

<목차>

- 1    그래프 설명
  - 1.1    Node
  - 1.2    Edge
  - 1.3    Node feature
  - 1.4    Edge feature
  - 1.5    한계점 및 해결방안
- 2    네트워크 분석  
      (property)
  - 2.1    Avg degree
  - 2.2    Degree distribution
  - 2.3    Avg clustering coefficient
  - 2.4    Connected component
  - 2.5    Degree centrality
  - 2.6    Closeness centrality
  - 2.7    Betweenness centrality
  - 2.8    Eigenvector centrality
- 3    그래프 시각화

## 1 그래프 설명

Suggest\_20.csv 파일을 이용하여 입법발의안 발의자 국회의원 간 무방향 가중치 그래프 (undirected weighted graph)를 만들었다.

### 1.1 Node

그래프의 node는 입법발의안 대표발의자 국회의원('RST\_PROPOSER')과 입법발의안 공동발의자 국회의원('PUBL\_PROPOSER')이다. node의 개수는 318개이다.

<그림 1>

```
NodeView([('박영선', '강훈식', '기동민', '김경수', '김두관', '김병욱',
```

<그림 1>은 파이썬 라이브러리 networkx를 이용하여 만든 그래프의 node의 일부이다.

### 1.2 Edge

그래프의 edge는 입법발의안을 공동으로 발의한 경우이다. 따라서 각 edge는 (대표발의자, 공동발의자) 형태를 띄고 있다. edge의 개수는 11889개이다.

<그림 2>

RST_PROPPUBL_PRCkeyword	COMMITTBILL_index
박영선	강훈식,기동민,김경수,김두관,김병욱,김상희,김성수,김영주,김영준,김종민,김종인,김현권,노웅래,

<그림 3>

```
EdgeView([('박영선', '강훈식'), ('박영선', '기동민'), ('박영선', '김경수'),
```

<그림 2>은 Suggest\_20.csv 파일의 첫 번째 행에 해당한다. 파이썬 라이브러리 networkx를 이용하여 만든 이 행의 edge의 일부는 <그림 3>과 같다.

### 1.3 Node feature

그래프의 node feature는 people\_20.csv 파일에 적혀있는 해당 node의 국회의원 소속 정당('party\_index')이다.

※ 뒤의 그래프 시각화 부분에서 밝히겠지만 people\_20.csv 파일에 2명의 국회의원('전혜숙', '어기구')의 소속 정당이 잘못 적혀있다.

### 1.4 Edge feature

그래프의 edge feature는 suggest\_20.csv 파일에 적혀있는 입법발의안('BILL\_index')과 가중치로 총 2가지를 시도하였고, 결론적으로 그래프에 반영된 edge feature는 가중치이다.

첫 번째 edge feature인 입법발의안 때문에 그래프를 만드는데 많은 고민을 하였다. 동일한 대표발의자와 공동발의자 쌍이 여러 개의 입법발의안을 발의한 경우가 많이 존재하였다. 예를 들어, <그림 3>의 첫 번째 edge인 ('박영선', '강훈식')은 BILL\_index 0번과 60번을 발의하였다. 위와 같은 경우 때문에 edge feature는 리스트가 되는 것이 적합하다는 판단하에 파이썬 라이브러리 networkx를 이용하여 edge feature('BILL\_index')를 아래의 <그림 4>처럼 만들었다.

<그림 4>

```
EdgeDataView([('박영선', '강훈식', {'bill_index': [0, 60]}), ('박영선', '기동민', {'bill_index': [0]}),
```

하지만 이렇게 edge feature를 리스트로 만들면 후에 gephi를 이용하여 그래프를 시각화하는 것이 불가능했다. 그래서 그래프 자체를 mutigraph로 다시 만드는 것이 적합한가를 고민하다 가중치(weight)를 도입하는 대안이 떠올랐다. 동일한 대표발의자와 공동발의자가 발의한 입법발의안의 개수를 가중치로 두어 해당 edge로 연결된 node끼리 입법발의안을 공동으로 발의한 빈도수 정보를 추가해주었다. 파이썬 라이브러리 networkx를 이용하여 가중치를 아래의 <그림 5>처럼 만들었다.

<그림 5>

```
EdgeDataView([('박영선', '강훈식', {'weight': 2}), ('박영선', '기동민', {'weight': 1}),
```

## 1.5 한계점 및 해결방안

- 한계점

People\_20.csv 파일을 이용해서 그래프의 node feature를 추가해서 node를 확인해본 결과 아래의 <그림 6>과 같이 이상한 node 2개가 보였다.

<그림 6>

```
[56] g.nodes.data()
NodeDataView({'박영선': {'party_index': 1}, '강훈식': {'party_index': 1},
( 최경환 3
( 최경환 4
Name: party_index, dtype: int64}, '박선숙': {'party_index': 3}, '윤호중':
( 김성태 2
( 김성태 2
```

원인을 찾던 도중 그래프의 node 개수는 318개인데 people\_20.csv 파일상에 존재하는 국회의원은 320명이였다. 입법 활동을 한번도 안한 국회의원은 없을 것이라는 가정하에 원인을 분석하면 동명이인 2쌍이 존재하는데 이를 suggest\_20.csv 파일에서 구분하고 있지 않아서 생긴 문제였다.

- 해결방안

중복값을 해결해주지 않고 그대로 node feature로 더해줄 경우, 후에 gephi를 이용하여 그래프를 시각화하는 과정에서 node와 node attribute가 맞지 않아 KeyError가 발생한다. 이를 해결하기 위해 차선책으로 파이썬 라이브러리 pandas에 존재하는 drop\_duplicates()를 이용하여 people\_20.csv 파일에 '의원명' column을 기준으로 중복값을 삭제한 후, node attribute를 추가해주었다.

## 2 네트워크 분석

- Property

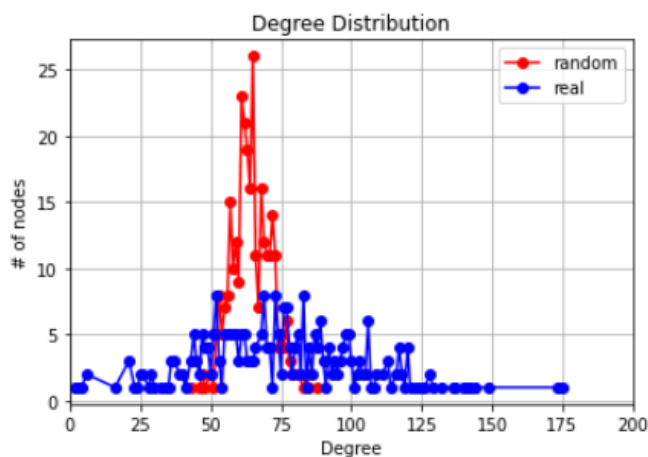
그래프 비교를 위해  $n=318$ ,  $p=0.20$ ,  $directed=False$ 인 Erdos-Renyi Random Graph를 생성하였다.

### 2.1 Avg degree

랜덤 그래프의 평균 degree는 64.516이고, 실제 그래프의 평균 degree는 74.774이다.

### 2.2 Degree distribution

<그래프 1>



<그래프 1>은 랜덤 그래프와 실제 그래프의 degree 분포를 나타낸다. 두 그래프가 형태는 비슷하지만 실제 그래프가 랜덤 그래프보다 더 넓은 범위의 degree를 갖는다는 점이 다르다.

### 2.3 Avg clustering coefficient

랜덤 그래프의 평균 cc는 0.204이고, 실제 그래프의 평균 cc는 0.527이다. 이러한 결과는 인간관계의 특성을 반영한 것 같다.

### 2.4 Avg shortest path

랜덤 그래프의 평균 path length는 1.796이고, 실제 그래프의 평균 path length는 1.823로 두 그래 모두 비슷한 값이 나왔다.

### 2.5 Connected Component

랜덤 그래프와 실제 그래프 모두 connected component가 1이 나왔다.

## ● Node centrality

### 2.6 Degree centrality

Degree centrality가 가장 높은 상위 10명은 다음과 같다.

```
[('박홍근', 0.5520504731861199),
 ('이철희', 0.5457413249211356),
 ('안민석', 0.4700315457413249),
 ('김성태', 0.45425867507886436),
 ('황주홍', 0.4479495268138801),
 ('서영교', 0.444794952681388),
 ('이찬열', 0.4416403785488959),
 ('박주민', 0.43217665615141954),
 ('이완영', 0.42902208201892744),
 ('노웅래', 0.41640378548895896)]
```

### 2.7 Closeness centrality

Closeness centrality가 가장 높은 상위 10명은 다음과 같다.

```
[('이철희', 0.6802575107296137),
 ('박홍근', 0.6788008565310493),
 ('안민석', 0.6522633744855967),
 ('황주홍', 0.6430020283975659),
 ('서영교', 0.6417004048582996),
 ('이찬열', 0.6417004048582996),
 ('김성태', 0.6404040404040404),
 ('박주민', 0.6352705410821643),
 ('이완영', 0.6327345309381237),
 ('노웅래', 0.6302186878727635)]
```

## 2.8 Betweenness centrality

Betweenness centrality가 가장 높은 상위 10명은 다음과 같다.

```
[('황주홍', 0.01443408821977514),  
 ('최인호', 0.013774154705553043),  
 ('김현아', 0.013086621458916115),  
 ('박홍근', 0.012291116973838235),  
 ('이철희', 0.011520104243502107),  
 ('김성태', 0.011421128035130528),  
 ('이찬열', 0.01136141827497521),  
 ('이완영', 0.010655921535207306),  
 ('나경원', 0.010337628648949061),  
 ('정병국', 0.009781770126644466)]
```

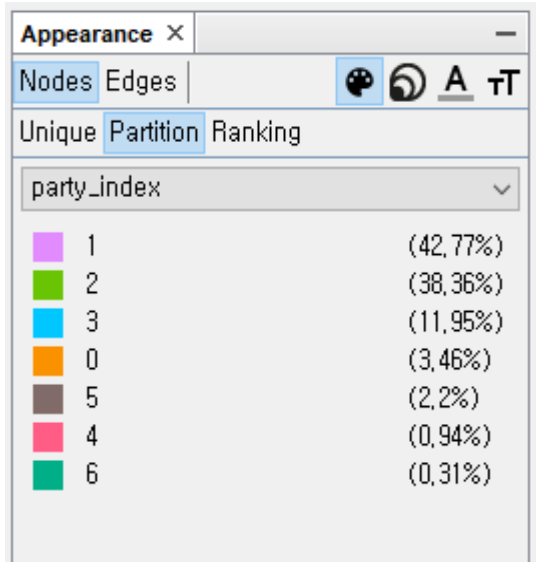
## 2.9 Eigenvalue centrality

Eigenvalue centrality가 가장 높은 상위 10명은 다음과 같다.

```
[('박홍근', 0.13317967311719),  
 ('이철희', 0.1314222380273776),  
 ('안민석', 0.1197098026087161),  
 ('서영교', 0.11275662992291674),  
 ('박주민', 0.11134321155679316),  
 ('박정', 0.10771201405645918),  
 ('우원식', 0.10581442678314361),  
 ('송옥주', 0.10538280172635693),  
 ('노웅래', 0.10494080780532103),  
 ('김해영', 0.10380752840454717)]
```

4개의 node centrality의 상위 10명을 보면 순위는 바뀌어도 구성원은 거의 비슷하다. 4개의 node centrality의 순위는 대체적으로 node의 degree에 의해서 정해졌다는 것을 알 수 있었다. 하지만 예외적으로 betweenness centrality에서는 degree가 상대적으로 낮은 node가 상위에 위치한 경우가 존재했다. betweenness centrality의 2등(최인호), 3등(김현아), 10등(정병국)은 모두 degree가 상대적으로 낮음에도 불구하고 betweenness centrality에서 상위 10명 안에 들었다. <그래프 2>를 보면 이들이 상대적으로 degree가 작아 node의 크기가 크지 않아도 그래프의 중앙에 위치해 있는 것을 볼 수 있다.

## 3 그래프 시각화

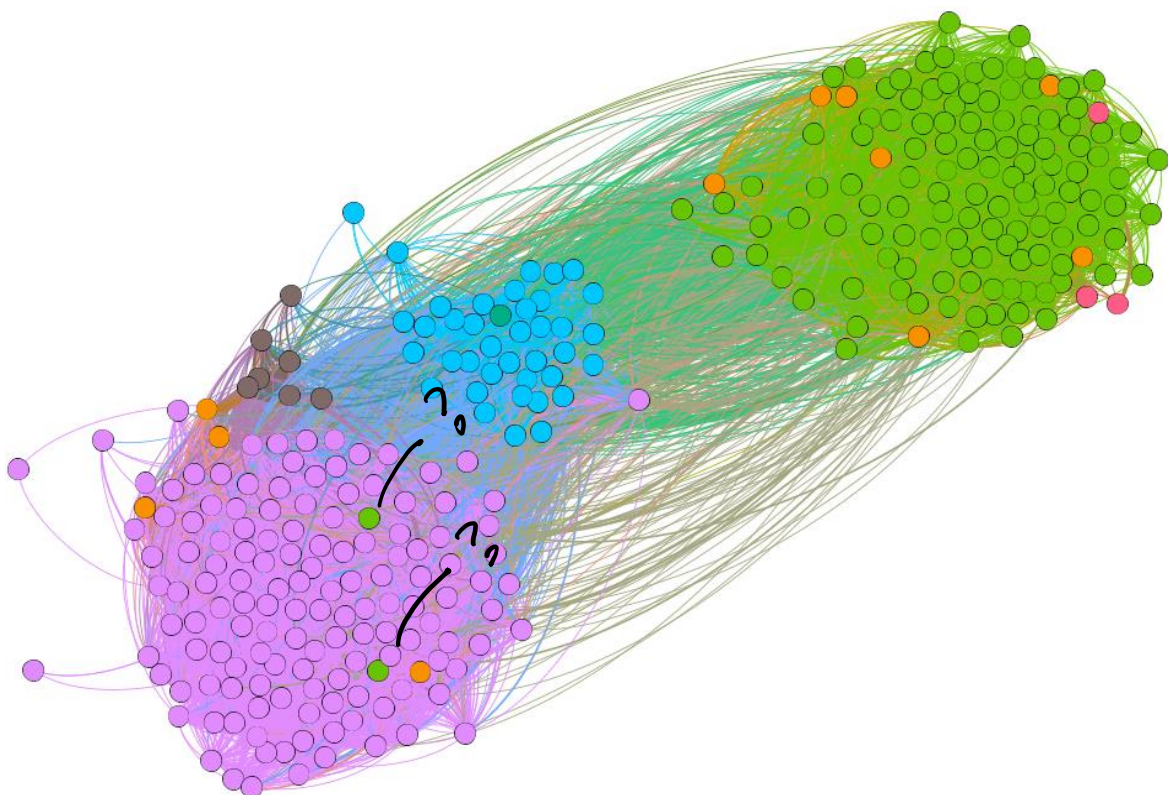


<정당 별 node 색상>	
분홍색	-1-더불어민주당
초록색	-2-새누리당
파란색	-3-국민의당
주황색	-0-무소속
갈색	-5-정의당
다홍색	-4-자유한국당
청록색	-6-바른미래당

그래프를 좀더 용이하게 분석하기 위해 node feature인 정당(party\_index)을 기준으로 node에 색을 입혀주었다. 각 색깔에 대응되는 정당은 다음과 같다.

<그래프 2>을 보면 무소속 국회의원인 주황색 노드를 제외하고는 정당별로 무리를 이루고 있는 것을 확인할 수 있다.

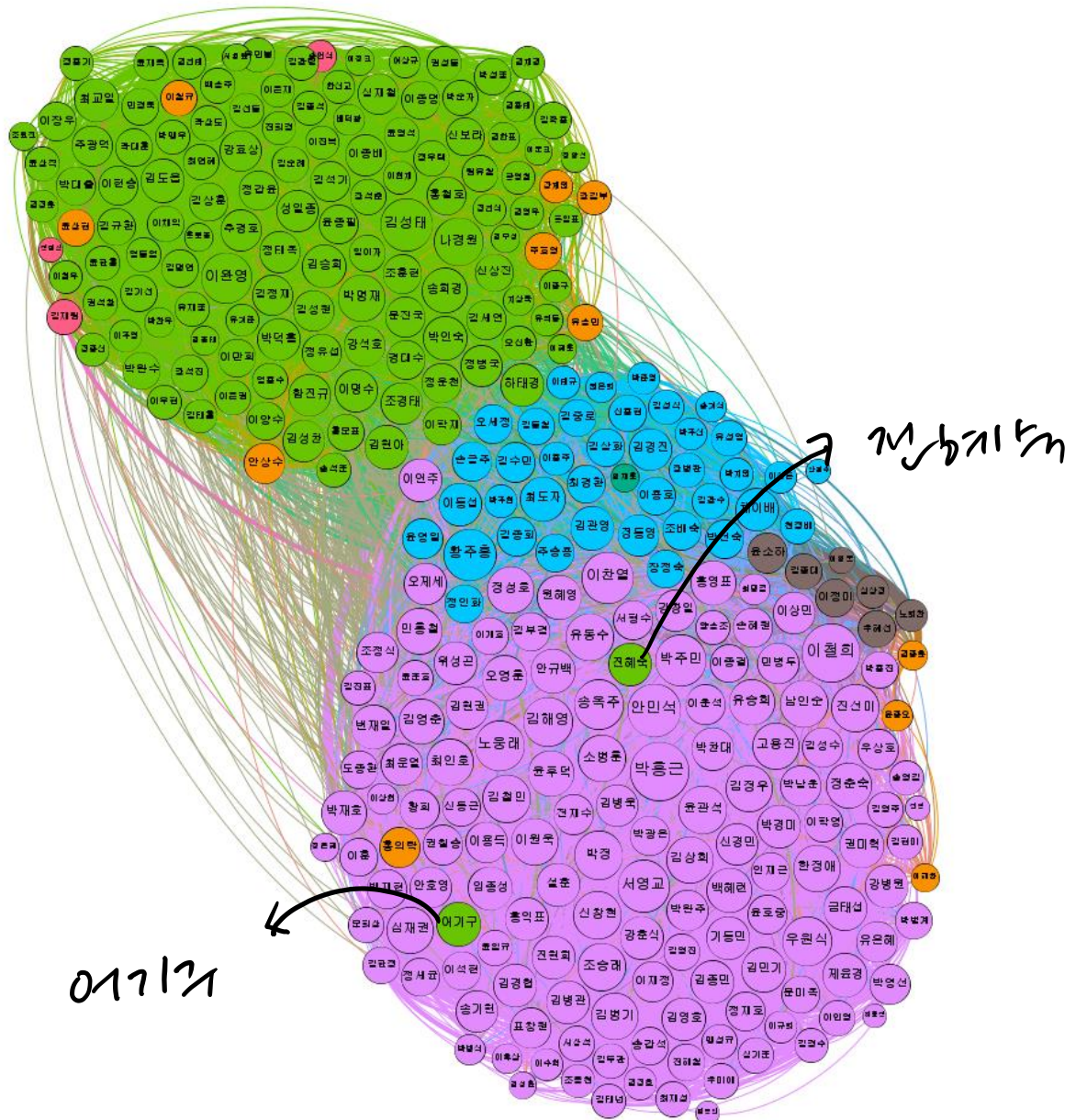
<그래프 2>





<그래프 3>는 더불어민주당(분홍색) 무리에 2명의 새누리당(초록색) 국회의원이 존재하는 것이 이상하여 2명의 신상정보를 확인하기 위해 node에 label을 붙이고 label adjust를 한 것이다.

<그래프 3>



그 결과, 더불어민주당(분홍색) 무리에 속해 있는 2개의 새누리당(초록색) node는 각각 '전혜숙' 국회의원과 '여기구' 국회의원인 것으로 나타났다. 네이버에서 두 국회의원의 신상정보를 검색해보니 두 국회의원 모두 새누리당이 아닌 더불어민주당 소속 국회의원으로 나타났다. people\_20.csv 파일의 정보가 잘못 입력되어 있었다.