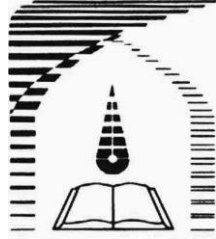


بِسْمِ اللَّهِ الرَّحْمَنِ الرَّحِيمِ



دانشگاه تربیت مدرس

دانشکده مهندسی برق و کامپیوتر

پایان نامه کارشناسی ارشد رشته مهندسی کامپیوتر گرایش نرم افزار

عنوان پایان نامه

کشف انجمن ها در شبکه های اجتماعی با استفاده از بهینه سازی کندوی زنبور عسل

نگارش

علیرضا قصبه

استاد راهنما:

دکتر محمد صنیعی آباده

استاد مشاور:

دکتر مهدی آبادی

بهمن ۱۳۹۳



تاییدیه اعضای هیات داوران حاضر در جلسه دفاع از پایان نامه کارشناسی ارشد

آقای علیرضا قصبه پایان نامه ۶ واحدی خود را با عنوان کشف انجمن ها در شبکه های اجتماعی با استفاده از بهینه سازی کندوی زنبور عسل در تاریخ ۱۳۹۳/۱۱/۱۸ ارائه کردند.

اعضای هیات داوران نسخه نهایی این پایان نامه را از نظر فرم و محتوا تایید کرده، پذیرش آنرا برای اخذ درجه کارشناسی ارشد مهندسی کامپیوتر نرم افزار پیشنهاد می کنند.

عضو هیات داوران	نام و نام خانوادگی	رتبه علمی	امضا
استاد راهنما	دکتر محمد صنیعی آباده	استادیار	
استاد مشاور	دکتر مهدی آبادی	استادیار	
استاد ناظر	دکتر نصراله مقدم چرکری	دانشیار	
استاد ناظر	دکتر آزاده شاکری	استادیار	
مدیر گروه (یا نماینده گروه تخصصی)	دکتر نصراله مقدم چرکری	دانشیار	

دستورالعمل حق مالکیت مادی و معنوی در مورد نتایج پژوهشهای علمی دانشگاه تربیت مدرس

مقدمه: با عنایت به سیاست‌های پژوهشی دانشگاه در راستای تحقق عدالت و کرامت انسانها که لازمه شکوفایی علمی و فنی است و رعایت حقوق مادی و معنوی دانشگاه و پژوهشگران، لازم است اعضای هیات علمی، دانشجویان، دانش‌آموختگان و دیگر همکاران طرح، در مورد نتایج پژوهشهای علمی که تحت عناوین پایان‌نامه، رساله و طرحهای تحقیقاتی که با هماهنگی دانشگاه انجام شده است، موارد ذیل را رعایت نمایند:

ماده ۱- حقوق مادی و معنوی پایان‌نامه‌ها / رساله‌های مصوب دانشگاه متعلق به دانشگاه است و هرگونه بهره‌برداری از آن باید با ذکر نام دانشگاه و رعایت آیین‌نامه‌ها و دستورالعمل‌های مصوب دانشگاه باشد.


ماده ۲- انتشار مقاله یا مقالات مستخرج از پایان‌نامه / رساله به صورت چاپ در نشریات علمی و یا ارائه در مجامع علمی باید به نام دانشگاه بوده و استاد راهنما مسئول مکاتبات مقاله باشد. تبصره: در مقالاتی که پس از دانش‌آموختگی بصورت ترکیبی از اطلاعات جدید و نتایج حاصل از پایان‌نامه / رساله نیز منتشر می‌شود نیز باید نام دانشگاه درج شود.

ماده ۳- انتشار کتاب حاصل از نتایج پایان‌نامه / رساله و تمامی طرحهای تحقیقاتی دانشگاه باید با مجوز کتبی صادره از طریق حوزه پژوهشی دانشگاه و بر اساس آیین‌نامه‌های مصوب انجام می‌شود.

ماده ۴- ثبت اختراع و تدوین دانش فنی و یا ارائه در جشنواره‌های ملی، منطقه‌ای و بین‌المللی که حاصل نتایج مستخرج از پایان‌نامه / رساله و تمامی طرحهای تحقیقاتی دانشگاه باید با هماهنگی استاد راهنما یا مجری طرح از طریق حوزه پژوهشی دانشگاه انجام گیرد.

ماده ۵- این دستورالعمل در ۵ ماده و یک تبصره در تاریخ ۱۳۸۴/۴/۲۵ در شورای پژوهشی دانشگاه به تصویب رسیده و از تاریخ تصویب لازم‌الاجرا است و هرگونه تخلف از مفاد این دستورالعمل، از طریق مراجع قانونی قابل پیگیری می‌شود.

نام و نام خانوادگی *علیرضا قصبه*
امضاء


۹۳، ۱۱، ۲۰

آیین نامه چاپ پایان نامه (رساله) های دانشجویان دانشگاه تربیت مدرس

نظر به اینکه چاپ و انتشار پایان نامه (رساله) های تحصیلی دانشجویان دانشگاه تربیت مدرس، مبین بخشی از فعالیتهای علمی - پژوهشی دانشگاه است بنابراین به منظور آگاهی و رعایت حقوق دانشگاه، دانش آموختگان این دانشگاه نسبت به رعایت موارد ذیل متعهد می شوند:

ماده ۱: در صورت اقدام به چاپ پایان نامه (رساله) ی خود، مراتب را قبلاً به طور کتبی به «دفتر نشر آثار علمی» دانشگاه اطلاع دهد.

ماده ۲: در صفحه سوم کتاب (پس از برگ شناسنامه) عبارت ذیل را چاپ کند:

«کتاب حاضر، حاصل پایان نامه کارشناسی ارشد نگارنده در رشته مهندسی کامپیوتر گرایش نرم افزار است که در سال ۱۳۹۳ در دانشکده مهندسی برق و کامپیوتر دانشگاه تربیت مدرس به راهنمایی جناب آقای دکتر محمد صنیعی آباده، مشاوره جناب آقای دکتر مهدی آبادی از آن دفاع شده است.»

ماده ۳: به منظور جبران بخشی از هزینه های انتشارات دانشگاه، تعداد یک درصد شمارگان کتاب (در هر نوبت چاپ) را به «دفتر نشر آثار علمی» دانشگاه اهدا کند. دانشگاه می تواند مازاد نیاز خود را به نفع مرکز نشر در معرض فروش قرار دهد.

ماده ۴: در صورت عدم رعایت ماده ۳، ۵۰٪ بهای شمارگان چاپ شده رابه عنوان خسارت به دانشگاه تربیت مدرس، تأدیه کند.

ماده ۵: دانشجو تعهد و قبول می کند در صورت خودداری از پرداخت بهای خسارت، دانشگاه می تواند خسارت مذکور را از طریق مراجع قضایی مطالبه و وصول کند؛ به علاوه به دانشگاه حق می دهد به منظور استیفای حقوق خود، از طریق دادگاه، معادل وجه مذکور در ماده ۴ را از محل توقیف کتابهای عرضه شده نگارنده برای فروش، تامین نماید.

ماده ۶: اینجانب علیرضا قصبه دانشجوی رشته مهندسی کامپیوتر گرایش نرم افزار مقطع کارشناسی ارشد تعهد فوق و ضمانت اجرایی آن را قبول کرده، به آن ملتزم می شوم.

نام و نام خانوادگی: علیرضا قصبه

تاریخ و امضا: ۱۳۹۳/۱۱/۲۰

تقدیم به

پدر بزرگوارم و مادر مهربانم

به پاس حمایت های بی دریغشان

به پاس فداکاری و گذشت بی پمانشان

به پاس سپید کشتن مویشان برای سپید کشتن رویان

به پاس قلب های بزرگشان و صبر و بردباری فراوانشان

به پاس عمر کرانقدر و پر بهایشان که به پای امنیت و آسایش و آرامش و کمال ما گذشت

پیشکش قطره ای در مقابل دریایی

سپاسگزاری:

خدای مهربانم با تعبیری از خودت شروع می‌کنم

اگر تمام درختان روی زمین قلم شوند

اگر تمام دریاها مرکب شوند و پس از آن هفت دریای دیگر به وجود آید

اگر تک تک لحظات عمرم را صرف کنم

باز هم نمی‌توانم شکر یکی از نعمت‌های تو را به جا آورم.

صمیمانه‌ترین تشکرها و قدردانی‌ها را تقدیم به جناب آقای دکتر محمد صنیعی، استاد راهنمای بزرگوارم می‌کنم. و از ایشان برای حسن خلقشان، برای راهنمایی‌های بی دریغشان، برای صرف وقت فراوانشان و امیدواری‌های دگرگون‌کننده‌شان نهایت سپاسگزاری را دارم.

از جناب آقای دکتر مهدی آبادی، استاد مشاور عزیزم به خاطر راهنمایی‌های ارزنده‌شان و دگرگونی‌هایشان کمال تشکر و قدردانی را دارم.

از همه‌ی کسانی که در طول زندگی از معلمین و اساتید بزرگوارم، از دوستان همیشه‌بمراهم، از خویشان مهربانم و از تک تک کسانی که حتی کوچکترین کاری را برای من انجام داده‌اند، تا من در این لحظه توفیق رسیدن به این مرحله از زندگی را داشته باشم، بی نهایت سپاسگزارم.

چکیده

شبکه‌های اجتماعی در واقع نوعی ساختار اجتماعی محسوب می‌شوند که از گره‌های متعددی تشکیل شده‌اند. این گره‌ها می‌تواند افراد حقیقی یا سازمان‌ها باشند. این گره‌ها توسط یک یا چند نوع خاص وابستگی مانند روابط دوستانه، خویشاوندی، همشهری بودن، روابط تجاری، روابط علمی، به یکدیگر متصل هستند. با گسترده شدن اینترنت و وب و همچنین افزایش چشمگیر گوشیهای هوشمند در سال‌های اخیر، شبکه‌های اجتماعی به یکی از ارکان جدایی ناپذیر زندگی ما، حداقل در بعد مجازی تبدیل شده‌اند. تحلیل شبکه‌های اجتماعی در واقع بررسی رابطه میان افراد، شامل بررسی ساختارهای اجتماعی، موقعیت اجتماعی، تحلیل نقش‌ها و موارد دیگر است. یکی از مهمترین ساختارهای اجتماعی انجمن‌ها هستند. انجمن به گروهی از رؤس گفته می‌شود که پیوندهای درونی افراد، بیشتر از پیوندهای آنها با خارج از گروه است.

کشف انجمن‌ها یکی از زمینه‌های تحقیقاتی است که محققان زیادی را در سالهای اخیر به خود جلب کرده است. پاسخ‌های بسیاری برای این مسئله ارائه شده است ولی هنوز این مسئله به طور رضایتمندانه‌ای حل نشده است. در این پایان‌نامه از ترکیب ایده‌ی خوشه‌بندی کلونی مورچه با بهینه‌سازی کندوی زنبور عسل استفاده شده است. خوشه‌بندی کلونی مورچه که یک جستجوی محلی است، توسط بهینه‌سازی کندوی زنبور عسل که یک راهکار سراسری است هدایت می‌شود. همچنین مدلی برای تخصیص زنبورهای رقصنده پیشنهاد شده است. راهکارهای پیشنهادی در کنار هم، باعث شده‌اند تا انجمن‌ها به طور دقیق‌تر و سریع‌تری کشف شوند. در واقع از زنبورهای رقصنده برای رد و بدل کردن اطلاعات میان گره‌ها، که گره در واقع همان مورچه در الگوریتم خوشه‌بندی کلونی مورچه است، استفاده شده است. نتایج بر روی شبکه‌های واقعی و گراف‌های مصنوعی تولید شده توسط کامپیوتر کارایی روش پیشنهادی را به خوبی نشان می‌دهد، زیرا که روش پیشنهادی توانست مصالحه‌ی خوبی میان دقت الگوریتم و نیز پیچیدگی زمانی برقرار نماید. همچنین الگوریتم بر روی مجموعه داده دلفین‌ها به دقت ۱۰۰ درصد دست پیدا کرد. و نیز در آزمون جی ان و لانیچیتی، زمانیکه پارامتر آمیزش از ۰,۵ گذر می‌کند، به دقت‌های بهتری نسبت به سایرین دست پیدا کرده است. همچنین در این پایان‌نامه نشان داده شد که هرچه ابعاد مجموعه داده‌ها بالاتر برود، الگوریتم دقت‌های بهتری را کسب می‌کند.

کلید واژه: شبکه‌های اجتماعی، کشف انجمن‌ها، خوشه‌بندی کلونی مورچه، بهینه‌سازی کندوی زنبور عسل،

گراف کاوی

فهرست مطالب

د	فهرست شکل‌ها	۵
و	فهرست جدول‌ها	۵
۱	فصل ۱- کلیات	۱
۱-۱	پیشگفتار	۲
۲-۱	تعریف مسئله	۴
۳-۱	پیچیدگی محاسبات	۷
۴-۱	اهداف پایان نامه و نوآوری‌ها	۹
۵-۱	ساختار پایان نامه	۱۰
۱۱	فصل ۲- مفاهیم پایه	۱۱
۱-۲	داده کاوی	۱۲
۲-۲	روش‌های توصیفی یا بدون نظارت	۱۵
۳-۲	خوشه‌بندی	۱۵
۴-۲	الگوریتم‌های فرامکاشفه‌ای زیستی	۱۸
۱-۴-۲	دورنمای برازش	۱۹
۲-۴-۲	قابلیت‌های پوشش و ارتفاع	۲۱
۳-۴-۲	طبقه‌بندی الگوریتم‌های فرامکاشفه‌ای	۲۳
۵-۲	الگوریتم‌های مبتنی بر هوش جمعی	۲۴
۱-۵-۲	روش‌های تقلید-محور	۲۵
۲-۵-۲	روش‌های علامت محور	۲۶
۶-۲	بهینه‌سازی کندوی زنبور عسل	۲۹
۷-۲	الگوهای گرافی در شبکه‌های واقعی	۳۳
۱-۷-۲	شبکه‌های مقیاس آزاد:	۳۳
۲-۷-۲	پدیده جهان کوچک	۳۴
۳-۷-۲	انجمن‌ها	۳۵
۸-۲	شبکه‌های اجتماعی	۴۱
۹-۲	تاریخچه کوتاهی از شبکه‌های اجتماعی بر خط	۴۲
۴۶	فصل ۳- پژوهش‌های پیشین	۴۶
۱-۳	مقدمه	۴۷
۲-۳	روش‌های سنتی	۴۷
۱-۲-۳	جزء بندی گراف	۴۷
۲-۲-۳	خوشه‌بندی سلسله مراتبی	۴۹

۵۲	خوشه‌بندی طیفی	۳-۲-۳
۵۳	الگوریتم‌های تقسیمی	۳-۳
۵۵	روشهایی بر پایه پودمانگی	۴-۳
۵۶	بهینه‌سازی پودمانگی	۱-۴-۳
۶۱	روش‌هایی مبتنی بر مکاشفه‌های زیستی	۵-۳
۶۴	مروری بر روش خوشه‌بندی کلونی مورچه	۶-۳
۶۴	جمع بندی	۷-۳
۶۶	روش پیشنهادی	۴
۶۷	توصیف کلی	۱-۴
۷۱	توصیف جزئیات الگوریتم به تفکیک	۲-۴
۷۱	محل زندگی	۱-۲-۴
۷۲	درک محل زندگی	۲-۲-۴
۷۷	پارامتر آستانه حرکت و تنظیم کردن سازگارانه‌ی آن	۳-۲-۴
۷۸	استراتژی حرکت	۴-۲-۴
۸۰	نحوه بروزسانی زنبورهای رقصنده و جدول رقص	۵-۲-۴
۸۵	پیچیدگی زمانی الگوریتم	۳-۴
۸۷	جمع بندی	۴-۴
۸۸	نتایج و ارزیابی روش پیشنهادی	۵
۸۹	مقدمه	۱-۵
۸۹	مجموعه داده‌ها	۲-۵
۸۹	شبکه‌های واقعی	۱-۲-۵
۹۲	شبکه‌های مصنوعی یا تولیدشده توسط کامپیوتر	۲-۲-۵
۹۴	زبان مدل کردن گراف	۳-۲-۵
۹۶	معیارهای ارزیابی	۳-۵
۹۷	تنظیم پارامترها	۴-۵
۹۸	بررسی تاثیر پارامتر آستانه حرکت بر الگوریتم	۱-۴-۵
۱۰۱	ارزیابی و مقایسه روش پیشنهادی با روش‌های دیگر	۵-۵
۱۰۱	مقایسه میان شبکه‌های واقعی	۱-۵-۵
۱۰۴	مقایسه با شبکه‌های تولید شده توسط کامپیوتر	۲-۵-۵
۱۰۹	کارآیی الگوریتم در ابعاد بالا	۶-۵
۱۱۰	نتیجه گیری و پیشنهادات	۶
۱۱۱	نتیجه گیری	۱-۶

۱۱۲.....کارهای آتی ۲-۶-

۱۱۴.....فهرست مراجع

۱۱۹.....واژه نامه‌ی فارسی به انگلیسی

فهرست شکل‌ها

صفحه

عنوان

۷	۱-۱ یک گراف ساده با سه انجمن که خط چین از هم جدا شده است.....
۱۲	۱-۲ سلسله مراتب داده تا خرد.....
۱۴	۲-۲ مراحل مختلف داده کاوی.....
۲۰	۳-۲ نمونه از یک دورنمای برازش یک بعدی که در آن هدف یافتن بلندترین قله است.....
۲۱	۴-۲ نمونه‌ای از دورنمای برازش در فضای سه بعدی.....
۲۸	۵-۲ روند کلی یک الگوریتم علامت-محور.....
۳۲	۶-۲ شبه کد ساده الگوریتم بهینه‌سازی کندوی زنبور عسل [۲].....
۳۴	۷-۲ نمایش توزیع‌های نرمال و قانون توانی [۳].....
۴۲	۸-۲ شبکه اجتماعی متشکل از افراد و ارتباطات آنها.....
۴۴	۹-۲ رشد قارچ گونه شبکه‌های اجتماعی برخط.....
۴۷	۱-۳ خط چین نمایش راه‌حل مسئله دوبخشی کمینه برای گراف ذکر شده است [۱].....
۵۱	۲-۳ یک دندروگرام یا درخت سلسله مراتبی. برش افقی یک جزء بندی را که یک معیار خاص را ارضا می‌نماید نشان می‌دهد [۶].....
۷۰	۱-۴ فلوچارت کلی الگوریتم در یک نگاه.....
۷۲	۲-۴ محل زندگی مورچه یا ماتریس باید به صورت کروی در نظر گرفته شود.....
۷۳	۳-۴ نمونه‌ای از یک مورچه در حال پردازش و همسایگی آن. در این مثال محدوده دید افقی و محدوده دید عمودی هر دو یک هستند.....
۷۶	۴-۴ اگر فاصله میان همه مورچه‌ها برابر صفر باشد، در این صورت برازش مورچه سیاه در هر دو حالت برابر یک خواهد بود.....
۷۹	۵-۴ طرح کلی حرکت مارپیچی در محیط زندگی یا همان ماتریس.....
۸۴	۶-۴ اعداد، درجه یا مرحله‌ی همسایگی مکان‌های همسایه را نسبت به گره i نشان می‌دهند.....
۸۶	۷-۴ شبه کد ساده الگوریتم.....
۹۰	۱-۵ شبکه باشگاه کاراته زاخاری. مربع یک انجمن و دایره هم انجمن دیگر را نمایش می‌دهد.....
۹۱	۲-۵ شبکه‌های دلفین‌های دست آموز، دایره نشان دهنده یک انجمن و مربع نشان دهنده انجمن دیگر است.....
۹۳	۳-۵ یک نمونه از آزمون لانیچینتی با ۵۰۰ گره [۵].....
۹۵	۴-۵ مثالی ساده از یک گراف در قالب <i>GML</i>
۹۹	۵-۵ نمایش تاثیر پارامتر آستانه حرکت در شبکه‌های واقعی.....
۱۰۰	۶-۵ اثر پارامتر آستانه حرکت بر روی شبکه‌های تولید شده توسط کامپیوتر.....
۱۰۴	۷-۵ نتایج آزمایشات بر روی آزمون GN. شکل الف از مرجع [۴] و شکل ب راست از [۷] و چپ [۸].....
۱۰۵	۸-۵ نتایج بدست آمده از الگوریتم پیشنهادی بر روی آزمون GN.....

- ۹-۵ نتایج آزمایشات الگوریتم‌های مختلف بر روی آزمون LFR با سایز 1000S, 1000B, 5000S, 5000B [۴]
۱۰۶.....
- ۱۰-۵ نتایج آزمایش الگوریتم CDHHO بر روی شبکه‌های تولید شده توسط روش LFR.....۱۰۷
- ۱۱-۵ مقایسه بین دقت‌های کسب شده برای ۵ شبکه با سایز متفاوت و پارامتر آمیزش متفاوت.....۱۰۹

فهرست جدول‌ها

صفحه

عنوان

۹۸.....	۱-۵ جدول پارامترهای الگوریتم در یک نگاه.....
۱۰۰.....	۲-۵ پارامترهای تنظیم شده برای الگوریتم.....
۱۰۱.....	۳-۵ مقایسه روش‌های مختلف با CDHHO بر روی مجموعه داده باشگاه کاراته.....
۱۰۲.....	۴-۵ مقایسه روش‌های مختلف با CDHHO بر روی مجموعه داده دلفین‌های دست آموز.....
۱۰۳.....	۵-۵ مقایسه روش‌های مختلف با CDHHO بر روی مجموعه داده فوتبال باشگاه‌های آمریکا.....
۱۰۸.....	۶-۵ مقایسه پیچیدگی زمانی الگوریتم‌هایی در آزمون LFR روی چهار شبکه شرکت داشته اند.....

فصل ۱- کلیات

۱-۱- پیشگفتار

علم شبکه‌ها، پیشرفت‌های قابل توجهی را در فهم ما از شبکه‌های پیچیده^۱ به ارمغان آورده است. یکی از پرمعناترین خاصیت‌های نمایش گرافی سیستم‌های حقیقی، ساختار انجمنی^۲ است. گروهی از رئوس که با یالهای زیادی به یکدیگر متصل هستند در حالیکه تعداد یالهای بسیار کمتری، رئوس این گروه را به باقی شبکه متصل می‌نماید، انجمن نامیده می‌شود. به طور منصفانه‌ای یک چنین خوشه‌ها یا انجمن‌هایی را می‌توان به عنوان اجزای مستقل یک گراف در نظر گرفت، که همان نقش بافت‌ها یا اعضا را در بدن انسان دارند. کشف انجمن‌ها اهمیت بسیاری در جامعه‌شناسی، بیولوژی و علوم کامپیوتر و هر جایی که سیستم‌ها را بتوان به صورت گراف مدل کرد دارد. کشف انجمن‌ها از جمله مسائل سخت محسوب می‌شود که هنوز به طور رضایت‌مندان‌های حل نشده است، در حالیکه دانشمندان بسیاری در یک دهه گذشته بر روی آن کار و تحقیق انجام داده‌اند.

تاریخ نظریه گراف تقریباً به سال ۱۷۳۶ باز می‌گردد و از آن موقع به بعد دانش بسیاری درباره گراف و ویژگی‌های ریاضی آن بدست آمد. گراف‌ها در قرن بیست و یکم در نمایش سیستم‌های گوناگونی در حوزه‌های علمی متفاوتی به طور گسترده مورد استفاده قرار گرفتند. شبکه‌های بیولوژیکی^۳، شبکه‌های اجتماعی^۴، شبکه‌های اطلاعاتی^۵ از جمله سیستم‌هایی بودند که به عنوان گراف مورد مطالعه قرار گرفتند، بنابراین تحلیل گراف در فهم ویژگی‌های این سیستم‌ها، بسیار پر اهمیت شد. به عنوان نمونه تحلیل

^۱ Complex Networks

^۲ Community Structure

^۳ Biological Networks

^۴ Social Networks

^۵ Information Networks

شبکه‌های اجتماعی در سال ۱۹۳۰ به عنوان یکی از موضوعات مهم در جامعه‌شناسی مطرح شد. در سالهای اخیر، انقلاب کامپیوترها، حجم عظیم از داده‌ها و منابع محاسباتی برای پردازش و تحلیل این داده‌ها در اختیار محققین قرار داده است. نیاز به درگیر شدن با چنین شبکه‌های بزرگی نیاز به تغییر مسیری است که در آن گراف‌ها مورد بررسی قرار می‌گرفته‌اند.

نمایش گرافی سیستم‌های حقیقی مثل توری^۱ خیلی با قاعده نیست. آنها اشیایی هستند که در آنها نظم همراه بی‌نظمی زندگی می‌کند. نمونه گراف‌های بی‌نظم گراف‌های تصادفی^۲ هستند، که در آن احتمال وجود یک یال میان یک جفت راس با همه جفت راس‌های ممکن برابر است. در گراف تصادفی توزیع یال‌ها میان رئوس، همگن^۳ است. برای مثال توزیع همسایه‌های یک راس یا توزیع درجه، دوجمله‌ای^۴ است بنابراین اکثر راس‌ها درجه یکسان یا شبیهی دارند.

شبکه‌های واقعی گراف تصادفی نیستند. آنها یک عمق زیادی از نظم و سازمان دهی را آشکار می‌کنند در حالیکه آنها ناهمگونی^۵ بزرگی را نمایش می‌دهند. توزیع درجه در آنها پهن است با یک دنباله‌ای که اغلب از قانون توانی^۶ پیروی می‌کند. بنابراین خیلی از راس‌های با درجه پایین با رئوسی با درجه‌های خیلی بالا همزیستی می‌کنند. علاوه بر این توزیع یال نه فقط به طور سراسری که به طور محلی نیز ناهمگون است، که این بدین معناست که تجمع بالایی از یالها میان گروه‌های خاصی از رئوس وجود دارد و یال‌ها متصل کننده این گروه‌ها در بین این گروه‌ها تمرکز و تجمع کمتری دارند. این ویژگی شبکه‌های اجتماعی

^۱ Lattice

^۲ Random Graphs

^۳ Homogeneous

^۴ Binomial

^۵ Inhomogeneity

^۶ Power Law

“ساختار انجمنی” نامیده می‌شود. انجمن‌ها خوشه یا ماژول^۱ نیز نامیده می‌شوند. انجمن‌ها در واقع گروهی از رؤس هستند که احتمالاً ویژگی‌های شبیهی را به اشتراک می‌گذارند و نقش‌های یکسانی در گراف دارند.

۱-۲- تعریف مسئله

تحلیل شبکه‌های اجتماعی^۲ (SNA) در واقع بررسی ارتباطات میان افراد است که شامل تحلیل ساختارهای اجتماعی، موقعیت‌های^۳ اجتماعی، تحلیل نقش‌ها^۴ و خیلی چیزهای دیگر است. به طور معمول ارتباط میان افراد یعنی خویشاوندی، دوستی، همسایگی، همکاری، هم کلاسی و ... به عنوان یک شبکه یا گراف قابل نمایش است. علوم اجتماعی بدین گونه عمل می‌کردند که با انتشار پرسشنامه‌هایی میان افراد و در خواست جواب از آنها، جزئیات تعاملاتشان با دیگران را استخراج می‌کردند. سپس یک شبکه براساس جواب آنها ساخته می‌شد که گره‌ها نمایانگر افراد و یالها نمایش دهنده‌ی تعاملات میان آنها بودند. این نوع جمع آوری داده تحلیل شبکه‌های اجتماعی سنتی را محدود به مقیاس‌های کوچک می‌کرد. با گسترش اینترنت و وب و پیشرفت فناوری بسیاری از سایت‌های شبکه‌های اجتماعی و رسانه‌های اجتماعی پدیدار گشتند. این پیشرفته‌ها باعث شد مردم به راحتی با یکدیگر در فضای مجازی ارتباط برقرار کنند. این پیشرفت باعث می‌شد که کار تحلیل شبکه‌های اجتماعی در مقیاس‌های بزرگتر نیز آسان شود. شبکه‌های ارتباطی ایمیلی، شبکه‌های پیغام دهی و گفتگو^۵، شبکه‌های موبایلی و شبکه‌های دوستی نمونه‌هایی از این شبکه‌های

^۱ Module

^۲ Social Network Analysis

^۳ Positions

^۴ Roles

^۵ Instant Messaging

اجتماعی مقیاس بزرگ هستند. این شبکه‌های واقعی و بزرگ دارای الگوهایی^۱ هستند که برای SNAها اهمیت زیادی دارند. یکی از پرکاربردترین و مهمترین این الگوها، ساختار انجمنی این شبکه‌ها هستند. افراد یا ایفاکنندگان نقش‌ها^۲ در شبکه‌های اجتماعی گروه‌هایی را تشکیل می‌دهند. بررسی ساختار و توپولوژی شبکه‌ها که منجر به کشف انجمن‌ها می‌شود یکی از وظایف مهم آنهاست.

داده کاوی، پایگاه‌ها و مجموعه‌های حجیم داده‌ها را در پی کشف و استخراج دانش مورد تحلیل و کند و کاوهای ماشینی (و نیمه ماشینی) قرار می‌دهد. این روش که با حداقل دخالت کاربران همراه است، اطلاعاتی را در اختیار آنها و تحلیلگران قرار می‌دهد تا براساس آنها تصمیمات مهم و حیاتی را در سازمان مربوطه اتخاذ نمایند. داده کاوی در سالهای اخیر، تأثیرات شگرفی در محیط‌های آکادمیک و صنعتی ایجاد کرده و کاربردهای فراوانی در زمینه‌های مختلف یافته است. به عنوان نمونه می‌توان به کاربردهای تجاری، مدیریت و کشف فریب، پزشکی، ورزشی، متن کاوی و وب کاوی اشاره نمود.

این فرایند از دو مرحله اصلی تشکیل شده است [۹]. مرحله اول پیش پردازش داده‌ها است که شامل پاکسازی، یکپارچه سازی، انتخاب صفات و تبدیل داده‌ها به قالب مورد استفاده برای داده کاوی میباشد. در مرحله دوم، داده‌های بدست آمده از مرحله اول به منظور تشخیص الگو مورد استفاده قرار می‌گیرند که این امر به کمک الگوریتم‌هایی نظیر دسته بندی^۳ و خوشه‌بندی^۴ صورت می‌گیرد. در این مرحله برای کشف الگو یک مدل یادگیری ایجاد می‌شود که این به مدل به مرور زمان و با تکرار فرآیند داده کاوی بهبود می‌یابد. خروجی این مرحله دانش کسب شده است که به کمک ابزارهای موجود قابل نمایش است.

^۱ Patterns

^۲ Actors

^۳ Classification

^۴ Clustering

به طور کلی تکنیک‌های داده کاوی به دو گروه با نظارت^۱ و بدون نظارت^۲ تقسیم بندی می‌شوند. اکثر روش‌های داده کاوی روشهای با نظارت می‌باشند که در این روشها یک متغیر هدف از قبل تعریف شده، وجود دارد. در این روشها نمونه‌های زیادی وجود دارند که مقدار متغیر هدف برای آنها از قبل مشخص میباشد، بنابراین الگوریتم میتواند به کمک آنها آموزش ببیند و دریابد که متغیرها و صفات توصیف کننده یک نمونه با کدام مقدار متغیر هدف متناظر میباشد. اما در روش‌های بدون نظارت متغیر هدفی تعریف نمیشود و الگوریتم داده کاوی همبستگی‌ها و ساختارهای بین تمام نمونه‌ها را جستجو میکنند. از مهمترین روشهای داده کاوی بدون نظارت، خوشه بندی را میتوان نام برد.

همانگونه که بیان شد شبکه‌های اجتماعی را می‌توان با یک گراف نمایش داد. و ما می‌خواهیم در این گراف به یافتن الگوهایی به نام انجمن بپردازیم. انجمن به گروهی از رئوس گفته می‌شود که تعداد اتصالات داخلی گروهی آنها بسیار زیادتر از اتصالات مابین گروهی آنهاست. این تعریف کلی انجمن است. بنابراین مسئله کشف انجمن در داده کاوی همان مسئله خوشه‌بندی است. اما تفاوت عمده و نکته اصلی این است که داده‌ها در اینجا گره‌های یک گراف هستند و باید به ارتباطات میان این گره‌ها در گراف برای خوشه‌بندی توجه شود.

^۱ Supervised

^۲ Unsupervised