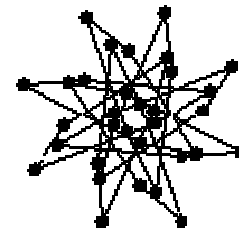


به نام خدا



عنوان پایان نامه
تشخیص اجتماعات با رویکرد ترکیبی در
شبکه‌های اجتماعی

ارائه دهنده: رسول حسین‌زاده
جناب آقای مهندس حسین علیزاده
جناب آقای دکتر اسلام ناظمی

اسفند ۹۱

فهرست مطالب

❖ مقدمه

❖ کاربردها و چالش‌ها

❖ کارهای مرتبط

❖ روش پیشنهادی

❖ تشخیص اجتماعات ترکیبی

❖ تشخیص تعداد اجتماعات

❖ نتایج آزمایشات

❖ نتیجه‌گیری و کارهای آینده

تشخیص اجتماعات

شبکه های اجتماعی



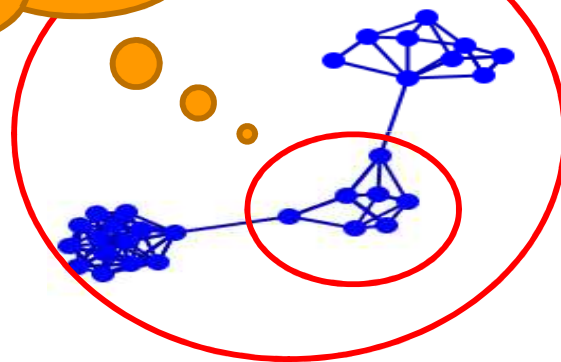
تشخیص اجتماعات

شبکه‌های اجتماعی

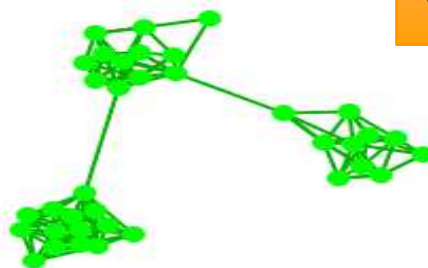
❖ یک اجتماع مجموعه‌ای از گره‌ها و لبه‌های بین آن‌ها می‌باشد که چگالی لبه‌ها داخل آن بیشتر از چگالی لبه‌های بین آن با سایر گره‌های شبکه است. (گروه، زیرگروه، خمشه و ...)

این گره‌های پرتراکم نیز یک اجتماع هستند

❖ به پیدا کردن یک اجتماع هستند شبکه اجتماعی و خروجی اجتماعات مختلف خواهد بود

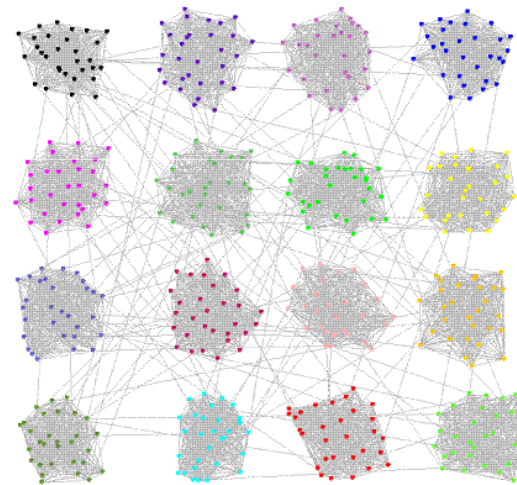
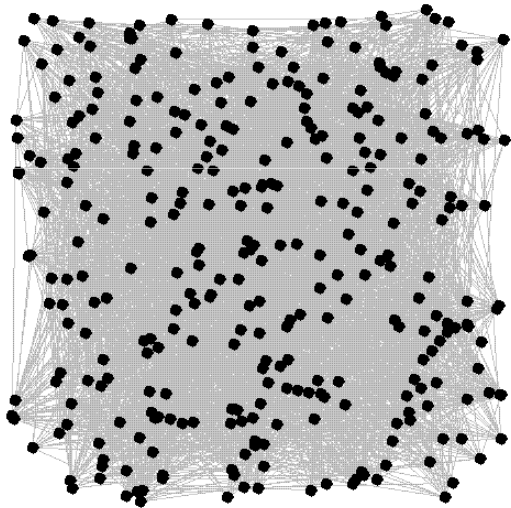


مفهوم اجتماع یک مفهوم ذهنی است و توافق آراییی از تعریف اجتماع وجود ندارد



در مسائل بسیاری از جمله:

- ❖ درک ساختار شبکه
- ❖ یافتن اجتماعاتی با ویژگی خاص
- ❖ قابل مشاهده کردن گراف
- ❖ بهبود موتورهای جستجو
- ❖ پیمایش موضوعی
- ❖ بازاریابی
- ❖ تبلیغات



❖ چگونه می‌توان یک شبکه را به اجتماعات اصلی آن تقسیم کرد؟

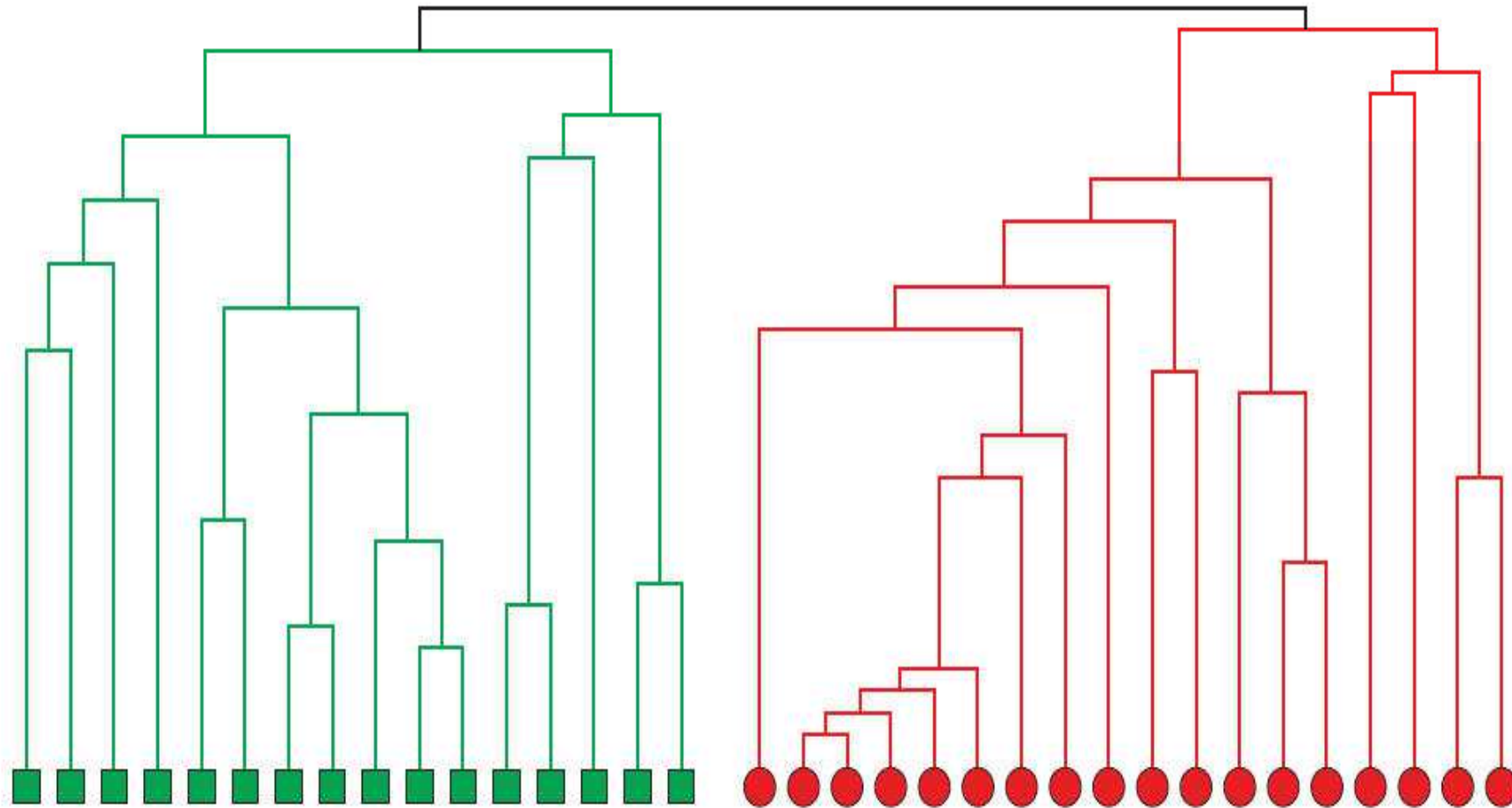
❖ در شبکه‌های واقعی هیچ اطلاعاتی درباره تعداد اجتماعات وجود ندارد، چگونه می‌توان تعداد صحیح اجتماعات را پیدا نمود؟

❖ چگونه می‌توان دقت نتایج روش‌های تشخیص اجتماع را بالا برد؟

خوشه‌بندی ترکیبی

پیمانه‌بندی

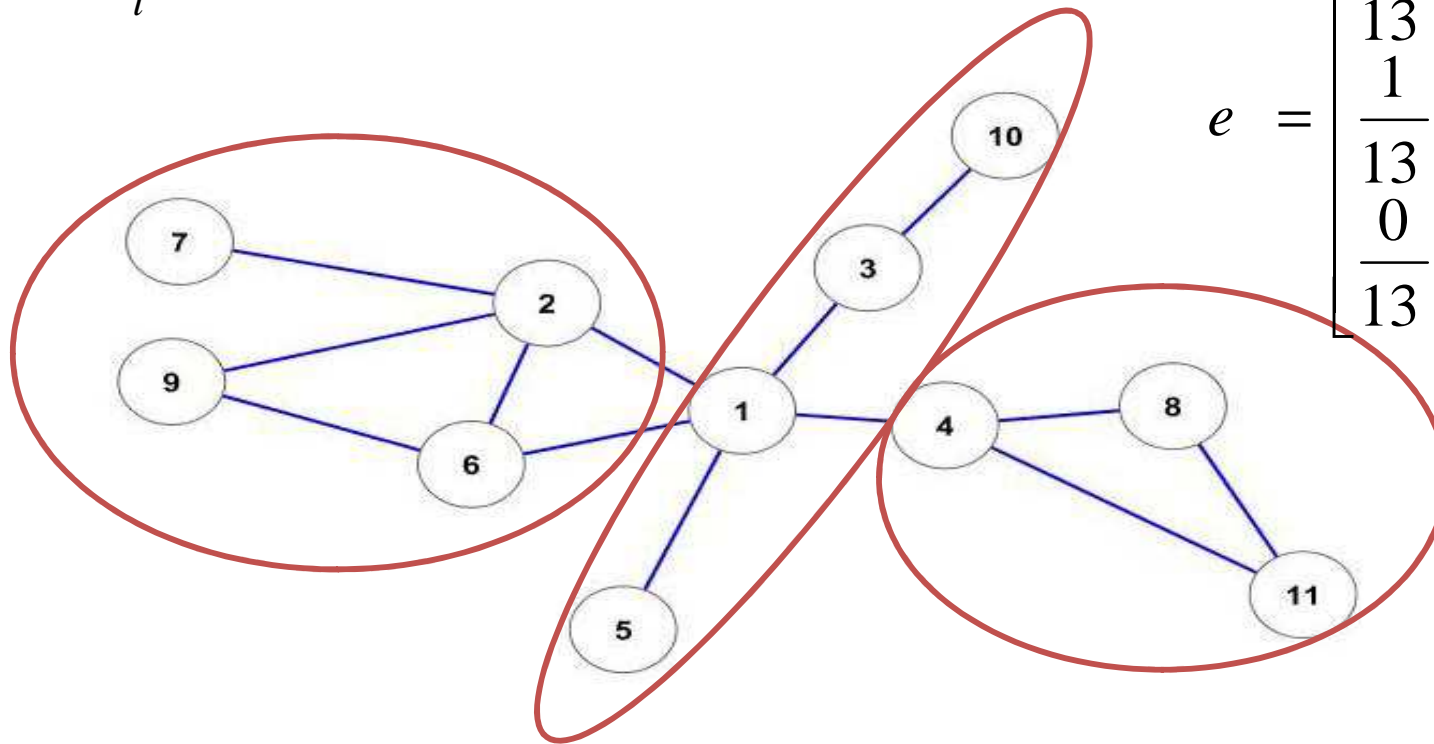
تشخیص اجتماعات



تجمعی: هر گره در ابتدا یک اجتماع و با یک معیار مشابهت به هم متصل می‌شوند.

پیمانه بندی (Modularity) یکی از معیارهای ارزیابی و تشخیص اجتماع می-باشد.

$$Q = \sum_i (e_{ii} - a_i^2)$$



$$e = \begin{bmatrix} \frac{4}{13} & \frac{1}{13} & \frac{0}{13} \\ \frac{1}{13} & \frac{3}{13} & \frac{0.5}{13} \\ \frac{0}{13} & \frac{0.5}{13} & \frac{3}{13} \end{bmatrix}$$

$$\frac{9}{23}$$

$$Q = \left[\left(\frac{4}{13} - \left(\frac{5}{13} \right)^2 \right) + \left(\frac{3}{13} - \left(\frac{4.5}{13} \right)^2 \right) + \left(\frac{3}{13} - \left(\frac{3.5}{13} \right)^2 \right) \right] = 0.451$$

خوشه‌بندی ترکیبی

پیمان‌بندی

تشخیص اجتماعات

مبتنی بر پیشینه‌سازی پیمان‌بندی: این روش که از پیمان‌بندی به عنوان یک معیار مشابهت استفاده می‌نماید و امکان ترکیب با هر روش دیگری را خواهد داشت.

الگوریتم‌های تکاملی: در این روش هر کروموزوم کل گره-های شبکه را شامل می‌شود. با تعریف توابع crossover و mutation گره-های اجتماع را جابجا می‌کند. معیار ارزیابی نیز پیمان‌بندی می‌باشد.

خوشه‌بندی ترکیبی

پیمانه‌بندی

تشخیص اجتماعات

❖ به خاطر پیچیدگی مسئله خوشه‌بندی و ضعف روش‌های خوشه‌بندی پایه، امروزه اکثر مطالعات به سمت روش‌های خوشه‌بندی ترکیبی هدایت شده است.

❖ نتایج خوشه‌بندی ترکیبی در مقابل خوشه‌بندی ساده

✓ استحکام (Robustness)

✓ نو بودن (Novelty)

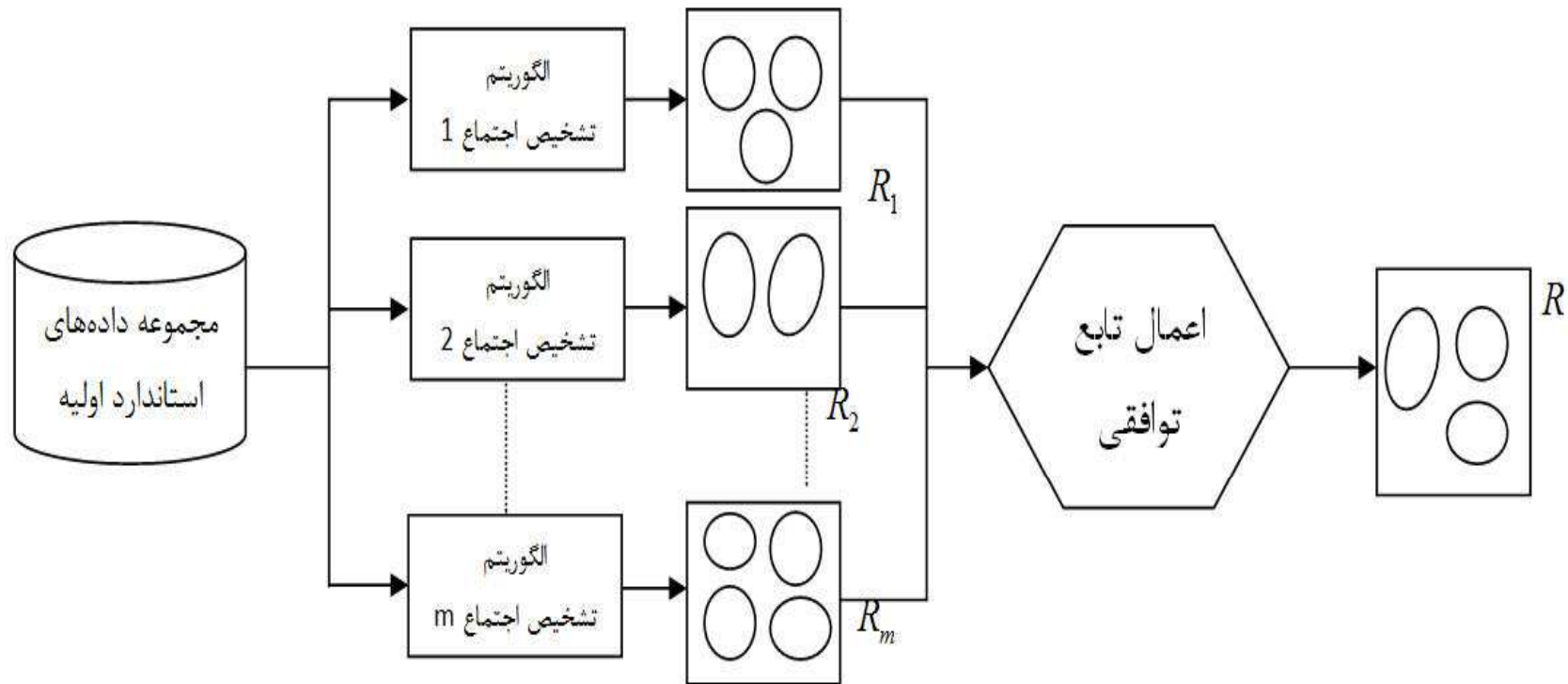
✓ پایداری (Stability)

✓ انعطاف‌پذیری (Flexibility)

تشخیص تعداد اجتماعات

تشخیص اجتماعات ترکیبی

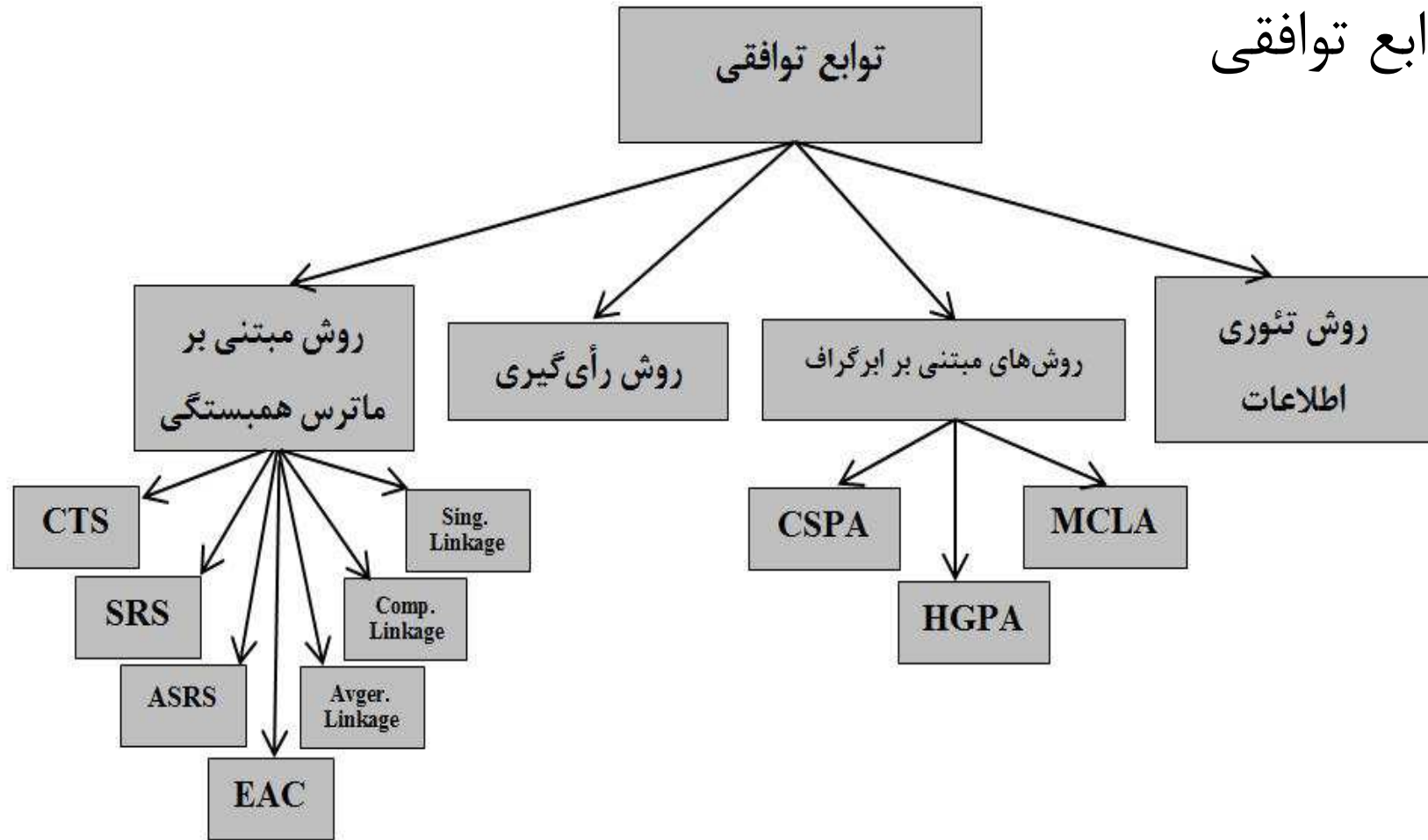
❖ شکل زیر روش پیشنهادی را نشان می دهد.



تشخیص تعداد اجتماعات

تشخیص اجتماعات ترکیبی

❖ توابع توافقی



❖ شبه کد روش پیشنهادی

Input: A(Adjacency matrix from Graph)

N(The Number of Community)= ComputeN

for Number of community detection method *do*

while (Can change initialization parameter) *do*

ECD+:=run(community detection method)

end while

if(Can change final state community detection method) *then*

ECD+:=run(community detection method)

end if

end for

ECD: Ensemble members of Community Detection methods

/*Each column show one result of community detection method*/

Co.association matrix:=run(Consensus function)

C:=run(Hierarchical clustering method)

Output:C(Final community detection result)

تشخیص تعداد اجتماعات

تشخیص اجتماعات ترکیبی

❖ شبه کد تعداد اجتماعات

Input: A(Adjacency matrix from Graph)

for (Number of base community detection method) *do*

NECD+=Number Of Community(run(community detection method))

end for

NECD: Number of Community in Ensemble community Detection methods

NECD:=Refine(NECD)

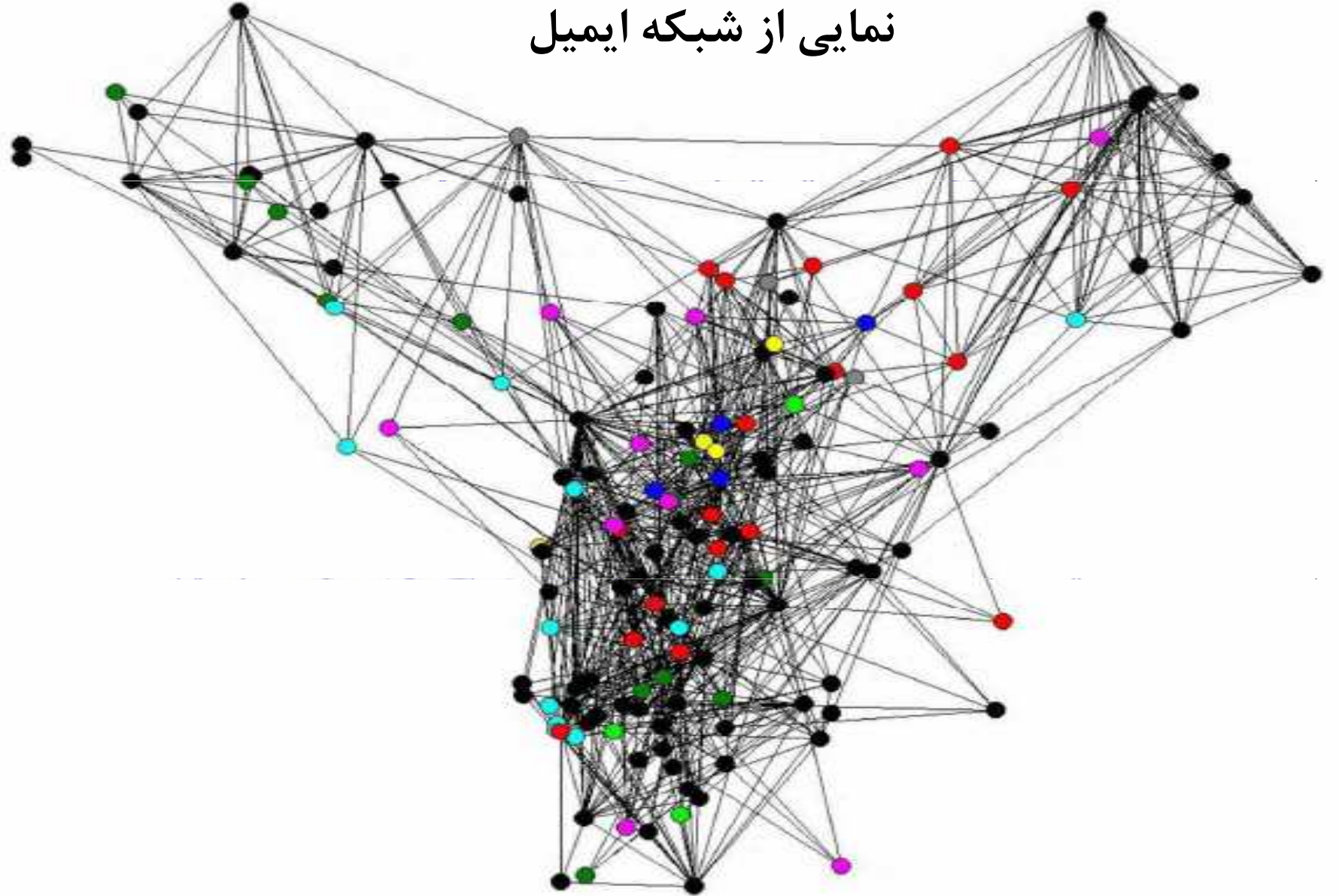
N:=Avg(NECD)

Output:N(Final community number)

نتایج آزمایشات

مجموعه داده

نمایی از شبکه ایمیل



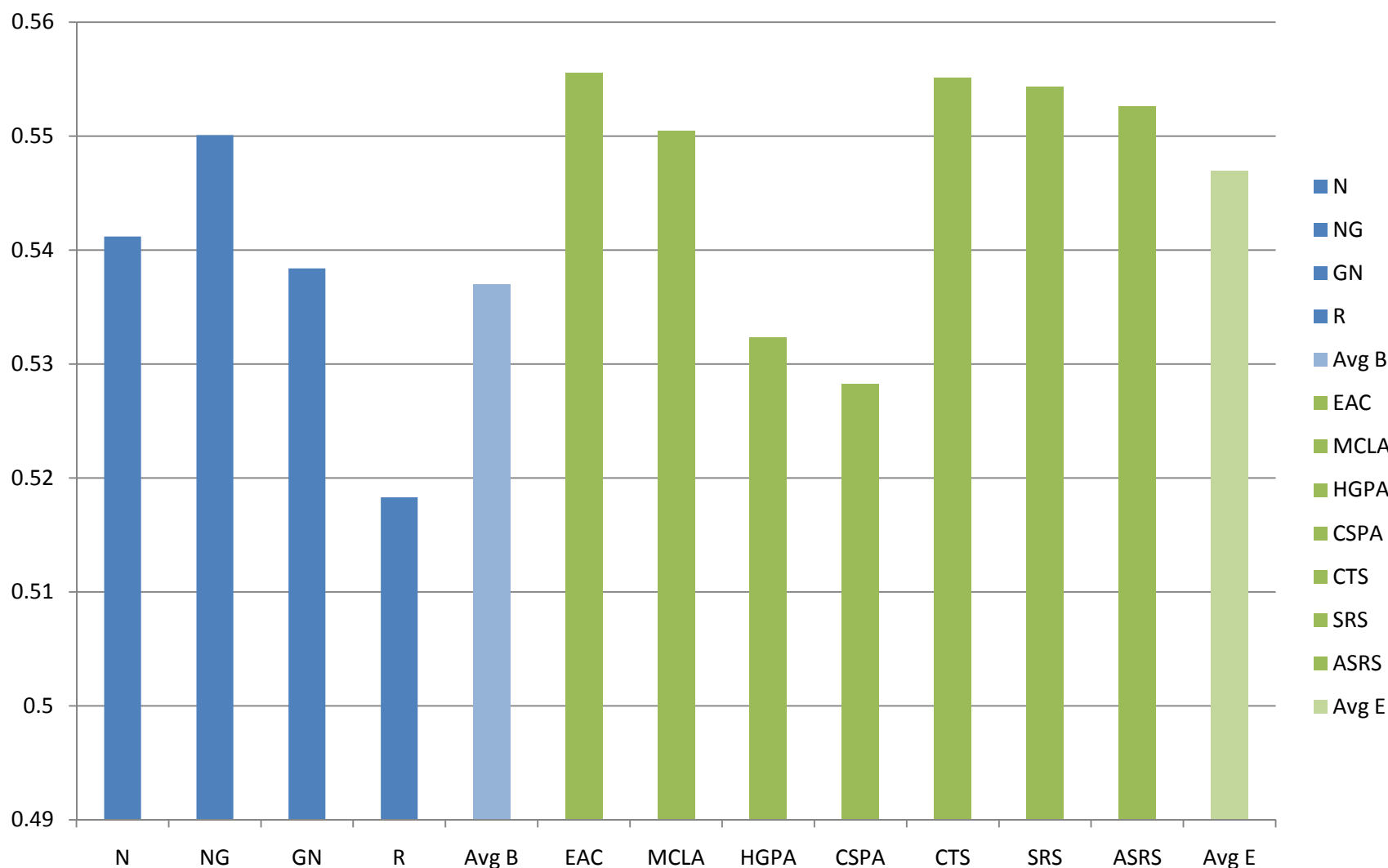
پیمانه بندی بدست آمده از روش های پایه و ترکیبی تشخیص اجتماعات بر روی داده های مختلف

| Avg E | ECD | | | | | | | Avg B | Basic Community Detection Method | | | | Data Set |
|--------|--------|--------|--------|--------|--------|--------|--------|--------|----------------------------------|--------|--------|--------|-----------|
| | ASRS | SRS | CTS | CSPA | HGPA | MCLA | EAC | | R | GN | NG | N | |
| 0.4519 | 0.4522 | 0.4518 | 0.4528 | 0.4522 | 0.4488 | 0.4528 | 0.4528 | 0.4266 | 0.3984 | 0.4304 | 0.4518 | 0.4258 | آزمایشی |
| 0.4068 | 0.4179 | 0.4197 | 0.4188 | 0.3836 | 0.3683 | 0.4197 | 0.4197 | 0.3982 | 0.3853 | 0.4009 | 0.3980 | 0.4086 | کاراته |
| 0.5922 | 0.5985 | 0.5985 | 0.6019 | 0.5687 | 0.5775 | 0.5988 | 0.6019 | 0.5777 | 0.5581 | 0.5714 | 0.6027 | 0.5788 | فوتبال |
| 0.4424 | 0.4445 | 0.4438 | 0.4447 | 0.4375 | 0.4390 | 0.4428 | 0.4450 | 0.4405 | 0.4394 | 0.4397 | 0.4442 | 0.4389 | جاز |
| 0.4294 | 0.4334 | 0.4374 | 0.4368 | 0.3989 | 0.4288 | 0.4334 | 0.4377 | 0.4238 | 0.4019 | 0.4334 | 0.4269 | 0.4332 | متابولیکی |
| 0.5566 | 0.5685 | 0.5763 | 0.5765 | 0.5137 | 0.5246 | 0.5597 | 0.5769 | 0.5437 | 0.5036 | 0.5463 | 0.5766 | 0.5485 | ایمیل |
| 0.9491 | 0.9534 | 0.9528 | 0.9546 | 0.9432 | 0.9394 | 0.9458 | 0.9551 | 0.9487 | 0.9419 | 0.9472 | 0.9511 | 0.9547 | علم شبکه |
| 0.5469 | 0.5526 | 0.5543 | 0.5551 | 0.5282 | 0.5323 | 0.5504 | 0.5555 | 0.5370 | 0.5183 | 0.5384 | 0.5501 | 0.5412 | AVG |

نتایج آزمایشات

مجموعه داده

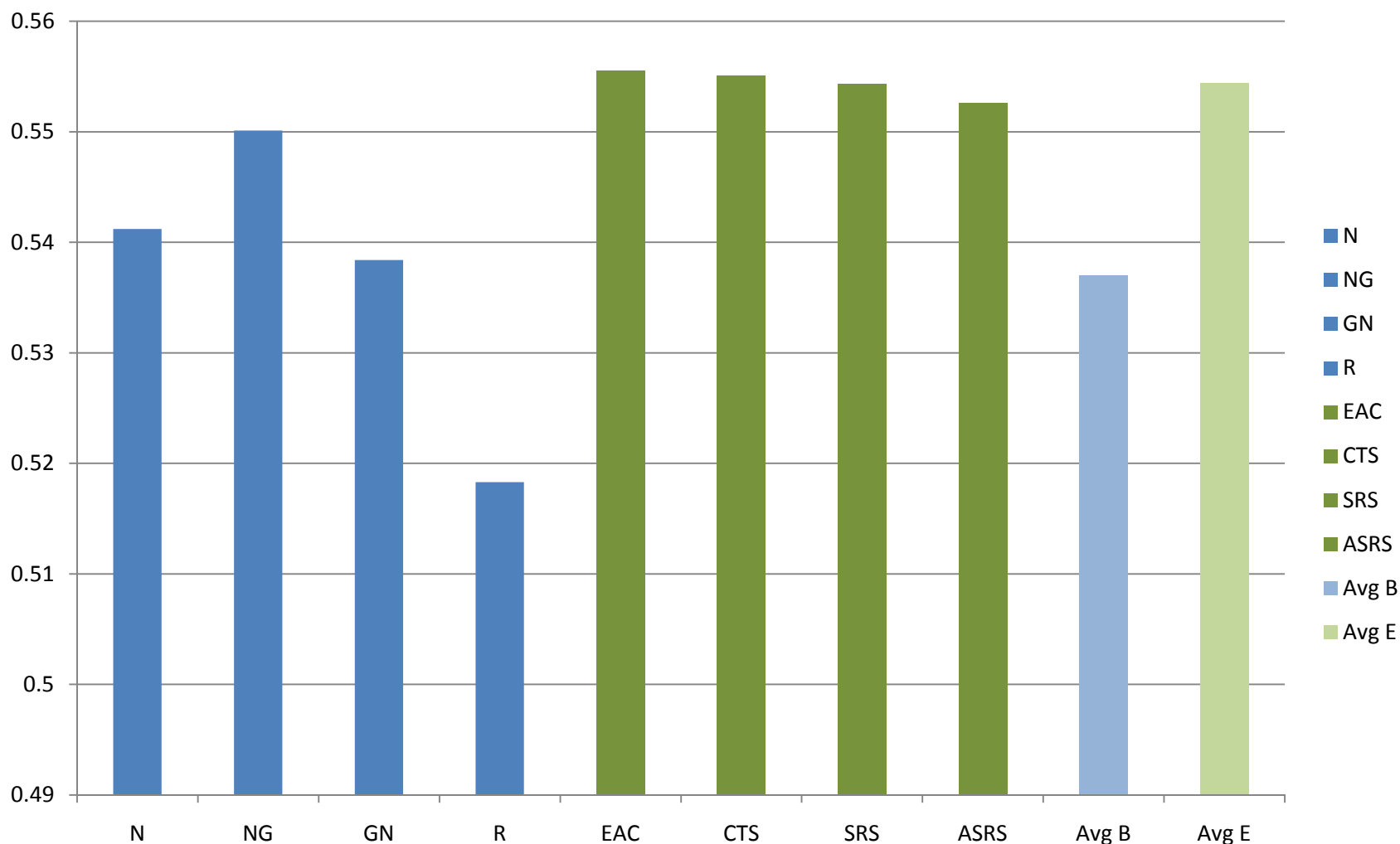
در نمودار زیر روش‌های مختلف پایه و ترکیبی تشخیص اجتماعات را مورد مقایسه قرار دادیم.



نتایج آزمایشات

مجموعه داده

در نمودار زیر بعضی روش های ترکیبی تشخیص اجتماعات را مورد مقایسه قرار دادیم.



۲۰
۲۳

کارهای آینده

نتیجه‌گیری

❖ با توجه به گسترش روز افزون وب و شبکه‌های اجتماعی به ابزارهایی برای غلبه بر پیچیدگی آنها نیاز است.

✓ تشخیص اجتماعات یکی از این ابزارها و موضوع‌ها جدید و دارای جای کار بسیار می‌باشد.

❖ روش پیشنهادی تشخیص اجتماعات ترکیبی

✓ دارای دو قسمت ایجاد نتایج اولیه و اجرای تابع توافقی می‌باشد.

❖ روش تشخیص تعداد اجتماعات

❖ مزایا و معایب

✓ دارای دقت بالا، مقیاس پذیری بالا، عدم نیاز به دانستن تعداد اجتماعات، دارای پایداری و

اطمینان در نتایج نهایی، قابل اعمال بر روی گراف‌های وزن دار، جهت دار می‌باشد.

✓ همچنین روش پیشنهادی قابل استفاده برای هر سازمان است.

- ❖ ارائه الگوریتم‌ها سریع‌تر برای روش‌های تشخیص اجتماعات پویا
- ❖ امکان ترکیب داده‌کاوی ساختار وب با داده‌کاوی استفاده از وب
- ❖ ارائه یک معیار جدید برای ارزیابی دقت اجتماعات

مراجع منتخب

- ✓ J. Leskovec, 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, pp. 631–640, 2010.
- ✓ S. Fortunato, “Community detection in graphs”, *Physics Reports*, Vol. 486, No. 3-5, pp. 75–174, 2010.
- ✓ D. Easley and J. Kleinberg, “Networks, Crowds, and Markets: Reasoning about a Highly Connected World”, *Cambridge University Press*, June 10, 2010.
- ✓ B. Liu, “Web DataMining Exploring Hyperlinks, Contents, and Usage Data”, *Springer*, 2007.
- ✓ M.A. Porter, J.P. Onnela and P.J. Mucha, “Communities in networks”, *Notices of the American Mathematical Society*, V. 56, No. 9, pp. 1082–1097, 2009.
- ✓ M.E.J. Newman, “Communities, modules and large-scale structure in networks”, *Nature Physics*, Vol. 8, January 2012.
- ✓ M.E.J. Newman, “Fast algorithm for detecting community structure in very large networks”, *Physical review E*, Vol. 69, 2004.



مراجع منتخب

- ✓ A. Lancichinetti, “Community detection algorithms: a comparative analysis”, *Physical Review E*, Vol. 80, No. 5, p. 056117, 2009.
- ✓ A. Fred and A.K. Jain, “Data Clustering Using Evidence Accumulation”, *In Conference on Pattern Recognition, ICPR02*, Quebec City, pp. 276 – 280, 2002.
- ✓ M.E.J. Newman and M. Girvan, “Finding and evaluating community structure in networks”, *Physical review E*, Vol. 69, No. 2, p. 26113, 2004.
- ✓ A. Clauset, M.E. J. Newman, and C. Moore, “Finding community structure in very large networks”, *Physical Review E*, Vol. 70, No. 6, p. 66111, 2004.
- ✓ A. Strehl and J. Ghosh, “Cluster Ensembles - A Knowledge Reuse Framework for Combining Multiple Partitions”, *Journal of Machine Learning Research*, pp. 583-617, 2003.
- ✓ X. S. Zhang et al., “Modularity optimization in community detection of complex networks”, *Euro Physics Letters(EPL)*, Vol. 87, p. 38002, 2009.
- ✓ S. Vega-Pons and J. Ruiz-Shulcloper, “A survey of clustering ensemble algorithms”, *International Journal of Pattern Recognition and Artificial Intelligence*, Vol. 25, No. 3, pp. 337-372, 2011.



مراجع منتخب

- ✓ M.E.J. Newman, “Modularity and community structure in networks”, Proceedings of the National Academy of Sciences, Vol. 103, No. 23, p. 8577, 2006.
- ✓ C. Pizzuti, “Ga-net: A genetic algorithm for community detection in social networks,” Parallel Problem Solving from Nature–PPSN X, pp. 1081–1090, 2008.
- ✓ M. Lipczak and E. Milios, “Agglomerative genetic algorithm for clustering in social networks”, In Proceedings of the 11th Annual conference on Genetic and evolutionary computation, pp. 1243–1250, 2009.
- ✓ N. Iam-on and S. Garrett, “LinkCluE: A MATLAB Package for Link-Based Cluster Ensembles”, In Journal of Statistical Software, Issue 9, Volume 36, August 2010.
- ✓ A. Lancichinetti and S. Fortunato, “Consensus clustering in complex networks”, Nature. Scientific Reports, March 2012.

✓ عزیزاده ح.، مینایی بیدگلی ب.، “خوشه-بندی ترکیبی مبتنی بر زیرمجموعه-ای از نتایج اولیه”، پایان نامه کارشناسی ارشد، دانشگاه علم و صنعت، اسفند ۱۳۸۷.





آماده پاسخگویی به سوالات شما هستیم