**STANDARDFORM OUTPUT OR EXPORT[[1]](#footnote-1)**

|  |  |
| --- | --- |
| Name *[name user]* | Thomas Emery |
| Telephone number *[telephone number where the user can be reached when there are questions about this output]* | 0644414900 |
| Institution *[name institution]* | Erasmus University Rotterdam |
| Date *[date on which the output is produced]* | 27/11/2023 |
| Project *[contractnumber + name research]* | 9424 |
| This output consists entirely of documentation *[syntax or text file without research results] If Yes, then you don’t need to fill in the rest of the form* | No |
| This output is an adjusted version of a stopped output. If Yes, then please fill in the date of the stopped output. | No |
| Data *[used datasets]* | BURENNETWERKTAB, COLLEGANETWERKTAB, HUISGENOTENNETWERKTAB, FAMILIENETWERKTAB,  KLASGENTOENNETWERKTAB,  INPATAB,  GBAPERSOONTAB |
| Purpose *[description of the purpose of the analysis in this output]* | The analysis uses network embeddings, which are generated with the 5 network files, to see whether they can be used to estimate income, death, or marriage at a specified date (in this instance 2019) |
| Relation with previous output *[description of the relation with previously submitted output]* |  |
| Content of the output  *[description of the output, for example: there will be frequencies or crosstabs or regression analysis[[2]](#footnote-2) in this output]* | In this project we have used a deep learning techniques to extract ‘embeddings’ for each individual in the network. These embeddings effectively describe an individuals network topography. The embeddings themselves are not included in this output. We have just produced a report that evaluates these embeddings. The output consists of a number of reports but they are all structured into the same 5 sections and use the same statistical metrics for evaluation.  The first section (Distance Distributions by Hop) provides graphs which show the cosine similarity of individuals embeddings in the graph under different specifications, firstly by looking at random pairs, then by people within 3 ‘steps/hops’ of an individual, then at 2 and then at 1. This helps identify that the embeddings we have generated are capturing something meaningful by showing that yoru embeddings are more similar to people close to you in you network than an individual drawn at random. These graphs have been edited to remove frequency counts on data points of less than 10. The second section of the report does the same as the first but using a slightly different process.  In the subsequent sections of the report we evaluate the embeddings by seeing whether they can be used to estimate common social outcomes including income and likelihood of marriage. These results are all expressed as R^2 and the N for each model is provided in a table below. None of the models us an N of less than 10. |
| 1 – Does this output contain zeros? | no |
| 2 – Does this output contain values < 10 or is it possible to recalculate values < 10? | no |
| 3 – Is there any group disclosure or dominance in this output? | no |
| 4 – Does this output contain information on 1 institution/company/household/etc? | no |
| 5 – Are there any other points where this output does not meet the [output guidelines](file:///K:\Utilities\Regels_Richtlijnen\Guidelines%20for%20RA%20Output.pdf) ? | no |
| If you have answered “Yes” to at least one of the above 5 questions, please give a short but clear explanation on why there is no risk of disclosure in your opinion. |  |

1. If this form is not fully or correctly filled in, the output will not be checked. [↑](#footnote-ref-1)
2. In case the output contains results from models, please clearly mention the underlying number of observations. [↑](#footnote-ref-2)