Identification of Receptors for Eating Disorders using Gap Statistic Methods

T. Venkata Sai Krishna
Research Scholar, J N T
University-Kakinada, India

Yesu Babu Adimulam, PhD
Professor and HoD, Department
of CSE, Sir C R Reddy College
of Engineering, Eluru, Andhra
Pradesh, India

R. Kiran Kumar, PhD
Assistant Professor, Department
of Computer Science, KRISHNA
UIVERSITY, Machillipatanam,
Andhra Pradesh, India

ABSTRACT

Eating disorders are an important cause of physical and psychosocial morbidity in adolescent girls and young adult women [1]. Eating disorders (EDs) are classified into three main types, anorexia nervosa (AN), bulimia nervosa (BN), and binge eating disorder. Other rare disorders include, Purging disorder (purging without binge eating) and Night eating syndrome (excessive night time food consumption), Avoidant/Restrictive Food Intake Disorder, Rumination Disorder etc [2]. Eating disorders and certain associated with genetic predisposition and traits. There is a clear and possibly substantial genetic contribution to both anorexia nervosa and bulimia nervosa. Factors that contribute EDs include several risk factors including biological, psychological, social and interpersonal factors.

Network analysis—the study of molecular interactions equates with the mathematical field of graph theory, in which the assembly of pair wise connections (edges) between discrete objects (nodes) coalesces to form a network, or graph [3].

Keywords

Eating Disorders, Bulimia Nervosa, Anorexia Nervosa, Molecular Interactions.

1. INTRODUCTION

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. Advancements in molecular and neurobiology revealed the regulatory mechanism of neurotransmitters, neuropeptides, neurohormones etc acting on the hypothalamus and cortical brain regions influencing intake of food materials, mood variations, response to stress and cognition. Few investigations in this field of eating disorders identified significant role of peripheral signals (serotonin, adipokine leptin and ghrelin) in influencing central nervous system (CNS) processes towards eating patterns.

Eating disorders (EDs) are classified into three main types, anorexia nervosa (AN), bulimia nervosa (BN), and binge eating disorder. Other rare disorders include, Purging disorder (purging without binge eating) and Night eating syndrome (excessive night time food consumption), Avoidant/Restrictive Food Intake Disorder, Rumination Disorder etc [1]. People with anorexia nervosa may see themselves as overweight, even when they are dangerously underweight. People with bulimia nervosa have recurrent and frequent episodes of eating unusually large amounts of food and feeling a lack of control over these episodes. People with binge-eating disorder lose control over his or her eating [5]. Although AN and BN are very similar, people with anorexia are usually very thin and underweight, but those with bulimia may be an average weight or can be overweight. EDs include extreme sensations, attitudes and behaviors with respect to body weight and food issues including stern sensitive and physical problems which might be life-threatening for either females or males.

1.1 Anorexia Nervosa

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. In most instances there is no true anorexia as such. In some patients, the restriction over food intake is also motivated by other psychological processes, including asceticism, competitiveness.

1.2 Bulimia Nervosa

The main feature that distinguishes bulimia nervosa from anorexia nervosa is that attempts to restrict food intake are punctuated by repeated binges (episodes of eating during which there is an aversive sense of loss of control and an unusually large amount of food is eaten).The amount consumed in these binges varies, but is typically between 4·2 MJ (1000 kcals) and 8·4 MJ(2000 kcals) [6].
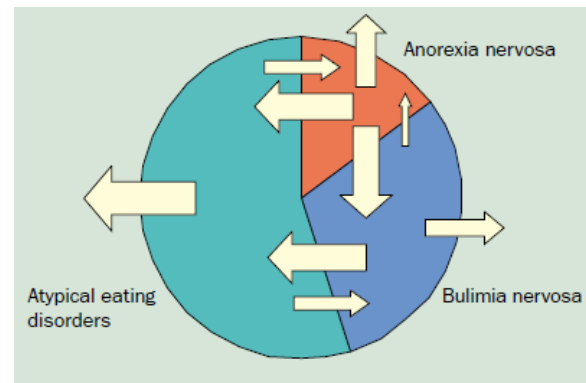

Figure 1: Representation of temporal movement between eating disorders

1.3 Genetics

Eating disorders and certain associated traits run in families [7]. There seems to be cross-transmission between anorexia nervosa, bulimia nervosa, and the atypical eating disorders, suggesting a shared familial liability [8].

There is a clear and possibly substantial genetic contribution to both anorexia nervosa and bulimia nervosa. Molecular genetic studies are being undertaken to identify the underlying loci and genes. Genetic association studies have focused on polymorphisms in serotonin (5-HT)-related genes, because this neurotransmitter system is important in regulation of eating and mood. Particular attention was drawn to the5-HT2AR (HTR2A) gene after an association was reported between allelic variation in the promoter region (-1438 A→G) and anorexia nervosa [9].

Main risk factors for anorexia nervosa and bulimia nervosa [10, 11, 12 and 13].

2. MATERIALS AND METHODS
2.1 K-Means Algorithm
Non hierarchical procedures do not involve the tree-like construction process. Instead, these methods assign objects into clusters once the number of clusters to be formed is specified. The number of clusters may be either be specified in advance or determined as part of the clustering procedure. Non hierarchical methods start from either from (1) an initial partition of items into groups or (2) an initial set of seed points, which will form the nuclei of clusters.

Gap statistic
gap statistic for kmeans
set.seed(123)

gap_stat <- clusGap(dfNorm, FUN = kmeans, K.max = 10, B = 50)

fviz_gap_stat(gap_stat)

gap statistic-PAM
set.seed(123)

gap_stat <- clusGap(dfNorm, FUN = pam, K.max = 10, B = 50)

fviz_gap_stat(gap_stat)

#Gap statistic for hierarchical clustering
set.seed(123)

gap_stat <- clusGap(dfNorm, FUN = hcut, K.max = 10, B = 50)

fviz_gap_stat(gap_stat)

3. RESULTS AND DISCUSSIONS
3.1 GAP Statistics
The gap statistic which can be applied to any clustering methodhas been published by R. Tibshirani et al. [14] The gap statistic compares the total within intracluster variation for different values of k with their expected values under null reference distribution of the data, i.e. a distribution with no obvious clustering.

The clus Gap function from the cluster package calculates a goodness of clustering measure, called the "gap" statistic. For each number of clusters k, it compares (W(k)) with E^*[(W(k))] where the latter is defined via bootstrapping, i.e. simulating from a reference distribution. The reference dataset is generated using Monte Carlo simulations of the sampling process. That is, for each variable ($x_i x_i$) in the data set we compute its range

$[min(x_i),max(x_j) min(x_i),max(x_j)]$ and generate values for the n points uniformly from the interval min to max.

The gap statistic for a given k is defined as follows:

$$Gap_n(k) = E_n^*\{log(W_k)\} - log(W_k)$$

Where E*n denotes the expectation under a sample of size nn from the reference distribution. E*n is defined via bootstrapping (B) by generating B copies of the reference datasets and, by computing the average log(W*k).

The output of gap statistic run is given below for kmeans.

Clustering Gap statistic ["clusGap"] from call:

clusGap(x = dfNorm, FUNcluster = kmeans, K.max = 10, B = 50)

B=50 simulated reference sets, k = 1..10; spaceH0="scaledPCA"

--> Number of clusters (method 'firstSEmax', SE.factor=1): 1

|	logW	E.logW	gap	SE.sim
[1,]	2.227211	2.646645	0.4194344	0.03298790
[2,]	2.075333	2.452160	0.3768280	0.03285850
[3,]	1.924599	2.339005	0.4144064	0.03045909
[4,]	1.816137	2.249723	0.4335862	0.03639677
[5,]	1.768342	2.175939	0.4075974	0.02913742
[6,]	1.639813	2.107224	0.4674101	0.02927924
[7,]	1.563933	2.046449	0.4825156	0.03387348
[8,]	1.480248	1.990852	0.5106043	0.03122305
[9,]	1.425718	1.939939	0.5142213	0.03141202
[10,]	1.325997	1.885395	0.5593979	0.03252470
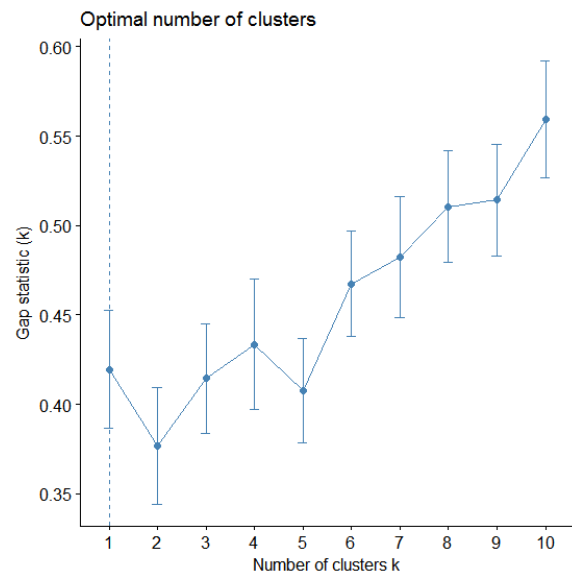
Figure 2: Gap statistic method for kmeans clustering suggesting 1 optimal cluster

Clustering Gap statistic ["clusGap"] from call:

clusGap(x = dfNorm, FUNcluster = pam, K.max = 10, B = 50)

B=50 simulated reference sets, k = 1..10; spaceH0="scaledPCA"

--> Number of clusters (method 'firstSEmax', SE.factor=1): 1

```
       logW   E.logW     gap      SE.sim
 [1,] 2.227211 2.647116 0.4199047 0.03948642
 [2,] 2.081683 2.464945 0.3832618 0.03367742
 [3,] 1.958471 2.354725 0.3962545 0.03552521
 [4,] 1.826088 2.266835 0.4407471 0.04045793
 [5,] 1.743325 2.190883 0.4475588 0.04326315
 [6,] 1.662996 2.124162 0.4611653 0.04200321
 [7,] 1.587480 2.059163 0.4716837 0.04026820
 [8,] 1.499710 2.000077 0.5003668 0.03903657
 [9,] 1.434125 1.946412 0.5122870 0.03993641
[10,] 1.358198 1.893584 0.5353859 0.04126343
```
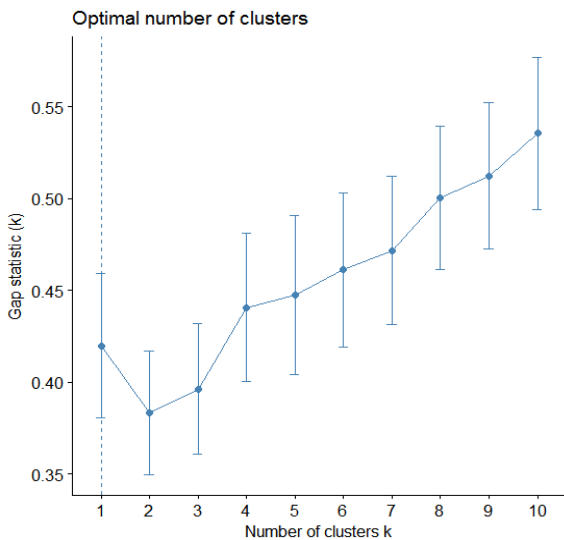

Figure 3: Gap statistic method for PAM clustering suggesting 1 optimal cluster

Clustering Gap statistic ["clusGap"] from call:

clusGap(x = dfNorm, FUNcluster = hcut, K.max = 10, B = 50)

B=50 simulated reference sets, k = 1..10; spaceH0="scaledPCA"

--> Number of clusters (method 'firstSEmax', SE.factor=1): 1

```
       logW   E.logW     gap      SE.sim
 [1,] 2.227211 2.647116 0.4199047 0.03948642
 [2,] 2.073187 2.458218 0.3850313 0.03389733
 [3,] 1.949309 2.341829 0.3925196 0.03882125
 [4,] 1.828577 2.253156 0.4245791 0.03883815
 [5,] 1.764041 2.176247 0.4122063 0.03855501
 [6,] 1.660960 2.107259 0.4462989 0.03805856
 [7,] 1.573579 2.044163 0.4705833 0.03780469
 [8,] 1.498186 1.983679 0.4854929 0.03799201
 [9,] 1.418882 1.928031 0.5091487 0.03811344
[10,] 1.355687 1.874133 0.5184456 0.03853961
```
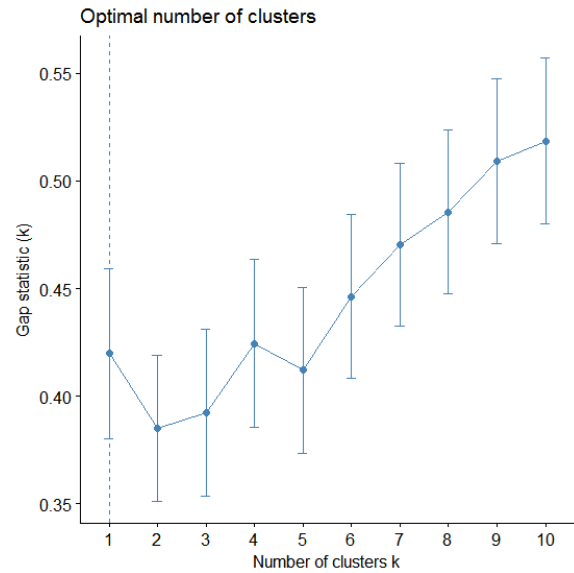

Figure 4: Gap statistic method for hierarchical clustering suggesting 1 optimal cluster

Finally from gap statistic, only one cluster solution is suggested. It should be noted that each method displayed different clusters. Table 8 given below compares the methods employed.

Table 1: Comparative data on optimal clusters generated by elbow, silhouette and gap statistic methods

Method	Number of optimal clusters		
	kmeans	PAM	Hierarchical
Elbow	3	3	3
Silhouette	10	7	2
Gap statistic	1	1	1

4. CONCLUSION

Scientific investigations in the field of eating disorders identified significant role of peripheral signals in influencing central nervous system (CNS) processes towards eating patterns. Eating disorders (EDs) are classified as, anorexia nervosa, bulimia nervosa and binge eating disorder. For the better clusters, determination of optimal number of clusters was carried out using gap statistic. The optimal number of clusters, k for the 5-HT receptor dataset was found to have 3 cluster solutions proposed by 7 indices. Therefore, initial value of k=3 was used to perform k-means, hierarchical followed by hybrid HK means algorithm on the dataset.

5. REFERENCES

[1] Li Y, Agarwal P (2009) A Pathway-Based View of Human Diseases and Disease Relationships. PLoS ONE 4(2): e4346.doi:10.1371/journal.pone.0004346.

[2] American Psychiatric Association's Fifth Edition of the Diagnostic and Statistical Manual of Mental Disorders (DSM-5), published in 2013 ISBN 978-0-89042-554-1.

[3] Barabasi, A.L. & Albert, R. Emergence of scaling in random networks. Science 286, 509–512 (1999).

[4] American Psychiatric Association's Fifth Edition of the Diagnostic and Statistical Manual of Mental Disorders (DSM-5), published in 2013 ISBN 978-0-89042-554-1.

[5] https://www.nimh.nih.gov/health/topics/eating-disorders

[6] Rosen JC, Leitenberg H, Fisher C, Khazam C. Binge-eating episodes in bulimia nervosa: the amount and type of food consumed. Int J Eat Disord 1986; 5: 255–67.

[7] Lilenfeld LR, Kaye WH. Genetic studies of anorexia and bulimia nervosa. In: Hoek HW, Treasure JL, Katzman MA, eds. Neurobiology in the treatment of eating disorders. Chichester: Wiley, 1998: 169–94.

[8] Strober M, Freeman R, Lampert C, Diamond J, Kaye W. Controlled family study of anorexia nervosa and bulimia nervosa: evidence of shared liability and transmission of partial syndromes. Am J Psychiatry 2000; 157: 393–401.

[9] Collier DA, Arranz MJ, Li T, et al. Association between 5-HT2A gene promoter polymorphism and anorexia nervosa. Lancet 1997; 350: 412.

[10] Connors ME. Developmental vulnerabilities for eating disorders. In: Smolak L, Levine MP, Striegel-Moore R, eds. The developmental psychopathology of eating disorders: implications for research, prevention and treatment. New Jersey: Lawrence Erlbaum, 1996:285–310.

[11] Fairburn CG, Welch SL, Doll HA, Davies BA, OConnor ME. Risk factors for bulimia nervosa: a community-based case-control study. Arch Gen Psychiatry 1997; 54: 509–17.

[12] Fairburn CG, Cooper Z, Doll HA, Welch SL. Risk factors for anorexia nervosa: three integrated case-control comparisons.Arch Gen Psychiatry 1999; 56: 468–76.

[13] Fairburn CG, Doll HA, Welch SL, et al. Risk factors for binge eating disorder: a community-based case-control study. Arch Gen Psychiatry 1998; 55: 425–32.

[14] Tibshirani, R., Walther, G. and Hastie, T. (2001). Estimating the number of data clusters via the Gap statistic. Journal of the Royal Statistical Society B, 63, 411–423.