

**Amin hassanZarei**

**401723094**

**Complex and dynamic networks final project report**

**Instructor : Dr. Rahmani**

**فهرست مطالب**

[1. مقدمه 3](#_Toc138446087)

[2. گام اول 3](#_Toc138446088)

[3. گام دوم 5](#_Toc138446089)

[4. گام سوم 7](#_Toc138446090)

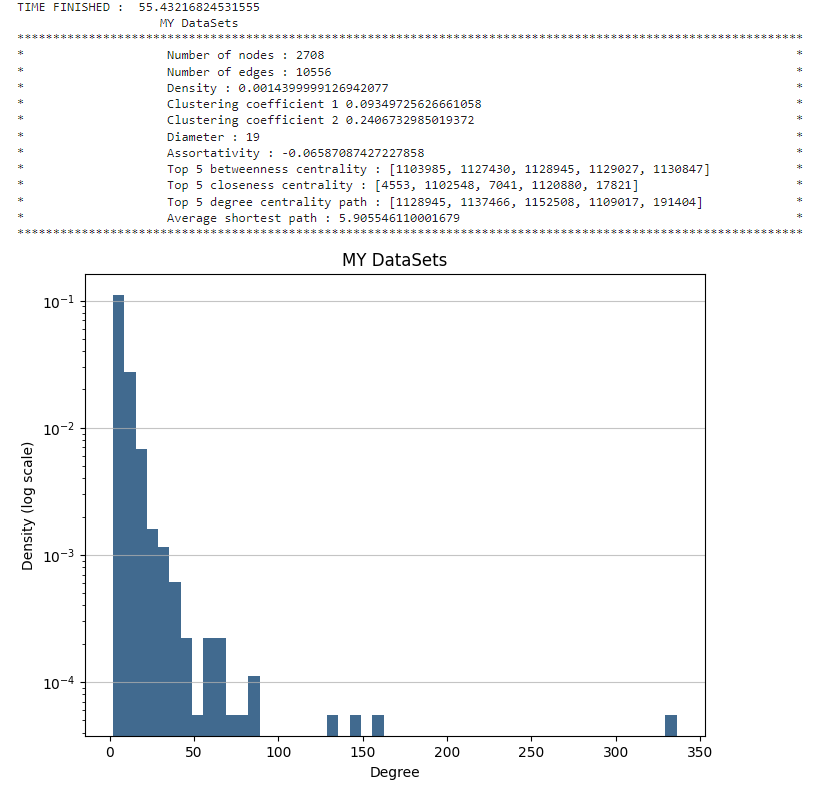
[5. گام چهارم 8](#_Toc138446091)

[6. گام پنجم 8](#_Toc138446092)

# مقدمه

در این گزارش تلاش بر این است که سوالات مطرح شده . در فایل پروژه پایانی درس شبکه‌های‌پیچیده‌پویا[[1]](#footnote-1) را پاسخ دهیم . زبان استفاده شده برای مجموعه‌داده‌ها[[2]](#footnote-2) پایتون می‌باشد البته برای نمایش داده از کتابخانه ژوپیتر[[3]](#footnote-3) استفاده شده‌است . برای راحتی کار مجموعه‌داده‌ها را به فایل csv تبدیل شده‌اند.

# گام اول

ابتدا با اجرای فایل setting\_one\_time.pyفایل‌های مجموعه‌داده‌ها را که در پوشه Datasets قرار داده‌ایم را تبدیل به csv می‌کنیم . سپس گراف متناظر جهت‌دار را با اجرای فایل complex\_network\_final.ipynb تشکیل‌داده و اطلاعات‌ آماری‌ و معیار‌های گراف را گزارش می‌دهیم . شکل ‏2.1   


شکل ‏2.1 نمایش اطلاعات آماری و معیار‌های گراف متناظر با داده و همچنین نمودار چگالی به درجه شبکه پیچیده متناظر با داده ها

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| تعداد نودها[[4]](#footnote-4) | تعداد لبه‌ها[[5]](#footnote-5) | چگالی[[6]](#footnote-6) | ضریب خوشه‌بندی 1[[7]](#footnote-7) | ضریب خوشه‌بندی 2[[8]](#footnote-8) | قطر[[9]](#footnote-9) | دسته‌بندی[[10]](#footnote-10) | 5 نود با بین‌بودن بالا[[11]](#footnote-11) | 5 نود با مرکزیت‌نزدیکی بالا[[12]](#footnote-12) | 5 نود با مرکزیت‌درجه بالا[[13]](#footnote-13) | میانگین کوتاه‌ترین مسیر[[14]](#footnote-14) |
| 2078 | 10556 | 0.0014 | 0.0934 | 0.2406 | 19 | 0.0658- | 1103985 | 4553 | 1128945 | 5.9055 |
|  |  |  |  |  |  |  | 1127430 | 1102548 | 1137466 |  |
|  |  |  |  |  |  |  | 1128945 | 7041 | 1152508 |  |
|  |  |  |  |  |  |  | 1129027 | 1120880 | 1109017 |  |
|  |  |  |  |  |  |  | 1130847 | 17821 | 191404 |  |

Table 1معیارها و داده‌های آماری برای گراف متناظرجهت‌دار

همچنین با مشاهده جدول چگالی به درجه می‌توان فهمید که نمودار به شدت از قانون‌توانی[[15]](#footnote-15) تبعیت می‌کند پس احتمال بالا این شبکه یک شبکه انسانی می‌باشد .

# گام دوم

با توجه به جدول Table 1 معیارها را تحلیل می‌کنیم :

- \*\*تراکم\*\*: این اندازه‌گیری نسبت تعداد یال‌ها را به حداکثر تعداد لبه‌های ممکن در شبکه محاسبه می‌کند. این نشان می‌دهد که شبکه چقدر به خوبی متصل است. تراکم بالاتر به معنای اتصالات بیشتر و پراکندگی کمتر است. این شبکه دارای چگالی 0.00144 است که بسیار کم است و نشان دهنده یک شبکه پراکنده است.

- \*\*ضریب خوشه بندی\*\*: این اندازه‌گیری میانگین احتمال اینکه دو همسایه یک گره همسایه یکدیگر باشند را محاسبه می کند. این نشان می دهد که شبکه چقدر خوشه‌بندی است. ضریب خوشه‌بندی بالاتر به معنای مثلث بیشتر و انسجام محلی بیشتر است. این شبکه دارای دو ضریب خوشه بندی است: 0.0935 و 0.2407. اولی بر اساس تعداد مثلث ها و دومی بر اساس تعداد سه ضلعی متصل است. هر دو مقدار نسبتاً پایین هستند و سطح پایینی از خوشه‌بندی را در شبکه شما نشان می‌دهند.

- \*\*قطر\*\*: این اندازه‌گیری طولانی‌ترین کوتاه ترین مسیر را بین هر جفت گره در شبکه محاسبه می‌کند. این نشان می‌دهد که گره‌ها چقدر از هم در شبکه فاصله دارند. قطر کمتر به معنی فواصل کوتاه‌تر و انتشار سریعتر اطلاعات است. قطر این شبکه 19 است که نسبتاً زیاد است و یک شبکه بزرگ با گره های دور را نشان می‌دهد.

- \*\*دسته‌بندی\*\*: این اندازه گیری همبستگی بین درجات گره‌های متصل در شبکه را محاسبه می‌کند. این نشان می‌دهد که گره‌ها از نظر اتصال چقدر شبیه یا متفاوت هستند. طبقه‌بندی مثبت به این معنی است که گره‌ها تمایل به اتصال با گره‌هایی با درجه مشابه دارند، در حالی که یک دسته‌بندی منفی به این معنی است که گره ها تمایل به اتصال با گره های درجات مختلف دارند. این شبکه دارای دسته‌بندی 0.0659- است که کمی منفی است و نشان دهنده تمایل ضعیف برای اختلاط نامناسب در شبکه است.

- \*\*مرکزیت‌بینی\*\*: این اندازه‌گیری تعداد کوتاه‌ترین مسیرهایی را که از یک گره می گذرد، شمارش می‌کند. این نشان می‌دهد که یک گره از نظر کنترل جریان اطلاعات در شبکه چقدر تأثیرگذار یا مرکزی است. مرکزیت بین‌المللی بالاتر به معنای قدرت بیشتر و پتانسیل بیشتر برای عمل به عنوان یک پل یا گلوگاه در شبکه است. این شبکه دارای پنج گره با مرکزیت میانی بالا است: 1103985، 1127430، 1128945، 1129027، 1130847. این گره ها ممکن است برای گسترش فرآیندها یا استحکام شبکه در شبکه مهم باشند.

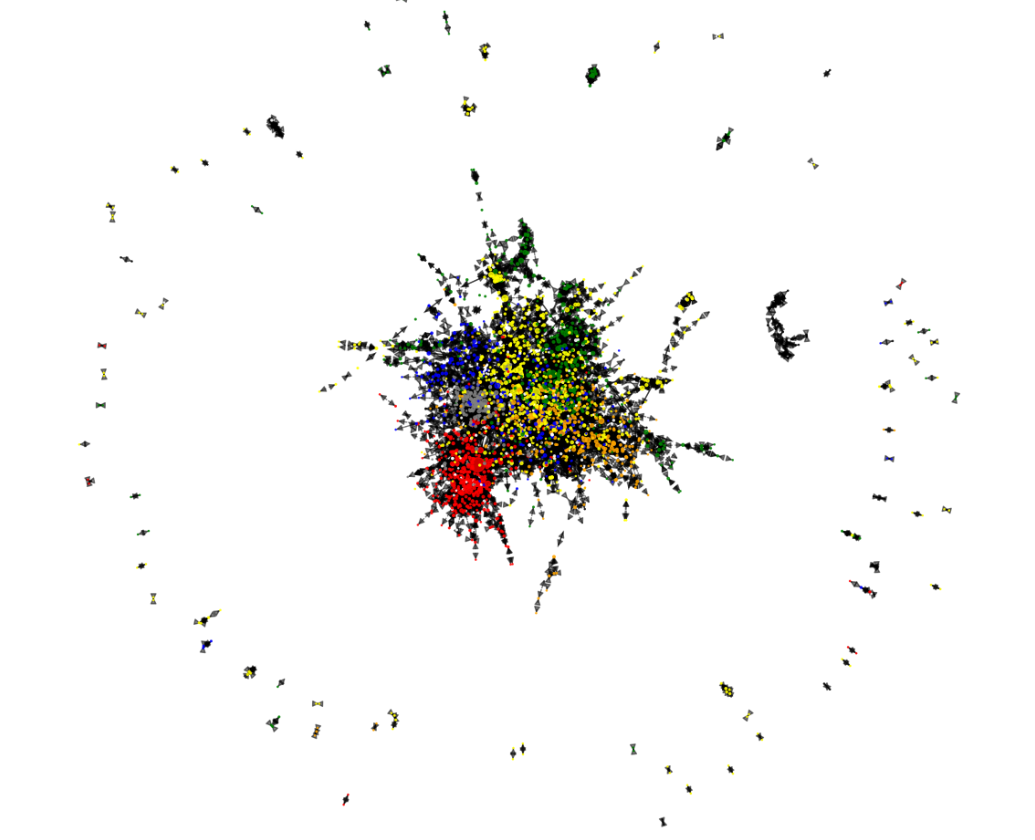
- \*\*مرکزیت‌نزدیکی\*\*: این اندازه گیری میانگین طول کوتاه‌ترین مسیر از یک گره به تمام گره‌های دیگر شبکه را محاسبه می‌کند. این نشان می‌دهد که یک گره از نظر دسترسی به سایر گره های شبکه چقدر در دسترس یا مرکزی است. مرکزیت‌نزدیکی بیشتر به معنای فواصل کوتاه‌تر و دسترسی سریعتر به اطلاعات در شبکه است. این شبکه دارای پنج گره با مرکزیت‌نزدیک بالا است: 4553، 1102548، 7041، 1120880، 17821. این گره‌ها ممکن است برای فرآیندهای ارتباطی یا هماهنگی در شبکه مهم باشند.

- \*\*مرکزیت‌درجه\*\*: این اندازه‌گیری تعداد لبه‌های متصل به یک گره را می‌شمارد. این نشان می‌دهد که یک گره از نظر داشتن اتصالات در شبکه چقدر محبوب یا مرکزی است. مرکزیت‌درجه بالاتر به معنای همسایگان بیشتر و پتانسیل بیشتر برای تعامل با گره‌های دیگر در شبکه است. این شبکه شما دارای پنج گره با مرکزیت‌درجه بالا است: 1128945، 1137466، 1152508، 1109017، 191404. این گره ها ممکن است برای فرآیندهای اجتماعی یا مشارکتی در شبکه مهم باشند.

- \*\*متوسط کوتاه‌ترین مسیر\*\*: این اندازه گیری میانگین طول کوتاه‌ترین مسیر بین هر جفت‌گره در شبکه را محاسبه می‌کند. این نشان می دهد که شبکه به طور کلی چقدر خوب متصل یا کارآمد است. میانگین کوتاه‌ترین مسیر به معنای فاصله‌های کوتاه‌تر و انتشار سریع‌تر اطلاعات در سراسر شبکه است. این شبکه میانگین کوتاه‌ترین مسیر 5.9055 دارد که نسبتاً زیاد است و یک شبکه بزرگ و پراکنده با چند مسیر طولانی را نشان می‌دهد.

بر اساس این معیارها، پنج گره مهم در این شبکه را می‌توان به‌عنوان 1103985، 1127430، 1128945، 1129027 و 1130847 شناسایی کرد، زیرا دارای بالاترین مرکزیت بینی هستند. با این حال، همچنین شایان ذکر است که گره هایی با بالاترین مرکزیت‌نزدیکی و مرکزیت‌درجه نیز ممکن است نقش مهمی در شبکه ایفا کنند.

# گام سوم



شکل ‏4.1 گراف رنگ آمیزی شده با استفاده از برچسب

اولین نکته‌ای که در گراف به چشم می‌خورد تراکم بالا در مرکزیت آن است که قانون‌توانی را نشان می‌دهد همچنین رنگ زرد بیشترین ارتباط را با باقی نودها دارا می‌باشد. و نکته جالب جدا بودن تقریبی دسته رنگ های قرمز و آبی از بقیه و تراکم آنها میان خودشان می‌باشد.

با استفاده از روش girvan\_newman ما 78 عدد community داریم و در روش Louvain حدود 730 عدد کامیونیتی داریم که شکل آنها در فایل complex\_network\_final.ipynb نمایش داده شده.

میانگین اینکه در هر community برچسب نودها برابر باشد در روش girvan\_newman 58 درصد است و در روش Louvain این مقدار برابر 61 درصد می‌باشد. پس میزان از روش های community detection با دقت حدود 60 درصد می‌توان گره‌ها را براساس برچسب طبقه بندی کرد البته میزان standard deviation برحسب تعداد نودهای از هر نوع در هر community هم محاسبه شده که دقت میانگین را در نطر می‌گیرد و در فایل ipynb قابل مشاهده است.

میزان modularity شبکه با روش Girvan\_Newman برابر 0.13072559406724654 می‌باشد ،که نشان دهنده وجود community در شبکه است. همچین میزان آن در روش Louvain برابر 0.5505540130904358 می‌باشد ، و چون بالاتر از 0.3 است که نشان‌دهنده وجود strong community است .

# گام چهارم

نسبت همسایگان گره‌هایی با برچسب "Unknown" که برچسب "L1" دارند را با بقیه گره‌ها در نظر می‌گیریم و آن را به عنوان‌شانس میزان ابتلا به بیماری گزارش می‌دهیم .

1152896 | 0.8

1153891 | 1.0

1153166 | 1.0

1153724 | 1.0

1153728 | 1.0

1153101 | 0.3333333333333333

1153150 | 0.3333333333333333

1152564 | 1.0

1153169 | 0.5

1152307 | 0.3333333333333333

1152448 | 0.3333333333333333

1152975 | 0.3333333333333333

1153703 | 0.3333333333333333

1153811 | 0.25

1138027 | 0.5

1153148 | 0.3333333333333333

1153003 | 0.25

1154500 | 1.0

1152308 | 1.0

1138091 | 1.0

1152244 | 1.0

1153877 | 1.0

1153879 | 1.0

1153889 | 1.0

1152436 | 1.0

# گام پنجم

در اینجا با استفاده از سه الگوریتم به پیشبینی برچسب هر گره می‌پردازیم و در نهایت آنها را به فایل گره‌ها وارد می‌کنیم. ما از سه الگوریتم knn, rbf, svm برای اینکار استفاده کردیم. که خروجی در کدها مشخص می‌باشد نکته جالب شبیه بودن پیشبینی rbf , svm است درحالی که knn مشخصا برای اینکار خوب نبوده و همه برچسب همه گره‌ها را "L2" پیشبینی کرده.

1. Complex and dynamic Network [↑](#footnote-ref-1)
2. Dataset [↑](#footnote-ref-2)
3. Jupiter notebook [↑](#footnote-ref-3)
4. Number of nodes [↑](#footnote-ref-4)
5. Number of edges [↑](#footnote-ref-5)
6. Density [↑](#footnote-ref-6)
7. Clustering coefficient 1 [↑](#footnote-ref-7)
8. Clustering coefficient 2 [↑](#footnote-ref-8)
9. Diameter [↑](#footnote-ref-9)
10. Assortativity [↑](#footnote-ref-10)
11. Top 5 betweenness centrality [↑](#footnote-ref-11)
12. Top 5 closeness centrality [↑](#footnote-ref-12)
13. Top 5 degree centrality [↑](#footnote-ref-13)
14. Average shortest path [↑](#footnote-ref-14)
15. Power law [↑](#footnote-ref-15)