# فصل سوم

# راهکار پیشنهادی

3-1 مقدمه

در این فصل در ارتباط باروش پیشنهادی در ارتباط با تشخیص ناهنجاری یا تقلب در تجویز پزشکی صحبت شده است. الگوریتمی که در این جا ارائه می شود تجویزکنندگان پزشکی را که اقدامات پزشکی انجام داده اند و تفاوتهای چشمگیری با سایر تجویزکنندگان دارای تخصص های مشابه دارند، شناسایی می کند. این تجویزکنندگان می توانند ناهنجاری هایی باشتد که باید برای کلاهبرداری و تقلب بررسی شود. فرض بر این است که بین ارائه دهندگان پزشکی که دارای تخصصهای مشابهی هستند، در الگوهای تجربی شباهت وجود دارد. الگوریتم تجویزکنندگان که این فرض را نقض می کنند، مشخص می کند.  
الگوریتم تشخیص ناهنجاری به شرح زیر خواهد بود:

1) شباهت های بین تجویزکنندگان پزشکی بر اساس نسخه های آنها محاسبه می شود.

2) گراف تشابه ایجاد می شود که گره ها نماینده آنها هستند و لبه ها نشان می دهند که دو ارائه دهنده مشابه هستند (شباهت بالاتر از یک آستانه مشخص است.)

3) سپس از این جا همسایگی ها توسط روش قوانین انجمنی با محاسبه امتیاز و سپس اجرای خوشه بندی محاسبه می شود.

4) تجویزکنندگان یعنی گره هایی با مقادیر منحصربفرد و با تخصص های مختلف نسبت به نمونه انتخاب شده به عنوان ناهنجاری مشخص می شوند. این تجویزکنندگان رتبه پایینتری در تشابه را دریافت می کنند.

در گراف شباهت تولید شده در مرحله 2، انتظار می رود ارتباطات بین تجویزکنندگان یک تخصص خاص متراکم تر باشد زیرا نسخه ها در بین تجویزکنندگان یک تخصص مشابه نسبت به تجویزکنندگان تخصص های مختلف بیشتر هستند. بنابراین، در قوانین انجمنی از نمرات یکسانی بهره برده و سپس در خوشه بندی احتمال قرار گرفتن این تجویزگنندگان در یک خوشه بالاتر می رود. اگر یک تجویزکننده که از یک تخصص متفاوت سود می برد، نمره بالایی در قوانین انجمنی داشته باشد، این نشان می دهد که تجویزکننده معالجات پزشکی را که معمولاً توسط تجویزکننده­گان دارای تخصص های مختلف تجویز می شوند، تجویز می کند. در چنین حالتی، بهتر است به این تجویزکنندگان توجه شود و بررسی نموده که آیا درمان وی مشروعیت دارد یا خیر. در بخش­های بعدی هر مرحله از الگوریتم تشخیص ناهنجاری با جزئیات بیشتری شرح داده شده است. همچنین معیارهای مرتبط با نفوذ ناهنجاری در یک شبکه گرافی مورد بررسی قرار گرفته است که برای تحلیل می­تواند کاربرد داشته یاشد.

از منظر گرافی یک شبکه از تجویزکنندگان را می توان به عنوان یک شبکه بدون وزن مقیاس ثابت G = (V، E) درنظر گرفت بطوری­که V = {v1, v2, ... vn} مجموعه ای از نقاط (رئوس) یا بازیگران و E = {e1, e2, ... em} مجموعه ای از یال ها یا اتصالات M است. ماتریس مجاورت A یک نهاد ریاضی است که نشان دهنده اتصالات شبکه است. در این پژوهش شبکه های بدون جهت در نظر گرفته می­شود، این بدان معناست i j ∈ V، Aij = Aji = 1 خواهد بود اگر i و j به هم متصل، و یا در غیر این صورت 0 درنظر گرفته می­شود. مسیر بین یک جفت رأس (i، j) یک توالی متوالی است که از i شروع می شود و به j متصل می شود، بنابراین هر رأس بیش از یک بار بازدید می شود. فاصله یا طول مسیر به عنوان تعداد یال های موجود در دنباله تعریف شده است. دو رأس همسایه هستند اگر توسط یک یال به طول 1 متصل شوند. کوتاه ترین فاصله بین دو رأس به عنوان کوتاهترین مسیر یا کوتاه ترین خطLij شناخته می شود. کوتاه ترین مسیر را می توان با الگوریتم هایDijkstra،BellmanFord و یا روش جستجوی اولیه بورد [36] محاسبه کرد. یک عنصر یا جزء[[1]](#footnote-1) بزرگترین زیر مجموعه ای از رأس های شبکه است که در آن حداقل یک مسیر بین هر جفت رأس وجود دارد اما هرگز به مولفه دیگری وصل نیست. یک شبکه متصل تنها یک جزء دارد. در صورتی که i و j متعلق به اجزای مختلف باشد، فرض می شود که L ij = ∞. به همین دلیل، در اینجا بزرگترین جزء شبکه در نظر گرفته شده است.

3-2 معیارهای مركزي

چندین معیار برای توصیف اهمیت یا مرکزی یک رأس در شبکه پیشنهاد شده است [37]. این معیارهای مركزی با در نظر گرفتن تعاریف خاصی از نفوذ تعریف شده است [38]. فرض بر این است که رأس ها با اندازه گیری مرکزی بالاتر، برای این­که بر نظرات دیگران تاثیر بگذارد، مناسب­تر هستند. در میان برخی از معیارها، محبوبیت یک رأس (مرکزیت درجه) [37]، نزدیکی یا نزدیک بودن فرد به دیگران (محوریت نزدیکی) [39]، رأس های مورد اعتماد در انتقال اطلاعات (محوریت بینش یا فیمابینی[[2]](#footnote-2)) [40]، نزديكي رأس ها به هسته شبكه (k-core) [41] و يا حتي شهرت فرد (محوریت رتبه صفحه[[3]](#footnote-3)) از جمله معیارها به­شمار می­آید. در زیر، برخی معیارهای مرکزی مهم ارائه می شود.

محدوده درجه[[4]](#footnote-4) (DG) تعداد اتصالات یا روابط یک رأس را بررسی می کند. مجموعه ای از رأس های متصل به یک رأس خاص i به عنوان محله تعریف شده و درجه اتصال DGi نشان دهنده اندازه محله آن است ]37[. این بدان معنی است که رأس با درجه بالاتر، محبوب تر است.

محدوده بینایی[[5]](#footnote-5) (BE) مربوط به ظرفیت انتقال اطلاعات رأس ها می باشد. برای یک رأس j این تعداد کمترین مسیرهایی را که از طریق j بین تمام جفت رأس ها (i، k) [40] عبور می کند را اندازه گیری میکند، با i، j و k متفاوت. این معیار مرکزی بیان می کند که چقدر گره یا راس j به عنوان پل عمل می کند، به این معنی است که اعتماد یا اطمینان در شبکه چیست.

محدوده خاص[[6]](#footnote-6) (EV) با توجه به اهمیت همسایگان خود، اهمیت ویژه ای دارد. مركز ويژه ويکتور اصلي­ترين ويژگي اصلي مربوط به بزرگترين ارزش ويژه ماتريس A مجذور است. این معیار اهميت ارقام را با توجه به کيفيت اتصالات آن توصيف مي کند]43[.

محوریت مرتبه صفحه[[7]](#footnote-7) (PR) از یک فرایند مارکووی ناشی می شود که از ناوبری مسیرتصادفی از طریق شبکه پیروی می کند. این اهمیت رأس را با توجه به احتمال ورود به رأس خاص بعد از تعداد زیادی از مراحل بیان می کند. ابتدا رتبه صفحه برای رتبه بندی صفحات وب پیشنهاد شد [42] و ایده این است که رفتار یک کاربر که در شبکه گشت و گذار میکند شبیه سازی شود. کاربر به دنبال پیوندهای موجود در صفحه فعلی حرکت می کند یا می تواند با تایپ URL جدید در مرورگر به صفحات دیگر پرش کند. در شبکه های اجتماعی می توان آن را مانند افراد ذکر شده یا مشهور نزدیک کرد.

محدوده نزدیکی[[8]](#footnote-8) (CL) میانگین مسیرهای کوتاه از هر رأس به بقیه شبکه است]39[. به صورت رسمی، تقریب نزدیک بودن معکوس از میانگین کوتاهترین مسیرها از i به تمام رأس­ها است، یعنی . به این ترتیب، رأسهایی که به دیگران نزدیکترند، مرکزیت نزدیکی بیشتری دارند. به منظور محاسبه مرکزیت برای شبکه های قطع شده، میانگین به صورت جزء در نظر گرفته شده است.

مركزیت(KC) k-core توپولوژي شبكه را با توجه به تجزيه زير شبكه در هسته توصيف مي كند. هسته مرتبه k (Hk) مجموعه ای از رأسهایی است که برای هر رأس i در Hk درجه آن ki ≥ k است. بدین معنا که k حداکثر هسته ای است که می توان آن را تعریف کرد، Kc (i) = k و Hk بزرگترین مجموعه ای از رأس ها با این ویژگی [41] است. هسته اصلی مجموعه ای از رأس ها با بزرگترین مقدار k-core از شبکه است و این رأس ها بیشترین مرکزی هستند. ستون های با مقادیر کم k-core معمولا در حاشیه شبکه قرار دارند. لازم نیست تمام رأسهای درجه بالا دارای مقادیر k-core بالاتر باشند. به عنوان مثال، مراکز هاپ ها در حاشیه ارزش های کوچکی از k-core دارند، و یا رأس ها با ارزش k-core بزرگتر می توانند به میزان زیاد نباشد [44].

ضریب خوشه بندی[[9]](#footnote-9) (CT) یا ترانزیتی یک ویژگی مشترک در شبکه های واقعی جهان است. در شبکه های اجتماعی به این معنی است که اگر A یک دوست داشته باشد که دوست B نیز باشد، گرایش قوی وجود دارد که A یک دوست B است. در شرایط توپولوژی، مثلث (چرخه سه مرتبه) در شبکه حضور دارد. ضریب خوشه بندی [45] برای یک راس CTi به عنوان تعداد مثلثی که در i بیش از حداکثر تعداد اتصالات ممکن برای i تعریف شده است تعریف می شود. CTi ارزش 1 را دارد اگر همسایه های i به هم متصل باشند.

حفره های ساختاری[[10]](#footnote-10) (HO) تمام رأس ها را به عنوان شبکه اگو[[11]](#footnote-11) در نظر می گیرند، که در آن اتصالاتی که نامربوط به هر رأس است بر روی آن اثر مستقیم ندارند. فاکتور کلیدی افزونگی است که هر رأس در همسایگی خود دارد، و اگر موقعیت و ارتباطات آن فرصت ها را به ارمغان بیاورد، ارزیابی می شود. این بدان معنی است که موفقیت یک شخص در یک شبکه گرافی یا سازمان مربوط به دسترسی به پل های محلی (افراد مورد اعتماد) است [46]. اگر i از شبکه خارج شود، یک حفره ساختاری در محلی اتفاق می افتد. این افراد برای اتصال به مناطق محلی مهم هستند.

به­طورکلی یک شبکه گرافی مفروض را هرچه باشد می­توان با میانگین اندازه­گیری­های مرکزی تمام رأس­ها یعنی مشخص کرد که یک فرمول آماری ساده مشابه با دیگر اندازه­گیری­ها است. جدول 3-1 مثال­هایی را از ویژگی­های ساختاری شبکه­های گرافی بزرگ نشان می­دهد که مقادیر ویژگی­هایی که پیش از این بیان شد در آن­ها ذکر شده است.

جدول 3-1: مثالی از ویژگی­های ساختاری شبکه­های پیچیده

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| شبکه | N | <DG> | <CT> | <HO> | <BE> | <CL> | <EV> | <KC> | <PR> |
| hamsterster | 2000 | 16.1 | 0.5399 | 0.3898 | 2.588\*103 | 1.43\*10-4 | 0.011 | 9.287 | 5.00\*10-4 |
| advogato | 5054 | 15.6 | 0.2525 | 0.3700 | 5.747\*103 | 6.19\*10-5 | 6.82\*10-3 | 8.137 | 1.98\*10-4 |
| google+ | 23613 | 3.32 | 0.1742 | 0.8112 | 3.581\*104 | 1.10\*10-5 | 2.30\*10-3 | 1.669 | 4.21\*10-5 |

3-3 فرایند انتشار (شایعه-ناهنجاری)

انتشار یک فرایند فراگیر در جامعه است و چند مدل برای شناختن انتشار ایده ها یا عقاید از طریق شبکه های گرافی توسعه یافته است. در مدل­های شایعه کلاسیک، افراد نادان یا غیرفعال (S) کسانی هستند که از اطلاعات نادیده گرفته می شوند، اسپردرها (I) کسانی هستند که ایده ها را منتشر می کنند و Stifler (R) کسانی هستند که این اطلاعات را می دانند اما علاقه خود را از دست می دهند گسترش آن تمام رأس ها برای احتمال متقاعد کردن همسایگان و احتمالات γ برای جلوگیری از فعال بودن آن به عنوان propagator یکسان هستند. مكی تامپسون (MT) [48] رویکرد شایعه برای را مدلسازی روند گسترش استفاده کرد. در MT هر زمانی که یک پخش کننده فعال i با یک ریسکه j غیرفعال می شود، دومی با یک احتمال ثابت β فعال می شود. در غیر این صورت، زمانی که j در مورد شایعه می داند، به این معنی است که j یک پخش کننده یا یک stifler است، تبدیل i به یک stifler با احتمال γ تبدیل می شود. رفتار زمانی که پخش کننده متوقف می شود درک می شود به این دلیل که اطلاعات بیش از حد شناخته شده (تماس با پخش کننده) و یا بدون تازگی هستند (تماس با stifler). قواعد کلی تماس می تواند به شرح زیر باشد:

معادله 3-1

که در آن i و j همسایه هستند و عملگر "+" به معنی تماس بین آنها است. از لحاظ اجرای مونت کارلو (MC)، یک جمعیت ثابت از رأس N در تمام مراحل زمان را در نظر بگیرید. هر راس می تواند تنها در یک حالت باشد، که Ii(t)=1 اگر i∈I، در غیر این صورت Ii(t)=0، و Si (T) + Ii (T) +Ri(t)=1 . کسر ماکروسکوپی بی اطلاع ψ (t) ، پخش کننده (φ (t) وstifler Φ(t) به صورت محاسبه می شود. فرض می­شود که آلودگی و بهبودی در همان پنجره یا گام زمانی مجزا اتفاق نمی افتد. همچنین مواردی که یک پخش کننده به صورت تصادفی با یک همسایه در هر واحد زمان تماس می گیرد، فرآیند تماس، تصویب شد. راه اندازی اولیه برای انتشار ψ (0) = 1 - 1 / N، φ (0) = 1 / N و φ (0) = 0. هر شبیه سازی با انتخاب یکنواخت از رأس ها طبق تعریف در تنظیم اولیه آغاز می شود. در هر گام زمانی، همه پخش کننده ها به طور یکنواخت یکی از همسایگان خود را انتخاب می کنند و سعی می کنند آن را با احتمال β آلوده کنند یا انتشار را با احتمال γ متوقف کنند. شبیه سازی ها تا پایان فرآیند انتشار به پایان می رسد، زمانی که φ∞ = 0.

مدل های نظری مختلف برای مدل سازی دینامیک شایعات در شبکه ها پیشنهاد شده است]49[. این مدل های تحلیلی، فرض هایی را درباره ساختار شبکه ای مانند همبستگی درجه یا توزیع، محفظه ها یا کلاس های رأس با همان احتمالات، اختلاط همگن یا تئوری میدان میدانی ارائه می دهند. نکته مشترک این روش­ها این است که همه آن­ها ادعا می کنند راه حل عددی آن­ها با شبیه سازی مونتوکارلو سازگاری دارد. درنتیجه در این پژوهش تا حد امکان سعی می­شود مدل و نتایج حتی­الامکان با این مدل هم­خوانی داشته باشد.

3-4 طراحی روش

در این بخش به تشریح روش ارائه شده پرداخته می­شود. انجام این تحقیق به تحلیل ناهنجاری در داده تجویز پزشکی با استفاده از روش­های داده­کاوی پرداخته می­شود و این خود شامل دو مرحله کلی می­باشد. در مرحله اول ارتباط افراد تجویزکننده با استفاده از قوانین انجمنی و نمرات آن مشخص شده که به صورت ارتباطات گرافی نمایش داده می شود و سپس بر اساس شباهت رفتاری اقدام به خوشه­بندی کاربران می­شود. خوشه­بندی در ادامه مرحله اول صورت می­پذیرد و در نهایت خروجی کار کاربرانی هستند که تجویزهایی شبیه به هم و از طرفی ویژگی­ها (مانند تخصص مشابه) را دارند و بر این اساس سیستم تحلیل ناهنجاری یا تقلب مشخص می­کند که تقلب در چه خوشه هایی اساسا می­تواند وجود داشته باشد. معمولا خوشه هایی احتمال تقلب بالاتری دارند که تعداد اعضای آن کم باشد و تخصص آن ها با تجویز صورت گرفته (بیماری) همخوانی نداشته باشد. این خوشه بندی درنهایت منجر به پیش بینی رتبه کاربران می شود. الگوریتم پیشنهادی با استفاده از تاثیر صورت گرفته از قواعد انجمنی و خوشه­بندی کاربران، می­تواند رتبه­بندی بهتری برای تجویزهای انجام شده هر تجویزکننده و تاثیر تجویزکنندگان بر روی داده های منتشرشده ارائه دهد. علاوه بر دو مرحله اصلی دو مرحله دیگر نیز در ابتدا و انتهای الگوریتم پیشنهادی وجود دارد که برای ارسال ورودی و پردازش خروجی به­کار می­رود. در فلوچارت زیر مراحل روش پیشنهادی در زمینه تحلیل ناهنجاری تجویز پزشکی با مشخص کردن ناهنجاری در تجویز تجویزکنندگان شرح داده شده است:

مرحله 1: استخراج کاربران (تجویزکنندگان) و داده­های مربوطه

فاز 2: اعمال قوانین همسایگی و وابستگی جهت انتخاب همسایگان

فاز 1: پیش پردازش داده‌ها

*مرحله 1: ساخت شبکه گراف شباهت تجویزکنندگان*

دیتاست

مرحله 2: مشخص کردن کاربران همسایه برای هر کاربر

مرحله 2: ذخیره اطلاعات کاربران و داده­ها در پایگاه داده

فاز 3: خوشه بندی K-means

مرحله 1: اعمال خوشه بندی بر روی کاربران برای افزایش دقت تشخیص همسایه

مرحله 2: براساس همسایه های هر کاربر، نحوه پیش بینی مشخص می شود

فاز 4: پیش بینی رتبه‌ کاربران (پایین ترین رتبه احتمال بالاتر ناهنجاری)

مرحله 1: جمع آوری رتبه‌ها نسبت به کاربران همسایه

مرحله 2: پیش­بینی براساس رتبه ها

شکل 3-1: مراحل کلی روش پیشنهادی برای تحلیل ناهنجاری

در این نمودار چهار فاز کلی آماده­سازی داده­ها، ساخت گراف شبکه و اعمال قوانین انجمنی (تعیین همسایه­ها)، خوشه­بندی بر روی کاربران و درنهایت پیش­بینی رتبه­های ناهنجاری کاربران براساس همسایه­های کاربر است. کاربران مشکوک به ناهنجاری کاربرانی هستند که در صدر رتبه ها قرار می گیرند. این مراحل می­تواند به­تنهایی برای یکسری کاربر صورت پذیرد و یا این­که برای تمام کاربران صورت پذیرد. زمانی که هر داده از تجویزکنندگان به سیستم وارد می شود، این داده جدید مربوط به تجویزکننده می تواند در فرایند خوشه بندی و پیش بینی وارد شود. در واقع این روش می­تواند هم به­شکل برخط[[12]](#footnote-12) و هم برون­خط[[13]](#footnote-13) اجرا شود که هرکدام معایب و مزایایی دارند. اگر به­شکل برخط اجرا شود ممکن است زمان تحلیل کمی طولانی شود اما می­تواند به­روزتر و موثرتر باشد. در صورت استفاده از سیستم برون­خط این سیستم در تحلیل ناهنجاری سریع­تر عمل کرده اما نیاز به حافظه اضافی برای ذخیره نتایج برای کل کاربران و نیز مشکل بروزرسانی دارد. اگرچه می­توان ترکیبی از این دو روش نیز صورت پذیرد تا نقاط ضعف پوشش داده شده و نقاط قوت پررنگ­تر شود. در این پژوهش فرض می­شود که روش برون­خط مورد استفاده قرار گرفته و با اجرای الگوریتم رتبه­بندی داده­ها برای تمام کاربران موجود و رتبه­بندی کاربران در تشخیص ناهنجاری صورت می­گیرد زیرا بدین ترتیب می­توان دقت الگوریتم را به­شکل کامل­تر مشخص نمود.

### 3-4-1 شرح روش پیشنهادی

یکی از روش‌هایی که تاکنون به منظور رفع مشکل انتخاب همسایگی در داده­های مختلف اینترنتی مورد استفاده قرار گرفته است، روش خوشه‌بندی می‌باشد. خوشه‌بندی باعث می‌شود که کاربران و یا اقلام (داده های) مشابه در یک خوشه قرار بگیرند. بنابراین، می‌توان از تعداد کاربران موجود در خوشه‌ی یک کاربر برای پیش‌بینی رتبه‌ی کاربران و تشخیص ناهنجاری استفاده کرد. اگرچه روش خوشه‌بندی می‌تواند راه‌حلی برای رفع مشکل انتخاب همسایگی در سیستم‌های اینترنتی باشد، اما این روش نیز با محدودیت‌ها و مشکلاتی مواجه است. تعیین اولیه تعداد خوشه‌ها، یک موضوع اصلی در روش‌های مبتنی بر خوشه‌بندی می‌باشد. زیرا کارایی این روش‌ها بستگی به تعریف اولیه‌ی تعداد خوشه‌ها دارد و در صورتی که تعداد اولیه مناسبی برای خوشه‌ها تعیین نشود، این روش‌ها نمی‌توانند کارایی خوبی داشته باشند. مشکل دیگر در مورد نامناسب بودن خوشه‌های تولید شده می‌باشد. نتایج ضعیف خوشه‌بندی، ممکن است باعث کاهش دقت پیش‌بینی و درنتیجه کاهش نرخ پوشش رتبه‌ها شود. این مشکل بیشتر زمانی اتفاق می‌افتد که در طی فرایند خوشه‌بندی، خوشه‌هایی بدست می‌آیند که تعداد اعضای این خوشه‌ها خیلی کم است و لذا این خوشه‌ها نمی‌توانند همسایگی‌های مناسبی را برای کاربران موجود در این خوشه‌ها فراهم کنند. اما دقیقا همین خوشه ها هستند که می­توانند در شناسایی ناهنجاری کمک بیشتری کنند. از طرف دیگر، بسیاری از سیستم‌های کاربری اینترنتی مبتنی بر خوشه‌بندی، فقط از معیار شباهت بین کاربران و یا آیتم‌ها (دراین­جا داده­ها یا تجویزها) برای خوشه‌بندی آن­ها استفاده می‌کنند. این کار باعث می‌شود که روش خوشه‌بندی نتواند خوشه‌بندی را به خوبی برای کاربرانی با نظرات کم و همچنین در داد‌ه‌های با تنکی بالا انجام دهد. بنابراین، استفاده از فاکتورهای دیگر به عنوان معیار شباهت، مانند روابط اعتماد، می‌تواند به این روش‌ها برای دسته‌بندی بهتر کاربران و یا آیتم‌های موجود در سیستم کمک کند. این روابط می­تواند تشابه بسیار مناسبی بین کاربران (تجویزکنندگان) بدست آورد.

در این فصل، یک روش با استفاده از روش‌های خوشه‌بندی و همچنین قوانین انجمنی به منظور تحلیل ناهنجاری در گراف متشکل از تجویزکنندگان دارویی مبتنی بر پیش­بینی رفتار کاربران و تاثیر آن­ها بر روی ناهنجاری ارائه شده است. برای این منظور، روش خوشه‌بندی k-means مبتنی بر تئوری‌ گراف برای دسته‌بندی کاربران و یا داده­های موجود در سیستم (تجویز) ارائه شده است. یکی از ویژگی‌های روش تحلیل ناهنجاری ارائه شده این است که این روش می­تواند بدون نیاز به برخی پیچیدگی­های موجود در روش­های پیشین تشخیص تقلب یا ناهنجاری در داده ها را انجام دهد. نتایج بدست آمده از آزمایشات بر روی مجموعه‌­های داده‌ای نشان خواهند که آیا روش پیشنهادی در مقایسه با سایر روش‌ها از عملکرد مناسبی برخوردار هست یا خیر اگرچه آن­چه مشخص است این روش در صورت داشتن مراحل مناسب و اجرای درست می­تواند بسیار موثر بوده و تحلیل مناسبی از داده های ناهنجار در تجویز پزشکی صورت دهد. در این پژوهش سعی شده است مراحل اجرای الگوریتم پیشنهادی منطبق بر ویژگی­های داده های تجویز پزشکی و قابلیت پیش­بینی رفتار تجویزکنندگان و تحلیل گراف در آن باشد.

در مرحله اول روش پیشنهادی، با نام پیش پردازش داده‌ها، می‌توانیم کاربران و داده­های متفاوتی را که در تجویز پزشکی وجود دارند، مشخص کنیم. در واقع، چنین کاربرانی با توجه به معیارهای متفاوتی مانند: تعداد تجویز شده، انحراف معیار بر روی داده‌ها و غیره مشخص می‌گردند. بنابراین داده‌های استخراج شده در این مرحله، در داخل یک پایگاه داده به‌منظور ورودی در فاز دوم و استفاده‌های بعدی ذخیره می‌شوند. جزئیات بیشتر در مورد انواع کاربران و داده‌ها به همراه نتایج بدست آمده برای هر یک از آنها در فصل ارزیابی نتایج مورد بررسی قرار خواهد گرفت.

بطورکلی دو مرحله در این فاز از اجرا وجود دارد. یکی استخراج برخی اطلاعات مهم از داده­های کاربران و سپس ذخیره آن­ها برای استفاده مجدد در مراحل مختلف. به عنوان مثال در یک سیستم ممکن است در یک زمان 100 تجویز وجود داشته باشد که کاربران آن­ها را صورت داده اند. از این میان ممکن است چند تجویز مثلا 7 قلم دقیقا تکراری باشد. درنتیجه 93 قلم جدید وجود دارد که در روش پیشنهادی مورد توجه قرار خواهد گرفت. همچنین به­عنوان مثال تعداد کاربران فعال که در تجویز شرکت نموده­اند نیز مشخص می­شود که داده های مربوط این کاربران در روش پیشنهادی مورد استفاده قرار خواهد گرفت. خروجی این مرحله اطلاعات مربوط به کاربران و داده ها است.

پس از این مرحله شبکه گراف کاربران محاسبه می شود. شبکه‌ی اعتماد کاربر هدف بر اساس ترکیب روابط اعتماد و معیار شباهت پیرسن[[14]](#footnote-14) به عنوان مقادیر نهایی شباهت ساخته می‌شود. این شبکه، در واقع یک گراف جهتدار و وزندار می‌باشد که همسایگان کاربر هدف رئوس این گراف را تشکیل می‌دهند و همچنین مقادیر نهایی شباهت بین کاربران، به عنوان وزن یال‌های این گراف در نظر گرفته می‌شوند. از رابطه‌ی (2-3) برای محاسبه‌ی اعتماد بین دو کاربر a و u استفاده شده است:

رابطه 3-2



که میزان اعتماد بین کاربر a و کاربر u می‌باشد، ماکزیمم طول انتشار اعتماد بین کاربران می‌باشد، و طول انتشار اعتماد بین کاربر a و کاربر u را نشان می‌دهد. مقدار به صورت تقریبی، برابر میانگین طول مسیر شبکه‌ی اعتماد می‌باشد. طول انتشار اعتماد برابر تعداد یال‌هایی است که در کوتاه‌ترین مسیر انتشار اعتماد، از کاربر اعتمادکننده[[15]](#footnote-15) به کاربر اعتمادشونده[[16]](#footnote-16) وجود دارد.



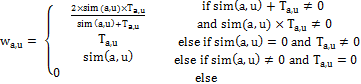
رابطه 3-3



که میانگین طول مسیر در شبکه‌ی اعتماد است، n اندازه‌ی شبکه‌ی اعتماد را مشخص می‌کند، و k میانگین درجه در شبکه‌ی اعتماد را نشان می‌دهد. سرانجام، وزن نهایی بین کاربر a و کاربر u که با نشان داده می‌شود، به صورت زیر محاسبه می‌گردد:



رابطه 3-4



که تابع شباهت بین کاربر a و کاربر u می‌باشد و با استفاده از تابع شباهت پیرسن به صورت زیر بدست می‌آید:



رابطه 3-5



که رتبه‌ی داده شده به داده i توسط کاربر u می‌باشد، میانگین رتبه‌های داده شده توسط کاربر u می‌باشد، و مجموعه‌ای از داده‌ها می‌باشد که توسط هر دو کاربر a و u رتبه داده شده‌اند.



پس از محاسبه‌ی وزن‌های شباهت نهایی بین کاربر هدف و دیگر کاربران، K نزدیک‌ترین همسایه برای کاربر هدف بر اساس این وزن‌های شباهت (رابطه‌ی 3-4) انتخاب می‌شوند. شبکه‌ی اعتماد کاربر هدف بر اساس این همسایگان انتخاب شده، ساخته می‌شود و از این شبکه‌ی اعتماد برای پیش‌بینی رتبه‌های کاربر هدف استفاده می‌گردد.

### 3-4-2 قواعد انجمنی

در شروع کار داده­هایی را که کاربران تجویز نموده‌اند را با استفاده از قوانین انجمنی مشخص نموده و از این طریق تجویزهای پرتکرار بین آنها پیدا می­شود. درنتیجه این مرحله همسایه­ کاربران برحسب علایق آن­ها در تجویز قابل تشخیص است. الگوریتم مد نظر شامل سه آیتم زیر می‌باشد که هر کدام به تفکیک تشریح شده‌اند و این مفاهیم در قوانین انجمنی به صورت زیر هستند [51][50]:

1 –پشتیبان[[17]](#footnote-17): نسبت تعداد تراکنش‌هایی که در آن اشیاءA وB هر دو حضور دارند، به کل تعداد رکوردها پشتیبان بدست می‌آید که دارای مقداری عددی بین صفر و یک می‌باشد و هر چه این میزان بیشتر باشد، نشان می‌دهد که این دو شیء بیشتر با هم در ارتباط هستند.

2 –اطمینان[[18]](#footnote-18)**:** معیار اطمینان نیز مقداری عددی بین صفر و یک می‌باشد، که هر چه این عدد بزرگتر باشد بر کیفیت قانون افزوده خواهد شد. استفاده از این معیار به همراه پشتیبان مکمل مناسبی برای ارزیابی قوانین انجمنی خواهد بود.

3 –Lift : معیارLift که با نام‌هایIntersect Factor یا Interestingness نیز شناخته می‌شود اشاره کرد، که این معیار میزان استقلال میان اشیاء A و B را نشان می‌دهد که می‌تواند مقدار عددی بین صفر تا بی نهایت باشد. در واقع Lift میزان هم اتفاقی بین ویژگی‌ها را در نظر می‌گیرد و میزان رخداد تکی بخش تالی قانون (یعنی شیءB) را در محاسبات خود وارد می‌کند.مقادیر نزدیک به عدد یک معرف این هستند که A و B مستقل از یکدیگر می‌باشند، بدین ترتیب نشان دهنده قانون جذابی نمی‌باشند. چنانچه این معیار از عدد یک کمتر باشد، نشان دهنده این است که A وB با یکدیگر رابطه منفی دارند. هر چه مقدار این معیار بیشتر از عدد یک باشد، نشان دهنده این است کهA اطلاعات بیشتری درباره B فراهم می‌کند که در این حالت جذابیت قانون بالاتر ارزیابی می‌شود. در ضمن این معیار نسبت به سمت چپ و راست قانون متقارن است در واقع اگر سمت چپ و راست قانون را با یکدیگر جابجا کنیم، مقدار این معیار تغییری نمی‌کند. ترکیب این معیار به همراه پشتیبان و اطمینان جزء بهترین روش‌های کاوش قوانین انجمنی است. استفاده از این قوانین در این راستا صورت می­پذیرد که بتوانیم کاربرانی را پیدا کنیم که بیشترین تشابه رفتاری را داشته باشد و داده­هایی که منتشر نموده­اند، بیشترین تشابه را بهم داشته باشند. بنابراین داریم:

F =  *رابطه 3-6*



استفاده از فرمول3-6 باعث می‌شود، داده­های مشترکی را که کاربران U وV ارسال نموده‌اند را مشخص نمائیم و میزان اشتراک دو کاربر بدست آید. rvi و rui بیانگر رتبه­بندی آیتم i توسط کاربر vو u می‌باشد.

در این تحقیق جهت مشخص کردن علایق یا عدم علایق کاربران به یک خبر یا شایعه و میزان تاثیرگذاری آن­ها در انتشار اخبار از دفعات ارسال آن به دیگران، استفاده شده و از این طریق مقایسه بین کاربران صورت می­پذیرد. بر این اساس اگر دفعات ارسال خبر مورد نظر کمتر از میانگین بدست آمده بود، آن خبر به عنوان عدم علاقه­مندی کاربر در نظر گرفته شده و اگر برابر و یا بیشتر از مقدار میانگین بود، آن خبر به عنوان علاقه­مندی کاربر در نظر گرفته می­شود. بر این اساس در رابطه 3-7 داریم:

A=رابطه 3-7



در رابطه 3-7، یعنی خبر u توسط کاربر r ارسال شده و امتیاز i را کاربر به آن داده است.n هم تعداد خبرهایی است که منتشر نموده است. حداقل مقدار i برابر با m می­باشد چون کاربر حداقل باید m خبر را ارسال نماید تا جزو کاربران فعال و تاثیرگذار محسوب شود. پارامتر m می­تواند با توجه به متوسط ارسال اخبار توسط کاربران وانحراف معیار آن مشخص شود. مثلا برای یک شبکه با متوسط 40 خبر و انحراف معیار 30 اگر یک کاربر بالای 10 خبر را انتشار بدهد، می­توان آن را کاربر فعال نامید. پس تعداد ارسال­هایی که کاربران (دارای ارتباط) بین خودشان انجام داده­اند را مورد مقایسه قرار داده و از این طریق علایق و عدم علاقه­مندی آنها نسبت به اخبار و شایعات موجود تشخیص داده می­شود چنانچه ارتباط مستقیم بین آنها برقرار باشد.

3-5 خوشه بندی کاربران

در این بخش به بررسی الگوریتم خوشه­بندی موردنظر برای اجرای روش پیشنهادی پرداخته می­شود. به­عنوان یک روش ساده الگوریتم خوشه­بندی k-means پیشنهاد می­شود اما یک روش خوشه­بندی گرافی به نام Louvain نیز مطرح خواهد شد که می­تواند به­عنوان جایگزین برای خوشه­بندی k-means بکارگرفته شود. اساسا هدف از این تحقیق ارائه طرح کلی تحلیل مسئله تقلب توسط خوشه­بندی بوده و الگوریتم خوشه­بندی مورد استفاده می­تواند بسته به کاربرد و نیازمندی­ها تغییر نماید.

3-5-1 روش خوشه بنديK – Means، C – Means يا C-Centeriod

اين روش علي رغم سادگي آن يك روش پايه و موثر محسوب مي شود که با این وجود که زمان بسیار زیادی از پیشنهاد آن گذشته است، همچنان مورد استفاده قرار می­گیرد اگرچه بهبودهای بسیاری در مورد آن مطرح شده است. اين روش روشي انحصاري و مسطح محسوب مي شود. براي اين الگوريتم شكل­هاي مختلفي بيان شده است. ولي همه آنها داراي روالي تكراري هستند كه براي تعدادي ثابت از خوشه ها سعي در تخمين موارد زير دارند :

* بدست آوردن نقاطي به عنوان مراكز خوشه ها اين نقاط در واقع همان ميانگين نقاط متعلق به هر خوشه هستند.
* نسبت دادن هر نمونه داده به يك خوشه كه آن داده كمترين فاصله تا مركز آن خوشه را دارا باشد.

در نوع ساده اي از اين روش ابتدا به تعداد خوشه هاي مورد نياز نقاطي به صورت تصادفي انتخاب مي شود. سپس در داده ها با توجه با ميزان نزديكي ( شباهت ) به يكي از اين خوشه ها نسبت داده مي شوند و بدين ترتيب خوشه هاي جديدي حاصل مي شود. با تكرار همين روال مي توان در هر تكرار با ميانگين گيري از داده ها مراكز جديدي براي آنها محاسبه كرد و مجدداً داده ها را به خوشه هاي جديد نسبت داد. اين روند تا زماني ادامه پيدا مي كند كه ديگر تغييري در داده ها حاصل نشود. تابع زير به عنوان تابع هدف مطرح است.

رابطه 3-9

كه || معيار فاصله بين نقاط و مركز خوشه j ام است.

الگوريتم زير الگوريتم پايه براي اين روش حسوب مي شود :

1.در ابتدا K نقطه به عنوان به نقاط مراكز خوشه ها انتخاب مي شوند.

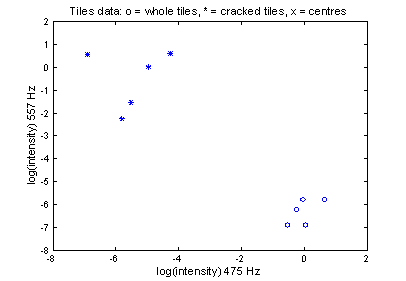
2. هر نمونه داده به خوشه اي كه مركز آن خوشه كمترين فاصله تا آن داده را داراست، نسبت داده مي شود.

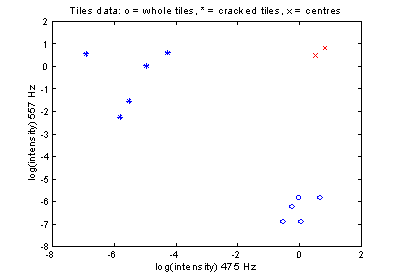
3. پس تعلق تمام داده ها به يكي از خوشه ها براي هر خوشه يك نقطه جديد به عنوان مركز محاسبه مي شود. ( ميانگين نقاط متعلق به هر خوشه )

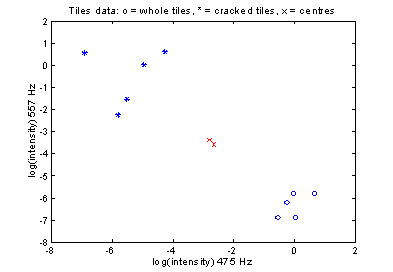
4. مراحل 2 و 3 تكرار مي شوند تا زماني كه ديگر هيچ تغييري در مراكز خوشه ها حاصل نشود.

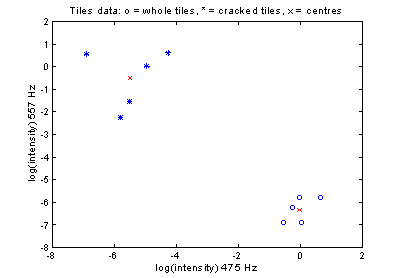
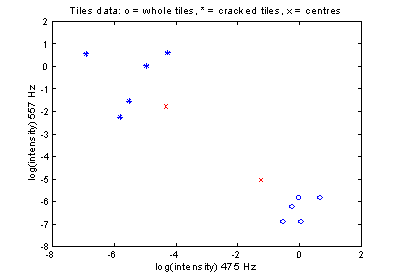
* مثالي براي روش خوشه بندي K-Means

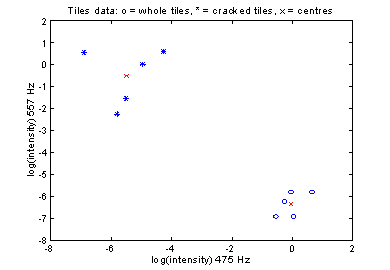
در شكل 3-2 نحوه اعمال اين الگوريتم خوشه بندي روي يك مجموعه داده كه شامل دوگروه داده است نشان داده شده است. يك گروه از داده ها با ستاره وگروه ديگر با دايره مشخص شده اند.(a). در مرحله اول نقطه اي به عنوان مركز خوشه ها انتخاب شده اند كه با رنگ قرمز نشان داده شده اند (b). سپس در مرحله دوم هر يك از نمونه داده ها به يكي از اين دوخوشه نسبت داده شده است و براي هر دسته جديد مركزي جديد محاسبه شده است كه در قسمت c نشان داده شده اند. اين روال تا رسيدن به نقاطي كه ديگر تغيير نمي كنند، ادامه پيدا كرده است( شكل 2-11).











شکل 3-2: مثالي براي روش خوشه‌بندي K-Means

### 3-5-2 الگوریتم خوشه­بندی Louvain

یکی از الگوریتم‌های سریع و کارا برای خوشه بندی گرافی، الگوریتم Louvain است که خوشه‌بندی گراف را با استفاده از حداکثر کردن تابع پودمانی انجام می‌دهد. در ابتدای شروع الگوریتم، هر گره به‌عنوان یک خوشه در نظر گرفته می‌شود و در ادامه، خوشه‌بندی در دو قدم تکرارشونده به‌صورت زیر انجام می‌شود:

قدم اول: برای هر گره منفعت نسبت دادن آن گره به خوشه با استفاده از رابطه زیر محاسبه می‌شود.



|  |
| --- |
|  |

رابطه 3- 8

که در آن مجموع وزن‌ها در خوشه است، مجموع وزن یال‌هایی است که به گره‌های خوشه وصل می‌شود مجموع یال‌های گره و نشان‌دهنده مجموع وزن یال‌هایی است که یک سر آن گرهi است و یک سر دیگر آن نیز خوشه است. همچنین m برابر مجموع وزن تمام یال‌های گراف است.



قدم دوم: هر گره به خوشه‌ای که تابع پودمانی را حداکثر کند نسبت داده می‌شود. سپس بر اساس این ساختار جدید، خوشه‌ها دوباره ساخته می‌شوند. این دو قدم به‌صورت پشت سر هم تا زمانی که دیگر تغییری در ساختار خوشه‌ها ایجاد نشود، تکرار می‌شوند.

از مزیت‌های الگوریتم تشخیص جوامع Louvain می‌توان به ساده و تکراری بودن آن اشاره کرد که تحلیل و پیاده‌سازی آن را بسیار آسان کرده است. همچنین تعداد خوشه‌ها به‌صورت خودکار تعیین شده و قبل از اجرای خوشه‌بندی نیازی نیست که در مورد ساختار داده‌ها اطلاعاتی وجود داشته یاشد. به‌علاوه، ازلحاظ پیچیدگی محاسباتی، الگوریتم Louvian بسیار کارا بوده و دارای پیچیدگی زمانی است کهn تعداد گره‌ها را نشان می‌دهد. به همین علت این الگوریتم برای گراف‌هایی با تعداد گره‌های بسیار زیاد و تا چند میلیون گره نیز قابل‌استفاده است. الگوریتم تشخیص جوامع Louvain در بسیاری از نرم‌افزارهای تحلیل شبکه مانند Gephi و NetworkX برای خوشه‌بندی گراف مورد استفاده قرار می‌گیرد.



### 3-5-3 روال خوشه­بندی

روال کار بدین صورت است که تعدادی از کاربرانی که تشابه رفتاری بین آنها برقرار است را بصورت تصادفی انتخاب می‌شود و برای هر کدام بازه­ای در نظر گرفته می­شود ( هر بازه به عنوان یک خوشه)؛ سپس داخل هر خوشه شباهت تاثیرگذاری کاربران با کاربر هدف در نظر گرفته می­شود و هر خوشه­ای که پروفایل اعضای آن بیشتر به کاربر نزدیک باشند، طبیعتاٌ بدان معنا است که تاثیرگذاری این کاربران هم مشابه هم می­باشد؛

بنابراین این خوشه انتخاب می­شود. جهت پیدا کردن شباهت تاثیرگذاری بین کاربران از فرمول زیر استفاده می‌شود [52] :

رابطه 3- 10

Rc , c' = rc, c'

با روش تجمیع[[19]](#footnote-19) شباهت بین کاربر *C* و *C'* بر اساس آیتم‌های پروفایل یعنی(C,C')sim اندازه‌گیری شده و بر حسب رتبه وزنی می‌باشد. هر اندازه کاربر *C* و *C'* به هم شبیه‌تر باشند، ضریب rc, c' تاثیر بیشتری در انتخاب کاربر نهایی دارد. این فرمول در این تحقیق با توجه به نوع کاربرد مقداری تغییر پیدا کرده است و در این تحقیق از این فرمول جهت شباهت بین تاثیرگذاری کاربران استفاده می‌شود.

بنابراین استفاده از فرمول (8-3) باعث می‌شود که تشابه رفتاری کاربران را تشخیص و بدست آورده و تجویزهای مشترکی را منتشر کردند را مشخص نمائیم و بکارگیری فرمول (10-3) باعث می‌شود کاربرانی را که بیشترین شباهت علاقمندی تجویز را با یکدیگر دارند را مشخص نموده و در نهایت کاربرانی پیدا می­شوند که بیشترین تشابه رفتاری (علایق و عدم علایق) و پروفایلی را با یکدیگر دارند. پس از آن کاربرانی که کمترین شباهت را با کاربران دیگر دارند مشخص شده و این کاربران گزینه مناسبی جهت تحلیل تقلب و ناهنجاری بوده و مشخص می­شود چه خوشه هایی از کاربران دارای اعضای کم است و این خوشه ها با احتمال بالاتری نشان دهنده ناهنجاری هستند.

3-6 پیش بینی رتبه‌ ناهنجاری

در این فاز با توجه به نتایج خوشه­بندی و تعیین کاربرانی که بیشترین شباهت را به یکدیگر دارند در مرحله قبل، رتبه کاربران (تجویزها) پیش­بینی خواهد شد و کمترین رتبه ها یعنی کمترین شباهت بین نسخه ها تعیین کننده موارد مشکوک به ناهنجاری است. چنان­چه بیان شد در این پژوهش مجموع رتبه تجویز در کاربران خوشه­ها، معرفی­کننده رتبه نهایی خواهد بود. بنابراین در مرحله اول از این فاز مجموع رتبه­ها برای تجویزهای مختلف یک خوشه محاسبه شده و خروجی مجموع رتبه­ها برای تجویز مختلف بوده و در مرحله دوم برای کاربران آن خوشه­ها، این مسئله تعیین کننده رتبه تجویز در سیستم پیش­بینی تقلب است. خروجی مرحله دوم نیز تجویز رتبه­بندی شده است که تجویز با بیشرین رتبه تجویز معمول و تجوز با کمترین رتبه ها موارد ناهنجار و مشکوک را نشان می دهد.

7-3 جمع­بندی

در این فصل در مورد روش پیشنهادی برای تحلیل ناهنجاری در تجویز پزشکی با استفاده از خوشه­بندی صحبت شد. این روش مبتنی بر داده­های بدست آمده از تجویز پزشکی براساس کاربران و رفتار آن­ها و ارتباطی که بین کاربران وجود دارد، بنا شد،ه قواعد انجمنی لازم را برای مشخص نمودن تجویزهای پرتکرار بکار می­گیرد و خوشه­بندی را با توجه به کاربران و تجویز منتشر شده توسط آنان صورت می­دهد. تجویزهای معمول را مشخص نموده، درنهایت مشخص می­شود که چه تجوزهایی غیرعادی می­توانند باشند که آن تجویزها معمولا در خوشه هایی با کمترین تعداد عضو و با رتبه پایین خواهند بود.

مراجع

[37] Vega-Oliveros, Didier A., et al. "A multi-centrality index for graph-based keyword extraction." *Information Processing & Management* 56.6 (2019): 102063.

[38] Schoch, David, and Ulrik Brandes. "Re-conceptualizing centrality in social networks." *European Journal of Applied Mathematics* 27.6 (2016): 971-985.

[39] Schoch, David, and Ulrik Brandes. "Re-conceptualizing centrality in social networks." *European Journal of Applied Mathematics* 27.6 (2016): 971-985.

[40] Li, Yong, et al. "Hierarchical decomposition for betweenness centrality measure of complex networks." *Scientific reports* 7 (2017): 46491.

[41] Eidsaa, Marius, and Eivind Almaas. "Investigating the relationship between k-core and s-core network decompositions." *Physica A: Statistical Mechanics and its Applications* 449 (2016): 111-125.

[42] Bhattacharjee, Sandipan, et al. "Medication prior authorization from the providers perspective: a prospective observational study." *Research in Social and Administrative Pharmacy* 15.9 (2019): 1138-1144.

[43] Parand, Fereshteh-Azadi, Hossein Rahimi, and Mohsen Gorzin. "Combining fuzzy logic and eigenvector centrality measure in social network analysis." *Physica A: Statistical Mechanics and its Applications* 459 (2016): 24-31.

[44] Al-garadi, Mohammed Ali, Kasturi Dewi Varathan, and Sri Devi Ravana. "Identification of influential spreaders in online social networks using interaction weighted K-core decomposition method." *Physica A: Statistical Mechanics and its Applications* 468 (2017): 278-288.

[45] Alshurafa, Nabil, et al. "Remote health monitoring outcome success prediction using baseline and first month intervention data." *IEEE journal of biomedical and health informatics* 21.2 (2016): 507-514.

[46] Burchard, Jake, and Benjamin Cornwell. "Structural holes and bridging in two-mode networks." *Social Networks* 55 (2018): 11-20.

[47] Chen, Yu, et al. "Network formation under random attack and probabilistic spread." *arXiv preprint arXiv:1906.00241* (2019).

[48] Agliari, Elena, et al. "Phase Transition for the Maki–Thompson Rumour Model on a Small-World Network." *Journal of Statistical Physics* 169.4 (2017): 846-875.

[49] Panagiotou, Konstantinos, and Leo Speidel. "Asynchronous rumor spreading on random graphs." *Algorithmica* 78.3 (2017): 968-989.

[50] Harahap, M., et al. "Mining association rule based on the diseases population for recommendation of medicine need." *Journal of Physics: Conference Series*. Vol. 1007. No. 1. IOP Publishing, 2018.

[51] Kareem, Saba, Rohiza Binti Ahmad, and Aliza Binit Sarlan. "Framework for the identification of fraudulent health insurance claims using association rule mining." *2017 IEEE Conference on Big Data and Analytics (ICBDA)*. IEEE, 2017.

[52] Meena, K., and K. Veena. "Performance evaluation of cybercriminal detection through cluster computing techniques." *Journal of Ambient Intelligence and Humanized Computing* (2019): 1-8.

[53] Datta, Subrata, and Kalyani Mali. "Trust: a new objective measure for symmetric association rule mining in account of dissociation and null transaction." *2016 Eighth International Conference on Advanced Computing (ICoAC)*. IEEE, 2017.

[54] Zheng, Baokun, et al. "Malicious bitcoin transaction tracing using incidence relation clustering." *International Conference on Mobile Networks and Management*. Springer, Cham, 2017.

1. Component [↑](#footnote-ref-1)
2. Betwenness centrality [↑](#footnote-ref-2)
3. Pagerank centrality [↑](#footnote-ref-3)
4. Degree centrality [↑](#footnote-ref-4)
5. Betweenness centrality [↑](#footnote-ref-5)
6. Eigenvector centrality [↑](#footnote-ref-6)
7. Pagerank centrality [↑](#footnote-ref-7)
8. Closeness centrality [↑](#footnote-ref-8)
9. Clustering coefficient [↑](#footnote-ref-9)
10. Structural Holes [↑](#footnote-ref-10)
11. ego [↑](#footnote-ref-11)
12. Online [↑](#footnote-ref-12)
13. Offline [↑](#footnote-ref-13)
14. Pearson Correlation Coefficient Measure [↑](#footnote-ref-14)
15. Trustor [↑](#footnote-ref-15)
16. Trustee [↑](#footnote-ref-16)
17. Support [↑](#footnote-ref-17)
18. Confidence [↑](#footnote-ref-18)
19. Aggregation [↑](#footnote-ref-19)