دانشگاه پیام نور

شناسایی اجتماع و کاوُش در رسانه‌های اجتماعی

لی تانگ و هوان لیو

ترجمهٔ

دکتر محمّدرضا ماشین‌چی

امروز کتابخوانی و علم‌آموزی نه تنها یک وظیفه‌ی ملی، که یک واجب دینی است[1].

مقام معظم رهبری

در عصر حاضر یکی از شاخصه‌های ارزیابی رشد، توسعه و پیشرفت فرهنگی هر کشوری میزان تولید کتاب، مطالعه و کتاب‌خوانی مردم آن مرز و بوم است. ایران اسلامی نیز از دیرباز تاکنون با داشتن تمدنی چندهزارساله و مراکز متعدد علمی، فرهنگی، کتابخانه‌های معتبر، علما و دانشمندان بزرگ با آثار ارزشمند تاریخی، سرآمد دولت‌ها و ملت‌های دیگر بوده و در عرصه فرهنگ و تمدن جهانی به‌سان خورشیدی تابناک همچنان می‌درخشد و با فرزندان نیک‌نهاد خویش هنرنمایی می‌کند. چه کسی است که در دنیا با دانشمندان فرزانه و نام‌آور ایرانی همچون ابوعلی سینا، ابوریحان بیرونی، فارابی، خوارزمی و ... همچنین شاعران برجسته‌ای نظیر فردوسی، سعدی، مولوی، حافظ و ... آشنا نباشد و در مقابل عظمت آنها سر تعظیم فرود نیاورد. تمامی این افتخارات ارزشمند، برگرفته از میزان عشق و علاقه فراوان ملت ما به فراگیری علم و دانش از طریق خواندن و مطالعه منابع و کتاب‌های گوناگون است. به شکرانه الهی، تاریخ و گذشته ما، همیشه درخشان و پربار است. ولی اکنون در این زمینه در چه جایگاهی قرار داریم؟ آمار و ارقام ارائه‌شده از سوی مجامع و سازمان‌های فرهنگی در مورد سرانه مطالعه هر ایرانی، برایمان چندان امیدوارکننده نمی‌باشد.

کتاب، دروازه‌ای به سوی گستره دانش و معرفت است و کتاب خوب، یکی از بهترین ابزارهای کمال بشری است. همه دستاوردهای بشر در سراسر عمر جهان، تا آنجا که قابل کتابت بوده است، در میان دست‌نوشته‌هایی است که انسان‌ها پدید آورده و می‌آورند. در این مجموعه بی‌نظیر، تعالیم الهی، درس‌های پیامبران به بشر، و همچنین علوم مختلفی است که سعادت بشر بدون آگاهی از آن‌ها امکان‌پذیر نیست. کسی که با دنیای زیبا و زندگی‌بخش کتاب ارتباط ندارد بی‌شک از مهم‌ترین دستاورد انسانی و نیز از بیشترین معارف الهی و بشری محروم است. با این دیدگاه، به‌روشنی می‌توان ارزش و مفهوم رمزی عمیق در این حقیقت تاریخی را دریافت که اولین خطاب خداوند متعال به پیامبر گرامی اسلام(ص) این است که »بخوان!« و در اولین سوره‌ای که بر آن فرستاده عظیم‌الشأن خداوند، فرود آمده، نام «قلم» به تجلیل یاد

1. https://farsi.khamenei.ir/message-content?id=2696

شده‌است: «إقْرَأ وَ رَبُّکَ الْاَکْرَمُ. اَلَّذی عَلَّمَ بِالْقَلَم» در اهمیت عنصر کتاب برای تکامل جامعه انسانی، همین بس که تمامی ادیان آسمانی و رجال بزرگ تاریخ بشری، از طریق کتاب جاودانه مانده‌اند.

دانشگاه پیام‌نور با گستره جغرافیایی ایران‌شمول خود با هدف آموزش برای همه، همه‌جا و همه‌وقت، به‌عنوان دانشگاهی کتاب‌محور در نظام آموزش عالی کشورمان، افتخار دارد جایگاه اندیشه‌سازی و خردورزی بخش عظیمی از جوانان جویای علم این مرز و بوم باشد. تلاش فراوانی در ایام طولانی فعالیت این دانشگاه انجام پذیرفته تا با بهره‌گیری از تجربه‌های گرانقدر استادان و صاحب‌نظران برجسته کشورمان، کتاب‌ها و منابع آموزشی درسی شاخص و خودآموز تولید شود. در آینده هم، این مهم با هدف ارتقای سطح علمی، روزآمدی و توجه بیشتر به نیازهای مخاطبان دانشگاه پیام‌نور با جدیت ادامه خواهد داشت. به‌طور قطع استفاده از نظرات استادان، صاحب‌نظران و دانشجویان محترم، ما را در انجام این وظیفه مهم و خطیر یاری‌رسان خواهد بود. پیشاپیش از تمامی عزیزانی که با نقد، تصحیح و پیشنهادهای خود ما را در انجام این وظیفه خطیر یاری می‌رسانند، سپاسگزاری می‌نماییم. لازم است از تمامی اندیشمندانی که تاکنون دانشگاه پیام‌نور را منزلگه اندیشه‌سازی خود دانسته و ما را در تولید کتاب و محتوای آموزشی درسی یاری نموده‌اند، صمیمانه قدردانی گردد. موفقیت و بهروزی تمامی دانشجویان و دانش‌پژوهان عزیز آرزوی همیشگی ما است.

دانشگاه پیام‌نور

فهرست مطالب

پیشگفتار ..... **نُه**
چکیده ..... **یازده**

فصل اول. رسانه‌های اجتماعی و محاسبات اجتماعی **۱**
۱-۱ رسانه‌های اجتماعی ۱
۱-۲ مفاهیم و تعاریف ۳
۱-۲-۱ شبکه‌ها و نحوهٔ نمایش آن‌ها ۳
۱-۲-۲ ویژگی‌های شبکه‌های بزرگ‌مقیاس ۶
۱-۳ چالش‌ها ۸
۱-۴ عملیات محاسبات اجتماعی ۱۰
۱-۴-۱ الگوسازی شبکه ۱۰
۱-۴-۲ تحلیل مرکزیت و الگوسازی ضریب‌تأثیر ۱۱
۱-۴-۳ شناسایی اجتماع ۱۲
۱-۴-۴ توصیه و دسته‌بندی ۱۴
۱-۴-۵ حریم خصوصی، هرزنامه و امنیت ۱۵
۱-۵ خلاصه ۱۶
فصل دوم ۱۹
گره‌ها، بندها و ضریب‌تأثیر ۱۹
۲-۱ اهمیت گره‌ها ۱۹
۲-۲ استحکام بندها ۲۶
۲-۲-۱ یادگیری از ساختار شبکه ۲۶
۲-۲-۲ یادگیری از ویژگی‌ها و تعاملات کاربر ۲۸
۲-۲-۳ یادگیری از سلسله فعالیت‌های کاربر ۲۹

۲-۳ الگوسازی ضریب‌تأثیر ۳۰
۲-۳-۱ الگوی آستانهٔ خطی ۳۱
۲-۳-۲ الگوی آبشار مستقل ۳۳
۲-۳-۳ بیشینه‌سازی ضریب‌تأثیر ۳۴
۲-۳-۴ تمایز میان ضریب‌تأثیر و همبستگی ۳۸

فصل سوم. شناسایی و ارزیابی اجتماع ۴۳
۳-۱ شناسایی اجتماع گره‌محور ۴۳
۳-۱-۱ اشتراک کامل ۴۳
۳-۱-۲ دسترس‌پذیری ۴۷
۳-۲ شناسایی اجتماع گروه‌محور ۴۸
۳-۳ شناسایی اجتماع شبکه‌محور ۵۰
۳-۳-۱ تشابه رأس ۵۰
۳-۳-۲ الگوهای فضای پنهان ۵۲
۳-۳-۳ تخمین الگوی بلوکی ۵۵
۳-۳-۴ خوشه‌بندی طیفی ۵۷
۳-۳-۵ بیشینه‌سازی تفکیک ۶۰
۳-۳-۶ فرایند یکپارچه ۶۲
۳-۴ شناسایی اجتماع مرتبه‌محور ۶۳
۳-۴-۱ خوشه‌بندی سلسله‌مرتبه‌ای تقسیمی ۶۳
۳-۴-۲ خوشه‌بندی سلسله‌مرتبه‌ای تجمیعی ۶۶
۳-۵ ارزیابی اجتماع ۶۷

فصل چهارم. اجتماعات در شبکه‌های ناهمگون ۷۵
۴-۱ شبکه‌های ناهمگون ۷۵
۴-۲ شبکه‌های چند-بعدی ۷۸
۴-۲-۱ ادغام شبکه ۸۰
۴-۲-۲ ادغام سودمندی ۸۱
۴-۲-۳ ادغام ویژگی ۸۴
۴-۲-۴ ادغام افراز ۸۷
۴-۳ شبکه‌های چند-حالتی ۹۰
۴-۳-۱ هم‌خوشه‌بندی شبکه‌های دو-حالتی ۹۰
۴-۳-۲ تعمیم به شبکه‌های چند-حالتی ۹۴

فصل پنجم. کاوُش رسانه‌های اجتماعی ۹۷
۵-۱ الگوهای تکامل در رسانه‌های اجتماعی ۹۷

۵-۱-۱ راه‌کار ساده‌بینانه برای بررسی تکامل اجتماع ۹۹
۵-۱-۲ سیر تحول اجتماع در شبکه‌های تکاملی آرام ۱۰۳
۵-۱-۳ خوشه‌بندی قطعه‌ای در شبکه‌های تکاملی ۱۰۷
۵-۲ دسته‌بندی داده‌های شبکه ۱۰۹
۵-۲-۱ دسته‌بندی جمعی ۱۱۰
۵-۲-۲ یادگیری مبتنی بر اجتماع ۱۱۴
۵-۲-۳ جمع‌بندی ۱۲۰

پیوست الف. جمع‌آوری داده ۱۲۱

پیوست ب. محاسبهٔ بینابینی ۱۲۵

پیوست پ. خوشه‌بندی k- میانگین ۱۲۹

واژه‌نامه انگلیسی ـ فارسی ۱۳۳

منابع ۱۴۱

پیشگفتار

جهان، عرصهٔ شبکه‌ها است و انسان، جزئی از این شبکهٔ عظیم محسوب‌می‌شود. «شبکه‌ها در همه جا حضور دارند» و برخی بر این باورند که امروز «دوران در هم تنیدگی» و «عصر قدرت شبکه‌ها» است؛ دوره‌ای که تغییر یک چیز، تأثیری ناگزیر بر سایر چیزها دارد. انسان از نخستین روزهای خلقتش، در میان شبکه‌های گوناگونی از اجتماعی، زیستی و فناورانه حضور داشته و برتری او نه صرفاً ناشی از اندازهٔ مغزش، بلکه بیشتر حاصل توانایی‌اش در برقراری ارتباط و شبکه‌سازی بوده. ارتباطات که نخستین گام‌های آن از طریق خط و نگارش شکل‌گرفته، با شبکه‌سازی‌های مداوم و پویا سبب رشد فناوری‌های پیشرفته گشته‌است. این فناوری‌ها در رفع سه چالش اساسی بشر یعنی خشکسالی، بیماری و جنگ، نقش کلیدی ایفاءکرده‌اند. تازه‌ترین آن که موضوع این کتاب نیز می‌باشد، فناوری‌های مبتنی بر هوش مصنوعی و الگوریتم‌های هوشمند است که می‌توانند در قالب دانش شبکه‌ها به حل مسائل کمک‌کنند.

گرچه وجود شبکه‌ها از دیرباز برای انسان قابل‌درک بوده؛ اما واژهٔ «شبکه»، برای اولین بار در قرون ۱۷ و ۱۸ میلادی به‌کاررفته. در کتاب‌های این دوران، شبکه‌های طبیعی مانند تارهای عنکبوت، رگ‌های بدن و غیره مورد توجه بوده‌اند. سپس این واژه در قرن ۱۹ میلادی به‌طور رسمی‌تر استفاده‌شده؛ چنان‌که نقشهٔ خطوط راه‌آهن و آب‌راهه‌ها، یا توصیف روابط انسان‌ها، به شبکه تشبیه‌شده‌اند. در قرن ۲۰، مفهوم شبکه‌ها به سرعت گسترش‌یافت و با آغاز از حوزه‌هایی همچون حمل و نقل، برق،

تلفن، تلویزیون و کامپیوتر، نهایتاً در حوزهٔ شبکه‌های اجتماعی آنلاین تثبیت‌شد. همگام با افزایش استفاده از شبکه‌ها، داده‌ها نیز افزایش‌یافتند و برای پردازش آن‌ها، نیاز به روش‌های نوین ایجادشد. این روش‌ها و الگوریتم‌های هوشمند آن‌ها می‌توانند با پردازش داده‌های انبوه، به‌طور مؤثر از ظرفیت دانش شبکه‌ها بهره‌ببرند؛ و این خود منجر به افزایش استفاده از این الگوریتم‌ها شده و تنیدگی آن‌ها را با زندگی انسان، روزافزون کرده‌است. هم‌زمان با این تنیدگی و فواید حاصل از برتری آن‌ها نسب به انسان، بخش‌های تاریک سیطرهٔ الگوریتم‌ها و صاحبان آنها نیز نمایان‌شده و به همین دلیل، شناخت، توسعه و بکارگیری صحیح آن‌ها ضرورت‌یافته‌است. در همین راستا، این کتاب به معرفی برخی الگوریتم‌های بنیادین و روش‌های تحلیل شبکه‌ها پرداخته‌است. بدیهی‌است که برای بهره‌گیری کامل از این دانش، باید مطالعات گسترده‌ای نسبت به منابع این کتاب و تازه‌های علمی این حوزه داشت.

کتاب حاضر در مقطع کارشناسی ارشد برای رشتهٔ مهندسی کامپیوتر با گرایش علوم داده، به عنوان منبع درس داده‌کاوی پیشرفته تعیین‌شده‌است. از آنجایی‌که این کتاب در میان کتب تخصصی فارسی، گشایشی در باب دانش شبکه‌ها می‌باشد، ترجمهٔ آن با دقت فراوان انجام‌شده و برخی مثال‌ها نیز برای فهم بهتر، بومی‌سازی و بازنویسی‌شده‌اند. همچنین واژه‌گزینی‌ها بر اساس مفاهیم تخصصی آنها صورت‌گرفته و گاه برای شفافیت یک موضوع، توضیحاتی در پاورقی اضافه شده‌است. با همهٔ این اوصاف، کمال مطلق از آنِ خالق شبکهٔ جهان هستی است و این کتاب، صرفاً برساختهٔ عضو کوچکی از شبکهٔ بزرگ اوست. بنابراین امید است که خوانندگان محترم، کاستی‌های احتمالی را بر مترجم بخشیده و برای بهبود نسخه‌های آتی، از ارسال نظرات ارزشمند خود دریغ ننمایند.

دکتر محمدرضا ماشین‌چی
r_mashinchi@yahoo.com
mashinchi@pnu.ac.ir

چکیده

در دهۀ گذشته، توسعه و تغییرات سریعی در حوزۀ وب مشارکتی و رسانه‌های اجتماعی رخ‌داده‌است و در پی آن از طریق نوآوری‌های مختلف، ارتباطات نزدیک‌تری میان افراد ایجادشده‌است. میلیون‌ها کاربر به‌صورت برخط در حال بازی، نشان‌گذاری، کار و معاشرت هستند و بدین‌ترتیب انواع جدیدی از همکاری‌ها، ارتباطات و هوشمندی‌های غیرقابل‌تصور شکل‌گرفته‌اند. رسانه‌های اجتماعی می‌توانند به بازنمایی الگوهای کسب‌وکار کمک‌کنند، بر روی احساسات و نظرات تأثیربگذارند، و تعاملات انسانی و رفتارهای جمعی را در مقیاس بی‌سابقه‌ای مورد مطالعه قراردهند. این کتاب ابتدا از منظر داده‌کاوی به معرفی ویژگی‌های رسانه‌های اجتماعی می‌پردازد و سپس با مرور عملیات اصلی موجود در محاسبات رسانه‌های اجتماعی، چالش‌ها را تشریح‌می‌کند. علاوه بر آن در کنار معرفی مفاهیم پایه، با ذکر مثال به ارایۀ الگوریتم‌های جدید پرداخته و روش‌هایی برای ارزیابی کارآمد پیشنهاد نموده‌است. در این راستا، به‌طور ویژه به فنون شناسایی اجتماع مبتنی بر گراف پرداخته‌شده و مطالب لازم برای کار با شبکه‌های پویا و ناهمگون در رسانه‌های اجتماعی تشریح‌شده‌است. این کتاب، نحوۀ استفاده از الگوهای اجتماعی کشف‌شده را برای کاوش رسانه‌های اجتماعی نشان‌می‌دهد. مفاهیم، الگوریتم‌ها و روش‌هایی که در این کتاب ارائه‌شده‌اند را می‌توان برای بهره‌برداری از قدرت شبکه‌های اجتماعی و ایجاد سیستم‌های هوش جمعی به‌کارگرفت. این کتاب، مقدمه‌ای بر «شناسایی اجتماع و کاوش در شبکه‌های اجتماعی»

است و مطالعهٔ آن برای افرادی مانند دانشجویان، پژوهشگران و یا فعالینی که از رسانه‌های اجتماعی به عنوان یک منبع داده در رشته یا حوزهٔ عملیاتی خود استفاده‌می‌کنند ضروری‌است. خوانندگان این کتاب می‌توانند با فهم مطالب، به‌طور مطلوب از مفاهیم استفاده‌نموده و به نوآوری دست‌یابند.

کلیدواژه‌ها

رسانه‌های اجتماعی، شناسایی اجتماع، کاوش رسانه‌های اجتماعی، تحلیل مرکزیت، استحکام بندها، الگوسازی ضریب‌تأثیر، انتشار اطلاعات، بیشینه‌سازی ضریب‌تأثیر، همبستگی، همسان‌گزینی، ضریب‌تأثیر، ارزیابی اجتماع، شبکه‌های ناهمگون، شبکه‌های چند بعدی، شبکه‌های چند–حالتی، تکامل اجتماع، دسته‌بندی جمعی، ابعاد اجتماعی، مطالعهٔ رفتاری.

فصل اول

رسانه‌های اجتماعی و محاسبات اجتماعی

۱-۱ رسانه‌های اجتماعی

در دهۀ گذشته شاهد توسعه و تغییرات سریعی در حوزۀ وب و اینترنت بوده‌ایم. برنامه‌های کاربردی مشارکتی و سایت‌های شبکه‌های اجتماعی، به‌طور مداوم افزوده‌می‌شوند و با اتصال افراد به یکدیگر، انواع جدیدی از همکاری‌ها و ارتباطات را میان آن‌ها شکل‌می‌دهند. اکنون بسیاری از افراد به‌طور داوطلبانه به نگارش مقالات آنلاین می‌پردازند که پیش از این از حیث محور و مقیاس، هرگز قابل‌تصور نبود. فروشگاه‌های آنلاین نیز برای فروش کالاهایشان از خِرد جمعی بهره‌می‌گیرند، و به خریدها و نظرات کاربران رجوع‌می‌کنند. حتی جنبش‌های سیاسی برای مشارکت‌ها و فعالیت‌های جمعی خود، از همین شیوه‌های جدید استفاده‌می‌کنند.

جدول ۱-۱، برخی از انواع رسانه‌های اجتماعی را نشان‌می‌دهد که شامل بلاگ‌ها، تالارهای گفت‌وگو، رسانه‌های بستر اشتراک‌گذاری، ریز بلاگ‌ها، شبکه‌های اجتماعی، اخبار اجتماعی، نشان‌گذاری‌های اجتماعی و ویکی‌ها است. در پس این موارد به ظاهر متفاوت، یک ویژگی مشترک نهفته‌است که سبب تمایز رسانه‌های اجتماعی از وب سنّتی می‌شود؛ و آن این است که «مصرف‌کنندگان محتوا، تولیدکنندۀ اطلاعات و دانش نیز می‌باشند».

در رسانه‌های سنّتی مانند رادیو، تلویزیون، فیلم و روزنامه، فقط بخش کوچکی از «مسئولین» یا «متخصصین» هستند که برای چیستی و چگونگی انتشار اطلاعات تصمیم‌می‌گیرند. در این رسانه‌ها، اکثر کاربران فقط نقش مصرف‌کننده دارند و از فرایند تولید حذف‌شده‌اند. این رسانه‌ها دارای الگوی ارتباطی یک‌طرفه بوده و جهت آن از تولیدکنندۀ مرکزی به انبوه مصرف‌کنندگان است.

جدول ۱-۱. انواع مختلف رسانه‌های اجتماعی.

بلاگ‌ها	Wordpress, Blogspot, LiveJournal, BlogCatalog
تالارهای گفت‌وگو	Yahoo! answers, Epinions
رسانه‌های بستر اشتراک‌گذاری	Flickr, YouTube, Justin.tv, Ustream, Scribd
ریزبلاگ‌سازی	Twitter, Foursquare, Google buzz
شبکه‌سازی اجتماعی	Facebook, MySpace, LinkedIn, Orkut, PatientsLikeMe
اخبار اجتماعی	Digg, Reddit
نشان‌گذاری اجتماعی	Del.icio.us, StumbleUpon, Diigo
ویکی‌ها	Wikipedia, Scholarpedia, Ganfyd, AskDrWiki

اما در رسانه‌های اجتماعی، مصرف‌کننده می‌تواند نقش تولیدکننده را نیز ایفاءکند. حضور صدها میلیون کاربر در انواع سایت‌های رسانه‌های اجتماعی، هر کاربر را به یک رسانهٔ خبری تبدیل‌کرده‌است (Shirky, 2008). در این نوع انتشار انبوه، اخبار به‌هنگام و اطلاعات مردمی تولیدمی‌شود و با قرارگرفتن در صدر «محتواهای تولیدشدهٔ کاربران»، «خِرد جمعی» شکل‌می‌گیرد. نمونهٔ آن، حملهٔ تروریستی لندن در سال ۱۳۸۳ است (Thelwall, 2006) که برخی از شاهدان، اقدام به ثبت خبرهای دست‌اول خود در وبلاگ‌ها کردند. رسانه‌های اجتماعی، نگارش مشارکتی را نیز فراهم‌نموده‌اند تا بتوان از طریق آن، آثار با کیفیتی را که بدون چنین نوع نگارشی ناممکن‌است تولیدکرد. مثلاً «پس از به‌وجودآمدن ویکی‌پدیا در سال ۱۳۷۹، این سایت به‌سرعت به یکی از بزرگ‌ترین وب‌سایت‌های مرجع تبدیل‌شده؛ و از سال ۱۳۸۷ تقریباً هرماه ۶۵ میلیون نفر بازدیدکننده داشته و در آن بیش از ۸۵،۰۰۰ نویسنده، در بیش از ۲۶۰ زبان، روی بیش از ۱۴،۰۰۰،۰۰۰ مقاله مشغول به فعالیت هستند[۱و۲].»

تفاوت دیگر رسانه‌های اجتماعی، غنای موجود در تعاملات کاربران آن‌ها می‌باشد. موفقیت این رسانه‌ها، وابسته به مشارکت کاربران آن‌ها بوده و افزایش مشارکت کاربران نیز حاصل افزایش تعاملات کاربران است، و به عکس. به عنوان مثال، فیسبوک مدعی‌است که تا مرداد ۱۳۸۹، بیش از ۵۰۰ میلیون کاربر فعال[۳] داشته‌است. مشارکت کاربران، یکی از عوامل اصلی موفقیت برای رسانه‌های اجتماعی

---

۱. اکنون در سال ۱۴۰۴، ویکی‌پدیا دارای ۱۱۸ هزار نویسنده در ماه، با بیش از ۳۰۰ زبان و ۶۴ میلیون مقاله است. ـ م.

2. http://en.wikipedia.org/wiki/Wikipedia:About

3. http://www.facebook.com/press/info.php?statistics

محسوب‌می‌شود. این عامل سبب‌شده تا هشت رسانهٔ اجتماعی جدول ۱-۲، در جمع ۲۰ سایت برتر قرارگیرند (به استناد گزارش اَلِکسا[1] برای ترافیک اینترنت در دوازدهم مرداد ۱۳۸۹). اتصال کاربران به یکدیگر، از طریق تعاملات میان آن‌ها صورت‌می‌گیرد و همین تعاملات باعث تشکیل شبکه‌های کاربران می‌شود. وجود این شبکه‌ها سبب‌شده تا امکان مطالعهٔ تعاملات انسانی و رفتارهای جمعی، در مقیاسی بی‌سابقه ولی با چالش‌های محاسباتی فراوان، مهیاشود که مستلزم توسعه فنون محاسباتی و الگوریتم‌های پیشرفته است.

در این درس، مفاهیم پایه‌ای شبکه‌های اجتماعی که در شناسایی اجتماع به‌کارمی‌روند توضیح‌داده‌می‌شوند و با مثال‌های ساده، الگوریتم‌های مرز دانش مورد استفاده در تحلیل داده‌های رسانه‌های اجتماعی، تشریح‌می‌شوند. این کتاب، یک منبع جامع است و به مباحث مهم شناسایی اجتماع در رسانه‌های اجتماعی می‌پردازد.

جدول ۱-۲. اسامی ۲۰ وب‌سایت برتر در آمریکا.

رتبه	سایت	رتبه	سایت
۱	google.com	۱۱	blogger.com
۲	facebook.com	۱۲	msn.com
۳	yahoo.com	۱۳	myspace.com
۴	youtube.com	۱۴	go.com
۵	amazon.com	۱۵	bing.com
۶	wikipedia.org	۱۶	aol.com
۷	craigslist.org	۱۷	linkedin.com
۸	twitter.com	۱۸	cnn.com
۹	ebay.com	۱۹	espn.go.com
۱۰	live.com	۲۰	wordpress.com

۱-۲ مفاهیم و تعاریف

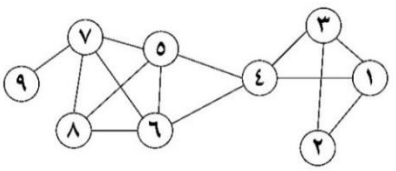
داده‌های شبکه از داده‌های مقدار ویژگی متفاوت‌اند و دارای خصوصیات ویژهٔ خود می‌باشند.

۱-۲-۱ شبکه‌ها و نحوهٔ نمایش آن‌ها

شبکهٔ اجتماعی، یک ساختار اجتماعی است که از گره‌ها (افراد یا سازمان‌ها) و یال‌ها

1. Alexa

تشکیل‌شده‌است. یال‌ها، گره‌ها را از طریق روابطی همچون دوستی، خویشاوندی و غیره به هم وصل‌می‌کنند. نمایش یک شبکه معمولاً دو روش صورت‌می‌گیرد: یکی، نمایش با گراف برای دیداری‌سازی مانند شبکهٔ شکل ۱–۱ با ۹ بازیگر و دیگری، نمایش با ماتریس («ماتریس اجتماعی» (Faust, 1994) یا «ماتریس مجاورت») مانند جدول ۳–۱. توجه‌کنید که یک شبکهٔ اجتماعی، غالباً دارای پراکندگی زیاد است و نمایش آن در جدول، همراه با تعداد زیادی صفر می‌باشد. این پراکندگی می‌تواند برای تحلیل کارآمد شبکه‌ها استفاده‌شود. در ماتریس مجاورت، درایه‌های روی قطر مشخص‌نمی‌شوند. مطابق تعریف، درایه‌های روی قطر نشان‌دهندهٔ خود–پیوندی گره‌ها هستند. خود–پیوندی نیز به‌معنای اتصال یک گره به خود می‌باشد. در تحلیل یک شبکه اصولاً درایه‌های روی قطر، با صفر مقداردهی‌می‌شوند؛ اما در برخی حالات نیز باید به یک تغییریابند. از این پس اگر مقدار درایه به‌صراحت ذکرنشود، بطور پیش‌فرض آن را با صفر مقداردهی‌می‌کنیم.



شکل ۱–۱. یک شبکهٔ اجتماعی متشکل از ۹ بازیگر و ۱۴ ارتباط. قطر شبکه، ۵ است و ضرایب خوشه‌بندی گره‌های ۱ تا ۹ عبارت‌اند از $C_۱ = ۲/۳$, $C_۲ = ۱$, $C_۳ = ۲/۳$, $C_۴ = ۱/۳$, $C_۵ = ۲/۳$, $C_۶ = ۲/۳$, $C_۷ = ۱/۳$, $C_۸ = ۱$, $C_۹ = ۰$. میانگین ضریب خوشه‌بندی ۰٫۶۱ است درحالی‌که ضریب خوشه‌بندی مورد انتظار برای یک گراف تصادفی با ۹ گره و ۱۴ یال، برابر ۰٫۱۹=(۹×۸/۲)/۱۴.

جدول ۱–۳. ماتریس مجاورت.

گره	۱	۲	۳	۴	۵	۶	۷	۸	۹
۱	–	۱	۱	۱	۰	۰	۰	۰	۰
۲	۱	–	۱	۰	۰	۰	۰	۰	۰
۳	۱	۱	–	۱	۰	۰	۰	۰	۰
۴	۱	۰	۱	–	۱	۱	۰	۰	۰
۵	۰	۰	۰	۱	–	۱	۱	۱	۰
۶	۰	۰	۰	۱	۱	–	۱	۱	۰
۷	۰	۰	۰	۰	۱	۱	–	۱	۱
۸	۰	۰	۰	۰	۱	۱	۱	–	۰
۹	۰	۰	۰	۰	۰	۰	۱	۰	–

شبکه‌ها دارای انواع مختلفی مانند وزن‌دار، علامت‌دار و یا جهت‌دار می‌باشند. در شبکه‌های وزن‌دار، یال‌ها دارای مقادیر عددی هستند؛ در شبکه‌های علامت‌دار، یال‌ها می‌توانند دارای روابط مثبت یا منفی باشند؛ در شبکه‌های جهت‌دار، یال‌ها دارای جهت هستند. مثلاً شبکهٔ شکل ۱-۱ بدون‌جهت است و لذا، ماتریس مجاورت آن نیز متقارن می‌باشد؛ اما در برخی سایت‌های رسانه‌های اجتماعی، تعاملات به‌صورت جهت‌دار است. برای مثال در توئیتر[1]، اگرچه کاربرِ x دنبال‌کنندهٔ کاربر y است اما کاربرِ y الزاماً دنبال‌کنندهٔ کاربرِ x نیست. در چنین حالتی، شبکهٔ دنبال‌کننده-دنبال‌شونده از نوع جهت‌دار و نامتقارن است. تمرکز این درس به‌جز در موارد خاص، بر ساده‌ترین شکل شبکه‌ها یعنی شبکه‌های غیرجهت‌دار با وزن یال بولی است. مثال جدول ۱-۳ نمونه‌ای از این شبکه‌ها است. درعین‌حال بسیاری از فنون ارائه‌شده در این درس را می‌توان به شبکه‌های وزن‌دار، علامت‌دار و جهت‌دار نیز تعمیم‌داد.

نکته: در اینجا برای اشاره به مجموعهٔ گره‌ها و یال‌ها در یک شبکه، به‌ترتیب از V و E استفاده‌می‌کنیم. تعداد گره‌ها در شبکه، n و تعداد یال‌ها، m می‌باشد. ماتریس $A \in \{۰,۱\}^{n\times n}$ نشان‌دهندهٔ ماتریس مجاورت برای یک شبکه است. درایهٔ $A_{i,j} \in \{۰,۱\}$ نشان‌دهندهٔ وجود یا عدم‌وجود پیوند میان گره‌های v_i و v_j است. یال میان دو گرهٔ v_i و v_j را با $e(v_i, v_j)$ نشان‌می‌دهیم. اگر $A_{ij} = ۱$ آنگاه گره‌های v_i و v_j در «مجاورت» یکدیگر قراردارند. N_i نشان‌دهندهٔ همهٔ گره‌های مجاور گرهٔ v_i است؛ یعنی گره‌های واقع در همسایگی گرهٔ v_i. تعداد گره‌های مجاور گرهٔ v_i را درجهٔ آن گره می‌نامند (که با d_i نشان‌داده‌می‌شود). مثلاً در شبکهٔ شکل ۱-۱، $d_۱ = ۳$ و $d_۴ = ۴$. اگر گرهٔ یک یال، از نوع پایانه‌ای باشد آنگاه آن یال، در «مجاورت» آن گره قراردارد. به عنوان مثال، $e(۱,۴)$ در مجاورت هر دو گرهٔ ۱ و ۴ قراردارد.

کوتاه‌ترین مسیر میان دو گره (مثلاً v_i و v_j)، «ژئودزیک» نامیده‌می‌شود. «فاصلهٔ ژئودزیک» میان دو گره (که با $g(v_i, v_j)$ نشان‌داده‌می‌شود)، تعداد پَرش‌های موجود در ژئودزیک است. در مثال حاضر، چون ژئودزیک $(۸-۶-۴-۳-۲)$ وجوددارد بنابراین $g(۲,۸) = ۴$. نمادهای پُرتکرار این کتاب، در جدول ۱-۴ آمده‌است.

1. Twitter

جدول ۱-۴. فهرست علائم.

نماد	مفهوم
A	ماتریس مجاورت یک شبکه
V	مجموعهٔ گره‌ها در شبکه
E	مجموعهٔ یال‌ها در شبکه
N	تعداد گره‌ها $(n = \|V\|)$
m	تعداد یال‌ها $(n = \|E\|)$
v_i	یک گرهٔ v_i
$\mathrm{e}(v_i, v_j)$	یال میان گره‌های v_i و v_j
A_{ij}	در صورت وجود یال میان گره‌های v_i و v_j، مقدار آن ۱ و در غیراین‌صورت ۰ است.
N_i	همسایگی (یعنی گره‌های همسایه) در گرهٔ v_i
d_i	درجهٔ گرهٔ v_i $(d_i = \|N_i\|)$
ژئودزیک	کوتاه‌ترین مسیر میان دو گره
فاصلهٔ ژئودزیک	طول کوتاه‌ترین مسیر
$g(v_i, v_j)$	فاصلهٔ ژئودزیک میان گره‌های v_i و v_j

۱-۲-۲ ویژگی‌های شبکه‌های بزرگ‌مقیاس

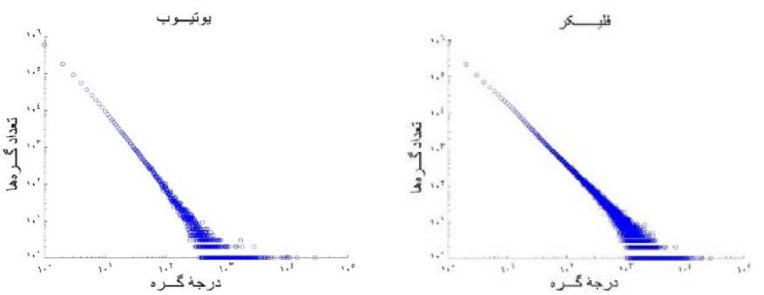
شبکه‌های رسانه‌های اجتماعی معمولاً شامل میلیون‌ها بازیگر و ارتباط می‌باشند. شبکه‌های بزرگ‌مقیاس دارای الگوهای مشترکی هستند که به‌ندرت در شبکه‌های کوچک دیده‌می‌شوند. مشهورترین الگوها عبارت‌اند از «توزیع بی‌مقیاس»، «اثر جهان کوچک» و «ساختار اجتماع قوی». شبکه‌هایی که حاوی ویژگی‌های هم‌پیوندی مبهم باشند را «شبکه‌های پیچیده» می‌نامند. دلیل این نام‌گذاری، ایجاد تمایز میان این شبکه‌ها و شبکه‌های ساده‌ای مانند گراف مُشبّکه یا گراف‌های تصادفی است.

معمولاً در شبکه‌های بزرگ‌مقیاس، درجهٔ گره‌ها از توزیع قانون توان پیروی‌می‌کند. مثلاً در شکل ۱-۲، میزان درجه به ازای تعداد گره در شبکه‌های یوتیوب[1] و فلیکِر[2] آورده‌شده. مطابق این شکل، اکثر گره‌ها درجهٔ کمی دارند و تنها تعداد اندکی از آن‌ها دارای درجهٔ به‌شدت زیاد هستند (مثلاً، درجه $> ۱۰^۴$). این دو شبکه در مقیاس لگاریتم-لگاریتم برای توزیع درجهٔ گره‌ها (درجهٔ گره‌ها نسبت به تعداد گره‌ها) دارای الگوی مشابه (تقریباً خطی یا خط مستقیم) می‌باشند. چنین الگویی را «توزیع قانون توان» یا «توزیع بی‌مقیاس» می‌نامند. به این دلیل که تغییر در مقیاس، منجر به تغییر

1. http://www.youtube.com/
2. http://www.flickr.com/

شکل توزیع، نمی‌شود. قابل‌توجه اینکه اگر بر روی دنبالهٔ آن دقت‌کنیم (مثلاً گره‌های دارای درجه<۱۰۰ را بررسی‌نمائیم) می‌توان همچنان توزیع قانون توان را مشاهده‌کرد. این خود-همانندی، مستقل از مقیاس است. به شبکه‌هایی که درجات گره‌های آن دارای توزیع قانون توان باشد، «شبکه‌های بی‌مقیاس» گفته‌می‌شود.

ویژگی برجستهٔ دیگر در شبکه‌های اجتماعی، «اثر جهان کوچک» است. در پژوهشی که توسط (Milgram, 1969) صورت‌گرفت، میانگین طول مسیر در شبکه‌های اجتماعی آمریکائیان بررسی‌شد. بدین‌ترتیب که از منتخبان خواسته‌شده‌بود یک نامهٔ زنجیره‌ای برای آشنایان خود ارسال‌کنند. چنانکه از فردی در شهر اوماها در ایالت نبراسکا یا شهر ویچیتا در ایالت کانزاس شروع‌شود و نهایتاً پس از طی مکان‌های مختلف، به دست فرد موردنظر در شهر بوستون در ایالت ماساچوست برسد. مطابق بررسی، ۶۴ نامه به مقصد رسید و میانگین طول مسیر آن‌ها ۵٫۵ یا تقریباً ۶ به‌دست‌آمد. درنتیجهٔ این پژوهش، قانون «تفکیک شش درجه‌ای» مشهور حاصل‌شد. این قانون اخیراً برای یک شبکهٔ پیام‌رسان فوری در مقیاس جهانی و با بیش از ۱۸۰ میلیون کاربر نیز مورد تأیید قرارگرفت. بدین‌صورت که میانگین طول مسیر میان هر دو فرد، ۶٫۶ به‌دست‌آمد (Faloutsos, 2006). دقت اثر جهان کوچک را می‌توان با سنجه‌ای به‌نام قطر شبکه افزایش‌داد. «قطر»، بزرگ‌ترین ژئودزیک در شبکه است (Faust, 1994). قطر شبکه در شکل ۱-۱، برابر ۵ می‌باشد (معادل با فاصلهٔ ژئودزیک میان گره‌های ۲ و ۹). اکثر شبکه‌های بزرگ‌مقیاس در دنیای واقعی، دارای قطر کوچکی هستند.



شکل ۱-۲. توزیع گره‌ها در دو شبکه. یکی، شبکه‌ای در یوتیوب با ۱۱۳۸۴۹۹گره و دیگری شبکه‌ای در فلیکر با ۱۷۱۵۲۵۵ یال است (براساس داده‌های مرجع (Bhattacharjee, 2007)). محورهای X و Y در مقیاس لگاریتم هستند. نمودار پراکندگی برای درجات گره‌ها برحسب توزیع قانون توان تقریباً یک خط مستقیم است. میانگین ضریب خوشه‌بندی برای هر نمودار برابر ۰٫۰۸ و ۰٫۱۸ می‌باشد. درحالی‌که اگر ارتباطات از نوع تصادفی یکنواخت باشند آنگاه ضریب مورد انتظار در یک گراف تصادفی برابر $۴٫۶ \times ۱۰^{-۶}$ و $۱٫۰ \times ۱۰^{-۵}$ خواهدبود.

شبکه‌های اجتماعی، دارای «ساختار اجتماعی» قوی هستند؛ بدین‌معنی که افراد علاقه‌دارند به‌جای ارتباط با خارج از گروه خود، با افراد درون گروه خود ارتباط برقرارکنند؛ چنان‌که دوستان یک دوست نیز احتمالاً با یکدیگر دوست هستند؛ بنابراین، این انتقال‌پذیری می‌تواند با «ضریب خوشه‌بندی» اندازه‌گیری‌شود. این ضریب، برابر با تعداد ارتباطات میان دوستان یک فرد نسبت به تعداد کل ارتباطات ممکن میان آن‌ها است. فرض‌کنید گرهٔ v_i دارای d_i همسایه است و تعداد k_i یال نیز میان این همسایگان وجوددارد. در این‌صورت ضریب خوشه‌بندی مطابق زیر است:

$$C_i = \begin{cases} \dfrac{k_i}{d_i \times \left(d_i - ۱\right)/۲} & d_i > ۱ \\ ۰ & d_i = ۰ \text{ یا } ۱ \end{cases} \qquad (۱-۱)$$

ضریب خوشه‌بندی، میزان تراکم ارتباطات میان دوستان یک فرد را اندازه‌گیری‌می‌کند. ضریب خوشه‌بندی در شبکهٔ حاوی اجتماع، دارای متوسط بسیار بالاتری نسبت به شبکهٔ تصادفی است. برای مثال در شکل ۱-۱، گرهٔ ۶ دارای چهار همسایهٔ ۴، ۵، ۷ و ۸ می‌باشد و میان آن‌ها چهار ارتباط وجوددارد که عبارت‌اند از $e\left(۴,۵\right)$، $e\left(۵,۷\right)$، $e\left(۵,۸\right)$ و $e\left(۷,۸\right)$. لذا ضریب خوشه‌بندی برای گرهٔ ۶، برابر ۲/۳ =۴/(۴×۳/۲). میانگین ضریب خوشه‌بندی شبکه نیز ۰٫۶۱ است؛ اما ضریب خوشه‌بندی مورد انتظار برای یک گراف تصادفی با همان تعداد گره و ارتباط، برابر ۰٫۱۹=۱۴/(۹×۸/۲).

۱-۳ چالش‌ها

میلیون‌ها کاربر به‌صورت آنلاین، مشغول بازی، کار و روابط اجتماعی هستند. این سیل عظیم از داده‌ها، فرصت بی‌سابقه‌ای برای تحلیل شبکه‌های اجتماعی بزرگ‌مقیاس با بیش از میلیون‌ها بازیگر را فراهم‌کرده. نمونه‌هایی از این شبکه‌ها عبارت‌اند از شبکه‌های ارتباطات نامه‌های الکترونیکی (Carley, 2005)، شبکه‌های پیام‌رسان‌های فوری (Horvitz, 2008)، شبکه‌های تماس‌های تلفن‌های همراه (Joshi, 2006) و شبکه‌های دوستی (Bhattacharjee, 2007). نمونه‌های دیگری از این شبکه‌ها نیز مورد آزمون قرارگرفته‌اند و الگوهای یکسانی را نشان‌داده‌اند. از جملهٔ آن‌ها شبکهٔ ارجاع یا نویسندهٔ

همکار، شبکهٔ بیولوژیکی، شبکهٔ مسیرهای متابولیک، و شبکهٔ تنظیم ژنتیکی و وب غذایی است (Newman, Barabási et al., 2011). بدین‌ترتیب، رسانه‌های اجتماعی دریچهٔ تازه‌ای را برای مطالعهٔ روابط انسانی گشوده‌اند.

هنگامی‌که شبکه‌های بزرگ‌مقیاس در بستر ویژگی‌های منحصربه‌فرد شبکه‌های اجتماعی قرارمی‌گیرند، کاوش شبکه‌های اجتماعی با چالش‌های جدیدی روبه‌رو می‌شود. در ادامه برخی از این چالش‌ها را بررسی‌می‌کنیم:

مقیاس‌پذیری: حجم شبکه‌های رسانه‌های اجتماعی می‌تواند مقیاسی برابر با میلیون‌ها بازیگر و صدها میلیون ارتباط داشته‌باشد. درحالی‌که تحلیل در شبکه‌های اجتماعی سنّتی معمولاً بر مبنای مقیاسی کمتر از چند صد سوژه است؛ بنابراین فنون موجود برای تحلیل شبکه در رویارویی مستقیم با چنین اندازهٔ نجومی از شبکه‌ها ناکارآمد است.

ناهمگونی: در دنیای واقعی، روابط متعددی میان افراد وجوددارد. دو فرد به‌طور هم‌زمان می‌توانند دوست و همکار باشند؛ بنابراین میان بازیگران یک مجموعهٔ واحد در شبکه، تعاملات گوناگونی برقرار است. ضمن اینکه ممکن‌است انواع متعددی از موجودیت‌ها نیز پدیدآیند. میان کاربران، نشان‌ها و محتواهای بسیاری از سایت‌های نشان‌گذاری اجتماعی و اشتراک‌گذاری رسانه‌ای، درهم‌تنیدگی وجوددارد. این درهم‌تنیدگی باعث ایجاد مقادیر ناهمگون در شبکه می‌شود. برای تحلیل شبکه‌های ناهمگون که حاوی موجودیت‌ها یا تعاملات ناهمگون هستند، نیاز به نظریه‌ها و ابزارهای نوین است.

تکامل: به‌هنگام‌بودن در شبکه‌های اجتماعی از اهمیت ویژه‌ای برخوردار است. مثلاً در سایت‌های اشتراک‌گذاری محتوا و دنیای وبلاگ‌نویسی، افراد به‌سرعت علاقهٔ خود را نسبت به محتواهای به‌اشتراک گذاشته‌شده یا مطالب وبلاگ از دست می‌دهند. همین ویژگی، شبکه‌های اجتماعی را از وب‌کاوی سنّتی متفاوت‌می‌کند. در شبکه‌های اجتماعی ابتدا کاربران جدید واردمی‌شوند، سپس ارتباطات تازه‌ای میان اعضاء شکل‌می‌گیرد و نهایتاً کاربران قدیمی، غیرفعال‌می‌شوند یا کاملاً آنجا را ترک‌می‌کنند. اکنون این پرسش‌ها مطرح‌می‌شوند که چگونه می‌توان تحرکات افراد شبکه را رصدکرد؟ آیا می‌توان اعضای ماندگار را که پایه‌های اجتماع هستند پیداکرد؟ آیا این افراد می‌توانند در ظهور و سقوط اجتماع خود نقش‌داشته‌باشند؟