An Overlapping Community Detection Algorithm Based on Density Peaks

این مقاله یک الگوریتم جدید به نام \*\*OCDDP\*\* را برای شناسایی جوامع همپوشان در شبکه‌ها ارائه می‌کند که بر اساس \*\*قله‌های چگالی\*\* کار می‌کند. این الگوریتم به طور خاص برای شبکه‌هایی طراحی شده که گره‌ها در آن‌ها ممکن است به چند جامعه مختلف تعلق داشته باشند، مانند شبکه‌های اجتماعی و شبکه‌های پروتئین-پروتئین.

فصل دوم و سوم این مقاله بر روی روش خوشه‌بندی بر اساس قله‌های چگالی و جزئیات الگوریتم پیشنهادی \*\*OCDDP\*\* متمرکز هستند.

**### فصل دوم: خوشه‌بندی بر اساس قله‌های چگالی**

روش خوشه‌بندی مبتنی بر قله‌های چگالی، که ابتدا توسط \*\*Rodriguez\*\* و \*\*Laio\*\* در سال 2014 پیشنهاد شد، بر این فرض استوار است که مراکز خوشه‌ها نقاطی با \*\*چگالی محلی بالا\*\* هستند که از سایر نقاط با چگالی بالاتر فاصله زیادی دارند. این روش مراحل زیر را دنبال می‌کند:

1. \*\*محاسبه چگالی محلی\*\*: چگالی هر نقطه محاسبه می‌شود تا مشخص شود هر نقطه تا چه اندازه با سایر نقاط در نزدیکی خود ارتباط دارد. این چگالی با استفاده از یک تابع بر اساس فاصله نقاط و یک مقدار آستانه محاسبه می‌شود.

2. \*\*انتخاب مراکز خوشه‌ها\*\*: نقاطی که دارای چگالی بالا و فاصله زیاد از نقاط با چگالی بالاتر هستند، به عنوان مراکز خوشه‌ها در نظر گرفته می‌شوند.

3. \*\*انتساب سایر نقاط\*\*: سایر نقاط به نزدیک‌ترین مرکز خوشه‌ای که چگالی بالاتری دارد، تخصیص می‌یابند.

این روش معمولاً برای **تشخیص خوشه‌های غیر همپوشان** استفاده می‌شود. اما در این مقاله با تغییراتی برای تشخیص جوامع همپوشان در شبکه‌ها به‌کار گرفته شده است.

**### فصل سوم: الگوریتم پیشنهادی (OCDDP)**

فصل سوم به معرفی دقیق‌تر الگوریتم \*\*OCDDP\*\* پرداخته و مراحل آن را به تفصیل شرح می‌دهد. این الگوریتم شامل مراحل زیر است:

1. \*\*ساخت ماتریس فاصله\*\*:

- در این مرحله، از ماتریس مجاورت شبکه که ارتباطات میان گره‌ها را نشان می‌دهد، یک ماتریس فاصله ساخته می‌شود. برای محاسبه فاصله بین گره‌ها، قدرت نسبی و مطلق لینک‌ها در نظر گرفته می‌شود تا تفاوت میان گره‌هایی که به طور مستقیم یا غیرمستقیم به یکدیگر متصل هستند، دقیق‌تر تعیین شود.

2. \*\*محاسبه و نرمال‌سازی چگالی محلی (ρ) و فاصله (δ)\*\*:

- چگالی محلی هر گره (ρ) و فاصله آن از نزدیک‌ترین گره با چگالی بالاتر (δ) محاسبه می‌شود. سپس، این مقادیر برای تشخیص هسته‌های جوامع نرمال‌سازی می‌شوند تا در شناسایی گره‌های مرکزی و گره‌های حاشیه‌ای در جوامع همپوشان دقت بیشتری حاصل شود.

3. \*\*انتخاب هسته‌های جامعه\*\*:

- گره‌هایی که چگالی و فاصله زیادی دارند، به عنوان هسته‌های جامعه انتخاب می‌شوند. همچنین، اگر یک گره فاصله زیادی از نزدیک‌ترین گره با چگالی بالاتر داشته باشد، آن نیز می‌تواند به عنوان یک هسته جامعه در نظر گرفته شود.

4. \*\*تخصیص گره‌ها به جوامع با استفاده از بردارهای عضویت\*\*:

- برای هر گره، یک بردار عضویت تعریف می‌شود که نشان‌دهنده احتمال تعلق آن به هر یک از جوامع است. این بردار بر اساس نزدیکی گره به هسته‌ها و چگالی همسایگان محاسبه می‌شود. هر گره می‌تواند به بیش از یک جامعه تعلق داشته باشد و با توجه به بردار عضویت، تخصیص آن به جوامع به‌صورت همپوشان صورت می‌گیرد.

### پیچیدگی زمانی الگوریتم

این الگوریتم از نظر زمانی بهینه است و در شرایط بدترین حالت دارای پیچیدگی زمانی \(O(n(m + n))\) است، که در آن \(n\) تعداد گره‌ها و \(m\) تعداد یال‌ها هستند.

**\*\*یافته‌های اصلی\*\*:**

این الگوریتم بر روی شبکه‌های مصنوعی و شبکه‌های اجتماعی واقعی آزمایش شده و نتایج نشان می‌دهد که \*\*OCDDP\*\* در شناسایی جوامع همپوشان کارآمدتر و پایدارتر از برخی روش‌های موجود است. به ویژه در شبکه‌های با ساختار ساده، عملکرد آن بهتر از سایر الگوریتم‌هاست و در شبکه‌های پیچیده نیز قابل مقایسه با الگوریتم‌های پیشرفته‌ای مانند \*\*SLPA\*\* و \*\*SpeakEasy\*\* است.

**\*\*نتیجه‌گیری\*\*:**

OCDDP به عنوان یک الگوریتم موثر در تشخیص جوامع همپوشان در شبکه‌ها معرفی شده و برای شبکه‌های اجتماعی واقعی از سایر روش‌های مقایسه شده بهتر عمل می‌کند.

"Density-Peak-Based Overlapping Community Detection Algorithm:

### خلاصه مقاله: الگوریتم تشخیص جوامع همپوشان مبتنی بر قله‌های چگالی (DPOCD)

#### 1. \*\*مقدمه و اهمیت موضوع\*\*

تشخیص جوامع همپوشان در شبکه‌های پیچیده یکی از موضوعات کلیدی در تحلیل شبکه‌ها است. جوامع همپوشان شامل گره‌هایی هستند که می‌توانند به چندین جامعه تعلق داشته باشند. این ساختارها در بسیاری از شبکه‌های اجتماعی، زیستی، و اطلاعاتی مشاهده می‌شوند. الگوریتم‌های سنتی مانند \*\*DPC\*\* (خوشه‌بندی با قله‌های چگالی) برای تشخیص جوامع غیرهمپوشان طراحی شده‌اند و محدودیت‌هایی در تشخیص جوامع همپوشان دارند. مقاله حاضر، الگوریتم \*\*DPOCD\*\* را معرفی می‌کند که با بهبود روش DPC، این محدودیت‌ها را برطرف کرده است.

---

#### 2. \*\*الگوریتم DPC و محدودیت‌های آن\*\*

\*\*DPC\*\* بر اساس دو ویژگی اصلی طراحی شده است:

1. مراکز خوشه‌ها دارای چگالی بالاتر نسبت به گره‌های اطراف هستند.

2. این مراکز فاصله بیشتری از سایر مراکز خوشه دارند.

گرچه این الگوریتم کارآمد است، اما دارای معایب زیر است:

- \*\*عدم توانایی در تشخیص جوامع همپوشان:\*\* این الگوریتم گره‌ها را فقط به یک خوشه اختصاص می‌دهد.

- \*\*انتخاب دستی مراکز خوشه:\*\* نیاز به تعیین دستی مراکز خوشه‌ها دارد که در داده‌های بزرگ غیرعملی است.

---

#### 3. \*\*ویژگی‌ها و نوآوری‌های الگوریتم DPOCD\*\*

الگوریتم DPOCD با هدف رفع محدودیت‌های DPC طراحی شده و نوآوری‌های زیر را ارائه می‌دهد:

##### 3.1. \*\*تعریف قدرت ارتباط گره و لبه:\*\*

- \*\*قدرت ارتباط گره (Point Link Strength):\*\* تأثیر روابط گره‌ها بر اساس تعداد گره‌های مشترک و درجه گره.

- \*\*قدرت ارتباط لبه (Edge Link Strength):\*\* تأثیر قدرت لبه‌ها در ارتباط بین گره‌ها.

این تعاریف برای محاسبه ماتریس فاصله استفاده می‌شوند، که اطلاعات دقیق‌تری درباره‌ی روابط میان گره‌ها ارائه می‌دهند.

##### 3.2. \*\*انتخاب خودکار مراکز جوامع با برازش خطی:\*\*

به جای انتخاب دستی، این الگوریتم از برازش خطی برای تعیین مراکز جامعه به صورت تطبیقی استفاده می‌کند. این کار باعث حذف تأثیرات ذهنی شده و انتخاب مراکز را برای داده‌های بزرگ بهینه می‌کند.

##### 3.3. \*\*استفاده از بردار احتمالات برای تعلق گره‌ها:\*\*

در DPOCD، هر گره با استفاده از یک بردار احتمالاتی نشان داده می‌شود که میزان تعلق آن به جوامع مختلف را تعیین می‌کند. این تعلق بر اساس چگالی محلی و فاصله گره‌ها محاسبه می‌شود.

---

#### 4. \*\*مزایا و نتایج\*\*

الگوریتم DPOCD در مقایسه با روش‌های قبلی دارای مزایای زیر است:

- \*\*تشخیص جوامع همپوشان:\*\* این الگوریتم به‌طور دقیق گره‌هایی را که در چندین جامعه عضو هستند شناسایی می‌کند.

- \*\*کارایی بالا:\*\* انتخاب خودکار مراکز و استفاده از معیارهای تطبیقی، این الگوریتم را برای داده‌های بزرگ مناسب کرده است.

- \*\*انعطاف‌پذیری در شکل جوامع:\*\* امکان تشخیص جوامع با اشکال غیرکروی.

---

#### 5. \*\*نتیجه‌گیری و کاربردها\*\*

الگوریتم DPOCD گامی مؤثر در بهبود روش‌های تشخیص جوامع همپوشان است. این روش می‌تواند در شبکه‌های اجتماعی، زیستی، و تحلیل داده‌های پیچیده کاربرد داشته باشد. با استفاده از مفاهیمی مانند قدرت ارتباط، برازش خطی، و بردار احتمالاتی، این الگوریتم دقت و انعطاف‌پذیری بیشتری ارائه می‌دهد و محدودیت‌های الگوریتم‌های سنتی را رفع می‌کند.

---

#### \*\*منبع مقاله:\*\*

Liping Sun, Tao Ye, Xiaoyu Duan, and Yonglong Luo. "Density-Peak-Based Overlapping Community Detection Algorithm." \*IEEE Transactions on Computational Social Systems,\* vol. 9, no. 1, pp. 88-100, 2021.