**공동발의 법안을 통한 국회의원 네트워크 분석**

**2013-10299 오동건**

**1. 서론.**

**1.1 문제제기 및 연구주제**

박근혜 전 대통령 탄핵 이후, 정당 대내외의 크고 작은 변화가 지속되고 있다. 분당, 탈당, 복당, 합당과 창당 등 다양한 방식으로 정당 구조의 재구성 되고 있고 2020년 치뤄질 총선을 앞두고, 대규모 재편이 이루어질 가능성이 크다. 이에 본 분석에서는 국회의원들간의 네트워크 분석을 통해 현 상황을 진단하여 앞으로의 변화에 대한 예측을 하고자 한다. .

정당의 재구성은 국회의원들의 이해관계와 친밀도 그리고 정치성향 등 다양한 요인을 통해 이루어진다. 그러나 이해관계, 정치적 성향, 친밀도 등과 같은 추상적인 지표는 측정하기가 매우 어렵다. 따라서 본 분석에서는 추상적인 지표를 측정하는 대안으로 공동 발의 법안 수를 사용하고자 한다. 하나의 법률안에는 공통된 사상, 가치관, 이해관계가 담기기 때문에, 동일한 법률안을 발의한 의원들끼리 유사한 가치관과 이해관계를 갖는다고 볼 수 있다.

**1.2 데이터 소개**

데이터는 2가지 종류를 사용했다. 첫번째로, 국회의원의 공동발의 법안이다. 2인 이상 공동 발의된 법률안이 대상이며, 2018년 12월 28일부터 2019년 6월 14일까지 총 169일의 데이터를 수집했다. 총 2900개의 법률안을 수집했다. 다만 해당 시기에 의원 자격이 소멸된 자는 제외하여 총 298명을 의원이 남았다. 국회의원 298명 목록은 다음과 같다. 두번째 데이터는 개별 국회의원 상세 정보이다. 지역구, 초선재선여부, 나이, 출생지, 학력 사항을 포함하며 정성적 분석에 활용했다.

**1.3** **데이터** **수집**

국회의원 공동발의 법안 데이터를 수집하기 위해서 동적 크롤링을 사용해야 했다. 수집해야 할 html 문서가 웹 상에 바로 나타나지 않고 클릭을 통해서 접근 해야하는 구조였다. 이에 클릭을 통한 접근은 Selenium을 사용했고 그 외의 parsing 및 데이터 수집에는 beautifulsoup 모듈을 사용했다. 또한 효율적인 수집을 위해 두 단계로 나누어 진행했다. 첫번째 단계에서는 의안 번호을 먼저 수집하여 접근 위치 목록을 만들고 다음 단계에서 접근 위치 목록을 바탕으로 데이터를 수집했다. 국회의원 상세 정보의 경우, Open API에서 제공해주는 정보에 한계가 있어서 지역구, 초선 여부만을 수집하고 네이버 인물 사전을 크롤링 하여 출생지, 나이, 학력 사항을 수집했다.

**1.4 전처리 과정**

동명이인 문제로 데이터 primary key 값은 한자명을 사용했다. 또한 심상정, 민병두 의원 같은 경우 한자에 한글이 포함되어 있어 오타를 수정했다. 전처리에서 가장 중요한 절차는 자료 구조를 분석 가능한 형태로 만드는 것이었다. 자료 구조는 아래 세 단계 과정을 거쳐 변화했다.

dataframe (2900,3) -> two\_mode\_matrix (298,2900)-> one\_mode\_matrix (298,298)

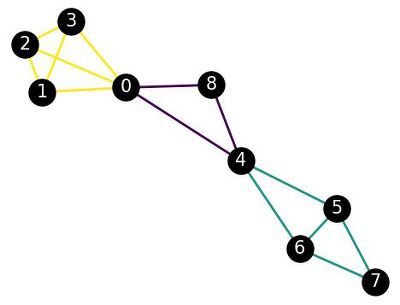
처음 수집된 데이터의 경우 의안 번호를 index 값으로 갖고 발의 참여한 의원들의 list가 들어가 있는 형태였다. 이를 one mode matrix로 만들어서 개별 의원이 row index를 의미하고 각 의안이 coloumn을 의미하는 298 by 2900 matrix를 만들었다. 이 매트릭스에서 value 값은 개별의원은 각 의안에 참여했는지 여부에 따라 1,0 값을 갖는다. 이렇게 만들어진 two mode matrix를 내적하여 one mode matrix를 만들었다. 이 매트릭스에서 대각 행렬은 개별 의원이 만든 총 법안 수를 의미하고 나머지 원소들은 서로 다른 의원들이 함께 참여한 의안 수를 의미한다.

**2. 분석방법**

본 장에서는 어떤 방법으로 국회의원의 네트워크 분석을 할 지를 설명한다. 유사도 측정의 단위와 측정방법을 설명하고, 측정된 유사도를 활용한 군집화 방법과 탐색적 자료 분석에 활용할 시각화 방법을 설명한다.

**2.1 유사도 측정 단위: Edge 기반의 링크 커뮤니티.**

본 분석에서는 Edge 기반으로 계층적 군집화를 하고 덴도그램, 차원축소를 활용해 탐색적 자료분석을 하고자 한다. 커뮤니티의 구조를 파악하는데, egde 기반의 링크 커뮤니티는 edge의 유사성을 기반으로 단계별로 군집화 할 수 있기 때문에 덴도그램을 통해 network의 관계를 세부적으로 파악할 수 있다. 반면, Modularity를 이용한 node 기반의 커뮤니티 추출은 betweeness centrality가 높은 Edge를 잘라가면서 Modularity를 최대화하는 방향으로 커뮤니티를 탐색하기 때문에, 노드들 사이의 연결은 고려하기 어렵다. 또한 modularity 기반의 community structure의 경우 resolution limit으로 인해 작은 집단을 파악하기 어렵다는 한계도 있다[[1]](#footnote-1). 따라서 edge기반의 링크 커뮤니티 탐지를 하고자 한다. 아래 그림은 링크 커뮤니티 탐지의 예이다.

Figure link community 예시[[2]](#footnote-2)

위 그림에서 확인할 수 있듯이 link 기반의 커뮤니티 탐지는 node로 커뮤니티를 나누지 않고[[3]](#footnote-3)edge를 기반으로 커뮤니티를 나누는 것을 확인할 수 있다. 만일 Modularity 방식을 사용했다면, 2와 4는 각각 노란색과 파란색 커뮤니티에 속하고 8은 어떠한 커뮤니티에도 속하지 못했을 것이다. 하지만 Edge 기반의 link community 방식에서는 0이 노란색, 보라색 커뮤니티 둘다 관계가 있음을 확인할 수 있다. 둘 중 어떤 방식이 옳다고는 할 수 없지만, 국회의원의 네트워크 분석에서는 후자가 사회현상의 다층적인 해석을 위한 정보를 제공한다고 판단했다.

**2.2 유사도 측정 방법: cosine 유사도**

Edge 간의 유사성은 cosine 유사도를 사용한다. link 커뮤니티 탐지에서는 jarccard coefficient 또는 tanimoto coeffient를 사용한다. 그러나 분석에서 사용된 데이터와 같이 node들이 특정 node들끼리만 자주 연결되어 있고 node들 사이의 공통된 edge가 적은 경우, 유사도에 대한 만족스러운 결과를 나타내기 어렵다[[4]](#footnote-4).

이에 대안이 되는 Cosine 유사도는 문서들 사이의 거리를 측정하는데 주로 사용되는 측도이다. 문서들 사이의 거리는 공통된 단어를 통해 두 문서 사이의 거리를 재는데, 서로 다른 주제의 문서의 경우 공통된 단어가 드물고 대부분이 0값을 갖는다. 또한 문서의 길이가 긴 문서는 공통된 단어의 출현 빈도가 상대적으로 높을 수 밖에 없다. 이 와 같이 sparse한 자료구조를 갖고 정규화가 반드시 필요한 데이터에 cosine 유사도가 주로 사용되고 있다.

본 분석에서 사용되는 데이터도 상기 서술한 특징을 갖는다. 국회의원들은 이전에 법안을 만들었던 사람들과 지속적으로 법안을 만드는 경향이 있고 그렇지 않은 사람과는 연결이 없다. 네크워크 매트릭스에서 0값의 비율이 31.5% 정도이고 3보다 작은 값의 비율은 64%에 이른다. 그리고 법안을 상대적으로 많이 만드는 국회의원의 경우, 다른 의원들과의 연결 강도가 높을 수 밖에 없기 때문에 공정한 비교를 위해서 정규화가 필수적이다.

**2.3 군집 방법과 군집의 수 결정: 계층적 군집 분석**

cosine 거리를 기반으로 측정된 Edge간의 유사도[[5]](#footnote-5)는 계층적 군집을 하는 데 활용된다. 군집을 연결하는 기준은 평균 연결법을 사용했다. 적절한 클러스터의 수의 경우, Elbow 기법을 통해 정했다. Elbow 방법을 통해 클러스터가 몇 개로 나누었을 때, 가장 최적의 결과를 줄 지 판단할 수 있다[[6]](#footnote-6). 이렇게 파악된 군집을 ‘상위 군집’으로 정의하고 각각의 상위 군집이 세부적으로 어떤 하위 구조를 가지고 있는지 파악한다.

**2.4 시각화 및 탐색적 자료분석: 덴도그램과 차원축소(MDS)**

탐색적 자료분석을 위한 시각화 방법은 덴도그램과 다차원척도법을 활용한 차원축소를 활용한다. 계층적 군집을 통해 군집이 결정되면 덴도그램을 통해 각각의 군집이 세부적으로 어떻게 연결되었는지 살펴볼 것이다. 그리고 차원 축소를 통해 298차원의 matrix를 2차원으로 표현하여 네트워크 구성의 전반을 살펴본다.

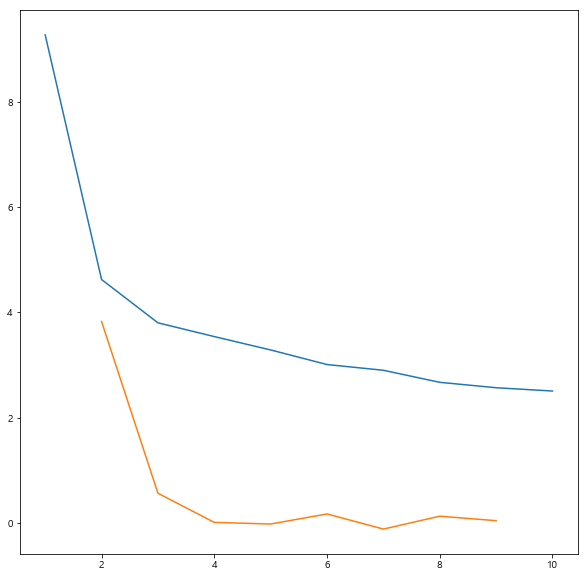
**3. 분석결과**

본 장에서는 분석결과를 설명한다. 클러스터링 결과를 설명하고 전체적인 양상을 탐색한 후, 상위 군집 각각을 세부적으로 탐색한다.

**3.1 군집수 결정과 군집화 결과**

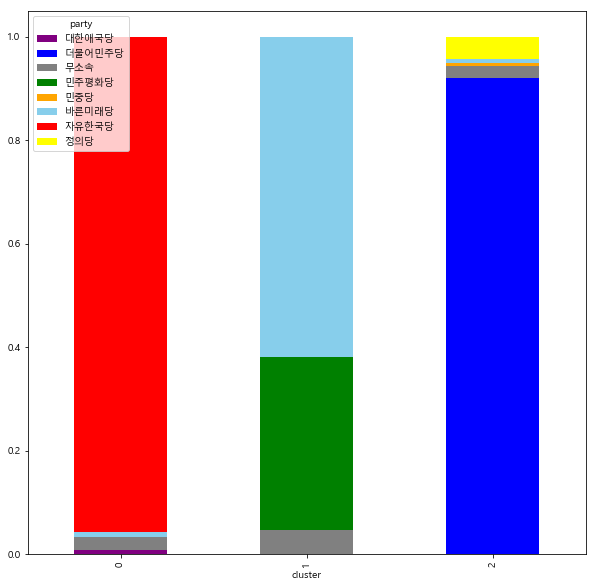
Elbow 기법을 통해, 최적의 커뮤니티의 수를 3개로 정하였다. 아래 그래프의 파랑색 선은 군집간의 거리의 감소량을 의미하고 주황색 선은 파랑색 선의 미분 값으로 거리가 감소하는 속도를 의미한다. 아래 그래프를 통해서 알 수 있듯이 클러스터의 개수가 3인 지점이 elbow임을 확인할 수 있다. 클러스터의 수가 3이상이 되면 더 이상 거리의 감소량이 크게 변하지 않고 감소 속도도 클러스터가 3인 경우를 기점으로 해서 확연하게 떨어진다.

**Figure 2 계층적 군집의 elbow 그래프**



아래 막대그래프는 상위 군집의 정당 비율에 대한 정보를 제공한다. 첫번째 군집은 대부분 자유한국당 의원들로 구성되어 있다. 그 외에도 대한애국당, 바른미래당, 무소속 의원들로 구성이 되었으며 총 117명의 의원이 포함된다. 두번째 군집은 민주평화당의원과 바른미래당 의원들로 대부분 구성되었으며 무소속의원들도 있다. 세 군집 중 가장 적은 42명의 의원이 속해있다. 세번째 군집은 더불어민주당이 가장 많고 무소속, 바른미래당, 민중당, 정의당 의원들이 속해있으며 139명의 의원들이 있다.

Figure 상위 군집의 정당 비율

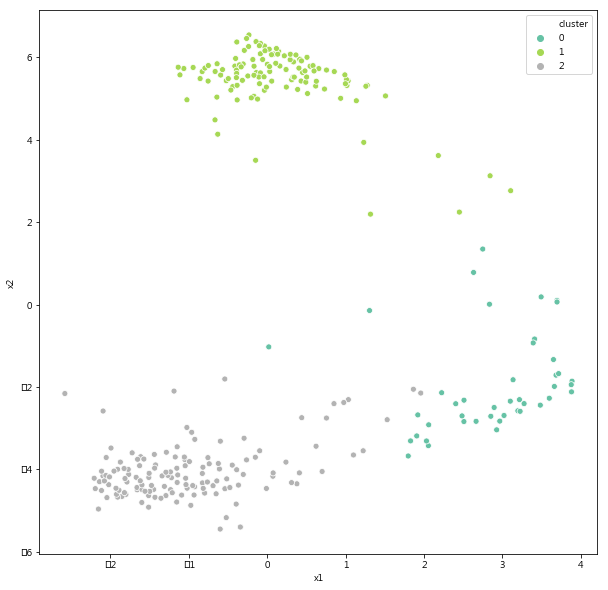
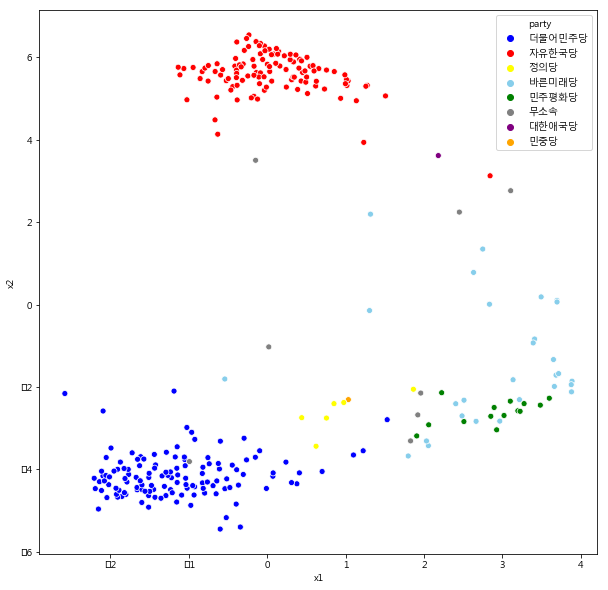


클러스터링 결과를 보면 소속정당에 따라 어느정도 구분됨을 확인할 수 있다. 다만 바른미래당과 민주평화당이 각각 자유한국당과 더불어 민주당으로 묶이거나 서로 다른 군집으로 구분되지 않는 점이 특징적이다. 이와 같은 특징은 군집의 수를 늘려도 여전히 유지된다. 이는 두 정당간의 연결이 다른 정당과의 연결에 비해 상대적으로 강하다고 추론해볼 수 있는데, 이와 관련해서는 다음 장에서 각각의 상위 군집을 세부적으로 분석할 때 자세하게 다루겠다.

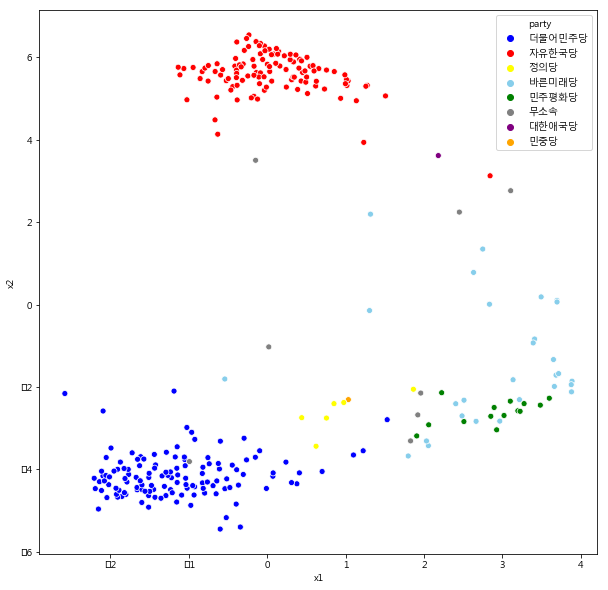
**3.2 차원 축소를 활용한 네트워크의 전반적인 양상 분석**

유사도 매트릭스를 다차원 축소법을 통해 2차원으로 그래프 상에 표현한 결과는 다음과 같다. 두 가지 기준으로 클러스터를 구분했는데, 왼쪽 그림은 정당에 따른 구분이고 오른쪽은 유사도에 따른 구분이다.

Figure 정당에 따른 구분 Figure 유사도에 따른 구분



정당에 따른 구분을 보면, 공동발의 법률안 데이터 만으로 특정한 위치임에도 불구하고 정당끼리 잘 모여 있다. 또한 민주평화당의원과 바른미래당 의원들이 혼재한 점도 특징적이다. 그리고 무소속의원들의 경우, 정당과 정당들 사이에 주로 위치하는 것이 주목할 만 한다. 이를 통해 무소속 의원들의 경우, 정당의 영향을 덜 받는 것을 추측해 볼 수 있다.



**Figure 6 무소속 의원들의 상대적인 위치**

그림 4에서 또 한가지 특징적인 부분은 정당에 따른 점들의 산포이다. 자유한국당, 민주평화당, 정의당, 더불어민주당은 비교적 유사한 위치에 국회의원들이 모여있다. 그러나 바른미래당 같은 경우, 점들의 산포가 상대적으로 크다. 이를 통해 바른미래당 소속 의원들의 내부 결속이 다른 정당에 비해서 약하다는 것을 추론해볼 수 있으나, 탐색적인 자료분석에 의한 감각적인 주장이기 때문에, centrality에 대한 정밀한 분석이 필요할 것이다.

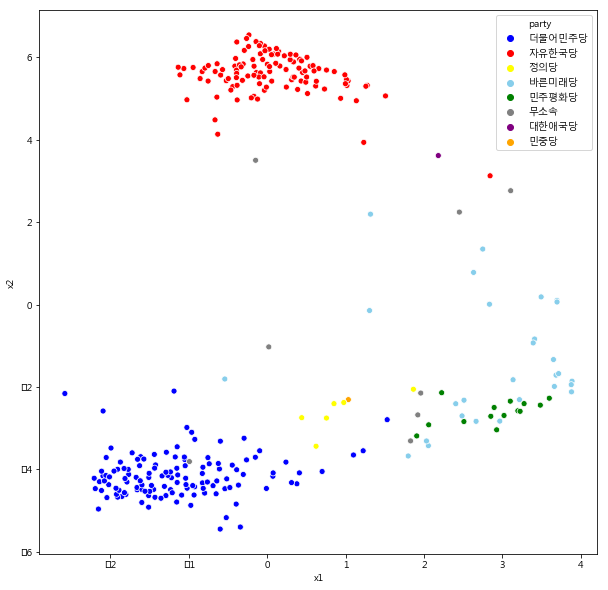


Figure 정당에 따른 산포

**3.3 덴도그램을 활용한 상위 군집의 세부적인 분석.**

본 장에서는 덴도그램을 통해 각각의 상위 군집이 어떤 구조를 가지는 지 확인한다. 전체 덴도그램은 다음과 같다. 자유한국당 의원들이 다수 포함된 상위 군집1과 더불어민주당 의원들이 다수 포함된 상위 군집2, 바른미래당 의원과 민주평화당 의원이 다수 포함된 상위 군집3으로 나뉜다. 각각의 군집은 초록, 빨강, 하늘색으로 표현되었다.

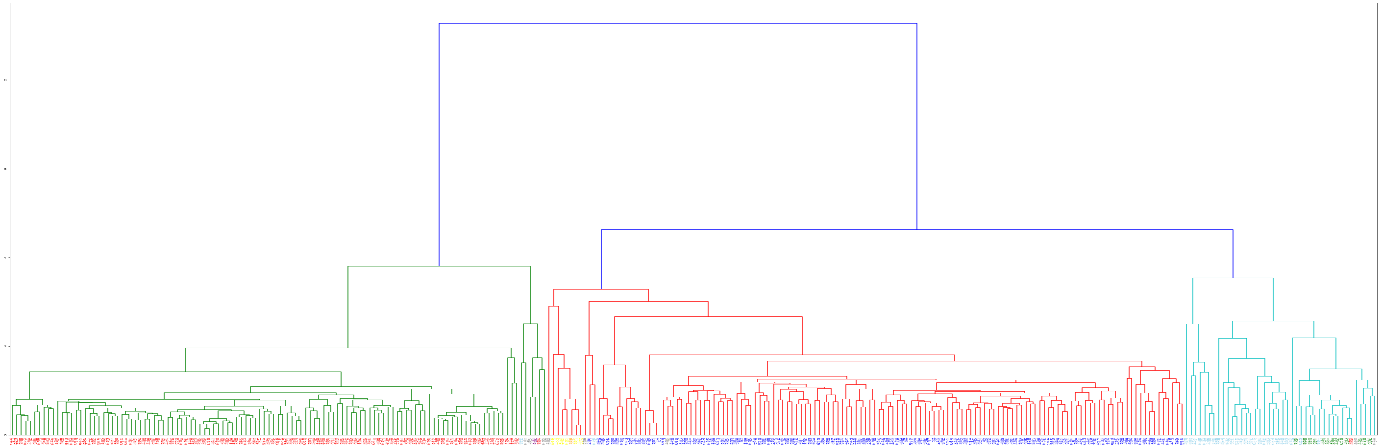


Figure 전체 덴도그렘

**3.3.1 상위 군집1 세부 분석**

상위 군집 1은 자유한국당 의원이 포함된 군집으로 총 117명의 의원들이 있으며 세부 구조는 아래와 같다.

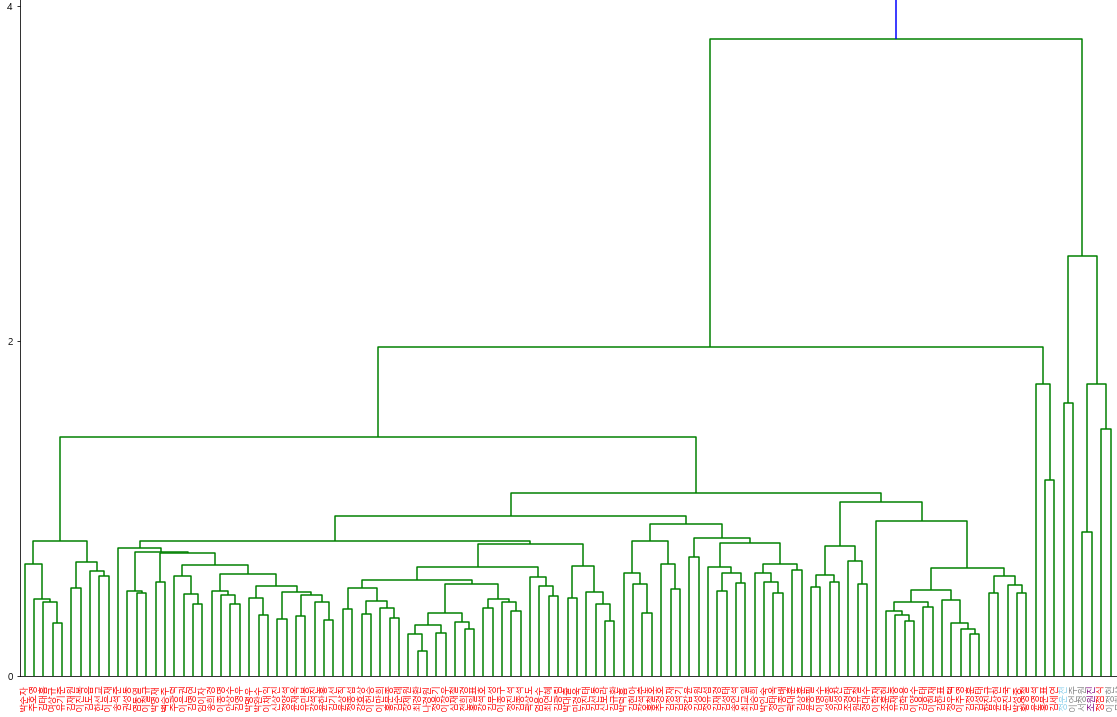
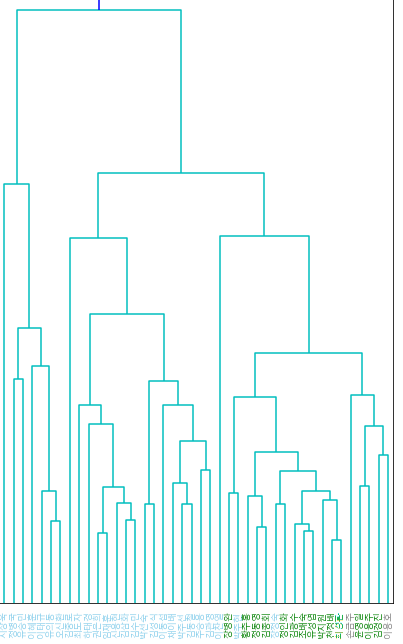


Figure 상위 그룹1의 세부구조

상위 그룹 1은 빨간색을 기준으로 크게 두개의 그룹을 나뉘는데, 오른쪽 그룹은 자유한국당 의원들의 대부분이 포함되어 있고, 왼쪽 그룹의 의원들은 차례대로 정운천, 이언주, 서청원, 조원진, 정점식, 이정현 의원이다. 이 그룹의 개별의원들은 자유한국당의 주류 그룹과 약한 연결을 갖는다는 특징이 있다. 개별 의원을 살펴보면 서청원, 조원진, 이정현은 대표적인 친박계 의원이며 정점식의원은 황교안 대표의 측근으로 분류할 수 있고 정운천 과거 새누리당의원 이었다가 바른미래당으로 당적을 변경했으며, 이언주 의원도 바른미래당 소속이었다 현재는 무소속이다. 개별의원들의 특징을 통해 3가지 가설을 세울 수 있는데, 첫 번 째는 자유한국당 내부에서 친박계 인사들과의 연결이 상대적으로 약해졌음을 추론해 볼 수 있다. 두 번 째는, 황교안 대표의 측근으로서 2019년 보궐선거에 정점식 의원이 주류 그룹에 포함되지 못했다는 점이다. 이는 당 내에서 황교안 대표의 영향력을 짐작해볼 수 있는 근거가 될 수 있다. 같은 시기 보궐선거로 당선된 정의당 여영국 의원은 소속당의 의원들과 강한 연결이 확인된다. 세번째는 바른미래당 의원들 중 일부 의원들은 자유한국당과의 연결이 더 강하는 점이다. 이언주 의원의 경우, 당 지도부와 마찰을 통해 갈등관계를 보여주었다는 점에서 자명한 결과로 해석된다. 정운천 의원의 경우, 정운천 의원의 지역구로 이를 설명할 수 있다. 정운천 의원의 지역구는 전북 전주로 정운천 의원은 탄핵정국이후 나타난 반새누리당 정서에 더욱 밀접하게 반응했을 것으로 보여진다. 이에 자유한국당과의 관계를 유지하면서 전략적인 선택으로 당적을 바른미래당으로 바꿨음을 추측해볼 수 있다.

**3.3.1 상위군집2 세부 분석**

상위 군집2는 민주평화당 의원과 바른미래당 의원이 혼재한 그룹으로 총 42명이 존재한다. 전체규모는 가장 작지만 가장 복잡한 세부 구조를 갖는다. 세부 구조는 아래와 같다[[7]](#footnote-7).

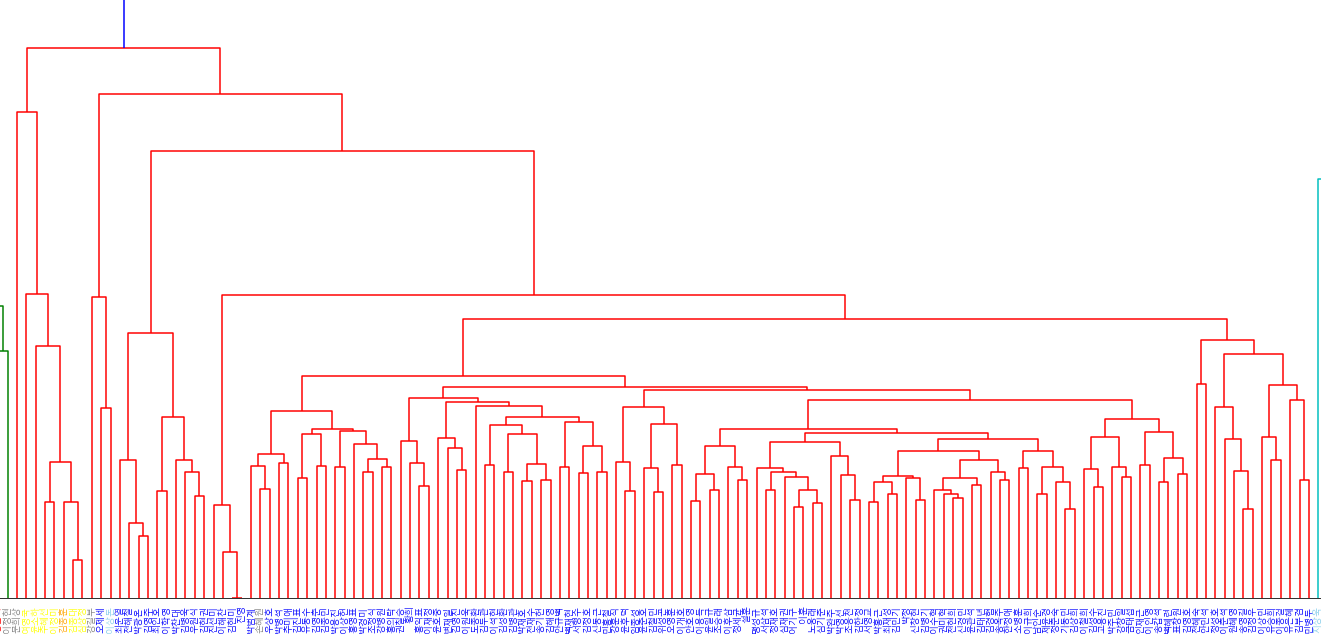


세부 구조를 살펴보면, 민주평화당과 바른미래당의 확연히 구분됨을 알 수 있다. 상위 군집2를다시 세 부분으로 나누어 보면 유승민의원이 포함된 첫번째 그룹은 유승민(대구)를 제외하고 모든 의원의 지역구가 경기도와 서울이다. 비례대표인 이태규의원의 경우 출생지가 경기도 양평으로 공통점을 갖는다. 두 번째 그룹의 경우, 바른미래당의 대부분의 비례대표의원이 포함되어있다. (9명) 그리고 지역구의원의 경우 호남계 의원이 상당수이다. 이 그룹에는 총 7명의 지역구 의원들이 존재하고 그 중 하태경 의원(부산 해운대구), 김성식 의원(서울 관악구)을 제외하고 5명의 의원이 호남을 지역구로 두고 있다. 마지막으로 비례대표의 경우 호남 지방이 4명 충청 지방이 4명 경기도 1명으로 주로 충청과 호남에 관련된 의원들이 많았다. 세번째 그룹의 경우 민주평화당 의원이 모두 포함되어 있으며 바른미래당의 박주현, 장정숙, 이찬열 의원과 무소속 손금주, 이용호 의원이 포함되어 있다. 바른미래당 의원의 경우 각각 군산, 대전, 화성이 고향으로 경향성이 없다. 다만 무소속인 손금주, 이용호의원은 고향이 모두 호남 지방이다. 두 의원의 경우, 2018년 12월 이후 더불어민주당으로 입당과 복당 신청을 했다가 반려된 경험[[8]](#footnote-8)이 있는데, 그 후 무소속 상태를 유지하며 민주평화당과 관계를 강화하고 있는 것으로 보인다.

상위 군집2의 구조를 보면 바른미래당과 민주평화당이 같은 군집으로 묶이지만 세부적으로 확연히 구분됨을 알 수 있다. 그리고 바른미래당 내부에서도 지역구와 고향에 따라서 그룹이 나누는 경향이 확인되었고 그 중 호남계가 다수 포함된 그룹은 수도권계 바른미래당 의원들 보다 민주평화당의원과 상대적으로 가깝다는 것이 확인되었다. 이러한 특징들이 앞서 확인했던 바른미래당 의원의 높은 산포와 관련됨을 추측해볼 수 있다.

**3.3.3 상위그룹3 세부 분석**

상위그룹3은 가장 많은 인원인 139명이 포함되었고 대부분 더불어민주당과 정의당의원이다. 세부 구조는 다음과 같다.



상위그룹3의 경우 크게 정의당과 더불어 민주당으로 나뉘며 세부 집단의 뚜렷한 경향성을 찾기 상위 군집3의 경우 세부 집단의 경향성을 찾기 어려웠다. 다만 문희상, 이상돈, 강길부, 오제세 의원 몇몇 개별 의원의 연결이 특이하다. 문희상의원은 정의당과 연결이 더불어민주당 의원들 보다 더 가까운데, 이는 문희상 의원이 후반기 국회의장이 됨에 따라 법률안을 발의 참여가 적어서 생긴 결과로 보여진다. 최근 6개월동안 문희상 의원이 포함된 발의 법안은 5건에 불과하다. 이상돈, 강길부, 오제세 의원의 경우 정당은 서로 다르지만 법률안 데이터 상으로는 유사한 패턴이 보인다. 세 의원들은 모두 더불어민주당과 법률안을 많이 만들지만 자유한국당, 바른미래당, 민주평화당등 다양한 정당의 의원과도 연결된다. 이러한 특징 때문에 세 의원이 서로 가까운 위치에 있는 것으로 보인다..

**4. 결론**

정치는 수 많은 요인들에 의해 이루어진다. 정치적 이해관계, 친밀감, 재선 가능성 등등 국회의원의 네크워크를 구성하는 데는 다양한 요인이 있다. 본 분석에서는 법률안을 통해 국회의원들의 네트워크를 단편적으로 파악할 수 있었다. 이에 결론에서는 앞선 분석결과에서 세운 가설을 통해 2020년 총선 전망을 예측하고 연구 한계와 향후 연구 방향에 대한 제언으로 마무리하고자 한다.

**4.1 친박 정당 출현과 황교안 대표의 불확실성**

공동발의 법률안을 기반으로 한 네트워크 상에서 서청원, 조원진 의원 같은 진성 친박계 의원들과 자유한국당의원들의 연결 그리 강하지 않음이 확인되었다. 이를 통해 최근 홍문종의원 탈당은 친박의원들간의 강한 유대를 바탕으로 진행된 것이 아니라 공천 탈락을 염두한 전략적 판단으로 보인다. 따라서 총선 공천과정에서 갈등이 빗어지지 않는 한 대규모 탈당이 일어날 가능성을 낮을 것으로 보여진다.

하지만 황교안 대표의 불확실성이 변수로 남는다. 2019년 보궐선거에서 황교안 대표가 공천한 정점식의원이 자유한국당 주류 그룹에 어울리지 못하고 있고 이를 통해 정당 내에서 황교안 대표의 영향력이 약함을 추론해볼 수 있다. 따라서 황교안 대표가 총선을 통해 자신 영향력을 키우기 위해 기존 의원들을 배제한 공천을 진행할 가능성이 있다. 하지만 이는 오히려 역효과를 불러올 가능성이 있기 때문에 신중하게 진행될 것이다. 공천으로 인한 마찰이 강화되면 어떠한 명분으로든 자유한국당의원들의 대규모 탈당 가능성이 존재하기 때문이다.

**4.2. 정당 간의 합당과 개별 의원의 당적변경.**

법률안을 바탕으로 살펴볼 때, 각각의 정당은 정당에 따라 차이는 존재하지만 어느정도 독자성을 갖게 된 것을 확인할 수 있다. 네트워크 상에서 서로 다른 정당이 무질서하게 혼재 되기 보다는 같은 정당 의원들끼리 네트워크가 모여 있음이 확인되었다. 따라서 총선을 앞두고 정당들 사이에 공식적인 합당은 이루어질 가능성이 적을 것으로 보인다. 다만 지역구 의원들 사이에 개별적인 단일화는 이루어질 가능성이 높다. 특히 호남지방에서 단일화가 이루어질 가능성이 크다. 바른미래당, 민주평화당, 더불어민주당 모두 호남 지역의 지역구 의원들이 존재하며 비례대표들 또한 고향이 호남인 의원들이 많다. 이에 세 정당 모두 호남에 유력한 후보를 낼 가능성이 있고 3인 구도로 총선이 이루어진다면 두 후보 사이에 단일화가 이루어질 가능성이 높다.

**4.3 요약 및 향후 연구 방향**

이상으로 분석결과를 바탕으로 현 국회의원 네트워크를 진단하고 2020년 치뤄질 총선 전망을 예측해보았다. 공동 발의된 법률안 만을 바탕으로 한 분석이기 때문에 모든 현상을 정확히 설명하기에는 어려움이 있었을 것이다. 하지만 법률안을 바탕으로 한 군집이 실제 정당과 상당히 유사하다는 측면에서 법률안을 통한 네트워크 분석의 타당성을 확인할 수 있었다.

다만, 6개월의 데이터를 모두 똑같은 비중으로 다루었고 한 시점에 대한 분석을 했다는 점은 한계로 남는다. 분석시점에서 법률안을 함께 만든 것과 6개월전에 법률안을 함께 만든 것은 서로 다른 비중으로 다루어야 할 것이다. 분석 시점과 가까운 시점에 이루어진 연결일수록 두 개체의 관계를 더욱 잘 설명할 수 있고 과거의 데이터일수록 현재를 설명하는데 어려움이 있다. 따라서 향후 연구는 시간에 경과에 따라 법률안에 가중치를 줘서 군집을 업데이트 하며 그 변화를 추적하는 방식을 시도해 볼 수 있을 것이다.

1. <https://en.wikipedia.org/wiki/Modularity_(networks)> [↑](#footnote-ref-1)
2. https://mons1220.tistory.com/108?category=804486 [↑](#footnote-ref-2)
3. 직관적으로 그래프에서 점이 아닌 선에 색이 칠해져 있는 것으로, 두 방식의 차이를 이해할 수 있다. Node 기반의 커뮤니티 탐지에서는 점에 색을 칠하는 방식으로 커뮤니티가 표현된다. [↑](#footnote-ref-3)
4. <http://mines.humanoriented.com/classes/2010/fall/csci568/portfolio_exports/sphilip/cos.html> 의 Analysis [↑](#footnote-ref-4)
5. 계층적 군집의 경우, 비유사도를 기반으로 하기 때문에 실제로는 1- 유사도의 값이 사용된다. [↑](#footnote-ref-5)
6. <https://joernhees.de/blog/2015/08/26/scipy-hierarchical-clustering-and-dendrogram-tutorial/> [↑](#footnote-ref-6)
7. 최경환 의원의 경우, 동명이인이 있어 빨간색(자유한국당)으로 표시되었는데 이를 수정했다. [↑](#footnote-ref-7)
8. <http://www.hani.co.kr/arti/politics/assembly/878216.html> [↑](#footnote-ref-8)