

## بسمه تعالی

### پروپوزال پروژه‌ی 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: ۴۰۲۰۱۹۱۳

\_\_ ایمیل: hashemierfa@gmail.com

### عنوان پروژه :

\_\_ تحلیل توزیع درجه در دیتاست Amazon MovieLens Ecommerce Graph

### چکیده:

در این پروژه، 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 پروژه مرتبط است و از یافته‌های این مقالات در این پروژه بهره گرفته می‌شود. این پروژه با تمرکز بر دیتاست MovieLens این مفاهیم را به‌صورت عملی بررسی می‌کند.

## روش پیشنهادی:

\_\_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 NetworkX و Python

## 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.

- 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