In the name of God



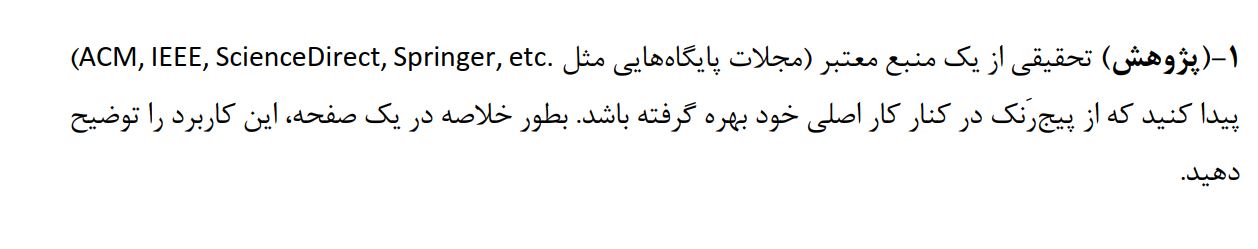
استاد : دکتر هراتی

دانشجو : توحید حقیقی سیس

شماره دانشجویی : 830598021

موضوع : تمرین چهارم

**تمرین اول :**

****

Centrality ranking in multiplex networks using topologically biased

random walks

Cangfeng Dinga,b

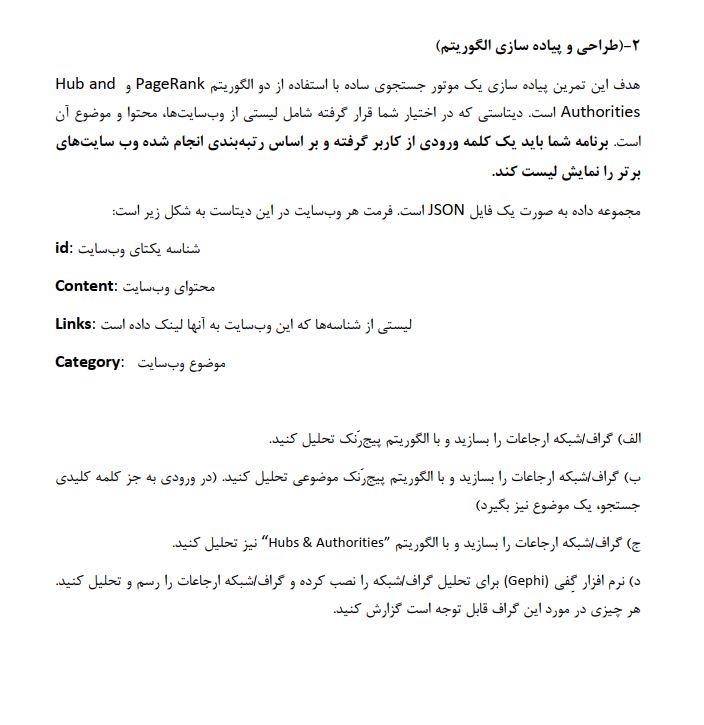
, Kan Lia,∗

مقاله بالا در مورد مرکزیت ها و محاسبه مرکزیت ها به وسیله الگوریتم page rank است در این الگوریتم بر اساس اهمیت پیج ها و hub بودن برخی نود ها .

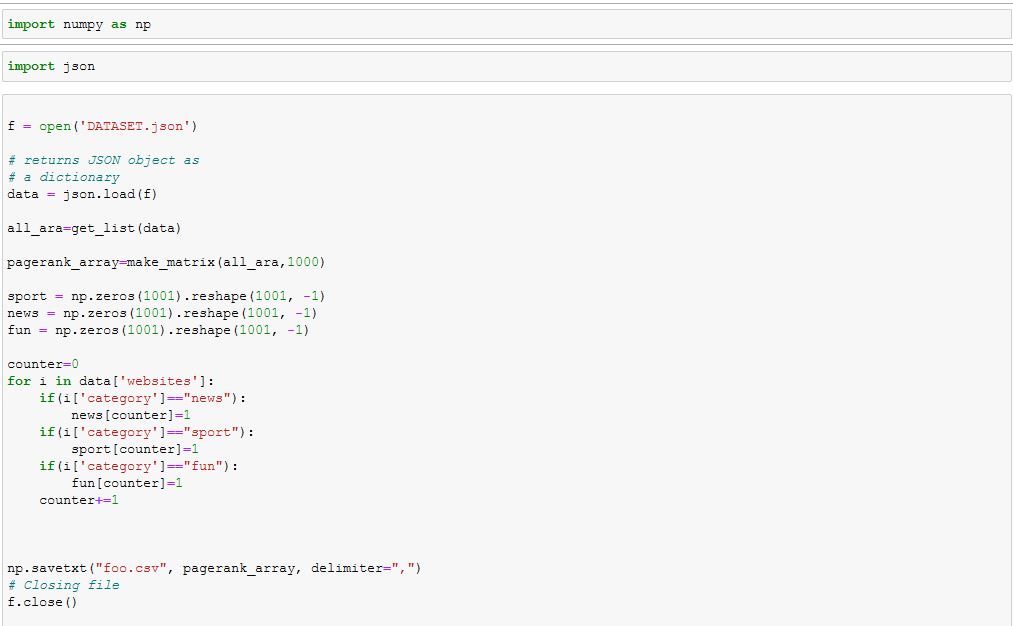
مشخص کردن مرکزیت آماری قابل توجه گره ها یکی از اهداف اصلی شبکه های مالتی پلکس است. با این حال ، اقدامات مرکزیت فعلی برای رتبه بندی گره ها فقط روی پیاده روی های تصادفی یا ساختار توپولوژیکی شبکه متمرکز هستند .

یک چالش اساسی نحوه سنجش مرکزیت گره ها در شبکه های مالتی پلکس است ، هم به توپولوژی شبکه و هم به انواع مغرضانه پیاده روی های تصادفی ، مانند پیاده روی های مغرضانه که با خواص هر گره به طور جداگانه در هر لایه انجام می شود یا پیاده روی های مغرضانه بستگی دارد. در عوض یک یا حتی بیشتر از ویژگیهای چندگانه ذاتی گره ورود است. در این مقاله ، با در نظر گرفتن این دو جنبه ، ما یک چارچوب ریاضی مبتنی بر پیاده روی تصادفی مغرضانه را پیشنهاد می کنیم ، به نام Multiplex چند جانبه Topologically PageRank ، که به شما امکان می دهد مرکزیت را محاسبه کرده و بر این اساس گره ها را در شبکه های مالتی پلکس رتبه بندی کنید. به طور خاص ، بسته به ماهیت تعصبات و برهم کنش گره ها بین لایه های مختلف ، موارد افزودنی ، ضرب و ترکیبی از Multiplex موضعی را با عنوان PageRank متمایز می کنیم. هر مورد با تنظیم پارامترهای بایاس ، نشان می دهد که چگونه درجه بندی مرکزی گره در یک لایه بر رتبه بندی که ماکت آن در لایه های دیگر به دست می آورد ، تأثیر می گذارد و میزان بازدید واکرها به طور ترجیحی از هاب ها یا گره های ضعیف متصل را ضبط می کند. آزمایشات انجام شده بر روی دو شبکه مالتی پلکس در دنیای واقعی نشان می دهد که مالتی پلکس چند جانبه ای که از نظر توپولوژیکی عمل می کند ، هم از نظر مورد بی طرفانه مربوطه و هم از روش های رتبه بندی فعلی بهتر عمل می کند و می تواند گره هایی را که در شبکه های مالتی پلکس به طور قابل توجهی در بالاترین رتبه قرار دارند .

**تمرین دوم :**

****

برای این سوال اول من اومدم از روی فایل های json ماتریس adjency matrix رو تولد کردم و همچین اومدم 3 تا وکتور مربوط به fun و sport و news رو هم ایجاد کردم تا در بخش دوم بتوانم از اون وکتور ها استفاده کنم .



همان طور که در شکل بالا میبینید از کتابخانه json برای کار با داده جسون استفاده کردم جیسون را خوانده و به یک تابع که در زیر هست پاس دادم تا ان را تبدیل به ارایه 2 بعدی کند و در مراحل بعدی بتوانم از ان استفاده کنم .



همان طور که از فایل جسون معلوم است ساختار به صورت درختی در زیر شاخه website و سایت هایی که ان سایت با ان ارتباط دارد در زیر شاخه links قرار دارد .

خروجی این تابع یک ارایه 2 بعدی است .

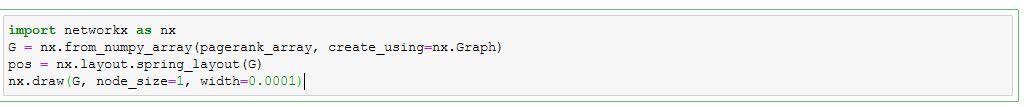
در مرحله بعدی عکس اول ارایه 2 بعدی را به تابع زیر میدهیم تا ان را تبدیل به ماتریس numpy و

از نوع adjency matrix کند .

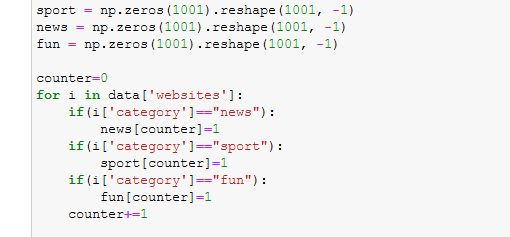


حالا ماتریس Adjency matrix را داریم و هر کاری خواستیم میتوانیم روی ان انجام دهیم .

میتوانیم با network گراف ان را رسم کنیم و یا میتوانیم با gephi ان را رسم کنیم در قسمت اخر عکس اول مقدار این ماتریس رو در یک فایل csv ذخیره کردم تا بتوانم از ان استفاده کنم .



در کد بالا ان را توسط network نوشتم که رسم کنم ولی ارور حافظه میدهد و توانایی رسم را ندارد بعد 1 ساعت رسم میکند ولی شبکه قابل تشخیص نیست.

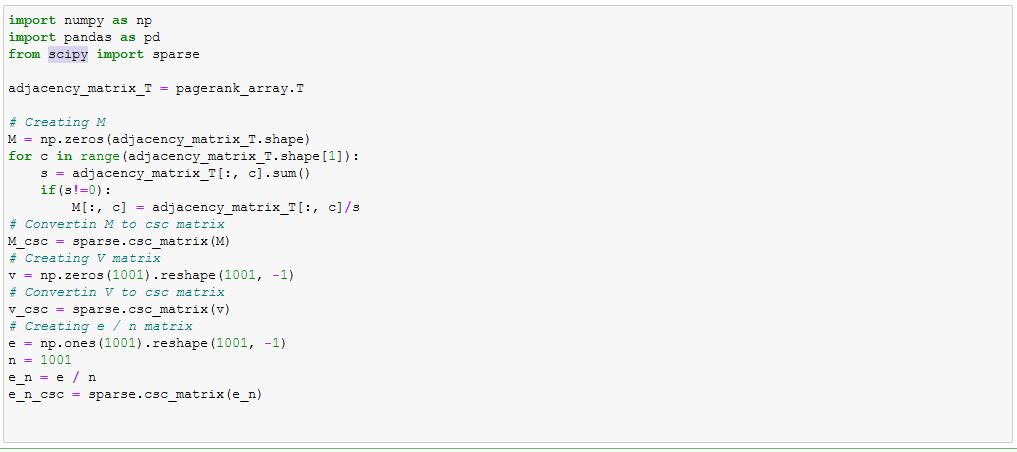


در کد بالا اومدم بر اساس دسته بندی ها و شاخه category در داده ها هر شاخه ای که مربوط به کدام بخش است را پیدا کردم و در 3 وکتور ان را قرار دادم.

تا اگر کاربر برای مثال دسته بندی ورزشی را سرچ کرد به جای e/n معمولی وکتور مخصوص ان را قرار دهیم .

در بخش اول سوال : page rank مطرح شده است کد ان به صورت زیر است :

از کتابخانه scipy برای افزایش سرعت ضرب استفاده کردم این کتابخانه خانه های صفر را نگه نمیدارد .



برای حل این الگوریتم چند تا متغیر اولیه لازم است چون فرمول ان به صورت زیر است .

B\*(MV)+(1-B)e/n

با توجه به فرمول بالا در حالت عادی به جای e/n یک وکتور که همه مقادیر ان 1000/1

است را قرار میدهیم .

و به جای B به گفته گوگل بهترین مقدار 0.85 است که ان را قرار میدهیم .

ولی در قسمت دوم سوال به جای e/n مقدار وکتور اختصاصی ان را قرار میدهیم . که در قسمت قبل ان ها را پیدا کردیم .

کد بالا به ترتیب اول تولید وکتور ها است بعد با کتابخانه spacy , ماتریس اولیه خود را به ان داده تا فقط مقادیر غیر صفر را نگه دارد و عمل ضرب را روی ان ها انجام دهد .

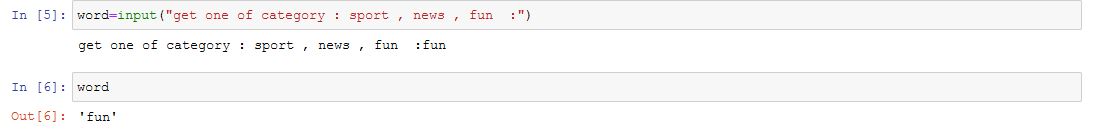
در قسمت زیر عملیات اصلی این الگوریتم را انجام میدهیم .



عملیات فرمول بالا رو با کد در شکل بالا انجام دادم 10 تا بهترین رتبه رو نمایش دادم که شکل بالا نمایان گر ان است .

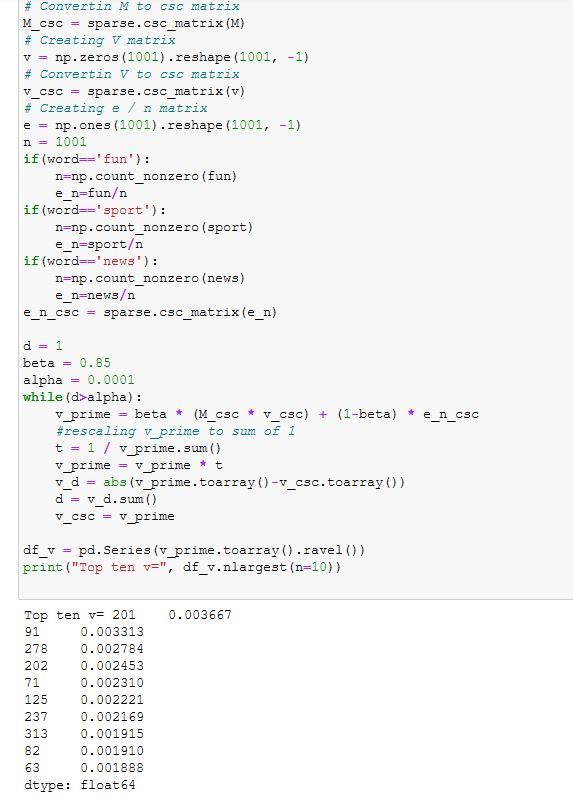
قسمت دوم سوال :

در این قسمت از کاربر یک متن گرفته و میگیرم که جزو کدام دسته است :



در زیر همان الگوریتم بالا رو بر اساس این قسمت e/n اختصاصی داده ام این به این معنی است که اگر کاربر اخبار سرچ کند بهش صفحه های خبری بیشتر پیشنهاد میدهد و ...

در زیر کد این قسمت اورده شده است .

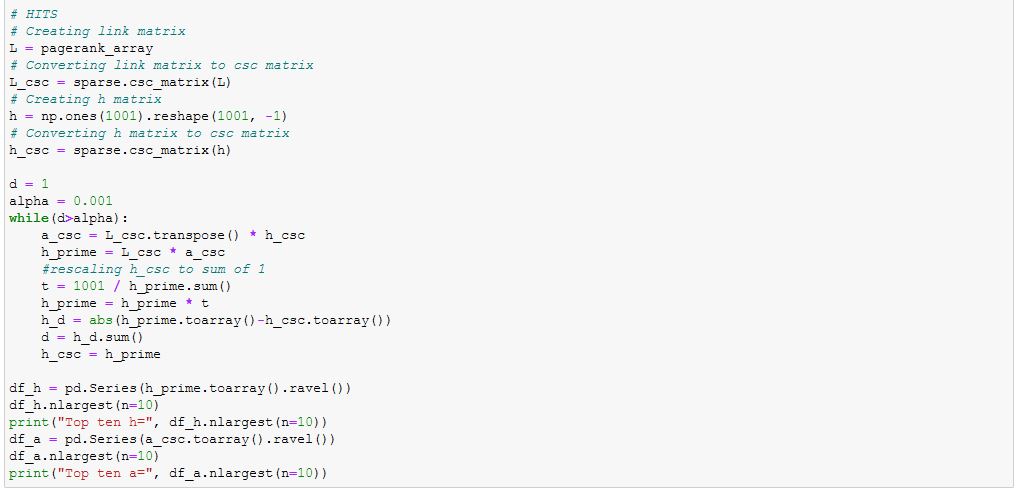


**تفاوت کد بالا با قسمت اول فقط در بخش موضوعی بودن ان است و در اخر هم 10 صفحه پر بازدید را نمایش میدهد .**

**در قسمت سوم این سوال از الگوریتم HITS استفاده شده است .**

**کد ان به صورت زیر اورده شده است .**

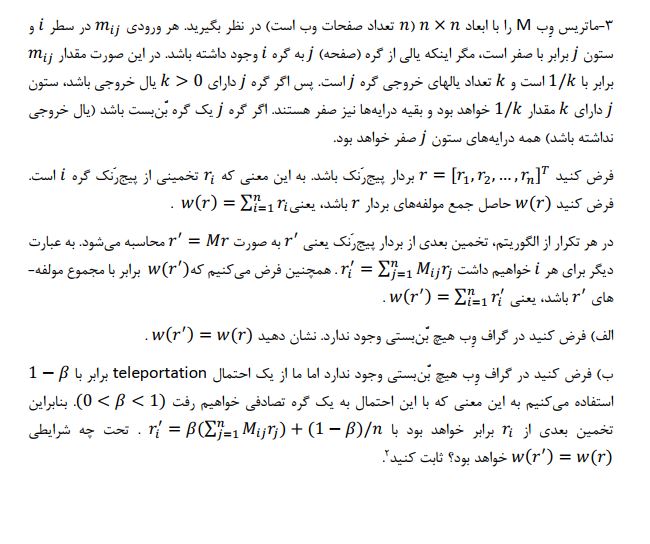
**و در اخر بر اساس h,a 10 صفحه برتر نیز نمایش داده شده است .**

****

****

**در قسمت چهارم : فایل تولید شده csv رو دادم به گفی memory error میده اجرا نشد .**

**تمرین سوم :**

****