



پژوهشکده پولی و بانکی بانک مرکزی جمهوری اسلامی ایران
کتاب گزلی بررسی مسائل امنیتی شرکت ملی انفورماتیک

یازدهمین همایش سالانه
بانکداری نوین و
نظام‌های پرداخت

زیست بوم بانکداری هوشمند؛ حکمرانی دیجیتال و نظارت فناوریانه

تهران، مرکز همایش‌های بین‌المللی برج میلاد - ۵ و ۶ آذر ۱۴۰۳

ISC مؤسسه استنادی و پایش
علم و فناوری جهان اسلام

شناسایی تراکنش‌های مشکوک پول- شویی با استفاده از یادگیری بازنمایی گراف

زهرا نوراله - دکتر محسن رضوانی

الله الرحمن الرحيم

- ۱ بیان مسئله، ضرورت و اهمیت تحقیق، اهداف و پیشینه تحقیق
- ۲ پیشینه پژوهش و کارهای مرتبط
- ۳ روش پیشنهادی
- ۴ مطالعه موردی - شناسایی زیر شبکه‌های مشکوک پولشویی
- ۵ ارزیابی، نتایج و یافته‌های پژوهش
- ۶ جمع بندی

بیان مساله، اهمیت و ضرورت انجام تحقیق

حجم زیاد و تنوع تراکنش‌های مالی موجب دشواری در تشخیص فعالیت‌های مشروع و مشکوک می‌شود.

پیچیده‌تر شدن ساختارهای مالی، کشف تقلب، تشخیص فعالیت‌های پول‌شویی و غیرقانونی مرتبط با آن برای سازمان‌های مالی و نظارتی یک چالش جدی به شمار می‌رود.



حفظ اعتماد مشتریان و
کاهش خطرات قابل
توجه برای مؤسسات
مالی

شناسایی زیرشبکه‌های مشکوک در تراکنش‌های مالی



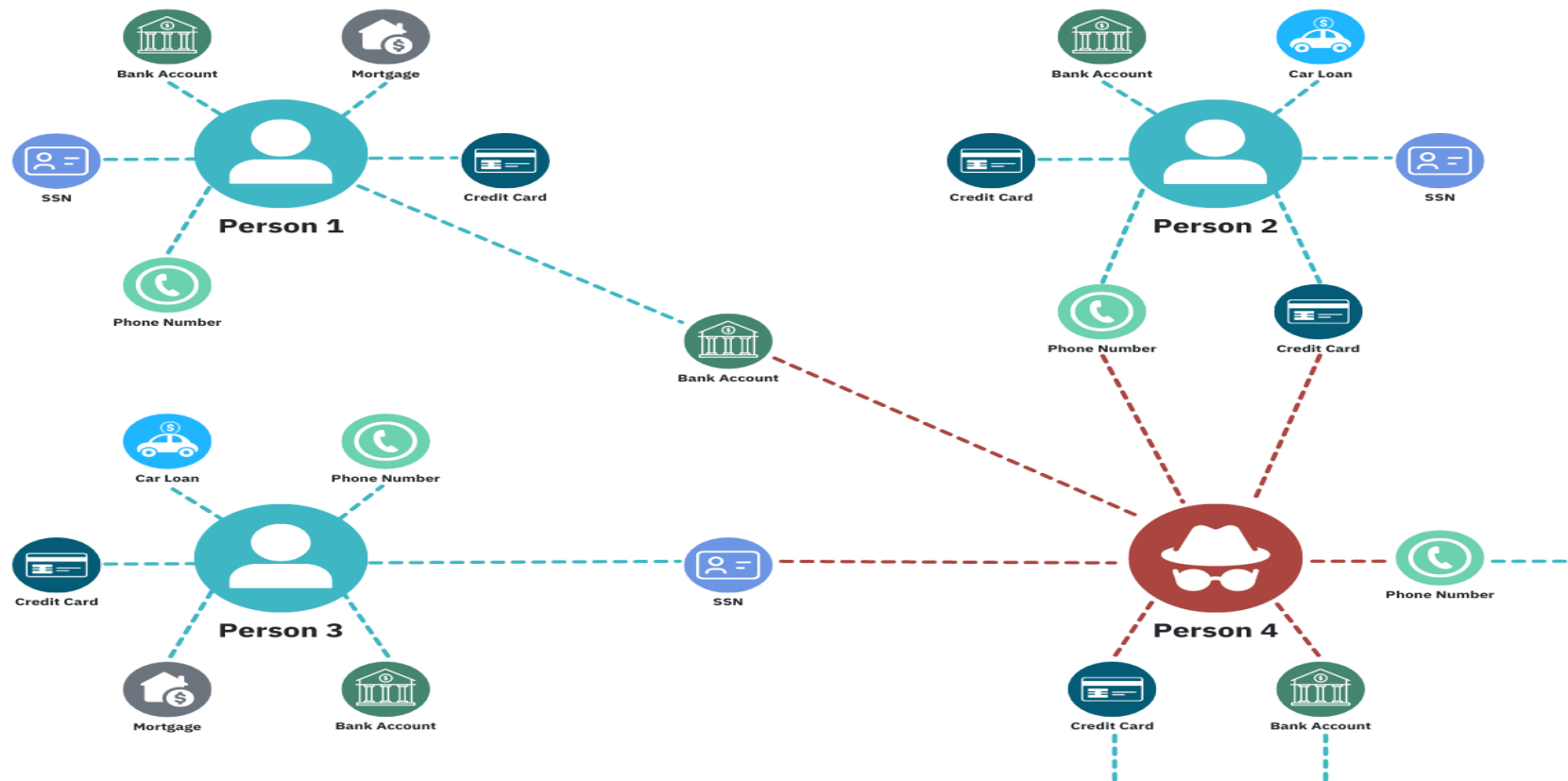
جلوگیری از تقلب و زیان‌های مالی

بیان مساله، اهمیت و ضرورت انجام تحقیق

در زندگی واقعی، ناهنجاری‌ها ممکن است با هم تبانی و رفتار جمعی با یکدیگر برای به دست آوردن منافع داشته باشند.



شناسایی و مدل سازی الگوهای همکاری و همدستی بین متقلبان در زیرگراف داده‌های تراکنش مالی از اهمیت ویژه ای برخوردار می باشد.



بیان مساله، اهمیت و ضرورت انجام تحقیق

فعالیت‌های پول‌شویی به دلیل پیامدهایی که برای اقتصاد و مؤسسات مالی در سراسر جهان دارند، در سطح بین‌المللی مورد توجه زیادی قرار گرفته‌اند.



فعالیت‌های پول‌شویی یکپارچگی سیستم‌های مالی را تضعیف می‌کند. ❁

منجر به از دست دادن کنترل بر سیاست‌های اقتصادی کشور و بی‌ثباتی در سرمایه‌گذاری‌ها می‌شود. ❁

درآمد مالیاتی کمتری نیز برای دولت به همراه دارد [2 و 1]. ❁

رشد روزافزون فعالیت‌های غیرقانونی و تغییرات مداوم در الگوهای آن، نیاز به بهبود و ارتقا ابزارهای تشخیصی را بیش از پیش نشان می‌دهد.

[1] ALUKO, Ayodeji; BAGHERI, Mahmood. The impact of money laundering on economic and financial stability and on political development in developing countries: The case of Nigeria. Journal of Money Laundering Control, 2012.

[2] CHEN, Zhiyuan, et al. Machine learning techniques for anti-money laundering (AML) solutions in suspicious transaction detection: a review. Knowledge and Information Systems, 2018

اهداف و سئولات پژوهش

تشخیص ناهنجاری نیازمند توسعه تکنیک‌های مؤثری است که بتواند بر عوامل مختلفی که در وقوع ناهنجاری‌ها موثر است تأثیرگذار باشد.

در پژوهش‌های پیشین برای شبکه‌های بزرگ مثلاً با میلیاردها گره، نمایش شبکه به صورت گراف با چالش‌هایی در آنالیز و پردازش شبکه مواجه است [1]. مانند:

✓ پیچیدگی محاسباتی بالا

✓ قابلیت موازی‌سازی کم

✓ عدم امکان استفاده از روش‌های یادگیری ماشین



اهداف و سؤالات پژوهش

بهبود الگوریتم های تشخیص ناهنجاری برای دقت تشخیص



شناسایی ویژگی ها و شاخص هایی از زیرگراف ها که بیشترین تاثیر را در تشخیص ناهنجاری مالی دارند.



استفاده از یادگیری عمیق و تکنیک های یادگیری ماشین برای بهبود تشخیص ناهنجاری در زیرگراف های تراکنش های مالی



شناسایی الگوهای تراکنش مشکوک به صورت سازمان یافته و در قالب زیرگراف های پیچیده



پیشینه پژوهش و کارهای مرتبط



پیشینه پژوهش و کارهای مرتبط

چالش‌های احصاء شده براساس مرور ادبیات موضوع

محدودیت‌های مدل‌های
یادگیری ماشین

روش‌های ارزیابی برای
تشخیص ناهنجاری

مقیاس پذیری و کاربرد در
دنیای واقعی

ساخت گراف و تحلیل‌های
مرتبط در سطح گروه

پیچیدگی تعاملات در
سطح گروه

کاربرد محدود تحلیل شبکه
برای درک روابط و تعاملات
پیچیده

Savage et al.(2016)

Baltoiu et al.(2019)

Francisquinia et al.(2022)

Safdari and De Bacco(2022)

LUO et al.(2022)

Sun et al.(2021)

Ai et al.(2024)



تشخیص ناهنجاری
در سطح زیرگراف



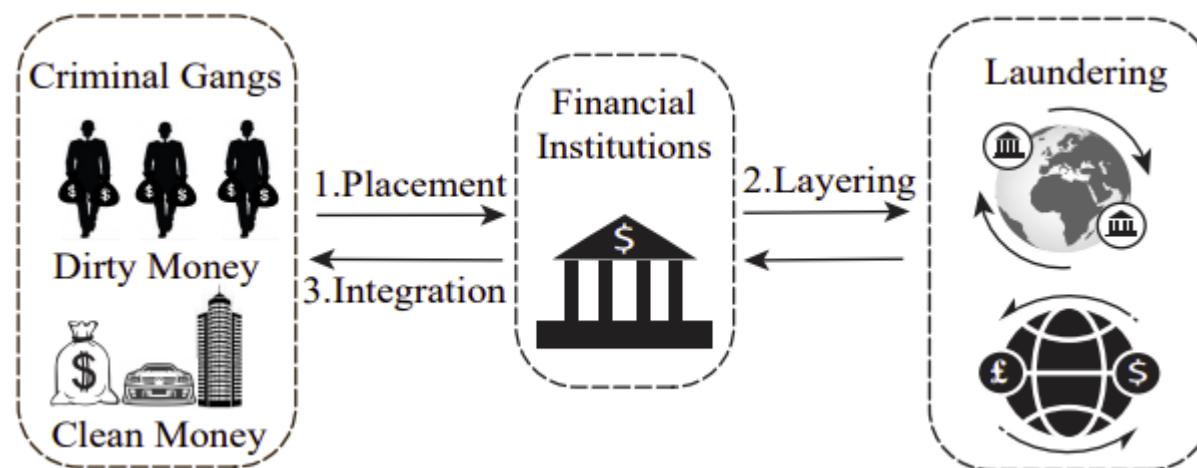
استفاده از رویکرد ادغام
ساختار انجمن در تشخیص
ناهنجاری (بدون نظارت)

استفاده از رویکرد ایجاد گراف
گروه مبتنی بر تجمیع گره
های مشکوک (نظارت شده)

Cheng et al.(2023)

Chen et al.(2023)

پولشویی به فعالیت‌های مربوط به پنهان کردن درآمد واقعی، فرار مالیاتی، یا تبدیل پول به دست آمده غیرقانونی برای استفاده مشروع اشاره دارد [۱].



فرآیند پولشویی شامل سه بخش جانمایی و قرار دادن، لایه‌بندی و یکپارچه سازی می‌باشد [۲].

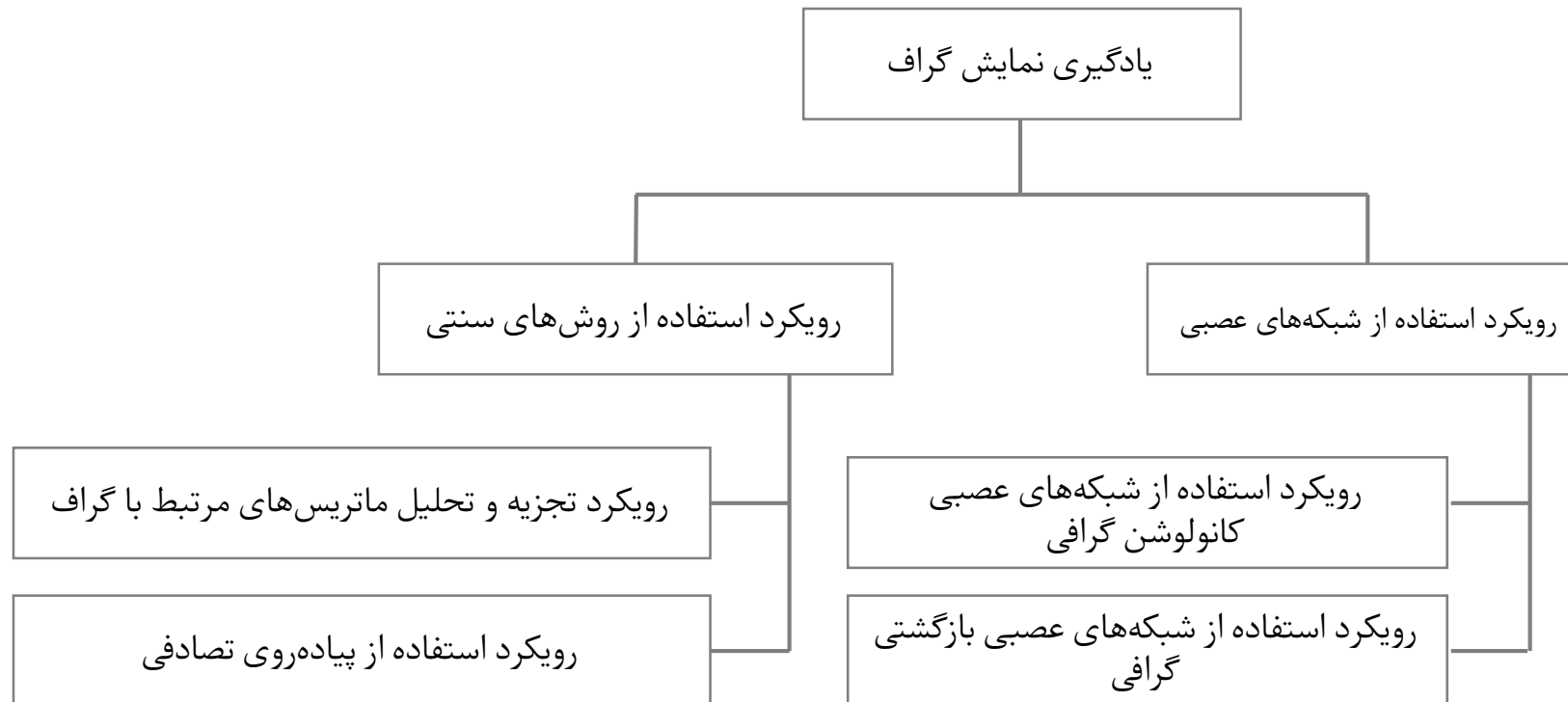
[1] DUMITRESCU, Bogdan; BĂLTOIU, Andra; BUDULAN, Ștefania. Anomaly detection in graphs of bank transactions for anti money laundering applications. IEEE Access, 2022

[2] Chopra S., Hadsell R., Le Cun Y. Learning a Similarity Metric Discriminatively, with Application to Face Verification; Proceedings of the 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05); San Diego, CA, USA. 20–25 June 2005

پیشینه پژوهش و کارهای مرتبط

تحلیل مفید و بهینه اطلاعات شبکه‌های بزرگ، وابستگی زیادی به نحوه‌ی نمایش و بازنمایی داده‌های شبکه دارد.

در فرآیند بازنمایی شبکه که با عنوان تغییر فضای شبکه یاد می‌شود، هدف اصلی، تبدیل رئوس و یال‌ها به یک فضای برداری است تا ویژگی‌ها و ارتباطات بین موجودیت‌ها به صورت عددی قابل استفاده باشند.



پیشینه پژوهش و کارهای مرتبط : نگاشت و یادگیری نمایش گراف

مزایا و معایب رویکردهای یادگیری نمایش گراف با توجه به تحقیقات انجام شده

عنوان دسته بندی	مزایا	معایب
رویکرد استفاده از روش‌های سنتی	✓ در روش‌های فاکتورسازی ساختار کلی گراف در نظر گرفته می‌شود. ✓ روش‌های گام‌های تصادفی در پیدا کردن مشابهت بین گره‌ها نسبتاً بهینه می‌باشند. ✓ بیان موثر روابط پیچیده در داده‌های ساختاریافته گراف، پیاده سازی آن‌ها آسان است. ✓ مقیاس پذیر در برخی از روش‌ها.	✓ قابل تعمیم به گره‌های دیده نشده نیست. ✓ روش‌های فاکتورسازی ماتریس هزینه زمانی و محاسباتی بالا دارد. ✓ روش‌های گام‌های تصادفی فقط ساختارهای محلی را در نظر می‌گیرد.
رویکرد استفاده از شبکه‌های عصبی	✓ آموزش قابل تعمیم به گره‌های دیده نشده است. می‌توانند آموزش را براساس وظیفه خاص و شباهت گره‌ها انجام دهند. ✓ کارایی موثر در اجرا دارند. ✓ غیر خطی بودن روش‌ها.	✓ قدرت تفسیر پذیری پایین و عدم مقیاس پذیری در برخی از روش‌ها. ✓ هزینه زمانی نسبتاً بالا دارند

تحلیل گراف با رویکرد تشخیص انجمن و یادگیری زیرگراف

« ویژگی ساختار انجمن، یکی از مهم‌ترین ویژگی‌های هر گراف می‌باشد. انجمن شامل گروهی از گره‌ها است که اتصالات رأس به رأس متراکم می‌باشد و شامل ویژگی‌های مشابه و الگوهای یکسان هستند و بین این گروه‌ها چگالی کمتری وجود دارد.

« توانایی تشخیص ساختار انجمن در یک شبکه به وضوح می‌تواند کاربردهای عملی داشته باشد.

« تشخیص انجمن، یک وظیفه اساسی برای تجزیه و تحلیل شبکه، با هدف تقسیم یک شبکه به زیرساخت‌های متعدد برای کمک به آشکار کردن عملکردهای نهفته آن‌ها است [1]

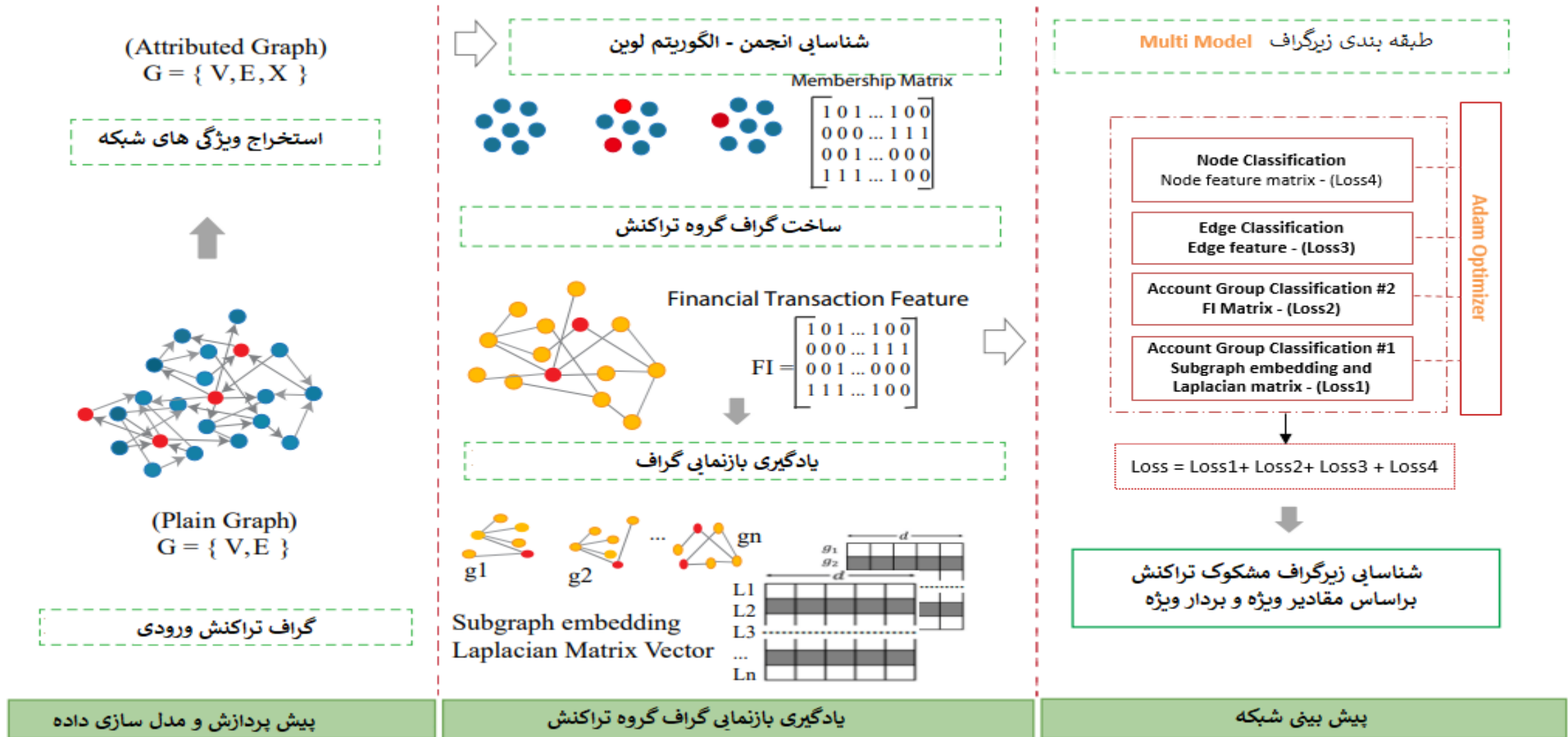
پیشینه پژوهش و کارهای مرتبط: تحلیل گراف

مزایا - معایب	کاربرد	مدل پیشنهادی
گروه‌هایی از گره‌ها را با هم پردازش می‌کند و باعث کاهش محاسبات و سربار حافظه می‌شود. با کاهش پیچیدگی محاسباتی از طریق پردازش گروهی، در نمودارهای بزرگ، مقیاس بهتری دارد. تعریف گروه‌های معنی دار از گره‌ها می‌تواند چالش برانگیز باشد و به دانش حوزه نیاز دارد.	بر اساس گراف شهر، گراف گروه شهر را برای مدل‌سازی وابستگی‌های فضایی و نهان بین شهرها ایجاد می‌کند. تا بتواند پیش‌بینی کیفیت هوای شهر را در سراسر کشور انجام دهد.	GAGNN [2]
پیچیدگی و منابع محاسباتی مورد نیاز برای آموزش و استقرار مدل شبکه عصبی گراف. بهینه‌سازی مدل‌های GAGNN ممکن است نیاز به تنظیم پارامترهای پیچیده پردازش گروهی داشته باشد که به طور بالقوه زمان‌های آموزشی را افزایش می‌دهد.	از یک رمزگذار انجمن محور و لایه بازنمایی گروهی عمیق برای تجزیه و تحلیل گراف تراکنش و شناسایی رفتارهای مجرمانه جمعی مانند پول‌شویی استفاده می‌کند.	GAT+eMRF (GAGNN) [1]
قابلیت تفسیر و قدرت تمایز وظایف طبقه‌بندی گراف را افزایش می‌دهد.	شناسایی و استخراج زیرگراف‌های کلیدی، گراف.	SUGAR [3]

[1] CHENG, Dawei, et al. Anti-money laundering by group-aware deep graph learning, 2023.

[2] CHEN, Ling, et al. Group-aware graph neural network for nationwide city air quality forecasting, 2023.

[3] Sun, Q., Li, J., Peng, H., Wu, J., Ning, Y., Yu, P.S. and He, L., Sugar: Subgraph neural network with reinforcement pooling and self-supervised mutual information mechanism. In Proceedings of the Web Conference 2021.



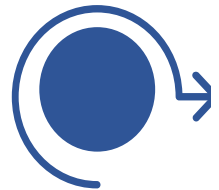


AUC



$$\frac{TP}{TP + FP} = \text{Precision}$$

$$\frac{TP}{TP + FN} = \text{Recall}$$

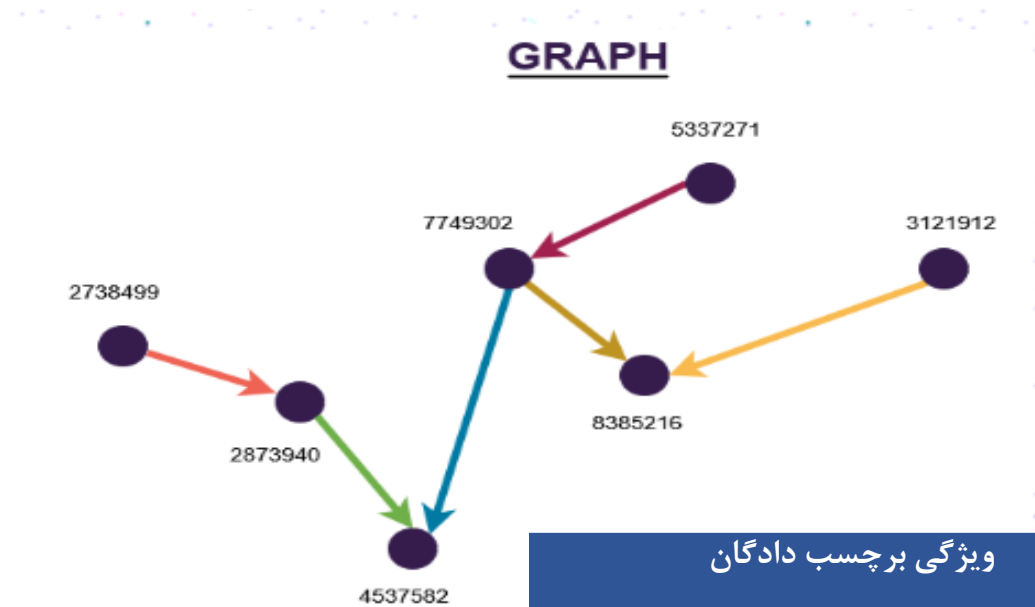
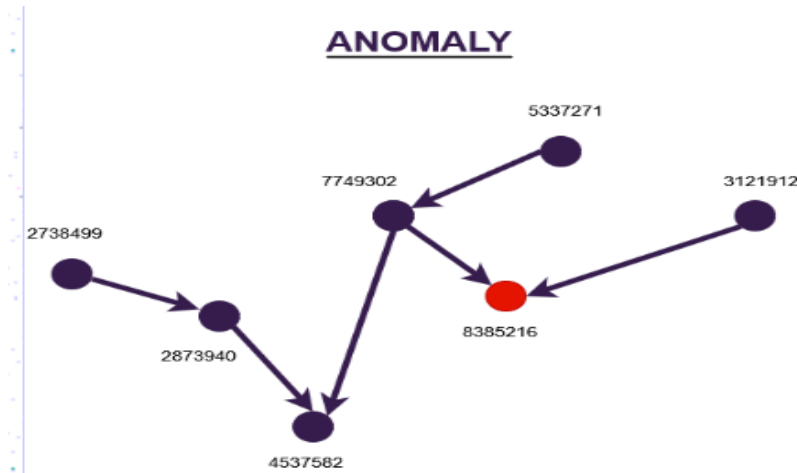


$$\frac{\text{Precision} * \text{Recall}}{\text{Precision} + \text{Recall}} * 2 = F - \text{Score}$$

مجموعه دادگان تراکنش‌های مالی لیبرابانک

بانک اینترنتی لیبرا، یک بانک رومانیایی است که فهرست کاملی از تراکنش‌ها را در طول سه ماه ارائه کرده است که در این پژوهش از آن استفاده شده است [۱].

ویژگی مجموعه دادگان	
4558805	تعداد تراکنش
3264	میانگین ارزش معامله
152	میانه ارزش معامله
42 میلیون	حداکثر ارزش معامله
0.01	حداقل ارزش معامله



ویژگی برچسب دادگان	
۵۱۷	تعداد تراکنش با برچسب هشدار
۱۱	تعداد تراکنش با برچسب گزارش
۶۰۰	تعداد گره‌های دارای برچسب هشدار
۱۵	تعداد گره‌های دارای برچسب گزارش

[1] DUMITRESCU, Bogdan; BĂLTOIU, Andra; BUDULAN, Ștefania. Anomaly detection in graphs of bank transactions for anti money laundering applications, 2022

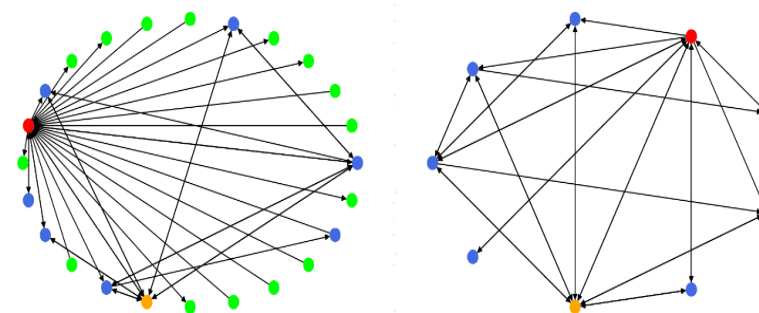
۸ درصد تراکنش‌ها آномال هستند.

مجموعه دادگان تراکنش‌های مالی لیبرابانک

ویژگی‌های استخراج شده از شبکه تراکنش‌های بانکی

عنوان ویژگی	شرح - محاسبه
تعداد همسایه‌های ورودی به یک گره	$N_i(k) = \{l \in V (l, k) \in E\}$ (k تعداد نودهای مشترک)
تعداد همسایه‌های خروجی	$N_o(k) = \{l \in V (k, l) \in E\}$ (k تعداد نودهای مشترک)
ویژگی‌های پایه	$F = \{in/out\ degree\}$ $\cup \{total\ in/out\ amount\}$ $\cup \{average\ in/out\ amount\}$ ----- $d_i = N_i(k) $ $d_o = N_o(k) $ $t_{ai}(k) = \sum_{l \in N_i(k)} s_{lk}$ $t_{ao}(k) = \sum_{l \in N_o(k)} s_{lk}$
نسبت بین تعداد لبه‌های egonet و تعداد گره‌های آن است	$E(k) = k \cup N_i(k) \cup N_o(k)$ egonet edge density

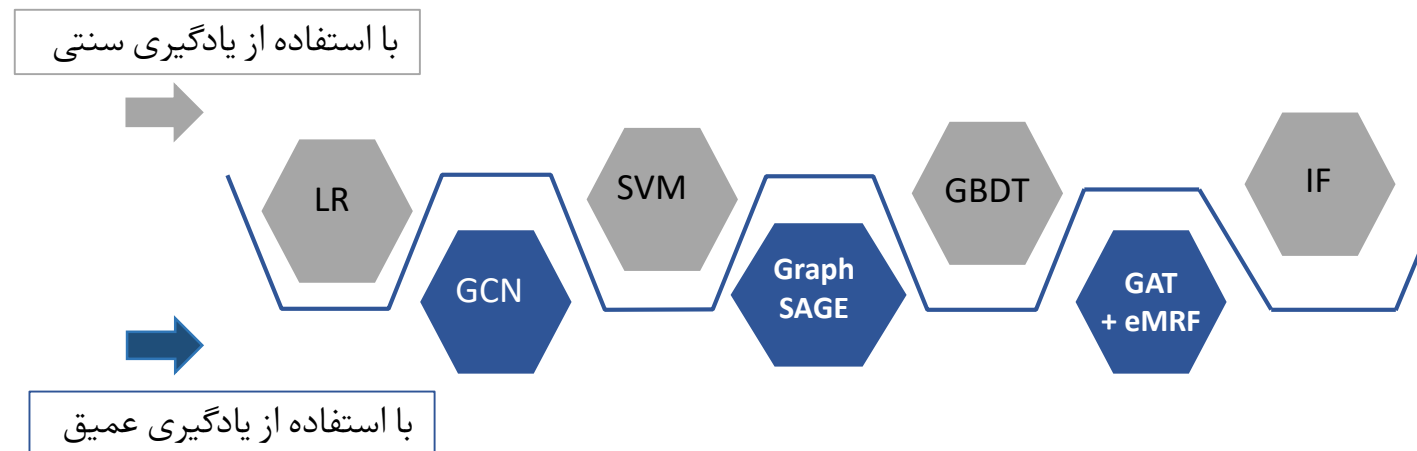
Features				
ID	Degree IN	Degree OUT	Amount IN	Amount OUT
2738499	0	1	0	200
2873940	1	1	200	679
3121912	0	1	0	380
4537582	2	0	699	0
5337271	0	1	0	1000
7749302	1	2	1000	44120
8385216	2	0	44480	0



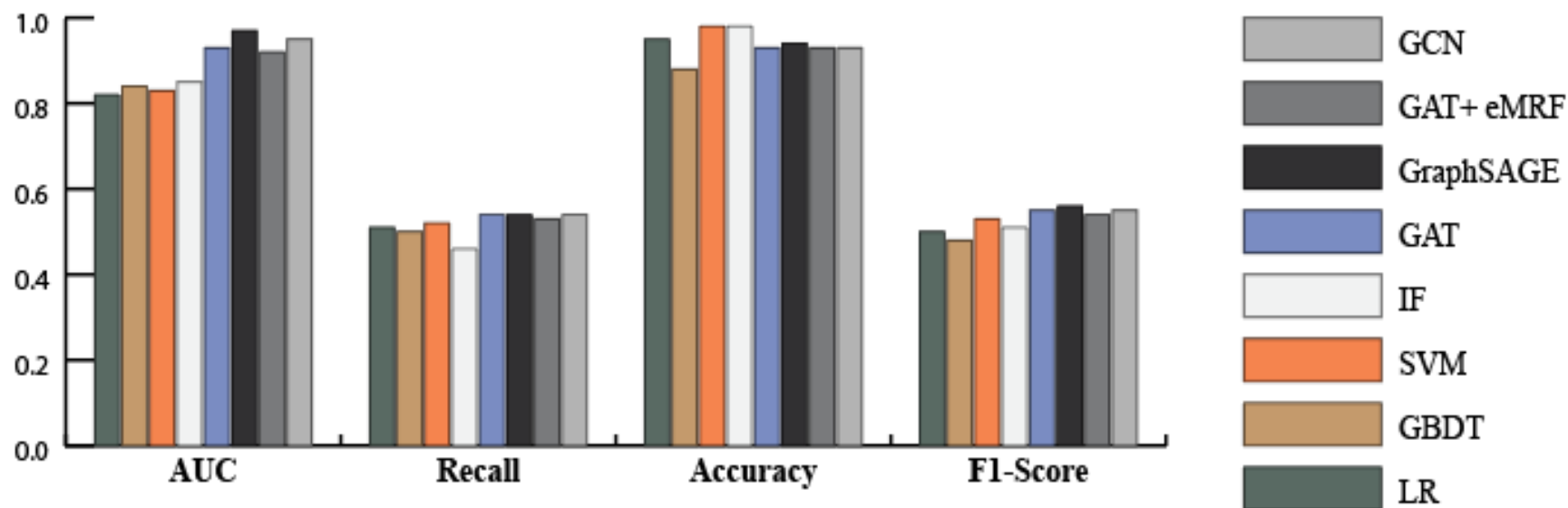
نمونه‌ای از آگونت و آگونت کاهش یافته

تشخیص ناهنجاری در سطح گره

برخی از روش‌های یادگیری ماشین که در پژوهش‌های پیشین مورد استفاده قرار گرفته است، برای تشخیص ناهنجاری در سطح گره برای تراکنش‌های مالی لیبرابانک استفاده شد تا اثربخشی روش پیشنهادی را برجسته نماید.

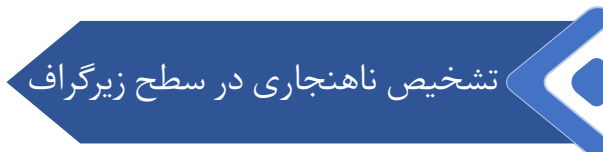


تشخیص ناهنجاری در سطح گره



F1-score	Precision	Accuracy	Recall	AUC	مدل
0.50	0.82	0.95	0.51	0.82	LR
0.48	0.84	0.88	0.50	0.84	GBDT
0.53	0.83	0.98	0.52	0.83	SVM
0.51	0.85	0.98	0.46	0.85	Isolation Forest
0.55	0.95	0.93	0.54	0.95	GCN
0.56	0.97	0.94	0.54	0.97	GraphSAGE
0.55	0.93	0.93	0.54	0.93	GAT
0.54	0.92	0.93	0.53	0.92	GAT + eMRF

تشخیص ناهنجاری در سطح زیرگراف



تراکنش‌های پول‌شویی به صورت پراکنده در کل گراف تراکنش توزیع نشده است و بین برخی از حساب‌های فردی متمرکز است.



تشخیص ناهنجاری در سطح زیرگراف

ساخت گراف گروه تراکنش

روش سنتی هنگام تجزیه و تحلیل قادر به تشخیص رفتار تراکنش در سطح گروه در ابعاد بالا نمی باشد.

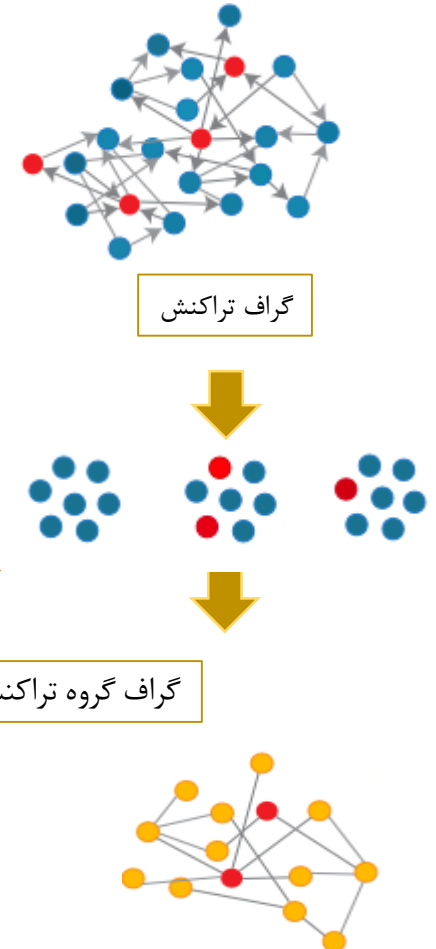
برای دستیابی به تجمیع گره ها و پیش بینی ناهنجاری در سطح گروه سیاست هایی معرفی شده است که با چالش هایی همراه می باشد:

همپوشانی گره ها در گروه ها در نظر گرفته نمی شود

با توجه به اینکه فرآیند پول شویی در برخی از موارد شامل مجموعه ای از تراکنش های عادی نیز می باشد، حساب هایی وجود دارند که با تراکنش عادی به حسابی که تراکنش های مشکوک دارد متصل می شود. لذا برخی از تراکنش ها و حساب هایی که عادی می باشند به اشتباه در گروه تراکنش های مشکوک قرار می گیرند.

مواردی از حساب هایی که ممکن است در ارتباط با پول شویی باشند و در گروه مشکوک قرار نگرفته باشد نیز وجود دارد.

مبنای تجمیع گره ها (حساب های مشکوک) تراکنش های مشکوک می باشد، لذا در صورت وجود تراکنش های عادی امکان تجمیع بهینه نمی باشد و و زمان اجرای این مدل ها نیز افزایش می یابد



استفاده از الگوریتم های تشخیص انجمن
Louvian Algorithm: الگوریتم لوین

تشخیص ناهنجاری در سطح زیرگراف

استفاده از الگوریتم های تشخیص انجمن
الگوریتم لوین: Louvian Algorithm

این الگوریتم تکرارشونده توسط بلاندل و همکاران [1] در دو فاز طراحی شده است.
یک الگوریتم کارآمد برای تقسیم گراف به انجمن ها است که بر اساس بهینه سازی معیار ماژولاریتی است.

$$Q = \frac{1}{2m} \sum_{i,j} \left[A_{i,j} - \frac{k_i k_j}{2m} \right] \delta(c_i, c_j)$$

مرحله ادغام: Aggregation Phase

در مرحله دوم، به شکل بازگشتی با شناسایی الگوهای ارتباطی حساب ها تجزیه و تحلیل می-شوند تا تقسیم بندی نهایی شبکه به انجمن ها به دست آید.

مرحله تقسیم: Partitioning Phase

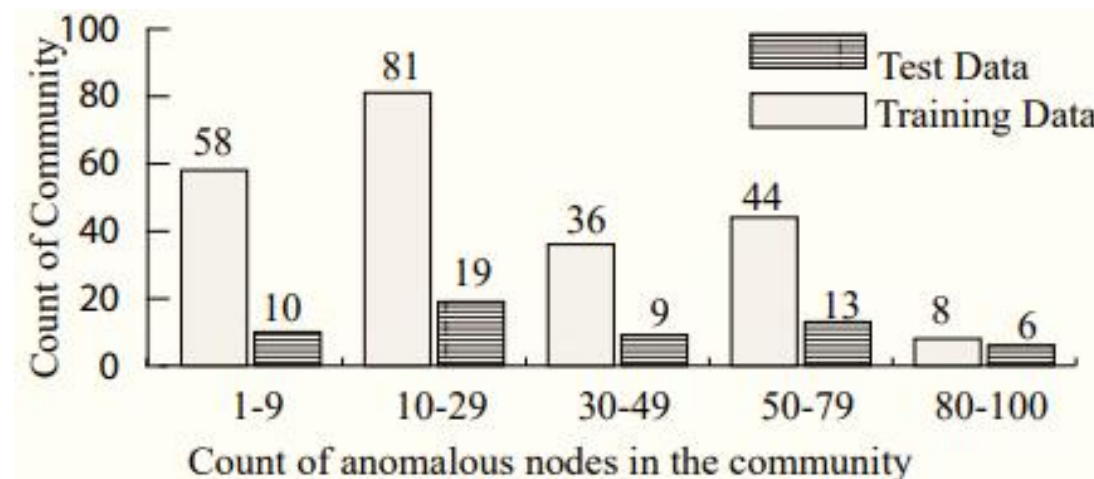
مطابق با الگوریتم لوین در مرحله نخست هر یک از حساب های مشتریان به عنوان یک انجمن در نظر گرفته می شود تا براساس معاملات بین حساب ها فرآیند الگوریتم تکمیل شود.

[1] BLONDEL, Vincent D., et al. Fast unfolding of communities in large networks. Journal of statistical mechanics: theory and experiment, 2008.

تشخیص ناهنجاری در سطح زیرگراف

۲۲۷	تعداد گره (انجمن) با برچسب مشکوک	ویژگی گراف گروه تراکنش آموزش
۸۳۱۹۴	تعداد یال‌ها (تراکنش)	
۱۳۵۴۰	تعداد گره (انجمن) با برچسب عادی	
۵۷	تعداد گره (انجمن) با برچسب مشکوک	ویژگی گراف گروه تراکنش آزمون
۴۶۶۴	تعداد یال‌ها (تراکنش)	
۲۰۴۱	تعداد گره (انجمن) با برچسب عادی	

میزان مشکوک بودن انجمن شناسایی شده با توجه به تعداد اعضای هر انجمن



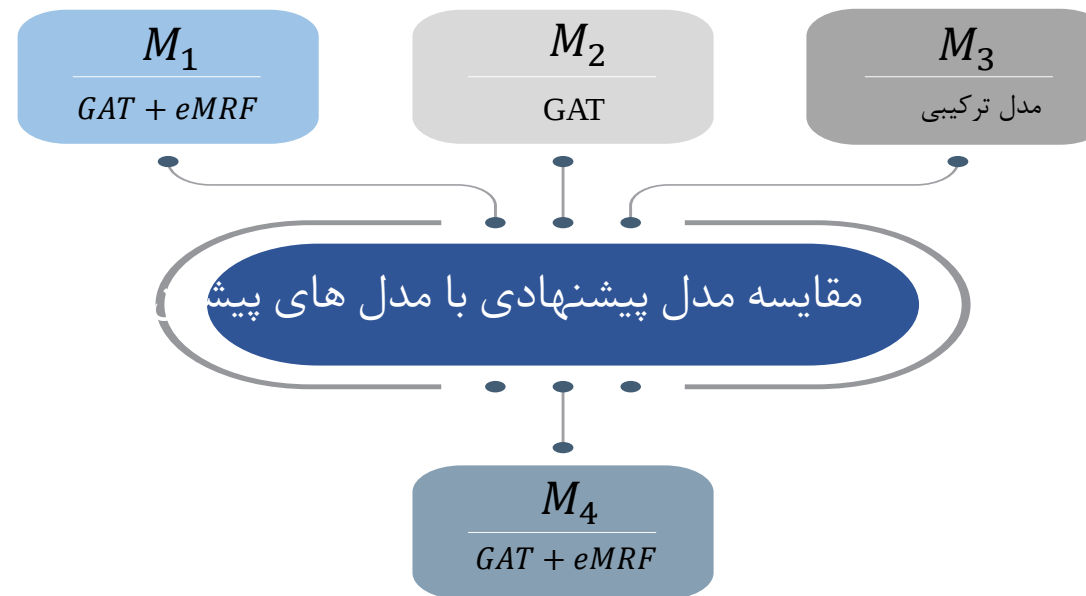
مقایسه مدل پیشنهادی با مدل های پیشین

رویکرد اول

استفاده از استراتژی بهینه سازی برای آموزش و یادگیری

رویکرد دوم

به جای تجمع گره ها براساس تراکنش مشکوک از الگوریتم دو مرحله ای لوین برای شناسایی انجمن ها استفاده گردید و براساس انجمن های شناسایی شده گراف گروه تراکنش ایجاد گردید



نتایج نهایی و ارزیابی اثر بخشی مدل پیشنهادی

رویکرد اول

استفاده از استراتژی بهینه‌سازی برای آموزش و یادگیری

رویکرد دوم

به جای تجمع گره‌ها براساس تراکنش مشکوک از الگوریتم دو مرحله‌ای لوین برای شناسایی انجمن‌ها استفاده گردید و براساس انجمن‌های شناسایی شده گراف گروه‌تراکنش ایجاد گردید

F-score	Precision	Accuracy	Recall	AUC	مدل
0.50	0.82	0.81	0.50	0.83	M_1
0.51	0.85	0.85	0.51	0.85	M_2
0.52	0.90	0.90	0.51	0.90	M_3
0.55	0.90	0.90	0.54	0.90	M_4

مدل پیشنهادی M_4 علاوه بر مرتفع نمودن چالش‌های عنوان شده از امتیاز و فراخوانی گروه مشکوک بالاتری نسبت به مدل پیشنهادی چنچ و همکاران M_2 در سال 2023 برخوردار است.



تشخیص ناهنجاری در سطح زیرگراف

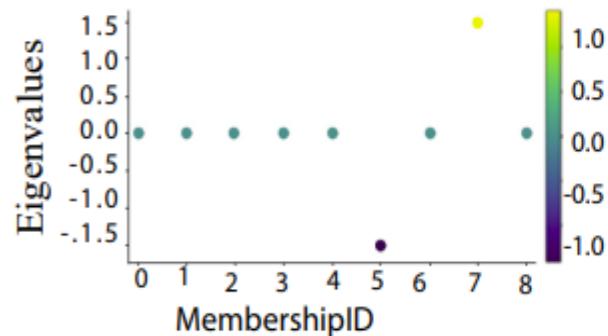
در زیرگراف‌هایی که با برچسب مشکوک طبقه‌بندی می‌شوند ممکن است حسابی که برچسب مشکوک دارد تراکنشی نیز داشته باشد که ماهیت آن عادی باشد. برای دقت در عملکرد مدل و حذف تراکنش‌های عادی از زیرگراف مشکوک، از مقادیر ویژه و بردار ویژه استفاده می‌شود.



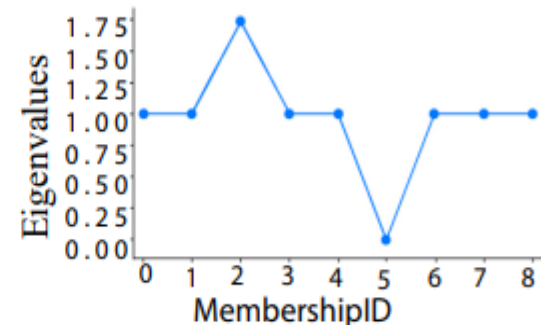
به کمک این ویژگی‌ها می‌توان الگوهای عادی را در داده‌ها شناسایی کرده و سپس انحرافات نسبت به این الگوها را در قالب امتیاز ناهنجاری تحلیل کرد.

پس از مشخص کردن امتیاز ناهنجاری با استفاده از یک مقدار حد آستانه تراکنش‌هایی که امتیاز ناهنجاری کمتری نسبت به مقدار حد آستانه دارند از زیرگراف حذف می‌گردد تا در نهایت زیرگراف مشکوک شناسایی شود.

مقادیر ویژه و بردار ویژه برای انجمن ۳۷۶۴



(a)



(b)

پیاده‌سازی

برای پیاده‌سازی از کتابخانه‌های پایتون استفاده گردید.

مسیر اجرایی و لینک کدهای پیاده‌سازی شده در مخزن گیت‌هاب قابل دسترسی می‌باشد.

<https://github.com/enznoor/AnomalyDetection.git>

چالش‌های مرتبط

با توجه به اینکه ساختار کل گراف برای مرحله آموزش دیده شده است، در زمان آزمون نتایج تحت تاثیر قرار می‌گیرد. راه حل استفاده شده در مراحل پیاده‌سازی این بود که دو گراف آموزش و آزمون به عنوان دو زیرگراف از گراف اولیه حاصل شود.

نتایج بدست آمده نشان داد که با این روش نیز تمام گره‌ها و یال‌ها در زمان آموزش دیده می‌شوند. لذا در نهایت برای آموزش و آزمون، دو گراف کاملاً مجزا در نظر گرفته شد تا نتایج بدست آمده در مرحله آزمون، براساس گره‌ها و یال‌های دیده نشده در مرحله آموزش بدست آید.

در فرآیند پیاده‌سازی تمام روش‌های مورد بررسی با یک مجموعه داده، مورد آموزش و آزمون قرار گرفت.

جمع بندی و پیشنهادهای کاربردی



شناسایی زیرشبکه‌های مشکوک می‌تواند به موسسات مالی کمک کند تا فعالیت‌های غیرقانونی مانند تقلب، اختلاس، یا فعالیت‌های مرتبط با جرایم سازمان‌یافته را متوقف کنند.



پیشرفت‌های اخیر در تشخیص ناهنجاری، به‌ویژه در حوزه تجزیه و تحلیل گراف مبتنی بر تکنیک‌های یادگیری عمیق، راه را برای استراتژی‌های مؤثرتر در کشف این شبکه‌های پیچیده مالی غیرقانونی هموار کرده است.



بررسی تاثیر استفاده از داده‌های کمکی مانند اطلاعات تکمیلی تراکنش‌ها مانند محل جغرافیایی تراکنش، زمان تراکنش، نوع تراکنش در تشخیص ناهنجاری‌های زیرگراف



با استفاده از الگوریتم‌های پیمایش گراف و ترکیب آن‌ها با روش‌های یادگیری عمیق، زیرگراف‌های مشکوک در شبکه‌ها شناسایی و تشخیص داده می‌شود.



استفاده از حاشیه‌نویسی و پیمایش گراف که تأثیر قابل توجهی بر روش‌های بازنمایی گراف دارند.



با سپاس

enznoor@gmail.com