

# تشخیص اجتماع مبتنی بر موضوع، گامی نوین در تحلیل شبکه‌های اجتماعی

علی ریحانیان

دانشکده فناوری اطلاعات، دانشگاه علوم و فنون مازندران، بابل، ایران

areihanian@ustmb.ac.ir

حسین علیزاده

دانشکده مهندسی کامپیوتر، دانشگاه علم و صنعت ایران، تهران، ایران

halizadeh@iust.ac.ir

بهروز مینایی

دانشکده مهندسی کامپیوتر، دانشگاه علم و صنعت ایران، تهران، ایران

b\_minaei@iust.ac.ir

## چکیده

یکی از مهم‌ترین مباحث در زمینه‌ی تحلیل شبکه‌های اجتماعی، مبحث تشخیص اجتماع می‌باشد. در این مقاله، از میان رویکردهای مختلفی که در پژوهش‌های علمی گوناگون به منظور تشخیص اجتماع پیشنهاد شده‌اند، دو رویکرد مهم بررسی می‌شود. رویکرد اول، تشخیص اجتماع موضوعی نامیده می‌شود. رویکرد دوم در این مقاله، تشخیص اجتماع با رویکرد ترکیبی نامیده می‌شود. به منظور تشخیص دقیق تر اجتماعات، رویکرد دوم، فرایند انجام شده در رویکردهای تشخیص اجتماع موضوعی و تشخیص اجتماع ساختاری – که در این مقاله بررسی نمی‌شود – را با هم ترکیب می‌کند و به عمل تشخیص اجتماع می‌پردازد. نتیجه‌ای که با بررسی رویکرد اول و همچنین در نظر گرفتن رویکرد تشخیص اجتماع ساختاری حاصل می‌شود این است که اجتماعات شناسایی شده توسط روش‌های تشخیص اجتماع ساختاری، اغلب شامل زیراجتماعاتی هستند که به لحاظ موضوعی متنوعند و از طرف دیگر، اجتماعات شناسایی شده توسط روش‌های تشخیص اجتماع موضوعی، اغلب شامل زیراجتماعاتی هستند که به لحاظ ساختاری متنوعند. این موضوع، لزوم وجود رویکرد ترکیبی را اثبات می‌کند. اجتماعاتی که با بکارگیری این رویکرد شناسایی می‌شوند نسبت به اجتماعات شناسایی شده توسط رویکردهای موضوعی و ساختاری با معناترند زیرا اعضای این اجتماعات هم به لحاظ ساختاری بصورت تنگاتنگ به یکدیگر متصلند و هم دارای علائق موضوعی یکسان می‌باشند.

## کلمات کلیدی:

شبکه اجتماعی، اجتماع، تشخیص اجتماع موضوعی، تشخیص اجتماع با رویکرد ترکیبی

## مقدمه

با اهمیت تبدیل شد [۱]. ساختار اجتماع<sup>۱</sup> چنین شبکه‌های

پیچیده‌ای، هم سازمان‌دهی این شبکه‌ها و هم ارتباطات

با فراگیر شدن شبکه‌های اجتماعی مثل facebook، تجزیه و

تحلیل داده‌های چنین شبکه‌هایی به یک موضوع تحقیقاتی

<sup>1</sup> Community structure

ساختار مقاله بدین صورت می‌باشد که در بخش بعدی، به بررسی برخی از مفاهیم اولیه در شبکه‌های پیچیده پرداخته می‌شود. در بخش سوم، به بررسی رویکرد تشخیص اجتماع موضوعی پرداخته شده و کارهای مرتبط انجام شده در این زمینه بررسی می‌شوند. در بخش چهارم، تشخیص اجتماع با رویکرد ترکیبی بررسی می‌شود. در بخش آخر نیز به بیان نتایج حاصل از این پژوهش پرداخته می‌شود.

## ۱- تعریف مفاهیم اولیه مورد نیاز

در این بخش به بررسی مفاهیم اولیه مورد نیاز جهت مطالعه‌ی شبکه‌های اجتماعی پرداخته می‌شود:

### ۱-۱- تعریف ریاضی شبکه

به عنوان یک نمایش ریاضی از یک شبکه، می‌توان آن را به صورت یک گراف  $G = \{V, E\}$  نمایش داد. این گراف، شامل مجموعه‌ای از گره‌ها به صورت  $V = \{v_1, v_2, \dots, v_n\}$  و مجموعه‌ای از یال‌ها به صورت  $E = \{e_1, e_2, \dots, e_m\}$  می‌باشد. در این گراف، هر یال  $e_j$  بصورت  $e_j = \{v_{j1}, v_{j2}\}$  تعریف می‌شود که در آن  $v_{j1} \neq v_{j2}$  و  $v_{j1}, v_{j2} \in V$  می‌باشد.

### ۱-۲- تعریف درجه گره

در گراف  $G = \{V, E\}$ ، همسایگان گره  $v \in V$ ، گره-هایی هستند که توسط یالی مانند  $e \in E$  به  $v$  متصلند. مجموع تمام یال‌هایی که به گره  $v$  متصلند را درجه آن گره گویند. درجه گره  $v$  را با  $k_v$  نمایش می‌دهند.

### ۱-۳- فاصله در شبکه

فاصله‌ی بین دو گره در شبکه، طول کوتاه‌ترین مسیر بین آن‌هاست. طول مسیر بین دو گره نیز برابر تعداد یال‌هایی است که در حرکت از یک گره به گره دیگر پیموده می‌شود.

### ۱-۴- نحوه نمایش شبکه

می‌توان شبکه‌ها را به ۲ صورت زیر نمایش داد [۳۵]:

پنهان بین اجزای تشکیل دهنده‌ی آن‌ها را آشکار می‌سازد [۲]. بنابراین تشخیص اجتماع، به عنوان یک موضوع مهم در تجزیه و تحلیل شبکه‌های اجتماعی مطرح می‌شود. نقش تشخیص اجتماع، جستجو برای پیدا کردن اجتماعات است [۳]. احتمال این‌که افرادی که در این شبکه‌ها به اجتماعات مشترکی متعلقند، دارای سرگرمی‌های مشترک، عملکردهای اجتماعی مشترک، دیدگاه‌های مشترک و ... باشند، بیشتر وجود دارد [۱]. بنابراین، اجتماعات شناسایی شده در چنین شبکه‌هایی، می‌توانند در پیشنهاد مشارکتی<sup>۱</sup>، انتشار اطلاعات<sup>۲</sup>، به اشتراک‌گذاری دانش<sup>۳</sup> و دیگر کاربردها مورد استفاده قرار گیرند.

از میان رویکردهای مختلفی که برای تشخیص اجتماع وجود دارد، می‌توان از ۳ رویکرد زیر نام برد: رویکرد اول، عمل تشخیص اجتماع را با در نظر گرفتن ساختار گرافی<sup>۴</sup> شبکه انجام می‌دهد که به این رویکرد، تشخیص ساختاری<sup>۵</sup> گفته می‌شود. همچنین به این رویکرد، تشخیص همبندی اجتماع نیز گفته می‌شود. رویکرد دوم، عمل تشخیص اجتماع را با در نظر گرفتن اطلاعات متنی<sup>۶</sup> گره-های شبکه انجام می‌دهد که به این رویکرد، تشخیص اجتماع موضوعی<sup>۷</sup> گفته می‌شود [۴]. رویکرد سوم، بیانگر یک حالت ترکیبی است، یعنی به نوعی دو رویکرد قبل را با هم ترکیب می‌کند و به عمل تشخیص اجتماع می‌پردازد. بنابراین می‌توان گفت که رویکرد سوم، هم ساختار گرافی شبکه و هم اطلاعات متنی گره‌های شبکه را در نظر می‌گیرد و با توجه به آن عمل تشخیص اجتماع را انجام می‌دهد. با توجه به اینکه رویکرد اول بسیار پرکاربرد می‌باشد و اکثر پژوهش‌هایی که در زمینه‌ی تشخیص اجتماع صورت گرفته‌اند از این رویکرد استفاده کرده‌اند، لذا در این مقاله به بررسی ۲ رویکرد بعدی پرداخته می‌شود.

<sup>1</sup> Collaborative recommendation

<sup>2</sup> Information spreading

<sup>3</sup> Knowledge sharing

<sup>4</sup> Graph structure

<sup>5</sup> Topological community detection

<sup>6</sup> Textual information

<sup>7</sup> Topical community detection

<sup>8</sup> Degree

<sup>9</sup> Distant

یک اجتماع شبکه که گاهی از اوقات به آن پیمانه<sup>۲</sup> یا خوشه<sup>۳</sup> گفته می‌شود، به عنوان یک گروه از گره‌ها در نظر گرفته می‌شود که تعاملات بهتر و بیشتری بین اعضای این گروه نسبت به اعضای این گروه و دیگر اعضای شبکه وجود دارد[۶].

تعریف دیگری بیان می‌کند که اجتماع، یک زیرشبکه‌ی<sup>۴</sup> متراکم<sup>۵</sup> در داخل یک شبکه‌ی بزرگتر می‌باشد، مثل یک گروه از دوستان صمیمی در یک شبکه اجتماعی یا گروهی از صفحات وب به هم پیوند داده شده در شبکه جهانی وب[۷].

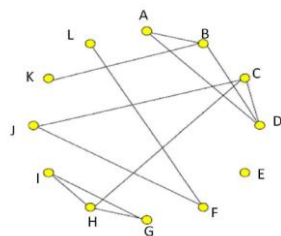
در [۸] آمده است که در یک اجتماع، چگالی<sup>۶</sup> یال‌های داخل آن بیشتر از چگالی یال‌های بین آن با سایر گره‌های شبکه است.

#### ۱-۶- خوشه‌بندی پایگانی<sup>۷</sup>

از آنجایی که برخی از روش‌های تشخیص اجتماع موضوعی از روش خوشه‌بندی پایگانی استفاده می‌کنند، در این قسمت به بررسی این روش پرداخته می‌شود.

در حالت کلی، اطلاعات بسیار کمی در خصوص ساختار اجتماع یک گراف وجود دارد. داشتن دانش قبلی در مورد تعداد خوشه‌های گراف یا در مورد عضویت رئوس یک گراف در خوشه‌ها، بسیار نامعمول است. برای حل مساله در چنین حالتی، بایستی مفروضاتی را در مورد تعداد و اندازه‌ی خوشه‌ها در نظر گرفت. از طرف دیگر، امکان دارد که گراف دارای ساختار پایگانی باشد. به این مفهوم که این گراف دارای سطوح مختلفی از گروه‌بندی<sup>۸</sup> رئوس باشد، یعنی خوشه‌های کوچک در خوشه‌های بزرگتر قرار داشته باشند و آن خوشه‌های بزرگتر نیز به همین ترتیب در خوشه‌های بزرگتر دیگری قرار داشته باشند و .... برای مثال، شبکه‌های اجتماعی، اغلب دارای ساختار پایگانی می‌باشند. زمانی که مساله دارای حالتی این‌چنینی باشد، می‌توان از الگوریتم-

۱. به صورت گراف: در این نوع نمایش از شبکه، ساختارهای گرافی، وظیفه متصل کردن گره‌ها و یال‌ها به یکدیگر را بر عهده دارند. شکل زیر، نمایی از این نوع نمایش از شبکه را نشان می‌دهد:



شکل ۱- نمایش گرافی شبکه [۳]

۲. به صورت ماتریس: برای نمایش گراف شبکه، می‌توان از ماتریس نیز استفاده کرد. به این نوع ماتریس، ماتریس مجاورتی<sup>۱</sup> گراف گفته می‌شود. ماتریس مجاورتی، یک ماتریس مربعی می‌باشد. ابعاد این ماتریس، برابر تعداد گره‌های گراف متناظر با آن بوده و در صورت وجود یالی بین دو گره، در درایه‌ی متناظر با دو گره در ماتریس مجاورتی، مقدار ۱ لحاظ شده و در غیر این‌صورت در این درایه، مقدار ۰ لحاظ می‌شود. شکل زیر، نمونه‌ای از ماتریس مجاورتی گراف ترسیم شده در شکل بالا را نشان می‌دهد:

|   | A | B | C | D | E | G | H | I | J | K | L |
|---|---|---|---|---|---|---|---|---|---|---|---|
| A | 0 | 1 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| B | 1 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 1 | 0 |
| C | 0 | 0 | 0 | 1 | 0 | 0 | 1 | 0 | 1 | 0 | 0 |
| D | 1 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| E | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| F | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 1 |
| G | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 0 | 0 | 0 |
| H | 0 | 0 | 1 | 0 | 0 | 1 | 0 | 1 | 0 | 0 | 0 |
| I | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 0 | 0 | 0 | 0 |
| J | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| K | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| L | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |

شکل ۲- نمایش ماتریس مجاورتی گراف ترسیم شده در شکل ۴ [۳]

#### ۱-۵- تعریف اجتماع

در مقالات و تحقیقات علمی مختلف، تعاریف متفاوتی از اجتماع ارائه شده است که در ادامه به ذکر برخی از این تعاریف پرداخته می‌شود:

<sup>2</sup> Module

<sup>3</sup> Cluster

<sup>4</sup> Subnetwork

<sup>5</sup> Dense

<sup>6</sup> Density

<sup>7</sup> Hierarchical Clustering

<sup>8</sup> Grouping

<sup>1</sup> Adjacency Matrix

های خوشه‌بندی پایگانی استفاده کرد. این الگوریتم‌ها، تکنیک‌های خوشه‌بندی هستند که ساختار چند سطحی گراف را آشکار می‌سازند.

نقطه‌ی شروع هر روش خوشه‌بندی پایگانی، تعریف یک معیار شباهت<sup>۱</sup> بین رئوس می‌باشد. پس از آن که این معیار تعریف شد، بایستی شباهت برای هر جفت راس محاسبه شود. در پایان این فرایند، یک ماتریس  $n \times n$  جدید به وجود می‌آید که به آن ماتریس شباهت<sup>۲</sup> گفته می‌شود. در واقع، هدف الگوریتم‌های خوشه‌بندی پایگانی، شناسایی گروه‌هایی از رئوس با میزان شباهت بالا می‌باشد. این الگوریتم‌ها می‌توانند به ۲ بخش زیر تقسیم شوند:

۱. الگوریتم‌های تجمعی<sup>۳</sup>: در این نوع الگوریتم‌ها، اگر شباهت خوشه‌ها به اندازه‌ی کافی بالا باشد، به صورت مکرر در هم ادغام می‌شوند.

۲. الگوریتم‌های تقسیم‌کننده<sup>۴</sup>: در این نوع الگوریتم‌ها، با حذف یال‌هایی که رئوس با شباهت پایین را به هم متصل می‌کنند، خوشه‌ها به طور مکرر از هم جدا می‌شوند.

این ۲ بخش، دارای فرایندهای متفاوتی می‌باشند. الگوریتم‌های تجمعی، یک فرایند پایین به بالا<sup>۵</sup> را دنبال می‌کنند، چون در آن‌ها، فرایند با در نظر گرفتن هر کدام از رئوس به عنوان خوشه‌های مجزا آغاز می‌شود و نهایتاً به گراف به عنوان یک خوشه‌ی یکتا ختم می‌شود. الگوریتم‌های تقسیم‌کننده، یک فرایند بالا به پایین<sup>۶</sup> را دنبال می‌کنند و مراحل انجام شده در فرایندشان برعکس فرایند قبلی است [۹].

## ۲- تشخیص اجتماع موضوعی

اجتماعات می‌توانند بر پایه‌ی عناوینی که از محتویات<sup>۷</sup> تولید شده توسط موجودیت‌های یک شبکه به دست می‌آیند تشخیص داده شوند. این محتویات، می‌توانند شامل مقالاتی که این موجودیت‌ها منتشر کرده‌اند، وبلاگ‌هایی که ایجاد

کرده‌اند (در آنها مطلبی نوشته‌اند) و .... باشند. محققینی که در بخش تشخیص اجتماعات موضوعی به پژوهش و تحقیق می‌پردازند، یک اصل کلی را دنبال می‌کنند. این اصل بیان می‌دارد که هرچه دو موجودیت تعداد کلمات مشابه بیشتری داشته باشند، به هم شبیه‌ترند [۴]. در واقع باید گفت که رویکردهای تشخیص اجتماع موضوعی، با در نظر گرفتن یک ماتریس شباهت یا یک ماتریس فاصله عمل می‌کنند. در گراف متناظر با این ماتریس‌ها، یال‌ها بیانگر اتصالات واقعی نیستند بلکه نشان دهنده‌ی شباهت دو گره می‌باشند. در ادامه، برخی از روش‌هایی که به منظور تشخیص اجتماع موضوعی در مقالات و پژوهش‌های علمی مختلف استفاده شدند، معرفی می‌شوند.

در رویکرد تشخیص اجتماع موضوعی نیز همانند رویکرد تشخیص اجتماع ساختاری، می‌توان از روش خوشه‌بندی پایگانی به عنوان یک روش معمول بهره جست. خوشه‌بندی پایگانی می‌تواند به منظور استخراج اطلاعات، از ماتریس شباهت استفاده کند [۱۰]. همان‌طور که در بخش قبل گفته شد، این روش نیاز به مشخص کردن تعداد و اندازه‌ی خوشه‌ها در ابتدا، ندارد. یکی از روش‌هایی که از خوشه‌بندی پایگانی به منظور تشخیص اجتماع موضوعی استفاده می‌کند، روش تحلیل ارجاعات متقابل نویسندگان (ACA)<sup>۸</sup> می‌باشد. این روش، به طور گسترده‌ای در مباحث مربوط به کتاب‌سنجی<sup>۹</sup> اعمال شد تا بتوان توسط نتایج حاصل از آن، ساختارهای ذهنی و فکری نویسندگان را کشف کرد و به تصویر کشید [۱۱]. این روش از خوشه‌بندی پایگانی، به منظور گروه‌بندی نویسندگان بر اساس میزان شباهت ارجاعات متقابلشان استفاده می‌کند [۱۲].

روش دیگری که به منظور تشخیص اجتماع موضوعی به کار می‌رود، روش تحلیل مفاهیم پنهان (LSA)<sup>۱۰</sup> می‌باشد. این رویکرد، بسیار پرکاربرد است و به منظور کشف و آشکارسازی ارتباطات مفهومی بین موجودیت‌ها در یک شبکه به کار می‌رود [۱۳]. هافمن یک گام مهم در زمینه‌ی

<sup>1</sup> Similarity measure

<sup>2</sup> Similarity matrix

<sup>3</sup> Agglomerative algorithms

<sup>4</sup> Divisive algorithms

<sup>5</sup> Bottom-up

<sup>6</sup> Top-down

<sup>7</sup> Content

<sup>8</sup> Author Co-citation Analysis

<sup>9</sup> Bibliometrics

<sup>10</sup> Latent Semantic Analysis

مباحث مربوط به غذا و نیمی دیگر، دربرگیرنده‌ی مباحث مربوط به حیوان باشند.

- این نویسنده، هر کلمه در مستند را به این صورت تولید می‌کند که ابتدا یک موضوع را انتخاب می‌کند (برای مثال، با توجه به مطالب بالا، این نویسنده، موضوع غذا را با ۵۰ درصد احتمال و موضوع حیوان را با ۵۰ درصد احتمال انتخاب می‌کند)، سپس از موضوع، برای تولید کلمه استفاده می‌کند. برای مثال، امکان دارد که موضوع حیوان، پاندا با ۳۰ درصد احتمال، ببر با ۲۰ درصد احتمال و .... را تولید کند. با فرض کردن این مدل تولیدی برای گروهی از مستندات، روش تخصیص دیریکله پنهان، سعی می‌کند که به صورت عقب‌گرد<sup>۹</sup>، از مستندات برای بدست آوردن مجموعه‌ی موضوعاتی که احتمال دارد این مستندات را تولید کرده باشند، استفاده کند.

روش تخصیص دیریکله پنهان پایگانی (HLDA)<sup>۱۰</sup>، روش تخصیص دیریکله پنهان را به یک خوشه‌بندی پایگانی بر پایه‌ی درخت تعمیم می‌دهد که این نوع خوشه‌بندی، شبیه به خوشه‌بندی پایگانی می‌باشد که در آن عناوین عمومی‌تر در قسمت بالای درخت و عناوین خاص‌تر در زیر درخت قرار می‌گیرند<sup>۱۱</sup> [۱۶]. روش تخصیص دیریکله پنهان یک روش پودمان<sup>۱۲</sup> بوده و براحتی قابل تعمیم می‌باشد.

در سال ۲۰۰۴، در مقاله‌ی [۱۷]، مدل نویسنده-موضوع<sup>۱۲</sup> معرفی شد. در این مدل، روش تخصیص دیریکله پنهان تعمیم داده شد تا اینکه اطلاعات نویسنده‌گی نیز در آن لحاظ شود. در این مدل، هر نویسنده با یک توزیع چندجمله‌ای بر روی عناوین همراه می‌شود و به این ترتیب خوشه‌های نویسندگان قابل شناسایی خواهد بود. این مدل، برای هر موضوع مخفی، خوشه‌ای از نویسندگان و خوشه‌ای از کلمات، تولید می‌کند. در سال ۲۰۱۰، گوین و همکاران با اعمال روش تخصیص دیریکله پنهان، اجتماعات را از فضای وبلاگی، با توجه به مطالب مطرح شده در وبلاگ‌های آن،

توسعه‌ی روش تحلیل مفاهیم پنهان برداشت. او روش تحلیل احتمالی مفاهیم پنهان (pLSA)<sup>۱</sup> را ارائه داد [۱۴]. نتیجه‌ی این روش، شناسایی اجتماعاتی می‌باشد که نسبت به اجتماعات شناسایی شده توسط روش تحلیل مفاهیم پنهان، موضوع‌گراتر هستند. به عنوان رویکرد دیگری که به منظور تشخیص اجتماع موضوعی به کار می‌رود، می‌توان از رویکرد مدل‌سازی موضوع<sup>۲</sup> نام برد. یکی از روش‌هایی که از این رویکرد استفاده می‌کند، روش تخصیص دیریکله پنهان (LDA)<sup>۳</sup> می‌باشد که توسط بلی و همکاران در سال ۲۰۰۳ معرفی شد [۱۵]. روش تخصیص دیریکله پنهان، یک روش تولیدی<sup>۴</sup> می‌باشد که به صورت تصادفی برای یکسری پارامترهای پنهان<sup>۵</sup> داده شده، داده‌های قابل مشاهده<sup>۶</sup> تولید می‌کند. اگر این داده‌های قابل مشاهده کلماتی باشند که از مستندات بیرون آمده باشند، این روش برای هر مستندی، این کلمات را به تعداد کوچکی از عناوین مخفی<sup>۷</sup>، مخفی<sup>۷</sup>، گروه‌بندی می‌کند. در واقع، روش تخصیص دیریکله دیریکله پنهان می‌تواند به عنوان یک رویکرد خوشه‌بندی مسطح<sup>۸</sup> دیده شود که موجودیت‌ها را بر اساس میزان شباهتشان گروه‌بندی می‌کند. این روش فرض می‌کند که مستندات بدین صورت ساخته شده‌اند:

- زمانی که یک نویسنده شروع به نوشتن یک مستند می‌کند، در مورد تعداد کلمات آن مستند تصمیم می‌گیرد.
- این نویسنده، تلفیق موضوعاتی که در آن مستند مطرح خواهد شد را مشخص می‌کند. برای مثال، در نظر بگیرید که یک نویسنده، دو موضوع غذا و حیوان را برای مستندش انتخاب کرده است. حال او تصمیم می‌گیرد که مثلاً نیمی از مستند او، در برگیرنده‌ی

<sup>1</sup> probabilistic latent Semantic Analysis

<sup>2</sup> Topic Modeling Approach

<sup>3</sup> Latent Dirichlet Allocation

<sup>4</sup> Generative model

<sup>5</sup> Hidden parameters

<sup>6</sup> Observable data

<sup>7</sup> Hidden topics

<sup>8</sup> Flat clustering

<sup>9</sup> Backtrack

<sup>10</sup> Hierarchical LDA

<sup>11</sup> Modularized

<sup>12</sup> Author-Topic Model

استخراج کردند. سپس آنها به منظور کشف زیراجتماعات<sup>۱</sup>، کلماتی که در هر موضوع وجود داشت را خوشه‌بندی کردند[۱۸].

در سال ۲۰۰۸، تانج و همکاران، به منظور در نظر گرفتن مکان انتشار مقالات، مدل نویسنده-موضوع را تعمیم دادند[۱۹]. در این روش، هر نویسنده با یک توزیع چند-جمله ای بر روی عناوین، کلماتی که نوشته است و کنفرانسی که در آن نوشته‌اش منتشر شده است، همراه خواهد بود. این مدل که به مدل نویسنده-کنفرانس-موضوع<sup>۲</sup> مشهور است، به همراه مدل نویسنده-موضوع، تعمیم یافته‌ی روش تخصیص دیریکله پنهان می‌باشند.

در سال ۲۰۰۷، میمنو و همکاران، مدل تولیدی به نام انجمن-نویسنده-موضوع<sup>۳</sup> پیشنهاد کردند[۲۰]. این مدل، هم متن و هم نویسندگان را بر اساس مفهوم اجتماع، خوشه‌بندی می‌کند. با توجه به نتایج حاصل از این مدل، میمنو و همکاران دریافتند که نویسندگان مطرح و قوی می‌توانند در اجتماعات مختلف پخش شده باشند. این امر، بیانگر این واقعیت می‌باشد که این نویسندگان، ممکن است که در مورد عناوین مختلف و با همکاری نویسندگان مختلف مطلب بنویسند.

در این بخش، به صورت فهرست وار، برخی از مهم‌ترین رویکردهای مبتنی بر موضوع به کار گرفته شده در مقالات و پژوهش‌های علمی مختلف مطرح شدند. تمامی این رویکردها، دارای مزایا و معایبی هستند که برخی از آنها ذکر شدند. با توجه به فرض اثبات شده در مقاله‌ی [۴]، به عنوان یکی از معایب رویکردهای بر پایه‌ی موضوع، می‌توان گفت که اجتماعاتی که توسط روش‌های تشخیص اجتماع موضوعی شناخته می‌شوند، تمایل به داشتن اجتماعاتی در درون خودشان دارند که به لحاظ ساختاری (همبندی) متنوعند.

### ۳- تشخیص اجتماع با رویکرد ترکیبی

در سال‌های اخیر، برخی محققین و پژوهش‌گران، تلاش‌هایی را در زمینه‌ی ترکیب دو رویکرد تشخیص اجتماع

ساختاری و تشخیص اجتماع موضوعی -که در بخش قبل بررسی شد- انجام دادند، که در ادامه به برخی از این تلاش‌ها اشاره می‌شود.

در سال ۲۰۰۴، گرول و همکاران، پی به این مطلب بردند که عناوینی که در فضای وبلاگی وجود دارند، با توسعه‌ی انجمن‌های اجتماعی رشد می‌کنند[۲۱]. بر مبنای این مطلب، در سال ۲۰۰۶، ژو و همکاران، یک مدل پویایی موضوع<sup>۴</sup> در مستندات اجتماعی<sup>۵</sup> معرفی کردند که این مدل، مدل، وابستگی موقت موضوع را به تعاملات اجتماعی، متصل و مربوط می‌کند[۲۲]. در واقع در این مدل، به نوعی عناوین به تعاملات موجود در شبکه ربط داده می‌شوند.

در مقاله‌ی دیگری که آن نیز در سال ۲۰۰۶ منتشر شد، ژو و گروه دیگری از همکارانش، دو مدل بیز مولد<sup>۶</sup> برای تشخیص اجتماع مفهومی در شبکه‌های اجتماعی پیشنهاد کردند[۲۳]. این مدل‌ها، مدل‌سازی احتمالی<sup>۷</sup> را با الگوریتم-الگوریتم‌های تشخیص اجتماع ترکیب می‌کنند. ژو و همکاران، با تست الگوریتم هایشان بر روی مجموعه داده‌ای که مربوط به پست الکترونیکی بود، پی به این مطلب بردند که رویکرد آنها، نه تنها قادر به شناسایی اجتماعات می‌باشد، بلکه توصیف مفهومی از موضوع را برای هر اجتماع فراهم می‌کند. آنها یک اجتماع مفهومی در یک شبکه را به صورت اجتماعی تعریف کردند که شامل کاربرانی می‌باشد که دارای علایق ارتباطاتی و موضوعی مشابه هم می‌باشند و این علایق، با تعاملات اجتماعی آنها وابستگی دارد. آنها مدلی به نام مدل اجتماع-کاربر-موضوع<sup>۸</sup> ارائه دادند که این مدل، روش تخصیص دیریکله پنهان را تعمیم بخشید تا این مطلب اثبات شود که یک اجتماع، زمانی شکل می‌گیرد که کاربران موجود در آن، به طور مکرر با هم ارتباط برقرار کنند و این کاربران عناوین موردعلاقه‌ی یکسانی داشته باشند. ۲ نسخه از مدل اجتماع-کاربر-موضوع وجود دارد؛ یکی از این نسخه‌ها، اجتماعات را با کاربران مدل‌سازی می‌کند که به

<sup>4</sup> Model of the topic dynamics

<sup>5</sup> Social documents

<sup>6</sup> Generative Bayesian models

<sup>7</sup> Probabilistic modeling

<sup>8</sup> Community-User-Topic model

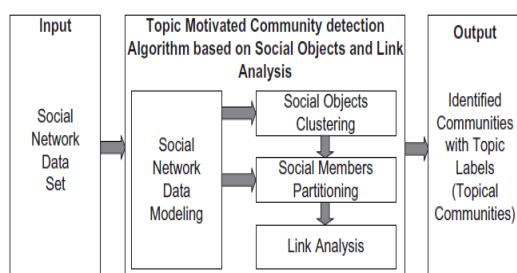
<sup>1</sup> Sub-communities

<sup>2</sup> Author-Conference-Topic model

<sup>3</sup> Community-Author-Topic model

مشترک دارند و همچنین، عناوین می‌توانند به چندین زیرموضوع<sup>۲</sup> تقسیم شوند و این زیرعناوین در طول زمان، بین اجتماعات مختلف پخش می‌شوند. آنها پی به این مطلب بردند که به نظر می‌رسد عناوین، عامل محرک برای تغییرات پویای اجتماعات<sup>۳</sup> می‌باشند، یعنی موضوعات باعث ظهور<sup>۴</sup>، مخلوط شدن<sup>۵</sup> و ناپدید شدن<sup>۶</sup> انجمن‌ها در طول زمان می‌باشند.

اما به عنوان آخرین مورد، به یکی از جدیدترین پژوهش‌هایی که در زمینه‌ی رویکرد ترکیبی انجام شده است، اشاره می‌شود. در مقاله‌ای که در سال ۲۰۱۲ توسط ژائو و همکاران نوشته شد [۱]، یک روش تشخیص اجتماع موضوع-گرا<sup>۷</sup> پیشنهاد شده است. این روش، می‌تواند اجتماعات موضوعی را که به طور همزمان منعکس کننده‌ی عناوین و قدرت اتصالات می‌باشند، شناسایی کند. چارچوب روش ارائه شده در این مقاله، در شکل زیر نشان داده شده است:



شکل ۳- چارچوب روش تشخیص اجتماع موضوع‌گرای ارائه شده توسط ژائو و همکاران [۱]

کل فرایند شناسایی انجمن در این روش توسط ۴ ماژول کلیدی انجام می‌شود، این ماژول‌ها عبارتند از: مدل‌سازی داده‌های شبکه اجتماعی<sup>۸</sup>، خوشه‌بندی اشیاء اجتماعی<sup>۹</sup>، افرازبندی اعضای اجتماعی<sup>۱۰</sup>، تحلیل پیوند<sup>۱۱</sup>. در ادامه، به بررسی ماژول‌های ذکر شده پرداخته می‌شود.

این نسخه، CUT1 گفته می‌شود و نسخه‌ی بعدی، اجتماعات را با عناوین مدل‌سازی می‌کند که به این نسخه، CUT2 گفته می‌شود. CUT1، یک اجتماع را به تنهایی به صورت یک توزیع چندجمله‌ای بر روی کاربران در نظر می‌گیرد و خیلی به اثرات اجتماع بر روی عناوین تولید شده توجه ندارد که این امر، منجر به ارتباط کم بین اجتماع و موضوع می‌شود. بنابراین می‌توان گفت که اجتماعاتی که توسط CUT1 شناسایی می‌شوند، شبیه به اجتماعاتی می‌باشند که توسط الگوریتم‌های تشخیص اجتماع ساختاری شناسایی می‌شوند. از طرف دیگر، CUT2 کاربران را، اگرچه برخی از آنان به ندرت با یکدیگر تعامل داشته باشند، بر اساس عناوین مشترکشان، در یک اجتماع قرار می‌دهد. تحلیل‌هایی که بر روی این دو نسخه از مدل اجتماع-کاربر-موضوع انجام شدند، نشان می‌دهند که هر یک از اجتماعاتی که توسط CUT1 شناسایی شدند، شامل عناوین مختلفی می‌باشند و همچنین، هر یک از اجتماعاتی که توسط CUT2 شناسایی شدند، شامل کاربرانی می‌باشند که از مکان‌های مختلفی گرد هم آمده‌اند و در نتیجه، دارای تعاملات کم اما دارای عناوین مشابه می‌باشند.

در سال ۲۰۱۰، لی و همکاران، با استفاده از یک ساز و کار استنتاجی<sup>۱</sup>، روش تخصیص دیریکله پنهان را با روش تشخیص اجتماع جیروان و نیومن ترکیب کردند [۲۴]. هر یک از اجتماعات بدست آمده توسط مدل پیشنهادی لی و همکاران، دربرگیرنده‌ی مجموعه‌ای از گره‌هایی بود که ارتباطات قوی‌تری با دیگر اعضای اجتماعشان نسبت به اعضای دیگر اجتماعات داشتند. بر اساس اطلاعات مربوط به هر یک از گره‌ها، احتمال توزیع هر یک از گره‌ها بر روی مجموعه‌ای از موضوعات، با استفاده از مدل تخصیص دیریکله پنهان، بدست آمد، در حالی که ساختار اجتماعی کاربران، توسط الگوریتم تشخیص اجتماع جیروان و نیومن آشکار شد. نتایج حاصل از تست الگوریتم پیشنهادی لی و همکاران، نشان داد که کاربرانی که در یک اجتماع هستند، در تمام دوره‌های زمانی، تمایل به داشتن مجموعه عناوین

<sup>2</sup> Sub-topics

<sup>3</sup> Dynamic changes of communities

<sup>4</sup> Emerging

<sup>5</sup> Blending

<sup>6</sup> Disappearing

<sup>7</sup> Topic oriented community detection approach

<sup>8</sup> Social network data modeling

<sup>9</sup> Social objects clustering

<sup>10</sup> Social members partitioning

<sup>11</sup> Link analysis

<sup>1</sup> Inference mechanism

شد و نتیجه این شد که روش بکار گرفته شده در مقاله، دارای دو مزیت زیر نسبت به روش دوم می‌باشد:

۱. کامل بودن اتصالات<sup>۱۱</sup> را در یک موضوع خاص حفظ می‌کند.

۲. کارایی تجزیه و تحلیل اتصالات را افزایش می‌دهد.

#### ۴- نتیجه

این مقاله، مروری کلی بر بخشی از رویکردهای مختلفی که در پژوهش‌های علمی مختلف به منظور تشخیص اجتماع در شبکه‌های پیچیده پیشنهاد شده‌اند، می‌باشد. در این مقاله، پس از بررسی مفاهیم اولیه مورد نیاز جهت مطالعه‌ی شبکه‌های اجتماعی، از میان رویکردهای مختلفی که تا کنون در مقالات و پژوهش‌های علمی مختلف به منظور تشخیص اجتماع در شبکه‌های پیچیده ارائه شده‌اند، به بررسی دو رویکرد مطرح پرداخته شد. رویکرد اول، رویکرد تشخیص اجتماع موضوعی می‌باشد که عمل تشخیص اجتماع را با در نظر گرفتن اطلاعات متنی گره‌های شبکه انجام می‌دهد. رویکرد دوم، به منظور تشخیص دقیق تر اجتماعات، فرایند انجام شده در رویکردهای تشخیص اجتماع موضوعی و تشخیص اجتماع ساختاری (که در این مقاله در مورد آن بحثی نشد) را با هم ترکیب می‌کند و به عمل تشخیص اجتماع می‌پردازد که در این مقاله، به آن رویکرد ترکیبی گفته شد. به عنوان یک نتیجه‌ی کلی از مباحث مطرح شده در این مقاله بایستی گفت که اجتماعاتی که توسط روش‌های تشخیص اجتماع ساختاری شناخته می‌شوند، به طور معمول دربرگیرنده‌ی زیراجتماعاتی هستند که این زیراجتماعات به لحاظ موضوعی متنوعند و از طرف دیگر، اجتماعاتی که توسط روش‌های تشخیص اجتماع موضوعی شناخته می‌شوند، به طور معمول دربرگیرنده‌ی زیراجتماعاتی هستند که این زیراجتماعات به لحاظ ساختاری متنوعند. این موضوع، بیانگر لزوم وجود رویکردی می‌باشد که هم یک شبکه را به لحاظ ساختاری و همبندی و هم به لحاظ محتوایی و موضوعی بررسی کرده و با توجه به این موارد به عمل

ماژول اول، ماژول مدل‌سازی داده‌های شبکه اجتماعی می‌باشد. هدف این ماژول، ساختاربندی مجموعه داده به مدل‌های فرمال<sup>۱</sup> برای پردازش می‌باشد. این مقاله، با در نظر گرفتن اشیاء اجتماعی<sup>۲</sup>، یک مدل گرافی فرمال برای توصیف شبکه اجتماعی پیشنهاد می‌دهد.

ماژول‌های خوشه‌بندی اشیاء اجتماعی و افرازبندی اعضای اجتماعی دومین و سومین ماژول می‌باشند. به طور کلی، مردم فعالیت‌های اجتماعی را بر روی اشیاء اجتماعی مثل ایمیل‌ها و وبلاگ‌ها انجام می‌دهند. این اشیاء، اغلب نمایانگر عناوینی که مردم به آنها علاقه مندند، می‌باشند. در این مقاله، تنها، اشیاء اجتماعی متنی<sup>۳</sup> در نظر گرفته شده‌اند.

چهارمین ماژول، ماژول تشخیص اجتماع موضوعی (یا تحلیل پیوند) می‌باشد. این ماژول به روی هر خوشه‌ی موضوعی اعمال می‌شود. اعضای هر کدام از خوشه‌های موضوعی، اغلب به یکدیگر با شدت‌های<sup>۴</sup> مختلفی متصلند. برای شناسایی اعضای که بسیار تنگاتنگ به هم متصلند<sup>۵</sup>، بایستی تحلیل اتصالات<sup>۶</sup> روی هر خوشه‌ی موضوعی انجام شود. در این مقاله به این فرایند، شناسایی انجمن موضوعی یا تجزیه و تحلیل اتصالات گفته می‌شود. در این فرایند، روش‌های تشخیص اجتماعات زیادی می‌توانند بکار گرفته شوند. در این مقاله، برای تجزیه و تحلیل اتصالات، از روش‌های بر پایه‌ی بیشینه‌سازی پودمانی<sup>۷</sup> استفاده شده است.

با توجه به موارد ذکر شده، روش ارائه شده توسط این مقاله، در ۳ مرحله انجام می‌شود: خوشه‌بندی اشیاء اجتماعی<sup>۸</sup>، افرازبندی کاربر بر مبنای موضوع<sup>۹</sup> و تشخیص اجتماع بر پایه‌ی اتصال<sup>۱۰</sup>. اما یک روش دیگر می‌تواند به این صورت باشد که ابتدا، تشخیص اجتماع بر پایه‌ی اتصالات انجام شود و سپس کاربران را بر مبنای عناوین افرازبندی کرد. در این مقاله، روش دومی، با روش بکار گرفته شده در مقاله مقایسه

<sup>1</sup> Formal models

<sup>2</sup> Social objects

<sup>3</sup> Text social objects

<sup>4</sup> Strength

<sup>5</sup> Tightly connected members

<sup>6</sup> Link analysis

<sup>7</sup> Modularity

<sup>8</sup> Social object clustering

<sup>9</sup> Topic-based user partitioning

<sup>10</sup> Link-based community detection

<sup>11</sup> Completeness of connections



- [13] Deerwester, S., et al., *Indexing by latent semantic analysis*, *Journal of the American society for information science*, 1990. 41(6): p. 391-407.
- [14] Hofmann, T. *Probabilistic latent semantic indexing*, in *Proceedings of the 22nd annual international ACM SIGIR conference on Research and development in information retrieval*. 1999. ACM.
- [15] Blei, D.M., A.Y. Ng, and M.I. Jordan, *Latent dirichlet allocation*, the *Journal of machine Learning research*, 2003. 3: p. 993-1022.
- [16] Blei, D.M., T.L. Griffiths, and M.I. Jordan, *The nested chinese restaurant process and bayesian nonparametric inference of topic hierarchies*, *Journal of the ACM (JACM)*, 2010. 57(2): p. 7.
- [17] Rosen-Zvi, M., et al. *The author-topic model for authors and documents*, in *Proceedings of the 20th conference on Uncertainty in artificial intelligence*. 2004. AUAI Press.
- [18] Nguyen, T., et al. *Hyper-community detection in the blogosphere*, in *Proceedings of second ACM SIGMM workshop on Social media*. 2010. ACM.
- [19] Tang, J., et al. *ArnetMiner: extraction and mining of academic social networks*, in *Proceedings of the 14th ACM SIGKDD international conference on Knowledge discovery and data mining*. 2008. ACM.
- [20] Mimno, D., H. Wallach, and A. McCallum. *Community-based Link Prediction with Text*, in *Proc. of NIPS*. 2007.
- [21] Gruhl, D., et al. *Information diffusion through blogspace*, in *Proceedings of the 13th international conference on World Wide Web*. 2004. ACM.
- [22] Zhou, D., et al. *Topic evolution and social interactions: how authors effect research*, in *Proceedings of the 15th ACM international conference on Information and knowledge management*. 2006. ACM.
- [23] Zhou, D., et al. *Probabilistic models for discovering e-communities*, in *Proceedings of the 15th international conference on World Wide Web*. 2006. ACM.
- [24] Li, D., et al. *Community-based topic modeling for social tagging*, in *Proceedings of the 19th ACM international conference on Information and knowledge management*. 2010. ACM.

تشخیص اجتماع بپردازد. برای رسیدن به این مهم، رویکرد ترکیبی مطرح شد. بنابراین می‌توان گفت که در این رویکرد، به نوعی ارتباط بین اجتماع و موضوع در نظر گرفته می‌شود. از آنجایی که این رویکرد هم متن و هم ساختار را در نظر می‌گیرد، می‌توان نتیجه گرفت که اجتماعاتی که به عنوان نتیجه‌ی این رویکرد شناسایی می‌شوند با معناترند، به این دلیل که اعضای آن هم دارای ارتباطات تنگاتنگ و تعاملات بالا و هم دارای علایق موضوعی یکسان خواهند بود.

## مراجع

- [1] Zhao, Z., et al., *Topic oriented community detection through social objects and link analysis in social networks*, *Knowledge-Based Systems*, 2012. 26: p. 164-173.
- [2] Lancichinetti, A. and S. Fortunato, *Consensus clustering in complex networks*, *Scientific reports*, 2012. 2.
- [3] Hosseinzadeh R., Alizadeh H., Nazemi E., **Community Detection Ensemble in Social Networks, The 11th Iranian Conf. on Intelligent Systems (ICIS13)**, Tehran, Feb. 27-28, 2013 (in Persian)
- [4] Ding, Y., *Community detection: topological vs. topical*, *Journal of Informetrics*, 2011. 5(4): p. 498-514.
- [5] Easley, D. and J. Kleinberg, *Networks, crowds, and markets*, Vol. 8. 2010: Cambridge Univ Press.
- [6] Leskovec, J., K.J. Lang, and M. Mahoney. *Empirical comparison of algorithms for network community detection*, in *Proceedings of the 19th international conference on World Wide Web*. 2010. ACM.
- [7] Newman, M., *Communities, modules and large-scale structure in networks*, *Nature Physics*, 2011.
- [8] Porter, M.A., J.-P. Onnela, and P.J. Mucha, *Communities in networks*, *Notices of the AMS*, 2009. 56(9): p. 1082-1097.
- [9] Fortunato, S., **Community detection in graphs**. *Physics Reports*, 2010. 486(3): p. 75-174.
- [10] Newman, M.E., *Detecting community structure in networks*, *The European Physical Journal B-Condensed Matter and Complex Systems*, 2004. 38(2): p. 321-330.
- [11] White, H.D. and K.W. McCain, *Visualizing a discipline: An author co-citation analysis of information science, 1972-1995*. *Journal of the American society for information science*, 1998. 49(4): p. 327-355.
- [12] Ding, Y., et al., *Bibliometric Information Retrieval System (BIRS): A Web search interface utilizing bibliometric research results*, *Journal of the American Society for Information Science*, 2000. 51(13): p. 1190-1204.