

تمرین ۳ درس شبکه‌های اجتماعی (گروه ۲)

دی ماه ۱۴۰۰

مهلت تحویل یک هفته

در این تمرین شما باید با استفاده از نرم افزار Gephi و ابزارهای آن نودهای گراف ضمیمه را تحلیل نمایید.

روش جمع آوری داده

گراف ضمیمه حاصل گردآوری اطلاعات واقعی از کانالهای تلگرامی در طول مدت یک ماه اخیر است. این کانالها در مورد شرط بندی مطلب یا مطالبی منتشر کرده اند. جهت اطلاع شما و برای اینکه تحلیل دقیقتری از نودهای گراف داشته باشید، کوئری جستجوی مطالب در کانالهای تلگرامی به صورت زیر بوده است:

- یا باید یکی از کلمات/عبارات زیر در مطلب آمده باشد (یعنی or عبارات زیر):

bet|شرط بندی|شرط بندی|مسابقه پوکر|بازی پوکر|پوکر بازی|کازینو|کازینوی آنلاین|پیشن پیش بینی|ضرایب شگفت انگیز|ضرایب بالا|بازی رولت|بازی جذاب رولت|بازی انفجار|بازی جذاب انفجار|فوربت|من و تو بت|هزار بت|هزار بت|نکس بت|نکسبت|بت رایون|بت رایون|ولف بت|بت فوروارد|بت فوروارد|فوروارد بت|فوروارد بت|ورزش|بت فوت|آریان بت|امل بت|کانون بت|کانن بت|وان کیک بت|وان ایکس بت|آی آر بت|ددی بت|بت وب|بت پلاس|بت ۹۰|بت پی|رومبت|روما بت|

- یا این که کلمه "بت" همراه با یکی از کلمات زیر آمده باشد (and کلمه بت همراه با or یکی از عبارات زیر):

مسابقه|مسابقات|شرط بازی|جایزه|جوایز|رولت|اسنوکر|انفجار|پوکر|پیشن|ضریب|ضرایب|لیگ|درآمد|سرگرمی|پرفکت مانی|برد

نکته مهمی که باید به آن توجه بفرمایید این است که اگر مطلبی از یک فرد در کانالی فوروارد شود و این فرد در تنظیمات حریم خصوصی خود به تلگرام اجازه نداده باشد که هویت او در هنگام فوروارد برای دریافت کنندگان مطلب مشخص شود، تلگرام آن فرد را با عنوان فرستنده مخفی (HiddenSender) معرفی میکند. لذا کلیه افرادی که مطلبی از ایشان در کانالها فوروارد شده، سرجمع، با یک نود HiddenSender در فایل ضمیمه ذکر شده اند. لطفا در تحلیلهای خود به این نکته توجه بفرمایید. در صورتی که به این نتیجه رسیدید که حضور این نود باعث اشتباه و بهم خوردن تحلیلهای میشود میتوانید آن را از جمع نودها حذف کنید.

ساختار گراف ضمیمه

اگر فایل ضمیمه را - که در قالب فایل‌های gdf ، یکی از قالب‌های معمول بیان گراف، ذخیره شده- را با یک ویرایشگر متنی باز کنید مشاهده میکنید که از دو قسمت اصلی تشکیل شده:

۱. Nodedef که نودهای گراف و ستونهای مربوط به ویژگیهای مختلف آنها را تعریف میکند و

۲. Edgedef که یالهای گراف را در قالب ستونهایی از ویژگیها معرفی میکند.

همه ستونهای هر بخش با کاما از هم جدا شده اند. قسمت تعریف نودها با عبارت زیر آغاز شده و در آن ۴ ستون بچشم میخورد:

`nodedef>name , label VARCHAR, count INT, views INT`

ستون اول شناسه عددی یکتای کانالهاست. سپس نام هر کانال در ستون label آمده و شما میتوانید محتوای این کانالها را با تایپ نام آن در هر اپ تلگرامی ببینید. ستون count تعداد مطالبی که هر کانال با موضوع شرط بندی منتشر کرده (اعم از مطالبی که خود کانال منتشر کرده یا از کانالهای دیگر فوروارد کرده) را نشان میدهد. ستون views هم جمع بازدید مطالبی که در موضوع شرط بندی در یک کانال منتشر شده را نشان میدهد (توجه کنید که در اینجا مشابه با رویه تلگرام، بازدید مطالب فورواردی به حساب کانالی که مطلب اصلی را منتشر کرده گذاشته میشود).

قسمت تعریف یالها با سطر زیر آغاز میشود:

`edgedef>node1 , node2 , directed BOOLEAN, weight DOUBLE`

ستون اول نود مبدا یال و ستون دوم نود مقصد یال است. ستون directed مشخص میکند که یال معرفی شده جهت دار است یا نه (در فایل ضمیمه همه یالها جهت دار هستند). ستون وزن هم تعداد دفعاتی که نود مبدا از نود مقصد مطلبی در کانال خود فوروارد کرده نشان میدهد. برای اینکه بتوانید یالها را بهتر تفسیر کنید توجه داشته باشید که جهت یال از کانال فوروارد کننده مطلب (یعنی کانال ارجاع دهنده) به کانال نویسنده اصلی مطلب (یعنی کانال مرجع) است.

سوال ۱ تمرین:

الف: توپولوژی گراف را بررسی کنید و با رسم نمودار مشخص کنید توزیع درجه نودها از چه توزیعی تبعیت میکند. شما باید پارامترهای توزیع را نیز به صورت دقیق محاسبه کنید.

ب و ج: همین کار را برای توزیع تعداد مطالب و جمع بازدید کانالها نیز انجام دهید.

برای اطمینان از درستی رویه انجام کار حتما بخش **Advanced Topic 3.B-Plotting Power-laws** از کتاب را مطالعه نمایید.

سوال ۲ تمرین:

برای این کار شما باید معیارهای مرکزیت را با استفاده از ابزارهای موجود در Gephi یا با کمک گرفتن از ابزارهای دیگر محاسبه کرده و ۱۰ نود برتر هر معیار را بیابید. سپس تفسیر کنید دلیل برتر بودن آنها در آن معیار چه بوده است. توجه داشته باشید که ممکن است برای تفسیر بهتر لازم باشد محتوای مطالب این کانالها را نیز بررسی نمایید. معیارهای مرکزیتی که باید حساب کنید:

۱- بینابینی (betweenness)

۲- نزدیکی (closeness)

۳- کارایی (efficiency) یا نزدیکی هارمونیک

۴- نامرکزیت (eccentricity): توجه داشته باشید که اعدادی که gephi برای معیار نامرکزیت حساب میکند موقع محاسبه نودهای مرکز باید عکس شوند.

۵- بردار ویژه (Eigenvector)

۶- HIT

۷- رتبه صفحه (Page Rank)

سوال ۳ تمرین:

پلیس فتا میخواهد کانالهایی را که در موضوع شرط بندی باعث کلاهبرداری و خروج پول از کشور میشوند را شناسایی کند. حال نکته اینجاست که تعداد چنین کانالهایی کم نیست و بررسی همه این کانالها و کنترل آنها به وقت زیادی احتیاج دارد. معیار مرکزیتی پیشنهاد دهید که با ترکیب شاخصها و اطلاعات مختلف بتواند کانالها را به ترتیبی متناسب با احتمال تاثیرگذاری بر روی کاربران مرتب کند.