JCWR2022
May 11-12, 2022; Tehran, Iran
8th International Conference on Web Research

Community Detection and Influencer Nodes in Information Propagation about the Coronavirus using Twitter Social Network Analysis

Faezeh Alizadeh[1], Zainabolhoda Heshmati*[2]

[1] PhD Student of Information Technology, Faculty of New Sciences and Technologies, University of Tehran, Tehran, Iran.
alizadeh.fa.72@gmail.com

[2] Assistant Professor of Department of Network Science and Technology, Faculty of New Sciences and Technologies, University of Tehran, Tehran, Iran.
zheshmati@ut.ac.ir

Abstract

The mass volume of information posted on social networks has attracted increasing attention due to the variety of applications that can be used. In fact, these virtual communities enable users to express their viewpoints as well as influence other users. Therefore, it is very important to identify the key influencers of the social network and the communities formed around a topic, using the method of social network analysis. The connection between different entities in the network will help to understand this issue. In this study, conversations about the new coronavirus have been examined on Twitter. The illustrated network is examined in Gephi software by one of the community detection algorithms such as Louvain. HITS and PageRank algorithms are also used to analyze the impact of information propagation and ranking of entities.

Keywords: Social networks, Influencer nodes, Coronavirus, Complex network analysis.

JCWR2022
May 11-12, 2022; Tehran, Iran
8th International Conference on Web Research

شناسایی انجمن‌ها و گره‌های موثر در انتشار اطلاعات مربوط به ویروس کرونا به کمک تحلیل شبکه اجتماعی توئیتر

فائزه علیزاده[1]، زینب الهدی حشمتی*[2]

[1] دانشجوی دکتری مهندسی فناوری اطلاعات، دانشکده علوم و فنون نوین، دانشگاه تهران، تهران
alizadeh.faezeh@ut.ac.ir

[2] استادیار گروه علوم و فناوری شبکه، دانشکده علوم و فنون نوین، دانشگاه تهران، تهران
zheshmati@ut.ac.ir

چکیده

حجم انبوه اطلاعات ارسال شده در شبکه‌های اجتماعی، به دلیل زمینه‌های کاربردی متنوع که می‌توان از آنها استفاده کرد، توجه فزاینده‌ای را به خود جلب کرده است. در واقع، این اجتماعات مجازی کاربران را قادر می‌سازند تا نظرات خود را بیان کنند و همچنین بر سایر کاربران تأثیر بگذارند. بنابراین، شناسایی تأثیرگذاران کلیدی شبکه اجتماعی و انجمن‌های تشکیل شده پیرامون یک موضوع، با بهره‌گیری از روش تحلیل شبکه‌های اجتماعی، بسیار مهم است. ارتباط میان موجودیت‌های مختلف در شبکه به درک این مسئله کمک خواهد کرد. در این پژوهش نظرات پیرامون ویروس جدید کرونا در شبکه‌ی توئیتر مورد بررسی قرار گرفته است. شبکه-ی مصور شده، در نرم‌افزار گفی، توسط یکی از الگوریتم‌های تشخیص انجمن همچون Louvain مورد بررسی قرار گرفته است. همچنین از الگوریتم‌های HITS و PageRank برای تجزیه و تحلیل تأثیر انتشار اطلاعات و رتبه‌بندی موجودیت‌ها استفاده شده است.

کلمات کلیدی

شبکه‌های اجتماعی، گره‌های موثر، ویروس کرونا، تحلیل شبکه‌های پیچیده.

۱- مقدمه

شبکه‌های اجتماعی شیوه ارتباط مردم را تغییر داده‌اند. میلیون‌ها نفر در سراسر جهان از رسانه‌های اجتماعی برای ایجاد، اشتراک‌گذاری و بحث در مورد اطلاعات و شبکه‌سازی استفاده می‌کنند. هر کسی که به اینترنت متصل است می‌تواند تجربیات خود را از طریق یک فیلم کوتاه، به اشتراک بگذارد، نظر شخصی خود را از طریق یک توئیت ارائه دهد، یا از طریق یک تصویر چیز جالبی را به جهان نشان دهد. این امر باعث شده است که تجزیه و تحلیل شبکه‌های اجتماعی در سال‌های گذشته به دلیل میزان داده‌هایی که به طور مداوم در اینترنت به اشتراک گذاشته می‌شود، به رشد تصاعدی برسد [1].

تحلیل شبکه‌های اجتماعی به عنوان روشی برای تحلیل داده‌ها با هدف پیدا کردن الگوهای ارتباطات میان یک گروه خاص است که برای تحلیل شبکه‌های شکل گرفته در علوم اجتماعی، اقتصادی، سیاسی، شبکه‌های کامپیوتری و شبکه‌های اجتماعی به کار می‌رود [2]. هر شبکه از گره‌ها و یال‌هایی ایجاد شده است. گره‌ها موجودیت یا اعضای هر شبکه و یال‌ها روابط میان این اعضا هستند. ایده اصلی در تحلیل شبکه اجتماعی ترسیم روابط میان موجودیت‌های مختلف در یک شبکه است. این ارتباط می‌تواند نشان-دهنده دوستی، خویشاوندی، تجارت، علاقه‌مندی به موضوعی خاص و موارد اینچنینی باشد. چگونگی ساختار این روابط در تحلیل شبکه‌های اجتماعی مهم است، زیرا باعث کشف الگوها، درک بهتر داده‌ها و تجسم اطلاعات کاربران می‌شود.

در شبکه‌های اجتماعی، پیرامون موضوعات مختلف بحث و گفتگو شکل می‌گیرد. یکی از موضوعاتی که این روزها در کشور و جهان در شبکه اجتماعی توئیتر مطرح شده است، نظرات مختلف پیرامون ویروس جدید کرونا است. انتشار اطلاعات در مورد این ویروس در شبکه‌های اجتماعی می‌تواند به

شدت بر رفتارهای مردم و بر اثربخشی اقدامات ایجاد شده توسط دولت‌ها و مقامات بهداشتی برای مهار ویروس تأثیر بگذارد.

برخی افراد در شبکه‌ی تشکیل شده از توئیت‌ها یا نظرات ویژگی‌هایی دارند، همچون تخصص یا ارتباطات زیاد، که آنها را از دیگران متمایز کرده و به آنها اجازه می‌دهد تا بر تعداد زیادی از افراد در شبکه خود تاثیر بگذارند [3]. این افراد می‌توانند در انتشار اطلاعات نادرست/درست در مورد همه‌گیری ویروس کرونا نقش داشته باشند. بنابراین شناسایی آنها مهم است، زیرا ممکن است باعث افزایش شایعات و اطلاعات مشکوک شوند.

در گفتگوهای تشکیل شده، افراد مخالف و موافق نظرات خود را بیان می‌کنند و از هشتگ‌هایی برای انتشار توئیت‌های خود استفاده می‌کنند. این باعث تشکیل انجمن‌هایی در شبکه می‌شود که کاربران مشابه ارتباطی بینشان برقرار می‌شود. مقالاتی پیرامون این موضوع منتشر شده است [6]–[4]. در این تحقیق، از یک منظر به دنبال کشف انجمن‌ها به کمک الگوریتم Louvain و بررسی روابط درون آنها هستیم. از منظر دیگر، کاربران موثر در انتشار اطلاعات مربوط به ویروس کرونا به کمک دو الگوریتم رتبه‌بندی معروف HITS[1][7] و PageRank[8]، در توئیت‌های فارسی، بررسی شده است. در این تجزیه و تحلیل بر تعاملات بین کاربران تمرکز شده که برخی از آنها شامل توئیت‌ها، بازتوئیت‌ها، پاسخ‌ها، هشتگ، نقل‌قول‌ها و منشن‌ها[2] است.

در ادامه، در بخش ۲ پیشینه تحقیق مورد بررسی قرار گرفته است. در بخش ۳ الگوریتم تشخیص انجمن Louvain و الگوریتم‌های رتبه‌بندی HITS و PageRank مطالعه شده است. بخش ۴ نتایج بررسی این الگوریتم-ها بر توئیت‌های فارسی با موضوع مرتبط با ویروس کرونا مورد تحلیل قرار گرفته است، و در بخش آخر نتیجه‌گیری و پیشنهادها آورده شده است.

۲- پیشینه تحقیق

در زمینه تحلیل و بررسی توئیت‌های با موضوع ویروس کرونا در شبکه اجتماعی توئیتر کارهای متفاوتی انجام شده است. در این بخش برخی پژوهش‌های انجام شده در زمینه کشف اطلاعات و محاسبه معیارهای مرکزیت، در توئیت‌های به زبان‌های مختلف، مطالعه شده است.

ساراسواسی[3] و همکاران [9]، از تجزیه و تحلیل شبکه اجتماعی برای مطالعه شیوع ویروس کرونا در کارناتاکا[4]، هند و ارزیابی آن به عنوان ابزاری برای نظارت و کنترل شیوع استفاده کرده‌اند. تعداد ۱۱۴۷ مورد مثبت کرونا (میانگین سنی ۳۴ سال و ۶۱٫۹۹٪ ۱۱ تا ۴۰ سال) تجزیه و تحلیل شد. ابزارهای نرم‌افزاری Cytoscape و Gephi برای مصورسازی شبکه و تعیین ویژگی‌های شبکه شامل گره‌ها و یال‌ها (یال‌های جهت‌دار از منبع به بیماران مورد هدف) استفاده شدند. درجه خروجی برای ۱۹۹ گره (۱۷٫۳۵٪) ۱تا۴۷ بود و برای ۸۹ گره (۷٫۷۶٪) معیار بینیت[5] ۰٫۵ تا ۸۷ بود. مردان، میانگین درجه بالاتر و زنان میانگین بینیت بالاتری داشتند. شهر بنگالور[6] دارای بیشترین حجم پرونده در ایالت (۲۲۹، ۲۰٪) بود، اما تشکیل خوشه نسبتاً کم بود. ۳۴ (۲٫۹۶٪) گره موثر (با درجه خروجی بزرگتر از ۵) باعث ۶۰٪ از انتقال‌ها شدند. تجسم شبکه اجتماعی می‌تواند به مدیران مراقبت‌های بهداشتی اجازه دهد تا نقاط مهم در حال تکامل را علامت‌گذاری کنند و گره‌های کلیدی در انتقال را مشخص کنند. اولویت‌بندی این مناطق و افراد برای مهار دقیق می‌تواند به حداقل رساندن هزینه منابع کمک کند و به طور بالقوه به کاهش قابل توجهی در انتقال ویروس کرونا دست یابد.

کینت[7] و همکارانش [10] از تجزیه و تحلیل شبکه‌های اجتماعی برای شناسایی الگوهای انتشار مکانی و زمانی ویروس کرونا در سراسر جهان استفاده کرده‌اند. همه‌گیری این ویروس به طور گسترده در سراسر جهان گسترش یافته است. مدل‌های ریاضی زیادی برای بررسی نقطه عطف[8] و الگوی گسترش این ویروس پیشنهاد شده‌اند. در این مقاله از تجزیه و تحلیل شبکه‌های اجتماعی برای خوشه‌بندی ویژگی‌های آنها استفاده شده است. هدف آنها نشان دادن استفاده از تحلیل شبکه برای شناسایی خوشه‌های انتشار ویروس کرونا بود. تعداد تجمعی موارد آلوده[9] در کشورها/مناطق از گیت‌هاب دانلود شد. مدل پاسخ[10] برای ایجاد یک مدل پیش‌بینی کلی برای هر کشور/منطقه استفاده شد. پارامترهای مکان در قاره‌ها، چین و ایالات متحده مقایسه شد. نتایج نشان داد که سه خوشه (آسیای شرقی و اروپا تا آمریکا) با استفاده از تحلیل شبکه اجتماعی از هم جدا شدند، الگوهای گسترش مکانی و زمانی را با استفاده از تحلیل شبکه و ضرایب همبستگی[11] خوشه‌بندی کردند. شناسایی الگوهای پراکندگی در دو دوره از ۲۲ ژانویه تا ۲۷ مارس ۲۰۲۰ و از ۱ ژانویه تا ۱۶ فوریه ۲۰۲۱ انجام شده است. اهداف این مطالعه؛ (۱) نشان دادن استفاده از تحلیل شبکه اجتماعی برای بررسی ارتباط مبتلایان روزانه در مناطق؛ (۲) بررسی مسیرهای انتشار ویروس کرونا از چین به غرب آسیا، اروپا، آمریکای شمالی و آمریکای جنوبی؛ (۳) شناسایی کشورها/مناطق با الگوی پراکندگی مشابه استان هبی در چین؛ و (۴) یک داشبورد برای تفسیر بهتر تعداد موارد آلوده بر اساس کشور/منطقه است.

محمد حبیبی و همکاران [11] رسانه‌های اجتماعی برای ارتباط مردم را در مورد همه‌گیری ویروس کرونا حیاتی‌تر شناخته‌اند. در رسانه‌های اجتماعی، هشتگ‌ها حاشیه‌نویسی اجتماعی هستند که اغلب برای نشان دادن محتوای پیام استفاده می‌شوند. این به عنوان یک ابزار بصری و انعطاف‌پذیر برای ایجاد مجموعه عظیمی از گفتگوها در توئیتر قابل جستجو است. هدف این مقاله تجزیه و تحلیل هشتگ توئیت‌های اندونزیایی مربوط به این ویروس، با استفاده از تحلیل شبکه اجتماعی است. از این تکنیک برای تجسم مدل‌های شبکه و اندازه‌گیری مرکزیت برای یافتن تأثیرگذارترین هشتگ در شبکه استفاده شده است. آنها ۵۰۰۰۰۰ توئیت را از توئیتر بر اساس کلمات کلیدی مربوط به ویروس کرونا جمع‌آوری و تجزیه و تحلیل کرده‌اند. بر اساس نتیجه اندازه‌گیری مرکزیت درجه، نزدیکی، بینیت و مرکزیت بردار ویژه، هشتگ #corona هشتگی با بیشترین ارتباط با هشتگ‌های دیگر است. هشتگ #COVID19 هشتگی است که بیشترین نزدیکی را با سایر هشتگ‌ها دارد. هشتگ #corona هشتگی است که بیشتر به عنوان پلی عمل می‌کند که می‌تواند جریان اطلاعات مربوط به این ویروس را کنترل کند. هشتگ #coronavirus بر اساس لینک آنها مهم‌ترین هشتگ است.

القریشی و همکارانش [12] اطلاعات مربوط به ویروس کرونا را از توئیت‌های عربی [13] شناسایی کرده‌اند. از زمان اولین مورد تایید شده کرونا، اطلاعات به میزان زیادی در پلتفرم‌های رسانه‌های اجتماعی پخش می‌شد. انتشار اطلاعات در مورد همه‌گیری این ویروس می‌تواند به شدت بر رفتار افراد تأثیر بگذارد. بنابراین، شناسایی تأثیرگذاران اطلاعات در طول همه‌گیری آن

گام مهمی در جهت درک واکنش‌های عمومی و انتشار اطلاعات است. در این کار، تحلیلی را بر روی توئیت‌های عربی که در طول همه‌گیری ویروس کرونا جمع‌آوری شده‌اند، ارائه کرده‌اند. از هر دو الگوریتم HITS و PageRank برای تجزیه و تحلیل تأثیر انتشار اطلاعات استفاده شده و رتبه‌بندی کاربران را مقایسه کرده‌اند. نتایج نشان می‌دهد که هم HITS و هم PageRank زیرمجموعه‌ای مشابه از گره‌های موثر در انتشار اطلاعات را کشف کردند.

حجم عظیمی از اطلاعات در شبکه‌ی اجتماعی توئیتر در حال انتشار است که مورد[12]و همکاران [14] محتوای داده‌های آن را برای کشف توئیت‌های گمراه‌کننده درباره ویروس کرونا و شناسایی منبع توئیت‌ها با استفاده از تکنیک‌های پردازش زبان طبیعی تجزیه و تحلیل کردند. پروفایل کاربران برای تعیین اینکه آیا کاربران تخصص پزشکی دارند یا خیر، تجزیه و تحلیل شد. ۸۳ درصد از کاربران هیچ تخصص پزشکی ندارند. این مطالعه تأثیر خطرناک انتشار اطلاعات نادرست و اهمیت استفاده از تأثیرگذاران خارج از حوزه پزشکی را برای کمک به انتشار اطلاعات صحیح درباره ویروس به تعداد زیادی از مردم توصیف می‌کند.

مطالعات قبلی بر تجزیه و تحلیل محتوای انگلیسی، عربی، هندی، اندونزی توئیتر متمرکز بود. در این کار، سعی شده است تحلیل را در مورد شناسایی انجمن‌ها و تأثیرگذاران در انتشار اطلاعات مربوط به ویروس کرونا از توئیت‌های فارسی ارائه شود.

۳- الگوریتم‌های تشخیص انجمن و رتبه‌بندی

بسیاری از شبکه‌های پیچیده ساختار خوشه‌پذیر دارند، به این معنا که دارای گروه‌هایی است که ارتباط درونی بالا و ارتباط بیرونی پایینی دارند. به جمعیت‌هایی که در ساختار شبکه‌های اجتماعی دارای ارتباط بالای درونی و ارتباط پایین بیرونی هستند، انجمن گویند [15]. چنین جوامعی می‌توانند به عنوان بخش‌های نسبتا مستقلی با خصوصیات خاص خود در ساختار یک شبکه اجتماعی در نظر گرفته شوند. تشخیص این جوامع در شبکه‌ها چالشی بزرگ است. در سال‌های اخیر و در شاخه‌های مختلف علوم مانند فیزیک، زیست‌شناسی، علوم اجتماعی، جامعه‌شناسی، زیست‌شناسی، علوم رایانه‌ای و غیره از اهمیت زیادی برخوردار است [16]. هر یک از روش‌ها و الگوریتم‌های کشف انجمن بر اساس معیارها و روش‌های متفاوتی به تشخیص انجمن در شبکه‌های اجتماعی می‌پردازند. در ادامه از دو نمونه از روش‌های معروف تشخیص انجمن در شبکه صحبت شده است.

تشخیص انجمن مبتنی بر یال: برخی از الگوریتم‌های تشخیص انجمن مبتنی بر حذف یال هستند. عملکرد این الگوریتم‌ها به این صورت است که با حذف یال‌های پرارتباط در شبکه، انجمن‌ها را در شبکه‌های اجتماعی شکل می‌دهند. گیروان–نیومن[13] یکی از الگوریتم‌های این دسته است که با حذف ارتباطات گره‌هایی که بیشترین مقدار مرکزیت بینیت را دارند، به تشخیص انجمن‌ها می‌پردازد. از مهم‌ترین ایرادهای این الگوریتم، هزینه محاسباتی زیاد و همچنین تشخیص جوامع کوچک و بی‌معنی در شبکه‌های اجتماعی است [17].

تشخیص انجمن مبتنی بر شاخص پیمانه‌بندی[14]: در رویکرد دیگر، بیشینه‌سازی مقدار پیمانه‌بندی مبنای شناسایی بهترین ترکیب انجمن‌ها قرار می‌گیرد. پیمانه‌بندی به میزان تفکیک‌پذیری شبکه به گروه‌ها اشاره دارد و برای یک انجمن‌بندی از تفاضل میزان پیوندهای موجود درون‌گروهی به مقدار احتمال وجود این پیوندها به صورت تصادفی در شبکه‌های اجتماعی محاسبه می‌شود. روش‌های مبتنی بر پیمانه‌بندی یکی از پرکاربردترین روش‌های تشخیص انجمن است و در این رویکرد، الگوریتم Louvain یکی از بهترین الگوریتم‌ها از نظر دقت و هزینه‌های محاسباتی است [18].

۳-۱- تشخیص انجمن در شبکه‌های اجتماعی با استفاده از الگوریتم Louvain

شاخص پیمانه‌بندی معیاری برای سنجش کیفیت تقسیم گره‌ها به انجمن‌های مختلف ارائه می‌کند که به‌دلیل ساده و مؤثربودن آن به یکی از پرکاربردترین معیارها برای اندازه‌گیری کیفیت الگوریتم‌های تشخیص انجمن تبدیل شده است [15]. ایده اولیه تشخیص انجمن به‌روش پیمانه‌بندی بسیار ساده است: یک زیرشبکه اجتماعی از شبکه‌های اجتماعی اصلیْ زمانی انجمن نامیده می‌شود که تعداد ارتباطات بین اعضای آن بیشتر از زمانی باشد که این ارتباطات به‌صورت تصادفی ایجاد شود.

الگوریتم Louvain یکی از پرکاربردترین الگوریتم‌های مبتنی بر بیشینه‌سازی شاخص پیمانه‌بندی است. این الگوریتم شامل دو گام است که به‌طور متناوب اجرا می‌شود. فرض کنید با یک شبکهٔ وزن‌دار با n رأس سروکار داریم. این الگوریتم ابتدا هر رأس را یک انجمن در نظر می‌گیرد؛ یعنی در شروع الگوریتم به تعداد گره، انجمن داریم. سپس الگوریتم برای هر رأس i، انجمن همسایه j را به‌نحوی می‌یابد که به‌ازای حذف i از انجمن خودش و ملحق‌کردن آن به انجمن j، شاخص پیمانه‌بندی بیشینه شود و سپس رأس i را به انجمن j اضافه می‌کند. این عمل تنها درصورتی انجام می‌شود که میزان پیمانه‌بندی افزوده شود؛ درغیر این‌صورت، رأس i در انجمن خودش باقی می‌ماند. این عمل به‌طور مکرر برای تمامی رئوس تکرار می‌شود تا زمانی که دیگر تغییری اعمال نشود. در این مرحله گام اول در یک نقطه بهینه محلی متوقف می‌شود. نقطه‌ای که پیمانه‌بندی بیشتر با تغییر انجمن هیچ رأسی به دست نمی‌آید. سپس در گام دوم الگوریتم با ادغام گروه‌های کوچک که توانایی ایجاد گروه‌های بزرگ‌تر را دارند، انجمن‌بندی را ادامه می‌دهد. این دو گام تا جایی ادامه می‌یابند که تغییری در انجمن‌ها ایجاد نشود و شاخص هم‌پیمانگی نیز به حالت بیشینهٔ خود دست یافته باشد. مقدار پیمانه‌بندی (Q) به روش فرمول (۱) بدست می‌آید:

$$Q = \frac{1}{2m}\sum_{ij}\left[A_{ij} - \frac{k_i k_j}{2m}\right]\delta(c_i, c_j) \qquad (۱)$$

که در آن، A_{ij} وزن یال بین گره‌های i و j را نشان می‌دهد. k_i و k_j مجموع وزن یال‌های متصل به گره‌های، به ترتیب، i و j هستند. m مجموع تمام وزن‌های یال در گراف است. c_i و c_j انجمن‌های گراف هستند. تابع δ در

صورتی که c_i و c_j برابر باشند یک می‌شود و در غیر اینصورت صفر. هر چه این مقدار Q به یک نزدیکتر باشد انجمن‌ها بهتر از هم جدا شده‌اند.

۲-۳- الگوریتم‌های رتبه‌بندی کاربران

- **الگوریتم PageRank**

PageRank یا رتبه صفحه، به الگوریتمی گفته می‌شود که بر پایه آن موتورهای جستجو، صفحات وبی که به هدف جستجوگر نزدیکترند را در رده‌های بالاتری نسبت به دیگران قرار می‌دهد. الگوریتم PageRank اولین بار در سال ۱۹۹۸ توسط Larry Page و Sergey Brin در دانشگاه استنفرد ارائه شده است. با این روش کاربرانی که کلمه ویژه‌ای را جستجو می‌کنند می‌توانند ابتدا صفحات وبی را ببینند که هم به خواسته آن‌ها نزدیکتر است و هم بازدید بیشتری داشته‌است.

این الگوریتم به هر صفحه وب، امتیازی اختصاص داده و از این امتیاز با درنظر گرفتن پرس‌وجوی کاربر، جهت رتبه‌بندی صفحات وب استفاده می‌کند. رتبه هر صفحه با اختصاص وزن به لینکی که به آن صفحه داده شده است به دست می‌آید که مقدار این وزن به کیفیت صفحه‌ای که لینک در آن قرار گرفته بستگی دارد. بدین صورت لینک‌های صفحات مهم‌تر وزن بیشتری می‌گیرند. جهت مشخص کردن کیفیت صفحه‌های رجوع کننده، از رتبه آن صفحه که به صورت بازگشتی تعیین می‌شود، استفاده می‌شود. می‌توان آن را به صورت ریاضی به صورت فرمول (۲) بیان کرد.

$$PR(A) = \frac{1-d}{N} + d \sum_{U \in P(A)} \frac{PR(U_i)}{P(U_i)} \qquad (۲)$$

PR(A) یعنی PageRank صفحه A و N تعداد کل گره‌ها است. P(Ui) تعداد لینک‌های خروجی از گره A به گره‌های U1.,.,.Un است. P(A) مجموعه گره هایی است که به گره A اشاره می‌کند. d یک فاکتور تضعیف است که معمولا برابر ۰٫۸۵ در نظر گرفته می‌شود.

- **الگوریتم HITS**

الگوریتم HITS یکی از الگوریتم‌های رایج برای رتبه‌بندی صفحات است که در سال ۱۹۹۸ توسط Kleinberg ارائه شد. در این نوع روش‌ها برای هر پرس‌وجو تحلیل لینک انجام می‌شود. برای انجام تحلیل لینک، ابتدا می‌بایست گراف خاص پرس‌وجو به نام گراف همسایگی ساخته شود. در حالت ایده‌آل این گراف تنها شامل صفحات مرتبط با پرس‌وجو است. برای ساخت گراف همسایگی، ابتدا یک مجموعه از صفحات مربوط با پرس‌وجو به وسیله موتور جستجو بازیابی می‌شوند. سپس این الگوریتم برای هر گره در گراف همسایگی، به طور تناوبی دو امتیاز مرجع و کانون را محاسبه می‌کند و سپس گره‌ها را با توجه به این امتیازات رتبه‌بندی می‌کند.

در این گراف دو نوع گره وجود دارد. گره کانون، صفحاتی که به عنوان لیست‌های منبع عمل می‌کنند. گره مرجع، صفحاتی که محتویات مهمی دارند. این الگوریتم فرض می‌کند صفحه‌ای که به صفحات وب دیگر بیشتری اشاره می‌کند، کانون خوبی است و صفحه‌ای که صفحات وب بیشتری به آن اشاره می‌کنند، مرجع خوبی است. به طور بازگشتی می‌توان به این نتیجه رسید که صفحه‌ای که به تعداد مرجع‌های خوب بیشتری اشاره می‌کند، کانون بهتری است و صفحه‌ای که کانون‌های خوب بیشتری به آن اشاره می‌کنند، مرجع بهتری است. در نتیجه یک صفحه به طور همزمان می‌تواند کانون و مرجع خوبی باشد. این الگوریتم بازگشتی دارای دو مرحله است.

۱. مرحله نمونه‌گیری: در این مرحله یک مجموعه از صفحات مرتبط با پرس‌وجو جمع‌آوری می‌شوند،

۲. مرحله بازگشتی: در این مرحله، کانون و مرجع صفحات بدست آمده در مرحله اول، محاسبه می‌شود.

برای محاسبه کانون یک صفحه مانند p از فرمول (۳) استفاده می‌شود که مجموع مرجع صفحات مرجع p است.

$$H_p = \sum_{q \in L(p)} A_q \qquad (۳)$$

در فرمول (۳)، H_p امتیاز کانون صفحه p است و A_q امتیاز مرجع صفحه q است.

برای محاسبه مرجع صفحه p از فرمول (۴) استفاده می‌شود که مجموع کانون صفحات مرجع صفحه p است.

$$A_p = \sum_{q \in B(p)} H_q \qquad (۴)$$

در فرمول (۴)، A_p امتیاز مرجع صفحه p است و H_q امتیاز کانون صفحه q است.

٤- کشف انجمن‌ها و گره‌های تاثیرگذار در انتشار اطلاعات مربوط به کرونا

در این بخش، ابتدا به نحوه جمع‌آوری داده‌ها و ساخت شبکه پرداخته شده است و در ادامه‌ی آن از الگوریتم‌های یاد شده برای یافتن انجمن‌ها و گره‌های موثر در انتشار اطلاعات درباره‌ی ویروس کرونا استفاده و نتایج بدست آمده تحلیل و بررسی شده است.

- **جمع‌آوری داده‌ها**

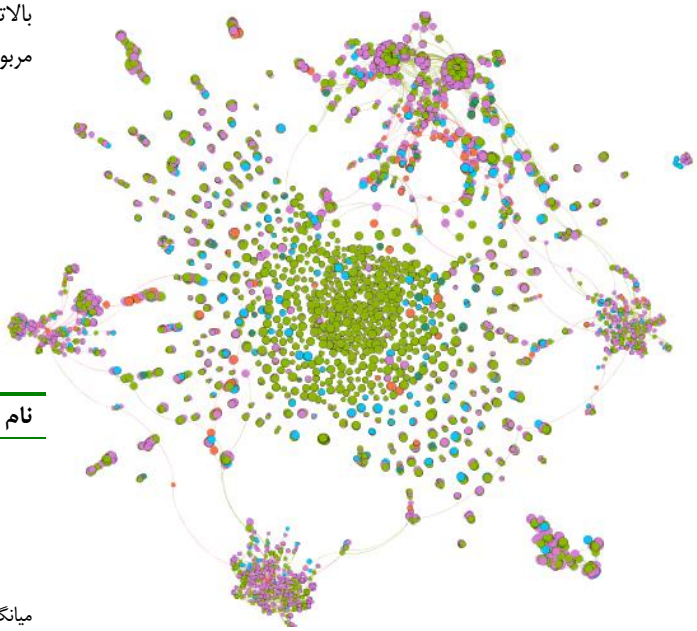
برای کشف انجمن‌ها و شناسایی گره‌های موثر در انتشارات اطلاعات مربوط به کرونا، از توئیت‌هایی از مجموعه داده فارسی در شبکه اجتماعی توئیتر استفاده شده است. مجموعه داده شامل ۵۶۷۰ گره و ۷۴۱۷ یال است. گره‌ها شامل کاربران، توئیت، رسانه، هشتگ و لینک هستند و یال بین آنها تعاملات بین کاربران است. این تعاملات شامل توئیت، بازتوئیت، هشتگ، لینک، رسانه، منشن و نقل قول است. به عنوان نمونه، اگر یک کاربر، توئیت کاربری دیگر را منتشر کرده باشد و یا پاسخ داده باشد، یک یال بین آنها صورت گرفته است. برای جمع‌آوری داده از افزونه API Streaming Importer در ابزار گفی استفاده شده است. توئیت‌ها بر اساس کلمات کلیدی #ویروس_کرونا، #کرونا، #کووید_۱۹، #کرونا_فریب_قرن، #ماسک و

#عوارض_کرونا گردآوری شده است. این مجموعه داده حاوی اطلاعاتی در مورد توئیت‌ها، پاسخ‌ها و بازتوئیت‌ها است که نشان دهنده‌ی یک گفتگوی عمومی بین کاربران در توئیتر است.

قبل از استفاده از داده، ستون‌هایی همانند long و place که خالی بوده و اطلاعاتی نداشت پاک شده‌اند. موقعیت مکانی ۸۲٫۸۴٪ از کاربران ذکر نشد و مقدار خالی در جدول داشته است. شکل (۱) گراف حاصل از گفتگوها است که بر اساس نوع گره تفکیک شده است و با استفاده از الگوریتم Force Atlas مصور شده است [19]. در جدول (۱) مشخصات داده‌های گردآوری شده آمده است.

- **ساخت شبکه**

در اینجا یک گراف جهت‌دار از $G = (V,E)$ تشکیل شده است که در آن V نشان‌دهنده‌ی کاربر است و E مجموعه‌ای از یال‌هایی است که نشان دهنده‌ی تعامل بین کاربران است. هر ارتباط در این گراف شامل مبدا و مقصد است که جهت بین آنها بر اساس نوع تعامل فرق می‌کند. به عنوان مثال، اگر کاربر A کاربر B را منشن می‌کند، یک یال از کاربر نویسنده A به کاربر B که منشن شده است ساخته می‌شود. هنگامی که یک توییت بازتوییت می‌شود یا توسط



شکل (۱). گراف تشکیل شده از گفتگوهای پیرامون ویروس کرونا با رنگ‌بندی بر اساس نوع گره

جدول (۱). مشخصات داده‌ی گردآوری شده

گره (تعداد)		یال (تعداد)	
توئیت	۲۶۴۸	هشتگ دارد	۸۹۸
		بازتوئیت	۱۱۷۰
کاربر	۲۴۲۱	توئیت	۲۶۴۸
		بازتوئیت از	۱۱۳۳
رسانه	۲۸۸	منشن دارد	۷۱۷
		رسانه دارد	۳۸۶
هشتگ	۲۵۰	نقل قول	۱۹۷
		نقل قول از	۱۹۷
لینک	۶۳	لینک دارد	۷۱

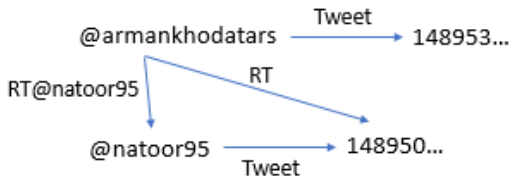
کاربر دیگری به اشتراک گذاشته می‌شود، شبکه‌ای از کاربران در معرض توئیت قرار می‌گیرند. بازتوئیت‌ها در شبکه با توئیت‌هایی که با *RT @user_name* شروع می‌شوند، شناسایی می‌شوند. یک نمونه تعامل در مثال شکل (۲) آمده است.

کاربری با نام @armankhodatars توئیتی را منتشر کرده است که توئیت آن با یک شماره یکتا مشخص است. کاربر دیگری با نام @natoor95 نیز توئیتی را منتشر کرده است. کاربر اول، کاربر دوم و توئیت وی را بازتوئیت کرده است و یک یالی بین آنها شکل گرفته است. شکل (۲) این تعاملات را نشان می‌دهد.

- **تجزیه و تحلیل داده‌ها**

در این بخش، به تجزیه و تحلیل داده‌ها به کمک معیارهای مرکزیت و الگوریتم شناسایی انجمن Louvain و الگوریتم‌های رتبه‌بندی HITS و PageRank پرداخته شده است. در جدول (۲)، مقدار میانگین درجات، الگوریتم پیمانه‌بندی و قطر شبکه آورده شده است.

در بررسی یافتن گره‌های موثر در انتشار اطلاعات پیرامون کرونا، گره-هایی از نوع هشتگ، توئیت و کاربر که امتیاز کانون و مرجع بدست آمده‌ی بالاتری داشتند شامل این موارد بوده‌اند: هشتگ با عنوان #کرونا، توئیت‌های مربوط به کمک در دوران همه‌گیری ویروس کرونا بوده که بیانگر امید و



شکل (۲). نمونه شبکه‌ی تشکیل شده از تعاملات کاربران

جدول (۲). اندازه‌گیری میانگین درجات، الگوریتم پیمانه‌بندی و قطر شبکه

نام معیار	مقدار	توضیحات
میانگین درجه	۱,۳۰۸	در اکثر شبکه‌های واقعی، توزیع درجات از قانون توان[۱۵] پیروی می‌کند. در اینجا نیز توزیع درجات را داریم که از این قانون پیروی می‌کند. تعداد زیادی گره داریم که درجه یک دارند و تعداد کمی گره داریم که درجه‌ی بالایی دارند. توزیع وزن‌دار درجات، برای یال‌های وزن‌دار در نظر گرفته می‌شود. در اینجا وزن یال یک است و نتیجه‌ی آن همان توزیع درجات می‌شود.
قطر شبکه	در حالت جهت‌دار ۱۰ است و میانگین طولانی-ترین مسیر ۲,۱۰۹ است. در حالت بدون جهت ۲۹ است و میانگین طولانی-ترین مسیر ۹,۹۵۸ است	قطر شبکه برابر است با طولانی‌ترین مسیری که در شبکه وجود دارد. علاوه بر این مرکزیت بینیت، مرکزیت نزدیکی، و بردار ویژه را بدست می‌آورد.
الگوریتم پیمانه‌بندی	۰,۹۳۴	این مقدار هرچه به عدد یک نزدیک باشد انجمن‌ها بهتر جدا شده‌اند. تعداد انجمن‌ها در این گراف ۷۵۴ تا شناسایی شده است.

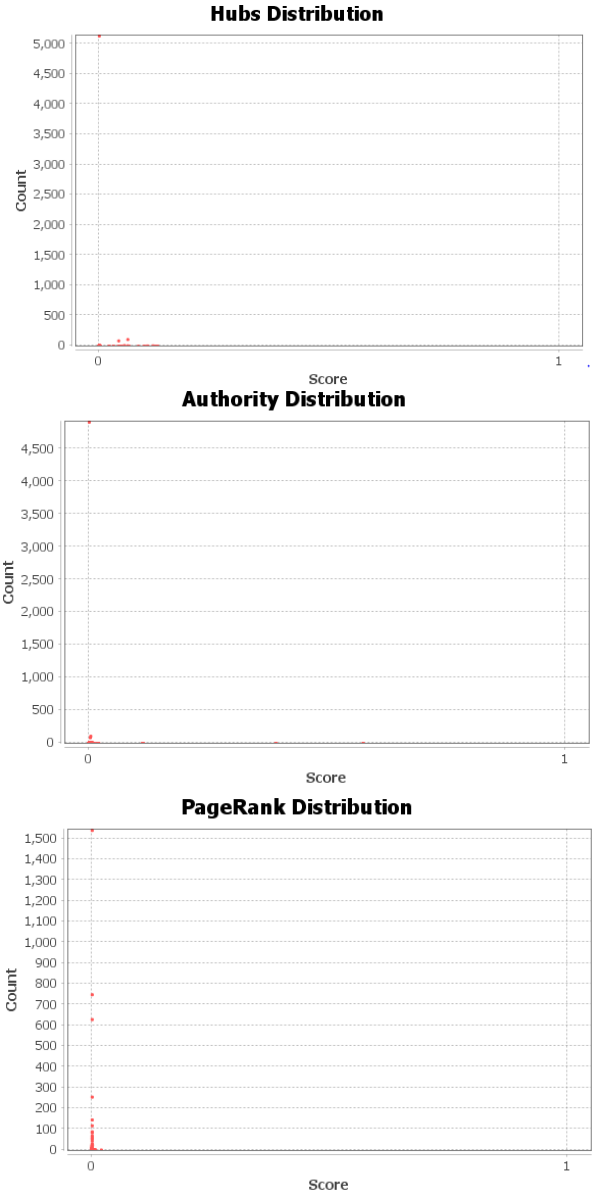
نشاط در کشور است، و تعدادی از کاربران، که تعداد دوستان و دنبال‌کننده‌های زیادی دارند، همچون کاربر @nahid52312 که ۲۸۲۹ دوست دارد و تعداد دنبال‌کننده‌های وی ۲۳۴۱ است. امتیاز PageRank نیز برای هشتگ #عوارض_واکسن، #سرایت_اُمیکرون و #گزارش_سیاسی، کاربران با تعداد دوستان و دنبال‌کننده‌های بالا و توئیت‌های مربوط به نگرانی‌هایی در مورد ویروس کرونا و عدم رعایت مردم و همچنین کمک‌های مومنانه، بالا بوده است. شکل (۳) نمودار توزیع خروجی این الگوریتم‌ها را نشان می‌دهد. توزیع الگوریتم‌ها نشان می‌دهد که تعداد گره‌های کمی مقدار کانون و مرجع بیشتری دارند و تعداد گره بیشتری مقادیر کمتری دارند. همچنین در محاسبه PageRank نیز این نتیجه حاصل شده است.

گره‌های از نوع توئیت، که البته بعضی از آنها بازتوئیت هستند، اطلاعات تخصصی در مورد بیماری کرونا نیستند و صحبت از مبتلا شدن به ویروس کرونا و اتفاقات و آمار روزانه در مورد آن است. همچون اینکه "معتادان کرونا نمی‌گیرند". اعتراضات در مورد بازگشایی مدارس و برگزاری امتحانات و نزدن ماسک، طنز در مورد کرونا و نظرات شخصی در مورد آن از جمله‌ی گفتگوها بوده است. که هر کدام از اینها انجمن‌هایی را تشکیل داده است. شکل (۴) به‌منظور استخراج انجمن‌های موجود در شبکه با استفاده از الگوریتم Louvain و به‌تفکیکِ رنگ هریک از انجمن‌ها نمایش داده شده است. همانطور که پیش‌تر بیان شد، این الگوریتم ۷۵۴ انجمن کشف کرده است که جدول (۳) محتوا و تعداد کاربران انجمن‌های شلوغ را نشان می‌دهد. در یک سری از انجمن‌ها امیدواری در مورد واکسن کرونا و بهبود وضعیت کشور در این دوره‌ی همه‌گیری صحبت شده و در یک سری از انجمن‌ها، برعکس آن، در مورد اعتراضات و ترکیب آن با بحث سیاسی بوده است. سایر انجمن‌ها از اعضای کمتری تشکیل شده‌اند و پیرامون گفتگوهای خودمانی بوده است.

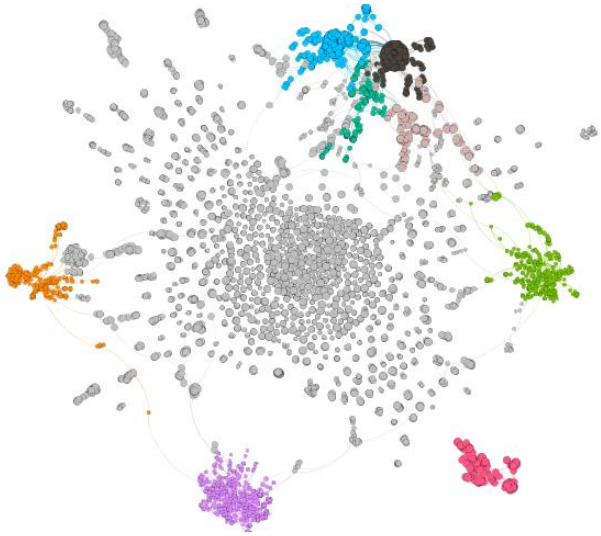
۵- نتیجه‌گیری و پیشنهادات

در این تحقیق، ابتدا مقالاتی در زمینه تحلیل و بررسی توئیت‌های با موضوع ویروس کرونا در شبکه اجتماعی توئیتر مطالعه شده است. برخی پژوهش‌های انجام شده در زمینه کشف انجمن‌ها و محاسبه معیارهای مرکزیت، در توئیت-های به زبان‌های مختلف، مورد بررسی قرار گرفته است. دو نوع معروف روش‌های تشخیص انجمن معرفی شده و در نهایت، گفتگوهای تشکیل شده به زبان فارسی پیرامون ویروس کرونا به کمک الگوریتم‌های معرفی شده، مورد تحلیل قرار گرفته است. همچنین، کاربران و توئیت‌های مهم و تاثیرگذار با الگوریتم‌های رتبه‌بندی شناسایی شده‌اند.

از بررسی‌های انجام شده و محاسبات انجام شده به کمک الگوریتم‌های رتبه‌بندی مشاهده می‌شود که حساب‌های کاربری و توئیت‌های موثر جزء حساب‌های فردی هستند و توئیت‌های آنها صرفا نظرات و تجربیات شخصی-شان از این ویروس است که بعضا با سیاست هم ترکیب شده است و مخالفان و موافقان حذب‌های مختلف در توئیت‌های خود درگیر مسائل سیاسی نیز شده‌اند. علاوه بر این، مسائل سیاسی و دولت را در اوضاع فعلی کشور در مورد شیوع این ویروس مقصر دانسته‌اند و حساب‌های سازمانی در اینجا فعالیتی نداشته‌اند و یا کمتر فعالیت داشته‌اند. این نشان می‌دهد، شبکه‌ای که میان این کاربران تشکیل می‌شود، حتی در موضوع بهداشتی و بیماری‌ها، می‌تواند



شکل (۳). توزیع الگوریتم HITS و PageRank



شکل (۴). انجمن‌های کشف شده از شبکه توئیت‌های پیرامون کرونا

جدول (۳). محتوای انجمن‌های کشف شده توسط الگوریتم Louvain

	رنگ	تعداد عضو	محتوا
۱	بنفش	۹٫۶۶٪ (۵۴۸)	هشتگ‌ها: #بزرگترین_کمک_مومنانه در این انجمن از کمک‌های مردمی در شرایط همه‌گیری کرونا صحبت شده که نشان‌دهنده‌ی امیدواری و نشاط در جامعه است. از کمک‌های بسیج و موسسات همدلی یاد شده است.
۲	سبز روشن	۵٫۸٪ (۳۲۹)	هشتگ‌ها: #وزارت_ضد_بهداشت، #ستاد_ایجاد_کرونا، #صدا_سیما، #واکسن_بخرید، #واکسن_نمی‌زنیم، #هواز، #نه_به_واکسن_اجباری، #دولت_انقلابی، #معجزات_واکسیناسیون در این انجمن در مورد واکسن زدن و عدم واکسن زدن اجباری گفتگو شکل گرفته است. همچنین اعتراضاتی در مورد واکسن بیان شده بود که به دولت مربوط می‌شد.
۳	آبی	۵٫۸٪ (۳۲۹)	هشتگ‌ها: #زبان_بدن در این انجمن بحث‌های سیاسی پیرامون مقامات خارجی و داخلی صورت گرفته است و کمتر در مورد کرونا صحبت شده است. در اینجا یکی از توئیت‌های با شماره ۱۴۸۹۵۵۱۱۸۶۴۲۰۱۹۵۳۳۰ مرکزیت بینیت ۱۳۵ دارد. کاربر iranintl@ درجه خروجی (۱۳۵) بالایی دارد و کاربر jhabiz@ درجه ورودی بالایی (۱۷) دارد.
۴	سیاه	۵٫۳۳٪ (۳۰۲)	هشتگ‌ها: #خبر در این انجمن بیشتر مباحث سیاسی مطرح شده و یک هشتگ با عنوان #اعتراضات_سراسری_معلمان نوشته شده که در مورد خودکشی یکی از معلمان گفته و ارتباطی به بحث کرونا نداشته است.
۵	نارنجی	۴٫۳۷٪ (۲۴۸)	هشتگ‌ها: #روایت_اقتدار، #از_انقلاب_راضیم، #عصر_شکوفایی، #ایران_قوی از ساخت واکسن در ایران صحبت شده است. امیدواری در مورد پیشرفت کشور با وجود تحریم‌ها، اعتماد به جوانان، گفتگو شکل گرفته است و از ایران به عنوان جزء ۶ کشور تولید واکسن نامبرده شده است.
۶	صورتی	۲٫۷۳٪ (۱۵۵)	در این انجمن بحث‌های خودمانی بین کاربرانی مبتلا به کرونا صحبت شده و از تجربیاتشان می‌گویند.
۷	سبز مات	۲٫۳۱٪ (۱۳۱)	هشتگ‌ها: #طنز، #خنده، #خنده_دار، #شانس و #جوک، #باکو، #آذربایجان، #ترکیه، #روسیه، #ایران، #آرژانتین، #اصفهان، #مشهد، #تهران، #مسعود_رجوی، #اعتراضات_ سراسری، #دادخواهی، #میکرون و #کوید۱۹ در این انجمن هشتگ‌های متنوع زیادی دیده شده است. همانطور که از هشتگ‌ها پیداست، بحث‌های سیاسی بیشتر از گفتگوهای کرونا بوده است.
۸	قهوه‌ای روشن	۲٫۰۸٪ (۱۱۸)	هشتگ‌ها: #پروتکل، #خبرگزاری_فارس، #رئیسی، #دیپلمات، #چین، #آمریکا، #پوتین گفتگو در مورد آخرین سفرهای خارجی دولت بوده و علاوه بر آن در مورد آمار کرونا و رعایت دستورالعمل‌های بهداشتی صحبت شده است.
۹	خاکستری	۱٫۶۸٪ (۹۵)	هشتگ‌ها: #موج_ششم، #فرهنگیان، #کرونا، #قیام_تنها_جوابه، #ماسک، #فاصله گذاری، # اُمیکرون در مورد اُمیکرون که به سرعت در حال انتشار است و از همه‌گیر شدن این سویه جدید نظراتی بیان شده است. همچنین، از اعتراض به دولت در بازگشایی مدارس و دانشگاه‌ها گفتگو شده است. در این گفتگو نام منبع خبر ذکر شده است. یوزر farzane25s@ درجه ورودی بالایی دارد و هشتگ #اُمیکرون درجه خروجی بالایی دارد.

اطلاعات نامناسبی را به جریان بیندازد و افراد نیز جدای از تخصصی که دارند اظهارنظر کنند. به طور کلی در هر انجمنی در کنار توئیت‌ها گلایه از مردم و دولت بوده است. در بازگشایی مدارس و دانشگاه‌ها و عدم رعایت دستورالعمل-های بهداشتی و کم‌کاری‌های دولت در واکسیناسیون از جمله‌ی اینها بوده است.

فعالیت حساب‌های سازمانی و توزیع اطلاعات تخصصی در انجمن‌هایی از گفتگوهایی که تشکیل می‌شود، می‌تواند اطلاعات مفیدی در اختیار کاربران قرار دهد تا از گمراهی آنها جلوگیری شود و موضوعات مختلف همچون سیاست و بهداشت، از هم تفکیک داده شوند.

همانطور که از محتوای انجمن‌ها پیداست، افراد زیادی به وضعیت فعلی با توجه به اثر مخرب کرونا بر کسب و کار، مرگ و میر زیاد، عدم رعایت دستورالعمل‌های بهداشتی توسط برخی مردم، معترض بوده‌اند. بنابراین، بهتر است اقدامات لازم و شفاف‌سازی صورت گیرد تا جامعه نسبت به موضوعات مختلف واکنش منفی نشان ندهد. در کنار این مسائل می‌توان مطالب علمی با هشتگ‌های مرتبط منتشر کرد تا افراد نسبت به اقدامات متقابل با ویروس و واکسیناسیون آگاهی کافی و لازم را داشته باشند.

تحلیل شبکه‌های اجتماعی، با توجه به فعالیت زیاد کاربران در آن، می-تواند در زمینه‌های مختلف به افراد خبره کمک کند تا صورت مسئله‌ای را بررسی کنند و راه‌حلی برای آن ارائه دهند.

مراجع

[1] G. Bello-Orgaz, J. J. Jung, and D. Camacho, “Social big data: Recent achievements and new challenges,” *Inf. Fusion*, vol. 28, pp. 45–59, 2016.

[2] D. Camacho, M. V. Luzón, and E. Cambria, “New research methods & algorithms in social network analysis,” *Future Generation Computer Systems*, vol. 114. Elsevier, pp. 290–293, 2021.

"ForceAtlas2, a continuous graph layout algorithm for handy network visualization designed for the Gephi software," *PLoS One*, vol. 9, no. 6, p. e98679, 2014.

زیرنویس‌ها

[1] Hyperlink-Induced Topic Search
[2] Mention
[3] Saraswathi
[4] Karnataka
[5] Betweenness
[6] Bangalore
[7] Kyent-Yon Yie
[8] Inflection Point
[9] Cumulative numbers of infected cases
[10] Item response model
[11] Correlation coefficients
[12] Mourad
[13] Girvan–Newman algorithm
[14] Modularity
[15] Power Law

[3] E. Bakshy, J. M. Hofman, W. A. Mason, and D. J. Watts, "Everyone's an influencer: quantifying influence on twitter," in *Proceedings of the fourth ACM international conference on Web search and data mining*, 2011, pp. 65–74.

[4] M. Zhang, C. Sun, and W. Liu, "Identifying influential users of micro-blogging services: A dynamic action-based network approach," in *Proceedings of the 15th Pacific Asia Conference on Information Systems (PACIS)*, 2011, pp. 1–12.

[5] M.-C. Yang, J.-T. Lee, S.-W. Lee, and H.-C. Rim, "Finding interesting posts in twitter based on retweet graph analysis," in *Proceedings of the 35th international ACM SIGIR conference on Research and development in information retrieval*, 2012, pp. 1073–1074.

[6] J. Weng, E.-P. Lim, J. Jiang, and Q. He, "Twitterrank: finding topic-sensitive influential twitterers," in *Proceedings of the third ACM international conference on Web search and data mining*, 2010, pp. 261–270.

[7] J. M. Kleinberg, "Authoritative sources in a hyperlinked environment," *J. ACM*, vol. 46, no. 5, pp. 604–632, 1999.

[8] L. Page, S. Brin, R. Motwani, and T. Winograd, "The PageRank citation ranking: Bringing order to the web.," Stanford InfoLab, 1999.

[9] S. Saraswathi, A. Mukhopadhyay, H. Shah, and T. S. Ranganath, "Social network analysis of COVID-19 transmission in Karnataka, India," *Epidemiol. Infect.*, vol. 148, 2020.

[10] K.-Y. Yie, T.-W. Chien, Y.-T. Yeh, W. Chou, and S.-B. Su, "Using Social Network Analysis to Identify Spatiotemporal Spread Patterns of COVID-19 around the World: Online Dashboard Development," *Int. J. Environ. Res. Public Health*, vol. 18, no. 5, p. 2461, 2021.

[11] M. Habibi, A. Priadana, and M. R. Ma'arif, "Hashtag Analysis of Indonesian COVID-19 Tweets Using Social Network Analysis," *IJCCS (Indonesian J. Comput. Cybern. Syst.*, vol. 15, no. 3, 2021.

[12] S. Alqurashi, A. Alashaikh, and E. Alanazi, "Identifying Information Superspreaders of COVID-19 from Arabic Tweets," *Preprints*, 2020.

[13] S. Alqurashi, A. Alhindi, and E. Alanazi, "Large arabic twitter dataset on covid-19," *arXiv Prepr. arXiv2004.04315*, 2020.

[14] A. Mourad, A. Srour, H. Harmanai, C. Jenainati, and M. Arafeh, "Critical impact of social networks infodemic on defeating coronavirus COVID-19 pandemic: Twitter-based study and research directions," *IEEE Trans. Netw. Serv. Manag.*, vol. 17, no. 4, pp. 2145–2155, 2020.

[15] M. Newman, "Networks: An Introduction - Oxford Scholarship," *Published to Oxford Scholarship Online: September 2010*, 2010. https://oxford.universitypressscholarship.com/view/10.1093/acprof:oso/9780199206650.001.0001/acprof-9780199206650 (accessed Feb. 10, 2022).

[16] S. Fortunato, "Community detection in graphs," *Phys. Rep.*, vol. 486, no. 3–5, pp. 75–174, 2010.

[17].M. E. J. Newman and M. Girvan, "Finding and evaluating community structure in networks," *Phys. Rev. E*, vol. 69, no. 2, p. 26113, 2004.

[18] X. Que, F. Checconi, F. Petrini, and J. A. Gunnels, "Scalable community detection with the louvain algorithm," in *2015 IEEE International Parallel and Distributed Processing Symposium*, 2015, pp. 28–37.

[19].M. Jacomy, T. Venturini, S. Heymann, and M. Bastian,