UNICORN VYSOKÁ ŠKOLA S.R.O.

Softwarový vývoj

EXTENDED SUMMARY

Modeling of chain-email spreading

Author: Tomáš Jirků

Supervisor: RNDr. Mgr. Horáček Jaroslav, Ph.D.

2023

Extended Summary

The spread of chain emails and their potentially dangerous content is still a relevant phenomenon. Although not all chain emails' content is necessarily. In the Czech Republic, they are often utilized as a propaganda tool by foreign countries, mainly Russia and China.

Empirical observation of chain emails is extremely difficult, if not nearly impossible, because emails are inherently private in their nature, and obtaining information about their dissemination through legal means is challenging. In contrast to chain emails, observing the spreading of information on social networks is easier due to their public nature.

There are not many models specifically designed for chain emails, but there are models available for social networks. Models for information propagation in social networks are often inspired by epidemic spreading models [4]. The structure of email networks and large social networks is relatively similar [2]. This raises the question of whether the spread of chain emails may follow the same or similar principles as the spread of messages in social networks.

In the year 2008, Jon Kleinberg and David Liben-Nowell published a unique study describing the structure of the spread of two internet petitions distributed via email. They also proposed a model that replicates this structure when applied to real networks. The spread structure produced by this model is very different from the structures produced by models of information propagation in social networks.

The main goal of this bachelor's thesis was to describe the structure and behavior of chain email spreading. I formulated the following steps to achieve this goal:

- Analysis of existing models
- Proposal of modifications or extensions to existing models to best describe the spread of chain emails.
- Implementation of an application for simulating selected models
- Implementation of an application for visualizing the spread of chain emails
- Analysis of the results obtained from the implemented models.

Afterwards I analyzed several rumor spreading models in social networks, model proposed by David Liben-Nowell and Jon Kleinberg (LNK model) and Galton-Watson model, which allows to imitate results of LNK model and serves as a good reference model for it. [3]

After the analysis I designed modified variant of the most recent and complex rumor spreading model from those analyzed. The modification incorporated average reaction type to conspiration theories and hoaxes amongst Czech population. I also designed modified variant of LNK model, which could describe the whole dissemination of chain email and does not only describe observable part of internet petitions.

I implemented those modified models together with Galton-Watson model, which serves as reference model for LNK model. I also implemented tool for visualization of structures of graphs produced by implemented models.

All modified models were tested on the same networks, which were reconstructed from a real-world email network.

I analyzed the structures and development in time of spreading of those implemented models and performed analysis on the results. The analysis showed that the structure of results of both modified was comparatively similar to the structure of results of general rumor spreading models. The structure of produced graphs showed that distances between nodes were relatively short, and the graphs weren't trees, which is in direct opposition with results of original variant of LNK model and reference Galton-Watson model which produce deep trees where distances between nodes are large.

However key difference between modified models was the dispersion of results in terms of size when compared within context of the same network. Both modified LNK and reference Galton-Watson model have very large number of results, where the dissemination of email stops very rapidly, but are also able to rarely produce fairly large structures. The modified rumor spreading model has much lower dispersion of results, the structures produced tend to concentrate much tightly around the average values. Majority of results produced by modified rumor spreading model show at least some tendency to spread, but the largest spreads are much closer to average than in case of modified LNK model and Galton-Watson model.

To answer the question if and how accurately the models implemented by me describe the actual propagation of chain emails is very difficult, because there is limited data on the actual propagation of chain emails, and acquiring it is highly challenging.

However, it is known that the majority of chain emails have a small reach, but occasionally, there are chain emails that have a large reach. [1] The propagation structure has not been conclusively described yet, but most studies on this topic tend to lean towards models derived from epidemiological spread, which produce graphs with short distances between individual nodes which are not trees.

My implementation of the rumor spreading model and the

LNK model both produce graphs with short distances between individual nodes. Both models thus fulfill the propagation structure generated by most models described in scientific studies on this topic. However only the modified LNK model satisfies the uneven distribution of email reach.

The added value of this thesis is the design and implementation of two modified models aiming to simulate the propagation of chain emails in a population as faithfully as possible. Both models I have proposed are at least partially consistent with available information on the topic of chain email spreading. Another added value is the implementation of visualization tool, which can display the results produced by implemented models. The main objective of this tool was to help intuitively understand the structure of results produced by implemented models.

Major Sources

(1) Liben-Nowell, D., & Kleinberg, J. (2008). Tracing information flow on a global scale using Internet chain-letter data. *Proceedings of the National Academy of Sciences of the United States of America*, *105*(12), 4633–4638. https://doi.org/10.1073/pnas.0708471105

(2) Barabási, A.-L., & Pósfai, M. (2017). *Network science*. Cambridge University Press. Also available from: http://networksciencebook.com/

(3) Golub, B., & Jackson, M. O. (2010). Using selection bias to explain the observed structure of internet diffusions. *Proceedings of the National Academy of Sciences*, 107(24),10833–10834.

(4) Nekovee, M., Moreno, Y., Bianconi, G., & Marsili, M. (2007). *Theory of rumour spreading in complex social networks. Physica A: Statistical Mechanics and Its Applications, 374(1), 457–470.*