دوفصلنامه محاسبات و سامانه‌های توزیع شده
سال هفتم، شماره ۲، شماره پیاپی ۱۴، صفحه ۲۵ تا ۳۱، سال ۱۴۰۳

تحلیل شبکه هم‌نویسندگی بزرگ مقیاس با یادگیری ماشین

زانیار کریمی[۱]، صادق سلیمانی[۲]
[۱] دانشجوی کارشناسی ارشد، آزمایشگاه تحلیل شبکه‌های اجتماعی و زیستی، دانشکده مهندسی کامپیوتر، دانشگاه کردستان، سنندج
[۲] دانشیار، آزمایشگاه تحلیل شبکه‌های اجتماعی و زیستی، دانشکده مهندسی کامپیوتر، دانشگاه کردستان، سنندج

چکیده

تحلیل شبکه‌های هم‌نویسندگی یک زمینه میان‌رشته‌ای است که اغلب از ابزارهای مختلف تحلیلی در مراحل گوناگون استفاده می‌کند. با این حال، با افزایش پیچیدگی و گستردگی شبکه‌ها، روش‌های سنتی محدودیت‌های قابل توجهی را تجربه می‌کنند. علاوه بر این، ادغام داده‌های متنی و منابع غیرمعمول فرصتی برای استخراج بینش‌های عمیق‌تر و پیچیده‌تر از این شبکه‌ها فراهم می‌کند. این فرایند نیازمند بهره‌گیری از تکنیک‌های پیشرفته مانند استخراج متن و پردازش زبان طبیعی است. این مقاله علاوه بر رسیدگی به این چالش‌ها، مرور سیستماتیکی از مطالعات موجود درباره تحلیل شبکه‌های هم‌نویسندگی از دیدگاه‌های مختلف ارائه می‌دهد.

این پژوهش به بررسی ویژگی‌های ساختاری و تکامل پویای شبکه‌های هم‌نویسندگی با استخراج داده‌های مربوط به کاربردهای مختلف آن از پایگاه‌های معتبر مانند PubMed، Scopus و Semantic Scholar می‌پردازد. مجموعه داده‌ای بیش از ۲،۰۰۰ مقاله مرتبط گردآوری شد و با استفاده از روش‌های یادگیری شبکه و الگوریتم‌های یادگیری ماشین مورد تحلیل قرار گرفت. این تحلیل شامل شاخص‌های عمومی شبکه هم‌تألیفی از قبیل تراکم و ضریب خوشه‌بندی، یافتن فعال‌ترین زمینه‌های کاربردی، نویسندگان و مجلات، خوشه‌بندی، بررسی پویایی زمانی تغییرات شبکه و ... بود. نتایج نشان‌دهنده گسترش چشمگیر جوامع هم‌نویسندگی است، به ویژه در حوزه‌های میان‌رشته‌ای مانند مهندسی، پزشکی و علوم کامپیوتر. این پژوهش اطلاعات ارزشمندی را برای پژوهشگرانی فراهم می‌کند که به شناسایی حوزه‌های کمتر بررسی‌شده و رفع شکاف‌های دانش در تحلیل شبکه‌های هم‌تألیفی کمک می‌کند.

کلمات کلیدی: شبکه‌های هم‌تألیفی، تحلیل شبکه پیچیده، مرکزیت شبکه، خوشه‌بندی، تحلیل شبکه، جوامع پژوهشی

تاریخچه مقاله:
تاریخ ارسال: ۱۴۰۳/۰۶/۰۳
تاریخ اصلاحات: ۱۴۰۳/۱۰/۲۷
تاریخ پذیرش: ۱۴۰۳/۱۲/۱۹
تاریخ انتشار: ۱۴۰۳/۱۲/۲۵
ایمیل نویسنده مسئول: zaniar.karimi@uok.ac.ir

۱ – مقدمه

شبکه‌های هم‌تألیفی یکی از مهم‌ترین ابزارها برای درک ساختار و پویایی‌های تعاملات علمی در حوزه‌های مختلف تحقیقاتی هستند. این شبکه‌ها نه‌تنها ساختار همکاری میان نویسندگان را به تصویر می‌کشند، بلکه امکان تحلیل دقیق جوامع علمی، شناسایی پژوهشگران تأثیرگذار و ارزیابی روندهای تحول پژوهش را نیز فراهم می‌آورند [۱، ۲]. از دهه ۱۹۶۰ به بعد، تحلیل شبکه‌های هم‌تألیفی به‌عنوان یکی از ابزارهای اصلی در کتاب‌سنجی و علم‌سنجی معرفی شده است [۵]. مطالعه این شبکه‌ها می‌تواند به درک عمیق‌تری از ساختار علمی، تعاملات بین‌رشته‌ای و توسعه علمی جهانی منجر شود [۶، ۷].

این پژوهش، برخلاف مطالعات پیشین که تنها به تحلیل‌های شبکه‌های هم‌تألیفی برای یک کاربرد خاص بسنده کرده‌اند، از رویکردهای داده‌کاوی پیشرفته و شاخص‌های شبکه به ارائه تصویری جامع از تعاملات علمی در میان مقالات مرتبط با خودِ تحلیل شبکه هم‌تألیفی استفاده کرده است. روش‌های مورد استفاده شامل بررسی شاخص‌های مرکزیت درجه، مرکزیت بینابینی، ضریب خوشه‌بندی و تراکم شبکه است که امکان تحلیل عمیق‌تر روندهای بین‌رشته‌ای و شناسایی نویسندگان کلیدی را فراهم می‌کند.

تحقیقات پیشین بر اهمیت شبکه‌های هم‌تألیفی در شناسایی نویسندگان با نفوذ، تحلیل خوشه‌های پژوهشی، و تعیین شاخص‌های کلیدی شبکه تأکید کرده‌اند. برای مثال، پژوهش‌هایی نشان داده‌اند که شاخص‌هایی همچون مرکزیت درجه، مرکزیت بینابینی، و تراکم شبکه می‌توانند تصویری جامع از نحوه انتقال دانش و همکاری‌های علمی ارائه دهند [۳، ۴]. با این حال، هنوز شکاف‌های قابل‌توجهی در درک روندهای تحول این شبکه‌ها و تأثیر آن‌ها بر پویایی‌های علمی وجود دارد [۸].

مطالعه حاضر با هدف پر کردن این شکاف‌ها و ارائه یک تحلیل جامع از ساختار و پویایی‌های شبکه‌های هم‌تألیفی انجام شده است. در این پژوهش، داده‌های استخراج‌شده از پایگاه‌های معتبر علمی مانند مختلفی مورد استفاده قرار گرفته‌اند [۶]. این داده‌ها شامل بیش از ۳۱،۰۰۰ مقاله بوده که پس از پالایش، مجموعه‌ای با کیفیت بالا شامل ۲۳۴۵ رکورد منحصربه‌فرد فراهم شد. فرآیند پالایش داده‌ها با

دقت بالا و با استفاده از ابزارهایی نظیر اُپن‌ریفاین[1] انجام شده است که امکان استانداردسازی نام نویسندگان، حذف مقالات تکراری و انتخاب مقالات مرتبط با کلیدواژه‌های خاص را فراهم کرده است. روش تحلیل این پژوهش مبتنی بر شاخص‌های کلیدی شبکه‌های اجتماعی، از جمله مرکزیت درجه، مرکزیت بینابینی، ضریب خوشه‌بندی و تراکم شبکه است که برای شناسایی نویسندگان کلیدی، تحلیل جوامع پژوهشی، و بررسی روندهای بین‌رشته‌ای به کار گرفته شده‌اند [۷، ۹].

این پژوهش چندین جنبه نوآورانه دارد:

۱. **استفاده از مجموعه داده‌های بزرگ و پالایش‌شده:** داده‌های این پژوهش از پایگاه‌های مختلف جمع‌آوری شده و با استفاده از ابزارهای ویژه مانند اُپن‌ریفاین پالایش شدند.

۲. **تحلیل روندهای زمانی:** شاخص‌های کلیدی در بازه‌های زمانی مختلف محاسبه شده‌اند تا روندهای تحول شبکه‌های هم‌تألیفی آشکار شوند [۹].

۳. **شناسایی خوشه‌های علمی:** با استفاده از الگوریتم‌های رایج مانند لووین[2]، جوامع پژوهشی شناسایی و تعاملات آن‌ها تحلیل شدند [۶، ۱۰].

این پژوهش علاوه بر شناسایی روندهای همکاری، تمرکز ویژه‌ای بر بررسی همکاری‌های بین‌رشته‌ای دارد. انتظار می‌رود حوزه‌های با ذات بین‌رشته‌ای نقش پررنگی در توسعه علم و فناوری نشان بدهند. همچنین میزان رشد همکاری‌های علمی در دهه‌های اخیر نشان دهنده اهمیت بررسی تحلیل شبکه هم‌تألیفی به صورت عام و فارغ از کاربردهای خاص است.

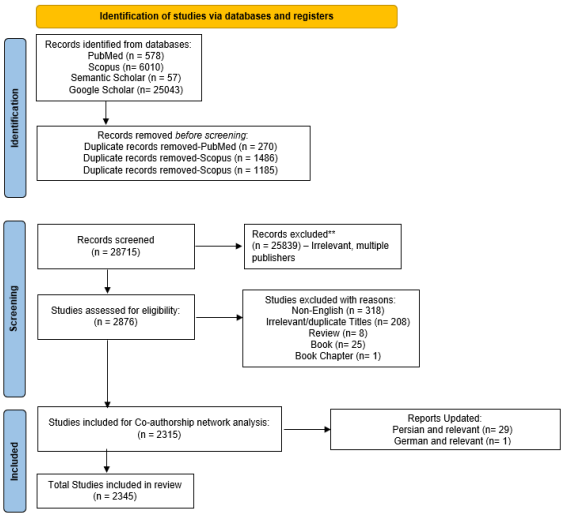
با توجه به این موارد، پژوهش حاضر با هدف ارائه تصویری جامع از شبکه‌های هم‌تألیفی و تحلیل ساختار و پویایی‌های آن‌ها طراحی شده است. نتایج این پژوهش می‌تواند به سیاست‌گذاران علمی و پژوهشگران کمک کند تا با درک بهتر ساختار تعاملات علمی، سیاست‌ها و راهبردهای مؤثرتری برای بهبود همکاری‌های بین‌رشته‌ای و توسعه علمی تدوین کنند.

۲ – روش شناسی

۲-۱: جمع‌آوری داده

برای این پژوهش، داده‌های کتاب‌سنجی از پایگاه‌های معتبر پاب‌مِد[3]، اسکوپوس[4]، سِمَنتیک اسکولار[5] و گوگل اسکولار[6] جمع‌آوری شدند [۶]. این داده‌ها شامل اطلاعات کتاب‌سنجی مرتبط با موضوع پژوهش و شبکه‌های هم‌تألیفی بود. از نرم‌افزار Publish or Perish برای استخراج داده‌ها استفاده شد که امکان جستجوی چندین پایگاه داده را به‌صورت هم‌زمان فراهم می‌کند. جستجو با استفاده از ۴ ترکیب متفاوت کلیدواژه "تحلیل شیکه هم‌نویسندگان" انجام شد [۷]. داده‌های انتخاب‌شده مربوط به مقالات پژوهشی و مرورهای سیستماتیک به زبان انگلیسی بودند که از پیدایش مبحث شبکه‌های هم‌تألیفی تا سال ۲۰۲۴ را پوشش می‌دادند.

فرآیند پالایش داده‌ها با رویکردی دقیق انجام شد تا اطمینان حاصل شود که مجموعه داده‌ها با معیارهای پژوهش کاملاً سازگار و مرتبط هستند؛ این فرآیند با استفاده از روش استاندارد پریسما[7] به‌صورت مرحله‌ای و شفاف پیاده‌سازی شد. این روش شامل چهار مرحله اصلی شناسایی، غربال‌گری، واجد شرایط بودن و انتخاب نهایی مقالات برای تحلیل بود. در مرحله نخست، ۳۱,۶۸۸ مقاله اولیه از پایگاه‌های مذکور شناسایی شدند. سپس، برای حذف مقالات تکراری که در پایگاه‌های مختلف مشترک نمایه شده بودند، پالایش داده انجام گرفت و ۲۵,۸۳۹ مقاله تکراری از مجموعه حذف شدند. پس از این مرحله، مقالات باقی‌مانده وارد فرآیند غربال‌گری شدند که در این مرحله ۴,۸۴۶ مقاله نامرتبط یا فاقد معیارهای پژوهشی حذف شدند (شکل ۱).



شکل-۱: نمودار پریسما

در نهایت، ۲۳۴۵ مقاله که تمامی معیارهای کیفی نظیر مرتبط بودن با موضوع پژوهش، کیفیت علمی و پوشش کامل کلیدواژه‌ها را برآورده می‌کردند، برای تحلیل نهایی انتخاب شدند. این مقالات نمایانگر داده‌هایی جامع و یکتا در زمینه شبکه‌های هم‌تألیفی بودند. فرآیند فوق به کمک نمودار پریسما نمایش داده شد که شفافیت و

[1] OpenRefine
[2] Louvain
[3] PubMed
[4] Scopus
[5] Semantic Scholar
[6] Google Scholar
[7] PRISMA

دقت فرآیند پژوهش را تضمین می‌کند و اطمینان حاصل می‌شود که داده‌های نهایی، معتبر و برای تحلیل شبکه مناسب هستند.

فرآیند پالایش داده‌ها با استفاده از ابزارهایی نظیر اُپن‌ریفاین انجام شد. این ابزار برای استانداردسازی اسامی نویسندگان، حذف خطاهای ناشی از تفاوت در نوشتار و شناسایی و حذف موارد تکراری بسیار مؤثر بود. داده‌های نهایی به گونه‌ای انتخاب شدند که از جامعیت و دقت کافی برای مدل‌سازی و تحلیل شبکه‌های هم‌تألیفی برخوردار باشند. علاوه بر این، نمودار پریسما فرآیند انتخاب داده‌ها را به‌صورت شفاف نمایش می‌دهد و تضمین می‌کند که داده‌های انتخاب‌شده برای تحلیل، معتبر و یکتا هستند. این نمودار به‌عنوان ابزاری ضروری در پژوهش‌های کتاب‌سنجی و علم‌سنجی شناخته می‌شود و نقش مهمی در تضمین قابلیت اطمینان یافته‌های پژوهشی ایفا می‌کند.

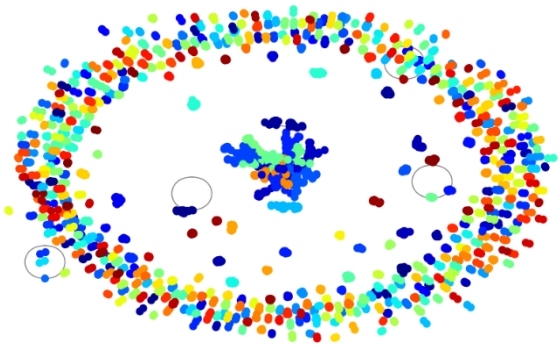
۲-۲: مرتب‌سازی و پالایش داده‌های دیتاست

پس از جمع‌آوری اولیه داده‌ها، فرآیند پالایش دقیقی به‌منظور حصول اطمینان از کیفیت و ارتباط داده‌ها با موضوع پژوهش انجام شد. ابتدا، بیش از ۲۵,۸۰۰ رکورد تکراری و نامرتبط که در پایگاه‌های مختلف جمع‌آوری شده بودند، حذف گردید. سپس، اسامی نویسندگان با استفاده از ابزار اُپن‌ریفاین استانداردسازی شدند تا از مشکلات مرتبط با تفاوت در نوشتار یا ابهامات در نام‌ها جلوگیری شود [۶]. در نهایت، مقالات مرتبط با استفاده از کلیدواژه‌های تخصصی شناسایی شدند و از میان رکوردهای باقیمانده، حدود ۲,۳۴۵ مقاله که معیارهای لازم را داشتند، برای تحلیل نهایی انتخاب گردیدند [۹]. این مقالات کیفیت بالایی داشتند و معیارهای پژوهش را به‌طور کامل پوشش می‌دادند.

۲-۳: ساخت و تحلیل شبکه

شبکه که نمود عملی حاصل از مفهوم نظری گراف می‌باشد در این بحث دارای گره‌ها و ارتباطات بین آن‌ها در دنیای هم‌تألیفی است [۶]. فرآیند ساخت شبکه به‌طور عمده شامل نمایه‌سازی روابط همکاری بین نویسندگان مختلف است. هر نویسنده به‌عنوان یک گره[8] در شبکه تعریف شده و روابط همکاری میان نویسندگان به صورت یال‌ها[9] بین این گره‌ها مشخص می‌شوند. این شبکه‌ها به‌طور خودکار روابط بین نویسندگی را شبیه‌سازی کرده و ساختار تعاملات علمی میان آن‌ها را تحلیل می‌کنند.

برای تحلیل این شبکه‌ها، از شاخص‌های استاندارد تحلیل شبکه با استفاده از کتابخانه NetworkX در زبان برنامه‌نویسی پایتون[10] استفاده شد، علاوه بر این، الگوریتم‌های یادگیری ماشین (مانند الگوریتم‌های خوشه‌بندی و تحلیل روند) برای تحلیل‌های عمیق‌تر مانند شناسایی الگوهای پیچیده همکاری استفاده گردید. این ابزارها امکان محاسبه شاخص‌های کلیدی شبکه و تجزیه‌وتحلیل ویژگی‌های ساختاری آن‌ها را فراهم می‌آورد [۷]. در این پژوهش، شاخص‌های مهمی نظیر مرکزیت درجه، مرکزیت بینابینی، ضریب خوشه‌بندی و تراکم شبکه برای تحلیل روابط میان نویسندگان و ساختار علمی شبکه محاسبه شدند. در بخش تحلیل خوشه‌ها، برای شناسایی و تحلیل جوامع علمی، از الگوریتم لووین استفاده شد. این الگوریتم به شناسایی گروه‌هایی از نویسندگان کمک کرد که در یک حوزه خاص به‌طور فعال همکاری داشتند. خوشه‌های شناسایی‌شده به‌طور خاص بر حوزه‌های میان‌رشته‌ای مانند مهندسی و علوم پزشکی تمرکز داشتند. برای هر خوشه، شاخص‌های کلیدی نظیر تعداد نویسندگان، میزان تعاملات، و ارتباطات بین خوشه‌ای محاسبه شد. به‌عنوان مثال، خوشه‌ای شامل نویسندگان در حوزه علوم پزشکی نشان‌دهنده ارتباط قوی با تحقیقات زیست‌پزشکی و مهندسی پزشکی بود. این نتایج بیانگر نقش مهم همکاری‌های میان‌رشته‌ای در رشد و گسترش دانش هستند [۸، ۹].



شکل-۲: شبکه ایجاد شده از دیتاست

مرکزیت درجه[11]: این شاخص تعداد ارتباطات مستقیم یک نویسنده با سایر نویسندگان را اندازه‌گیری می‌کند. نویسندگان با مرکزیت درجه بالا به‌عنوان گره‌های مرکزی شناخته می‌شوند که همکاری‌های بسیاری با سایر نویسندگان دارند و از این طریق می‌توانند انتقال سریع‌تری از دانش و اطلاعات در شبکه ایجاد کنند.

مرکزیت بینابینی[12]: این شاخص نشان‌دهنده میزان ارتباط یک نویسنده به‌عنوان پل ارتباطی میان دیگر نویسندگان است. نویسندگانی که دارای مرکزیت بینابینی بالا هستند، معمولاً نقش مهمی در برقراری ارتباطات میان جوامع علمی مختلف دارند و به‌عنوان تسهیل‌کنندگان انتقال دانش در شبکه عمل می‌کنند [۷].

[8] Node
[9] Edges
[10] Python
[11] Degree Centrality
[12] Betweenness Centrality

ضریب خوشه‌بندی[13]: این شاخص میزان تعاملات نزدیک و همکاری‌های متراکم میان نویسندگان یک گروه خاص را اندازه‌گیری می‌کند. نویسندگان با ضریب خوشه‌بندی بالا معمولاً در جوامع علمی خاصی فعالیت دارند که همکاری‌های فشرده‌تری با یکدیگر دارند.

تراکم شبکه[14]: این شاخص نسبت یال‌های موجود در شبکه به تعداد یال‌های ممکن را نشان می‌دهد. هرچه تراکم شبکه بالاتر باشد، نشان‌دهنده این است که نویسندگان بیشتری در یک شبکه هم‌پوشانی دارند و ارتباطات در میان آنها بیشتر است.

این تحلیل‌ها به پژوهشگران کمک می‌کند تا نویسندگان و جوامع کلیدی در شبکه‌های علمی را شناسایی کرده و روندهای همکاری علمی را در طول زمان بررسی کنند. در این پژوهش، علاوه بر تجزیه‌وتحلیل عمومی شبکه، به شناسایی جوامع علمی و روندهای همکاری بین‌رشته‌ای پرداخته شد که می‌تواند به درک بهتری از پویایی‌های علم و تکنولوژی در سطح جهانی منجر شود.

۳ - نتایج

۳-۱: شاخص‌های کلی شبکه

ارزیابی شاخص‌های مرکزیت، نقش محوری نویسندگان در ساختار شبکه همکاری‌های علمی را آشکار می‌سازد. نویسندگانی که از مرکزیت درجه بالایی برخوردارند، غالباً در کانون تعاملات علمی قرار داشته و به‌عنوان گره‌های تأثیرگذار در شبکه عمل می‌کنند. تحلیل روند زمانی شاخص‌های شبکه (جدول ۱)، رشد قابل‌ملاحظه شبکه‌های هم‌تألیفی را در بازه ۲۰۰۵ تا ۲۰۲۴ نشان می‌دهد. در حالی که تعداد گره‌ها (نویسندگان) و یال‌ها (همکاری‌ها) به‌طور پیوسته افزایش یافته است، کاهش نسبی تراکم شبکه، همزمان با رشد کلی، می‌تواند نشان‌دهنده گسترش شبکه به حوزه‌های تخصصی‌تر و افزایش پیچیدگی ساختاری آن باشد؛ به طوری که همکاری‌ها بیشتر در درون خوشه‌های تخصصی متمرکز می‌شوند. این پدیده، با بالا ماندن ضریب خوشه‌بندی در طول دوره که بیانگر تمایل پایدار پژوهشگران به تشکیل گروه‌های همکاری منسجم و متمرکز است، همخوانی دارد و مبادله دانش را تسهیل می‌کند.

نتایج تحلیل شاخص‌های شبکه نشان داد که رشد همکاری‌های علمی در حوزه‌های بین‌رشته‌ای مانند مهندسی و علوم پزشکی ناشی از ذات میان‌رشته‌ای این حوزه‌ها و نیاز به تعاملات علمی گسترده‌تر است. برای مثال، رشد خوشه‌های مهندسی در دهه اخیر نشان‌دهنده تقویت همکاری‌های پژوهشی در زمینه‌های مشترک با علوم کامپیوتر و فناوری‌های نوظهور بوده است. در مقابل، حوزه پزشکی به دلیل نیاز به تحقیقات چندجانبه، تعاملات گسترده‌ای با حوزه‌های زیست‌شناسی و مهندسی پزشکی داشته است. این روندها نشان می‌دهند که همکاری‌های علمی نه تنها به توسعه دانش کمک کرده، بلکه مرزهای رشته‌ها را نیز به چالش کشیده است.

جدول-۱: شاخص‌های کلی شبکه هم‌تالیفی

بازه زمانی	تعداد گره‌ها	تعداد یال‌ها	تراکم شبکه	ضریب خوشه‌بندی
۲۰۰۵-۲۰۰۹	۴۵	۴۳	۰.۰۴۳	۰.۷۱۱
۲۰۱۰-۲۰۱۴	۲۳۲	۲۷۶	۰.۱۰	۰.۷۸۴
۲۰۱۵-۲۰۱۹	۶۱۷	۸۰۳	۰.۰۳	۰.۷۳۶
۲۰۲۰-۲۰۲۴	۱۲۰۴	۱۸۱۲	۰.۰۲	۰.۷۵۴

۳-۲ : برترین نویسندگان

تحلیل شاخص مرکزیت بینابینی، نویسندگانی را شناسایی نمود که نقش کلیدی در اتصال خوشه‌های علمی مجزا و تسهیل جریان دانش در شبکه ایفا می‌کنند. به‌عنوان نمونه، نویسندگانی مانند فریده عصاره[15] و ژوفِنگ‌وانگ[16] با کسب بالاترین مقادیر این شاخص، به‌عنوان پل‌های ارتباطی حیاتی میان جوامع پژوهشی مختلف عمل می‌کنند. این نقش واسطه‌ای، اهمیت بسزایی در تسریع انتشار دانش میان‌رشته‌ای، ایجاد فرصت‌های همکاری نوین و تقویت انسجام کلی شبکه علمی از طریق اتصال حوزه‌های دانشی مجزا دارد.

۳-۳: تحلیل خوشه‌ها

تحلیل خوشه‌ها در این پژوهش با استفاده از الگوریتم لووین[17]، یکی از روش‌های رایج برای شناسایی ساختار جوامع در شبکه‌ها، انجام شد. این الگوریتم با بهینه‌سازی معیار ماژولاریتی،گروه‌هایی از نویسندگان (خوشه‌ها) را شناسایی می‌کند که همکاری‌های متقابل و پیوسته‌ای با یکدیگر دارند و معمولاً نمایانگر یک گروه پژوهشی فعال در زمینه‌ای خاص هستند. در این پژوهش، خوشه‌های شناسایی‌شده بر اساس میزان همکاری نویسندگان طی دو دهه گذشته، تمرکز پژوهش در حوزه‌ها و موضوعات علمی خاص را نشان دادند.

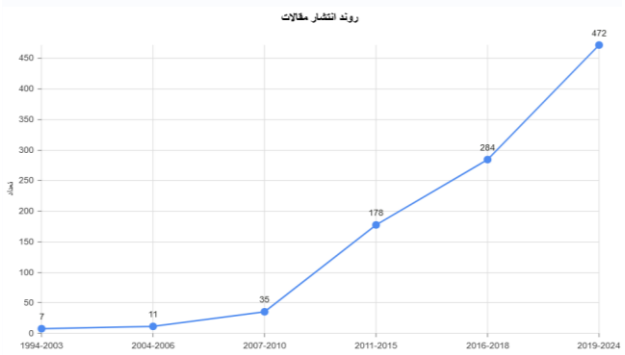
نتایج تحلیل نشان داد که تعاملات بین‌رشته‌ای در سال‌های اخیر به‌طور قابل‌توجهی افزایش یافته است. جوامع علمی که پیش‌تر به‌طور عمده در مرزهای رشته‌های خاص محدود بودند، اکنون تعاملات بیشتری در زمینه‌های مختلف دارند. این تغییرات می‌تواند به گسترش همکاری‌های بین‌رشته‌ای و شناسایی مسیرهای جدید تحقیقاتی منجر شود. با شناسایی جوامع علمی، پژوهشگران و

[13] Clustering Coefficient
[14] Network Density
[15] Farideh Osareh
[16] Xuefeng Wang
[17] Louvain

سیاست‌گذاران علمی می‌توانند به درک بهتری از روندهای علمی و پویایی‌های همکاری در علوم مختلف برسند.

۳-۴: افزایش انتشار مقالات هم‌تالیفی

همان‌گونه که در نمودار شکل ۳ مشهود است، روند انتشار مقالات مرتبط با تحلیل شبکه هم‌تألیفی طی دو دهه گذشته، روندی قویاً صعودی داشته است. در بازه زمانی ۲۰۰۵ تا ۲۰۰۹، تعداد مقالات منتشرشده حدود ۴۵ مقاله بود که نشان‌دهنده آغاز روند توجه به شبکه‌های هم‌تألیفی و استفاده از آن‌ها در تحلیل‌های علمی بود. این تعداد مقالات در بازه ۲۰۱۰ تا ۲۰۱۴ به ۲۳۲ مقاله افزایش یافت که نمایانگر شروع رشد اولیه و گسترش علاقه به تحلیل‌های شبکه‌ای در علم بود. از سال ۲۰۱۵ تا ۲۰۱۹، تعداد مقالات به ۶۷۰ مقاله رسید و این نشان‌دهنده توجه روزافزون به این حوزه و گسترش همکاری‌های علمی در موضوعات مختلف است. در نهایت، در بازه ۲۰۲۰ تا ۲۰۲۴ تعداد مقالات به ۱,۲۰۴ مقاله رسید که این رشد شتاب‌دار، بیانگر گسترش قابل‌توجه شبکه‌های هم‌تألیفی در سطح جهانی است. این روند افزایشی واضحاً نشان‌دهنده توسعه گسترده‌تر همکاری‌های علمی در سراسر جهان است و نشان می‌دهد که شبکه‌های هم‌تألیفی به ابزاری اساسی در تحلیل‌های علمی و پیشبرد تحقیقات بین‌رشته‌ای تبدیل شده‌اند. افزایش مقالات در سال‌های اخیر، علاوه بر اینکه به رشد و پیشرفت علمی کمک کرده است، مسیر جدیدی را برای شناخت جوامع پژوهشی و روندهای نوین همکاری‌های علمی فراهم کرده است.



شکل-۳: روند انتشار مقالات

۳-۵: حوزه‌های برتر پژوهشی

تحلیل محتوایی داده‌ها با استفاده از روش‌های متن‌کاوی نشان داد که برخی حوزه‌های پژوهشی، توجه بیشتری را در شبکه‌های هم‌تألیفی به خود جلب کرده‌اند. با استفاده از روش‌های متن‌کاوی و تحلیل روندهای زمانی، ده حوزه پژوهشی برتر که بیشترین تعداد مقالات هم‌تألیفی را به خود اختصاص داده‌اند، شناسایی شدند (جدول ۲). این حوزه‌ها به‌طور عمده شامل مهندسی، پزشکی، علوم کامپیوتر، و اقتصاد بودند که هرکدام به نوبه خود رشد قابل‌توجهی در همکاری‌های علمی نشان دادند. به‌ویژه، مهندسی با ۱۳۴ مقاله، پزشکی با ۱۱۹ مقاله، و علوم کامپیوتر با ۱۱۵ مقاله در صدر قرار داشتند. این حوزه‌ها نمایانگر تمرکز بالای پژوهش‌ها و همکاری‌های علمی در این زمینه‌ها است.

تحلیل داده‌ها با استفاده از متن‌کاوی نشان داد که حوزه‌های زیر بیشترین تعداد مقالات هم‌تألیفی را به خود اختصاص داده‌اند:

جدول-۲: برترین حوزه‌های پژوهشی شناسایی شده به ترتیب فراوانی مقالات

حوزه پژوهشی	تعداد مقالات
مهندسی	۱۳۴
پزشکی	۱۱۹
علوم کامپیوتر	۱۱۵
اقتصاد	۱۱۳
فیزیک	۱۰۶
زیست‌شناسی	۹۷
علوم اجتماعی	۹۶
ریاضیات	۸۱
شیمی	۷۵
روان‌شناسی	۷۰

با بررسی روندهای زمانی مشخص شد که در دهه اخیر، برخی از این حوزه‌ها مانند علوم اجتماعی و زیست‌شناسی رشد چشم‌گیری در تعداد مقالات هم‌تألیفی داشته‌اند. این تغییرات نشان‌دهنده تحولاتی در فضای تحقیقاتی و ظهور زمینه‌های جدید پژوهشی است که پژوهشگران را به تعاملات بیشتر و همکاری‌های بین‌رشته‌ای ترغیب می‌کند. نتایج این تحلیل می‌تواند به سیاست‌گذاران علمی کمک کند تا بر روی حوزه‌های کلیدی و پرتانسیل تمرکز کنند و از روندهای پیشرو در زمینه‌های پژوهشی حمایت نمایند.

۳-۶: نشریات برتر

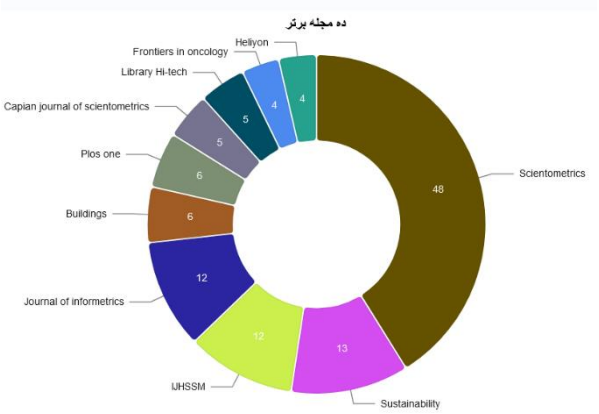
بررسی ۱۰ مجله برتر در حوزه شبکه‌های هم‌تألیفی نشان داد که برخی نشریات نقش کلیدی در انتشار مقالات مرتبط با این حوزه ایفا کرده‌اند. مجله علم‌سنجی[18] بیشترین سهم را در این زمینه داشت و با انتشار بیش از ۴۱ درصد از مقالات، به‌عنوان مرجعی پیشرو در علم‌سنجی شناخته می‌شود. این مجله به‌ویژه بر تحلیل‌های کمی و ساختاری شبکه‌های هم‌تألیفی تمرکز دارد و نقشی اساسی در شناسایی روندهای علمی ایفا می‌کند. مجلات پایداری[19] و

[18] Scientometrics

[19] Sustainability

علم‌سنجی[20] به ترتیب با ۱۱/۳۰٪ و ۱۰/۴۳٪ از مقالات هم‌تألیفی، در جایگاه‌های دوم و سوم قرار دارند. این نشریات بر مطالعات میان‌رشته‌ای و روش‌های نوین تحلیل همکاری‌های علمی تمرکز دارند و بخش عمده‌ای از مقالات خود را به تحلیل‌های پیچیده شبکه‌های اجتماعی علمی اختصاص داده‌اند.

نشریات پلوس وان[21] و کتابخانه های-تک[22] نیز با تمرکز بر کاربردهای عملی و پروژه‌های علمی در زمینه‌های مختلف، نقش مهمی در گسترش تحلیل‌های هم‌تألیفی در علوم اجتماعی، فناوری اطلاعات و پزشکی ایفا می‌کنند. از دیگر نشریات برجسته می‌توان به پیشرفت‌های نوین در سرطان‌شناسی[23]، هِلیون[24]، و مجله علم‌سنجی کاسپین[25] اشاره کرد که هرکدام در زمینه‌های خاص خود به انتشار مقالات هم‌تألیفی پرداخته‌اند و روندهای علمی جدید را معرفی کرده‌اند.



شکل-۴: ده مجله برتر نشر کننده آثار

۴- نتیجه‌گیری

این پژوهش، تصویری جامع از ساختار و پویایی شبکه‌های هم‌تألیفی علمی ارائه می‌دهد و نشان می‌دهد که همکاری‌های علمی، به ویژه در حوزه‌های میان‌رشته‌ای، در دو دهه اخیر به‌طور قابل‌توجهی افزایش یافته است. نویسندگان کلیدی در اتصال میان جوامع علمی نقش حیاتی ایفا کرده و تحلیل خوشه‌ها نمایانگر تعاملات بین‌رشته‌ای در حال گسترش است. یافته‌ها تأکید می‌کنند که درک این شبکه‌ها برای سیاست‌گذاری علمی و شناسایی روندهای نوظهور ضروری است. پیشنهاد می‌شود که در پژوهش‌های آینده، تحلیل شبکه‌های هم‌تألیفی با سایر شاخص‌های کتاب‌سنجی (مانند تحلیل استنادی یا تحلیل محتوایی عمیق‌تر) و روش‌های پیشرفته‌تر یادگیری ماشین ترکیب شوند تا درک عمیق‌تری از پویایی‌های پیچیده علمی و الگوهای همکاری حاصل شود.

۶- منابع

[1] *X. Sun, J. Kaur, S. Milojević, A. Flammini, and F. Menczer, "Social dynamics of science," Scientific Reports, vol. 3, no. 1, Jan. 2013, doi: 10.1038/srep01069.*

[2] *W. Glänzel and A. Schubert, "Analysing Scientific Networks Through Co-Authorship," in Kluwer Academic Publishers eBooks, 2006, pp. 257–276. doi: 10.1007/1-4020-2755-9_12.*

[3] *L. Leydesdorff, "Betweenness centrality as an indicator of the interdisciplinarity of scientific journals," Journal of the American Society for Information Science and Technology, vol. 58, no. 9, pp. 1303–1319, May 2007, doi: 10.1002/asi.20614.*

[4] *C. Chen, F. Ibekwe-SanJuan, and J. Hou, "The structure and dynamics of cocitation clusters: A multiple-perspective cocitation analysis," Journal of the American Society for Information Science and Technology, vol. 61, no. 7, pp. 1386–1409, Mar. 2010, doi: 10.1002/asi.21309.*

[5] *F. Momeni and P. Mayr, "Evaluating Co-authorship Networks in Author Name Disambiguation for Common Names," in Lecture notes in computer science, 2016, pp. 386–391. doi: 10.1007/978-3-319-43997-6_31.*

[6] *Y. Ding, "Scientific collaboration and endorsement: Network analysis of coauthorship and citation networks," Journal of Informetrics, vol. 5, no. 1, pp. 187–203, Nov. 2010, doi: 10.1016/j.joi.2010.10.008.*

[7] *H. Hou, H. Kretschmer, and Z. Liu, "The structure of scientific collaboration networks in Scientometrics," Scientometrics, vol. 75, no. 2, pp. 189–202, Dec. 2007, doi: 10.1007/s11192-007-1771-3.*

[8] *L. J. Ducharme, K. Fujimoto, J. Kuo, J. Stewart, B. Taylor, and J. Schneider, "Collaboration and growth in a large research cooperative: A network analytic approach," Evaluation and Program Planning, vol. 102, p. 102375, Sep. 2023, doi: 10.1016/j.evalprogplan.2023.102375.*

[9] *K. W. Boyack and R. Klavans, "Co-citation analysis, bibliographic coupling, and direct citation: Which citation approach represents the research front most accurately," Journal of the American Society for Information Science and Technology, vol. 61, no. 12, pp. 2389–2404, Oct. 2010, doi: 10.1002/asi.21419.*

[10] *M. E. J. Newman, "The Structure and Function of Complex Networks," SIAM Review, vol. 45, no. 2,*

[20] Journal of Informetrics
[21] PLOS One
[22] Library Hi-Tech
[23] Frontiers in Oncology
[24] Heliyon
[25] Caspian Journal of Scientometrics

Large Scale Co-Authorship Network Analysis Using Machine Learning

Z. Karimi[1], S. Sulaimany[1]
[1]Social and Biological Network Analysis Laboratory (SBNA), Department of Computer Engineering, University of Kurdistan, Sananadaj, Iran

Abstract

Co-authorship network analysis is a multidisciplinary endeavor, often utilizing various analytical tools across multiple phases. However, as network complexity and scale increase, conventional methodologies encounter significant limitations. Moreover, the integration of textual and non-standard data sources provides an opportunity to extract deeper, more nuanced insights from these networks. This process necessitates the application of advanced techniques such as text mining and natural language processing. This paper aims to address these methodological challenges by systematically reviewing existing literature on co-authorship network analysis from diverse perspectives.

The study investigates the structural properties and dynamic evolution of co-authorship networks, drawing on data sourced from renowned databases, including PubMed, Scopus, and Semantic Scholar. A dataset encompassing over 2,000 relevant publications was compiled and analyzed using machine learning algorithms. This analysis encompassed general metrics of the co-authorship network, such as density and clustering coefficient, identification of the most active application fields, authors, and journals, clustering, temporal dynamics of network changes, and more. Results indicate substantial expansion in co-authorship communities, particularly within interdisciplinary fields like engineering, medicine, and computer science. This research provides valuable insights for academics and policymakers by identifying areas requiring further exploration and addressing existing gaps in the study of co-authorship networks.

pp. 167–256, Jan. 2003, doi: 10.1137/s003614450342480.

روش ارجاع: ز.کریمی، ص.سلیمانی، تحلیل شبکه هم‌نویسندگی بزرگ مقیاس با یادگیری ماشین، دوفصلنامه محاسبات و سامانه‌های توزیع شده، سال هفتم، شماره۲، شماره پیاپی ۱۴، صفحه ۲۵ تا ۳۱، سال ۱۴۰۳.

How to cite: Z.Karimi, S.Sulaimany, Large Scale Co-Authorship Network Analysis Using Machine Learning, Journal of Distributed Computing and Systems (JDCS), Vol 7, Issue 2, Pages 25 – 31, 2025.

زانیار کریمی دانشجوی کارشناسی ارشد هوش مصنوعی و رباتیکز در دانشگاه کردستان می باشد. زمینه های موردعلاقه ایشان مدل‌های زبانی بزرگ، گراف دانش، یادگیری عمیق، شبکه‌های پیچیده زیستی و اقتصادی و پردازش احساساست می باشد، و عضو آزمایشگاه تحلیل شبکه‌های زیستی، پزشکی و اجتماعی (SBNA) می‌باشد و همچنین تعدادی مقاله در دست نگارش و ارسال به چاپ در مجلات و کنفرانس در این موضوعات می‌باشد.

نشانه رایانامه ایشان عبارتند از:

Zaniar.karimi@uok.ac.ir

صادق سلیمانی عضو هیات علمی گروه کامپیوتر دانشگاه کردستان و مسئول آزمایشگاه تحلیل شبکه‌های اجتماعی و زیستی است. ایشان مدرک دکتری خود را در سال ۱۳۹۵ در رشته بیوانفورماتیک از دانشگاه تهران دریافت کرد و تاکنون در مباحث نظری شبکه‌های پیچیده و کاربردهای مختلف آن در حیطه‌های مختلف اعم از شبکه‌های اجتماعی، زیستی، پزشکی، اقتصادی و ... مقالات مختلفی منتشر کرده است. به طور خاص ایشان در حیطه پیش‌بینی پیوند و گراف‌های پدیداری، تمرکز ویژه داشته است.

نشانی رایانامه ایشان عبارتند از:

S.Sulaimany@UoK.ac.ir