Identifying Dissimilarity Matrix Using Network Analysis as a Method in Eating Disorders

T. VENKATA SAI KRISHNA[1], Dr. YESU BABU ADIMULAM[2], Dr. R. KIRAN KUMAR[3]
[1]*Research Scholar, J N T UNIVERSITY-Kakinada, INDIA*
[*2]*Professor & HoD, Dept. of CSE, Sir C R Reddy College of Engineering, Eluru, Andhra Pradesh, INDIA*
[3]*Assistant Professor, Dept. of CSE, Krishna University, Machillipatanam, Andhra Pradesh, INDIA*

Abstract **-** 'Eating Disorders' published in various journal resources were considered in the study. A dataset of 900 articles associated with the disease. Data in the form of article title information was only considered and was extracted from Malacards human disease database. The dataset with titles, author names and year segregated and used as input file. 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*. Data exploration was done through displaying nodes and edges in various layouts such as fruchterman-reingold and kamada-kawai. The dataset extracted from Malacards database was visually inspected to identify common words such as 'study', 'effect', 'associated', 'test' etc and are excluded from the study. 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. Term document matrix function was utilized to create a term document matrix (TDM) from the corpus which defines the frequency of terms that occur in the collected database. 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. A distance based agglomerative hierarchical clustering technique was implemented to identify which groups of terms appeared in each cluster with k=5. Further, word cloud data was obtained from dataset as well as from word frequencies as obtained in term matrix data. 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.

Keywords **-** Eating Disorders, Malacards, TDM, Clustering.

I. INTRODUCTION

Several factors have been identified as being associated with the prevalence and progression of eating disorders in humans. Advancements in molecular and neurobiology revealed the regulatory mechanism of neuro-transmitters, neuro-peptides, neuro-hormones etc acting on the hypothalamus and cortical brain regions influencing intake of food materials, mood variations, response to stress and cognition.

A. NETWORK ANALYSIS

Network analysis is the interdisciplinary study of social relations and has roots in anthropology, sociology, psychology, and applied mathematics. It conceives of social structure in relational terms, and its most fundamental construct is that of a social or biological network. The nodes or members of the network can be groups or organizations as well as people. Network analysis involves a combination of theorizing, model building, and empirical research, including (possibly) sophisticated data analysis. The goal is to study network structure. Network structure can be studied at many different levels: the dyad, triad, subgroup, or even the entire network. Furthermore, network theories can be postulated at a variety of different levels. Although this multilevel aspect of network analysis allows different structural questions to be posed and studied simultaneously, it usually requires the use of methods that go beyond the standard approach of treating each individual as an independent unit of analysis. This is especially true for studying a complete or whole network: a census of a well-defined population of social actors in which all ties, of various types, among all the actors are measured. Such analyses might study structural balance in small groups, transitive flows of information through indirect ties, structural equivalence in organizations, or patterns of relations in a set of organizations.

Network analysis allows a researcher to model the inter dependencies of organization members. The paradigm provides concepts, theories and methods to investigate how informal organizational structures intersect with formal bureaucratic structures in the unfolding flow of work-related actions of organizational members and in their evolving sets of knowledge and beliefs [1].

B.

B. COMPLETE NETWORKS

In complete network studies, a census of network ties is taken for all members of a prespecified population of network members. A variety of methods may be used to observe the network ties (e.g., survey, archival, participant observation), and observations may be made on a number of different types of network tie. Studies of complete networks are often

appropriate when it is desirable to understand the action of network members in terms of their location in a broader social system (e.g., their centrality in the network, or more generally in terms of their patterns of connections to other network members).

C. NOTATION

In the simplest case, network studies involve a single type of directed or non-directed tie measured for all pairs of a node set N = {1,2, . . . ,n} of individual objects. The observed tie linking node i to node j (i,j Î N) can be denoted by x_{ij} and is often defined to take the value 1 if the tie is observed to be present and 0 otherwise. The network may be either directed (in which case x_{ij} and x_{ji} are distinguished and may take different values) or nondirected (in which case x_{ij} and x_{ji} are not distinguished and are necessarily equal in value). Other cases of interest include the following:

1. Valued networks, where x_{ij} takes values in the set {0,1, . . . ,C – 1}
2. Time-dependent networks, where x_{ijt} represents the tie from node i tonode j at time t
3. Multiple relational or multivariate networks, where x_{ijk} represents the tie of type k from node i to node j (with k Î R ={1,2, . . . , r}, a fixed set of types of tie)

In most of the statistical literature on network methods, the set N is regarded as fixed and the network ties are assumed to be random. In this case, the tie linking node i to node j may be denoted by the random variable X_{ij} and the n ´ n array X = [X_{ij}] of random variables can be regarded as the adjacency matrix of a random (directed) graph on N. The state space of all possible realizations of these arrays is W_n. The array x = [x_{ij}] denotes a realization of X.

II. LITERATURE REVIEW

A substantial literature dedicated to the analysis of biological networks has emerged in the last few years, and some significant progress has been made on identifying and interpreting the structure of such networks. Due to recent advances in high-throughput technologies, large-scale maps of protein interaction networks [2] [3], metabolic networks [4] and transcriptional regulatory networks [5] have been constructed for a number of simple organisms. A bipartite, topological and clustering graph analysis in order to gain a better understanding of the relationships between human genetic diseases and the relationships between the genes has been reported [6]. Various types of biological data have been used to infer associations between diseases. One of the most commonly used biological data is disease-gene association. Networks have been used to model large-scale biological data, and network topology is beginning to provide insights into diseases and their associations [7] [8]. By considering the inter-connectivity of bio-molecules in the cell, the topology of biological networks is expected to have various biological and clinical applications [9]. Network approaches have been useful in dissecting and providing insight into the underlying mechanism leading to concurrent diseases. Remarkably, analysis of the human metabolic network revealed that connected diseases with metabolic links displayed higher co morbidity than those with no metabolic links [10]. Network approaches have successfully identified biomarkers with clinical applicability. Using a well characterized set of genes, a network approach identified biomarkers for progressive supranuclear palsy [11].Topological network analysis of gene-disease associations uncovers important properties of the nature of mendelian diseases [12].

Data clustering algorithm helps to find groups in data that share a common pattern. It has been used to automatically find clusters in a collection without any user supervision. The main goal of the clustering is to find meaningful groups so that the analysis of all the documents within clusters is much easier compared to viewing it as a whole collection. Some of the most common applications of clustering are in information retrieval, document organization, genetics, weather forecasting, medical imaging, etc [13]. There are different ways to cluster documents. But two common types of clustering methods are used: Partitional and Hierarchical clustering.

III. MATERIALS AND METHODS

A. MALACARDS

MalaCards is an integrated database of human maladies and their annotations, modeled on the architecture and richness of the popular GeneCards database of human genes [14].

The MalaCards disease and disorders database is organized into "disease cards", each integrating prioritized information, and listing numerous known aliases for each disease, along with a variety of annotations, as well as inter-disease connections, empowered by the GeneCards relational database, searches, and GeneAnalytics set-analyses. Annotations include: symptoms, drugs, articles, genes, clinical trials, related diseases/disorders and more. An automatic computational information retrieval engine populates the disease cards, using remote data, as well as information gleaned using the GeneCards platform to compile the disease database. The MalaCards disease database integrates both specialized and general disease lists, including rare diseases, genetic diseases, complex disorders and more (http://www.malacards.org).

MalaCards disease sections are populated by:

[1] Directly interrogating disease resources, to establish integrated disease names, synonyms, summaries, drugs/therapeutics/treatments, clinical features, genetic tests, and anatomical context;

[2] Searching GeneCards for related publications, and for associated genes;
[3] Analyzing disease-associated gene-sets in GeneAnalytics to yield affiliated pathways, phenotypes, compounds, and GO terms;
[4] Searching within MalaCards itself, e.g. for additional related diseases/disorders.

The MalaCards project constitutes an attempt to generate a complete lexicon of all human diseases. The MalaCards naming process provides a capacity to analytically compare disease coverage among different databases. Each disease defined by the naming and unification process is subsequently assigned a hierarchy of disease relatedness layers as follows:
(i) Disease aliases (synonyms)
(ii) Disease families [15]
(iii) Related diseases [16]
(iv) Disease Super Paths [17]

Another form of disease-disease connection shown in the Related Diseases section is co-morbidity. A set of disease co-morbidity relationships (with P<0.01) was obtained from the Phenotypic Disease Network (PDN) [18]. MalaCards employs several different methods to annotate its disease cards:
(i) Direct mining of relevant text from a 'named' target source, i.e. one for which the unification process has generated a relationship between a MalaCards name and the source's disease name. This is exemplified by summaries from Genetic Home Reference or symptoms from Disease Ontology (DO).
(ii) Text mining for the MalaCards name in a target source, followed by mining of the required information, e.g. publications from PubMed, whereby the MalaCards name is matched with in the PubMed title to associate publications with a disease.
(iii) Identifier links connecting a MalaCard to a record target source, followed by information mining, as exemplified by variations from ClinVar.
(iv) Manual curation of specific sections in a target source followed by obtainment of specific annotations. This is done in the case of disease-related drugs obtained from FDA.gov.
(v) Set enrichment analysis via Gene Analytics, by probing the overlap between genes associated with an entity in Gene Cards (e.g. pathways, GO terms and mouse phenotypes) and disease-related genes.

B. 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 Malacards human disease database. The dataset with titles, author names and year were segregated and used as csv file input.

C. NETWORK ANALYSIS

Within the fields of biology and medicine, potential applications of network analysis include for example drug target identification, determining a protein's or gene's function, designing effective strategies for treating various diseases or providing early diagnosis of disorders.

(i) PROTEIN-PROTEIN INTERACTION (PPI) NETWORKS

Mainly hold information of how different proteins operate in coordination with others to enable the biological processes within the cell. Despite the fact that for the majority of proteins the complete sequence is already known, their molecular function is not yet fully determined. Predicting protein function is still a bottleneck in computational biology research and many experimental and computational techniques have been developed in order to infer protein function from interactions with other bio-molecules [19].

IV. RESULTS AND DISCUSSIONS

A. NETWORK ANALYSIS

Further, an undirected network graph was plotted based on terms that appeared in the term matrix with sparsity less than 96% (Figure 1). A graph G can be defined as a pair (V, E) where V is a set of vertices representing the nodes and E is a set of edges representing the connections between the nodes [20]. The degree of a node in an undirected graph is the number of connections or edges the node has to other nodes. The main data structure used to store network graph is given by adjacency matrix. Given a graph $G = (V, E)$ the adjacency matrix representation consists of a $|V|x|V| = nxn$ matrix $A = (a_{ij})$ such that $a_{ij} = 1$ if $(i, j) \in V$ or $a_{ij} = 0$. For undirected graphs the matrix is symmetric. The graph density shows how sparse or dense a graph is according to the number of connections per node set and is defined as:

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

A sparse graph is a graph where $|E| = O(|V|^k)$ and $2 > k > 1$ or otherwise when $|E| \approx |V|$. Dense is a graph where $|E| \approx |V|^2$. It has been reported and mentioned that biological networks are generally sparsely connected, as this confers an evolutionary advantage to preserve robustness. Leclerc RD (2008) [21] reported that the transcriptional regulatory networks of *S. cerevisiae*, *E. coli*, *D. melanogaster* all have connectivity densities lower than 0.1. 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.

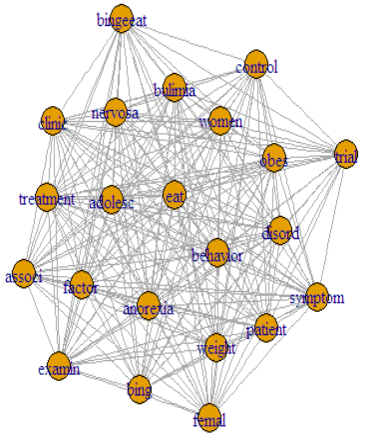

Fig.1: Network analysis graph showing top terms that appeared in term matrix with sparsity less than 96%.

B. CLUSTER ANALYSIS

In this step, a hierarchical and k-means clustering followed by a hybrid hierarchical-k-means (HHK) algorithm was implemented. Before performing cluster analysis on a set of 5-HT receptor bound drugs were extracted from Malacards website and used further as dataset. Table 1 given below identifies the list of 52 drugs that are available in market towards reducing effect of 5-HT receptor activation.

Table 1: List of parameters selected for 5-HT receptor target.

row.names	Mwt	logP	Heavy_ Atoms	HBD	HBA	tPSA	RB	half_life
paroxetine	329.371	3.327	24	1	3	44	4	21
sertraline	306.236	5.18	20	1	0	16	2	24
citalopram	324.399	3.813	24	1	2	37	5	35
Clomipramine	314.86	4.528	22	1	1	7	4	32
Escitalopram	324.399	3.813	24	1	2	37	5	27
Fluoxetine	309.331	4.435	22	1	1	25	6	1
Fluvoxamine	318.339	3.202	22	1	3	58	9	15.6
Cocaine	303.358	1.868	22	1	4	57	3	0.5
Desipramine	266.388	3.533	20	1	1	19	4	7
duloxetine	297.423	4.631	21	1	2	25	6	12
imipramine	280.415	3.875	21	1	1	7	4	16
Methamphetamine	149.237	1.837	11	1	0	16	3	4
Methylphenidate	233.311	2.085	17	1	2	42	3	1
Milnacipran	246.354	1.771	18	1	1	47	5	6
Nortriptyline	263.384	3.826	20	1	0	16	3	16
Phentermine	149.237	1.966	11	1	0	27	2	7
Venlafaxine	277.408	3.036	20	2	2	33	5	5

Vilazodone	441.535	4.03	33	3	4	103	7	25.4
Amoxapine	313.788	3.429	22	1	3	41	0	8
Atomoxetine	255.361	3.725	19	1	1	25	6	5
Desvenlafaxine	263.381	2.733	19	3	2	44	4	10
Dexfenfluramine	231.261	3.246	16	1	0	16	4	32
Doxepin	279.383	3.962	21	1	1	13	3	6
Minaprine	298.39	2.196	22	1	5	50	5	2
Nefazodone	470.017	3.552	33	0	7	55	10	2
Protriptyline	263.384	4.302	20	1	0	16	4	6
Sibutramine	279.855	4.738	19	1	0	4	5	1.1
Tramadol	263.381	2.635	19	2	2	33	4	6.3
Trazodone	371.872	2.362	26	0	6	45	5	3
Trimipramine	294.442	4.121	22	1	1	7	4	11
amitriptyline	277.411	4.169	21	1	0	4	3	10
Mirtazapine	265.36	2.479	20	0	3	19	0	20
Mazindol	284.746	2.609	20	1	3	35	1	10
Pseudoephedrine	165.236	1.328	12	2	1	36	3	9
Vortioxetine	298.455	3.864	21	1	2	19	3	66
Dexmethylphenidate	233.311	2.085	17	1	2	42	3	2
Dextromethorphan	271.404	3.383	20	1	1	13	1	3
Mianserin	264.372	3.084	20	0	2	6	0	10
Amphetamine	135.21	1.576	10	1	0	27	2	10
Dopamine	153.181	0.599	11	3	2	68	2	0.02
Meperidine	247.338	2.213	18	1	2	30	3	3
verapamil	454.611	5.093	33	1	5	65	13	2.8
Loxapine	327.815	3.771	23	0	4	28	0	4
Olanzapine	312.442	1.746	22	1	5	30	0	21
Ondansetron	293.37	3.129	22	0	4	39	2	5.7
Quetiapine	383.517	2.856	27	1	6	48	5	6
Ribavirin	324.186	2.894	21	3	11	195	5	9.5
Phenelzine	136.198	0.692	10	2	2	38	3	1.2
Alitretinoin	300.442	5.603	22	0	2	40	5	2
Tegaserod	301.394	2.815	22	4	2	87	7	11
fenfluramine	231.261	3.246	16	1	0	16	4	20
Amineptine	337.463	4.499	25	1	2	56	8	0.48

C. ASSESSING THE CLUSTERABILITY - HOPKINS STATISTIC

The function get_clust_tendency() in factoextra shall be used to assess whether the dataset can be clustered. This can be achieved by computing *Hopkins statistic. Hopkins statistic* is used to assess the 'clustering tendency' of a dataset by measuring the probability that a given dataset is generated by a uniform data distribution. In other words it tests the 'spatial randomness' of the data.

Hopkins statistic (H) is calculated as the mean nearest neighbor distance in the random dataset divided by the sum of the mean nearest neighbor distances in the real and across the simulated dataset, given by the formula:

$$H = \frac{\sum_{i=1}^{n} y_i}{\sum_{i=1}^{n} x_i + \sum_{i=1}^{n} y_i}$$

A value of H about 0.5 means that $\sum_{i=1}^{n} y_i$ and $\sum_{i=1}^{n} x_i$ are close to each other, and thus the data D is uniformly distributed. If the value of *Hopkins statistic* is close to zero, then we can reject the null hypothesis and conclude that the dataset is significantly a clusterable data.

```
res <- get_clust_tendency(dfNorm, 51, graph = TRUE)
# Hopskin statistic
res$hopkins_stat
res$plot
> res$hopkins_stat
Hopskin statistic: 0.2357666
```

The value of *Hopkins statistic* is significantly < 0.5, indicating that the data is highly clusterable. Additionally, it is observed that the ordered dissimilarity image (Figure 2) contains patterns (i.e., clusters).The ordering of dissimilarity matrix is done using hierarchical clustering.

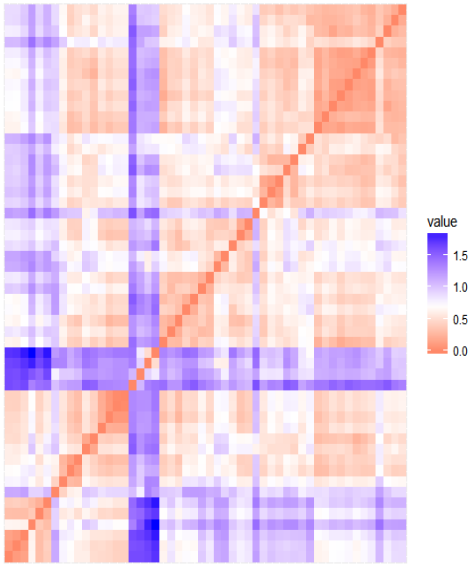

Fig.2: Dissimilarity matrix of the dataset

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-2). 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. 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. A fan type plot is shown in Figure 3.

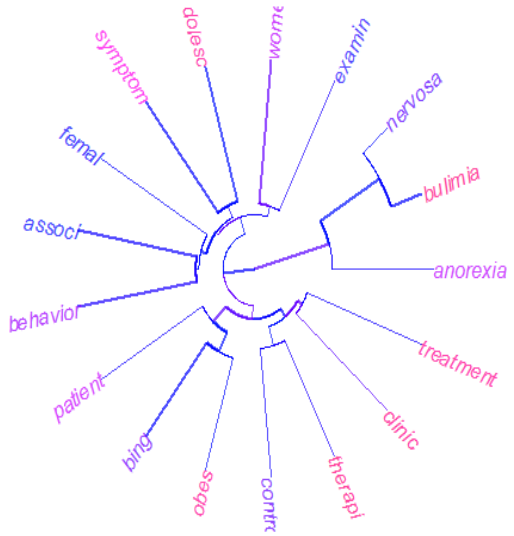

Fig.3: Fan type plot of clustering analysis of frequent terms in matrix.

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.

Table 2: Appearance of 17 terms in document summary represented as term matrix.

	Docs																												
Terms	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	1	1	0	0	0			
Anorexia	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	1	0	0	0	0	0	0	0	0	0	0	
Behaviour	0	0	0	1	0	0	0	1	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	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	1	0	0	0	1		
Control	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	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	
Nervosa	0	0	0	0	0	1	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	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	
Symptom	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	
Therapy	0	0	0	0	0	0	0	0	0	1	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	1	0	0	0	0	1	0	0	
Women	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	

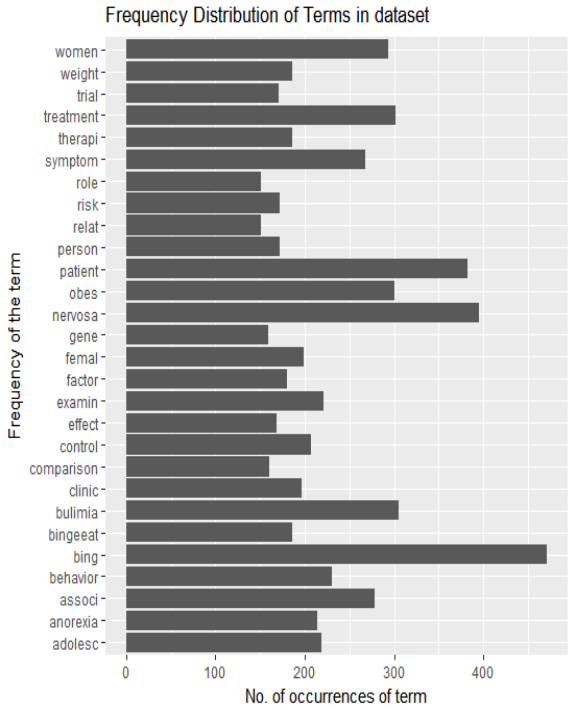

Fig.4: Frequency distribution of terms in the dataset with frequency more than 150.

V. 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 graph theories towards investigation of prime features of a group of objects representing similar nature from a significant cluster. 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 Degreecentrality, Closeness centrality, Eigenvectorc entrality and betweenness centrality resulted in values within the limits.

VI. REFERENCES

[1]. Carrington, P. J., Scott, J., & Wasserman, S. (Eds.). (2003). Models and methods in social network analysis. New York: Cambridge University Press

[2]. I. Xenarios et al. DIP: the database of interacting proteins. Nucleic Acids Research, 28(1):289-291, 2000

[3]. H. Mewes et al. MIPS: a database for genomes and protein sequences. Nucleic Acids Research, 30(1):31-34, 2002

[4]. R. Overbeek et al. Wit: integrated system for high-throughput genome sequence analysis and metabolic reconstruction. Nucleic Acids Research, 28(1):123-125, 2000

[5]. A. Tong et al. Global mapping of the yeast genetic interaction network. Science, 303:808-813, 2004

[6]. Kontou PI et al. Network analysis of genes and their association with diseases. Gene. 2016 Sep 15;590(1):68-78

[7]. Goh KI, Cusick ME, Valle D, Childs B, Vidal M, Barabasi AL: The human disease network. Proc Nat Acad Sci. 2007, 104 (21): 8685-8690

[8]. Milenković T, Memišević V, Bonato A, Pržulj N: Dominating biological networks. PloS one. 2011, 6 (8): e23016

[9]. Ideker T, Sharan R: Protein networks in disease. Genome Res. 2008, 18 (4): 644-652
[10]. Lee DS, Park J, Kay KA, Christakis NA, Oltvai ZN, et al. (2008) The implications of human metabolic network topology for disease comorbidity. Proc Natl Acad Sci U S A 105: 9880-9885
[11]. Santiago JA, Potashkin JA (in press) A Network Approach to Diagnostic Biomarkers in ProgressiveSupranuclear Palsy. MovDisord
[12]. Goh KI, Cusick ME, Valle D, Childs B, Vidal M, Barabasi AL: The human disease network. Proc Nat Acad Sci. 2007, 104 (21): 8685-8690
[13]. Pang-Ning Tan, Michael Steinbach, and Vipin Kumar. Introduction to Data Mining, (First Edition), volume chapter 8. May 2005
[14]. Rappaport et al, MalaCards: an amalgamated human disease compendium with diverse clinical and genetic annotation and structured search, NAR 2017, Vol. 45, Database issue D877–D887
[15]. Amberger,J.S., Bocchini,C.A., Schiettecatte,F., Scott,A.F. and Hamosh,A. (2015) OMIM.org: Online Mendelian Inheritance inMan (OMIM(R)), an online catalog of human genes and geneticdisorders.Nucleic Acids Res.,43, D789–D798
[16]. Ben-Ari Fuchs,S., Lieder,I., Stelzer,G., Mazor,Y., Buzhor,E.,Kaplan,S., Bogoch,Y., Plaschkes,I., Shitrit,A., Rappaport,N.et al.(2016) GeneAnalytics: an integrative gene set analysis tool for nextgeneration sequencing, RNAseq and microarray data.OMICS,20,139–151
[17]. Belinky,F., Nativ,N., Stelzer,G., Zimmerman,S., Iny Stein,T.,Safran,M. and Lancet,D. (2015) PathCards: multi-sourceconsolidation of human biological pathways.Database: J. Biol.Dtabases Curation,2015, bav006
[18]. Hidalgo,C.A., Blumm,N., Barabasi,A.L. and Christakis,N.A. (2009) A dynamic network approach for the study of human phenotypes.PLoS Comput. Biol.,5, e1000353
[19]. Pellegrini Matteo, Haynor David, Johnson JM: Protein interaction networks. Expert Rev Proteomics. 2004, 1 (2):
[20]. Pavlopoulos, Georgios A et al. Using graph theory to analyze biological networks. BioData Mining. 2011, 4:10
[21]. Leclerc RD: Survival of the sparsest: robust gene networks are parsimonious. Mol Syst Biol. 2008, 4: 213