

# کشف اجتماعات با استفاده از شبکه‌های پیچیده

محمد مهدی دلیری خمامی<sup>۱</sup>، علیرضا رضوانیان<sup>۲</sup> و محمدرضا میبدی<sup>۳</sup>

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

<sup>۲</sup>آزمایشگاه محاسبات نرم، دانشگاه صنعتی امیرکبیر، تهران، ایران، a.rezvanian@aut.ac.ir

<sup>۳</sup>آزمایشگاه محاسبات نرم، دانشگاه صنعتی امیرکبیر، تهران، ایران، mmeybodi@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 وزن دهنده یال‌های شبکه است به طوریکه 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 دهنده 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 مازوچریتی و خلوص و معیارهای مشابهت حاکی از برتری نسبی روش پیشنهادی است. کلید واژه - شبکه‌های پیچیده، شبکه‌های اجتماعی، کشف اجتماعات، ساختار اجتماعی، مازوچریتی

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 نویسنده‌گان مقالات توسط شناسایی اجتماعات می‌توان به شناسایی اجتماع نویسنده‌گان 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 پیشنهاده 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 قراردارند مینیمم می‌باشد. الگوریتم‌های افزار

## ۱. مقدمه

بسیاری از شبکه‌های موجود در دنیا واقعی می‌توانند توسط یک شبکه پیچیده مدل شوند. شبکه‌های پیچیده توسط گرافی از نودها (معرف افراد) و یال‌ها (بینگر 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 سال-های 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 ساختارشبکه و مدل کردن ارتباطات بین شبکه می‌باشد. یکی از مهم‌ترین ویژگی‌های شبکه‌های اجتماعی وجود ساختار مازوچر (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 تقویت کننده 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 ۱-۲ و الگوریتم شناسایی اجتماعات در بخش ۲-۲ تابع وزن پیشنهادی با شرح جزئیات در بخش ۳-۲ معرفی می‌شود.

### ۱-۲ انتخاب نودهای مرکزی

در ابتدا کلیه 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  $i$  اضافه می‌گردد. این رویه برای نود  $v_i$  تا جایی ادامه می‌باید که دیگر نتوان نودی را به مجموعه  $i$  اضافه نمود. در این رویه نودهایی که بازدید می‌شوند بعنوان نود بازدید شده لحاظ می‌شوند همچنین نودی که دارای بیشترین میزان مشابهت باشد

گراف، الگوریتم‌های خوشبندی طیفی [۵] و الگوریتم‌های خوشبندی سلسله مراتبی [۱] دسته دیگری از الگوریتم‌های شناسایی اجتماعات می‌باشند. الگوریتم‌های دیگری مبتنی بر بهینه‌سازی شامل بهینه‌سازی مازولاریتی، بهینه‌سازی خارجی، بهینه‌سازی طیفی برای حل مسئله کشف اجتماعات پیشنهاد شده است. الگوریتم‌های سلسله مراتبی دسته دیگری از الگوریتم‌ها برای شناسایی اجتماعات می‌باشد. 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 خوشبندی سلسله مراتبی می‌باشد [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 افزار اولیه دلخواه از گراف می‌باشد. در این روش‌های معمولاً در ابتدا هرنود 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 مازولاریتی  $Q$  بیان نموده است [۶]. با اینحال مازولاریتی 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 مازولاریتی را برای حل مسئله مازولاریتی معرفی نمودند ماکریم نمودن این تابع باعث می‌شود که مازولاریتی وزن دار نتیجه بهتری نسبت به نسخه اولبه مازولاریتی ارائه کند. در روش مازولاریتی وزن دار 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. شبکه 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: شبکه 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 نماییم. ایده ای 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 برسی قرار می گیرد. با توجه به افزایش فاصله در همسایگان یک نود می توان از یک ضربی 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.

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

در این رویه بعنوان نود مرکزی (هاب) شناخته می شود. این رویه تا زمانی ادامه می یابد که کلیه نودها بازدید شوند. شبکه 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  $v_i$ 
         $N_i \leftarrow \underset{v_i \in V}{\operatorname{argmax}} \{ \text{degree}(v_i) \}$ 
    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  $N_i$ 
         $N_i \leftarrow \underset{v_i \in V}{\operatorname{argmax}} \{ \text{degree}(C_i) \}$ 
    While( $N_i \neq \text{Empty} \text{ } \text{ } \text{ } \text{ }$ )
    End While
End Algorithm

```

شکل ۱: شبکه 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 تکرای 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

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

معیار سالتون ( $S_{v_i, v_j}^{Salton}$ ): شاخص سالتون [۴] تحت نام

کوسینوسی نیز شناخته می شود و بصورت زیر تعریف می شود.

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

معیار گسترش شاخص هاب ( $S_{v_i, v_j}^{Hpl}$ ): معیار گسترش

شاخص هاب [۴] 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 معیار یالهایی که نزدیک هاب می باشند مقدار بیشتری به آنها داده می شود .

### ۲-۳ معیار ارزیابی

در این مقاله از معیار مازوچاریتی [۶] و خلوص برای ارزیابی اجتماعات معرفی می گردد. معیار مازوچاریتی بعنوان معیاری برای ارزیابی کیفیت خوش بندی بکار می رود. این معیار که در فرمول (۹) معرفی گردیده دارای مقداری بین ۰ و ۱ می باشد. هر چقدر مقدار مازوچاریتی به ۱ نزدیک باشد نشان دهنده کیفیت مناسب خوشبندی می باشد. گراف ( $G = (V, E)$ ) بعنوان یک گراف بدون خوشبندی می باشد. گراف ( $G = (A_{ij}, m = |E|$ ) تعداد یال های  $A = (A_{ij})$  معرفی می باشد. گراف  $G$  توسط ماتریس مجاورت ( $A_{ij}$ ) توصیف می شود. هرگاه نود  $i$  با نود  $j$  متصل باشد درایه متناظر با آنها در ماتریس مجاورت برابر ۱ و در غیر اینصورت برابر صفر خواهد بود. همچنین  $K_i$  درجه نود  $i$  می باشد. مازوچاریتی برای پارهیز بندهای  $C_1, C_2, \dots, C_l$  از گراف  $G$  بصورت زیر معرفی می گردد.

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

در فرمول (۷) اگر نود  $i$  و  $j$  در اجتماع مشترک باشند مقدار  $\delta_{ij}$  برابر ۱ می باشد در غیر اینصورت مقدار  $\delta_{ij}$  برابر صفر می باشد. معیار خلوص بعنوان معیار دیگری برای ارزیابی کیفیت اجتماعات معرفی شده است. این معیار برای هر اجتماع کلاسی را در نظر می گیرد در هر کلاس ماقریزم مجموعه نودهای که بدرستی در آن کلاس قرار دارد را محاسبه می کند. معیار خلوص در فرمول (۸) بیان شده است.

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

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

بطوریکه  $p_i$  مسیرهای مجزا (مسیرهایی که دارای هیچ یال مشترکی با مسیرهای قبلی نداشته باشند) به طول  $n$  مابین نودهای  $s$  و  $d$  است و  $n$  تعداد نودهای موجود در گراف مورد نظر است که در رابطه (۱) مسیرهای به طور حداقل  $n-1$  در نظر گرفته می شود.  $w_i$  وزن متناظر با هر یک از مسیرها می باشد که برای سادگی  $w_i = I/i$  در نظر گرفته شده است، بنابراین می توان داشت:

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

با توجه به رابطه (۲)، هر چقدر تعداد همسایگان یک نود رفاسله دورتری از یک نود قرار گیرند، دارای تاثیر کمتری در عضویت آنها به یک اجتماع می باشند. قابل توجه است که به منظور کاهش پیچیدگی محاسبات و زمانبند بودن آن برای مسیرهای طولانی می توان از برخی از مسیرهای وزن ناچیز صرف نظر کرد.

### ۳. آزمایشات

جهت ارزیابی از معیارهای مشابه معرفه شکاره، سورنسن، سالتون و شاخص گسترش هاب برای مقایسه با تابع پیشنهادی برای وزندهی یال های شبکه استفاده شده است که به طور محتص در بخش ۳-۱ ارائه شده است. سپس در بخش ۳-۲-۳ معیار ارزیابی کیفی مازوچاریتی و خلوص برای مقایسه روش پیشنهادی معرفی شده است. آزمایشات مختلفی در بخش ۳-۳ برای ارزیابی روش پیشنهادی بر روی شبکه های معروف آزمایش شده است.

### ۱-۳ معیارهای مشابه

در این مقاله برای مقایسه تابع وزن دهی پیشنهادی از ۴ معیار مشابه معروف استفاده شده است. هر یک از معیارهای موجود در ادامه یک وزن برای مشابه یال مابین نودهای  $v_i$  و  $v_j$  ارائه می دهد. در روابط ذکر شده در ذیل ( $i$ )  $n(v_i)$  بعنوان مجموعه همسایگان نود  $v_i$  تعریف می شود. معیارهای موجود به شرح زیر است:

معیار جکارد ( $S_{v_i, v_j}^{jaccard}$ ): معیار جکارد که در [۴] معرفی شده است برای اندازه اشتراک مجموعه نودها تقسیم بر اجتماع آنها می باشد.

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

معیار سورنسن ( $S_{v_i, v_j}^{Soresnse}$ ): شاخص سورنسن [۴] برای شناسایی اجتماعات اکلوزیکی معرفی شده است و بصورت زیر تعریف می شود:

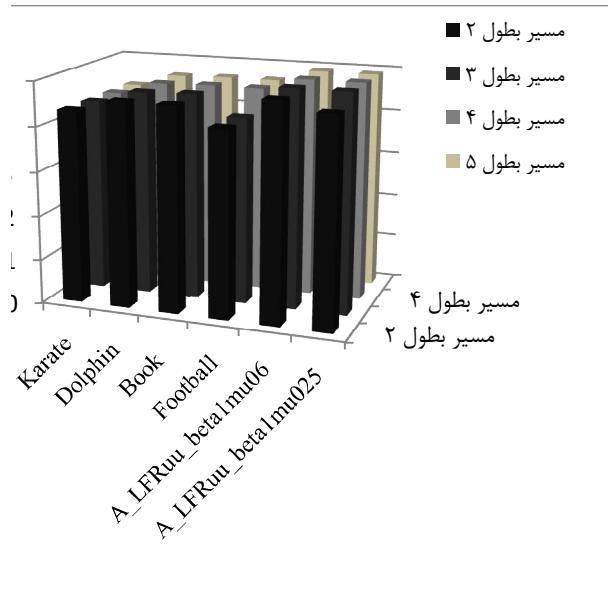
### آزمایشات ۳-۳

برای شبیه سازی نتایج الگوریتم پیشنهادی از مجموعه داده های Political Book، Football، Dolphin، Karate 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 maxk ماکزیمم 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	قطر	میانگین درجات	یال ها	رؤس
Karate	۵	۴.۵۹	۷۸	۳۴
Football	۴	۱۰.۶۶	۶۱۳	۱۱۵
Dolphin	۸	۵.۱۳	۱۵۹	۶۲
Political Book	۷	۸.۴	۴۴۱	۱۰۵
A_LFRuu_beta1mu06	۴	۱۹.۷۷	۹۸۸۴	۱۰۰۰
A_LFRuu_beta1mu025	۵	۱۹.۴۲	۹۷۱۱	۱۰۰۰

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 مسیرهای به طول ۲ تا ۵ مورد بررسی قرار گرفته است. در حالت کلی بدليل وجود ساختار تنک و فاقد یال مستقیم برخی از شبکه‌ها در آزمایشات از مسیر بطول ۱ (یال مستقیم) صرفنظر نموده ایم همچنین بدليل 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 مسیرهای صرفنظر شده است. نتایج آزمایش بر روی مسیرهای مختلف و مقدار مازوپاریتی 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 مازوپاریتی در مسیر بطول ۳ دارای نتایج مشابهی نسبت به مسیرهای با طول بیشتر می باشد، بنابراین بدليل پیچیده بودن محاسبات در تابع وزن مسیر بطول ۳ برای مقایسه با سایر روش‌ها در نظر گرفته شده است. در آزمایش دوم هدف، بررسی و مقایسه مقدار مازوپاریتی تعداد اجتماعات بدست آمده توسط تابع وزن پیشنهادی در مقایسه با تابع 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 مازوپاریتی برای تابع وزن های پیشنهادی در مقایسه با توابع وزن معرفی شده.

دادگان آزمایشی	دادگان آزمایشی	چکاره	سالتوون	سورنسون	گسترش هاب	تابع وزن پیشنهادی
Karate	Karate	~0.37	~0.42	~0.38	~0.43	~0.43
Football	Football	~0.25	~0.25	~0.17	~0.32	~0.46
Dolphin	Dolphin	~0.17	~0.17	~0.33	~0.16	~0.42
Political Book	Political Book	~0.08	~0.11	~0.09	~0.09	~0.46
A_LFRuu_beta1mu06	A_LFRuu_beta1mu06	~0.22	~0.13	~0.17	~0.17	~0.49
A_LFRuu_beta1mu025	A_LFRuu_beta1mu025	~0.32	~0.12	~0.17	~0.17	~0.49

همانطور که در جدول (۲) مشاهده می شود معیار پیشنهادی در این مقاله بدليل 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\_beta1mu06، A\_LFRuu\_beta1mu025 منجر به نتایج بهتر مازوپاریتی نسبت به بسیاری از معیار های مشابه است.

باشد که برای رسیدن به مقدار مأذولاریتی نزدیکتر به الگوریتم لوین باید فاصله های با طول بیشتر را لحاظ نمود، از طرفی بدلیل سربار محاسباتی بالا در تعیین فاصله با طول بیشتر از این مسئله می توان صرفنظر 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 دهنی 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 دهنی به یال های موجود در شبکه معرفی شده است. نتایج اجرای الگوریتم ۲ مرحله ای حاکی از برتری نسبی الگوریتم در مقایسه با معیارهای مشابه محلی بدلیل درنظر گرفتن فاصله در همسایگی ها و در نظر گرفتن معیار فاصله در محاسبه تعلق نودها به اجتماعات می باشد، که این موضوع در معیارهای مشابه 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 دادهای کاراته و مجموعه داده های تصنیعی الگوریتم پیشنهادی دارای برتری می باشد. در مجموعه داده های بوک، دلفین، فوتbal 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 بر محاسباتی صرفنظر 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 جکارد، سالتون، سورنسون بدلیل ساختار و نوع توبولوژی 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 صرفنظر از نودهای با فاصله بیشتر منجر به نتایج ضعیفتر نسبت به تابع وزن پیشنهادی شده است. همچنین همانطور که در جدول (۳) مشاهده می شود از حيث 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: مقایسه خلوص برای تابع وزن های پیشنهادی در مقایسه با تابع وزن معرفی شده.

دادگان آزمایشی	جکارد	سالتون	سورنسون	گسترش هاب	تابع وزن پیشنهادی
Karate	0.82	0.93	0.84	0.96	0.96
Football	0.65	0.65	0.31	0.59	0.86
Dolphin	0.31	0.21	0.6	0.29	0.85
Political Book	0.11	0.15	0.12	0.12	0.6
A_LFRuu_beta1mu06	0.6	0.24	0.22	0.32	0.93
A_LFRuu_beta1mu025	0.6	0.22	0.31	0.31	0.92

در آزمایش چهارم به مقایسه الگوریتم پیشنهادی با الگوریتم های معروف از قبیل گرون و نیومن [۱۴] و لوین [۱۵] و کلازت- نیومن- 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	0.4151	0.4012	0.3806	0.4253
Football	0.5241	0.5994	0.5488	0.463
Dolphin	0.5266	0.5193	0.5145	0.4663
Political Book	0.5986	0.5099	0.5019	0.4207
A_LFRuu_beta1mu06	0.4456	0.1254	0.2799	0.4923
A_LFRuu_beta1mu025	0.4785	0.4155	0.6436	0.4922

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.