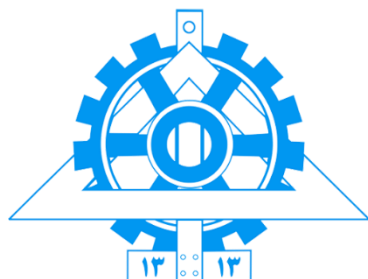


به نام خداوند جان و خرد



دانشگاه تهران

دانشکده فنی

دانشکده مهندسی برق و کامپیوتر

شبکه‌های اجتماعی

پاسخ‌نامه آزمون پایان‌ترم

نام و نام خانوادگی: **علی خرم‌فر**

شماره دانشجویی: **۸۱۰۱۰۲۱۲۹**

دی‌ماه ۱۴۰۲

فهرست مطالب

۱	پاسخ مسئله شماره ۱	۱
۲	پاسخ مسئله شماره ۲	۲
۲-۱	توضیحات الگوریتم	۲
۳	پارامتر بر پایه قدرت ارتباط بین دو نود:	۳
۳	پارامتر بر پایه شناسایی انجمن‌های گراف	۳
۴	پارامتر بر پایه شناسایی Structural holes ها	۴
۲-۲	بررسی مرتبه زمانی روش قدرت ارتباط بین دو نود	۴
۳	پاسخ مسئله شماره ۳	۵
۴	پاسخ مسئله شماره ۴	۵
۴-۱	حالت ۳ علامت مثبت	۶
۴-۲	حالت ۲ علامت مثبت و ۱ علامت منفی	۶
۴-۳	حالت ۱ علامت مثبت و ۲ علامت منفی	۷
۴-۴	حالت ۳ علامت منفی	۷
۵	پاسخ مسئله شماره ۵	۸
۵-۱	پاسخ قسمت الف	۸
۹	پاسخ قسمت a	۹
۱۰	پاسخ قسمت b	۱۰
۱۰	پاسخ قسمت c	۱۰
۱۱	پاسخ قسمت d	۱۱
۵-۲	پاسخ قسمت ب	۱۱
۵-۳	پاسخ قسمت ج	۱۱
۶	پاسخ مسئله شماره ۶	۱۱
۶-۱	محاسبه PageRank در حالتی که α صفر باشد	۱۲
۱۲	محاسبه PageRank گره‌های درون Clique	۱۲
۱۳	محاسبه PageRank گره خارجی	۱۳
۶-۲	محاسبه PageRank در حالتی که α یک باشد	۱۳
۶-۳	محاسبه PageRank در حالتی که α بین صفر و یک باشد	۱۳
۷	پاسخ مسئله شماره ۷	۱۴

۷-۱_ بررسی فاکتورهای تاثیرگذار در Modularity	۱۴
۷-۲_ یافتن انجمن‌ها با بهینه‌سازی Modularity	۱۴
مرحله اول یافتن انجمن‌ها	۱۵
مرحله دوم یافتن انجمن‌ها	۱۵
مرحله N ام یافتن انجمن‌ها	۱۵
۸_ پاسخ مسئله شماره ۸	۱۶
۸-۱_ پاسخ قسمت الف	۱۶
۸-۲_ پاسخ قسمت ب	۱۷
۹_ مراجع	۱۷

فهرست اشکال

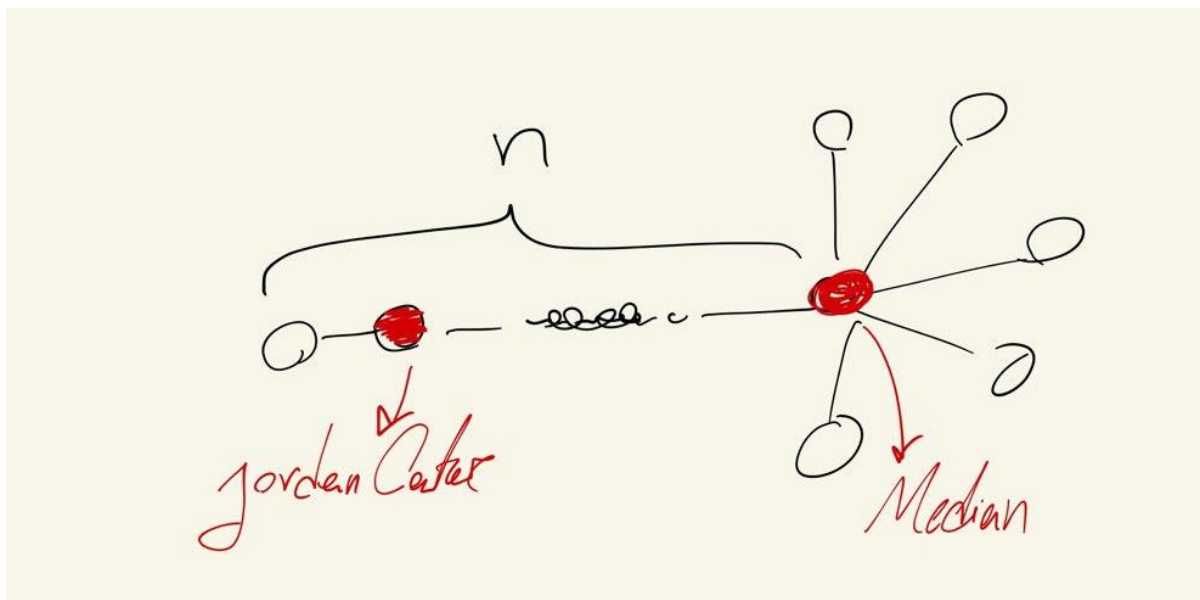
شکل ۱_ گراف با Median و JordanCenter با فاصله d	۱
شکل ۲_ گراف با Median و JordanCenter با فاصله 1	۲
شکل ۳_ ویژگی Strong Triadic Closure در گراف	۳
شکل ۴_ حالات مختلف گراف اولیه	۵
شکل ۵_ حالت ۳ علامت مثبت در گراف اولیه	۶
شکل ۶_ حالت ۲ علامت مثبت و ۱ علامت منفی در گراف اولیه	۶
شکل ۷_ حالت ۱ علامت مثبت و ۲ علامت منفی در گراف اولیه	۷
شکل ۸_ حالت ۳ علامت منفی در گراف اولیه	۷
شکل ۹_ حالت اولیه مسئله با انتخاب مجموعه I به عنوان آغازگر	۸
شکل ۱۰_ سود هر گره در گراف اولیه	۹
شکل ۱۱_ اضافه شدن گره k به استفاده‌کنندگان فناوری a	۹
شکل ۱۲_ اضافه شدن گره l به استفاده‌کنندگان فناوری a	۱۰
شکل ۱۳_ گسترش فناوری a با دو گره و یک یال اضافه	۱۱

۱- پاسخ مسئله شماره ۱

در این مسئله نیاز است که گرافیکی تشکیل دهیم که در آن فاصله بین Median و Jordan Center به اندازه d بوده و این d حداقل برابر ۱ باشد تا این دو نقطه بر یکدیگر منطبق نباشند. JordanCenter مجموعه ای از نقاط در گراف است که ماکزیمم مینیمم فاصله آن نود از تمام دیگر نودهای گراف کمینه باشد که در مسئله بیمارستان این موضوع بررسی شد. همچنین در Median مینیمم فاصله یک نقطه تا دیگر نقاط محاسبه شده و مجموع آن‌ها برای این نقطه در گراف باید کمینه باشد که در مسئله مرکز فروش آن را بررسی کردیم. برای حل این مسئله می‌توان حالتی را در نظر گرفت که یک Path در نظر گرفته می‌شود که دارای n گره می‌باشد. این زیرگراف را به یک زیر گراف با توپولوژی ستاره متصل میکنیم که این زیرگراف دارای m گره متصل به گره اصلی است. m به صورت زیر تعریف می‌شود:

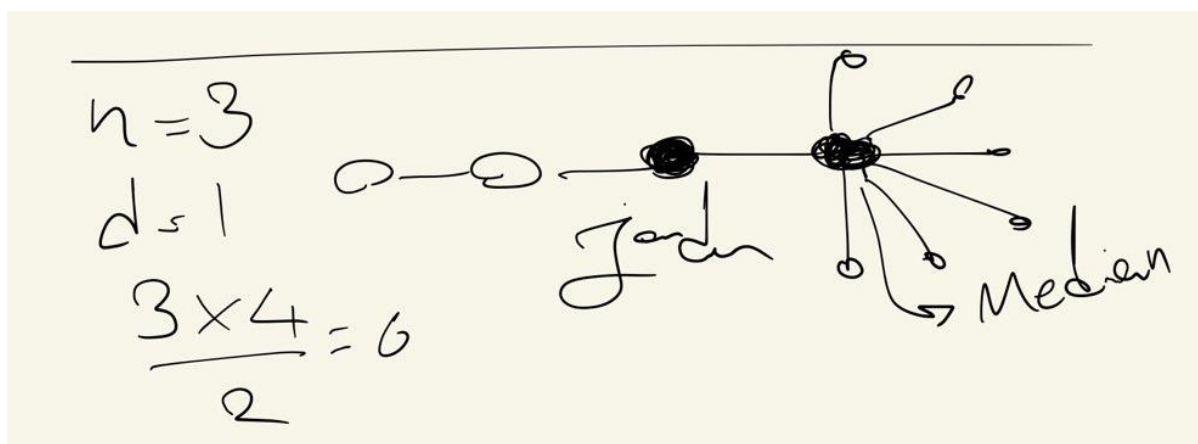
$$m = \frac{n(n+1)}{2}$$

حال این دو زیرگراف را به یکدیگر متصل می‌کنیم. و نتیجه حاصل یک گراف است که فاصله Median تا JordanCenter آن برابر با d خواهد بود. مقدار m به این علت تنظیم می‌شود تا نقطه Median همیشه در مرکز ستاره قرار بگیرد و با اضافه کردن گره به زیرگراف اولیه، مقدار فاصله d از JordanCenter به مقدار دلخواه تغییر یابد. در این حالت JordanCenter در وسط قرار میگیرد. کافی است برای هر ۱ فاصله اضافه در d ۲ گره به n اضافه شود.



شکل ۱ گراف با Median و JordanCenter با فاصله d

حالت پایه این گراف زمانی است که فاصله ۱ باشد که در شکل زیر مشاهده می‌شود. که به اندازه ۶ به نود مرکزی گراف ستاره اضافه شده‌است.



شکل ۲ شکل ۱ گراف با Median و JordanCenter با فاصله ۱

۲- پاسخ مسئله شماره ۲

مسئله پیش‌بینی ارتباطات جدید برای گراف، یکی از مسائل مهم حوزه شبکه‌های اجتماعی است که روش‌های مختلفی بسته به پارامترهای ارائه‌شده در شبکه قابل بررسی است. این حوزه به شناسایی احتمالی ارتباطات بین گره‌ها در یک شبکه بر اساس ارتباطات فعلی شبکه و سایر ویژگی‌های گره‌ها می‌پردازد. تکنیک‌های پیش‌بینی پیوند می‌توانند در طیف گسترده‌ای از کاربردها، همچون ارتباطات دوستی استفاده می‌شود. در مقاله‌ی (Wang et al., 2014) برخی از این روش‌ها بررسی شده‌اند که هر کدام پارامتر خاصی از شبکه را بررسی می‌کنند که این روش‌ها به دو گروه مبتنی بر ویژگی‌های شبکه و استفاده از یادگیری ماشین تقسیم شده‌اند. در این مسئله تمرکز اصلی بر روی ویژگی‌های شبکه خواهد بود و الگوریتمی ارائه می‌شود که از ویژگی‌های ساختاری، مانند میزان قوی بودن ارتباط دو گره و یا درجه گره‌ها، برای پیش‌بینی ارتباطات جدید بین دو گره استفاده می‌شود.

۲-۱- توضیحات الگوریتم

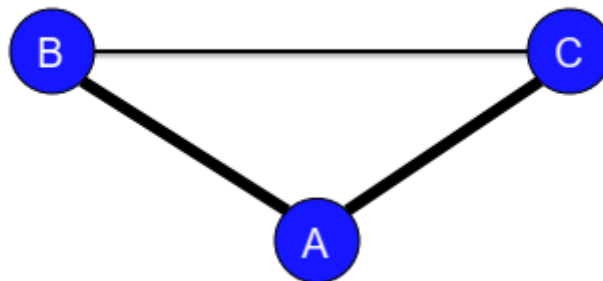
در این الگوریتم یک رویکرد کلی وجود دارد که گراف $G(V,E)$ در زمان t به صورت ورودی داده می‌شود و گراف $G(V,E)$ در زمان $t+1$ خروجی داده می‌شود. در این گراف خروجی یک یال اضافه شده که این یال بین دو گره که بیشترین امتیاز را دارند برقرار می‌شود. به این منظور هر جفت گره‌ی که به یک دیگر متصل نیستند به صورت زوج مرتب (x,y) در یک لیست مرتب قرار داده می‌شوند. سپس امتیاز هر گره محاسبه شده و با وزن خاصی که برای هر پارامتر در نظر می‌گیریم، زوج مرتب بین گره‌هایی که ارتباط ندارند تشکیل

داده که برابر ترکیبی از امتیاز دو گره هستند. سپس ماکزیمم جفت گرهی که در لیست مرتب شده قرار دارند به عنوان دو سر یال جدید انتخاب شده و پیش‌بینی می‌کنیم این یال اولین یالی است که به این شبکه اضافه خواهد شد.

در ادامه برخی از پارامترهایی که در امتیاز این زوج مرتب ها تاثیر دارند را بررسی می‌کنیم:

پارامتر بر پایه قدرت ارتباط بین دو نود:

برای محاسبه این پارامتر به این صورت در نظر می‌گیریم که بر اساس قضیه Strong Triadic Closure اگر بین دو گره A و C و همچنین دو گره A و B ارتباط قوی وجود داشته باشد، قطعاً یک ارتباط (ضعیف یا قوی) بین دو گره B و C وجود خواهد داشت.



شکل ۳ ویژگی Strong Triadic Closure در گراف

برای شبیه‌سازی این حالت در این مسئله فرض می‌کنیم که ارتباطاتی که وزن آن‌ها که معیار قوی بودن آن‌هاست، از میانه (یا در حالاتی میانگین) وزن‌ها بیشتر باشد، در این صورت این نوع ارتباط به عنوان قوی یا S در نظر گرفته می‌شود. در غیر این صورت و کمتر بودن وزن یال از میانه، ارتباط به عنوان ضعیف در نظر گرفته می‌شود. پس از آن می‌توان بر حسب میزان قوی بودن امتیازی به جفت همسایه‌ی یک گره که ارتباط قوی دارند اضافه کرد.

تا این مرحله اگر که فقط از طریق روش بالا برای پیش‌بینی استفاده کنیم مرتبه الگوریتم از $O(n^3)$ خواهد بود.

پارامتر بر پایه شناسایی انجمن‌های گراف

یکی دیگر از عواملی که در پیش‌بینی ارتباطات آینده موثر است این است که درست است که ارتباطات درون یک انجمن نسبت به خارج آن چگالی بیشتری دارد ولی به هر حال برخی از گره‌ها در یک انجمن ممکن است با یکدیگر ارتباط مستقیم نداشته باشند. در این صورت می‌توان ابتدا با توجه به ساختار گراف ورودی و

همچنین میزان قوی بودن ارتباطات آن‌ها، انجمن‌ها شناسایی شده و سپس به جفت گره‌هایی که در یک انجمن وجود داشته ولی ارتباط مستقیم ندارند امتیاز اضافه شود.

این مورد همان تمایل ارتباط یک گره با افرادی است که به آن‌ها نزدیک و همسایه که با محاسبه Neighborhood overlap قابل بررسی است.

برای محاسبه امتیاز یک روش دیگر Random Walk روی گره مورد نظر است تا همسایگان نزدیک که ارتباط با آن‌ها وجود نداشته و در یک انجمن هستند شناسایی شوند.

در صورتی که از روش بالا استفاده شود، بسته به روشی که برای شناسایی انجمن‌ها استفاده می‌شود از $O(n^2)$ تا $O(n \log n)$ متغیر خواهد بود.

پارامتر بر پایه شناسایی Structural holes

یکی دیگر از عواملی که در پیش‌بینی ارتباطات آینده و تشخیص یال بعدی موثر هستند، Structural holes ها در شبکه هستند. بر این اساس گره‌هایی که ارتباطات قوی در دو انجمن مجزا دارند تمایل دارند به واسطه برقراری Bridge بین آن‌ها ارتباط برقرار شده و به عنوان واسطه بین دو انجمن باشند. که این مورد نیز می‌تواند به امتیاز هر جفت گره در لیست مرتب شده مذکور لحاظ شده و در انتخاب جفت گره با ماکزیمم امتیاز موثر باشند.

علاوه بر عوامل بالا، پارامترهای مختلفی همچون درجه هر گره نیز ممکن است تاثیرگذار باشد که باتوجه به تمرکز مسئله بر روی روش‌های از جنس موارد مذکور از بررسی روش‌های مبتنی بر ساختار گراف صرف نظر شد.

۲-۲. بررسی مرتبه زمانی روش قدرت ارتباط بین دو نود

در این روش ابتدا میانه وزن‌های گراف محاسبه شده که $O(E \log E)$ خواهد بود. سپس برای هر دو یال متصل به یک گره، برای محاسبه امتیاز یک جفت، برچسب‌زنی Strong و Weak انجام می‌شود که این عمل نیز در مرتبه $O(V^3)$ خواهد بود. که به طور کلی الگوریتم ارائه در صورت استفاده فقط از شیوه امتیازدهی اول $O(E \log E + V^3)$ خواهد بود.

۳_ پاسخ مسئله شماره ۳

این ارتباط ضعیف است.

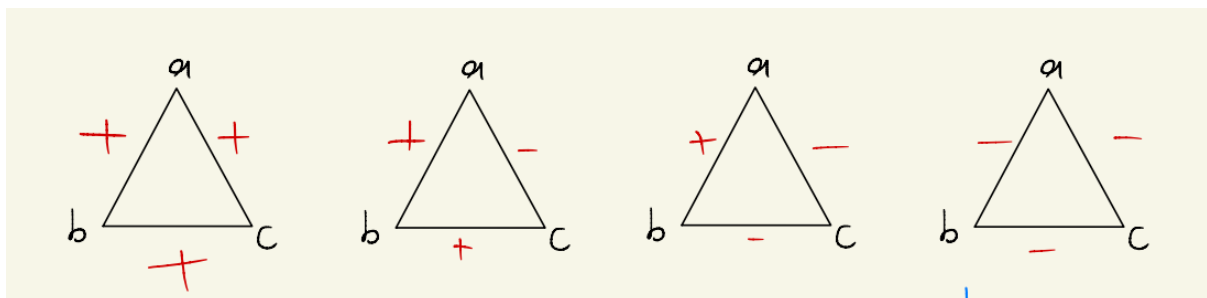
بر اساس قضیه Strong Triadic Closure اگر بین دو گره A و C و همچنین دو گره A و B ارتباط قوی وجود داشته باشد، قطعاً یک ارتباط (ضعیف یا قوی) بین دو گره B و C وجود خواهد داشت. در گراف ارائه شده در این مسئله، فرض کنیم که ارتباط بین دو گره b و c قوی باشد. پس در این صورت باتوجه به اینکه خاصیت Strong Triadic Closure در این گراف داریم و همچنین باتوجه به اینکه در این شرایط ارتباط گره b با گره های e و c قوی خواهد بود، پس در نتیجه حداقل یک یال ضعیف بین دو گره e و c خواهیم داشت که این مورد برخلاف مسئله بوده و فرض اولیه نادرست بوده و خلاف فرض اثبات می شود. پس ارتباط بین b و c ضعیف است.

همین مورد برای ارتباط قوی بین گره c و گره های f و b برقرار است.

۴_ پاسخ مسئله شماره ۴

برای بررسی این مسئله ابتدا تمامی شرایط ممکن برای گراف اولیه را تشکیل داده و یک گره d به نحوی که با تمامی ۳ گره ارتباط دارد تشکیل می دهیم. سپس به بررسی علامت یال بین گره d و هر کدام از گره های گراف اولیه خواهیم پرداخت.

باتوجه به اینکه در این مسئله گراف ممکن است Balance باشد یا نباشد، پس ۴ حالت برای ارتباطات بین این ۳ گره خواهیم داشت.

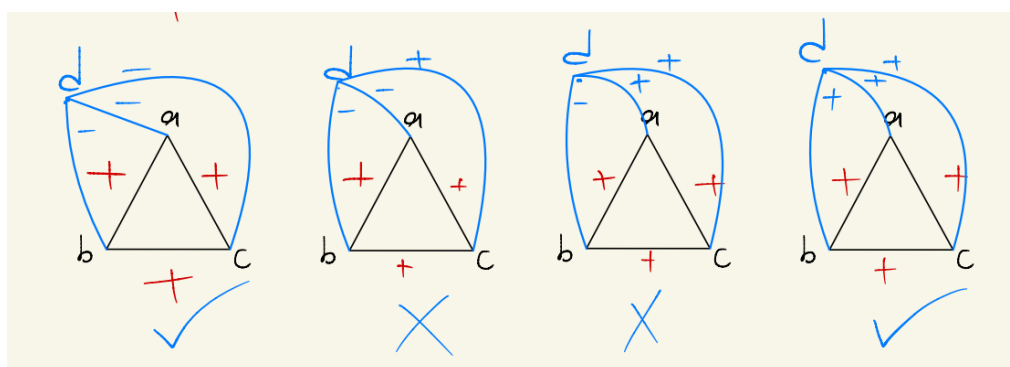


شکل ۴ حالات مختلف گراف اولیه

بدیهیست که حالاتی که علامت بین گره های گراف اولیه ممکن است جابجا شوند که حالت هایی بیشتری ایجاد شوند ولی همگی این حالات جدید ایزومورف همان ۴ حالت هستند که در ادامه بررسی می کنیم.

۴-۱. حالت ۳ علامت مثبت

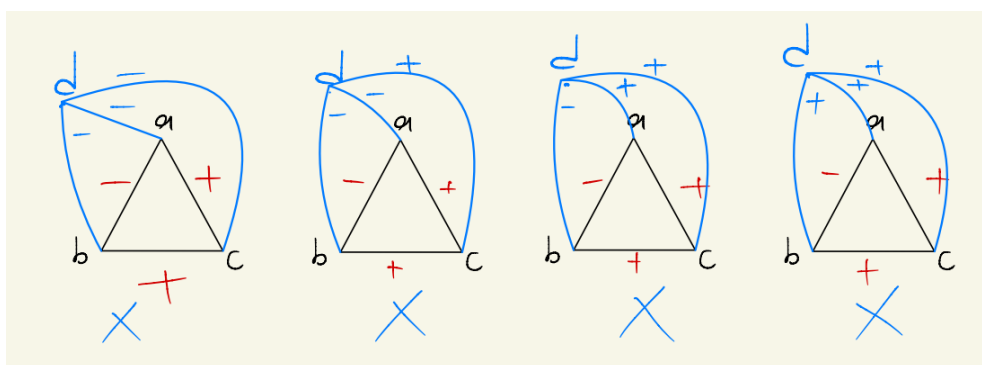
در این حالت همانگونه که در شکل زیر مشاهده می‌شود، در صورتی که هر سه علامت یال‌های ایجاد شده توسط گره d مثبت باشند، در ارتباطات سه تایی که d در آن‌ها شرکت دارد Balance خواهیم داشت. یک حالت دیگر نیز زمانی است که تمام ارتباطات d منفی باشند. باتوجه به اینکه ارتباطات منفی هستند باید به دو سمت گرهی متصل شوند که ارتباط بین آن‌ها مثبت است. در این صورت Balance برقرار خواهد شد. در غیر این صورت حداقل ۱ حالت unbalance وجود خواهد داشت که در شکل زیر دو حالت وسط از این موارد هستند.



شکل ۵ حالت ۳ علامت مثبت در گراف اولیه

۴-۲. حالت ۲ علامت مثبت و ۱ علامت منفی

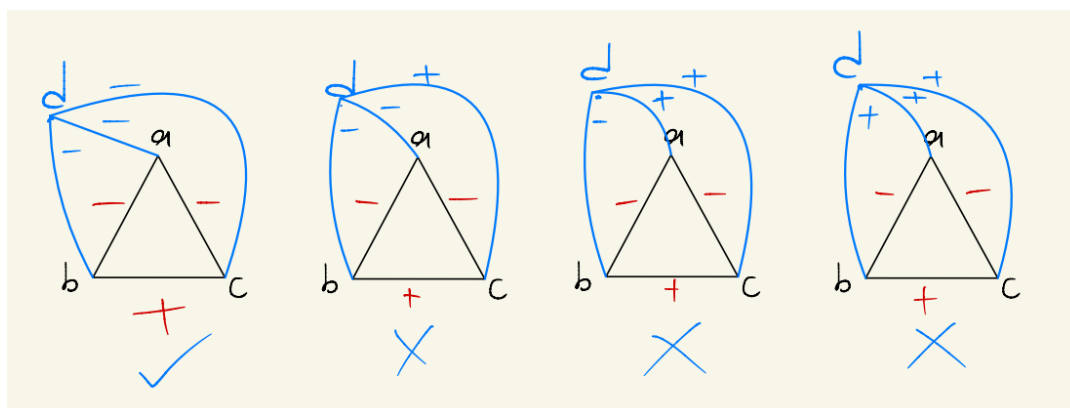
در این حالت باتوجه به گراف‌های رسم‌شده که در شکل زیر مشاهده می‌شود، در همه حالات مثال نقض در ارتباطات سه تایی ایجاد شده توسط d داریم پس در هیچکدام از موارد شرط مسئله برقرار نخواهد شد. زیرا که ارتباطاتی که برقرار شده‌اند دارای حالتی هستند که عامل unbalance شدن گراف می‌شوند. همانند حالت ۲ علامت مثبت و یک علامت منفی. پس راه حلی در این حالت از گراف اولیه وجود ندارد.



شکل ۶ حالت ۲ علامت مثبت و ۱ علامت منفی در گراف اولیه

۳-۴_ حالت ۱ علامت مثبت و ۲ علامت منفی

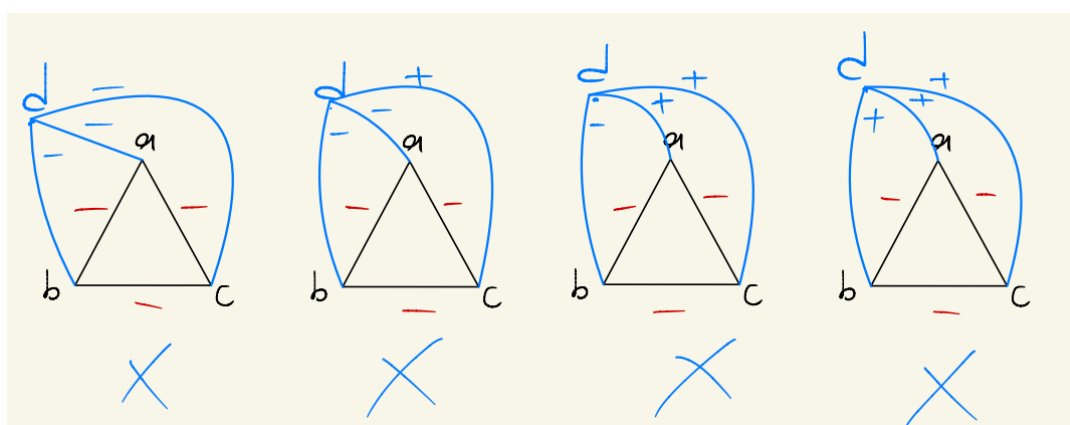
در این حالت همانگونه که در شکل زیر مشاهده می‌شود، در صورتی که هر سه علامت یال‌های ایجاد شده توسط گره d منفی باشند، در ارتباطات سه تایی که d در آن‌ها شرکت دارد Balance خواهیم داشت. در غیر این صورت حداقل ۱ حالت unbalance وجود خواهد داشت که مانع برقراری راه‌حل برای مسئله خواهد شد که در شکل زیر سه حالت وسط از این موارد هستند.



شکل ۷ حالت ۱ علامت مثبت و ۲ علامت منفی در گراف اولیه

۴-۴_ حالت ۳ علامت منفی

در این حالت باتوجه به گراف‌های رسم‌شده که در شکل زیر مشاهده می‌شود، در همه حالات مثال نقض در ارتباطات سه تایی ایجاد شده توسط d داریم پس در هیچکدام از موارد شرط مسئله برقرار نخواهد شد. زیرا که ارتباطاتی که برقرار شده‌اند دارای حالتی هستند که عامل unbalance شدن گراف می‌شوند. همانند حالت ۲ علامت مثبت و یک علامت منفی. پس راه حلی در این حالت از گراف اولیه وجود ندارد.



شکل ۸ حالت ۳ علامت منفی در گراف اولیه

۵_ پاسخ مسئله شماره ۵

در این مسئله فرض شده که گره‌ها در گراف اولیه از فناوری b استفاده می‌کنند، حال مجموعه I شامل دو گره e و f به عنوان آغازگر استفاده از فناوری a انتخاب شده‌اند. با توجه به مباحث مطرح شده در قسمت Networked Coordination Games در صورتی یک گره فناوری جدیدی مانند A را می‌پذیرد که:

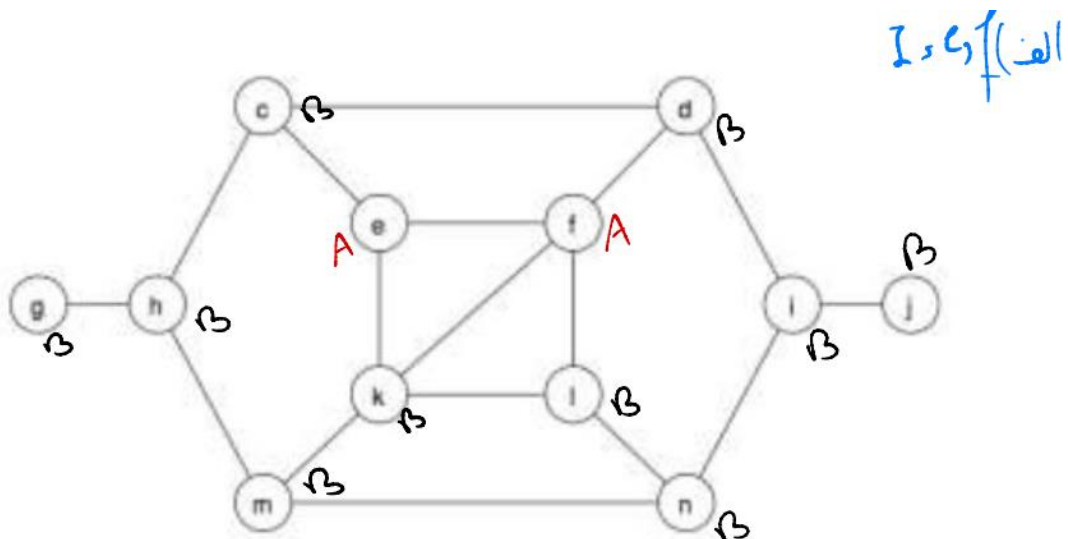
$$p d a \geq (1 - p) d b \rightarrow p \geq \frac{b}{a + b}$$

هرچند که در این مسئله موضوع پیشینه‌کردن سود نیز در همین راستا مطرح شده‌است. در ادامه به بررسی حالاتی که در آینده برای این شبکه پیش‌بینی می‌شود می‌پردازیم.

۵-۱_ پاسخ قسمت الف

مجموعه I شامل دو گره e و f به عنوان آغازگر استفاده از فناوری a انتخاب شده‌اند. در صورتی که دو همسایه فناوری a را انتخاب کرده باشند، میزان سود هر کدام ۳ و در غیر این صورت و انتخاب فناوری b هر کدام به مقدار ۲ سود خواهند برد. در این حالت در هر مرحله مقدار p نیز محاسبه می‌کنیم:

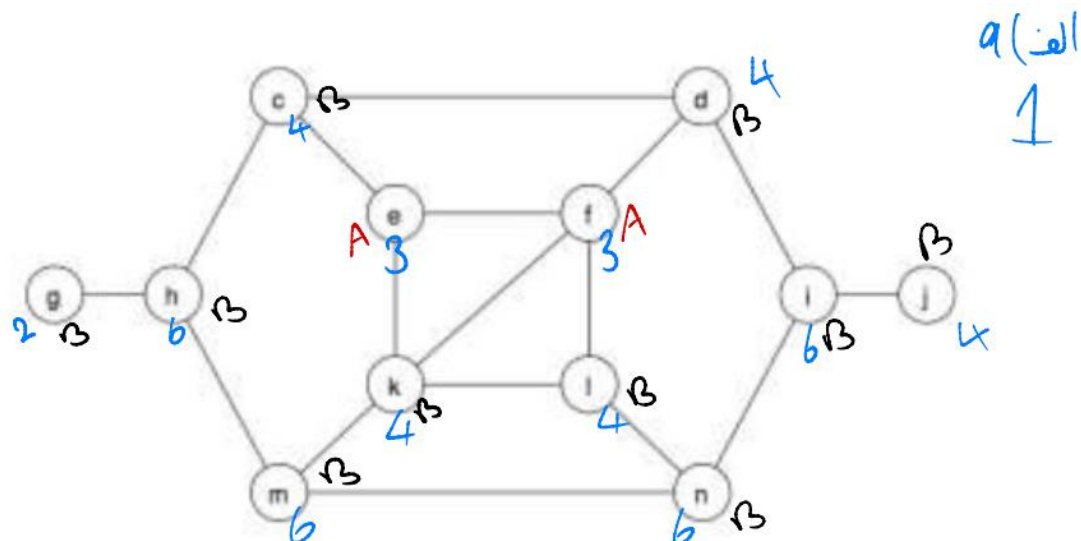
$$p \geq \frac{b}{a+b} \rightarrow p \geq \frac{4}{10}$$



شکل ۹ حالت اولیه مسئله با انتخاب مجموعه I به عنوان آغازگر

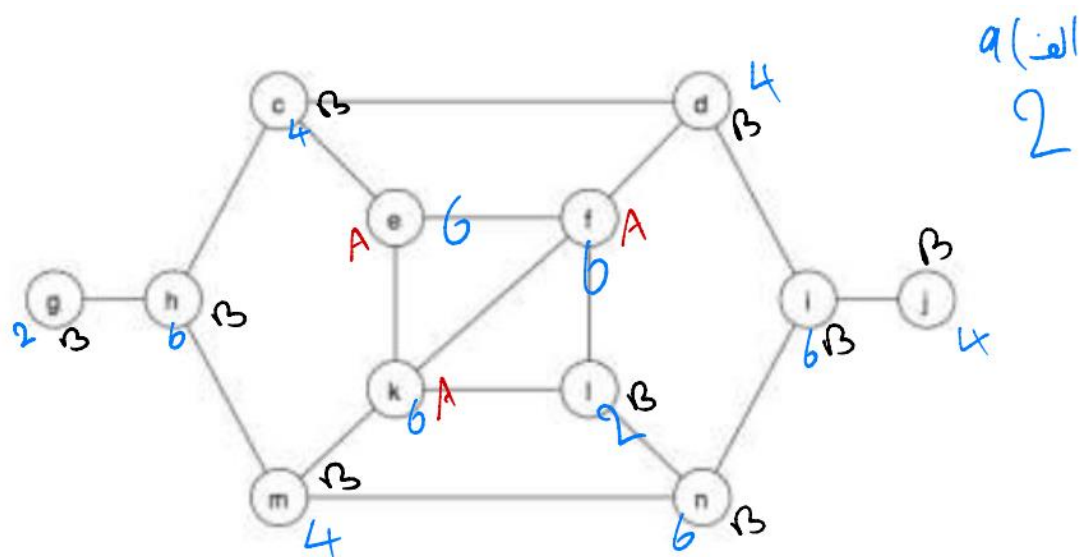
پاسخ قسمت a

گره‌هایی که در هر مرحله مقدار p بیشتر از این میزان دارند را به همراه سود در هر مرحله محاسبه شده که در شکل زیر نتیجه هر مرحله مشاهده می‌شود:



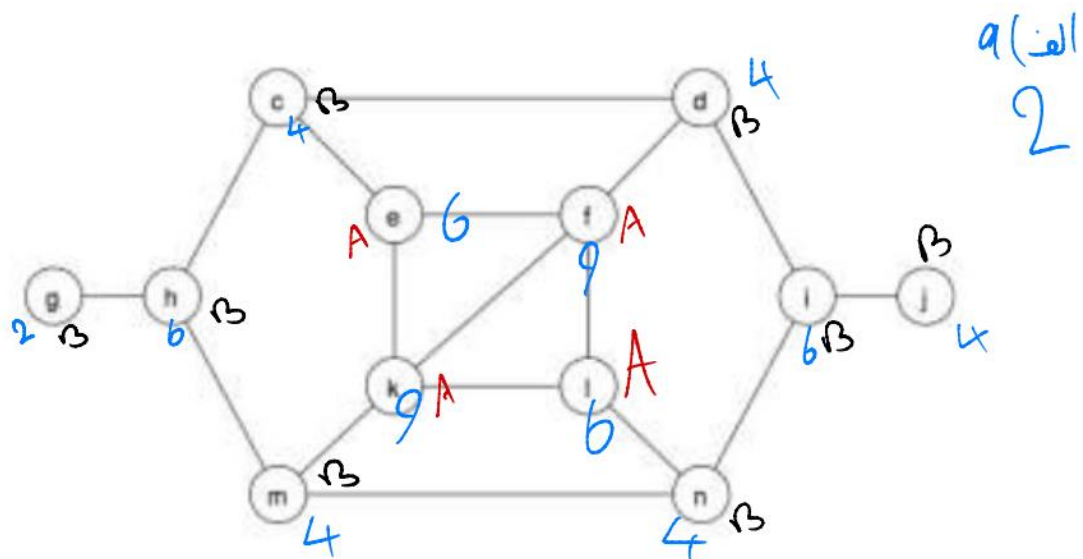
شکل ۱۰ سود هر گره در گراف اولیه

مرحله اول) گره K نیز به فناوری a روی می‌آورد.



شکل ۱۱ اضافه شدن گره k به استفاده‌کنندگان فناوری a

مرحله دوم) گره 1 نیز به فناوری a روی می آورد:



شکل ۱۲ اضافه شدن گره 1 به استفاده کنندگان فناوری a

پس از این مرحله به دلیل اینکه مقدار p برای گره های دیگر که از b استفاده می کنند، به حد مورد نیاز نمی رسد، پس گسترش a متوقف می گردد.

پاسخ قسمت b

باتوجه به تعاریف انجام شده، در صورتی که به یک انجمن با چگالی $1-t$ برخورد کنیم، گسترش فناوری متوقف می شود. پس چگالی انجمنی که فناوری b را حفظ می کنند برابر است با:

$$t = \frac{4}{10}$$

$$1 - t > \frac{6}{10}$$

باتوجه به اینکه در سوال مطرح شده که در صورت برابر بودن ارزش ها از فناوری a استفاده خواهد شد پس علامت به صورت بزرگتر خواهد بود و خود مقدار $0/6$ جز حداقل چگالی نیست.

پاسخ قسمت c

حداقل ۲ یال - به دلیل وجود گره های i و h تنها در صورتی می توانیم تکنولوژی a را به این دو گره منتقل کنیم که دو سر آنها نیز برقرار باشد. و اینکار نیازمند برقراری دو یال است. در غیر این صورت گسترش در یکی از این دو نقطه متوقف می شود.

پاسخ قسمت d

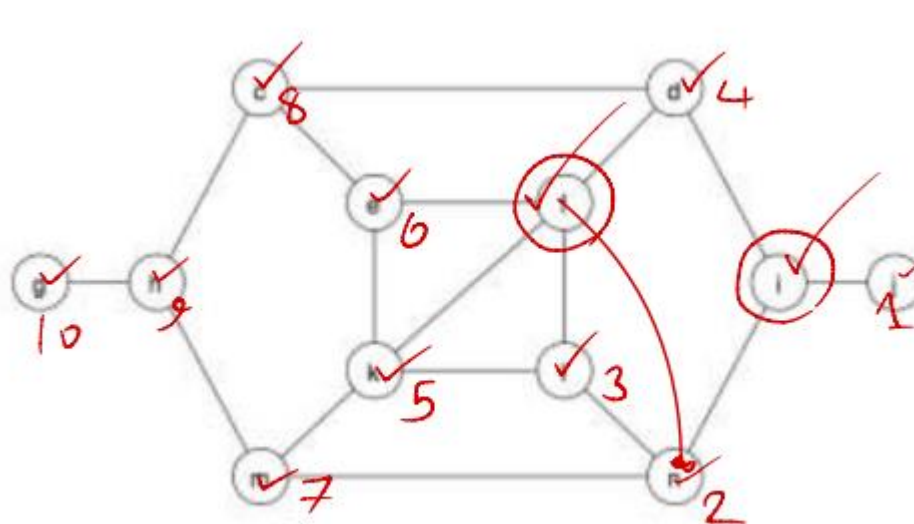
بله - با اضافه شدن یکی از دو گره i یا h که در قسمت c علت آن بررسی شد، مشکل حل می شود.

۵-۲_ پاسخ قسمت ب

حداقل ۳ گره نیاز است. که باتوجه به بررسی های انجام شده می توان همان مجموعه اولیه I با گره های e و f را به همراه یکی از گره هایی بررسی شده در قسمت d مثلاً i به عنوان گره آغازین انتخاب کنیم.

۵-۳_ پاسخ قسمت ج

قسمت های مختلفی می توان این یال را برقرار کرد که بستگی به انتخاب گره های آغازگر دارد. که در شکل زیر یک نمونه آن رسم شده و مشخص شده پس آن تمام گره ها به فناوری a ملحق می شوند.



شکل ۱۳ گسترش فناوری a با دو گره و یک یال اضافه

۶_ پاسخ مسئله شماره ۶

در این مسئله یک Clique به اندازه n ارائه شده که یک زیرگراف کامل n تایی است که یک گره به تمامی گره های این Clique متصل شده است. جهت تمامی این یال ها از سمت گره های کلیک به سمت گره خارجی است. پس در این مسئله گراف دارای $n+1$ گره است.

حال به بررسی مطلوب مسئله می پردازیم. برای محاسبه PageRank در هر لحظه از زمان فرمولی ارائه شده که بر حسب آن مقدار PageRank هر گره در لحظه $t+1$ محاسبه می شود. در لحظه $t=0$ مقدار ۱ بین تمامی گره ها تقسیم می شود. پس در ابتدا هر گره دارای PageRank معادل $1/n$ می باشد.

$$p_i(t+1) = \beta \sum_j \left[\frac{a_{ij} p_i(t)}{k_j} \right] + \frac{1-\beta}{n}$$

این فرمول بیان می‌کند که در هر دوره زمانی، مقدار PageRank یک گره برابر جمع ۲ حالت است.

یا رندم جامپ خواهیم داشت به احتمال $1-\beta$ که در آن صورت میزان مقداری به کل گره‌های گراف به اندازه $1-\beta$ در هر مرحله اضافه می‌شود.

در غیر این صورت پیج رنک برابر جمع پیج رنک تمامی گره‌هایی است که به سمت آن گره یال دارند. که البته میزان پیج رنک گره‌های همسایه تقسیم بر درجه آن‌ها می‌شود.

در این مسئله فرمول PageRank به صورت زیر بر اساس مفروضات مسئله بازنویسی می‌شود:

$$p_i(t+1) = (1-\alpha) \sum_j \left[\frac{a_{ij} p_i(t)}{k_j} \right] + \frac{\alpha}{n+1}$$

حال در این مسئله بر اساس اینکه مقدار α به چه میزانی باشد، ممکن است شرایط مختلفی رخ دهد که هر کدام را در حال حدی بررسی می‌کنیم. بدیهی است که در زمان صفر مقدار PageRank هر گره در این مسئله برابر است با:

$$p_i(0) = \frac{1}{n+1}$$

۱-۶. محاسبه PageRank در حالتی که α صفر باشد

در این مسئله باتوجه به اینکه همه گره‌های Clique یال خروجی به سمت گره خارجی دارند پس مانند مثال ارائه شده در محتوای درس یک حالت حلقه تشکیل شده که در آن تمامی PageRank ها در یک نقطه رسوب می‌کنند. حال در این حالت PageRank را برای هر کدام از گره‌ها محاسبه می‌کنیم.

محاسبه PageRank گره‌های درون Clique

در هر گره $n-1$ گره ورودی وجود دارد.

$$p_i(1) = (1-0) \sum_j \left[\frac{a_{ij} p_i(t)}{k_j} \right] + \frac{0}{n+1} = \frac{1}{n+1} * \frac{1}{n+1} + \dots = \frac{n-1}{(n+1)(n+1)}$$

پس هربار مقدار PageRank گره‌های کلیک کم می‌شود. پس در حالت حدی در صورتی که RandomJump نداشته باشیم مقدار آن‌ها برابر صفر خواهد شد.

$$p_i(t+1) = (1-0) \sum_j \left[\frac{a_{ij} p_i(t)}{k_j} \right] + \frac{0}{n+1} \sim t \rightarrow \infty = 0$$

محاسبه PageRank گره خارجی

در این حالت چه در صورتی که α صفر باشد چه زمانی که غیرصفر باشد یک مشکل وجود خواهد داشت. و این مورد این است که مقدار PageRank تغییری نمی‌کند.

$$p_i(t+1) = (1-\alpha) \sum_j \left[\frac{a_{ij} p_j(t)}{k_j} \right] + \frac{\alpha}{n+1} = \frac{1-\alpha}{n+1} + \frac{\alpha}{n+1} = \frac{1}{n+1}$$

که این میزان برابر همان $p_i(0)$ تخصیصی در اول مسئله است. این مورد یکی از مشکلات تعریف فعلی از PageRank است که باعث می‌شود گره‌هایی که یال خروجی نداشته باشند، باعث خطای محاسباتی شوند که به این نوع صفحات dangling pages گفته می‌شود. (Berkhin, 2005)

روش‌های مختلفی برای حل این مشکل پیشنهاد شده‌است. برای نمونه در مقاله (Berkhin, 2005) پیشنهاد شده این صفحات حذف شوند. اما در این مسئله این مورد امکان‌پذیر نیست. هرچند که در محاسبات در فضای ماتریس پیشنهاد شده یال‌های جدیدی اضافه شود ولی ما از راه حل ارائه شده در مقاله (Bianchini, 2005) استفاده می‌کنیم و برای گره خارجی یک حلقه به سمت خودش قرار می‌دهیم تا در هر مرحله برای محاسبات مقادیر به سمت خودش اضافه شوند. پس باتوجه به موارد بررسی شده در حالت حدی تمامی مقادیر به سمت این گره خارجی خواهند رفت.

$$p_i(t+1) = (1-0) \sum_j \left[\frac{a_{ij} p_j(t)}{k_j} \right] + \frac{0}{n+1} \sim t \rightarrow \infty = \frac{n+1}{n+1} = 1$$

۲-۶. محاسبه PageRank در حالتی که α یک باشد

در این حالت هیچ‌گونه تقسیم‌پذیری بین یال‌های خروجی وجود نخواهد داشت و برخلاف قسمت قبل گره خارج از کلیک مزیتی ندارد. پس تمام مقادیر به صورت مساوی بین آن‌ها تقسیم می‌شود.

۳-۶. محاسبه PageRank در حالتی که α بین صفر و یک باشد

در این حالت باتوجه به اینکه کماکان برتری بررسی شده در حالت اول برای گره خارجی برقرار است، پس تمامی مقادیر به سمت این گره جاری خواهند شد و در آن‌جا خواهند ماند. ولی در این حالت میزان α مقداری بین صفر و یک دارد پس RandomJump خواهیم داشت و مقدار PageRank در حالت به میزان $\frac{\alpha}{n+1}$ که مربوط به ترم رندم جامپ در فرمول است به همه‌ی دیگر گره‌ها خواهد رسید و این میزان هر بار از گره خارجی کم می‌شود. (همانند مثال مالیات ارائه شده در درس)

$$p_i(t+1) = (1-\alpha) \sum_j \left[\frac{a_{ij} p_j(t)}{k_j} \right] + \frac{\alpha}{n+1} \sim t \rightarrow \infty = \frac{\alpha}{n+1}$$

۷ پاسخ مسئله شماره ۷

در این مسئله فرض می‌شود که از روش (Newman & Girvan, 2004) برای تقسیم انجمن‌های شبکه استفاده می‌شود. در این روش به صورت حریصانه هربار دو گره که میزان Q را بیشینه می‌کنند به یکدیگر متصل شده و این مرحله مکرراً انجام می‌پذیرد تا به Clustering مطوب دست یابیم.

۷-۱. بررسی فاکتورهای تاثیرگذار در Modularity

میزان Modularity از فرمول زیر محاسبه می‌شود:

$$Q = \frac{1}{2m} \sum_{ij} \left(A_{ij} - \frac{k_i k_j}{2m} \right) \delta(c_i, c_j)$$

در این رابطه m تعداد یال‌های گره است.

$\delta(c_i, c_j)$ برابر ۱ است اگر که در یک انجمن فرض شوند، در غیر این صورت برابر ۰ است.

احتمال $\frac{k_i k_j}{2m}$ این است که در صورت تصادفی بودن تخصیص یال تعداد بین گره i و j قرار بگیرند. این مورد به درجه نودهای i و j بستگی دارد.

A_{ij} در صورتی که بین دو گره در گراف اصلی یال باشد برابر ۱ و در غیر این صورت صفر است.

* پس نتیجه می‌گیریم اگر که هرچه درجه گره i و j بیشتر باشد، مقدار Q کمتر است.

حال به بررسی شرایط این مسئله می‌پردازیم.

۷-۲. یافتن انجمن‌ها با بهینه‌سازی Modularity

در این مسئله تعداد C کلیک به هم متصل شده‌اند که هر کدام دارای m گره هستند. باهدف اینکه این m با m در فرمول اشتباه نشود، تعداد هر گره در هر کلیک n فرض می‌شود. پس تعداد یال‌ها در فرمول بالا برابر است با تعداد یال‌های درون کلیک که در گراف کامل برابر $n(n-1)/2$ که c کلیک داریم و همچنین بعلاوه c که تعداد یال‌های متصل کننده کلیک‌ها به یکدیگر خواهد بود.

$$m = \frac{Cn(n-1)}{2} + C$$

همچنین در این گراف مقدار درجه گره‌های مرزی یعنی گره‌هایی که عامل اتصال دو کلیک به یکدیگر هستند برابر است با:

$$K_{Outer} = n$$

دلیل این مورد این است که این گره‌ها علاوه بر اتصال با $n-1$ گره داخلی درون کلیک یک ارتباط بیرون کلیک هم دارند. پس در صورتی که دو گره بررسی شده عامل اتصال دو کلیک متفاوت باشند مقدار $k_i k_j$ برابر است با:

$$k_i k_j = n * n$$

همچنین در این گراف مقدار درجه گره‌های درونی هر کلیک برابر است با:

$$K_{inner} = n - 1$$

دلیل این مورد این است که این گره‌ها با $n-1$ گره داخلی درون کلیک ارتباط دارند. پس در صورتی که دو گره بررسی شده درون یک کلیک باشند مقدار $k_i k_j$ برابر است با:

$$k_i k_j = (n - 1)(n - 1)$$

*** پس نتیجه می‌گیریم در حین انجام الگوریتم Clustering اگر دو نود برای اتصال بررسی شد، در صورتی که در یک کلیک باشند، $k_i k_j$ کمتر از حالتی است که دو گره از دو کلیک متفاوت بوده و نقش گره مرزی را برای اتصال دو کلیک ایفا کنند.

باتوجه به نتایج حاصل * و **، پس در حالت اول Q بیشتر خواهد شد.

مرحله اول یافتن انجمن‌ها

در مرحله اول هر نود ایزوله به عنوان یک Cluster در نظر گرفته می‌شود.

مرحله دوم یافتن انجمن‌ها

در این مرحله دو گره که مقدار Q بیشتری ارائه دهند متصل شده و یک انجمن دوتایی تشکیل می‌دهند. باتوجه به بررسی‌های انجام شده این دو گره درون کلیک بوده و گره مرزی جز این اتصال نیست.

مرحله Nام یافتن انجمن‌ها

الگوریتم به صورت حریصانه تکرار می‌شود و در نهایت باتوجه به نتیجه‌گیری‌های انجام شده، انجمن‌هایی که بهینه‌سازی modularity برای تشخیص انجمن در این گراف اجرا می‌شود، برابر با Clique های گراف اولیه خواهند بود.

۸ پاسخ مسئله شماره ۸

۸-۱ پاسخ قسمت الف

رویکردهای مختلفی برای شناسایی انجمن‌های شکل گرفته در این وبسایت وجود دارد. آنگه که محتوا را از کاربران جداسازی کنیم پس در این گراف Focal Point خواهیم داشت. که ارتباط کاربران با این Focal Point ها به صورت منفی یا مثبت خواهد بود. یعنی در این گراف علامت منفی و یا مثبت بین هر گره و Foci خواهیم داشت. در Adjacency matrix علاوه بر ۰ و ۱، منفی ۱ هم خواهیم داشت. در Affiliation network نوع ارتباط به نحوی بود که کاربرانی که به یک focal point مشترک متصل بودند احتمال بیشتری داشت که با توجه به Homophily بر یکدیگر تاثیر گذاشته و ارتباط داشته باشند و به معنای دیگر تشکیل انجمن دهند. در حالت خاص این مسئله، علاوه بر طرفداران یک محتوا، منتقدان یک محتوا هم دسته‌بندی می‌شود. برای مثال اگر یک کاربری محتوای سیاسی با حزب راست تولید می‌کند، در عین حالی که ممکن است هیچ کاربری محتوای نزدیک به حزب چپ تولید نکند، ولی با شناسایی منتقدان تولیدکننده محتوای حزب راست، می‌توان انجمنی برای آن‌ها تشکیل داده و آن‌ها را بشناسیم.

اگر از روش اول که گفته شد، با کمک Affiliation matrix کاربران موافق و مخالف یک محتوا تشخیص داده شوند، در ادامه می‌توان انجمن‌هایی بر همین اساس تقسیم‌بندی کرد.

روش دیگر این است که می‌توان از تکنیک‌های دیگر Clustering مطرح شده استفاده کرد و یک حالت سلسله‌مراتب بر اساس علایق کاربران تشکیل داد و کاربران را بر اساس علایق آن‌ها و تعداد رای مثبت در یک انجمن قرار داد. و باید توجه شود که اگر کاربری از محتوای یک انجمن رای منفی داده شود احتمال کمتری باشد که عضو آن انجمن قرار داده شود. همچنین می‌توان روش‌های مختلفی برای Clustering استفاده کرد و لزوماً یک Clustering مورد بررسی قرار نگیرد.

همچنین در این سیستم نظر همه افراد ممکن است به یک مقدار وزن نداشته باشد و بر اساس فعالیت و تعداد رای‌شان وزن خاص به آن‌ها نسبت دهیم.

یک نکته که در این سوال وجود دارد این است که مطالب بر اساس رای مرتب می‌شوند، پس مشخص است مطالبی که رای بالایی دارند لزوماً بهترین مطالب نیستند. زیرا دلایل متعددی وجود دارد که یک مطلب رای بالایی داشته باشد مثلاً اثر Information cascade که برای تشخیص انجمن‌ها لازم است اگر که وزندهی خاصی انجام می‌شود این موارد هم در نظر گرفته شود تا Outlier ها چه از سمت مثبت چه از سمت منفی باعث خطا در تشخیص انجمن‌ها نشوند.

۲-۸_ پاسخ قسمت ب

امروزه روش‌های مختلفی برای تشخیص کاربران تقلبی استفاده می‌شود ولی باتوجه به اینکه در فرض این مسئله اطلاعات کمی نسبت به کاربران داریم لذا روش‌های تشخیص نیز محدود خواهند شد. همچنین در این گراف حالت بررسی Static فرض شده‌است. پس فرکانس رای دهی نیز نخواهیم داشت.

یکی از فاکتورهای مهم قابل بررسی این موضوع است که کاربری به محتوای تولید شده از سمت یک کاربر دیگر فقط امتیاز منفی بدهد. در این صورت گمان خواهد رفت که این کاربر تقلبی بوده از انجمن رقیب باشد. و یا برعکس ممکن است کاربری فقط به محتوای یک کاربر خاص رای مثبت دهد. در این صورت احتمال داده می‌شود که همان کاربر تولید کننده محتوا یک کاربر تقلبی هم دارد.

یکی دیگر از موارد تشخیص انجمن‌هایی است که رفتار قابل پیش‌بینی و دارای الگوی خاص دارند که احتمال می‌رود تمام کاربران آن تقلبی باشند. به طور کلی روش‌های تشخیص مختلف هستند که در شبکه‌های اجتماعی نیز به کار می‌روند (Yang et al., 2015) در صورتی که فاکتورهای مورد بررسی برای یک کاربر از حدی بالاتر برود حساب آن کاربر مسدود شده و تمامی رای‌های صادر شده نیز باطل می‌شود.

۹_ مراجع

Berkhin, P. (2005). A survey on PageRank computing. *Internet mathematics*, 2(1), 73-120 .

Bianchini, M., Gori, M., & Scarselli, F. (2005). Inside pagerank. *ACM Transactions on Internet Technology (TOIT)*, 5(1), 92-128 .

Newman, M. E., & Girvan, M. (2004). Finding and evaluating community structure in networks. *Physical review E*, 69(2), 026113 .

Wang, P., Xu, B., Wu, Y., & Zhou, X. (2014). Link prediction in social networks: the state-of-the-art. *arXiv preprint arXiv:1411.5118* .

Yang, Z., Xue, J., Yang, X., Wang, X., & Dai, Y. (2015). VoteTrust: Leveraging friend invitation graph to defend against social network sybils. *IEEE Transactions on dependable and secure computing*, 13(4), 488-501 .