

بسمه تعالی

پروپوزال پژوهشی درس گراف کاوی

تاریخ تحویل: ۱۴۰۴/۰۹/۱۹

درس: گراف کاوی

استاد: دکتر زینب ملکی

اطلاعات دانشجو:

نام و نام خانوادگی: عرفان هاشمی

شماره دانشجویی: ۱۹۱۳۰۲۰۴

ایمیل: hashemierfa@gmail.com

عنوان پژوهش:

تحلیل توزیع درجه در دیتاست Amazon MovieLens Ecommerce Graph

چکیده:

در این پژوهش، هدف بررسی ساختار شبکه‌ی تعاملات کاربران و فیلم‌ها در دیتاست Amazon MovieLens است. با مطالعه‌ی الگوی درجه، می‌توان ویژگی‌هایی مانند مقیاس‌پذیری، وجود گره‌های مهم و رفتار شبکه در برابر حذف یا اضافه شدن یال‌ها را تحلیل کرد. این تحلیل به درک بهتر ساختار گراف و رفتار کاربران و محصولات کمک می‌کند.

مسئله و انگیزه:

توزیع درجه یکی از بنیادی‌ترین ویژگی‌های هر گراف است و نقش مهمی در تحلیل ساختار شبکه‌های واقعی دارد. بسیاری از شبکه‌های تجارت الکترونیک رفتار Scale-Free دارند و تعداد کمی از گره‌ها دارای درجه‌ی بسیار بالا هستند.

تحلیل این ویژگی‌ها در دیتاست یادشده می‌تواند به درک بهتر الگوهای تعامل کاربران و محبوبیت محصولات کمک کند. همچنین این تحلیل برای کاربردهای مختلفی مانند پیشنهاددهنده‌ها و بهینه‌سازی شبکه مفید است.

اهداف:

استخراج و تمیزسازی دیتاست و ساخت گراف نهایی

محاسبه درجه هر گره و تحلیل آماری آن

رسم نمودارهای توزیع درجه‌ی (Degree Distribution) و (Log-Log Plot)

کارهای مرتبط:

مقالات مختلف نشان داده‌اند که شبکه‌های واقعی مانند شبکه‌های تجاری، اجتماعی و بیولوژیک دارای توزیع درجه‌ی قانون توانی (Power Law) هستند. مقالات آقای باراباشی در زمینه شبکه‌های Scale-Free با این پژوهه مرتبط است و از یافته‌های این مقالات در این پژوهه بهره گرفته می‌شود. این پژوهه با تمرکز بر دیتاست MovieLens این مفاهیم را به صورت عملی بررسی می‌کند.

روش‌بیشنهادی:

دیتاست: Amazon MovieLens شامل کاربران، محصولات و یال‌هایی که رفتارهای خرید یا امتیازدهی را نشان می‌دهند.

– نوع گراف: دوبخشی

– تعداد تقریبی گره‌ها:

– کاربران: ۶ تا ۱۰ هزار گره

– فیلم‌ها: ۴ تا ۶ هزار گره

– تعداد یال‌ها: ۱۰۰ تا ۲۰۰ هزار یال

– پیش‌پردازش: حذف گره‌های تکراری، ساخت گراف دوتایی (User-Item).

تکنیک‌ها و الگوریتم‌ها:

- محاسبه درجه

- رسم هیستوگرام توزیع درجه

- استفاده از Python و NetworkX

طرح ارزیابی:

– مقایسه توزیع درجه کاربران و فیلم‌ها

– بررسی وجود الگوی Long-Tail یا Scale-Free

– استخراج رفتار کاربران خاص (بسیار فعال یا بسیار غیرفعال)

چالش‌ها:

– اندازه بزرگ دیتاست ممکن است پردازش را زمان‌بر کند.

– نیاز به پاکسازی دقیق داده‌ها و بررسی یال‌های غیرمعتبر داریم.

منابع:

- 1: Newman, M. (2010). *Networks: An Introduction*. Oxford University Press.
- 2: Barabási, A.-L. (1999). Emergence of scaling in random networks. *Science*.
- 3: Barabási, A.-L. (2016). *Network Science*. Cambridge University Press.
- 4: Yang, J., McAuley, J., & Leskovec, J. (2013). Modeling User Interaction in Online Recommendation Systems.

Student Signature: ErfanHashemi_12.11.2025