واحد بین‌الملل

پایان‌نامه کارشناسی ارشد

در رشته مهندسی فناوری اطلاعات (مدیریت سیستم‌های اطلاعاتی)

ارائه مدلی کارا برای یافتن الگوی انتشار اطلاعات در شبکه‌های اجتماعی

به کوشش

صبا باباخانی

استاد راهنما:

دکتر علی حمزه

شهریورماه ۱۳۹۳

بـه نـام خـدا

اظهارنامه

اینجانب صباباخانی (۹۱۶۰۰۴۳) دانشجوی رشته‌ی مهندسی فناوری اطلاعات گرایش مدیریت سیستم های اطلاعاتی دانشکده واحد بین الملل اظهار می‌کنم که این پایان‌نامه حاصل پژوهش خودم بوده و در جاهایی که از منابع دیگران استفاده کرده‌ام، نشانی دقیق و مشخصات کامل آن را نوشته‌ام. همچنین اظهار می‌کنم که تحقیق و موضوع پایان‌نامه‌ام تکراری نیست و تعهد می‌نمایم که بدون مجوز دانشگاه دستاوردهای آن را منتشر ننموده و یا در اختیار غیر قرار ندهم. کلیه حقوق این اثر مطابق با آیین‌نامه مالکیت فکری و معنوی متعلق به دانشگاه شیراز است.

نام و نام خانوادگی: صبا باباخانی

تاریخ و امضا:

به نام خدا

ارائه مدلی کارا برای یافتن الگوی انتشار اطلاعات در شبکه‌های اجتماعی

به کوشش : صبا باباخانی

پایان نامه‌ی

ارائه شده به تحصیلات تکمیلی دانشگاه به عنوان بخشی از فعالیت‌های تحصیلی لازم
برای اخذ درجه کارشناسی ارشد

در رشته‌ی

مهندسی فناوری اطلاعات- مدیریت سیستم های اطلاعاتی

از دانشگاه شیراز
شیراز
جمهوری اسلامی ایران

ارزیابی شده توسط کمیته پایان نامه با درجه: عالی

دکتر علی حمزه، استادیار بخش مهندسی و علوم کامپیوتر (استاد راهنما) .....

دکتر شهرام جعفری، استادیار بخش مهندسی و علوم کامپیوتر (استاد مشاور) .....

دکتر مصطفی فخراحمد، استادیار بخش مهندسی و علوم کامپیوتر(داور متخصص داخلی) .....

مهرماه 1393

نخست تقدیم به:

پدر و مادر عزیز و خواهر مهربانم

که ضرب‌آهنگ قلبم بسته به حضور گرم آن‌ها در کنار من است ...

و سپس تقدیم به او که دوستش دارم ...

سپاسگزاری

تشکر و سپاس فراوان از :
استاد عزیز جناب آقای دکتر علی حمزه

که در طی دوران تحصیلم و به خصوص در طی انجام این پایان‌نامه ، دلسوزانه به من آموختند و کلامشان روشنگر راه من بود، و مشوق حرکت به سمت جلو؛ و نفس گرمشان باعث دلگرمی و پیمودن ادامه مسیر است، و نوید بخش امید به آینده ...
همچنین سپاس فراوان از **سرکار خانم نیلوفر مظفری**، راهنما و دوست خوبم در به ثمر رساندن این پایان نامه ...

چکیده

ارائه مدلی کارا برای یافتن الگوی انتشار اطلاعات در شبکه های اجتماعی

به کوشش

صبا باباخانی

دنیای اینترنت، وب۲ و تکنولوژی‌های آن باعث به وجود آمدن گونه‌های جدیدی از ارتباط میان انسان‌ها شده است. این تکنولوژی‌ها به راحتی در دسترس کاربران قرار دارند و به آنها اجازه می‌دهند تا از طریق وبلاگ، ایمیل، پیام‌های متنی فوری و شبکه اجتماعی با سایر کاربران تعامل برقرارکنند. کاربران از طریق شبکه‌های اجتماعی قادر خواهند بود تا عقاید، توصیه‌ها و نظرات خود را در مورد کالاها و خدمات و اخبار به اشتراک گذارند و از نظرات و عقاید سایر افراد باخبر گردند. شبکه‌های اجتماعی، علیرغم تفاوت در ساختار و نوع تعاملات اجتماعی بین اعضای خود، ویژگی‌های آماری مشترکی دارند. جدا از ویژگی‌های آماری مشترک میان شبکه‌های اجتماعی، ویژگی دیگری را نیز می‌توان در نظر گرفت ،این ویژگی، فرآیند انتشار در شبکه های اجتماعی نام دارد. انتشار اطلاعات یک تعریف عمومی است که شامل هر چیزی که در یک شبکه گسترش می-یابد، می‌شود.

همان‌طور که اشاره گردید یکی از جدیدترین و مهم‌ترین موضوعات تحقیقی در حوزه انتشار اطلاعات که در سال‌های اخیر توجه متخصصان تحلیل شبکه‌های اجتماعی را به خود جلب کرده است، نحوه انتشار اطلاعات در شبکه‌های اجتماعی، مسیرهای انتشاری و الگوی انتشار اطلاعات در این شبکه‌ها می‌باشد. درحقیقت تمرکز این مبحث بر روی پیدا کردن الگوی کارایی است تا بتواند نحوه‌ی انتشار اطلاعات را برطبق واقعیت پیش بینی و مدل نماید که کاربردهای عمده‌ای در زمینه‌های متفاوتی چون خرید محصولات توسط مردم، انتشار ویروس-ها و آلودگی‌های کامپیوتری، پست‌گذاری در بلاگ‌ها ، انتشار شایعات در شبکه‌های اجتماعی و ... می‌تواند داشته باشد. برای یافتن الگوی انتشار، نودهایی که در گذشته تحت تاثیر یک خبر منتشر شده قرار گرفته‌اند در نظر گرفته شده و براساس یک سری از پارامترها، نودهایی که در آینده تحت تاثیر خبر قرارخواهند گرفت، به عنوان تابعی از زمان پیش بینی می‌گردند. یکی از پارامترهای مورد استفاده در این مدل محبوبیت خبر است که براساس این پارامتر، تاثیرگذاری بر سایر نودها در آینده انجام می‌گیرد. در این پایان‌نامه دو مدل برای مدل-سازی الگوی انتشار اطلاعات ارائه شده‌است. مدل اول که به اختصار MIDM(Mathematical Information Diffusion Model) نام دارد، مسئله الگوی انتشار اطلاعات را با در نظرگرفتن مقادیر متفاوت محبوبیت خبر در زمان‌های مختلف مدل‌سازی می‌نماید و روش دوم که RLIDM (Reinforcement Learning Information Diffusion Model) است، با استفاده از سامانه یادگیری XCSF الگوی انتشار اطلاعات در شبکه‌های اجتماعی واقعی Memetracker، Twitter و Google را مدل می‌کند. نتایج مقایسه دو روش پیشنهادی در مقابل مدل اصلی نشان‌دهنده برتری مدل‌های پیشنهادی می‌باشند.

واژگان کلیدی: شبکه های اجتماعی، الگوی انتشار اطلاعات، محبوبیت خبر، نودهای تحت تاثیر خبر، سامانه یادگیر طبقه بند، XCSF

فهرست مطالب

عنوان **صفحه**

فصل اول : مقدمه

۱-۱. مقدمه ۲
۱-۲. هدف از انجام این پایان نامه ۸
۱-۳. نگاه کلی به فصول رساله ۱۴

فصل دوم : پیشینه ی تحقیق

۲-۱. مقدمه ۱۶
۲-۲. تعریف مسئله ۱۷
۲-۳. روش های اخیر جهت مدل سازی الگوی انتشار اطلاعات ۱۸
۲-۳-۱. روش های قدیمی (انتشار و فراگیری اطلاعات و نوآوری) ۱۹
۲-۳-۲. روش های جدید ۲۰
۲-۳-۲-۱. مدل انتشار در حوزه وبلاگ نویسی ۲۱
۲-۳-۲-۲. مدل C-S ۲۲
۲-۳-۲-۳. مدل LIM ۲۳
۲-۳-۲-۴. مدل SpikeM ۲۶

فصل سوم : ارائه راه حل و روش های پیشنهادی

۳-۱. مقدمه ۲۹
۳-۲. بررسی فرآیند الگوی انتشار اطلاعات در شبکه های اجتماعی ۳۱
۳-۳. توزیع توانی و رفتار پریودیک ۳۳
۳-۳-۱. توزیع توانی ۳۴
۳-۳-۲ رفتار پریودیک ۳۵
۳-۴. مروری بر مدل SPIKEM ۳۵

۳-۵. مروری بر سامانه طبقه بند یادگیر ۳۸
۳-۵-۱. سیستم طبقه بند یادگیر XCS ۳۹
۳-۵-۱-۱. نمایش دانش XCS ۴۰
۳-۵-۱-۲. فعل و انفعالات یادگیری XCS ۴۲
۳-۵-۱-۳. سامانه یادگیری XCS ۴۴
۳-۵-۱-۴. سامانه کشف طبقه بندها XCS ۴۷
۳-۵-۱-۵. عمگر پوشش ۴۸
۳-۵-۱-۶. حذف پوششی ۴۹
۳-۵-۱-۷. رفتار متقابل کاوش و بهره وری ۴۹
۳-۵-۲. سیستم طبقه بند یادگیری XCSF ۵۰
۳-۶. روش های پیشنهادی ۵۰
۳-۶-۱. روش اول : مدل جدید MIDM (براساس مدل SpikeM) ۵۱
۳-۶-۱-۱. پارامترها و محاسبه مقادیر پارامترها ۵۵
۳-۶-۱-۲. فرضیات مدل MIDM ۵۵
۳-۶-۲.روش دوم : مدل الگوی انتشار اطلاعات بر اساس روش یادگیری XCSF ۵۵
۳-۶-۲-۱. نمایش دانش XCSF ۵۷
۳-۶-۲-۲. فعل و انفعالات یادگیری XCSF ۵۸
۳-۶-۲-۳. سامانه یادگیری XCSF ۵۹
۳-۷. جمع بندی ۶۰

فصل چهارم : نتایج تجربی

۴-۱. مقدمه ۶۱
۴-۲. مجموعه داده ها ۶۲
۴-۲-۱. مجموعه داده های واقعی ۶۲
۴-۲-۱-۱. Memetracker ۶۲
۴-۲-۱-۲. Twitter ۶۳
۴-۲-۱-۳. Google Trends ۶۴
۴-۳.. معیار ارزیابی ۶۵
۴-۴. نتایج و تحلیل ها ۶۶
۴-۴-۱. نتایج مدل MIDM بر روی مجموعه های دادهای ۶۶
۴-۴-۱-۱. نتایج مدل MIDM بر روی مجموعه دادهای Memetracker ۶۸
۴-۴-۱-۲. نتایج مدل MIDM بر روی مجموعه دادهای Twitter ۷۲
۴-۴-۱-۳. نتایج مدل MIDM بر روی مجموعه دادهای Google ۷۷
۴-۴-۲. نتایج مدل RLIDM بر روی مجموعه های دادهای ۸۱
۴-۴-۲-۱. نتایج مدل RLIDM بر روی مجموعه دادهای Memetracker ۸۱
۴-۴-۲-۲. نتایج مدل RLIDM بر روی مجموعه دادهای Twitter ۸۳
۴-۴-۲-۳. نتایج مدل RLIDM بر روی مجموعه دادهای Google ۸۶
۴-۴-۳ توانایی مدل RLIDM در یادگیری الگوی انتشار اطلاعات ۸۶
۴-۴-۴. مقایسه بین سه مدل SpikeM، MIDM وRLIDM ۸۷

۴-۵. جمع بندی ۹۰

فصل پنجم : جمع بندی

۵-۱. نتیجه گیری ۹۲
۵-۲. پیشنهادها و کارهای آتی ۹۷
فهرست منابع ۹۸

فهرست جداول

عنوان **صفحه**

جدول ۴-۱.ویژگی های مجموعه داده ایMEMETRACKER ۶۳
جدول ۴-۲.ویژگی های مجموعه داده ای TWITTER..... ۶۴
جدول ۴-۳.ویژگی های آماری مجموعه داده ایTWITTER ۶۵
جدول ۴-۴.لسیت اصطلاحات مجموعه داده ای MEMETRACKER برای مدلMIDM ۶۸
جدول ۴-۵.لیست هشتگ های مجموعه دادهای TWITTER برای مدلMIDM ۷۲
جدول ۴-۶.لیست اصطلاحات جست وجو شده در GOOGLE برای مدلMIDM ۷۸
جدول ۴-۷.لیست اصطلاحات مجموعه داده ای MEMETRACKER برای مدلRLIDM ۸۲
جدول ۴-۸.لیست هشتگ های مجموعه داده ای TWITTER برای مدلRLIDM ۸۴
جدول ۴-۹.لیست اصطلاحات جست وجو شده در GOOGLE برای مدلRLIDM ۸۶
جدول ۴-۱۰.مقایسه بین دو حالت PRIOR TRAINING و TRAINING برروی سه مجموعه داده ای ۸۶
جدول ۴-۱۱.مقایسه بین سه مدل SPIKEM، MIDM و RLIDM برروی مجموعه داده ای MEMETRACKER... ۸۹
جدول ۴-۱۲.مقایسه بین سه مدل SPIKEM، MIDM و RLIDM برروی مجموعه داده ای TWITTER..... ۸۹
جدول۴-۱۳.مقایسه بین سه مدل SPIKEM، MIDM و RLIDM برروی مجموعه داده ای GOOGLE ۸۹

فهرست شکل ها

عنوان **صفحه**

شکل ۱-۱.نمایی از شبکه اجتماعی با مجموعه رأس ها و ارتباط هایشان ۷
شکل ۲-۱.مدل نشان دهنده ساختار یک وبلاگ است ۲۲
شکل ۲-۲.تابع اثر که در آن تعداد نودهای تحت تاثیر قرار گرفته در آینده است ۲۴
شکل ۳-۱.روند کارکرد سیستم XCS به صورت شماتیک ۴۳
شکل ۳-۲.توصیف کلی نحوه انتشار یک خبر در طول شبکه در زمان های مختلف ۵۲
شکل ۳-۳.روند یادگیری مدل RLIDM برای یادگیری الگوی انتشار اطلاعات ۵۷
شکل ۴-۱.نتایج حاصل از مدل MIDM بر روی مجموعه داده ای MEMETRACKER در مقیاس خطی ۶۹
شکل ۴-۲.نتایج حاصل از مدل MIDM بر روی مجموعه داده ای MEMETRACKER در مقیاس لگاریتمی ۷۰
شکل ۴-۳.مقایسه ای بین نتایج حاصل از روش پیشنهادی MIDM و روش SPIKEM بر روی MEMETRACKER.... ۷۲
شکل ۴-۴.نتایج حاصل از مدل MIDM بر روی مجموعه داده ای MEMETRACKERدر مقیاس خطی ۷۴
شکل ۴-۵.نتایج حاصل از مدل MIDM بر روی مجموعه داده ای TWITTER در مقیاس لگاریتمی ۷۵
شکل ۴-۶.مقایسه ای بین نتایج حاصل از روش پیشنهادی MIDM و روش SPIKEM بر رویTWITTER ۷۷
شکل ۴-۷.نتایج حاصل از مدل MIDM بر روی مجموعه داده ای GOOGLE در مقیاس خطی ۷۸
شکل ۴-۸.نتایج حاصل از مدل MIDM بر روی مجموعه داده ای GOOGLE در مقیاس لگاریتمی ۷۹
شکل ۴-۹.مقایسه ای بین نتایج حاصل از روش پیشنهادی MIDM و روش SPIKEM بر رویGOOGLE ۸۰
شکل ۴-۱۰.نتایج حاصل از مدل RLIDM بر روی مجموعه داده ای MEMETRACKER در مقیاس خطی ۸۳
شکل ۴-۱۱.نتایج حاصل از مدل RLIDM بر روی مجموعه داده ای TWITTER در مقیاس خطی ۸۵
شکل ۴-۱۲.نتایج حاصل از مدل RLIDM بر روی مجموعه داده ای GOOGLE در مقیاس خطی ۸۷

فصل اول

۱. مقدمه

۱-۱. مقدمه

تعامل انسان با کامپیوتر[1] از زمان ایجاد اولین کامپیوترها همواره مورد توجه بوده است و شامل برنامه ریزی و طراحی رابطه بین مردم (کاربران) و کامپیوترها است. معمولا از HCI به عنوان نقطه تقاطع علوم کامپیوتر، علوم رفتاری[2]، علم طراحی[3] و چند زمینه‌ی دیگر یاد می-شود. این اصطلاح برای اولین بار توسط کارد، موران و نیوول در کتاب " روانشناسی تعامل انسان با کامپیوتر" مطرح شده است و دلالت ضمنی بر این مطلب دارد که بر خلاف ابزارهایی که کاربرد محدود دارند، کامپیوتر دارای مزایا و کاربردهای بی‌شماری بوده که در یک دیالوگ بدون انتها بین آن و کاربر انجام می شوند[۱].

متخصصان این حوزه در ابتدا به دنبال تولید سخت افزارهایی با ارگونومی مناسب برای راحتی انسان بودند. طی دهه‌ی ۱۹۸۰، این حوزه با تحولی پارادایمیک مواجه شد و تمرکز اصلی آن به جنبه‌های ادراکی رفتار کاربر و ایجاد نرم‌افزارهای کاربر پسند سوق یافت. اما طولی نکشید که در دهه ۱۹۹۰، موج جدیدی گفتمان غالب متخصصان این حوزه را متحول ساخت. در این گفتمان جدید، افزایش کیفیت ارتباط میان انسان‌ها با کامپیوترها هدف نبود، بلکه در این دیدگاه کامپیوتر به عنوان ابزاری برای ایجاد تعاملات انسانی نگاه می‌شد. با توجه به این رویکرد، شبکه‌های اجتماعی اینترنتی به عنوان عامل ایجاد تعامل میان انسان‌ها در فضای مجازی از اهمیت خاصی برخوردار گشت. یکی از این تاثیرات ظهور مفاهیم جدید، چون وب اجتماعی۲ [4] بود[۲].

با گسترش وب اجتماعی، نیاز به تحلیل ساختارها و رفتارهای شبکه‌های اجتماعی، به عنوان یکی از نیازمندی‌های اساسی شرکت‌های تجاری مبدل گشت. تحلیل شبکه‌های اجتماعی در بسیاری از کاربردها از جمله مدیریت شبکه اجتماعی، تحلیل گرایش بازار،

[1] Human- Computer Interaction (HCI)
[2] Behavior Science
[3] Design Science
[4] Web 2.0

شناسایی افراد تاثیرگذار و حامیان، ارتقاء کارایی سامانه‌های توصیفگر و ... قابل استفاده است. نیازمندی‌های تجاری باعث شده است در سال‌های اخیر در بعد آکادمیک توجه زیادی به تحلیل شبکه‌های اجتماعی گردد. امروزه این ابزار قدرتمند نه تنها مورد متخصصان فناوری اطلاعات می‌باشد، بلکه پژوهشگران سایر رشته‌هایی چون علوم تربیتی، زیست شناسی، علوم ارتباطات، اقتصاد و... به عنوان یک تکنیک کلیدی از تحلیل شبکه اجتماعی بهره می‌برند.

برای تحلیل شبکه‌ها، از معیارها و نرم‌افزارهای متفاوتی استفاده می‌شود. نرم‌افزارهای تجزیه و تحلیل شبکه‌های اجتماعی جهت شناسایی، تجزیه و تحلیل، تجسم و شبیه سازی رأس‌ها و یال‌ها از انواع مختلف داده‌های ورودی (رابطه‌ای و غیر رابطه‌ای)، از جمله مدل‌های ریاضی شبکه‌های اجتماعی است. ابزار تجزیه وتحلیل شبکه به محققان اجازه می دهد تا شبکه-هایی با اندازه‌های مختلف (شبکه‌های کوچک مانند خانواده و شبکه‌های بزرگ مانند اینترنت) را بررسی کنند این نرم‌افزارها با فراهم آوردن ابزارهای مختلف اجازه اعمال رویه‌های ریاضی و آماری را روی مدل شبکه می‌دهند. این نرم‌افزارها با نمایش‌های بصری شبکه‌های اجتماعی به درک و تحلیل نتایج کمک زیادی می کنند. معیارهای زیر در تحلیل شبکه‌های اجتماعی کاربرد وسیعی دارند:

Betweenness : تعداد افرادی در شبکه که یک شخص بطور غیر مستقیم از طریق خطوط مستقیم آن‌ها متصل شده است.

Closeness : تنوع مجموعه کوتاهترین مسیرها بین هر فرد و دیگر افراد در شبکه.

Centrality Degree : محاسبه میزان پیوندهایی که فرد با دیگر افراد در شبکه دارد.

Centralization: تفاوت بین تعداد پیوندها برای هر نود تقسیم شده توسط بیشترین مجموع تفاوت‌ها. یعنی در یک شبکه همیشه رأس‌هایی وجود دارند که نسبت به دیگر رأس‌ها تعداد پیوندهای بیشتری دارند. در شبکه‌ای که دچار عدم تمرکز است تفاوت کمی بین پیوندهای هر نود وجود دارد.

Cohesion : اشاره به درجه‌ای دارد که افراد بطور مستقیم با همدیگر ارتباط دارند.

Path Length : مسافت بین هر دو نود در یک شبکه را می‌گویند، میانگین path length در واقع میانگین مسافت‌های بین تمامی جفت رأس‌ها است.

Structural Hole : تعداد کمی از افراد که اگر از گروه خارج شوند گروه از همدیگر جدا می‌شوند و اتصالات قطع می‌شود.

شبکه اجتماعی ساختاری اجتماعی است که از گره‌هایی (که عموما فردی یا سازمانی هستند) تشکیل شد است که توسط یک یا چند نوع خاص از وابستگی به هم متصل اند، برای مثال قیمت‌ها، الهامات، ایده‌ها و تبادلات مالی، دوست‌ها خویشاوندی، تجارت ، لینک‌های وب، سرایت بیماری‌ها (اپیدمیولوژی) یا مسیرهای هواپیمایی. ساختارهای حاصل اغلب بسیار پیچیده هستند. تحلیل شبکه‌های اجتماعی روابط اجتماعی را با اصطلاحات رأس و یال می-نگرد. رأس‌ها بازیگران فردی درون شبکه‌ها هستند و یال‌ها روابط میان این بازیگران هستند. انواع زیادی از یال‌ها می‌تواند میان رأس‌ها وجود داشته باشد. تحقیق در تعدادی از زمینه‌های آکادمیک نشان داده است که شبکه‌های اجتماعی در بسیاری از سطوح به کار گرفته می‌شوند از خانواده‌ها گرفته تا ملت‌ها و نقش مهمی در تعیین راه حل مسائل، اداره کردن تشکیلات و میزان موفقیت افراد در رسیدن به اهدافشان ایفا می‌کند. در ساده ترین شکل یک شبکه اجتماعی نگاشتی از تمام یال‌های مربوط، میان رأس‌های مورد مطالعه است. شبکه‌ی اجتماعی هم‌چنین می‌تواند برای تشخیص موقعیت اجتماعی هر یک از بازیگران مورد استفاده گیرد. این مفاهیم غالبا در یک نمودار شبکه‌ی اجتماعی نشان داده می‌شوند که در آن، نقطه‌ها رأس‌ها هستند و خط‌ها نشانگر یال‌ها[۳]. اما آن‌چه که در این تحلیل، پایه و اساس محسوب می‌شود، نظریه‌ی گراف‌ها است. که در کنار کاربردهای بیشمار آن، در تحلیل شبکه‌های اجتماعی نیز نقش مهمی ایفا می‌کنند. در ادامه به شرح خلاصه‌ای از تاریخچه گراف و مواردی از کاربرد آن خواهیم پرداخت.

نظریه گراف شاخه‌ای از ریاضیات بوده که درباره‌ی گراف‌ها بحث می‌کند و در واقع شاخه-ای از توپولوژی است که با جبر و نظریه ماتریس‌ها پیوند مستحکم و تنگاتنگی دارد. نظریه‌ی گراف بر خلاف شاخه‌های دیگر ریاضیات خاستگاه مشخصی دارد و آن، حل معمای پل‌های کونیگزبرگ[5] در سال ۱۷۳۶ و توسط اویل ریاضیدان سوئیسی است[۴]. از آن زمان تاکنون تحقیقات زیادی بر روی گراف‌ها انجام شده که این امر منجر به جمع‌آوری مطالب بسیاری در مورد آن‌ها شده است. از جمله‌ی این تحقیقات می‌توان به موارد زیر اشاره کرد.

در سال ۱۸۴۷، گوستاوکیرشهف نوع خاصی از گراف‌ها به نام درخت را مورد بررسی قرار داد. کیرشهف این مفهوم را هنگام تعمیم قوانین اهم برای جریان الکتریکی در کاربردهایی که

[5] Puzzle of Konigsberg's bridges

حاوی شبکه‌های الکتریکی بودند به کار گرفت. ده سال بعد، آرتور کیلی همین نوع گراف را برای شمارش ایزومرهای متمایز هیدروکربن‌های اشباع شده‌ی C_nH_{2n+2} به کاربرد. در همین دوران دو ایده‌ی دیگر در صحنه حضور پیدا کردند. ایده‌ی اول حدس چهار رنگ بود که نخستین بار توسط فرانسیس گوثری در حدود ۱۸۵۰ مورد تحقیق قرار گرفت. این مسئله سرانجام در سال ۱۹۷۶، توسط کنث اپل و ولفگانگ هیکن و با استفاده از یک تحلیل رایانه‌ای پیچیده حل شد. ایده‌ی مهم دوم، دور همیلتونی بود. این دور به افتخار سر ویلیام روآن همیلتون نامگذاری شده است. او این ایده را در سال ۱۸۵۹ برای حل معمای جالبی حاوی یال-های یک دوازده وجهی منتظم به کار گرفت. یافتن جوابی برای این معما چندان دشوار نیست، ولی ریاضیدانان هنور در پی یافتن شرایطی لازم و کافی هستند که گراف‌های بیسوی حاوی مسیر یا دور همیلتونی را مشخص کنند. پس از این کارها تا بعد از سال ۱۹۲۰ فعالیت اندکی در این زمینه صورت گرفت. مسئله‌ی مشخص کردن گراف‌های مسطح را کازیمیر کوراتوفسکی ، ریاضیدان لهستانی، در سال ۱۹۳۰ حل کرد. نخستین کتاب درباره‌ی نظریه گراف در سال ۱۹۳۶ منتشر شد. این کتاب را ریاضیدان مجارستانی، دنش کونیگ، که خود محقق برجسته‌ای در این زمینه بود، نوشت. از آن پس فعالیت‌های بسیاری در این زمینه صورت گرفته و کامپیوتر نیز در چهار دهه‌ی اخیر به یاری این فعالیت‌ها آمده است[۵].

پیشرفت‌های اخیر در ریاضیات، به ویژه در کاربردهای آن موجب گسترش چشمگیر نظریه‌ی گراف شده است به گونه‌ای که هم اکنون نظریه‌ی گراف ابزار بسیار مناسبی برای تحقیق در زمینه‌های گوناگون مانند نظریه کدگذاری، تحقیق در عملیات، آمار، شبکه‌های الکتریکی، علوم رایانه، شیمی، زیست شناسی، علوم اجتماعی و سایر زمینه‌ها گردیده‌است.

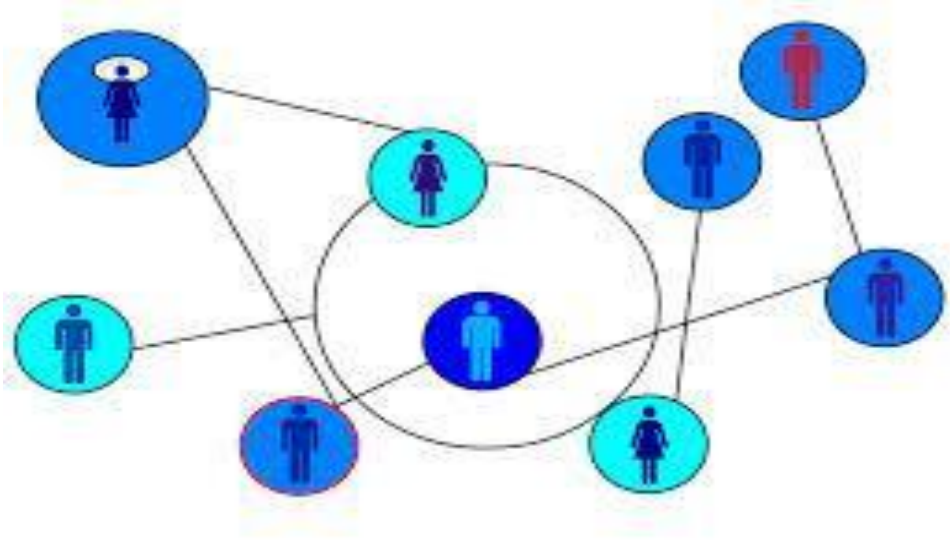
از گراف‌ها برای حل مسائل زیادی در ریاضیات و علوم کامپیوتر استفاده می‌شود. ساختارهای زیادی را می‌توان به کمک گراف‌ها به نمایش در آورد. برای مثال برای نمایش چگونگی رابطه وب سایت‌ها به یکدیگر می توان از گراف جهت‌دار استفاده کرد. به این صورت که هر وب سایت دیگری بود، یک یال جهت‌دار از این رأس به رأسی که وب سایت دیگر را نمایش می‌دهد وصل می کنیم. از گراف‌ها همچنین در شبکه‌ها، طراحی مدارهای الکتریکی، اصلاح هندسی خیابان‌ها برای حل مشکل ترافیک و... استفاده می‌شود. مهم‌ترین کاربرد گراف مدل‌سازی پدیده‌های گوناگون و بررسی بر روی آنهاست. با گراف می‌توان به راحتی یک نقشه بسیار بزرگ یا شبکه‌ای عظیم را در درون یک ماتریس وقوع گراف دخیره کرد و یا الگوریتم-

های مناسب مانند الگوریتم دایکسترا یا الگوریتم کروسکال و ... را بر روی آن اعمال نمود. در این جا به بررسی گراف‌هایی می‌پردازد که می‌توان آن‌ها را به نحوی روی صفحه کشید که یال-ها جز در محل رأس‌ها یکدیگر را قطع نکنند. این نوع گراف در ساخت جاده‌ها و حل مساله کلاسیک و قدیمی سه خانه و سه چاه آب به کار می‌رود. کاربرد گراف بازه‌ها از گراف‌ها برای حل مسائل زیادی در ریاضیات و علوم کامپیوتر استفاده می‌شود. ساختارهای زیادی را می‌توان به کمک گراف‌ها به نمایش در آورد. درخت و ماتریس درخت در رشته‌های مختلفی مانند شیمی مهندسی برق و علم محاسبه کاربرد دارد. کیرشهف در سال ۱۸۴۷ میلادی هنگام حل دستگاه‌های معادلات خطی مربوط به شبکه‌های الکتریکی درخت‌ها را کشف و نظریه درخت‌ها را بارور کرد. کیلی در سال ۱۸۵۷ میلادی درخت‌ها را در ارتباط با شمارش ایزومرهای مختلف هیدروکربن‌ها کشف کرد و. اگر هزینه کشیدن مثلا راه آهن بین هر دو شهر از p شهر مفروض مشخص باشد ارزان‌ترین شبکه‌ای که این p شهر را به هم وصل می‌کند با مفهوم یک درخت از مرتبه ارتباط نزدیک دارد. به جای مساله مربوط به راه آهن می‌توان وضعیت مربوط به شبکه-های برق رسانی و لوله کشی نفت و گاز و ایجاد کانال‌های آب‌رسانی را در نظر گرفت. برای تعیین یک شبکه با نازلترین هزینه از قاعده‌ای به نام الگوریتم صرفه جویی استفاده می‌شود که کاربردهای فراوان دارد. از گراف‌ها می‌توان به عنوان کدهای کمکی نام برد که به DVD player ها در بالا بردن قابلیت‌های آن‌ها کمک می‌کنند. گراف‌ها دارای مزایای مختلفی هستند که شفاف‌تر کردن و واضح‌تر کردن تصویر و کاهش مصرف CPU به عنوان یکی از اصلی‌ترین مزایای آن‌ها به شمار می‌رود[۵].

یکی دیگر از کاربردهای مهم گراف‌ها در نمایش سیستم‌های پیچیده طبیعی و اجتماعی و در قالب یک شبکه[۶] است؛ مجموعه‌ای از رأس‌ها و یال‌هایی که این رأس‌ها را به هم متصل می-کنند[۶]. به عنوان نمونه‌هایی از شبکه‌ها، می‌توان به شبکه‌های اجتماعی[۷] [۷]، شبکه‌های فنی[۸] [۸] و شبکه‌های زیستی مثل شبکه‌های عصبی[۹] اشاره کرد. رأس‌ها در این شبکه‌ها، موجودیت‌ها و یال‌ها، ارتباط بین آن‌ها را نشان می‌دهند. مثلاً در شبکه اینترنت، کامپیوترها یا

[6] Network
[7] Social Network
[8] Technological Network

مسیریاب‌ها و در شبکه اجتماعی، مردم را با رأس‌ها و ارتباط‌های داده‌ای بین کامپیوترها و یا روابط دوستی بین مردم را با یال‌ها نمایش می‌دهند.



شکل ۱-۱.نمایی از شبکه اجتماعی با مجموعه رأس‌ها و ارتباط‌هایشان

شبکه‌ها دارای ویژگی‌های آماری مشترکی هستند. یکی از ویژگی‌ها، ویژگی پدیده دنیای کوچک[۹] [۱۰] بوده که به ۶ درجه جدایی[۱۰] [۱۱] نیز معروف است و بیان می‌کند که در یک شبکه، فاصله متوسط بین رأس‌ها، کوتاه و معمولا تابعی لگاریتمی از تعداد آن‌ها است. ویژگی-های دیگر، ویژگی توزیع‌های درجه اریب به رأس[۱۱] [۱۲] بوده و بیان می‌کند که درجه بیشتر رأس‌های یک شبکه، کم و تنها تعداد محدودی از آن‌ها درجه بالا دارند و توزیع درجات، غالبا به فرم نمایی یا قانون توانی[۱۲] [۶] می‌باشد. ویژگی بعدی، ویژگی خوشه بندی[۱۳] یا انتقال پذیری شبکه[۱۴] است و بیان می‌کند که دو رأس مجاور با رأس سوم، با احتمال زیادی با یکدیگر مجاور بوده و با ضریب خوشه بندی مقدار دهی می‌شود[۱۳]. اما مهم‌ترین ویژگی شبکه که

[9] Small World Phenomenon
[10] Six Degrees of separation
[11] Right-skewed degree distributions
[12] Power-law
[13] Clustering
[14] Network Transivity

توجه بیشتری را به خود جلب کرده است، ویژگی ساختار تشکل[15] می‌باشد[۱۴و۱۵] که به معنی وجود گروه‌های متراکم از رأس‌ها و ارتباطات تنک بین این گروه‌ها است (به این ویژگی، خوشه بندی نیز گفته می‌شود، اما برای جلوگیری از تداخل با ویژگی قبل، از این نام استفاده نمی‌شود.). این گروه‌ها، معمولا خوشه[16]، تشکل[17]، گروه‌های به هم پیوسته[18] و یا مدول[19] نامیده می‌شوند.

جدا از ویژگی‌های آماری مشترک میان شبکه‌های اجتماعی، ویژگی دیگری را نیز می‌توان در نظر گرفت که به تازگی کانون توجهات را به خود جلب کرده و از آن به عنوان راه حلی برای بسیاری از مشکلات موجود در شبکه‌های اجتماعی نام دارد. انتشار اطلاعات[20] را می‌توان به عنوان یکی از نمونه‌های فرآیند انتشار نام برد. انتشار اطلاعات یک تعریف عمومی است که شامل هر چیزی که در یک شبکه گسترش می‌یابد می‌شود. بیشینه کردن گسترش اعتبار[21] مانند یک ایده‌ی برتر و یا حتی تشخیص سریع یک فاجعه مانند دزدی. همه این موارد نمونه-هایی از انتشار اطلاعات هستند.

در نهایت از فرآیند انتشاری که در شبکه‌های اجتماعی رخ می‌دهد، از جمله انتشار اطلاعات، جهت ارائه‌ی مدلی برای دسته‌ای از پدیده‌ها در این شبکه‌ها استفاده می‌شود. پدیده-هایی مانند گسترش ویروس‌ها و بدافزارها در کامپیوترها، گسترش اطلاعات مربوط به کالاها در میان مردم و غیره. ارائه این مدل‌های آماری می‌تواند راهنمایی باشد جهت بررسی بیشتر ساختار شبکه‌های اجتماعی، تشخیص تاثیرگذارترین رئوس در این شبکه‌ها [۱۶ و ۱۷] و همچنین نحوه گسترش و انتشار اطلاعات در آن‌ها؛ مبحثی که به تازگی مورد توجه بسیاری از محققین قرار گرفته است.

۱-۲. هدف از انجام این پایان‌نامه

[15] Community structure
[16] Cluster
[17] Community
[18] Cohesive groups
[19] Module
[20] Diffusion Process
[21] Maximizing Spread of Influence