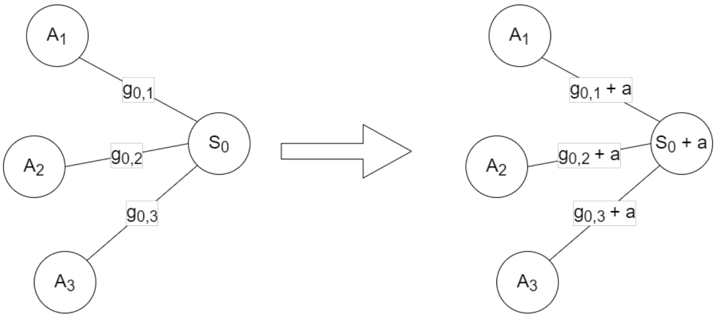

Figure 4.1: Shift Invariance

4.1 Shift Invariance

Definition 4.17 (Shift invariance)**.** If all agents shift their individual grade for agent j by $+a$ the the score of agent j also shifts by $+a$.

Shift invariance makes sure that the accuracy of the grading agents stays the same, if all the grading agents shift their individual grades by the same amount, since the distance to the fixed point doesn't change.

Example 4.18. The weighted average, $S_j = \frac{\Sigma_{i=1}^n g_{j,i} * A_i}{\Sigma_{i=1}^n A_i}$, is shift invariant.

Proof. $\frac{\Sigma_{i=1}^n (g_{j,i}+a) * A_i}{\Sigma_{i=1}^n A_i} = \frac{\Sigma_{i=1}^n g_{j,i} * A_i + a * A_i}{\Sigma_{i=1}^n A_i} = \frac{\Sigma_{i=1}^n g_{j,i} * A_i}{\Sigma_{i=1}^n A_i} + \frac{a * \Sigma_{i=1}^n A_i}{\Sigma_{i=1}^n A_i}$
$= \frac{\Sigma_{i=1}^n g_{j,i} * A_i}{\Sigma_{i=1}^n A_i} + a = S_j + a$ □

Remark 4.19. In fact, there is shift invariance for any arbitrary weight function of the accuracy, $t(A_i)$.

Proof. $\frac{\Sigma_{i=1}^n (g_{j,i}+a) * t(A_i)}{\Sigma_{i=1}^n t(A_i)} = \frac{\Sigma_{i=1}^n g_{j,i} * t(A_i) + a * t(A_i)}{\Sigma_{i=1}^n t(A_i)} = \frac{\Sigma_{i=1}^n g_{j,i} * t(A_i)}{\Sigma_{i=1}^n t(A_i)} + \frac{a * \Sigma_{i=1}^n t(A_i)}{\Sigma_{i=1}^n t(A_i)}$
$= \frac{\Sigma_{i=1}^n g_{j,i} * t(A_i)}{\Sigma_{i=1}^n t(A_i)} + a = S_j + a$ □

Remark 4.20. The form $S_j = \frac{\Sigma_{i=1}^n u(g_{j,i}) * t(A_i)}{\Sigma_{i=1}^n t(A_i)}$ is also shift invariant, with $u(.)$ being an arbitrary function with $u(x + a) = u(x) + a$, and $t(.)$ is an arbirtrary function.

Proof. $\frac{\Sigma_{i=1}^n u(g_{j,i}+a) * t(A_i)}{\Sigma_{i=1}^n t(A_i)} = \frac{\Sigma_{i=1}^n u(g_{j,i}) * t(A_i) + a * t(A_i)}{\Sigma_{i=1}^n t(A_i)} = \frac{\Sigma_{i=1}^n u(g_{j,i}) * t(A_i)}{\Sigma_{i=1}^n t(A_i)} + \frac{a * \Sigma_{i=1}^n t(A_i)}{\Sigma_{i=1}^n t(A_i)}$
$= \frac{\Sigma_{i=1}^n u(g_{j,i}) * t(A_i)}{\Sigma_{i=1}^n t(A_i)} + a = S_j + a$ □

Theorem 4.21. $u(x + a) = u(x) + a \Rightarrow u(x) = x + u(0)$

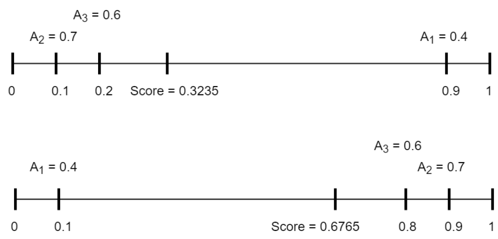

Figure 4.2: Symmetry example

Proof. Simply plug in $x = 0$ and $a = x'$ on the right side.

$u(0 + x') = u(x') = u(0) + x'$ □

Theorem 4.22. *The form* $S_j = \frac{\Sigma_{i=1}^n u(g_{j,i}) * t(A_i)}{\Sigma_{i=1}^n t(A_i)}$ *is shift invariant and fulfills unanimity, if* $u(0) = 0$*.* $t(.)$ *is an arbitrary function.*

Proof. We know from *Theorem 4.21* that $u(x) = x + u(0)$. In order to fulfill unanimity we know that S_j has to be 0 if all $g_{j,i}$ are 0. If all $g_{j,i}$ are 0, we have $S_j = 0 = \frac{\Sigma_{i=1}^n u(0) * t(A_i)}{\Sigma_{i=1}^n t(A_i)}$ for any A_i.

Since $t(.)$ is arbitrary we conclude that $u(0) = 0$. □

Corollary 4.23. *With* Theorem 4.21 *and* Theorem 4.22 *we conclude that for the form* $S_j = \frac{\Sigma_{i=1}^n u(g_{j,i}) * t(A_i)}{\Sigma_{i=1}^n t(A_i)}$ *to be shift invariant and fulfill unanimity we require that* $u(x) = x$*.*

4.2 Symmetry

Definition 4.24 (Symmetry)**.** If all individual grades are flipped, then all the scores are also flipped. And by flipped is meant, the maximum grade minus the individual grade and the maximum grade minus the score, e.i., $g_{max} - grade + g_{min}$, respectively $g_{max} - score + g_{min}$.

Example: Let us assume agents $1, 2$ and 3 grade the agent 0 on a scale of 0 to 1. The score is calculated with a weighted average, where the weights are the accuracies, $t(A_i) = A_i$. The individual grades are $g_{0,1} = 0.1, g_{0,2} = 0.2$ and $g_{0,3} = 0.9$. The accuracies of the grading agents are $A_1 = 0.7, A_2 = 0.6$ and $A_3 = 0.4$. Therefore, the score is $S_0 = \frac{\Sigma_{i=1}^n g_{0,i} * t(A_i)}{\Sigma_{i=1}^n t(A_i)} = \frac{0.1*0.7+0.2*0.6+0.9*0.4}{0.7+0.6+0.4} = 0.3235$.

Now the individual grades are flipped, e.g., for agent 1 we have $g_{0,1} = 1 - 0.1 = 0.9$. The other new individual grades are $g_{0,1} = 0.8$ and $g_{0,3} = 0.1$. The new score is $S_0 = \frac{0.9*0.7+0.8*0.6+0.1*0.4}{0.7+0.6+0.4} = 0.6765 = 1 - 0.3235$ [*Figure 4.2*].

Theorem 4.25. *The weighted average with an arbitrary weighting function $t(.)$ as scoring rule is symmetric.*

Proof. The score is $S_j = \frac{\Sigma_{i=1}^{n} g_{j,i} * t(A_i)}{\Sigma_{i=1}^{n} t(A_i)}$.
$\frac{\Sigma_{i=1}^{n}(g_{max}-g_{j,i}+g_{min})*t(A_i)}{\Sigma_{i=1}^{n} t(A_i)} = \frac{\Sigma_{i=1}^{n}(g_{max}+g_{min})*t(A_i)-g_{j,i}*t(A_i)}{\Sigma_{i=1}^{n} t(A_i)} = \frac{(g_{max}+g_{min})*\Sigma_{k=1}^{n} t(A_k)-\Sigma_{i=1}^{n} g_{j,i}*t(A_i)}{\Sigma_{i=1}^{n} t(A_i)}$
$= \frac{(g_{max}+g_{min})*\Sigma_{k=1}^{n} t(A_k)}{\Sigma_{i=1}^{n} t(A_i)} - \frac{\Sigma_{i=1}^{n} g_{j,i}*t(A_i)}{\Sigma_{i=1}^{n} t(A_i)} = g_{max} - S_j + g_{min}$ □

4.3 Strategyproofness

Definition 4.26 (Strategyproofness)**.** An agent fares not worse by grading the other agents truthfully.

Definition 4.27 (Impartiality (Sel))**.** Agents can't influence their own score.

This work encourages agents to grade truthfully with the an incentive, called accuracy. Accuracy monotonicity makes sure that better grading gets rewarded accordingly.

Thanks to score independence, impartiality is achieved.

4.4 Independence

Definition 4.28 (Score Independence)**.** Changing $g_{j,i}$ can not change the score of agent i, $\forall(i,j), i \neq j$.

Definition 4.29 (Accuracy Independence)**.** Changing $g_{j,i}$ can not change the accuracy of agent j, $\forall(i,j), i \neq j$.

If an agent had no influence on their own total grade they would have no reason to be dishonest in their review of the other agents. Thus, an agent should have as little influence as possible on their total grade. While the individual grades an agent gives have of course an impact on this agents accuracy they shouldn't have an impact on their score. This is the idea behind *score independence.*

On the other hand, the reception of individual grades should not influence the accuracy of an agent which is reflected in the *accuracy independence.*

This work proves later that score independence and accuracy independence are actually equivalent. *Independence* merges the score and the accuracy independence into one property.

Definition 4.30 (Independence)**.** Changing $g_{j,i}$ can neither change the score of agent i nor the accuracy of agent j, $\forall(i,j), i \neq j$.

4.4.1 Independence $\iff$ Bipartition

A group of peers are given a *task*. This group of peers is afterwards asked to grade the work of their peers. In the following we will call the peers *agents*.

The final evaluation of an agent is based on two parts. The first part is based on how well the agent performed at the given task. And the second part is based on how well the agent performed in grading their peers.

Definition 4.31 (Individual Grade)**.** The individual grade $g_{j,i}$ is the grade agent j receives from agent i, $i \neq j$.

Definition 4.32 (Score)**.** The *Score* is the evaluation of the work an agent did.

Definition 4.33 (Accuracy)**.** The *Accuracy* evaluates the ability of an agent to grade its peers.

For the rest of the proof we will make two assumptions.

Assumption 4.34 (No Dictators)**.** Every agent that is graded is graded by at least two other agents.

Definition 4.35 (No Score Dummy)**.** If agent j changes their evaluation of agent i and all other individual grades stay the same, i's score will change.

Assumption 4.36 (No Dummy)**.** We assume that we have no score dummy and accuracy monotonicity.

Lemma 4.37. *Score independence and accuracy independence are equivalent. Thus we get the property called* Independence*.*

Theorem 4.38. *Independence is equivalent to a bipartite S-A-graph (score-accuracy-graph).*

Definition 4.39 (Score-Accuracy-Graph)**.** The score-accuracy-graph is an undirected graph where for each agent i we have an *Accuracy Node* A_i and a *Score Node* S_i. Both A_i and S_i are connected to the *dummy node* d_i. If $g_{j,i}$ exists then we add an edge (A_i, S_j).

Remark 4.40. From *Assumption 4.36* we know that, if $g_{j,i}$ changes then S_j changes and thus A_i also changes.

Proof. First of all let us prove *Lemma 4.37*.

We prove that accuracy and score indepence are equivalent by showing that if we don't have accuracy independence, respectively score independence, then we also can't have the other one. *No accuracy independence* means there exists a pair (i, j) where changing $g_{j,i}$ *can* change the accuracy of agent j. And *no score independence* means there exists a pair (i, j) where changing $g_{j,i}$ *can* change the

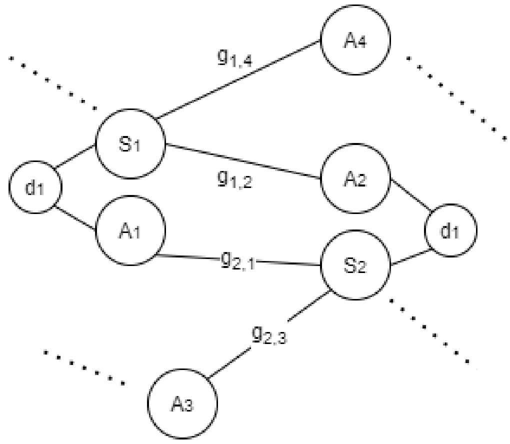

Figure 4.3: Score-Accuracy-Graph

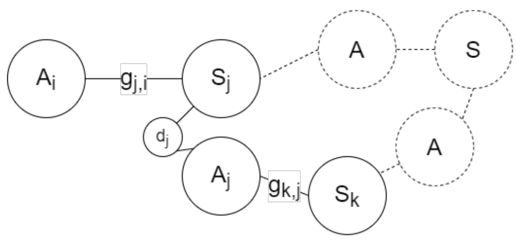

Figure 4.4: No Accuracy Independence

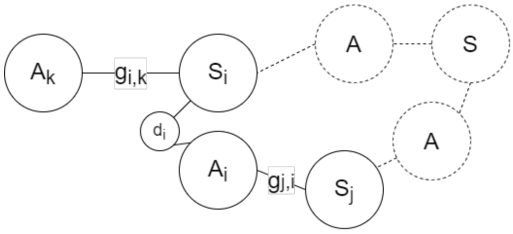

Figure 4.5: No Score Independence

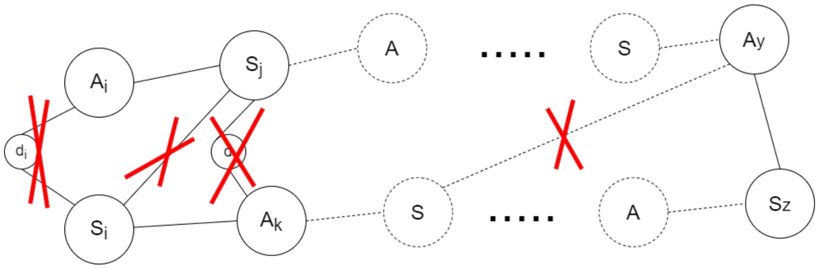

Figure 4.6: Odd Path

score of agent i. We will use the same variables as in a S-A-graph, namely A_i and S_i, as the accuracy and the score of agent i respectively.

Let us assume that we have no accuracy independence. Thus there exists an individual grade $g_{j,i}$ which can influence A_j. Clearly $g_{j,i}$ influences S_j. And S_j influences not only A_i but also the accuracy of at least one other agent, since we don't have a dictatorship. Because we assumed that we don't have accuracy independence we know that A_j also gets influenced by $g_{j,i}$. This has to happen through the change of a score of agent k, without loss of generality. S_k is also influenced through the accuracy of another agent, apart from A_j. This accuracy node is connected to another score node, and so forth until we reach S_j. Therefore $g_{k,j}$ would also be able to influence S_j, and thus we have no score independence.

Now let us assume we have no score independence. $g_{j,i}$ is now able to influence not only S_j but only S_i. S_i is influenced by the accuracy of at least two other agents. Let us call one of these agents agent k. Clearly, $g_{i,k}$ can also influence A_i since we don't have score independence. Therefore we also can't have accuracy independence. □

Lemma 4.41. *There exists an odd-length path from A_i to S_i in the S-A-graph if and only if independence* doesn't *hold.*

Proof. We first assume that independence doesn't hold and show that an odd path from A_i to S_i exists for at least one agent i. Since we have no independence we know that there is a path from S_i to A_i that doesn't go over any dummy nodes. We can't use dummy nodes in our path, since they do not propagate any influence. If no dummy nodes are allowed all paths from an accuracy node to a score node have an odd length. Therefore the assumed path from S_i to A_i also has to be odd.

Next, we build an odd-length path from A_i to S_i and try to keep the S-A-graph independent. We obviously can't use d_i since this path would have a length of two and thus be even. A_i could directly connect to S_i. This path would be odd-length but this graph would clearly not fulfill independence. A_i can also be connect to the score node of an agent j, S_j. Agent j wasn't in our path so far. S_i is connected to at least two accuracy nodes, we call one of them A_k. If $k = j$ we have the same problem that we started with, just with different indices. Therefore, we assume $j \neq k$. A_k and S_j could now directly be connected. This would give us an odd path from S_i to A_i but we would also have no independence since A_i could directly influence S_i. Or both nodes can be connected to new agents. This can go on until we run out of new agents. At the end we have two nodes, A_y and S_z. A_y can now either be either be connected to a score node of another agent and destroy the independence. Or it can be connected to a dummy node and form an even path. Of course A_y has to pick a score node of another agent and thus we have no independence. Similarly, S_z can either pick an accuracy node of another agent or a dummy node. And again S_z has to pick an accuracy node and we have no independence. Therefore, if there exists an odd-length path from A_i to S_i for at least one agent i then independence can't hold.

□

Lemma 4.42. *We have an odd-length path from A_i to S_i in the S-A-graph if and only if the S-A-graph is* not *bipartite.*

Proof. It is known that a bipartite graph is equivalent to a graph that doesn't have any odd cycles [6]. An odd-length path from A_i to S_i is easily made into an odd cycle with the inclusion of the dummy node d_i. Thus, if we have an odd-length path from A_i to S_i in the S-A-graph then the S-A-graph is *not* bipartite.

Next, we prove that if there exists an odd cycle in a S-A-graph, then there exists an agent i where we have an odd-length path from A_i to S_i. Let us assume we have the odd cycle c in the S-A-graph. First, we prove by contradiction that there is an agent i such that A_i and S_i are both on the cycle.

Let us assume there is no such agent i. From this follows that there is no dummy no in c and, therefore, for every agent k in c, we either have A_k or S_k in c. As c is odd, we either have an edge between two accuracy nodes or two score nodes. This however contradicts the definition of an S-A-graph. Therefore, we conclude that there exists an agent i such that both A_i and S_i are in c, which is a contradiction to our assumption.

Since the cycle c is odd and an agent i with both A_i and S_i in c exists, there is both an odd and an even path from S_i to A_i in order to form an odd cycle. □

Combining *Lemma 4.41* and *Lemma 4.42* proves *Theorem 4.38*.

4.5 Score Monotonicity

Score Monotonicity is fulfilled if the scoring rule is a simple average of the individual grades, $S_i = \frac{\sum_{i=1}^{m} g_{j,i}}{m}$, where m is the number of agents that grade agent j. However, it is less obvious for more complicated scoring rules whether they are score monotonic or not.

Below the score monotonicity of a very simple peer grading mechanism is proven.

4.5.1 Score Monotonicity Proof

First, we have to take some assumptions.

Assumption 4.43. The individual grades are in the domain $[0, 1]$.

Assumption 4.44. We look at the scoring rule $S_j = \frac{\sum_{i=1}^{m} g_{j,i} A_i}{\sum_{i=1}^{m} A_i} \in [0, 1]$, where m is the number of agents that grade agent j.

Assumption 4.45. The accuracy, $a_{j,i}$, of the individual grade $g_{j,i}$ is in the domain $[0, 1]$.

Assumption 4.46. The accuracy of agent i is calculated with $A_i = \frac{\sum_{j=1}^{m} a_{j,i}}{7m} + \frac{6}{7} \in [\frac{6}{7}, 1]$.

In the following we will look at a scenario, where all individual grades stay the same, except for agent 1, which increases by D_g. The change of an individual grade is called D_g and the following change of accuracy for agent j is called D_{aj}. We want to show that with this change the score of agent i will increase. We mark the variables after the change with a prime, e.g. A'_3.

We assume that the accuracy of agent 1 after the change in the individual grade $g_{j,1}$ can be anywhere in the domain of the accuracy if D_g is at its maximum of 1. However, the accuracy of the other agents is a factor of m smaller and thus bound by $\frac{1}{7m} D_g$.

Assumption 4.47. $\frac{1}{7} D_g \geq |D_{a1}|$

Assumption 4.48. $\frac{1}{7m} D_g \geq |D_{aj}|, j \neq 1$

Thus, $A'_1 - A_1 = D_{a1} \in [-\frac{1}{7}D_g, \frac{1}{7}D_g]$ and $A'_j - A_j = D_{aj} \in [-\frac{1}{7m}D_g, \frac{1}{7m}D_g], j \neq 1$. From this follows $\sum_{j=2}^{m} A'_j - A_j = \sum_{j=2}^{m} D_{aj} \in [-\frac{m-1}{7m}D_g, \frac{m-1}{7m}D_g]$.

Theorem 4.49. *A peer grading mechanism with the above assumptions is score monotonic.*

Proof. We show that $S'_i > S_i$ if the individual grade $g_{i,1}$ increased to $g'_{i,1} = g_{i,1} + D_g$. We use if $a > b$, $A \geq a$ and $b \geq B$ then $A > B$.

$S'_i = \frac{\sum_{i=1}^{m} g'_{i,j} A'_j}{\sum_{j=1}^{m} A'_j} > \frac{\sum_{j=1}^{m} g_{i,j} A_j}{\sum_{j=1}^{m} A_j} = S_i$

$\iff \sum_{j=1}^{m} A_j (\sum_{j=2}^{m} g_{i,j} A'_j + g'_{i,1} A'_1) > \sum_{j=1}^{m} g_{i,j} A_j \sum_{j=1}^{m} A'_j$

We subtract $\sum_{j=1}^{m} A_j \sum_{j=1}^{m} g_{i,j} A_j$ from both sides.

$\sum_{j=1}^{m} A_j (\sum_{j=2}^{m} g_{i,j} A'_j + g'_{i,1} A'_1 - \sum_{j=1}^{m} g_{i,j} A_j) > \sum_{j=1}^{m} g_{i,j} A_j (\sum_{j=1}^{m} A'_j - \sum_{j=1}^{m} A_j)$

$\iff \sum_{j=2}^{m} g_{i,j} A'_j + g'_{i,1} A'_1 - \sum_{j=1}^{m} g_{i,j} A_j > S_i (\sum_{j=1}^{m} A'_j - A_j)$

We want to maximize the term $\sum_{j=1}^{m} A'_j - A_j$ on the right-hand side (RHS). We split the term up into two parts $\sum_{j=1}^{m} A'_j - A_j = \sum_{j=2}^{m} A'_j - A_j + A'_1 - A_1$. $A'_1 - A_1$ can be at most $\frac{1}{7}D_g$ according to our assumptions. And $\sum_{j=1}^{m} A'_j - A_j$ at most $\frac{m-1}{7m}D_g$. With this second term we go even further and simplify $\frac{m-1}{7m}D_g$ to $\frac{1}{7}D_g$ which is even bigger. So, we get $\sum_{j=1}^{m} A'_j - A_j > \frac{2}{7}D_g$.

We plug this result into our inequation and also split up $\sum_{j=1}^{m} g_{i,j} A_j$ into $\sum_{j=2}^{m} g_{i,j} A_j + g_{i,1} A_1$ on the left-hand side (LHS).

$\sum_{j=2}^{m} g_{i,j} (A'_j - A_j) + g'_{i,1} A'_1 - g_{i,1} A_1 > S_i \frac{2}{7} D_g$

Again, we look at some terms on the RHS in isolation, where we try to minimize the RHS.

$A'_j - A_j$ has to be at least $-\frac{1}{7m}D_g$.

On the other term we have $g'_{i,1} A'_1 - g_{i,1} A_1 = (g_{i,1} + D_g)(A_1 + D_{a1}) - g_{i,1} A_1$

$= g_{i,1} A_1 + g_{i,1} D_{a1} + D_g A_1 + D_g D_{a1} - g_{i,1} A_1 = (g_{i,1} + D_g) D_{a1} + D_g A_1$. D_{a1} is minimal with $D_{a1} = -\frac{1}{7}D_g$. Therefore, we maximize $g_{i,1}$ with $g_{i,1} = 1$ in order to minimize the whole term $(g_{i,1} + D_g) D_{a1}$ and we get

$g'_{i,1} A'_1 - g_{i,1} A_1 \geq (-1)(1 + D_g) \frac{1}{7} D_g + D_g A_1$.

We plug our new found results into our inequation.

$\sum_{j=2}^{m} g_{i,j} (-\frac{1}{7m} D_g) - (1 + D_g) \frac{1}{7} D_g + D_g A_1 > S_i \frac{2}{7} D_g$

Next, both sides get multiplied by $\frac{7}{D_g}$.

$-\frac{\sum_{j=2}^{m} g_{i,j}}{m} - (1 + D_g) + 7A_i > 2S_i$

The RHS is maximized if $S_i = 1$, the LHS is minimized with the individual accuracies being 1, $D_g = 1$ and $A_i = \frac{6}{7}$.

$-\frac{m-1}{m} - (1+1) + 7 * \frac{6}{7} > 2$

We simplify $-\frac{m-1}{m}$ to -1 which is even smaller and we arrive at

$-1 - 2 + 6 = 3 > 2$

which is of course always true. □

Remark 4.50. We would achieve equality in the proof above if the accuracy were to be in the domain of $[\frac{5}{6}, 1]$.

Chapter 5

Approach and Model

5.1 Accuracy and Score

In PeerRank [1] the grade X_i^n of agent i determines how influential agent i is for the grades of i's peers. Such an approach assumes that agents knowledgeable enough to solve the given task well also are knowlegeable enough to assess their peers accurately. The idea in our work is however that agents with good grades can also be bad at grading. Therefore, the final grade is split up into two parts, the score and the accuracy. The score reflects a consensus grade that others give to an agent. The accuracy indicates the grading ability of an agent, which is similar to the incentive in other works [3, 4]. At the end the score and the accuracy are aggregated into a final grade, the total grade, which can be reported back to the agents, e.g., in form of a school report for pupils. Since the accuracy is part of the total grade it is in the agents interest to predict the score of the peers they grade correctly.

5.2 About truth

Remark 5.1 (Grading Scheme). An agent wants to achieve a high total grade. Therefore, the agent wants its score and its accuracy to be as high as possible.

Definition 5.2 (Ground Truth)**.** The ground truth (GT) is the score of an agent, if all of its peers grade it truthfully.

Assumption 5.3 (Untruthfulness assumption)**.** If an agent grades untruthfully, the given grade is *further* from the GT than if the agent was honest.

Definition 5.4 (Truthfulness)**.** If an agent is not untruthful, it is truthful.

Definition 5.5 (Accuracy Monotonicity)**.** The closer an individual grade $g_{j,i}$ is to the fixed point $F_{j,i}$ of the receiving agent j, the better is the accuracy of the assessing agent i.

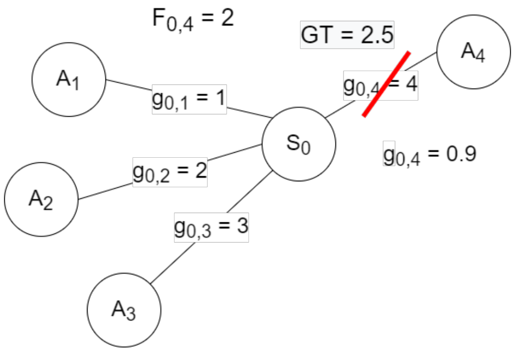

Figure 5.1: Example

Definition 5.6 (Fixed point)**.** The fixed point (FP) $F_{j,i}$ is the score agent j would receive without the individual grade of agent i, $g_{j,i}$. Thus, if $g_{j,i}$ was $F_{j,i}$ the score of agent j wouldn't change and be $F_{j,i}$.

Example 5.7. The agent $1, 2, 3$ and 4 grade their peer 0. Agents 1 to 3 give the following individual grades: $g_{0,1} = 1, g_{0,2} = 2$ and $g_{0,3} = 3$. This gives us a fixed point of $F_{0,4} = 2$ if the scoring rule is either the average or the median. If agent 4 were the only agent grading 0 it would choose $g_{0,4} = 4$. But if agent 4 gives $g_{0,4} = 0.9$ instead, it will receive a higher accuracy, even though 0.9 is further from the GT of 2.5 than 4. Thus agent 4 grades untruthfully. But agent 4 also benefits from grading untruthfully since they receive a higher accuracy.

However, we also have to consider what information the agents have when they are grading their peers. They only know the work their peer did. Nothing else. No grades of other agents, no fixed points, no ground truths.

Let us now assume a grading scale from 0 to 10. An agent can upper bound the error to the FP by 5 if he always gives a 5 no matter the work. With minimal effort the agent could already say if the work presented is bad, good or something in between. If it is terrible he can expect that the FP will be 5 or below, and thus he can upper bound his error to 2.5 by giving bad work the grade 2.5. If the work is great he can expect a FP of 5 or above. Thus, by giving a 7.5 he can upper bound the error to 2.5. If the agent has trouble grading the work or it is somewhere in between he can give a 5 and again upper bounding the error to 5.

We try a new definition for GT and a new untruthfulness assumption.

Definition 5.8 (Grount truth)**.** The GT is the score of an agent, if all of its peers grade it, as if they were the only ones to grade the agent.

One can look at the GT as the score of an agent that gets graded by its peers, as if all of the peers thought themselves as dictators.

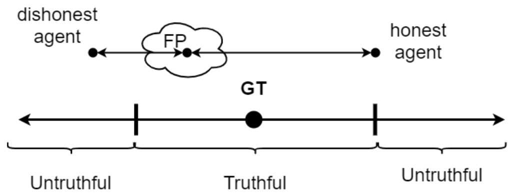

Figure 5.2: Untruthfulness Domain

Definition 5.9 (Untruthfulness)**.** If an agent grades untruthfully, the given grade is *further* from the GT than the given grade were as if the agent were the only one to grade its peer.

The rest we leave as we had it, namely accuracy monotonicity and fixed points. This gives us a domain where the agent can be untruthful, and also a domain where to agent grades truthfully.

5.3 Fixed Points

Definition 5.10 (Fixed point)**.** The fixed point (FP) $F_{j,i}$ is the score agent j would receive without the individual grade of agent i, $g_{j,i}$. Thus, if $g_{j,i}$ was $F_{j,i}$ the score of agent j wouldn't change and be $F_{j,i}$.

For a one-dimensional problem, at the beginning nothing is known except the individual grades. Therefore, the accuracies can be assumed to be all equal. The Fixed Points can calculated. This leads to new accuracies which then leads to new Fixed points, and so forth [*Figure 5.3*].

For a multi-dimensional problem not much changes, the different dimensions can be looked as simply another agent even though the actually belong to the same agent. The only difference will be that the accuracy of an agent will probably be multi-dimensional, as well as the score (and the weights in a weighted average).

Example 5.11. We look a situation where the individual grades are two-dimensional and the agents $1, 2$ and 3 grade agent 0. We assume the three individual grades form an equilateral triangle. Here, the fixed point of agent i is always between $g_{0,k}$ and $g_{0,l}$, $i, k, l \in \{1, 2, 3\}, i \neq k \neq l$. Thus the distance between the fixed point and the given individual grade is the same for all agents, and thus also the accuracy. Therefore, the score of agent 0 is always right in the middle of the triangle.

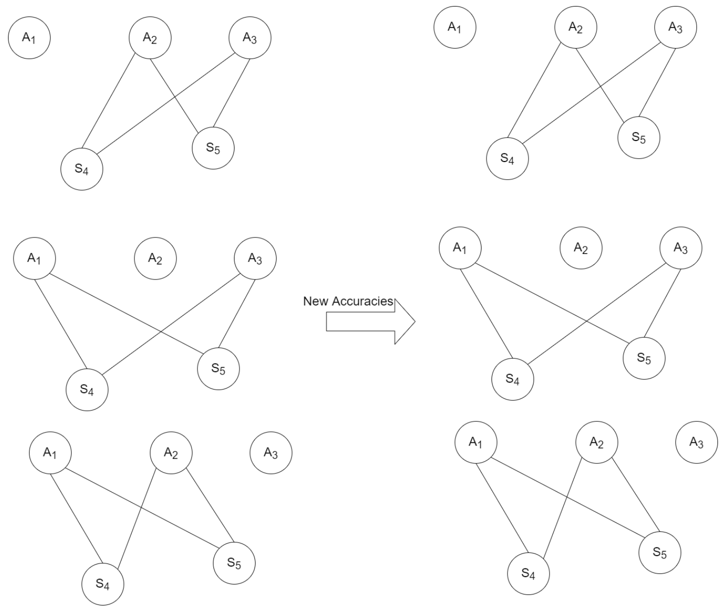

Figure 5.3: Calculating the Accuracy after each iteration

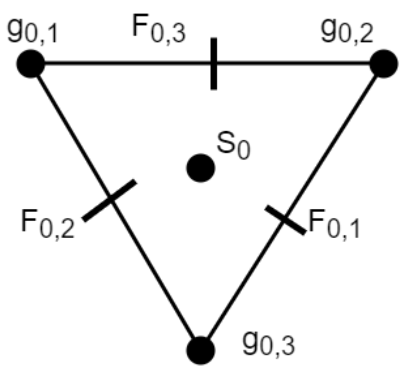

Figure 5.4: Triangle

5.4 Variance

Does the variance of the partial accuracies an agent receives matter?

Example 5.12. Let's look at two agents, one which is extremely accurate, except in a few cases where they are extremely inaccurate. The other agent we look at always has an accuracy that is a bit better than if they graded at random. From this we can see that agent 1 mostly spot on with hitting the fixed point except in a few cases. But if we trust agent 1 too much, when they are very inaccurate they perturb the score by alot. On the other hand, agent 2 never has very good accuracy, but the direction they give is always correct. Both are overall equally important for the score, thus the variance doesn't really matter for us.

The discussion on the influence of variance can be expanded in future works.

5.5 Model

We make some assumptions on the model of this work.

1. *Memorylessness:* The total grade is not dependant on previous assignements.
2. *Many-to-many:* Only a subset of agents should have to grade an agent.
3. *No spot-checks:* The grading scheme requires *no spot-checks*, e.i., no instructor/TA grading is required.

CHAPTER 6

Accuracy

The goal of this chapter is to find a family of functions that calculate the accuracy of agent i from the distances between the individual grades agent i gave and their fixed points.

6.1 Accuracy Monotonicity

Definition 6.1 (Accuracy Monotonicity)**.** If agent i grades agent j 'better' and the other agents grade agent j the same way, agent i's accuracy should increase. 'Better' means that i's individual grade is closer to the fixed point.

The point of accuracy monotonicity is so that more accurate grading actually gets rewarded accordingly.

6.2 Negative Accuracies

Being able to have negative accuracy would have two main impacts. First, a negative accuracy would lead to a handicap in the total grade for the agent with the negative accuracy. Second, score of the agents that get graded by a peer with negative accuracy would get pushed further from the individual grade given instead of pulled towards it.

The accuracy should be the highest when we have a distance of 0. The accuracy should be zero, if the distance is as if the individual grade was given at random. A negative accuracy would then mean that the distance is worse than at random. If one grades worse than at random they seem to intentionally try to skew the rating. Thus, we give them a negative rating.

One issue would be, how we can see that an agent grades at random.

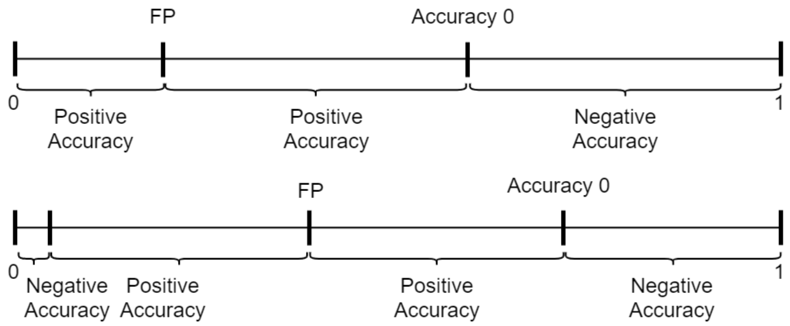

Figure 6.1: Different FPs lead to different domains of positive/negative accuracies

6.2.1 At random

What does it mean when we say at 'at random'?

Definition 6.2 (At random)**.** At random is in expectation be half of distance between a given fixed point and its furthest grade in the grading domain.

This definition means that the distance where the Accuracy is 0 can be dependent on where the fixed point is [*Figure 6.1*].

6.2.2 Integration rule

Our goal is to have a rule such that when an agent grades uniformly at random they get an accuracy of 0 in expectation in the case where negative accuracies exist. We achieve this by saying that the integral of the accuracy over the distance between the fixed point and the individual grade has to be 0. With this rule the expectation of the accuracy is 0 if picked uniformly at random.

Definition 6.3 (Integration rule)**.** The integral of the accuracy over the domain of the distance is 0.

6.3 Accuracy Aggregation

The agents receive accuracies from giving individual grades to their peers. These accuracies need to be aggregated to give an assessment on how accurate the agents grade overall. These total accuracies are used to find the fixed point in the next iteration and are part of the total grade.

But how are the accuracies aggregated? Multiplication of accuracies won't work if we want negative accuracies to be possible, since the multiplication of two negative accuracies would be the same if they both were positive. However, the summation of accuracies could work.

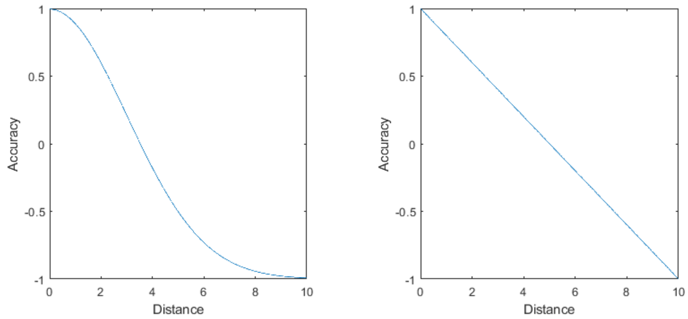

Figure 6.2: Some examples for valid accuracy functions with negative accuracies

Since we the order of the accuracies shouldn't matter, due to anonimity, the aggregation has to treat all partial accuracies the same. There are two new things we could want from the accuracy. First, if all partial accuracies are the same, then the aggregated accuracy should also have this value, which we could call accuracy unanimity. Second, that if one partial accuracy increases and the other partial accuracies stay the same, then the aggregated accuracy should also increase. This two rules would pretty much force the aggregated accuracy to be the average of the partial accuracies. Another advantage is that this rule would also work for negative partial accuracies.

Another idea could be that accuracies are not aggregated at all. This would mean that the accuracies are more specific for a given agent.

6.4 Accuracy Function

With this the accuracy as a function of the distance is quite narrowed down. The Integration rule and the Accuracy Monotonicity are quite strong in itself [*Figure 6.2*].

CHAPTER 7

Putting everything together

7.1 Properties that can be fulfilled

It is possible to construct a cardinal peer grading mechanism that meets all properties listed in chapter 3 *(Preliminaries)*.

Independence is special property, since it is equivalent with a bipartite graph. Therefore, it can be automatically fulfilled if and only if a model with two groups of agents that grade each other, but not among themselves, is taken.

7.2 Relationship

What is the relationship between scores and accuracies? And what is the relation of the accuracy of an agent to the accuracies of other agents?

Should the score be closest to the agent with the highest accuracy?

Example 7.1. Let us assume the agent $1, 2, 3$ and 4 grade the agent 0 on a scale of 0 to 10. The score is calculated simply with a weighted average, where the weights are the accuracies. The individual grades are $g_{0,1} = 10, g_{0,2} = 1, g_{0,3} = 2$ and $g_{0,4} = 3$. The accuracies of the grading agents are $A_1 = 0.9, A_2 = 0.8, A_1 = 0.3$ and $A_1 = 0.1$. Therefore, the score is $S_0 = \frac{\Sigma_{i=1}^{n} g_{0,i} * t(A_i)}{\Sigma_{i=1}^{n} t(A_i)} = \frac{10*0.9+1*0.8+2*0.3+3*0.1}{0.9+0.8+0.3+0.1} = 5.0952$. The score is not closest to the individual grade with the highest accuracy, but actually closest to the one with the *lowest* accuracy in this example [*Figure 7.1*].

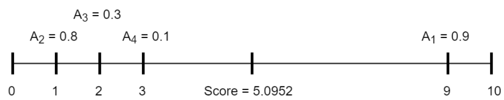

Figure 7.1: The agent with the highest accuracy is not the closest

7.3 Accuracy Example

One of the cases that would fulfill all the properties is the example used in the proof for score monotonicity.

$$A_i = \frac{\sum_{j=1}^{m} dist_{max} - |g_{j,i} - F_{j,i}|}{7m} + \frac{6}{7}$$

7.4 Score Example

Again, a scoring rule that would fulfill all properties and even be score monotonic is the example used in the proof for score monotonicity.

$S_j = \frac{\Sigma_{i=1}^{n} g_{j,i} * A_i}{\Sigma_{i=1}^{n} A_i}$, where n is the number of grades that an agent receives.

7.5 Total Grade Example

At the very end, the total grade can be calculated. The total grade should be both depending on the score and the accuracy, where a higher score or accuracy respectively increase the total grade. Otherwise, the instructor or teacher is quite free in how to calculate the total grade. But it should be stated that the total grade has to stay in its own domain.

The most intuitiv formula would be $G_j^n = (1 - \alpha) * S_j^n + \alpha * A_j^n$, where α is how much of the total grade should be coming from the accuracy.

Chapter 8

Related Work

Most related work on peer grading is either in the context of course evaluation for students or the selction of papers for scientific conferences. These problems are mostly tackled in three ways. Cardinal Peer Grading, where each grader give feedback in form of a number, Ordinal Peer Grading, where the peers are put into a ranking of the received work, and Peer Selection where the top k peers get selected. Most work is either for the selection problem [8, 9, 5, 10] or on ordinal peer grading [11, 12, 13, 14] while there is less literature on cardinal peer grading [1, 15, 16].

All the papers on selection mentioned solve the problem in a strategyproof way. According to Luca de Alfaro et al [16, 3] ordinal peer grading is less popular with students than cardinal peer grading. While the mentioned work on cardinal peer grading says nothing about strategyproofness, they all use spot-checks and incentives for the students to grade honestly. There are multiple approaches on how to construct an incentives for the students as can be seen in [3] and [4].

Toby Walsh's PeerRank Method [1] stands out from the crowd by being one of the few mechanisms for cardinal grading and with its elegant solution for achieving remarkable results. However, according to the author PeerRank is not score monotonic. The PeerRank Method has been assessed in other works and even some changes have been proposed and tested [17, 18].

Chapter 9

Conclusion

In this work properties for cardinal peer grading were examined. The grades are split into two parts, accuracy and score. Accuracy measures how good an agent is at grading its peers and the score aggregates the received individual grades. This works further includes a proof on the equality of independence and bipartite graphs, and a proof on the feasibility of score monotonicity. Additionally, a number of approaches and models for cardinal peer grading are examined. We found a family of functions for the aggregation of accuracies that fulfill the properties for the accuracy.

Bibliography

[1] T. Walsh, "The peerrank method for peer assessment," *CoRR*, vol. abs/1405.7192, 2014. [Online]. Available: http://arxiv.org/abs/1405.7192

[2] V. Todd and J. C. Hudson, "Using graded peer evaluation to improve students writing skills, critical thinking ability, and comprehension of material in a principles of public relations course," *Journal of College Teaching & Learning (TLC)*, vol. 4, no. 10, 2007.

[3] L. De Alfaro, M. Shavlovsky, and V. Polychronopoulos, "Incentives for truthful peer grading," *arXiv preprint arXiv:1604.03178*, 2016.

[4] G. Radanovic, B. Faltings, and R. Jurca, "Incentives for effort in crowdsourcing using the peer truth serum," *ACM Transactions on Intelligent Systems and Technology (TIST)*, vol. 7, no. 4, pp. 1–28, 2016.

[5] H. Aziz, O. Lev, N. Mattei, J. S. Rosenschein, and T. Walsh, "Strategyproof peer selection using randomization, partitioning, and apportionment," *Artificial Intelligence*, vol. 275, pp. 295 – 309, 2019. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0004370219301523

[6] A. S. Asratian, T. M. J. Denley, and R. Häggkvist, "Bipartite graphs and their applications," 1998.

[7] M. Vault, "The definitive glossary of higher mathematical jargon," https://mathvault.ca/math-glossary/#onto, accessed: 2021-01-19.

[8] N. Alon, F. Fischer, A. Procaccia, and M. Tennenholtz, "Sum of us: Strategyproof selection from the selectors," in *Proceedings of the 13th Conference on Theoretical Aspects of Rationality and Knowledge*, ser. TARK XIII. New York, NY, USA: Association for Computing Machinery, 2011, p. 101–110. [Online]. Available: https://doi.org/10.1145/2000378.2000390

[9] H. Aziz, O. Lev, N. Mattei, J. Rosenschein, and T. Walsh, "Strategyproof peer selection: Mechanisms, analyses, and experiments," *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 30, no. 1, Feb. 2016. [Online]. Available: https://ojs.aaai.org/index.php/AAAI/article/view/10038

[10] Y. Xu, H. Zhao, X. Shi, J. Zhang, and N. B. Shah, "On strategyproof conference peer review," 2020.

[11] D. Berga and R. Gjorgjiev, "Impartial social rankings," *Manuscript*, p. 8, 2014.

[12] K. Raman and T. Joachims, "Methods for ordinal peer grading," in *Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining*, 2014, pp. 1037–1046.

[13] A. E. Waters, D. Tinapple, and R. G. Baraniuk, "Bayesrank: A bayesian approach to ranked peer grading," in *Proceedings of the Second (2015) ACM Conference on Learning@ Scale*, 2015, pp. 177–183.

[14] K. Raman and T. Joachims, "Bayesian ordinal peer grading," in *Proceedings of the second (2015) acm conference on learning@ scale*, 2015, pp. 149–156.

[15] J. R. Wright, C. Thornton, and K. Leyton-Brown, "Mechanical ta: Partially automated high-stakes peer grading," in *Proceedings of the 46th ACM Technical Symposium on Computer Science Education*, 2015, pp. 96–101.

[16] L. De Alfaro and M. Shavlovsky, "Crowdgrader: A tool for crowdsourcing the evaluation of homework assignments," in *Proceedings of the 45th ACM technical symposium on Computer science education*, 2014, pp. 415–420.

[17] M. CIKLA and E. BEKTAS, "Evaluation of the generalized peerrank method for peer assessment," 2015.

[18] J. Kline, "Analysis of the peerrank method for peer grading," 2015.