# خوشه بندی نود های تاثیر گذار در  شبکه های اجتماعی با استفاده از SOM


## چکیده

در طول دهه گذشته، تعداد شبکه های اجتماعی به سرعت افزایش یافته اند. میلیون ها نفر در شبکه های اجتماعی از قبیل فیس بوک و توییتر شرکت می کنند. فیس بوک به طور ویژه، بیش از یک میلیارد کاربر فعال در سال ۲۰۱۲ داشته است. همانطور که تعداد کاربران افزایش می یابد، پیچیدگی ارزیابی شبکه های اجتماعی نیز بیشتر می شود. در این مقاله خوشه بندی شبکه های اجتماعی بررسی شده و با برخی از روش های موجود در این زمینه مقایسه شده است. نتایج حاصل از مقایسه نشان می دهد الگوریتم پیشنهادی نتایج بهتری را ارائه نموده است.



کلمات کلیدی: شبکه های اجتماعی، نود های تاثیر گذار در شبکه های اجتماعی، شبکه SOM

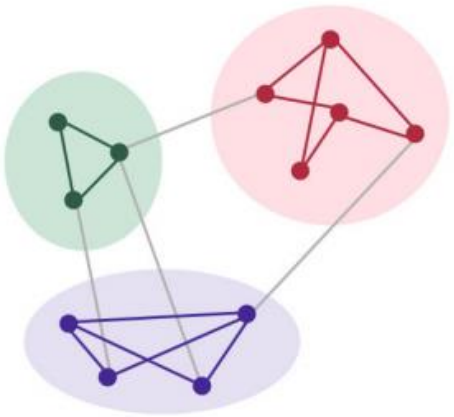

## ۱- مقدمه

در طول دهه گذشته، تعداد شبکه های اجتماعی[۱] به سرعت افزایش یافته اند. میلیون ها نفر در شبکه های اجتماعی از قبیل فیس بوک[۲] و توییتر[۳] شرکت می کنند[۱]. فیس بوک به طور ویژه، بیش از یک میلیارد کاربر فعال در سال ۲۰۱۲ داشته است. همانطور که تعداد کاربران افزایش می یابد، پیچیدگی نیز هنگامی که ارزیابی شبکه های اجتماعی بیشتر می شود، بالا می رود. علاوه بر آن حوزه ی شبکه های اجتماعی نیز گسترده تر می شود[۲]. هدف عمده ی شبکه های اجتماعی اتصال افراد است تا هر کاربر در شبکه اجتماعی بتواند یک لینک هدفمند در شبکه های اجتماعی برقرار کند. این اتصالات و ارتباطات در شبکه های اجتماعی تحت عنوان یک گراف G مدل سازی می شود که به عنوان یک مجموعه ی $G = (V, E)$ تعریف می شود که V یک مجموعه از N نود $(V = \{V_1, V_2, \dots, V_n\})$ و $E \subseteq V \times V$ مجموعه ی یال هایی است که نود $V_i$ و $V_j$ را به هم متصل می کند[۱, ۲]. به عبارت دیگر V×V یک ماتریس مجاورت[۴] $E = [E_{ij}]$ ، $i, j \in V$ است که $V_{ij} \in \{0,1\}$ قابلیت دسترسی یک یال از نود i به نود j را نشان می دهد. وزن یال $E_{ij} > 0$، شدت تعامل[۵] را نشان می دهد و گراف $G(V, E)$ در این مورد یک گراف وزن دار[۶] نامیده می شود. گراف، اگر $E_{ij} \neq E_{ji}$، برای همه ی $i, j \in V$، غیر

---

۱ Social Network
۲ Face Book
۳ Tweeter
۴ Adjacency Matrix
۵ Intensity Of Interaction
۶ Weighted Graph

جهت دار است[۱]. اکثر شبکه های اجتماعی، قابلیت هایشان را رایگان در اختیار کاربران قرار می دهند، اگرچه بعضی از شبکه های اجتماعی نیاز دارند تا کاربران در آنها ثبت نام کنند تا به تمام قابلیت ها دسترسی داشته باشند. اطلاعات شخصی درباره ی هر کاربر در پروفایل وی ذخیره می شود که یک پروفایل مجموعه ای از اطلاعات کاربر است که هویت فرد[۷] و ویژگی های شخصی دیگر را از قبیل علایق اش را شکل می دهد[۳]. هدف اصلی شبکه های اجتماعی، ارتباط افراد است بنابراین هر کاربر در شبکه اجتماعی می تواند یک لینک با کاربران دیگر در شبکه برقرار کند. شکل (۱) انواع ارتباط هایی را که در شبکه های اجتماعی رخ می دهد نشان می دهد. یک مثال مفهوم «مرا دنبال کن»[۸] در توییتر باشد که یک کاربر (ایجاد کننده) می تواند کاربران دیگر (هدف) را فالو کند[۳, ۴]. یک ارتباط کامل بین ایجاد کننده و هدف در صورتی که هر دو یکدیگر را فالو کنند، برقرار می شود. در مثال توییتر، یک اتصال کامل قابلیت های اضافی از قبیل توانایی ارسال پیام خصوصی بین کاربران را امکان پذیر می کند. کاربران این ارتباط ها را برقرار می کنند تا مشارکت دیگر را فالو کنند به ویژه آنها علائق مشترک داشته باشند[۴].



شکل (۱) نمونه ای از ساختار ارتباطی در شبکه های اجتماعی

بسیاری از شبکه های اجتماعی اجازه می دهند تا هر کاربری پروفایل کاربران دیگر را ببینند، اگرچه برخی شبکه های اجتماعی از قبیل فیس بوک کاربرانی با سطح خصوصی[۹] ایجاد می کند که به آنها اجازه می دهد تا فقط به گروه خاصی از پروفایل ها دسترسی داشته باشند[۵].

---

۷ Identity
۸ Follow-Me
۹ Privacy Level

## ۲- خوشه بندی شبکه های اجتماعی

تقسیم بندی افراد در گروه های مختلف در ذات بشر است. در گذشته افراد، خوشه بندی را به منظور مطالعه پدیده ها[۱۰] و مقایسه آن ها با یکدیگر براساس مجموعه ی خاصی از قوانین، به کار می برند. خوشه بندی به گروه بندی اشیاء شبیه یکدیگر اطلاق می شود[۶, ۷]. هر گروه یک خوشه نامیده می شود. هر خوشه از اشیایی تشکیل می شود که شباهت های مشترک دارند و به اشیاء موجود در گروه شبیه هستند. اکثر تعاریف برای خوشه بندی پارتیشن بندی داده ها در گروه ها براساس معیارهای معین است که این داده های گروه بندی شده در خوشه باید مشترک باشند. این معیارها شامل شباهت های مشترک است که با استفاده از اندازه گیری فاصله محاسبه شده است. می توان مفهوم خوشه بندی را در شبکه های اجتماعی دنیای واقعی با گروه بندی افراد با ارتباط دوستانه بالا از نظر درونی و پراکنده شده از نظر بیرونی، تعریف کرد[۱, ۸]. هر الگوریتم خوشه بندی یک فاکتور شباهت (ماتریس مجاورت[۱۱]) دارد تا اشیا مشابه را با یکدیگر سازماندهی کند. درک معیار تشابه بسیار مهم است. تاکنون الگوریتم های خوشه بندی زیادی برای دسته بندی داده براساس معیارهای مختلف پیشنهاد شده است. الگوریتم های مرکزی از روش تکراری پیروی می کنند که از طریق تشریح کل شبکه و یا هر یک از گره ها به عنوان یک جامعه، کار خود را انجام می دهند. وقتی که آنها از کل شبکه استفاده کنند، کار خود را با استفاده از متد های تقسیمی شروع می کنند و در حالتی که از گره ها استفاده کنند،که خود را با استفاده از الگوریتم های متراکم کننده انجام می دهند. الگوریتم پیشنهادی از یک روش جدید تخصیص وزن استفاده می کند که در آن یک شبکه اجتماعی براساس روابط بین گره های شبکه عصبی به زیر شبکه های کوچکتری تقسیم می شوند. برای بررسی و ارزیابی متد پیشنهادی الگوریتم بر روی پایگاه داده های مشهوری که توسط بسیاری از نمونه های دیگر بهره برداری شده اند به کار گرفته شده است. برای این هدف پایگاه داده های زیر استفاده شده اند :

۱-باشگاه کاراته

۲-فوتبال دانشگاه امریکا

۳-کتابهای سیاست های آمریکا

دلیل استفاده از این پایگاه های داده آن است که هر یک از آنها یک شبکه اجتماعی را ارائه می کنند و در آنها ساختار، تعداد و اعضای هر یک از جوامع به وضوح تعریف شده است. در مقایسه با سایر الگوریتم های دیگری که

---

۱۰ Phenomena

۱۱ Proximity Matrix

برای شناسایی جوامع در شبکه های اجتماعی تعریف شده اند این الگوریتم دقت بالاتری دارد. این یک مدل غیر نظارتی است. در این مدل گره ها یا عصب ها در فضای دو بعدی قرار داده می شوند و رابطه بین آنها نیز نقش SOM را تعریف می کند. این نقش، تخمین تابع توزیع است. در این متد تعدادی برای تعداد عصب های خروجی انتخاب می شوند و با استفاده از یک منطق ساده، فاصله هندسی را محاسبه می کنند. عصب های ورودی و خروجی با مقادیر باینری نشان داده می شوند. شبکه کار خودرا با کاهش فاصله بین خودش و الگو های ورودی انجام می دهد. الگوریتمی براساس نقشه خودسازماندهی (SOM ) پیشنهاد شد. این الگوریتم ساختار یک شبکه اجتماعی را به عنوان ورودی دریافت و گراف مجاورت آن را شکل می داد. با به کارگیری تغییرات در فاز یادگیری SOM ، از طریق تنظیم وزن عصب های شبکه، شبکه اجتماعی به کلاستر های مختلفی تقسیم می شد و براساس نتایج حاصل از به کارگیری این الگوریتم بر روی شبکه های اجتماعی مختلف، می توان دریافت که این الگوریتم توانایی خوشه بندی یک شبکه اجتماعی را به خوبی دارد. سپس با اعمال تغییراتی در الگوریتم به کار گرفته شده سعی داریم الگوریتم را برای تنظیم وزن عصب ها نظیر روش محاسبه عصب پرنده را اصلاح کنیم. همچنین الگوریتم باید به گونه ایی اصلاح گردد که پارامتر هایی نظیر چگالی، ارتباطات شبکه و غیره را نیز در نظر بگیرد. انواع زیادی از الگوریتم های خوشه بندی وجود دارد. این الگوریتم ها براساس ساختار خوشه بندی (سلسله مراتبی، پارتیشنی) و ساختار و انواع داده ای (عددی و دسته ای) یا سایز داده ها (مجموعه داده های بزرگ) می توان دسته بندی کرد[۹]. به طور کلی، روش های خوشه بندی می تواند به چهار نوع سلسله مراتبی[۱۲]، پارتیشنی[۱۳]،کنترل شده فراجستجویی[۱۴] و مبتنی بر تراکم[۱۵] تقسیم می شوند.

## ۳- بررسی کارهای پیشین

مورتاق[۱۶] گراف نزدیک ترین همسایه (NN)[۱۷] را به عنوان جمع آوری نقاط تعریف کرد که NN(P) نزدیک ترین همسایه ها از نقطه P را نشان می دهد. اگر برای نقطه p و q، داشته باشیم:

$$NN(q) = p \quad , NN(p) = q \qquad (۱)$$

---

۱۲ Hierarchical
۱۳ Partitional
۱۴ Meta-Search Controlled
۱۵ Density-Based
۱۶ Murtagh
۱۷ Nearest-Neighbor

آنگاه نقاط p و q تحت عنوان نزدیک ترین همسایه های متقابل (RNN)[۱۸] تعریف می شوند. یک زنجیره NN، زنجیری است که از NN(P) تشکیل می شود که از یک نقطه ی دلخواه شروع می شود و در RNN ها خاتمه می یابد. در تحقیقات قبل از مورتاق، از این تعریف ها استفاده و چهار الگوریتم پیشنهاد شده است. الگوریتم اول همه ی زوج RNN ها را می یابد و آن هایی را که به هم نزدیک تر هستند، به هم متصل می کند. دومین و سومین الگوریتم یک زنجیره ی NN می یابد و سپس نزدیکترین نقاط را بعد از ایجاد زنجیره، به هم متصل می کند. چهارمین الگوریتم همه ی زنجیره های NN را می یابد و سپس هر دو نزدیکترین زنجیره NN را به هم متصل می کند[۱۰].

الگوریتم خوشه بندی دیگر، الگوریتم اسکن[۱۹] نامیده می شود که توسط گروه تحقیقی یو[۲۰] معرفی شد. این الگوریتم اختلاف های ساختاری یک دیاگرام شبکه را بررسی می کند تا خوشه بندی را انجام دهد و دو نقش ویژه تعریف می کند: یک هاب که دو خوشه یا بیشتر را به هم متصل می کند و داده های پراکنده[۲۱] که ادغام سطح پایین تری با اعضای دیگر در یک خوشه دارند. [۱۱, ۱۲]. الگوریتم های گروه بندی یا خوشه بندی خودکار که برای شبکه های اجتماعی استفاده می شوند، در کار تحقیقی گروه اسلامی مورد بحث قرار گرفته اند. در این تحقیق، محققان پیشنهاد می کنند که خوشه بندی متقابل سربار کاربران شبکه های اجتماعی هستند. بنابراین کاربران شبکه اجتماعی مورد نظر از الگوریتم های خوشه بندی خودکار برای انجام گروه بندی سریع دوستان شبکه های اجتماعی استفاده می کنند[۱۳, ۱۴]. در میان این فاکتورها، متداول ترین معیار استفاده شده، دسته ها و گروه های اجتماعی و بعد از آن دوام رابطه می باشد. بعد از اینکه وظیفه ذخیره سازی کارت تکمیل شد، از الگوریتم اسکن برای خوشه بندی داده ها از فیس بوک استفاده می شود. در نهایت، شباهت نتایج حاصل با استفاده از متد ذخیره سازی کارت و متد اسکن با هم مقایسه می شوند. ابتدا، نتایج خوشه بندی ذخیره سازی کارت را با مجموعه $C = \{C_1, C_2, C_3, \cdots, C_m\}$ و نتایج خوشه بندی اسکن با مجموعه ی $G = \{G_1, G, G_3, \cdots, G_m\}$ نشان داده می شود[۱۵]. شباهت بین $C_i (1 \le i \le m)$ و $G_j (1 \le j \le m)$ توسط $S_{ij}$ نشان داده شده است. این متد، نمره درصد تشابه را محاسبه می کند به طوری که تعداد دوستان در هر دو گروه (اشتراک $C_i$ و $G_j$) با جمع اعضای متمایز گروه های مختلف (اجتماع $C_i$ و $G_j$) تقسیم شده است. مقدار تشابه نتایج آزمایشی بین ۱۸/۱ و ۷۹/۵٪ با

[۱۸] Reciprocal Nn
[۱۹] Scan
[۲۰] Xu
[۲۱] Outlier

میانگین ٪۴۴/۸ قرار می گیرد. به دلیل اینکه این متد فقط ساختار شبکه را در نظر می گیرد، دقت آن کافی نیست. بنابراین این متد نامناسب تشخیص داده می شود[۱۵, ۱۶].

$$S_{\mathrm{im}}(\mathrm{C},\mathrm{G}) = \frac{\sum_{i\le m\le n} S_{ij}}{\mathrm{Max(m,n)}} \qquad (۲)$$

این آزمایش هم چنین نشان می دهد که هنگامی که افراد، دوستان فیس بوکی را دسته بندی کردند، چندین دوست آن ها مضطرب می شوند طوری که افراد نمی دانند چگونه دوستان را خوشه بندی کرده اند. اکثر این افراد توسط الگوریتم اسکن به عنوان هاب یا داده پراکنده شناخته می شوند. پیشنهاد می شود که این نوع از افراد مشخص شوند تا از پروسه ی خوشه بندی مبتنی بر الگوریتم حذف نشوند و فقط برای خوشه بندی متقابل[۲۲] استفاده شوند. در چندین مقاله دیگر گروه های تحقیقاتی بارات[۲۳] و فان بین[۲۴] و جلداستن دویستین[۲۵] آنالیزهای شبکه ی وزن دار را ارائه کردند و سه تایی باز[۲۶] که از دو لبه تشکیل می شود و سه تایی بسته[۲۷] که متشکل از سه لبه است را تعریف کرده اند. راه های زیادی برای محاسبه وزن هر سه تایی وجود دارد که شامل میانگین حسابی[۲۸] و میانگین هندسی[۲۹]، ماکزیمم و مینیمم است[۱۷]. هر چند که بیشتر محققان از میانگین هندسی استفاده می کنند[۱۷, ۱۸]. ضریب خوشه بندی وزن دار به صورت زیر تعیین می شود:

$$C_{\omega} = \frac{\text{مقدار کل سه تایی های بسته}}{\text{مقدار کل سه تایی ها}} = \frac{\sum_{\tau}\Delta^{\omega}}{\sum_{\tau}\omega} \qquad (۳)$$

اسکات وال[۳۰] و جان شپرد[۳۱] روشی پیشنهاد کردند که براساس روش خوشه بندی طیفی محققان قبلی، گروه جوردن[۳۲] و هم چنین گروه ژانگ[۳۳] می باشد. در این روش ابتدا ترکیب طیف شبکه باید تعیین شود که به تعیین سطح سلسله مراتب در شبکه و تعداد خوشه ها در هر سطح سلسله مراتب کمک می کند. با تعداد خوشه ها در سطح پایه ی تعیین شده، یک خوشه بندی اولیه روی ماتریس مجاورت مربعی A انجام می شود. کلاوسد[۳۴] و

---

[۲۲] Matual Clustering
[۲۳] Barrat
[۲۴] Phanbinh
[۲۵] Fjeldstadoystein
[۲۶] Open Triplet
[۲۷] Closed Triplet
[۲۸] Arithmetic Mean
[۲۹] Geometric Mean
[۳۰] Scott Wahl
[۳۱] John Sheppard
[۳۲] Jordan
[۳۳] Zhang
[۳۴] Clauset

نیومن[۳۵] و مور[۳۶] الگوریتم دیگری به نام CNM پیشنهاد کردند. الگوریتم CNM یک متد خوشه بندی جمع کننده پایین به بالا است که از تکنیک حریصانه برای ترکیب و ادغام خوشه ها در شبکه استفاده می کند. در بدترین حالت کارایی زمان این الگوریتم $O(nlog^{۲}_{n})$ می باشد[۱۹, ۲۰]. گروه پژوهشی خدیوی[۳۷] مراحل پیش پردازش و پس پردازش[۳۸] را برای بهبود الگوریتم نیومن و گیروان[۳۹] پیشنهاد کردند که تحت نام نیومن سریع[۴۰] شناخته می شود. آن ها وزن ها را برای هر یال $e_{ij}$ که هر دو راس i , j را در گراف به هم وصل می کند، محاسبه کردند. بعد از محاسبه ی وزن ها، الگوریتم نیومن سریع از متد حریصانه برای به حداکثر رسانی تابع Q استفاده می کند. الگوریتم نیومن سریع با نمایش هر راس به عنوان یک انجمن شروع می شود. سپس انجمن ها ادغام می شوند تا مقدار Q را به حداکثر برساند. این الگوریتم هنگامی که Q نمی تواند بیشتر بهبود یابد یا همه نود یک انجمن را شکل داده باشند، خاتمه می یابد[۲۱]. بررسی شبکه های اجتماعی بطور کامل در [۲۲] انجام شده است. در این منبع، شبکه های اجتماعی از دیدگاه های مختلف بررسی شده و سپس نحوه تبادل داده و خوشه بندی آن ها مورد بررسی قرار گرفته است. مطالعه [۲۲] چندین روش خوشه بندی در این نوع شبکه را نشان می دهد که با استفاده از آن ها می توان ارتباط بین اعضا و مدل های رفتاری آن ها را نمایش داد. در [۲۳] بررسی ایده ها و نظرات مربوط به افراد مختلف کارآ تلقی شده و شناسایی نظرات مختلف با توجه به دسته بندی افراد در خوشه های مجزا بررسی شده است. سپس یک مدل رفتاری با توجه به ارتباطات و علایق افراد ارائه شده نحوه توزیع نظرات مختلف و گاها متناقض در شبکه های اجتماعی بررسی شده است.

## ۴- معرفی الگوریتم پیشنهادی

در شبکه خود سازمانده، از روش یادگیری رقابتی برای آموزش استفاده می شود و مبتنی بر مشخصه های خاصی از مغز انسان توسعه یافته است. سلول ها در مغز انسان در نواحی مختلف طوری سازمان دهی شده اند که در نواحی حسی مختلف، با نقشه های محاسباتی مرتب و معنی دار ارائه می شوند. برای نمونه، ورودی های حسی لامسه –شنوائی و ... با یک ترتیب هندسی معنی دار به نواحی مختلف مرتبط هستند[۲۴]. واحد ها در یک فرآیند یادگیری رقابتی نسبت به الگوهای ورودی منظم می شوند. محل واحدهای تنظیم شده در شبکه به گونه ای نظم می یابد که برای ویژگیهای ورودی، یک دستگاه مختصات معنی دار روی شبکه ایجاد شود. لذا یک نقشه ی خود

---

[۳۵] Newman
[۳۶] Moore
[۳۷] Khadivi
[۳۸] Post-Processing
[۳۹] Girvan
[۴۰] Newman Fast

سازمان ده، یک نقشه ی توپوگرافیک از الگوهای ورودی را تشکیل می دهد که در آن، محل قرار گرفتن واحدها، متناظر ویژگیهای ذاتی الگوهای ورودی است[۲۵]. یادگیری رقابتی که در این قبیل شبکه ها بکار گرفته می شود بدین صورت است که در هر قدم یادگیری، واحدها برای فعال شدن با یکدیگر به رقابت می پردازند، در پایان یک مرحله رقابت تنها یک واحد برنده می شود، که وزن های آن نسبت به وزن های سایر واحدها به شکل متفاوتی تغییر داده می شود. این نوع از یادگیری را یادگیری بی نظارت می نامند. شبکه های خود سازمانده به لحاظ ساختاری به چند دسته تقسیم می شوند که در ادامه به هر یک از آن ها به صورت مختصر پرداخته خواهد شد[۲۵, ۲۶]. نگاشت خودسازنده یک مدل یادگیری غیر نظارت شده[۴۱] است. در این مدل، نودها یا نورون ها در یک فضای دو بعدی مرتب شده اند و تعامل بین آن ها، نقش نگاشت خودسازنده را تعیین می کند. این نگاشت برای تخمین یک تابع توزیع به کار می رود. در این متد، یک عدد برای تعداد نورون های خروجی انتخاب می شود و با استفاده از یک منطق ساده، فاصله ی هندسی مدل محاسبه می شود. نورون های ورودی و خروجی با مقادیر دودویی مقداردهی اولیه می شوند. شبکه با کاهش فاصله ی بین خودش و الگوهای ورودی کار می کند. بردار $X \in R^n$ در نظر گرفته می شود که هر یک از عناصرش تراکم احتمال[۴۲] $P_i(x)$ دارد. نمونه ها به صورت دوره ای و تصادفی از این فضای تراکم انتخاب می شوند و روی شبکه اعمال می شوند. براساس پارتیشن بندی بردار ورودی در $R^n$، وزن های سلول ها برحسب الگوریتم تغییر می کند. این تغییر در چنین روشی که در نهایت بردارهای وزن سلول ها در تراکم احتمال فضای ورودی توزیع می شوند. بنابراین شبکه تراکم احتمال فضای ورودی را با توزیع سلول هایش در آن تخمین می زند. توزیع سلول ها در فضای احتمال ورودی می تواند به عنوان فشرده سازی داده در نظر گرفته شود. زیرا در حال حاضر هر سلول که در محدوده ی خاص است، یک تخمین از یک محدوده ی خاص را در فضای $R^n$ نشان می دهد. الگوریتم شبکه ی نگاشت خودسازنده به شرح زیر است:

ورودی: یک مجموعه از بردارهای ورودی $V = \{v_1, v_2, v_3, \dots, v_m\}$

خروجی: یک مجموعه از بردارهای خروجی $Z = \{z_1, z_2, \dots, z_m\}$

گام ۱: وزن ها توسط $w_i^o$ مقداردهی اولیه می شوند. پارامترهای همسایگی مربوط به برنده[۴۳] تعریف می شوند و نرخ یادگیری تعیین می شود.

---

[۴۱] Unsupervised learning model
[۴۲] Probability density
[۴۳] Winner

گام ۲: هنگامی که شرط خاتمه نادرست است، گام ۳ تا ۹ را تکرار کن.

گام ۳: برای هر بردار ورودی X، گام های ۴ تا ۶ را تکرار کن.

گام ۴: نورون J به صورت زیر تعیین می شود:

$$D(j) = \sum_i (w_{ij} - x_i)^2 \quad (۴)$$

گام ۵: اندیس j که نزدیک ترین فاصله را به الگوی ورودی[۴۴] دارد، تعیین می شود (مقدار مینیمم $D(j)$).

گام ۶: بردارهای وزن هر دو نورون برنده و همسایه اش به صورت زیر محاسبه می شود:

$$w_{ij}^{new} = (1 - \alpha) w_{ij}^{old} + \alpha x_i \quad (۵)$$

گام ۷: نرخ یادگیری $\alpha$ به روزرسانی می شود.

گام ۸: شعاع همسایگی در زمان مشخص شده کاهش می یابد.

گام ۹: شرط خاتمه بررسی می شود.

## ۵- شبیه سازی الگوریتم پیشنهادی

پایگاه داده کتاب های سیاسی ایالات متحده آمریکا در سال ۲۰۰۹ توسط کربز[۴۵] جمع آوری شده است. این پایگاه داده شامل چهارصد و چهل و یک رکورد از یکصد و پنج کتاب می باشد. یکی از این کتاب ها در آمازون فروخته شده است. عناوین این کتاب، سیاست های ایالت متحده آمریکا می باشد. از طرف دیگر این شبکه یکصد و پنج راس و چهارصد و چهل و یک یال دارد. یک ویژگی در آمازون نشان می دهد که مشتریانی که کتاب می خرند، کتاب های دیگر را نیز خریده اند. براساس این ویژگی ها یال های بین رئوس در این شبکه ارتباط بین خرید کتاب ها را به صورت متوالی نشان می دهد. رئوس در شبکه با علامت های L، N و C برچسب گذاری شده اند که به ترتیب، لیبرال[۴۶]، بی طرف[۴۷] و محافظه کار[۴۸] را نشان می دهند. این برچسب گذاری رئوس توسط نیومن در سال ۲۰۰۹ و براساس شرح و توضیح کتاب ها انجام شده است و در آمازون پست شده اند. سه دسته از کتاب ها، شامل

---

[۴۴] Input pattern
[۴۵] Krebs
[۴۶] Liberal
[۴۷] Neutral
[۴۸] Conservative

کتاب های لیبرال، کتاب های بی طرفانه و کتاب های محافظه کار وجود دارد. هدف، تمایز این سه دسته از کتاب هاست. در این الگوریتم ابتدا دیتاست مورد نظر خوانده شده سپس عملیات نرمال سازی روی آن انجام می گیرد. در حالت معمول داده های موجود در دیتاست دارای مقادیر مختلف هستند که کار کردن با این مقادیر ساده نیست و داده های نویز دار موجود در دیتاست باعث کاهش دقت نتایج خروجی می شود. اما با نرمال سازی داده ها محدوده تمام مقادیر در بازه صفر و یک قرار می گیرند که این کار انجام عملیات محاسباتی را ساده کرده و دقت نتایج حاصل شده را بالا می برد. فرایند نرمال سازی داده ها با استفاده از رابطه (۶) انجام می شود که در آن x نشان دهنده داده غیر نرمال بوده و $min_x$ و $max_x$ به ترتیب کمترین و بیشترین مقدار موجود در داده ها می باشند و داده نرمال شده نیز با $x_n$ نشان داده شده است.

$$x_n = ۲ * \frac{x - min_x}{max_x - min_x} - ۱ \qquad (۶)$$

سپس شبکه SOM مربوطه ساخته شده و با استفاده از داده های موجود در دیتاست آموزش داده خواهد شد. پس از آموزش کلاس مربوط به هریک از داده های ورودی مشخص شده و در قالب یک تصویر نشان داده می شود. در ادامه نتایج حاصل از اجرای الگوریتم پیشنهادی با دو روش دیگر مقایسه خواهد شد. برای مقایسه الگوریتم پیشنهادی از روش گیروان-نیومن[۴۹] و روش ژاو[۵۰] استفاده شده است و نتایج حاصل از اجرای هر سه الگوریتم بر دیتاست کتاب های سیاسی ایالات متحده آمریکا با هم مقایسه شده اند. جدول (۱) تعداد داده هایی که به غلط دسته بنده شده را در اجرای هر سه الگوریتم بر دیتاست کتاب های سیاسی ایالات متحده آمریکا نشان می دهد.

جدول (۱) تعداد دسته بندی نادرست در الگوریتم های مختلف

| نام دیتاست | گیروان-نیومن | ژاو | الگوریتم پیشنهادی |
|---|---|---|---|
| کتاب های سیاسی | ۱۷ | ۱۷ | ۰,۲ |

با توجه به اینکه تعداد داده های موجود در دیتاست مشخص هستند می توان دقت انجام عملیات را بصورت رابطه (۷) تعریف نمود و الگوریتم های گفته شده را نیز بر اساس دقت هر یک دسته بندی نمود. جدول (۲) دقت هر یک از الگوریتم های گفته شده را نشان می دهد. با بررسی نتایج حاصل از اجرای هر سه الگوریتم مشاهده می شود که روش پیشنهادی نتایج بهتری را ارائه کرده است.

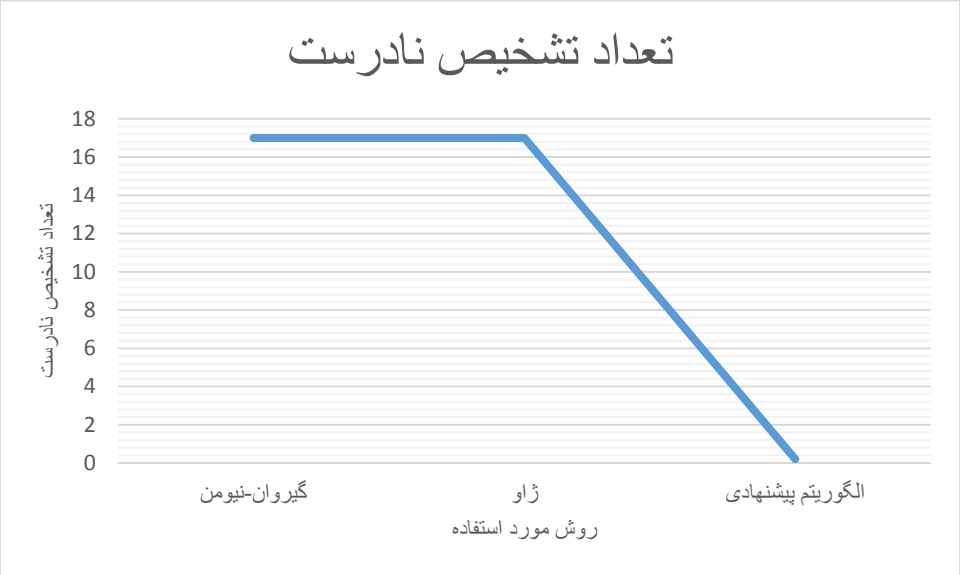
$$\text{دقت} = \frac{\text{تعداد نمونه هایی که درست تشخیص داده شده}}{\text{تعداد کل نمونه ها}} \qquad (۷)$$

[۴۹] Girvan and Newman
[۵۰] Zhao

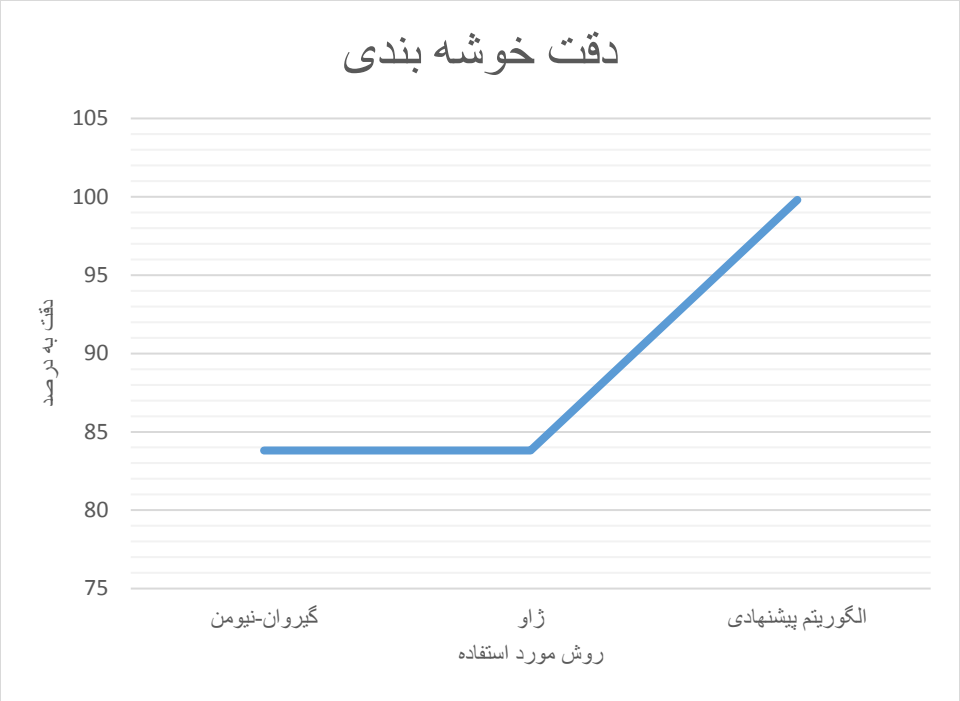
جدول (۲) مقایسه دقت الگوریتم های مختلف

| معیار ارزیابی | گیروان-نیومن | ژاو | الگوریتم پیشنهادی |
|---|---|---|---|
| دقت | ۸۳,۸۰۹۵۲۳۸۱ | ۸۳,۸۰۹۵۲۳۸۱ | ۹۹,۸ |

شکل های (۲) و (۳) به ترتیب نتایج حاصل از جداول (۱) و (۲) را نشان می دهند. بررسی این شکل ها نیز نشان می دهد الگوریتم پیشنهادی نسبت به برخی از الگوریتم های موجود در این زمینه کارایی بهتری داشته است.




شکل (۲) تعداد تشخیص نادرست در الگوریتم های مورد بررسی




شکل (۳) دقت الگوریتم های مورد بررسی

## ۶- نتیجه گیری

در این مقاله شبکه های اجتماعی بررسی شده و سپس خوشه بندی اطلاعات موجود در شبکه های اجتماعی مبتنی بر شبکه های خود سازمانده انجام شده است. نتایج حاصل از الگوریتم پیشنهادی نشان می دهد می توان با استفاده از شبکه های SOM شبکه های اجتماعی را خوشه بندی نمود که این خوشه بندی نسبت به برخی از الگوریتم های موجود در این زمینه از دقت بالاتری برخوردار است.

## ۷- منابع

[۱] S. Gole and B. Tidke, "A survey of big data in social media using data mining techniques," in *Advanced Computing and Communication Systems, 2015 International Conference on*, ۲۰۱۵, □□. ۱-۶.

[۲] B. Colaco and S. S. Khan, "Privacy preserving data mining for social networks," in *Advances in Communication and Computing Technologies (ICACACT), 2014 International Conference on*, ۲۰۱٤, pp. ۱-٤.

[۳] M. E. Newman, "Fast algorithm for detecting community structure in networks," *Physical review E*, vol. ٦٩, no. ٦ ,p. ۰٦٦۱۳۳, ۲۰۰٤.

[۴] S. Zhang, R.-S. Wang, and X.-S. Zhang, "Identification of overlapping community structure in complex networks using fuzzy c-means clustering," *Physica A: Statistical Mechanics and its Applications,* vol. ۳۷٤, no. ۱, pp. ٤۸۳-٤۹۰, ۲۰۰۷.

[۵] J. Leskovec, K. J. Lang, A. Dasgupta, and M. W. Mahoney, "Community structure in large networks: Natural cluster sizes and the absence of large well-defined clusters," *Internet Mathematics,* vol. ٦, no. ۱, pp. ۲۹-۱۲۳, ۲۰۰۹.

[۶] R. S. Burt, "Network items and the general social survey," *Social networks,* vol. 6, no. 4, pp. ۲۹۳-۳۳۹, ۱۹۸٤.

[۷] N. B. Ellison, "Social network sites: Definition, history, and scholarship," *Journal of Computer-Mediated Communication,* vol. ۱۳, no. ۱, pp. ۲۱۰-۲۳۰, ۲۰۰۷.

[۸] J. Scott, *Social network analysis*. Sage, ۲۰۱۲.

[۹] M. Burke, C. Marlow, and T. Lento, "Social network activity and social well-being," in *Proceedings of the SIGCHI conference on human factors in computing systems*, 2010, pp. ۱۹۰۹-۱۹۱۲: □□□.

[۱۰] F. Murtagh and P. Contreras, "Algorithms for hierarchical clustering: an overview," *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery,* vol. 2, no. 1, pp. 86-۹۷, ۲۰۱۲.

[۱۱] D. Gómez, E. Zarrazola, J. Yáñez, and J. Montero, "A divide-and-link algorithm for hierarchical clustering in networks," *Information Sciences,* vol. ۳۱٦, pp. ۳۰۸-۳۲۸, ۲۰۱۵.

[۱۲] X. Xu, N. Yuruk, Z. Feng, and T. A. Schweiger, "Scan: a structural clustering algorithm for networks," in *Proceedings of the 13th ACM SIGKDD international conference on Knowledge discovery and data mining*, ۲۰۰۷, pp. ۸۲٤-۸۳۳: ACM.

[۱۳] E. Turkina, A. Van Assche, and R. Kali, "Structure and evolution of global cluster networks: evidence from the aerospace industry," *Journal of Economic Geography,* p. lbw۰۲۰, ۲۰۱۶

[۱۴] H.-T. Chang, Y.-W. Li, and N. Mishra, "mCAF: a multi-dimensional clustering algorithm for friends of social network services," *SpringerPlus,* vol. ٥, no. ۱, p. ۷٥۷, ۲۰۱٦.

[۱۵] R. Irfan *et al.*, "A survey on text mining in social networks," *The Knowledge Engineering Review,* vol. ۳۰, no. ۰۲, pp. ۱٥۷-۱۷۰, ۲۰۱٥.

[۱۶] C. Grieco, L. Michelini, and G. Iasevoli, "Measuring value creation in social enterprises a cluster analysis of social impact assessment models," *Nonprofit and voluntary sector quarterly,* vol. ٤٤, no. ٦, pp. ۱۱۷۳-۱۱۹۳, ۲۰۱٥.

[١٧] K. Singh, H. K. Shakya, and B. Biswas, "Clustering of people in social network based on textual similarity," *Perspectives in Science,* ٢٠١٦.

[١٨] J. J. Jung, "Exploiting geotagged resources for spatial clustering on social network services," *Concurrency and Computation: Practice and Experience,* vol. 28, no. 4, pp. ١٣٥٦-١٣٦٧, ٢٠١٦.

[١٩] S. Wahl and J. Sheppard, "Hierarchical Fuzzy Spectral Clustering in Social Networks using Spectral Characterization," in *FLAIRS Conference,* ٢٠١٥, pp. ٣٠٥-٣١٠: Citeseer.

[٢٠] P. Sarkar and A. W. Moore, "Dynamic social network analysis using latent space models," *ACM SIGKDD Explorations Newsletter,* vol. ٧, no. ٢, pp. ٣١-٤٠, ٢٠٠٥.

[٢١] S. H. Javadi, S. Khadivi, M. E. Shiri, and J. Xu, "An ant colony optimization method to detect communities in social networks," in *Advances in Social Networks Analysis and Mining (ASONAM), 2014 IEEE/ACM International Conference on*, 2014, pp. 200-203: IEEE.

[٢٢] J. Scott, *Social network analysis*. Sage, ٢٠١٧.

[٢٣] H. Chen, H. Yin, X. Li, M. Wang, W. Chen, and T. Chen, "People Opinion Topic Model: Opinion based User Clustering in Social Networks," presented at the Proceedings of the ٢٦□□ □□□□□□□□□□□□□□ □□□□□□□□□□ □□ □□□□□ □□□□ □□□ □□□□□□□□□□, □□□□□, □□□□□□□□□□.٢٠١٧ ,

[٢٤] M. C. Pham, Y. Cao, R. Klamma, and M. Jarke, "A Clustering Approach for Collaborative Filtering Recommendation Using Social Network Analysis," *J. UCS,* vol. 17, no. 4, pp. ٥٨٣-٦٠٤, ٢٠١١.

[٢٥] A. Shepitsen, J. Gemmell, B. Mobasher, and R. Burke" ,Personalized recommendation in social tagging systems using hierarchical clustering," in *Proceedings of the 2008 ACM conference on Recommender systems*, ٢٠٠٨, pp. ٢٥٩-٢٦٦: ACM.

[٢٦] W. B. Frakes and R. Baeza-Yates, "Information retrieval: data structures and algorithms," ١٩٩٢.