پیش‌بینی لینک (Link Prediction) یکی از مسائل مهم در تحلیل شبکه‌هاست که هدف آن پیش‌بینی یال‌های جدید یا مفقود بین گره‌ها بر اساس ساختار فعلی شبکه است. این مسئله در حوزه‌هایی مانند شبکه‌های اجتماعی، بیوانفورماتیک و سیستم‌های توصیه‌گر کاربرد دارد.

در این تمرین، چند روش متداول پیش‌بینی لینک روی یک گراف نمونه (گراف باشگاه کاراته زاخاری) پیاده‌سازی و ارزیابی شده‌اند.

### توضیح کد : تقسیم داده‌ها

تابع train\_test\_split\_graph گراف ورودی را به سه بخش تقسیم می‌کند:

**Train Graph**: شامل درصدی از یال‌های اصلی (برای آموزش)

**Test Positive**: یال‌های واقعی که در آموزش حذف شده‌اند (برای تست مثبت)

**Test Negative**: جفت گره‌هایی که در گراف آموزشی یال ندارند (برای تست منفی)

**روش‌های امتیازدهی (Scoring Methods)**

تابع score\_method پنج روش رایج را پیاده‌سازی می‌کند:

Common Neighbors تعداد همسایه‌های مشترک

Jaccard Coefficient نسبت همسایه‌های مشترک به کل همسایه‌ها

Adamic-Adar Index وزن‌دهی به همسایه‌های مشترک بر اساس درجه آن‌ها

Preferential Attachment حاصل‌ضرب درجات دو گره

Resource Allocation Index مشابه Adamic-Adar ولی با وزن‌دهی متفاوت

**ارزیابی روش‌ها:**

تابع evaluate\_method برای هر روش معیارهای زیر را محاسبه می‌کند:

:**ROC AUC** مساحت زیر منحنی ROC

:**PR AUC** مساحت زیر منحنی Precision-Recall

:**Precision@k** دقت در انتخاب k پیش‌بینی برتر

همچنین داده‌های لازم برای رسم منحنی ROC و PR ذخیره می‌شوند.

**مقایسه روش‌ها :**

تابع compare\_link\_prediction همه روش‌ها را روی یک گراف اجرا کرده و نتایج را در قالب یک جدول خلاصه و یک لیست جزئیات برمی‌گرداند.

**ترسیم نمودارها:**

**سه تابع برای ترسیم نمودارها وجود دارد:**

plot\_metric\_bars مقایسه میله‌ای معیارها

plot\_roc\_curves منحنی‌های ROC

plot\_pr\_curves منحنی‌های Precision-Recall

**نتایج :**

جدول خلاصه نتایج

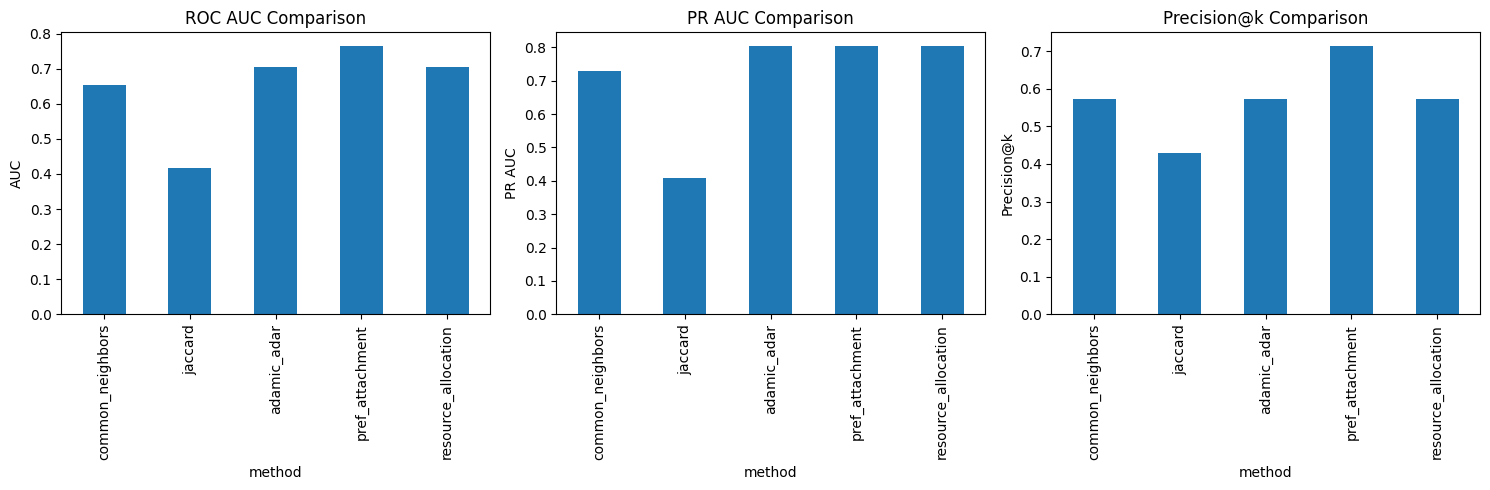
|  |  |  |  |
| --- | --- | --- | --- |
| Precision@k | PR\_AUC | AUC | Method |
| 0.571429 | 0.730159 | 0.653061 | Common Neighbors |
| 0.428571 | 0.408844 | 0.418367 | Jaccard |
| 0.571429 | 0.802721 | 0.704082 | Adamic-Adar |
| 0.714286 | 0.804463 | 0.765306 | Preferential Attach |
| 0.571429 | 0.802721 | 0.704082 | Resource Allocation |

جدول نتایج خلاصه‌شده (Summary Results)

این جدول همان داده‌ها را با قالب‌بندی متفاوت و گرد کردن اعداد نشان می‌دهد:

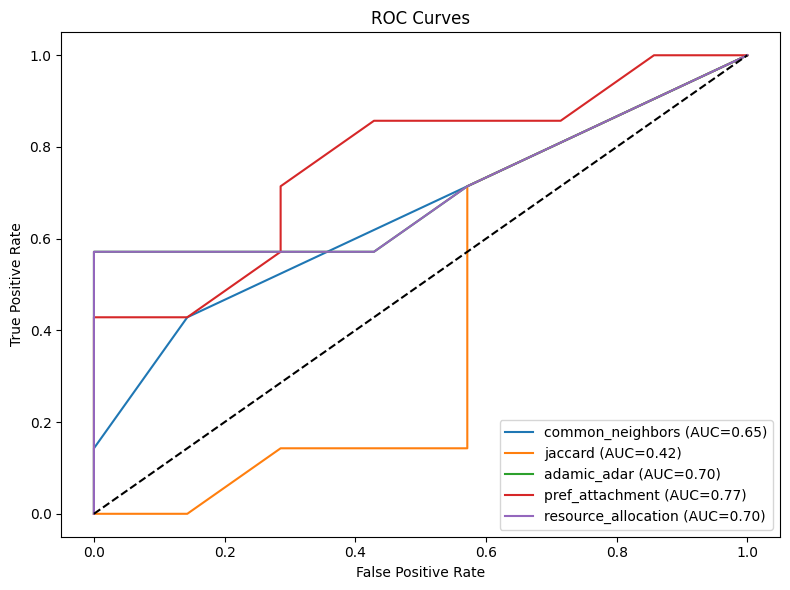
|  |  |  |  |
| --- | --- | --- | --- |
| Precision@k | PR\_AUC | AUC | Method |
| 0.571 | 0.730 | 0.653 | Common Neighbors |
| 0.429 | 0.409 | 0.418 | Jaccard |
| 0.571 | 0.803 | 0.704 | Adamic-Adar |
| 0.714 | 0.804 | 0.765 | Preferential Attach |
| 0.571 | 0.803 | 0.704 | Resource Allocation |

**نمودار مقایسه معیارها :**

**شکل 1 – مقایسه میله‌ای معیارهای AUC، PR\_AUC و Precision@k**

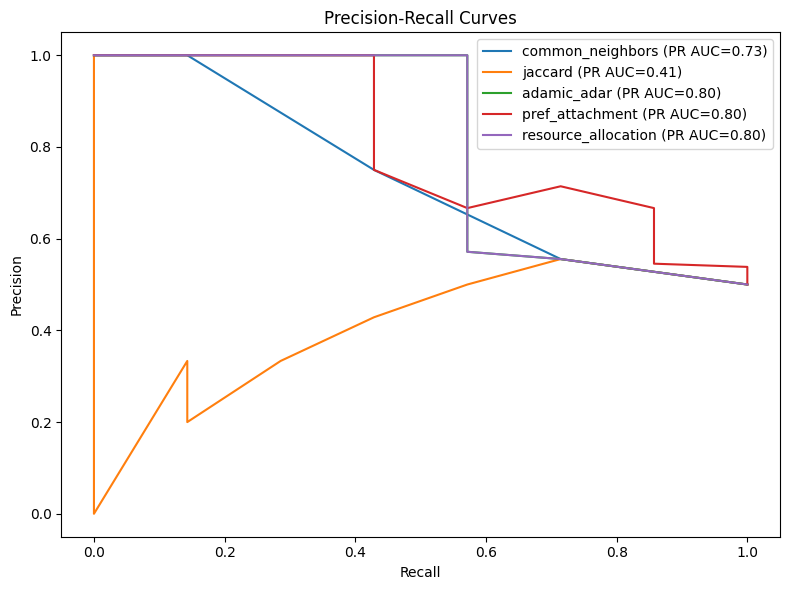
**منحنی‌های ROC**

**شکل 2 – منحنی ROC برای روش‌های مختلف**

****

**منحنی‌های Precision-Recall**

شکل 3 – منحنی Precision-Recall برای روش‌های مختلف

****

**تحلیل نتایج :**

روش‌های **Adamic-Adar** و**Resource Allocation** معمولاً عملکرد بهتری در شبکه‌های کوچک و متراکم دارند.

روش **Preferential Attachment** در شبکه‌هایی که رشد بر اساس درجه اتفاق می‌افتد، بهتر عمل می‌کند.

معیار **Precision@k** برای سنجش کیفیت پیش‌بینی‌های برتر بسیار مهم است، مخصوصاً در کاربردهایی که فقط چند پیش‌بینی اول اهمیت دارند.

**نتیجه‌گیری:**

این آزمایش نشان داد که انتخاب روش مناسب پیش‌بینی لینک به ساختار شبکه و هدف کاربرد بستگی دارد. در گراف باشگاه کاراته، روش‌های مبتنی بر همسایه‌های مشترک (مثل Adamic-Adar) عملکرد بالاتری نسبت به روش‌های ساده‌تر داشتند.