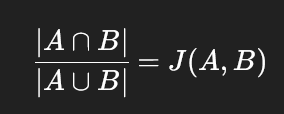
برای نوشتن یک داکیومنت برای پروژه **Link Prediction** و استفاده از معیارهای مختلف مانند **Jaccard**, **Adamic-Adar**, **Preferential Attachment (PA)**، و **Resource Allocation (RA)**، ابتدا باید توضیحاتی در مورد هر یک از این معیارها بدهیم و بعد توضیح دهیم که چرا از این معیارها برای این پروژه استفاده کرده‌ایم و مزایای آن‌ها چیست.

**1. Jaccard Similarity**

**تعریف:**

معیار **Jaccard Similarity** یکی از رایج‌ترین معیارها برای پیش‌بینی لینک‌ها است. این معیار برای دو مجموعه A و B به صورت زیر تعریف می‌شود:



که در آن:

* ∣A∩B∣|A \cap B| تعداد اعضای مشترک در دو مجموعه است (تقاطع).
* ∣A∪B∣|A \cup B| تعداد کل اعضای در هر دو مجموعه است (اتحاد).

**مزایا:**

* ساده و قابل فهم است.
* به راحتی می‌توان آن را برای هر دو مجموعه کاربر و فیلم پیاده‌سازی کرد.
* برای پیش‌بینی لینک‌ها، Jaccard به ما کمک می‌کند که بفهمیم دو گره (کاربر یا فیلم) چقدر مشابه هم هستند.

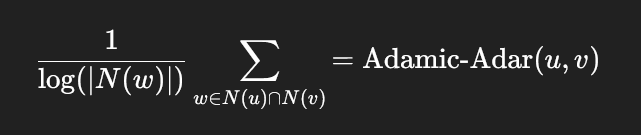
**استفاده در پروژه:**

برای پیش‌بینی فیلم‌هایی که ممکن است یک کاربر به آن‌ها علاقه‌مند باشد، از این معیار استفاده کردیم تا شباهت بین کاربران را بسنجیم. اگر دو کاربر مجموعه‌ای مشابه از فیلم‌ها را دیده باشند، احتمالاً علاقه‌مندی‌های مشابهی خواهند داشت.

**2. Adamic-Adar**

**تعریف:**

معیار **Adamic-Adar** به طور خاص بر روی گره‌هایی که در مجاورت یکدیگر قرار دارند، تمرکز دارد. این معیار به این صورت محاسبه می‌شود:



که در آن:

* N(u)N(u) همسایگان گره uu هستند.
* N(w)N(w) تعداد همسایگان گره ww است.
* ww تنها گره‌هایی هستند که هم همسایه‌های uu و هم همسایه‌های vv هستند.

**مزایا:**

* تمرکز بیشتری بر روی گره‌های نادر دارد. گره‌هایی که تعداد همسایگان کمی دارند، برای پیش‌بینی لینک‌ها اهمیت بیشتری پیدا می‌کنند.
* مدل پیچیده‌تر از Jaccard است و به دقت بالاتری برای داده‌های پیچیده‌تر می‌انجامد.

**استفاده در پروژه:**

از این معیار برای پیش‌بینی فیلم‌هایی استفاده کردیم که کاربر ممکن است به آن‌ها علاقه‌مند باشد، به خصوص زمانی که همسایگان مشترک بین کاربران زیاد نیستند. این معیار می‌تواند برای گره‌های نادر (کاربران با علاقه‌مندی‌های خاص) عملکرد خوبی داشته باشد.

**3. Preferential Attachment (PA)**

**تعریف:**

معیار **Preferential Attachment** یک ایده ساده ولی مؤثر است که می‌گوید گره‌هایی که درجه بالاتری دارند، احتمال بیشتری برای اتصال به گره‌های جدید دارند. این معیار برای گراف‌های مقیاس‌پذیر و شبکه‌های اجتماعی مناسب است و به صورت زیر تعریف می‌شود:



که در آن:

* ∣N(u)∣|N(u)| درجه (تعداد همسایگان) گره uu است.
* ∣N(v)∣|N(v)| درجه (تعداد همسایگان) گره vv است.

**مزایا:**

* در گراف‌های مقیاس‌پذیر و شبکه‌های اجتماعی بسیار مؤثر است.
* ساده است و می‌تواند برای داده‌هایی که تعداد زیادی ارتباط دارند، عملکرد خوبی داشته باشد.

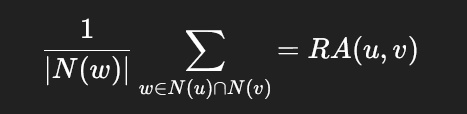
**استفاده در پروژه:**

از PA برای پیش‌بینی فیلم‌هایی استفاده کردیم که کاربران با درجه بالاتری (یعنی تعداد بیشتری از فیلم‌ها را مشاهده کرده‌اند) احتمال بیشتری برای علاقه‌مندی به فیلم‌های دیگر دارند. این معیار برای شبکه‌هایی که کاربران و فیلم‌ها تعداد زیادی ارتباط دارند مفید است.

**4. Resource Allocation (RA)**

**تعریف:**

معیار **Resource Allocation** مشابه معیار Adamic-Adar است، اما به جای استفاده از درجه گره‌ها، به میزان تخصیص منابع به گره‌های مشترک توجه می‌کند. این معیار به صورت زیر محاسبه می‌شود:



که در آن:

* N(u)N(u) همسایگان گره uu هستند.
* ∣N(w)∣|N(w)| تعداد همسایگان گره ww است.

**مزایا:**

* به گره‌هایی که تعداد کمی همسایه دارند توجه بیشتری می‌کند.
* به سادگی می‌توان آن را پیاده‌سازی کرد و عملکرد خوبی در شبکه‌های بزرگ دارد.

**استفاده در پروژه:**

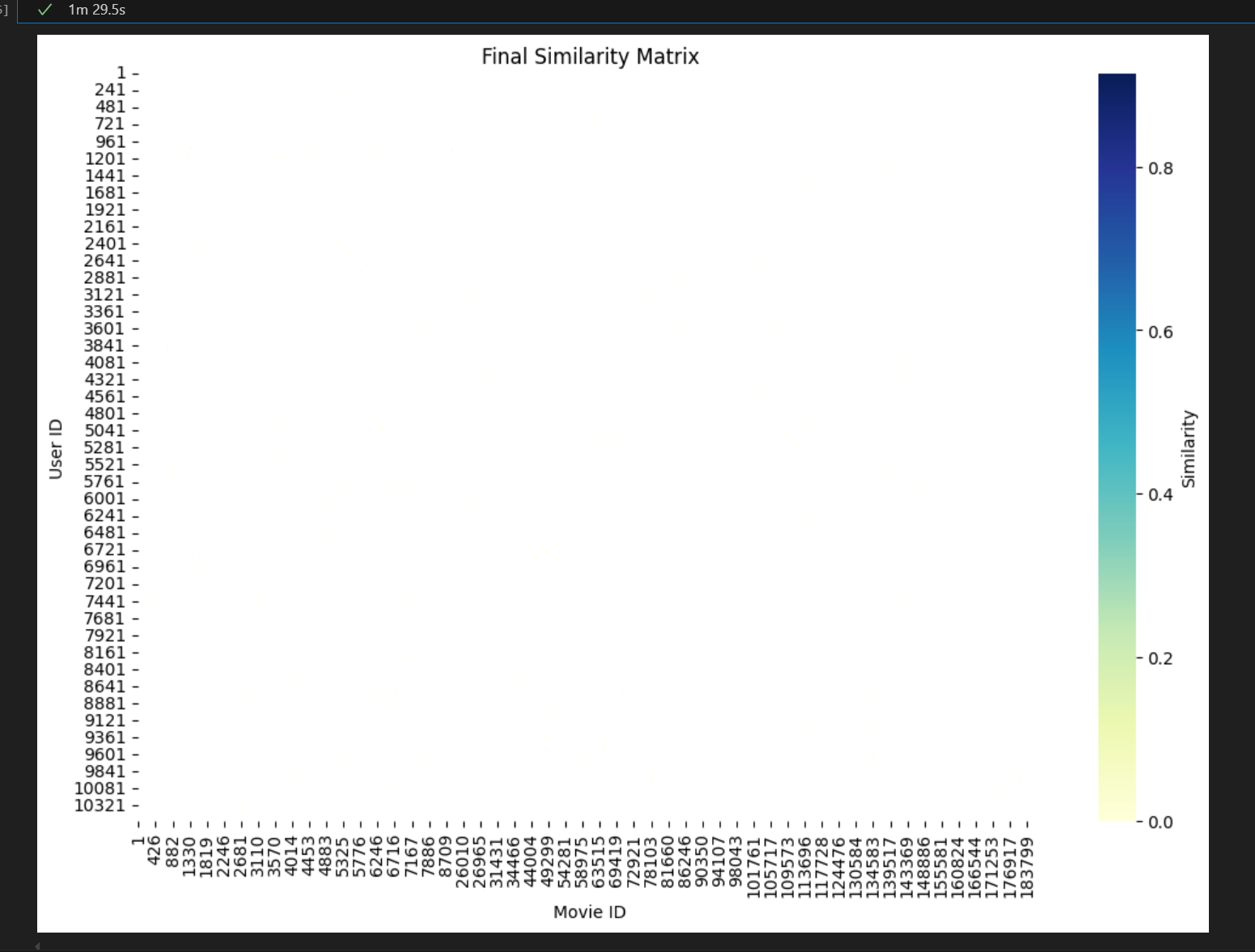
این معیار به ما کمک می‌کند تا فیلم‌هایی را پیش‌بینی کنیم که کاربران ممکن است به آن‌ها علاقه‌مند باشند، با تمرکز بر روی گره‌هایی که همسایگان مشترک کمی دارند ولی هرکدام اهمیت بیشتری دارند.

**چرا از این 4 معیار استفاده کردیم؟**

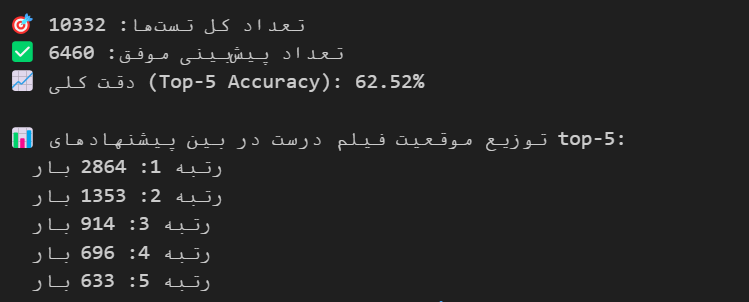
این معیارها برای Link Prediction استفاده می‌شوند چرا که:

* **Jaccard** برای محاسبه شباهت ساده و سریع بین گره‌ها به کار می‌رود.
* **Adamic-Adar** بر روی گره‌های نادر تمرکز دارد که می‌تواند برای گراف‌های پیچیده و کم‌داده بسیار مفید باشد.
* **PA** برای شبکه‌های مقیاس‌پذیر و اجتماعی که کاربران و فیلم‌ها تعداد زیادی ارتباط دارند، مؤثر است.
* **RA** مشابه Adamic-Adar است اما به منابع و همسایگان گره‌ها توجه بیشتری دارد و می‌تواند در شبکه‌های با تراکم پایین بهتر عمل کند.

این ترکیب از معیارها به ما اجازه می‌دهد تا از مزایای هرکدام از آن‌ها بهره‌مند شویم و پیش‌بینی دقیق‌تری از فیلم‌های مورد علاقه هر کاربر داشته باشیم. هر معیار ویژگی‌های خاص خود را دارد که در برخی سناریوها می‌تواند به بهبود پیش‌بینی‌ها کمک کند.



ماتریس شباهت کلی



نتیجه اجرا : دقت 62 درصد