Angry M&M’s: Aggression Modeling and Minimization in Online Social Networks

Marinos Poiitis

Elaboration of dissertation as part of Web and Data Science M.Sc. Programme

Faculty of Applied Sciences
Computer Science Department

March 17, 2020
Statement Of Authorship

I, Marinos Poiitis, declare that the thesis with title ‘Aggression Modeling and Minimization in Social Networks’ and the work presented in it are a genuine work of mine. I ensure that:

- This work has taken place wholly or mainly during my candidacy for graduate studies at this university.

- Wherever any part of this dissertation has previously been used for the acquisition of a degree or other qualification in this or another university, this is clearly stated.

- Wherever I have consulted the third-party work published, this is correctly attributed.

- Wherever I have cited from third-party work, the source is always given. With the exception of these quotes, this thesis is entirely my personal work.

- I have cited all the utility sources.

- Wherever this thesis is based on collaborative work of my own and of others, I have made it clear which pieces have been made by others and how I contributed.

Signature: Marinos Poiitis
Date: July 15, 2019
The more I learn, the less I realize I know.

Sokrates
Online aggressive behavior has become a common phenomenon in social networks, already counting many victims who suffer both physically and mentally. Aggression impact in online social networks (OSNs) is becoming a crucial problem since aggression diffusion is intense and unpredictable. This dissertation addresses this problem by proposing a systematic approach to model and minimize aggression diffusion in social networks.

The thesis addresses aggression modeling by adjusting conventional influence maximization models such as the Independent Cascade (IC) and the Linear Threshold (LT) which are the most popular ones in the literature. These state-of-the-art models are exploited in this thesis to proceed with a choice of an appropriate model configuration. More specifically, graph weighting, seed selection and aggression transferring processes are applied and under extensive experimentation and testing, the best models’ configurations are detected. Furthermore, the thesis shows that aggression diffusion can be represented by the proposed models, which are extended with consideration of aggression propagation minimizing effects. The cascading models enable the use of competitive cascades as a minimization technique, a method that supposes the existence of two or more competing diffusion processes where each one tries to dominate over the other. Hence, competitive cascades are exploited and compared with the state-of-the-art strategies used by the various social media platforms. It is shown that the proposed methods succeed in reducing the overall aggression score of the network, in contrast to the algorithms currently in use which mitigate the effect of aggression diffusion on a much smaller scale.

More specifically, Chapter 1 outlines the problem addressed and the concept of the thesis, placing emphasis on people vulnerability and exposure states in OSNs. People nowadays tend to use social media as an inextricable part of their everyday life to interact with others, thus online social circles evolve continuously and dynamically. Users diverse origins and scope drive such social circles focus and since several malicious
users are present, negative phenomena have emerged in OSNs. Cyber-aggression among such phenomena, evolves rapidly in OSNs, with cases such as cyberbullying, racism and online expressed discrimination becoming dominant and disturbing. Such phenomena are harmful for users regardless of age, educational or family background, leading the potential victims to marginalization, emotions of deep sadness or even to actions that can hurt themselves. This chapter delineates and offers a top-level description of the cyber-aggression in OSNs, highlights the open critical problems and presents the main contributions of this work.

Chapter 2 covers the theoretical and foundational basis required for understanding the literature’s formulation, with emphasis on the methods which are mostly relevant with the goals of this thesis. The most relevant definitions, entities and parameters are summarized to support the next Chapters readability and comprehension.

Chapter 3 summarizes the literature of the most related research topics and themes. Emphasis is given on earlier work in the areas of information diffusion, influence maximization, optimization for competitive cascades and immunization methods, with focus on aggression modeling and minimization which is the focus of this work. A comparative analysis is provided to highlight earlier approaches bottlenecks and to indicate the most crucial open issues which are addressed in this thesis.

Chapter 4 presents the theoretical framework for the proposed aggression modeling methods. For each method, a pseudo-code algorithmic presentation outlines aggression modeling procedures. The aggression diffusion modeling pipeline is then described, based on the initial nodes selection under various seed selection strategies. Then, different graph weighting schemes are analyzed, to highlight users relationships role in the seed selection and the overall diffusion process and finally the diffusion models are detailed and discussed. Emphasis is placed on indicating IC and LT-based models variations, since they largely affect nodes activation. The IC-LT variations are due to their different view of the factors which impact the propagation processes. IC is based on specific activation criteria which exploit either the aggressiveness of a user or the cumulative effect of the neighborhood in order to activate the user, meaning that they are considered as influenced by the aggressive behavior. Respectively, LT is based on the need to define user threshold specification strategies which will act beneficially towards the overall information dissemination process. These strategies define the nature of the influence that the user is susceptible to, thus they should be fine-grained towards aggression. After the above extensive experiments the parameter configuration that best describes the aggression diffusion process is selected.
Similarly, in Chapter 5, the theoretical framework and the algorithms pseudo-code with respect to aggression minimization are specified. By exploiting the best configurations indicated in Chapter 4, both for IC and LT, emphasis is placed on the two best minimization approaches which reach the highest cosine similarity (based on ground truth validation vector). Initially, based on the fact that aggression modeling through IC and LT is valid - according to the similarity with ground truth - competitive cascades are introduced as a minimization technique. This approach considers the existence of a positive educational cascade that aims to mitigate the effect of the aggressive one. A range of healing mechanisms is proposed to describe the effect of the positive cascade. Specifically, the healing mechanisms considered refer to the total immunization of the user, the healing based on a time-decaying factor and a combination of the previous two methods. Then, the blocking aggression minimization problem is addressed by considering tailored immunization techniques. The proposed immunization techniques are in accordance with the ones used in popular OSNs, and advances with respect to the state-of-the-art are discussed. The consideration of immunization techniques offers an alternative to the competitive cascades as they are subject to aggression modeling through cascade models, an approach that is rather new.

Chapter 6 summarizes the results of the experimental process. The experimentation validates the aggression modeling and minimization efficiency with respect to the proper seed size selection and the appropriate threshold setting in the processes of categorizing a user as aggressive or not. For each phase - aggression modeling and aggression minimization - all the possible configurations of parameters described in the previous chapters are created. First, the aggression modeling configurations are put into test and in continuation, the minimization approaches are compared. As it is shown from the results and the high cosine similarity to the ground truth vector, a cascading model (such as IC and LT) can indeed model aggression diffusion properly, and offer comprehensive insights about the aggression phenomenon. Additionally, the promising results on aggression modeling enable the study of competitive cascades as a minimization technique, as they require the modeling through a cascading process. It is shown that this approach can minimize the overall aggression of the network - measured as a per user score - on a much bigger scale than the methods currently in use. Finally, Chapter 7 has an overall discussion on the topic under examination and proposes some potential future directions to investigate and experiment on.
Αριστοτέλειο Πανεπιστήμιο Θεσσαλονίκης

Περίληψη
Σχολή Θετικών Επιστημών
Τμήμα Πληροφορικής
Μεταπτυχιακό Πρόγραμμα Επιστήμης και Διδασκόντων Ιστού
Ποιητής Μαρίνος - ΑΕΜ: 17

Angry M&M’s: Μοντελοποίηση και Ελαχιστοποίηση της Επιθετικότητας στα διαδικτυακά κοινωνικά δίκτυα

Σήμερα, η επιθετική συμπεριφορά στο διαδίκτυο είναι ένα κοινό φαινόμενο στα κοινωνικά δίκτυα και μετρά πολλά θύματα που υποφέρουν τόσο σωματικά όσο και ψυχικά. Αυτή η διατριβή παρουσιάζει μια συστηματική προσπάθεια να μοντελοποιηθεί και να ελαχιστοποιηθεί η διάδοση της επιθετικότητας στα κοινωνικά δίκτυα. Αρχικά, η εργασία ασχολείται με τη μοντελοποίηση της επιθετικότητας μέσω κλασικών μοντέλων μεγιστοποίησης επιρροής όπως το ανεξάρτητο καταρράκτη (Independent Cascade) και των γραμμικών κατωφλίων (Linear Thresholds). Προκειμένου να προσαρμοστούν αυτά τα μοντέλα στο πρόβλημα που εξετάζεται, η διατριβή προτείνει στρατηγικές που αντλούνται από τον τομέα της επιθετικότητας και είναι προσαρμοσμένες σε αυτόν για την επιλογή του τρόπου απόδοσης βαρών στους γράφους, την επιλογή των κόμβων από τους οποίους θα εκκινήσει η διαδικασία και τη μεταβίβαση της επιθετικότητας, που υποστηρίζονται από εκτεταμένα πειράματα για τον εντοπισμό της καταλληλότερης παραμετροποίησης. Στη συνέχεια, η διατριβή δείχνει ότι η διάδοση της επιθετικότητας μπορεί να μοντελοποιηθεί από τα προτεινόμενα μοντέλα και επιπρόσθετα στοχεύει στην ελαχιστοποίηση της επίδρασης αυτής της διάδοσης. Για το σκοπό αυτό, γίνεται αξιοποίηση μοντέλων από τον τομέα των ανταγωνιστικών μοντέλων καταρράκτη (competitive cascades) και σύγχρονης με τις πιο σύγχρονες στρατηγικές αποκλεισμού κόμβων και ακμών (node/edge blocking) που χρησιμοποιούν τα διάφορα κοινωνικά μέσα. Αποδυκνείται ότι οι προτεινόμενες μέθοδοι επιτυγχάνουν τη μείωση του συνολικού βαθμού επιθετικότητας και εκτιμάται ότι οι συχνότερες εμφανίσεις ή ακόμη και να μετριάσουν απλώς την επίδραση της διάχυσης της επιθετικότητας.

Ειδικότερα, το κεφάλαιο 1 παρουσιάζει τη γενική ιδέα αυτής της εργασίας, συζητώντας τα θέματα της ανάπτυξης και της επικράτησης των κοινωνικών μέσων στο διαδίκτυο καθώς
και το βαθμό στον οποίο αυτά επηρεάζουν τους χρήστες. Οι άνθρωποι σήμερα έχουν την τάση να χρησιμοποιούν τα κοινωνικά μέσα ενημέρωσης ως αναπόσπαστο μέρος της καθημερινής τους ζωής και ρουτίνας για να γνωρίζουν νέους ανθρώπους, να έρχονται σε επαφή με τους φίλους τους αλλά και για να διαχειριζούνται τον κοινωνικό τους κύκλο γενικότερα. Ωστόσο, υπάρχουν διαφορετικά υπόβαθρα χρηστών και έτσι η επίδραση μιας τέτοιας έκθεσης ποικίλλει επίσης.

Ένα από τα παρατηρούμενα αρνητικά φαινόμενα του σύγχρονου κόσμου είναι η επιθετικότητα στον κυβερνοχώρο, όπως ο ηλεκτρονικός εκφοβισμός, ο ρατσισμός και κάθε είδους διαδικτυακή διάκριση. Αυτά τα φαινόμενα είναι επιβλαβή για τους χρήστες ανεξάρτητα από την ηλικία, το εκπαιδευτικό ή το οικογενειακό υπόβαθρο, οδηγώντας τα δυνητικά θύματα σε περιθωριοποίηση, συναισθήματα βαθιάς θλίψης ή ακόμα και σε δράσεις που μπορούν να βιάσουν τον ιδιο τους τον εαυτό. Ως αποτέλεσμα, το κεφάλαιο αυτό περιγράφει και παρέχει μια ανάλυση υψηλότερου επιπέδου της επιθετικότητας στον κυβερνοχώρο και στα κοινωνικά μέσα και παρουσιάζει τις τρεις συνεισφορές αυτού του έργου.

Το Κεφάλαιο 2 καλύπτει τη θεωρητική και μαθηματική βάση που απαιτείται για την κατανόηση και τη διατύπωση των φορμαλισμών που παρουσιάζονται στην προϋπάρχουσα σχετική βιβλιογραφία, δίνοντας έμφαση στις μεθόδους που είναι σχετικότερες με το πλαίσιο αυτής της διατριβής. Οι πιο σχετικοί ορισμοί, έννοιες και παράμετροι ομαδοποιούνται και παρουσιάζονται για να υποστηρίξουν το περιεχόμενο των επόμενων κεφαλαίων.

Το κεφάλαιο 3 ανασκοπεί τη βιβλιογραφία για όλα τα σχετικά ερευνητικά θέματα και τομείς, από τη διάχυση πληροφοριών, τις τεχνικές μεγιστοποίησης επιρροής (influence maximization) και τις μεθόδους βελτιστοποίησης (optimization) μέχρι τα competitive cascades και τις τεχνικές ανοσοποίησης (immunization), με κατάληξη και απώτερο σκοπό την κατανόηση και προσομοίωση της επιθετικότητας (aggression modeling) όπως και την ελαχιστοποίηση αυτής (minimization), αντικείμενα που αποτελούν και τους στόχους της παρούσας εργασίας. Επίσης, παρέχεται μια συγκριτική ανάλυση για να αναδείξει τα προβλήματα των προηγούμενων προσεγγίσεων και να υποδείξει τα ανοιχτά ζητήματα που αντιμετωπίζονται σε αυτή τη διπλωματική εργασία.

Το τελευταίο κεφάλαιο παρουσιάζει την θεωρητική ανάλυση σχετικά με τη μοντελοποίηση της επιθετικότητας. Επιπροσθέτως, παρέχει ψευδοκώδικα για κάθε μέθοδο που υλοποιήθηκε για τη μοντελοποίηση της επιθετικότητας και τεχνικές επεξήγησης (technique of explanation). Αναφορικά με τη μοντελοποίηση της επιθετικότητας, η ακολουθία μοντέλων περιγράφουν ξεκινώντας από τις διάφορες στρατηγικές επιλογής των κόμβων-σπόρων, ή άλλως το πώς πρέπει να επιλέγοντας οι αρχικοί κόμβοι από τους οποίους θα εκκινήσει το μοντέλο διάχυσης. Στη συνέχεια αναλύονται και συζητούνται τα διαφορετικά σχήματα απόδοσης βαρών στο γράφο, οι σχέσεις, δηλαδή, που εκφράζουν τον τρόπο με τον οποίο οι χρήστες συνδέονται, επηρεάζοντας ολόκληρη τη διαδικασία αλλά και τα ίδια τα μοντέλα διάχυσης. Επιπρόσθετα, παρουσιάζονται τα μοντέλα διάχυσης πληροφορίας και
παραλλαγές αυτών και δίνεται έμφαση στις διαφορές που εντοπίζονται ανάμεσα στις τεχνικές που βασίζονται στο IC και σε αυτές που ακολουθούν το LT μοντέλο, που προκύπτουν από το διαφορετικό τρόπο με τον οποίο επηρεάζουν έναν κόμβο κατά τη διαδικασία της διάχυσης. Οι παραλλαγές IC-LT αφεύλονται στην διαφορετική τους οπτική για τους παράγοντες που επηρεάζουν τις διαδικασίες διάδοσης. Από τη μία, το IC βασίζεται σε συγκεκριμένα κριτήρια ενεργοποίησης που εκμεταλλεύονται είτε την επιθετικότητα είτε την αθροιστική επιρροή της γειτονιάς αυτού για να ενεργοποιήσουν τον χρήστη, δηλαδή να θεωρείται ως επηρρεασμένος από την επιθετική πληροφορία που μεταδίδεται. Αντίστοιχα, το LT βασίζεται στην ανάγκη καθορισμού στρατηγικών επιλογής κατωφλίου για κάθε χρήστη που θα λειτουργούν ευεργετικά προς τη συνολική διαδικασία διάδοσης της πληροφορίας. Αυτές οι στρατηγικές καθορίζουν τη φύση της επίδρασης που δέχεται ο χρήστης, επομένως πρέπει να είναι προσαρμοστούν ώστε να αντικατοπτρίζουν τη διάχυση της επιθετικότητας. Μετά τα παραπάνω εκτεταμένα πειράματα εκτελείται εκείνο το σύνολο των παραμέτρων που μοντελοποιεί καλύτερα τη διαδικασία διαχύσης της επιθετικότητας.

Στη συνέχεια, το Κεφάλαιο 5 σε αντιστοιχία με το Κεφάλαιο 4 παρέχει τις θεωρητικές έννοιες καθώς και την περιγραφή της υλοποίησης των μεθόδων που αφορούν την ελαχιστοποίηση της επιθετικότητας. Χρησιμοποιώντας τα δύο καλύτερα σύνολα παραμέτρων, όπως προέκυψαν από τον πειραματισμό του Κεφαλαίου 4, δοκιμάζεται η εφοδιασμός των αντίστοιχων διαφορετικών κομμάτων με την επιθετική πληροφορία. Αρχίζει να δοκιμάζεται μια σειρά μηχανισμών ίασης με σκοπό την περιγραφή της επιθετικής κατωφλίου και την κατανόηση του πειραματισμού. Συγκεκριμένα, η εξέταση των τεχνικών ανοσοποίησης προσφέρει μια εναλλακτική λύση στα competitive cascades, προσφέροντας τεχνικές ανοσοποίησης που εργάζονται με τους διαφορετικούς τρόπους επιδράσεων της επιθετικής κατωφλίου. Επομένως, η εξέταση των τεχνικών ανοσοποίησης προσφέρει μια εναλλακτική λύση στα competitive cascades, προσφέροντας τεχνικές ανοσοποίησης που εφαρμόζονται σε διάφοροι τρόπους επιδράσεων της επιθετικής κατωφλίου. Επομένως, η εξέταση των τεχνικών ανοσοποίησης προσφέρει μια εναλλακτική λύση στα competitive cascades, προσφέροντας τεχνικές ανοσοποίησης που εφαρμόζονται σε διάφοροι τρόπους επιδράσεων της επιθετικής κατωφλίου.
δημιουργούνται όλοι οι πιθανοί συνδυασμοί παραμέτρων. Πρώτον, τα διαφορετικά σύνολα παραμέτρων στη διαδικασία της μοντελοποίησης της επιθετικότητας τίθενται σε δοκιμή και εν συνεχεία, συγκρίνονται οι προσεγγίσεις που αφορούν την ελαχιστοποίηση. Όπως φαίνεται από τα αποτελέσματα και την υψηλή ομοιότητα με το διάνυσμα των πραγματικών δεδομένων, ένα μοντέλο καταρράκτη, όπως το IC και το LT, μπορεί όντως να μοντελοποιήσει τη διάδοση της επιθετικότητας καταλλήλως, παρέχοντας έτσι τα μέσα για την καλύτερη κατανόηση του φαινομένου της επιθετικότητας. Επιπλέον, αυτά τα αποτελέσματα επιτρέπουν τη χρήση των competitive cascades ως τεχνική ελαχιστοποίησης, καθώς αυτά απαιτούν την μοντελοποίηση μέσω ενός μοντέλου cascade. Αποδεικνύεται ότι η προσέγγιση αυτή αποτελεί μια τεχνική που μπορεί να ελαχιστοποιήσει τα συνολικά επίπεδα επιθετικότητας στο δίκτυο σε πολύ μεγαλύτερη κλίμακα από τις μεθόδους που χρησιμοποιούνται σήμερα. Τέλος, το κεφάλαιο 7 περιλαµβάνει μια γενική συζήτηση για το εξεταζόμενο θέμα και προτείνει ορισµένες πιθανές μελλοντικές κατευθύνσεις για διερεύνηση και πειραµατισµό.
Acknowledgements

The elaboration of the present research work presented many difficulties and was particularly time consuming. However, with the right people next to me I was able to have it completed.

First and foremost I would like to thank my supervisor - Pr. Athena Vakali - both for the guidance and the support she has shown me throughout the thesis. Even at times when I lost my faith, she was present and helped me to be reassured, so none of it would be a reality without her own decisive contribution.

Additionally I would like to thank Dr. Nicolas Kourtellis, research of Telefonica I+D department in Barcelona, Spain who stood by my side explaining every ambiguity that was arising during the elaboration of the thesis and providing me with insightful ideas and solutions.

Furthermore, I would like to thank Ilias Dimitriadis, Ph.D. candidate in CS department of Aristotle University of Thessaloniki who helped me with practical ideas during the implementation and writing process and supported me the whole time.
## Contents

**Statement Of Authorship**  i

**Abstract**  iii

**Abstract (Greek)**  vi

**Acknowledgements**  x

**List of Figures**  xiii

**List of Tables**  xv

**Abbreviations**  xvi

### 1 Introduction
1.1 Problems addressed  4
1.2 Thesis Contribution  6
1.3 Thesis Structure  7

### 2 Theoretical Background - Fundamentals
2.1 Network Basics  10
2.2 Information Diffusion and Influence Maximization  11
2.3 Statistical Tests
   2.3.1 ANOVA  14
   2.3.2 Tukey’s HSD post-hoc  15

### 3 Literature Review
3.1 Information Diffusion Models  17
3.2 Influence Maximization  19
3.3 Competitive Cascades
   3.3.1 Competitive Influence Maximization  22
   3.3.2 Influence Blocking Maximization / Influence Minimization  23
3.4 Immunization  24
3.5 Aggression Minimization  26
4 Aggression Diffusion Modeling
  4.1 Seed Selection
    4.1.1 Theoretical Analysis
    4.1.2 Specifications
  4.2 Weighting Schemes
    4.2.1 Theoretical Analysis
    4.2.2 Specifications
  4.3 Diffusion Model (Activation Criteria and Aggression Transferring)
    4.3.1 Theoretical Analysis
      4.3.1.1 Independent Cascade
      4.3.1.2 Linear Threshold
    4.3.2 Specifications

5 Aggression Minimization
  5.1 Competitive Aggression Minimization
    5.1.1 Theoretical Analysis
    5.1.2 Specifications
  5.2 Blocking Aggression Minimization
    5.2.1 Theoretical Analysis
    5.2.2 Specifications

6 Experimentation - Results
  6.1 Dataset Description
  6.2 Results
    6.2.1 Aggression Modeling
      6.2.1.1 Seed size and Aggression Threshold
      6.2.1.2 Seed Selection Strategy
      6.2.1.3 Weighting Scheme
      6.2.1.4 Activation criterion and Threshold strategy
      6.2.1.5 Snapshot evaluation
    6.2.2 Aggression Minimization
      6.2.2.1 Competitive Aggression Minimization
      6.2.2.2 Blocking Aggression Minimization

7 Discussion and Future Work

A Mathematical Notation
  A.1 General Notation
  A.2 Set Notation
  A.3 Logic Notation
  A.4 Statistics Notation

Bibliography
# List of Figures

1.1 Number of social media worldwide ........................................... 1
1.2 Social media usage by age group and country ................................. 2
1.3 Social media cyberbullying intensities ........................................... 4
1.4 Holistic process pipeline .............................................................. 7
2.1 A simple undirected graph ............................................................ 10
2.2 A directed graph ........................................................................... 10
2.3 A weighted graph ......................................................................... 10
2.4 Different types of graph ................................................................. 10
4.1 Overall aggression modeling process .............................................. 27
4.2 Seed Selection component .............................................................. 28
4.3 Weighting Scheme component ....................................................... 33
4.4 Activation Criteria component ....................................................... 36
4.5 Aggression Transferring component .............................................. 36
5.1 Aggression minimization process ...................................................... 41
6.1 Cosine similarities Tukey’s HSD pairwise significance test for IC experiments with regard to the seed strategy ........................................... 54
6.2 Cosine similarities Tukey’s HSD pairwise significance test for "Single Discount" seed strategy on IC experiments with regard to the weighting scheme ........................................................................... 56
<table>
<thead>
<tr>
<th>Section</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>6.3</td>
<td>Cosine similarities Tukey’s HSD pairwise significance test for &quot;Single Discount&quot; seed strategy and &quot;Jaccard Overlap&quot; weighting scheme on IC experiments with regard to activation criterion</td>
</tr>
<tr>
<td>6.4</td>
<td>Snapshot evolution for the best IC and LT models and three different metrics: Cosine Similarity, Pearson and Spearman coefficients</td>
</tr>
<tr>
<td>6.5</td>
<td>Aggression evolution for different seed strategies on IC models with respect to the case where no healing is applied. Healing strategies are: Decaying transfer, Hybrid, Transfer and Vaccination</td>
</tr>
<tr>
<td>6.6</td>
<td>Activations of the IC configurations</td>
</tr>
<tr>
<td>6.7</td>
<td>Aggression evolution for different seed strategies on LT models with respect to the case where no healing is applied. Healing strategies are: Decaying transfer, Hybrid, Transfer and Vaccination</td>
</tr>
<tr>
<td>6.8</td>
<td>Activations of the LT configurations</td>
</tr>
<tr>
<td>6.9</td>
<td>Aggression evolution for different blocking mechanisms based on node or edge removal and normal or aggression-based matrix for IC (left) and LT (right) models</td>
</tr>
</tbody>
</table>
List of Tables

2.1 Basic Symbols Notation ................................................. 9

3.1 Method classification for information diffusion ...................... 19

3.2 Method classification for the seed selection process .......... 22

3.3 Method classification for minimization techniques .......... 25

6.1 Configuration encoding and parameter options .................. 52

6.2 Cosine Similarities for LT experiments ......................... 55

6.3 Configuration encoding and parameter options of aggression minimization experimental phase ........................................ 60

6.4 Activated nodes for different blocking mechanisms on IC and LT models  65
### Abbreviations

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Expanded Form</th>
</tr>
</thead>
<tbody>
<tr>
<td>OSN</td>
<td>Online Social Network</td>
</tr>
<tr>
<td>LT</td>
<td>Linear Threshold</td>
</tr>
<tr>
<td>IC</td>
<td>Independent Cascade</td>
</tr>
<tr>
<td>ANOVA</td>
<td>ANalysis Of VAriance</td>
</tr>
<tr>
<td>HSD</td>
<td>Honestly Significant Difference</td>
</tr>
<tr>
<td>SA</td>
<td>Simulated Annealing</td>
</tr>
<tr>
<td>CELF</td>
<td>Cost Effective Lazy Forward selection</td>
</tr>
<tr>
<td>DAG</td>
<td>Directed Acyclic Graph</td>
</tr>
<tr>
<td>PHCA</td>
<td>Predictive Hill Climbing Algorithm</td>
</tr>
<tr>
<td>DD</td>
<td>Degree Discount</td>
</tr>
<tr>
<td>SD</td>
<td>Single Discount</td>
</tr>
<tr>
<td>CAM</td>
<td>Competitive Aggression Minimization</td>
</tr>
<tr>
<td>BAM</td>
<td>Blocking Aggression Minimization</td>
</tr>
<tr>
<td>MCICM</td>
<td>Multi-Campaign Independent Cascade Model</td>
</tr>
<tr>
<td>COICM</td>
<td>Campaign-Oblivious Independent Cascade Model</td>
</tr>
</tbody>
</table>
Devoted to my parents, brother and my good friend Ilias Dimitriadis who supported me and were by my side during the whole time of this dissertation...
Chapter 1

Introduction

Nowadays, social media have become an extension of our everyday routine and habits. For example Facebook alone, which is currently the most used and largest social media platform in the world, counts a total of approximately 2.4 billion users. Other well-known social media applications and platforms such as Whatsapp and Youtube also count a tremendous number of users who are over than 1 billion. The aggregated statistics of these facts are presented in [58] and depicted in Figure 1.1 [58]. These numbers are huge but yet they drive today’s online social reality since in a global population of 7.5 billion people worldwide [51], more than 3.5 billion are online [48]. These numbers express the fact that social media platforms are not only used by more than two-thirds of the total amount of internet users but also by more than a third of the human population.

![Number of social media worldwide](image)

**Figure 1.1:** Number of social media worldwide
Social media have changed the world! The rapid and vast adoption of these technologies is changing several activities such as: how we find information, how we access information from the news, how we interact with others and how we impact political change. We open our Facebook or Instagram profile to post our new achievements or to share our thoughts on an incident on our way to work. We use Twitter to encore the public speak of a politician or to comment on a friend’s opinion. A lot of people check on their social media wall first thing in the morning as well as before going to sleep, let aside their exposure during the day. Of course, this situation is dominant not only between adults but even more among children and teenagers. Today’s youth generations are born into the world of social media and get used to being interconnected quite early in their lives, sometimes even before learning how to operate and behave in the frame of society.

The numbers arising indicate a great amount of heterogeneity across different platforms since the various social media platforms attract different population groups and thus are more popular among specific targeted groups. Generally, young people feel a special affinity towards social media than older people and hence they tend to use them a lot more as shown in Figure 1.2. But some platforms are much more popular among younger people. On the other hand, apart from age differences, we observe gender or even income-related ones too. Pinterest, for example, is much more popular among women compared to men while social media usage in Sweden or the United Kingdom is much bigger than in Mexico as it is again depicted in the second image of figure 1.2.

The rise of online aggression phenomena: Except for the communication horizons or the dynamic and worldwide interconnection that social media have made possible, many dystopic phenomena have also risen. More specifically, the large heterogeneity discussed above in combination with the improper education of people regarding social media usage leads to phenomena of online aggressive behaviors such as cyberbullying, racism or any other kind of discrimination. In fact, in many cases, these dangers are very latent and barely noticeable as they gradually become a daily behavior without anyone noticing the negative effects. For younger ages, verbally harassing others in
public while avoiding personal confrontation is a hallmark of passive-aggressive behavior. This form of **aggression** is a masked but deliberate way of expressing feelings of anger, distrust or sadness. Oftentimes it is carried out online through active behaviors such as posting embarrassing photos or through passive and subconscious ones, such as failing to stop the spread of online gossip. Furthermore, Figure 1.3 shows in which social media cyberbullying phenomena are more prevalent with their intensities been highlighted. As it can be observed, the problem is not restricted to a specific platform but it is rather omnipresent and it has a critical global impact.

With respect to the above considerations, a more clear definition of online aggression and its most prevalent form of expression, cyberbullying, should be provided. However, up to now there is no uniform definition of online aggression - as discussed in [19]. Trying to approximate proper definitions of online aggression and cyberbullying, as they are critical in detecting and facing the roots of online aggression, [29] is the reference point. Specifically, online aggression is defined as the intentional harm delivered via electronic means to a person or a group of people who perceive such acts as offensive, derogatory, harmful or unwanted. Furthermore, according to the same article, cyberbullying typically denotes repeated and hostile behavior performed by a group or an individual.

**The effects of online aggression:** But why are these phenomena that critical? The content every individual shares online - whether it is any kind of personal content or any negative and possibly hurtful post - gradually outlines a permanent public record of their activity, ideas and points of view. In other words, it impacts users online reputation, which is easily accessible to schools, employers, any kind of group, and others who might be interested in an individual either in the present or in the future [27]. Online aggression phenomena can harm the online reputations of not only the individual that gets bullied but also of the perpetrators, affecting everyone that gets involved in such an incident [23]. The negative effect of aggressive phenomena is amplified by three of their core traits [6]:

1. **Online permanence:** as the information that gets published online stays online and can be accessible by anyone, if it does not get reported or removed. Thus negative information can influence many different aspects of one’s life such as college or work approval.

2. **Hidden influence:** since parents and teachers may not notice, overhear or observe aggressive phenomena manifesting, it is more difficult to recognize them.

3. **Omnipresence:** digital devices and social media provide us with the ability to communicate immediately and continuously 24 hours a day, making it difficult for
victims (children or adults) experiencing a form of online aggression expression to find relief from it.

**Aggression Spreading**: In addition to the negative effects of online aggression on both victims and perpetrators, another factor constituting dealing with them crucial is their **dynamic and fast spreading behavior** over the users of the social network. In [33] the authors have pointed that users can be influenced by their social circle to exhibit aggressive behaviors and even bully others due to elevated toxicity and aggression levels in this exact circle of influence. In a similar fashion, this kind of behavior can easily manifest in the online world as well. Thus, aggression can spread over the network by propagating from one user and neighborhood to another, and therefore, raise an overall negative user experience. To reinforce this statement of the necessity of immediate response to online aggression spreading - or diffusion - there are some early works in the domains of psychology and sociology that have already proposed models of online digital abuse phenomena based on the well-studied theories of social learning and bonding, as well as the theory of planned behavior [39].

### 1.1 Problems addressed

The two characteristics, negative effects and dynamic and fast spreading behavior, of online aggression presented above emphasize the importance of taking countermeasures to restrain the manifestation of aggressive online phenomena. The enormous negative results of aggressive behaviors in social networks and in the cyber-world generally are
delineated by the special attention and big effort on the specific topic in academia. Researchers have studied aggression and cyberbullying from the perspective of social psychology and sociology [7, 52] proposing a solid theoretical framework for an integrative, bio-social-cognitive, developmental approach to understand aggression. These frameworks consist the foundational basis of the notions implemented in this thesis. On the contrary, there are also algorithmic efforts in the domain of computational social science [12, 20, 25, 60] that create applied models able to help in the mitigation of the negative effects of the online aggression phenomenon. This category of models provide the predictive mechanisms for the distinction of users to aggressive and normal and clarify the computational methodology on which this thesis is based. The different approaches in combination with the lack of a uniform definition suggest that there are various ways to address online aggression depending on the severity of the behavior, the power of the aggressor over the victim and other factors [19]. As a result, this ambiguity introduces a polarity. On one hand, it creates obstacles in online aggression confrontation as it does not provide a concrete basis to construct an approach on, which will be able to model and better understand the mechanics of online aggression. On the other hand, it offers the opportunity to study online aggression and its confinement from different points and leverage the most efficient methods towards the single goal of online aggression restriction.

The importance of online aggressive behavior phenomena confrontation however, is already noticeable by the various social media platforms. These platforms try to mitigate - if not eliminate - the negative effects of online aggression by blocking abusive users, reporting and content removal [57]. To detect and ultimately block abusive users, they exploit methods such as NetShield [55] and NetMelt [54] which aim to detect those users (or relations) that are most likely to be infected by an epidemic spread of a virus (or spread a virus) - where virus likens to online aggression. Nevertheless, the policy that these methods enforce is strict and affects negatively user experience and satisfaction. Additionally, the methods currently in use are effective in detecting abusive behavior generally instead of focusing on the specific domain of aggression detection.

In summary the problems addressed in this thesis are:

- **Problem 1**: How can online aggression diffusion be modeled in order to understand the phenomenon and its underlying spreading mechanisms.
- **Problem 2**: How can the overall aggression level in the social network be minimized while user experience and satisfaction are preserved.
1.2 Thesis Contribution

Based on the above considerations, in this thesis we investigate the two introduced problems, namely aggression diffusion modeling and aggression minimization, from the perspective of a cascading diffusion process and specifically its two basic models, Independent Cascade (IC) and Linear Threshold (LT). The selection of this approach is based on two factors. First, cascades are processes that evolve in a front-like manner, meaning that a user gets influenced by many neighbors at once, a property that matches the way that an aggressive post is spread through social media [33]. Second, if cascades can indeed model aggression diffusion then we can also exploit them in aggression minimization. Hence, we can substitute the strict methods used today, such as banning a user due to a post, with a cascade-based one improving user experience. For example, the implementation of a cascade-based minimization method consists from an educational post, which targets specific users to spread it to their social circle with ultimate goal to eliminate aggressive behaviors. Thus, not only the aggressiveness is reduced but also users are more satisfied as they do not get penalized that strictly.

The approach of the thesis can be described through a pipeline of processes, which is presented in Figure 1.4. It starts with the data collection and ground truth creation processes, which are necessary both in implementing the various aggression diffusion models and in comparing them to the original state of the network. Then, it continues with the investigation of the best aggression modeling method based on the similarity to ground truth, meaning that the model that best approximates the real network status is the most suitable to describe aggression dissemination. Last, the pipeline ends with the detection of the best aggression minimization method, where competitive cascade are compared to immunization or blocking methods. Thus, this process addresses both problems by finding the model that best describes aggression propagation and leveraging its properties to achieve the best possible aggression minimization in the social network.

The main contributions of this work are presented:

1. C1: We show that both IC and LT are valid and appropriate models for aggression diffusion in social networks, meaning that they can efficiently and robustly describe aggression propagation. To test that, we compare the different IC and LT-based configurations to the ground truth vector using three different metrics, cosine similarity (the main metric), Pearson R and Spearman R statistics. Additionally, we observe that all configurations converge to a stable similarity score during the last steps of the propagation rather than oscillating indefinitely proving the consistency of the two diffusion models.
2. **C2**: We propose common strategies - among IC and LT - for both of these models to better understand the dynamics of the aggression phenomenon. Specifically, we show that a strategy that focuses on central - in terms of influence - network users is appropriate for the initialization of the diffusion process and the relations between users should be weighted according to the similarity of their neighborhoods.

3. **C3**: We extend our study to aggression minimization by exploiting methods from the competitive cascade domain and show that we can not only mitigate but also reduce the overall users’ aggressiveness at the end of the process while we also apply a more tolerant user policy that enables user satisfaction. Extensive experimentation on the proposed methods displays a reduction of approximately 50% for the IC and 15% for the LT models in comparison to the blocking methods currently in use.

4. **C4**: We validate our results through statistical analysis and tests showing that there are significant differences between the tested parameter values. That is, there are specific and well-defined parameter values that comprise a concrete best aggression diffusion model and minimization method.

### 1.3 Thesis Structure

Next, in Chapter 2 the fundamentals and theoretical background needed to understand this work will be presented. Chapter 3 presents the related literature and the methods on which the proposed models are grounded. Chapter 4 explains the methodology and the specifics for the algorithmic implementation of aggression diffusion modeling while Chapter 5 presents the same notions from the aggression minimization point of view. Next, Chapter 6 demonstrates the results of the experimentation on both aggression
modeling and minimization and last, Chapter 7 introduces a discussion on the overall work and future research directions.
Chapter 2

Theoretical Background -
Fundamentals

Chapter 1 presented the overall problem of aggressive behavior in online social media and clarified the great challenges involved. Before we proceed to our proposed method of aggression modeling and aggression minimization we define the various concepts and notations of social networks, information diffusion and minimization that will be exploited to address the problem under investigation. Additionally, Table 2.1 contains the notations presented below.

This chapter’s purpose is to create the theoretical basis and background of social networks and information diffusion, as well as to clarify the critical differences among them and examine the sub notions and definitions that underlie the aforementioned topics.

<table>
<thead>
<tr>
<th>Symbol</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>G</td>
<td>Graph or Network</td>
</tr>
<tr>
<td>V</td>
<td>Set of graph vertices</td>
</tr>
<tr>
<td>E</td>
<td>Set of graph edges</td>
</tr>
<tr>
<td>A</td>
<td>Adjacency matrix of graph</td>
</tr>
<tr>
<td>N_u</td>
<td>Set of neighbors of node u</td>
</tr>
<tr>
<td>k</td>
<td>Budget or seed size</td>
</tr>
<tr>
<td>S</td>
<td>Set of seed nodes</td>
</tr>
<tr>
<td>σ</td>
<td>Influence spread</td>
</tr>
<tr>
<td>θ</td>
<td>Node threshold</td>
</tr>
</tbody>
</table>

Table 2.1: Basic Symbols Notation
2.1 Network Basics

The core and basic entity to define is the structure of an underlying social network, which is captured by a so called "network" or a "graph". A formal definition of a graph is given below:

**Definition 1 (Graph).** A graph $G$ is an ordered pair $G = (V, E)$, where $V$ is a set of vertices or nodes, and $E \subseteq \{(x, y)\mid (x, y) \in V^2 \land x \neq y\}$ is a set of edges or links.

In the case of social networks, a social network is a graph whose vertices are the users of the corresponding social media, such as Twitter, Facebook etc., and its edges are the relationships between users. The way we formulate these relationships has an immediate impact on the structure and the information that the network provides. Enriching the generic definition of a Graph, the below two notions arise:

**Definition 2 (Undirected Graph).** An undirected graph $G$ is an ordered pair $G = (V, E)$, where $V$ is a set of vertices or nodes, and $E \subseteq \{(x, y)\mid (x, y) \in V^2 \land x \neq y\}$ is a set of edges or links, while an edge is an unordered set of vertices meaning that the relation between vertices $v_1$ and $v_2$ that are connected through an edge $e$ is reciprocal.

**Definition 3 (Directed Graph).** A directed graph $G$ is an ordered pair $G = (V, E)$, where $V$ is a set of vertices or nodes, and $E \subseteq \{(x, y)\mid (x, y) \in V^2 \land x \neq y\}$ is a set of edges or links, while an edge is an ordered set of vertices meaning that the relation between vertices $v_1$ and $v_2$ that are connected through an edge $e$ is a one-way relationship with $v_1$ as a source and $v_2$ as the target.

Each social media platform defines the relationships between users in a different way, e.g. Facebook prefers to create undirected edges capturing the notion of "friend" while Twitter sticks to the directed edges as they express the phenomenon of "following" and
not "friend" someone. There is a plethora of different graph types such as acyclic graphs, multigraphs etc., but they fall outside the scope of this work and thus they are omitted.

Having discussed the direction of an edge’s expressed relation the intensity of this exact relation is now considered.

**Definition 4** (Weighted Graph). A weighted graph $G$ is an ordered pair $G = (V, E)$, where $V$ is a set of vertices or nodes, and $E \subseteq \{(x, y, w)\} | (x, y) \in V^2 \land x \neq y \land w \in \mathbb{R}^+ \}$ is a set of edges or links, with each edge having a weight $w$ showing the intensity of the interaction between the two nodes.

From definition 4 it is evident that the bigger the weight, the more intense the interaction of the two users. Figure 2.4 shows the discussed types of graphs. Specifically, the third one is a weighted graph. A weighted graph can be either directed or undirected. In addition, weights can be integers or floats, supposing they follow a global fashion for the whole graph. Weight calculation is not restricted to a specific type and it could be anything from a uniformly picked number to a more sophisticated notion that captures the community structure of the interacting nodes.

Having described the various graph types a way to represent a given graph is also needed so as to make algebraic manipulations and extract useful knowledge. The most common representation of a graph, and by extension of a social network, is the adjacency matrix presented in definition 5. This representation will be useful for the implementation of the immunization and blocking techniques of section 5.2.

**Definition 5** (Adjacency Matrix). Given an undirected and unweighted network with $n$ users, its adjacency matrix $A \in \mathbb{R}^{n \times n}$ is a symmetric matrix where cell $(i, j) = 1$ if there is a link between users $i$ and $j$. In the directed network case, matrix $A$ is not symmetric and cell $(i, j) = 1$ only if there is a link from user $i$ to user $j$. Regarding the case of a weighted matrix, the value of a cell equals the weight of the corresponding edge.

### 2.2 Information Diffusion and Influence Maximization

A social network is represented by a weighted directed graph $G = (V, E, p)$ where $V$ represents a set of nodes, $E$ a set of edges and $p$: $V \times V \rightarrow [0, 1]$ the probability of an edge to propagate information. Additionally, for each node $N_u$ denotes the set of $u$’s neighbors, $N_u = \{v | (u, v) \in E\}$. 
Having defined the different types of networks that are related to the topic of this dissertation, we should now discuss the process of information diffusion. Generally speaking, information diffusion is the process of information spreading from an initial node, or an initial set of nodes, to the rest of the network taking advantage of the existing links and relations between the network’s nodes. The main goal of information diffusion is to create a model that best describes the propagation of information taking into consideration multiple and diverse factors such as time \cite{15, 13}, predetermined attitude against the specific type or instance of information \cite{14} or even the existence of a number of competing information sources \cite{61, 14}. Alongside information diffusion, another tightly related term is influence maximization, whose aim is to find several influential users that will lead to the maximization of information propagation in social networks. In order to measure the influence of a node set a proper function is needed:

**Definition 6** (Influence Spread). Given an initial seed set \( A \subseteq V \) in network \( G \), the influence spread of \( A \), denoted by \( \sigma(A, G) \), is the expected number of active nodes at the end of the process.

Given a set of \( S \) edges not in \( E \), we also denote by \( G(S) \) the graph augmented by adding the edges in \( S \) to \( G \), \( G(S) = (V, E \cup S) \). Thus a formal definition of the problem of influence maximization can now be provided:

**Definition 7** (Influence Maximization). Given a social network \( G = \{V, E\} \), a vertex set \( A \subseteq V \) and an integer \( k \) (budget), the problem of influence maximization lays in finding a set \( S \) of edges incident to the nodes in \( A \) (that is \( S \subseteq \{(\alpha, v) : v \in V \setminus N_{\alpha}, \alpha \in A\} \)) such that \(|S| \leq k \) and \( \sigma(A, G(S)) \) is maximum.

The maximization of a function such as the influence spread function, given a limited budget \( k \) proves to be too cumbersome to solve and it has been shown that it is an NP-hard problem \cite{38}. Fortunately, under Independent Cascade and Linear Threshold models (that will be discussed later), two basic function properties can be exploited to approximate the optimal solution, monotonicity and submodularity.

An influence spread function \( f \) is monotone in the sense that adding an element to a set cannot cause \( f \) to decrease: \( f(S \cup \{v\}) \geq f(S) \) for all elements \( v \) and sets \( S \). In addition, \( f \) is submodular if it satisfies a "diminishing returns" property: the marginal gain from adding an element to a set \( S \) is at least as high as the marginal gain from adding the same element to a super set of \( S \). Formally, a submodular function \( f \) satisfies

\[
 f(S \cup \{v\}) - f(S) \geq f(T \cup \{v\}) - f(T),
\]

for all elements \( v \) and all pairs of sets \( S \subseteq T \).
With these two properties present an approximation of the optimum result within a factor of \((1 - 1/e)\) can be achieved \[46\], where e is the base of the natural logarithm. The algorithm that achieves this approximation is a greedy algorithm, starts with the empty set and repeatedly adds the element that gives the maximum marginal gain.

Generally, all of the methods and approaches used to address influence maximization suppose that the nodes lay in two categories, active and inactive, which describe whether a node has been influenced by the specific information instance or not. Thus, consequently the methods that are going to be implemented in this work: Independent Cascade Model and Linear Threshold Model shall be described.

In Independent Cascade (IC) model the information flows over the network through a cascade. Nodes appear in two states, active: node has been influenced by the information instance, and inactive: node is unaware of the information instance. The process evolves in discrete steps. Initially, few nodes, the seed nodes, are influenced by the information instance, thus they are regarded as active. In each following discrete step, an active node attempts to influence one of its inactive neighbors. Regardless of its success, the specific node is not given a second chance to activate the same inactive neighbor. The success of a node in activating one of its neighbors depends on the propagation probability of the corresponding edge. Finally the process terminates when no further nodes get activated.

In Linear Threshold (LT) model, there are the same two node states: active and inactive. Analogously to IC, LT starts with a seed set comprised of active nodes, but in contrast to IC, every node in the network has a threshold \(\theta\) that is selected uniformly at random. At each discrete step \(t\), an inactive node becomes active if the sum of the weights of the edges with active neighbors exceeds the threshold \(\theta\).

\[
\sum_{w \in \text{adj}[v]} b_{v,w} \geq \theta_v
\]

The LT process terminates in a similar fashion to IC, when no further nodes get activated.

### 2.3 Statistical Tests

The last issue to address from a theoretical aspect is the statistical tests used during the experimental process in Chapter 6. There we make use of two statistical tests in order to verify the validity and significance of the results. The first test is called analysis of variance (ANOVA) and it is used to determine whether there is a significant effect of a specific parameter value when a plethora of parameter values is under comparison. The
second is called Tukey’s Honestly Significant Difference (HSD) post-hoc test and used in case of ANOVA’s null hypothesis rejection in order to determine between which pair of parameter values the difference is spotted. Below we analyze thoroughly these two tests.

2.3.1 ANOVA

ANOVA let us compare differences of means between two or more groups. To do so, it examines the variations in the data and where exactly they are spotted. In other words, ANOVA compares the inter-variation with the intra-variation. Inter-variation is the variation between groups while intra-variation is the amount of variation within groups.

When we sample from a population there usually is a sampling error in the mean value due to the fact that we do not use the entire population. However, ANOVA points out whether the difference between the means of the various groups is greater than the expected just by chance or due to a true difference in the population mean.

The mathematical formula of ANOVA is given below:

\[ x_{ij} = \mu_i + \epsilon_{ij} \]

where \( x \) are the data points, \( i \) and \( j \) denote the group and the individual observation, \( \epsilon \) is the unexplained variation and \( \mu \) are the population means for each group. Thus, each point \( x_{ij} \) is its population mean plus an error.

In ANOVA, as well as other classical statistical tests, we calculate a test statistic called \( F \)-ratio with which we can obtain the probability \( P \)-value of obtaining the data assuming the null hypothesis. As null hypothesis we define the case where all population means are equal and hence the alternative hypothesis is that there is at least one population mean which is significantly different from the rest. A significant \( P \)-value - often taken as \( P<0.05 \) - suggests that the null hypothesis can be safely rejected.

ANOVA splits the calculation into two steps. The first one contains the calculation of between group variation which is calculated by comparing the mean of each group with the overall mean of the data, that is:

\[ \text{BetweenSS} = n_1(\bar{x}_1 - \bar{x})^2 + n_2(\bar{x}_2 - \bar{x})^2 + ... + n_m(\bar{x}_m - \bar{x})^2 \]

assuming \( m \) groups. Then, we divide BetweenSS by the number of degrees of freedom - similar to sample size, with the difference that it is \( n-1 \), because the deviations must
sum to zero, and once you know n-1, the last one is also known - and as a result we get the estimate of the between groups mean variation.

The second step contains the within group variation which is the variation of each observation from its group’s mean. The definition is:

$$SS_R = s_{1}^2(n_{1} - 1) + s_{2}^2(n_{2} - 1) + ... + s_{m}^2(n_{m} - 1)$$

again for m groups. The above formula expresses the summation of variance of each group times the degrees of freedom of each group. Last we divide once more by the total degrees of freedom to get the mean variation within groups.

Having calculated the between and within groups variation we can now compute the F-ratio as the fraction: $\frac{\text{MeanBetween}}{\text{MeanSS}_R}$. If the average difference within groups is similar to that of between groups then F-ratio is approximately 1. The bigger the average difference between groups comparing to that of within groups. the larger the F-ratio. Additionally, the P-value is obtained by testing the F-distribution of a random variable with the degrees of freedom associated with the numerator and denominator of the ratio. Then, P-value expresses the probability to acquire this F-ratio or a greater one. This means that the larger the F-ratios the smaller the P-values.

Concluding, we should present the assumptions of ANOVA test. First of all, it supposes that the response data is normally distributed. It perceives the within variance of different groups to be similar. Last, it supposes that the data points are independent.

### 2.3.2 Tukey’s HSD post-hoc

Tukey’s HSD post-hoc test is a statistical test to determine if there is a significant difference in the relationship between two data sets. In other words, it tests if there is a strong possibility that a numerical change in one value is causally related to a change observed in another value. It is used after the null hypothesis of ANOVA test has been rejected to spot exactly between which parameter group the significant difference is spotted. Moreover, its null hypothesis states that all means under comparison are from the same population. As a result the alternative hypothesis states that there are means belonging to different populations.

Tukey’s HSD value is calculated using the absolute value of the difference between pairs of means divided by the standard error of the mean as determined by the previous
ANOVA test. In terms of mathematical definition, the HSD value for each pair of means is given by the formula:

\[ HSD = \frac{|\mu_i - \mu_j|}{SE} \]

where \( \mu_i - \mu_j \) is the difference between the pair of means and \( SE \) is the standard error of the sum of means.

Similar to ANOVA test, Tukey’s HSD has also some assumptions. Specifically, the groups for each mean in the test are normally distributed and observations are independent both within and among groups. Furthermore, there is equal within-group variance across the groups associated with each mean in the test, a property called homogeneity of variance.

**Conclusion:** In this chapter all the required notions that will be helpful in building the information diffusion models in order to formulate the aggression dissemination process as well as its minimization have been presented and explained. Additionally, this information will be beneficial in understanding the research effort that has been made in the various related domains - presented in Chapter 3 upon which the proposed methods will be grounded.
Chapter 3

Literature Review

Chapter 3 reviews all the existing methods and algorithms pointing out their intuition, best features but their drawbacks as well. In this work, we are addressing two different problems, aggression modeling and aggression minimization. To model aggression propagation efficiently, a good understanding of the methods related to information diffusion, influence maximization and their optimization is needed. For the purpose of aggression diffusion modeling, cascading models are chosen as they inherently contain the notion of team and neighborhood-based influence that is applied on a node when a post is spread through social media. Proceeding to the second problem that this thesis addresses, aggression minimization, the presented methods lay in the domains of competitive cascades, information minimization and immunization algorithms. The reason behind this selection emerges from the results of the aggression modeling process. If cascades prove to model aggression diffusion sufficiently, then competitive cascades can be exploited to minimize the overall aggression level of the network. However, if this is not the case, then immunization techniques - the approach that is currently preferred by the social media platforms with algorithms such as NetShield [55] - can be leveraged. Additionally, it is noted here that the immunization methods should be adjusted on the specific minimization task of aggression to achieve the best results as they currently aim to detect and limit abusive behaviors in general.

3.1 Information Diffusion Models

Cascade Models: A lot of attention has been placed in modeling information diffusion for the last two decades. Initial efforts, such as the work of Domingos and Richardson [22, 50], have approached the problem with probabilistic models. Based on these works, Kempe et al. [38] addressed the issue as a discrete optimization problem. Specifically,
they proved that the problem is NP-hard and proposed a greedy algorithm that approximates the optimal solution within a $(1 - 1/e)$ error boundary. They show that this boundary can be ensured due to two properties of the algorithm’s objective function, called: (a) monotonicity which is the well known mathematical property and (b) submodularity which takes advantage of the diminishing returns property [10], which means that the value gained by a single piece of information is greater the smaller the previous exposure to similar information is. In this context, Kempe et al. proposed the Independent Cascade (IC) and Linear Threshold (LT) diffusion models. These two models let the process unfold in discrete time steps, while they address each node only once and affect a set of nodes rather than a single one, thus they develop in a front-like manner. They are also characterized as progressive models due to the single, non-repetitive impact on each node. Moreover, they are considered as the fundamental components inspiring many other works which aim to improve greedy algorithm’s efficiency [10, 16, 18, 37] and will be the core of the aggression modeling methodology followed in this thesis. The reason behind this choice is their inherent property to address diffusion as a holistic process, where in each step a node gets affected by a set of neighbors instead of just a single one. This fact is in accordance with the property of online aggression phenomena to be transmitted to users through the cumulative influence of their social circle [33].

**Opinion Dynamics Models:** A second category of information diffusion models address the dissemination process from the perspective of opinion dynamics field. This category contains the well studied Voter model [24] where a user gets influence by one of its neighbors through a random "voting" process, as well as the SIS and SIR and SIRS models [30, 36, 56] from the field of epidemics. Regarding Voter model, it is categorized in the domain of non-progressive symmetric models. More specifically, the symmetric nature of these models suggests that when a node gets affected by a piece of information, this state is not final but it may change in a future step of the diffusion process. Thus, these models require specific conditions in order to ensure opinion convergence or termination of the process in contrast to LT, IC and their derivatives. SIS and SIR on the other hand, comprise a broader umbrella term and contain all those models that categorize the users into Susceptible and Infected (for SIS) or even Recovered (for SIR and SIRS) users. In fact, IC is a special case of the SIR model. However, despite the fact that they are able to simulate the diffusion process itself, their application mainly aims at explaining the structure of the network and the dynamics of opinion formation, rather than the propagation process, rendering them inappropriate for the problem under investigation. Howbeit, the investigation of dynamics is useful for determining the factors that affect the diffusion process and in extention the aggression dissemination studied in the thesis.
Summing up the above models for information diffusion, Table 3.1 presents the type of model as well as how these methods approach information diffusion. Following, a specific topic of information diffusion, namely Influence Maximization, is discussed with special emphasis on the seed strategy selection mechanisms as they comprise a crucial part of this work.

<table>
<thead>
<tr>
<th>Model Type</th>
<th>Information Diffusion Approach</th>
<th>Method</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cascade</td>
<td>Progressive</td>
<td>IC and LT [38]</td>
</tr>
<tr>
<td>Opinion Dynamics</td>
<td>Symmetric</td>
<td>Voter [24], SIS, SIR, SIRS [30, 36, 56]</td>
</tr>
</tbody>
</table>

Table 3.1: Method classification for information diffusion

### 3.2 Influence Maximization

The methods discussed in the previous section are sufficient to model information diffusion in a controlled environment. However, in many real world scenarios there is a scarcity in available resources and the utmost purpose of the diffusion models is to maximize the effect of the propagating piece of information. This is the case for digital marketing ads or even vaccination of infected populations. Out of this necessity, a different research topic has emerged called influence maximization [9, 14, 53, 42]. The main goal of the influence maximization approaches is to maximize the expected influence in the network or even find the most influential users. Under this context, the limitation in resources is translated into a specific number of seed nodes that can be initially affected by the information. Thus in the case of expected influence maximization the aim is to find the best method starting from the specific seed nodes, while influential node detection completes this process by selecting as seed nodes these specific influencers.

**Occurring phenomena:** Notwithstanding, there are many phenomena to take into consideration while addressing influence maximization. Chen et al. [14] aim at maximizing the influence of a piece of information when negative opinions are present by introducing a quality factor. An example of this phenomenon is the case of a deficit product that causes a user to turn negative towards it. Additionally, Tong et al. [53] propose a model where the selection of the nodes to start the diffusion process is dynamic instead of offline and adapts to network’s state at each time step. In this way, it captures the volatility that the diffusion process inserts into the corresponding network. Moreover, [9] and [42] investigate the problem in a competitive cascade context where two or more competing cascades are present.
Seed Selection Process: Apart from addressing the various phenomena underlying the maximization process, there have been several other approaches focusing on optimizing the seed selection process itself. This category of methods is defined as minimum cost seed selection. In this approach, the main objective is to minimize the cost needed for the selection of nodes that will act as seeds for the overall information cascade, while the spread of the cascade is kept above a specified threshold. Thus, these mechanisms can be exploited to enable the overall aggression diffusion process that is investigated in this work. Specifically, Kempe et al. in their work have shown that greedy algorithm outperforms methods based on network and node properties. For this reason, recent influence maximization techniques turn to the usage of greedy algorithm in order to select the appropriate seed set that can achieve the largest node activation at the end of the process. However, the simple version of the greedy technique is infeasible and for that many optimization methods have been proposed.

Generic Strategies: The most significant efforts address both types of cascading models - IC and LT - and thus they consist an important reference point. Leskovec et al. take advantage of submodularity and diminishing returns property - where the effect of a piece of information is inversely proportional to the previous exposure to similar pieces of information - and proposed Cost-Effective Lazy Forward selection (CELF) achieving an 700% improvement in execution time over the greedy algorithm. Additionally, Chen et al. propose Single Discount (SD) and Degree Discount (DD) heuristics which run more than six orders of magnitude faster than CELF and maintain a high influence spread, constituting them appropriate for big data analysis where time constraints are significant. Specifically, SD selects seed nodes based on their degree but rather than considering the initial degree of each node, it recomputes it in every time step, reducing it by 1 for every neighbor node already included in the seed set. DD is fine-grained towards IC model and exploits the influence probability of a seed node to a neighbor v, that is with probability, v will be influenced by u so there is no need to add v in the seed set.

Natural Strategies: Beyond the pure IC and LT generic seed strategies, there are also methods that have followed a different approach inspired by natural phenomena dealing either with the seed selection. For example, in a process simulating metal annealing, the authors optimize the spread function of the diffusion process both in terms of time and magnitude. To achieve that they leverage a search algorithm in the close neighborhood which aims to discover the seed node that maximizes influence spread. The difference, though, from the simple greedy strategy is that Simulated Annealing may probabilistically accept a worse solution escaping possible local optima. A different approach in leverages a mechanism from the field of Ferromagnetism, namely heat diffusion. In this manner, influential users are likened to heat sources which
diffuse their heat to their neighboring lower temperatures positions that represent currently unaffected users and the selection of the seed nodes is based on a greedy algorithm that maximizes the heat diffusion function. Additionally, this work is one of the first attempts to capture the diffusion of negative information too, due to the endogenous property of the proposed heat model. Last, Heat Diffusion introduces the notion of time in the amount a user affects its neighbors from which the Decaying Aggression Transfer healing mechanism of the minimization process of Chapter is inspired. Considering the above approaches, we conclude that the significance of seed selection is highlighted by the large and disparate variety of methods addressing it. These works, introduce features such as the optimization process and the negative information, that are useful in the modeling process of aggression propagation studied in this work.

**IC-exclusive Strategies**: Proceeding to the methods that are exclusive to the IC model, [16] propose PMIA optimization which exploits a tree structure called Maximum Influence Arborescence (MIA) and a prefix excluding strategy. In particular, it calculates MIA structure for every node u in the graph which contains all the edges through which seed set S can propagate influence to u. As a result they can find which nodes can influence u just by checking the local neighborhood of u. The prefix strategy preterns in the seed set only those nodes u whose influence is not blocked by some other node v, meaning that the influence of u is more direct than v’s. Last, IRIE is proposed by Jung et al. [37], which proposes a global node influence ranking computed in a small number of iterations and makes use of a tree graph influence propagation mechanism to achieve good estimates of the influence spread function. Furthermore, IRIE runs up to two orders of magnitude faster than PMIA through an influence estimation method. PMIA and IRIE achieve better influence spread than any other previous method, but for a large scale data analysis scenario SD and DD are still preferable due to better execution time and close, yet inferior, influence spread.

**LT-exclusive Strategies**: Turning to LT model, there are also exclusive strategies here. Chen et al. [18] provide a linear-time algorithm computing exact influence spread in directed acyclic graphs (DAGs). However, real social networks are not DAGs typically, making direct DAG application infeasible. Hence, they use a local DAG called LDAG for each node u to approximate its influence in the original network. However, LDAGs have to be carefully selected in order to cover a significant portion of the original network and for this purpose they use a Dijkstra shortest-path algorithm. Moreover, Goyal et al. [28] propose SIMPATH, an improvement of LDAG. It builds upon CELF but instead of complex Monte Carlo simulations it exploits simple paths - paths where no nodes are repeated - to compute spread. Specifically, it calculates simple paths in a small neighborhood due to LT’s property where probabilities of paths diminish rapidly as they get longer. However, despite that the above two algorithms achieve high influence spread
they do not provide theoretical approximation guarantees and present inferior execution times in comparison to DD and SD.

The discussion on the proper seed strategy selection depicts that the proposed approaches differ in various ways, from the diffusion model they apply to, to the optimization methodology followed. Table 3.2 summarizes this information to help the reader categorize the different research directions.

<table>
<thead>
<tr>
<th>Model Type</th>
<th>Optimization</th>
<th>Method</th>
</tr>
</thead>
<tbody>
<tr>
<td>Generic</td>
<td>CELF</td>
<td>CELF [40]</td>
</tr>
<tr>
<td></td>
<td>Degree</td>
<td>Single Discount, Degree Discount [17]</td>
</tr>
<tr>
<td>Natural</td>
<td>Probabilistic Greedy</td>
<td>Simulated Annealing [35]</td>
</tr>
<tr>
<td></td>
<td>Greedy</td>
<td>Heat Diffusion [45]</td>
</tr>
<tr>
<td>IC exclusive</td>
<td>MIA</td>
<td>PMIA [16]</td>
</tr>
<tr>
<td></td>
<td>Influence Ranking/Estimation</td>
<td>IRIE [37]</td>
</tr>
<tr>
<td>LT exclusive</td>
<td>DAG</td>
<td>LDAG [18]</td>
</tr>
<tr>
<td></td>
<td>CELF + simple paths</td>
<td>SIMPATH [28]</td>
</tr>
</tbody>
</table>

Table 3.2: Method classification for the seed selection process

3.3 Competitive Cascades

In both information diffusion and influence maximization, there have been multiple mentions on the presence of a second, sometimes negative, cascade alongside the first one. The investigation of information diffusion under these circumstances falls under the category of Competitive Cascades. Depending on the goal of the competition, competitive cascades can be exploited either in influence maximization [8] or in influence minimization [10]. Specifically, the two processes depict that the aim of the investigation is to maximize the spread of one cascade over the presence of the other or to minimize its spread through appropriate manipulations on the second cascade respectively.

3.3.1 Competitive Influence Maximization

Competitive IC models: In particular, in [8] the authors investigate information diffusion under IC when multiple competing cascades are present, through a game theory lens. They provide a greedy algorithm where the node with the largest marginal gain is added in the activated node set achieving a factor \((1 - 1/e)\) of the best response. They also consider some first mover strategies in a duopoly scenario - which is the strategy of the first player when two cascades are present and competing - and by utilizing dynamic
programming alongside a tree structure they propose an improvement of the greedy algorithm. Chen et al. extend competitive IC model in [14] by adding a quality factor representing the natural tendency of people turning negative to a product due to its deficits. They also make a reference to the minimization of a negative information by tweaking the spread function of their proposed algorithm by incorporating both positive and negative spread alongside MIA optimization algorithm for performance boost. However, in this case, submodularity and monotonicity are achieved only for specific values of the quality factor.

**Competitive LT models**: On the contrary, Borodin et al. [9] turn to the case of competitive linear threshold models where they show that submodularity fails to hold for some of the proposed maximization functions and thus a greedy algorithm of the bounded approximation guarantees is not possible. For this reason they study the separated threshold model and the competitive threshold model with forcing. In the former each individual becomes active based on separate criteria for each piece of information while in the latter each inactive node chooses randomly. Despite all these modifications, submodularity can not be achieved. Thus they propose the OR model where they exploit two probability functions satisfying monotonicity and submodularity and as a result preserving these properties in the original problem too.

### 3.3.2 Influence Blocking Maximization / Influence Minimization

In competitive cascades, apart from influence maximization, there are considerable efforts in proposing models where the objective is to find a positive seed set that can minimize the spread of a negative cascade. This problem is called *Influence Blocking Maximization*. Specifically, He et al. [31] propose a greedy algorithm under the competitive LT scenario but due to the time-consuming Monte Carlo simulations they also extend their efforts to a CLDAG algorithm, the competitive cascade analogy of the LDAG optimization discussed in the corresponding section of this work. By carefully selecting the LDAG during the process and by utilizing a dynamic programming algorithm they show that CLDAG achieves the best performance among all tested heuristics especially when negative influence is strong.

Furthermore, Zhang et al. [62] address misinformation containment - different title for the same problem - on LT model by proposing a greedy algorithm for identifying the most influential nodes and another one that uses the CELF optimization to find the best seed candidates for the positive cascade. They introduce **preference** - a metric defined by the node thresholds and the edge weights of each node’s in-neighbors - which enables nodes to make different decisions upon receiving same information. Still, the seed
selection process introduces a trade-off between limiting the spread of misinformation or expanding the effect of positive information and hence it does not focus purely on the minimization aspect.

Budak et al. [10] approach the same problem of influence minimization under a different title, namely *Eventual Influence Limitation* and investigate two sub-problems called *Multi-Campaign Independent Cascade Model (MCICM)* and *Campaign-Oblivious Independent Cascade Model (COICM)* respectively. In MCICM, the positive information is spread with probability $= 1$ reflecting the fact that good information is diffused by authorities, users with high validity - also called *high effectiveness* property - while misinformation by less influential users and hence the negative spread is usually much smaller than that of the good information. On the other hand, COICM addresses any two diffusion processes in a similar manner by setting their propagation probability to be equal. However, in the case of simultaneous influence by the two cascades, the positive one dominates. Moreover, except for the experimentation on the usual greedy approach, they also propose a *predictive hill climbing algorithm* (PHCA) to tackle the case of missing information. They do so by predicting the current state of all nodes in the network using only a small known fraction of nodes. Howbeit, a limiting factor of their work is the hypothesis that the negative campaign’s spread starts from a single node rather than a set of nodes and it is detected after a delay $r$.

More recently, Wu et al. [61] address the above MCICM and COICM problems by investigating influence blocking minimization under competitive IC model. Based on the MIA optimization of the simple IC model they propose their competitive derivatives *CMIA-H* and *CMIA-O* respectively. The use of the above optimization structure enables them to outperform the greedy algorithm of [10]. This is possible due to the high effectiveness property in the case of CMIA-H and a dynamic programming method to efficiently compute the negative activation probability of any node in CMIA-O case. Additionally, they overcome the limitation of the single starting node for the negative cascade. Notwithstanding, due to the lack of submodularity these methods lack approximation guarantee.

### 3.4 Immunization

The previous section discusses influence minimization under competitive cascades. However, competitive cascades presuppose that the information diffusion process can be sufficiently modeled by a cascade model. Social media nowadays prefer to address abusive behavior restriction by blocking abusive users or posts. This preference is grounded
on either the simpler nature of those algorithms or because of their more straightforward application. In the current work, these methods are utilized in two ways: (a) If cascading diffusion methods can indeed model aggression diffusion - thus competitive cascade minimization technique is applicable - immunization techniques are leveraged as a comparison baseline and (b) if cascades prove to be insufficient, then an examination of whether immunization algorithms can be fine-grained towards aggression specific minimization rather than the broader influence or epidemic spread minimization is attempted.

State-of-the-art methods in the domain of immunization contain algorithms based on linear algebra manipulation to exploit interesting users. Specifically, for a large family of diffusion processes, Prakash et al. [49] find that the only network parameter that determines whether this diffusion would become an epidemic or not is the largest eigenvalue $\lambda$ of the adjacency matrix $A$. Tong et al. propose NetShield algorithm [55] which exploits this property to introduce a ranking score called Shield score and to detect the most important abusive users. Furthermore, in [51] the authors take advantage of the same property to propose NetMelt, an algorithm that transfers the problem to edge deletion, stating that oftentimes node deletion may be more radical than necessary. Notwithstanding, these techniques detect abusive behavior in an offline manner, meaning that they preprocess the adjacency matrix to expose the important elements, whether they are nodes or edges. Our proposed methodology is online and adjusts to the specifics of the diffusion process during its execution.

<table>
<thead>
<tr>
<th>Method Classification</th>
<th>Minimization Approach</th>
<th>Optimization strategy</th>
<th>Method</th>
</tr>
</thead>
<tbody>
<tr>
<td>IC Competitive Cascades</td>
<td>Greedy</td>
<td>First Mover [8], Greedy + PHCA [10]</td>
<td></td>
</tr>
<tr>
<td></td>
<td>MIA</td>
<td>MIA-N [14], CMIA-H and CMIA-O [61]</td>
<td></td>
</tr>
<tr>
<td>LT Competitive Cascades</td>
<td>Greedy</td>
<td>Separated-Threshold, Forcing and OR models [9]</td>
<td></td>
</tr>
<tr>
<td></td>
<td>LDAG, CELF</td>
<td>CLDAG [31]</td>
<td></td>
</tr>
<tr>
<td>Node Blocking</td>
<td>Largest Eigenvalue</td>
<td>NetShield [55]</td>
<td></td>
</tr>
<tr>
<td>Edge Blocking</td>
<td>Largest Eigenvalue</td>
<td>NetMelt [54]</td>
<td></td>
</tr>
</tbody>
</table>

Table 3.3: Method classification for minimization techniques

The discussed methods from the domains of Competitive Influence Maximization, Influence Blocking Maximization and Immunization with their corresponding interesting features are also presented in Table 3.3.
3.5 Aggression Minimization

Having analyzed all the related works on minimizing the spread of negative information, the scope should now be limited to aggression minimization itself. To the best of the author’s knowledge, this problem has yet to be studied by the research community. There is only one work by Kourtellis et al. - currently under submission and peer review process - which attempts to model the aggression diffusion process. However, there are two differences between this work and ours. First of all, they model aggression propagation by leveraging models of the Opinion Dynamics domain such as Deffuant, Friedkin & Johnsen (FJ) and Hegselmann & Krause (HK) models [21, 26, 32], which simulate the case of symmetric information diffusion where a node gets affected by the same piece of information multiple times. Considering the scenario of aggression propagation in social networks where a post spreads the aggression, progressive models are more suitable than symmetric ones to describe aggression modeling. The second point of deviation is that the specific thesis is the first to address aggression minimization. Having succeeded in discovering a model and a set of parameters that best describe aggression diffusion, the effort continues one step further and introduces methods to limit its negative effect on nodes. Actually, two different methods of minimizing aggression are investigated using (a) competitive cascades and (b) node/edge blocking respectively. The second method is what social media such as Twitter use nowadays, but in Chapter 6 is shown that competitive cascades stand for a more efficient solution to the problem.

Conclusion: Aggression diffusion modeling and minimization depend on multiple domains, from the general information diffusion and influence maximization models to the maximization/minimization techniques of competitive cascades and blocking or immunization methods. Despite the various domains involved, to the best of our knowledge there are no prior established methods addressing these problems directly. This is the gap that this thesis fills.

With all the existing literature reviewed, the following chapters present the analysis of the proposed methodology both in a macroscopic-theoretical level as well as in a microscopic-implementation one where the specifics of each algorithm will be provided. Chapter 4 presents the various diffusion models, seed strategies, activation and threshold criteria as well as aggression transfer mechanisms and provides the corresponding algorithmic implementations in pseudo code. Chapter 5 is related to aggression minimization where the two problems under consideration - competitive aggression minimization and blocking aggression minimization - are defined and the healing strategies are specified. Also, similar to Chapter 4 it provides the algorithmic implementation discussed.
Chapter 4

Aggression Diffusion Modeling

Chapter 4 is dedicated to the analysis of the general methodology, the components that compose the proposed algorithms as well as the description of the specifics and implementation of each component via provided pseudo code with regard to the first problem that the thesis address, aggression diffusion modeling. Every section is comprised of two parts. First, a shallow skimming of the components of each method is pursued where the general ideas are discussed, leaving the implementation details for the second part. There, for every component and implemented method, pseudo code is provided in order to explain the algorithm and the underlying details.

The initial goal of this work is to examine if there is an information diffusion model laying in the category of cascade (IC or LT) models - methods mostly used in the domain

![Figure 4.1: Overall aggression modeling process](#)
The methodology unfolds by initially discovering the best seed selection strategy, moving to the weighting schemes and ending with the model specific components, which are the activation criteria and the aggression transferring mechanisms. These steps do not need to be executed in this order, however this pipeline enables the process to move from the most generic to the most specific configuration parameter.

### 4.1 Seed Selection

![Seed Selection diagram](image)

**Figure 4.2: Seed Selection component**

Under aggression diffusion modeling scenario the seed selection process is the first crucial component to describe. The term seed refers to these nodes from which the diffusion process will kick off. Thus, the ultimate result of the propagation model is tightly related to these initial seed nodes.
4.1.1 Theoretical Analysis

The necessity of the seed selection process arises from the fact that the overall diffusion model is restricted by a given budget k, meaning that only a specific amount of initial stimulation or seed nodes can be afforded. As a result, a sophisticated seed selection strategy could enable the diffusion process and lead to the increase of the number of activated nodes. The proposed alternative strategies are shown in figure 4.2 and listed below:

- **All Aggressive:** If budget $k \geq |\text{aggressive users}|$ choose as seeds all of the aggressive users. This method is straightforward and aims at capturing the aggression diffusion by directly addressing the source of the phenomenon, that is the most aggressive users.

- **Top Aggressive:** If budget $k < |\text{aggressive users}|$ choose top k users based on aggression score. Considering the limited budget that is available, the seeding strategy should also be adjusted accordingly. Top aggressive method aims at capturing the most effective nodes with regard to the above considerations.

- **Degree Discount and Single Discount:** DD and SD algorithms were proposed by Chen in [17]. Degree discount heuristic is slightly inferior to the greedy algorithm, however it runs several orders of magnitudes faster. Nevertheless, it is fine-tuned for the IC model. That is why Single Discount, a more generic approach is also put to the test.

- **Random:** Choose seeds randomly. This method is considered as a baseline.

- **Greedy:** If budget $k < |\text{aggressive users}|$ choose from aggressive users those k that greedily maximize the marginal gain with respect to spread function’s definition. It has been shown [38] that this kind of greedy algorithms outperform methods based on network properties with regard to the number of activated nodes at the end of the propagation process. However, this efficiency comes at the cost of bigger and sometimes even infeasible execution time. The results shown that the execution time of this method is prohibiting and thus it will not be discussed further.

4.1.2 Specifications

Continuing with the specifics of each seed strategy, the first method to be presented is "Random" seed strategy. Algorithm 1 provides the pseudo code for this. This method is used as a baseline and it unfolds iteratively by picking a node randomly and adding it to the seed set until the required size has been reached. Lines 2-6 show this process.
Algorithm 1: Random seed strategy

Data: graph G, seed size k

Result: The seed set \( S \)

1. \( i \leftarrow 0 \)
2. while \( i < k \) do
3.     \( \text{choice} \leftarrow \text{random(}\text{graph nodes}\text{)} \)
4.     \( S \leftarrow S \cup \{\text{choice}\} \)
5.     \( i \leftarrow i + 1 \)
6. end
7. return \( S \)

Next, the aggression related methods are discussed. Here, Algorithms 2 and 3 present the all-aggressive and top-aggressive strategies correspondingly. "All Aggressive" is not based on a specific seed size but instead on an initial aggression threshold which needs to match the results of the prediction algorithm used to create the ground truth. The specifics of these values are given in the corresponding section of Chapter 6. Next, the algorithm fills the seed set by selecting the users with an aggression score bigger or equal to the given threshold (lines 3-5 of Algorithm 2). On the other hand, "Top Aggressive" fills the gap of providing a seed size smaller than that of "All Aggressive", in other words a seed size which corresponds to a threshold smaller than that provided by "All Aggressive". Hence, a necessity to make a sophisticated selection of aggressive users to include in the seed set emerges. For this purpose, it sorts all users in descending order from the most to the least aggressive and picks the top \( k \), shown in lines 7-11 of Algorithm 3. For completeness purposes, it is noted that in case of a seed size \( \geq \) threshold provided by the "All Aggressive" algorithm, these two methods behave identically.
Algorithm 2: All-aggressive seed strategy

Data: graph G, threshold

Result: The seed set S

1. aggression threshold ← threshold
2. for node ∈ graph nodes do
3.     if node/aggression score/ ≥ aggression threshold then
4.         S ← S ∪ {choice}
5.     end
6. end
7. return S

Algorithm 3: Top-aggressive seed strategy

Data: graph G, seed size k, threshold

Result: The seed set S

1. aggression_threshold = get_threshold(k)
2. if aggression_threshold ≥ threshold then
3.     S ← Algorithm 2
4. else
5.     sorted_nodes ← sort(graphnodes)
6.     i = 0
7.     while i < k do
8.         choice ← sorted_nodes[i]
9.         S ← S ∪ {choice}
10.        i ← i + 1
11.     end
12. end
13. return S

After the aggression related methods, the implementation of single discount and degree discount follows, presented in algorithms 4 and 5. These methods exploit the out-degree of the graph nodes. The first one - "Single Discount" - is worse than "Degree Discount" in terms of efficiency as it achieves a complexity of $O(kNn)$, in contrast to $O(k \ast \log(n) + m)$ \cite{17}, but yet it generalizes to both IC and LT while the "Degree Discount" is applicable only to IC. It uses a priority queue where nodes are added and the one with the higher degree is prioritized. Until the specified seed size has been reached nodes are popped from the priority queue and added to the seed set (lines 7 and 8) while at the same time the degrees of the node’s neighbors are updated (line 12). This update has a magnitude of 1 in case of an unweighted graph or equal to the corresponding edge weight when a weighted graph is used.
Algorithm 4: Single Discount seed strategy

Data: graph $G$, seed size $k$

Result: The seed set $S$

1. $q \leftarrow$ priority queue with respect to the out-degree
2. for node $\in$ graph nodes do
3.   $q \leftarrow q \cup \{\text{node}\}$
4. end
5. $i \leftarrow 0$
6. while $i < k$ do
7.   $u \leftarrow q$.pop()
8.   $S \leftarrow S \cup \{\text{node}\}$
9.   $N_u \leftarrow \text{neighbors of } u$
10. for $v \in N_u$ do
11.     if $v \notin S$ then
12.         $v\_degree \leftarrow v\_degree - 1$
13.     end
14. end
15. $i \leftarrow i + 1$
16. end
17. return $S$

Algorithm 5: Degree Discount seed strategy

Data: graph $G$, seed size $k$

Result: The seed set $S$

1. $q \leftarrow$ priority queue with respect to the out-degree
2. for node $\in$ graph nodes do
3.   $q \leftarrow q \cup \{\text{node}\}$
4.   $\text{trace}[\text{node}] \leftarrow 0$
5. end
6. $i \leftarrow 0$
7. while $i < k$ do
8.   $u \leftarrow q$.pop()
9.   $S \leftarrow S \cup \{\text{node}\}$
10. $N_u \leftarrow \text{neighbors of } u$
11. for $v \in N_u$ do
12.     if $v \notin S$ then
13.         $p \leftarrow (u, v)$ edge weight
14.         $\text{trace}[v] \leftarrow \text{trace}[v] + 1$
15.         $v\_priority \leftarrow v\_degree - 2 \times \text{trace}[v] - (v\_degree - \text{trace}[v]) \times \text{trace}[v] \times p$
16.     end
17. end
18. $i \leftarrow i + 1$
19. end
20. return $S$

On the contrary Degree Discount leverages the fact that in case of a seed node its neighbor $u$ can be neglected with probability at least equal to their corresponding edge weight, as the latter will possibly get influenced by the former. We once again point out that this observation is only applicable to IC model. This is shown in line 15 of Algorithm 5.

Moving forward, the thesis proceeds to the mechanisms related with the initial step of the overall diffusion process, the weighting scheme of the graph.
4.2 Weighting Schemes

Figure 4.3: Weighting Scheme component

Having defined the different seed selection strategies and mechanisms the next component to discuss is the weighting scheme selection method. This parameter, the weighting scheme expresses the meaning and intensity of the formed links between users in the social network. Thus, being able to distinguish a proper scheme is of major importance to formulate a valid network context upon which the diffusion process will unfold.

4.2.1 Theoretical Analysis

Each one of the proposed weighting scheme captures real world phenomena (neighborhood similarity, user credibility etc) that tend to relate to the diffusion of aggression. Thus, their implementation in the diffusion process should enable the overall aggression propagation in the network as well as the modeling methodology. Below, the various schemes are described while figure 4.3 illustrates the corresponding component.

- **Jaccard Overlap**: Given two nodes \( u, v \) and their corresponding sets of neighbors \( N_u \) and \( N_v \), the Jaccard Overlap is defined as \( w_{uv} = \frac{N_u \cap N_v}{N_u \cup N_v} \in [0,1] \). In social media, the friends and followers of a user delineate and outline his own beliefs. If two users appear to have similar or even identical social circles this could mean that they share the same beliefs. Jaccard overlap emulates exactly this observation of neighborhood similarity, also discussed in [33].
• **Power Score Ratio:** Given a node $u$, the power score $P_u$ is defined as the ratio of the in-degree over out-degree, $P_u = \frac{\text{inDegree}_u}{\text{outDegree}_u} \in [0, 1]$. The bigger the power score the more dominant the influence that a node receives from its in-neighbors in comparison to the degree that it influences its out-neighbors. The weighting magnitude should be pairwise though. Therefore, given an edge from node $u$ to node $v$ we use $P_{uv} = \frac{P_v}{P_u} \in [0, 1]$. The intuition emerges from the definition of power score. The bigger the power score of $v$, the bigger the influence from its in-neighbors. Accordingly, the smaller the power score of $u$, the bigger its influence to its out-neighbors. In this way, the credibility of a user is taken into account.

• **Weighted Overlap:** Given two nodes $u,v$, their Jaccard overlap $w_{uv}$ and their power score ratio $P_{uv}$, weighted overlap is defined as $P_{wuv} = P_{uv} \times w_{uv} \in [0, 1]$. This metric considers both of the already proposed magnitudes and combines them in order to track their simultaneous effect on the aggression diffusion process. As each one of the separate properties consist part of the behavior of an abusive user, it is expected that their combination will do so too.

4.2.2 Specifications

Proceeding to the specifics and implementation of weighting scheme selection strategies - as it has been already mentioned in Section 4.2 - three weighting schemes are discerned, namely "Jaccard Overlap", "Power Score" and "Weighted Overlap". Below, the pseudo code related with these schemes is provided and discussed:

Specifically, in Algorithm 6, Jaccard Overlap is set as the edge weight according to the neighborhoods of its source and destination nodes. Lines 2-3 show the selection of the corresponding neighborhoods, while line 4 expresses the total weight as the fraction of their intersection divided by their union.

**Algorithm 6:** Jaccard Overlap weighting scheme  

**Data:** An unweighted directed graph  

**Result:** The weighted directed graph according to Jaccard Overlap  

1. for edge $\in$ graph edges do  
   2. $N_s \leftarrow$ set of neighbors of source node;  
   3. $N_d \leftarrow$ set of neighbors of dest node;  
   4. edge weight $\leftarrow (N_s \cap N_d)/(N_s \cup N_d)$  
5. end
Additionally, in Algorithm 7 which presents the power score based weighting scheme, first of all, individual node power scores are calculated according to node’s in and out-degrees (line 2). Then, the computation of edge’s power score takes place as the ratio of destination over source power score (line 5).

Algorithm 7: Power Score weighting scheme

Data: An unweighted directed graph

Result: The weighted directed graph according to Power Score

1 for node ∈ graph nodes do
2 \[ \text{power\_score}[\text{node}] \leftarrow \text{indegree}/\text{outdegree} \]
3 end

4 for edge in graph edges do
5 \[ \text{edge weight} \leftarrow \text{power\_score}[\text{dest}]/\text{power\_score}[\text{source}] \]
6 end

Last, in Algorithm 8 the previous techniques are combined to derive a single score, called "Weighted Overlap". This score is calculated as the multiplication of edge Jaccard overlap by the corresponding power score as shown in lines 5-9. Thus, all of these techniques encapsulate the notions of neighborhood and credibility of a node with respect to its neighbors.

Algorithm 8: Weighted Overlap weighting scheme

Data: An unweighted directed graph

Result: The weighted directed graph according to Weighted Overlap

1 for node ∈ graph nodes do
2 \[ \text{power\_score}[\text{node}] = \text{indegree}/\text{outdegree} \]
3 end

4 for edge in graph edges do
5 \[ \text{N}_s \leftarrow \text{set of neighbors of source node}; \]
6 \[ \text{N}_d \leftarrow \text{set of neighbors of dest node}; \]
7 \[ \text{jaccard\_overlap} \leftarrow (\text{N}_s \cap \text{N}_d)/(\text{N}_s \cup \text{N}_d) \]
8 \[ \text{power\_score} \leftarrow \text{power\_score}[\text{dest}]/\text{power\_score}[\text{source}] \]
9 \[ \text{edge weight} \leftarrow \text{jaccard\_overlap} \times \text{power\_score} \]
10 end

Continuing, given a constant budget k, two distinct cases - referring to IC and LT diffusion models respectively - are described.
4.3 Diffusion Model (Activation Criteria and Aggression Transferring)

After describing all the necessary seed selection strategies and weighting schemes we are now ready to continue to the actual diffusion process. For this purpose there are two possible alternatives, the IC and LT models. Each one is governed by specific rules regarding the diffusion process as well as the aggression transferring mechanisms.

4.3.1 Theoretical Analysis

Considering the different mechanics of IC-based and LT-based models, it is necessary to present the specifics of each one separately. In this way, the comparison of two dissimilar models is avoided but instead the important points of each model are discussed.

4.3.1.1 Independent Cascade

We are now ready to describe the propagation of aggression under the IC model. At time step $t = 0$ the process starts with an initial seed set $S$ of active users with $|S| \leq k$ and proceeds in discrete time steps according to the following stochastic rule. Each active node $u$, that got activated in time step $t$, has a single chance to activate each one of its inactive neighbors $v$ with probability $p$ which obeys some specific rules defined below. We consider the case where node $u$ tries to activate node $v$ and we define the possible alternatives for the activation probability $p$, which are also illustrated in figure 4.4.

- **Jaccard Overlap**: The bigger the Jaccard overlap, the more similar their neighborhoods are and thus the more related the two nodes are. As a result it is easier for $u$ to activate $v$. Hence, $v$ gets activated with probability $p \leq w_{uv}$.
• **Power Score Ratio**: Again the bigger the power score ratio, as defined previously in section 4.2, the easier for node $u$ to activate its neighbor $v$. Thus, $p \leq P_{uv}$.

• **Weighted Overlap**: Given the above two observations activation probability could be subject to both Jaccard Overlap and Power Score Ratio. Thus, we make use of the weighted overlap for which $p \leq P_{wuv}$, in order to capture the potential simultaneous effect of the aforementioned notions.

Given that a node succeeds in activating one of its neighbors, under the aggressive IC model that means that $u$ transfers its aggression score to neighbor $v$, that is $A_v = A_u$. However, for the diffusion process itself a node is considered either active or inactive in a binary fashion, similar to the traditional IC model, and thus the aggression transferring mechanism does not complicate the overall process any further. If multiple nodes try to activate the same node simultaneously and succeed, the decision of whose aggression score will be transferred follows one of the below strategies:

• **Top**: Keep the activation of the most aggressive node. This strategy should be biased towards a bigger spread of aggression both in terms of activated nodes and the overall network’s aggression score increase.

• **Random**: The activation order is arbitrary the aggression score of the last node is the one that finally dominates over the rest.

• **Cumulative**: Keep the cumulative aggression score of all nodes according to their contribution. If $S = \{u_1, u_2, u_3\}$ the set of nodes that succeeded then $A_v = \sum_{i \in S} w_i \cdot A_i$ with $w_i = \frac{A_i}{\sum_{j \in S} A_j}$ and $A_i$ their respective aggression scores. This strategy arises from the community behavior that aggressive users present according to [7].

The above alternatives are also shown schematically in figure 4.5. If $u$ succeeds, then node $v$ gets activated in step $t+1$; but regardless of $u$’s success, it cannot make any further attempts to activate $v$ in a following step. The process terminates when there are not any new activations.

### 4.3.1.2 Linear Threshold

Apart from the IC model, the case of LT model is also considered due to its inherent property to better capture the neighborhood effect on a specific node. According to this, each node $v$ is assigned a threshold $\theta_v$. For the aggression modeling scenario, two alternatives are presented:
• Thresholds are equivalent to the aggression scores of the nodes. That means that 
  \( \theta_v = A_v \forall v \in V \).

• Thresholds are equivalent to the power scores of the nodes. Similar to the previous case, 
  \( \theta_v = P_v \forall v \in V \).

The process starts with an initially activated seed set \( S \) with \( |S| \leq k \), analogously to the IC process. In step \( t \), a node \( v \) is influenced by each activated neighbor \( u \in N_v \) according to a weight which should respect the node threshold selection. Specifically, the activation criteria at time step \( t \) are:

- **Aggression:** When aggression scores are used as node thresholds: 
  \[ \theta_v \leq \sum_{u \in N_v \cap \text{Active}^t} A_u \]
  where \( \text{Active}^t \) is the set of active nodes at time step \( t \).

- **Power:** When power scores are used as node thresholds: 
  \[ \theta_v \leq \sum_{u \in N_v \cap \text{Active}^t} P_u \]

The diffusion process unfolds until there are no new activations in a specific time step. Finally, under the aggressive Linear Threshold model, activation means that the neighbors of node \( v \) transfer their average aggression score to \( v \), that is 
  \[ A_v = \frac{\sum_{u \in N_v} A_u}{|N_v|} \in [0, 1] \]

4.3.2 Specifications

The last component of aggression modeling, the diffusion model itself, should also be described in a microscopic level. Thus in this section, the algorithmic process for the two tested models, IC and LT is presented.

The first model to discuss - in Algorithm 9 - is aggressive IC. This model is similar to the original IC model with the difference that there is a need to keep track of the activators of each node, as shown in line 9. This step is necessary in order to transfer aggression properly in step 17 according to the criteria discussed in Section 4.3.1.1. An important observation here is the fact that the process evolves as long as there are new activations, shown in line 3. This fact expresses the single chance that each node has to activate a specific neighbor.

Next, in algorithm 10 the presentation of the second model, namely aggressive LT, takes place. Once again, the algorithm is similar to the original LT model and close to the aggressive IC presented above. The main difference is that in LT we also have
Algorithm 9: Aggressive Independent Cascade model

**Data:** graph $G$, seed set $S$

**Result:** The activated node set $S$

1. $new\_activations \leftarrow S$
2. activators $\leftarrow$ set of activators for each node
3. **while** $\exists new\_activations$ **do**
   4. **for** $u \in new\_activations$ **do**
      5. **for** $v \in N_u$ **do**
         6. **if** $v \not\in S$ **then**
            7. $p \leftarrow$ random probability
            8. **if** $p > (u,v)$-edge weight **then**
               9. activators$[v] \leftarrow$ activators$[v] \cup \{u\}$
               10. $new\_activations \leftarrow new\_activations \cup \{v\}$
            11. **end**
         12. **end**
      13. $new\_activations \leftarrow new\_activations \setminus \{u\}$
   14. **end**
   15. $S \leftarrow S \cup new\_activations$
   16. **transfer aggression according to activation criterion**
17. **end**
18. **return** $S$

To track the influence of each one of the activators on the corresponding node, as its threshold may not be surpassed on the specific time step. This is what lines 8-11 stand for. Hence, in contrast to IC model and its single-chance activation logic, here a node gets activated when the cumulative influence of its activated neighbors is larger than its preset threshold (line 9).
Algorithm 10: Aggressive Linear Threshold model

Data: graph G, seed set S

Result: The activated node set S

1. new_activations ← S
2. activators ← set of activators for each node
3. while ∃ new_activations do
   4. for u ∈ new_activations do
      5. for v ∈ N_u do
         6. if v /∈ S then
            7. activators[v] ← activators[v] ∪ {u}
            8. node_influences[v] ← node_influences[v] + (u, v) − edgeweight
            9. if node_influences[v] > thresholds[v] then
               10. new_activations ← new_activations ∪ {v}
      11. end
   12. end
   13. new_activations ← new_activations \ {u}
14. end
15. end
16. S ← S ∪ new_activations
17. return S

Conclusion: In this chapter, all the necessary information regarding each separate component of aggression modeling has been given. Each distinct combination of those components alongside the choice of the underlying diffusion model consists a well defined configuration. These configurations were examined thoroughly to understand the nature and the insights of aggression diffusion over a given social network. The results of these experiments will be discussed in the corresponding chapter of this work. The schematic representation of the overall aggression modeling process given in figure 4.1 shows that indeed IC and LT models can capture aggression diffusion through a well defined pipeline, fulfilling Contribution 1 presented in Chapter 1 and in extension to address Contribution 2 by finding the best configuration. Specifically for Contribution 2, Chapter 6 provides the exact configurations that achieve the best results.
Chapter 5

Aggression Minimization

In this chapter, the problem of aggression minimization is addressed by means of two different approaches called competitive and blocking aggression minimization respectively. Competitive cascades approach is possible due to the fact that aggression can indeed be modeled by a cascading model, as results of Chapter 6 depict, while blocking minimization is a common approach of many social media platforms like Twitter, in order to restrain improper and abusive user behavior, but not necessarily related to aggression. However, for this problem an extension of the state-of-the-art algorithms is also provided in order to investigate their effect on aggression minimization specifically.

**Figure 5.1:** Aggression minimization process

In order to better understand the aggression minimization process, Figure 5.1 presents the methodology followed in order to investigate the most suitable model for minimizing the overall aggression score of the network. There are two main approaches, as already discussed, minimization via competitive cascades or immunization (blocking) techniques.
Under competitive cascades, there are two crucial factors: (a) the cascade of the positive or educational piece of information and (b) the healing strategy when the positive cascade reaches and activates a node. Last, under blocking minimization, the decision to use either a node or an edge blocking approach should be taken into consideration. Both approaches will be addressed in more detail during the following sections of this chapter.

5.1 Competitive Aggression Minimization

Focusing on Competitive Aggression Minimization (CAM), two competing diffusion processes can be observed, a negative one - also called aggressive - and a positive or educational cascade whose goal is to minimize the spread of the aggressive. It should be noted here that had the cascade processes discussed in Chapter 4 not been able to model the diffusion of aggression, competitive cascades would not have been a valid aggression minimization technique. Hence, Blocking Aggression Minimization presented in Section 5.2 would have been a one-way route. However the results of Chapter 6 are positive and under the context of CAM, the specific models as well as their corresponding parameter configurations for both of the competing cascades should be presented.

5.1.1 Theoretical Analysis

Diving deeper into the specifics of the negative cascade and hence the aggression diffusion, the model with the best performance during the experimental phase of aggression modeling is chosen to be leveraged. In a similar manner, for the educational process the same cascading model is used and the various alternative parameter configurations are put into test. In this way, the comparison of negative and positive cascades is equitable without favoring one over the other. Below we define the problem in a more formal way.

Problem 1 (CAM). Given a directed network $G = (V,E)$, with $V$ and $E$ denoting the node and edge sets respectively, an integer $k$ and two competing diffusion processes namely, aggressive and educational with the later having a set of parameters $\theta(\cdot)$, find a positive seed set $S$ with $|S| \leq k$ and a $\theta^*(\cdot)$ s.t. $\theta^*(\cdot) = \arg \min_\theta A(G)$ w.r.t. $S$ where $A(G) = \sum_{v \in V} A_v$ is the overall aggression score of the network.

We now have to define the rules that describe the educational cascade and the diffusion process. Under IC the activation probability strategy has to be clarified. We should take in mind that in contrast to the negative cascade, here user aggression score is a limiting factor and hence we can exploit it to better capture the properties of the cascade. Specifically, we could use:
• Any of the previous Jaccard overlap, power score or weighted overlap scores, analogously to the aggression modeling scenario

• A strategy based on the fact that the more aggressive a user is the more difficult to sensitize him. Thus, if node u tries to activate v then $p \geq A_v \forall v \in V$.

However, under LT, the activation criterion as well as node thresholds should depend on power scores only. Aggression scores would be inappropriate as thresholds because the educational cascade intents to mitigate or even reduce the effect of aggression on the network instead of maximize its overall magnitude. As a result, using aggression scores as node thresholds and activation criterion would undermine our ultimate goal as they would enable the flow of aggression over the network.

Consequently we describe the competitive process. Using a predefined seed set for the negative cascade and calculating a positive seed set of the same size, according to the best strategy of aggression modeling scenario, let the two processes, negative and positive, unfold simultaneously. If at time step t both cascades reach the same node v, then v gets positively activated simulating the fact that most probably an educated or aware user would stop manifesting aggressive behavior. The competitive process is completed when there are no nodes left to activate for any of the two cascades.

Proceeding to the meaning of activation from the perspective of the positive cascade, we discern four possible cases:

• **Vaccination**: activating user v results to $A'_v \rightarrow 0$. This hypothesis is very strict.

• **Aggression transfer**: activating user v results to:
  
  \begin{itemize}
  \item IC: $A'_v \rightarrow A_u$ with u being the user that activates v
  \item LT: $A'_v \rightarrow \frac{\sum_{u \in N_v} A_u}{|N_u|}$
  \end{itemize}

• **Decaying aggression transfer**: activating user v is affected by a decaying factor $\lambda$ capturing the distance from the source of the information. Hence, $\lambda = \frac{1}{\text{hops}}$ and:
  
  \begin{itemize}
  \item IC: $A'_v \rightarrow \lambda \times A_u$ with u being the user that activates v
  \item LT: $A'_v \rightarrow \frac{\sum_{u \in N_v} \lambda \times A_u}{|N_u|}$
  \end{itemize}

This mechanism is based on the effect that time has on the influence maximization process according to [45].

• **Hybrid**: using a combination of 1. and 3.:

\[
A'_v = \begin{cases} 
\text{Vaccination} & \text{if } p \geq A_v \\
\text{Decaying aggression transfer} & \text{otherwise}
\end{cases}
\]
This means that the lower the aggression score of a node the easier to immunize it.

Finally, we should define the possible seed selection strategies. It is clarified here that each seed node $s$ will stop being aggressive with probability equal to its current aggression score. That is:

$$A'_s = \begin{cases} 
0 & \text{if } p \geq A_s \\
A_s & \text{otherwise} 
\end{cases}$$

This function simulates the effect of the hybrid approach introduced above but with no option of using decaying transfer as long as the seed nodes are the source of the educational cascade.

Proceeding to the positive seed set selection strategy itself, we distinguish the below possible cases:

- Use the same strategy as the negative cascade
- Use one of the strategies proposed in aggression modeling section
- Choose seeds randomly

This whole methodology acts as the tool to address Contribution 3 presented in the Introduction, if a cascade model is indeed appropriate to capture aggression propagation. However, if Contributions 1 and 2 are not fulfilled we can also address Contribution 3 by approaching and solving the Blocking Aggression Minimization problem presented below.

### 5.1.2 Specifications

The methods addressing CAM are quite similar to the aggressive IC and LT models presented before in Chapter 4. In fact they are the same models applied twice, once for the positive and once for the negative cascade. However, for the negative cascade we use the configuration that achieved the best performance during the experimental phase.

In algorithm 11, the competitive IC model used in aggression minimization is presented, while algorithm 12 pertains to competitive LT. In lines 5, 24 and 25 of both algorithms the healing step takes place, first for the initial seeds and then for every new activation of the positive cascade. Additionally, line 8 of both algorithms essentially shows that if there are new negative activations a step of the aggressive IC, and LT respectively, algorithms should be executed according to algorithms 9 and 10 addressed in Chapter 4. Moreover lines 10-26 of both algorithms are almost identical to aggressive IC
and LT algorithms with the only difference that instead of the aggression transferring, a healing step takes place - depicted in lines 24 and 25 - considering that they both refer to the positive cascades.

The algorithmic process makes clear that a node can get activated by both negative and positive cascades. Thus, the influence of one cascade does not make the node unapproachable. As a result, the outcome of the process depends solely on the healing strategy and the parameters discussed in Chapter 4 for both the negative and the positive cascade. This observation is important as it clarifies that the order of the cascade diffusion does not affect the ultimate result of the competitive process.

Algorithm 11: Competitive Independent Cascade model

Data: graph G, negative seed set $NS$, positive seed set $PS$

Result: The negative and positive activated node sets $NS$ and $PS$

1. $new\_negative\_activations \leftarrow NS$
2. $new\_positive\_activations \leftarrow PS$
3. $negative\_activators \leftarrow$ set of negative activators for each node
4. $positive\_activators \leftarrow$ set of positive activators for each node
5. healed $\leftarrow$ heal($NS$) according to healing strategy
6. while $\exists$ new_negative_activations or new_positive_activations do
7. if $\exists$ new_negative_activations then
8. one iteration of Algorithm 9 for new_negative_activations, negative_activators and $NS$
9. end
10. if $\exists$ new_positive_activations then
11. for $u \in new\_positive\_activations$ do
12. for $v \in N_u$ do
13. if $v \notin PS$ then
14. $p \leftarrow$ random probability
15. if $p > (u,v)$-edge weight then
16. positive_activators[$v$] $\leftarrow$ positive_activators[$v$] $\cup$ {$u$}
17. $new\_positive\_activations \leftarrow new\_positive\_activations \cup$ {$v$}
18. end
19. end
20. $new\_positive\_activations \leftarrow new\_positive\_activations \setminus$ {$u$}
21. end
22. $PS \leftarrow PS \cup new\_positive\_activations$
23. healed $\leftarrow$ healed $\cup$ heal($new\_positive\_activations$) according to healing strategy
24. end
25. $NS \leftarrow NS \setminus$ healed
26. end
27. return $NS$, $PS$
Algorithm 12: Competitive Linear Threshold model

Data: graph G, negative seed set NS, positive seed set PS

Result: The negative and positive activated node sets NS and PS

1. `new_negative_activations ← NS`
2. `new_positive_activations ← PS`
3. `negative_activators ← set of negative activators for each node`
4. `positive_activators ← set of positive activators for each node`
5. `healed ← heal(NS) according to healing strategy`
6. `while ∃ new_negative_activations or new_positive_activations do`
7.     `if ∃ new_negative_activations then`
8.         `one iteration of Algorithm 10 for new_negative_activations, negative_activators and NS`
9.     `end`
10.    `if ∃ new_positive_activations then`
11.       `for u ∈ new_positive_activations do`
12.           `for v ∈ N_u do`
13.               `if v ∉ PS then`
14.                   `positive_activators[v] ← positive_activators[v] ∪ {u}`
15.                   `positive_node_influences[v] ← positive_node_influences[v] + (u, v) - edgeweight`
16.                   `if positive_node_influences[v] > thresholds[v] then`
17.                       `new_positive_activations ← new_positive_activations ∪ {v}`
18.                   `end`
19.               `end`
20.       `new_positive_activations ← new_positive_activations \ {u}`
21.     `end`
22.     `PS ← PS ∪ new_positive_activations`
23.     `healed ← healed ∪ heal(new_positive_activations) according to healing strategy`
24.     `NS ← NS \ healed`
25. `end`
26. `return NS, PS`

5.2 Blocking Aggression Minimization

Proceeding to the second part of the proposed aggression minimization methodology, Blocking Aggression Minimization (BAM) is presented. In contrast to CAM, here there is only one cascade, the aggressive. Additionally, the aim is to target specific nodes or edges to immunize or block, which means that their removal would optimally suppress aggression diffusion over the network. However, to the best of the author’s knowledge, currently all the known immunization techniques address abusive user and behavior discovery rather than the more specific aggressive behavior discovery that this work
discusses. Hence, in the theoretical analysis of this chapter an aggression-based variation of the tested methods is also presented.

5.2.1 Theoretical Analysis

Following the general discussion on BAM, there is a necessity to address why these methods are needed. In the case of social networks, node removal is equal to banning a user which is a drastic but yet very famous measure. This is why the case of edge removal is also considered. Many social media platforms such as Twitter exploit these methods in order to find abusive users, but even the more relaxed version of edge blocking is still very intrusive and has a negative effect on user satisfaction. Thus, the methods addressing BAM are implemented in order to be compared to the competitive scenario and find out which one works better in the case of aggression minimization. Before proceeding to the specifics of BAM, however, the definitions regarding the two versions of the problem should be presented:

Problem 2 (BAM-N). (node version)

Given a directed network \( G = (V, E) \), with \( V \) and \( E \) denoting the node and edge sets respectively and an integer \( k \) find a node subset \( S \subseteq V \) with \(|S| = k \) s.t. \( S = \arg \min_{s \in P(V)} A(G) \) where \( A(G) = \sum_{v \in V} A_v \) is the overall aggression score of the network and \( P(V) \) is the powerset of \( V \).

Problem 3 (BAM-E). (edge version)

Given a directed network \( G = (V, E) \), with \( V \) and \( E \) denoting the node and edge sets respectively and an integer \( k \) find an edge subset \( S \subseteq E \) with \(|S| = k \) s.t. \( S = \arg \min_{s \in P(E)} A(G) \) where \( A(G) = \sum_{v \in V} A_v \) is the overall aggression score of the network and \( P(E) \) is the powerset of \( E \).

Based on the major finding of [59] and [19], for a large family of diffusion processes the only network parameter that determines whether this diffusion would become an epidemic or not is the largest eigenvalue \( \lambda \) of the adjacency matrix \( A \). For this reason we use NetShield [55] for the BAM-N problem and NetMelt [54] for BAM-E, algorithms that are based on this observation. Driven by the results of those two and competitive minimization, presented in the corresponding section later, we also implement an aggression-related variation. There, instead of using the initial edge weights we exploit the product of the aggression scores of source and destination nodes, that is for \((u,v)\) pair in the adjacency matrix the cell value now becomes \( A_u \times A_v \).
5.2.2 Specifications

As already mentioned in the previous section, apart from competitive cascades we have also exploited immunization techniques in order to address aggression minimization. Below, offline algorithms NetShield \[55\] and NetMelt \[54\] address BAM-N and BAM-E respectively. Both of these methods exploit the largest eigenvalue of the adjacency matrix to calculate a score able to exploit the most abusive nodes.

**Algorithm 13: Aggressive NetShield**

**Data:** graph $G$, budget $k$

**Result:** The node set $S$ to block

1. $G' \leftarrow G$ with weights the $A_u * A_v$ for $(u,v)$ - edge
2. $A \leftarrow$ adjacency of $G$
3. $\lambda \leftarrow$ the largest eigenvalue of $A$
4. $u \leftarrow$ the eigenvector corresponding to $\lambda$
5. $n \leftarrow$ number of nodes in $G$
6. $S \leftarrow$ empty
7. for $j=1$ to $n$ do
   8. \hspace{1em} $v[j] \leftarrow (2 * \lambda - A[j,j]) * u[j]^2$
8. end
9. for $i=1$ to $k$ do
   10. \hspace{1em} $B \leftarrow A[:, S]$
11. \hspace{1em} $b \leftarrow B * u[S]$
12. \hspace{1em} for $j \leftarrow 1$ to $k$ do
13. \hspace{2em} if $j \in S$ then
14. \hspace{3em} \hspace{1em} $score[j] \leftarrow -1$
15. \hspace{2em} else
16. \hspace{3em} \hspace{1em} $score[j] \leftarrow v[j] - 2 * b[j] * u[j]$
17. \hspace{2em} end
18. \hspace{1em} end
19. \hspace{1em} $i \leftarrow \text{argmax}_{j} score[j]$
20. \hspace{1em} $S \leftarrow S \cup \{i\}$
21. end
22. return $S$

**Algorithm 14: Aggressive NetMelt**

**Data:** graph $G$, budget $k$

**Result:** The node set $S$ to block

1. $G' = G$ with weights the $A_u * A_v$ for $(u,v)$ - edge
2. $A =$ adjacency of $G$
3. $\lambda =$ the largest eigenvalue of $A$
4. $u =$ the left eigenvector corresponding to $\lambda$
5. $v =$ the left eigenvector corresponding to $\lambda$
6. if $\min_{i=1,...,n} u[i] < 0$ then
   7. \hspace{1em} $u \leftarrow -u$
6. end
7. if $\min_{i=1,...,n} v[i] < 0$ then
   8. \hspace{1em} $v \leftarrow -v$
7. end
8. for edge $e_x : (i_x,j_x)$
   9. \hspace{1em} $e_x = 1,..., m i_x,j_x = 1,...,n$ do
10. \hspace{2em} $score[e_x] \leftarrow u[i_x] * v[j_x]$
11. end
12. for edge $e_x : (i_x,j_x)$
13. \hspace{1em} $e_x = 1,..., m i_x,j_x = 1,...,n$ do
14. \hspace{2em} $S \leftarrow S \cup \{\text{top-k edges with highest $score[e_x]$}\}$
15. return $S$
In algorithm 13, the exploitation of the largest eigenvalue is shown in lines 8 and 17, while in algorithm 14 line 13 fulfills the same purpose. It is noted that these two algorithms present the aggressive variation of NetShield and NetMelt, as depicted in line 1 of both algorithms. This means that the weights of the adjacency matrix (or the graph) are adjusted according to the aggression score of the nodes where an edge exists. This modification is based on the fact that the original implementations address abusive behavior blocking in general rather than the aggressive one’s.

**Conclusion:** In this chapter the theoretical analysis as well as the implementation and algorithms of the two aggression minimization approaches, competitive and blocking aggression minimization were presented fulfilling contribution 3 presented in introduction. Specifically, there is a double contribution from the above formulation. First, competitive cascades act as an alternative, less intrusive - in comparison to the blocking techniques - approach to the minimization of the social network’s overall aggression score. Second, blocking techniques, alongside their aggression-based modifications act as a safeguard in case competitive cascades can not be exploited as a minimization technique, due to the inadequacy of the cascading techniques to model aggression diffusion. Continuing, in Chapter 6 the results that arose from the experimental evaluation of the different configurations, both for aggression modeling and for aggression minimization will be discussed and analyzed completing Contributions 2 and 4 presented in introduction chapter 1. However, the first concept to address will be the data set used, as it is one of the most crucial factors both in the formalization of the process itself and in the outcome of the experimental phase. Additionally, there will also be a presentation of the technologies and programming libraries leveraged.
Chapter 6

Experimentation - Results

In this chapter the results of the experimental process are presented beginning with aggression modeling and proceeding to aggression minimization. However, the first thing to discuss is the technologies that were used to implement the theoretical models and strategies. For this purpose, the code was written in Python. Specifically, NetworkX \cite{1} was used, a library dedicated in network and graph manipulation. Additionally, Pandas, a library responsible for data manipulation and handling \cite{2} as well as Numpy \cite{3} for efficient mathematical operations were also exploited. Furthermore, there are some scripts written in batch code, so as to be able to simultaneously execute a plethora of experiments in an automated way from the command line. Last but not least, the various plots that will be presented during this chapter were produced by Matplotlib \cite{4} and Seaborn \cite{5}, the two most used python libraries for data visualization tasks.

Having presented the various technologies used in this thesis, the description of the leveraged data set follows. This step is crucial as the data set acts as a basic factor both in the formalization of the diffusion process itself and in the outcome of the experimental phase outlined below in this chapter.

6.1 Dataset Description

In this work, Twitter social media platform is investigated due to the fact that it provides more meaningful bidirectional follower-follower relationships, rather than the simpler friend relation. This means that an edge is directed as it expresses the will of a user to follow another user but not vice versa. If the opposite is also true, that is the second user wants to follow back the first, then a bidirectional edge is formed.
Regarding the nature of the dataset, we choose to leverage an existing, unlabeled but yet large social network of users on Twitter in which users connect with their neighbors through directional edges since they can be either followers or followed by others. The data set consists of 81,306 users or nodes, and 1,768,149 directed edges or links between them. During the experimental phase, however, it is decided to use the strongly connected component of the graph, which contains 68,413 nodes and 1,685,163 edges so as to get rid of isolated nodes or small groups that communicate only in the border limited by their reach. Edges between users can have weights on them, based on the intensity of interaction between the two users that an edge refers to, or based on the affinity or closeness of the two users. Following previous work on network analysis by Zuo et al., the weights applied on the edges are based on the three different weighting schemes - namely "Jaccard overlap", "Power score" and "Weighted overlap" - which were described previously in Section 4.2.

However, an unlabeled network can not be helpful to distinguish which users are aggressive and which are not. Thus, a way to identify this property is needed. For this purpose, the below process to label users as aggressive or normal is followed:

1. Based on a past, small Twitter network which contains labeled users - aggressive and normal through a crowdsourcing annotation task - a basic classifier leveraging users’ network features is trained to predict the likelihood of a user being aggressive in an unlabeled network.

2. The same user properties are extracted from the original unlabeled data set described above.

3. Last, the classifier is applied on this data set to label users based on some threshold.

The above process results in 8.5% or 5820 aggressive users with probability higher than 0.5. We also verify that these extracted aggressive users present similar distributions of network properties with the initial aggressive users of the small data set in where the classification algorithm achieved an AUC of 0.907. Moreover, the predicted user score is selected as the aggression score of each user for the modeling process. However, by selecting different aggression thresholds $T_A$ we can re-define the threshold that categorizes a user as aggressive. This experimentation is analyzed in Chapter 6.

\footnote{The numbers of user’s followers and followees, and their ratio, and the user’s clustering coefficient score, hub score, authority score, and eigenvector score.}
6.2 Results

The necessary tools to discuss the extracted results have been presented in the previous sections. Thus, following the experimental phase of both aggression modeling and aggression minimization is described.

6.2.1 Aggression Modeling

<table>
<thead>
<tr>
<th>Configuration encoding</th>
<th>Options</th>
</tr>
</thead>
<tbody>
<tr>
<td>{m}<em>{ws}</em>{ss}_{at</td>
<td>ts}</td>
</tr>
</tbody>
</table>

- **Parameter** | **Options**                  |
  - m: model     | ic: independent cascade     |
  -             | lt: linear threshold       |
  - ws: weighting scheme | J: jaccard overlap       |
  -             | P: power score ratio       |
  -             | W: weighted overlap        |
  - ss: seed strategy | r: random                  |
  -             | aa: all aggressive         |
  -             | ta: top aggressive         |
  -             | sd: single discount        |
  -             | dd: degree discount        |
  - at: aggression transferring | r: random                |
  -             | t: top aggressive          |
  -             | c: cumulative              |
  - ts: threshold strategy | a: aggression score        |
  -             | p: power score ratio       |

Each experiment is described by a set of parameters, the combination of which leads to a specific configuration. We refer to this configuration using the notation \{m\}_\{ws\}_\{ss\}_\{at|ts\}. The alternatives for each configuration parameter are presented in Table 6.1 exhaustively. For example, the configuration ic\_P\_r\_r means that in the corresponding experiment we use the IC model with power scores as weighting scheme picking seed nodes randomly and setting the aggression transferring to be random as well.

The aggression score of each user is calculated using the prediction process described in Section 6.1. Additionally, the accuracy of each configuration is measured using their proposed validation vector. Specifically, we compute the vector corresponding to the final state of the graph and compare it to the ground truth vector using cosine similarity.
Of course, we examine various alternatives as well, such as Pearson R and Spearman R statistics while we also experiment on calculating the validation vector on an intermediate step rather than the end of the process.

The presentation of the experimental results depends on the underlying diffusion model. In order to investigate the various IC based configurations we run every experiment 10 times due to the inherent randomized nature of the activation process. Additionally, when "random" is used as the seed strategy - further discussion will follow in the corresponding section - it inserts an additional random factor in the overall process. In this case we run each experiment an additional 10 times, leading to a total of 100 experiments for the specific configurations. To validate whether there is a significant parameter we use One-Way ANOVA followed by a pairwise post-hoc Tukey’s HSD test to spot the exact value of the significant parameter. Last, we present these results using cumulative density function (CDF). However, in LT case, there is no probabilistic step and thus we run each experiment only once except for the case of "random" seed strategy where we execute each one 10 times.

6.2.1.1 Seed size and Aggression Threshold

The experimentation process is comprised of all the possible combinations of the above parameters. Howbeit, the first step of the process is to determine the best number of seed nodes, the budget k, if this is necessary. Even though we can choose any seed size, we decide to use a size of 5594 corresponding to approximately 8% of the nodes, as long as this was the percentage of aggressive users predicted by the algorithm of Section 6.1. This number of seeds is not hard-coded but derives from an initial aggression threshold of 0.695 as this is the proper value to acquire the specified percentage of nodes. Thus, we can either pick the aggressive users or any combination of other users as the initial seed set.

Except for the seed size, we initially experiment on different aggression threshold values, denoted as $T_A$, used in deciding when a user turns from normal to aggressive. Although aggression thresholds of $T_A = \{0.1, 0.2, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9\}$ were tested, only thresholds $\leq 0.4$ depicted interesting and varying behavior throughout the experiments. However, the smaller the threshold the smaller the achieved cosine similarity to the ground vector and hence we decide to keep $T_A = 0.4$ for the rest of the experimental phase which corresponds to a cosine similarity of 0.69. Here we have to clarify that thresholds $\geq 0.4$ - even though they achieve extremely high cosine similarity (0.97) - do not present any interest because they are unreachable considering the fact that the majority of nodes has an aggression score of 0.3 and the changes during the diffusion...
process are affecting the second and third decimal point. Thus, using a $T_A$ in that range is similar to freezing the network on its initial state during the whole diffusion process.

### 6.2.1.2 Seed Selection Strategy

The first thing to observe from the experimental results is the effect that seed strategies have on the overall efficiency of the models. We distinguish the results of IC and LT-based models due to the different presentation processes. By now, we have reduced the range of experimental spectrum to $T_A = 0.4$ and seed size = 5594. Having decided to keep the specific seed size, "top aggressive" strategy is not an option, as all aggressive users can be used without leaving someone aside. Figure 6.1 presents the results regarding IC-based models and also shows the outcome of the significance tests in order to decide if the observed dominance of a specific value is indeed significant. It is noted that for brevity reasons we do not present the exhaustive list of experiments, however each one of them depicts similar results to the presented.

**Figure 6.1:** Cosine similarities Tukey’s HSD pairwise significance test for IC experiments with regard to the seed strategy

Figure 6.1 shows a dominance of network-feature strategies, meaning "Single Discount" and "Degree Discount", over the others. There is an exception in the third image of the figure, where "random" strategy behaves similarly to "Degree Discount" but yet "Single Discount" is the most prevalent. However, despite the above observations, we should also refer to the significance tests. ANOVA tests calculated an F-statistic $> 23.41$ and probability $p < 4.15 * 10^{-12}$, meaning that we can reject the null hypothesis of the lack of significant difference between the tested parameter values as long as this hypothesis can occur with probability $= p$. Tukey’s HSD found many significantly different pairs but we focus on the ones achieving the best results. As the three bottom images of Figure 6.1 depict, both in Power graph and Weighted overlap cases "Single Discount" and "Degree Discount" are not significantly different and thus both of them are dominant...
strategies. Notwithstanding in the case of Jaccard graphs "Single Discount" is prevalent. Hence, choosing as seed strategy "Single Discount" is the way to go regardless the other parameters.

Table 6.2: Cosine Similarities for LT experiments

<table>
<thead>
<tr>
<th></th>
<th>lt_J_aa_a</th>
<th>lt_P_aa_a</th>
<th>lt_W_aa_a</th>
<th>lt_J_aa_p</th>
<th>lt_P_aa_p</th>
<th>lt_W_aa_p</th>
<th>lt_J_r_a</th>
<th>lt_P_r_a</th>
<th>lt_W_r_a</th>
<th>lt_J_r_p</th>
<th>lt_P_r_p</th>
<th>lt_W_r_p</th>
<th>lt_J_sd_a</th>
<th>lt_P_sd_a</th>
<th>lt_W_sd_a</th>
<th>lt_J_sd_p</th>
<th>lt_P_sd_p</th>
<th>lt_W_sd_p</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0.6882</td>
<td>0.6896</td>
<td>0.6896</td>
<td>0.6883</td>
<td>0.6896</td>
<td>0.6896</td>
<td>0.6905</td>
<td>0.6896</td>
<td>0.6896</td>
<td>0.6907</td>
<td>0.6897</td>
<td>0.6897</td>
<td>0.6907</td>
<td>0.6897</td>
<td>0.6897</td>
<td>0.6905</td>
<td>0.6897</td>
<td>0.6897</td>
</tr>
</tbody>
</table>

We now investigate the effect of seed strategy on LT-based configurations. We note that "degree discount" although applicable it is not theoretically appropriate for the case of the LT model and thus we do not trace its results. Furthermore, in the case of "random" strategy, we present the mean similarity achieved by a repetition of 10 experiments due to its inherent probabilistic nature, as pointed out at the beginning of this section. By looking at Table 6.2 and focusing again on the configurations of Jaccard-weighted graphs it is clearly shown that similar to the case of IC the most dominant strategy is "Single Discount". Continuing our analysis one step further, we can see that in Power Score-weighted graphs "Single Discount" is again prevalent regardless of the threshold criterion, while in Weighted Overlap-weighted graphs there is no significant difference between the various seed strategies. Thus, once again, similar to the case of IC we select "Single Discount" as the most prevalent seed strategy for every LT configuration.

6.2.1.3 Weighting Scheme

In the previous section, we confined our investigation scope even more by selecting "Single Discount" as the most appropriate seed strategy. We now continue to the experimentation on the different weighting schemes.

Figure 6.2 shows a clear dominance of the "Jaccard overlap" weighting scheme over the rest, regardless of the activation criterion. Thus a significant test is quite unnecessary but we present it for completeness purposes. Specifically, we validate that indeed "Jaccard overlap" is significantly superior to the other schemes. ANOVA showed an F-statistic bigger than $2.072 \times 10^4$ with probability $p < 9.65 \times 10^{-44}$ for all cases, validating the crystal-clear supremacy of Jaccard weighting scheme.

Moving to the LT-based configurations and experiments, Table 6.2 again depicts that the best performance is achieved by Jaccard-weighted graphs. In fact, depending
on the seed strategy Jaccard-weighted graphs may be the worst-case scenario. However, we have already observed that "Single Discount" is the best seed strategy and for this scenario, the best efficiency is achieved. Additionally, weighted overlap and power score as weighting schemes do not present any actual fluctuations - deviations are observed on the sixth decimal point - constituting them inappropriate weighting schemes. The above two reasons lead to a straightforward choice of "Jaccard overlap" as the best weighting scheme, just like the case of IC-based models.

### 6.2.1.4 Activation criterion and Threshold strategy

Until now, we have examined the whole spectrum of configuration from a macroscopic point of view and we have seen that aggression diffusion modeling is heavily related to the right selection of weighting scheme, considering that this is the factor that captures node interactions. Moreover, counter-intuitively a simple seed selection strategy like "Single Discount", which is highly dependent on basic user network features such as node’s degree, is also more preferable than picking a method that addresses aggression directly.

Specifically, Figure 6.3 shows that "cumulative" activation strategy is the most dominant while there is no clear distinction between "top" and "random" strategies. In addition, this observation is validated by ANOVA test too. Specifically, it resulted in an $F_{statistic} = 10.64$ and a probability $p = 3.92810^{-4}$ to accept the null hypothesis of lack of significance. For interpretation purposes we present the results of Tukey’s HSD test in the second part of the figure where it is shown that the value ranges of "top" and "random" overlap with each other while "cumulative" is obviously superior. Moreover,
the choice of "cumulative" is rational considering the fact that a user’s community behavior is a crucial factor of overlining their own behavior as it is also pointed out by the General Aggression Model [7].

In a similar manner, for LT-based configurations, we should again refer to Table 6.2 to test the efficiency of each threshold strategy. Again, we focus on the experiments regarding "Jaccard overlap" weighted graphs with "Single Discount" as the seed selection mechanism. There we can observe that both "power" and "aggression score" thresholds operate sufficiently, but the latter is the one presenting slightly better performance. Thus, in LT-based configurations, "aggression score" threshold should be used to model aggression propagation in the best possible way. This observation is intuitive regarding the fact that the overall LT model should enable the diffusion of aggression while at the same time the usage of the aggression threshold expresses this exact notion and idea.

![Figure 6.3: Cosine similarities Tukey’s HSD pairwise significance test for "Single Discount" seed strategy and "Jaccard Overlap" weighting scheme on IC experiments with regard to activation criterion](image)

Wrapping it all up, we observe that both for IC and LT-based models the best seed strategy is "Single Discount" meaning that the most accurate models of aggression diffusion initiate their process from the nodes with higher degree, meaning the most central points in the network. Moreover, for both cases the best performing weighting scheme is "Jaccard overlap" which means that nodes form their relations based on their neighborhoods and their similarity. Furthermore, for IC-based models activation criterion should be set to "cumulative" enabling the whole neighborhood of a node to affect its aggression state, while for LT-based ones, "aggression score" threshold strategy is the most prevalent. Once again, "aggression score" is a rational choice considering that it is governed by the same notions that enable the diffusion of aggression during the propagation process of the LT model.
6.2.1.5 Snapshot evaluation

Until now, we have fully covered the set of parameters and the possible configurations. It is noted that in the above scenarios we let the diffusion process converge in order to calculate the similarity with ground truth. However, there might be a configuration that can better model aggression diffusion at an intermediate step of the propagation process. For this purpose, for each experiment we create snapshots of the diffusion process at the end of every time step.

Figure 6.4 shows how the performance in each time step is affected in relation to three different similarity metrics, cosine similarity, Pearson and Spearman correlations. In this way, we can also investigate whether there is a more appropriate metric for the experimental process. We notice that in the case of IC, the performance in terms of cosine similarity and Pearson correlation decreases within the first snapshots and gradually increases, to stabilize in the last snapshots. The exact opposite behavior is observed for Spearman coefficient due to the transformation of the first variables to rank variables according to metric’s definition. The similar trend of Pearson correlation coefficient with cosine similarity makes sense regarding that the former is a demeaned version of the latter. Additionally, the similar metric pattern across models depicts that either of them can be exploited in the performance analysis process. For the LT case, the exact opposite behavior is observed with an early Cosine and Pearson increase and a fall in the following steps, to ultimately reach stability and an early decrease for the Spearman coefficient followed by an increase until stabilization, once again.

Regarding the snapshots themselves we observe that there is indeed an earlier snapshot where the corresponding models achieve better efficiency, which is spotted on the second step for the case of IC and on the third step for the LT. Thus, we could force the process to stop on these specific time steps rather than unfold until convergence. However, during the last steps of the diffusion process a more stable behavior is observed and hence the decision to keep the convergence value is rational.

Conclusion: Combining the above experimental observations, it is observed that aggression diffusion over social networks can be modeled with a cascading model, such as IC or LT. Additionally, Jaccard overlap is the weighting scheme that best captures the interactions between users with respect to aggression. Moreover, a simple seed strategy that targets the most central users early on is efficient in approximating the real network’s state. Last, a cumulative neighborhood behavior enables the IC diffusion process while the LT one benefits from aggression score based thresholds. These models can either converge to achieve a stable state or forcibly cut during the initial steps to achieve the
best modeling similarity. Last, the above observations fulfill Contribution 2 while the statistical results complete Contribution 4.

6.2.2 Aggression Minimization

6.2.2.1 Competitive Aggression Minimization

By presenting the results of the aggression modeling section, we are now emphasizing on the importance of minimizing the network’s overall aggression score. Table 6.3 clarifies the notation used for each experiment of this section, similar to the previous one. Specifically, we exploit a competitive cascade process in order to address CAM problem and mitigate or even decrease the final aggression score of the network. For this purpose, we choose the competitive positive cascade to be similar to the negative one. This means that for both IC and LT models, the chosen weighting scheme is ‘Jaccard overlap’, negative cascade’s seed strategy is ‘Single Discount’ and both positive’s and negative’s cascade seed size is equal to 5594. For the IC scenario, the activation strategy is ‘cumulative’ while for LT we use ‘aggression score’ as threshold strategy. In this way, we do not favor one of them against the other but rather we investigate the rest of the parameters and their effect on aggression minimization. Specifically, we experiment on different seed strategies for the positive cascade as well as various healing mechanisms for the nodes influenced by the positive piece of information under diffusion.

We focus on the best configuration for each one of the IC and LT models found so far which is the ic$_{-J_{sd}}$c and lt$_{-J_{sd}}$a. Figure 6.5 presents aggression evolution during the competitive cascade process for the various seed and healing strategies on the IC-based models. Y-axis presents the ratio of loss or gain with respect to the case where no healing is applied First of all, regarding the seed strategy of the positive cascade, it is
Table 6.3: Configuration encoding and parameter options of aggression minimization experimental phase

<table>
<thead>
<tr>
<th>Configuration encoding</th>
<th>Parameter</th>
<th>Options</th>
</tr>
</thead>
<tbody>
<tr>
<td>{m}<em>{ws}</em>{nss}_{hs}</td>
<td>m: model</td>
<td>ic: independent cascade</td>
</tr>
<tr>
<td></td>
<td></td>
<td>lt: linear threshold</td>
</tr>
<tr>
<td></td>
<td>ws: weighting scheme</td>
<td>J: jaccard overlap</td>
</tr>
<tr>
<td></td>
<td>nss: negative seed strategy</td>
<td>r: random</td>
</tr>
<tr>
<td></td>
<td></td>
<td>aa: all aggressive</td>
</tr>
<tr>
<td></td>
<td></td>
<td>sd: single discount</td>
</tr>
<tr>
<td></td>
<td></td>
<td>dd: degree discount</td>
</tr>
<tr>
<td></td>
<td>hs: healing strategy</td>
<td>v: vaccination</td>
</tr>
<tr>
<td></td>
<td></td>
<td>t: transfer</td>
</tr>
<tr>
<td></td>
<td></td>
<td>d: decay</td>
</tr>
<tr>
<td></td>
<td></td>
<td>h: hybrid</td>
</tr>
</tbody>
</table>

Depicted that, "Random" presents the best results for every healing mechanism reaching 57.8% reduction, while "Degree Discount" follows with 57.3%. "Single discount" and "all aggressive" come last with 55.3% and 53.5% respectively. Thus the seed strategy for the positive cascade should be set to "Degree Discount". Proceeding to the healing strategies, transfer method presents the worst performance, as it decreases the overall aggression score of only by 8%, while in the case of "all aggressive" it even increases it by 15%. This observation makes sense considering that this method - as its name suggests - simply transfers the aggression score of the activating neighbor without being aware of its effect on the node under investigation. Regarding the rest of the healing mechanisms, it is shown that they perform similarly regardless the seed strategy. For "Degree Discount" which is the best strategy regarding the inherent uncertainty of "Random" method, "Vaccination" is the most dominant mechanism achieving 57.3% aggression reduction while "Decaying transfer" and "Hybrid" follow with 57% roughly. These results are intuitive as "Vaccination" makes the strong assumption that nodes get completely healed, "Decaying transfer" is a more relaxed and realistic version of "Vaccination" and "Hybrid" lays in between these two.

Another critical factor of the healing strategy's efficiency is the number of nodes that get activated by the negative cascade. This is important considering that a healing strategy is efficient if the overall aggression score is low while there are a lot of activated nodes, as this would mean that the negative piece of information indeed propagated but nodes were well-prepared. Figure 6.6 shows that all healing mechanisms except "Transfer" activate approximately the same number of nodes. The big difference with
Figure 6.5: Aggression evolution for different seed strategies on IC models with respect to the case where no healing is applied. Healing strategies are: Decaying transfer, Hybrid, Transfer and Vaccination.

"Transfer" lays in the fact that it is not a healing mechanism per se as it simply transfers aggression not considering if it is more or less than the current one. Thus, the effect of healing is depicted in the rest of the mechanisms where the number of activated nodes is drastically reduced in all cases. Combining the above observations, we note that "Vaccination" is the best performing mechanism but due to its mostly theoretical nature we point out that "Decaying Transfer" or ideally "Hybrid" could be used to minimize aggression in IC-based models.

Proceeding to the LT-based configurations, we present the respective results in Figures 6.7 and 6.8. Figure 6.7 specifically, presents a more clear fashion in comparison to the IC case. Here, "Degree Discount" is not an option as it has been explained already. For the rest of the strategies we note that "Single Discount" strategy is dominant achieving 92% reduction with "Random" following pretty close at 90%. Moving to the healing strategies, "Vaccination" achieves the best reduction of approximately 92% but here there is a big prevalence in relation to "Hybrid" and "Decaying transfer", as the prior does not only presents the best performance but it is also the only one that preserves the reduction during the whole process. This can be explained by the fact that "Vaccination" completely deactivates the corresponding nodes and thus they have zero
effect on their neighbors during the following steps. On the contrary, the rest methods even though they achieve a reduction of 90% they cannot preserve this efficiency, as LT-models ground their spreading mechanism on the cumulative effect of the neighborhood. However, considering that the nodes are sensitized by the healing mechanism, the overall aggression score is still reduced - in "Single Discount" and "Random" - strategies. "All aggressive" case, however consists an exception as these strategies lead to an increase in the relative aggression score. Hence, while the best results are achieved by "Vaccination" we observe that in LT-based models the most possible reduction that can be achieved starts from 5% and can raise up to 20% using as mechanism the "Hybrid" solution. However, if we could ensure the conditions to use "Vaccination" as the healing mechanism in a real case scenario, then we would be able to almost eradicate the aggression from the network.

Regarding the activated nodes during each propagation process, Figure 6.8 depicts the intuitive fact that "Transfer" mechanism activates a significant amount of nodes while "Vaccination" the least. This small number of activated nodes in the case of "Vaccination" is an enabling factor to achieve a high reduction as observed before. Additionally, the fact that "Hybrid" activates approximately the same amount of nodes
Figure 6.7: Aggression evolution for different seed strategies on LT models with respect to the case where no healing is applied. Healing strategies are: Decaying transfer, Hybrid, Transfer and Vaccination.

with "Vaccination" in all cases but performs analogously to "Decaying transfer" clarifies that between "Hybrid" and "Decaying transfer" the latter is the way to go.

To sum it all up, we observe that competitive cascades can indeed reduce the overall aggression score of the network, but differently concerning the underlying aggression model. The best reduction in IC-based models is achieved when "Random" or "Degree Discount" seed strategy is selected for the positive cascade in combination with "Decaying transfer" healing mechanism, achieving a reduction of 57%. On the other hand, the LT-based models are more robust and resistant to aggression minimization methods. The most realistic method should be to use "Single Discount" seed strategy alongside "Hybrid" healing to achieve a reduction of 5% to 20%. However, if we can ensure that "Vaccination" can be applied in a real-world scenario then we can almost wipe out aggression as the achieved reduction reaches 92%.
6.2.2.2 Blocking Aggression Minimization

Apart from the competitive cascade method presented above we also experiment on blocking mechanisms that may help in decreasing the overall aggression level of the network. Figure 6.9 shows the results of four different methods. Based on whether we block nodes or edges we test NetShield [55] and NetMelt [54] algorithms correspondingly. Additionally, driven by the results of competitive aggression minimization we implement a variation of the above algorithms where the adjacency matrix used, in position \((u, v)\) contains the product of \(u\)'s and \(v\)'s aggression scores rather than the initial edge weight.

Specifically, Figure 6.9 depicts that in the case of IC the variation on the weighting of the adjacency matrix does not present any important improvements as the differences are observed on the first decimal point. Moreover, it is shown that node removal can reduce the overall aggression score by 8% but on the contrary, edge removal increases it by 100%, despite the steep drop at the end of the process. This behavior is explained by the combination of the fact that edge removal interferes with the degree of the nodes in the network and that the negative cascade selects its seeds using "Single Discount", a
degree-based strategy. Node removal is not the same as it discards whole nodes and not just edges.

However, for the LT case, different behavior is observed. Here edge removal presents a periodical behavior in relation to the zero case of no immunization, meaning that initially it decreases the aggression score by 1% but afterwards it increases it in a similar manner while at the end it drops to the initial levels. These fluctuations, though, are too negligible to produce a significant outcome. Regarding node removal, there is a better performance in terms of aggression minimization as the reduction reaches 6% for the normal adjacency and 8% for the aggression-based matrix. Generally speaking, the aggression alternative of the adjacency matrix seems to benefit the process but further research is needed.

For completeness purposes in Table 6.4 we present the number of nodes that each one of the aforementioned immunization methods get to activate. We can observe that both for the IC and LT models, the aggression reduction is tightly related to the small number of activated nodes when node blocking is leveraged in contrast to edge blocking. However, in comparison to the competitive cascades the immunization methods tend to activate 4x times the nodes that the former affect, as it can been extracted from the combination of Figures 6.6, 6.8 and Table 6.4.

**Table 6.4:** Activated nodes for different blocking mechanisms on IC and LT models

<table>
<thead>
<tr>
<th></th>
<th>IC</th>
<th>LT</th>
</tr>
</thead>
<tbody>
<tr>
<td>ic_J_edge_c_normal</td>
<td>38134</td>
<td>50716</td>
</tr>
<tr>
<td>ic_J_edge_c_aggression</td>
<td>38421</td>
<td>50716</td>
</tr>
<tr>
<td>ic_J_node_c_normal</td>
<td>28844</td>
<td>39212</td>
</tr>
<tr>
<td>ic_J_node_c_aggression</td>
<td>27493</td>
<td>39760</td>
</tr>
</tbody>
</table>
**Conclusion:** It is observed that competitive cascade minimization outperforms blocking techniques in terms of reducing the overall aggression score of the network. Specifically, it achieves a reduction of 57% of the initial total aggression score of the network on the IC models, while blocking methods - in the best case scenario - reach a reduction of 8%. Concerning LT models, again competitive cascades are prevalent as they achieve a 5-20% reduction in contrast to the 8% of the blocking methods. That means that instead of banning and blocking users from social media, which is the current methodology adopted, social media platforms could instead educate users through posts or campaigns on topics related to aggression and similar phenomena. Additionally, these observations cover Contribution 3 of Chapter 1.

Chapter 6 concludes the contributions of this work by discussing the experimental results. In Chapter 7 we discuss the findings of this thesis and we also propose some potential directions for future research.
Chapter 7

Discussion and Future Work

This chapter discusses the results that have arisen from the current work. In particular, the dual problem of aggression modeling and aggression minimization in online social networks and specifically Twitter has been addressed. Even though a lot of effort has been made in the domains of information diffusion and influence maximization, there are no prior works currently addressing aggression modeling. An exception is a work of Kourtellis et al. - currently under peer review - which approaches aggression modeling through opinion dynamics, but this work has yet to be published. Regarding aggression minimization, howbeit, we are the first to address this problem. Specifically, models from the domains of competitive cascades and immunization techniques are exploited and put into test.

First of all, it has been shown that aggression diffusion can indeed be modeled with both IC and LT-based propagation processes as they can reach a decent cosine similarity of approximately 70% with the ground truth. In order to achieve this performance, the various parameter values that affect the model have to be carefully selected. In particular, it has been depicted that "Jaccard Overlap" is the weighting scheme that best captures the interactions between users with respect to aggression for both types of diffusion processes. This observation is based on the fact that nodes form their relations based on their neighborhoods and their similarity. Additionally, again both for IC and LT, "Single Discount" is the most proper seed selection strategy denoting that aggression diffusion benefits from the most central nodes in a network rather than the most aggressive ones.

Diving deeper into the specifics of each approach, it has been observed that the "cumulative" is the best activation criterion for the IC model as it captures the notion and structure of the community and team effort to transfer the underlying piece of information. On the other hand, in the case of LT model, "aggression score" threshold is
the best threshold strategy, an observation that makes sense regarding that this strategy is governed by the same notions that enable the diffusion of aggression during the propagation process of the LT model.

According to the results of this thesis, with respect to aggression modeling, some interesting future directions would be to:

1. implement further theoretical features and characteristics of aggression in the diffusion models in order to increase the similarity with the ground truth.

2. investigate the various thresholds that enable an epidemic spread of aggression over a social network under different virus spreading models such as SIR, SIS or SIRS and their deviations.

Proceeding to the aggression minimization discussion, here it has been shown that both competitive cascades and immunization methods can reduce the overall aggression score of the network. However the former is prevalent in comparison to the latter as it can achieve higher reduction rates. Specifically, competitive cascades reach a total of 57% reduction in IC and 5-20% in LT model, while immunization or blocking methods achieve an 8% reduction in both IC and LT. Here, it is necessary to clarify that regarding blocking mechanisms, edge blocking is inappropriate and social media platforms should only depend on node deletion. A possible explanation of this observation is the fact that removing edges from the network intervenes with the seed selection strategy of the aggressive cascade leading to unpredictable behavior, considering that "SD" and "DD", two degree based strategies were shown to better describe aggression diffusion. This result validates today’s methods of dealing with online aggression, as social media such as Twitter prefer to block and ban abusive users than following a different approach. However, these approaches are rather intrusive and affect user experience and satisfaction negatively.

The intrusive nature of blocking techniques alongside the better results of competitive cascades explain the authors’ preference on competitive cascades as an aggression minimization method. Diving deeper into the mechanics of competitive cascades, the experimental phase has shown that for the positive diffusion process, seed strategies highly dependent on users’ network features are the most useful approach in aggression minimization too. In particular, "Degree Discount" produces the best results with regard to the IC model, while in the case of LT - where "Degree Discount" is not valid - "Single Discount" has exhibited the highest performance. Regarding the healing mechanism, it has been discovered that "Decaying Transfer" is the most appropriate even though it does not achieve a better performance than "Vaccination". This happens because
"Vaccination" presumes that the positive piece of information, such as a cyberbullying sensitizing campaign, affects users in a way that they couldn’t slip back to abusive behavior. Obviously, this assumption is rather ideal and thus not easily applicable in a real-world scenario. However, if the conditions of achieving the results of "Vaccination" could be ensured then, in LT case, almost an eradication of aggressive phenomena from the social network could be achieved.

Combining the observations on competitive cascades and blocking techniques, the usage of a consciousness raising campaign is proposed, such as a post on racism extinction instead of the banning of a user to prevent the outbreak of aggression in online social networks.

Under aggression minimization there is a lot of work to be done, as this thesis is the first to address the problem, to the best of the author’s knowledge. Thus, some interesting future ideas for further investigation are proposed below:

1. Node blocking techniques seem to reduce the overall aggression score of the network and the proposed aggression-based adjacency matrix has an even bigger reduction. For this reason, further research has to be done on the proper manipulation of the adjacency matrix such as how to combine the aggression scores of two users to produce the equivalent entry of the matrix.

2. The proposed blocking techniques are based on the exploitation of the first eigenvalue of the adjacency matrix. Possible research would be to investigate how other spectral methods perform on the task of aggression minimization and if they can be adjusted to match the challenges that it introduces.

3. In combination with the investigation of the virus spreading models proposed in aggression modeling discussion, some interesting results would be extracted from the epidemic thresholds and the behaviors observed during the process. Could they be leveraged to reduce the network’s aggression score?
Appendix A

Mathematical Notation

In this appendix we present the mathematical notations used in the specific thesis.

A.1 General Notation

\( \in \) Belongs to
\( \sum \) Sum
\( \mathcal{O} \) Worst-case complexity
\( \leftarrow \) Left deduction

A.2 Set Notation

\( \subseteq \) Subset
\( \subset \) Pure Subset

A.3 Logic Notation

\( \exists \) Exists

A.4 Statistics Notation

\( \epsilon \) Unexplained Variation
\( \mu \) Population Mean Value
\( s \) Standard Deviation
Bibliography

[1] https://networkx.github.io/


[12] Despoina Chatzakou, Nicolas Kourtellis, Jeremy Blackburn, Emiliano De Cristofaro, Gianluca Stringhini, and Athena Vakali. Mean birds: Detecting aggression and


ourworldindata.org/internet#growth-of-the-internet


[51] Max Roser, Hannah Ritchie, and Esteban Ortiz-Ospina. World pop-
ulation by world regions. https://ourworldindata.org/grapher/
world-population-by-world-regions-post-1820


[54] Hanghang Tong, B Aditya Prakash, Tina Eliassi-Rad, Michalis Faloutsos, and Christos Faloutsos. Gelling, and melting, large graphs by edge manipulation. In Pro-


[57] Kurt Wagner. Twitter says it’s getting better at detecting abusive
twitter-abuse-update-health-technology-harassment

social-media-users/

[59] Yang Wang, Deepayan Chakrabarti, Chenxi Wang, and Christos Faloutsos. Epi-
demic spreading in real networks: An eigenvalue viewpoint. In 22nd International
2003.

[60] Zeerak Waseem and Dirk Hovy. Hateful symbols or hateful people? predictive
features for hate speech detection on twitter. In Proceedings of the NAACL student
research workshop, pages 88–93, 2016.

[61] Peng Wu and Li Pan. Scalable influence blocking maximization in social networks
under competitive independent cascade models. Computer Networks, 123:38–50,
2017.

[62] Huiyuan Zhang, Huiling Zhang, Xiang Li, and My T Thai. Limiting the spread of
misinformation while effectively raising awareness in social networks. In Interna-

[63] Xiang Zuo, Jeremy Blackburn, Nicolas Kourtellis, John Skvoretz, and Adriana
Iamnitchi. The power of indirect ties. Computer Communications, 73:188–199,
2016.