دانشگاه شیخ‌بهائی

دانشکده فنی‌مهندسی

پایان‌نامه کارشناسی ارشد رشته مهندسی کامپیوتر- نرم افزار

ارائه روش بهبود یافته شناسایی اجتماعات معنادار همپوشان در شبکه‌های اجتماعی

پژوهشگر

فریبرز نحوی

استاد راهنما

دکتر محمدرضا خیام باشی

مهر۱۳۹۲

بسم الله الرحمن الرحیم

پایان نامه کارشناسی ارشد رشته مهندسی کامپیوتر - نرم افزار

ارائه روش بهبودیافته شناسایی اجتماعات معنادار همپوشان در شبکه‌های اجتماعی

چکیده

با توجه به گسترش شبکه‌های اجتماعی، شناسایی اجتماعات که به عنوان بستر پنهان شبکه‌های پیچیده بشمار می‌آیند، از اهمیت خاصی برخوردار شده است. اکثر روش‌های کلاسیک، با دید ماکروسکوپی، کل ساختار گراف را در نظر گرفته و اقدام به شناسایی اجتماعات می‌نمایند. این روش‌ها، علیرغم دقت بالا، زمان اجرا بسیار بالایی نیز داشته و در شبکه‌های بزرگ در دنیای واقعی، قابل استفاده نمی‌باشند. بمنظور رفع این مشکل، الگوریتم‌های شناسایی محلی اجتماعات توسعه یافتند. از آنجایی که الگوریتم‌های محلی، تک‌تک گره‌های گراف را در نظر گرفته و ارتباطات بین آن‌ها را بررسی می‌نمایند، پیچیدگی زمانی پایین‌تری نسبت به روش‌های سراسری دارند. لیکن، پایین بودن کیفیت این روش‌ها، علیرغم سرعت بالا، آن‌ها را به سمت استفاده از معیارهای سراسری گراف جهت افزایش دقت سوق داده است که تهدیدی برای ماهیت الگوریتم‌های محلی بشمار می‌آید.

در این پژوهش، ابتدا الگوریتمی بمنظور شناسایی محلی اجتماعات همپوشان با ایده گرفتن از روش‌های بهینه‌سازی حریصانه محلی معرفی گردیده است. شروع الگوریتم از یک هسته اولیه متراکم، عدم امکان انحراف فرآیند شناسایی اجتماعات محلی حول یک گره، قابلیت شناسایی گره‌های دورافتاده و قطب، کاهش تعداد چرخه‌های الگوریتم و در نتیجه کاهش پیچیدگی زمانی آن، شناسایی گره‌های مشترک و اجتماعات همپوشان بروش سراسری و عدم نیاز به تعیین مقادیر، پیش از شروع الگوریتم، از جمله قابلیت‌های الگوریتم معرفی شده می‌باشد.

همچنین، الگوریتم پیشنهادی، با استفاده از خاصیت هموفیلی شبکه‌های پیچیده، ایده استفاده از ویژگی چگالی همسایگان مشترک دو گره در شبکه‌های اجتماعی را مطرح نموده است. این ایده که بر مبنای خاصیت هموفیلی شبکه‌های پیچیده بنا نهاده شده و توسط معیار مشابهت ساختاری جاکارد قابل محاسبه است، دو فرد در یک شبکه اجتماعی را هر چقدر دارای دوستان مشترک بیشتری باشند، به لحاظ شخصیتی شبیه‌تر می‌پندارد. از دیدگاه الگوریتم پیشنهادی، مشابهت گره‌های گراف، نه به عنوان معیاری مجزا، بلکه به عنوان بستر و زیرساخت ارتباط بین گره‌ها می‌باشد. به همین دلیل، الگوریتم پیشنهادی، با وزن‌دار نمودن یال‌های گراف، در شناسایی اجتماعاتی که نه تنها به لحاظ ساختاری با یکدیگر ارتباط قوی دارند، بلکه به لحاظ معنایی نیز با یکدیگر مرتبطند اقدام می‌کند.

نتایج حاصل از مقایسه زمان اجرای الگوریتم پیشنهای و روش‌های مشابه بر روی مجموعه داده‌های آزمایشی دنیای واقعی و نیز محک تحلیلی، حاکی از بهبود زمان اجرای الگوریتم می‌باشد. همچنین، استفاده از مشابهت ویژگی‌های گره‌ها در شناسایی اجتماعات، منجر به کاهش آنتروپی و همگن‌تر شدن اجتماعات شناسایی شده نسبت به روش‌های مشابه و نیز نسبت به الگوریتم پیشنهادی، بدون استفاده از ویژگی‌ها می‌باشد.

کلمات کلیدی

اجتماعات همپوشان محلی، معیار مشابهت ویژگی ، خاصیت هموفیلی، اجتماعات معنادار.

فهرست مطالب

۱- مقدمه ۲

۱-۱- تکامل مفهوم اجتماع ۳

۱-۲- مسئله تحقیق ٤

۱-۳- اهداف تحقیق ٥

۱-٤- اهمیت و ارزش تحقیق ٥

۱-٥- کاربرد نتایج تحقیق ٦

۱-٦- فرضیه‌ها ۷

۱-۷- جمع‌بندی ۷

۱-۸- ساختار پایان‌نامه ۷

۲- ادبیات تحقیق ۹

۲-۱- گره‌ها ۱۰

۲-۱-۱- گره همسایه ۱۰

۲-۱-۲- گره حاشیه‌ای ۱۰

۲-۱-۳- گره پل / قطب ۱۰

۲-۱-٤- گره دورافتاده ۱۱

۲-۲- اجتماعات ۱۲

۲-۲-۱- اجتماع ضعیف ۱۲

۲-۲-۲- اجتماع قوی ۱۲

۲-۳- شبکه‌ها ۱۲

۲-۳-۱- شبکه های متراکم ۱۳

۲-۳-۲- شبکه‌های متصل کامل ۱۳

۲-۳-۳- شبکه های تنک ۱۳

۲-۳-٤- شبکه‌های همگن و ناهمگن ۱۳

۲-٤- شبکه‌های پیچیده ۱۳

۲-٤-۱- ضریب Assortativity ۱٤

۲-٤-۲- هموفیلی ۱٤

۲-٤-۳- انتقال‌پذیری ۱۵
۲-٤-٤- جهان کوچک ۱۵
۲-٤-۵- ساختار سلسله‌مراتبی ۱۵
۲-٤-٦- ساختار اجتماع ۱٦
۲-٤-۷- ارتجاع شبکه ۱٦
۲-۵- انواع شبکه‌های پیچیده ۱٦
۲-۵-۱- شبکه‌های تصادفی ۱۷
۲-۵-۲- شبکه‌های Power-Law ۱۸
۲-۵-۳- شبکه‌های جهان کوچک ۱۹
۲-٦- ساختار اجتماع محلی ۲۰
۲-٦-۱- یال‌های اجتماع محلی ۲۱
۲-٦-۲- چگالی زیرگراف ۲۱
۲-۷- معیارهای توقف گسترش روش‌های محلی ۲۱
۲-۷-۱- معیار اجتماع قوی ۲۲
۲-۷-۲- معیار P-Strong ۲۲
۲-۷-۳- معیار Trailing ۲۳
۲-۸- معیار پیمانه‌ای ۲٤
۲-۸-۱- محاسبه معیار پیمانه‌ای ۲۵
۲-۸-۲- مشکلات معیار پیمانه‌ای ۲۵
۲-۹- معیارهای بازیابی اطلاعات ۲٦
۲-۹-۱-Precision ۲۷
۲-۹-۲-Recall ۲۷
۲-۹-۳-F-Measure ۲۸
۲-۹-٤-Accuracy ۲۸
۲-۱۰- محک NMI ۲۸
۲-۱۱- محک NMI همپوشان ۲۹
۲-۱۲- مشابهت ساختاری گره‌ها ۳۰

۲-۱۲-۱- معرفی معیارهای مشابهت ساختاری ۳۱
۲-۱۲-۲- مقایسه معیارهای مشابهت ساختاری ۳۳
۲-۱۳- محاسبه مشابهت ویژگی‌ها ۳٤
۲-۱۳-۱- معیار مشابهت ویژگی NAS ۳٤
۲-۱٤- جمع‌بندی و نتیجه‌گیری ۳۵
۳- پیشینه تحقیق ۳۷
۳-۱- روش‌های شناسایی اجتماعات معنادار ۳۸
۳-۱-۱- روش شناسایی بر مبنای نزدیکی ۳۸
۳-۱-۲- روش بر مبنای مدل ۳۸
۳-۱-۳- روش SCAN ۳۹
۳-۱-٤- روش K-SNAP ۳۹
۳-۱-۵- روش SA-Cluster ۳۹
۳-۱-٦- روش SAC1 ٤۰
۳-۱-۷- روش SAC2 ٤۰
۳-۲- روش‌های بهینه‌سازی محلی حریصانه ٤۱
۳-۲-۱- معیار R ٤۱
۳-۲-۲- معیار R′ ٤۳
۳-۲-۳- معیار M ٤۳
۳-۲-٤- معیار L ٤۵
۳-۲-۵- معیار T ٤٦
۳-۲-٦- شناسایی بر مبنای مشابهت یال‌ها ٤٦
۳-۲-۷- شناسایی بر مبنای هسته محلی ٤۷
۳-۳- روش‌های بهینه‌سازی محلی غیر حریصانه ٤۸
۳-۳-۱- معیار برون‌گرایی ٤۸
۳-۳-۲- روش CONGO ٤۹
۳-۳-۳- روش بهینه‌سازی نهایی ٤۹
۳-٤- روش‌های بهینه‌سازی محلی حریصانه همپوشان ۵۰

۳-٤-۱- روش Baumes ۵۰
۳-٤-۲- روش LFM ۵۰
۳-٤-۳- روش LOFO ۵۱
۳-٤-٤- روش Wu ۵۳
۳-٤-۵- روش MONC ۵٤
۳-٤-٦- روش Chen ۵٤
۳-۵- روش‌های شناسایی محلی همپوشان بر مبنای دسته ۵۵
۳-۵-۱- روش EAGLE ۵۵
۳-۵-۲- روش GCE ۵۵
۳-۵-۳- روش EM-BOAD ۵٦
۳-٦- معیار پیمانه‌ای همپوشان ۵۷
۳-٦-۱- معیار پیمانه‌ای همپوشان وزن‌دار Qo ۵۷
۳-٦-۲- معیار پیمانه‌ای همپوشان Li ۵۸
۳-٦-۳- معیار پیمانه‌ای همپوشان Shen ۵۸
۳-۷- جمع‌بندی و نتیجه‌گیری ۵۹
٤- راه‌حل پیشنهادی ٦۱
٤-۱- ساختار روش‌های بیشینه‌سازی محلی حریصانه ٦۲
٤-۲- روند اجرای راه‌حل پیشنهادی ٦۳
٤-۳- مشکلات روش‌های جاری ٦۵
٤-۳-۱- انتخاب گره شروع ٦۵
٤-۳-۲- عدم شناسایی اجتماع محلی حول گره ٦٦
٤-۳-۳- افزودن گره‌های دورافتاده به اجتماع ٦٦
٤-۳-٤- نیاز به تعیین پارامترها پیش از شروع الگوریتم ٦۷
٤-۳-۵- گره‌های بی‌اجتماع ٦۸
٤-٤- معیار مشابهت گره‌ها در روش پیشنهادی ٦۸
٤-٤-۱- مشکلات روش‌های موجود ۷۰
٤-٤-۲- همبستگی معیار مشابهت ویژگی و معیار مشابهت ساختاری ۷۰

٤-٤-٣- انتخاب معیار مشابهت گره‌ها ٧٢
٤-٤-٤- معیار شناسایی هسته اولیه ٧٣
٤-٤-٥- معیار برازندگی F' ٧٥
٤-٥- قوانین روش پیشنهادی ٧٨
٤-٥-١- قانون ١: انتخاب گره شروع ٧٨
٤-٥-٢- قانون ٢ : تعیین هسته اولیه ٧٩
٤-٥-٣- قانون ٣: تعیین مجموعه کاندید ٧٩
٤-٥-٤- قانون ٤: وزن دار نمودن همسایگان اجتماع ٧٩
٤-٥-٥- قانون ٥: به‌روزرسانی مجموعه‌ها ٧٩
٤-٥-٦- قانون ٦: شناسایی سراسری گره‌های قطب ٨٠
٤-٥-٧- قانون ٧: شناسایی گره‌های بی اجتماع ٨٠
٤-٥-٨- قانون ٨: شناسایی گره‌های همپوشان ٨٠
٤-٦- جمع‌بندی و نتیجه‌گیری ٨٢
٥- ارزیابی و تحلیل نتایج ٨٤
٥-١- محاسبه پیچیدگی زمانی ٨٥
٥-١-١- ارزیابی پیچیدگی زمانی ٨٦
٥-٢- ارزیابی زمان اجرای روش پیشنهادی ٨٦
٥-٣- مقایسه پیچیدگی زمانی با روش‌های شناسایی سراسری اجتماعات ٨٨
٥-٤- ارزیابی تعداد اجتماعات شناسایی شده ٨٩
٥-٥- ارزیابی تعداد گره‌های پل شناسایی شده ٨٩
٥-٦- ارزیابی ادغام اجتماعات همپوشان ٩٠
٥-٧- ارزیابی کمینه، میانگین و بیشینه اندازه اجتماعات شناسایی شده ٩١
٥-٨- ارزیابی روش بر روی محک‌های تحلیلی ٩٢
٥-٩- ارزیابی روش بر روی مجموعه داده واقعی ٩٣
٥-١٠- ارزیابی مشابهت ویژگی‌های اجتماعات ٩٦
٥-١١- ارزیابی کیفیت بخش‌بندی روش پیشنهادی٢ ..... ٩٧
٥-١٢- ارزیابی تعداد گره‌های دورافتاده شناسایی شده ٩٩

۵-۱۳- نتیجه‌گیری ۹۹
٦- نتیجه‌گیری و راه‌کارهای آینده ۱۰۳
٦-۱- فرضیات تحقیق ۱۰٤
٦-۲- کارهای آینده ۱۰۵
٦-۲-۱- پیشنهاد اول ۱۰۵
٦-۲-۲- پیشنهاد دوم ۱۰۵
٦-۲-۳- پیشنهاد سوم ۱۰٦
مراجع ۱۰۷

فهرست شکل‌ها

شکل ۱-۱- نمایی از سه اجتماع فرضی به همراه یال‌های بین آن‌ها ۳
شکل ۱-۲- نمایی از چهار نوع رسانه‌های اجتماعی مختلف ٤
شکل ۱-۳- چهار فاز الگوریتم پیشنهادی ٥
شکل ۲-۱- گره حاشیه‌ای A ۱۰
شکل ۲-۲- گره ٦ گره قطب و گره ۱۳ گره دورافتاده است ۱۱
شکل ۲-۳- گره دورافتاده L در اجتماع سمت راست ۱۱
شکل ۲-٤- نمایی از دو شبکه با ضرایب متفاوت ASSORTAITIVITY ۱٤
شکل ۲-٥- نمایی از ساختار سلسله مراتبی گراف نمونه ۱٥
شکل ۲-٦- شبکه دارای ساختار اجتماع در سمت چپ و شبکه بدون ساختار اجتماع در سمت راست ۱٦
شکل ۲-۷- توزیع درجه گره‌ها در شبکه های تصادفی ۱۷
شکل ۲-۸- نمایی از یک شبکه تصادفی، اکثر گره‌ها به سه یا چهار گره متصل هستند ۱۸
شکل ۲-۹- توزیع درجه گره‌ها در شبکه های واقعی ۱۹
شکل ۲-۱۰- در شبکه‌های SCALE-FREE درصد کمی از گره‌ها تعداد بسیاری یال دارند ۱۹
شکل ۲-۱۱- نمایی از یک شبکه جهان کوچک ۱۹
شکل ۲-۱۲- اجزای مختلف اجتماع محلی C ۲۰
شکل ۲-۱۳- تغییر دقت الگوریتم با تغییر مقدار Z_{OUT} و کاهش وضوح مرزهای اجتماع ۲۲
شکل ۲-۱٤- تغییر دقت روش با تغییر مقدار Z_{OUT} و مقایسه آن با مقدار M_{OUT} ۲۳
شکل ۲-۱٥- کاهش معیار پیمانه‌ای در اثر ابهام در مرزهای اجتماعات ۲٥
شکل ۲-۱٦- نمایی از گراف با سه نوع کره ۲۷
شکل ۲-۱۷- دو گره هم ارز ۸ و ۹ ..... ۳۱
شکل ۳-۱- نمایی از اجتماع محلی D، هسته مرکزی C و گره‌های حاشیه‌ای B ٤۱
شکل ۳-۲- ارتباط بین درجه گره مبدأ و مقدار پیمانه‌ای محلی ٤۳
شکل ۳-۳- مقادیر مختلف معیار M پس از افزودن گره ۱۱ ..... ٤٥
شکل ۳-٤- مثالی از انتخاب بین گره‌های حاشیه‌ای ۱ و ۱ جهت افزوده شدن به اجتماع محلی داخلی ٤۹
شکل ۳-٥- نمایی از یک دسته 5-CLIQUE ٥٥
شکل ٤-۱- زیرمجموعه‌های روش‌های بیشینه سازی محلی حریصانه ٦۲
شکل ٤-۲- شبه کد شناسایی اجتماعات محلی در الگوریتم‌های حریصانه ٦۳
شکل ٤-۳- نمودار روند اجرای راه‌حل پیشنهادی به همراه فازهای آن ٦٤
شکل ٤-٤ -نمایش گره‌های با مرکزیت بالا در مقابل گره‌های با مرکزیت پایین ٦٦
شکل ٤-٥- گراف آزمایشی ۲ ..... ٦۷
شکل ٤-٦- مشکل وجود سطح آستانه در تعیین وضعیت گره‌های همپوشان بین دو اجتماع ٦۷
شکل ٤-۷- گراف همکاری نویسندگان به همراه سه روش خوشه‌بندی ٦۹
شکل ٤-۸- رابطه بین تعداد کلمات مشابه و اجتماعات در فلیکر ۷۱

شکل ٤-٩- میانگین مشابهت دو دوست در مقایسه با دو فرد تصادفی ٧١
شکل ٤-١٠- همبستگی مقادیر پیمانه‌ای مشابهت ساختاری و ویژگی ٧٢
شکل ٤-١١- ضریب خوشه‌بندی گره پایین گراف ٧٥
شکل ٤-١٢- نمایی از اجتماع محلی D و بخش‌های مختلف آن ٧٨
شکل ٥-١- مقایسه زمان اجرای روش پیشنهادی بر حسب دقیقه در سی بار اجرا ٨٧
شکل ٥-٢- مقایسه تعداد اجتماعات شناسایی شده در سی بار اجرای روش ٨٩
شکل ٥-٣- مقایسه تعداد گره‌های پل شناسایی شده در چهار روش ٩٠
شکل ٥-٤- مقایسه تعداد اجتماعات شناسایی شده پیش و پس از ادغام ٩١
شکل ٥-٥- مقایسه بیشینه، کمینه و میانگین اندازه اجتماعات شناسایی شده ٩١
شکل ٥-٦- نمایش اجرای روش پیشنهادی بر روی مجموعه داده ZACHARY ٩٤
شکل ٥-٧- اجرای روش LFM بر روی مجموعه ZACHARY ٩٥
شکل ٥-٨- اجرای روش Wu بر روی مجموعه ZACHARY ٩٥
شکل ٥-٩- اجرای روش EM-BOAD بر روی مجموعه ZACHARY ٩٥
شکل ٥-١٠- اجرای روش REES بر روی مجموعه ZACHARY ٩٦
شکل ٥-١١- مقایسه مقادیر آنتروپی در حالت استفاده و عدم استفاده از مشابهت گره‌ها ٩٧
شکل ٥-١٢- ایده تکثیر گره‌های مشترک جهت محاسبه معیار پیمانه‌ای همپوشان ٩٨
شکل ٥-١٣- مقایسه معیار پیمانه‌ای همپوشان الگوریتم پیشنهادی ٢ ..... ٩٨
شکل ٥-١٤- مقایسه تعداد گره‌های دورافتاده شناسایی شده بر روی شبکه NCAA ٩٩

فهرست جداول

جدول ۲-۱- ماتریس آشفتگی..... ۲۷
جدول ۲-۲- معیارهای مشابهت ساختاری گراف..... ۳۲
جدول ۵-۱- مقایسه پیچیدگی زمانی روش‌های محلی حریصانه همپوشان ۸۶
جدول ۵-۲- مقایسه پیچیدگی زمانی الگوریتم‌های سراسری..... ۸۸
جدول ۵-۳- پارامترهای تنظیمی محک LFR همپوشان جهت‌دار ۹۳
جدول ۵-٤- مقایسه تعداد و نوع گره‌های شناسایی شده الگوریتم‌ها..... ۹٤

فصل اول

مقدمه

۱- مقدمه

شبکه‌های همکاری[1]، اینترنت، وب و شبکه‌های زیست‌شناختی[2] نمونه‌هایی از شبکه‌های پرکاربرد هستند که از دو بخش مهم یال‌ها و گره‌ها تشکیل شده‌اند. در این شبکه‌ها، گره‌ها نماد موجودیت و یال‌ها نمایانگر ارتباط بین آن‌ها می‌باشد. به مجموعه ای از گره‌های شبکه که دارای ارتباطات قوی در بین خود و ارتباطات تنک[3] با گره‌های بیرونی هستند، اجتماع اطلاق می‌گردد و یا در تعریفی دیگر، در فرهنگ لغات "وبستر[4]"، مجموعه‌ای از افراد که بر اساس علاقه مشترکی به یکدیگر مرتبط شده‌اند اجتماع نامیده می‌شوند. برخلاف کاربرد وسیع اجتماعات در حوزه فیزیک و کامپیوتر، هنوز تعریف جامعی در این خصوص ارائه نشده است. مفهوم اجتماع، به دو صورت صریح و ضمنی[5] کاربرد دارد. در مدل صریح، فرد با اراده خود، به عضویت اجتماع درآمده و یا از آن خارج می‌گردد. حال آن‌که در مدل ضمنی، عضویت، بر اساس محاورات و فعالیت‌های فرد شکل می‌گیرد و نیازمند روش‌های خاصی جهت کشف الگوهای ارتباطی می‌باشد. این الگوها تحت عنوان الگوریتم‌های شناسایی اجتماعات شناخته می‌شوند.

---

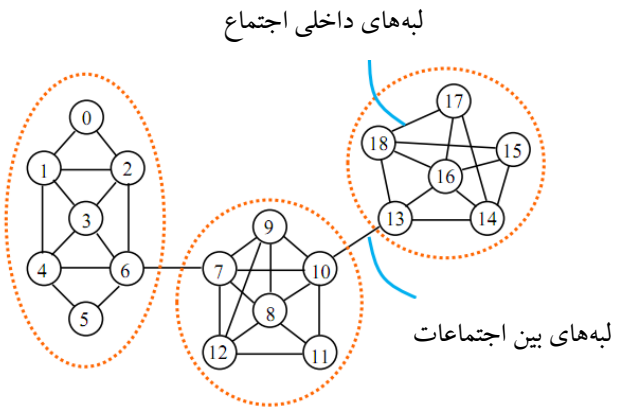
1 Collaboration Network
2 Biological Network
3 Sparse
4 Webster
5 Implicit

۱-۱- تکامل مفهوم اجتماع

با تکامل اینترنت سه نوع اجتماع به وجود آمدند[۵].

- اجتماع تحت وب[1]: در وب یک، اجتماع، به مردمی اطلاق می‌گردید که دور هم جمع شده و با داشتن یک هدف یکسان، صفحات وبی را تشکیل و به یکدیگر ارتباط می‌دادند.
- در وب ۲، به مجموعه‌ای از اشیا تحت وب شامل مستندات و کاربران که دارای ساختار منطقی و معنایی خاص خود هستند، اجتماع گفته می‌شود.
- ساختار اجتماع، گروهی از گره‌ها هستند که ویژگی‌های مشترک و یا نقش مشترکی را در عملیات سازمان ایفا می‌کنند.

وجود ارتباط بین اعضای اجتماع جزو اصول اولیه اجتماع است. بنابراین مردمی که در یک ایستگاه اتوبوس قراردارند ولی هیچ‌گونه محاوره‌ای با یکدیگر ندارند نمی‌توانند تشکیل یک اجتماع دهند. درحالی‌که مردمی که از طریق یک رسانه اجتماعی[2] با یکدیگر در حال ارتباط هستند، تشکیل یک اجتماع می‌دهند. بنابراین، اجتماع را می‌توان زیرمجموعه‌ای از کنش‌گران[3] دارای رابطه نسبتاً قوی، مستقیم و تکراری دانست. با توجه به موضوع، می‌توان تعاریف متفاوتی از اجتماع ارایه نمود. برای مثال، در شبکه‌های اجتماعی، به گروهی از مردم که علاقه‌ای را به اشتراک می‌گذارند، اجتماع اطلاق می‌گردد. در شبکه‌های زیستی، گروه‌های کارکردی[4] نشان‌دهنده اجتماع هستند و در وب، یک خوشه موضوعی از صفحات وب، یک اجتماع را تشکیل می‌دهند. در شکل ۱-۱ نمایی از نحوه ارتباط سه اجتماع فرضی، دارای یال‌های بین اجتماع کم و یال‌های داخلی متراکم نمایش داده شده است.



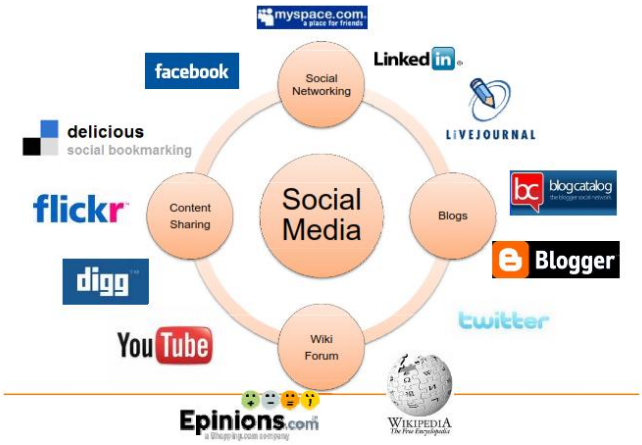
شکل ۱-۱- نمایی از سه اجتماع فرضی به همراه یال‌های بین آن‌ها [۴۳]

[1] Web Community
[2] Social Media
[3] Actor
[4] Functional Group

شناسایی مجموعه گره‌های با ویژگی‌های فوق را شناسایی اجتماعات، گروه‌بندی، خوشه‌بندی و یا یافتن زیرگروه‌های منسجم[1] می‌نامند. کاربردهای شناسایی اجتماعات عبارتند از:

- ارایه پیشنهاد به کاربران، مانند پیشنهاد دوستان از طریق تحلیل آماری انتخاب‌های گذشته بقیه اعضای اجتماع محلی حول وی
- تحلیل روش و سرعت شیوع بیماری‌های واگیردار در سطح جامعه
- مانیتور نمودن رفتار یک گره از طریق بررسی اجتماع محلی حول وی
- تحلیل و شناسایی رفتار مردم
- تشکیل بستری به منظور فعالیت‌های دیگر نظیر داده‌کاوی
- توانایی فهم و تحلیل گروه‌ها از طریق بصری نمودن[2] شبکه

در شکل ۱-۲، انواع رسانه‌های اجتماعی شامل شبکه‌های اجتماعی، اجتماعات ویکی[3]، وبلاگ‌ها و سایت‌های اشتراک محتوا[4] نمایش داده شده است.



شکل ۱-۲- نمایی از چهار نوع رسانه‌های اجتماعی مختلف [۳]

۱-۲- مسئله تحقیق

علیرغم روش‌های مختلفی که تاکنون جهت شناسایی اجتماعات ارایه شده است، روش بهینه‌ای که در همه کاربردها بهترین جواب را بدهد وجود ندارد. بلکه روش مورد استفاده بر اساس نوع کاربرد، شبکه و منابع محاسباتی متغیر می‌باشد. در یک شبکه می‌توان اجتماعات مختلفی را شناسایی نمود، لیکن، هدف از ارایه روش‌های مختلف شناسایی اجتماعات، یافتن بهترین ساختار اجتماع می‌باشد. لیکن رسیدن به این هدف نیازمند بررسی کلیه حالت‌ها می‌باشد تا از طریق آن بتواند بهترین ساختاری که دارای مقدار پیمانه‌ای بیشینه است را شناسایی نماید.

---

1 Cohesive Subgroups
2 Visualization
3 Wiki
4 Content Sharing

دست‌یابی به چنین اجتماعی، یک مسئله NP-Complete به‌شمار می‌آید. بنابراین بایستی روشی ارایه گردد که بتواند در زمان منطقی اجتماعات با کیفیتی را شناسایی نماید.

هدف از این پژوهش ارایه روشی چندفازه شامل فازهای شروع، گسترش، ارزیابی مجدد و توقف می‌باشد که علاوه بر استفاده از ویژگی‌های[1] گره‌های گراف به منظور شناسایی اجتماعات معنادار، همپوشانی بین اجتماعات را نیز در نظر بگیرد. لیکن با توجه به آنکه در شبکه‌های اجتماعی، تعداد گره‌های گراف بسیار زیاد می‌باشد، امکان شناسایی این اجتماعات با روش‌های کلاسیک به صورت ترتیبی و در زمان منطقی وجود ندارد. به همین دلیل، از معیار پیمانه‌ای محلی[2] استفاده می‌شود. خروجی چنین الگوریتمی اجتماعات معنادار و همپوشانی هستند که با استفاده از معیار پیمانه‌ای محلی در مدت زمان منطقی و قابل قبولی شناسایی می‌شوند.

۱-۳- اهداف تحقیق

هدف از این پژوهش آنست که با استفاده از معیار پیمانه‌ای محلی که پایه و اساس روش‌های شناسایی موازی هستند، روشی ارایه گردد که اجتماعات معنادار همپوشان در شبکه‌های اجتماعی را در مدت زمان قابل قبول بسته به نوع کاربرد شناسایی نماید. بدین منظور، روش ارایه شده، علاوه بر استفاده از ساختار گراف، از ویژگی‌های گره‌ها نیز جهت خوشه‌بندی گراف استفاده می‌کند تا اجتماعات شناسایی‌شده به لحاظ معنایی با یکدیگر مرتبط باشند. همچنین با شناسایی اجتماعات همپوشان، همانند دنیای واقعی که در آن یک فرد می‌تواند همزمان عضو چندین اجتماع باشد، مفهوم اجتماعات شناسایی شده به اجتماعات دنیای واقعی نزدیک‌تر می‌گردد. از آنجایی که پیاده‌سازی چنین الگوریتمی با روش‌های کلاسیک که هر دو ویژگی معنادار بودن و همپوشانی را به صورت همزمان پشتیبانی نماید، پیچیدگی زمانی بسیار بالایی خواهد داشت، از معیار بهینه محلی استفاده می‌گردد. چهار فاز الگوریتم در شکل ۱-۳ نمایش داده شده است.

شکل ۱-۳- چهار فاز الگوریتم پیشنهادی

۱-٤- اهمیت و ارزش تحقیق

در جامعه گروه‌های مختلفی مانند خویشاوندان، همکاران و دوستان وجود دارند. از سوی دیگر، پراکندگی اینترنت منجر به ایجاد گروه‌های مجازی مانند گروه دوستداران، گروه موسیقی و غیره شده است که بر بستر وب در حال فعالیت می‌باشند. برخی از این گروه‌ها به صورت مشخص ایجاد شده، کاربران عضو آن‌ها می‌شوند و فعالیت

[1] Attribute
[2] Local Modularity

می‌کنند. لیکن برخی دیگر جنبه پنهان تحلیل شبکه اجتماعی را نشان می‌دهند که کاربران بدون آنکه مطلع باشند از جنبه‌های مختلف عضو گروه‌های متفاوتی می‌باشند.

با توجه به گسترش روز افزون ابعاد شبکه های اجتماعی و تعداد کاربران و محتوای تولید شده توسط آن‌ها اعم از متن، فیلم، عکس و ... امکان تحلیل این شبکه‌ها به روش‌های سنتی تحلیل شبکه های اجتماعی وجود ندارد. بنابراین نیاز به وجود روش‌هایی می‌باشد که با شناسایی اجتماعات در کم‌ترین زمان ممکن بتواند بهترین تصمیمات را بگیرد. برای مثال تحلیل برخط شبکه های موبایل امری بسیار ضروری می‌باشد تا شرکت‌های مخابراتی با تغییر روش‌های مسیریابی خود، ترافیک را در بخش‌های مختلف شبکه به صورت پارامتریک تغییر دهند. یکی از این روش‌های شناسایی اجتماعات در کمترین زمان ممکن، شناسایی آن‌ها به صورت موازی می‌باشد.

از طرف دیگر، در حال حاضر اکثریت روش‌های شناسایی اجتماعات بر اساس ساختار شبکه اقدام به تصمیم‌گیری می‌کنند و ویژگی‌های گره‌های شبکه هیچ نقشی در این خصوص ندارد. برای مثال در روش‌های تحلیل شبکه اجتماعی فعلی، روش شناسایی یک شبکه مصنوعی از مستندات تگ شده و یا شبکه‌ای از پروتئین‌های داخل سلول‌های بدن انسان یکسان می‌باشد. این امر موجب از بین رفتن بخش‌های با معنی شبکه می‌شود.

۱-۵- کاربرد نتایج تحقیق

کاربردهای اصلی این پژوهش عبارتند از: سیستم‌های پیشنهاد دهنده، تبلیغات هدفمند در شبکه‌های اجتماعی از طریق شناسایی کنش‌گران اصلی شبکه، پیشگویی و مدل کردن رفتار کاربران، مطالعه بیماری‌های مسری و فیلترینگ جمعی[1] بر اساس علایق کاربران مشابه.

[1] Collaborative Filtering

۱-٦- فرضیه‌ها

در این پژوهش، سعی می‌گردد با اثبات فرضیه‌ها (سوالات تحقیق) ذیل، اولین روش شناسایی محلی اجتماعات همپوشان با استفاده از مشابهت ویژگی‌های گره‌های گراف ارایه گردد.

- آیا معیار پیمانه‌ای محلی، اجتماعات موجود در شبکه‌های اجتماعی را در زمانی قابل قبول بسته به نوع کاربرد و با پیچیدگی زمانی کمتر نسبت به معیار پیمانه‌ای سراسری شناسایی می‌نماید؟
- آیا خوشه‌بندی بر اساس مشابهت گره‌ها، موجب شناسایی اجتماعات معنی‌دار به لحاظ محتوایی می‌گردد؟
- آیا به‌کارگیری مکانیزم شناسایی زیرگرافها، منجر به شناسایی اجتماعات همپوشان شبکه‌های اجتماعی می‌گردد؟

۱-۷- جمع‌بندی

در این فصل، ابتدا اهداف، اهمیت و ارزش پژوهش حاضر ارایه گردید. سپس برخی کاربردهای روش‌های مشابه بیان شده و در پایان، فرضیه‌هایی که پژوهش حاضر بر مبنای آن‌ها شکل گرفته و در صدد پاسخ به آن‌ها می‌باشد ارایه شده است. در فصل بعد، مفاهیم پایه‌ای مورد استفاده در این‌دسته از روش‌ها بیان می‌گردد.

۱-۸- ساختار پایان‌نامه

در فصل دوم، ادبیات موضوع تحقیق شامل مفاهیم پایه‌ای مورد استفاده در روش‌های شناسایی اجتماعات بیان می‌گردد. فصل سوم به بررسی روش‌های مشابه در حوزه شناسایی اجتماعات، شامل روش‌های شناسایی محلی، روش‌های شناسایی اجتماعات همپوشان و روش‌های معنادار اختصاص دارد. در فصل چهارم، ابتدا مشکلات و کمبودهای روش‌های موجود با جزییات کامل مورد بررسی قرار گرفته و سپس روش پیشنهادی ارایه شده است. در فصل پنجم، روش پیشنهادی مورد ارزیابی قرار گرفته و نتایج حاصل از این ارزیابی با روش‌های مشابه و مرجع مقایسه شده است. فصل ششم نیز به نتیجه‌گیری و ارایه دیدگاه در خصوص روش‌های آینده اختصاص دارد.