

# پیشنهاد یک الگوریتم برای خوشه‌بندی شبکه‌های اجتماعی مبتنی بر یال‌های مابینیت

محسن کجیاف<sup>1</sup>، سعید پارسا<sup>2</sup>

<sup>1</sup> دانشجوی کارشناسی ارشد کامپیوتر، گرایش نرم افزار، دانشگاه آزاد اسلامی، واحد بروجرد- بروجرد، ایران

آدرس پست الکترونیک: m\_kajbaf@yahoo.com

<sup>2</sup> عضو هیات علمی گروه کامپیوتر دانشگاه علم و صنعت ایران - تهران، ایران

آدرس پست الکترونیک: parsa@iust.ac.ir

## چکیده

امروزه شبکه‌های اجتماعی به‌طور فزاینده‌ای محبوب شده‌اند، زیرا کانال‌های ارتباطی سریعی هستند که افراد می‌توانند از طریق وسایل الکترونیکی شخصی در هر مکان و در هر زمان با دوستان، آشنایان و همکاران و ... ارتباط برقرار کنند. برای مدل‌سازی این شبکه‌ها، از گراف استفاده می‌کنیم که در آن افراد گره‌ها را تشکیل می‌دهند و روابط بین افراد با یال‌ها نمایش داده می‌شود. در این بین یک گراف بزرگ به‌وجود می‌آید. در این مقاله برآنیم که با استفاده از الگوریتم‌های گراف کاوی، راه‌حلی برای خوشه‌بندی این‌گونه شبکه‌ها (موسوم به شبکه‌های پیچیده) پیشنهاد دهیم. در پایان، الگوریتم پیشنهاد شده را بر روی دیتاست استاندارد بازیکنان فوتبال، اعمال کرده‌ایم و نتایج حاصل شده در قالب خوشه‌های به-دست‌آمده، نمایش داده شده‌اند.

## کلمات کلیدی

شبکه‌های اجتماعی، خوشه‌بندی، الگوریتم

## 1. مقدمه

همه نوع نظیر خانواده‌ها، سکونتگاه‌ها، مذهب‌ها، سازمان‌ها و گاهی نیز اجتماع‌های مجازی (در این سمینار رسانه‌ها و مخصوصاً شبکه‌های اجتماعی مجازی) پیوند برزند [8].

یک شبکه اجتماعی، یک ساختار اجتماعی ساخته شده از افراد (یا سازمان‌ها) است که گره‌ها را تشکیل می‌دهند. این گره‌ها به وسیله‌ی انواع خاصی از وابستگی‌های متقابل مانند دوستی، خویشاوندی، منافع مشترک، روابط جنسی، باورهای مشترک و ... به هم متصل شده‌اند. تجزیه و تحلیل شبکه‌های اجتماعی، بیانگر روابط اجتماعی است که به کمک تئوری گراف مدل می-شوند [8]. شبکه‌های اجتماعی شامل گره‌ها (افراد یا سازمان‌ها) و روابط (یال‌ها، پیوندها یا اتصالات) می‌باشد. گره‌ها بازیگران فردی درون شبکه‌ها هستند و علاقمندی‌ها همان روابط بین بازیگران

انسان‌ها با افراد غریبه بیشتر از سایر موجودات دیگر همکاری کرده و اطلاعات مبادله می‌کند و این امر همراه با ورود به جوامع است و از جمع این جوامع یک شبکه جهان‌گستر تشکیل می-شود. جامعه‌گرایی افراد موجب می‌شود که آن‌ها بتوانند در فعالیت‌های معدودی متخصص شوند و آن‌ها را به دلیل ارضای نیازها و باقی خواسته‌هایشان به دیگران وابسته می‌نماید. در نتیجه افراد را می‌توان اتم‌های شناوری در بازارهایی که در آن کالا و خدمات مبادله می‌نمایند، به حساب آورد [10]. اما در بازبینی‌های دقیق‌تر می‌توان گفت که انسان‌ها تا حدودی ترجیح می‌دهند خودشان را به سایر افراد در گروه‌ها و اجتماع‌هایی از

هستند، بنابراین ساختارهای مبتنی بر گراف خیلی پیچیده هستند. در بخش 2 این مقاله به بررسی خوشه‌بندی شبکه‌های اجتماعی خواهیم پرداخت. در بخش 3 الگوریتم پیشنهادی را شرح خواهیم داد. در بخش 4 یک مطالعه موردی و اعمال الگوریتم پیشنهادی بر روی آن را شرح می‌دهیم. در بخش 5 نتیجه‌گیری و کارهای آینده را مورد بررسی قرار می‌دهیم.

## 2. شرح مقاله

### 2.1. خاصیت‌های یک خوشه‌بندی مطلوب

خوشه‌بندی گراف، گروه‌بندی راس‌های گراف به خوشه‌ها با در نظر گرفتن ساختار یال‌های گراف، به گونه‌ای است که باید یال‌های درون هر خوشه ماکزیمم و بین خوشه‌ها حداقل تعداد یال موجود باشد. خوشه‌بندی گراف به معنای گروه‌بندی گره‌های گراف ورودی به خوشه‌ها است که موضوع این مقاله می‌باشد.

گرچه یافتن جوامعی که چگالی درونی آن‌ها بالاتر از چگالی بین جوامع باشد، یک کار ساده به نظر می‌رسد، اما از نظر پیچیدگی محاسباتی یک مسئله NP-Hard است، زیرا در گراف‌های بزرگ هم‌پوشانی بین گروه‌ها زیاد است [۲،۳]. بنابراین به روشنی مشخص است که یافتن راه‌حلی که به صورت نسبی خوشه‌بندی را به خوبی (کمترین میزان هم‌پوشانی در خوشه‌ها) و دسته‌های با اندازه بزرگ‌تر را کشف و پارامترهای خوشه‌بندی نظیر ضریب خوشه‌بندی را بهبود بخشد، ضروری به نظر می‌رسد.

متأسفانه هیچ تعریف واحدی از یک خوشه در گراف به‌طور همگانی پذیرفته نشده است و انواع تعاریف در ادبیات مختلف موجود هستند [6]. در محیط (فضای) گراف، هر خوشه باید مستقیماً متصل باشد؛ باید حداقل یک مسیر چندگانه متصل برای هر جفت از گره‌ها درون یک خوشه وجود داشته باشد. اگر راس  $U$  از راس  $V$  قابل دسترس نباشد، آن‌ها نباید در خوشه‌های یکسان (مشابه) گروه‌بندی شوند. علاوه بر این، مسیرها باید درون خوشه‌ها باشد؛ علاوه بر این گره مجموعه  $e$  باید به  $e$  وصل شود. زیرگراف استخراج‌شده به وسیله  $e$  باید به خودی خود متصل باشند، به این معنی که برای دو گره  $U$  و  $V$  در  $e$  به وسیله یک مسیر که از گره‌های  $V \setminus e$  گذر کند، کافی نیست. اما آن‌ها به

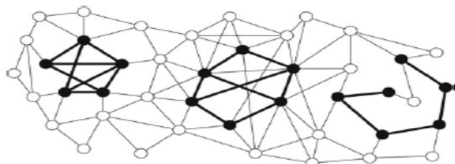
اتصال به کمک یک مسیر که فقط گره‌های شامل  $e$  را ملاقات کند، نیاز دارد. به عنوان یک نتیجه، وقتی خوشه‌بندی یک گراف غیرمتصل با اجرایش شناخته شود، خوشه‌بندی باید معمولاً روی هر جز به‌طور جداگانه، انجام بپذیرد. مگر این که برخی از محدودیت‌های کل درون خوشه‌های حاصله، تحمیل شوند. در برخی از کاربردها ممکن است بخواهیم خوشه‌های با ترتیب (نظم-یکسان) و یا چگالی یکسان به دست آوریم که در این صورت خوشه‌ی محاسبه‌شده در یکی از اجزا، تحت تأثیر خوشه‌بندی‌های اجزای دیگر قرار می‌گیرد. ما یال‌های متلاقی در  $V \in e$  را به 2 گروه طبقه‌بندی می‌کنیم: یال‌های درونی که  $V$  را به دیگر گره‌های در  $G$  متصل می‌کنند و یال‌های خارجی که  $V$  را به گره‌های که در خوشه  $G$  نیستند [۳،۵].

$$\deg_{int}(V, G) = |\Gamma(V) \cap G| \quad (1)$$

$$\deg_{ext}(V, G) = |\Gamma(V) \cap (V \setminus G)| \quad (2)$$

$$\deg(V) = \deg_{int}(V, G) + \deg_{ext}(V, G) \quad (3)$$

مشخص است که وقتی  $\deg_{ext}(V) = 0$  دلالت بر این دارد که  $G$  شامل  $V$  می‌تواند یک خوشه خوب باشد، چون که  $V$  هیچ اتصال بیرونی (خارجی) از آن ندارد. به‌طور مشابه، اگر  $\deg_{int}(V) = 0$  باشد،  $V$  نباید در  $C$  به صورت غیرمتصل به دیگران، برای هر گروه دیگر، وجود داشته باشد. به‌طور کلی بر روی یک زیرمجموعه از گره‌ها، به شکل یک خوشه خوب مورد توافق است، اگر شامل زیرگراف چگال، (پرتراکم) بشود. اما تعداد نسبتاً کمی اتصالات، شامل گره‌ها به گره‌ها در بقیه گراف باشد. به شکل (1) نگاه کنید.



شکل (1): گراف با سه خوشه انتخاب همه اعضای خوشه در سیاه و سفید و لبه‌های داخلی خود کشیده ضخیم‌تر نسبت به لبه‌های دیگر از نمودار کشیده شده است. خوشه‌ای که در سمت چپ است با کیفیت خوب، متراکم‌تر است.

الگوریتم پیشنهادی:

1. The betweenness of all existing edges in the network is calculated first.
2. The edge with the highest betweenness is removed.
3. The betweenness of all edges affected by the removal is recalculated.
4. Steps 2 and 3 are repeated until no edges remain.

#### 4. مطالعه موردی: باشگاه فوتبال

##### 4.1 مشخصات دیتاست و محیط شبیه سازی

این مجموعه دیتاست با انتخاب 35 بازیکن فوتبال جام-جهانی 1998 از تیم‌های مختلف و بررسی ارتباطات اعضا دیتاست با یکدیگر، ایجاد شده است. سپس با توجه به ارتباط یا عدم وجود ارتباط بین اعضا، گراف شبکه اجتماعی آن رسم شده است. در شکل 2، این گراف نمایش داده شده است.

- تعداد افراد عضو شبکه (تعداد گره‌ها شبکه) برابر 35 می‌باشد.
- تعداد ارتباطات بین اعضا (تعداد یال‌های موجود در شبکه) برابر 118 می‌باشد.
- نوع یال‌ها (ارتباطات) جهت‌دار می‌باشد.

شبیه‌سازی این الگوریتم روی سیستم کامپیوتری با مشخصات زیر انجام گرفته است:

- Genuine Intel(R) CPU T2060 @ 1.60GHz
- Memory(RAM): 1.00 GB
- OS: Windows 7 - 32-bit

از ابزار NodeXL که ابزاری با قابلیت برنامه‌نویسی به زبان VBA برای تحلیل شبکه‌های اجتماعی است، استفاده کرده‌ایم.

گراف اولیه بازیکنان فوتبال در زیر آمده است.

یک معیار اندازه‌گیری که به ارزیابی تنگی (کم‌پشت‌بودن) اتصالات از خوشه به بقیه گراف، کمک می‌کند، اندازه برش  $C(G, V \setminus G)$  است. اندازه برش کوچک‌تر بهتر بودن جدای خوشه است. با محاسبه چگالی گراف به‌دست می‌آید، تعیین می‌شود که یک خوشه چقدر متراکم است. چگالی خوشه را به عنوان چگالی داخل یا درون خوشه‌ای در نظر می‌گیریم.

$$\delta_{int}(G) = \frac{|\{V, u\} | V \in G, u \in G\}}{|G|(|G| - 1)} \quad (4)$$

چگالی درون خوشه‌ای، خوشه‌بندی یک گراف مفروض  $G$  به  $k$  خوشه‌ی  $G_1$  و  $G_2$  و .... برابر است با متوسط (میانگین) چگالی‌های درون خوشه‌ای، خوشه‌های دربرگرفته شده است.

$$\delta_{int}(G|G_1, \dots, G_k) = \frac{1}{k} \sum_{i=1}^k \delta_{int}(G_i) \quad (5)$$

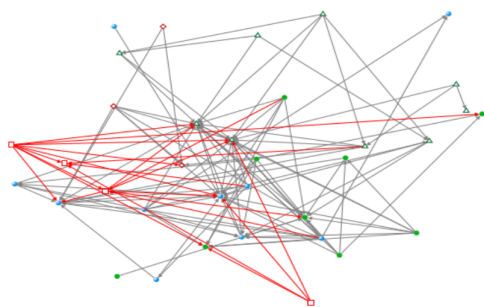
چگالی درون یا داخل خوشه‌ای یک خوشه‌بندی برابر است با نسبت یال‌های درون خوشه به حداکثر تعداد یال‌های ممکن درون خوشه، که به‌طور موثر مجموع اندازه‌های برش تمام خوشه‌ها می‌باشد که در محدوده‌ی  $[0,1]$  است.

$$\delta_{ext}(G|G_1, \dots, G_k) = \frac{|\{V, u\} | V \in G_i, u \in G_j, i \neq j\}}{n(n-1) - \sum_{L=1}^k (|G_L|(|G_L| - 1))} \quad (6)$$

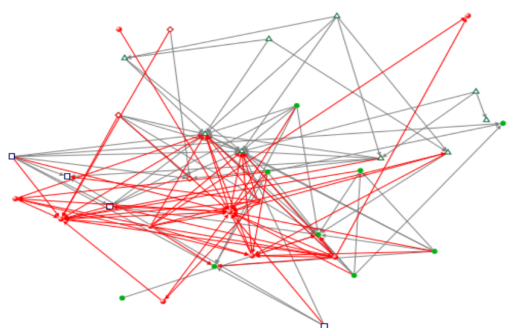
#### 3. الگوریتم پیشنهادی

در این بخش از مقاله یک الگوریتم برای خوشه بندی شبکه‌های اجتماعی شرح می‌دهیم. الگوریتم پیشنهادی ابتدا با به‌دست آوردن کوتاه‌ترین مسیر بین جفت گره‌ها و تعیین درجه مابینیت یال‌ها (تعداد دفعاتی که یک یال در کوتاه‌ترین مسیر بین جفت گره‌ها به عنوان بخشی از مسیر در نظر گرفته می‌شود) عمل می‌کند. پس از تعیین درجه مابینیت یال‌ها، آن‌ها را مرتب می‌کنیم. اکنون با توجه به تعداد خوشه‌ها مدنظر، یال‌ها با بالاترین درجه مابینیت را قطع می‌کنیم. اگر تعداد خوشه‌های به‌دست‌آمده از تعداد مورد-نظر کمتر بود، عملیات فوق را مجدداً تکرار می‌کنیم. در ادامه شبهه کد الگوریتم را آورده‌ایم.

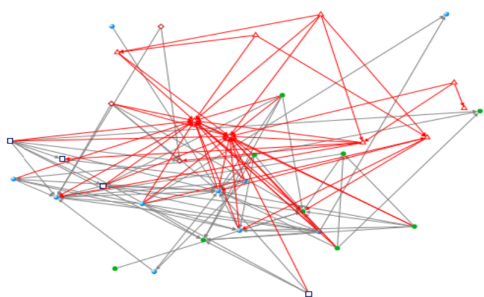
## 4.2 خوشه‌های حاصل شده



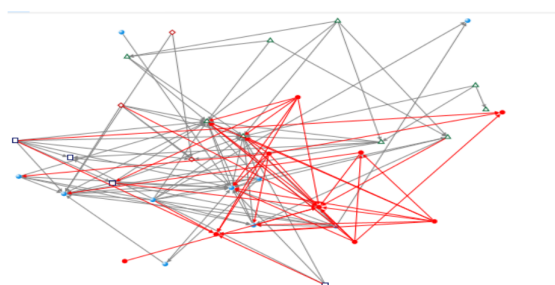
شکل(3): نمایش خوشه G1



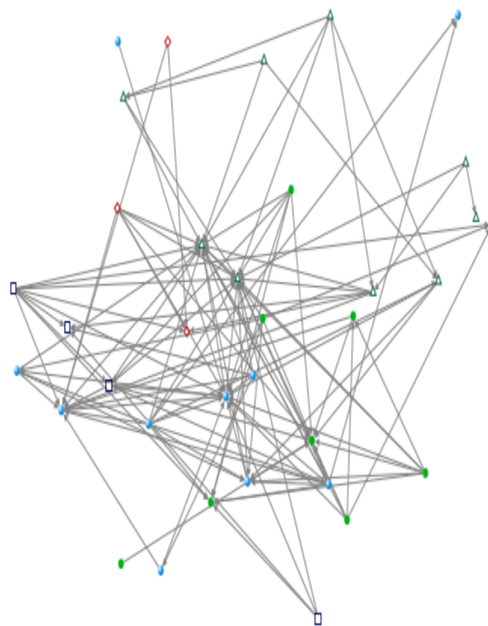
شکل(4): نمایش خوشه G2



شکل(5): نمایش خوشه G3



شکل(6): نمایش خوشه G4



شکل(2): شکل اولیه گراف شبکه اجتماعی باشگاه فوتبال

در ادامه خوشه‌های به دست آمده از الگوریتم نشان داده شده‌اند.

## 4.1 شبیه سازی

جدول(1): ارزیابی خوشه های به دست آمده

شماره خوشه	تعداد کل گره‌های موجود در خوشه	تعداد یال-های ممکن	تعداد یال-های موجود	چگالی درون خوشه ای
G1	5	20	9	0.55
G2	9	72	34	0.37
G3	9	72	35	0.48
G4	8	56	30	0.53
G5	4	12	9	0.75

که چگالی درون خوشه ای از رابطه زیر به دست می آید:

$$\frac{\text{Number of exist edge}}{N(N-1)} \quad (7)$$

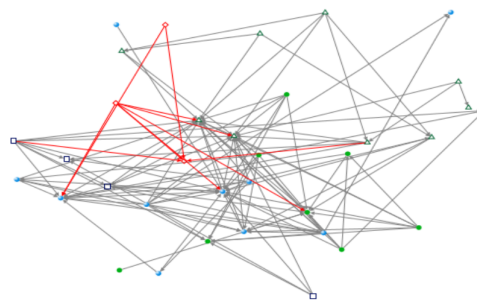
در رابطه فوق N تعداد گره‌های هر خوشه می‌باشد. طبق گفته مرجع 6 هرچه چگالی درون خوشه‌ای به 1 نزدیک‌تر باشد، نشان‌دهنده بالاتر بودن کیفیت خوشه‌بندی می‌باشد. هر چند در شرح مقاله ذکر گردید که هیچ معیار استاندارد و فراگیری برای تعیین خوشه‌ها، وجود ندارد.

از زحمات استاد فرهیخته کشور، جناب آقای دکتر سعید پارسا که همواره ما را با حمایت‌های خویش یاری نموده‌اند، کمال تشکر و قدردانی می‌نماییم.

## 7. مراجع

- [1]. Reichardt J, Bornholdt S, "Detecting fuzzy community structures in complex networks with a potts model," *Physical Review Letters*, 2004, 93(19) : 218701.
- [2]. Yang B, Cheung WK, Liu J, "Community mining from signed social networks," *IEEE Trans. on Knowledge and Data Engineering*, 2007, 19(10) : 1333-1348.
- [3]. Brandes U, Delling D, Gaertler M, Gorke R, Hoefer M, Nikoloski Z, Wagner D, "On modularity clustering," *IEEE Trans. On Knowledge and Data Engineering*, 2008, 20(2) : 172-188.
- [4]. Cartozo CC, Rios PDL, Piazza F, Lio P, "Bottleneck genes and community structure in the cell cycle network of *Spombe*," *PLOS Computational Biology*, 2007, 3(6) : e 103.
- [5]. Shiga M, Takigawa I, Mamitsuka H, "A spectral clustering approach to optimally combining numerical vectors with a modular network," In : Berkhin P, Caruana R, Wu X, eds. *Proc. of the 13th ACM SIGKDD Int'l Conf. on Knowledge Discovery and Data Mining*. New York: ACM Press, 2007. 647-656.
- [6]. S. Boccaletti, V. Latora, Y. Moreno, M. Chavez, and D. Hwang, "Complex networks: Structure and dynamics," *Physics Reports*, 2006.
- [7]. L. Danon, A. Diaz-Guilera, J. Duch, and A. Arenas, "Comparing community structure identification," *Journal of Statistical Mechanics: Theory and Experiment*, 2005.
- [8]. M. Girvan and M. Newman, "Community structure in social and biological networks," in *Proceedings of the National Academy of Sciences*, 2002.
- [9]. J. Reichardt and S. Bornholdt, "Statistical mechanics of community detection," *Physical Review E*, 2006.

[10]. محسن کجباف "پیشنهاد یک الگوریتم برای تعیین تعیین تعارضها درون جوامع در شبکه‌های اجتماعی مجازی مبتنی بر عدد رنگی گراف"، کنفرانس روز رسانه‌های اجتماعی، تهران، تیرماه 91.



شکل (7): نمایش خوشه G5

## 5. نتیجه‌گیری

در این مقاله با مدل کردن شبکه‌های اجتماعی به یک گراف که در آن گره‌ها همان افراد، سازمان‌ها یا گروه‌ها هستند، و یال‌ها همان روابط بین افراد، سازمان‌ها و گروه‌ها می‌باشند، یک گراف بسیار بزرگ شامل تعداد زیادی گره و یال ایجاد شد. با خوشه‌بندی این گراف، و ارزیابی نتایج حاصل شده از این الگوریتم و بررسی خاصیت چگالی درون خوشه‌ای، و با توجه به این‌که گفته شد که هر چه چگالی درون خوشه‌ای به 1 نزدیک‌تر باشد، کیفیت خوشه‌بندی بالاتر است، نتیجه می‌گیریم که الگوریتم پیشنهاد شده برای گراف‌های با اندازه کوچک‌تر، کاراتر است (با توجه به جدول 1). زیرا هر چه گراف کوچک‌تر باشد، احتمالاً خوشه‌های کشف شده در گراف، دارای تعداد گره کمتری هستند و بنابراین چگالی درون خوشه‌ای حتی با تعداد کمی یال، سریع‌تر به 1 میل خواهد کرد.

## 6. کارهای آینده

زمینه تحقیقاتی آینده در این حوزه، کشف خوشه‌های هم‌پوشان در شبکه‌های اجتماعی، کشف گره‌های رهبر درون خوشه‌ها در شبکه‌های اجتماعی، خوشه‌بندی مقیاس‌پذیر برای شبکه‌های اجتماعی آنلاین و خوشه‌بندی دینامیک می‌باشد.

## سپاسگذاری