Utrecht University

DIPP: Information Diffusion for Privacy in Multi-agent Systems

Master's Thesis

Artificial Intelligence

Department of information and Computing Sciences

Project Supervisor:
prof. dr. Pinar Yolum Birbil
Daily Supervisor:
Onuralp Ulusoy, MSc

Author:
Albert Mwanjesa
6641016

First examiner:
prof. dr. Pinar Yolum Birbil
Second examiner:
dr. Shihan Wang

April 8, 2021

Abstract

Ensuring the privacy of users is a key component in collaborative computer systems, where users can access private information of others. Social networks are a prime example of such systems. Therefore, privacy has to be ensured not only by the administrators but also by users in collaboration. The content users choose to share may conflict with the privacy preferences of their own or those of others, given the context of the content. Thus, a decision to share or not to share can be seen as a privacy decision. To manage privacy preferences better, it is important to understand how they appear and disappear on social networks. However, it is also important to understand how the privacy preferences spread throughout the network. Given this understanding, one can reason about the implications of the spreading mechanism has on mitigating or promoting certain privacy preferences.

In this work, the diffusion of infectious privacy preferences (DIPP) model is proposed to investigate how these privacy preferences spread on social networks. An epidemic model is used to model the spread of privacy preferences. Simulations of social network interactions are used to investigate various circumstances, such as the rarity of a privacy preference and opposition of a privacy preference, and their effect on the diffusion of privacy preferences. Furthermore, we investigate the effect of using a trust model to model trust between agents. The results show that the DIPP provides a stable foundation to further research the spread of privacy preferences in online social networks.

Contents

1 Introduction 5

2 Related work 6

2.1 Privacy in online social networks . . . 6

2.1.1 How do users behave with regard to privacy in OSNs? . . 7

2.1.2 Privacy preservation and norms . . . 7

2.2 Information Diffusion . . . 9

2.2.1 Background . . . 9

2.2.2 Influence models . . . 9

2.2.3 Game Theory model . . . 11

2.3 Epidemic models . . . 12

2.3.1 Flexibility . . . 14

2.3.2 Resilience of epidemics . . . 15

2.3.3 Attention and Competition . . . 16

2.3.4 Source Detection . . . 16

2.4 Data . . . 18

2.4.1 Semantically Analysed Metadata of Tumblr Posts and Bloggers . . . 18

2.4.2 Stack Overflow temporal network . . . 18

2.4.3 Interaction-Based Behavioural Analysis in Twitter Social Network . . . 19

2.4.4 Stanford Network Analysis Project . . . 19

2.4.5 Anonymized Instagram network data from Amsterdam and Copenhagen . . . 19

2.4.6 An exploration of the Facebook social networks of smokers and non-smokers . . . 19

2.4.7 Dynamics of Instagram Users . . . 20

2.5 Trust in multi-agent systems . . . 20

3 Modelling privacy preference diffusion as information diffusion 20

3.1 Context of content shared . . . 23

3.2 SIR model for privacy preference diffusion . . . 24

3.3 Opposing privacy preferences . . . 25

3.4 Trust . . . 27

3.5 Co-ownership . . . 30

4 Agent-based modelling of privacy preference diffusion 31

4.1 Agent . . . 31

4.1.1 Agent initialization . . . 31

4.1.2 Agent Actions . . . 34

4.2 Model . . . 36

5 Experiments 37

5.1 Measurements . . . 37

5.1.1 Influence . . . 37

5.1.2 Epidemic endings . . . 38

5.1.3 Epidemic peaks . . . 38

5.2 Simulations . . . 39
5.2.1 Basic settings . . . 39
5.2.2 Rare privacy preference with high degree influencers . . . 40
5.2.3 No Trust . . . 40
5.2.4 Static Trust . . . 41
5.2.5 Dynamic Trust . . . 41
5.2.6 Dynamic trust with two levels of privacy violations . . . 42
5.2.7 Opposing privacy preference dynamic . . . 43
5.3 Social Network Data . . . 43

6 Results 43
6.1 Including opposing preference dynamics . . . 45
6.1.1 Baseline simulations . . . 45
6.1.2 Rare privacy preference with high degree influencers . . . 53
6.1.3 No Trust . . . 57
6.1.4 Static Trust . . . 59
6.1.5 Dynamic Trust . . . 64
6.1.6 Dynamic trust with two levels of privacy violations . . . 67
6.2 Pro privacy preference dynamics only . . . 70
6.2.1 Baseline simulations . . . 71
6.2.2 Rare privacy preference with high degree influencers . . . 76
6.2.3 No Trust . . . 80
6.2.4 Static Trust . . . 81
6.2.5 Dynamic Trust . . . 84
6.2.6 Dynamic trust with two levels of privacy violations . . . 86
6.3 Summary . . . 89

7 Discussion 90
7.1 Main properties of the state dynamics . . . 90
7.1.1 Rooted in information diffusion . . . 91
7.1.2 Privacy preferences that sell themselves . . . 91
7.1.3 What to trust? . . . 91
7.1.4 Influencers beware! . . . 91
7.1.5 Opposing privacy preferences: static or dynamic . . . 92
7.2 Validity of methodology . . . 92
7.2.1 Trust only as a shield? . . . 92
7.2.2 Protection against influencers . . . 92
7.2.3 Actions per step . . . 93
7.2.4 Definition of time . . . 93
7.2.5 Initial spread of privacy preferences . . . 93
7.2.6 Customization . . . 93

8 Future Work 93
8.1 Opposing privacy preference dynamics . . . 94
8.2 Different epidemic models . . . 94
8.3 Allowing more interactions . . . 94
8.4 Not all friends are equal . . . 95

9 Conclusion 95

Acknowledgements

This master's thesis would not have been possible without the help of various individuals. Firstly, I would like to thank my first supervisor Onuralp Ulusoy for his guidance, all the insightful discussions and his overall commitment throughout the whole project. I would also like to thank professor Pinar Yolum for her useful insights and reassuring oversight throughout the research period.

I, additionally, want to express my gratitude to Asun Alsina Lopez, Elfia Bezou-Vrakatseli and Sinan Robillard for their assistance in reviewing this document and making its current state achievable.

1 Introduction

With the rise of social media and the prevalence of smartphones, humans are more connected online than they have ever been before. In general, this means people can easily communicate with each other no matter their location, physical social circles and so on. People have been able to leverage social media to more efficiently advertise businesses due to the ease of communication. Governments can reach their citizens quicker and more directly than before. Furthermore, many families spread out across the world can stay in touch with ease. This is only a subset of the positive aspects of social media. For researchers, social networks provide many opportunities to investigate human behaviour on a large scale. A negative side, however, is the implications of online social networks (OSNs) on the privacy of the users.

The content shared by users can reveal a lot about them, either explicitly or implicitly. In the past decade, governments have used social media to influence their citizens. Using user data from OSNs, citizens can be targeted for propaganda advertisements. Facebook is an example of a social network that has been found to sell data without informed consent from users[1]. This is an example of a privacy breach by way of the administrator of the social network. In this case, the administrator was willing to participate in the breach, but hackers can also breach the administrators' security to get user data from online platforms.

Different measures have since been put into place to secure user information. One example is the EU's general data protection regulation[2], which focuses on anonymization of the user data to mitigate the impact of data breaches on the privacy of users as well as requiring informed consent to utilize user data.

People can also have their online privacy violated by fellow users. When people share content that reveals private information about you without your consent, your privacy is violated. With fellow users, the intent is not always to harm. It may, however, be difficult for users to know what other users prefer when it comes to their privacy. Social norms for privacy guide people concerning this problem, as with much of our social interactions. Social norms can be seen as rules that coordinate the behaviour of members of a society, as defined by Young in [34]. Deviating from the social norms, generally, comes with negative consequences. Yet, these norms may not have the desired precision. A user can, for example, choose to only share information about others that he or she is comfortable sharing about themselves. It is evident that one's level of openness is not the same for all and violations are likely to occur. Social norms can clash with personal privacy preferences. The following section will show how more fine-grained norms can come about. Privacy preferences can be derived from the privacy decisions users make. The context of content shared by users describes their privacy preference, as will be elaborated on in the latter sections.

In this work, the main focus is on how the privacy preferences move across networks. An attempt is made to understand under which circumstances privacy preferences spread in OSNs.

- How does a privacy preference of posting content that fits in a given context become prevalent?

[1] https://www.bbc.com/news/technology-45976300
[2] https://ec.europa.eu/info/law/law-topic/data-protection

- Who are the most influential figures in the spreading process of the norms?
- What is the influence of trust modelling in the spread of privacy preferences?

To investigate these questions, knowledge is borrowed from the field of information diffusion. This is a field that tries to understand how information spread through social networks. The field has produced useful applications for the *viral marketing* and social network analysis [11, 19]. Viral marketing is a marketing strategy advertising on a social network in hopes that the advertisement goes viral positively. This project will contribute to a new field of research alongside the fields of privacy in multi-agent systems and information diffusion in OSNs. These contributions will consist, but are not limited to, of the following:

- Investigation of whether epidemic models are accurate models to model the spread of privacy preferences
- Analysis of the influence of different circumstances in a social network on this diffusion process
- A simulation of the diffusion of privacy decisions
- A method for users to protect themselves against opposed privacy preferences, namely using trust modelling.

Research from the field of information diffusion modelling provides a foundation upon which the proposed field of privacy preference diffusion modelling can be built. The subsequent sections introduce related work on privacy management and privacy norms. Secondly, sources from information diffusion research are laid out to provide the foundation referenced above. Thirdly, a method is sketched to explain the different variables of the simulations and their motivation. Following the method, the results of the simulations are laid with a reference to the goals stated here and the results are discussed. Finally, future work is proposed to continue the application of epidemic models to the domain of privacy preference diffusion.

2 Related work

In this section, various research projects are explored that have inspired this work. This exploration starts with some insights into privacy in online social networks. This is followed by overviews of important work in the field of information diffusion, epidemic modelling and trust in multi-agent systems. Finally, it ends with a look at useful data sets for the modelling of privacy preferences.

2.1 Privacy in online social networks

Privacy in OSNs is an active research field. The problem of privacy preservation can be viewed from various scientific angles due to its interdisciplinary properties. Sociologists, computer scientists and psychologist can all have a say on the matter. This section will lay out some important work done on the topic of privacy in OSNs.

2.1.1 How do users behave with regard to privacy in OSNs?

In work by Dupree *et al.* [7], the researchers investigate privacy personas in online social networks. The authors try to find an alternative to Westin's categories for users concerning their behaviour towards privacy and security. Using a survey, data are gathered on user behaviour towards privacy and security. A cluster analysis shows that five categories cover the data; Fundamentalists, Lazy Experts, Technicians, Amateurs and the Marginally Concerned. The authors then perform a secondary survey with a larger group to investigate the robustness of the five categories, the results of which show that the clusters indeed hold. Their research does not only cover interpersonal privacy issues as the work in this project.

A source of influence in privacy decision-making is an agent's trust towards their neighbours. A qualitative study by Lampinen *et al.* [16] has found that privacy management in content sharing decisions is mostly based on trust. Users expect each other to understand the way they want to represent themselves in OSNs. To ensure this, users are said to utilize mental/behavioural strategies as well as preventive and corrective strategies. The first dimension includes reciprocity concerning trusting other users as well as self-censoring and the division of the social network into different spaces for different types of content. Preventive and corrective strategies include the negotiation with other agents for the removal of tags or only allowing certain people to view your shared content. After the analysis of data, the researchers find there is another dimension to the strategies, namely individual and collaborative strategies. Although collaborative strategies are found to be more successful than individual ones, the subjects do not apply them enough.

2.1.2 Privacy preservation and norms

Researchers have investigated the problem of privacy preservation in OSNs using approaches based on multi-agent systems. As stated, in this work, the focus lies on privacy norm breaches caused by fellow users on a network. Norms are known to guide human behaviour and preserving them is often a collaborative effort. Norms in the context of privacy are characterized by privacy decisions, decisions to share or not share certain content in a given context. There is a difference between personal norms as well as social norms. In online social networks, it is key to understand both social norms and the personal norms of people you interact with to preserve harmony. It is, however, evident that with vast connectivity of online social networks personal norms cannot always be common knowledge. For instance, let's assume that Jerry and Sandy are friends on Facebook. Jerry never shares his political preferences online, but many others do on his network thus it is not frowned upon. When Sandy posts a photo of Jerry and her at a political rally, she infringes on Jerry's privacy without knowing because the social norm and Jerry's norm do not match.

The example given shows a direct violation of a user's privacy norm. Some researchers have pointed out that less direct violations occur after inference. For example, inferring Jerry's location from a geo-tagged photo in which Jerry is tagged, without Jerry wanting the location to be public. To detect both forms of violation PriGuard has been suggested in work by Kokciyan and Yolum [15], a framework that uses commitments between user agents and the social network administrator to detect privacy violations on Facebook. Agents collaborate by committing to certain norms. Commitments are captured using description logic.

To prevent violations from happening privacy auctioning for OSNs (PANO), by Ulusoy and Yolum [28], has been proposed. Agents, that represent users involved in a piece of content, bid on whether the content should be shared given their privacy preferences, using PANO. To achieve this, the authors employ the Clarke-Tax mechanism. This scheme is truth-dominant, meaning agents have no incentive to lie about the preferences for the sake of better utility return. User preferences are captured for different types of content as well as different audiences. Bidding is limited to group bids, minimum and maximum bids to mitigate abuse.

In work by Ulusoy and Yolum [29], researchers show that given the ability of agents to collaborate, personal privacy norms can become social norms over time. This occurs when a set of agents have similar notions of privacy. In this work, agents in the simulation start with some personal norms (m-norms) as well as common-knowledge system-wide norms (s-norms). Agents make decisions whether to post content that has co-owners given a certain context. A co-owner is someone whose private information may be revealed if the content were to be shared. The context is key as it provides users with the ability to be precise about the preferences. A user may not be opposed to appearing drunk in a photo at a friend's party, but opposed to appearing drunk in a photo at an office party. All s-norms and m-norms can be created, updated and removed from their norm base. When privacy decisions are made during the simulation, clustering of this content is used to determine what type of content has become normative. The results show that over time given this framework the ratio of s-norms and m-norms move in the favour of s-norms. This is, in turn, presents the novelty of the work. Agents do not need to agree on every piece of content before sharing rather they evaluate the s-norms to find the correct decision. There is less direct collaboration and thus less overhead in the privacy management system given this framework.

Ulusoy's and Yolum's work in [29] is the main inspiration for this research project as it shows that privacy can be managed more centrally given the emergence of social norms, but it is unclear what underlies this emergence. It is useful to understand what the circumstances are in which a privacy norm can emerge as a social norm. Understanding this enables an administrator to take measures to mitigate privacy norms that they deem undesirable. For example, in a social network for a high school, posting content in which another student's grade can be inferred, can be deemed undesirable. Given an understanding of the spread of this norm, teachers can prevent privacy violations. The alternative can be to restrict, in many ways, the type of content that can be shared by students, but this results in a social network with less rich content.

In this work, we assume that the spread of privacy preferences, just like pieces of information, can be modelled using information diffusion models. These models assume that agents in a network can be influenced by their neighbours to disseminate information or in this case to take certain privacy preferences. Instead of being moved to share by the semantics of a piece of information, it can be argued that in this case users are moved by trend following and the need to stay relevant to take certain privacy preferences. How this manifests itself, is one of the topics of the project. As specified by the last two project goals, a simulation method is proposed that should provide insight into how different circumstances in a social network influence the diffusion of privacy preferences.

2.2 Information Diffusion

The general goal of information diffusion on OSNs research is to understand the inner workings of how information spreads in these networks [11]. Information spread on OSNs has proven influential to everyday life. For example, numerous protests and revolutions that occurred over the past decade would not have had the same impact without social media. The spread of information depends on human behaviour and seems to entice human behaviour. Thus understanding the spread of information on OSNs can prove useful to fight negative information or elevate positive messages, depending on the context.

2.2.1 Background

The diffusion of information depends on social influence. Social influence can be experienced or put forth by members of a society in the form of behaviour, leading to imitation by other members. In OSNs, imitation can be explicit, for example, in the form of sharing a post on Facebook. Diffusion can occur in the form of herd behaviour when a group of people spread information given their own beliefs as well as under influence from others. Diffusion also occurs as an information cascade. In this scenario, users sequentially make similar decisions to others according to their observations [11].

In modelling information diffusion on OSNs, users are often seen as nodes and interactions between them are represented by edges. These edges may be weighted depending on the interaction and task at hand. Concerning models, there is a distinction between two approaches; explanatory and predictive models as surveyed in work by Li *et al.* [19]. Explanatory models try to shed light on the circumstances under which information spreads in OSNs. They can quantify the influence of individuals nodes as well as groups of nodes. These models also allow *replays* of the diffusion process for analysis. On the other hand, predictive models aim to predict the impact of a piece of information and where it will spread in a network. Consider the Arab spring revolution of 2011. For many governments involved, it would've been useful to be able to predict how the encouragements to protest would spread across users from their country on OSNs. This information could've prepared them better for what was coming. However, after the revolution, it is still useful to learn who were the most influential users that spread information about the revolution online. This is where explanatory models are useful. They can help identify users that maximize the spread of information, in other words, the most influential nodes on the network. Although this distinction exists, it is not the case that the two types of models are not related. To understand this, a few applications are discussed in the next sections. These sections also showcase the inspirations that are the foundation for the method put forth in the latter sections.

2.2.2 Influence models

Influence maximization research seeks to find techniques that can identify k nodes in a network that maximize the spread of a piece of information. Viral marketing is one field that benefits from influence maximization techniques. Marketing agencies can choose a set of influencers on social media to advertise their product as to maximize their reach to customers.

The influence maximization is generally formulated as a combinatorial optimization problem, since the inception of the first approach in a study by Kempe *et al.* as reviewed in [27]. The base of all approaches in the independent cascade model (IC). Given a weighted graph G with edges $(u, v) \in E$ and nodes $v \in V$, the weights represent the probability that one node can affect a directly neighbouring node. Although influence maximization is an explanatory method, the IC model is predictive. The predictive abilities are used to search for the seed set S for which the IC model predicts maximum spread.

$$S \in V, \|V\| = k$$

An exhaustive algorithm would calculate the spread for each subset of nodes. In Kempe *et al.* [14], the Greedy algorithm is proposed that starts with an empty set and nodes are added incrementally if they maximize the increase of the expected influence of the seed set. Accuracy of the estimation of the expected influence of a candidate seed set is the computational bottleneck of the Greedy approach and its successors. In a paper by Tang *et al.* [27], Two-phase Influence Maximization (TIM) is presented as a successor of the Greedy algorithm. TIM also uses sampling to estimate the expected influence of a candidate seed set. However, TIM's search is more targeted with two phases: parameter estimation and node selection. The authors show that TIM achieves an $(1 - 1/e - \varepsilon)$ accurate solution with a confidence $1 - n^{-\ell}$. Furthermore, TIM runs in

$$O\left((k + \ell)(m + n) \log n/\varepsilon^2\right)$$

compared to

$$O\left(kmnr\right)$$

,

where $n = \|V\|, m = \|E\|, k = \|S\|$ and r is the number of samples taken to estimate the increase of expected influence and ε is a constant related to the graph G and r. With this result, the authors show that influence maximization can be practical while still providing theoretical guarantees.

Other researchers have proposed the use of evolutionary algorithms to solve the influence maximization problem. Using a multi-objective evolutionary algorithm, Bucur *et al.* [5] have been able to find the seed set that not only maximizes the influence but also is the smallest. The results are promising as for intermediate values k the algorithm outperforms state-of-art heuristic algorithms, but yields mixed results for more extreme values of k.

The IC model, within the influence maximization domain, shows exactly the usefulness of a predictive model for information diffusion. Another prominent predictive model is the game theory model which is discussed next. As stated earlier, one of the things this project will try to investigate is the most influential nodes in the diffusion process of privacy norms. This related work will help with the project.

2.2.3 Game Theory model

In game theory, rational players play a game to maximize utility. Game theory is often used to model human behaviour, as is the case in information diffusion. To predict information diffusion processes, nodes in a social network are seen as rational agents that try to maximize their utility. There is an evolutionary component as agents can adjust their strategy over time. The prediction of the diffusion process is finished when the strategies of the agents reach an equilibrium. The environment is self-referential, one agent's choice to propagate or to not propagate information influences the other agents and thus the environment. The reader may recall that this property is essential discussed in the basics section.

Using the game theory, researchers have investigated the diffusion process given judgement schemes for users to imitate their neighbours' strategies, as documented in a paper by Li *et al.* [18]. The authors hypothesize that while imitating a neighbour's strategy to spread or not to spread information, a user does not trust everyone equally. Given this, they implement an evolutionary game that takes into account the strength of ties between different neighbours. Furthermore, they test the parameters that allow for punishing of agents deemed to be influential in the diffusion process. The researchers let the software run on different real-world data sets and find that giving the ability to judge neighbours to the agents can lead to diffusion processes that are easier to control. For example, they find that when an agent imitates exclusively weaker-tied neighbours or strongly-tied neighbours the spread of information can be limited sooner. On the other hand, when agents randomly select a neighbour to imitate the spread of information seems to be promoted. It is also found that punishing a node with higher degrees is the most effective method to limit the diffusion process.

The game theory model has also been used to predict future links between users in online microblogs, in this case, Twitter, as described in Liu *et al.* [20]. The story is one of rumours spreading in a network. The researchers assume network homophily, namely that nodes with the same attributes are more likely to become connected. They also assume that the closer two nodes are the more likely they are to become connected. A distance is defined that takes the structure of the network into account as well as the interactions between users, the social distance measure. The framework incorporates a coalition game, a game in which agents can collaborate to achieve higher payoffs against a possible competing coalition of agents. The game is played to maximize social-welfare and the result is a prediction of new links. The results of their experiments show the superior performance of the social distance measure compared to commonly used measures that only take the structure of the network into account. This result suggests that interactions between users may be important features to consider in prediction future network structure.

Competition between pieces of information is another phenomenon that the game theory model can make predictions on. In a study by Sun *et al.* [26], the authors investigate competitive information diffusion using a framework based on the coordination game. The game story is one of two companies that want to advertise for two competing products via viral marketing. The authors vary the attitude of the agents towards the two products on the network as well as the different network structures. To capture the attitude of agents towards the products, the researchers introduce measures for brand loyalty and self-perceived knowledge. These measures are taken as part of the utility function for each agent along with a third component, namely the popularity of the agent.

Giving each of these three component weights enables the authors to create personas: *experts, conformists and sworn followers.* The agents choose a neighbour's strategy to imitate using a probability based on the difference between two agents utilities in previous rounds. The authors compared diffusion processes in small-world and scale-free networks. Scale-free networks are characterized by a power law distribution of node degrees, while small-world networks are characterized by local node clusters with short path lengths between them [31]. From their results, they conclude that scale-free is the most efficient type of network for information diffusion. Equilibria are reached earlier in scale-free networks compared to small-world networks. They also find having a different ratio of different personalities leads to different diffusion processes. The more agents with higher self-perceived knowledge, the easier it is for information to diffuse and defeat competing information. That is to say, the more people on an OSN that perceive themselves to be knowledgeable about product A, the easier it is for information about product A to spread on said OSN.

The strength of these game theory models lies in their ability to incorporate behavioural aspects of users in the diffusion process predictions. In this work, however, analysis of the diffusion process is the focal point. Simulations are being run and conclusions are being drawn from the analysis of these simulations. Since there is no real prior knowledge of privacy preference diffusion from research, there is no real basis for a predictive model. In this project, we speculate about and explore diffusion processes. Furthermore, game theory models assume rationality in agents. There is no basis to say that rationality is also at the basis of our privacy preferences as these preferences are dynamic and change. These preferences may change as the network, time, relationships or various other factors change. It is also not certain that equilibria will be present i.e. that all agents will agree on strategies for privacy preference diffusion. To this end, the project focuses on the epidemic models from the information diffusion domain. These are explanatory models and are discussed in the subsequent sections.

2.3 Epidemic models

As suggested by the name, epidemic models were originally intended to model the spread of epidemics of infectious diseases. Their application in information diffusion follows from the analogy that on social networks users can be infected by the information spread by other users, leading to further spread of this information. They are seen as explanatory models as surveyed by Li *et al.* [19], but we will also see that prediction using epidemic models is not impossible.

In the basic epidemic model for networks, people can be (S)usceptible to infection or (I)nfected. An infection rate of λ represents the probability that a random susceptible user can become infected. At each time step t, $i(t)$ and $s(t)$ represent the infected and susceptible proportions of the population, at each time step $i(t) + s(t) = 1$ holds. N is the number of members of the population and this number does not change during the diffusion process. The diffusion process is then governed by the different equations of $i(t)$ and $s(t)$ with respect to t.

$$\frac{di}{dt} = \lambda i(1-i)$$
$$i(0) = i_0$$

The basic model is useful but very limited. It assumes that there is no transition out of the (I)nfected state, which is not realistic. To incorporate this, researchers introduced the SIS model. The SIS model introduces the possibility of transitioning from the (I)nfected state back to the (S)usceptible state. This model adds a cure rate parameter μ that represents the probability that a random infected user will be cured at any t. The change in dynamics are as follows; we now have to subtract, from the growth of the infected proportion, the cured proportion at each time step.

$$\frac{di}{dt} = \lambda i(1-i) - \mu i$$

The SIS model does not cover the concept of immunity, thus users can recover from disease only to be susceptible to the same disease again. This may be true for some diseases like the flu, but it certainly does not hold for all diseases. To incorporate the concept of immunity, the SIR model is introduced. The (R)emoved state represent immunity, there is no transition from this state to S or I. Now we have that a proportion of the population can be immune it follows that $i(t) + s(t) + r(t) = 1$, $r(t)$ represent the proportion of the population that is immune. Variable μ is now the rate at which infected people can become immune. The dynamics now also have to cover the increase of the decrease of susceptible people as $s(t) = 1 - i(t)$ no longer holds. This gives the following equations.

$$\frac{d\mathrm{s}}{dt} = -\lambda si$$
$$\frac{di}{dt} = \lambda i(1-i) - \mu i$$
$$\frac{dr}{dt} = \mu i$$

The R state can also be interpreted as a state in which an agent has recovered but is not necessarily immune to the infectious disease. This means that there is a probability α that an agent will move from the (R)ecovered state to the (S)usceptible state. The differential equations now exhibit the growth of $s(t)$ given α.

$$\frac{ds}{dt} = -\lambda si + \alpha r$$
$$\frac{di}{dt} = \lambda i(1-i) - \mu i$$
$$\frac{dr}{dt} = \mu i - \alpha r$$

These four models have been an inspiration for many variations of the epidemic models for information diffusion. Also, several properties of the epidemic models are discussed with examples and their relation to this project.

One example of the power of the SIR model is described in a study by Woo *et al.* [32]. In this work, the authors try to predict the spread of different topics on online for a given data from the earlier stages of the diffusion process. The researchers make a distinction between chatter and spiky topics. Chatter topics are topics that are part of the everyday exchange whereas spiky topics are topics that come about due to some sudden interest in them. The data sets used are crawled from the *Yahoo! Finance Walmart* message board and *US Politics Online Breaking News in Politics* political forum. The authors identify topics using a latent Dirichlet allocation model and clustering methods. To extract only the spiky topics, the researchers analyse the time-series patterns of their longitudinal data sets. The SIR model parameters are then estimated using a genetic algorithm. The SIR model is evaluated using the mean squared error (MSE) and R-square metrics. They find that for major topics the smallest R-squared value achieved is 0.43. This means that in the worst case, the SIR model explains 43% in the time series of the test data.

This is an example of a setting where the epidemic model can apply to prediction even with the basic SIR model. The next model discussed also performs predictions and, more importantly, shows the flexibility of epidemic models.

2.3.1 Flexibility

One such model, that illustrates the flexibility of epidemic models, is the SEIR model. The (E)exposed state captures agents that are exposed to the information diffusion in a diffusion process. From this state, they can go on to the (I)nfected state, diffuse the information themselves or become immune and not spread the information, (R)emoved. Researchers have used this model to investigate the effect of information value on the diffusion process, as documented by Xu *et al.* [33]. They introduce the S-SEIR model. The information value is captured by an audit function that decides on the value of information before exposure. The researchers also incorporate the ability for users to share information on different social network leading to increase people exposed to the information. They find that, in certain circumstances, that the higher the value of the information, the more people are exposed to it. They also find that the more people share on other platforms, the more susceptible users there are. As the researchers also note, their definition of information value needs more support from empirical research.

A variation on the SIR can be used to model the process of adaptation and abandonment of a social media platform from a study by Cannarella *et al.* [6]. Researchers introduce the notion of infectious recovery with the infectious recovery model irSIR. This notion is used to describe the downfall of MySpace. The researchers hypothesize that, in the case of social media platform adaptation and abandonment, people *recovering* and leaving a platform can be seen as influential behaviour on the people that still use the platform, the infected. This concept is reflected in the equations that describe the different states.

$$\frac{d\mathrm{s}}{dt} = -\lambda si$$

$$\frac{di}{dt} = \lambda i(1-i) - \mu ir$$

$$\frac{dr}{dt} = \mu ir$$

It is important to note that the infected proportion of the population decreases with a factor that depends on the proportion of the population that has become immune. This means that the higher the number of people that have left the platform the quicker the number of users on the platform decreases, thus leaving is now infectious.

To model abandonment, the authors use Google Trend data. They draw the analogy that the fewer people search for a social media platform on the Google search engine the less prominence this platform has. Their assumption is justified by the data found on MySpace searches. The social media platform died down between 2009 and 2011. They evaluate the irSIR against the original SIR model. They find a decrease of 75% in the sum of squared error (SSE) when going from the SIR to the irSIR model. The irSIR provides a better fit under this metric. The researchers then try to predict the demise of Facebook and find that the irSIR predicts Facebook will disappear between 2016 and 2020, most likely in 2017. As of writing, it is 2021 and Facebook is still one of the most popular social networks. Of course, this can be due to numerous factors. For example, Facebook is no longer only a single social network but has also acquired other successful social networks such as WhatsApp and Instagram. Facebook has also heavily invested in the development of artificial intelligence applications.

This study shows the malleability of epidemic models. In essence, the authors have adopted an epidemic model to perform time-series fitting and forecasting. This ability to adapt and tweak is valuable when entering a completely new field as is the case with this project.

2.3.2 Resilience of epidemics

Researchers have tried to understand the effect of interventions on the diffusion process. In a paper by Lu *et al.* [21], the authors investigate what the effect of a change in infection rate has on the diffusion process under the SIS model. To do this, a control stage is applied during the diffusion process with a second infection rate, λ_1, λ_2. The infection rate λ_2 is applied during the control stage before reverting to λ_1. The setting, in general, is that of $\lambda_2 < \lambda_1$. The network structure, infection rate settings, the start time of the intervention ct and the duration of the control stage cd are varied. Generated Erdős–Rényi networks and scale-free networks are used as well as three real-world networks. An Erdős–Rényi network is a random network that has a binomial degree distribution.

From the results of various simulations, the authors show that even after a control stage an epidemic can spread as much as before the control stage, thus resilience is possible. They show that it is possible to derive the critical value for the control duration cd_{max} to ensure the epidemic does not survive. This critical control duration value is also found to be different for the various network structures. This indicates to the authors that the network structure determines cd_{max}, specifically the authors derive that cd_{max} is related to the diameter of the network $cd_{\max} \sim d^{\alpha}$. Through analysis, they also define the probability of resilience given ct and cd as follows.

$$P = \begin{cases} 0, & \rho(ct + cd) \leq \frac{1}{N} \\ 1, & \rho(ct + cd) > \frac{1}{N} \end{cases}$$

These are the keys findings for this project. As stated before, for system administrators not all privacy norms are desirable. To mitigate such norms, an administrator would likely perform an intervention. The administrator's goal is to derive the appropriate ct and cd to ensure that $P = 0$, according to the formula above. To this end, the administrator needs to find cd_{max}, since any $cd \geq cd_{max}$ ensures $P = 0$. This way, in theory, there is no chance for a second epidemic.

2.3.3 Attention and Competition

Information on a social network comes in vast quantities and vastly different types. Thus, researchers have argued that the accuracy of epidemic models in explaining diffusion processes cannot always be guaranteed, as documented by Feng *et al.* [8]. The authors find that using a data set from micro-blog site Sina Weibo, the average number of neighbours infected before a node gets infected k is less than or equal to 2. This is in contrast to what they find using the SIR model to model epidemic diseases in scale-free networks. They conclude that there is a key difference that pertains to the attention of the users on social networks. They hypothesize that users on social networks pay limited attention to their neighbours and that the higher the number of neighbours the less attention a user pays. In other words, the neighbours compete for attention and the more neighbours the user has, the less attention can be paid to all.

To capture this notion, the authors adopt an infection rate that depends on the degree of a node, γ/k where γ is the infection rate and k the degree. The method is called the fractional SIR model, FSIR. FSIR also incorporates guaranteed immunization after τ time steps. Validation on the Sina Weibo data set shows the superiority of the FSIR model against the SIR model. They also define a critical threshold below which information can produce an outbreak but no epidemic. The authors link this threshold to the value of the information. Thus, when different pieces of information compete for attention, those with a high value produces an epidemic while other pieces only get briefly shared in high volume.

These findings indicate that it is not always the case that nodes with a high degree have the most impact in all diffusion processes. Of course, when a node with a high degree shares information they increase the number of susceptible users significantly. However, the choice to share depends on the value of the information given competing information as well.

2.3.4 Source Detection

An important part of the analysis of an information diffusion process is the detection of the source. The importance of source detection is motivated by the spread of rumours, information that is not strictly true. The rise of *fake news* has made it more difficult for people to find the ground truth about several subjects online. In 2020, during the COVID-19 pandemic, a rumour was spread about how the existence of the COVID-19 was directly related to the rise of

5G networking technology in the Netherlands[3]. This led to several radio towers being set on fire by people that believed this rumour. Understanding where the rumour started can provide government officials with a source to validate information or in the case that the information is false, someone to hold on accountable for the damages caused by the rumour.

Research around source detection has been covered in a survey by Shelke and Attar [24]. In information diffusion research, there are general key steps in the process of source detection.

1. Gathering data on information being shared between users of an OSN
2. Pre-processing the data in question
3. Constructing a network from the data
4. Identifying an appropriate diffusion model and evaluation metrics
5. Applying models and metric to classify sources
6. Validation and post-analysis of results

As we have already discussed some relevant diffusion models, this is not needed exclusively. The key steps to understand are those about evaluation metrics and classification. Classification in source detection can have two goals; identifying one single source or multiple sources. The key feature used in classification is the centrality of a node on the network. Centrality is one of the main structural properties of a network that is used to quantify node influence. Centrality measures come in different flavours: degree centrality, betweenness centrality, eigenvector centrality, PageRank and so on [19]. To define the performance of a classifier, four main metrics are used: accuracy, rank, distance error and the time taken to find the sources. The F-score is generally used to define accuracy.

$$\text{F-score} = \frac{2 \times \text{precision} \ \times \ \text{recall}}{\text{precision} \ + \ \text{recall}}$$

The precision measures defines what ratio of the actual sources are in the set of source found by the classifier over the all classified sources. The recall measure defines what ratio of the actual sources are in the set of source found by the classifier over the set of actual sources in the diffusion process.

$$\text{precision} = \frac{|\{ \text{ classified sources } \} \cap \{ \text{ actual sources } \}|}{|\{ \text{ classified sources } \}|}$$

$$\text{recall} = \frac{| \text{ classified sources } \} \cap \{ \text{ actual sources } \}|}{|\{ \text{ actual sources } \}|}$$

The rank measure is the index of the actual source(s) in a sorted list of nodes, that is descending concerning the score of the node. The distance error refers to the shortest distance between the classified source and the actual source on the network.

In Shelke and Attar's survey [24], the techniques used to classify are categorized by network topology and diffusion model. In the case of this project,

[3] https://eenvandaag.avrotros.nl/item/extra-bewaking-moet-betreurenswaardige-en-bizarre-branden-5g-masten-voorkomen/

the diffusion models are the epidemic and, as we'll see in the subsequent sections, the network topology is generic, as it is a graph and not a tree. The data set used is a snapshot of a social network. For these settings, the existing approaches use the rumour centrality metric to classify sources. The rumour centrality is described as the number of diffusion paths starting from the origin node. The higher the rumour centrality of a node, the more likely it is to be the source of the information.

All the work presented here shows the maturity of research on information diffusion modelling using epidemic models. This is the maturity that also enables the exploration using these models to model the spread of privacy preferences.

2.4 Data

To guide the simulations in this work, real world data from OSNs is used. For a data set to be useful for privacy preference diffusion modelling, it should contain a graph of a real life social network. Furthermore, the data set should capture user attributes like; age, content preferences among other properties. Finally, the data set should contain content being shared on the network. No data set was found that had all these features. The following data sources were considered for use in privacy preference diffusion modelling.

2.4.1 Semantically Analysed Metadata of Tumblr Posts and Bloggers

The *Semantically Analysed Metadata of Tumblr Posts and Bloggers* [2] data set is an SQL dump consisting of 6 tables that describe 2224 Tumblr users. The tables cover features of a blogger such as the number of posts they have liked, the number of posts made by the blogger, title and description of their blog as well as if they allow ask interactions on their blog. Information about a blogger's posts comes in the form of the number of notes, timestamp of the post, number of tags, title and more. Using semantic analysis of the posts, the creators of the data set also provide the tone and topic classification as well as sentiment score for each post. This data set is rich in information about the content shared by users. It lacks, however, a social network graph. It is unclear whether the bloggers are connected. There is a column in the posts table that define where a post is re-blogged from, but inspection showed that this column was always empty. If the re-blogged relationship were present, it would be possible to infer a (partial) network from this data set.

2.4.2 Stack Overflow temporal network

The *Stack Overflow temporal network* data set is part of the SNAP repository and features multiple temporal networks that illustrate answer-to-questions, comment-to-question and comment-to-answer interactions of the question and answering website Stack Overflow. Programmers can ask other programmers questions concerning anything to do with programming. This data set could be useful to model the change in the interaction inside a network over time. This requirement is, however, secondary to having rich content information. THe data set shows the existence of one of three types of interaction at timestamp but no details of the content being shared at that point.

2.4.3 Interaction-Based Behavioural Analysis in Twitter Social Network

This data set was collected via the Twitter API to use in the classification of user types in different categories, namely popular-active, observer-passive, and spam-bot-malicious [1]. It provides content shared by users over time. Using the data set, a network could be inferred using the retweet and mention interactions captured in the tweet text. To use this information, the project would have to include a plan to derive information from these plain text tweets. As an exploration of a new field of privacy decisions diffusion, this is beyond the scope of this study.

2.4.4 Stanford Network Analysis Project

The Stanford network analysis project (SNAP) [17] provides various networks for network analysis purposed. The repository also contains social networks that were considered for this project. Most of the networks are network graphs with users as nodes. Edges exist between two users if there is some form of interaction between them. The interactions range mutual friends and following relationships to user actions on a massive open online course (MOOC) and administratorship votes on Wikipedia. A drawback of these data sets is their lack of node attributes.

2.4.5 Anonymized Instagram network data from Amsterdam and Copenhagen

Another candidate data set was the *Anonymized Instagram network data from Amsterdam and Copenhagen* data set [4]. It captures the mutual liking and commenting on Instagram among users in Amsterdam and Copenhagen. It could be argued that mutual likes and comments capture what two users like. However, in the end, other candidate data sets provided higher resolution data to derive user information from.

2.4.6 An exploration of the Facebook social networks of smokers and non-smokers

This data set is from a study by Jacobs *et al.* [9]. This study examined the social network of Facebook users that smoke versus social networks of users that do not smoke. The authors found that the network structures for the two group differed significantly on various properties. This data set could provide data about the user's smoking habit, age, likes count, country, gender and wall post count. Furthermore, since the study investigated different partial networks for each (family, groups, photos and friend), the data capture different properties of these networks such as; modularity, communities, clusters, diameter, isolates, transitivity and more. The biggest downside to this data set is the lack of a network and interactions between users with which to derive a network.

2.4.7 Dynamics of Instagram Users

This data set was originally used to model the dynamic of users on Instagram [1]. It captures several features of approximately 1000 Instagram users. The data set captures number of posts, number of followers, number of followings, number of self-presenting posts from nine previous posts and gender. Furthermore, the number of likes is captured for the tenth, eleventh and twelfth previous posts. This data set does lack a network and any interactions between users.

2.5 Trust in multi-agent systems

Trust modelling is multi-agent systems is used to model the faith agents have in other agents' abilities to complete certain tasks successfully. Trust modelling can be one-dimensional or multi-dimensional. Both are useful for privacy preference diffusion. Here, we take a look at two key approaches in multi-dimensional modelling.

In a study by Griffiths [10], the first description of an experience-based multidimensional trust model is given. This enables agents in multi-agent systems to reason more accurately about tasks in delegating trust. For example, imagine an agent delegating a task and receiving feedback upon completion of the task. It may be the case that although the task was successfully completed it cost more than expected or took longer than expected to complete. Cost and duration are dimensions for which the agent might want to model the trust of agents that have served them.

In a paper by Reece *et al.* [22], this idea is extended to allow for correlation between the different trust dimension. The authors show that, while using a separate beta distribution for all dimensions can be useful, information of the interactions between the dimensions is mostly lost. To mitigate this, they propose the multi dimensional extension of the beta distribution, Dirichlet distribution, to model trust. This results in better information capture for the interactions between the trust dimensions. Furthermore, the authors propose a method to decentralize the storage of reputation based on trust. This allows for more accurate sharing of trust information between agents.

In the following sections, the settings of the simulations with the DIPP model will be made preciser. The choice of an epidemic model and states will be motivated. Furthermore, the goals of the project are made precise by defining them in terms of measurable quantities.

3 Modelling privacy preference diffusion as information diffusion

In order to model the diffusion of privacy preferences, this work models an online social network (OSN) as a multi-agent system. The agent in this system will represent users of an OSN. The agents will be able to share content, mimicking the content sharing by human users on OSNs. On OSNs, content can be defined by the context it portrays.

When a piece of content on an OSN is observed, us humans can perceive different properties that describe the content. These properties may describe something worthy of our public appreciation in the form of a *like*. However, the properties could also describe something we do not like to perceive or do not care about either way. Imagine a photo of someone taking a self-portrait photo of a user at the pet zoo with a llama in the morning. The description of the scene of the photo captures its *context*. This context defines location (i.e., zoo), people in the picture (i.e., user), what is happening in the picture (i.e., posing with a llama) and the time of day (i.e., the morning). For an intelligent agent, this description can be determined using image caption models that take images and provide a description with words, e.g. using work by Vinyals *et al.*[30], or by user manually documenting the properties. On an OSN, a user can come across this picture and like it, of course. However, the user could be the lady in the picture, who didn't know that this picture was shared without her explicit consent. Does she approve of this content of hers being shared in this context? Whether she approves or not, where does her opinion come from? This project investigates how this opinion can come about as an infection from interacting with other users in the OSN.

In this work, the context of the content shared by a user on an OSN is assumed to reveal the privacy preferences of that user on said OSN. This follows from the assumption that a user only shares content that they want to be seen by their friends on an OSN. It could be argued that users can be moved to share content by external factors [3]. However, in the end, it is always the user's decision whether to share content.

Definition 3.1 (Privacy preference)**.** A context description that can be ascribed to the content shared by an agent with the assumption that agents only share content they do not find to be private.

A difficult point of describing the context of content is resolution, since the shared content exhibits context describable by various properties. There are many details that can describe content leading to many fine-grained inferred privacy preferences. This challenge is beyond the scope of this project. The focus of this project is to propose a novel framework that can describe the spread of privacy preferences. Dealing with the resolution of the context in the content shared on an OSN can happen later and does not limit the DIPP model. The resolution of the context used to describe content will become clear in the latter sections.

Example: Alice and Bob are university students and friends on an OSN. During the Christmas period, Bob shares pictures of his family's Christmas party while Alice does not. After observing Bob's Christmas posts, Alice infers that Bob is comfortable sharing content containing family members, at family parties. This is one of his privacy preferences, the combination of the context of the content and the fact that Bob shares this content.

The problem with context resolution is that one can also ask: should Alice infer that Bob is particularly comfortable sharing content in the shirt he is wearing in all the Christmas posts? This detail might matter towards defining a privacy preference, or it might not. However, answering such questions is not an imperative for the present project and the framework it proposes.

Given the content shared by a user, we can define the agent's privacy preferences as the different contexts revealed by the content they have shared. It should be clear that this definition might not be exhaustive, as the user could hold privacy preferences that are hidden to other users on the OSN. It could be the case that Alice would be willing to share video clips of her dance rehearsals but has just not had the opportunity. Alice's friends on an OSN can not infer that sharing video clips of dance rehearsals is a privacy preference of hers. Furthermore, a user can also hold privacy preferences that represent content that they believe should never be shared, thus these are *opposed privacy preferences*.

The DIPP model takes into account the opposed privacy preferences in two ways. In one setting, it is assumed that the opposed privacy preferences spread similarly to the supported privacy preferences. In a second setting, the opposed privacy preferences are assumed to be static over time. It is highly probable that both these assumptions are inaccurate, however, as the opposed privacy preferences are not derivable from content, we opt for these two. The first setting captures the fact that there is a dynamic to the opposed privacy preferences and the second one captures makes no assumption on the dynamic covering the hidden feature of the preferences. Although this project does not aim to contribute to the accurate modelling of opposed privacy preferences, they are paramount in capturing privacy violations.

Privacy violations in OSNs occur when content reveals information about a user that the user would want to keep private. Opposed privacy preferences can be deemed to describe content that a user believes should be private. Privacy violations can occur when a user shares content that is co-owned by another user and, in doing so, reveals private information of said user. Content is co-owned when it contains information that can be traced back to more than one person. For example, a group picture is co-owned by everyone in the group in the picture, even though one person takes and stores the picture. The sharing of such a picture, we theorize, makes the co-owners more likely to be infected with privacy preference exhibited by the piece of content. This could be the case if a co-owner has never considered sharing such content and does not mind it being shared. Furthermore, it could be the case that a co-owner notices other users engaging in more positive interactions with their content of the same type. This may cause the co-owner to go along with a similar privacy preference even if they originally opposed it.

Using privacy violations, we can model trust between agents on an OSN. When Alice's privacy is violated by Bob on an OSN, this should affect the relationship between the two agents, at least concerning the OSN. This could manifest itself as Alice becoming less likely to agree to share co-owned content with Bob. Alice could even decide to not make content with Bob in the future altogether. All this could happen because Bob has damaged the trust between Alice and him. The notion of trust will be part of the model created in this project. Using the model, we will investigate if the use of trust values can help agents protect themselves from opposed privacy preferences.

These are the underpinning of this research; privacy preferences are infectious via the medium of the content sharing habits of friends. Users perceive friends sharing certain content and are deemed more likely to adapt the same habits. In this section we have outlined the global theories on how these infections come about and how they could be influenced.

Example: Alice shares a selfie of Bob and her after a night out without consulting Bob. Bob, however, prefers to keep his party life away from social media, thus Alice has violated Bob's privacy. In the future, Bob might trust Alice less with content they both own.

Definition 3.2 (Co-owned content)**.** Content is co-owned when it represents more than one user. Representation could be by appearance in a picture, sound of a voice in a video or audio recording, explicit tagging of users and more.

Definition 3.3 (Privacy violation)**.** A privacy violation in an OSN is an instance in which content is shared without the explicit consent of a user. Furthermore, this content is in a context that the user opposes sharing content of.

The next sections present a model that simulates the spread of privacy preferences and the varying factors that are believed to affect this dynamic in OSNs. Opposed privacy preferences, co-ownership, privacy violations and trust are believed to impact the diffusion of privacy preferences in reality. The model presented will incorporate these factors. As previously stated, epidemic models will be used to model privacy preference diffusion. In original epidemic models, an infectious disease is the infectious entity spreading in a network. In the case of this research, the infectious disease is substituted for a privacy preference. This substitution comes about after the review of various epidemic models used for modelling information diffusion. In line with information diffusion models, we believe privacy preferences can be infectious. Users are exposed to the infectious entities through observing content. Users can become infected due to various factors. In this work, we consider opposed privacy preferences, co-ownership, privacy violations and trust as factors. The following sections will elaborate on the different circumstances under which the diffusion processes will be investigated.

3.1 Context of content shared

Context, in this study, is described using three locations and four times of day. The four times of day are *morning, afternoon, evening* and *night.* The three locations are at *work, the beach* and *the mall.* The contexts are represented as 2-tuples with 12 possible combinations, for example: $< night, work >$. Given these definitions, a user has a certain privacy preference if they are infected with the content described by these tuples. A privacy decision is then a binary decision representing whether to share or not to share content that is represented by the context tuple. In reality, a privacy preference is often unknown to fellow social network users and only sometimes known to the social network administrator. Privacy preferences are often private, thus users' friends rarely know them. Some OSNs provide users with tools to protect their privacy on OSN that involve the user providing some description of their privacy preferences. One of the few ways for users to have a sense of the privacy preferences of others is through perception of the privacy decisions others make.

It should be noted that these 12 variable values are easily interchangeable as the semantics they carry are of no relevance in the model developed here. These 12 values were chosen as they provide clear descriptions of contexts in two clear dimensions, location and time of day. There was no need to add more values.

Furthermore, concerns about the computational performance of the model made us choose not to include more values. One could consider a model for diffusion of privacy preferences in which the semantics of these values are part of the equation, but that is beyond the scope of this project.

Definition 3.4 (Privacy decision)**.** A binary decision representing whether to share or not to share content that is represented by a context description.

3.2 SIR model for privacy preference diffusion

The SIR model is an epidemic model that is governed by an infection rate and a recovery rate. The agents, part of the model, can be in three states; (S)usceptible, (I)nfected and (R)ecovered. When an agent is in the susceptible state, it is susceptible to the infectious entity that is spreading. When the infectious entity infects an agent, the agent moves to the infected state. If and when an agent recovers from this infectious entity, the agent is in the recovered state. Being in the recovered state is synonymous with being immune to the infectious entity as there is no state transition out of the recovered state. The state dynamics of the SIR model are governed by an infection rate i and recovery rate r. The dynamics are visualized in Figure 1.

The infectious entity in the SIR model was originally any disease that would spread among humans. However, researchers noted that information could be seen as infectious as well. Thus, the SIR model was adopted to model information diffusion [19]. In work by John and Joshua [6], the researchers even model adoption and abandonment of OSNs using the SIR model.

To model privacy preference diffusion, we will use the SIR model. The motivation behind this is best explained looking at the different states of the SIR model. Concerning the (S)usceptible state, agents are deemed susceptible to imitate certain behaviour if they witness it from their neighbours. In this work, we assume that the same holds in the context of privacy preferences. Users on OSNs sees other users share content that depict certain privacy preferences. They can then be infected if they have never thought of sharing such content and willing to do so, thus adopting the privacy preference. They could also be infected because their friends start adopting certain privacy preferences that they do not hold.

If an agent observes another agent, to whom they are connected, take a certain privacy decision within a context, they will be susceptible to become (I)nfected by that behaviour and thus imitating it and adapting that privacy preference. This will be reflected by an infection rate for that privacy preference.

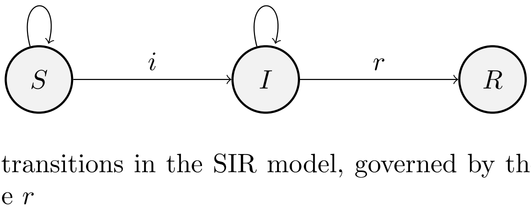

Figure 1: State transitions in the SIR model, governed by the infection rate i and recovery rate r

The last state of the SIR model is the (R)esistant state. Agents in this state have recovered from an infection and are immune to a privacy preference, thus they will not share content aligned with said privacy decision. From these models, the key measurements are population proportions for each state. This is common practice in diffusion modelling as these measurements allow for a relatively complete analysis of the diffusion process. The willingness of an agent to share content that represents a certain context is depicted in Table 3.2. The state determines the willingness of an agent to share content described by a certain context. Note that the states are in line with the SIR epidemic model, the infectious entity is the willingness of an agent to share certain content. An agent is considered infected when they are willing to share content in that context. In the susceptible state, an agent can become willing to share that type of content as they are vulnerable for infection. The resistant state has no agents that are willing to share content represented by that type of context. These state transitions are governed, mainly, by the infected rate i and recovery rate r and can be seen in Figure 1.

This initial description of the epidemic model used is similar to epidemic models used previously in information diffusion research. The main difference is that this model carries multiple infectious entities rather than the common one or two. To make the model more accurate, in terms of modelling the spread of privacy decisions, concepts such as opposing infectious entities and co-ownership of content are added.

Context	**State**
<Morning, Work>	INFECTED
<Morning, Mall>	SUSCEPTIBLE
<Morning, Beach>	INFECTED
<Afternoon, Work>	INFECTED
<Afternoon, Mall>	SUSCEPTIBLE
<Afternoon, Beach>	SUSCEPTIBLE
<Evening, Work>	INFECTED
<Evening, Mall>	SUSCEPTIBLE
<Evening, Beach>	INFECTED
<Night, Work>	INFECTED
<Night, Mall>	SUSCEPTIBLE
<Night, Beach>	INFECTED

Table 1: State representation of an agent's willingness to share content that represents a certain context

3.3 Opposing privacy preferences

The model described above carries information on an agent's willingness to share content that describes a certain context. However, one can imagine social network users having personal convictions regarding, not only what constitutes a right privacy preference, but also what constitutes a wrong one. Thus, the DIPP model also carries information on an agent's opposition towards certain privacy preferences. The representation is the same as in Table 3.2. It follows that, instead of 12 infectious entities, an agent keeps track of 24 simultaneous ones. Consequently, there are co-occurring states, *pro* and *opposing*. This can lead to issues when, for example, an agent is infected in both the pro and opposing settings of a context. However, in reality, a social network user can only be in favour of one of the two settings. If Bob believes photos taken at the beach should stay private, he cannot share photos of himself at the beach. Since, otherwise, that would be contradicting. Furthermore, it would also be impossible for Bob to note privacy violations if he supports and opposes this privacy preference at the same time.

Definition 3.5 (Pro side)**.** The epidemic of privacy preferences that are supported by agents. These are the privacy preferences that underlie the infected agents' content sharing habits.

Definition 3.6 (Anti side)**.** The epidemic of privacy preferences that are opposed by agents. Infected agents on this side believe that the content, described by the privacy preference they are infected with, should not be shared.

Table 2 shows how the two sides' states can co-occur in the DIPP model. The only combined state that is impossible is one where an agent is infected on both the pro and opposing sides. The rest of the combined states are deemed possible because of the following explanations.

- **Infected and Susceptible** An agent can be infected on one side and susceptible on the other because it is not impossible for someone to oppose content sharing behaviour and slowly gravitate to it to the point of crossing over and sharing the type of content themselves. For example, with the rise of selfies in the early 2010s, there are bound to be people who found this type of content to be inappropriate, but as the masses continued sharing said content the opposing opinions dwindled resulting in (even) more selfies being shared.

  How an agent can come to such a crossover point is an interesting question. In this work, we investigate trust as a contributing factor.

- **Infected and Resistant** In this combined state, the agent has become immune to one side and is infected by the other side. This is a plausible state as one can imagine a social network user so convinced that sharing pictures of exam results is not appropriate that they would never even consider the possibility of sharing such content despite their peers sharing their results on the social network. One could also consider the possibility of immune agents becoming susceptible once again with a small probability, but this first iteration of the DIPP model does not cover this factor.

- **Susceptible and Resistant** In this combined state, the agent has become immune to one side and is susceptible on the other side. This is a plausible state as an agent could be totally opposed to sharing a type of content and thus be in the resistant state on the pro side. However, this opposition means the agent is also susceptible to becoming an actively opposing agent by becoming infected on the opposing side. One can imagine someone never sharing revenge pornography on a social network who, as time goes by, might actively oppose such content and advocate for measures to be taken when such content shared.

- **Susceptible and Susceptible** In this combined state, the agent has no opinion on this specific type of content. This state could represent a person that has never encountered the type of content before. Thus, when they perceive the content, there is a possibility they might agree with the sharing or disagree.
- **Resistant and Resistant** In this combined state, the agent is immune to both side of the type of content being shared. This means there are indifferent to this type of content, but will also never share such content.
- **Influence on state dynamics** The combined states should clearly have influence on the way agents change states. It should be harder to infect someone if their infected with the opposite opinion towards a type of content. To capture this, the infection rate i is multiplied by a random number

  $$0.5 > h > 0.1$$

  that represents an agent's resilience to someone trying to change their opinion. From this it follows that the infection rate $i = i * h$ when the agent we are trying to infect is infected by the opposing opinion. Firstly, h is a random number to capture the fact that not all agents in a social network have the same resilience in terms of changing their opinion and this resilience is also not static over time. Secondly, the choice for the range captures, in this scenario, that the infectious agent is only half as potent at most and one-tenth times as potent at least. The state dynamics given opposing privacy preferences are visualized in Figure 2.

  The other combined states have no influence on the state dynamics in the DIPP model. There are probably ways in which these states could influence the dynamics, but those are beyond the scope of this work. In the next section, trust in this agent-based model is discussed.

Pro state	**Opposing state**	**Possibility**
INFECTED	INFECTED	Impossible
INFECTED	SUSCEPTIBLE	Possible
INFECTED	RESISTANT	Possible
SUSCEPTIBLE	INFECTED	Possible
SUSCEPTIBLE	SUSCEPTIBLE	Possible
SUSCEPTIBLE	RESISTANT	Possible
RESISTANT	INFECTED	Possible
RESISTANT	SUSCEPTIBLE	Possible
RESISTANT	RESISTANT	Possible

Table 2: Possible states for agents to be in with regard to being for or against sharing content described by context

3.4 Trust

In multi-agent systems, trust is used as a way for agents to quantify the faith they have in the fact that other agents can successfully complete a certain task.

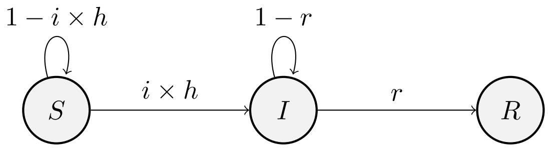

Figure 2: State transitions in the SIR model, governed by the infection rate i and recovery rate r with h in the case of trying to change agent's established opinion

When Alice takes the bus to go to campus, she trusts the bus driver to take her to her destination safely and timely. A model that simulates whatever aspects of using public transport, could incorporate trust between passengers and the people that steer the vehicles they travel with. Trust can be one dimensional, but that is not necessary. Consider bus driver John. Alice knows that John drives her 8 am bus to campus on Mondays and Thursday. Alice always gets to campus safely on Mondays and Thursdays, thus she trusts John to take her safely to her destination. However, John enjoys speaking with passengers a lot which, in turn, means that Alice is always 5 minutes late to campus even though she caught the bus at the right time. So, Alice does not trust John as much when it comes to taking her to her destination in a timely manner.

Trust is a part of the DIPP model because it has been found to be an important factor in making privacy decisions. Research by Lampinen *et al.* [16], shows that privacy management in content sharing on online social networks is mostly based on trust of neighbours inside the network.

The DIPP model implements one-dimensional trust modelling. One-dimensional trust refers to an agent's faith in another agent to complete a task with one measurement of success [22]. In a one-dimensional trust model on public transport safety, Alice would trust John as her experiences with John have always been safe. However, if we make the model two-dimensional by adding the measurement of success lateness, Alice would not trust John as much as in the one-dimensional model before.

In the DIPP model, trust is based on interactions between agents of which they are two: perceiving another agent's shared content and being a co-owner of the content shared by another agent. These two interactions can lead to privacy violations that negatively influence the trust value one agent has towards another agent. Firstly, when an agent perceives a neighbour exhibiting behaviour consistent with a privacy preference they are opposed to, that should cause that agent to negatively update their trust towards said neighbour. In other words, the agent trusts their neighbour less because they do not agree with the privacy preference the neighbour exhibits. For example, let us think of two university students, Alice and Bob, who are friends on Facebook. Alice sees Bob posting a photo in which he is pictured kissing Angela at a party. Alice would never post a picture of herself kissing someone at a party. This leads Alice to distrust Bob's privacy management, unbeknown to Bob.

A second interaction, which causes an agent to trust a friend less, is when the agent's friend decides to share content with a context that the agent opposes. For example, imagine that Alice also believes that content captured at the beach should stay private. If Bob decides to share last June's photo of Alice and him at the beach then Alice will not only be upset, but she will also trust Bob less.

The reader should also notice that there seems to be a clear difference of the severity between the two negative interactions. When Bob shares the picture of Alice and him at the beach, Alice is affected directly. In the first interaction, however, Angela is the one directly affected, whether the effect is negative or positive depends on Angela's privacy preferences. The severity of the two types of privacy violations will be considered in the DIPP model.

To model trust, the beta distribution is used. This probability distribution is governed by two variables α, β. Given that the outcome of each interaction can only be one without a privacy violation or one with a privacy violation, we have binary outcomes. The beta distribution is well suited to model the uncertainty over

$$p_{mr}(o = 1)$$

, where o represents a positive outcome, which is the probability Alice holds for a positive interaction with Bob. The actual distribution can never be known, thus Alice has to estimate

$$\hat{p}_{mr}(o = 1)$$

, given the finite experience she has in interacting with Bob on the social network. The two parameters of the beta distribution a, b can be used to represent positive outcomes and negative outcomes respectively. The expectation of the beta distribution

$$\mathrm{E}[X] = \frac{\alpha}{\alpha + \beta} = \hat{p}_{mr}(o = 1)$$

then provides Alice with an estimate of the likelihood of positive interaction with Bob or, in other words, of her trust in Bob not to violate her privacy. In practice, this amounts to

$$\hat{p}(o = 1) = \frac{n + 1}{N + 2}$$

[22] being the trust value every agent holds for each of their neighbours. In this formula, n represents the number of interactions without privacy violation and N the total number of interactions. We will see that N amounts to the current number of steps in a simulation as each agent shares content at each step and every friend of an agent perceives said content. As previously stated, the severity of a violation differs between the two possible types. To account for this, the violations are assigned levels. Level 1 is the violation where an agent perceives behaviour they do not agree with and level 2 is the direct violation of an agent's trust by sharing co-owned content. Given this, the trust value formula for level 1 violation comes to be

$$tr_1 = \frac{n + 2}{N - n_2 + 2}$$

, where n_2 is the number of level 2 violations, conversely this means

$$tr_2 = \frac{n + 2}{N - n_1 + 2}$$

where n_1 is the number of level 1 violations and $n = N - n_1 - n_2$. Finally, these two trust values are integrated into one with weights agent uses to express what they see as more severe

$$tr = s_1 \times tr_1 + s_2 \times tr_2$$

with $s_1 + s_2 = 1$.

Definition 3.7 (Level 1 privacy violation)**.** A level 1 violation occurs when an agent perceives another agent sharing content that portrays a privacy preference that they oppose.

Definition 3.8 (Level 2 privacy violation)**.** A level 2 violation occurs when, for example, Bob shares content that is co-owned by Alice without Alice's consent.

The trust value directly influences the infection rate as can be seen in Figure 3. In this project, the severity values are static over all simulations.

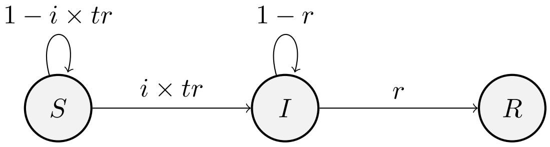

Figure 3: State transitions in the SIR model, governed by the infection rate i and recovery rate r with tr representing the trust in the neighbour trying to infect the agent

It can be noted that trust is modelled, in the DIPP model, with a keen focus on its negative effect on interactions between agents. This is because of the focus on investigating whether agents can protect themselves against opposing privacy preferences using trust modelling. In multi-agent systems, trust is a factor that can also strengthen relationships between agents, as is true for the trust model above as well. However, this aspect is not the focus here and is, thus, no investigated.

3.5 Co-ownership

To add more realism, the concept of co-ownership is also incorporated. Co-ownership is the ideal that content is sometimes owned by multiple people even though only one person makes the choice to share the content. A group photo is a form of content with co-ownership. All the people in the photo stand to gain or lose, with regard to privacy, whenever the content is shared by one of them. In the DIPP model, co-owners are assigned at random with at most 5 co-owners for each piece of content shared.

It is clear that an agent should be affected differently when they are co-owners of the content being shared by a neighbour than when they are passively witnessing the content. This concept manifests itself in the model as another factor that alters the infection rate. In this case,

$$1.5 > co > 1.1$$

is a random number that represents how impressionable an agent is, with regard to sharing content they have doubts over. As with h, co is a random number to capture the fact that not all agents in a social network can be persuaded on the same level and this resilience is also not static over time. Furthermore, the choice for the range captures, in this scenario, the fact that the infectious agent is 50% more potent at most and 10% more potent at least. The state dynamics given content co-ownership are visualized in Figure 4.

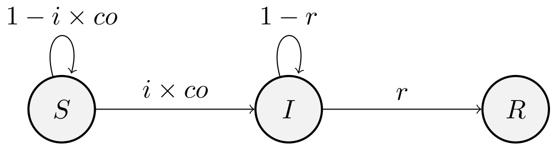

Figure 4: State transitions in the SIR model, governed by the infection rate i and recovery rate r with co in the case of an infection attempt via co-ownership

4 Agent-based modelling of privacy preference diffusion

In this section, DIPP, the agent-based model is formally defined. Firstly, we consider the agents, the various decisions they can take and how these decisions come about. Secondly, the aspects of the model that do not concern the agent are described. The entire model is written on top of the Mesa library for Python 3[4]. The time steps are discrete.

4.1 Agent

In the DIPP model, an agent represents a user on an online social network that shares content from which other agents can derive said agent's privacy preferences. To this end, each agent must have the ability to share content and all their friends should be able to perceive the content. This section covers the details of how these actions are taken by agents. A schematic view of an agent in the DIPP model can be seen in Figure 5.

4.1.1 Agent initialization

At the time of instantiating the simulation, each agent is assigned privacy preferences on the pro and anti sides. As stated previously, for each privacy preference, an agent can be in three states: susceptible, infected and resistant. In the DIPP model, that states are randomly generated. For each privacy context, the agent is randomly assigned either the state susceptible or infected, with a probability of 0.5 of both, as can be seen in the loop from line 5 in code block 1. The state resistant is ignored here as it is a final state and would thus limit the dynamic in the experiments set out. If an agent starts out in recovered, then the agent is effectively removed from the experiment. Since the aim is to investigate the state dynamics, it makes no sense to limit these dynamics beforehand.

For the epidemic dynamic to work, infection rates and recovery rates have to also be assigned for each of the privacy preferences. These can be single numeric values that hold for either pro or anti side. However, the rates can also be defined as a mapping between a context and a numeric value, allowing for precise custom assignment of rates for any experimental setting. Furthermore, assigning the infection and recovery rates at the level of an agent means that a researcher can predefine different rates for different agents if an experimental setting requires this. An impression of this assignment process can be seen in code block 2.

[4] https://mesa.readthedocs.io/en/master/overview.html

Algorithm 1 Assigning states for privacy preferences of an agent

```
procedure $(l, t)$                                   ▷ locations and times of day
    $proStateStore \leftarrow \{\}$
    $antiStateStore \leftarrow \{\}$
    $c \leftarrow product(l, t)$                     ▷ c represents the context tuples
    for $l_c, t_c \in c$ do
        $s \leftarrow randomChoice(S, I)$            ▷ (S)usceptible, (I)nfected
        $proStateStore[(l_c, t_c)] \leftarrow s$
        if $s = S$ then
            $antiStateStore[(l_c, t_c)] \leftarrow I$
        else
            $antiStateStore[(l_c, t_c)] \leftarrow S$
        end if
    end for
end procedure
```

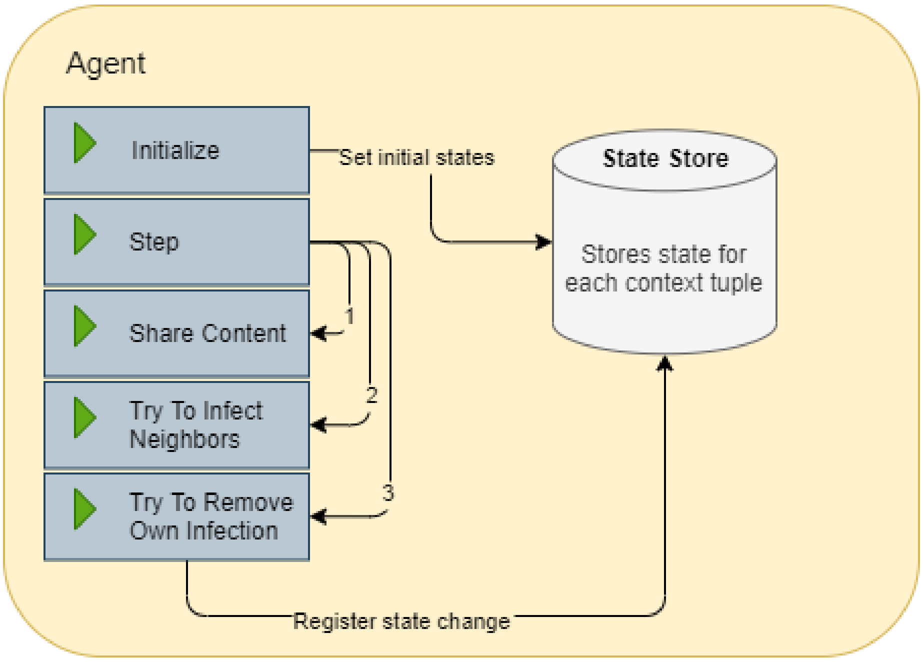

Figure 5: Schematic view of the components of an agent in the SIR model for privacy preference diffusion.

Algorithm 2 Assigning infection and recovery rates for privacy preferences of an agent

```
function setRates($agentRates, contextTuples, rate$)
    for $l_c, t_c \in contextTuples$ do
        $agentRates[(l_c, t_c)] \leftarrow rate$
    end for
return $agentRates$
end function

procedure $(l, t, iPro, rPro)$        ▷ locations, times of day, infection rates, recovery rates
    $iProAgent \leftarrow \{\}$
    $rProAgent \leftarrow \{\}$
    $c \leftarrow$ product$(l, t)$        ▷ c represents the context tuples
    if isType$(iPro, Mapping)$ then
        $iProAgent \leftarrow iPro$
    else
        $iProAgent \leftarrow$ setRates$(iProAgent, c, iPro)$
    end if
    if isType$(rPro, Mapping)$ then
        $rProAgent \leftarrow rPro$
    else
        $rProAgent \leftarrow$ setRates$(rProAgent, c, rPro)$
    end if
end procedure
```

4.1.2 Agent Actions

In the DIPP model, an agent can try to infect other agents, with which they are in a friendship relationship, with their privacy preferences by sharing. An agent can also recover from an infection at any point. These key actions are described in detail here.

Every time an agent takes a *step* in this model, they do at least three things: share content if possible on pro side, share content if possible of anti side and try to recover from an infection. An agent shares content by first checking their state store. The agent lists every context tuple for which they are infected and choose randomly from this list a context tuple. The content shared by the agent is then represented by this context tuple and other agents can perceive this. The step function can be seen in code block 3.

Algorithm 3 Actions an agent takes at each step

```
function STEP()
    $allInfectedProContexts \leftarrow$ findInfections($self.proStateStore$)
    $contentToSharePro \leftarrow$ randomChoice($allInfectedContexts$)
    tryToInfectNeighbours($contentToSharePro$)

    $allInfectedProContexts \leftarrow$ findInfections($self.proStateStore$)
    $contentToSharePro \leftarrow$ randomChoice($allInfectedContexts$)
    tryToInfectNeighbours($contentToSharePro$)
    tryToRemoveInfection($anti = False$)
    tryToRemoveInfection($anti = True$)
end function
```

To infect neighbours, an agent firsts lists neighbours that are susceptible for the content it wants to share. To capture the concept of co-ownership, a random number of neighbours, from zero to five, is randomly chosen to represent co-owners. Note that this means more often than not that the content shared in this model is owned by more than one person. As explained in Section 3.5, the infection rates of chosen neighbours are affected in a way of making them more likely to be infected. Subsequently, an attempt is made to infect each susceptible neighbour with heightened infection rates for co-owners. The infection rates of susceptible neighbours that are infected with the opposing privacy preference are lowered, as detailed in Section 3.3. Finally, the trust value changes the infection rate once again, as explained in Section 3.4 The resulting infection rate is then the probability that the neighbour will get infected. The infection action is captured in code block 4 from line 7 to 29.

After the agent has attempted to infect all susceptible neighbours, all neighbours, that are not susceptible get to perceive the shared content and register a positive or negative experience according to their preferences, from line 29 of code block 4.

The final part of a step is the *attempt* to recover from an infection. In this part, a context tuple is chosen randomly from all the context tuple for which the agent is infected. The context tuple's recovery rate is the probability that the agent will recover in this step, see code block 5.

Algorithm 4 Infecting neighbours

```
function tryToInfectNeighbours($contextTuple$)
    $susceptibleNeighbours \leftarrow$ getSusceptibleNeighbours($contextTuple$)
    $nonSusceptibleNeighbours \leftarrow$ getNonSusceptibleNeighbours($contextTuple$)
    $nCoOwners \leftarrow$ randomFloat$(0, 5)$
    $coOwners =$ randomSample($susceptibleNeighbours, nCoOwners$)

    $criticalValue \leftarrow$ randomFloat$(0, 1)$
    for $neighbour \in susceptibleNeighbours$ do
        $iPro \leftarrow neighbour$.getInfectionRate($contextTuple$)
        if $neighbour \in coOwners$ then
            $iPro \leftarrow iPro \times$ randomFloat$(1.1, 1.5)$
        end if
        if $neighbour$.getState($contextTuple, anti = True$) $= I$ then
            $iPro \leftarrow iPro \times$ randomFloat$(0.1, 0.5)$
        end if
        $iPro \leftarrow iPro \times neighghbor$.getTrustValue($self$)
        if $criticalValue < iPro$ then
            $neighbour$.setState($contextTuple, I$)
            if $neighbour$.getState($contextTuple, anti = True$) $= I$ then
                $neighbour$.setState($contextTuple, S, anti = True$)
            end if
            $neighbour$.registerPositiveExperience()
        else
            if $neighbour$.getState($contextTuple, anti = True$) $= I$ then
                $neighbour$.registerViolation($level = 2$)
            end if
        end if
    end for
    for $neighbour \in susceptibleNeighbours$ do
        if $neighbour$.getState($contextTuple, anti = True$) $= I$ then
            $neighbour$.registerViolation($level = 2$)
        else
            $neighbour$.registerPositiveExperience()
        end if
    end for
end function
```

Algorithm 5 Recovery from an infection

```
function tryToRemoveInfection()
    $allInfectedProContexts \leftarrow$ findInfections($self.proStateStore$)
    $contextToRecoverFrom \leftarrow$ randomChoice($allInfectedContexts$)
    $criticalValue \leftarrow$ randomFloat$(0, 1)$
    $rPro \leftarrow self$.getInfectionRate($contextToRecoverFrom$)
    if $criticalValue < rPro$ then
        $self$.setState($contextToRecoverFrom, R$)
    end if
end function
```

4.2 Model

The DIPP model simulates the spread of privacy preferences on a OSNs using real-world OSN data. In essence, it is the environment that facilitates the agent described in the previous section and collects data on the interactions that occur in a simulation. This section discusses a few key components. A schematic view of the DIPP model can be seen in Figure 6.

Role in simulations

For a researcher, the model is the interface to interact with to run any simulations. It creates agents for each node on the social network provided to it. Furthermore, it handles the stepping mechanism. Specifically, the DIPP model makes sure that the step execution is random every time. This way the order in which agents take a step is random each time and not the same sequence repeatedly. The model collects for each context tuple the number of agents in each of the three states. This allows for a clear picture of the SIR model dynamic. It also collects data on the number of state changes per step. Finally, the model also keeps track of the infection chains for each context tuple.

Customization

The model also provides an interface to customize the initial outbreak of privacy preferences. If a rare privacy preference is useful in an experiment, this can be defined with a mapping from context tuple to fractions representing a fraction of the population that should be infected with regard to the defined context tuples. The interface also provides the ability to specify the state of a specific agent regarding a specific context tuple.

Stopping criteria

It is possible to run the model for any number of steps. However, there is also an option to run the model until the state dynamics have reached a stable condition. Stability is reached when a predefined number of steps has passed and there have been fewer state changes than a predefined fraction of the population. For example, if one defines a network of 300 agents, the look back steps to be 10 and the fraction of the population to be 0.1, then the model will stop taking steps when it finds that the total number of changes in the last 10 steps has been less than 30.

Network

In the next section, the different simulations with the model are discussed. It is important to note that there is a basic model that does not have the concept of trust implemented. As such, this model uses an undirected network to represent the social network. In the settings with trust, a model that requires a directional graph is used, as the edges in this graph also carry information of the positive and negative experiences agents have with each other. This information is needed to derive the trust values between agents.

This specification of an epidemic model is believed to be suited to model the spread of privacy preferences through privacy decisions. In the next section, the measurable factors are explored and the suitability clam is tested with simulations of varying settings.

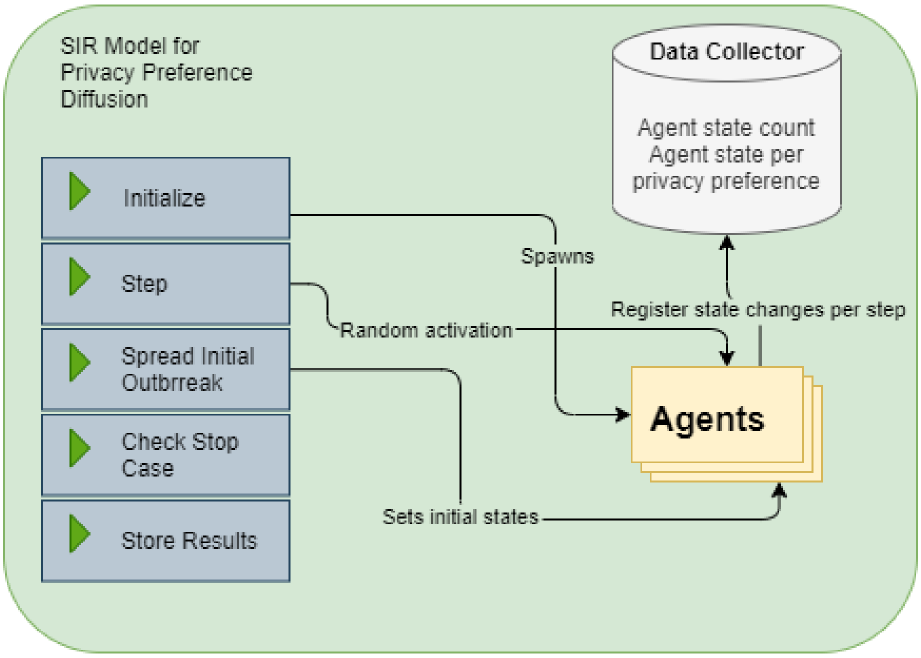

Figure 6: Schematic view of the components of the SIR model for privacy preference diffusion. It shows the main functions and their relationship to the agents.

5 Experiments

In this section, the inner workings of all the experiments are covered and all measurement are outlined.

5.1 Measurements

As stated in the previous section, the DIPP model keeps track of infection chains across all privacy preferences as well as agent count for each state of each privacy preference. It is clear that agent counts are useful to understand the dynamic of the SIR model in this state. The infection chains are used to derive an influence rating for each agent for each privacy preference. The influence rating is one of the main measured factors. Infection endings and number of crossovers are also part of this group of factors. This section delves into these measurable factors and defines the specific questions for the simulations.

5.1.1 Influence

In information diffusion modelling, influence has always been important. Sometimes influential users in a network are sought after for different reasons. It might be because they could be used to promote information on the network. In other cases, they might be deemed *bad influences*.

Definition 5.1 (Influence rating)**.** The influence rating of an agent is the count of each descendant proportional to the length of the shortest path between the agent and the descendant in the infection chain.

In this work, we propose a metric to measure the influence of agent in the spread of privacy preferences using infection chains. For each privacy preference, an infected chain is created. This infection chain is defined as a directed graph $G = (V, E)$ with a set of nodes V and a set of edges E.

$$E \subseteq V \times V$$

$$(a_i, a_j) \in E$$

$$\{a_i, a_j\} \subseteq V$$

Any agent on the social network could be a member of the set V for any privacy preference. The set of edges describes the occurrence of one agent infecting another agent.

To derive the influence of an agent, all the shortest paths between node pairs in G are calculated. This provides the shortest path length for each node to a node that they have directly or indirectly infected,

$$d(a_m, a_r)$$

for agents Alice and Bob. If we were to count the number of descendants Bob has in the infection chain, we would get a measure for influence that assumes that each descendant Bob in the infection chain is directly infected by Bob, which may well not be the case. Thus, Bob's influence rating is defined as

$$\sum_{\forall a_i \in V, a_i \neq a_r} 0.5^{d(a_m, a_r)}$$

, in which we count each descendant proportional to the length of the shortest path between the agent and the descendant in the infection chain.

5.1.2 Epidemic endings

One of the goals of this project is to explore whether trust values can be used by agents to protect themselves against opposed privacy preference. To this end, the epidemic endings are measured. An epidemic ending of a privacy preference is the step at which no agent on the network is infected with said privacy preference. If, at step 31, no more agents are infected the privacy preference to share content from a night at the beach, then that is the epidemic ending of $< mall, night >$.

5.1.3 Epidemic peaks

A subsequent goal of this project is to explore whether trust values can be used by agents to protect themselves against infections of opposed privacy preferences. To this end, the epidemic peaks are measured. An epidemic peak of a privacy preference is the maximum number of infected agents, infected with that privacy preference. If, at some step of the simulation, the number of infected agents with privacy preference of $< mall, night >$ is 600 and the epidemic never reaches a higher number of infected agents at one time, then 600 is the epidemic peak of the privacy preference $< mall, night >+$.

5.2 Simulations

For the purposes of answering the research questions of this project, simulations are run with the model described in the previous section. Some settings or setups are constant across all simulations. These setting will be explained first. After this, the specific simulations are put forth. Finally, the expectations from these experiments are discussed.

5.2.1 Basic settings

All the simulations use flat infection rates and recovery rates. This means that all the privacy preferences have the same rates on both pro and opposing sides. Three rates are chosen, $\{0.25, 0.50, 0.75\}$. These three values are varied over the four variables: pro infection rate, pro recovery rate, anti infection rate and anti recovery rate. These lead to a total of $3^4 = 81$ combinations of rates to run. Every simulation is executed 100 times to account for the stochastic nature of the model's dynamic. 100 repetitions is deemed to be the right compromise between correctness of the results and the run time of the experiments. This means that one experimental setting amounts to 8100 simulation runs. A simulation is run until a stable condition is reached. Stability is defined using a look back of 20 steps and a fraction of the population of 0.05. This means that a simulation ends when in the last 20 steps there have been fewer state changes than $0.05 \times 800 = 40$, as is summarized in Table 3.

In the case of disregarding the opposing preference dynamic, there is no need to include the anti infection and recovery rates. Consequently, there are $3^2 = 9$ combinations of rates to run. The number of repetitions stays 100 in this case as well as the stability criteria. These basic settings are used to run the baseline simulation setting. The following sections describe simulation settings that incorporate different factors previously discussed. The baseline will be used to compare and recognize any impact of said factors.

There are few general characteristics that are expected to be present in the results of simulations with basic settings. It is expected that:

- **1a** a privacy preference epidemic will last longer when the infection rate of said privacy preference is higher than its recovery rate,
- **1b** this setting will show a positive correlation between the degree of an agent on the network and its influence on the spread of privacy preference.

These assertions need to hold for the model to, at least, be a valid epidemic model. Thus, they are the foundations of the DIPP model.

A node's degree has been shown previously to correlate with the influence of that node in information diffusion [19]. That feature is expected to stay the same in the context of privacy preference diffusion as modelled here.

The following sections will elaborate on the rest of the simulation settings used to investigate the dynamic of privacy preference epidemics, as summarized in Table 4.

Parameter	**Values**
Pro infection rate	$\in \{0.25, 0.50, 0.75\}$
Pro recovery rate	$\in \{0.25, 0.50, 0.75\}$
Anti infection rate	$\in \{0.25, 0.50, 0.75\}$
Anti recovery rate	$\in \{0.25, 0.50, 0.75\}$
Iterations	100
Look back	20 steps
Population fraction	0.05

Table 3: Basic settings of each simulation. Note that given the varying rates, each experiment has a total of 81 settings.

5.2.2 Rare privacy preference with high degree influencers

Experiments in this setting are performed to investigate the ability of a few of the most influential users to spread privacy preferences that are rare.

The privacy preferences that are set to be rare are represented by the content tuples $< night, work >$ on the pro side and $< afternoon, beach >$. Rarity is defined as 0.2% of the network having the privacy preference. In other words, 0.2% of the agents will be infected for the context tuples above. Furthermore, the top 1% of the agents, with regard to degree, are assigned the rare privacy preference. This setting emulates the common phenomenon of commercial companies recruiting users on social media with a high number of followers to promote their product. The product, in this case, is a privacy preference.

With this setting, it is hypothesized that:

- **2a** the influence of the top 1% of agents on the spread of the rare privacy preference will be significantly higher than in the baseline simulation setting.

In the baseline setting, anyone can be an *influencer*, as everyone has an equal probability of starting the simulation infected with any privacy preference. The agents of the top 1% are expected to still have a high influence in the baseline setting due to their degree. In this setting, however, it is hypothesized that their influence will increase, because they have fewer agents competing for influence in the spreading of the rare privacy preference.

5.2.3 No Trust

Experiments in this setting are performed to investigate what happens when social network users don't trust each when it comes to privacy preferences. This setting incorporates a static trust value for each agent towards each of their friends on the network. The trust value is always 0. This setting is mostly used to prove that the workings of the mechanism. In this setting, it is hypothesized that:

- **3a** no infections are expected to take place in this setting as all infection rates will be 0 given the trust values of 0.

5.2.4 Static Trust

Experiments in this setting are performed to investigate how privacy preferences spread when agents trust each other to some degree and this level of trust also remains static throughout an experiment.

This setting is an extension of the no trust setting. This setting incorporates a static trust value for each agent towards each of their friends on the network. The trust value is a random number between 0 and 1. This setting portrays a scenario in which OSN users could assign a trust value to each of their friends once. It is expected that:

- **4a** the epidemics will last shorter than in the baseline simulations as the inclusion of trust negatively influences the infection rate,
- **4b** agents' influence on the spread of privacy preferences will be diminished compared to the baseline simulations, since how much an agent is trusted by their neighbours now becomes a factor as well,
- **4c** epidemic peaks will be lower than in the baseline simulations.

It should be noted that trust values are only presented on the pro side of the epidemic. This is the side where content is actually shared and can be perceived by users of the OSN. This applies to all simulation settings that incorporate trust.

The hypotheses also show the fact that agents would be able to protect themselves from opposed privacy preferences by limiting the impact of the epidemics.

Experiment	**Key properties**
Baseline	Basic settings
Rare privacy preference with high degree influencers	Rare privacy preference
Static Trust	Trust
No Trust	Trust always zero
Dynamic Trust	Updated Trust
Dynamic Trust with different violation levels	Levels, Updated Trust

Table 4: All the experimental settings of this work

5.2.5 Dynamic Trust

Experiments in this setting are performed to investigate how privacy preferences spread when agents trust each other to some degree and this level of trust changes over time. When an agent's privacy is violated, they trust the perpetrator less than before. Each type of violation leads to the same decrease of trust. When Alice perceives Bob sharing content that represents a privacy preference that she opposes, Alice will register a violation. Furthermore, when Bob shares content that is co-owned by Alice without Alice's consent, Alice also registers a privacy violation. Alice will then trust less in the future and be less influenced by Bob's content sharing habits.

This setting is an extension of the static trust setting. This is the first setting that incorporates privacy violations. At each time step, an agent can perceive the two above-mentioned types of privacy violations. However, in this setting, there is no distinction between the two types. This means the trust value is equal to

$$tr = \frac{n+2}{N+2}$$

, where n is the number of interactions without privacy violations and N the number of interactions between two agents. In contrast to static trust setting, this setting starts off with a trust value of 1, as

$$tr = \frac{0+2}{0+2} = 1$$

.

The hypotheses for this setting are similar to the other settings with trust modelling, namely that:

1. **5a** the epidemics will last shorter than in the baseline simulations as the inclusion of trust negatively influences the infection rate,

2. **5b** agents' influence on the spread of privacy preferences will be diminished compared to the baseline simulations, since how much an agent is trusted by their neighbours now becomes a factor as well,

3. **5c** epidemic peaks will be lower than in the baseline simulations.

5.2.6 Dynamic trust with two levels of privacy violations

Experiments in this setting are performed to investigate how privacy preferences spread when agent model trust based on violations with different levels of severity. This setting is similar to the dynamic trust setting. Agents can adjust their trust value based on experiences they have with their friends, either positive or negative experiences. However, in this setting the agents calculate trust value based on the severity they assign to the two levels of privacy violations.

The severity weights are set to be $s_1 = 0.2$ and $s_2 = 0.8$. This makes that a level 2 violation is much more severe than level 1 version. This means that every agent feels that level 2 violations bear more weight. These values are also intuitive, because with a level 2 privacy violation, private information is actually revealed to other friends inside the OSN. This is more severe than perceiving privacy preferences that are not in line with agents. There is no investigation in the influence on the epidemic dynamic of the severity values, given the assumption mentioned.

The hypotheses for this setting are similar to the other settings with trust modelling, namely that:

- **6a** the epidemics will last shorter than in the baseline simulations as the inclusion of trust negatively influences the infection rate,

- **6b** agents' influence on the spread of privacy preferences will be diminished compared to the baseline simulations, since how much an agent is trusted by their neighbours now becomes a factor as well,

- **6c** epidemic peaks will be lower than in the baseline simulations.

5.2.7 Opposing privacy preference dynamic

At the beginning of this section, it was explained that opposing privacy preferences have to be present in any model for privacy preference diffusion as a way of modelling the users' beliefs regarding what should stay private. However, as stated, it is also the case that this work does not investigate the dynamics of opposing privacy preference epidemics. The dynamics are not perceivable in an OSN. That is why the aforementioned experiments were run twice.

In the first instance, the experiments are run with the assumption that the dynamic of the spread of opposing privacy preferences is the same as that of their counterparts on the pro side. The second situation assumes that the anti side is static. Agents in this instance have opposing privacy preferences, but they do not change over time. In both instances, trust is only a factor on the pro side of the privacy preference diffusion.

5.3 Social Network Data

The experiments outlined here are done on a real online social network. Researchers from the Technical University of Denmark have created the *Copenhagen Networks Study interaction data* set. This data set represents a multilayer temporal network. The data were collected at the Technical University of Denmark and subjects are freshmen at the same university, as documented by Sapiezynski *et al.* [23]. The data set captures four networks covering the subjects; a Facebook friendship network, an SMS text message network, a call network and a Bluetooth proximity network. The Facebook friendship network is an adjacency list in which each element represents a friendship relationship that lasts during the whole experiment during which the data were collected. The SMS text message network is an adjacency list in which each element represents an instance of an agent sending a text message to another agent. These data also capture the timestamp of each text message. The timestamp defines the number of seconds since the start of the experiment.

The starting point for this project is the Facebook friendship network, but the other networks can help to derive similarities between different members of the social network for extensions of the project in the future.

The Facebook friendship network has a total of 800 subjects whose friendships are represented in a static manner. The degree distribution of this network can be seen in Figure 7. The average degree in this network is 16.07 while the minimum and maximum degrees are 1 and 101 respectively. Every user in this network is an agent in the forthcoming experiments. This network is small, compared to the size of the entire Facebook social network, but, because of the realism of the network and the promise of the accompanying networks, it is the best option for this research.

6 Results

This section provides a review of the experiments that were run. As previously stated, each parameter setting is run 100 times to account for the stochastic nature of the state dynamics in the DIPP model. The goal of the simulations is to capture data of the metrics: influence rating of agents, infection endings

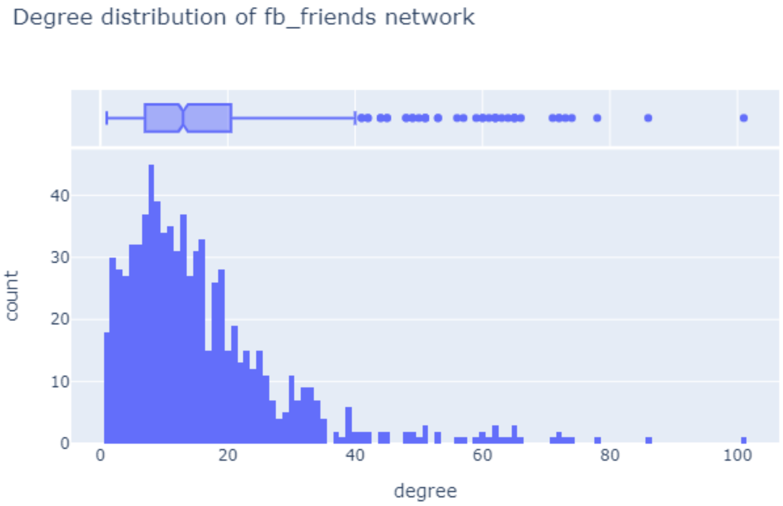

Figure 7: Network of Facebook friendships from *The Copenhagen Networks Study interaction data* set [23]

and infection peaks. These data will help in testing the hypotheses set out for the various simulations settings. Ultimately, an attempt is made to understand the influences of various circumstances on the diffusion of privacy preferences. Furthermore, it is investigated whether the introduction of trust modelling limits the impact of privacy preference epidemics.

The software, created to run the simulations, is written in Python, version 3.8.3. The Mesa library is used as the backbone of the simulations. It provides customizable agent-model dynamic, random activations mechanisms and data collection. These properties can also be seen in Table 5.

Software Properties	
Programming Language	Python 3.8.3
Notables Libraries	Mesa
State Stores	Python Dictionaries
Data Collector	Python Dictionary

Table 5: Software Properties of the DIPP model

The review of the results will build up from the most basic simulation settings to the most elaborate settings, as previously in the method section. Each experimental setting has been run with the opposing preference dynamic and without. The experiments were run with flat rates for all privacy preferences for each agent. The epidemics are also randomly spread throughout the network, as previously stated. Thus, the results will be discussed with focus on the privacy preference $< night, work >$. This is the only privacy preference that has an alternative initial spread, namely in the setting of rare privacy preferences with high degree influencers. We will first explore the results from the experiments in which the opposing privacy preferences are not static before following up with the results of the experiments where the opposing preferences are static over all steps. In the following sections, there will be references to a digital appendix here.

6.1 Including opposing preference dynamics

In this section, the results are presented for the simulations in which the opposing privacy preferences states are assumed to follow the same dynamic as the privacy preferences on the pro side.

6.1.1 Baseline simulations

In the baseline simulations, all parameters settings are as defined in the basic settings. These simulations are performed to record a baseline impression of the DIPP model with the basic settings. This baseline is later used for comparison when experiments are performed with more elaborated versions of the DIPP model. These comparisons will, in turn, show the effect of including different factors in the DIPP model. Flat infection and recovery rates are used for both the pro and the anti side. Agents hold opposing privacy preferences and can share co-owned content. Trust is not a factor in these simulations.

In Figures 9 and 8, the dynamics of the $< night, work >$ privacy preference are shown. For each state, the box plot of the number of agents is plotted against the step in the simulation. In Figure 8, it shows the rapid rise of immune agents from step 0 while infected and susceptible agents decrease in number from the same step.

In the extremes of the box plot for the susceptible state, the effect of the pro side can be observed. It shows that in some experiments the number of agents susceptible to the $< night, work >$ opposing privacy preference increase shortly from step 0 before starting to decrease around step 6. This happens as agents are being infected by the pro side privacy preference and thus becoming susceptible to this anti side preference. This can be seen in Figure 9, where in the same region of the graph as in Figure 8. The number of infected agents rises up until around step 5 of the simulation before decreasing to zero.

More contrasting dynamics can be seen in Figures 10 and 12. This figure visualizes the aggregated dynamic of the $< night, work >$ privacy preference when the pro infection rate is 0.75 and the anti recovery rate is 0.75. The privacy preference is more infectious on the pro side while agents are more likely to recover from it on the anti side. In this setting, it can be noted the number of infected agents on the pro side increases more dramatically at the start of the simulation and the epidemic seems to end at around the same step as in the previous parameter setting. Due to the rapid increase of infected agents, there is also a dramatic decrease of susceptible agent on the pro side in the first 20 steps.

On the anti side, the number of susceptible agents increase rapidly in the first 5 steps or so only to later decrease and stabilize at a median value of roughly 300 agents. The spread of this stabilization number is wider than in the dynamic where all rates were 0.25. The number of resistant agents increases quickly at the start of the simulation only to stabilize at a median value of roughly 500

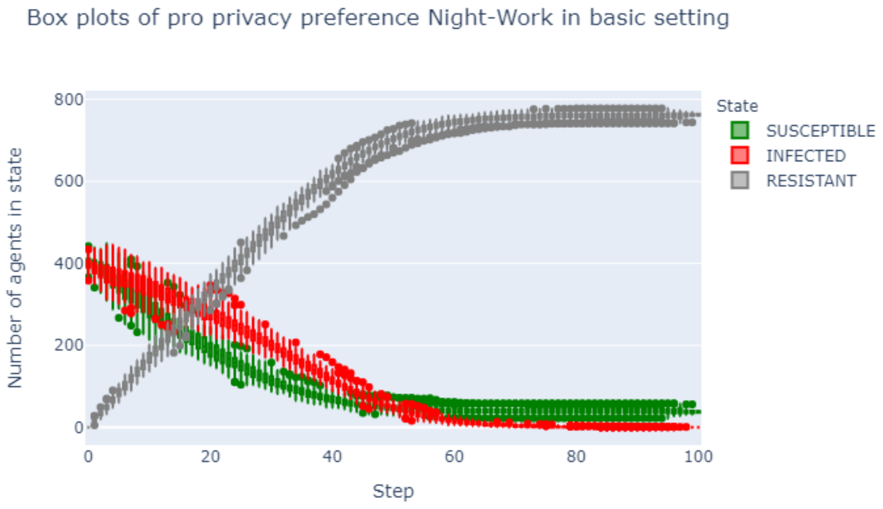

Figure 8: Dynamic of the anti side $< night, work >$ privacy preference with pro infection rate 0.25, pro recovery rate 0.25, anti infection rate 0.25, anti recovery rate 0.25

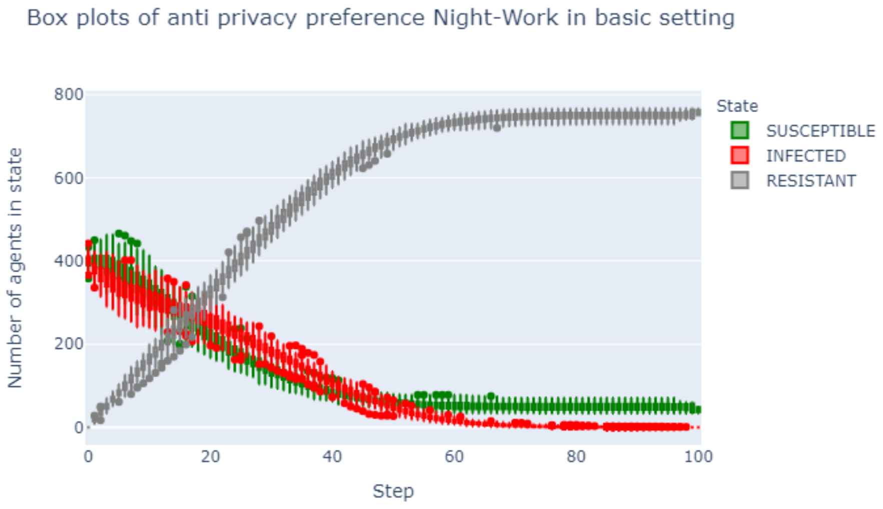

Figure 9: Dynamic of the pro side $< night, work >$ privacy preference with pro infection rate 0.25, pro recovery rate 0.25, anti infection rate 0.25, anti recovery rate 0.25

agents. However, once again, the spread of this stability value seems much wider than in the dynamic where all rates were 0.25. After this general overview of the dynamic, with an illustration of the effect of the parameter setting, the focus switches to the hypotheses for this setting in the next sections.

On the pro side, Figures 10 and 11 depict the two most extremes. Figure 10 displays data from simulations with most favourable parameter settings for a privacy preference on the pro. The recovery rate is the lowest and infection rate the highest. Furthermore, in 11 the least favourable parameter settings is depicted, the infection rate of the opposing privacy preference is lowest and the recovery rate of said preference is highest. This can be seen in the results as the peak number of infected in Figure 11 is lower than in Figure 10. The peak in question is also at step 0 and has a median value of 400.5. This means this peak stems from the initial infection spread only. It is also noticeable that the number of recovered agents in the least favourable setting never reaches the same heights as in the most favourable setting. This seems a manifestation of the strength of the opposing privacy preference dynamic. When an agent gets infected with a privacy preference for which the agent is already infected on the opposing side, the agent become susceptible on this opposing side. This effect is clearly visible in Figure 11 as at the beginning of the simulation the number of susceptible agents shortly and sharply increases before dropping to a stable number.

These observations follow intuitively from the theory behind the DIPP model and experimental setting. This is a positive point as the baseline seems, from observations only, to provide a stable base for the latter experiments.

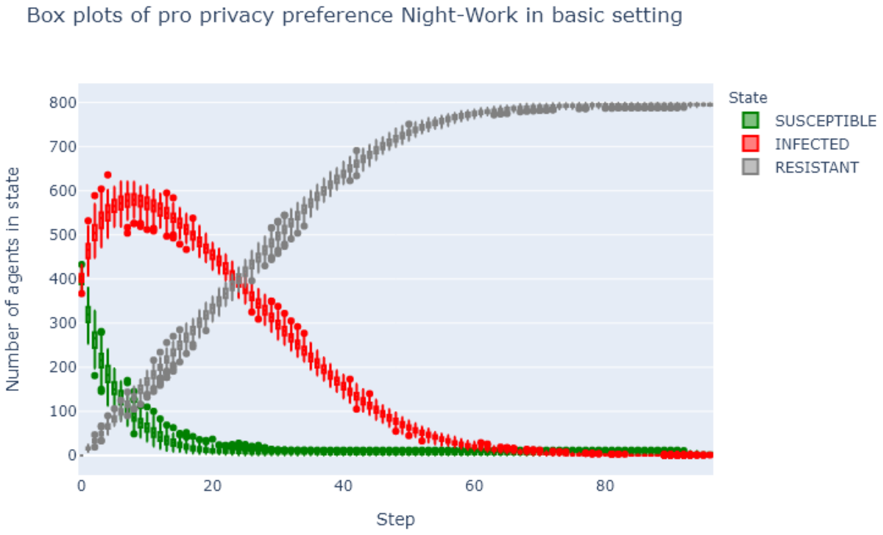

Figure 10: Dynamic of the pro side $< night, work >$ privacy preference with pro infection rate 0.75, pro recovery rate 0.25, anti infection rate 0.25, anti recovery rate 0.75

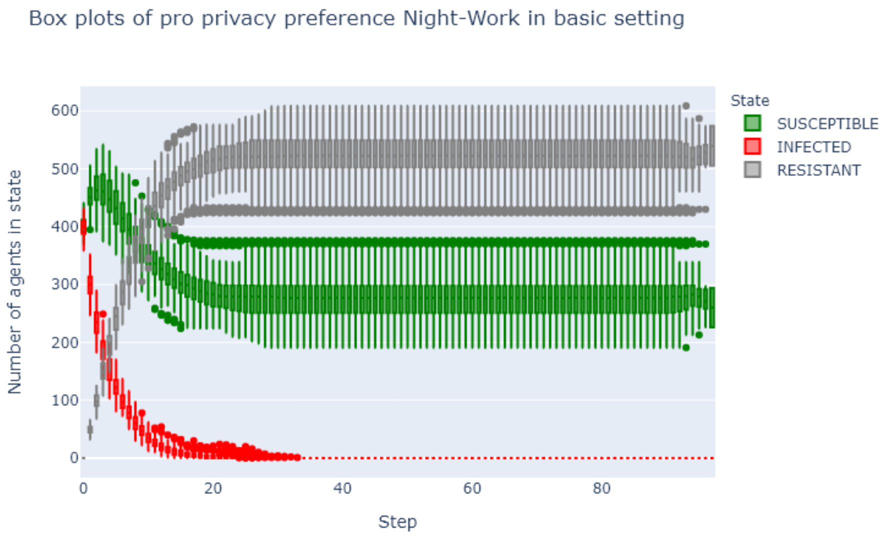

Figure 11: Dynamic of the pro side $< night, work >$ privacy preference with pro infection rate 0.25, pro recovery rate 0.75, anti infection rate 0.75, anti recovery rate 0.25

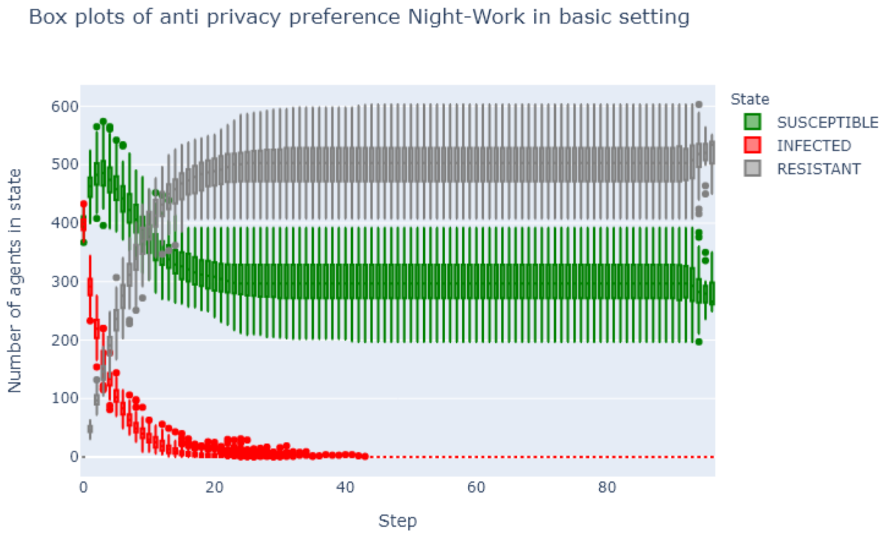

Figure 12: Dynamic of the pro side $< night, work >$ privacy preference with pro infection rate 0.75, pro recovery rate 0.25, anti infection rate 0.25, anti recovery rate 0.75

6.1.1.1 Infection Rate vs. Recovery Rate

Claim 1a for the baseline simulations was that a privacy preference epidemic will last longer when the infection rate of this privacy preference is higher than its recovery rate. To this end, data were collected on when epidemics end. For each parameter setting, there are 100 repetitions. It follows that for each privacy preference, there are 100, possibly different, steps at which its epidemic endings. The data collected are then segregated in data from the pro side and anti side. For each side, there are three groups; one group where the infection rate is higher than the recovery. The second group contains data from parameter settings in which the infection rate is the same as the recovery rate. Finally, the final group contains data from settings in which the infection rate is lower than the recovery rate. Using the paired Wilcoxon signed rank test, it is then tested whether the steps, at which infections end, in group 1 are significantly later than those in the latter two groups.

The paired Wilcoxon signed rank test is a non-parametric statistical test. The main idea of the test is to calculate the sign and the absolute difference between two paired values. The absolute differences are ranked and the rank is multiplied by the previously calculated sign of the difference. The statistic calculated is the sum of the ranks of the differences between two data samples. The sum of the signed ranks is then used to infer whether hypothesis H_0 can be rejected or not given a p value. This test is appropriate, because it is a non-parametric test. This means there is no need to check whether the collected data are normally distributed.

The results of these tests can be seen in tables 6 and 7. The results show that for every privacy preference and side it holds that an epidemic ends later when the infection rate is higher than the recovery rate as the p-values are all well below 0.05.

Scenario	Privacy Preference	p-value	Statistic
1	Afternoon-Beach	0.0000	1268113.5
2	Afternoon-Beach	0.0000	1183497.0
1	Afternoon-Mall	0.0000	1316022.5
2	Afternoon-Mall	0.0000	1117460.5
1	Afternoon-Work	0.0000	1304395.5
2	Afternoon-Work	0.0000	1125978.5
1	Evening-Beach	0.0000	1206737.0
2	Evening-Beach	0.0000	1032806.0
1	Evening-Mall	0.0000	1291506.5
2	Evening-Mall	0.0000	1114230.5
1	Evening-Work	0.0000	1260973.5
2	Evening-Work	0.0000	1117252.5
1	Morning-Beach	0.0000	1314371.0
2	Morning-Beach	0.0000	1144354.0
1	Morning-Mall	0.0000	1295205.5
2	Morning-Mall	0.0000	1084476.5
1	Morning-Work	0.0000	1267104.0
2	Morning-Work	0.0000	1150338.5
1	Night-Beach	0.0000	1243992.0
2	Night-Beach	0.0000	1162820.5
1	Night-Mall	0.0000	1293419.0
2	Night-Mall	0.0000	1179924.5
1	Night-Work	0.0000	1277388.0
2	Night-Work	0.0000	1058577.0

Table 6: Results of paired Wilcoxon signed rank tests that test whether an infection, on the pro side, ends later when the infection rate is higher than the recovery rate of privacy preferences. The alternatives tested against are when the infection rate is the same as the recovery rate (scenario 1) and when the infection rate is lower than the recovery rate (scenario 2)

Scenario	Privacy Preference	p-value	Statistic
1	Afternoon-Beach	0.0000	3041008.5
2	Afternoon-Beach	0.0000	3478445.5
1	Afternoon-Mall	0.0000	3037178.0
2	Afternoon-Mall	0.0000	3476607.5
1	Afternoon-Work	0.0000	3036123.5
2	Afternoon-Work	0.0000	3466150.0
1	Evening-Beach	0.0000	3021424.0
2	Evening-Beach	0.0000	3471660.5
1	Evening-Mall	0.0000	3059523.5
2	Evening-Mall	0.0000	3488633.0
1	Evening-Work	0.0000	3037852.5
2	Evening-Work	0.0000	3465699.0
1	Morning-Beach	0.0000	3004843.5
2	Morning-Beach	0.0000	3470422.0
1	Morning-Mall	0.0000	3034753.5
2	Morning-Mall	0.0000	3475077.5
1	Morning-Work	0.0000	3006659.0
2	Morning-Work	0.0000	3466522.0
1	Night-Beach	0.0000	3046827.0
2	Night-Beach	0.0000	3468488.0
1	Night-Mall	0.0000	3025085.0
2	Night-Mall	0.0000	3461368.5
1	Night-Work	0.0000	3023021.0
2	Night-Work	0.0000	3460549.0

Table 7: Results of paired Wilcoxon signed rank tests that test whether an epidemic, on the anti side, ends later when the infection rate is higher than the recovery rate of privacy preferences. The alternatives tested against are when the infection rate is the same as the recovery rate (scenario 1) and when the infection rate is lower than the recovery rate (scenario 2)

6.1.1.2 Agent Influence

The influence rating measure, previously defined, is used to measure an agent's influence on the spread of privacy preference. After each simulation, the influence ratings of each agent on each privacy preference on each side are calculated and the 100 ratings are averaged for each setting. In Figures 13 and 14, the influence ratings are plotted against the node degree of an agent on the social network. The graph suggests the existence of the aforementioned positive correlation between an agent's degree in a social network and its influence on the spread of a privacy preference.‘

To find evidence for this, hypothesis 1b, the Pearson correlation coefficient is calculated between the influence ratings and the corresponding agent's degree on the social network. The results can be seen in Table 8. All correlation coefficients were found to be significant with $p \leq 0.05$. This shows that there is a positive correlation between an agent's degree in the network and their influence in the spreading of privacy preferences.

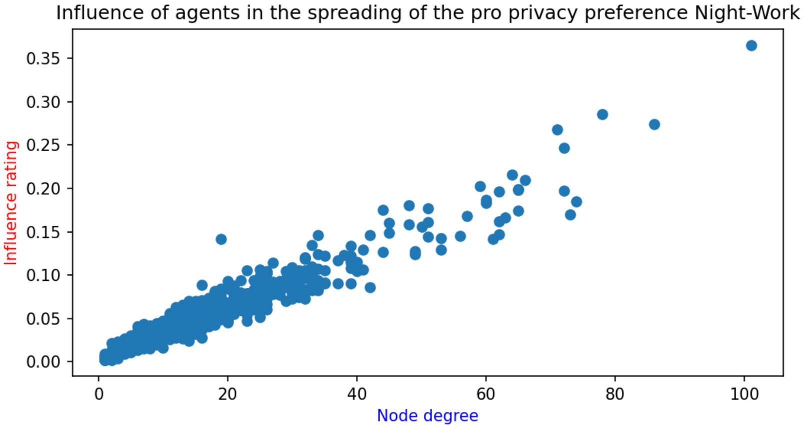

Figure 13: Average agent influence on the spread of the pro side $< night, work >$ privacy preference with pro infection rate 0.75, pro recovery rate 0.25, anti infection rate 0.25, anti recovery rate 0.75. Captured from results of the baseline simulations. Plotted against the degree of the agent on the social network

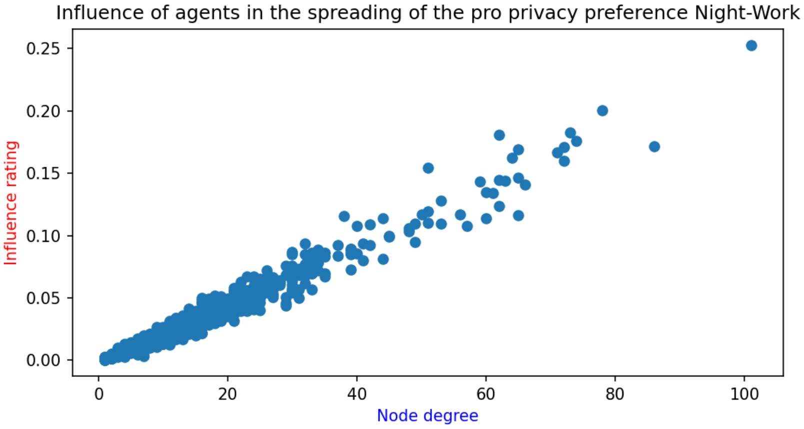

Figure 14: Average agent influence on the spread of the pro side $< night, work >$ privacy preference with pro infection rate 0.25, pro recovery rate 0.75, anti infection rate 0.75, anti recovery rate 0.25. Captured from results of the baseline simulations. Plotted against the degree of the agent on the social network

Median r	0.48
Mean r	0.48
Maximum r	0.62
Minimum r	0.49

Table 8: Summary of the Pearson correlation coefficients between the influence rating of an agent and the degree of said agent on the social network

6.1.2 Rare privacy preference with high degree influencers

Experiments in this setting are performed to investigate the ability of a few of the most influential users to spread privacy preferences that are rare. Specifically, the top 1% of agents, with regard to degree, are assigned privacy preference infections of a privacy preference with an initial spread of 0.2% of the agents in the network. The influence ratings of these agents are analysed to see whether they are more influential in this setting that in the baseline simulations.

The $< night, work >$ pro privacy preference is the rare privacy preference in this experimental setting. In Figures 15 and 16, the aggregates of the state dynamics of the $< night, work >$ pro privacy preference are shown, for the most favourable and least favourable parameter settings.

In Figure 15, one can see that the peak in the number of infected agents seems to be later than in the baseline simulations in Figure 10. Furthermore, it is noticeable that there are a lot more outliers in the box plots of this experimental setting compared to the baseline setting.

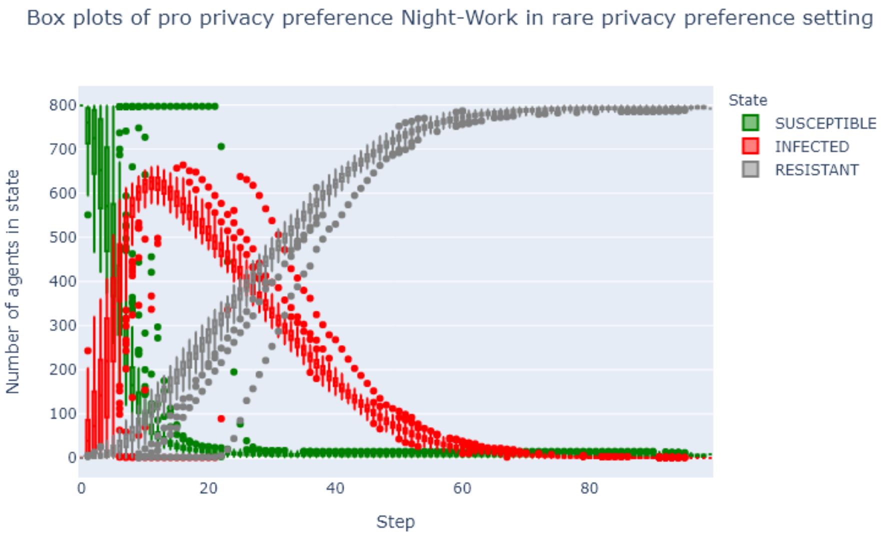

Figure 15: Dynamic of the pro side $< night, work >$ privacy preference with pro infection rate 0.75, pro recovery rate 0.25, anti infection rate 0.25, anti recovery rate 0.75

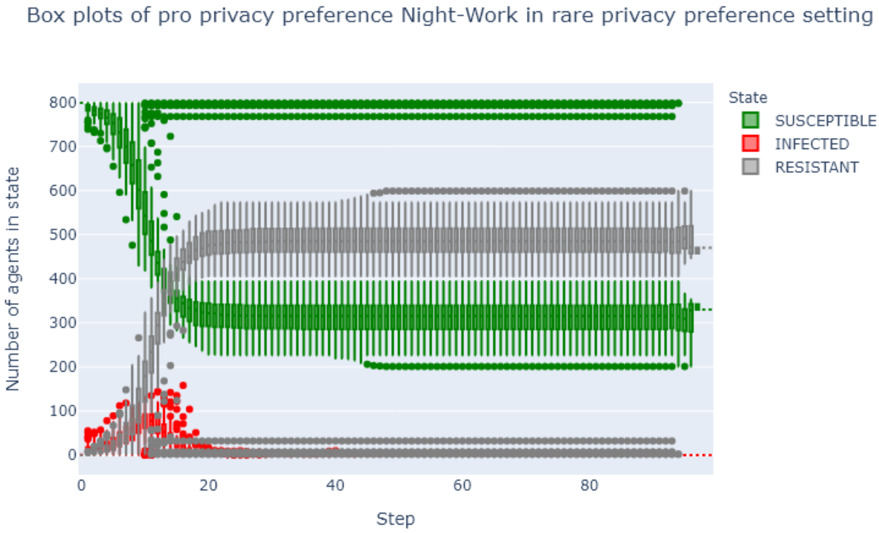

Figure 16: Dynamic of the pro side $< night, work >$ privacy preference with pro infection rate 0.25, pro recovery rate 0.75, anti infection rate 0.75, anti recovery rate 0.25

In Figure 16, once again, there are more outliers than observed in the previous setting. The number of infection rates peeks lower due to the lower infection rate. It can also be noted that the number of recovered agents peaks lower than when infection rate is 0.75 in Figure 15. This is most likely due to the high infection rate of the opposing privacy preference. This effect is similar to that seen in Figure 11. However, compared to Figure 11, in Figure 16 there is still a peak to be seen in the number of infected agents between step 10 and 20 of the simulations. This likely shows the power the small number of well-connected influencers, that try to spread this rare privacy preference, have.

In previous sections, the claim (2a) was made that the influence of the top 1% of agents on the social network, with regard to degree, increases when they are the only ones spreading a rare privacy preference. To prove this claim, data are collected on the influence of each agent on the spread of a privacy preference. In Figures 17 and 18, the average influence, on the spread of the $night, work$ privacy preference pro side, of an agent is plotted against the agent's degree on the social network. All agents in these figures are the top 1% of agents, with regard to degree. These influence figures can be compared to the figures of agent influence spreading the same privacy preference in the basic setting, as seen in Figures 14 and 13. It is noticeable that the average influence of the top 1% of agents rises when the privacy preference is rare. This suggests that these agents become more influential when they spread rare privacy preferences.

To find evidence for this claim, the paired Wilcoxon signed rank test is once again used to test whether the influence of the top 1% of agents, for the privacy preference $night, work$ on the pro side, is higher when the privacy preference is

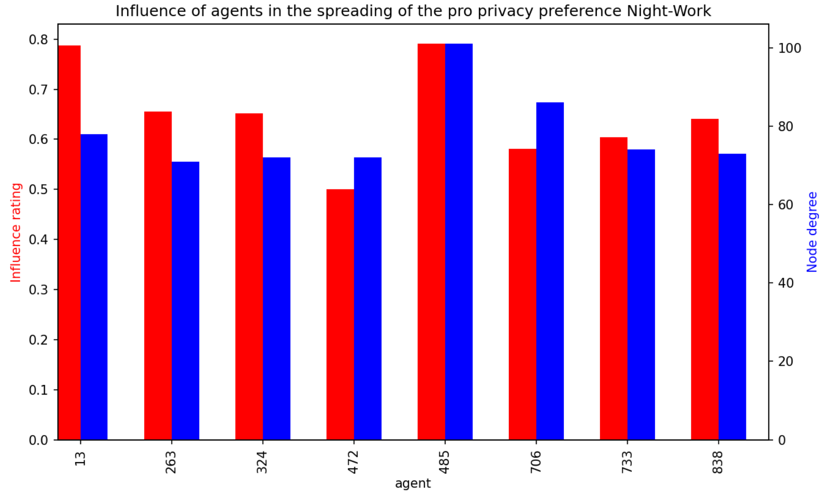

Figure 17: Average agent influence on the spread of the pro side $< night, work >$ privacy preference with pro infection rate 0.75, pro recovery rate 0.25, anti infection rate 0.25, anti recovery rate 0.75. $night, work$ is a rare privacy preference in this setting. Plotted against the degree of the agent on the social network

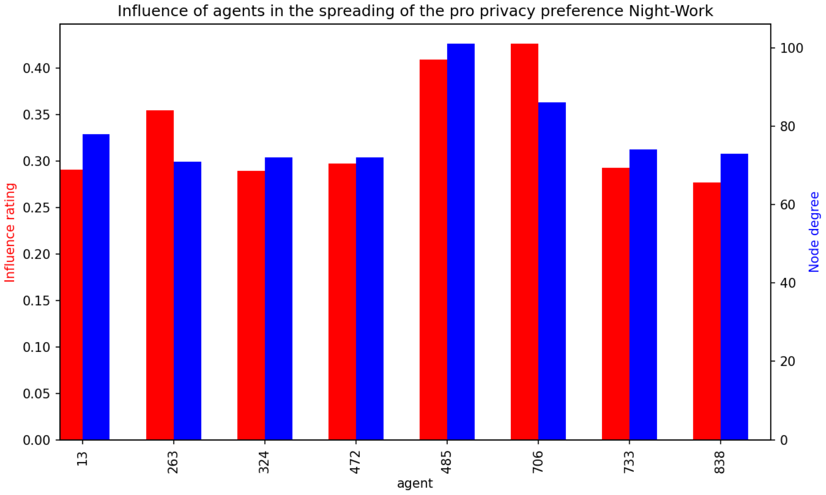

Figure 18: Average agent influence on the spread of the pro side $< night, work >$ privacy preference with pro infection rate 0.25, pro recovery rate 0.75, anti infection rate 0.75, anti recovery rate 0.25. $night, work$ is a rare privacy preference in this setting. Plotted against the degree of the agent on the social network

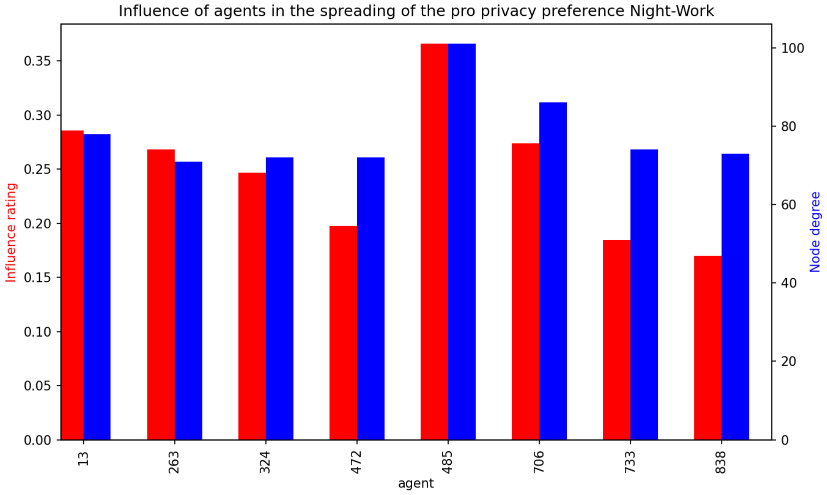

Figure 19: Average agent influence on the spread of the pro side $< night, work >$ privacy preference with pro infection rate 0.75, pro recovery rate 0.25, anti infection rate 0.25, anti recovery rate 0.75. Captured from baseline simulations. Plotted against the degree of the agent on the social network

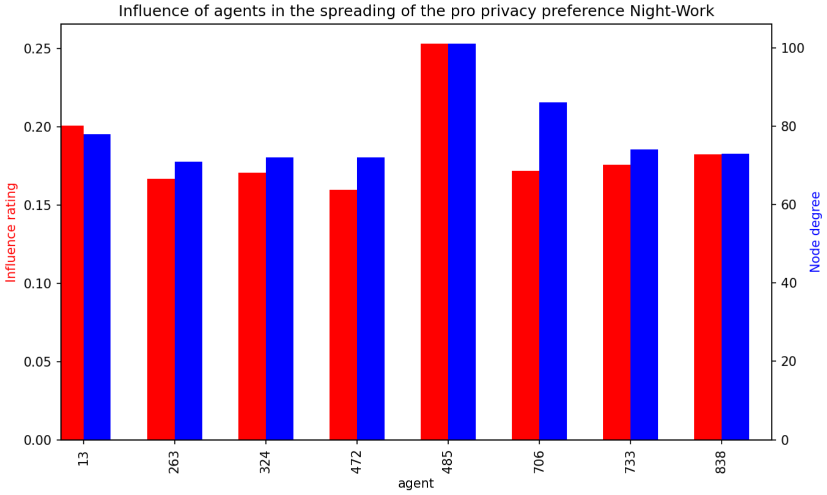

Figure 20: Top 1% of agents, with regard to degree, and their average agent influence on the spread of the pro side $< night, work >$ privacy preference with pro infection rate 0.25, pro recovery rate 0.75, anti infection rate 0.75, anti recovery rate 0.25. Captured from baseline simulations. Plotted against the degree of the agent on the social network

rare. Since the test is done per parameters setting, it seems tedious to display results from all 81 parameter settings. Thus, in Figure 9, the parameter settings are shown for which the claim does not hold, $p \geq 0.5$ for these settings. This means that the influence rating of the top 1% of agents does not increase in all parameters settings when the privacy preference they spread is rare. Thus, Claim 2a does not hold in all parameter settings. All parameters settings for which the claim does not hold have a pro side infection rate of 0.75. This is the maximum possible infection rate. This suggests that when the infection rate is at its highest there is no increase in top 1% of agents' influence ratings when spreading rare privacy preferences. However, there is still evidence that, in 76 out of 81 parameter settings the influence rating of the best-connected agents in the social network increases when they share content that represents a rare privacy preference.

Anti Recovery Rate	**Anti Infection Rate**	**Pro Recovery Rate**	**Pro Infection Rate**	**Statistic**	**p-value**
0.5	0.25	0.25	0.75	530258.5	0.244110
0.5	0.50	0.25	0.75	543861.0	0.239291
0.75	0.50	0.25	0.75	466986.0	0.093379
0.5	0.75	0.25	0.75	550419.5	0.221511
0.5	0.50	0.50	0.75	512343.5	0.226137

Table 9: Parameter settings for which the influence of the top 1% of agents, with regard to degree, is not higher when the privacy preference $night, work$ becomes rare.

6.1.3 No Trust

In this setting, the infection rate of privacy preference is proportional to the trust between two agents with a static trust value of 0. The data on the epidemics in this simulation setting show no rise in infection on the pro side, as expected. This phenomenon can be seen in Figures 21 and 22, with data collected from the two most extreme setting once again.

It is noticeable from the graphs that the number of infected agent only decrease after step 0. In the least favourable settings for the pro side norms, the number of infected agent declines even faster as more agents are likely to adapt the opposing privacy preference due to its high infection rate. These results are in line with Claim 3a, namely that no infections were expected to take place in this setting.

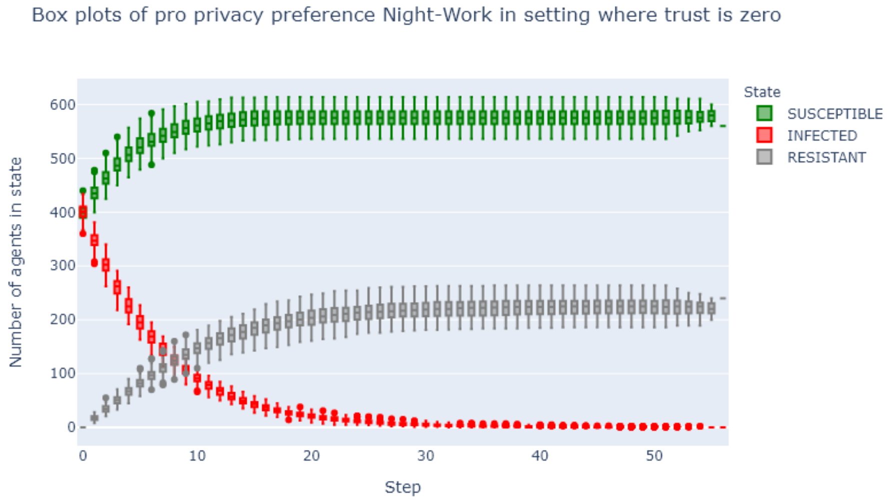

Figure 21: Dynamic of the pro side $< night, work >$ privacy preference with pro infection rate 0.75, pro recovery rate 0.25, anti infection rate 0.25, anti recovery rate 0.75

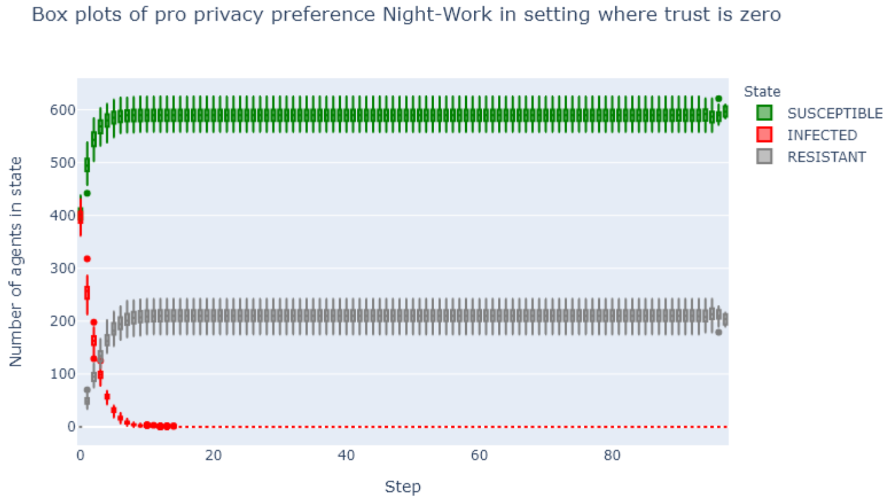

Figure 22: Dynamic of the pro side $< night, work >$ privacy preference with pro infection rate 0.25, pro recovery rate 0.75, anti infection rate 0.75, anti recovery rate 0.25

6.1.4 Static Trust

The simulations of this setting reflects a situation in which an agent can specify trust values to other agents they are connected to in a social network. In contrast to the no trust setting, agents in these simulations are assigned random static trust values at the start of each simulation. The purpose is to find out if epidemics are less impactful when static trust is introduced.

In Figures 23 and 24, the state dynamics of the two most extreme settings are displayed once again. It can be noted that in the most favourable setting for the pro side, Figure 23, the number of infected agents always peak below 600 agents while in the baseline simulations the peak at times is even higher than 500. This is in line with the expectations for this setting. It was also expected that the epidemics would last shorter in this setting, however, there is no clear indication of that in these figures.

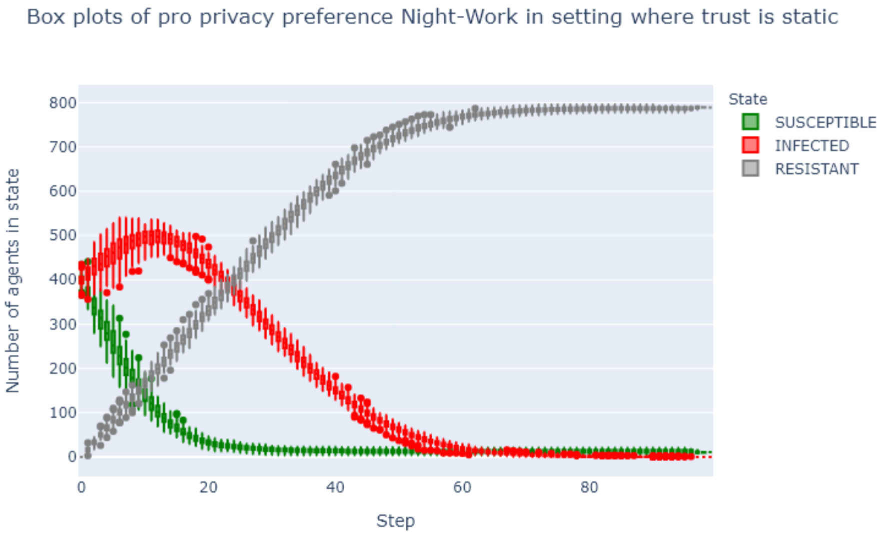

Figure 23: Dynamic of the pro side $< night, work >$ privacy preference with pro infection rate 0.75, pro recovery rate 0.25, anti infection rate 0.25, anti recovery rate 0.75. Taken from simulations in which agent holds static trust values.

In Figures 13 and 26, the average influence of agents is plotted against the degree of the agents on the social network. Although there still seems to be a correlation between the degree of an agent and its influence in the spread of a network, the influence seems to be lower compared to the baseline simulation results in Figures 25 and 26. Using the paired Wilcoxon signed rank test, the hypothesis (4b) that are agents' influences are diminished in the presence of static trust values is tested. The results of these tests can be found in the appendix. The results show that for each parameter setting there is significant evidence that the influence of agents is diminished when static trust values are introduced compared to the baseline simulations.

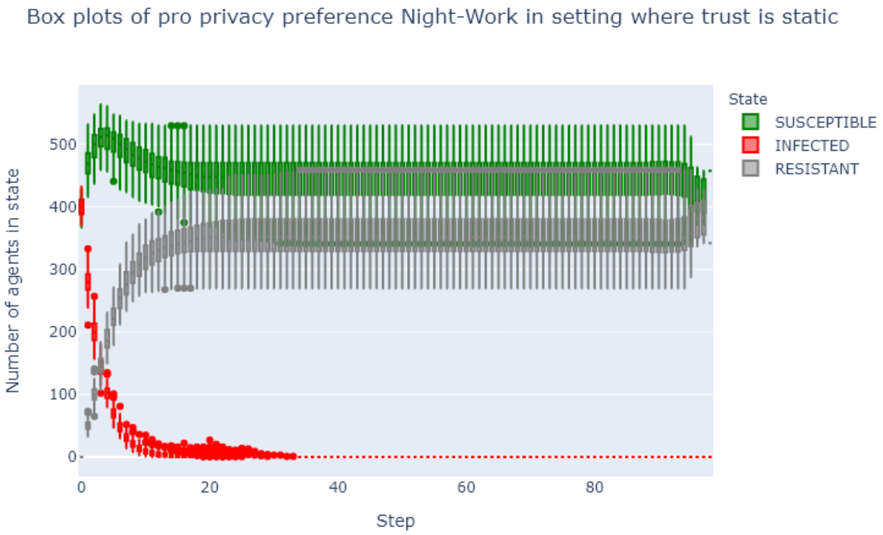

Figure 24: Dynamic of the pro side $< night, work >$ privacy preference with pro infection rate 0.25, pro recovery rate 0.75, anti infection rate 0.75, anti recovery rate 0.25. Taken from simulations in which agent holds static trust values.

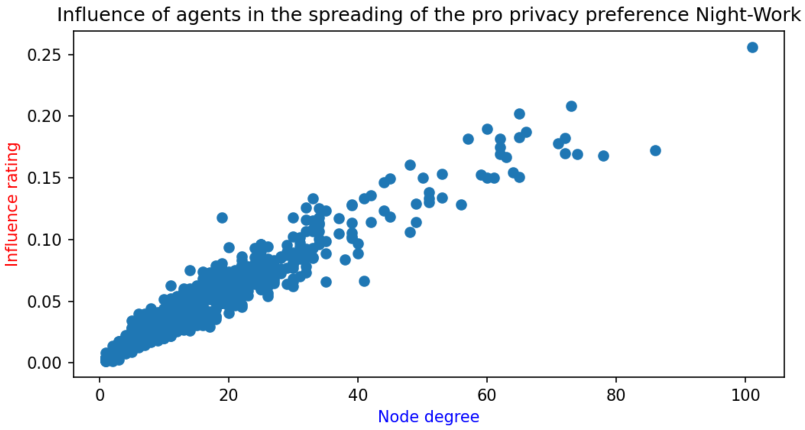

Figure 25: Average agent influence on the spread of the pro side $< night, work >$ privacy preference with pro infection rate 0.75, pro recovery rate 0.25, anti infection rate 0.25, anti recovery rate 0.75. Captured from results of the simulations in which trust is a static random value. Plotted against the degree of the agent on the social network

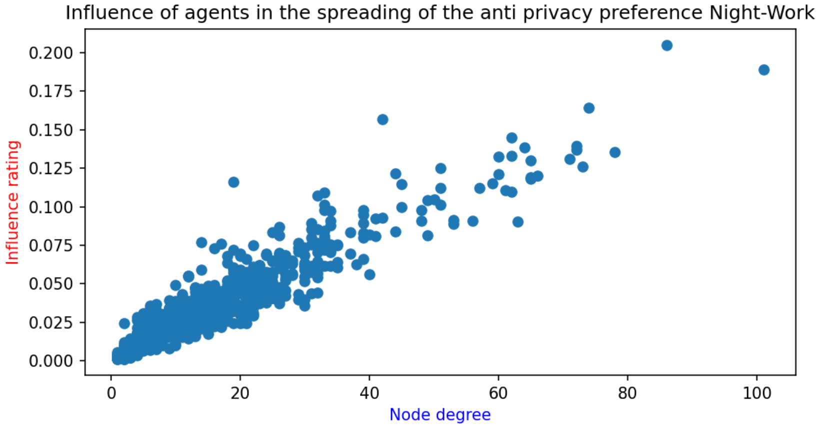

Figure 26: Average agent influence on the spread of the pro side $< night, work >$ privacy preference with pro infection rate 0.25, pro recovery rate 0.75, anti infection rate 0.75, anti recovery rate 0.25. Captured from results of the simulations in which trust is a static random value. Plotted against the degree of the agent on the social network

The same statistical test is used to determine whether a privacy preference epidemic endings sooner when static trust is introduced, Claim 4b. The epidemic endings are the step at which a privacy preference is no longer present on the network as the number of agents infected by it is 0. Including all 81 parameter settings would be tedious and, thus, only the significant results are shown in Table 10. All test results that are not significant will be included in the appendix. There are 30 out of 81 parameter settings for which the claim holds. Thus, in 30 out of 81 parameter settings, the epidemics end sooner when agents hold static trust values towards each other compared to the baseline simulations in which trust is not a factor.

Finally, Claim 4c, it was expected that the peak in the number of infected agents would be lower in this setting compared to the baseline. To investigate this, all epidemic peaks are collected over all parameter settings for all privacy preferences. Using a paired Wilcoxon signed rank test, it is tested whether the epidemic peak is, in fact, lower when static trust values are introduced. The results of these tests are shown in Table 11. It shows that there are 53 parameter settings for which the claim holds. Thus, in 53 out of 81 parameters settings the peak in the number of infected agents during a privacy preference epidemic is lower when agents hold static trust values towards each other compared to the baseline simulations in which trust is not a factor.

For the claims 4a, 4b and 4c, mixed results appear. While there is overwhelming proof for Claim 4b, the other two claims only holds in some of the 81 parameters settings.

Pro Infection Rate	**Pro Recovery Rate**	**Anti Infection Rate**	**Anti Recovery Rate**	**p-value**	**Statistic**
0.25	0.50	0.25	0.25	0.0014	1375145.5
0.25	0.50	0.25	0.75	0.0000	1316763.0
0.25	0.50	0.50	0.75	0.0063	1129559.5
0.25	0.75	0.25	0.25	0.0000	1522890.5
0.25	0.75	0.25	0.75	0.0005	1074578.0
0.25	0.75	0.50	0.25	0.0000	1652887.0
0.25	0.75	0.75	0.25	0.0000	1711084.5
0.50	0.25	0.25	0.25	0.0008	1411722.5
0.50	0.25	0.25	0.50	0.0013	1357900.5
0.50	0.50	0.25	0.25	0.0057	1292855.0
0.50	0.50	0.25	0.50	0.0000	1327058.5
0.50	0.50	0.25	0.75	0.0000	1459217.5
0.50	0.50	0.50	0.75	0.0018	1171178.5
0.50	0.50	0.75	0.75	0.0047	1116017.5
0.50	0.75	0.25	0.25	0.0073	1254611.5
0.50	0.75	0.25	0.50	0.0004	1193142.5
0.50	0.75	0.25	0.75	0.0000	1163499.0
0.50	0.75	0.50	0.25	0.0018	1289181.0
0.75	0.25	0.25	0.25	0.0000	1415608.0
0.75	0.25	0.25	0.50	0.0000	1538447.0
0.75	0.25	0.50	0.25	0.0160	1404503.0
0.75	0.25	0.50	0.50	0.0401	1323529.5
0.75	0.50	0.25	0.25	0.0239	1294681.5
0.75	0.50	0.25	0.50	0.0009	1290205.0
0.75	0.50	0.25	0.75	0.0000	1539400.0
0.75	0.50	0.50	0.75	0.0000	1246636.0
0.75	0.75	0.25	0.25	0.0049	1174675.5
0.75	0.75	0.25	0.50	0.0002	1160257.5
0.75	0.75	0.25	0.75	0.0000	1254801.0
0.75	0.75	0.50	0.75	0.0148	995425.0

Table 10: Significant ($p < 0.05$) results of paired Wilcoxon signed rank tests that test whether an epidemic, on the pro side, ends sooner when agents hold static trust values towards instead of no inclusion of trust at all in the baseline. All parameter settings that are not included returned insignificant results.

Pro Infection Rate	Pro Recovery Rate	Anti Infection Rate	Anti Recovery	p-value	Statistic
0.25	0.25	0.25	0.25	0.0000	389536.0
0.25	0.25	0.25	0.50	0.0000	614583.5
0.25	0.25	0.25	0.75	0.0000	720597.0
0.25	0.25	0.50	0.75	0.0000	670555.5
0.25	0.25	0.75	0.75	0.0000	568809.0
0.25	0.75	0.25	0.75	0.0401	362132.5
0.50	0.25	0.25	0.25	0.0000	719361.5
0.50	0.25	0.25	0.50	0.0000	720600.0
0.50	0.25	0.25	0.75	0.0000	720600.0
0.50	0.25	0.50	0.25	0.0000	576418.5
0.50	0.25	0.50	0.50	0.0000	716883.5
0.50	0.25	0.50	0.75	0.0000	720600.0
0.50	0.25	0.75	0.25	0.0058	377975.5
0.50	0.25	0.75	0.50	0.0000	599863.5
0.50	0.25	0.75	0.75	0.0000	720600.0
0.50	0.50	0.25	0.25	0.0000	614542.5
0.50	0.50	0.25	0.50	0.0000	676682.0
0.50	0.50	0.25	0.75	0.0000	716045.5
0.50	0.50	0.50	0.25	0.0069	378482.0
0.50	0.50	0.50	0.50	0.0000	415646.5
0.50	0.50	0.50	0.75	0.0000	477696.0
0.50	0.50	0.75	0.75	0.0041	376865.5
0.50	0.75	0.25	0.25	0.0000	416714.0
0.50	0.75	0.25	0.50	0.0000	462744.5
0.50	0.75	0.25	0.75	0.0000	500557.5
0.50	0.75	0.75	0.50	0.0307	362922.5
0.75	0.25	0.25	0.25	0.0000	720581.0
0.75	0.25	0.25	0.50	0.0000	720600.0
0.75	0.25	0.25	0.75	0.0000	720600.0
0.75	0.25	0.50	0.25	0.0000	718979.5
0.75	0.25	0.50	0.50	0.0000	720598.5
0.75	0.25	0.50	0.75	0.0000	720599.0
0.75	0.25	0.75	0.25	0.0000	662259.5
0.75	0.25	0.75	0.50	0.0000	720572.0
0.75	0.25	0.75	0.75	0.0000	720599.0
0.75	0.50	0.25	0.25	0.0000	719168.0
0.75	0.50	0.25	0.50	0.0000	720578.0
0.75	0.50	0.25	0.75	0.0000	720600.0
0.75	0.50	0.50	0.25	0.0000	606398.0
0.75	0.50	0.50	0.50	0.0000	669907.5
0.75	0.50	0.50	0.75	0.0000	712045.5
0.75	0.50	0.75	0.25	0.0000	456615.5
0.75	0.50	0.75	0.50	0.0000	505900.0
0.75	0.50	0.75	0.75	0.0000	609258.0
0.75	0.75	0.25	0.25	0.0000	667746.5
0.75	0.75	0.25	0.50	0.0000	700142.5
0.75	0.75	0.25	0.75	0.0000	713277.5
0.75	0.75	0.50	0.25	0.0000	478361.5
0.75	0.75	0.50	0.50	0.0000	529129.0
0.75	0.75	0.50	0.75	0.0000	570944.5
0.75	0.75	0.75	0.25	0.0001	387697.5
0.75	0.75	0.75	0.50	0.0003	390430.0
0.75	0.75	0.75	0.75	0.0000	403065.5

Table 11: Significant ($p < 0.05$) results of paired Wilcoxon signed rank tests that test whether an epidemic, on the pro side, peaks lower when agents hold static trust values towards instead of no inclusion of trust at all in the baseline. All parameter settings that are not included returned insignificant results.

6.1.5 Dynamic Trust

In this simulation setting, the agents keep track of privacy violations during simulations and use these records to calculate trust values they have towards other agents they are connected to. The two different types of privacy violations, described previously, hold the same weight in this setting. The impact of epidemics is expected to be reduced with the introduction of dynamic trust values.

Figures 27 and 28, show the aggregated state dynamics of the simulations from this setting in the most favourable and least favourable setting respectively. In the most favourable setting, it is noticeable that the rise of infected agents is less steep than in the same graph from the baseline simulations. This is likely due to the fact that infection rates are limited by the introduction of trust. Similarly to the static trust setting, it is noticeable that in the least favourable setting the number of recovered agent never reaches the same height as the same graph from the baseline simulations. This is likely due to the fact that infection are so limited and the number of infected only really decreases over time and so there are no agents that need to recover.

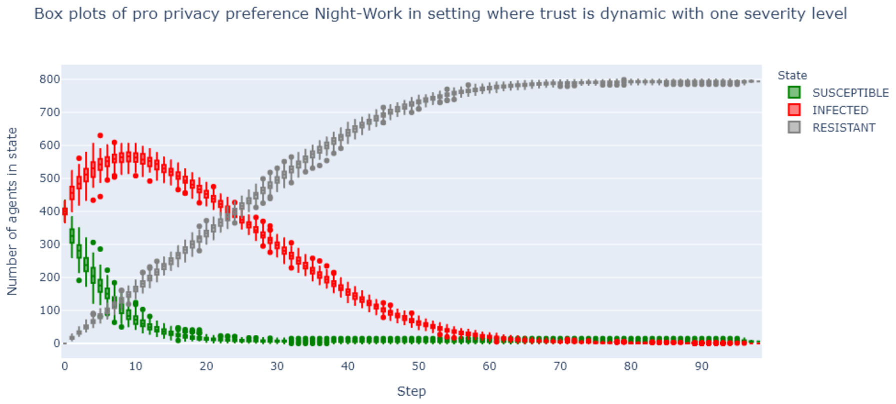

Figure 27: Dynamic of the pro side $< night, work >$ privacy preference with pro infection rate 0.75, pro recovery rate 0.25, anti infection rate 0.25, anti recovery rate 0.75. Taken from simulations in which agent holds dynamic trust values derived from single level privacy violations.

In terms of influence, Claim 5b, the agents in these simulations were expected to have less influence on the spread of privacy preferences. Using a paired Wilcoxon signed rank test, this claim is tested for each parameter setting. The results show that this claim holds for all but one parameter setting. When the pro infection rate, pro recovery rate, anti infection rate and anti recovery rate are 0.75, 0.25, 0.25 and 0.75, respectively, the influence of agents on the spread privacy preferences is not greater in the baseline setting than in this setting dynamic trust values with single level privacy violations. This setting is once again the most favourable setting for the pro side epidemics. Thus, in 80 out of 81 parameters settings the influence ratings of agents in the spread of privacy

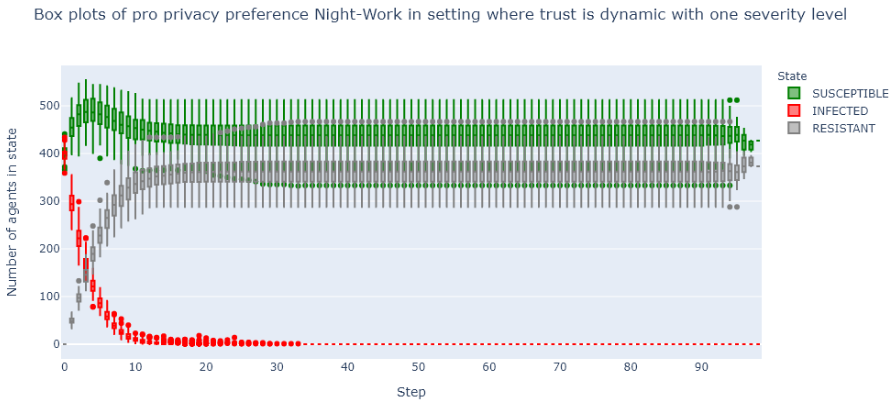

Figure 28: Dynamic of the pro side $< night, work >$ privacy preference with pro infection rate 0.75, pro recovery rate 0.25, anti infection rate 0.25, anti recovery rate 0.75. Taken from simulations in which agent holds dynamic trust values derived from single level privacy violations.

preferences are lower when agents hold dynamic trust values towards each other compared to the baseline simulations in which trust is not a factor at all.

With regard to epidemic endings and peaks, claims 5a and 5c, these were expected to earlier and lower, respectively, compared to the baseline setting. Using a paired Wilcoxon signed rank test, this claim is tested for each parameter setting. The significant results of these tests can be seen in tables 12 and 13. All test results that are not significant will be included in the appendix. In the tables, we can observe that claims 5a and 5c do not hold in all parameter settings. Claim 5a holds in 20 out of 81 parameter settings. Claim 5c holds in 45 out of 81 parameter settings.

Thus, in 20 out of 81 parameters settings the privacy preference epidemics end sooner when agents hold dynamic trust values towards each other compared to the baseline simulations in which trust is not a factor at all. Furthermore, in 45 out of 81 parameters settings the peak in the number of infected agents during a privacy preference epidemic is lower when agents hold dynamic trust values towards each other compared to the baseline simulations in which trust is not a factor at all. It follows that these two hypothesis cannot be accepted for all parameter settings and more investigation is needed to find what the relationship is between the parameters settings and whether these claims hold.

Pro Infection Rate	Pro Recovery Rate	Anti Infection Rate	Anti Recovery Rate	p-value	Statistic
0.25	0.50	0.25	0.25	0.0000	1429791.0
0.25	0.75	0.25	0.25	0.0000	1465631.0
0.50	0.25	0.25	0.25	0.0016	1411103.0
0.75	0.25	0.25	0.25	0.0332	1373228.5
0.75	0.25	0.25	0.50	0.0012	1383678.0
0.50	0.50	0.25	0.75	0.0002	1275602.5
0.75	0.50	0.25	0.75	0.0169	1279388.0
0.25	0.50	0.50	0.25	0.0000	1428274.5
0.25	0.75	0.50	0.25	0.0000	1729185.0
0.50	0.75	0.50	0.25	0.0000	1311028.5
0.75	0.25	0.50	0.25	0.0421	1397084.5
0.75	0.75	0.50	0.25	0.0036	1211893.0
0.25	0.75	0.50	0.50	0.0000	1322085.5
0.50	0.50	0.50	0.75	0.0043	1146188.5
0.75	0.25	0.50	0.75	0.0023	1289108.0
0.75	0.50	0.50	0.75	0.0007	1220783.5
0.25	0.50	0.75	0.25	0.0000	1549867.5
0.25	0.75	0.75	0.25	0.0000	1882449.5
0.50	0.75	0.75	0.25	0.0000	1363735.5
0.25	0.75	0.75	0.50	0.0000	1396217.5

Table 12: Significant ($p < 0.05$) results of paired Wilcoxon signed rank tests that test whether an epidemic, on the pro side, ends sooner when agents keep track of privacy violations to derive dynamic trust values towards instead of no inclusion of trust at all in the baseline. All parameter settings that are not included returned insignificant results.

Pro Infection Rate	Pro Recovery Rate	Anti Infection Rate	Anti Recovery Rate	p-value	Statistic
0.25	0.25	0.25	0.25	0.0026	379229.0
0.25	0.25	0.25	0.50	0.0000	584633.0
0.25	0.25	0.25	0.75	0.0000	716551.5
0.25	0.25	0.50	0.75	0.0000	673631.0
0.25	0.25	0.75	0.75	0.0000	555729.5
0.25	0.75	0.50	0.50	0.0401	367650.0
0.50	0.25	0.25	0.25	0.0000	620028.0
0.50	0.25	0.25	0.50	0.0000	688641.0
0.50	0.25	0.25	0.75	0.0000	696939.5
0.50	0.25	0.50	0.25	0.0000	491182.5
0.50	0.25	0.50	0.50	0.0000	688366.5
0.50	0.25	0.50	0.75	0.0000	698815.0
0.50	0.25	0.75	0.50	0.0000	560635.5
0.50	0.25	0.75	0.75	0.0000	707288.5
0.50	0.50	0.25	0.25	0.0000	430332.0
0.50	0.50	0.25	0.50	0.0000	508052.5
0.50	0.50	0.25	0.75	0.0000	606747.5
0.50	0.50	0.50	0.50	0.0159	374779.5
0.50	0.50	0.50	0.75	0.0000	431475.5
0.50	0.75	0.25	0.50	0.0197	373143.0
0.50	0.75	0.25	0.75	0.0010	380889.0
0.75	0.25	0.25	0.25	0.0000	598686.5
0.75	0.25	0.25	0.50	0.0000	634084.0
0.75	0.25	0.25	0.75	0.0000	623018.0
0.75	0.25	0.50	0.25	0.0000	583421.5
0.75	0.25	0.50	0.50	0.0000	656580.5
0.75	0.25	0.50	0.75	0.0000	650906.5
0.75	0.25	0.75	0.25	0.0000	515754.0
0.75	0.25	0.75	0.50	0.0000	657032.0
0.75	0.25	0.75	0.75	0.0000	670303.0
0.75	0.50	0.25	0.25	0.0000	486273.0
0.75	0.50	0.25	0.50	0.0000	534984.0
0.75	0.50	0.25	0.75	0.0000	583811.5
0.75	0.50	0.50	0.25	0.0000	412814.0
0.75	0.50	0.50	0.50	0.0000	470826.5
0.75	0.50	0.50	0.75	0.0000	547485.5
0.75	0.50	0.75	0.25	0.0171	368852.0
0.75	0.50	0.75	0.50	0.0000	397294.5
0.75	0.50	0.75	0.75	0.0000	482050.5
0.75	0.75	0.25	0.25	0.0000	411639.5
0.75	0.75	0.25	0.50	0.0000	442081.0
0.75	0.75	0.25	0.75	0.0000	469817.0
0.75	0.75	0.50	0.50	0.0223	370046.0
0.75	0.75	0.50	0.75	0.0000	417810.5
0.75	0.75	0.75	0.75	0.0020	378351.5

Table 13: Significant ($p < 0.05$) results of paired Wilcoxon signed rank tests that test whether an epidemic, on the pro side, peaks lower when agents keep track of privacy violations to derive dynamic trust values towards instead of no inclusion of trust at all in the baseline. All parameter settings that are not included returned insignificant results.

6.1.6 Dynamic trust with two levels of privacy violations

In this simulation setting, the agents keep track of privacy violations during simulations and use these records to calculate trust values they have towards other agents they are connected to. The two different types of privacy violations, described previously, hold different weights in this setting. The impact of epidemics is expected to be reduced with the introduction of dynamic trust values.

Figures 29 and 30, show the aggregated state dynamics of the simulations from this setting in the most favourable and least favourable setting respectively. In the most favourable setting, it is noticeable that the rise of infected agents is less steep than in the same graph from the baseline simulations. This is likely due to the fact that infection rates are limited by the introduction of trust. When the pro infection rate is 0.75, it is noticeable that the epidemic peak is lower than in the setting where there was only one severity for all privacy violations.

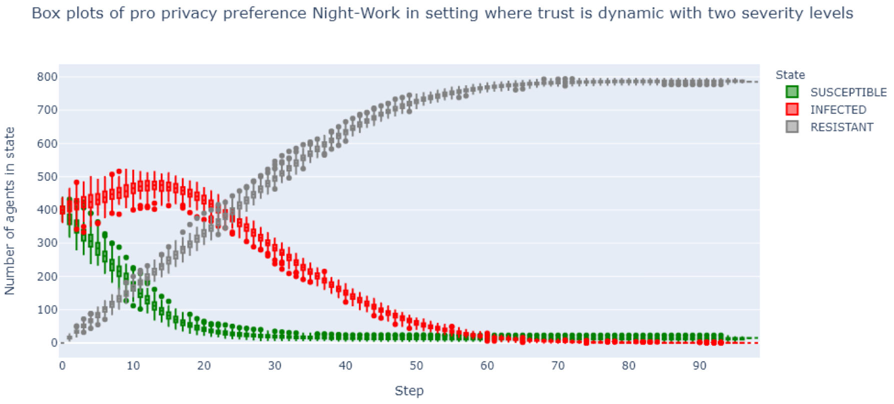

Figure 29: Dynamic of the pro side $< night, work >$ privacy preference with pro infection rate 0.75, pro recovery rate 0.25, anti infection rate 0.25, anti recovery rate 0.75. Taken from simulations in which agent holds dynamic trust values derived from two level privacy violations.

In terms of influence, Claim 6b, the agents in these simulations were expected to have less influence on the spread of privacy preferences. Using a paired Wilcoxon signed rank test, this claim is tested for each parameter setting. The results show that this claim holds for all parameters. Thus, all agents have less influence when dynamic trust values are used that are derived from two levels of privacy violations compared to the baseline simulations in which trust is not a factor at all.

With regard to epidemic endings and peaks, claims 6a and 6c, these were expected to earlier and lower, respectively, compared to the baseline setting. Using a paired Wilcoxon signed rank test, this claim is tested for each parameter setting. The significant results of these tests can be seen in tables 14 and 15. All test results that are not significant will be included in the appendix. In

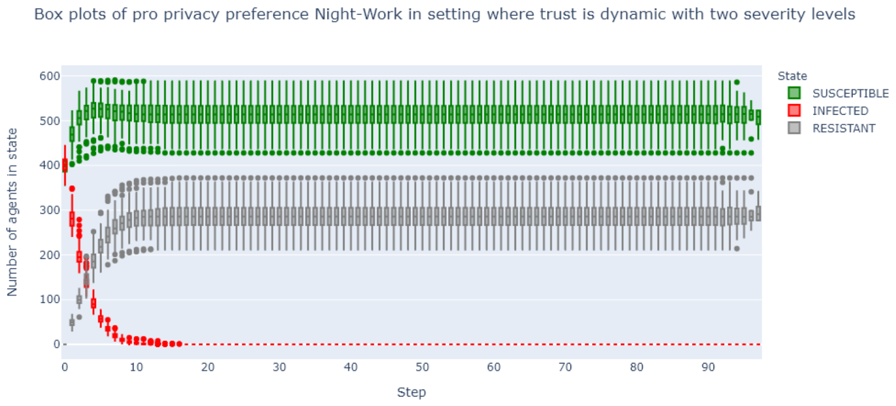

Figure 30: Dynamic of the pro side $< night, work >$ privacy preference with pro infection rate 0.75, pro recovery rate 0.25, anti infection rate 0.25, anti recovery rate 0.75. Taken from simulations in which agent holds dynamic trust values derived from two level privacy violations.

these tables, we can observe that claims 6a and 6c do not hold in all parameter settings. Claim 6a holds in 34 out of 81 parameter settings. Claim 6c holds in 53 out of 81 parameter settings. Thus, in 34 out of 81 parameters settings the privacy preference epidemics end sooner when agents hold dynamic trust values towards each other compared to the baseline simulations in which trust is not a factor at all. Furthermore, in 53 out of 81 parameters settings the peak in the number of infected agents during a privacy preference epidemic is lower when agents hold dynamic trust values, that are derived from two levels of privacy violation, towards each other compared to the baseline simulations in which trust is not a factor at all. It follows that these two hypothesis cannot be accepted for all parameter settings and more investigation is needed to find what the relationship is between the parameters settings and whether these claims hold.

Pro Infection Rate	**Pro Recovery Rate**	**Anti Infection Rate**	**Anti Recovery Rate**	**p-value**	**Statistic**
0.25	0.50	0.25	0.25	0.0000	1748632.5
0.25	0.50	0.25	0.75	0.0000	1301283.5
0.25	0.50	0.50	0.25	0.0000	1858450.5
0.25	0.50	0.75	0.25	0.0000	2062514.0
0.25	0.75	0.25	0.25	0.0000	1828664.0
0.25	0.75	0.25	0.50	0.0000	1633911.0
0.25	0.75	0.25	0.75	0.0000	1196497.5
0.25	0.75	0.50	0.25	0.0000	1986458.0
0.25	0.75	0.50	0.50	0.0000	1888760.5
0.25	0.75	0.75	0.25	0.0000	2063965.5
0.25	0.75	0.75	0.50	0.0000	2089486.0
0.50	0.25	0.25	0.25	0.0065	1398616.5
0.50	0.25	0.25	0.50	0.0001	1377698.0
0.50	0.50	0.25	0.25	0.0000	1409139.0
0.50	0.50	0.25	0.50	0.0000	1329216.0
0.50	0.50	0.25	0.75	0.0000	1455341.5
0.50	0.75	0.25	0.25	0.0000	1407603.0
0.50	0.75	0.25	0.50	0.0242	1165258.5
0.50	0.75	0.25	0.75	0.0000	1069027.5
0.50	0.75	0.50	0.25	0.0000	1622416.5
0.50	0.75	0.75	0.25	0.0000	1686290.5
0.75	0.25	0.25	0.25	0.0083	1382614.0
0.75	0.25	0.25	0.50	0.0000	1579416.0
0.75	0.25	0.25	0.75	0.0047	1322711.5
0.75	0.50	0.25	0.50	0.0000	1270375.0
0.75	0.50	0.25	0.75	0.0000	1601083.0
0.75	0.50	0.50	0.50	0.0212	1215522.0
0.75	0.50	0.50	0.75	0.0000	1291858.5
0.75	0.50	0.75	0.75	0.0197	1119988.0
0.75	0.75	0.25	0.25	0.0000	1317299.0
0.75	0.75	0.25	0.50	0.0404	1116817.0
0.75	0.75	0.25	0.75	0.0000	1238330.5
0.75	0.75	0.50	0.25	0.0000	1330048.0
0.75	0.75	0.75	0.25	0.0000	1408565.5

Table 14: Significant ($p < 0.05$) results of paired Wilcoxon signed rank tests that test whether an epidemic, on the pro side, ends sooner when agents keep track of two levels of privacy violations to derive dynamic trust values towards instead of no inclusion of trust at all in the baseline. All parameter settings that are not included returned insignificant results.

Pro Infection Rate	Pro Recovery Rate	Anti Infection Rate	Anti Recovery Rate	p-value	Statistic
0.25	0.25	0.25	0.25	0.0000	403785.0
0.25	0.25	0.25	0.50	0.0000	612682.5
0.25	0.25	0.25	0.75	0.0000	720599.0
0.25	0.25	0.50	0.75	0.0000	672473.5
0.25	0.25	0.75	0.75	0.0000	565549.0
0.25	0.75	0.75	0.50	0.0293	371148.5
0.50	0.25	0.25	0.25	0.0000	719339.0
0.50	0.25	0.25	0.50	0.0000	720600.0
0.50	0.25	0.25	0.75	0.0000	720600.0
0.50	0.25	0.50	0.25	0.0000	586555.0
0.50	0.25	0.50	0.50	0.0000	716854.0
0.50	0.25	0.50	0.75	0.0000	720600.0
0.50	0.25	0.75	0.25	0.0000	397227.0
0.50	0.25	0.75	0.50	0.0000	592471.5
0.50	0.25	0.75	0.75	0.0000	720600.0
0.50	0.50	0.25	0.25	0.0000	617399.5
0.50	0.50	0.25	0.50	0.0000	681034.0
0.50	0.50	0.25	0.75	0.0000	715144.5
0.50	0.50	0.50	0.25	0.0020	378337.0
0.50	0.50	0.50	0.50	0.0000	417980.0
0.50	0.50	0.50	0.75	0.0000	463582.0
0.50	0.75	0.25	0.25	0.0000	412355.5
0.50	0.75	0.25	0.50	0.0000	454664.0
0.50	0.75	0.25	0.75	0.0000	523773.5
0.50	0.75	0.50	0.25	0.0496	363411.5
0.50	0.75	0.75	0.75	0.0219	372623.0
0.75	0.25	0.25	0.25	0.0000	720600.0
0.75	0.25	0.25	0.50	0.0000	720600.0
0.75	0.25	0.25	0.75	0.0000	720600.0
0.75	0.25	0.50	0.25	0.0000	717325.0
0.75	0.25	0.50	0.50	0.0000	720600.0
0.75	0.25	0.50	0.75	0.0000	720600.0
0.75	0.25	0.75	0.25	0.0000	668418.5
0.75	0.25	0.75	0.50	0.0000	720544.0
0.75	0.25	0.75	0.75	0.0000	720600.0
0.75	0.50	0.25	0.25	0.0000	719020.0
0.75	0.50	0.25	0.50	0.0000	720593.5
0.75	0.50	0.25	0.75	0.0000	720600.0
0.75	0.50	0.50	0.25	0.0000	612469.0
0.75	0.50	0.50	0.50	0.0000	676432.5
0.75	0.50	0.50	0.75	0.0000	715463.5
0.75	0.50	0.75	0.25	0.0000	451849.0
0.75	0.50	0.75	0.50	0.0000	509765.5
0.75	0.50	0.75	0.75	0.0000	609984.5
0.75	0.75	0.25	0.25	0.0000	680696.5
0.75	0.75	0.25	0.50	0.0000	702383.0
0.75	0.75	0.25	0.75	0.0000	712960.0
0.75	0.75	0.50	0.25	0.0000	453909.5
0.75	0.75	0.50	0.50	0.0000	517608.5
0.75	0.75	0.50	0.75	0.0000	563601.5
0.75	0.75	0.75	0.25	0.0000	400012.0
0.75	0.75	0.75	0.50	0.0000	409119.0
0.75	0.75	0.75	0.75	0.0000	416155.0

Table 15: Significant ($p < 0.05$) results of paired Wilcoxon signed rank tests that test whether an epidemic, on the pro side, peaks lower when agents keep track of two levels of privacy violations to derive dynamic trust values towards instead of no inclusion of trust at all in the baseline. All parameter settings that are not included returned insignificant results.

6.2 Pro privacy preference dynamics only

In this section, the results are presented for the simulations in which the opposing privacy preferences states are assumed to only change given infections on the pro side.

6.2.1 Baseline simulations

Similar to before, box plots are utilized to illustrate the state dynamics of each privacy preference. In Figures 31 and 32, the state dynamics of the privacy preference $night, work$ are displayed for a pro infection rate of 0.25 and pro recovery rate of 0.25. Comparing the pro side dynamics between this setting and the baseline simulations with fully dynamic privacy preferences in Figure 9, one can see that in this setting the epidemic grows strongly before declining and this is not the case in the previous baseline simulations.

In terms of the anti side, it is noticeable that the dynamic involves no recoveries only agents becoming infected with the pro preference and thus becoming susceptible to the opposing privacy preference. This can be seen in the fact the dynamic is solely the decline of the number of infected agents mirrored the rise of susceptible agents.

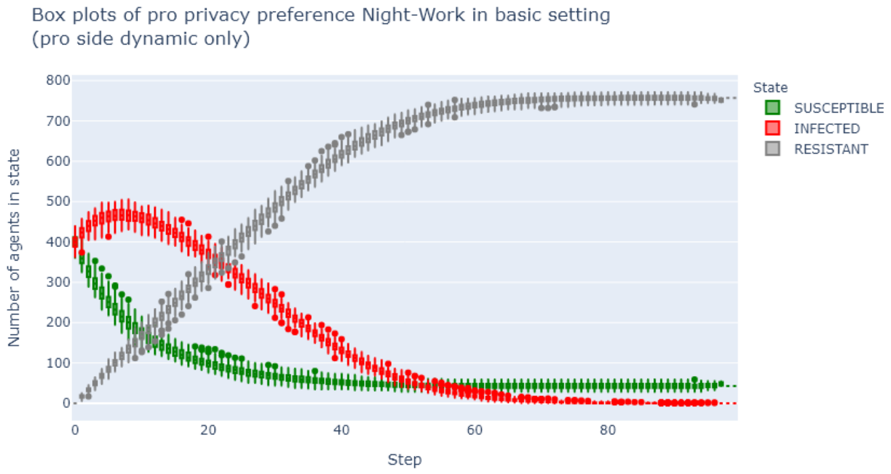

Figure 31: Dynamic of the pro side $< night, work >$ privacy preference with pro infection rate 0.25, pro recovery rate 0.25, anti infection rate 0.25, anti recovery rate 0.25. Taken from simulations in which the opposing privacy preferences states are assumed to only change given infections on the pro side.

In terms of the most favourable setting for the pro side, in Figure 33, it is noticeable that the rise of the number of infected agents is more rapid that when both the infection rate and recovery rate are 0.75. Conversely, the decline of the number of susceptible agents is also more steep. On the contrary in the least favourable setting, in Figure 34, the number of susceptible agents does not reach the same heights as in the baseline simulations with fully dynamic opposing privacy preferences in Figure 11. Furthermore, the epidemic seems to also last longer.

After these initials observations, the hypotheses for these simulations were tested.

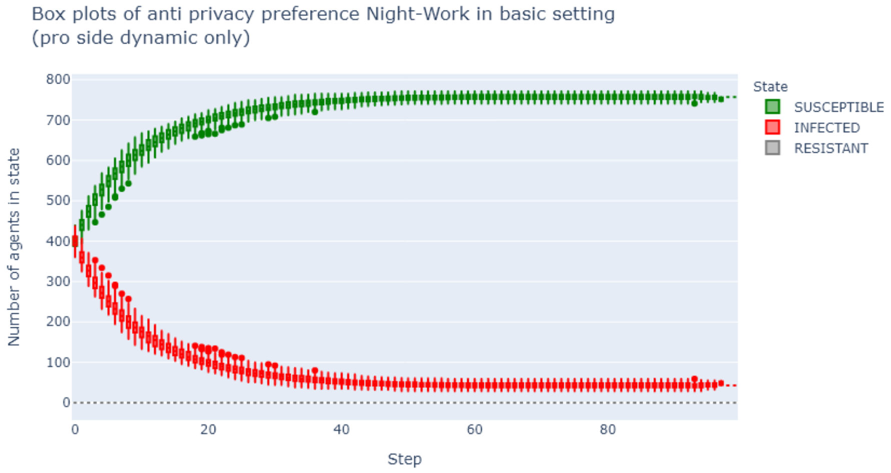

Figure 32: Dynamic of the anti side $< night, work >$ privacy preference with pro infection rate 0.25, pro recovery rate 0.25. Taken from simulations in which the opposing privacy preferences states are assumed to only change given infections on the pro side.

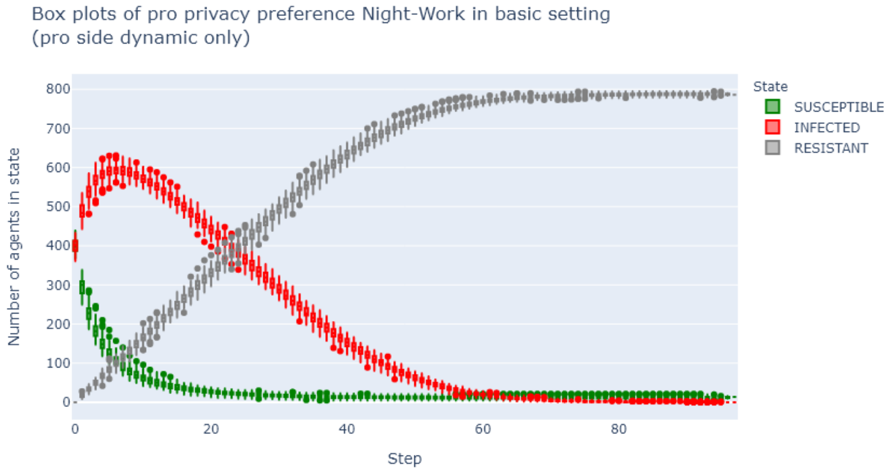

Figure 33: Dynamic of the pro side $< night, work >$ privacy preference with pro infection rate 0.75, pro recovery rate 0.25. Taken from simulations in which the opposing privacy preferences states are assumed to only change given infections on the pro side.

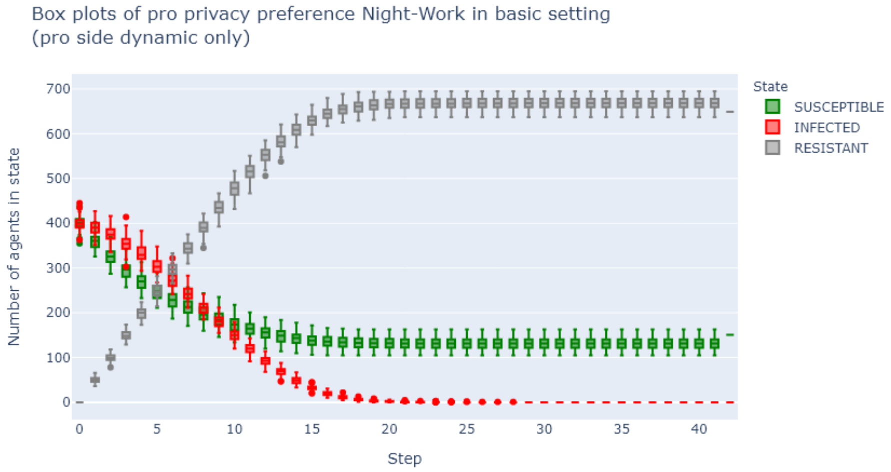

Figure 34: Dynamic of the pro side $< night, work >$ privacy preference with pro infection rate 0.25, pro recovery rate 0.75. Taken from simulations in which the opposing privacy preferences states are assumed to only change given infections on the pro side.

6.2.1.1 Infection rate vs. Recovery rate

It is expected that privacy preference epidemics last longer when the infection rate of a privacy preference is higher than the recovery rate of said privacy preference. To test this Claim 1a, once again, all epidemic endings are recorded and using a paired Wilcoxon signed rank test. The results of these tests can be seen in Table 16. For all privacy preferences, it is found that their epidemics end sooner when the infection rate is higher than the recovery rate.

Scenario	Privacy Preference	Statistic	p-value
1	Afternoon-Beach	34662.5	0.0000
2	Afternoon-Beach	41586.0	0.0000
2	Afternoon-Mall	42437.5	0.0000
1	Afternoon-Mall	36873.5	0.0000
1	Afternoon-Work	38506.5	0.0000
2	Afternoon-Work	43430.0	0.0000
1	Evening-Beach	37018.5	0.0000
2	Evening-Beach	42240.0	0.0000
2	Evening-Mall	42785.0	0.0000
1	Evening-Mall	37459.5	0.0000
1	Evening-Work	36771.0	0.0000
2	Evening-Work	42673.0	0.0000
1	Morning-Beach	36854.0	0.0000
2	Morning-Beach	42763.0	0.0000
1	Morning-Mall	37863.5	0.0000
2	Morning-Mall	42984.5	0.0000
1	Morning-Work	36470.0	0.0000
2	Morning-Work	42010.5	0.0000
1	Night-Beach	36433.5	0.0000
2	Night-Beach	41933.0	0.0000
1	Night-Mall	39337.5	0.0000
2	Night-Mall	43736.0	0.0000
1	Night-Work	37007.0	0.0000
2	Night-Work	42107.5	0.0000

Table 16: Results of paired Wilcoxon signed rank tests that test whether an infection, on the pro side, ends later when the infection rate is higher than the recovery rate of privacy preferences. The alternatives tested against are when the infection rate is the same as the recovery rate (scenario 1) and when the infection rate is lower than the recovery rate (scenario 2)

6.2.1.2 Agent Influence

To investigate Claim 1b, the influence rating measure, previously defined, is used. In Figures 35 and 36, the influence ratings are plotted against the node degree of an agent on the social network. The graphs suggest the existence of the aforementioned positive correlation between an agent's degree in a social network and its influence on the spread of a privacy preference.

Using the captured, the Pearson correlation coefficient is calculated along with the p-value. For all parameter settings, a positive correlation coefficient is

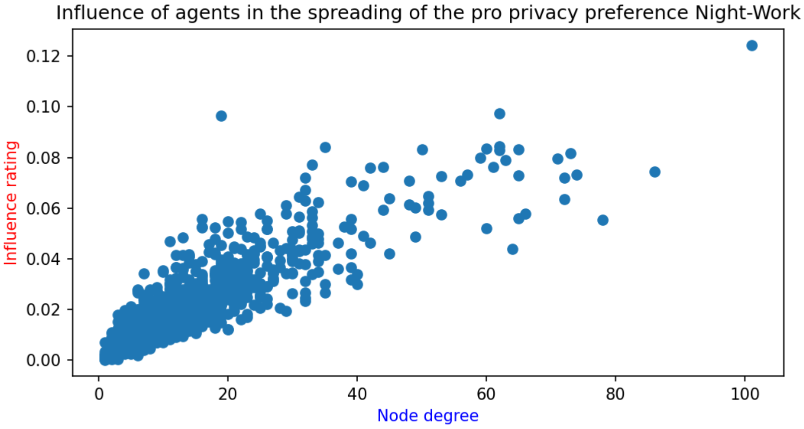

Figure 35: Average agent influence on the spread of the pro side $< night, work >$ privacy preference with pro infection rate 0.75, pro recovery rate 0.25. Captured from results of the baseline simulations in which the opposing privacy preferences states are assumed to only change given infections on the pro side. Plotted against the degree of the agent on the social network

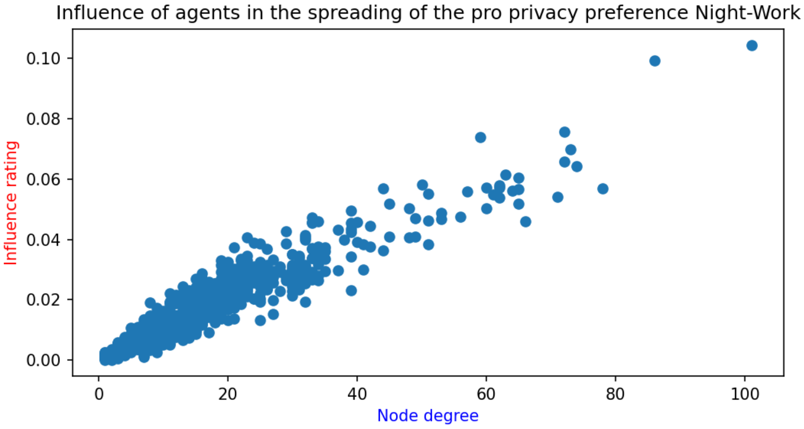

Figure 36: Average agent influence on the spread of the pro side $< night, work >$ privacy preference with pro infection rate 0.25, pro recovery rate 0.75. Captured from results of the baseline simulations in which the opposing privacy preferences states are assumed to only change given infections on the pro side. Plotted against the degree of the agent on the social network

found. Furthermore, all correlation coefficients come with $p < 0.05$. Thus, it is once again the case that there is a positive correlation between the degree of an agent in a social network and their influence in the spread of privacy preferences in the DIPP model. The results can be found in Table 17.

Pro Infection Rate	Pro Recovery Rate	r	p-value
0.25	0.25	0.3402	0.0
0.25	0.05	0.3334	0.0
0.25	0.75	0.3234	0.0
0.05	0.25	0.3112	0.0
0.05	0.05	0.3115	0.0
0.05	0.75	0.3057	0.0
0.75	0.25	0.2899	0.0
0.75	0.05	0.2893	0.0
0.75	0.75	0.2876	0.0

Table 17: Pearson correlation coefficients between the influence rating of an agent and the degree of said agent on the social network.

6.2.2 Rare privacy preference with high degree influencers

Experiments in this setting are performed to investigate the ability of a few of the most influential users to spread privacy preferences that are rare. Specifically, the top 1% of agents, with regard to degree, are assigned privacy preference infections of a privacy preferences with an initial spread of 0.2% of the agents in the network. The influence ratings of these agents are analysed to see whether they are more influential in this setting that in the baseline simulations. However, now this happens with fully dynamic opposing privacy preferences.

The $< night, work >$ pro privacy preference is the rare privacy preference in this experimental setting. In Figures 37 and 38, the aggregates of the state dynamics of the $< night, work >$ pro privacy preference in the most two infection and recovery settings extreme settings. Compared to the baseline simulations, it is noticeable that the peak in the number of infected agents happens later in the most favourable setting. Furthermore, in the least favourable setting, there is a rise in the epidemic before dying out, something that is not the case in the baseline results.

With regard to Claim 2a, the question is whether these top 1% of agents are more influential in this setting than in the baseline simulations. In Figures 39 and 42, the comparison between the average influence figures between the two settings show that the top 1% of agents seem to enjoy more influence when the $night, work$ privacy preference is rare. For further proof, a paired Wilcoxon signed rank test is used to test whether the claim that the influence of the most influential agents rises when the privacy preference is rare. The results of the statistical tests are shown in Table 18. It is clear that for all parameter settings it is found that the influence of the top 1% of agents, with regard to degree, rises when the privacy preference they hold becomes rare.

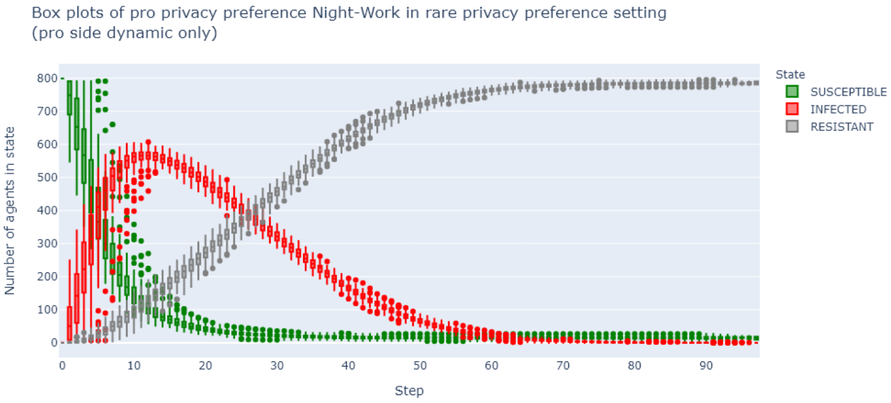

Figure 37: Dynamic of the pro side $< night, work >$ privacy preference with pro infection rate 0.75, pro recovery rate 0.25. Taken from simulations in which the opposing privacy preferences states are assumed to only change given infections on the pro side.

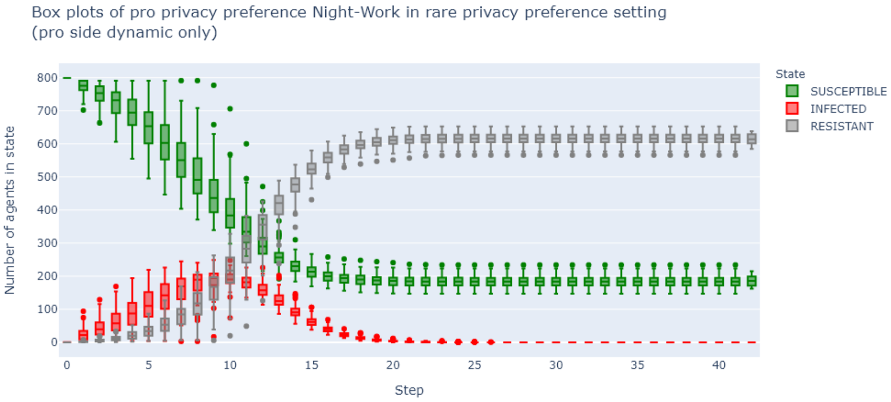

Figure 38: Dynamic of the pro side $< night, work >$ privacy preference with pro infection rate 0.25, pro recovery rate 0.75. Taken from simulations in which the opposing privacy preferences states are assumed to only change given infections on the pro side.

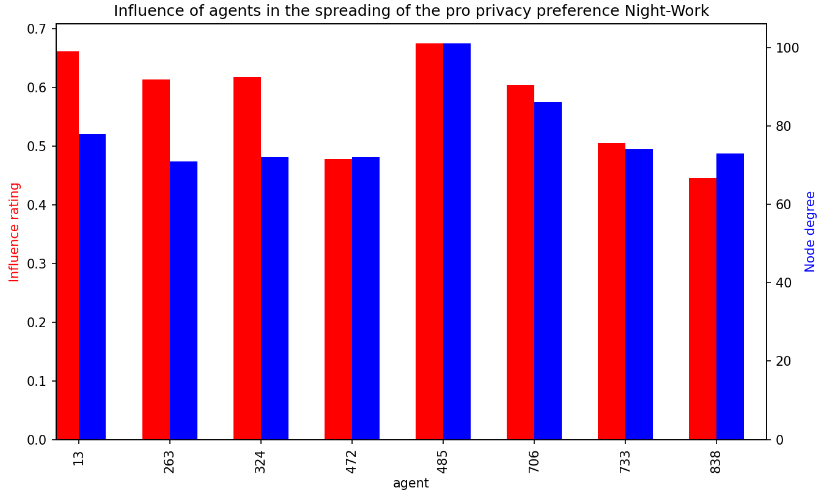

Figure 39: Average agent influence on the spread of the pro side $< night, work >$ privacy preference with pro infection rate 0.75, pro recovery rate 0.25. $night, work$ is a rare privacy preference in this setting. Plotted against the degree of the agent on the social network.

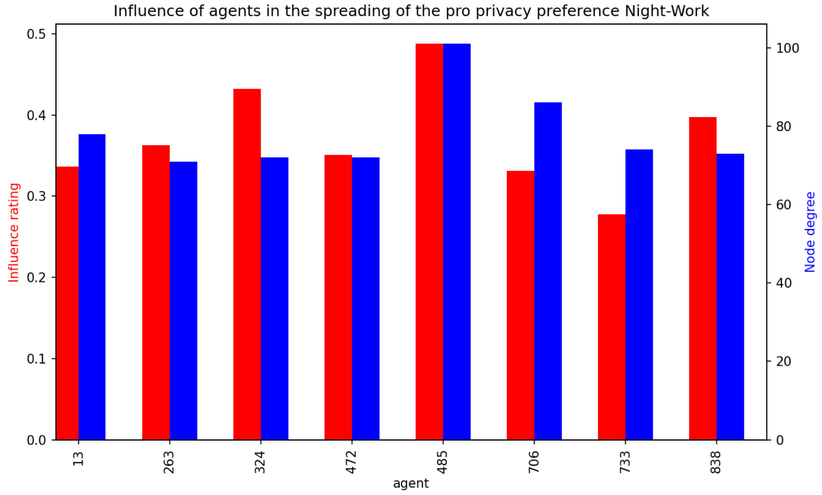

Figure 40: Average agent influence on the spread of the pro side $< night, work >$ privacy preference with pro infection rate 0.25, pro recovery rate 0.75. $night, work$ is a rare privacy preference in this setting. Plotted against the degree of the agent on the social network.

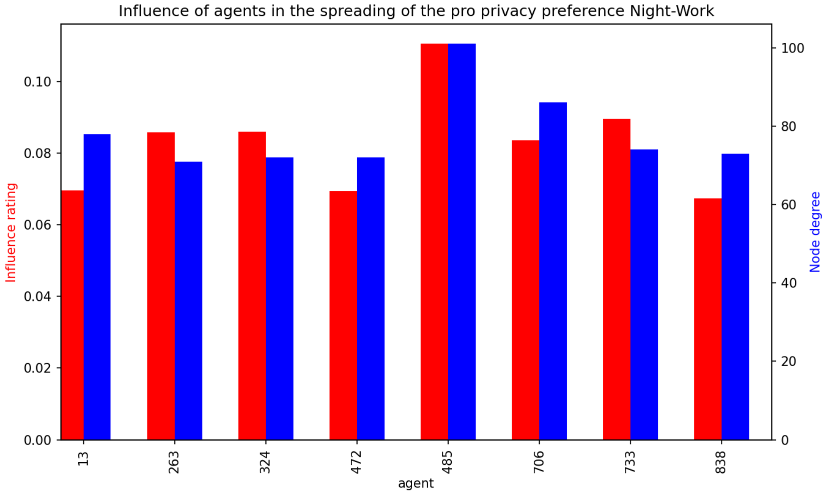

Figure 41: Average agent influence on the spread of the pro side $< night, work >$ privacy preference with pro infection rate 0.75, pro recovery rate 0.25. Captured from baseline simulations. Plotted against the degree of the agent on the social network.

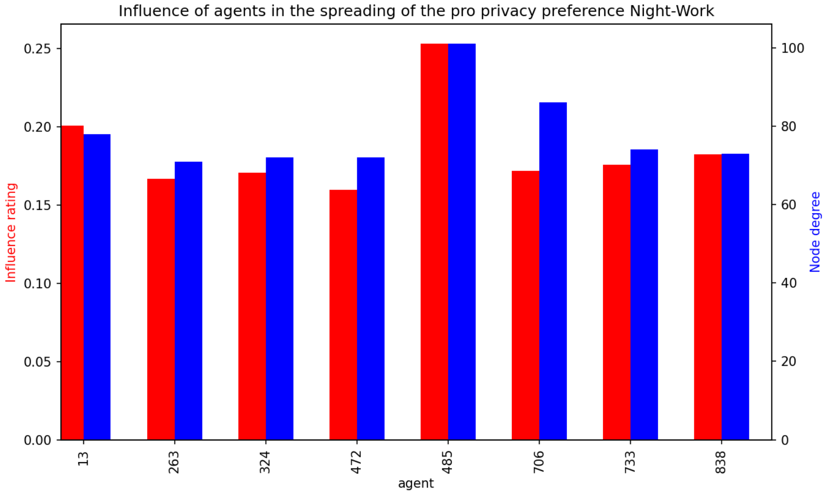

Figure 42: Top 1% of agents, with regard to degree, and their average agent influence on the spread of the pro side $< night, work >$ privacy preference with pro infection rate 0.25, pro recovery rate 0.75. Captured from baseline simulations. Plotted against the degree of the agent on the social network.

Pro Infection Rate	Pro Recovery Rate	p-value	Statistic
0.25	0.25	0.0000	25284.5
0.25	0.50	0.0000	27863.5
0.25	0.75	0.0000	23778.5
0.50	0.25	0.0000	32823.0
0.50	0.50	0.0000	33731.0
0.50	0.75	0.0000	28941.5
0.75	0.25	0.0000	32851.0
0.75	0.50	0.0000	43561.5
0.75	0.75	0.0000	39626.5

Table 18: Results for the tests whether the influence of the top 1% agents, with regard to degree, is higher when the privacy preference $night, work$ becomes rare

6.2.3 No Trust

In this no trust setting, the agents hold a static trust value of $tr = 0$ towards each other. No new infections are expected in this setting. In Figures 43 and 44, we can see this expectation come true. After step 0, agents only become immune to the privacy preference $night, work$ and none of the susceptible agents get infected.

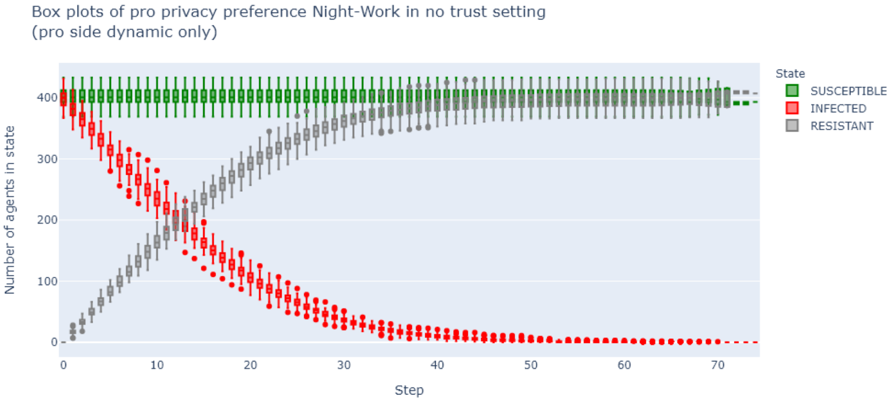

Figure 43: Dynamic of the pro side $< night, work >$ privacy preference with pro infection rate 0.75, pro recovery rate 0.25. Taken from simulations in which the opposing privacy preferences states are assumed to only change given infections on the pro side.

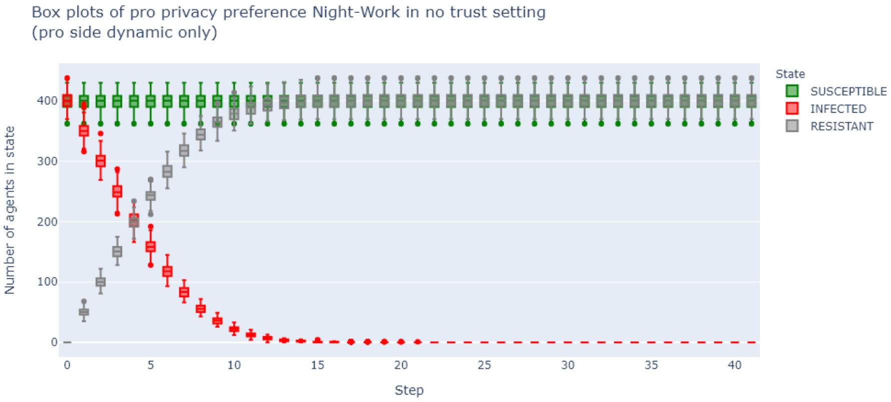

Figure 44: Dynamic of the pro side $< night, work >$ privacy preference with pro infection rate 0.25, pro recovery rate 0.75. Taken from simulations in which the opposing privacy preferences states are assumed to only change given infections on the pro side.

6.2.4 Static Trust

While in the no trust setting agents do not trust each other at all, in this setting agents a static randomly chosen trust value to each of their friends on the social network. In the Figures 45 and 46, the aggregated state dynamics for privacy preference $night, work$ are shown, for the most favourable and least favourable parameter settings.

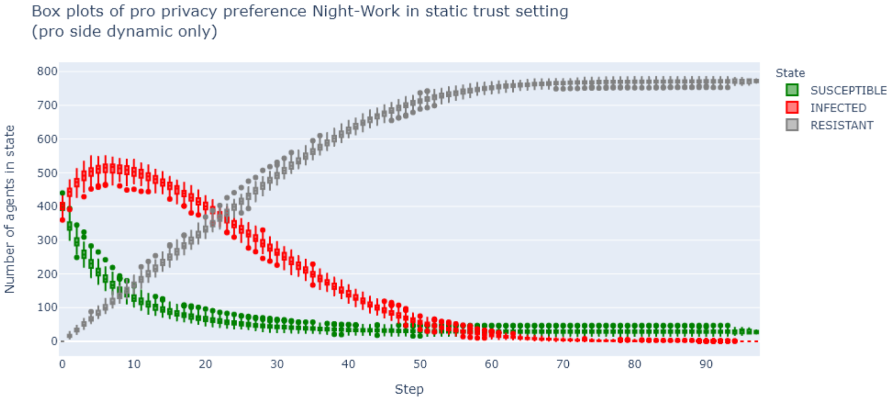

Figure 45: Dynamic of the pro side $< night, work >$ privacy preference with pro infection rate 0.75, pro recovery rate 0.25. Taken from simulations in which the opposing privacy preferences states are assumed to only change given infections on the pro side.

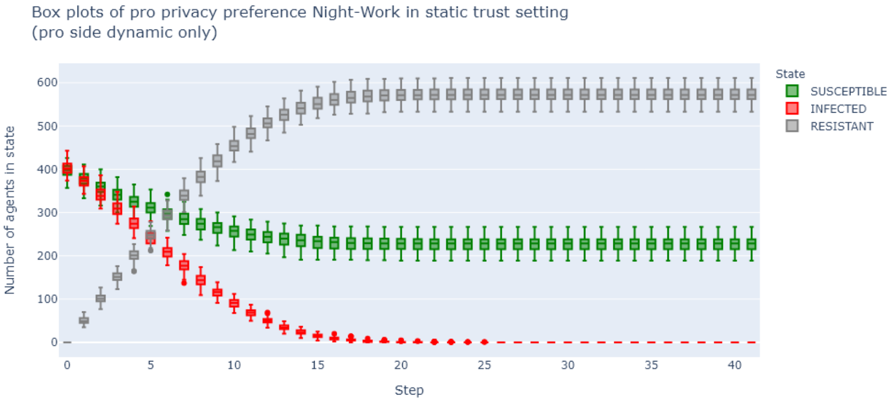

Figure 46: Dynamic of the pro side $< night, work >$ privacy preference with pro infection rate 0.25, pro recovery rate 0.75. Taken from simulations in which the opposing privacy preferences states are assumed to only change given infections on the pro side.

It was expected that the epidemics in this setting would end sooner than those in the baseline simulations, Claim 4a. To test this claim, the steps, at which epidemics end, are compared using a paired Wilcoxon signed rank test. The results of these tests can be seen in Table 19. For all settings, there is a statistically significant result to the query whether epidemics end sooner when static trust is included in the simulation. Thus, the epidemics ends sooner than in the baseline simulations.

Pro Infection Rate	**Pro Recovery Rate**	**p-value**	**Statistic**
0.25	0.25	0.0000	387432.00
0.25	0.50	0.0000	428161.00
0.25	0.75	0.0000	423688.00
0.50	0.25	0.0000	373025.50
0.50	0.50	0.0000	371188.50
0.50	0.75	0.0000	319764.00
0.75	0.25	0.0267	341279.50
0.75	0.50	0.0010	319342.00
0.75	0.75	0.0000	278089.00

Table 19: Significant ($p < 0.05$) results of paired Wilcoxon signed rank tests that test whether an epidemic, on the pro side, ends sooner when agents hold random static trust values towards instead of no inclusion of trust at all in the baseline.

With regard to Claim 4b, it was also expected that in this setting the influence of each agent on the spread of privacy preferences would be diminished compared to the baseline simulations. The influence ratings were compared to the ones from the baseline simulations using paired Wilcoxon signed rank tests. The results of these tests can be seen in Table 20. Only in the most favourable setting is there no significant evidence that agents have diminished influence. Thus all agents have a lower influence ratings when they hold static trust values towards their friends compared to the baseline simulations in which trust is not a factor.

Pro Infection Rate	**Pro Recovery Rate**	**p-value**	**Statistic**
0.25	0.25	0.0000	47373126169.50
0.25	0.50	0.0000	36823532911.00
0.25	0.75	0.0000	28252727929.00
0.50	0.25	0.0000	48414344123.00
0.50	0.50	0.0000	43168048322.50
0.50	0.75	0.0000	38068307484.00
0.75	0.25	0.8100	46117640955.00
0.75	0.50	0.0000	42882639707.00
0.75	0.75	0.0000	39561364297.50

Table 20: Significant ($p < 0.05$) results of paired Wilcoxon signed rank tests that test whether agents' influence on the spread of privacy preference is diminished when agents hold random static trust values towards instead of no inclusion of trust at all in the baseline.

Finally, testing Claim 4c, epidemics were also expected to peak lower in this setting than in the baseline simulations. The epidemic peaks were compared to the ones from the baseline simulations using paired Wilcoxon signed rank tests. The results of these tests can be seen in Table 21. Only in the least favourable parameter setting do we find that there is no significant evidence for epidemics peaking lower when agents hold static random trust values.

Pro Infection Rate	**Pro Recovery Rate**	**p-value**	**Statistic**
0.25	0.25	0.0000	720594.5
0.25	0.50	0.0000	573665.5
0.25	0.75	0.0539	364153.5
0.50	0.25	0.0000	720600.0
0.50	0.50	0.0000	719353.0
0.50	0.75	0.0000	697995.0
0.75	0.25	0.0000	720598.0
0.75	0.50	0.0000	720597.5
0.75	0.75	0.0000	719166.0

Table 21: Significant ($p < 0.05$) results of paired Wilcoxon signed rank tests that test whether an epidemic, on the pro side, peaks lower when agents hold static trust values towards instead of no inclusion of trust at all in the baseline. All parameter settings that are not included returned insignificant results.

6.2.5 Dynamic Trust

In this simulation setting, the agents keep track of privacy violations during simulations and use these records to calculate trust values they have towards other agents they are connected to. Privacy violations of all types have the same weights when calculation trust. Figures 47 and 48, show the aggregated state dynamics of the simulations from this setting in the most favourable and least favourable setting respectively. In the most favourable setting, it is noticeable that the rise of infected agents is less steep than in the same graph from the baseline simulations. This is likely due to the fact that infection rates are limited by the introduction of trust.

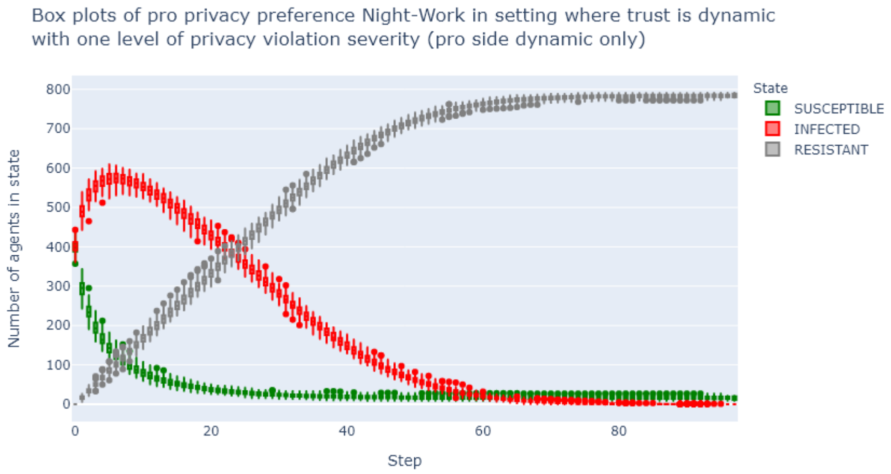

Figure 47: Dynamic of the pro side $< night, work >$ privacy preference with pro infection rate 0.75, pro recovery rate 0.25. Taken from simulations in which the opposing privacy preferences states are assumed to only change given infections on the pro side.

It was expected that the epidemics in this setting would end sooner than those in the baseline simulations, Claim 5a. To test this claim, the steps, at which epidemics end, are compared using a paired Wilcoxon signed rank test. The results of these tests can be seen in Table 22. The table shows that, in 5 out of 9 parameter settings, the epidemics end sooner in when trust is included and is dynamic compared to the baseline simulations in which trust is not factor.

Another hypothesis for the simulations was that the epidemics would peak lower, when dynamic trust is introduced than in the baseline simulations, Claim 5c. Using a paired Wilcoxon signed rank test, the relevant from the two simulation settings was compared to see if the claim stated above holds. The results of these tests can be seen in Table 23. Only the parameter setting in which the infection rate and recovery rate are both 0.5 is missing and thus shows no proof for the hypothesis.

Finally, the last hypothesis stated that agents would enjoy less influence on

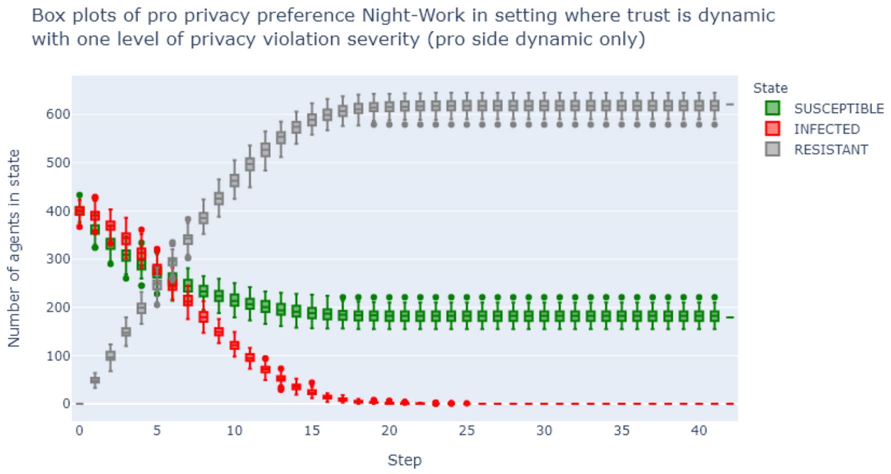

Figure 48: Dynamic of the pro side $< night, work >$ privacy preference with pro infection rate 0.25, pro recovery rate 0.75. Taken from simulations in which the opposing privacy preferences states are assumed to only change given infections on the pro side.

Pro Infection Rate	Pro Recovery Rate	p-value	Statistic
0.25	0.25	0.0064	356502.50
0.25	0.50	0.0023	326049.50
0.25	0.75	0.0000	317403.00
0.50	0.25	0.0011	354350.00
0.50	0.75	0.0015	270638.50

Table 22: Significant ($p < 0.05$) results of paired Wilcoxon signed rank tests that test whether an epidemic, on the pro side, ends sooner when agents keep track of privacy violations to derive dynamic trust values towards instead of no inclusion of trust at all in the baseline. All parameter settings that are not included returned insignificant results.

Pro Infection Rate	Pro Recovery Rate	p-value	Statistic
0.25	0.25	0.0000	659925.5
0.25	0.50	0.0000	413352.0
0.50	0.25	0.0000	664161.0
0.50	0.50	0.0000	563700.0
0.50	0.75	0.0000	410693.5
0.75	0.25	0.0000	631538.0
0.75	0.50	0.0000	544983.0
0.75	0.75	0.0000	458303.5

Table 23: Significant ($p < 0.05$) results of paired Wilcoxon signed rank tests that test whether an epidemic, on the pro side, peaks lower when agents keep track of privacy violations to derive dynamic trust values towards instead of no inclusion of trust at all in the baseline. All parameter settings that are not included returned insignificant results.

the spread privacy preferences in this setting, Claim 5b. To test this, influence ratings from this setting and the baseline simulations are compared using a paired Wilcoxon signed rank test to see if the influence ratings are higher in the baseline simulations. The results of these test can be seen in Table 24. Only the parameter setting, in which the infection rate and recovery rate are 0.75 and 0.25 respectively, is missing and thus shows no proof for the hypothesis. Thus, in eight parameter of setting, the influence ratings of agents in the spread of privacy preferences is lower when agents hold dynamic trust values compared to the baseline simulations in which trust is not a factor.

Pro Infection Rate	Pro Recovery Rate	p-value	Statistic
0.25	0.25	0.0000	47366650025.00
0.25	0.50	0.0000	37140037497.00
0.25	0.75	0.0000	28874302970.00
0.50	0.25	0.0000	47861636185.00
0.50	0.50	0.0000	42931336440.50
0.50	0.75	0.0000	38143577174.00
0.75	0.50	0.0000	42097857349.00
0.75	0.75	0.0000	38799968462.00

Table 24: Significant ($p < 0.05$) results of paired Wilcoxon signed rank tests that test whether agents' influence on the spread of privacy preference is diminished when agents keep track of privacy violations to derive dynamic trust values towards instead of no inclusion of trust at all in the baseline. All parameter settings that are not included returned insignificant results.

6.2.6 Dynamic trust with two levels of privacy violations

In this simulation setting, the agents keep track of privacy violations during simulations and use these records to calculate trust values they have towards other agents they are connected to. The two different types of privacy violations, described previously, hold different weights in this setting.

Figures 49 and 50, show the aggregated state dynamics of the simulations from this setting in the most favourable and least favourable setting respectively. In the most favourable setting, it is noticeable that the rise of infected agents is less steep than in the same graph from the baseline simulations. This is likely due to the fact that infection rates are limited by the introduction of trust.

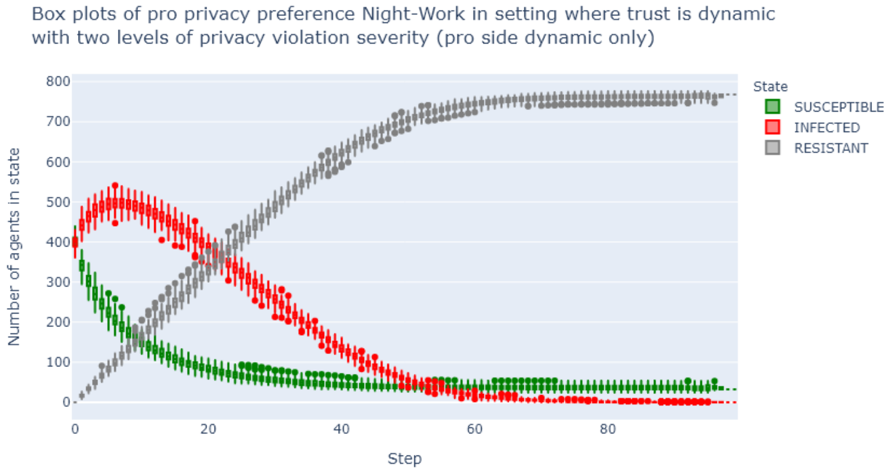

Figure 49: Dynamic of the pro side $< night, work >$ privacy preference with pro infection rate 0.75, pro recovery rate 0.25. Taken from simulations in which the opposing privacy preferences states are assumed to only change given infections on the pro side.

Once again, Claim 6a, that epidemics ends sooner in this setting compared to the baseline, is tested using a paired Wilcoxon signed rank test. The results of these tests can be seen in Table 25. Epidemics end sooner than the baseline epidemics in all parameters settings.

The peaks of the epidemics are then compared to see if the peaks in this setting are lower than the peaks in the baseline simulations, Claim 6c. These tests are performed using the paired Wilcoxon signed rank test. The results of these test can be seen in Table 26. Epidemics peak lower than the baseline epidemics in all parameters settings.

Finally, it was expected that the influence of agents would be diminished in this setting compared to the baseline, Claim 6b. Using all influence ratings from the two simulation settings, this claim is tested using a paired Wilcoxon signed rank test. The results of these tests can be seen in Table 27. Agents have lower influence ratings compared to the baseline simulations in all parameters settings.

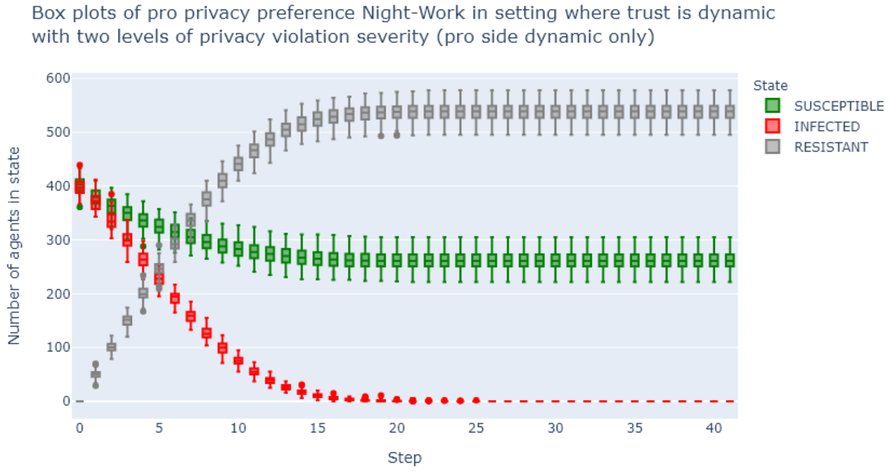

Figure 50: Dynamic of the pro side $< night, work >$ privacy preference with pro infection rate 0.25, pro recovery rate 0.75. Taken from simulations in which the opposing privacy preferences states are assumed to only change given infections on the pro side.

Pro Infection Rate	Pro Recovery Rate	p-value	Statistic
0.25	0.25	0.0000	47261132239.00
0.25	0.50	0.0000	36256319076.50
0.25	0.75	0.0000	27621709459.50
0.50	0.25	0.0000	48597725915.50
0.50	0.50	0.0000	43060275850.50
0.50	0.75	0.0000	37820937324.00
0.75	0.25	0.0000	46383402040.00
0.75	0.50	0.0000	43034315683.50
0.75	0.75	0.0000	39455031935.50

Table 25: Significant ($p < 0.05$) results of paired Wilcoxon signed rank tests that test whether an epidemic, on the pro side, ends sooner when agents keep track of two levels of privacy violations to derive dynamic trust values towards instead of no inclusion of trust at all in the baseline. All parameter settings that are not included returned insignificant results.

Pro Infection Rate	Pro Recovery Rate	p-value	Statistic
0.25	0.25	0.0000	720599.0
0.25	0.50	0.0000	572043.5
0.25	0.75	0.0440	362236.5
0.50	0.25	0.0000	720600.0
0.50	0.50	0.0000	719389.5
0.50	0.75	0.0000	695915.0
0.75	0.25	0.0000	720600.0
0.75	0.50	0.0000	720600.0
0.75	0.75	0.0000	719323.0

Table 26: Significant ($p < 0.05$) results of paired Wilcoxon signed rank tests that test whether an epidemic, on the pro side, peaks lower when agents keep track of two levels of privacy violations to derive dynamic trust values towards instead of no inclusion of trust at all in the baseline. All parameter settings that are not included returned insignificant results.

6.3 Summary

The baseline simulations exhibit the expected behaviour consistent with previous work with epidemic models in the information diffusion domain. An epidemic lasts longer when the infection is higher than the recovery[19]. The connectivity of an agent correlates positively with their ability to spread a privacy influence. These results provided the initial foundation for the DIPP model. These results are also intuitive, as we would expect well-connected agents to spread information diffusion easily due to their reach and this property also should hold in privacy preference diffusion. Furthermore, it is intuitive that when people are more likely to be infected by an infectious entity than they are to recover from said entity. Thus, the epidemic caused by the infectious entity will last longer.

The results also show that when the best-connected agents are tasked with spreading a rare privacy preference, the influence rating measure captures their increased influence in the spread of this privacy preference due to its rarity. This does not hold in every parameter setting when the opposing privacy preferences are fully dynamic. This is in line with the common methods of viral marketing on OSNs in which influential users are chosen to market a product to their current audience and a possible new audience. These influencers can be perceived as pioneers of the use of the product they are advertising. Even if other influential users catch on to the trend, the fact that the pioneers started the trend attributes to them more influence.

Overall, it is found that the introduction of trust limits the spread of privacy preferences albeit not in all parameters settings. Furthermore, the experiments, in which opposing privacy preferences are static, show more consistent ability for agents to protect themselves against privacy preference epidemics than the experiments in which the dynamics of the pro and anti side are modelled the same way. This result implies that the DIPP shows an ability to model the fact that trust is at the heart of privacy decisions, as seen in Lampinen *et al.* [16]. The influence agents have, on the spread of a privacy preference, is also limited when trust is introduced. This claim seems to hold in almost all parameter settings, except when dealing with a high infection rate, relative to the other rates. This foundation is useful to be build upon, as the way trust is modelled can obviously be varied in future work.

Finally, we have seen that privacy preferences can become prevalent in social networks under epidemic models due to their infection and recovery rates. The best-connected agents are the most influential in the spread of privacy preferences. However, agents can employ a trust model to limit the impact of privacy preference epidemics.

Pro Infection Rate	**Pro Recovery Rate**	**p-value**	**Statistic**
0.25	0.25	0.0000	47261132239.00
0.25	0.50	0.0000	36256319076.50
0.25	0.75	0.0000	27621709459.50
0.50	0.25	0.0000	48597725915.50
0.50	0.50	0.0000	43060275850.50
0.50	0.75	0.0000	37820937324.00
0.75	0.25	0.0000	46383402040.00
0.75	0.50	0.0000	43034315683.50
0.75	0.75	0.0000	39455031935.50

Table 27: Significant ($p < 0.05$) results of paired Wilcoxon signed rank tests that test whether agents' influence on the spread of privacy preference is diminished when agents keep track of two levels of privacy violations to derive dynamic trust values towards instead of no inclusion of trust at all in the baseline. All parameter settings that are not included returned insignificant results.

7 Discussion

The results provide a lot of material to unpack. In this section, the results are discussed along with their implications.

7.1 Main properties of the state dynamics

The data from all experiments show that the foundation of the DIPP model for privacy preference diffusion is stable. The infection and recovery rates govern the state dynamics as expected. Epidemics last shorter when the recovery rate is higher than the infection rate. There are also noticeable influences from the anti side dynamic, whether the dynamic is static or not. This is a solid foundation for the purpose at hand. The most influential agents in the diffusion process are, as expected, the agents with the highest degree on the social network. This is in line with previous research results from information diffusion models. In most parameter settings over all experiments, trust modelling gives agents the ability to limit the spread of privacy preferences.

7.1.1 Rooted in information diffusion

As the proposed method is heavily influenced by theory from information diffusion. It was expected that this would be reflected in the results. We saw that when the infection rate is higher than recovery rate a privacy preference epidemic lasts longer. This is intuitive and provides a stable foundation. The same goes for the property that there is a positive correlation between an agent's degree and their influence rating in the spread of privacy preferences.

7.1.2 Privacy preferences that sell themselves

In the setting with rare privacy preferences and high degree influecers, we saw that the influence rating of the influecers does not always increase when the privacy preference they spread is rare. This is notably the case in the parameter settings that have the highest possible infection rate. This seems to suggest that in these setting the privacy preference is so infectious that is spread as easily as in the baseline simulations.

7.1.3 What to trust?

The results of the experiments are mostly as expected when it comes to limiting the impact of epidemics using trust models. This is apart from the experimental settings in which in the introduction of trust modelling did not allow the agents to limit the impact of a privacy preference epidemic. It is important to explore where this difference of results comes from. On the other hand, all results from the experiments, in which the anti side privacy preferences are mostly static, are very conclusive.

The obvious point to make is that in the experiments, in which the anti side privacy preferences are fully dynamic, the anti side influences the pro side by design. If anti side privacy preference has a high recovery rate, this is more beneficial to its pro side counterpart. In this case, the pro side privacy preference endures less resistance as fewer people stay infected with the opposing privacy preference over time. The opposite holds for those cases in which the infection rate of anti side privacy preference is high. The number of infected agents can even rise in this case, leading to more resistance for the pro side privacy preference. This dynamic does not take place in the experiments without dynamic anti side privacy preferences. The pro side epidemics can proceed with only limited resistance from the anti side, as agents still hold opposing norms.

7.1.4 Influencers beware!

In all experiments, it was found that the influence of agents on the spread of privacy preferences is diminished when agents employ a trust model in all but the most favourable setting for the privacy preference epidemics. This exception was also only found in the experiments in which the opposing privacy preference dynamics were fully dynamic. While this was mostly expected, it is still a surprising result. It was expected that influence would be reduced mostly due to the accompanied lowering of the epidemic peaks and earlier epidemic endings. Because the higher the number of agents infected, directly and indirectly, the higher an agent's influence rating. But as we have seen in the previous section, lower the epidemic peaks and earlier epidemic endings were only present in a limited number of parameter settings in the experiments in which the opposing privacy preference dynamics were fully dynamic. But even in these settings, the influence ratings of agents were reduced.

7.1.5 Opposing privacy preferences: static or dynamic

The conducted experiments were never meant to help in choosing between opposing privacy preferences that mostly static and dynamic ones, in modelling privacy preference diffusion. But the differences we have seen in the results from experiments from the two conditions only further raise the question as to which one of the two is the most accurate. The truth mostly lies between the two assumed approaches.

7.2 Validity of methodology

The methodology, put forth in this document, is meant to be a starting point for a new field of research. This section explores the validity of the method. Validation on real life data is regrettably impossible as a data set with spreading privacy preferences has not been found, The points in this section are focus around the topic of realism, because the perfect simulation of privacy preference diffusion would be as close to a theorized reality as possible. As a starting point, the DIPP model, inevitably, lacks some aspects of realism. While this is not ideal, the framework is also modular enough to allow for improvements without the need to start afresh.

7.2.1 Trust only as a shield?

One of the main goals of this project is to investigate whether trust modelling can ensure that agents can limit the impact of privacy preference epidemics. Since trust is based on positive and negative (i.e. privacy violations) experiences, the higher the number of negative experiences with a certain, the less said agent will be trusted. The results show that the introduction of trust modelling does limit the impact of privacy preference epidemics. However, trust should also enhance the relationship between two agents when their interactions have been positive. This side of the trust coin has not been explored. Does trust modelling provide the agents with the ability to strengthen relationships as well diminishing them? Answering this question would, in turn, provide more explicit evidence for the idea that privacy preference epidemics are less impactful due to the negative experiences agents have with opposed privacy preferences.

7.2.2 Protection against influencers

The results show that the influence an agent has in the spread of privacy preferences is reduced when agents employ a trust model for privacy violations. While this is a good result, it may not imply that agents can protect themselves from highly influential figures (i.e. agents with a high degree in the OSN). It may be the case, that since everybody employs a trust model in the simulations, the overall influence is diminished since all agents are less likely to be infected. It would be good to identify agents that are vulnerable to the most influential agents and quantify the difference in influence exerted on these vulnerable agents.

7.2.3 Actions per step

In the current model, an agent share content and tries to recover from an infection at each step in the simulation. But this may not be realistic. A study of Facebook user sharing habits[5] finds that 16% of Facebook users use Facebook at most a few times a week, while 58% of users use the OSN multiple times a day. Women are found 5% more likely to share content at least once a day. Of the 2000 surveyed Facebook users, 9% shares no content on Facebook regularly. These statistics shows that the implicitly assumed uniformity in sharing behaviour of all agents is not realistic. One could imagine a sharing behaviour parameter that, for example, dictates the probability at each step that an agent shares content. If the agent does not share content, they have the ability to recover from privacy preferences they are infected with.

7.2.4 Definition of time

The steps in the simulation are abstract as put forth here. It is unclear whether a step is an hour, a day or a month. The definition of this resolution could provide the method with more external validity. The steps, at this moment, do not translate to any definition of time in real life.

7.2.5 Initial spread of privacy preferences

In real life social networks, the initial spread of privacy preferences could be argued to be focused around a small group of people that had adopted/been infected with a privacy preference. In the experimental setup, however, the initial spread of privacy preference infections would be around half of the social network. It is unlikely for half of all user on an OSN to start sharing video clips of the birth of their children without such content being perceived on the network before. The initial spread in the current experimental setup is only realistic if the scenario is that data recording start in the midst of an ongoing epidemic.

7.2.6 Customization

The previous points on initial spread of infections, the nature of the anti side dynamics and sharing behaviour of users on the social network, are important points to address in future work. The power of the method lies in the ease with which new ideas can be adopted in the DIPP model. For example, given statistics on how to initially spread privacy preference on the social network, a researcher can easier specify these using a Python Dictionary with the privacy preferences as keys and the proportion of the population to infect with the privacy preference. Furthermore, the model allows for the manual changing of states of each agent for each privacy preference.

8 Future Work

There are various research avenues that can follow from the DIPP model. This section delves into the possible research angles to explore in order to improve

[5] https://www.frac.tl/work/marketing-research/facebook-user-sharing-habits-study/

the DIPP model.

8.1 Opposing privacy preference dynamics

As previously mention, the DIPP model allows for assumptions of the opposing privacy preference state dynamics, fully dynamic and static. The truth is somewhere in the middle. Research into how opposing privacy preference come about can make the DIPP model a more accurate model. For example, this research might be a qualitative study into what offline discussions on privacy on OSNs contain.

8.2 Different epidemic models

As we have seen in the related work section, epidemic models have shown flexibility to model infection in different contexts. The DIPP model is based on the SIR epidemic model. However, other epidemic models can be considered. One could imagine a situation in which recovered users become susceptible once again due to some phenomenon in the OSN, the SIRS model [13]. For example, if Alice decides to stop sharing her political opinion anymore and content of political opinions only become more popular, will she stand by her decision or come back on it? There is clearly a chance that Alice might come back on her decision. This could be taken into account in further iterations of the DIPP model.

Peer pressure exists in social networks, as modelled Hui and Buchegger in [12]. Using epidemic models, it can be modelled in privacy preference diffusion research as well. As John and Joshua's work in [6], shows the concept of infectious recovery. They set the recovery rate to be proportional to the number of agents that have already recovered divided by the total number of agents. This means that the higher the number of agents that have recovered, the more likely other agents are to recover as well. Peer pressure could be introduced by, for example, setting the infection rate to the number of friends that have already been infected with a privacy preference divided by the total number of friends. Thus, the higher the number of friends that have been infected, the more likely an agent becomes to be infected.

8.3 Allowing more interactions

The DIPP allows two types of interactions at the moment; observing shared content and sharing content. However, social networks allow for many more types of interactions. Users can often like or dislike content. Users can block users they would not like to content from. They can also directly re-share the content shared by others. These are all interactions that can affect the spread of privacy preferences.

While the DIPP model is tested with flat infection rates, these can be made dynamic. The first step has already been done, with the inclusion of trust values to make infection rates dynamic. Infection rates define how likely users are to get infected. It can be argued that content that is liked more, is also more infectious. Research by Sherman *et al.* [25] finds that like interactions (receiving and giving likes) leads to brain activity in regions of the brain associated with rewards among other regions.

The ability to block users ensures that a user can not see content from a user they have blocked. The blocked user can also not perceive content from the user that has blocked them. Since we hypothesize that privacy preferences spread through shared content, providing the ability to block users that share content that is opposed seems an important next step in the evolution of the DIPP model.

8.4 Not all friends are equal

In OSNs, users maybe connected to users as friends. But not every friend holds the same type of relationship with the user offline. And it is intuitive that a close friend's influence on a user is different to the influence of a distant relative they are friends with on an OSN. This difference in relationship strength can provide a different view on how privacy preferences spread than what we now hold.

To explore the strength of relationships, the *Copenhagen Networks Study interaction data* data set can be used. As previously stated, the data set contains data depicting the offline calls and text messages sent between students. These data could provide researchers with a measure of relationship strength. Furthermore, the similarity of users can also be considered as a factor in the strength of their relationship.

9 Conclusion

This project aims to propose a new framework to model the spread of privacy preferences using an epidemic model. The DIPP model uses the SIR model to model diffusion of privacy preferences. Simulations are run using the DIPP model and data from an actual OSN. Under this model, we observe the expected properties that are consistent with the models from information diffusion DIPP's creation is inspired by. It is also found that how well an agent's degree in the OSN, determines, at least partly, their influence in the spread of a privacy preference.

Agent can also employ a trust model to limit the impact of privacy preference epidemics. The introduction of trust modelling ensures that in all cases the influence rating of agents is diminished, in most cases the epidemics end sooner and in most cases the peak number of infections is lower, compared to simulations without trust models.

These results can be strengthened with the exploration of relationship strength based on trust as the goal of trust modelling is not only to destroy relationships between agents, but, also to strengthen them.

Future work can improve the realism of the DIPP model by considering more interaction, more dimensions to relationship strength and different epidemic model variations. Finally, there is a need for a more accurate model for opposing privacy preferences and their dynamics.

References

[1] İş, H., and Tuncer, T. Interaction-Based Behavioral Analysis in Twitter Social Network. `https://doi.org/10.5281/zenodo.3490938`, Oct. 2019.

[2] Agarwal, S., and Sureka, A. Semantically Analyzed Metadata of Tumblr Posts and Bloggers. `https://data.mendeley.com/datasets/hd3b6v659v/2`, Apr. 2016.

[3] Barth, S., and De Jong, M. D. The privacy paradox–investigating discrepancies between expressed privacy concerns and actual online behavior–a systematic literature review. *Telematics and informatics 34*, 7 (2017), 1038–1058.

[4] Boy, J. D. Anonymized Instagram network data from Amsterdam and Copenhagen, Pajek format. `https://doi.org/10.5281/zenodo.45272`, Jan. 2016.

[5] Bucur, D., Iacca, G., Marcelli, A., Squillero, G., and Tonda, A. Multi-objective evolutionary algorithms for influence maximization in social networks. In *European conference on the applications of evolutionary computation* (2017), Springer, pp. 221–233.

[6] Cannarella, J., and Spechler, J. A. Epidemiological modeling of online social network dynamics. *arXiv preprint arXiv:1401.4208* (2014).

[7] Dupree, J. L., Devries, R., Berry, D. M., and Lank, E. Privacy personas: Clustering users via attitudes and behaviors toward security practices. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems* (2016), pp. 5228–5239.

[8] Feng, L., Hu, Y., Li, B., Stanley, H. E., Havlin, S., and Braunstein, L. A. Competing for attention in social media under information overload conditions. *PLOS ONE 10*, 7 (07 2015), 1–13.

[9] Fu, L., Jacobs, M. A., Brookover, J., Valente, T. W., Cobb, N. K., and Graham, A. L. An exploration of the facebook social networks of smokers and non-smokers. *PLOS ONE 12*, 11 (11 2017), 1–15.

[10] Griffiths, N. Task delegation using experience-based multi-dimensional trust. In *Proceedings of the fourth international joint conference on Autonomous agents and multiagent systems* (2005), pp. 489–496.

[11] Guille, A., Hacid, H., Favre, C., and Zighed, D. A. Information diffusion in online social networks: A survey. *ACM Sigmod Record 42*, 2 (2013), 17–28.

[12] Hui, P., and Buchegger, S. Groupthink and peer pressure: Social influence in online social network groups. In *2009 International Conference on Advances in Social Network Analysis and Mining* (2009), IEEE, pp. 53–59.

[13] Jin, Y., Wang, W., and Xiao, S. An sirs model with a nonlinear incidence rate. *Chaos, Solitons & Fractals 34*, 5 (2007), 1482–1497.

[14] Kempe, D., Kleinberg, J., and Tardos, É. Maximizing the spread of influence through a social network. In *Proceedings of the ninth ACM SIGKDD international conference on Knowledge discovery and data mining* (2003), pp. 137–146.

[15] Kökciyan, N., and Yolum, P. Priguardtool: A web-based tool to detect privacy violations semantically. In *International Workshop on Engineering Multi-Agent Systems* (2016), Springer, pp. 81–98.

[16] Lampinen, A., Lehtinen, V., Lehmuskallio, A., and Tamminen, S. We're in it together: interpersonal management of disclosure in social network services. In *Proceedings of the SIGCHI conference on human factors in computing systems* (2011), pp. 3217–3226.

[17] Leskovec, J., and Krevl, A. SNAP Datasets: Stanford large network dataset collection. `http://snap.stanford.edu/data`, June 2014.

[18] Li, D., Ma, J., Tian, Z., and Zhu, H. An evolutionary game for the diffusion of rumor in complex networks. *Physica A: Statistical Mechanics and its Applications 433* (2015), 51–58.

[19] Li, M., Wang, X., Gao, K., and Zhang, S. A survey on information diffusion in online social networks: Models and methods. *Information 8*, 4 (2017), 118.

[20] Liu, D., Wang, Y., Jia, Y., Li, J., and Yu, Z. From strangers to neighbors: Link prediction in microblogs using social distance game. *Diffusion Networks and Cascade Analytics, WSDM* (2014).

[21] Lu, D., Yang, S., Zhang, J., Wang, H., and Li, D. Resilience of epidemics for sis model on networks. *Chaos: an interdisciplinary journal of nonlinear science 27*, 8 (2017), 083105.

[22] Reece, S., Rogers, A., Roberts, S., and Jennings, N. R. Rumours and reputation: Evaluating multi-dimensional trust within a decentralised reputation system. In *Proceedings of the 6th international joint conference on Autonomous agents and multiagent systems* (2007), pp. 1–8.

[23] Sapiezynski, P., Stopczynski, A., Lassen, D. D., and Lehmann, S. Interaction data from the copenhagen networks study. *Scientific Data 6*, 1 (2019), 1–10.

[24] Shelke, S., and Attar, V. Source detection of rumor in social network–a review. *Online Social Networks and Media 9* (2019), 30–42.

[25] Sherman, L. E., Hernandez, L. M., Greenfield, P. M., and Dapretto, M. What the brain 'likes': neural correlates of providing feedback on social media. *Social cognitive and affective neuroscience 13*, 7 (2018), 699–707.

[26] Sun, Q., and Yao, Z. Evolutionary game analysis of competitive information dissemination on social networks: An agent-based computational approach. *Mathematical Problems in Engineering 2015* (06 2015).

[27] Tang, Y., Xiao, X., and Shi, Y. Influence maximization: Near-optimal time complexity meets practical efficiency. In *Proceedings of the 2014 ACM SIGMOD international conference on Management of data* (2014), pp. 75–86.

[28] Ulusoy, O., and Yolum, P. Pano: Privacy auctioning for online social networks. In *Proceedings of the 17th International Conference on Autonomous Agents and MultiAgent Systems* (Richland, SC, 2018), AAMAS '18, International Foundation for Autonomous Agents and Multiagent Systems, p. 2103–2105.

[29] Ulusoy, O., and Yolum, P. Emergent privacy norms for collaborative systems. In *International Conference on Principles and Practice of Multi-Agent Systems* (2019), Springer, pp. 514–522.

[30] Vinyals, O., Toshev, A., Bengio, S., and Erhan, D. Show and tell: Lessons learned from the 2015 mscoco image captioning challenge. *IEEE transactions on pattern analysis and machine intelligence 39*, 4 (2016), 652–663.

[31] Wang, X. F., and Chen, G. Complex networks: small-world, scale-free and beyond. *IEEE circuits and systems magazine 3*, 1 (2003), 6–20.

[32] Woo, J., and Chen, H. Epidemic model for information diffusion in web forums: experiments in marketing exchange and political dialog. *Springer-Plus 5*, 1 (2016), 66.

[33] Xu, R., Li, H., and Xing, C. Research on information dissemination model for social networking services. *International Journal of Computer Science and Application (IJCSA) 2*, 1 (2013), 1–6.

[34] Young, H. P. The evolution of social norms. *Annu. Rev. Econ 7*, 1 (2015), 359–387.