Islamic Azad University , Shiraz Branch
نشریه تحلیل مدارها، داده ها و سامانه ها
Journal of Circuits, Data and Systems Analysis
sanad.iau.ir/journal/jcdsa

A method for identifying social influence using embedded similarity and homophily in location-based social networks

Zohreh Sadat Akhavan-Hejazi[1] ,Mahdi Esmaeili[2] ,Mostafa Ghobaei-Arani[3*] ,Behrouz Minaei-Bidgoli[4]

[1] Department of Computer Engineering, Qo.C., Islamic Azad University, Qom, Iran
z.akhavanhejazi@iau.ac.ir
[2] Department of Computer Engineering, Kas.C, Islamic Azad University, Kashan, Iran
mesmaeili@iau.ac.ir
[3] Department of Computer Engineering, Qo.C., Islamic Azad University, Qom, Iran
mo.ghobaei@iau.ac.ir
[4] Department of Computer Engineering, Iran University of Science &Technology, Tehran, Iran
b_minaei@iust.ac.ir

Abstract: Location-based social networks are very important in the field of targeted commerce and advertising due to their valuable location data and user relationships. A key issue in these networks is determining the social influence of users to identify influential users. Research has proven the role of node similarity in finding the social influence of users. Among the methods for calculating node similarity, many existing methods ignore multiple aspects of similarity between users. In this paper, by combining two criteria of homophily and embedded and basic similarities, the social influence of users in location-based networks is investigated. The results show that combined similarity increases the accuracy in identifying more influential users, especially when users have similar characteristics. This means that the personal and social characteristics of users, in addition to their structural relationships in the network, play a decisive role in analyzing social behaviors. The use of node embedding models improved the accuracy of structural similarities and the detection of complex relationships between nodes. By applying the proposed method on real datasets, it was shown that the use of node embedding models, due to their simplicity and lack of time spent on data training, is computationally and cost-competitive in processing more complex and larger data compared to advanced methods.

Keywords: Social influence, influential users, embedded similarity, homophily, location-based social networks.

JCDSA, Vol. 3, No. 3, Autumn 2025	Online ISSN: 2981-1295	Journal Homepage: https://sanad.iau.ir/en/Journal/jcdsa
Received: 2025-06-10	Accepted: 2025-09-12	Published: 2025-12-21

CITATION

Akhavan-Hejazi,ZS. ,et. al., "A method for identifying social influence using embedded similarity and homophily in location-based social networks", Journal of Circuits, Data and Systems Analysis (JCDSA), Vol. 3, No. 3, pp. 12-27, 2025.
DOI: 10.82526/jcdsa.2026.1209488

* Corresponding author

Extended Abstract

1- Introduction

Location-Based Social Networks (LBSNs), by integrating spatial data and social relationships, provide a crucial platform for analyzing human interactions, targeted marketing, and effective information dissemination. Identifying influential users in such networks is a fundamental challenge aimed at enhancing the spread of information and improving communication efficiency. In this context, structural similarity between nodes and the phenomenon of homophily—people's tendency to interact with others who are similar—play key roles in determining social influence. Existing methods often focus on only one of these aspects, neglecting the potential benefits of combining both. The proposed framework uses a hybrid algorithm to calculate structural similarity and homophily, including embedded and classical methods. The innovation of this approach is the emphasis on the important role of node characteristics on social influence over their topological structure in the network, which uses a combination of two embedded methods (for the friendship graph) and two classical methods due to the existence of a two-layer network including a friendship graph and a user login graph.

2- Methodology

This study proposes a multi-step approach to identify influential users in location-based social networks by integrating structural similarity, homophily, and spatiotemporal influence patterns. Initially, social interaction data and users' location visits are collected and preprocessed, forming two interconnected graphs: a user friendship graph and a bipartite user-location visit graph. To capture the complex relationships between users, both classical similarity measures such as Jaccard and Adamic/Adar indices, and advanced graph embedding techniques like Node2Vec and DeepWalk are employed to compute node similarities. These embedding methods leverage random walk simulations to generate vector representations of nodes, enabling more nuanced similarity assessments through cosine similarity.

In parallel, homophily—reflecting users' tendency to connect with others sharing similar attributes—is quantified using spatial similarity matrices derived from user features. A weighted combination of structural similarity and homophily, controlled by a parameter α, balances the contribution of each factor in calculating overall user similarity. The social influence is then estimated based on temporal patterns of location visits, where a user's influence is defined by the number of friends who subsequently visit the same locations within a specified time window. This spatiotemporal influence captures both the social and geographic proximity effects on information diffusion. Finally, the proposed NSH algorithm integrates these components by applying the combined similarity measures to candidate users extracted from the social and location graphs, and quantifies their influence through the defined temporal-spatial metrics. By varying α, the model evaluates the relative importance of structural connections and homophilic traits in determining social influence within location-based networks. This comprehensive methodology provides a robust framework to analyze and predict influential actors in complex social and spatial environments.

3- Results and discussion

The proposed algorithm was evaluated using real-world location-based social network datasets including Gowalla, Foursquare, and Brightkite. These datasets contain user interactions, locations, and social ties collected via public APIs. After preprocessing and filtering data for key regions such as Texas, California, and New York, the algorithm's performance was analyzed using Pearson correlation and linear regression to measure the relationship between estimated similarity and actual social influence. The NSH approach combines node embeddings (from methods like Node2Vec and DeepWalk) with homophily measures, effectively capturing both structural and geographic user similarities. Experiments on filtered regional data showed a positive correlation between NSH's similarity scores and users' social influence, with stronger correlations observed when homophily is given higher weight. The results consistently demonstrated that similarity based on shared traits significantly affects social influence in these networks.

Linear regression analysis confirmed the statistical significance of the relationship, supporting the predictive validity of NSH-derived similarities. Compared to baseline methods—including CSH (which combines structural similarity and homophily), AGCN-IM (a deep learning influence maximization model), and IGCN-POI (a GCN-based recommendation model)—NSH offered a strong balance between accuracy, interpretability, and efficiency. While AGCN-IM slightly outperformed NSH in accuracy, it is more complex and less interpretable, making NSH a more practical choice for unsupervised and real-time applications. In terms of computational complexity, NSH efficiently scales to large datasets by leveraging optimized calculations of spatial similarity and random walk embeddings.

4- Conclusion

This study showed that combining structural similarity with users' attribute-based homophily improves the accuracy of identifying influential users. Increasing the weight of homophily enhances social influence detection, highlighting the importance of user attribute similarity, especially in LBSNs where geographic and social factors impact behavior. Node embedding techniques outperformed traditional similarity measures by better capturing complex network relationships and handling larger datasets more efficiently. The results confirmed that both structural relations and user attributes significantly affect social influence. Future research should explore directed graphs, varied edge weights in user-location bipartite graphs, and include more diverse user features to further improve influence prediction

Islamic Azad University , Shiraz Branch
نشریه تحلیل مدارها، داده ها و سامانه ها
Journal of Circuits, Data and Systems Analysis
sanad.iau.ir/journal/jcdsa

روشی برای شناسایی تاثیر اجتماعی با استفاده از شباهت تعبیه شده و هوموفیلی در شبکه های اجتماعی مکان مبنا

زهره سادات اخوان حجازی[1]، مهدی اسماعیلی[2]، مصطفی قبائی آرانی[3]*، بهروز مینایی بیدگلی[4]

۱- گروه مهندسی کامپیوتر، واحد قم، دانشگاه آزاد اسلامی، قم، ایران (Z.akhavanhejazi@iau.ac.ir)
۲- گروه مهندسی کامپیوتر، واحد کاشان، دانشگاه آزاد اسلامی، کاشان، ایران (mesmaeili@iau.ac.ir)
۳- گروه مهندسی کامپیوتر، واحد قم، دانشگاه آزاد اسلامی، قم، ایران (mo.ghobaei@iau.ac.ir)
۴- گروه مهندسی کامپیوتر، دانشگاه علم و صنعت ایران، تهران، ایران (b_minaei@iust.ac.ir)

چکیده: شبکه‌های اجتماعی مکان‌مبنا به دلیل داده‌های ارزشمند موقعیت مکانی و روابط کاربران، در حوزه تجارت و تبلیغات هدفمند بسیار مهم هستند. یک موضوع کلیدی در این شبکه‌ها، تعیین نفوذ اجتماعی کاربران برای شناسایی افراد تأثیرگذار است. تحقیقات اخیر، نقش شباهت گره‌ها را در تعیین تأثیر اجتماعی کاربران به روشنی نشان داده است. از بین روشهای محاسبه شباهت گره‌ها، بسیاری از روش‌های موجود جنبه‌های متعدد شباهت بین کاربران را نادیده می‌گیرند. در این مقاله، با ترکیب دو معیار هوموفیلی و شباهت‌های تعبیه‌شده و پایه، تأثیر اجتماعی کاربران درشبکه‌های مکان‌مبنا بررسی شده است. نتایج نشان می‌دهد شباهت ترکیبی به‌ویژه در مواقعی که کاربران دارای ویژگی‌های مشابه هستند، موجب افزایش دقت در شناسایی کاربران با نفوذ بیشتر می‌شود. این بدان معناست که ویژگی‌های شخصی و اجتماعی کاربران، علاوه بر روابط ساختاری آنها در شبکه، نقش تعیین‌کننده‌ای در تحلیل رفتارهای اجتماعی ایفا می‌کنند. استفاده از مدل‌های تعبیه گره باعث بهبود دقت شباهت‌های ساختاری و تشخیص روابط پیچیده میان گره‌ها شد. با بکارگیری روش پیشنهادی برروی مجموعه داده‌های واقعی، نشان داده شد که استفاده از مدل‌های تعبیه گره بدلیل سادگی و عدم صرف زمان برای آموزش داده‌ها، در پردازش داده‌های پیچیده‌تر و بزرگ‌تر درمقایسه با روش‌های پیشرفته، از نظر محاسباتی و هزینه قابل رقابت می‌باشند.

واژه های کلیدی: تاثیر اجتماعی، کاربران اجتماعی، شباهت تعبیه شده، هوموفیلی، شبکه‌های اجتماعی مکان مبنا.

DOI: 10.82526/jcdsa.2026.1209488

نوع مقاله: پژوهشی
تاریخ ارسال مقاله: ۱۴۰۴/۰۳/۲۰ تاریخ پذیرش مقاله: ۱۴۰۴/۰۶/۲۱ تاریخ چاپ مقاله: ۱۴۰۴/۰۹/۳۰

۱- مقدمه

شبکه‌های اجتماعی به عنوان بستری گسترده برای ارتباط و اشتراک‌گذاری اطلاعات میان کاربران با ویژگی‌ها و علایق مشابه، نقش مهمی در زندگی دیجیتال امروز ایفا می‌کنند. رشد سریع این شبکه‌ها موجب افزایش تحقیقات در حوزه تحلیل نفوذ اجتماعی شده است، به‌ویژه با توجه به حجم عظیم داده‌های آنلاین، شناسایی کاربران تأثیرگذار به یک چالش مهم تبدیل شده است [۱و۲]. شبکه‌های اجتماعی به‌عنوان گراف‌هایی مدل‌سازی می‌شوند که در آن‌ها گره‌ها نمایانگر کاربران و یال‌ها نشان‌دهنده ارتباطات میان آن‌ها هستند. برای تحلیل تأثیر در این شبکه‌ها، رتبه‌بندی گره‌ها و محاسبه شباهت میان آن‌ها ضروری است. با توجه به بزرگ و پیچیده‌تر شدن این گراف‌ها، توسعه الگوریتم‌های کارا برای محاسبه سنجه‌های شباهت از اهمیت بالایی برخوردار است. تحقیقات نشان می‌دهند که تاثیر اجتماعی[2] افراد یکسان نیست و ساختار شبکه و تعاملات کاربران در تعیین این نفوذ نقش اساسی دارند [۳و۴]. دو عامل کلیدی در انتشار اطلاعات، نفوذ اجتماعی و هوموفیلی[3] هستند، که در شبکه‌های اجتماعی مکان‌مبنا[4] اهمیت بیشتری پیدا می‌کنند. این شبکه‌ها با افزودن بعد مکانی به تعاملات کاربران، امکان اشتراک‌گذاری مکان‌های فیزیکی را فراهم می‌کنند [۵-۷]. شناسایی کاربران تأثیرگذار در این شبکه‌ها نیازمند در نظر گرفتن عوامل ساختاری و رفتاری است، چرا که اقدامات یک کاربر می‌تواند بر دوستانش تأثیر بگذارد. روش‌های متعددی برای تعیین نفوذ

* نویسنده مسئول

[2] Social influence
[3] Homophily
[4] Locaion Based Social Networks(LBSNs)

کاربران طراحی شده‌اند که عمدتاً بر پیوندهای بین کاربران یا ویژگی‌های همبندی شبکه تمرکز دارند [۶-۱۰]، اما ترکیب ساختار شبکه و رفتار کاربر کمتر مورد توجه قرار گرفته است.

با توجه به تنوع شبکه‌های اجتماعی و دیدگاه‌های تحقیقاتی، معیارهای ارزیابی اهمیت گره‌ها نیز متفاوت است، که این موضوع نیاز به توسعه روش‌های نوین و کارا برای تحلیل این شبکه‌ها را نشان می دهد. در این تحقیق، تمرکز بر تحلیل شبکه‌های اجتماعی مکان‌مبنا و شناسایی کاربران تأثیرگذار با در نظر گرفتن عوامل نفوذ اجتماعی، هوموفیلی، و شباهت مکانی-زمانی است. با در نظر گرفتن یک شبکه دو لایه شامل شبکه ارتباط اجتماعی و شبکه بازدید از مکان، تاثیر کاربران محاسبه می‌شود و سپس با تحلیل نفوذ بین کاربران، روشی ممکن برای یافتن گره‌های تاثیرگذار با ترکیب شباهت‌های درون شبکه پیشنهاد می‌شود. در اکثر روش‌های پیشنهادی، آن دسته از کاربرانی که در موقعیت همبندی مناسب شبکه قرار دارند، به عنوان کاربران تأثیرگذار در نظر گرفته می‌شوند. این الگوریتم‌ها معمولاً ویژگی‌های کاربران را در نظر نمی‌گیرند و آن روابط را دودویی فرض می‌کنند. بنابراین، این پژوهش رویکرد جدیدی را برای تعیین کاربران تأثیرگذار در یک شبکه اجتماعی با درنظر گرفتن هوموفیلی کاربران ارائه می‌کند. هوموفیلی به معنای تمایل افراد به برقراری ارتباط با کسانی است که ویژگی‌های مشابهی دارند. این پدیده می‌تواند به شکل‌گیری الگوهای خاصی در تعاملات اجتماعی منجر شود که در نهایت بر نفوذ اجتماعی کاربران تأثیر می‌گذارد [۶و۱۱]. در این تحقیق قصد داریم تا یک رویکرد و چارچوب بهینه برای نفوذ اجتماعی در یک شبکه اجتماعی مکان‌مبنا ارائه دهیم که بتواند از دو ویژگی ساختار همبندی گراف و صفات خاصه‌ی گره-های آن در تعیین شباهت گره‌ها و مدیریت نفوذ اجتماعی آنها استفاده نماید. در این راستا از تکنیک‌های شباهت تعبیه شده بجای روش‌های ساختاری پایه استفاده می‌شود [۱۲]. ادامه این مقاله به شرح زیر سازماندهی شده است؛ در بخش ۲، کارهای مرتبط جدید در رابطه با پژوهش ما مورد مطالعه قرار می‌گیرند. در بخش ۳، راه حل پیشنهادی با جزئیات بیشتر شرح داده می‌شود؛ در بخش ۴، ارزیابی ارائه شده و نتایج تجربی نمایش داده می‌شود. در نهایت، نتیجه‌گیری در بخش ۵ ارائه می‌شود.

۲- پیشینه تحقیق

شبکه‌های اجتماعی مکان‌مبنا به عنوان بسترهایی برای تعاملات اجتماعی و تبادل اطلاعات میان کاربران به سرعت در حال توسعه هستند. این شبکه‌ها به کاربران اجازه می‌دهند تا در مکان‌های خاص به اشتراک‌گذاری تجربیات بپردازند و به ایجاد ارتباطات جدید بپردازند. تحقیقات نشان می‌دهد که این پلتفرم‌ها نه تنها به عنوان ابزارهای ارتباطی، بلکه به عنوان منابع مهم اطلاعاتی و تجاری نیز شناخته شده‌اند [۴و۱۱]. شناسایی کاربران تأثیرگذار یکی از موضوعات تحقیقاتی مهم در حوزه شبکه‌های اجتماعی است. از آنجایی که تأثیرگذاران می‌توانند نقش کلیدی در انتشار اطلاعات و شکل‌دهی به رفتارهای اجتماعی ایفا کنند، بهینه‌سازی روش‌های شناسایی آنها حائز اهمیت است. تحقیقات پیشین به بررسی ویژگی‌های کاربران تأثیرگذار از جمله درجه اتصال و فعالیت‌های شبکه‌ای و شباهت ساختاری آنها پرداخته‌اند [۷و۱۲و۱۳]. معیارهای مرکزیت[1] کلاسیک مانند درجه[2]، نزدیکی[3]، میانگی[4]، بردار ویژه[5] و غیره به عنوان ابزارهایی برای تعیین تأثیر هر گره در شبکه مورد استفاده قرار گرفتند. این روشها اگرچه ساده و کارآمد بودند، اما در شبکه‌های بزرگ با پیچیدگی محاسباتی مواجه می‌شدند و قادر به در نظرگرفتن ویژگی‌های رفتاری کاربران نبودند. اکثر پژوهش‌ها از روش های ترکیبی با در نظرگرفتن ساختار شبکه‌ای گره‌ها جهت شناخت ویژگی‌های تأثیرگذاران به شناسایی استراتژی‌های مؤثر برای بهره‌برداری از توانمندی‌های آنها استفاده نمودند [۱۴-۱۷]. در [۶و۱۲] نشان داده شد که هوموفیلی به عنوان تمایل افراد به برقراری ارتباط با کسانی که ویژگی‌های مشترکی دارند، یکی از عواملی است که می‌تواند بر تعاملات اجتماعی تأثیر بگذارد. تحقیقات نشان دادند که کاربران تمایل دارند با افرادی ارتباط برقرار کنند که از نظر ویژگی‌های فردی، علایق یا رفتارهای اجتماعی مشابه هستند. این پدیده نه تنها بر ساختار شبکه‌های اجتماعی تأثیر می‌گذارد، بلکه در فرآیند انتشار اطلاعات و نفوذ اجتماعی نیز مؤثر است، بنابراین، درک هوموفیلی می‌تواند به شناسایی الگوهای رفتاری کاربران و شناسایی تأثیرگذاران کمک کند [۱۸-۲۰]. در [۶،۱۱،۲۰] نشان داده شد که کاربران تمایل دارند با افرادی ارتباط برقرار کنند که از نظر جغرافیایی یا اجتماعی مشابه هستند. با پیشرفت تکنولوژی و ابزارهای یادگیری ماشین، امکان تحلیل دقیق‌تری از داده‌های شبکه‌های اجتماعی فراهم شده است. الگوریتم‌های یادگیری ماشین، به محققین این امکان را می‌دهند که به شناسایی الگوهای پیچیده در داده‌های حجیم، پیش‌بینی رفتار کاربران و شناسایی تأثیرگذاران بپردازند. تکنیک‌های تعبیه گره‌ها[6] به عنوان یکی از روش-های نوین در شناسایی کاربران تأثیرگذار مورد توجه قرار گرفتند. این روش‌ها با تبدیل گره های شبکه به بردارهایی در فضای چندبعدی، امکان محاسبه دقیقتر شباهت بین کاربران را فراهم کردند [۲۱-۲۵].

در [۲۱] روش DeepWalk معرفی شد که با انجام پیاده‌روی‌های تصادفی در میان گره‌های گراف، توالی گره‌ها را جمع آوری کرده و از این توالی ها برای یادگیری جاسازی گره ها استفاده می‌کند. در [۲۵] روش Node2Vec معرفی شد که با استفاده از پیاده روی‌های تصادفی و تنظیم پارامترهای جستجو، جاسازی گره‌ها را یاد می‌گیرد و می‌تواند ویژگی‌های ساختاری و رفتاری کاربران را به طور همزمان در نظر بگیرد. برخی مراجع به بررسی استفاده از تکنیکهای یادگیری عمیق [۲۶] و شبکه‌های عصبی گرافی [۲۷] برای شناسایی کاربران تأثیرگذار در شبکه‌های اجتماعی مبتنی بر مکان پرداخته و نشان دادند که چگونه

[1] Centrality Measures
[2] Degree Centrality
[3] Closeness Centrality
[4] Betweenness Centrality
[5] Eigenvector Centrality
[6] Nodes embedding

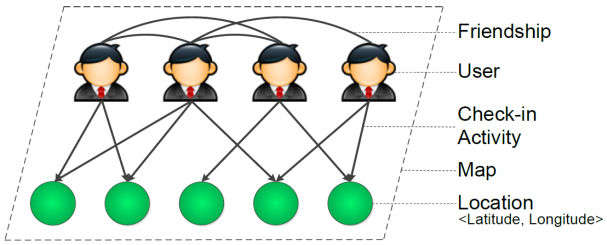
این روش‌ها می‌توانند به بهبود دقت در شناسایی کاربران تأثیرگذار کمک کنند. در [۲۸] الگوریتم NodeSim معرفی گردید که شباهت گره‌ها و ساختار جامعه را حفظ می‌کرد و در پیش‌بینی لینک‌ها بسیار دقیق عمل کرد. در پژوهشی دیگر، محققان یک روش جدید جاسازی گراف با معیارهای شباهت بررسی نمودند. آنها با استفاده از یک معیار شباهت متقارن مبتنی بر توجه و جمع‌آوری Top-k کاربران، ارتباط بین گره ها را کشف نموده و از تابع معیار شباهت متقارن بهمراه اطلاعات توپولوژیکی هر گره جهت رمزگزاری استفاده نمودند [۲۹]. در [۳۰] نویسندگان با استفاده از مدل‌های تعبیه‌ای LBSN2vec و PLBSN2vec، کاربران تأثیرگذار در موقعیت‌های مکانی را شناسایی کردند. هدف، شناسایی کاربران LBSN بود که توانایی جذب بازدیدکنندگان بیشتر به مکان هدف در ابتدای کمپین تبلیغاتی را داشتند. پژوهشگران در [۳۱و۳۲] دو مدل LBSN2Vec و LBSN2Vec++ را با استفاده از هایپرگراف‌ها و رندمواک با توقف، وابستگی پیچیده میان تحرک کاربران و روابط اجتماعی مدل‌سازی کردند و عملکرد بهتری نسبت به روش‌های پیشرفته موجود ارائه دادند. نویسندگان در [۳۳] مسئله حداکثرسازی تأثیر را در شبکه‌های اجتماعی با استفاده از شبکه‌های پیچشی گراف [1] (GCN) پیاده‌سازی نمودند. آن‌ها استدلال کردند که مدل‌های سنتی نمی‌توانند ساختارهای شبکه پنهان را به طور مؤثر ثبت کنند و از همپوشانی تأثیر بین گره‌های با مرکزیت بالا رنج می‌برند. برای رفع این محدودیت‌ها، آن‌ها یک مدل GCN تطبیقی (AGCN) را معرفی کردند که ویژگی‌های مرکزیت گره و مکانیسم‌های انتشار مبتنی بر توپولوژی را ادغام می‌کند. چارچوب آن‌ها به صورت پویا وزن‌های تأثیر را تنظیم می‌کند که منجر به بهبود مقیاس‌پذیری و دقت پیش‌بینی تأثیر می‌شود. مقایسه عملکرد نشان داد که رویکرد آنها در به حداکثر رساندن نفوذ اجتماعی، از روش‌های اکتشافی کلاسیک بهتر عمل می‌کند.

در پژوهشی دیگر [۳۴]، نویسندگان بر تأثیر اجتماعی ضمنی در سیستم‌های توصیه‌گر با استفاده از شبکه‌های عصبی گراف [2] تمرکز کرده‌اند. آن‌ها چالش‌های کلیدی در مدل‌سازی تأثیر اجتماعی، مانند عدم تجمیع تعاملات تاریخی و مدل‌سازی تأثیر محدود مبتنی بر همسایه را شناسایی کردند. نویسندگان IIEGNN (تخمین تأثیر ضمنی با GNN) را پیشنهاد کردند، مدلی که بردارهای جاسازی کاربر را بر اساس تعاملات تاریخی تولید می‌کند و روابط همسایه اجتماعی را از طریق یک چارچوب گراف ساختاریافته در بر می‌گیرد. نتایج تجربی نشان داد که مدل‌سازی تأثیر مبتنی بر جاسازی، دقت توصیه را در مقایسه با روش‌های مرسوم مبتنی بر شباهت به طور قابل توجهی افزایش می‌دهد. مرجع [۳۵] روشی به نام RARE معرفی می‌کند که با استفاده از اطلاعات نقش و اجتماع گره‌ها، مسیرهای تصادفی آگاه از نقش ایجاد می‌کند. این روش قادر است به‌طور هم‌زمان نزدیکی گره‌ها و شباهت ساختاری آن‌ها را در تکنیک تعبیه حفظ کند. نتایج آزمایش‌ها نشان می‌دهد که RARE در وظایفی مانند طبقه‌بندی گره، پیش‌بینی لینک و بصری‌سازی شبکه، نسبت به روش‌های متداول عملکرد بهتری دارد .

در [۳۶] یک تکنیک بهبود یافته برای توصیه POI [3] در LBSN طراحی شد که شامل ساخت گراف‌های تعامل کاربر-POI، POI-POI و کاربر-کاربر بر اساس تاریخچه داده‌های ورود کاربر است. اطلاعات مشارکتی از سه نوع روابط استخراج شد و بردارهای جاسازی مشارکتی برای کاربر و POI جهت افزایش عملکرد توصیه آموخته شدند. با استفاده از بردار جاسازی مرتبه بالاتر از نمودار کاربر-کاربر، تأثیر کاربران و دوستان اجتماعی آنها بر یکدیگر، و همچنین تأثیر همسایگان جامعه بر کاربر و همسایگان نزدیک آنها، برای بهبود پراکندگی داده‌ها در نظر گرفته شد. در پژوهشی دیگر [۳۷] چارچوبی برای ترکیب هوموفیلی و نفوذ اجتماعی در پیش‌بینی محبوبیت پیشنهاد شد. یک شبکه عصبی گراف بدون نظارت برای تولید بازنمایی نفوذ اجتماعی با ترکیب مکانیسم توجه برای بدست آوردن روابط اجتماعی ناهمگن و جهت‌دار استفاده شد. این موضوع اهمیت مدلهای ترکیب نفوذ اجتماعی با هوموفیلی را در تعیین پیش بینی‌ها نشان داد. با مرور تحقیقات پیشین مشاهده می‌شود که دو ویژگی ساختار همبندی گراف‌ها و صفات خاصه‌ی گره‌های آن بطور توام در تعیین شباهت گره‌ها و تعیین نفوذ اجتماعی آنها استفاده نشده است. با استفاده از روش‌های نوین مانند تکنیک‌های تعبیه گره‌ها و ترکیب ویژگی‌های ساختاری و هوموفیلی، می‌توان به نتایج دقیق‌تر و کارآمدتری دست یافت. این تحقیقات نه تنها به درک بهتر پدیده‌های اجتماعی در شبکه‌های مبتنی بر مکان کمک می‌کنند، بلکه می‌توانند در کاربردهای عملی مختلف نیز مورد استفاده قرار گیرند.

۳- روش شناسی پژوهش

این مطالعه یک رویکرد جدید برای شناسایی کاربران تأثیرگذار در شبکه‌های اجتماعی مبتنی بر مکان با ترکیب شباهت ساختاری و ویژگی‌های هموفیلی پیشنهاد می‌کند. شکل (۱) ساختار گراف شبکه‌های مبتنی بر مکان مورد استفاده در تحقیق را نشان می‌دهد. رویکرد مورد استفاده در پژوهش، از یک فرآیند چند مرحله ای شامل جمع آوری و پیش پردازش داده ها، ساخت گراف، محاسبه انواع شباهت، محاسبه نفوذ اجتماعی و ارزیابی نتایج پیروی می کند.



شکل (۱): گراف ارتباط کاربران و مکان‌ها در یک شبکه اجتماعی مکان مبنا [۱۱]

[1] Graph Convolutional Network(GCN)
[2] Graph Neural Networks(GNN)
[3] Point-Of-Interest(POI)

۳-۱- رویکرد پیشنهادی

شکل (۲) چارچوب پیشنهادی این پژوهش را نشان می‌دهد. مطابق این چهارچوب مراحل انجام تحقیق تشریح می‌شود. در ادامه با رسم گراف شبکه اجتماعی، تحلیل شبکه به منظور شناسایی تاثیر اجتماعی و بررسی الگوهای تعاملات اجتماعی انجام می‌شود. در این مرحله، از تکنیک‌های زیر استفاده شده است:

- **محاسبه شباهت با تکنیک‌های پایه** : با استفاده از معیارهای پایه شباهت همچون ژاکارد و آدامیک/آدار، برای تعیین شباهت کاربران تأثیرگذار محاسبه انجام می‌گیرد.
- **محاسبه شباهت گره با تکنیک‌های تعبیه شده:** این بخش با استفاده از تکنیک‌های یادگیری عمیق مانند Node2Vec و DeepWalk برای استخراج ویژگی‌های پیچیده از داده‌های شبکه و سپس محاسبه شباهت بردارهای جانمایی گره‌ها با استفاده از شباهت کسینوسی انجام می‌شود.
- **تعیین هوموفیلی کاربران:** در این مرحله بررسی تأثیر هوموفیلی بر تعاملات کاربران با استفاده از روش‌های تعیین شباهت صفات خاصه صورت می‌گیرد.
- **محاسبه نفوذ اجتماعی کاربران:** در شبکه‌های اجتماعی مکان‌مبنا، نفوذ اجتماعی کاربران با استفاده از اطلاعات دوستانی که از مکان‌های مشابه بازدید می‌کنند، ارزیابی می‌شود. کاربران تأثیرگذار معمولاً با گروه‌های بیشتری در ارتباط هستند و تعداد دوستانی که از مکان‌های مشابه بازدید کرده‌اند، به عنوان معیاری برای محاسبه تأثیر اجتماعی آن‌ها استفاده می‌شود. این موضوع نشان می‌دهد که وقتی کاربران در مکان‌های مشترک قرار می‌گیرند، تأثیر بیشتری از خود نشان می‌دهند. نفوذ اجتماعی به دو صورت غیرمکانی و مکانی رخ می‌دهد. نفوذ اجتماعی غیرمکانی زمانی رخ می‌دهد که تأثیرات اجتماعی بدون محدودیت مکانی اتفاق بیفتند. برای مثال، اگر کاربری در یک کشور روی دکمه "لایک" یک صفحه فیس‌بوک کلیک کند و دوست او در کشور دیگری همین کار را انجام دهد، این نوع تأثیر به عنوان نفوذ غیرفضایی شناخته می‌شود. این نوع نفوذ به دلیل کمبود داده‌های مکانی، کمتر مورد تحلیل قرار می‌گیرد. نفوذ اجتماعی مکانی، برعکس، به مکان کاربران وابسته است. زمانی که کاربران در مجاورت یکدیگر قرار دارند، احتمال نفوذ اجتماعی افزایش می‌یابد. برای مثال، اگر کاربر و دوستانش در یک شهر زندگی کنند، تأثیر اجتماعی بیشتر خواهد بود. این نوع نفوذ را می‌توان به عنوان تأثیر اجتماعی مکانی-زمانی نیز در نظر گرفت، زیرا اثرات آن با زمان تغییر می‌کنند [۶].

۳-۲- الگوریتم پیشنهادی

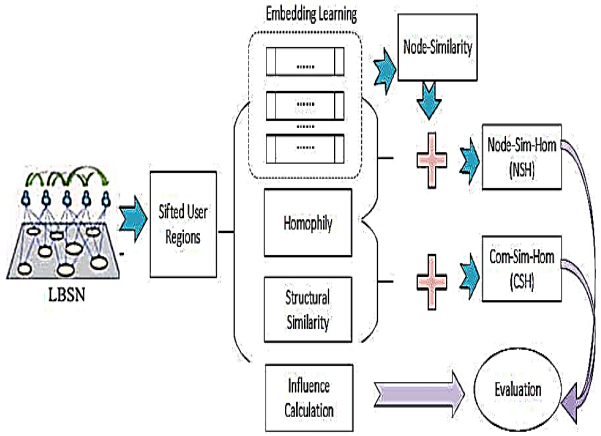
پس از تعیین محدوده داده‌ها و انتخاب کاربران از منطقه مورد نظر، نوبت به اجرای الگوریتم می‌رسد. در ابتدا، الگوریتم NSH شامل ترکیبی از هموفیلی و شباهت گره‌ها با استفاده از تکنیک‌های تعبیه گراف محاسبه می‌شود. ورودی این الگوریتم یک گراف بدون جهت از ارتباطات اجتماعی کاربران و یک گراف دوبخشی از بازدیدهای مکانی کاربران است و روی تمام کاربران شبکه اجرا می‌شود. معیارهای مبتنی بر یال مفاهیم مرتبط با نفوذ را بیان می‌کنند و بر روی یک جفت گره تعریف می‌شوند. آنها تعاملات ساده بین گره ها را توصیف می‌کنند. برای انجام این مرحله، جداسازی افرادی که با یکدیگر دوستی دارند (گراف دوستی G1) و به مکانی مشترک رفته‌اند (گراف بازدید مکانی G2) ضروری است. الگوریتم ۲ گره های کاندید را جهت انجام مراحل بعدی کار شناسایی می نماید. الگوریتم از رابطه کاربران مبنی بر وجود یا عدم دوستی کاربر استفاده می‌کند و سپس مکان‌های مشترک بین این کاربران را شناسایی می‌کند (خطوط ۷-۸). خروجی این الگوریتم کاربران کاندیدی هستند که برای مرحله بعدی بعنوان ورودی هستند. در روش پیشنهادی از دو تکنیک برای محاسبه شباهت ساختاری و ترکیب آن با هوموفیلی استفاده می‌شود، سپس روشهای مورد نظر بررسی و ارزیابی خواهند شد.

۳-۲-۱- محاسبه شباهت ساختاری با تکنیک های پایه

ابتدا شباهت هر جفت کاربر در گراف دوستی محاسبه می‌شود. به طور خاص، جفت دوستانی که قبلاً متصل شده بودند در نظر گرفته می‌شوند، بنابراین آنها یک سطح غیر صفر از شباهت را نشان دادند. ماتریس شباهت برای هر جفت کاربر در G1 ایجاد می‌شود. برای محاسبه شباهت از دو روش اساسی استفاده می‌شود. اولین شباهت الگوریتم Adamic/Adar است (رابطه ۱). این الگوریتم مبتنی بر این واقعیت است که همسایگان مشترک با تعداد همسایه‌های زیاد با وزن کمتری در محاسبه شباهت استفاده می‌شوند [۱۹و۱].

$$AD2(i,j)=\sum_{k\in N(i)\cap N(j)}\frac{1}{\log|N(k)|} \qquad (۱)$$

روش دوم شباهت به روش تشابه ارائه شده توسط زارعی [۱۳] نزدیک است. او با در نظر گرفتن اتصال مستقیم و غیرمستقیم گره‌ها، شباهت ساختاری هر دو گره را بر اساس تعداد همسایگان مشترک و همبستگی بین بردارهای همبستگی گره‌ها محاسبه کرد.



شکل (۲): رویکرد پیشنهاد شده

Algorithm 1: Pseudocode for Node-Sim-Hom (NSH)

1:***Input:*** *G1: friendship network, G2: check-ins network, Δt: threshold time*
2:***Output:*** *NSH similarity & Influence for each pair user*
3:***Begin***
4: ***for*** (every u_i at LBSN) ***do***
5: ***Begin***
7: ***Sifted Region_user (u_i)***
8: ***Find-Similar-Venue-Users (u_i)***
9: ***Embedding learning(u_i) : (Node2vec & DeepWalk)***
10: ***Calculate Combine Similarity (u_i, u_j)***
11: ***Calculate Influence (u_i, u_j)***
12: ***End for***
13: ***End***

Algorithm 2:Pseudocode for Find-Similar-Venue-Users

1:***Input:*** *G1: friendship network, G2: check-ins network*
2:***Output: Candid_List:*** *A set of candid_users*
3:***Begin***
4: ***while*** (Ch_i ∈ *Check-ins*) ***do***
5: ***Begin***
6: ***Candid-list = [,]***
7: ***for*** (each u_i in sifted area at LBSN) ***do***
8: ***Begin***
9: ***If (friend (u_i, u_j) .AND. (u_i.location = u_j.location)***
Candid_list = Candid_list ∪ (u_i , u_j)
10: ***Else***
11: ***Continue***
12: ***End for***
13: ***End while***
14: ***End***

Algorithm 3: Pseudocodes for Node2vec Technique

1: ***Learn_Features*** *(G1 = (V, E, W)*, *Dimensions d, Walks per node r, Walk length l, Context size k, Return p, In-out q)*
2: *π =* ***PreprocessModifiedWeights*** *(G, p, q)*
3: *G' = (V, E, π)*
4: ***Initialize walks*** *= 0*
5: ***for*** *(iter = 1 to r)* ***do***
6: ***for*** *(all nodes u ∈ V)* ***do***
7: ***Begin***
8: *walk =* ***Node2vecWalk(G', u, l)***
9: *Append walk to walks*
10: ***End***
11: *f =* ***StochasticGradientDescent(k, d, walks)***
12: ***Return f***

1: ***Node2vecWalk(*** *G1: friendship network (V, E , π), start node u, Length l):*
2: ***Initialize walk to [u]***
3: ***for (walk_ityer = 1 to l do***
4: ***Begin***
5: ***Curr = walk[-1]***
6: ***Vcurr = GetNeighbors(curr , G1)***
7: ***S = AliasSample(Vcurr*** *, π)*
8: ***Append S to walk***
9: ***End***
10: ***Return walk***

بر اساس این روش، از تکنیک مشابهی برای در نظر گرفتن ارتباط مستقیم بین گره ها استفاده می‌شود. بردارهای محاسبه همبستگی، بردارهای ردیفی از ماتریس مجاورت هر جفت گره در نظر گرفته می‌شوند. سپس در (۳) از همسایگان مشترک گره استفاده شده است. در نهایت، (۴) شباهت ساختاری مورد نظر را محاسبه می‌کند. نتیجه این رویکرد با سایر رویکردهای مبتنی بر شباهت سازگار است.

$$Corr(i,j)=\frac{\sum_{k\in A_i\cup A_j}\left(A_i[k]-\ddot{A}_i\right)\left(A_j[k]-\ddot{A}_j\right)}{\sqrt{\sum\left(A_i[k]-\ddot{A}_i\right)^2}\sqrt{\left(A_j[k]-\ddot{A}_j\right)^2}}: \quad (۲)$$

$$\forall\ k=1,2,...,|v|$$

$$\ddot{A}_i=\frac{\sum_{k\in(A_i\cup A_j)}A_i[k]}{|A_i\cup A_j|}$$

$$CN(i,j)=\left(\Gamma(u_i)\cap\Gamma(u_j)\right) \quad (۳)$$

$$St-Sim(i,j)=\left(1+CN(i,j)\right)\left(1+Corr(i,j)\right) \quad (۴)$$

در (۲)، A ماتریس مجاورت دوستی کاربران است. $\bar{A}$ میانگین مقادیر بردار A است و $\mathrm{Corr}(u_i,u_j)$ همبستگی دو گره را نشان می‌دهد. $\mathrm{CN}(u_i,u_j)$ در (۳) روش همسایگان مشترک می‌باشد[۱۳].

۳-۲-۲- محاسبه شباهت با تکنیک‌های تعبیه گره

در این رویکرد به روش‌های پیاده‌روی تصادفی توجه شده و کاربرد آن‌ها در شناخت نفوذ اجتماعی مورد بررسی قرار گرفته است. پیاده روی تصادفی یک رویکرد موثر برای انجام تعبیه گره‌ها و تعیین شباهت گره‌ها است. دو روش معروف تعبیه گره که در یادگیری ماشین جزء اولین روش‌های محاسبه بردار گره ها بوده اند در اینجا مورد استفاده قرار می‌گیرد.

الگوریتمNode2vec

این یک روش جاسازی گراف است که در آن نمایش عددی گره‌های گراف را براساس قدم‌زن تصادفی مرتبه دوم ایجاد می‌کند. قدم‌زن تصادفی مرتبه دوم نسخه اصلاح شده‌ی قدم‌زن تصادفی مرتبه اول است. این قدم‌زن در هر موقعیتی با توجه به وضعیت فعلی و قبلی خود گره بعدی را انتخاب می‌کند. در این الگوریتم با توسعه چنین قدم‌زنی، همسایگان گره‌ها به روش جستجوی سطحی[1] و یا با روش جستجوی عمقی[2] کاوش می‌شوند. الگوریتم ۳ مراحل کار را در این روش نمایش می‌دهد. [۲۵].

الگوریتم DeepWalk

DeepWalk تعبیه‌های اجتماعی (رئوس گراف محصور شده) را از طریق شبیه‌سازی یک مدل پیاده‌روی تصادفی کوتاه[3] می‌آموزد. ویژگی‌های پنهان رئوس، عضویت در جامعه و شباهت همسایگی را با گرفتن بازنمایی‌های اجتماعی نشان می‌دهند[۱۶]. الگوریتم ۴ (DeepWalk) شامل مراحل زیر است: یک رأس تصادفی v_i به طور تصادفی توسط تولیدکننده گام تصادفی نمونه برداری شده و به عنوان ریشه Wvi پیاده

[1] Breadth First Search (BFS)
[2] Depth First Search (DFS)
[3] Short random walk

روی تصادفی استفاده می‌شود. برای رسیدن به حداکثر طول (t)، نمونه‌های پیاده‌روی از آخرین همسایه‌های راس به طور مداوم جمع‌آوری می‌شوند. مدل Skip-Gram (الگوریتم ۵)، به بررسی هر گونه ترکیب احتمالی که ممکن است در طول یک پیاده‌روی تصادفی که در پنجره w یافت می‌شود رخ دهد، می‌پردازد. برای هر یک، هر رأس v_j به بردار نمایش فعلی خود $\Phi(v_j) \in R^d$ نگاشت می‌شود. هدف، به حداکثر رساندن احتمال همسایگان v_j در راه رفتن (خط ۳) از طریق نمایش آن است. استفاده از طبقه‌بندی‌کننده‌های متعدد رویکردی برای یادگیری چنین توزیع پسینی‌است[۲۱].

پس از محاسبه بردارهای تعبیه شده گره‌ها، نوبت به محاسبه شباهت گره‌ها با کمک روش‌های مبنای محاسبه شباهت برداری می‌رسد. در این قسمت از روش تشابه کسینوسی استفاده شد. اگر شباهت بین بردارهای جاسازی دو کاربر زیاد باشد، در نظر گرفته می‌شود که کاربر شانس بیشتری برای بازدید از مکان‌های مشابه دارد. معیارهایی مانند معیارهای فاصله وجود دارد که فاصله یا نزدیکی داده‌ها را از یکدیگر محاسبه می‌کند. بدیهی است که معیارهای تشابه با معیارهای فاصله رابطه معکوس دارند و به عبارت دیگر هر چه تشابه بیشتر باشد فاصله دو جسم کمتر می‌شود. یک کاربر با بردار به دست آمده با استفاده از روش تعبیه شده گره از ارتباطات اجتماعی خود مرتبط است. از آنجایی که نمی‌توانیم مستقیماً بردارها را مقایسه کنیم، بر روش‌شناسی متفاوتی تکیه می‌کنیم. برای اهداف خود، باید یک معیار تشابه بین بردارها تعریف کنیم. در این کار از شباهت کسینوسی[1] استفاده شد[۱۱]. تشابه کسینوسی معیاری است که برای تعیین شباهت اسناد صرفنظر از اندازه آنها استفاده می‌شود. از نظر ریاضی شباهت کسینوسی، کسینوس زاویه بین دو بردار را که در یک فضای چند بعدی نمایش داده می‌شود، اندازه‌گیری می‌کند. به طور خاص، شباهت بین دو کاربر u و v به صورت زیر تعریف می‌شود [۱۱و۱۸]:

$$Node\_Sim(u_i,u_j) = \frac{A.B}{||A||.||B||} = \frac{\sum_{i=1}^{n} A_i \times B_i}{\sqrt{\sum_{i=1}^{n} A_i^2} \times \sqrt{\sum_{i=1}^{n} B_i^2}} \quad (۵)$$

۳-۲-۳- تعیین شباهت خاصه

بخش دوم ترکیب مورد استفاده در رویکرد پیشنهادی یک روش شباهت صفت خاصه به نام هوموفیلی است. انواع مختلف رویکردها برای محاسبه هوموفیلی طراحی شده‌اند [۶و۱۱و۱۸و۱۹و۲۰]. در رویکرد پیشنهادی، برای محاسبه هوموفیلی از ماتریس شباهت فضایی ورودی کاربران استفاده می‌شود (خروجی الگوریتم ۱). گراف G2 ویژگی‌های کاربر را محاسبه می‌کند (به دلیل محدودیت داده‌های کاربران). سطرها و ستون‌های این ماتریس کاربران مورد نظر هستند. اجزای این ماتریس همپوشانی هر کاربر را با سایر کاربران نشان می‌دهد. رابطه (۶) هوموفیلی کاربران جفت را اعلام می‌کند.

$$Hom\_Sim(u_i,u_j) = \frac{|loc(u_i) \cap loc(u_j)|}{|loc(u_i) \cup loc(u_j)|} \quad (۶)$$

۳-۲-۴- الگوریتم شباهت ترکیبی

با استفاده از هر یک از دو تکنیک معرفی شده و روشهای تشابه مورد استفاده، ترکیب تشابه ساختاری و هوموفیلی، مطابق با روابط (۷) و (۸) محاسبه می‌شود:

$$CSH(u_i,u_j) = \alpha * Com_Sim(u_i,u_j) + (1-\alpha) * H\_sim(u_i,u_j) \quad (۷)$$

ضریب α یک نسبت ثابت است $0 \le \alpha \le 1$. بطور مثال $\alpha = 0.5$ به این معنی است که شباهت‌های ساختاری و هوموفیلی به طور مساوی در محاسبه شباهت دو کاربر نقش دارند. مقادیر به طور مداوم در بازه [۰،۱] محاسبه و نمایش داده می‌شوند. در این فرمول، شباهت از دو روش استفاده می‌کند که به صورت جداگانه محاسبه می‌شوند. حال با استفاده از الگوریتم ۶ و محاسبه فرمول‌ها، شباهت ترکیبی CSH برای هر جفت کاربر محاسبه می‌شود. در ادامه، روش NSH شامل ترکیب هوموفیلی و شباهت گره‌ها با استفاده از تکنیک‌های تعبیه گراف معرفی می‌شود. در رویکرد قبلی، تأثیر ترکیب شباهت ساختاری و هوموفیلی را در تعیین تأثیر اجتماعی کاربران با استفاده از روش‌های ترکیب شباهت پایه بررسی کردیم [۱۲]. در این تحقیق با استفاده از هر یک از دو روش تشابه جاسازی شده معرفی شده، ترکیب شباهت و هوموفیلی بردار گره‌ها محاسبه شده است.

$$NSH(u_i,u_j) = \alpha * Node\_Sim(u_i,u_j) + (1-\alpha) * H\_sim(u_i,u_j) \quad (۸)$$

Algorithm 4: Pseudocode for DeepWalk

```
Input: G1: friendship network, window size w,
embedding size d walks per vertex γ, walk length t
Output: matrix of vertex representations Φ ∈ R^{|V|×d}
Begin
Initialization: Sample Φ from U^{|V|×d}
Build a binary Tree T from V
for i = 0 to γ do
  Begin
  O = Shuffle(V)
  for each vi ∈ O do
    Begin
      Wvi = RandomWalk(G1, vi,t)
      SkipGram(Φ, Wvi , w)
    end for
  end for
end
```

Algorithm 5: Pseudocode for SkipGram

```
for each v_j ∈ Wv_i do
    for each u_k ∈ Wv_i [j − w : j + w] do
        J(Φ) = − log Pr(u_k | Φ(v_j))
        Φ = Φ − α * ∂J / ∂
    end for
end for
```

[1] Cosine similarity

Algorithm 6: Pseudocode to Combine-Similarity

1:***Input:*** *Candid_list from Algorithm2, G1: Friendship Network, G2: Check-ins Network*
2:***Output: NSH: Combine two Node&Attribute Similarity***
3: ***Begin***
4: α: [0,1]
5: ***for*** (*each* u_i *in Candid_List*), ***do***
6: ***Begin***
7: ***Calculate Nodes vector similarity*** (u_i ,u_j) ***by Eq.(1)***
8: ***Calculate Structural Similarity*** (u_i ,u_j) ***by Eq.(2-5)***
9: ***Calculate Homophily*** (u_i , u_j) ***using Eq.(6)***
10: ***Calculate NSH*** (u_i , u_j) ***using Eq.(7)***
11: ***Calculate CSH*** (u_i , u_j) ***using Eq.(8)***
12: ***End for***
13:***End***

Algorithm 7:Pseudocode for Influence Calculation

1:***Input:****Gc*: *Candid_list* , *G2*: *check-ins network* , *Δt*
2:***Output:*** *influence of pair users*
3:***Begin***
4: ***Initialization parameter:*** Δt = 30
5: ***for*** (*each* u_i *in Candid_List at the interval* Δ*t*) ***do***
6: ***Begin***
7: ***Compute Count_Chk using (Eq. 4)***
8: ***If*** (u_i ***.qloc*** = u_j ***.qloc*** ***) .AND.*** ($ch_i.u_i$!= $ch_j.u_j$) ***.AND.*** ($|u_i.timecheckin - u_j.timecheckin| < \Delta t$)
9: ***Calculate the influence*** (u_i,u_j) ***using Eq. (10)***
10: ***End for***
11: ***End***

روش Node_Sim به طور جداگانه بر روی بردارهای تعبیه‌شده از دو روش Node2vec و DeepWalk طبق الگوریتم های ۳ و ۴ محاسبه شد و سپس نتایج مورد تجزیه و تحلیل قرار گرفت. ضریب α یک نسبت ثابت و ۱≤ α≤۰ می‌باشد. با تغییر مقدار α در محدوده [۰,۱]، می‌توان سهم هر یک از بردارهای شباهت گره‌های حاصل از روش‌های تعبیه‌شده و نیز شباهت صفات خاصه یا هوموفیلی را بررسی کرد. در روش ترکیبی ارائه شده ۰.۵ = α به این معنی است که شباهت بردار گره‌ها و هوموفیلی به طور مساوی در محاسبه شباهت دو کاربر نقش دارند. مقادیر به طور مداوم در بازه [۰,۱] محاسبه و نمایش داده می‌شوند. همانطور که مشاهده می‌شود، ورودی الگوریتم ۶، لیستی از کاربران کاندید از الگوریتم ۲ و گراف‌های *G1* و *G2* است، پس از محاسبه شباهت ساختاری با هر یک از روش‌های معرفی شده در معادلات (۱) تا (۵) (خط ۸) و به صورت هوموفیلی تعریف شده در رابطه (۶) (خط ۹)، با تغییر مقدار α در بازه [۰,۱] (خط ۴)، مشارکت هر دو شباهت و هوموفیلی در رابطه‌های (۷و۸) بررسی می‌شود (خط ۱۰و۱۱).

۴-۲-۴- محاسبه نفوذ اجتماعی

تعیین نفوذ اجتماعی، با محاسبه تعداد دوستان کاربر که برای اولین بار بعد از او مکانی را در یک بازه زمانی Δt بازدید کرده‌اند، مشخص می‌شود (۹). تعداد به دست آمده برای تمام زمان‌های موثر بازدید کاربر بر تعداد مکان‌های بازدید شده توسط دو کاربر تقسیم شد و سپس عدد به دست آمده برای همه زمان‌های موثر بازدید کاربر در نظر گرفته شد (۱۰). به عنوان مثال، اگر u_2 از سه مکان متداول بعد از u_1 بازدید کرده باشد و مجموع مکان های بازدید شده توسط هر دوی آنها ۶ باشد، اثر ۰.۵ خواهد بود.

$$Count\_Chk\left(u_i, loc(k), \partial t\right) = \sum_{k=1}^{N}\left|Checkins\left(\Gamma(u_i), loc(k), \partial t\right)\right| \quad (۹)$$

N تعداد مکان‌هایی را نشان می‌دهد که u_iو به دنبال آن u_j بازدید کرده است. رابطه (۱۰) تأثیر کاربر اول روی کاربر دوم را محاسبه می‌کند:

$$Influence(u_i, u_j) = \begin{cases} \dfrac{Count-Chk\left(u_i, loc(k), \partial t\right)}{\left|loc(u_i) \cup loc(u_j)\right|} & If\ \ u_j\ \ check\_ins\ \ loc(k)\ \ after\ \ u_i \\ 0 & otherwise\ \ 0 \end{cases} \quad (۱۰)$$

برای محاسبه تأثیر الگوریتم پیشنهادی بر روی ناحیه مورد نظر، گره‌ها دارای رابطه دوستی هستند و در زمان‌های مختلف به مکان‌های مشابه رفته‌اند (خروجی الگوریتم ۲). اگر ui در زمان T از مکانی k بازدید کرده باشد و uj در بازه زمانی تعیین شده Δt و در زمان T+1 (بعد از ui) به Lock رفته باشد، می‌توان نتیجه گرفت که تأثیر رخ داده است. برای تعیین تأثیر یک کاربر بر کاربر دیگر، برای هر مکان بازدید شده یکی به شمارنده Count_Chk اضافه شد. سپس عدد بدست آمده بر تعداد مکان‌های بازدید شده توسط هر دو کاربر تقسیم شد. اگر نتیجه ۱ بود، ui در جاهایی که با uj به آن مراجعه نموده اند، uj را تحت تأثیر قرار داده اند. اگر صفر بود یعنی تاثیری بر روی یکدیگر نداشته‌اند. عدد تأثیر ۰.۵ به این معنی است که کاربر در نیمی از مکان‌هایی که با هم رفته اند، uj را تحت تأثیر قرار می‌دهد.در این قسمت محاسبه نفوذ کاربر توسط الگوریتم ۷ استفاده شده است. در اینجا از لیست کاندیدها با مقدار تشابه ترکیبی که خروجی الگوریتم ۲ بوده به عنوان ورودی استفاده شده است. سپس برای تمامی داوطلبان دارای شرایط مورد نیاز (خط ۱) تعداد ورودهای کاربر توسط رابطه (۹) محاسبه شد (خط ۷) و محاسبه نفوذ کاربر در صورت وجود شرایط توسط رابطه (۱۰) انجام شد (خط ۹). در آزمایش‌ها، دوره آستانه زمانی Δt برای ۳۰ روز در نظر گرفته شد (خط۴). بنابراین، اگر u2 (دوست کاربر u1) بعد از u1 در دوره آستانه به یک مکان مشترک رفته و قبل از آن بازدیدی از آن مکان صورت نگرفته باشد، می‌توان نتیجه گرفت که نفوذ رخ داده است.

۴- ارزیابی عملکرد و نتایج

در این بخش الگوریتم پیشنهادی با استفاده از معیارهای ارزیابی مختلف در این زمینه مورد مقایسه قرار می‌گیرد. در ادامه، ابتدا تنظیمات شبیه‌سازی و سپس چند معیار برای ارزیابی نتایج روش پیشنهادی معرفی می‌گردد. همچنین برای ارزیابی الگوریتم ارائه شده و مقایسه آن با سایر روش‌ها، مشخصات مجموعه داده‌های واقعی مورد استفاده آورده شده است. نتایج ارائه شده در این بخش با استفاده از کتابخانه‌های ارزشمند زبان پایتون به عنوان ابزاری برای تجزیه و

تحلیل داده‌ها و گراف‌های شبکه‌های اجتماعی توسعه یافته است. همچنین برای آزمایشات از پردازنده ۲.۸۰ گیگاهرتزی corei7 با حافظه ۱۶ گیگابایتی استفاده شده است. برای انجام این تحقیق، از داده‌های واقعی مربوط به شبکه‌های اجتماعی مکان‌مبنا استفاده شده است. داده‌ها شامل اطلاعات مربوط به کاربران، مکان‌ها و تعاملات آنها در سکوهای مختلف مانند Foursquare و Brightkite و Gowallaمی‌باشد. این داده‌ها به صورت عمومی در دسترس بوده و از طریق API های مربوطه جمع‌آوری شده‌اند. قبل از تحلیل داده‌ها، مراحل پیش‌پردازش انجام شده است. این مراحل شامل: حذف داده‌های ناقص و غیرمعتبر، استخراج ویژگی‌های کلیدی کاربران و مکان‌ها، از جمله تعداد دوستان، تعداد بازدیدها و ویژگی‌های جغرافیایی و غربال کردن داده‌های مربوط به سه منطقه پربازدید می باشد.

عملکرد رویکرد پیشنهادی از طریق چند معیار ارزیابی می شود: معیار همبستگی پیرسون که همبستگی نفوذ اجتماعی و شباهت‌ها را تحلیل می‌کند. همچنین رگرسیون خطی که تاثیر متغیرهای پیش‌بینی را بر متغیر وابسته تحلیل می‌کند. در ادامه روش مورد نظر با چندین روش جایگزین مقایسه و نتایج را مورد بررسی و تحلیل قرار می‌دهیم.

۴-۱- ضریب همبستگی پیرسون[1]

ضریب همبستگی پیرسون پارامتری برای مقایسه عملکرد روش‌های مختلف است. هنگام تعیین درجه، نوع و جهت رابطه بین دو متغیر مورد بررسی، از این ضریب محاسبه استفاده می‌شود. این ضریب برای داده‌هایی با توزیع نرمال و اعداد زیاد استفاده می‌شود. محاسبه مقدار ضریب پیرسون بر اساس رابطه (۱۱) انجام می‌شود:

$$r = \frac{\sum(x_i - x')(y_i - y')}{\sqrt{\sum(x_i - x')^2 \sum(y_i - y')^2}} \quad (11)$$

که در آن x و y مقادیر متغیر x و y در داده های نمونه هستند. همچنین x و y میانگین مقادیر آنها هستند [۱۱]. ضریب همبستگی پیرسون برای داده‌های پیوسته و نرمال استفاده می‌شود. فرض می‌کند که رابطه بین دو متغیر خطی است. به مقادیر دقیق داده‌ها حساس است، از میانگین و انحراف معیار برای محاسبه استفاده می‌کند و به سادگی محاسبه می‌شود.

۴-۲- رگرسیون خطی[2]

رگرسیون خطی به معنای پیش‌بینی یک متغیر وابسته بر اساس یک یا چند متغیر مستقل است. در این روش تأثیر همزمان متغیرهای پیش‌بینی کننده بر متغیر وابسته بررسی می‌شود. به عبارت دیگر، در این روش بر خلاف روش همبستگی، تنها روابط دو به دو در نظر گرفته نشده و همه متغیرها با هم تحلیل می‌شوند. در رگرسیون خطی ساده، اگر Y را به عنوان متغیر وابسته و X را به عنوان متغیر مستقل در نظر بگیریم، معادله خط رگرسیون را می‌توان به صورت (۱۲) نوشت:

$$Y = A + B X \quad (12)$$

A و B پارامترهای خط نامیده می‌شوند و می‌توان آنها را تخمین زد [۱]. متغیر مستقل متغیری است که توسط آزمایشگر کنترل شده و با X نمایش داده می‌شود و متغیر وابسته متغیری است که مقدار آن به X بستگی دارد و با Y نشان داده می‌شود و متغیر اثر نامیده می‌شود.

۴-۳- مجموعه داده

سه مجموعه داده در اینجا تحلیل می‌شوند: اولین مجموعه داده از Gowalla جمع‌آوری شد، یک وب‌سایت شبکه اجتماعی مبتنی بر مکان که در آن کاربران مکان‌های خود را از طریق چک کردن به اشتراک می‌گذارند[۳۸]. شبکه دوستی یک گراف بدون جهت است که با استفاده از API عمومی آنها ساخته شده و از ۱۹۶,۵۹۱ گره و ۹۵۰,۳۲۷ یال تشکیل شده است. این مجموعه داده شامل ۶,۴۴۲,۸۹۰ کاربر است که بین فوریه ۲۰۰۹ تا اکتبر ۲۰۱۰ ثبت حضور کرده‌اند. مجموعه داده دوم از Brightkite جمع‌آوری شد که زمانی ارائه دهنده خدمات شبکه اجتماعی مبتنی بر مکان بود که در آن کاربران مکان‌های خود را به اشتراک می‌گذاشتند[۳۹]. شبکه دوستی با استفاده از API عمومی آنها ارائه شد و شامل ۲۲,۸۵۸ گره و ۲۱۴,۰۷۸ یال بدون جهت بود. مجموعه داده همچنین شامل ۴,۴۹۱,۱۴۳ بررسی این کاربران از آوریل ۲۰۰۸ تا اکتبر ۲۰۱۰ است. مجموعه داده بعدی Foursquare شامل ۲,۱۵۳,۴۷۱ کاربر، ۱,۱۴۳,۰۹۲ مکان بازدید شده، ۱,۰۲۱,۹۷۰ اعلام حضور، و ۲۷,۰۹۸,۴۹۰ از ارتباطات اجتماعی کاربران با کمک API آن جمع‌آوری شده است[۴۰]. جزئیات ساختار شبکه این سه نوع مختلف مجموعه داده طبق جدول (۱) آمده است.

۴-۴- آزمایشات و تحلیل نتایج

در این بخش، رویکرد ارائه شده در بخش ۳-۱ مورد ارزیابی قرار گرفت. ابتدا تنظیمات شبیه‌سازی و معیارهای عملکرد تعریف شده و سپس نتایج تجربی مورد بحث قرار می‌گیرد. برای تحلیل و اجرای رویکرد NSH از روش‌های زیر استفاده شد:

- ***DeepWalk*** [۲۱]: توالی گره‌ها را با استفاده از راه رفتن‌های تصادفی ایجاد می‌کند، و سپس آنها را به مدل *Skip-Gram* برای خروجی جاسازی گره ها تغذیه می‌کند. طول راه رفتن را l = ۴۰، تعداد راه رفتن در هر گره را r = ۸۰، *workers* =۱ و اندازه پنجره زمینه را برای مدل *Skip-Gram* روی k = ۱۰ قرار می‌دهیم.
- ***Node2vec*** [۲۵]: با معرفی یک روش پیاده‌روی تصادفی پارامتریک برای متعادل‌کردن استراتژی‌های جستجوی عرض-اول (پارامتر بازگشتی p) و جستجوی عمق-اول (پارامتر داخل-خارج q) گسترش می‌دهد تا ساختارهای گراف غنی‌تر را به تصویر بکشد. با توجه به پیشنهادات در مقاله اصلی، ابعاد=۶۴ را تنظیم کرده و سایر پارامترها را *r=200*، *l=10*، *k=10* و *workers* =۱ تنظیم می‌کنیم.

[1] Pearson correlation

[2] Linear regression

جدول (۱): مجموعه داده ها

مجموعه داده‌ها	دوره زمانی	نودها	یال ها	بازدیدها
Gowalla	۲۰۰۹ تا ۲۰۱۱	۵۹۱,۱۹۶	۳۲۷,۹۵۰	۸۹۰,۴۴۲,۶
Brightkite	۲۰۰۹ تا ۲۰۱۰	۲۲۸,۵۸	۰۷۸,۲۱۴	۱۴۳,۴۹۱,۴
Foursquare	۲۰۱۰ تا ۲۰۱۲	۴۷۱,۱۵۳,۲	۴۹۰,۰۹۸,۲۷	۹۷۰,۰۲۱,۱

جدول (۲): نواحی غربال شده از سه مجموعه داده

مجموعه Gowalla			
ناحیه	نودها	یال ها	بازدیدها
Texas	۱,۳۸۸	۲۷,۸۹۴	۲۴۳,۸۶۲
California	۱,۳۲۸	۱۹,۶۱۰	۱۷۶,۹۱۰
New York	۶۵۴	۷,۷۱۰	۴۲,۰۷۲
Brightkite			
ناحیه	نودها	یال ها	بازدیدها
Texas	۶۱۴	۷,۵۷۳	۵۶,۱۴۵
California	۱,۰۶۸	۱۴,۴۸۸	۱۵۵,۰۰۵
New York	۵۴۲	۴,۹۹۵	۳۱,۵۱۲
Foursquare			
ناحیه	نودها	یال ها	بازدیدها
Texas	۲۳۹	۷,۵۷۳	۵۶,۱۴۵
California	۱۴۹۵	۱۴,۴۸۸	۱۵۵,۰۰۵
New York	۱۱۶۴	۴,۹۹۵	۳۱,۵۱۲

CSH [۱۲]: تشابه ساختاری ترکیبی و هوموفیلی که با استفاده از روشهای پایه شباهت ساخته شده شده است. در این روش ترکیبی، شباهت ساختاری گره‌ها با استفاده از دو روش پایه به طور جداگانه به‌همراه شباهت ویژگی‌های خاص آنها محاسبه می‌شود. در این آزمایش برای به دست آوردن نتایج دقیق‌تر، از داده‌های سه منطقه‌ی با بیشترین بازدید مکانی و بیشترین تعداد کاربر، یعنی تگزاس، نیویورک و کالیفرنیا در شبیه‌سازی استفاده شد. جدول۲ داده‌های سه منطقه را برای مجموعه داده‌های Gowalla، Brightkite و Foursquare نشان می‌دهد.

AGCN [۳۳]: یک نوع تطبیقی از GCN که به صورت پویا ساختار گراف را بر اساس الگوهای نفوذ به‌روزرسانی می‌کند و امکان انتخاب دقیق‌تر کاربران تأثیرگذار را در سناریوهای مبتنی بر انتشار فراهم می‌کند. این روش، یادگیری ساختاری و مدل‌سازی نفوذ را از طریق اصلاح تکراری گراف متعادل می‌کند.

IGCN-POI [۳۶]: یک شبکه کانولوشن گراف بهبود یافته (IGCN) برای توصیه نقاط مورد علاقه شخصی (POI) در شبکه‌های اجتماعی مبتنی بر مکان. با ساخت یک گراف دوبخشی کاربر-POI و ادغام الگوهای بازدید مشترک مکانی، این مدل نمایش‌های جاسازی را یاد می‌گیرد که نشان‌دهنده تحرک و ترجیحات مکانی کاربران است. در حالی که در درجه اول برای توصیه طراحی شده است، جاسازی‌های کاربر آموخته شده، روابط اجتماعی و مکانی نهفته را رمزگذاری می‌کنند و این رویکرد را برای چارچوب‌های تخمین نفوذ همانند چارچوب پیشنهادی که شباهت ساختاری و جغرافیایی را ترکیب می‌کنند، مرتبط می‌سازد.

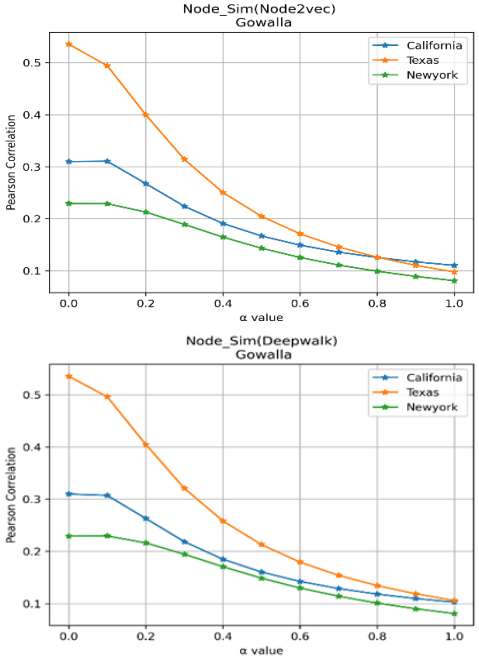
۴-۵- تحلیل همبستگی نتایج

برای ارزیابی رویکرد پیشنهادی، کارایی روش تحت نواحی انتخابی مختلف بر روی سه مجموعه داده بررسی شده، کاربران مشترک در گرافهای *G1* و *G2* برای دریافت نتایج قابل قبول انتخاب می‌شوند. برای افزایش دقت نتایج، آزمایش‌هایی بر روی نواحی غربال شده از سه مجموعه داده نمونه انجام می‌شود (جدول ۲). داده‌های کاندید بدست آمده از ناحیه انتخاب شده مربوط به هر یک از مجموعه داده‌ها، تحت ترکیب هوموفیلی و هر یک از دو تکنیک تشابه معرفی‌شده در بخش ۳-۱ محاسبه می‌شوند و نتایج همبستگی با تأثیر اجتماعی کاربران تجزیه و تحلیل می‌شود. از آنجایی که داده‌های فقط شامل بررسی‌ها و ارتباطات دوستی می‌شود، داده‌های حاوی نمایه‌های کاربر یا اطلاعات دیگر در دسترس نیستند.

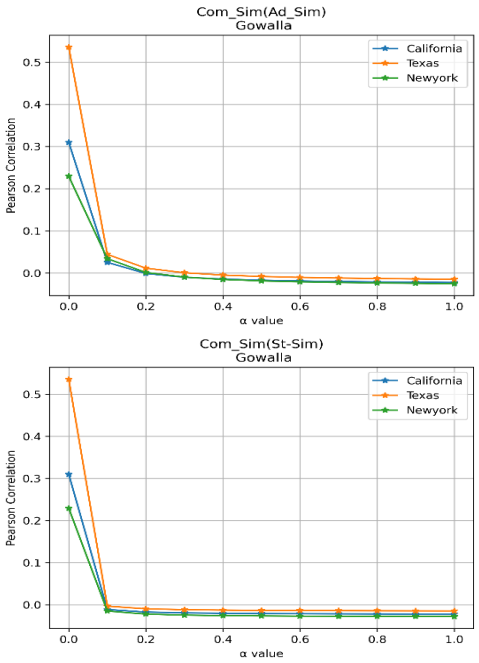
ابتدا شباهت ساختاری و شباهت بردار گره‌ها در گراف *G1* محاسبه شد. پس از ترکیب بردار شباهت گره‌ها و هوموفیلی برای هر جفت کاربر نامزد، میزان نفوذ اجتماعی برای آنها با استفاده از گراف *G2* محاسبه گردید. خروجی الگوریتم ۱ ترکیب شباهت برداری گره‌ها و هوموفیلی جفت‌های کاربر را در قالب الگوریتم NSH نشان می‌دهد. در ادامه، تأثیر اجتماعی جفت کاربر محاسبه شده و نتایج با کمک ضریب همبستگی برای نشان دادن شدت رابطه و نوع رابطه (مستقیم / معکوس) بررسی شد. با محاسبه مقدار همبستگی نفوذ اجتماعی و روش NSH با استفاده از ضریب همبستگی پیرسون، نمودارهای زیر در اشکال (۳ تا ۸) گزارش شده است. با در نظر گرفتن مقادیر α در محدوده $[0,1]$، به این نتیجه رسیدیم که شباهت ترکیبی و هوموفیلی کاربران طبق مشاهدات با نفوذ اجتماعی همبستگی مثبت دارد. با افزایش این مقدار، همبستگی کاهش می‌یابد، تا زمانی که برای شباهت برداری گره ها به صفر نزدیک می‌شود. نتایج روش پیشنهادی NSH، از جمله ترکیب بردار شباهت گره‌ها با هوموفیلی در مناطق مختلف انتخاب شده از هر سه مجموعه داده، نرخ همبستگی خوبی را نسبت به روش‌های دیگر نشان داد.

شکل (۳) نشان می‌دهد که هرچه مقدار α به صفر نزدیکتر باشد، تأثیر اجتماعی و روش ترکیبی پیشنهادی همبستگی مثبت خواهند داشت. در واقع تأثیر اجتماعی و روش پیشنهادی در یک جهت حرکت می‌کنند. بنابراین، هوموفیلی در روش پیشنهادی تأثیر مهمی دارد و با افزایش سطح آن، همبستگی بین روش پیشنهادی و تأثیر اجتماعی افزایش می‌یابد. این بدان معناست که نتایج به‌دست‌آمده از روش‌های ترکیبی تعبیه‌شده نیز به وضوح تأثیر شباهت ویژگی‌های گره‌ها را در تأثیر اجتماعی آنها، مشابه روش‌های کلاسیک نشان می‌دهد. شکل (۴) همبستگی تأثیر اجتماعی و روش CSH را برای مناطق انتخاب شده از مجموعه داده Gowalla را تحت دو روش شباهت پایه نشان می‌دهد. در تمامی مناطق غربال شده، نتایج نشان می‌دهد که در α بزرگتر از ۰.۱ نتایج ثابت می‌ماند و با افزایش هوموفیلی، میزان تاثیر اجتماعی افزایش می‌یابد. یعنی شباهت فضایی افراد در تأثیرگذاری اجتماعی مؤثر خواهد بود.

شکل (۵) همبستگی بین نفوذ اجتماعی و روش NSH را در سه منطقه نمونه از مجموعه داده Brightkite بر اساس انواع دو روش شباهت انتخابی و هوموفیلی نشان می‌دهد. همچنین مشابه دو تحلیل انجام شده قبلی نتایج در هر سه منطقه انتخاب شده روند مشابهی دارند. مشاهده میشود که نتایج به دست آمده با روش های Nod2vec و DeepWalk بسیار نزدیک است. شکل (۶) مشابه شکل (۵) است که همبستگی بین نفوذ اجتماعی و روش CSH را در مجموعه داده Brightkite با روش‌های Ad_Sim و St_Sim نشان می‌دهد.

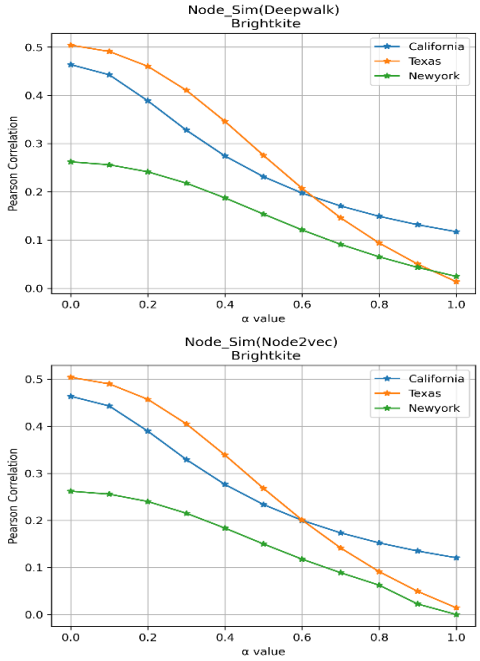


شکل (۳): همبستگی نفوذ اجتماعی و روش NSH درمجموعه داده Gowalla

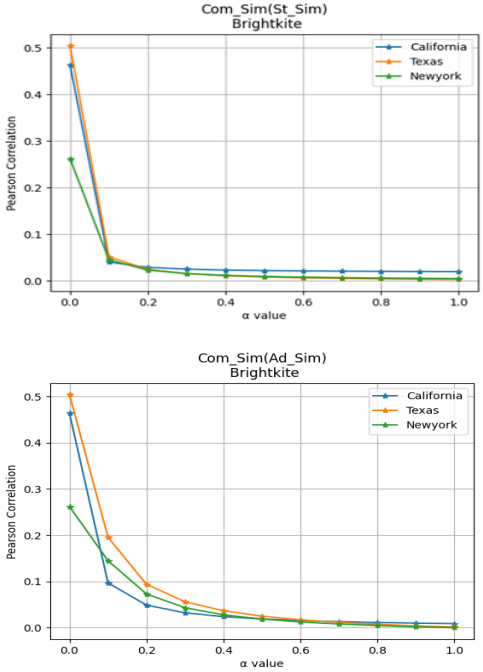


شکل (۴): همبستگی نفوذ اجتماعی و روش CSH درمجموعه داده Gowalla

در کلیه این نمودارها تأثیر هوموفیلی کاربران بر تأثیر اجتماعی قابل مشاهده است.شکل‌های (۷) و (۸) همبستگی تأثیر اجتماعی را برای سه منطقه نمونه انتخاب شده از مجموعه داده‌های Foursquare به ترتیب، با روش های NSH و CSH نشان می‌دهند. می‌توان دریافت که نتایج به دست آمده در هر سه منطقه این مجموعه داده مشابه دو مجموعه داده قبلی است که نشان‌دهنده این واقعیت است که میزان تأثیر اجتماعی با افزایش هوموفیلی افزایش می‌یابد. یعنی تشابه صفات خاصه افراد در نفوذ اجتماعی مؤثر خواهد بود.



شکل (۵): همبستگی نفوذ اجتماعی و روش NSH درمجموعه داده Brightkite



شکل (۶): همبستگی نفوذ اجتماعی و روش CSH درمجموعه داده Brightkite

با مشاهده نتایج شکل‌های (۷) و (۸) می‌توان دریافت که نمودارهای به دست آمده در هر سه منطقه این مجموعه داده مشابه دو مجموعه داده قبلی است که نشان‌دهنده این واقعیت است که میزان تأثیر اجتماعی با افزایش هوموفیلی افزایش می‌یابد. یعنی تشابه صفات خاصه افراد در نفوذ اجتماعی مؤثر خواهد بود.

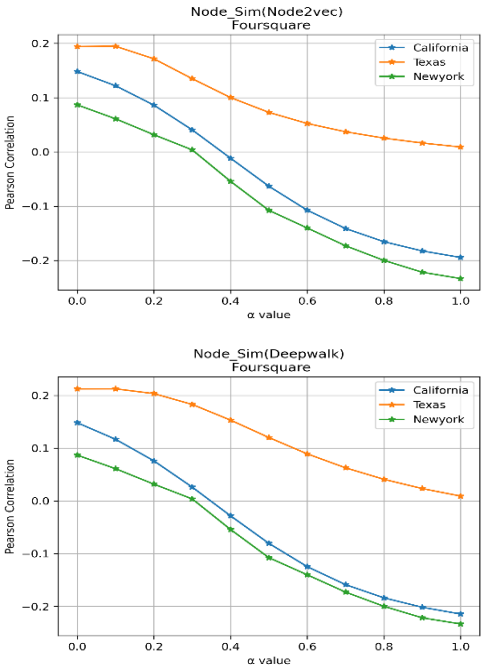
۴-۵-۱- تحلیل رگرسیون

ابتدا رابطه بین متغیرها با محاسبه ضریب همبستگی اثبات شد و سپس شدت آن‌ها با استفاده از رگرسیون اندازه‌گیری شد. معادله خط رگرسیون در (۱۲) در بخش (۴-۲) بیان شد. در اینجا، متغیر مستقل شباهت پیشنهادی و متغیر وابسته تأثیر اجتماعی در نظر گرفته می‌شود و متغیر اثر نام دارد. با توجه به مجموعه داده‌های مورد مطالعه، با محاسبه رگرسیون برای ستون شباهت پیشنهادی با توجه به تأثیر اجتماعی، نتایج به دست آمده صحت همبستگی به دست آمده را برای دو متغیر مورد مطالعه نشان می‌دهد. شکل (۹) تجزیه و تحلیل رگرسیون را برای روش NSH نشان می‌دهد که از داده‌های Nod2Vec برای ایالت تگزاس از مجموعه داده Gowalla به دست آمده است.

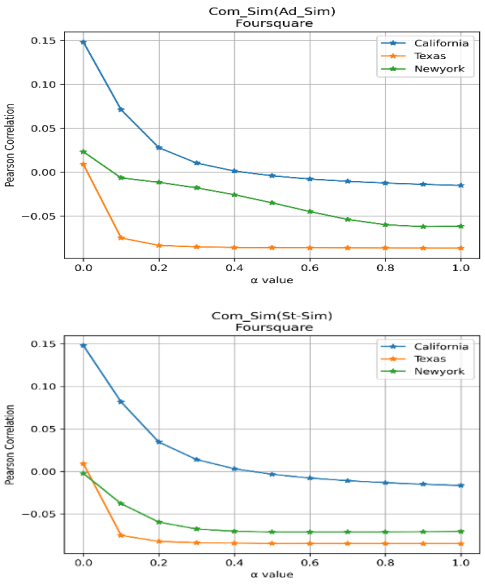
برای بررسی درستی نتایج باید مقدار P و F را بررسی نماییم. اگر این مقادیر کمتر از ۰.۰۵ باشند نتیجه میگیریم که نتایج بدست آمده معنی‌دار هستند. از نتایج شکل (۹) می‌توان دریافت که از آنجایی که مقدار P یا خطا برابر با صفر است، یعنی ضریب همبستگی معنی‌دار بوده و رگرسیون به دست آمده اعتبار دارد. بنابراین نتیجه می‌گیریم که رابطه X و Y و همچنین معادله Y= 0.029+1.364X قابل قبول است. به همین ترتیب، برای بقیه داده‌ها تحلیل و کارایی روش پیشنهادی مورد ارزیابی قرار گرفت.

۴-۵-۲- مقایسه با سایر روش ها

برای نشان دادن اثربخشی نسبی چارچوب NSH پیشنهادی، آن را با سه روش جایگزین مقایسه کردیم (مطابق با جدول ۳). CSH (مدل ترکیبی شباهت پایه و هوموفیلی)، AGCN-IM (حداکثرسازی تأثیر مبتنی بر یادگیری عمیق) وIGCN-POI. هر روش با استفاده از دو معیار عملکرد ارزیابی شد: (۱) همبستگی پیرسون بین شباهت تخمینی و تأثیر تجربی، و (۲) امتیاز توافق که سازگاری سطح رتبه را اندازه‌گیری می‌کند. مدل NSH ما، عملکرد خوبی با همبستگی ۰.۲۱ و امتیاز توافق ۸۰.۲٪ نشان داد. این موضوع نشان می‌دهد که ترکیب جاسازی ساختاری با هموفیلی جغرافیایی به طور مؤثر هر دو ابعاد توپولوژیکی و مکانی تأثیر اجتماعی را در بر می‌گیرد. برخلاف مدل‌هایی که نیاز به نظارت دارند، NSH همچنان قابل تفسیر و کارآمد باقی می‌ماند. CSH که به شباهت‌های ساختاری و ویژگی سنتی متکی است، عملکرد کمی پایین‌تری داشت (همبستگی ۰.۱۸-۰.۱۹، توافق ۷۴-۷۶٪)، که مجدداً بر مزیت یادگیری بازنمایی نسبت به معیارهای شباهت دست‌ساز تأکید می‌کند. AGCN-IM، یک مدل کانولوشن گراف با مکانیزم‌های توجه، بالاترین دقت (همبستگی ۰.۲۳، توافق ۸۲.۷٪) را تولید کرد، اما به قیمت پیچیدگی آموزش و تفسیرپذیری پایین. ماهیت یادگیری عمیق



شکل (۷): همبستگی نفوذ اجتماعی و روش NSH در مجموعه داده Foursquare



شکل (۸): همبستگی نفوذ اجتماعی و روش CSH درمجموعه داده Foursquare

Regression Statistics								
Multiple R	0.535642							
R Square	0.386912							
Adjusted R Square	0.386871							
Standard Error	0.036771							
Observations	17170							
ANOVA								
	df	*SS*	*MS*	*F*	*Significance F*			
Regression	1	9.339628339	9.339628	6907.577	0			
Residual	17168	23.21258697	0.001352					
Total	17169	32.55221531						
	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*Lower 95.0%*	*Upper 95.0%*
Intercept	0.029013	0.000309598	93.71091	0	0.0284059	0.02962	0.028406	0.02962
influence	1.364853	0.016421881	83.11184	0	1.33266413	1.397041	1.332664	1.397041

شکل (۹): تحلیل رگرسیون نفوذ اجتماعی و روش NSH در Gowalla

روش، آن را برای کاربردهای بدون نظارت یا بلادرنگ کمتر کاربردی می‌کند. مدل IGCN-POI، در حالی که برای توصیه مکان توسعه داده شده است، شباهت ساختاری با NSH در استفاده از جاسازی‌های مبتنی بر GCN دارد. همبستگی تخمینی (۰.۱۹) و توافق (۷۵.۰٪) آن، کاربرد بالقوه آن را برای تخمین تأثیر نشان می‌دهد، اگرچه کمتر برای این کار مناسب است. به طور خلاصه، NSH یک تعامل مؤثر بین دقت، تفسیرپذیری و کارایی ایجاد می‌کند و یک راه حل قوی و مقیاس‌پذیر برای تخمین تأثیر در شبکه‌های اجتماعی مبتنی بر مکان ارائه می‌دهد.

جدول (۳): مقایسه روش پیشنهادی با سایر روشها

روش	متغیر نفوذ	ترکیب شباهت	همبستگی پیرسون	امتیاز توافق (%)
NSH (Node2Vec + H_sim)	نفوذ نرمال شده بر مبنای ورود کاربران	ترکیب وزن‌دار تکنیک‌های تعبیه (Node2vec هوموفیلی +(فضایی	0.20	78.4
NSH (DeepWalk + H_sim)	نفوذ نرمال شده بر مبنای ورود کاربران	ترکیب وزن‌دار تکنیک‌های تعبیه (Deepwalk هوموفیلی +(فضایی	0.21	80.2
CSH (Ad_Sim + H_sim)[7]	نفوذ نرمال شده بر مبنای ورود کاربران	شباهت توپولوژیکال + هوموفیلی فضایی	0.18	74.0
CSH (St_Sim + H_sim)[7]	نفوذ نرمال شده بر مبنای ورود کاربران	شباهت توپولوژیکال + هوموفیلی فضایی	0.19	76.5
AGCN-IM [33]	مدل IC یاد گرفته شده	شباهت هسته‌بندی شده از طریق GCN + مکانیسم توجه	0.23	82.7
IGCN-POI[36]	تأثیر ضمنی از طریق تعامل کاربر-مکان (POI)	شباهت مبتنی بر جاسازی از بهبود GCN یافته	0.19	75.0

۴-۵-۳- پیچیدگی زمانی

پیچیدگی زمانی الگوریتم‌های روش پیشنهادی به شرح زیر است: ماتریس شباهت فضایی کاربران در الگوریتم ۲ با پیچیدگی زمانی O(nm+mk) تعیین می‌شود، که در آن n تعداد کاربران، m دوستان کاربر و k تعداد مکان‌های ورود کاربران است. الگوریتم‌های ۳ و ۴ که روش‌های جاسازی‌شده را محاسبه می‌کنند، پیچیدگی اجرای پیاده‌روی‌های تصادفی به تعداد گره‌ها و طول پیاده‌روی‌ها بستگی دارد. با داشتن n گره و اجرای r پیاده روی تصادفی به طول l، پیچیدگی کلی برای پیاده روی تصادفی O(n * r * l) است.همچنین پیچیدگی الگوریتم‌های ۶و۷ O(mk) است.

۵- نتیجه‌گیری

این پژوهش نشان داد که ترکیب شباهت‌های ساختاری و ویژگی‌های صفات خاصه کاربران (هوموفیلی) موجب شناسایی دقیق‌تر کاربران با نفوذ اجتماعی بالا می‌شود. در این میان، هرچه وزن هوموفیلی در شباهت ترکیبی افزایش یابد، تأثیر اجتماعی نیز بیشتر می‌شود. این دانش نشان‌دهنده اهمیت شباهت ویژگی‌های کاربران در مدیریت نفوذ اجتماعی آنها است. به‌ویژه، از آنجایی که در بسیاری از شبکه‌های اجتماعی نظیر LBSN ها همبستگی‌های جغرافیایی و اجتماعی میان کاربران اثرات قابل توجهی بر رفتارهای اجتماعی دارند، ترکیب این شباهت‌ها دقت مدل را بهبود می‌بخشد. استفاده از تکنیک‌های تعبیه گره در تحلیل ساختار شبکه و شباهت‌های ساختاری، نتایج مثبتی را در مقایسه با روش‌های شباهت معیار نشان داد. در حقیقت، استفاده از مدل‌های تعبیه گره باعث بهبود دقت شباهت‌های ساختاری و تشخیص بهتر روابط پیچیده میان گره‌ها شد. هرچند که در برخی موارد، روش‌های مبتنی بر شباهت معیار نیز توانسته‌اند نتایج مشابهی ارائه دهند، ولی مدل‌های تعبیه گره در پردازش داده‌های پیچیده‌تر و بزرگ‌تر از نظر محاسباتی مزیت دارند. نتایج تحلیل هوموفیلی نشان داد که این عامل در تعیین نفوذ اجتماعی کاربران شبکه‌های مکان‌مبنا تأثیر قابل توجهی دارد. ترکیب شباهت ساختاری و هوموفیلی به‌ویژه در مواقعی که کاربران دارای ویژگی‌های مشابه از نظر جغرافیایی یا اجتماعی هستند، موجب افزایش دقت در شناسایی کاربران با نفوذ بیشتر می‌شود. این یافته‌ها بیانگر این است که ویژگی‌های شخصی و اجتماعی کاربران، علاوه بر روابط ساختاری آنها در شبکه، نقش تعیین‌کننده‌ای در تحلیل رفتارهای اجتماعی ایفا می‌کنند. با توجه به یافته‌های این پژوهش، پیشنهاد می‌شود که برای بهبود دقت مدل‌های شناسایی تأثیر اجتماعی، استفاده از گراف‌های جهت‌دار و در نظر گرفتن وزن‌های متفاوت برای یال‌ها در گراف دوبخشی (کاربر-مکان) مورد بررسی قرار گیرد. همچنین، استفاده از ویژگی‌های متنوع‌تر کاربران در تحلیل شباهت‌ها می‌تواند به‌ویژه در شبکه‌های اجتماعی بزرگ‌تر و پیچیده‌تر، کارایی مدل را افزایش دهد. در نهایت، تحقیقات آینده می‌تواند به طراحی مدل‌های تعبیه گره بهینه‌شده برای گراف‌های مکان‌مبنا پرداخته و به تحلیل‌های پیچیده‌تر در سطح داده‌های کلان بپردازد.

مراجع

[1] C. C. Aggarwal, *An Introduction to Social Network Data Analytics*, Springer US, 2011, doi: 10.1007/978-1-4419-8462-3_1.

[2] R. Panchendrarajan and A. Saxena, "Topic-based influential user detection: a survey," *Applied Intelligence*, vol. 53, no. 5, pp. 5998-6024, 2023, doi: 10.1007/s10489-022-03831-7.

[3] R. Rabade, N. Mishra, and S. Sharma, "Survey of influential user identification techniques in online social networks," in Recent Advances in Intelligent Informatics: Proceedings of the Second International Symposium on Intelligent Informatics (ISI'13), Mysore, India, Aug. 2013, pp. 359-370, Springer, 2014, doi: 10.1007/978-3-319-01778-5_37.

[4] S. Peng, Y. Zhou, L. Cao, S. Yu, J. Niu, and W. Jia, "Influence analysis in social networks: A survey," *Journal of Network and Computer Applications*, vol. 106, pp. 17-32, 2018, doi: 10.1016/j.jnca.2018.01.005.

[5] X. Wei, Y. Qian, C. Sun, J. Sun, and Y. Liu, "A survey of location-based social networks: problems, methods, and future research directions," *GeoInformatica*, vol. 26, no. 1, pp. 159-199, 2022, doi: 10.1007/s10707-021-00450-1.

[6] S. Khetarpaul, "Mining location based social networks to understand the citizen's check-in patterns," *Computing*, vol. 103, no. 12, pp. 2967-2993, 2021, doi: 10.1007/ 10.1007/s00607-021-01020-x.

[7] X. Yang, "Identification of influential spreaders in geo-social network," in *Proc. 25th Int. Conf. on Geoinformatics*, Aug. 2017, pp. 1-4, doi: 10.1109/GEOINFORMATICS.2017.8090941.

[8] V. Iswarya, V. Govindasamy, and V. Akila, "A survey on the identification of influential spreaders in complex networks," in *Proc. 2nd Int. Conf. on Computer, Communication and Control (IC4),* Feb. 2024, pp. 1-4, doi: 10.1109/IC457434.2024.10486364.

[9] X. Pan, R. Hu, and D. Li, "Social-IFD: Personalized influential friends discovery based on semantics in LBSN," in *Proc. IEEE Int. Conf. on Communications (ICC)*, Jun. 2020, pp. 1-6, doi: 10.1109/ICC40277.2020.9148703.

[10] K. Ali, C. T. Li, and Y. S. Chen, "Joint selection of influential users and locations under target region in location-based social networks," *Sensors*, vol. 21, no. 3, p. 709, 2021, doi: 10.3390/s21030709.

[11] K. Zhang and K. Pelechrinis, "Understanding spatial homophily: the case of peer influence and social selection," in *Proc. 23rd Int. Conf. on World Wide Web*, Apr. 2014, pp. 271-282, doi: 10.1145/2566486.2567990.

[12] Z. S. Akhavan-Hejazi, M. Esmaeili, M. Ghobaei-Arani, and B. Minaei-Bidgoli, "Identifying influential users using homophily-based approach in location-based social networks," *The Journal of Supercomputing*, pp. 1-36, 2024, doi: 10.1007/s11227-024-06228-0.

[13] A. Zareie and R. Sakellariou, "Similarity-based link prediction in social networks using latent relationships between the users," *Scientific Reports*, vol. 10, no. 1, p. 20137, 2020, doi: 10.1038/s41598-20-76799-4.

[14] X. Chen, L. Deng, Y. Zhao, X. Zhou, and K. Zheng, "Community-based influence maximization in location-based social network," *World Wide Web*, vol. 24, pp. 1903-1928, 2021, doi: 10.1007/s11280-021-00935-x.

[15] N. U. Khan, W. Wan, R. Riaz, S. Jiang, and X. Wang, "Prediction and classification of user activities using machine learning models from location-based social network data," *Applied Sciences*, vol. 13, no. 6, p. 3517, 2023, doi: 10.3390/app13063517.

[16] J. Zhao, Y. Song, F. Liu, and Y. Deng, "The identification of influential nodes based on structure similarity," *Connection Science*, vol. 33, no. 2, pp. 201-218, 2021, doi: 10.1080/09540091.2020.1806203.

[17] A. M. U. D. Khanday, M. A. Wani, S. T. Rabani, and Q. R. Khan, "Hybrid approach for detecting propagandistic community and core node on social networks," *Sustainability*, vol. 15, no. 2, p. 1249, 2023, doi: 10.3390/su15021249.

[18] J. A. Smith, M. McPherson, and L. Smith-Lovin, "Social distance in the United States: Sex, race, religion, age, and education homophily among confidants, 1985 to 2004," *American Sociological Review*, vol. 79, no. 3, pp. 432-456, 2014, doi: 10.1177/0003122414531776.

[19] R. Zafarani, M. A. Abbasi, and H. Liu, *Social Media Mining: An Introduction*, Cambridge University Press, 2014, doi: 10.1017/CBO9781139088510.

[20] K. Z. Khanam, G. Srivastava, and V. Mago, "The homophily principle in social network analysis: A survey," *Multimedia Tools and Applications*, vol. 82, no. 6, pp. 8811-8854, 2023, doi: 10.1007/s11042-021-11857-1.

[21] B. Perozzi, R. Al-Rfou, and S. Skiena, "Deepwalk: Online learning of social representations," *Proc. 20th ACM SIGKDD Int. Conf. on Knowledge Discovery and Data Mining*, 2014, pp. 701-710, doi: 10.1145/2623330.2623732.

[22] M. Grohe, "word2vec, node2vec, graph2vec, x2vec: Towards a theory of vector embeddings of structured data," *Proc. 39th ACM SIGMOD-SIGACT-SIGAI Symp. on Principles of Database Systems*, 2020, pp. 1-16, doi: 10.1145/3375395.3387641.

[23] P. Goyal and E. Ferrara, "Graph embedding techniques, applications, and performance: A survey," Knowledge-Based Systems, vol. 151, pp. 78-94, 2018, doi: 10.1016/j.knosys.2018.03.022.

[24] J. Zhou, L. Liu, W. Wei, and J. Fan, "Network representation learning: from preprocessing, feature extraction to node embedding," ACM Computing Surveys (CSUR), vol. 55, no. 2, pp. 1-35, 2022, doi: 10.1145/3491206.

[25] A. Grover and J. Leskovec, "node2vec: Scalable feature learning for networks," Proc. 22nd ACM SIGKDD Int. Conf. on Knowledge Discovery and Data Mining, 2016, pp. 855-864, doi: 10.1145/2939672.2939754.

[26] C. Zhang, T. Li, Y. Gou, and M. Yang, "KEAN: Knowledge embedded and attention-based network for POI recommendation," *Proc. IEEE Int. Conf. on Artificial Intelligence and Computer Applications (ICAICA)*, 2020, pp. 847-852, doi: 10.1109/ICAICA50127.2020.91822385.

[27] A. Kosmatopoulos, K. Loumponias, D. Chatzakou, T. Tsikrika, S. Vrochidis, and I. Kompatsiaris, "Random-walk graph embeddings and the influence of edge weighting strategies in community detection tasks," Proc. 2021 Workshop on Open Challenges in Online Social Networks, 2021, pp. 9-13, doi: 10.1145/3472720.3483621.

[28] A. Saxena, G. Fletcher, and M. Pechenizkiy, "Nodesim: node similarity-based network embedding for diverse link prediction," EPJ Data Science, vol. 11, no. 1, p. 24, 2022, doi: 10.1140/epjds/s13688-022-00336-8.

[29] T. Tao, Q. Wang, Y. Ruan, X. Li, and X. Wang, "Graph Embedding with Similarity Metric Learning," Symmetry, vol. 15, no. 8, p. 1618, 2023, doi: 10.3390/sym15081618.

[30] D. Chen, P. Du, B. Fang, D. Wang, and X. Huang, "A node embedding-based influential spreaders identification approach," *Mathematics*, vol. 8, no. 9, p. 1554, 2020, doi: 10.3390/math8091554.

[31] D. Yang, B. Qu, J. Yang, and P. Cudre-Mauroux, "Revisiting user mobility and social relationships in LBSN: a hypergraph embedding approach," The Web

Conference (WWW), 2019, pp. 2147-2157, doi: 10.1145/3308558.3313635.

[32] D. Yang, B. Qu, J. Yang, and P. Cudré-Mauroux, "Lbsn2vec++: Heterogeneous hypergraph embedding for location-based social networks," IEEE Trans. on Knowledge and Data Engineering, vol. 34, no. 4, pp. 1843-1855, 2022, doi: 10.1109/TKDE.2020.2997869.

[33] W. Liu, S. Wang, and J. Ding, "Influence Maximization Based on Adaptive Graph Convolution Neural Network in Social Networks," *Electronics*, vol. 13, no. 16, p. 3110, 2024, doi: 10.3390/electronics13163110.

[34] Z. Liu, P. Yang, Q. Hao, W. Zheng, and Y. Xiao, "Exploring implicit influence for social recommendation based on GNN," Soft Computing, pp. 1-14, 2024, doi: 10.1007/s00500-024-09898-3.

[35] H. Zhang, G. Kou, Y. Peng, and B. Zhang, "Role-aware random walk for network embedding," Information Sciences, vol. 652, p. 119765, 2024, doi: 10.1016/j.ins.2023.119765.

[36] J. Liu, H. Yi, Y. Gao, and R. Jing, "Personalized Point-of-Interest Recommendation Using Improved Graph Convolutional Network in Location-Based Social Network," *Electronics*, vol. 12, no. 16, p. 3495, 2023, doi: 10.3390/electronics12163495.

[37] Y. Shang, B. Zhou, X. Zeng, Y. Wang, H. Yu, and Z. Zhang, "Predicting the popularity of online content by modeling the social influence and homophily features," Frontiers in Physics, vol. 10, p. 915756, 2022, doi: 10.3389/fphy.2022.915756.

[38] Stanford SNAP, "Gowalla Dataset," [Online]. Available: https://snap.stanford.edu/data/loc-gowalla.html. Accessed: Jan. 14, 2016.

[39] Stanford SNAP, "Brightkite Dataset," [Online]. Available: https://snap.stanford.edu/data/loc-brightkite.html. Accessed: Jan. 14, 2016.

[40] J. J. Levandoski, M. Sarwat, A. Eldawy, and M. F. Mokbel, "Lars: A location-aware recommender system," *Proc. IEEE 28th Int. Conf. on Data Engineering (ICDE)*, 2012, pp. 450-461, doi: 10.1109/ICDE.2012.54.