تشخیص ناهنجاری در گراف‌های پویا

دانشجو: امیرحسین عبیری

استاد راهنما: دکتر بهنام بهرک

دانشکده مهندسی برق و کامپیوتر، دانشگاه تهران

خلاصه

تشخیص ناهنجاری و رفتارهای نامتعارف در داده‌ها از مباحث بسیار پرکاربرد علوم داده است. یکی از شاخه‌های این مساله تشخیص ناهنجاری در داده‌های شبکه‌های اجتماعی و به طور کلی هر نوع داده‌ای است که بتوان آن را توسط گراف مدل‌سازی کرد. ناهنجاری در یک شبکه بانکی می‌تواند نشانگر پول‌شویی و در یک شبکه‌ی اجتماعی می‌تواند نشانگر فعالیت‌های مخرب باشد. از این رو تشخیص آن اهمیت بسیار بالایی دارد. در این تحقیق به تعریف ناهنجاری، بررسی کلی روشهای تشخیص ناهنجاری و طبقه‌بندی آن‌ها پرداخته و سپس روشی برای بهبود یک الگوریتم تشخیص ناهنجاری ارائه شده است.

ساختار

در این پروژه ضمن دسته‌بندی الگوریتم‌های تشخیص ناهنجاری گراف به پنج دسته‌ی کلی، دو الگوریتم شاخص این حوزه پیاده‌سازی شده‌است. دسته‌بندی‌ها عبارتند از:

- روش‌های مبتنی بر شناسایی جامعه
- روش‌های مبتنی بر فشرده‌سازی
- روش‌های مبتنی بر تجزیه‌ی ماتریس
- روش‌های مبتنی بر سنجش فاصله
- روش‌های مبتنی بر احتمالات

در ادامه دو الگوریتم شاخص Oddball و Autopart پیاده‌سازی و تحلیل شد. با توجه به این که الگوریتم Autopart مبتنی بر فشرده‌سازی است، تنها بر روی گراف‌های بدون وزن کار می‌کند. با ایجاد تغییراتی در محاسبه‌ی هزینه‌ی این الگوریتم، پارامتر وزن یال را به الگوریتم اضافه شد و الگوریتم برای گراف‌های وزن‌دار بهینه شد.

این مهم با تغییر در تعریف پارامتر چگالی انجام شد که در الگوریتم به این صورت تعریف می‌شد:

$$P_{ij} = \frac{w_{ij}}{n_{ij}}$$

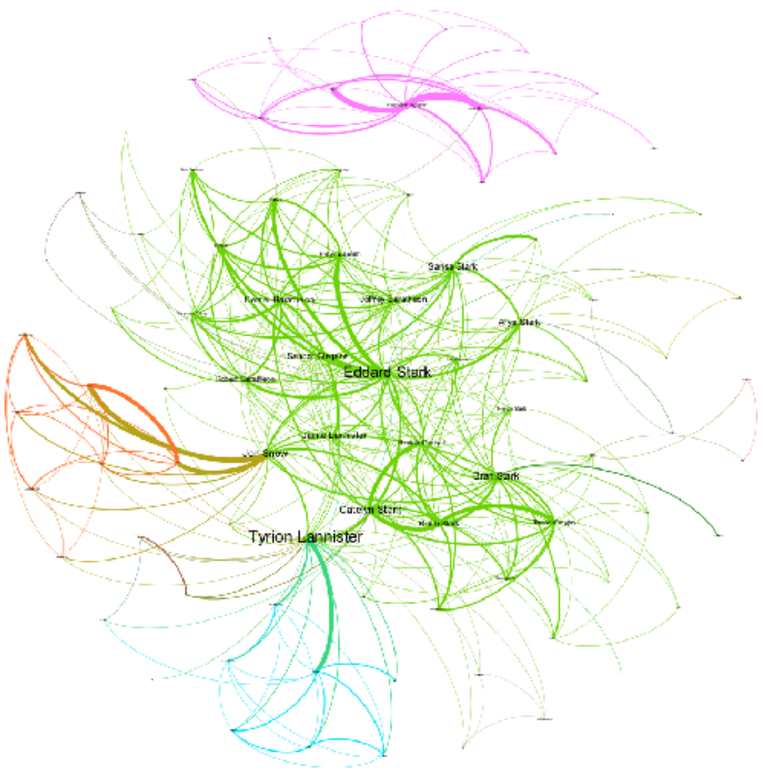
با تغییر در تعریف این پارامتر به صورت زیر الگوریتم Autopart بر روی گراف‌های وزن‌دار قابل پیاده‌سازی شد. و نتایج بررسی شد.

$$P_{ij} = \frac{sum\ of\ edge\ weights\ with\ weight\ more\ than\ threshold}{sum\ of\ all\ edge\ weights}$$

نتایج

با تغییر الگوریتم Autopart برای بهبود کارکرد آن در گراف‌های وزن‌دار، الگوریتم جدید را بر روی گراف‌های موجود اجرا کردیم. نتیجه‌ی تشخیص ناهنجاری و خوشه‌بندی گراف‌ها در مقایسه با سایر الگوریتم‌های موجود در این زمینه قابل دفاع بود. یک نمونه از اجرای الگوریتم خوشه‌بندی بر روی گراف شخصیت‌های یک سریال معروف در ذیل آورده شده است.

نمونه‌ی خوشه‌بندی انجام شده بر روی گراف شخصیت‌های سریال Game of Thrones



جمع بندی

با توجه به اهمیت تشخیص رویدادهای مهم در مجموعه داده‌ها و همچنین تشخیص داده‌های دارای رفتار ناهنجار، در این پروژه دو الگوریتم شاخص تشخیص ناهنجاری پیاده‌سازی و بر روی مجموعه داده‌های مختلف اجرا شد و نتیجه مورد بررسی قرار گرفت. همچنین با ایجاد تغییراتی در الگوریتم مبتنی بر فشرده‌سازی، الگوریتم مورد بهبود قرار گرفت.

کاربرد های صنعتی:

الگوریتم‌های تشخیص ناهنجاری می‌توانند در تشخیص پولشویی‌های بانکی، تشخیص فعالیت‌های غیرمعمول در شبکه‌های اجتماعی و تشخیص خرید و فروش‌های نامعمول در پلتفرم‌های فروش آنلاین به کار روند.

مراجع اصلی

1. Ranshous, S., Shen, S., Koutra, D., Harenberg, S., Faloutsos, C. and Samatova, N. (2015). Anomaly detection in dynamic networks: a survey. *Wiley Interdisciplinary Reviews: Computational Statistics*, 7(3), pp.223-247.
2. Chakrabarti, D. (2004). AutoPart: Parameter-Free Graph Partitioning and Outlier Detection. *Lecture Notes in Computer Science*, pp. 112-124.
3. Akoglu, L., McGlohon, M. and Faloutsos, C. (2018). *oddball: Spotting Anomalies in Weighted Graphs*.