بسم الله الرحمن الرحیم

دانشگاه یزد

دانشکده ریاضی

گروه علوم کامپیوتر

پایان‌نامه

جهت دریافت درجه کارشناسی ارشد

علوم کامپیوتر

مقایسهٔ مدلهای ارائه شده برای شبکه‌های اجتماعی برخط با محوریت انجمن‌های تشکیل شده

استاد راهنما:

دکتر مهدیه هاشمی‌نژاد

استاد مشاور:

دکتر علی دولتی

پژوهش‌گر:

حنیف امام‌قلی‌زاده

مهر ۱۳۹۳

تقدیم به

پدر و مادر عزیزم

و همه کسانی که می‌اندیشند.

سپاس‌گزاری

سپاس از اساتیدم که نگرشی نو به ریاضی را به من آموختند. از آنان آموختم که از ریاضی لذت ببرم، بپرسم و راضی نشوم.

چکیده

چگونگی توسعه شبکه‌های اجتماعی و تغییرات ساختاری آن‌ها یکی از موارد مورد بحث در مورد شبکه‌های اجتماعی است. چندین مدل برای پیش‌بینی ویژگی‌های شبکه‌های اجتماعی ارائه شده‌است.

در سال ۲۰۰۹ با در نظرگرفتن شبکه‌های اجتماعی ایمیل و وب سایت last.fm، مقایسه‌ای بین این مدل‌ها صورت گرفته است، که با توجه به عدم فراگیری این شبکه‌ها، در این مقایسه کمبودهایی دیده می‌شود. این کمبودها تنها به فراگیری شبکه‌های اجتماعی محدود نمی‌شود، در پژوهش مذکور، دو داده‌ی جهان واقعی با ویژگی‌های تقریبا مشابه یعنی میانگین درجه و ضریب خوشه‌بندی تقریبا کوچک تشکیل می‌شد. اما با متراکم شدن شبکه‌های اجتماعی نظیر فیس بوک، توییتر و گوگل پلاس نیاز به بازنگری دوباره در این مقایسه احساس می‌شود. با ظهور این شبکه‌های متراکم با ضریب خوشه‌بندی بالا، عیب‌های مدل‌های ارائه شده برای شبکه‌های اجتماعی، بیشتر آشکار می‌شود.

در این پژوهش، بعد از معرفی دقیق هرکدام از مدل‌ها، با در نظر گرفتن شبکه اجتماعی برخط فیس بوک و وب سایت ویکی‌وُت این مدل‌ها پیاده‌سازی، مورد بررسی و مقایسه قرار گرفته‌اند. در قسمت مقایسه نشان خواهیم داد، که اغلب این مدل‌ها توانایی شبیه‌سازی ویژگی‌های شبکه‌های اجتماعی را ندارند.

فهرست مطالب

فهرست تصاویر

فهرست جداول

فصل ۱

مقدمات

۱.۱ آشنایی مقدماتی با شبکه‌های اجتماعی

به [۱] هر ساختار اجتماعی از افراد که بر اساس یک رابطه اجتماعی ایجاد می‌شود، یک شبکه اجتماعی می‌گوییم. بنابراین هر شبکه اجتماعی شامل مجموعه‌ای از انسان‌ها و روابط اجتماعی بین آن‌هاست. لذا هر شبکه اجتماعی از دو عنصر تشکیل شده است: موجودیت‌های شرکت کننده در ارتباط و ارتباط بین این موجودیت‌ها. در علوم اجتماعی به موجودیت‌های شرکت کننده در ارتباط بازیگر و به ارتباطات بین این موجودیت‌ها رابطه گفته می‌شود [۱]. شبکه های اجتماعی به دو نوع برخط و برون‌خط تقسیم می‌شود. از شبکه‌های برون‌خط می‌توان به شبکه دوستان، شبکه همکاران و شبکه همکلاسی‌ها اشاره کرد. از شبکه های برخط می توان به شبکه های اجتماعی نظیر فیس بوک[۲] توییتر [۳] و گوگل پلاس [۴] اشاره کرد.

شبکه‌های اجتماعی از قرن نوزدهم مورد توجه قرار گرفت. پژوهش‌ها در این حوزه از دهه چهل به بعد با تعریف ابزارهایی چون گراف اجتماعی [۲] شتاب بیشتری گرفت. در سال ۱۹۹۴ واسرمن[۵] با چاپ کتاب تحلیل شبکه‌های اجتماعی [۱] این زمینه از علم را وارد دوره جدیدی کرد، و پس از آن شبکه‌های اجتماعی به صورت جدی در زیرمجموعه‌های علوم اجتماعی و ریاضی مورد بررسی قرار گرفت.

بحث‌های جسته و گریخته‌ای از سال ۱۹۶۰ در مورد شبکه‌های اجتماعی برخط به راه افتاد. نخستین شبکه اجتماعی در سال ۱۹۹۷ با نام سیکس دیگریز [۶] راه اندازی شد. اما انقلاب عظیم در هزاره دوم میلادی به وقوع پیوست جایی که از سال ۲۰۰۲ به بعد شبکه‌هایی نظیر فرنداستر [۷] اورکات [۸] و لینکداین [۹] روی وب قرار گرفتند. پدیده بزرگ دیگر شبکه اجتماعی فیس‌بوک بود که در سال ۲۰۰۴ توسط مارک زاکربرگ [۱۰] دانشجوی دانشگاه هاروراد بنا نهاده شد، که بعد از ده سال این شبکه اجتماعی با بیش از یک میلیارد نفر عضو به بزرگترین کشور چند ملیتی جهان تبدیل شده است. بعد از این سال شبکه‌هایی اجتماعی مجازی نظیر توییتر و گوگل پلاس نیز به وجود آمدند و بسیار مورد توجه قرار گرفتند.

[۱] برای دریافت داده‌ها و فایل‌های پیاده سازی شده با ایمیل h.emamgholizadeh@gmail.com تماس حاصل فرمایید.
[۲] www.facebook.com
[۳] www.twitter.com
[۴] www.plus.google.com
[۵] Wasserman
[۶] SixDegrees.com
[۷] Friendster
[۸] Orkut
[۹] LinkedIn
[۱۰] Mark Zuckerberg

جدول ۱-۱: ماتریس مربوط به شبکه اجتماعی با سه راس

Node	Bob	Carol	Ted
Bob	—	۰	۱
Carol	۱	—	۰
Ted	۰	۱	—

۱.۱.۱ شبکه‌های اجتماعی در ریاضی و کامپیوتر

برای مطالعه و تحلیل یک شبکه اجتماعی نیاز به ساختارهایی قابل فهم در علوم کامپیوتر و ریاضیات است که بازیگران شبکه و روابط آنها را نمایش دهد. به صورت عمده دو شیوه برای نمایش شبکه‌های اجتماعی مورد استفاده قرار می‌گیرد: گراف‌ها و ماتریس‌ها در این بخش به معرفی این دو ابزار می‌پردازیم.

ماتریس

می‌توانیم شبکه‌های اجتماعی را به صورت آرایش مربعی از اعداد نشان داد. در ماتریس هر سطر متعلق به یک راس (یا بازیگر) و به همان ترتیب هر ستون نیز متعلق به یک راس است. در ماتریس مربوطه به یک شبکه اجتماعی مولفه a_{ij} برابر یک است، اگر از راس i به راس j یک یال (ارتباط) موجود باشد. در غیر این صورت a_{ij} برابر صفر قرار داده می‌شود. برای مثال جدول ۱-۱ سه راس و ارتباط بین آن‌ها را در یک شبکه اجتماعی نشان می‌دهد.

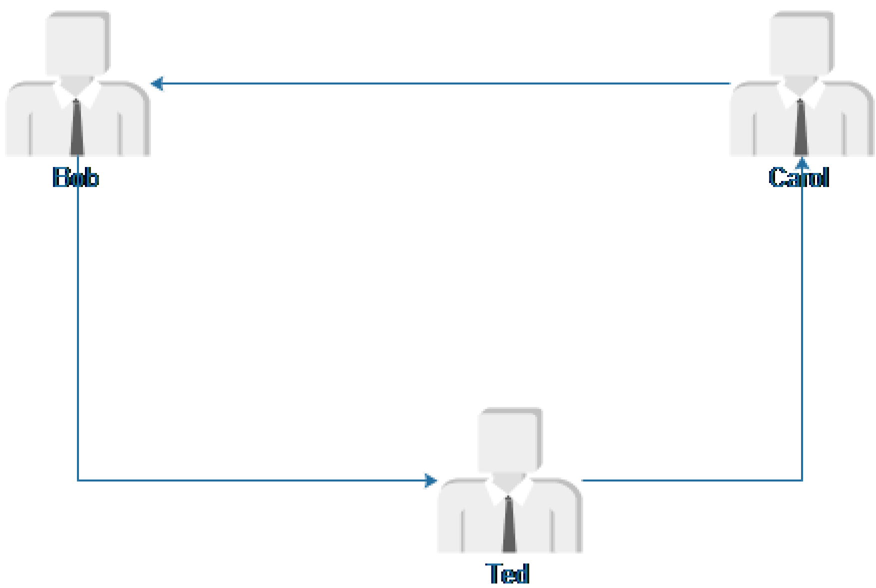
ماتریس‌های نشان دهنده شبکه‌های اجتماعی لزوما ماتریس‌های متقارنی نیستند. یعنی روابطی که توسط این ماتریس بیان می‌شود می‌تواند در یک جهت یا در هر دو جهت باشد. برای نشان دادن این موضوع ماتریس نشان داده شده در جدول ۱-۱ را بررسی می‌کنیم. اگر از هر یک از این سه نفر خواسته شود تنها یک نفر از دوستانشان را به عنوان نزدیکترین دوست گزارش کنند. بدین ترتیب باب، تد را نزدیکترین دوست خود می‌داند در حالی که تد کارول را نزدیکتر دوست خود معرفی کرده است. نزدیکترین دوست کارل بنابر آنچه گزارش شده باب می‌باشد.

در صورتی که ماتریس ارائه شده توسط یک شبکه اجتماعی متقارن باشد، آن شبکه اجتماعی دو طرفه در نظر گرفته می‌شود. می‌توان این موضوع را به صورت واضح در شبکه‌های اجتماعی مشاهده کرد. برای مثال شبکه

اجتماعی فیس بوک شبکه اجتماعی است که دو نوع ارتباط را امکان پذیر ساخته است اما به خاطر نسبت بسیار بالای روابط دو طرفه (دوست بودن) به رابطه یک طرفه (دنبال کردن) این شبکه اجتماعی، بیشتر به عنوان شبکه اجتماعی دو طرفه در نظر گرفته می‌شود. اما در شبکه‌هایی نظیر تویتر یا گوگل پلاس نیاز به ارتباط رابطه دو طرفه نیست و دنبال کردن به صورت یک طرفه انجام می‌شود. هرچند که با توجه به اینکه امروزه اغلب ارتباطات در شبکه‌های اجتماعی دو طرفه هستند در تحلیل این شبکه‌ها اجتماعی با توجه به نسبت کم ارتباطات یک طرفه به ارتباطات دو طرفه می‌توان همه ارتباطات را دو طرفه در نظر گرفت.

گراف

گراف‌ها به واسطه نمایش دادن بصری شبکه‌های اجتماعی به عنوان دیگر ابزار ریاضی برای نشان دادن و تحلیل شبکه اجتماعی به کار می‌روند. در این گراف‌ها بازیگران به عنوان راس‌های گراف در نظر گرفته می‌شوند و ارتباطات بین آن‌ها توسط یال‌های گراف به نمایش گذاشته می‌شود. گراف‌ها نیز همچون ماتریس‌ها برای نمایش هر دو نوع شبکه‌های اجتماعی (یک‌طرفه و دوطرفه) به کار برده می‌شوند. از گراف‌های جهت دار برای نمایش شبکه‌های اجتماعی یک طرفه و از گراف بدون جهت برای نمایش شبکه‌های اجتماعی دو طرفه استفاده می‌شود. برای مثال اگر ماتریس جدول ۱-۱ در نظر بگیریم. گراف متناظر آن به صورت گراف نشان داده شده در شکل ۱.۱ را خواهد بود.



شکل ۱.۱: گراف مربوط به ماتریس جدول ۱.۱

۲.۱.۱ تحلیل شبکه‌های اجتماعی

در ادامه این بخش به معرفی تحلیل شبکه‌های اجتماعی می‌پردازیم.

۳.۱.۱ آشنایی با تحلیل شبکه‌های اجتماعی

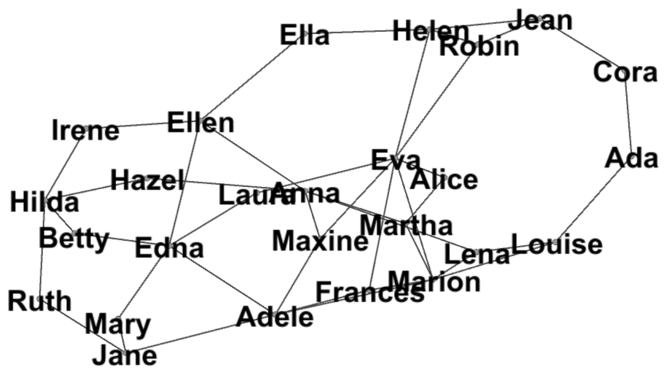
تحلیل شبکه‌های اجتماعی (SNA[۱۱]) خوانده می‌شود. منظور از تحلیل شبکه‌ها، مطالعه و تجزیه و تحلیل آنهاست. برای تحلیل شبکه روش‌های متفاوتی وجود دارد. این روش‌ها در ابتدا ریشه در جامعه‌شناسی و ریاضی (نظریه گراف) داشتند، اما امروزه علاوه براین دو رشته در سایر علوم نیز مورد استفاده قرار می‌گیرد[۳]. با استفاده از تحلیل شبکه‌های اجتماعی می‌توانیم به اطلاعات مفیدی از قبیل میانگین دوستان یک فرد در یک شبکه دوستی، یا میانگین فاصله هر دو نفر در یک شبکه همکاران دست یابیم.

۴.۱.۱ کاربردهای تجزیه و تحلیل شبکه‌های اجتماعی

توجه به این شاخه از دانش و پیشرفت سریع آن در محافل آکادمیک، به دلیل موارد استفاده و کاربرد آن می‌باشد. تجزیه و تحلیل شبکه‌های اجتماعی علاوه بر ریاضی و علوم‌اجتماعی در رشته‌هایی نظیر بیوانفورماتیک نیز کاربرد دارد. به عنوان یک مثال، شیوع بیماری مسری را بین افراد در نظر می‌گیریم. با توجه به محدودیت منابع سعی داریم از کم هزینه‌ترین روش‌ها برای جلوگیری از شیوع بیماری استفاده کنیم. شکل ۲.۱ را در نظر بگیرید. این شکل نشان دهنده یک گراف با ۲۶ راس و ۴۲ یال است. در این گراف اگر قرار باشد تنها سه نفر را در برابر ویروس واکسینه کنیم، واکسینه کردن، کدام افراد بیشترین بازده را برایمان درپی خواهد داشت؟ بهترین راه برای مقابله، واکسینه کردن اوا، مارینو و آنا است. زیرا اوا داری درجه ارتباطی ۷، مارینو دارای درجه ارتباطی ۶ و آنا دارای درجه ارتباطی ۵ است. حال اگر بخواهیم شایعه‌ای را در این گروه دوستی پخش کنیم باید دقیقا این سه فرد را مورد هدف قرار دهیم. این موارد تنها دو مورد از استفاده‌های پیش‌پا افتاده تحلیل شبکه‌های اجتماعی است. موارد حرفه‌ای تری را در علم داده می‌توان مشاهده نمود. داده‌ها در شبکه‌های اجتماعی بسیار بزرگ مانند فیس‌بوک با بیش از یک میلیارد راس و ده‌ها میلیارد یال به صورت گراف سامان دهی می‌شوند و تحلیل شبکه‌های اجتماعی برای تحلیل این داده‌ها به کار می‌رود.

---

[۱۱] Social Network Analysis



شکل ۲.۱: گراف ارتباطی با ۲۶ راس و ۴۲ یال

۵.۱.۱ مهمترین عناصر تحلیل شبکه‌های اجتماعی

در این قسمت می‌خواهیم به تعریف مفاهیمی بپردازیم که در تحلیل شبکه‌های اجتماعی به کار می‌روند. از این جمله می‌توان به شعاع، قطر، توزیع درجات، توزیع درجات گره‌های متصل، ویژگی استقلال از مقیاس، اشتراک‌پذیری، ضریب خوشه‌بندی، مرکزیت ارتباطی، ماجولاریتی[۱۲] و مولفه همبندی اشاره کرد، که در ادامه این بخش به تعریف این مفاهیم خواهیم پرداخت.

شعاع و قطر: برای تعریف شعاع و قطر باید نخست مفهوم گریز از مرکز یک راس را تعریف کرد. گریز از مرکز بیشترین فاصله‌ای است که آن راس از راس‌های دیگر می‌تواند داشته باشد. فاصله دو راس از هم برابر با طول کوتاهترین مسیر بین آن دو راس از میان تمام مسیرهاست. شعاع یک گراف کمترین مقدار گریز از مرکز کل گراف است. بدین معنا که شعاع نشان دهنده بیشنه فاصله مرکزی‌ترین راس از گراف با دیگر راس‌های گراف است. قطر یک گراف برابر با بیشترین مقدار گریز از مرکز در کل گراف است. یعنی قطر نشان دهنده بیشترین فاصله یک راس از راس‌های گراف که در دورترین فاصله نسبت به مرکز گراف قرار دارد[۴].

توزیع درجه: توزیع درجه یک گراف تابع $P(k)$ می‌باشد که نشان دهنده نسبت تعداد راس‌های با درجه k به تعداد کل راس‌های گراف است [۴]. توزیع درجه یکی از ویژگی‌های مهم هر گراف است. همانطور که در ادامه خواهیم دید گراف‌های مربوط به شبکه‌های اجتماعی ویژگی‌های مشابه و خاصی را از خود بروز می‌دهند. از جمله این‌که گراف‌های متناظر با شبکه‌های اجتماعی دارای توزیع درجه توانی هستند.

[۱۲] modularity

توزیع درجه راس‌های متصل: تابع توزیع درجه راس‌های متصل با $J(k,m)$ نشان داده می‌شود. تابع $J(k,m)$ نشان دهنده نسبت تعداد یال‌های بین راس‌های با درجه m و k به کل یال‌های گراف است. در گراف‌های جهت دار $J(k,m)$ نشان دهنده نسبت یال‌های جهت دار از راس‌ای با درجه k به راس‌ای با درجه m به کل یال‌های گراف است. تابع توزیع درجه راس‌های گراف که به اختصار با (JDD[۱۳]) نیز نشان داده می‌شود، معیاری است برای نشان دادن اینکه تا چه اندازه راس‌ها به راس‌های مشابه خود وصل شده‌اند.

اصولا محاسبه JDD نتایج آشفته و غیرقابل درکی را به همراه دارد. تابع $k_{nn}(k)$ برای بدست آوردن تقریبی از JDD به کار می‌رود؛ که نشان دهنده میانگین درجه راس‌های همسایه است [۵]. این تابع ابزاری مناسب برای نشان دادن میزان شرکت پذیری گراف است که در ادامه به تعریف این کمیت خواهیم پرداخت. اگر برای k های بزرگ مقادیر بزرگی برای $k_{nn}(k)$ بدست آید یعنی راس‌های با درجه بالا تمایل دارند با راس‌های با درجات بالای دیگر ارتباط داشته باشند و همانطور که در ادامه خواهیم دید، این ویژگی یکی از ویژگی‌های اصلی برای کوچک بودن میانگین کوتاه‌ترین فاصله بین دو راس در گراف است. اصولا شبکه‌های اجتماعی شرکت‌پذیر هستند، و در آن‌ها راس‌های با درجه بزرگتر تمایل دارند به راس‌های با درجه بزرگ دیگر متصل شوند.

ویژگی استقلال از مقیاس: یکی دیگر از معیارها که می‌توان توسط آن اشتراک‌پذیری یک گراف را بررسی کرد، ویژگی استقلال از مقیاس است. ویژگی استقلال از مقیاس $s(G)$ نخستین بار توسط لی جینیانگ [۱۴] و همکاران [۶] معرفی شد. $s(G)$ یک گراف مقداری است که مستقیما از روی توزیع درجه راس‌های متصل محاسبه می‌شود و مقداری بین ۰ و ۱ دارد. برای محاسبه $s(G)$ باید نخست $s'(G)$را تعریف کرد

$$s'(G) = \sum_{(i,j)\in E} d_i d_j \tag{۱-۱}$$

در این رابطه d_i و d_j به ترتیب درجه راس i و j را نشان می‌دهد و ، G نیز گراف مورد بررسی می‌باشد. معیار استقلال از مقیاس به صورت زیر تعریف می‌شود:

$$s(G) = \frac{s'(G)}{s'_{max}} \tag{۲-۱}$$

که در آن s'_{max} نشان دهنده بیشترین مقدار s' از بین تمام گراف‌هایی است که با G توزیع یکسانی دارند [۵]. بدین ترتیب هرچه $s(G)$ بزرگتر باشد، راس‌های با درجات بالا تمایل دارند با راس‌های با

[۱۳] joint degree distribution

[۱۴] Lee Jinyang

درجه بالای دیگر همسایگی داشته باشد، و هرچه این مقدار به ۰ نزدیک شود، راس‌های با درجات بالا تمایل دارند با راس‌های با درجه پایین در ارتباط باشند.

شرکت‌پذیری: شرکت‌پذیری اغلب به همبستگی بین یک جفت از راس‌ها تعبیر می‌شود. دو روش برای محاسبه شرکت‌پذیری وجود دارد که عبارتند از ضریب شرکت‌پذیری و اتصال همسایگی که در اینجا به معرفی ضریب شرکت‌پذیری که توسط نیومن[۱۵] [۷] تعریف شد، می‌پردازیم.

ضریب شرکت‌پذیری، ضریب همبستگی پیرسون [۱۶] درجات، بین جفت راس‌هایی است که در همسایگی هم قرار دارند. مقدار مثبت r همبستگی بین راس‌های با درجات مشابه را نشان می‌دهد. ضریب شرکت‌پذیری به صورت زیر نشان داده می‌شود:

$$r = \frac{\sum_{jk}(jk(e_{jk} - q_j q_k))}{\sigma_q^۲} \qquad (۳-۱)$$

که در آن q_k نشان دهنده توزیع درجات باقی راس‌هاست. بدین معنا که q_k نشان دهنده تعداد یال‌های خروجی از راس‌هایی است که جزو راس‌های مد نظر ما نیستند. مقدار نرمال شده q_k برابر است با:

$$q_k = \frac{(k+۱)p_k}{\sum_j jp_j} \qquad (۴-۱)$$

e_{jk} توزیع احتمال ارتباط راس‌هاست. یعنی احتمال اینکه با انتخاب یالی به تصادف درجه راس‌های دو سر آن برابر j و k باشد که

$$\sum_{jk} e_{jk} = ۱ \qquad \sum_j e_{jk} = q_k$$

اگر شبکه شرکت‌پذیری پایینی را از خود بروز دهد مقدار e_{jk} مقداری برابر $q_j q_k$ خواهد داشت. اگر گراف دارای شرکت‌پذیری بالایی باشد e_{jk} مقداری متفاوت از مقدار $q_j q_k$ خواهد داشت و e_{jk} را می‌توانیم از رابطه زیر بدست آوریم:

$$\langle jk \rangle - \langle j \rangle \langle k \rangle = \sum_{jk}(e_{jk} - q_j q_k) \qquad (۵-۱)$$

[۱۵] Newman

[۱۶] ضریب همبستگی پیرسون یک کمیت آماری است که همبستگی خطی بین دو متغیر را بدست می‌دهد، حاصلش مقداری بین ۱− و ۱ است و هرچه از ۱− به ۱ نزدیکتر می‌شویم همبستگی خطی مثبت بیشتر می‌شود.برای آشنایی بیشتر به ۳.۲.۱ مراجعه کنید.

که $\langle jk \rangle$ تعداد راس‌های در ارتباط با درجه j و k و j تعداد راس‌های با درجه jمی‌باشد. در انتها انحراف معیار را بدست می‌آوریم:

$$\sigma_q^2 = \sum_k k^2 q_k - [\sum_k k q_k]^2 \tag{۱-۶}$$

هرچقدر ضریب شرکت‌پذیری از ۱− به سمت ۱ نزدیک شود راس‌ها با راس‌های مشابه بیشتری در ارتباط هستند، و هرچه این مقدار به ۱- نزدیکتر شود راس‌ها با راس‌های مشابه کمتری در ارتباط هستند.

ضریب خوشه‌بندی: ضریب‌خوشه بندی مقیاسی است برای بررسی اینکه به چه نسبتی همسایه‌های یک راس با هم در ارتباط هستند. برای یک گراف، ضریب خوشه‌بندی، احتمال همسایه بودن دو راسی است که یک همسایه مشترک دارند. به صورت رسمی می‌توانیم ضریب خوشه‌بندی را به این طریق تعریف کنیم [۴]: ضریب خوشه‌بندی راس i، که با $c(i)$ نشان داده می‌شود، توسط نسبت بین تعداد یال‌هایی که بین دو راس همسایه i وجود دارد به تعداد کل یال‌هایی که می‌تواند بین کل همسایه‌های i موجود باشد، نشان داده می‌شود. لذا اگر بین همسایه‌های راس i به تعداد n یال موجود باشد، ضریب‌خوشه بندی آن برابر است با:

$$c(i) = \frac{n}{d_i(d_i - 1)} \tag{۱-۷}$$

ضریب‌خوشه بندی یک گراف میانگین ضریب خوشه بندی تک تک راس‌های گراف است که به صورت زیر تعریف می‌شود

$$C(G) = \frac{\sum_{v \in V} c(v)}{|v|}. \tag{۱-۸}$$

ضریب‌خوشه بندی یک گراف عددی مابین ۰ و ۱ است. زمانی که ضریب خوشه بندی یک راس برابر ۱ باشد، آن راس با همسایگان خود تشکیل یک دسته با سه راس (دسته به یک مجموعه راس گفته می‌شود که دو به دو با هم در ارتباط هستند) می‌دهد. زمانی که ضریب‌خوشه بندی یک گراف ۰ است هیچ مثلثی بین راس‌های آن گراف تشکیل نمی‌شود. منظور از مثلث یک مجموعه سه راسی است، که هرکدام با دو راس دیگر در ارتباط باشند.

مرکزیت ارتباطی: نیومن و گیروان[۱۷] [۸] برای ارائه الگوریتمی جهت یافتن مجموعه‌ای از راس‌ها که در یک گراف ارتباطات داخلی زیادی دارند، مفهوم مرکزیت ارتباطی برای یک گراف را تعریف کردند. مرکزیت

[۱۷] Girvan

ارتباطی برای یک یال عبارت است از تعداد کوتاهترین مسیرهای بین هر دو راس گراف که از این یال عبور می‌کند. برای دو راس از یک گراف که چندین کوتاهترین مسیر دارند برای هر مسیر وزنی منتسب می‌کنیم، بطوری که وزن مجموع مسیرها برابر با ۱ می‌باشد. مرکزیت ارتباطی یال e را می توانیم به صورت زیر نشان دهیم:

$$B(e) = \sum_{u\in V, v\in V} \frac{\sigma_e(u,v)}{\sigma(u,v)} \qquad (۱-۹)$$

که در این عبارت، $\sigma(u,v)$ تعداد کل کوتاهترین مسیرها بین راس‌های u و v و e نشان دهنده تعداد کوتاهترین مسیرهایی است که از e عبور می‌کنند. مرکزیت ارتباطی معیاری است برای نشان دادن اهمیت یک یال در گراف، هرچه یک یال بیشتر به عنوان قسمتی از کوتاهترین مسیرهای یک گراف استفاده شود، نقش مهمتری را در گراف بازی می‌کند. نیومن و گیروان [۸] نشان دادند که هرچه ضریب مرکزیت ارتباطی یک یال بیشتر باشد به احتمال بیشتری یک پل بین دو انجمن [۱۸] وجود دارد. تعریف مشابهی نیز برای راس ارائه شده است. مرکزیت ارتباطی یک راس تعداد کوتاهترین مسیرهایی بین راس‌های یک گراف که از یک راس خاص می‌گذرند.

جهان کوچک: واتس [۱۹] [۴۵] به معرفی خاصیت جهان کوچک شبکه‌های پیچیده پرداخت. زمانی که یک شبکه پیچیده دارای میانگین کوتاه‌ترین مسیرهای کوچک و ضریب‌خوشه‌بندی بزرگی باشد دارای خاصیت جهان کوچک است.

مولفه‌های همبندی: برای یک گراف بدون جهت یک مولفه همبندی عبارت است از مجموعه‌ای از راس‌ها که بین هر دوتای آن‌ها حداقل یک مسیر وجود دارد. اما برای گراف جهت دار باید بین دو مولفه همبندی قوی (SCC[۲۰]) و مولفه همبندی ضعیف (WCC[۲۱]) تفاوت قائل شد. مولفه همبندی قوی، مولفه‌ای است که بین هر دو راس یک مسیر جهت دار وجود داشته باشد. اما مولفه همبندی ضعیف، مولفه‌ای است که بین هر دو راس مولفه، بدون در نظر گرفتن جهت یال‌ها مسیری وجود داشته باشد.

نظریه پیوستگی: نظریه پیوستگی توسط باربارزی [۲۲] و آلبرت [۲۳] [۱۰] ارائه شد. نظریه پیوستگی، وابستگی درجه یک راس i نسبت به زمان را محاسبه می‌کند. درجه راس i یعنی k_i، هر گام زمانی با ورود یک

[۱۸] تعریف انجمن‌ها به طور کامل در ۳.۲.۱ آمده است

[۱۹] Watts

[۲۰] Strongly connected component

[۲۱] Weakly connected component

[۲۲] Barabási

[۲۳] Albert

راس جدید به سیستم و اتصال آن به i افزایش می‌یابد، احتمال این فرایند برابر $\Pi(k_i)$ می‌باشد. با فرض اینکه k_i یک متغیر حقیقی پیوسته است، انتظار می‌رود احتمال تغییر درجه i با $\Pi(k_i)$ متناسب باشد:

$$\frac{\partial k_i}{\partial t} = m\Pi(k_i) \tag{۱۰-۱}$$

با این کار یک متغیر گسسته به یک متغیر پیوسته تبدیل می‌شود. بدین وسیله می‌توانیم با انتگرال و مشتق گیری از آن تابع تجمعی و تابع احتمال این متغیر را بدست آوریم.

معادلات سرآمد: معادلات سرآمد توسط دوروگوتسف [۲۴] و همکاران [۱۱] ارائه شد. این روابط $p(k, t_i, t)$ احتمال برابر بودن درجه راس i با k در زمان t می‌باشد، در صورتی که راس i در زمان t_i به گراف اضافه شده باشد. به کمک این روش نیز می‌توانیم متغیرهای گسسته را به متغیرهای پیوسته تبدیل کنیم.

معادلات رشد: معادلات رشد که توسط کرابوسکی [۲۵] [۱۲] معرفی شد، روی میانگین تعداد راس‌ها با درجه k در زمان t تمرکز می‌کند ($N_k(t)$) و احتمال نمو این مقدار در یک بازه زمانی را بدست می‌آورد. این روش متغیر گسسته را به متغیر پیوسته تبدیل می‌کند.

تابع مولد: تابع مولد معمولی دنباله a_n برابر است با:

$$G(a_n; x) = \sum_{n=0}^{\infty} a_n x^n \tag{۱۱-۱}$$

تابع مولد توزیع درجات یک گراف نیز برابر است با:

$$G_0(x) = \sum_{k=0}^{\infty} P(k) x^k \tag{۱۲-۱}$$

یکی از ویژگی‌های مهم تابع مولد رابطه زیر می‌باشد:

$$\langle k \rangle = \sum_k kP(k) = G'_0(1) \tag{۱۳-۱}$$

یعنی میانگین درجات گراف با مشتق تابع مولد برای $x = 1$ برابر است.

حوزه میانی: زمانی که در شبکه‌های اجتماعی تکنیک حوزه میانی به کار می‌بریم در واقع چندین پیش‌فرض را در مورد شبکه‌های اجتماعی قبول می‌کنیم:

---
[۲۴] Dorogovtsev
[۲۵] Krapivsky

۱. نوسات احتمالاتی هیچ تاثیری بر روی مدل ندارند.

۲. فرض می‌کنیم درجه راس تنها ویژگی است که رفتار راس‌ها را تعیین می‌کند.

۳. فرض می‌کنیم راس‌ها هیچ وابستگی به هم ندارند.

زنجیره مارکوف: می‌توان زنجیره مارکوف را به این صورت تعریف کرد: مجموعه‌ای از حالت‌های $S = \{s_۱, s_۲, \ldots, s_r\}$ را داشته باشیم، فرایند از یکی از حالت‌ها شروع کرده و به صورت پی‌درپی به حالت‌های بعدی می‌رود. هر تغییر حالت را یک گام می‌نامیم. اگر زنجیره در حالت s_i باشد و در گام بعدی به حالت s_j برود احتمال این تغییر حالت را با p_{ij} نشان می‌دهیم. این احتمال به حالتی که زنجیره پیش از این قرار داشت ارتباطی ندارد (مستقل از آن است).

۶.۱.۱ نمونه‌ای از پژوهش‌های مشابه

در سال ۲۰۰۹ تویونن [۲۶] و همکاران [۱۳] مقایسه‌ای بین مدل‌های ارائه شده و داده‌های واقعی انجام دادند. تویونن و همکاران در این تحقیق از داده‌های وب‌سایت http://www.last.fm و شبکه ایمیل‌های فرستاده شده استفاده کردند. این تحقیق بر روی شبکه‌های اجتماعی برخط انجام شده است اما به نظر نمی‌رسد که پژوهش بیانگر تمام وجوه شبکه‌های اجتماعی برخط باشد. یکی از دلایل این ادعا مجموعه داده‌های مورد بررسی این پژوهش است. شبکه اجتماعی http://www.last.fm که یک وب‌سایت اشتراک گذاری موسیقی است نمی‌تواند برآیند و نشان دهنده شبکه‌های اجتماعی برخط همگانی مانند فیس‌بوک، توییتر و گوگل پلاس باشد، چون در این وب‌سایت‌ها افراد با تمام سلایق حضور دارند. لذا برای دستیابی به یک مدل بهتر و البته مقایسه واضح‌تر نیاز به بررسی مدل‌ها و مقایسه آن‌ها با داده‌های شبکه‌های اجتماعی فراگیرتری احساس می‌شود. این مورد بخصوص در بررسی‌ها و پژوهش‌های جامعه‌شناسان می‌تواند بسیار مفیدتر باشد. چون آن‌ها برای پیش‌بینی ویژگی‌ها و ساختارهای جامع، نیاز به مدل‌هایی دارند که انطباق بیشتری با شبکه‌های اجتماعی فراگیر داشته باشد. از طرف دیگر شبکه اجتماعی ایمیل دارای نقصان مضاعفی است. آنچنان که واضح است این شبکه اجتماعی نمی‌تواند اطلاعات چندانی از ساختارهای شبکه‌های اجتماعی بدهد، چون اساسا شبکه ایمیل‌های ارسالی بین افرادی که آشنایی پیشین دارند یا به نحوی از طریق چیزی غیر از فضای مجازی آشنایی دارند، تشکیل می‌شود. اما در شبکه‌های اجتماعی واقعی افراد از اقصی نقاط جهان با هم در ارتباطند بدون آنکه هیچ آشنایی پیشینی داشته باشند. بیشتر این ارتباطات به واسطه علایق مشترک هستند.

[۲۶] Toivonen

از طرف دیگر داده‌های واقعی مورد استفاده در پژوهش تویونن و همکاران دارای میانگین درجه و ضریب خوشه‌بندی کوچکی می‌باشند، لذا برای بررسی جامع‌تر ویژگی‌های مدل‌های ارائه شده، باید داده‌های واقعی با ضریب خوشه‌بندی و میانگین بزرگ‌تر از داده‌های واقعی مورد بررسی تویونن و همکاران تحلیل و بررسی شوند.

۷.۱.۱ اهداف پژوهش

در این پژوهش سعی داریم که نقصان‌های اشاره شده در بررسی تویونن و همکاران را رفع کنیم. بدین منظور از مجموعه داده‌هایی استفاده می‌کنیم که از شبکه‌های اجتماعی فراگیر بدست آمده باشد. وب‌سایت فیس‌بوک طبق آمار [۲۷] منتشر شده، از لحاظ تعداد کاربران در جایگاه نخست قرار دارند. در این پژوهش از داده‌های ارائه شده مربوط به شبکه اجتماعی فیس بوک در وب‌سایت دانشگاه آکسفورد [۲۸] استفاده شده است؛ تا بدین طریق مقایسه واقع‌بینانه‌تری بین مدل‌ها انجام شود. از طرف دیگر محوریت کار بر روی مقایسه بین انجمن‌های تشکیلی بنانهاده شده است، تا بدین وسیله انجمن‌های تشکیلی، که از مهمترین ویژگی‌های شبکه‌های اجتماعی است، مورد بررسی قرار دهیم. در بررسی انجمن‌های تشکیلی از برنامه CFinder [۲۹] استفاده شده است که دلیل این انتخاب در بخش ۲.۱ که به تفصیل به معرفی و مقایسه الگوریتم‌های ارائه شده برای یافتن انجمن‌های گراف پرادخته، ذکر خواهد شد.

در این پژوهش در انتخاب داده‌های واقعی علاوه بر عمومیت، معیارهای میانگین درجه و ضریب خوشه‌بندی را نیز مورد توجه قرار دادیم، به طوری که علاوه‌بر فیس بوک که عمویت، ضریب خوشه‌بندی و میانگین درجه بالایی داشت، از داده‌های مربوط به ویکی وُت [۳۰] که زیر مجموعه‌ای از وب سایت ویکی‌پدیا می‌باشد، استفاده کردیم، زیرا داده‌های مربوط به این شبکه میانگین بالا و ضریب خوشه‌بندی پایین‌تر (نسبت به داده‌های مربوط به پژوهش تویونن و همکاران) را داراست(داده‌ها از وب سایت http://snap.stanford.edu/data استخراج شده‌است).

۲.۱ انجمن و انجمن یابی

ماجولاریتی و انجمن نیز همانند اکثر مفاهیمی که در این بخش به آن‌ها پرداختیم از ویژگی‌های ساختاری و توپولوژیکی گراف‌هاست.

---
[۲۷] http://www.statista.com/statistics/272014/global-social-networks-ranked-by-number-of-users/

[۲۸] http://snap.stanford.edu/data/

[۲۹] Cfinder.org

[۳۰] http://www.wikipedia.org/

الگوریتم‌های انجمن یابی برای گراف‌های وزن‌دار و بی‌وزن، جهت‌دار و بی‌جهت و گراف‌های حاصل از ترکیب این‌ها ارائه شده است. ما به مرور از الگوریتم‌های ارائه شده برای انواع گراف‌ها خواهیم پرداخت. اما به علت نیازهای این پژوهش بیشتر بر روی گراف‌های بی‌وزن و بی‌جهت متمرکز خواهیم شد.
دو نوع انجمن وجود دارد انجمن‌هایی که با هم همپوشانی دارند و انجمن‌هایی که همپوشانی ندارند. گراف‌های مربوط به شبکه‌های اجتماعی اغلب گراف‌هایی هستند که همپوشانی دارند یعنی یک راس می‌تواند داخل چندین انجمن قرار بگیرد. در این بخش به مرور الگوریتم‌های ارائه شده برای یافتن انجمن‌های همپوشان و غیر همپوشان خواهیم پرداخت. در نهایت الگوریتم انتخابی خود را با دلیل انتخاب به صورت مبسوط شرح خواهیم داد.

۱.۲.۱ آشنایی با انجمن‌ها و تاریخچه آن‌ها

یکی از سخت‌ترین کارهای مربوط به انجمن‌یابی در همان گام نخست روی می‌دهد؛ سوال اساسی و پایه‌ای مربوط به تعریف انجمن‌هاست. شکل ۳.۱ تصویر یک زیرگراف از گراف فیس‌بوک است که از ۴۰۳۹ راس و ۸۸۲۳۴ یال تشکیل شده می‌باشد. هرچقدر به لحاظ بصری یافتن انجمن‌های این گراف راحت به نظر می‌رسد. اما به لحاظ علم ریاضی و نظریه گراف ارائه تعریف و یافتن این انجمن‌ها پیچیده می‌نماید. در گراف شکل ۳.۱ توده‌های مجتمع راس‌ها، انجمن‌ها را تشکیل می‌دهند. شکل ۴.۱ یک گراف تصادفی است با تعداد راس‌ها و یال‌های مساوی با گراف فیس‌بوک که برای هر یال دو راس را به طور تصادفی انتخاب کرده و به هم وصل می‌کند. همانطور که مشخص است. به لحاظ شهودی و تصویری گراف‌های مربوط به شبکه اجتماعی از گروه‌های متراکم زیادی تشکیل شده است اما گراف مربوط به گراف تصادفی تنها از یک توده تشکیل شده‌است. این دو گراف تفاوت‌های دیگری نیز دارند که در ادامه به آن خواهیم پرداخت، از جمله این تفاوت‌ها وجود چند راس (قطب [۳۱]) است که نقش ارتباطی را بازی می‌کنند.

به صورت شهودی توزیع یال‌ها در گراف‌های واقعی همانند گراف فیس‌بوک نه به صورت سراسری که به صورت محلی است. بدین معنی که تعداد یال‌های توزیع شده میان گروهی از راس‌ها بسیار بیشتر از تعداد یال‌های توزیع شده بین این گروه از راس‌ها با بقیه راس‌های گراف است. این ویژگی که در گراف‌های مربوط به داده‌های واقعی می‌توان مشاهده کرد، انجمن خوانده می‌شود. در برخی از منابع انجمن، خوشه یا ماجول نیز خوانده می‌شود. به عبارت دیگر انجمن‌ها مجموعه‌ای از راس‌هاست که با احتمال بیشتری نسبت به بقیه گراف ویژگی‌های مشترکی را به اشتراک می‌گذارند. انجمن‌ها و انجمن‌یابی چندین سال است که به صورت گسترده مورد مطالعه قرار می‌گیرد [۱۱ ، ۱۲ ، ۱۳ ، ۱۴]. در این پژوهش به گروه‌بندی که بین راس‌های گراف انجام می‌شود تقسیم‌بندی،

[۳۱] Hub

به هریک از این گروه‌ها در طول فرایند انجمن‌یابی خوشه و پس از اتمام فرایند انجمن می‌گویند. از کاربردهای انجمن‌یابی می‌توان به تبلیغات و بازاریابی اشاره کرد. از آنجایی که افراد حاضر در انجمن‌های تشکیل شده در یک شبکه اجتماعی به احتمال زیاد علایق مشترکی دارند، می‌توان با یافتن علایق آن‌ها از این اطلاعات به منظور تبلیغ محصولات خاص استفاده کرد. کاربردهای زیاد دیگری نیز می‌توان برای انجمن‌ها نام برد. همین کاربرد فراوان در زمینه‌هایی همچون زیست‌شناسی، مهندسی کامپیوتر، اقتصاد این شاخه از علم شبکه‌های اجتماعی و تئوری گراف را به زمینه‌ای محبوب برای پژوهشگران جهت تحقیق تبدیل کرده‌است.

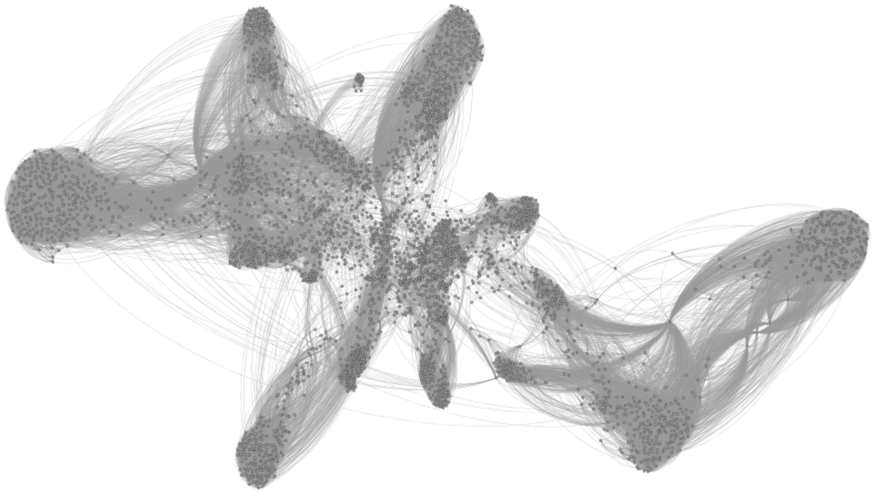
برخی شبکه‌های اجتماعی (و اغلب شبکه‌های اجتماعی برخط) که خاصیت انجمنی (یعنی دارای انجمن‌هایی هستند) را از خود بروز می‌دهند، دارای ساختار سلسله مراتبی هستند. بدین معنا که انجمن‌های بزرگتر از یک مجموعه انجمن‌های کوچکتر تشکیل شده و این انجمن‌های کوچکتر خود از یک دسته انجمن‌های کوچکتر دیگر تشکیل شده‌است. شبکه‌های اجتماعی اغلب این خاصیت را از خود بروز می‌دهند. برای مثال فرض کنید انجمن شامل دانش‌آموزان مربوط به یک مدرسه باشد. این گروه از اعضای یک شبکه دارای ارتباطات بیشتری نسبت به بقیه اعضای حاضر در یک شبکه‌اجتماعی هستند، حال در این مدرسه دانش‌آموزان مربوط به رده‌های تحصیلی مختلف ارتباطات بیشتری دارند. برای مثال دانش‌آموزان مربوط به کلاس‌های اول ارتباطات بیشتری باهم دارند و انجمن کوچکتری را تشکیل می‌دهند. در همین انجمن کوچکتر دانش‌آموزانی که در یک کلاس خاص (مثلا اول یک) هستند نیز ارتباطات بیشتری با هم دارند و تشکیل انجمن کوچکتری را می‌دهند که انجمن بزرگتر اجتماعی از این مجموعه‌هاست. برخی از الگوریتم‌های پیشنهاد شده برای انجمن‌یابی دقیقا به کاوش در این ساختار سلسله مراتبی می‌پردازند. هرچه گام‌های بیشتری اجرا شوند انجمن‌های کوچکتر را کشف می‌کنند.

به طور خلاصه می‌توان گفت که هدف الگوریتم‌های انجمن‌یابی کشف ساختار انجمنی، و در صورت وجود یافتن ویژگی سلسله مراتبی گراف‌ها تنها با اتکا به ساختار توپولوژیکی و اطلاعات ارائه شده توسط گراف است [۹].

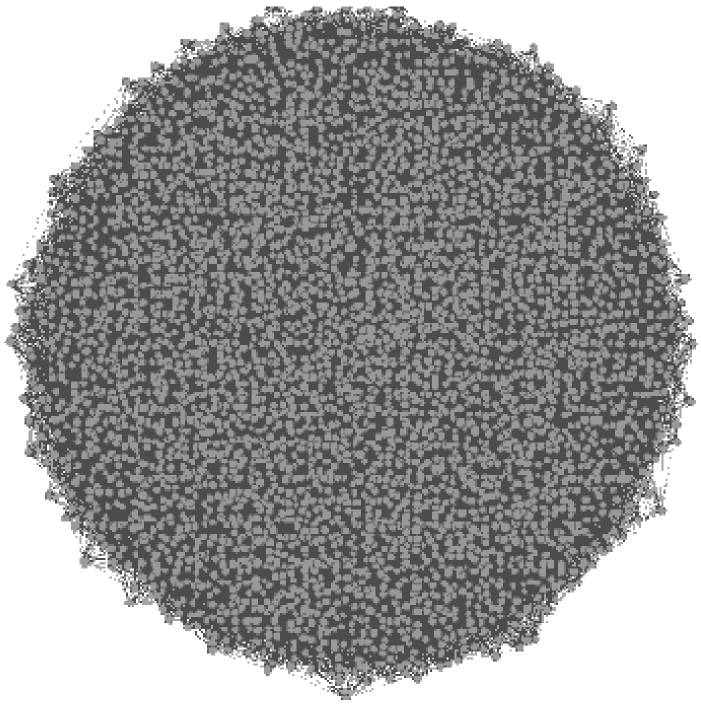
ریشه‌های انجمن‌یابی را می‌توان در [۱۸] پی گرفت که در آن ویس [۳۲] و جاکوبسن[۳۳] برای یافتن کارگروه‌های یک شرکت نخستین بار این مفهوم را به کار گرفتند. اهمیت این مقاله به این خاطر است که صورت اصلاح شده‌ای از الگوریتم ارائه شده توسط ویس و جاکوب امروزه در برخی از الگوریتم‌های انجمن‌یابی استفاده می‌شود. با ارائه مقاله [۸] و ارائه شیوه جدید انجمن‌یابی با استفاده از مرکزیت‌ارتباطی، انجمن‌یابی وارد فاز تازه‌ای شد و با یافتن ماجولاریتی بنیان‌های این علم محکم‌تر گردید.

[۳۲] Weiss

[۳۳] Jacobson



شکل ۳.۱: زیرگرافی از گراف کامل فیس‌بوک با ۴۰۳۹ راس و ۸۸۲۳۴ یال



شکل ۴.۱: گراف تصادفی ۴۰۳۹ راس و ۸۸۲۳۴ یال

۲.۲.۱ انواع گراف‌ها به عنوان ورودی الگوریتم‌های انجمن‌یابی

براساس انواع مختلف گراف انواع مختلفی از انجمن‌یابی‌ها وجود دارد. برای مثال برای گراف‌های جهت‌دار، الگوریتم‌های خاصی ارائه شده یا الگوریتم‌های ارائه شده برای گراف‌های بدون‌جهت به منظور استفاده در گراف‌های جهت دار اصلاح شده‌اند. برای گراف‌های وزن‌دار نیز یا الگوریتم‌های خاص ارائه شده یا الگوریتم‌های ارائه شده برای گونه‌های دیگر برای این نوع از گراف‌ها نیز سازگار شده‌اند.

علاوه بر نوع گراف که تعیین کننده نوع الگوریتم‌های مورد استفاده است، خود انجمن‌های تشکیل شده در گراف‌ها نیز به دو نوع تقسیم می‌شوند. انجمن‌های غیر همپوشان و انجمن‌های همپوشان. چون تعریف تقسیم‌بندی شامل راس‌هایی است که تنها به یک انجمن تعلق دارند، برای انجمن همپوشان به جای واژه قسمت از واژه پوشش استفاده می‌کنیم. حال برای گراف‌هایی که دارای انجمن‌های همپوشان هستند نیاز به الگوریتم‌های خاصی می‌باشد.

اغلب شبکه‌های اجتماعی برخط دارای انجمن‌های همپوشان هستند. یعنی این شبکه‌ها از انجمن‌هایی تشکیل شده‌اند که با هم همپوشانی دارند و یک راس می‌تواند به بیش از یک انجمن تعلق داشته باشد. گراف‌های تولید شده توسط مدل‌ها و همچنین گراف فیس بوک دارای انجمن‌های همپوشان هستند. معروفترین الگوریتم تعریف شده برای انجمن‌های همپوشان الگوریتم پالا [۳۴] [۱۹] می‌باشد. که به عنوان الگوریتم مورد استفاده ما برای یافتن انجمن‌ها در بخش ۶.۲.۱ معرفی خواهد شد.

۳.۲.۱ تعریف انجمن

همانطور که در ابتدای بخش ۱.۲.۱ نیز اشاره شد، یکی از دشواری‌های مربوط به انجمن‌ها ارائه تعریفی برای آن است. تعریف انجمن‌ها نیازمند مقداری چشم پوشی و اختیار است، تا از دشواری‌های کار کاسته شود. از طرفی اصولا تعریف دقیق انجمن‌ها برای کاربردهای مختلف می‌تواند متفاوت باشد.

اگر بین هر دو راس تابعی به عنوان تابع فاصله وجود داشته باشد، می‌توانیم انجمن‌ها را به عنوان مجموعه‌ای از راس‌ها تعریف کنیم که به فاصله کمی از هم قرار دارند. این تعریف بیشتر در خوشه‌بندی داده‌ها استفاده می‌شود. اما همچنان که پیش‌تر عنوان شد انجمن‌ها مجموعه‌ای از راس‌هاست که تعداد یال‌های داخل آن مجموعه (یعنی یال‌هایی که هر دو راس انتهایی آن داخل انجمن باشند) از یال‌های آن مجموعه، با دیگر راس‌های گراف به مراتب بیشتر باشد. این تعریف به عنوان پایه‌ای برای اغلب تعاریف دیگر به حساب می‌آید.

[۳۴] Palla

به منظور رسیدن به تعریف ریاضی‌وار از انجمن‌ها نیاز به آشنایی بیشتر با برخی مفاهیم داریم. فرض کنیم زیر گراف C از گراف G را داشته باشیم. که در آن $|C| = n_c$ و $|G| = n$، که به ترتیب نشان دهنده تعداد راس‌های زیرگراف C و تعداد راس‌های G است. درجه داخلی و خارجی $v \in C$ را با k_v^{in} و k_v^{ext} نشان داده می‌شوند که به ترتیب عبارتند از تعداد یال‌های خروجی از v که به راس‌های دیگر C متصل شده است، و تعداد یال‌های v که به راس‌های خارج از زیر گراف C متصل شده است. اگر $k_v^{ext} = ۰$ راس v تنها دارای همسایه‌هایی درون C است. لذا زیرگراف مذکور می‌تواند انجمن خوبی برای راس v باشد. اگر $k_v^{in} = ۰$ راس به صورت کامل از زیر گراف C جداست و باید به دنبال زیرگراف دیگری برای راس v بود. بنابراین، درجه خروجی زیرگراف C برابر مجموع درجات خروجی راس‌هایش $\sum_{k \in C} k_v^{ext}$ می‌باشد. مجموع درجه داخلی زیرگراف C برابر مجموع درجات داخلی راس‌هایش یعنی $\sum_{k \in C} k_v^{in}$ است. مجموع درجات راس‌های زیر گراف هم از جمع این دو مقدار بدست می‌آید.

چگالی داخل انجمن زیرگراف C را با $\delta_{int}(C)$ نشان می‌دهیم که برابر نسبت بین یال‌های داخلی زیرگراف C و کل تعداد یال‌های ممکن زیرگراف است:

$$\delta_{int}(C) = \frac{\#\ internal\ edge\ of\ C}{n_c(n_c - ۱)/۲} \qquad (۱۴-۱)$$

به طور مشابه، چگالی بین انجمنی زیرگراف C که با $\delta_{ext}(C)$ نشان داده می‌شود برابر نسبت بین تعداد یال‌هایی که از زیرگراف خارج شده است به تعداد کل یال‌هایی که راس‌های عضو C می‌توانستند با بقیه راس‌های موجود در گراف ایجاد کنند:

$$\delta_{ext}(C) = \frac{\#\ inter - cluster\ edge\ of\ C}{n_c(n - n_c)}. \qquad (۱۵-۱)$$

برای اینکه زیرگراف C را به عنوان انجمن در نظر بگیریم انتظار داریم $\delta_{int}(C)$ به صورت معناداری از چگالی یال‌های گراف G، $\delta(G)$ بزرگتر باشد. $\delta(G)$ برابر نسبت بین یال‌ها گراف به کل یال‌های ممکن برای گراف آن است:

$$\delta(G) = \frac{\#edge\ of\ G}{n(n - ۱)/۲}. \qquad (۱۶-۱)$$

از طرف دیگر در یک انجمن δ_{ext} باید بطور قابل توجهی از $\delta(G)$ کوچکتر باشد. یافتن زیرگراف‌هایی که مقدار بزرگتری برای $\delta_{int}(C)$ و مقدار کوچکتری برای $\delta_{ext}(C)$ داشته باشند از اهداف الگوریتم‌های انجمن‌یابی است.

یکی از ساده‌ترین شیوه‌های بیشینه کردن تفاضل این دو مقدار یعنی $\delta_{int}(C) - \delta_{ext}(C)$ می‌باشد [۲۰]. از طرف دیگر یکی از پایه‌ای‌ترین ویژگی‌های انجمن‌ها مولفه همبندی بودن زیرگراف است یعنی هیچ راس تنهایی (بدون یال) در زیرگراف موجود نباشد [۹].

در [۹]، سه نوع تعریف انجمن مشخص شده‌است: تعریف محلی، تعریف سراسری، تعریف برپایه شباهت راس‌ها. در ادامه این بخش به طور مفصل به این سه نوع تعریف خواهیم پرداخت که خلاصه شده مفاهیمی است که در مقاله مذکور آمده است.

تعاریف محلی

انجمن‌ها بخشی از گراف هستند که با بقیه گراف ارتباط(یال) محدودی دارند. در برخی موارد می‌توان ادعا کرد که آن‌ها ایالت‌هایی خودگردان هستند. بدین معنی که می‌توان آن‌ها را جدا از کل گراف در نظر گرفت. تعاریف محلی برروی زیرگراف‌های مورد مطالعه و احتمالا همسایه‌های بلافصل آن متمرکز می‌شوند، و بقیه گراف را در نظر نمی‌گیرند.

در این تعاریف چهار نوع معیار مطرح هستند:

۱. تقابل کامل

۲. دستیابی‌پذیری

۳. درجه راس‌ها

۴. مقایسه همبستگی داخلی و خارجی [۳۵]

زیر گرافی که نسبت به حداقل یکی از این ویژگی‌ها بیشینه باشد، یک انجمن است. بدین معنی که زیرگراف بزرگتری که شامل این زیرگراف باشد و برخی از این معیارها در آن‌ها صدق کند را نتوان یافت. در یک حالت به شدت سخت‌گیرانه انجمن را می‌توان به صورت زیرگرافی از یک گراف در نظر گرفت که همه راس‌های آن با هم در ارتباط هستند(دسته‌ها). این تعریف به عنوان تقابل کامل [۲۱] نیز شناخته می‌شود. این تعریف همانگونه که ذکر شد سختگیرانه است. حالتی را در نظر بگیرید که تمام راس‌های یک زیر گراف با هم در ارتباط باشند این زیر گراف یک دسته را تشکیل می‌دهد که مسلما می‌توان انجمن نامید. حال اگر یکی از یال‌ها را حذف کنیم باز این زیرگراف با توجه به تفاضل $\delta_{int}(C) - \delta_{ext}(C)$ انجمنی را تشکیل می‌دهد اما دیگر در

[۳۵] comparison of internal versus external cohesion

تعریف دسته صدق نمی‌کند. یعنی با توجه به تعریف ذکر شده دیگر زیرگراف حاصل نشان دهنده یک انجمن نیست. یکی دیگر از مشکلاتی که این معیار به وجود می‌آورد این است که در یک دسته همه راس‌ها قرینه هم هستند. اما از مباحث پیشین می‌دانیم که برخی از انجمن‌ها دارای ساختار سلسله مراتبی هستند لذا این معیار نمی‌تواند بیانگر همه ویژگی‌های یک انجمن باشد. یکی از روش‌ها برای بهبود اوضاع استفاده از تعاریف شبه-دسته‌هاست. یکی از این شبه دسته ها n-دسته‌ها هستند که جزو معیارهای دستیابی‌پذیری دسته‌بندی می‌شوند. یک n-دسته یک زیرگراف بیشینه است که فاصله هر دو راس آن در گراف اصلی بیش از n نباشد [۲۲]. این تعریف مقداری از سخت‌گیری‌های موجود کم می‌کند، اما هنوز مقداری از محدودیت‌های مربوط به فاصله بین دو راس در گراف را داراست. یکی از این محدودیت‌ها مربوط به این موضوع می‌شود که ممکن است یک مسیر از راس‌هایی استفاده کند که جزو راس‌های زیرگراف ما نیستند. این مسئله دارای دو مشکل است؛ نخست آنکه ممکن است قطر زیرگراف بیشتر از n باشد؛ دوم آنکه زیرگراف ممکن است اصلا زیرگرافی متصل نباشد. برای رهایی از این مشکلات مُککن[۳۶] [۲۳] دو جایگزین برای n-دسته ها پیشنهاد کرد این دو جایگزین عبارتند از n-قبیله و n-باشگاه. یک n-قبیله یک n-دسته است که قطر آن بیشتر از n نباشد. n-باشگاه یک زیر گراف بیشینه با قطر n است. این دو تعریف شباهت زیادی با هم دارند: تفاوت بین این دو در این است که n-قبیله زیرگراف بیشینه تحت قید یک زیرگراف n-دسته بودن است. در حالی کهn-باشگاه زیرگراف بیشینه تحت قید طول قطر می‌باشد.

معیارهای همبستگی بر پایه همسایگی یا مجاورت بنا نهاده شده‌است. ایده‌ی کلی این است که هر راس موجود در زیرگراف باید حداقل با یک کمینه تعداد راس، در زیرگراف همسایه باشد. بدین معنا که برای هر راس از بین همه همسایه‌های آن راس حداقل k تای آن‌ها عضو زیرگراف مورد نظر باشد. k-پیچیده [۲۴] به عنوان بیشینه زیرگرافی تعریف می‌شود که در آن هر راس با همه راس‌ها دیگر به جز حداکثر k راس در زیرگراف همسایه باشد. به صورت مشابه k-هسته یک زیرگراف بیشینه است، که در آن هر راس زیرگراف حداقل با k راس از زیرگراف همسایه می‌باشد [۲۵]. انجمن‌هایی که با توجه به این زیرگراف‌ها تشکیل می‌شوند به دلیل وجود یال‌های داخلی زیادتر دارای همبستگی بیشتری می‌باشند. k-هسته مشابه p-quasi است که زیرگرافی را تعریف می‌کند که درجه هر راس آن از $p(k-1)$ بزرگتر باشد که در آن p عددی حقیقی در بازه [۰, ۱] و k مرتبه زیرگراف می‌باشد[۲۶].

با در نظر گرفتن تراکم زیرگراف، یک زیرگراف انجمن خوانده می‌شود، هرگاه همبستگی داخل زیرگراف از همبستگی آن با بقیه گراف بیشتر باشد. بنابراین، باید همبستگی داخلی و خارجی زیرگراف را محاسبه کنیم.

[۳۶] Mokken

این یکی از معیارهایی است که در الگوریتم‌های ارائه شده اخیر استفاده شده‌است. دو نوع انجمن را می‌توان از هم تفکیک نمود: انجمن قوی و انجمن‌های ضعیف. یک LS-مجموعه [۲۷] یا یک انجمن‌قوی، زیرگرافی است که درجه داخلی هر راس زیرگراف از درجه خارجی کل زیرگراف بزرگتر باشد. این تعریف کمی سخت‌گیرانه به نظر می‌رسد و می‌توان تعریفی دیگر ارائه داد، که تا این حد سختگیرانه نباشد. یک زیرگراف، انجمن ضعیف نامیده می‌شود اگر درجه داخلی زیرگراف از درجه خارجی آن بزرگتر باشد.

تعریف دیگری از انجمن‌ها روی پایداری انجمن در برابر حذف یال‌ها و استفاده از مفهوم همبندی تمرکز دارد. همبندی یال برای یک جفت از راس‌های یک گراف، کم‌ترین تعداد یال‌هایی است که با حذف آن‌ها از گراف آن دو راس از هم جدا می‌شوند (یعنی مسیری برای رسیدن از یکی به دیگری وجود ندارد). یک مجموعه لامبدا [۳۷] زیرگرافی است که هر جفت راس متعلق به آن همبندی یال بزرگتری نسبت به دو راسی دارند که یکی متعلق یک زیرگراف باشد، ولی دیگری به آن زیرگراف تعلق نداشته باشد. آنچه که در مجموعه لامبدا نسبت به LS-مجموعه متفاوت است وجود احتمال جدا بودن راس‌ها در مجموعه لامبدا می‌باشد. به علاوه در این مجموعه، راس‌ها می‌توانند در هر فاصله‌ای از همدیگر واقع شوند. بدین ترتیب می‌توان انجمن را یک زیرگراف لامبدا تعریف کرد.

یکی دیگر از معیارهایی که برای شناسایی انجمن‌ها مورد استفاده قرار می‌گیرد، مجموعه معیار تناسب است. هرچه مقدار یک معیار تناسب برای یک زیرگراف بیشتر شود آن زیرگراف نمایانگر انجمن قوی‌تری است. یکی از معیارهای تناسب معیار چگالی داخلی زیرگراف [۳۸]، می‌باشد که با $\delta_{int}(C)$ نشان داده می‌شود. برای مثال یک زیرگراف را می‌توانیم انجمن بنامیم، هرگاه چگالی داخلی آن زیرگراف از یک مقدار مشخص مثلا ξ بیشتر باشد. معیار تناسب به همراه همبندی نیز می‌تواند برای یافتن انجمن‌ها به کار رود. یک انجمن خوب انجمنی است که برش مینیمم [۳۹] آن اندازه کوچکتری داشته باشد.

تعاریف سراسری

در تعاریف محلی زیرگراف به عنوان یک جزء به صورت مجزا بررسی می‌شد. اما انجمن‌ها می‌توانند با توجه به کل گراف نیز تعریف شوند. بسیاری از تعاریف سراسری، به صورت غیرمستقیم به شناسایی انجمن‌ها می‌پردازند؛ بدین ترتیب که برخی ویژگی‌های سراسری گراف در الگوریتم‌ها بکار می‌روند و خروجی این الگوریتم‌ها به عنوان

[۳۷] lambda

[۳۸] intra-cluster density

[۳۹] تعداد یال‌هایی که با حذف آن یک ارتباط بین راس‌های یک زیرگراف مورد مطالعه با بقیه گراف قطع می‌شود. در ارتباط با زیرگراف، تعداد یال‌هایی است که با حذف آن‌ها زیرگراف از بقیه گراف جدا می‌شود.

انجمن در نظر گرفته می‌شود. همچنین یک دسته از تعاریف وجود دارند، که انجمن‌ها را براساس میزان تفاوتی که گراف داده شده با گراف تصادفی دارد شناسایی می‌کنند؛ زیرا انتظار نمی‌رود که گراف تصادفی دارای انجمن باشد. براساس آنچه در معرفی مدل‌ها خواهد آمد الحاق امتیازی به وجود آورنده راس‌های قطب [۴۰] می‌باشند. زمانی که افزودن یک یال با توجه به الحاق امتیازی انجام شود، راس‌های با درجه بیشتر با احتمال زیادتری به عنوان مقصد یا مبدا یال جدید انتخاب می‌شوند. قطب‌ها نیز به نوبه خود به وجود آورنده انجمن‌ها می‌باشند. همین موضوع باعث به وجود آمدن تفاوت بین گراف‌های تصادفی، که در آن یال‌ها به تصادف و احتمال مساوی بین راس‌ها ایجاد می‌شوند، و گراف‌هایی دارای ویژگی الحاق امتیازی می‌شود. می‌توان مدل‌های محض را در نظر گرفت که در آن‌ها مدل محض دارای برخی از ویژگی‌های ساختاری (برای مثال توزیع درجات مشابه) گراف اصلی می‌باشد. اما بدون در نظر گرفتن این ویژگی‌ها مدل محض به یک گراف تصادفی تبدیل می‌شود. این مدل محض به عنوان معیاری برای شناسایی انجمن‌ها استفاده می‌شود. برای مثال الگوریتم نیومن و گیروان [۸] برپایه یک مدل محض به شناسایی انجمن‌ها می‌پردازد.

تعریف انجمن‌ها برپایه شباهت گره‌ها

این دسته از تعاریف برپایه این فرض بنا نهاده‌شده‌اند که انجمن‌ها گروهی از راس‌ها هستند که به هم شبیه‌اند. بدین منظور لازم است معیارهایی برای تعیین میزان شباهت راس‌ها درنظر گرفته شود.

اگر بتوانیم راس‌ها را به روی یک فضای اقلیدسی n-بعدی نگاشت کنیم، می‌توانیم معیار شباهت را فاصله دو راس از هم در نظر بگیریم. هرچه فاصله راس‌ها از هم بیشتر شود شباهت آن‌ها به هم کمتر می‌شود. برای مثال می‌توانیم از نرم دوم برای محاسبه فاصله بین دو رأس A و B استفاده کرد.

اگر $A = (a_۱, a_۲, \dots, a_n)$ و $B = (b_۱, b_۲, \dots, b_n)$ باشند، فاصله بین A و B با کمک نرم دو به این صورت محاسبه می‌شود:

$$d^E_{AB} = \sqrt{\sum_{k=۱}^{n} (a_k - b_k)^۲} \qquad (۱۷-۱)$$

معیارهای دیگری نظیر نرم یک و نرم بی‌نهایت برای محاسبه فاصله بین دو راس نیز وجود دارد.

اگر امکان نگاشت راس‌ها برروی صفحه اقلیدسی وجود نداشته باشد. می‌توانیم از ارتباطات مجاورتی، به منظور محاسبه معیاری از شباهت دو رأس با یکدیگر، استفاده کرد فاصله دو راس i و j بدون نگاشت روی صفحه به

[۴۰] قطب: راس‌هایی با تعداد همسایه‌های بسیار زیاد می‌باشد.

وسیله رابطه زیر محاسبه می‌شود:

$$d_{i,j} = \sqrt{\sum_{k\neq i,j}(A_{ik} - A_{jk})^{۲}} \tag{۱-۱۸}$$

که در این رابطه A ماتریس مجاورت می‌باشد. این معیار براساس شباهت ساختاری تعریف می‌شود. دو راس به لحاظ ساختاری یکسان هستند، هرگاه همسایه‌های یکسانی داشته باشند حتی اگر با هم در ارتباط نباشند. لذا اگر i و j به لحاظ ساختاری معادل هم باشند $d_{ij} = ۰$ می‌شود. هرچه d_{ij} کوچکتر باشد این دو راس به لحاظ ساختاری به هم شبیه‌تر هستند. هرچه درجه دو راس بزرگتر و همسایه‌ها متفاوت‌تر باشند مقدار بزرگتری برای d_{ij} بدست می‌آید، یعنی دو راس در فاصله زیادتری نسبت به هم قرار می‌گیرند. معیار دیگری از شباهت دو راس، به صورت نسبت بین تعداد همسایه‌های همپوشان به مجموع همسایه‌های دو راس تعریف می‌شود، اگر $\Gamma(i)$ و $\Gamma(j)$ مجموعه همسایه‌های راس‌های i و j باشند داریم:

$$\omega_{ij} = \frac{|\Gamma(i) \cap \Gamma(j)|}{|\Gamma(i) \cup \Gamma(j)|} \tag{۱-۱۹}$$

معیار دیگری از شباهت دو راس که مشابه معیارهای پیشین است، همبستگی پیرسون بین ستون‌ها یا سطرهای ماتریس مجاورت می‌باشد:

$$C_{ij} = \frac{\sum_k (A_{ik} - \mu_i)(A_{jk} - \mu_j)}{n\sigma_i\sigma_j} \tag{۱-۲۰}$$

که در آن میانگین برابر $\mu_i = (\sum_j A_{ij})/n$ و انحراف از معیار آن برابر $\sigma_i = \sqrt{\sum_j (A_{ij} - \mu_i)^{۲}/n}$ می‌باشد.

دسته‌ای دیگر از معیارهای شباهت معیارهایی هستند که ویژگی‌های مرتبط با گشت عابر تصادفی در گراف را به کار می‌گیرند. یکی از ویژگی‌ها زمان گشت عابر تصادفی بین دو راس گراف است، که به صورت میانگین تعداد گام‌های مورد نیاز عابر تصادفی برای رسیدن از یک راس به دیگری و بازگشت به نقطه شروع تعریف می‌شود. هرچه این معیار برای دو راس بزرگتر باشد فاصله بین آن دو بیشتر است.

۴.۲.۱ معیارهای کیفیت انجمن‌های یافت شده

بعد از یافتن انجمن‌ها نیاز به معیاری داریم تا به ما نشان بدهد که انجمن بدست آمده دارای چه کیفیتی است. این کار توسط توابع کیفیت انجام می‌شود. می‌توان توابع کیفیت را به صورت توابعی در نظر گرفت که برای هر انجمنی یک مقدار به عنوان خروجی در نظر می‌گیرد و بر اساس این خروجی ما توانایی ارزیابی کیفیت یک تقسیم را داریم. اگر $q(C)$ تابع کیفیت باشد، کیفیت یک تقسیم برابر است با:

$$Q(P) = \sum_{C \in P} q(C) \tag{۲۱-۱}$$

در این تابع C معرف خوشه‌های یک گراف است. از معیارهای تناسب می‌توانیم به عنوان توابع کیفیت خوشه‌ها استفاده کرد.

یکی از معروف‌ترین توابع کیفیت تابع عملکرد P می‌باشد. این تابع نسبت بین، تعداد جفت راس‌های موجود در یک انجمن که بین آن‌ها یالی وجود دارد به اضافه تعداد جفت راس‌هایی که در دو انجمن متفاوت قرار دارند و بین آن‌ها یالی وجود ندارد، به کل تعداد یال‌های ممکن گراف را به عنوان معیار کیفیت انجمن در نظر می‌گیرد برای تقسیم‌بندی G داریم:

$$P(G) = \frac{|\{(i,j) \in E, C_i = C_j\}| + |\{(i,j) \notin E, C_i \neq C_j\}|}{n(n-۱)/۲}. \tag{۲۲-۱}$$

تابع پوشش یکی دیگر از توابعی است که می‌توان از آن به عنوان تابع کیفیت استفاده کرد این تابع نسبت بین تعداد یال‌های داخل انجمن را به تعداد کل یال‌های گراف به عنوان خروجی تابع کیفیت ارائه می‌دهد.

اما محبوب‌ترین تابع کیفیت [۹] ماجولاریتی است که که توسط نیومن و گیروان در سال ۲۰۰۴ [۸] ارائه شد. در یک شبکه اجتماعی که دارای انجمن‌هایی است، احتمال وجود یال بین راس‌ها یکسان نیست. زیرا می‌دانیم که احتمال به اشتراک گذاشتن یک یال بین دو راسی که در یک انجمن قرار دارند از احتمال اشتراک گذاری یک یال بین دو راسی که در انجمن‌های متفاوتی قرار دارند بیشتر است. پس می‌توان انتظار داشت که در یک گراف تصادفی هیچ انجمنی وجود نداشته باشد. ماجولاریتی براساس این ایده ارائه شد. لذا با مقایسه بین چگالی زیرگراف معادل یک گراف تصادفی نسبت به چگالی یک خوشه در گراف اصلی، می‌توان به معیاری برای کیفیت انجمن رسید. گراف تصادفی مورد نیاز برای مقایسه را می‌توان با کمک مدل محض بدست آورد. مدل محض یک گراف تصادفی است که برخی از ویژگی‌های گراف مورد بررسی را حفظ می‌کند. چون یال‌ها به صورت تصادفی اضافه می‌شود انتظار داریم که مدل محض هیچ انجمنی نداشته باشد. می‌توانیم تابع ماجولاریتی را به

صورت زیر نوشت:

$$Q = \frac{۱}{۲m} \sum_{ij} (A_{ij} - P_{ij}) \delta(C_i, C_j) \quad (۱-۲۳)$$

که مجموع روی همه جفت راس‌ها اجرا می‌شود، A ماتریس مجاورت، و m تعداد یال‌های گراف می‌باشد، و P_{ij} مقدار مورد انتظار وجود یال بین راس i و j را در مدل محض نشان می‌دهد. در این رابطه تابع δ برابر ۱ است هرگاه C_i و C_j یعنی خوشه‌ای که i در آن قرار دارد و خوشه‌ای که j در آن قرار دارد یکی باشد، یعنی i و j در یک خوشه باشند، در غیر این صورت مقدار این تابع برابر صفر است.

انواع مختلفی از مدل‌های محض را می‌توانیم معرفی کنیم از جمله می‌توانیم تعدا کل یال‌های گراف مدل محض را با گراف مدل مورد بررسی یکی بگیریم که در این صورت مقدار مورد انتظار را برای وجود یال بین دو راس i و j را به صورت زیر بدست آوریم:

$$P_{ij} = p = ۲m/[n(n-۱)], \forall i, j \quad (۱-۲۴)$$

این مدل محض گرافی تصادفی را ارائه می‌دهد چون گراف تصادفی دارای توزیع درجات پوآسون می‌باشد نمی‌تواند به خوبی ویژگی‌های یک گراف مربوط به شبکه اجتماعی را بازسازی کند. زیرا گراف‌های مربوط به شبکه‌های اجتماعی اصولا توزیع درجات توانی با $۲ \leq \gamma \leq ۱$ را دارا هستند[۹]. لذا نیاز به مدل محضی داریم تا ویژگی توزیع درجات گراف را به خوبی بروز دهد. بدین منظور می‌توانیم از مدل‌های محضی استفاده کنیم که دارای توزیع درجات یکسانی با گراف مورد بررسی باشد.

در مدل محض استانداردی که در ماجولاریتی استفاده می‌شود، مقدار مورد انتظار برای دنباله درجات با دنباله درجات گراف یکسان است. در این مدل محض یک راس می‌تواند به هر راس دیگری متصل باشد و احتمال اتصال دو راس به واسطه یک یال به درجه راس‌ها وابسته است. برای ارتباط راس‌های i و j در مدل محض باید دو نیم یال خروجی از این راس‌ها به هم برسند. احتمال اینکه راس i با درجه k_i دارای نیم یال باشد برابر $k_i/۲m$ و احتمال اینکه راس j با درجه k_j دارای نیم یال باشد، برابر $k_j/۲m$ می‌باشد، بدین ترتیب احتمال وجود یالی بین دو راس i و j برابر با ضرب احتمال‌های وجود نیم یال $p_i p_j$ می‌باشد، که این مقدار برابر $k_i k_j/۴m^۲$ است، حال مقدار مورد انتظار وجود یال بین دو راس i و j برابر است با $۲m p_i p_j = k_i k_j/۲m$.
با توجه به این مقدار ماجولاریتی برابر است با:

$$Q = \frac{۱}{۲m} \sum_{ij} (A_{ij} - \frac{k_i k_j}{۲m}) \delta(C_i, C_j) \quad (۱-۲۵)$$

ماجولارتی همیشه دارای مقداری کوچکتر از یک است. ماجولاریتی، همچنین می‌تواند دارای مقداری منفی نیز باشد. برای مشاهده این حالت تقسیمی را در نظر بگیرید که هر راس آن یک انجمن مجزا باشد در این حالت مجموع روی تمام راس‌ها محاسبه می‌شود، که هر یک دارای مقدار صفر یا منفی است. لذا می‌توان گفت، گرافی که دارای هیچ تقسیمی با مقدار مثبت نباشد، هیچ انجمنی ندارد.

می‌توانیم به جای محاسبه کل جفت راس‌ها فقط جفت راس‌هایی که در یک انجمن قرار دارند را در محاسبات وارد کنیم. در این صورت می‌توان فرمول ۱-۲۵ را به صورت زیر بازنویسی کرد:

$$Q = \sum_{c=۱}^{n_c}\Big[\frac{l_c}{m} - \Big(\frac{d_c}{۲m}\Big)^۲\Big] \tag{۱-۲۶}$$

که در این رابطه n_c تعداد خوشه‌ها، l_c تعداد کل یال‌هایی است که راس‌های موجود در خوشه c را به هم متصل می‌کند. d_c مجموع درجات راس‌های خوشه c است. حال از این رابطه استفاده می‌کنیم تا بیشینه مقدار ماجولاریتی را بدست آوریم:

$$\begin{aligned} Q &= \max_p \Big\{ \sum_{c=۱}^{n_c} \Big[\frac{l_c}{m} - \Big(\frac{d_c}{۲m}\Big)^۲\Big] \Big\} \\ &= \frac{۱}{m} \max_p \Big\{ \sum_{c=۱}^{n_c} \big[l_c - Ex(l_c)\big] \Big\} \\ &= -\frac{۱}{m} \min_p \Big\{ -\sum_{c=۱}^{n_c} \big[l_c - Ex(l_c)\big] \Big\} \end{aligned} \tag{۱-۲۷}$$

که $Ex(l_c) = \frac{d_c}{۲m}$ می‌باشد. بدین‌ترتیب می‌توانیم بیشنه مقداری را برای ماجولاریتی بدست آوریم. این مقدار بیشینه می‌تواند دو کاربرد برای ما داشته باشد کاربرد اول در الگوریتم‌های انجمن‌یابی است. پس از معرفی ماجولاریتی یک دسته از الگوریتم‌ها معرفی شدند که به کمک آن‌ها می‌توان انجمن‌های یک گراف را بدست آورد. بهینه‌سازی مقدار ماجولاریتی یکی از پایه‌های اصلی این الگوریتم‌هاست. کاربرد دوم مقایسه مقدار ماجولاریتی بیشینه با مقدار ماجولاریتی انجمن حاصل از یک الگوریتم به منظور بررسی کارایی الگوریتم و کیفیت انجمن بدست آمده می‌باشد.

ماجولاریتی مشکلاتی را به همراه دارد. یکی از مشکلات مربوط به بیشینه مقدار می‌باشد و اینکه این بیشینه

مقدار تا چه اندازه می‌تواند مورد اعتماد باشد. بیشینه مقدار باید بین صفر و یک باشد. زیرا می‌دانیم بیشینه مقدار کمتر از ۱ است، این مورد با بررسی فرمول ماجولاریتی واضح است. از طرف دیگر بیشینه مقدار باید بزرگتر از صفر باشد، زیرا اگر گراف را به صورت یک انجمن در نظر بگیریم، مقدار ماجولاریتی صفر خواهد داشت. یکی از موارد اساسی که به صورت پیش فرض در ارائه رابطه‌ای برای محاسبه ماجولاریتی در نظر گرفته شد، عدم وجود انجمن در گراف‌های تصادفی است. اما همانطور که گویمرا [۴۱] و همکاران [۲۸] نشان دادند گراف‌های تصادفی ممکن است دارای مقدار ماجولاریتی بزرگی باشد. این پدیده بدلیل نوساناتی که در توزیع یال‌ها بین راس‌ها در گراف تصادفی روی می‌دهد به وجود می‌آید. بدین طریق تمرکز یال‌ها در یک بخش از گراف باعث به وجود آمدن انجمن‌ها می‌شوند.

برای حل این مشکل می‌توانیم کیفیت بیشینه مقدار بدست آمده را از رابطه زیر بدست آوریم:

$$z = \frac{Q_{max} - \langle Q \rangle_{NM}}{\sigma_Q^{NM}} \qquad (۲۸-۱)$$

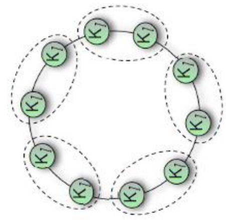
در این رابطه Q_{max} بیشنه مقدار ماجولاریتی برای گراف، $\langle Q \rangle_{NM}$ میانگین بیشنه مقدار ماجولاریتی روی تعداد مشخصی از گراف‌های تصادفی تولید شده به وسیله مدل محض و σ_Q^{NM} انحراف از معیار استاندارد می‌باشد. اگر $z \gg ۱$ آن‌گاه Q_{max} نشان‌دهنده یک تقسیم با انجمن قوی است. z در اکثر موارد مقداری بین ۲ و ۳ را داراست.

اما مشکل بنیادی‌تر به این برمی‌گردد که ماجولاریتی تا چه اندازه قابلیت کشف و شناسایی انجمن‌های خوب را دارد. این مشکل نخستین بار توسط فورتناتو [۴۲] و بارتلمی [۴۳] در سال ۲۰۰۷ [۲۹] مورد بررسی قرار گرفت. مدل محضی که برای محاسبه ماجولاریتی به کار می‌رود، فرض می‌کند که هر راس i بتواند به هر راس دیگر j دسترسی داشته باشد، و مقدار مورد انتظار برای وجود یالی بین آن‌ها برابر $p_{ij} = k_i k_j / ۲m$ می‌باشد. بطور مشابه مقدار مورد انتظار وجود یال بین دو خوشه A و B با مجموع درجات K_A و K_B برابر $P_{AB} = K_A K_B / ۲m$ است. حال اختلاف بین ماجولاریتی که از ادغام خوشه A و B بدست آمده است با ماجولاریتی که بدون ادغام این دو خوشه بدست آمده برابر است با:

$$\Delta Q_{AB} = \frac{l_{AB}}{m} - \frac{K_A K_B}{۲m^۲} \qquad (۲۹-۱)$$

در این رابطه l_{AB} تعداد یال‌هایی است که خوشه A را به خوشه B متصل می‌کند. اگر $l_{AB} = ۱$ باشد، یعنی فقط یک یال دو خوشه را به هم متصل کند، آن‌گاه به صورت واضح این دو خوشه از هم جدا هستند و انتظار

[۴۱] Guimera
[۴۲] Fortunato
[۴۳] Barthelemy



شکل ۵.۱: این شکل در [۲۹] برای نشان دادن مشکل تفکیک‌پذیری در ماجولاریتی ارائه شد.

داریم ماجولاریتی گراف در حالی که دو خوشه از هم جدا هستند، بیشتر از ماجولاریتی گراف در حالت ادغام دو خوشه باشد. برای مثال، اگر $K_AK_B/۲m < ۱$ آن‌گاه $\Delta Q_{AB} > ۰$. حال اگر به منظور راحت‌تر شدن محاسبات $K_A \sim K_B = K$ بدین معنا که زیرگراف‌ها تقریبا مجموع درجات برابری را دارا هستند. حال می‌توان نتیجه گرفت، اگر $K <\sim \sqrt{۲m}$ آن‌گاه ماجولاریتی گرافی که از ادغام دو خوشه بدست می‌آید، بزرگ‌تر از ماجولاریتی گرافی است که دو خوشه در آن از هم مستقل هستند [۲۹] .
به صورت شهودی می‌توان گفت: اگر تعداد یال‌های موجود بین دو خوشه بیش از مقدار مورد انتظار یال‌ها باشد، یک همبستگی ساختاری نیرومند بین زیرگراف‌ها برقرار است. اما اگر زیرگراف‌ها به مقدار کافی کوچک باشند (به لحاظ مجموع درجات)، مقدار مورد انتظار برای وجود یال‌ها بین دو زیرگراف می‌تواند کوچکتر از ۱ باشد. لذا حتی با وجود تنها یک یال بین دو زیرگراف این دو زیرگراف بعد از ادغام دارای ماجولاریتی بزرگتری هستند. نکته جالب این است که این خاصیت مستقل از ساختار زیرگراف می‌باشد زیرا این مشکل حتی برای دسته‌ها که قوی‌ترین خوشه‌ها را تشکیل می‌دهند، نیز صادق است. در شکل ۵.۱، گراف از n_c دسته مساوی تشکیل شده‌است که هر دسته دارای l راس است. این دسته‌ها به وسیله یک یال، چنان که در شکل نشان داده می‌شود به هم متصل شده‌اند. هرگاه n_c از $l^۲$ بزرگتر باشد، ماجولاریتی برای تقسیمی که همه این خوشه‌ها را در یک گروه قرار دهد، بیشتر از دیگر تقسیم‌ها خواهد بود.

این مشکل ماجولاریتی به خاطر تعریف مدل محض می‌باشد. نقطه ضعف مدل محض در این است که راس‌ها با همه راس‌های دیگر قادر به تعامل هستند، بدین معنا که هر قسمت از گراف همه اطلاعات مربوط به دیگر قسمت‌های گراف را در گستره دیدش دارد. باید محدودیتی برای وسعت دید هر قسمت از گراف قائل باشیم، اما تا کنون مدل محضی که دارای این ویژگی باشد ارائه نشده است.
با وجود همه این موارد که به عنوان نواقص تابع کیفیت ماجولاریتی برشمردیم تا کنون ماجولارتی پرکاربردترین و بهترین تابع کیفیت ارائه شده می‌باشد [۹].

۵.۲.۱ انواع الگوریتم‌های انجمن یابی

الگوریتم‌های ارائه شده برای انجمن‌یابی را می‌توان به شش گروه تقسیم کرد:

۱. الگوریتم‌های تقسیم
۲. الگوریتم‌های براساس ماجولاریتی
۳. الگوریتم‌های براساس شعاع طیفی مقادیر ویژه
۴. الگوریتم‌های پویا
۵. الگوریتم‌های براساس استنباط‌های آماری
۶. سایر روش‌ها

هر کدام از الگوریتم‌های ارائه شده برای انجمن‌یابی در یکی از این دسته‌ها قرار می‌گیرد. در ادامه این بخش به تعریف این گروه‌ها خواهیم پرداخت و برای آن‌ها معروف‌ترین الگوریتم را شرح خواهیم داد.

الگوریتم‌های تقسیم

فلسفه کلی این دسته از الگوریتم‌ها یافتن یال‌های بین انجمنی و حذف آن‌ها است. اگر تمام یال‌های بین انجمن‌ها را حذف کنیم، آن انجمن‌ها مولفه‌های جدا از هم را تشکیل می‌دهند.
معروف‌ترین و محبوب‌ترین الگوریتم که در دسته الگوریتم‌های تقسیم قرار می‌گیرد الگوریتم، گیروان نیومن [۸] [۳۰] است. این الگوریتم برای یافتن یال‌های بین انجمنی از مفهوم مرکزیت ارتباطی استفاده می‌کند. ایده کلی این الگوریتم این است یال‌هایی که بین انجمن‌ها قرار می‌گیرند، دارای مقدار مرکزیت ارتباطی بزرگ‌تری هستند. این الگوریتم از چهار مرحله تشکیل می‌شود:

۱. محاسبه مرکزیت ارتباطی هر یال،
۲. یال با بزرگترین مقدار مرکزیت ارتباطی را حذف می‌کنیم،
۳. مرکزیت ارتباطی یال‌ها را دوباره محاسبه می‌کنیم

۴. به مرحله دوم باز می‌گردیم.

نخست این الگوریتم در مقاله [۳۰] معرفی شد. با توجه به اینکه زمان محاسبه مرکزیت ارتباطی برای هر یال در هر مرحله برابر $O(n^2)$ (برای ماتریس‌های تنک) پیچیدگی زمانی این الگوریتم برابر $O(n^3)$ بود. در مقاله‌ای که نیومن و گیروان در سال ۲۰۰۴ [۸] دو سال پس از مقاله اول منتشر کردند؛ معیاری برای یافتن بهترین تقسیم ارائه دادند. در نسخه اولیه تمام نمودار سلسله مراتبی الگوریتم بدست می‌آمد و با برش از یک سطح انجمن‌ها مشخص می‌شد اما در الگوریتم اصلاح شده، تقسیمی که بزرگترین مقدار ماجولاریتی را ارائه می‌داد، به عنوان تقسیم مورد نظر مشخص می‌کردند.

این الگوریتم با وجود بهبودهای حاصل شده پیچیدگی زمانی زیادی داشت و مهمتر از همه نمی‌توانست انجمن‌های همپوشان را کشف کند.

الگوریتم‌های بر اساس ماجولاریتی

هرچقدر مقدار ماجولاریتی یک تقسیم به مقدار ماجولاریتی ماکزیمم نزدیک‌تر شود، تقسیم حاصل تقسیم خوب‌تری خواهد بود. همین مساله ایده اصلی این دسته از الگوریتم‌ها هستند. الگوریتم‌های مربوط به این دسته به چهار نوع زیر تقسیم می‌شوند

- الگوریتم تبرید شبیه‌سازی شده
- بهینه‌سازی خارجی
- بهینه‌سازی شعاع طیفی
- دیگر شیوه‌های بهینه‌سازی

الگوریتم‌های متفاوتی برای بیشینه‌سازی ماجولاریتی ارائه شده که به یکی از چهار دسته بالا تعلق دارند. اما از مهمترین الگوریتم‌های معرفی شده الگوریتم حریصانه‌ای است که توسط نیومن [۳۱] ارائه شد. الگوریتم با قرار دادن هر راس در یک انجمن مجزا کار خود را شروع می‌کند. در شروع کار هیچ یالی وجود ندارد، با افزودن یک به یک یال‌ها انجمن‌هایی که در دو سر این یال قرار دارند در صورت افزایش ماجولاریتی تقسیم در هم ادغام می‌شوند. ماجولاریتی تقسیم از روی گراف کامل محاسبه می‌شود یعنی گرافی که یال‌ها به آن اضافه می‌شود،

تنها حکم نشانگر انجمن‌ها را دارد. اگر افزودن یک یال ادغامی در انجمن‌ها به وجود نیاورد، آن یال یک یال درون انجمنی است. لذا مقدار ماجولاریتی را تغییر نخواهد داد. تعداد تقسیم‌های یافت شده در طول فرایند برابر n یعنی تعداد راس‌هاست. هرکدام از این تقسیم‌ها دارای یک مقدار ماجولاریتی هستند در نهایت بعد از افزودن یال‌ها تقسیمی که بزرگترین ماجولاریتی را دارد به عنوان خروجی مشخص می‌شود. پیچیدگی زمانی این الگوریتم برای یک گراف تنک یعنی گراف‌هایی که تعداد یال‌های آن از تعداد کل یال‌های ممکن که برای n راس برابر $n(n-۱)/۲$ است بسیار کمتر (از مرتبه خطی)می‌باشد، مساوی $O(n^۲)$ است.

الگوریتم‌های براساس شعاع طیفی مقادیر ویژه

این الگوریتم‌ها از شعاع طیفی مقادیر ویژه ماتریس مربوط به ماتریس‌های مجاورت (یعنی ماتریس‌هایی که به طریقی از ماتریس مجاورت استخراج می‌شوند) استفاده می‌کنند. نخستین تحقیقات بر روی شعاع طیفی خوشه‌ها توسط دوناث [۴۴] و هوفمن [۴۵] [۳۲] انجام شد. در این مقاله از بردار ویژه ماتریس مجاورت برای تقسیم‌بندی گراف استفاده شد. در همان سال فیدلر[۴۶] [۳۳] نشان داد که بردار ویژه دومین مقدار ویژه کوچک ماتریس لاپلاسین به احتمال زیاد تقسیمی ارائه می‌دهد که مینیمم برش این تقسیم اندازه بسیار کوچک‌تری دارد.

گراف ساده G را با n یال داریم، ماتریس لاپلاسین آن $L = (l_{i,j})_{n\times n}$ به صورت زیر تعریف می‌شود:

$$L = D - A$$

که در آن D ماتریس درجات راس‌هاست. برحسب کاربرد می‌توان ماتریس لاپلاسین را به صورت زیر نیز تعریف کرد:

$$l_{i,j} = \begin{cases} deg(v_i) & \text{اگر } i = j \\ -۱ & \text{اگر } i \neq j \ and\ v_i\ adjacent\ to\ v_j \\ ۰ & otherwise \end{cases}$$

تاکنون لاپلاسین پرکاربردترین ماتریس برای الگوریتم‌های شعاع طیفی بوده‌است [۹]. دونتی[۴۷] و مونز[۴۸] [۳۴] روشی، رابرپایه بردار ویژه ماتریس لاپلاسین ارائه دادند. چون مقدار بردار ویژه مولفه‌ها برای راس‌هایی که در

---

[۴۴] Donath

[۴۵] Hoffman,

[۴۶] Fiedler

[۴۷] Donetti

[۴۸] Munoz

یک انجمن قرار دارند، مقدار نزدیکی به هم دارند، می‌توان با استفاده از بردارهای ویژه انجمن‌ها را کشف کرد. بدین معنی که اگر از m بردار ویژه استفاده کنیم، می‌توانیم راس‌ها را در یک فضای m بعدی قرار دهیم که انجمن‌ها به صورت راس‌هایی که در گروه‌های نزدیک به هم در این فضا قرار دارند مشخص می‌شوند. هرچقدر تعداد بردارهای ویژه به کار رفته بیشتر باشد، انجمن‌ها به صورت واضح‌تری در فضا مشخص می‌شود. الگوریتم ارائه شده توسط دونتی و مونز شامل گروه‌بندی نقاط و استخراج تقسیم می‌باشد. دونتی و مونز از خوشه‌بندی سلسله مراتبی، با این محدودیت که تنها خوشه‌هایی که حداقل یک یال بین خوشه‌ای در گراف اصلی دارند با هم ادغام می‌شوند، استفاده می‌کند. از بین همه تقسیم‌های استخراج شده، تقسیمی که بزرگ‌ترین مقدار ماجولاریتی را دارد به عنوان خروجی مشخص می‌شود. پیچیدگی زمانی این روش برابر $O(n^۳)$ می‌باشد.

الگوریتم‌های پویا

این دسته از الگوریتم‌ها بصورت مستقیم بر روی گراف کار می‌کنند و به این طریق انجمن‌ها را استخراج می‌کنند. در اینجا به تشریح الگوریتم عابر تصادفی که توسط ژو[۴۹] [۳۵] ارائه شده‌است و در این دسته از الگوریتم‌ها طبقه‌بندی می‌شود، می‌پردازیم. عابر تصادفی که نخستین بار توسط هوگز[۵۰] [۳۶] معرفی شد، عابری تصادفی است که به صورت تصادفی روی گراف حرکت می‌کند و در هر راس با توجه به یال‌های موجود به هر کدام از راس‌های مجاور بصورت تصادفی می‌رود. ایده الگوریتم ژو این است که عابر تصادفی به علت چگالی زیاد یال‌ها در داخل انجمن زمان بیشتری را داخل انجمن مصرف خواهد کرد. ژو از عابر تصادفی برای تعریف فاصله بین دو راس استفاده کرد: فاصله d_{ij} بین دو راس i و j میانگین تعداد یال‌هایی است که یک عابر تصادفی برای رسیدن از j به i باید از آن‌ها عبور کند. راس‌ها نزدیک به هم احتمالا به یک انجمن تعلق دارند و جذب کننده سراسری راس i را به عنوان نزدیک‌ترین همسایه این راس تعریف می‌کند (راسی که کوچکترین مقدار d_{ij} را داشته باشد). او همچنین جذب کننده محلی راس i را به عنوان راسی تعریف می‌کند که i نزدیک‌ترین همسایه آن باشد. دو نوع انجمن براساس جذب کننده سراسری و جذب کننده محلی وجود دارد: راس i با تمام راس‌هایی که جذب کننده سراسری (محلی) هستند در یک انجمن قرار می‌گیرد و همه راس‌هایی که i جذب کننده سراسری (محلی) آن هستند نیز در انجمنی قرار می‌گیرند که i در آن قرار دارد. انجمن زیرگراف کمینه است، یعنی هیچ زیرگراف کوچک‌تری که شرایط مورد نظر در آن صدق کند وجود ندارد. پیچیدگی زمانی این

[۴۹] Zhou

[۵۰] Hughes

الگوریتم برابر $O(n^۳)$ است.

الگوریتم‌های براساس استنباط‌های آماری

استنباط‌های آماری به منظور استخراج ویژگی‌های مجموعه‌ای از داده‌ها با استفاده از یک دسته مشاهدات و مقایسه آن‌ها با مدل‌های فرضی تولید شده، مورد استفاده قرار می‌گیرند. اگر مجموعه داده یک گراف باشد، مدل مجموعه‌ای از راس‌هاست که به وسیله یال‌هایی به هم متصل می‌شوند، این مدل تولید شده با توپولوژی گراف اصلی منطبق است. این دسته از الگوریتم‌ها از این مدل‌ها استفاده کرده و انجمن‌های موجود در گراف را پیش‌بینی می‌کند.

استنباط بیزی یکی از روش‌های استنباط‌های آماری است که در مدل‌سازی گراف‌های واقعی مانند شبکه‌های اجتماعی به کار می‌رود، در این بخش به معرفی الگوریتم هاستینگ[۵۱] [۳۷] می‌پردازیم که از روش استنباط بیزی استفاده می‌کند.

استنباط بیزی از مشاهدات به منظور تخمین احتمال درستی یک فرض استفاده می‌کند. استنباط بیزی شامل دو جزء است: شواهد، که عبارت است از اطلاعات D که می‌توان از سیستم بدست آورد؛ و یک مدل آماری با پارامتر $\{\theta\}$. استنباط بیری از محاسبه احتمال $P(D|\{\theta\})$ که برابر احتمال مشاهده شواهد در مدل مورد نظر با پارامتر $\{\theta\}$ است، شروع می‌شود. هدف مشخص کردن مقدار $\{\theta\}$ای است که مقدار $P(\{\theta\}|D)$ را حداکثر کند.

در ارتباط با گراف، شواهد توسط ساختار گراف (ماتریس مجاورت یا ماتریس وزن) بدست می‌آید. در این مورد یک جزء دیگر علاوه بر اجزاء ذکر شده وجود دارد و آن گروه‌بندی گراف‌ها به وسیله قرار دادن راس‌ها داخل گروه‌هاست. این گروه‌بندی اطلاعات پنهانی است که انتظار داریم از مدل انتخابی به وسیله پارامتر بدست آوریم.

در تمام روش‌هایی که از استنباط بیزوی استفاده می‌کنند، هدف بیشینه کردن $P(\{\theta\}|D)$ می‌باشد که در آن مدل شامل ساختار گراف مشاهده شده، با مقداری محدویت اعمال شده می‌باشد. در ارتباط با گراف‌ها، پارامتر $\{\theta\}$ توسط سه گانه $(\{q\}, \{\theta\}, k)$ مشخص می‌شود که در آن $\{q\}$ انجمن‌هایی که توسط انتساب راس‌ها به یال‌ها مشخص شده را نشان می‌دهد. $\{\theta\}$ پارامتر مدل و k تعداد خوشه‌هاست.

الگوریتم هاستینگ از مدلی با نام مدل طراحی شده استفاده می‌کند که در این مدل، n راس به q گروه منتسب می‌شوند: راس‌هایی که در یک گروه قرار دارند با احتمال p_{in} با هم در ارتباط هستند (یالی بینشان قرار دارد)، در حالی که راس‌هایی متعلق به گروه‌های مجزا با احتمال p_{out} با هم ارتباط دارند. اگر $p_{in} > p_{out}$ گراف

[۵۱] Hastings

دارای انجمن می‌باشد. کلاس‌بندی گراف با مجموعه برچسب‌های $\{q_i\}$ مشخص می‌شود. احتمال اینکه با توجه به گراف داده شده کلاس‌بندی $\{q_i\}$ یک کلاس‌بندی مناسب، متناسب با مدل داده شده باشد برابر است با:

$$p(\{q_i\}) \propto \{\exp[-\sum_{\langle ij \rangle} J\delta_{q_i q_j} - \sum_{i \neq j} J'\delta_{q_i q_j}/۲]\}^{-۱} \qquad (۳۰-۱)$$

که در آن $J = \log\{[p_{in}(۱-p_{out})]/[p_{out}(۱-p_{in})]\}$ و $J' = \log\{[(۱-p_{in})]/[(۱-p_{out})]\}$ و اولین مجموع روی نزدیک‌ترین همسایه‌ها اجرا می‌شود.

سایر روش‌ها

بقیه الگوریتم‌های ارائه شده برای یافتن انجمن‌ها در این دسته قرار می‌گیرند. از بین این الگوریتم‌ها، الگوریتم راقوان[۵۲] و همکاران [۳۸] را معرفی خواهیم کرد.

راقوان و همکاران الگوریتم ساده و سریعی طراحی کردند که با نام الگوریتم گسترش برچسب‌ها شناخته می‌شود. راس‌ها در ابتدای کار با برچسب‌های منحصر به فردی از هم مجزا می‌شوند. در هر مرحله یک بار مورد بررسی قرار می‌گیرند و برچسب‌شان با کمک برچسب راس‌های اطراف مشخص می‌شود؛ بدین ترتیب که هر راس برچسب خود را به برچسبی که بیشترین تعداد تکرار را در بین راس‌های همسایه دارد، تغییر می‌دهد. اگر چندین برچسب با تعداد تکرار مساوی در همسایگی راسی قرار داشته باشد، یکی به تصادف انتخاب می‌شود. بدین وسیله برچسب‌ها در گراف گسترش می‌یابند: اغلب برچسب‌ها ناپدید می‌شوند و برخی دیگر به برچسب غالب گراف تبدیل می‌شوند. در مرحله‌ای که دیگر تغییری رخ نمی‌دهد، الگوریتم پایان می‌یابد. به صورت ساختاری هر راس همسایه‌های بیشتری را در بین راس‌های انجمنی که در آن قرار دارد، نسبت به راس‌های متعلق به دیگر انجمن‌ها داراست. این الگوریتم احتمالا جواب‌های مختلفی را برای یک گراف در تکرار (الگوریتم) بدست می‌دهد. الگوریتم برای هر مرحله زمان $O(m)$ را مصرف می‌کند.

۶.۲.۱ الگوریتم‌های انجمن‌یابی برای یافتن انجمن‌های همپوشان

الگوریتم‌های مختلفی برای یافتن انجمن در ارتباط با گراف‌های جهت‌دار، وزن‌دار و برای یافتن انجمن‌های همپوشان ارائه شده‌است. برخی از این الگوریتم‌ها با تغییر روی الگوریتم‌هایی که برای گراف‌های ساده ارائه شدند به دست آمدند و برخی دیگر برای این نوع گراف‌ها به صورت تخصصی ارائه شدند. با توجه به اینکه گراف‌های

[۵۲] Raghavan

بدست آمده برای شبکه‌های اجتماعی اغلب گراف‌های بی‌وزن و بی‌جهت و دارای انجمن‌های همپوشان می‌باشند.

در این بخش به صورت اختصاصی به معرفی الگوریتم‌های یافتن انجمن‌ها همپوشان می‌پردازیم.

الگوریتم‌های پالا و همکاران [۱۹]، بومز [۵۳] و همکاران [۳۹]، نپوسز[۵۴] و همکاران [۴۰] برای یافتن انجمن‌های همپوشان ارائه شدند. لانچیچنتی [۵۵] و فورتوناتو[۵۶] [۴۱] مقایسه‌ای بین الگوریتم‌های انجمن‌یابی انجام داده‌اند. با توجه به این مقایسه برای انجمن‌های همپوشان تنها الگوریتم پالا و همکاران [۱۹] نتایج واقع بینانه‌ای را به عنوان خروجی ارائه می‌دهد. برای تحلیل انجمن‌های مدل‌های ارائه شده از الگوریتم پالا استفاده کرده‌ایم. لذا در ادامه این بخش به معرفی و بررسی الگوریتم پالا می‌پردازیم.

برای بحث در مورد این الگوریتم نیاز داریم که برخی مفاهیم را به صورت واضح تعریف کنیم:

- **k-دسته**[۵۷]: زیرگرافی با k راس که به صورت کامل متصل باشد؛ یعنی بین هر دو راسی حتما یالی باشد[۵۸].

- **k-دسته‌های مجاور**[۵۹]: دو k-دسته مجاور هم هستند، هرگاه $k-1$ راس را به اشتراک بگذارند. یعنی تنها در یک راس با هم متفاوت باشند.

- **زنجیره k-دسته**[۶۰]: یک زیرگراف از گراف اصلی است، که از یک دنباله k-دسته مجاور تشکیل شده‌است به طوری که هر دو k-دسته متوالی مجاور باشند.

- **k-دسته متصل**[۶۱]: دو k-دسته، k-دسته متصل نامیده می‌شوند هرگاه بخشی از یک زنجیره k-دسته باشد.

- **انجمن k-دسته** [۶۲]: انجمن‌های k-دسته عبارت است ازبزرگ‌ترین زیرگراف متصل بدست آمده از اجتماع یک k-دسته و همه k-دسته‌های متصل به آن.

---

[۵۳] Bomze

[۵۴] Nepusz

[۵۵] Lancichinetti

[۵۶] Fortunato

[۵۷] k-clique:

[۵۸] این تعریف با تعریف مربوط به n-دسته که در بخش ۳.۲.۱ ارائه شد فرق دارد

[۵۹] k-clique adjacencent

[۶۰] k-clique chain

[۶۱] connected k-clique

[۶۲] k-clique community

یک الگو k-دسته یک همریختی [۶۳] با یک گراف کامل با k راس است. غلتیدن عبارت است از حرکت یک الگوی k-دسته از یک k-دسته به k-دسته دیگر که در مجاورت آن قرار دارد با حفظ همه k-1 راس مشترک آن. با این تعاریف می‌توان یک انجمن k-دسته را بوسیله غلتیدن یک الگوی k-دسته روی گراف‌های مجاور بدست آورد. انجمن‌های k-دسته می‌توانند راس‌ها را به اشتراک بگذارند، لذا می‌توانند انجمن‌های همپوشان را کشف کنند. در این الگوریتم برخی از راس‌ها ممکن است توسط هیچ‌کدام از k-دسته‌ها یافت نشوند، مانند راس‌هایی که درجه یک دارند. برای کشف انجمن‌های k-دسته می‌توان نخست به دنبال دسته‌های بیشینه بود. سپس یک ماتریس همپوشان دسته-دسته O ساخت. این ماتریس یک ماتریس $n_c \times n_c$ می‌باشد که در آن n_c تعداد دسته‌هاست؛ و O_{ij} تعداد راس‌هایی است که توسط دسته i و j به اشتراک گذاشته می‌شود. برای یافتن یک k-دسته کافی است درایه‌هایی که مقادیر بزرگتر یا مساوی $k-۱$ دارند مشخص کنیم و مقدار بقیه درایه‌ها را برابر صفر قرار دهیم و در نهایت مولفه‌های همبندی ماتریس حاصل را بدست آوریم. این الگوریتم به لحاظ محاسباتی پیچیدگی زمانی زیادی دارد (نمایی) اما به لحاظ عملی این الگوریتم زمان قابل قبولی را ارائه می‌دهد. حتی با قرار دادن t برابر ۲ (یعنی زمان جستجو برای هر راس برابر ۲ ثانیه)می‌توان در زمان قابل قبول تقریب خوبی از انجمن‌ها بدست آورد.

۳.۱ مدلسازی شبکه‌های اجتماعی

مدل، مدلسازی و استفاده از آن در بسیاری از علوم چه طبیعی و تجربی، و چه انسانی نقش مهمی را برعهده دارند. در اغلب علوم انجام پژوهش برروی کل یک سیستم مورد پژوهش عقلانی و قابل اجرا نیست. مثلا موردی را در نظر بگیرید که در علم زلزله شناسی میزان آسیبی که زلزله به یک منزل مسکونی وارد می‌کند مورد تحقیق باشد، بدون مدل‌سازی دانشمندان مجبورند کل یک سیستم مورد تحقیق را در معرض زلزله قرار بدهند و مقدار خسارت را گزارش کنند. یا در علوم تجربی، فرض کنید زیست‌شناسان بخواهند تاثیر تغییرات اکوسیستم را بر روی گونه خاصی از حیوانات بررسی کنند، بدون مدلسازی دانشمندان باید کل اکوسیستم را برای پژوهش تخریب کنند.

در مورد شبکه‌های اجتماعی استفاده از مدل و مدلسازی گریزناپذیر می‌نماید. برای مثال شبکه‌اجتماعی برخط فیس‌بوک را در نظر بگیرید. این شبکه دارای بیش از یک میلیارد عضو هست، یعنی گراف این شبکه دارای بیش از یک میلیارد راس می‌باشد. اگر فرض کنیم هر راس حداقل با ۱۰ نفر دیگر در ارتباط است؛ گراف این شبکه

[۶۳] دو گراف همریخت است هرگاه یک تابع پوشا وجود داشته باشد که یکی را به دیگری تبدیل کند.

دارای بیش از ده میلیارد یال است، یعنی بیش از جمعیت کل کره زمین. حال در نظر بگیرید بخواهیم برخی از ویژگی‌های این گراف را محاسبه کنیم (مثلا میانگین درجه راس‌های گراف، یا انجمن‌های تشکیل شده)، بی‌تردید اجرای الگوریتم‌هایی با درجه پیچیدگی حداقل $O(n^۲)$ برای محاسبه ویژگی‌ها، زمان بسیار زیادی را می‌طلبد. از این گذشته برای پیش‌بینی برخی از ویژگی‌های گراف نظیر مقدار فضای مورد نیاز به گرافی است که از طریق مدل‌ها به وجود بیایند.

با توجه به حجم زیاد داده‌ها و نیاز به پیش‌بینی برخی از ویژگی‌ها (مثلا الگوریتم‌های پیشنهاد)، نمونه‌گیری و مدلسازی به بخش جدانشدنی از تجزیه و تحلیل شبکه‌های اجتماعی بدل شده‌است.

در این بخش نخست به تعریف مدلسازی به صورت عمومی خواهیم پرداخت. سپس به بررسی مدلسازی در شبکه‌های اجتماعی می‌پردازیم. نهایتا در مورد پارامترها و انواع مدل‌ها بحث خواهیم کرد.

۱.۳.۱ انواع مدلسازی

مدل خلاصه‌ای از واقعیت است. مدل چیزی است که برای توصیف یک مورد کلی‌تر به کار می‌رود. مدل در معنای نمونه و شبیه می‌باشد. مدلسازی همان شبیه‌سازی است، شبیه‌سازی در مقیاس کوچک‌تر نسبت به یک شیء بزرگ. مدل را می‌توان به معنای دستگاهی برای اندیشیدن نیز به کار برد، لذا مدلسازی ایجاد دستگاه و سیستمی است برای اندیشیدن.

مدل در ساده‌ترین شکل دستگاهی است که به تشریح چگونگی فرآیند تکامل موجودیت‌ها و تعامل میان موجودیت‌ها می‌پردازد. می‌توان با استفاده از مدل به تخمین برخی یا همه ویژگی‌های یک سیستم بزرگ‌تر رسید.

می‌توان انواع مدل‌ها را به صورت زیر دسته‌بندی کرد:

- مدل فیزیکی
- مدل ذهنی
- مدل ریاضی

مدل‌های فیزیکی: مدل‌های فیزیکی، یک مدل با مقیاس کوچک‌تر از یک سیستم بزرگ‌تر است. مانند ماکت یک پل که توسط مهندسان ساخته می‌شود.

مدل‌های ذهنی: مدل‌هایی که بیشتر در ذهن ما وجود دارند و بیشتر در علوم انسانی به کار می‌روند. مانند

مدل‌های مربوط به توسعه در جوامع مختلف.

مدل ریاضی: مدل ریاضی عبارت است توصیف یک سیستم با توجه به ابزارها ریاضی نظیر قضیه‌ها و نمادهای ریاضی.

مدل‌های ارائه شده برای شبکه‌های اجتماعی جزو مدل‌های ریاضی می‌باشند. مدل‌های شبکه‌های اجتماعی، یک شبکه‌ی اجتماعی را با کمک ابزارهای ریاضی، یعنی گراف و ماتریس توصیف می‌کنند.

۲.۳.۱ آشنایی با مدلسازی شبکه‌های اجتماعی

۳.۳.۱ ویژگی‌های مورد نیاز برای شبیه سازی شبکه‌های اجتماعی

همانطور که گفته شد شبکه‌های اجتماعی توسط ابزارهای ریاضی یعنی ماتریس‌ها و گراف‌ها مدل می‌شوند. هدف از مدلسازی شبکه‌های اجتماعی تولید گرافی است که بیشترین انطباق را با شبکه‌ی اجتماعی مورد نظر ما داشته باشد. مدل‌های ارائه شده اغلب سعی دارند که ساختار توپولوژیک شبکه‌های اجتماعی را شبیه‌سازی کنند. ابزارهای مدلسازی راس‌ها و یال‌های بین‌شان هستند.

مدل‌های ارائه شده سعی دارند تا در طی یک فرایند راس‌ها یا یال‌ها و یا هر دو را به گراف به شکلی اضافه یا کم کنند که مدل حاصل به بهترین نحو ساختار توپولوژیک شبکه‌های اجتماعی را در خود داشته باشد.

شبکه‌های اجتماعی ویژگی‌های خاصی را ار خود بروز می‌دهند. مدل‌های مختلف ارائه شده سعی در شبیه‌سازی این ویژگی‌ها دارند. میسلف و همکاران [۴] در سال ۲۰۰۹ با بررسی طیف وسیعی از شبکه‌های اجتماعی برخط نظیر فیلیکر[۶۴]، یوتیوب[۶۵]، لیوجورنال [۶۶]، و اورکوت به نتایج زیر در ارتباط با شبکه‌های اجتماعی آنلاین دست یافت:

- **توزیع درجه:** اغلب شبکه‌های اجتماعی برخط توزیع درجه توانی [۶۷] دارند.

- **همبستگی درجات ورودی و خروجی:** در یک شبکه اجتماعی جهت‌دار یال‌هایی که درجه خروجی بالایی دارند، دارای درجه ورودی بالاتری هستند. به عبارت دیگر با استفاده از ادبیات شبکه‌های اجتماعی بازیگران فعال محبوبیت بالایی نیز دارند. همپوشانی راس‌هایی که جزو ۱٪ راس‌هایی با بیشترین ورودی

[۶۴] Flickr.com

[۶۵] youtube.com

[۶۶] LiveJournal.com

[۶۷] در آمار توزیع توانی تابعی است میان دو متغیر که تغییرات یکی از متغیرها برپایه توان متغیر دیگر است: $f(x) = ax^{\gamma}$

هستند، با راس‌هایی که جزو ۱٪ راس‌های با بیشترین درجه خروجی هستند بیشتر از ۶۵٪ است.

- **میانگین کوتاه‌ترین مسیرها:** شبکه‌های اجتماعی دارای شعاع، قطر، و میانگین کوتاه‌ترین مسیر، بسیار کوچکی هستند.

- **توزیع درجات راس‌های متصل:** راس‌های با درجات بالاتر تمایل دارند به دیگر راس‌های با درجات بالا متصل شوند. بدین طریق هسته‌های [۶۸] شبکه اجتماعی را تشکیل دهند.

- **ویژگی استقلال از مقیاس:** شبکه‌های اجتماعی ویژگی استقلال از مقیاس بالایی را از خود بروز می‌دهند به این معنا که راس‌های با درجه بالا به راس‌های با درجه بالای دیگر متصل می‌شوند. همچنین راس‌هایی که درجه پایین تمایل دارند به راس‌های با درجه پایین دیگر مرتبط باشند.

- **شرکت‌پذیری:** شبکه‌های اجتماعی شرکت‌پذیری بالایی را از خود به نمایش می‌گذارند.

- **هسته‌ها و حاشیه‌های شبکه‌های اجتماعی آنلاین:** با در نظر گرفتن خاصیت‌های شرکت‌پذیری و استقلال از مقیاس می‌توان ادعا کرد که اغلب شبکه‌های اجتماعی از راس‌هایی با اتصالات زیاد به عنوان هسته تشکیل شده‌است که راس‌هایی با درجات پایین‌تر را به عنوان شاخه‌هایی در حاشیه خود دارد.

- **ضریب خوشه‌بندی:** شبکه‌های اجتماعی ضریب خوشه‌بندی بزرگ‌تری را نسبت به گراف‌های تصادفی معادلشان دارند.

مدل‌های ارائه شده برای شبکه‌های اجتماعی باید دارای این ویژگی‌ها باشند. به علاوه مدل‌ها باید دارای انجمن‌ها نیز باشند.

۴.۳.۱ تاریخچه مدلسازی شبکه‌های اجتماعی

تاریخچه مدلسازی شبکه‌های اجتماعی نشان‌دهنده سعی دانشمندان برای ایجاد مدل‌هایی است که ویژگی‌های اشاره شده به اضافه خاصیت وجود انجمن را از خود بروز دهند. نخستین مدلی که برای مدلسازی گراف شبکه‌های

[۶۸] هسته زیرمجموعه کمینه از راس‌هاست که دارای دو خاصیت باشد: وجودش برای متصل بودن شبکه ضروری باشد، یعنی با حذف این مجموعه شبکه به شبکه‌های جدا از هم تقسیم شود. این مجموعه از راس‌ها تعداد یال‌های زیادی را باید با هم به اشتراک بگذارند یعنی اتصال قوی داشته باشند که این خاصیت باعث کوتاه بودن قطر می‌شود، لذا هسته باید دارای قطر نسبتا کوچکی داشته باشد.

اجتماعی مورد استفاده قرار گرفت مدل تصادفی اردوش [۶۹] و رنیی [۷۰] [۴۲] بود. در این مدل بین هر دو راس یالی با احتمال p وجود دارد. لذا مدل اردوش و رنیی دارای توزیع درجات پوآسون [۴۳] می‌باشد. اما با توجه به آنچه گفته شد، شبکه‌های اجتماعی دارای توزیع درجات توانی هستند. ژئونگ [۷۱] و همکاران [۴۴] مدلی را ارائه کردند که بر این اصل شبکه‌های اجتماعی استوار بود که محبوب‌ها با سرعت بیشتری محبوب‌تر می‌شوند. به صورت شهودی در یک شبکه اجتماعی درخواست دوستی بیشتری برای یک فرد مشهور نسبت به یک فرد گمنام ارسال می‌شود. این مدل برای برآورده کردن این ویژگی از خاصیت اتصال امتیازی استفاده می‌کند، یعنی احتمال اینکه یک یال ایجاد شده توسط یک راس به راس دیگری متصل شود، متناسب با درجه راس هدف است. یعنی راس‌هایی که درجه بیشتری دارند به احتمال بیشتری به عنوان انتهای دیگر یک یال انتخاب می‌شوند. با کشف ویژگی جهان کوچک در شبکه‌های اجتماعی برآورده کردن این نیاز در دستور کار مدلسازان شبکه‌های اجتماعی قرار گرفت. مدلی که از الحاق امتیازی برای افزودن یال به گراف استفاده می‌کرد، جستجو را در سطح کل گراف انجام می‌داد. یعنی تنها تفاوت بین راس‌ها برای ایجاد یال درجات آن‌ها بود. اما ضریب خوشه‌بندی بالای شبکه‌های اجتماعی نشانگر وجود جستجوی محلی (جستجویی که به یک فضای خاص محدود باشد) برای انتخاب راس هدف است، بصورت شهودی نیز این جستجوی محلی منطقی است، احتمال اینکه دوستِ دوست [۷۲] شما با شما دوست باشد بیشتر از احتمال دوست بودن شما با فردی دیگر در یک شبکه اجتماعی خاص می‌باشد. بنابراین باید احتمال ایجاد یال بین دو راس که یک همسایه مشترک دارند بیشتر از دو راسی باشد که با هم هیچ همسایه مشترکی ندارند. این خاصیت باعث به وجود آمدن شبکه‌ای با ضریب خوشه‌بندی بزرگی شد. از طرفی خاصیت الحاق امتیازی در مدل‌ها حفظ شد. همانطور که اشاره شد، الحاق امتیازی باعث به وجود آمدن راس‌های قطب و هسته‌ها شده، و وجود قطب‌ها و هسته‌ها باعث کوتاه شدن میانگین کوتاهترین مسیرهای یک گراف می‌شود. با کنار هم قرار دادن این موارد مدل‌ها دارای خاصیت جهان کوچک می‌شوند. با مطرح شدن بحث وجود انجمن‌ها در شبکه‌های اجتماعی نیاز به معرفی مدل‌هایی که دارای انجمن‌ها باشند، احساس شد. بعد از آن مدل‌ها سعی داشتند که علاوه بر ویژگی جهان کوچک دارای انجمن نیز باشند.

---

[۶۹] Erdös

[۷۰] Rényi

[۷۱] Jeong

[۷۲] friend of friend

۵.۳.۱ تقسیم‌بندی مدل‌ها

همانند تقسیم‌بندی تویونن و همکاران [۱۳] می‌توان انواع مدل‌های ارائه شده را در دو دسته کلی قرار داد: مدل‌های تکامل شبکه ،که به اختصار NEM[۷۳] گفته می‌شوند، و مدل‌های مبتنی بر ویژگی‌های راس که به اختصار NAM[۷۴] خوانده می‌شوند.

مدل‌های تکاملی شبکه

مدل‌های تکاملی شبکه‌ای به دو زیردسته مدل‌های تکاملی شبکه رشد و مدل‌های تکاملی شبکه‌ای پویا تقسیم می‌شوند:

مدل‌های تکاملی شبکه‌ای رشد

به مدل‌های تکاملی شبکه‌ای، که در آن راس‌ها تا زمانی که تعداد آن‌ها به یک مقدار از پیش تعریف نشده نرسیده، به گراف اضافه می‌شوند، مدل‌های تکاملی شبکه‌ای رشد گفته می‌شود. در هر گام یک راس اضافه می‌شود یا یک یال حذف و یا یک یال به گراف اضافه می‌شوند.

مدل‌های تکاملی شبکه‌ای پویا

به مدل‌های تکاملی شبکه‌ای که در آن‌ها تعداد راس‌ها ثابت است و یال‌ها تا زمان اتمام فرایند به حذف و اضافه شدن ادامه می‌دهند، مدل‌های تکاملی شبکه‌ای پویا گفته می‌شود.
تویونن و همکاران ویژگی‌های زیر را برای مدل‌های تکاملی شبکه برشمرده‌اند:

۱. مدل‌های تکاملی شبکه برای تولید گراف G از یک فرایند تکراری تولید و حذف یال‌ها و راس‌ها استفاده می‌کند. فرایند بر روی یک ساختار اولیه $G(t_0)$ آغاز می‌شود. فرآیندهای مدل‌های تکاملی شبکه‌ای پویا اغلب از یک شبکه خالی شروع می‌شوند، اما مدل‌های تکاملی شبکه‌ای رشد از یک هسته کوچک شبکه شروع به کار می‌کنند.

۲. فرایند تولید مدل‌های تکاملی شبکه شامل یک دسته از قانون‌های تصادفی است که در هر مرحله باعث تکاملی شبکه تولیدی می‌شود. این قوانین زیرمجموعه‌ای از یال‌ها و راس‌ها را انتخاب می‌کند و در هر

[۷۳] network evolution models
[۷۴] nodal attribute models

مرحله یال‌ها یا راس‌هایی به این زیرگراف اضافه یا از آن کم می‌کند. این قوانین سعی در شبیه‌سازی قوانین جهان واقعی دارند. برای مثال با ایجاد بستارهای مثلثی (سه راس که هرکدام به دوتای دیگر متصل است) سعی در شبیه‌سازی روابط دوست بودن دو فرد با یک دوست مشترک را دارد. قوانین در هر مرحله باعث گذار گراف از یک حالت $G(t_{k-1})$ به حالت بعدی یعنی $G(t_k)$ می‌شود. فرایند در نهایت در حالت $G = G(t_{end})$ متوقف می‌شود.

۳. برای دو دسته مدل‌های تکاملی شبکه، دو معیار توقف وجود دارد:

- مدل‌های تکاملی شبکه‌ای رشد تا زمانی که تعداد راس‌های آن به مقدار مشخصی نرسیده ادامه پیدا می‌کنند و به محض رسیدن به تعداد راس مشخص متوقف می‌شود.
- مدل‌های تکاملی شبکه پویا زمانی متوقف می‌شوند، که شبکه تولیدی به لحاظ آماری به ثبات برسد.

مدل‌های مبتنی بر ویژگی‌های راس

در مدل‌های مبتنی بر ویژگی‌های راس احتمال وجود یال e_{ij} بین راس‌های i و j تابعی از ویژگی‌های راس‌های i و j است. این مدل‌ها را می‌توان با استفاده از نگاشتن راس‌ها به صورت نقاطی در یک فضا تولید کرد. به طوری که ویژگی‌های راس‌ها در مکان نگاشت منعکس می‌شوند. فاصله دو راس نشان‌دهنده شباهت‌های آن‌ها با هم می‌باشد. هرچه فاصله کمتر باشد، شباهت بیشتر و هر چه فاصله بیشتر باشد، شباهت کمتر می‌شود.

تکامل مدل‌ها

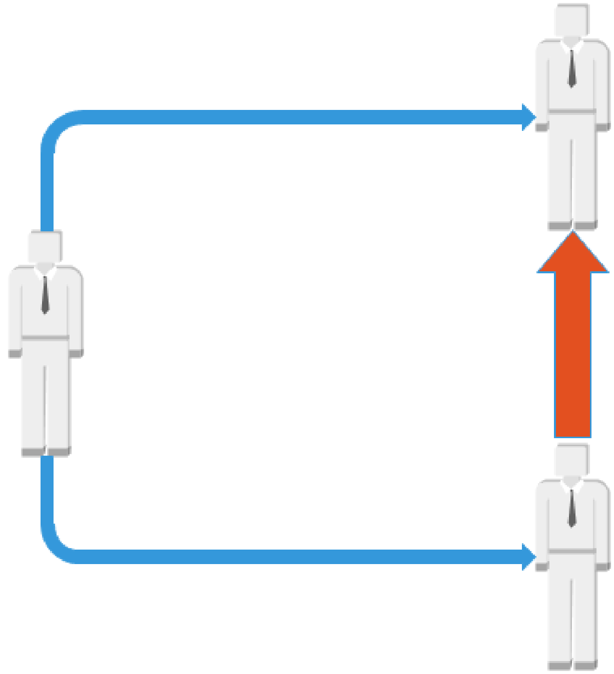
افزودن

برای رشد مدل‌ها نیاز به افزودن یال‌ها و راس‌ها وجود دارد. برای شبیه‌سازی رشد شبکه‌های اجتماعی نیاز به دو نوع جستجو هست:

۱. جستجوی سراسری

۲. جستجوی محلی

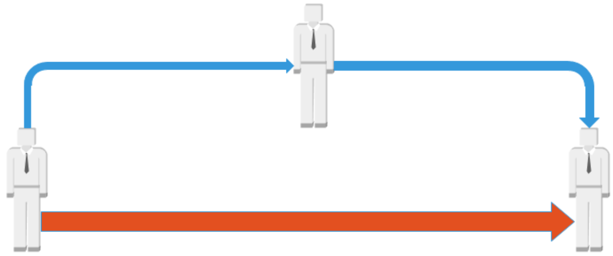
افراد و بازیگران یک شبکه اجتماعی تعدادی از دوستان خود را به تصادف بدون هیچ پیش‌زمینه‌ای انتخاب می‌کنند. دو نفر به صورت تصادفی طی حادثه‌ای با هم آشنا می‌شوند و بدین وسیله رابطه دوستی بین آن‌ها



شکل ۶.۱: یال افزایی نوع اول در این نوع از یال افزایی دو نفر که دوست مشترکی دارند به هم معرفی می‌شوند. فلش‌های آبی (نازک‌تر) نشان دهنده رابطه‌های قدیمی‌تر و فلش قرمز (ضخیم‌تر) رابطه جدید ایجاد شده می‌باشند.

شکل می‌گیرد. این کار در مدل‌های مربوط به شبکه‌های اجتماعی توسط جستجوی سراسری انجام می‌شود. بین دو راس که به تصادف انتخاب شده‌اند در صورت عدم وجود یال، یالی ایجاد می‌شود.

بیشتر دوستی‌ها و ارتباطات شکل گرفته شده در یک شبکه اجتماعی بین افرادی است که پیش‌زمینه‌ای نسبت به هم برای برقراری ارتباط دارند. برای مثال احتمال ایجاد ارتباط بین دو نفر که دوست مشترکی دارند، بیشتر از احتمال ایجاد ارتباط بین دو نفر می‌باشد که هیچ دوست مشترکی ندارند. جستجوی محلی سعی در شبیه‌سازی این نوع ارتباط دارد. مدل‌های مبتنی بر ویژگی‌های راس، یال‌ها را بین راس‌ها بر اساس شباهت راس‌ها اضافه می‌کند، لذا تنها یک نوع افزودن یال وجود دارد. اما مدل‌های تکاملی شبکه‌ای دو نوع افزودن یا را پیشنهاد داده‌اند. در یال افزایی نوع اول که در شکل ۶.۱ مشاهده می‌شود دو فرد که دوست مشترکی دارند، به وسیله دوست مشترک به هم معرفی می‌شوند. در یال افزایی نوع دوم که در شکل ۷.۱ قابل مشاهده است یک فرد با دوست دوستش ارتباط برقرار می‌کند.



شکل ۷.۱: یال افزایی نوع دوم در این نوع از یال افزایی یک رابطه دوستی با دوست دوست ایجاد می‌شود. فلش‌های آبی (نازک‌تر) نشان دهنده رابطه‌های قدیمی‌تر و فلش قرمز (ضخیم‌تر) رابطه جدید ایجاد شده می‌باشند.

حذف

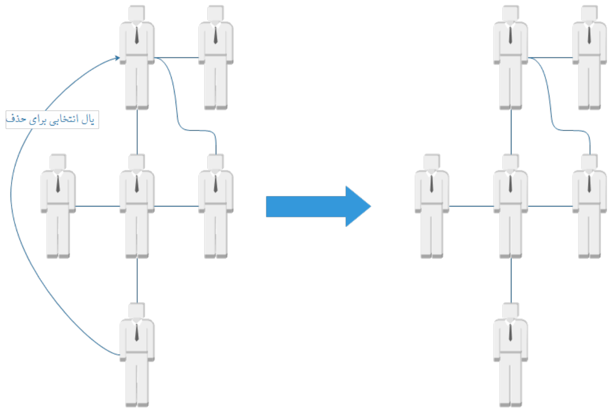
برای شبیه‌سازی شبکه‌های اجتماعی واقعی از دو روش حذف استفاده می‌شود:

۱. حذف یال

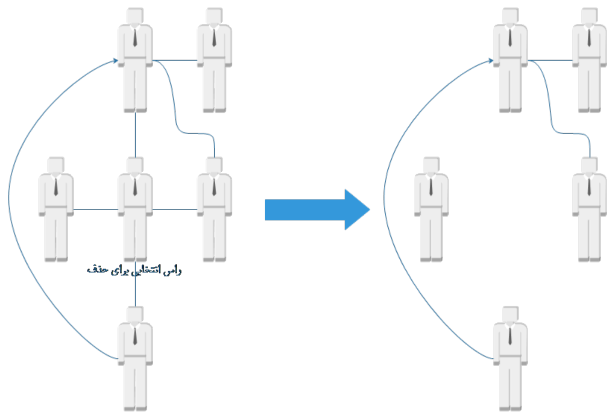
۲. حذف راس

در یک شبکه اجتماعی مانند شبکه دوستی، ممکن است به خاطر نزاع بین دو دوست رابطه دوستی آن‌ها از بین برود. برای شبیه‌سازی این نوع از بین رفتن رابطه از حذف یال استفاده می‌کنیم. شکل ۸.۱ این نوع حذف را نشان می‌دهد. یک یال به تصادف انتخاب شده و حذف می‌شود.

حالتی را در نظر بگیرید که در یک گروه همکاران یکی از همکاران به علتی گروه را ترک می‌کند. برای شبیه سازی این حالت، از حذف راس استفاده می‌کنیم. یک راس به تصادف انتخاب شده و راس انتخابی به همراه همه روابط حادثش حذف می‌شود. این نوع حذف در شکل ۹.۱ نشان داده می‌شود. در فصول بعد، در طی فرایند مدلسازی موارد استفاده هریک از این روش‌ها را مشاهده خواهیم کرد.



شکل ۸.۱: حذف یال: در این روش یک یال به تصادف انتخاب شده و حذف می‌شود. شکل سمت چپ شبکه اجتماعی پیش از حذف و شکل سمت راست شبکه اجتماعی پس از حذف را نشان می‌دهد.



شکل ۹.۱: حذف راس در این روش یک راس به تصادف انتخاب شده و راس انتخابی به همراه تمام روابط از شبکه حذف می‌شود. شکل سمت چپ شبکه اجتماعی پیش از حذف و شکل سمت راست شبکه اجتماعی پس از حذف را نشان می‌دهد.

فصل ۲

روند تکاملی مدل‌ها

۱.۲ ظهور ویژگی جهان کوچک بواسطه تعامل محلی

در این بخش به معرفی یکی از نخستین مدل‌های ارائه شده برای شبکه‌های اجتماعی می‌پردازیم، مدل DEB [۱] اولین مدلی بود که از جستجوی محلی برای ایجاد یال‌های جدید استفاده کرد. اما پیش از بررسی کامل مدل و برای توضیح مدل‌هایی که زیربنای کاری این مدل جدید را تشکیل می‌دادند، نیاز است تا دو مدل گراف تصادفی اردوش و رنیی [۴۲] و الحاق امتیازی [۴۷] معرفی گردد. بعد از معرفی این دو مدل، مدل DEB را به صورت کامل تحلیل و مورد بررسی قرار خواهیم داد.

۱.۱.۲ گراف تصادفی اردوش-رنیی

می‌توان گفت که نخستین مدل ارائه شده در مدلسازی شبکه‌های پیچیده، (گراف‌های بسیار بزرگ که شبکه اجتماعی زیرمجموعه این گراف‌هاست) مدل گراف تصادفی اردوش رنیی بود. این مدل که در سال ۱۹۵۹ ارائه شد در بین تمامی مدل‌ها ساده‌ترین مدل می‌باشد. الگوریتم تکامل آن به شکل زیر است:

> برای تولید یک گراف تصادفی اردوش رنیی با N راس و n یال، از بین $\frac{N(N-۱)}{۲}$، n یال به تصادف انتخاب می‌کنیم.

لذا گراف تصادفی اردوش رنیی دست به انتخاب یکی از گراف‌ها از بین $C^n_{\frac{N(N-۱)}{۲}}$ گراف ممکن می‌زند. گراف دودویی که معادل گراف اردوش رنیی هست را نیز می‌توان به این صورت تعریف کرد:

> با شروع از N راس برای هرجفت از راس‌ها یالی با احتمال p اضافه می‌شود.

با در نظر داشتن این تعریف تعداد یال‌های مورد انتظار مساوی $E(|e|) = p\frac{N(N-۱)}{۲}$ می‌باشد. احتمال اینکه گرافی با N راس و n یال به عنوان گراف حاصل از الگوریتم گراف تصادفی دودویی باشد، برابر است با

$$P(G) = p^n(۱-p)^{\frac{N(N-۱)}{۲}-n}$$

که در آن G گراف مورد نظر می‌باشد.

نظریه گراف تصادفی ویژگی‌های احتمالاتی گرافی با N راس را زمانی که $N \to \infty$ مورد مطالعه قرار می‌دهد. بسیاری از ویژگی‌های چنین گراف‌های تصادفی را می‌توان بوسیله محاسبات آماری بدست آورد [۴۸]. احتمال وجود این ویژگی‌ها در گراف با $N \to \infty$ به ۱ نزدیکتر می‌شود. با اتکا به این ویژگی گراف‌ها اردوش و رنیی

[۱] نامگذاری مدل‌ها براساس نام ارائه دهندگان مدل انجام شده‌است.

به محاسبی برخی از ویژگی‌های آماری گراف تصادفی پرداختند. مهمترین هدف نظریه گراف تصادفی مشخص کردن این موضوع است که در چه احتمال pای برخی از ویژگی‌های مورد نظر پدیدار می‌شوند. بزرگترین یافته اردوش و رنیی این بود که تعداد زیادی از مهمترین ویژگی‌های نظریه گراف به صورت ناگهانی پدیدار می‌شوند. بدین معنا که تقریبا همه گراف‌های حاصل از فرایند مدلسازی اردوش رنیی با یک احتمال خاص، یا دارای ویژگی خاصی (مثلا وجود مسیری بین هر دو راس) هستند، یا هیچکدام دارای این ویژگی نیستند. گذار از حالتی که احتمال وجود ویژگی بسیار کم باشد به حالتی که احتمال وجود همان ویژگی بسیار زیاد باشد، خیلی سریع رخ می‌دهد. برای اغلب ویژگی‌ها یک احتمال بحرانی $p_c(N)$ وجود دارد. بدین معنا که $P(N)$ با سرعت کمتری نسبت به $p_c(N)$ رشد می‌کند، در این حالت زمانی که $N \to \infty$ تقریبا همه گراف‌هایی که با احتمال ارتباطی $P(N)$ تولید می‌شوند، دارای یک ویژگی خاص خواهد بود. لذا احتمال اینکه یک گراف با N راس و احتمال ارتباطی $p = p(N)$ دارای ویژگی Q باشد برابر است با [۴۸]

$$\lim_{N\to\infty} P_{N,p}(Q) = \begin{cases} \circ & \text{اگر } \frac{p(N)}{p_c(n)} \to \infty \\ ۱ & \text{اگر } \frac{p(N)}{p_c(n)} \to \circ \end{cases} \qquad (۱-۲)$$

با در نظر گرفتن این موضوع و با استفاده از محاسبات آماری می‌توان به نتیجه زیر دست یافت:

- میانگین درجه گراف برابر است با:

$$\langle k \rangle = ۲n/N = p(N-۱) \simeq pN \qquad (۲-۲)$$

علاوه بر این می‌توان در مورد وجود زیرگرافی خاص مانند درخت یا گراف کامل با استفاده از محاسبات آماری بحث کرد.

۲.۱.۲ مدل الحاق امتیازی

توزیع درجات پوآسون در گراف تصادفی دودویی و توزیع درجات توانی در شبکه‌های پیچیده و خاصیت استقلال از مقیاس این شبکه‌ها نیاز به ارائه مدل جدیدی را نمایان ساخت. در سال ۱۹۹۹ باربارزی و آلبرت [۴۷] مدل جدیدی را ارائه دادند که توزیع درجات نمایی را دارا بود. این مدل از الحاق امتیازی استفاده می‌کرد. وجود الحاق امتیازی باعث به وجود آمدن گراف‌هایی می‌شود که هم دارای توزیع درجات توانی هستند، و هم خاصیت استقلال از مقیاس را از خود بروز می‌دهند:

در مدل الحاق امتیازی فرض بر این است که احتمال اتصال راسی به راس i متناسب با درجه k راس i است.

این مدل با مدل اردوش رنیی دو تفاوت عمده دارد: نخست آنکه در مدل الحاق امتیازی $\Pi(k)$ (احتمال اینکه راس i دارای درجه k باشد) وابسته به k است در حالی که در مدل اردوش رنیی این مقدار $\Pi(k) = p$ می‌باشد. دوم آنکه، تابع $\Pi(k)$ برای مدل الحاق امتیازی برحسب k خطی است. این مدل هر دو نیاز ذکر شده در ابتدای این بخش را برآورده کرد.

۳.۱.۲ مدل DEB

با کشف این نکته که شبکه‌های اجتماعی خاصیت جهان کوچک را از خود بروز می‌دهند، لزوم ارائه مدلی برای برآورده کردن این نیازها حس شد. همانطور که اشاره شد گرافی که دارای خاصیت جهان کوچک است دو ویژگی را از خود بروز می‌دهد:

۱. ضریب خوشه‌بندی بزرگ نسبت به گراف تصادفی،

۲. کوچک بودن میانگین کوتاهترین مسیر بین دو راس (میانگین کوتاهترین مسیرها باید نسبت به تعداد راس‌ها حداکثر از درجه لگاریتمی باشد).

از مهمترین عوامل کوتاه بودن مسیر بین راس‌ها حضور راس‌هایی است که درجه بسیار بالا دارند. این راس‌ها که قطب خوانده می‌شوند، در شبکه‌های اجتماعی با شرکت‌پذیری بالا با دیگر قطب‌ها در ارتباطند. همین مسئله باعث کوچک‌تر شدن میانگین طول کوتاهترین مسیر می‌شوند.

اما تولید گرافی که دارای ضریب‌خوشه بندی بزرگی باشد نیاز به جستجوی محلی دارد. نخستین بار دیویدسون و همکاران (۲۰۰۱) مدلی ارائه دادند که به جستجوی محلی می‌پرداخت و ضریب خوشه بندی بزرگی داشت. مدل DEB در دسته مدل‌های تکاملی پویا قرار می‌گیرد. تعداد راس‌ها از پیش مشخص است؛ و یال‌ها بوسیله قوانینی به آن افزوده و یا از آن کم می‌شوند.

الگوریتم تکامل مدل DEB را می‌توان به صورت زیر تشریح کرد:

این مدل از یال افزایی نوع اول که در بخش ۵.۳.۱ بخش پیش تشریح شد، استفاده می‌کند. یعنی دو نفر توسط یک دوست مشترک به هم معرفی می‌شوند. همچنین این مدل از حذف یال

که در ۵.۳.۱ معرفی شد، استفاده می‌کند. طریقه حذف، درج یال‌ها و تکامل مدل به صورت زیر است. در هر مرحله:

۱. یک راس به تصادف انتخاب شده و دو دوست خود را به تصادف انتخاب کرده و در صورتی که آن دو راس با هم در ارتباط نباشند، یالی بینشان رسم می‌شود. در حالتی که راس انتخابی کمتر از دو همسایه داشت، او خودش را به یک راس دیگر معرفی می‌کند. یعنی یک راس به تصادف انتخاب می‌کند و در صورت عدم وجود یال بینشان، یالی بین این دو راس ایجاد می‌شود.

۲. یک راس به تصادف انتخاب می‌شود، با احتمال p به همراه کلیه یال‌هایش از گراف حذف می‌شود. چون در این مدل تعداد راس‌ها ثابت است بلافاصله یک راس به گراف اضافه می‌شود، و یک یال بین این عضو جدید و یکی از اعضای قدیمی که به تصادف انتخاب می‌شوند، ایجاد می‌شود.

ویژگی‌های این مدل عبارتند از [۴۶]:

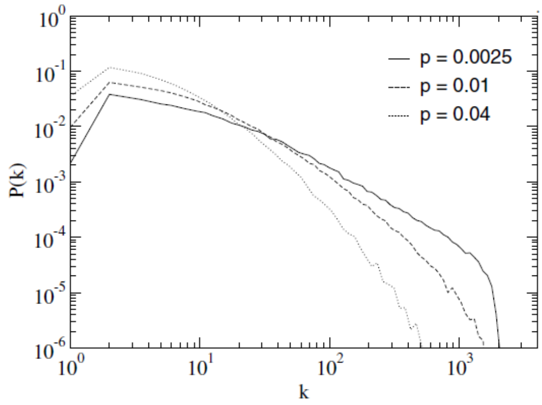
۱. تعداد راس‌های این مدل ثابت است. مدل از تغییراتی که ممکن است در تعداد اعضای یک شبکه اجتماعی رخ بدهد، چشم پوشی می‌کند.

۲. طول عمر محدود راس‌ها باعث می‌شود بعد از مدتی شبکه به یک حالت پایدار برسد. ادامه تکرارها باعث تغییر در ویژگی‌های گراف نشود.

۳. مدل ویژگی‌های توزیع درجات توانی و استقلال از مقیاس را از خود بروز می‌دهد.

۴. اگر تعداد راس‌ها به اندازه کافی زیاد باشد، توزیع درجات گراف، به تنها متغیر قابل تغییر یعنی p وابسته می‌شود.

همانطور که در ویژگی‌های مدل هم اشاره شد، مقدار p نقش مهمی را در این مدل به عنوان تنها متغیر آزاد بازی می‌کند. در واقع در هر شبکه اجتماعی روابط در عرض چند دقیقه یا چند ساعت ایجاد می‌شوند، این در حالی است که مقیاس زمانی اضافه شدن یا کم شدن اعضا به شبکه اجتماعی می‌تواند شامل ماه یا سال باشد. بنابراین مقدار $p \ll ۱$ در نظر گرفته می‌شود [۴۶]. در شکل ۱.۲ که از [۴۶] استخراج شده‌است، توزیع درجات برای مقادیر مختلف p نشان داده شده است. به خاطر محدود بودن طول عمر هر راس در گراف، تعداد راس‌های



شکل ۱.۲: نمودار توزیع درجات راس‌ها در گراف تولیدی توسط مدل DEB برای $N = ۷۰۰۰$

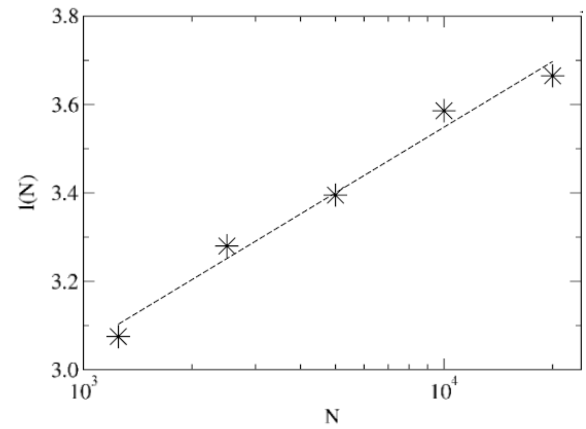
جدول ۲-۱: مقایسه ضریب خوشه‌بندی مدل ارائه شده با گراف تصادفی و بیشنه مقدار ضریب‌خوشه بندی برای گراف تصادفی برای $N = ۷۰۰۰$. [۴۶]

p	$\langle k \rangle$	$\langle k^2 \rangle$	C	C'	C_{rand}
0.04	14.9	912	0.45	0.036	0.0021

همسایه یک راس از نظر تعداد رشد همیشگی ندارد، و در نهایت به یک دامنه مقادیر مشخصی محدود می‌شود. این موضوع را می‌توان در نمودار شکل ۱.۲ مشاهده کرد، که در آن نقطه برش [۲] برای مقادیر بزرگ k نشان دهنده طول عمر محدود هر راس است. برای $p \ll ۱$ توزیع درجات وابسته به فرآیند یال افزایی است، که در نتیجه نمودار برای مقادیر کوچکتر p شیب کمتری را داراست. برای مقادیر بزرگتر p فرایند حذف و درج برای تاثیر گذاری بر روی نمودار توزیع درجات با فرایند یال افزایی به رقابت می‌پردازد، و در نتیجه در توزیع درجات منعکس می‌شود و در نهایت برای $p \approx ۱$، با توجه به خاصیت جستجوی سراسری حذف و درج راس، نمودار به نمودار پواسون(دارای شیب تند) تبدیل می‌شود. لذا، مدل DEB گرافی تولید می‌کند که دارای توزیع درجات توانی است. برای مقادیر بزرگ N توزیع درجات $P(k)$ فقط به مقدار پارامتر آزاد یعنی p وابسته است.

جدول ۲-۱ ضریب خوشه بندی و میانگین درجات راس‌های گراف را برای گراف حاصل از این مدل (C)، گراف تصادفی (C_{rand}) و گرافی با بیشینه مقدار ضریب‌خوشه بندی که توسط گراف تصادفی بدست آمده (C') را نشان می‌دهد. مقادیر گراف تصادفی و گراف مدل DEB دارای تعداد راس‌ها و میانگین درجات یکسانی هستند.

[۲] cutoff



شکل ۲.۲: نمودار میانگین کوتاه‌ترین مسیرها نسبت به تعداد رئوس برای مدل DEB

بیشینه مقدار ضریب‌خوشه بندی برای گراف تصادفی توسط نیومن و همکاران [۴۹] محاسبه شده‌است:

$$C' = \frac{۱}{\langle k\rangle N}\Big(\frac{\langle k^{۲}\rangle}{\langle k\rangle} - ۱\Big)^{۲} \quad (۲-۳)$$

همانطور که جدول ۲-۱ نشان می‌دهد، ضریب‌خوشه بندی برای مدل بیشتر از ضریب خوشه‌بندی برای مدل تصادفی و حتی بیشتر از مقدار بیشینه ضریب خوشه‌بندی برای گراف تصادفی است. لذا مدل یکی از ویژگی‌های مورد نیاز جهان کوچک را تامین می‌کند.

ویژگی دومی که جهان کوچک نیاز دارد میانگین کوتاه‌ترین مسیرهای گراف است. این مقدار باید نسبت به اندازه گراف یا تعداد راس‌ها از درجه لگاریتمی باشد. نمودار شکل ۲.۲ که از [۴۶] آورده شده است نشان‌دهنده رشد لگاریتمی میانگین کوتاهترین مسیر نسبت به اندازه گراف است.

نکته‌ای که شاید ابهام آمیز باشد وجود خاصیت توزیع توانی درجات بدون وجود خاصیت الحاق امتیازی در گام‌های الگوریتم است. باید توجه داشت که خاصیت الحاق امتیازی که مهمترین عامل ایجاد شبکه‌های دارای توزیع درجات توانی است به صورت ضمنی در این مدل موجود است. در گام اول یکی از راس‌ها به تصادف انتخاب می‌شود و اگر درجه‌اش کمتر از ۲ باشد، یک یال بین آن و یک راس دیگر به تصادف ایجاد می‌شود. در غیر این صورت در گام یال افزایی از بین همسایگان یک راس، یکی به تصادف انتخاب می‌شود. دو راس i و j را در نظر بگیرید. به طوری که $d_i > d_j$ باشد. یکی از راس‌های گراف با احتمال $\frac{۱}{N}$ انتخاب می‌شود که N تعداد راس‌های گراف است. اگر درجه راس i برابر d_i باشد؛ احتمال اینکه راس انتخاب شده برای یال افزایی راس i یا j باشد برابر است با:

$$p(C_i) = \frac{۱}{N}d_i > p(C_j) = \frac{۱}{N}d_j \quad (۲-۴)$$

همانطور که مشاهده می‌شود احتمال انتخاب نسبت به d_v یا درجات راس‌ها خطی است. که این نیز نشان‌دهنده حضور الحاق امتیازی در الگوریتم است.

این مدل با هدف شبیه‌سازی خاصیت جهان کوچک مدلی را طراحی کرد، و توانست این نیازها را برآورده کند. یک سال بعد از ارائه این مدل توسط دیویدسون و همکاران، واسکوز [3] [۵۰]نشان داد که جستجوی محلی همچنین می‌تواند باعث ضریب خوشه‌بندی بالا و همبستگی درجه-درجه یا همان شرکت‌پذیری گراف می‌شود.

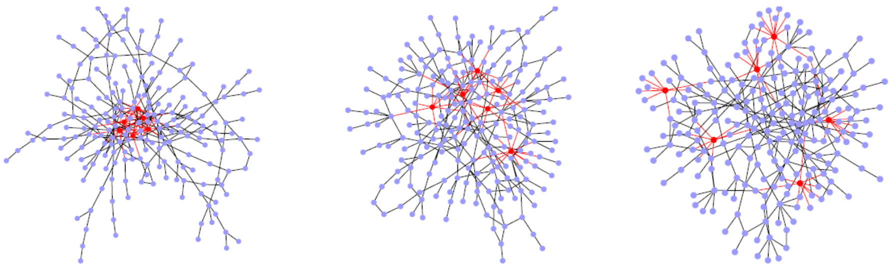
۲.۲ رشد شبکه‌ها با قوانین محلی

همانطور که در بخش پیش اشاره شد، رفته رفته با کشف ویژگی‌هایی که اغلب شبکه‌های اجتماعی از خود بروز می‌دهند، نیاز به ارائه مدل‌هایی که بازتاب دهنده این ویژگی‌ها باشد، احساس شد. مدل DEB (۲۰۰۲) به منظور شبیه‌سازی ویژگی‌های جهان کوچک ارائه شد. سال بعد یعنی در سال ۲۰۰۳ واسکوز [۵۰] در مقاله‌ای با عنوان "رشد شبکه‌ها با قوانین محلی: الحاق امتیازی، سلسله مراتب خوشه‌بندی، و درجه همبستگی" به بررسی تاثیر جستجوی محلی بر روی ویژگی‌هایی چون توزیع درجات، ضریب‌خوشه‌بندی و همبستگی بین راس‌ها با استفاده از توزیع درجات راس‌های همسایه پرداخت. در این مقاله چند مدل برپایه عابر تصادفی ارائه و مقایسه‌ای بین داده‌های حاصل و داده‌های واقعی انجام شده‌است. در نهایت مدل اصلی که از هر دو روش یال افزایی استفاده می‌کند ارائه شده و تاثیر این نوع یال افزایی‌ها برای شبیه‌سازی ویژگی‌های شبکه‌های اجتماعی نشان‌داده شده‌است. به‌علاوه برای هر مدل سه ویژگی توزیع درجات، ضریب‌خوشه‌بندی و همبستگی درجات، که بر مبنای میانگین درجات نزدیکترین همسایه بنانهاده شده، بررسی شده‌است. در این بخش به بررسی و مقایسه دو مدل از مدل‌های ارائه شده برای شبیه‌سازی شبکه‌های اجتماعی که بیش از همه ویژگی‌های شبکه‌های اجتماعی را شبیه‌سازی می‌کنند، پرداخته می‌شود.

۱.۲.۲ همبستگی درجات

یک شبکه معمولا وقتی دارای همبستگی درجه به درجه خوانده می‌شود که راس‌ها با الگوی خاصی به هم متصل شده باشند. وقتی راس‌های با درجه بالا به هم و راس‌های با درجه پایین به هم متصل باشند، شرکت‌پذیری

[3] Vazquez



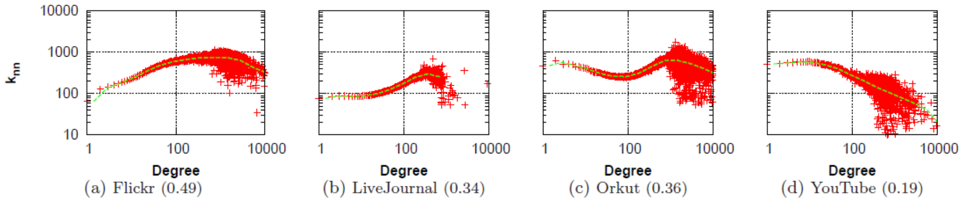
شکل ۳.۲: سه شبکه سه حالت اشتراک‌پذیر، اشتراک‌ناپذیر و غیروابسته درجه به درجه را نشان می‌دهند. راس‌های قرمز نشان‌دهنده قطب‌ها هستند. در شکل سمت چپ قطب‌ها به هم متصل هستند و راس‌های با درجه پایین‌تر به هم، لذا این گراف شرکت پذیر است و در شکل وسط هیچ وابستگی بین اتصالات وجود ندارد و در شکل سمت راست قطب‌ها به راس‌های با درجه پایین‌تر متصل است لذا گراف اشتراک‌ناپذیر است.

مثبت و هرگاه راس‌های با درجات بالا به راس‌های با درجه پایین متصل باشد شرکت‌پذیری منفی (اشتراک‌ناپذیر) است.

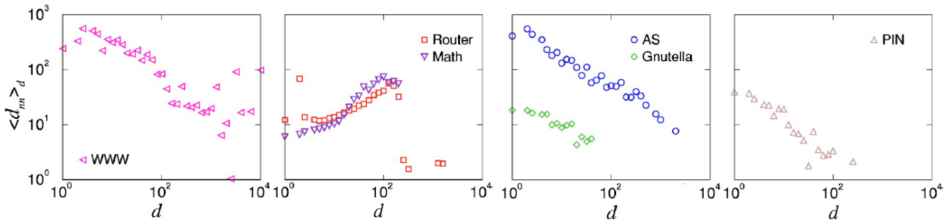
اغلب وابستگی درجه به درجه با ضریب شرکت‌پذیری شبکه که توسط نیومن در [۷] معرفی شد، محاسبه می‌شود. ضریب شرکت‌پذیری در واقع ضریب همبستگی پیرسون بین بردارهای درجات در دو سوی یک یال، به عنوان تابعی از همه یال‌ها می‌باشد. به عبارت دیگر اگر ضریب اشتراک پذیری مثبت یا منفی باشد، تابع وابستگی درجات دارد، در غیر این صورت تابع وابستگی درجات ندارد. به زبان ساده‌تر هرگاه احتمال اتصال دو یال به هم به درجات آن‌ها وابسته باشد (یعنی درجات بزرگ به بزرگ و کوچک به کوچک یا درجات بزرگ به کوچک) گراف هبستگی درجه به درجه دارد و اگر احتمال اتصال وابسته به درجه راس‌ها نباشد، گراف همبستگی درجه به درجه ندارد.

در شکل ۳.۲ سه گراف مشاهده می‌کنیم که در آن راس‌های قرمز نشان‌دهنده قطب‌ها و راس‌های آبی نشان دهنده راس‌های با درجات پایین‌تر هستند. در گراف سمت چپ قطب‌ها به هم متصل هستند و راس‌های با درجه پایین‌تر به هم وصل می‌شوند. لذا این گراف نمایانگر یک گراف اشتراک‌پذیر است. از طرف دیگر در گراف وسط هیچ چیزی در مورد اتصال قطب‌ها به هم یا به راس‌های دیگر نمی‌توان گفت، لذا این گراف، گرافی است که همبستگی درجه به درجه ندارد. در نهایت در گراف سمت راست، راس‌های قطب با راس‌های با درجه پایین‌تر مرتبط هستند، لذا این گراف یک گراف اشتراک‌ناپذیر است.

یکی از بهترین معیارها برای محاسبه وابستگی درجه به درجه، میانگین درجات همسایه‌ها است که می‌توان به صورت $\langle d_{nn} \rangle$ نشان داد. شکل ۴.۲ که از [۲۰] گرفته شده‌است نشان دهنده نسبت میانگین درجات همسایه‌ها



شکل ۴.۲: نمودار میانگین درجات همسایه برای شبکه‌های اجتماعی [۲۰]



شکل ۵.۲: نموادر میانگین درجات برای شبکه‌های پیچیده (غیر اجتماعی) [۵۰]

به درجه راس‌ها می‌باشد. در این شکل چهار شبکه اجتماعی فلیکر، لیوجورنال، اورکوت و یوتیوب بررسی شده است.

شکل ۵.۲ چهار شبکه پیچیده را نشان می‌دهد. روتر [۴] گراف اینترنت در سطح روتر می‌باشد؛ یعنی راس‌های این گراف را روترها تشکیل می‌دهند و ارتباط فیزیکی بین روترها به عنوان یال‌های گراف در نظر گرفته می‌شود. AS[۵] گرافی است که در آن هر راس یک ارائه دهنده سرویس اینترنت می‌باشد و ارتباط بین این سرویس دهنده‌ها در جهان واقعی در گراف به یال تبدیل می‌شوند. WWW نشان دهنده گرافی است که هر راس آن یک صفحه وب و هر یال آن گراف نشان دهنده لینک‌های بین این صفحات می‌باشد (یال‌های جهت دار، بدون جهت در نظر گرفته شده‌است). gnutella گرافی است که نمایانگر شبکه‌های دو به دو متصل با نام یکسان هستند. در این گراف هر راس نشان دهنده یک کاربر و هر یال نشان دهنده یک ارتباط بین این کاربرها می‌باشد. PIN گرافی است از تعامل پروتئین‌ها، که هر راس نشان دهنده یک پروتئین و هر یال نشان دهند تعامل بین پروتئین‌هاست. Math گرافی است که همکاری بین پژوهشگران ریاضی برای ارائه مقاله را نشان می‌دهد. هر راس نشان‌دهنده یک پژوهشگر می‌باشد و یالی بین این راس‌ها وجود دارد اگر و تنها اگر پژوهشگران متناظر دارای مقاله‌ای مشترک باشند.

همانطورکه مشاهده می‌شود، اغلب همه شبکه‌ها، همبستگی درجه به درجه دارند. همه شبکه‌های اجتماعی به

[۴] Router
[۵] autonomous system

جز یوتیوب دارای همبستگی درجات با شرکت‌پذیری مثبت هستند؛ یعنی اغلب قطب‌ها با هم ارتباط دارند و راس‌های با درجه پایین‌تر با هم در ارتباط هستند. اما شبکه‌های پیچیده به جز روتر دارای وابستگی درجه به درجه با شرکت‌پذیری منفی هستند. البته روتر نیز دارای وابستگی درجه به درجه با شرکت‌پذیری مثبت است. پژوهشگران نمایش متفاوت یوتیوب را در [۲۰] به خاطر ویژگی محبوبیت محور این شبکه دانسته‌اند که در آن یک عده افراد مشهور از سوی تعداد زیادی مردم عادی دنبال می‌شوند.

این دو نوع وابستگی به صورت کامل در [۵۰] مورد بررسی قرار گرفته و براساس این ویژگی مدل‌های مختلف ارائه شده‌است. در ادامه این بخش به بررسی دو مدل (از بین مدل‌های مختلف) ارائه شده توسط واسکوز می‌پردازیم. این مدل‌ها را هم به لحاظ آماری و هم با شبیه‌سازی و از طریق داده‌های عددی بررسی می‌نماییم.

۲.۲.۲ مدل عابر تصادفی

معرفی مدل عابر تصادفی

مدل ارائه شده با عنوان عابر تصادفی شامل دو گام افزودن و قدم‌زدن است. این مدل از یک حالت اولیه شروع به کار کرده و به مرور زمان تکامل می‌یابد.

- **حالت اولیه:** مدل با یک راس بدون یال شروع به کار می‌کند و مراحل زیر را تا زمان اتمام مدلسازی تکرار می‌نماید.
- **افزودن:** یک راس جدید ایجاد می‌شود و با یک یال به یکی از راس‌ها که به تصادف انتخاب شده‌است، متصل می‌شود.
- **قدم‌زدن:** اگر راس i به وسیله یالی به راس j متصل است با احتمال q_e، یک یال بین i و یکی از همسایه‌های j (به غیر از i) که به تصادف انتخاب شده‌است ایجاد می‌گردد. هرگاه یالی ایجاد نشد به مرحله افزودن برمی‌گردیم.

برای بررسی آماری این مدل می‌توانیم از حرکت عابر تصادفی بر روی گراف شبکه جهانی وب استفاده کنیم. نخست احتمال اینکه یک راس در یک جستجو چه سراسری (به وسیله موتور جستجو در وب) و چه محلی (به وسیله دنبال کردن لینک‌های ارائه شده در صفحه وب) ملاقات شود برابر است با:

$$v_i = \frac{۱ - q_e}{N} + q_e \sum_j J_{ij} \frac{v_j}{d_j^{ou}} \qquad (۵-۲)$$

در این رابطه q_e احتمال ایجاد یال به واسطه جستجوی محلی و $۱ - q_e$ احتمال ایجاد راس جدید، J_{ij} یک درایه از ماتریس مجاورت، d_j^{ou} درجه خروجی راس j است. به کمک تخمین میدان میانگین می‌توانیم رابطه ۲-۵ را به صورت زیر بازنویسی کنیم:

$$v_i = \frac{۱ - q_e}{N} + q_e \Theta d_i^{in} \tag{۲-۶}$$

که در این رابطه Θ میانگین احتمال مشاهده یک راس است که یالی به i دارد و d_i^{in} درجه ورودی راس i می‌باشد. با توجه به [۵۰] در نهایت

$$\Theta = \frac{v_a}{q_v v_s N} \tag{۲-۷}$$

که در آن v_a تعداد راس‌های افزوده شده در هر مرحله افزودن، v_s تعداد عابران تصادفی و q_v احتمال اضافه شدن یک یال زمانی که یک راس با احتمال q_e انتخاب می‌شود. q_v به این خاطر افزوده شده‌است که احتمالا یک یال پیش از این و توسط عابر دیگری به گراف افزوده شده‌است.

برای یافتن تعداد راس‌های با درجات مختلف نیاز داریم تا احتمال اینکه درجه یک راس با درجه $d^{(in)}$ هنگام عبور عابر تصادفی یک واحد افزایش یابد را محاسبه کنیم. لذا احتمال افزایش یک واحدی درجه راسی با درجه $d^{(in)}$ برابر است با $A(d^{(in)}) = q_v v(d^{(in)})$ که در این رابطه $v(d^{(in)})$ نشان‌دهنده احتمال انتخاب شدن برای ایجاد یال جدید برای راس‌های با درجه $d^{(in)}$ است. با توجه روابط ۲-۶ و ۲-۷ داریم:

$$A(d^{(in)}) = \frac{۱}{N}\left[q_v(۱ - q_e) + q_e \frac{v_a}{v_s} d^{(in)}\right] \tag{۲-۸}$$

باید توجه داشت که مدل عابر تصادفی، الحاق امتیازی را در خود نهفته دارد. در رابطه ۲-۸، یک رابطه خطی بین احتمال رشد درجه راس با درجه آن پیش از افزودن یال وجود دارد، یعنی هرچه درجه یک راس بیشتر باشد احتمال رشد درجه آن نیز بیشتر می‌شود. حال می‌توانیم تعداد راس‌ها با درجه ورودی d^{in} را به صورت زیر محاسبه کنیم.

$$\frac{\partial n_{d^{in}}}{\partial t} = v_s A_{d^{in}-۱} n_{d^{in}-۱} - v_s A_{d^{in}} n_{d^{in}} + v_a \delta_{d^{in}۰} \tag{۲-۹}$$

در این رابطه $v_a \delta_{d^{in}۰}$ احتمال ایجاد راسی با درجه d می‌باشد. می‌توانیم نرخ رشد تعداد صفحات اضافه شده را با اندازه گراف متناسب در نظر بگیریم، یعنی تعداد صفحات افزوده شده با تعداد کل گراف متناسب است ($v_a \propto N$). همچنین می‌توانیم نرخ رشد تعداد عابران را با نرخ رشد گراف متناسب در نظر بگیریم اگر $v_s \propto N$.

آنگاه داریم:

$$\frac{v_s}{v_a} = \alpha \tag{۲-۱۰}$$

که در آن α مقداری ثابت است. باید توجه داشت که ۲-۱۰ همیشه برای شبکه‌هایی که نرخ رشد ثابتی دارند، صادق است. اگر این نرخ رشد موجود باشد توزیع درجه یال‌های ورودی به یک حالت پایدار می‌رسد و برابر $n_{d^{in}}(t) = Np_{d^{in}}$، که در این رابطه $p_{d^{in}}$ احتمال این است که در حالت پایدار یک زاس انتخابی دارای درجه d^{in} باشد. با انتگرال گیری از رابطه ۲-۹ و جایگزاری این مقدار در آن در نهایت بدست می‌آوریم:

$$p_{d^{in}} \sim (d^{in})^{\gamma} \tag{۲-۱۱}$$

که در این رابطه

$$\gamma = ۱ + \frac{۱}{q_e}. \tag{۲-۱۲}$$

لذا مشاهده می‌کنیم که مدل عابر تصادفی دارای یک توزیع درجات توانی برای یال‌های ورودی با ضریب توانی $\gamma \geq ۲$ می‌باشد.

حال با بدون جهت در نظر گرفتن یال‌ها قصد داریم تا تعداد یال‌ها بین همسایه‌های یک یال را محاسبه کنیم. فرض کنیم e_i نشان دهنده تعداد یال‌هایی باشد که بین همسایگان i وجود دارد. با توجه به این که تنها عامل تکامل، تحرک عابر تصادفی است داریم:

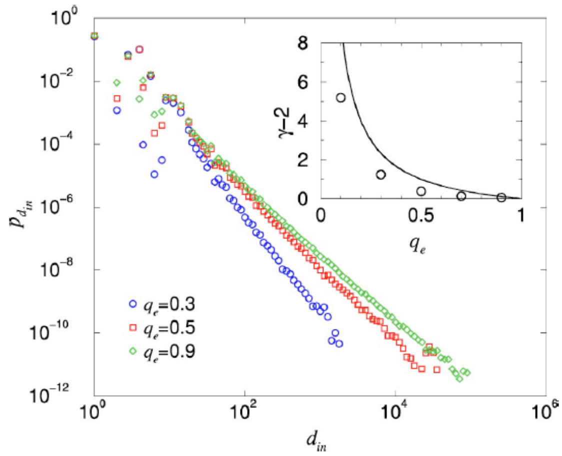
$$\frac{\partial e_i}{\partial t} = q_v(q_e \Theta d_i^{in} + q_e v_i) \tag{۲-۱۳}$$

در این رابطه در سمت راست جمله اول احتمال انتخاب شدن یک راس در گراف است که یالی به i دارد و عنصر دوم احتمال انتخاب شدن راس j است که راس i یک یال خروجی به آن دارد. حال با استفاده از روابط ۲-۵، ۲-۷ و ۲-۸ و با در نظر گرفتن این‌که $\partial_t d_i^{in} = A(d_i^{in})$ داریم:

$$\frac{\partial e_i}{\partial t} \approx (۱ + q_e)\frac{\partial d_i^{in}}{\partial t} \tag{۲-۱۴}$$

که در این رابطه از عنصر اول در سمت راست رابطه ۲-۸ چشم‌پوشی کرده‌ایم. با انتگرال‌گیری از رابطه ۲-۱۴ با شرط مرزی $e(d^{in} = ۰) = ۰$ ضریب خوشه بندی را بدست می‌آوریم:

$$\langle c \rangle_d = \frac{۲e(d)}{d(d-۱)} = \frac{۲(۱+q_e)}{d} + \frac{۲(۱+q_e)(۱-d^{ou})}{d(d-۱)} \tag{۲-۱۵}$$



شکل ۶.۲: نمودار توزیع درجات ورودی برای مدل تولید شده با 10^6 راس. این نمودار میانگین بروی ۱۰۰ گراف تولید شده می‌باشد[۵۰].

که برای مقادیر بزرگ d داریم:

$$\langle c \rangle \approx \frac{2(1+q_e)}{d} \qquad (16\text{-}2)$$

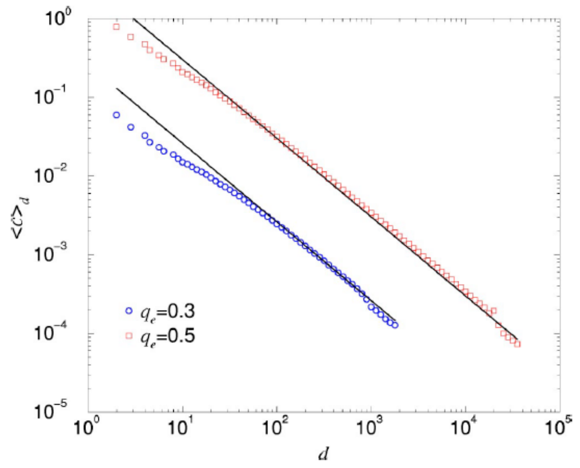
از رابطه ۲-۱۶ می‌توان به رابطه عکس ضریب‌خوشه بندی با درجه راس‌ها پی برد.

با این اوصاف برای ساده‌سازی الگوریتم عابر تصادفی نخست مقادیر $v_a = 1 - q_e$ و $v_s = 1$ قرار می‌دهد، و هر عابری پس از عدم توانایی، در ایجاد راس کار خود را متوقف کرده و یک عابر جدید تولید شده و شروع به پیش‌روی می‌کند. همچنین مقدار $q_v = 1$ در نظر گرفته می‌شود.

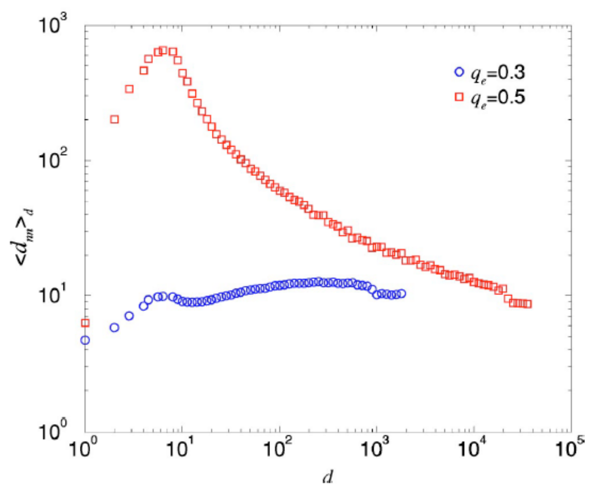
در شکل ۶.۲ میانگین ۱۰۰ بار اجرای الگوریتم مدل عابر تصادفی برای تولید توزیع درجات گراف با 10^6 برای مقادیر مختلف q_e نشان داده شده‌است. نمودار ضریب توانی γ را برای مقادیر مختلف q_e که در شبیه‌سازی عددی بدست آمده را نشان می‌دهد. همانطورکه از نمودار داخلی (نمودار کوچک) شکل ۶.۲ واضح است، زمانی که $q \to 0$، ضریب توزیع توانی مقادیر بسیار بزرگی را دارا می‌باشند. لذا نمی‌توانیم توزیع توانی را از توزیع نمایی متمایز کنیم، اما با $q \to 1$ مقدار γ به کمترین مقدار خود یعنی $\gamma = 2$ نزدیک می‌شود.

رابطه ۲-۱۶ نشان داد که ضریب خوشه بندی راس‌ها نسبت به درجه آن‌ها دارای رابطه معکوس است. اثبات عددی این نکته در شکل ۷.۲ قابل مشاهده است. در این نمودار نسبت بین ضریب خوشه‌بندی به درجه راس‌ها برای دو مقدار q_e نشان داده شده‌است.

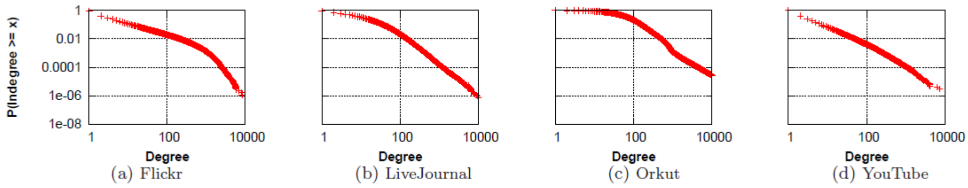
همانطور که واسکوز نیز اشاره نموده‌است، بدست آوردن رابطه‌ی آماری برای پیش‌بینی میانگین درجات همسایه‌ها



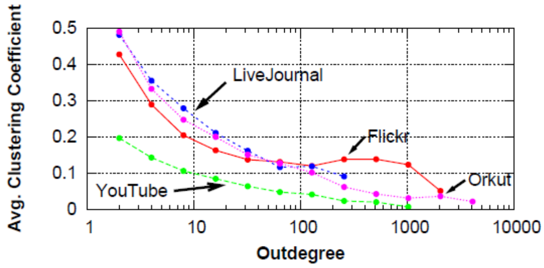
شکل ۷.۲: رابطه بین ضریب‌خوشه‌بندی و درجه راس‌ها برای مدل عابر تصادفی با ۱۰^۶ راس که نمودار حاصل میانگین محاسبه شده برای ۱۰۰ گراف تولید شده می‌باشد[۵۰].



شکل ۸.۲: رابطه بین میانگین درجات همسایه با درجه راس برای گرافی با اندازه ۱۰^۶. نمودار حاصل میانگین محاسبه شده برای ۱۰۰ گراف تولید شده می‌باشد[۵۰].



شکل ۹.۲: [۲۰].توزیع درجات ورودی گراف شبکه‌های اجتماعی برخط فلیکر، لیوجورنال، اورکوت و یوتیوب



شکل ۱۰.۲: نمودار رابطه بین ضریب‌خوشه‌بندی به درجه راس‌ها برای شبکه‌های اجتماعی فلیکر، لیوجورنال، اورکوت و یوتیوب [۲۰].

کار بسیار دشواری است. به همین دلیل بدون بدست آوردن رابطه آماری برای این ویژگی، به بررسی ویژگی‌های مدل پیاده‌سازی شده عابر تصادفی می‌پردازیم. در شکل ۸.۲ رابطه بین میانگین درجات همسایه با درجه راس برای دو مقدار q_e نشان داده شده‌است. برای مقادیر کوچک q_e هیچ وابستگی خاصی بین میانگین درجات همسایه و درجه راس‌ها وجود ندارد. اما برای مقادیر بزرگ q_e، مدل شرکت‌پذیری منفی را از خود بروز می‌دهد. در این نمودار بیشترین مقدار میانگین درجات همسایه در $d = ۱۰$ وجود دارد.

جمع‌بندی مدل عابر تصادفی

با مقایسه شکل ۶.۲ و ۹.۲ درمی‌یابیم که با توجه به وجود خاصیت الحاق امتیازی این مدل توزیع درجات توانی را داراست. با توجه به وابستگی مقدار γ به مقدار q_e می‌توانیم مدل را طوری تولید کنیم که بتواند خاصیت توزیع درجات یک مدل واقعی را با دقت خوبی شبیه‌سازی کند.

از طرف دیگر با مقایسه شکل ۷.۲ و شکل ۱۰.۲ در می‌یابیم که این مدل می‌تواند رابطه بین ضریب‌خوشه‌بندی و درجه راس‌ها را برای شبکه‌های اجتماعی شبیه‌سازی کند. همانطور که در شکل ۱۰.۲ مشاهده می‌شود این مقادیر نسبت عکس با هم دارند. این مدل نیز همچنان که در شکل ۷.۲ دیده می‌شود، رابطه عکسی را بین

ضریب‌خوشه‌بندی و درجه راس‌ها شبیه‌سازی می‌نماید.

اما مهمترین تفاوت بین این مدل و شبکه‌های اجتماعی در رابطه بین میانگین درجات همسایه درجه‌ها و راس‌ها نمایان می‌شود. همانطور که در ابتدا این بخش بیان شد این مدل سعی دارد، که شبکه‌های پیچیده‌ای مانند شبکه پروتئین‌ها یا روترها را شبیه‌سازی نماید، این شبکه‌ها چنانکه که در شکل ۵.۲ نشان‌داده شده‌است، دارای وابستگی درجه به درجه با شرکت‌پذیری منفی هستند. اما مدل‌های شبکه‌های اجتماعی (شکل ۴.۲) وابستگی درجه به درجه با شرکت‌پذیری مثبت دارند. این مدل سعی در مدل‌سازی شبکه‌های پیچیده دارد؛ لذا با توجه به شکل ۸.۲ برای مقادیر بزرگ q_e وابستگی درجه به درجه با شرکت‌پذیری منفی را تولید می‌کند. بنابراین مدل عابر تصادفی از بروز ویژگی شرکت‌پذیری مثبت که از خاصیت‌های مهم شبکه‌های اجتماعی است ناتوان می‌باشد.

۳.۲.۲ مدل عابر تصادفی با جستجوی بازگشتی

می‌توان بر پایه مدل عابر تصادفی واسکوز مدل دیگری را نیز ارائه داد. در این مدل مرحله قدم‌زدن کمی متفاوت است:

- **قدم‌زدن:** زمانی که یک یال به یک راس ایجاد شد با احتمال q_e یالی به هریک از همسایگان آن ایجاد می‌شود. در صورت عدم تولید یال، به مرحله افزودن باز می‌گردد.

این مدل ویژگی‌های تقریبا یکسانی را با مدل عابر تصادفی از خود بروز می‌دهد.

۴.۲.۲ مدل ارتباط نزدیک‌ترین همسایگان

بررسی آماری

واسکوز با در نظر گرفتن این موضوع که " اغلب یال‌ها در یک شبکه اجتماعی بین راس‌هایی ایجاد می‌شود که یک همسایه مشترک دارند. " و با الهام گرفتن از مدل DEB مدل جدیدی را با نام مدل ارتباط نزدیک‌ترین همسایه‌ها معرفی کرد (بعد از این هرگاه به مدل Vaz اشاره کردیم منظورمان این مدل است).
این مدل سعی دارد تا هر دو نوع یال‌افزایی که در بخش ۵.۳.۱ معرفی شدند را بکار ببرد. برای تشریح این مدل نخست باید یال‌های بالقوه را تعریف کرد. می‌گوییم دو راس به وسیله یک یال بالقوه با هم در ارتباط هستند هرگاه:

۱. راس‌ها با هیچ یالی به هم وصل نشده باشند.

۲. دو راس حداقل یک همسایه مشترک داشته باشند.

رشد و تکامل گراف با استفاده از تغییر حالت جفت راس‌ها بین سه حالت جدا (s)، متصل با استفاده از یال بالقوه (p) و متصل با استفاده از یال‌های واقعی (e) تعریف و بررسی می‌شود. فرض می‌کنیم d_i^* تعداد یال‌های بالقوه راس i است. برای رسیدن به رابطه‌ای که بیانگر سیر تکامل این مدل باشد از رهیافت پیوستگی استفاده می‌کنیم، بدین معنا که با استفاده از زنجیره مارکوف و حوزه میانی تغییرات درجه و ساختار گراف را به صورت پیوسته در نظر می‌گیریم:

$$\frac{\partial d_i}{\partial N} = v_{s\to e}\hat{d}_i + v_{p\to e}d_i^* - (v_{e\to s} + v_{e\to p})d_i \tag{۲-۱۷}$$

$$\frac{\partial d_i^*}{\partial N} = v_{s\to p}\hat{d}_i + v_{e\to p}d_i - (v_{p\to s} + v_{p\to e})d_i \tag{۲-۱۸}$$

که در این دو رابطه

$$\hat{d}_i = N - d_i - d_i^*$$

می‌باشد. $v_{x\to y}$ نرخ تبدیل یال‌ها از حالت x به حالت y به ازای هر واحد از N است و $\hat{d}_i$ تعداد راس‌های باقی‌مانده است که نه به وسیله یال واقعی و نه به وسیله یال بالقوه به راس i متصل نشده‌است. ایجاد (حذف) یک یال بالقوه به یک راس زمانی اتفاق می‌افتد که یک یال به یکی از راس‌های همسایه آن اضافه (یا از آن حذف) شود. برای مثال زمانی که یک راس جدید i به یکی از راس‌های از پیش موجود j وصل شود؛ آنگاه یال‌های بالقوه بین راس i و همه راس‌های همسایه آن ایجاد می‌شود. بنابراین:

$$v_{s\to p} = v_{s\to e}d_i \tag{۲-۱۹}$$

$$v_{p\to s} = v_{e\to s}d_i$$

این روابط اساس مدل اتصال نزدیک‌ترین همسایه را تشکیل می‌دهند.

واسکوز برای ارزیابی آمار خود از هرگونه حذف یال چشم پوشی می‌کند، با این استدلال که در برخی از مدل‌ها مانند مدل همکاران پژوهشی، یال ایجاد شده هیچ‌گاه از بین نمی‌رود. البته به نظر می‌رسد که این محدویت برای شبکه‌های اجتماعی نادرست باشد، زیرا در شبکه اجتماعی احتمال از بین رفتن یک رابطه حتی خارج شدن یک فرد از یک شبکه و از بین رفتن تمام یال‌های آن وجود دارد.

تبدیل از حالت یال بالقوه به یک یال واقعی نسبت به تبدیل از حالت یال جدا به یال واقعی دارای احتمال بیشتری است. در واقع، ایجاد یال بین دو راس جدا از هم بدون همسایه مشترک فرآیندی است که ایجاد ارتباط

اجتماعی بین دو موجودیت شبکه‌های اجتماعی که به تصادف انتخاب شده‌اند را شبیه‌سازی می‌کند. فرض می‌کنیم:

$$v_{s\to e} = \frac{\mu_0}{N^2} \tag{۲-۲۰}$$

از طرف دیگر، ایجاد یک یال بین دو راس با یک همسایه مشترک (دارای یال بالقوه بینشان)، ایجاد یک رابطه بین دوستِ دوست را شبیه‌سازی می‌کند. برای این حالت نیز داریم:

$$v_{p\to e} = \frac{\mu_1}{N} \tag{۲-۲۱}$$

با داشتن این تخمین‌ها روابط ۲-۱۷ و ۲-۱۸ به صورت زیر در می‌آید:

$$\begin{aligned} N\frac{\partial d_i}{N} &= \mu_0 + \mu_1 d_i^* \\ N\frac{\partial d_i^*}{N} &= \mu_0 d_i + \mu_1 d_i^* \end{aligned} \tag{۲-۲۲}$$

لذا وجود الحاق امتیازی در این مدل (نرخ رشد خطی تعداد یال‌های واقعی و بالقوه گراف برای هر راس با نسبت d_i و d_i^*) باتوجه به رابطه ۲-۲۲ واضح است. بنابراین انتظار داریم این مدل دارای توزیع توانی باشد. با انتگرال گیری از این معادله دیفرانسیل برای $N \gg N_i$ داریم:

$$d_i(N) = d_0\Big(\frac{N}{N_i}\Big)^{\beta} \qquad d_i^*(N) = d_0^*\Big(\frac{N}{N_i}\Big)^{\beta} \tag{۲-۲۳}$$

که در این رابطه N_i، اندازه گراف زمانی که راس i افزوده می‌شود، می‌باشد، و

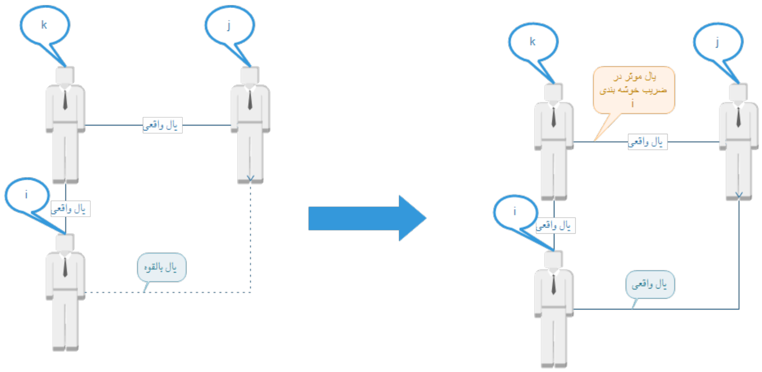
$$\beta = \frac{\mu_1}{2}\Big(-1 + \sqrt{1 + 4\frac{\mu_0}{\mu_1}}\Big) \tag{۲-۲۴}$$

حال اگر فرض کنیم راس‌ها در بازه‌های زمانی یکسانی افزوده شوند یعنی داشته باشیم $P(N_i = N) = 1/N$. می‌توانیم رابطه زیر را بدست آوریم:

$$\begin{aligned} P(d_i > d) &= P\Big[d_0\Big(\frac{N}{N_i}\Big)^{\beta} > d\Big] \\ &= \int_0^N \frac{dN_i}{N}\Theta\Big[d_0\Big(\frac{N}{N_i}\Big)^{\beta} - d\Big] \end{aligned} \tag{۲-۲۵}$$

در نتیجه داریم:

$$p_d = \frac{\partial P(d_i > d)}{\partial d} \sim d^{-\gamma} \tag{۲-۲۶}$$



شکل ۱۱.۲: تاثیر تبدیل یال بالقوه به یال معمولی در ضریب‌خوشه بندی

که در آن

$$\gamma = 1 + \frac{1}{\beta} \qquad (۲-۲۷)$$

باید در نظر داشت که آنچه باعث به وجود آمدن توزیع توانی می‌شود رابطه ۲-۱۹ می‌باشد. اما اگر $v_{s\to p}$ مستقل از درجه راس‌ها باشد توزیع درجات تبدیل به توزیع نمایی خواهد شد.

حال برای محاسبه رابطه آماری برای پیش‌بینی ضریب خوشه‌بندی راس‌ها، مهم‌ترین نقش برعهده تبدیل حالت از یال بالقوه به یال معمولی می‌باشد. در واقع، اگر یک یال بالقوه، که راس i را به راس j متصل می‌کند -که این یال بالقوه به خاطر وجود یک همسایه مشترک به نام k است- به یک راس واقعی تبدیل می‌شود، در واقع راس i علاوه بر دریافت یک یال و افزایش درجه، به واسطه وجود یال بین j و k یک یال به یال‌هایی که همسایگان i را به هم وصل کرده افزوده می‌شود. این مسئله در شکل ۱۱.۲ نشان داده شده‌است. لذا داریم:

$$\frac{\partial e_i}{\partial N} = v_{p\to c} d_i^* = \mu_1 \frac{d_i^*}{N} \qquad (۲-۲۸)$$

با انتگرال‌گیری از این رابطه با کمک رابطه ۲-۲۳ داریم:

$$\langle c \rangle_d = \frac{2e(d_i)}{d(d-1)} \approx \frac{2\mu_1}{d} \qquad (۲-۲۹)$$

لذا بار دیگر به رابطه معکوس ضریب خوشه‌بندی و درجه راس می‌رسیم.

پیش از بیان الگوریتم کامل باید ذکر کنیم که این مدل از هر دو نوع یال افزایی ذکر شده در بخش ۵.۳.۱ استفاده می‌کند. گرچه این دو نوع یال افزایی عملا تفاوت ذاتی چندانی با هم ندارند، اما هر یک از مدل‌هایی که پیش از این ذکر شد؛ یکی از روش‌های یال افزایی به کار می‌گیرد. اما این مدل با تبدیل یال‌های بالقوه به یال‌های واقعی عملا هر دو نوع یال افزایی را شبیه‌سازی می‌کند.

الگوریتم مدل ارتباط نزدیک‌ترین همسایگان

با شروع از یک راس تنها و بدون یال به عنوان گراف یکی از مراحل زیر را به صورت تکراری تا خاتمه مدلسازی انجام می‌دهیم:

۱. با احتمال $u - ۱$ یک راس جدید را به گراف اضافه می‌کنیم و از این راس یک یال به یکی از راس‌های دیگر گراف j، که به تصادف انتخاب می‌شود، رسم می‌کنیم. این کار باعث ایجاد یال‌های بالقوه بین راس i و همسایه‌های j می‌شود.

۲. با احتمال u یکی از یال‌های بالقوه را به یال معمولی تبدیل می‌کنیم.

واسکوز در این مدل سعی در ارائه مدلی دارد که مدل DEB را شبیه‌سازی کند. اما در این مدل برخلاف مدل DEB تعداد راس‌های گراف در طول فرایند ثابت نیست. البته باید توجه داشت همانطور که واسکوز نیز اشاره کرده برای گراف‌های بسیار بزرگ هر دو، ویژگی‌های یکسانی را از خود بروز می‌دهند.
با قرار دادن

$$\mu_{۰} = ۱, \qquad \mu_{۱} = \frac{u}{u - ۱} \tag{۲-۳۰}$$

و با استفاده از روابط ۲-۲۴ و ۲-۲۷ داریم:

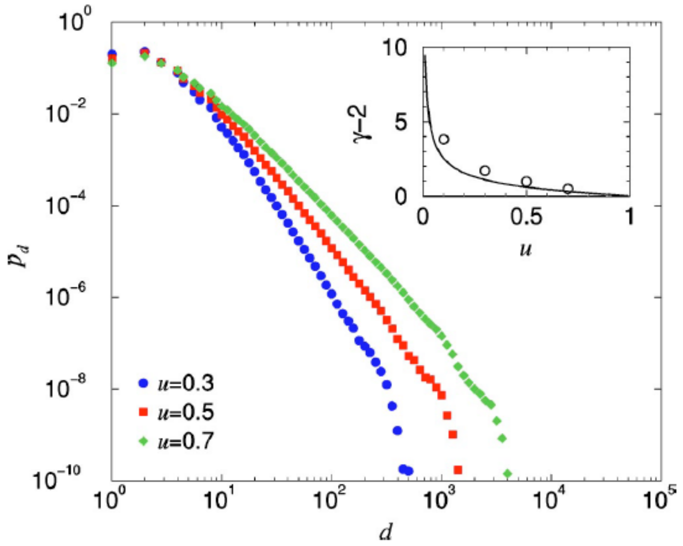
$$\gamma(u) = ۱ + \frac{۲(۱ - u)}{u}\Big(- ۱ + \sqrt{۱ + ۴\frac{۱ - u}{u}}\Big)^{-۱} \tag{۲-۳۱}$$

لذا داریم:

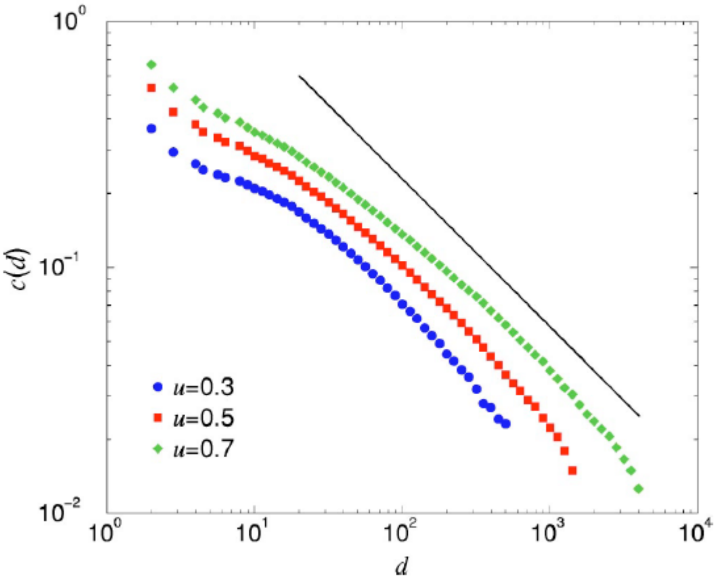
$$\gamma(۰) = \infty, \qquad \gamma(۱) = ۲ \tag{۲-۳۲}$$

بنابرای ضریب توزیع توانی کمترین مقدار خود را $u \to ۱$ داراست. یعنی در حالتی که هیچ راس افزایی نداشته باشیم.

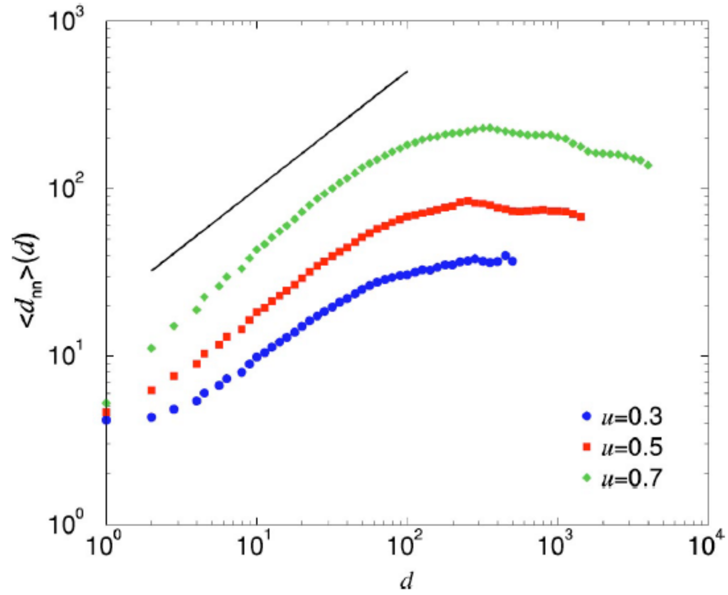
شکل ۱۲.۲ حاوی نمودار توزیع درجات مدل می‌باشد. همانطور که پیش‌بینی‌های آماری نشان می‌دهد، این مدل توزیع درجه توانی را داراست. شکل ۱۳.۲ رابطه بین ضریب خوشه‌بندی و درجه راس‌ها را نشان می‌دهد که همانطور که روابط آماری نیز نشان می‌داد، رابطه‌ای معکوس بین این دو وجود دارد. اما مهمترین مورد برای این مدل در شکل ۱۴.۲ نشان داده می‌شود. برخلاف مدل عابر تصادفی این مدل همبستگی درجه به درجه



شکل ۱۲.۲: نمودار توزیع درجات برای مدل اتصال نزدیک‌ترین همسایگان برای گرافی به اندازه 10^6. نمودار میانگین ۱۰۰ بار اجرای مدل می‌باشد [۵۰].



شکل ۱۳.۲: نمودار میانگین نزدیک‌ترین همسایه به درجه راس، برای مدل اتصال نزدیک‌ترین همسایگان برای گرافی به اندازه 10^6. نمودار میانگین ۱۰۰ بار اجرای مدل می‌باشد [۵۰].



شکل ۱۴.۲: نمودار ضریب خوشه‌بندی به درجه راس برای مدل اتصال نزدیک‌ترین همسایگان برای گرافی به اندازه ۱۰^۶. نمودار میانگین ۱۰۰ بار اجرای مدل می‌باشد [۵۰].

را با شرکت‌پذیری مثبت از خود نشان می‌دهد؛ که این مسئله مطابق با شکل ۴.۲ شرکت‌پذیری مثبت را برای شبکه‌های اجتماعی نشان می‌دهد. با این اوصاف می‌توان گفت که مدل اتصال نزدیک‌ترین همسایه‌ها نسبت به مدل عابر تصادفی ویژگی‌های شبکه‌های اجتماعی را بهتر مدلسازی می‌کند.

۳.۲ فراز و فرود شبکه‌های اجتماعی

بعد از ارائه مدل DEB و Vaz ، مارسلی [۶] و همکاران در مقاله‌ای تحت عنوان "فراز و فرود شبکه‌های اجتماعی: یک مدل رسمی" [۵۱] سعی کردند با تغییراتی که در یال‌افزایی و حذف در مدل DEB به وجود می‌آورند، به نتایج بهتری برای شبیه‌سازی شبکه‌های اجتماعی دست یابند. مارسلی و همکاران در این مدل به جای یال افزایی نوع اول که در DEB استفاده شده بود از یال‌افزایی نوع دو استفاده کردند. به‌علاوه به جای استفاده از حذف یال از حذف راس برای قرار دادن مدل در یک حالت پایدار استفاده نمودند. همچنین آن‌ها در این مقاله به بررسی تغییرهای ناگهانی که در این مدل (به عنوان حالتی خاص از مدل‌های تصادفی) رخ می‌دهد، می‌پردازند. در ادامه این بخش به معرفی این مدل ،که با نام MVS شناخته می‌شود، خواهیم پرداخت. در نهایت با تحلیل آماری برآوردی از ویژگی‌های این مدل خواهیم داشت.

[۶] Marsili

۱.۳.۲ الگوریتم مدل MVS

این مدل از یال‌افزایی نوع دوم استفاده می‌کند. همانطور که پیش‌تر نیز اشاره شد، در این نوع یال‌افزایی یک راس از یکی از دوستان خود که به تصادف یا براساس یک توزیع احتمالاتی خاص انتخاب شده‌است می‌خواهد که او را به یکی از دوستانش معرفی کند. بعد از دیویدسون و همکاران [۴۶] که یال‌افزایی نوع اول را معرفی کردند، واسکوز [۵۰] این نوع یال‌افزایی را پیشنهاد داد. این دو یال افزایی به صورت ذاتی با هم تفاوت چندانی ندارند. یعنی می‌شود یکی از انواع یال‌افزایی را به کمک دیگری شبیه‌سازی کرد. برای مثال اگر i از j بخواهد که او را به یکی از دوستانش مثلا k معرفی کند، همانند این است که از j خواسته شود که دو عدد از دوستانش را به هم معرفی کند. تفاوت تنها در این است که در هر مرحله در یال‌افزایی نوع دوم درجه راس i با احتمال ξ یک واحد اضافه می‌شود. در حالی که در یال افزایی نوع اول درجه دو تا از همسایگان i در هر مرحله با احتمال ξ افزایش می‌یابد. با این حال به دلیل سهولت استنتاج ویژگی‌های آماری در اغلب مدل‌ها از یال افزایی نوع دوم که واسکوز معرفی کرد استفاده می‌شود.

مدل از n راس که از پیش مشخص شده‌است ساخته می‌شود. این مدل در هر لحظه از زمان، به وسیله یک گراف (غیر جهت‌دار) $\Gamma(t) = \{N, g(t)\}$ مشخص می‌شود که در آن $N = \{۱, ۲, \ldots, n\}$ تعداد راس‌هاست و $g(t) \subseteq N \times N$ نشان‌دهنده یال‌های گراف می‌باشد.

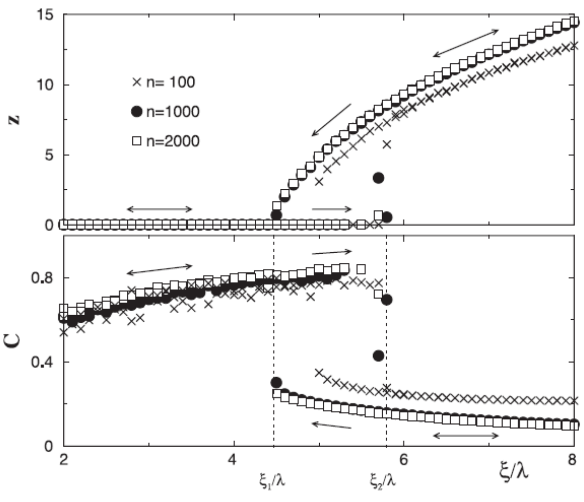
الگوریتم ساخت مدل:

۱. در هر بازه زمانی $[t, t + dt)$ هر یال موجود $ij \in g(t)$ با احتمال λdt از گراف حذف می‌شود.

۲. با احتمال ηdt برای هر راس i، اجازه جستجوی سراسری داده می‌شود. به این معنی که بین راس i و راس j که به تصادف انتخاب شده‌است، یالی ایجاد می‌شود.

۳. با احتمال ξ برای هر راس امکان جستجوی محلی فراهم می‌شود. برای مثال با احتمال ξ به راس i امکان جستجوی محلی داده می‌شود. این راس به صورت تصادفی از یکی از همسایگان j خود می‌خواهد که او را (i) به یکی از دوستانش یعنی k معرفی کند. در صورتی که i یا j هیچ یالی نداشته باشند، یالی افزوده نمی‌شود.

همانطور که از الگوریتم مشخص است، مدل MVS دارای سه پارامتر (λ, η, ξ) می‌باشد که در بازه زمانی $[t + dt)$ بر روی مدل اثر می‌گذارند. می‌توانیم با تنظیم کردن تیک‌های زمانی یکی از این پارامترها را از مدل حذف کنیم.



شکل ۱۵.۲: تغییرات میانگین درجات و ضریب خوشه بندی نسبت به تغییرات ξ/λ بدین منظور $\eta = ۰/۰۰۱$ و $\lambda = ۰/۱$ قرار داده شده و اندازه گراف $n = ۱۰۰, ۱۰۰۰, ۲۰۰۰$ می‌باشد [۵۱].

مارسلی و همکاران از مقیاس میانگین درجه $z(t)$ به جای مقیاس چگالی (نسبت بین تعداد یال‌های گراف به تعداد کل یال‌های ممکن)، برای نشان دادن مقدار تراکم یال‌ها در گراف استفاده کردند. آن‌ها با استفاده از این معیار تغییرات تراکم گراف تولیدی را نسبت به ξ/λ نشان دادند.

پیش از بررسی آماری مدل به تغییراتی که با تغییر ξ/λ در ضریب خوشه بندی C و میانگین درجات گراف z حاصل می‌شود، نگاهی می‌اندازیم. شکل ۱۵.۲ از [۵۱] در اینجا آورده شده‌است. این شکل دو نمودار ضریب‌خوشه بندی و میانگین درجات را نسبت به تغییرات ξ نشان می‌دهد. برای $\xi \ll \lambda$ گراف تولیدی مدل مانند تبدیل یک گراف تصادفی با میانگین درجات $z = ۲\eta/\lambda$ می‌شود. برای ξ/λ کوچک مدل از تعداد زیادی مولفه‌های جدا از هم تشکیل شده‌است. برای مقدار کوچک ξ، طریقه رشد گراف تنها توسط جستجوی سراسری η و نرخ حذف یال λ تعیین می‌شود. اغلب مولفه‌ها از دو راس به هم متصل تشکیل می‌شوند، و در غیر این صورت جستجوی محلی به سرعت یال‌ها را از بین گزینه‌های محدود دوستِ دوست (به علت تعداد محدود دوستان) ایجاد می‌کند. به علت محدود بودن گزینه‌ها (دوست دوست) و همچنین احتمال ایجاد یال، جستجوی محلی نقشی تعیین کننده در ساختار گراف ندارد. ناگهان در یک نقطه بحرانی $\xi_۲$ یک مولفه عظیم (مولفه عظیم یک مولفه است که درصد زیادی از کل گراف را شامل می‌شود) بخش زیادی از گراف را در بر می‌گیرد و در واقع توده عظیم به هم پیوسته به وجود می‌آید. میانگین درجات گراف در این نقطه از نمودار ناگهان به صورت جهشی زیاد می‌شود. گراف با افزایش ξ چگال‌تر می‌شود. همه این‌ها در یک حالت پایدار رخ می‌دهد. یعنی بعد از گذر زمان مدل با درج و حذف یال‌ها حالت ناپایدار را پشت سر می‌گذارد و به یک حالت پایدار می‌رسد.

ضریب خوشه بندی برای مقادیر مختلف ξ/λ رفتاری غیر بدیهی از خود نشان می‌دهد. در حالتی که گراف از تعداد زیادی مولفه کوچک تشکیل شده، با رشد ξ مقدار ضریب خوشه‌بندی گراف رشد می‌کند. همانطورکه اشاره شد، در این حالت اغلب گراف از تعداد زیادی مولفه کوچک تشکیل شده که تعداد کمی راس را به هم متصل کرده‌است. در این حالت ضریب خوشه‌بندی طبق انتظار باید زیاد باشد چون مولفه‌های همبندی اغلب تعداد کمی راس را دارند که به هم متصل‌اند و این مولفه‌های کوچک ضریب خوشه بندی بزرگی را دارا هستند. از سوی دیگر در حالتی که گراف چگالی بالایی دارد، ضریب خوشه بندی مقدار نسبتا کوچکی را داراست. با افزایش ξ ضریب خوشه بندی کاهش می‌یابد.

اما نمودار شکل ۱۵.۲ به نکته دیگری نیز اشاره دارد. اگر نمودار مربوط به میانگین درجات گراف را در نظر بگیریم، با حرکت از چپ به راست و افزایش ξ میانگین درجات تا زمان رسیدن به ξ_2 سیر یکنواختی را طی می‌کند. اگر قبل از رسیدن به ξ_2/λ (مدل درحالت پایدار قرار دارد)، ξ/λ را افزایش یا کاهش دهیم، میانگین درجات سیری قابل پیش‌بینی خواهد داشت. حال اگر گراف برای $\xi/\lambda > \xi_2/\lambda$ مدلسازی شود یعنی حالتی که مولفه عظیم ایجاد شده، و سپس رفته رفته ξ/λ را کمتر کرده و به کمتر از ξ_2/λ برسانیم تا جایی، که $\xi/\lambda \sim \xi_2/\lambda$ آن‌گاه با پدیده جالبی مواجه می‌شویم. در این حالت تا زمانی که به ξ_1/λ نرسیده‌ایم مولفه عظیم از بین نمی‌رود. در این نقطه مولفه عظیم از هم پاشیده و دوباره گراف از مولفه‌های کوچک جدا از هم تشکیل می‌شود. همین حالت برای ضریب خوشه بندی نیز رخ می‌دهد، یعنی تا زمان تشکیل مولفه عظیم یعنی ξ_2/λ با افزایش ξ/λ ضریب خوشه‌بندی افزایش می‌یابد. اما بعد از تشکیل مولفه عظیم و جهش منفی که ضریب خوشه‌بندی انجام می‌دهد اگر ξ/λ را کم کنیم، هیچ جهش عظیمی تا زمان رسیدن به ξ_1/λ رخ نمی‌دهد. بعد از این جهش باز ضریب خوشه‌بندی سیر منطقی و طبیعی را پی می‌گیرد. مارسلی و همکاران ناحیه $[\xi_1/\lambda, \xi_2/\lambda]$ را ناحیه بحرانی نامیدند.

موارد ذکر شده در مورد میانگین درجه گراف و ضریب خوشه بندی نمایانگر ثبات گراف متراکم (گرافی که در آن مولفه عظیم تشکیل شده است) در ناحیه بحرانی می‌باشد. یعنی مدل در قبال تغییرات متغیر حذف یال، مقاوم است و این مقاومت حتی در ناحیه‌ای که در آن گراف متراکم تشکیل نمی‌شود، نیز موجود می‌باشد.

۲.۳.۲ تحلیل آماری مدل MVS

محاسبه توزیع درجات گراف

با تنظیم مقیاس زمانی و قرار دادن $\lambda = ۱$ می‌توانیم با استفاده از معادلات سرآمد، نرخ تبدیل را برای محاسبه میانگین درجه و ضریب خوشه بندی به صورت زیر بنویسیم:

$$w(z_i \to z_i + ۱) = ۲\eta + \beta\theta(z_i) + \gamma z_i \qquad (۳۳\text{-}۲)$$

و

$$w(z_i \to z_i - ۱) = z_i \qquad (۳۴\text{-}۲)$$

که در این رابطه داریم:

$$\theta(x) = \begin{cases} ۰ & \text{اگر } x \leq ۰ \\ ۱ & \text{اگر } x > ۰ \end{cases}$$

در رابطه ۲-۳۳ اولین عنصر در سمت راست نشان دهنده احتمال افزایش درجه با کمک جستجوی سراسری است. ضریب ۲ به این خاطر است که هر راس می‌تواند هم به عنوان مبدا و هم به عنوان مقصد یک جستجوی سراسری باشد. یک راس با احتمال η به عنوان مبدا یک جستجوی محلی یالی را ایجاد می‌کند. همچنین این راس با احتمال $\eta\frac{۱}{N-۱}$ به عنوان مقصد احتمالی یک یال ایجاد شده به واسطه جستجوی سراسری توسط یکی از راس‌های دیگر گراف انتخاب می‌شود و چون این کار در هر گام برای $N - ۱$ راس تکرار می‌شود لذا احتمال اینکه یک راس در یک مرحله به عنوان مقصد یک جستجوی سراسری باشد η می‌باشد. دومین عنصر نشان دهنده احتمال افزایش ضریب خوشه‌بندی به خاطر جستجوی محلی است، که این حالت زمانی روی می‌دهد که درجه راس $z_i > ۱$ باشد. در

$$\beta = \xi(۱ - C)P\{z_j > ۱|ij \in g\} \qquad (۳۵\text{-}۲)$$

در این عبارت $(۱ - C)$ احتمال نبود یالی بین راس i و دوستِ دوست آن یعنی k می‌باشد. می‌دانیم که ضریب خوشه‌بندی احتمال وجود یالی بین همسایه‌های یک راس است. دو راس i و k همسایه‌های راس j می‌باشند. لذا احتمال وجود یالی بین آن‌ها برابر C و احتمال عدم وجود یال $۱-C$ می‌باشد. عبارت $P\{z_j > ۱|ij \in g\}$ احتمال این است که درجه راس $j > ۱$ باشد. نهایتا آخرین عنصر، احتمال راس هدف بودن i در یک جستجوی محلی می‌باشد. برای γ داریم:

$$\gamma = \xi(۱ - C)\langle z_k^{-۱}\rangle \qquad (۳۶\text{-}۲)$$

که این عبارت احتمال این است که راس k با i در ارتباط نباشد و k راس j را انتخاب نماید. همچنین عبارت $\langle z_k^{-۱} \rangle$ نشان دهنده درجه راس k می‌باشد. هم β و هم γ می‌توانند به صورت درون سازگار محاسبه شوند. می‌توان تابع تولید این مدل را با استفاده از معادله سرآمد محاسبه کرد:

$$\pi(s) = \sum_s s^z p(z) = \frac{\beta + ۲\eta(۱ - \gamma s)^{-\mu}}{\beta + ۲\eta(۱ - \gamma)^{-\mu}} \tag{۳۷-۲}$$

که در این رابطه $\mu = (۲\eta + \beta)/\gamma$ می‌باشد. با محاسبات جبری می‌توان نشان داد:

$$p(z) = \frac{۱}{\beta + ۲\eta(a - \gamma)^{-\mu}} \Big[\beta \delta_{z,۰} + \frac{۲\eta \Gamma(\mu + z)}{\Gamma(\mu)!} \gamma^z \Big] \tag{۳۸-۲}$$

لذا $p(z) \sim z^\mu$ برای مقادیر کوچک z دارای دنباله توانی است، همچنین برای مقادیر بزرگ z شیب دنباله نمایی و توزیع درجات برابر $p(z) \sim e^{-|ln\gamma|z}$ می‌شود.

رابطه ۲-۳۷ به ما این امکان را می‌دهد که احتمال $P\{z_j | ij \in g\} = \tilde{p}(u)$ را برای درجه z_j برای راس j همسایه i محاسبه کنیم. هرچه درجه j بیشتر باشد با احتمال بیشتر j همسایه i است لذا $\tilde{p}(u) \propto up(u)$، که با در نظر گرفتن تابع تولید $\tilde{\pi}(s) = s\pi'(s)/\pi'(۱)$ می‌توانیم رابطه زیر را بدست آوریم:

$$P\{z_j > ۱ | ij \in g\} = ۱ - \tilde{\pi}'(۰) = ۱ - \frac{\pi'(۰)}{\pi'(۱)} \tag{۳۹-۲}$$

$$\langle z_k^{-۱} \rangle = \frac{۱ - \pi(۰)}{\pi'(۱)} \tag{۴۰-۲}$$

حال با جایگذاری ۲-۳۹ و ۲-۶۷ در روابط ۲-۳۵ و ۲-۳۶ داریم:

$$\beta = \xi(۱ - C)\Big[۱ - \frac{\pi'(۰)}{\pi'(۱)}\Big] \tag{۴۱-۲}$$

$$\gamma = \xi(۱ - C)\frac{۱ - \pi(۰)}{\pi'(۱)} \tag{۴۲-۲}$$

با داشتن این روابط با قرار دادن $\xi \to ۰$ داریم $\beta \to ۰$ و $\gamma \to ۰$، مدل مشابه مدل تصادفی می‌شود. در نتیجه با توجه به روابط حاصله، توزیع درجات مدل به توزیع نمایی با میانگین ۲η تبدیل می‌شود. با افزایش احتمال جستجوی محلی $\xi > ۰$ مدل، توزیع نمایی خود را از دست می‌دهد.

ضریب خوشه‌بندی

برای محاسبه ضریب خوشه بندی بر روی یک راس خاص i تمرکز می‌کنیم تا تعداد راس‌های همسایه i که با هم در ارتباط هستند $Q_i = C_i z_i (z_i - ۱)/۲$ را محاسبه نماییم. جستجوی محلی به دو طریق تعداد Q_i

را افزایش می‌دهد. نخست زمانی است که خود راس i به عنوان مبدا جستجوی محلی باشد. که این احتمال مطابق آنچه پیش از این بحث شد برابر است با:

$$W_1(Q_i \to Q_i + 1) = \beta \tag{۲-۴۳}$$

i با j همسایه است و k نیز با j همسایه می‌باشد. زمانی که یک یال از i به j وصل می‌شود، در واقع ضریب خوشه‌بندی i یکی افزایش می‌یابد، زیرا پیش از این j با k در ارتباط بودند. لذا رابطه ۲-۴۳ احتمال افزایش ضریب خوشه بندی در هر جستجوی محلی که i مبدا آن است را برآورد می‌کند. شیوه دوم حالتی است که یکی از دوستان i به طور مثال j از i می‌خواهد او را به یکی از دوستانش k معرفی کند. نرخ رشد این حالت برابر است با:

$$W_2(Q_i \to Q_i + 1) = \xi \langle z_i \theta(z_i - 1) \rangle \langle z_j^{-1} \rangle (1 - C) \tag{۲-۴۴}$$

در این رابطه $1/z_j$ احتمال انتخاب i از سوی j از بین همه همسایه‌هایش می‌باشد. $1 - C$ احتمال در ارتباط نبود j و k است. ضریب $\langle z_i \theta(z_i - 1) \rangle$ به این خاطر به رابطه اضافه شده که زمانی که $i \geq 2$ باشد برای هر یک از همسایگانش این جستجوی محلی می‌تواند رخ دهد. نهایتا باید نرخ کاهش تعداد یال‌ها را به واسطه حذف یال گراف محاسبه کنیم:

$$W_\lambda(Q_i \to Q_i - 1) = \frac{\langle z_i \theta(z_i - 1) \rangle C}{2} \tag{۲-۴۵}$$

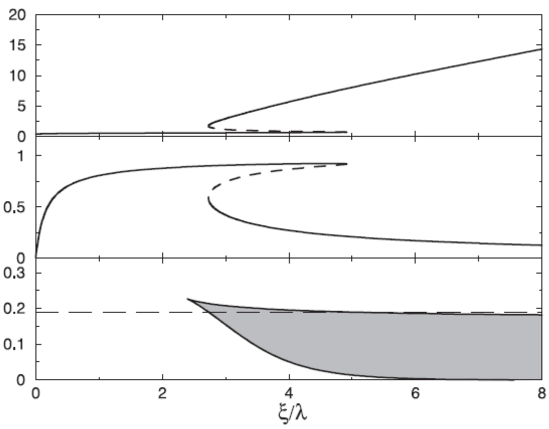
با محاسبه میانگین z_i و z_j با احتمال‌های p و $\tilde{p}$، نهایتا به یک حالت پایدار می‌رسیم:

$$\langle \Delta Q_i \rangle = W_1 + W_2 - W_\lambda = 0 \tag{۲-۴۶}$$

و نهایتا با محاسبات آماری داریم:

$$\frac{C}{2}\pi''(1) = \xi(1 - C)[2 - \pi(0)]\Big[1 - \frac{\pi'(0)}{\pi'(1)}\Big] \tag{۲-۴۷}$$

نتایج بدست آمده در شکل ۱۶.۲ دقیقا تایید کننده نتایج عددی بدست (شکل ۱۵.۲) می‌باشند. شکل ۱۶.۲ همچنین دارای نمودار پیش‌بینی فاز به منظور شناسایی ناحیه بحرانی که در آن توپولوژی‌های مختلفی همزیستی دارند، می‌باشد. در قسمت هاشور خورده روابط ۲-۴۱ ، ۲-۴۲و ۲-۴۷، دارای سه جواب است که دو تای آن‌ها پایدار و معادل دو حالت همزیست در نمودار و جواب سوم ناپایدار است که در نمودار با خط چین نشان داده شده‌است. این حالت ناپایدار حوزه دو جواب پایدار را از هم جدا می‌کند.



شکل ۱۶.۲: تغییرات میانگین درجات و ضریب خوشه بندی نسبت به تغییرات ξ/λ با توجه به روابط استخراجی برای میانگین درجه و ضریب خوشه‌بندی. به علاوه نمودار پایین تاثیر پارامترهای مدل به تشکیل ناحیه بحرانی را نشان می‌دهد. [۵۱].

به طور خلاصه تغییرات غیر بدیهی ضریب خوشه بندی را با تغییر پارامترها مشاهده نمودیم. به علاوه این مدل به طور کامل، با وجود استفاده از جستجوی محلی، توزیع درجات توانی را از خود بروز نمی‌دهد و دنباله نمودار توزیع نمایی پیروی می‌کند.

مدل MVS دومین مدل (بعد از DEB) از نوع مدل‌های تکاملی شبکه پویا بود. نوع دیگر این دسته، که در ادامه معرفی خواهد شد، با استفاده از انتصاب وزن به یال‌ها، شبکه‌های اجتماعی را مدلسازی می‌کند.

۴.۲ ظهور خوشه‌بندی، همبستگی و انجمن‌ها

بوگونا [۷] و همکاران در سال ۲۰۰۳ برای شبیه‌سازی ویژگی‌هایی شبکه اجتماعی، که آن‌ها به عنوان امضای شبکه‌های اجتماعی واقعی از آن یاد می‌کنند، نظیر ضریب‌خوشه‌بندی، همبستگی درجات و ساختار انجمنی مدلی (BPDA) را ارائه کردند [۵۴]. این مدل سعی داشت خاصیت وجود انجمن را در گراف خود (برخلاف سه مدل قبلی) وارد کند. به علاوه مدل ارائه شده برای شبیه‌سازی به صورت بنیادین با تمام مدل‌هایی که تا آن زمان ارائه شده بود متفاوت می‌نمود. مدل‌هایی که تا پیش از معرفی این مدل معرفی شده بودند، همگی یال‌ها را به صورت تصادفی یا با احتمال خاصی که به ویژگی‌های محلی راس مربوط بود اضافه می‌کردند. اما این مدل به کلی رهیافت متفاوتی در پیش گرفت.

[۷] Boguna

طبیعی‌است احتمال دوست بودن دو نفر که در ایران زندگی می‌کنند از احتمال دوست بودن یک نفر در ایران و یک نفر در آمریکا بیشتر باشد. حال در ایران احتمال اینکه دو دوست جنسیت یکسانی داشته باشند از احتمال اینکه دو نفر جنسیت مختلفی داشته‌باشند بیشتر است. این موارد را می‌توان گسترش داد. همچنین ترکیب کرد. برای مثال احتمال اینکه دو نفر در ایران در مازندران، در ساری، دارای علاقه به فلسفه با گرایش فلسفه غرب با هم دوست باشند، بسیار بیشتر از دو نفر در کشور ایران و ترکیه است که دارای علایق مشترکی نیستند. همین موضوع الهام بخش بوگونا و همکاران شد. آن‌ها هر یک از زمینه‌های مشترک از جمله ملیت، تابعیت، جنسیت و غیره را به عنوان یک بعد از فضا در نظر گرفتند. راس‌ها را براساس خصوصیت‌هایشان در این فضا پراکندند. ارتباط راس‌ها وابسته به موقعیتی بود که در این فضا اشغال می‌کردند.

این مدل برخلاف مدل‌های تکاملی شبکه از خصوصیت راس‌ها برای ایجاد ارتباط کمک می‌گیرد و با در نظر گرفتن ابعاد مختلف می‌توان آن را برای یک شبکه خاص شخصی‌سازی کرد. در ادامه این بخش نخست به معرفی مدل و الگوریتم آن خواهیم پرداخت و سپس ویژگی‌های مدل را به لحاظ آماری بررسی، آن را با داده‌های عددی مقایسه کرده و در نهایت تاثیر پارامترهای مدل را بر روی ویژگی‌های مدل نشان می‌دهیم.

۱.۴.۲ الگوریتم مدل BPDA

هدف ارائه مدلی است برای شبیه‌سازی شبکه‌های اجتماعی، که در آن یال‌ها (افراد) در یک فضای اجتماعی در یک نقطه خاص جای گرفته‌اند. در این فضا، مکان یک فرد بازتاب دهنده ویژگی‌های اوست که بر اساس هر کدام از این ویژگی‌ها یک نقطه از یک بعد خاص را اشغال می‌کنند. در نهایت جایگاه فرد در فضا بازتاب دهنده ویژگی‌های او و البته تفاوت و شباهت او با دیگر افراد جامعه است. افراد با احتمالی که با بیشتر شدن فاصله کمتر می‌شود، یال‌ها (ارتباطات) را بینشان به اشتراک می‌گذارند.

به طور کلی این مدل گرافی با ضریب خوشه‌بندی نسبتا بزرگ، به اضافه همبستگی درجه به درجه با شرکت‌پذیری مثبت را از خود بروز می‌دهد. به‌علاوه برای یک بازه خاص از احتمال ارتباط، مدل گرافی دارای خاصیت انجمنی با انجمن‌های (تشابه به لحاظ ساختاری) خود متشابه را تولید می‌کند.

الگوریتم مدل برای تولید گراف به صورت زیر است:

فرض کنید که N راس جدا از هم که در فضای اجتماعی $\mathcal{H}$ قرار گرفته‌اند را داریم.
مکان این نقاط براساس چگالی $(\rho\vec{h})$ و به‌صورت تصادفی مشخص می‌شود. برای راس i داریم:

که در این بردار $\vec{h}_i(h_i^1, ..., h_i^{d_{\mathcal{H}}})$ تعداد ابعاد فضای $d_{\mathcal{H}}$ می‌باشد. هر زیر فضا از $\mathcal{H}$ (که به وسیله مختصات بردار $\vec{h}$ مشخص می‌شود) ارائه دهنده یک ویژگی خاص اجتماعی، نظیر حرفه، دین، موقعیت جغرافیایی و غیره است. این ویژگی‌ها به وسیله متغیرهای پیوسته‌ای که توسط دامنه‌یشان با افزایش جمعیت، افزایش می‌یابد برای الگوریتم قابل استفاده می‌شوند. انتخاب یک نقطه این قیدِ جامعه واقعی را در خود دارد که "هیچ دو فرد یکسانی در جامعه وجود ندارد"، لذا با افزایش راس‌ها گرافی با تنوع بیشتر به وجود می‌آید. همچنین در این مدل زیرفضاها را مستقل از هم در نظر می‌گیریم. بنابراین می‌توانیم چگالی کلی را به این صورت بدست آوریم:

$$\rho(h) = \prod_{n=1}^{d_{\mathcal{H}}} \rho_n(h^n) \qquad (۴۸-۲)$$

با در نظر گرفتن استقلال زیرفضاها، احتمال وجود یال بین دو راس $\vec{h_i}$ و $\vec{h_j}$ برابر رابطه زیر است:

$$r(\vec{h_i}, \vec{h_j}) = \sum_{n=1}^{d_{\mathcal{H}}} \omega_n r_n(h_i^n, h_j^n) \qquad (۴۹-۲)$$

در این رابطه ω_n یک ضریب نرمال برای وزن‌دهی به هر ویژگی براساس اهمیت هرکدام از ویژگی‌هاست. مهم‌ترین نکته این مدل مفهوم فاصله دو راس در یک زیر فضاست. فرض می‌کنیم، برای دو راس i و j با مختصات $\vec{h_i}$ و $\vec{h_j}$ فاصله دو نقطه در یک زیرفضا را بدست آوریم: $d_n(h_i^n, h_j^n) \in [0.\infty), n = 1, \dots d_{\mathcal{H}}$. انتظار داریم احتمال ایجاد یال بین دو راس با افزایش فاصله کاهش یابد. لذا می‌توانیم احتمال وجود یال بین دو راس را به صورت زیر بدست آوریم:

$$r_n(h_i^n, h_j^n) = \frac{1}{1 + [b_n^{-1} d_n(h_i^n, h_j^n)]^{\alpha_n}} \qquad (۵۰-۲)$$

در این رابطه b_n مشخصه مقیاس طول است. از b_n در ادامه جهت کنترل میانگین درجه استفاده خواهیم کرد. $\alpha > 1$ ضریب هموفیلی است [۵۳]. ضریب هموفیلی احتمال ارتباط افراد متشابه در یک شبکه اجتماعی است. هرچقدر دو نفر در یک جامعه شباهت بیشتری داشته‌باشند فاصله کمتری از هم دارند، لذا این ضریب باعث دادن احتمال بیشتر به راس‌های متشابه برای به اشتراک گذاشتن یال می‌شود.

در ادامه با یک مثال پیاده‌سازی الگوریتم را به کار خواهیم برد.

۲.۴.۲ تحلیل آماری مدل BPDA

توزیع درجات شبکه را می‌توانیم با استفاده از احتمال شرطی $g(k|\vec{h})$ که عبارت از احتمال اینکه یک نقطه با مختصات اجتماعی $\vec{h}$ دارای k یال می‌باشد [۵۴] بدست آورد. لذا می‌توانیم توزیع درجات گراف را به صورت زیر بدست آوریم:

$$P(k) = \int \rho(\vec{h}) g(k|\vec{h}) dh \qquad (۲\text{-}۵۱)$$

با استفاده از توزیع دوجمله‌ای منفی می‌توانیم مقدار $g(k|\vec{h})$ را به صورت زیر بدست بیاوریم:

$$g(k|\vec{h}) = \binom{N-1}{k} \Big(\frac{\bar{k}(\vec{h})}{N-1}\Big)^k \Big(1 - \frac{\bar{k}(\vec{h})}{N-1}\Big)^{N-1-k} \qquad (۲\text{-}۵۲)$$

در این رابطه $\bar{k}(\vec{h})$ میانگین درجه یک راس زمانی که در نقطه $\vec{h}$ قرار داشته باشد، است. برای زیرفضاهای مستقل[۸]، میانگین درجات به این صورت بدست می‌آید:

$$\bar{k}(\vec{h}) = (N-1) \sum_{n=1}^{d_{\mathcal{H}}} \omega_n \int \rho_n(h'^n) r_n(h^n, h'^n) dh'^n \qquad (۲\text{-}۵۳)$$

در حالت‌هایی که ماتریس تنک و میانگین درجه ثابت باشد [۵۴]، احتمال شرطی ۲-۵۱ به شکل پوآسون در می‌آید و می‌توان توزیع درجات را با استفاده از ۲-۵۱، ۲-۵۲ و ۲-۵۳ به شکل زیر نوشت:

$$P(k) = \frac{1}{k!} \int \rho(\vec{h}) [\bar{k}(\vec{h})]^k e^{-\bar{k}(\vec{h})} dh \qquad (۲\text{-}۵۴)$$

چون افراد در فضا به صورت همگون و یکنواخت توزیع شده‌اند، توزیع درجه راس‌ها کراندار خواهد بود.
برای محاسبه ضریب خوشه‌بندی نخست نیاز داریم تا احتمال اینکه یک راس با مختصات $\vec{h}$ با یک راس دیگر با مختصات $\vec{h'}$ یالی را به اشتراک بگذارند را محاسبه کنیم. این احتمال به وسیله احتمال شرطی $p(\vec{h'}|\vec{h})$ بدست می‌آید:

$$p(\vec{h'}|\vec{h}) = (N-1)\rho(\vec{h'}) \frac{r(\vec{h}, \vec{h'})}{\bar{k}(\vec{h})} \qquad (۲\text{-}۵۵)$$

با در نظر گرفتن استقلال احتمال افزوده شدن یال بین دو راس، ضریب خوشه بندی یک راس برابر است با:

$$c(\vec{h}) = \int\int p(\vec{h'}|\vec{h}) r(\vec{h'}, \vec{h''}) p(\vec{h''}|\vec{h}) dh' dh'' \qquad (۲\text{-}۵۶)$$

---

[۸] یک زیرفضا نسبت به یک بردار مستقل است، هرگاه برای بردار $\vec{v} = (v_1, v_2, \ldots, v_n)$ در آن زیر فضا و برای عناصر $(a_1, a_2, \ldots, a_n)$ از همان زیر فضا، رابطه $a_1v_1 + a_2v_2 + \ldots + a_nv_n$ تنها زمانی برابر صفر باشد، که تمام عناصر $(a_1, a_2, \ldots, a_n)$ صفر باشد.

حال می‌توانیم میانگین ضریب خوشه بندی یال را محاسبه کنیم:

$$\langle c\rangle = \int \rho(\vec{h})c(\vec{h})dh \tag{۲-۵۷}$$

۳.۴.۲ تولید گراف به کمک مدل BPDA و بررسی تاثیر پارامترها

برای تولید گراف، ساده‌ترین نوع از مدل یعنی مدلی که دارای تنها یک بعد باشد، را در نظر می‌گیریم. فضای $\mathcal{H}$ را یک بعدی و محدود به بازه $[۰, h_{max}]$ در نظر می‌گیریم. راس‌ها را در این بازه به طور یکنواخت توزیع می‌کنیم، یعنی داریم: $\rho(h) = ۱/h_{max}$. بدین ترتیب چگالی توزیع راس‌ها در فضا یا چگالی فضای اجتماعی برابر $\delta = N/h_{max}$ می‌باشد. به این ترتیب فاصله بین دو راس برابر $d(h_i, h_j) \equiv |h_i - h_j|$ است. لذا تنها متغیر تاثیر گذار برای کنترل مدل α می‌باشد. تصاویر سمت چپ شکل ۱۷.۲ تغییرات ساختار گراف را با تغییر α نشان می‌هد. همان‌طور که در شکل هم واضح است هرچه α بزرگتر می‌شود، راس‌های بیشتری به راس‌هایی که در یک دسته خاص قرار دارند، وصل می‌شوند. راس‌هایی که در یک دسته قرار دارند، تشابه بیشتری به هم دارند.

زمانی که $\lim_{h_{max}\to\infty}$ مدل یکنواخت می‌باشد، یعنی ویژگی‌های هر راس مستقل از موقعیت آن در صفحه می‌باشد. بنابراین می‌توانیم میانگین درجات گراف را با کمک رابطه

$$\langle k\rangle = \lim_{max\to\infty} \bar{k}(h = \frac{h_{max}}{۲}) \tag{۲-۵۸}$$

و با استفاده از رابطه ۲-۵۳ بدست آوریم:

$$\langle k\rangle = \frac{۲\delta b\pi}{\alpha \sin \pi/\alpha} \tag{۲-۵۹}$$

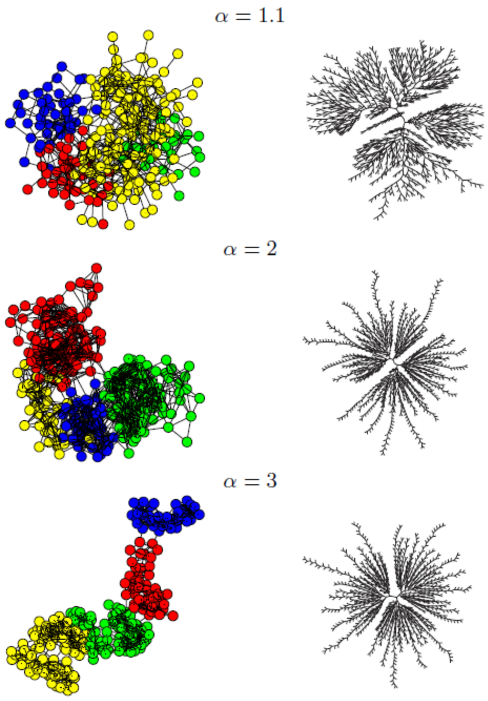
رابطه ۲-۵۹ نشان می‌دهد که با استفاده از b مناسب می‌توان گرافی با α متفاوت و میانگین درجه یکسان ساخت.

ضریب خوشه بندی را با استفاده از رابطه ۲-۵۷ بدست می‌آید:

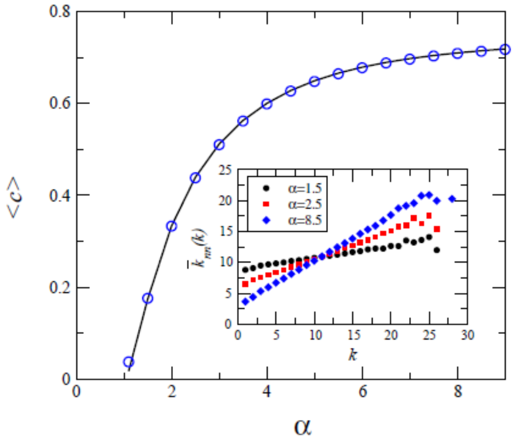
$$\langle c\rangle = \frac{\alpha^۲}{۴\pi^۲} f(\alpha) \sin^۲ \frac{\pi}{\alpha} \tag{۲-۶۰}$$

که در این رابطه

$$f(\alpha) = \int_{-} \int_{-} \frac{dxdy}{(۱ + |x|^\alpha)(۱ + |x-y|^\alpha)(۱ + |y|^\alpha)} \tag{۲-۶۱}$$



شکل ۱۷.۲: قطعه چپ تاثیر α بر روی ساختار گراف را نشان می‌دهد. قطعه سمت راست سلسله مراتب انجمن‌های استخراج شده به کمک الگوریتم نیومن گیروان را نشان می‌دهد[۵۲].



شکل ۱۸.۲: نمودار تغییر ضریب خوشه بندی و میانگین درجه نزدیک‌ترین همسایه نسبت به α [۵۲].

با بررسی شکل ۱۸.۲ مقادیر عددی تایید کننده روابط آماری است. مقدار ضریب خوشه بندی با نزدیک شدن α ، به ۱ به صفر میل می‌کند و با بزرگ شدن α ، در نهایت به ۳/۴ همگرا می‌شود.

همانطور که در شکل ۱۸.۲ نمودار داخلی واضح است، گراف دارای همبستگی درجه به درجه با شرکت‌پذیری مثبت می‌باشد. همانطور که پیش از این نیز بحث شد، شبکه‌های اجتماعی واقعی نیز این خاصیت را از خود بروز می‌دهند.

برای یافتن انجمن‌های گراف‌های تولیدی از الگوریتم نیومن گیروان [۳۰] استفاده شده است. قطعه سمت چپ شکل ۱۷.۲ نشان دهنده گراف و قطعه سمت راست نشان دهنده سلسله مراتب استخراجی از الگوریتم نیومن گیروان است. همانطور که مشاهده می‌شود با افزایش α انجمن‌ها دارای یال‌های بین انجمنی کمتری می‌باشند.

۵.۲ بهسازی مدل‌های تکاملی شبکه‌ای رشد

تویونن و همکاران [۵۵] به منظور تولید ویژگی‌های شبکه‌های اجتماعی نظیر توزیع درجات توانی، میانگین کوتاه‌ترین مسیر کوچک، ضریب خوشه بندی بزرگ، همبستگی درجه به درجه با شرکت‌پذیری مثبت و خاصیت انجمنی، مدل (TOSHK) را ارائه دادند که در دسته مدل‌های تکاملی شبکه‌ای رشد قرار می‌گیرد. این مدل سعی دارد کمبودها و نقص‌های مدل ارائه شده توسط واسکوز [۵۰] را در تولید انجمن‌ها رفع کند. این مدل با انجام یک یا بیش از یک جستجوی سراسری و یک یا بیش از یک جستجوی محلی مدل‌هایی تولید کند که هم دارای انجمن باشد (به واسطه جستجوی محلی) و هم یال‌هایی را بین انجمن‌ها داشته باشد (زمانی که بیش از یک جستجوی سراسری انجام می‌شود)، به این یال‌های بین انجمن‌ها پل می‌گوییم. بدین ترتیب انجمن‌های تولیدی شباهت بیشتر را با انجمن‌های شبکه‌های اجتماعی واقعی پیدا می‌کنند.

این مدل نیز مانند سایر مدل‌ها دو نوع جستجو را در دستور کار خود قرار می‌دهد. جستجوی سراسری و جستجوی محلی، که جستجوی محلی نقش الحاق امتیازی را نیز برعهده دارد.

تویونن و همکاران هدف خود از ارائه مدل را دست‌یابی به ویژگی‌های ذکر شده برای شبکه‌های اجتماعی و ارائه مدلی ساده ذکر کردند. هدف از ساده‌سازی مدل ایجاد توانایی تحلیل آماری می‌باشد. هرچند یافتن رابطه‌ای به منظور پیش‌بینی ضریب همبستگی بسیار دشوار می‌باشد بطوری که تقریبا هیچکدام از مدل‌های ارائه شده نتوانستند رابطه‌ای آماری به منظور پیش‌بینی این ویژگی شبکه اجتماعی ارائه دهند و تنها با پیاده‌سازی و با استفاده از داده‌های عددی این ویژگی را با داده‌های واقعی مقایسه کرده‌اند.

این مدل به عنوان اخرین مدل ارائه شده برای شبکه‌های تکاملی رشد می‌باشند. در این بخش نخست به معرفی

الگوریتم مدل خواهیم پرداخت. سپس روابط آماری را برای مدل بدست خواهیم آورد. در نهایت به مقایسه داده‌های بدست آمده از پیاده‌سازی مدل و روابط آماری می‌پردازیم.

۱.۵.۲ معرفی الگوریتم مدل TOSHK

این مدل نیز مانند مدل‌هایی که پیش از این معرفی شدند از دو نوع جستجوی سراسری و محلی استفاده می‌کند. جستجوی محلی از دنبال کردن راس‌هایی که در جستجوی سراسری انتخاب شدند،راس‌هایی را برمی‌گزینند. ماهیت محلی جستجوی نوع دوم باعث می‌شود خاصیت‌های خوشه‌بندی، شرکت‌پذیری مثبت و انجمنی در گراف به وجود بیاید. همانطور که در ادامه مشاهده خواهیم کرد توزیع توانی درجات گراف وابسته به تعداد یال‌هایی است که در جستجوی محلی ایجاد می‌شوند.

الگوریتم مدل TOSHK

۱. با یک هسته اولیه شامل $N_{\circ}$ راس اولیه شروع می‌کنیم.

۲. به طور میانگین تعداد $m_r \geq ۱$ راس را از بین راس‌های گراف به تصادف انتخاب می‌شود(جستجوی سراسری).

۳. بطور میانگین $m_s \geq ۰$ راس از همسایه‌های هرکدام از راس‌های انتخاب شده در جستجوی سراسری را انتخاب می‌شود (جستجوی محلی).

۴. یک راس ایجاد و بین راس ایجاد شده و هرکدام از راس‌های انتخاب شده در جستجوی محلی و سراسری یالی ایجاد می‌شود.

۵. مراحل ۲ تا ۴ را تا زمان رسیدن به تعداد راس‌های دلخواه به صورت تکراری ادامه می‌یابد.

یکی از موارد خاصی که ممکن است روی دهد، حالتی است که راس در مرحله ۲ انتخاب شده و در مرحله ۳ مقدار m_s بیشتر از درجه راس باشد. این حالت را حالت اشباع شده می‌گوییم. در پیاده‌سازی این مدل با توجه به انتخاب بیشینه مقدار m_s، این حالت به ندرت روی می‌دهد و قابل چشم‌پوشی است.
به منظور ساختن انجمن‌های مشابه با جهان واقعی، لازم است تعداد یال‌هایی که در جستجوی محلی ایجاد می‌شود، در هر مرحله متفاوت باشد. این یکی از تفاوت‌هایی است که این مدل با مدل واسکوز دارد که پیش‌تر به آن پرداخته شد. در مدل واسکوز در جستجوی محلی یک راس یا همه راس‌های همسایه انتخاب می‌شد.

همچنین نیاز است تا در جستجوی سراسری یک، دو یا بیشتر راس انتخاب شود. این راس‌ها نقش پل را در انجمن‌ها بازی خواهند کرد. در پیاده‌سازی این مدل از یک توزیع یکنواخت $n_{2nd} \sim U[0, k], k = 1, 2, 3$ برای تعداد راس‌های انتخاب شده در جستجوی محلی استفاده می‌کنیم. برای جستجوی محلی با احتمال $p_1 = 0.95$، یک راس و با احتمال $p_2 = 0.05$، دو راس انتخاب می‌شود. این احتمال باعث می‌شود ارتباط بین انجمن‌ها کم باشد. توزیع یکنواخت n_{2nd} جهت ساده‌سازی انتخاب شده است. با افزایش k می‌توان به دسته‌های بزرگتر و انجمن‌های قوی‌تری دست یافت.

۲.۵.۲ تحلیل آماری

۳.۵.۲ توزیع درجات

برای بدست آوردن تخمین آماری ویژگی‌های مدل از روابط رشد و حوزه میانگین استفاده می‌کنیم. مدل‌های تکاملی ارائه شده برای شبکه‌های پیچیده نظیر [۵۷]، [۵۸]، [۵۹] که از دو روش ایجاد یال الحاق تصادفی و الحاق امتیازی استفاده می‌کنند، دارای توزیع توانی $p(k) \sim k^{\gamma}$ با $2 < \gamma < \infty$ می‌باشند. مدل‌های ارائه شده نشان داده‌است که اگر افزودن یال بین راس ایجاد شده و راس‌های موجود باشد، آن‌گاه کران پایین γ برابر ۳ خواهد بود و اگر یال‌ها بین راس‌های از پیش موجود ایجاد شوند، γ مقداری بین ۲ و ۳ را نیز می‌تواند داشته باشد.

زمانی که وابستگی درجه به درجه وجود داشته باشد، با احتمال وجود یال بین راس‌هایی با درجات تقریبا مشابه(راس‌ها با درجه بزرگ به هم و راس‌ها با درجه کوچک به هم وصل شوند)، بیشتر از احتمال وجود یال بین راس‌های با درجه نامشابه است. در غیر این صورت هرچه درجه راس بیشتر باشد، با احتمال بیشتر یکی از دو سر یک یال به تصادف انتخاب شده، ظاهر می‌شود. حضور وابستگی درجه به درجه در مدل باعث انحراف از ویژگی‌های الحاق امتیازی محض می‌شود. در مدل با شروع از یک راس با درجه کم با احتمال زیادتری با انتخاب یک یال به یک راس با درجه کم دیگر خواهیم رسید. بنابراین در این مدل راس‌های قطب دارای درجه کم‌تری نسبت به مدل‌های دارای الحاق امتیازی محض هستند. همین انحراف از الحاق امتیازی محض و با توجه به اینکه در بررسی‌های آماری همبستگی درجه به درجه مورد توجه قرار نخواهد گرفت، نمودار بدست آمده از داده‌های عددی کمی متفاوت از نمودار بدست آمده از تخمین‌های آماری خواهد بود. هرچند به صورت مجانبی این دو نمودار نهایتا با هم منطبق می‌شوند.

نخست معادلات رشد، که تغییرات درجه یک راس را در طول یک گام زمانی با توجه به فرآیند رشد مدل شرح

می‌دهد را بدست می‌آوریم. این تابع فقط تغییرات درجه را مورد توجه قرار می‌دهد، لذا درجه یک راس v_i با توجه به دو فرآیند زیر رشد می‌کند:

۱. یک راس مستقیما (با استفاده از جستجوی سراسری) به راس تازه متولد شده متصل می‌شود. احتمال انتخاب راس v_i برابر m_r/t می‌باشد. چون روی هم رفته (بدون در نظر گرفتن هسته اولیه) تا لحظه t تعداد راس‌های افزوده شده برابر t می‌باشد، و تعداد راس‌هایی که به عنوان نتیجه جستجوی سراسری انتخاب می‌شوند برابر m_r می‌باشد.

۲. راس v_i در طی جستجوی محلی انتخاب شود. (در ادامه برای بدست آوردن معادلات آماری فرض می‌کنیم که احتمال انتخاب یک راس در جستجوی محلی با درجه راس ارتباط خطی دارد، و این دقیقا باعث بروز تفاوت در نمودار تحلیل‌های آماری و داده‌های عددی می‌شود که پیش‌تر به آن پرداخته شد. به این طریق در تخمین‌هایمان از وجود همبستگی درجه به درجه چشم پوشی خواهیم کرد). به طور میانگین m_s همسایه m_r راس انتخاب شده در جستجوی سراسری انتخاب می‌شوند.

با قرار دادن نتایج در رابطه داریم:

$$\frac{\partial k_i}{\partial t} = m_r\Big(\frac{۱}{t} + m_s\frac{k_i}{\sum_i k}\Big) = \frac{۱}{t}\Big(m_r + \frac{m_s}{۲(۱+m_s)}k_i\Big) \qquad (۶۲\text{-}۲)$$

که در این رابطه $۲m_r(۱+m_s)t$ به جای $\sum k$ قرار گرفته‌است. با توجه به اینکه درجه راس اولیه $k_{init} = m_r(۱+m_s)$ می‌باشد لذا درجه گراف بعد از t گام برابر $۲m_r(۱+m_s)t$ است. با انتگرال گیری می‌توان تابع رشد درجه را برای یک راس بدست آوریم:

$$k_i(t) = B\Big(\frac{t}{t_i}\Big)^{\frac{۱}{A}} - C \qquad (۶۳\text{-}۲)$$

که در این رابطه $A = ۲(۱+m_s)/m_s$ ، $B = m_r(A+۱+m_s)$ ، $C = Am_r$ می‌باشد.

با استفاده از ۲-۶۳ می‌توانیم توزیع درجات $p(k)$ را به وسیله تابع توزیع تجمعی $F(k)$ و گرفتن مشتق از تابع توزیع تجمعی با توجه به k بدست آوریم. چون در تخمین حوزه میانگین تابع $k_i(t)$ برای راس v_i از زمان به وجود آمدن آن t_i صعودی است، تعداد راس‌های تابع که درجه راس‌هایش در t کمتر از $k_i(t)$ باشد برابر با تعداد راس‌هایی است که بعد از t_i متولد شده‌اند. چون در هر گام زمانی یک راس متولد می‌شود این تعداد راس برابر است با $(t-t_i)$ که نسبت آن با کل گراف $(t-t_i)/t$ می‌باشد. لذا برای تابع توزیع تجمعی داریم:

$$F(k_i) = P(\tilde{k} \leq k_i) = P(\tilde{t} \geq t_i) = \frac{۱}{t}(t-t_i) \qquad (۶۴\text{-}۲)$$

با حل $t_i = t_i(k_i, t) = B^A(k_i + C)^{-A}t$ از رابطه ۲-۶۳ و قرار دادن آن در ۲-۶۴ و گرفتن مشتق از $F(k_i)$ نسبت به k_i و جایگزین کردن k_i با k در رابطه حاصل، می‌توانیم توزیع درجات گراف را بدست آوریم:

$$p(k) = AB^A(k + C)^{-۲/m_s-۳} \tag{۶۵-۲}$$

که A, B, C در بالا تعریف شده‌اند. لذا برای kهای بسیار بزرگ، توزیع به توزیع توانی $p(k) \propto k^{-\gamma}$ تبدیل می‌شود که برای $m_s > ۰$ مقدار γ بزرگتر از ۳ ($۳ < \gamma < \infty$) می‌شود. در مدل به خاطر حضور جستجوی سراسری $\gamma = ۳$ هرگز به دست نمی‌آید. اما هر قدر از اهمیت جستجوی سراسری کم و به اهمیت جستجوی محلی افزوده شود، مقدار γ به مقدار ۳ نزدیک‌تر می‌شود. به صورت تئوری، می‌توان مقدار $\gamma = ۳$ را بدست آورد که این موضوع به خاطر حضور الحاق امتیازی محض در برآوردهای آماری است و در عمل برای مدل TOSHK این حالت چنانکه اشاره شد روی نمی‌دهد.

۴.۵.۲ ضریب خوشه بندی

وابستگی ضریب خوشه بندی به درجه راس را نیز می‌توانیم با استفاده از معادلات رشد بدست آوریم. می‌خواهیم تغییرات تعداد مثلث‌های ایجاد شده E_i حول راس v_i را با گذشت زمان بدست آوریم. مثلث‌های اطراف راس v_i به دو طریق به وجود می‌آیند:

۱. راس v_i با احتمال m_r/t به عنوان نتیجه جستجوی سراسری انتخاب می‌شود و راس جدید انتخاب شده به تعدادی ازهمسایه‌های v_i متصل می‌شود. بدین طریق مثلث‌هایی حول راس v_i ساخته می‌شود.

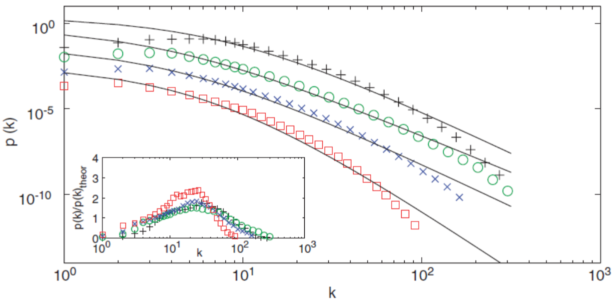
۲. راس v_i به عنوان نتیجه جستجوی محلی انتخاب می‌شود و بدین طریق مثلثی حول v_i ساخته می‌شود.

علاوه بر این حالات دیگری نیز برای افزایش تعداد مثلث حول v_i وجود دارد، نظیر حالتی که دو راس همسایه به عنوان راس‌های انتخابی جستجوی سراسری انتخاب شوند. ولی با توجه به ناچیز بودن احتمال وقوع این حالت، می‌توان از آن در محاسبات چشم‌پوشی کرد. لذا داریم:

$$\frac{\partial E_i(k_i, t)}{\partial t} = \frac{m_r m_s}{t} + m_r m_s \frac{k_i}{\sum k} = \frac{\partial k_i}{\partial t} + \frac{m_r(m_s - ۱)}{t} \tag{۶۶-۲}$$

این رابطه با قرار داردن $۲m_r(۱ + m_s)t$ به جای $\sum k$ و با استفاده از رابطه ۲-۶۲ بدست می‌آید. با انتگرال گیری از این رابطه نسبت به t و حالت اولیه $E_i(k_i nit, t_i) = m_r(۱ + m_s)$ تابع رشد تعداد مثلث‌های حول راس v_i با رشد زمان را بدست می‌آوریم:

$$E_i(t) = k_i(t) + m_r(m_s - ۱) \ln\left(\frac{t}{t_i}\right) - m_s \tag{۶۷-۲}$$



شکل ۱۹.۲: توزیع درجات داده‌های عددی بدست آمده از پیاده‌سازی مدل با $N = ۱۰^۶$ که نتایج حاصل میانگین محاسبه شده برای ۱۰۰ بار تولید گراف است. برای (+) تعداد راس انتخاب شده در جستجوی سراسری دارای یک توزیع یکنواخت $U[۱,۳]$ و تعداد راس‌های انتخاب شده در جستجوی محلی $U[۰,۲]$. (○): $p(n_{init} = ۱) = ۰/۰۹۵$، $p(n_{init} = ۲) = ۰/۰۵$ و $n_{۲nd} \sim U[۰,۳]$. (×): $p(n_{init} = ۱) = ۰/۰۹۵$، $p(n_{init} = ۲) = ۰/۰۵$ و $n_{۲nd} \sim U[۰,۲]$. (□): $p(n_{init} = ۱) = ۰/۰۹۵$، $p(n_{init} = ۲) = ۰/۰۵$ و $n_{۲nd} \sim U[۰,۱]$

با حل $\ln(t_i/t)$ نسبت به k_i با استفاده از رابطه ۲-۶۳ و قرار دادن نتیجه حاصل در ۲-۶۷ برای بدست آوردن $E_i(k_i)$ و تقسیم آن به بیشنه تعداد مثلث‌های ممکن یعنی $k_i(k_i - ۱)/۲$ می‌توانیم ضریب خوشه‌بندی را برای راس v_i بدست آوریم:

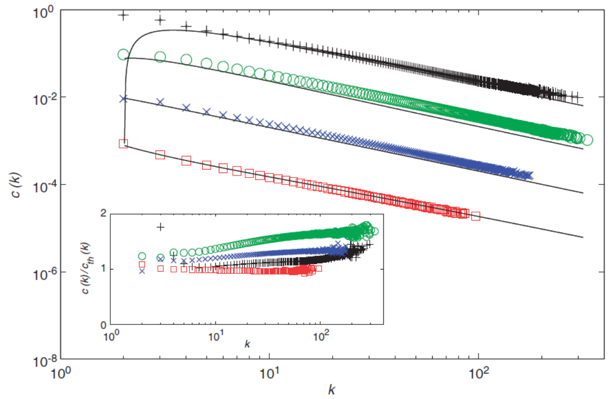
$$c_i(k_i) = \frac{۲E_i(k_i)}{k_i(k_i - ۱)} = ۲\frac{k_i + D\ln(k_i + C) - F}{k_i(k_i - ۱)} \qquad (۶۸-۲)$$

که $C = Am_r, D = C(m_s - ۱), F = D\ln B + m$. برای درجات بزرگ ضریب خوشه بندی وابسته به درجه راس است $c(k) \sim ۱/k$.

۵.۵.۲ مقایسه روابط بدست آمده با داده‌های پیاده‌سازی

شکل ۱۹.۲ توزیع درجات مدل تولید شده برای $N = ۱۰^۶$ را که نتایج میانگین ۱۰۰ بار تولید گراف است را نشان می‌دهد. همچنین این شکل نمودار تولید شده توسط برآورد آماری رابطه ۲-۶۵ را نشان می‌دهد. برآوردهای آماری در حالت مجانبی به یک توزیع توانی $p(k) \propto k^{-\gamma}$ می‌رسد (از بالا به پایین) $\gamma = ۵, ۴/۳۳, ۵$و۷. همانطور که پیش از این ذکر شد، داده‌های شبیه‌سازی با روابط آماری دقیقا منطبق نیستند، اما می‌توان انتظار داشت این دو نهایتا با هم منطبق گردند.

شکل ۲۰.۲ نمودار ضریب خوشه بندی داده‌های عددی و روابط آماری را نشان می‌دهد، همانطور که مشاهده



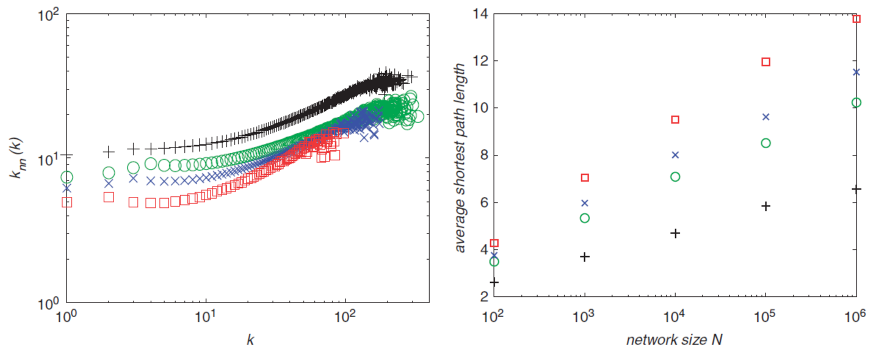
شکل ۲۰.۲: ضریب خوشه‌بندی داده‌های عددی بدست آمده از پیاده‌سازی مدل با $N = ۱۰^۶$ که نتایج حاصل میانگین محاسبه شده برای ۱۰۰ بار تولید گراف است. برای (+) تعداد راس انتخاب شده در جستجوی سراسری دارای یک توزیع یکنواخت $U[۱,۳]$ و تعداد راس‌های انتخاب شده در جستجوی محلی $U[۰,۲]$. (○): $p(n_{init} = ۱) = ۰/۰۹۵$ ، $p(n_{init} = ۲) = ۰/۰۵$ و $n_{۲nd} \sim U[۰,۳]$. (×): $p(n_{init} = ۱) = ۰/۰۹۵$ ، $p(n_{init} = ۲) = ۰/۰۵$ و $n_{۲nd} \sim U[۰,۲]$. (□): $p(n_{init} = ۱) = ۰/۰۹۵$ ، $p(n_{init} = ۲) = ۰/۰۵$ و $n_{۲nd} \sim U[۰,۱]$

می‌شود این دو نمودار تقریبا منطبق بر هم هستند. نمودار رابطه عکس ضریب خوشه بندی و درجه راس را نشان می‌دهد. ضریب خوشه بندی گراف (از بالا به پایین) برابر با: $\langle c \rangle = ۰/۴۳, ۰/۵۴, ۰/۵۸, ۰/۳۰$ است. لذا برای افزایش ضریب خوشه بندی کافی است تعداد راس‌های انتخابی در جستجوی محلی را افزایش دهیم.

در نمودار سمت چپ شکل ۲۱.۲ می‌توانیم شرکت‌پذیری مثبت تولید شده توسط مدل را که از ویژگی‌های شبکه‌های اجتماعی است مشاهده کنیم. همانطور که در شکل مشخص است، مدل دارای همبستگی درجه به درجه با شرکت‌پذیری مثبت می‌باشد.

نمودار سمت راست شکل ۲۱.۲ میانگین فاصله کوتاه‌ترین مسیر برای گراف با اندازه‌های متفاوت را نشان می‌دهد. همانطور که مشاهده می‌شود رشد کوتاه‌ترین مسیر نسبت به اندازه گراف تقریبا لگاریتمی است، و این مطابق ویژگی کوتاه‌ترین مسیر شبکه‌های اجتماعی واقعی می‌باشد.

هر دو نوع جستجو در ظهور انجمن در مدل دخیل هستند. زمانی که یک راس و در جستجوی محلی چند راس همسایه آن به عنوان نتیجه جستجوها انتخاب می‌شوند، (در گراف‌های متراکم) با احتمال زیاد همه این یال‌ها به یک انجمن تعلق دارند. لذا با افزودن یال بین راس‌های انتخاب شده و راس جدید، انجمنِ درگیر بزرگ‌تر می‌شود. حال اگر درجستجوی سراسری بیش از یک راس انتخاب شود و این راس‌ها به انجمن‌های متفاوتی

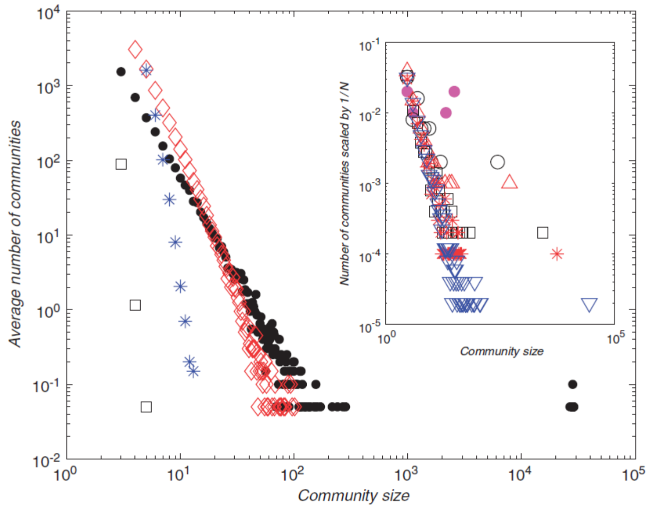


شکل ۲۱.۲: نمودار میانگین درجات همسایگان (چپ) و نمودار میانگین کوتاه‌ترین مسیرها. داده‌های عددی بدست آمده از پیاده‌سازی مدل با $N = ۱۰^۶$ که نتایج حاصل میانگین محاسبه شده برای ۱۰۰ بار تولید گراف است. نمادها همانند نمادهای استفاده شده در شکل قبل می‌باشد

متعلق باشند، راس جدید می‌تواند نقش پل را بازی کند و دو انجمن را مشابه انجمن‌های واقعی به هم متصل کند. بدین ترتیب هر دو نوع جستجو در تولید انجمن‌های مشابه انجمن‌های واقعی کمک می‌کنند.

برای یافتن انجمن‌ها اینجا از الگوریتم ارائه شده توسط پالا و همکاران [۱۹] که در بخش ۶.۲.۱ شرح داده شد، استفاده شده‌است. توزیع انجمن‌های k–دسته دارای توزیع توانی است (شکل ۲۲.۲). زمانی که از دسته‌های k تایی برای یافتن انجمن‌ها استفاده می‌شود، می‌توانیم مشاهده کنیم یک انجمن عظیم تقریبا نصف گراف را پوشش می‌دهد. شکل ۲۲.۲ تعداد انجمن‌ها با تعداد مشخصی راس را نشان می‌دهد. این مقادیر همانطور که در شکل نشان داده شده برای مقادیر مختلف k محاسبه شده‌است. با مقایسه اندازه انجمن‌ها با گراف تصادفی با توزیع درجات مشابه مدل می‌توانیم به حضور انجمن‌ها در این مدل پی‌ببریم.

با جمع‌بندی موارد بحث شده می‌توان گفت، این مدل که جزء دسته مدل‌های شبکه تکاملی رشد بود، ویژگی‌های مورد نیاز را از خود بروز می‌داد. با گسترش بحث انجمن یابی و اهمیت آن در شبکه‌های اجتماعی این مدل جزو اولین مدل‌هایی بود که ایجاد انجمن‌هایی که به انجمن‌های موجود در شبکه‌های اجتماعی واقعی نزدیک‌تر باشد را در دستور کار خود قرار داد، و با استفاده از دو مدل جستجو توانست تا حدی در ایجاد مدلی که انجمن‌های واقعی را در خود داشته باشد، موفق باشد.



شکل ۲۲.۲: نمودار میانگین تعداد انجمن‌ها برای الگوریتم پالا [۱۹] با ($\bullet : k = ۳, \diamond : k = ۴, \star k = ۵$) که برای مدل با اندازه $N = ۵۰,۰۰۰$ و $p(n_{init} = ۱) = ۰/۰۹۵$، $p(n_{init} = ۲) = ۰/۰۵$ و $n_{۲nd} = U[۰, ۳]$.

۶.۲ مدل فضایی برای شبکه‌های اجتماعی

همانطور که در بخش ۴.۲ و در تشریح الگوریتم مدل BPDA نیز اشاره شد، هموفیلی یا همانندی ،که در [۶۲] به صورت کامل توسط مک پیرسون [۹] و همکاران بحث شده‌است، به صورت رایج در شبکه‌های اجتماعی موجود است. ما تمایل داریم با کسانی ارتباط ایجاد کنیم که شبیه ما هستند. برای مثال در مدل BPDA سعی شد با فرض وجود فضای یک بعدی و پراکندن راس‌ها به طور یکنواخت بر روی این فضا و برحسب فاصله، تابع احتمالی برای وجود یال بین دو راس بدست آوریم. اما در مقاله [۶۰] وونگ [۱۰] و همکاران همانندی‌های موجود را به دو دسته همانندی مبنایی و همانندی رفتاری- نژادی (به اختصار رفتاری) تقسیم کرده‌اند. آن‌ها به همانندی‌هایی که ناشی از فاصله جغرافیایی، همانندی‌های نژادی و قومی و مانند این‌ها هستند، همانندی‌های مبنایی و به همانندی‌هایی که مربوط به علایق و خصوصیات رفتاری یک فرد می‌شود همانندی رفتاری می‌گویند. آن‌ها مشکل مدل‌های ارائه شده (که در دسته مدل‌های مبتنی بر ویژگی‌های راس قرار می‌گیرند) را با توجه به همانندی رفتاری و عدم توجه به همانندی مبنایی می‌دانند. آن‌ها یال‌های ایجاد شده به عنوان ارتباط افراد را به یک حوض یال‌های بالقوه تشبیه کرده‌اند که ارتباطات از درون این حوض انتخاب می‌شوند. از مهم‌ترین منابع همانندی فاصله جغرافیایی است. این معیار هم در تعداد ارتباطات شکل گرفته و هم در هزینه نگه‌داری ارتباط به وجود آمده تاثیر دارد.

[۹] McPherson
[۱۰] Wong

مدل ارائه شده توسط مک وونگ و همکارن [۶۰] سعی دارد عنصر همانندی مبنایی بخصوص فاصله جغرافیایی را به صورتی به مدل پیشنهادی خود وارد کند. آن‌ها با در نظر گرفتن فضای دو بعدی و توزیع پواسون راس‌ها در این فضا، سعی نمودند مدلی را ارائه دهند که علاوه‌بر دخیل کردن همانندی مبنایی، ویژگی‌های اصلی شبکه‌های اجتماعی که در فصول پیش به آن‌ها اشاره شد را در گراف تولید شده توسط مدل، نهفته داشته باشد.

۱.۶.۲ مدل گراف تصادفی فضایی

همانند دیگر مدلی که جزو دسته مدل‌ها بر مبنای ویژگی‌های راس قرار دارد، کار نخست تعبیه کردن راس در فضاست. در این مدل از یک فضای دو بعدی $\mathbb{R}^۲$ جهت تعبیه راس‌ها در آن استفاده می‌شود. همچنین یک تابع برای محاسبه فاصله بین دو راس در فضا را تعریف می‌کنیم: $d : \mathbb{R}^۲ \times \mathbb{R}^۲ \to \mathbb{R}$. تابع d در نامساوی مثلثی صدق می‌کند و همچنین برای تمام مقادیر در فضا، مقداری بزرگتر یا مساوی صفر را به عنوان خروجی به ما می‌دهد. برای هر راس در مدل یک مختصات $(x_i, y_i)^t$ در فضا بر اساس یک تابع توزیع خاص اختصاص داده می‌شود. بردار $\chi = (x_۱, y_۱)^t, (x_۲, y_۲)^t, \dots$ مختصات کل راس‌ها بر روی فضا می‌باشد. برای ساده‌سازی از این به بعد از $d_{ij} = d((x_i, y_i), (x_j, y_j))$ برای نشان دادن فاصله دو راس در فضا استفاده می‌شود. راس‌ها در مدل به صورت تصادفی در فضا پراکنده شده‌اند، به این معنا که این پراکندگی‌ها مستقل از هم می‌باشد. مکان یک راس به مکان راس‌های قبل و بعد وابسته نیست. در این مدل برای توزیع نقاط در فضا از فرایند توزیع پواسون یکنواخت استفاده می‌شود. فرایند توزیع پواسون نقاط با نرخ $\rho < \infty$ در یک فضای d-بعدی اقلیدسی $\mathbb{R}^d$، فرایندی است که:

- برای همه زیرمجموعه‌های مجزا $A_۱, A_۲, \dots, A_k \subset \mathbb{R}^d$، متغیر تصادفی که تعداد نقاط را در هر زیر مجموعه نشان می‌دهد، $N(A_۱), N(A_۲), \dots, N(A_k)$ دو به دو مستقل از هم می‌باشند.

- $N(A)$ دارای توزیع پواسون می‌باشد.

- $E[N(A)] = \rho|A|$ برای همه $A \subset \mathbb{R}^d$ می‌باشد.

همانند مدل BPDA در این مدل نیز احتمال وجود یال بین راس‌ها وابسته به فاصله آن‌ها از هم می‌باشد. برای یک توزیع خاص راس‌ها در فضا χ داریم: $P(x_{ij} = ۱|\chi) = f(d_{ij})$ که در آن $f : \mathbb{R} \to [۰, ۱]$ می‌باشد. با

توجه به آنچه اشاره شد تابع احتمال وجود یال بین دو راس را به صورت زیر تعریف می‌کنیم:

$$P(x_{ij}=۱|\chi)=\begin{cases} p+p_b & \text{اگر}\quad d_{ij}\leq H \\ p-\Delta & \text{اگر}\quad d_{ij}>H \end{cases} \tag{۶۹-۲}$$

که در این رابطه p میانگین تراکم شبکه، H شعاع همسایگی، p_b ضریب تاثیر همسایگی بر احتمال وجود یال می‌باشد؛ یعنی p_b تاثیر فاصله در رابطه را نشان می‌دهد. هرچه این مقدار بزرگ‌تر باشد با احتمال بیشتری راس‌های نزدیک‌تر با هم در ارتباط هستند. این ضریب برای همه راس‌ها یکسان در نظر گرفته می‌شود. علاوه بر این $\Delta=\Delta(p_b,H|\chi)$ به عنوان تصحیح کننده کل رابطه مورد استفاده قرار می‌گیرد. از Δ برای حفظ میانگین تراکم کل گراف p استفاده می‌شود. برای تراکم گراف رابطه زیر را داریم:

$$E\Big[\frac{۱}{N-۱}\sum_{i<j}x_{ij}\Big|\chi\Big]=p \tag{۷۰-۲}$$

بدون ثابت نگه داشتن تراکم، پیش‌بینی ویژگی‌های راس با توجه به تغییر تعداد یال‌ها در مدل و بررسی تاثیر p_b و H بر روی این ویژگی‌ها دشوار است.

برای بدست آوردن Δ، نخست باید تعداد یال‌های ممکن که طول کوتاه‌تر از شعاع H دارند را در شبکه χ محاسبه کنیم. تعداد یال‌های ممکن با طول کوتاه‌تر از H را با $S_{\leq H}(\chi)$ نشان می‌دهیم. زمانی که N مقدار بزرگی دارد و با چشم‌پوشی از تاثیراتی که نواحی مرزی روی مدل می‌گذارد داریم:

$$S_{\leq H}(\chi)=\frac{N\pi\rho H^۲}{۲} \tag{۷۱-۲}$$

با تنظیم ρ و H می‌توان یکی از این متغیرها را برحسب دیگری نوشت. برای مثال با قرار دادن $H=\sqrt{\rho}$ می‌توان به رابطه زیر رسید:

$$S_{\leq H}(\chi)=\frac{N\pi\rho^۲}{۲} \tag{۷۲-۲}$$

رابطه اخیر به عنوان پایه کاری برای بدست آوردن ادامه روابط مورد استفاده قرار می‌گیرد. حال با توجه به روابط ۲-۷۱ و ۲-۷۲ تعداد کل یال‌های ممکن گراف، که فاصله بین راس‌های دخیل در ارتباط بیش از H باشد برابر است با:

$$S_{>H}(\chi)=\binom{N}{۲}-S_{\leq H}(\chi) \tag{۷۳-۲}$$

با توجه به تعریف و رابطه ۲-۶۹ تراکم مورد انتظار در داخل همسایگی H برابر $p+p_b$ می‌باشد. لذا تعداد واقعی یال‌ها در داخل همسایگی H برابر $(p+p_b)S_{\leq H}$ است. حال برای نگه داشتن تراکم برابر مقدار مورد

انتظار باید تعداد یال‌های خارج از همسایگی H برابر $\binom{N}{2} - (p+p_b)S_{\leq H}$ باشد. لذا داریم:

$$p - \Delta = \frac{1}{S_{>H}}\Big[\binom{n}{2}p - (p+p_b)S_{\leq H}\Big] = p - \frac{S_{\leq H}}{\binom{N}{2} - S_{\leq H}}p_b \tag{۲-۷۴}$$

با توجه به اینکه احتمال $P(x_{ij}=1|\chi)$ توسط ۰ و ۱ از پایین و بالا کراندار شده است، لذا $0 \leq p+p_b \leq 1$ و $0 \leq \Delta \leq 1$ می‌باشد. این روابط کران بالا و پایینی را برای p_b با توجه به p و H مشخص بدست می‌دهد.

برای بیان روابط آماری مدل، در این بخش از شیوه نمایی که در [۶۱] معرفی شده استفاده می‌کنیم. پیش از تحلیل آماری مدل به معرفی روش نمایی برای تحلیل می‌پردازیم.

دنباله $Z = (Z_1, \ldots, Z_m)$ که شامل m متغیر تصادفی گسسته می‌باشد را در نظر می‌گیریم. فرض کنید R_i بازه فضای Z_i و R بازه فضای Z باشد ($|R| < \infty$). تابع احتمال Z را با $Pr(Z=z) = P(z)$ برای $z \in R$ نشان می‌دهیم. تابع احتمال P به صورت زیر تعریف می‌شود:

$$P(z) = c^{-1}\exp Q(z) \tag{۲-۷۵}$$

که در آن

$$c = \sum_{z\in R}\exp Q(z) \tag{۲-۷۶}$$

برای هر تابع حقیقی Q و R می‌باشد. حال از این روابط استفاده می‌کنیم تا مدل خود را تحلیل کنیم.

فرض کنیم x یک نمونه از گراف تولید شده توسط این مدل باشد که مختصات راس‌هایش با χ نشان داده می‌شود. آنگاه تابع احتمال اینکه الگوریتم مدل گراف x را تولید کند برابر است با:

$$P(X=x|\chi) = \frac{1}{Z(\chi)}\exp[-\mathcal{H}(x|\chi)] \tag{۲-۷۷}$$

که $\mathcal{H}(x|\chi)$ تابع همیلتونی گراف x است که در ادامه تشریح خواهد شد. در این رابطه برای تابع $Z(\chi)$ داریم:

$$Z(\chi) = \sum_{x}\exp[-\mathcal{H}(x|\chi)] \tag{۲-۷۸}$$

تابع همیلتون می‌تواند هر تابعی باشد که وابستگی‌های تولید یال‌های گراف را بازتاب دهد. ساده‌ترین فرم برای تابع همیلتون $-\mathcal{H}(x) = \theta_0 L(x|\chi)$ است که در آن $L(x|\chi) = L(x)$ تعداد یال‌های گراف x که مستقل از χ هستند، می‌باشد و θ_0 پارامتر همراه است. با تنظیم θ_0 می‌توانیم تراکم مورد انتظار در یک گراف را تغییر دهیم. برای تعریف $\mathcal{H}$ نخست $L_{\leq}(x|\chi) = \sum_{i<j,d_{ij}<H} x_{ij}$ و $L_{>}(x|\chi) = \sum_{i<j,d_{ij}>H} x_{ij}$ که به ترتیب نشان دهنده یال‌های کوتاه‌تر و بلندتر از H است، را تعریف می‌نماییم. به این ترتیب برای تابع همیلتون داریم:

$$-\mathcal{H}(x|\chi) = \theta_{\leq}L_{\leq}(x|\chi) + \theta_{>}L_{>}(x|\chi) \tag{۲-۷۹}$$

که در این رابطه $\theta_{\leq}$ و $\theta_{>}$ را می‌توان از رابطه ۲-۶۹ بدست آورد:

$$\theta_{\leq} = logit(p + p_b) \qquad \theta_{>} = logit(p - \Delta) \tag{۲-۸۰}$$

که در آن $logit\ q = log[q/(۱ - q)]$.

با استفاده از رابطه همیلتون می‌توانیم $Z(\chi)$ را هم بدست آورد:

$$\begin{aligned} Z_N(\chi) &= \sum_x \exp[\mathcal{H}(x)] = \sum_x \exp(\theta_{\leq} L_{\leq}(x) + \theta_{>} L_{>}(x)) \\ &= \sum_x \exp\left(\theta_{\leq} \sum_{i<j, d_{ij} \leq H} x_{ij} + \theta_{>} \sum_{i<j, d_{ij} > H} x_{ij}\right) \\ &= (۱ + e^{\theta_{\leq}})^{S_{\leq}(\chi)} + (۱ + e^{\theta_{>}})^{\binom{N}{۲} - S_{\leq}(\chi)} \end{aligned} \tag{۲-۸۱}$$

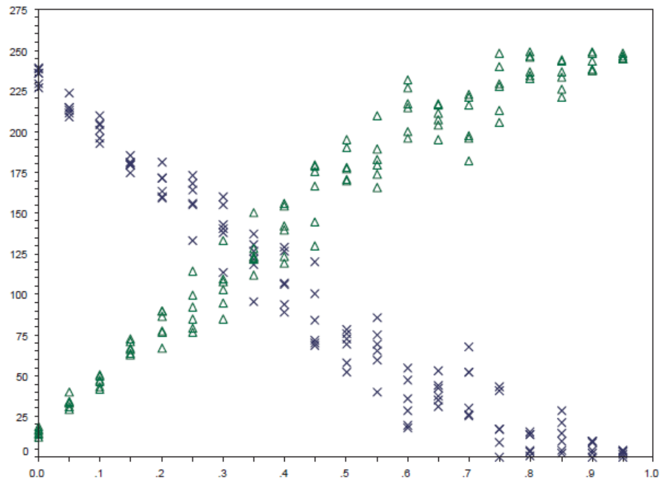
به صورت نظری می‌توانیم هر برآورد آماری $\langle Q(x|\chi)\rangle$ را با استفاده از این روابط (از جمله ضریب خوشه‌بندی) بدست آوریم. بدین منظور یک عنصر کمکی را به تابع همیلتون می‌افزاییم $\Delta\mathcal{H}(x) = yQ(x|\chi)$. لذا داریم:

$$\begin{aligned} \langle Q(x|\chi)\rangle &= \frac{۱}{Z(\chi)} \sum_x Q \exp[-\mathcal{H}|\chi - yQ(x|\chi) \\ &= \frac{۱}{\chi} \left.\frac{Z(\chi)}{\partial y}\right|_{y=۰} \end{aligned} \tag{۲-۸۲}$$

همانطور که ونگ و همکاران نیز اشاره کردند، بدست آوردن ضریب‌خوشه بندی با استفاده از این روابط بسیار دشوار است. ونگ و همکاران به خاطر همین دشواری به جای تحلیل آماری ویژگی‌های گراف به پیاده‌سازی و بررسی داده‌های عددی پرداخته‌اند.

۲.۶.۲ پیاده‌سازی و بررسی داده‌های عددی

در پیاده‌سازی مدل به خاطر تشابه و ارتباط ρ و H می‌توان یکی از این متغیرها را حذف کرد. بدین منظور در پیاده‌سازی $\rho = ۱$ در نظر گرفته خواهد شد و با تغییر H، تغییر ویژگی‌های گراف پیگیری می‌شود. زمانی که H به مقدار کافی بزرگ باشد، انتصاب مقداری خاص به آن مسئله بحرانی نیست. به‌علاوه زمانی که H بسیار کوچک باشد، اکثریت یال‌ها دارای طولی بیشتر از شعاع H هستند، در این حالت مدل به مدل گراف تصادفی اردوش و رنیی [۴۲] تبدیل می‌شود. در پیاده‌سازی مقدار $H = ۳/۲$ در نظر گرفته می‌شود. هدف این بخش بررسی تغییرات ویژگی‌های مدل با تغییر p_b می‌باشد.



شکل ۲۳.۲: نمودار تغییر تعداد یال‌های کوتاه و بلند با تغییر p_b. ($\times$) تعداد یال‌های بلند، ($\triangle$) تعداد یال‌های کوتاه

تعداد یال‌های کوتاه و بلند

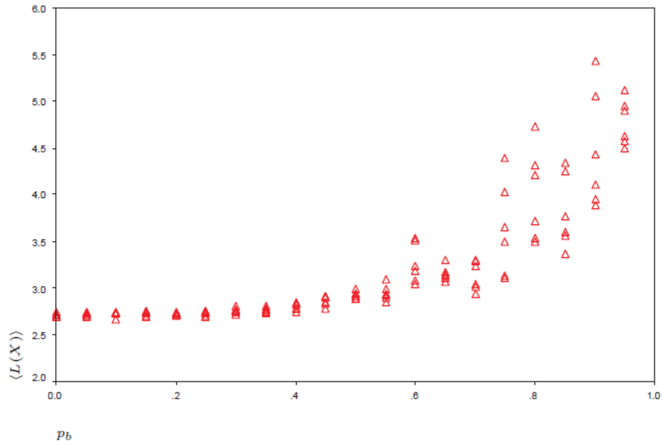
شکل ۲۳.۲ نمودار میانگین تعداد یال‌های کوتاه‌تر $\langle L_{\leq}(x)\rangle$ و بزرگتر $\langle L_{>}(x)\rangle$ از H را نشان می‌دهد، که آن را به ترتیب یال‌های کوتاه و یال‌های بلند می‌نامیم. زمانی که $p_b = ۰$ تمام یال‌ها، یال‌های بلند می‌باشند با افزایش p_b همانطور که مشاهده می‌شود یال‌های بلند رو به کاهش گذاشته و تعداد یال‌های کوتاه افزایش می‌یابد.

ویژگی جهان کوچک

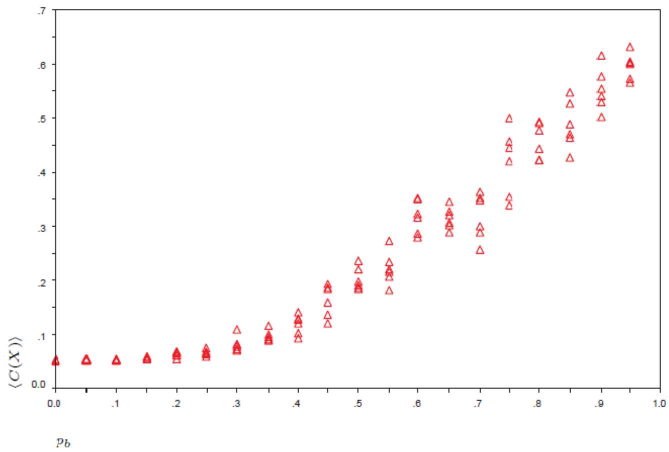
شکل ۲۴.۲ نمودار تغییر میانگین طول کوتاه‌ترین مسیر را نسبت به تغییرات p_b نشان می‌دهد. از شکل می‌توان دریافت که برای بازه بزرگی از p_b میانگین طول کوتاه‌ترین مسیر مقدار کوچکی باقی می‌ماند. اما بعد از یک نقطه بحرانی طول میانگین کوتاه‌ترین مسیر ناگهان افزایش می‌یابد. این نقطه بحرانی نشان دهنده تغییر از یک گراف تصادفی ساده به یک گراف شبکه‌های اجتماعی می‌باشد.

شکل ۲۵.۲ نشان دهنده تغییرات ضریب خوشه‌بندی گراف نسبت به p_b است. همانطور که مشاهده می‌شود مقدار ضریب خوشه‌بندی کاملا وابسته به مقدار p_b می‌باشد. این افزایش ضریب خوشه‌بندی به این دلیل مشاهده می‌شود که مثلث‌ها اصولا بین راس‌های نزدیک به هم تشکیل می‌شوند.

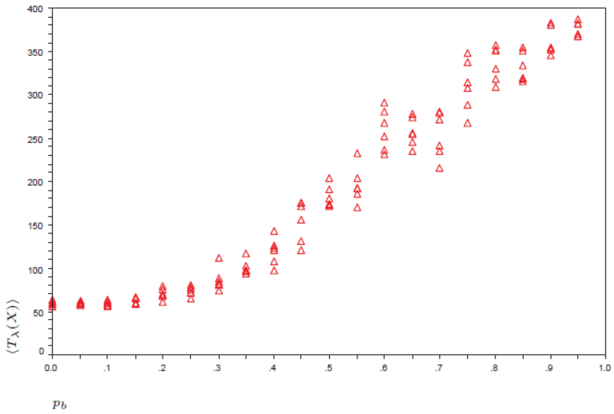
همانطور که مشاهده می‌شود برای یک بازه از p_b گراف هم دارای میانگین طول کوتاه‌ترین مسیر کوچک و هم دارای ضریب خوشه‌بندی بزرگ می‌باشد. پس در این بازه گراف ویژگی جهان کوچک را از خود بروز می‌دهد.



شکل ۲.۲۴: نمودار تغییر میانگین طول کوتاه‌ترین مسیر به p_b



شکل ۲.۲۵: نمودار تغییر ضریب خوشه‌بندی با تغییر p_b مدل WPR



شکل ۲۶.۲: نمودار تغییر $\langle T_\lambda(x)\rangle$ با تغییر p_b مدل WPR

انجمن

در این مدل، نویسندگان برای نشان دادن انجمن‌ها از k-دسته استفاده می‌کنند. تعداد کل k-دسته‌ها از رابطه زیر بدست می‌آید:

$$T_\lambda(G) = 3t_1(G) - \frac{t_2(G)}{\lambda} + \frac{t_3(G)}{\lambda^2} - \ldots + (-1)^{(n-3)}\frac{t_{n-2}}{\lambda^{n-3}} \qquad (۸۳-۲)$$

$$= \lambda \sum_{i<j}\left[1 - \left(1 - \frac{1}{\lambda}\right)^{g_{ij}(G)}\right] \qquad (۸۴-۲)$$

که در این رابطه $g_{ij}(G)$ تعداد مسیرهای دو طرفه بین v_i و v_j و λ مقداری ثابت می‌باشد. شکل ۲۶.۲ نشان دهنده تغییرات T_2 نسبت به p_b می‌باشد. همانطور که انتظار می‌رود با افزایش p_b انجمن‌های تشکیل شده بزرگ‌تر و متراکم‌تر می‌شوند. این موضع به دلیل خاصیت p_b می‌باشد که با افزایش آن یال‌ها در یک محدوده خاص بین یال‌های خاص متمرکز شده و باعث پدید آمدن انجمن‌ها می‌شوند.

سایر ویژگی‌های این مدل در فصل بعد بیشتر بررسی خواهد شد. اما این مدل با استفاده از یک فضای دو بعدی سعی داشت علاوه بر ویژگی‌های رفتاری ویژگی‌های مبنایی را در تشکیل یال‌ها دخیل کند و بدین وسیله به مدلی واقعگرایانه‌تر از شبکه‌های اجتماعی حقیقی دست پیدا کند.

۷.۲ مدلی مبتنی بر گراف وزن‌دار

همه مدل‌هایی که تاکنون معرفی شده‌اند، گراف‌هایی تولید می‌کردند که یال‌هایشان دارای وزن نبود. اما در برخی شبکه‌های اجتماعی و شبکه‌های پیچیده علاوه بر اهمیت حضور یال بین دو راس، وزن یال نیز دارای اهمیت می‌باشد. برای مثال شبکه‌ی اجتماعی در نظر بگیرید که بین افراد و به وسیله ارتباطات تلفنی ایجاد شده است. با توجه به مسافت بین دو فرد هزینه تماس تغییر می‌کند، لذا گرافی که این شبکه ایجاد می‌کند یک گراف وزن‌دار می‌باشد. در بسیار از موارد وزن‌دار بودن یال بر روی ویژگی‌های ساختاری گراف تاثیر می‌گذارد. طبیعی است انتظار داشته باشیم حضور یال‌های وزن‌دار بر روی انجمن‌های تشکیلی نیز اثر بگذارد.

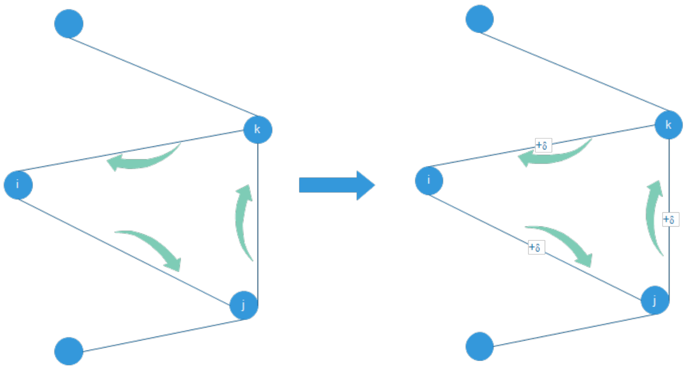
بعضی مدل‌ها که از ویژگی‌های وزن-ساختار در الگوریتم استفاده می‌کنند، برای برخی شبکه‌های پیچیده دیگر نظیر مدلسازی شبکه حمل و نقل [۶۳] ارائه شده‌است. اما این مدل‌ها یکی از مهمترین ساختارهای موجود در شبکه‌های اجتماعی یعنی انجمن‌ها را در گراف تولیدی نهفته ندارند. مدلی که در این بخش معرفی می‌شود و توسط کومپولا [۱۱] و همکاران [۶۴] ارائه شده و با نام KOSKK شناخته می‌شود، سعی دارد با دخیل کردن وزن یال در مدل و استفاده از آن گراف‌هایی تولید کند که دارای انجمن باشند.

گرانوتر [۱۲] و همکاران [۶۵] نشان دادند که شبکه‌های اجتماعی در مقیاس بزرگ، در فرضیه اتصالات ضعیف صدق می‌کنند. یعنی لینک‌های ضعیف‌تر گراف را متصل نگه می‌دارند در حالی که یال‌های قوی تشکیل دهنده انجمن‌ها هستند. همانند دیگر مدل‌ها مدل KOSKK سعی در شبیه سازی دو نوع جستجو و ایجاد ارتباط در شبکه‌های اجتماعی را دارد. کومپولا و همکاران دو نوع بستار را دلیل ایجاد ارتباط بین افراد در شبکه اجتماعی می‌داند:

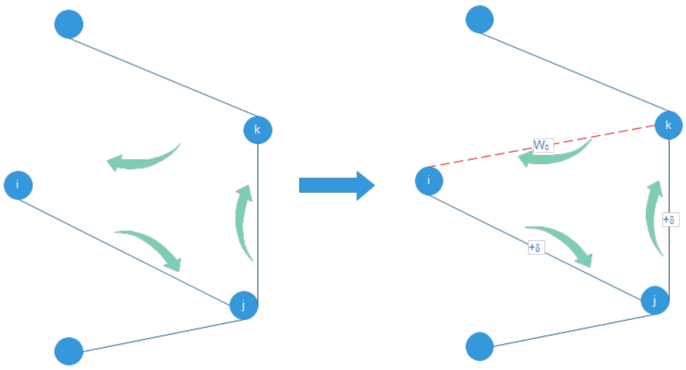
- بستار چرخه‌ای
- بستار کانونی

بستار چرخه‌ای عامل ایجاد ارتباطات بین دوستِ دوست می‌باشد. احتمال این نوع بستار وابسته به فاصله بین عوامل درگیر هست. هرچه دو فرد از هم دور باشند احتمال ایجاد ارتباطی توسط بستار چرخه‌ای کم می‌شود. این نکته با توجه به آنچه پیشتر گفته شده بدیهی می‌نماید. بستار کانونی عامل ایجاد ارتباطاتی می‌شود که مستقل از فاصله بین افراد درگیر است. به خاطر وجود برخی ویژگی‌های مشترک بین دو فرد به وجود می‌آید که این دو نوع ایجاد ارتباط پایه‌های جستجوهای مدل KOSKK برای ایجاد ارتباط را تشکیل می‌دهند.

[۱۱] Kumpula
[۱۲] Granovetter



شکل ۲۷.۲: جستجوی محلی بدون نیاز به افزودن یال مدل KOSKK



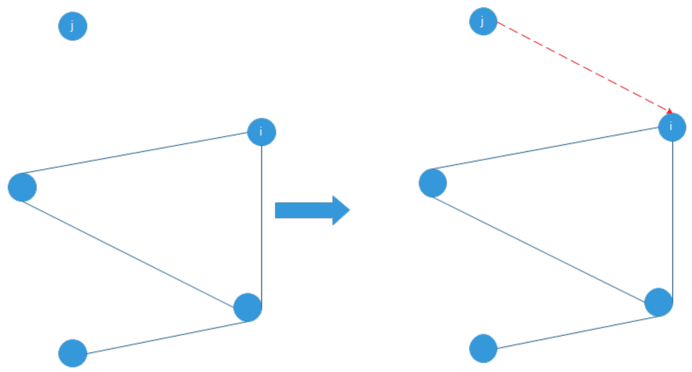
شکل ۲۸.۲: جستجوی محلی با یال افزایی مدل KOSKK

۱.۷.۲ الگوریتم مدل KOSKK

الگوریتم ارائه شده یال‌ها را براساس وزن یال‌های دیگر منتهی به یک راس، اضافه می‌کند، و هر بار با انتخاب یک یال برای ادامه جستجو یال را قوی‌تر (یعنی وزن آن را بیشتر) می‌کند. الگوریتم مدل KOSKK به شرح زیر است:

فرض می‌کنیم گراف از تعداد ثابتی راس N تشکیل شده باشد. یال‌ها در این مدل به دو طریق ایجاد می‌شوند:

۱. در هر بازه زمانی ΔT هر راسی که حداقل یک همسایه دارد، جستجوی محلی وزن‌دار خود را برای یافتن یک همسایه جدید آغاز می‌کند (شکل۲۷.۲ و ۲۸.۲). راس i یکی از همسایگان خود j را با احتمال w_{ij}/s_i انتخاب می‌کند، که در این احتمال w_{ij} وزن یال بین i، j و $s_i = \sum_j w_{ij}$ وزن راس i، که

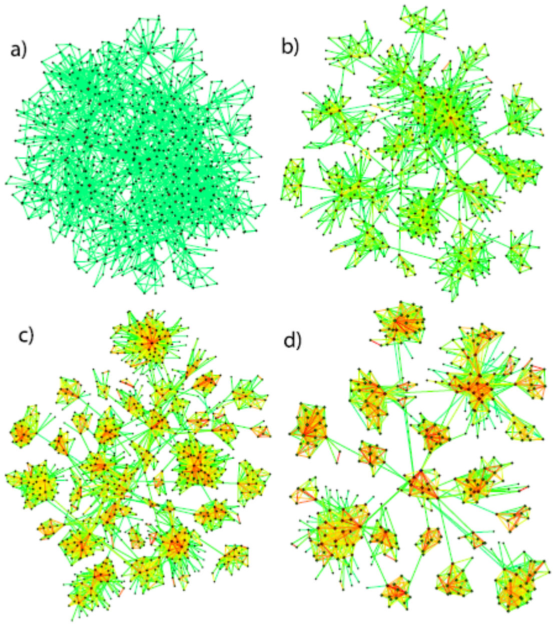


شکل ۲۹.۲: افزودن یال بین یک راس تنها و یکی از راس‌های گراف که به تصادف انتخاب شده است مدل KOSKK

مجموع وزن همه یال‌های حادث به i می‌باشد. اگر راس j همسایه‌ای به جز i داشته باشد یکی از آن‌ها (k) را با احتمال $w_{jk}/(s_j - w_{ij})$ انتخاب می‌کند. بنابراین جستجو علاقه‌مند به یال‌های قوی می‌باشد. اگر یالی بین i و k وجود نداشته باشد (شکل ۲۸.۲) یک یال با احتمال $p_\Delta \Delta T$ بین i و k افزوده می‌شود. وزن یال برابر یک مقدار اولیه W_0 قرار می‌گیرد. اگر یالی بین i و k موجود باشد (شکل ۲۷.۲) وزن آن به مقدار δ افزایش می‌یابد. علاوه‌براین در طی فرایند، هر بار یالی برای ادامه فرآیند انتخاب شد وزن‌اش به اندازه δ افزایش می‌یابد. به این جستجو، جستجوی محلی LA می‌گوییم و معادل بستار چرخه‌ای است که پیش از این شرح دادیم.

۲. نوع دوم جستجو، برای حالتی است که یک راس هیچ یالی ندارد. تحت این شرایط یک یال بین راس تنها و یکی از راس‌های دیگر که به تصادف انتخاب شده‌است ایجاد می‌شود و وزن یال برابر مقدار اولیه قرار می‌گیرد. در غیر این صورت با احتمال $p_r \Delta t$ یک یال بین آن و یک راس به تصادف انتخاب شده ایجاد می‌شود (شکل ۲۹.۲). وزن یال جدید ایجاد شده w_0 قرار داده می‌شود. این نوع جستجوی با نام جستجوی سراسری (GA) شناخته می‌شود و بستار کانونی را شبیه‌سازی می‌کند.

نهایتا هر راس در هر گام زمانی با احتمال $p_d \Delta T$ به همراه تمام یال‌های حادث حذف می‌شود (ND) و یک راس جدید تنها به جای آن به گراف افزوده می‌شود تا تعداد کل راس‌ها ثابت بماند. حذف راس تنها مکانیزم حذف یال در گراف می‌باشد و باعث می‌شود بازه زمانی زندگی راس‌ها و یال‌ها دارای یک توزیع توانی باشد. به طوری که برای میانگین دوره زندگی برای راس $\langle \tau \rangle = p_d^{-1}$ و برای یال $\langle \tau_w \rangle = (2p_d)^{-1}$ می‌باشد.

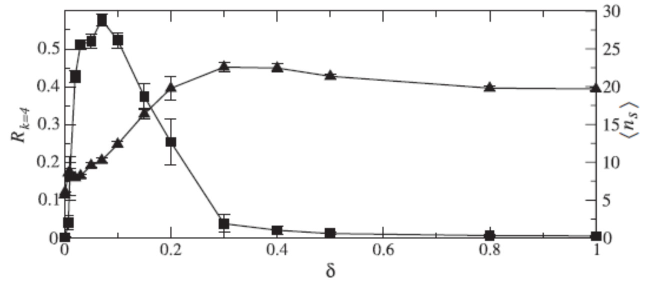


شکل ۲.۳۰: به وجود آمدن انجمن‌ها با تغییر اندازه δ (a) $\delta = ۰$، (b) $\delta = ۰/۱$ (c) $\delta = ۰/۵$ و (d) $\delta = ۱$ [۶۴].

۲.۷.۲ پیاده‌سازی مدل KOSKK

مدل با N عدد راس تنها شروع شده و به صورت پی در پی با استفاده از LA و GA یال‌هایی به گراف افزوده و توسط ND راس‌هایی از آن را حذف می‌کند. در طول فرایند تعداد راس‌ها ثابت می‌ماند. می‌توانیم برای راحتی کار و بدون از دست دادن جامعیت Δ و $w_۰$ را برابر ۱ قرار دهیم. نرخ حذف راس‌ها را $p_d = ۱۰^{-۳}$ و $p_r = ۵ \times ۱۰^{-۴}$ قرار می‌دهیم، بدین طریق طول دوره زندگی یک راس $\langle \tau \rangle = ۱۰^۳$ گام زمانی می‌باشد و با توجه به p_r در طول دوره زندگی، به طور میانگین یک بار به وسیله جستجوی محلی، یالی به آن اضافه می‌شود. با این شرایط مدل پس از مدتی به پایداری می‌رسد.

گراف وزن‌دار با کمک پارامتر δ به وجود می‌آید، برای در نظر گرفتن تاثیر این پارامتر در به وجود آمدن انجمن‌ها، میانگین درجه گراف را ثابت در نظر می‌گیریم $\langle k \rangle = ۱۰$. بدین طریق با توجه به ثابت بودن تعداد راس‌ها در گراف تعداد یال‌ها برای δهای مختلف تقریبا ثابت است. ثابت نگه داشتن میانگین درجات با استفاده از تنظیم دو مقدار P_Δ و δ انجام می‌گیرد. در حالتی که $\delta = ۰$ است، گراف تبدیل به گراف تصادفی اردوش و رنیی می‌شود که پیش از این شرح داده شد (گراف (a) از شکل ۲.۳۰). با افزایش δ همانطور که در گراف‌های b، c و d در شکل ۲.۳۰ نشان داده می‌شود مدل شروع به ساختن انجمن‌ها می‌کند و هرچه δ افزایش پیدا می‌کند، تراکم یال‌های درون انجمن بیشتر می‌شود و تعداد یال‌های بین انجمن کمتر و انجمن‌ها مجزا تر می‌شود. عامل



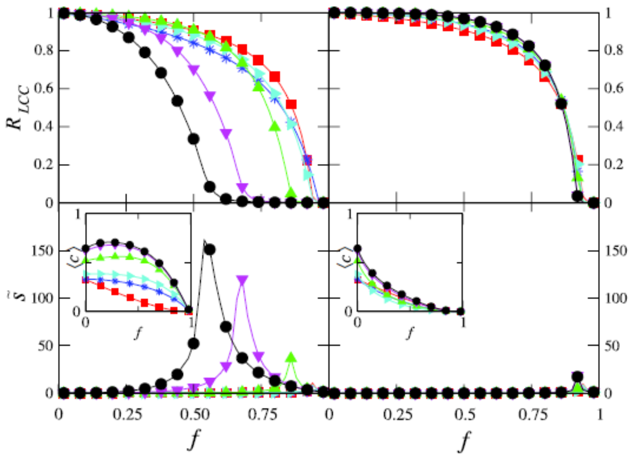
شکل ۳۱.۲: $R_{k=۴}$ ($\square$) و $\langle n_s \rangle$ ($\triangle$) برای $N = ۵ \times ۱۰^۴$ و δهای مختلف [۶۴].

این تغییر در ساختار گراف گیر افتادن جستجوی محلی در تله‌ی راس‌های قوی‌تر است. هرچه یک راس وزن بیشتری داشته باشد بیشتر و بیشتر به عنوان انتخاب یک جستجوی محلی مورد پیمایش قرار گرفته و قوی‌تر می‌شود. این باعث ایجاد یال‌هایی حول آن راس (تبدیل شدن به هسته یا قطب) و در نتیجه ایجاد انجمن‌ها می‌شود.

همانند بخش پیش در این بخش نیز از الگوریتم پالا [۱۹] برای یافتن انجمن‌ها استفاده می‌شود. چون LA منبع اصلی تولید یال‌ها در هر مرحله یک مثلث ایجاد می‌کند، لذا از الگوریتم پالا با ۴-دسته استفاده می‌شود. شکل ۳۱.۲ تغییرات بزرگترین انجمن تولید شده توسط گراف ($R_{k=۴}$) و میانگین اندازه بقیه انجمن‌ها (به جز بزرگ‌ترین انجمن) یعنی $\langle n_s \rangle$ نسبت به δ در بازه $\delta \in [۰, ۱]$ را نشان می‌دهد. زمانی که $\delta = ۰$ و انجمن‌ها بسیار کوچک باشد، $\langle n_s \rangle \sim ۶$ که بزرگترین انجمن دارای تعداد حدودا ۵۰ راس می‌باشد. با افزایش δ ساختار انجمن‌ها تغییر می‌کند. با تغییر δ نخست شبکه به گرافی دارای ساختار یکنواخت تبدیل می‌شود، اما با رشد بیشتر δ انجمن‌هایی با تراکم یال زیاد (در درون انجمن) ایجاد می‌شود. به وجود آمدن و قوی‌تر شدن انجمن‌ها را در افزایش $\langle n_s \rangle$ و کاهش $R_{k=f}$ می‌توان مشاهده کرد. این دو مقدار با افزایش δ تا یک مقدار مشخص تغییر (افزایش و کاهش) پیدا می‌کنند و پس از رسیدن به یک مقدار مشخص از δ، دیگر افزایش مقدار δ تاثیری روی این دو کمیت نخواهد گذاشت.

زمانی که δ مقدار بزرگی را داراست، جستجوی محلی اغلب در درون یک انجمن حرکت و راس انتخاب می‌کند و به این طریق می‌توان گفت که ادغام انجمن‌ها بعد از پایدار شدن مدل به ندرت اتفاق می‌افتد. لذا می‌توانیم رابطه زیر را برای تغییرات اندازه انجمن S در طول زمان بدست آوریم:

$$\frac{dN_s}{dt} = -p_d N_s + p_d N \frac{N_s}{N} = ۰ \qquad (۲\text{-}۸۵)$$

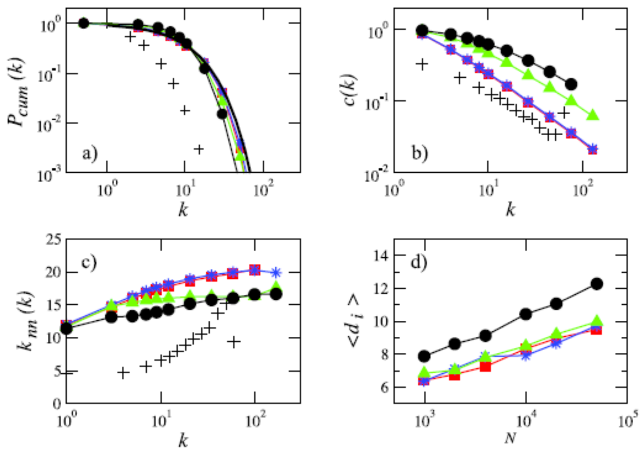


شکل ۳۲.۲: تغییرات R_{LCC} و $\tilde{s}$ نسبت به حذف یال‌ها نمودارهای سمت چپ نشان دهنده حذف یال به صورت صعودی و نمودارهای سمت راست نشان دهنده حذف یال به صورت نزولی می‌باشد. در این نمودار (□):۰ ، ($\star$):$۱۰^{-۳}$، ($\triangleright$): $۱۰^{-۲}$، ($\triangle$): ۰.۱، ($\triangledown$):۰.۵ و (●):۱ [۶۴].

که در این رابطه عنصر اول در سمت راست مربوط به حذف راس‌ها در هر بازه زمانی از بین راس‌های داخل می‌شود. عنصر دوم، اضافه شدن یال به گراف (در پی حذف) و اتصال آن به یکی از راس‌های انجمن را نشان می‌دهد. مورد اخیر از این نکته حاصل می‌شود که پس از اتصال یک راس به یک انجمن با توجه به بزرگ بودن δ جستجوی محلی داخل انجمن در تله‌ای گیر می‌افتد و به ندرت از آن خارج می‌شود. همانطور که در شکل ۳۰.۲ و شکل ۳۱.۲ نشان داده شد، بعد از $\delta = ۰/۵$ با افزایش δ تغییر خاصی در ساختار گراف رخ نمی‌دهد.

حال با توجه به فرضیه یال‌های ضعیف که پیش از این عنوان شد، می‌خواهیم تاثیر حذف یال‌ها را به ترتیب صعودی (از یال با وزن کم به یال با وزن بیشتر) و نزولی (از یال با وزن بیشتر به یال با وزن کمتر) را بر روی بزرگترین مولفه و کمیت حساسیت نرمال $\tilde{s} = \sum n_s s^۲/N$ (n_s تعداد مولفه‌های با اندازه s) بررسی کنیم. نمودار شکل ۳۲.۲ تغییرات این دو کمیت به اضافه ضریب خوشه‌بندی را نسبت به حذف یال صعودی (نمودارهای سمت چپ) و نزولی (نمودارهای سمت راست) نشان می‌دهد. برای مقادیر کوچک δ تقریبا هیچ انجمنی وجود ندارد، و مولفه عظیم با حذف پی در پی یال‌ها در یک نقطه (در هر دو ترتیب حذف) فرو می‌ریزد. در نقطه فروریزی هنوز $\tilde{s}$ مقدار کوچکی را داراست. هر چقدر $\delta \leq ۰/۱$ باشد فروریزی با سرعت بیشتری روی می‌دهد. این نکته تایید کننده این موضوع می‌باشد که یال‌های ضعیف نقش پل را در گراف بازی می‌کنند. تغییرات ضریب خوشه بندی نیز حکایت از این دارد که با حذف نزولی ضریب خوشه‌بندی گراف با سرعت بیشتری (نسبت به ترتیب صعودی) افت می‌کند و کوچکتر می‌شود.

در شکل ۳۳.۲ ویژگی‌هایی چون توزیع درجات (a)، ضریب خوشه بندی (c)، میانگین درجات همسایگان (c)



شکل ۳۳.۲: نمودار (a):توزیع درجات گراف، (b): ضریب خوشه‌بندی، (c): میانگین درجات همسایه‌ها و (d): قطر گراف. برای گرافی با اندازه $N = ۱۰^۴$ و میانگین درجه $\langle k \rangle = ۳/۱۴$ که علائم همانند علائم شکل ۳۲.۲ می‌باشد [۶۴].

و قطر گراف (d) را برای گراف تولید شده توسط مدل مشاهده می‌کنیم. همانطور که مشاهده می‌شود توزیع درجات دارای توزیع توانی است که از مشخصه‌های اصلی شبکه‌های اجتماعی می‌باشد. به علاوه همانند شبکه‌های اجتماعی ضریب خوشه‌بندی با درجه راس رابطه عکس دارد. همچنین مدل همبستگی درجه به درجه با اشتراک پذیری مثبت از خود بروز می‌دهد. به‌علاوه قطر گراف نیز با رشد اندازه گراف سیر صعودی از خود نشان می‌دهد.

فصل ۳

مقایسه مدل‌ها با داده‌های واقعی

۱.۳ مقدمه

از سال ۲۰۰۹ که تویونن و همکاران [۵۵] به مقایسه شبکه‌های اجتماعی واقعی و مدل‌های ارائه شده برای آن پرداختند، هم رویکرد مردم به این شبکه‌ها و هم ساختار این شبکه‌ها با تغییر جدی مواجه شده‌است. داده‌هایی که تویونن و همکاران به عنوان داده‌های واقعی مورد بررسی قرار دادند تراکم بسیار کمی داشتند. آن‌ها دو شبکه اجتماعی وب سایت last.fm و شبکه اجتماعی ایمیل را مورد بررسی قرار دادند. اطلاعات این شبکه‌ها در جدول ۳-۱ موجود می‌باشد. گراف‌هایی که توسط تویونن و همکاران مورد بررسی قرار گرفتند، گراف‌هایی با میانگین درجه کوچک و همچنین ضریب خوشه بندی کمتر از ۰.۳۱ بودند.

اما در این پژوهش ما به مقایسه مدل‌های ارائه شده برای شبکه‌های اجتماعی با شبکه‌های اجتماعی فیس

جدول ۳-۱: مشخصات شبکه‌های اجتماعی وب سایت last.fm و شبکه اجتماعی ایمیل که در [۵۵] مورد بررسی قرار گرفته‌اند. به اضافه شبکه اجتماعی فیس بوک که در این فصل مورد بررسی قرار خواهد گرفت.

نام گراف	تعداد راس‌ها	تعداد یال‌ها	میانگین درجه	ضریب خوشه بندی	تعداد مولفه همبندی
lastfm	۸۰۰۳	۱۶۸۲۴	۴.۲۰	۰.۳۱	۱
email	۱۱۳۳	۵۴۵۱	۹.۶۲	۰.۲۲	۱
facebook	۴۰۳۹	۸۸۲۳۴	۴۳.۶۹۱	۰.۶۰۵	۱
wiki-Vote	۷۱۱۵	۱۰۳۶۸۹	۲۹.۱	۰.۱۴	۲۴

بوک و ویکی‌وُت می‌پردازیم. دلیل انتخاب این شبکه‌های اجتماعی از بین شبکه‌های اجتماعی دیگر در ادامه تشریح خواهد شد. اما پیش از آن باید به تفاوت‌هایی که شبکه اجتماعی منتخب ما با شبکه‌های اجتماعی انتخاب شده توسط تویونن و همکاران وجود دارد، اشاره کرد. همانطور که در جدول ۳-۱ مشاهده می‌کنید گراف شبکه اجتماعی فیس بوک تراکم بسیار بیشتری نسبت به شبکه‌های اجتماعی بررسی شده در [۵۵]، دارد. میانگین درجه گراف شبکه اجتماعی فیس بوک ۴۳.۶۹۱ می‌باشد، در حالی که میانگین درجه گراف شبکه اجتماعی ایمیل، ۹.۶۲ و میانگین درجه گراف شبکه اجتماعی last.fm، تنها ۴.۲۰ می‌باشد. این تراکم باعث آشکار شدن برخی از معایب مدل‌های ارائه شده می‌شود. اما نکته بسیار مهم ضریب خوشه‌بندی گراف‌های مربوط به شبکه‌های اجتماعی است. همانطور که پیش از این نیز اشاره شد شبکه‌های اجتماعی ضریب خوشه بندی بالایی دارند. شبکه‌های اجتماعی last.fm و ایمیل به ترتیب ضریب خوشه‌بندی ۰.۳۱ و ۰.۲۲ دارند. اما

شبکه اجتماعی فیس بوک، ضریب خوشه‌بندی بسیار بالاتر یعنی ۰.۶۰ را داراست. این ضریب خوشه‌بندی بالا یکی از عیب‌های بزرگ مدل‌های تکاملی را آشکار کرد، و آن اینکه این شبکه‌ها توانایی تولید گرافی با چنین ضریب خوشه‌بندی بزرگی را ندارند.

شبکه‌های اجتماعی زیادی به صورت آنلاین به مخاطبانشان سرویس ارائه می‌دهند. اما هیچ کدام از این شبکه‌ها در سطح عمومی به اندازه فیس بوک محبوب نیستند. طی نظر سنجی که موسسه PEW [۱] انجام داده ۷۳٪ از افراد مورد پرسش از شبکه‌های اجتماعی استفاده می‌کردند. که شبکه اجتماعی فیس بوک مورد استفاده ۷۱٪ از پرسش شدگان قرار می‌گرفت. این یکه تازی شبکه اجتماعی فیس بوک یکی از اصلی‌ترین دلایل انتخاب آن به عنوان شبکه اجتماعی واقعی در این پژوهش بود. از دلایل دیگر می‌توان به همگانی و عمومی بودن آن اشاره کرد. دلایل ذکر شده فیس بوک را به موردی مطلوب برای انجام پژوهش‌های این چنینی تبدیل می‌کند.

در کنار این عمومیت و ضریب خوشه‌بندی بالا به منظور بررسی توانایی مدل‌ها در تولید ضریب خوشه‌بندی کوچک، داده واقعی ویکی وُت را مورد بررسی و تحلیل قرار دادیم. این مجموعه داده نسبت به داده‌های مورد بررسی در پژوهش تویونن و همکاران ضریب خوشه‌بندی کوچک و میانگین درجه بزرگ‌تری را داراست.

در این فصل نخست به تحلیل و بررسی شبکه اجتماعی فیس بوک می‌پردازیم و سپس مدل‌های ارائه شده را بررسی و با داده‌های مربوط به شبکه اجتماعی مقایسه می‌کنیم، و در نهایت انجمن‌های تولید شده توسط گراف‌های حاصل از مدل‌ها را با هم مقایسه خواهیم کرد. سپس تمام فرایند فوق را برای مجموعه داده ویکی وُت و مدل‌ها تکرار می‌کنیم.

برای پیاده‌سازی مدل‌ها از نرم افزار شبیه‌سازی نت لوگو [۲] استفاده کرده‌ایم. داده‌های تولید شده توسط نت لوگو را با استفاده از نرم افزار Cytoscape تحلیل کردیم. و از نرم افزار Cfinder برای یافتن انجمن‌ها استفاده نمودیم.

۲.۳ تحلیل گراف مربوط به شبکه اجتماعی فیس بوک

همانطور که اشاره شد گراف مربوط به شبکه اجتماعی فیس بوک [۳] دارای تراکم یال بسیار زیادی می‌باشد. اطلاعات عمومی مربوط به شبکه اجتماعی فیس بوک را در جدول ۳-۲ مشاهده می‌کنید. همانگونه که انتظار می‌رود این گراف دارای ضریب خوشه بندی بزرگ و میانگین کوتاه‌ترین مسیر کوچکی می‌باشد.

[۱] PEW.com

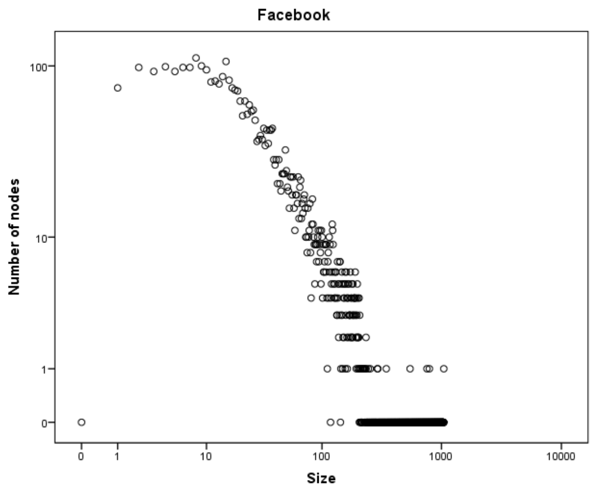
[۲] NetLogo

[۳] این داده‌ها از وب‌سایت /http://snap.stanford.edu/data بدست آمده است.

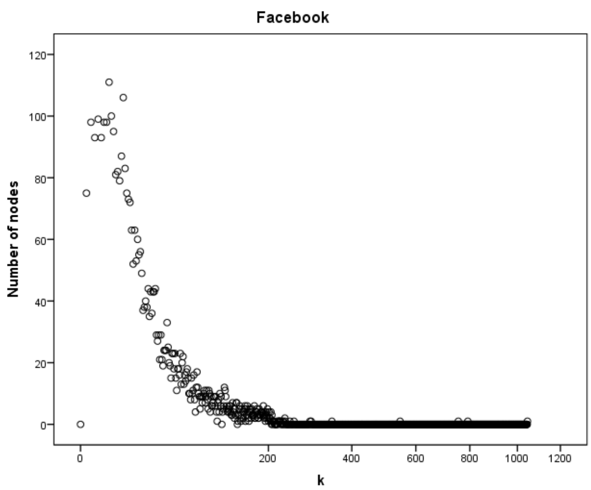
جدول ۳-۲: اطلاعات مربوط به شبکه اجتماعی فیس بوک

نام گراف	تعداد راس‌ها	تعداد یال‌ها	میانگین درجه گراف	ضریب خوشه بندی	تعداد مولفه همبندی	شعاع	قطر	میانگین کوتاه‌ترین مسیر
facebook.com	۴۰۳۹	۸۸۲۳۴	۴۳.۶۹۱	۰.۶۰۵	۱	۵	۸	۳.۶۹۳

در شکل ۱.۳ نمودار ۱.۳(آ) نمودار لگاریتمی توزیع درجات و شکل ۱.۳(ب) نمودار توانی توزیع درجات گراف فیس بوک را نشان می‌دهد. همانطور که انتظار می‌رفت، شبکه اجتماعی فیس بوک دارای توزیع درجات توانی می‌باشد. از ویژگی‌های اساسی دیگری که باید مورد بررسی قرار گیرد توزیع ضریب خوشه بندی نسبت به



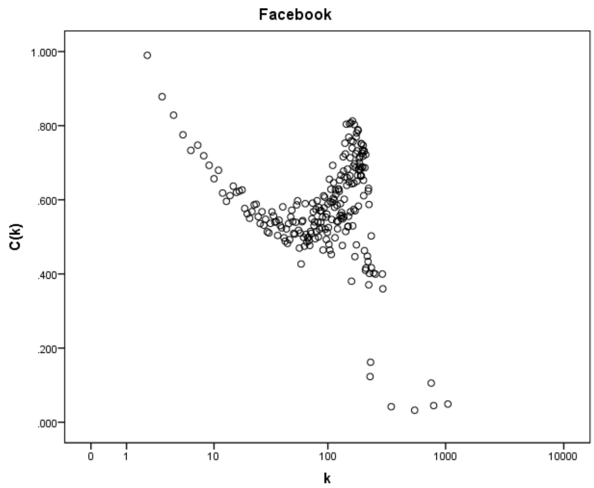
(آ) نمودار لگاریتمی توزیع درجات



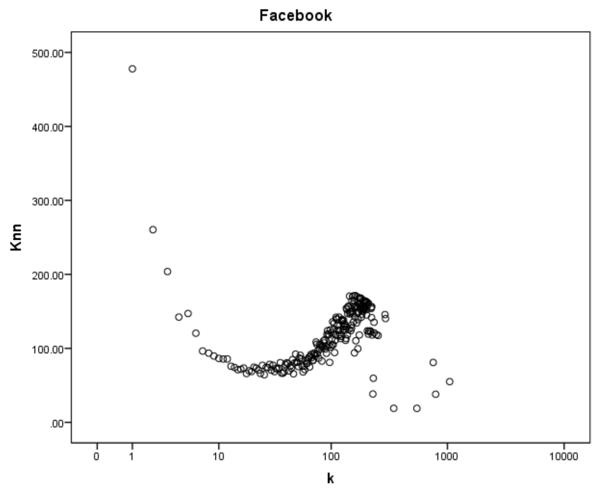
(ب) نمودار توانی توزیع درجات

شکل ۱.۳: توزیع درجات لگاریتمی و توانی گراف مربوط به شبکه اجتماعی فیس بوک

درجه راس‌ها و توزیع میانگین درجات همسایه‌ها نسبت به درجه راس می‌باشد. در شکل ۱۹.۳ توزیع ضریب خوشه‌بندی نسبت به درجه (۲.۳(آ)) و نمودار میانگین درجات همسایه‌ها نسبت به درجه راس (۲.۳(ب)) را مشاهده می‌کنیم. برخلاف انتظار، با بزرگ شدن درجه، ضریب خوشه‌بندی بزرگ می‌شود و این مخالف چیزی است که انتظار می‌رود شبکه‌های اجتماعی از خود بروز بدهند. همچنین این گراف دارای شرکت‌پذیری مثبت می‌باشد.



(آ) نمودار ضریب خوشه‌بندی نسبت به درجه راس



(ب) نمودار میانگین درجات همسایه‌ها نسبت به درجه راس

شکل ۲.۳: نمودار توزیع ضریب خوشه‌بندی و میانگین درجات همسایه

در کنار این ویژگی‌ها که به صورت طبیعی در تحلیل شبکه‌های اجتماعی مورد بررسی قرار می‌گیرد، ما در اینجا به بررسی ویژگی دیگری نیز می‌پردازیم که به نظر بررسی آن مهم است، و آن ویژگی‌ها عبارتند از تعداد

جدول ۳-۳: جدول مربوط به تعداد و درجه راس‌های قطب گراف فیس بوک

ترتیب	۱	۲	۳	۴	۵	۶	۷	۸	۹	۱۰
درجه	۱۰۴۵	۷۹۲	۷۵۵	۵۴۷	۲۹۴	۲۹۱	۲۵۴	۲۳۵	۲۳۴	۲۳۱

و چگونگی قطب‌ها (یا راس‌هایی با درجه بسیار بالا). جدول ۳-۳ ده قطب با بیشترین درجه را در گراف فیس بوک نشان می‌دهد. همانطور که مشاهده می‌شود قطب‌ها درجه بسیار بزرگی را دارا هستند. قطب‌ها اهمیت ویژه‌ای در شبکه‌های اجتماعی و پایدار بودن آن دارند. با حذف قطب‌ها، شبکه اجتماعی از هم متلاشی شده و به چندین مولفه همبندی با تعداد کمی راس تجزیه می‌شود.

اما پیش از بررسی مدل‌ها باید اشاره کنیم. که با توجه به تعداد پارامترها موجود در مدل‌ها برای تولید گراف‌هایی مطابق با گراف فیس بوک باید برخی از ویژگی‌های گراف‌ها را با هم همسان کنیم. دو مدل DEB و Vaz تنها یک پارامتر آزاد دارند لذا تنها با همسان کردن درجه‌ی این مدل‌ها با گراف فیس بوک، می‌توانیم گرافی حاصل کنیم که شباهت بیشتری را با گراف فیس بوک داشته باشد. اما مدل‌های دیگر با توجه به دارا بودن تعداد بیشتر پارامتر نیاز دارند به منظور تولید گرافی که بیشترین تشابه را با گراف فیس بوک داشته باشد، علاوه به میانگین درجات یکسان، ضریب خوشه‌بندی تقریبا مساوی با گراف شبکه اجتماعی حقیقی داشته باشند. اما حتی یکسان سازی این ویژگی نیز برای تولید گرافی با بیشتر شباهت با داده واقعی برای مدل WPR کافی نیست. برای یکسان کردن این شبکه اجتماعی براستی هیچ ابزار دیگری در دست نداریم. بدین منظور سعی کردیم با همسان کردن میانگین کوتاه‌ترین مسیرهای گراف WPR و گراف فیس بوک شبکه‌ای تولید کنیم که بیشتر شباهت را با داده واقعیمان داشته باشد. تا انتهای این فصل تمامی مدل‌ها ده بار تولید شده؛ نتایج و نمودارها میانگین این ده بار اجرا می‌باشد.

۳.۳ تحلیل گراف حاصل از مدل DEB

همانطور که اشاره شد تنها پارامتر آزاد برای تغییر در این مدل احتمال حذف است. با توجه به اینکه در این مدل از حذف راس برای رسیدن گراف به حالت پایدار استفاده می‌شود، با قرار دادن احتمال حذف $p = ۰/۰۱۱$ گراف تولید کردیم که دارای ویژگی‌های زیر می‌باشد:

جدول ۳-۴ ویژگی‌های گراف DEB را نشان می‌دهد همانطور که مشاهده می‌شود این مدل ضریب خوشه

جدول ۳-۴: اطلاعات مربوط به گراف تولید شده توسط مدلDEB

نام گراف	تعداد راس‌ها	تعداد یال‌ها	میانگین درجه گراف	ضریب خوشه بندی	تعداد مولفه همبندی	شعاع	قطر	میانگین کوتاه‌ترین مسیر
DEB	۴۰۳۹	۸۸۹۵۸	۴۴.۰۵	۰.۳۸	۱	۴	۶.۳	۲.۷۸

جدول ۳-۵: جدول مربوط به تعداد و درجه راس‌های قطب گراف DEB

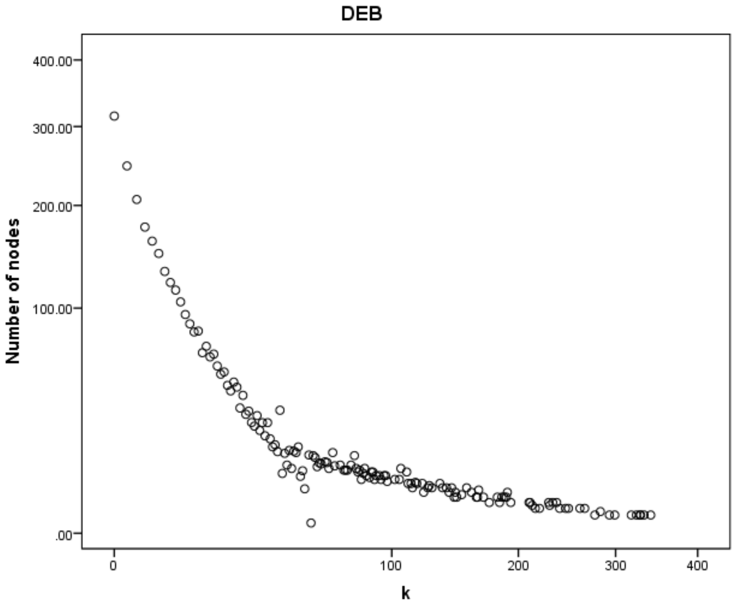
ترتیب	۱	۲	۳	۴	۵	۶	۷	۸	۹	۱۰
درجه	۳۸۶.۲	۳۵۳.۶	۳۴۰.۳	۳۳۰.۴	۳۱۹.۱	۳۱۳.۴	۳۰۸	۳۰۴.۲	۲۹۸.۹	۲۹۶.۲

بندی بسیار کوچک‌تر از ضریب خوشه‌بندی گراف واقعی تولید می‌کند. در مدل تولید شده ما تنها حدود ۴۱۰۰ یال از طریق جستجوی سراسری ایجاد شده‌اند و بقیه یال‌ها یعنی حدود ۸۴۰۰۰ یال از طرق جستجوی محلی ایجاد شده‌اند. با این وجود ضریب خوشه‌بندی گراف بسیار کوچک‌تر از ضریب خوشه بندی گراف فیس بوک است.

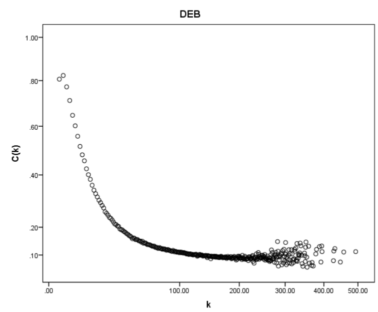
در شکل ۲۰.۳ نمودار ۳.۳(آ) نمودار توزیع درجات را نشان می‌دهد. همانطور که مشاهده می‌کنید با توجه به وجود ویژگی الحاق امتیازی مطابق انتظار، گراف دارای توزیع توانی می‌باشد. نمودار ۳.۳(ب) نشان دهنده رابطه ضریب خوشه بندی راس با درجه آن در گراف می‌باشد. همانطور که مشاهده می‌شود رابطه‌ی عکس بین آن‌ها برقرار است. نمودار ۳.۳(ج) نشان دهنده شرکت‌پذیر بودن مدل می‌باشد. با توجه به اینکه میانگین تعداد همسایه‌های مدل با افزایش درجه افزایش پیدا می‌کند، این مدل دارای همبستگی درجه به درجه با شرکت‌پذیری مثبت است.

همانطور که در شکل ۳.۳(آ) مشاهده می‌شود نمودار توزیع درجات گراف دارای شیب تندی است و درجه راس‌هایش حداکثر به ۴۵۰ می‌رسد. این در حالی است که نمودار توزیع درجات مربوط به گراف فیس بوک دارای شیب کندتری است و درجه راس‌های آن حتی به بیش از ۱۰۰۰ می‌رسد. همانطور که در جدول ۳-۲۳ مشاهده می‌کنید این مدل دارای قطب‌هایی با درجه نسبتا کوچک می‌باشد.

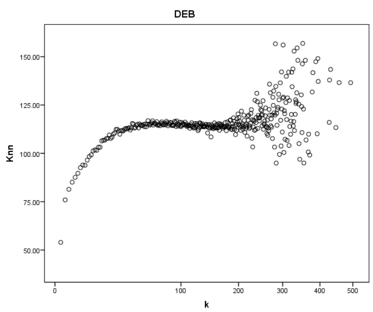
این مدل نه توانایی تولید گرافی با ضریب خوشه‌بندی هم اندازه با گراف فیس بوک را دارد و نه می‌تواند قطب‌هایی با آن درجه تولید کند. مدل می‌تواند رابطه معکوس بین ضریب خوشه‌بندی و درجه را مانند



(آ) نمودار توزیع درجات گراف DEB



(ب) نمودار ضریب خوشه‌بندی نسبت به درجه راس گراف DEB



(ج) نمودار میانگین درجات همسایه‌ها نسبت به درجه راس DEB

شکل ۳.۳: نمودارهای توزیع درجات، ضریب خوشه بندی و میانگین درجه همسایه‌های گراف DEB

شبکه‌های اجتماعی و (بر خلاف گراف فیس بوک) تولید کند. همچنین شرکت‌پذیری مثبت، که از ویژگی‌های مهم شبکه‌های اجتماعی است در گراف‌های تولیدی در این مدل موجود می‌باشد.

۴.۳ تحلیل مدل Vaz

مدل Vaz نیز تنها دارای یک پارامتر آزاد می‌باشد که آن عبارت است از احتمال اضافه شدن یک یال بر اساس جستجوی محلی، یعنی تبدیل یال بالقوه به یال واقعی (u) و یا افزوده شدن یک راس و ایجاد یک یال براساس جستجوی سراسری بین راس تازه افزوده شده و یکی از راس‌های گراف با احتمال $1 - u$. برای رسیدن به

جدول ۳-۶: اطلاعات مربوط به گراف تولید شده توسط مدل Vaz

نام گراف	تعداد راس‌ها	تعداد یال‌ها	میانگین درجه گراف	ضریب خوشه بندی	تعداد مولفه همبندی	شعاع	قطر	میانگین کوتاه‌ترین مسیر
Vaz	۴۰۳۹	۸۷۸۸۹	۴۳.۵۲	۰.۳۴۲	۱	۴.۷	۸.۸	۳.۶۶

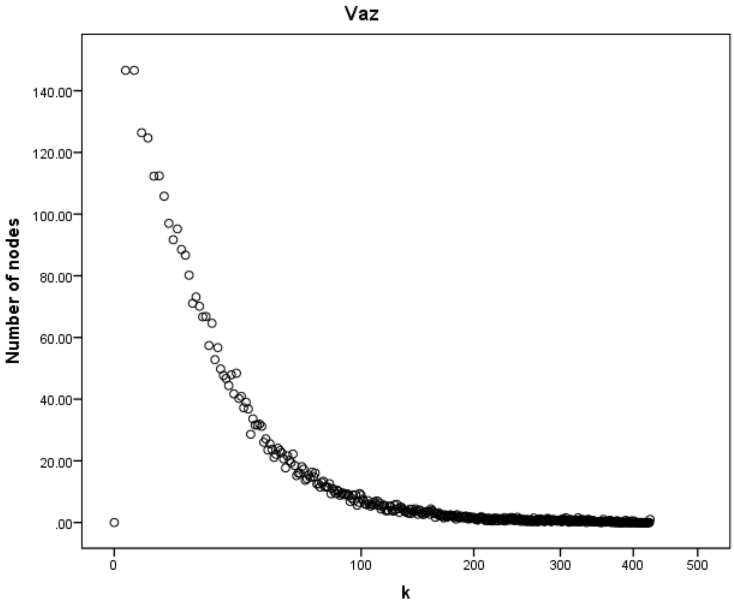
جدول ۳-۷: جدول مربوط به تعداد و درجه راس‌های قطب گراف Vaz

ترتیب	۱	۲	۳	۴	۵	۶	۷	۸	۹	۱۰
درجه	۳۶۱.۷	۳۵۳	۳۴۶.۴۰۲	۳۴۲.۳	۳۳۷	۳۳۳.۶	۳۳۲.۱	۳۲۹.۸	۳۲۷.۷	۳۲۴.۷

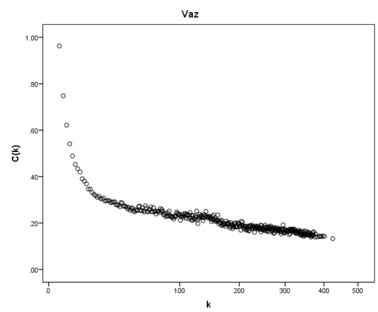
میانگین درجه حدود $k =$ ۴۳٫۵، احتمال افزوده شدن یال را $u =$ ۰٫۹۶۶ و احتمال افزوده شدن یک راس را $1 - u =$ ۰٫۰۳۴ قرار دادیم.

جدول ۳-۶ مشخصات مربوط به گراف تولید شده توسط مدل Vaz می‌باشد. همانطور که مشاهده می‌شود این مدل ضریب خوشه‌بندی کوچک‌تری را نسبت به گراف DEB تولید می‌کند. میانگین کوتاه‌ترین مسیر گراف Vaz بزرگتر از گراف DEB می‌باشد، اما تقریبا مساوی با گراف فیس بوک است.

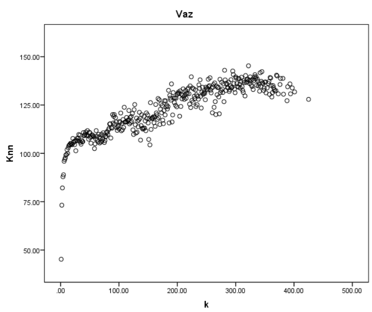
شکل ۲۱.۳ نمودارهای توزیع درجه ۴.۳(آ)، ضریب خوشه بندی نسبت به درجه ۴.۳(ب) و میانگین درجات نزدیک‌ترین همسایه نسبت به درجه ۴.۳(ج) را نشان می‌دهد. همانطور که مشاهده می‌شود، نمودار توزیع درجات تشابه بسیار زیادی با نمودار توزیع درجات فیس بوک دارد با این تفاوت که شیب نمودار نسبت به توزیع درجات فیس بوک بسیار تندتر است. این مدل راس‌هایی با حداکثر درجه ۴۵۰ تولید می‌کند. همانطور که پیش از این نیز اشاره شد درجه قطب‌ها در گراف فیس بوک بسیار بزرگ‌تر است. در شکل نیز مشاهده می‌شود که گراف‌های تولیدی برای نسبت میانگین درجات همسایگان و درجه راس (مشابه DEB) رفتاری مشابه با گراف فیس بوک دارند. این مدل نیز مانند مدل قبل رابطه عکسی بین ضریب خوشه‌بندی و درجه راس‌هایش دارد. همچنین در این گراف، قطب‌هایی با درجه بزرگ‌تر نسبت به مدل DEB تولید می‌شود(جدول ۳-۸).



(آ) نمودار توزیع درجات گراف Vaz



(ب) نمودار ضریب خوشه‌بندی نسبت به درجه راس گراف Vaz



(ج) نمودار میانگین درجات همسایه‌ها نسبت به درجه راس Vaz

شکل ۴.۳: نمودارهای توزیع درجات، ضریب خوشه بندی و میانگین درجه همسایه‌های گراف Vaz

۵.۳ تحلیل مدل MVS

برای تولید مدل MSV، احتمال جستجوی محلی $p_l =$ ۰٫۰۹۳۴ و جستجوی سراسری $p_g =$ ۰٫۰۰۰۵ و احتمال حذف را برابر ۰.۰۰۴ قرار دادیم. پس از رسیدن مدل به حالت پایدار گرافی تولید شد که، اطلاعات مربوط به این گراف در جدول ۳-۸ موجود می‌باشد.

برای این گراف با توجه به ابزارها و پارامترهای آزاد سعی کردیم ضریب خوشه‌بندی را مساوی ضریب خوشه‌بندی گراف فیس بوک قرار دهیم، اما این کار امکان پذیر نشد. ما جستجوی محلی را چندین برابر بزرگ‌تر جستجوی سراسری انتخاب کردیم. برای این کارمان دو دلیل داشتیم نخست آنکه با توجه به درجه میانگین گراف هر یال

جدول ۳-۸: اطلاعات مربوط به گراف تولید شده توسط مدلMVS

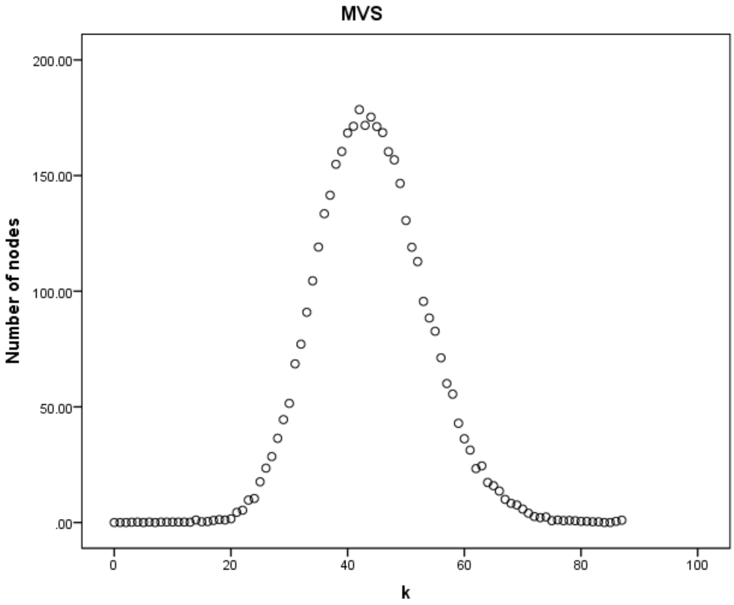
نام گراف	تعداد راس‌ها	تعداد یال‌ها	میانگین درجه گراف	ضریب خوشه بندی	تعداد مولفه همبندی	شعاع	قطر	میانگین کوتاه‌ترین مسیر
MVS	۴۰۳۹	۸۸۸۱۷	۴۳.۹۸	۰.۰۳۵	۱	۳	۴	۲.۶۰۷

جدول ۳-۹: جدول مربوط به تعداد و درجه راس‌های قطب گراف MVS

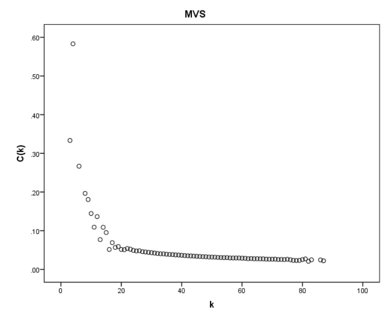
ترتیب	۱	۲	۳	۴	۵	۶	۷	۸	۹	۱۰
درجه	۸۳.۷	۸۰.۵	۷۹.۷	۷۸.۲	۷۷.۴	۷۴.۱	۷۳.۳	۷۳.۸	۷۲.۷	۷۲.۶

حداقل باید یک یال با جستجوی سراسری داشته باشد تا امکان جستجوی محلی وجود داشته باشد. از طرفی با کوچک‌تر کردن جستجوی سراسری راس‌های تنهایی تولید می‌کردند که دیگر یالی طی جستجوی سراسری برای این راس‌ها ایجاد نمی‌شد. در نهایت توانستیم در بهترین شرایط گرافی تولید کنیم که ضریب خوشه بندی ۰.۰۳۵ داشته باشد، و این مطابق ویژگی‌های این مدل بود که ضریب خوشه‌بندی با بزرگ شدن درجه به ۰.۰۱ همگرا می‌شد. این گراف همچنین میانگین کوتاه‌ترین مسیری کوچک‌تر از گراف Vaz و تقریبا مساوی گراف مدل DEB دارد.

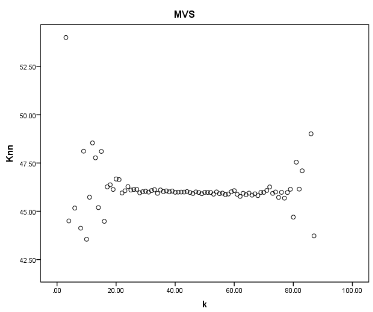
در شکل ۲۲.۳ نمودارهای توزیع درجات ۵.۳(آ)، نسبت ضریب خوشه‌بندی به درجه ۵.۳(ب) و میانگین درجه همسایگان ۵.۳(ج) را مشاهده می‌کنید. نمودار توزیع درجات گراف کمی متفاوت از آن چیزی است که در مدل‌های پیشین و همچنین داده واقعی خود مشاهده کردیم. البته این ارائه دهندگان در [۵۱]، هیچ تاکیدی بر وجود این تشابه ندارند. این مدل همانند گراف تصادفی اردوش-رنیی دارای درجه توزیع پواسون می‌باشد. در قسمت نزولی نمودار، نمودار شیب بسیار تندی را دارد و حداکثر درجه راس‌هایش به ۱۰۰ می‌رسد. جدول ۳-۹ نشان می‌دهد که در این مدل قطب‌ها نسبت به دو مدل پیشین و البته گراف فیس بوک درجه بسیار کوچک‌تری را دارا هستند.



(آ) نمودار توزیع درجات گراف MVS



(ب) نمودار ضریب خوشه‌بندی نسبت به درجه راس گراف MVS



(ج) نمودار میانگین درجات همسایه‌ها نسبت به درجه راس MVS

شکل ۵.۳: نمودارهای توزیع درجات، ضریب خوشه بندی و میانگین درجه همسایه‌های گراف MVS

۶.۳ تحلیل مدل BPDA

در این مدل با قرار دادن فضایی که نقاط بر روی آن پرا کنده می‌شوند برابر $[۰, ۱۱۰۰۰]$ و $\beta = ۴۴$ و $\alpha = ۳$ توانستیم گرافی تولید کنیم که میانگین درجه و ضریب خوشه‌بندی تقریبا یکسانی با گراف فیس بوک داشته باشد.

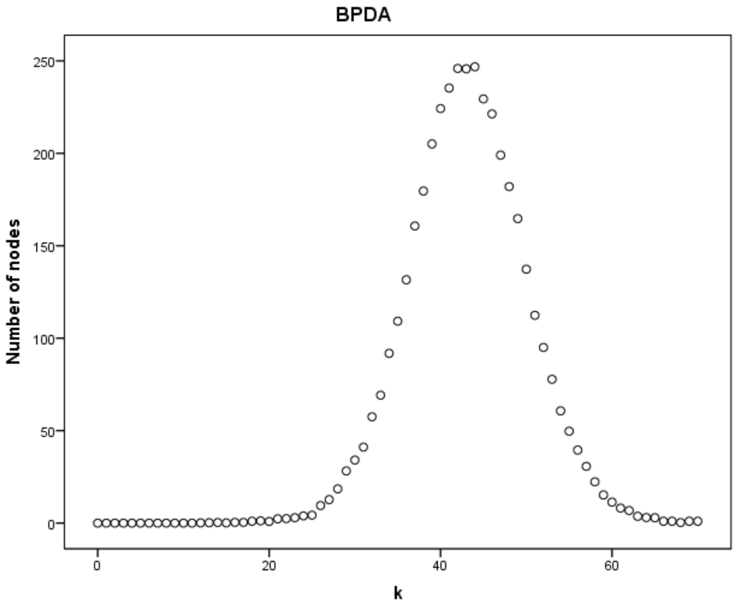
در این مدل همه چیز طبیعی است به جز کمیت‌های مربوط به فاصله. تویونن و همکاران نیز به منظور شبیه‌سازی گراف وب سایت last.fm که دارای میانگین کمترین فاصله‌ی ۷.۴ بود، توسط مدل BPDA، گرافی تولید کردند که دارای میانگین کمترین فاصله ۲۳.۹ بود. لذا به نظر می‌رسد بزرگ بودن مقدار میانگین کمترین

جدول ۳-۱۰: اطلاعات مربوط به گراف تولید شده توسط مدلBPDA

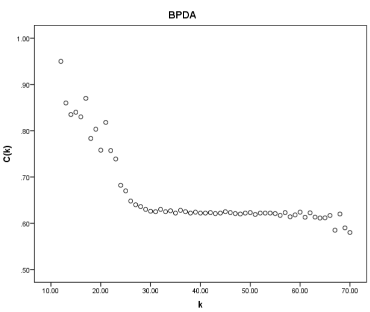
نام گراف	تعداد راس‌ها	تعداد یال‌ها	میانگین درجه گراف	ضریب خوشه بندی	تعداد مولفه همبندی	شعاع	قطر	میانگین کوتاه‌ترین مسیر
BPDA	۴۰۳۹	۸۸۲۸۶	۴۳.۱۴	۰.۶۲	۱	۳۹.۵	۷۸.۶	۲۶.۷۶

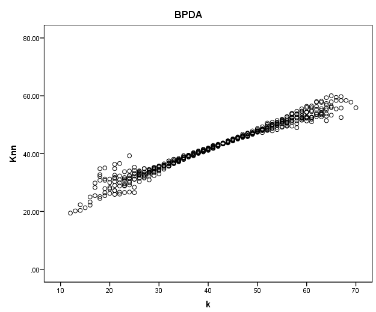
فاصله بین راس‌ها مربوط به ذات مدل باشد.

شکل ۲۳.۳ حاوی نمودارهای توزیع درجات ۶.۳(آ)، ضریب خوشه‌بندی ۶.۳(ب) و میانگین درجات همسایه‌ها



(آ) نمودار توزیع درجات گراف BPDA





(ب) نمودار ضریب خوشه‌بندی نسبت به درجه راس گراف BPDA

(ج) نمودار میانگین درجات همسایه‌ها نسبت به درجه راس BPDA

شکل ۶.۳: نمودارهای توزیع درجات، ضریب خوشه بندی و میانگین درجه همسایه‌های گراف BPDA

جدول ۳-۱۱: جدول مربوط به تعداد و درجه راس‌های قطب گراف BPDA

ترتیب	۱	۲	۳	۴	۵	۶	۷	۸	۹	۱۰
درجه	۶۶	۶۴.۸	۶۳.۵	۶۳.۱	۶۲.۹	۶۲.۷	۶۲.۶	۶۲.۲	۶۲	۶۱.۸

۶.۳(ج) است. این مدل نیز مانند مدل MVS، توزیع درجات پوآسون است، این مدل حتی بزرگترین درجه کمتری نسبت به مدل MVS دارد. به طوری که حداکثر درجه راس‌های آن کمتر از ۸۰ می‌باشد. اما ضریب خوشه بندی رفتاری مشابه داده حقیقی دارد. نمودار ضریب خوشه‌بندی، نشان می‌دهد با افزایش درجه ضریب خوشه‌بندی تفاوتی نمی‌کند. در جدول ۳-۱۱ درجه قطب‌های تولید شده توسط این مدل را مشاهده می‌کنید.

همانطور که مشاهده شد این مدل توانایی تولید گرافی با ضریب خوشه‌بندی بالا را دارد، اما به خاطر زیاد بودن میانگین کوتاه‌ترین درجه حتی توانایی تولید گرافی با ویژگی جهان کوچک را ندارد.

۷.۳ تحلیل مدل TOSHK

این مدل نیز همچون دیگر مدل‌های تکاملی توانایی تولید گرافی با ضریب خوشه‌بندی بسیار بزرگ را ندارد. همانطور که پیش از این نیز اشاره شد این مدل دارای جستجوی محلی و سراسری است. با توجه به اینکه ضریب خوشه‌بندی با افزایش جستجوی محلی افزایش می‌یابد. در این مدل با احتمال بسیار کم یعنی با احتمال $1 - p = ۰/۰۰۱$، دو راس را در جستجوی سراسری انتخاب می‌کنیم. لذا با احتمال $p = /۹۹۹$، یک راس را در جستجوی محلی انتخاب می‌شود، تا اکثر یال‌های ایجاد شده توسط جستجوی محلی باشد. بدین منظور در انتخاب تعداد راس‌های ثانویه برای جستجوی محلی یک عدد را به تصادف بین$U[۰, ۱۸۹[$ انتخاب می‌کنیم. بدین طریق از حدود ۸۹۰۰۰ یال ایجاد شده و موجود در مدل، ۴۰۳۹ عدد به واسطه جستجوی سراسری ایجاد شده بود یعنی حداقل تعداد ممکن و مابقی یال‌ها حاصل جستجوی محلی هستند. با اعمال تمام این شرایط باز مدل نتوانست گرافی با ضریب خوشه‌بندی بزرگ تولید کند، با وجود اینکه کمترین تعداد ممکن یال را از طریق جستجوی سراسری تولید شد، اما مدل نتوانست گرافی با ضریب خوشه‌بندی بیش از ۰.۳۵ تولید کند. در جدول ۳-۱۲، مشخصات گراف تولید شده را مشاهده می‌کنید. علاوه بر ضریب خوشه‌بندی که پیش از این اشاره شد دیگر کمیت‌ها مقادیر قابل قبول را دارند. برخلاف مدل BPDA در این مدل شعاع و قطر کوچک، نشانی

جدول ۳-۱۲: اطلاعات مربوط به گراف تولید شده توسط مدلTOSHK

نام گراف	تعداد راس‌ها	تعداد یال‌ها	میانگین درجه گراف	ضریب خوشه بندی	تعداد مولفه همبندی	شعاع	قطر	میانگین کوتاه‌ترین مسیر
TOSHK	۴۰۳۹	۸۸۷۹۷	۴۳.۹۷	۰.۳۵	۱	۴	۸.۸	۳.۶۶

جدول ۳-۱۳: جدول مربوط به تعداد و درجه راس‌های قطب گراف TOSHK

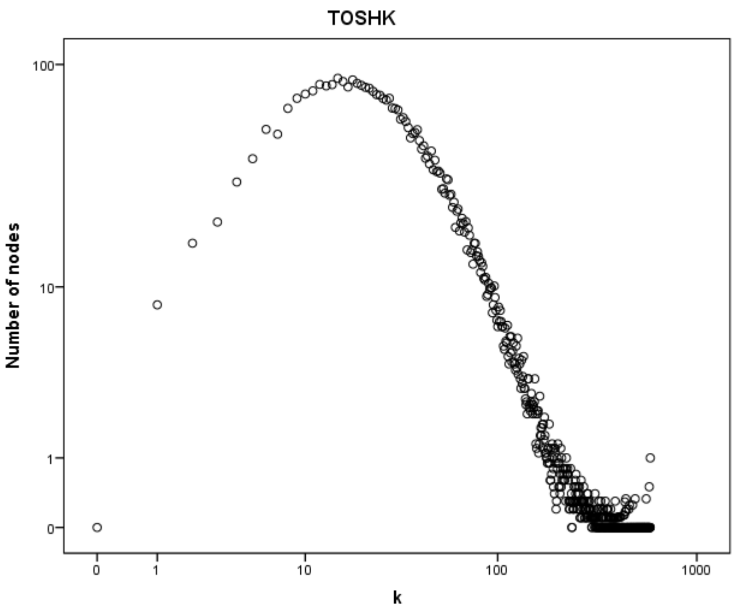
ترتیب	۱	۲	۳	۴	۵	۶	۷	۸	۹	۱۰
درجه	۴۷۸.۳	۴۱۹.۷	۳۹۶.۹	۳۶۹.۲	۳۴۸.۴	۳۳۶.۴	۳۲۶.۲	۳۱۸.۳	۳۰۷.۸	۲۹۶.۶

از کوچک بود میانگین کوتاه‌ترین مسیر است. میانگین کوتاه‌ترین مسیر در این مدل تقریبا مساوی کوتاه‌ترین مسیر گراف فیس بوک می‌باشد.

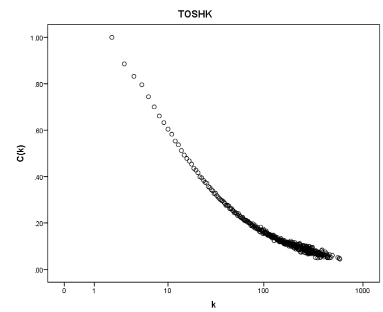
شکل ۲۴.۳، نمودارهای توزیع درجات ۷.۳(آ)، ضریب خوشه‌بندی ۷.۳(ب) و میانگین درجات همسایه‌های راس ۷.۳(ج) را نشان می‌دهد. همانطور که در شکل واضح است توزیع درجات مدل، توزیع توانی و مشابه داده واقعی می‌باشد، البته با شیبی کمی تندتر به طوری که بیشتر درجه راس‌ها به حدود ۸۰۰ هم می‌رسد. از این نظر این مدل از بین مدل‌های دیگر به داده واقعی شبیه‌تر است. ضریب خوش بندی هم رابطه عکس با درجه راس دارد. این مورد البته مشابه داده واقعی نیست، اما آن چیزی است که اصولا از شبکه‌های اجتماعی انتظار می‌رود. میانگین درجه همسایه‌ها هم تا حدودی مشابه داده واقعی است و حکایت از شرکت‌پذیری مثبت مدل دارد.

جدول ۳-۱۳ درجه قطب‌های گراف را نشان می‌دهد. این مدل قطب‌هایی تولید می‌کند که درجه نسبتا بزرگی دارند. البته هنوز درجه این قطب‌ها بسیار کم‌تر از داده واقعی است.

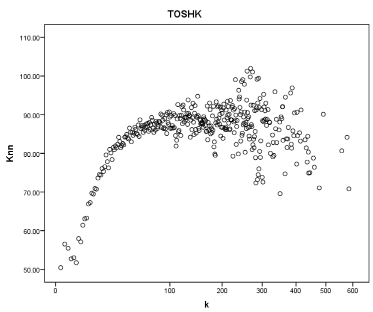
همانطور که اشاره شد این مدل گرافی تولید می‌کند که برخی از ویژگی‌های داده واقعی را شبیه‌سازی می‌کند، اما در بازتولید برخی ویژگی‌های دیگر ناتوان است. اما به جز این مورد مدل بسیاری دیگر از ویژگی‌های شبکه‌های اجتماعی را تولید می‌کند. هرچند رابطه بین ضریب خوشه‌بندی و درجه راس مشابه نیاز شبکه‌های اجتماعی است، اما چندان به داده واقعی مورد بررسی ما شباهت ندارد.



(آ) نمودار توزیع درجات گراف TOSHK



(ب) نمودار ضریب خوشه‌بندی نسبت به درجه راس گراف TOSHK



(ج) نمودار میانگین درجات همسایه‌ها نسبت به درجه راس TOSHK

شکل ۷.۳: نمودارهای توزیع درجات، ضریب خوشه بندی و میانگین درجه همسایه‌های گراف TOSHK

۸.۳ تحلیل مدل WPR

مدل WPR برخلاف مدل‌های تکاملی و مشابه دیگر مدل مبتنی بر ویژگی راس، توانایی تولید گرافی با میانگین درجه ۴۳.۵ و ضریب خوشه‌بندی بیش از ۰.۵۹ را دارد. برای این منظور $p =$ ۰٫۰۷۰۱۸ و احتمال ایجاد یال درون شعاع خاص را $p_b =$ ۰٫۵۷ (این مقدار به p افزوده می‌شود) و احتمال ایجاد یال خارج از محدود همسایگی (این مقدار از p کم می‌شود) $p_\Delta =$ ۰٫۰۶۹۴۸ انتخاب کردیم. نقاط با توجه به ویژگی مدل به صورت تابع پوواسون با میانگین ۷.۵ در فضای دو بعدی پخش شده و شعاع همسایگی را $H =$ ۲ قرار دادیم. گراف حاصله دارای ویژگی‌هایی است که در جدول ۳-۱۴ قابل مشاهده‌است. از بین مدل‌هایی که تاکنون بررسی شده برخی

از ویژگی‌های این مدل به نظر می‌رسد تشابه بیشتری با داده واقعی داشته باشد. البته این تشابه تنها به نزدیک بودن مقدار شعاع قطر و میانگین کوتاه‌ترین فاصله (در کنار ضریب خوشه‌بندی) محدود نمی‌شود.

شکل ۲۵.۳ حاوی نمودارهای توزیع درجات ۸.۳(آ)، ضریب خوشه‌بندی ۸.۳(ب) و میانگین درجات همسایه‌ها

جدول ۳-۱۴: اطلاعات مربوط به گراف تولید شده توسط مدلWPR

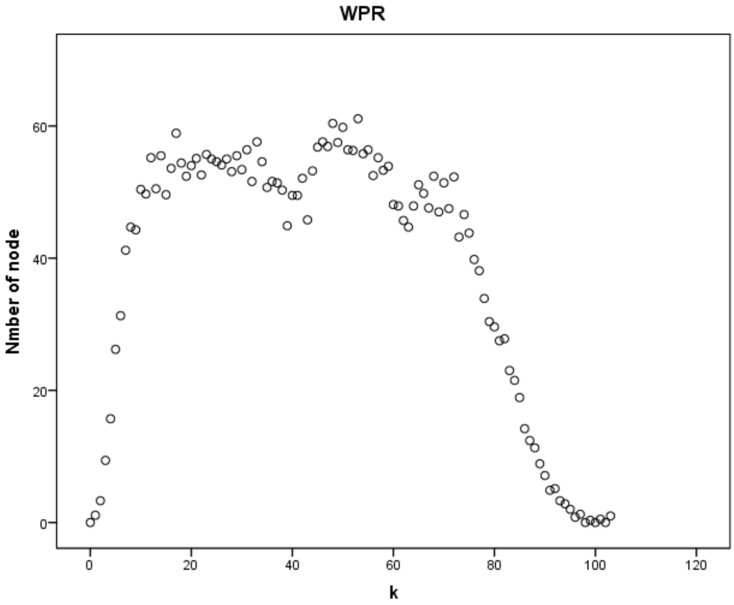
نام گراف	تعداد راس‌ها	تعداد یال‌ها	میانگین درجه گراف	ضریب خوشه بندی	تعداد مولفه همبندی	شعاع	قطر	میانگین کوتاه‌ترین مسیر
WPR	۴۰۳۹	۸۷۹۶۸	۴۳.۵۶	۰.۵۹	۱	۴	۵.۵	۳.۰۶

۸.۳(ج) است. نمودار توزیع درجات مطابق انتظار (مشابه مدل تولید شده توسط تویونن در [۱۳]) زیاد مشابه توزیع درجات داده واقعی نیست. البته در مقاله ارائه شده برای معرفی این مدل هیچ تاکیدی بر روی توزیع درجات نشده بود، اما شبیه‌سازی تویونن و همکاران و ما نشان داد این مدل چندان توانایی در باز تولید توزیع درجات گراف ندارد، اما به جز این مورد نمودارهای دیگر تقریبا مشابه نمودار داده واقعی است. این مدل برخلاف شبکه‌های اجتماعی پیش از این بررسی شده (و البته مشابه داده واقعی ما) دارای رابطه عکس بین ضریب خوشه‌بندی و درجه نیست. مشابه سایر شبکه‌های اجتماعی (و داده واقعی ما) دارای شرکت‌پذیری مثبت می‌باشد.

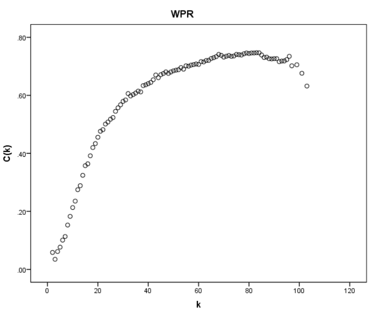
با تمام این اوصاف این مدل توانایی تولید مدلی که در آن قطب‌ها دارای تعداد زیادی همسایه باشند را ندارد. جدول ۳-۱۵ درجه قطب‌ها نشان داده شده‌است. این مدل نیز مانند دیگر مدل در دسته مدل‌های مبتنی بر ویژگی راس توانایی تولید راس با درجه بالا را ندارد.

جدول ۳-۱۵: جدول مربوط به تعداد و درجه راس‌های قطب گراف WPR

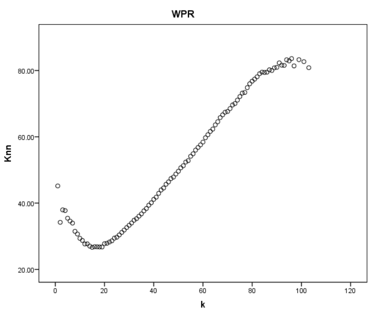
ترتیب	۱	۲	۳	۴	۵	۶	۷	۸	۹	۱۰
درجه	۹۵.۴	۹۳	۹۲.۴	۹۲	۹۱.۵	۹۱.۲	۹۰.۸	۹۰.۴	۹۰.۲	۹۰.۱



(آ) نمودار توزیع درجات گراف WPR



(ب) نمودار ضریب خوشه‌بندی نسبت به درجه راس گراف WPR



(ج) نمودار میانگین درجات همسایه‌ها نسبت به درجه راس WPR

شکل ۸.۳: نمودارهای توزیع درجات، ضریب خوشه بندی و میانگین درجه همسایه‌های گراف WPR

۹.۳ تحلیل مدل KOSKK

همانطور که اشاره شد این مدل توسط گراف‌های وزن‌دار سعی در تولید گرافی داشت که انجمن‌های بهتری تولید بکند. به منظور شبیه‌سازی گراف فیس بوک احتمال افزوده شدن یال توسط جستجوی محلی را برابر ۰.۸۰، احتمال افزوده شدن یال توسط جستجوی سراسری را برابر ۰.۰۰۰۰۰۱ و احتمال حذف راس ۰.۰۰۷ قرار دادیم. به منظور تولید گرافی با ضریب خوشه‌بندی بزرگ و انجمن‌های قوی $\delta = ۱$ و مقدار اولیه وزن را نیز $W_۰ = ۱$ انتخاب نمودیم. با این وجود باز هم مدل ضریب خوشه‌بندی گراف مد نظر ما را تولید نکرد. مشخصات این مدل در جدول ۳-۱۶ ذکر شده‌است.

جدول ۳-۱۶: اطلاعات مربوط به گراف تولید شده توسط مدلKOSKK

نام گراف	تعداد راس‌ها	تعداد یال‌ها	میانگین درجه گراف	ضریب خوشه بندی	تعداد مولفه همبندی	شعاع	قطر	میانگین کوتاه‌ترین مسیر
KOSKK	۴۰۳۹	۸۸۵۱۴	۴۳.۸۳	۰.۲۲	۱	۱.۳	۷.۴	۲.۷۳

جدول ۳-۱۷: جدول مربوط به تعداد و درجه راس‌های قطب گراف KOSKK

ترتیب	۱	۲	۳	۴	۵	۶	۷	۸	۹	۱۰
درجه	۲۰۵.۸	۱۷۳.۶	۱۶۶.۵	۱۶۲.۵	۱۵۷.۹	۱۵۵.۹	۱۵۴.۲	۱۵۲.۱	۱۵۰.۲	۱۴۸.۲

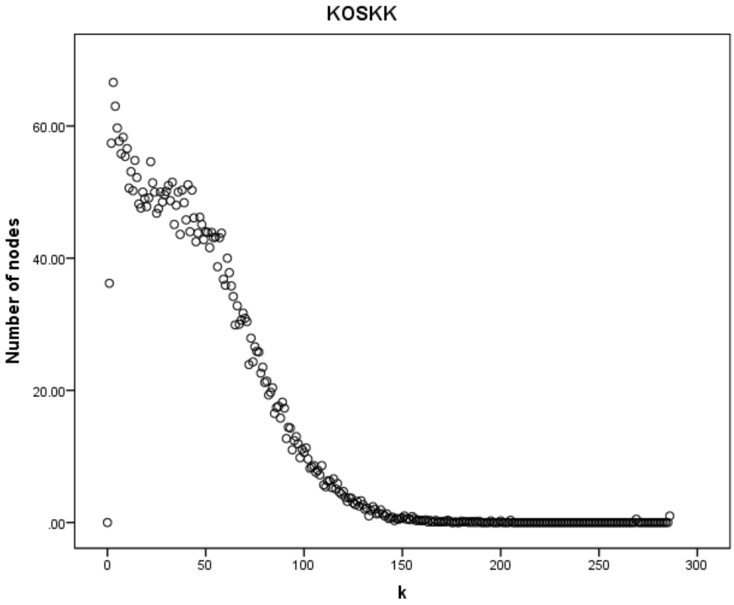
این مدل شعاع قطر و میانگین نزدیک‌ترین مسیر قابل قبولی را تولید می‌کند، اما همچنان که ذکر شد، مدل ضریب خوشه‌بندی نسبتا کوچکی دارد. در نمودار توزیع درجات این مدل (شکل ۹.۳(آ)) نشان می‌دهد مدل توزیع درجات توانی دارد. همچنین مدل توانسته است تا نمودار میانگین نزدیک‌ترین همسایگان (شکل ۹.۳(ج)) مشابه گراف فیس بوک را تولید کند. اما گراف نمودار ضریب خوشه‌بندی ((شکل ۹.۳(ب))) مشابه با گراف فیس بوک ندارد. قابل ذکر است که این مدل توانایی تولید قطب‌های نسبتا قدرتمندی را دارد. در جدول ۳-۳۵ درجه قوی‌ترین قطب‌های گراف آمده‌است.

۱۰.۳ مقایسه انجمن‌های تشکیل شده توسط مدل‌ها

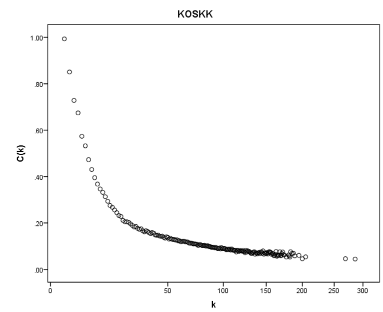
برای بدست آوردن انجمن‌های تشکیلی با توجه به محدودیت سخت افزاری مجبور شدیم از یک نمونه ۱۰۰۰ تایی از مدل‌ها و گراف فیس بوک استفاده کنیم. لذا در ادامه تمام نمودارها برای گراف‌های با تعداد ۱۰۰۰ راس بدست آمده است.

۱.۱۰.۳ انجمن‌های گراف فیس بوک

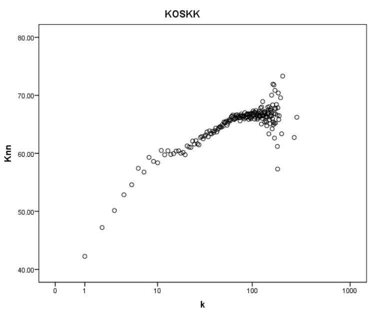
نخست به بررسی جدول گراف فیس بوک می‌پردازیم. شکل ۱۰.۳(آ) و شکل ۱۰.۳(آ)، نمودار تعداد انجمن‌های تشکیلی به ترتیب برای $k = ۴$ و $k = ۵$ می‌باشد. گراف فیس بوک برای $k = ۴$ از یک انجمن بزرگ با تعداد ۱۶۰ راس و تعدادی انجمن کوچکتر تشکیل شده‌است. برای $k = ۵$ نیز فیس بوک یک انجمن با تعداد ۷۹



(آ) نمودار توزیع درجات گراف KOSKK



(ب) نمودار ضریب خوشه‌بندی نسبت به درجه راس گراف KOSKK



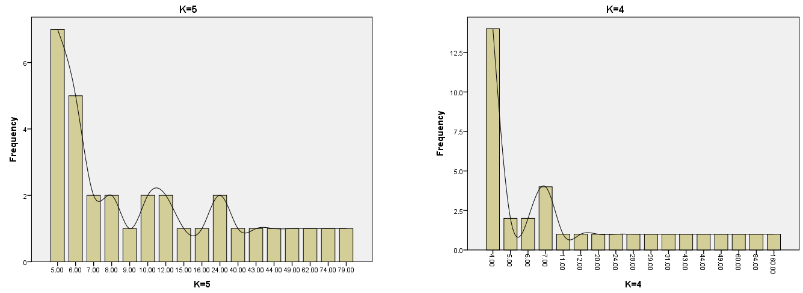
(ج) نمودار میانگین درجات همسایه‌ها نسبت به درجه راس KOSKK

شکل ۹.۳: نمودارهای توزیع درجات، ضریب خوشه بندی و میانگین درجه همسایه‌های گراف KOSKK

راس و تعدادی انجمن کوچک‌تر تشکیل می‌شود.

انجمن‌های تشکیلی DEB

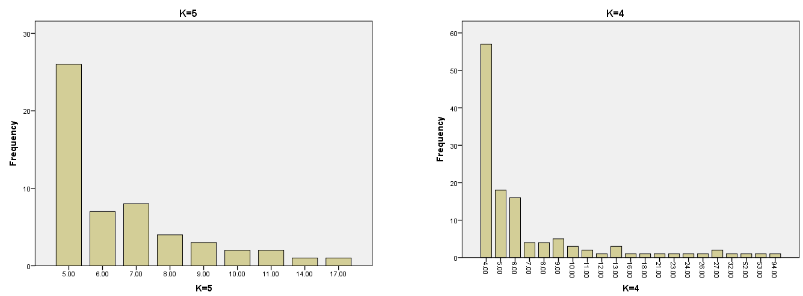
این مدل به علت تراکم زیاد یال توانسته تا حدودی انجمن‌های مقبولی تشکیل دهد. البته انجمن‌های این مدل نسبت به مدل فیس بوک از تعداد راس کمتری برخوردارند. شکل ۱۱.۳ تعداد انجمن‌های تشکیلی برای $K = ۴$ و $K = ۵$ را نشان می‌دهد.



(آ) نمودار تعداد انجمن‌ها برای $k =$ ۴ گراف فیس بوک

(ب) نمودار تعداد انجمن‌ها برای $k =$ ۵ گراف فیس بوک

شکل ۱۰.۳: نمودار تعداد انجمن‌های تشکیلی برای گراف فیس بوک



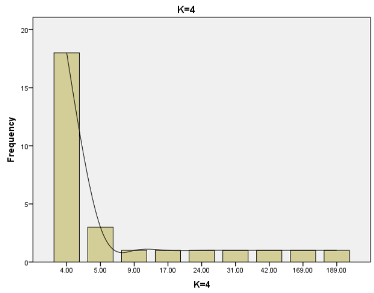
(آ) نمودار تعداد انجمن‌ها برای $k =$ ۴ گراف DEB

(ب) نمودار تعداد انجمن‌ها برای $k =$ ۵ گراف DEB

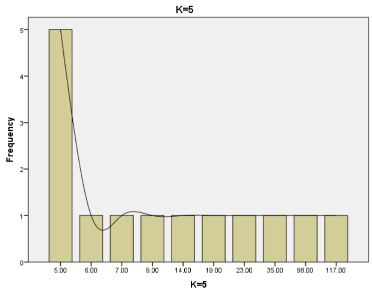
شکل ۱۱.۳: نمودار تعداد انجمن‌های تشکیلی برای گراف DEB

انجمن‌های تشکیلی Vaz

نمودارهای شکل ۱۲.۳، تعداد انجمن‌های تشکیلی برای مدل Vaz را نشان می‌دهد. این مدل باز به علت تراکم زیاد یال توانسته برای $k =$ ۴، دو انجمن به نسبت بزرگ بدست آورد. برای $k =$ ۵، در قیاس با انجمن‌های فیس بوک این مدل دو مولفه بزرگ تشکیل داده ولی تعداد انجمن‌های آن به صورت قابل توجهی نسبت به گراف فیس بوک کمتر است. می‌توان گفت که به علت میانگین درجه بالا دو مولفه به نسبت بزرگ تشکیل شده‌است. ویژگی مهم انجمن‌های گراف فیس بوک تعداد زیاد انجمن‌ها می‌باشد که البته تعداد راس‌های زیادی در هر انجمن حاضر هستند. هم این مدل و هم مدل Vaz بخصوص برای $k =$ ۵ توانایی تولید این تعداد انجمن را ندارند.



(آ) نمودار تعداد انجمن‌ها برای $k = ۴$ گراف Vaz

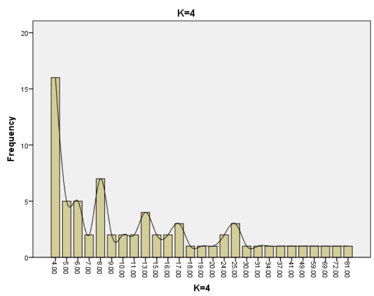


(ب) نمودار تعداد انجمن‌ها برای $k = ۵$ گراف Vaz

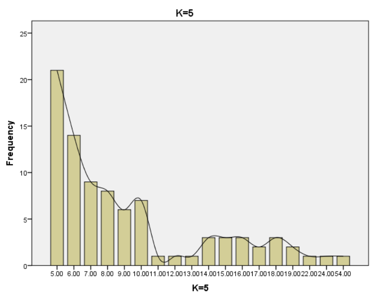
شکل ۱۲.۳: نمودار تعداد انجمن‌های تشکیلی برای گراف Vaz

انجمن‌های تشکیلی MVS

همانطور که برای دو مدل پیش مشاهده کردیم گراف فیس بوک برای $k = ۵$ تعداد بیشتری انجمن تولید می‌کند. شکل ۱۳.۳ تعداد انجمن‌های تشکیلی برای مدل MVS را نشان می‌دهد. برای $k = ۴$ و $k = ۵$ تعداد انجمن بیشتری نسبت به دو مدل پیش تولید می‌شود اما تعداد راس‌های موجود در این انجمن‌ها نسبت به گراف فیس بوک کمتر است. برای مثال سه انجمن بزرگ برای گراف فیس بوک ($k = ۵$) به ترتیب برابر ۷۹،۷۴،۶۲ ولی برای گراف MVS به ترتیب برابر ۵۴،۲۴ و ۲۲ می‌باشد. همین موضوع برای $k = ۴$ نیز صادق است. بزرگ‌ترین انجمن برای گراف فیس بوک ($k = ۴$) ۱۶۰ راس دارد، ولی بزرگ‌ترین انجمن برای گراف MVS دارای ۸۱ راس می‌باشد.



(آ) نمودار تعداد انجمن‌ها برای $k = ۴$ گراف MVS

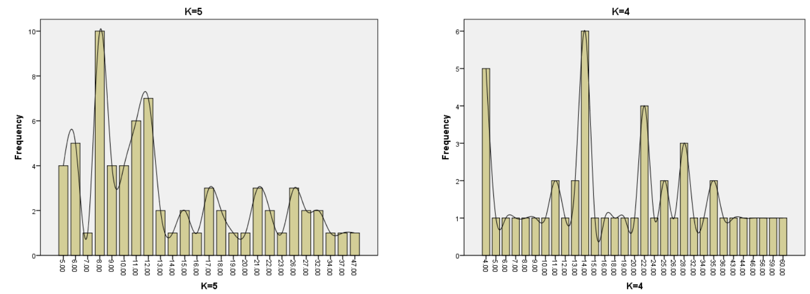


(ب) نمودار تعداد انجمن‌ها برای $k = ۵$ گراف MVS

شکل ۱۳.۳: نمودار تعداد انجمن‌های تشکیلی برای گراف MVS

انجمن‌های تشکیلی BPDA

این مدل نیز همانند مدل دارای تعداد نسبتا زیادی انجمن کوچک است. شکل ۱۴.۳، نمودار فراوانی انجمن‌های تشکیلی مدل BPDA می‌باشد. برای $k = ۴$ گراف فیس بوک دارای انجمنی با اندازه ۱۶۰ می‌باشد اما بزرگ‌ترین انجمن BPDA دارای ۶۰ راس می‌باشد. همین امر برای $k = ۵$ نیز صادق است، بزرگ‌ترین انجمن فیس بوک دارای ۷۹ راس می‌باشد در حالی که بزرگ‌ترین انجمن دارای تنها ۴۷ راس می‌باشد.



(آ) نمودار تعداد انجمن‌ها برای $k = ۴$ گراف BPDA

(ب) نمودار تعداد انجمن‌ها برای $k = ۵$ گراف BPDA

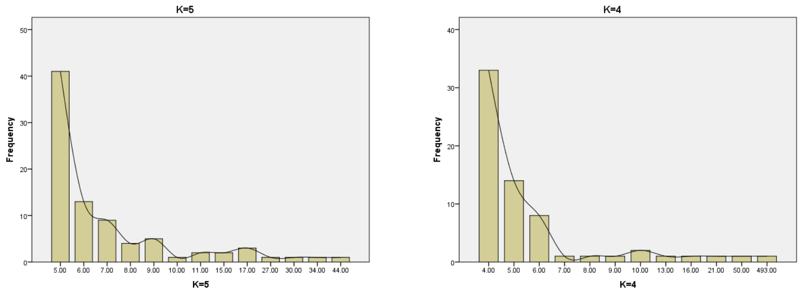
شکل ۱۴.۳: نمودار تعداد انجمن‌های تشکیلی برای گراف BPDA

انجمن‌های تشکیلی TOSHK

این مدل به لحاظ انجمن‌های تشکیلی با مدل‌های پیشین کمی تفاوت دارد(۱۵.۳). این مدل برای $K = ۴$ با توجه به تعداد زیاد جستجوی محلی که برای هر راس انجام دادیم یک انجمن بزرگ با ۴۹۳ راس و تعدادی انجمن بسیار کوچک‌تر می‌سازد. اما دلیل تشکیل این انجمن بزرگ نوع جستجوی ماست که چهار راس را بخاطر تعداد جستجوی محلی بسیار زیاد دو به دو به هم متصل می‌کند. اما برای $k = ۵$ بزرگ‌ترین انجمن دارای ۴۴ راس می‌باشد، برای گراف فیس بوک بزرگ‌ترین انجمن دارای ۷۹ راس می‌باشد.

انجمن‌های تشکیلی WPR

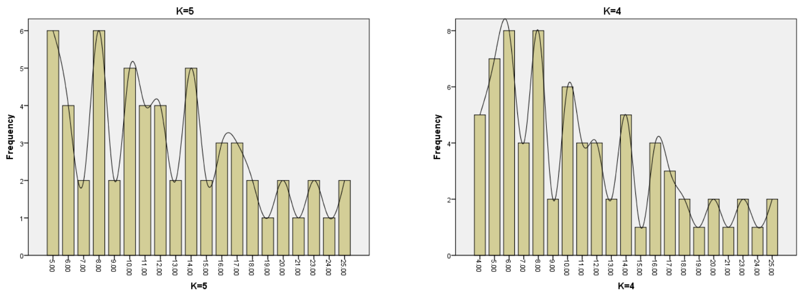
مدل به خاطر نوع جستجو و مقادیر انتخابی ما به عنوان پارامتر از تعداد بسیار زیادی انجمن که دارای تعداد نسبتا کمی راس هستند، تشکیل می‌شود(شکل ۱۶.۳). هم برای $k = ۴$ و هم برای $k = ۵$، ۴ انجمن بزرگ با ۲۵، ۲۴، ۲۳ راس هستند، که از انجمن‌های متناظر گراف فیس بوک بسیار کم‌تر می‌باشند. تعداد انجمن‌های کشف شده برای $k = ۴$ در گراف فیس بوک ۳۵ و برای گراف مدل WPR بیش از ۷۰ انجمن می‌باشد.



(آ) نمودار تعداد انجمن‌ها برای $k = ۴$ گراف TOSHK

(ب) نمودار تعداد انجمن‌ها برای $k = ۵$ گراف TOSHK

شکل ۱۵.۳: نمودار تعداد انجمن‌های تشکیلی برای گراف TOSHK



(آ) نمودار تعداد انجمن‌ها برای $k = ۴$ گراف WPR

(ب) نمودار تعداد انجمن‌ها برای $k = ۵$ گراف WPR

شکل ۱۶.۳: نمودار تعداد انجمن‌های تشکیلی برای گراف WPR

انجمن‌های تشکیلی KOSKK

این مدل نیز همانند مدل TOSHK از یک انجمن بزرگ با تعداد بسیار زیاد راس (برای $k = ۴$ تعداد ۲۴۸ راس) و تعدادی انجمن کوچک تشکیل شده‌است اما تفاوت این مدل با مدل TOSHK در تعداد انجمن‌های کوچک است که مدل TOSHK نسبت به این مدل تعداد بیشتری انجمن دارد.

۱۱.۳ جمع‌بندی تحلیل آماری مربوط به گراف فیس بوک

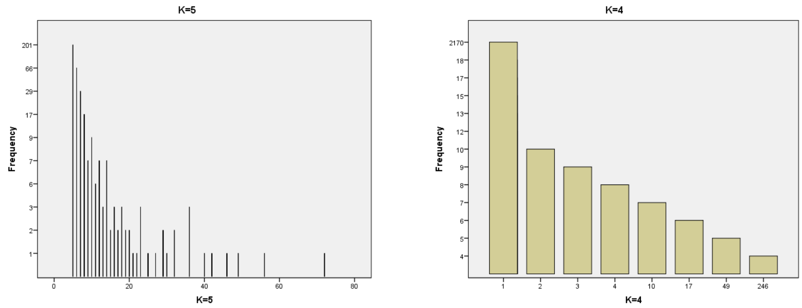
در جدول‌های ۳-۱۸ و ۳-۱۹ مقایسه‌ای کلی بین داده‌های واقعی و مدل‌های ارائه می‌دهد.

جدول ۳-۱۸: جدول مقایسه مدل‌ها

نام گراف	تعداد راس‌ها	تعداد یال‌ها	میانگین درجه گراف	ضریب خوشه بندی	تعداد مولفه همبندی	شعاع	قطر	میانگین کوتاه‌ترین مسیر
facebook.com	۴۰۳۹	۸۸۲۳۴	۴۳.۶۹۱	۰.۶۰	۱	۵	۸	۳.۶۹۳
DEB	۴۰۳۹	۸۸۹۵۸	۴۴.۰۵	۰.۳۸	۱	۴	۶.۳	۲.۷۸
Vaz	۴۰۳۹	۸۷۸۸۹	۴۳.۵۲	۰.۳۴	۱	۴.۷	۸.۸	۳.۶۶
MVS	۴۰۳۹	۸۸۸۱۷	۴۳.۹۸	۰.۰۳۵	۱	۳	۴	۲.۶۰۷
BPDA	۴۰۳۹	۸۸۲۸۶	۴۳.۱۴	۰.۶۱	۱	۳۹.۵	۷۸.۶	۲۶.۷۶
TOSHK	۴۰۳۹	۸۸۷۹۷	۴۳.۹۷	۰.۳۵	۱	۴	۸.۸	۳.۶۶
WPR	۴۰۳۹	۸۷۹۶۸	۴۳.۵۶	۰.۵۹	۱	۴	۵.۵	۳.۰۶
KOSKK	۴۰۳۹	۸۸۵۱۴	۴۳.۸۳	۰.۲۲	۱	۱.۳	۷.۴	۲.۷۳

جدول ۳-۱۹: جدول مقایسه تطبیقی مدل‌ها. برای معیارهای شعاع و قطر بازه‌ی نوسان $\pm$۱ برای ضریب خوشه‌بندی $\pm$۰٫۰۱ و برای کوتاه‌ترین مسیر $\pm$۰٫۶ را در نظر گرفته شده است.

نام گراف	تعداد راس‌ها	تعداد یال‌ها	میانگین درجه گراف	ضریب خوشه بندی	تعداد مولفه همبندی	شعاع	قطر	میانگین کوتاه‌ترین مسیر	قطب
facebook.com	۴۰۳۹	۸۸۲۳۴	۴۳.۶۹۱	۰.۶۰۵	۱	۵	۸	۳.۶۹۳	۱۰۴۵
DEB	✓	✓	✓	×	✓	✓	×	×	×
Vaz	✓	✓	✓	×	✓	✓	✓	✓	×
MVS	✓	✓	✓	×	✓	✓	×	×	×
BPDA	✓	✓	✓	✓	✓	×	×	×	×
TOSHK	✓	✓	✓	×	✓	✓	✓	✓	×
WPR	✓	✓	✓	✓	✓	✓	×	✓	×
KOSKK	✓	✓	✓	×	✓	×	✓	×	×



(آ) نمودار تعداد انجمن‌ها برای $k = ۴$ گراف KOSKK

(ب) نمودار تعداد انجمن‌ها برای $k = ۵$ گراف KOSKK

شکل ۱۷.۳: نمودار تعداد انجمن‌های تشکیلی برای گراف KOSKK

جدول ۳-۲۰: اطلاعات مربوط به شبکه گراف ویکی وُت

نام گراف	تعداد راس‌ها	تعداد یال‌ها	میانگین درجه گراف	ضریب خوشه بندی	شعاع	قطر	میانگین کوتاه‌ترین مسیر
wiki-Vote	۷۱۱۵	۱۰۳۶۸۹	۲۹.۱۴	۰.۱۴۱	۱	۷	۳.۲۴۸

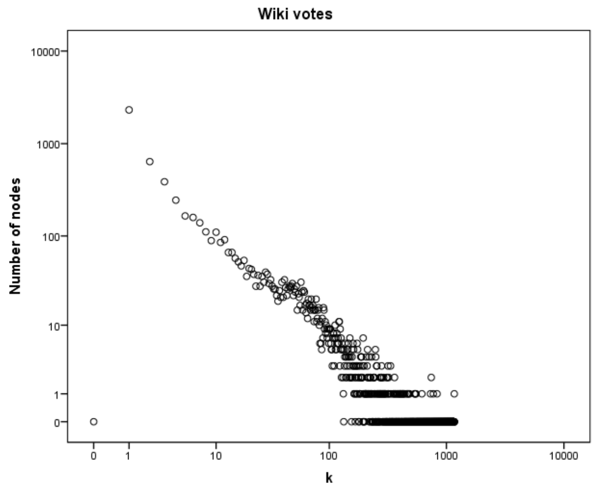
۱۲.۳ تحلیل گراف مربوط به ویکی وُت

در ادامه به بررسی تفاوت‌ها و تشابه‌های گراف‌های ارائه شده توسط مدل‌های شبکه‌های اجتماعی، با داده واقعی بدست آمده از وب سایت ویکی وت می‌پردازیم. ویژگی‌های گراف ویکی وت در جدول ۳-۲۰ آمده‌است.

در شکل ۱۸.۳ نمودار توزیع درجات گراف ویکی وت آمده‌است. همانطور که انتظار می‌رود این مدل دارای توزیع درجات توانی می‌باشد. همچنین در شکل ۱۹.۳(آ) نمودار نسبت بین ضریب خوشه‌بندی و درجه راس آمده است. آنچنان که برای شبکه‌های اجتماعی انتظار داریم، این گراف دارای رابطه عکس بین درجه راس و ضریب خوشه‌بندی آن می‌باشد. اما برخلاف مدل فیس بوک این مدل با توجه به شکل ۱۹.۳(ب) دارای وابستگی درجه به درجه با شرکت‌پذیری منفی می‌باشد.

جدول ۳-۲۱: جدول مربوط به تعداد و درجه راس‌های قطب گراف ویکی وت

ترتیب	۱	۲	۳	۴	۵	۶	۷	۸	۹	۱۰
درجه	۱۱۶۷	۸۳۲	۷۷۳	۷۴۳	۷۳۲	۶۱۸	۵۵۱	۵۴۳	۵۳۸	۵۳۷



شکل ۱۸.۳: نمودار توزیع درجات گراف ویکی وت

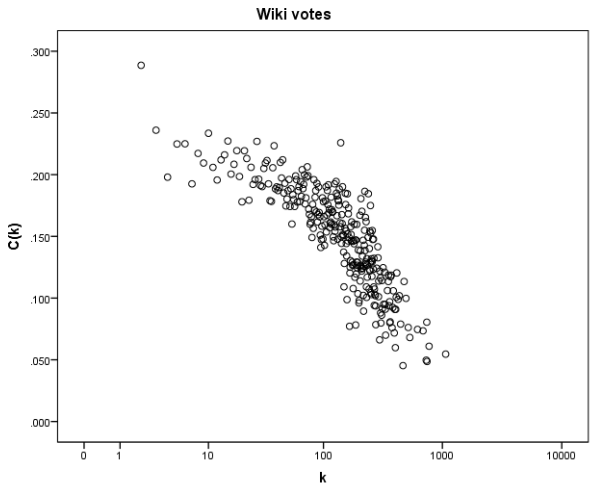
جدول ۳-۲۲: اطلاعات مربوط به گراف تولید شده توسط مدلDEB

نام گراف	تعداد راس‌ها	تعداد یال‌ها	میانگین درجه گراف	ضریب خوشه بندی	شعاع	قطر	میانگین کوتاه‌ترین مسیر
DEB	۷۱۱۵	۱۰۲۷۷۶	۲۸.۸۹	۰.۳۶۸	۳.۸	۷.۴	۳.۱۵

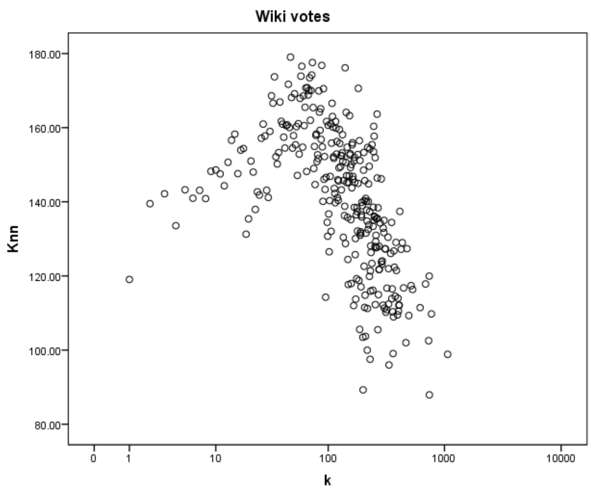
۱۳.۳ تحلیل گراف حاصل از مدل DEB

برای تولید گراف مشابه با گراف ویکی وت ۱۸ $p = $ ٪۰ قرار دادیم. ویژگی‌های مربوط به گراف تولیدی در جدول ۳-۲۲ آمده‌است. این گراف ضریب خوشه‌بندی بزرگ‌تری از گراف ویکی وت را داراست. این گراف همچنین، قطر و میانگین کوتاه‌ترین مسیر مشابه داده واقعی را دارد، اما شعاع آن بزرگ‌تر از شعاع داده واقعی است. نمودارهای این مدل مشابه آنچیزی است که در مورد شبیه سازی گراف فیس بوک توسط DEB به آن اشاره شد. نمودرا توزیع درجات توانی، رابطه بین درجه و ضریب خوشه‌بندی معکوس و گراف وابستگی درجه به درجه با شرکت‌پذیری مثبت دارد. که البته نمودار میانگین درجات مشابه نمودار داده واقعی نیست (نمودارهای شکل ۲۰.۳(آ) ، ۲۰.۳(ب) و ۲۰.۳(ج)).

همانند آنچه پیش‌تر در مورد شبیه‌سازی گراف فیس بوک هم اشاره شد این مدل توانایی تولید قطب‌هایی، به قدرت قطب‌های گراف فیس بوک را ندارد، اما می‌تواند قطب‌های به نسبت قدرتمندی تولید کند. که حداکثر درجاتش به ۳۵۰ می‌رسد.



(آ) نمودار ضریب خوشه‌بندی نسبت به درجه راس



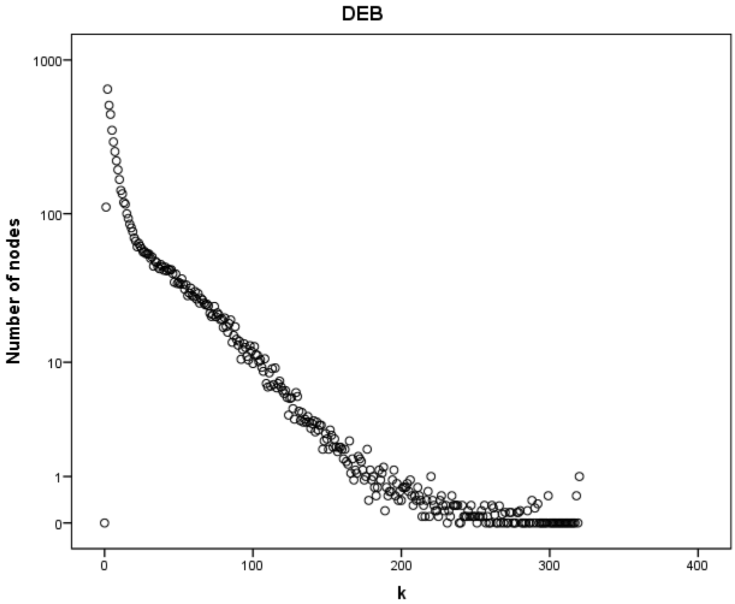
(ب) نمودار میانگین درجات همسایه‌ها نسبت به درجه راس

شکل ۱۹.۳: نمودار توزیع ضریب خوشه‌بندی و میانگین درجات همسایه

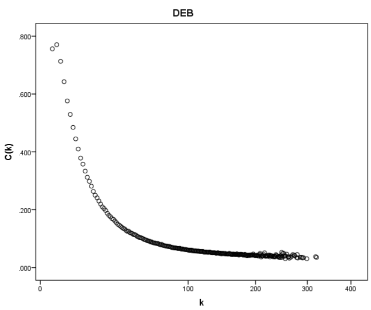
۱۴.۳ تحلیل مدل Vaz

در مدل Vaz احتمال افزوده شدن یال $۱ - u = ۰/۰۵۵$ و احتمال افزوده شدن یال در جستجوی محلی را برابر $u = ۰/۹۴۵$ قرار دادیم. ویژگی‌های گراف تولیدی در جدول ۳-۲۴ آمده‌است. این مدل ضریب خوشه‌بندی بزرگ‌تری دارد، شعاع، قطر و میانگین کوتاه‌ترین مسیرهای آن نیز همانند ضریب خوشه‌بندی از داده واقعی بیشتر است.

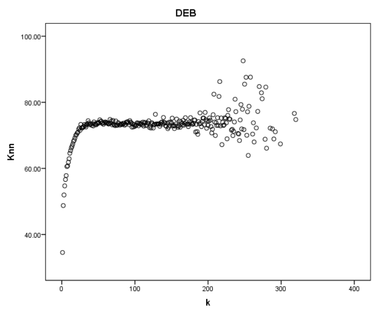
نمودارهای توزیع درجات، نسبت درجه با ضریب خوشه‌بندی و میانگین درجات همسایه‌ها در شکل‌های



(آ) نمودار توزیع درجات گراف DEB



(ب) نمودار ضریب خوشه‌بندی نسبت به درجه راس گراف DEB



(ج) نمودار میانگین درجات همسایه‌ها نسبت به درجه راس DEB

شکل ۲۰.۳: نمودارهای توزیع درجات، ضریب خوشه بندی و میانگین درجه همسایه‌های گراف DEB

۲۱.۳(آ)، ۲۱.۳(ب) و ۲۱.۳(ج) آمده است. که مشابه آنچیزی است که در مدلسازی فیس بوک نیز مطرح شد، یعنی نمودار توزیع درجات توانی، نسبت معکوس ضریب خوشه‌بندی به درجه، و وابستگی درجه به درجه با شرکت پذیری مثبت از ویژگی‌های گراف تولید شده توسط این مدل می‌باشد. این مدل نیز قطب‌های نسبتا قدرتمندی تولید می‌کند که البته ضعیف‌تر از قطب‌های داده واقعی می‌باشد.

جدول ۳-۲۳: جدول مربوط به تعداد و درجه راس‌های قطب گراف DEB

ترتیب	۱	۲	۳	۴	۵	۶	۷	۸	۹	۱۰
درجه	۲۸۴	۲۶۵.۲	۲۵۴.۹	۲۴۶.۳	۲۳۹.۴	۲۳۴.۹	۲۲۹.۹	۲۲۵.۷	۲۲۳.۷	۲۱۷.۴

جدول ۳-۲۴: اطلاعات مربوط به گراف تولید شده توسط مدل Vaz

نام گراف	تعداد راس‌ها	تعداد یال‌ها	میانگین درجه گراف	ضریب خوشه بندی	شعاع	قطر	میانگین کوتاه‌ترین مسیر
Vaz	۷۱۱۵	۱۰۲۷۰۵	۲۸.۷۸	۰.۳۲	۵.۴	۱۰.۴	۴.۰۲

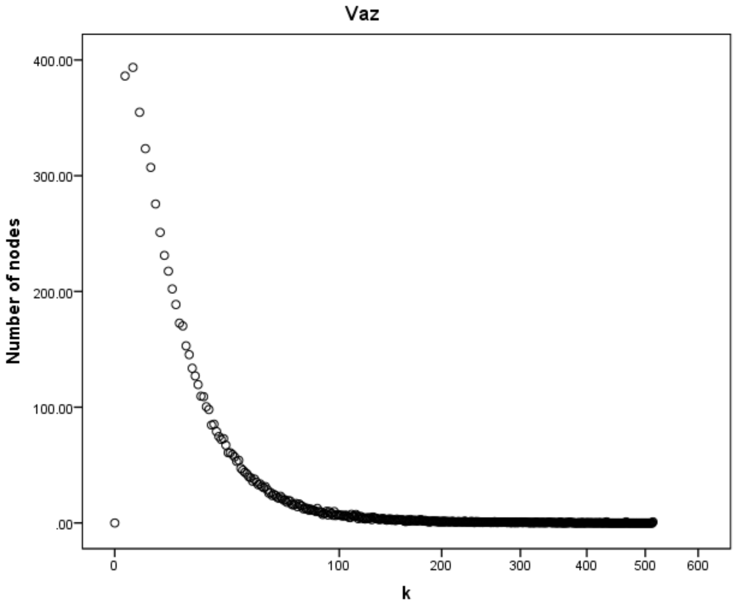
۱۵.۳ تحلیل مدل MVS

برای تولید گراف در این مدل جستجوی سراسری را $p_g = ۰/۰۰۰۱۵$ ، جستجوی محلی را $p_l = ۰/۰۶۶۴$ و احتمال حذف را نیز ۰.۰۰۴۳ در نظر گرفتیم. ویژگی‌های این مدل در جدول ۳-۲۶ آمده‌است. در این مدل در حالت پایدار با بزرگ شدن درجه ضریب خوشه‌بندی به ۰/۰۱ میل می‌کند. در نهایت در حالت پایدار ضریب خوشه بندی برای این گراف برابر ۰/۰۴۲ می‌شود، که از ضریب خوشه‌بندی گراف داده واقعی کمتر است. شعاع گراف حاصل از این مدل از شعاع گراف ویکی وت بیشتر اما قطر آن از گراف داده واقعی کمتر است. این مدل میانگین کوتاه‌ترین مسیر کوچک‌نری نسبت به داده واقعی تولید می‌کند.

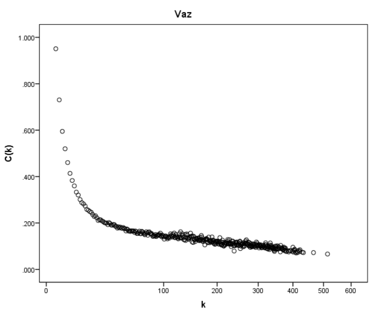
نمودارهای توزیع درجات، ضریب خوشه‌بندی و میانگین درجات همسایه‌ها به ترتیب در شکل‌های ۲۲.۳(آ)، ۲۲.۳(ب) و ۲۲.۳(آ) آمده است. این مدل توزیع درجات پوآسون دارد، ضریب خوشه‌بندی آن همانند داده واقعی رابطه عکس با درجه دارد، و گراف حاصل از این مدل مانند داده واقعی دارای شرکت‌پذیری منفی می‌باشد. جدول ۳-۲۷ نشان می‌دهد که گراف حاصل از این مدل قطب‌های بسیار ضعیفی تولید می‌کند.

جدول ۳-۲۵: جدول مربوط به تعداد و درجه راس‌های قطب گراف Vaz

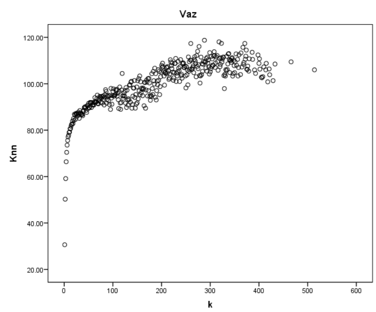
ترتیب	۱	۲	۳	۴	۵	۶	۷	۸	۹	۱۰
درجه	۳۹۶.۱	۳۷۹.۷	۳۶۹.۷	۳۵۹.۴	۳۵۷.۲	۳۵۲.۵	۳۴۷.۹	۳۴۲.۲	۳۳۸	۳۳۵.۳



(آ) نمودار توزیع درجات گراف Vaz



(ب) نمودار ضریب خوشه‌بندی نسبت به درجه راس گراف Vaz



(ج) نمودار میانگین درجات همسایه‌ها نسبت به درجه راس Vaz

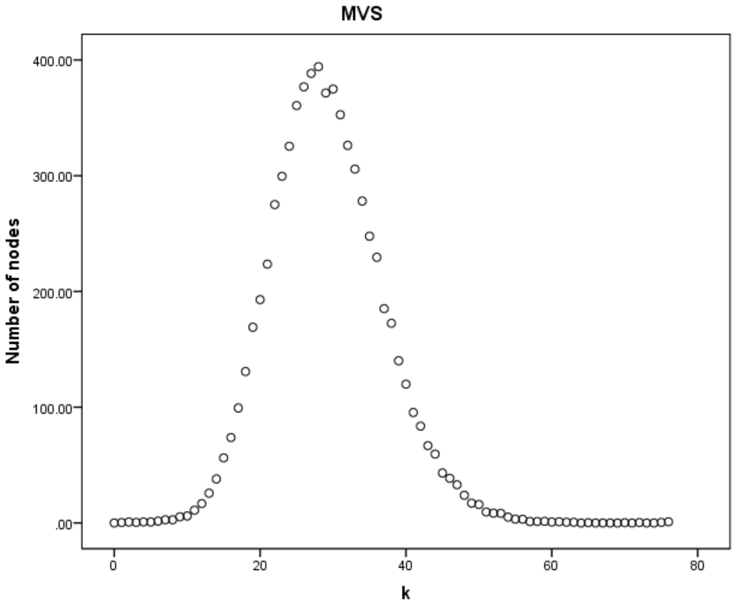
شکل ۲۱.۳: نمودارهای توزیع درجات، ضریب خوشه بندی و میانگین درجه همسایه‌های گراف Vaz

۱۶.۳ تحلیل مدل BPDA

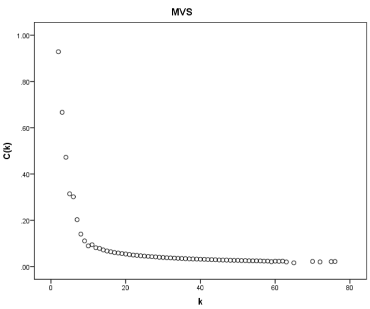
اگر برخی از مدل‌های برای تولید ضریب خوشه‌بندی بسیار بزرگ ناتوان بودند، این مدل در تولید ضریب خوشه‌بندی کوچک ناتوان است. با انتخاب $\alpha = ۱$ یعنی کوچکترین مقدار ممکن و قرار داد $\beta = ۳۲۰$ برای تنظیم میانگین و توزیع راس‌ها به صورت یکنواخت در فضای یک بعدی $[۱ \ldots ۷۰۰۰۰۰]$ گرافی تولید کردیم که دارای ضرب خوشه‌بندی ۰/۱۵۴ بود. این گراف همانند آنچه در مورد مدلسازی گراف فیس بوک اشاره شد شعاع، قطر و میانگین کوتاه‌ترین مسیر بسیار بزرگی تولید می‌کند. ویژگی‌های گراف این مدل در جدول ۳-۲۸ آمده است.

جدول ۳-۲۶: اطلاعات مربوط به گراف تولید شده توسط مدلMVS

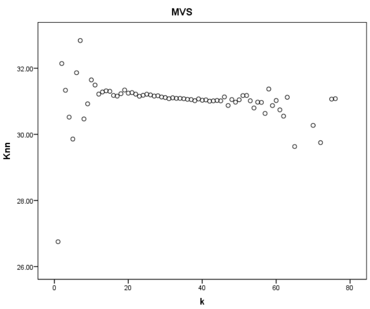
نام گراف	تعداد راس‌ها	تعداد یال‌ها	میانگین درجه گراف	ضریب خوشه بندی	شعاع	قطر	میانگین کوتاه‌ترین مسیر
MVS	۷۱۱۵	۱۰۴۰۲۱	۲۹.۲۴	۰.۰۴۲	۴	۴	۲.۹۴



(آ) نمودار توزیع درجات گراف MVS



(ب) نمودار ضریب خوشه‌بندی نسبت به درجه راس گراف MVS



(ج) نمودار میانگین درجات همسایه‌ها نسبت به درجه راس MVS

شکل ۲۲.۳: نمودارهای توزیع درجات، ضریب خوشه بندی و میانگین درجه همسایه‌های گراف MVS

این مدل دارای توزیع درجات پواسون می‌باشد (۲۳.۳(آ)) می‌باشد. در این گراف نسبت بین ضریب خوشه‌بندی و درجه راس ثابت می‌باشد (۲۳.۳(ب)). در نهایت این مدل همبستگی درجه به درجه با شرکت‌پذیری مثبت

جدول ۳-۲۷: جدول مربوط به تعداد و درجه راس‌های قطب گراف MVS

ترتیب	۱	۲	۳	۴	۵	۶	۷	۸	۹	۱۰
درجه	۶۶.۱۰	۶۱.۷	۵۸.۹	۵۷.۷	۵۶.۸	۵۶.۱	۵۵.۷	۵۵.۲	۵۴.۸	۵۴.۷

جدول ۳-۲۸: اطلاعات مربوط به گراف تولید شده توسط مدلBPDA

نام گراف	تعداد راس‌ها	تعداد یال‌ها	میانگین درجه گراف	ضریب خوشه بندی	شعاع	قطر	میانگین کوتاه‌ترین مسیر
BPDA	۷۱۱۵	۱۰۳۳۴۵	۲۹.۰۵	۰.۱۵	۱۴	۲۸	۲۹.۰۵

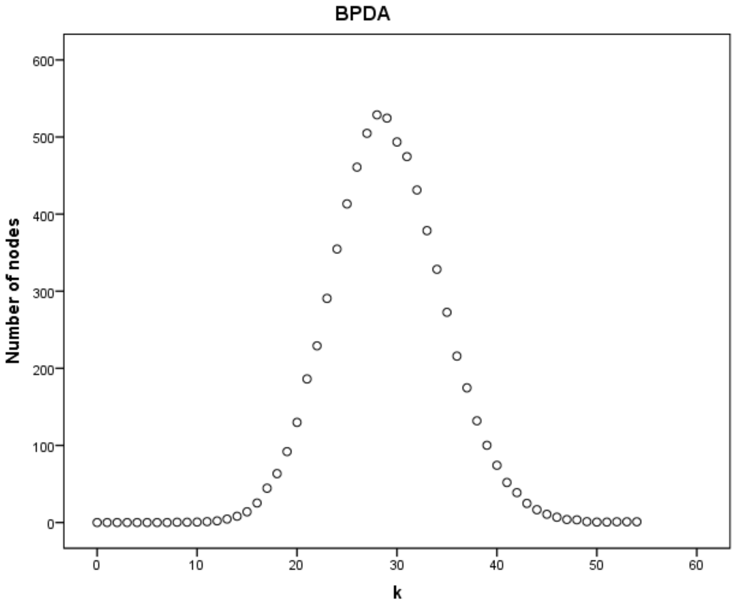
دارد(۲۳.۳(ج)). همچنین این مدل دارای قطب‌های با درجه بسیار کوچک می‌باشد.

۱۷.۳ تحلیل مدل TOSHK

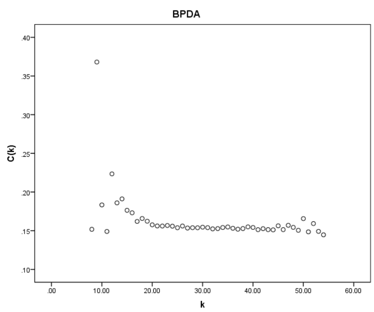
این مدل در حالت عادی یعنی حالتی که حداکثر دو راس را در جستجوی سراسری انتخاب می‌کرد، توانایی تولید گرافی با ضریب خوشه بندی کمتر از ۰/۱۷ را نداشت، اما تعریف این مدل انعطاف‌پذیری زیادی دارد. برای تولید مدلی با ضریب خوشه‌بندی کوچک جستجوهای سراسری را حداکثر به ۴ افزایش دادیم، یعنی با احتمال ۰/۰۱ یک راس، با احتمال ۰/۰۱ دو راس، با احتمال ۰/۷۵ سه راس و با احتمال ۰/۲۳ چهار راس را در جستجوی سراسری انتخاب می‌کنیم. برای جستجوی محلی نیز تعداد $U[۱, ۷]$ همسایه هر راس انتخاب شده در جستجوی سراسری انتخاب شده و یال‌هایی بین راس تولید شده و آن‌ها رسم می‌شود. نهایتا مدل ضریب خوشه‌بندی ۰/۱۴ را تولید کرد. جدول ۳-۳۰ حاوی ویژگی‌های گراف تولیدی توسط این مدل می‌باشد. شکل ۲۴.۳، نمودارهای توزیع درجات ۲۴.۳(آ)، ضریب خوشه‌بندی ۲۴.۳(ب) و میانگین درجات همسایه‌های راس ۲۴.۳(ج) را نشان می‌دهد. مدل دارای توزیع درجات توانی، نسبت معکوس ضریب خوشه‌بندی و شرکت‌پذیری

جدول ۳-۲۹: جدول مربوط به تعداد و درجه راس‌های قطب گراف BPDA

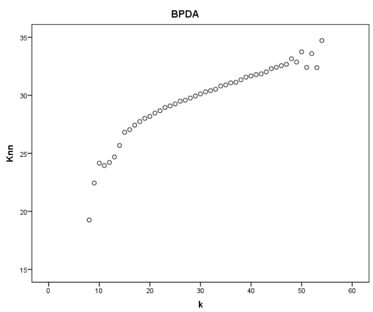
ترتیب	۱	۲	۳	۴	۵	۶	۷	۸	۹	۱۰
درجه	۵۲	۵۰.۵	۴۹.۶	۴۸.۴	۴۷.۸	۴۷.۶	۴۷.۲	۷۴.۱۵	۴۶.۹	۴۶.۶



(آ) نمودار توزیع درجات گراف BPDA



(ب) نمودار ضریب خوشه‌بندی نسبت به درجه راس گراف BPDA



(ج) نمودار میانگین درجات همسایه‌ها نسبت به درجه راس BPDA

شکل ۲۳.۳: نمودارهای توزیع درجات، ضریب خوشه بندی و میانگین درجه همسایه‌های گراف BPDA

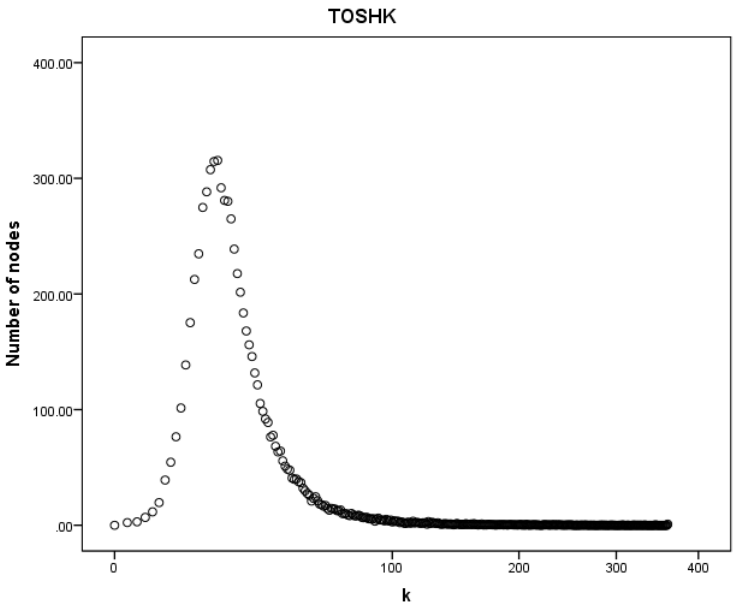
مثبت می‌باشد. این مدل نیز قطب‌های نسبتا قدرتمندی تولید می‌کند (جدول ۳-۳۱) که البته ضعیف‌تر از قطب‌های داده واقعی می‌باشد.

۱۸.۳ تحلیل مدل WPR

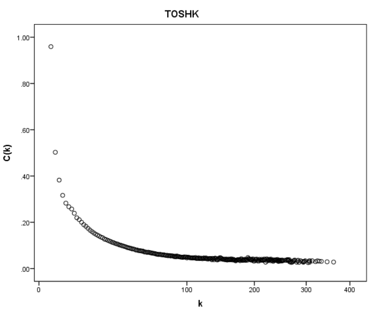
در این مدل راس‌ها را در یک فضای دو بعدی توسط توزیع پوآسون با میانگین ۷٫۵ توزیع کردیم. سپس برای شعاع $H = ۲$ بقیه متغییرها را به صورت مقابل مقداردهی نمودیم: $p = ۰٫۰۶۸۵$ ، $p_b = ۰٫۰۸۲۱۹$ و

جدول ۳-۳۰: اطلاعات مربوط به گراف تولید شده توسط مدلTOSHK

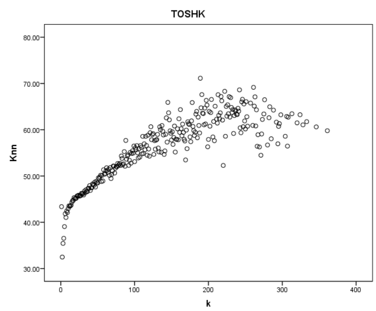
نام گراف	تعداد راس‌ها	تعداد یال‌ها	میانگین درجه گراف	ضریب خوشه بندی	شعاع	قطر	میانگین کوتاه‌ترین مسیر
TOSHK	۷۱۱۵	۱۰۱۳۱۷	۲۸.۸۴	۰.۱۴	۳	۵.۳	۲.۹۵



(آ) نمودار توزیع درجات گراف TOSHK



(ب) نمودار ضریب خوشه‌بندی نسبت به درجه راس گراف TOSHK



(ج) نمودار میانگین درجات همسایه‌ها نسبت به درجه راس TOSHK

شکل ۲۴.۳: نمودارهای توزیع درجات، ضریب خوشه بندی و میانگین درجه همسایه‌های گراف TOSHK

$p_{\Delta} = ۰/۰۶۸$. نهایتا گراف تولید شده دارای ویژگی‌های موجود در جدول ۳-۳۲ می‌باشد. همانطور که در شکل ۲۵.۳(آ) آمده این مدل دارای توزیع درجات پوآسون می‌باشد نسبت مستقیم بین ضریب خوشه‌بندی و درجه و شرکت‌پذیری مثبت از نمودارهای ۲۵.۳(ب) و ۲۵.۳(ج) قابل شناسایی می‌باشد. این مدل دارای قطب‌های با

جدول ۳-۳۱: جدول مربوط به تعداد و درجه راس‌های قطب گراف TOSHK

ترتیب	۱	۲	۳	۴	۵	۶	۷	۸	۹	۱۰
درجه	۳۱۶.۶	۲۹۵.۲	۲۷۹.۴	۲۶۷.۵	۲۵۸.۴	۲۴۹.۷	۲۴۳.۵	۲۳۹.۲	۲۳۶.۳	۲۳۲.۶

جدول ۳-۳۲: اطلاعات مربوط به گراف تولید شده توسط مدلWPR

نام گراف	تعداد راس‌ها	تعداد یال‌ها	میانگین درجه گراف	ضریب خوشه بندی	تشعاع	قطر	میانگین کوتاه‌ترین مسیر
WPR	۷۱۱۵	۱۰۲۱۰۰	۲۸.۷	۰.۱۴	۴	۵.۶	۳.۳۱

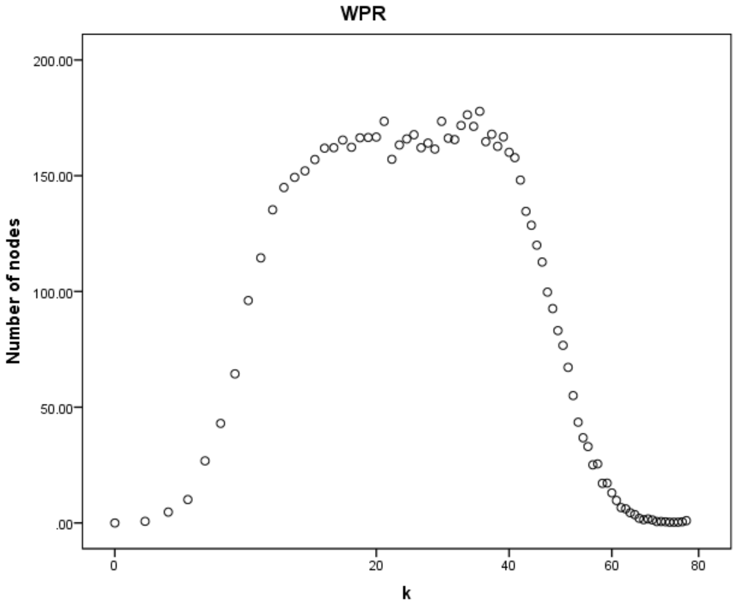
درجه بسیار پایین می‌باشد (جدول ۳-۳۳).

۱۹.۳ تحلیل مدل KOSKK

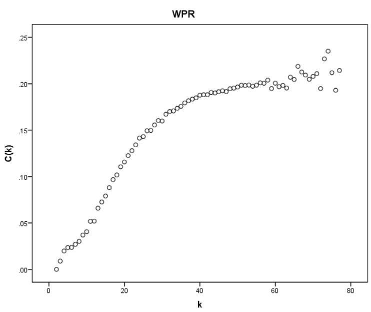
در این مدل جستجوی سراسری را برابر ۶٪۰ جستجوی محلی را برابر ۸۵٪۰ و $\delta = 1$، $W_0 = 1$ قرار دادیم. احتمال حذف نیز برابر ۹۰۰٪۰ قرار دادیم. ویژگی‌های این مدل در جدول ۳-۳۴ ذکر شده است. این مدل ضریب خوشه‌بندی، قطر و میانگین کوتاه‌ترین مسیر قابل قبولی را تولید می‌کند. نمودار شکل ۲۶.۳(آ) نشان دهنده توزیع توانی گراف این مدل می‌باشد. این مدل همچنین دارای رابطه عکس بین ضریب خوشه بندی و درجه، و شرکت‌پذیری مثبت می‌باشد (شکل ۲۶.۳(ب) و ۲۶.۳(ج)). گراف این مدل قطب‌های نسبتا قدرتمندی تولید می‌کند، که البته در مقایسه با قطب‌های داده واقعی ضعیف‌تر است.

جدول ۳-۳۳: جدول مربوط به تعداد و درجه راس‌های قطب گراف WPR

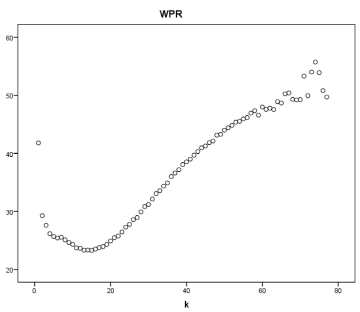
ترتیب	۱	۲	۳	۴	۵	۶	۷	۸	۹	۱۰
درجه	۷۱.۳	۶۸.۳	۶۷	۶۶.۷	۶۵.۹	۶۵.۵	۶۵.۱	۶۴.۶	۶۴.۵	۶۴.۱



(آ) نمودار توزیع درجات گراف WPR



(ب) نمودار ضریب خوشه‌بندی نسبت به درجه راس گراف WPR



(ج) نمودار میانگین درجات همسایه‌ها نسبت به درجه راس WPR

شکل ۲۵.۳: نمودارهای توزیع درجات، ضریب خوشه بندی و میانگین درجه همسایه‌های گراف WPR

۲۰.۳ جمع‌بندی تحلیل آماری مربوط به گراف ویکی وت

در جدول‌های ۳-۳۶ و ۳-۳۷ مقایسه‌ای کلی بین داده‌های واقعی و مدل‌های ارائه می‌دهد.

همانطور که از مقایسه مدل‌های ارائه شد (جدول ۳-۱۸ ، ۳-۱۹، ۳-۳۶ و ۳-۳۷) واضح است، با توجه به جداول مذکور (در جدول تطبیقی برای معیارهای شعاع و قطر بازه‌ی نوسان ±۱ برای ضریب خوشه‌بندی ±۰٫۰۵ و برای کوتاه‌ترین مسیر ±۰٫۶ را در نظر گرفته‌ایم) و انجمن‌های تشکیل شده، هنوز مدلی که بتوان آن را برای انطباق با یک شبکه اجتماعی ساختاربندی کرد، وجود ندارد. برخی مدل‌ها اصولا توانایی تولید مدلی با ضریب خوشه‌بندی مورد انتظار را ندارند، و برخی دیگر توانایی تولید انجمن‌های مشابه با داده واقعی را ندارند. بعضی از

جدول ۳-۳۴: اطلاعات مربوط به گراف تولید شده توسط مدلKOSKK

نام گراف	تعداد راس‌ها	تعداد یال‌ها	میانگین درجه گراف	ضریب خوشه بندی	شعاع	قطر	میانگین کوتاه‌ترین مسیر
KOSKK	۷۱۱۵	۱۰۴۵۹۰	۲۹.۴	۰.۱۴	۳.۴	۶.۳	۲.۹۴

جدول ۳-۳۵: جدول مربوط به تعداد و درجه راس‌های قطب گراف KOSKK

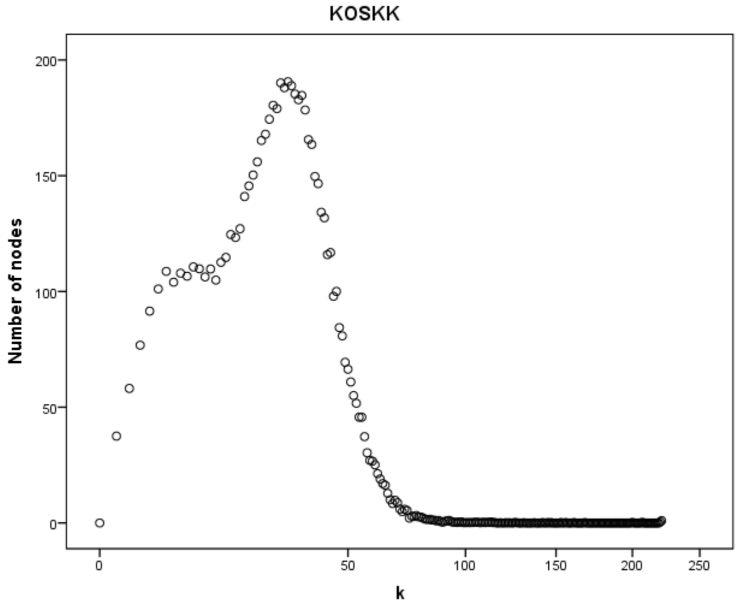
ترتیب	۱	۲	۳	۴	۵	۶	۷	۸	۹	۱۰
درجه	۱۷۰	۱۲۴.۲	۱۱۳	۱۰۸.۲	۱۰۴	۱۰۰.۱	۹۵	۹۲.۴	۹۱	۸۹.۸

جدول ۳-۳۶: جدول مقایسه مدل‌ها

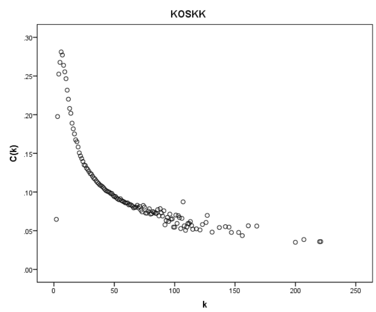
نام گراف	تعداد راس‌ها	تعداد یال‌ها	میانگین درجه گراف	ضریب خوشه بندی	شعاع	قطر	میانگین کوتاه‌ترین مسیر
wiki-Vote	۷۱۱۵	۱۰۳۶۸۹	۲۹.۱۴	۰.۱۴۱	۱	۷	۳.۲۴۸
DEB	۷۱۱۵	۱۰۲۷۷۶	۲۸.۸۹	۰.۳۶۸	۳.۸	۷.۴	۳.۱۵
Vaz	۷۱۱۵	۱۰۲۷۰۵	۲۸.۷۸	۰.۳۲	۵.۴	۱۰.۴	۴.۰۲
MVS	۷۱۱۵	۱۰۴۰۲۱	۲۹.۲۴	۰.۰۴۲	۴	۴	۲.۹۴
BPDA	۷۱۱۵	۱۰۳۳۴۵	۲۹و۰۵	۰.۱۵	۱۴	۲۸	۲۹.۰۵
TOSHK	۷۱۱۵	۱۰۱۳۱۷	۲۸.۸۴	۰.۱۴	۳	۵.۳	۲.۹۵
WPR	۷۱۱۵	۱۰۲۱۰۰	۲۸.۷	۰.۱۴	۴	۵.۶	۳.۳۱
KOSKK	۷۱۱۵	۱۰۴۵۹۰	۲۹.۴	۰.۱۴	۳.۴	۶.۳	۲.۹۴

جدول ۳-۳۷: جدول مقایسه تطبیقی مدل‌ها. برای معیارهای شعاع و قطر بازه‌ی نوسان ۱$\pm$ برای ضریب خوشه‌بندی ۰٫۰۵$\pm$ و برای کوتاه‌ترین مسیر ۰٫۶$\pm$ را در نظر گرفته شده است.

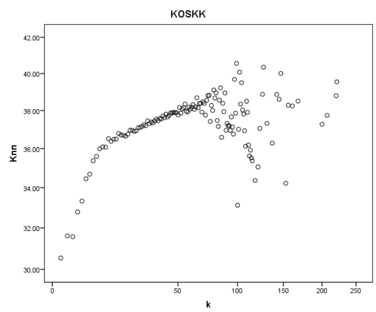
نام گراف	تعداد راس‌ها	تعداد یال‌ها	میانگین درجه گراف	ضریب خوشه بندی	شعاع	قطر	میانگین کوتاه‌ترین مسیر
wiki-Vote	۷۱۱۵	۱۰۳۶۸۹	۲۹.۱۴	۰.۱۴۱	۱	۷	۳.۲۴۸
DEB	$\checkmark$	$\checkmark$	$\checkmark$	$\times$	$\times$	$\checkmark$	$\checkmark$
Vaz	$\checkmark$	$\checkmark$	$\checkmark$	$\times$	$\times$	$\times$	$\times$
MVS	$\checkmark$	$\checkmark$	$\checkmark$	$\times$	$\times$	$\checkmark$	$\checkmark$
BPDA	$\checkmark$	$\checkmark$	$\checkmark$	$\times$	$\times$	$\times$	$\times$
TOSHK	$\checkmark$	$\checkmark$	$\checkmark$	$\checkmark$	$\times$	$\times$	$\checkmark$
WPR	$\checkmark$	$\checkmark$	$\checkmark$	$\checkmark$	$\times$	$\times$	$\checkmark$
KOSKK	$\checkmark$	$\checkmark$	$\checkmark$	$\checkmark$	$\times$	$\checkmark$	$\checkmark$



(آ) نمودار توزیع درجات گراف KOSKK



(ب) نمودار ضریب خوشه‌بندی نسبت به درجه راس گراف KOSKK



(ج) نمودار میانگین درجات همسایه‌ها نسبت به درجه راس KOSKK

شکل ۲۶.۳: نمودارهای توزیع درجات، ضریب خوشه بندی و میانگین درجه همسایه‌های گراف KOSKK

مدل‌ها قطب‌هایی با درجه کم تولید می‌کنند، برخی توزیع درجات پوآسون دارند که در تضاد با داده‌های واقعی است و برخی دیگر کوتاه‌ترین مسیر طولانی دارند. با در کنار هم قرار دادن این موارد نیاز به ارائه مدل‌های جدیدی احساس می‌شود.

می‌توان به منظور کنترل بیشتر بر روی قطب‌ها و انجمن‌های تولیدی و بر طبق قانون ۸۰-۲۰ که در مورد شبکه‌های اجتماعی می‌توان اینگونه در نظر گرفت که ۸۰ درصد محتویات را فقط ۲۰ درصد کاربران تولید می‌کنند (باید اشاره کرد که این قانون، قانونی تایید شده در حوزه شبکه‌های اجتماعی نیست، اما به لحاظ شهودی می‌توان این قانون را در شبکه‌های اجتماعی مشاهده کرد). ما به این اقلیتی که اکثریت محتویات را

تولید می‌کنند، کاربران پرسرو صدا می‌گوییم. اصولا این عده از کاربران در معرض روابط بیشتری قرار می‌گیرند. براساس همین ایده می‌توان با تغییراتی در مدل KOSKK به مدل جدیدی رسید که قطب‌های بیشتری می‌کند، علاوه بر پارامترهای موجود در مدل KOSKK، پارامتر N_n درصدی از جمعیت که پر سر و صدا هستند، D_n احتمال مرگ و میر اعضای پر سر و صدا و δ_n وزن پر سر و صداها در نظر می‌گیریم. با تنظیم این پارامترها می‌توان مدلی تولید کرد، تعداد و درجه قطب‌ها تا حدودی قابل کنترل باشد.
با توجه به اینکه می‌توان گفت قطب‌ها هسته‌های انجمن‌ها را تشکیل می‌دهند، به نظر می‌رسد این مدل انجمن‌های قابل کنترلی داشته باشد.

۲۱.۳ کارهای آینده

با توجه به عدم تولید ضریب خوشه‌بندی بسیار بزرگ در مدل‌های تکاملی و تولید ضریب خوشه‌بندی بزرگ برای مدل‌های مبتنی بر ویژگی‌های راس، به نظر می‌رسد، مشکل مدل‌های تکاملی در جستجوهای سراسری باشد که باعث کم شدن ضریب خوشه‌بندی می‌شود. اصولا نمی‌توان جستجوی سراسری را از مقدار مشخصی کمتر کرد، چون برای جستجوی محلی حداقل به یکی از این یال‌ها (حاصل جستجوی سراسری) نیاز است. اما در مدل‌های مبتنی بر ویژگی‌های راس می‌توان جستجوهای سراسری را نیز به ناحیه‌ای خاص محدود کرد.
به نظر می‌رسد با در هم آمیختن این دو نوع مدل و استفاده از ویژگی‌های مثبت هر دو به مدلی بهتر رسید. برای مثال یکی از کارهایی که در ادامه به آن خواهیم پرداخت، قرار دادن نقاط در فضا، استفاده از فاصله این نقاط برای جستجوی سراسری و استفاد از جستجوی دوست دوست به عنوان جستجوی محلی برای رسیدن به مدلی با ضریب خوشه‌بندی بالا می‌باشد.

واژه‌نامه فارسی به انگلیسی

ا

اتصال همسایگی . . . . . neighbor connectivity

استقلال از مقیاس . . . . . scale free

اشتراک‌پذیری . . . . . assortivity

اشتراک‌ناپذیر . . . . . disassortativity

الحاق امتیازی . . . . . preferential attachment

الگوریتم تبرید شبیه‌سازی شده . . . . . simulated annealing

انجمن . . . . . community

انجمن قوی . . . . . strong community

انجمن‌های ضعیف . . . . . weak community

انجمن‌های همپوشان . . . . . overlapping communities

ب

بازیگر . . . . . actor

برخط . . . . . online

برون‌خط . . . . . offline

بستار چرخه‌ای . . . . . cyclic clousre

بستار کانونی . . . . . focal clousre

بهینه‌سازی خارجی . . . . . extremal optimization

بهینه‌سازی شعاع طیفی .... spectral optimization

پ

پل .... bridge

پوشش .... cover

پیشنهاد .... suggestion

ت

تابع عملکرد .... performance function

تقابل کامل .... complete mutuality

تقسیم‌بندی .... partitioning

توزیع درجات .... Degree distribution

توزیع درجات گره‌های متصل .... Joint degree distribution

ج

جذب کننده سراسری .... global attractor

جذب کننده محلی .... local attractor

جستجوی سراسری .... global search

جستجوی محلی .... local search

ح

حذف راس .... node deletion

حذف یال .... edge deletion

حوزه میانی .... mean field

خ

خود متشابه .... self similar

خوشه‌بندی داده‌ها .... data clustering

د

درون سازگار .... self consistent

دسته .... clique

دستیابی‌پذیری .... reachability

ر

رابطه .... relation

راس .... node

رفتاری- نژادی .... inbreeding

ز

زمان گشت .... commute time

زنجیره مارکوف .... Markov Chain

ش

شعاع .... radius

ض

ضریب خوشه‌بندی .... clustering coefficient

ع

عابر تصادفی . . . . . random walker

علم داده . . . . . data scince

ف

فرضیه اتصالات ضعیف . . . . . weak links hypothesis

ق

قسمت . . . . . patition

قطر . . . . . diameter

گ

گراف اجتماعی . . . . . sociogram

گریز از مرکز . . . . . eccentricity

م

مبنایی . . . . . baseline

محبوبیت محور . . . . . celebrity deriven

مدل‌های محض . . . . . null models

مرکزیت ارتباطی . . . . . betweenness centrality

معادلات رشد . . . . . rate equation approach

معادلات سرآمد . . . . . master equation approach

معیارهای تناسب . . . . . fitnesses measure

مولفه عظیم . . . . . giant component

ن

نظریه پیوستگی . . . . continuum theory

ه

همپوشانی . . . . overlaping

ی

یال . . . . edge

فهرست اختصارات

A

AS . . . . Autonomous system

J

JDD . . . . Joint degree distribution

N

NAM . . . . Nodal attribute models

NEM . . . . Network evolution models

S

SCC . . . . Strongly connected component

SNA . . . . Social Network Analysis

W

WCC . . . . Weakly connected component

کتاب‌نامه

[1] Wasserman, S. (1994). Social network analysis: Methods and applications (Vol. 8). Cambridge university press.

[2] Moreno, J. L. (1953). Who shall survive? Foundations of sociometry, group psychotherapy and socio-drama.

[3] Freeman, L. C. (2004). The development of social network analysis: A study in the sociology of science (Vol. 1). Vancouver: Empirical Press.

[4] Mislove, E.A. (2009) Online Social Networks: Measurement, Analysis, and Applications to Distributed Information System (Huston, Texas)

[5] Pastor-Satorras, R., Vázquez, A., Vespignani, A. (2001). Dynamical and correlation properties of the Internet. Physical review letters, 87(25), 258701.

[6] Jinyang, L., Boon, T.L., Hellerstein, J., Kaashoek, F., Karger, D.R, Morris, R. (2003) On the feasibility of peer-to-peer web indexing and search.(In Proceedings of the 2nd International Workshop on Peer-to-Peer Systems (IPTPS'03, Berkeley, CA).

[7] Newman, M. E. (2002). Assortative mixing in networks. Physical review letters, 89(20), 208701.

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

[9] Fortunato, S. (2010). Community detection in graphs. Physics Reports, 486(3), 75-174.

[10] Barabási, A. L., Albert, R. (1999). Emergence of scaling in random networks. Science, 286(5439), 509-512.

[11] Dorogovtsev, S. N., Mendes, J. F. F., Samukhin, A. N. (2000). Structure of growing networks with preferential linking. Physical Review letters, 85(21), 4633.

[12] Krapivsky, P. L., Redner, S., Leyvraz, F. (2000). Connectivity of growing random networks. Physical review letters, 85(21), 4629.

[13] Toivonen, R., Kovanen, L., Kivelä, M., Onnela, J. P., Saramäki, J., Kaski, K. (2009). A comparative study of social network models: Network evolution models and nodal attribute models. Social Networks, 31(4), 240-254.

[14] Coleman, J. S. (1964). Introduction to mathematical sociology. London Free Press Glencoe.

[15] Freeman, L. C. (2004). The development of social network analysis: A study in the sociology of science (Vol. 1). Vancouver: Empirical Press.

[16] Kottak, C. P. (2004) Cultural Anthropology (McGraw-Hill, New York, USA).

[17] Moody, J., White, D. R. (2003). Structural cohesion and embeddedness: A hierarchical concept of social groups. American Sociological Review, 103-127.

[18] Weiss, R. S., Jacobson, E. (1955). A method for the analysis of the structure of complex organizations. American Sociological Review, 661-668.

[19] Palla, G., Derényi, I., Farkas, I., Vicsek, T. (2005). Uncovering the overlapping community structure of complex networks in nature and society. Nature, 435(7043), 814-818.

[20] Mancoridis, S., Mitchell, B. S., Rorres, C., Chen, Y., Gansner, E. R. (1998, June). Using automatic clustering to produce high-level system organizations of source code. In International Conference on Program Comprehension (pp. 45-45). IEEE Computer Society. Workshop on Program Comprehension (IEEE Computer Society, Washington, DC, USA).

[21] Luce, R. D., Perry, A. D. (1949). A method of matrix analysis of group structure. Psychometrika, 14(2), 95-116.

[22] Alba, R. D. (1973). A graph theoretic definition of a sociometric clique. Journal of Mathematical Sociology, 3(1), 113-126.

[23] Mokken, R. J. (1979). Cliques, clubs and clans. Quality Quantity, 13(2), 161-173.

[24] Seidman, S. B., Foster, B. L. (1978). A graph theoretic generalization of the clique concept. Journal of Mathematical sociology, 6(1), 139-154.

[25] Seidman, S. B. (1983). Network structure and minimum degree. Social networks, 5(3), 269-287.

[26] Matsuda, H., Ishihara, T., Hashimoto, A. (1999). Classifying molecular sequences using a linkage graph with their pairwise similarities. Theoretical Computer Science, 210(2), 305-325.

[27] Luccio, F., Sami, M. (1969). On the decomposition of networks in minimally interconnected subnetworks. Circuit Theory, IEEE Transactions on, 16(2), 184-188.

[28] Guimera, R., Sales-Pardo, M., Amaral, L. A. N. (2004). Modularity from fluctuations in random graphs and complex networks. Physical Review E, 70(2), 025101.

[29] Fortunato, S., Barthelemy, M. (2007). Resolution limit in community detection. Proceedings of the National Academy of Sciences, 104(1), 36-41.

[30] Girvan, M., Newman, M. E. (2002). Community structure in social and biological networks. Proceedings of the National Academy of Sciences, 99(12), 7821-7826.

[31] Newman, M. E. (2004). Fast algorithm for detecting community structure in networks. Physical review E, 69(6), 066133.

[32] Donath, W. E., Hoffman, A. J. (1973). Lower bounds for the partitioning of graphs. IBM Journal of Research and Development, 17(5), 420-425.

[33] Fiedler, M. (1973). Algebraic connectivity of graphs. Czechoslovak Mathematical Journal, 23(2), 298-305.

[34] Donetti, L., Munoz, M. A. (2004). Detecting network communities: a new systematic and efficient algorithm. Journal of Statistical Mechanics: Theory and Experiment, (10), P10012.

[35] Zhou, H. (2003). Distance, dissimilarity index, and network community structure. Physical review e, 67(6), 061901.

[36] Hughes, B. D. (1996). Random walks and random environments. Oxford: Clarendon Press.

[37] Hastings, M. B. (2006). Community detection as an inference problem. Physical Review E, 74(3), 035102.

[38] Raghavan, U. N., Albert, R., Kumara, S. (2007). Near linear time algorithm to detect community structures in large-scale networks. Physical Review E, 76(3), 036106.

[39] Bomze, I. M., Budinich, M., Pardalos, P. M., Pelillo, M. (1999). The maximum clique problem. In Handbook of combinatorial optimization (pp. 1-74). Springer US.

[40] Nepusz, T., Petróczi, A., Négyessy, L., Bazsó, F. (2008). Fuzzy communities and the concept of bridgeness in complex networks. Physical Review E, 77(1), 016107.

[41] Lancichinetti, A., Fortunato, S. (2009). Community detection algorithms: a comparative analysis. Physical review E, 80(5), 056117.

[42] Erdös, P., Rényi, A. (1961). On the strength of connectedness of a random graph. Acta Mathematica Hungarica, 12(1), 261-267.

[43] Newman, M. E., Strogatz, S. H., Watts, D. J. (2001). Random graphs with arbitrary degree distributions and their applications. Physical Review E, 64(2), 026118.

[44] Jeong, H., Néda, Z., Barabási, A. L. (2003). Measuring preferential attachment in evolving networks. EPL (Europhysics Letters), 61(4), 567.

[45] Watts, D. J. (1999). Small worlds: the dynamics of networks between order and randomness. Princeton university press.

[46] Davidsen, J., Ebel, H., Bornholdt, S. (2002). Emergence of a small world from local interactions: Modeling acquaintance networks. Physical Review Letters, 88(12), 128701.

[47] Barabási, A. L., Albert, R. (1999). Emergence of scaling in random networks. science, 286(5439), 509-512.

[48] Albert, R., Barabási, A. L. (2002). Statistical mechanics of complex networks. Reviews of modern physics, 74(1), 47.

[49] Newman, M. E., Watts, D. J., Strogatz, S. H. (2002). Random graph models of social networks. Proceedings of the National Academy of Sciences, 99(suppl 1), 2566-2572.

[50] Vázquez, A. (2003). Growing network with local rules: Preferential attachment, clustering hierarchy, and degree correlations. Physical Review E, 67(5), 056104.

[51] Marsili, M., Vega-Redondo, F., Slanina, F. (2004). The rise and fall of a networked society: A formal model. Proceedings of the National Academy of Sciences of the United States of America, 101(6), 1439-1442.

[52] Boguná, M., Pastor-Satorras, R., Diaz-Guilera, A., Arenas, A. (2003). Emergence of clustering, correlations, and communities in a social network model. arXiv preprint cond-mat/0309263.

[53] Watts, D. J., Dodds, P. S., Newman, M. E. (2002). Identity and search in social networks. science, 296(5571), 1302-1305.

[54] Boguná, M., Pastor-Satorras, R. (2003). Class of correlated random networks with hidden variables. Physical Review E, 68(3), 036112.

[55] Toivonen, R., Onnela, J. P., Saramäki, J., Hyvönen, J., Kaski, K. (2006). A model for social networks. Physica A: Statistical Mechanics and its Applications, 371(2), 851-860.

[56] Cohen, R., Erez, K., Ben-Avraham, D., Havlin, S. (2000). Resilience of the Internet to random breakdowns. Physical review letters, 85(21), 4626.

[57] Krapivsky, P. L., Redner, S. (2001). Organization of growing random networks. Physical Review E, 63(6), 066123.

[58] Evans, T. S., Saramäki, J. P. (2005). Scale-free networks from self-organization. Physical Review E, 72(2), 026138.

[59] Dorogovtsev, S. N., Mendes, J. F. F., Samukhin, A. N. (2000). Structure of growing networks with preferential linking. Physical Review letters, 85(21), 4633.

[60] Wong, L. H., Pattison, P., Robins, G. (2006). A spatial model for social networks. Physica A: Statistical Mechanics and its Applications, 360(1), 99-120.

[61] Frank, O., Strauss, D. (1986). Markov graphs. Journal of the American Statistical Association, 81(395), 832-842.

[62] McPherson, M., Smith-Lovin, L., Cook, J. M. (2001). Birds of a feather: Homophily in social networks. Annual review of sociology, 415-444.

[63] Barrat, A., Barthélemy, M., Vespignani, A. (2004). Weighted evolving networks: coupling topology and weight dynamics. Physical review letters, 92(22), 228701.

[64] Kumpula, J. M., Onnela, J. P., Saramäki, J., Kaski, K., Kertész, J. (2007). Emergence of communities in weighted networks. Physical review letters, 99(22), 228701.

[65] Granovetter, M. S. (1973). The strength of weak ties. American journal of sociology, 1360-1380.

Abstract

The manner of development and topological changing of social networks is one of the disputed issue in social network science. Several models for anticipation behavior of social network have been presented.

In 2009, a comparison based on last.fm and email social network have been done. Since those social networks are not so comprehensive, there are some defect in that comparison. This defect is not only limited in choosing imperfect data set, but also, defect appears in choosing data sets, which were selected with almost same characteristics, I mean, that data had small graph degree average and small clustering coefficient. With emerging dense social network like facebook, twitter and google plus a new review in result of that comparison seems necessary. Dense social network with high clustering coefficient clearly reveal that comparison defect.

In this research, after defining all the models, based on the online social networks ”facebook” and ”Wiki-Vote”, all the models have been implemented and compared. In analysis part, we show that almost all of models, are unable to simulate real social networks characteristics.

Yazd University

Faculty of Mathematics

Department of Computer Science

Thesis submitted

for the degree of Master of Science

Title:

Comparison of presented online social network models with focus on constructed communities

Supervisor:

Dr. Mahdieh Hasheminezhad

Advisor:

Dr. Ali Dolati

By:

Hanif Emamgholizadeh

October, 2014