# Selective Anti-Phishing Training using Centralities of an Email Trust Network

Paul Workman*

## ABSTRACT

Using Generative AI, attackers automatically produce massive amounts of email traffic, luring users into spoof, maleficent links and potentially gaining access to a company's email accounts. From there, they may hijack employee trust to propagate malware further. Phishing education is more important now than ever. Therefore, we must look at any new way of allocating resources to increase the resistance of networks. This project uses SEIR simulations over an infection network based on email communications to find new ways of selecting employees to train using the centrality values found within the trust network. I find that selecting users who are not central but who may hinder initial phishing expansion significantly decreases the infection rate and total number of infections across a network when attacked. I further discuss potential new directions of related research which may be of use to produce more accurate models and better phishing training allocation.

## KEYWORDS

Network analysis, email-based trust network, phishing education, SEIR simulation, cybersecurity

## 1 INTRODUCTION

With most modern companies having complex and sensitive cyber architecture, the damage caused by cyber-attacks can range from irritating to devastating. 94% of companies in 2023 incurred an email-based security incident within the previous year [5], and 79% of account takeover (ATO) attacks were initiated through phishing, a process of using inauthentic URL addresses to convince users to input secure data. With the current innovation and prevalence of Generative AI, creating automated, convincing spoof emails to acquire users' secrets is becoming more accessible for attackers.

One method of preventing ATOs through employee training is phishing simulation, whereby a trusted group or company sends a safe spoof email to employees. Employees who follow the link receive mandatory training to reduce the likelihood of falling victim to an attack. The efficacy of the selection of candidates is potentially flawed — approximately 67.6% of phishing simulation emails go unopened by their recipients, who may fall victim to more relevant phishing attacks [15]. As 35.1% of those who opened the email clicked the spoof link and, therefore, failed the test, 1-in-3 of those who missed the link still need training but are being neglected. Following an ATO, attackers may use the victim's email account to hijack the trust employees have for each other, increasing the risk of another employee falling victim to the attack.

This project aims to find alternative methods of selecting users to receive phishing education. This method provides better protection against the malicious spread of phishing attacks by modelling the email connections of employees within a company. The email network can then be used to construct a trust network, allowing for a more accurate account of each employee's likelihood of falling victim to a phishing attack. This new model can then apply selective training and complete a SEIR phishing spread simulation to evaluate the impact of an attacker on the company. The selection method can then be evaluated, allowing for a comparison of each method and thereby improving the protection these institutions have against cyber attackers.

## 2 LITERATURE REVIEW

Phishing Education primarily focuses on preventing ATO attacks by training employees to notice malicious links [9]. A 2012 experiment involving multiple groups of 300 members of the United States Military determined that the best method of persistent phishing training was an embedded response immediately when an employee clicked a safely constructed phishing link [4]. The embedded response reduced the fail rate of phishing tests from 47.5% to 24.5% over three months. As Military members may have a higher level of suspicion over emails, it is worth further researching untrained users' susceptibility to phishing attacks. A 2017 book on persistent training techniques for phishing education found that 54% of untrained users fail when tested against a single safe spoof link[14]. While understanding the impact of anti-phishing training is required to accurately model learning across a network, as the project aims to model improvements over some time, studying the decay of employee knowledge around phishing education techniques is also a necessity. A 2009 paper detailing the use of a custom anti-phishing training program named PhishGuru found that users retain the information for at least 28 days [11]. Following interaction with anti-phishing application NoPhish, users were found to retain various levels of knowledge for up to 5 months [2].

The research around measuring social network trust using email databases is more sparse. A 2010 paper discussed converting outgoing and incoming email frequency and time to produce a social trust network [7]. This method relies on two stages to complete the network. The first stage determines the trust between users who have previously corresponded, using the equality and quantity of emails to determine a level of trust. The second stage occurs when the users are unknown to each other, requiring a central server to produce the shortest and most trustworthy path using shared contacts. This method was further applied within a vehicular network, where cars exchange data with each other to improve each understanding of road conditions [8].

SEIR simulations for malware spread across a network are well-utilised within the field. The model presumes a single node can be in one of four states: susceptible (available to be infected), exposed (directly able to become infected), infectious (spreading malware across a network), and recovered (no longer spreading malware, and not susceptible to reinfection) [3]. A 2016 method uses a SEIRS-V model to represent how persistent malware may be across a network [6]. It presumes each node may become susceptible after infection, or they may be "vaccinated" and unable to return to

*pw466@exeter.ac.uk.

the infectious state. This paper used mathematical analysis over networks to determine ratios of re-susceptibility and vaccination where malware is continually allowed to propagate or eventually become extinct. The malware propagation model MalSEIRS was used in 2021 to determine offensive and defensive methods to protect against malware attacks by modelling different characteristics of malware [13]. Following or during a cyber attack, the analysts can alter the parameters of the MalSEIRS model to more accurately represent the current activity, and then determine the best course of action.

# 3 PROJECT DESIGN

## 3.1 Enron Email Corpus Data Set

The project utilises networks deriving from the 2015 Enron Email Corpus. The corpus data is an updated catalogue of Enron's internal emails released by the Federal Energy Regulatory Commission following the company's bankruptcy and subsequent investigation, with all the relevant email addresses and contents. An example entry within the Enron Corpus is shown in A.1. This project uses the KONECT 2017 Enron Email dataset, which represents 1,000,000 emails from 1999-2003 [1, 10, 12]. The dataset is a directed graph containing loops. Each node within the network represents a pseudonymised email address, and each directed edge represents an email from the sender to the receiver. This network is highly suited for this project as it is an accurate and representative network of a company's extensive email communication network. Therefore, this project can derive findings from a genuine company's cyber architecture. As each edge represents an email, rather than whether the sender has corresponded with the receiver, it allows for the construction of a trust network, utilising both inbound and outbound messages. A snippet of the network file is given in A.2.

## 3.2 Network Processing

*3.2.1 Cleaning the Network.* Several preprocessing steps must be completed to produce a feasible network to simulate malware spread over the company's infrastructure. The program collapses all edges with identical start and end nodes into a single edge with a weight vector representing the previous number. This produces a new network where the edges remain directional. To make the data set more wieldy, all nodes with less than 200 outgoing and 200 incoming emails over the four years contained within the dataset (corresponding to one email a week) are removed. This process decreases the data set to 1000 nodes.

As the project aims to examine the propagation of malware within a network, having neighbourhoods of nodes which act as sinks may increase the variability of the experiments, particularly if many starting infected nodes are within such a sink. Therefore, the program removes all nodes not appearing in the largest strongly connected network. This additionally removes any non-connected components of the network. This step reduces the size of the network to 802 nodes. A visualisation of the adjacency matrix of this strongly connected network is shown in A.3.

*3.2.2 Creating the Infection-Rate Network.* This connected email graph is then converted to a trust network utilising the functions created by Dijiang Huang et al. [7]. The equations used are as follows, with a minor adjustment to 2:

$$\gamma_{ij} = \frac{2}{\frac{Lij}{Lji} + \frac{Lji}{Lij}} \cdot \frac{Lij}{Lij + Lji}\beta. \tag{1}$$

$$T_{ij} = [\log_2 (N_{ij} + 2)]^{\gamma_{ij}} - 1 \tag{2}$$

Where $L_{ij}$ is the number of emails sent from $i$ to $j$, $\beta = 1/0.553$, and $N_{ij}$ is the total number of emails sent between $i$ and $j$ (that is $N_{ij} = L_{ij} + L_{ji}$).

The parameters used to model infection rate are $\theta$, each node's susceptibility to phishing attacks, and $\eta$, each node's weight for trust with its contacts. The formula to produce the survival rate when one node sends a phishing link to another is as follows:

$$S_{ij} = \max(1, \min[0, 1 - (\theta_j - T_{ij} \cdot \eta_j)]) \tag{3}$$

Where $S_{ij}$ is the rate of survival of $j$ given an email sent from $i$.

## 3.3 Node Training

To simulate education, upon a node being selected for training, the personal survival rate $\theta$ is increased. The formula represents a decrease of infection by 40% it models the efficacy of the best method of anti-phishing training within the 2012 Military study [4]. The formula to update this value is represented as follows:

$$\theta_i = \max(0.95, 1 - (1 - \theta_i) \cdot 0.40) \tag{4}$$

As each timestep is used to model a month, following five successive months without training, a user's susceptibility score from their training will completely decay to the original, untrained score, modelling the findings within the NoPhish paper [2]. This is achieved by reducing the knowledge each node retains by 14% each month. The following equation is used to model the information retention of nodes that do not undergo a phishing awareness course.

$$\theta_i = \max(0.42, \theta \cdot 0.860) \tag{5}$$

## 3.4 SEIR Malware Spread Simulation

This program is based on a SEIR spread simulation with minor modifications. The classifications for this project are as follows: safe (an email address unknown to the attacker), exposed (an email address known to the attacker), infectious (currently spreading malware to its contacts) and recovered (unsusceptible to phishing attack). When the simulation starts, all but ten randomly chosen nodes will initialise in the safe state; the other ten initialise in an infectious state. A safe node will only become exposed if a node that has previously contacted it, becomes infected. Following this, an infected edge has a 2% likelihood of transmitting a phishing link each timestep. If the node receives a phishing link, the network uses the probability of the node surviving the attack using the given $S_i j$. If the node fails, it becomes infected, exposing all nodes receiving emails from it. Once the node has been infected for four timesteps, it transfers to the recovered state. The simulation stops when no nodes are infectious. Contrary to the standard SEIR simulation. Once a node has become exposed, it remains so for the duration of the program. While this does not functionally change the program, it does allow for the analysis of information on the structure of the email-based communication network a user now has based on the contacts found on infected machines.

## 4 EXPERIMENT STRUCTURE

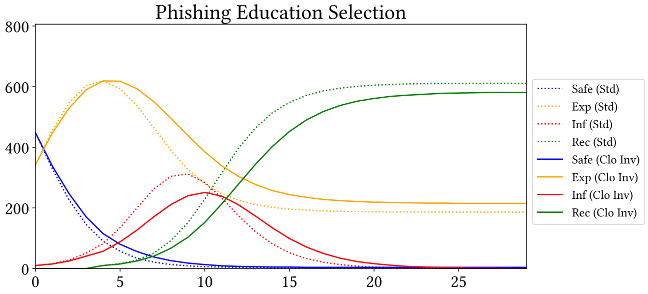
Seven selection methods are used to model employee training over a year, represented by 12 rounds of potential training. Each network starts with each node's survival rate at an identical 42% with trust coefficient $\eta$ remaining at 0.1 for the entire process.

The control group uses the failure response method, selecting users who fail tests to give mandatory phishing training and decaying those who pass for 12 continuous rounds. Three methods complete a single round of failure response, followed by two subsequent rounds of centrality-based selection methods. The program determines the centrality of the trust network, selecting 400 nodes with the highest centrality score, training one half one round, then the next half the following. Completing this, the program repeats the previous steps until 12 rounds have passed. Each method uses a different centrality measure, which, for this program, are Degree, Closeness and Eigenvector. The final three methods complete a single round of failure response, again followed by two rounds of centrality-based selection methods. However, the program chooses the 400 least central nodes for these methods to improve anti-phishing awareness, training half one round and the final 200 the next. The program repeats these steps until all 12 training rounds have been completed. The measures of centrality are Degree, Closeness and Eigenvector.

The program will generate 20 networks for each method and complete 25 infection simulations across each network. Starting positions of infected nodes will be identical across all methods but change each simulation. The program will record the timestep when 10%, 25% and 50% of the network has become infected, the timestep when most nodes are infected, and the total number of infections. Additionally, it will produce the average number of the node stages for each timestep over all simulations on each method.

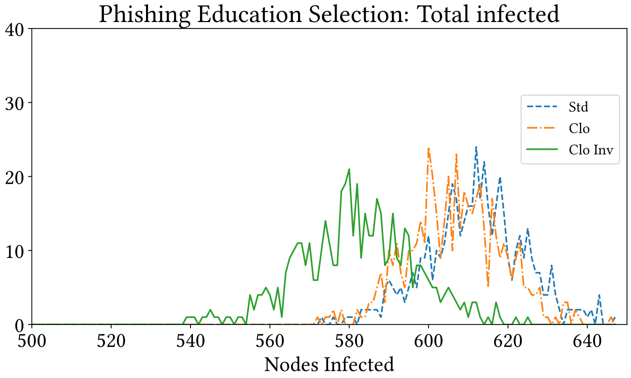
## 5 RESULTS AND ANALYSIS

### 5.1 Average Composition



While the selection methods by choosing nodes with the highest centrality did not significantly deviate from the average composition created through failure selection, the inverse methods significantly altered the composition curves throughout the lifetime of the simulation. All inverse methods reduced the infection and recovery rates across the entire program duration, resulting in a smaller infection curve, pushing the centre of the curve later. This could improve analysts' response to a phishing attack, as they have more time to evaluate the cause and, additionally, are not as overworked recovering secrets as would be with the failure or most central methods. As expected, all methods retained the general SEIR composition and structure, and no method produced worse infection rates than the failure selection method.

### 5.2 Final Infection Total Analysis



The Closeness selection method decreased the average final infection total but did not alter it significantly enough to draw any conclusion. However, the inverse closeness did reduce the average total infection by approximately 35 nodes or approximately 5.7%. This indicates that if a large-scale, unavoidable attack occurs or a highly aggressive phishing attack rapidly spreads across a network, the final state, if the nodes were selected using the inverse Closeness method, would be less damaging to the network. However, as the potential gain is relatively small, this may only recoup significant damages on a large network. Therefore, a more extensive network analysis is required to determine the utility of this method.

## 6 CONCLUSION AND FUTURE WORK

This project has covered multiple methods of selecting users to complete phishing education through the use of trust networks, SEIR simulations, and the Enron email network. The experiments show that counterintuitively assigning education to nodes that are not well connected can significantly increase the time taken to reach peak infections, decrease the total number of infections, and lower the overall infection rate when compared to the current failure-based selection method. Additionally, educating nodes using least-centrality increases the time for a phishing attack to gain 10%, 25% and 50% of all nodes on a network. Most central methods do not significantly deviate in performance for any metric than the failure-based selection method. Using the centrality method does not significantly affect the selection method's performance.

The findings of this project show there is significant efficacy to gain from novel methods of selecting users to participate in anti-malware training. However, further research needs to be completed to assert whether these methods are successful in the real world. Further studies could involve larger, more recent email-based networks. Additionally, the trust-based network within this project remained static for all selection methods. Therefore, further research into trust networks and methods of increasing employee email scepticism may provide more realism for the networks produced in this project.

# REFERENCES

[1] 2017. Enron network dataset – KONECT. http://konect.cc/networks/enron

[2] Gamze Canova, Melanie Volkamer, Clemens Bergmann, and Roland Borza. 2014. NoPhish: an anti-phishing education app. In *Security and Trust Management: 10th International Workshop, STM 2014, Wroclaw, Poland, September 10-11, 2014. Proceedings 10*. Springer, 188–192.

[3] Nakul Chitnis. 2017. Introduction to SEIR models. In *Workshop on mathematical models of climate variability, environmental change and infectious diseases, Trieste, Italy*.

[4] Ronald Dodge, Kathryn Coronges, and Ericka Rovira. 2012. Empirical Benefits of Training to Phishing Susceptibility, Vol. 376. https://doi.org/10.1007/978-3-642-30436-1_37

[5] Egress Software. 2024. 2024 Egress Email Security Risk Report. https://pages.egress.com/whitepaper-email-risk-report-01-24.html

[6] Soodeh Hosseini and Mohammad Abdollahi Azgomi. 2016. A model for malware propagation in scale-free networks based on rumor spreading process. *Computer Networks* 108 (2016), 97–107. https://doi.org/10.1016/j.comnet.2016.08.010

[7] Dijiang Huang and Vetri Arasan. 2010. Email-based social network trust. In *2010 IEEE Second International Conference on Social Computing*. IEEE, 363–370.

[8] Dijiang Huang, Zhibin Zhou, Xiaoyan Hong, and Mario Gerla. 2010. Establishing email-based social network trust for vehicular networks. In *2010 7th IEEE Consumer Communications and Networking Conference*. IEEE, 1–5.

[9] Daniel Jampen, Gürkan Gür, Thomas Sutter, and Bernhard Tellenbach. 2020. Don't click: towards an effective anti-phishing training. A comparative literature review. *Human-centric Computing and Information Sciences* 10, 1 (09 Aug 2020), 33. https://doi.org/10.1186/s13673-020-00237-7

[10] Bryan Klimt and Yiming Yang. 2004. The Enron Corpus: A New Dataset for Email Classification Research. In *Proc. Eur. Conf. on Mach. Learn.* 217–226.

[11] Ponnurangam Kumaraguru, Justin Cranshaw, Alessandro Acquisti, Lorrie Cranor, Jason Hong, Mary Blair, and Theodore Pham. 2009. School of Phish: A Real-World Evaluation of Anti-Phishing Training. *SOUPS 2009 - Proceedings of the 5th Symposium On Usable Privacy and Security*. https://doi.org/10.1145/1572532.1572536

[12] Jérôme Kunegis. 2013. KONECT – The Koblenz Network Collection. In *Proc. Int. Conf. on World Wide Web Companion*. 1343–1350. http://dl.acm.org/citation.cfm?id=2488173

[13] Isabella Martínez Martínez, Andrés Florián Quitián, Daniel Díaz-López, Pantaleone Nespoli, and Félix Gómez Mármol. 2021. MalSEIRS: Forecasting Malware Spread Based on Compartmental Models in Epidemiology. *Complexity* 2021 (27 Dec 2021), 5415724. https://doi.org/10.1155/2021/5415724

[14] Jordan Schroeder. 2017. *Chapter 4: Persistent Training*. Apress, 40–48.

[15] Wang-Sheng Lee Fadi Al Jafari William Yeoh, He Huang and Rachel Mansson. 2022. Simulated Phishing Attack and Embedded Training Campaign. *Journal of Computer Information Systems* 62, 4 (2022), 802–821. https://doi.org/10.1080/08874417.2021.1919941 arXiv:https://doi.org/10.1080/08874417.2021.1919941

# A APPENDIX

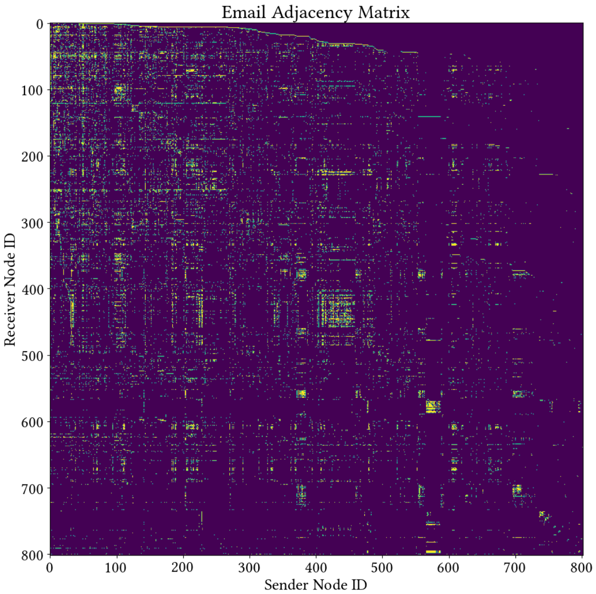
## A.1 Enron Corpus Data Set Entry

```
...
 -----Original Message-----
From:     Dunton, Heather
Sent:     Wednesday, December 05, 2001 1:43 PM
To:       Allen, Phillip K.; Belden, Tim
Subject:            FW: West Position

Attached is the Delta position for 1/16,
1/30, 6/19, 7/13, 9/21



 -----Original Message-----
From:     Allen, Phillip K.
Sent:     Wednesday, December 05, 2001 6:41 AM
To:       Dunton, Heather
Subject:            RE: West Position

Heather,

This is exactly what we need. Would it
possible to add the prior day for each of
the dates below to the pivot table. In
order to validate the curve shift on the
dates below we also need the prior days
ending positions.

Thank you,

Phillip Allen
...
```

## A.2 KONECT 2017 Enron Data Set Entry

```
...
1000 966 1 1004411463
1000 966 1 1004411913
100 100 1 1010781654
100 10756 1 1010687491
100 10756 1 1010781654
100 10756 1 1011379321
1001 1002 1 1006901795
1001 1077 1 1002320021
100 111 1 1010515702
100 111 1 1011379321
1001 129 1 1002320021
...
```

## A.3 Strong Network Adjacency Matrix


Email Adjacency Matrix
Receiver Node ID
Sender Node ID