Phishing Training Strategies: Useful or Wasteful?

Elizabeth Walkup
Stanford University
ewalkup@stanford.edu

1. INTRODUCTION

In the security community, it is commonly known that, no matter how well-built your network, humans will always be a weak link. In particular, the classic phishing email attack is still widely used by attackers trying to gain access to an organization's computing resources. Although phishing may seem simplistic, it can be extremely effective. In a phishing quiz taken by over 19,000 Americans in the last year (2015), Intel found that over 80% fell for at least one phishing email[1]. In its 2013 Data Breach Investigation Report, Verizon found that with just 12 emails, attackers statistically have an almost 100% chance of getting someone to click[2].

Because of this, many security teams pour a huge amount of time and effort into stopping phishing before it reaches the user, but they cannot catch everything. It has become common to give users some level of cybersecurity training to help them recognize and respond appropriately to phishes on their own. However, in any given group, the technical knowledge of its members will vary widely. There are some people who fall for phishes time and time again despite having the same training as everyone else. Unfortunately for companies, these people tend to be management or administrative personnel who are highly connected within the email graph. With that in mind, the goal of this project was to model the spread of a simple phishing attack and examine the pros and cons of extra targeted cybersecurity training within an organization.

2. RELATED WORK

Almost no has been done on modeling the spread of phishing specifically, and many phishing training techniques have been proposed, they have almost never been evaluated experimentally against each other. However, work has been done in modeling malware and spam spread, of which phishing is often a component. The most similar work is a 2010 paper by Tong et. al. [10] which looked at theoretical ways to express the vulnerability of an arbitrary graph to a virus. They developed a selective shielding defense that targeted the most vulnerable nodes with high betweenness.

There are examples of malware infection modeling in systems unrelated to email. Pradip et. al. [9] modeled infections in wireless sensor networks starting with a single compromised node. Work has been done on containing social network malware and spam infections by immunizing the nodes with the highest PageRank. [11] Graph infection theory is even used to predict which hosts in a network will be infected next by a botnet. [12]

There is a bit more material where relevant network theory is concerned. Phishing spread is similar to other contagion models, except that more infected neighbors does not necessarily lead to a higher infection rate. Pastor-Satorras [3] examined targeted immunization strategies based on node connectivity, a strategy that inspired some of the training models used here. Jackson et. al. [4] examined diffusion and contagions among graphs whose nodes were heterogeneous agents with different vulnerabilities, and found that the most vulnerable clusters could drag down the more resilient ones. Also more recently, a study was done on the effect of heterogenous thresholds on social graph infections [13] - it was shown that a large variation in node thresholds caused continuous and easily predictable changes in the cascades. However, this study examined the impact of multiple initial infection points, rather than the ideal threshold distribution.

3. THE DATA

When examining the problem of phishing, three different email graphs were considered:

1. Enron Email [7] - This data set was gathered from the Enron emails made public during the federal investigation. It is the smallest graph.

2. EU Research Group Email [8] - This graph comes from a European research organization, and is much larger than the Enron graph.

3. Filtered EU Research Group Email [8] - This graph is what is left after all non-reciprocal edges are removed. The idea is that this would leave behind only the stronger ties. Roughly one quarter of the edges were removed, but the number of nodes is the same.

3.1 Relevant Characteristics

Table 1 is a full list of graph characteristics. Notably, the Enron email network has both the highest clustering coefficient and the highest average degree, despite being the smallest graph. It also has a strongly connected component (SCC) and weakly connected component (WCC) equal to over 90% of the graph, making it much easier to spread infections. The SCC of the EU graphs are tiny in comparison, but the graphs themselves are larger, so graph infections are only compared within techniques on the same graph. The fact that an infection on an EU graph is larger than an infection on the Enron graph is generally just a result of graph size differences.

Node eccentricity, the largest shortest-path distance from a given to any other node, is also much higher in the Enron graph. So while it may it may have clusters, it also has pockets that take much longer to reach.

Table 1: Graph Characteristics

	Enron	EU Orig.	EU Filtered
Nodes	36,692	265,214	265,214
Edges	183,831	420,045	164,560
Average Degree	10.02	3.17	1.24
Max Degree	1383	7636	1460
Largest SCC	91.8%	12.9%	12.3%
Largest WCC	91.8%	84.8%	22.2%
Diameter	13	11	11
Modularity	0.0	0.25	0.25
Clustering Coefficient	0.4970	0.0671	0.0196
Avg. Eccentricity	8.153	5.410	1.331

4. THE EMPLOYEE MODELS

Each email graph represents a company. In these models, three types of company employees are considered. Note that these are only generalizations for modeling purposes - in real life, there is a wide range of skill levels.

1. Managers - on average more well connected, but generally less technically knowledgeable and more likely to be the target of and to fall for phishing.

2. Technical Employees - these employees generally have a computer science or engineering background, and so are less likely to fall for phishes.

3. Regular Employees - employees without any background knowledge about phishing, but less likely to be the target of a phishing attack than a manager.

Although these are model generalizations, they have some basis in real life. Managers, and top-level employees titles like CEO and Vice President, do not generally have a technical background since they are more concerned with the business side of things. While it appears no formal study has been done on the technical expertise of managers and above, there is plenty [14] of anecdotal evidence[15] to support the stereotype of a somewhat technically clueless overseer. This stereotype is one almost ingrained in our culture.

Nodes in the graphs are assigned an employee category and then a susceptibility based on their category's susceptibility distribution. All distributions are Gaussian: managers $\sim N(0.6, 0.1)$, regular $\sim N(0.7, 0.2)$, technical $\sim N(0.75, 0.1)$. These choices are based mainly on an Intel survey of over 19,000 visitors from more than 140 countries. [1] Their results, pictured in Figure 1, roughly fit a normal distribution centered at 0.7. These results are also seen in a smaller study done by Montclair State University [17].

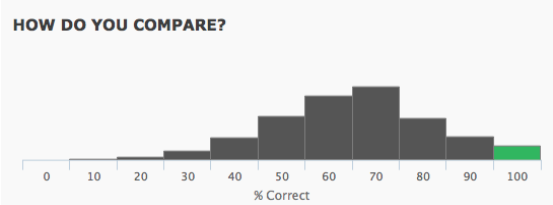

Figure 1: Plot of Intel phishing scores

The Regular Employee category is meant to reflect this distribution, as they are representative of 'the average person' in this study. The other two categories represent theoretical distributions to reflect the high and low spectrums of technical skill.

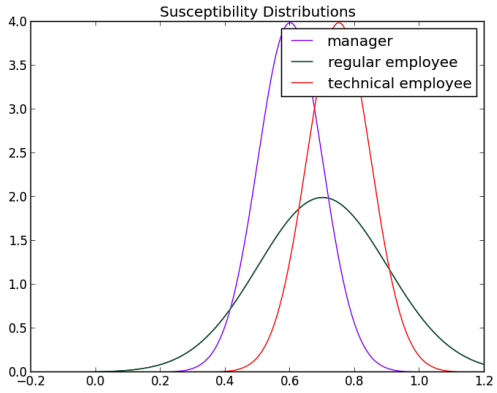

Figure 2: Plot of susceptibility distributions

Note that the higher the number, the less likely a node is to be infected - for example, a node with a susceptibility of 0.7 has a 30% chance of getting infected.

Analysis of the Enron employee data [16] shows that those in managerial positions are more likely to have a higher degree, although the difference is not too stark. Managers make up 50% of the top 10% by degree and 66% of the upper middle degree range (30-50 percentile), while they make up only 35% of the lowest 10% of degrees. Manager labels in these experiments have been adjusted to roughly follow these real-life percentages.

According to a survey done on companies from 2006 to 2011, the average number of employees to managers is 23:1. [5] But there are also on average 4 to 5 layers of management, which inflates the number greatly. In the Enron data, around 50% of the employees have a managerial title. Since companies very seldom release this information publicly, the percentage of managers used (50%) reflects the real data that can be seen. The ratio of technical employees to regular employees is determined by the type of company being modeled.

4.1 A Tech Company

Many companies in the Bay Area of California are built around new technology. These kinds of companies will have a higher ratio of technical workers than an average US company. So when modeling a tech company, the employees

are split roughly half and half, with 25% of the nodes being labeled as technical employees and 25% being labeled as regular employees. The missing 50% are managers, as mentioned above.

4.2 A Normal Company

Based on surveys done, the average company has a much lower ratio of technical employees [6]. So when modeling an 'average' company, technical employees are in a ratio of 1:5. Regular employees make up 45% of nodes and technical employees make up 5%.

4.3 The Training Models

The main point of interest in this study is the training models. Each model is evaluated for its effectiveness in reducing cascade size.

1. No training - this is the control baseline. All other training performances will be compared to this.
2. General training for everyone - this models the normal company-wide cybersecurity training. Every node's susceptibility is reduced by $\alpha = 0.03$.
3. Targeted training for the 'worst offenders' - this is more concentrated, more effective training for those with the worst susceptibilities. The worst 10% of nodes have their susceptibility reduced by $\beta = 0.09$.
4. Targeted training for worst offenders AFTER training for everyone - in this case, both (2) and (3) are applied.
5. Training for most connected nodes - the top 10% of nodes with the highest degree have their susceptibility reduced by $\beta = 0.09$.
6. Training for most connected nodes AFTER training for everyone - in this case, both (2) and (5) are applied.

Targeted training methods are modeled as 3 times more effective than the general training. The susceptibility training is what tests the goal of the paper, since this is what companies are often already looking at. The degree training is there to compare how that fares against a more graph-based logical feature. In practice, it would be difficult for a company to isolate people by email degree.

5. THE INFECTION MODELS

The infection model dictates the way the phishing infection spreads. There are two classic infection models that were considered. Other infections models, such as ones that infect a node based on the number of infected neighbors, were deemed unrealistic for a phishing scenario.

5.1 Susceptible-Infected (SI)

In this model, all nodes start out susceptible, and once they become infected, they stay infected and cannot be infected a second time. This is the main model considered because, for security purposes, once a computer is infected, the damage is done. Infecting the computer with the same thing a second time does not gain the attacker anything.

For the purposes of this study, this is also equivalent to the Susceptible-Infected-Recovered (SIR) infection model, because it does not allow multiple infections. Once infected a node is considered 'immune', as is the case with recovered nodes.

5.2 Susceptible-Infected-Susceptible (SIS)

In this model, all nodes start out susceptible, but after they become infected, they become susceptible again after 2 rounds. This models someone cleaning the malware infection from the phish off their computer, and then somehow falling for the phish again. This is a very pessimistic view of user intelligence, and, in practice, did not seem to be a very helpful model for this problem. The extra susceptibility would make things run much longer as infections became established, and there was no good criteria for a stopping point. In light of this, this type of infection was eventually scrapped to focus on more meaningful results.

6. THE EXPERIMENT

An experiment runs as follows:

1. The susceptibilities are drawn for each node from the distributions, as described in the Employee Models section.
2. For each training model:
   (a) The relevant training is applied.
   (b) For 25 runs:
      i. A node is chosen at random for the initial infection
      ii. Each time a node is infected, all its neighbors have a single chance to be infected in the next round.
      iii. The infection ends when no nodes are infected in the final round.
      iv. If the infection was a cascade, its results are recorded. For this study, a cascade is defined as an infection lasting more than five rounds and having size ≥ 100.

Multiple experiments are run and averaged to obtain the results.

7. RESULTS

7.1 The Tech Company Model

Table 2: Tech: % of times a cascade was found

	Enron	EU Orig.	EU Filtered
None	100.0%	63.3%	54.5%
Targeted Susceptibility	93.3%	66.7%	73.3%
Targeted Degree	100.0%	80.0%	63.3%
General	100.0%	73.3%	56.7%
General + Sus	100.0%	73.3%	63.3%
General + Degree	100.0%	70.0%	36.7%

Figures 3 and 4 contain average results for a tech company and all the different training models under an SI infection model. The most interesting result is between the degree targeting and the susceptibility targeting. When done without general training, targeting by degree is slightly better for all graphs, nearly equal to general training on its own. But when paired with general training, targeting nodes by

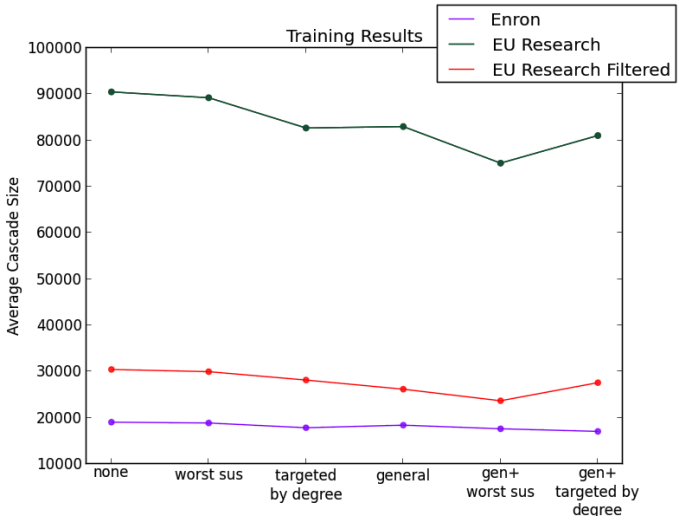

Figure 3: Tech Company: Cascade sizes

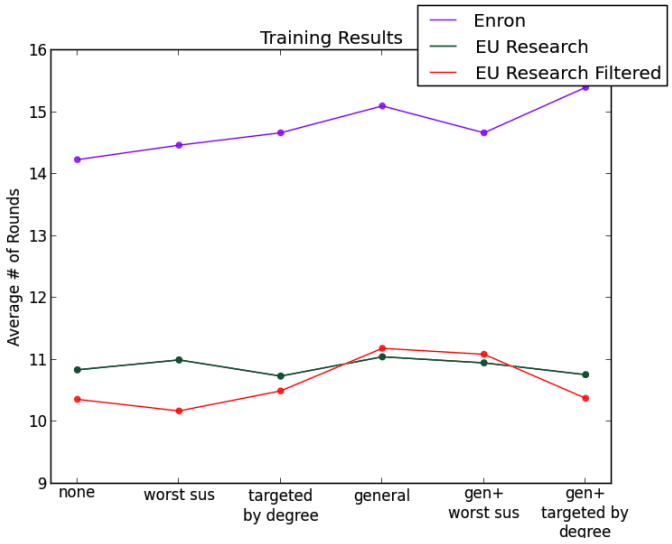

Figure 4: Tech Company: Rounds per cascade

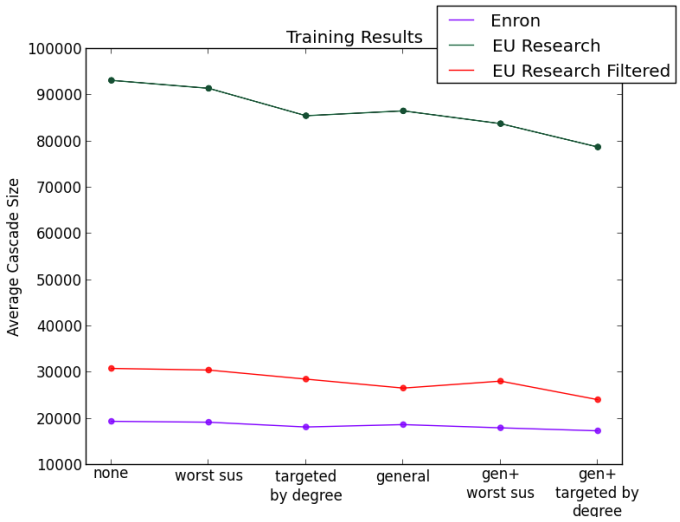

Figure 5: Normal Company: Cascade sizes

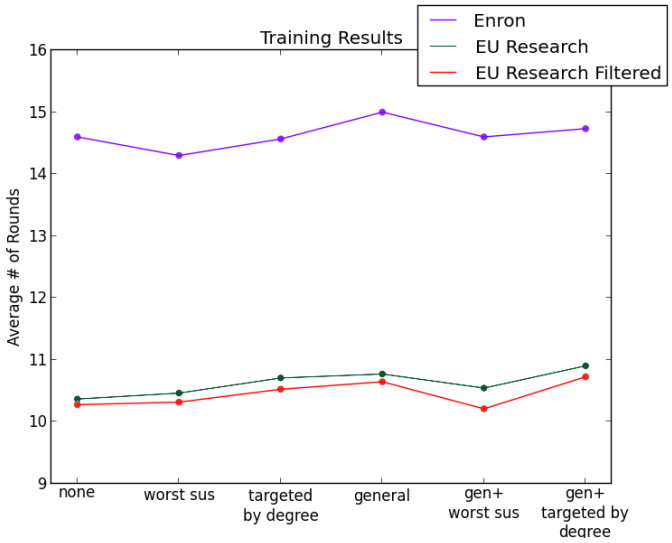

Figure 6: Normal Company: Rounds per cascade

susceptibility outperformed degree targeting significantly in both EU graphs. It performs roughly the same on the Enron graph. Since this effect is much more pronounced in the unfiltered EU graph, it seems that it is likely that this is related to the number of edges. As a side note, in earlier trials, when targeted types of training acted on a relatively small (1000 nodes) set, not much difference was noticed, but when the targeted set was upped to 10% of the graph, differences became more apparent.

On an individual training level, targeting by degree is the most effective for the Enron graph and the unfiltered EU graph. However, general training is the most effective on the filtered EU graph.

The graphs do not show the entire story. Looking at Table 2, it is easy to see that Enron's highly connected nodes come into play here. During each experiment (with 25 tries), it almost never fails to find a cascade on the Enron graph. However, looking at the EU graph, training can in fact help prevent cascades from starting at all. In this case, the general + degree targeted training is extremely effective in preventing cascades on the filtered EU graph, much more so that general + susceptibility targeting. It also provides a very small benefit in the original EU graph, but not enough to be statistically meaningful.

Despite having the smallest cascades and the highest connectivity, the Enron graph has the most rounds. The edge filtering of the EU graph does not seem to strongly affect how many rounds it has. Interestingly, as cascades get smaller, in general the number of rounds goes up, not down. This suggests that training primarily reduces the size of each round, rather than blocking rounds from happening once a cascade occurs.

7.2 The Normal Company Model

Table 3: Normal: % of times a cascade was found

	Enron	EU Orig.	EU Filtered
None	100.0%	83.3%	73.3%
Targeted Susceptibility	100.0%	66.7%	56.7%
Targeted Degree	100.0%	76.7%	53.3%
General	100.0%	66.7%	53.3%
General + Sus	100.0%	70.0%	70.0%
General + Degree	100.0%	84.8%	36.7%

Figures 5 and 6 contain average results for the normal company model, once again with an SI infection model. This time, combining degree targeting with general training beat general + targeted susceptibility, although susceptibility training added some benefit as well.

As in the tech company case, training by degree is the most effective individual training for the Enron and original EU graphs. General training continues to slightly lead on the filtered EU graph.

Again, as with the tech company, the Enron graph infections always manage to trigger a cascade. Cascades occur slightly more often for the EU graphs here than they did in Table 1, but not in a significant amount. Degree targeting + general training works extremely well at preventing cascades in the filtered EU graph, but here it does worse in the original graph.

Unlike the tech company results, where the number of rounds typically increased, this time general + susceptibility training caused a noticeable drop in the number of rounds for all graphs, which no other training did. However, overall the number of rounds still fits within the same ranges as those of the tech company experiment. Since targeted susceptibility lost to targeted degree training in these experiments, this suggests that the number of rounds does not have as large an impact on the final cascade size as would seem logical.

8. CONCLUSIONS

Clearly, general training for the entire staff offers a substantial benefit - this is just another bit of evidence that points to the importance of having company-wide cybersecurity training. But the focus of this experiment was on the other kinds of training.

The original question was: would targeted susceptibility training for users help a company's security? In this case, targeted susceptibility training works best at a company where a good number of the employees are technically knowledgeable (about phishing). This is because when more of your company is mostly immune, plugging the experience gaps makes it much less likely that the cascade will spread. This is why it works even better when paired with general training. Susceptibility training had no significant impact on cascade occurrence in any experiment.

On the other hand, in the modeled 'normal' company, targeted susceptibility training did not reduce cascade size as much as targeted degree training. Targeted degree training was also better at preventing cascade occurrence in the sparse, filtered EU graph in both company models. However, this kind of targeted training would be more difficult from a practical standpoint, since a company would have to build their own graph to know who to target.

The final conclusion is that targeted susceptibility training does indeed have merit, but to get a good return on investment a company needs to have the right conditions. Since the tech company model saw a larger benefit than the normal company model, it helps to have at least one-fifth of the staff reasonably trained in phishing to being with. It also helps more if the company's communication graph is not very dense, as seen in the Enron versus EU graph results.

9. FUTURE WORK

When exploring the problem further, other things to look at would be the effectiveness of other kinds of targeting features, like betweenness or PageRank. It would also be worth looking at other metrics that can more easily be obtained in practice - for instance, targeting by volume of email sent and received.

In the interest of making the models more accurate, a survey of phishing scores divided up by job category would be helpful for making sure the manager and technical employee models are realistic. Data from simulating phishing spread in a real email network would also be useful.

The main thing that is needed is more data, especially more real company email graphs. The variation in the results between the three graphs studied here shows that what works on one graph may not work as well on another. So while targeted training methods are worth looking into, each company would have to do their own individual assessment. The ideal assessment would be to combine a company email graph with the knowledge of each person's performance on phishing training tests (since this information would likely be considered proprietary, it probably will not be done in a published paper).

10. REFERENCES

[1] Amanda Schupak *Majority of Americans fall for email phishing scams*, (CBS News) (2015) `http://www.cbsnews.com/news/majority-of-americans-fall-for-email-phishing-scams-cbs-intel-security-quiz/`

[2] Verizon. *2013 Data Breach Investigations Report* (Verizon) (2013) `http://www.verizonenterprise.com/resources/reports/rp_data-breach-investigations-report-2013_en_xg.pdf`

[3] R. Pastor-Satorras, A. Vespignani. *Immunization of complex networks*, (Physical Review E) (2002) `http://snap.stanford.edu/class/cs224w-readings/pastorsatorras02immunization.pdf`

[4] M.O. Jackson and D. Lopez-Pintado. *Diffusion and contagion in networks with heterogeneous agents and homophily*, (Network Science 1.01) (2013): 49-67. `http://arxiv.org/pdf/1111.0073.pdf`

[5] V. Smeets, M. Waldman, and F. Warzynski *Performance, Career Dynamics, and Span of Control*, (Aarhus University) (2013). `http://www.cbs.dk/files/cbs.dk/performance_career_dynamics_and_span_of_control.pdf`

[6] Mercer Human Resource Consulting and ITAA *Ratio of IT Staff to Employees*, (Workforce.com) (2003). http://www.workforce.com/articles/ratio-of-it-staff-to-employees

[7] William Cohen *Enron email network*, (SNAP) http://snap.stanford.edu/data/email-Enron.html

[8] *EU email communication network*, (SNAP) http://snap.stanford.edu/data/email-EuAll.html

[9] PRADIP DE, YONGHE LIU, and SAJAL K. DAS *Deployment-Aware Modeling of Node Compromise Spread in Wireless Sensor Networks Using Epidemic Theory*, (ACM Transactions on Sensor Networks (TOSN)) (May 2009):Volume 5 Issue 3, May 2009 Article No. 23. http://dl.acm.org/citation.cfm?id=1525861

[10] Tong et. al. *On the Vulnerability of Large Graphs*, (Data Mining (ICDM), 2010 IEEE 10th International Conference) (2010):pp. 1091 - 1096 . http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=5694090&tag=1

[11] J. Kong, M. Han *Malware Containment Using Weight based on Incremental PageRank in Dynamic Social Networks*, (KSII Transactions on Internet and Information Systems (TIIS)) (2015):Vol.9 No.1, pp. 421-433 . http://www.dbpia.co.kr/Journal/ArticleDetail/NODE06139375

[12] J. Rrushi, E. Mokhtari, and A. Ghorban *Early Stage Botnet Detection and Containment via Mathematical Modeling and Prediction of Botnet Propagation Dynamics*, (University of New Brunswick) (2010). http://ftp.cs.unb.ca/tech-reports/documents/TR10-206.pdf

[13] P.Karampourniotis, S. Sreenivasan, B. Szymanski, G. Korniss *The Impact of Heterogeneous Thresholds on Social Contagion with Multiple Initiators*, (Cornell University) (2015). arxiv.org/pdf/1506.00986v1.pdf

[14] B. Brooks *Clueless Managers and Their Technology Bumblings*, (The Brooks Review) (2010). https://brooksreview.net/2010/05/clueless-managers-and-their-technology-bumblings/

[15] B. Sullivan *CEOs Are Clueless About What Their Information Managers Actually Do*, (TechWeek Europe) (2015). http://www.techweekeurope.co.uk/e-management/ceos-clueless-information-managers-actually-165173

[16] A. Schulz *Enron Email Data*, http://www.ahschulz.de/enron-email-data/

[17] J. Ragucci *Learn to Catch the Phish: A Tool for User Education* , (Montclair State University) (2005). http://pages.csam.montclair.edu/~robila/SECURITY/F2005_P/P4/Ragucci_Phishing_Paper.pdf

APPENDIX

A. GRAPH VISUALIZATIONS

These visualizations of the email networks made in Gephi using the Force Atlas 2 layout. Nodes are colored and sized by degree.

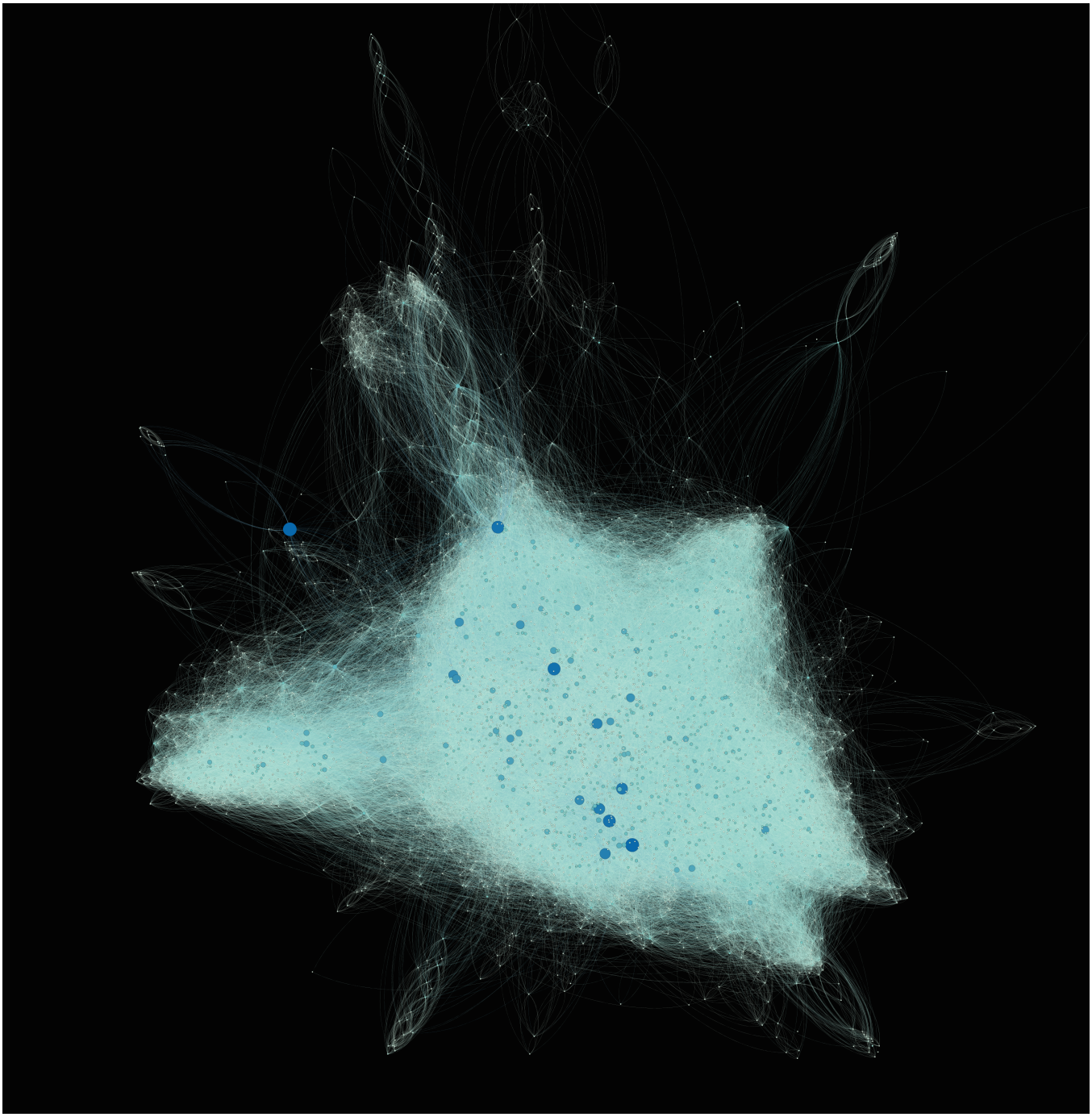

Figure 7: Enron Graph, filtered to nodes with degree ≥ 15

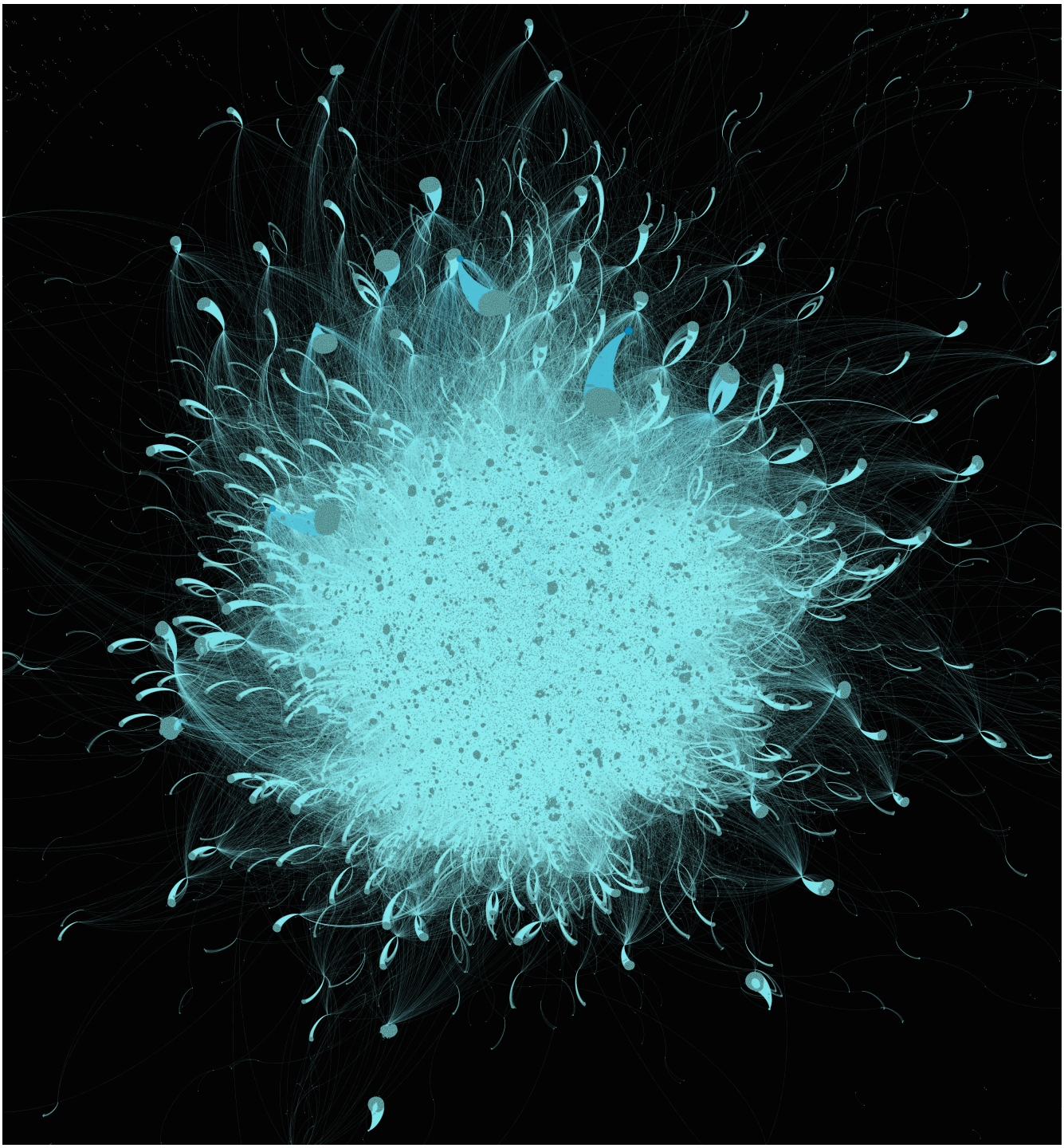

Figure 8: EU Graph, unfiltered