WEB ANALYTICS

HUMAN DISEASE-SYMPTOM NETWORK

Saumya Agarwal
Sun Qianqian
Wang Yanhang
Ye Jialiang

CONTENTS

01 Data Pre-processing

02 Community Detection

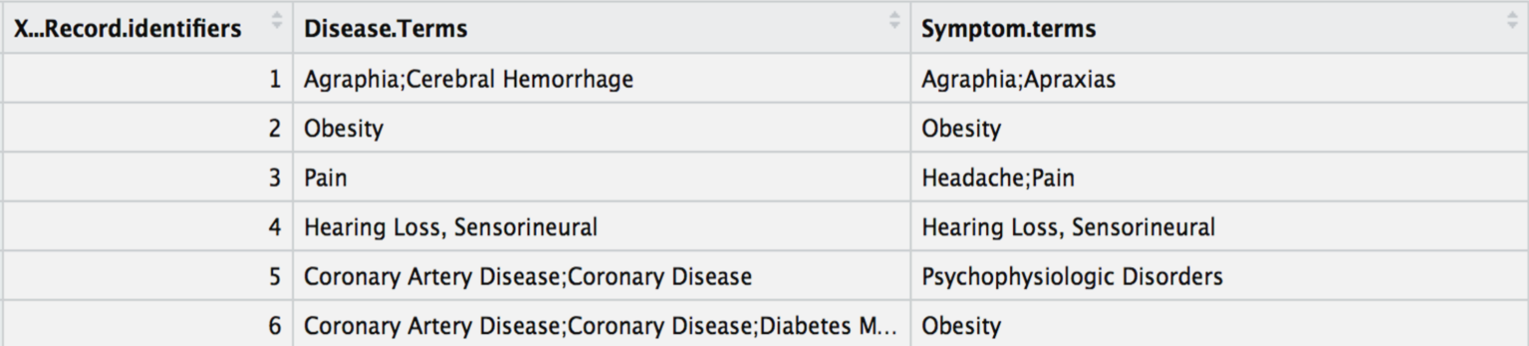
03 Keyplayer Detection

04 Visualization & Insights

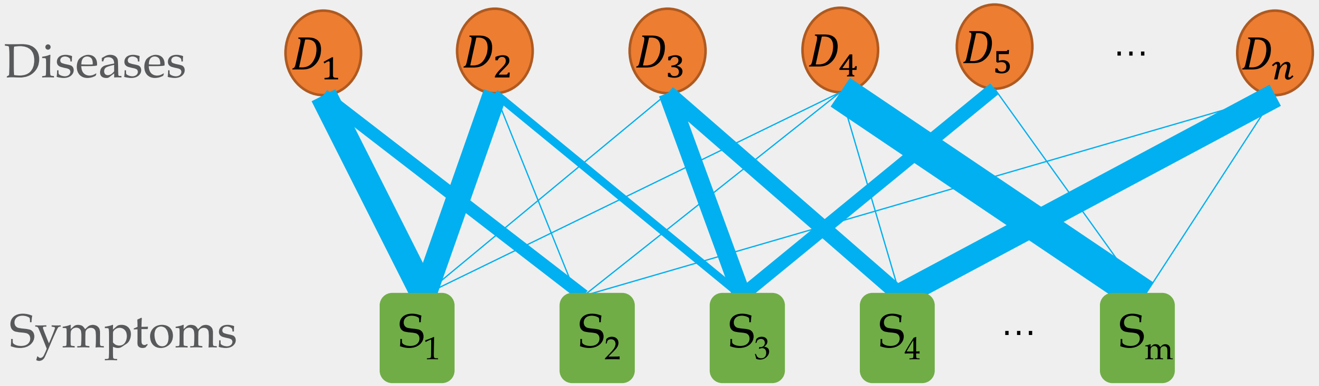
Data Pre-processing 01

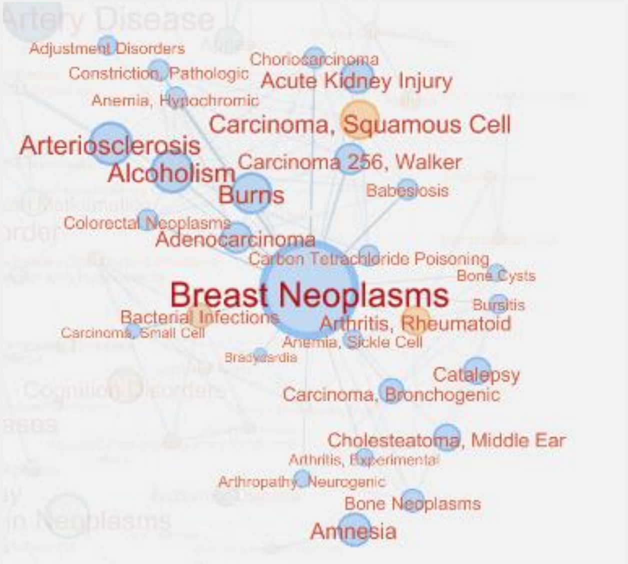
Introduction

Disease and Symptom



X...Record.identifiers	Disease.Terms	Symptom.terms
1	Agraphia;Cerebral Hemorrhage	Agraphia;Apraxias
2	Obesity	Obesity
3	Pain	Headache;Pain
4	Hearing Loss, Sensorineural	Hearing Loss, Sensorineural
5	Coronary Artery Disease;Coronary Disease	Psychophysiologic Disorders
6	Coronary Artery Disease;Coronary Disease;Diabetes M...	Obesity





Data Pre-processing

Record ID	Diseases	Symptoms
1	Agraphia; Cerebral Hemorrhage	Agraphia; Apraxias
2	Coronary Artery Disease; Coronary Disease; Diabetes Mellitus, Type 2	Obesity
3	Neuralgia	Facial Pain; Low Back Pain; Neuralgia

```
dis<-cSplit(disease,"Disease.Terms",";",direction = "long")

dis<-cSplit(dis,"Symptom.terms",";",direction = "long")
```

Split Disease and Symptom

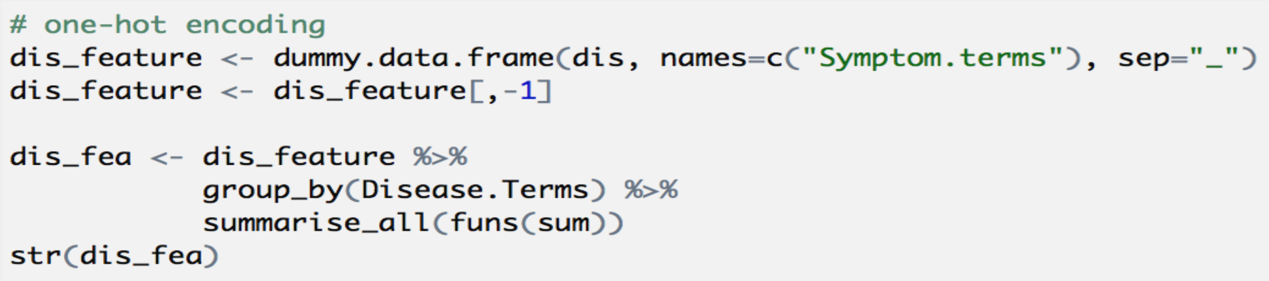
Record ID	Diseases	Symptoms
1	D1,D2	S1
2	D3	S1,S3

↓

Record ID	Diseases	Symptoms
1	D1	S1
1	D2	S1
2	D3	S1
2	D3	S3

One-hot Encoding

Record ID	Diseases	Symptoms
1	D1	S1
1	D2	S1
2	D3	S1
2	D3	S3

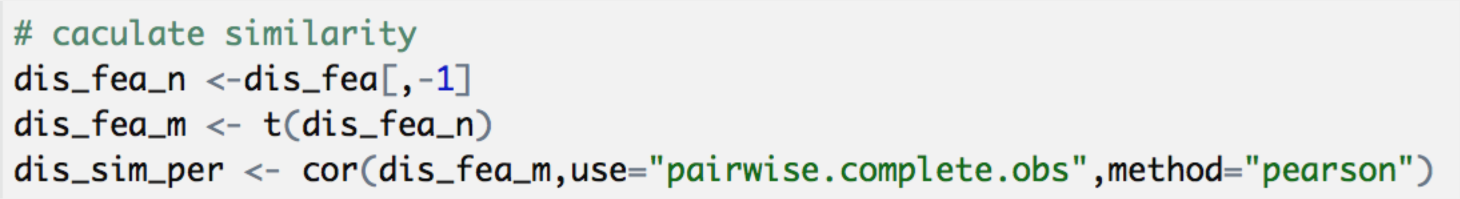


```
# one-hot encoding
dis_feature <- dummy.data.frame(dis, names=c("Symptom.terms"), sep="_")
dis_feature <- dis_feature[,-1]

dis_fea <- dis_feature %>%
           group_by(Disease.Terms) %>%
           summarise_all(funs(sum))
str(dis_fea)
```

Disease	Symptoms_S1	Symptoms_S2	Symptoms_S3
D1	1	0	0
D2	1	0	0
D3	1	0	1

Calculate the Similarity

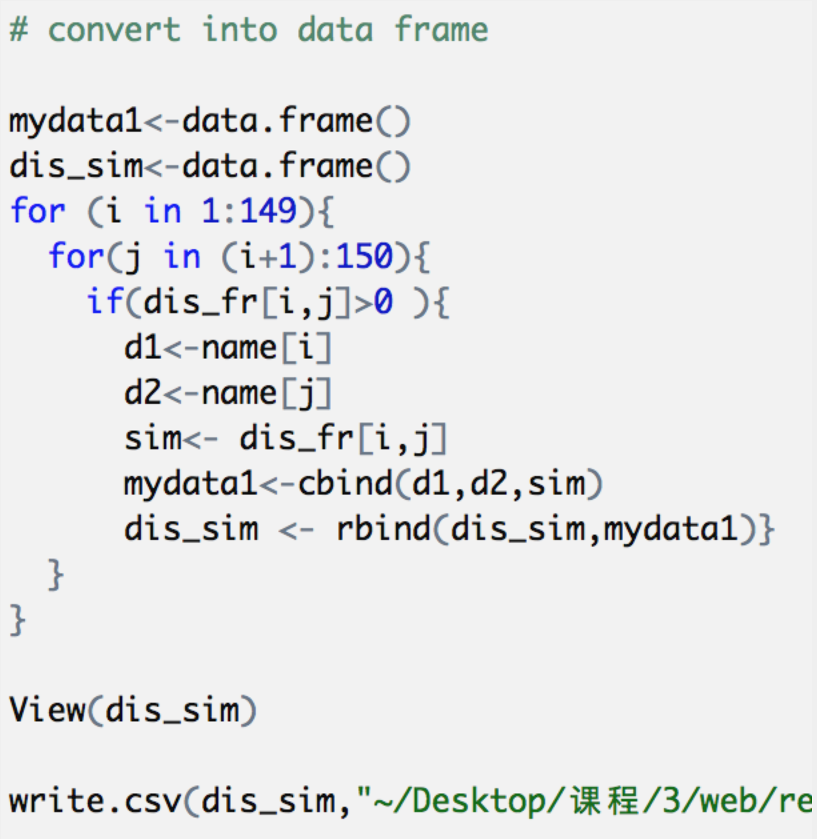
Disease	Symptoms_S1	Symptoms_S2	Symptoms_S3
D1	1	0	0
D2	1	0	0
D3	1	0	1



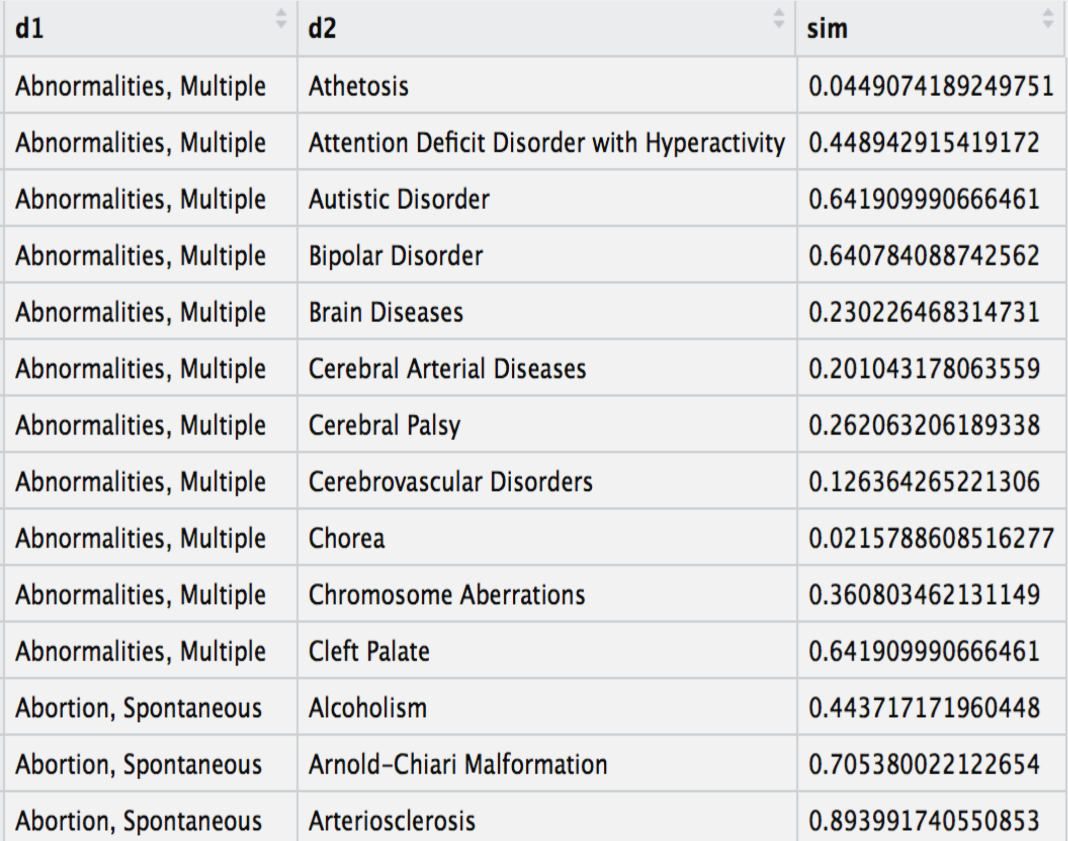
```
# caculate similarity
dis_fea_n <-dis_fea[,-1]
dis_fea_m <- t(dis_fea_n)
dis_sim_per <- cor(dis_fea_m,use="pairwise.complete.obs",method="pearson")
```

	D1	D2	D3
D1	1	0.006915490	-0.007144623
D2	-0.006915490	1	0.372935908
D3	-0.007144623	0.372935908	1

Re written as a standard form



```
# convert into data frame

mydata1<-data.frame()
dis_sim<-data.frame()
for (i in 1:149){
  for(j in (i+1):150){
    if(dis_fr[i,j]>0 ){
      d1<-name[i]
      d2<-name[j]
      sim<- dis_fr[i,j]
      mydata1<-cbind(d1,d2,sim)
      dis_sim <- rbind(dis_sim,mydata1)}
  }
}

View(dis_sim)

write.csv(dis_sim,"~/Desktop/课程/3/web/re
```



d1	d2	sim
Abnormalities, Multiple	Athetosis	0.0449074189249751
Abnormalities, Multiple	Attention Deficit Disorder with Hyperactivity	0.448942915419172
Abnormalities, Multiple	Autistic Disorder	0.641909990666461
Abnormalities, Multiple	Bipolar Disorder	0.640784088742562
Abnormalities, Multiple	Brain Diseases	0.230226468314731
Abnormalities, Multiple	Cerebral Arterial Diseases	0.201043178063559
Abnormalities, Multiple	Cerebral Palsy	0.262063206189338
Abnormalities, Multiple	Cerebrovascular Disorders	0.126364265221306
Abnormalities, Multiple	Chorea	0.0215788608516277
Abnormalities, Multiple	Chromosome Aberrations	0.360803462131149
Abnormalities, Multiple	Cleft Palate	0.641909990666461
Abortion, Spontaneous	Alcoholism	0.443717171960448
Abortion, Spontaneous	Arnold-Chiari Malformation	0.705380022122654
Abortion, Spontaneous	Arteriosclerosis	0.893991740550853

Community Detection 02

Community Detection

Vertex Count: 127 + 4 unconnected

Transitivity: 0.4602273

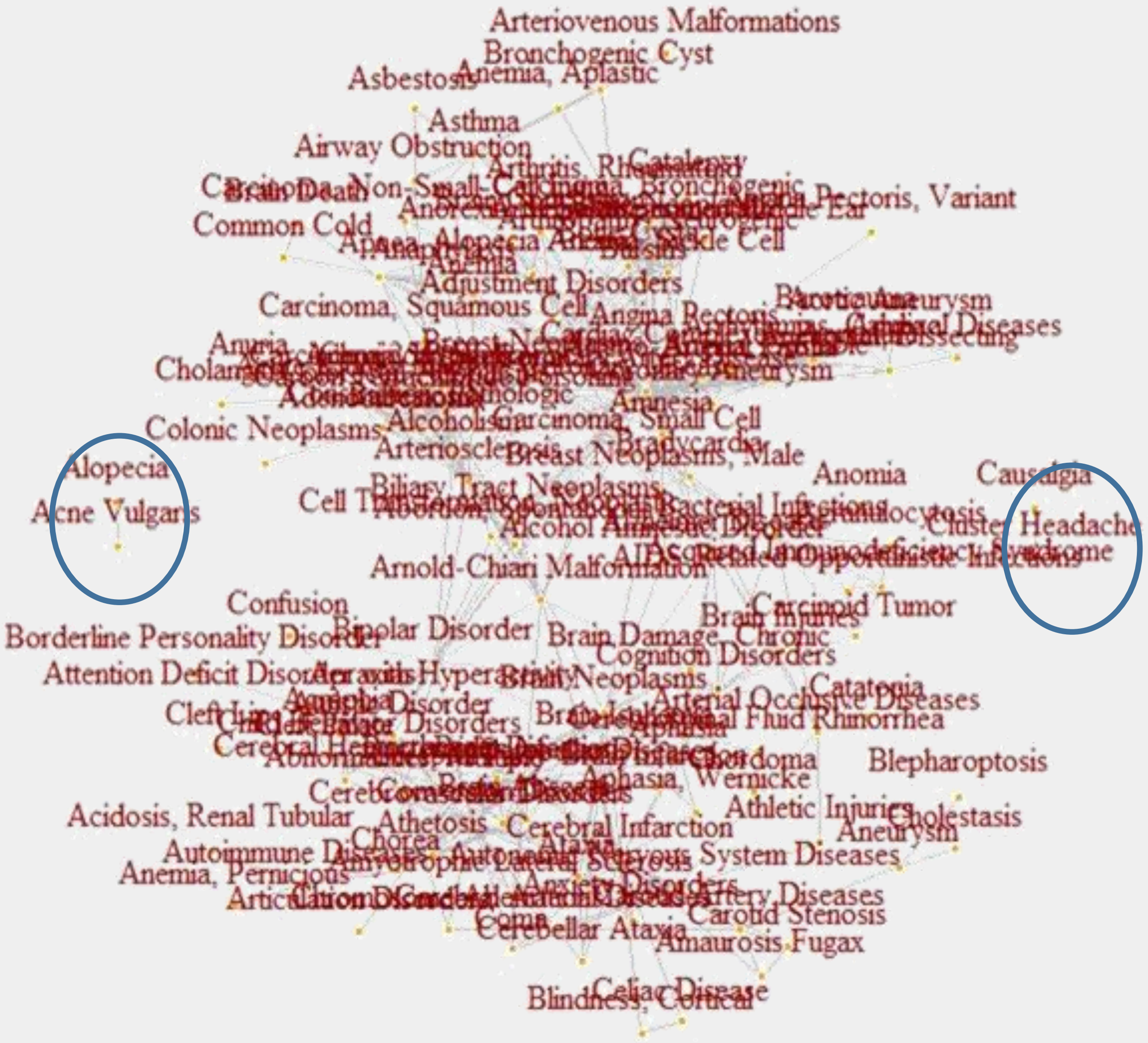
Density: 5.961755

Diameter: 2.188508

Clique Number: 9

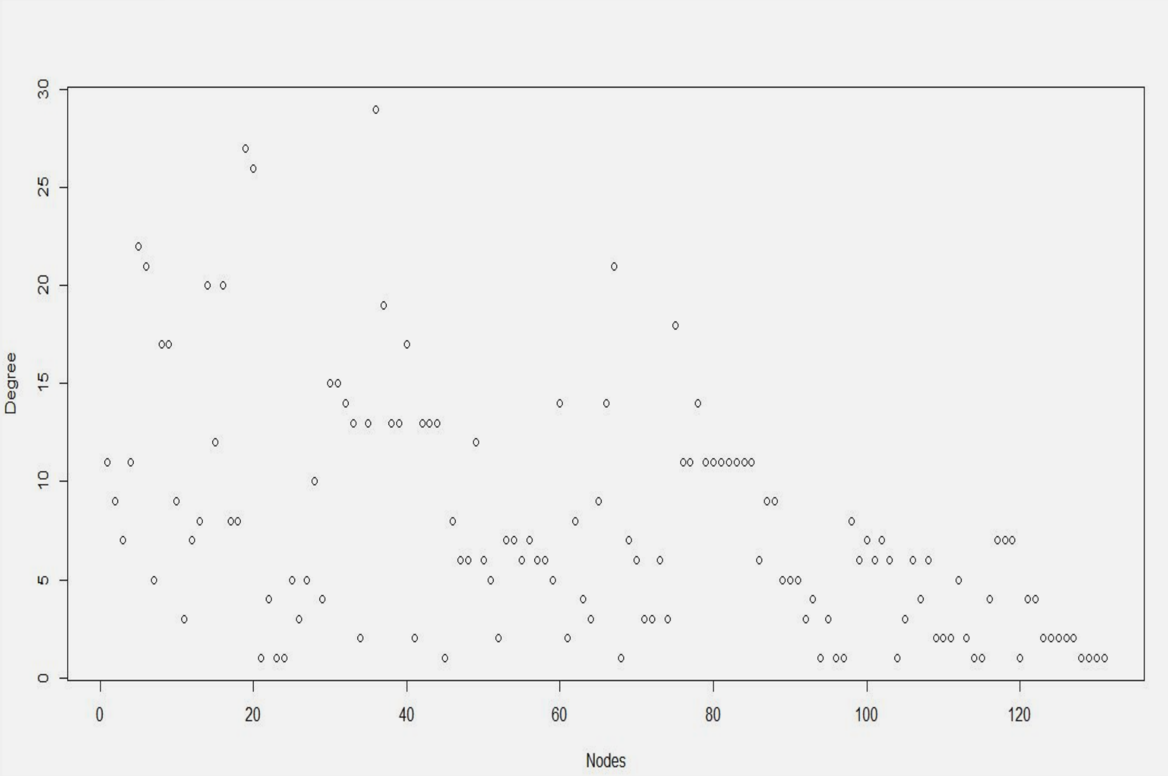
Vertex with Maximum Betweenness value: **Coronary Artery Disease**

Vertex with Maximum Closeness centrality: **Coronary Artery Disease**

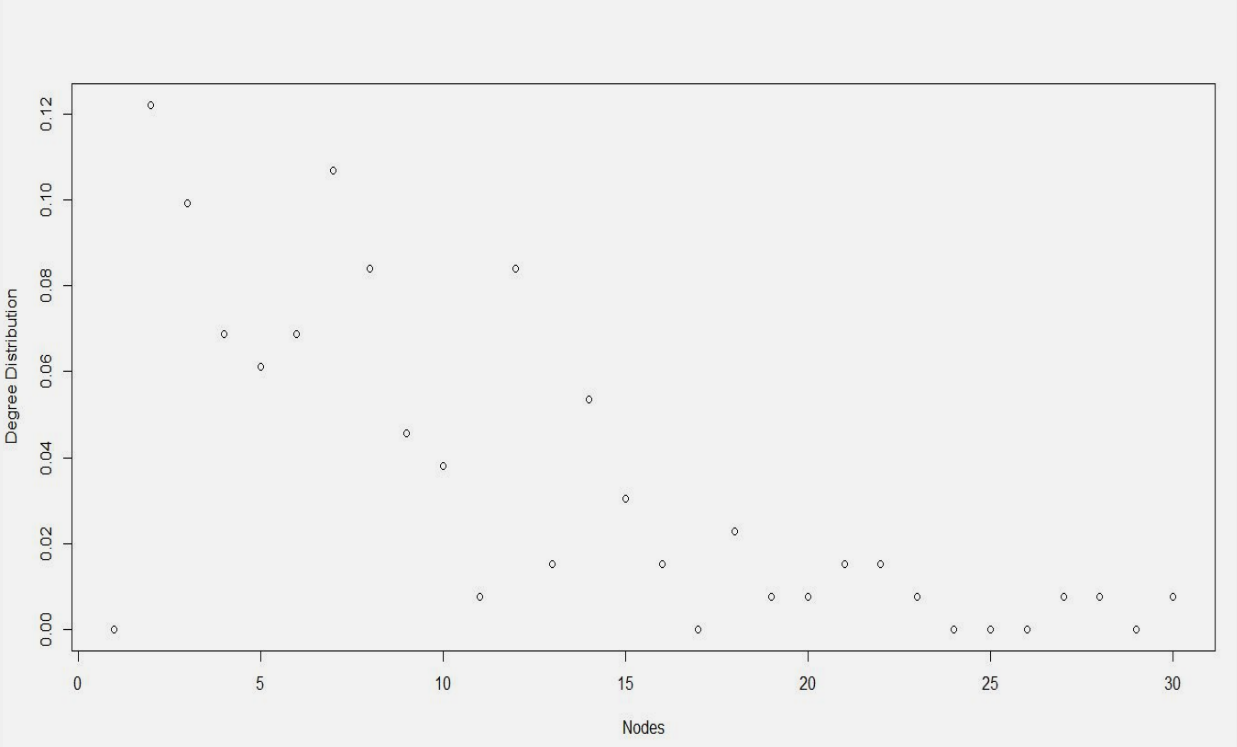


Basic Concepts of Graph

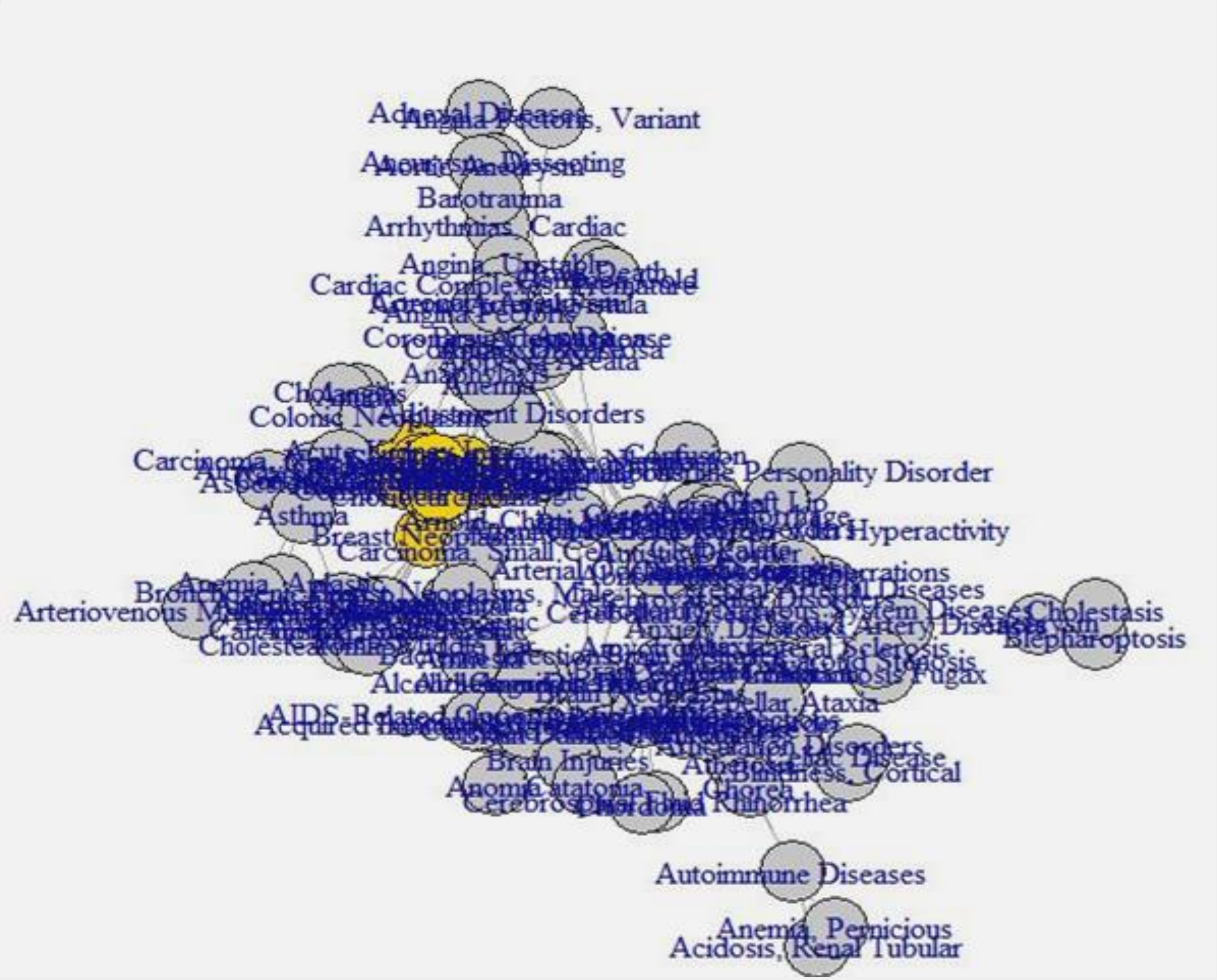
Degree Distribution: Given a network graph G, we define f_d to be the fraction of vertices v ∈ V with degree dv = d. The collection { fd}d≥0 is called the degree distribution of G, and is simply a rescaling of the set of degree frequencies



Degree: The degree dv of a vertex v, in a network graph G = (V,E), counts the number of edges in E incident upon v.



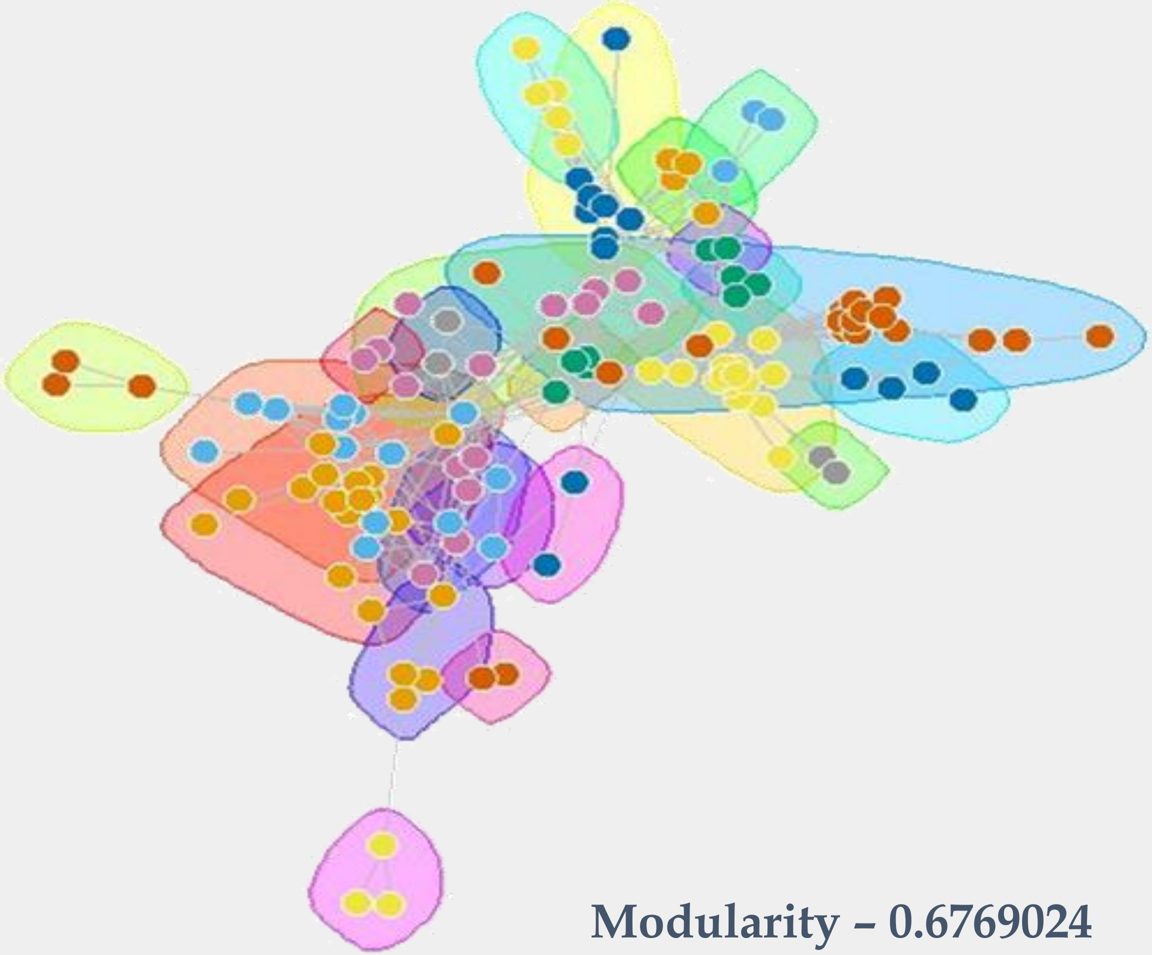
Clique-Based Community Detection



```
cliques(g, min=9, max=9)
largest.cliques(g)
```

A clique, C, in an undirected graph G = (V, E) is a subset of the vertices, C ⊆ V, such that every two distinct vertices are adjacent. This is equivalent to the condition that the induced subgraph of G induced by C is a complete graph.

Label propagation community Detection

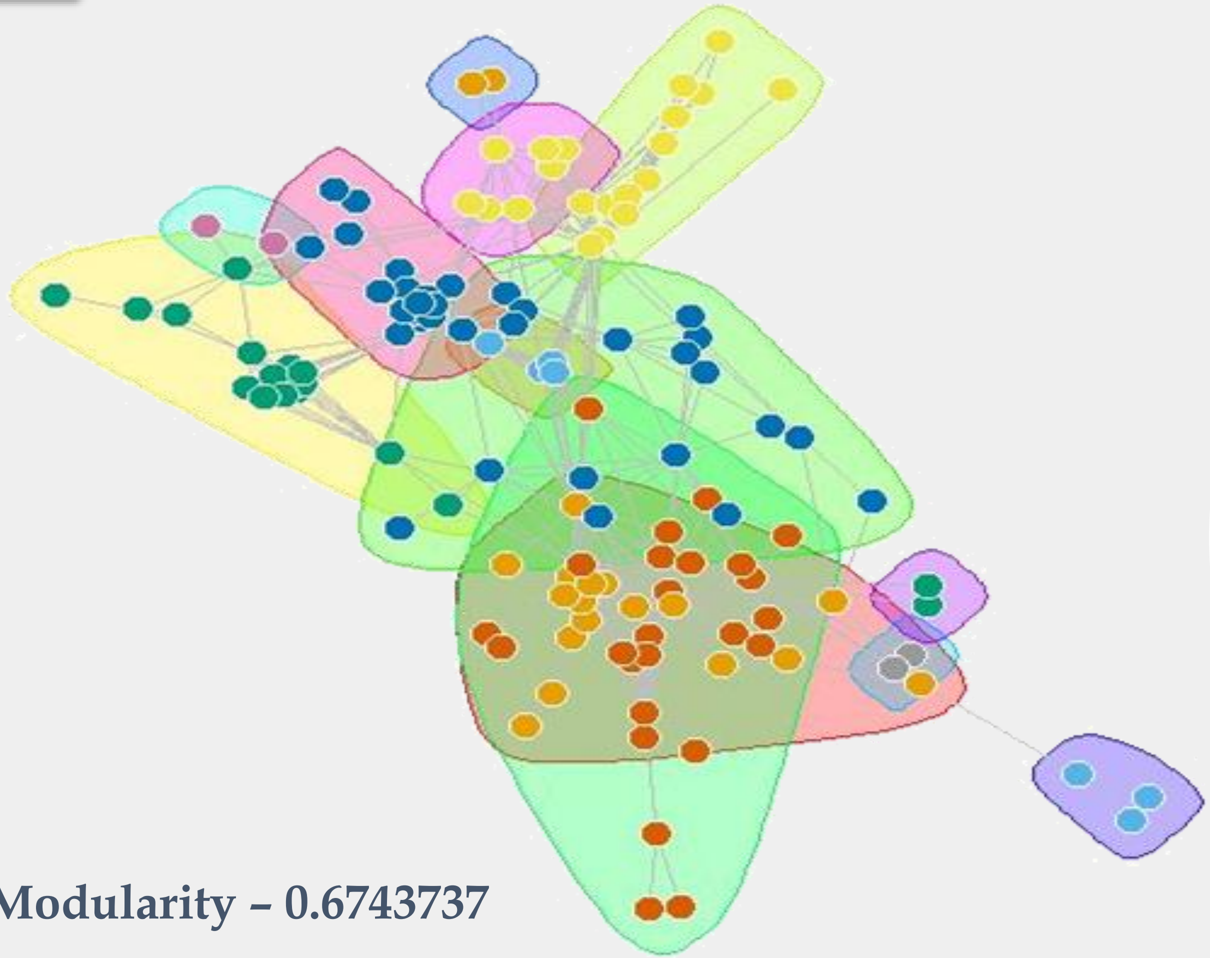


Modularity – 0.6769024

lab <- label.propagation.community(g)
modularity(lab)

The label propagation algorithm uses an iterative process to find stable communities in a graph. The method begins by giving each node in the graph a unique label. Then, the algorithm iteratively simulates a process in which each node in the graph adopts the label most common amongst its neighbours. The process repeats until the label of every node in the graph is the same as the label of maximum occurrence amongst its neighbours.

Leading eigenvector community Detection

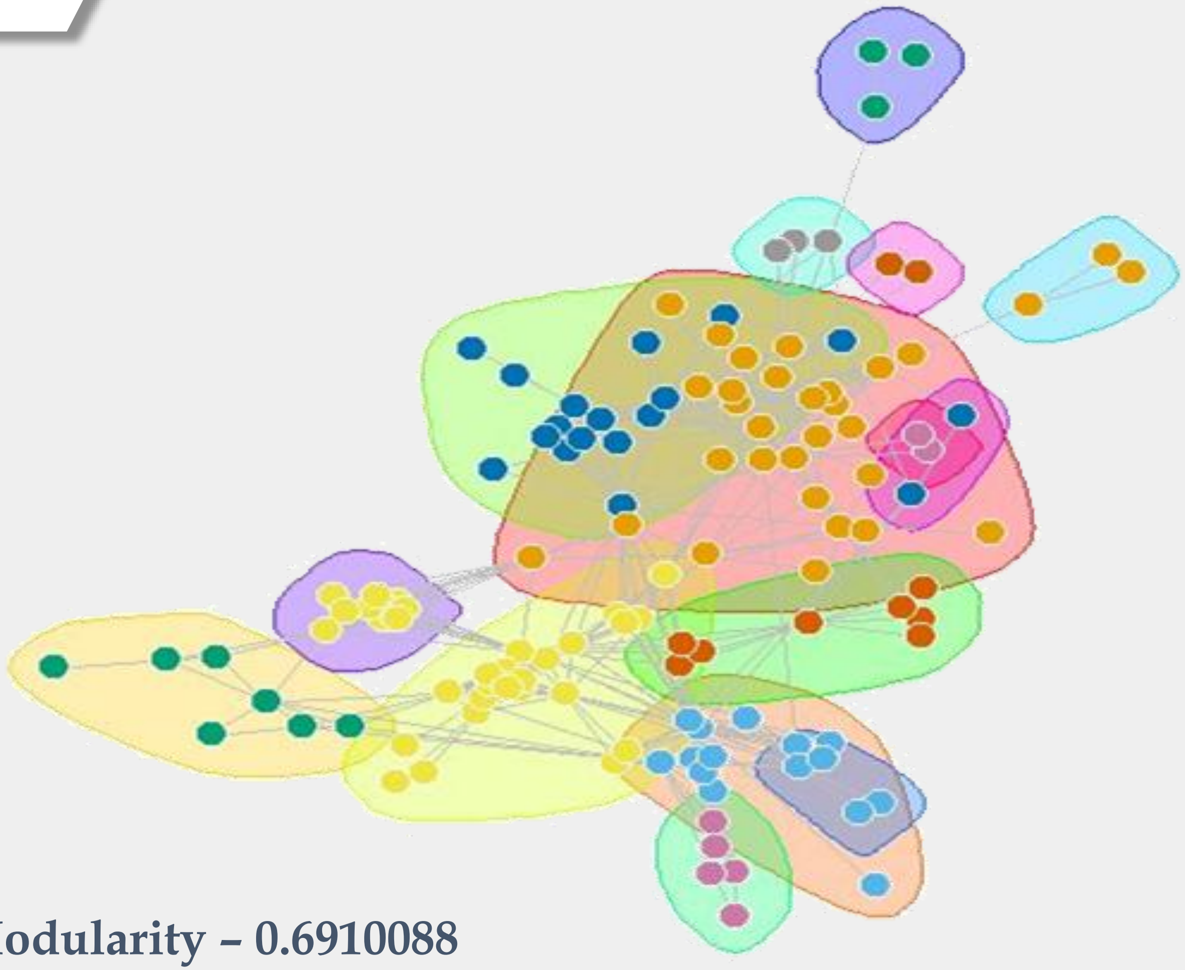


Modularity – 0.6743737

The leading eigenvector method works by calculating the eigenvector of the modularity matrix for the largest positive eigenvalue and then separating vertices into two community based on the sign of the corresponding element in the eigenvector. If all elements in the eigenvector are of the same sign that means that the network has no underlying community structure.

lec <- leading.eigenvector.community(g,options=list(maxiter=1000000))
modularity(lec)

Walktrap Community Detection

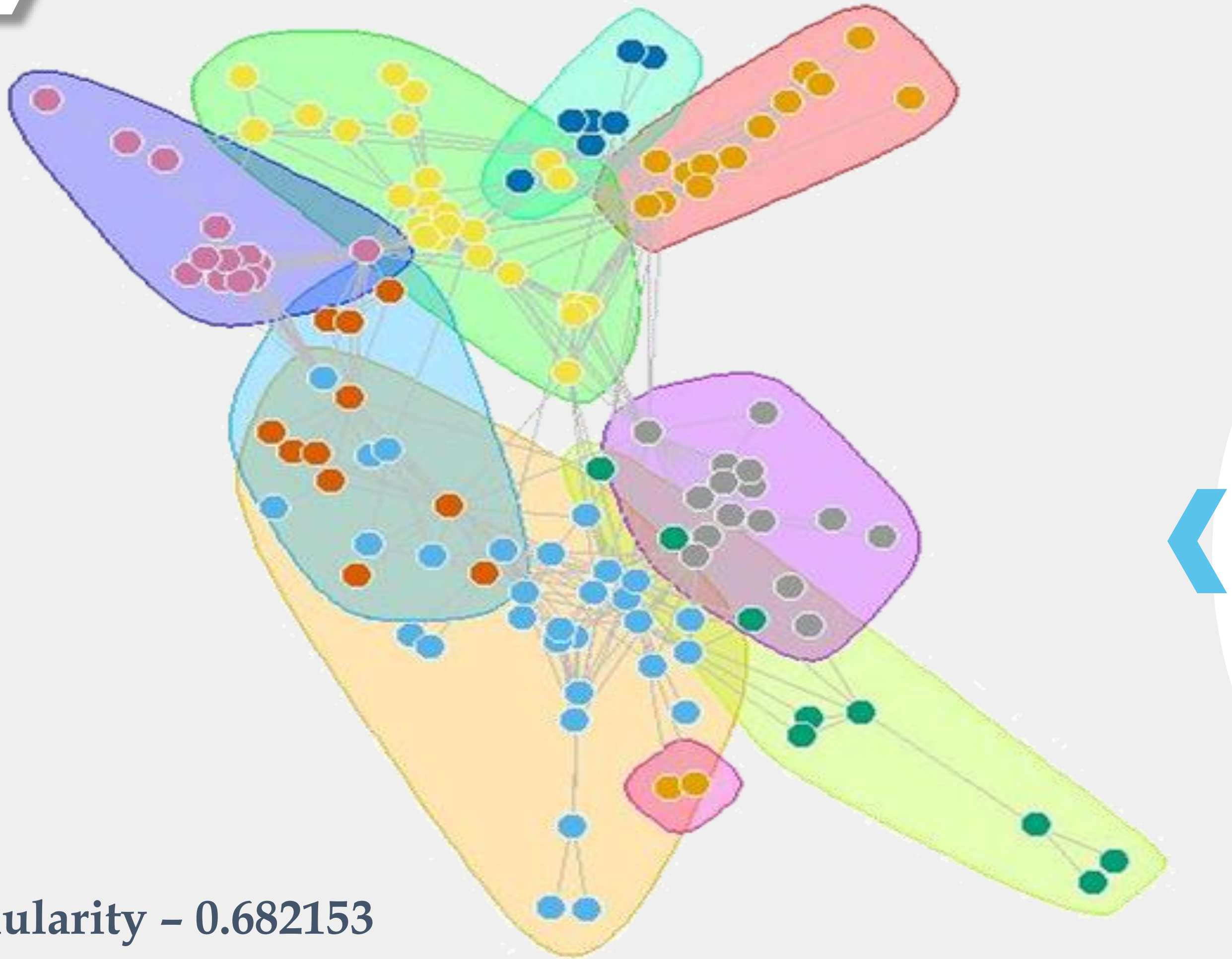


Modularity – 0.6910088

wc <- walktrap.community(g)
modularity(wc)

This algorithm finds densely connected subgraphs by performing random walks. The idea is that random walks will tend to stay inside communities instead of jumping to other communities.

Fastgreedy community Detection

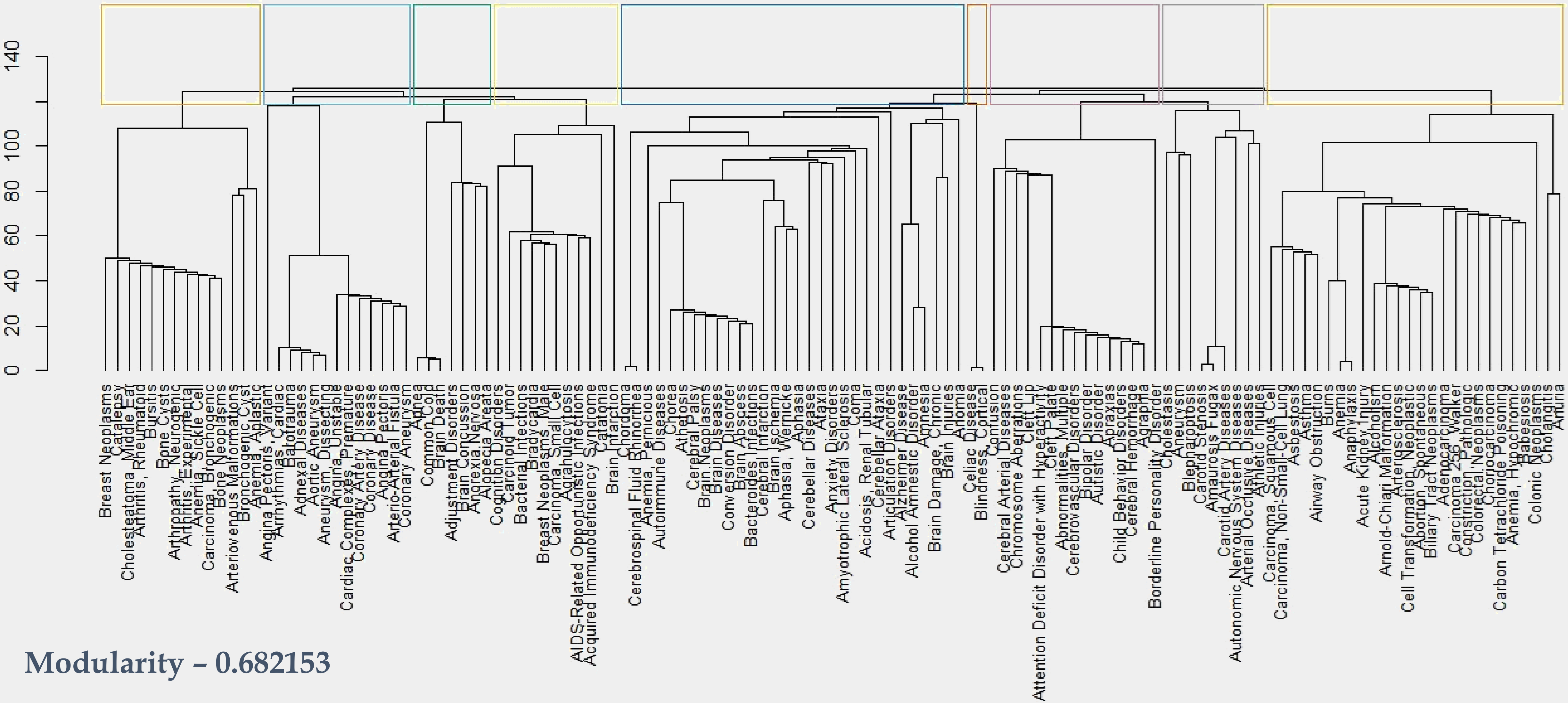


Modularity – 0.682153

g1 <- simplify(g)
fc <- fastgreedy.community(g1)

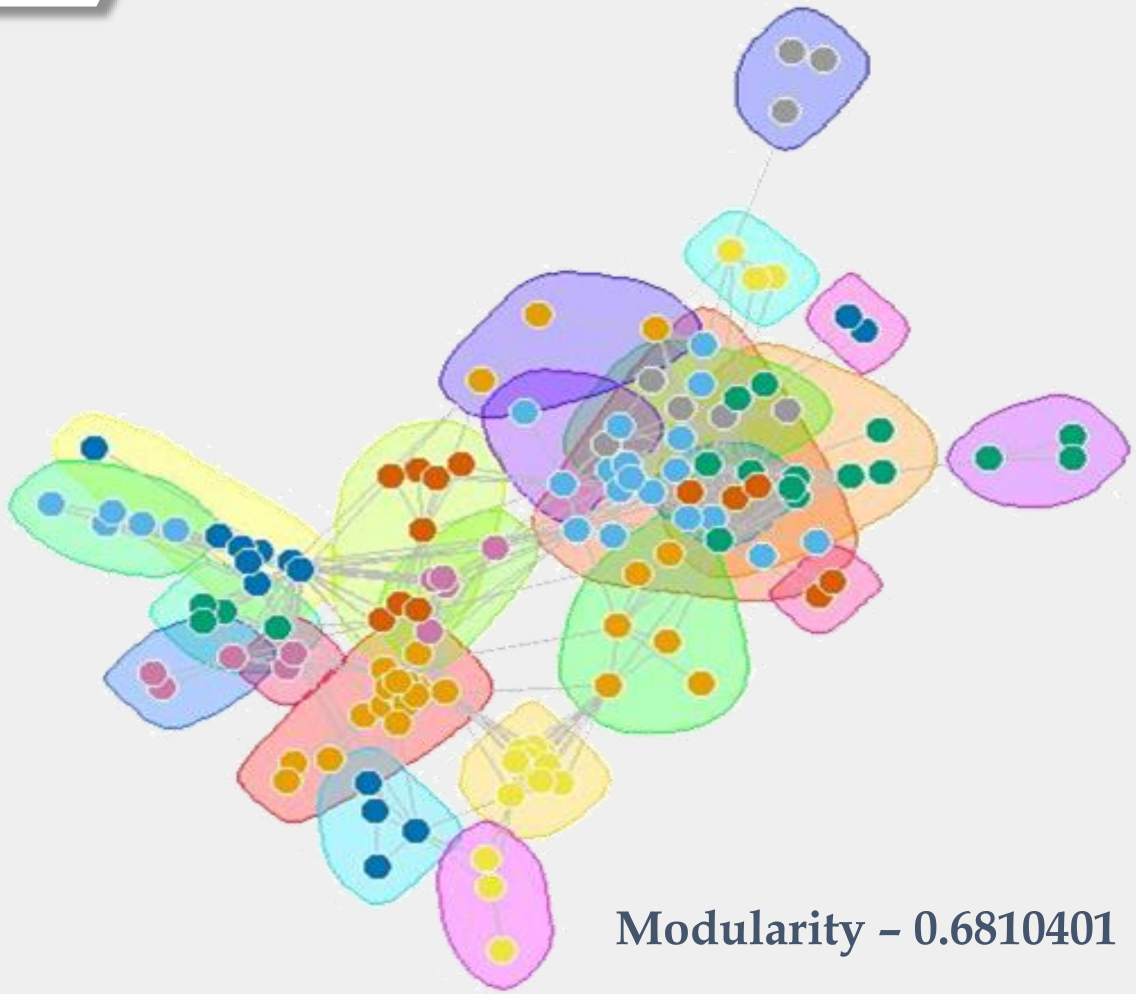
In this case the algorithm is agglomerative. At each step two groups merge. The merging is decided by optimising modularity. This is a fast algorithm, but has the disadvantage of being a greedy algorithm. Thus, is might not produce the best overall community partitioning.

Fastgreedy community Detection



Modularity – 0.682153

Infomap Community Detection

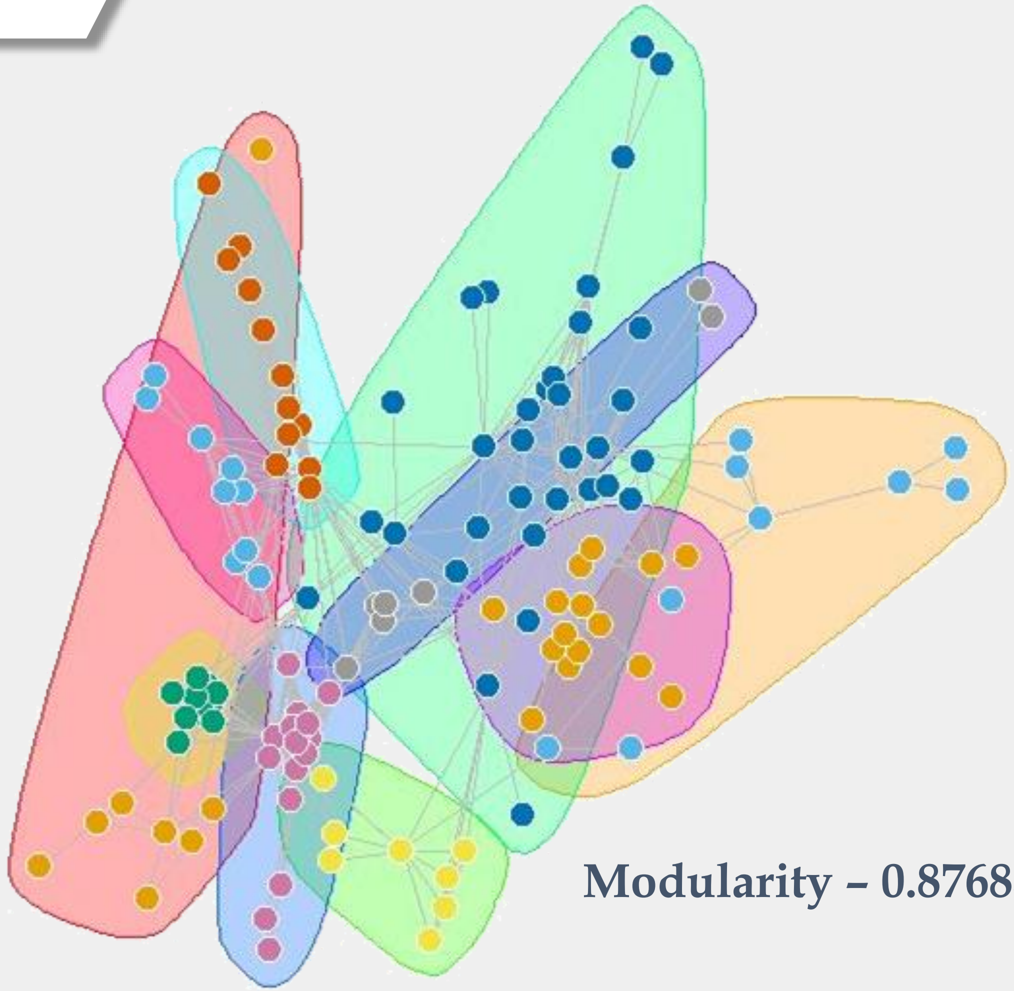


Modularity – 0.6810401

ic <- infomap.community(g1)
modularity(ic)

The Infomap algorithm is based on the principles of information theory. Infomap characterizes the problem of finding the optimal clustering of a graph as the problem of finding a description of minimum information of a random walk on the graph. The algorithm maximizes an objective function called the Minimum Description Length, and in practice an acceptable approximation to the optimal solution can be found quickly.

Spinglass community Detection



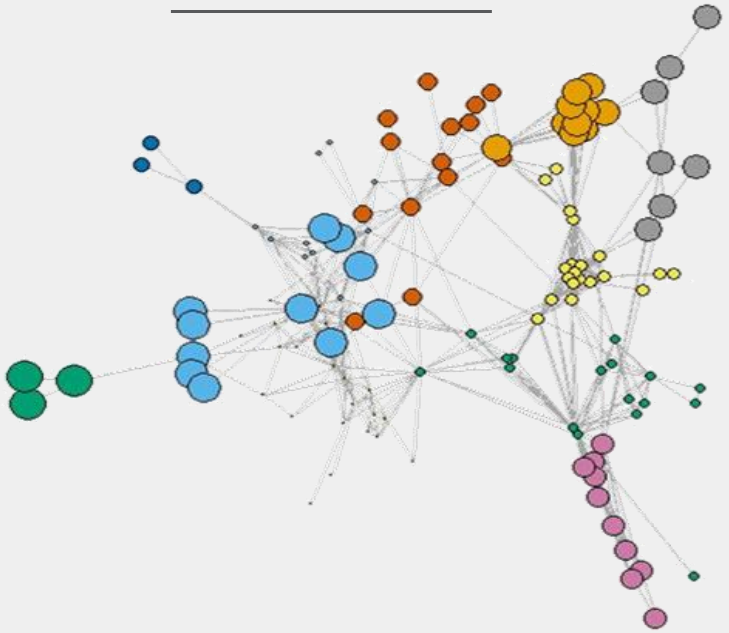
Modularity – 0.8768559

set.seed(1234)
sc <- spinglass.community(g, spins=10)

This algorithm uses as spin-glass model and simulated annealing to find the communities inside a network. The community structure of the network is interpreted as the spin configuration that minimizes the energy of the spin glass with the spin states being the community indices. We elucidate the properties of the ground state configuration to give a concise definition of communities as cohesive subgroups in networks that is adaptive to the specific class of network under study.

Comparison of Different Algorithms

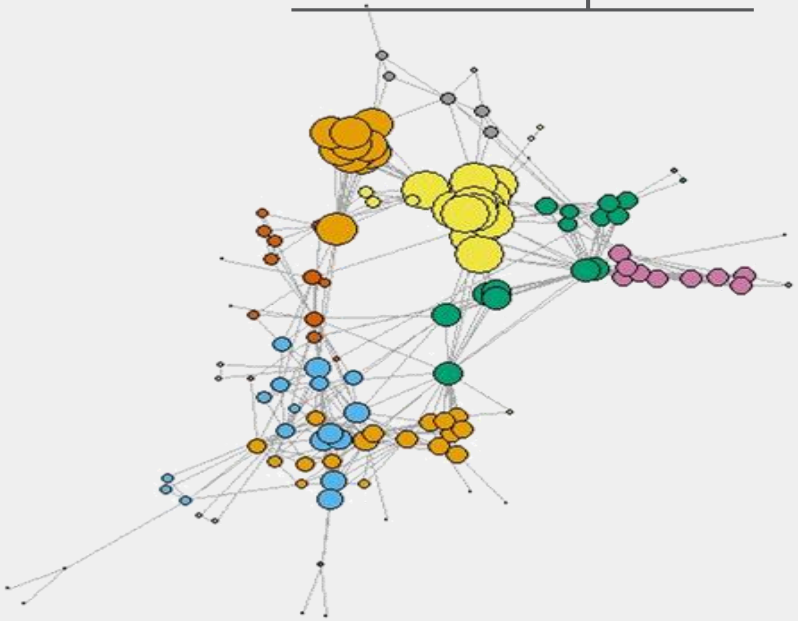
Girvan-Newman



Modularity – 0.5729917

```
mods <- sapply(0:ecount(g), function(i){
  g2 <- delete.edges(g,
ebc$removed.edges[seq(length=i)])
  cl <- clusters(g2)$membership
```

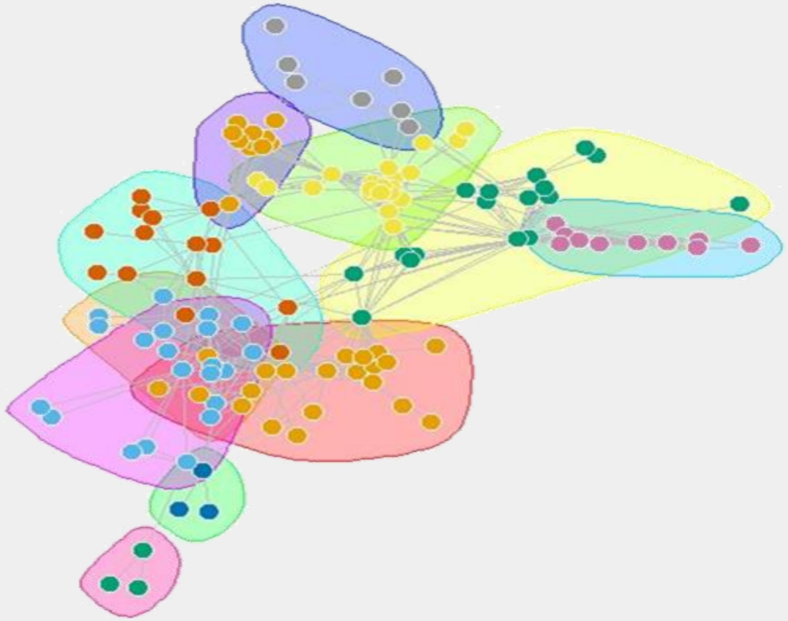
K-Core Decomposition



Modularity – 0.4544663

```
kc <- coreness(g, mode="all")
modularity(g,kc)
```

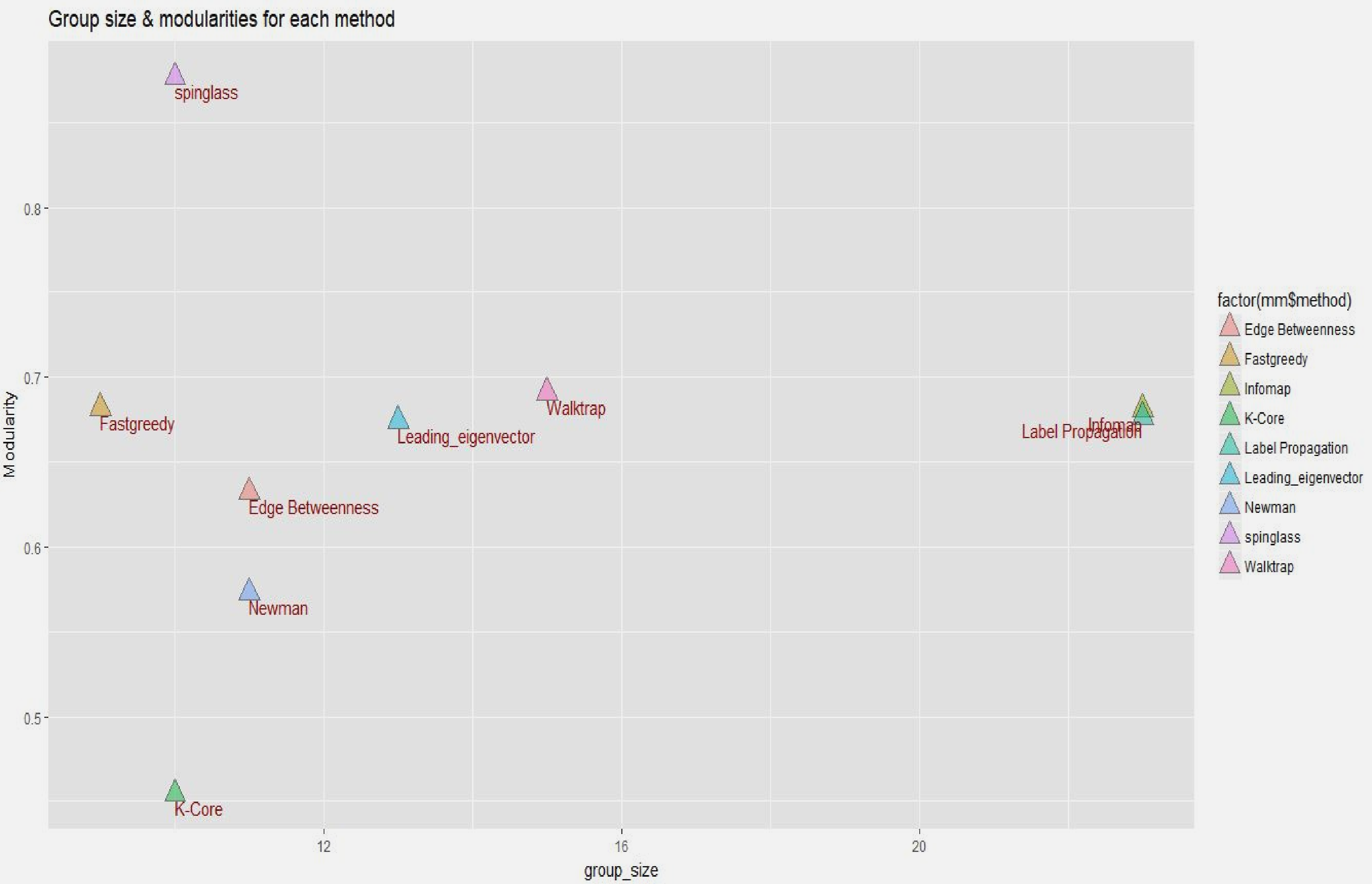
Edge Betweenness



Modularity – 0.6322804

```
eb <- edge.betweenness.community(g)
modularity(eb)
```

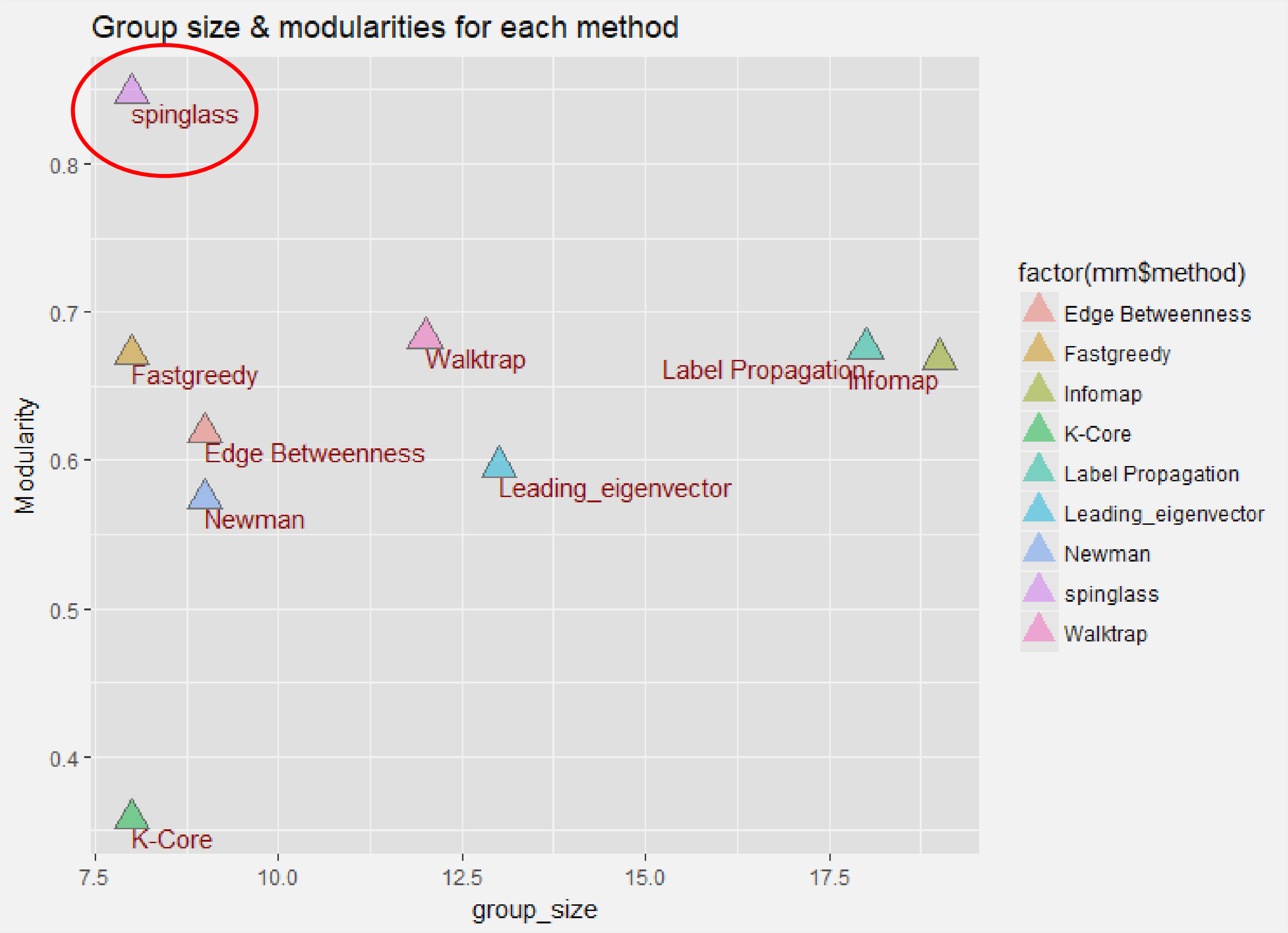
Comparison of Different Algorithms



High modularity for a partitioning reflects dense connections within communities and sparse connections across communities.

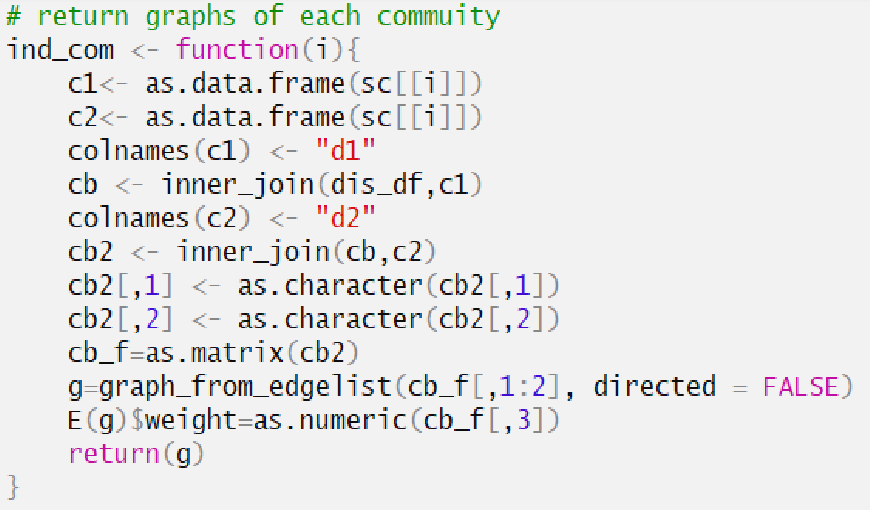
Keyplayer Detection 03

Keyplayer Detection



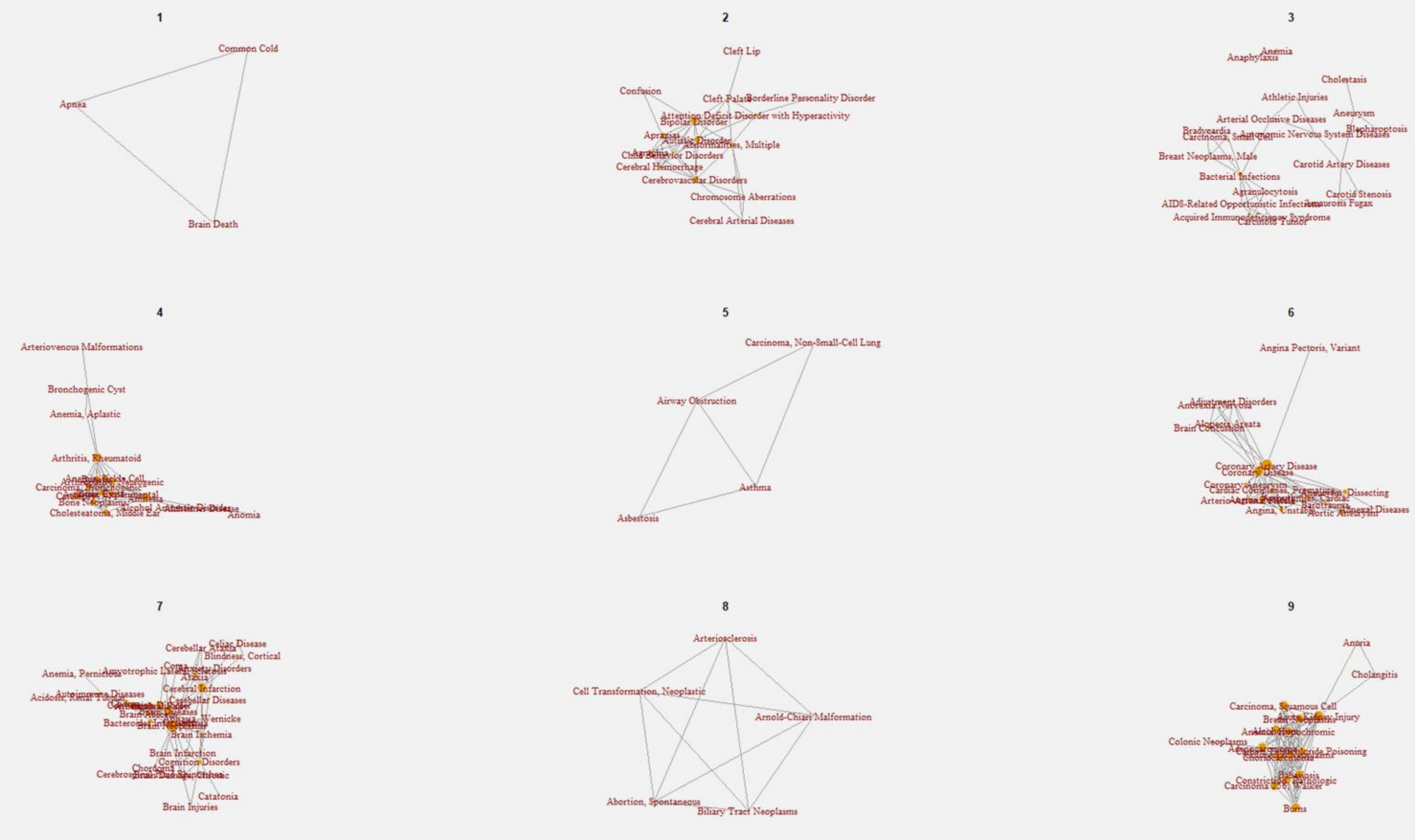
Use the community results obtained by springlass methods

- higher modularity: 0.878~0.907
- smaller group size : 9 communities



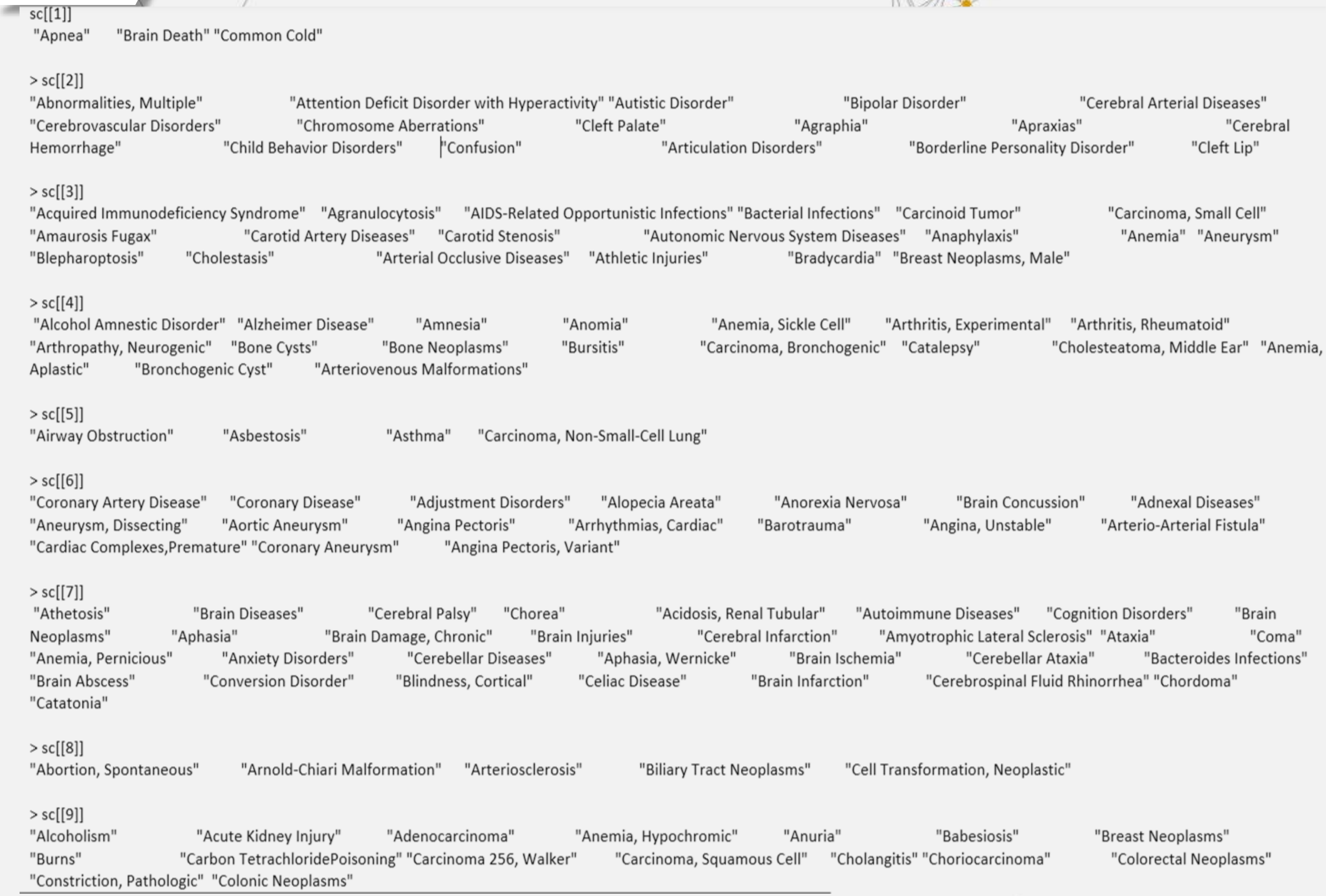
```
# return graphs of each commuity
ind_com <- function(i){
    c1<- as.data.frame(sc[[i]])
    c2<- as.data.frame(sc[[i]])
    colnames(c1) <- "d1"
    cb <- inner_join(dis_df,c1)
    colnames(c2) <- "d2"
    cb2 <- inner_join(cb,c2)
    cb2[,1] <- as.character(cb2[,1])
    cb2[,2] <- as.character(cb2[,2])
    cb_f=as.matrix(cb2)
    g=graph_from_edgelist(cb_f[,1:2], directed = FALSE)
    E(g)$weight=as.numeric(cb_f[,3])
    return(g)
}
```

9 Communities Formed by Spinglass



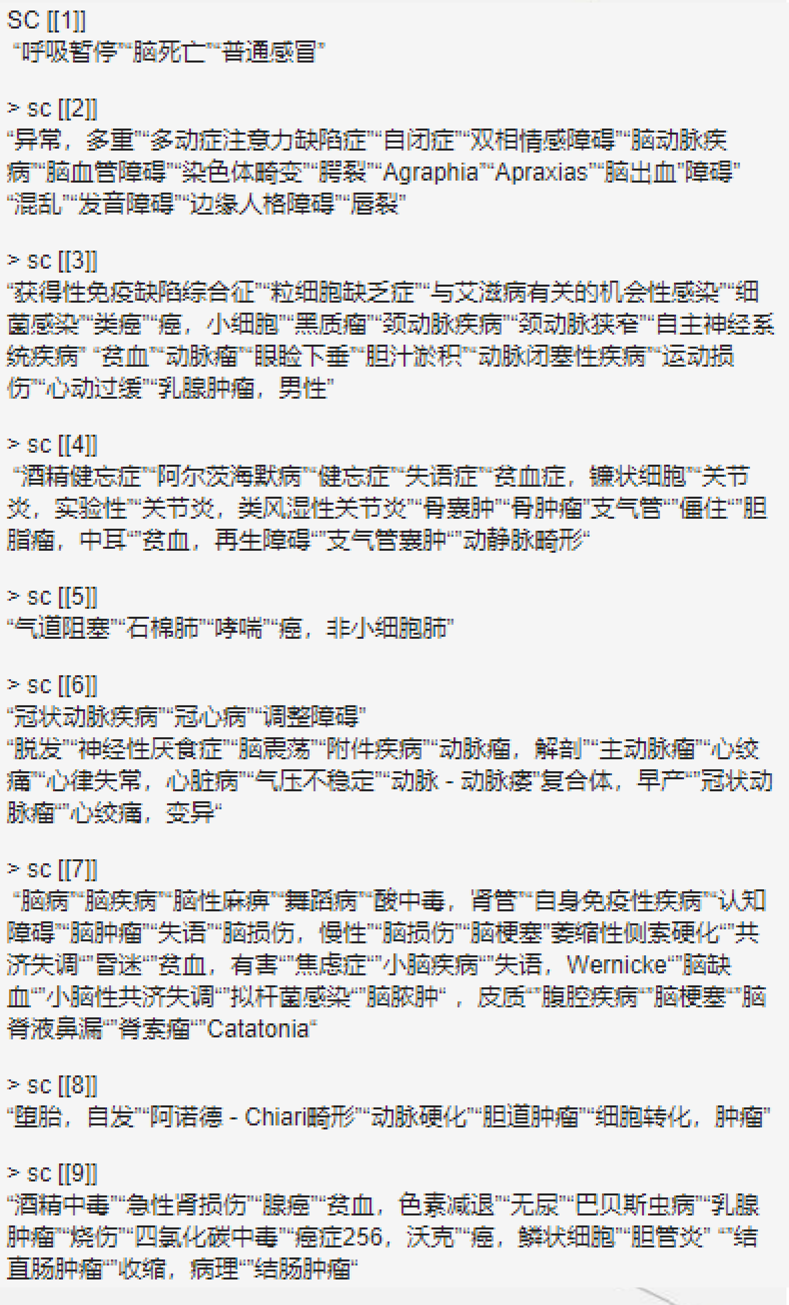
9 Communities Formed by Spinglass

2



```
sc[[1]]
 "Apnea"      "Brain Death" "Common Cold"

> sc[[2]]
"Abnormalities, Multiple"              "Attention Deficit Disorder with Hyperactivity" "Autistic Disorder"              "Bipolar Disorder"              "Cerebral Arterial Diseases"
"Cerebrovascular Disorders"            "Chromosome Aberrations"          "Cleft Palate"               "Agraphia"                  "Apraxias"                  "Cerebral
Hemorrhage"             "Child Behavior Disorders"          "Confusion"                 "Articulation Disorders"           "Borderline Personality Disorder"        "Cleft Lip"

> sc[[3]]
"Acquired Immunodeficiency Syndrome"   "Agranulocytosis"    "AIDS-Related Opportunistic Infections" "Bacterial Infections"   "Carcinoid Tumor"             "Carcinoma, Small Cell"
"Amaurosis Fugax"             "Carotid Artery Diseases"   "Carotid Stenosis"             "Autonomic Nervous System Diseases"   "Anaphylaxis"             "Anemia"   "Aneurysm"
"Blepharoptosis"        "Cholestasis"               "Arterial Occlusive Diseases"    "Athletic Injuries"            "Bradycardia"  "Breast Neoplasms, Male"

> sc[[4]]
 "Alcohol Amnestic Disorder"  "Alzheimer Disease"      "Amnesia"            "Anomia"             "Anemia, Sickle Cell"      "Arthritis, Experimental"    "Arthritis, Rheumatoid"
"Arthropathy, Neurogenic"    "Bone Cysts"           "Bone Neoplasms"       "Bursitis"            "Carcinoma, Bronchogenic"  "Catalepsy"            "Cholesteatoma, Middle Ear"   "Anemia,
Aplastic"         "Bronchogenic Cyst"        "Arteriovenous Malformations"

> sc[[5]]
"Airway Obstruction"        "Asbestosis"              "Asthma"     "Carcinoma, Non-Small-Cell Lung"

> sc[[6]]
"Coronary Artery Disease"    "Coronary Disease"       "Adjustment Disorders"      "Alopecia Areata"         "Anorexia Nervosa"        "Brain Concussion"        "Adnexal Diseases"
"Aneurysm, Dissecting"      "Aortic Aneurysm"        "Angina Pectoris"         "Arrhythmias, Cardiac"      "Barotrauma"             "Angina, Unstable"         "Arterio-Arterial Fistula"
"Cardiac Complexes,Premature" "Coronary Aneurysm"        "Angina Pectoris, Variant"

> sc[[7]]
 "Athetosis"            "Brain Diseases"           "Cerebral Palsy"     "Chorea"              "Acidosis, Renal Tubular"      "Autoimmune Diseases"      "Cognition Disorders"        "Brain
Neoplasms"         "Aphasia"               "Brain Damage, Chronic"       "Brain Injuries"          "Cerebral Infarction"        "Amyotrophic Lateral Sclerosis"  "Ataxia"               "Coma"
"Anemia, Pernicious"        "Anxiety Disorders"         "Cerebellar Diseases"        "Aphasia, Wernicke"         "Brain Ischemia"           "Cerebellar Ataxia"          "Bacteroides Infections"
"Brain Abscess"           "Conversion Disorder"        "Blindness, Cortical"       "Celiac Disease"          "Brain Infarction"          "Cerebrospinal Fluid Rhinorrhea" "Chordoma"
"Catatonia"

> sc[[8]]
"Abortion, Spontaneous"       "Arnold-Chiari Malformation"     "Arteriosclerosis"          "Biliary Tract Neoplasms"       "Cell Transformation, Neoplastic"

> sc[[9]]
"Alcoholism"            "Acute Kidney Injury"        "Adenocarcinoma"          "Anemia, Hypochromic"        "Anuria"                "Babesiosis"              "Breast Neoplasms"
"Burns"              "Carbon TetrachloridePoisoning" "Carcinoma 256, Walker"      "Carcinoma, Squamous Cell"    "Cholangitis" "Choriocarcinoma"          "Colorectal Neoplasms"
"Constriction, Pathologic"  "Colonic Neoplasms"
```

3



```
SC [[1]]
 “呼吸暂停”“脑死亡”“普通感冒”

> sc [[2]]
“异常，多重”“多动症注意力缺陷症”“自闭症”“双相情感障碍”“脑动脉疾
病”“脑血管障碍”“染色体畸变”“腭裂”“Agraphia”“Apraxias”“脑出血”障碍”
“混乱”“发音障碍”“边缘人格障碍”“唇裂”

> sc [[3]]
“获得性免疫缺陷综合征”“粒细胞缺乏症”“与艾滋病有关的机会性感染”“细
菌感染”“类癌”“癌，小细胞”“黑质瘤”“颈动脉疾病”“颈动脉狭窄”“自主神经系
统疾病” “贫血”“动脉瘤”“眼睑下垂”“胆汁淤积”“动脉闭塞性疾病”“运动损
伤”“心动过缓”“乳腺肿瘤，男性”

> sc [[4]]
 “酒精健忘症”“阿尔茨海默病”“健忘症”“失语症”“贫血症，镰状细胞”“关节
炎，实验性”“关节炎，类风湿性关节炎”“骨囊肿”“骨肿瘤”支气管”“僵住”“胆
脂瘤，中耳”“贫血，再生障碍”“支气管囊肿”“动静脉畸形”

> sc [[5]]
“气道阻塞”“石棉肺”“哮喘”“癌，非小细胞肺”

> sc [[6]]
“冠状动脉疾病”“冠心病”“调整障碍”
“脱发”“神经性厌食症”“脑震荡”“附件疾病”“动脉瘤，解剖”“主动脉瘤”“心绞
痛”“心律失常，心脏病”“气压不稳定”“动脉 - 动脉瘘”复合体，早产”“冠状动
脉瘤”“心绞痛，变异“

> sc [[7]]
 “脑病”“脑疾病”“脑性麻痹”“舞蹈病”“酸中毒，肾管”“自身免疫性疾病”“认知
障碍”“脑肿瘤”“失语”“脑损伤，慢性”“脑损伤”“脑梗塞”萎缩性侧索硬化”“共
济失调”“昏迷”“贫血，有害”“焦虑症”“小脑疾病”“失语，Wernicke”“脑缺
血”“小脑性共济失调”“拟杆菌感染”“脑脓肿“ ，皮质”“腹腔疾病”“脑梗塞”“脑
脊液鼻漏”“脊索瘤”“Catatonia“

> sc [[8]]
“堕胎，自发”“阿诺德 - Chiari畸形”“动脉硬化”“胆道肿瘤”“细胞转化，肿瘤”

> sc [[9]]
“酒精中毒”“急性肾损伤”“腺癌”“贫血，色素减退”“无尿”“巴贝斯虫病”“乳腺
肿瘤”“烧伤”“四氯化碳中毒”“癌症256，沃克”“癌，鳞状细胞”“胆管炎” “结
直肠肿瘤”“收缩，病理”“结肠肿瘤“
```

9 Communities Formed by Spinglass

2

3

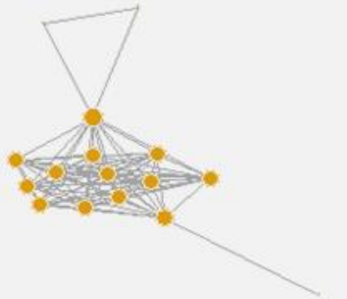
Community	Represent	Size
1	No certain pattern (small size)	3
2	Mental disorder	16
3	Immunological diseases	19
4	Hydatoncus or inflammation	17
5	Respiratory system related diseases	4
6	Heart diseases	17
7	Cerebral disease	30
8	No certain pattern (small size)	5
9	Tumor/cancer/damage to organs	16

4

6

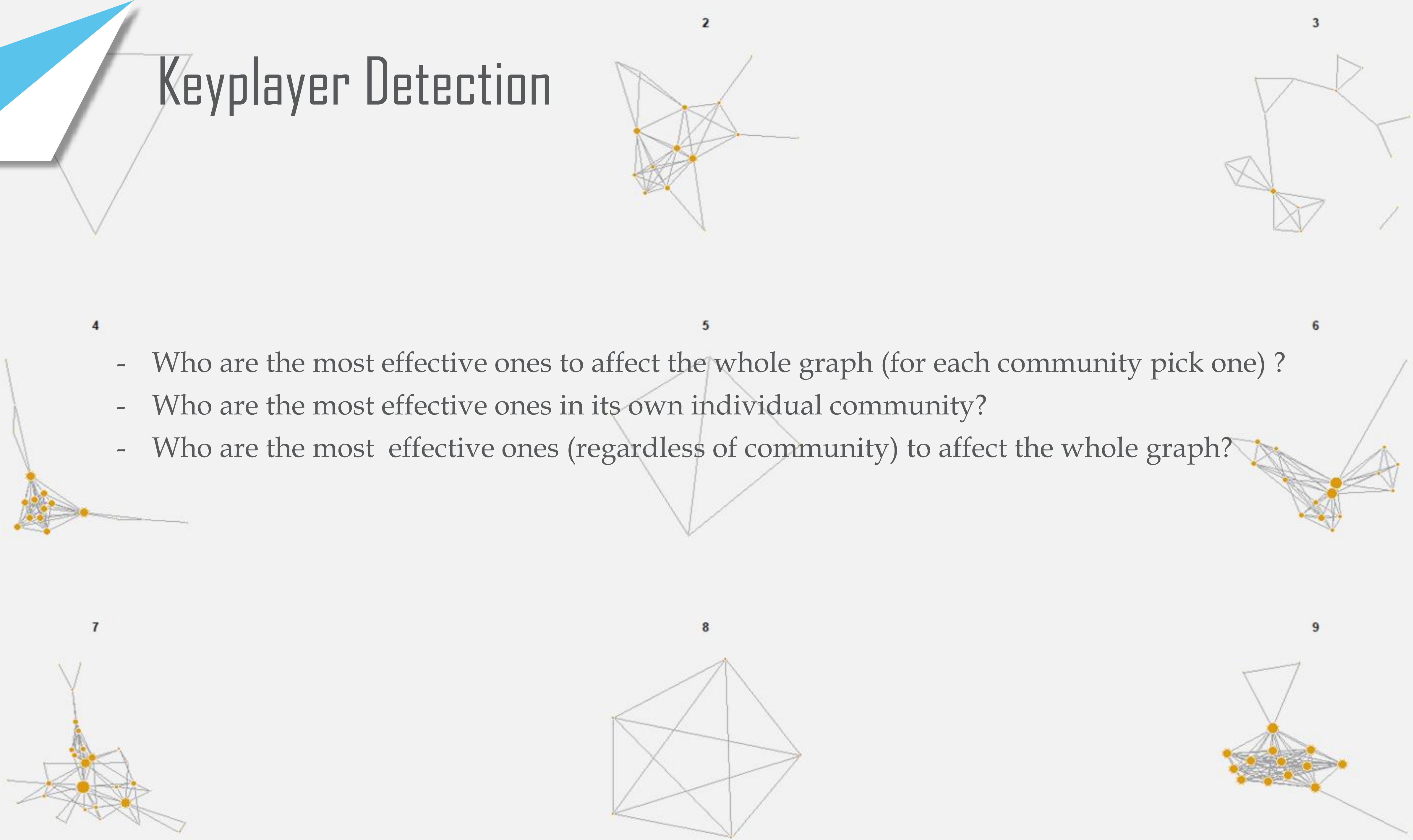
7

9



Keyplayer Detection

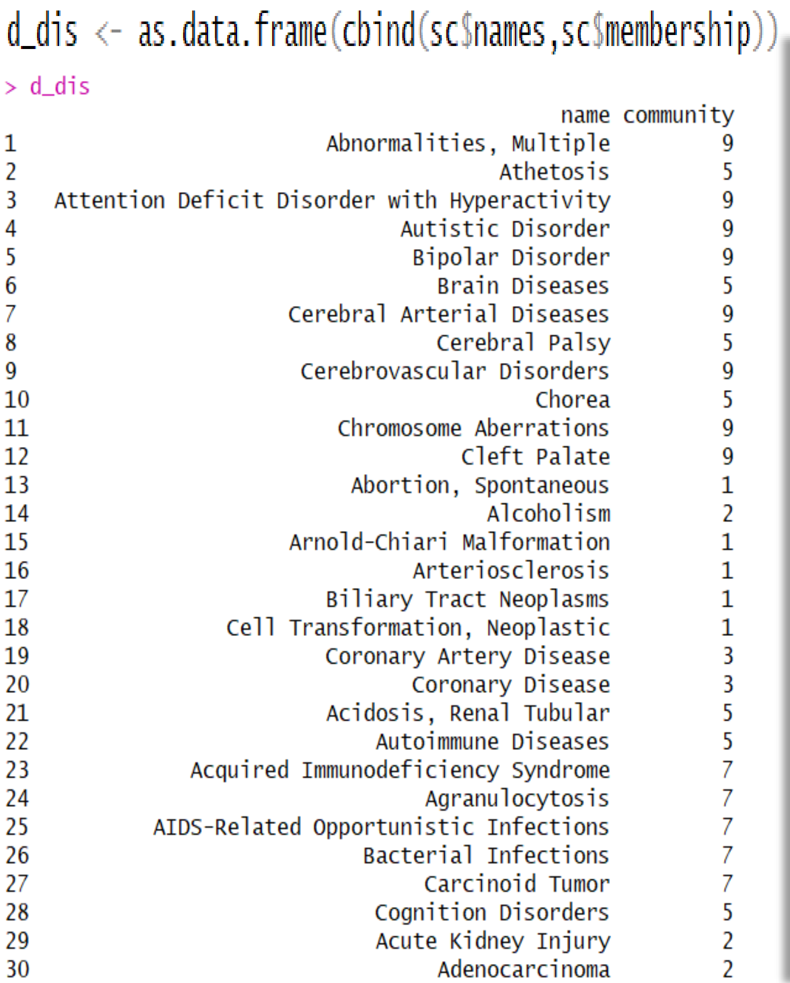
- Who are the most effective ones to affect the whole graph (for each community pick one) ?
- Who are the most effective ones in its own individual community?
- Who are the most effective ones (regardless of community) to affect the whole graph?



Keyplayer Detection Using 4 Measurements

4 measurements: Degree/Betweenness/Closeness/Weight (use degree-based as example)

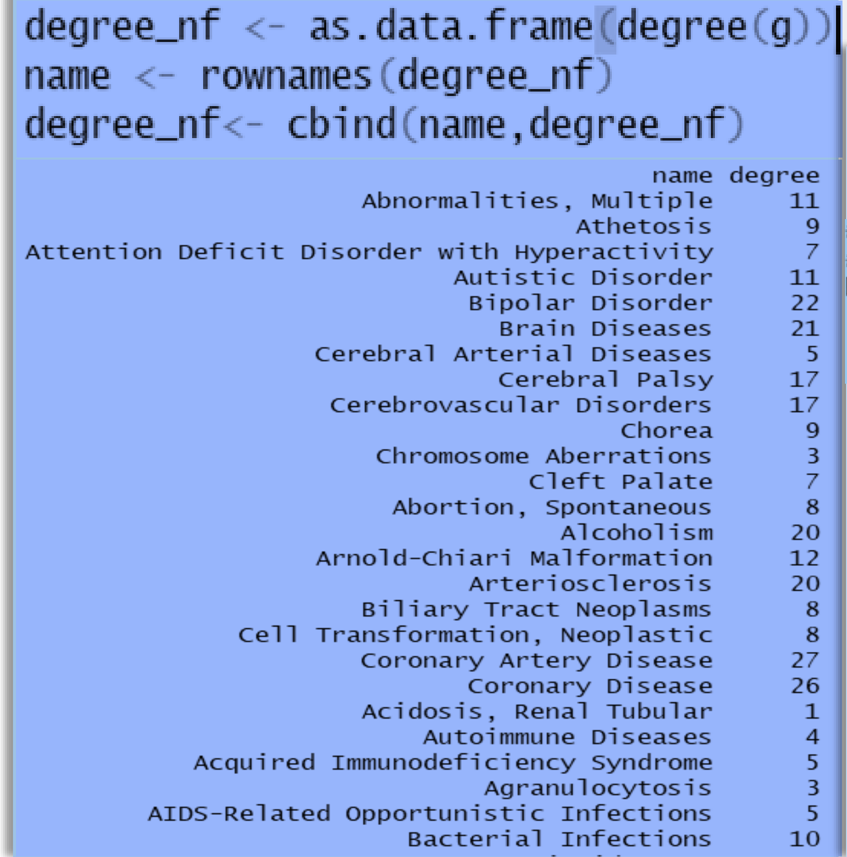
1. Assign diseases into each community



```
d_dis <- as.data.frame(cbind(sc$names,sc$membership))
```

```
> d_dis
                                      name community
1                  Abnormalities, Multiple         9
2                                Athetosis         5
3   Attention Deficit Disorder with Hyperactivity  9
4                        Autistic Disorder         9
5                         Bipolar Disorder         9
6                           Brain Diseases         5
7                Cerebral Arterial Diseases        9
8                           Cerebral Palsy         5
9                 Cerebrovascular Disorders        9
10                                  Chorea         5
11                  Chromosome Aberrations         9
12                           Cleft Palate          9
13                  Abortion, Spontaneous          1
14                              Alcoholism         2
15                Arnold-Chiari Malformation       1
16                      Arteriosclerosis           1
17                  Biliary Tract Neoplasms        1
18             Cell Transformation, Neoplastic     1
19                 Coronary Artery Disease         3
20                        Coronary Disease         3
21                  Acidosis, Renal Tubular        5
22                      Autoimmune Diseases        5
23          Acquired Immunodeficiency Syndrome     7
24                         Agranulocytosis         7
25       AIDS-Related Opportunistic Infections     7
26                     Bacterial Infections        7
27                         Carcinoid Tumor         7
28                     Cognition Disorders         5
29                      Acute Kidney Injury        2
30                           Adenocarcinoma        2
```

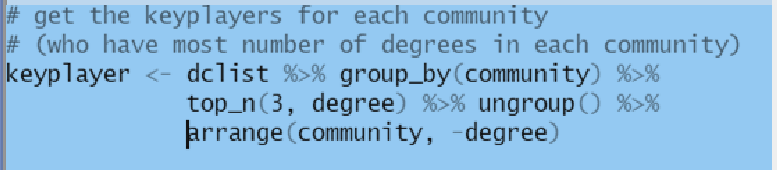
2. Find number of degrees for every disease



```
degree_nf <- as.data.frame(degree(g))
name <- rownames(degree_nf)
degree_nf<- cbind(name,degree_nf)
```

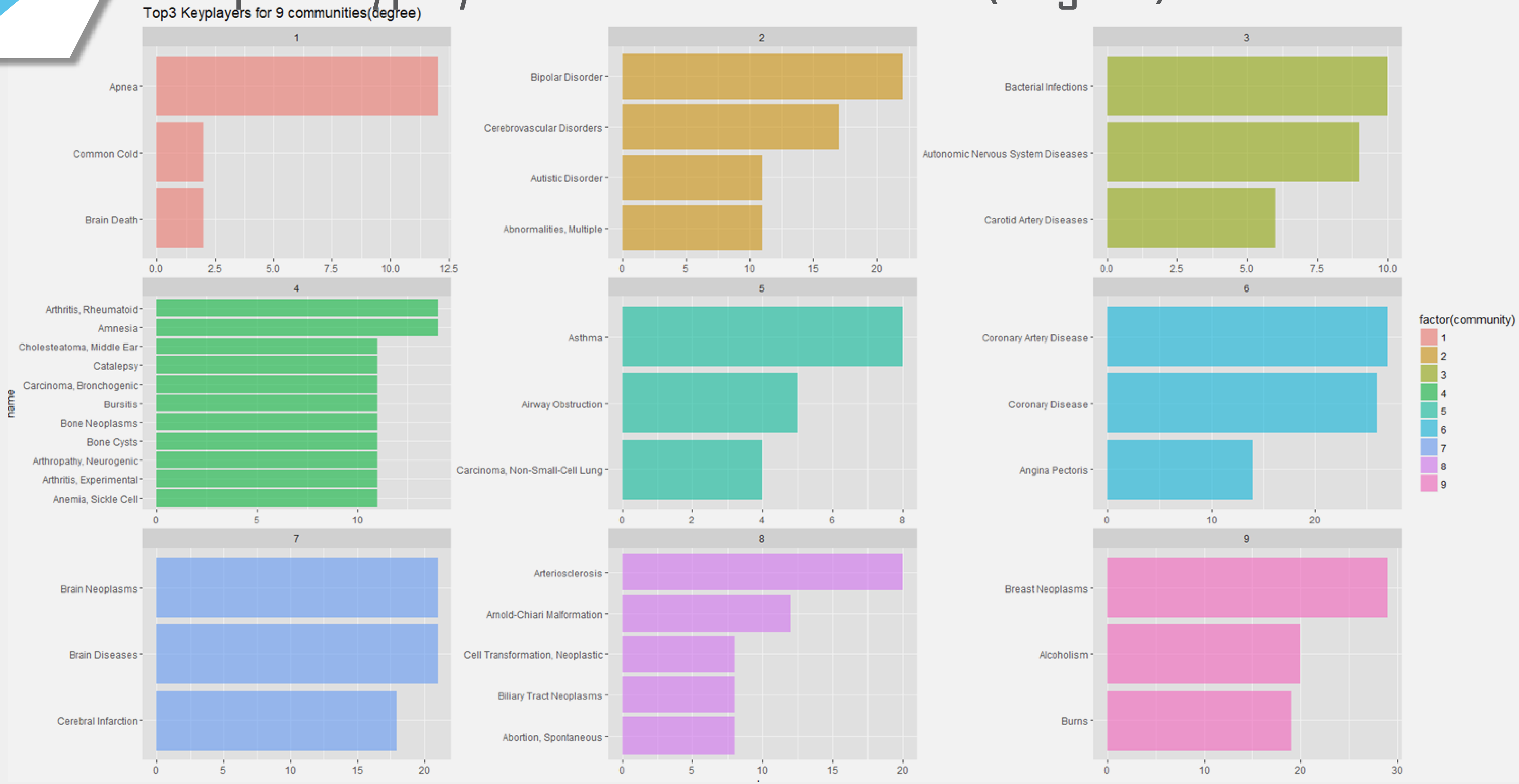
```
                                       name degree
                    Abnormalities, Multiple     11
                                  Athetosis      9
Attention Deficit Disorder with Hyperactivity    7
                          Autistic Disorder     11
                           Bipolar Disorder     22
                             Brain Diseases     21
                 Cerebral Arterial Diseases      5
                             Cerebral Palsy     17
                  Cerebrovascular Disorders     17
                                     Chorea      9
                     Chromosome Aberrations      3
                               Cleft Palate      7
                      Abortion, Spontaneous      8
                                 Alcoholism     20
                 Arnold-Chiari Malformation     12
                           Arteriosclerosis     20
                    Biliary Tract Neoplasms      8
            Cell Transformation, Neoplastic      8
                    Coronary Artery Disease     27
                           Coronary Disease     26
                    Acidosis, Renal Tubular      1
                        Autoimmune Diseases      4
         Acquired Immunodeficiency Syndrome      5
                            Agranulocytosis      3
      AIDS-Related Opportunistic Infections      5
                       Bacterial Infections     10
```

3. Join two tables and get the keyplayers for each community



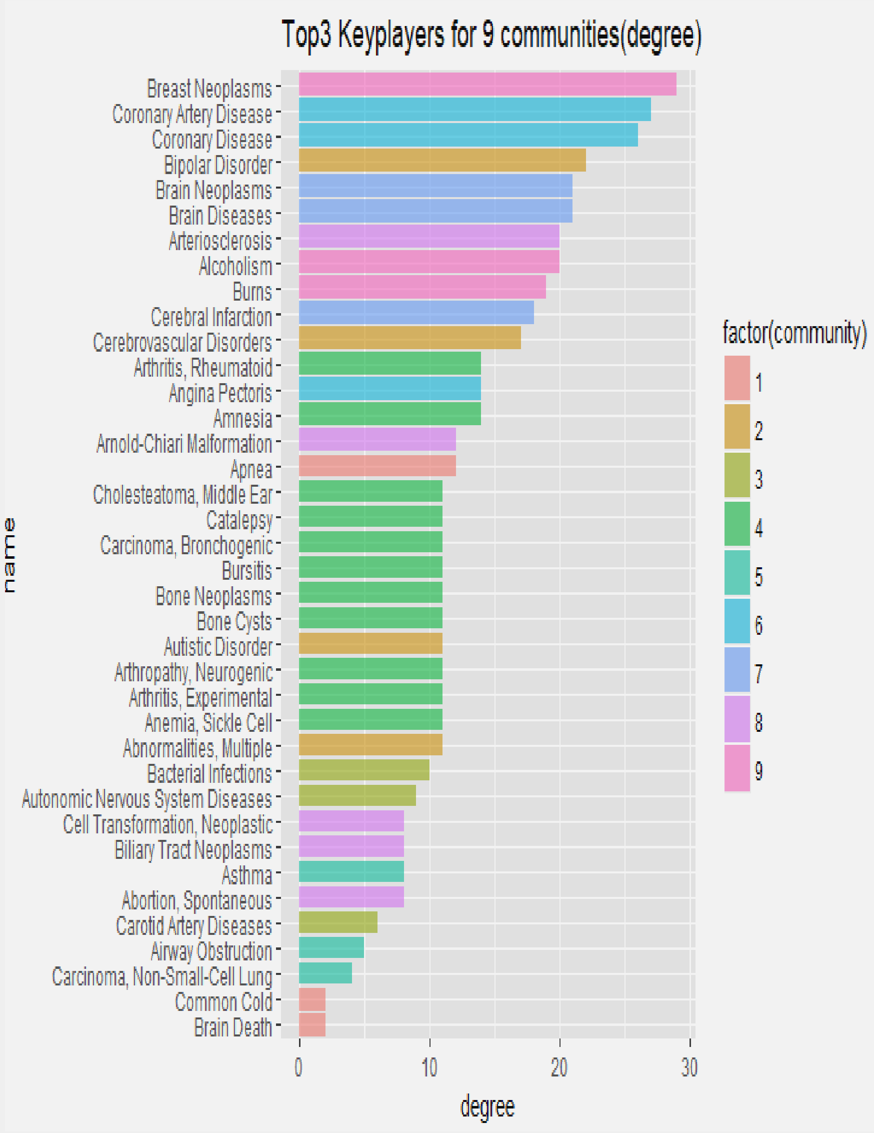
```
# get the keyplayers for each community
# (who have most number of degrees in each community)
keyplayer <- dclist %>% group_by(community) %>%
             top_n(3, degree) %>% ungroup() %>%
             arrange(community, -degree)
```

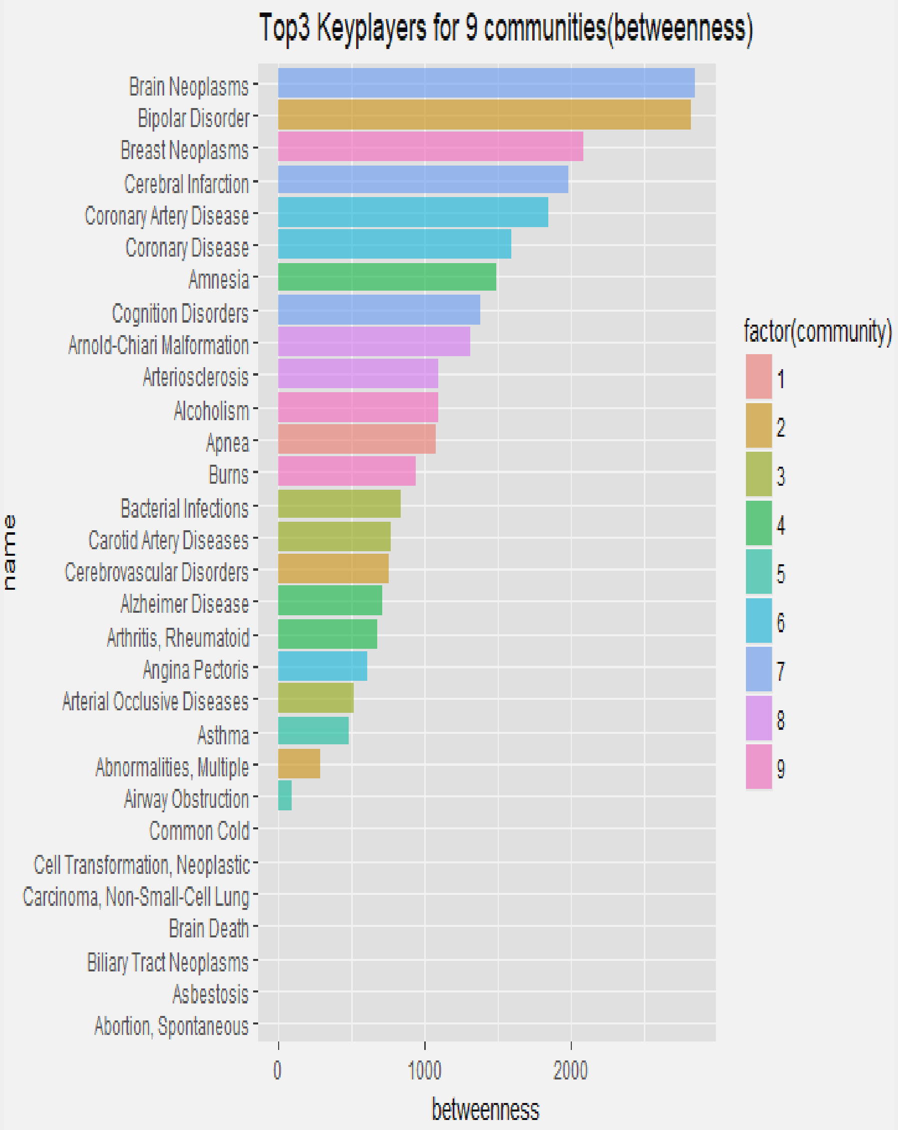
Top 3 Keyplayers For 9 Communities (Degree)



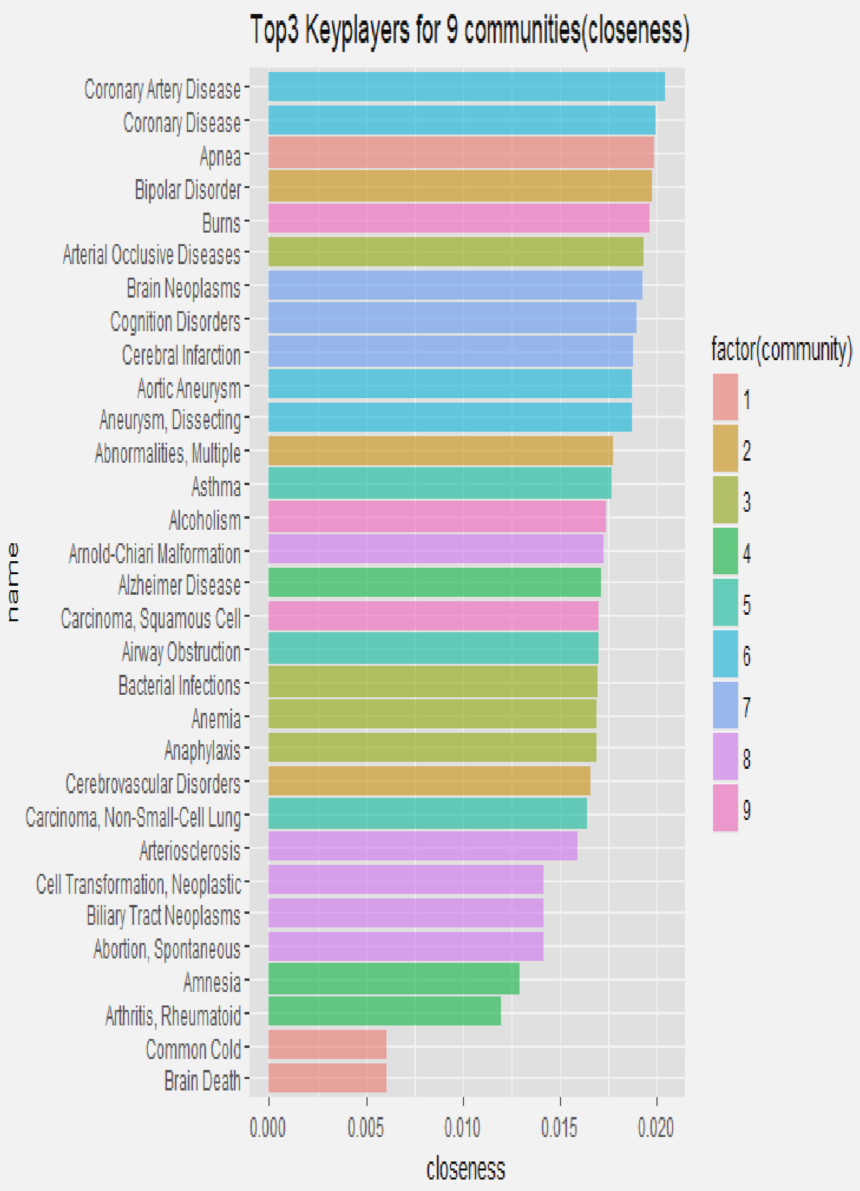
Degree

Betweenness

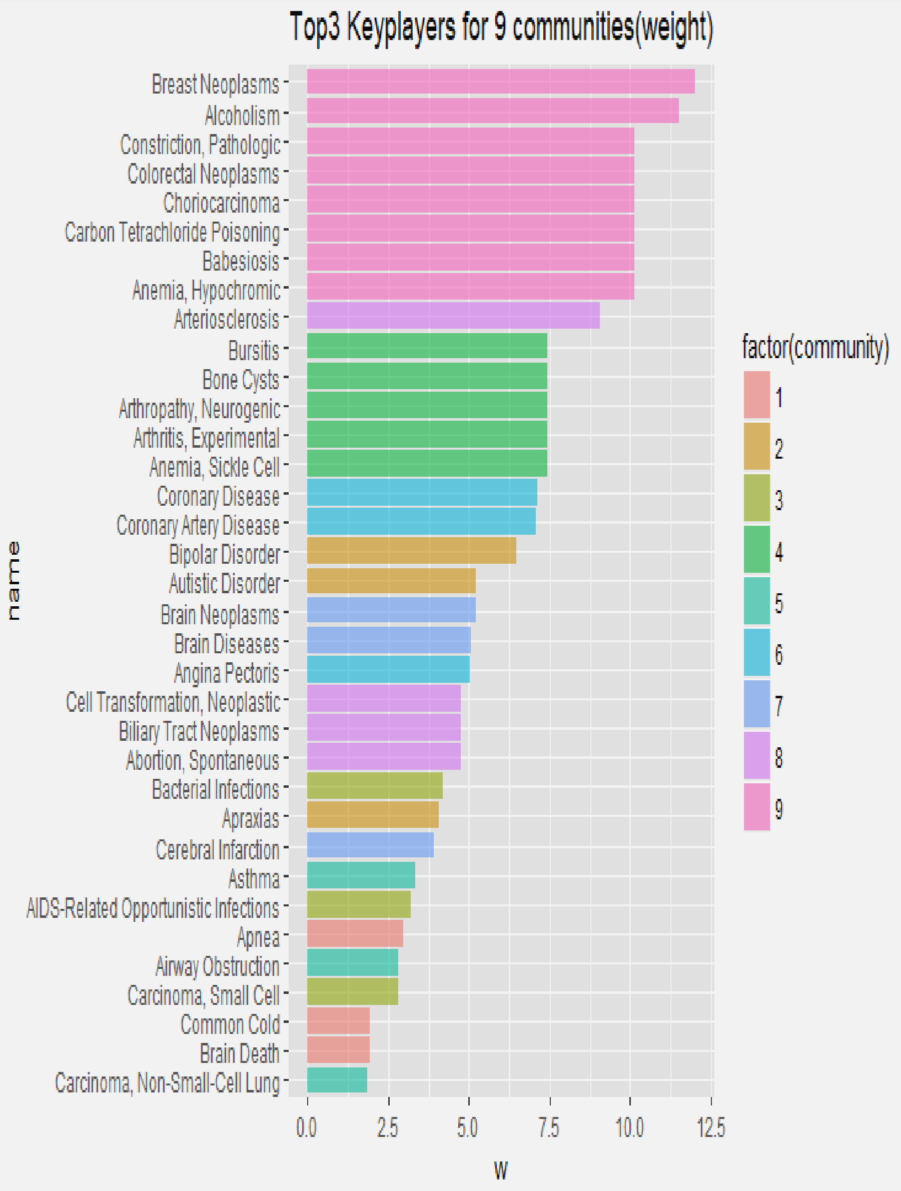




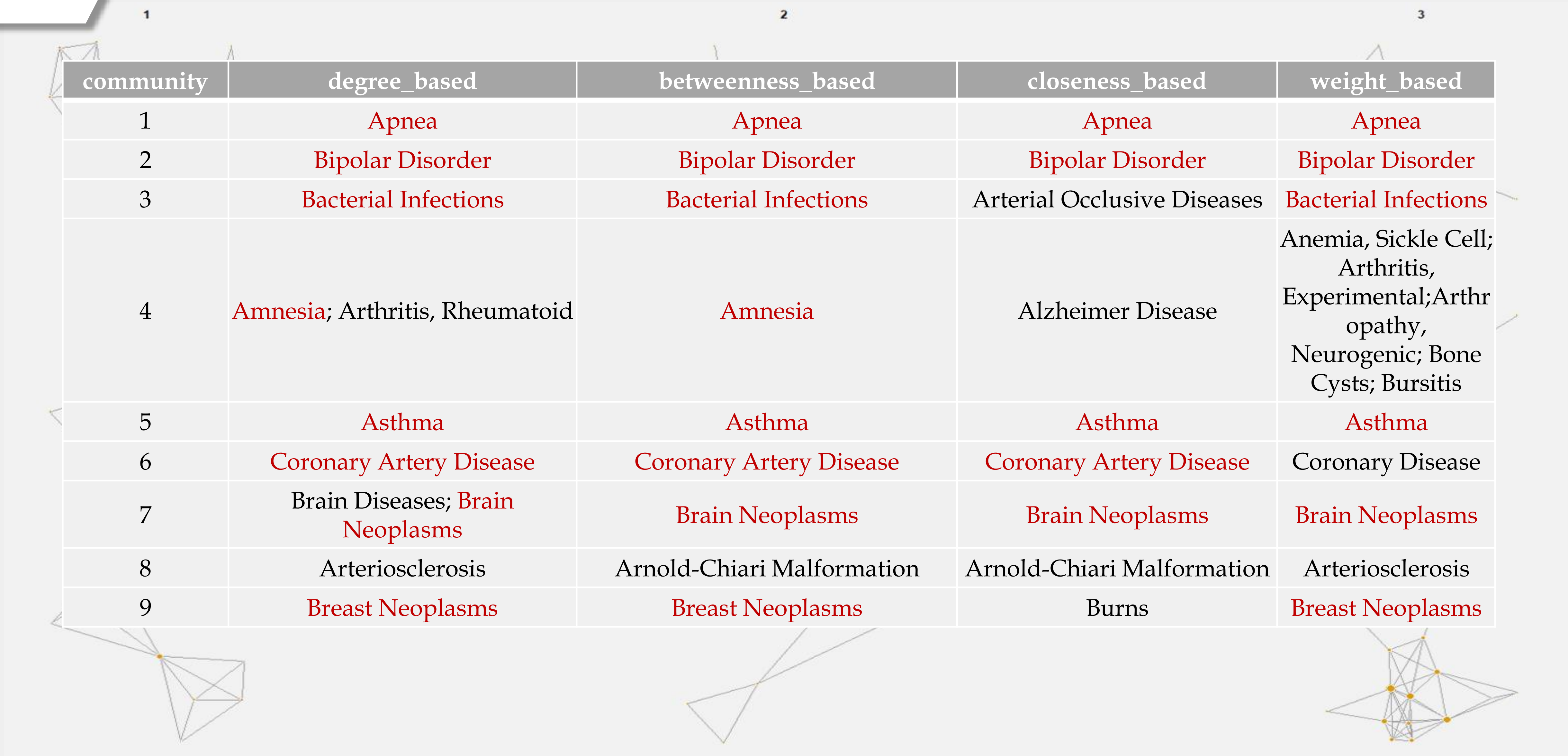
Closeness



Weight



Top 1 Keyplayer in Each Community (Different Measurements)



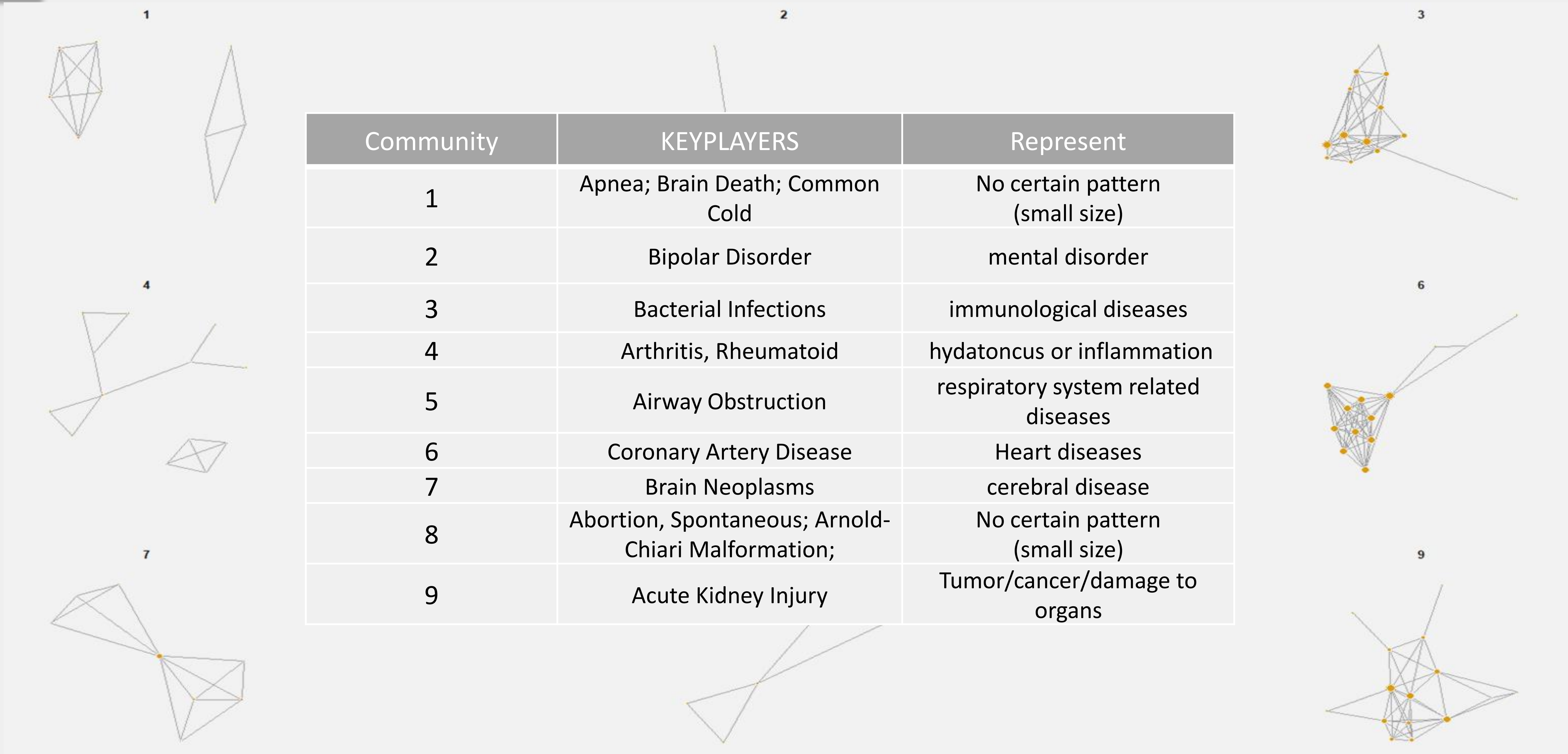
community	degree_based	betweenness_based	closeness_based	weight_based
1	Apnea	Apnea	Apnea	Apnea
2	Bipolar Disorder	Bipolar Disorder	Bipolar Disorder	Bipolar Disorder
3	Bacterial Infections	Bacterial Infections	Arterial Occlusive Diseases	Bacterial Infections
4	Amnesia; Arthritis, Rheumatoid	Amnesia	Alzheimer Disease	Anemia, Sickle Cell; Arthritis, Experimental;Arthropathy, Neurogenic; Bone Cysts; Bursitis
5	Asthma	Asthma	Asthma	Asthma
6	Coronary Artery Disease	Coronary Artery Disease	Coronary Artery Disease	Coronary Disease
7	Brain Diseases; Brain Neoplasms	Brain Neoplasms	Brain Neoplasms	Brain Neoplasms
8	Arteriosclerosis	Arnold-Chiari Malformation	Arnold-Chiari Malformation	Arteriosclerosis
9	Breast Neoplasms	Breast Neoplasms	Burns	Breast Neoplasms

Find the Most Effective Ones in Each Community

View each community as an individual graph (break the links between communities, and then to get the most effective ones in each community)

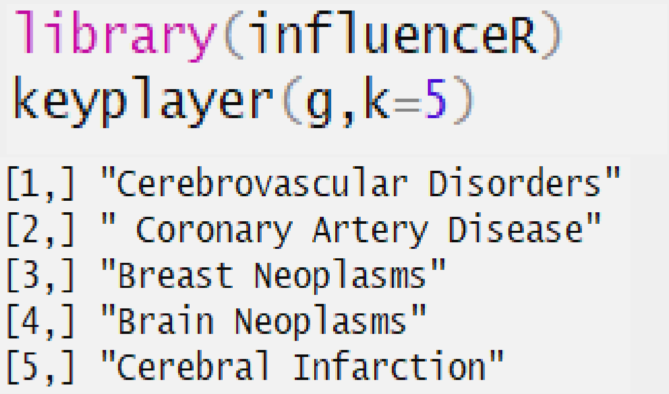
Community	degree_based	betweenness_based	closeness_based	weight_based
1	Apnea; Brain Death; Common Cold	Apnea; Brain Death; Common Cold	Apnea; Brain Death; Common Cold	Apnea; Brain Death; Common Cold
2	Bipolar Disorder	Bipolar Disorder	Bipolar Disorder	Autistic Disorder
3	Bacterial Infections	Bacterial Infections	Arterial Occlusive Diseases	Bacterial Infections
4	Amnesia; Arthritis, Rheumatoid	Amnesia; Arthritis, Rheumatoid	Amnesia	Anemia, Sickle Cell; Arthritis, Experimental; Arthropathy, Neurogenic; Bone Cysts; Bursitis
5	Airway Obstruction; Asthma	Airway Obstruction; Asthma	Asthma	Airway Obstruction
6	Coronary Artery Disease	Coronary Artery Disease	Aneurysm, Dissecting; Aortic Aneurysm	Coronary Artery Disease
7	Brain Neoplasms	Brain Neoplasms	Chorea	Brain Neoplasms
8	Abortion, Spontaneous; Arnold-Chiari Malformation; Arteriosclerosis; Biliary Tract Neoplasms; Cell Transformation, Neoplastic	Abortion, Spontaneous; Arnold-Chiari Malformation; Arteriosclerosis; Biliary Tract Neoplasms; Cell Transformation, Neoplastic	Arnold-Chiari Malformation	Abortion, Spontaneous; Biliary Tract Neoplasms; Cell Transformation, Neoplastic
9	Acute Kidney Injury	Acute Kidney Injury	Acute Kidney Injury	Anemia, Hypochromic; Babesiosis; Carbon Tetrachloride Poisoning; Choriocarcinoma; Colorectal Neoplasms; Constriction, Pathologic

Top 1 Keyplayer in Each Community (Different Measurements)



Community	KEYPLAYERS	Represent
1	Apnea; Brain Death; Common Cold	No certain pattern (small size)
2	Bipolar Disorder	mental disorder
3	Bacterial Infections	immunological diseases
4	Arthritis, Rheumatoid	hydatoncus or inflammation
5	Airway Obstruction	respiratory system related diseases
6	Coronary Artery Disease	Heart diseases
7	Brain Neoplasms	cerebral disease
8	Abortion, Spontaneous; Arnold-Chiari Malformation;	No certain pattern (small size)
9	Acute Kidney Injury	Tumor/cancer/damage to organs

Find the Most Effective Ones in the Whole Graph



```r
library(influenceR)
keyplayer(g,k=5)
```

```
[1,] "Cerebrovascular Disorders"
[2,] " Coronary Artery Disease"
[3,] "Breast Neoplasms"
[4,] "Brain Neoplasms"
[5,] "Cerebral Infarction"
```

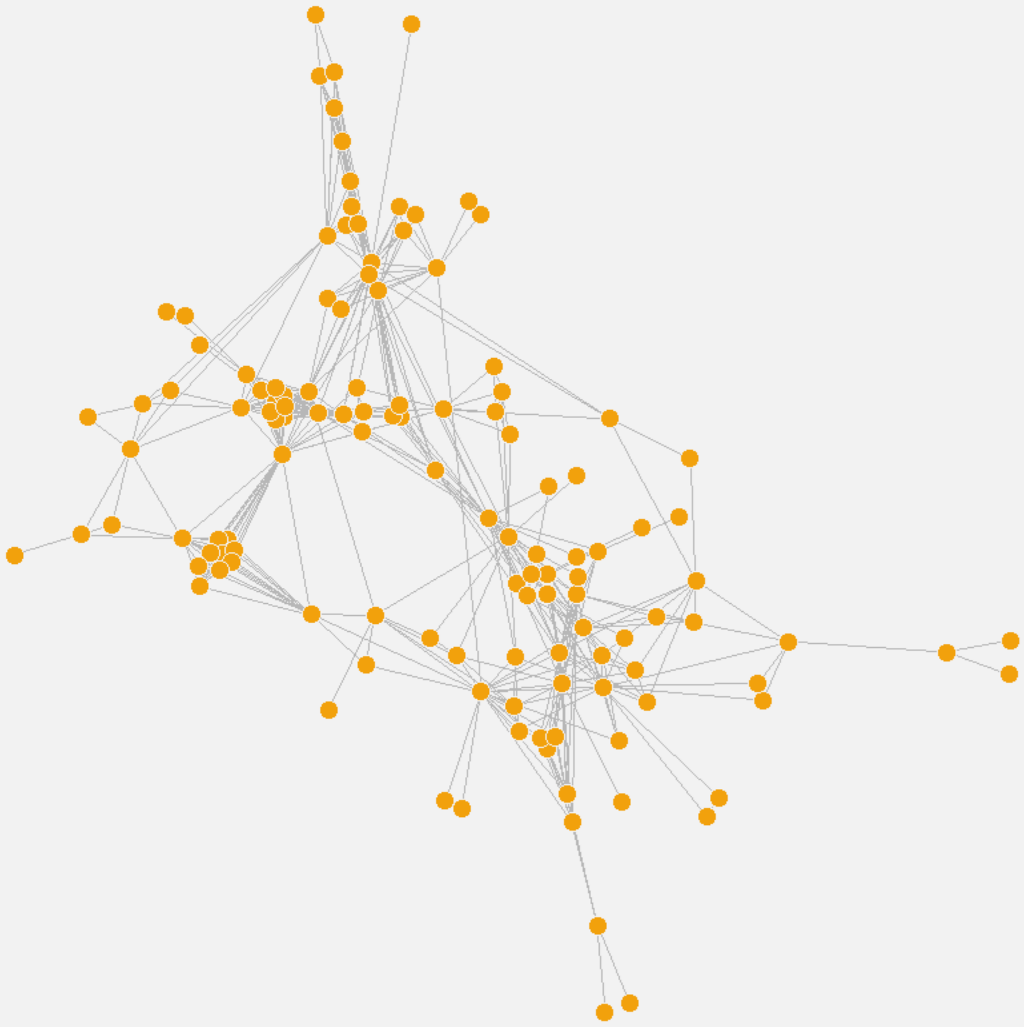
“脑血管障碍”

“冠状动脉疾病”

“乳腺肿瘤”

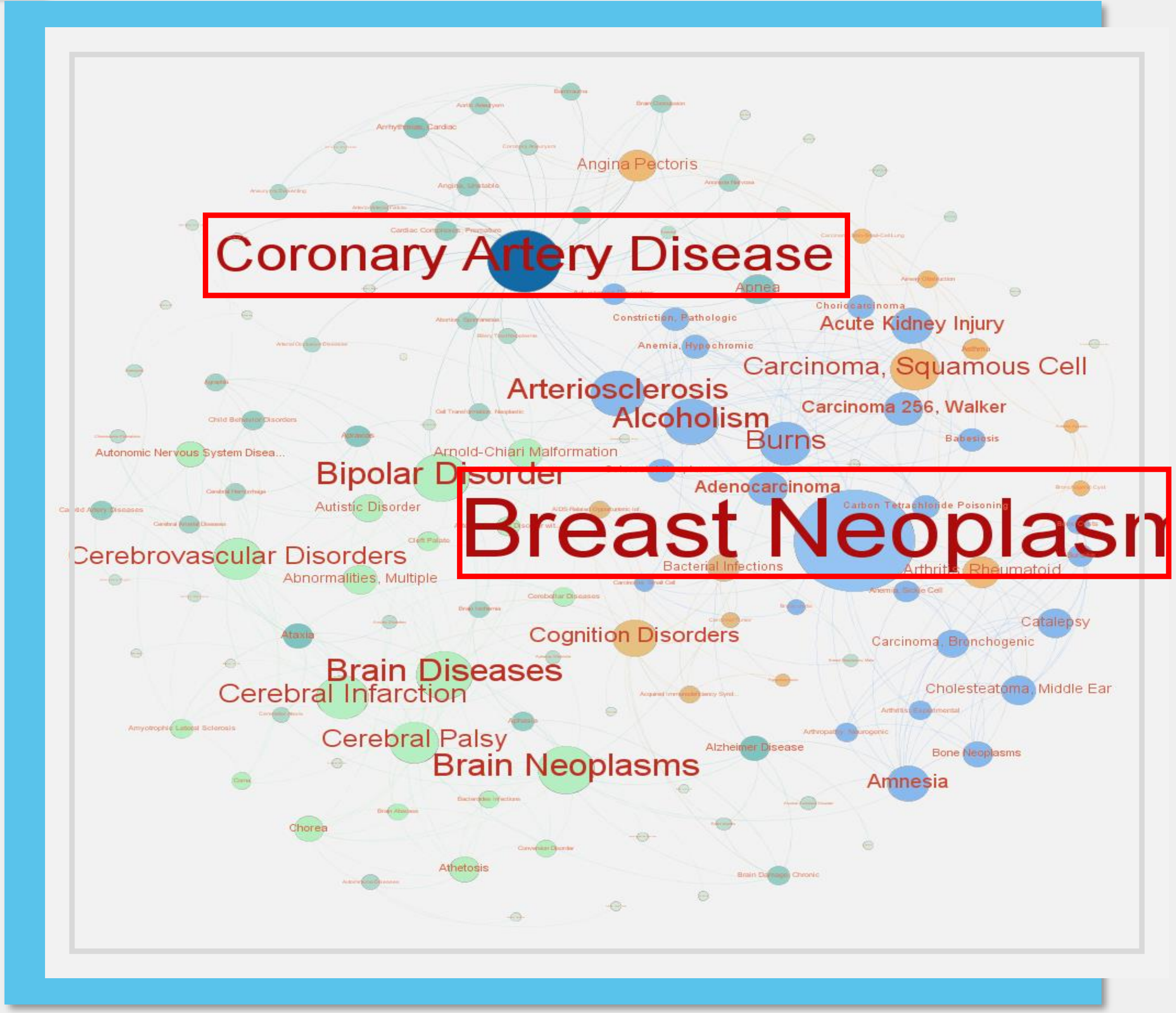
“脑肿瘤”

“脑梗塞”



Visualization & Insights 04

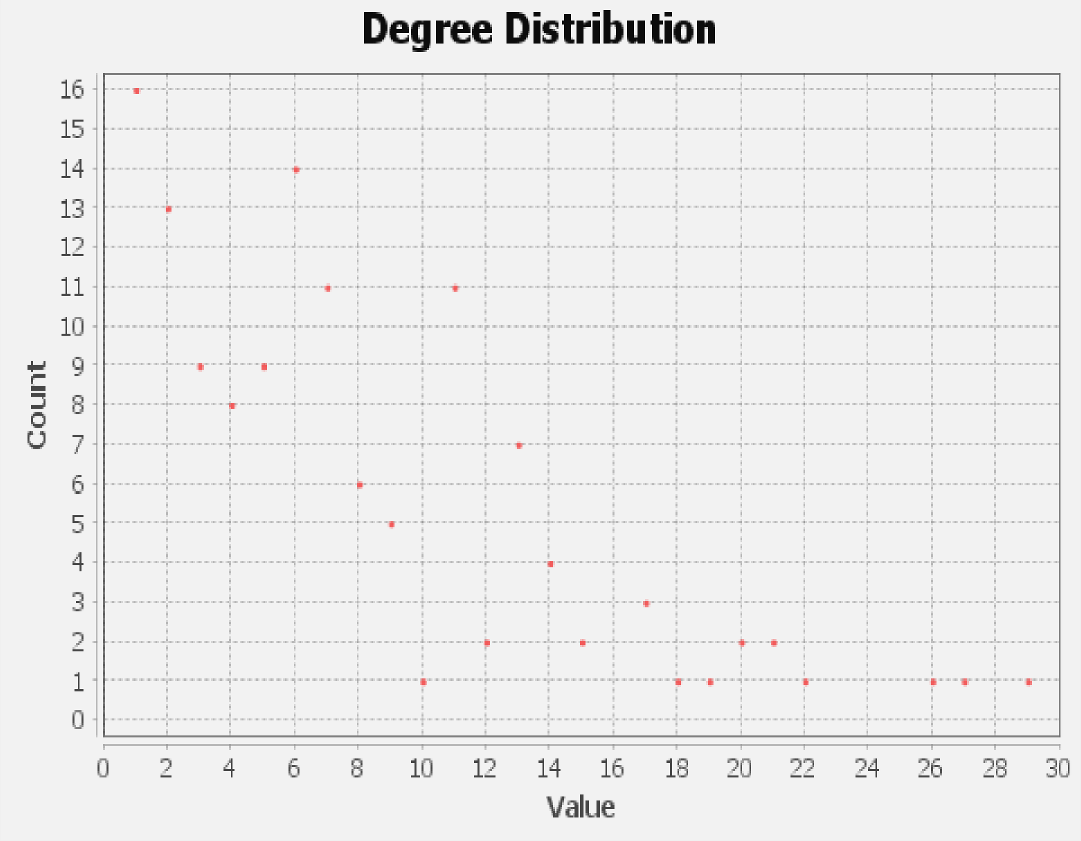
Backbone of Disease-Symptom Network

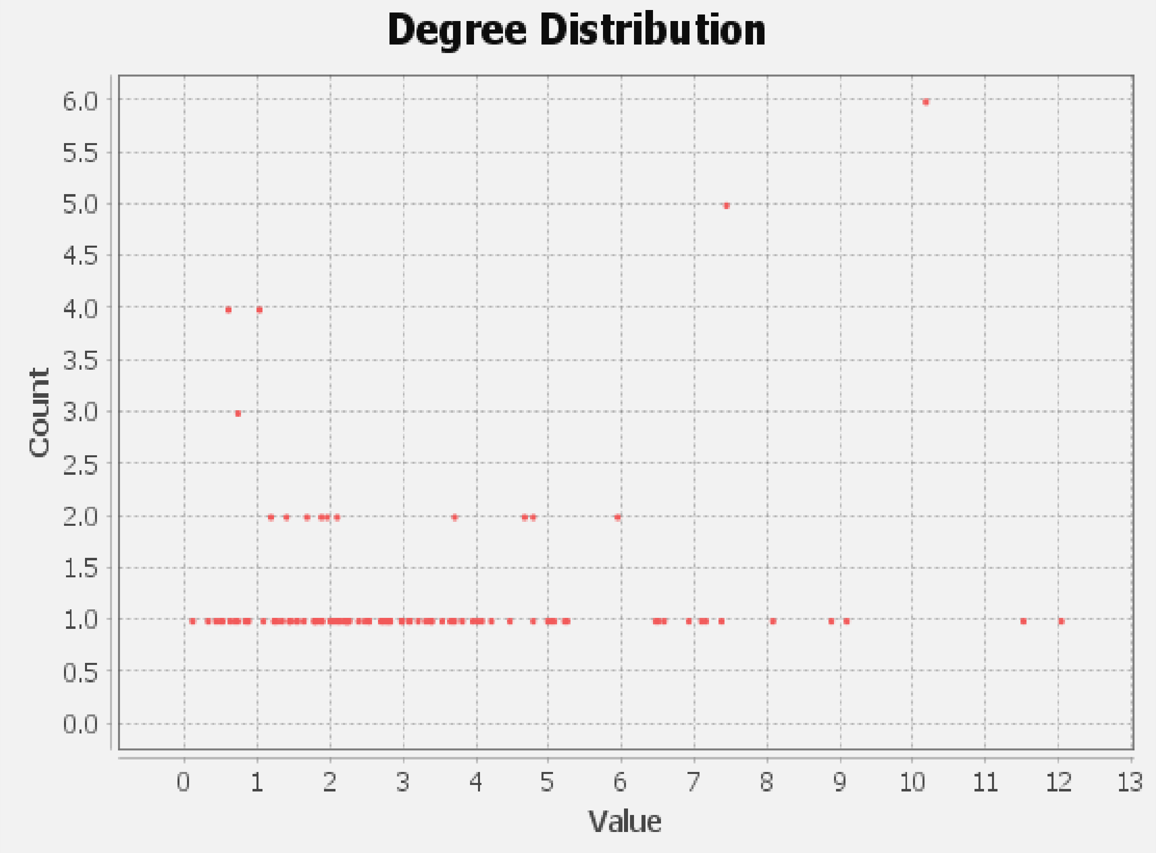


OVERVIEW

Gephi is a useful software for network analytics. Using *Fruchterman Reingold* layout, we get this circular network. You can see that **Breast Neoplasms** has the highest degree, **Coronary Artery Disease** comes the second.

Average (Weighted) Degree of Disease-Symptom Network

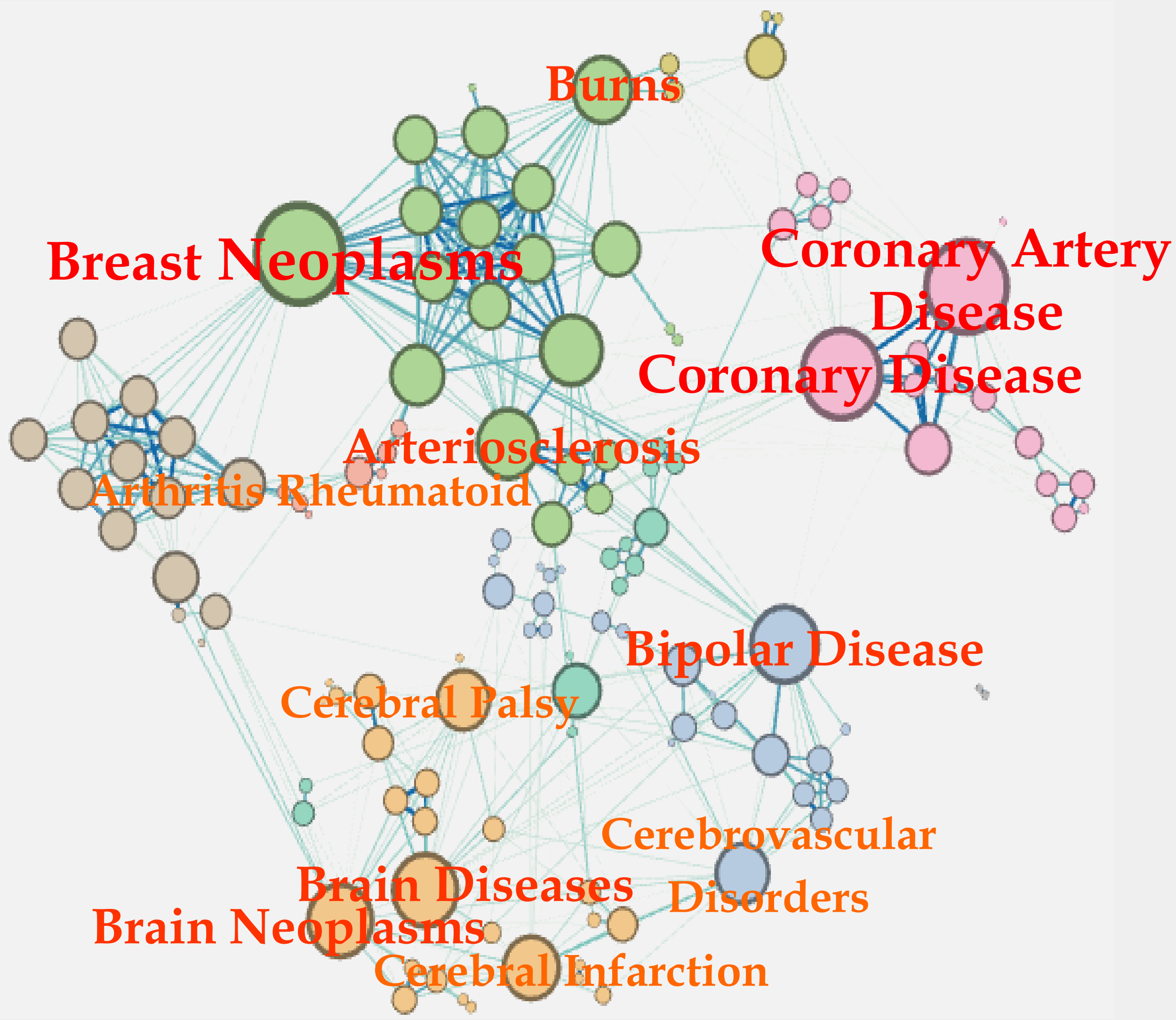




Average Degree: 7.679

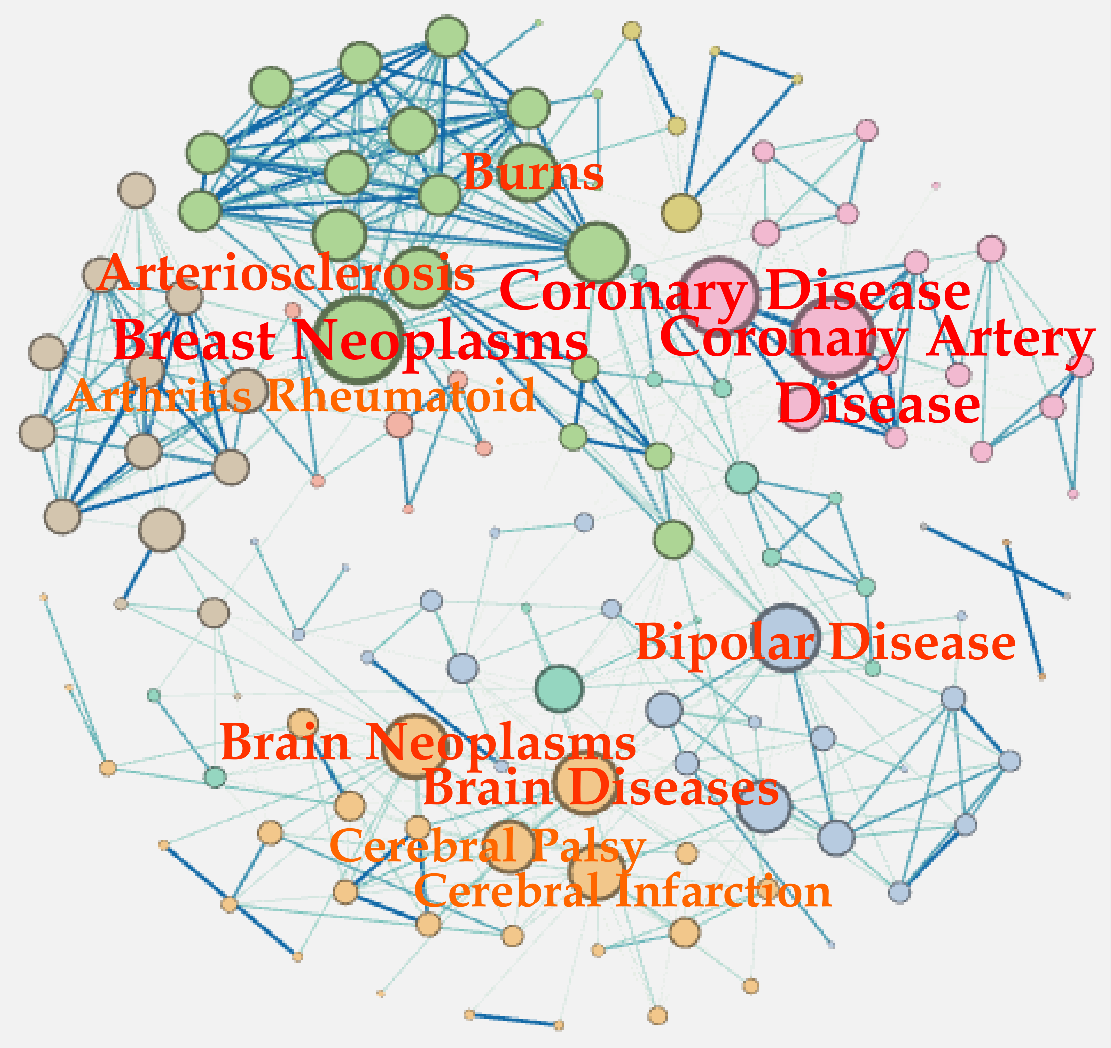
Average Weighted Degree: 3.424

Force Atlas 2 Layout Algorithm



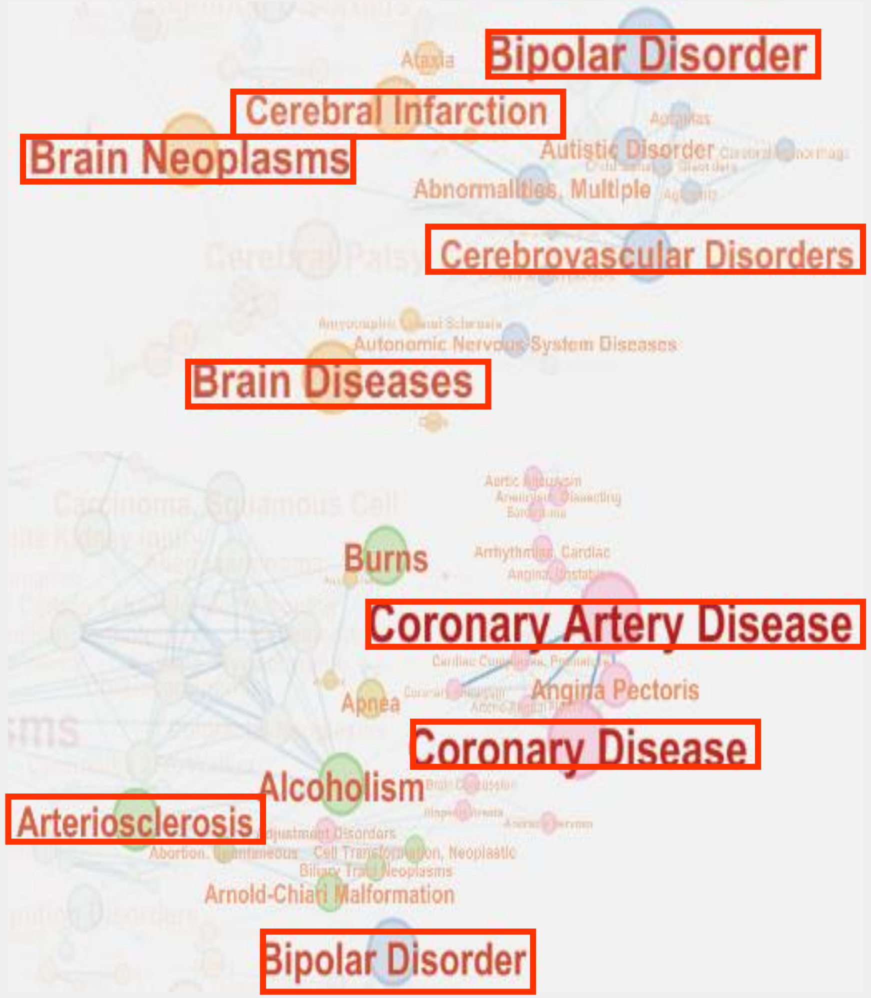
- Force Atlas 2 is an algorithm in the set of force-directed algorithms available in Gephi. It attempts to make a balance between the quality of the final layout and the speed of the computation algorithm.

Fruchterman Reingold Layout Algorithm



- The Fruchterman Reingold layout algorithm belongs to the class of force-directed algorithms. It is one of the standard algorithms in Gephi and is made use of quite often.
- In the Fruchterman Reingold layout algorithm, the nodes are assumed to be entities made of steel and the edges are assumed to be springs. The attractive force between the nodes mimics the spring force, whereas the repulsive force between the nodes is analogous to the electrical force.
- This algorithm does not take into consideration the edge weight to come up with an optimal layout.

Results



Brian

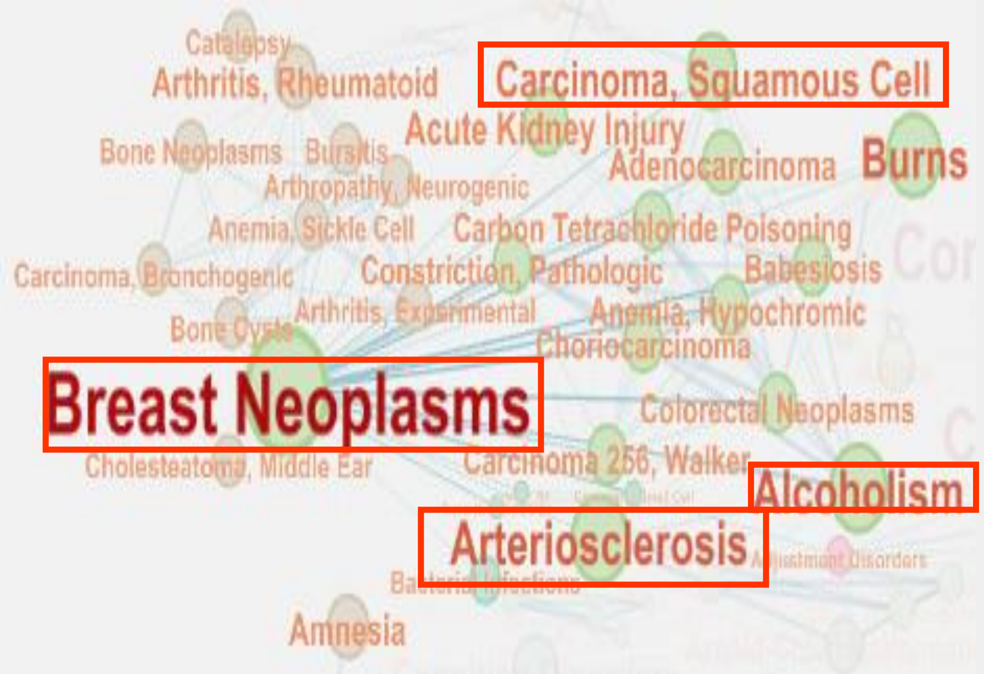
Adult

Arterial Blood Vessels

Bipolar Disorder → Coronary Disease

Arteriosclerosis

Results


Catalepsy
Arthritis, Rheumatoid
Carcinoma, Squamous Cell
Bone Neoplasms
Bursitis
Acute Kidney Injury
Adenocarcinoma
Burns
Arthropathy, Neurogenic
Anemia, Sickle Cell
Carbon Tetrachloride Poisoning
Carcinoma, Bronchogenic
Constriction, Pathologic
Babesiosis
Arthritis, Experimental
Anemia, Hypochromic
Bone Cysts
Choriocarcinoma
Breast Neoplasms
Colorectal Neoplasms
Cholesteatoma, Middle Ear
Carcinoma 256, Walker
Alcoholism
Arteriosclerosis
Adjustment Disorders
Bacterial Infections
Amnesia

✖

Why?

THANKS

FOR YOUR WATCHING