

Mínim Conjunt Dominador d'Influència Positiva –Minimum Positive Influence Dominating Set– (MPIDS)

GRAU A Q1 CURS 2021-2022

Departament de Ciències de la Computació
Universitat Politècnica de Catalunya

Resum

Aquest projecte té com a objectiu la implementació d'heurístiques i metaheurístiques per a un problema d'optimització combinatoria i la consegüent experimentació que permetrà comparar els resultats obtinguts pels diferents grups.

El projecte es farà en grups de **4 persones**, formats (en la mesura del possible) per estudiants matriculats al mateix subgrup de problemes. Per formalitzar els grups us heu d'apuntar al fitxer compartit **GRAU-A.grupsProjecte** (seguiu l'enllaç).

El lliurament dels materials demanats a aquest projecte es faran en línia via el **Racó FIB** i s'haurà de dur a terme abans de les 23:59 hores del dia **28 de novembre de 2021**.

En qualsevol moment durant el procés de correcció podríeu ser contactats per part d'algun professor de l'assignatura per tal de resoldre dubtes o fer aclariments sobre el vostre treball.

Totes les comunicacions públiques referents al projecte es duran a terme mitjançant el **canal d'Slack #projecte**. Per a altres qüestions, contacteu privadament per l'Slack amb la professora Maria J. Blesa.

1 Introducció

Amb el creixement continu de les xarxes socials, els individus obtenen una plataforma des de la qual poden difondre les seves opinions i idees sense restriccions i sense adonar-se que permetem que aquestes opinions influeixin en les nostres opinions sociològiques, polítiques i personals. Ara és molt més fàcil emetre una idea que té el potencial de difondre àmpliament la influència positiva o negativa a través d'una xarxa social. En un intent d'abordar la influència en aquest context, recentment va aparèixer la noció de "*conjunt que domina la influència positiva*" [5, 3]. Donat un graf o xarxa $G = (V, E)$ que representa una xarxa social (on V són els actors de la xarxa i les arestes de E representen les relacions entre actors), direm que un conjunt de nodes $D \subseteq V$ domina la influència positiva a G si tots nodes de V tenen un percentatge suficient de veïns a D . Com a percentatge suficient s'acostuma a considerar el 50%, donat que representa la majoria simple.

El corresponent problema d'optimització busca un conjunt d'influència positiva de mida mínima que domini positivament tot el conjunt. En termes tècnics, el problema es pot descriure de la següent forma: **Donat un graf connex i no dirigit $G = (V, E)$, el problema consisteix a trobar un conjunt dominador de vèrtexos $D \subseteq V$ (en anglès, *dominating set*) de cardinalitat mínima, tal que almenys la meitat dels veïns de cadascun dels vèrtexos de G formin part del conjunt dominant D .** Aquest problema rep el nom de *Mínim Conjunt Dominador d'Influència Positiva*, amb acrònim MPIDS, i és un problema NP-hard [7, 8]. Aquest problema també es planteja en altres àrees, com ara el software d'aprenentatge electrònic [9], els negocis en línia [6], o problemes socials relacionats amb alcoholisme, tabaquisme i drogues [7].

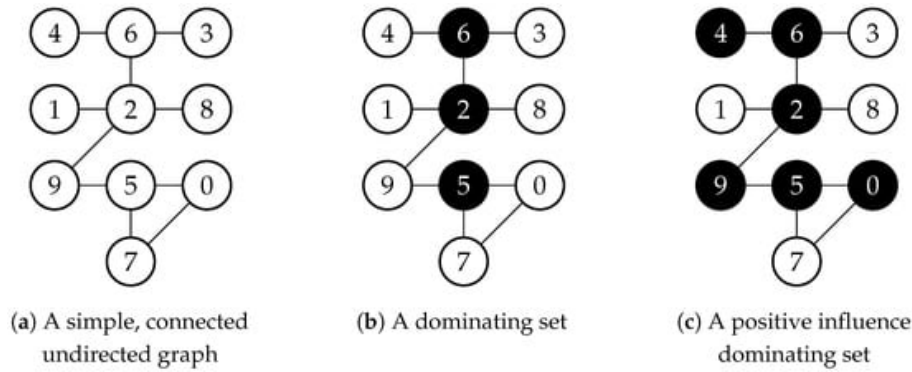


Figura 1: Exemples de (a) Graf G connex, simple i no dirigit; (b) El conjunt $\{2, 5, 6\}$ és dominador a G , però no dominador d'influència positiva; (c) El conjunt $\{0, 2, 4, 5, 6, 9\}$ sí és dominador d'influència positiva a G . (Exemple extret de [1].)

El problema del MPIDS es pot plantejar fàcilment en termes de programació lineal entera (ILP). Considerem una variable binària x_i associada a cada vèrtex $v_i \in V$ que indica l'elecció o no del vèrtex en qüestió per pertànyer al conjunt dominador. Aleshores, el problema de MPIDS es pot formalitzar de la següent forma:

$$\text{Minimitzar} \quad \sum_{v_i \in V} x_i \quad (1)$$

$$\text{Subjecte a} \quad \sum_{v_j \in N(v_i)} x_j \geq \left\lceil \frac{\deg(v_i)}{2} \right\rceil \quad \forall v_i \in V \quad (2)$$

$$x_i \in \{0, 1\} \quad (3)$$

On $N(v_i)$ és el veïnatge de v_i al graf G d'entrada i $\deg(v_i) = |N(v_i)|$ és el seu grau. L'Equació (1) comptabilitza el nombre de vèrtexos de V que pertanyen al conjunt dominador (és el que volem minimitzar); l'Equació (2) assegura que una solució factible conté, com a mínim, la meitat dels veïns de cada vèrtex $v_i \in V$. Finalment, l'Equació (3) indica que les variables x_i són binàries.

2 Objectius

Els continguts d'aquest projecte es poden organitzar en dues parts. A la primera part es demana una resolució del problema d'optimització mitjançant tècniques algorísmiques senzilles. La segona part pretén sofisticar la resolució mitjançant tècniques més complexes per tal d'obtenir solucions de millor qualitat.

Part 1:

La primera part del projecte té un pes del 60% sobre la qualificació total i consta dels següents tres apartats:

- Donat un graf $G = (V, E)$ i un conjunt $D \subseteq V$, comprovar si D és conjunt dominador d'influència positiva a G i, en cas de ser-ho, comprovar si és minimal.

Donat un graf $G = (V, E)$, heu de trobar un bon conjunt dominador d'influència positiva a G fent servir les següents tècniques:

- (b) Dissenyeu i implementeu un algorisme voraç (greedy) determinista.
- (c) Dissenyeu i implementeu un algorisme de cerca local. L'algorisme pot ser determinista (p.ex. hill climbing) o estocàstic (p.ex. simulated annealing).

Part 2:

La segona part contempla el 40% restant de la qualificació i consta dels següents dos apartats:

- (d) Implementeu una metaheurística de la vostra elecció.
- (e) Implementeu el model ILP i resoleu-lo fent servir CPLEX.

Aquests dos darrers apartats només es qualificaran si prèviament s'han realitzat els de la primera part del projecte. Si no es realitzen els apartats de la primera part la qualificació final del projecte serà 0.

Experimentació:

S'ha de dur a terme un estudi experimental dels algorismes proposats. L'objectiu d'aquests experiments és comprovar i comparar la qualitat de les solucions obtingudes amb les diferents tècniques que cadascú implementa, i també comparar amb els resultats obtinguts pels altres grups de classe. Per tal de dur a terme aquestes comparacions de manera correcta, els experiments s'haurien de dur a terme d'una manera estandaritzada (mateix llenguatge de programació, mateix hardware, mateixes configuracions dels experiments, mateixes instàncies, etc.). Donat que sou grups separats i treballareu en condicions diferents aquesta comparació de resultats no la farem tan escrupolosament. Només fixarem les instàncies sobre les quals provar els algorismes i un temps màxim d'execució (o un nombre determinat d'aplicacions). El benchmark d'instàncies que heu de fer servir se us proporcionarà al canal d'Slack **#projecte**, on s'aniran donant també tots els detalls referents a la configuració dels experiments.

Dels grups que facin l'apartat (d) (una metaheurística o més), aquell que obtingui les millors solucions pel problema tindrà 1 punt extra a la seva qualificació (amb una qualificació màxima possible de 10 punts). S'habilitarà un lloc web on podreu anar recopil·lant tots els resultats.

3 Detalls de l'entrega

El nivell de sofisticació i esforç dedicat a la pràctica és opcional i es tindrà en compte a l'hora d'avaluar-la. Tingueu en compte que la documentació entregada ens ha de permetre valorar el nivell d'assoliment de la competència transversal que hem d'avaluar: **Capacitat d'autoaprenentatge**. En el context del projecte hi han molts aspectes rellevants relacionats amb aquesta competència: des de l'estudi de noves tècniques i models algorísmics, fins al disseny i anàlisi d'experiments, i la documentació d'aquests tipus de treballs de recerca. La qualificació final del projecte reflectirà la qualitat del vostre aprenentatge, de l'experimentació feta i de la documentació. La qualitat del codi entregat (programes) es pressuposa i representarà una part molt petita de la qualificació final.

La documentació ha de recollir i presentar la feina feta, les fonts que s'han consultat, el que heu après i els resultats de l'experimentació. En particular és molt important que reflecteixi

de forma succinta el que heu après. Si no es compleix aquesta condició, la qualificació final del projecte reflectirà la qualitat de la presentació i no es tindrà en consideració la resta de material lliurat.

Què cal lliurar

Caldrà lliurar una carpeta comprimida (.zip) que contingui, d'una banda, els programes corresponents:

- Per a la part 1:
 - Un fitxer `basics.<EXTENSIÓ>`¹ per l'apartat (a).
 - Un fitxer `greedy.<EXTENSIÓ>` per l'apartat (b).
 - Un fitxer `local_search.<EXTENSIÓ>` per l'apartat (c). Si implementeu més d'una versió de cerca local, poseu-les totes en aquest fitxer. Podeu implementar una cerca local determinista que segueixi el criteri *first-improvement* i/o una que segueixi el criteri *best-improvement*. També podeu implementar una cerca local estocàstica (p.ex., un simulated annealing). Per qualsevol dels casos, penseu que podeu generar la solució inicial de manera aleatòria o aprofitar l'algorisme golafre realitzat a l'apartat (b) per generar-ne una de millor.
 - Altres fitxers addicionals necessaris per a la compilació i execució.
- Per a la part 2:
 - Un fitxer `metaheuristic.<EXTENSIÓ>` per l'apartat (d). Observeu que per a implementar una metaheurística podeu aprofitar l'algorisme golafre implementat a l'apartat (b) i/o la cerca local implementada a l'apartat (c).
 - Una carpeta `ILP_CPLEX` on s'incloguin tots els fitxers necessaris per implementar el model ILP del problema MPIDS per a ésser executat a CPLEX (apartat (e)). Donant el temps necessari, el CPLEX us resoldrà el problema de manera òptima.
 - Altres fitxers addicionals necessaris per a la compilació i execució.

Apart dels codis dels programes s'ha d'incloure també un informe sobre la feina realitzada. Aquest informe ha de documentar acuradament els algorismes desenvolupats i els experiments realitzats. La qualitat d'aquest informe és un factor molt important per a la qualificació final del projecte.

4 Alguns punters

Aquest document és intencionadament vague i s'espera que investigueu pel vostre compte totes les tècniques algorísmiques i models que es mencionen aquí i que, per manca de temps, no veiem aquest quadrimestre a l'assignatura de GRAU-A. Hi ha bibliografia magnífica al respecte i no us costarà gens trobar-ne informació. A mode d'ajuda inicial (i apart dels llibres [2, 4] de la bibliografia bàsica de l'assignatura), us mencionem aquí alguns punters d'informació que us poden ser d'utilitat per començar la vostra recerca:

- S. Kirkpatrick, C. D. Gelatt, and M. P. Vecchi. Optimization by simulated annealing. *Science*, 220(4598):671–680, 1983.

¹L'<EXTENSIÓ> dependrà del llenguatge de programació utilitzat.

- Jon Kleinberg and Éva Tardos. *Algorithm Design*. Chp 12: Local Search. Pearson/Addison-Wesley, 2006.
- Christian Blum and Andrea Roli. Metaheuristics in combinatorial optimization: Overview and conceptual comparison. *ACM Computing Surveys*, 35(3):268–308, 2003.
- Thomas H. Cormen, Charles E. Leiserson, Ronald L. Rivest, and Clifford Stein. *Introduction to Algorithms, 3rd Edition*. Chp 29: Linear Programming. MIT Press, 2009.

Us adrecem també a treballs recents sobre el problema MPIDS molt relacionats amb aquest projecte i que us poden servir d'inspiració. Tingueu en compte, però, que el que se us demana a vosaltres és quelcom molt més senzill.

- Feng Wang, Hongwei Du, Erika Camacho, Kuai Xu, Wonjun Lee, Yan Shi, and Shan Shan. On positive influence dominating sets in social networks. *Theoretical Computer Science*, 412(3):265–269, 2011.
- Hassan Raei, Nasser Yazdani, and Masoud Asadpour. A new algorithm for positive influence dominating set in social networks. In *2012 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining*, pages 253–257. IEEE, 2012.
- Ruizhi Li, Shuli Hu, Huan Liu, Ruiting Li, Dantong Ouyang, and Minghao Yin. Multi-start local search algorithm for the minimum connected dominating set problems. *Mathematics*, 7(12):article number 1173, 2019.
- Jiehui Pan and Tian-Ming Bu. A fast greedy algorithm for finding minimum positive influence dominating sets in social networks. In *IEEE INFOCOM 2019-IEEE Conference on Computer Communications Workshops*, pages 360–364. IEEE, 2019.
- Yupeng Zhou, Jinshu Li, Yang Liu, Shuai Lv, Yong Lai, and Jianan Wang. Improved memetic algorithm for solving the minimum weight vertex independent dominating set. *Mathematics*, 8(7):article number 1155, 2020.
- Salim Bouamama and Christian Blum. An improved greedy heuristic for the minimum positive influence dominating set problem in social networks. *Algorithms*, 14(3):79, 2021.

Marc de software

Podeu fer els vostres programes en el llenguatge de programació que més us agradi, sempre i quan mantingueu l'estructura de fitxers demanada a l'entrega. A mode d'exemple, i per facilitar parcialment la posterior comparació de resultats, us proporcionarem uns fitxers plantilla en C++ que incorporen la lectura d'instàncies i la impressió de resultats. **Vosaltres podeu decidir lliurement si els feu servir o no, i com.** Com hem dit abans, donat que no farem una comparació acurada ni estadísticament significativa dels resultats podem permetre heterogeneïtat a les implementacions. Aquestes plantilles es proporcionaran via el canal d'Slack.

IBM ja ha tret la versió 20.1 de CPLEX que disposa d'un entorn gràfic. Com és habitual, es posen versions simplificades a disponibilitat de la comunitat científica de manera gratuïta. Nosaltres no necessitem l'entorn gràfic i tenim prou amb versions anteriors, per això us proporcionem (també via Slack) el fitxer `ILOG.tgz` que conté la versió 12.2 de CPLEX per a ésser utilitzada des de programes en C++. Teniu total llibertat per fer servir versions més noves si així ho decidiu.

Referències

- [1] Salim Bouamama and Christian Blum. An improved greedy heuristic for the minimum positive influence dominating set problem in social networks. *Algorithms*, 14(3):79, 2021.
- [2] Thomas H. Cormen, Charles E. Leiserson, Ronald L. Rivest, and Clifford Stein. *Introduction to Algorithms, 3rd Edition*. MIT Press, 2009.
- [3] Dilek Güneş, Subramanian Raghavan, and Rui Zhang. Least-cost influence maximization on social networks. *INFORMS Journal on Computing*, 32(2):289–302, 2020.
- [4] Jon Kleinberg and Éva Tardos. *Algorithm Design, Pearson New International Edition*. Pearson Education Limited, 2013.
- [5] Cheng Long and Raymond Chi-Wing Wong. Minimizing seed set for viral marketing. In *2011 IEEE 11th International Conference on Data Mining*, pages 427–436. IEEE press, 2011.
- [6] Amir Afrasiabi Rad and Morad Benyoucef. Towards detecting influential users in social networks. In *International Conference on E-Technologies*, pages 227–240. Springer, 2011.
- [7] Feng Wang, Erika Camacho, and Kuai Xu. Positive influence dominating set in online social networks. In *International Conference on Combinatorial Optimization and Applications*, pages 313–321. Springer, 2009.
- [8] Feng Wang, Hongwei Du, Erika Camacho, Kuai Xu, Wonjun Lee, Yan Shi, and Shan Shan. On positive influence dominating sets in social networks. *Theoretical Computer Science*, 412(3):265–269, 2011.
- [9] Guangyuan Wang. *Domination problems in social networks*. PhD thesis, University of Southern Queensland, 2014.