International Journal of Engineering & Technology, 7 (2.7) (2018) 841-847

International Journal of Engineering & Technology

Website: www.sciencepubco.com/index.php/IJET

Research Paper



Network analysis as a method of assessing terms in a dataset comprising eating disorders

Dr. Adimulam Yesu Babu[1], Dr. Deepak Nedunuri[2*], T. Venkata Sai Krishna[3]

[1]*Head and Professor, Department of CSE, Sir C R Reddy College of Engineering, India*
[2*]*Associate Professor, Department of CSE, Sir C R Reddy College of Engineering, India*
[3]*Research Scholar, Department of CSE, JNTU-KAKINADA*
[*] *Corresponding author E-mail: nedunurideepak@gmail.com*

Abstract

Eating disorders are central reason of physical and psycho-social morbidity. Several factors have been identified as being associated with the prevalence and progression of eating disorders in humans. Scientific investigation was carried out to assess the usage of terms in manuscript titles of nearly 900 published articles followed by network analysis and network centralities using R programming. The tm package, term document matrix function was utilized to create a term document matrix (TDM) from the corpus. A binary word matrix comprising 17 terms was created based on higher probability of occurring a term in a column. An agglomerative hierarchical clustering technique using ward clustering algorithm was presented. A data frame from the TDM was created to store data and used to plot word cloud based on word frequencies. An undirected network graph was plotted based on terms that appeared in the term matrix. Centralization measures such as Degree centrality, Closeness, Eigenvector and betweenness Centrality were reported.

Keywords: *Eating disorder, R programming, Term matrix, Word cloud, Clustering, Network analysis*

1. Introduction

Classification of eating disorders is three main types, anorexia nervosa (AN), bulimia nervosa (BN), and binge eating disorder. In anorexia nervosa, the pursuit of weight loss is successful in that a very low weight is achieved. This loss of weight is primarily the result of a severe and selective restriction of food intake, with foods viewed as fattening being excluded. The main feature that distinguishes bulimia nervosa from anorexia nervosa is that attempts to restrict food intake are punctuated by repeated binges. A typical eating disorders closely resemble anorexia nervosa and bulimia nervosa, and many are as severe and long lasting.

Factors that contribute EDs include several risk factors [1] including biological (unbalanced controlling of genes related to hunger, appetite and digestion in brain), psychological (depression, anxiety, anger, stress etc), social (Cultural pressures, body shape, beautifications etc) and interpersonal factors (family troubles, emotions and feelings, abuse (etc) [2].
Pertaining to the severity of the problem of EDs, several studies have been focussed on the role of genetics in the etiology of EDs. The two methodologies involving molecular investigations are linkage and association analyses [3]. Linkage analysis represents studying paternal and maternal localized disease genes and their influence on offspring. Further, a genetic association study involves comparing a trait of interest against a control who does not express the trait. Bulik et al 2003b [4], identified chromosome 10 and linkage on chromosome 14 in a genome wide analysis of 316 families with at least two subjects with BN. On the other hand, genetic association studies revealed association of EDs with Brain-Derived Neurotropic Factor (BDNF), Serotonin transporter, Monoamine oxidase A (MAO-A), cannabinoid receptor and potassium channel (KCNN3) genes. Evidence suggests that BDNF playsan important role in the patho physiology of psychiatric diseases, including mood disorders, and in the mechanism of action of therapeutic agents [5], [6].

Several literature resources reported nearly 900 articles on all types and aspects of eating disorders which suggests a significant importance and impact of EDs in human community. Hence, an investigation was carried out to assess the usage statistics of terms in manuscript titles of published articles followed by network analysis and network centralities using R programming.

2. Materials and Methods

2.1 R Program

Recent advances in molecular biology extended our chances to discover and comprehend the complexity of biological systems. Data integration has been made possible with the advent of computational tools, statistical methods and allied softwares [7]. R program is a free software environment for statistical computing and graphics. Packages are being developed to address the specific needs of systems biology approaches, such as network modeling, simulation formal verification, and graph visualization [8]. Castelo and Roverato [9] developed an R package for molecular network discovery using micro array data. Le Meur and Gentleman [10] identified multi protein complexes and pairs of multi protein complexes that share a unusually high number of synthetic genetic interactions. An R program was written with all dependencies and functionalities to perform network analysis and visualization of data. Many different add-on packages have been used such as word

Copyright © 2018 Authors. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

cloud-to create word cloud, tm-text mining of data, gg plot-data visualization, Rcolor Brewer-coloring words in word cloud, ape-dendrogram package and I graph-for network analysis graph.

2.2 Dataset

A dataset of 900 articles associated with the disease, 'Eating Disorders' published in various journal resources were considered in the study. Data in the form of article title information was only considered and was extracted from Mala cards human disease database. Mala Cards is an integrated searchable database of human maladies and their annotations, modelled on the architecture and richness of the popular Gene Cards database of human genes [11]. The dataset with titles, author names and year were segregated and used as csv file input.

2.3 Network Analysis

Network analysis represents the use of network and graph theories towards investigation of prime features of a group of objects representing similar nature from a significant cluster. Networks are created in terms of nodes and edges. Visual depiction of networks is imperative to understand the network data and deliver the outcome of analysis [12].Data exploration was done through displaying nodes and edges in various layouts such as fruchterman-reingold [13] and kamada-kawai [14], respectively.

3. Results and Discussion

The dataset extracted from Mala cards database was visually inspected to identify common words that appear more times in title subjects. Few words such as 'study', 'effect', 'associated', 'test' etc were excluded as commonly occurring words. Hyperlinks, if any, punctuations, numbers and extra spaces between words were also excluded from the dataset. Initially a data frame is created followed by creating a corpus of words that appear repeatedly and more number of times in the dataset. Further, tm package, term document matrix function was utilized to create a term document matrix (TDM) from the corpus. TDM defines the frequency of terms that occur in the collected database. In a TDM, rows correspond to documents and columns correspond to terms. The TDM obtained for the current dataset is given below and the frequency of terms is given in Table-1. With a frequency limit of 150, nearly 28 terms appeared and the top 15 term frequencies are given in Table-2 and Figure 1.

<<TermDocumentMatrix (terms: 1639, documents: 900)>>
Non-/sparse entries: 8635/1466465
Sparsity : 99%
Maximal term length: 30
Weighting : SMART ltn (SMART)

Table 1: Frequency of Terms that appeared in TDM with a frequency limit of 150.

[1]	"adolesc" "anorexia" "associ"
	"behavior" "bing"
[6]	"bingeeat" "bulimia" "clinic"
	"comparison" "control"
[11]	"effect" "examin" "factor" "femal" "gene"
[16]	"nervosa" "obes" "patient"
	"person" "relat"
[21]	"risk" "role" "symptom"
	"therapi" "treatment"
[26]	"trial" "weight" "women"

Table 2: Frequency of all 28 terms that showed against a frequency limit of 150.

WORD	FREQ
bing	471.2751
nervosa	395.0534
patient	382.1630
bulimia	304.8765
treatment	302.2320
obes	300.0046
women	293.7795
associ	278.1366
symptom	268.1728
behavior	230.8917
examin	220.6365
adolesc	219.1803
anorexia	213.9146
control	207.2121
femal	198.8087
clinic	197.1308
therapi	185.7542
bingeeat	185.7015
weight	185.6635
factor	179.6741
person	171.8227
risk	171.8227
trial	170.3601
effect	168.6419
comparison	160.6948
gene	158.8548
relat	150.6471
role	150.6471

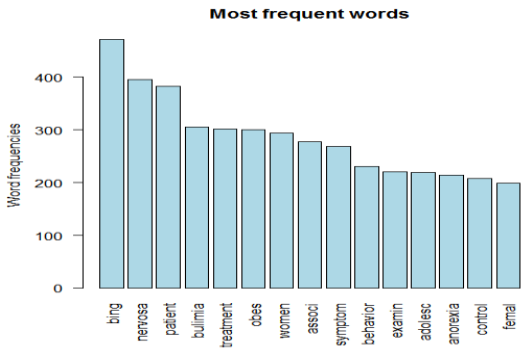


Figure 1: Graphical representation of top 15 terms with a frequency limit of 150.

In order to avoid cluttering of terms, a frequency of > 99% percentile which is otherwise referred as less than 1% sparse was employed on TDM which resulted in 17 terms as binary word matrix (Table-3). However, with about 98% probability, nearly 33 terms appeared in the matrix. Prior constructing a binary matrix, documents which does not contain any terms are excluded and the document for which a term is repeated more than once is counted as 1 entry. Binary method is used to find the relative similarity of those terms that have higher probability of occurring together in a column.

Table 3: Appearance of 17 terms in document summary represented as term matrix.

Terms	Docs 1	2	4	5	6	8	10	11	12	14	15	17	18	20	21	22	24	25	28	29	30	31	33	34
adolesc	0	0	0	1	0	0	0	1	0	0	0	0	0	0	0	0	0	0	1	1	0	0	0	0
anorexia	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	1	0	0	0	0	1	0
associ	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0
behavior	0	0	0	1	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
bing	1	1	0	0	0	1	0	0	0	1	0	1	0	1	0	1	0	0	0	0	1	1	1	0
bulimia	0	0	0	0	0	1	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	1	1	0
clinic	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	1
control	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0
examin	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	1	0	1	0	0	0	0
femal	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
nervosa	0	0	0	0	0	1	0	0	0	0	0	0	0	0	1	0	0	1	0	0	0	1	1	0
obes	0	0	0	0	1	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0
patient	0	0	0	0	0	0	0	0	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0
symptom	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0
therapi	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
treatment	0	1	0	0	0	0	1	0	0	0	0	0	0	0	0	0	1	0	0	0	0	1	0	0

Considering all terms in the binary matrix, a distance based agglomerative hierarchical clustering technique was implemented to identify which groups of terms appeared in each cluster with k=5. Figures 2a and 2b represent clusters of 17 terms appeared in term matrix when a probability of 99% is used, whereas a probability of 98% resulted in 33 terms. A ward clustering algorithm (ward1- "on a scale of squared distances" and ward2 - "on a scale of distances") was employed, where both the clusters appear similar. It is evidenced that the eating disorder types, anorexia, bulimia and nervosa appear under one clade. Bing eating disorder is significant in obese patients, hence all these three terms appearing as one group in cluster-4 is justified.

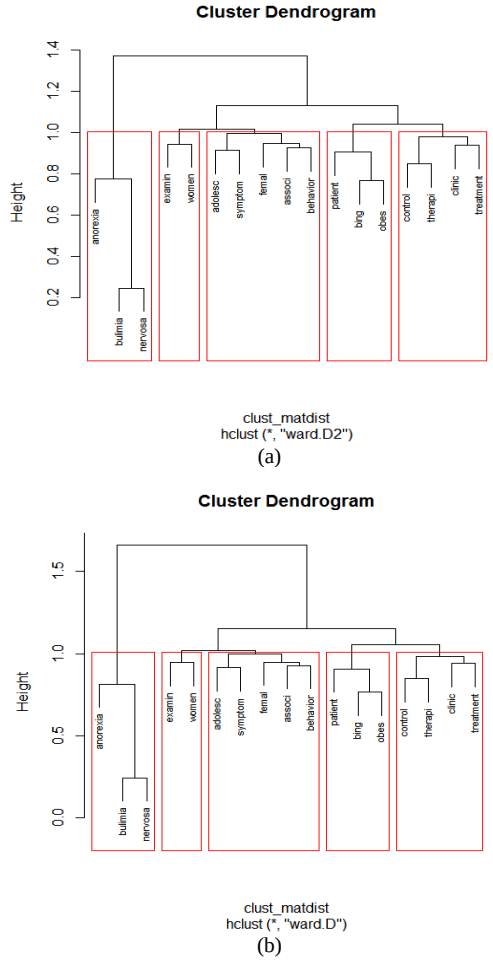


Figure 2: Ward D1 and D2 clustering analysis of frequent terms in matrix.

In the next step, word cloud data was obtained from dataset as well as from word frequencies as obtained in term matrix data. A data frame from the TDM was created to store data and used to plot word cloud based on word frequency. The specific high frequency words appear big in the word cloud and therefore, from Figure-3, it is observed that cluster-4 elements appeared bigger which represent that these terms are used more number of times in publications on eating disorders. Apart from 17 frequent terms in term matrix, the more concentrated areas in eating disorders were found to have other terms also as evidenced from a word cloud of complete dataset, shown in Figure 4.Frequency distribution of terms with a maximum frequency of 150 times a particular term is found in the dataset was plotted and showed in Figure 5.



Figure 3: Wordcloud showing specific word appearances of term matrix data.



Figure 4: Word cloud showing specific word appearances of complete dataset.

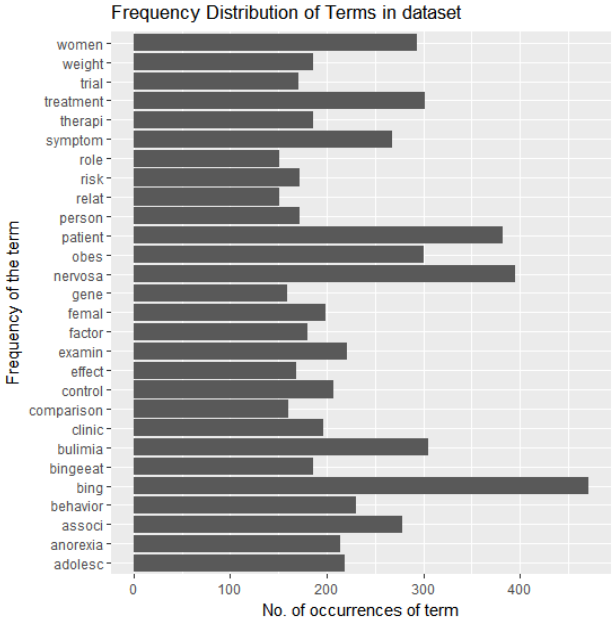


Figure 5: Frequency distribution of terms in the dataset with frequency more than 150.

3.1 Network Analysis

Further, an undirected network graph was plotted based on terms that appeared in the term matrix with sparsity less than 96% (Figure 6). The graph density can be calculated using the equation:

$$density = \frac{2|E|}{|V|(|V|-1)}$$

Our analysis result was found to be in line with the agreement of Leclerc RD (2008) where the connectivity densities for terms in the matrix were found to be less than 0.1 (Figure 7).

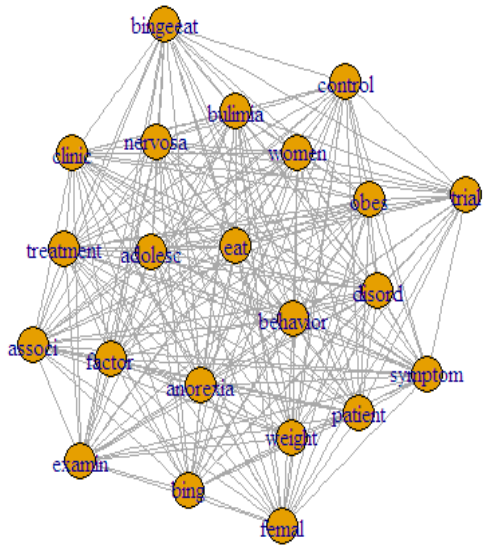


Figure 6: Network analysis graph showing top terms that appeared in term matrix with sparsity less than 96%.

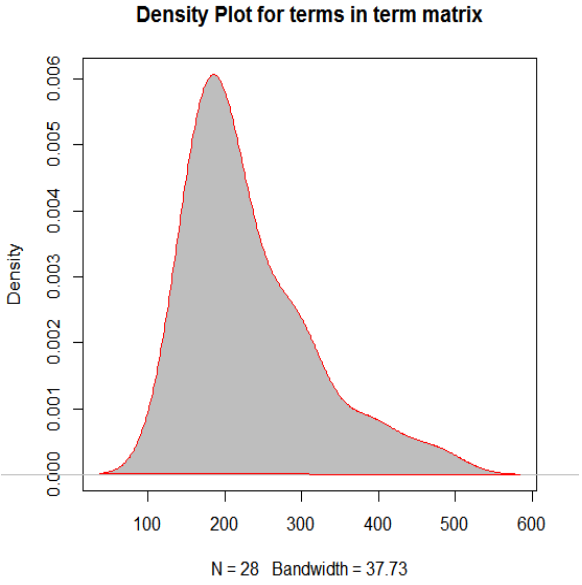


Figure 7: Density plot of term matrix showing a sparse graph.

3.2 Network Centralities

In general, central nodes or intermediate nodes affect the topology of the network. Some points are not central to the data, however, might have crucial role, hence it is important to detect such nodes using Degree centrality, Closeness centrality and Eigenvector Centrality. Centralization measures whether a network has a star-like topology or has on average the same connectivity. If the centralization value is nearer to 1, the network represents a star-like topology. If the value nearer to 0, it is more likely that the nodes of the network have on average the same connectivity.

Degree Centrality shows that an important node is involved in a large number of interactions. For a node i, the degree centrality is calculated as C d (i) = deg(i).Nodes with very high degree centrality are called hubs since they are connected to many neighbours. Degree centrality is the simplest network centrality measure. It only takes into account the degree of a node, which is the number of nodes that a given node is connected to [15].

Closeness Centrality indicates important nodes that can communicate quickly with other nodes of the network. Let G = (V, E) be an undirected graph. Then, the centrality is defined as

$$C_{clo}(i) = \frac{1}{\sum_{t \in V}^{|V|} dist(i,j)}$$

where dist(i, j) denotes the distance or else the shortest path p between the nodes i and j.

Eigenvector Centrality ranks high values to the nodes that are connected to important neighbours [16]. Let G = (V, E) be an undirected graph and A the adjacency matrix of network G. The eigenvector centrality is the eigenvector C eiv- of the largest eigen value λ max in absolute value such that λC eiv = AC eiv. The values pertaining to centrality measurements were given in Table 4.

Measurement	Value
Degree centrality	0.0083
Closeness centrality	0.017
Eigenvector centrality	0.794

Betweenness centrality displays high ranks for nodes which are intermediate between neighbours [17]. These nodes are crucial in communicating between two neighbours. Hence, betweenness centrality shows significant nodes on a high fraction of paths between other nodes in the network. The centralities of all terms representing nodes are given below, where it is evidenced that node disorder is more central.

Table 4: Degree centrality, Closeness and Eigenvector centrality measurements of network graph.

adolesc	anorexia	associ	behavior	bing	bulimia	control	disord
0	0	0	0	0	0	0	104
eat	examin	nervosa	obes	patient	symptom	treatment	women
0	0	0	0	0	0	0	0

Further, it was identified that many genes are found to be expressed in eating disorder disease state/conditions. Nearly 14 genes such as SLC6A4, LEP, BDNF, COMT, DRD2, POMC, ADIPOQ, CCK, CNR1, GHRL, ALB, HTR2C, LEPR and NPI were extracted from Malacards database. The list contained one or more genes associated with other diseases, for example, few genes expressed in obesity (ADIPOQ, ALB, BDNF, CCK, CNR1, GHRL) are found to be expressed in diabetes mellitus. Diabetic encephalopathy and periampullary adenoma were found to be affiliated with similar genes such as ADIPOQ, LEP, LEPR, SLC6A4 respectively. Table 5 contains the data.

Table 5: Eating disorder related genes and diseases.

S. No.	Related Disease	Affiliated Genes
1	obsessive-compulsive personality disorder	BDNF, COMT, HTR2C, SLC6A4
2	avoidant personality disorder	BDNF, CCK, CNR1, COMT, DRD2, HTR2C
3	schizophrenia	BDNF, CCK, CNR1, COMT, DRD2, HTR2C
4	obesity	ADIPOQ, ALB, BDNF, CCK, CNR1, GHRL
5	neurodegeneration with optic atrophy, childhood onset	COMT, SLC6A4
6	basal cell carcinoma	DRD2, SLC6A4
7	splenic disease	DRD2, SLC6A4
8	erysipelas	COMT, DRD2
9	meckel's diverticulum	BDNF, COMT, SLC6A4
10	vogt-koyanagi-harada disease	BDNF, COMT, SLC6A4
11	phaeohyphomycosis	BDNF, COMT, SLC6A4
12	rabies	CCK, SLC6A4
13	malt worker's lung	CCK, NPY
14	hemorrhagic cystitis	COMT, DRD2, SLC6A4
15	anosognosia	COMT, DRD2, SLC6A4
16	tongue disease	COMT, DRD2, SLC6A4
17	conjunctival degeneration	HTR2C, SLC6A4
18	west syndrome	BDNF, CNR1, NPY
19	viral gastritis	CCK, GHRL, SLC6A4
20	jaw cancer	CNR1, COMT
21	epidermolysisbullosaacquisita	BDNF, POMC
22	hidradenocarcinoma	LEP, POMC
23	gestational choriocarcinoma	GHRL, LEP, NPY
24	lipedematous scalp	COMT, DRD2
25	active vestibular meniere's disease	BDNF, POMC
26	parkinsonism-dystonia, infantile	COMT, DRD2, SLC6A4
27	dental fluorosis	HTR2C, LEP, SLC6A4
28	tropical sprue	GHRL, LEP, PYY
29	pediatricmesenchymalchondrosarcoma	POMC, SLC6A4
30	peripheral nervous system neoplasm	CCK, DRD2, GHRL
31	hemopneumothorax	LEP, POMC
32	pure mitochondrial myopathy	ADIPOQ, LEP
33	pancreatic cystadenocarcinoma	CCK, PYY
34	lymphadenitis	BDNF, POMC, SLC6A4
35	hypothalamic neoplasm	NPY, POMC
36	ulcerative colitis	BDNF, COMT, DRD2, SLC6A4
37	maternal uniparentaldisomy of chromosome 13	BDNF, COMT, DRD2, SLC6A4
38	vulvar melanoma	CCK, PYY, SLC6A4
39	pelvic inflammatory disease	DRD2, HTR2C, SLC6A4
40	alcoholic liver cirrhosis	BDNF, CNR1, DRD2, SLC6A4

41	retinal dystrophy with inner retinal dysfunction and ganglion cell abnormalities	CCK, GHRL, PYY
42	transsexualism	BDNF, CNR1, POMC
43	acinar cell carcinoma	BDNF, DRD2, NPY, SLC6A4
44	pyodermagangrenosum	COMT, POMC, SLC6A4
45	central sleep apnea	HTR2C, SLC6A4
46	sexual disorder	BDNF, COMT, DRD2, SLC6A4
47	autoimmune disease of central nervous system	CCK, CNR1, POMC
48	laryngeal carcinoma	CCK, LEP, POMC
49	intestinal disaccharidase deficiency	ALB, POMC
50	semantic agnosia	CCK, CNR1, POMC
51	tracheal calcification	ADIPOQ, GHRL, LEP
52	parasitic ectoparasitic infectious disease	CCK, CNR1, POMC
53	nondystrophicmyotonia	ADIPOQ, LEP
54	hemoglobinopathy	ALB, POMC
55	cervix uteri carcinoma in situ	COMT, GHRL, LEP
56	tenosynovial giant cell tumor	ADIPOQ, LEP
57	albinism, oculocutaneous, type v	DRD2, HTR2C, SLC6A4
58	lipoic acid synthetase deficiency	LEP, LEPR
59	vulvar disease	BDNF, CNR1, COMT, DRD2, SLC6A4
60	phlegmonousdacryocystitis	COMT, DRD2, POMC, SLC6A4
61	brain meningioma	ALB, BDNF, COMT, DRD2
62	parkinson disease, late-onset	BDNF, CNR1, COMT, DRD2, SLC6A4
63	cocaine dependence	BDNF, CNR1, DRD2, POMC
64	fechtner syndrome	CCK, COMT, POMC, SLC6A4
65	potato nose	ADIPOQ, LEP
66	chronic myelomonocyticleukemia	ALB, LEP, POMC
67	antidepressant type abuse	ADIPOQ, ALB, LEP
68	transvestism	ADIPOQ, SLC6A4
69	patellofemoral pain syndrome	ADIPOQ, LEP, POMC
70	keratoacanthoma	ADIPOQ, LEP, POMC
71	chronic cervicitis	BDNF, COMT, DRD2, HTR2C, SLC6A4
72	vernal conjunctivitis	BDNF, COMT, DRD2, HTR2C, SLC6A4
73	aicardi-goutieres syndrome 2	BDNF, COMT, DRD2, HTR2C, SLC6A4
74	craniosynostosis, adelaide type	BDNF, COMT, DRD2, HTR2C, SLC6A4
75	hypertrichosis	ADIPOQ, ALB, LEP
76	alveoli adenoma	ALB, GHRL, LEP, POMC
77	angelman syndrome	ADIPOQ, GHRL, LEP, PYY
78	heroin dependence	BDNF, CNR1, DRD2, POMC, SLC6A4
79	constrictive pericarditis	ALB, NPY
80	enthesopathy	BDNF, CCK, COMT, DRD2, NPY, SLC6A4
81	hypoglycemic coma	BDNF, COMT, HTR2C, POMC, SLC6A4
82	jmp syndrome	LEPR, MC4R, POMC
83	hypertension, essential	ADIPOQ, ALB, LEP, NPY
84	short bowel syndrome	ADIPOQ, LEP, LEPR
85	isthmus cancer	ADIPOQ, LEP, LEPR
86	diabetic encephalopathy	ADIPOQ, LEP, LEPR, SLC6A4
87	periampullary adenoma	ADIPOQ, LEP, LEPR, SLC6A4
88	femoral neuropathy	ADIPOQ, ALB, LEP, POMC
89	pervasive developmental disorder	BDNF, CCK, CNR1, COMT, GHRL, POMC
90	purpura	BDNF, CNR1, COMT, DRD2, NPY, POMC
91	gemistocytic astrocytoma	ADIPOQ, ALB, BDNF, LEP, POMC
92	hordeolumexternum	ADIPOQ, ALB, LEP, LEPR
93	dysthymic disorder	BDNF, CCK, GHRL, HTR2C, LEP, NPY
94	deafness, x-linked 5	ADIPOQ, CNR1, GHRL, LEP, LEPR, MC4R
95	hepatic adenoma, somatic	ADIPOQ, ALB, GHRL, LEP, LEPR, NPY
96	narcolepsy 2	BDNF, CCK, CNR1, COMT, DRD2, GHRL
97	survival motor neuron spinal muscular atrophy	ADIPOQ, ALB, GHRL, LEP, LEPR, MC4R
98	mucolipidoses	ADIPOQ, ALB, GHRL, LEP, LEPR, MC4R
99	angiomatous meningioma	ADIPOQ, CNR1, GHRL, LEP, LEPR, MC4R
100	diabetes mellitus, noninsulin-dependent	ADIPOQ, ALB, CNR1, GHRL, LEP, LEPR
101	acatalasemia	ADIPOQ, BDNF, CCK, COMT, GHRL, LEP
102	lymphosarcoma	ADIPOQ, ALB, BDNF, CCK, CNR1, COMT

Figure 8 given below is an example of numbering genes which have frequencies of appearance in other diseases; however, associated with eating disorders (Table 6).The graph is generated based on closely related terms. In other words, the 14 genes which are associated with 102 diseases with eating disorders in the dataset are grouped together based on the frequency of terms. Assessing the graph elements at the clustered centres reveal the presence of related terms while those at the periphery are less relevant and are random.

Table 6: List of 14 genes associated with eating disorders.

SLC6A4	LEP	BDNF	COMT	DRD2	POMC	ADIPOQ	CCK	CNR1	GHRL	ALB	HTR2C	LEPR	NPI
41	36	35	35	31	29	28	21	21	20	18	14	13	11

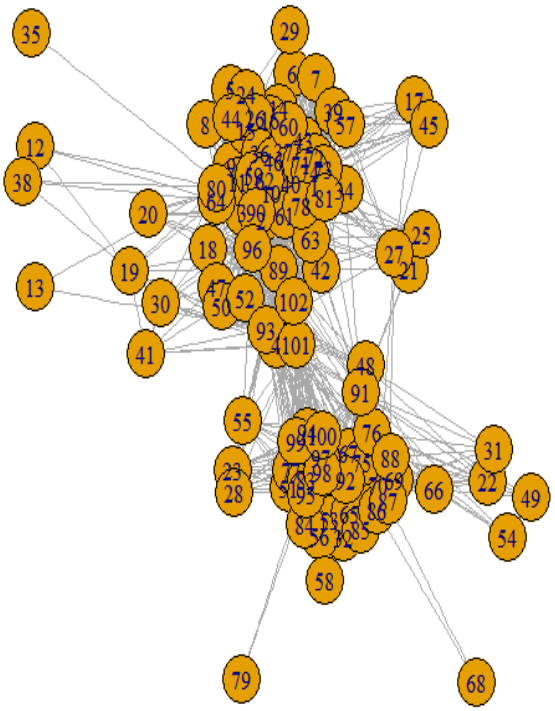


Figure 8: Numbered network graph of all 102 associated diseases in eating disorder.

4. Conclusion

Assessing the usage of terms, network analysis and centrality statistics of terms used in manuscript titles of nearly 900 published articles extracted from Malacards human disease database was carried out using R programming. An undirected network graph plotted based on terms that appeared in the term matrix followed by a density plot, where the connectivity densities for terms in the matrix were found to be less than 0.1 which suggested that the nodes of the network have on average the same connectivity Centralization measures such as Degree centrality, Closeness centrality, Eigenvector centrality and betweenness centrality resulted in values within the limits. Further analysis has been extended considering genes associated with eating disorders which further represent possible genes which might co-exist in the eating disorder disease patterns.

References

[1] Jacobi, C., Hayward, C., de Zwaan, M., Kraemer, H. &Agras, W.S. (2004). Coming to terms with risk factors for eating disorders: Application of risk terminology and suggestions for a general taxonomy. Psychological Bulletin, 130, 1, 19-65.

[2] Jacobi, C. &Fittig, E. (2010). Psychosocial risk factors for eating disorders.In Agras, W.S. (Ed.), Oxford Handbook of Eating Disorders. Oxford University Press: N.Y.

[3] P.E. Slagboom and I. Meulenbelt. Organisation of the human genome and our tools for identifying disease genes. Biological Psychology 61 (2002) 11-31.

[4] Bulik CM, Devlin D et al. (2003b) Significant linkage on chromosome 10p in families with bulimia nervosa. Americal Journal of Human Genetics, 72: 200-7.

[5] Duman, R.S., 2002. Synaptic plasticity and mood disorders. Mol. Psychiatry 7, S29–S34][Coyle, J.T., Duman, R.S., 2003. Finding the intracellular signaling pathways affected by mood disorder treatments. Neuron 38, 157–160.

[6] Hashimoto, K., Shimizu, E., Iyo, M., 2004. Critical role of brain-derived neurotrophic factor in mood disorders. Brain Res. Brain Res. Rev. 45, 104–114.

[7] Nolwenn Le Meur and Robert Gentleman. Analyzing Biological Data Using R: Methods for Graphs and Networks. Chapter 19.

[8] Huber W, Carey VJ, Long L, Falcon S, Gentleman R. (2007) Graphs in molecular biology. BMC Bioinformatics, 8(6):S8.

[9] Castelo R, Roverato A. (2009) Reverse engineering molecular regulatory networks from microarray data with qp-graphs. J Computational Biol, 16(2):213–227.

[10] Le Meur N, Gentleman R. (2008) Modeling synthetic lethality. Genome Biol, 9(9):R135.

[11] http://www.malacards.org/

[12] Linton C. Freeman. "Visualizing Social Networks". Journal of Social Structure.

[13] Fruchterman, Thomas M. J.; Reingold, Edward M. (1991), "Graph Drawing by Force-Directed Placement", Software – Practice & Experience, Wiley, 21 (11): 1129–1164.

[14] Kamada, Tomihisa; Kawai, Satoru (1989), "An algorithm for drawing general undirected graphs", Information Processing Letters, Elsevier, 31 (1): 7–15.

[15] Freeman LC. Centrality in social networks: conceptual clarification, Soc. Networks, 1979, vol. 1 (pg. 215-239.

[16] Newman MEJ. The structure and function of complex networks, SIAM Rev, 2003, vol. 45 pg. 167.

[17] Freeman LC. A set of measures of centrality based on between's, Sociometry, 1977, vol. 40 (pg. 35-41.