Raport ştiinţific final

(2020 - 2022)

Competiţia:	**Proiect experimental demonstrativ - PED 2019**
Nr. contract:	287PED/2020
Cod proiect:	**PN-III-P2-2.1-PED-2019-3248**
Domeniul de cercetare:	Tehnologia informaţiei şi a comunicaţiilor, spaţiu şi securitate
Titlul :	Graphomaly - bibliotecă software pentru detectarea anomaliilor în grafuri ce modelează tranzacții financiare
Acronim:	Graphomaly
Dată începere proiect:	3/08/2020
Dată finalizare proiect:	28/04/2022
Durata (luni):	20 luni
Buget total:	669 825 lei
Sursa 1 Bugetul de stat	600 000 lei
Sursa 2 Alte surse atrase (cofinanţare):	69 825 lei
Pagina web proiect:	http://graphomaly.upb.ro/
Instituţia coordonatoare:	Universitatea Politehnica București
Director de proiect:	Bogdan Dumitrescu
Partener 1 proiect (P1):	Universitatea București
Partener 2 proiect (P2):	Tremend Software Consulting SRL

1. **Prezentare generală a realizării obiectivelor proiectului, cu punerea în evidență a rezultatelor și gradul de realizare a obiectivelor. Prezentarea trebuie să includă explicații care să justifice diferențele (dacă există) dintre activitățile preconizate și cele realizate**

Obiectivele principale ale proiectului, așa cum reies din propunerea inițială, au fost:

- O1. Implementarea unei biblioteci Python pentru detecția de anomalii într-o listă de tranzacții bancare.
- O2. Dezvoltarea de noi algoritmi
- O3. Testare și validare

O1. Prezentăm în continuare modul și gradul de realizare a obiectivelor, cu referire și la detalierea obiectivelor descrisă în propunere.

Din punct de vedere funcțional, biblioteca Graphomaly a fost implementată conform modelului Scikit-learn [SciLe11], pentru interoperabilitate cu alte pachete din domeniul învățării

automate. Modelele și algoritmii sunt integrați în clase, cu implementarea metodelor tipice fit(), transform(), predict(). Biblioteca este publică [Glib], se instalează transparent cu comanda uzuală *pip install graphomaly* și se utilizează în programe Python cu *import graphomaly*. Documentația se găsește la [Gdoc]. Modul de lucru este deci standard, profesional și facil.

Similar, a fost implementată în Python și biblioteca dictlearn [DLlib] pentru învățarea dicționarelor, o dezvoltare a programelor ce însoțesc cartea [DI18], utilizată și pentru detecția de anomalii. Documentația se găsește la [DLdoc]. Dictlearn este utilizată în Graphomaly pentru detecția de anomalii, împreună cu alte tipuri de metode.

Conținutul bibliotecii Graphomaly și celelalte programe realizate dar neintegrate în bibliotecă (acestea se găsesc la https://gitlab.com/unibuc/graphomaly/aml) se încadrează în așteptările și planurile din propunerea de proiect. Descriem în continuare cele cinci categorii de algoritmi pe care le-am propus inițial pentru dezvoltare și modul de realizare.

- *Preprocesare*. Au fost implementate transformările din listă de tranzacții în graf și din oricare din aceste două formate, în diverse trăsături. Sunt acceptate date în mai multe formate standard (graf Networkx, dataframe Pandas, matrice). Pe lângă trăsăturile cu elemente inovative, ce vor fi descrise mai târziu și care sunt extrase din structura grafului de tranzacții și din sumele transferate, am implementat și trăsături derivate din informații adiționale, precum data și ora tranzacției, trăsături statistice și trăsături legate de evoluția istorică a unor câmpuri.
- *Detecție de comunități*. A fost realizată o interfață unică pentru mai mulți algoritmi de detecție de comunități, din biblioteci diferite sau cu implementări distincte. Teste cu grafuri artificiale având caracteristici inspirate din tranzacțiile bancare au permis selectarea unor algoritmi (cei mai buni fiind Louvain [BGL08] și EdMot [LHWL19]) și eliminarea altora.
- *Detecție de anomalii cu structură cunoscută*. Cele mai bune rezultate au fost obținute prin extragerea de trăsături inspirate din structuri de graf cunoscute pentru similaritatea cu cele din cazurile de spălare de bani (clici, inele, stele), dar și din analiza datelor reale oferite de Libra Internet Bank. Trăsăturile de tip egonet, egonet redus sau random walk au totuși un caracter general. Ele au fost implementate în biblioteca Graphomaly, împreună cu diverse trăsături derivate, în număr total de peste 40.
- *Detecție de anomalii cu structură necunoscută*. Detecția de anomalii propriu-zisă a fost implementată în Graphomaly printr-o interfață unică spre i) metode deja existente, în special cele din PyOD [ZNL19]; ii) o metodă de tip învățare de dicționare (DL – dictionary learning) [DI18]. Alte metode DL au fost propuse și implementate [IADu22a] și vor fi integrate în bibliotecă.
- Combinații de metode. A fost implementată o metodă de votare generală, în care se descriu metodele de detecție de anomalii folosite și criteriul de performanță luat în considerare. De asemenea, este facilitată optimizarea parametrilor prin metode *grid search*.

Grad de realizare al O1: complet. Chiar dacă există mici diferențe în lista de algoritmi propusă inițial și cea implementată, numărul și importanța algoritmilor implementați sunt superioare.

O2. Am propus mai mulți algoritmi noi în contextul detecției de anomalii în grafuri sau în general pentru detecția de anomalii. De asemenea, cercetarea la acest proiect a produs rezultate conexe în domeniul DL. Articolele publicate sunt enumerate în Sec.5 (inclusiv cu etichetele pentru citare).

În propunerea de proiect am identificat două direcții de inovare, algoritmii online și cei distribuiți. Algoritmii propuși de noi pentru detecția de anomalii în grafuri se bazează pe extragerea de trăsături. Articolul [DBS22] este cel mai reprezentativ în această privință, trăsăturile fiind de tip egonet, egonet redus și random walk. Toate aceste trăsături se pot calcula independent pentru fiecare nod, deci au un clar caracter distribuit. Variante online ale

acestui algoritm se pot obține relativ simplu, prin recalcularea trăsăturilor nodurilor implicate în tranzacții, dar timpul de calcul depinde de mărimea egoneților.

O altă categorie de algoritmi propuși se referă la DL, care în propunere a fost văzută ca tehnica de detecție de anomalii pe care o vom investiga în primul rând. Detecția de anomalii este subiectul principal în [IRP21], [IADu22a] și [IRSP21], în ultima lucrare fiind vorba explicit despre grafuri. Alte probleme DL sunt abordate și rezolvate în [IADu21], [MiDu21], [RuIr21], [IADu22b].

Grad de realizare al O2: deși la algoritmii online am realizat mai puțin decât am dorit inițial, am compensat prin inovații în detecția de anomalii cu DL și în alți algoritmi DL, publicând mai multe lucrări decât prevăzut în propunere.

O3. Testarea a fost preconizată pe date simulate, publice și disponibile de la partenerul BRD. Am realizat testarea în primul rând pe date obținute de la noul partener Libra Internet Bank, care a început cooperarea cu echipa Graphomaly după startul proiectul, și care a oferit un set foarte mare de date reale acoperind doi ani de tranzacții. Din acestea, am izolat un set acoperind trei luni, adnotat cu două nivele de anomalii numite alerte, pentru tranzacțiile care au necesitat investigații, și raportări, pentru tranzacții care au fost raportate ca suspicioase la ONPCSB (Oficiul Național de Prevenire și Combatere a Spălării Banilor). Deoarece aceste date depășesc ca amploare și specificitate datele publice (care se referă la alte tipuri de fraude, nu la spălarea de bani) și am putut obține explicații suplimentare cu privire la semnificația unor rezultate, ne-am concentrat în mod natural pe ele. Am folosit și date artificiale cu anomalii specifice spălării de bani, precum și date de la partenerul BRD într-o aplicație separată de identificare de fraude.

Considerăm că testarea pe aceste date a validat metodele propuse, care sunt robuste și funcționează bine, depășind unele metode consacrate, atât pe datele reale de la Libra cât și pe cele artificiale.

Grad de realizare al O3: complet. Considerăm că testarea și-a atins scopul, validând algoritmii propuși cu date suficient de diverse, atât reale cât și artificiale.

2. Prezentarea şi argumentarea nivelului de maturitate tehnologică (TRL) la finalul proiectului

TRL propus în proiect a fost 4, tehnologie validată în laborator. Considerăm că acest nivel a fost atins, argumentele fiind următoarele:

- Biblioteca Graphomaly, realizată în Python și având un caracter deschis, are acum toate componentele care permit detecția de anomalii într-o listă de tranzacții bancare: transformare din listă în graf, extragere de trăsături (din graf sau din lista de tranzacții), detecția propriu-zisă.
- Datorită numărului mare de trăsături calculate, sunt posibile multe combinații, în funcție și de specificul anomaliilor căutate. Am găsit deja combinații de trăsături cu rezultate foarte bune în detectarea anomaliilor tipice spălării de bani.
- Interfața are structura binecunoscutului pachet scikit-learn, extrem de popular în comunitatea machine learning, deci utilizarea modulelor bibliotecii este facilă pentru cei cu minimă experiență în domeniu.
- Biblioteca a fost testată pe date reale, anume un set de tranzacții pe trei luni de la Libra Internet Bank, obținând rezultate bune în privința calității, în timpi de execuție adecvați utilizării zilnice.
- Testarea a fost realizată și pe date simulate în care au fost inserate tranzacții cu structuri tipice de spălare de bani. Rezultatele obținute sunt excelente, comparabile

sau mai bune cu ale unor metode dedicate (care obțin rezultate semnificativ mai slabe pe datele Libra).

Desigur, sunt necesare dezvoltări ulterioare pentru a integra algoritmii dezvoltați în fluxul de prelucrare în timp real al unei bănci. Acestea țin mai degrabă de interfațare și reorganizare a operațiilor de prelucrare decât de algoritmii implementați, care sunt suficienți. Am demonstrat deja că algoritmii dezvoltați pot rezolva probleme de detecție de operațiuni suspecte de spălare de bani, dar testarea pe alte date bancare decât cele de la Libra (greu de obținut datorită reticenței băncilor de a partaja date reale) este încă necesară pentru a trece la un TRL superior. Totuși, apreciem că trecerea la TRL 5 este aproape și vom încerca păstrarea legăturilor între universități și partenerul industrial pentru a coopera cu Libra Internet Bank la realizarea unui software funcțional online pe datele reale ale băncii, validând astfel tehnologia noastră și într-un mediu relevant.

3. Gradul de atingere a rezultatelor estimate (prezentarea produsului/tehnologiei sau a serviciului rezultat al proiectului)

Scopul principal al proiectului a fost realizarea în limbajul Python a unei biblioteci pentru detecția de anomalii în grafuri ce modelează tranzacții financiare, care să funcționeze atât pe date reale cât și pe date artificiale cu structuri tipice pentru spălarea de bani.

Biblioteca se numește Graphomaly și a fost implementată complet. Sursele sunt publice [Glib], la fel și documentația [Gdoc]. Modul de lucru este ilustrat în Figura 1 și constă în transformări ale datelor (listă de tranzacții sau graf) pentru obținerea de trăsături specifice, pe care se aplică funcții de detecție de anomalii.

Modulele principale ale bibliotecii sunt:

- Preprocesare: extragere de trăsături. Au fost implementate un număr de 44 de trăsături extrase din graf și 21 extrase direct din lista de tranzacții, la acestea din urmă numărul fiind în fapt mult mai mare, deoarece unele trăsături sunt statistici ce pot fi aplicate oricărui câmp numeric, iar altele au un caracter istoric, cu ferestre de timp la alegere.
- Interfață cu modelele de detecție de anomalii.
- Interfață generală pentru prelucrarea completă a datelor, de la forma primară de listă de tranzacții sau graf, până la metode de votare sau grid search (pentru optimizarea parametrilor).

Biblioteca este structurată astfel încât să permită utilizarea facilă de către o persoană, de preferat familiarizată cu modelul scikit-learn, fără a cunoaște prelucrările efectuate în interior, dar și de un specialist care extrage diverse funcții sau module specializate, în cunoștință de cauză.

Elemente de originalitate pot fi găsite în

- forma generală a bibliotecii, prima de acest fel, care să permită detecția de anomalii în grafuri într-un mod simplu și unitar, integrabil în modul de lucru scikit-learn;
- unii algoritmi de extragere de trăsături, în particular cele de tip egonet redus, care nu sunt menționate în literatura de specialitate.

Testele prezentate în rapoartele pe 2021 și 2022, și în special în [DBS22], arată un comportament foarte bun pe date reale și artificiale.

Biblioteca Graphomaly conține exemple de utilizare, sub formă de programe Python sau notebooks, care ilustrează funcționarea pe date la alegerea utilizatorului sau luate de pe site-ul http://graphomaly.upb.ro/, unde se găsesc datele Libra și date artificiale, precum și programe de generare a acestora.

Considerăm că forma actuală a bibliotecii Graphomaly corespunde planului inițial și chiar îl depășește. În plus, structura bibliotecii permite dezvoltarea prin adăugarea de module. De exemplu, adăugarea unor funcții care generează noi trăsături se poate face extrem de simplu. Sperăm deci că și alți dezvoltatori vor fi atrași și că biblioteca se va îmbogăți pe măsură ce va fi utilizată.

4. Impactul rezultatelor obținute, cu sublinierea celui mai semnificativ rezultat obținut

Din punctul de vedere al publicațiilor, am reușit:

- Publicarea unui articol [DBS22] într-o revistă Q2, IEEE Access, ceea ce va asigura o bună vizibilitate a rezultatelor descrise acolo, anume detecția de anomalii în grafuri folosind trăsături de tip egonet și egonet redus.
- Publicarea a două articole în conferințele generale cele mai reprezentative în prelucrarea semnalelor, ICASSP [IRP21] și EUSIPCO [MiDu21].

Pe lângă articolul în IEEE Access, cel mai reprezentativ rezultat este chiar biblioteca Graphomaly, descrisă în secțiunea anterioară, care a fost făcută publică de curând. Pentru a ilustra impactul ei, prezentăm un exemplu reprezentativ de rezultate numerice obținute pentru o selecție de trei luni din datele Libra. Figura 3 prezintă curba True Positive Rate (TPR) pentru trei seturi de trăsături, contribuția noastră fiind în trăsăturile de tip egored (egonet redus) și random walk. Reamintim că TPR reprezintă numărul de anomalii găsite raportat la numărul total de anomalii, nodurile fiind ordonate în ordine descrescătoare a scorurilor de anormalitate furnizate de un algoritm general de detecție de anomalii pe setul de trăsături ales.

Graful corespunzător datelor Libra are 385100 noduri, dintre care 600 sunt adnotate ca alerte (tranzacții posibil suspicioase) și 15 ca raportări (tranzacții suspicioase după investigația de către un expert). Figura 1 prezintă TPR pentru primele 1% noduri ordonate după scorul de anormalitate de Isolation Forest [LTZ08]. Observăm, de exemplu, că TPR pentru primele 500 de noduri este aproximativ 0.45, ceea ce înseamnă că sunt găsite aproximativ 270 de alerte, ceea ce înseamnă un număr foarte rezonabil de fals pozitive. Subliniem că investigarea de către un expert a clienților cu tranzacții suspicioase este consumatoare de timp pentru lucrători extrem de calificați, deci este important ca numărul de fals pozitive să nu fie prea mare.

Pe lângă rezultatele numerice, trebuie să ținem seama de două considerente calitative importante în aprecierea algoritmilor noștri. Acești algoritmi trebuie văzuți ca un complement, nu un substitut al algoritmilor existenți bazați pe reguli și deci:

- Nodurile cu scor mare de anormalitate, dar deocamdată fals pozitive, pot fi totuși demne de investigat, deoarece pot semnala conexiuni suspecte în graf care nu pot fi identificate de reguli.
- Nu dorim ca TPR să ajungă repede la valoarea optimă 1, adică nu vrem să emulăm cât mai bine regulile prin metoda noastră. Considerăm că găsirea a 80% din alerte în primele 1% noduri suspecte este un compromis rezonabil între adecvarea la specificul anomaliilor din datele Libra și buna modelare a structurilor suspecte de graf, ignorând deci anomaliile ce rezultă din tranzacții izolate.

Mai multe detalii despre rezultatele pe datele Libra pot fi găsite în [DBS22].

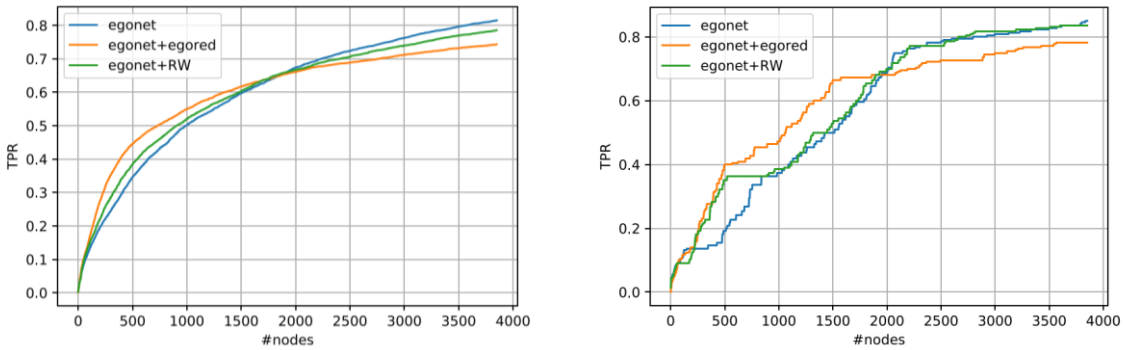

Figura 1. Curba TPR pentru datele Libra: alerte (stânga) și raportări (dreapta)

5. Detalii privind exploatarea și diseminarea rezultatelor proiectului

În cadrul proiectului au fost publicate sau trimise spre publicare articolele următoare.

Articole în reviste:

- [DBS22] B.Dumitrescu, A.Băltoiu, Ş.Budulan, "Anomaly Detection in Graphs of Bank Transactions for Anti Money Laundering Applications", IEEE Access, accepted April 2022 (Q2).
- [PaIr22] A.Pătraşcu, P.Irofti, "Computational complexity of Inexact Proximal Point Algorithm for Convex Optimization under Holderian Growth", https://arxiv.org/abs/2108.04482, to be submitted.

Lucrări la conferințe (primele 5 acceptate, celelalte 2 trimise):

- [IADu21] B.Dumitrescu, D.Ilie-Ablachim, "Classification with Incoherent Kernel Dictionary Learning", Int. Conf. Control Systems and Computer Science, Bucharest, May 2021.
- [MiDu21] F.I.Miertoiu, B.Dumitrescu, "Shape Parameter and Sparse Representation Recovery under Generalized Gaussian Noise", EUSIPCO, Dublin, Ireland, Aug. 2021.
- [RuIr21] C.Rusu, P.Irofti, "Efficient and Parallel Separable Dictionary Learning", IEEE 27th International Conference on Parallel and Distributed Systems (ICPADS), Beijing, China, Dec. 2021.
- [IRP21] P.Irofti, C.Rusu, A.Pătraşcu, "Dictionary Learning with Uniform Sparse Representations for Anomaly Detection", ICASSP, 2022.
- [IADu22a] D.C.Ilie-Ablachim, B.Dumitrescu, "Anomaly Detection with Selective Dictionary Learning", 8th Int. Conf. Control, Decision and Information Technology (CoDIT), Istanbul, Turkey, May 2022.
- [IRSP21] P.Irofti, L.Romero-Ben, F.Stoican, V.Puig, "Data-driven Leak Localization in Water Distribution Networks via Dictionary Learning and Graph-based Interpolation", submitted to IEEE Conference on Control Technology and Applications, Trieste, Italy, Aug. 2022.
- [IADu22b] D.C.Ilie-Ablachim, B.Dumitrescu, "Reduced Kernel Dictionary Learning", submitted to 30th European Signal Processing Conference (EUSIPCO), Belgrade, Serbia, Aug. 2022.

Pe lângă aceste articole, dorim realizarea unei lucrări care să descrie biblioteca Graphomaly, cu accent pe structura software și modul de lucru. Vom putea astfel face mai multă publicitate bibliotecii și deci atrage noi utilizatori. Lucrarea va fi realizată în cursul anului 2022.

O componentă importantă a diseminării este cea software, dată de bibliotecile Graphomaly și dictlearn, care sunt publice, ușor accesibile și bine documentate, așa cum am explicat în Sec.1 și 3. Ele pot fi folosite izolat sau în cadrul unor lanțuri de procesare conținând alte operații asupra datelor.

Colaborarea cu băncile Libra și BRD a dus la cunoașterea de către acestea a tehnicilor propuse de noi pentru datele furnizate de bănci, precum și la validarea unor rezultatele. Este de așteptat ca această colaborare să continue în viitor, canalele de comunicare universitate–dezvoltator software–bancă fiind acum stabilite și funcționale. Proiecte viitoare vor urma pe această linie. Ne așteptăm la integrarea în verificările AML ale Libra a unor rezultate obținute în acest proiect, plecând de la implementările disponibile în biblioteca Graphomaly [Glib].

6. Prezentarea livrabilelor/indicatorilor obținuți la finalul proiectului comparativ cu cei propuşi

Nr. crt.	Livrabile/indicatori planificați	Nr.	Livrabile/indicatori realizați	Nr.
1.	Bibliotecă Python pentru detecția de anomalii în grafuri	1	Biblioteca Graphomaly: surse publice [Glib], documentație [Gdoc]	1
2.	Articole în reviste	1	Acceptat: [DBS22] (Q2) Trimis: [PaIr22]	1+1
3.	Articole la conferințe nivel ICASSP, EUSIPCO	1	Acceptate: [MiDu21], [IRP21]. Trimise: [IADu22b]	2+1
4.	Articole la alte conferințe	2	Acceptate: [IADu21], [RuIr21], [IADu22a] Trimise: [IRSP21]	3+1

Menționăm în plus că biblioteca Graphomaly, socotită în tabel ca un singur livrabil, conține mulți algoritmi și posibilitatea alegerii lor după cunoștințe și eventual intuiție. Structura ei, descrisă în secțiunile 1 și 3, este diversă, complexă și superioară viziunii noastre din propunerea de proiect.

Prezentare succintă a rezultatelor

Proiectul Graphomaly a avut ca scop principal producerea unei biblioteci de programe pentru *detecția anomaliilor în grafuri* ce modelează tranzacții financiare; prin anomalii înțelegem, în primul rând, posibile tranzacții frauduloase cu caracter complex, cum ar fi spălarea de bani, care se manifestă prin structuri tipice în graful de tranzacții.

Biblioteca a fost realizată în limbajul Python. Structura prelucrărilor este ilustrată în Figura F1. Lista tranzacțiilor bancare dintr-o perioadă de timp este transformată în graf, nodurile fiind clienții, iar arcele tranzacțiile cumulate. Din graf se extrag trăsături, cu metode inovative ce constituie aportul teoretic al proiectului. Un succes deosebit l-au avut trăsăturile de tip egonet (graful format de vecinii unui nod central) și egonet redus (egonet din care sunt eliminate nodurile care au doar o conexiune, cea cu centrul). Figura F2 ilustrează aceste concepte pentru un caz real. Nodul roșu este centrul egonetului, iar tranzacțiile dintre el și nodul portocaliu sunt suspecte de fraudă. După eliminarea nodurilor verzi se obține egonetul redus. Ansamblul trăsăturilor, cuprinzând printre altele densitatea arcelor, sume transferate, număr de conexiuni ale nodului, pentru egonet și egonetul redus, permite detecția, cu metode de învățare automată, a anomaliilor.

Validarea experimentală a bibliotecii Graphomaly a fost efectuată pe două tipuri de date: i) date reale furnizate de Libra Internet Bank; ii) date sintetice, constând din grafuri aleatoare în care sunt inserate structuri tipice de anomalii (clici, inele, stele). Rezultatele sunt foarte bune, pe ambele tipuri de date, spre deosebire de alte metode, care performează doar în anume condiții, în special pe date artificiale. Timpii de execuție pe date mari, adică grafuri cu sute de mii de noduri, sunt adecvați utilizării în aplicații bancare. Având în vedere și structura software a bibliotecii, similară unor pachete populare în învățarea automată și deci ușor de utilizat pentru cei specializați în acest domeniu, putem afirma că am produs un pachet de programe eficient, flexibil și robust.

TRANSACTIONS LIST

ID Ordinator	ID Beneficiary	Amount	Currency	Date	Transaction type
2738499	2873940	200	EUR	12/03/21	PX
7749302	8385216	44100	RON	12/03/21	CV
2873940	4537582	679	RON	12/03/21	CV
5337271	7749302	1000	RON	12/03/21	AA
7749302	4537582	20	RON	12/03/21	CV
3121912	8385216	380	EUR	12/03/21	PX

GRAPH

FEATURES

ID	Degree IN	Degree OUT	Amount IN	Amount OUT
2738499	0	1	0	200
2873940	1	1	200	679
3121912	0	1	0	380
4537582	2	0	699	0
5337271	0	1	0	1000
7749302	1	2	1000	44120
8385216	2	0	44480	0

ANOMALY

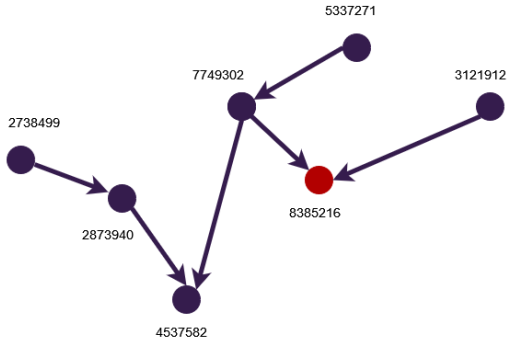

Figura F1. Schema prelucrărilor în biblioteca Graphomaly

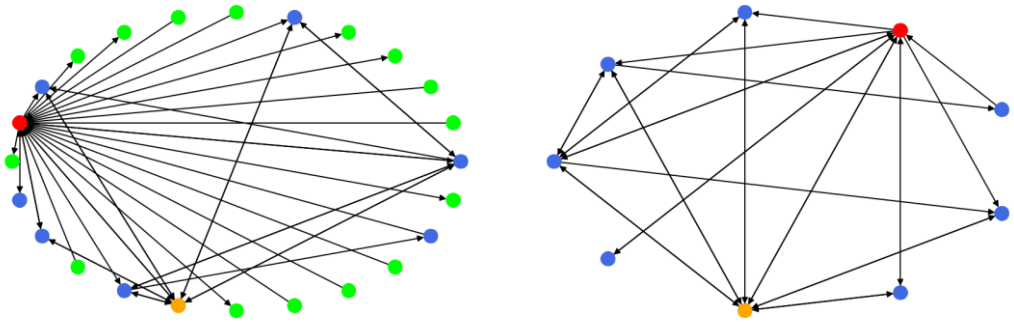

Figura F2. Egonet (stânga) și egonet redus (dreapta) pentru tranzacții frauduloase reale

Bibliografie

[AMF10] L. Akoglu, M. McGlohon, C. Faloutsos. “Oddball: Spotting anomalies in weighted graphs”. In: Pacific-Asia Conference on Knowledge Discovery and Data Mining, 410–421, 2010.

[BGL08] V. D. Blondel, J.-L. Guillaume, R. Lambiotte, E. Lefebvre. Fast unfolding of communities in large networks, J. Stat. Mech., 10, P10008, Oct. 2008.

[DLdoc] Dictlearn documentation, https://unibuc.gitlab.io/graphomaly/dictionary-learning/

[DLlib] Dictlearn library, https://gitlab.com/unibuc/graphomaly/dictionary-learning

[DI18] B.Dumitrescu, P.Irofti. Dictionary Learning Algorithms and Applications, Springer, 2018.

[Ell19] A. Elliott et al. Anomaly detection in networks with application to financial transaction networks. *arXiv:1901.00402*, 2019.

[Gdoc] Graphomaly documentation, https://unibuc.gitlab.io/graphomaly/graphomaly/

[Glib] Graphomaly library, https://gitlab.com/unibuc/graphomaly/graphomaly

[LHWL19] P.-Z. Li, L. Huang, C.-D. Wang, J.-H. Lai. EdMot: An Edge Enhancement Approach for Motif-aware Community Detection,” Proc. 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, 479–487, 2019.

[LTZ08] F.T. Liu, K.M. Ting, Z.H. Zhou, Isolation Forest. In *8th IEEE International Conference on Data Mining*, pp. 413-422, Dec. 2008.

[SciLe11] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas. Scikit-learn: Machine learning in Python. The Journal of Machine Learning Research, 12, 2825-2830, 2011.

[ZNL19] Y. Zhao, Z. Nasrullah and Z. Li, (2019). *PyOD: A Python Toolbox for Scalable Outlier Detection*. Journal of Machine Learning Research (JMLR), 20(96), pp.1-7.

Director de proiect
prof. Bogdan Dumitrescu

Data: 28.04.2022