데이터 과학 기초

08

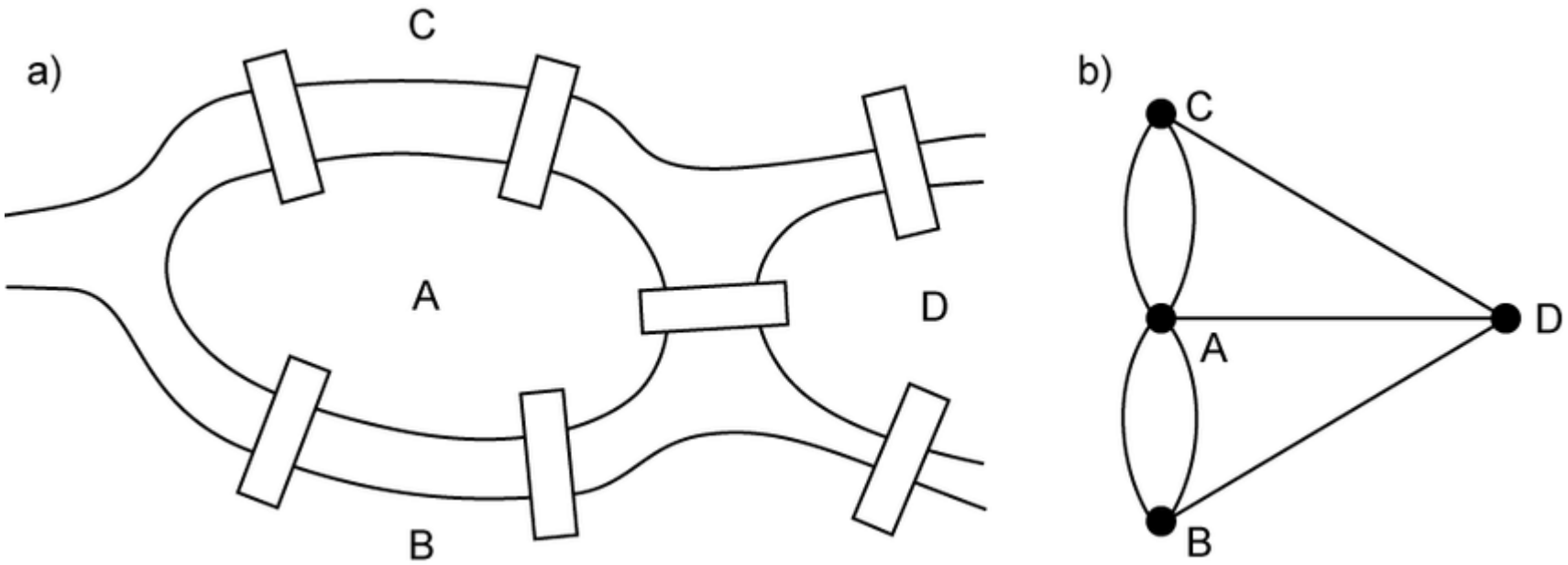
네트워크 분석

경북대학교 배준현 교수

(joonion@knu.ac.kr)

08. 네트워크 분석

- **그래프 이론:** *Graph Theory*
  - 정점의 집합과 간선의 집합으로 구성된 그래프를 연구하는 수학의 한 분야
  - 그래프: $G = (V, E)$
    - V: 정점의 집합, E: 간선의 집합

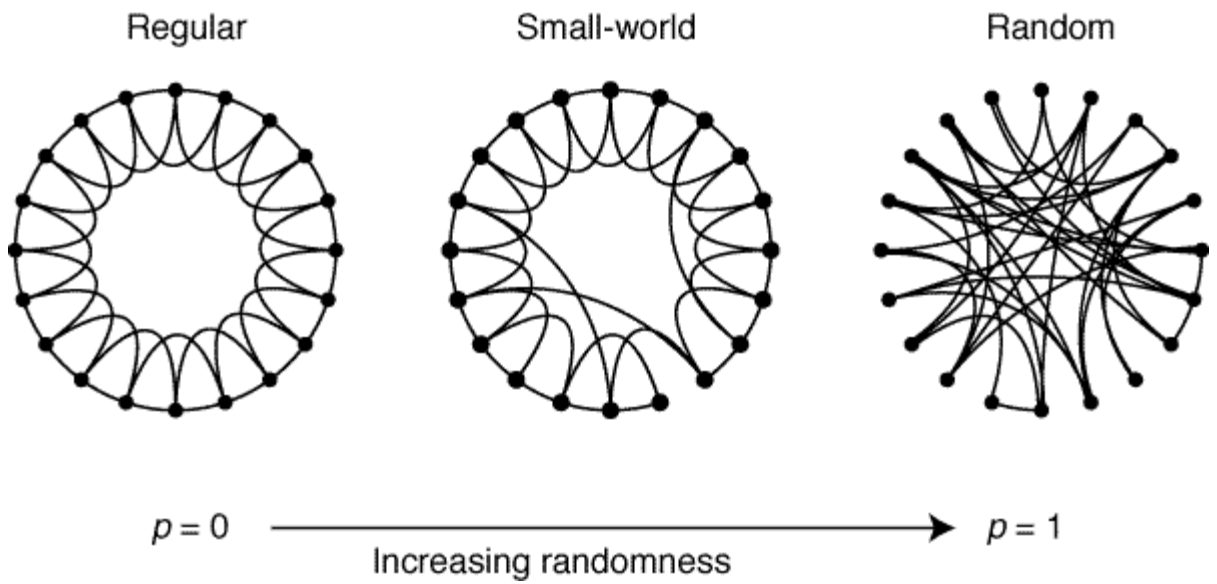


08. 네트워크 분석

- **네트워크 과학: *Network Science***
  - 다양한 학문 분야에 펼쳐져 있던 **복잡계**의 연구 대상들이
    - '**네트워크**'라는 하나의 주제로 통일되면서 발생한 **학제간** 연구 분야
  - 복잡계 네트워크: *Complex Network*
    - 사회 현상의 탐구: 소셜 네트워크 - 사람과 사람 사이의 관계 분석
    - 생명 현상의 탐구: 단백질 네트워크 - 분자와 분자 사이의 관계 분석
    - 자연 현상의 탐구: 상전이 현상 – 네트워크 동역학으로 분석 가능

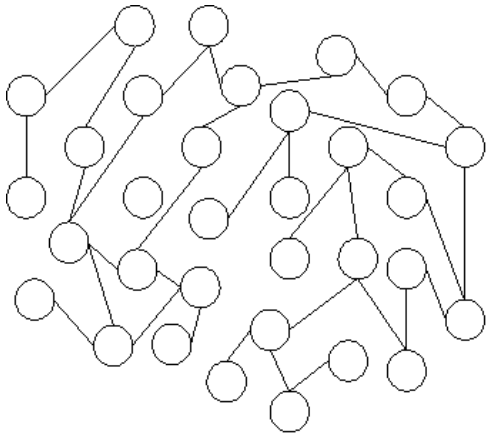
08. 네트워크 분석

- **작은 세상 네트워크: *Small-World* Network**
  - 6단계의 분리: 세상이 참 좁은 이유에 대한 과학적 설명
  - 소수의 원거리 연결만으로도 전체 네트워크의 평균 거리가 크게 짧아짐

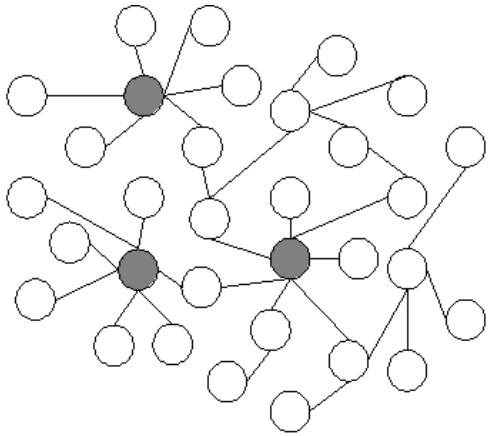


08. 네트워크 분석

- **척도 없는 네트워크: *Scale-Free* Network**
  - **빈익빈 부익부: 네트워크에서 허브가 생겨나는 이유에 대한 과학적 설명**
  - **선호적 연결에 의해 평균 이상으로 많은 링크를 가진 허브가 탄생함**

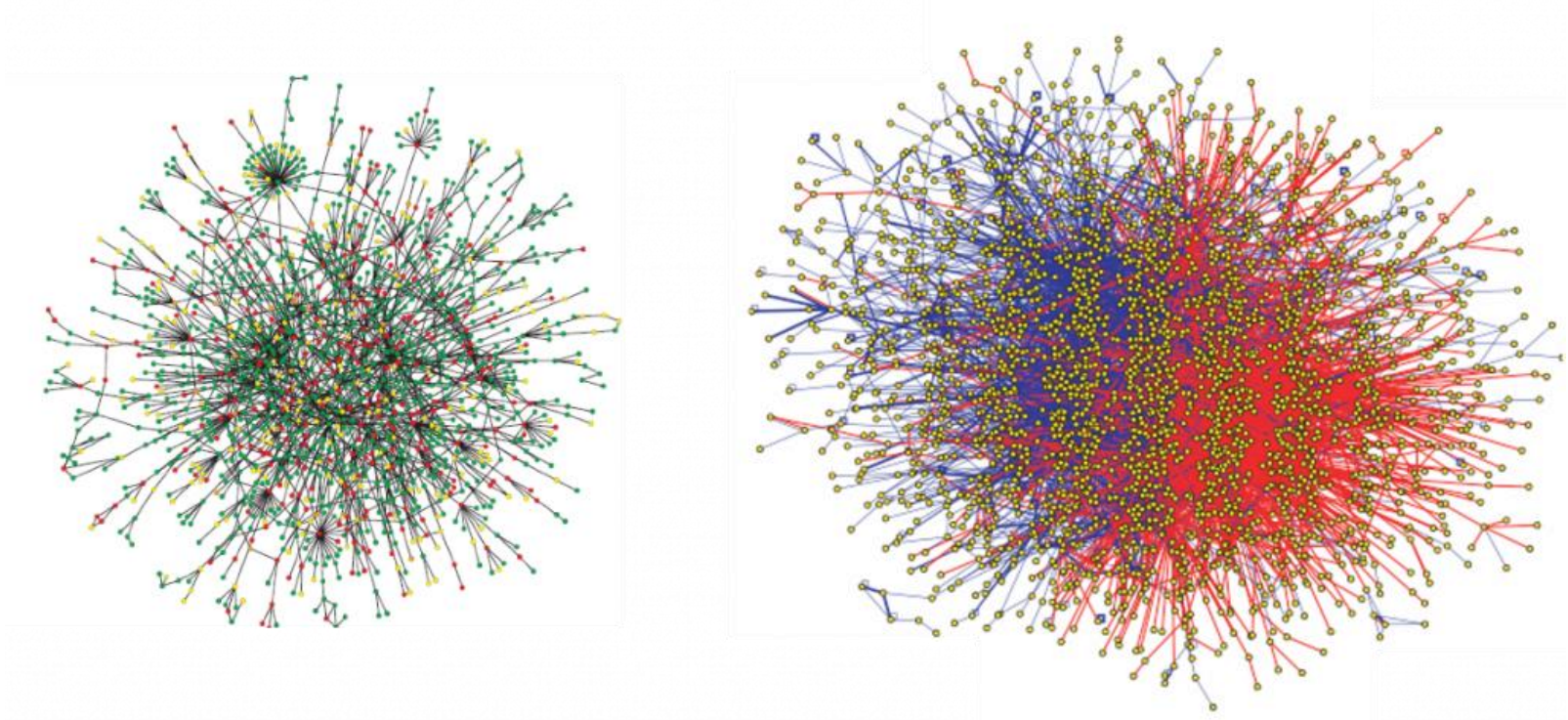


(a) Random network



(b) Scale-free network

- **단백질 접힘 문제:** *Protein Folding* Problem
  - 단백질의 아미노산 서열이 어떻게 고유한 접힌 구조를 결정하는가?
  - 주어진 아미노산 서열만으로 단백질의 접힌 구조를 예측할 수 있을까?

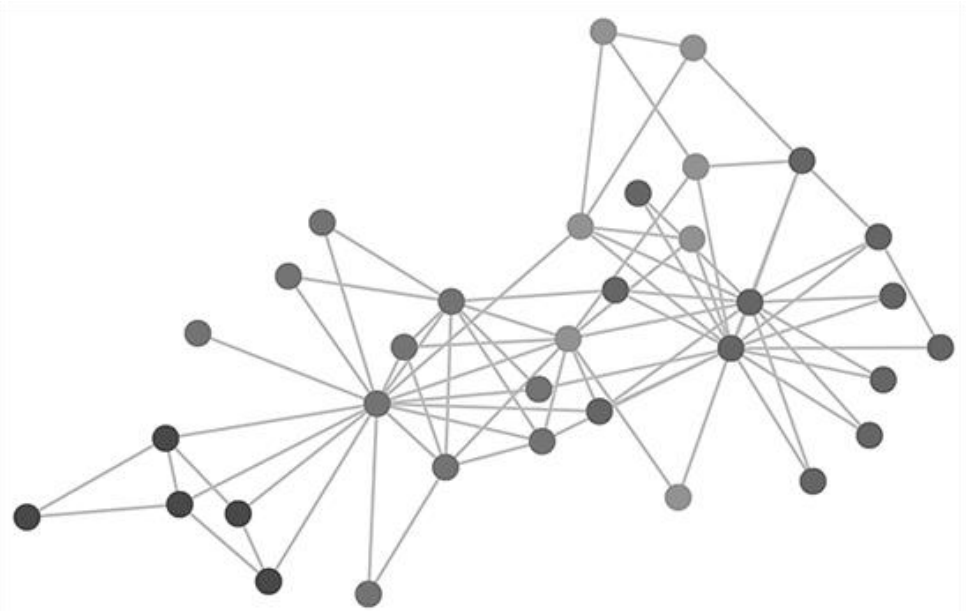


08. 네트워크 분석

- **네트워크 분석**: *Network Analysis*
  - 자연과학, 생명과학, 사회과학 등 모든 분야를 아우르는 데이터 과학 분야
    - 소셜 네트워크 분석: *SNA*, Social Network Analysis
    - 복잡계 네트워크 분석: Complex Network Analysis
  - 주요 탐구 주제:
    - 중심성 분석: *Centrality* Analysis
    - 군집 분석: *Cluster* Analysis

08. 네트워크 분석

- **네트워크 분석을 위한 기본 용어:**
  - **방향 그래프, 무방향 그래프:** *directed* and undirected
  - **가중치 그래프, 가중치 없는 그래프:** *weighted* and unweighted
  - **인접 행렬, 인접 리스트:** *adjacency* matrix and list
  - **단순 경로, 경로 길이, 랜덤 워크:** simple *path*, path length, *random walk*

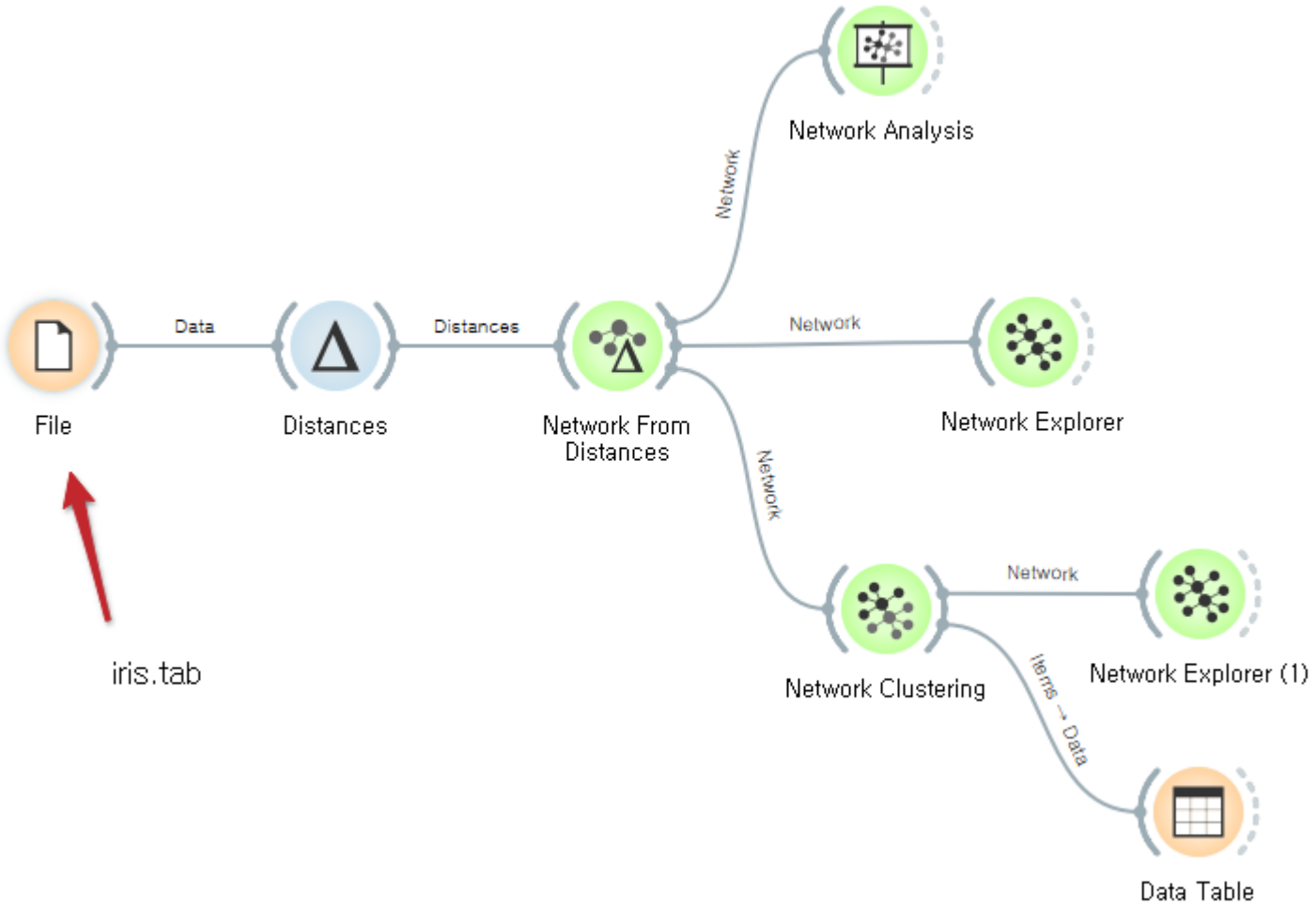


08. 네트워크 분석

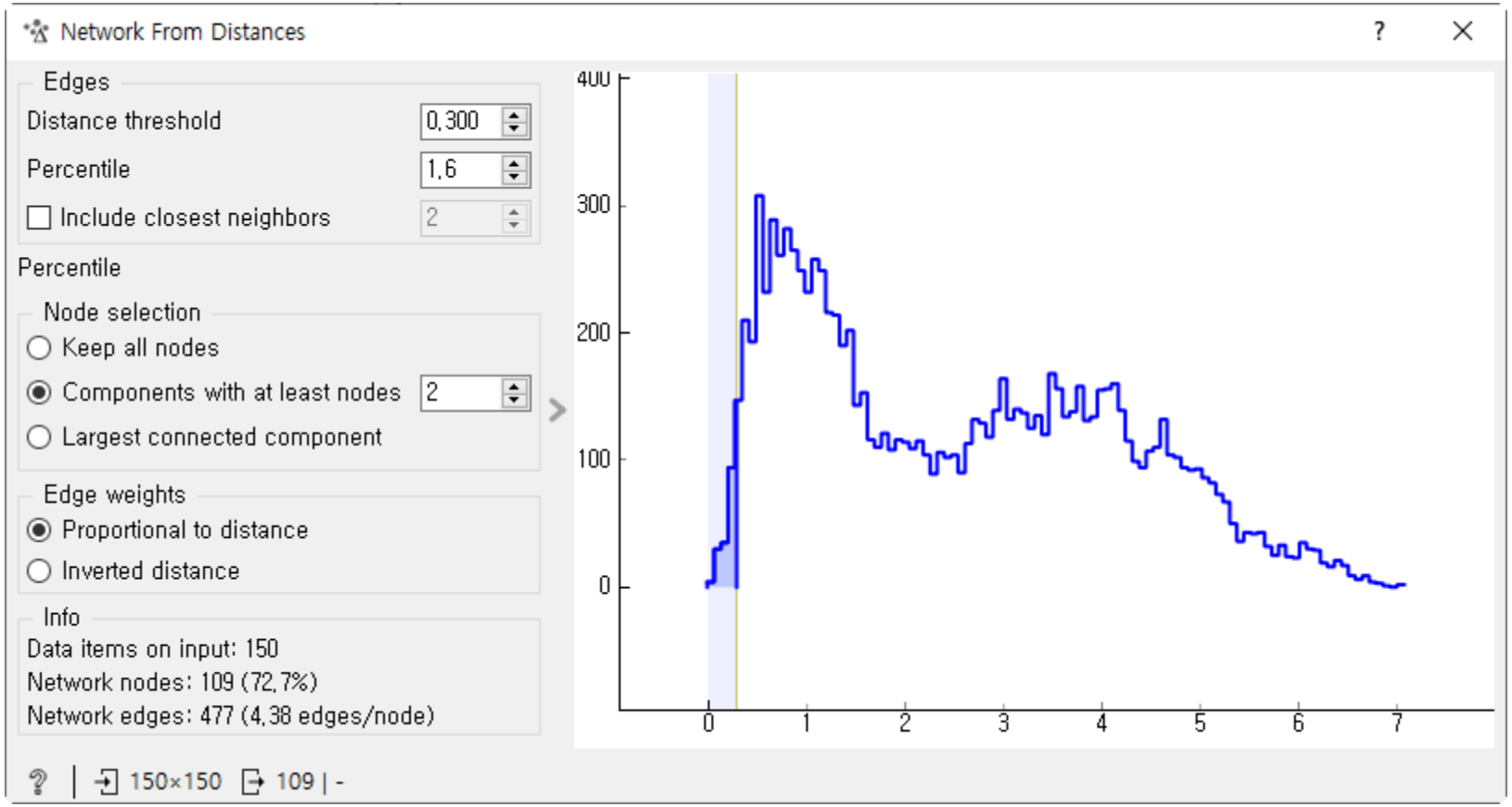
- **네트워크 분석을 위한 기본 정보:**
  - **노드(정점)와 연결(간선)의 수**: Number of *nodes* and *edges*
  - **평균 차수**: Average *degree*
  - **밀도, 반경, 반지름**: *Density*, *Diameter*, Radius
  - **평균 최단거리**: Average *shortest path* length
  - **강연결/약연결 컴포넌트**:
    - Number of *Strongly Connected* Components (SCC)
    - Number of *Weakly Connected* Components (WCC)

08. 네트워크 분석

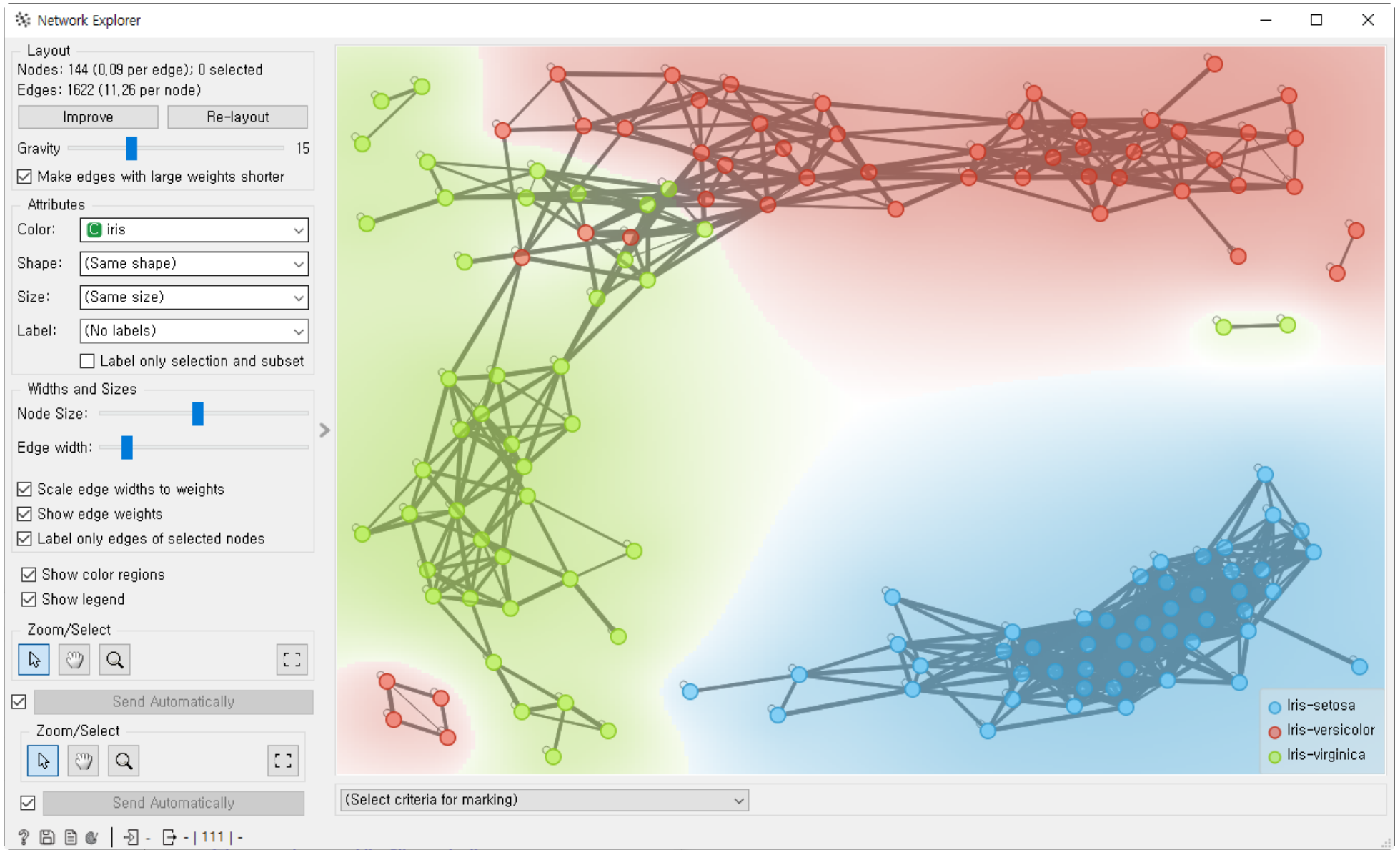
- Orange: Networks



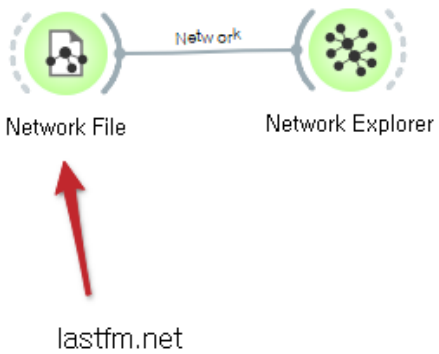
08. 네트워크 분석

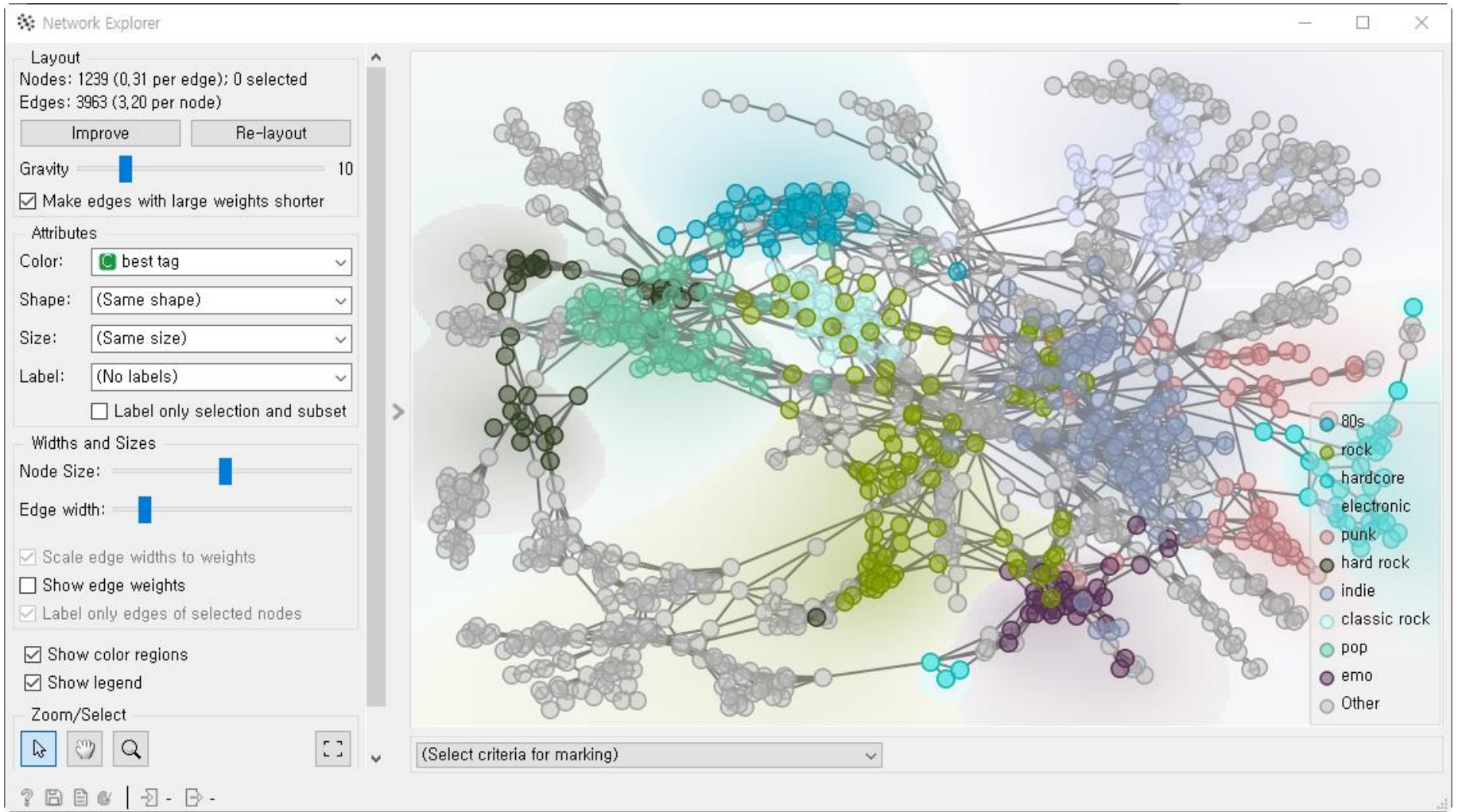

Network From Distances
Edges
Distance threshold
0,300
Percentile
1,6
Include closest neighbors
2
Percentile
Node selection
Keep all nodes
Components with at least nodes
2
Largest connected component
Edge weights
Proportional to distance
Inverted distance
Info
Data items on input: 150
Network nodes: 109 (72,7%)
Network edges: 477 (4,38 edges/node)
150×150
109 | -
400
300
200
100
0
0
1
2
3
4
5
6
7

08. 네트워크 분석


Network Explorer
Layout
Nodes: 144 (0,09 per edge); 0 selected
Edges: 1622 (11,26 per node)
Improve
Re-layout
Gravity
15
Make edges with large weights shorter
Attributes
Color:
iris
Shape:
(Same shape)
Size:
(Same size)
Label:
(No labels)
Label only selection and subset
Widths and Sizes
Node Size:
Edge width:
Scale edge widths to weights
Show edge weights
Label only edges of selected nodes
Show color regions
Show legend
Zoom/Select
Send Automatically
Zoom/Select
Send Automatically
Iris-setosa
Iris-versicolor
Iris-virginica
(Select criteria for marking)

08. 네트워크 분석


Network
Network File
Network Explorer
lastfm.net

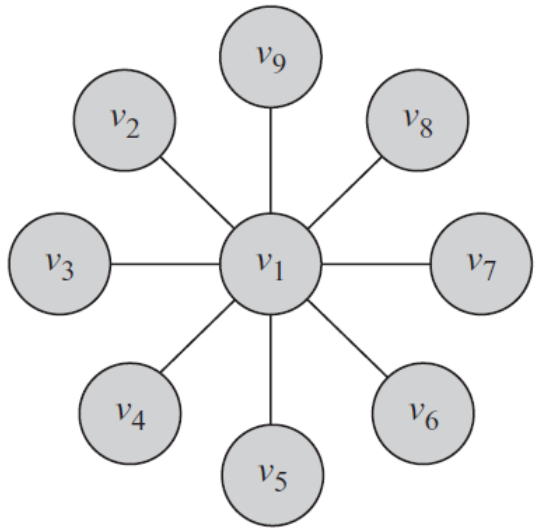

Network Explorer
Layout
Nodes: 1239 (0,31 per edge); 0 selected
Edges: 3963 (3,20 per node)
Improve
Re-layout
Gravity
10
Make edges with large weights shorter
Attributes
Color:
best tag
Shape:
(Same shape)
Size:
(Same size)
Label:
(No labels)
Label only selection and subset
Widths and Sizes
Node Size:
Edge width:
Scale edge widths to weights
Show edge weights
Label only edges of selected nodes
Show color regions
Show legend
Zoom/Select
(Select criteria for marking)
80s
rock
hardcore
electronic
punk
hard rock
indie
classic rock
pop
emo
Other

08. 네트워크 분석

- **중심성 분석:** *Centrality* Analysis
  - 네트워크에서 중심부에 위치한 (가장 중요한) 노드를 찾기 위한 방법
    - 예) 인스타그램에서 10대 남성에게 가장 영향력이 높은 인플루언서는?
  - 중심성 지수: Centrality Measures
    - 차수 중심성: *Degree* Centrality
    - 인접 중심성: *Closeness* Centrality
    - 매개 중심성: *Betweenness* Centraliity
    - 페이지랭크: PageRank

08. 네트워크 분석

- **차수 중심성:** *Degree* Centrality
  - **한 노드에 연결된 모든 연결의 개수로 중심성을 평가하는 지수**
    - $C_d(v_i) = d_i$
  - **방향 그래프에서는** in-degree**와** out-degree**로 구분**



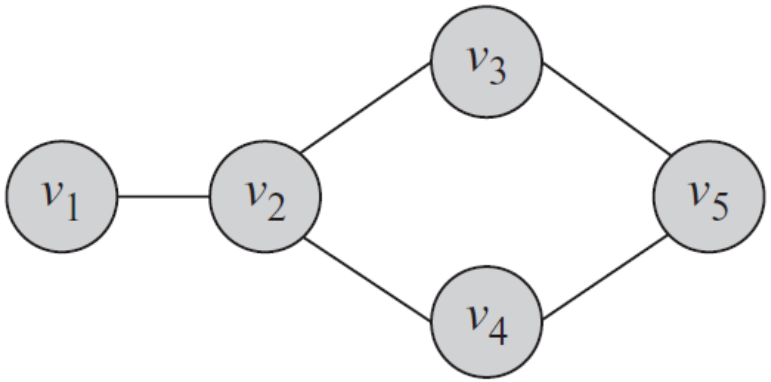
Source: Zafarani, Reza, Mohammad Ali Abbasi, and Huan Liu. *Social media mining: an introduction*. Cambridge University Press, 2014.

08. 네트워크 분석

- **인접 중심성:** *Closeness* Centrality
  - 한 노드가 다른 노드들로 가는 최단 경로가 얼마나 짧은 지를 평가
    - $C_c(v_i) = \frac{1}{\bar{l}_{v_i}}, \bar{l}_{v_i} = \frac{1}{n-1}\sum_{v_j \neq v_i} l_{i,j},$
      - $\bar{l}_{v_i}$: 노드 v_i에서 다른 모든 노드들로 가는 최단 경로 길이의 평균
  - 가정: 다른 모든 노드들로 가는 경로가 짧을수록 중심성이 높을 것이다.
    - 평균 최단 경로 길이가 짧은 노드일수록 중심성이 높다.

매개 중심성: *Betweenness* Centrality

- 한 노드가 네트워크의 다른 노드 간의 연결에 얼마나 기여하는 지를 평가
  - $C_b(v_i) = \sum_{s \neq t \neq v_i} \frac{\sigma_{st}(v_i)}{\sigma_{st}} d_i$,
    - σ_{st}: 노드 s에서 t로 가는 최단 경로의 개수
    - $\sigma_{st}(v_i)$: 노드 s에서 t로 가는 최단 경로 중 의 v_i를 거쳐 가는 경로의 개수



$$C_b(v_2) = 2 \times (\underbrace{(1/1)}_{s=v_1,t=v_3} + \underbrace{(1/1)}_{s=v_1,t=v_4} + \underbrace{(2/2)}_{s=v_1,t=v_5} + \underbrace{(1/2)}_{s=v_3,t=v_4} + \underbrace{0}_{s=v_3,t=v_5} + \underbrace{0}_{s=v_4,t=v_5})$$
$$= 2 \times 3.5 = 7, \quad (3.34)$$
$$C_b(v_3) = 2 \times (\underbrace{0}_{s=v_1,t=v_2} + \underbrace{0}_{s=v_1,t=v_4} + \underbrace{(1/2)}_{s=v_1,t=v_5} + \underbrace{0}_{s=v_2,t=v_4} + \underbrace{(1/2)}_{s=v_2,t=v_5} + \underbrace{0}_{s=v_4,t=v_5})$$
$$= 2 \times 1.0 = 2, \quad (3.35)$$

Source: Zafarani, Reza, Mohammad Ali Abbasi, and Huan Liu. *Social media mining: an introduction*. Cambridge University Press, 2014.

08. 네트워크 분석

페이지랭크: *PageRank*

- 고유벡터 중심성: *Eigenvector* Centrality
  - 단순히 친구가 많은 노드와 핵인싸 친구가 많은 노드를 비교하면?
  - $\lambda \boldsymbol{C}_e = A^T \boldsymbol{C}_e$, λ는 상수, $\boldsymbol{C}_e$는 인접행렬 A^T의 고유벡터
- Katz 중심성: 고유벡터 중심성을 방향 그래프에 적용
  - $\boldsymbol{C}_{\text{Katz}} = \alpha A^T \boldsymbol{C}_{\text{Katz}} + \beta \mathbf{1}$, $\mathbf{1}$은 모든 원소가 1인 벡터
- 페이지랭크 중심성:
  - 핵인싸와 친구인 노드가 모두 인싸라고 할 수 있을까?
  - 특정 노드의 영향력은 그 노드의 차수에 반비례해서 전파되어야 함
  - $\boldsymbol{C}_{\text{PageRank}} = \alpha A^T D^{-1} \boldsymbol{C}_{\text{PageRank}} + \beta \mathbf{1}$,
    - $D = diag(d_1^{out}, d_2^{out}, \cdots, d_n^{out})$: 차수의 주대각선 행렬

- **중심성 지수의 비교:**

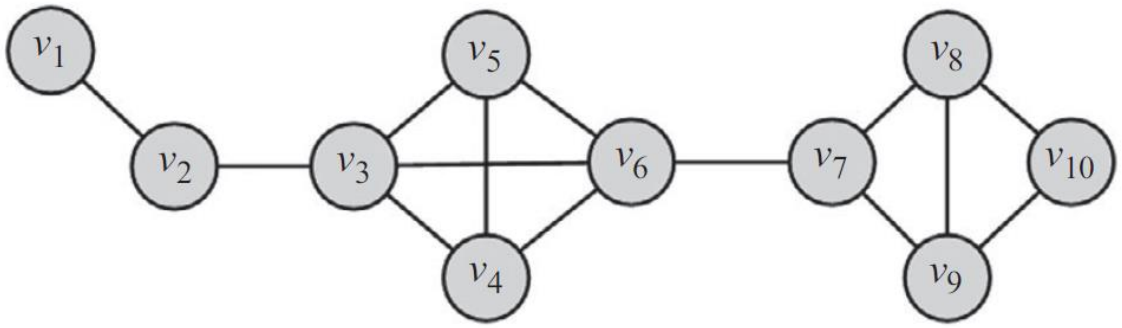


Table 3.1. *A Comparison between Centrality Methods*

	First node	Second node	Third node
Degree Centrality	v_3 or v_6	v_6 or v_3	$v \in \{v_4, v_5, v_7, v_8, v_9\}$
Eigenvector Centrality	v_6	v_3	v_4 or v_5
Katz Centrality: $\alpha = \beta = 0.3$	v_6	v_3	v_4 or v_5
PageRank: $\alpha = \beta = 0.3$	v_3	v_6	v_2
Betweenness Centrality	v_6	v_7	v_3
Closeness Centrality	v_6	v_3 or v_7	v_7 or v_3

08. 네트워크 분석

- **페이지랭크와 구글 검색엔진:**

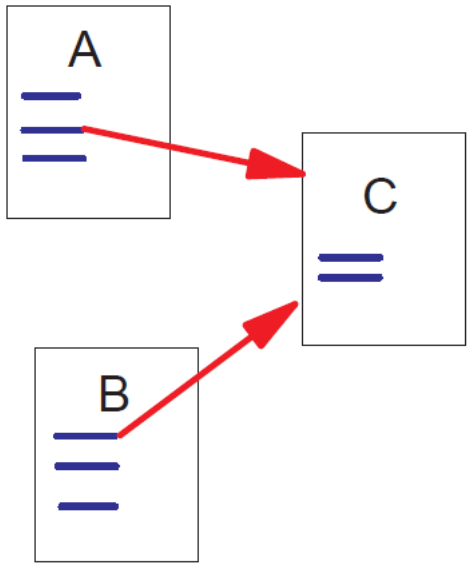


Figure 1: A and B are Backlinks of C

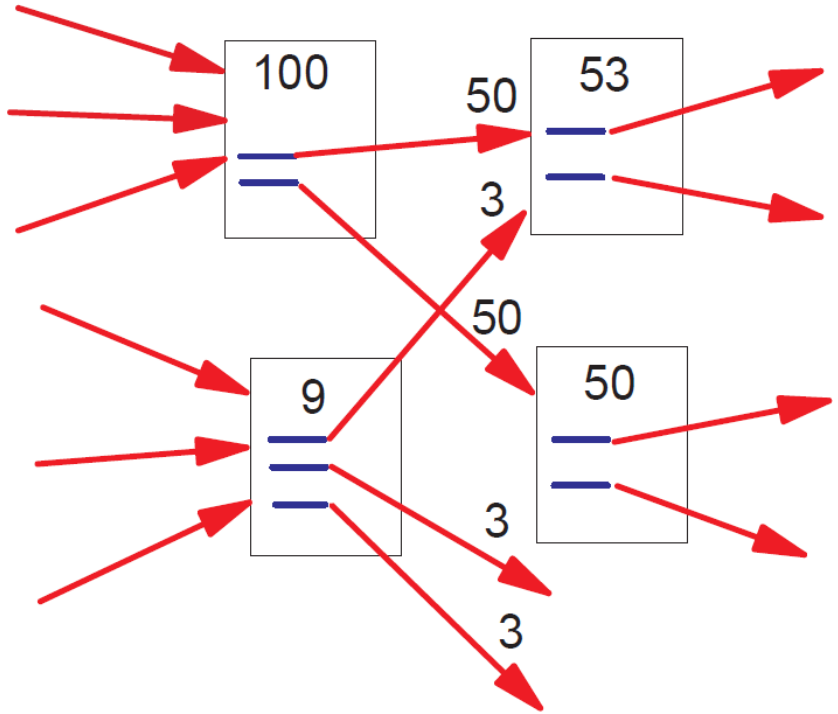


Figure 2: Simplified PageRank Calculation

Source: Brin, Sergey, and Lawrence Page. "The anatomy of a large-scale hypertextual web search engine." (1998).

08. 네트워크 분석

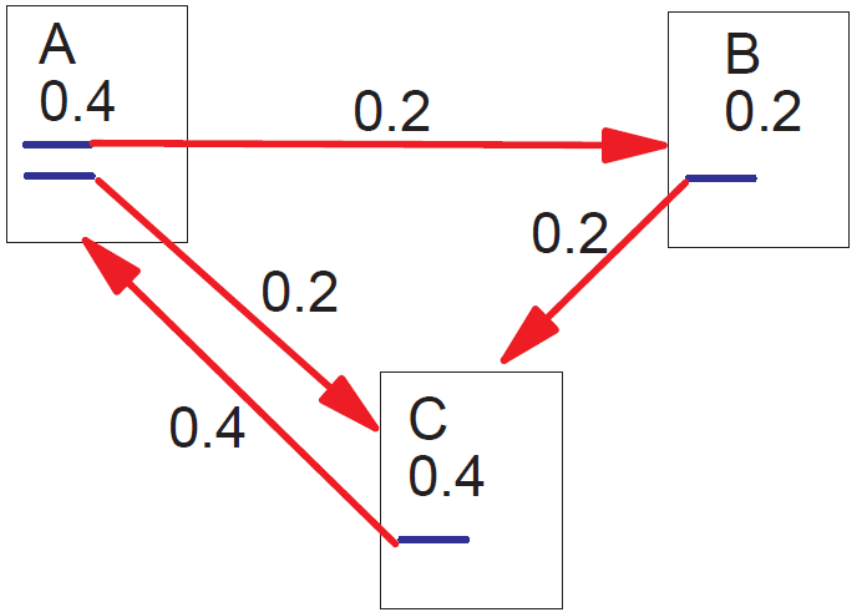
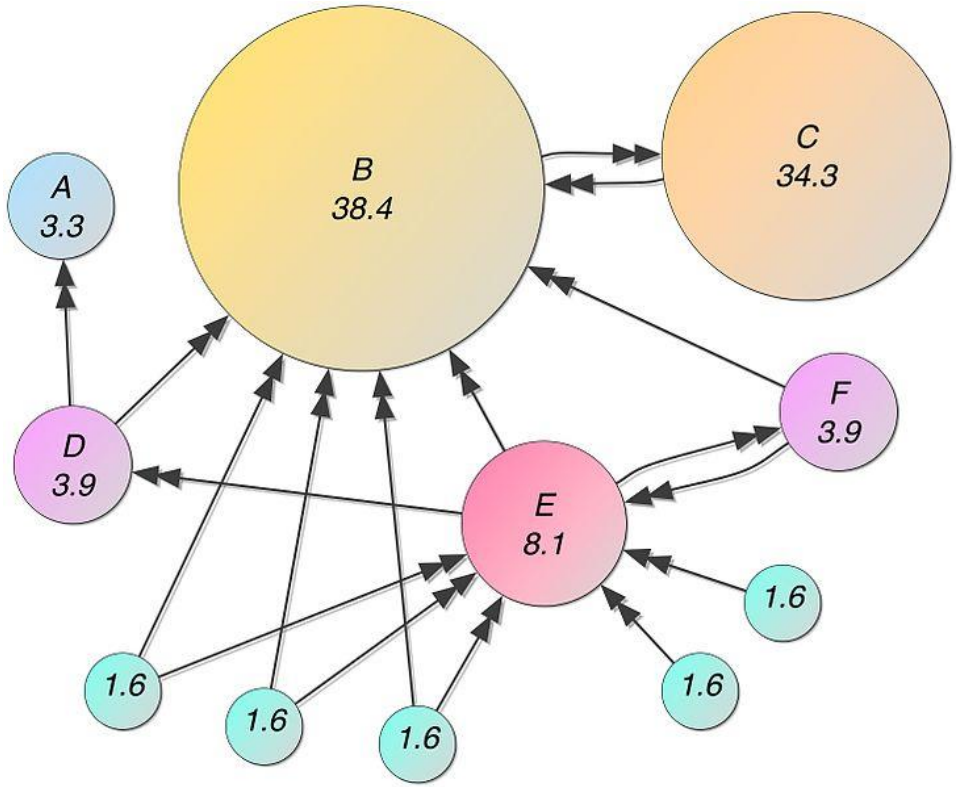


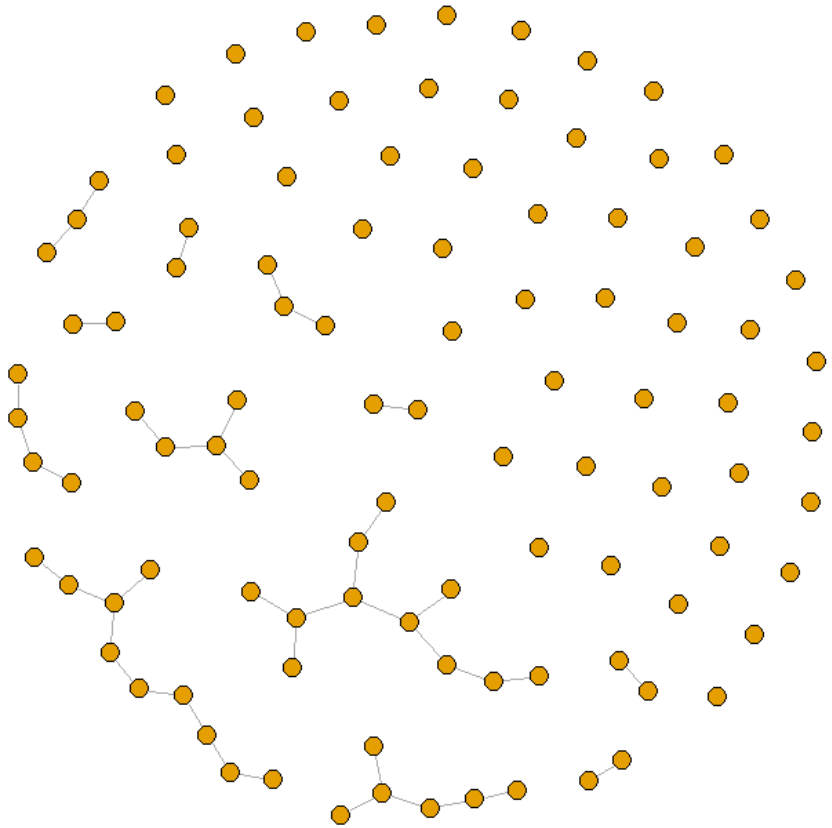
Figure 3: Simplified PageRank Calculation



- R: igraph

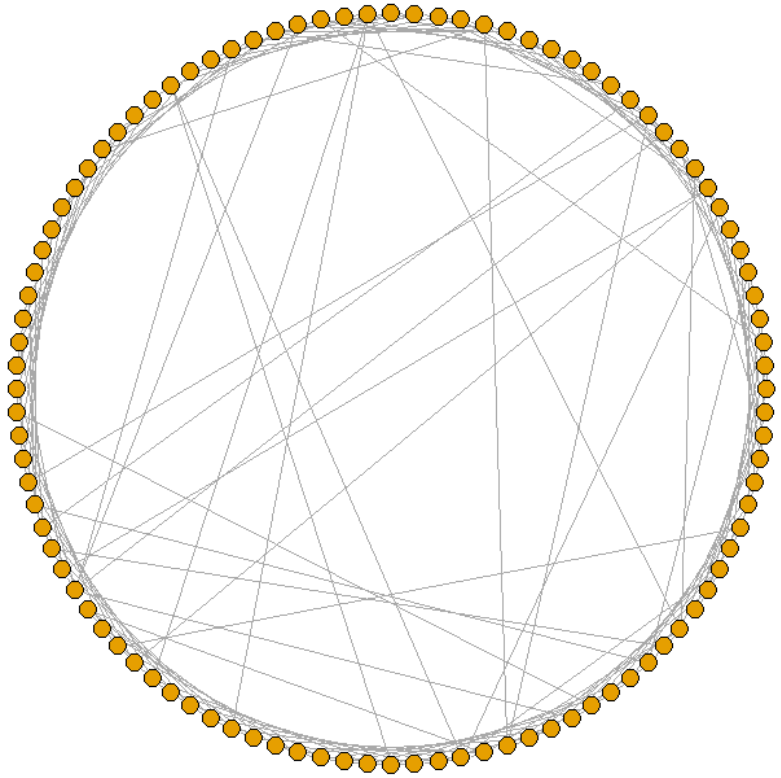
```
#install.packages("igraph")
library(igraph)

# Erdos-Renyi: Random Graph Model
er <- sample_gnm(n=100, m=40)
plot(er, vertex.size=5, vertex.label=NA)
```



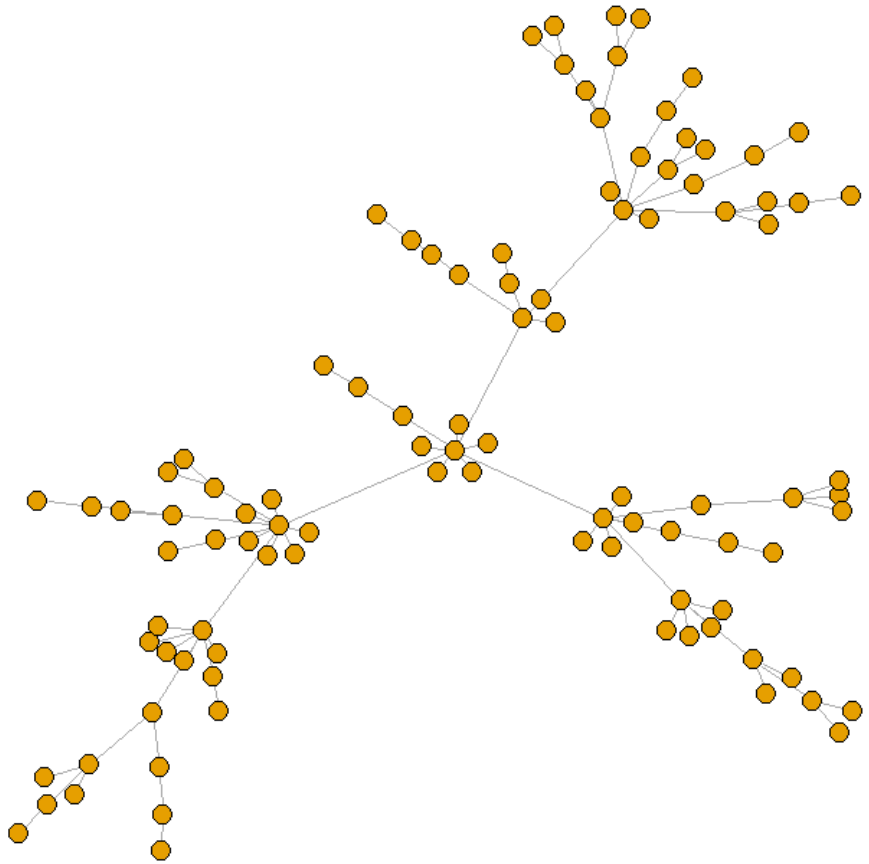
08. 네트워크 분석

```
# Watts-Strogatz: Small-World Model
sw <- sample_smallworld(dim=2, size=10, nei=1, p=0.1)
plot(sw, vertex.size=5, vertex.label=NA, layout=layout_in_circle)
```



08. 네트워크 분석

```
# Barabasi-Albert: Scale-Free Network Model (Preferential Attachment)
ba <- sample_pa(n=100, power=1, m=1, directed=F)
plot(ba, vertex.size=5, vertex.label=NA)
```



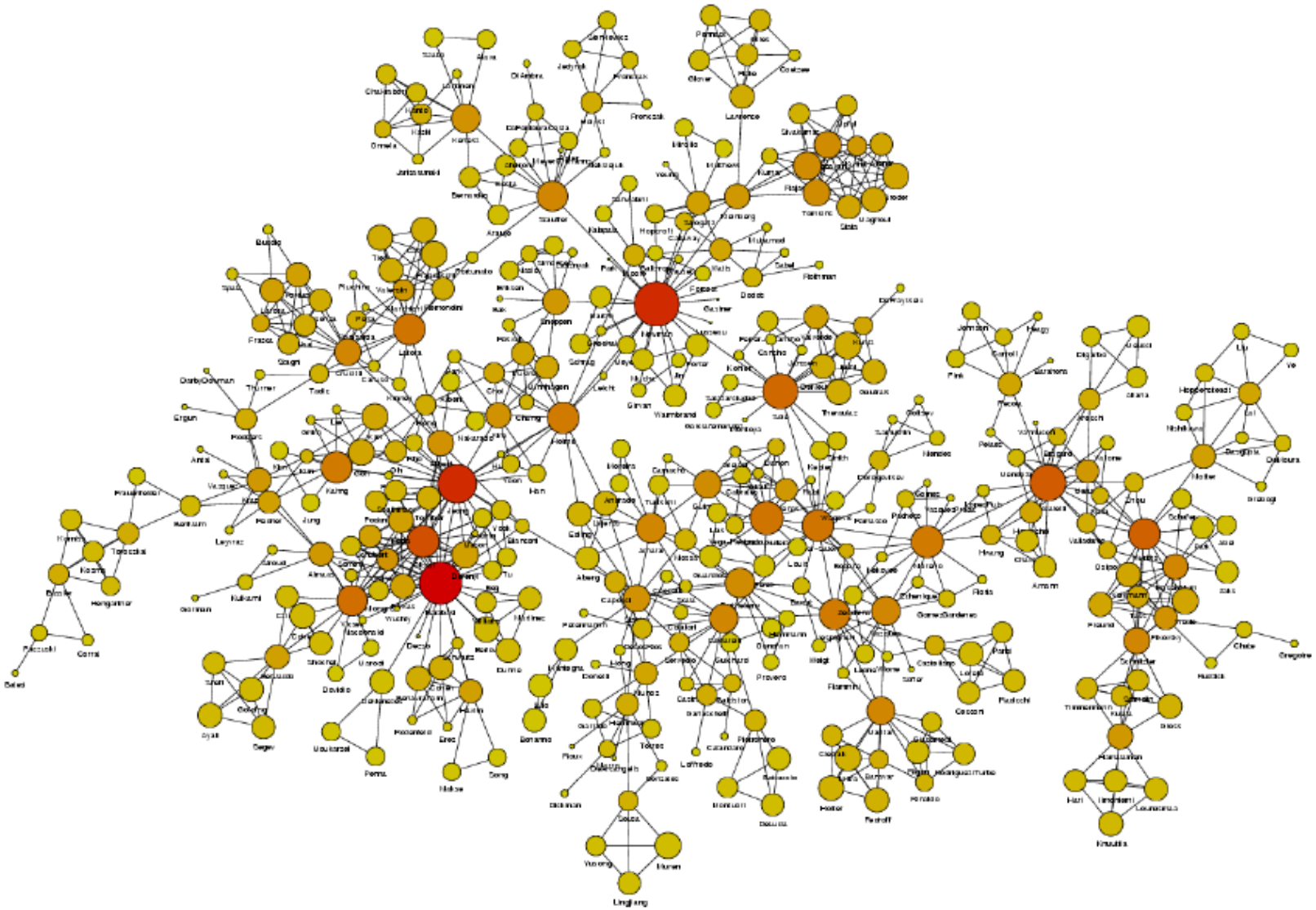
08. 네트워크 분석

- Dataset: NetScience Collaboration Network
  - Coauthorship: **네트워크 과학 분야 논문 공저자들의 네트워크**
    - **저자를 정점으로 두고, 공저자일 경우에 연결을 생성**
    - Pajek **포맷**: netscience.net

```
*vertices 1589
1 "Kuperman, M"
2 "Acebron, J"
3 "Bonilla, L"
4 "Perezvicente, C"
5 "Ritort, F"
6 "Spigler, R"
……
```

```
*edges
1 1589 2.5
3 2 0.25
4 2 0.25
4 3 0.25
5 2 0.25
5 3 0.25
……
```

08. 네트워크 분석



08. 네트워크 분석

```r
# Newman's NetScience Collaboration Network
netsci <- read.graph("netscience.net", "pajek")
netsci

# Degree Centrality
Cd <- sort(igraph::degree(netsci), decreasing = T)
head(Cd, 10)
```

```
> head(Cd, 10)
  Barabasi, A      Jeong, H     Newman, M     Oltvai, Z       Young, M
           34            27            27            21            20
      Uetz, P     Cagney, G  Mansfield, T       Alon, U Boccaletti, S
           20            20            20            19            19
```

08. 네트워크 분석

```
# Closeness Centrality
Cc <- sort(igraph::closeness(netsci), decreasing = T)
head(Cc, 10)
```

```
> head(Cc, 10)
        Holme, P        Newman, M        Edling, C       Liljeros, F        Derenyi, I
5.198596e-07      5.198539e-07      5.198483e-07      5.198483e-07      5.198483e-07
        Jeong, H       Stanley, H          Yoon, C           Han, S Pastorsatorras, R
5.198463e-07      5.198443e-07      5.198432e-07      5.198432e-07      5.198410e-07
```

08. 네트워크 분석

```r
# Betweenness Centrality
Cb <- sort(igraph::betweenness(netsci), decreasing = T)
head(Cb, 10)
```

```
> head(Cb, 10)
        Holme, P         Jeong, H       Newman, M        Boguna, M        Moreno, Y
        24773.92         24507.99        23669.24         22928.72         19900.72
Pastorsatorras, R   Boccaletti, S       Arenas, A      Stanley, H         Sole, R
        17299.21         16986.33        16588.32         16375.72         14124.71
```

08. 네트워크 분석

```r
# PageRank Centrality
PR <- page_rank(netsci)$vector
head(PR, 10)
```

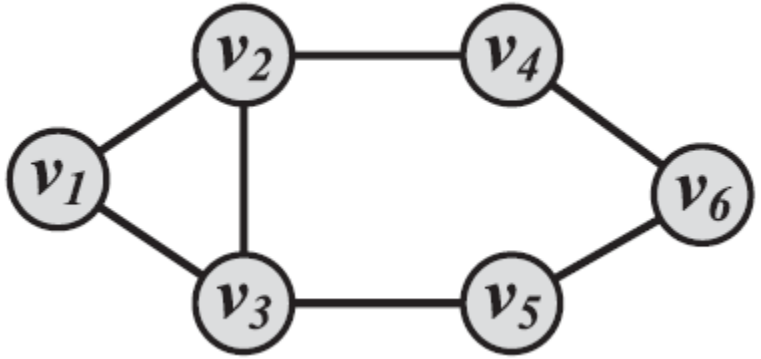
```
> head(PR, 10)
    Kuperman, M      Acebron, J      Bonilla, L Perezvicente, C       Ritort, F
   0.0011511973    0.0006755844    0.0006755844    0.0006755844    0.0006755844
     Spigler, R       Adamic, L          Adar, E     Huberman, B       Lukose, R
   0.0006755844    0.0012519094    0.0002786913    0.0016896052    0.0006485216
```

08. 네트워크 분석

- **네트워크에서의 유사도 척도: *Similarity* Measures**
  - **그래프 내부의 인접한 두 정점 간의 유사도를 측정하려면?**
    - **두 노드의 이웃이 얼마나 겹치는 지를 평가**
    - $\sigma(v_i, v_j) = |N(v_i) \cap N(v_j)|$,
    - $N(v_i)$: **노드** v_i**의 이웃 노드들의 집합**

08. 네트워크 분석

- **유사도 척도의 정규화:** *Normalization*
  - 유사도 척도를 $[0, 1]$ 구간의 값으로 정규화 하는 방법
  - 자카드 유사도: *Jaccard* Similarity
    - $\sigma_{\text{Jaccard}}(v_i, v_j) = \frac{|N(v_i) \cap N(v_j)|}{|N(v_i) \cup N(v_j)|}$
  - 코사인 유사도: *Cosine* Similarity
    - $\sigma_{\text{Cosine}}(v_i, v_j) = \frac{|N(v_i) \cap N(v_j)|}{\sqrt{|N(v_i)| \times |N(v_j)|}}$

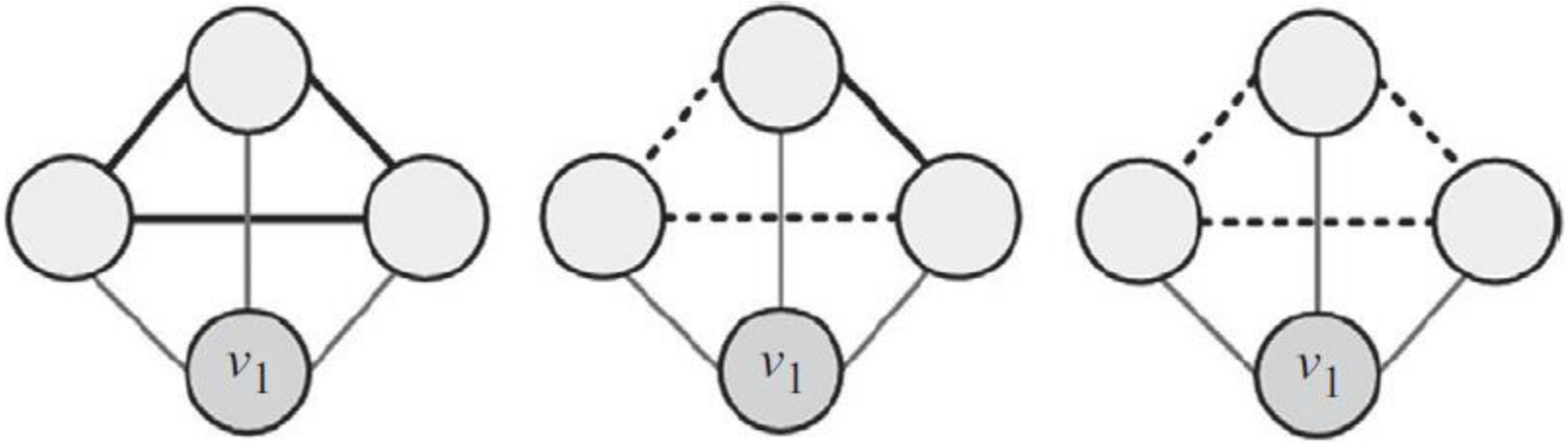


$$\sigma_{\text{Jaccard}}(v_2, v_5) = \frac{|\{v_1, v_3, v_4\} \cap \{v_3, v_6\}|}{|\{v_1, v_3, v_4, v_6\}|} = 0.25,$$

$$\sigma_{\text{Cosine}}(v_2, v_5) = \frac{|\{v_1, v_3, v_4\} \cap \{v_3, v_6\}|}{\sqrt{|\{v_1, v_3, v_4\}||\{v_3, v_6\}|}} = 0.40.$$

08. 네트워크 분석

- **네트워크의 군집도:**
  - **전이적 연결:** *Transitive Linking*
    - 유유상종: 내 친구의 친구는 내 친구일 가능성이 높다.
    - $(v_i, v_j) \wedge (v_j, v_k) \rightarrow (v_i, v_k)$
  - **군집 계수:** *Clustering Coefficient*
    - 노드들이 얼마나 서로 똘똘 뭉쳐 있는가를 평가
    - $C(v_i) = \frac{Number\ of\ Pairs\ of\ Neighbors\ of v_i\ that\ are\ Connected}{Number\ of\ Pairs\ of\ Neighbors\ of\ v_i}$



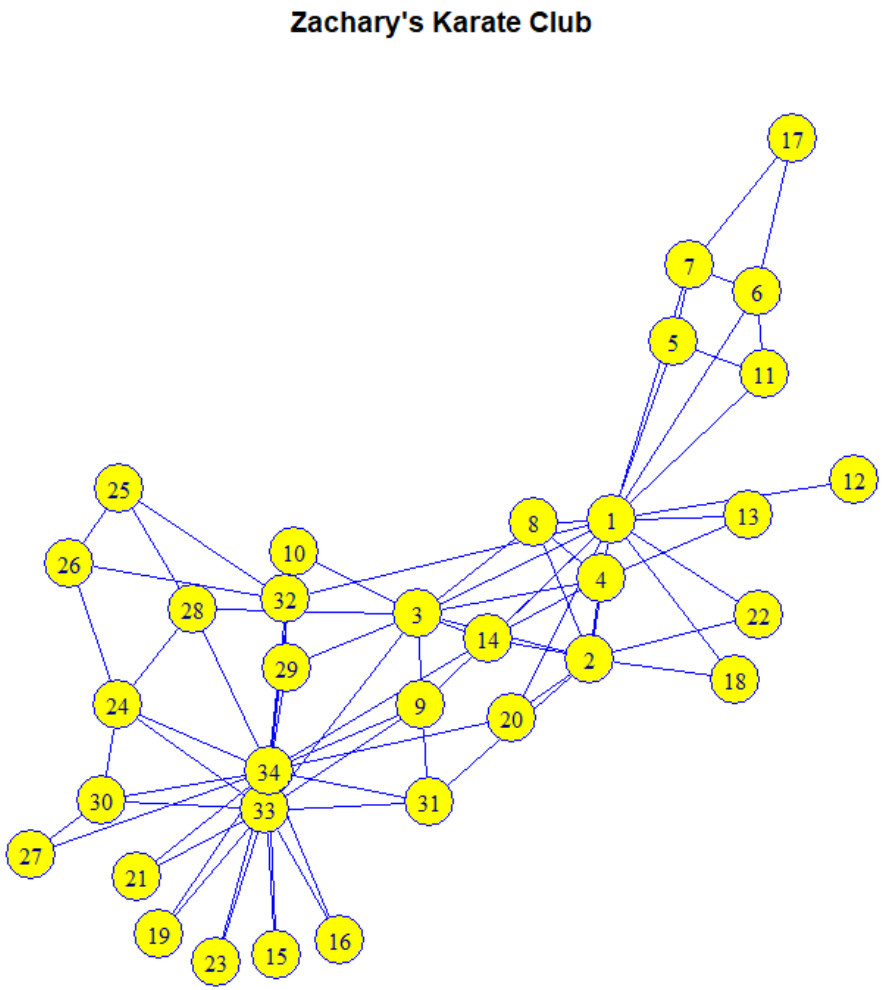
$C(v_1) = 1$ $C(v_1) = 1/3$ $C(v_1) = 0$

08. 네트워크 분석

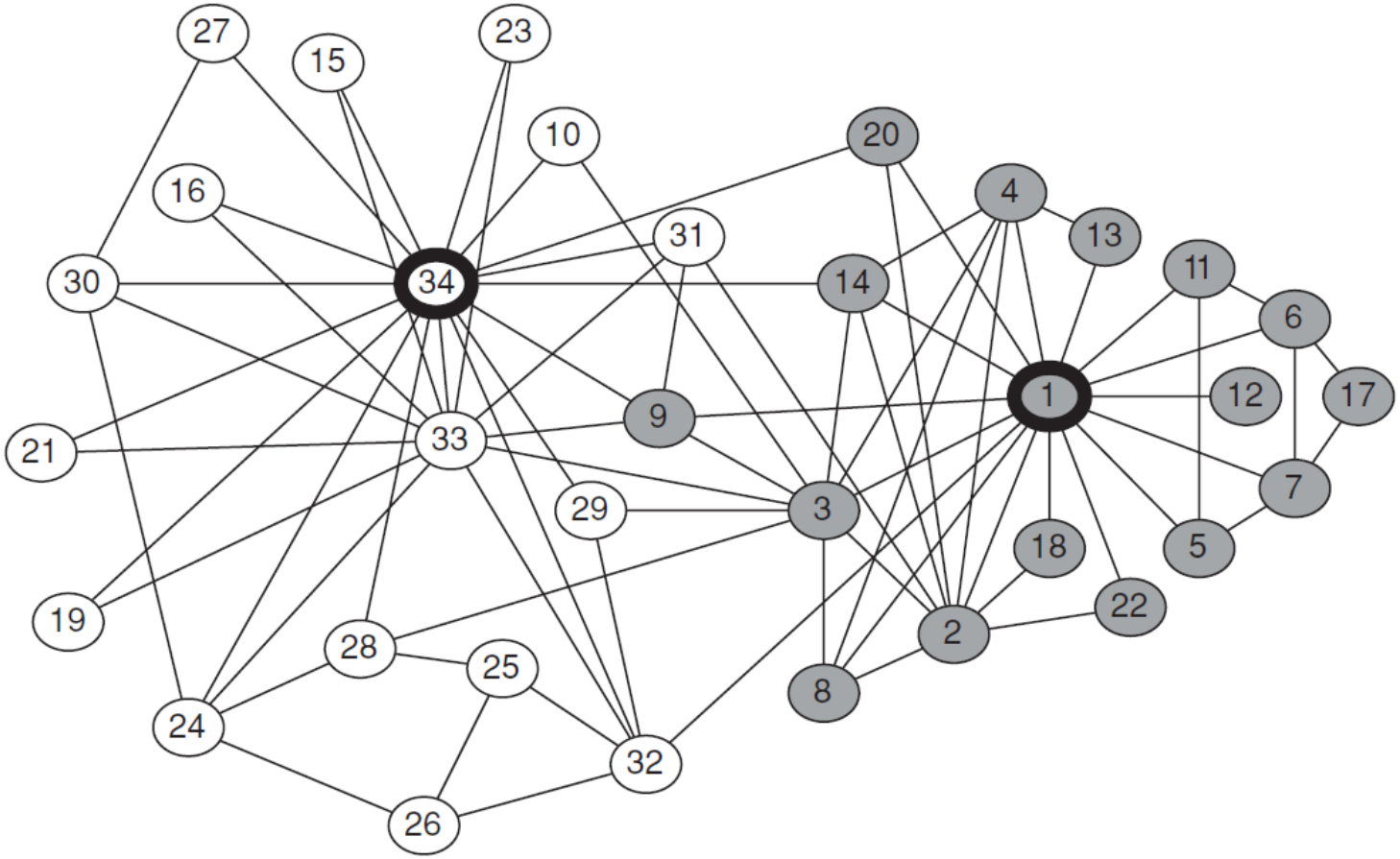
- **커뮤니티 분석: *Community Analysis***
  - **군집 .vs. 커뮤니티: *Cluster* .vs. *Community***
    - 본질적으로는 같은 의미이나, 목적과 수단이 서로 다른 배경에서 출발
    - 군집: 특성에 의해 나누어진 군집을 발견하는 것이 목적 (IRIS)
    - 커뮤니티: 나누어진 군집을 발견해서 특성을 이해하는 것이 목적 (SNS)
  - **커뮤니티 발견: *Community Detection***
    - 네트워크의 연결 구조를 이용한 클러스터링 알고리즘
    - 때로는 고정된 커뮤니티가 아닌, 진화하는 커뮤니티 발견이 문제일 수 있음

자카리의 가라테 클럽: Zachary's Karate Club Network

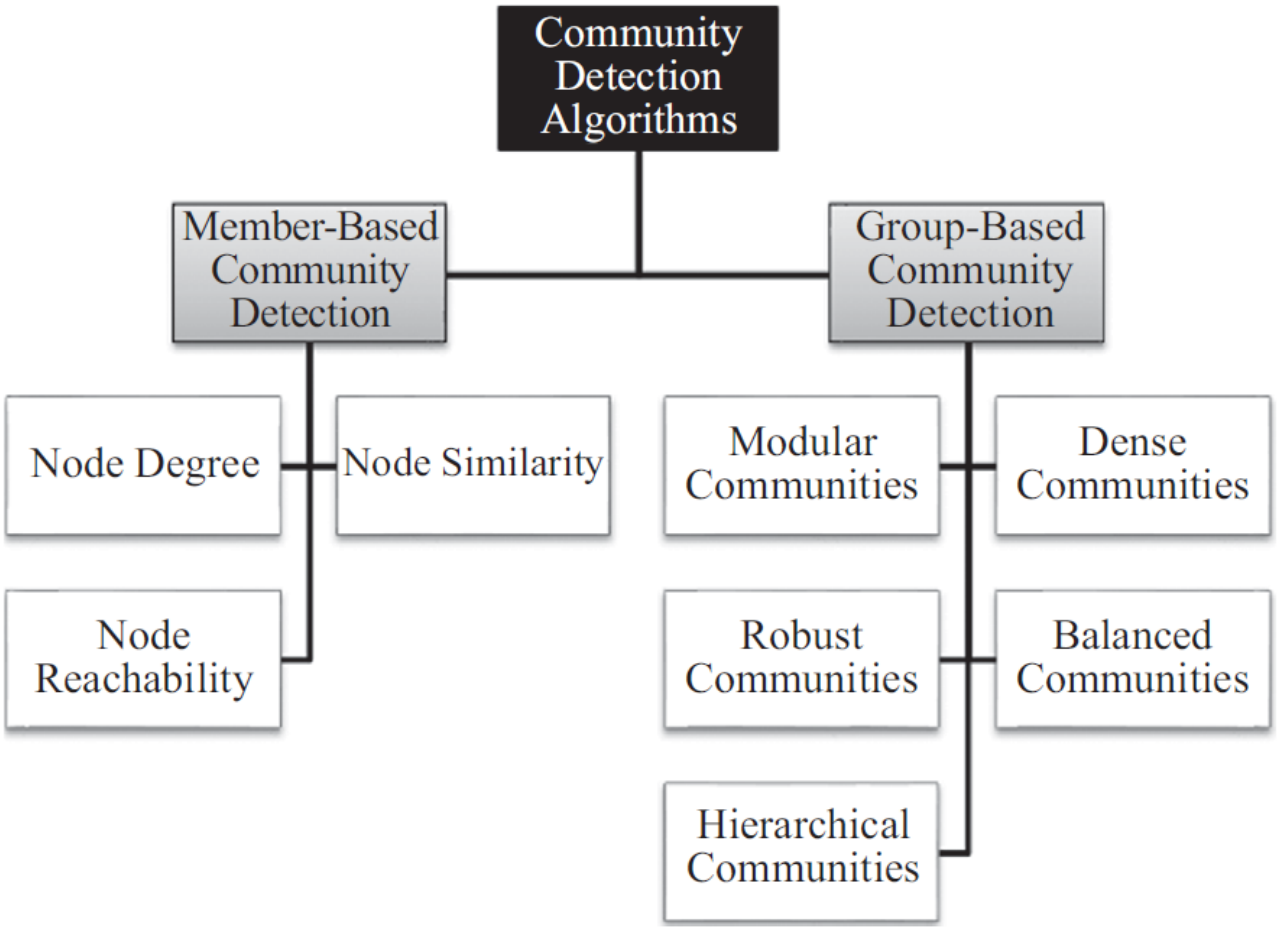
```
# Zachary's Karate Club
karate <- graph("Zachary")
plot(karate,
     main = "Zachary's Karate Club",
     vertex.size = 12,
     vertex.color = "yellow",
     vertex.frame.color = "blue",
     edge.color= "blue")
```



08. 네트워크 분석


27
15
23
10
20
16
4
31
13
30
34
14
11
6
1
12
17
9
21
33
7
29
3
18
5
19
22
2
28
25
8
24
32
26

- Community Detection Algorithms:



08. 네트워크 분석

- **커뮤니티 구조의 평가:**
  - 지도 학습:
    - 실제로 해당 노드가 어떤 커뮤니티에 속하는 지를 아는 경우
    - 혼동 행렬을 통한 평가 지표를 사용할 수 있음
  - 비지도 학습:
    - 해당 노드가 어떤 커뮤니티에 속하는 지를 모르는 경우
    - 커뮤니티의 구조가 네트워크의 구조를 얼마나 잘 반영했는가를 평가

08. 네트워크 분석

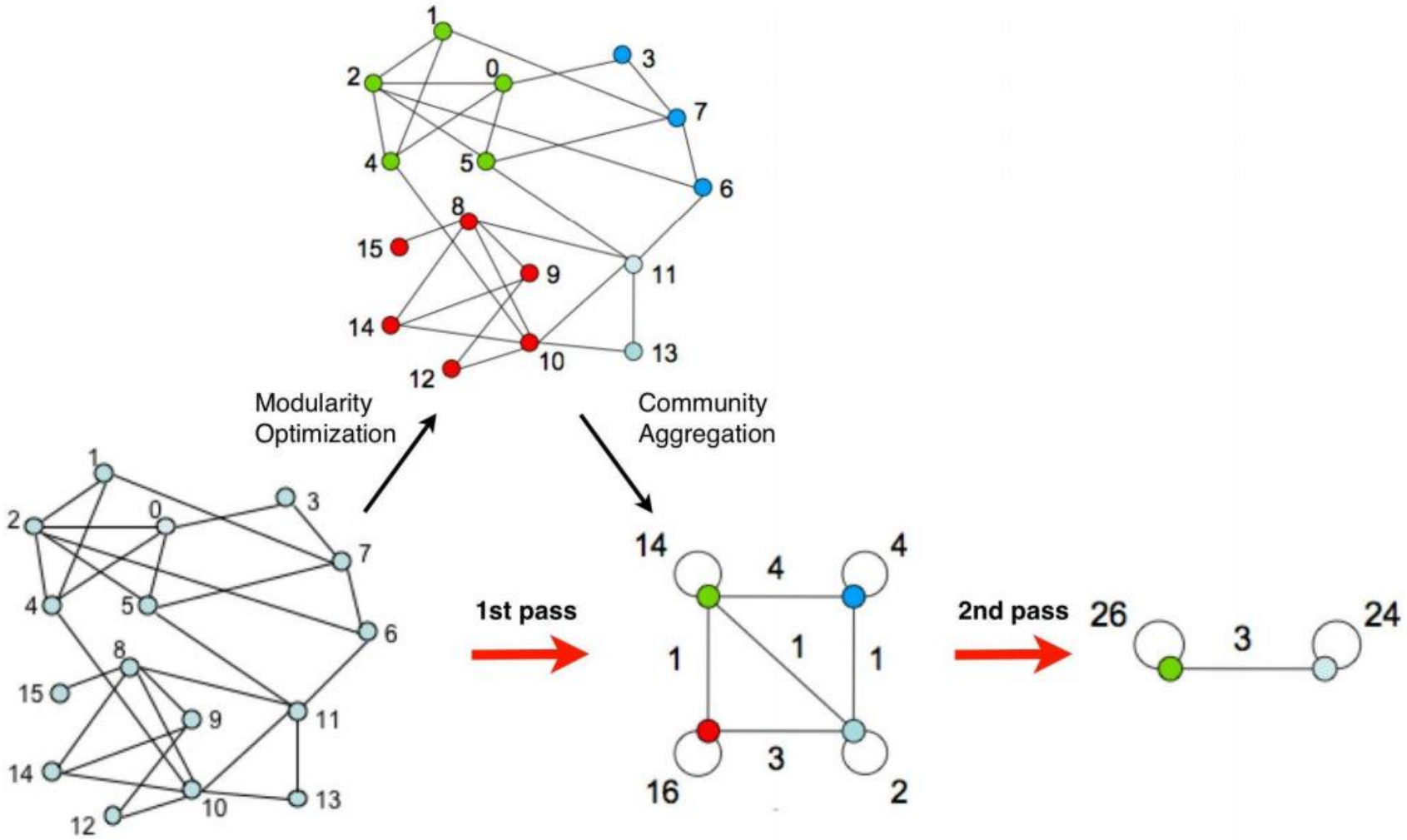
모듈성: *Modularity*

- 기본 아이디어: 커뮤니티 구조는 랜덤 네트워크의 구조와는 달라야 한다.
- 가중치 없는 무방향 그래프 $G = (V, E)$에서,
  - 전체 간선의 수가 $|E| = m$이라면,
  - 차수가 d_i, d_j인 두 노드 v_i, v_j 가 서로 연결되어 있을 확률은 얼마인가?
- 뉴먼 모듈성 지수: Newman's Modularity
  - $Q = \frac{1}{2m}\left(\sum_{x=1}^{k}\sum_{v_i,v_j \in P_x} A_{ij} - \frac{d_i d_j}{2m}\right)$
  - k는 파티션(커뮤니티)의 개수, P_x는 x번째 파티션, A_{ij}는 인접행렬의 원소

08. 네트워크 분석

- **루베인 알고리즘: *Louvain* Method**
  - 뉴먼의 모듈성 지수를 최적화하는 알고리즘
  - 1단계: 한 노드를 현재 커뮤니티에서 빼내어 인접한 커뮤니티에 재배치
    - 모듈성 지수가 증가하면 해당 커뮤니티에 배치, 아니면 원래 커뮤니티에 둠
  - 2단계: 1단계에서 생성된 커뮤니티를 하나의 노드로 하는 그래프를 새로 만듦
    - 커뮤니티의 내부 링크는 self-loop의 가중치로,
    - 커뮤니티 간 연결은 해당 커뮤니티 간의 연결 가중치의 합으로 결정
  - 위와 같은 pass(1단계+2단계)를 모듈성 지수 증가가 없을 때까지 반복

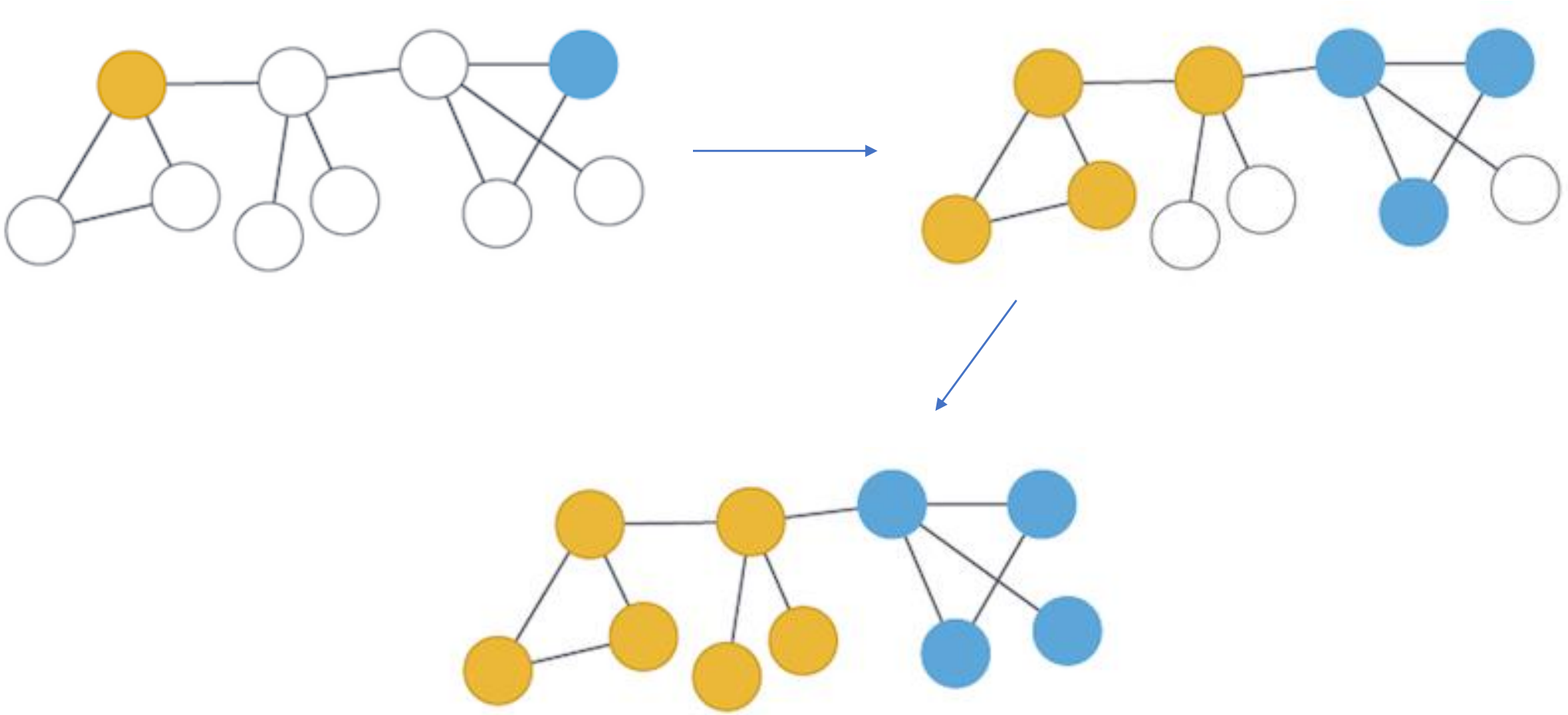
08. 네트워크 분석



Source: Blondel, Vincent D., et al. "Fast unfolding of communities in large networks." *Journal of statistical mechanics: theory and experiment* 2008.10 (2008): P10008.

- **라벨 전파 알고리즘**: LPA, *Label Propagation* Algorithm
  - 가정: 내가 속한 커뮤니티는 내 이웃들이 많이 속해 있는 커뮤니티와 같다.
  - 각자가 가진 라벨을 네트워크에 전파하여 마지막에 남은 라벨로 커뮤니티 결정.
  - LABEL-PROPAGATION-ALGORITHM:
    - 모든 노드가 각자의 고유한 라벨을 소유함
    - 임의의 순서로 각 노드는 이웃 노드들이 가장 많이 가진 라벨로 업데이트함
    - 모든 노드의 라벨이 이웃 노드들 중 가장 많은 라벨로 구성되면 종료함
    - 이웃 노드의 라벨의 개수가 동일하면 임의로 하나 선정함

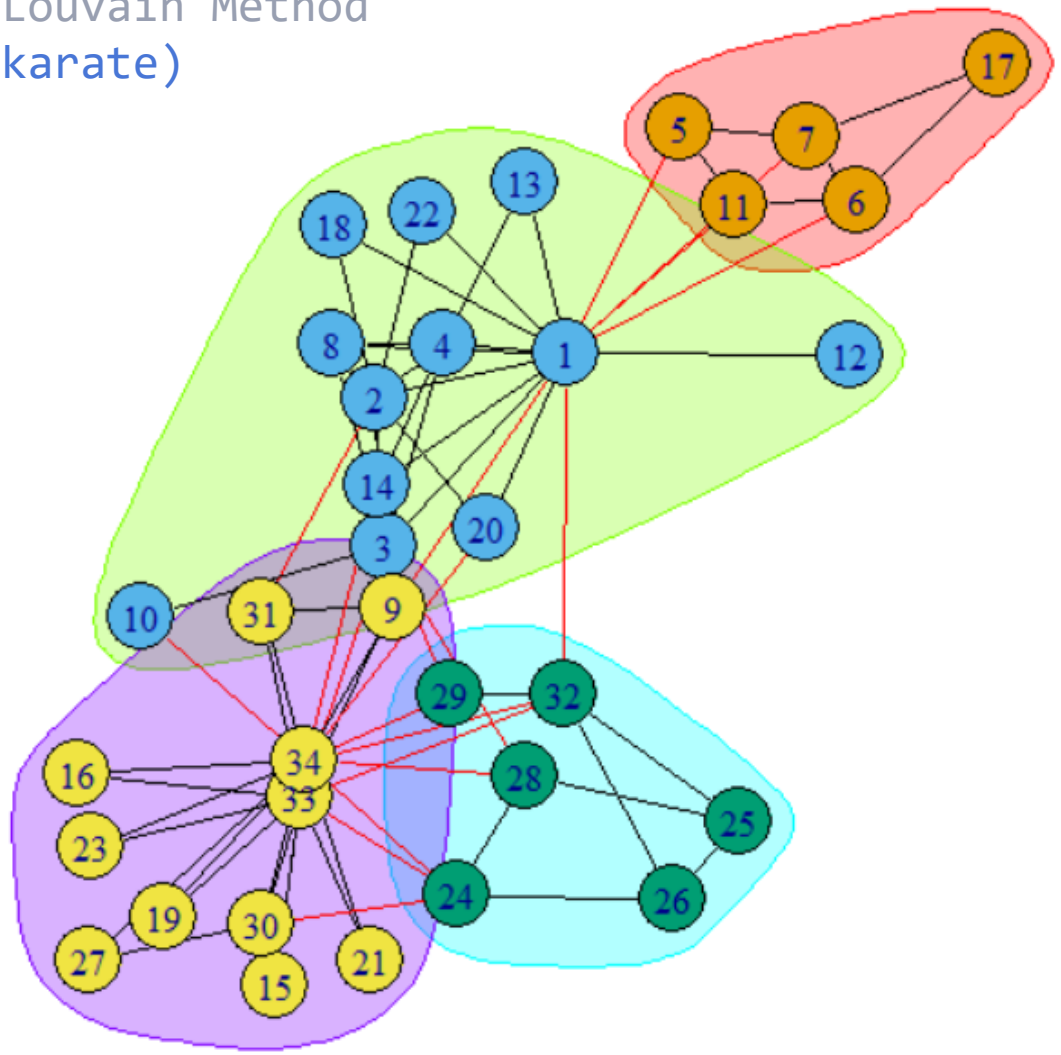
08. 네트워크 분석



Source: Raghavan, Usha Nandini, Réka Albert, and Soundar Kumara. "Near linear time algorithm to detect community structures in large-scale networks." Physical review E 76.3 (2007): 036106.

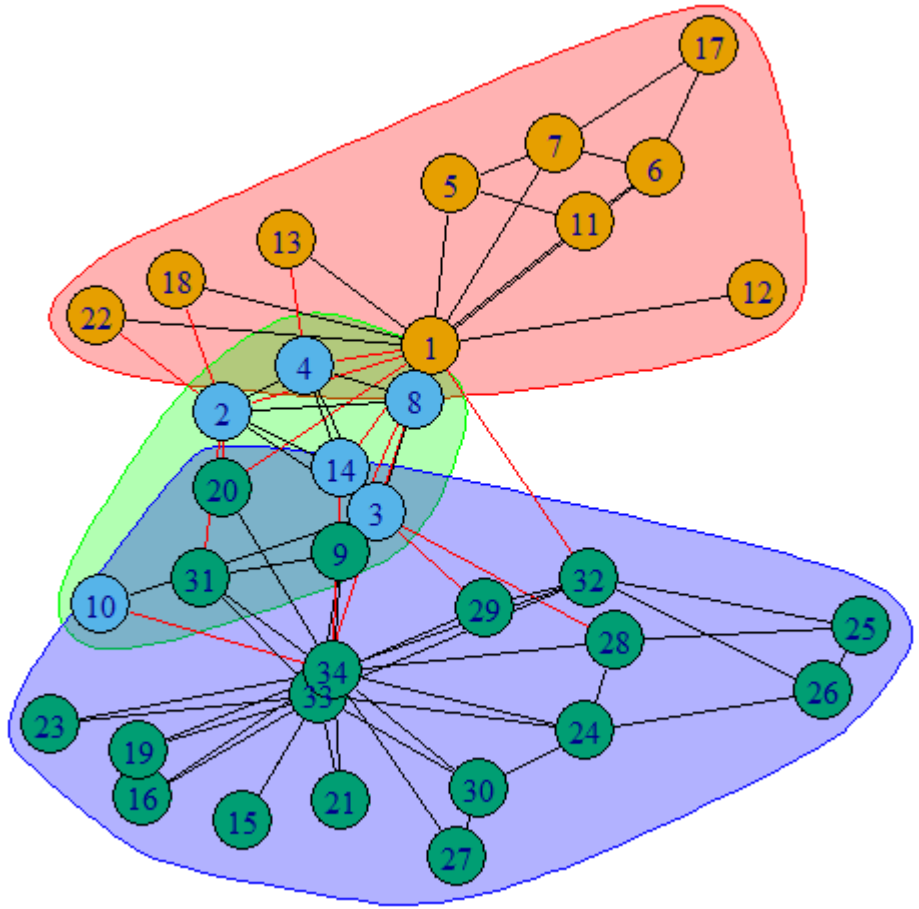
08. 네트워크 분석

```
# Community Detection with the Louvain Method
communities <- cluster_louvain(karate)
modularity(communities)
plot(communities, karate)
```



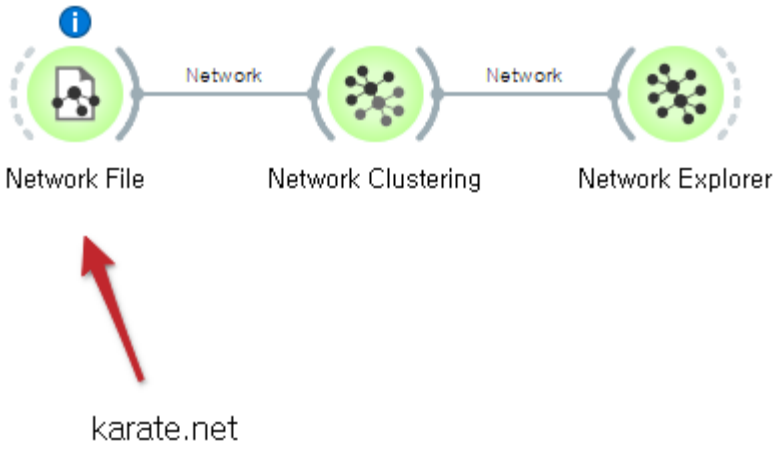
08. 네트워크 분석

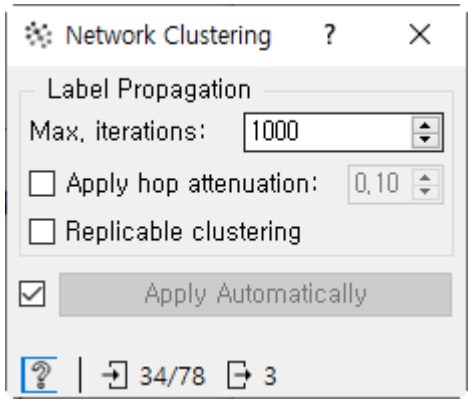
```
# Community Detection with the Label Propagation Algorithm
communities <- cluster_label_prop(karate)
modularity(communities)
plot(communities, karate)
```



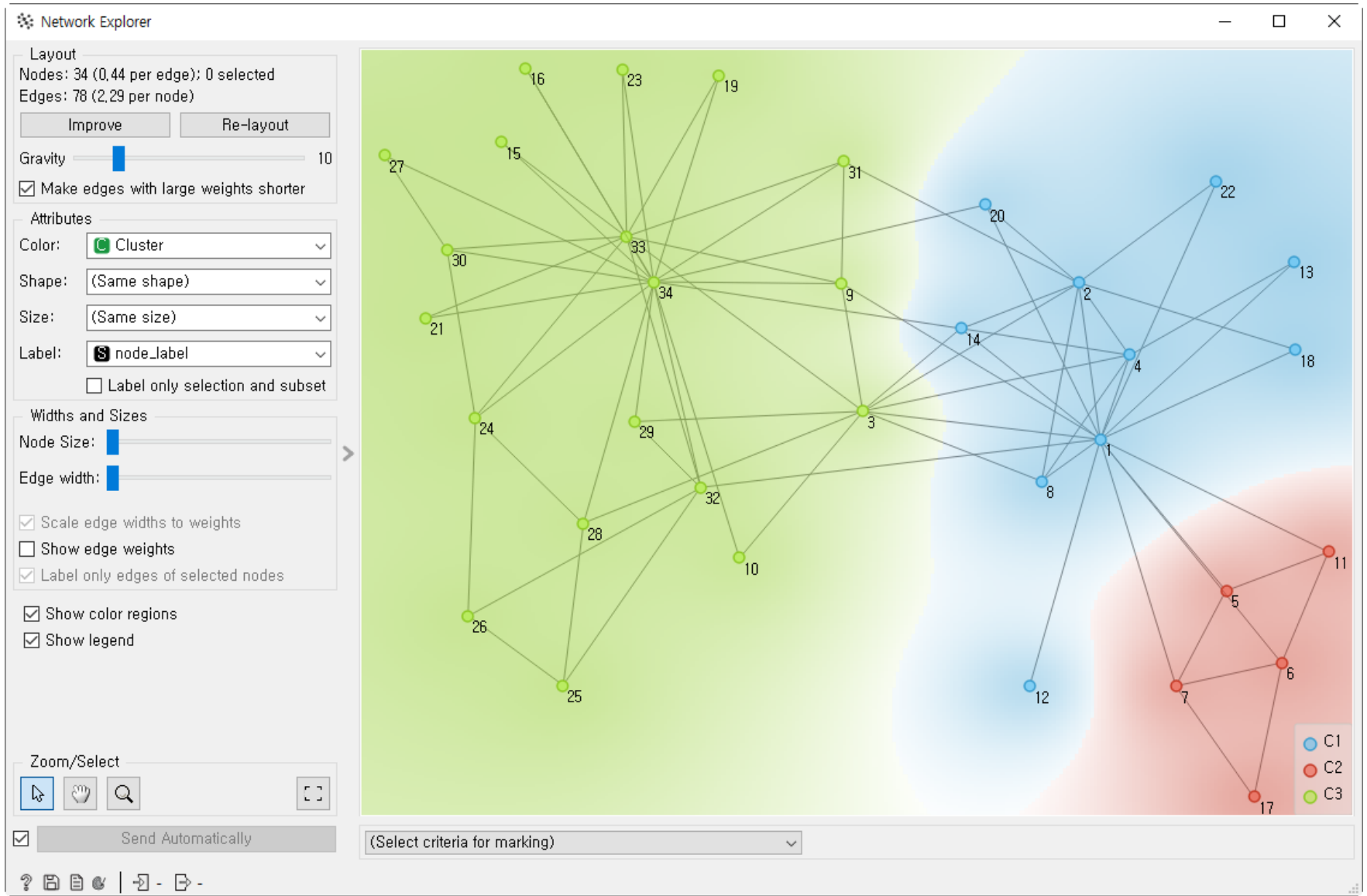
08. 네트워크 분석

- Orange: Network Clustering

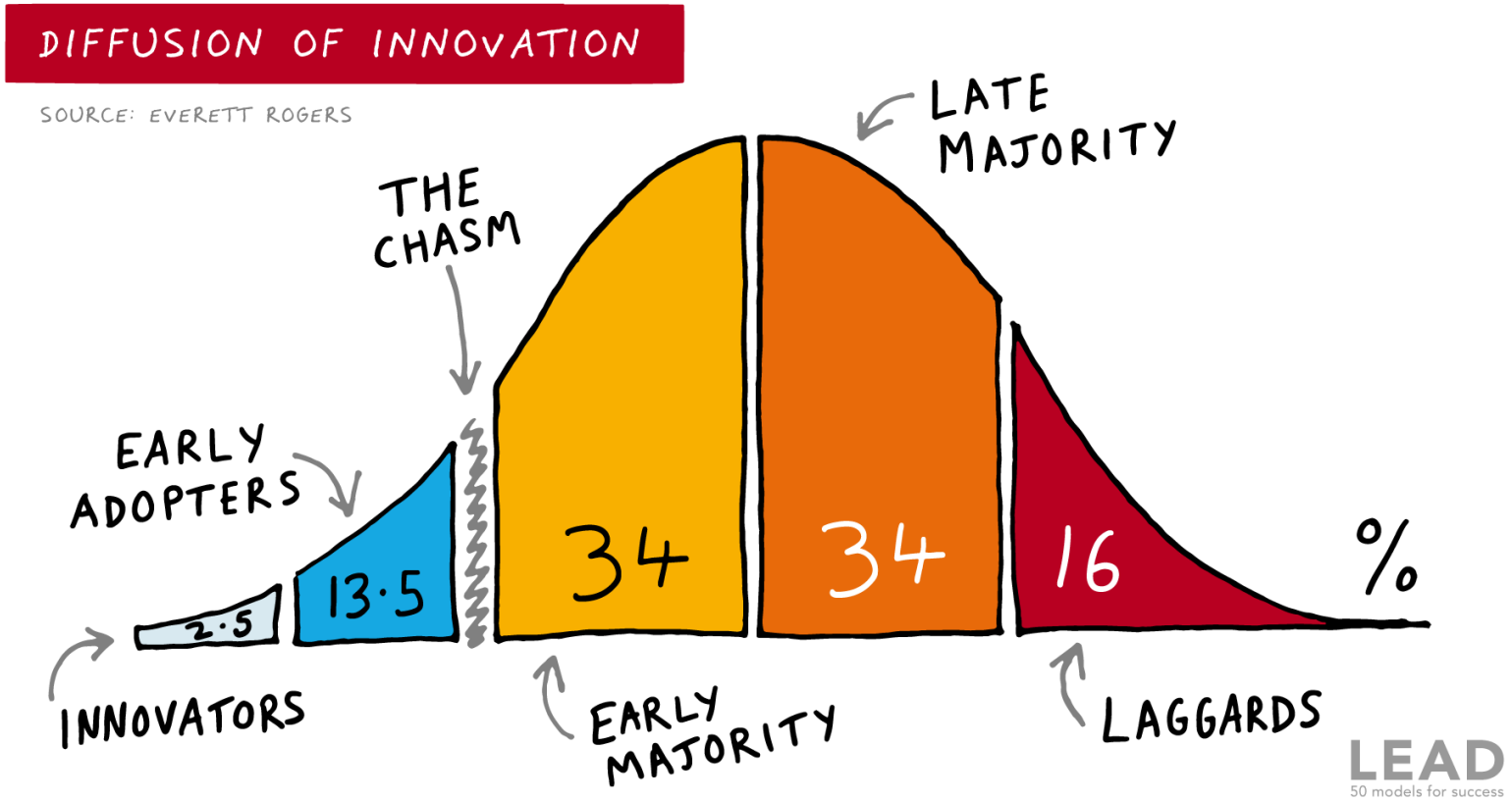




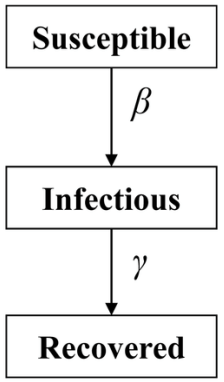
08. 네트워크 분석


Network Explorer
Layout
Nodes: 34 (0,44 per edge); 0 selected
Edges: 78 (2,29 per node)
Improve
Re-layout
Gravity
10
Make edges with large weights shorter
Attributes
Color:
Cluster
Shape:
(Same shape)
Size:
(Same size)
Label:
node_label
Label only selection and subset
Widths and Sizes
Node Size:
Edge width:
Scale edge widths to weights
Show edge weights
Label only edges of selected nodes
Show color regions
Show legend
Zoom/Select
Send Automatically
(Select criteria for marking)
C1
C2
C3

- **개혁의 확산: *Diffusion of Innovation***
  - 에버렛 로저스: 사람들이 신기술을 받아들이는 데에는 일정한 패턴이 있다.



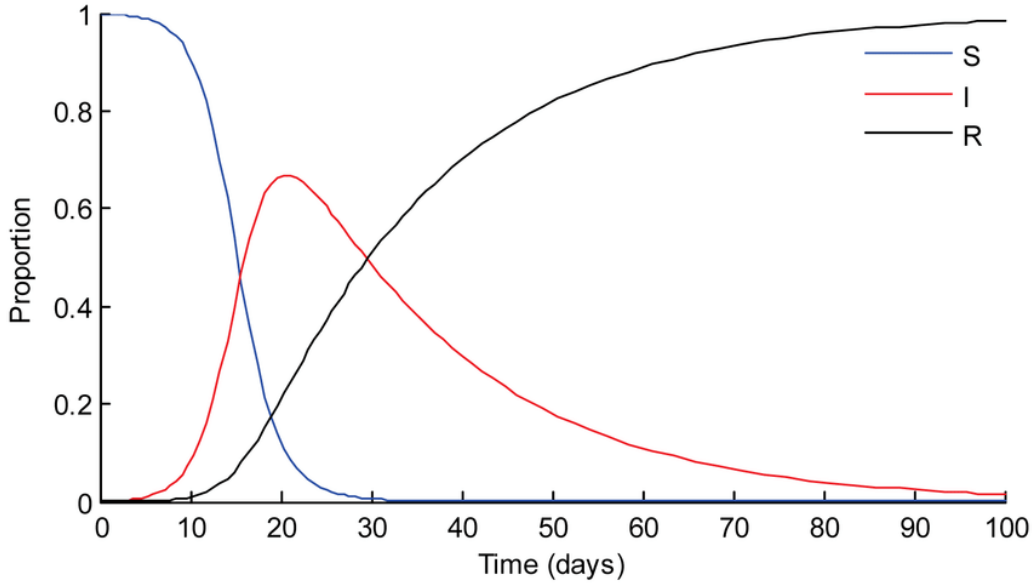
- **질병의 확산 모델**: The Spread of Disease, *Epidemics*
  - SIR 모델: 전염병의 확산을 분석하는 가장 기본적인 모델
    - Susceptible: 감염될 수 있는 개체
    - Infected: 감염된 개체
    - Recovered: 면역된 개체



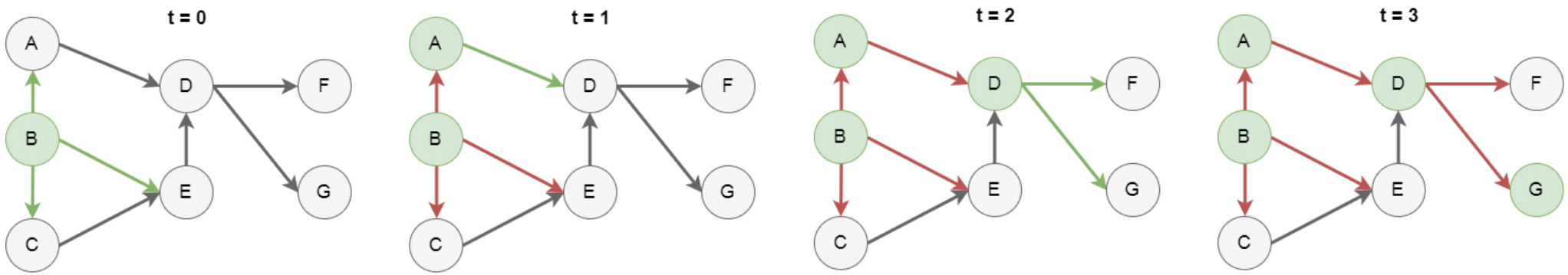
$$\frac{dS}{dt} = -\beta SI$$

$$\frac{dI}{dt} = \beta SI - \gamma I$$

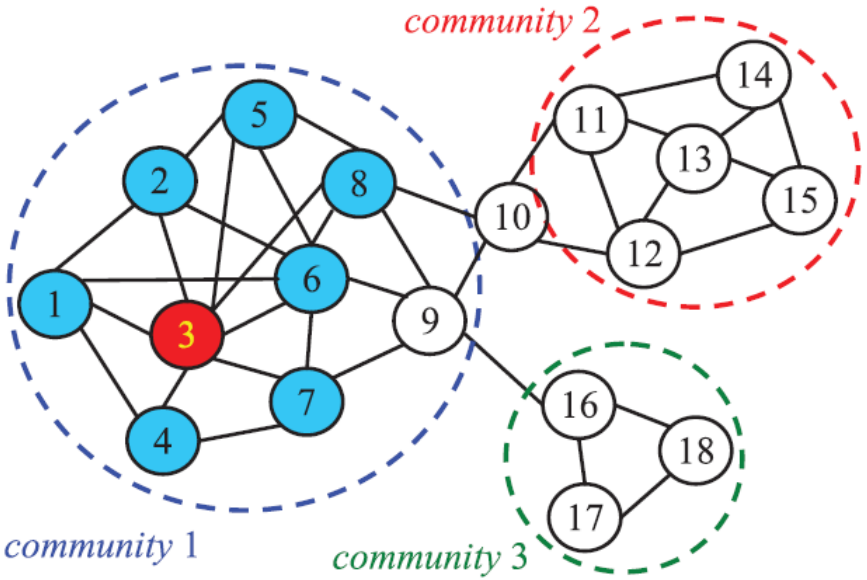
$$\frac{dR}{dt} = \gamma I$$

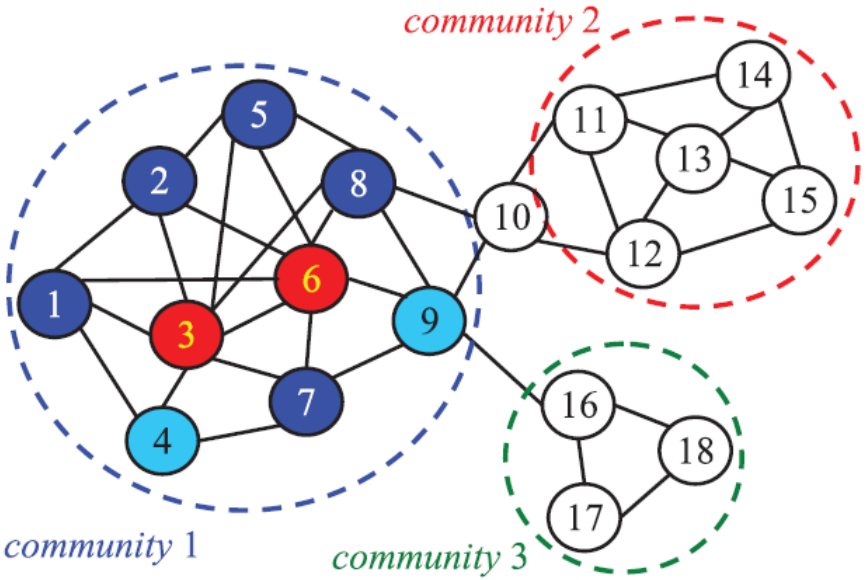


- **정보의 캐스케이드: *Information Cascade***
  - 101마리 원숭이 현상: 다수의 사람들이 동시에 같은 결정을 내리게 되는 현상
  - 소문의 확산, 가짜 뉴스의 확산 등을 모델링할 때
  - ICM: *Independent Cascade* Model
    - 정보의 확산을 독립된 개체의 확률적 결정으로 판단



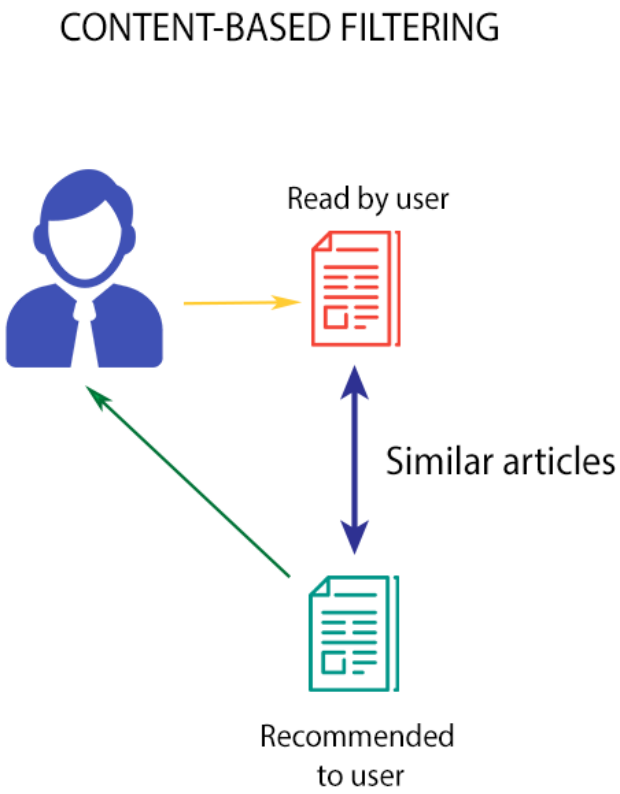
- **영향력 극대화 문제: *Influence Maximization* Problem**
  - 소셜 네트워크에서 영향력을 극대화할 수 있는 사람들의 집합을 찾는 문제
  - k가 주어졌을 때, 가장 확산 범위가 높은 k개의 노드 선택하기

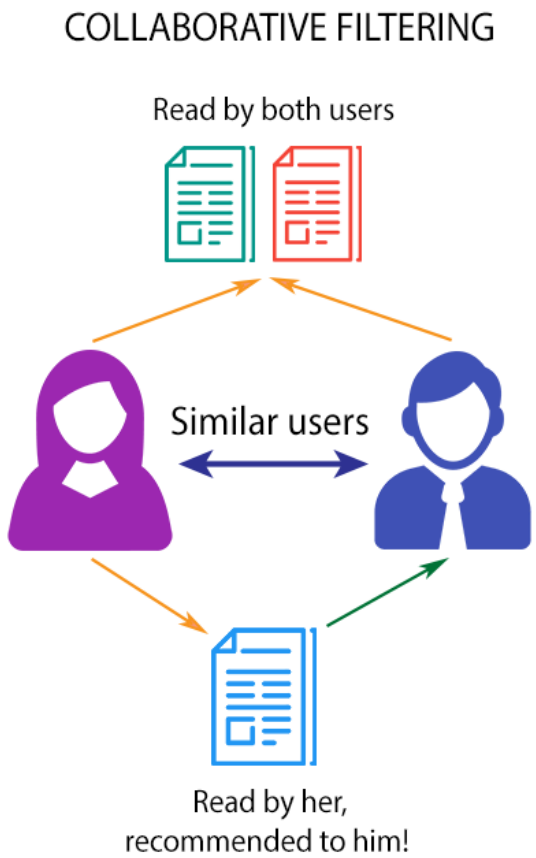




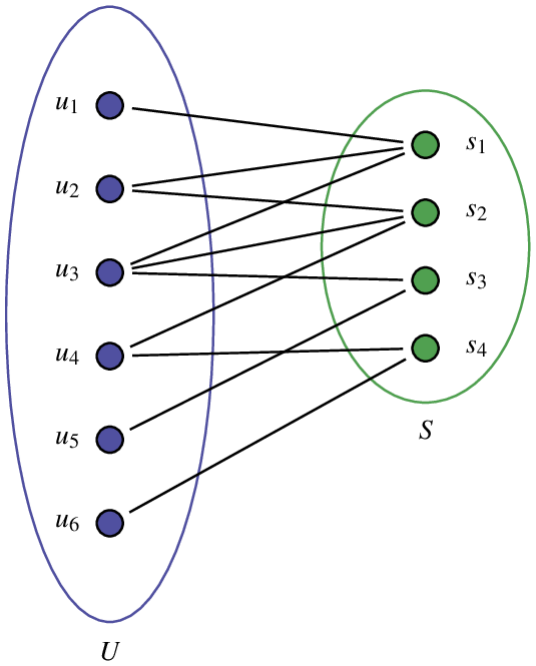
08. 네트워크 분석

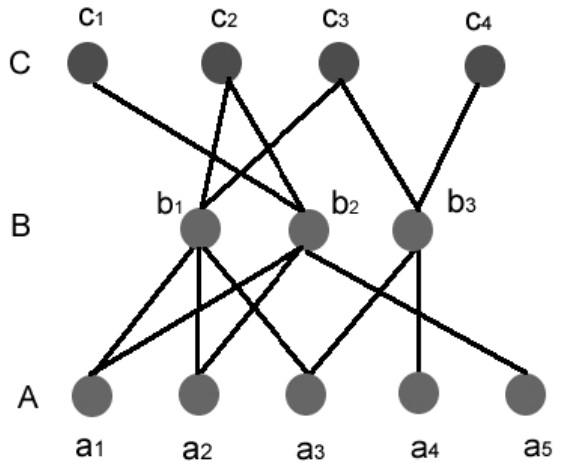
- **추천 알고리즘**: *Recommendation* Algorithms
  - **사용자가 선호할 아이템을 예측하여 해당 항목을 추천하는 알고리즘**
  - **콘텐츠 기반 추천**: Content-based Recommendations
    - **아이템이 사용자의 성향에 잘 맞는가를 판단함 (사용자-아이템 유사도)**
  - **협업 필터링**: Collaborative Filtering
    - **유사한 사용자들이 선호하는 아이템을 추천 (사용자-아이템 행렬)**


CONTENT-BASED FILTERING
Read by user
Similar articles
Recommended
to user


COLLABORATIVE FILTERING
Read by both users
Similar users
Read by her,
recommended to him!

- **추천 시스템을 위한 네트워크 모델:**
  - **이분할 그래프:** *Bipartite* Network
  - **삼분할 그래프:** *Tripartite* Network





Any Questions?