/ئ Overlapping community detection using expansion with contraction

روش‌های متعددی برای شناسایی جوامع غیرهم‌پوشان به سطح پیشرفته‌ای دست یافته‌اند. در سال‌های اخیر، برخی روش‌های شناسایی جوامع هم‌پوشان نیز پیشنهاد شده‌اند، اما این روش‌ها عموماً توانایی تنظیم میزان هم‌پوشانی جوامع در عین حفظ کیفیت شناسایی را ندارند. برای حل این مسئله، این مقاله یک روش جدید با عنوان \*\*روش انبساط و انقباض برای شناسایی جوامع هم‌پوشان (ECOCD)\*\* ارائه می‌دهد.

به طور خاص، روش ECOCD ابتدا جوامع غیرهم‌پوشان را با استفاده از \*\*تحلیل ماتریس غیرمنفی (NMF)\*\* به دست می‌آورد و سپس فرآیند انبساط و انقباض (شامل دو فرآیند انبساط و انقباض) را اجرا می‌کند. در هر مرحله از این فرآیند، یک جامعه به صورت تصادفی انتخاب شده و فرآیندهای انبساط و انقباض به صورت متوالی روی آن اجرا می‌شوند.

- \*\*فرآیند انبساط:\*\* گره‌ها را بر اساس \*\*درجه وابستگی\*\* جذب می‌کند.

- \*\*فرآیند انقباض:\*\* گره‌ها را بر اساس معیار \*\*پایداری (Permanence)\*\* حذف می‌کند.

علاوه بر این، پیچیدگی محاسباتی روش ECOCD به صورت نظری مورد تحلیل قرار گرفته است. این روش با امکان تنظیم میزان هم‌پوشانی، برای انواع مختلفی از شبکه‌ها با ویژگی‌های گوناگون قابل اعمال است و از کیفیت بالایی در شناسایی جوامع هم‌پوشان برخوردار است.

نتایج آزمایش‌های گسترده روی شبکه‌های مصنوعی و واقعی نشان می‌دهند که روش ECOCD نسبت به یازده روش پیشرفته دیگر در چهار معیار عملکرد برتری دارد و کارایی، اثربخشی و استحکام آن را تأیید می‌کند.

**خلاصه مقاله: شناسایی جوامع هم‌پوشان با استفاده از انبساط و انقباض**

روش‌های متعددی برای شناسایی جوامع غیرهم‌پوشان به سطح پیشرفته‌ای دست یافته‌اند. در سال‌های اخیر، برخی روش‌های شناسایی جوامع هم‌پوشان نیز پیشنهاد شده‌اند، اما این روش‌ها عموماً توانایی تنظیم میزان هم‌پوشانی جوامع در عین حفظ کیفیت شناسایی را ندارند. برای حل این مسئله، این مقاله یک روش جدید با عنوان \*\*روش انبساط و انقباض برای شناسایی جوامع هم‌پوشان (ECOCD)\*\* ارائه می‌دهد.

به طور خاص، روش ECOCD ابتدا جوامع غیرهم‌پوشان را با استفاده از \*\*تحلیل ماتریس غیرمنفی (NMF)\*\* به دست می‌آورد و سپس فرآیند انبساط و انقباض (شامل دو فرآیند انبساط و انقباض) را اجرا می‌کند. در هر مرحله از این فرآیند، یک جامعه به صورت تصادفی انتخاب شده و فرآیندهای انبساط و انقباض به صورت متوالی روی آن اجرا می‌شوند.

- \*\*فرآیند انبساط:\*\* گره‌ها را بر اساس \*\*درجه وابستگی\*\* جذب می‌کند.

- \*\*فرآیند انقباض:\*\* گره‌ها را بر اساس معیار \*\*پایداری (Permanence)\*\* حذف می‌کند.

علاوه بر این، پیچیدگی محاسباتی روش ECOCD به صورت نظری مورد تحلیل قرار گرفته است. این روش با امکان تنظیم میزان هم‌پوشانی، برای انواع مختلفی از شبکه‌ها با ویژگی‌های گوناگون قابل اعمال است و از کیفیت بالایی در شناسایی جوامع هم‌پوشان برخوردار است.

نتایج آزمایش‌های گسترده روی شبکه‌های مصنوعی و واقعی نشان می‌دهند که روش ECOCD نسبت به یازده روش پیشرفته دیگر در چهار معیار عملکرد برتری دارد و کارایی، اثربخشی و استحکام آن را تأیید می‌کند.

**بخش ۳: مبانی**

#### تعریف مسئله

در شبکه‌های اجتماعی، هر گره می‌تواند به چندین جامعه تعلق داشته باشد که این جوامع ممکن است دارای هم‌پوشانی باشند. مسئله شناسایی جوامع هم‌پوشان به معنای یافتن مجموعه‌ای از جوامع است که برخی گره‌ها می‌توانند در چند جامعه عضو شوند.

#### تحلیل ماتریس غیرمنفی (NMF)

\*\*NMF\*\* به‌عنوان یکی از ابزارهای اصلی این مقاله، ماتریس مجاورت شبکه \( A \) را به دو ماتریس غیرمنفی \( W \) و \( H \) تقسیم می‌کند به طوری که:

\[ A \approx WH \]

- \*\*\( W \):\*\* نشان‌دهنده وابستگی گره‌ها به جوامع مختلف است.

- \*\*\( H \):\*\* تعامل بین جوامع را مدل می‌کند.

برای اندازه‌گیری کیفیت فاکتورها، از حداقل‌سازی فاصله اقلیدسی استفاده شده است. الگوریتم \*\*NNDSVD\*\* برای مقداردهی اولیه ماتریس‌ها به کار گرفته شده که با کاهش باقی‌مانده اولیه، سرعت همگرایی را افزایش می‌دهد.

#### انتخاب بذر (Seed Selection)

در روش‌های کلاسیک، برای شروع فرآیند شناسایی، انتخاب بذرهای مناسب ضروری است. در این مقاله به جای انتخاب چند گره پراکنده به‌عنوان بذر، خوشه‌های اصلی به عنوان نقاط شروع انتخاب شده‌اند که باعث بهبود کیفیت جوامع اولیه می‌شود.

---

**بخش ۴: روش ECOCD**

روش ECOCD در دو مرحله اصلی طراحی شده است:

#### ۱. شناسایی جوامع اصلی (Core Communities)

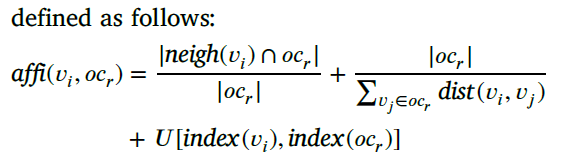
ابتدا ماتریس \( A \) با استفاده از \*\*NMF\*\* به دو ماتریس \( W \) و \( H \) فاکتورگیری می‌شود. سپس، هر گره به جامعه‌ای اختصاص می‌یابد که دارای بیشترین وابستگی به آن است. این جوامع اولیه به‌عنوان \*\*جوامع اصلی (Core Communities)\*\* شناخته می‌شوند.

- محدودیت: گره‌ها نمی‌توانند از جامعه اصلی خود حذف شوند، اما می‌توانند به جوامع دیگر اضافه شوند.

#### ۲. فرآیند انبساط و انقباض

پس از شناسایی جوامع اصلی، فرآیند انبساط و انقباض در چندین تکرار انجام می‌شود تا جوامع هم‌پوشان نهایی به دست آید.

##### (۱) فرآیند انبساط

در این فرآیند، گره‌هایی که بیشترین وابستگی را به جامعه دارند به آن اضافه می‌شوند. معیار وابستگی (\( affi \)) به صورت زیر تعریف شده است:

- \*\*ترم اول:\*\* شباهت جاکارد بین گره و جامعه.

- \*\*ترم دوم:\*\* معکوس میانگین فاصله کوتاه‌ترین مسیر از گره به اعضای جامعه.

- \*\*ترم سوم:\*\* قدرت ارتباط گره با جامعه بر اساس ماتریس \( U \) استخراج‌شده از NMF.

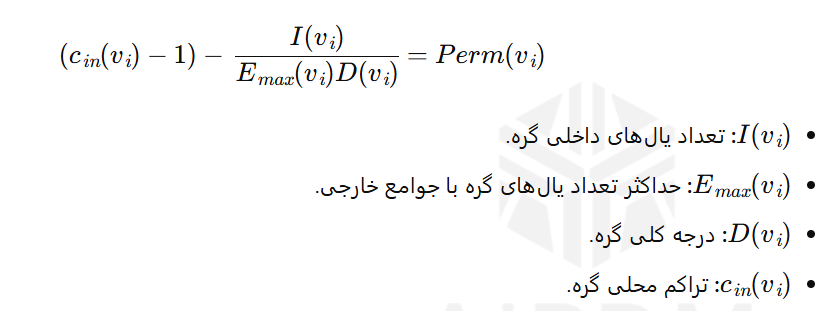
مراحل انبساط:

1. انتخاب مجموعه‌ای از گره‌های کاندید از همسایه‌های گره‌های موجود در جامعه.

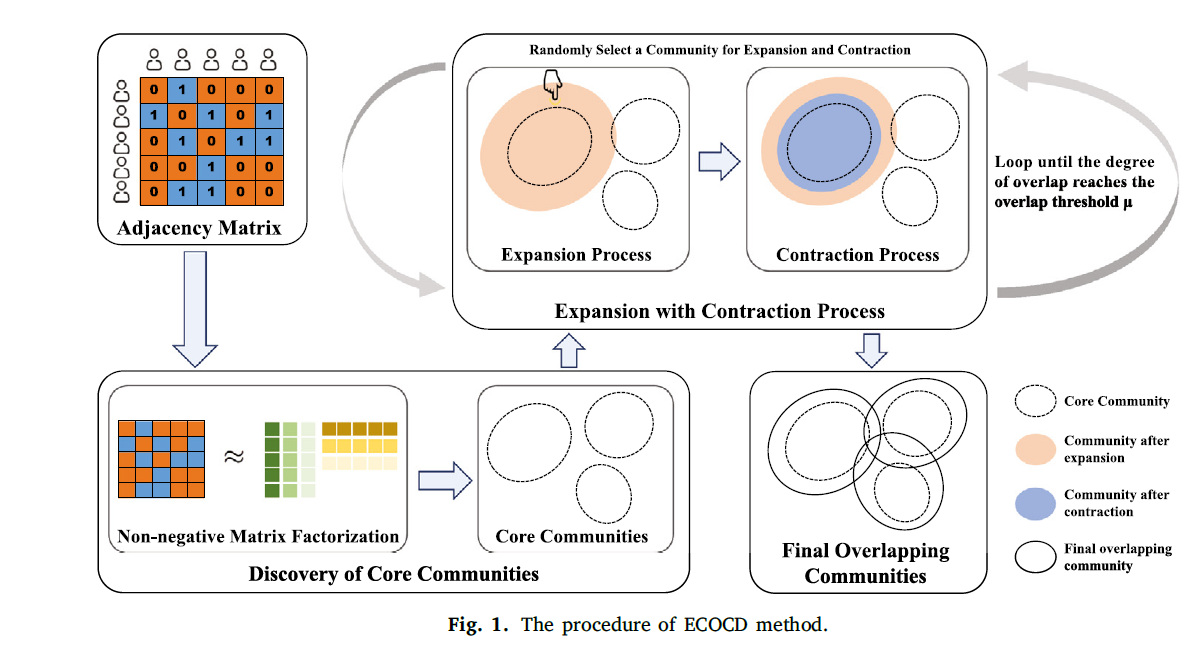
2. افزودن گره با بالاترین مقدار وابستگی به جامعه.

##### (۲) فرآیند انقباض

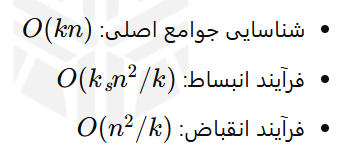
هدف از انقباض حذف گره‌هایی است که به طور ضعیفی به جامعه مرتبط هستند. معیار حذف (\( Perm \)) به صورت زیر تعریف می‌شود:



در هر تکرار، گره‌ای که کمترین مقدار \( Perm \) را دارد و در جامعه اصلی نیست، حذف می‌شود.



### تحلیل پیچیدگی زمانی

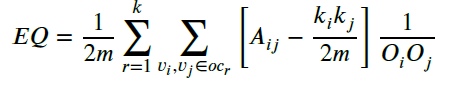


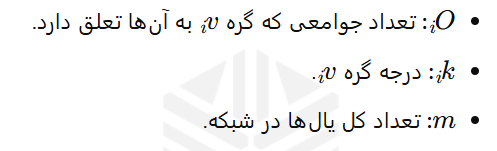
**### \*\*۵.۳.۲. معیارهای ارزیابی\*\***

در این مقاله، برای ارزیابی عملکرد روش ECOCD در شناسایی جوامع هم‌پوشان در شبکه‌های واقعی، از سه معیار اصلی استفاده شده است. این معیارها به بررسی کیفیت و دقت جوامع شناسایی‌شده می‌پردازند:

#### \*\*۱. ماژولاریتی هم‌پوشان گسترده (Extended Overlapping Modularity - EQ)\*\*

این معیار نسخه‌ای گسترش‌یافته از ماژولاریتی کلاسیک است که برای ارزیابی جوامع هم‌پوشان طراحی شده است. ماژولاریتی هم‌پوشان میزان تراکم یال‌ها درون جوامع در مقایسه با حالتی که یال‌ها به طور تصادفی توزیع شوند را اندازه‌گیری می‌کند. فرمول این معیار به شکل زیر است:

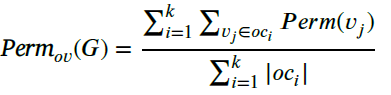




\*\*نکته:\*\* هرچه مقدار \( EQ \) بالاتر باشد، کیفیت جوامع شناسایی‌شده بهتر است.

#### \*\*۲. پایداری اصلاح‌شده (Modified Permanence - \( \text{Perm}\_{ov}(G) \))\*\*

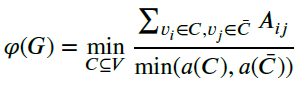
پایداری، میزان وابستگی یک گره به جامعه خود و توانایی شبکه در برابر اغتشاشات را اندازه‌گیری می‌کند. در نسخه اصلاح‌شده برای جوامع هم‌پوشان، تعریف پایداری به‌گونه‌ای گسترش یافته که شامل گره‌های مشترک بین جوامع باشد. فرمول این معیار به شکل زیر است:

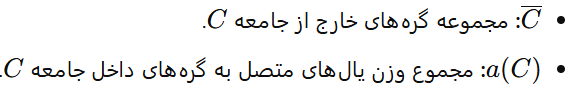


.

\*\*نکته:\*\* مقادیر بالاتر برای این معیار نشان‌دهنده استحکام بیشتر و وابستگی قوی‌تر گره‌ها به جوامع شناسایی‌شده است.

#### \*\*۳. رسانایی (Conductance - \( \phi(G) \))\*\*

رسانایی معیاری است که سرعت انتقال تصادفی از یک جامعه به گره‌های خارج از آن را اندازه‌گیری می‌کند. رسانایی پایین‌تر نشان‌دهنده انسجام بیشتر یک جامعه و تفکیک بهتر بین جوامع است. فرمول رسانایی به شکل زیر است:



\*\*نکته:\*\* مقدار کمتر رسانایی نشان‌دهنده کیفیت بالاتر جوامع شناسایی‌شده است.

**### نتایج و دستاوردها**

1. ECOCD توانایی تنظیم میزان هم‌پوشانی را با کیفیت بالا دارد.

2. این روش در شبکه‌های مصنوعی و واقعی عملکرد بهتری نسبت به ۱۱ الگوریتم پیشرفته دیگر داشته است.

3. ECOCD در شبکه‌های با هم‌پوشانی بالا (مانند LFR) عملکرد بهتری نسبت به الگوریتم‌های دیگر در معیارهای مانند ONMI نشان داده است.

Overlapping Community Detection Based on Weak Equiconcept

**خلاصه مقاله: شناسایی جوامع همپوشان بر اساس مفهوم‌های ضعیف**

این مقاله یک روش نوآورانه برای شناسایی جوامع همپوشان در شبکه‌های پیچیده ارائه می‌دهد که بر اساس **تحلیل مفهوم رسمی (FCA)** و استفاده از **مفهوم‌های ضعیف** است.

**اهداف و مشکلات اصلی:**

* شناسایی جوامع همپوشان، جایی که یک گره می‌تواند به چندین جامعه تعلق داشته باشد.
* رفع مشکلات مربوط به انتخاب دستی گره‌های آغازین (بذرها) و کاهش کیفیت نتایج در روش‌های موجود.
* بهبود کیفیت کشف جوامع حاشیه‌ای که در شبکه‌ها کمتر شناسایی می‌شوند.

**روش پیشنهادی:**

1. **انتخاب گره‌های بذر (Seed Selection):**
   * با استفاده از مفهوم‌های ضعیف (Weak Equiconcept) در بستر تحلیل مفهوم رسمی، بذرهایی با بیشترین نمایندگی و استحکام انتخاب می‌شوند.
   * این روش به طور خودکار، گره‌های کلیدی شبکه را شناسایی کرده و وابستگی به تنظیمات دستی را کاهش می‌دهد.
2. **گسترش جوامع (Community Expansion):**
   * یک الگوریتم بهبود‌یافته PageRank با در نظر گرفتن اطلاعات ساختاری و ویژگی‌های گره‌ها برای گسترش بذرها استفاده می‌شود.
   * اطلاعات مربوط به شباهت ویژگی‌های گره‌ها در روند گسترش در نظر گرفته می‌شود.
3. **بهینه‌سازی جوامع (Community Optimization):**
   * با استفاده از آنتروپی اطلاعات، همپوشانی میان جوامع کاهش داده می‌شود تا نتایج بهینه‌تری حاصل شود.

**یافته‌ها:**

* روش پیشنهادی، ساختارهای جامعه‌ای دقیق‌تر و باکیفیت‌تری نسبت به روش‌های موجود ارائه می‌دهد.
* در آزمایش‌های انجام‌شده روی داده‌های واقعی (مانند شبکه‌های اجتماعی و علمی)، الگوریتم جدید عملکرد بهتری در معیارهایی چون دقت (Precision)، فراخوانی (Recall) و نمره F1 داشت.

**نتیجه‌گیری:**

این تحقیق نشان داد که ترکیب اطلاعات ساختاری و ویژگی‌های گره‌ها با تحلیل مفهوم رسمی می‌تواند به شناسایی بهتر جوامع همپوشان کمک کند. از دیگر مزایای این روش، افزایش خودکارسازی فرآیند شناسایی و کاهش وابستگی به تنظیمات دستی است.

**پیشنهادات آینده:**

* توسعه روش‌هایی برای بهبود سرعت تولید مفهوم‌های ضعیف و محاسبه ضریب پایداری.
* بررسی شبکه‌های چندلایه و تعاملات پیچیده میان خوشه‌ها برای گسترش قابلیت‌های روش در تحلیل شبکه‌های پیچیده‌تر.