

소셜네트워크 분석

Jinseog Kim

Dongguk University

jskim1986@gmail.com

2018-03-13

Contents

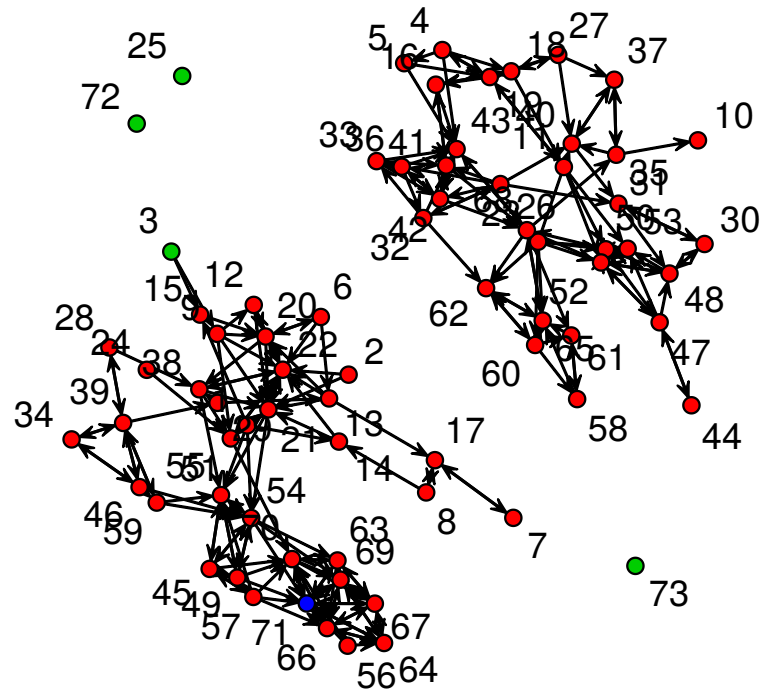
1	소셜네트워크 자료 예제	4
2	소셜 네트워크 및 자료	6
2.1	그래프(네트워크)의 구성	6
2.2	그래프의 표현	6
2.3	R에서 그래프 객체의 생성	8
2.4	다양한 그래프 외부 파일 형식(포맷)	9
2.5	Pajek(파이엑) 포맷	9
2.6	gml: 보다 단순한 XML	10
2.7	예시: 돌고래 네트워크	11
3	그래프의 속성 및 특성치	14
3.1	그래프의 속성	14
3.2	노드레벨 중심성(centrality) - 네트워크에 대한 노드의 영향력	15
3.3	R예제 : 노드레벨 중심성(centrality)	16
3.4	네트워크의 특성치 - 중심성과 밀도	18
3.5	R예제 - 네트워크의 특성치	19
4	네트워크의 군집분석 - 커뮤니티 탐색	21
5	Friendship network 예제	23

5.1 plot 23
5.2 중심성 - centrality 25
5.3 community 탐색 28

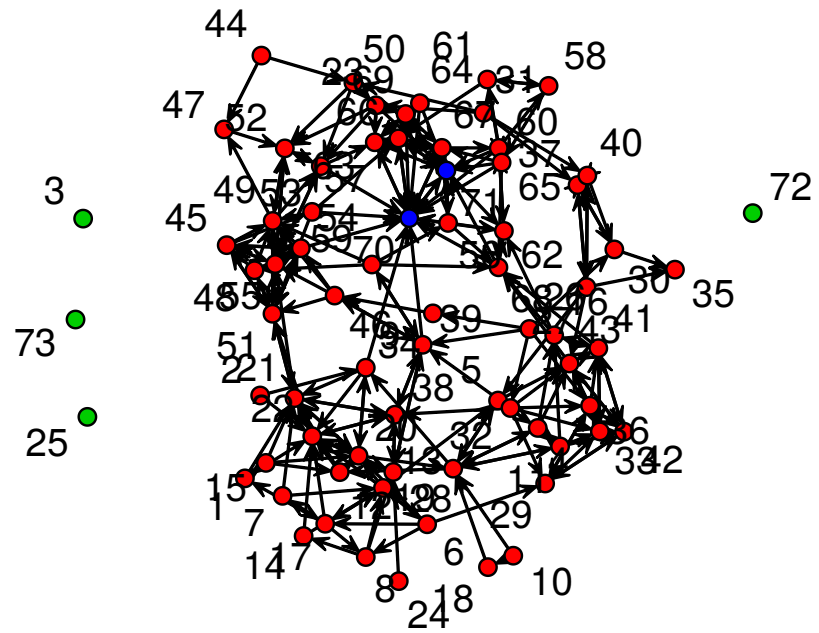
1 소셜네트워크 자료 예제

□ Coleman's High School Friendship Network

Fall(1)



Spring(1)



2 소셜 네트워크 및 자료

2.1 그래프(네트워크)의 구성

그래프 또는 네트워크, $G = G(V, E, W)$ 의 구성요소

- V 꼭지점(vertices), 노드 (node)
- E 변(edge), link : E 는 두 개의 노드 (a, b) 로 표현
 - (a, b) 는 노드 a 와 b 가 연결한다는 의미

□ Example

$$V = \{A, B, C, D, E, F, G\},$$
$$E = \{(1, 2), (1, 3), (2, 4), (4, 5), (3, 5), (4, 5), (5, 6), (6, 7)\}$$

2.2 그래프의 표현

□ 소시오-행렬(socio-matrix) 또는 인접행렬(adjacency matrix)

```
## 7 x 7 sparse Matrix of class "dgCMatrix"
```

```
##  A B C D E F G
## A . 1 1 . . . .
## B 1 . . 1 . . .
## C 1 . . . 1 . .
## D . 1 . . 1 1 .
## E . . 1 1 . . .
## F . . . 1 . . 1
## G . . . . . 1 .
```

□ 에지-리스트(edge-list): 에지를 노드의 쌍으로 표현한 two-컬럼 행렬

```
##      [,1] [,2]
## [1,] "A"  "B"
## [2,] "A"  "C"
## [3,] "B"  "D"
## [4,] "D"  "E"
## [5,] "C"  "E"
## [6,] "D"  "F"
## [7,] "F"  "G"
```

2.3 R에서 그래프 객체의 생성

1. 에지-리스트에서 igraph의 make_graph() 함수를 이용하여 그래프 객체를 형성

```
e1 <- c("A", "B", "A", "C", "B", "D", "D", "E", "C", "E", "D", "E", "F", "G")
g <- make_graph(e1, directed=FALSE)
g
```

```
## IGRAPH 8b98e8e UN-- 7 7 --
## + attr: name (v/c)
## + edges from 8b98e8e (vertex names):
## [1] A--B A--C B--D D--E C--E D--E F--G
```

2. 그래프 생성 문법을 이용 igraph의 graph_from_literal()

```
g2 <- graph_from_literal(A--B, A--C, B--D, D--E, C--E, D--E, F--G)
g2
```

```
## IGRAPH 0ebbabe UN-- 7 6 --
## + attr: name (v/c)
## + edges from 0ebbabe (vertex names):
## [1] A--B A--C B--D C--E D--E F--G
```


2.4 다양한 그래프 외부 파일 형식(포맷)

1. Pajek(파이엑) 포맷
2. graphML: XML형식
3. gml: graphML의 단순한 형태
4. 에지리스트

2.5 Pajek(파이엑) 포맷

*Vertices 9

*Edges

1 2

1 9

2 9

2 3

2 8

3 8

3 4

4 5

4 7

5 7

5 6

6 4

2.6 gml: 보다 단순한 XML

graph

```
[  
  directed 0  
  node  
  [  
    id 0  
    label "Kim"  
  ]  
  node  
  [  
    id 1  
    label "Lee"  
  ]  
  ...  
  edge  
  [  
    ]  
  ]
```

```
    source 1
    target 2
  ]
  ...
]
```

2.7 예시: 돌고래 네트워크

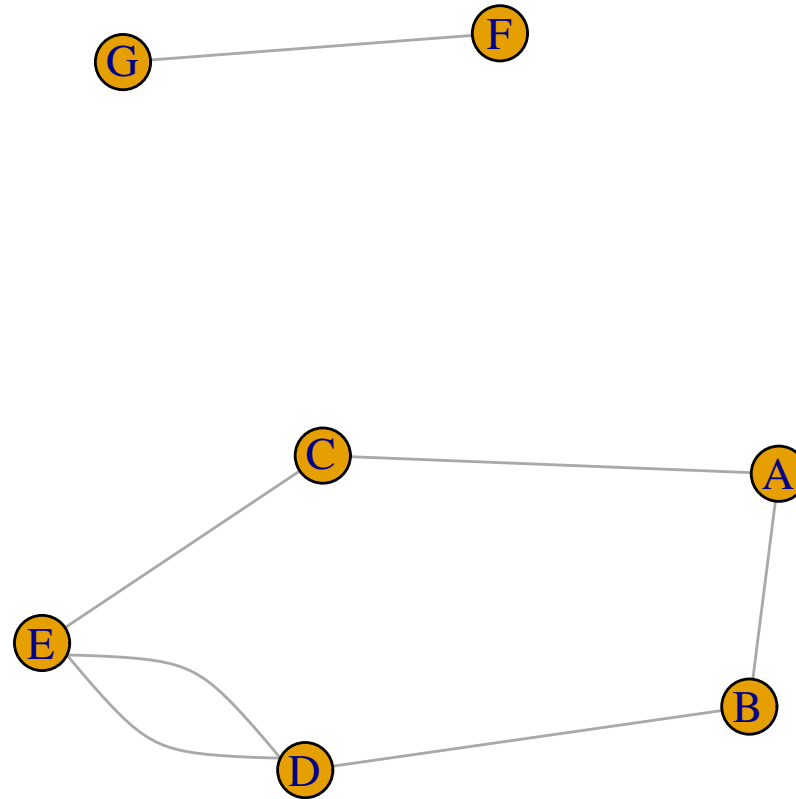
```
dolphins <- read.graph("dolphins.gml","gml")
dolphins
```

```
## IGRAPH ad98570 U--- 62 159 --
## + attr: id (v/n), label (v/c)
## + edges from ad98570:
## [1] 4-- 9 6--10 7--10 1--11 3--11 6--14 7--14 10--14 1--15 4--15
## [11] 1--16 15--17 2--18 7--18 10--18 14--18 16--19 2--20 8--20 9--21
## [21] 17--21 19--21 19--22 18--23 15--25 16--25 19--25 18--26 2--27 26--27
## [31] 2--28 8--28 18--28 26--28 27--28 2--29 9--29 21--29 11--30 19--30
## [41] 22--30 25--30 8--31 20--31 29--31 18--32 10--33 14--33 13--34 15--34
## [51] 17--34 22--34 15--35 34--35 30--36 2--37 21--37 24--37 9--38 15--38
```

```
## [61] 17--38 22--38 34--38 35--38 37--38 15--39 17--39 21--39 34--39 37--40
## [71] 1--41 8--41 15--41 16--41 34--41 37--41 38--41 2--42 10--42 14--42
## + ... omitted several edges
```

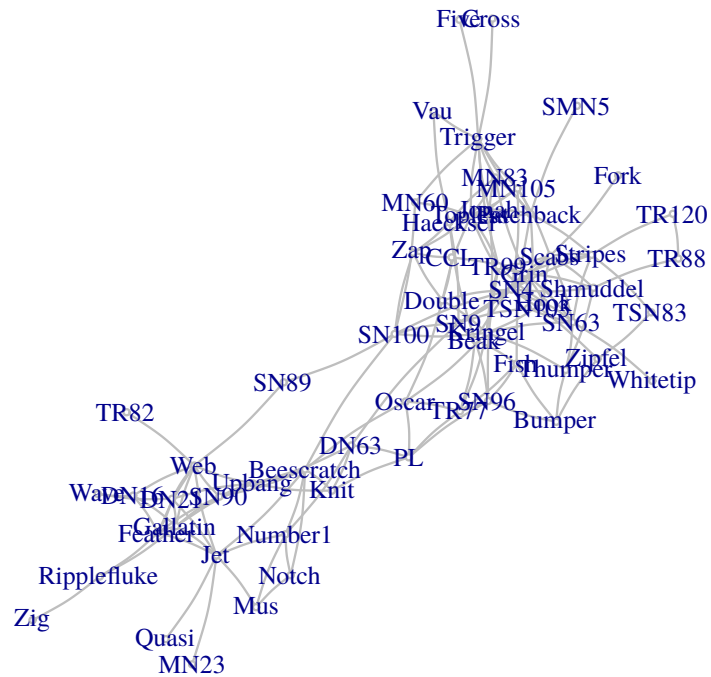
1. 시각화

```
plot(g)
```



```
plot(dolphins, edge.curved=.1, vertex.label.cex=0.7, edge.color="grey",
     vertex.frame.color="grey", vertex.size=0, vertex.color="white", asp=1,
     main=paste0("돌고래 네트워크"))
```

돌고래 네트워크



3 그래프의 속성 및 특성치

3.1 그래프의 속성

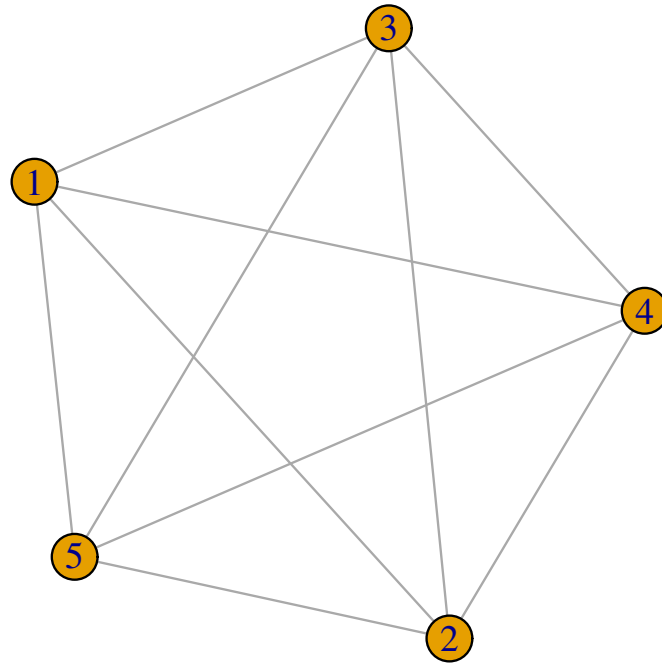
▣ size(크기) : 노드의 수 $|V|$

▣ order : 링크의 수 $|E|$

▣ 공그래프(empty graph) : $|E| = 0$

▣ 완전그래프(complete graph) : $n(n - 1)/2 = \binom{n}{2}$

complete graph



3.2 노드레벨 중심성(centrality) - 네트워크에 대한 노드의 영향력

1. 연결중심성 (degree) - 노드 n 의 인접노드(직접 연결된 노드)의 수

2. 근접중심성 (closeness) - 다른 노드들과의 최단거리의 합의 역수

$$C_n = (|V| - 1) / \left[\sum_{k=1}^{|V|} d(n, k) \right]$$

3. 매개중심성 (betweenness) - 서로 다른 두 노드의 최단거리 상에 자신의 노드가 포함되는 비율

$$B_n = \sum_{(j,k) \in E} d(j, n, k) / d(j, k)$$

4. 고유벡터중심성(eigen-vector) - Bonacich(1987)은 인접행렬의 고유벡터를 이용

$$e_n : A\mathbf{v} - \lambda_{max}\mathbf{v} = \mathbf{0},$$

여기서, A 는 인접행렬, λ_{max} 는 A 의 최대 고유치

5. PageRank - Brin and Page(1998) 인접행렬을 아래와 같이 마코프 추이행렬로 변환한 후, 정상분포를 구함

$$A^* = \alpha A + (1 - \alpha) \mathbf{1}\mathbf{1}' \frac{1}{n}, \quad \alpha = 0.85$$

3.3 R예제 : 노드레벨 중심성(centrality)


```
igraph::degree(g)
```

```
## A B C D E F G  
## 2 2 2 3 3 1 1
```

```
igraph::closeness(g)
```

```
##           A           B           C           D           E           F  
## 0.05000000 0.05000000 0.05000000 0.05000000 0.05000000 0.02777778  
##           G  
## 0.02777778
```

```
igraph::betweenness(g)
```

```
## A B C D E F G  
## 1 1 1 1 1 0 0
```

```
centr_eigen(g)$vector
```

```
## [1] 0.4384472 0.5615528 0.5615528 1.0000000 1.0000000 0.0000000 0.0000000
```

```
page_rank(g)$vector
```

```
##           A           B           C           D           E           F           G  
## 0.1263561 0.1234441 0.1234441 0.1705207 0.1705207 0.1428571 0.1428571
```

3.4 네트워크의 특성치 - 중심성과 밀도

▣ 그래프 중심성 (centralization) : 전체 네트워크에서 가장 영향력 있는 노드의 다른 노드들에 대한 상대적인 영향력을 측정

$$C_D = \frac{\sum_{i=1}^{|V|} (C_D^* - C_D(i))}{(|V| - 1)(|V| - 2)}$$

- $C_D(i)$: 노드레벨 중심성, C_D^{max} 노드레벨 중심성 최대값

▣ 밀도(density) : 네트워크내에 있는 노드들의 밀집도를 평가

$$\Delta = \frac{\text{네트워크의 링크수}}{\text{가능한 모든 링크 수}} = \frac{|E|}{|V|(|V| - 1)/2} = \frac{\bar{d}}{|V| - 1}$$

3.5 R예제 - 네트워크의 특성치

▣ 그래프 중심성

```
g1 <- graph_from_adjacency_matrix(coleman[1,,])  
g2 <- graph_from_adjacency_matrix(coleman[2,,])  
centr_degree(g1)$centralization
```

```
## [1] 0.06577932
```

```
centr_degree(g2)$centralization
```

```
## [1] 0.09008488
```

```
centr_clo(g1)$centralization
```

```
## [1] 0.004962646
```

```
centr_clo(g2)$centralization
```

```
## [1] 0.01617278
```

```
#centr_betw(g)$centralization
```

```
#centr_eigen(g)$centralization
```

▣ 그래프 밀도

```
edge_density(g1)
```

```
## [1] 0.04623288
```

```
edge_density(g2)
```

```
## [1] 0.05003805
```

4 네트워크의 군집분석 - 커뮤니티 탐색

- 네트워크의 군집분석: 네트워크의 노드들의 연결강도를 이용하여 서로 밀집된 서브-그래프를 탐색
- 모듈성(modularity, Newman, 2006): 서브그래프의 노드들간 연결 수에서 동일 노드들이 무작위로 가질 수 있는 링크수의 기댓값을 뺀 것으로 양의 값이 클수록, 네트워크에 커뮤니티가 존재할 가능성이 높음

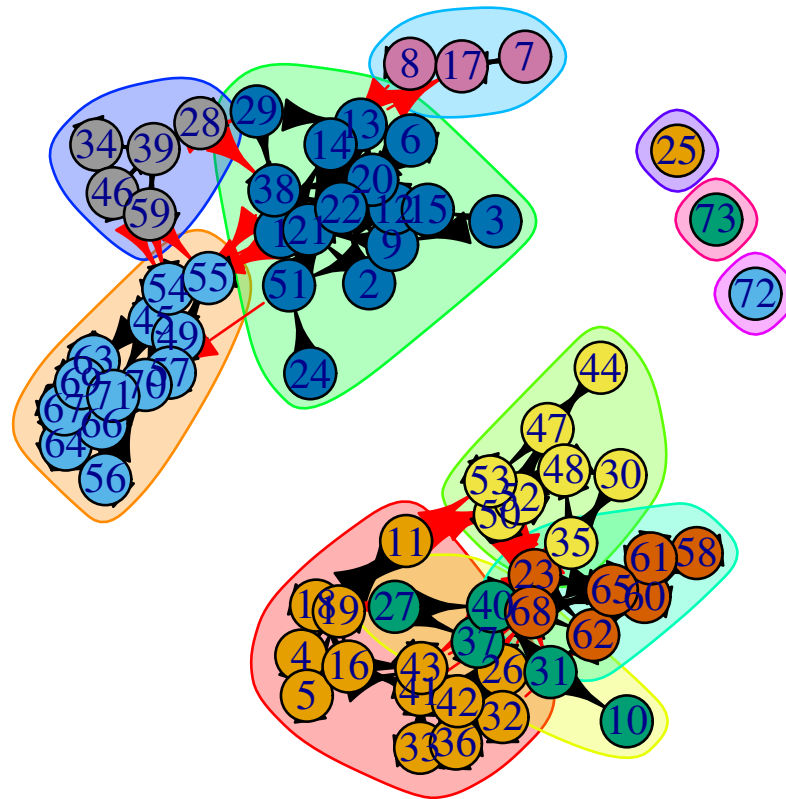
$$Q = \frac{1}{4m} \sum \left((A_{ij} - \frac{k_i k_j}{2m}) (s_i s_j - 1) \right)$$

```
wc <- cluster_walktrap(g1)
modularity(wc); membership(wc)
```

```
## [1] 0.6851006
```

```
## 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25
## 5 5 5 1 1 5 7 7 5 3 1 5 5 5 5 1 7 1 1 5 5 5 6 5 9
## 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 42 43 44 45 46 47 48 49 50
## 1 3 8 5 4 3 1 1 8 4 1 3 5 8 3 1 1 1 4 2 8 4 4 2 4
## 51 52 53 54 55 56 57 58 59 60 61 62 63 64 65 66 67 68 69 70 71 72 73
## 5 4 4 2 2 2 2 6 8 6 6 6 2 2 6 2 2 6 2 2 2 10 11
```

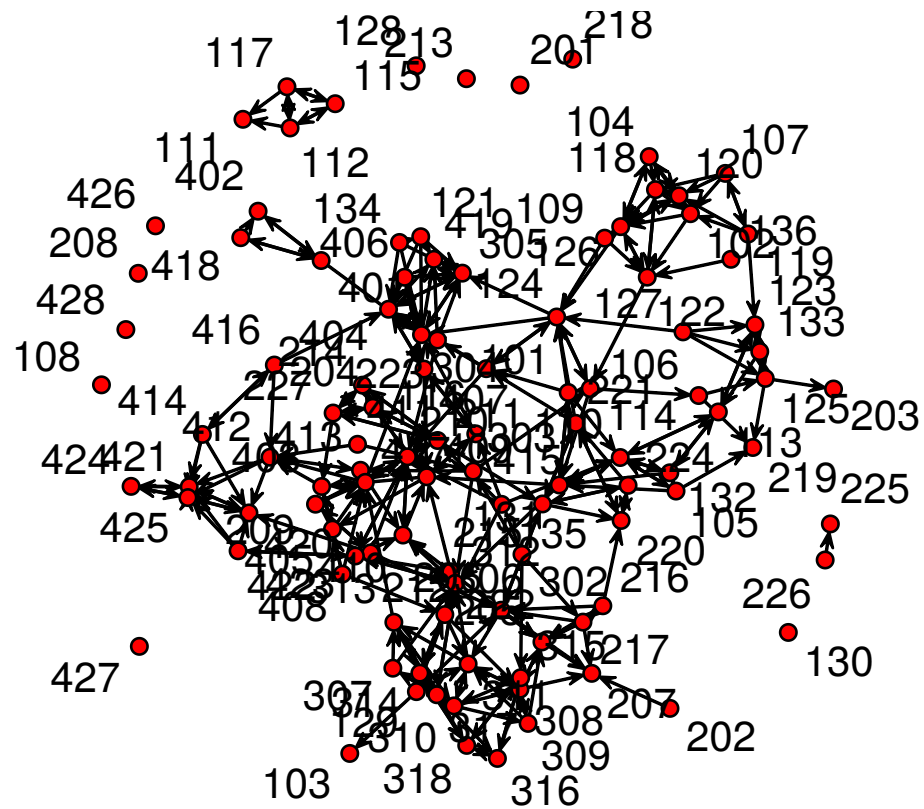
plot(wc, g1)



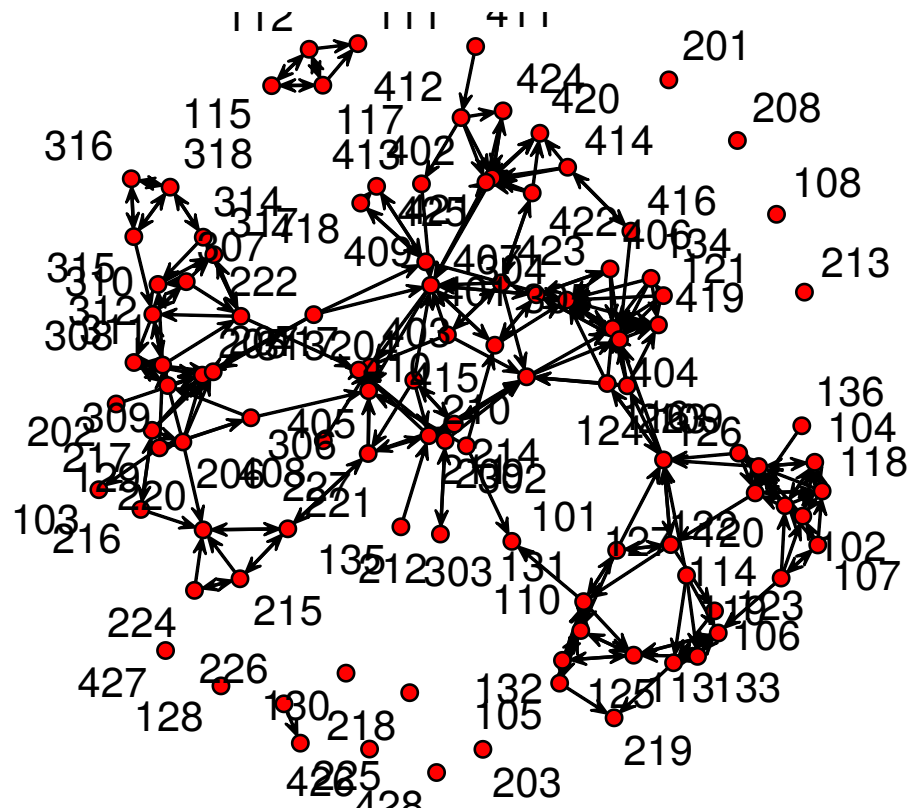
5 Friendship network 예제

5.1 plot

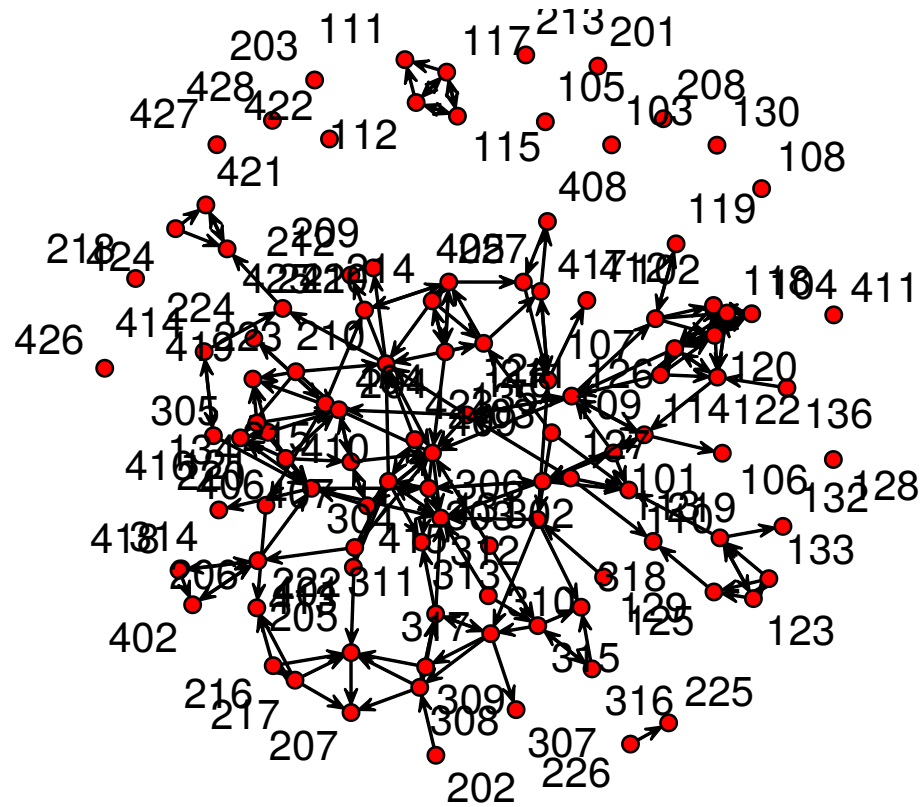
1. Q1: 감정네트워크(기쁠 때)



2. Q2: 감정네트워크(슬플 때)



3. Friendship network : Q3: 협업네트워크



5.2 중심성 - centrality

1. Q1: 기별 때

	id	grade	sex	in.degree	betweenness	closeness
66	304	3	남	9	338.59762	0

	id	grade	sex	in.degree	betweenness	closeness
84	404	4	남	9	515.89841	0
46	210	2	남	8	1657.30000	0
74	312	3	여	8	1312.47857	0
24	124	1	여	7	1157.04206	0
67	305	3	남	7	26.89127	0
10	110	1	여	6	374.28333	0
20	120	1	남	6	28.16667	0
26	126	1	남	6	227.91905	0
41	205	2	남	6	24.43333	0

2. Q2: 슬플 때

	id	grade	sex	in.degree
84	404	4	남	10
89	409	4	남	9
66	304	3	남	8
67	305	3	남	8
46	210	2	남	7
24	124	1	여	6
26	126	1	남	6

	id	grade	sex	in.degree
41	205	2	남	6
85	405	4	남	6
105	425	4	여	6

3. Q3: 협업

	id	grade	sex	in.degree
74	312	3	여	8
89	409	4	남	8
26	126	1	남	6
46	210	2	남	6
66	304	3	남	6
84	404	4	남	6
10	110	1	여	5
20	120	1	남	5
22	122	1	남	5
24	124	1	여	5

5.3 community 탐색

