Graduate Project (Network Analysis CSC 607, Fall 2022)
Venkata Vijay Krishna Gabbula

- Implemented with reviews greater than 130.
- Merged the reviews and business datasets (extracted from the Yelp Dataset)
- **(i).** Sample data of the Reviews dataset after removing the unwanted columns

	review_id	user_id	business_id
0	KU_O5udG6zpxOg-VcAEodg	mh_-eMZ6K5RLWhZyISBhwA	XQfwVwDr-v0ZS3_CbbE5Xw
1	BiTunyQ73aT9WBnpR9DZGw	OyoGAe7OKpv6SyGZT5g77Q	7ATYjTIgM3jUlt4UM3IypQ
2	saUsX_uimxRlCVr67Z4Jig	8g_iMtfSiwikVnbP2etR0A	YjUWPpI6HXG530lwP-fb2A
3	AqPFMleE6RsU23_auESxiA	_7bHUi9Uuf5__HHc_Q8guQ	kxX2SOes4o-D3ZQBkiMRfA
4	Sx8TMOWLNuJBWer-0pcmoA	bcjbaE6dDog4jkNY91ncLQ	e4Vwtrqf-wpJfwesgvdgxQ
...	...	...	...
6990275	H0RIamZu0B0Ei0P4aeh3sQ	qskILQ3k0I_qcCMI-k6_QQ	jals67o91gcrD4DC81Vk6w
6990276	shTPgbgdwTHSuU67mGCmZQ	Zo0th2m8Ez4gLSbHftiQvg	2vLksaMmSEcGbjI5gywpZA
6990277	YNfNhgZlaaCO5Q_YJR4rEw	mm6E4FbCMwJmb7kPDZ5v2Q	R1khUUxidqfaJmcpmGd4aw
6990278	i-I4ZOhoX70Nw5H0FwrQUA	YwAMC-jvZ1fvEUum6QkEkw	Rr9kKArrMhSLVE9a53q-aA
6990279	RwcKOdEuLRHNJe4M9-qpqg	6JehEvdoCvZPJ_XlxnzIIw	VAeEXLbEcI9Emt9KGYq9aA

6990280 rows × 3 columns

- Sample data from the business dataset after removing unwanted columns

	business_id	name	address	city	state	postal_code	latitude	longitude	stars	review_count	is_open
0	Pns2l4eNsfO8kk83dixA6A	Abby Rappoport, LAC, CMQ	1616 Chapala St, Ste 2	Santa Barbara	CA	93101	34.426679	-119.711197	5.0	7	0
1	mpf3x-BjTdTEA3yCZrAYPw	The UPS Store	87 Grasso Plaza Shopping Center	Affton	MO	63123	38.551126	-90.335695	3.0	15	1
2	tUFrWirKiKi_TAnsVWINQQ	Target	5255 E Broadway Blvd	Tucson	AZ	85711	32.223236	-110.880452	3.5	22	0
3	MTSW4McQd7CbVtyjqoe9mw	St Honore Pastries	935 Race St	Philadelphia	PA	19107	39.955505	-75.155564	4.0	80	1
4	mWMc6_wTdE0EUBKIGXDVfA	Perkiomen Valley Brewery	101 Walnut St	Green Lane	PA	18054	40.338183	-75.471659	4.5	13	1
...	...	...	...	...	...	...	...	...	...	...	...
150341	lUQopTMmYQG-qRtBk-8QnA	Binh's Nails	3388 Gateway Blvd	Edmonton	AB	T6J 5H2	53.468419	-113.492054	3.0	13	1
150342	c8GjPlOTGVmlemT7j5_SyQ	Wild Birds Unlimited	2813 Bransford Ave	Nashville	TN	37204	36.115118	-86.766925	4.0	5	1
150343	_QAMST-NrQobXduilWEqSw	Claire's Boutique	6020 E 82nd St, Ste 46	Indianapolis	IN	46250	39.908707	-86.065088	3.5	8	1
150344	mtGm22y5c2UHNXDFAjaPNw	Cyclery & Fitness Center	2472 Troy Rd	Edwardsville	IL	62025	38.782351	-89.950558	4.0	24	1
150345	jV_XOycEzSlTx-65W906pg	Sic Ink	238 Apollo Beach Blvd	Apollo beach	FL	33572	27.771002	-82.394910	4.5	9	1

150346 rows × 14 columns

- From the merged dataset extracting the total number of states.
- From the total number of states available in the dataset, Firstly I consider ‘Louisiana’ state.
- From ‘Louisiana state’ considering ‘KENNER’ metropolitan area

- **(ii)** For the list of businesses with common reviewers I implemented a groupby function Implementing a aggregate function while grouping the 'business_id' and 'review_count'
- By this we will eliminate re-occurrence of business_id in the dataframe.
- Below is the Sample list of businesses with common reviewers

	business_id	review_count
0	-4VQum5gCgEfSZcFycmAHw	5.0
1	-8luB5pJ7d9UOoiF7wikkw	182.0
2	-BAbmwcDslfiYVNizbjSnw	5.0
3	-FXxCFnPPya9o5_8wAVSGQ	30.0
4	-H5v2-mADBj8_n2yeACjLA	116.0
...	...	...
579	zMX25T0IpPzitVFmke51YQ	8.0
580	zOSAZOTA8wbV2P7T9RPbAQ	19.0
581	zPBTvq6lg0XAmnf7BczgTQ	11.0
582	zolbWK0zAk6lY-DqiZPLjA	29.0
583	zvDYG_6f9omEcKW-20IUCQ	9.0

584 rows × 2 columns

- **(iii) For the reviewers count > 130, for KENNER metropolitan area of Louisiana state**

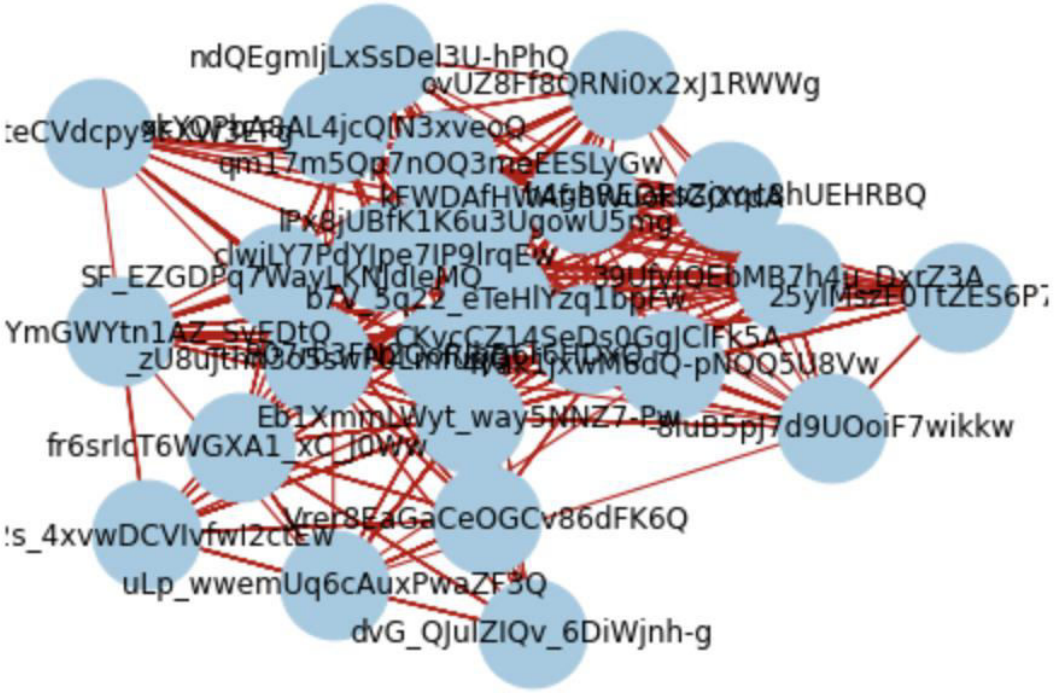


Fig: Number of Edges in the Graph: 1657
Number of Nodes in the Graph: 25

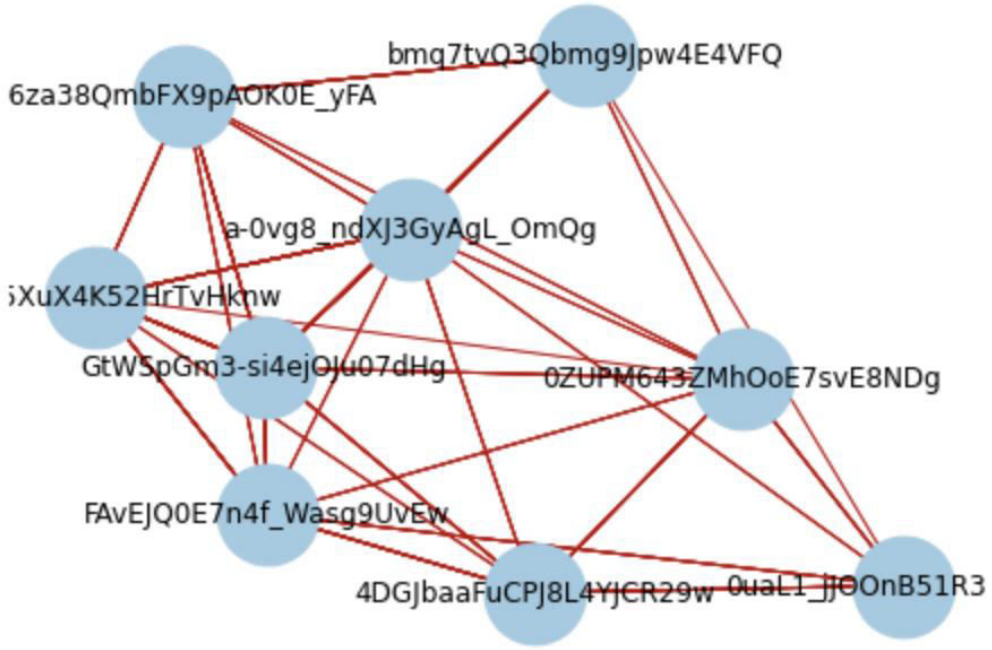
- edges and the business_IDs are extracted, and we get a list of business and user IDs
- In the graph we take unique business IDs as nodes.
- Edges list has tuples () with business IDs with one business ID as source and other Business ID as destination
- Edges are total list of s-t pairs of the graph.

```
Degree Centrality is :
 {'-8luB5pJ7d9UOoiF7wikkw': 2.625, 'CKvcCZ14SeDs0GgJClFk5A': 7.0, 'uLp_wwemUq6cAuxPwaZF3Q': 3.1666666666666665, 'b7V_
5q22_eTeHlYzq1bpFw': 15.583333333333332, 'Eb1XmmLWyt_way5NNZ7-Pw': 14.375, 'clwjLY7PdYJpe7IP9lrqEw': 9.291666666666666
6, 'h4ghREOFsZjxqc8hUEHRBQ': 2.958333333333333, 'qm17m5Qp7nOQ3meEESLyGw': 10.75, 'xkYOPbA8AL4jcQIN3xveoQ': 5.875, '4v
ak1jxwM6dQ-pNQQ5U8Vw': 6.416666666666666, '_zU8uJthn3o5swP0LmfuqQ': 6.833333333333333, '39UfyJQEbMB7h4u_DxrZ3A': 3.87
5, '25ylMszF0TtZES6P7m6fvA': 1.9166666666666665, 'R07rb3FN2QoRIbB6r6HDxQ': 6.5, 'SF_EZGDPq7WayLKNJdIeMQ': 5.0, 'kFWDA
fHWAf-BWuokIGOYdA': 6.625, 'Vrer8EaGaCeOGCv86dFK6Q': 3.7916666666666665, 'fr6srIcT6WGXA1_xC_J0Ww': 3.333333333333333,
'ovUZ8Ff8QRNi0x2xJ1RWWg': 3.0, 'ndQEgmIjLxSsDel3U-hPhQ': 2.833333333333333, 'lPx8jUBfK1K6u3UgowU5mg': 8.25, 'WStVCYmG
WYtn1AZ_SvEDtQ': 2.833333333333333, 'uK0rzqteCVdcpy9FXW3EFg': 1.3333333333333333, 'dvG_QJuIZIQv_6DiWjnh-g': 1.6666666
666666665, 'aZ2s_4xvwDCVIvfwI2ctEw': 2.25}
------
Values of Degree centrality is :
 dict_values([2.625, 7.0, 3.1666666666666665, 15.583333333333332, 14.375, 9.291666666666666, 2.958333333333333, 10.7
5, 5.875, 6.416666666666666, 6.833333333333333, 3.875, 1.9166666666666665, 6.5, 5.0, 6.625, 3.7916666666666665, 3.333
333333333333, 3.0, 2.833333333333333, 8.25, 2.833333333333333, 1.3333333333333333, 1.6666666666666665, 2.25])
------
Sum of all the degree centrality values is:
 138.08333333333334
------
Total nodes of the graph is:
 25
------
Average Degree Centrality is :
 5.523333333333333
```

Average Degree Centrality for the above Graph is: 5.23

Repeating the experiment for a different metropolitan area for the same state (Louisiana)

- I consider city 'HARVEY' and review count again > 130
-



-

Fig: Number of Nodes in the graph is: 9
Number of Edges in the graph is: 198

```
Degree Centrality is
 {'0ZUPM643ZMhOoE7svE8NDg': 3.125, 'bmq7tvQ3Qbmg9Jpw4E4VFQ': 4.0, 'a-0vg8_ndXJ3GyAgL_OmQg': 9.25, 'FAvEJQ0E7n4f_Wasg9
UvEw': 6.0, 'GtWSpGm3-si4ejOJu07dHg': 8.75, '0uaL1_jJOOnB51R3XIYrKw': 2.125, '4DGJbaaFuCPJ8L4YJCR29w': 4.0, 'XbobIU5X
uX4K52HrTvHknw': 8.625, 'S6za38QmbFX9pAOK0E_yFA': 3.625}
------
Values of Degree centrality is :
 dict_values([3.125, 4.0, 9.25, 6.0, 8.75, 2.125, 4.0, 8.625, 3.625])
------
Sum of all the degree centrality values is:
 49.5
------
Total nodes of the graph is:
 9
------
Average Degree Centrality is :
 5.5
------
```

(V) Differences in two values?

I consider the state LOUISIANA State for analysis, first metropolitan area is **KENNER** for this city we have a greater number of businesses and more review counts. In the graph we have a total of 25 nodes (Unique business ID's) and total number of Edges in graph is 1657 (total connections if we have same reviewer for two businesses).

Similarly, 2nd metropolitan area considered is **HARVEY** and the graph contains a total of 9 nodes (Unique business ID's) and total number of Edges in graph is 198.

Average Degree Centrality For graph (KENNER metropolitan area): 5.523
Average Degree Centrality For graph (HARVEY metropolitan area): 5.5

We can observe that even with more businesses and user reviews in KENNER city than in HARVEY city we have same average degree centrality. So we can conclude that degree centrality increases if more users give more reviews rather than the number of businesses available. (As we consider a criteria of connecting two businesses if reviewed by the same user).