

کشف اجتماعات با استفاده از شباهت یال در شبکه‌های پیچیده

محمد مهدی دلیری خمایی^۱، علیرضا رضوانیان^۲ و محمدرضا میبیدی^۳

^۱ دانشجوی کارشناسی ارشد، گروه مهندسی نرم افزار، دانشگاه آزاد اسلامی، واحد قزوین، قزوین، ایران، m.daliri@qiau.ac.ir

^۲ آزمایشگاه محاسبات نرم، دانشگاه صنعتی امیرکبیر، تهران، ایران، a.rezvanian@aut.ac.ir

^۳ آزمایشگاه محاسبات نرم، دانشگاه صنعتی امیرکبیر، تهران، ایران، mmebybodi@aut.ac.ir

چکیده - یکی از مهمترین ویژگی‌های شبکه‌های پیچیده وجود ساختارهای اجتماعی می‌باشد. بطور مشخص شناسایی این ساختارها در شبکه‌های پیچیده به تحلیل ویژگی‌های ساختاری شبکه کمک می‌کند. در سال‌های اخیر الگوریتم‌های متعددی برای کشف اجتماعات در شبکه‌های پیچیده پیشنهاد شده است. با توجه به ویژگی‌های این اجتماعات، یکی از روش‌های موجود برای شناسایی اجتماعات ارائه الگوریتم‌هایی برای وزن‌دهی یال‌های شبکه است به طوری که وزن یال‌های درون اجتماعات افزایش و بطور همزمان وزن یال‌های مابین اجتماعات کاهش یابد تا تمایز میان اجتماعات به سادگی قابل شناسایی باشند. در این مقاله، یک الگوریتم دو مرحله‌ای به صورت سلسله مراتبی پیشنهاد شده است. در مرحله اول، الگوریتم پیشنهادی سعی در شناسایی نودهای مرکزی دارد و در مرحله دوم با استفاده از وزن دهی یال‌ها بر مبنای معیار مشابهت در یک روال تکراری به ادغام میان نودهای مشابه می‌پردازد. فرآیند ادغام به صورت پایین به بالا تا رسیدن به یک اجتماع واحد از کل شبکه ادامه می‌یابد. به منظور ارزیابی روش پیشنهادی، آزمایشات متعددی بر روی مجموعه داده‌های استاندارد شبکه‌های پیچیده صورت گرفته است. نتایج آزمایشات از لحاظ معیارهای ارزیابی ماژولاریتی و خلوص و معیارهای مشابهت حاکی از برتری نسبی روش پیشنهادی است. کلید واژه - شبکه‌های پیچیده، شبکه‌های اجتماعی، کشف اجتماعات، ساختار اجتماعی، ماژولاریتی

یکی از چالش‌ها در تجزیه و تحلیل شبکه‌های پیچیده شناسایی و استخراج اجتماعات می‌باشد. از شناسایی اجتماعات برای تحلیل ویژگی‌های ساختاری شبکه‌های پیچیده استفاده می‌شود که برای تحلیل برخی از ویژگی‌های ضمنی نیز مورد استفاده قرار می‌گیرد [۱-۲]. مثلاً در شبکه اجتماعی نویسندگان مقالات توسط شناسایی اجتماعات می‌توان به شناسایی اجتماع نویسندگان همکار با موضوعات مشابه دست یافت. همچنین در شبکه‌های متابولیک، پروتئین‌های دارای عملکرد مشابه اجتماعات را تشکیل می‌دهد. از کاربردهای شناسایی اجتماعات می‌توان به دسته‌بندی تصویر، دسته‌بندی صفحات وب، دسته‌بندی در شبکه‌های سنسور اشاره نمود. در بسیاری از شبکه‌های اجتماعی ارتباطات داخلی قوی مابین نودهای همان اجتماع در مقایسه با سایر نودها از اجتماعات مشاهده می‌شود. ارتباطات داخلی قوی افرادی که ارتباطات محکمی با یکدیگر دارند را شامل می‌شود. در سال‌های اخیر الگوریتم‌های متعددی برای شناسایی اجتماعات پیشنهاد شده است. یکی از الگوریتم‌های اولیه شناسایی اجتماعات الگوریتم برش کمینه می‌باشد. در الگوریتم برش کمینه شبکه به تعداد بخش‌های از قبل تعیین شده تقسیم می‌شود. در این روش تعداد یال‌هایی که بین اجتماعات قرار دارند مینیمم می‌باشد. الگوریتم‌های افراز

۱. مقدمه

بسیاری از شبکه‌های موجود در دنیای واقعی می‌توانند توسط یک شبکه پیچیده مدل شوند. شبکه‌های پیچیده توسط گرافی از نودها (معرف افراد) و یال‌ها (بیانگر ارتباط مابین نودها) نمایش داده می‌شوند. شبکه‌های پیچیده اعم از شبکه‌های بیولوژیکی، اینترنتی، اطلاعاتی و اجتماعی دارای ویژگی‌های مشابهی مانند جهان کوچک، مستقل از مقیاس و دارای ساختارهای پیمانه‌ای (گروهی) می‌باشند، از این حیث در سال‌های اخیر مطالعه شبکه‌های پیچیده توجه بسیاری از محققان را به خود جلب نموده است [۱]. تجزیه و تحلیل شبکه‌های اجتماعی منجر به استخراج اطلاعات مفید می‌گردد. در واقع هدف از تحلیل شبکه‌های اجتماعی، شناسایی ساختار شبکه و مدل کردن ارتباطات بین شبکه می‌باشد. یکی از مهم‌ترین ویژگی‌های شبکه‌های اجتماعی وجود ساختار ماژولار (پیمانه‌ای) می‌باشد. برای بررسی ویژگی‌های اطلاعات در ساختار پیمانه‌ای، شبکه از چندین پیمانه یا اجتماع تشکیل شده است که این اجتماعات دارای ارتباطات داخلی قوی مابین نودهای همان اجتماع در مقایسه با سایر نودها از اجتماعات دیگر است [۱]

می‌کنند، در واقع یال‌های ضعیف یکپارچگی سراسری را حفظ می‌کنند در حالیکه یال‌های قوی تقویت‌کننده ارتباطات در داخل اجتماعات می‌باشند. با اینحال مسئله مهم چگونگی انتساب وزن‌ها به گراف می‌باشد که به عنوان امری مهم تلقی می‌شود معیارهای مشابهت ابزاری قدرتمند برای تمایز بین ارتباطات در شبکه می‌باشند. اخیراً معیارهای زیادی معرفی شده است در واقع معیارهای مشابهت برای رسیدن به نتایج بهتری کشف اجتماعات می‌تواند کمک کند. در ادامه این مقاله به صورت زیر سازماندهی شده است: بخش دوم الگوریتم پیشنهادی همراه با جزئیات و معیارهای مشابهت معرفی شده است در بخش سوم معیارهای ارزیابی معرفی شده است و در بخش چهارم مجموعه داده‌های آزمایشی و نتایج آزمایشات بیان می‌شود و در نهایت در بخش پایانی نتیجه‌گیری و معرفی کارهای آینده ارائه شده است.

۲. الگوریتم پیشنهادی

در الگوریتم پیشنهادی برای شناسایی اجتماعات ابتدا به شناسایی نودهای مرکزی می‌پردازیم. پس از شناسایی نودهای مرکزی، برای شناسایی اجتماعات از ویژگی‌های توپولوژیکی شبکه برای اندازه‌گیری میزان شباهت مابین نودها استفاده می‌شود. برای اندازه‌گیری و مقایسه نودها براساس میزان شباهت، یک تابع وزنی پیشنهاد شده است که در این تابع نودهایی با بیشترین شباهت (کمترین فاصله) با یکدیگر ادغام می‌شوند. الگوریتم شناسایی نودهای مرکزی در بخش ۲-۱ و الگوریتم شناسایی اجتماعات در بخش ۲-۲ و تابع وزن پیشنهادی با شرح جزئیات در بخش ۲-۳ معرفی می‌شود.

۱-۲ انتخاب نودهای مرکزی

در ابتدا کلیه‌ی نودها در حالت بازدید نشده در نظر گرفته می‌شود (توسط لیست L_i) و در هر تکرار بخشی از نودهای گراف مورد بازدید قرار می‌گیرد. در ابتدای تکرار از بین نودهای بازدید نشده یک نود (v_i) به تصادف انتخاب شده و همسایگان نود انتخابی با درجات بیشتر به عنوان کاندید قرار گرفتن در اجتماع معرفی می‌گردد، سپس از بین همسایگان، شبیه‌ترین نود براساس نسبت تعداد همسایگان مشترک بر روی تعداد کل همسایگان مشترک و غیر مشترک محاسبه می‌گردد، انتخاب شده و به مجموعه v_i اضافه می‌گردد. این رویه برای نود v_i تا جایی ادامه می‌یابد که دیگر نتوان نودی را به مجموعه v_i اضافه نمود. در این رویه نودهایی که بازدید می‌شوند بعنوان نود بازدید شده لحاظ می‌شوند همچنین نودی که دارای بیشترین میزان شباهت باشد

گراف، الگوریتم‌های خوشه‌بندی طیفی [۵] و الگوریتم‌های خوشه‌بندی سلسله مراتبی [۱] دسته دیگری از الگوریتم‌های شناسایی اجتماعات می‌باشند. الگوریتم‌های دیگری مبتنی بر بهینه‌سازی شامل بهینه‌سازی ماژولاریتی، بهینه‌سازی خارجی، بهینه‌سازی طیفی برای حل مسئله کشف اجتماعات پیشنهاد شده است. الگوریتم‌های سلسله مراتبی دسته دیگری از الگوریتم‌ها برای شناسایی اجتماعات می‌باشند. در الگوریتم‌های سلسله مراتبی معیاری برای شباهت جفت نودها معرفی می‌شود. بنابراین نودهای با بیشترین معیار تشابه در یک اجتماع قرار می‌گیرند. دسته دیگری از الگوریتم‌ها گسترده‌ترین تکنیک مورد استفاده در شناسایی اجتماعات الگوریتم‌های خوشه‌بندی سلسله مراتبی می‌باشند [۱]. روش‌های سلسله مراتبی اساساً به دو دسته عمده تقسیمی و ادغامی دسته‌بندی می‌شوند. الگوریتم‌های تقسیمی که یال‌های خاصی که متصل به اجتماعات دیگر باشند را شناسایی و حذف می‌کنند. نیومن الگوریتمی تقسیمی مبتنی بر بینیت برای شناسایی اجتماعات پیشنهاد داده است [۵] که در این روش یالی که دارای مقدار بینیت بالایی می‌باشد حذف می‌گردد. روش شناسایی اجتماعات ارائه شده توسط نیومن در شبکه‌های با مقیاس بزرگ از کارایی لازم برخوردار نیست. در اکثر روش‌های تجمعی گروه‌بندی راس‌های گراف با شروع از یک افراز اولیه دلخواه از گراف می‌باشد. در این روش‌های معمولاً در ابتدا هر نود بعنوان یک جامعه در نظر گرفته می‌شود سپس براساس معیار مشابهت سعی در تشکیل اجتماع در هر مرحله دارند. در این روش‌ها برای تعیین ارزیابی اجتماعات نیازمند معیاری در شبکه می‌باشند. یکی از روش‌های پیشنهادی روش نیومن [۳] است که مسئله کشف اجتماعات را بعنوان یک مسئله بهینه‌سازی با در نظر گرفتن معیاری بنام ماژولاریتی Q بیان نموده است [۶]. با اینحال ماژولاریتی برای گروه خاصی از گراف‌ها جواب قابل قبول ارائه می‌کند [۹]. به عنوان مثال در یک گراف متشکل از چند کلیک، ممکن است اجتماعات به درستی تشخیص داده نشود. لی و همکاران [۱۱] معیار چگالی ماژولاریتی را برای حل مسئله ماژولاریتی معرفی نمودند ماکزیمم نمودن این تابع باعث می‌شود که ماژولاریتی وزن‌دار نتیجه بهتری نسبت به نسخه اولیه ماژولاریتی ارائه کند. در روش ماژولاریتی وزن‌دار استخراج اجتماعات به ارتباطات شبکه و همچنین وزن ارتباطات نیز وابسته می‌باشد. در [۱۲] نظریه‌ای مبنی بر نقش یال‌های مختلف در شبکه بیان شده است که یال‌های ضعیف (یال‌های ارتباطی مابین اجتماعات) و یال‌های قوی (یال‌های موجود در میان اجتماع) نقش متفاوتی در رفتار شبکه ایفا

و یا اینکه نود مجزایی در شبکه وجود نداشته باشد. شبه کد الگوریتم برای شناسایی اجتماعات در شکل ۲ ارائه شده است.

Algorithm 2: Finding Community Based on Edge Similarity

Input: a Network $G(V, E)$ and Core nodes
Output: a Set of communities

Assumptions
Let S_i be the core node and vertices that member of core node
Let V_{S_i} be the similar vertex of S_i .
Let Flg to keep the structure of community.

Begin Algorithm
Calculate the weight of edges in graph based on weight function & assign to edges
do
 For each Core Node S_i Do Parallel
 $V_{S_i} \leftarrow$ Select an adjacency and Non-repetitive Node with S_i randomly
 If (V_{S_i} Have Most Similarity with S_i) Then
 $S_i \leftarrow S_i \cup V_{S_i}$
 Flg \leftarrow change
 Else
 $S_i \leftarrow S_i \setminus V_{S_i}$
 Flg \leftarrow No change
 End if
 End for
While (Graph changed topologically)
End Algorithm

شکل ۲: شبه کد شناسایی اجتماعات توسط الگوریتم پیشنهادی

۳-۲ تابع وزنی پیشنهادی

با توجه به اهمیت وزن دهی در فرآیند کشف اجتماعات، در این مقاله ما از یک معیار فاصله برای انتساب وزن به یال های شبکه استفاده می نمایم. ایده ی اصلی این معیار براساس تخمینی از فاصله می باشد، در واقع نودهای درون اجتماع دارای فاصله کمتری نسبت به نودهای بین اجتماعات می باشند. در تابع وزنی پیشنهادی برای نودهای مرتبط مستقیم وزن بیشتری نسبت به نودهای مرتبط همراه با واسطه در نظر گرفته شده است. بنابراین، در حالت مستقیم یال میان هر دو نود و در حالت غیر مستقیم کلیه مسیرهای موجود مابین هر دو نود مورد بررسی قرار می گیرد. با توجه به افزایش فاصله در همسایگان یک نود می توان از یک ضریب وزنی برای میزان تاثیر فاصله برای عضویت همسایگان یک نود استفاده نمود. با توجه به موارد مذکور می توان رابطه زیر را برای وزن دهی یال ها در نظر گرفت. برای هر زوج نود v_s و v_d تابع وزن پیشنهادی بصورت زیر خواهد بود.

$$S_{v_s d} = \sum_{i=1}^{n-1} w_i \cdot p_i \quad (1)$$

در این رویه بعنوان نود مرکزی (هاب) شناخته می شود. این رویه تا زمانی ادامه می یابد که کلیه نودها بازدید شوند. شبه کد الگوریتم برای شناسایی نودهای مرکزی در شکل (۱) ارائه شده است.

Algorithm1: Finding core node

Input: An undirected graph $G=(V, E)$
Output: A set of nodes as core node

Assumptions
Let N_i be a list of visited nodes from i
Let C_i be a list of the most appropriate selected node from candidate node
Let T_i be a flag which shows the status of each node and initially set to 0

Begin Algorithm
While (all nodes are not visited) **Do**
 Begin
 Select a node randomly as v_i
 $N_i \leftarrow \underset{v_i \in V}{argmax} \{degree(v_i)\}$
 Do
 For each $v_j \in N_i$ Do
 begin
 Similarity(v_i, v_j) $\leftarrow \frac{|\Gamma(v_i) \cap \Gamma(v_j)|}{|\Gamma(v_i) \cup \Gamma(v_j)|}$
 Visited[v_j] \leftarrow visit
 End For
 Add maximum similar node C_i based on similarity measure
 Clear elements of N_i
 $N_i \leftarrow \underset{v_i \in V}{argmax} \{degree(C_i)\}$
 While($N_i \neq$ Empty)
 End While
End Algorithm

شکل ۱: شبه کد شناسایی نودهای مرکزی

۲-۲ الگوریتم شناسایی اجتماعات

پس از شناسایی نودهای مرکزی در مرحله اول از الگوریتم پیشنهادی، فرآیند وزن دهی برای شناسایی اجتماعات آغاز می گردد. در این مرحله از اجرای الگوریتم، در هر تکرار تعلق نودهای مجاور با نود مرکزی در یک رویه تکراری حذف و اضافه شدن نودها انجام می گیرد. رویه حذف و اضافه بصورت تکراری صورت می پذیرد بطوریکه در رویه اضافه شدن، نود مجاور به نود مرکزی در تکرار جاری اضافه می گردد، سپس در این مرحله فرآیند جستجو برای یافتن نودی با بیشترین میزان مشابهت به نود مرکزی بصورت محلی آغاز می گردد، چنانچه نودی دارای بیشترین میزان مشابهت (کمترین فاصله) از نود جاری وجود داشته باشد نود اضافه شده حذف می گردد و نود دارای بیشترین میزان مشابهت اضافه می گردد. روال اجرای الگوریتم تا جایی ادامه می یابد که امکان تغییر جدیدی در شبکه امکان پذیر نباشد

$$S_{v_i, v_j}^{Soresnsen} = \frac{2|n(v_i) \cap n(v_j)|}{n(v_i) + n(v_j)} \quad (4)$$

معيار سالتون (S_{v_i, v_j}^{Salton}): شاخص سالتون [4] تحت نام کوسینوسی نیز شناخته می شود و بصورت زیر تعریف می شود.

$$S_{v_i, v_j}^{Salton} = \frac{|n(v_i) \cap n(v_j)|}{\sqrt{n(v_i) + n(v_j)}} \quad (5)$$

معيار گسترش شاخص هاب (S_{v_i, v_j}^{Hpl}): معيار گسترش شاخص هاب [4] برای کمی نمودن جفت نودهای همپوشان در شبکه متابولیک بکار رفته است و بصورت زیر تعریف می شود:

$$S_{v_i, v_j}^{Hpl} = \frac{|n(v_i) \cap n(v_j)|}{\min\{n(v_i), n(v_j)\}} \quad (6)$$

در این معيار يال‌هایی که نزدیک هاب می باشند مقدار بیشتری به آنها داده می شود.

۲-۳ معيار ارزیابی

در این مقاله از معيار ماژولاریتی [6] و خلوص برای ارزیابی اجتماعات معرفی می‌گردد. معيار ماژولاریتی بعنوان معياری برای ارزیابی کیفیت خوشه بندی بکار می‌رود. این معيار که در فرمول (9) معرفی گردیده دارای مقداری بین 0 و 1 می باشد. هر چقدر مقدار ماژولاریتی به 1 نزدیک باشد نشان دهنده کیفیت مناسب خوشه‌بندی می‌باشد. گراف $G=(V,E)$ بعنوان یک گراف بدون جهت که $n=|V|$ برابر تعداد راس‌ها و $m=|E|$ تعداد يال‌های گراف می باشد. گراف G توسط ماتریس مجاورت $A=(A_{ij})$ توصیف می شود. هرگاه نود i با نود j متصل باشد درایه متناظر با آنها در ماتریس مجاورت برابر 1 و در غیر اینصورت برابر صفر خواهد بود. همچنین K_i درجه نود i می باشد. ماژولاریتی برای پارتیشن بندی $C = \{C_1, C_2, \dots, C_l\}$ از گراف G بصورت زیر معرفی می‌گردد.

$$Q(C) = \frac{1}{2m} \sum_{i,j} (A_{ij} - \frac{K_i K_j}{2m}) \delta_{ij} \quad (7)$$

در فرمول (7) اگر نود i و j در اجتماع مشترک باشند مقدار δ_{ij} برابر 1 می باشد در غیر اینصورت مقدار δ_{ij} برابر صفر می باشد. معيار خلوص بعنوان معيار دیگری برای ارزیابی کیفیت اجتماعات معرفی شده است. این معيار برای هر اجتماع کلاسی را در نظر می‌گیرد در هر کلاس ماکزیمم تعداد نودهای که بدرستی در آن کلاس قرار دارد را محاسبه می‌کند. معيار خلوص در فرمول (10) بیان شده است.

$$Purity(C_{RS}, C_{CDA}) = \frac{1}{N} \sum \max |C_{RS} \cap C_{CDA}| \quad (8)$$

C_{RS} مجموعه کلاس‌های در نظر گرفته شده و C_{CDA} خروجی الگوریتم کشف اجتماعات می باشد.

بطوریکه p_i مسيرهای مجزا (مسيرهایی که دارای هیچ يال مشترکی با مسيرهای قبلی نداشته باشند) به طول i مابین نودهای s و d است و n تعداد نودهای موجود در گراف مورد نظر است که در رابطه (1) مسيرهای به طور حداکثر $n-1$ در نظر گرفته می‌شود. w_i وزن متناظر با هر یک از مسيرها می‌باشد که برای سادگی $w_i=1/i$ در نظر گرفته شده است، بنابراین می‌توان داشت:

$$S_{v_{sd}} = \sum_{i=1}^{n-1} \frac{p_i}{i} \quad (2)$$

با توجه به رابطه (2)، هر چقدر تعداد همسایگان یک نود در فاصله دورتری از یک نود قرار گیرند، دارای تاثیر کمتری در عضویت آنها به یک اجتماع می‌باشند. قابل توجه است که به منظور کاهش پیچیدگی محاسبات و زمانبر بودن آن برای مسيرهای طولانی می‌توان از برخی از مسيرهای وزن ناچیز صرف نظر کرد.

۳. آزمایشات

جهت ارزیابی از معيارهای مشابهت معروف جکارد، سورنسن، سالتون و شاخص گسترش هاب برای مقایسه با تابع پیشنهادی برای وزندهی يال‌های شبکه استفاده شده است که به طور مختصر در بخش 3-1 ارائه شده است. سپس در بخش 3-2 معيار ارزیابی کیفی ماژولاریتی و خلوص برای مقایسه روش پیشنهادی معرفی شده است. آزمایشات مختلفی در بخش 3-3 برای ارزیابی روش پیشنهادی بر روی شبکه‌های معروف آزمایش شده است.

۱-3 معيارهای مشابهت

در این مقاله، برای مقایسه تابع وزن‌دهی پیشنهادی از 4 معيار مشابهت معروف استفاده شده است. هریک از معيارهای موجود در ادامه یک وزن برای مشابهت يال مابین نودهای v_i و v_j ارائه می‌دهد. در روابط ذکر شده در ذیل $n(v_i)$ بعنوان مجموعه همسایگان نود v_i تعریف می شود. معيارهای موجود به شرح زیر است:

معيار جکارد ($S_{v_i, v_j}^{jaccard}$): معيار جکارد که در [4] معرفی شده است برابر اندازه اشتراک مجموعه نودها تقسیم بر اجتماع آنها می باشد.

$$S_{v_i, v_j}^{jaccard} = \frac{|n(v_i) \cap n(v_j)|}{|n(v_i) \cup n(v_j)|} \quad (3)$$

معيار سورنسن ($S_{v_i, v_j}^{Soresnsen}$): شاخص سورنسن [4] برای شناسایی اجتماعات اکولوژیکی معرفی شده است و بصورت زیر تعریف می‌شود:

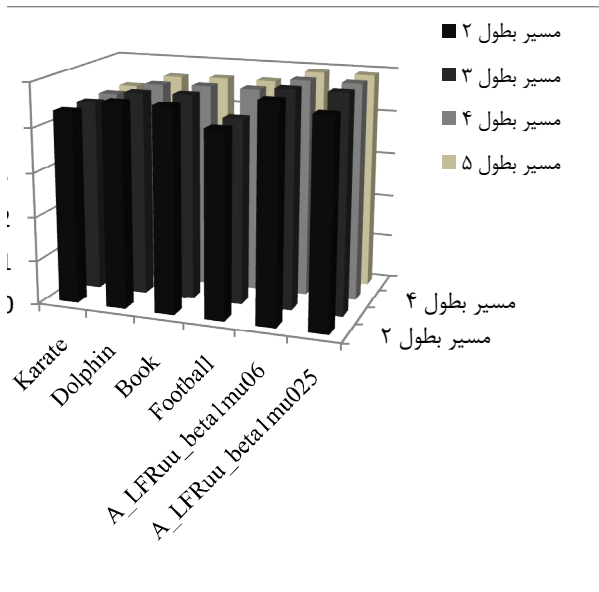
۳-۳ آزمایشات

برای شبیه سازی نتایج الگوریتم پیشنهادی از مجموعه داده های Karate, Dolphin, Football, Political Book استفاده شده است. همچنین از دو مجموعه داده مصنوعی LFR [۸] برای ارزیابی توسط روش پیشنهادی مورد بررسی قرار گرفته است. مجموعه داده LFR برای ایجاد گراف مصنوعی بکار می رود در این مجموعه داده N برابر تعداد نودها k میانگین درجات max کمازیم درجات داخلی می باشد (جزئیات بیشتر پارامترها مربوط به ایجاد گراف مصنوعی LFR در مرجع [۸] معرفی شده است). جزئیات مربوط به هریک از این دادگان در جدول (۱) معرفی شده است.

جدول ۱: شرح جزئیات مربوط به دادگان آزمایشی

دادگان آزمایشی	رئوس	یال ها	میانگین درجات	قطر
Karate	۳۴	۷۸	۴.۵۹	۵
Football	۱۱۵	۶۱۳	۱۰.۶۶	۴
Dolphin	۶۲	۱۵۹	۵.۱۳	۸
Political Book	۱۰۵	۴۴۱	۸.۴	۷
A_LFRuu_beta1mu06	۱۰۰۰	۹۸۸۴	۱۹.۷۷	۴
A_LFRuu_beta1mu025	۱۰۰۰	۹۷۱۱	۱۹.۴۲	۵

هدف از آزمایش اول بررسی تاثیر مسیرهای با طول مختلف و تعیین مسیر با طول مناسب در مجموعه دادگان آزمایشی می باشد. بدین منظور مجموع مسیر های به طول ۲ تا ۵ مورد بررسی قرار گرفته است. در حالت کلی دلیل وجود ساختار تنک و فاقد یال مستقیم برخی از شبکه ها در آزمایشات از مسیر بطول ۱ (یال مستقیم) صرف نظر نموده ایم همچنین بدلیل در نظر گرفتن مسیرهای متمایز و مستقل در شبکه و عدم وجود مسیر متمایز با فاصله های بیشتر و برتری نسبی در بررسی مسیرهای بطول ۲ تا ۵ از سایر مسیر های صرف نظر شده است. نتایج آزمایش بر روی مسیرهای مختلف و مقدار ماژولاریتی بدست آمده مورد ارزیابی قرار گرفته و در شکل شماره (۱) نشان داده شده است.



شکل ۳: نتایج بررسی تاثیر مسیرهای مختلف بر روی مجموعه دادگان

همانطور که مشاهده می شود در بسیاری از مجموعه دادگان مقدار ماژولاریتی در مسیر بطول ۳ دارای نتایج مشابهی نسبت به مسیرهای با طول بیشتر می باشد، بنابراین بدلیل پیچیده بودن محاسبات در تابع وزن مسیر بطول ۳ برای مقایسه با سایر روش ها در نظر گرفته شده است. در آزمایش دوم هدف، بررسی و مقایسه مقدار ماژولاریتی تعداد اجتماعات بدست آمده توسط تابع وزن پیشنهادی در مقایسه با توابع وزن معرفی شده در این مقاله می باشد. نتایج بدست آمده در برای ۱۰ بار اجرای الگوریتم برای توابع وزن معرفی شده در جدول (۲) گزارش شده است.

جدول ۲: مقدار ماژولاریتی برای تابع وزن های پیشنهادی در مقایسه با توابع وزن معرفی شده.

دادگان آزمایشی	چکارد	سالتون	سورنسون	گسترش هاب	تابع وزن پیشنهادی
Karate	۰.۳۷	۰.۴۲	۰.۳۸	۰.۴۳	۰.۴۳
Football	۰.۳۵	۰.۳۵	۰.۱۷	۰.۳۲	۰.۴۶
Dolphin	۰.۱۷	۰.۱۷	۰.۳۳	۰.۱۶	۰.۴۲
Political Book	۰.۰۸	۰.۱۱	۰.۰۹	۰.۰۹	۰.۴۶
A_LFRuu_beta1mu06	۰.۳۲	۰.۱۳	۰.۱۷	۰.۱۷	۰.۴۹
A_LFRuu_beta1mu025	۰.۳۲	۰.۱۲	۰.۱۷	۰.۱۷	۰.۴۹

همانطور که در جدول (۲) مشاهده می شود معیار پیشنهادی در این مقاله بدلیل در نظر گرفتن همسایگی های و عدم تمرکز بر همسایگی های محلی و همچنین ساختار نسبتا تنک و دارای میانگین درجات پایین برخی از مجموعه داده ها از قبیل Political Book, Dolphin, Football, A_LFRuu_beta1mu025, A_LFRuu_beta1mu06 منجر به نتایج بهتر ماژولاریتی نسبت به بسیاری از معیار های مشابهت

باشد که برای رسیدن به مقدار ماژولاریتی نزدیکتر به الگوریتم لوین باید فاصله های با طول بیشتر را لحاظ نمود، از طرفی دلیل سربار محاسباتی بالا در تعیین فاصله با طول بیشتر از این مسئله می توان صرف نظر نمود.

۴. نتیجه گیری

بسیاری از شبکه های موجود در دنیای واقعی می توانند توسط یک شبکه پیچیده مدل شوند. بررسی ساختار شبکه های پیچیده منجر به معرفی اجتماعات گردیده است. از آنجائیکه این اجتماعات دارای ارتباطات داخلی قوی مابین نودهای همان اجتماع در مقایسه با سایر نودها از اجتماعات دیگر است. ارائه معیاری برای وزن دهی و شناخت بهتر این اجتماعات امری ضروری تلقی می شود. در این مقاله یک الگوریتم ۲ مرحله ای برای شناسایی نودهای مرکزی و تشکیل اجتماعات و معیاری برای وزن دهی به یال های موجود در شبکه معرفی شده است. نتایج اجرای الگوریتم ۲ مرحله ای حاکی از برتری نسبی الگوریتم در مقایسه با معیارهای مشابهت محلی بدلیل در نظر گرفتن فاصله در همسایگی ها و در نظر گرفتن معیار فاصله در محاسبه تعلق نودها به اجتماعات می باشد، که این موضوع در معیارهای مشابهت نادیده گرفته می شود. همچنین از جهت برتری نسبت به سایر الگوریتم های معرفی شده از قبیل گرون- نیومن، لوین، کلازت- نیومن- مور در مجموعه داده های کاراته و مجموعه داده های تصنعی الگوریتم پیشنهادی دارای برتری می باشد. در مجموعه داده های بوک، دلفین، فوتبال بدلیل ساختار تنک این شبکه ها الگوریتم می توان فاصله های بیشتری را برای شناسایی جوامع در نظر گرفت که این مسئله دارای سربار محاسباتی زیادی بوده و از سر بار محاسباتی صرف نظر شده است.

۵. مراجع

- [1] Santo Fortunato, "Community detection in graphs", Proceedings of the National Academy of Sciences of the United States of America, Vol.486; PP 75-174; 2010.
- [2] Peng Gang Sun, "Weighting links based on edge centrality for community detection", Proceedings of Physica A : Statistical Mechanics and its Applications, Vol.394; PP 346-357; 2014.
- [3] A. Clauset, M. Newman, and C. Moore, "Finding community structure in very large Networks",

گردیده است. حتی در مواردی در معیارهای جکارا، سالتون، سورنسون بدلیل ساختار و نوع توپولوژی شبکه معیار پیشنهادی در مجموعه داده بوک و مجموعه داده های LFR از برتری بیشتری برخوردار می باشد. در واقع در معیارهای محلی بدلیل اولویت به نودهای محلی که در همسایگی مستقیم قرار دارند و صرف نظر از نودهای با فاصله بیشتر منجر به نتایج ضعیف تر نسبت به تابع وزن پیشنهادی شده است. همچنین همانطور که در جدول (۳) مشاهده می شود از حیث معیار خلوص و دقت دسته بندی بدلیل در نظر گرفتن نودهای مرکزی و شناسایی نودهایی که در مرکز اجتماعات قرار گرفته اند، تابع وزن پیشنهادی در الگوریتم ارائه شده نسبت به معیارهای محلی از دقت بالاتری برخوردار می باشد.

جدول ۳: مقدار خلوص برای تابع وزن های پیشنهادی در مقایسه با توابع وزن معرفی شده.

دادگان آزمایشی	جکارا	سالتون	سورنسون	گسترش هاب	تابع وزن پیشنهادی
Karate	۰.۸۲	۰.۹۳	۰.۸۴	۰.۹۶	۰.۹۶
Football	۰.۶۵	۰.۶۵	۰.۳۱	۰.۵۹	۰.۸۶
Dolphin	۰.۳۱	۰.۳۱	۰.۶	۰.۲۹	۰.۸۵
Political Book	۰.۱۱	۰.۱۵	۰.۱۲	۰.۱۲	۰.۶
$A_{LFRuu_beta1mu06}$	۰.۶	۰.۲۴	۰.۳۲	۰.۳۲	۰.۹۳
$A_{LFRuu_beta1mu025}$	۰.۶	۰.۲۲	۰.۳۱	۰.۳۱	۰.۹۲

در آزمایش چهارم به مقایسه الگوریتم پیشنهادی با الگوریتم های معروف از قبیل گرون و نیومن [۱۴] و لوین [۱۵] و کلازت و نیومن و مور [۱۶] پرداخته شده است.

جدول ۴: مقایسه الگوریتم پیشنهادی دو مرحله ای براساس تابع وزن معرفی شده با سایر الگوریتم ها

مجموعه دادگان آزمایشی	لوین	گرون و نیومن	کلازت و نیومن و مور	الگوریتم دو مرحله ای پیشنهادی
Karate	۰.۴۱۵۱	۰.۴۰۱۲	۰.۳۸۰۶	۰.۴۳۵۳
Football	۰.۵۲۴۱	۰.۵۹۹۴	۰.۵۴۸۸	۰.۴۶۳
Dolphin	۰.۵۲۶۶	۰.۵۱۹۳	۰.۵۱۴۵	۰.۴۶۶۳
Political Book	۰.۵۹۸۶	۰.۵۰۹۹	۰.۵۰۱۹	۰.۴۲۰۷
$A_{LFRuu_beta1mu06}$	۰.۴۴۵۶	۰.۱۲۵۴	۰.۲۷۹۹	۰.۴۹۲۳
$A_{LFRuu_beta1mu025}$	۰.۴۷۸۵	۰.۴۱۵۵	۰.۶۴۳۶	۰.۴۹۲۲

همانطور که در جدول (۴) مشاهده می شود الگوریتم پیشنهادی در مجموعه داده کاراته و مجموعه داده مصنوعی ال اف آر حاکی از برتری روش پیشنهادی نسبت به سایر الگوریتم های معرفی شده می باشد. در مجموعه داده های فوتبال، دلفین و بوک روش لوین دارای برتری نسبی نسبت به الگوریتم پیشنهادی می باشد، یکی از دلایل برتری روش لوین نسبت به الگوریتم های معرفی شده ساختار نسبتا تنک این شبکه ها می -

- [10] P.G. Sun, L. Gao, Y. Yang, "Maximizing modularity intensity for community partition and evolution", *Information Sciences*, Vol.236,2013.
- [11] M.S. Granovetter, "Economic action and social structure: the problem of embeddedness", *American journal of sociology*, Vol.91,No.3,1985.
- [12] W.W. Zachary, "An information flow model for conflict and fission in small groups", *Journal of Anthropological Research*, Vol. 33, 1977.
- [13] V.D. Blondel, J.-L. Guillaume, R. Lambiotte, E. Lefebvre, "Fast unfolding of communities in large networks", *proceeding of Iopscience*, No.10,2008.
- [14] M. Girvan, M.E.J. Newman, "Community structure in social and biological networks ", *Proceeding of National Academy Scosity*, Vol.99,no 12,2002.
- [15] Fast unfolding of communities in large networks, Vincent D Blondel, Jean-Loup Guillaume, Renaud Lambiotte, Etienne Lefebvre, *Journal of Statistical Mechanics: Theory and Experiment 2008 (10)*, P10008.
- [16] J. Duch and A. Arenas, "Community detection in complex networks using extremal optimization," *Physical Review E*, vol. 72, no. 2, p. 027104, 2005.
- Proceedings of Statistical Nonlinear and Soft Matter Physics ,Vol.70,2004.
- [4] L. Lu, T. Zhou, "Link Prediction in Complex Networks: A Survey", *Proceedings of Physica A: Statistical Mechanics and its Applications*, Vol.390, N.6, 2011.
- [5] T. Hastie, R. Tibshirani, J.H. Friedman, *The Elements of Statistical Learning*, Springer, Berlin, Germany, 2nd edition, P.P 282-745, 2009.
- [6] R. Guimera, M. Sales-Pardo, L.A.N. Amaral, "Modularity from fluctuations in random graphs and complex networks", *Proceedings of Statistical Nonlinear and Soft Matter Physics*, vol.70,2004.
- [7] M. Girvan, M.E.J. Newman, "Community structure in social and biological networks ", *Proceeding of National Academy Scosity*, Vol.99,no 12,2002.
- [8] A. Lancichinetti, S. Fortunato, F. Radicchi, "Benchmark graphs for testing community detection algorithms", *Proceeding of Statistical Nonlinear And Soft Matter Physics*, Vol. 78, 2008.
- [9] S. Fortunato, M. Barthelemy, "Resolution limit in community detection", *Proceedings of the National Academy of Sciences of the United States of America* ,Vol.2007, USA 2007.