1.1)

The problem is the following: find k complaint officers that maximise the set of complaint officers' friends. Thus, we can see it like a maximum coverage problem. To solve it we can apply a Greedy approach, which tries to maximise at each step the number of elves with a complaint officer as a friend. Let's call F_e the list of friends for each elf $e \in W$, and C the set of complaint officers selected.

- At each step of the *while,* we select the elf e', not already in C, who is one of the elves (in the set E') with the highest number of friends not yet "covered". We used an intuitive function *first* to retrieve the first of the set, but any of the elves in E' would have been fine. Then we add e' to C. Of course, the first e' selected will be one of the elves who have the longest list of friends.
- At the end the algorithm returns the k elves that we selected as complaint officers, which is the set C.

$C = \emptyset$
$i = 0$
$while\ i < k:$
    $E' = argmax_{e \in \{W-C\}}\ (Covered(C \cup \{e\}) - Covered(C))$
    $e' = first(E')$
    $C = C \cup \{e'\}$
$return\ C$

$Covered(S):$

    $return\ |\cup_{e \in S}\ F_e|$

1.2)

Notation: Cov_i is the number of "covered elves" (elves who have a complaint officer as a friend up to this point) after iteration i; OPT is the maximum possible (w.r.t W) number of elves that can be covered with k complaint officers, therefore the optimal solution; r_i is the number of "remaining elves w.r.t. OPT" (which means $r_i = OPT - Cov_i$) after iteration i; *SOL* is the returned value of our greedy algorithm, which corresponds to Cov_k. The first thing to observe is that the number of elves that the algorithm adds to covered elves at iteration $i+1$ must be greater than or equal to $\frac{r_i}{k}$. This is an application of the pigeonhole principle, indeed, at each iteration there are some F_e which are greater than or equal to $\frac{r_i}{k}$, to ensure that the optimal solution can cover OPT elves in k iterations. Our greedy algorithm, by picking always the elf which brings to the biggest increment of covered elves, confirms the observation. Thanks to this first observation we obtain that:

$$r_{i+1} \le r_i - \frac{r_i}{k} = r_i * (1 - \frac{1}{k})$$

If we substitute with $k = i + 1$:

$$r_k \le r_{k-1} * \left(1 - \frac{1}{k}\right) \le r_{k-2} * \left(1 - \frac{1}{k}\right)^2 \le r_{k-3} * \left(1 - \frac{1}{k}\right)^3 \le \cdots$$

Continuing until the k^{th} iteration, we obtain that:

$$r_k \le r_0 * \left(1 - \frac{1}{k}\right)^k$$

We know that from the definition of r_i that $r_0 = OPT$, because $Cov_0 = 0$, then:

$$r_k \le OPT * \left(1 - \frac{1}{k}\right)^k$$

If we consider big values of k, we obtain:

$$\lim_{k \to inf} \left(1 - \frac{1}{k}\right)^k = \frac{1}{e} \quad \rightarrow \quad r_k \le OPT * \frac{1}{e}$$

By definition, the number of remaining elves w.r.t OPT after iteration k is $r_k = OPT - SOL$, therefore:

$$OPT - SOL \le OPT * \frac{1}{e}$$

$$SOL \ge \left(1 - \frac{1}{e}\right) * OPT$$

Considering that $OPT \ge SOL$ and $\left(1 - \frac{1}{e}\right) < 1$, the algorithm is a $\left(1 - \frac{1}{e}\right)$-approximation.

2.a)

Since the goal is to choose a path P_{ij} for every item g_i such that the maximum number of paths through a segment $e \in E$ is as small as possible, we can formulate the ILP problem with the following objective function:

$$min\left\{ max_{e\in E}\left\{\sum_{P_{ij}} x_{ij} \mid e \in P_{ij}\right\}\right\}$$

Of course we have to put some constraints to respect the problem specifications:

$$x_{i1} + x_{i2} + x_{i3} = 1 \quad i \in \{1..n\}$$

$$x_{ij} \in \{0,1\} \qquad i \in \{1..n\} \;\; j \in \{1..3\}$$

$x_{ij} = 1$ indicates that the path P_{ij} is chosen and 0 otherwise; i is used to iterate over the n items, while j is used to iterate over the three companies. To relax the ILP as a fractional LP problem, we changed the possible range of values of the x_{ij} variables:

$$x_{ij} \in \left[0, \frac{1}{3} - \varepsilon\right] \cup \left[\frac{1}{2} + \varepsilon, 1\right]$$

Where $\varepsilon > 0$ and infinitesimally small. The rounding is explained in the section 2.b.

2.b)

The rounding algorithm is the following: consider a path P_{ij} as "chosen" if $x_{ij} > \frac{1}{2}$. Let's call C the set of chosen paths:

$$C = \left\{P_{ij} \mid x_{ij} > \frac{1}{2}\right\}$$

Now we have to show that with this rounding we can obtain a 3-approximation of the optimal solution. The first thing to notice is that the optimal solution of the ILP would correspond to a set C' of chosen paths, which makes as small as possible the max number of chosen paths passing through a road segment (considering integer x_{ij} values). In the LP instead, we minimise over more possible values of x_{ij} (according to the fractional range written above).

Therefore, the solution of the LP must be lesser than or equal to the optimal integer solution:

$$maxSum_{LP} \leq maxSum_{OPT}$$

Then we must show that with this rounding we obtain solutions for the problem. We can observe that if an x_{ij} is greater than $\frac{1}{2}$, then the $x_{ij'}$ with $j' \neq j$, can't be greater than $\frac{1}{2}$ because this will violate the first constraint. This observation guarantees that only one truck can be chosen for every item i. Moreover, another similar observation is that, at least one truck for each item i MUST be chosen, because if all the three variables ($x_{i1},\ x_{i2},\ x_{i3}$) are lesser than $\frac{1}{2} + \varepsilon$, the first constraint is violated.

Now we can proceed with some intuitive steps:

$$maxSum_{LP} = max_{e\in E}\ \left\{\sum_{P_{ij}} x_{ij} \mid e \in P_{ij}\right\} \geq max_{e\in E}\ \left\{\ \sum_{P_{ij}} x_{ij}\ \mid e \in P_{ij}\ ,\ P_{ij} \in C\right\} \geq$$

$$\geq max_{e\in E}\left\{\sum_{P_{ij}}\left(\frac{1}{2} + \varepsilon\right) \mid e \in P_{ij}\ ,\ P_{ij} \in C\right\} = max_{e\in E}\ \left\{\left(\frac{1}{2} + \varepsilon\right) * \sum_{P_{ij}} 1\ \mid\ e \in P_{ij}\ ,\ P_{ij} \in C\right\} =$$

$$= \left(\frac{1}{2} + \varepsilon\right) * max_{e\in E}\left\{\sum_{P_{ij}} 1\ \mid e \in P_{ij}\ ,\ P_{ij} \in C\right\} = \left(\frac{1}{2} + \varepsilon\right) * maxSum_C$$

Where $maxSum_C$ is the solution when considering the paths in $C.$ Finally, we can conclude with:

$$maxSum_C \leq \frac{1}{\left(\frac{1}{2}+\varepsilon\right)} * maxSum_{LP} = \frac{2}{1+2\varepsilon} * maxSum_{LP} \leq \frac{2}{1+2\varepsilon} * maxSum_{OPT}$$

Therefore, the algorithm is a $\left(\frac{2}{1+2\varepsilon}\right)$-approximation. Since $\frac{2}{1+2\varepsilon} < 3$, is also a 3-approximation.

3.a)

To answer we have to reason on the first $n-\sqrt{n}$ steps of the algorithm. From the theory of the original contraption algorithm, we know that at each step the probability to pick an edge from the ones which compose the min cut is lesser than or equal to $2/currentNumberOfNodes$. We can write the probability P, that the question is asking, as the probability that in the first $n-\sqrt{n}$ steps the algorithm doesn't contracts edges from the min cut:

$$P \geq \left(1-\frac{2}{n}\right)*\left(1-\frac{2}{n-1}\right)*\ldots*\left(1-\frac{2}{\sqrt{n}+2}\right)*\left(1-\frac{2}{\sqrt{n}+1}\right)=$$
$$=\left(\frac{n-2}{n}\right)*\left(\frac{n-3}{n-1}\right)*\ldots*\left(\frac{\sqrt{n}}{\sqrt{n}+2}\right)*\left(\frac{\sqrt{n}-1}{\sqrt{n}+1}\right)=$$

Considering the products at numerator and denominator in a factorial form, we can follow with:

$$=\frac{\frac{(n-2)!}{(\sqrt{n}-2)!}}{\frac{n!}{(\sqrt{n})!}}=\frac{(n-2)!}{(\sqrt{n}-2)!}*\frac{(\sqrt{n})!}{n!}=\frac{\sqrt{n}\,*\,(\sqrt{n}-1)}{n\,*\,(n-1)}\geq\frac{n-\sqrt{n}}{n^2}$$

3.b)

In this case, we can see the probability that the question is asking as $P=(1-P^{fail})$; where P^{fail} is the probability that the algorithm fails in returning the correct min cut set. Let's call P^{first} the probability that the algorithm doesn't contract edges from the min cut in the first $n-\sqrt{n}$ steps (which is the one computed in 3.a). And let's call $P^{copiesFail}$ the probability that all the $\sqrt{n}$ copies of the graph G_k would fail. Thus, we can proceed:

$$P^{fail}=\left(P^{first}*P^{copiesFail}\right)+\left(1-P^{first}\right)$$

$$P^{copiesFail}=\left(P^{individualCopyFails}\right)^{\sqrt{n}}$$

$$P^{individualCopyFails}=1-P^{individualCopyCorrect}$$

$$P^{individualCopyCorrect}\geq\left(1-\frac{2}{\sqrt{n}}\right)*\left(1-\frac{2}{\sqrt{n}-1}\right)*\ldots*\left(1-\frac{2}{4}\right)*\left(1-\frac{2}{3}\right)=$$

$$=\left(\frac{\sqrt{n}-2}{\sqrt{n}}\right)*\left(\frac{\sqrt{n}-3}{\sqrt{n}-1}\right)*\ldots*\left(\frac{2}{4}\right)*\left(\frac{1}{3}\right)=\frac{(\sqrt{n}-2)!}{\left(\frac{(\sqrt{n})!}{2}\right)}=(\sqrt{n}-2)!*\frac{2}{(\sqrt{n})!}=\frac{2}{\sqrt{n}*(\sqrt{n}-1)}\geq\frac{2}{n}$$

$$\textit{Thanks to the HINT, we can say that: }\left(1-\frac{2}{n}\right)^{\sqrt{n}}\leq 1-\frac{2\sqrt{n}}{n}\textit{. Then:}$$

$$P^{fail}\leq\left(P^{first}*\left(1-\frac{2}{n}\right)^{\sqrt{n}}\right)+\left(1-P^{first}\right)\leq\left(P^{first}*\left(1-\frac{2\sqrt{n}}{n}\right)\right)+\left(1-P^{first}\right)=$$

$$=P^{first}\left(1-\frac{2\sqrt{n}}{n}-1\right)+1=\left(-P^{first}*\frac{2\sqrt{n}}{n}\right)+1\leq\left(-\frac{n-\sqrt{n}}{n^2}*\frac{2\sqrt{n}}{n}\right)+1=1-\left(\frac{2n\sqrt{n}-2n}{n^3}\right)=1-\left(\frac{2\sqrt{n}-2}{n^2}\right)$$

Finally:

$$P\geq 1-\left(1-\left(\frac{2\sqrt{n}-2}{n^2}\right)\right)=\frac{2\sqrt{n}-2}{n^2}$$

3.c)

About the number of contractions (with $n>2$):

Modified Alg: $(n-\sqrt{n})+\sqrt{n}*(\sqrt{n}-2)=2n-3\sqrt{n}$

Original Alg twice: $2*(n-2)=2n-4$

The two are of the same order of magnitude, but specifically the original algorithm twice does more contractions.

About the error probability (with $n>2$):

Modified Alg: $P^{fail}\leq 1-\left(\frac{2\sqrt{n}-2}{n^2}\right)$ (computed before)

Original Alg Twice: $P^{fail}\leq\left(1-\frac{2}{n^2}\right)^2$ (both runs fail)

With $n\geq 9$, the modified algorithm has a smaller upper bound on the error probability then the one of the original algorithm twice. With $2<n<9$ is the opposite.

4.a)

Our set of s_{ij}'s, that satisfies both the assumption and the point 4.a request, can be represented as a $k \times n$ matrix M in which an element in position (i, j) is the cost s_{ij} for the reviewer i to write a review for paper j:

$$M = \begin{pmatrix} \gamma & T-\alpha & T+\beta & T+\beta & \cdots & T+\beta \\ \gamma & T+\beta & T-\alpha & T+\beta & \cdots & T+\beta \\ \gamma & T+\beta & T+\beta & T-\alpha & \cdots & T+\beta \\ \gamma & T+\beta & T+\beta & T+\beta & \cdots & T+\beta \\ \gamma & T+\beta & T+\beta & T+\beta & \cdots & T-\alpha \\ \vdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ \vdots & \cdots & \cdots & \cdots & \cdots & \cdots \end{pmatrix} \qquad M' = \begin{pmatrix} 0.1 & 0.9 & 0.96 & 0.96 \\ 0.1 & 0.96 & 0.9 & 0.96 \\ 0.1 & 0.96 & 0.96 & 0.9 \\ 0.1 & 0.9 & 0.96 & 0.96 \end{pmatrix}$$

1) General case

2) Example with $n = k = 4, \gamma = 0.1, T = 0.95, \alpha = 0.05, \beta = 0.1$

Where all rows must have exactly one element $T - \alpha$, while the columns (except the first one) must have at least one element $T - \alpha$. We also need the following constraints:

I. $k \geq n \geq 1$

II. $\frac{1}{k} - \gamma > 1 - (T - \alpha)$

III. $\gamma + (T - \alpha) > T \;\; \rightarrow \;\; \gamma > \alpha$

IV. $\alpha,\ \beta,\ T > 0$

V. $0 < \gamma < \frac{1}{k}$

A combination of strategies which satisfies the assumption is the one where: $(n-1)$ reviewers will review only the paper which costs $T - \alpha$ for them, while the remaining reviewer will review only the paper which costs γ. The constraint V is the one to guarantee that the reviewer who takes the paper which cost γ has positive utility.

About the PNE scenario, let's analyze the best choice to take in the case of a reviewer i while considering that the other reviewers are all choosing to review the paper which costs γ. The constraint III and the constraint on the T given in the text ($\sum_{j \in S_i} s_{ij} \leq T$), force the reviewer i to two possibilities: the one in which he reviews only the paper of cost γ (together with all the other reviewers), and the one in which he reviews only the paper of cost $T - \alpha$ (alone). Thanks to the II constraint, the reviewer i best choice will be always to review only the paper of cost γ, because his utility in that case would be the highest. We can conclude by saying that at PNE, the only reviewed paper is the one with cost γ, thus the fraction of reviewed papers is $1/n$.

A final observation is that this instance is not "strict", in the sense that it could work even when γ, α *and* β are specific for each reviewer, of course after the generalization of the constraints.

4.b)

In this case we will show an instance with $3n + 1$ papers, to get a fraction of reviewed papers at PNE close to n. This time, our set of s_{ij}'s, that satisfies both the assumption and the point 4.b request, can be represented as a $k \times (3n + 1)$ matrix M, in which an element in position (i, j) is the cost s_{ij} for the reviewer i to write a review for paper j, with some constraints:

$$M = \begin{pmatrix} \gamma & \cdots & \gamma & 2+\beta & \cdots & 2+\beta \\ \gamma & \cdots & \gamma & 2+\beta & \cdots & 2+\beta \\ \gamma & \cdots & \gamma & 2+\beta & \cdots & 2+\beta \\ \gamma & \cdots & \gamma & 2+\beta & \cdots & 2+\beta \\ \gamma & \cdots & \gamma & 2+\beta & \cdots & 2+\beta \\ \vdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ \vdots & \cdots & \cdots & \cdots & \cdots & \cdots \end{pmatrix}$$

$$\underbrace{\qquad\qquad}_{n} \qquad \underbrace{\qquad\qquad\qquad}_{2n+1}$$

I. $k, n, \beta > 0$

II. $\left[n * \left(\frac{2}{k} - \gamma\right)\right] + \left[\frac{2n+1}{k} * (2 - (2 + \beta))\right] > 0 \rightarrow n * \left(\frac{2}{k} - \gamma\right) > \frac{2n+1}{k} * \beta$

III. $T \geq n * \left(\frac{2}{k} - \gamma\right)$

IV. $0 < \gamma < \frac{2}{k}$

The assumption is satisfied when the strategies are chosen in this way: all the reviewers will review the n papers with cost γ, while the $2n + 1$ remaining papers are divided among all the reviewers (in conclusion each reviewer will review $n + \frac{2n+1}{k}$ papers).

Thanks to II and IV constraints, the utility of each reviewer will be positive, and thus the assumption is respected. The III constraint allow us to take a T which is bigger enough.

For PNE, let's see which the best strategy for a reviewer i is while considering that the other reviewers are all choosing to review all the papers which costs γ. To maximize his utility, the reviewer i will review all the papers which cost γ (because each of them would increase his utility by $\frac{2}{k} - \gamma$) and none of the papers which cost $2 + \beta$ (because each of them would reduce his utility by $2 + \beta$). Therefore, the papers that receive a review, at PNE, are close to n.

The final observation did at the end of ex 4.a, holds even in this case.