

Cosponsorship networks in the 17th National Assembly of Republic of Korea

Chanmoo Park^a · Woncheol Jang^{a,1}

^aDepartment of Statistics, Seoul National University

(Received March 8, 2017; Revised April 16, 2017; Accepted April 16, 2017)

Abstract

In this paper, we investigate cosponsorship networks found in the 17th National Assembly of Republic of Korea. New legislation should be sponsored by at least 10 legislators including one main sponsor. Cosponsorship networks can be constructed, using directional links from cosponsors of legislation to its main sponsor; subsequently, these networks indicate the social relationships among the legislators. We apply Exponential Random Graph Model (ERGM) for valued networks to capture structural properties and the covariate effects of networks. We find the effect of the same party has the greatest influence on the composition of the network. Mutuality also plays an important role in the cosponsorship network; in addition, the effect of the number of elections won by a legislator has a small but significant influence.

Keywords: centrality, cosponsorship, exponential random graph model, mutuality, party match

1. 서론

한 나라의 정치적인 구조를 연구하고 그 특성을 살펴보려는 노력이 정치학을 비롯한 각종 분야에서 많이 이루어지고 있다. 그 중에서도 특히 입법 활동의 중심에 있는 국회와, 국회를 구성하는 국회의원들에 대한 연구 또한 활발히 이루어져 왔다.

반면, 국회의원들이 입법하는 행태에 관하여 실제 자료를 이용한 통계적 분석 연구는 상대적으로 그 수가 적다. Lee와 Youm (2009)는 17대 국회에서 발의된 전체 5,728개의 법안 중에서 보건복지위원회 소속 의원들 간에 주고받은 공동발의를 대상으로 네트워크를 구성하였다. 공동발의 자료는 법안을 제안하는 대표발의자와 대표발의자의 제안을 받는 공동발의자들로 구성되어있어, 대표발의자와 공동발의자 간의 방향성 있는 네트워크를 구성할 수 있다. 구체적으로 공동 발의 제안의 유무를 이용하여 네트워크를 구성하여 Exponential Random Graph Model (ERGM)을 이용하여 그 구성원리에 대해 분석을 하였다.

This work was carried out with the support of the Basic Science Research Program through the National Research Foundation (NRF) of Korea grant by the Ministry of Education (No. 2013R1A1A2010065) and the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIP) (No. 2014R1A4A1007895).

¹Corresponding author: Department of Statistics, Seoul National University, 1 Gwanak-ro, Gwanak-gu, Seoul 08826, Republic of Korea. E-mail: wjang@snu.ac.kr

본 연구에서는 공동발의라는 행위가 발생의 유무 뿐만 아니라 횟수가 있는 행위임에 착안하여 본래의 ERGM에서 국회의원들 사이의 공동발의 횟수를 추가적으로 고려한 모형인 ERGM for Valued Network (Krivitsky, 2012)를 이용하여 17대 국회의원의 입법 행태를 분석하고자 한다.

이 논문의 나머지 부분의 구성은 다음과 같다. 2장에서는 이 연구에 사용된 17대 국회의원의 법안 발의 자료와 공동발의의 개념, 그리고 공동발의라는 개념을 통해 정의한 공동발의 네트워크에 대하여 소개한다. 3장에서는 이 연구에서 공동발의 네트워크를 분석하는데 이용한 모형인 ERGM의 개념과 구체적인 모형으로 edge가 이항형(binary)일 때와 횟수형(count)일 때의 ERGM 모형을 간단히 소개한다. 4장에서는 공동발의 네트워크를 설명하는데 주요한 효과가 있을 것으로 기대되는 여러 가지 네트워크 충분통계량(sufficient statistics of network)들에 대해 소개하고 이들을 이용하여 설계한 모형들에 대해 소개한다. 또한, 모형들의 분석결과를 제시하고 결과에 대한 해석과 논의를 진행한다.

2. 데이터 설명

본 연구에서 사용된 데이터는 국회의안정보시스템에서 구할 수 있는 17대 국회의 법안 발의 정보와, 국회의원의 당적변경 자료이다. 법안 발의 정보에는 법안의 내용, 발의 시기, 발의 결과, 대표발의자, 공동발의자 등이 포함되어 있다. 이 장에서는 공동발의(cosponsorship)에 대한 설명과, 공동발의라는 개념을 이용하여 구축할 수 있는 공동발의 네트워크(cosponsorship network)에 대하여 정의하도록 하겠다.

2.1. 공동발의

국회법 제79조(의안의 발의 또는 제출)에 따르면, 대한민국 국회에서 법안의 발의가 유효하기 위해서는 최소 10명 이상의 국회의원이 공동발의를 하여야 한다. 이 10명에는 법안을 대표로 발의하는 대표발의자와 그 발의를 지지하는 공동발의자들을 포함한다. 따라서, 법안 발의시 공동발의자들이 대표발의자를 지원하는 행위를 공동발의라고 정의한다.

미국의 경우에도 공동발의에 요구되는 국회의원의 수가 최소 25명이며, Fowler (2006)와 같이 공동발의 네트워크를 통해 국회 내의 사회적 관계를 살펴보려는 연구들이 정치학계에서 이미 활발하게 이루어지고 있는 추세이다.

2.2. 공동발의 네트워크

2.1절의 공동발의에 대한 개념을 살펴보면, 이를 통해 국회의원 사이의 관계를 생각해볼 수 있다. 예를 들면, 어떤 A라는 국회의원이 대표발의한 법안에 B라는 국회의원이 공동발의하는 행위는 해당 법안의 발의를 유효하게 하기 위하여 B라는 국회의원이 A라는 국회의원을 지원하는 행위로 해석할 수 있다. 그렇다면, 마치 SNS에서 한 이용자가 다른 이용자에게 취하는 행동을 기준으로 이용자들 사이의 방향성 있는 네트워크를 정의할 수 있듯이, 국회에서도 법안에 대한 공동발의를 지원하는 행위를 기준으로 국회의원들 사이의 방향성 있는 네트워크를 정의할 수 있을 것이다.

이러한 원리로, 위의 Figure 2.1과 같이 하나의 법안 발의에 대하여 공동발의자들로부터 대표발의자로 향하는 방향성 있는 네트워크를 정의할 수 있다. 이 방법을 이용하여, 발의된 모든 법안들에 대하여 만든 네트워크를 공동발의 네트워크라고 정의할 수 있으며, 이러한 공동발의 네트워크는 국회에서 가장 의미있는 정치적 활동인 법안 발의라는 행동을 통해 본 국회의원들의 관계를 나타내므로, 어느정도 국회의원들의 정치적 관계를 유추할 수 있는 자료로 볼 수 있다. 17대 국회에서 발의된 법안에 대해 국회의원들의 공동발의 네트워크를 그려보면 다음 Figure 2.2과 같다.

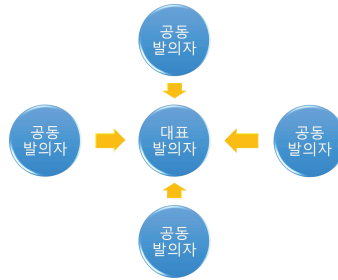


Figure 2.1. Formation of cosponsorship networks.

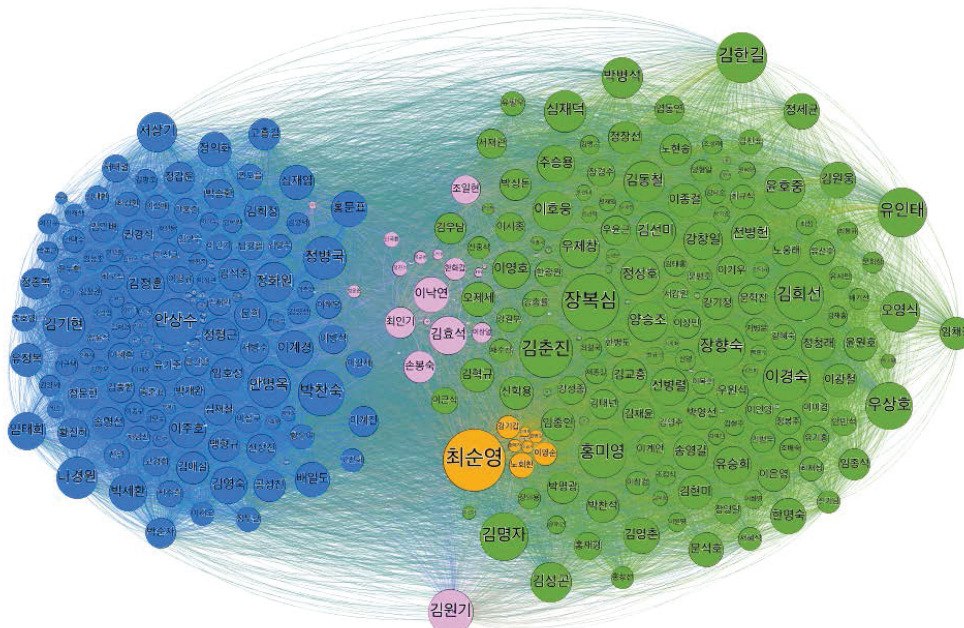


Figure 2.2. Cosponsorship networks of the 17th National Assembly members.

Figure 2.2는 Java 기반의 오픈소스 네트워크 Visualization software인 Gephi (Bastian 등, 2009)를 이용하여 그렸다. 원들은 각각 네트워크를 구성하는 국회의원들을 의미하며 원의 크기는 공동발의 지원을 받은 횟수에 비례하도록 그려진 그림이다. 연두색은 당시 열린우리당 소속, 파란색은 한나라당, 노란색은 민주노동당, 분홍색은 나머지 정당들에 소속된 국회의원들을 의미한다. 랜덤한 위치에서 시작하여, 주고받은 공동발의 지원 횟수를 국회의원 사이의 관계의 강도로 가정하여, 강한 강도의 관계를 가진 국회의원들끼리 가깝도록 위치를 수렴시키는 forceatlas2 알고리즘 (Jacomy 등, 2014)을 이용하여 각 vertex의 위치를 수렴시켰다.

이와 같은 그림을 통해 정량적인 분석을 하지 않고도 어느정도 같은 정당끼리의 공동발의가 활발함을 시각적으로 확인할 수 있으며, 우리가 뉴스나 신문 기사 등을 통해 자주 접했던 인물들의 vertex의 크기가 돋보임을 쉽게 알 수 있다. 하지만 이 연구에서는 그림을 통한 정성적인 분석이 아닌 모형을 통한 정량적인 분석을 진행해보려고 한다.

3. 지수 랜덤 그래프 모형 (ERGM)

ERGM은 edge에 대한 확률을 모형화하는 Random Graph Model의 한 종류로, 지수족(exponential family) 확률모형을 이용하는 모형이다. Random Graph Model의 가장 기본적인 가정은 현재 분석하려는 네트워크를 구성하는 vertex들 사이의 edge가 확률적으로 생성된 결과 즉, 랜덤 표본이라는 가정이다. 다음으로는 랜덤 표본으로 가정한 edge가 어떠한 분포로부터 추출된 것인지에 대한 가정을 추가하여 여러 가지 Random Graph Model로 특정할 수 있으며, ERGM의 경우 그 분포를 지수족에 속한 확률분포를 이용한다. 이 장에서는 ERGM의 일반적인 모형에 대한 정의 및 설명과, 가장 기본적으로 사용되는 이항형 edge에 대한 모형과 이 논문에서와 사용될 횡수형 edge에 대한 모형을 소개하겠다.

3.1. 지수 랜덤 그래프 모형이란?

네트워크에 존재하는 vertex의 갯수를 n 이라 하자. n 개의 vertex에서 발생할 수 있는 모든 edge들의 집합을 \mathbb{Y} 라 하면, $\mathbb{Y} = \{(i, j) : i, j = 1, \dots, n\}$ 로 표시할 수 있다. 각 edge가 가질 수 있는 값의 집합을 E 라 하면 (이항형 edge의 경우는 $E = \{0, 1\}$, 횡수형 edge의 경우는 $E = \{0\} \cup \mathbb{N}$, \mathbb{N} 은 자연수 집합), 발생할 수 있는 모든 네트워크의 집합은 멱집합(power set) $\mathcal{Y} = E^{\mathbb{Y}}$ 로 정의할 수 있으며 이 \mathcal{Y} 를 네트워크 공간이라고 정의하겠다. 네트워크 공간 \mathcal{Y} 에서 정의되는 네트워크 확률변수 \mathbf{Y} 의 형태는 네트워크에 속한 edge들이 가지는 값(Y_{ij})들의 집합 ($\mathbf{Y} = \{Y_{ij} \mid i, j = 1, \dots, n\}$)이며, Y_{ij} 역시 E 에서 정의되는 확률변수이다. \mathbf{Y} 에 대한 ERGM의 확률밀도함수는 네트워크 관측치 \mathbf{y} 에 대하여 다음과 같이 일반적으로 표현할 수 있다.

$$P(\mathbf{Y} = \mathbf{y}) = \frac{h(\mathbf{y}) \exp(\boldsymbol{\eta}(\boldsymbol{\theta}) \cdot \mathbf{g}(\mathbf{y}))}{\kappa(\boldsymbol{\theta})}, \quad \mathbf{y} \in \mathcal{Y}.$$

여기서, h 는 지수족 중에서 어떤 분포를 이용할 것인지를 정해주는 함수이며, $\boldsymbol{\eta}$ 는 모수 $\boldsymbol{\theta}$ 의 정준모수(canonical parameter)로의 사영이다. 일반적으로 선형성을 띠는 지수족을 빈번하게 사용하기 때문에 항등함수 $\boldsymbol{\eta}(\boldsymbol{\theta}) = \boldsymbol{\theta}$ 를 많이 사용한다. 네트워크의 관측치의 함수인 $\mathbf{g}(\mathbf{y})$ 는 네트워크를 형성하는 구조적 특성이나 공변량을 나타내는 네트워크 충분 통계량(sufficient statistics of network)이다. $\mathbf{g}(\mathbf{y})$ 와 이에 대응하는 모수들 $\boldsymbol{\theta}$ 는 ERGM에서 주관심사가 되는 부분으로, 해당하는 네트워크의 구조적 특성이 우연에 의해 발생한 것이 아니라 강한 확률로 등장하는 것인지, 혹은 공변량이 네트워크의 형성에 유의한 영향을 미치는지에 대한 분석을 위한 가설 설계를 통해 정해지는 부분이다. 마지막으로, κ 는 확률밀도함수의 조건을 만족시키기 위한 위한 정규화 상수로, 풀어쓰자면 다음과 같다.

$$\kappa(\boldsymbol{\theta}) = \sum_{\mathbf{y} \in \mathcal{Y}} h(\mathbf{y}) \exp(\boldsymbol{\eta}(\boldsymbol{\theta}) \cdot \mathbf{g}(\mathbf{y})).$$

이때, κ 의 계산은 네트워크 공간 \mathcal{Y} 에서 가능한 모든 경우의 수에 대한 계산인데, \mathcal{Y} 가 $E^{\mathbb{Y}}$ 꼴의 멱집합이며, \mathbb{Y} 는 $O(n^2)$ 을 따르는 것을 감안한다면, vertex의 크기가 조금만 커져도 매우 복잡한 계산을 알 수 있다. 따라서, 실제 추정에서는 이 계산을 Monte Carlo Markov Chain 접근 방법을 이용하여 계산한다는 것이 알려져 있으며, R의 `ergm` 패키지 (Hunter 등, 2008), `ergm.count` 패키지 (Krivitsky, 2016)에서 구현이 되어있다.

3.2. 이항형 네트워크를 위한 지수 랜덤 그래프 모형

네트워크 중에서 edge에 대해 추가적인 공변량이 없는 경우, 즉 vertex와 vertex가 연결이 되어 있는지의 여부만 나타내는 경우를 이항형 네트워크(binary network)라고 한다. 이러한 네트워크의 경우에는

이항형 edge를 가지기 때문에 각 edge가 가지는 값에 대한 확률변수인 Y_{ij} 에 대한 지수족 확률분포으로써 베르누이 분포를 이용할 수 있다. 이 경우에 네트워크 공간 \mathcal{Y} 에서 정의되는 네트워크 확률변수 \mathbf{Y} 에 대한 ERGM의 확률밀도함수는 다음과 같다.

$$P(\mathbf{Y} = \mathbf{y}) = \frac{\exp(\boldsymbol{\theta} \cdot \mathbf{g}(\mathbf{y}))}{\kappa(\boldsymbol{\theta})}.$$

3.3. 가중치 네트워크를 위한 지수 랜덤 그래프 모형

3.2절과는 달리 edge에 대해 공변량이 있는 네트워크를 가중치가 있는 네트워크(weighted network or valued network)라고 하며, 대 네트워크 형성에 이용된 관계의 횟수가 그 공변량이 되는 경우가 많다. 이러한 네트워크의 경우에는 횟수형 edge를 가지기 때문에, Y_{ij} 에 대한 지수족 확률분포로써 포아송 분포를 이용할 수 있다. 이 경우에 네트워크 공간 \mathcal{Y} 에서 정의되는 네트워크 확률변수 \mathbf{Y} 의 ERGM 확률 밀도함수는 다음과 같다.

$$P(\mathbf{Y} = \mathbf{y}) = \frac{h(\mathbf{y}) \exp(\boldsymbol{\theta} \cdot \mathbf{g}(\mathbf{y}))}{\kappa(\boldsymbol{\theta})}, \quad h(\mathbf{y}) = \prod_{i,j} (y_{ij}!)^{-1}.$$

4. 자료 분석

앞 장에서 소개한 ERGM을 이용한 네트워크 모형에서는 적절한 가설 설정을 통해 정한 네트워크 충분 통계량 $\mathbf{g}(\mathbf{y})$ 와, 이에 대응하는 $\boldsymbol{\theta}$ 의 추정을 통하여, 네트워크의 구성원리에 대한 추론과 해석이 가능해진다. 즉, 공동발의 네트워크에서 공동발의 행위가 어떠한 상황에서 더 잘 이루어지는가의 질문에 대한 답을 얻을 수 있게 된다. 이번 장에서는 이와 같이 공동발의 네트워크의 구성원리를 설명하기 위해 이 연구에서 세운 가설에 포함되는 ERGM 모형의 네트워크 충분 통계량 $\mathbf{g}(\mathbf{y})$ 들과 통계량들을 이용한 모형을 설명하고, 그 분석 결과를 해석하려고 한다. 또한, 그에 앞서 공동발의 네트워크를 구성하는 기본적인 기술통계량(descriptive statistics)을 통하여 기본적인 네트워크의 특성을 알아보겠다.

4.1. 기술통계량 분석

주된 분석에 앞서, 17대 국회의 공동발의 네트워크에 대하여 다음과 같은 기술통계량을 구해보았으며, 각 기술통계량들에서 가장 높은 5명을 Table 4.1에 나타냈다. 네트워크를 구성하는 vertex 집합 V 에 속하는 i 의 Centrality는 다음과 같이 여러 가지로 정의할 수 있다.

- Degree Centrality (Indegree, Outdegree)

$$c_{Id}(i) = \frac{d_i^{\text{in}}}{\sum_{i \in V} d_i^{\text{in}}}, \quad c_{Od}(i) = \frac{d_i^{\text{out}}}{\sum_{i \in V} d_i^{\text{out}}}.$$

Degree Centrality의 경우 위의 식과 같이 vertex i 의 Indegree 값(d_i^{in})과 Outdegree 값(d_i^{out})의 정규화 값을 의미하며, 공동발의 네트워크에서 어떤 국회의원이 상대적으로 얼마나 많이 법안 발의 지원을 주고 받는지에 대한 지표로써 활용될 수 있다.

- Closeness Centrality (Indegree, Outdegree)

$$c_{Cl}(i) = \frac{1}{\sum_{j \in V} \text{dist}(j, i)}.$$

Table 4.1. Descriptive statistics for cosponsorship networks in the 17th National Assembly.

Indegree	Outdegree	Incloseness	Outcloseness	Betweenness	Eigenvector
최순영(0.748)	엄호성(0.729)	최순영(0.799)	이정일(0.079)	엄호성(0.0132)	최순영(0.124)
장복심(0.685)	안상수(0.651)	장복심(0.761)	김홍일(0.078)	안상수(0.0126)	장복심(0.121)
김춘진(0.639)	정성호(0.648)	김춘진(0.735)	이인제(0.078)	장복심(0.0124)	김춘진(0.113)
김한길(0.607)	이해봉(0.642)	김한길(0.718)	박창달(0.078)	배일도(0.0101)	김한길(0.110)
유인태(0.589)	박상돈(0.620)	유인태(0.709)	정덕구(0.078)	정성호(0.0097)	홍미영(0.107)

Closeness Centrality의 경우 위의 식과 같이 정의되며, 여기서 $\text{dist}(j, i)$ 는 vertex j 와 i 사이의 최단 거리를 의미한다. 공동발의 네트워크에서 Closeness Centrality는 어떤 국회의원이 다른 국회의원들과 법안 발의 관계에서 얼마나 가까운지를 나타내주며, Closeness Centrality가 높은 국회의원은 많은 국회의원들과 직접적으로 혹은 다른 국회의원을 사이에 두고 간접적으로 법안 발의 관계를 맺었다고 볼 수 있다.

- Betweenness Centrality

$$c_B(i) = \sum_{j \neq k \neq i \in V} \frac{\sigma(j, k | i)}{\sigma(j, k)}.$$

Betweenness Centrality의 경우 위의 식과 같이 정의되며, 여기서 $\sigma(j, k)$ 는 vertex j 와 k 사이의 최단 거리의 개수를, $\sigma(j, k | i)$ 는 j 와 k 사이의 최단 거리 중에서 i 를 지나는 최단 거리의 개수를 의미한다. 공동발의 네트워크에서 Between Centrality가 높다는 것은 해당 국회의원을 거치지 않으면 발의되지 않았을 법안이 많았다는 의미로, 법안발의 관계를 형성하는 과정에서 자신에 대한 타인의 의존성이 높다는 의미를 지닌다.

- Eigenvector Centrality

$$c_{Ei}(i) = \frac{\sum_{(i,j): y_{ij} \in \mathbf{Y}} c_{Ei}(j)}{\lambda}.$$

Eigenvector Centrality의 경우 위의 식과 같이 정의되며, c_{Ei} 는 인접행렬 \mathbf{A} 의 가장 큰 고유값 λ 에 연관된 고유벡터를 의미하며, 즉 $\mathbf{A}c_{Ei} = \lambda c_{Ei}$ 의 해를 의미한다. 공동발의 네트워크에서 어떤 국회의원의 Eigenvector Centrality가 높다는 것은 법안발의에 대한 영향력이 높은 것은 물론 해당 국회의원과의 관계를 맺는 다른 국회의원들도 그 영향력이 높다는 것을 의미한다.

Table 4.1의 결과는 Eigenvector Centrality를 제외하고는 선행연구인 Chang (2011)의 결과와 일치했다. 이러한 기술통계량들에 대한 Chang (2011)의 해석을 소개하자면 다음과 같다. Indegree Centrality와 Closeness Centrality에서 강점을 보였던 인물들은 민주노동당에 속한 최순영 의원을 제외하면 대부분 당시 여당이었던 열린우리당 소속이었으며, 이는 열린우리당 의원들이 그만큼 공동발의 지원을 평균적으로 많은 횟수로 받았다는 것을 의미한다. 특히, Incloseness에서 강점을 보였던 의원들은 많은 의원들로부터 법안 발의를 받는 최단 거리가 짧았기 때문에 공동발의 네트워크에서 높은 인기를 대변할 수 있다는 해석도 가능하며, Outcloseness에서 강점을 보였던 의원들은 다른 의원들에게 법안 발의를 하는 최단 거리가 짧았기 때문에 많은 사람들을 지원하면서 세력을 확장하려는 시도를 보였다고 해석할 수 있다. 또한, 가장 높은 통계량을 보여주었던 최순영 의원의 경우 민주노동당에서 발의한 법안에 대해 동일 당원들의 지지를 전폭적으로 받았다는 것을 알 수 있다. 하지만 Outdegree Centrality와 Betweenness Centrality에서 강점을 보였던 인물들은 열린우리당에 속했던 박상돈을 제외하고 대부분 한나라당

에 속했던 의원들로, 한나라당 의원들의 경우 평균적으로 더 많은 횟수의 공동발의 지원을 해주었으며, 자신들에 대한 타인의 의존성을 높이는 방법으로 국회 내의 영향력을 발휘하였다는 것을 알 수 있다. 한편, Outcloseness에서 높은 값을 가졌던 국회의원들은 다른 Centrality와는 다른 경향을 보였는데, 대부분 당시 야당과 제1여당에 속해있지 않고 소수 정당에 속해있던 의원들이었다. 이는 해당 의원들이 공동발의 지원을 해주었던 의원들도 비교적 많은 다른 의원들에게 공동발의 지원을 해주었다는 의미로 해석할 수 있다.

4.2. 네트워크 충분 통계량

공동발의 네트워크를 구성하는 공동발의 행위에 영향을 미칠 수 있는 네트워크 충분 통계량 $g(\mathbf{y})$ 들을 다음과 같이 선정할 수 있었다.

4.2.1. 상호 호혜성(mutuality) 가장 먼저 생각해볼 수 있는 효과는 네트워크의 구조적 효과 중 하나인 상호 호혜성(mutuality)이다. 공동발의 네트워크에서의 상호 호혜성이 유의하다는 것은, A라는 국회의원이 B라는 국회의원에게 공동발의 지원을 해준 경험이 있다면, B라는 국회의원도 반대로 A라는 국회의원에게 공동발의 지원을 해줄 가능성이 높다는 것을 의미한다. Valued Network에서의 상호 호혜성을 나타내는 네트워크 충분 통계량은 다음과 같다 (Krivitsky, 2012).

$$g(\mathbf{y}) = \sum_{i,j} \min(y_{ij}, y_{ji}).$$

4.2.2. 같은 정당 효과(party match) 다음으로 생각해볼 수 있는 효과는 공동발의 네트워크의 vertex에 해당하는 국회의원의 소속 정당의 효과이다. 공동발의 네트워크에서 같은 정당 효과가 유의하다는 것은, A라는 국회의원과 B라는 국회의원의 소속 정당이 같을 경우, 서로 공동발의 지원을 주고받을 가능성이 높다는 것을 의미한다. Valued Network에서 vertex의 covariate에 해당하는 같은 정당 효과를 나타내는 네트워크 충분 통계량은 다음과 같다는 것이 알려져 있다 (Krivitsky, 2012).

$$g(\mathbf{y}) = \sum_{i,j} y_{ij} I(i \text{와 } j \text{가 같은 정당 소속}).$$

4.2.3. 당선 횟수(number of election) 마지막으로 고려해볼 수 있는 효과는 국회의원의 당선 횟수에 대한 효과이다. 공동발의 네트워크에서 당선 횟수 효과가 유의하다는 것은 공동발의 지원을 주는 국회의원이나 받는 국회의원의 당선 횟수가 공동발의 지원 행위에 영향을 준다는 것을 의미한다. Valued Network에서 vertex의 covariate에 해당하는 당선 횟수를 나타내는 네트워크 충분 통계량은 다음과 같다는 것이 알려져 있다 (Krivitsky, 2012).

$$g(\mathbf{y}) = \left(\sum_{(i,j):y_{ij} \in \mathbf{Y}} n_i, \sum_{(i,j):y_{ij} \in \mathbf{Y}} n_j \right).$$

여기서, n_i 는 i 번째 국회의원의 당선 횟수를 의미한다.

4.2.4. 절편항과 시기별 네트워크 분석 위에서 언급한 모형에서 주 관심사가 되는 효과 외에도, 네트워크의 edge 관측값인 y_{ij} 를 통해 sparsity와 overdispersion을 보정해줄 필요가 있음을 알 수 있었다.

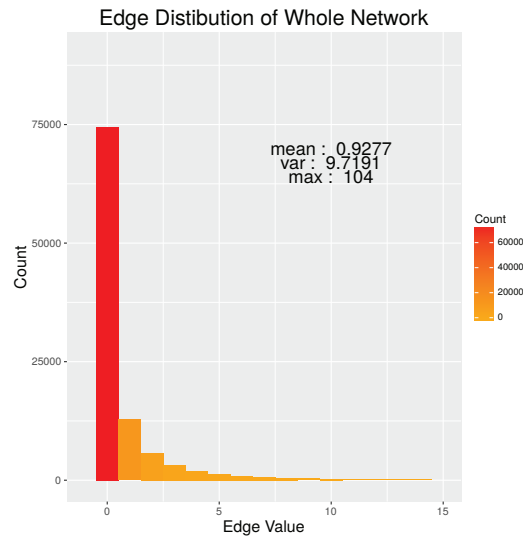


Figure 4.1. Histogram for y_{ij} of cosponsorship networks in the 17th National Assembly.

전체 공동발의 네트워크의 $y_{i,j}$ 값들의 분포를 나타낸 Figure 4.1을 살펴보면, 우선 0값의 출현빈도가 다른 값들에 비해 현저하게 높은 것으로 보아 이 네트워크가 매우 sparse한 네트워크이며, 일반적인 횃수형 자료를 모형화하는 포아송 회귀분석에서는 이런 상황에서 zero-inflated 포아송 분포를 이용하여야 하는 상황임을 알 수 있다. ERGM for Value Network에서도 이와 비슷한 방법으로 이를 위한 보정항을 네트워크 충분 통계량에 추가함으로써 해결할 수 있다 (Krivitsky, 2012). 이에 더하여, $y_{i,j}$ 의 평균(0.9277)에 비해 분산(9.7191)이 과도하게 높게 나타난 것으로 보아, 포아송 분포를 따른다는 가정에 무리가 있다고 볼 수 있었다.

이 문제를 해결하고 더 나은 분석을 하기 위한 대안으로, 공동발의 네트워크를 Chang (2011)에서 제안한 것과 같이 네 시기로 나누어서 살펴보았다. Chang (2011)에서는 17대 국회를 정국의 변화에 따라 다음과 같이 4개의 시기로 나누었다. 시기(1)은 2004년 6월부터 2005년 1월까지(237-251회기)로, 열린우리당이 4대 개혁입법(국가보안법, 사학법, 언론법, 과거사법)을 추진하던 시기였으며, 시기(2)는 2005년 2월부터 2006년 1월까지(252-257회기)로, 노무현 대통령과 열린우리당의 관계가 악화되었던 시기였다. 다음으로 시기(3)은 2006년 2월부터 2007년 1월까지(258-264회기)로 한미 FTA로 대표되는 신자유주의적 개혁의 시대라고 할 수 있으며, 마지막 시기(4)는 2007년 2월부터 2008년 5월까지(265-284회기)로 17대 대선과 18대 총선이 맞물려있던 17대 국회의 끝무렵이라고 볼 수 있다. 이러한 구분이 위와 같이 사전 연구인 Chang (2011)에서 정치학적 근거로 나누어진 구분이기 때문에, 이 논문에서 또한 같은 구분을 사용하여 살펴보았다.

전체 공동발의 네트워크를 위의 4개의 시기로 나눈 후 각 네트워크의 $y_{i,j}$ 값들의 분포를 나타낸 Figure 4.2를 살펴보면(왼쪽 위부터 오른쪽 아래로 1시기, 2시기, 3시기, 4시기를 나타낸다), 각 시기마다 평균과 분산이 매우 상이함을 알 수 있었다. 그 중에서 가장 안정적인 분산을 가진 시기는 1시기였으며, 나머지 시기들은 분산이 평균의 5배에서 최대 20배에까지 이르렀다. 이론적으로 지나치게 높은 분산을 가진 횃수형 변수의 경우에는 포아송 분포가 아닌 음이항분포(negative binomial distribution)를 이용해야 함이 맞으나, R의 패키지 `ergm`, `ergm.count`에는 현재 구현이 되어있지 않았다. 따라서, 본 연구에는 다소 제한적이지만 1시기의 네트워크에 대하여 overdispersion과 sparsity를 보정하여 포아송 분포를

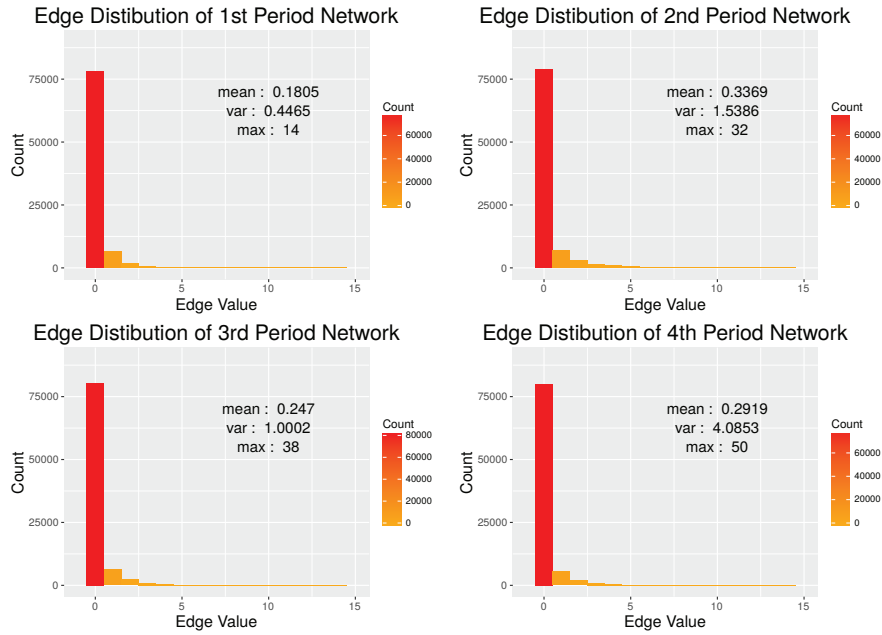


Figure 4.2. Histogram for y_{ij} of cosponsorship networks in the 17th National Assembly for each period.

이용한 ERGM을 적합시켜보기로 하였다.

모형의 Intercept에 해당하는 네트워크 충분 통계량은 다음과 같다는 것이 알려져 있다 (Krivitsky, 2012).

$$g(\mathbf{y}) = \sum_{i,j} y_{ij}.$$

다음으로 네트워크의 sparsity를 보정하는 네트워크 충분 통계량은 다음과 같으며 해당 통계량에 대응하는 모수가 양의 값으로 유의하다면 네트워크가 sparse한 상태임을 나타낸다는 것이 알려져 있다 (Krivitsky, 2012).

$$g(\mathbf{y}) = \sum_{(i,j):y_{ij} \in \mathbf{Y}} I(y_{ij} \neq 0).$$

마지막으로, edge 값들의 overdispersion을 보정하는 네트워크 충분 통계량은 다음과 같으며 해당 통계량에 대응하는 모수가 음의 값으로 유의하다면 네트워크의 분산이 과분산 상태임을 나타낸다는 것이 알려져 있다 (Krivitsky, 2012).

$$g(\mathbf{y}) = \sum_{(i,j):y_{ij} \in \mathbf{Y}} \sqrt{y_{ij}}.$$

4.3. 모형 설명

4.2절에서 소개한 네트워크 충분통계량 $g(\mathbf{y})$ 들의 유의성 검정과 추정을 위해 다음과 같은 5개의 모형을 설계하였다.

Table 4.2. Parameter estimates of Network Models

Model	Estimates (Std.Errors)				
	B	BM	BP	BN	BMPN
Nonzero	3.78 (0.11)	3.53 (0.10)	2.68 (0.12)	3.66 (0.12)	2.73 (0.11)
Sum	3.21 (0.01)	3.00 (0.04)	1.72 (0.06)	3.38 (0.05)	1.94 (0.05)
Overdispersion	-9.44 (0.15)	-9.14 (0.14)	-7.34 (0.16)	-9.24 (0.16)	-7.49 (0.16)
Mutuality		1.03 (0.02)			0.77 (0.02)
Party Match					
열린우리당			0.95 (0.02)		0.83 (0.02)
한나라당			0.87 (0.02)		0.80 (0.02)
민주노동당			1.68 (0.04)		1.26 (0.03)
기타			1.23 (0.04)		1.10 (0.04)
Number of Election					
indegree				-0.09 (0.01)	-0.04 (0.01)
outdegree				-0.07 (0.01)	-0.06 (0.01)
AIC	-104553	-106339	-109142	-104868	-111548

- 1) Model B : Baseline only
- 2) Model BM : Baseline with Mutuality
- 3) Model BP : Baseline with Party Match
- 4) Model BN : Baseline with Number of Election
- 5) Model BMPN : Baseline with all terms

위의 모형들은 순서대로 네트워크 보정을 위한 부분과 intercept를 포함하는 Baseline 만을 가지는 모형 B, Baseline에 상호 호혜성(mutuality)을 나타내는 네트워크 충분 통계량을 포함하는 모형 BM, Baseline에 같은 정당 효과(party match)를 나타내는 충분 통계량을 포함하는 모형 BP, Baseline에 당선 횟수(number of election)을 포함하는 모형 BN, 모든 효과를 고려한 모형 BMPN을 의미한다.

모형 B를 통해서 기본적으로 네트워크의 sparsity와 overdispersion의 정도를 살펴볼 수 있으며, 모형 BM, BP, BN을 통해서 세 가지의 관심 요인들이 단독으로 추가되었을 때의 효과를 알 수 있다. 마지막으로 모든 효과를 고려했을 때에는 모든 효과가 통제되었을 때는 단독으로 추가되었을 때에 비해서 어떤 변화를 보이는지를 살펴볼 수 있다. 최종적으로 5가지 모형의 로그 가능도를 Akaike's Information Criteria (AIC)를 통해 비교하여 자료에 대한 모형의 적합성도 비교한다.

4.4. 분석 결과

4.3절에서 세운 5가지의 모형에 대하여 포아송 분포를 이용한 ERGM모형을 적합한 결과는 Table 4.2과 같다. 관심있었던 요인인 상호 호혜성, 같은 정당 효과, 당선 횟수가 모두 양의 값으로 유의했으며, 5개의 모형 중에서 AIC 기준으로 자료를 가장 잘 설명하는 모형은 모든 요인이 포함된 모형 BMPN임을 알 수 있었다.

AIC 기준으로 가장 자료를 잘 설명했던 모형 BMPN을 기준으로 그 계수에 대한 해석은 다음과 같다.

가장 먼저 overdispersion 모수가 -7.49 로 유의하였기 때문에 히스토그램으로 살펴보았듯이 y_{ij} 들의 분포가 과분산 상태임을 알 수 있다.

상호 호혜성(mutuality)의 경우 0.77의 계수가 유의했으며, 이 계수의 해석은 다음과 같은 예시로 설명할 수 있다. 만약 현재시점까지 A라는 국회의원이 B라는 국회의원에게 3번의 법안발의 지원을 해주었고, B라는 국회의원이 A라는 국회의원에게 5번의 법안발의 지원을 해준 상태라면 A라는 국회의원이 B라는 국회의원에게 다음 법안을 지원해줄 확률은 교류가 전혀 없는 경우와 비교해서 $(5 - 3) \times \exp(0.77)$ 배 (약 4.32배) 만큼 증가한다는 뜻이다.

같은 정당 효과(party match)의 경우 당별로 열린우리당(0.83), 한나라당(0.80), 민주노동당(1.26)의 계수가 유의하였으며, 기타 당은 여러 당이 혼재하기에 그 해석에는 큰 의미가 없다. 이 계수의 해석은 법안 발의 지원을 주고 받는 의원이 같은 당일 경우에는 그렇지 않을 때보다 평균적인 법안 발의 지원 횟수가 열린우리당의 경우 $\exp(0.83)$ 배 (약 2.29배), 한나라당의 경우 $\exp(0.80)$ 배 (약 2.22배), 민주노동당의 경우 $\exp(1.26)$ 배 (약 3.53배) 정도 많았다는 것을 의미한다. 이 결과는 앞의 4.1절에서 소개한 Chang (2011)의 기술통계량 중 Indegree Centrality와 Incloseness Centrality에 대한 해석과도 어느정도 일맥상통한 결과임을 알 수 있다.

당선 횟수의 경우 법안을 지원받는 국회의원은 -0.04 , 법안을 지원해주는 국회의원은 -0.06 으로 유의했다. 이는 각각 당선 횟수가 1번 증가할 경우 평균적으로 법안을 지원받는 횟수는 $\exp(-0.04)$ 배 (약 0.96배), 법안을 지원하는 횟수는 $\exp(-0.06)$ 배 (약 0.94배) 정도로 매우 소폭 감소한다는 것을 의미한다. 이는 곧 당선 횟수가 높은 다선 의원일 수록 입법 행위 자체에 대한 활동이 소폭 감소하며, 이에 대한 원인은 다선 의원의 활발한 주요 당직 활동으로 유추해볼 수 있다.

전반적으로 국회에서 법안 발의의 지원이 이루어지는 매커니즘에 가장 큰 영향을 주는 효과는 지원받는 국회의원과 지원하는 국회의원이 같은 당 소속인지의 여부였다. 그 중에서도, 민주노동당의 경우가 당시 여당이었던 열린우리당과 제1야당이었던 한나라당보다 그 효과가 컸다는 것을 알 수 있다. 이는 국회에서 큰 비중을 차지한 당은 아니었지만 법안 발의에 있어서의 결속이 강했다는 점과, 노동자들을 위한 정당이었다는 점에서 당원들이 법안 발의에 있어서 이해관계가 확실히 일치했다는 점이 반영된 것으로 이해해볼 수 있다. 그 다음으로는 상호 호혜성이 크게 영향을 미쳤는데, 눈여겨보아야 할 것은 모형 BM과 모형 BMPN에서의 계수의 차이점이다. 모형 BM에서는 1.03이었던 상호 호혜성의 계수가 모형 BMPN에서 0.77로 줄어드는데 가장 큰 영향을 미쳤던 것은 같은 정당 효과가 추가된 점이라고 할 수 있다. 이는 국회의원들끼리의 상호 호혜성은 단독으로도 강하게 유의했지만 그 효과 중 약 25%는 해당 국회의원들이 같은 당이었기 때문에 비롯된 것으로 해석할 수 있다. 당선 횟수의 경우 유의한 감소 효과가 있었지만 그 효과의 크기가 매우 미미하였기에 큰 의미는 두기 힘들다고 할 수 있다.

5. 결론

본 논문에서는 대한민국의 17대 국회에서 발의한 법안에 대한 자료를 이용하여 국회의원들 사이의 공동 발의 네트워크를 정의하고, 이를 포아송 분포를 이용한 ERGM 모형을 통해 분석해 보았다. 네트워크의 Sparsity와 Overdispersion을 보정한 후 법안 발의에 있어서 국회의원들의 상호 호혜성, 같은 정당 효과, 당선 횟수가 미치는 영향에 대해 알아보았다.

분석에 앞서 포아송 분포를 사용하기 위하여 네트워크의 edge 값들의 분포를 살펴본 결과, 분산이 평균에 비해 상당히 큰 과분산 상태임을 알 수 있었다. 따라서, Chang (2011)의 연구를 바탕으로 17대 국회를 4개의 시기로 나누어 다시 한번 edge 값들의 분포를 확인하였으며, 그 결과 첫 번째 시기가 포아송 분포를 사용하기에 무리가 없다고 판단되어 첫 번째 시기에 대하여 분석을 진행하였다.

분석 결과 국회에서 공동발의가 발생할 때 가장 큰 영향을 주는 요인은 같은 정당 효과였다. 어떤 국회의원이 다른 국회의원이 발의한 법안을 지원할 때 같은 정당에 속한 국회의원이라면 그렇지 않을 때보다

평균 지원횟수가 최대 4배 이상 많음을 확인할 수 있었다. 다음으로는 상호 호혜성 효과 역시 두드러졌다. 이는 국회의원들끼리 서로 공동발의 지원을 주고받는 행위가 결코 우연이 아니며, 지원을 받은 만큼 지원을 해주려는 경향이 강하게 나타났음을 확인할 수 있었다. 당선 횟수의 경우 다선 의원일 수록 오히려 법안을 발의하여 지원을 받거나 다른 국회의원을 공동발의 행위로 지원해주는 경우가 유의하게 낮다는 결과를 보였지만, 그 효과의 크기가 매우 작았기 때문에 영향력이 크다고 보기는 힘들었다.

공동발의 네트워크를 ERGM을 이용하여 분석하는 데에 있어서, 과도한 분산 때문에 모든 시기에 대한 분석이 이루어지지 못했으며, 음이항 분포와 같은 대안을 사용할 수 없었던 것이 이 연구의 한계점이었다. 하지만 기존의 기술통계량 위주의 네트워크 분석에서 한걸음 더 나아가 직접적으로 가중치가 있는 네트워크의 생성 매커니즘에 대해서 분석하였다는 데에 의의가 있다. 또한, 그 분석 결과가 실제로 국회에서 발생하는 국회의원들의 공동발의 행태를 합리적으로 설명하였기 때문에 추후 더 다양한 Random Graph Model을 이용한 국회 연구에 대한 동기부여가 될 것으로 예상된다.

References

- Bastian, M., Heymann, S., and Jacomy, M. (2009). Gephi: an open source software for exploring and manipulating networks, *International AAAI Conference on Weblogs and Social Media*, **8**, 361–362.
- Chang, D. (2011). Policy networks in the 17th National Assembly of Korea: an analysis of co-sponsorship and friendship Ties, *Bulletin of the Korean Association of Party Studies*, **10**, 157–178.
- Fowler, J. H. (2006). Connecting the Congress: A study of cosponsorship networks. *Political Analysis*, 456–487.
- Hunter, D. R., Handcock, M. S., Butts, C. T., Goodreau, S. M., and Morris, M. (2008). ergm: A package to fit, simulate and diagnose exponential-family models for networks, *Journal of Statistical Software*, **24**, n1pha 54860.
- Jacomy, M., Venturini, T., Heymann, S. and Bastian, M. (2014). ForceAtlas2, a continuous graph layout algorithm for handy network visualization designed for the Gephi software, *PLoS ONE*, **9**, e98679.
- Krivitsky, P. N. (2012). Exponential-family random graph models for valued networks, *Electronic Journal of Statistics*, **6**, 1100–1128.
- Krivitsky, P. N. (2016). ergm.count: Fit, Simulate and Diagnose Exponential-Family Models for Networks with Count Edges, *The Statnet Project* <http://www.statnet.org>, R package version 3.2.2, <http://CRAN.R-project.org/package=ergm.count>.
- Lee, B. K. and Youm, Y. S. (2009). Identifying the structure of co-signing networks among the 17th Korean congressmen in the standing committee of health and welfare: By using p-net modeling, *Journal of Contemporary Society and Culture*, **29**, 33–60.

17대 국회의 공동법안발의에 관한 네트워크 분석

박찬무^a · 장원철^{a,1}

^a서울대학교 통계학과

(2017년 3월 8일 접수, 2017년 4월 16일 수정, 2017년 4월 16일 채택)

요약

본 논문에서는 대한민국 17대 국회의 공동발의 네트워크에 대하여 연구한다. 대한민국 국회에서 발의되는 법안은 대표발의자를 포함하여 10명의 공동발의자의 동의가 있어야 그 효력이 유의하다. 따라서, 공동발의라는 개념을 이용하여 공동발의자들로부터 대표발의자로 향하는 방향성 있는 네트워크를 구성할 수 있으며, 이 네트워크는 곧 국회 내의 국회의원들 간의 사회적 관계를 나타낸다고 볼 수 있다. 우리는 이 네트워크에 가중치가 있는 네트워크를 위한 지수 랜덤 그래프 모형을 적합하여 네트워크의 구성원리에 미치는 네트워크 구조적 성질의 영향과, 공변량의 효과에 대해서 알아보고자 한다. 분석 결과, 17대 국회의 공동발의 네트워크의 구성에 가장 큰 영향을 미치는 것은 같은 정당 효과였다. 상호 호혜성 역시 공동발의 네트워크 구성에 중요한 역할을 하였으며, 당선 횟수의 효과는 작지만 유의한 영향을 보였다.

주요용어: 중심도, 공동발의, 상호호혜성, 같은 정당, 지수 랜덤 그래프 모형

본 논문은 2013년도 정부(교육부)와 2014년도 정부(미래창조과학부)의 재원으로 한국연구재단 지원을 받아 수행된 기초연구사업임 (No. 2013R1A1A2010065, No. 2014R1A4A1007895).

¹교신저자: (08826) 서울특별시 관악구 관악로1, 서울대학교 통계학과. E-mail: wcjang@snu.ac.kr