

بررسی تأثیر رتبه نویسندگان در شبکه‌های کتاب‌شناختی بر عملکرد مدل سند-محور بازیابی تخصص

* هاشم عطاپور: استادیار علم اطلاعات و دانش‌شناسی، دانشگاه تبریز، تبریز، ایران (نویسنده مسئول) hashematapour@tabrizu.ac.ir
فاطمه فهیم‌نیا: دانشیار علم اطلاعات و دانش‌شناسی، دانشگاه تهران، تهران، ایران

چکیده

دریافت: ۱۳۹۶/۵/۱۸
پذیرش: ۱۳۹۷/۱۰/۱

زمینه و هدف: این پژوهش به بررسی تأثیر رتبه نویسندگان در شبکه‌های کتاب‌شناختی بر مدل سند-محور بازیابی تخصص می‌پردازد. هدف آن پی بردن به این نکته است که کدام یک از شبکه‌های کتاب‌شناختی نویسندگان می‌تواند عملکرد مدل سند-محور بازیابی تخصص را بهبود ببخشد.

روش پژوهش: روش پژوهش تجربی مقدماتی است. برای انجام پژوهش، مجموعه آزمونی متشکل از ۵۵ پرس‌وجو و ۹۶۳۷۵ سند ایجاد شد: پرس‌وجوهایی که توسط دانشجویان و فارغ‌التحصیلان دکتری علم اطلاعات و دانش‌شناسی ایران ساخته شد و اسنادی که از مقالات نمایه‌شده علم اطلاعات و دانش‌شناسی در پایگاه وب آو ساینس تشکیل شده بود. پرس‌وجوها به پیکره آزمون عرضه، و مدل بازیابی DLH13 برای بازیابی اسناد به کار گرفته شد. ۱۰۰ سند اول بازیابی شده برای هر پرس‌وجو انتخاب شد. سپس اسامی افراد موجود در آنها استخراج و پردازش، و بر اساس ۵ شاخص مورد استفاده در تحلیل شبکه‌های اجتماعی رتبه‌بندی گردید. ۱۰ نتیجه اول هر روش انتخاب و در انبوهه نویسندگان پرس‌وجوی مربوطه قرار گرفت. افراد موجود در سیاهه مورد قضاوت ربط قرار گرفتند تا مقایسه عملکرد رتبه‌بندی‌های افراد از حیث یافتن خبرگان میسر شود.

یافته‌ها: یافته‌ها نشان داد عملکرد مدل‌های مبتنی بر رتبه‌بندی نویسندگان در شبکه‌های استنادی تفاوت معناداری با مدل سند-محور بازیابی تخصص ندارند، اما عملکرد مدل مبتنی بر رتبه‌بندی نویسندگان در شبکه هم‌تألیفی ضعیف‌تر از مدل سند-محور بازیابی تخصص بوده و آن را کاهش می‌دهد.

نتیجه‌گیری: در مقایسه با شبکه‌های پدیدآوری، موقعیت افراد در شبکه‌های استنادی شاهد بهتری برای تخصص افراد در حوزه‌های موضوعی مختلف به شمار می‌رود.

کلیدواژه‌ها: ارزیابی، بازیابی اطلاعات، خبره‌یابی، شبکه استنادی، شبکه هم‌استنادی، شبکه هم‌تألیفی، مجموعه‌آزمون‌ها

تعارض منافع: گزارش نشده است.
منع حمایت‌کننده: حامی مالی نداشته است.

شبیه استناد به این مقاله

APA: Atapour, H., Fahimnia, F., (2018). Investigating the Impact of Authors' Rank in Bibliographic Networks on Expertise Retrieval. *Human Information Interaction*. 5(1);71-82. (Persian)

Vancouver: Atapour H, Fahimnia F. Investigating the Impact of Authors' Rank in Bibliographic Networks on Expertise Retrieval. *Human Information Interaction*. 2018;5(1):71-82. (Persian)



انتشار مجله تعامل انسان و اطلاعات با حمایت مالی دانشگاه فوارزمی انجام می‌شود.

انتشار این مقاله به صورت دسترسی آزاد مطابق با [CC BY-NC-SA 1.0](https://creativecommons.org/licenses/by-nc-sa/4.0/) صورت گرفته است.

Investigating the Impact of Authors' Rank in Bibliographic Networks on Expertise Retrieval

***Hashem Atapour:** Assistant Professor of Knowledge and Information Studies, Tabriz University, Tabriz, Iran
(Corresponding author) hashematapour@tabrizu.ac.ir

Fatemeh Fahimnia: Associate Professor of Knowledge and Information Studies, University of Tehran, Tehran, Iran

Received: 08/09/2017

Accepted: 12/22/2018

Abstract

Background and Aim: This research investigates the impact of authors' rank in Bibliographic networks on document-centered model of Expertise Retrieval. Its purpose is to find out what kind of authors' ranking in bibliographic networks can improve the performance of document-centered model.

Methods: Current research is an experimental one. To operationalize research goals, a new test collection was developed which includes 55 queries and 96375 documents. The queries were made by Iran Knowledge and Information Science PhD students, and the documents were papers indexed in the Web of Science database under Library Science and Information Science category. The queries were submitted to the database consisting of test collection documents, and then DLH13, a known IR model, were used to retrieve documents from database. The first 100 documents retrieved by DLH13 model for each query were chosen for second stage. All people names occurred in the retrieved documents were extracted, processed, and ranked in 5 different ways based on micro metrics of Social Network Analysis. The top 10 results of every method accumulated in a pool of authors. After relevance judgment on authors' expertise, the expert finding performance of every ranking method was measured.

Results: Results showed that performance of authors' ranking in citation networks hadn't significant difference with document-centered model, whereas authors' ranking in co-authorship networks was weaker than document-centered model, and impact it negatively.

Conclusion: Compared with author-based networks, citation-based networks are better evidence for individual's expertise in different subject areas.

Keywords: Citation network, Co-citation network, Co-authorship network, Evaluation, Expert finding, Test collection

Conflicts of Interest: None

Funding: None

How to cite this article

APA: Atapour, H., Fahimnia, F., (2018). Investigating the Impact of Authors' Rank in Bibliographic Networks on Expertise Retrieval. Human Information Interaction. 5(1);71-82. (Persian)

Vancouver: Atapour H, Fahimnia F. Investigating the Impact of Authors' Rank in Bibliographic Networks on Expertise Retrieval. Human Information Interaction. 2018;5(1):71-82. (Persian)



مقدمه

قرار می‌گیرد. چنانچه دامنه اسناد مورد بررسی به انتشارات دانشگاهی محدود گردد، افراد یافته‌شده توسط مدل با عنوان نویسندگان خبره قلمداد می‌شوند که در پاره‌ای از مطالعات حوزه بازیابی تخصص نظیر بوگرس، کوکس، و وان دن بوش^۵ (۲۰۰۸) و افران^۶ (۲۰۰۹) شاهد آن هستیم. در پژوهش حاضر نیز، به دلیل پیاده‌سازی مدل بازیابی تخصص بر روی انتشارات دانشگاهی افراد بازیابی‌شده به عنوان نویسندگان خبره در نظر گرفته می‌شوند.

در سویی دیگر و در متون علم‌سنجی، برای شناسایی و رتبه‌بندی نویسندگان مقالات دانشگاهی روش‌های متفاوتی دنبال می‌شود که یکی از آنها استفاده از تحلیل شبکه‌های ارتباطی نویسندگان بر اساس شاخص‌های مختلف مرکزیت می‌باشد. در میان اسامی افراد واقع شده در انتشارات دانشگاهی، انواع مختلفی از شبکه‌های ارتباطی قابل شکل‌گیری هست که از جمله آنها می‌توان به شبکه‌های استنادی^۷، هم‌استنادی^۸، و هم‌تألیفی^۹ (برنر، ۲۰۱۰) اشاره نمود. در شبکه‌های مذکور، شاخص‌های مختلفی مانند برای تحلیل شبکه به کار گرفته می‌شود که در دو دسته شاخص‌های کلان و شاخص‌های خرد جای می‌گیرند. شاخص‌های کلان تحلیل شبکه‌های اجتماعی به بررسی پیکربندی و ویژگی‌های کلی شبکه‌ها می‌پردازند و شامل شاخص‌هایی مثل چگالی^{۱۰}، ضریب خوشه‌بندی^{۱۱}، مؤلفه‌های تشکیل‌دهنده شبکه^{۱۲} و میانگین فاصله در شبکه^{۱۳} می‌شود. در مقابل، عملکرد هر یک از گره‌های موجود در شبکه با استفاده از شاخص‌های خرد مورد بررسی قرار می‌گیرد. مرکزیت از شاخص‌های خرد در تحلیل شبکه‌های اجتماعی بوده و شامل انواع مرکزیت درجه‌ای^{۱۴}، مرکزیت بینابینی^{۱۵}، مرکزیت نزدیکی^{۱۶}، مرکزیت بردار ویژه^{۱۷}، و پیچ‌رنک^{۱۸} (نوعی خاص از مرکزیت بردار ویژه) است (عرفان‌منش و بصیریان جهرمی، ۱۳۹۲). از میان شاخص‌های خرد تحلیل شبکه‌های اجتماعی، مرکزیت درجه‌ای به دلیل تأکید بر کمیّت ارتباطات، و پیچ‌رنک به دلیل تأکید همزمان بر کمیّت و کیفیت ارتباطات در پژوهش حاضر مورد استفاده قرار می‌گیرند که در ادامه شرح مختصری از آنها ارائه می‌شود.

پژوهش‌هایی که در دهه ۱۹۶۰ میلادی در علم اطلاعات و دانش‌شناسی در حوزه رفتار اطلاع‌یابی کارکنان دانشی مانند پژوهشگران و مهندسان صورت گرفت، نشان داد که مراجعه به خبرگان انسانی یکی از روش‌های کسب اطلاعات کارکنان دانشی بوده است (بالوگ، فانگ، دورایک، سردیوکوف، و سی^۱، ۲۰۱۲). با این وجود حدود سه دهه طول کشید تا نیاز کاربران به یافتن خبرگان انسانی مورد توجه نظام‌های اطلاعاتی قرار گیرد. در دهه ۱۹۹۰ میلادی توجه سازمان‌ها به مدیریت دانش، باعث شد تا آنها بخشی را به نظام‌های اطلاعاتی خود اضافه کنند که از بازیابی تخصص پشتیبانی نماید: نظام‌هایی که افراد صاحب‌دانش در حوزه موضوعی خاص را شناسایی و ارائه کند یا حوزه‌های تخصصی هر فرد را سیاهه نماید. نظام‌های اولیه بازیابی تخصص دستی بودند که در ایجاد و نگهداری آنها مشکلاتی وجود داشت. بنابراین، از همان ابتدا تلاش‌ها بر روی ایجاد نظام‌هایی معطوف گشت که بتوانند خبرگان را به طور خودکار پروفایل کرده و در مواقع نیاز بازیابی نمایند. بدین منظور، پژوهشگران با دوری جستن از چالش‌های شناختی مفهوم تخصص بر روی اسناد متنی به عنوان شاهد تخصص افراد تمرکز نموده و چنین تعریفی را ارائه کردند: هر چقدر فرد با اسناد یک حوزه موضوعی ارتباط مستحکم داشته باشد، بیشتر می‌توان بر روی او به عنوان خبره آن موضوع تأکید کرد. بدین ترتیب، مسئله بازیابی تخصص به انباشت اسناد متنی و کشف روابط افراد با آن اسناد تقلیل یافت. برای کشف روابط بین اسناد و افراد، مدل‌های متعددی ارائه شده‌اند که مدل سند-محور^۲ بالوگ^۳ یکی از موفق‌ترین مدل‌ها بوده و مبنای بسیاری از مدل‌های بعدی قرار گرفته است. در مدل سند-محور برای شناسایی و رتبه‌بندی افراد خبره در حوزه موضوعی یک پرس‌وجو، ابتدا با استفاده از یک مدل بازیابی اطلاعات، اسنادی که گمان می‌رود با پرس‌وجو مرتبط هستند بازیابی می‌شوند. سپس، افراد مرتبط با اسناد بازیابی شده به عنوان خبرگان احتمالی حوزه موضوعی پرس‌وجو در نظر گرفته می‌شوند. امتیاز هر فرد در ارتباط با پرس‌وجو، از طریق جمع امتیاز اسناد مرتبط با وی یا جمع تعداد اسناد مرتبط با وی یا جمع معکوس رتبه اسناد مرتبط با وی به دست می‌آید (بالوگ، آزوپاردی و دو رایک^۴، ۲۰۰۶). افراد بر اساس امتیازات حاصل شده رتبه‌بندی می‌شوند و پس از قضاوت ربط، عملکرد مدل بازیابی تخصص بر اساس سنجه‌های ارزیابی نظام‌های بازیابی اطلاعات مورد سنجش

⁵ Bogers, Kox, Van Den Bosch

⁶ Efron

⁷ Citation network

⁸ Co-citation network

⁹ Co-authorship network

¹⁰ Density

¹¹ Clustering coefficient

¹² Components

¹³ Mean distance

¹⁴ degree Centrality

¹⁵ Betweenness Centrality

¹⁶ Closeness Centrality

¹⁷ eigenvector centrality

¹⁸ Page Rank

¹ Balog, Fang, de Rijke, Serdyukov, & Si

² Document-Centered Model

³ Balog

⁴ Balog, Azzopardi, & de Rijk

مرکزیت درجه‌ای: در یک گراف، رئوسی که دارای بیشترین یال‌ها هستند، از بالاترین مرکزیت درجه‌ای برخوردارند، به عبارتی دیگر، رئوس دارای بیشترین پیوند مستقیم با سایر رئوس موجود در گراف را رئوس با مرکزیت درجه‌ای بالا می‌خوانند. مرکزیت درجه‌ای در یک شبکه اجتماعی، نشان‌دهنده تعداد ارتباطات آن گره با سایر گره‌های تشکیل‌دهنده شبکه است. شاخص مرکزیت گره k یا P_k از طریق معادله زیر محاسبه می‌شود:

$$C_D(P_k) = \sum_{i=1}^n a(p_i, p_k)$$

که در آن n تعداد گره‌های موجود در شبکه و $a(p_i, p_k)$ صورت اتصال دو گره P_i و P_k برابر با یک و در غیر این صورت برابر با صفر است (همان).

الگوریتم پیچ‌رنک: این الگوریتم که در سال ۱۹۹۸ توسط پیچ و برین^۱ ارائه شده است، بر مبنای اتصالات بین صفحات وب عمل می‌کند. پیچ‌رنک، رتبه‌بندی استاتیک صفحات وب است و محاسبه آن بر اساس ایده محاسبه اعتبار^۲ یک عامل در شبکه‌های اجتماعی است. بر اساس مفاهیم مطرح شده در باره اعتبار در شبکه‌های اجتماعی: الف) وجود یک پیوند از یک صفحه به صفحه دیگر، به طور ضمنی نشان‌دهنده نفوذ^۳ صفحه مقصد است. بنابراین، پیوند ورودی بیشتر برای صفحه i نشان‌دهنده نفوذ بیشتر این صفحه است. ب) صفحاتی که به i اشاره می‌کنند، خود نیز دارای اعتبارهای متفاوتی هستند. پیوند رسیده از یک صفحه با اهمیت بالاتر، ارزش بیشتری نسبت به پیوند رسیده از یک صفحه با ارزش کمتر دارد. به عبارتی، صفحه‌ای مهم است که از صفحات مهم دیگر پیوند گرفته باشد (طیبی، هاشمی، و محدث خراسانی، ۱۳۸۶). برای محاسبه پیچ‌رنک در میان N صفحه، در مرحله اول پیچ‌رنک اولیه تمامی صفحات $\frac{1}{N}$ فرض می‌شود. در مرحله دوم پیچ‌رنک اولیه هر صفحه به طور مساوی (بی‌وزن) یا نسبی (وزن‌دار) در میان صفحاتی که به آنها پیوند داده‌اند توزیع می‌شود تا پیچ‌رنک ثانویه هر صفحه به دست آید. مرحله دوم به طور متناوبی تکرار می‌شود تا زمانی که پیچ‌رنک هر صفحه به عددی ثابت همگرا شود. اگر صفحه‌ای هیچ پیوند خروجی نداشته باشد، پیچ‌رنک آن به طور مساوی در میان همه صفحات توزیع می‌شود. معادله پیچ‌رنک به صورت ذیل است:

$$PR_i = \frac{1-d}{N} + d \sum_{j \in B_i} \frac{PR_j}{N_j}$$

که در آن N تعداد کل صفحات موجود در وب، PR_j تعداد پیوندهای خروجی صفحه j می‌باشد که با صفحه i در ارتباط است، N_i نشانگر پیوندهای خروجی صفحه i ، و B_i نشانگر پیوندهای ورودی صفحه i است. همچنین d یک ضریب تعدیل‌کننده است که عددی بین صفر و یک اختیار می‌کند. در پیچ و برین (۱۹۹۸) این ضریب 0.85 تعیین شده است. همان‌طور که از قسمت دوم معادله پیداست، امتیاز پیچ‌رنک هر صفحه پیونددهنده را بر تعداد پیوندهای خروجی آن تقسیم می‌کنند تا اولاً رتبه هر صفحه به طور عادلانه‌ای در میان خروجی‌هایش تقسیم شده، و ثانیاً جمع رتبه‌ها به عدد یک نرمال‌سازی شود (پیچ و برین، ۱۹۹۸). هدف اصلی از ارائه الگوریتم پیچ‌رنک، استفاده از آن برای رتبه‌بندی صفحات وب بوده است. لیکن پژوهش‌هایی مانند لیو، بُلن، نلسون، و سُمیل^۴ (۲۰۰۵)، بوگرس، کوکس، و وان دن بوش (۲۰۰۸)، دینگ، یان، فراژو، و کاورلی^۵ (۲۰۰۹)، دینگ (۲۰۱۱)، و زوارقی (۱۳۹۱) این شاخص را برای رتبه‌بندی نویسندگان مورد استفاده قرار داده‌اند.

نکته قابل ذکر در تحلیل شبکه‌های ارتباطی نویسندگان این است که تحلیل‌های مذکور غالباً در حوزه‌های موضوعی کلی صورت گرفته، و تا کنون تحلیلی در حوزه‌های خاص موضوعی مانند حوزه موضوعی یک پرس‌وجو انجام نپذیرفته است. همچنین، نویسندگان رتبه‌بندی شده بر اساس شاخص‌های تحلیل شبکه‌های اجتماعی، از نظر خبرگی در حوزه موضوعی مورد تحلیل، در معرض قضاوت ربط قرار نگرفته‌اند. بنابراین مشخص نیست کدام یک از شاخص‌های تحلیل شبکه‌های اجتماعی، از عملکرد بهتری در بازیابی و رتبه‌بندی خبرگان یک حوزه موضوعی برخوردار می‌باشد. برای رفع خلأ موجود، پژوهش حاضر مدل سند-محور بازیابی تخصص را با شبکه‌های ارتباطی نویسندگان مقالات دانشگاهی در جهت یافتن و رتبه‌بندی نویسندگان خبره ترکیب می‌کند: به طوری که ابتدا نظام بازیابی اطلاعات اسنادی را که گمان می‌کند با پرس‌وجو مرتبط هستند بازیابی نموده، سپس شبکه‌های مختلف ارتباطی نویسندگان در میان مقالات بازیابی شده برای هر پرس‌وجو تشکیل شده، و نویسندگان بر حسب شاخص‌های خرد تحلیل شبکه‌های اجتماعی رتبه‌بندی شده، و مورد قضاوت ربط قرار می‌گیرند. شناسایی عملکرد مدل بازیابی تخصص در شرایطی که افراد واقع شده در اسناد بر اساس شاخص‌های مرکزیت درجه‌ای و پیچ‌رنک در شبکه‌های استنادی، هم‌استنادی و هم‌تألیفی نویسندگان رتبه‌بندی می‌شوند، هدف اصلی پژوهش است. آگاهی از عملکرد هر کدام از این رتبه‌بندی‌ها که بر حسب سنج‌های ارزیابی نظام‌های بازیابی اطلاعات اندازه‌گیری می‌شود، به شناسایی روش یا روش‌هایی که

¹ Page & Brin

² Prestige

³ Authority

نیکزاد (۱۳۹۰)، حسن‌زاده، خدادوست و زندیان (۱۳۹۱)، زوارقی (۱۳۹۱)، عصاره، سهیلی، فرج‌پهلوی، و معرف‌زاده (۱۳۹۱)، عرفان‌منش و بصیریان جهرمی (۱۳۹۲)، سهیلی، عصاره، و خادمی (۱۳۹۲)، و عصاره، صراطی شیرازی و خادمی (۱۳۹۳) از جمله این پژوهش‌ها هستند. در پژوهش‌هایی که توصیف آنها گذشت، نویسندگان پرکار، پراستناد، و با مرکزیت بالا در شبکه‌های هم‌تألیفی، استنادی و هم‌استنادی به طور تلویحی مترادف با خبره موضوعی خاص قلمداد نشده‌اند. حتی اگر پژوهش‌های پیشین، نویسندگان با شرایط گفته‌شده را به عنوان افراد خبره در یک حوزه موضوعی خاص قلمداد می‌کردند، از آنجایی که تحلیل‌های صورت گرفته در یک حوزه موضوعی نسبتاً کلی انجام می‌گرفت، نمی‌توانست اهداف در نظر گرفته شده برای این پژوهش را محقق کند. چرا که در این پژوهش شبکه‌های ارتباطی نویسندگان در سطح مقالات بازیابی‌شده برای هر پرس‌وجو تشکیل شده، و نویسندگان بر حسب شاخص‌های خرد تحلیل شبکه در شبکه‌های مذکور رتبه‌بندی شده، و در نهایت عملکرد هر مدل از نظر یافتن نویسندگان خبره با استفاده از سنجه‌های ارزیابی نظام بازیابی اطلاعات سنجیده می‌شود.

پیش از آنکه مدل‌های بازیابی تخصص توسط بالوگ و همکارانش فرمالیزه شود، استفاده از شبکه‌های ارتباطی برای یافتن خودکار خبرگان رایج بود. اولین تلاش‌ها در این زمینه را می‌توان در پژوهش‌هایی ردیابی کرد که سعی می‌کردند خبرگان نهفته در ارتباطات ایمیلی را کشف کنند. از آنجا که اسناد ایمیلی به طور طبیعی فعالیت‌ها و علایق افراد را در بر داشتند، برای عمل خبره‌یابی مناسب قلمداد می‌شدند. پژوهش شوارتز و وود^۱ (۱۹۹۳) اولین مطالعه‌ای است که برای یافتن افراد، به مشاهده الگوهای ارتباطات ایمیلی پرداخته است. در آن پژوهش، فیلدهای "از" و "به" از لاگ‌های ایمیلی ۱۵ وبسایت به مدت دو ماه جمع‌آوری شد و برای تولید گرافی که حدوداً ۵۰ هزار نفر را در بر داشت، مورد استفاده قرار گرفت. سپس تعدادی از الگوریتم‌های اکتشافی گراف برای خوشه‌بندی افراد دارای علایق مشترک به کار گرفته شد: یعنی زیرمجموعه‌ای از گره‌ها که ارتباطات بسیار بالایی با هم داشتند، مورد جستجو قرار گرفت. انگیزه مؤلفان برای توسعه این فنون، مشارکت در توسعه ابزارهای پویایی بود که به دنبال شناسایی افراد دارای علایق مشترک بودند. کشف افرادی که در باره موضوعی خاص صاحب دانش هستند، به عنوان یکی از کاربردهای بالقوه این پژوهش شناسایی شد.

اثربخشی بهتری دارند، منجر می‌شود. هر چقدر نظام بازیابی نتایج مربوط به بیشتری را بازیابی نموده و تعداد نتایج مربوط در میان نتایج اول بازیابی شده بیشتر باشد، اثربخشی نظام بیشتر خواهد بود. اثربخشی در بطن سنجه‌های ارزیابی نظام‌های بازیابی اطلاعاتی مورد توجه قرار گرفته است. نتایج پژوهش می‌تواند در طراحی نظام‌های اطلاعاتی که سرویس جستجوی خبرگان را ارائه می‌دهند مورد استفاده قرار گیرد.

سوال پژوهش

۱. عملکرد مدل سند-محور بازیابی تخصص در امر یافتن نویسندگان خبره به چه صورت است؟
۲. رتبه‌بندی افراد در شبکه استنادی بر اساس سنجه‌های مرکزیت درونی و پیچ‌رنک، چه تأثیری بر مدل سند-محور بازیابی تخصص می‌گذارد؟
۳. رتبه‌بندی افراد در شبکه هم‌استنادی بر اساس سنجه‌های مرکزیت و پیچ‌رنک، چه تأثیری بر مدل سند-محور بازیابی تخصص می‌گذارد؟
۴. رتبه‌بندی افراد در شبکه هم‌تألیفی بر اساس سنجه مرکزیت، چه تأثیری بر مدل سند-محور بازیابی تخصص می‌گذارد؟

پیشینه پژوهش

جستجو در پایگاه‌های اطلاعاتی داخل کشور آشکار کرد که تا کنون در ایران در ارتباط با امر یافتن خبرگان دو پژوهش صورت گرفته است. پرنور و رضایی‌نور (۱۳۹۴) با این فرض که رتبه‌بندی خبرگان با استفاده از روش‌های ترکیبی بهتر از رتبه‌بندی خبرگان تنها بر اساس یکی از روش‌های محتوای محور یا شبکه‌محور است، به مقایسه نتایج روش ترکیبی با روش محتوای محور پرداختند. روش ترکیبی یا همان مدل پیشنهادی پژوهشگران شامل ترکیب مدل فضای برداری با ماتریس همستگی کلمات و الگوریتم پیچ‌رنک، و مدل محتوای محور صرفاً شامل مدل فضای برداری بود. نتایج پژوهش نشان داد که عملکرد مدل ترکیبی بهتر از مدل محتوای محور صرف می‌باشد. عطاپور و فهیم‌نیا (۲۰۱۸)، مدل سند-محور بازیابی تخصص را برای یافتن نویسندگان خبره به کار گرفته، و در محاسبه احتمال ارتباط افراد با اسناد، نوع وقوع اسامی نویسندگان در مقالات را مورد توجه قرار داده‌اند. از سوی دیگر دسته‌ای از مطالعات هستند که اخیراً در ایران رواج زیادی پیدا کرده‌اند و آن استفاده از شاخص‌های خرد تحلیل شبکه‌های اجتماعی مانند مرکزیت درجه‌ای، مرکزیت بینایی، مرکزیت نزدیکی و مرکزیت بردار ویژه برای رتبه‌بندی نویسندگان در شبکه‌های استنادی، هم‌تألیفی و هم‌استنادی استفاده می‌شود. پژوهش‌های حریری و

¹ Schwartz & Wood

کنار یک نفر دیگر انجام شد. نتایج نشان داد که در رتبه‌بندی نویسندگان، عملکرد مدل قدم‌زنی تصادفی در ماتریس غیرهمگن اسناد و نویسندگان، بهتر از عملکرد رتبه‌بندی‌های مبتنی بر تعداد استنادات، انتشارات، مرکزیت‌درجه‌ای و پیچ‌رنک در شبکه نویسندگان است.

بوگرس، کوكس، و وان دن بوش (۲۰۰۸) در پژوهشی به بررسی قابلیت استفاده از شبکه‌های استنادی برای خبره‌یابی پرداختند. آنها ایده خود را بر روی ۱۴۷ مقاله از مقالات کارگروه یادگیری ماشین در دانشگاه تیلبورگ پیاده کردند. پژوهشگران رویکرد سند-محور خبره‌یابی، که نوعی رویکرد محتوای محور است، را به عنوان سطح پایه اتخاذ کردند. سپس به بررسی مرکزیت درجه‌ای درونی و پیچ‌رنک نویسندگان در شبکه استنادی مقالات و همچنین شبکه استنادی نویسندگان پرداختند. آنها این کار را یک بار به صورت مستقل از پرس‌وجو (با نام رتبه‌بندی ایستا) و بار دیگر به صورت وابسته به پرس‌وجو انجام دادند. یکی از نتایج پژوهش این بود که عملکرد رتبه‌بندی ایستا تفاوت معناداری با رتبه‌بندی وابسته به پرس‌وجو ندارد. توجه پژوهشگران این است که چون مجموعه آزمون مورد استفاده بسیار تخصصی است، همان نویسندگانی که در اسناد مربوط به موضوع پرس‌وجو رتبه‌های خوبی به دست می‌آورند، از نظر شاخص‌های مرکزیت درونی و پیچ‌رنک در شبکه استنادی کل مقالات نیز رتبه‌های خوبی به دست می‌آورند. نتیجه دیگر آنها این بود که رتبه‌بندی نویسندگان در شبکه‌های استنادی، به بهبود مدل پایه بازیابی تخصص نمی‌انجامد.

دینگ، هان، لیو، و ایروین^۹ (۲۰۱۲) برای یافتن نویسندگان خبره، در بطن مدل سند-محور بازیابی تخصص به ترکیب شبکه‌های کتابشناختی ناهمگن شامل سند-سند، نویسنده-سند، و هم‌تألیفی پرداختند. آنها ایده خود را بر روی پایگاه کتابشناسی علوم کامپیوتر^{۱۰} پیاده‌سازی کردند. نتایج پژوهش نشان داد مدل پیشنهادی عملکرد مدل سند-محور بازیابی تخصص را به طور معناداری بهبود می‌بخشد.

نشاطی، هاشمی و بیگی (۲۰۱۴) به امر یافتن نویسندگان خبره در شبکه‌های کتابشناختی پرداختند. آنها روش‌های قبلی خبره‌یابی را که امتیاز یکسانی برای همه نویسندگان مرتبط با یک سند تخصیص می‌دهند ناکافی دانسته و معتقدند که نویسندگان بانفوذ در حوزه موضوعی مقاله باید امتیاز بیشتری را دریافت کنند. بدین منظور، آنان سه گروه از مشخصه‌ها - شامل مشخصه‌های ساختاری (مانند تعداد هم‌تألیفی)، زمانی (مانند سال‌های فعالیت در حوزه موضوعی)، و فعالیت-محور (مانند تعداد استناد) را که به نظر آنها از

کمپبل، ماگلیو، کازی، و دام^۱ (۲۰۰۳) برای شناسایی افراد با نفوذ، ساختار پیوندی به وجود آمده از دریافت‌کنندگان و ارسال‌کنندگان ایمیل‌ها را با استفاده از الگوریتم هیتس^۲ مورد تحلیل قرار دادند. گراف تخصص با استفاده از فیلدهای "از" و "به" ایجاد شد، که شامل افراد به عنوان گره و ایمیل‌ها به عنوان یال بود. آنها نشان دادند که استفاده از امتیازات نفوذ الگوریتم هیتس برای رتبه‌بندی کاندیداها، در مقایسه با روش محتوا-محور از دقت بالاتر و جامعیت کمتری برخوردار است.

علاوه بر ارتباطات ایمیلی، فنون تحلیل شبکه‌های اجتماعی برای امر خبره‌یابی در سایر منابع ایجاد شبکه‌های اجتماعی نیز به کار گرفته شده است. این منابع شامل لاگ‌های چت (رلیچ، لین، و گریفث^۳، ۲۰۰۷)، فوروم‌های بحث (ژانگ، آکرمان، و آدامیک^۴، ۲۰۰۷)، و اطلاعات شبکه‌های هم‌تألیفی در پایگاه‌های کتابشناختی (لی، تانگ، ژانگ، لو، لیو، و هونگ^۵، ۲۰۰۷؛ ژانگ، تانگ، و لی^۶، ۲۰۰۷) می‌باشد. برای نمونه، ژانگ، آکرمان، و آدامیک (۲۰۰۷) در جهت شناسایی کاربرانی که خبرگی بالایی دارند، به تحلیل اجتماعی بزرگ از برنامه‌نویسان جاوا پرداختند. گراف اجتماعی شامل تعاملات پرسش/پاسخ کاربران بود که به منظور تشویق فعالیت‌های پاسخ‌دهی، یال‌ها از سوی پرسش‌ها به پاسخ‌ها جهت‌دار بودند. سه سنجه هیتس، پیچ‌رنک و نسبت پاسخ‌ها به پرسش‌ها مورد مقایسه قرار گرفتند. نتایج نشان داد که سنجه "نسبت پاسخ‌ها به پرسش‌ها" رتبه‌بندی بهتری از خبرگان ارائه کرده و در مقایسه با سایر روش‌ها، عملکرد مطلوبتری دارد. در نتیجه، پاسخ‌دهی کاربرانی که خودشان پرسش‌های زیادی مطرح می‌کنند، فعالیتی نیست که حاکی از خبرگی بالای فرد باشد.

ژو، اورشانسکی، ژا، و گیلز^۷ (۲۰۰۷) در پژوهشی به مسئله رتبه‌بندی اسناد و نویسندگان پرداختند. آنها به ارائه یک الگوریتم جدید پرداختند که مبتنی بر قدم‌زنی تصادفی^۸ در ماتریس غیرهمگن اسناد و نویسندگان بود. قدم‌زدن‌های تصادفی مکرر، به بردار ویژه‌ای منجر می‌شود که شامل رتبه‌بندی نویسندگان و رتبه‌بندی اسناد است. درایه‌های ماتریس مشتمل بر روابط نویسنده-نویسنده، سند و سند-نویسنده بود. فرضیه پژوهشگران این بود که نویسندگان قوی و اسناد قوی همدیگر را تقویت می‌کنند. آنها الگوریتم خود را بر روی مقالات گردآوری شده از پایگاه CiteSeer آزمون کردند. قضاوت ربط توسط خود پژوهشگران در

¹ Campbell, Maglio, Cozzi, & Dom

² HITS algorithm

³ Ehrlich, Lin, & Griffiths

⁴ Zhang, Ackerman, & Adamic

⁵ Li, Tang, Zhang, Luo, Liu, & Hong

⁶ Zhang, Tang, & Li

⁷ Zhou, Orshanskiy, Zha, & Giles

⁸ Random walking

⁹ Deng, Han, Lyu, & King

¹⁰ DBLP

قابلیت تمییز نویسندگان بانفوذ برخوردار هستند- شناسایی کردند. سپس برای محاسبه نفوذ نویسندگان، با اتکا به مشخصه‌های مذکور دو الگوریتم یادگیری را توسعه داده و در بطن مدل خبره‌یابی مورد استفاده قرار دادند. نتایج پژوهش نشان داد استفاده از شاخص نفوذ نویسندگان در محل ارتباط اسناد و اشخاص عملکرد مدل‌های پایه خبره‌یابی را به طور معناداری بهبود می‌بخشد.

پاتول، کومار، چاودیوری، و ناندی^۱ (۲۰۱۵) در پژوهشی با استفاده از الگوریتم پیچ‌رنک به ارائه روشی جدید برای رتبه‌بندی نویسندگان در شبکه کتابشناختی چندلایه پرداختند. روش رتبه‌بندی آنها بدین صورت بود که ابتدا مقالات و نویسندگان موجود در مجموعه در طی فرایندهای تکرارشونده رایج در پیچ‌رنک به ترتیب بر اساس اسناد و هم‌تألیفی امتیازدهی و رتبه‌بندی شدند. سپس سهم هر نویسنده در مقالات هم‌تألیف به شیوه‌های گوناگون تعیین، و در نهایت رتبه‌بندی نویسندگان ایجاد شد. نتایج نشان داد بین روش رتبه‌بندی آنها و رتبه‌بندی بر اساس شاخص اچ در رتبه‌های بالا همبستگی قوی‌تری وجود دارد، اما همبستگی آن دو در رتبه‌های پایین کمتر بوده، و روش پیشنهادی پژوهش در معرفی نویسندگان جوان عملکرد بهتری داشت.

پژوهش حاضر با پژوهش بوگرس، کوکس، و وان دن بوش (۲۰۰۸) ارتباط نزدیکی دارد، لیکن از جهاتی با آن متفاوت است. در حالی که پژوهش حاضر برای یافتن نویسندگان خبره، در چارچوب مدل بازیابی تخصص از شبکه‌های هم‌تألیفی، هم‌استنادی، و استنادی نویسندگان استفاده نموده، و عملکرد آنها را از نظر یافتن نویسندگان خبره مورد ارزیابی قرار می‌دهد، بوگرس و همکاران عملکرد یافتن نویسندگان خبره را تنها با استفاده از شبکه‌های استنادی نویسندگان و اسناد مورد ارزیابی قرار داده‌اند. آنها گرچه از روش هم‌استنادی نیز استفاده کردند، اما رویکرد آنها بر بهبود رتبه نویسندگان در صورت هم‌استنادی با نویسندگان رتبه برتر تمرکز داشت. نهایتاً اینکه پژوهش آنها بر روی یک مجموعه کوچک (۱۴۷ مقاله از مقالات کارگروه یادگیری ماشین) انجام، و داده‌های استنادی آن از پایگاه گوگل اسکالر گرفته شده است. در حالیکه پژوهش حاضر بر روی حجم نسبتاً انبوهی از اسناد (۹۶۳۷۵ مقاله در حوزه علم اطلاعات و کتابداری) صورت گرفته و استنادهای آن مبتی بر ارجاعات همان مقالات می‌باشد.

روش پژوهش

این پژوهش از نوع تجربی مقدماتی است. پیاده‌سازی این پژوهش مستلزم حضور پرس‌وجوها، بازیابی اسناد برای پرس‌وجوها، تشکیل

^۲ Terrier

^۳ برای آگاهی از علت انتخاب این مدل بازیابی به عطاپور و فهم‌نیا (۲۰۱۸) مراجعه کنید. همچنین علاقمندان برای کسب اطلاعات کامل در مورد این مدل و سایر مدل‌های بازیابی اطلاعات می‌توانند به آملی و وان‌رایزبرگن (۲۰۰۲)، و مانینگ، رغوان و شوتز (۲۰۰۹) مراجعه کنند. اطلاعات کامل کتابشناختی این آثار در بخش منابع آمده است.

^۴ نرم‌افزار بیباکسل (Bibexcel) توسط یکی از پژوهشگران مشهور حوزه علم‌سنجی با نام Olle Persson تهیه شده و از آدرس <http://www.umu.se/inforsk/Bibexcel> به صورت رایگان در دسترس است.

^۵ یو سی آی نت (Ucinet) یکی از مشهورترین نرم‌افزارهای تحلیل شبکه‌های اجتماعی است که توسط سه تن از متخصصان علوم اجتماعی به نام‌های بورگانی، اورت، و فرمین (۲۰۰۲) تهیه شده است. از طریق آدرس <https://sites.google.com/site/ucinetsoftware/> می‌توان به طور رایگان به این نرم‌افزار دسترسی پیدا کرد.

^۶ نرم‌افزار Sci2 توسط کنی برنر و همکارانش در دانشگاه ایندیانا ای آمریکا تهیه شده و مناسب مصورسازی اطلاعات و محاسبه شاخص‌های خرد تحلیل شبکه‌های اجتماعی است. این نرم‌افزار از طریق آدرس <https://sci2.cns.iu.edu> به طور رایگان در دسترس است.

^۱ Paul, Kumar, Choudhury, & Nandi

پرس‌وجو بود، توسط پژوهشگر به دو نفر فوق اضافه شد. معیار مربوط بودن هر نویسنده به پرس‌وجو، کسب حداقل ۲ رأی مربوط بودن از جانب ۳ قضاوت‌کننده بود. در مجموع قضاوت ربط در مورد خبرگی نویسندگان برای ۳۰ پرس‌وجو تکمیل شد. پس از قضاوت ربط، عملکرد روش‌ها مورد مقایسه قرار گرفت تا روش یا روش‌های برتر شناسایی شوند.

برای ارزیابی عملکرد مدل بازیابی تخصص در امر یافتن نویسندگان خبره، از سنجه‌های ارزیابی عملکرد بازیابی اطلاعات استفاده شد که در ارزیابی‌های رسمی ترک^۱ نیز از آنها استفاده می‌شود. این سنجه‌ها عبارتند از دقت در سطح پنج نتیجه اول، میانگین متوسط دقت، و میانگین معکوس رتبه.

دقت در سطح ۵ نتیجه اول ($P@5$): دقت، عبارت است از نسبت اسناد مربوط بازیابی شده به کل اسناد بازیابی شده برای یک پرس‌وجو (بیزا بیتس و ریرو-تنو، ۱۳۸۴، ص. ۱۸). به طور معمول، دقت در بین n (۵، ۱۰، ۱۵، ۲۰، ۳۰، و ۱۰۰) سند اول بازیابی شده محاسبه می‌شود. در پژوهش حاضر، دقت در سطح ۵ نتیجه اول ($n=5$) مد نظر قرار می‌گیرد.

میانگین متوسط دقت (MAP): عبارت است از مجموع امتیازات متوسط دقت ($AveP(q)$) برای هر پرس‌وجو تقسیم بر تعداد کل پرس‌وجوها. متوسط دقت با استفاده از فرمول زیر محاسبه می‌شود (تورپین و شولر^۳، ۲۰۰۶).

$$AveP(q) = \frac{\sum_{k=1}^n (P(k) \times rel(k))}{\text{تعداد اسناد مرتبط}}$$

در معادله فوق $p(k)$ برابر با دقت در سطح بازیابی k می‌باشد. $rel(k)$ تابع شاخصی است که مقدار یک یا صفر به خود می‌گیرد. اگر آیتمی که در رتبه k قرار دارد، مرتبط باشد، $rel(k)$ برابر با یک، در غیر این صورت برابر با صفر منظور می‌شود.

میانگین معکوس رتبه (MRR): منظور از معکوس رتبه یک پرس‌وجو، معکوس رتبه اولین پاسخ صحیح (مربوط) است. میانگین معکوس رتبه، از تقسیم مجموع معکوس رتبه همه پرس‌وجوها بر تعداد کل پرس‌وجوها به دست می‌آید (ورهیس^۵، ۱۹۹۹).

در نهایت، برای سنجش پایایی قضاوت ربط در مورد نویسندگان خبره از روش آزمون‌های موازی (قضاوت همزمان چند نفر) استفاده شد که ضریب همبستگی ۰/۸۹ بین قضاوت‌ها، حاکی از پایایی قضاوت‌ها است.

یافته‌ها

پاسخ سؤال اول پژوهش: جدول ۱ که عملکرد مدل سند-محور بازیابی تخصص را نشان می‌دهد در امر بازیابی تخصص معمولاً به عنوان مدل پایه در نظر گرفته می‌شود. در این مدل، امتیاز هر سند از طریق مدل‌های بازیابی اطلاعات، و امتیاز هر فرد از جمع امتیاز اسناد مرتبط با وی یا جمع امتیاز نمایی اسناد مرتبط با وی به دست می‌آید. همان طور که مشاهده می‌شود، مدلی که امتیاز اسناد مرتبط با فرد را به وی اختصاص می‌دهد، در هر سه شاخص عملکرد بهتری داشته است.

پاسخ سؤال دوم پژوهش: جدول ۲ عملکرد مدل بازیابی تخصص را در شرایطی نشان می‌دهد که رتبه افراد در شبکه استنادی بر اساس سنجه‌های مرکزیت و پیچ‌رنک مبنای رتبه‌بندی قرار گرفته است. همان طور که مشاهده می‌شود از نظر سه شاخص عملکرد بازیابی، مدل مبتنی مرکزیت درجه‌ای درونی عملکرد بهتری در مقایسه با مدل مبتنی بر پیچ‌رنک داشته است.

سؤال دوم پژوهش تأثیر رتبه‌بندی افراد در شبکه استنادی بر مدل سند-محور بازیابی تخصص را مورد سؤال قرار می‌دهد. بدین منظور لازم است عملکرد این بخش با مدل پایه بازیابی تخصص که در سؤال اول مورد بررسی قرار گرفته، مقایسه شود. در اینجا، صرفاً به مقایسه مدل برتر پایه (مدلی که امتیاز اسناد مرتبط با فرد را به وی اختصاص می‌دهد) با مدل برتر رتبه‌بندی در شبکه استنادی (مرکزیت درجه‌ای درونی) پرداخته می‌شود. با توجه به نرمال بودن توزیع داده‌های $P@5$ ، AP ، و RR برای مقایسه مدل‌ها بر حسب شاخص‌های مذکور، از آزمون تی مستقل استفاده شد. جدول ۳ نتایج این آزمون را نشان می‌دهد. با توجه به اینکه در هر سه مورد معیار تصمیم بیشتر از ۰/۰۵ است، می‌توان نتیجه گرفت که میانگین $P@5$ ، AP ، و RR مدل‌ها تفاوت معناداری ندارند. یعنی با وجود اینکه شاخص‌های $P@5$ ، AP ، و RR مدل رتبه‌بندی افراد در شبکه‌های استنادی بر اساس مرکزیت درونی اندکی بهتر از مدل پایه بازیابی تخصص است، اما این تفاوت از نظر آماری معنادار نیست. نتیجه کلی این بخش آن است که رتبه‌بندی افراد بر اساس سنجه‌های مرکزیت در شبکه استنادی تأثیری در مدل پایه بازیابی تخصص از نظر میانگین $P@5$ ، MAP ، و MRR نمی‌گذارد.

پاسخ سؤال سوم پژوهش: جدول ۴ عملکرد مدل بازیابی تخصص را در شرایطی نشان می‌دهد که رتبه افراد در شبکه هم‌استنادی بر اساس سنجه‌های مرکزیت و پیچ‌رنک مبنای رتبه‌بندی قرار گرفته است. همان طور که مشاهده می‌شود در اینجا نیز از نظر سه شاخص عملکرد بازیابی، مدل مبتنی مرکزیت درجه‌ای درونی عملکرد بهتری در مقایسه با مدل مبتنی بر پیچ‌رنک

¹ TREC: Text REtrieval Conference

² MAP: Mean Average Precision

³ Turpin & Scholer

⁴ Mean Reciprocal Rank

⁵ Vorhees

جدول ۱- عملکرد مدل سند-محور بازیابی تخصص در امر یافتن نویسندگان خبره

MRR	MAP	P@5	مدل خبره‌یابی
۰/۸۶۸	۰/۲۵۳	۰/۷۴۰	مجموع امتیاز اسناد مرتبط با فرد
۰/۸۶۴	۰/۲۰۴	۰/۶۴۷	مجموع امتیاز نمایی اسناد مرتبط با فرد

جدول ۲- عملکرد مدل بازیابی تخصص بر مبنای رتبه افراد در شبکه استنادی

MRR	MAP	P@5	مدل خبره‌یابی
۰/۹۱۷	۰/۲۸۱	۰/۷۴۷	مرکزیت درجه‌ای درونی در شبکه استنادی نویسندگان
۰/۶۸۲	۰/۱۶۵	۰/۵۴۵	پیچ‌رنک افراد در شبکه استنادی نویسندگان

جدول ۳- نتایج آزمون تی مستقل برای مقایسه میانگین AP ، $P@5$ و RR مدل پایه بازیابی تخصص با مدل مبتنی بر رتبه‌بندی افراد در شبکه استنادی

نتیجه سنج	آماره F	درجه آزادی	معیار تصمیم
P@5	۱/۸۲۵	۵۸	۰/۹۱۴
AP	۱/۴۱۳	۵۸	۰/۲۵۳
RR	۳/۴۱۳	۵۸	۰/۳۹۷

جدول ۴- عملکرد مدل بازیابی تخصص بر مبنای رتبه افراد در شبکه هم‌استنادی

MRR	MAP	P@5	مدل خبره‌یابی
۰/۸۹۱	۰/۲۵۴	۰/۷۰۷	مرکزیت درجه‌ای افراد در شبکه هم‌استنادی نویسندگان
۰/۸۵۳	۰/۲۲۵	۰/۶۶۰	پیچ‌رنک افراد در شبکه هم‌استنادی نویسندگان

داشته است.

سؤال سوم پژوهش تأثیر رتبه‌بندی افراد در شبکه هم‌استنادی بر مدل سند-محور بازیابی تخصص را مورد سؤال قرار می‌دهد. بدین منظور لازم است عملکرد این قسمت با مدل پایه بازیابی تخصص مقایسه شود. در این قسمت نیز، صرفاً به مقایسه مدل برتر پایه (مدلی که امتیاز اسناد مرتبط با فرد را به وی اختصاص می‌دهد) با مدل برتر رتبه‌بندی در شبکه استنادی (مرکزیت درجه‌ای درونی) پرداخته می‌شود. با توجه به نرمال بودن توزیع داده‌های $P@5$ ، AP و RR برای مقایسه مدل‌ها بر حسب شاخص‌های مذکور، از آزمون تی مستقل استفاده شد. جدول ۵ نتایج این آزمون را نشان می‌دهد. با توجه به اینکه در هر سه مورد معیار تصمیم بیشتر از ۰/۰۵ است، می‌توان نتیجه گرفت که میانگین $P@5$ ، AP و RR مدل‌ها تفاوت معناداری ندارند. در اینجا نیز با وجود اینکه شاخص‌های $P@5$ ، AP و RR مدل رتبه‌بندی افراد در شبکه هم‌استنادی بر اساس مرکزیت درونی اندکی بهتر از مدل پایه بازیابی تخصص است، اما

این تفاوت از نظر آماری معنادار نیست. نتیجه کلی این بخش آن است که رتبه‌بندی افراد بر اساس سنج‌های مرکزیت در شبکه هم‌استنادی تأثیری در مدل پایه بازیابی تخصص از نظر میانگین $P@5$ ، MAP و MRR نمی‌گذارد.

پاسخ سؤال چهارم: جدول ۶ عملکرد مدل بازیابی تخصص را در شرایطی نشان می‌دهد که رتبه افراد در شبکه هم‌تألیفی بر اساس سنج مرکزیت درجه‌ای قرار گرفته است.

سؤال چهارم پژوهش تأثیر رتبه‌بندی افراد در شبکه هم‌تألیفی بر مدل سند-محور بازیابی تخصص را مورد سؤال قرار می‌دهد. بنابراین لازم است عملکرد این بخش با مدل پایه بازیابی تخصص مقایسه شود. در این قسمت نیز، صرفاً به مقایسه مدل برتر پایه (مدلی که امتیاز اسناد مرتبط با فرد را به وی اختصاص می‌دهد) با عملکرد مدل بازیابی تخصص مبتنی بر رتبه افراد در شبکه هم‌تألیفی پرداخته می‌شود. با توجه به غیر نرمال بودن توزیع داده‌های $P@5$ و AP برای مقایسه مدل‌ها بر حسب شاخص‌های

جدول ۵- نتایج آزمون تی مستقل برای مقایسه میانگین AP ، $P@5$ و RR مدل پایه بازیابی تخصص با مدل مبتنی بر رتبه‌بندی افراد در شبکه هم‌استنادی

نتیجه سنج	آماره F	درجه آزادی	معیار تصمیم
P@5	۰/۳۸۱	۵۸	۰/۵۹۸
AP	۲/۱۴۰	۵۸	۰/۹۴۱
RR	۰/۶۳۴	۵۸	۰/۷۰۴

جدول ۶- عملکرد مدل بازیابی تخصص بر مبنای رتبه افراد در شبکه هم‌تألیفی

MRR	MAP	P@5	مدل خبره‌یابی
۰/۵۸۶	۰/۰۹۹	۰/۳۷۳	مرکزیت درجه‌ای افراد در شبکه هم‌تألیفی مقالات

جدول ۷- نتایج آزمون مان-ویتنی برای مقایسه میانگین P@5 و AP مدل پایه بازیابی تخصص با مدل مبتنی بر رتبه‌بندی افراد در شبکه هم‌تألیفی

نتیجه	آماره آزمون	تعداد کل	معیار تصمیم
P@5	۷۰۴/۰۰۰	۶۰	۰/۰۰۰
AP	۷۸۳/۰۰۰	۶۰	۰/۰۰۰

جدول ۸- نتایج آزمون تی مستقل برای مقایسه میانگین RR مدل پایه بازیابی تخصص با مدل مبتنی بر رتبه‌بندی افراد در شبکه هم‌تألیفی

نتیجه	آماره F	درجه آزادی	معیار تصمیم
RR	۳۴/۰۵	۵۸	۰/۰۰۴

رتبه‌بندی نویسندگان شبکه‌های استنادی تفاوت معناداری با مدل سند-محور بازیابی تخصص ندارد، اما هر دوی آن‌ها از رتبه‌بندی نویسندگان در شبکه هم‌تألیفی عملکرد بهتری دارند. عدم بهبود در مدل بازیابی تخصص به واسطه استفاده از رتبه‌بندی نویسندگان شبکه‌های استنادی همراستا با نتایج پژوهش بوگرس، کوکس و وان دن بوش (۲۰۰۸) می‌باشد، اما کاهش عملکرد مدل بازیابی تخصص در اثر استفاده از رتبه‌بندی نویسندگان در شبکه هم‌تألیفی با نتایج پژوهش پیش گفته و همچنین پژوهش‌های دنگ، هان، لیو، و ایروین (۲۰۱۲) و نشاطی، هاشمی و بیگی (۲۰۱۴) همخوانی ندارد. به طوری که مدل مبتنی بر هم‌تألیفی در پژوهش بوگرس، کوکس تفاوت معناداری با مدل پایه بازیابی تخصص نداشته و در دو پژوهش بعدی، مدل استفاده شده به بهبود عملکرد مدل پایه بازیابی تخصص منجر شده است. دلیل آن را می‌توان به استفاده از الگوریتم هم‌تألیفی در ترکیب سایر الگوریتم‌ها مانند استناد و نفوذ دانست. استفاده از رتبه‌بندی افراد بر اساس مرکزیت درجه‌ای در شبکه هم‌تألیفی بر مبنای این ایده بود که فرد با مرکزیت درجه‌ای بالا به واسطه خبرگی در آن حوزه موضوعی، توانسته همکاری تعداد زیادی از سایر افراد را جلب کند. اما نتایج مشخص کرد که این ایده در عمل رتبه‌بندی نویسندگان خیره عملکرد خوبی در مقایسه با سایر روش‌ها نداشته است. مراجعه به فایل اسناد حاوی نویسندگانی که در شبکه هم‌تألیفی از مرکزیت درجه‌ای بالا برخوردار بودند، پرده از این واقعیت برداشت که نویسندگان مزبور در تألیف مقالات دارای نویسنده زیاد نقش داشتند، امری که مرکزیت درجه‌ای آنها را بالا برده است. بنابراین مرکزیت درجه‌ای بالای آنها، به خاطر همکاری‌های مکرر آنها با افراد مختلف نبوده است، بلکه بخاطر حضور آنها در تألیف مقالاتی بوده است که نویسندگان زیادی داشته است.

مذکور، از آزمون مان-ویتنی استفاده شد. جدول ۷ نتایج این آزمون را نشان می‌دهد. با توجه به اینکه در هر دو مورد معیار تصمیم کمتر از ۰/۰۵ است، نتیجه‌گیری می‌شود که میانگین P@5 و AP مدل‌ها تفاوت معناداری دارند. با مقایسه جداول ۶ و ۷ می‌توان پی برد که از حیث شاخص‌های میانگین P@5 و MAP عملکرد مدل پایه بازیابی تخصص از مدل مبتنی بر رتبه‌بندی افراد شبکه هم‌تألیفی بهتر بوده است. از طرفی دیگر، با توجه به اینکه داده‌های RR از توزیع نرمال برخوردار بودند، برای مقایسه عملکرد مدل پایه بازیابی تخصص با مدل مبتنی بر رتبه‌بندی در شبکه هم‌تألیفی از نظر شاخص مذکور، از آزمون تی مستقل استفاده گردید. جدول ۸ نتایج این آزمون را نشان می‌دهد. با توجه به اینکه معیار تصمیم کمتر از ۰/۰۵ است، می‌توان نتیجه گرفت که میانگین RR مدل‌ها تفاوت معناداری ندارند. در اینجا نیز، با مقایسه جداول ۶ و ۷ می‌توان پی برد که از حیث شاخص میانگین RR، عملکرد مدل پایه بازیابی تخصص از مدل مبتنی بر رتبه‌بندی افراد شبکه هم‌تألیفی بهتر بوده است. نتیجه کلی که از پاسخ سؤال چهارم می‌توان گرفت این است که رتبه‌بندی افراد بر اساس سنجه مرکزیت درجه‌ای در شبکه هم‌تألیفی بر عملکرد مدل پایه بازیابی تخصص از نظر میانگین P@5، MAP، و MRR اثر گذاشته و آن را کاهش می‌دهد.

بحث و نتیجه‌گیری

رویکردی که در این پژوهش برای رتبه‌بندی نویسندگان خبره استفاده شد، تشکیل شبکه‌های نویسندگان در اسناد بازیابی‌شده و رتبه‌بندی نویسندگان بر مبنای شاخص‌های مرکزیت درجه‌ای و پیچ‌رنک در شبکه‌های مذکور بود. اصلی‌ترین یافته پژوهش آن است که در امر یافتن نویسندگان خبره، عملکرد مدل مبتنی بر

- Hassanzadeh, M., Khodadust, R., Zandian, F. (2012). Analysis of Co-Authorship Indicators, Betweenness Centrality and Structural Holes of the Iranian Nanotechnology Researchers in Science Citation Index (1991-2011). 28(1), 223-249. (Persian)
- Soheili, F., Osareh, F., Khademi, R., (2013). The Scientific Structure of Biology in Iran During 1990 – 2008: A Co-citations Analysis. Studies on Library & Information Science. 5(12), 83-102. (Persian)
- Zavaraqi, R. (2013). Visualizing, and interpreting the intellectual structure of Iranian Science (1990-2009): a critical viwe. (PhD thesis), University of Tehran.
- Tayebi, MA., Tashakkori Hashemi, M., Mohades Khorasani, A., (2007). Web-analysis based ranking algorithms and its justified version for blog space. Paper presented at the first data mining conference, Tehran, Iran, 19-20 November.
- Erfanmanesh, M., Basirian Jahromi, R., (2013). The Co-authorship Network of the Articles Published in the National Studies on Librarianship and Information Organization Journal Using Social Networks Analysis Indexes. National Studies on Librarianship and Information Organization, 24(2), 76-96.
- Osareh, F., Soheili, F., Farajpahlou, A., moarefzadeh, A. (2013). A survey on centrality measure in co-authorship networks in information science journals. Library And Information Science Research (LISRJ), 2(2).
- Osareh, F., Serati Shirazi, M., Khademi, R. (2014). A Survey on Co-authorship Network of Iranian Researchers in the field of Pharmacy and Pharmacology in Web of Science during 2000-2012. JHA. 17(56), 33-45. (Persian)
- Atapour, H., Fahimnia, F. (2018). Applying Expert Retrieval Model for Finding Expert Authors. 33(4), 1713-1744. (Persian)
- Amati, G., van Rijsbergen, C. J., (2002). Probabilistic models of information retrieval based on measuring the divergence from randomness. ACM Trans. Inf. Syst. 20(4), 357-389.
- Balog, K., Azzopardi, L., & de Rijke, M. (2006). Formal models for expert finding in enterprise corpora. In Proceedings of the 29th annual international ACM SIGIR conference on Research and development in information retrieval (pp. 43-50). ACM.
- Balog, K., Fang, Y., de Rijke, M., Serdyukov, P., & Si, L. (2012). Expertise retrieval. Foundations and Trends in Information Retrieval, 6(2-3), 127-256.
- Bogers, T., Kox, K., & van den Bosch, A. (2008). Using citation analysis for finding experts in workgroups. In Proc. DIR (pp. 21-28).
- Borner, K. (2010). "Network Science: Theory, Tools, and Practice". In W. S. Bainbridge (Ed.), Leadership in Science and Technology: A

یکی دیگر از شاخص‌هایی که مبنای رتبه‌بندی در شبکه‌های استنادی و هم‌استنادی به کار گرفته شد، شاخص پیچ‌رنک بود. این شاخص بین استندهای دریافتی افراد تفاوت قائل شده، و به استندهای دریافتی از افراد پراستناد اهمیت بیشتری می‌دهد. نتایج نشان می‌دهد که در اغلب موارد این شاخص تفاوت معناداری با رتبه‌بندی افراد بر مبنای مرکزیت درجه‌ای در شبکه استنادی ندارد. حتی از دیدگاه آمار توصیفی عملکرد مرکزیت درجه‌ای بهتر از پیچ‌رنک است. با توجه به اینکه محاسبه این شاخص نیازمند استفاده از زنجیره مارکوف و انجام محاسبات پیچیده است، از نظر کارایی در جانب نظام‌بازیابی اطلاعات استفاده از مرکزیت درجه‌ای نویسندگان در شبکه استنادی منطقی‌تر به نظر می‌رسد. چرا که به کارگیری آن موجب می‌شود نظام اطلاعاتی زمان کمتری را صرف انجام محاسبات کند.

نکته آخر این که فنون تحلیل شبکه‌های اجتماعی برای شناسایی و یافتن افراد با نفوذ در حوزه‌های موضوعی نسبتاً کلی به کار گرفته می‌شود و سنج‌های تحلیل شبکه‌های اجتماعی در جاهایی که قرار است کنشگران اجتماعی مهم شناسایی شوند، خوب عمل می‌کنند. در این پژوهش، این فنون برای یافتن نویسندگان خبره در سطح پرس‌وجوهایی نسبتاً خاص مورد استفاده قرار گرفت. کشف یک اجتماع هم‌جنس در حوزه موضوعی برخی پرس‌وجوهای خاص ممکن است دشوار باشد. این امر بویژه در مورد موضوعات جدید که تازه ظهور کرده‌اند، بیشتر صدق می‌کند. لازمه موفقیت رویکرد به کار گرفته شده در پژوهش حاضر آن است که نظام اطلاعاتی بتواند برای هر پرس‌وجو به تعداد کافی سند مربوط ارائه نماید.

تعارض منافع

گزارش نشده است.

منبع حمایت کننده

گزارش نشده است.

References

- Baeza-Yates, R. (1961). Modern information retrieval. Ghasemi, A. H., Azadi, S., Translation; Tehran, Chapar Publication. (Persian)
- Pornour, E., Rezaenoor, J. (2015). A new Expert Finding model based on Term Correlation Matrix. 30 (4), 1147-1171. (Persian)
- Hariri, N., Nikzad, M. (2011). Co-authorship networks of Iranian articles in library and information science, psychology, management and economics in ISI during 2000-2009. 26(4), 825-844. (Persian)

- Publications.
- Brin, S., & Page, L. (1998). The Anatomy of Large Scale Hypertextual Web Search Engine. *Computer Networks*, 30 (1-7), pp. 107-117.
- Campbell, C. S., Maglio, P. P., Cozzi, A., & Dom, B. (2003). Expertise identification using email communications. In *Proceedings of the twelfth international conference on Information and knowledge management*(pp. 528-531). ACM.
- Deng, H., Han, J., Lyu, M. R., & King, I. (2012, June). Modeling and exploiting heterogeneous bibliographic networks for expertise ranking. In *Proceedings of the 12th ACM/IEEE-CS joint conference on Digital Libraries* (pp. 71-80). ACM.
- Ding, Y. (2011). Applying weighted PageRank to author citation networks. *Journal of the American Society for Information Science and Technology*, 62(2), 236-245.
- Ding, Y., Yan, E., Frazho, A., & Caverlee, J. (2009). PageRank for ranking authors in co-citation networks. *Journal of the American Society for Information Science and Technology*, 60(11), 2229-2243.
- Efron, M. (2009). Finding expert authors in institutional repositories. *Proceedings of the American Society for Information Science and Technology*, 46(1), 1-15.
- Ehrlich, K., Lin, C. Y., & Griffiths-Fisher, V. (2007). Searching for experts in the enterprise: combining text and social network analysis. In *Proceedings of the 2007 international ACM conference on Supporting group work* (pp. 117-126). ACM.
- Li, J., Tang, J., Zhang, J., Luo, Q., Liu, Y., & Hong, M. (2007). Eos: expertise oriented search using social networks. In *Proceedings of the 16th international conference on World Wide Web* (pp. 1271-1272). ACM.
- Liu, X., Bollen, J., Nelson, M.L., Sompel, H.V.. (2005). Co-authorship networks in the digital library research community. *Information Processing and Management*, 41 (6), 1462-1480.
- Manning, C. D., Raghavan, P. & Schütze, H. (2009). *Introduction to Information Retrieval*. Cambridge University Press.
- Reference Handbook. Thousand Oaks, CA: Sage
- Neshati, M., Hashemi, S. H., & Beigy, H. (2014). Expertise finding in bibliographic network: Topic dominance learning approach. *IEEE transactions on cybernetics*, 44(12), 2646-2657.
- Paul, P. S., Kumar, V., Choudhury, P., & Nandi, S. (2015, January). Temporal analysis of author ranking using citation-collaboration network. In *Communication Systems and Networks (COMSNETS), 2015 7th International Conference on* (pp. 1-6). IEEE.
- Schwartz, M. F. and Wood, D. C. M. (1993). Discovering shared interests using graph analysis. *Commun. ACM*, 36(8), 78-89.
- Small, H. (1973). Co-citation in the scientific literature: A new measure of the relationship between two documents. *Journal of the American Society for information Science*, 24(4), 265-269.
- Turpin, A., & Scholer, F. (2006). User performance versus precision measures for simple search tasks. In *Proceedings of the 29th annual international ACM SIGIR conference on Research and development in information retrieval* (pp. 11-18). ACM.
- Voorhees, E.M. (1999). TREC-8 Question Answering Track Report. In *Proceedings of the 8th Text Retrieval Conference* (Vol. 99, pp. 77-82.)
- Zhang, J., Ackerman, M. S., & Adamic, L. (2007). Expertise networks in online communities: structure and algorithms. In *Proceedings of the 16th international conference on World Wide Web* (pp. 221-230). ACM.
- Zhang, J., Tang, J., Li, J. (2007). Expert finding in a social network. In *Advances in Databases: Concepts, Systems and Applications* (pp. 1066-1069). Springer Berlin Heidelberg.
- Zhou, D., Orshanskiy, S., Zha, H., & Giles, C. L. (2007, October). Co-ranking authors and documents in a heterogeneous network. In *Data Mining, 2007. ICDM 2007. Seventh IEEE International Conference on* (pp. 739-744). IEEE.