Politecnico di Milano

Facoltà di Ingegneria
Scuola di Ingegneria Industriale e dell'Informazione
Dipartimento di Elettronica, Informazione e Bioingegneria

Master of Science in
Computer Science and Engineering

A Fancy Title for a Fancy Thesis

Supervisor:
Prof. Name Surname

Assistant Supervisor:
Dr. Name Surname

MSc Thesis by:
First Author
matr. 000000
Second Author
matr. 111111

Academic Year 20MM/20MN

Insert here your dedication.
Optional.

Abstract

Here goes the abstract, a summary of your thesis work. You may add some keyword at the end that clearly identify the research field of the thesis. The abstract should not contain any reference to related works.

Keywords keyword1; keyword2; keyword3

Sommario

In questo capitolo puoi inserire l'abstract della tesi in italiano.

Parole chiave keyword1; keyword2; keyword3

Acknowledgements

In this chapter you can acknowledge the people that were somehow helpful in the realization of the thesis. It is better to keep this chapter formal, you can add the friendly thanks at the end (chapter Thanks).

Contents

List of Figures

List of Tables

List of Algorithms

Introduction 1

1.1 Context

This is basically an extension of the abstract. Here you provide context for the problem faced. Keep in mind that even if you now have gained expertise on it, most of the readers are no so inside the problem as you are. Start from the basics and explain clearly. You can also introduce here some hints about the methodology and your contribution. For this purpose, you may also decide to add more sections.

1.2 Thesis outline

Here you explain the structure of the thesis.

> “The thesis is structured in the following way:

- In chapter 2, we present ... .
- In chapter 3, we formulate the problem we address in the thesis and ... .
- In chapter 4, we present our solution for ... .
- In chapter 5, we show experimental results of our proposed methods in different settings ... .
- Finally, in chapter 6, we present our conclusions and possible future paths toward which our work could be extended."

Preliminaries and State of the Art

2

2.1 Preliminary notions

Notation	Description
G	Graph
V	set of nodes of G
E	set of edges of G
W	set of weights corresponding to each edge in E
$w_{u,v}$	weight of edge (u, v)
n	$\|V\|$, number of nodes
m	$\|E\|$, number of edges

Table 2.1: Graph notation.

"In this section, we introduce the preliminary notions at the base of our study. We start by briefly introducing the problem, and then we provide the necessary concepts and the notation used."

You may insert a subsection for each of the most relevant features of your problem. You can add some reference if needed, but just to explain the problem. The references with the solutions of the problem should be put in the next section.

You can keep a notation table for the notation used in this chapter as Table 2.1. Everything inside the notation table must be written at least once inside this chapter. You can put an extended notation for the whole thesis in the appendix.

It is likely that you have to present definitions, theorems or propositions. We suggests to use the following environments. You can cite them as Definition 2.1, Theorem 1, Proposition 2.1.

Definition 2.1 ((α, β)-approximation)**.** An (α, β)-*approximation* algorithm outputs with success probability β a solution which is at least α fraction of the optimal solution, for some $\alpha, \beta \leq 1$.

Algorithm 1 Combinatorial Thompson Sampling

Input: Directed graph $G(V, E)$, budget constraint k, time horizon T

for $i = 1$ **to** $|E|$ **do**
  $\alpha_i = 1, \beta_i = 1$ ▷ Assign a Beta distribution Beta(1,1) to each edge
for $t = 1$ **to** T **do**
  For each arm i, draw a sample $\theta_i(t) \sim \text{Beta}(\alpha_i, \beta_i)$
  Let $\boldsymbol{\theta}(t) = (\theta_1(t), \dots, \theta_m(t))$
  $S_t \leftarrow \texttt{oracle}(G(V, E, \boldsymbol{\theta}(t)), k)$
  Run cascade with S_t as the seed set and collect the feedback F_t
  Update the Beta distributions of the edges involved using F_t

Theorem 1. *For a non-negative, monotone, submodular function $f(\cdot)$, let S_k be a set obtained by selecting k elements one at a time, each time choosing an element that provides the largest marginal increase in the function value. Let S_k^* be a set that maximizes the value of $f(\cdot)$ over all k-sized sets. Then,*

$$f(S_k) \geq \left(1 - \frac{1}{e}\right) \cdot f(S_k^*).$$

Proposition 2.1. *If the diffusion process starting with S is simulated independently at least $r = \Omega\left(\frac{n^2}{\varepsilon^2} \ln\left(\frac{1}{\delta}\right)\right)$ times, then the average number of activated nodes over these simulations is a $(1 \pm \varepsilon)$-approximation to $\sigma(S)$, with probability at least $1 - \delta$.*

2.2 State of the Art

"In this section, we survey the most relevant works for the problems of ... ."

If you face a problem that has more than one macro-topic, you may choose to add a subsection for each of these topics (better no more than 2-3), like *Related works on Topic 1*, etc.

List the works in chronological order and cite only the most important and pertinent ones, avoid 100 citations for a master thesis.

You can insert pseudocodes of algorithms as Algorithm 1.

3 Problem formulation

This chapter is dedicated to the formal presentation of the problem with the technical details. Here your should put also the figure of merit you use to compare your solution with the ones of the related works you presented.

"In this thesis, we address the problem of ..."

"The figure of merit ..."

4 Design and Implementation

“This approach is similar to the one proposed in a recent work by Nuara et al. [2], in which ..."

5 Experimental evaluation

"In this chapter, we present experimental results on the algorithms proposed and we compare them with state of the art methods."

5.1 Experimental setting

Name	n	m	Average degree	Max $In(v)$	Max $Out(v)$
Email	971	3466	3.57	4	47

Table 5.1: Dataset used for the experiment.

Datasets Table 5.1 shows the characteristics of the dataset used for the experiment. The dataset is a subgraph obtained from a real dataset provided by SNAP [1], *email-Eu-core*, generated using email data from a large European research institution. A directed edge (u, v) means that person u sent an e-mail to person v.

As widely done in literature, we assigned the ground-truth influence probabilities according to the weighted cascade model, that is, $p_{u,v} = \frac{1}{|In(v)|}$, for each edge $(u, v) \in E$.

Algorithms For the learning process, we use the Thompson Sampling (TS) as principal exploration strategy (Algorithm 1). For the sake of completeness, we show also results with a Pure Exploitation (PE) approach, in which the oracle is fed with the mean estimates of the influence probabilities at each round.

5.2 Results

Experiment 1 In this experiment, we test the algorithms on Email over a time horizon of $T = 100$ rounds. The objective is to show the performances of the algorithms. The results have been averaged over 30 runs. Figure 5.1 shows the cumulative regret, with a 95% confidence interval.

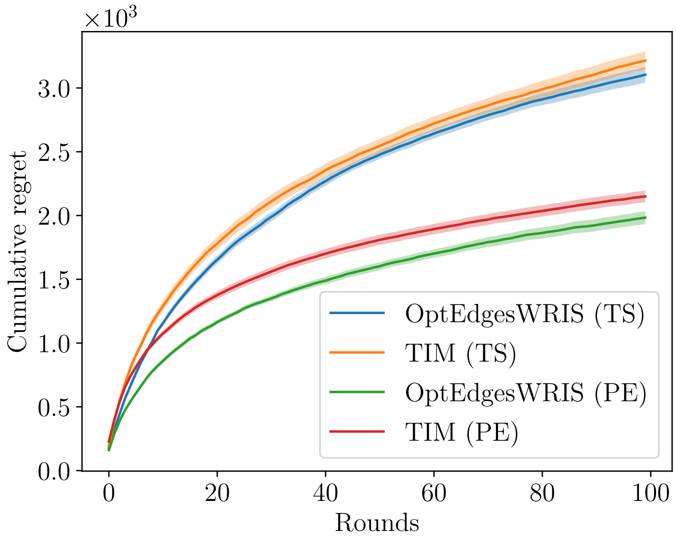

Figure 5.1: Cumulative regret in Email-In4 with 95% confidence interval.

As shown in the plot, our algorithm performs better with both the exploration strategies. However, the gain on TIM is more evident with the PE strategy, specially in the first rounds.

6 Conclusions and Future Works

In this chapter, you present the conclusions of your thesis and a couple of possible future works to extend your results. First of all, you should briefly repeat the problem you addressed in the thesis. Then, you report your achievements and how they improve the state of the art.

6.1 Conclusions

“In this thesis, we analyzed ..."
“We proposed a new approach that ..."
“We tested this method on ..."
“Reported results show that our proposal outperforms the state of the art method."

6.2 Future works

“There are several appealing paths for future works ... ."
“A possible extension could be to ..."

Notation

A

Notation	Description
G	influence graph
V	set of nodes of G
E	set of edges of G
W	set of influence weights corresponding to each edge in E
$w_{u,v}$	weight of edge (u, v)
n	$\|V\|$, number of nodes
m	$\|E\|$, number of edges

Table A.1: Notation used.

Acknowledgments

Here you can insert optionally the acknowledgments for who had a significant importance for the accomplishment of this goal. These acknowledgments are less formal than the ones at the beginning of the thesis and are not listed in the table of contents.

Bibliography

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

[2] Alessandro Nuara, Francesco Trovò, Dominic Crippa, Nicola Gatti, and Marcello Restelli. Driving exploration by maximum distribution in gaussian process bandits. In *Proceedings of the 19th International Conference on Autonomous Agents and MultiAgent Systems*, pages 948–956, 2020.