

بسمه تعالی

شبکه‌های پیچیده – تمرین دوم

* خروجی این تمرین یک فایل pdf و یک فایل اسکریپت پایتون است (۲ فایل)

۱- سه شبکه تصادفی (ER) به صورت زیر ایجاد کنید:

a. $G(500, 0.1)$

b. $G(1000, 0.2)$

c. $G(2000, 0.25)$

الف) هر شبکه را به صورت بصری نمایش داده در داکيومنت قرار دهید.

ب) شبکه‌ها را به فرمت مناسب ذخیره و هر یک را در Gephi مجدداً باز کنید و با layout مناسب هر یک را بازنمایش و در داکيومنت قرار دهید

ج) تمامی معیارهای قابل محاسبه در Gephi را روی هر شبکه به دست آورده و در یک جدول در داکيومنت گزارش دهید

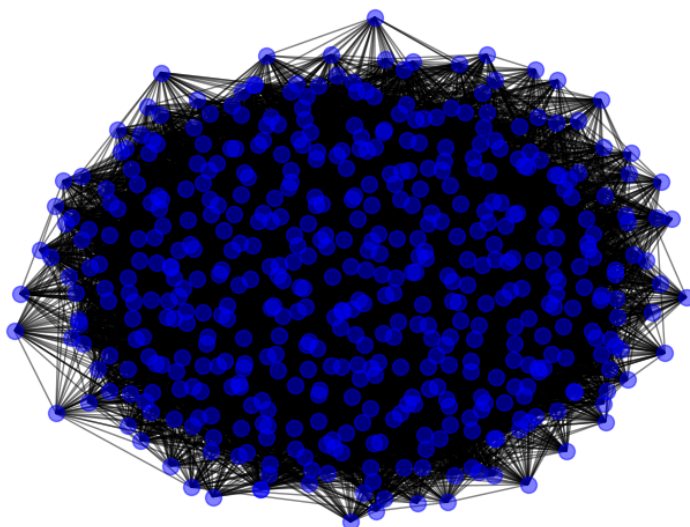
د) سعی کنید نتایج به دست آمده در جدول را تحلیل (مقایسه) کنید (بر حسب پارامترهای n, p).

لطفاً هر دو فایل را زیپ کرده و در lms (فقط lms) بارگذاری کنید.

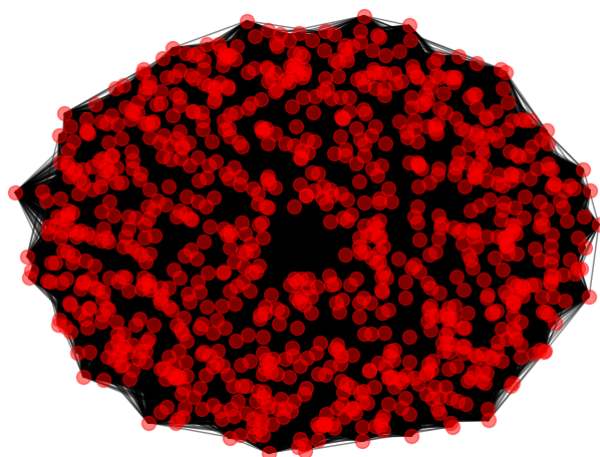
پاسخ ها:

پاسخ الف) کدهای پایتون در فایل زیپ قرار دارند.

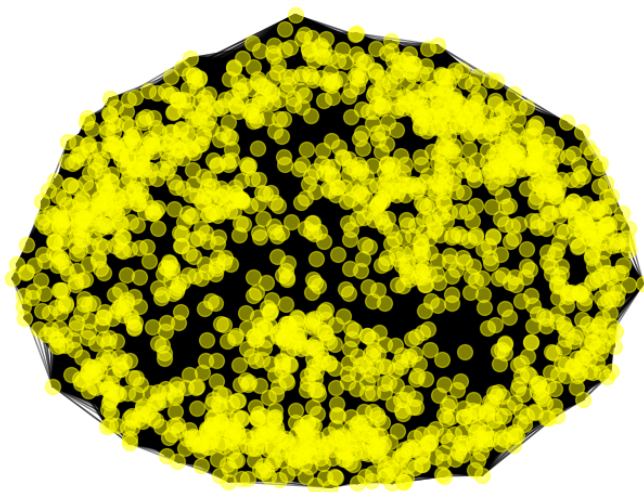
شبکه تصادفی a:



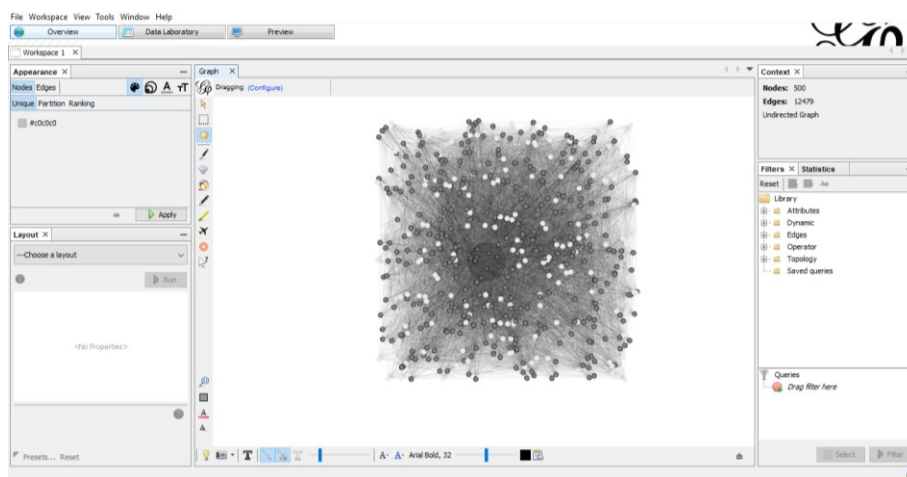
شبکه تصادفی b:



شبکه تصادفی c:



پاسخ ب) در خروجی فایل پایتون یک فایل با پسوند gexf. به وجود می آید که این فرمت برای نرم افزار Gephi یک فرمت deprecated است بنابراین یکبار فایل را در نرم افزار باز می کنیم و به کمک save as آن را با فرمت مناسب gephi. ذخیره می کنیم و مجدداً فایل جدید را با نرم افزار Gephi باز می کنیم.

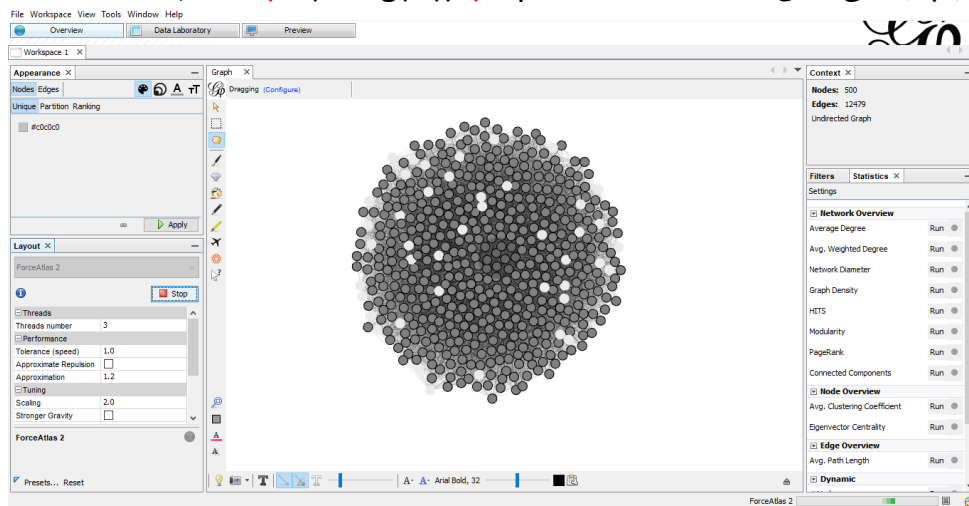


برای انتخاب Layout مناسب ، با توجه به راهنمای

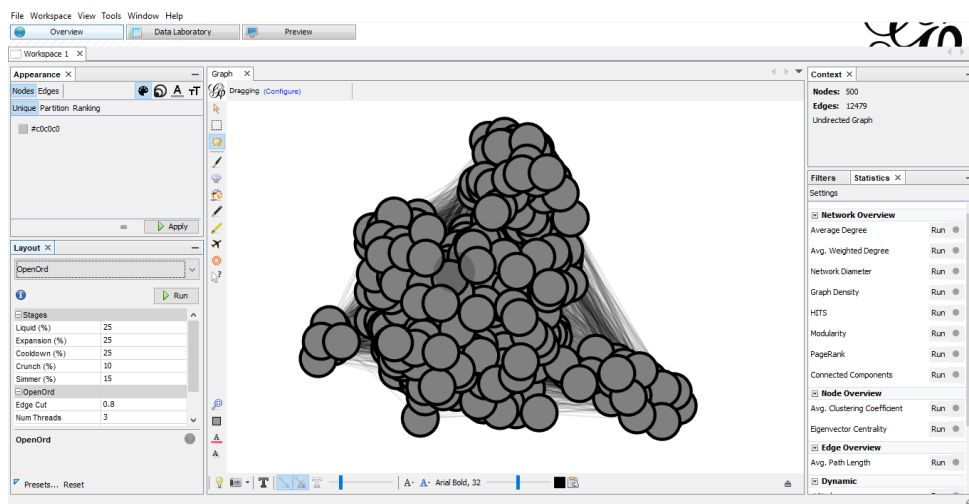
<https://gephi.org/tutorials/gephi-tutorial-layouts.pdf>

و ابعاد شبکه های مورد نظر ، بیشتر Layout ها قابل اعمال بر روی گراف های این تمرین خواهند بود، اما ما در این جا با توجه به امکانات پردازشی و سخت افزاری خود، OpenOrd layout را انتخاب میکنیم زیرا این Layout در نرم افزار gephi نسبتاً سریع تر از بقیه است. البته توصیه می شود برای گراف هایی با تعداد راس های کمتر از 10K از ForceAtlas Layout یا ForceAtlas 2 استفاده شود اما همان طور

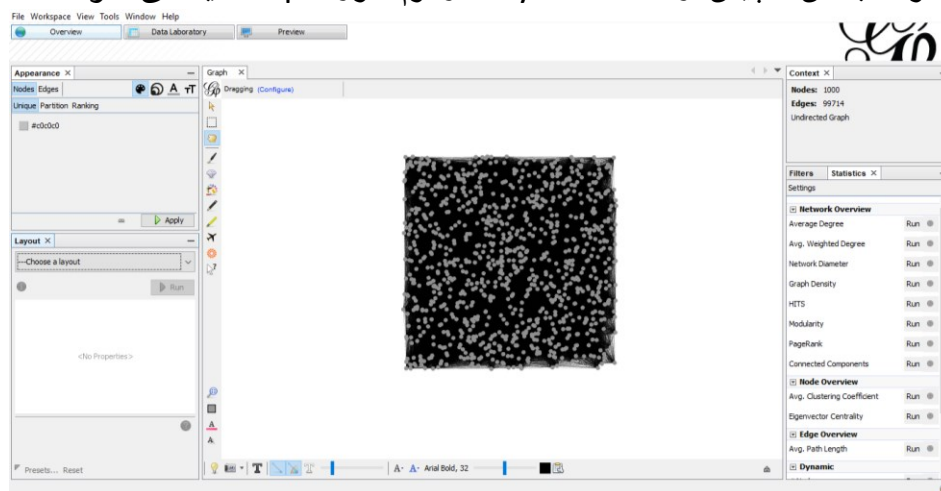
که در تصویر زیر مشخص است اجرای این Layout در سیستم مورد آزمایش با زمان زیادی همراه بود در نتیجه الگوریتمی با پیچیدگی زمانی کمتر مانند OpenOrd ($O(N \log(N))$) انتخاب شد.



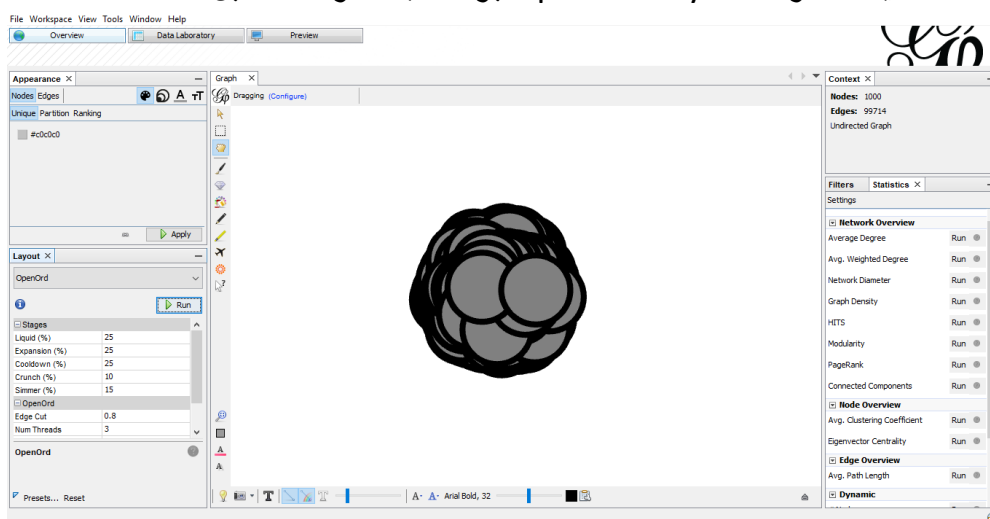
در تصویر زیر نتیجه اجرای OpenOrd layout برای شبکه گراف a قابل مشاهده است.



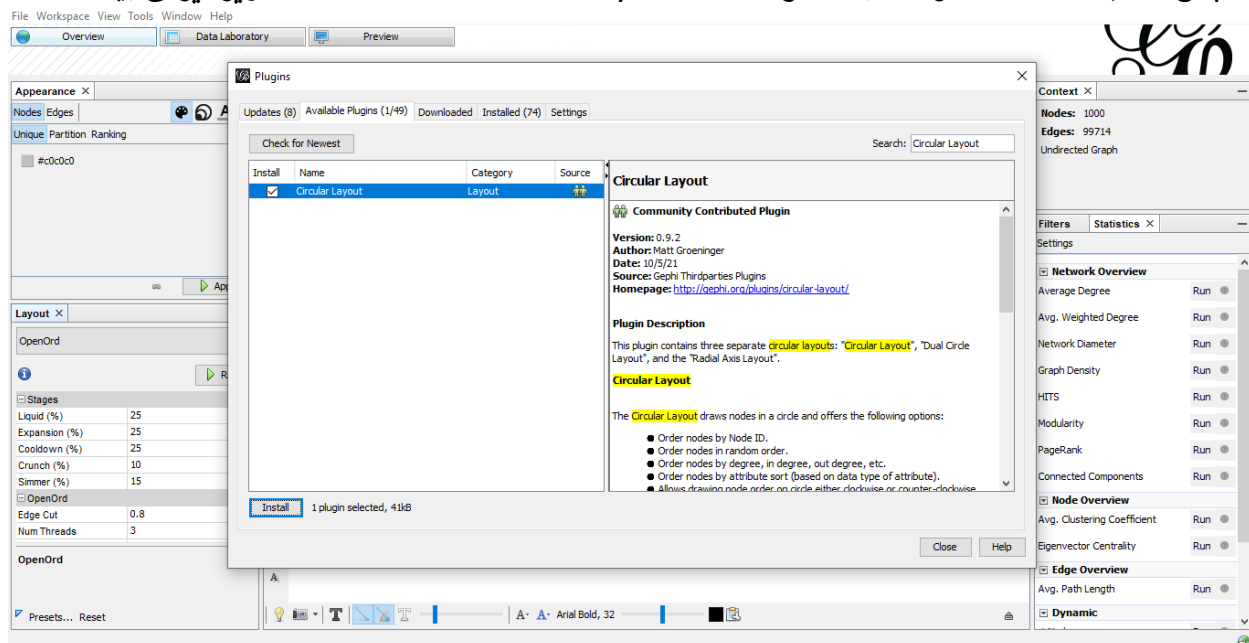
در تصویر زیر گراف بخش b پیش از اعمال Layout در نرم افزار Gephi دیده می شود.



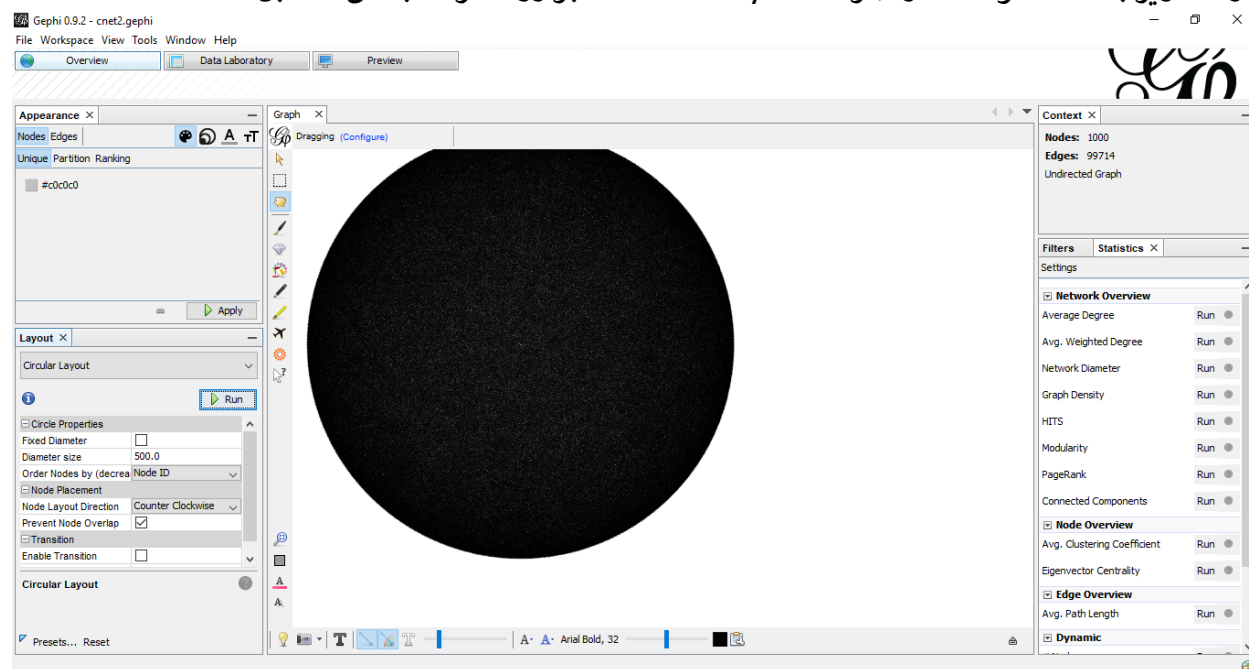
در تصویر بعدی، نتیجه اجرای OpenOrd layout برای شبکه گراف b قابل مشاهده است.

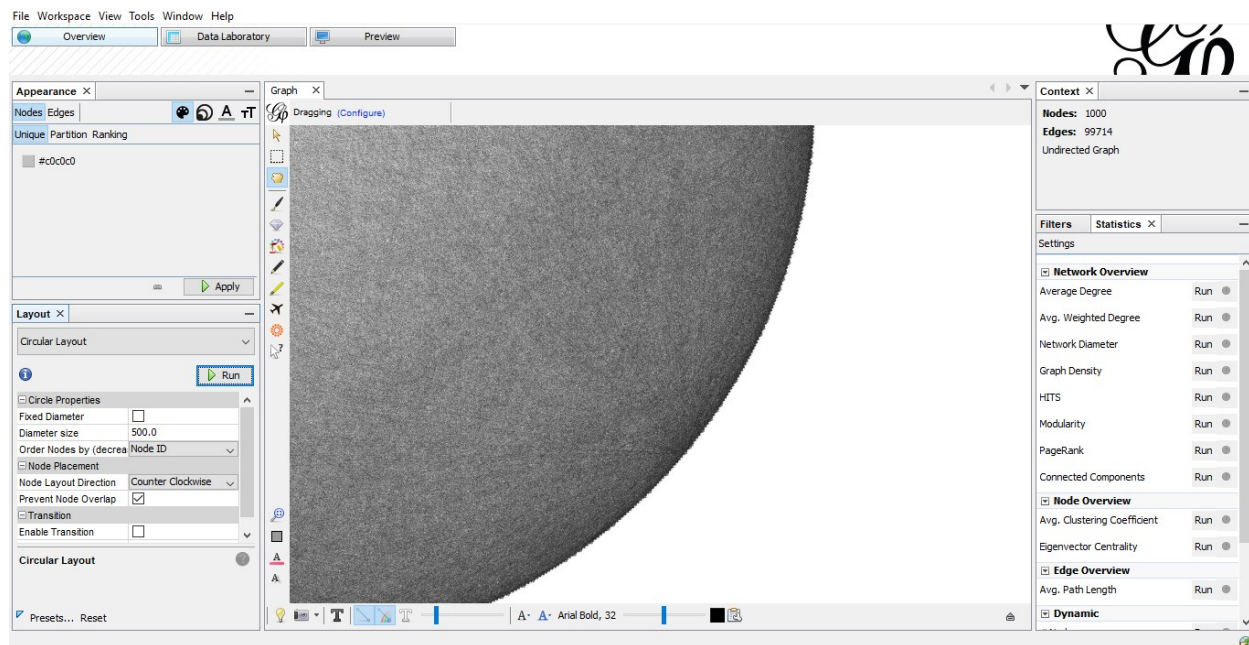


همچنین میتوان از طریق Tools > Plugins > Available Plugins بازنمایش های دیگری را در برنامه gephi نصب و استفاده کرد که یک نمونه از آن Circular Layout است که در تصویر زیر می بینید:

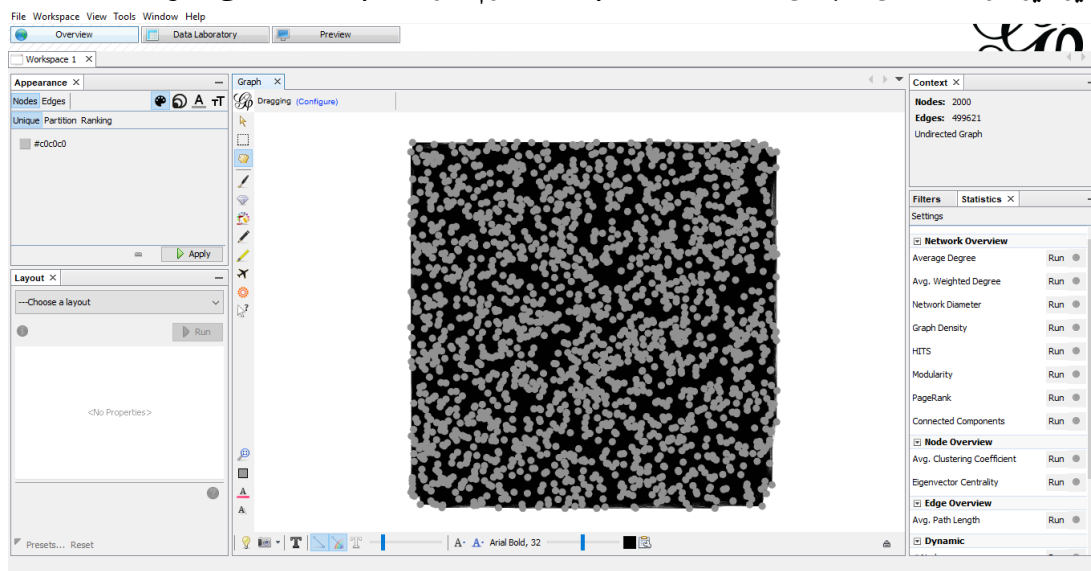


در تصاویر بعدی، نمونه ای از اجرای Circular Layout بر روی گراف بخش b قابل مشاهده است.

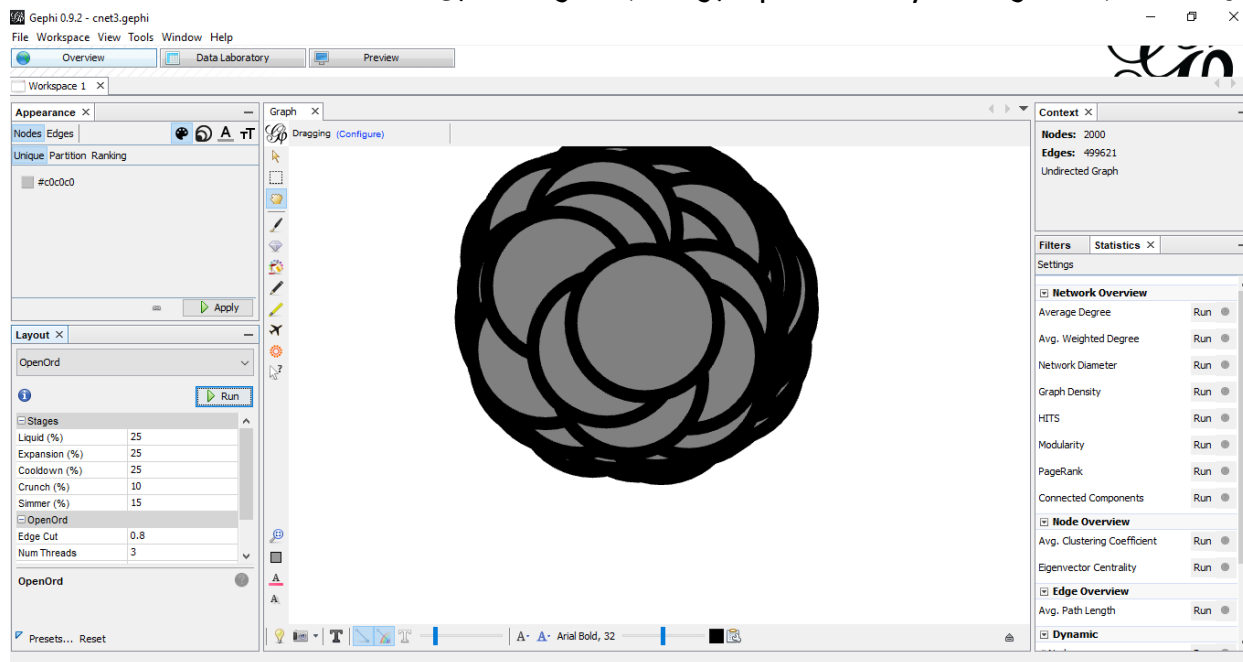




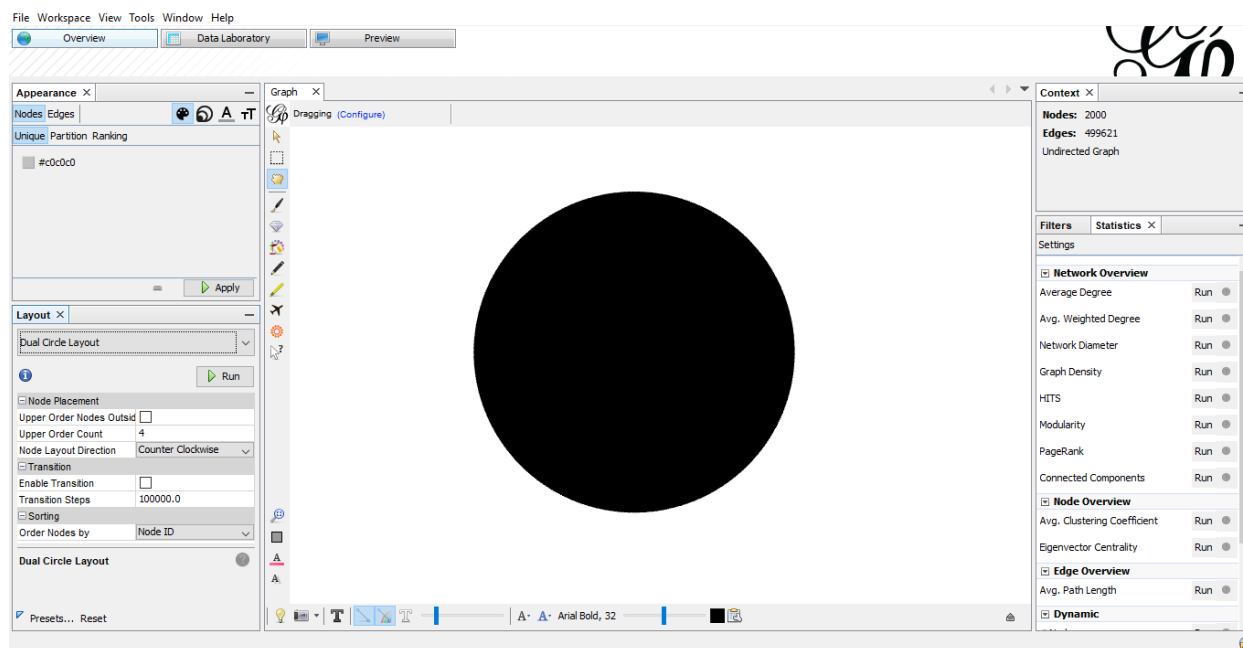
در تصویر زیر گراف بخش c پیش از اعمال Layout در نرم افزار Gephi دیده می شود.



در ادامه، نتیجه اجرای OpenOrd layout برای شبکه گراف c قابل مشاهده است.



همچنین در تصویر بعد، نتیجه اجرای Dual Circle layout برای شبکه گراف c قابل مشاهده است.



پاسخ ج) برای اعمال و محاسبه معیارهای شبکه از قسمت statistics در نرم افزار Gephi استفاده می کنیم.

در دو تصویر زیر، نتیجه اعمال محاسبات معیارهای شبکه را برای گراف شبکه بخش a نشان می دهد. لازم به ذکر است این محاسبات پس از اعمال layout های OpenOrd و Yifan Hu صورت گرفته و همان طور که دیده می شود فقط تفاوت اندکی در بخش Modularity مشاهده می شود.

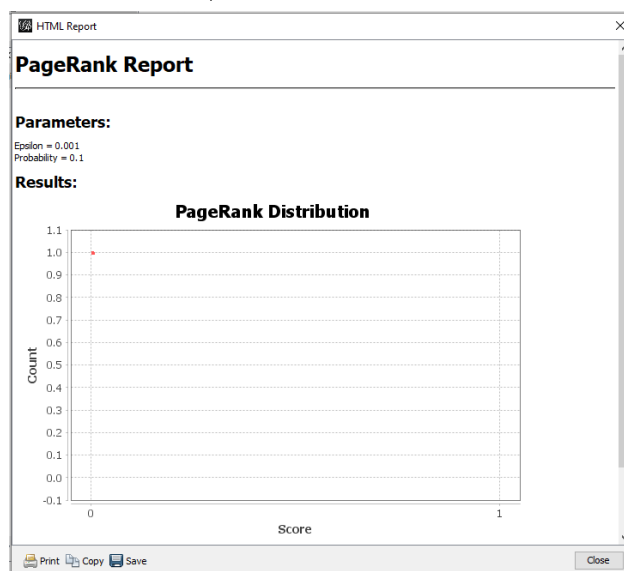
Filters	Statistics ×	—
Settings		
Network Overview		
Average Degree	49.916	Run ?
Avg. Weighted Degree	49.916	Run ?
Network Diameter	3	Run ?
Graph Density	0.1	Run ?
HITS		Run ?
Modularity	0.112	Run ?
PageRank		Run ?
Connected Components	1	Run ?
Node Overview		
Avg. Clustering Coefficient	0.101	Run ?
Eigenvector Centrality		Run ?
Edge Overview		
Avg. Path Length	1.905	Run ?
Dynamic		

OpenOrd

Filters	Statistics ×	—
Settings		
Network Overview		
Average Degree	49.916	Run ?
Avg. Weighted Degree	49.916	Run ?
Network Diameter	3	Run ?
Graph Density	0.1	Run ?
HITS		Run ?
Modularity	0.116	Run ?
PageRank		Run ?
Connected Components	1	Run ?
Node Overview		
Avg. Clustering Coefficient	0.101	Run ?
Eigenvector Centrality		Run ?
Edge Overview		
Avg. Path Length	1.905	Run ?
Dynamic		

Yifan Hu

در ادامه، نتیجه محاسبات PageRank برای گراف a را می بینیم.



۲ تصویر بعدی مربوط به نتایج محاسبات معیارهای شبکه برای گراف شبکه بخش b است. مشابه گراف قبلی نتایج تقریباً یکسان است و تفاوت اندکی صرفاً در بخش Modularity دیده می‌شود.

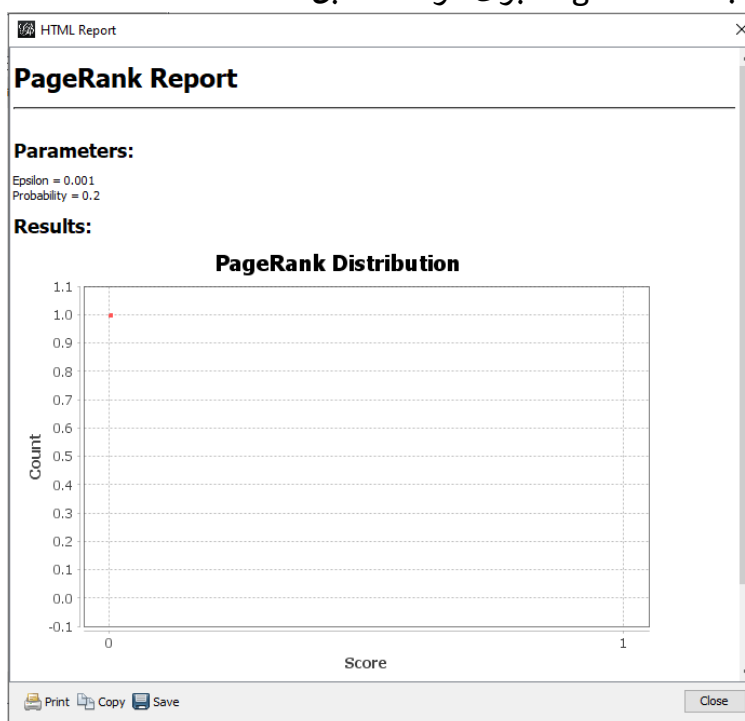
Filters	Statistics	×	—
Settings			
<input checked="" type="checkbox"/> Network Overview			
Average Degree	199.428	Run	?
Avg. Weighted Degree	199.428	Run	?
Network Diameter	2	Run	?
Graph Density	0.2	Run	?
HITS		Run	●
Modularity	0.052	Run	?
PageRank		Run	?
Connected Components	1	Run	?
<input checked="" type="checkbox"/> Node Overview			
Avg. Clustering Coefficient	0.2	Run	?
Eigenvector Centrality		Run	●
<input checked="" type="checkbox"/> Edge Overview			
Avg. Path Length	1.8	Run	?
<input checked="" type="checkbox"/> Dynamic			

openOrd

Filters	Statistics	×	—
Settings			
<input checked="" type="checkbox"/> Network Overview			
Average Degree	199.428	Run	?
Avg. Weighted Degree	199.428	Run	?
Network Diameter	2	Run	?
Graph Density	0.2	Run	?
HITS		Run	●
Modularity	0.054	Run	?
PageRank		Run	?
Connected Components	1	Run	?
<input checked="" type="checkbox"/> Node Overview			
Avg. Clustering Coefficient	0.2	Run	?
Eigenvector Centrality		Run	●
<input checked="" type="checkbox"/> Edge Overview			
Avg. Path Length	1.8	Run	?
<input checked="" type="checkbox"/> Dynamic			

Circular Layout

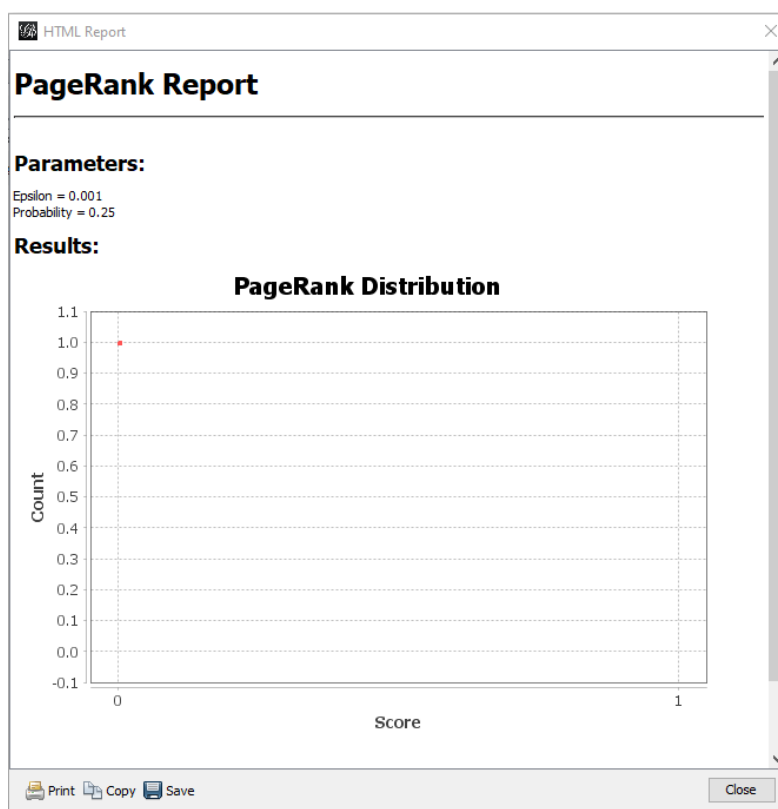
در ادامه، نتیجه محاسبات PageRank برای گراف b قابل مشاهده است.



در تصویر زیر، نتیجه اعمال محاسبات معیارهای شبکه برای گراف بخش c دیده می شود. این محاسبات بدون اعمال Layout خاصی صورت گرفته است.

Filters	Statistics	×	—
Settings			
<input checked="" type="checkbox"/> Network Overview			
Average Degree	499.621	Run	?
Avg. Weighted Degree	499.621	Run	?
Network Diameter	2	Run	?
Graph Density	0.25	Run	?
HITS		Run	?
Modularity	0.031	Run	?
PageRank		Run	?
Connected Components	1	Run	?
<input checked="" type="checkbox"/> Node Overview			
Avg. Clustering Coefficient	0.25	Run	?
Eigenvector Centrality		Run	?
<input checked="" type="checkbox"/> Edge Overview			
Avg. Path Length	1.75	Run	?
<input checked="" type="checkbox"/> Dynamic			

همچنین در ادامه، نتیجه محاسبات PageRank برای گراف c را می بینیم.



تمامی نتایج فوق به صورت خلاصه در جدول زیر گزارش شده است:

شبکه (گراف) معیار شبکه	a	b	c
Average Degree	49.916	199.428	499.621
Average Weighted Degree	49.916	199.428	499.621
Network Diameter	3	2	2
Graph Density	0.100	0.200	0.250
Modularity	0.112	0.052	0.031
Connected component	1	1	1
Avg. clustering coefficient	0.101	0.200	0.25
Avg. path length	1.905	1.8	1.75

پاسخ د)

در گراف تمامی بخش ها داریم $p > \frac{1}{n}$ بنابراین p از حد آستانه عبور کرده و ما در همه گراف ها شاهد یک giant component هستیم که ویژگی های کلی شبکه را در بر دارد. (Connected component) با توجه به نتایج جدول فوق Average Degree با بیشتر شدن احتمال تشکیل یال (p) بالاتر رفته است. رابطه ای که برای پیدا کردن میانگین درجات گراف وجود دارد به صورت زیر است:

$$\langle k \rangle = p (N-1)$$

اگر مقادیر n, p گراف ها را در این رابطه قرار دهیم :

a. $\langle k \rangle = 0.1 (500-1) = 49.9$

b. $\langle k \rangle = 0.2 (1000-1) = 199.8$

c. $\langle k \rangle = 0.25 (2000-1) = 499.75$

از مقایسه نتایج بدست آمده و نتایج موجود در جدول مشاهده می شود که اعداد تقریباً یکسان هستند و اختلاف اندک حاصل احتمالاً مربوط به تصادفی بودن شبکه ها می شود.

به طور مشابه Average Weighted Degree افزایش یافته است. البته لازم به ذکر است که در اینجا گراف ها بدون وزن در نظر گرفته شده اند پس نتایج همان مقادیر قبلی خواهند شد.

متوسط مسیر در گراف (Avg. path length) با افزایش p کاهش پیدا کرده. فرمول محاسبه ی این معیار بر حسب n, p مطابق زیر است:

$$\text{Avg. path length} = (\log N) / \log \langle k \rangle$$

مطابق این فرمول با افزایش احتمال وجود یال p مقدار $\langle k \rangle$ افزوده می شود و چون این مقدار در مخرج کسر قرار دارد بنابراین میزان متوسط مسیر در گراف کاهش پیدا می کند که در نتایج موجود در جدول هم همین نشان داده شد. اگر مقادیر n, p گراف ها را در این رابطه قرار دهیم:

a. $\text{Avg. path length} = \log 500 / \log 49.9 = 1.58$

b. $\text{Avg. path length} = \log 1000 / \log 199.8 = 1.30$

c. $\text{Avg. path length} = \log 2000 / \log 499.75 = 1.22$

که این مقادیر با تقریبی مشابه نتایج به دست آمده در نرم افزار Gephi هستند. Network Diameter همان طولانی ترین shortest path length در گراف است که محاسبات آن مشابه Avg. path length می باشد.

برای محاسبه متوسط Clustering coefficient در شبکه های تصادفی در این ناحیه، از رابطه زیر استفاده می کنیم:

$$\text{Clustering coefficient} = \langle k \rangle / N$$

با قرار دادن مقادیر n, p گراف ها را در این رابطه داریم:

a. $\text{Clustering coefficient} = 49.9 / 500 = 0.098$

b. $\text{Clustering coefficient} = 199.8 / 1000 = 0.199$

c. $\text{Clustering coefficient} = 499.75 / 2000 = 0.249$

با مقایسه در می یابیم که این مقادیر و مقادیر موجود در جدول تقریباً نزدیک بهم هستند. چگالی گراف Graph Density برابر رابطه ی زیر است:

$$\text{Graph Density} = \frac{\text{number of edges}}{\text{total number of possible edges.}}$$

که این رابطه همان نشانگر احتمال وجود یال (p) هم می تواند در نظر گرفته شود که در صورت مساله ذکر شده است و در نتایج حاصل از نرم افزار هم همین بدست آمد.

منابع:

- <https://www.geeksforgeeks.org/saving-a-networkx-graph-in-gexf-format-and-visualize-using-gephi/>
- <https://math.stackexchange.com/questions/1526372/what-is-the-definition-of-the-density-of-a-graph/1526421#:~:text=It's%20right%20that%20graph%20density,total%20no%20of%20possible%20edges.>
- <https://towardsdatascience.com/a-quick-tutorial-on-gephi-layouts-daa87fec5a20>
- <https://gephi.org/tutorials/gephi-tutorial-layouts.pdf>