Université de Montréal

Simulation of fake news propagation

par

Rigo, Arpad Botond

Département de l'informatique et de recherche opérationnelle
Faculté des arts et des sciences

Mémoire présenté en vue de l'obtention du grade de
Maître ès sciences (M.Sc.)
en Discipline

October 21, 2023

Université de Montréal

Faculté des arts et des sciences

Ce mémoire intitulé

Simulation of fake news propagation

présenté par

Rigo, Arpad Botond

a été évalué par un jury composé des personnes suivantes :

Nom du président du jury
(président-rapporteur)

Nom du directeur de recherche
(directeur de recherche)

Nom du membre de jury
(membre du jury)

Résumé

...sommaire et mots clés en français...

Abstract

...summary and keywords in english...

Contents

List of Tables

List of Figures

Abbreviations

GCN	a subclass of artificial intelligence in the neural network category *Graph Convolutional Network*
URL	link to website *Uniform Resource Locator*
DAU	Average of daily active users *Daily Active Users*
mDAU	Average of daily active users who have seen ads on Twitter *monetizable Daily Active Users*
API	an application that can be called or used in an other app *Application Programming Interface*

Special Thanks

Introduction

The propagation of news and especially fake news is a reality on social media. Some are easier to detect than others but there is always the human factor that makes it difficult to define the real impact as for its spreading. Furthermore, to be able to detect them and establish a way to stop the spreading of these false information, it is important to understand the mechanism behind the way those information and these news in particular can spread.

But first of all, let us define what is a fake news. A fake news, in this paper, will represent a piece of false information that users on social media share between themselves. Those may or may not know if the shared information is false, even more, some of them are social media bots with the goal to spread the fake news faster. As for the human users themselves, it's important to consider factor of closeness and trustworthiness to define if a user will believe more or less easily the shared information.

Then, we should take in consideration the fact that social media platforms use different types of algorithm to filter the received information and what will be on the top of the page for the user. To have a certain variety, the focus will be on 3 social media : Facebook, Twitter and Reddit.

With the aid of Graph Neural Networks, the main goal behind this simulation is to provide a template of how fake news spread and an assistance to fake news detection algorithms by considering the closeness, knowledge and trustworthiness of users on social media platforms on a larger scale such as the propagation across multiple media platforms.

Chapter 1

Literature Review

There has been a lot of research done about social media and especially fake news spreading in the last few years in particular due to COVID-19 which caused a significant increase of average time spend on media platforms. For instance, according to Statista [1], the average time users from the United States spent on social media increased from an average of 54 minutes per day to 65 minutes which represent an increase of 20% compared to the average of the year before. With this increased time spent on social media, it was necessary to consider its impacts on social media such as information overload and the echo chamber effect.

To simulate the spreading of fake news on social media, the epidemiological model is a common way to represent it mainly for its disease-like propagation. Indeed, the spreading of a fake news is based on the *infected* agents and how those will react to the *infection*. The way the agent as well as the Network structure are defined play key roles as for the spreading itself. The authors in [2] use four type of agents, the *Susceptible*, the *Incubator*, the *Spreaders* and the *Recovered*. In this model, once the agent received the information, it enters in the *Susceptible* category. Depending on the momentary tendency (*Incubator* state) of the agent, its behaviour will be either to believe or not the information and to propagate or not either the false or truth about it, which is the *Spreaders* state. The authors in [2] explain that each agent can come back in the *Susceptible* category depending on its state and environment at that particular moment. Also, there is also a state called *Stifler* which represent the agents that are not inclined to share information whether true or false. Agents in this category is may become *Susceptible* again but also choose to never spread again as shown in the 2.1.

However, its important to consider the structure of the network to define how the propagation will happen. Some of media platforms, such as Reddit, will focus on the reviews and likes of the posts as well as the time when the post was made to establish the best feeds

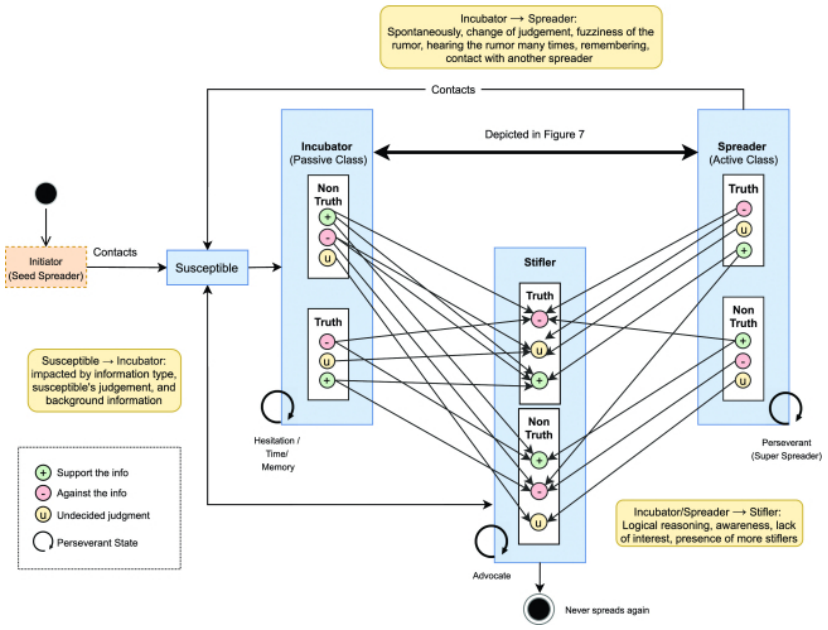

Figure 1.1. State changes in the epidemic model from [2]

to users **Reddit**. On the other hand, big social medias, especially Facebook, do a filtering of their inventory to decide what is appropriate for the user, then calculate the score of the filtered elements in a way to provide information/news that triggers the interest of the target users as well as to have enough variety to keep the attention on a long term while increasing the chance of interaction between users[3]. By calculating the interest of the target users, this algorithm focus on the long term impact and satisfaction of the audience with a stable tendency toward different topics. This includes also the degree of closeness between agents as well as the the groups triggering the interest of a specific agent for a better comprehension of the persona and its need vs side topics.

Chapter 2

Solution

To simulate the propagation on several media platforms, its necessary to understand the algorithm behind each of them and create an environment that replicates it. To do so, each agent will be tagged to define which social media it's associated with. In real time, multiple agents (accounts) can have the same owner so the behaviour be almost the same. The main difference will be the type of connections and feed algorithm these agents will face, changing slightly the behaviour of the person depending of media platform. As for the connection between them, a GCN (Graph Convolutional Network) will be used represent the relation and sharing across the platforms[4].

2.1. Media Platforms

2.1.1. Social Media demography

To simulate fake news spreading across different social media, 3 media platforms is replicated on a smaller scale: Facebook, Twitter, Reddit. The choice behind these social medias is due to the difference between their algorithm and the demographic difference between users. For instance, Facebook uses a scoring algorithm [3] to please its audience. In fact, Facebook's demography shows that 51.4% of users are between 18 and 34 from a total of 3 billion people, as well as 11.9% are between 45 and 54 years old [5]. However, it's also shown that younger people are slowly leaving this media platform to change for a different platform such as TikTok whom's audience is composed at 38.9% of people between 18 and 24 years old.

For Twitter, the algorithm provides a mixed feed from users that the agent is connected to and news/tweets with the most appeal [6]. A new way to represent interaction between agents is provided by [6] which lets the users choose between the subjects that appeal them the most. For this, it is important to consider that over 59% of users were find between the age range of 25 and 49 in 2021 as for the younger and older generations, 17.1% of them

	Facebook	Reddit	Twitter
Age demography	29.9% between 25-34	36% between 18-29	28.4% between 35-44
Feed algorithm	Scoring process by refining the inventory for each user [8].	Multiple recommendation algorithms are possible such as "home", "popular", "latest" and "all" where each algorithm uses different level of filtering	Mix between in and out of network in a 50%-50% way [9].
Post difference	Mostly public posts with tags that let users refine their search.	Posts and discussions are separated into "Subreddits" or groups where users may or may not have to be members.	Mostly open discussions or tweets with tags.
Average daily active users	A grown from 1.284 billion daily users in 2017 to 2.064 billion users daily in 2023. [10]	A grown from an average of 52 million daily user in mid-2021 to 57 million daily users at the end of 2022. [11]	A grown from an average of 109 million mDAU (monetizable daily active users) in 2017 to an average of 237.8 million mDAU in mid-2022. [12]
Level of education	64% of people who took part of the study in 2021(in the U.S.) with a High School education or less use Facebook, while 73% of people with a higher education such as college or more declared to use it.[13]	9% of respondents in 2021 with a High School or less said using Reddit while 26% of people with a College degree or more said to use it (all participants are from the US). [13]	14% of people who answered the 2021 survey while having a High School degree or less said to use Twitter while 33% of participants with a College degree or more(all participants are from the US). [13]

Tableau 2.1. Caption

were in between 18-25 and another 17.1% were over 50. As for Reddit, most of the users are between 18 and 29 year old, namely 36% [7].

2.1.2. Modeling the feed algorithm

To model our feed algorithm, texts and URLs need to be classified. For this, it is possible to use an open-source API such as Twinword giving a certain score for different topics about a given text or web page 2.1. With the scoring of an information, especially fake news, it's possible to define a certain target audience that will have a bigger interest and/or knowledge

```
"topic": {
  "law": 0.22976767934643857,
  "crime": 0.18381414347715086,
  "knowledge": 0.18381414347715086,
  "criminal": 0.18381414347715086,
  "felon": 0.16083737554250702,
  "announcement": 0.16083737554250702,
  "work": 0.16083737554250702,
  "people": 0.16083737554250702,
  "judge": 0.16083737554250702,
  "bad": 0.16083737554250702
},
```


Figure 2.1. E.g. : Result from the classification of the text A.1 from A using Twinword Tagging system.

about the topic.

The first feed algorithm is a representation of Facebook's feed algorithm. The agents' interest range are a probabilistic model so the sum of importance of each topics equals to 1. Because the focus of this simulation is fake news propagation, especially in a text format, some elements, such as market information, music recommendations and videos(excepted scripts), are not considered in the computation. Then, agents' social activity need to be defined in order to point out influencers as well as other important people in the network. A similar modal has already been put in place by [14] to define the importance of connection between agents based on the social activity between them:

$$social\_activity\_set(user_x) = \{P(user_x),S(user_x),C(user_x),L(user_x),I(user_x),M(user_x),PL(user_x),PC(user_x)\} 2.2$$

Based on the model from [14] to measure trust between agents, the formula to calculate the user interest towards a news can be defined as :

$$prob\_interest_{userx} = \frac{\Sigma(social\_activity * topic\_prob)}{\Sigma social\_activity}$$

where the "social_activity" is a set of interests or category and "topic_prob" defines the probability that the given information touches a certain topic. By including the concept of closeness, time laps and feedback count from different agents, a specific score is established for each user of a same social platform.

For Twitter, it's mostly the same algorithm as Facebook with the the difference that the feeds are divided in 2 categories, "In-Network" and "Out-Network"[9]. *In-Network* recommendations are from followed accounts on which a Real Graph is used [9] in order to predict the likely engagement rate. As for the Out-Network [9], different types of scoring systems are used. The ranking scoring system relays on the retweets, liking and replies to rank recommendations which can be considered similar to Facebook's feed algorithm.

List label	Description of the list	Weight assigned to the list
$\boldsymbol{P}(user_x)$	List of friends who are tagged on a same photo with the Facebook $user_x$	w_P
$\boldsymbol{S}(user_x)$	List of friends who write on the Facebook $user_x$'s *Wall*	w_S
$\boldsymbol{C}(user_x)$	List of friends who leave comments on the Facebook $user_x$'s *Wall*	w_C
$\boldsymbol{L}(user_x)$	List of friends who like posts on the Facebook $user_x$'s *Wall*	w_L
$\boldsymbol{I}(user_x)$	List of friends who write to the Facebook $user_x$'s inbox	w_I
$\boldsymbol{M}(user_x)$	List of friends on whose *Walls* the Facebook $user_x$ writes or comments	w_M
$\boldsymbol{PL}(user_x)$	List of friends who like the Facebook $user_x$'s photos	w_{PL}
$\boldsymbol{PC}(user_x)$	List of friends who leave comments on the Facebook $user_x$'s photos	w_{PC}
$\boldsymbol{F}(user_x)$	List of all friends of the Facebook $user_x$	–

Figure 2.2. Legend for the social activity set computation by [14]

Other scoring systems, such as social graphs and embedding space, are also used to study interaction between users and tweets to regroup people with similar or common interests.

As for Reddit, there are 4 main recommendation algorithms offered to the user; "home", "popular", "all" and "latest". The *home* algorithm is based on communities joined by the user as for the *latest* and *popular* are based on popularity (up votes) and the time span of the up votes. The *all* option is a more open recommendation method which uses less filtering and analysis unlike the precedent ones.

2.2. Agent modeling

When it comes to agent modeling, it's important to take in account that there is a lot of human factors that may interfere to an accurate spreading simulation. Some of these factors, such as the timing of information reception and the user's mood, cannot be changed or accurately predetermined. However, other factors, like the impulsivity of the user or the subjects/topics he prefers. The impulsivity is based on the average of reply or reaction delay starting at the reception time which may be different for each connection and topic for each user. Authors [2] developed a model that represent the average engagement toward different connections (Fig. 2.3) on a media platform.

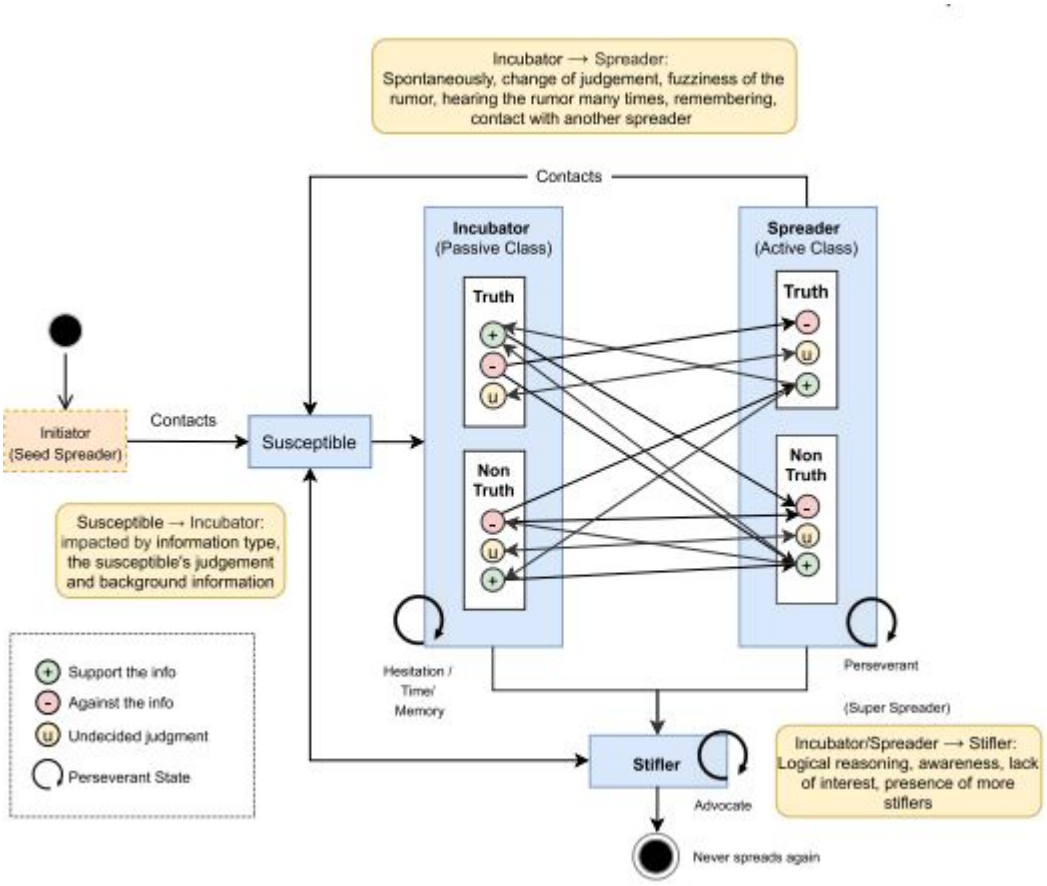

Figure 2.3. Model of propagation and spreading based on user-spreader interaction from [2]

In this model, the spreaders enter in contact with their connection and send information to them, those connections then become either interested in the given information and are in the incubation phase or they are not interested and enter in the Stifler state. The Incubation phase is a passive state where users either believe I propose a similar modal which one takes in account a multi-platform type of propagation. Because of the difference between feed algorithms, some information have a lower propagation on a social media than on another. The Fig. 2.4 shows a model where spreading depends not only on the state of the connected nodes (contacts), but also on the platform choice and message type. In instance, a video may not only be seen by direct contacts but also by people that are indirectly connected such as the contact or subscription of close friends and groups or saved tags.

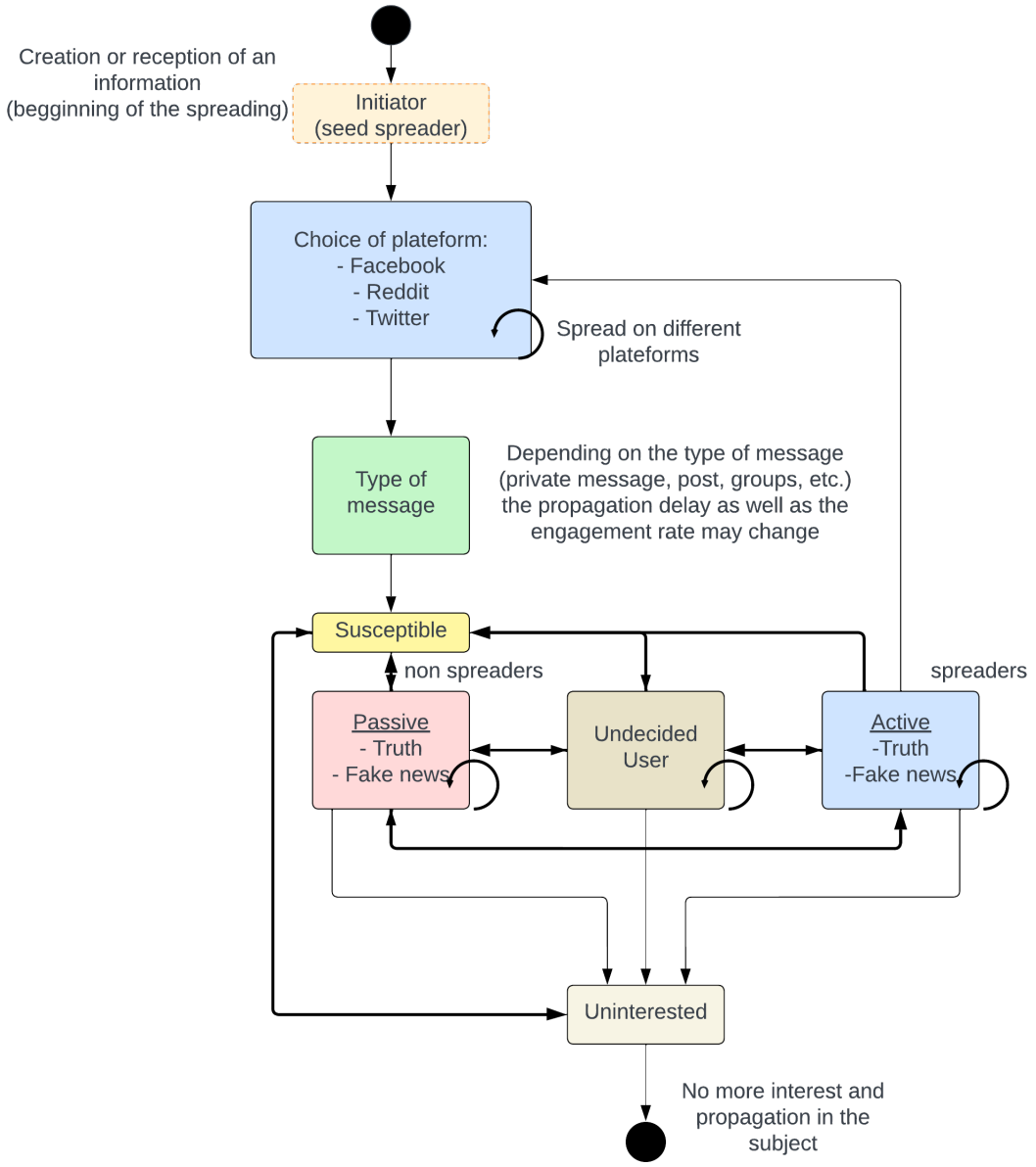

Figure 2.4. Multi-platform propagation model 2.2

Once a user enters the spreader state, he chooses a platform to spread the information based on the accounts linked to the pincoming account(i.e : the infected person was infected by Twitter and can choose to spread the information trough Twitter again or change to a linked Facebook account). This choice leads to multiple way of propagation ultimately to multiple infected platforms due to the fact that some platforms have different sharing system which may change the spreading delay unlike if it was a single social media.

Initiator	The first user to enter in contact with the fake news. For the needs of this paper, the Initiator will automatically spread the information.
Passive	Category of users that either believe or not the given information but are inactive and do not spread it.
Active	Category of users that either believe or not the given information and choose to spread it across the networks.
Undecided User	Category of users that are uncertain whether the information is true or false.
Susceptible	Users that are not yet infected.
Uninterested	Users that are not interested in the actual information(if a new information appear, they will may be back to susceptible category).
Choice of platform	The platform that the spreaders will choose to spread the information. (a spreader can choose multiple platforms at condition that he owns an account on that platform)
Type of message	Tagging the message as a post, private message or other which may change the spreading speed.

Tableau 2.2. Definitions

Chapter 3

Conclusion

To conclude, I propose the 2.4 as a multi-platform propagation model based on 2.3 suggested by [2] in order to simulate a cross-platform propagation by taking in account the time and algorithm complexity of each platform as well as the engagement rate and the content type which can be evaluated by multiple API. The demographic data sets could also be subject to studies due to the fact that they may influence the type of content circulating on social media as well as its engagement rate. Finally, it is important to consider that people may have multiple accounts across and even inside social media platforms which could possibly affect the propagation and spreading complexity.

Références bibliographiques

[1] S. Dixon, "Social media use during covid-19 worldwide - statistics and facts", *Statista*, Oct. 2022. [Online]. Available: https://www.statista.com/topics/7863/social-media-use-during-coronavirus-covid-19-worldwide/#topicOverview, (accessed: 03.06.2023).

[2] S. Raponi, Z. Khalifa, G. Oligeri, and R. Di Pietro, "Fake news propagation: A review of epidemic models, datasets, and insights", *ACM Trans. Web*, vol. 16, no. 3, Sep. 2022, ISSN: 1559-1131. DOI: 10.1145/3522756. [Online]. Available: https://doi.org/10.1145/3522756.

[3] Meta, *How facebook distributes content*. [Online]. Available: https://www.facebook.com/business/help/718033381901819?id=208060977200861, (accessed: 05.06.2023).

[4] M. Karavokyris and S. Sioutas, "Graph neural networks for affective social media: A comprehensive overview", *CIKM'22*, 2022. [Online]. Available: https://dl.acm.org/doi/abs/10.1145/3566097.3568345.

[5] S. Dixon, "Distribution of facebook users worldwide as of january 2023, by age and gender", *Statista*, Feb. 2023. [Online]. Available: https://www.statista.com/statistics/376128/facebook-global-user-age-distribution/ (visited on 06/01/2023).

[6] A. Sharma, J. Jiang, P. Bommannavar, B. Larson, and J. Lin, "Graphjet: Real-time content recommendations at twitter", *Proc. VLDB Endow.*, vol. 9, no. 13, pp. 1281–1292, Sep. 2016, ISSN: 2150-8097. DOI: 10.14778/3007263.3007267. [Online]. Available: https://doi.org/10.14778/3007263.3007267.

[7] O. Samanta, *Reddit statistics | revenue, user count growth 2023*, Mar. 2023. [Online]. Available: https://prioridata.com/data/reddit-statistics/#Reddit_Users.

[8] Meta, *Our approach to facebook feed ranking*, Jun. 2023. [Online]. Available: https://transparency.fb.com/features/ranking-and-content/.

[9] Twitter, *Twitter's recommendation algorithm*, Mar. 2023. [Online]. Available: https://blog.twitter.com/engineering/en_us/topics/open-source/2023/twitter-recommendation-algorithm.

[10] S. J. Dixon, "Number of daily active facebook users worldwide as of 2nd quarter 2023", *Statista*, Jul. 2023. [Online]. Available: https://www.statista.com/statistics/346167/facebook-global-dau/, (accessed: 02.08.2023).

[11] S. J. Dixon, "Average daily active users (dau) of reddit in june 2021 and december 2022", *Statista*, Jun. 2023. [Online]. Available: https://www.statista.com/statistics/1324264/reddit-daily-active-users/, (accessed: 02.08.2023).

[12] S. J. Dixon, "Number of monetizable daily active twitter users (mdau) worldwide from 1st quarter 2017 to 2nd quarter 2022", *Statista*, Aug. 2023. [Online]. Available: https://www.statista.com/statistics/970920/monetizable-daily-active-twitter-users-worldwide/, (accessed: 28.08.2023).

[13] M. Anderson, L. Rainie, and H. Nolan, "Social media use in 2021", *Pew Research Center*, Apr. 2021. [Online]. Available: https://www.pewresearch.org/internet/wp-content/uploads/sites/9/2021/04/PI_2021.04.07_Social-Media-Use_FINAL.pdf, (accessed: 28.08.2023).

[14] V. Podobnik, D. Striga, A. Jandras, and I. Lovrek, "How to calculate trust between social network users?", in *SoftCOM 2012, 20th International Conference on Software, Telecommunications and Computer Networks*, 2012, pp. 1–6. [Online]. Available: https://ieeexplore.ieee.org/document/6347635.

Annexe A

Systeme de score

A.1. Twinword scoring

Twinword text example no. 1

As the publication of such false news of his Majesty, has a tendency to disquiet the minds of his subjects, hurt public credit, and diminish the regard and duty which they owe him, I think the doing it with such views is an offence punishable at Common Law, and for which an Indictment or Information can lye [sic]. And the frequency of such publications is evidence of such wicked designs. But as every false report of this kind which may arise from mistake only cannot be charged as a crime, so it is very difficult to say how often it must be repeated in the paper to make it criminal... I don't know any method to prevent this practice but by prosecuting the offenders when they are guilty.

A.2. Example of multi-platform links

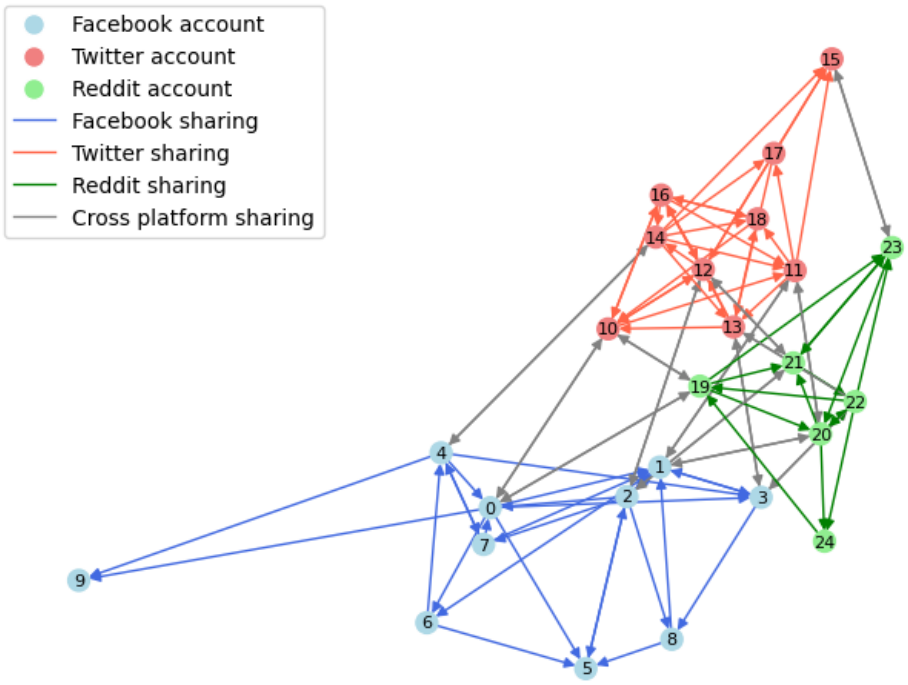

Figure A.1. Multi-platform graph model randomly generated