

# کاوش شبکه های مجازی اینستاگرام و توییتر به همراه تحلیل آنها

استاد : دکتر قائمی بافقی

دانشجو : علی عادلخواه

# روش های کاوش فضای مجازی

- استفاده از API توییتر از آدرس  
<https://developer.twitter.com/en/apply/user>
- استفاده از نرم افزار های تجاری کاوش و تحلیل فضای  
مجازی مانند NodeXL
- کاوش با استفاده از مرورگر شخصی و کتابخانه های وب  
گردی مانند selenium

# کاربرد های تحلیل فضای مجازی

- اقتصادی

1. Market Research

2. Customer Service

3. Competitive Intelligence

- سیاسی

1. Political Communication

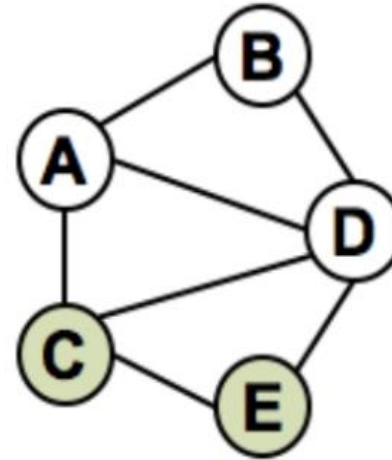
2. Election Forecasting

3. Political Risk Management

# تطبيق شبكة ارتباطات به صورت گراف

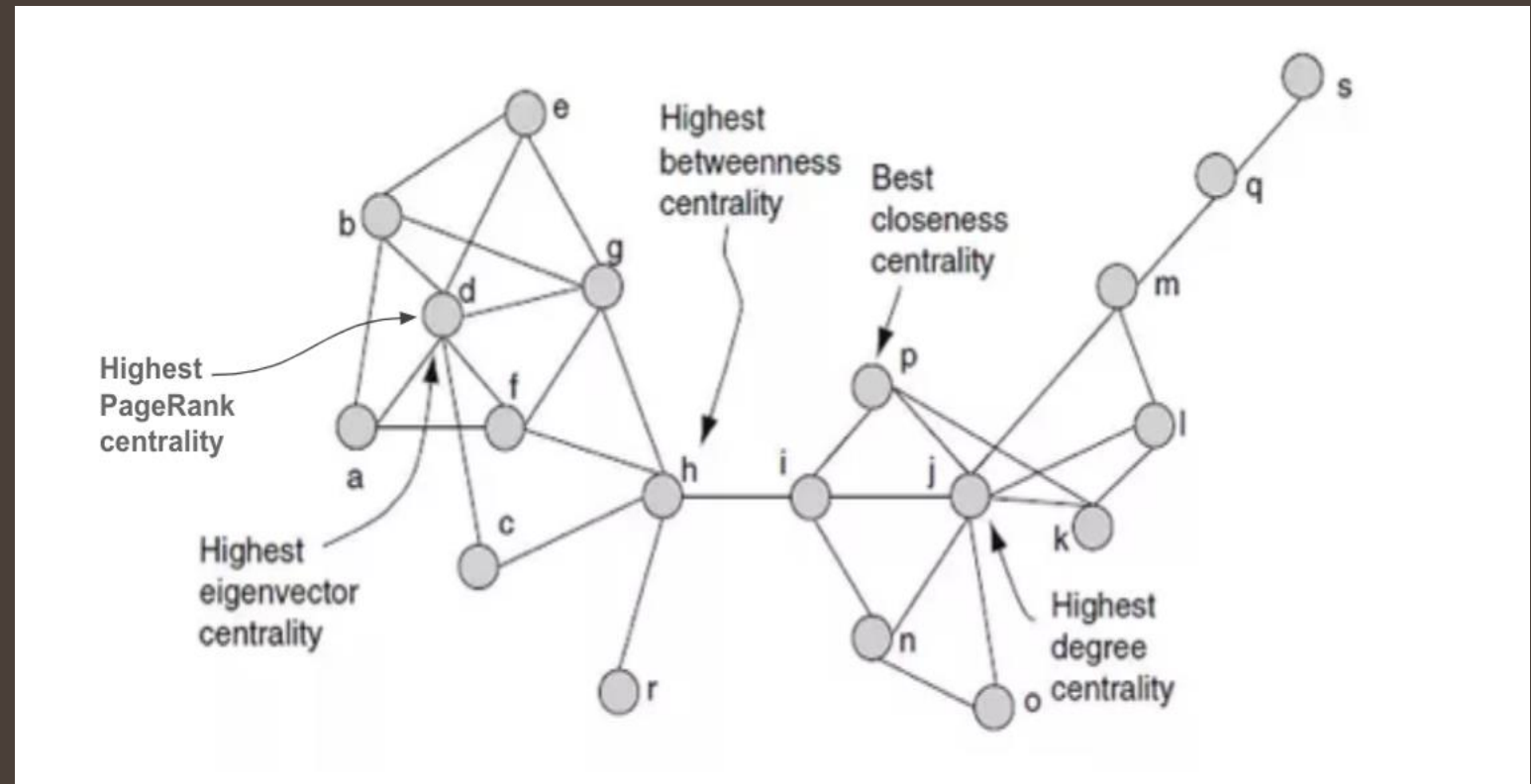


Nodes



Edges

# معیار مرکزیت



$$g(v) = \sum_{s \neq v \neq t} \frac{\sigma_{st}(v)}{\sigma_{st}}$$

- betweenness : بر اساس تعداد راه هایی که بین نود ها از این راس عبور میکنند.

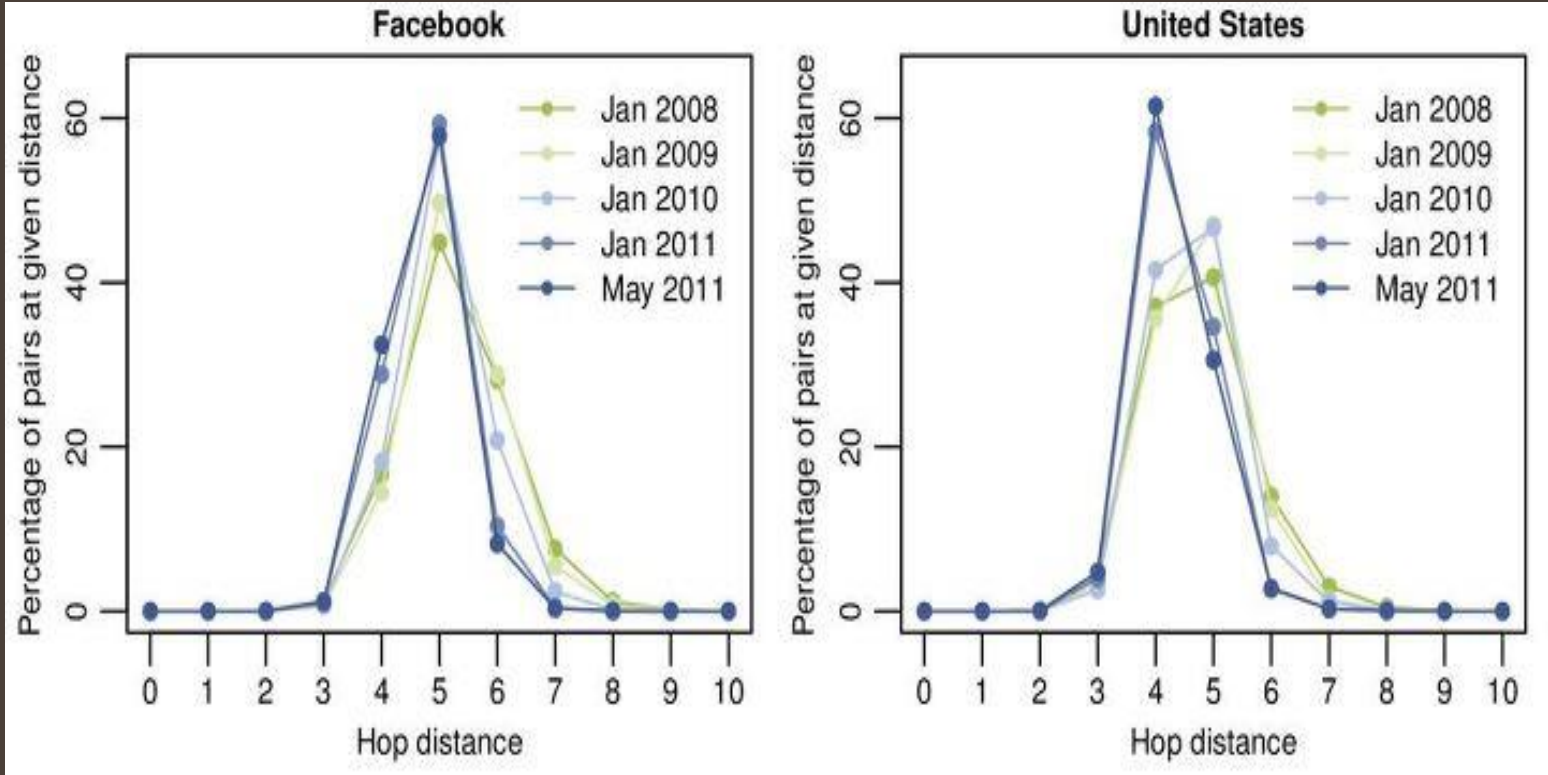
- degree : بر اساس تعداد همسایه های يك راس.

- closeness : بر اساس کمترین فاصله ای که دور ترین راس دارد.

$$PR(u) = \sum_{v \in B_u} \frac{PR(v)}{L(v)},$$

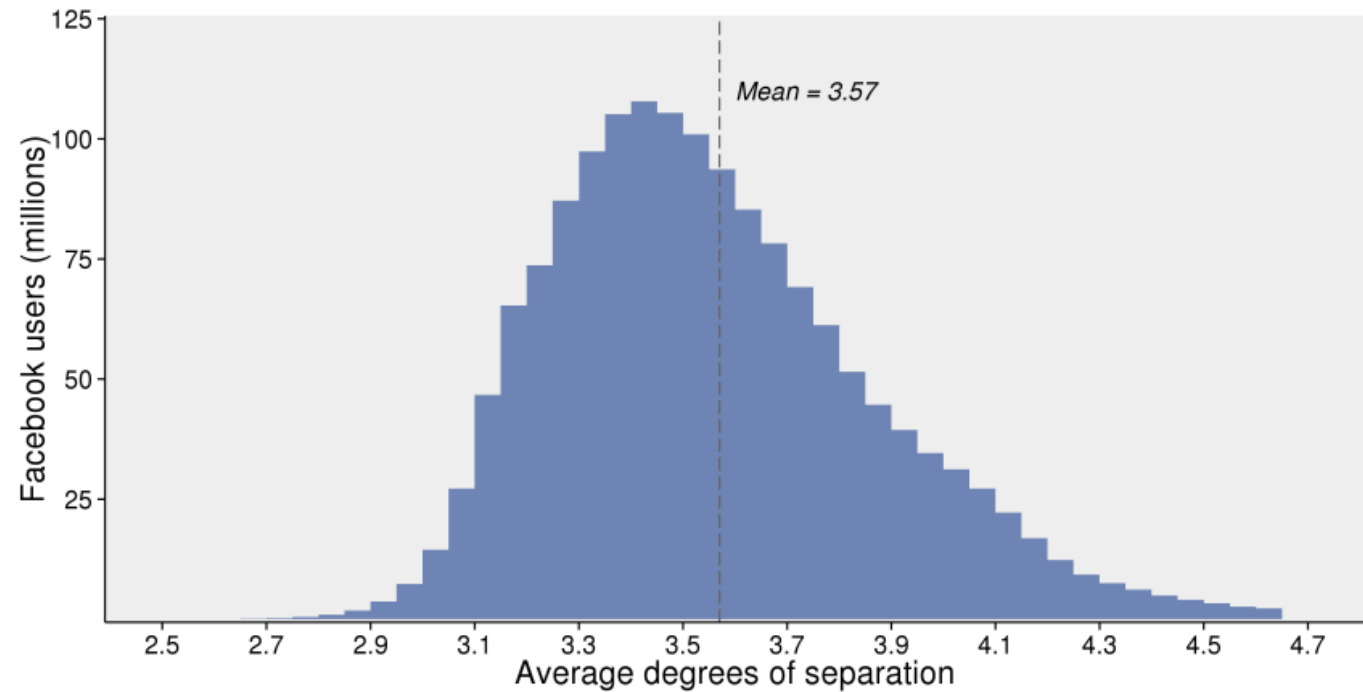
- pagerank : يك روش بازگشتي است که pagerank يك نود از pagerank همسایه های آن بدست مي آید.

تحقیق فیسبوک بر روی 721 میلیون کاربر در سال 2012  
Average Distance = 4.74



# تحقیق فیسبوک بر روی 1.6 میلیارد کاربر در سال 2016

Average Distance = 3.57





## Modularity

$$\sum_u \left( \frac{m_u}{m} - \left( \frac{k_u}{2m} \right)^2 \right)$$

خوشه بندی گراف (NP Hard)

Algorithm	$\xi$	Complexity
Label propagation	N	$O(V)$
Louvain modularity	D	$O(V \log V)$
Infomap	N	$O(V \log V)$
Greedy opt. of modularity	D	$O(V \log^2(V))$
Leading eigenvector	D	$O(V^2 \log V)$
Walktrap	N	$O(V^2 \log V)$
Edge betweenness	D	$O(V^3)$

# مقایسه الگوریتم های خوشه بندی بر حسب Modularity

Table 1. The comparison of modularity from different algorithms on four datasets

Dataset	EB	FG	Infomap	LP	LE	Louvain	Walktrap
2019 AAG related to GIS	0.8475	0.8066	0.8052	0.7626	0.8091	0.8499	0.8183
1999 AAG	0.6702	0.6666	0.6194	0.6279	0.5842	0.6859	0.5940
2009 AAG	0.5140	0.5122	0.4692	0.4389	0.4381	0.5447	0.4430
2015-2019 AAG	-	0.4127	0.3377	0.0581	0.3045	0.4239	0.3320

# مقایسه الگوریتم های خوشه بندی بر حسب Time

Table 2. The comparison of processing time (in seconds) from different algorithms on the four datasets

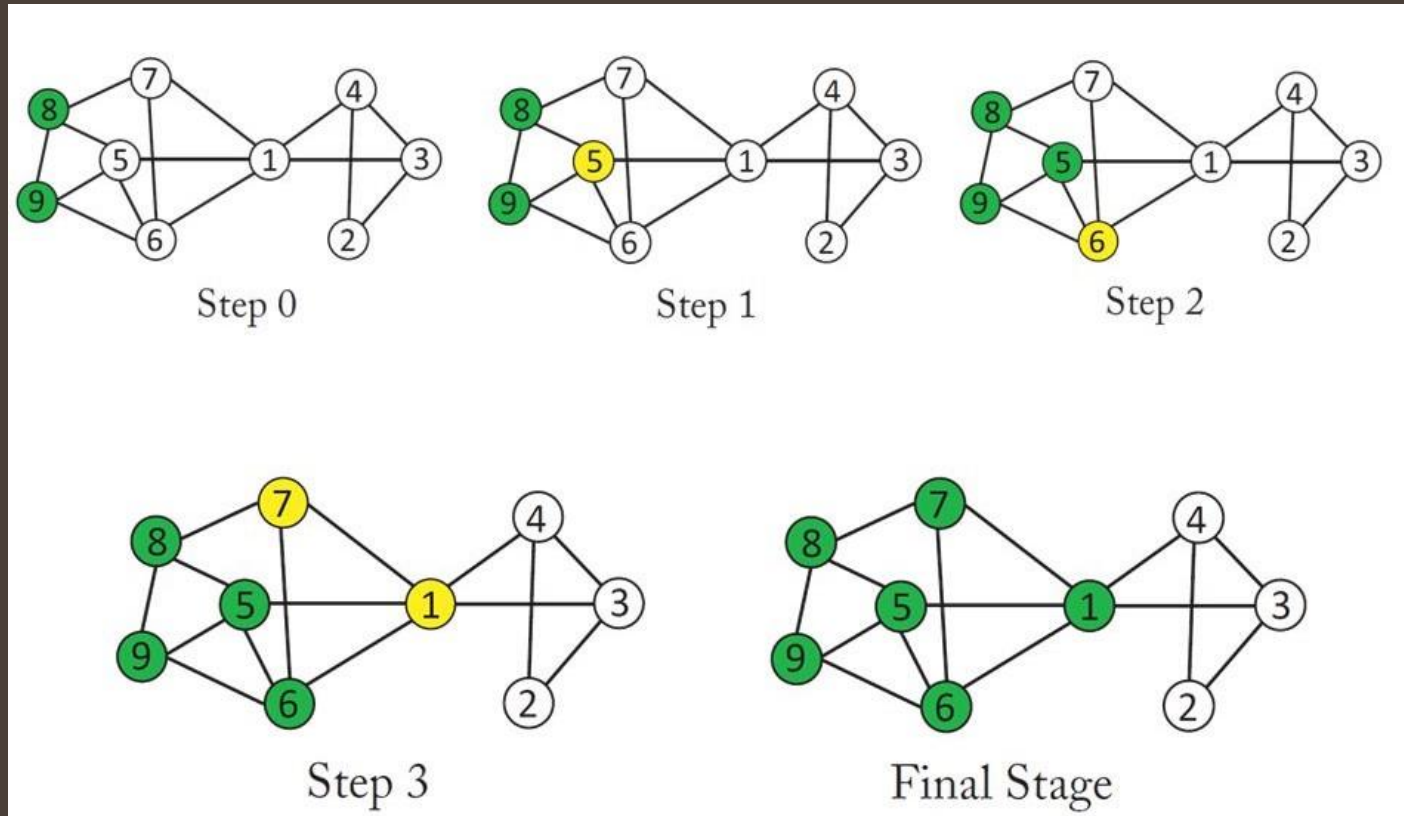
Dataset	EB	FG	Infomap	LP	LE	Louvain	Walktrap
2019 AAG related to GIS	39	0.005	0.313	0.004	0.215	0.006	0.043
1999 AAG	2437	0.042	2.421	0.049	0.938	0.021	0.283
2009 AAG	181326	0.742	12.751	0.078	2.713	0.097	3.186
2015-2019 AAG	$\infty$	29.609	137.501	0.424	8.523	0.565	119.938

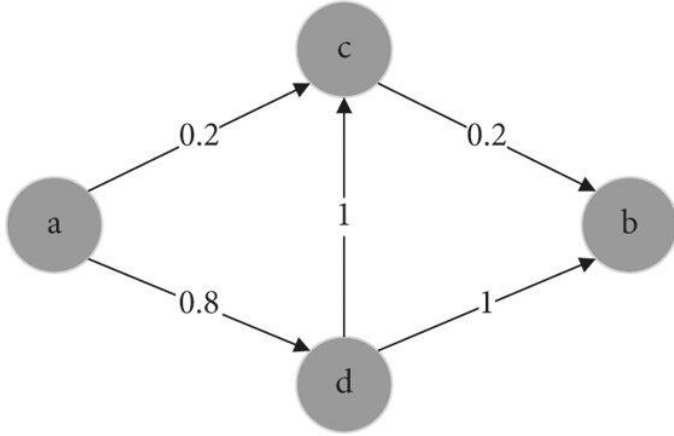


# Influence Maximization (NP Hard)

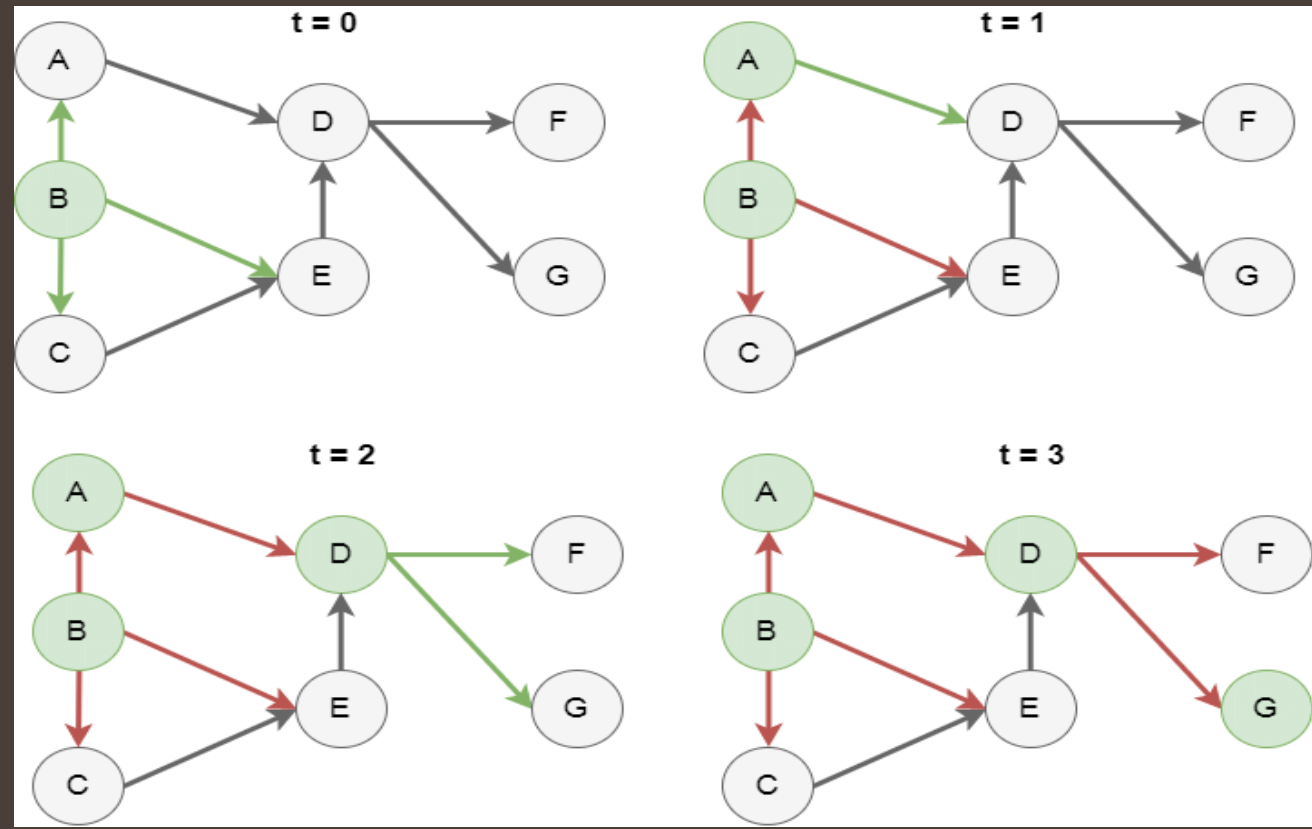


# Linear Threshold

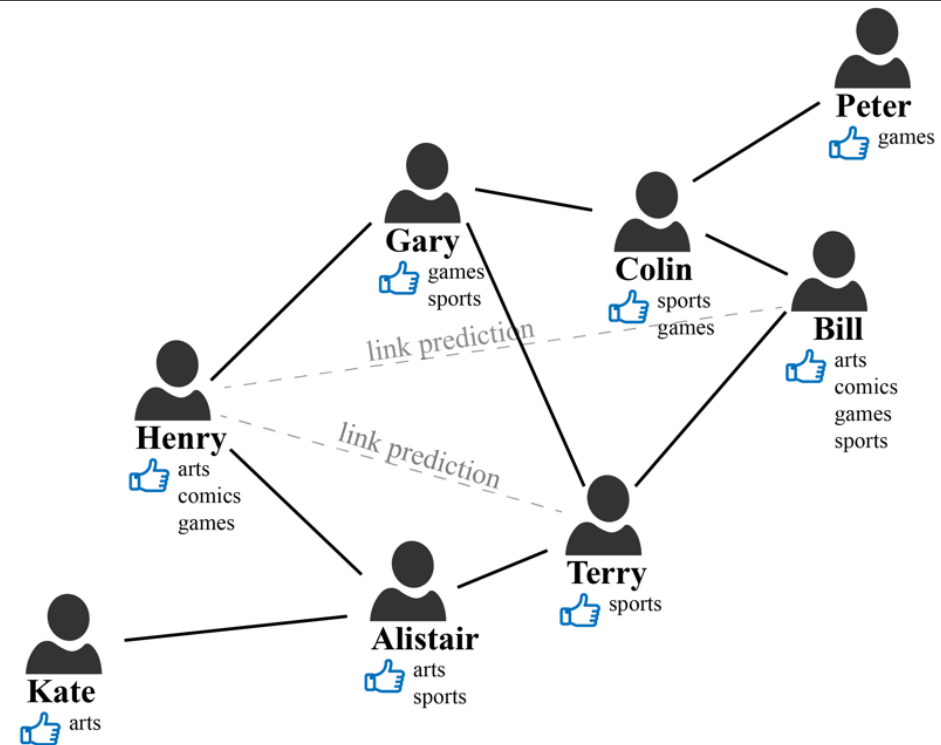




# Independent Cascade



# Link Prediction

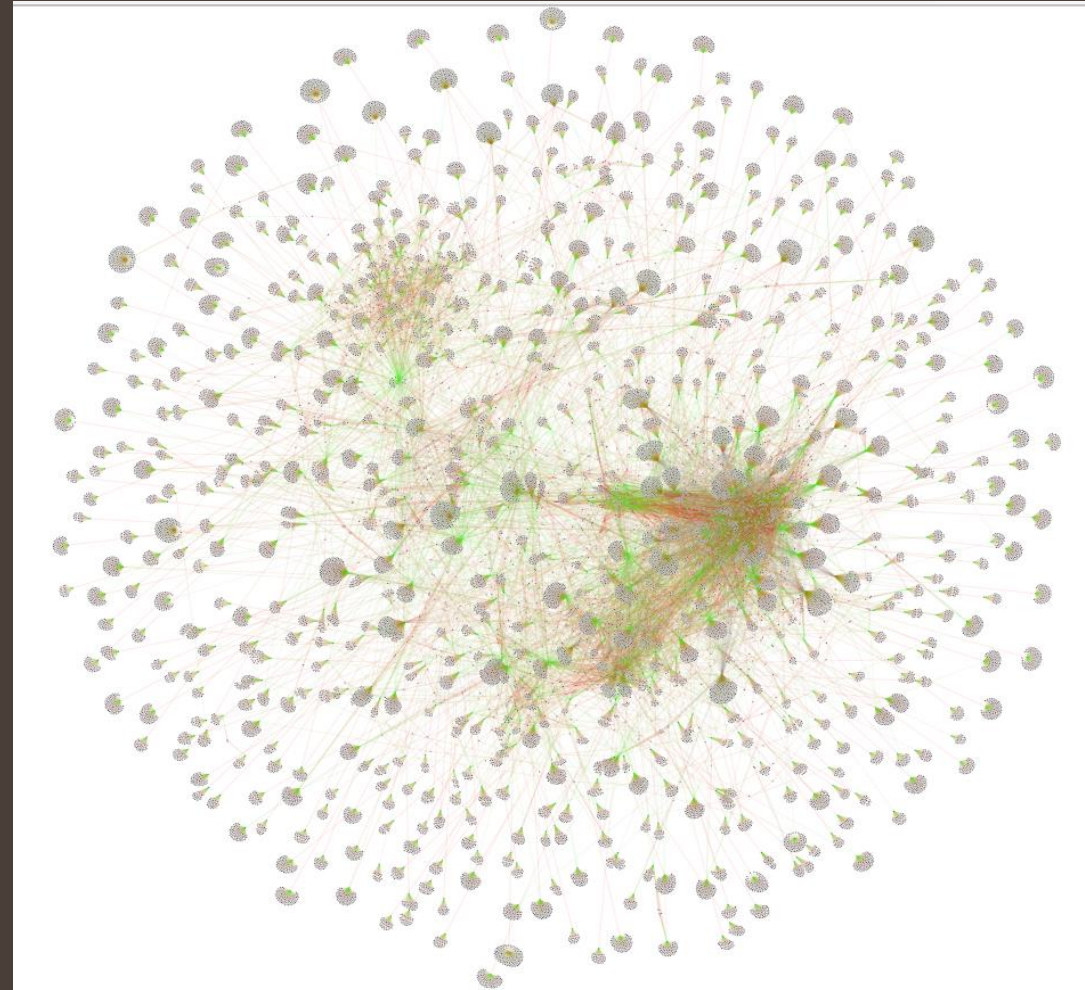
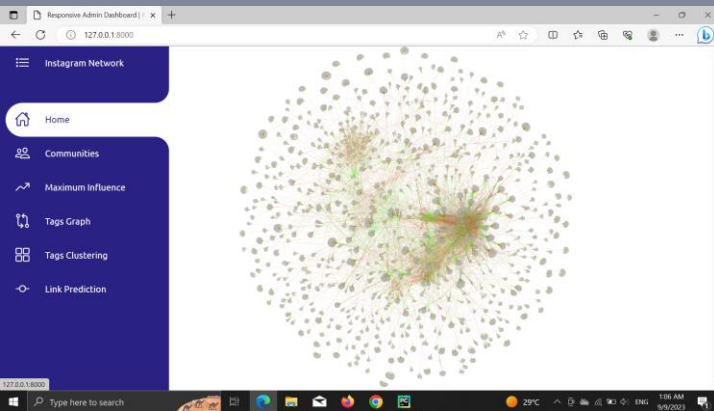


# پیاده سازی

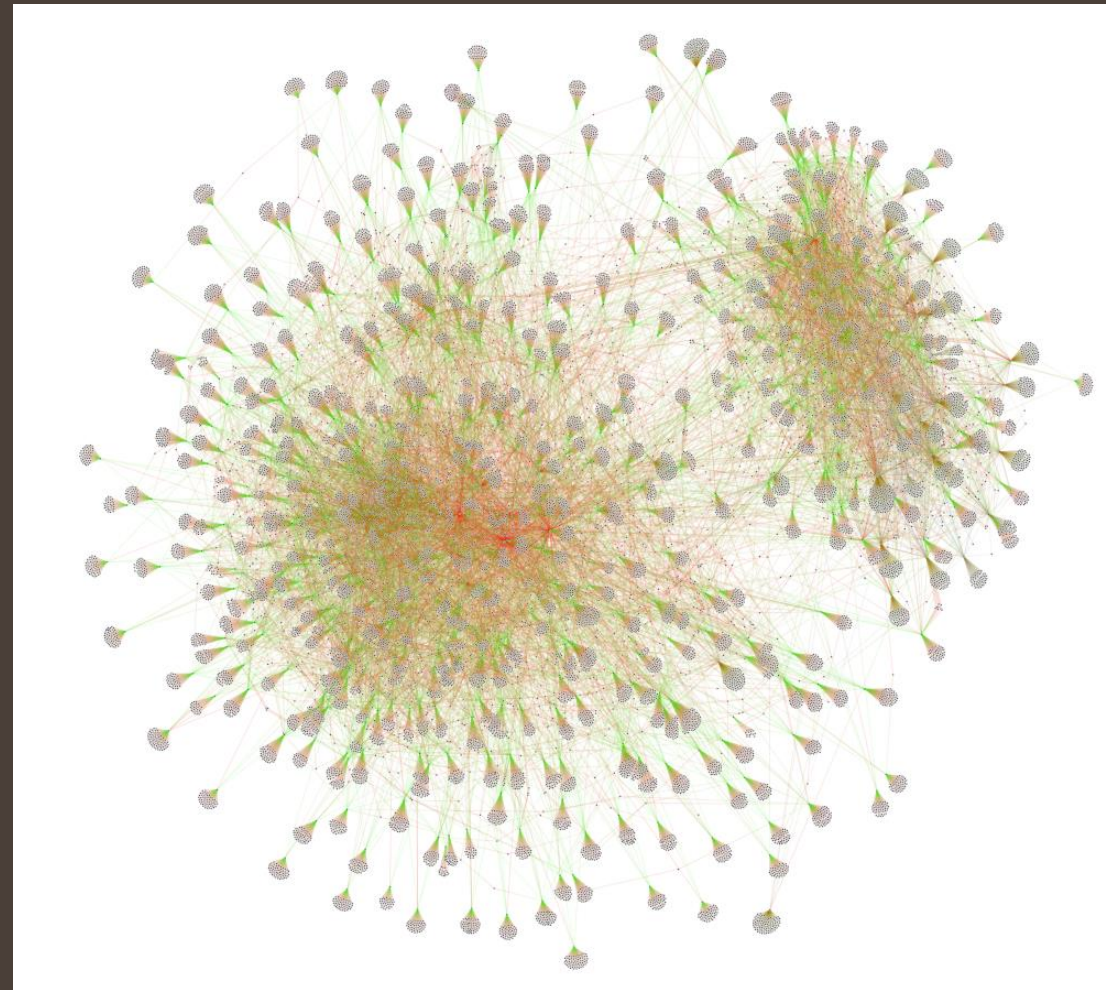
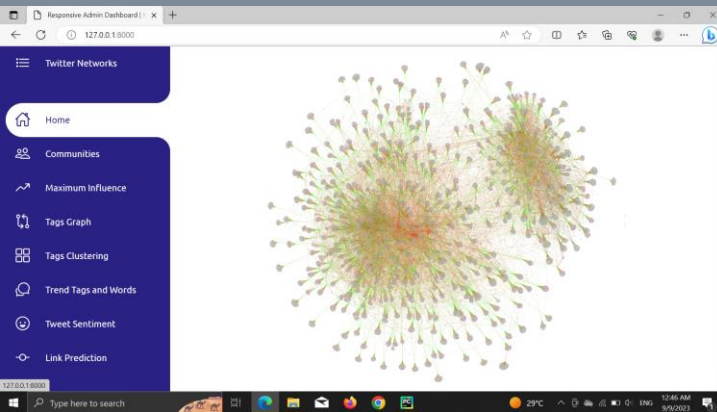
- کاوش 22 هزار کاربر توییتر
- کاوش 29 هزار کاربر اینستاگرام
- نمایش گراف ارتباطات با جاوا اسکریپت و کتابخانه d3
- خوشه بندی گراف با الگوریتم Louvain
- مقایسه معیار های مرکزیت در مسئله Influence Maximization
- مقایسه دو الگوریتم celf و greedy در مسئله Influence Maximization
- نمایش گراف تگ ها
- کلمات پر تکرار در کنار تگ ها و دسته بندی احساسی توییت ها برای هر تگ با استفاده از هوش مصنوعی
- پیشبینی پیوند های جدید با استفاده از هوش مصنوعی



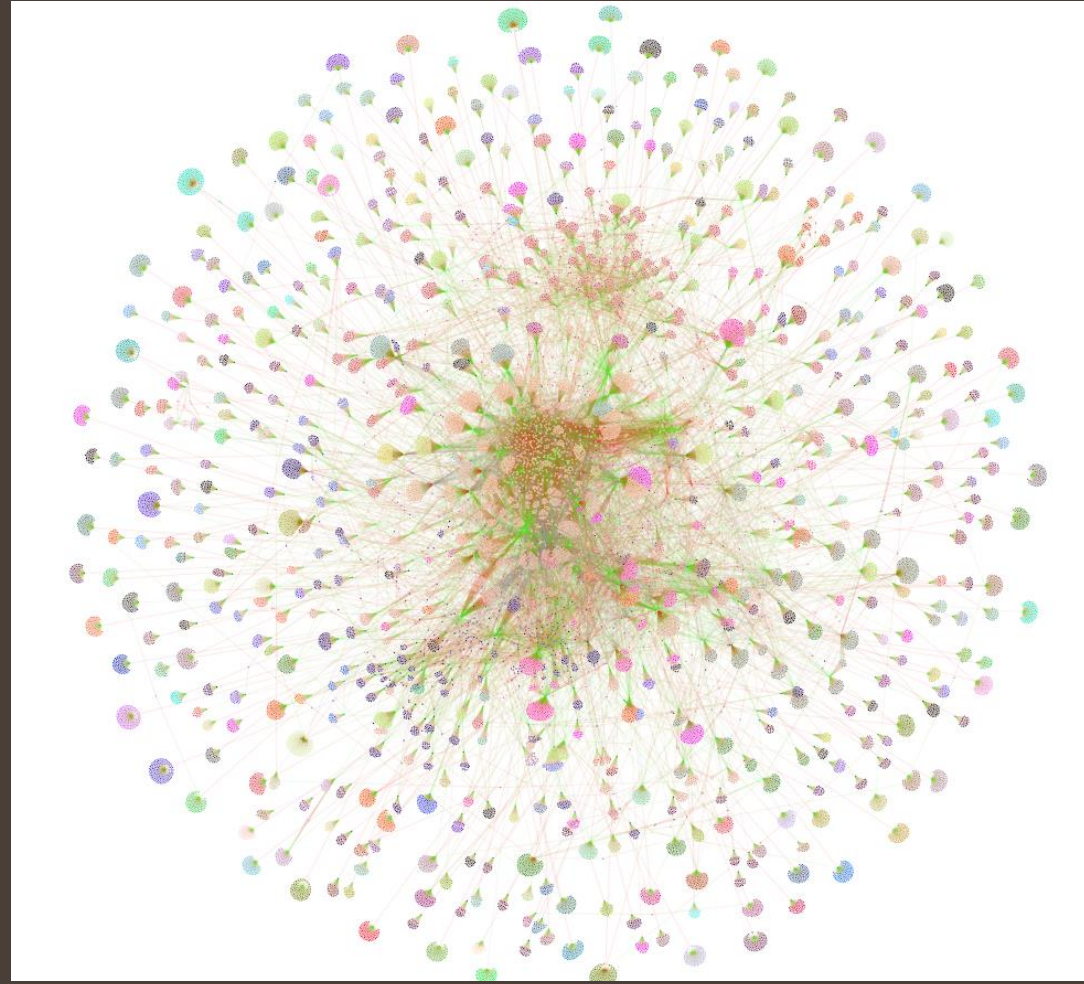
# گراف 29 هزار کاربر اینستاگرام



# گراف 22 هزار کاربر توییتر

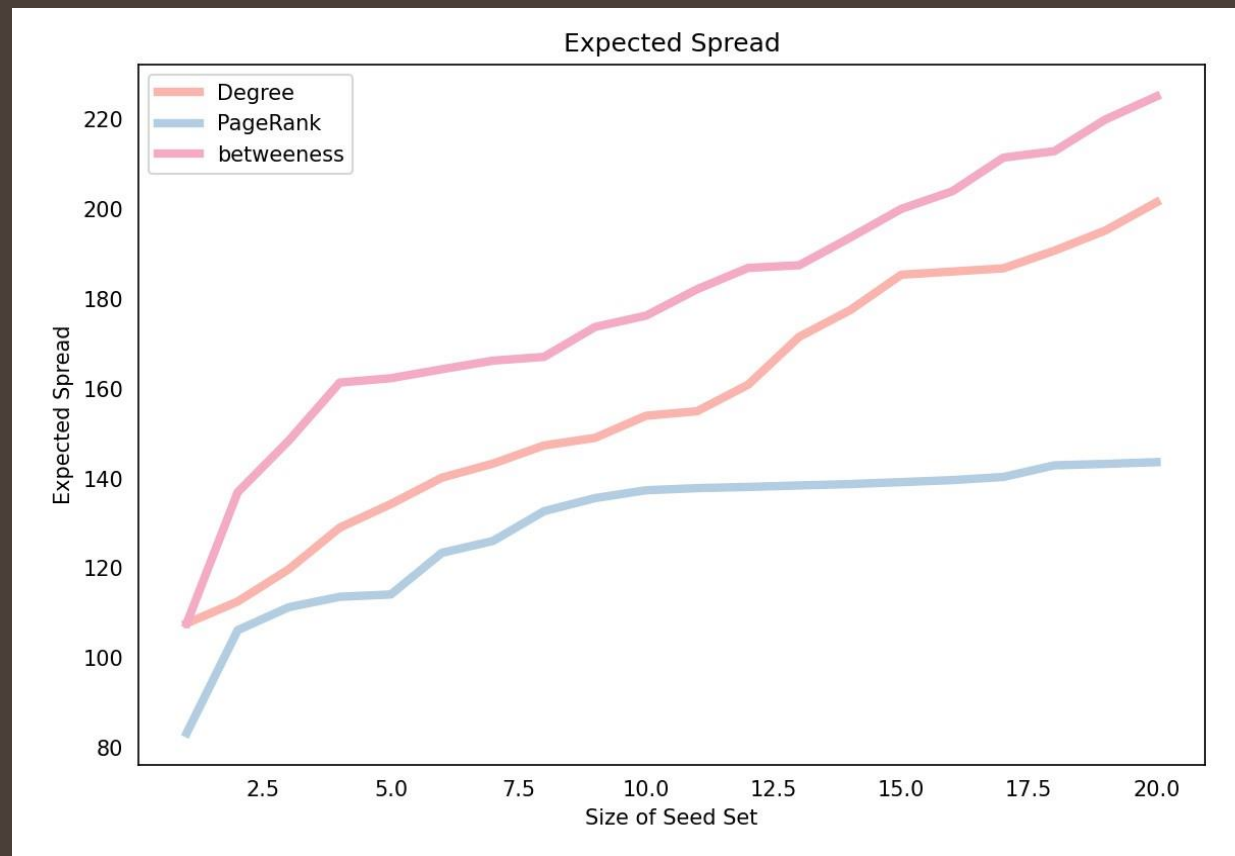


# خوشه بندی گراف

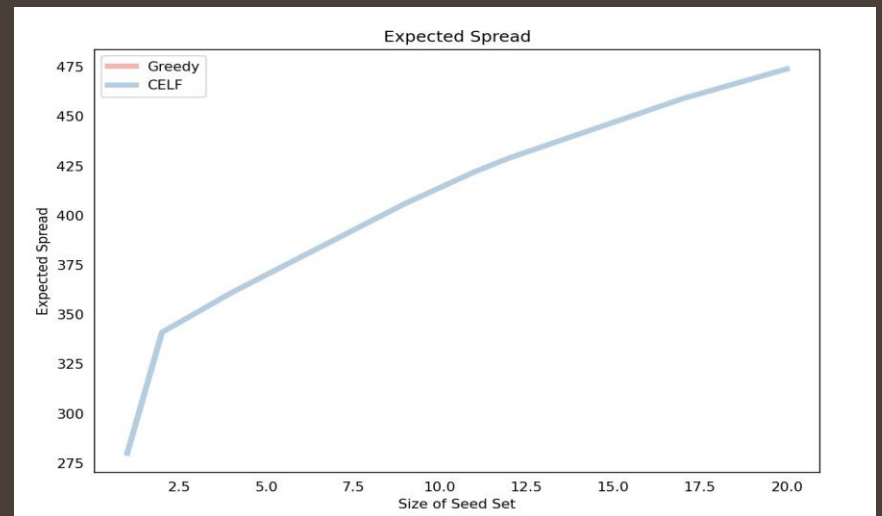
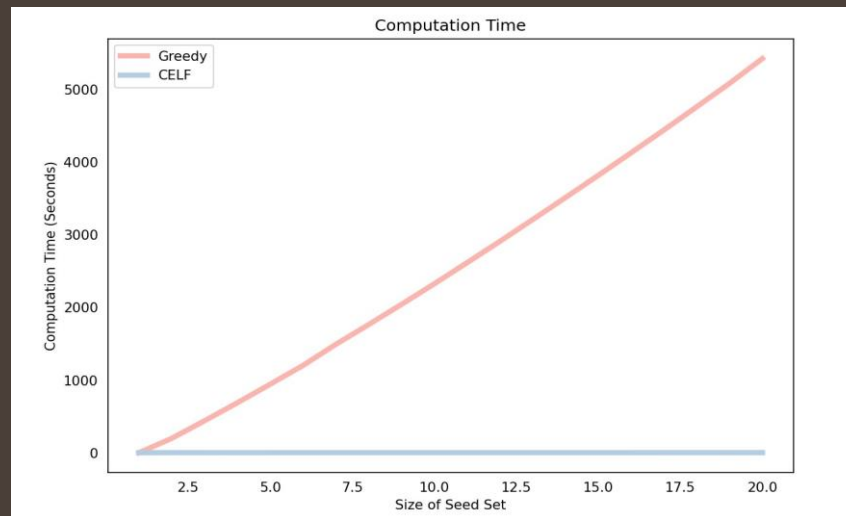




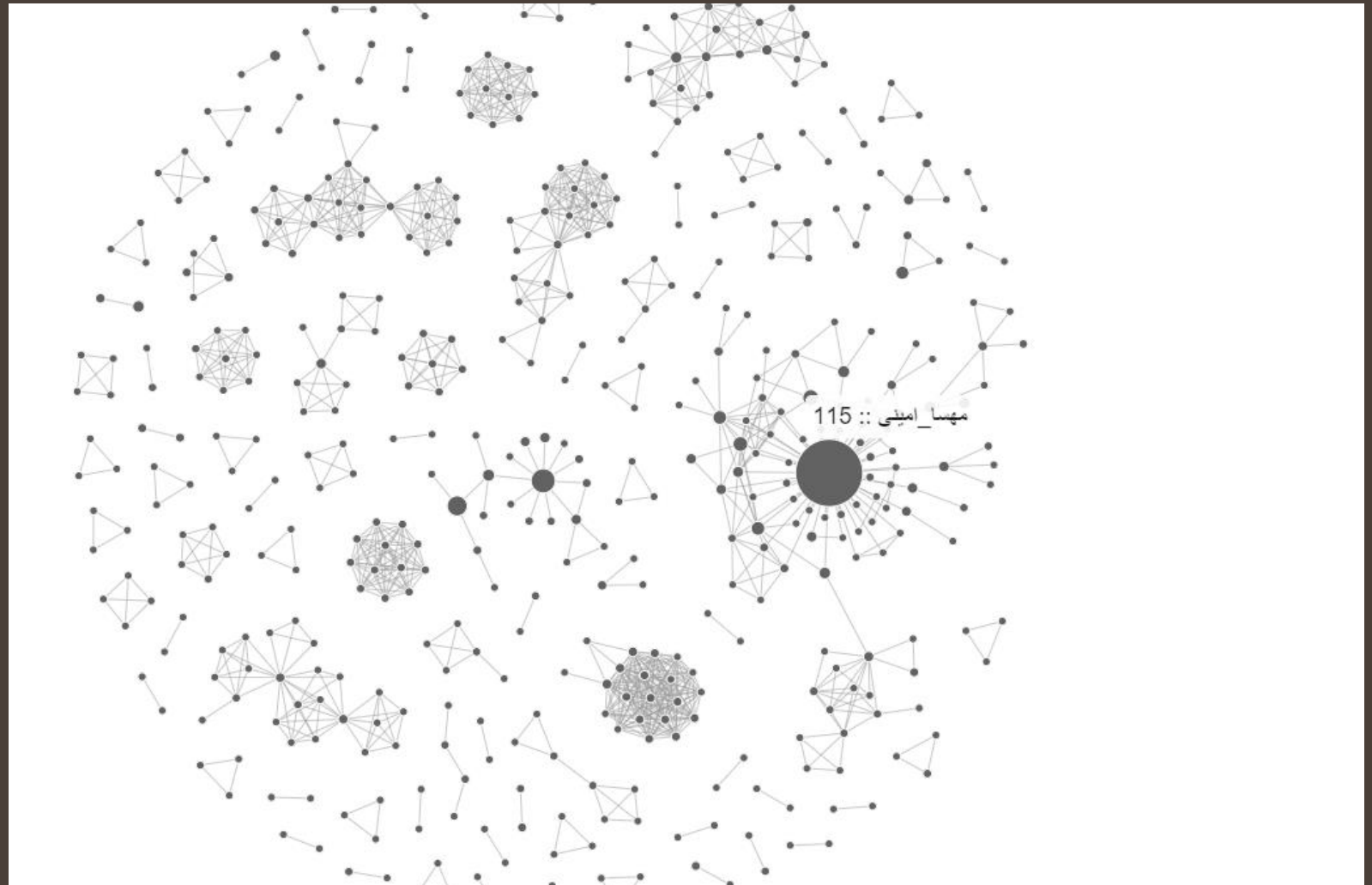
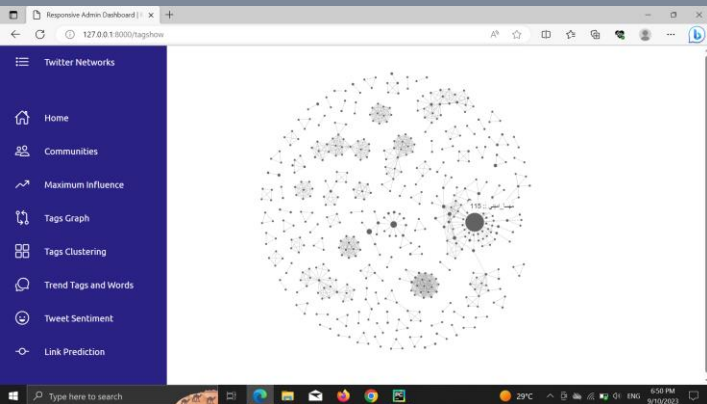
## مقایسه 3 معیار مرکزیت بر روی گراف توپتر



# مقایسه دو الگوریتم greedy و celf



# گراف تگ ها در توییتر



# کلمات پر تکرار در کنار تگ ها

Responsive Admin Dashboard | x +

127.0.0.1:8000/showtagwords

Twitter Networks

Home

Communities

Maximum Influence

Tags Graph

Tags Clustering

Trend Tags and Words

Tweet Sentiment

Link Prediction

## Trend Tags

Tag	N	words
مهسا_امینی	115	مهسا_امینی# (53) برای (6) جمهوری (5) اسلامی (5) کردن
مهسا_امینی	115	جمهوری اسلامی (5) را به (4) فقط با (2) اون روز (2) ثابت کنه (2)
حب_الحسین_یجمعنا	32	حب_الحسین_یجمعنا# (9) کربلا (3) دارم (2) میگم (1) بوسه (1)
حب_الحسین_یجمعنا	32	تو بگو (1) بگو عشق (1) عشق +من (1) من میگم+ (1) میگویم بوسه (1)
Arbaeen2023	24	#Arbaeen2023 (7) اربعین (4) تمام (4) امام (4) ظهور (3)
Arbaeen2023	24	بعد از (2) کتیبه بالای (1) بالای ورودی (1) ورودی باب (1) باب القبله (1)
مهدی_یراحی	16	مهدی_یراحی# (9) صدای (5) برای (3) مهدی (3) یراحی (3)
مهدی_یراحی	16	مهدی یراحی (3) کاپوس خامنه‌ای (2) خامنه‌ای رو (2) رو دستگیر (2) دستگیر (2) کردن (2)
توماج_صالحی	14	توماج_صالحی# (11) #ToomajSalehi (4) life (4) trial (4) artist (3)
توماج_صالحی	14	life is (4) is in (4) He is (3) His life (3) in dange (3)
MahsaAmini	12	#MahsaAmini (6) توماج_صالحی# (2) trial (2) with (2) against (2)
MahsaAmini	12	His life (2) life is (2) is in (2) توماج_صالحی# (2) Today, (1) dissiden (1)

127.0.0.1:8000/showtagwords

Type here to search

29°C

5:55 PM 9/10/2023

# • دسته بندی احساسی توییت ها برای هر تگ با استفاده از هوش مصنوعی

Responsive Admin Dashboard | K x +

127.0.0.1:8000/showtagsentiment

Twitter Networks

Home

Communities

Maximum Influence

Tags Graph

Tags Clustering

Trend Tags and Words

**Tweet Sentiment**

Link Prediction

### Tweets Sentiment Per Tags

[View All](#)

Tag	N	joy	sad	hate	love	angry	fear	none
مهسا_امینی	115	7	41	22	28	6	10	1
حب_الحسن_یجمعنا	32	0	11	5	10	0	1	5
Arbaeen2023	24	0	6	4	10	0	1	3
مهدی_پراچی	16	1	7	4	3	1	0	0
توماج_صالحی	14	1	3	0	7	0	0	2
MahsaAmini	12	1	2	2	6	1	0	0
اعتصاهاات_سراسری	12	0	0	3	6	0	2	1
ریتوییت	11	0	4	3	4	0	0	0
جواد_روحی	9	1	3	3	2	0	0	0
اربعین	9	1	3	3	2	0	0	0
سالگرد_مهسا	8	0	3	2	3	0	0	0
نیکا_شاکرمی	8	0	6	0	1	0	1	0

29°C 4:45 PM 9/11/2023



# • پیشبینی پیوند ها جدید با هوش مصنوعی

Responsive Admin Dashboard | K x +

127.0.0.1:8000/linkpredict

Twitter Networks

- Home
- Communities
- Maximum Influence
- Tags Graph
- Tags Clustering
- Trend Tags and Words
- Tweet Sentiment
- Link Prediction

### Users Similarity [View All](#)

User1	User2	Common Neighbors	Average Tweets Similarity
arashsm79	hammeredrq7	28	0.6833159150304021
arashsm79	SGMCX1	28	0.6963452439186937
hammeredrq7	SGMCX1	28	0.689046027525416
RoozbeehSa	SGMCX1	26	0.6762859159045749
MmdrezaR	SGMCX1	23	0.6762859159045749
SGMCX1	andromedaROCL	23	0.6847394898533821
MmdrezaR	RoozbeehSa	22	0.5618375301361084
SGMCX1	AmirrezaFYL	22	0.6521328474794116
arashsm79	Parsa_etm_1601	21	0.6653339172664442
YasaminZarei	SGMCX1	21	0.6482740732339712
nemigholam	SGMCX1	21	0.6521328474794116
arashsm79	dokhtar_amo	20	0.6772148800747735
hammeredrq7	MmdrezaR	20	0.6242639223734537

127.0.0.1:8000/linkpredict

Type here to search

29°C 1:02 AM 9/9/2023

## References

- <https://d3js.org/>.
- <https://pypi.org/project/hazm/>.
- <https://pypi.org/project/imgbeddings/>.
- <https://pypi.org/project/langdetect/>.
- <https://pypi.org/project/selenium/>.
- <https://pypi.org/project/sent2vec/>.
- Lars Backstrom, Paolo Boldi, Marco Rosa, Johan Ugander, and Sebastiano Vigna. Four degrees of separation. In Proceedings of the 4th Annual ACM Web Science Conference, pages 33–42, 2012.
- Sergey Edunov, Carlos Diuk, Ismail Onur Filiz, Smriti Bhagat, and Moira Burke. Three and a half degrees of separation. Research at Facebook, 694, 2016.
- [g] Dmitri Goldenberg. Social network analysis: From graph theory to applications with python. arXiv preprint arXiv:2102.10014, 2021.

- Petros Iosifidis, Mark Wheeler, Petros Iosifidis, and Mark Wheeler. The political economy of social media. *Public Spheres and Mediated Social Networks in the Western Context and Beyond*, pages 39–64, 2016.
- David Kempe, Jon Kleinberg, and Éva Tardos. Maximizing the spread of influence through a social network. In *Proceedings of the ninth ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 137–146, 2003.
- Youngho Lee, Yubin Lee, Jeong Seong, Ana Stanescu, and Chul Sue Hwang. A comparison of network clustering algorithms in keyword network analysis: A case study with geography conference presentations. *International Journal of Geospatial and Environmental Research*, 7(3):1, 2020.
- Jure Leskovec and Eric Horvitz. Planetary-scale views on a large instant-messaging network. In *Proceedings of the 17th international conference on World Wide Web*, pages 915–924, 2008.
- Jure Leskovec, Andreas Krause, Carlos Guestrin, Christos Faloutsos, Jeanne VanBriesen, and Natalie Glance. Cost-effective outbreak detection in networks. In *Proceedings of the 13th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 420–429, 2007.