Share •••

ما یک دیتابیس از متادیتای مقالات داریم که یکی از ستون های جدول articles، ستونی هست به نام authors_json و این ستون اطلاعات نوسندگان را دارد. نمونه هایی از این ستون در زیر هستند که هر کدام برای یک ردیف جداگانه هستند

given": "Lin", "family": "Yu", "orcid": "https://"}]
orcid.org/0000-0003-0971-2947", "sequence": "first",
"affiliations": ["School of Ocean and Civil Engineering
Shanghai Jiao Tong University Shanghai China", "Key
Laboratory of Marine Intelligent Equipment and System
given": "Lei", "} ,{Ministry of Education Shanghai China"

Show more

با ستون `authors_json` می‌توان یک تحلیل بسیار ارزشمند از **ساختار اجتماعی، نهادی و جغرافیایی یک حوزه پژوهشی** انجام داد. نکته مهم این است که برخی تحلیل‌ها فقط با اطلاعات نویسندگان ممکن‌اند، اما برای رسیدن به تصویری عمیق‌تر باید این ستون را با ستون‌هایی مثل `journal` ، `keywords` ، `abstract` ، `title` ، `publication_year` و تعداد استناد ترکیب کرد.

۱. شناسایی گروه‌های همکاری ثابت

مهم‌ترین تحلیل، ساخت **شبکه هم‌نویسندگی یا Co-authorship Network** است.

در این شبکه:

- هر نویسنده یک گره است.
- اگر دو نویسنده حداقل یک مقاله مشترک داشته باشند، بین آن‌ها یک یال ایجاد می‌شود.
- وزن یال برابر تعداد مقالات مشترک آن دو نفر است.

مثلاً اگر سه نویسنده در ۸ مقاله با یکدیگر همکاری کرده باشند، می‌توان نتیجه گرفت که احتمالاً یک تیم پژوهشی پایدار، آزمایشگاه یا پروژه مشترک را تشکیل می‌دهند.

این تحلیل به شما امکان می‌دهد موارد زیر را شناسایی کنید:

- تیم‌های پژوهشی پایدار
- گروه‌هایی که فقط یک همکاری مقطعی داشته‌اند
- هسته‌های اصلی تولید دانش
- گروه‌های مستقل و جدا از یکدیگر
- همکاری‌های جدید یا در حال شکل‌گیری

برای کلاسترکردن این شبکه می‌توان از الگوریتم‌هایی مانند موارد زیر استفاده کرد:

- Louvain Community Detection
- Leiden Algorithm
- Label Propagation
- Girvan–Newman
- Infomap

خروجی مطلوب مثلاً می‌تواند چنین باشد:

> کلاستر شماره ۳ شامل ۱۲ پژوهشگر است که از سال ۲۰۲۱ تا ۲۰۲۶، در ۲۳ مقاله مشترک فعالیت داشته‌اند و عمدتاً بر Digital Twin در حوزه مهندسی دریایی کار کرده‌اند.

---

۲. شناسایی نویسندگان کلیدی حوزه

شبکه هم‌نویسندگی فقط برای کلاسترکردن نیست. با معیارهای شبکه می‌توان نقش هر پژوهشگر را نیز تشخیص داد.

Degree Centrality

تعداد نویسندگان متفاوتی که یک فرد با آن‌ها مقاله مشترک دارد.

نویسنده‌ای با Degree بالا معمولاً:

- شبکه همکاری گسترده‌ای دارد؛
- در پروژه‌های متعدد شرکت می‌کند؛
- یا عضو یک گروه تحقیقاتی بزرگ است.

Weighted Degree

مجموع تعداد همکاری‌های نویسنده با دیگران است.

ممکن است یک فرد با افراد زیادی فقط یک بار همکاری کرده باشد، ولی فرد دیگری با چهار نفر، هرکدام ده مقاله مشترک داشته باشد. Weighted Degree این تفاوت را نشان می‌دهد.

Betweenness Centrality

مشخص می‌کند چه نویسنده‌ای ارتباط میان چند گروه پژوهشی را برقرار کرده است.

چنین فردی ممکن است:

- میان دو دانشگاه همکاری ایجاد کرده باشد؛
- دو زیرحوزه علمی را به یکدیگر متصل کند؛
- یا نقش هماهنگ‌کننده پروژه‌های بین‌رشته‌ای داشته باشد.

این نویسندگان را می‌توان **Bridge Author** یا نویسنده پل نامید.

Eigenvector Centrality یا PageRank

نویسنده‌ای را مهم می‌داند که با نویسندگان مهم دیگر همکاری کرده باشد.

بنابراین فردی که فقط تعداد زیادی همکاری دارد، لزوماً از فردی که با هسته اصلی حوزه کار می‌کند، مهم‌تر نیست.

Closeness Centrality

نشان می‌دهد یک نویسنده با چه سرعتی می‌تواند از طریق شبکه همکاری به سایر پژوهشگران حوزه متصل شود.

---

۳. تشخیص هسته مرکزی و پیرامون حوزه

با تحلیل Core–Periphery می‌توان نویسندگان را تقریباً به سه دسته تقسیم کرد:

1. **هسته مرکزی:** نویسندگان فعال، متصل و دارای همکاری‌های مکرر
2. **لایه میانی:** نویسندگانی که با چند گروه ارتباط دارند
3. **پیرامون:** نویسندگان تازه‌وارد یا افرادی که فقط یک یا دو مقاله دارند

این تحلیل نشان می‌دهد که حوزه موردنظر:

- توسط چند گروه محدود کنترل می‌شود؛
- یا بین تعداد زیادی تیم توزیع شده است؛
- یا هنوز حوزه‌ای نوظهور و پراکنده است.

برای اندازه‌گیری تمرکز حوزه می‌توان سهم مقالات تولیدشده توسط ۱۰ یا ۲۰ نویسنده برتر را محاسبه کرد.

مثلاً:

۱۵ نویسنده برتر، ۴۲ درصد مقالات حوزه را تولید کرده‌اند.

این عدد نشان‌دهنده تمرکز نسبتاً زیاد تولید علمی است.

---

۴. شناسایی رهبران علمی و نقش نویسندگان

فیلد sequence مشخص می‌کند نویسنده در مقاله اول بوده یا نویسنده اضافی.

از این اطلاعات می‌توان معیارهای زیر را استخراج کرد:

- تعداد مقالات هر نویسنده
- تعداد دفعات حضور به‌عنوان نویسنده اول
- نسبت نویسندگی اول به کل مقالات

- تعداد همکاری‌ها به‌عنوان نویسنده غیر اول
- تداوم حضور به‌عنوان نویسنده اول
- تغییر نقش فرد در طول زمان

مثلاً ممکن است یک پژوهشگر ابتدا نویسنده اول مقالات باشد و در سال‌های بعد بیشتر به‌عنوان نویسنده غیر اول ظاهر شود. این الگو ممکن است نشان‌دهنده حرکت او از نقش پژوهشگر اجرایی به سمت نقش سرپرست گروه باشد؛ البته برای تأیید سرپرستی، اطلاعات نویسنده مسئول یا ترتیب نویسنده آخر نیز لازم است.

شاخص ساده‌ای مانند زیر قابل تعریف است:

$$FLR_i = \frac{\text{تعداد مقالات نویسنده اول برای فرد } i}{\text{کل مقالات فرد } i}$$

که می‌توان آن را «نسبت رهبری در نویسندگی اول» در نظر گرفت.

---

۵. تحلیل پایداری تیم‌های پژوهشی

برای هر گروه نویسندگان می‌توان بررسی کرد:

- از چه سالی همکاری را آغاز کرده‌اند؛
- چند سال متوالی همکاری داشته‌اند؛
- در چند مقاله مشترک حضور داشته‌اند؛
- اعضای اصلی گروه چه کسانی‌اند؛
- چه افرادی وارد یا خارج شده‌اند؛
- آیا گروه در حال رشد، افول یا ثبات است.

می‌توان برای یک تیم، شاخص پایداری تعریف کرد:

$$TeamStability = \frac{\text{تعداد همکاری‌های تکرارشونده}}{\text{کل همکاری‌های گروه}}$$

همچنین می‌توان شباهت اعضای تیم را بین دو سال با ضریب جاکارد اندازه‌گیری کرد:

$$J(A_t, A_{t+1}) = \frac{|A_t \cap A_{t+1}|}{|A_t \cup A_{t+1}|}$$

اگر مقدار جاکارد بالا باشد، اعضای گروه در طول زمان ثابت مانده‌اند.

---

۶. شناسایی زوج‌های همکاری قدرتمند

برای هر دو نویسنده می‌توان تعداد مقالات مشترک را محاسبه کرد:

$$w_{ij} = \text{تعداد مقالات مشترک نویسندگان } i \text{ و } j$$

اما تعداد خام ممکن است گمراه‌کننده باشد. برای مثال دو نویسنده پرکار ممکن است تصادفاً چند همکاری داشته باشند. بنابراین معیارهای نرمال‌شده نیز مفید هستند.

ضریب جاکارد همکاری

$$J_{ij} = \frac{|P_i \cap P_j|}{|P_i \cup P_j|}$$

که در آن P_i مجموعه مقالات نویسنده i است.

اگر دو نویسنده تقریباً تمام مقالاتشان را با یکدیگر منتشر کرده باشند، ضریب جاکارد آن‌ها بالا خواهد بود.

Collaboration Dependency

$$D_{i \to j} = \frac{|P_i \cap P_j|}{|P_i|}$$

این معیار نامتقارن است. ممکن است ۸۰ درصد مقالات یک پژوهشگر با فرد دیگری باشد، ولی فقط ۱۰ درصد مقالات فرد دوم با پژوهشگر اول منتشر شده باشد.

این موضوع وابستگی علمی یا سازمانی میان پژوهشگران را نشان می‌دهد.

---

۷. تحلیل مؤسسات و دانشگاه‌های فعال

در `affiliations` اطلاعات سازمانی نویسندگان وجود دارد. با پاک‌سازی این داده‌ها می‌توان استخراج کرد:

- فعال‌ترین دانشگاه‌ها و مؤسسات
- تعداد مقالات هر مؤسسه
- تعداد نویسندگان هر مؤسسه
- مؤسساتی با بیشترین نویسنده اول
- مؤسساتی با بیشترین همکاری خارجی
- گروه‌های پژوهشی وابسته به هر مؤسسه
- رشد یا کاهش فعالیت هر مؤسسه در طول زمان

مثلاً از نمونه دوم می‌توان تشخیص داد که بخش بزرگی از تیم از مؤسسه `GEUS` است، ولی یک نویسنده از `SDFI` حضور دارد. بنابراین این مقاله یک همکاری بین‌سازمانی محسوب می‌شود.

یک شبکه جداگانه نیز می‌توان ساخت که در آن:

- گره‌ها مؤسسات باشند؛
- یال میان دو مؤسسه به معنی مقاله مشترک باشد؛
- وزن یال تعداد مقالات مشترک مؤسسات باشد.

این شبکه مشخص می‌کند کدام دانشگاه‌ها یا مراکز تحقیقاتی قطب همکاری در حوزه هستند.

---

۸. تحلیل همکاری‌های ملی و بین‌المللی

از متن affiliation معمولاً می‌توان کشور و گاهی شهر را استخراج کرد. سپس می‌توان مقالات را به سه دسته تقسیم کرد:

- تک‌مؤسسه‌ای
- چندمؤسسه‌ای درون یک کشور
- بین‌المللی

شاخص همکاری بین‌المللی می‌تواند به‌صورت زیر تعریف شود:

$$ICR = \frac{\text{تعداد مقالات دارای نویسنده از بیش از یک کشور}}{\text{کل مقالات}}$$

همچنین می‌توان بررسی کرد:

- کدام کشورها بیشترین تولید علمی را دارند؛
- کدام کشورها بیشتر با یکدیگر همکاری می‌کنند؛
- کدام کشور نقش پل میان چند منطقه را دارد؛
- آیا حوزه از نظر جغرافیایی متمرکز است؛
- چه کشورهایی اخیراً وارد حوزه شده‌اند.

خروجی شبکه کشورها می‌تواند جریان همکاری علمی را به‌خوبی نشان دهد.

---

۹. تشخیص گروه‌های محلی و همکاری‌های خارجی

برای هر مؤسسه می‌توان نسبت همکاری داخلی و خارجی را محاسبه کرد.

$$ExternalCollaborationRate = \frac{\text{مقالات مشترک با مؤسسات دیگر}}{\text{کل مقالات مؤسسه}}$$

با این شاخص می‌توان مؤسسات را به شکل زیر دسته‌بندی کرد:

- گروه‌های درون‌گرا با همکاری عمدتاً داخلی
- مؤسسات دارای شبکه همکاری بین‌المللی
- مراکز واسط و پل
- مؤسسات تازه‌وارد
- مراکز دارای یک شریک خارجی ثابت

---

۱۰. تحلیل تغییرات زمانی شبکه پژوهشگران

اگر ستون `publication_year` وجود داشته باشد، می‌توان شبکه را برای هر سال یا بازه زمانی

جداگانه ساخت.

مثلاً:

- ۲۰۱۵ تا ۲۰۱۸
- ۲۰۱۹ تا ۲۰۲۲
- ۲۰۲۳ تا ۲۰۲۶

سپس می‌توان بررسی کرد:

- چه کلاسترهایی ایجاد شده‌اند؛
- چه تیم‌هایی منحل شده‌اند؛
- چه نویسندگانی به مرکز شبکه رسیده‌اند؛
- کدام پژوهشگران از حوزه خارج شده‌اند؛
- ارتباط میان کدام گروه‌ها افزایش یافته است؛
- آیا حوزه از حالت پراکنده به یک جامعه منسجم تبدیل شده است.

این تحلیل یکی از بهترین روش‌ها برای فهم **جریان تکامل حوزه پژوهشی** است.

---

۱۱. شناسایی نویسندگان تازه‌وارد و پژوهشگران پایدار

برای هر نویسنده می‌توان موارد زیر را محاسبه کرد:

- سال اولین مقاله
- سال آخرین مقاله
- طول دوره فعالیت
- تعداد سال‌های فعال
- تعداد مقالات در هر سال
- بیشترین فاصله زمانی میان دو مقاله
- استمرار فعالیت

مثلاً:

$$ResearchSpan_i = LastYear_i - FirstYear_i + 1$$

و:

$$ActivityContinuity_i = \frac{\text{تعداد سال‌های دارای مقاله}}{\text{طول دوره فعالیت}}$$

بر اساس این معیارها نویسندگان را می‌توان به گروه‌های زیر تقسیم کرد:

- تازه‌وارد
- فعال موقت
- پژوهشگر پایدار
- پژوهشگر در حال رشد

- پژوهشگر کم‌فعال‌شده
- پژوهشگر خارج‌شده از حوزه

---

۱۲. تشخیص رشد یا افول گروه‌های پژوهشی

برای هر نویسنده، کلاستر یا مؤسسه می‌توان روند تعداد مقالات را محاسبه کرد.

مثلاً اگر تعداد مقالات یک گروه طی چهار سال به شکل زیر باشد:

$$2,\ 4,\ 7,\ 11$$

می‌توان آن را یک گروه در حال رشد در نظر گرفت.

در مقابل:

$$10,\ 7,\ 3,\ 1$$

می‌تواند نشانه کاهش فعالیت گروه باشد.

برای این کار می‌توان شیب رگرسیون تعداد مقالات برحسب سال را محاسبه کرد:

$$P_t = \beta_0 + \beta_1 t$$

- $\beta_1 > 0$: رشد
- $\beta_1 < 0$: افول
- $\beta_1 \approx 0$: ثبات

---

۱۳. تحلیل اندازه تیم‌های پژوهشی

از تعداد عناصر آرایه JSON می‌توان تعداد نویسندگان هر مقاله را به دست آورد.

تحلیل‌های مفید عبارت‌اند از:

- میانگین تعداد نویسنده در هر مقاله
- میانه و توزیع اندازه تیم
- تغییر اندازه تیم در طول زمان
- مقالات تک‌نویسنده‌ای
- مقالات دارای تیم‌های بسیار بزرگ
- تفاوت اندازه تیم میان موضوعات یا مجلات

اگر میانگین اندازه تیم در طول زمان افزایش یابد، ممکن است نشان‌دهنده موارد زیر باشد:

- پیچیده‌ترشدن مسائل حوزه
- افزایش بین‌رشته‌ای‌بودن

- وابستگی بیشتر به داده، آزمایشگاه یا تجهیزات
- افزایش پروژه‌های مشترک بین‌المللی

---

۱۴. ترکیب نویسندگان با موضوعات مقالات

بیشترین ارزش زمانی ایجاد می‌شود که داده نویسندگان را با `abstract` ، `title` یا `keywords` ترکیب کنید.

در این حالت می‌توان مشخص کرد:

- هر نویسنده روی چه موضوعاتی کار می‌کند؛
- تخصص غالب هر کلاستر چیست؛
- کدام گروه‌ها روی موضوع مشابه ولی جدا از هم کار می‌کنند؛
- چه نویسندگانی بین چند موضوع پل هستند؛
- هر موضوع در اختیار کدام گروه‌های پژوهشی است؛
- چه موضوعاتی در حال رشد هستند؛
- چه تیم‌هایی اخیراً جهت تحقیق خود را تغییر داده‌اند.

برای هر نویسنده می‌توان یک بردار موضوعی ساخت:

$$\mathbf{v}_i = [p_{i1}, p_{i2}, \ldots, p_{ik}]$$

که p_{ij} سهم مقالات نویسنده i در موضوع j است.

سپس می‌توان نویسندگان را نه‌فقط براساس همکاری، بلکه براساس **شباهت موضوعی** نیز کلاستر کرد.

این موضوع دو شبکه متفاوت ایجاد می‌کند:

1. شبکه همکاری واقعی
2. شبکه شباهت پژوهشی

ممکن است دو گروه هیچ مقاله مشترکی نداشته باشند، اما دقیقاً روی یک مسئله مشابه کار کنند. این مورد برای شناسایی رقبا یا فرصت‌های همکاری بسیار مهم است.

---

۱۵. شناسایی گروه‌های بین‌رشته‌ای

اگر کلمات کلیدی، موضوع مجله یا دسته‌بندی علمی در دسترس باشد، می‌توان برای هر نویسنده یا تیم تنوع موضوعی محاسبه کرد.

برای مثال با آنتروپی شانون:

$$H_i = -\sum_{k=1}^{K} p_{ik} \log(p_{ik})$$

- مقدار پایین: نویسنده بسیار تخصصی
- مقدار بالا: نویسنده بین‌رشته‌ای یا دارای موضوعات متنوع

می‌توان بررسی کرد آیا نویسندگان پل در شبکه همکاری، از نظر موضوعی نیز میان چند زیرحوزه قرار دارند یا خیر.

---

۱۶. تشخیص انتقال دانش بین مؤسسات

اگر نویسنده‌ای در سال‌های مختلف affiliation متفاوتی داشته باشد، می‌توان جابه‌جایی سازمانی او را بررسی کرد.

مثلاً:

- ابتدا در دانشگاه A
- سپس در مرکز تحقیقاتی B
- بعد همکاری میان A و B

این الگو ممکن است نشانه انتقال دانش، تشکیل همکاری جدید یا جابه‌جایی یک پژوهشگر کلیدی باشد.

برای این تحلیل باید affiliation هر نویسنده به تفکیک مقاله و سال نگهداری شود، نه اینکه فقط یک affiliation نهایی برای او در نظر گرفته شود.

---

۱۷. تحلیل ORCID

وجود `orcid` یک مزیت بسیار مهم است، زیرا مشکل تشابه اسامی را تا حد زیادی برطرف می‌کند.

از ORCID می‌توان تحلیل کرد:

- چه درصدی از نویسندگان ORCID دارند؛
- نویسندگان کلیدی حوزه تا چه اندازه قابل شناسایی دقیق هستند؛
- آیا یک نویسنده با شکل‌های مختلف نام در دیتابیس ثبت شده است؛
- چه رکوردهایی احتمالاً به یک فرد واحد تعلق دارند.

شاخص پوشش ORCID:

$$ORCIDCoverage = \frac{\text{تعداد نویسندگان یکتای دارای ORCID}}{\text{کل نویسندگان یکتا}}$$

پوشش پایین ORCID به معنی آن است که بخش مهمی از تحلیل باید با الگوریتم‌های رفع ابهام نام

انجام شود.

---

۱۸. تشخیص نویسندگان هم‌نام و داده‌های تکراری

قبل از هر تحلیل، باید مسئله **Author Name Disambiguation** حل شود.

مثلاً ممکن است موارد زیر یک فرد باشند:

- `J. Wang`
- `Jian Wang`
- `Wang, Jian`
- `Jian Z. Wang`

و از سوی دیگر، ممکن است دو فرد کاملاً متفاوت هر دو با نام `Wei Zhang` وجود داشته باشند.

برای ادغام یا تفکیک نویسندگان می‌توان از این ویژگی‌ها استفاده کرد:

- ORCID
- نام و نام خانوادگی کامل
- affiliation
- همکاران مشترک
- موضوع مقالات
- ایمیل، در صورت وجود
- کشور
- الگوی زمانی فعالیت

ترتیب پیشنهادی شناسه‌سازی نویسنده:

1. ORCID معتبر
2. نام نرمال‌شده همراه با affiliation
3. نام همراه با همکاران مشترک
4. مدل احتمالی رفع ابهام

بدون این مرحله، شبکه همکاری می‌تواند به‌شدت اشتباه باشد.

---

۱۹. شاخص‌های ساختاری کل حوزه

برای کل شبکه هم‌نویسندگی می‌توان معیارهای زیر را محاسبه کرد:

Density

نشان می‌دهد چه درصدی از ارتباطات ممکن واقعاً وجود دارند:

$$Density = \frac{2E}{N(N-1)}$$

که در آن N تعداد نویسندگان و E تعداد روابط همکاری است.

Average Clustering Coefficient

نشان می‌دهد همکاران یک نویسنده تا چه اندازه با یکدیگر نیز همکاری دارند.

مقدار بالا معمولاً نشان‌دهنده گروه‌های بسته و منسجم است.

Number of Connected Components

تعداد بخش‌های جدا از هم شبکه را نشان می‌دهد.

اگر تعداد مؤلفه‌ها زیاد باشد، حوزه احتمالاً پراکنده یا چندپارچه است.

Giant Component Ratio

$$GCR = \frac{\text{تعداد نویسندگان بزرگ‌ترین مؤلفه}}{\text{کل نویسندگان}}$$

مقدار بالا یعنی بخش بزرگی از جامعه پژوهشی به‌طور مستقیم یا غیرمستقیم به هم متصل است.

Modularity

شدت تفکیک شبکه به اجتماعات مختلف را نشان می‌دهد.

Modularity بالا معمولاً یعنی حوزه شامل چند جامعه پژوهشی نسبتاً مستقل است.

---

۲۰. شناسایی مقالات یا پروژه‌های اتصال‌دهنده گروه‌ها

گاهی یک مقاله شامل نویسندگانی از چند کلاستر متفاوت است. این مقاله می‌تواند یک **مقاله پل** یا پروژه اتصال‌دهنده باشد.

چنین مقالاتی ممکن است:

- آغاز یک همکاری جدید باشند؛
- دو زیرحوزه را به هم متصل کنند؛
- نتیجه یک پروژه بین‌المللی باشند؛
- یا نشان‌دهنده ظهور یک موضوع بین‌رشته‌ای باشند.

می‌توان برای هر مقاله تعداد کلاسترهای نویسندگان حاضر در آن را محاسبه کرد. هرچه این عدد بیشتر باشد، مقاله نقش اتصال‌دهنده بیشتری دارد.

---

۲۱. تشخیص رفتار حلقه بسته در گروه‌های پژوهشی

برخی تیم‌ها تقریباً همیشه فقط با اعضای ثابت خود مقاله منتشر می‌کنند. برخی دیگر دائماً نویسنده جدید جذب می‌کنند.

برای هر گروه می‌توان محاسبه کرد:

- نسبت نویسندگان تکراری
- نسبت نویسندگان جدید در هر سال
- نرخ همکاری خارج از کلاستر
- تنوع مؤسسات همکار
- تنوع کشورها

این تحلیل گروه‌ها را به دسته‌های زیر تقسیم می‌کند:

- بسته و پایدار
- باز و توسعه‌پذیر
- پروژه‌محور
- بین‌المللی
- محلی و مؤسسه‌محور

---

۲۲. شناسایی نسل‌های علمی و روابط احتمالی استاد–دانشجو

فقط با اطلاعات نویسندگان نمی‌توان رابطه استاد–دانشجو را قطعی تشخیص داد؛ اما می‌توان الگوهای احتمالی را یافت.

مثلاً:

- نویسنده A در تعداد زیادی مقاله حضور دارد؛
- نویسندگان اول مرتب تغییر می‌کنند؛
- affiliation مشترک است؛
- نویسندگان اول پس از چند سال مستقل می‌شوند.

این الگو می‌تواند نشان‌دهنده یک گروه تحت هدایت یک پژوهشگر ارشد باشد.

اما باید نتیجه را به‌صورت «رابطه احتمالی سرپرستی پژوهشی» گزارش کرد، نه رابطه قطعی استاد–دانشجو.

---

۲۳. تحلیل کیفیت و اثرگذاری گروه‌ها

اگر تعداد استناد، `is-referenced-by-count` ، شاخص‌های Altmetric یا اطلاعات مجله وجود داشته باشد، می‌توان تولید علمی را با اثرگذاری ترکیب کرد.

مثلاً:

- تعداد مقالات هر نویسنده
- مجموع استنادات
- میانگین استناد هر مقاله
- h-index در دیتاست حوزه
- استناد نرمال‌شده براساس سال
- اثرگذاری هر کلاستر
- اثرگذاری هر مؤسسه
- اثرگذاری همکاری‌های بین‌المللی در مقایسه با داخلی

این تحلیل مشخص می‌کند چه کسانی فقط پرکارند و چه کسانی واقعاً جریان علمی حوزه را هدایت می‌کنند.

---

۲۴. تحلیل «جریان‌های کاری» حوزه

برای فهم جریان‌های اصلی کار در حوزه، پیشنهاد می‌شود سه نوع کلاستر هم‌زمان ساخته شود:

کلاستر اجتماعی

براساس اینکه چه کسانی با یکدیگر مقاله می‌نویسند.

کلاستر موضوعی

براساس شباهت عنوان، چکیده و کلمات کلیدی مقالات.

کلاستر سازمانی

براساس دانشگاه، مؤسسه و کشور.

ترکیب این سه سطح می‌تواند نتایجی مانند زیر تولید کند:

> جریان اول حوزه توسط گروه‌هایی در چین و کره جنوبی هدایت می‌شود، بیشتر روی Digital Twin صنعتی و تولید هوشمند متمرکز است و دارای همکاری‌های درون‌کشوری قوی است.
>
> جریان دوم عمدتاً شامل دانشگاه‌های اروپای شمالی است، روی مدل‌سازی محیطی و هیدرولوژیکی تمرکز دارد و ساختار همکاری بین‌سازمانی بیشتری دارد.
>
> جریان سوم جدیدتر است، از سال ۲۰۲۳ رشد کرده و AI، پردازش تصویر و Digital Twin را به هم متصل می‌کند.

این سطح از تحلیل دقیقاً برای فهم «چه گروه‌هایی، روی چه موضوعاتی، در چه زمانی و با چه روابطی کار می‌کنند» مناسب است.

خروجی‌های کاربردی پیشنهادی

در نهایت، تحلیل شما بهتر است حداقل این خروجی‌ها را تولید کند:

1. شبکه هم‌نویسندگی کل حوزه
2. کلاسترهای اصلی نویسندگان
3. جدول نویسندگان کلیدی و معیارهای مرکزی‌بودن
4. زوج‌های دارای بیشترین همکاری تکراری
5. تیم‌های پایدار در بازه‌های زمانی
6. شبکه همکاری مؤسسات
7. شبکه همکاری کشورها
8. نویسندگان و گروه‌های تازه‌وارد
9. گروه‌های در حال رشد یا افول
10. ارتباط میان کلاسترهای نویسندگی و موضوعات مقالات
11. نویسندگان پل میان موضوعات یا مؤسسات
12. نمودار تحول شبکه در طول زمان
13. نقشه جغرافیایی همکاری‌ها
14. کیفیت داده و میزان پوشش ORCID
15. شاخص تمرکز و پراکندگی حوزه

اولویت پیشنهادی برای شروع

برای مرحله اول، این پنج تحلیل بیشترین ارزش را دارند:

1. نرمال‌سازی و شناسایی یکتای نویسندگان
2. ساخت شبکه وزنی هم‌نویسندگی
3. کشف کلاسترها با Louvain یا Leiden
4. محاسبه Degree، Betweenness، PageRank و تعداد مقالات
5. ترکیب کلاسترهای نویسندگان با سال انتشار و موضوع مقالات

این ترکیب نه‌فقط نشان می‌دهد «چه کسی با چه کسی کار کرده»، بلکه روشن می‌کند:

> چه تیم‌هایی جریان اصلی حوزه را می‌سازند، هر تیم روی چه موضوعی متمرکز است، چه کسانی میان جریان‌ها پل هستند و ساختار حوزه در طول زمان چگونه تغییر کرده است.

Sources